r statistics
for time-domain cross correlation
on burst candidate events

Laura Cadonati

LIGO-MIT

LSC collaboration meeting, LLO march 03

LIGO-G030084-00-Z

LIGO

Preface

- The burst analysis pipeline uses Event Trigger Generators (ETGs) to flag times when “something” occurs (burst candidates).
- Triggers from different interferometers are brought together in coincidence.
- All cuts in the pipeline are generous, in order to preserve sensitivity: we don’t really know what we are looking for…

- We want to use the full power of a coincident analysis:
  - » What confidence can we put on the coincidence candidate?
  - » Are the waveforms consistent?

- The ETG outputs (Δt, BW, amplitude/power/SNR) do not provide enough information, we need to go back to the time series

Cross correlation in time domain

- Typical duration of coincident events from the burst pipeline: tens of seconds from SLOPE, 0.5-1 sec from TFCLUSTERS.

- Load 5 sec of data from the two interferometers, 2 sec before event start
- High-pass at 100 Hz
- Whitening/line removal: train an adaptive filter (linear predictor filters – studied by Shourov Chatterji) over the first second of loaded data, apply to the rest. This is fundamental to bypass the problem of non-stationary lines.
- Use the r-statistics over "small" time intervals and implement time lags to evaluate the linear cross correlation.

Pearson's r statistics

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}}$$

NULL HYPOTHESIS:
the two (finite) series {xi} and {yi} are uncorrelated

⇒ Their linear correlation coefficient (**Pearson's r**) is normally distributed around zero, with standard deviation = 1/sqrt(N)
where N is the number of points in the series (N >> 1)

S = erfc (|r| sqrt(N/2)) = double-sided significance of the correlation
probability that |r| should be larger than what measured, if {xi} and {yi} are uncorrelated

C = - log10(S) = confidence that the null hypothesis is FALSE
(that is: confidence that the two series are in fact correlated)

Reference: Numerical Recipes in C

What does a large confidence mean?

Confidence C = 1 ⇔ significance S = 0.1 (10%)

3 different ways to say the same thing:
⇒ 10% probability that the null hypothesis is true, OR
⇒ 90% probability the hypothesis of no correlation is false (events are correlated) OR
⇒ 10% probability this is a false coincidence

If we can assign a confidence to the coincident event pair, we can define a cut on it. The cut defines the false probability in the analysis.

confidence=1	==> significance=0.1 (10%)	==> 90%	correlation probability
confidence=1.3	==> significance=0.05 (5.0%)	==> 95%	correlation probability
confidence=1.6	==> significance=0.025 (2.5%)	==> 97.5%	correlation probability
confidence=2	==> significance=0.01 (1.0%)	==> 99%	correlation probability
confidence=3	==> significance=0.001 (0.1%)	==> 99.9%	correlation probability

Assign a confidence to the pair of coincident events (burst candidate)

We suspect a burst happened in a 0.5 sec time window. We do not know how long the burst is (1 ms?) WHEN it happens within the interval and at what delay between sites.

Let's shift one of the two series by one data point at a time and calculate a series of:

- coefficients r_k
- significances S_k
- confidences C_k

$$r_k = \frac{\sum_i (x_i - \bar{x})(y_{i+k} - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (y_{i+k} - \bar{y})^2}}$$

…and look for a peak in confidence.

In order to reduce fluctuations, decimate the confidence series by 4.
Max confidence = confidence in the correlation of the event pair
Time shift for max confidence = delay between IFOs

