Estimating statistical significance of the candidate events in LSC compact binary coalescence search.

Ruslan Vaulin[1]

for the

LIGO Scientific Collaboration

[1]University of Wisconsin-Milwaukee

GWDAW 12, Dec 13-16, 2007 Cambridge, MA

General description of the search (see also Drew Keppel's presentation):

- The search is based on match filtering.
- The current analysis pipeline consists of two main stages with multiple cuts designed to filter out the noise events.
- The output of the pipeline is a set of the candidate event triggers.

Some of the open issues (see also Romain Gouaty's presentation):

- The final output contains both noise and signal.
- One would like to have a quantitative estimate of the statistical significance for each of the candidates.
- The data streams generated by the LIGO detectors are dominated by non-gaussian, non-stationary noise which is difficult to model.

Estimating the significance of the candidate

- The key question:

Given a candidate event trigger “c” what is the probability that it is a GW signal “h” ?

- The related quantity p(c,h) is the probability that the signal “h” is contained in the data and it resulted in the candidate event trigger “c”.
- Using Bayes’ theorem we can express the probability we are interested in as

$$p(h|c) = \frac{p(c|h)p(h)}{\int p(c|h')p(h')dh' + p(c|0)p(0)}$$

where

$p(c|h)$ is the detection probability

$p(c|0)$ is the false alarm probability

$p(h)$ is probability that there is “h” GW signal in the data

$p(0)$ is probability that there is no GW signal in the data

- Defining the new quantity,

likelihood ratio: $$\Lambda = \frac{p(c|h)}{p(c|0)}$$

- The formula becomes

$$p(h|c) = \frac{\Lambda(c,h)p(h)}{\int \Lambda(c,h')p(h') + p(0)}$$

- Naturally, candidates can be ranked based on the quantity,

$$\Lambda(c) = \int \Lambda(c,h)p(h)dh$$

Calculation of detection and false alarm probabilities

- Our main task is to calculate the total detection probability, $\int p(c|h)p(h)dh$ and the false alarm probability, $p(c|0)$ for each of the candidates.
- We "measure" these quantities by sampling the results of the multiple runs of the search pipeline over the two types of data

a) The actual data containing GW signal (simulated by software injections)

b) The actual data containing only noise (time slides)

- Thus, we arrive at a phenomenological method of estimating the likelihood of the candidates to be a GW signal.

Parameter spaces of candidate events and GW signals

- Candidate event triggers and GW signals "live" in their corresponding multidimensional parameter spaces.

Candidate event trigger parameters:

- ✓ Signal to Noise Ratio (SNR)
- ✓ Value of χ^2 - test
- ✓ Chirp Mass
- ✓ Effective distance
- ✓ Value of ellipsoid coincidence test
- ✓ …

GW signal parameters:

- ✓ Binary's masses
- ✓ Binary's spins and angular momentum
- ✓ Inclination, polarization angles
- ✓ Physical distance
- ✓ Sky position

- The search pipeline maps one space into the other.
- The mapping is non-linear with stochastic terms due to the non-gaussian, non-stationary noise in LIGO data.

Introducing ε-ball

- We calculate the detection and false alarm probabilities only for the candidates.

- In this case the main practical question one has to answer is:

Given a candidate with the set of parameters $\{\lambda_c^i\}$ how many foreground (background) events were found in its vicinity?

- The vicinity (neighborhood) is usually defined in terms of the ε-ball.
- Having no natural metric on the parameter space of candidate triggers we make the following anzats for it

$$\{\lambda_e\} : \frac{1}{N_p} \sum_i \left(1 - \frac{\lambda_e^i}{\lambda_c^i}\right)^2 < \epsilon^2$$

where $\{\lambda_e^i\}$ is the set of the (background or foreground) event parameters

