SURF Final Report

THE DATA ANALYSIS FOR SHORT-TERM GRAVITATIONAL-WAVE BURST SIGNALS WITH A MODIFIED MAXIMUM LIKELIHOOD DETECTION METHOD

by

Bence Kocsis* **and Merse Előd Gáspár***

LIGO-T030213-00-D

Advisor: **Szabolcs Márka**

September 2003, LIGO Caltech

* Eötvös Loránd University of Sciences, Hungary

Contents

Abstract

The method of maximum likelihood detection is proposed for the search for short period gravitational-wave signals in the data of currently operating gravitational wave detectors as LIGO. The measurement of these signals with realistic sampling frequencies gives a small finite number of data points, allowing to make analytic probabilistic estimates on the efficiency of the various burst-search algorithms. Matched filtering is analyzed for white noise and is shown to be optimal in certain regards. A modification of this method is proposed to decrease the false alarm rate by matching the time differences and signal strengths between the different gravitational-wave detectors.

1 Introduction

The objective of this study is to give a theoretical review of the optimal detection methods regarding the detection efficiency for short period gravitational-wave signals. Short period gravitational wave signals are expected by the most interesting phenomena, such as supernova explosions [1], gamma-ray bursts [2], and black hole collisions [3]. The understanding of the detection results are vital to comprehend the limitations of the data analysis of the currently operating gravitational wave detectors.

Other gravitational wave sources include the periodic sources primarily from pulsars, the chirps from binary inspirals, and the stochastic background from the big bang. These sources are longer in duration allowing a long integration of the data stream, leading to the relative attenuation of noise. The stochastic background [4] is measured with the analysis of the cross-correlation between the separate gravitational wave detectors, whereas the periodic sources [5] are investigated by matched filtering . The evaluation is effective in Fourier space.

Short-range gravitational wave burst signals are analyzed by various different means. The physics behind the generation of these signals is generally not well understood, and the size of the family of possible signals is enormous. The second difficulty is that the length of the burst signals is typically too short and long integrations are not possible. For this reason, instead of matched filtering other techniques are used in practice. The algorithms use the correlation between the detectors and examine the statistics of the data distribution. Examples of these algorithms are the SLOPE [6]

and the TFCLUSTERS [7]. The SLOPE algorithm searches for instances where the smoothed amplitude increases rapidly, and TFCLUSTERS locates continuous enhanced regions in the time-frequency domain. The results are compared with the random noise statistics. There is also an analogous time domain method [8], where the integration length is varied in the calculation of the correlation integral between the detectors and the relative time shift is checked to reduce the false alarm rate. A fourth method [9] examines the change in the variance of the data, and creates a trigger if there is a significant increase. A fifth technique is the search for outlier spikes in the statistical distribution of the data samples. The results are used to make upper limits for the statistical presence of burst signals.

This study attempts to derive probabilistic calculations to examine what the optimally sensitive method is for the detection of distinct short period gravitational wave burst signals. The theoretical evaluations are capable to give an objective description of the detection efficiency [10]. In comparison, the evaluations of numerical simulations aiming to examine the efficiency requires much care. The generating of an unbiased sample regarding the detection is an important nontrivial task. Nevertheless, the results should be compared with Monte Carlo simulations for a check.

The remainder of this article is organized as follows. First the possible inferences are derived for a single detector with white Gaussian noise. The various optimal decision making criteria are described for a single data observation. The result is then generalized for longer data streams, and for multiple alternate hypotheses. The material included in these introductory sections are not original, but are included to aid the later calculations. We then describe a geometrical picture to visualize the problem geometrically. Next, we derive the optimal multiple detector method and extend to the case of arbitrary non-white Gaussian noise statistics. The limitations of the derived estimator are calculated in the following section. The resulting method is optimal in sensitivity in the mathematical sense, but can be further improved with the use of physical requirements. Our proposals for these are described in the next section. Finally, a possible template family is defined for burst searches. In the last section the conclusions are drawn and the future plans are listed.

2 Observations with a single detector

With the desire for completeness let us fist review the classic results of statistics for making inferences about a single measurement with one detector. This overview gives the opportunity to highlight what one can obtain from a short sampled data set. We shall then build on these fundamentals, and show in what context the proposed method is thought of as optimal.

This section is structured in four subsections. First we will discuss the results for single data observations, and review useful decision making criteria. Next we shall expand the results for multiple alternate hypothesis. Then we shall treat measurements of data as a function of time. Finally we shall explain what changes are necessary for a non-white noise background.

2.1 Single data observations

First we shall examine the detection of a binary signal in a noise background, where the $P(N)$ distribution of the noise is given. Let $Y(t)$ denote the variable measured, which assumes some value y at an instance t_0. Similarly for the signal and noise, S and N denote the variables, s and n their corresponding values. For simplicity we shall not make the distinction between the variables and their values when it doesn't lead to confusion. The measured signal is the sum of the true signal value and the random value assumed by noise.

$$y = s + n \tag{1}$$

Since N is a random variable, Y is also a random variable for each fixed s signal value denoted by the conditional probability distribution $P(Y \mid S = s)$. It is possible to explicitly calculate these distributions in practice. A Gaussian noise with

$$P(n) = \frac{1}{\sqrt{2\pi}} \exp(-n^2/2) \tag{2}$$

gives

$$P(y \mid s) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y-s)^2}{2}\right) \tag{3}$$

Let us assume we would like to decide between two possible outcomes of the experiment. The null hypothesis (H_0) is the event that no signal is present, i.e. $S = 0$, and the alternative hypothesis (H_1) is the event that $S = 1$. The

decision rule is to choose H_1 if

$$P(H_1 \,|\, y) \geq P(H_0 \,|\, y) \tag{4}$$

These probabilities are called *a posteriori probabilities*, so this is called the *maximum a posteriori probability criterion*. To make use of Eq. (3) we shall use the Bayes theorem for conditional probabilities

$$P(H_i \,|\, y) = \frac{P(y \,|\, H_i)P(H_i)}{P(y)} \tag{5}$$

Substituting in Eq. (4) and simplifying with the common denominators of the two sides, the decision rule becomes

$$P(y|S=1)P(S=1) > P(y|\, S=0)P(S=0) \tag{6}$$

or equivalently choose H_1 if

$$\frac{P(y|S=1)}{P(y|\, S=0)} \geq \frac{P(S=0)}{P(S=1)} \tag{7}$$

The probabilities on the RHS are referred to as *a priori* probabilities, and are often not known exactly. The LHS is called the likelihood ratio.

Generally many other criteria for goodness involves equations for the likelihood ratio as well, only the quantity on the RHS of Eq. (7) changes. Eq. (7) was the decision rule for the maximum posteriori criterion (4) and envolved the a priori probabilities. Common examples for criterions, that do not depend on a priori probabilities are the Neyman-Pearson criterion and the minimax criterion. We shall now briefly discuss these two criteria.

For the Neyman-Person and minimax criteria one needs to introduce conditional decision probabilities and cost functions. Let D_1 denote the event when the decision is made in favor of H_1, and D_0 the event for a decision for H_0. The sum $P(D_1 \,|\, H_1)P(H_1) + P(D_0 \,|\, H_0)P(H_0)$ is the probability that the correct decision was made. $P(D_1 \,|\, H_0)$ is referred to as the probability of false alarms. The quantities C_{ij} are assigned cost functions associated to the the D_i, H_j pairs[1]. The average risk or cost for the decision procedure is given by

$$\bar{C} = \sum_{ij} C_{ij} P(D_i \,|\, H_j)P(H_j) \tag{8}$$

[1]The diagonal elements C_{ii} correspond to the correct decisions and are usually assigned a zero value.

The objective of the Neyman-Pearson criterion is to maximize the probability of detection for a given probability of false alarm. In other words it maximizes $P(D_1 \mid H_1)$ subject to the constraint $P(D_1 \mid H_0) = \alpha$. It can be shown that this can accomplished by using a likelihood ratio test. Specifically, there exists some nonnegative number λ_0 such that hypothesis H_1 should be chosen if and only if

$$\lambda(y) = \frac{P(y|S=1)}{P(y|\,S=0)} \geq \lambda_0 \tag{9}$$

The threshold λ_0 is chosen to satisfy the false alarm probability constraint.

The last criterion we shall discuss is the minimax criterion. The purpose of the minimax criterion is to minimize the maximum possible $\bar{C}$ expected cost when the a priori probability of each hypothesis is unknown. Therefore the solution involves finding the least favorable a priori probability for which the average cost is the maximum. Substituting $P(H_1) = 1 - P(H_0)$ in Eq. (8) and differentiating with respect to $P(H_0)$ produces the condition for the minimax solution:

$$C_{10}P(D_1 \mid H_0) + C_{00}P(D_0 \mid H_0) = C_{01}P(D_0 \mid H_1) + C_{11}P(D_1 \mid H_1) \tag{10}$$

Let R_0 be the range of y values where D_0 decision is made and the complementary range, R_1, is where the decision is D_1. The decision probabilities can be written in the form

$$P(D_i \mid H_j) = \int_{R_i} P(y|H_j)\mathrm{d}y \tag{11}$$

Substituting this in (10) gives the equation for the R_0, R_1 intervals for the minimax criterion. Using the fact that both H_0, H_1 and D_0, D_1 are complementary respectively, we obtain an equation for the likelihood ratio.

All three of the above examples for decision criteria gave constraints on the likelihood ratio for a unique test of a stationary random process. The goodness of the test was always a monotonic function for the λ likelihood ratio, a higher λ value always satisfies the conditions for a lower λ. Therefore there is a one-to-one correspondance between the likelihood ratio and the confidence level of a given decision.

2.2 Multiple alternate hypotheses

For multiple alternate hypotheses we are dealing with a family of possible events: $H_0, H_1, \ldots, H_I$. The objective can be either to decide simply whether

there was any signal present in the data, or to decide which of the given hypotheses is most likely to describe the results of the measurement. For a simple trigger generation it is sufficient to do I different independent single alternate hypothesis tests for all pairs (H_0, H_i) for $i \in [1 \dots I]$. The a posteriori, Neyman-Pearson and minimax detection criteria is thus the same as before. If any of the tests of these pairs result in a non-D_0 decision, a trigger is generated marking the possibility of the presence of a possible signal.

The same line of argument as in the case of a single alternate hypothesis leads to the following maximum likelihood multiple hypotheses a posteriori decision rule. Choose the event H_i for which

$$P(y \mid H_i)P(H_i) = maximum \tag{12}$$

where $P(H_i)$ is the a priori probability of the event H_i. If all different events are *believed to* have equal a priori probabilities, then $P(H_i)$ is the same for all i. The decision rule is then choose the event H_i for which $P(y \mid H_i) = maximum$. For the case of white noise and H_i events corresponding to $S = s_i$, we get

$$\frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(y - s_i)^2}{2\sigma^2}\right] = max \tag{13}$$

or equivalently, s_i is most likely if among $i \in [1 \dots I]$

$$ys_i - \frac{1}{2}s_i^2 = max \tag{14}$$

This is usually referred to the as the *maximum likelihood detection*.

The value of an estimate is dependent on the observations. As new sets of measurements are taken, the numerical value of the estimate changes. Hence the estimator itself is a random variable. When constructing a detection algorithm for obtaining an estimator, it is desirable for the estimator to have the minimum variance. It can be shown that the minimum variance unbiased[2] estimator is unique [12]. Assume the likelihood function is written as $P(y \mid s)$ where s is the parameter for to be estimated. Let the estimate be denoted by $\hat{s}$. There exists a theoretical lower bound, called Cramer-Rao bound, that gives the minimum variance of an estimate [12]

$$\sigma_{\hat{s}}^2 \geq \frac{1}{E\left[\left(\frac{\partial \ln P(y \mid s)}{\partial \alpha}\right)^2\right]} \tag{15}$$

[2]If the mean of an estimate is equal to the true value it is said to be unbiased.