Magic number: 10 ms

LIGO

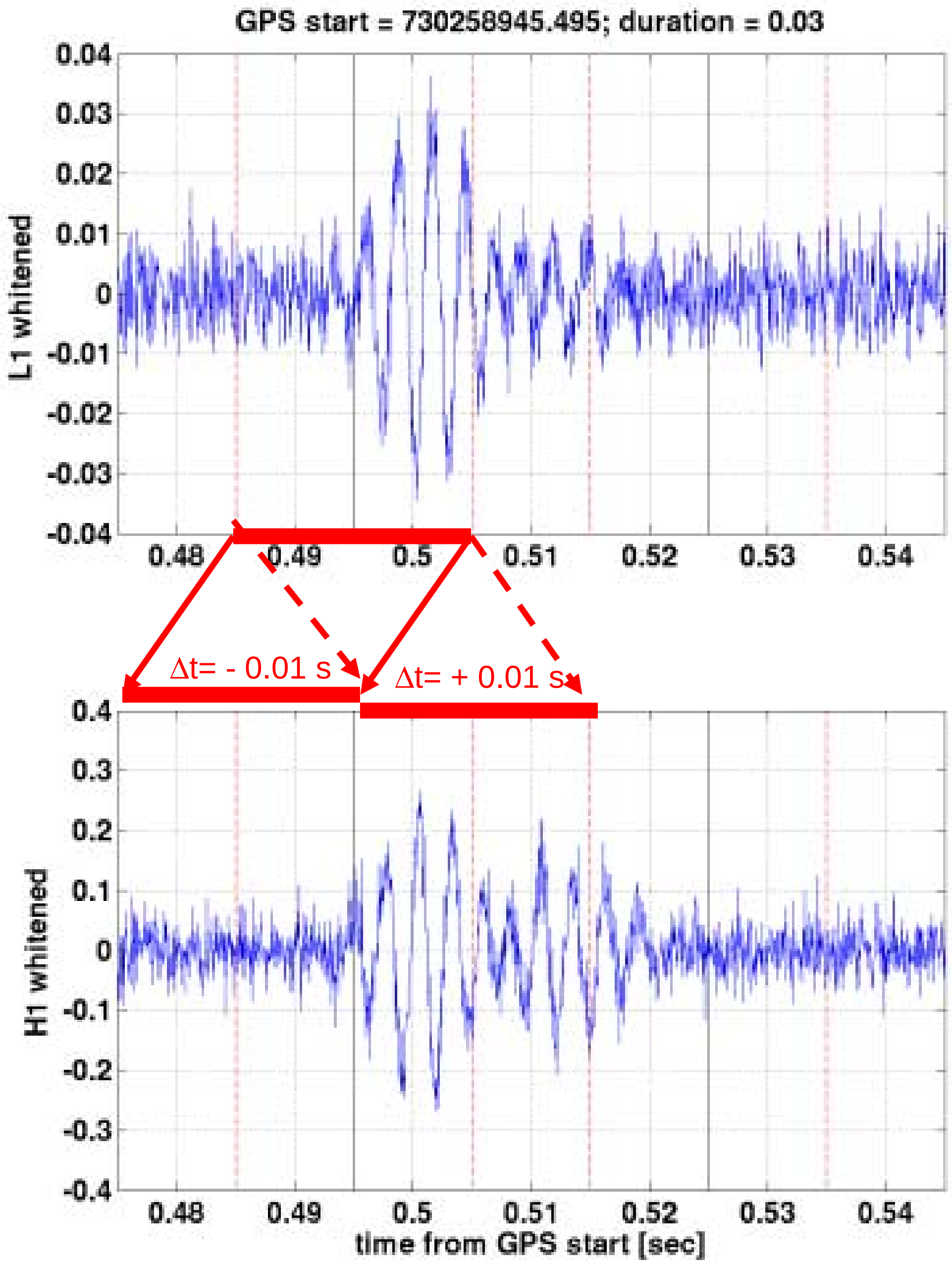
Example: S2 hardware injection From Feb 25, 2003 – H1-L1 (sine gaussian, 361Hz, Q=9)

Divide in 20 ms data segments, shift by +/- 10 ms and find C_{max}
Move 10 sec forwards and repeat the iteration

This way we allow for bursts with separation up to 10 ms at any point within the trigger duration.

20 ms intervals work well for burst injections – needs some tuning, though.

If the segment is much larger, the correlation washes out.



