LASER INTERFEROMETER GRAVITATIONAL WAVE OBSERVATORY
- LIGO –
CALIFORNIA INSTITUTE OF TECHNOLOGY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Technical Note	LIGO-T030112-00-D	04/22/2003

A study of the cross-correlation coefficient distribution in the presence of additive signals

S. Mohanty, R. Rahkola, Sz. Márka, S. Mukherjee, R. Frey

Max Planck Institut für Gravitationsphysik
Am Mühlenberg 1, D14476, Germany
Phone +49-331-567-7220
Fax +49-331-567-7298
E-mail: office@aei.mpg.de

California Institute of Technology
LIGO Laboratory - MS 18-34
Pasadena CA 91125
Phone (626) 395-212
Fax (626) 304-9834
E-mail: info@ligo.caltech.edu

Massachusetts Institute of Technology
LIGO Laboratory - MS 16NW-145
Cambridge, MA 01239
Phone (617) 253-4824
Fax (617) 253-7014
E-mail: info@ligo.mit.edu

www: http://www.ligo.caltech.edu/

This note is a continuation of [1], which should be read for background information.

1 Statement of problem

In order to use the cross-correlation coefficient, r, for confidence interval estimation, we need to know its probability density function, $p(r; h_x, h_y)$, given additive signals $h_x[i]$ and $h_y[i]$, $i = i_0, \ldots, i_0 + M - 1$, in the two IFO time series $x[k]$ and $y[k]$, $k = 0, \ldots, N-1$, respectively. In the following, $0 \leq i_0 \leq N-1$ and $0 \leq i_0 + M - 1 \leq N$. We report a study of $p(r; h_x, h_y)$ in the context of its use in a triggered search.

2 Analytical results

For simplicity, we consider only the identical signal case $h_x[k] = h_y[k] = h[k]$ here. Let the noise in each detector be a Gaussian, zero mean, white random process.

First, we explore how the signal waveform influences the distribution $p(r; h)$ of r. The main result is that there are only three gross quantities associated with a waveform that are relevant to this problem. The three quantities are,

1. The signal mean μ_h,
$$\mu_h = \frac{1}{M} \sum_{i=0}^{M-1} h[i + i_0] \; . \tag{1}$$

2. The signal duty cycle ϵ,
$$\epsilon = \frac{M}{N} \; . \tag{2}$$

3. The norm ρ_h of the signal *after mean removal*,
$$\rho_h^2 = \sum_{i=0}^{M-1} \left(h[i + i_0] - \mu_h \right)^2 \; . \tag{3}$$

Theorem – The distribution $p(r; h)$ depends only on a single quantity, ρ_r,
$$\rho_r^2 = \rho_h^2 + M \mu_h^2 (1 - \epsilon) \; . \tag{4}$$

Proof – Let $n_x[k]$ and $n_y[k]$ denote the noise components of $x[k]$ and $y[k]$ respectively. Let $s[k]$ be the signal component, $s[k] = h[k]$, $k = i_0, \ldots, i_0 + M - 1$ and zero otherwise. Then,

$$r = \frac{\sum_{i=0}^{N-1} (n_x[i] + s[i] - \widehat{\nu}_x - \epsilon\mu_h)(n_y[i] + s[i] - \widehat{\nu}_y - \epsilon\mu_h)}{||n_x + s - \widehat{\nu}_x - \epsilon\mu_h|| \, ||n_y + s - \widehat{\nu}_y - \epsilon\mu_h||} \; ,$$

$$\widehat{\nu}_x = \frac{1}{N}\sum_{i=0}^{N-1} n_x[i] \; , \tag{5}$$

$$\widehat{\nu}_y = \frac{1}{N}\sum_{i=0}^{N-1} n_y[i] \; , \tag{6}$$

Temporarily redefine $n_x[i] \to n_x[i] - \widehat{\nu}_x$, $n_y[i] \to n_y[i] - \widehat{\nu}_y$ and $s[i] \to s[i] - \epsilon\mu_h$. Then,

$$r = \frac{\langle n_x, n_y\rangle + ||s||\left(||s|| + \langle n_x, \widehat{s}\rangle + \langle n_y, \widehat{s}\rangle\right)}{\sqrt{||n_x||^2 + ||s||\left(||s|| + \langle n_x, \widehat{s}\rangle\right)}\sqrt{||n_y||^2 + ||s||\left(||s|| + \langle n_y, \widehat{s}\rangle\right)}} \; , \tag{7}$$

$$\widehat{s} = \frac{s}{||s||} \; . \tag{8}$$

Consider the terms $\langle n_x, \widehat{s}\rangle$ and $\langle n_y, \widehat{s}\rangle$. Reverting back to the original definitions for n_x and n_y,

$$\langle n_x, \widehat{s}\rangle \to \langle n_x, \widehat{s}\rangle - \widehat{\nu}_x \sum_{i=0}^{N-1} \widehat{s}[i] \; . \tag{9}$$

