Gravity Exploration Institute

Sefydliad Archwilio Disgyrchiant

MLY PIPEPLINE

Vassilis Skliris, Patrick Sutton, Michael Norman, Kyle Willets, Wasim Javed

https://arxiv.org/abs/2009.14611 (Methods Paper)

Background and Motivation

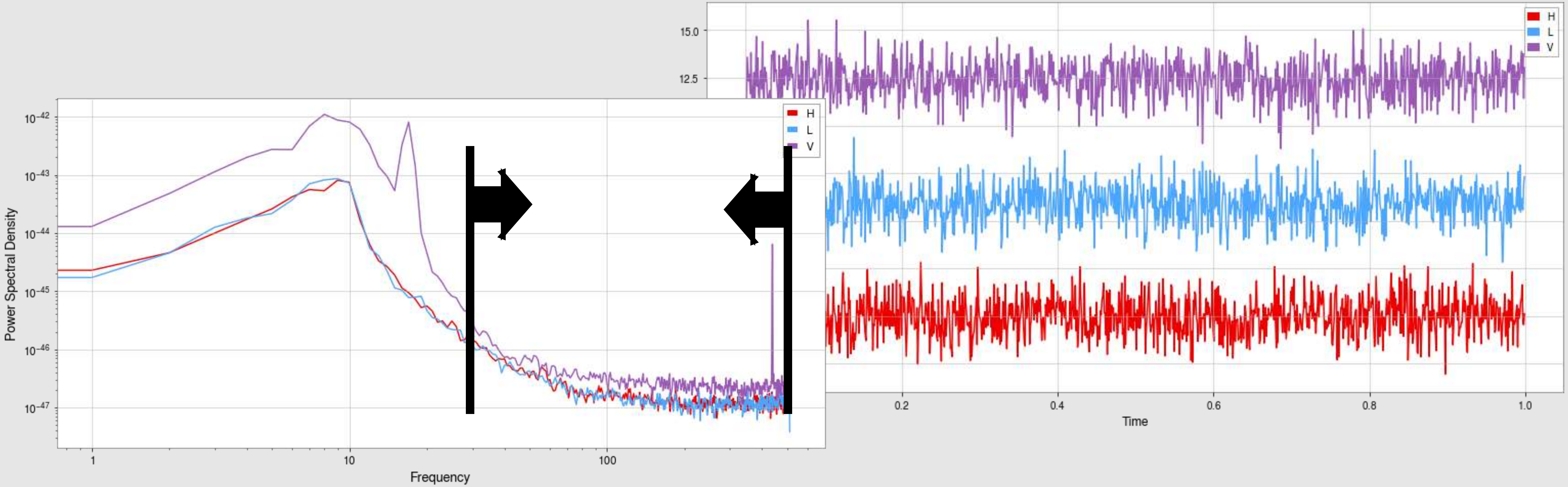
- It was the end of O2 everybody was talking about how the future will be overwhelming with detections rates of 1/day in the future.
- The first papers discussing applications of ML on signal detection (CBCs only) were been published.
- There was still justified skepticism about ML in LIGO due to some very enthusiastic claims by some of those first papers.
- Mostly because the efficiencies claimed were set on high, unusable false alarm rate.

Goals

- We needed an ML search for **transient signals**, not only CBCs. – MLy Pipeline
- It needs to be trustworthy. – We need to reach low FAR on detector noise before we do any claims.
- Compare our results on the same basis as analytical methods. – Comparison with CWb.
- It needs to be easily reproducible – Creating a framework of analyzing the data.

Building blocks

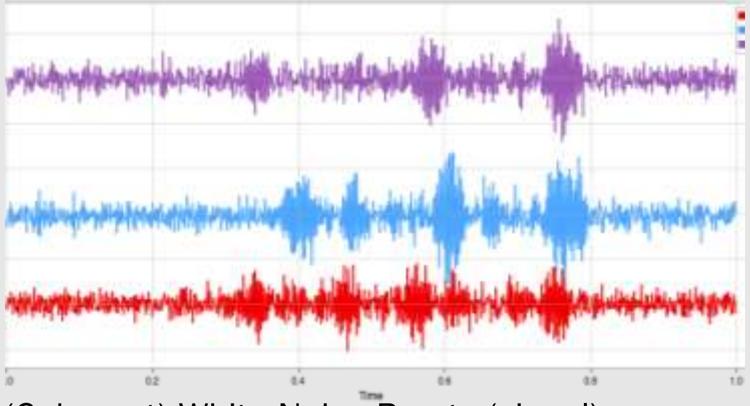
1. Simulated gaussian Noise 20-1024Hz



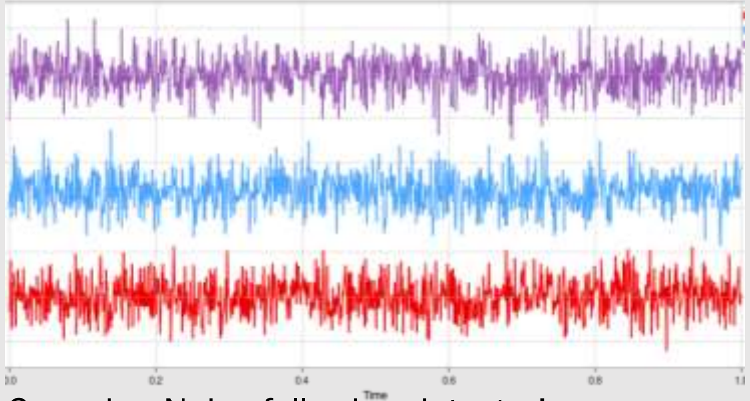
2. Whine Noise Bursts (WNBs) F_{min} , F_{max} , T

Model 1

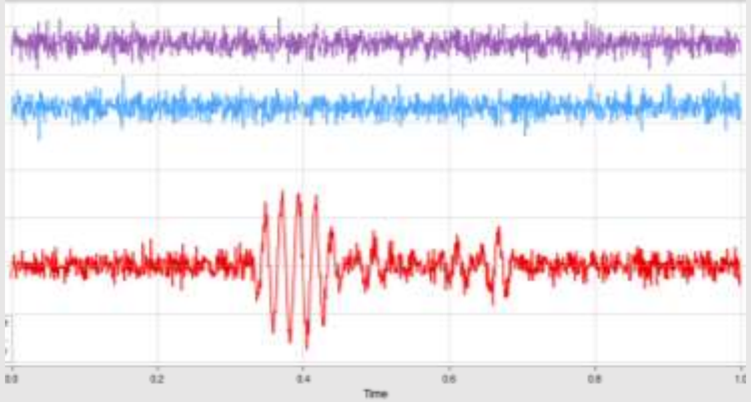
Is there a signal to **at least** two detectors ?



