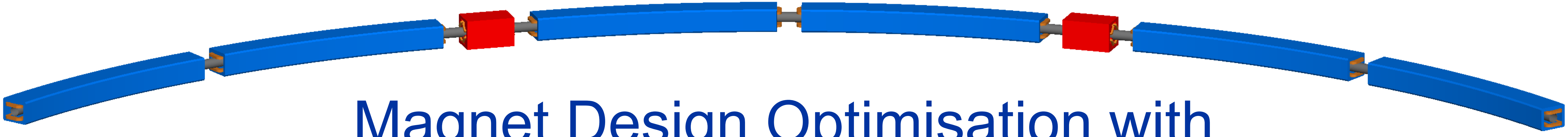


Magnet Design Optimisation with Supervised Deep Neural Networks

F. Stummer[1,2], E. Andersen[1], D. Banerjee[1], A. Baratto Roldan[1], J. Bernhard[1], S. T. Boogert[3], M. Brugger[1], N. Charitonidis[1], M. Deniaud[2], L. A. Dyks[1], L. Gatignon[1,4], S. Gibson[2], A. Goillot[1], M. Jebramcik[1], A. Keyken[2], F. Metzger[1], R. Murphy[1,2], L. J. Nevay[1], E. Parozzi[1], B. Rae[1], S. Schuh-Erhard[1], W. Shields[2], L. Suette[1], M. Van Dijk[1], A. Visive[1,5], T. Zickler[1]

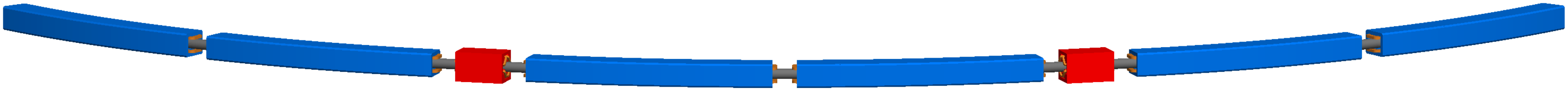
[1] CERN, 1211, Meyrin, Switzerland

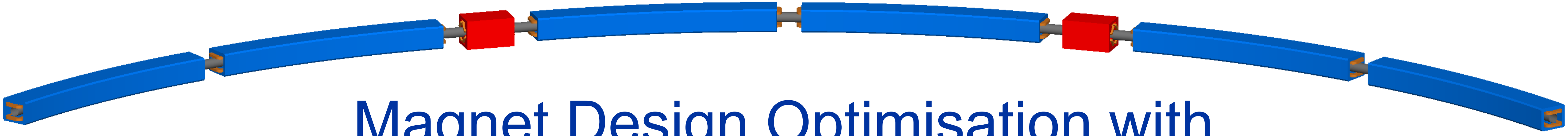
[2] JAI at Royal Holloway University of London, TW20 0EX, Egham, United Kingdom

