Confidence Test for Waveform Consistency of LIGO Burst Candidate Events

Laura Cadonati
LIGO Laboratory
Massachusetts Institute of Technology
Preface

The LIGO Burst Search pipeline uses *Event Trigger Generators* (ETGs) to flag times when “something anomalous” occurs in the strain time series

⇒ burst candidate events ($\Delta t$, $\Delta f$, SNR)

Events from the three LIGO interferometers are brought together in coincidence (time, frequency, power).

In order to use the full power of a coincident analysis:

» Are the waveforms consistent? To what confidence?

» Can we suppress the false rate in order to lower thresholds and dig deeper into the noise?

Cross correlation of coincident events
Assigning a Correlation Confidence to Coincident Candidate Events

Pre-process

1. Load time series from 2 interferometers (2 sec before event start)
2. Decimate and high-pass \(\Rightarrow 100-2048\text{Hz}\)
3. Remove predictable content (effective whitening/line removal): train a linear predictor filter over 1 s of data (1 s before event start), apply to the rest.
   \(\Rightarrow\) emphasis on transients, avoid non-stationary, correlated lines.
4. Apply an r-statistic test to quantify the correlation between interferometer pairs
**r-statistic**

\[
r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}
\]

**Null Hypothesis:** the two (finite) series \(\{x_i\}\) and \(\{y_i\}\) are uncorrelated

\[\Rightarrow\text{Their linear correlation coefficient (Pearson’s } r\text{) is normally distributed around zero, with } \sigma = 1/\sqrt{N} \text{ where } N \text{ is the number of points in the series (} N \gg 1\text{)}\]

\[S = \text{erfc } (|r| \sqrt{N/2})\]

double-sided significance of the null hypothesis

i.e.: probability that \(|r|\) is larger than what measured, if \(\{x_i\}\) and \(\{y_i\}\) are uncorrelated

\[C = - \log_{10}(S)\]

confidence that the null hypothesis is FALSE \(\Rightarrow\) that the two series are correlated
5th Edoardo Amaldi Conference - July 9, 2003

Delay and Integration Time

What delay?

Shift \{y_i\} vs \{x_i\} and calculate: r_k, S_k, C_k

…then look for the maximum confidence C_M

Time shift for C_M = delay between IFOs

Shift limits: ±10 ms (LLO-LHO light travel time)

Integration time \(\tau\):

> If too small, we lose waveform information and the test becomes less reliable

> If too large, we wash out the waveform in the cross-correlation

Test different \(\tau\)'s and do an OR of the results  (20ms, 50ms, 100ms)

\[
 r_k = \frac{\sum_i (x_i - \bar{x})(y_{i+k} - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_{i+k} - \bar{y})^2}}
\]
Simulated Sine-Gaussian $Q=9$, $f_0=554\text{Hz}$ (passed through IFO response function)

- L1-H1 pre-processed waveforms and r-statistic plot
  - integration time $\tau = 20\text{ ms}$,
  - centered on the signal peak time.

- a Kolmogorov-Smirnov (KS) test states the $\{r_k\}$ distribution is NOT consistent with the null hypothesis.
  - there is less than 0.1% probability that this distribution is due to uncorrelated series.

- On to the calculation of the confidence series and of its maximum $C_M(j)$

Noise only (no added signal)

- KS stat = 9.9% significance=37.6% KS passed
Scanning the Trigger Duration $\Delta T$

» Partition trigger in $N_{\text{sub}}=(2\Delta T/\tau)+1$ subsets and calculate $C_M(j)$ (j=1.. $N_{\text{sub}}$)

» Use $\Gamma_{ab} = \max_j(C_M(j))$ as the correlation confidence for a pair of detectors over the whole event duration
Each point: max confidence $C_M(j)$ for an interval $\tau$ wide (here: $\tau = 20\text{ms}$)

Define a cut (pattern recognition?):

2 IFOs:
$$\Gamma = \max_j(C_M(j)) > \beta_2$$

3 IFOs:
$$\Gamma = \frac{\max_j(C_M^{12} + C_M^{13} + C_M^{31})}{3} > \beta_3$$

In general, we can have $\beta_2 \neq \beta_3$

$\beta = 3$: 99.9% correlation probability
Sample Performance

Test on 26 simulated events:
Sine-Gaussians $f_0=361$ Hz or $f_0=554$ Hz ; $Q=9$
Gaussians $\tau = 1\text{ms}$
ETGs as in S1 (~1 Hz single rate)
h$_0$=signal peak amplitude with 50% efficiency for triple coincidence event analysis (S1-style)
$\beta_3=3$ cut in the r-statistics test

Out of the 26 test points:
No background event passes the r-statistic test

All pass r-statistics test for $h_{\text{peak}} \geq h_0$

⇒ background suppression at no cost for sensitivity (so far)

⇒ ETG thresholds can be lowered (sensitivity increase?)
Summary & Outlook

r-statistic test for cross correlation in time domain – allows to:

» Assign a confidence to coincidence events at the end of the burst pipeline
» Verify the waveforms are consistent
» Suppress false rate in the burst analysis, allowing lower thresholds

In progress:

» Method tune-up on hardware injections
» Ongoing investigation with simulated signals
» Exploring implementation in frequency domain
» Coordination of the test implementation with the externally triggered search
  (see talk by Mohanty)
Simulated Sine-Gaussian $Q=9$, $f_0=361$Hz (passed through IFO response function)

Simulated 1 ms Gaussian (passed through IFO response function)

KS stat = 41.2%
significance < 0.1% KS fails

Max confidence:
$C_M(t_0) = 16.6$
at $t_2 - t_1 = 2.9$ ms

KS stat = 68.4%
significance < 0.1% KS fails

Max confidence:
$C_M(t_0) = 10.7$
at $t_2 - t_1 = 0$ ms
Background

Added signal (50% efficiency level)