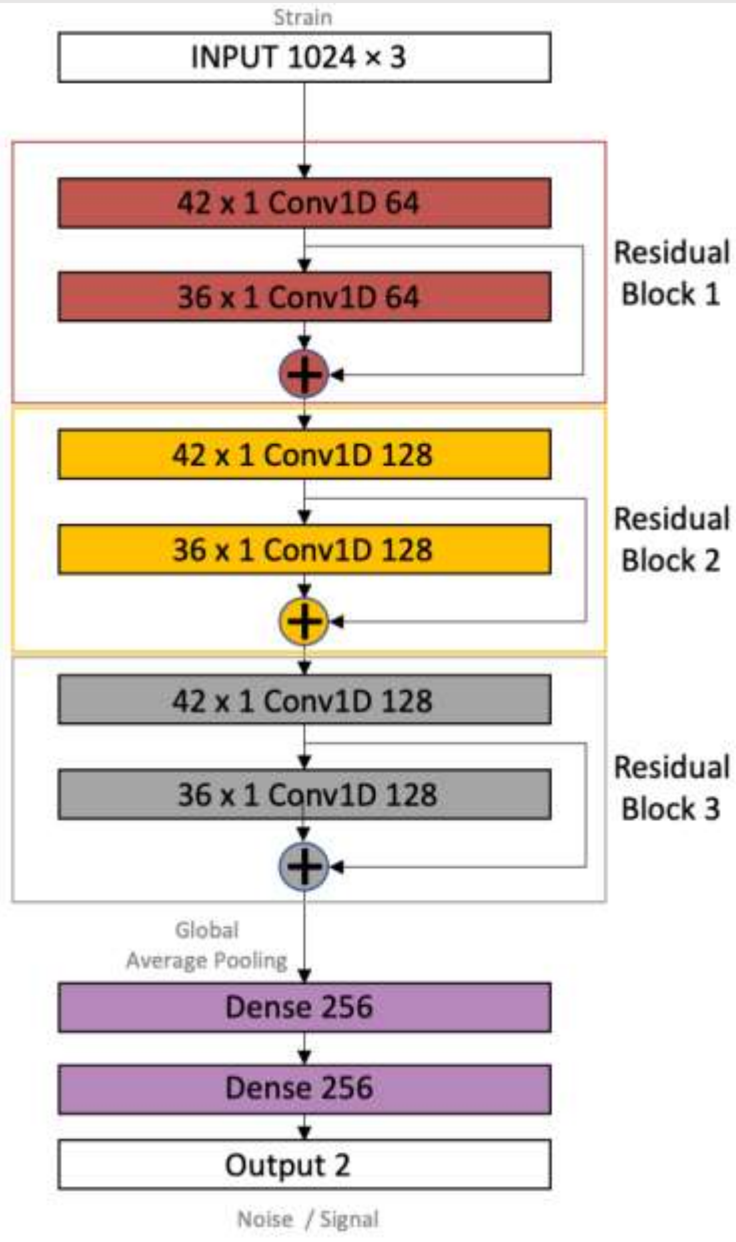
(Coherent) White Noise Bursts (signal)



Gaussian Noise following detector's response

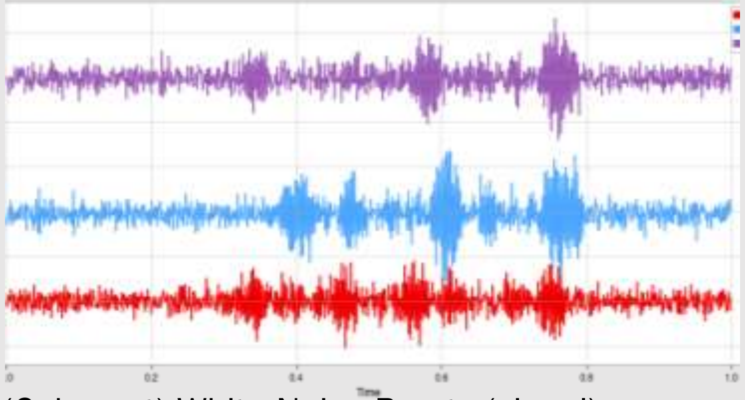


WNB in one detector (glitch behavior)

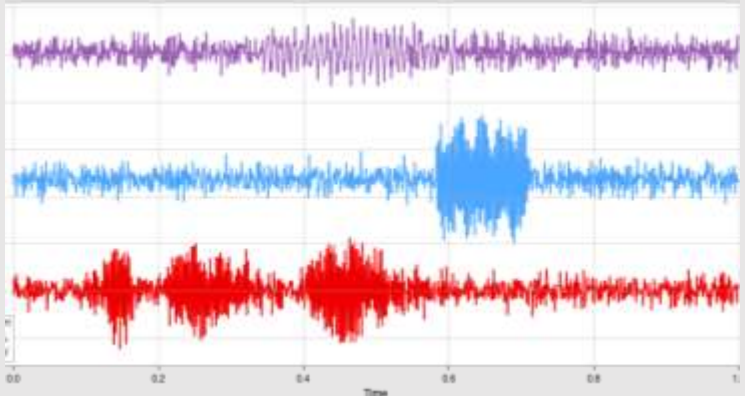


Model 2

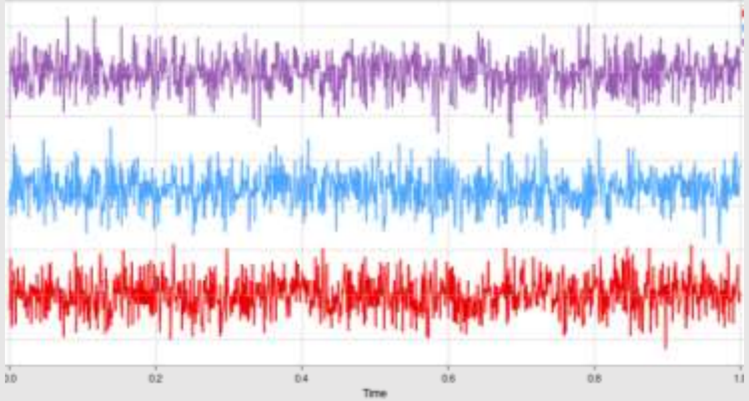
Is there any **coherency** between any detectors ?



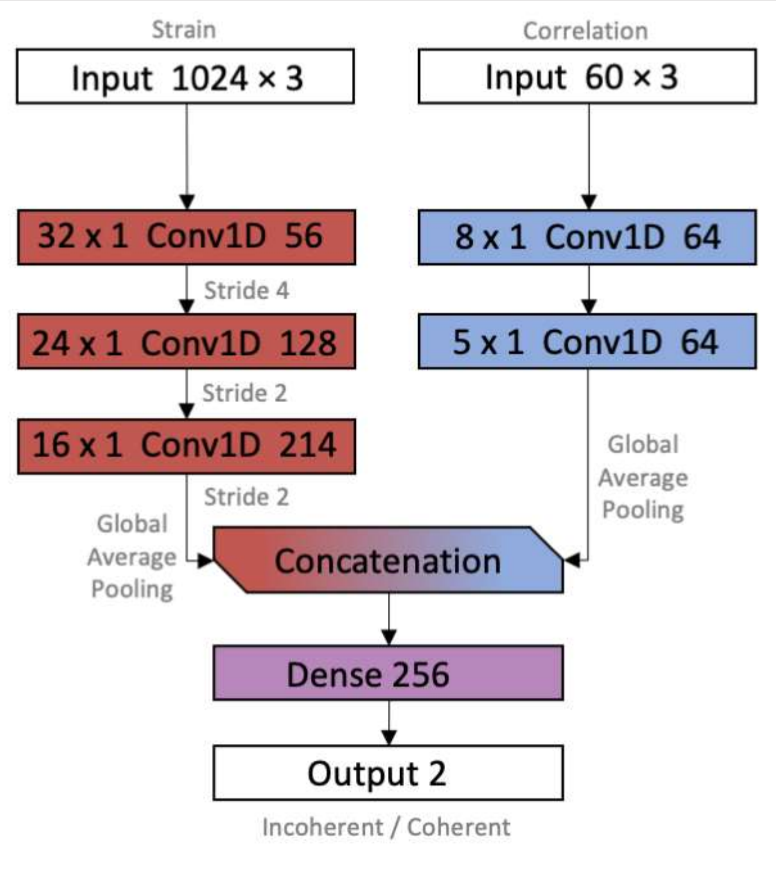
(Coherent) White Noise Bursts (signal)



