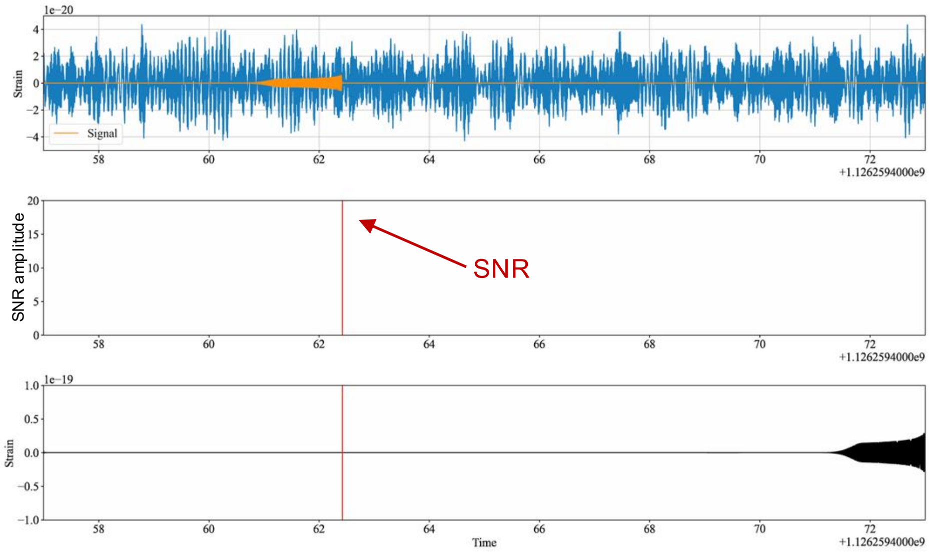
Neural network classifiers for distinguishing signals from instrumental noise

Melissa Lopez
ICERM 2025

Modelled searches: matched filtering (MF) for CBC

What is matched filtering?

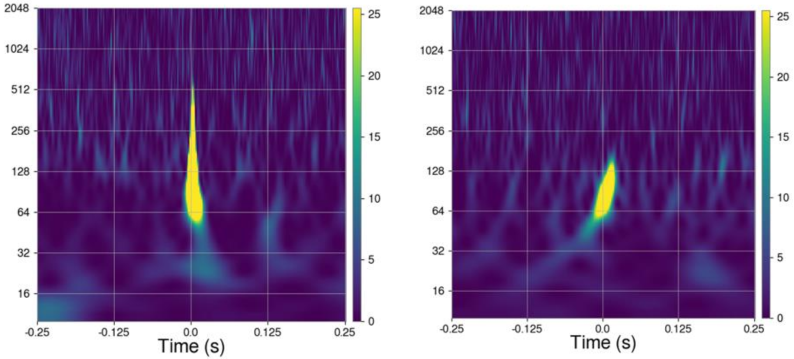
Unknown signal * Template Bank (simulated GW) → Triggers



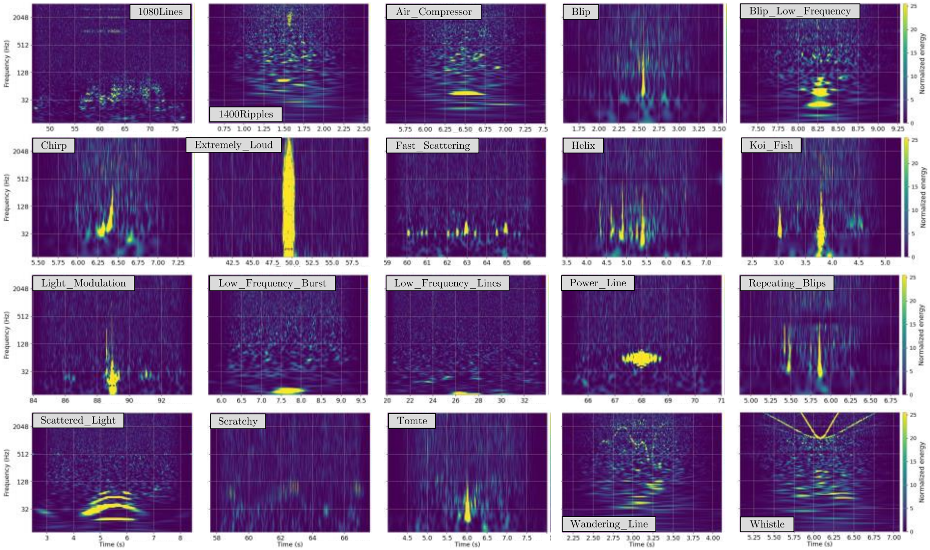
Idea: unknown signals generate *multiple* triggers. Can we find *patterns* with Machine Learning?

