Machine Learning to Boost Multimessenger Astrophysics

Zsuzsa Marka

Columbia Astrophysics Laboratory

MG17, Pescara, July 12, 2024

Circa 2006

Prospects of gravitational wave data mining and exploration via evolutionary computing

M. Lightman, J. Thurakal, J. Dwyer, R. Grossman, P. Kalmus, L. Matone, J. Rollins, S. Zairis, S. Márka

704 Pupin Laboratory, Physics Department Columbia University New York, NY 10027

The expected signals have a very low signal to noise ratio (SNR) with uncertain and varied waveforms.

Evolutionary computing based search algorithms can be trained to find signals in non-Gaussian noise.

We used realistic interferometric detector noise as opposed to Gaussian noise to evolve the population of our algorithm.

Tests with Burst injections (SG, Q:8.9, 250Hz) at 90%CL:

generation 100:	2.7×10^{7} false alarms / year for peak amplitude 3.5
	2.3×10^{7} false alarms / year for peak amplitude 5
generation 199:	1.6×10^{5} false alarms / year for peak amplitude 3.5
	6.5×10^{-6} false alarms / year for peak amplitude 5

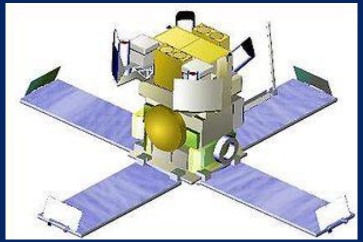
It took roughly five days to complete 199 generations on a 64 bit, 1.8 GHz processor with a 1 MB cache and 1 GB DDR PC3200 RAM.

J. Phys.: Conf. Ser. **32** 58 (2006)

Basic Glossary: Multimessenger Approaches

"Multi-messenger astrophysics": connecting different kinds of observations of the same astrophysical event or system

"ExtTrig" strategy:

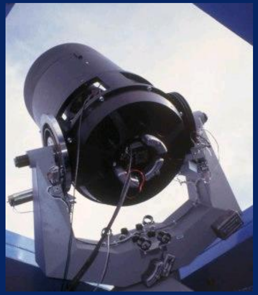


Telescopes, Satellites or other external entities → Flow of trigger information → GW Search

First exttrig search GRB030329

"Follow-Up" strategy:

GW Data → Flow of trigger information → Telescopes, Satellites or other external entities

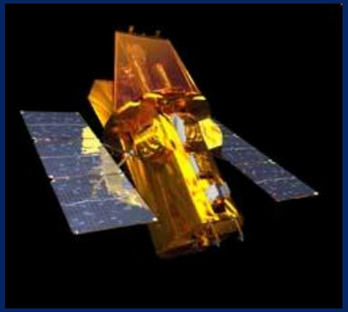


First EM follow campaign: summer 2007 pilot project

"joint search" strategy:

GW detector →
GRB detector →
HEN detector →
++ →

Low **L**atency **A**lgorithm for **M**ultimessenger **A**strophysics



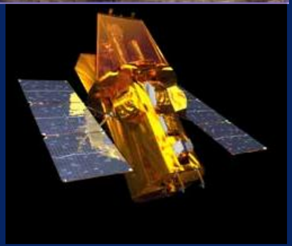
First search concept: 2006 in ***low-latency*** since O2 (2017)

Realtime Data Analysis Infrastructure

Concurrent Observations → Analyze Disparate Streams => Metadata → Combine Datastreams → Model-based Prediction → Trigger Additional Observations

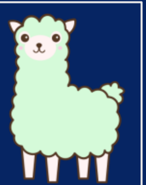
Dig deep, below noise level => subthreshold trigger sets

“Multi-messenger astrophysics”: connecting different kinds of observations of the same astrophysical event or system



GW detector →
GRB detector →
HEN detector →
++ →

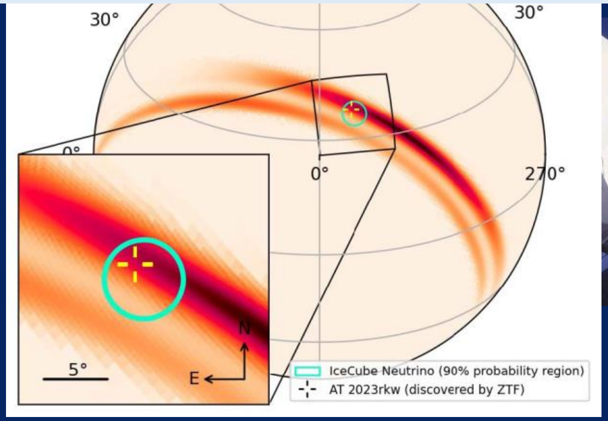
Low Latency Algorithm for Multimessenger Astrophysics



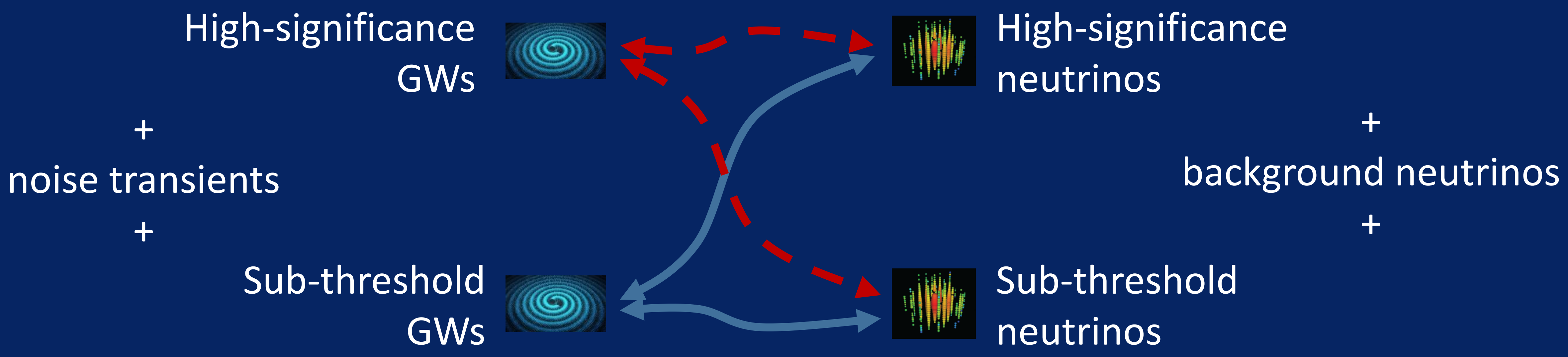
Telescopes, Satellites or other external entities →

How to Search for Multiple Messengers—A General Framework Beyond Two Messengers
Veske et. al., ApJ **908** 216 (2021)

Example: EM search triggered by GCN Circular 34616 - LIGO/Virgo/KAGRA S230904n: 1 counterpart neutrino candidate from IceCube neutrino searches



Multimessenger Search for GW+HEN sources with complete neutrino and GW datastreams

