PyGRB: A matched filtering triggered gravitational-wave search pipeline

Sebastian Gomez Lopez - July 11th, 2024

La Sapienza Università di Roma
INFN sezione di Roma

SAPIENZA UNIVERSITÀ DI ROMA

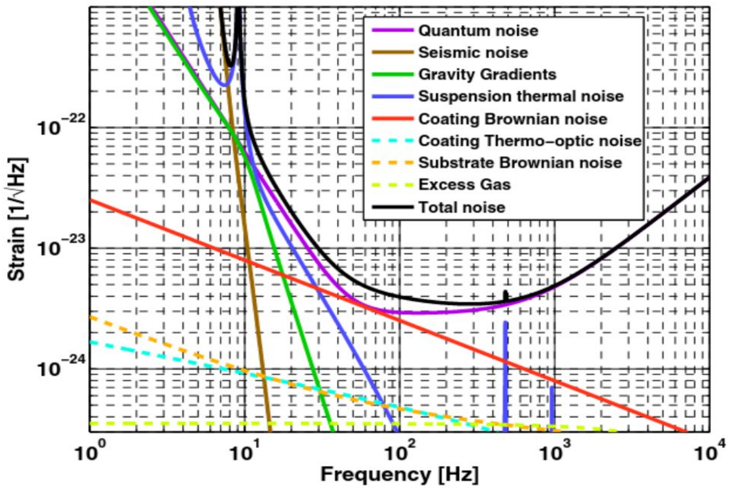
PyCBC

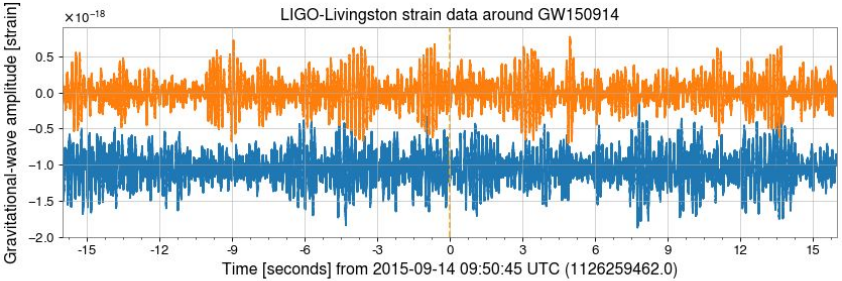
INFN ROMA
Istituto Nazionale di Fisica Nucleare
Sezione di Roma

I’m giving this talk on behalf of active PyGRB developers:

Jacob Buchanan
Marco Cusinato
Erin Vincent
Stephanie Hoang
Marion Pillas
Prasia Pakuni
Francesco Pannarale
Tito dal Canton

However, we are “walking in the shoulders of giants” on this project…


Quantum noise
Seismic noise
Gravity Gradients
Suspension thermal noise
Coating Brownian noise
Coating Thermo−optic noise
Substrate Brownian noise
Excess Gas
Total noise
Strain [1/√Hz]
Frequency [Hz]


LIGO-Livingston strain data around GW150914
Gravitational-wave amplitude [strain]
Time [seconds] from 2015-09-14 09:50:45 UTC (1126259462.0)

The problem of GW astronomy

1. Search for a signal u(t) buried in stationary, frequency-dependent, Gaussian noise n(t).

   NOTE: This talk will focus on the case of signals u(t) of known morphology.

2. Decide whether it is of astrophysical origin or not

Hypothesis 1

$$h(t) = n(t)$$

Hypothesis 2

$$h(t) = u(t) + n(t)$$

Note: not only very sensitive detectors are required, but optimal and computationally affordable data analysis techniques.

-> The optimal linear filter to find q in a noisy frequency dependent data stream h(t) is:

$$\langle h, q \rangle = 4Re\left(\int_0^\infty \frac{h(f) \cdot q^*(f)}{S_n(f)} \cdot df\right)$$

Preprocessed Interferometric data

Known signal model. "Template"

Estimate of the noise power spectral density

$$\langle n(f), n(f') \rangle = \delta(f' - f) \cdot S_n(f)$$

- A “sort of” complex valued noise weighted cross correlation:

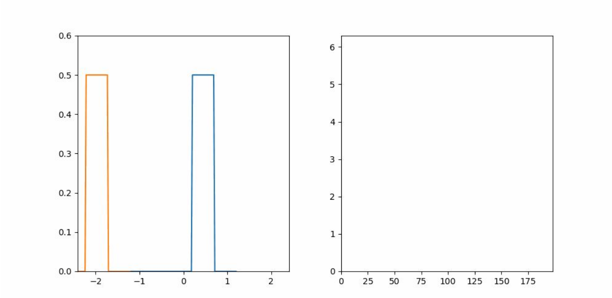
$$\langle h, q e^{i(2\pi f\tau+\phi_0)}\rangle_{\phi_0=\phi_{opt}} = 4 \cdot \left(\left| \int_{-\infty}^{\infty} \theta(f) \frac{h(f)\cdot q^*(f)}{S_n(|f|)} \cdot e^{i2\pi f\tau} df \right| \right)$$

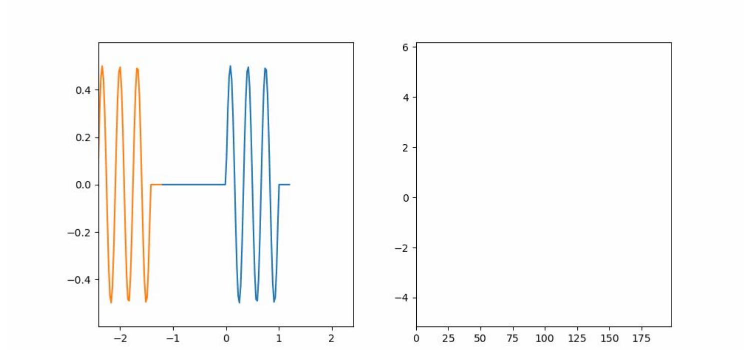
$$\rho(\tau) = \frac{\langle h, q e^{i(2\pi f\tau+\phi_0)}\rangle_{\phi_0=\phi_{opt}}}{\sqrt{\langle q, q\rangle}}$$