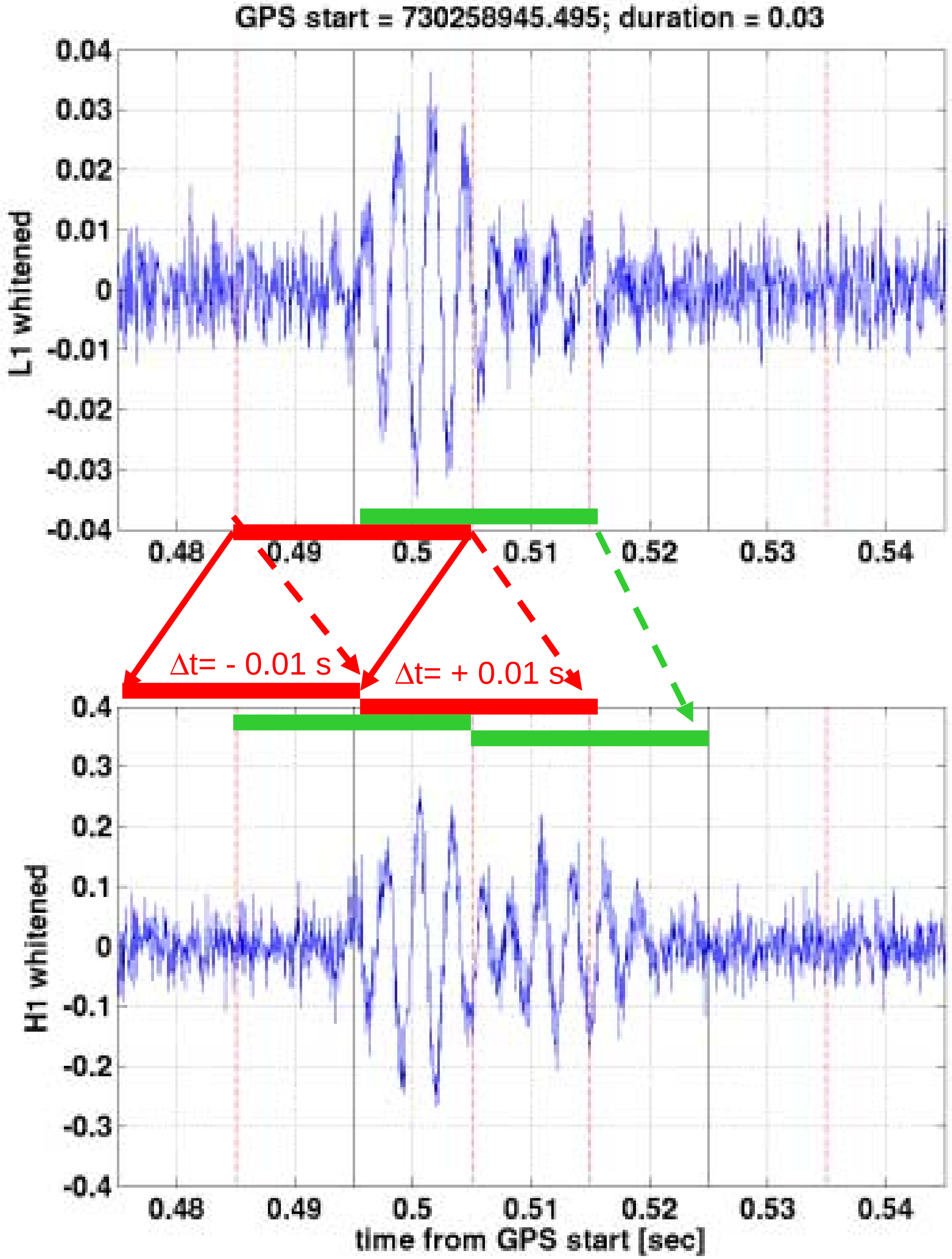
Example: S2 hardware injection From Feb 25, 2003 – H1-L1 (sine gaussian, 361Hz, Q=9)