But $\sum_{i=0}^{N-1} \widehat{s}[i] = 0$ and the remaining term is the projection of n_x on some fixed unit vector. However, n_x, being a white noise sequence will have an isotropic distribution in an N dimensional vector space. Therefore, the distribution of its projection on a fixed unit vector will not depend on the direction of the unit vector. (The same argument goes through for $\langle n_y, \widehat{s}\rangle$.) Hence, examining Eq. 7, we see that none of the random variables, $\langle n_x, n_y\rangle$, $||n_x||$, $||n_y||$, $\langle n_x, \widehat{s}\rangle$ and $\langle n_y, \widehat{s}\rangle$, depend on the signal in their distribution. The distribution of r, therefore, only depends on $||s||$,

$$\begin{aligned} ||s||^2 &= \sum_{i=0}^{M-1} (h[i] - \epsilon\mu_h)^2 + \sum_{i=M}^{N-1} (\epsilon\mu_h)^2 \\ &= \rho_h^2 + M\mu_h^2(1-\epsilon) \; . \end{aligned} \tag{10}$$

Q.E.D.

The above result means that the r statistic is as efficient for Gaussians as for sine-Gaussians or for any other signal as long as they have the same ρ_r. It also means that an interval/point estimate of ρ_r is independent of signal waveform, within random errors, **even if the estimation is done using signal injection with specific waveform types**. The only way we can tune ourselves to one signal type versus another is to first pre-process the data such that only one type of signal survives. For example, band pass filtering the data can choose between sine-Gaussians with different carrier frequencies. However, even with pre-processing, the estimate of ρ_r applies to *whatever signal survives the processing.*

Thus, the purpose of using signal injection for estimating ρ_r is *not* to arrive at different limits for different signal types but to arrive at a limit on ρ_r for whatever signal is present in the conditioned data, albeit with a different treatment of noise and a different set of assumptions about the data. For instance, the systematic errors may be different for methods that use signal injections from the ones that do not or the response to non-stationarity may be different. The astrophysical interpretation of the signal injection derived limit on ρ_r will not be any different from waveform independent methods for limiting ρ_r.

3 Numerical results for identical signals

The proof that $p(r;h)$ depends only on ρ_r is demonstrated using MC simulation with different signal types. The MC simulation consists of generating pairs, $x[k]$ and $y[k]$, $k = 0,\ldots,N-1$, of white Gaussian noise sequences. Each sequence is zero mean and each sample has unit variance. The *same* signal waveform $h[k]$, $k = 0,\ldots,M-1$, $(M \leq N)$ is injected into each sequence and r is calculated repeatedly over N_{trials} independent trials. The distribution $p(r;h)$ is then estimated from this set of r values. Signals are injected with amplitude A such that ρ_r has some prescribed value.

The signal types used are,

Gaussian pulse

DFM waveforms From the Dimmelmeier, Font, Müller catalog [2, 3].

Figure 1 shows the result. As can be seen, $p(r;h)$ depends only on ρ_r within statistical error.

References

[1] Soumya D. Mohanty *et al*, LIGO-T030111-00-D.

[2] Dimmelmeier, H., Font, J.A., and Müller, E., "Relativistic simulations of rotational core collapse. I. Methods, initial models, and code tests", Astron. Astrophys., 388, 917-935 (2002); astro-ph/0204288.

[3] Dimmelmeier, H., Font, J.A., and Müller, E., "Relativistic simulations of rotational core collapse. II. Collapse dynamics and gravitational radiation", Astron. Astrophys., 393, 523-542 (2002); astro-ph/0204289.