Transient noise burst (glitches)

- Caused by instruments or environment (known or unknown)
- Diminish scientific data available
- Hinder GW detection (mask and/or mimic)

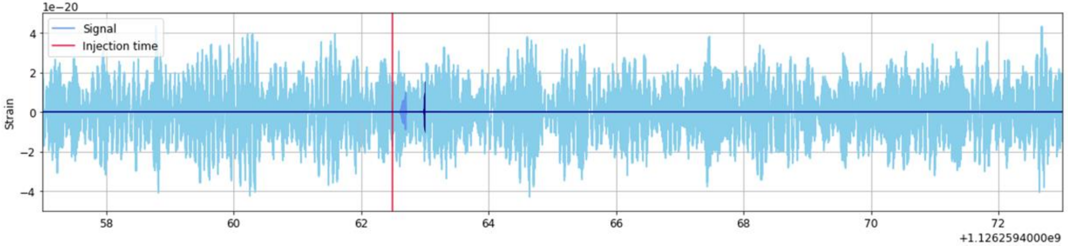


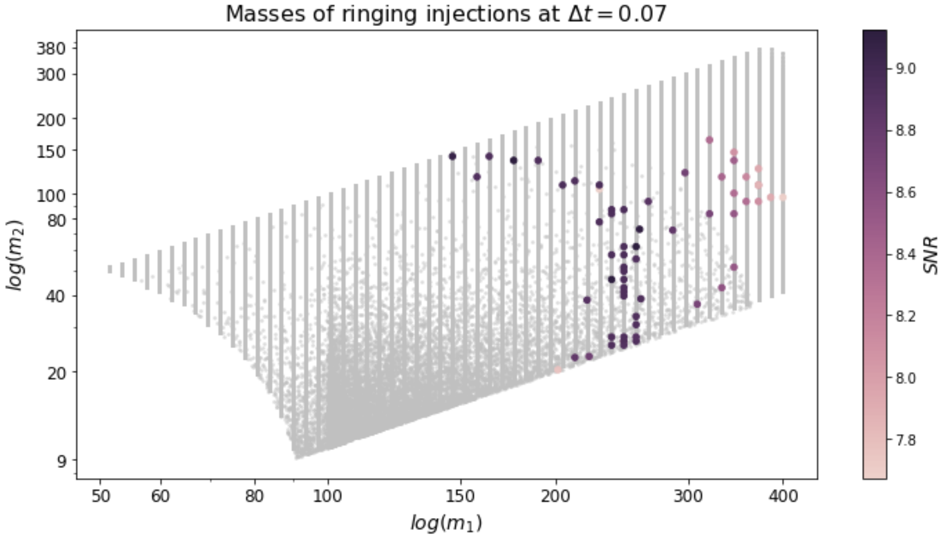
Example of a blip glitch (left) and a intermediate-mass black hole (right)