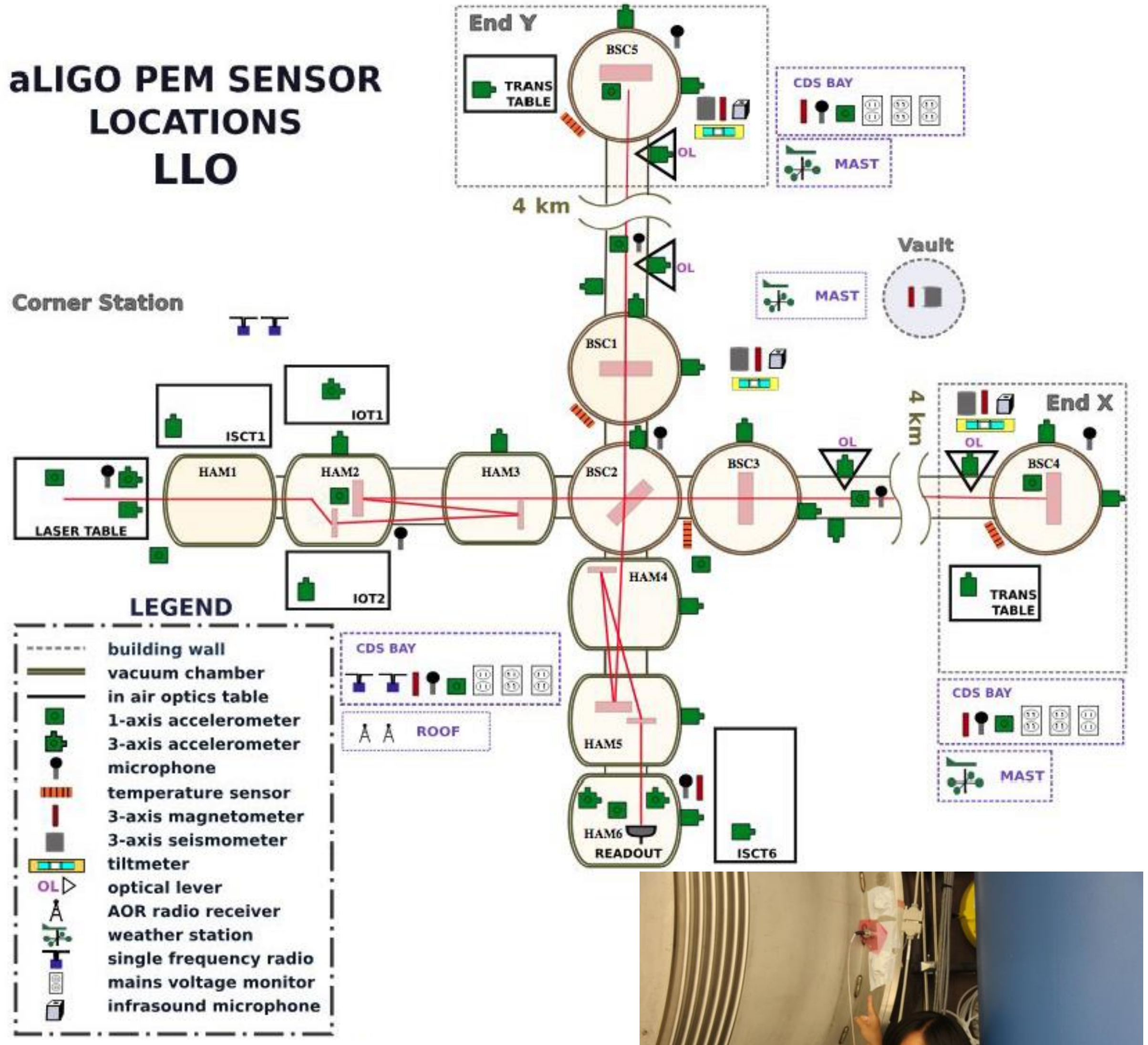


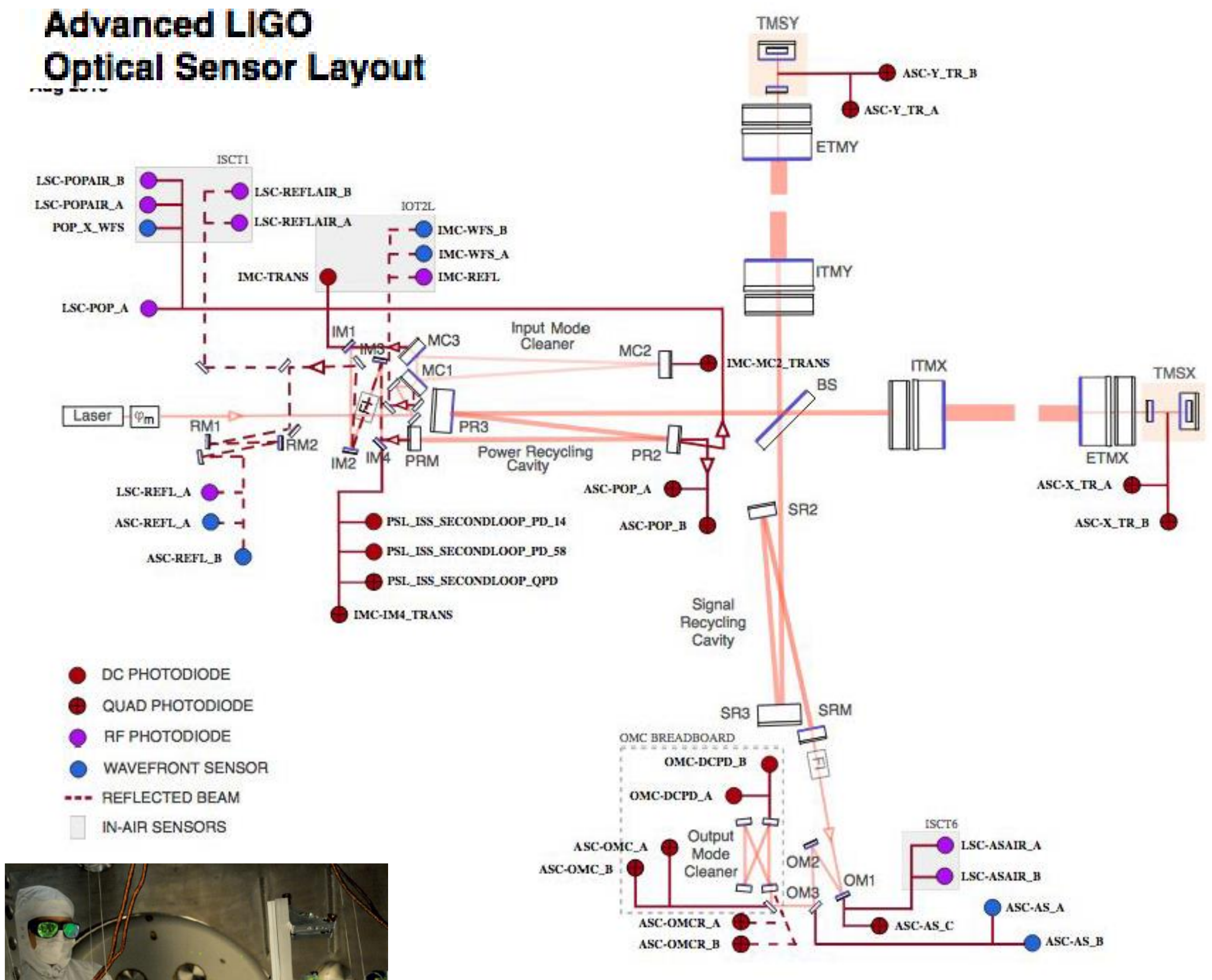
Statistical properties of the ***detector noise background*** intertwined with common detections and even ***more sub-threshold GW events***

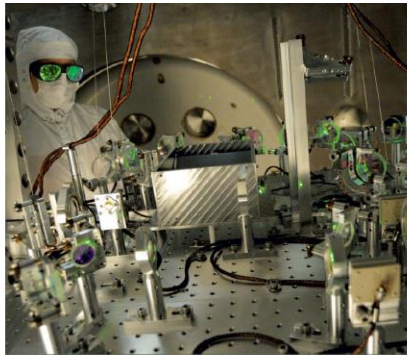
=> how to determine the background => key in estimating of significance / joint significance

Machine Learning approaches can help in both!

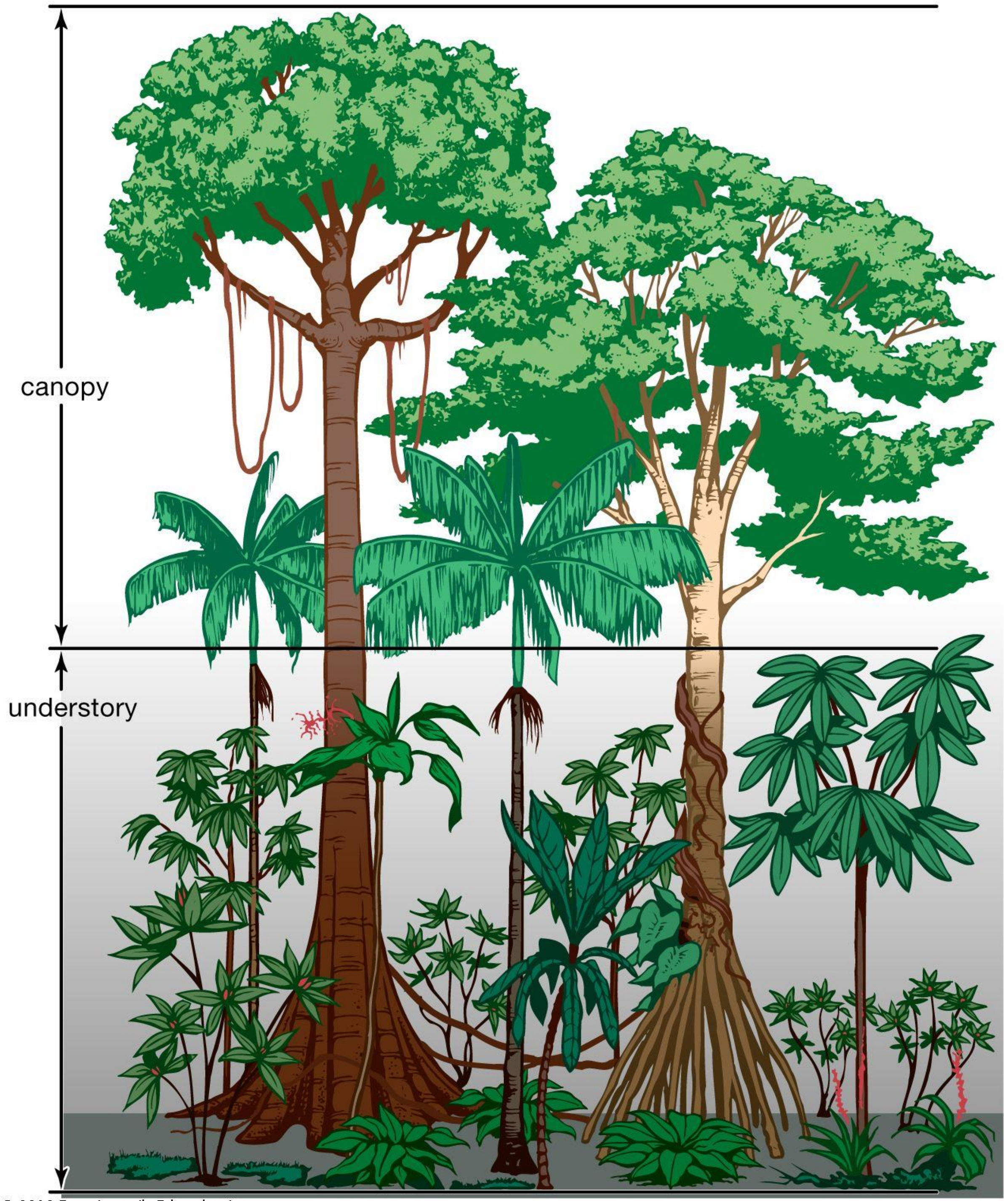
>200,000 auxiliary channels per site (both statistical and computational challenge)





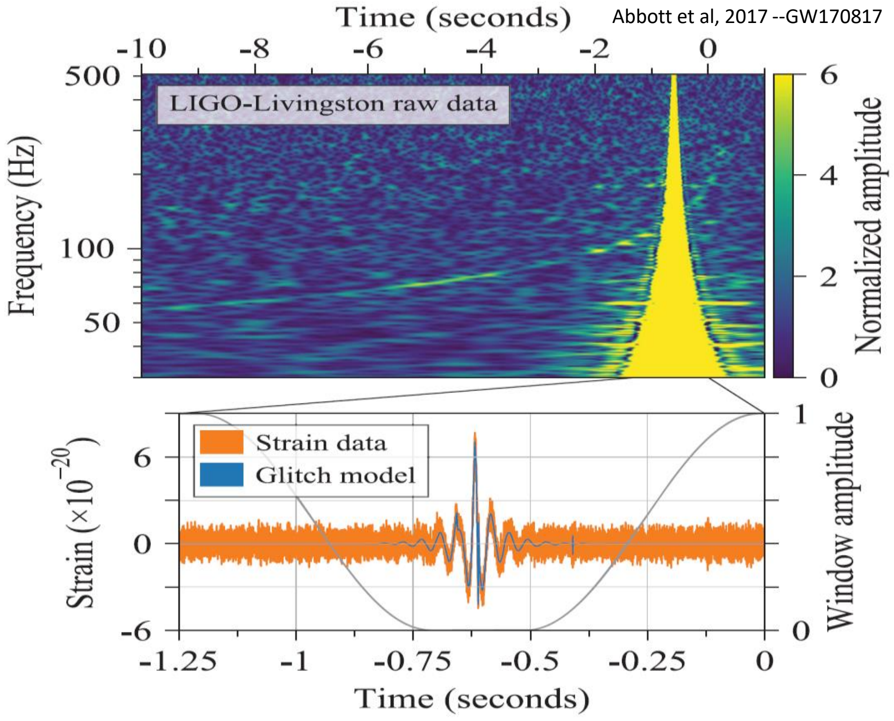


from LIGO Document G1801206, R Corley et al



Glitches

Abbott et al, 2017 --GW170817



Current Status Quo

- Observe glitches (what an interesting tree?)
- Find correlations with aux channels (certain trees grow better in specific environment)
- **veto** a section of h(t) (cut a tree when it bothers us, but keep growing the type, as we hope we can recognize it again)

Ideally..

- If we know what makes a certain type grow we remove the cause

$\Rightarrow$ Lets get rid of the glitches for forever

or at least establish a glitch model and remove its effect from data

Can we find glitches without looking at GW strain?

Auxilliary channels

-witness

-cause

Not sensitive to GWs, so won’t mistakenly identify a true event as a glitch

Can we use all >200k aux channels?

Independent corroboration

Information about origin not just diagnostics

Glitches

Generally identified via excess power methods on the time frequency plane

OmegaScan (J. Rollins thesis –Columbia U., 2010)

Omicron (F. Robinet et al, 2015)