This sets an effective limit on the precision of an estimator.

2.3 Multiple-point measurements

Let us now consider the measurement of K subsequent data samples with one given detector. We shall first assume stationary white noise $P(n) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-n^2/2\sigma^2)$ and treat the more general case of an arbitrary nearly-stationary Gauss background later. Let the value of the separate data points of measurement, signal and noise values be denoted by y_k, s_k and n_k where $k \in [0, 1, \ldots, K]$. The null hypothesis, H_0, is $s_k = 0$ for all k, and the alternate hypothesis, H_1, is a given K element long sequence of values, $(s_k \,|\, k \in [1 \ldots K])$. Given that H_1 is true $y_k - s_k$ has the distribution of noise for every k. Since the probability of statistically independent events are the product of the individual probabilities, the likelihood ratio corresponding to (7) is given by

$$\lambda(\boldsymbol{y}) = \frac{P(y_1, y_2, \ldots, y_K \,|\, H_1)}{P(y_1, y_2, \ldots, y_K \,|\, H_0)} = \frac{p_1(\boldsymbol{y})}{p_0(\boldsymbol{y})} \tag{16}$$

where

$$p_1(\boldsymbol{y}) = \frac{1}{\sqrt{2\pi}\sigma}\prod_{k=1}^{K}\exp\left[-\frac{(y_k - s_k)^2}{2\sigma^2}\right] \tag{17}$$

and

$$p_0(\boldsymbol{y}) = \frac{1}{\sqrt{2\pi}\sigma}\prod_{k=1}^{K}\exp\left[-\frac{y_k^2}{2\sigma^2}\right] \tag{18}$$

After simplifications the likelihood ratio becomes

$$\lambda(\boldsymbol{y}) = \exp\left[\frac{1}{2\sigma^2}\left(\sum_{k=1}^{K} 2y_k s_k - \sum_{k=1}^{K} s_k^2\right)\right] \tag{19}$$

and the decision rule is: choose H_1 if $\lambda(\boldsymbol{y}) \geq \lambda_0$, with λ_0 set according to one of the criteria treated above. With some manipulations this is equivalent to choose H_1 precisely when

$$\ln\lambda(\boldsymbol{y}) = \boldsymbol{s}\cdot\boldsymbol{y} - \frac{1}{2}\boldsymbol{s}^2 \geq \lambda_0' \tag{20}$$

Note that the scalar products of these discretely sampled functions transform to integrals in the continuous limit. In that approximation the evaluation of

the cross correlation integral of the measured and expected signals is necessary.

To get an insight on the essence of this decision rule one can visualize the result for the case of $K = 3$. (The $K = 2$ case is plotted on Figure 1.) The given sampled signal function for H_1 is $\boldsymbol{s} = (s_1, s_2, s_3)$, one point in the three dimensional space. For H_0 the corresponding signal is the origin. The probability distribution of the measurement can either be distributed around the point $\boldsymbol{s}$ or the origin depending upon whether H_1 is true or H_0. A likelihood decision rule as (20) corresponds to choosing H_1 if the measured $\boldsymbol{y}$ value is in a given half-space. The limiting plane between the half-spaces has a normal vector parallel to the position vector $\boldsymbol{s}$. The position of the plane is given by the likelihood constraint λ_0'. By increasing λ_0', it is shifted in the direction of $\boldsymbol{s}$. For an arbitrary $K > 3$, the equation defines a hyperplane in the K dimensional space. The decision is made according to which side the measured $\boldsymbol{y}$ value takes.

The multiple hypothesis detection for more data points is handled similarly. The maximum likelihood estimate corresponding to a choice of equal a priori probabilities gives the equation

$$\ln \lambda(\boldsymbol{y}) = \max_i \left[\boldsymbol{s}_i \cdot \boldsymbol{y} - \frac{1}{2} \boldsymbol{s}_i^2 \right] \geq \lambda_0' \tag{21}$$

where the $\boldsymbol{s}_i \,|\, i \in [1..I]$ is the family of signal templates. The event H_* is chosen for the index, where the argument on the RHS is the highest.

This can also be visualized in the vector space geometrically (see Figure 1). Mark the points $\boldsymbol{s}_i$ in the vector space. For each point color the half-spaces of the single signal point case described above by (20). The region where some alternate hypothesis will be chosen is the union of these colored half-planes. The regions corresponding to a given $\boldsymbol{s}_j$ template is bounded by hyperplanes with normal vectors $s_j - s_i$ with $i \in [0, \ldots, I]$. Therefore, it is to be emphasized that a decision in favor of a given $\boldsymbol{s}_j$ point is really the choice for a given region of the space of functions. Any representative member of that region, i.e. a point in the K dimensional vector space, which is a discrete sampled function in time, is fitted with the same function $\boldsymbol{s}_j$, but could also have been fitted with signals corresponding to the other points in the region. Indeed to avoid a biased estimate, one has to use a random template set with a Monte Carlo method rather than a fixed template family. Let us also point out that the refinement of the template family is only efficient to the point

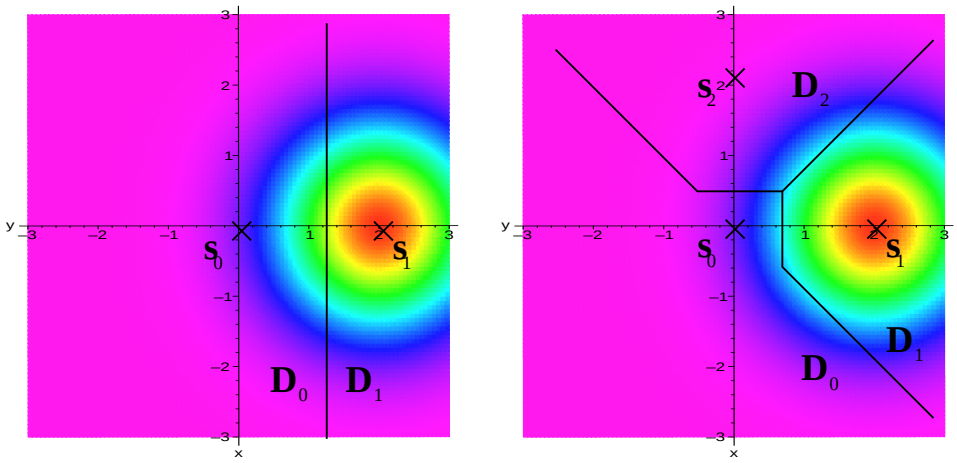

Figure 1: The plots depict the likelihood decision rules in the $K = 2$ dimensional case. The probability distribution of the measurement is plotted when the $\boldsymbol{s}_1$ signal is present in a Gaussian noise background. The boundaries separating the decisions D_0, D_1, and D_2 are marked by black lines. For the single alternate hypothesis *(left)*, the decision regions are half-spaces. The multiple alternate hypotheses acceptance domains *(right)* are bounded by hyperplanes, that can be approximated by chopped cones in the higher dimensional cases.

where the cell size reaches the scale which is comparable to the minimum variance of the estimator given above by (15).

The primary objective is to decide whether a given waveform is present in the data stream. For the marginal detection of the lowest strength signals, it is sufficient to work with signal templates of identical power. Therefore the signal templates $\boldsymbol{s}_i$ lie on a given spherical surface in the K-dimensional vector space. The radius of the sphere is

$$r = ||\boldsymbol{s}_i|| = (\text{the marginally detectable signal strength}) \tag{22}$$

The variable parameters of the method is q_K, the number of templates on the spherical surface, and m, the limit under which always D_0 null-decision is chosen. The adjustment of these parameters changes $P(D_0 \,|\, H_i)$, the false alarm, $P(D_i \,|\, H_j,\, i \neq j)$ the confusion, and $P(D_i \,|\, H_i)$, the correct detection probabilities. This is examined in Section 3 in detail.

Therefore, we have carried out the probabilistic considerations which led to the optimal detection method of signals. We have made strong use of

the fact that the noise was Gaussian and statistically independent. The relaxation of this condition will be discussed in the next chapter. Notice, that a method can be more sensitive than matched filtering only if there were additional information available for the detected signals. These could include multiple coincident detection of the gravitational waves, or a cooperation with other facilities, like the optical or neutrino detectors. In any case, the information available is well defined. For a given set of conditions, the optimal method and its limitations can be derived theoretically. After carrying out the corresponding calculations, there will later be no need for any improvements of the data analysis algorithms in terms of sensitivity.

2.4 Non-white Gaussian noise

Let us now take the more general case of non-white Gaussian noise[3]. In this case the distinct data samples in the data stream of the detector will be correlated, and the data points will not be statistically independent. This means the previous derivation fails, and the equations (17) and (18) do not hold.

For signals with Fourier transforms, eq. (20) will be valid with the redefinition of the scalar product (see e.g. [11]). It is common to use the following inner product in the literature

$$\langle g, h \rangle = 2 \int_{-\infty}^{\infty} \frac{\tilde{g}^*(f)\tilde{h}(f)}{S_n(f)} \mathrm{d}f \tag{23}$$

where tildes denote the Fourier transforms and $S_n(f)$, the one-sided noise power spectral density, is given by

$$\overline{\tilde{n}^*(f_1)\tilde{n}(f_2)} = \frac{1}{2}\delta(f_1 - f_2)S_n(f_1) \quad \text{for } f_1 > 0 \tag{24}$$

For our purposes we would like an algorithm fully in time domain. In this case it can be shown [11] the inner product is

$$\langle g, h \rangle = \int_0^T \mathrm{d}t \int_0^T \mathrm{d}\tau g(t) R_n^{-1}(t - \tau) r(\tau) \tag{25}$$

[3]Note, that for our purposes, for the detection of short-period burst signals, it is sufficient to assume the noise to be stationary, if the detection criteria is set to the actual noise level.

Where R_n^{-1} is defined by the equation

$$\int_0^T \mathrm{d}\tau R_n^{-1}(t-\tau)R_n(\tau-z) = \delta(t-z) \tag{26}$$

where $R_n(\tau)$ is the autocorrelation function of the noise. In particular, for white noise we have $R_n(\tau) = (N_0/2)\delta(\tau)$ and $R_n^{-1}(\tau) = (2/N_0)\delta(\tau)$, which when substituted to (25) gives back the original scalar product (20) up to a constant of proportionality.

The integrals used above are useful in the continuous case, however we are dealing with a discretely sampled finite data set. In this case one needs to use the discrete Karhunen-Loeve expansion [11], analogous to the discrete Fourier transforms. Let the length of observation be T. The procedure consists of 4 main steps.

- One first obtains the noise autocorrelation function $R_n(\tau)$ in the interval with the given sampling frequency F_s. This will effectively be $K = TF_s$ number of points.

- Second, the following eigenvalue problem of the R_n kernel needs to be solved. That is, the following equation should be solved

$$\mu_j f_j(t) = \int_0^T \mathrm{d}s f_j(s)R_n(t-s) \tag{27}$$

  for the eigenfunctions[4] f_j and eigenvalues μ_j. The integral is to be understood as a finite sum corresponding to the F_s sampling frequency. This is effectively a system of K linear equations. There are generally as many eigenvectors as the dimension of the linear space, i.e. K vectors.

- Third, the measured data sample and matched signals are expanded in terms of these eigenvectors.

$$r_k = \int_0^T \mathrm{d}t f_k(t)r(t) \tag{28}$$

  The integrals are again practically finite sums over the allowed points. The data points in this basis are statistically independent. The previous results are therefore valid in this representation.

[4]The eigenfunctions are to have unit norm.

- The likelihood functions are then

$$p(y \mid H_i) = \prod_{k=1}^{K} \frac{1}{\sqrt{2\pi\mu_k}} \exp\left[-\frac{(r_k - s_{ik})^2}{2\mu_k}\right] \tag{29}$$

  and the maximum likelihood estimator for identical a priori templates is the highest among these.