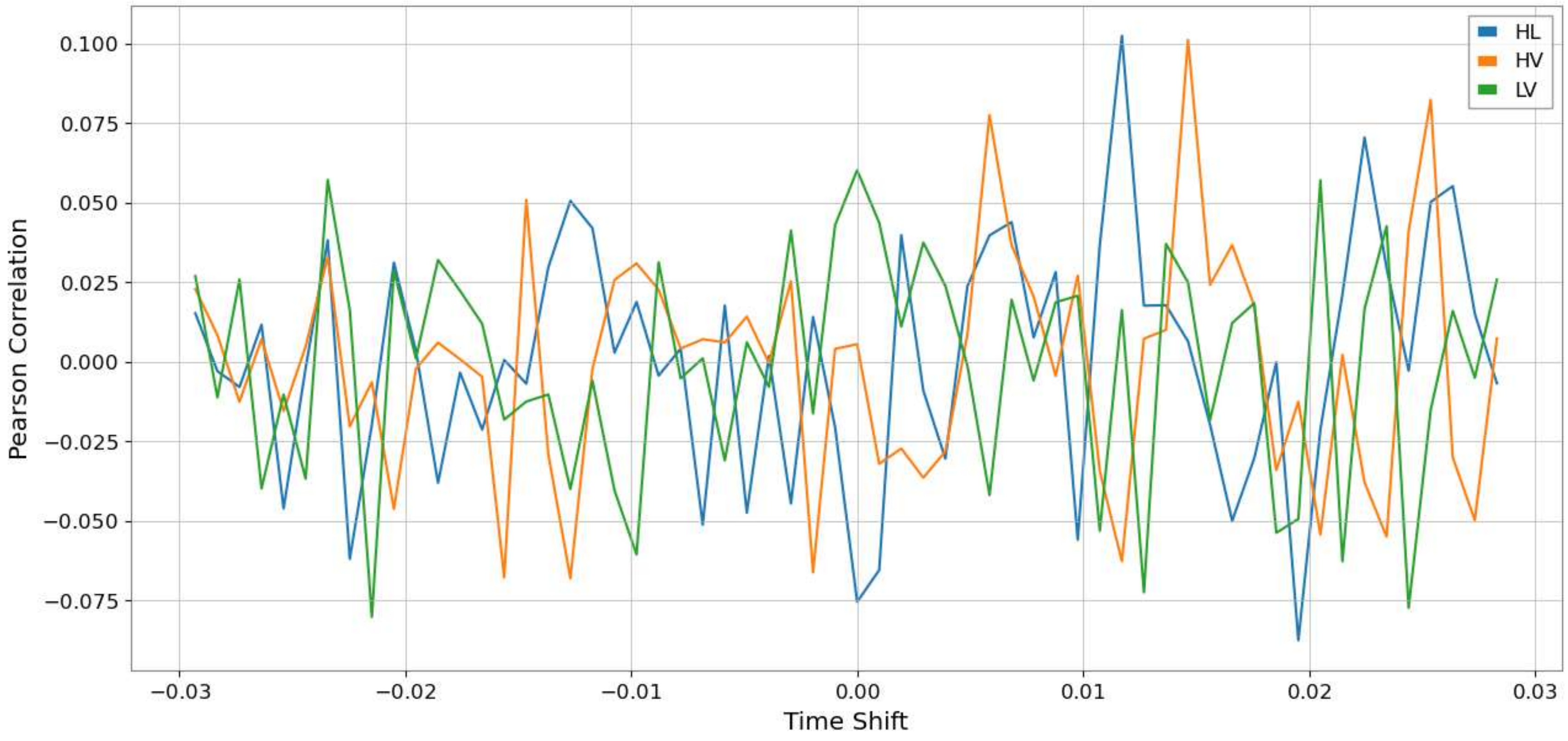
Incoherent White Noise Bursts



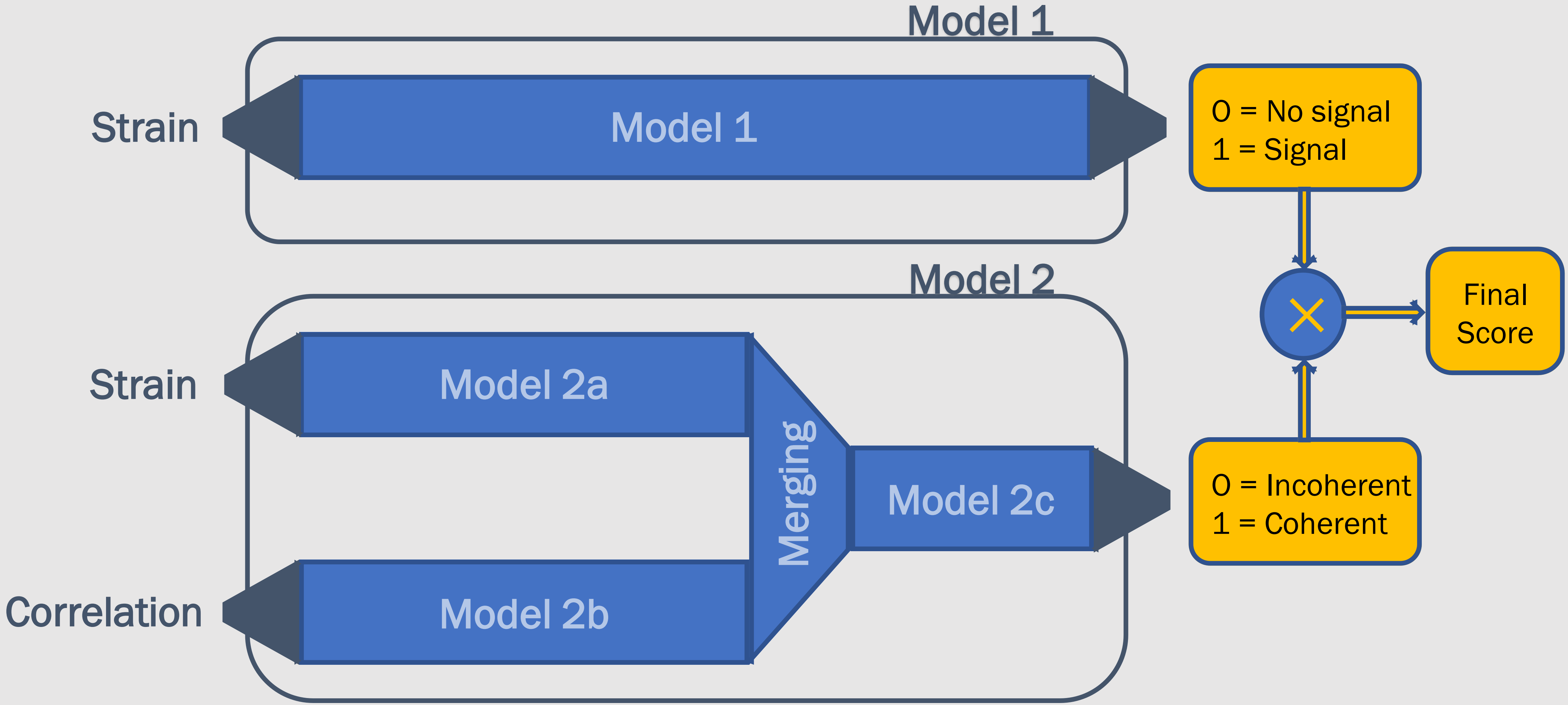
Gaussian Noise following detector's response