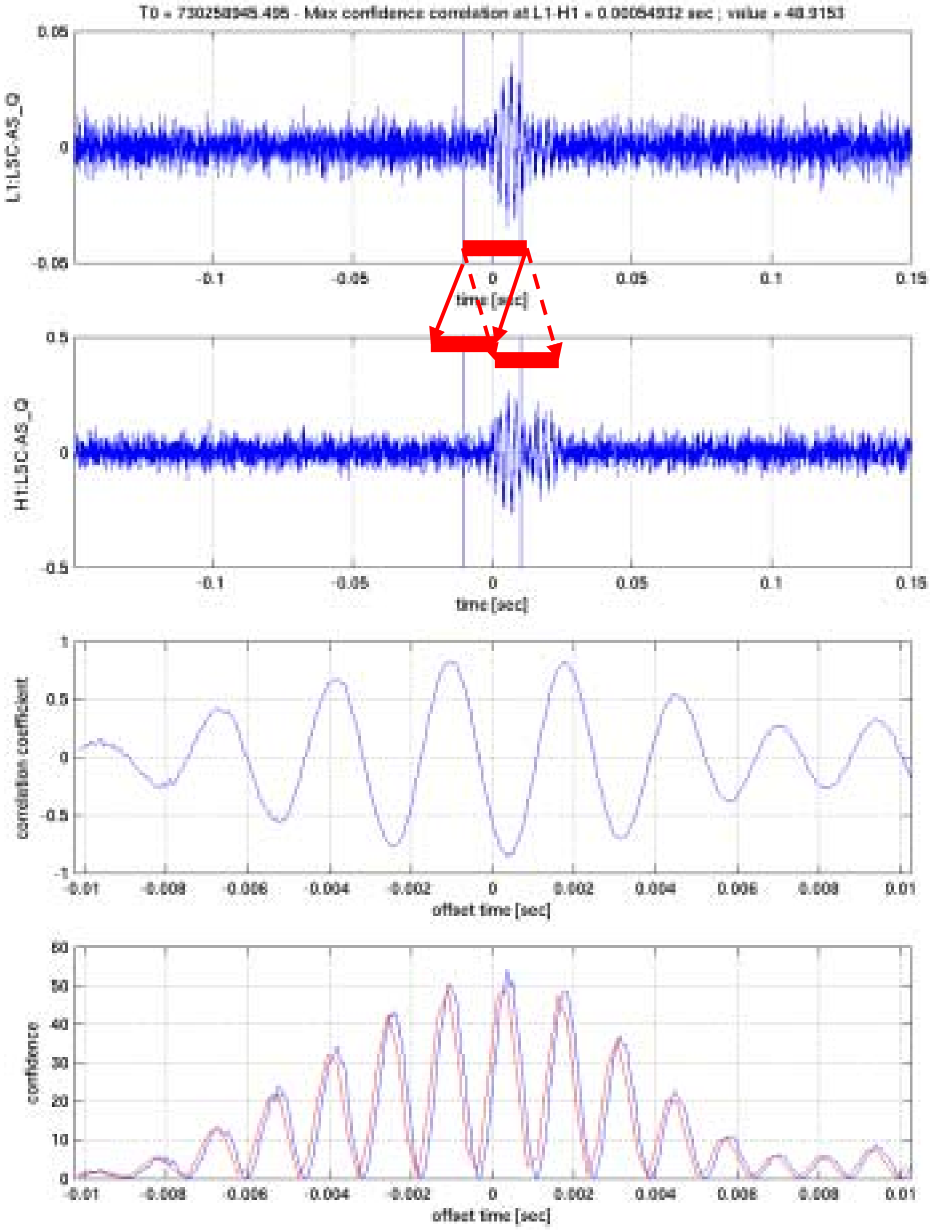
Divide in 20 ms data segments, shift by +/- 10 ms and find C_{max}
Move 10 sec forwards and repeat the iteration