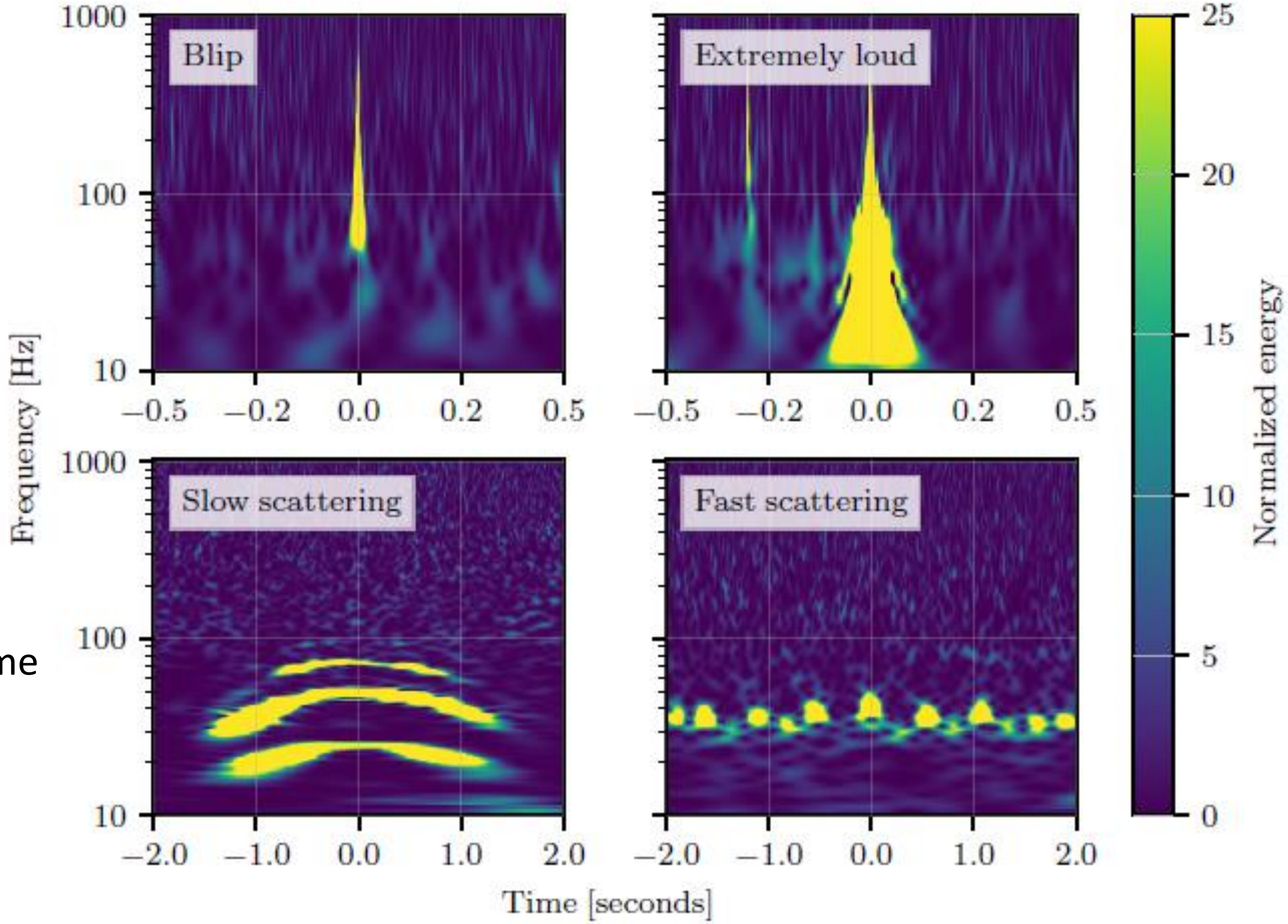
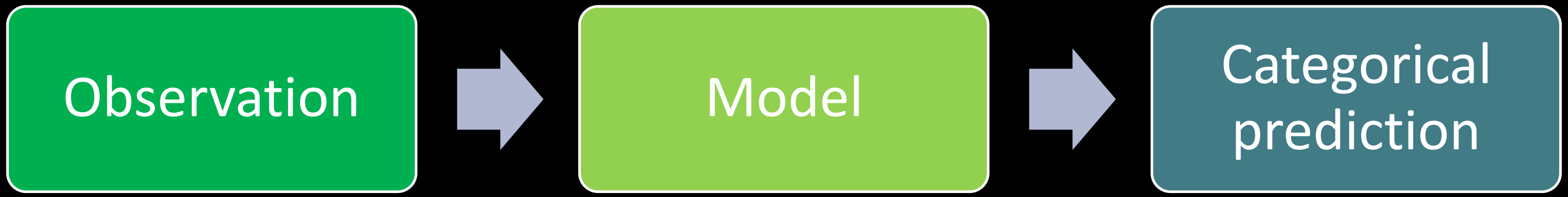


Figure from D Davis et al, 2021 – uses GravitySpy (Zevin etal 2017)

The binary classification problem



The binary classification problem

Observation → Model → Categorical prediction

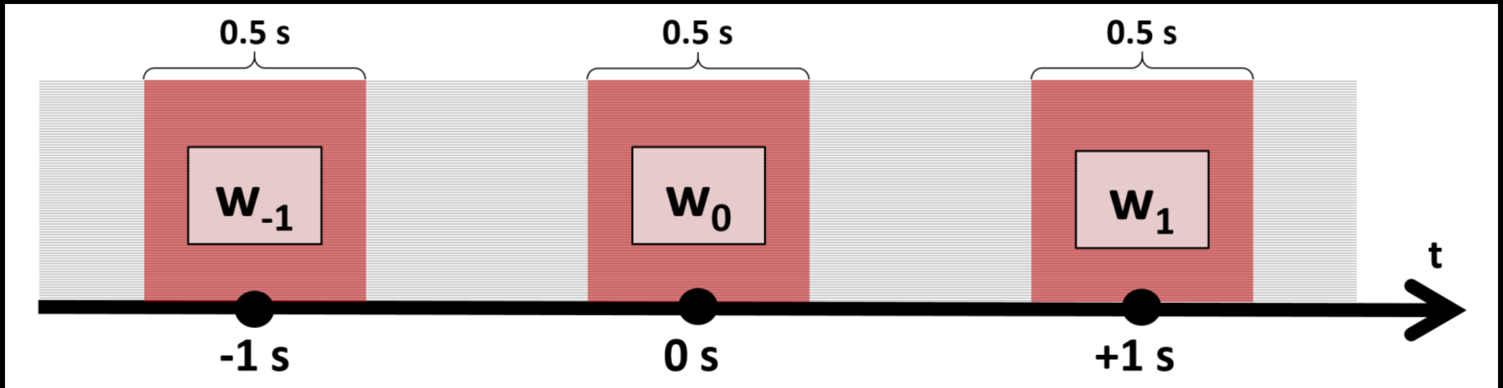
Glitch identification from auxiliary channels using time series data

- Data at time t: glitch (1) or not (0)?
- Our pipeline:

Channel reduction → Feature extraction → Normalization → Model → Prediction (glitch or not)

Colgan, Robert E., K. Rainer Corley, Yenson Lau, Imre Bartos, John N. Wright, Zsuzsa Márka, and Szabolcs Márka. "Efficient gravitational-wave glitch identification from environmental data through machine learning." *Physical Review D* 101.10 (2020): 102003.

Feature extraction



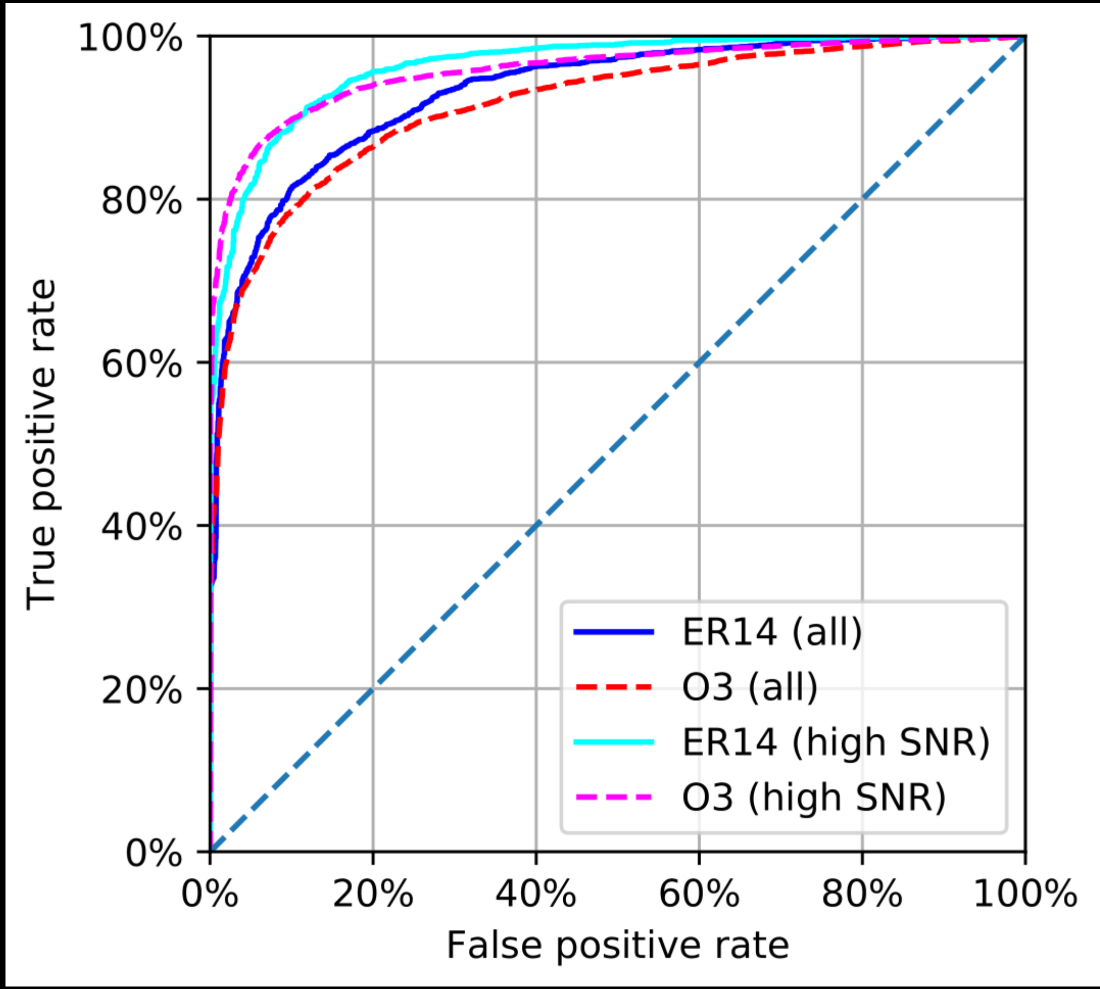
- 3 windows around t_0 to capture surrounding behavior
- μ_i and σ_i: mean and standard deviation over window w_i
- 10 features per channel at each time:

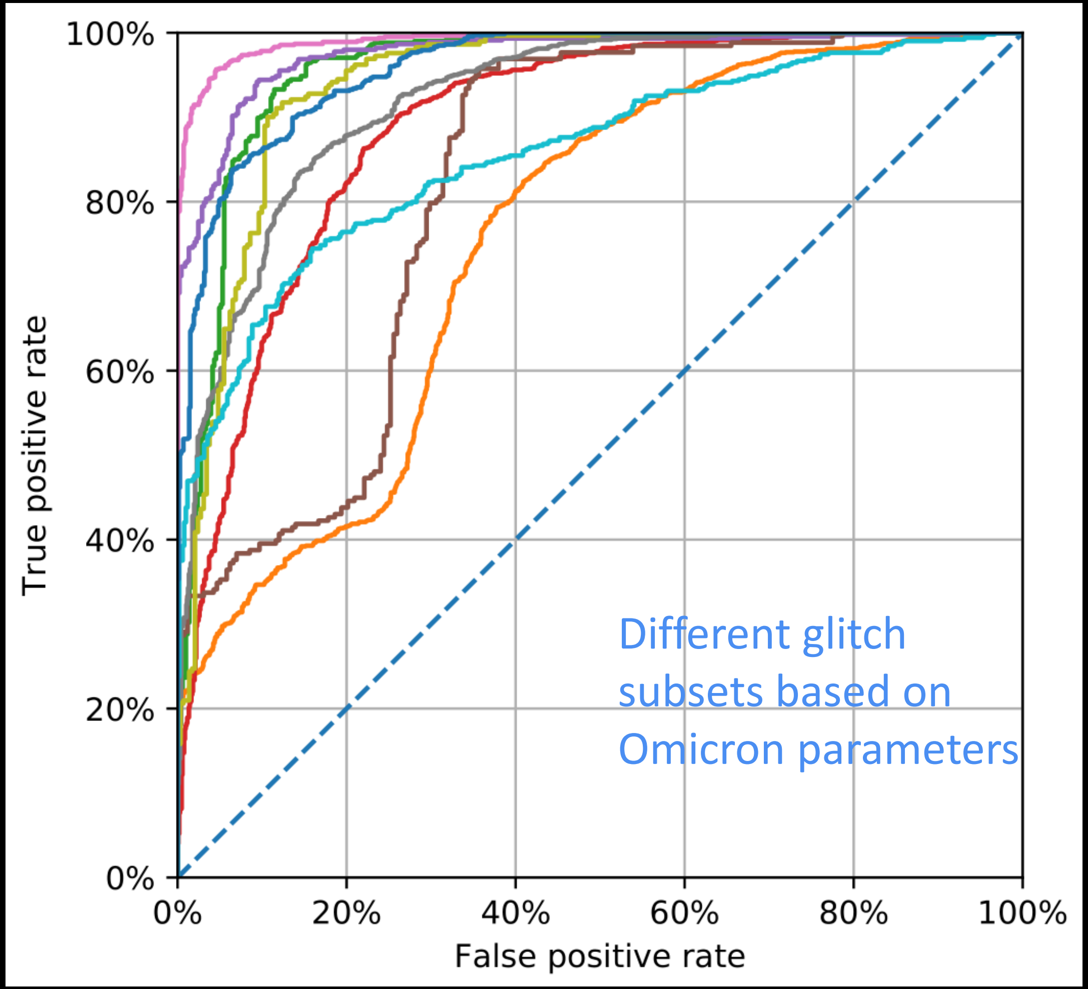
$$\mu_{-1} \qquad \mu_0 \qquad \mu_1 \qquad \mu_1 - \mu_{-1} \qquad \mu_0 - \frac{\mu_1 + \mu_{-1}}{2}$$

$$\sigma_{-1} \qquad \sigma_0 \qquad \sigma_1 \qquad \sigma_1 - \sigma_{-1} \qquad \sigma_0 - \frac{\sigma_1 + \sigma_{-1}}{2}$$

Colgan, Robert E., K. Rainer Corley, Yenson Lau, Imre Bartos, John N. Wright, Zsuzsa Márka, and Szabolcs Márka. "Efficient gravitational-wave glitch identification from environmental data through machine learning." *Physical Review D* 101.10 (2020): 102003.

