NSF HDR ML A3D3: Detecting Anomalous Gravitational Wave Signals

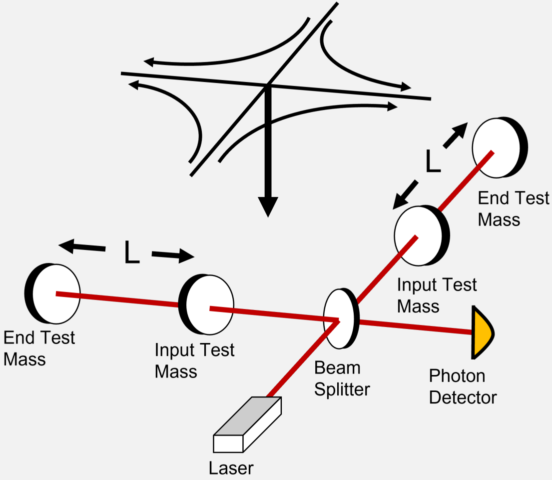
Chia-Jui Chou

National Yang Ming Chiao Tung University, Taiwan

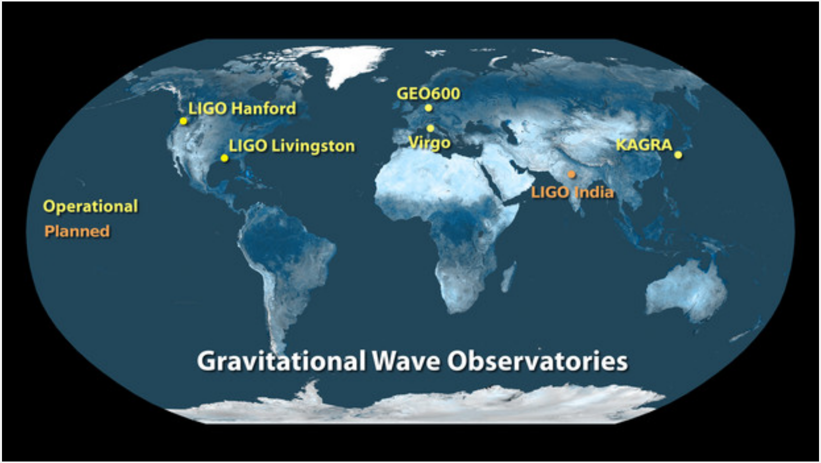
NSF HDR Hackathon in Taiwan@2024/12/23

Accelerated AI Algorithms for Data-Driven Discovery

Gravitational Wave Detectors



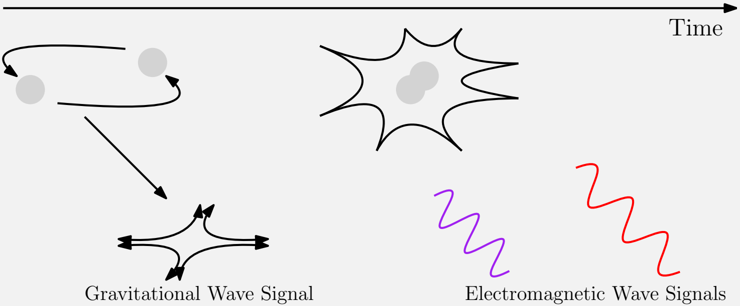
Gravitational Wave Detectors



Sources of Gravitational Waves

- Compact Binary Coalescence:
  - Black Hole-Black Hole Mergers,
  - Neutron Star-Neutron Star Mergers,
  - Black Hole-Neutron Star Mergers.
- Bursts:
  - Core-collapse Supernovae, Neutron Star Glitches, etc.
- Continuous Waves:
  - Spinning Neutron Stars,etc.
- Stochastic Background:
  - Cosmological background, Astrophysical background, etc.

Multi-Messenger Astronomy (MMA)



Sky localization of the "Known GW Sources" from their GW signals and send out alerts in low latency for the EM telescopes to capture the follow-up EM wave signals.

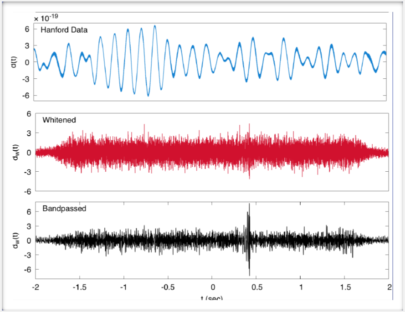
Searching for Anomalous GW Signals

- The unknown "Unknown GW Sources" are called "Anomalous". The conventional way to search such unmodelled GW signals is "Coherence Search" which analyze the coherenece between different detectors simultaneously.
- Now we would like to develop a semi-supervised approach to discover anomalous signals without explicit modelling.

Data Preprocessing

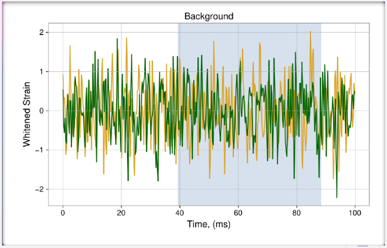
- Whitening: The original time series are transformed by the Power Spectral Density (PSD) so different frequency components are comparably scaled.
- Bandpassing: The filter is then applied to remove the low-frequency ($< 30Hz$) and high-frequency ($> 1500Hz$) components which are too noisy.

- 2-second time series with the sampling rate = 4096 Hz



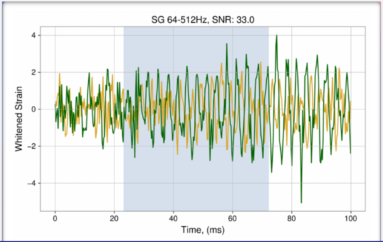
Background Dataset

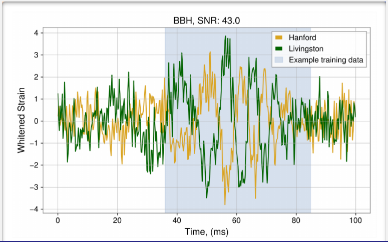
- The data is diveded into segments of 50 milliseconds, which contains 200 data points ($50 \times 10^{-3} s \times 4096\ samples/s \approx 200 samples.$)
- The dimension of the input data is $(N, 200, 2)$ where N is the number of the data segments. The last 2 corresponds to the data streams from the 2 LIGO detectors: Hanford (H1) and Livingston (L1).



Signal Dataset

- The data is diveded into segments of 50 milliseconds, which contains 200 data points
  $(50 \times 10^{-3} s \times 4096\ samples/s \approx 200 samples.)$
- The dimension of the input data is $(N, 200, 2)$ where N is the number of the data segments. The last 2 corresponds to the data streams from the 2 LIGO detectors: Hanford (H1) and Livingston (L1).





Example Submission

```python
import numpy as np

class Model:
    def __init__(self):
        # You could include a constructor to initialize your model here, but all calls will be made to the load method
        self.clf = None

    def predict(self, X):
        # This method should accept an input of any size (of the given input format) and return predictions appropriately
        return np.array([0 for _ in range(len(X))])

    def load(self):
        # This method should load your pretrained model from wherever you have it saved
        pass
```

Github Link

References

- Codabench: https://www.codabench.org/competitions/2626/
- Notebook: https://colab.research.google.com/drive/1hatkYT5Xq6qauDXY6xl
- Paper: MLST 10.1008/2632-2153/ad3a31
- Github issue: https://github.com/a3d3-institute/HDRchallenge/issues

Thank you