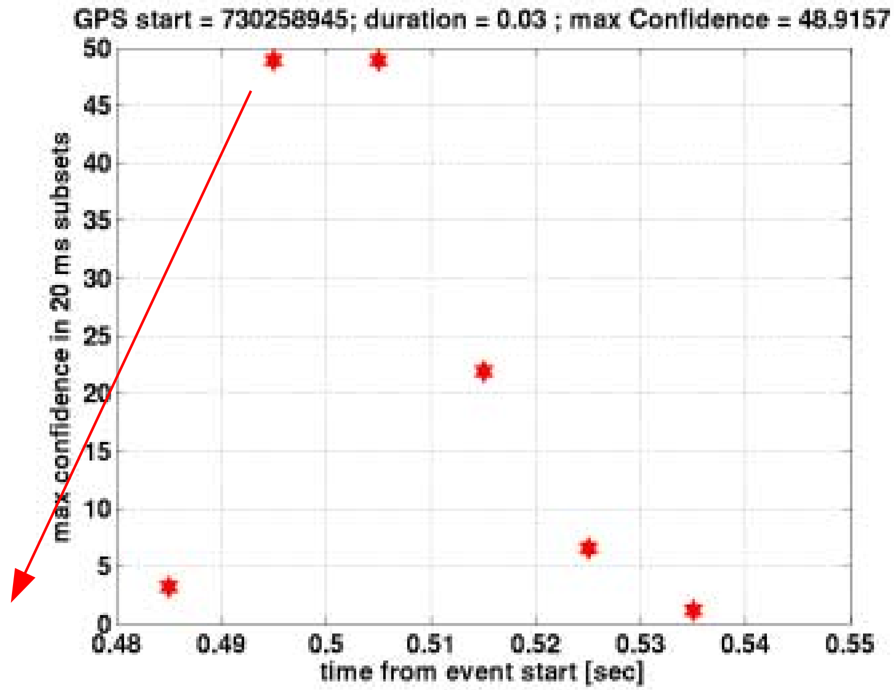
This way we allow for bursts with separation up to 10 ms at any point within the trigger duration.

20 ms intervals work well for burst injections – needs some tuning, though.

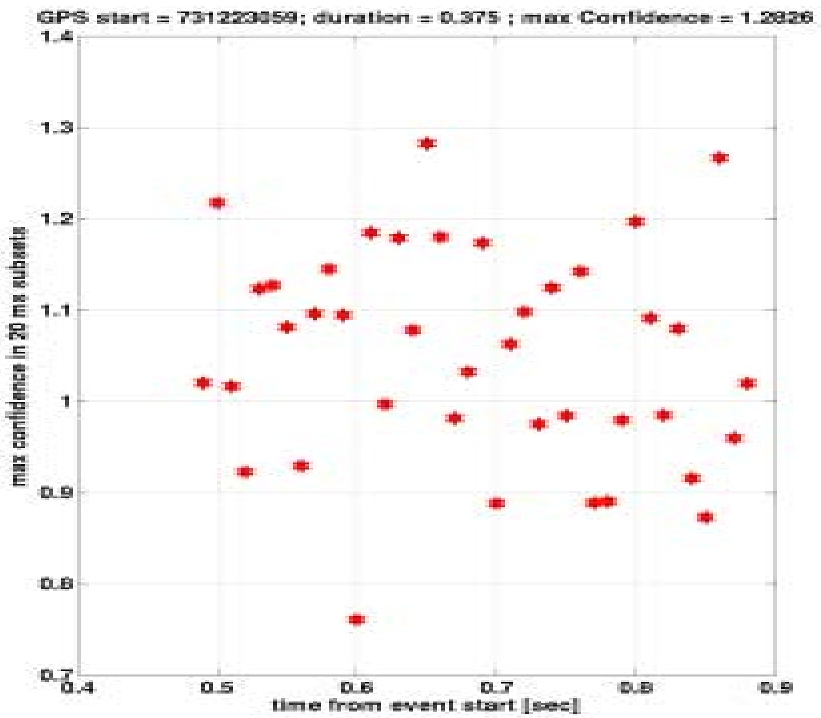
If the segment is much larger, the correlation washes out.