The main advantage of this algorithm is that it gives a statistically independent sample for a given noise statistics. Fourier space methods such as equation (23) assume that the Fourier modes of various frequencies are independent in the detector noise. This assumption holds only for ideally linear detectors with only narrowband noise sources.

The Fourier representation of (23) also breaks down numerically for band limited or notched data, where the domain of integration need to be chosen. Depending on how this was chosen the Fourier representation leads to correlated data sets, introducing further errors.

In conclusion, the algorithm derived in the previous sections define the optimal time domain one-detector algorithm. It uses all of the information available for the noise and data. With no prior preference for any signal template, it assigns the relative probability for each template. This is the relative probability distribution one believes the measured signal belongs to that choice of template, which is the most one can hope for using a single detector.

3 Limitations on the estimate

As we have pointed out previously, finite number of measurements of time samples cannot yield a definite waveform, but leads to a given result only with an associated probability distribution. Let us now examine the measure of this uncertainty.

3.1 Single alternate hypothesis

We shall work with a Gaussian distributed statistically independent noise. The measured time-sampled signal $\boldsymbol{y}$ is a K dimensional vector. Let the possible signal input be either $\boldsymbol{s}_1$ or $\boldsymbol{s}_0 = \boldsymbol{0}$. As shown earlier, it is sufficient to set the decision rule upon the value of the likelihood ratio $\lambda(\boldsymbol{y})$. We

shall choose the $\boldsymbol{s}_1$-present hypothesis if and only if $\lambda(\boldsymbol{y}) > \lambda_0$. Let $\boldsymbol{e}_1$ be the normal vector parallel to $\boldsymbol{s}_1$. The normal case letters shall denote the component in this direction. Let $\boldsymbol{m}$ be the measurement for which

$$\lambda(\boldsymbol{m}) = \lambda_0 \tag{30}$$

and $\boldsymbol{m} = m\boldsymbol{e}_1$. Visually, m is the magnitude of the signal from which we believe the signal to be present. For smaller $\boldsymbol{e}_1$ component of the measurement $\boldsymbol{y}$, the null-decision is made, and for a higher y component, an $\boldsymbol{s}_1$ signal is believed to be present.

The false alarm probability is given by

$$P(D_1 \,|\boldsymbol{s}_0) = \int\limits_{\lambda(\boldsymbol{y})>\lambda_0} \frac{\mathbf{d}\boldsymbol{y}}{(\sqrt{2\pi})^K} \exp\frac{-\boldsymbol{y}^2}{2} = \int\limits_{m}^{\infty} \frac{\mathrm{d}y}{\sqrt{2\pi}} \exp\frac{-y^2}{2} = F(-m) \tag{31}$$

where in the second equality we have changed to the variable y for which $\boldsymbol{y} = y\boldsymbol{e}_1$ and evaluated all other integrals[5]. The value of m is defined by defining the desired false alarm rate in this equation.

Let us derive the probability of an error of the first kind to be made. Assume that $\boldsymbol{s}_1$ is indeed present, but the decision misses the detection. Subtracting $\boldsymbol{s}_1$ from $\boldsymbol{y}$ leaves the noise with the given Gaussian distribution,

$$P(D_0 \,|\boldsymbol{s}_1) = \int\limits_{\lambda(\boldsymbol{y})<\lambda_0} \mathbf{d}\boldsymbol{y} \frac{1}{(\sqrt{2\pi})^K} \exp\left[-\frac{(\boldsymbol{y}-\boldsymbol{s}_1)^2}{2}\right] \tag{32}$$

Again, the integrals on the orthogonal compliment of $\boldsymbol{e}_1$ are trivial. What remains is

$$P(D_0 \,|\boldsymbol{s}_1) = \int\limits_{-\infty}^{m} \mathrm{d}y \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(y-s_1)^2}{2}\right] = F(-s_1+m) \tag{33}$$

The detection confidence for $\boldsymbol{s}_1$ is

$$P(D_1 \,|\boldsymbol{s}_1) = 1 - P(D_0 \,|\boldsymbol{s}_1) = F(s_1 - m) \tag{34}$$

Therefore a measurement of $\boldsymbol{y}$ gives false alarms with a probability $F(-m)$ and a correct detection with $F(s_1-m)$. The results (34) and (31) are plotted

[5]Note that y is not the norm of $\boldsymbol{y}$, just a projection in the K dimensional vector space.

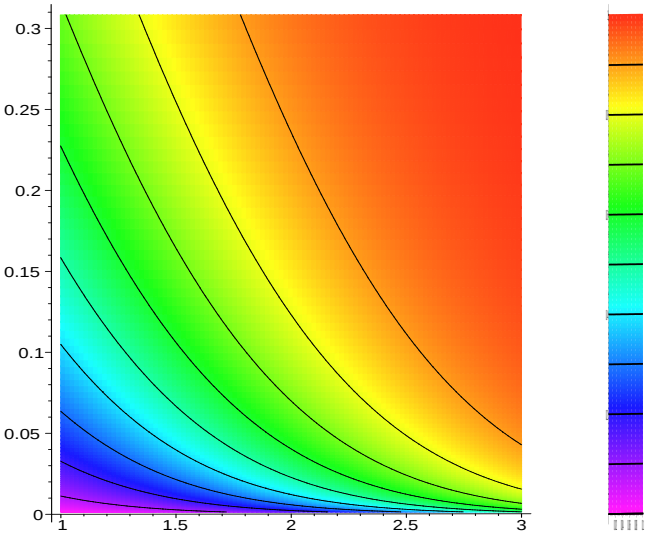

Figure 2: The single alternate hypothesis detection confidence is plotted as a function of the template signal strength $||\boldsymbol{s}_1||$ (x axis) and the false alarm rate (y axis). The noise variance is 1 for each data point, and the result is independent of K, the length of the time sample. Note however, that the total noise variance is $\sqrt{K}$, so the detection of a signal of unit order is a great achievement for large K values.

on Figure 2. Note the peculiarity, that since F is the unit variance Gaussian cumulative probability function, this gives a sensitivity scale of order 1, independent of the length of the analyzed data stream. There are two solutions for this peculiarity.

Firstly, for longer data streams the same signal has a greater norm, but the noise is equally enhanced, as shown on Figure 3. The increase of both the signal and noise strength is by a factor of $\sqrt{K}$. However the single component in the $\boldsymbol{e}_1$ direction does not change for noise. The detection of a unit strength signal in a larger vector space means the detection of the original signal of strength $1/\sqrt{K}$. Also note, that the unit strength signals of longer lengths contain more information, i.e. more details. Therefore, their detection is an improvement compared to the corresponding lower dimensional unit strength signal case.

The second notice is that a shorter sample fits more times in a data stream of length T. For a signal $\boldsymbol{s}_1^\sigma$ with a Gaussian envelope of variance σ, the detection of the corresponding signal with a relative time difference

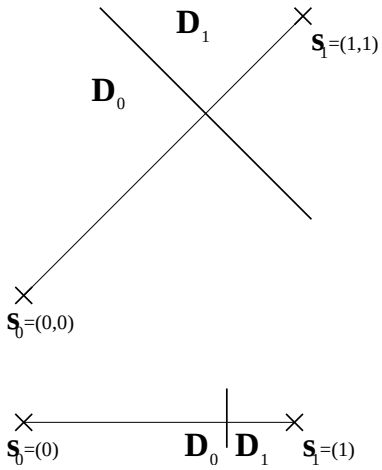

Figure 3: The visualization of the peculiarity for $K = 2$. The signal is $\boldsymbol{s}_1 = (1, 1)$ or $\boldsymbol{s}_0 = (0, 0)$, the measurement is distributed around either of these points with unit variance. The decision is made according to which half-plane the measurement takes. Compare this to the $K = 1$ dimensional problem, where $\boldsymbol{s}_1 = (1)$ and $\boldsymbol{s}_0 = (0)$. Intuitively, the two-point measurement shall be more effective than the one point measurement. It is indeed more effective, but only due to the fact, that it corresponds to a one-dimensional signal of length $\sqrt{2}$. However, the normalized signals' detections are identical.

of 3σ will be uncorrelated. The support of these functions are disjunct, and are therefore statistically independent. Thus, for a total detection time T, there are $n_\sigma = T/(3\sigma)$ independent time shifts of the same function. The false alarm rate for $\boldsymbol{s}_1^\sigma$ is enlarged by this factor. Therefore a longer signal can be detected with a better sensitivity for the same false alarm rate in a given data stream.

There is a third aspect to consider regarding the apparent insensitivity of the efficiency of the decision to the sample size. It is to realize, that in the above case we were only considering between the two possibilities $\boldsymbol{s}_0$ and $\boldsymbol{s}_1$, and one of these events was always chosen, even if the measured $\boldsymbol{y}$ was far away from both hypotheses. In the later case, we shall be aware that a very low probability data stream has been detected for the given signal. The inference, that some non-$\boldsymbol{s}_0$ signal was present, is valid, but the decision in the favor of $\boldsymbol{s}_1$ is questionable.

To avoid this problem, one could expand the size of the signal template set, and make inferences according to a multiple alternate hypotheses detection scheme such as the maximum likelihood detection. The nonzero signal templates have equal strength, and are evenly distributed on a spherical surface in the vector space. The amplitude of the detected signal can be determined later, separately.

3.2 Multiple alternate hypotheses

The maximum likelihood detection presumes the belief that each signal is equally probable. The decision for a given measurement $\boldsymbol{y}$ is made according to (21), the nearest signal $\boldsymbol{s}_* \in \{\boldsymbol{s}_i \,|\, i \in [1 \dots I]\}$. The signal templates $\boldsymbol{s}_i$ are chosen with identical strengths. Geometrically, they lie on the surface of a K-dimensional sphere of radius $r = ||\boldsymbol{s}_i||$. They need not cover the whole of the sphere, just the subspace spanned by the burst waveforms.

Compared to the single hypothesis case treated previously, this method reduces the false alarm rates $P(D_i \,|\, H_0)$ for each signal $\boldsymbol{s}_i$ by a significant amount. However the detection probabilities $P(D_i \,|\, H_i)$ are also decreased, due to the possible confusion between the alternate hypotheses. The confusion probability is defined as $P(D_i \,|\, H_j,\, i \neq j)$. In the remaining of this section we give the theoretical values for these uncertainties.

The separation between the signal templates is determined by the magnitude of the noise. Since the noise is assumed to have a unit normal distribution for each data point, it has a variance $\sqrt{K}$ around the true signal. This separates as a unit variance in the direction of $\boldsymbol{s}_1$ and $\sqrt{K-1}$ in the orthogonal complement. One has to choose templates with mutual separations of this order of magnitude. A finer resolution does not carry more information for the observation of only one event. For such an effective template family, the maximum number of templates of lowest strength, q_K, will be the number of points available on the $\sqrt{K}$ radius K-dimensional sphere having a $\sqrt{K-1}$ separation. In Appendix A. it is proved that a sufficient approximation is $q_K = 2K$.

We can now return to the question of detecting a given signal $\boldsymbol{s}_1$ of the signal family in $\sqrt{K}$ variance Gaussian noise. Presuming that this signal is present, the measured $\boldsymbol{y}$ sample will have a probability distribution with variance $\sqrt{K}$ around $\boldsymbol{s_1}$. The maximum likelihood decision criteria is to have $\boldsymbol{y}$ in the vicinity of this point in the K-dimensional vector space. The amplitude matching can be done separately, first we shall only determine the

most likely direction of the signal from the origin. Compared to the simple single-hypothesis decision rule discussed in the beginning of the section, this narrows the acceptance domain of D_1 from the half-space, to a chopped cone with axis along $\boldsymbol{s}_1$, peak at the origin $\boldsymbol{s}_0$, radius $\sqrt{K-1}$, and the top cutoff at $m\boldsymbol{e}_1$. A measurement, $\boldsymbol{y}$, at a larger angle from the origin, corresponds to a different template of the unit sphere, and the tip of the cone corresponds to the no signal case $\boldsymbol{s}_0$ (see Figure 1).