Signal-to-noise ratio “SNR”

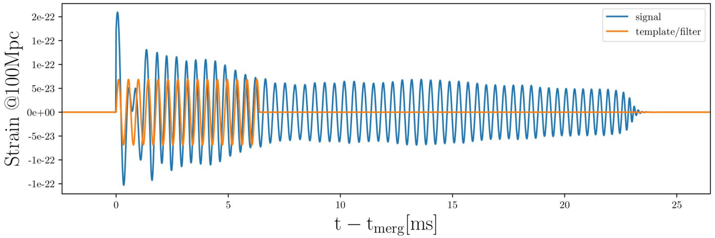
(Its a complex timeseries, not a single scalar quanity)

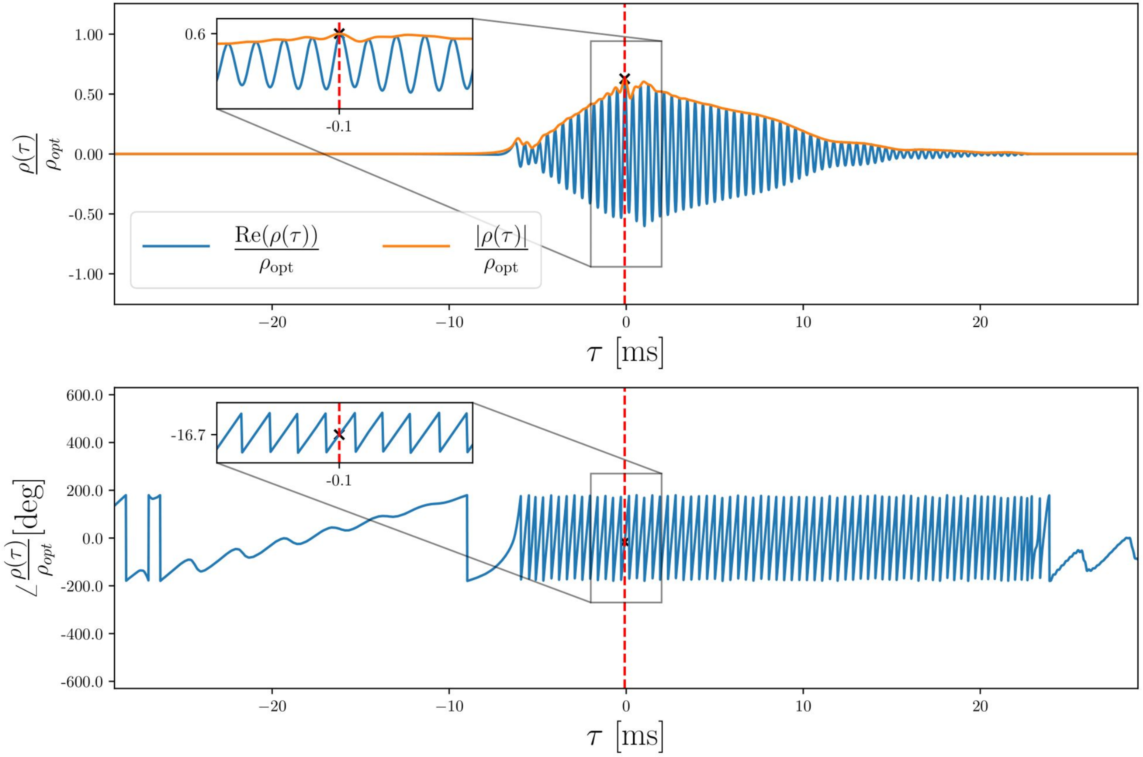
$$\rho(\tau) = \frac{4 \cdot \left| \mathrm{IFFT}\left(\theta \cdot \frac{h\cdot q^*}{S_n}\right)\right|(\tau)\cdot \Delta f}{\sqrt{4 \cdot \sum_{-\infty}^{\infty} \frac{q\cdot q^*}{S_n} \Delta f}}$$





- Gives info about:
  - Signal power
  - Signal phase***





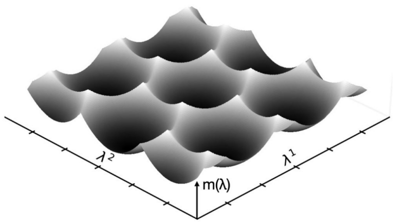
Doing a real search becomes high dimensional problem

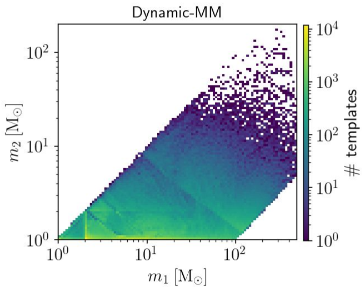
$$\langle h, qe^{i(2\pi f\tau+\phi_0)}\rangle_{a_1,\ldots,a_n,\phi_0=\phi_{opt}} = 4\cdot\left(\left|\int_{-\infty}^{\infty}\theta(f)\frac{h(f)\cdot q^*_{a_1,\ldots,a_n}(f)}{S_n(|f|)}\cdot e^{i2\pi f\tau}df\right|\right)(\tau)$$

$$q_{a_1,\ldots,a_n}(t)$$

Template banks

- Can not be defined continuously at every point of $a_1 \times a_2 \ldots \times a_n$
- Geometrically motivated ways have been invented to sample that space discretely without losing too much SNR. (Using manifold theory)