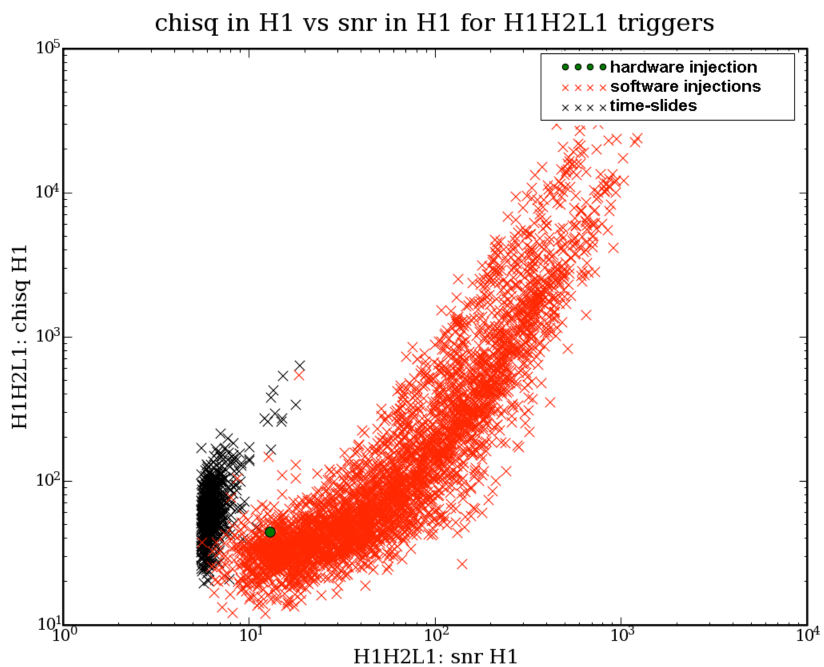
and N_p is the number of dimensions of the parameter space.

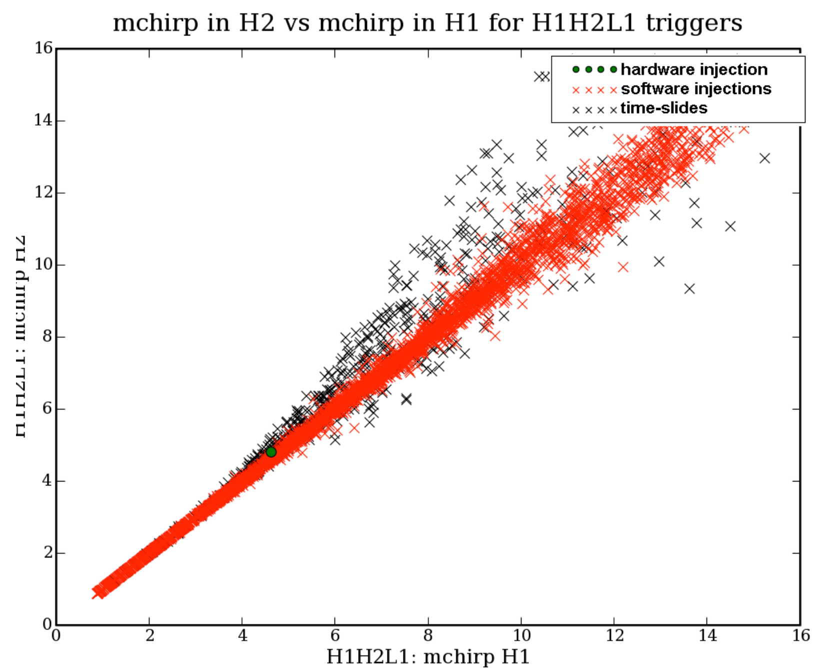
Distributions: time-slides vs. software injections

The plots below should be treated as illustrative examples only!

On the plots: typical distributions of software injections (red) and time-slides (black) triggers in LIGO data in the (M_{chirp}, SNR, χ^2) parameter space.

The green circle is a hardware injection.