The variable parameters of the method are θ and m. θ is given by the density of the templates on the spherical surface and m is the limit under which the null hypothesis D_0 is chosen. The adjustment of these parameters changes the false alarm rate, the confusion between the alternate hypotheses, and the correct detection probability for a given signal. The following observations can be drawn from inspection:

$$(\text{Increasing } m) \Rightarrow \left(\begin{array}{c} \text{decreases false alarms} \\ \text{decreases detection probability} \\ \text{leaves confusion rate unchanged} \end{array} \right)$$

$$(\text{Increasing } \theta) \Rightarrow \left(\begin{array}{c} \text{leaves the total false alarm rate unchanged} \\ \text{leaves the total detection probability unchanged} \\ \text{increases the individual false alarm rates} \\ \text{increases the individual detection probabilities} \\ \text{decreases the confusion rate} \end{array} \right)$$

Notice that the increase of θ increases the sensitivity by every means. The extreme value $\theta = \pi$ corresponds to the case where only one signal direction $\boldsymbol{s}_1$ is distinguished, with a similar decision rule as discussed previously in 3.1.

Therefore, the main differences in the single and multiple alternate hypothesis tests can be summarized in the following points[6].

1. Detection probability of single alternate hypothesis is higher.
2. False alarm rate of single alternate hypothesis is higher.
3. Have to know *the one* signal prior to the measurement
4. Cannot distinguish more than one burst
5. Cannot be sure whether the measured signal is a true burst

[6]The signal strengths are assumed to equal in the two cases.

Appendix B derives the false alarm probability for the maximum likelihood method. The result is

$$P^{FA} = \frac{I}{q_K}\frac{\Gamma(K/2, m^2/2)}{\Gamma(K/2)} \approx \frac{I}{2K}F[\sqrt{2K} - \sqrt{2}m)] \tag{35}$$

where $\Gamma(a,b)$ is incomplete Gamma function, $\Gamma(a)$ is the complete Gamma function, and $F(a)$ is the cumulative Gaussian probability distribution. I denotes the number of templates associated to burst signals, $q_K = 2K$ for K sample size. The parameter m can be adjusted to reach the necessary maximum false alarm probability.

We also derived the detection probability for multiple alternate hypotheses in Appendix C. The nearly exact formula is

$$P = \frac{1}{2}\mathrm{erf}(\sqrt{K})F(s_1 - m) + \int_{m-s_1}^{\infty} \frac{\mathrm{d}y}{2\sqrt{2\pi}} \exp(-y^2/2)\mathrm{erf}(y + s_1 - \sqrt{K}) \tag{36}$$

and Appendix C also gives the simple analytic approximation for this function. The final result is plotted on Figure 4. Equation (36) shows that the detection probability depends strongly on s_1, the signal amplitude. The detection limit is $\propto \sqrt{K}$.

The limitations on the estimate are summarized by the equations (35) and (36). The adjustable free variables are the minimum signal strength s_1 and the detection margin m. Observe in (35) the dependance on the F Gaussian cumulative probability density function, which is negligible for negative arguments. This leads to the constraint that both s_1 and m have to scale as $\sqrt{K}$. The minimum detectable signal to noise ratio cannot be less than approximately 1, for as low false alarm rates with one detector. This is not surprising, since the detection rule was essentially to have the $\boldsymbol{s}_i$ component of the signal larger than all other components. The RMS uncertainty of a measurement is $\sqrt{K}$, which is therefore the minimum expected value for the detectable signal.

What one could alter in the method is the density of the template family. This practically means to group arrays of dimensions together and treat the resulting smaller set of data points. These data have higher strength (i.e. compared to noise). One can carry out the maximum likelihood method for the resulting set similarly. If the size of the data bins was N, this results in a noise attenuation of $1/\sqrt{N}$ and an effective sample size $K' = K/N$. The substitution in (35) and (36) is straightforward, which describe how the

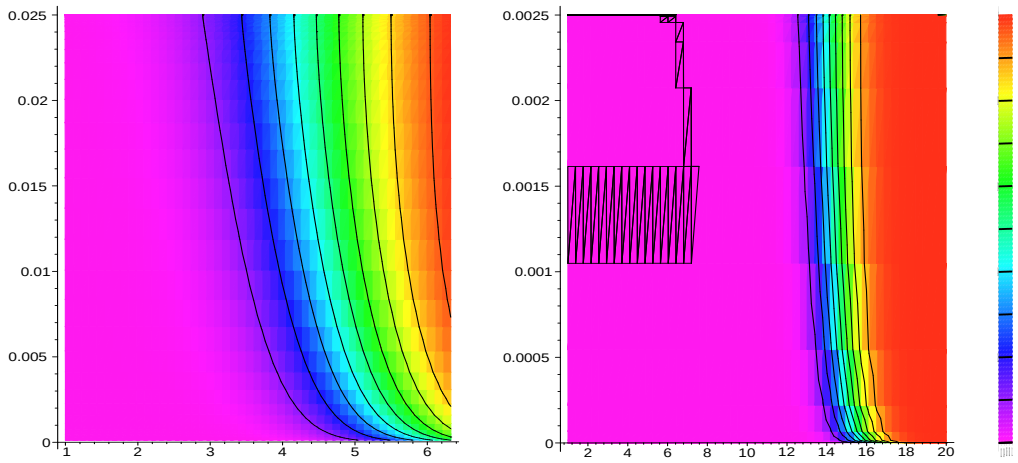

Figure 4: The multiple alternate hypothesis maximum likelihood detection confidence is plotted as a function of the template signal strength $||\boldsymbol{s}_1||$ (x axis) and the false alarm rate (y axis). The length of the data stream is $K = 20$ (*left*), and $K = 200$ (*right*). Observe that the detection is unlikely for $s < \sqrt{K}$, i.e. under 4.5 (left) or 14 (right).

resulting false alarm rate and true detection rate increase. The best detection rate can be obtained by setting the parameter N to the preferred value. Note, that the method for simple single alternate hypothesis case discussed in 3.1 is the case where there is only one bin of size $N = K$. The detection probability was of order $1/\sqrt{K}$, but the distinction between the alternate hypotheses was impossible.

So far in this section, we were considering the statistical error due to the superposition of random noise $\boldsymbol{n}$, but there is also a possibility for systematic errors. Observe, that we were dealing with the detection of a signal family $\{\boldsymbol{s}_i\}$, where the signals are chosen prior to the measurement. The detection probability given by (36) corresponds to an incoming signal among this family. What happens if a signal $\boldsymbol{x}$, not in this family arrives? The algorithm creates a burst trigger if the measurement is in the vicinity of one of the templates. By choosing an appropriately large number of templates, the burst space can be spanned by the templates. The burst signal $\boldsymbol{x}$ within the interior of this space will be attributed to one of the burst templates. The detection probability, that the measured incoming signal in noise, $\boldsymbol{x} + \boldsymbol{n}$, belongs to its appropriate cell in the vector space is generally less than the

detection probability (36). Nevertheless, the probability that *any* neighboring burst trigger is generated for $\boldsymbol{x}+\boldsymbol{n}$ is well approximated with the result (36). The quantity that is highly influenced by exactly which element of the burst space $\boldsymbol{x}$ assumes, is the confusion probability $P(D_j \,|\, H_i, i \neq j \neq 0)$. This question is revisited in Appendix B.

4 Truncation of the data stream

For gravitational-wave burst search, one is generally dealing with wave-forms, that are localized in the time domain, i.e. have a finite support in time. A gravitational wave burst signal, $\boldsymbol{s}_1$, with T data sample length, lies in a T dimensional subspace. Therefore, an optimal burst search method is also invariant for the projection on that subspace, and the vector components of a measurement along the orthogonal directions can be neglected. The truncation of the vector space is necessary to improve the algorithm performance.

4.1 Single alternate hypothesis

The truncation of the data stream is straightforward for the case of a single alternate hypothesis. The likelihood ratio, given by (20) can be evaluated within the significant subspace, where the signal is nonzero. However, since there is typically a smooth cutoff near the edges of burst signals, this truncation leads to errors. We shall now calculate the bias of the approximation, when the data points near those edges are neglected.

We shall calculate the error due to the truncation as compared to the original signal efficiency (34). Each single hypothesis case treated above, (i.e. the posteriori, the Neyman-Pearson, and the minimax criterion) have led to a rule (20), that depended only on the likelihood ratio. The event H_1 is chosen exactly if $\boldsymbol{e}_1 \cdot \boldsymbol{y} > m$, for a given m. This divided the vector space of the measurements in two half-spaces, with a dividing plane given by m. This plane was perpendicular to $\boldsymbol{s}_1$, and included the point $m\boldsymbol{e}_1$, where $\boldsymbol{e}_1$ was the normal basis vector parallel to $\boldsymbol{s}_1$. The detection confidence was given by

$$P(D_1 \,|\boldsymbol{s}_1) = \int_{\boldsymbol{e}_1\cdot\boldsymbol{y}>m} \frac{\mathbf{d}\boldsymbol{y}}{\sqrt{2\pi}^K} \exp\left[-\frac{(\boldsymbol{y}-\boldsymbol{s}_1)^2}{2}\right] \tag{37}$$

We shall calculate the deviation from this level. Let $\boldsymbol{s}_1' = \boldsymbol{e}_1' \cdot \boldsymbol{s}_1$ denote the truncated signal. The decision rule will be altered by substituting $\boldsymbol{s}_1'$ for

the original $\boldsymbol{s}_1$ signal: choose the event $S = s_1$ exactly when $\boldsymbol{e}_1' \cdot \boldsymbol{y} > m$. This effectively rotates the dividing plane that is between the decision half-planes, around the origin. The detection confidence is described by the same integral formula (37), only the integration domain changes. Let us switch to an integration variable $\boldsymbol{x} = \boldsymbol{y} - \boldsymbol{s}_1$.

$$P(D_1 \,|\boldsymbol{s}_1) = \int\limits_{\boldsymbol{e}_1' \cdot \boldsymbol{x} > m - s_1'} \frac{\mathbf{d}\boldsymbol{x}}{\sqrt{2\pi}^K} \exp\left[-\frac{\boldsymbol{x}^2}{2}\right] \tag{38}$$

The integrals over the subspace orthogonal to $\boldsymbol{e}_1'$ are separable and equal 1. Only the $\boldsymbol{e}_1'$ component remains. This will be denoted by x.

$$P(D_1 \,|\boldsymbol{s}_1) = \int_{m-s_1'}^{\infty} \frac{\mathrm{d}x}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right] = F(s_1' - m) \tag{39}$$

The false alarm rate only depends on the magnitude $\boldsymbol{y}^2$, and is independent of $\boldsymbol{s}_1$. This is therefore invariant under a rotation of the limiting plane, leading to an unchanged false alarm rate

$$P(D_1 \,|\boldsymbol{s}_0) = F(-m) \tag{40}$$

Comparing with (34) shows that the truncation of the data stream reduces the detection probability from $F(s_1 - m)$ to $F(s_1' - m)$, for the same false alarm rate $F(-m)$. This is not at all a surprise, since it is just the simple consequence of the essence of the method: the measurement is projected along the signal, and the components in the orthogonal subspace are neglected.