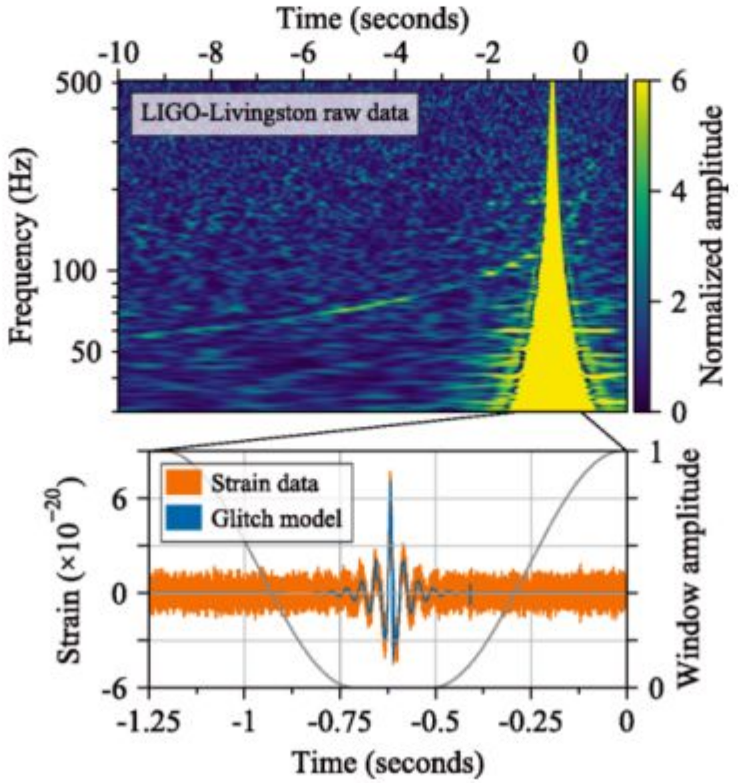
B. Allen 2021

LVK collaboration 2023

But GW interferometric data deviates from our initial assumption rather frequently:

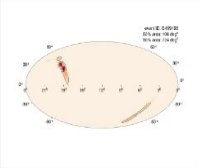
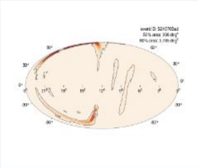
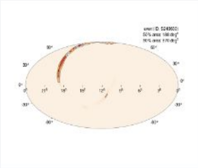
- Non-gaussian transients appear
- Noise artifacts look like GW signals, and can trigger templates with high SNR values.

- Autogating using CAT1, CAT 2 vetoes produced by detector scientists
- Chi squared vetoing



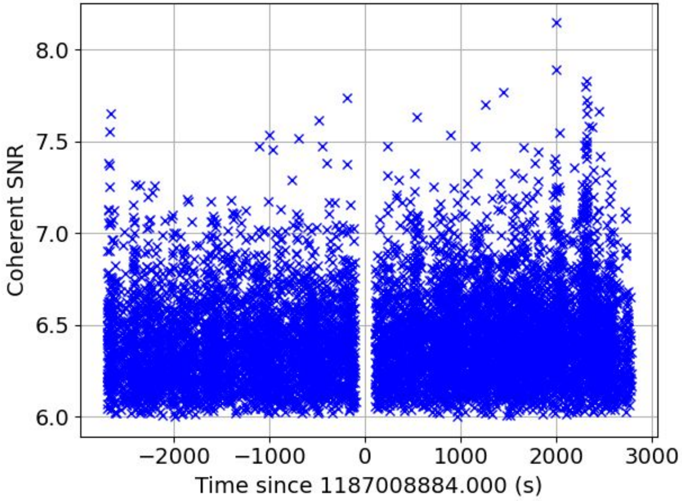
Source: GW170817 discovery paper

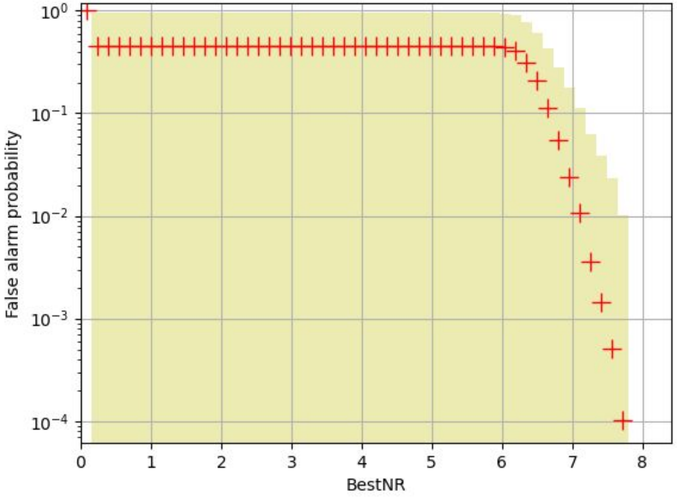
Assessing astrophysical significance requires more than just the SNR value…

Event ID	Possible Source (Probability)	Significant	UTC	GCN	Location	FAR
S240705at	BBH (>99%)	Yes	July 5, 2024 05:32:15 UTC	GCN Circular Query Notices | VOE		1 per 4.4755e+07 years
S240703ad	BBH (>99%)	Yes	July 3, 2024 19:13:55 UTC	GCN Circular Query Notices | VOE		1 per 2.6751e+05 years
S240630t	BBH (>99%)	Yes	June 30, 2024 10:17:03 UTC	GCN Circular Query Notices | VOE		1 per 16736 years