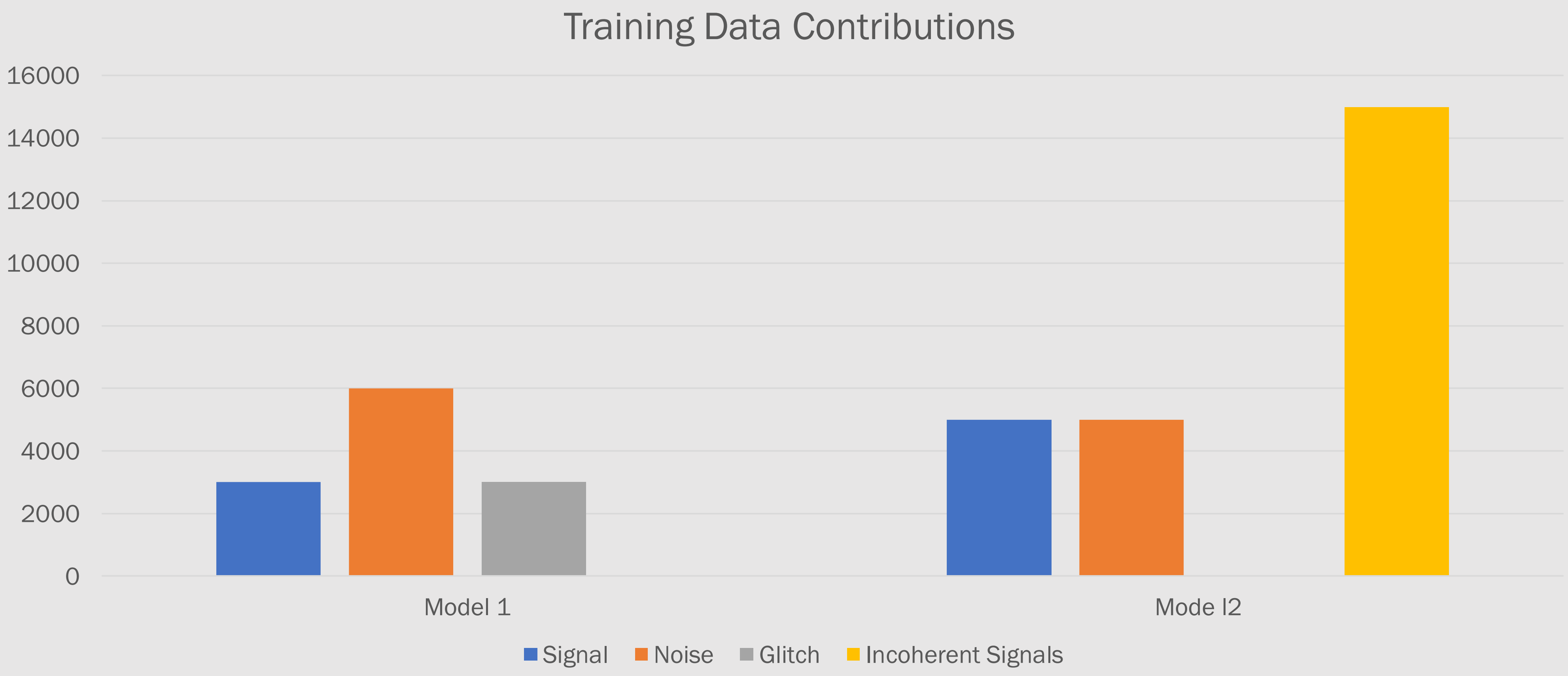
Model 2
Correlation data



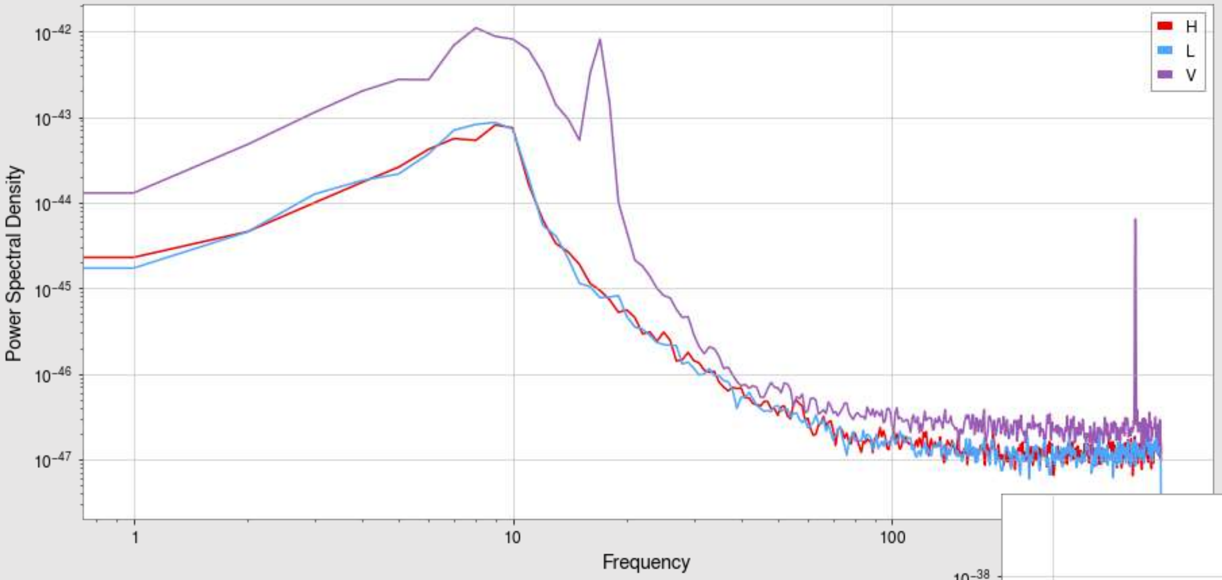
Detection Algorithm



The power of data type ratios

