Hearing the Signal through the Static: Real-time Noise Reduction in the Hunt for Binary Black Holes and other Gravitational Wave Transients

Sydney Chamberlin[1], Reed Essick[2], *Patrick Godwin*[1], **Chad Hanna[1]**, Erik Katsavounidis[3], *Duncan Meacher*[1], Madeline Wade[4]

[1]*The Pennsylvania State University, University Park, PA, 16801*
[2]*University of Chicago, Chicago, IL 60637*
[3]*Massachusetts Institute of Technology, Cambridge, MA 02139*
[4]*Kenyon College, Gambier, OH 43022*

At a glance:

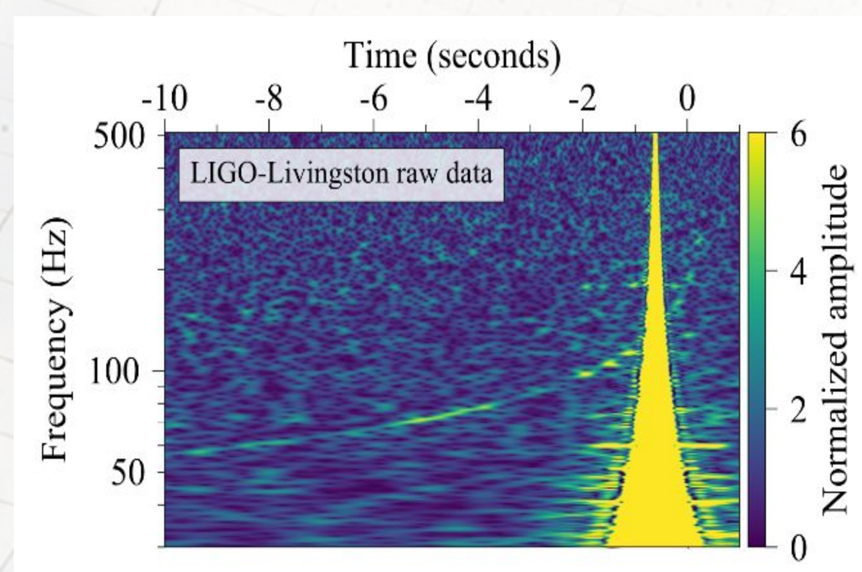
- 2000 sensor channels
- 2 TB / day
- 300 CPU cores
- 3 Data Centers
- 7s latency
- >99% uptime
- 2000 commits
- Conda installable

We now have cyberinfrastructure that is a start for a framework to do production machine learning on RAW LIGO data in real-time

Goal: Identify non-stationary noise in LIGO gravitational wave (GW) detectors in real-time.

LIGO detects GWs from black hole and binary neutron star collisions providing real-time alerts. Non-stationary noise limits our ability to detect GWs in real-time.

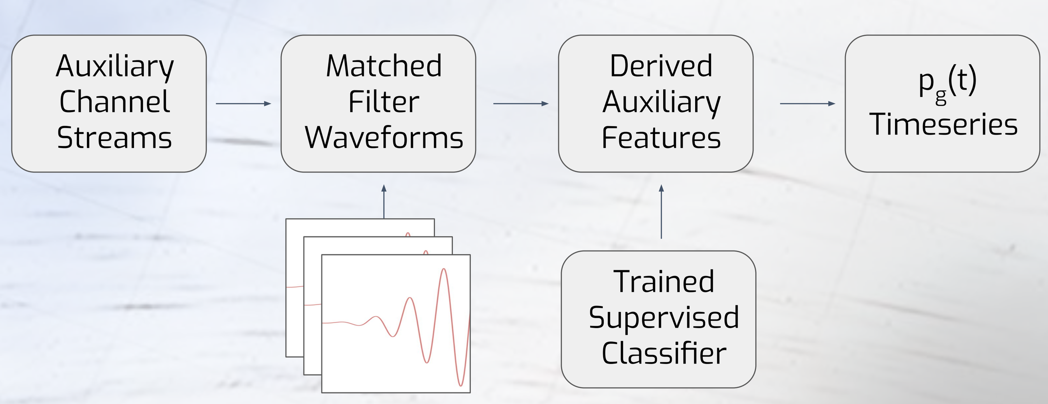
LIGO also has many auxiliary data streams recording environmental and instrumental noise.