Source: gracedb.ligo.org

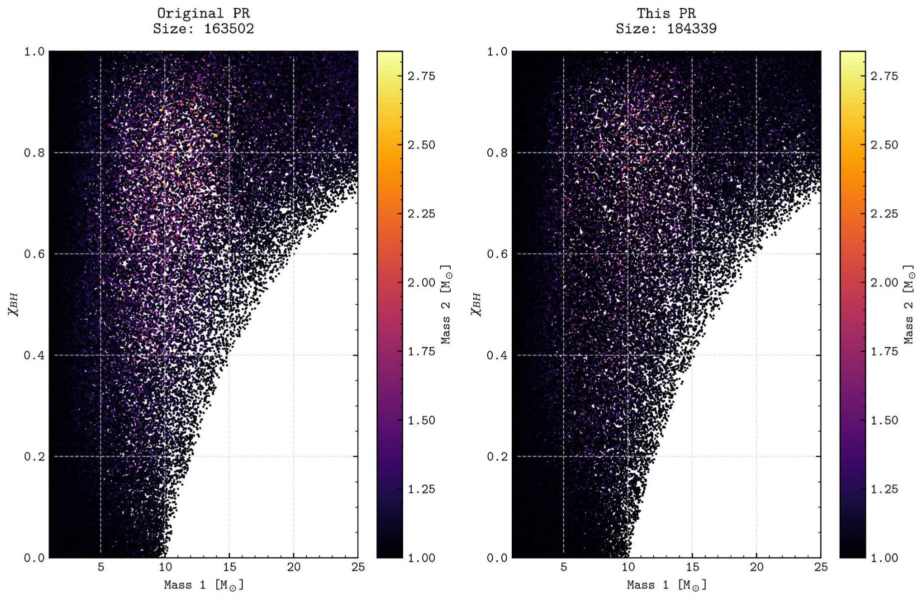
1. Define a background and a foreground dataset.
2. Study the statistical properties of the background by counting triggers produced during the matched filtering stage.
3. Create a histogram to understand the chances of having a false alarm, i.e. build the False alarm probability of the background.





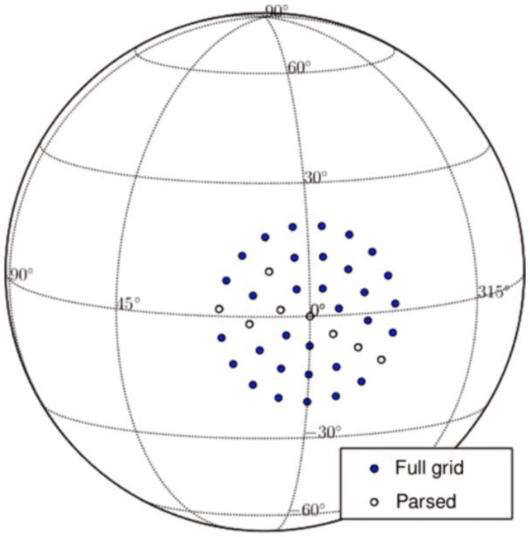
PyGRB: Matched filtering pipeline designed to search for electromagnetic transients coincident with GWs

EM bright bank



Source: github.com - [PR#4299](PR#4299) by Marco Cusinato

Tight skypatch



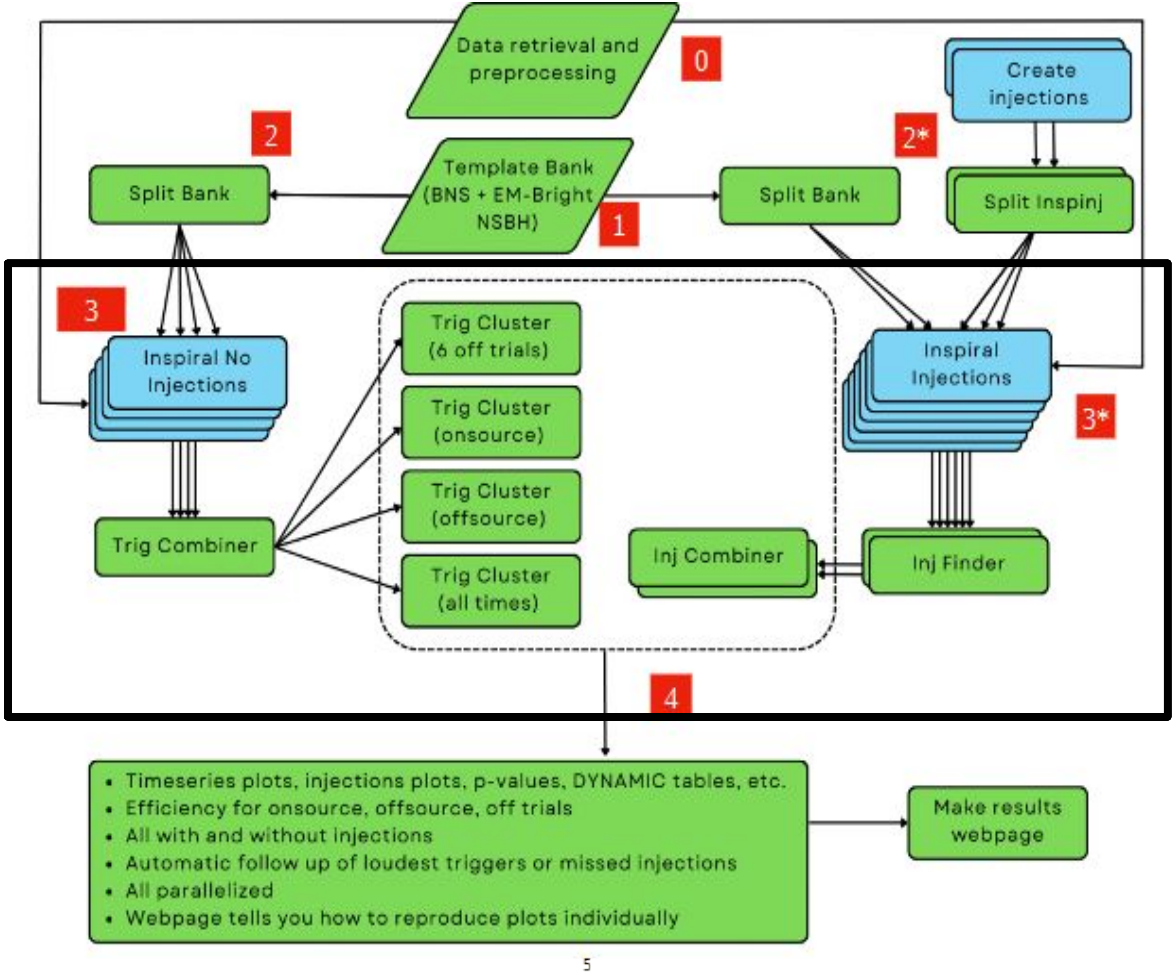
Source: Iain Dorrington's PhD thesis, 2019

