PySTAMPAS - A long-duration transients GW pipeline

Adrian Macquet, ARTEMIS
Marie-Anne Bizouard, ARTEMIS
Michael Coughlin, University of Minnesota

Sources of long duration GW transients

Many potential sources, but rough modelling. Use of several waveform models to estimate the pipeline sensitivity :

- van Putten accretion disk instabilities and fragmentation (ADI) [1]
- Rotational instabilities in proto-neutron star (PNS) remnants [2]
- Proto-neutron star convection [3]
- Fallback accretion on neutron stars [4]
- Instabilities in central magnetars [5]
- BNS post-merger
- Neutron star r-modes

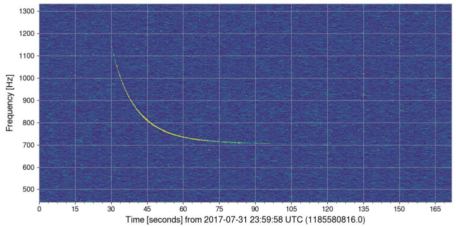
[1] van Putten, PRL 87, 091101 (2001)
[2] Ott et al, Astrophys. J. 625, L119 (2005)
[3] Kotake et al, Astrophys. J. 697, L133 (2009)
[4] Piro and Thrane, Astrophys. J. 761, 63 (2012)
[5] Corsi and Meszaros, Astrophys. J. 702, 1171 (2011)

Signals have expected duration of 10 - 1000s

Long duration GW transients searches

- Potential source waveforms are poorly known : no matched filtering.
- General principle : look for excess of power in time-frequency maps
  - Use cross-correlation between 2 detectors - Need to take account of sky localization of the source.
  - 3 detector pairs : H1L1, V1H1, L1V1
  - Background noise is generated by time-shifting the data between detectors.
- STAMPAS developed in the 00s in LVC is one of the pipelines used to search for long-duration signals.
  - Produce coherent TF-maps for different potential sky positions.
  - Apply a seed-based clustering algorithm to build triggers.
  - Compute a coherent SNR that is used as detection statistic.

TF-map containing an injected signal (ISCOchirpB)



Motivation of this development

- PySTAMPAS (python) is a new version of STAMPAS (matlab).
- STAMPAS bottleneck : coherent cross-correlation and clustering
  - Limited number of time-slides for background estimation (FAR ~ 1 / 50 yr)
  - Limited number of sky positions tested → loss in sensitivity (especially for high freq signals)
- PySTAMPAS implements a new method and new features :
  - Lonetrack method [1] : separate the analysis into 2 stages (incoherent & coherent)- computationally cheaper.
  - Multi-resolution time-frequency maps for better signal reconstruction.
  - New detection statistic to improve detection efficiency.
- Goals :
  - Reduce computational cost to be able to reach 5 σ false-alarm probability.
  - Increase detection sensitivity

[1] Coughlin and Thrane, PRD 89, 063012 (2014)