Elastic net logistic regression model – Test Results





Colgan, Robert E., K. Rainer Corley, Yenson Lau, Imre Bartos, John N. Wright, Zsuzsa Márka, and Szabolcs Márka. "Efficient gravitational-wave glitch identification from environmental data through machine learning." *Physical Review D* 101.10 (2020): 102003.

Can we learn the behavioral signatures in auxiliary channels that give rise to glitches in the GW strain?

Previously:

Some channels have shorter or longer timescale correlations with certain types of glitches.

Some channels might consistently behave in a specific pattern before (or after) a glitch, but there is no reason to suspect it would be the same such pattern for all of them—or that such patterns could easily be captured by the features we happened to select

Hand-designed, inflexible features: probably *suboptimal*

Feature Learning:

Recent ML progress driven by learned representations, end-to-end models trained on raw data

Features should capture most useful properties of raw data behavior

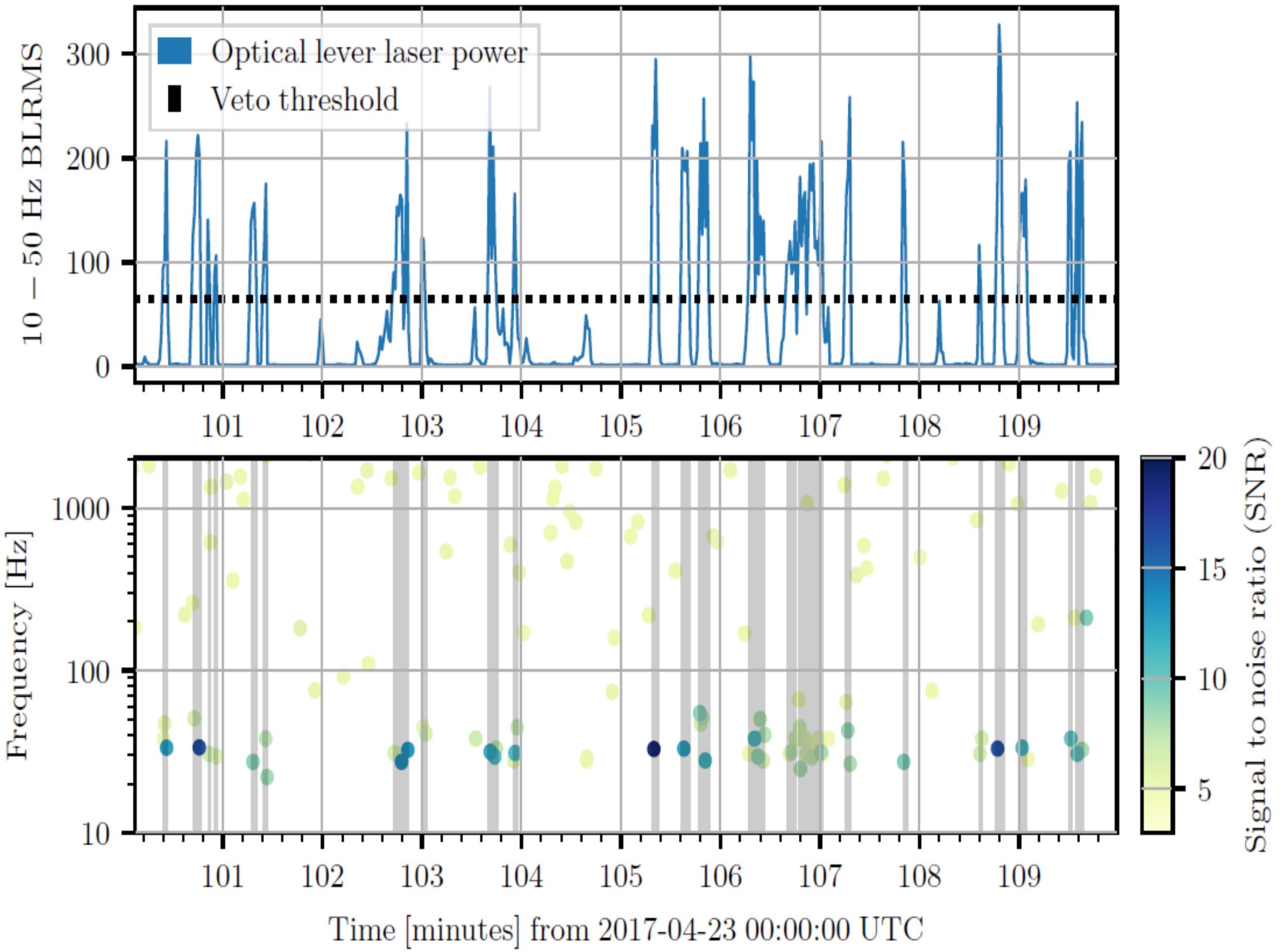


Figure from D Davis et al, 2021

Architectural Optimization and Feature Learning for High-Dimensional Time Series Datasets

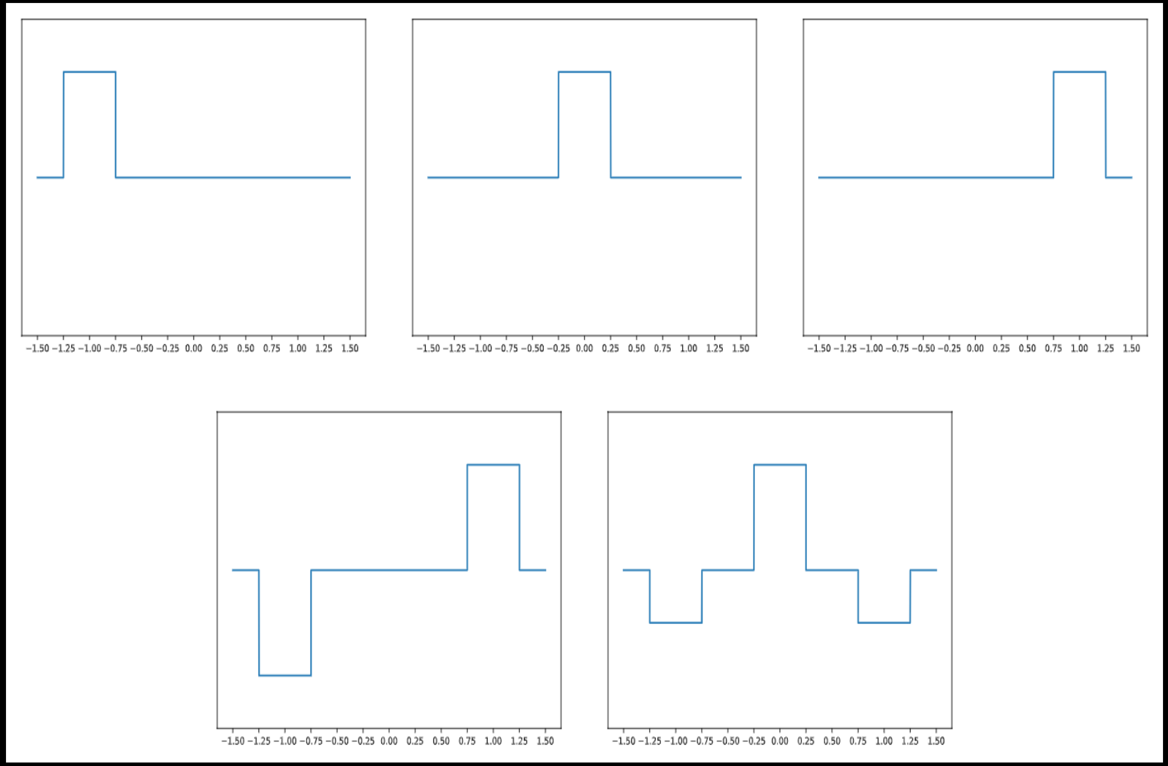
Convolution: measures similarity between two signals

Convolutional Neural Networks: learn convolutional filters from raw data, then aggregate for decision

Especially well-suited to data with temporal or spatial structure
- Images
- Time series (e.g. LIGO channels)

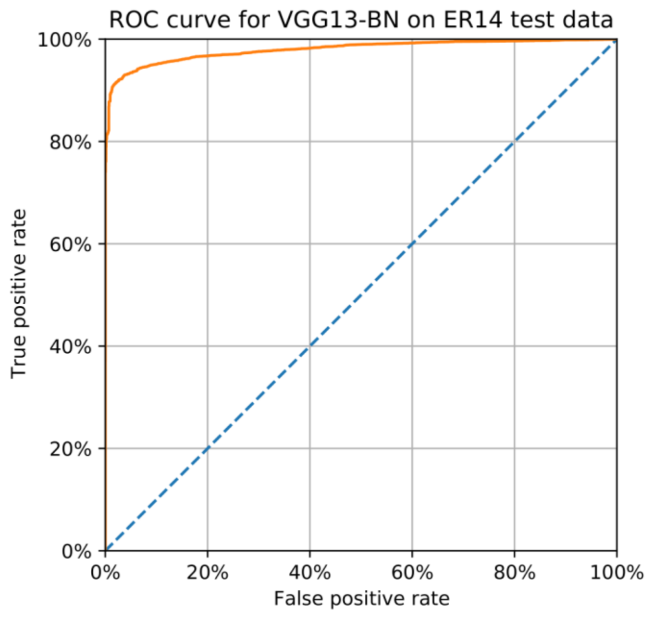
Similar training to classical ML methods
- Use labeled data to compute model error
- Gradient descent
- But more parameters to optimize

The 5 mean-based features from our previous study visualized as convolutional filters



Colgan, Robert E., Jingkai Yan, Zsuzsa Márka, Imre Bartos, Szabolcs Márka, and John N. Wright. "Architectural Optimization and Feature Learning for High-Dimensional Time Series Datasets” arXiv:2202.13486 2022

Deeper models



Increased depth consistently shown to improve performance over equivalent size shallower models

Model	Feature Learning?	Depth	Nonlinear?	Pooling?	Best Val Loss (Acc)	Best Val Acc (Loss)	Test Acc
Fixed Features FF	✗	1	✗	✗	0.3392 (85.9%)	86.0% (0.3567)	85.8%
LF	✓	1	✗	✗	0.2376 (90.4%)	90.9% (0.2423)	89.6%
1Hid	✓	2	✗	✗	0.2385 (90.5%)	91.2% (0.2486)	89.3%
1HidReLU	✓	2	✓	✗	0.2330 (91.0%)	91.0% (0.2330)	91.0%
VGG6	✓	6	✓	✓	0.2010 (91.9%)	93.0% (0.2050)	94.0%
VGG13	✓	13	✓	✓	0.1956 (93.4%)	93.4% (0.1956)	93.6%
VGG13-BN	✓	13	✓	✓	0.1732 (93.1%)	93.6% (0.1822)	94.7%

Deeper models

Colgan, Robert E., Jingkai Yan, Zsuzsa Márka, Imre Bartos, Szabolcs Márka, and John N. Wright. "Architectural Optimization and Feature Learning for High-Dimensional Time Series Datasets" arXiv:2202.13486 2022

How can these methods be useful in investigating instrumental/environmental origin of glitches?