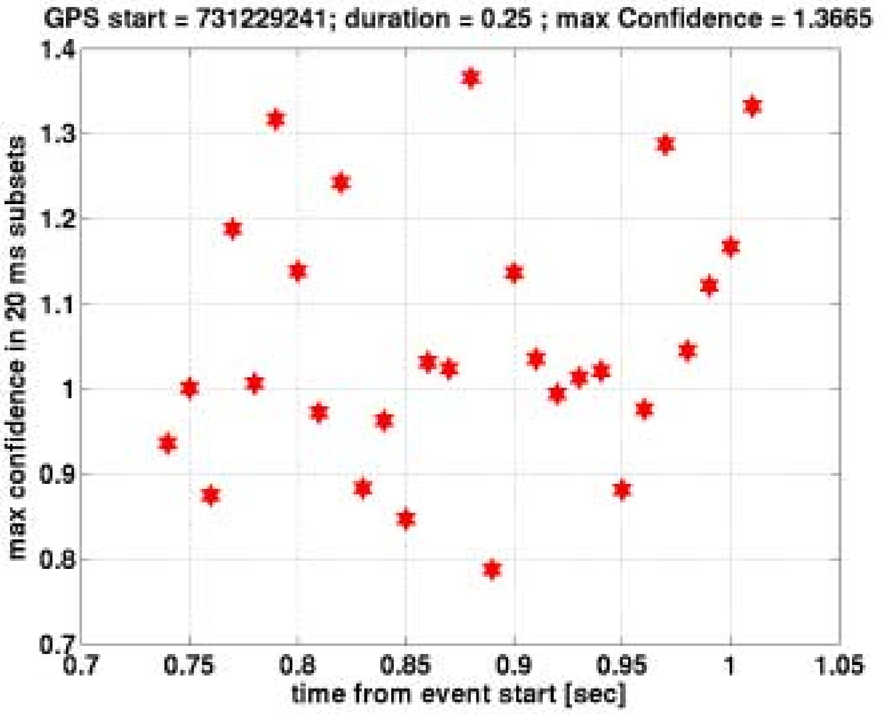



Max confidence correlation at L1-H1
L1:LSC-AS_Q
H1:LSC-AS_Q
time [sec]
correlation coefficient
offset time [sec]
confidence


GPS start = 730258945; duration = 0.03 ; max Confidence = 48.9157
max confidence in 20 ms subsets
time from event start [sec]