- For short GRBs

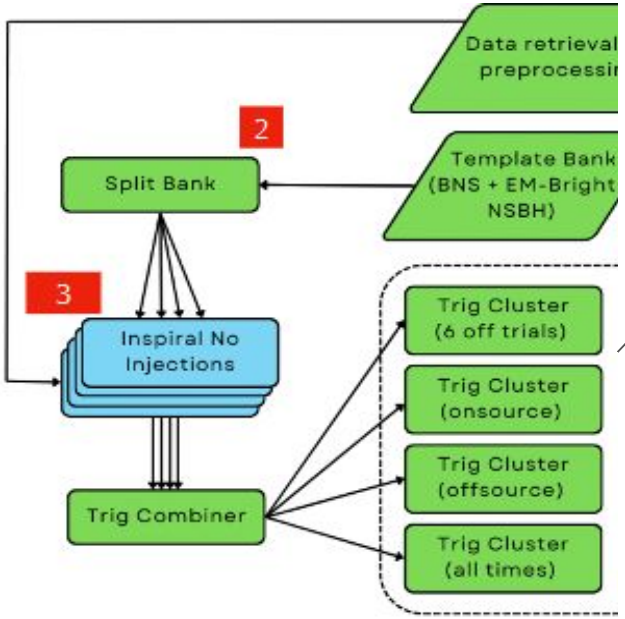
We get input information about the GRB trigger time and the skypatch where it was localized.

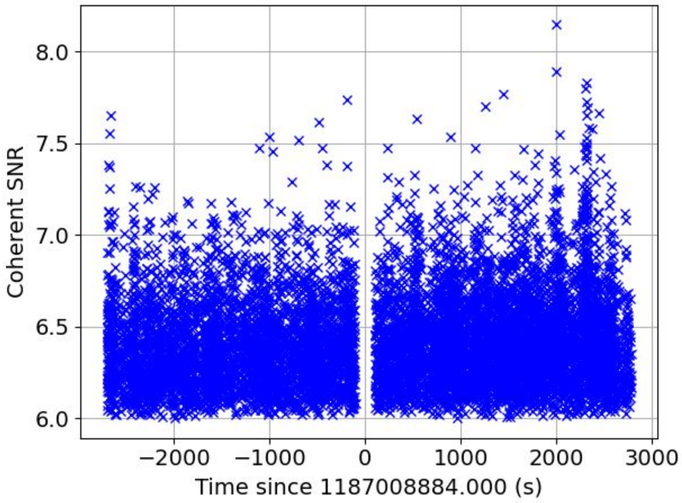
- Background: About 90 minutes(5000+ s) of data around the trigger are processed to estimate the background False alarm probability.
- Foreground: [-6s, +1s] around trigger time.

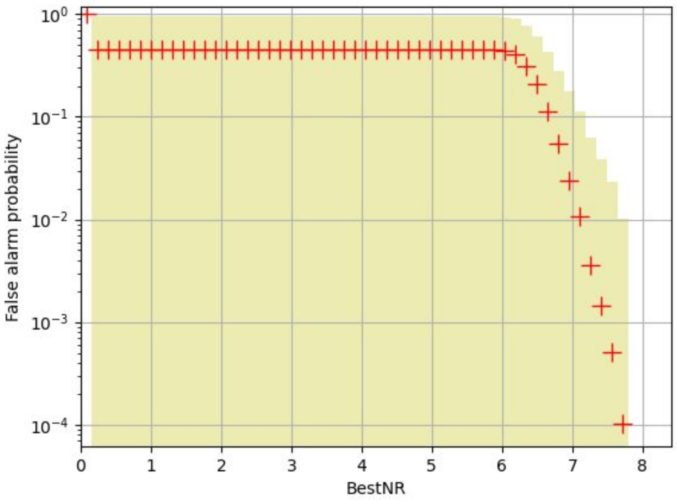
- FRBs, neutrinos, …


Data retrieval and preprocessing
0
Create injections
2
2*
Split Bank
Template Bank (BNS + EM-Bright NSBH)
1
Split Bank
Split Inspinj
3
Inspiral No Injections
Trig Cluster (6 off trials)
Trig Cluster (onsource)
Trig Cluster (offsource)
Trig Cluster (all times)
Inspiral Injections
3*
Trig Combiner
Inj Combiner
Inj Finder
4
• Timeseries plots, injections plots, p-values, DYNAMIC tables, etc.
• Efficiency for onsource, offsource, off trials
• All with and without injections
• Automatic follow up of loudest triggers or missed injections
• All parallelized
• Webpage tells you how to reproduce plots individually
Make results webpage
5