Strain-independent glitch verification

- Auxiliary channels not astrophysically sensitive
- Reduce likelihood of misidentifying a true event as a glitch

Channel selection

- Which channels are most often associated with glitches?
- Sparsity: identify a few tens to hundreds out of 200,000+

Learned features

- Behavioral signatures in auxiliary channels?

Individual glitch analysis

- **Interpretable model** indicates which channels contribute most to individual classifications

Application example to specific glitch type =>

"Scattered Light"

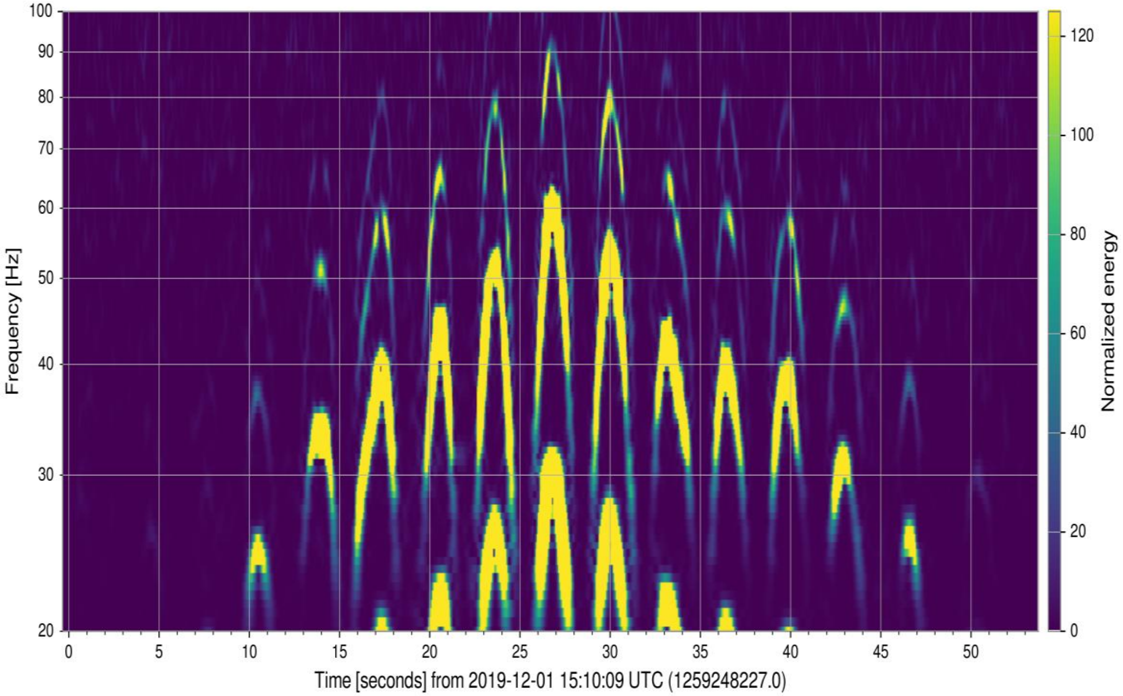
Very prevalent at Livingston site

Wide arches in low-frequency band

Previously observed to correlate with

increased microseismic activity (0.03-0.5 Hz)

Very high rate on 12/1/2019



Colgan, Robert E., Zsuzsa Márka, Jingkai Yan, Imre Bartos, John N. Wright, and Szabolcs Márka. "Detecting and Diagnosing Terrestrial Gravitational-wave Mimics Through Feature Learning." arXiv:2203.05086 2022

Applying our models

Train slightly modified LF (flat) model to distinguish between “Scattered Light” and “no glitch” (of any type)