Background events

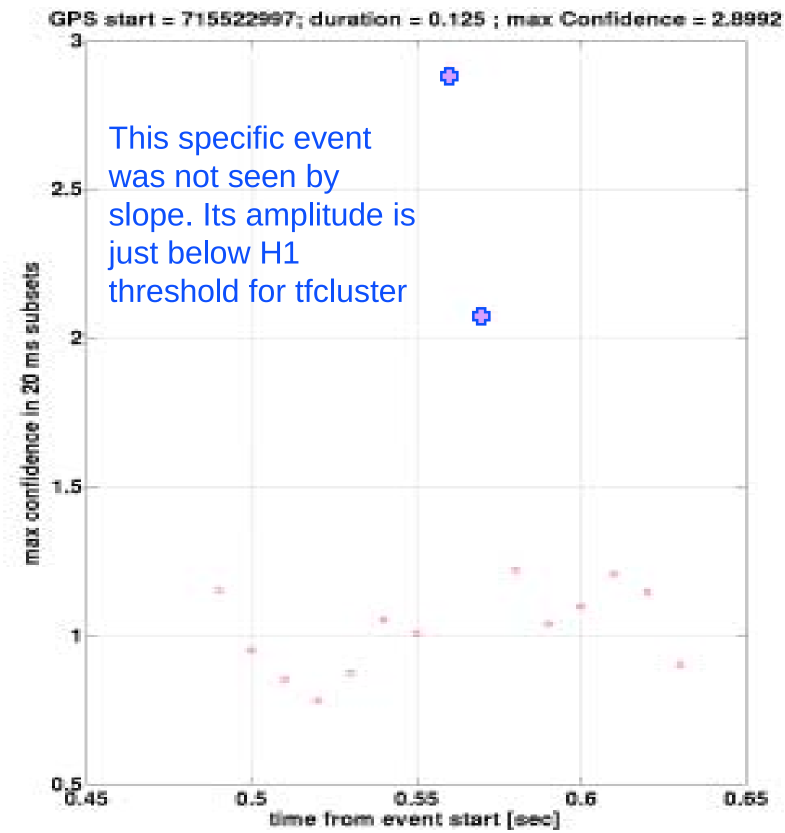




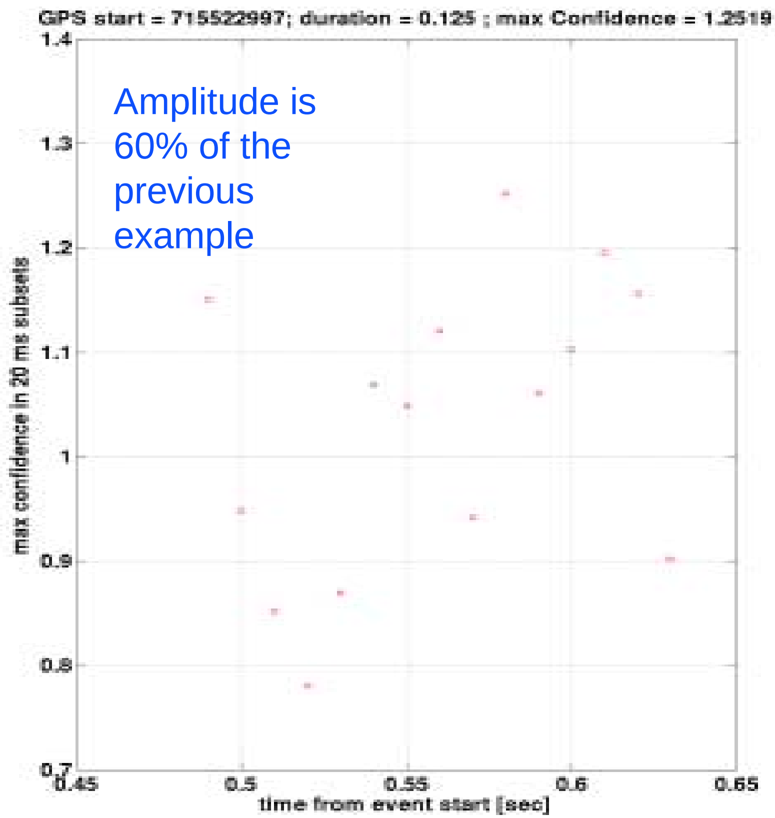
2 instances of L1H1 coincidence in the S2 playground: max confidence = 1.3 (significance = 0.05)

Need to sample more of these to tune the cut. 1.6-2 seems reasonable on the basis of what seen so far.

S1 software injections (1 ms Gaussian)



This would pass confidence cut

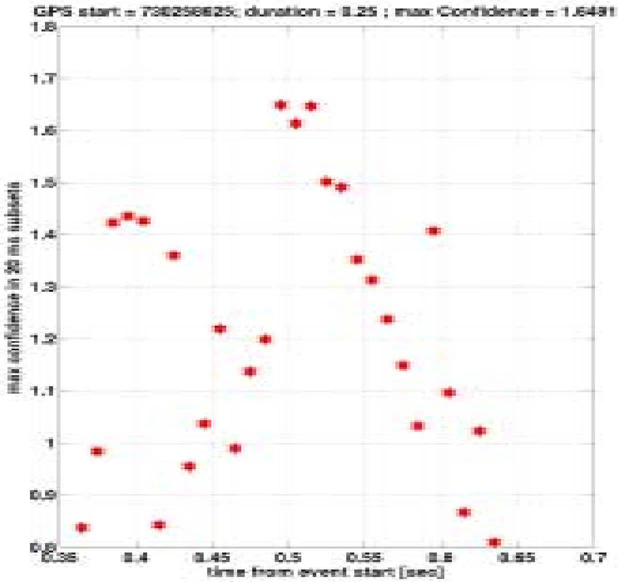


No correlation here

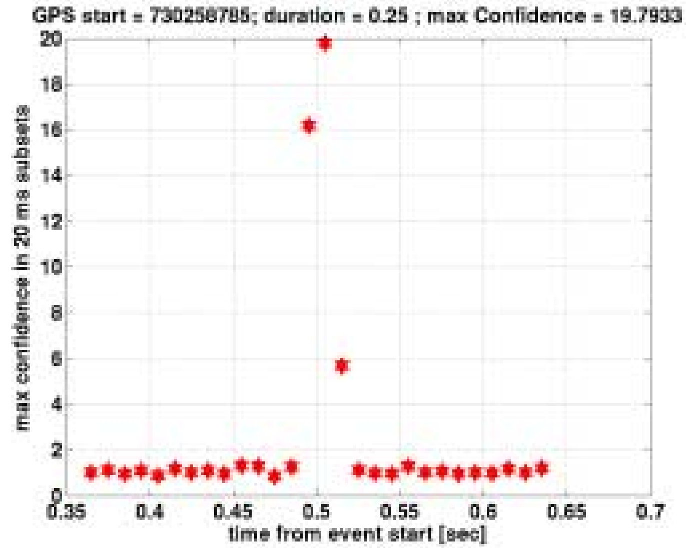
S2 hardware injections (Feb 25, 2003)