4.2 Multiple alternate hypotheses

We shall now examine the consequences of the truncation for the multiple alternate hypotheses detection, the maximum likelihood method. Section 2.3. has used the assumption that the signal templates were of identical length. If the signal length varies from template to template, the result for the likelihood ratio (21) needs revision. The changes in the derivation are straightforward, the relative probability of an event corresponding to the given signal should be calculated for different signal strengths. Let us recall the definition of the likelihood ratio from (16):

$$\lambda(\boldsymbol{y}) = \max_i \frac{P(\boldsymbol{y} \,|\, H_i)}{P(\boldsymbol{y} \,|\, H_0)} = \max_i \exp\left[-\frac{1}{2\sigma}(\boldsymbol{y} - \boldsymbol{s}_i)^2 + \frac{1}{2\sigma}\boldsymbol{y}^2\right] \tag{41}$$

Making a decomposition in the truncated and orthogonal subspaces yields

$$2\sigma \ln \lambda(\boldsymbol{y}) = \max_i \left(\boldsymbol{y'} \cdot \boldsymbol{s'_i} - \frac{1}{2}\boldsymbol{s'_i}^2 + \boldsymbol{y}^\perp \cdot \boldsymbol{s}_i^\perp - \frac{1}{2}{\boldsymbol{s}_i^\perp}^2 \right) \tag{42}$$

The ${\boldsymbol{s}^\perp}^2$ term is to second order small, which can be neglected. Using the Cauchy-Schwarz identity $-||\boldsymbol{a}||\,||\boldsymbol{b}|| \leq \boldsymbol{a} \cdot \boldsymbol{b} \leq ||\boldsymbol{a}||\,||\boldsymbol{b}||$ gives two bounds on the RHS.

$$2\sigma \ln \lambda(\boldsymbol{y}) \approx \max_i \left(\boldsymbol{y'} \cdot \boldsymbol{s'_i} - \frac{1}{2}\boldsymbol{s'_i}^2 \pm ||\boldsymbol{y}^\perp||\,||\boldsymbol{s}_i^\perp|| \right) \tag{43}$$

If the result depends on the choice of the signs $\pm$ for given indexes, then the truncation approximation cannot distinguish between those hypotheses. The term $||\boldsymbol{y}^\perp||$ can be approximated by the RMS value of the noise in the orthogonal direction, $||\boldsymbol{y}^\perp|| \approx \sqrt{K-N}$. To have the last term in (43) small means

$$\frac{||\boldsymbol{s}_i^\perp||}{||\boldsymbol{s'_i}||} \ll \frac{1}{\sqrt{K-N}} \tag{44}$$

Therefore relative squared integral of the signal template on the neglected domain has to be small as compared to $1/(K-N)$. With this criterion the likelihood function becomes

$$2\sigma \ln \lambda(\boldsymbol{y}) \approx \max_i \left(\boldsymbol{y'} \cdot \boldsymbol{s'_i} - \frac{1}{2}\boldsymbol{s'_i}^2 \right) \tag{45}$$

The signal template $\boldsymbol{s}_i$ should be chosen iff the the argument on the RHS is maximum for that index, and the likelihood function exceeds the limit λ_0.

The calculation time is given by the evaluation of each $\boldsymbol{y'} \cdot \boldsymbol{s'_i}$ correlation sum for the template family, each of which calculated on the corresponding truncated time interval. This is considerably less then the time needed for the evaluation for sums over the full data stream.

The error caused by the approximation can be estimated according to (45). The effective signal strengths decrease to s'. The false alarm and detection confidence probabilities change with the substitution of s' for s in (35) and (36). The effective signal to noise ratio will be less by a factor of order s'_1/s_1.

5 Multiple detectors

With the three LIGO detectors and the other gravitational-wave detectors built around the world, we are fortunate to have more than one measurement

for a given gravitational wave reaching Earth. Using a coincident analysis of the multiple detectors effectively can improve the detection sensitivity with orders of magnitude. We shall now examine the case with J independent detectors.

We have used several subtle approximations for our algorithm. Firstly, we shall only take into account the LIGO network, consisting of three nearly aligned gravitational-wave detectors. The deviation from exact alignment is taken negligible, so the detectors practically measure the same polarization of the gravitational waves. Secondly, the detectors are generally described by different transfer characteristics and bias, resulting in dissimilar output even for no noise and identical input. For the sake of the theoretical considerations these differences are neglected. These two approximations are not crucial for the proposed method, their relaxation is straightforward to carry out. Third, we shall neglect the possible correlations in the noises present in the different detectors. For a better model, one should apply a transformation similar to the Karhunen-Loeve expansion described above, thereby obtaining a statistically independent set of data samples. Forth, the noise is assumed to be Gaussian and stationary. The preprocessed signal of the LIGO detectors are indeed highly Gaussian, and are stationary on the timescale of seconds. Fifth, the noise is assumed to be white. The preprocessing of the detector output generally produces a bandlimited nearly white noise, with notched single frequencies. The white noise approximation provides a rough sense of the efficiency of the method. Future improvements should dissolve this problem by the use of the Karhunen-Loeve expansion.

We shall use the same probabilistic description as in the single detector case. Since the detectors are assumed to be totally independent, the probability of a given observation is just the product of the probabilities of the single detectors. The likelihood functions are therefore

$$\lambda(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_J) = \max_{i \in [1,\ldots,I]} \prod_{j=1}^{J} \frac{P(\boldsymbol{y}_j \mid H_i)}{P(\boldsymbol{y}_j \mid H_0)} = \max_i \prod_{j=1}^{J} \frac{p_i(\boldsymbol{y}_j)}{p_0(\boldsymbol{y}_j)} \tag{46}$$

where

$$p_i(\boldsymbol{y}_j) = \frac{1}{\sqrt{2\pi}\sigma_j} \prod_{k=0}^{K} \exp\left[-\frac{(y_{jk} - s_{ik})^2}{2\sigma_j^2}\right] \tag{47}$$

and

$$p_0(\boldsymbol{y}_j) = \frac{1}{\sqrt{2\pi}\sigma_j} \prod_{k=0}^{K} \exp\left[-\frac{y_{jk}^2}{2\sigma_j^2}\right] \tag{48}$$

Note, that equation (47) can be written as

$$p_i(\boldsymbol{y}_j) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left[-\frac{1}{2\sigma_j^2}\sum_{k=0}^{K}(y_{jk} - s_{ik})^2\right] \tag{49}$$

Substituting in (46) gives the likelihood ratio for multiple detectors.

$$\ln\lambda(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_J) = \max_i \sum_{j=0}^{J}\sum_{k=0}^{K}(y_{jk} - \frac{1}{2}s_{ik})\frac{s_{ik}}{\sigma_j^2} \tag{50}$$

Let $*$ denote the index for which the argument on the RHS is maximal, indicating that $\boldsymbol{s}_*$ is the signal closest to most $\boldsymbol{y}_j$. Abbreviating the k-sum with the scalar product, we get

$$\ln\lambda(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_J) = \sum_{j=0}^{J}\frac{1}{\sigma_j^2}(\boldsymbol{y}_j \cdot \boldsymbol{s}_* - \frac{1}{2}\boldsymbol{s}_*^2) \tag{51}$$

This equation can be simplified in the case of identical detectors, when $\sigma_j = \sigma$ for all j.

$$\ln\lambda(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_J) = \frac{1}{\sigma^2}\left[\left(\sum_{j=0}^{J}\boldsymbol{y}_j\right) \cdot \boldsymbol{s}_* - \frac{J}{2}\boldsymbol{s}_*^2\right] \tag{52}$$

The log likelihood function therefore only depends on the sum of the individual detectors' measurements, and is otherwise independent of the individual $\boldsymbol{y}_j$ values. The use of multiple detectors improves the detection efficiency by the increased number of data points. Instead of N data points, there are JN data points for J detectors. The results of Section 3 can be substituted directly for this number of dimensions. Equation (46) and (50) show that the detection with multiple detectors is analogous to the detection with an increased sampling rate. The detection sensitivity increases with a factor of $\sqrt{J}$.

6 False alarm rejection methods

The previous chapters defined the optimal mathematical algorithm for detecting a signal in noise. The detection was limited by the probabilistic nature of the noise, and the derived methods built on the exact form of the

probability distribution function, the normal distribution. The algorithm was optimal if the expected signal waveforms were assumed known and the detectors were assumed to have stationary noise.

An interesting result of the likelihood detection is that it only depends on $\sum_j \boldsymbol{y}_j$ the sum of the measurements in the individual detectors. The measurement, when only one detector provides all of the contribution to the likelihood function, is treated equally with the measurement when each detector see an equal amount, $1/J$ of the same signal. The former measurements can be vetoed on intuitive grounds. The elimination of these measurements, decrease the false alarm rate.

Two measurable quantities are the most likely time positions and signal strengths of the signal $\boldsymbol{s}_1$ in the individual detectors. The relative time shift check is important when the synchronization of the detectors is not available, i.e. the a priori direction of the signal is unknown. The elimination of the unphysical decisions can be the next step of the burst search algorithm.

6.1 Relative time shift check

Assume that the multiple decision rules (20) and (21) were evaluated for signals $\boldsymbol{s}_i$ with all of their time-shifted versions $L_t\,\boldsymbol{s}_i$, and a nonzero decision D_1 was made for a signal $\boldsymbol{s}_1$. The variable t is a continuous index in theory, but is restricted to discrete values according to the sampling frequency of the measurement. Next, we evaluate the test separately for each detector for the $L_t\,\boldsymbol{s}_1$ signals to obtain the most likely estimate of the arrival times $\hat{t}_j$, $(j \in \{1 \dots J\})$ in the various detectors. This is obtained with probability distributions

$$P(\hat{t}_j \,|\, \boldsymbol{s}_1, D_1) \tag{53}$$

This can be calculated theoretically for the given $\boldsymbol{s}$ signal waveform, F_s sampling frequency, $P(n)$ noise statistics for the maximum likelihood detection. Let us define the relative time shifts $\hat{\tau}_{ij} = \hat{t}_i - \hat{t}_j$. Therefore each $\hat{\tau}_{ij}$ is a random variable. Let $\Delta\hat{\tau}$ denote its variance.

There are two possible kinds of measurements that can be accomplished with gravitational wave detectors:

1. The detection in cooperation with other facilities, such as optical and neutrino detectors.

2. The detection only in the gravitational wave channel.

In the first case, there is an extra information for the gravitational waves, their exact direction is known. Therefore the relative time shifts between the gravitational wave detectors is calculable. By preforming a shift of the data streams with these magnitudes, the detectors can be synchronized. The synchronized signal templates have

$$t_i = t_j \tag{54}$$

for each pair of detectors. In Section 5, this is exactly what was assumed for an approximation, the individual probabilities were calculated with the same $\boldsymbol{s}_*$ hypothesis. The multiple detector maximum likelihood detection rule (52) already uses this information, and no fundamental improvements should be awaited for with the relative time shift check.

In the second case there is a weaker constraint: the relative time shifts have to be below a theoretical limit $\tau^{\max}$ for each detector pair.

$$\tau_{ij}^{\max} = \frac{1}{c} d_{ij} \tag{55}$$

where d_{ij} is the spacial separation of the detectors and c is the speed of light.