- Scattered Light => GravitySpy
- No glitch => Omicron

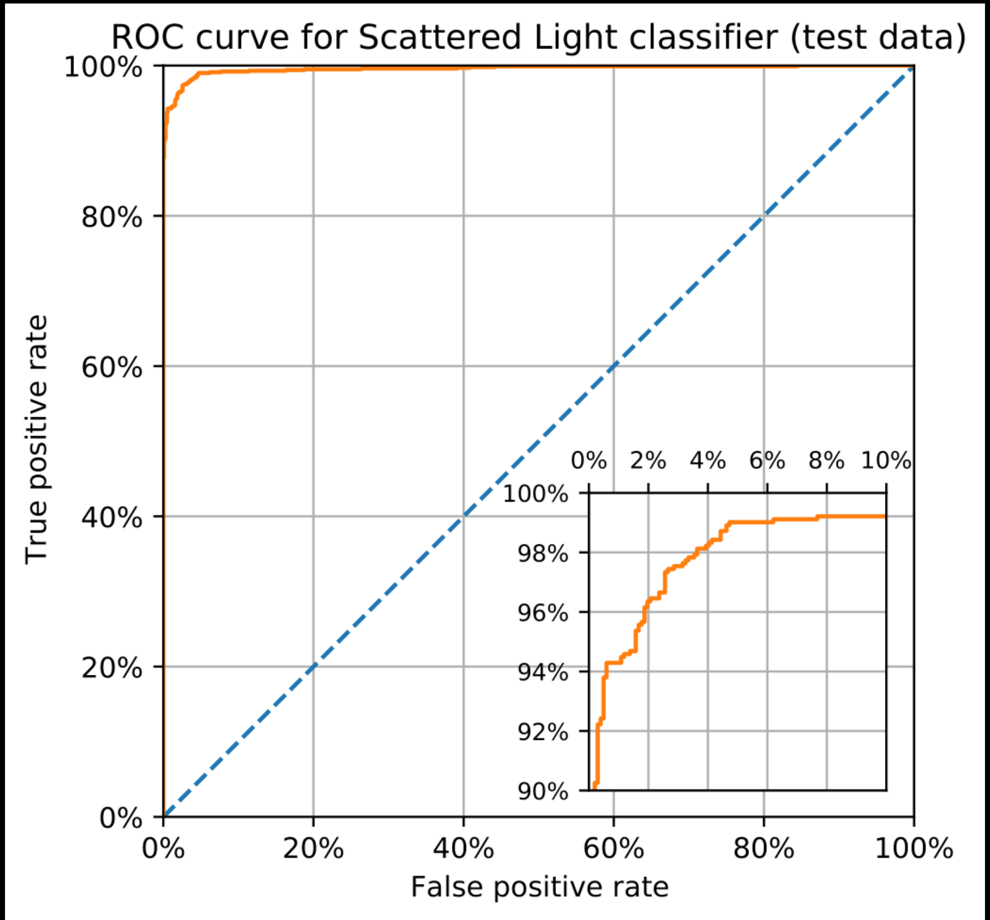
97.1% test accuracy

- 98.4% true positive rate
- 95.8% true negative rate

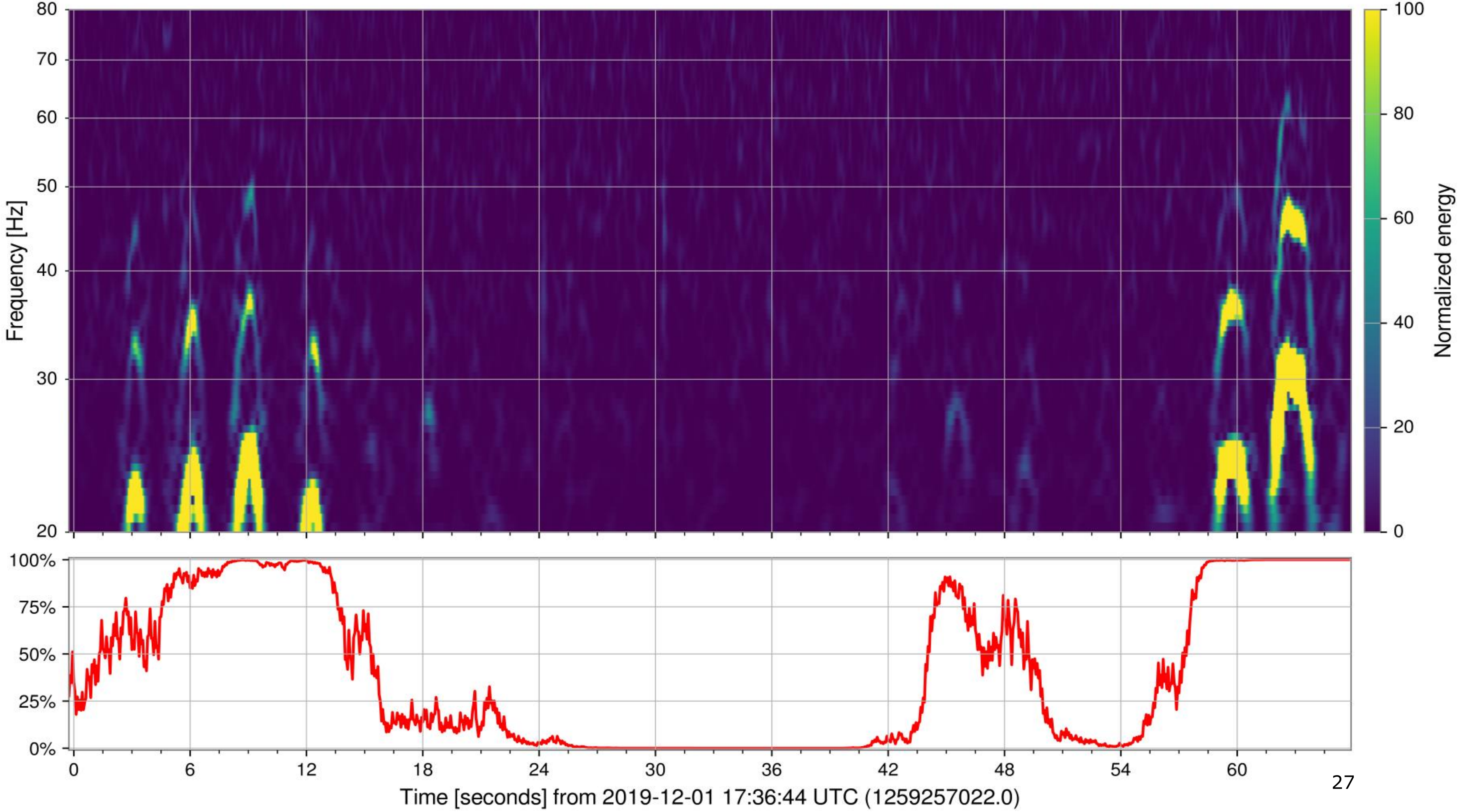
25 of 39,147 nonzero channels selected

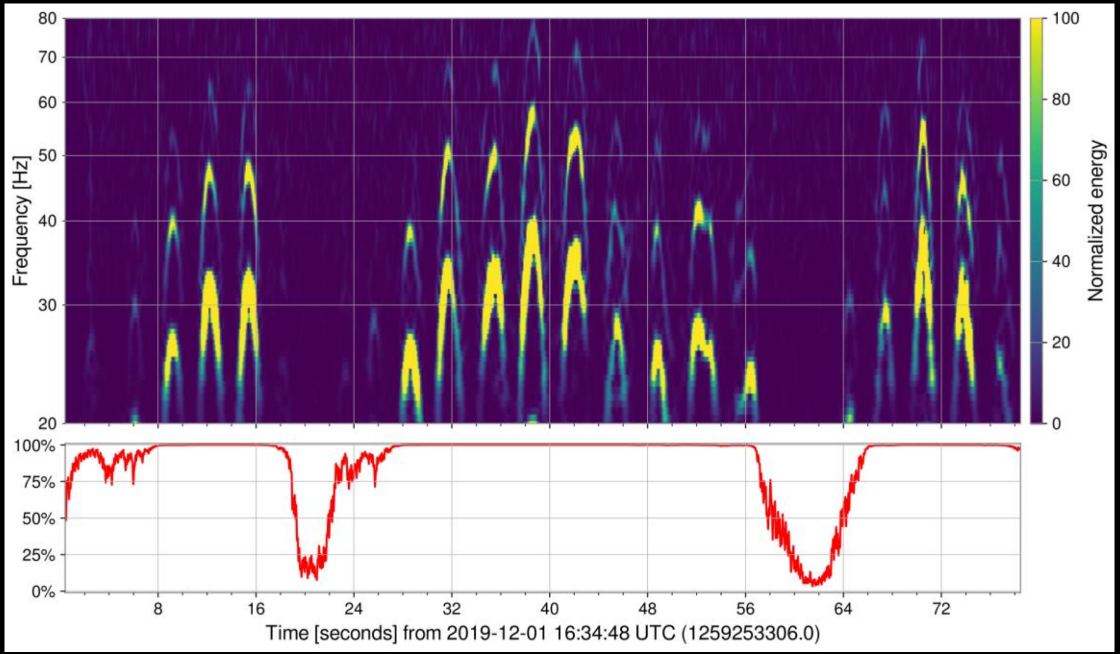
SUS-ETMX_L2 OSEM

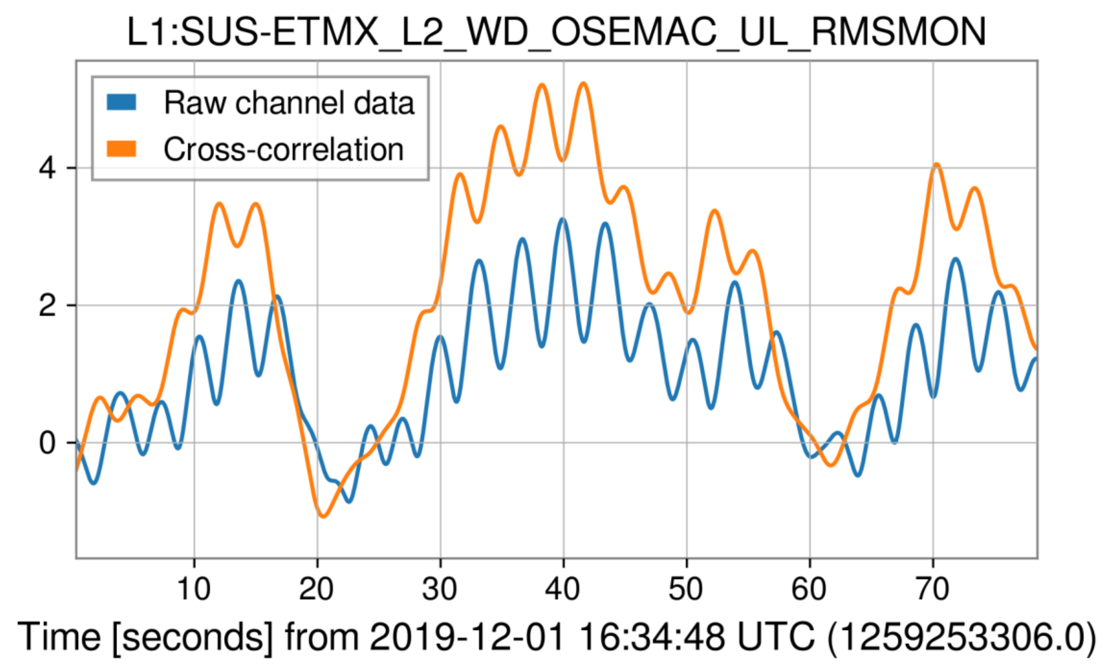
- SUS: suspension system
- ETMX: X arm end test mass
- L2: penultimate stage (immediately above test mass)
- OSEMAC: **O**ptical **S**ensing and **E**lectro**M**agnetic **AC**tuator



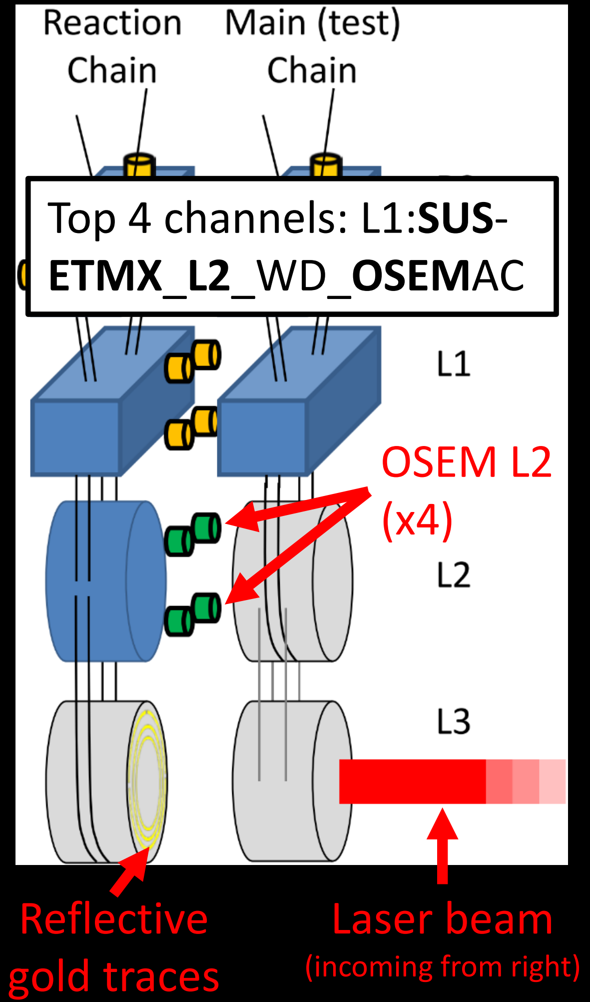
Colgan, Robert E., Zsuzsa Márka, Jingkai Yan, Imre Bartos, John N. Wright, and Szabolcs Márka. "Detecting and Diagnosing Terrestrial Gravitational-wave Mimics Through Feature Learning." arXiv:2203.05086 2022


Frequency [Hz]
Normalized energy
Time [seconds] from 2019-12-01 17:36:44 UTC (1259257022.0)
0%
25%
50%
75%
100%





Interpretation



Also see Soni, S., et al. "Reducing scattered light in LIGO's third observing run." *Classical and Quantum Gravity* 38.2 (2020): 025016.
Image based on LIGO document G1100866 (2011).

PHYSICAL REVIEW D **105**, 043006 (2022)

Generalized approach to matched filtering using neural networks

Jingkai Yan,[1,4] Mariam Avagyan,[1,4] Robert E. Colgan,[2,4] Doğa Veske,[3] Imre Bartos,[6] John Wright,[1,4] Zsuzsa Márka,[4,5] and Szabolcs Márka[3,4]

[1]*Department of Electrical Engineering, Columbia University in the City of New York, 500 W. 120th St., New York, New York 10027, USA*
[2]*Department of Computer Science, Columbia University in the City of New York, 500 W. 120th St., New York, New York 10027, USA*
[3]*Department of Physics, Columbia University in the City of New York, 538 W. 120th St., New York, New York 10027, USA*
[4]*Data Science Institute, Columbia University in the City of New York, 550 W. 120th St., New York, New York 10027, USA*
[5]*Columbia Astrophysics Laboratory, Columbia University in the City of New York, 538 W. 120th St., New York, New York 10027, USA*
[6]*Department of Physics, University of Florida, PO Box 118440, Gainesville, Florida 32611-8440, USA*

Template –based searches are limited by

LIGO noise
- non-Gaussian
- time-varying noise distribution

Density and coverage issues in the template bank

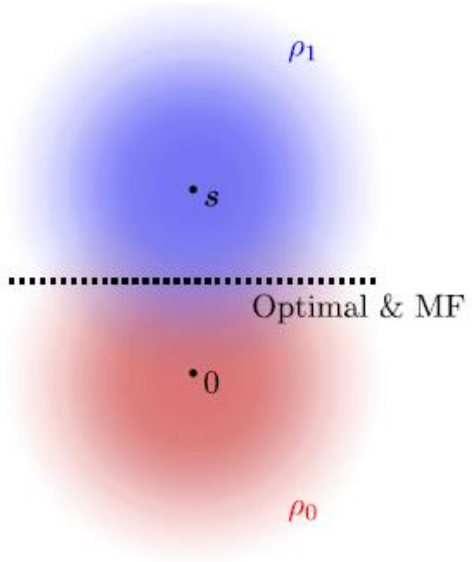
The Detection Problem and Matched Filtering

For a **single target signal**

$$H_0: \quad \boldsymbol{x} = \boldsymbol{z}$$
$$H_1: \quad \boldsymbol{x} = \boldsymbol{s} + \boldsymbol{z}$$

MF decision rule

$$\delta(\boldsymbol{x}) = 1 \quad \text{iff } \langle \boldsymbol{x}, \boldsymbol{s} \rangle > \tau$$



The **parametric detection problem**

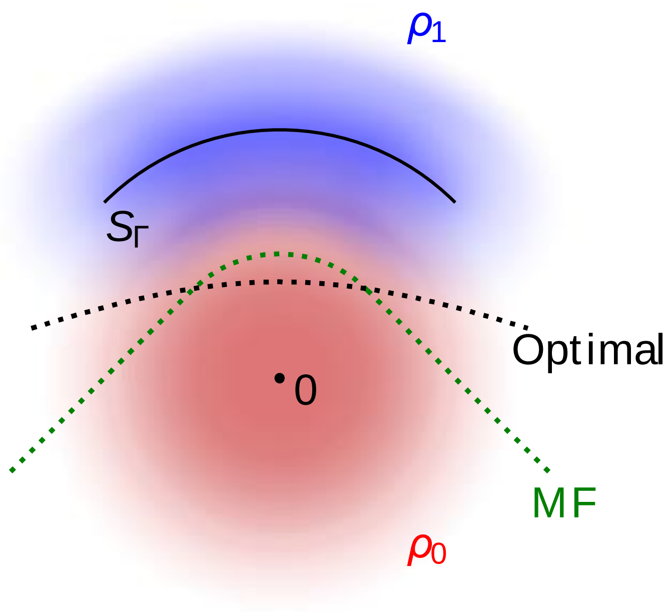
$$H_0: \quad \boldsymbol{x} = \boldsymbol{z}$$
$$H_1: \quad \boldsymbol{x} = \boldsymbol{s}_\gamma + \boldsymbol{z} \quad \text{for some } \boldsymbol{s}_\gamma \in S_\Gamma$$

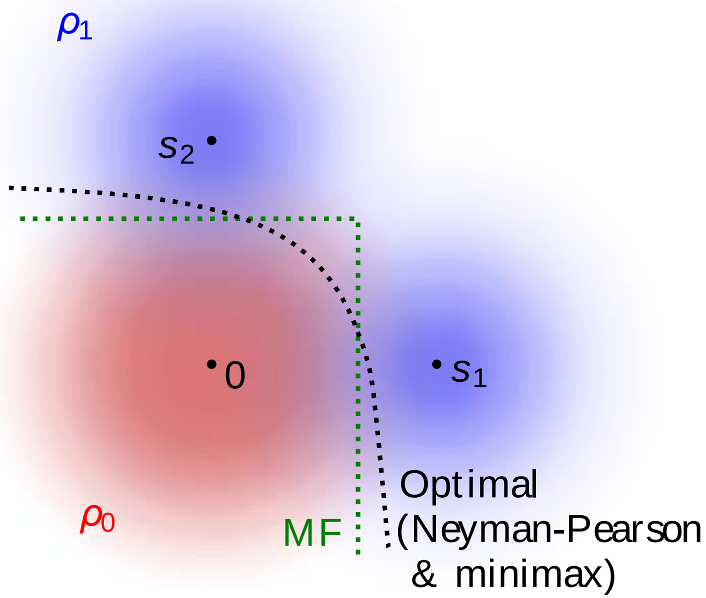
MF decision rule

$$\delta(\boldsymbol{x}) = 1 \quad \text{iff} \quad \max_{\gamma \in \Gamma} \langle \boldsymbol{x}, \boldsymbol{s}_\gamma \rangle > \tau$$