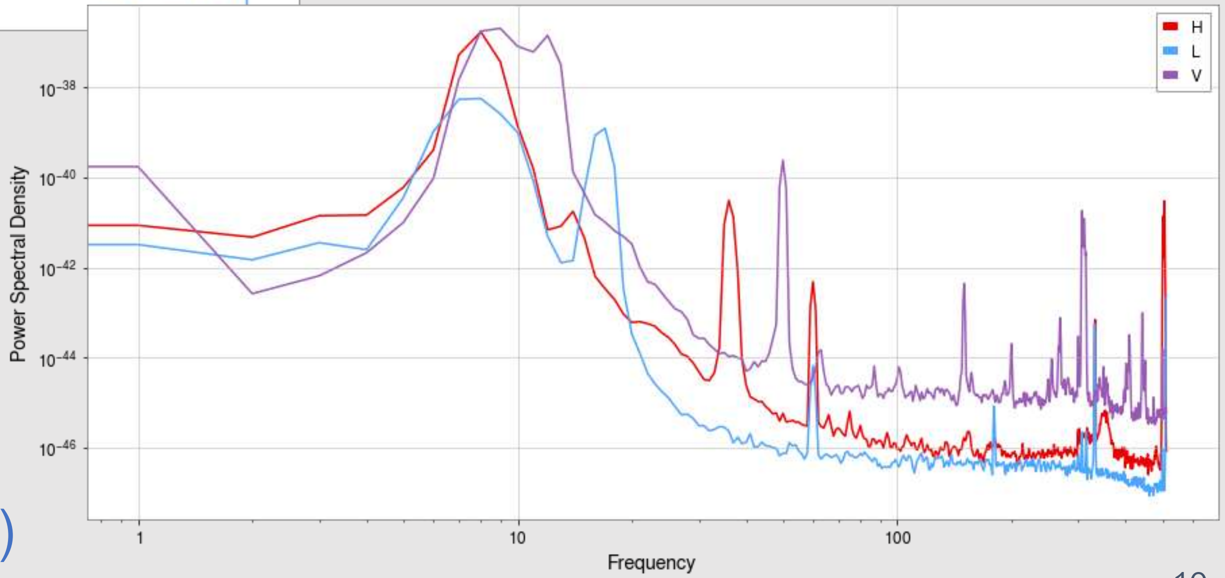


Detector contribution balance



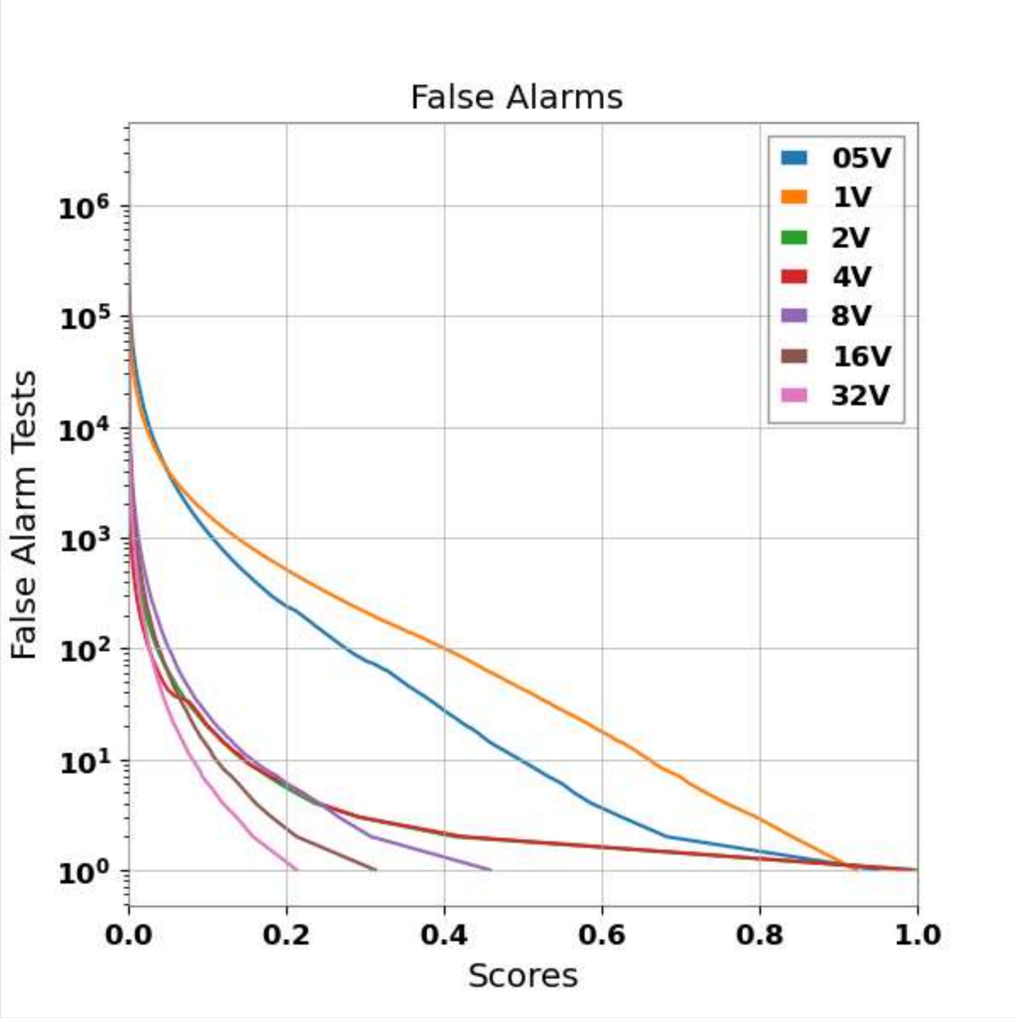
Training Data

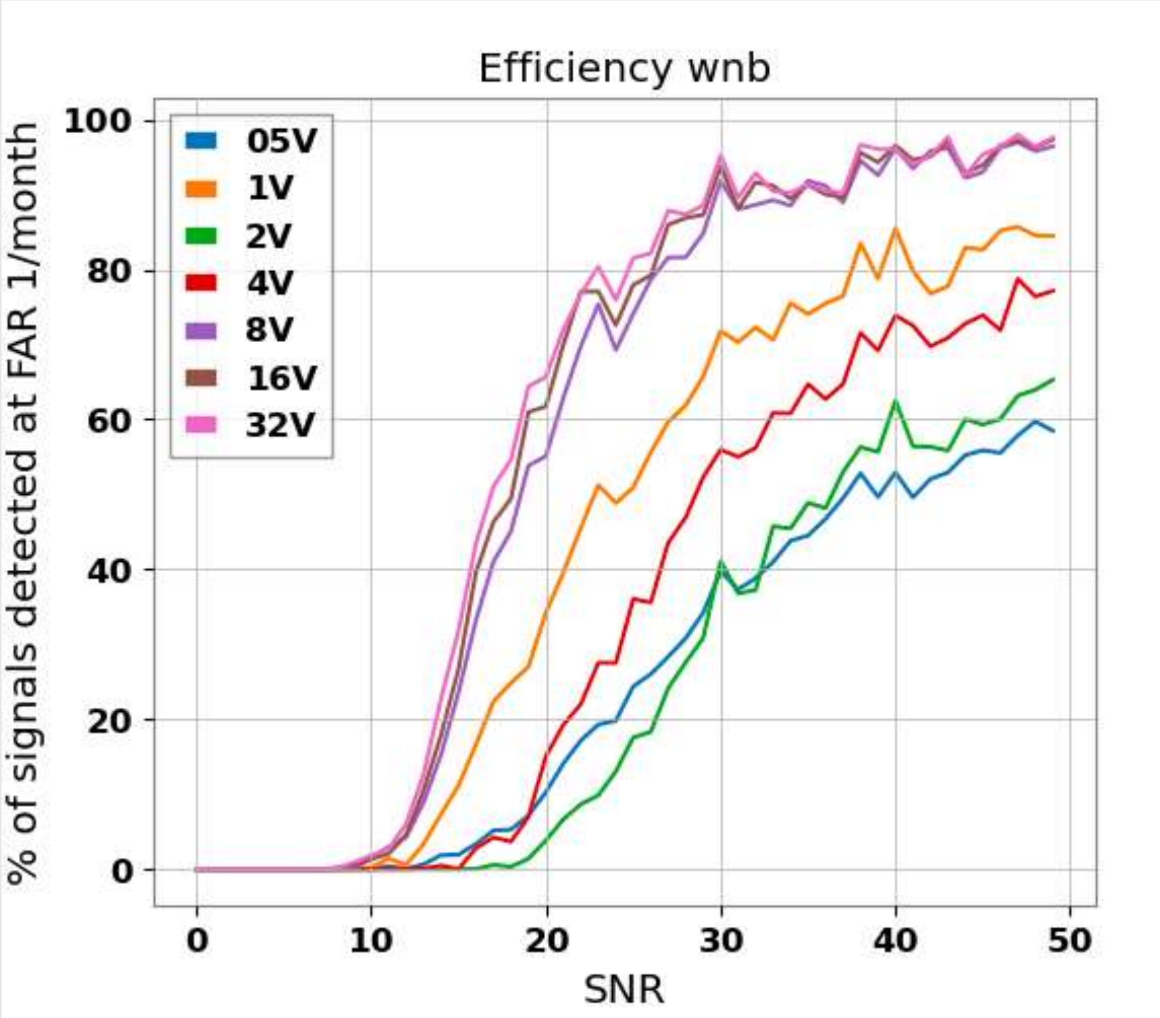