Let us approximate the $\hat{\tau}_{ij}$ probability distribution with a normal distribution of variance

$$\sigma_{ij} = \begin{cases} \Delta\hat{\tau}_{ij} & \text{cooperation with other channels} \\ \Delta\hat{\tau}_{ij} + \tau^{\max} & \text{only gravity wave detectors} \end{cases} \tag{56}$$

in the two cases respectively. The probability distribution is therefore

$$P(\hat{\tau}_{ij} \,|\, \boldsymbol{s}_1, D_1) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\sum_{\text{different pairs}} -\frac{\hat{\tau}_{ij}^2}{2\sigma_{ij}^2} \right] \tag{57}$$

The index $q \in \{1 \dots Q\}$ will be used for the detector pairs, and $\boldsymbol{\tau} = \{\hat{\tau}_q \,|\, q \in \{1 \dots Q\}\}$ to abbreviate notation, where $Q = \frac{1}{2} J(J-1)$ for J detectors. For the sake of this estimate we will assume $\sigma = \sigma_{ij}$ for all pairs.

$$P(\hat{\boldsymbol{\tau}} \,|\, \boldsymbol{s}_1, D_1) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2} \hat{\boldsymbol{\tau}}^2 \right] \tag{58}$$

The uncertainty in $\hat{\boldsymbol{\tau}}$ is

$$\tau_{RMS} = \sqrt{\Delta\hat{\boldsymbol{\tau}}^2} = \sqrt{Q}\,\sigma \tag{59}$$

The probability distribution when no signal is present (i.e. $\boldsymbol{S} = \boldsymbol{s}_0$) is approximated by a uniform distribution for the arrival time estimators, $\hat{t}_j$, of the signal.

$$P(\hat{\boldsymbol{t}} \,|\, \boldsymbol{s}_0, D_1) \approx \frac{1}{T^Q} \tag{60}$$

where T is the smallest of the following two parameters: the observation time and the typical cutoff scale of the $\hat{\tau}_{ij}$ distributions.[7]

The likelihood function can be associated to the given $\hat{\boldsymbol{t}}$ estimators:

$$\lambda(\hat{\boldsymbol{t}}) = \frac{P(\hat{\boldsymbol{\tau}} \,|\, \boldsymbol{s}_1, D_1)}{P(\hat{\boldsymbol{t}} \,|\, \boldsymbol{s}_0, D_1)} \geq \lambda_0 \tag{61}$$

where $\hat{\boldsymbol{\tau}}$ is calculable for a given $\hat{\boldsymbol{t}}$ using $\hat{\tau}_{ij} = \hat{t}_j - \hat{t}_i$. The value of λ_0 can be chosen to eliminate the triggers which are most likely false alarms.

For a measurement of length T with J detectors, the detections with $||\hat{\boldsymbol{t}}|| > \tau_0$ can be rejected. ($\tau_0 = 3\tau_{RMS}$ is a reasonable choice.) This is a simple decision rule, for which the false alarm rate and decision probabilities can be easily calculated.

If there were no signal present (i.e. $\boldsymbol{s}_0$), the false alarm probability is obtained from the uniform distribution. It is proportional to the measure of the region $D^\tau{}_1$ in the parameter space:

$$P^\tau_{FA} = P(D^\tau_1 \,|\, \boldsymbol{s}_0, D_1) = \sqrt{J} S_{J-1} \frac{\tau_0^{J-1}}{T^{J-1}} \tag{62}$$

where D^τ_1 is the event when the decision D_1 is not vetoed after the test, and S_{J-1} is the volume of the $J-1$ dimensional sphere, given by (83). The indices FA and τ refer to the false alarm probability with the time shift test. Explicitly for $J = 2$ gravitational wave detectors we get $P^\tau_{FA} = 2\sqrt{2}(\tau_0/T)$ or $J = 3$ detectors yields $P^\tau_{FA} = \sqrt{2}\pi(\tau_0/T)^2$.

The detection confidence is calculable for the distribution given by (58).

$$P^\tau_{TD} = P(D^\tau_1 \,|\, \boldsymbol{s}_1, D_1) = 2F\left(\frac{\tau_0}{\sqrt{Q}\sigma}\right) - 1 \tag{63}$$

where F is the cumulative unit Gaussian distribution function and the indices TD and τ refer to the true detection with the time shift test.

[7] It is expected, that for low signal strengths this distribution will indeed be flat for a scale $T >> \hat{\tau}_{RMS}$, where $\hat{\tau}_{RMS}$ is the variance when the signal $\boldsymbol{s}_1$ is there. If this condition does not hold, the time shift check will not be able to improve the detection method.

Equations (62) and (63) are the probabilities of the false alarms and good detections. It is important that this test is executed after the original detection rule, it is used after a given trigger was generated. The final false alarm probability is thus the original multiplied by this value $P_{FA}P_{FA}^{\tau}$, and similarly the true detection probability $P_{TD}P_{TD}^{\tau}$. Therefore both the false alarms and true detections are decreased by this method. Nevertheless, equations (62) and (63) show that the false alarms can be decreased by a huge factor, whereas the good detections stay nearly the same. The only condition is that the total observation time has to be large compared to the uncertainty: $T \gg \tau_0$ and $\tau_0 \gg \tau_{RMS}$.

6.2 Relative signal strength check

The second physical constraint can be to reject the triggers with inconsistent signal strength estimates over the various detectors. Recall that the original methods derived in this study are only sensitive to the sum $\sum_i \boldsymbol{y}_i$ of the individual measurements, but are otherwise insensitive to the distribution among the detectors. Also note, that the derived methods depended only on the waveforms, but not on the value of the signal strength, provided that it crossed a given threshold. Introducing this extra constraint eliminates the alarms which have large signal strengths in only a negligible number of detectors.

Our conjecture is that from this additional requirement, the corrections in the decisions will be insignificant with a low number of detectors. The exact calculations are left to future studies.

7 Burst search templates

In Appendix A it is shown that the effective template family consists of orthogonal signal templates, if the number of templates is greater than 7. Additional templates do not carry significant additional information, and cannot increase the sensitivity. Two methods were obtained from the probabilistic considerations and were compared in Section 3.2. The first method was to decide between the null and a single alternate hypothesis only and the second method was to decide between multiple alternate hypotheses. The later method was shown to be more reliable in certain regards, but needed a much larger signal to noise ratio ($s/n \approx 1/\sqrt{J}$) than the single hypothesis

test ($s/n \approx 1/\sqrt{JN}$).[8] For the multiple hypothesis method, the construction of a template family is necessary.

Let B denote the region of the K-dimensional space, which is the union of all waveforms called bursts. The acceptance domain of a burst template $\boldsymbol{s}_* \in B$ was a chopped cone in the K-dimensional vector space of the measurement (see Figure 1). Let this region be denoted by S_*. Any element of S_* has the largest component in the $\boldsymbol{s}_*$ direction in any orthonormal coordinate system. The template family is effectively is a choice of basis[9] $\{\boldsymbol{s}_i \,|\, i \in \{1 \dots I\}\}$ for which

$$\bigcup_{i \in \{1 \dots I\}} S_i \subset B \tag{64}$$

where S_i is the region attributed to $\boldsymbol{s}_i$. An effective template family is a choice of templates which is minimal and "spans" B, i.e.

$$\biguplus_{i \in \{1 \dots I\}} S_i = B \tag{65}$$

All effective template families are identical in terms of sensitivity and performance, therefore any given choice is sufficient to use.

Let us use the definition, that a burst waveform is a discrete sampled Gabor function, which is within the sensitive bands of the LIGO detectors. Thus a burst signal can be written as

$$s_k(A_1, A_2, f, \sigma, t_0) = [A_1 \sin(2\pi f t_k) + A_2 \cos(2\pi f t_k)] \exp\left[-\frac{t_k^2}{2\sigma^2}\right] \tag{66}$$

[8] Recall that J was the number of detectors and N was the effective data length of the signal.

[9] With this definition, the linear combination of bursts are not necessarily bursts. What one burst basis (template) "spans" is a chopped tube, and the "span" of several burst bases is the union of those solids in the K-dimensional vector space.

where the parameters are subject to the following constraints

$$t_k = t_0 + \frac{1}{F_s}k \tag{67}$$

$$k = \{1, 2, \ldots, K\} \tag{68}$$

$$0 < t_0 < T \tag{69}$$

$$||\boldsymbol{s}|| = 1 \quad \Leftrightarrow \quad A_1^2 + A_2^2 \approx 1/\sqrt{2\pi\sigma^2} \tag{70}$$

$$100 = f_0 < \frac{f}{\mathrm{Hz}} < f_1 = 1000 \tag{71}$$

$$0.01 = \sigma_0 < \frac{\sigma}{\mathrm{s}} < \sigma_1 = 1 \tag{72}$$

The effective template family should be minimal and orthonormal. We will construct a signal family with a small nonzero overlap in the continuous limit, which will become infinitesimal after the truncation of the data stream. Note that one can apply the Gauss Jordan elimination to obtain a truly orthonormal base.

In the parameter family $\{A_1, t_0, \sigma, f\}$ let us observe how many parameters need to be varied. First generate all possible t_0 time shifts, for a signal with a given f and σ. The number of different time shifts will be

$$n_{t_0} \approx \frac{T}{6\sigma} \tag{73}$$

Next we can vary the f frequencies with a linear resolution between the bounds f_0 and f_1.[10] The number of different f values is

$$n_\sigma = \frac{f_1}{f_0} \tag{74}$$

For each of these signals σ can be varied, with a step size increasing as a geometrical series. This way the scalar product (i.e. cross correlation) of the neighboring functions in the σ series will be equal. The number of different σ parameter values is

$$n_\sigma = \frac{\ln(\sigma_1/\sigma_0)}{\ln 6} \tag{75}$$

Finally the A_1 value can be given a binary value, to have the signal norm fixed. If $A_1 = 0$ then A_2 should be assigned the appropriate value with (70), otherwise $A_2 = 0$. Therefore

$$n_A = 2 \tag{76}$$

[10] For very short σ signals, the restriction is that $f > 2\sigma$.

Therefore the size of the proposed template family is

$$I = n_A n_f n_\sigma \sum_\sigma n_{t_0} = 2 \left(\frac{f_1}{f_0}\right) \left(\frac{\ln(\sigma_1/\sigma_0)}{\ln 6}\right) \left(\frac{T}{\sigma_0}\frac{1}{5}\right) \tag{77}$$

where we have evaluated the sum of the geometrical series. If the limits of the parameter choices are given according to (67-72), this gives $I \approx 1000{\cdot}T/\mathrm{s}$. For J independent detectors with effective sampling frequencies $F_s = 4096\,\mathrm{Hz}$, the number of data points is $K = 4096 \cdot JT/\mathrm{s}$.

Substituting these numbers for the false alarm and true detection probabilities (35) and (36), the false alarm rates become calculable, for a given choice of margin m.

$$\lambda_{FA} = \frac{1}{8J} F(\sqrt{2JK} - \sqrt{2}m) \tag{78}$$

This false alarm rate has a corresponding detection probability given by (36) or (114), which can be calculated exactly. Executing the burst search algorithm yields a detection rate, λ_D, for these burst signals. The a priori probability of the given bursts are then calculable

$$P(\boldsymbol{s}_i \,|\, i \in \{1 \ldots I\}) = \frac{\lambda_D - \lambda_{FA}}{P(\boldsymbol{s}_i \,|\, i \in \{1 \ldots I\}, \neg D_0)} \tag{79}$$

In fact, this calculation can be carried out for any subset of the burst signals leading to the a priori probabilities of those signals. If $\lambda_D \approx \lambda_{FA}$, then the result can be used to obtain an upper limit for the event rate. Since false alarms follow a Poison distribution, the uncertainty is

$$\Delta\lambda = \sqrt{\lambda_{FA}} \tag{80}$$

Therefore the upper limit for the true detection rate is

$$\lambda_{TD} < \lambda_{FA} + \sqrt{\lambda_{FA}} \tag{81}$$

By applying the two additional procedures of Section 6, namely the relative time shift and signal strength check, the false alarm rates can be reduced. This leads to a stronger upper limit for the burst rate.

8 Conclusions

In this study, we have investigated the theory of optimal signal detection in noise. A probabilistic interpretation was associated to the problem. The single alternate hypothesis and the multiple alternate hypotheses maximum likelihood detection methods were defined which were shown to give the greatest detection probability for a fixed false alarm rate. The single alternate hypothesis method such as the Neyman-Pearson detection can only be used effectively if there was an a priori belief that there is only one possible signal waveform present in the data stream. In the realistic case however, a large population of signal templates have to be used. With no a priori information, the maximum likelihood detection method was shown to be optimal and the necessary and sufficient size of the template family was defined.