A simulated GW through a detection pipeline

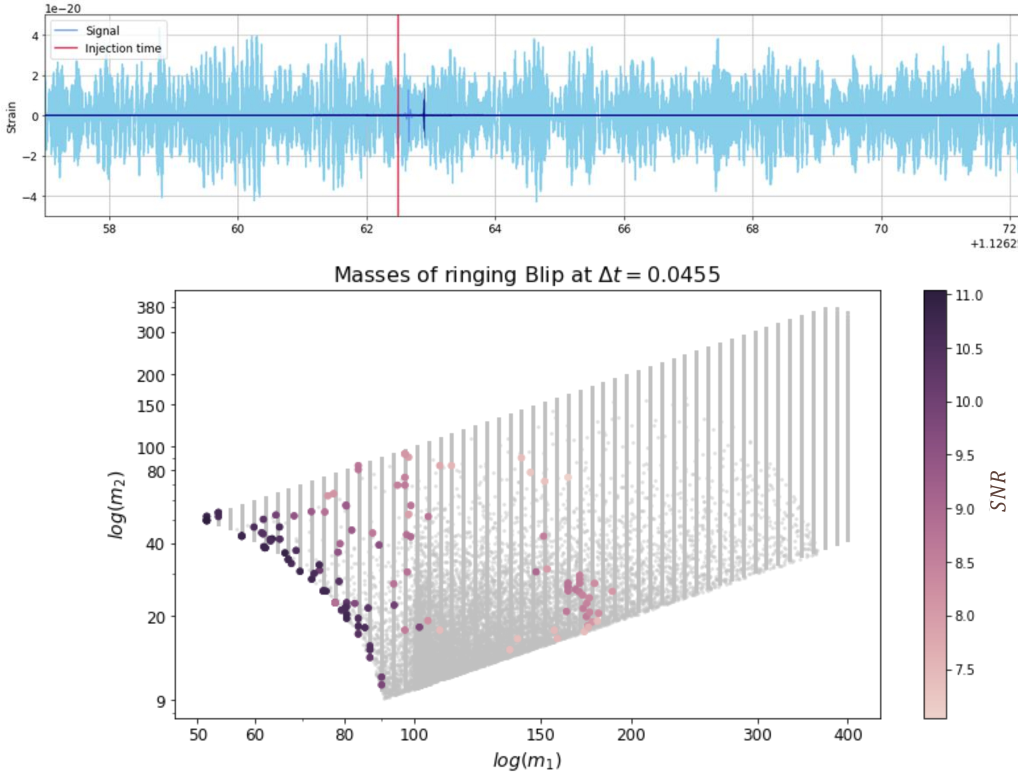
Δt: time when trigger happened – time when GW signal was added to the noise