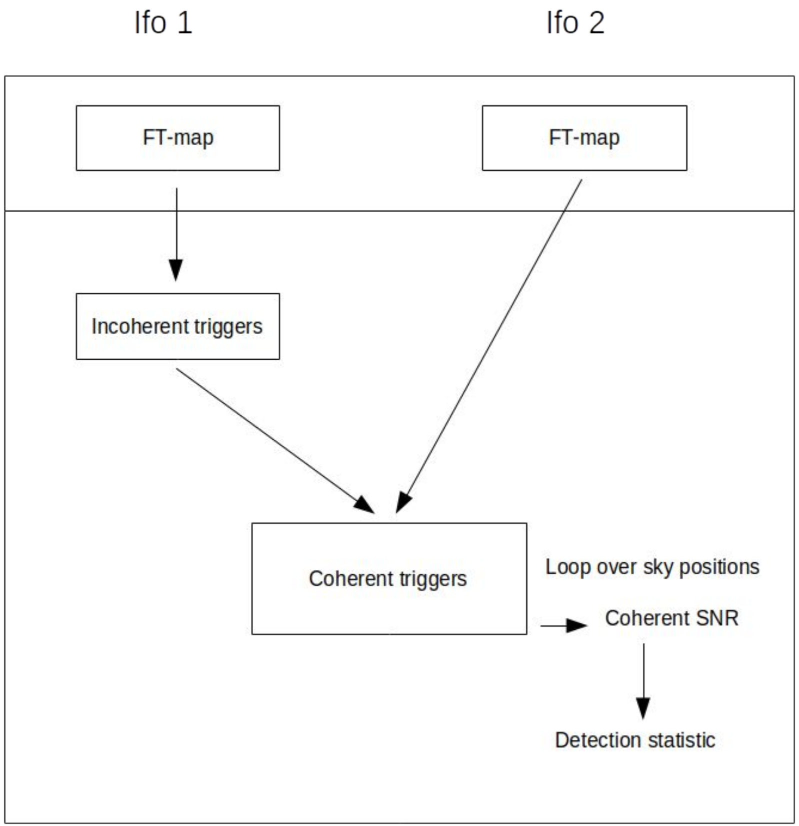
Pipeline flowchart

- TF-maps are built from strain time series.
- Clustering is performed on single detector TF-maps.
- Pixels from ifo 1 clusters are matched with corresponding pixels from Ifo 2 - cross correlation produces coherent variables (“statistic”).
- Output : coherent triggers

Preprocessing

<u>Stage 1</u> : single-ifo clustering

<u>Stage 2</u> : cross-correlation



Coherent analysis

A GW event should induce a signal which is correlated in every detector of the network*, contrary to a transient noise event. → Compute correlation product between two detectors (IJ).

- Pixels in ft-maps represent the value of the FFT of the signal for a given time segment and frequency (whitened by the PSD): $\tilde{y}(t;f)$

- Cross-correlation needs to take into account the delay of arrival of the signal between the detectors τ : $\mathrm{SNR}(t;f|\hat{\Omega}) \equiv e^{2i\pi f\tau}\tilde{y}_I^\star(t;f)\tilde{y}_J(t;f)$
  - ➢ $\hat{\Omega}$ represents the unit sky position vector of the source, which induces the delay τ.
  - ➢ To perform an all-sky search, it is necessary to test multiple sky positions and keep the highest SNR. Computational gains allow us to test more positions, thus reconstruct signals better.

*The amplitude will depend on the detector orientation and intrinsic sensitivity, but that does not affect the phase.

Detection statistic

Clusters of excess power pixels are selected on single-detector data. Thus loud glitches can be selected despite having no counterpart in the second detector.

- Define a “residual energy” for each detector $E^I(t; f, \hat{\Omega}) \equiv \mathrm{SNR}(t; f, \hat{\Omega}) - |\tilde{y}_I(t; f)|^2$
  - Can be seen as the “incoherent” component
  - Should be small in both detectors only for a coherent signal.
- Construct a statistic that takes into account both the SNR and these residual energies :

$$\Lambda \equiv \frac{\mathrm{SNR}_\Gamma}{\mathrm{SNR}_\Gamma + E^I_\Gamma + E^J_\Gamma}.$$

$\mathbf{\Gamma}$ means the quantity is summed over all pixels

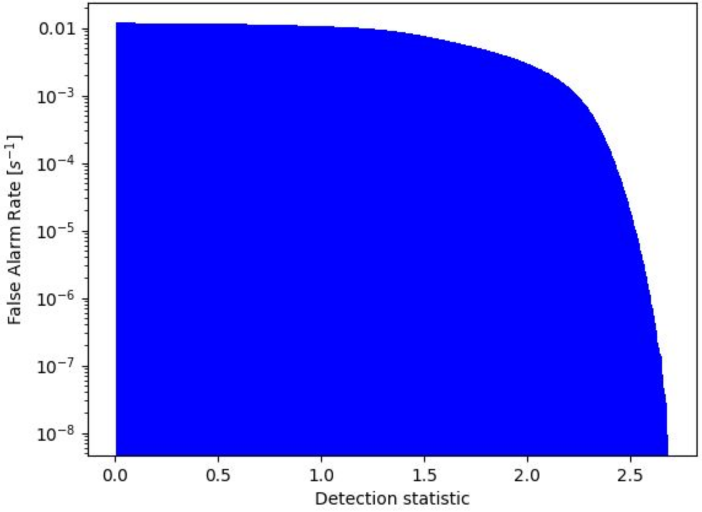
$\Lambda \rightarrow 1$ for a coherent event as we expect $\mathrm{SNR}_\Gamma >> E^0_\Gamma$.
For commodity we use $p_\Lambda \equiv -\log(|1-\Lambda|)$ to estimate the significance of an event.

Performances of the pipeline - Background study

We simulated a search over 2 weeks of gaussian noise colored with O2 nominal PSD.

The distribution of background events is estimated by performing time-slides of one detector data stream with respect to the other. 1280 time-slides are done, simulating 50 years of data.

The distribution of background triggers follows a Gaussian distribution. The cumulative distribution allows to assign a false-alarm rate to a given trigger.