The measurement probability distribution was imagined in a K dimensional vector space. This gave a basis independent visualization. Using this aid, we could carry out the calculations to describe the efficiency of the method. We derived the analytic approximations to the detection confidence and the false alarm probabilities. This defined the optimal detection expectations for a given waveform. The results naturally depended on K, the length of the data stream, which demonstrated how the efficiency improved for longer waveforms.

The calculations on the limitations of the proposed estimate were carried out only for the white noise case. Nevertheless it was pointed out, that using the exact form of the autocorrelation function of noise is possible with the Karhunen-Loeve expansion, which leads to a statistically independent realization of the measurement. The KL transformation is a unitary transformation that corresponds to a rotation of the basis vectors in the K dimensional vector space. The calculation results are therefore valid in the general case after the KL transformation.

We have also generalized to the case of multidetector signal detection. The likelihood function was derived assuming that the detectors were independent. The corresponding optimal detection efficiency with multiple detectors were identical with the single detector case with a larger sampling frequency. This way the calculations for the single detectors could be applied also in the general case.

Finally we have shown that the maximum likelihood detection method has also the advantage that the physical requirements can be easily incorporated. Since each individual detector measures the same gravitational wave signal,

the relative signal strengths and time shifts are known additionally. This way, a given detection trigger can be checked for these requirements, and the unphysical triggers can be vetoed. This improvement increases the efficiency by decreasing the false alarm rate.

Future plans include the numerical check of the theoretical calculations with Monte Carlo simulations. The optimal code can be applied to search the data streams of the available science runs of the LIGO detectors. In particular, selected time intervals with associated optical triggers can be examined closely to see what consequences can be drawn from those measurements for their gravitational radiation signatures. Also, the systematic data analysis of the gravitational wave detectors' statistics for the full science runs can be used to obtain upper limits for the statistical presence of burst signals.

9 Acknowledgements

We are grateful for the Caltech SURF program and the LIGO collaboration, who made our work possible. The authors gratefully acknowledge the support of the United States National Science Foundation for the LIGO Laboratory. This material is based upon work supported by National Science Foundation under Grant [or Cooperative Agreement] No. (NSF grant or cooperative agreement number). This document has been assigned LIGO Laboratory document number LIGO-T030213-00-D.

This paper is one of the three papers submitted for our final report of the LIGOs 2003 SURF program.[11] We also gratefully thank our mentor, Szabolcs Marka, for his help and supervision during the work.

10 Appendix A.

For an effective template family, the q_K number of lowest strength templates will be the number of points available on the $\sqrt{K}$ radius K-dimensional sphere having a $\sqrt{K-1}$ separation. Hereby we present the derivation of the formula for q_K.

It seems evident, that the number of such points q_K is proportional to

[11] The other two papers are [13] and [14].

S_K, the surface area of the unit sphere. Explicitly,

$$\sqrt{K-1}^{K-1} q_K = \sqrt{K}^{K-1} S_K = \frac{2\sqrt{K}^{K-1}\pi^{K/2}}{\Gamma(K/2)} \tag{82}$$

For $K \gg 1$, we can simplify with the $K-1$ power factors.

$$q_K = S_K = \frac{2\pi^{K/2}}{\Gamma(K/2)} \tag{83}$$

However, $\Gamma(n)$ grows faster than $n^n e^{-n}$ for large K, therefore S_K approaches 0. This is an interesting feature of multidimensional Euclidean geometry: above the 7^{th} dimension the surface of the unit sphere decreases, and vanishes faster than exponentially with the number of dimensions. Naturally the number of unit separation points q_K on the unit sphere is a monotonic function of the dimensions. For dimensions $k > 7$, one point can always be chosen for each dimension along the corresponding basis vectors and one in the opposite direction. The distance between these points and the selected points in the previous subspaces are surely greater than one. Therefore for large number of dimensions, where $S_K - S_{K-1} < 2$ the number q_K increases[12] as $2K$.

$$q_K = \begin{cases} 19 + 2K & \text{for } K \geq 7 \\ S_K & \text{otherwise} \end{cases} \tag{84}$$

This is therefore the number of different unit-strength signals one can distinguish in unit variance noise[13]. The approximation of $q_K = 2K$ corresponds to a template family with a minimal mutual separation of $\sqrt{2K}$. For larger strength signals with norm $s_1 > \sqrt{K}$, the surface of the sphere is larger leading to a generally larger number of detectable signals.

$$s_1^{K-1} S_K = \frac{2 s_1^{K-1}\pi^{K/2}}{\Gamma(K/2)} \tag{85}$$

Although larger than (83), the s_1 radii sphere's surface also approaches zero with the number of dimensions. Therefore, just as in the unit norm case, the increase of the number of distinct templates q_K can be estimated by $2K$

[12] For lower dimensions K<7, eq. (83) is valid. The q number increases as 2, 6, 12, 19, 26, 31, 33 for $k \leq 7$.

[13] If the opposite sign signals are identified, then $q_K/2$ is the number of signals we can measure.

above a certain number of dimensions. This number, corresponding to 7 for $s_1 = \sqrt{K}$, will generally depend on s_1. We shall only work with the $q_K = 2K$ lower bound approximation in this study.

11 Appendix B.

Here we derive the multiple alternate hypotheses false detection probability for Section 3.2.

Recall from 3.2, that the decision is made in favor of a template $\boldsymbol{s}_i$ exactly if the measurement $\boldsymbol{y}$ assumed a point within a chopped cone solid. The cone has an axis along $\boldsymbol{s}_1$, peak at the origin $\boldsymbol{s}_0$, radius $\sqrt{K-1}$, and the top cutoff at $m\boldsymbol{e}_1$. A measurement, $\boldsymbol{y}$, at a larger angle from the origin, corresponds to a different template of the unit sphere, and the tip of the cone corresponds to the no signal case $\boldsymbol{s}_0$.

We have used the distinction between confusion and false alarm probabilities. The former was $P_{ij}^C = P(D_i \,|\, H_j,\, i \neq j)$, whereas the later was defined by $P_i^{FA} = P(D_i \,|\, H_0)$ for each $i \in [1 \ldots I]$. The individual false alarm probabilities are equal, so the total false alarm probability is $P^{FA} = IP_1^{FA}$, where I is the number of alternate hypotheses. The expected total false detection is

$$< \text{False detection} >= IP_1^{FA}P(H_0) + \sum_{ij} P_{ij}^C P(H_j) \tag{86}$$

Since for the a priori probabilities $P(H_0) \gg P(H_j)$, the false detection is dominated by the false alarms over the confusion terms.

The false alarm probability associated with the $\boldsymbol{s}_1$ signal is

$$P_1^{FA} = P(D_1 \,|\boldsymbol{s}_0) = \int\limits_{\text{chopped cone}} \frac{\mathbf{d}\boldsymbol{y}}{(\sqrt{2\pi})^K} \exp \frac{-\boldsymbol{y}^2}{2} \tag{87}$$

Let us now make a small digression in order to work out the correct intuition one needs to better comprehend the task (87). There is a counter-intuitive consequence of multidimensional geometry: the radial distribution[14] of the distribution $\sqrt{2\pi}^{(-1)} \exp(-\boldsymbol{y}^2/2)$ is negligible within the unit sphere and is only non-zero for radii around $\kappa = \sqrt{K}$. This latter region for high K

[14]The radial distribution is the probability density of a measurement to fall on the spherical surface of radius r and thickness dr.

dimensions is far away from the unit sphere. The reason for this peculiarity is that the surface area of a K dimensional sphere grows with r^{K-1}, a high power of the radius. When calculating the contributions of spherical surfaces away from the origin, the far away regions are enhanced by a great amount. As we shall see, this is the reason for the relative insensitivity of the multiple hypotheses detection compared to the single alternate hypothesis method of 3.1.

Changing to spherical coordinates, the angle dependant integrals can be evaluated as

$$\int_{\dots} \frac{\mathbf{d}\boldsymbol{y}}{(\sqrt{2\pi})^K} = \int_m^\infty \frac{\mathrm{d}y}{(\sqrt{2\pi})^K} y^{K-1} \frac{S_K}{q_K} \tag{88}$$

since $\boldsymbol{s_1}$ occupies $1/q_K$ of the total spherical surface.

$$P(D_1 \,|\boldsymbol{s}_0) = \int_m^\infty \frac{\mathrm{d}y}{(\sqrt{2\pi})^K} y^{K-1} \frac{S_K}{q_K} \exp \frac{-y^2}{2} = \tag{89}$$

$$= \frac{1}{2\sqrt{\pi}^K} \frac{S_K}{q_K} \int_{m^2/2}^\infty \mathrm{d}z z^{\frac{K}{2}-1} \exp(-z) = \tag{90}$$

$$= \frac{1}{q_K} \frac{\Gamma(K/2, m^2/2)}{\Gamma(K/2)} \tag{91}$$

Where in the first equation we have changed to the variable[15] $z = y^2/2$, and in the second we have used the definition of S_K from (83) and the incomplete Gamma function $\Gamma(n, \mu)$. Note that $\Gamma(n) = \Gamma(n, 0)$.

To make this more visible, let us express the answer in terms of the Gaussian cumulative probability density $F(x)$. Recall that the complete and the incomplete Gamma functions are defined with the same integral formula, only the the bounds of the integral domains are different. The kernel function is $\phi(z, n) = z^{n-1} \exp(-z)$. For large n, one can approximate the kernel with a Gaussian using the saddle point method. Carrying out the calculations and using the Stirling formula for the Γ function, we get

$$\phi(z, n) \approx \frac{\Gamma(n)}{\sqrt{2\pi}\sqrt{n-1}} \exp\left[-\frac{(z-(n-1))^2}{2(n-1)}\right] \tag{92}$$

The relative error of this approximation is of order 10^{-3} for $n = 10$, and decreases further for larger n. Substituting this in the defining integrals of

[15] Notice that z is a random variable with Poison distribution. **Nezzem meg valahol.**

the Γ functions, we obtain

$$\frac{\Gamma(n,\mu)}{\Gamma(n)} = \int_{\mu}^{\infty} \frac{\mathrm{d}z}{\sqrt{2\pi}\sqrt{n-1}} \exp\left[-\frac{(z-(n-1))^2}{2(n-1)}\right] \tag{93}$$

$$= \int\limits_{\frac{\mu-(n-1)}{\sqrt{n-1}}}^{\infty} \frac{\mathrm{d}z}{\sqrt{2\pi}} \exp(-z^2/2) = F\left[\frac{(n-1)-\mu}{\sqrt{n-1}}\right] \tag{94}$$

Substituting this approximations for the false alarm rate (89) we get

$$P(D_1\,|\boldsymbol{s}_0) = \frac{1}{q_K} F\left[\frac{-\mu+\kappa}{\sqrt{\kappa}}\right] \approx \frac{1}{2K} F\left[\frac{-\mu+\kappa}{\sqrt{\kappa}}\right] \tag{95}$$

where $\mu = m^2/2$ and $\kappa = \frac{K}{2} - 1$, and eq. (84) was used for q_K assuming that K is large.