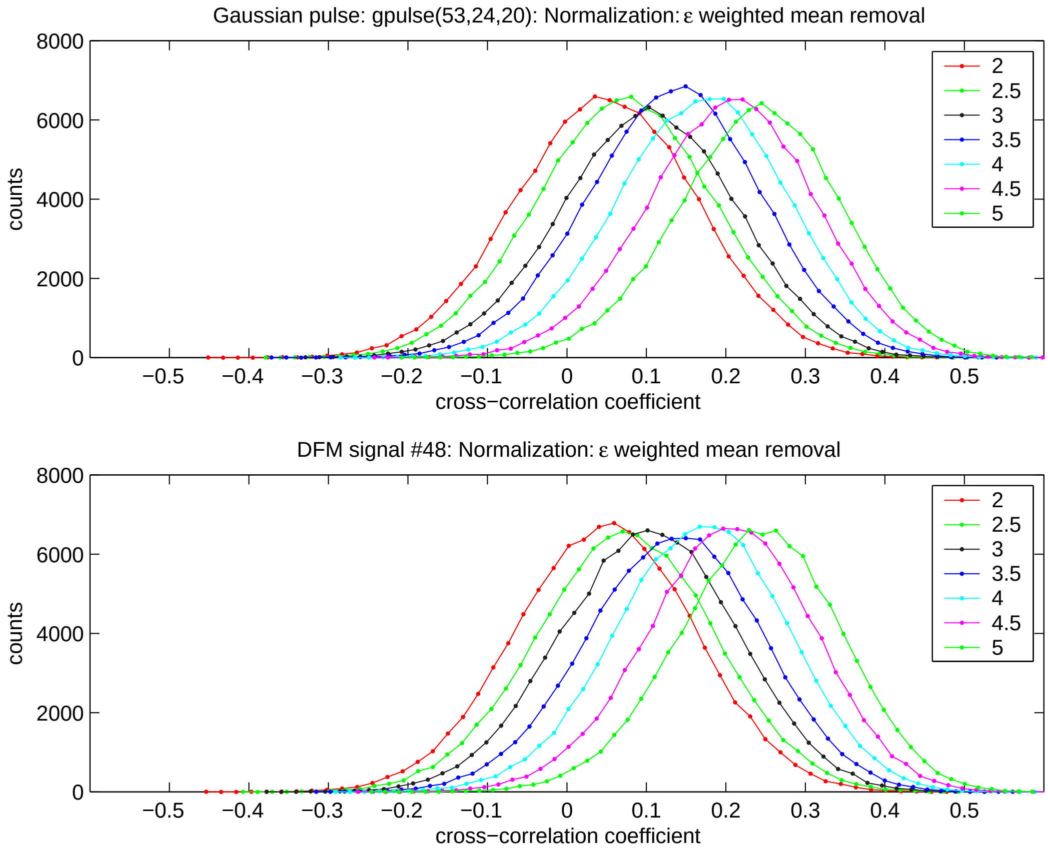

Figure 1: The probability density of r for different signal types as a function of ρ_r. The top panel is for a Gaussian pulse that is 50 msec wide. The bottom panel is for a signal (A4B4G5_R) from the DFM catalog [2, 3]. The signals are shown in Fig. 2.

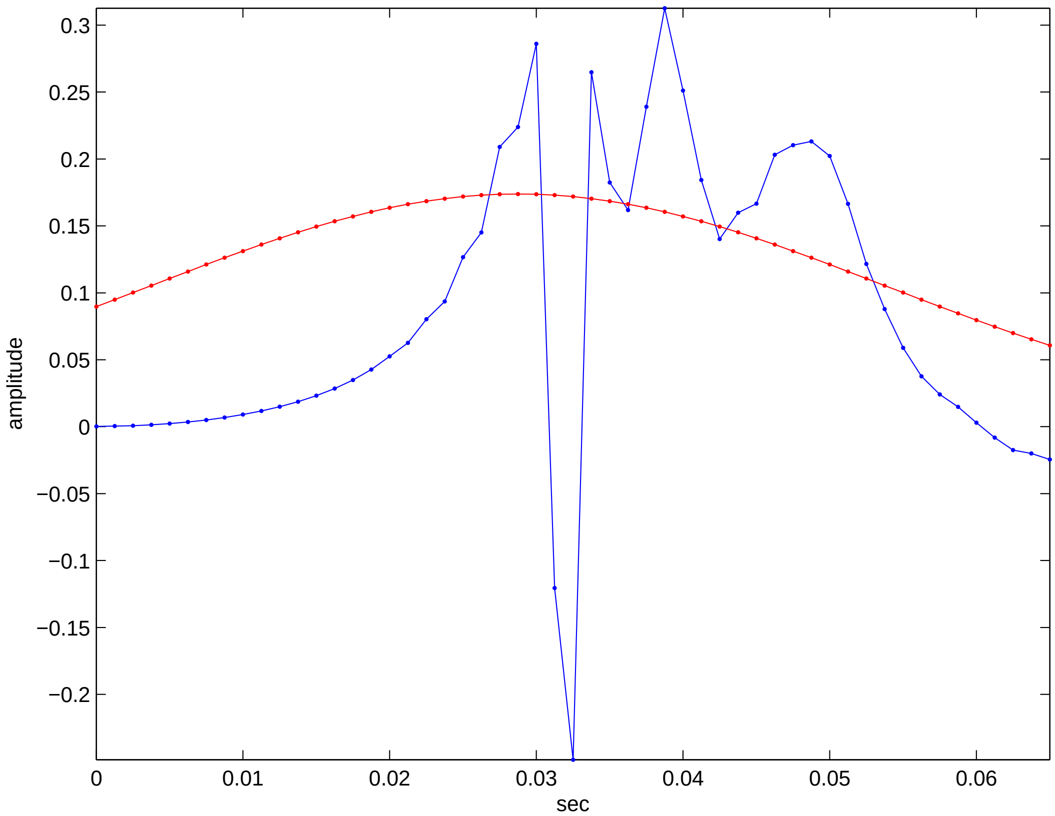

Figure 2: The signals used in the simulation (c.f. Fig. 1).fig2)