False-alarm rate as a function of the detection statistic

Performances of the pipeline - Efficiency comparison

Waveform	Dist@50% STAMPAS (Mpc)	Dist@50% PySTAMPAS (Mpc)	Gain
adiA	19.52	35.8	+ 83 %
adiB	91.76	129.7	+ 41 %
ISCOchirpA	0.55	0.63	+ 14 %
ISCOchirpB	8.48	7.35	- 13 %
ISCOchirpC	59.43	44.1	- 25 %
NCSACAM_A	24.1	42.7	+ 79 %
NCSACAM_B	23.66	24.1	+ 1 %
NCSACAM_C	18.70	33.4	+ 78 %
NCSACAM_D	53.19	90.6	+ 70 %
NCSACAM_E	58.02	81.2	+ 40 %
NCSACAM_F	31.09	57.2	+ 84 %

Distance for which 50% of the injections are recovered for a set of waveforms used in long-duration searches.

Injection : waveforms are injected in strain time series with random sky position and polarization angle with several amplitudes.

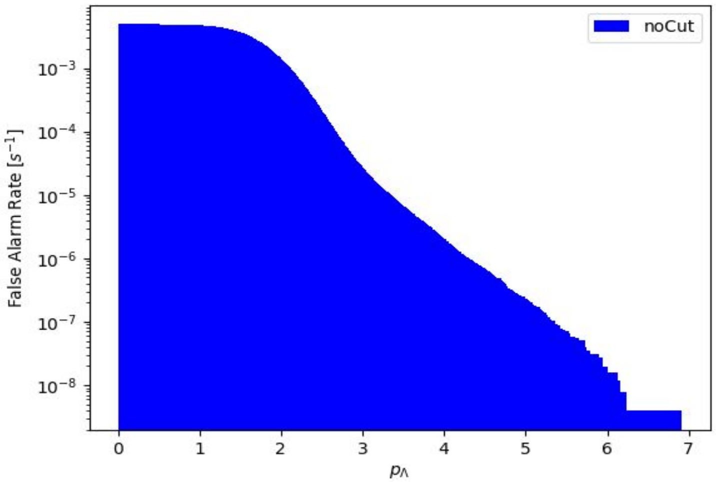
An injection is considered detected if it generates a trigger with FAR < 1e-8 / s.

Comparison with the current STAMPAS implementation shows an increase in detection efficiency for most of the waveforms tested.

PySTAMPAS applied on real GW data

Real data from terrestrial interferometers are more difficult to analyze because of non-Gaussian noise features : instrumental lines, short glitches...

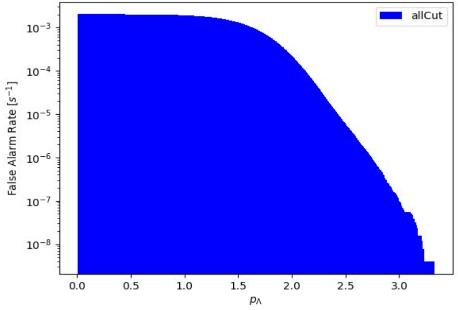
Background study over 2 weeks of data from O2 H1/L1 : the triggers distribution presents a tail that decreases drastically the sensitivity of the search.



False-alarm rate as a function of the detection statistic

Dealing with non-gaussian events

- Frequency notch : mask frequency bins corresponding to instrumental lines.
- Gating : find and remove peaks in strain time-series.
- Data Quality Flags : use auxiliary channels to identify noisy periods.
- Post-processing vetos : identify and remove “suspicious” triggers.
    - ➢ SNRfrac : fraction per time bin of the trigger’s energy
    - ➢ Rveto : measures the “unbalanced” energy between the two detectors (works for aligned detectors).



FAR curve obtained after all these procedures

Challenge : identify features of non-gaussian noise events that distinguish them from a potential signal.

Waveform	Dist@50% STAMPAS (Mpc)	Dist@50% PySTAMPAS (Mpc)
adiB	103.45	140.39
ISCOchirpA	0.55	0.71
ISCOchirpB	7.64	8.58
ISCOchirpC	35.94	55.03
sgC	2.98	20.45

Efficiency comparison on O2 data confirms the sensitivity increase.

Conclusion and perspectives

- Gain in computational performances
  - 2 stages method allows to perform time-slides more quickly.
  - More background simulated and lower false-alarm rates reached.
  - Analysis is faster.
- Increased sensitivity
  - Test more sky positions to better reconstruct coherent signals.
  - Multi-resolution ft-maps can adapt to a wider range of signal morphologies.
  - New coherent statistic helps to distinguish coherent signal from noise.
- The pipeline is currently analyzing O3 data.
- Targeted searches to look for potential GW counterpart in GRBs
- Method paper in preparation
- Open project : develop more intelligent post-processing vetos