MF searches in GW detection use increasingly denser sampling (millions of waveforms)

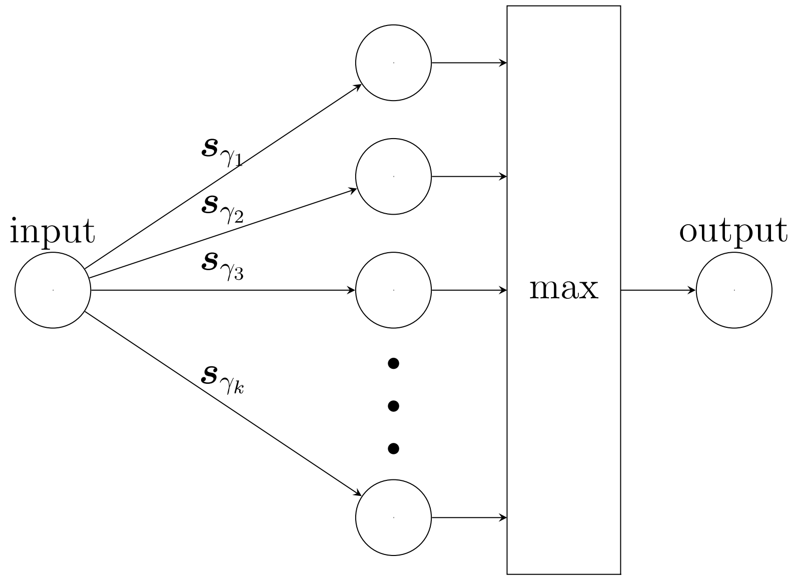
$\boldsymbol{s}_\gamma$ - astrophysical signal
$\boldsymbol{z}$ - noise (with distribution ρ_0)
$\boldsymbol{s}_\gamma \in S_\Gamma$ - signal belongs to a parametric family of signals
γ - masses, orbits, and spins, etc.

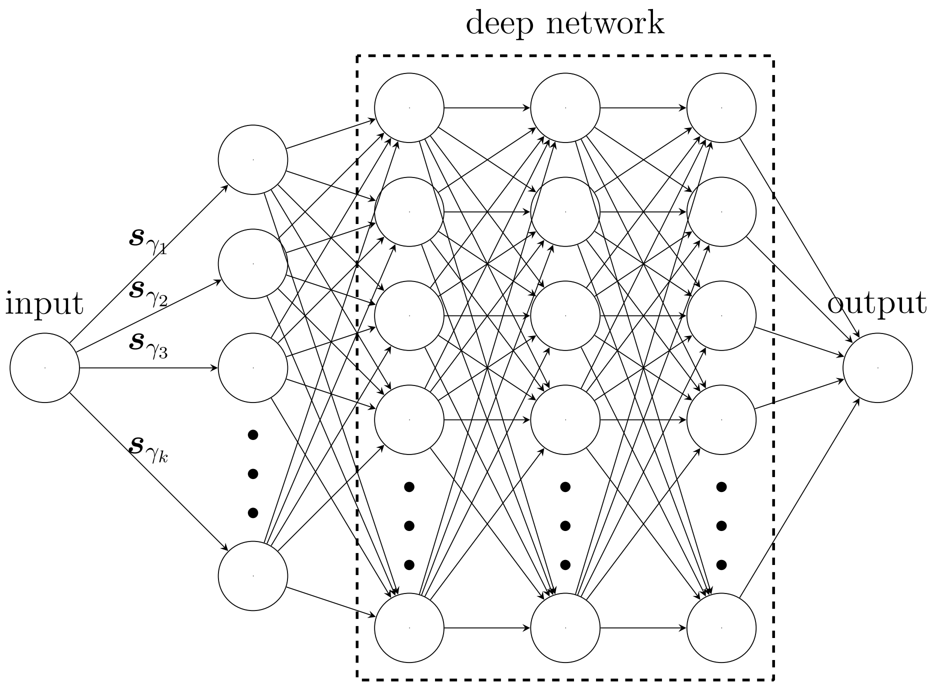




MF Is a Particular NN

- MF with a given set of templates can be constructed analytically as an equivalent NN.
- *MNet-Shallow*
  - Exact replication of MF.
- *MNet-Deep*
  - Replaces the “max” operation with a specially designed deep ReLU network.
  - Advantages: more flexible and can handle a wider range of distributions.



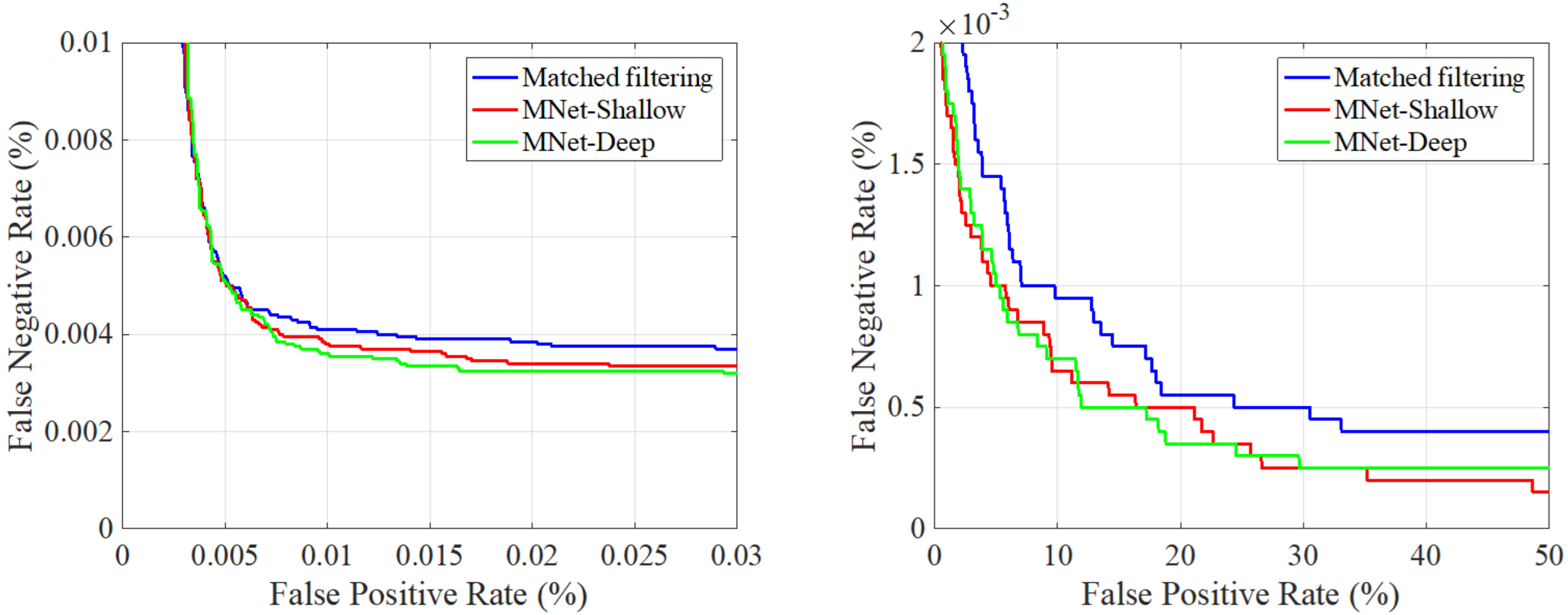


NN Can be Further Improved by Training

- We have constructed NNs that are initialized to be equivalent to MF.
- They can be further improved by training on data.
- Neyman-Pearson scenario (prior given)
  - With certain loss functions, the NN training process is aimed at learning the statistically optimal decision rule.

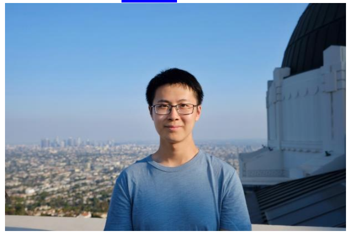
Experimental Results

- LIGO GWOSC O2 public data, 8/1/2017 — 8/25/2017.
- Synthetic waveforms
  - IMRPhenomD, mass 40~50 $M_{\odot}$, no spin, plus polarization.
  - SNR=9. Two panels below show the same curves with different axis ranges.



Hierarchical Detection (Neural)Networks and Complexity

Yan et al PRD 106.063008

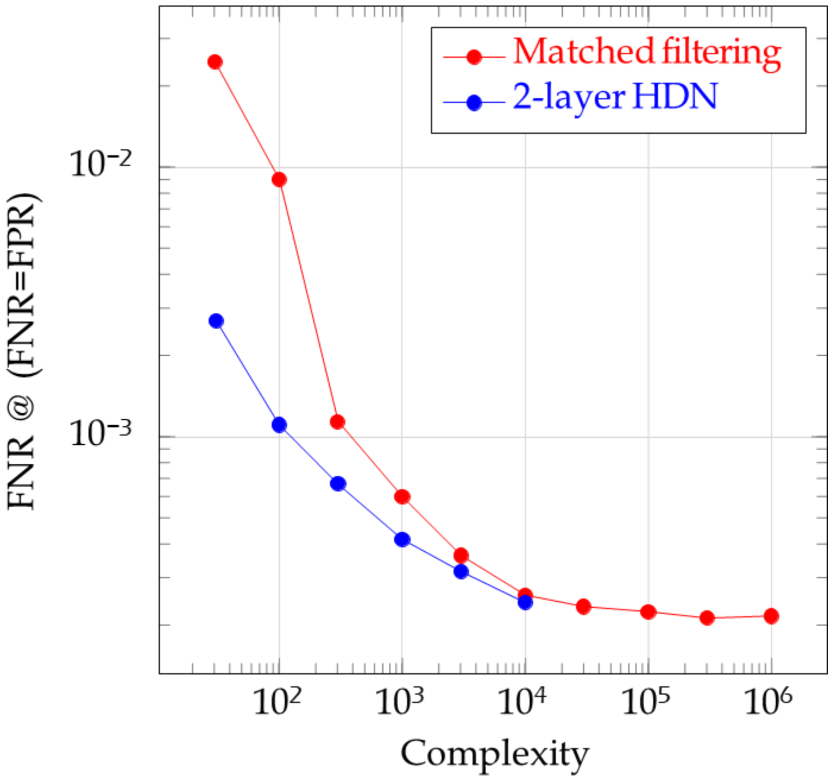


Demonstrated hierarchical detection networks that improve accuracy of search, while significantly reducing search complexity, resulting in efficient detection

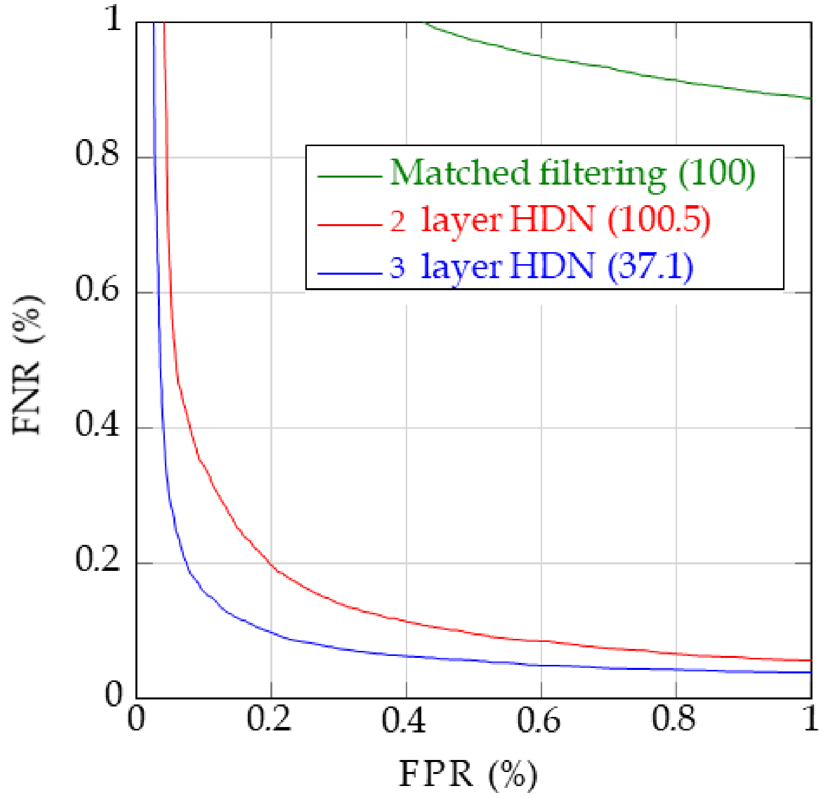
„Boosting the efficiency of parametric detection with hierarchical neural networks"

LIGO O2 public data, synthetic injections, SNR=9. Compare error rates at different complexities.