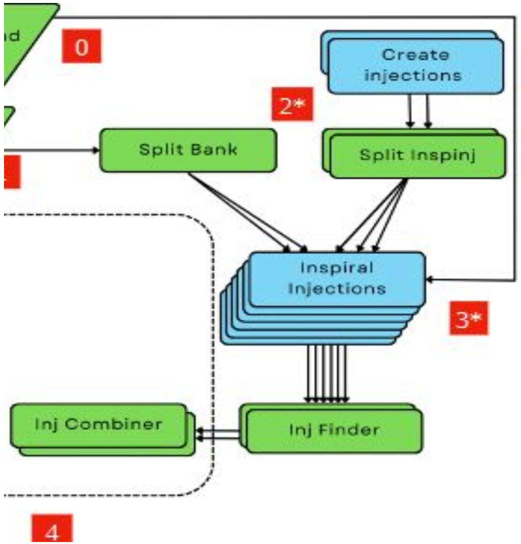
Build the background FAP distribution

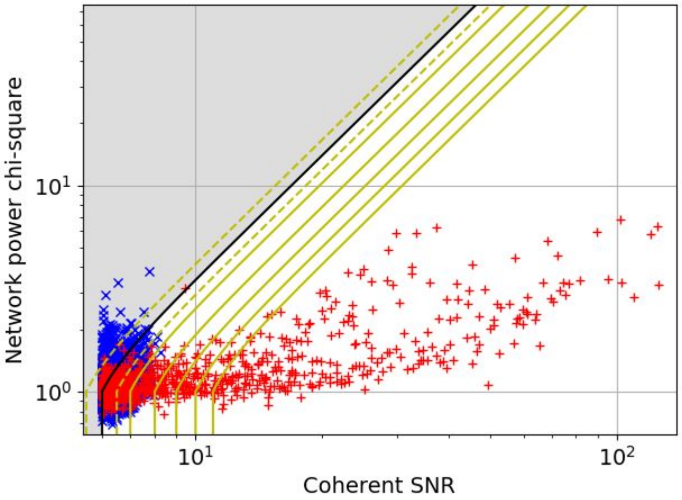


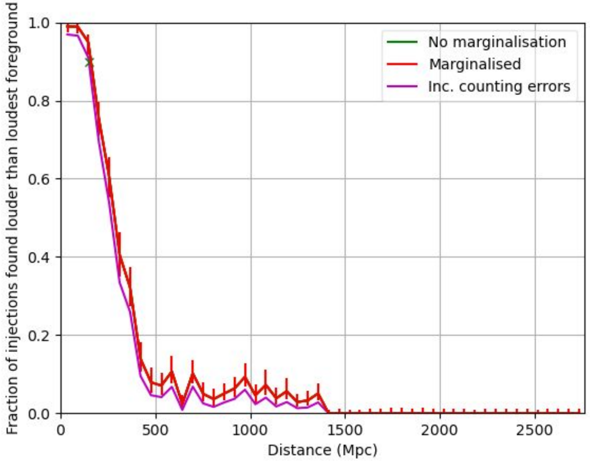


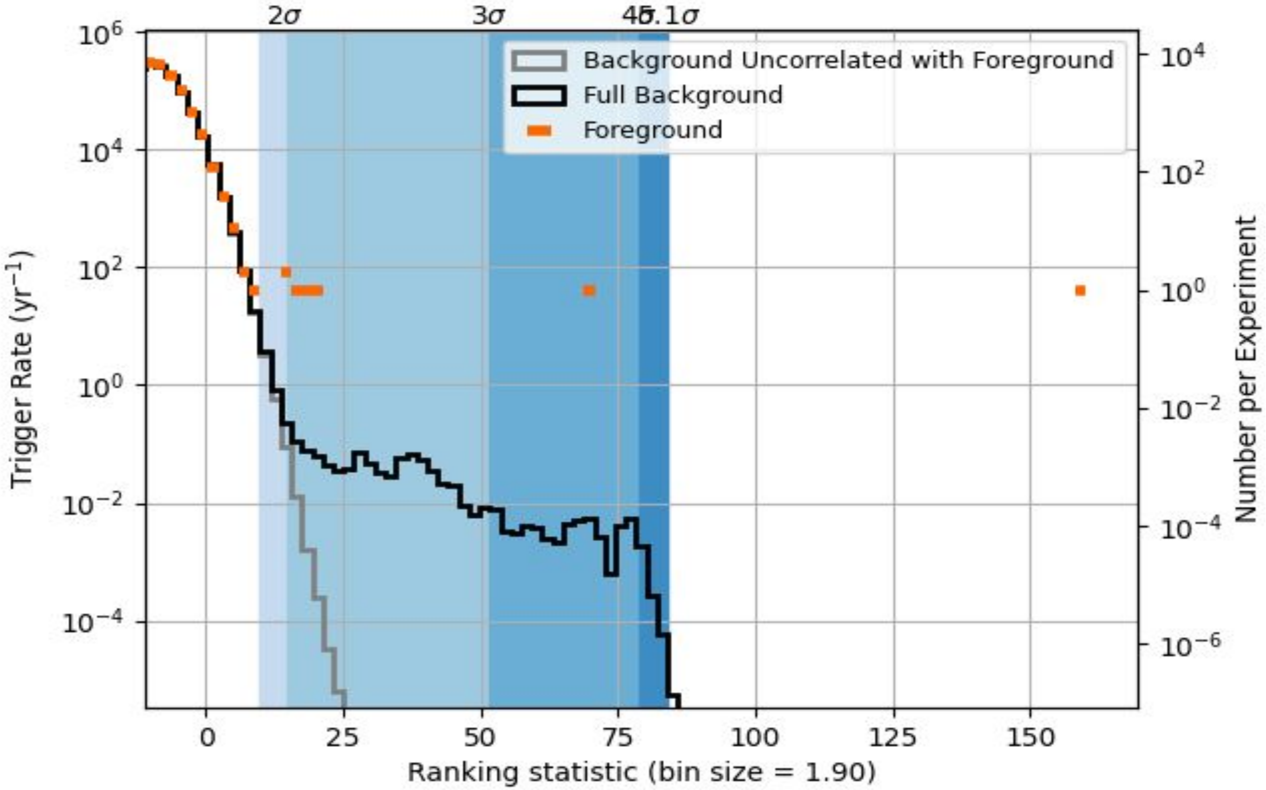


Inject thousands of signals to understand the sensitivity of the pipeline



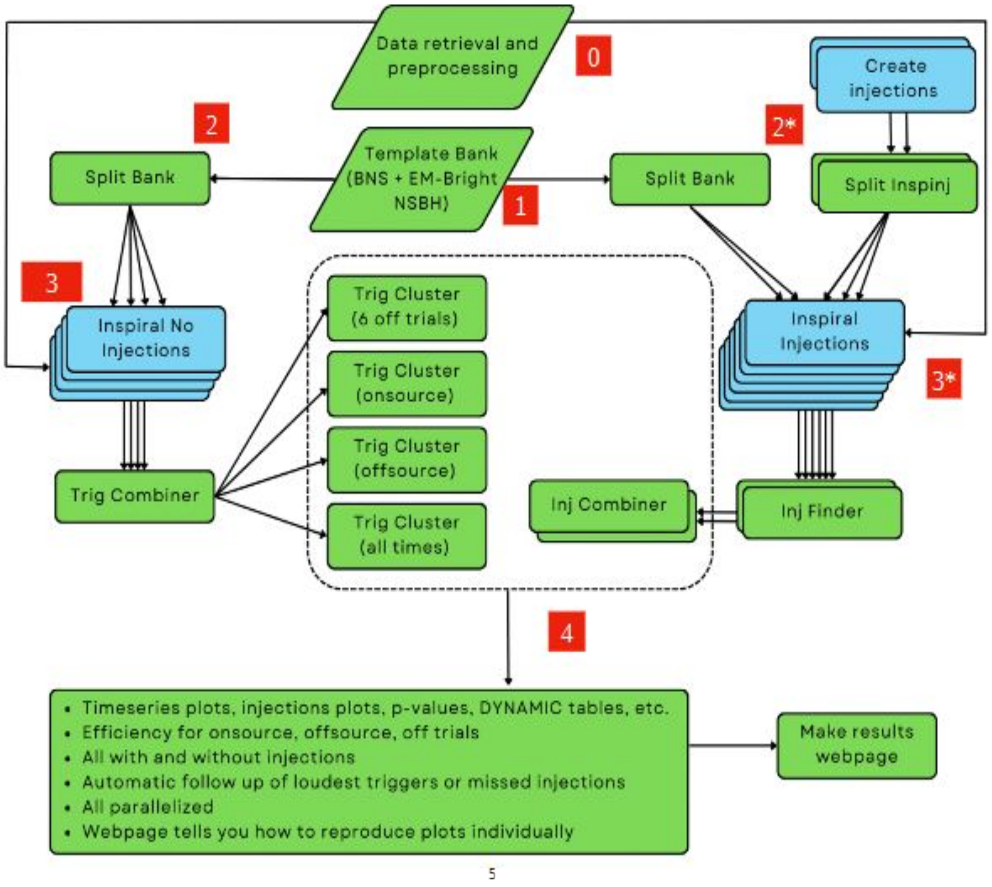






Credit: PyCBC allsky search team

Current development

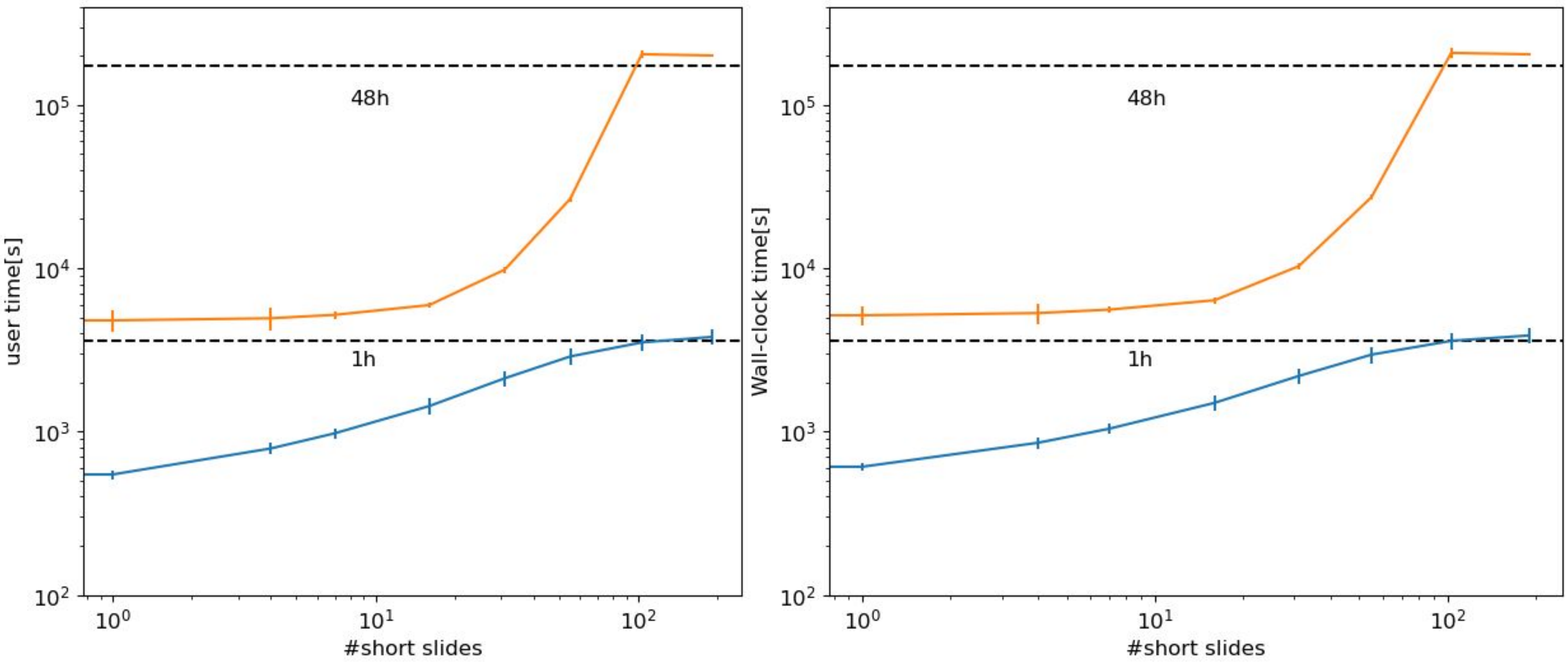


PyCBC

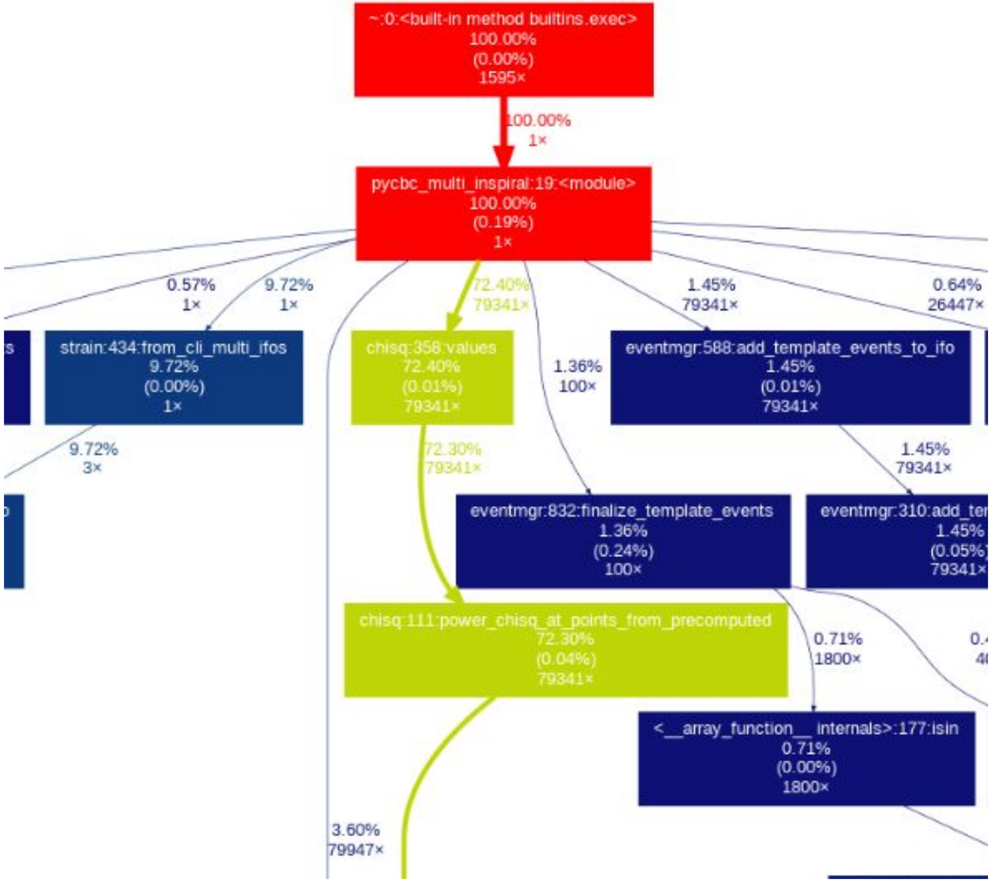
Performance gains…

lalapps_coh_PTF vs pycbc_multi_inspiral @CIT
Intel(R) Xeon(R) Gold 6154 CPU @ 3.00GHz

HLV network, T_bank_size=100
.gwf length-> 5648s
block duration=5632s, segment duration=256s



Bottleneck investigations…



Summary:

- Even though the signal to noise ratio(SNR) inner product is still at the core of our searches, our typical datasets require a larger infrastructure to assess statistical/astrophysical significance.

- PyGRB can and has been used for other transients.

- Current development is focused on the usage of modern open-source software and file formats and improve performance.

- PyGRB uses the same framework of the all-sky pycbc searches.

References:

- Harry, I. W. and Fairhurst, S., “Targeted coherent search for gravitational waves from compact binary coalescences”, Physical Review D, vol. 83, no. 8, APS, 2011. doi:10.1103/PhysRevD.83.084002.

- Iain Dorrington, Cardiff U., 2019. Improved methods for the detection of gravitational waves associated with gamma-ray bursts

- Sathyaprakash, B. S. and Schutz, B. F., “Physics, Astrophysics and Cosmology with Gravitational Waves”, Living Reviews in Relativity, vol. 12, no. 1, 2009. doi:10.12942/lrr-2009-2.