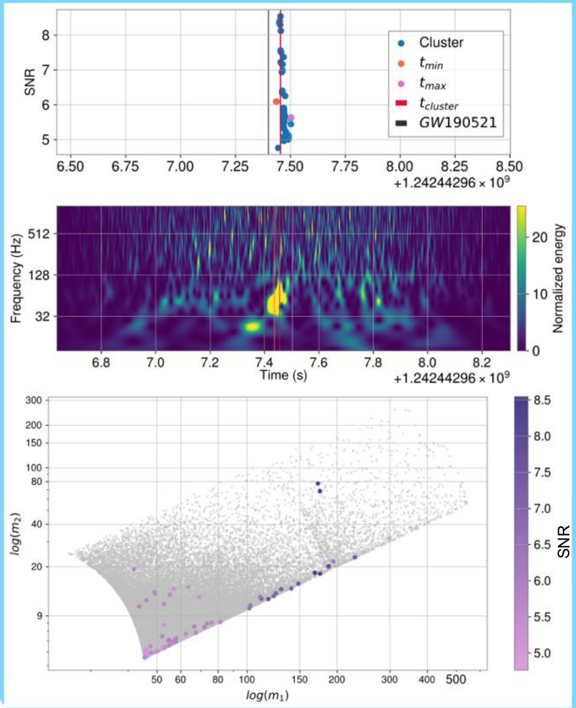
A glitch through a detection pipeline

Δt: time when trigger happened – time when glitch happened

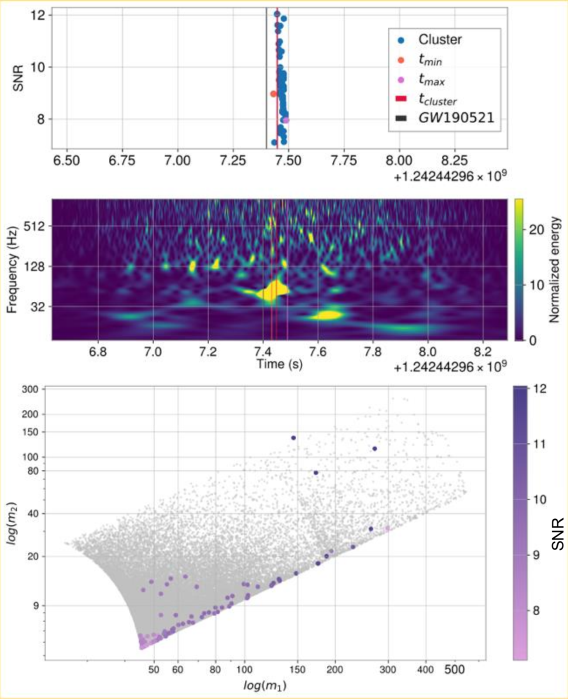


GW190521

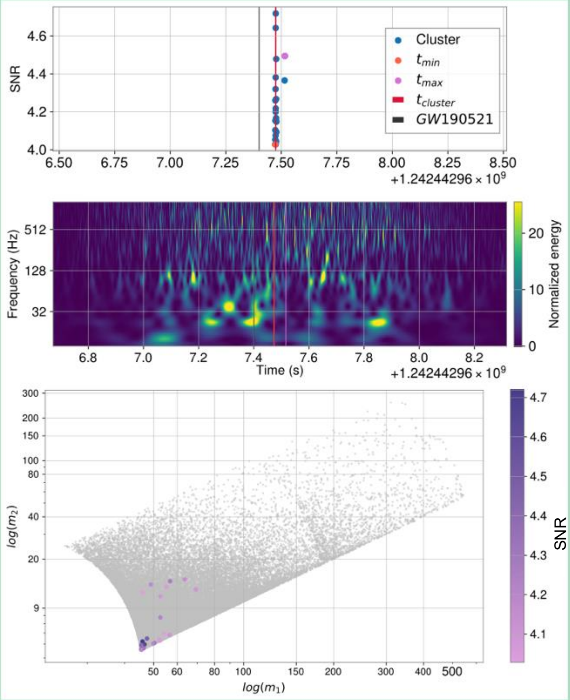
LIGO Hanford (H1)



LIGO Livingston (L1)



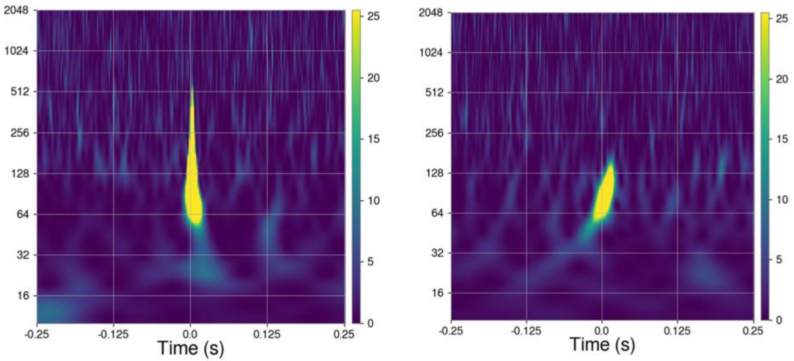
Virgo (V1)



Motivation

Context: intermediate-mass black holes (IMBH) are the missing link between stellar black holes and supermassive black holes, but they are hard to detect!

Idea: use triggers from matched filtering (free information) from detection algorithms to learn the background (glitches) and foreground (GW signals) with ML



Example of a blip glitch (left) and a IMBH (right)

- MF searches use *strict* conditions for detection.
- Can we *relax* the search with the interpolation ability of ML?