GPS	cycle	frequency	L1tfclu	H1tfclu	L1-H1	confidence	visible peak?
730258565	1	100	0	0	-2.3ms	0.8	NO
730258585	1	153	1	0	-2.3ms	11.6	YES
730258605	1	235	1	0	8.7ms	2.2	YES
730258625	1	361	-1	0	8.4ms	1.6	NO
730258645	1	554	0	0	-2.8ms	1.0	NO
730258665	1	850	0	0	-6.5ms	1.2	NO
730258685	1	1304	-1	0	-6.1ms	2.0	YES
730258705	1	2000	0	0	-7.0ms	1.8	NO
730258725	2	100	-1	0	9.6ms	9.7	YES
730258745	2	153	1	1	1.6ms	36.8	YES
730258765	2	235	1	1	0.1ms	23.7	YES
730258785	2	361	1	1	0.2ms	19.3	YES
730258805	2	554	1	-1	-0.1ms	12.9	YES
730258825	2	850	-1	0	2.1ms	3.0	YES
730258845	2	1304	1	0	0.4ms	14.6	YES
730258865	2	2000	1	0	0.1ms	6.6	YES
730258885	3	100	0	0	3.5ms	43.0	YES
730258905	3	153	1	1	-1.5ms	67.6	YES
730258925	3	235	1	1	-3.7ms	55.8	YES
730258945	3	361	1	1	0.6ms	52.3	YES
730258965	3	554	1	1	-0.9ms	42.3	YES
730258985	3	850	-1	1	0.4ms	25.8	YES
730259005	3	1304	1	1	0.4ms	31.0	YES
730259025	3	2000	1	1	0.1ms	25.9	YES

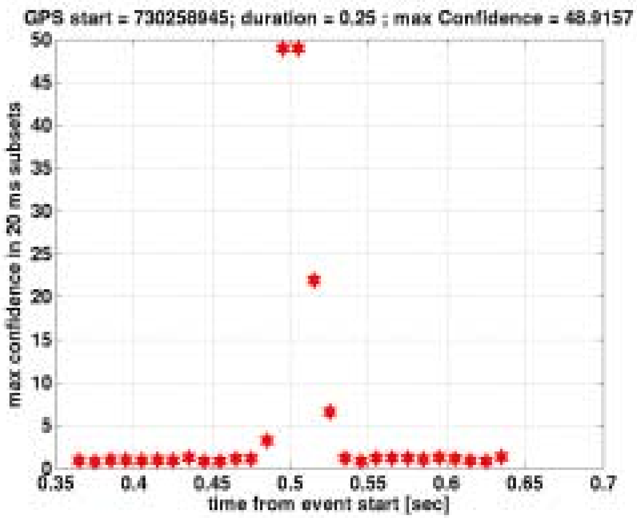
S2 hardware injections (02/25/03) (554 Hz Sine Gaussian)



554Hz , weak
Not detected



554Hz
Power x10
detected



554Hz
Power x100
detected

Summary

Suggested method: cross correlation in time domain.

- » Assigns a confidence to coincidence events at the end of the burst pipeline.
- » Verifies the waveforms are consistent.
- » Computationally expensive (present MATLAB implementation: 10 minutes for a 0.5 sec event – could do better), but manageable, since it does not act on the raw data flow but to a finite number of (short) time intervals.
- » Reduces false rate in the burst analysis

- » TO-DO list:
  - – Run over a suite of S2 playground events
  - – Tune a confidence cut (S1 as playground for S2?)
  - – Verify effect on burst detection sensitivity, with burst Montecarlo
  - – Try more interesting waveforms