Experiments: Two-Layer



Experiments: Multi-Layer



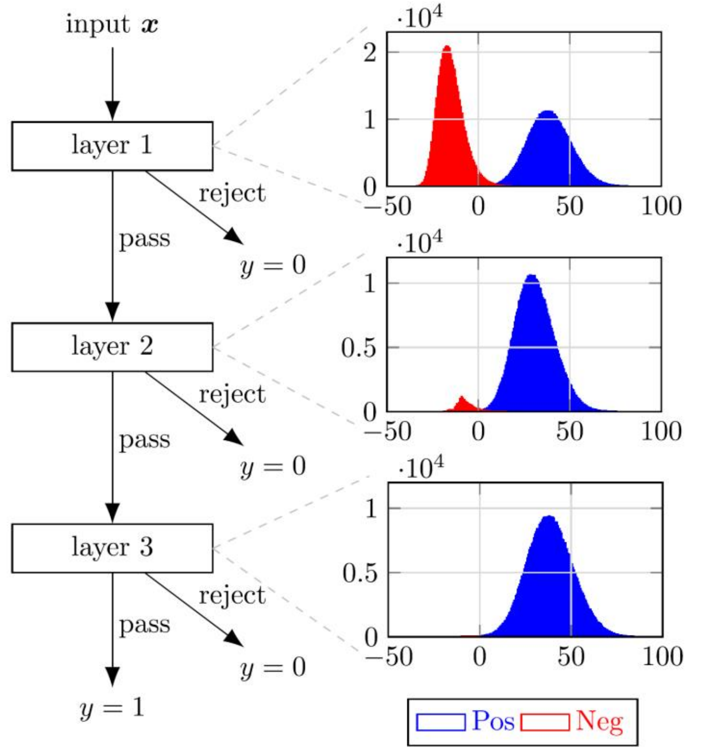
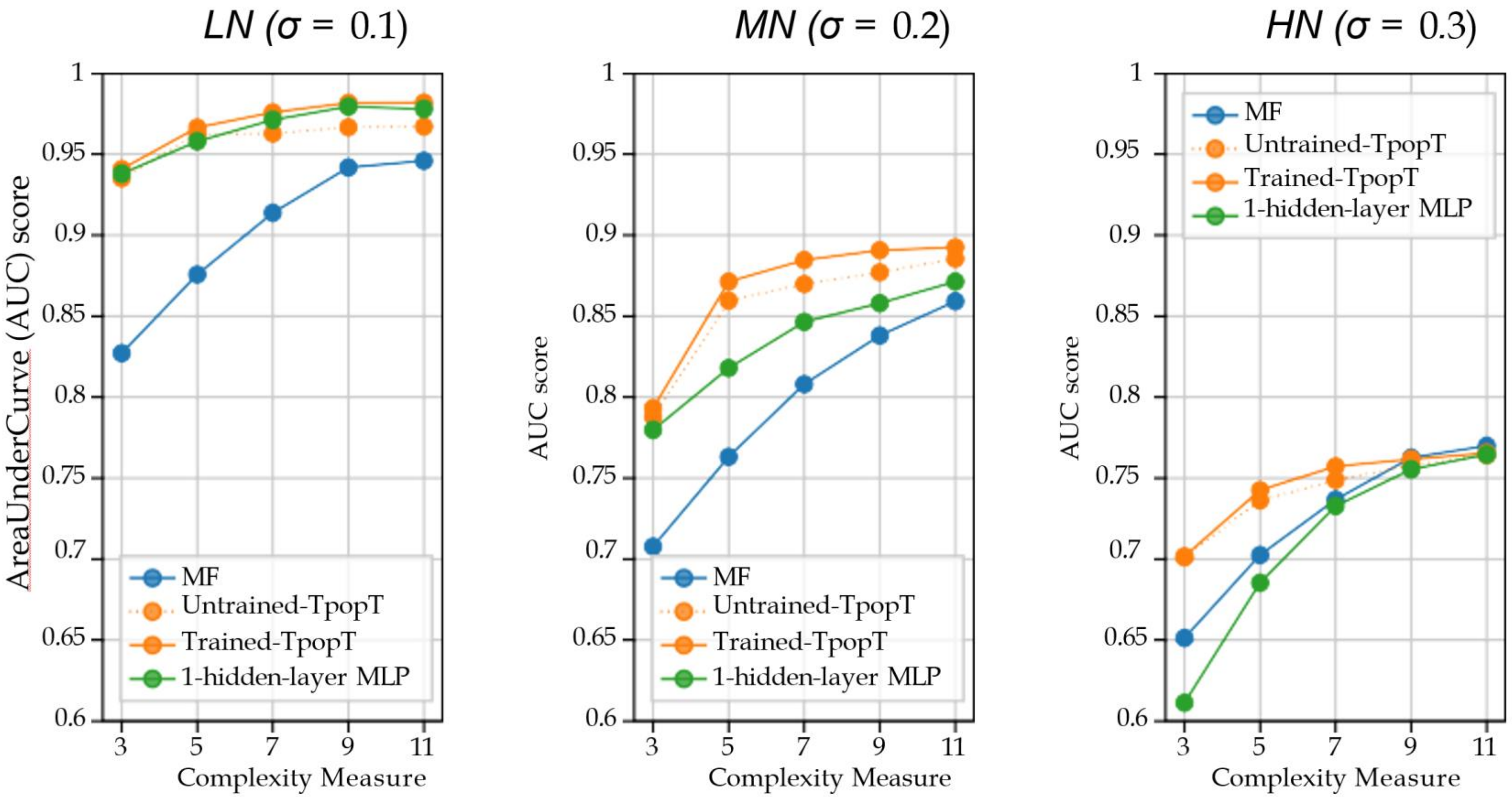


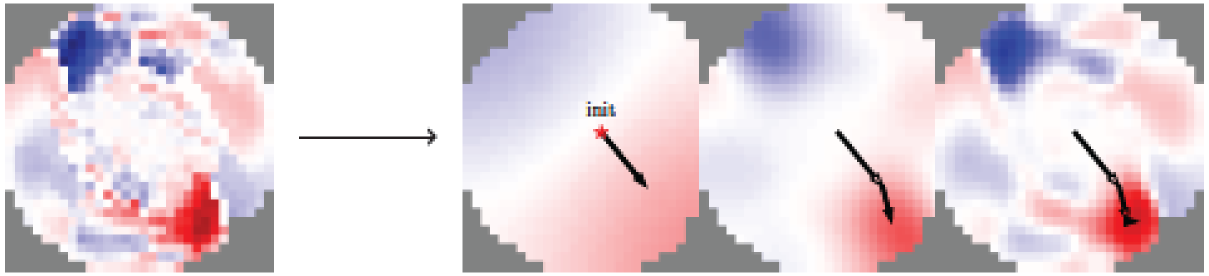
FIG. 6. Illustration of the three-layer architecture and the output densities on the test data from each layer. Only data entries that reach a given layer are shown. We see that each layer successfully rejects the vast majority of incoming negative data, and barely any negative data reach the last layer.

TpopT (TemPlate OPTimization)

- Leverage the geometric properties of the signal in order avoid the majority of unnecessary templates.
- Realization: Riemannian gradient descent for TpopT is exponentially more efficient than MF
- Treats the iterations of an optimization method as layers of a neural network – trainable
- Significantly improved complexity-accuracy tradeoffs



Nonparametric TpopT extension




Jingkai Yan, Shiyu Wang et al., arXiv:2310.10039

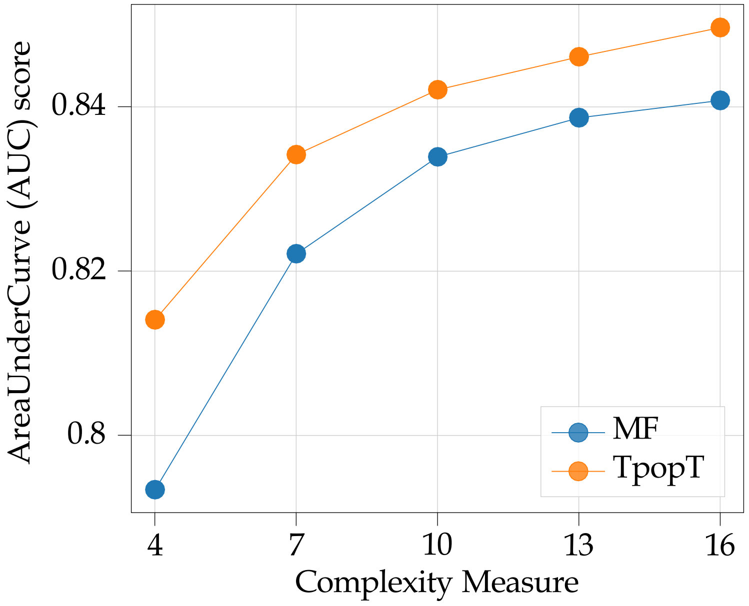
correlation landscape

optimization steps with multi-level smoothing

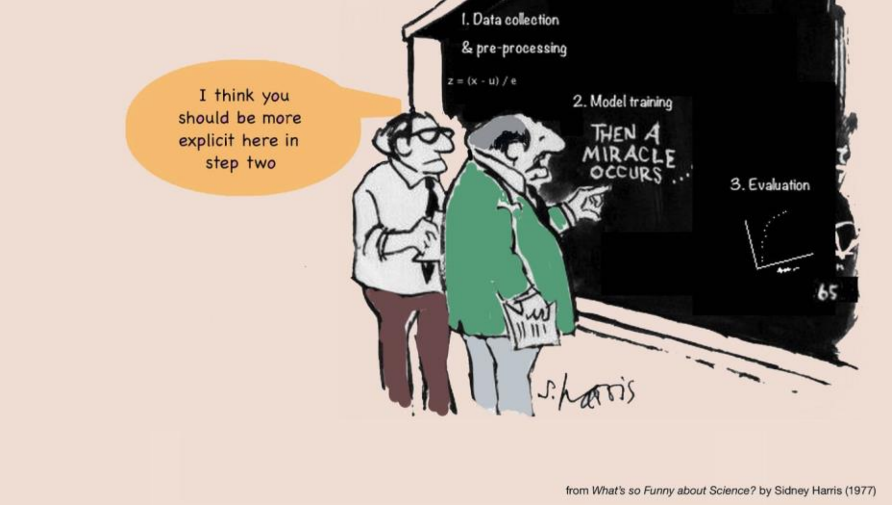
Figure 2: Illustration of 2-dim signal embeddings and the parameter optimization procedure for gravitational wave signals.

- **Task:** Detect the digit ***'3'*** from all other digits

Example: Handwritten Digit Recognition Case



Summary: INTERPRETABILITY



The LIGO detectors and their data represent highly complex engineered systems and are inspiring new ML models and methods

-- We shown applications involving high-dimensional data to find the experimental basis of noise artifacts

-- Gravitational wave data has inspired theoretical analyses of deep learning

neural network and MF equivalence

complexity and accuracy tradeoff

Applications beyond GW science

Hyperparameter optimization

- Validation dataset to choose best α, λ
- Grid search
- Training set: 7,500 glitches, 7,500 glitch-free
- Validation set: 2,500 glitches, 2,500 glitch-free