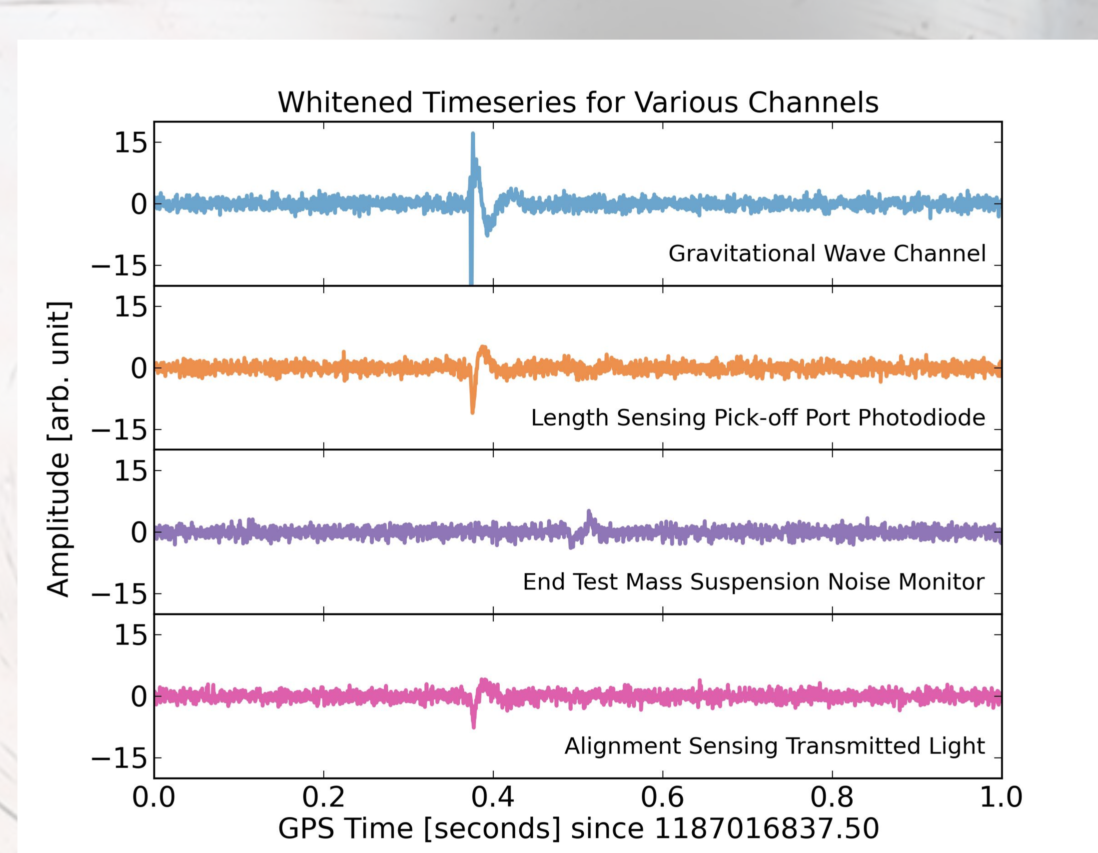
Phys Rev Lett. 119, 161101 (2017)

We aim to infer the presence of nonstationary noise using auxiliary channel information in **real-time** to improve the reliability of automated GW alerts and remove the need for human vetting.

Method: Use stream-based feature extraction and machine learning.



We use a combination of signal processing and machine learning techniques on auxiliary channels to infer the presence of nonstationary noise in gravitational wave data. This is done in two steps:

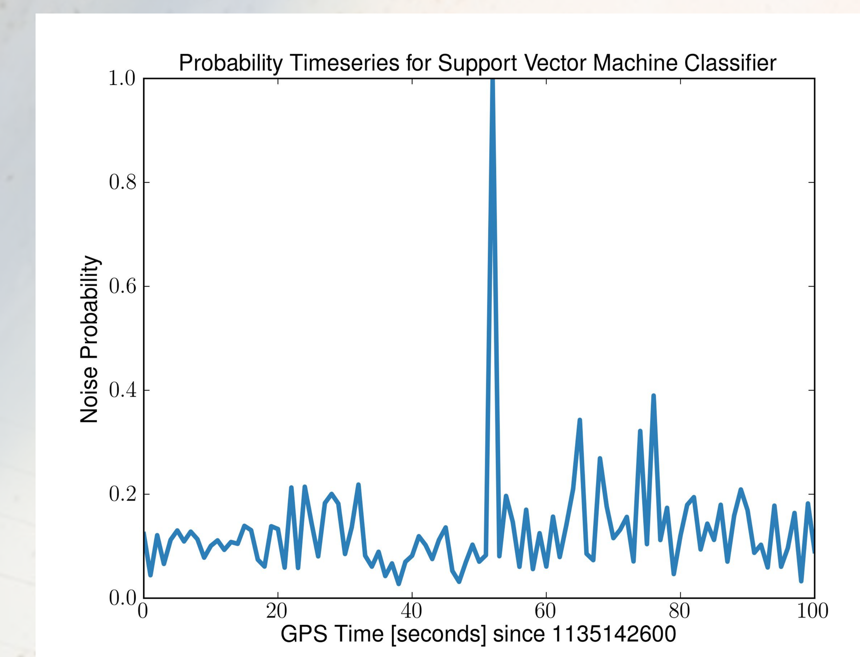


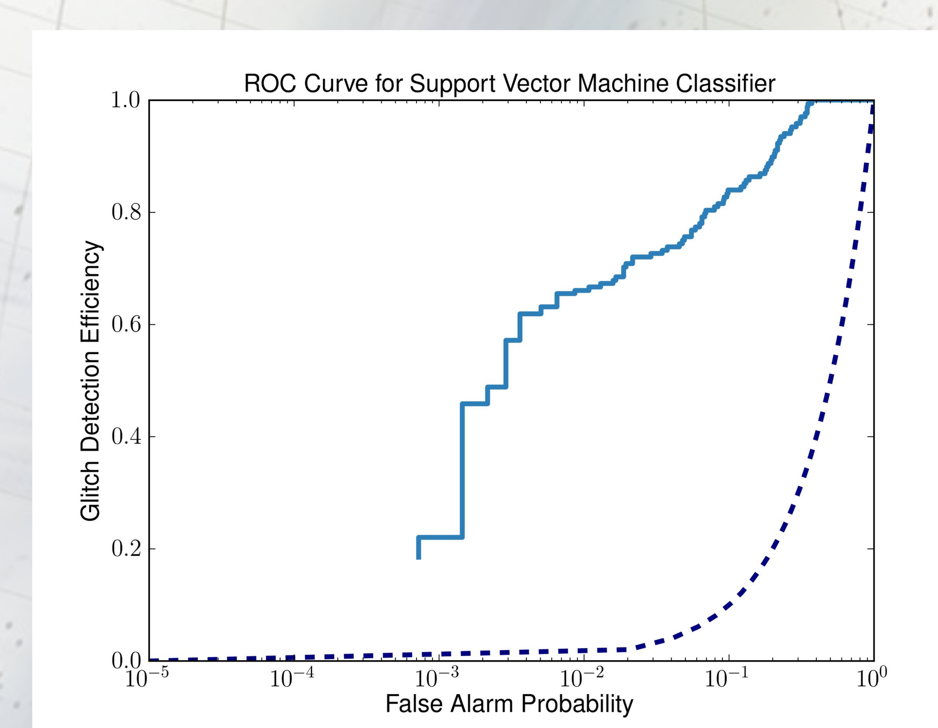
Nonstationary noise is present in auxiliary channels. The coupling to gravitational wave data can be inferred by supervised classifiers.

1) Apply matched filtering on streaming auxiliary channel timeseries using noise waveforms to extract meaningful quantities that describe important features of the noise.
2) Feed these derived features into supervised learning classifiers that infer the presence of noise in the gravitational wave channel. The end product is a probability of the presence of nonstationary noise in gravitational wave data, given auxiliary channel data, produced as a streaming timeseries and available in real-time.

Results:

Supervised classifiers were trained on 500000 seconds of data taken during O1 containing about 5500 training samples. Depicted here are results from a Support Vector Machine classifier using a radial basis function kernel. Evaluation was done on 70000 seconds of data, shown on the right for a 100 second timespan.





Performance metrics described by a Receiver Operator Characteristic (ROC) Curve is shown on the left, and characterizes the performance of a classifier in distinguishing noise from gravitational waves.

Current Status:

- Real-time classification has been running since April 2019.
- The classification time-series is distributed with LIGO strain data as a separate channel. End-to-end latency is O(10s).
- IDQ products show up in automated data quality pages.
- IDQ is not yet automatically applied to search results - this step will be enabled in advanced LIGO's next observing run.
- IDQ is being used for offline results and has shown great promise in helping to classify black hole detections when only one of the LIGO gravitational wave detectors is operating.
- This project has developed substantial cyber-infrastructure that has operated robustly and provided the start of a framework to run other machine learning applications on real-time LIGO data in a production environment.

Details:

https://git.ligo.org/lscsoft/gstlal
https://git.ligo.org/reed.essick/iDQ
https://docs.ligo.org/reed.essick/iDQ/

PHY-1454389
PHY-1700765
OAC-1642391