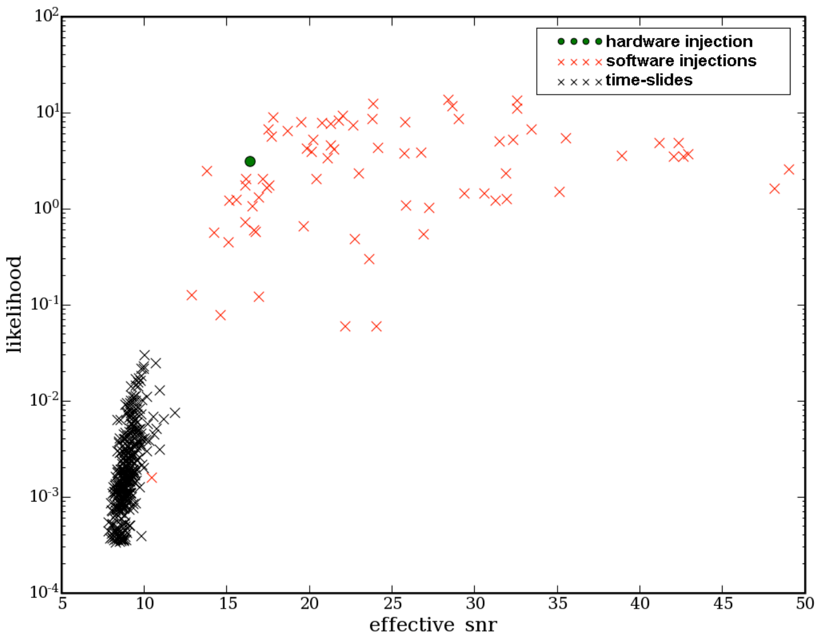




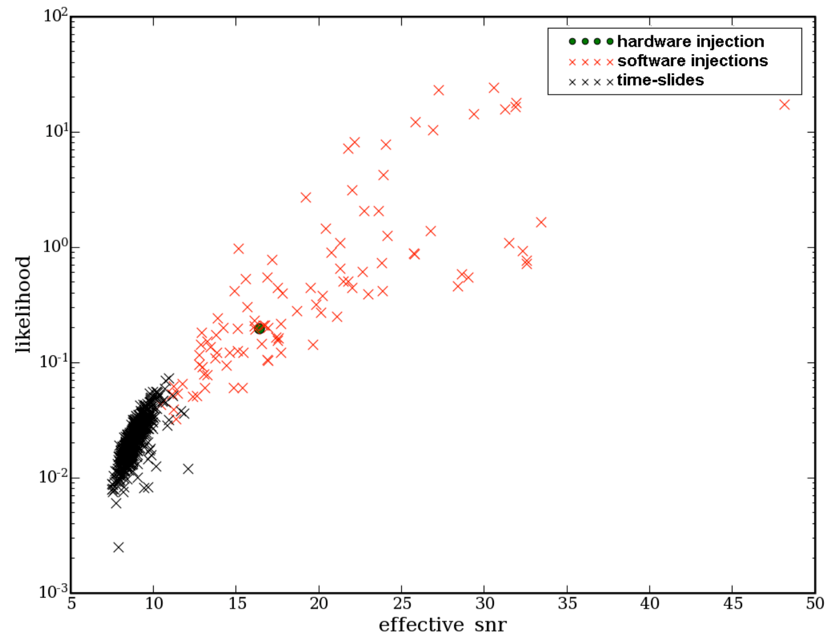
Example: likelihood for hardware injection

In the calculations illustrated by the plots below we used the following parameters:
SNR, χ^2-test and M_{chirp} as measured in each of the interferometers.
Altogether nine parameters. The only adjustable parameter was ε-radius.

Run with ε=0.3



Run with ε=0.5



Summary of the main features of the method

- This is a work in progress but …
- There are indications that the method can be used as a sensitive signal/noise discriminator.
- It allows to consider multidimensional distributions.
- The method is intrinsically phenomenological.
- it uses the actual data without relying on any theoretical models of noise.
- It is “flexible” and “tunable”

Future developments and investigations

- Integrating it into the existing pipeline.
- Thorough investigation of the parameter space of candidate events.
- Consider other metrics for ε-ball.
- Assess the dependence of the final results on the properties of the population of software injections.
- Consider using this framework for parameter estimations of the candidates.