Multi-class classification demo

Demo data from PhysRevD.111.103020 - arXiv 2412.17169

SCAN ME

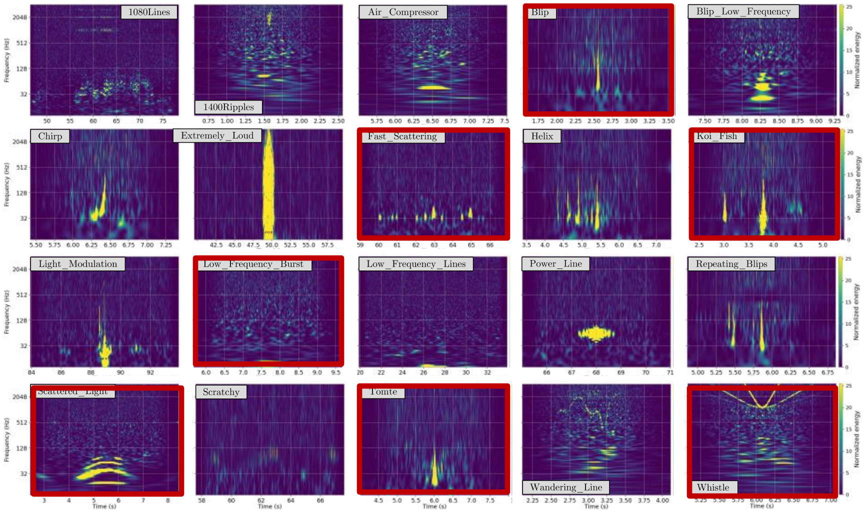
Task: Distinguish IMBH from different glitch classes in <u>single detector</u> → we have 3 detectors!

Algorithm: Multi-layer perceptron (MLP)

Input: Adding time is hard, so let's simplify the problem. Each template is defined by $m_1, m_2, s_{1z}, s_{2z}, \chi^2, SNR$. We weight average by SNR to get the feature vector

$$\mu(m_1, m_2, s_{1z}, s_{2z}, \chi^2, SNR)$$

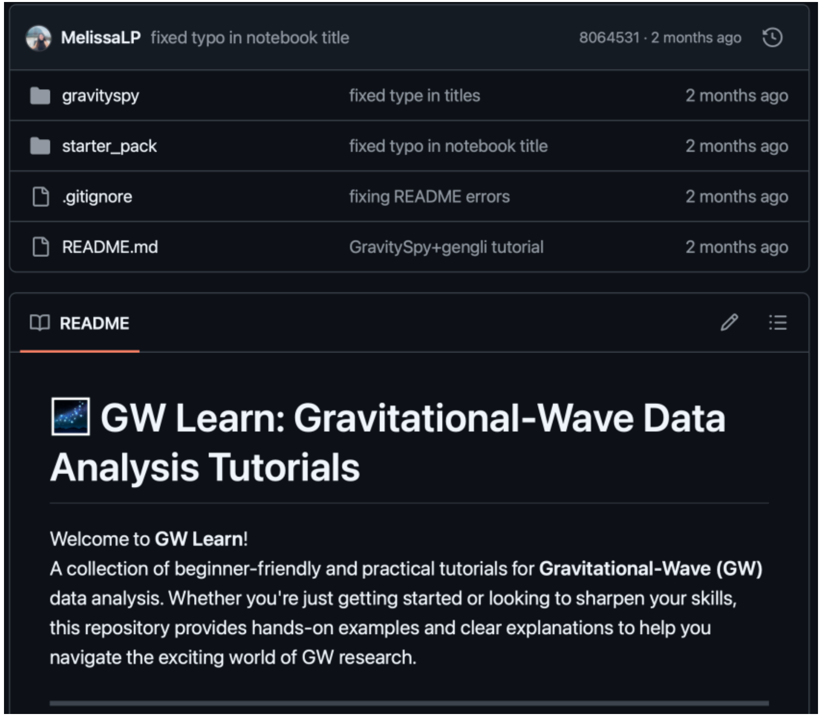
Output: class probability



Idea: MLP differentiates **6 classes**: 5 different types of background (glitches) and single foreground (GW signals). It uses only **6 parameters** in **single detector**