The noise level of Virgo in artificial data is much closer to H1 and L1 than in the real data.



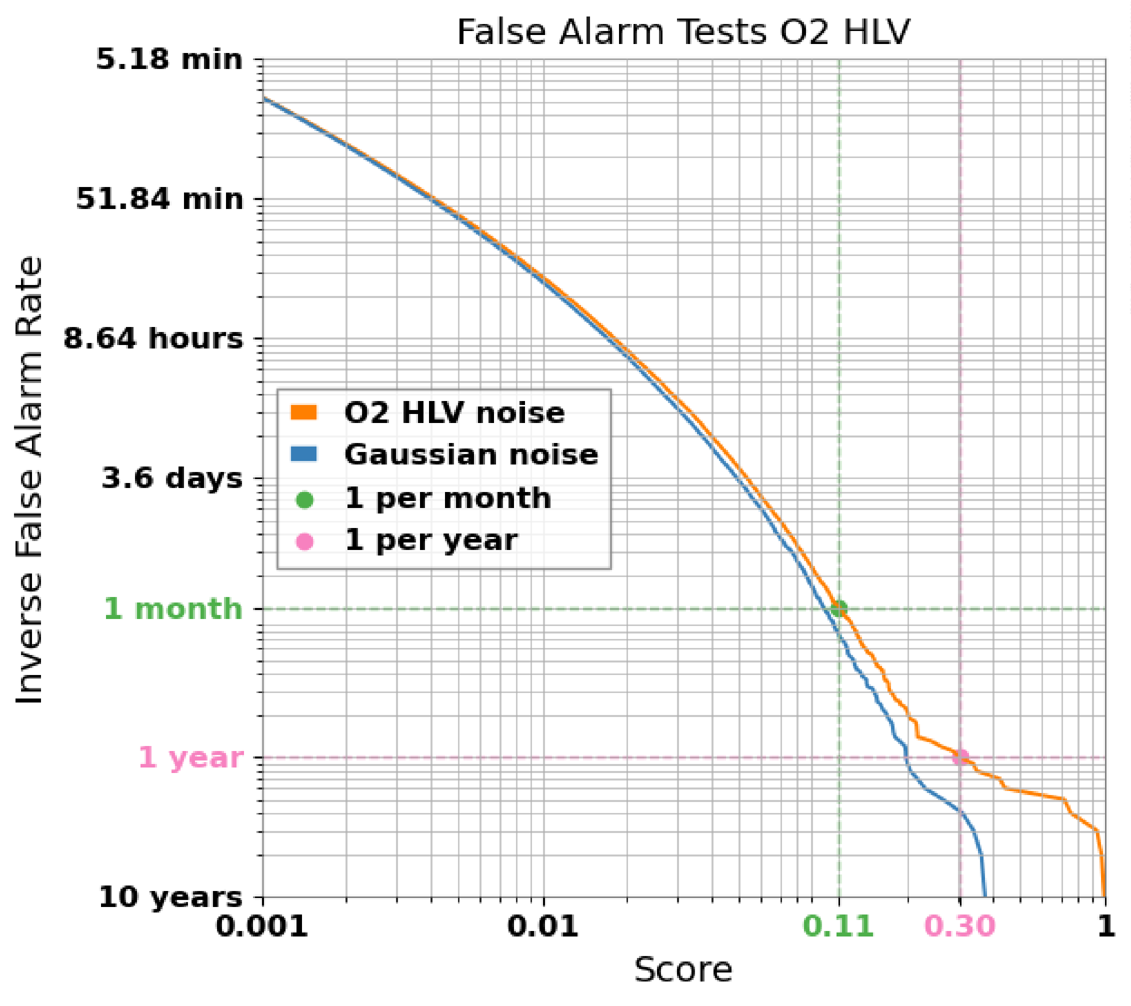
Testing Data (O2)