In the previous solution, we calculated the result by approximating the analytical answer (89). It turns out that the Gaussian approximation of the original integral formula leads to a smaller error. We only present the result of that calculation for $K \gg 1$ here.

$$\int_m^{\infty} \frac{\mathrm{d}y}{(\sqrt{2\pi})^K} y^{K-1} \frac{S_K}{q_K} \exp\frac{-y^2}{2} \approx \frac{1}{q_K} \int_m^{\infty} \frac{\mathrm{d}y}{\sqrt{\pi}} \exp[-(y-\sqrt{K})^2] \tag{96}$$

Recall that y is the polar coordinate representation of the measured signal[16], or equivalently its empirical standard deviation. We have obtained that for K-long data samples, this has an approximately normal distribution of mean $\sqrt{K}$ and variance $1/\sqrt{2}$. [17]

After changing the integration variable, this simplifies to

$$\frac{1}{q_K} \int_{m-\sqrt{K}}^{\infty} \frac{\mathrm{d}y}{\sqrt{\pi}} \exp[-y^2] = \frac{1}{q_K} \int_{\sqrt{2}(m-\sqrt{K})}^{\infty} \frac{\mathrm{d}y}{\sqrt{2\pi}} \exp[-y^2/2] \tag{97}$$

Therefore the result for the false alarm rate is

$$P_1^{FA} = P(D_1\,|\boldsymbol{s}_0) = \frac{1}{2K} F[\sqrt{2K} - \sqrt{2}m)] \tag{98}$$

[16] We are assuming that there is no true signal present for this calculation. $S = \boldsymbol{s}_0$

[17] Let us point out that this result does not contradict the central limiting distribution theorem. That theorem applies for the arithmetic mean of identical distributions, whereas we have calculated the distribution of the second order mean. The second order mean has a variance $1/\sqrt{2}$ instead of 1, the variance for the simple mean.

Equation (98) is a better approximation of the exact answer (89) than the previously obtained (98) result. The false alarm probability for each single template decreases as K^{-1}. This is a big advantage in favor of long long data streams.

There is a second aspect, which shows why longer signal samples are better in terms of detection. The noise strength grows just as the signal strength, but according to equation (96) its relative uncertainty decreases as $1/\sqrt{2K}$. Observe in Figure 4 that the detection threshold can be adjusted very close to $s = \sqrt{K}$ (i.e. $s/n = 1$) for long (large K) data streams.

12 Appendix C.

Here we derive the correct signal detection probability for Section 3.2. The signal $\boldsymbol{s}_1$ is accepted whenever the measurement $\boldsymbol{y}$ falls in the region of the previous case D_1

$$P(D_1\,|\boldsymbol{s}_1) = \int\limits_{D_1} \frac{\mathbf{d}\boldsymbol{y}}{(\sqrt{2\pi})^K} \exp\left[\frac{-(\boldsymbol{y}-\boldsymbol{s}_1)^2}{2}\right] = \tag{99}$$

$$= \int\limits_{D_1-\boldsymbol{s}_1} \frac{\mathbf{d}\boldsymbol{y}}{(\sqrt{2\pi})^K} \exp\frac{-\boldsymbol{y}^2}{2} = \tag{100}$$

$$= P(D_1\,|\boldsymbol{s}_0) + \int\limits_{(D_1-\boldsymbol{s}_1)\setminus D_1} \frac{\mathbf{d}\boldsymbol{y}}{(\sqrt{2\pi})^K} \exp\frac{-\boldsymbol{y}^2}{2} \tag{101}$$

$$= P(D_1\,|\boldsymbol{s}_0) + I \tag{102}$$

This extent of which the detection probability will be higher than the false alarm rate is therefore determined by the integral over the region $(D_1 - \boldsymbol{s}_1) \setminus D_1$. We shall abbreviate this formula with I. Recall that D_1 is a chopped cone with an upper cutoff at a distance m from the cone's peak.

Notice the rotational symmetry around the axis parallel to $\boldsymbol{s}_1$, $K-2$ integrals can be evaluated in cylindrical coordinates. What remains are integrals over a two dimensional domain

$$I = \int_m^\infty \mathrm{d}r \frac{r^{K-2} S_{K-1}}{\sqrt{2\pi}^{K-1}} \int_{r-s_1}^{r} \frac{\mathrm{d}y}{\sqrt{2\pi}} \exp(-r^2/2) \exp(-y^2/2) \tag{103}$$

$$+ \int_0^m \mathrm{d}r \frac{r^{K-2} S_{K-1}}{\sqrt{2\pi}^{K-1}} \int_{m-s_1}^{m} \frac{\mathrm{d}y}{\sqrt{2\pi}} \exp(-r^2/2) \exp(-y^2/2) \tag{104}$$

The remaining two integrals can be evaluated numerically for a given (s_1, m) pair[18]. We shall instead use analytic approximations to get a more visible functional form. Firstly, we can apply the saddle point approximation (96), and use $K \gg 1$ to finally obtain

$$I = \int_m^\infty \frac{\mathrm{d}r}{\sqrt{\pi}} \int_{r-s_1}^r \frac{\mathrm{d}y}{\sqrt{2\pi}} \exp[-(r-\sqrt{K})^2] \exp(-y^2/2) \tag{106}$$

$$+ \int_0^m \frac{\mathrm{d}r}{\sqrt{\pi}} \int_{m-s_1}^m \frac{\mathrm{d}y}{\sqrt{2\pi}} \exp[-(r-\sqrt{K})^2] \exp(-y^2/2) \tag{107}$$

The order of magnitude of s_1 which is possible to detect is $\sqrt{K}$. Therefore the boundary m is between 0 and $s_1 \propto \sqrt{K}$. Shifting the r coordinate by $\sqrt{K}$, and rescaling by $\sqrt{2}$, we get

$$I = \int_{\sqrt{2}(m-\sqrt{K})}^\infty \frac{\mathrm{d}r}{\sqrt{2\pi}} \int_{\frac{r}{\sqrt{2}}+\sqrt{K}-s_1}^{\frac{r}{\sqrt{2}}+\sqrt{K}} \frac{\mathrm{d}y}{\sqrt{2\pi}} \exp\left[-\frac{r^2+y^2}{2}\right] \tag{108}$$

$$+ \int_{-\sqrt{2K}}^{\sqrt{2}m-\sqrt{2K}} \frac{\mathrm{d}r}{\sqrt{2\pi}} \int_{m-s_1}^m \frac{\mathrm{d}y}{\sqrt{2\pi}} \exp\left[-\frac{r^2+y^2}{2}\right] \tag{109}$$

Let us denote the two parts of this sum as I_1 and I_2 respectively. The second integral can be obtained analytically

$$I_2 = \left[F(\sqrt{2K}-\sqrt{2}m) - F(\sqrt{2K})\right] \left[F(m) - F(m-s_1)\right] \tag{110}$$

For the first integral I_1, we can make use of the fact that the integrand is rotationally symmetric. By rotating the axes in the "parallel" and "perpendicular" dimensions of the domain, it is possible to give two bounds on the integral I_1.[19]

$$I_{1-} < I_1 < I_{1+} \tag{111}$$

[18] One of these can be calculated analytically. The remaining integral

$$\frac{1}{2}\mathrm{erf}(+\sqrt{K})F(s_1-m) + \int_{m-s_1}^\infty \frac{\mathrm{d}y}{2\sqrt{2\pi}} \exp(-y^2/2)\mathrm{erf}(y+s_1-\sqrt{K}) \tag{105}$$

is plotted on Figure 4.

[19] The domain of the integral I_1 is an infinite trapeziod, which can be bounded by two infinite rectangles from above and from below respectively. The integrals over rectangles are separable. Their values will be denoted by I_{1-} and I_{1+}.

Explicitly,

$$I_{1-} = F\left[\sqrt{\frac{2}{3}}(\sqrt{K} - 2m)\right]\left\{F\left[\frac{m + s_1 - \sqrt{2K}}{\sqrt{3}}\right] - F\left[\frac{m - \sqrt{2K}}{\sqrt{3}}\right]\right\} \tag{112}$$

$$I_{1+} = F\left[\sqrt{\frac{2}{3}}(s_1 + \sqrt{K} - 2m)\right]\left\{F\left[\frac{m + s_1 - \sqrt{2K}}{\sqrt{3}}\right] - F\left[\frac{m - \sqrt{2K}}{\sqrt{3}}\right]\right\} \tag{113}$$

For the sake of an estimate we shall use the central value between the bounds. The results can hereby be substituted in (101) to give

$$\begin{aligned} P(D_1\,|\boldsymbol{s}_1) =& \frac{1}{2K}F(\sqrt{2K} - \sqrt{2}m) \\ &+ \left[F(\sqrt{2K} - \sqrt{2}m) - F(\sqrt{2K})\right]\left[F(m) - F(m - s_1)\right] \\ &+ F\left[\sqrt{\frac{2}{3}}\left(\frac{s_1}{2} + \sqrt{K} - 2m\right)\right]\left\{F\left[\frac{m + s_1 - \sqrt{2K}}{\sqrt{3}}\right]\right. \\ &\qquad \left. - F\left[\frac{m - \sqrt{2K}}{\sqrt{3}}\right]\right\} \end{aligned} \tag{114}$$

Whereas the false alarm rate was given by (98)

$$P(D_1\,|\boldsymbol{s}_0) = \frac{1}{2K}F[\sqrt{2K} - \sqrt{2}m)] \tag{115}$$

Equations (114) and (115) gives the detection confidence and false alarm rate for the maximum likelihood method, for a given strength signal $\boldsymbol{s}_1$. We have assumed the noise to be unit variance Gaussian for each data sample point and of much larger a priori probability than the alternate hypotheses. The alternate hypotheses templates were evenly distributed on the $\sqrt{K}$-radius sphere with a minimum separation of $\sqrt{2K}$. Comparing with (31) and (34), it is obvious that the false alarm rate is indeed decreased for the multiple hypotheses detection by a large amount, i.e. by a factor of $\frac{1}{2K}$, however the the true detection probability is also decreased. Therefore the detection sensitivity will be much lower in this case (see Figure 4 and Figure 2) with a signal to noise ratio of $s/n \propto 1$.

References

[1] K. C. B. New, "Gravitational Waves from Gravitational Collapse" http://xxx.lanl.gov/abs/gr-qc/0206041

[2] M.H.P.M. van Putten, A. Levinson, H. K. Lee, T. Regimbau, M. Punturo, G. Harry, "Gravitational radiation from gamma-ray bursts as observational opportunities for LIGO and VIRGO," LIGO publication LIGO-P030041-00-D, 2003.

[3] Buonnanno, Chen, Vallasneri, "Detection template families for gravitational waves from the final stages of binary-black-hole inspirals: Nonspinning case" *Phys. Rev. D* **67** 024016, 2003.

[4] LIGO-I authors, "Analysis of First LIGO Science Data for Stochastic Gravitational Waves," LIGO publication LIGO-P030009-A-Z, 2003.

[5] LIGO-I authors, "Setting upper limits on the strength of periodic gravitational waves using the first science data from the GEO600 and LIGO detectors," LIGO publication LIGO-P030008-05-Z, 2003.

[6] LIGO-I authors, "First upper limits on gravitational wave bursts from LIGO," LIGO publication LIGO-P030011-00-Z, 2003.

[7] J. Sylvestre, "Time-frequency detection algorithm for gravitational wave bursts," *Phys. Rev. D* **66**, 102004, 2002.

[8] S. Mohanty el al., "Gamma ray bursts and gravitational waves: triggered search," accepted for publication in Classical and Quantum Gravity

[9] L. S. Finn, "Issues in gravitational wave data analysis," http://www.lanl.gov/abs/gr-qc/9709077

[10] G. J. Feldman and R. D. Cousins, "Unified approach to the classical statistical analysis of small signals," *Phys. Rev. D* **66**, 3873, 1998.

[11] A. D. Whalen, "Detection of Signals in Noise" *Academic Press*, New York, 1971.

[12] M. G. Kendall and A. Stuart, "Inference and Relationship, The Advanced Theory of Statistics," Vol. 2. *Hafner Publ.*, New York, 1961.

[13] B. Kocsis and M. Gaspar, "Expectations on the gravitational wave signals associated with cosmic brehmsstahlung events," LIGO technical note LIGO-T030136-00-D, 2003.

[14] M. Gaspar and B. Kocsis, "The development of a digital camera with a high dynamic range" LIGO technical note LIGO-T030232-00-D, 2003.