About today's tutorial

⚠️ Promotion time! gw_learn tutorials

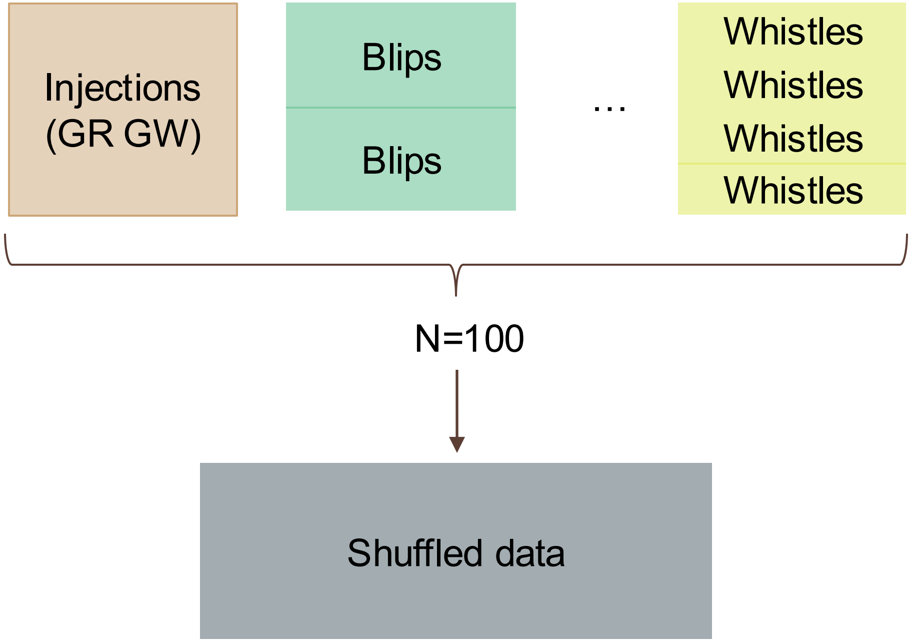


Acess tutorial of today:

https://shorturl.at/vxMH4

Dealing with imbalanced data

1. Accounting for imbalanced data (boostrapping with replacement)



Mentimeter

Exercise 1: What is bootstrapping and why is it important?

This is a trick question -- E. T. Jaynes dixit

I'm not sure — I'm a physicist and I've ever heard of bootstrapping. Is it like dimensionality reduction or some kind of data compression technique?

Resampling

Computing uncertainties when you don't know what else to do

resampling

A way to estimate the unknown sampling/data distribution using the current data to derive estimators for target (usually Frequentist) statistics.

When you want to report uncertainties but don't have the ability/compute/will to draw additional samples from the distribution of interest.

I think it's important for uncertainty quantification

3

14

Mentimeter

Quote

– *Who*

Mentimeter

Exercise 2: What are the main problems with this network?

Too shallow

Not wide enough

too small.

Data are not normalized

It’s very small — it only has two layers. Maybe include more layers and also include a variable learning rate, as well as dropout. Maybe it’s overfitting? I noticed the accuracy goes up only to .75.

Input data not scaled appropriately

Neural network architecture is a "prior" over the structure you expect in your data. MLP assumes just a general curve, and with few parameters/layers the functions are probably too simple.

Change the activation to RELU may better performance.

4

11

Mentimeter

What would you like to talk about? Short link: https://shorturl.at/Blil6

Incorporating diagnostic data channels from GW detectors in searches directly by assimiliating them using ML stages, and using that to discriminate signals from noise transients invariantly

How to design a search pipeline around ML in practice?

Calibration uncertainty reduction using ML

At which SNR levels will template-banks-based methods become too computationally expensive? Will they at all?

How would we actually verify a detection of an unmodeled signal by a ML algorithm?

What can traditional searches learn from ML?

Is an end to end joint model for noise transients and astrophysical sources feasible with ML?

how often does the distribution of noise change during the run roughly? Or how often do ML models need to be trained/recalibrated during the obs run?

5

27

What would you like to talk about? Short link: https://shorturl.at/BIil6

ML for unmodelled/unknown signals

5

27