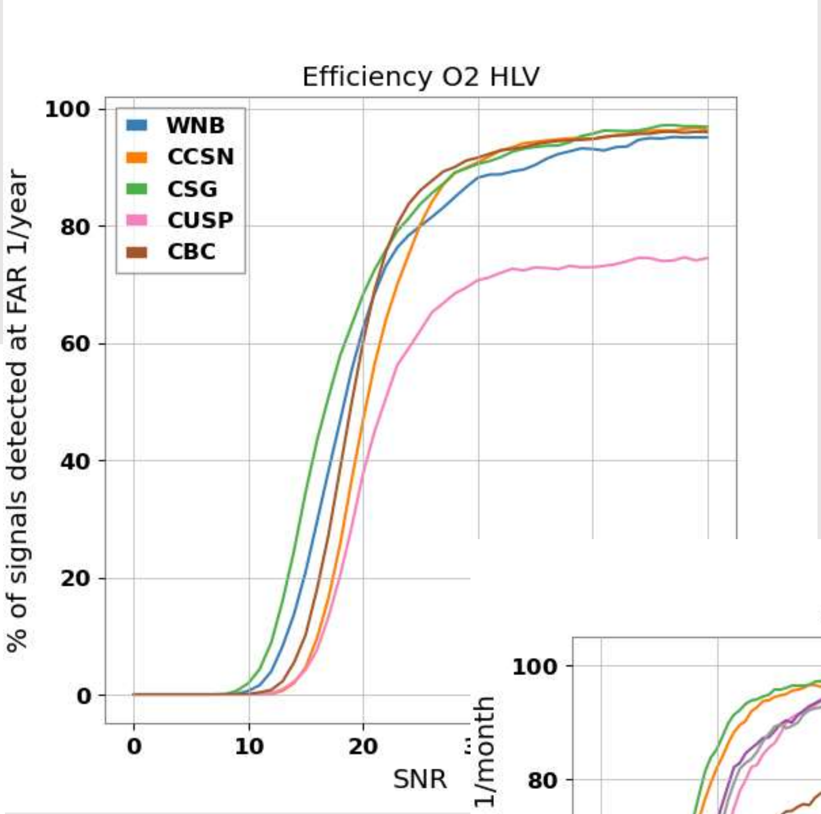
Detector contribution balance

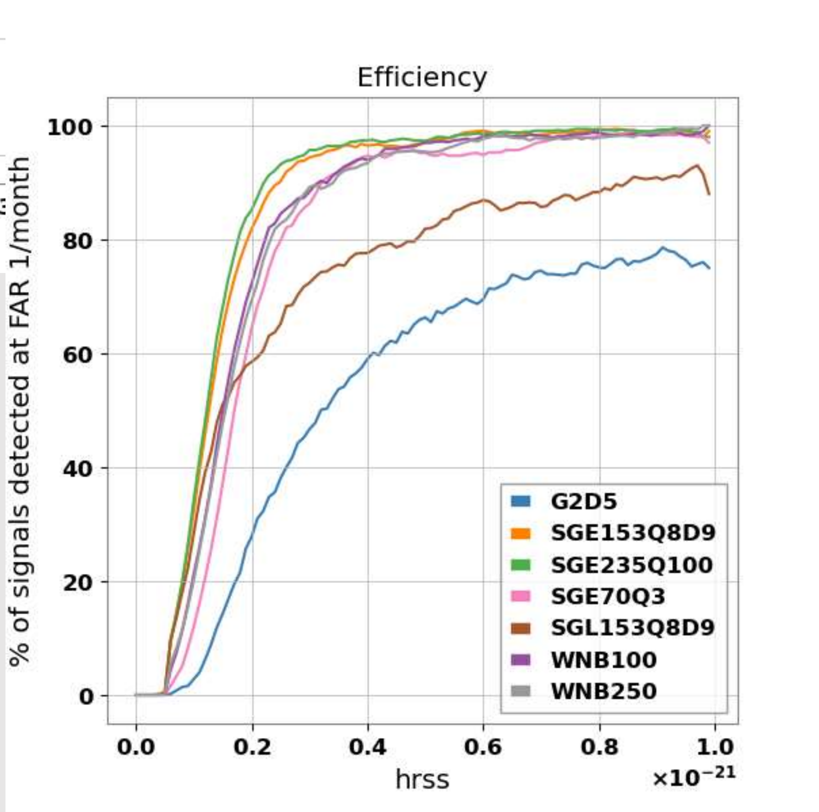




O2 performance

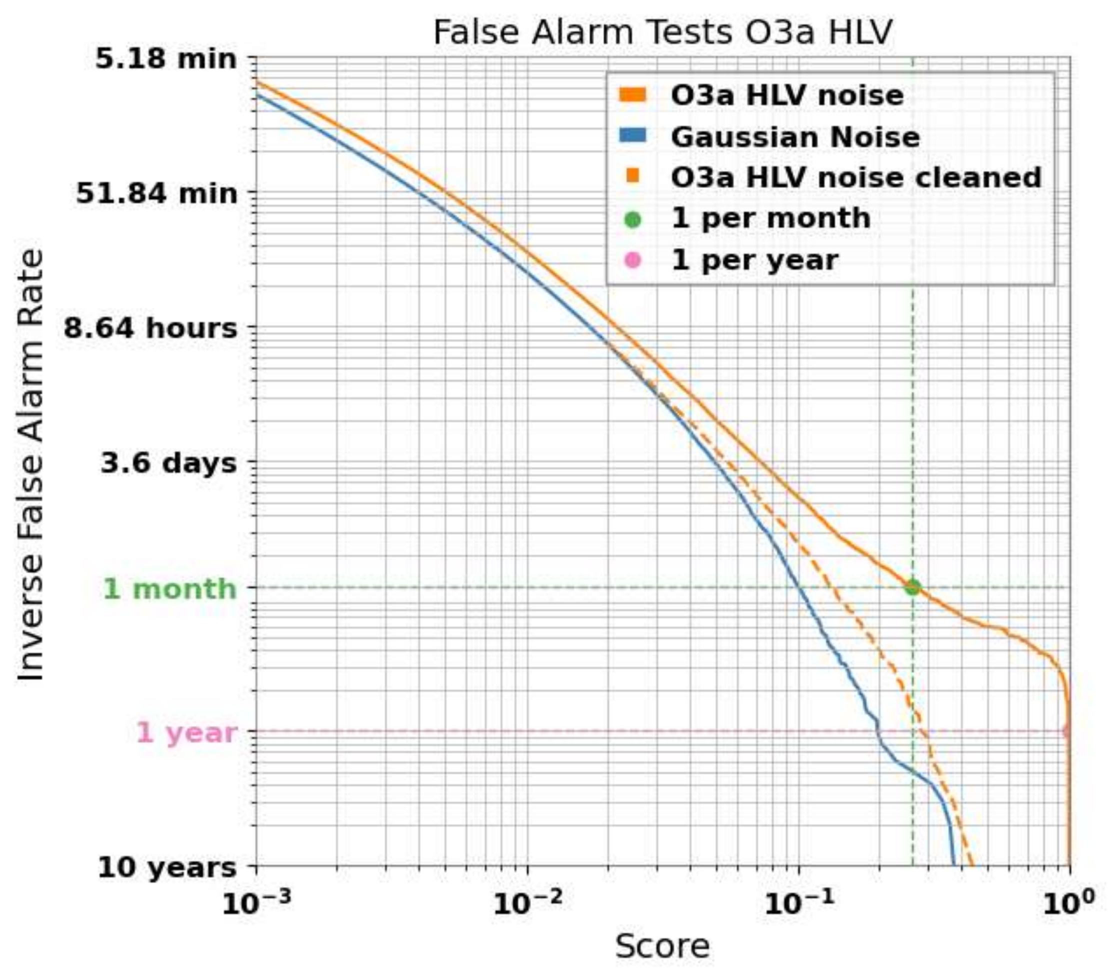


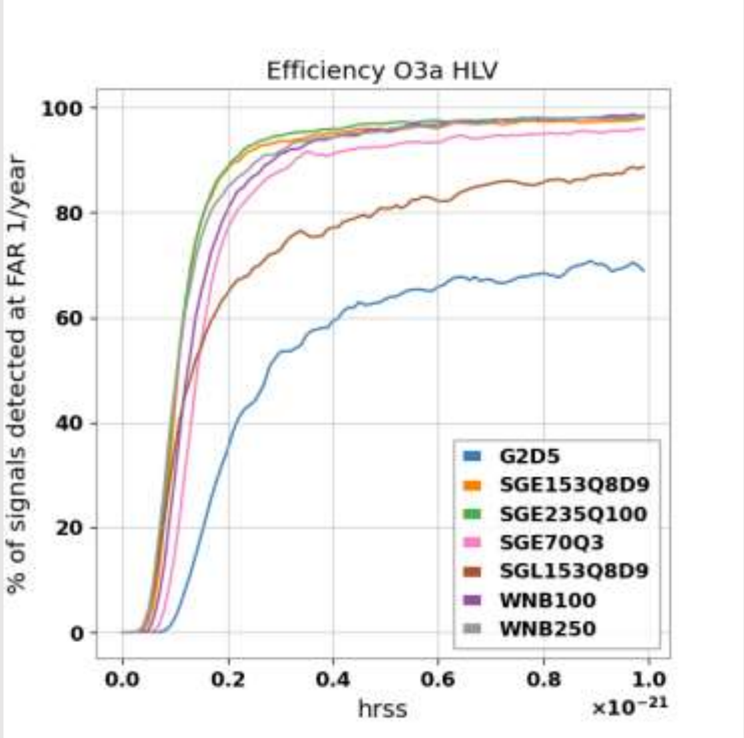




O3a performance

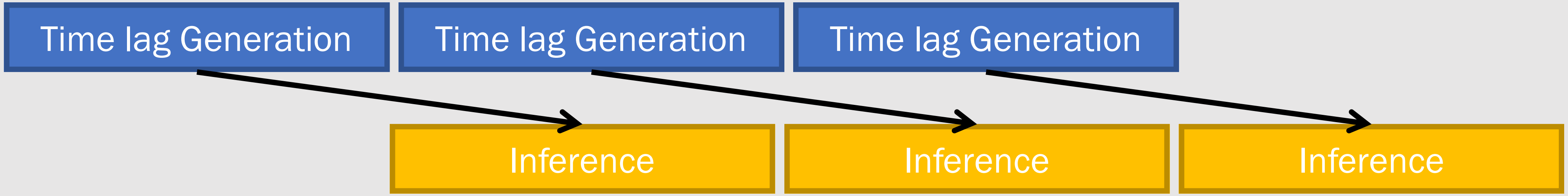
Performance comparison on the O3a HLV data set		
Morphology	cWB [4] ($10^{-22}Hz^{-1/2}$)	MLy ($10^{-22}Hz^{-1/2}$)
Gaussian pulses		
t=2.5 ms	1.6	2.9
Sine-Gaussian wavelets		
f_0=70 Hz, Q=3	0.9	1.5
f_0=153 Hz, Q=8.9	0.6	1.2
f_0=235 Hz, Q=100	0.6	1.1
White-Noise Bursts		
f_{low}=100 Hz, Δf=100 Hz, t=0.1 s	0.8	1.3
f_{low}=250 Hz, Δf=100 Hz, t=0.1 s	0.8	1.1





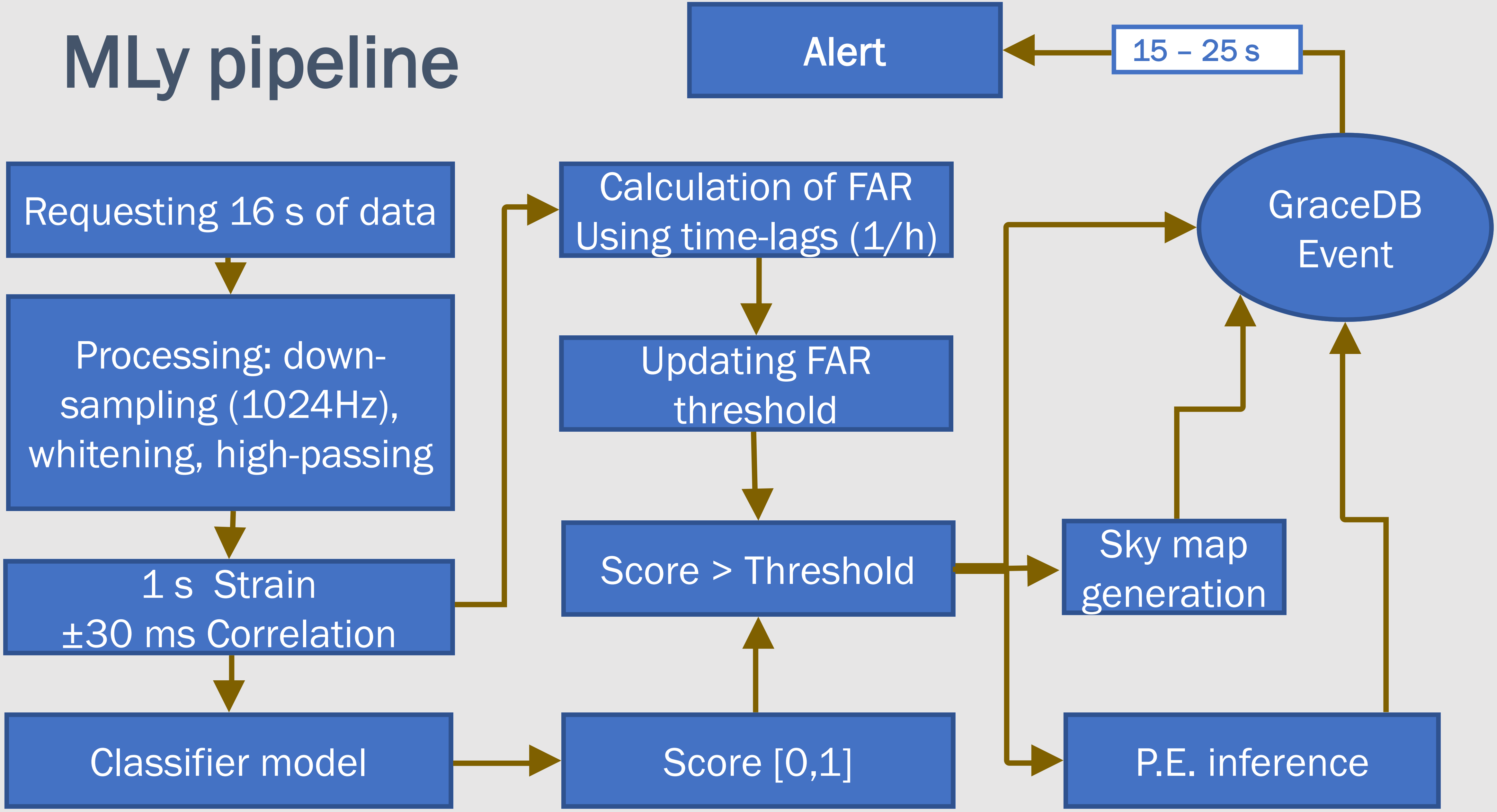
Low latency application

- The background data are becoming available every second.
- Noise behavior changes, so we need to be able to able to update our FAR and eventually our thresholds in real time. We need continuous FAR estimation.
- From every hour of data, we can roughly get 40 days of background simulation. **We use Hermes increase our throughput up to 1000 inferences/s.**



- **And finally, the most important, latency of order of seconds.**

MLy pipeline



Future plans

- Off-line search. Demonstrating low computational costs.
- Overlapping segments during inference. Evolve our significance metric.
- Continuous training : Fine-tuning of the models as new data become available. Utilization of real noise.
- Expansion of search parameter space (higher frequencies, longer duration)
- Low latency glitch rejection. Reduction of FAR.

Thank You