[3] UMIST, Manchester, M6O 1QD, England, United Kingdom

[4] Lancaster University, LA1 4YW, Lancaster, United Kingdom

[5] KTH Royal Institute of Technology, 114 28, Stockholm, Sweden



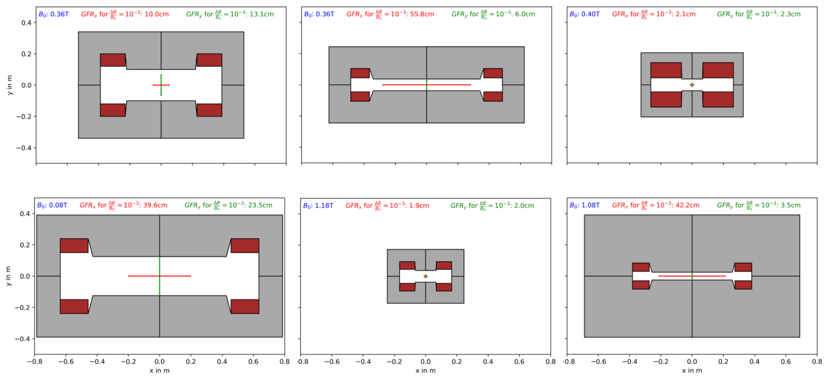


Magnet Design Optimisation with Supervised Deep Neural Networks

Magnet Dataset → train → Deep Neural Network → plug into → Optimiser

Magnet Dataset

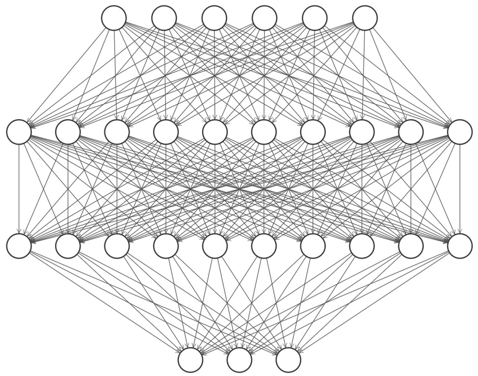
60.000 samples created and simulated



Extract Specifications that they meet from Finite Element Method Simulation

Deep Neural Network

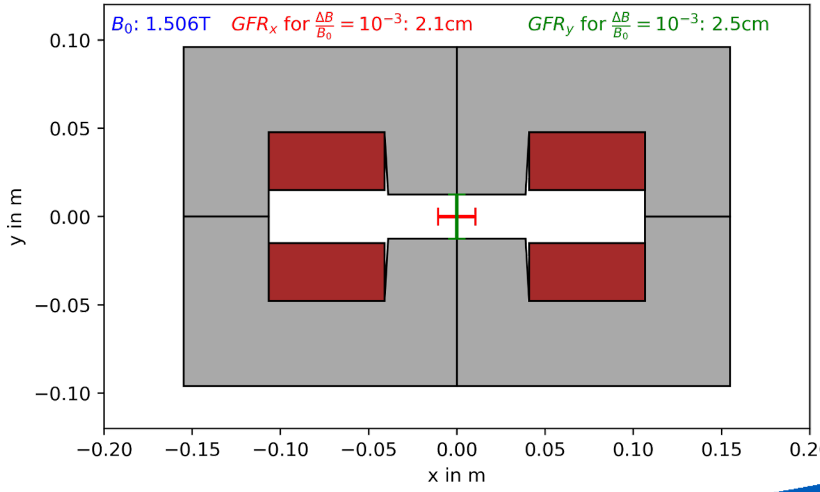
Design Parameters

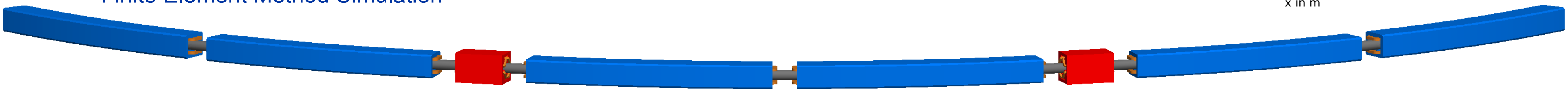


Specifications

Optimiser

Design magnet with Specificationss:
B_0=1.5 T, $GFR_x > 2$ cm, $GFR_y > 1$ cm for $\frac{\Delta B}{B_0} = 10^{-3}$





EUROPEAN AI FOR FUNDAMENTAL PHYSICS CONFERENCE EuCAIFCon 2024

Learning new physics with a kernel machine

Marco Letizia

Machine Learning Genoa Center and INFN

In collaboration with: P. Cappelli (UniPd), G. Grosso (IAIFI-MIT), N. Lai (UniPd), M. Pierini (CERN), L. Rosasco (UniGe-MaLGa), A. Wulzer (IFAE), M. Zanetti (UniPd).

Learning new physics with a kernel machine

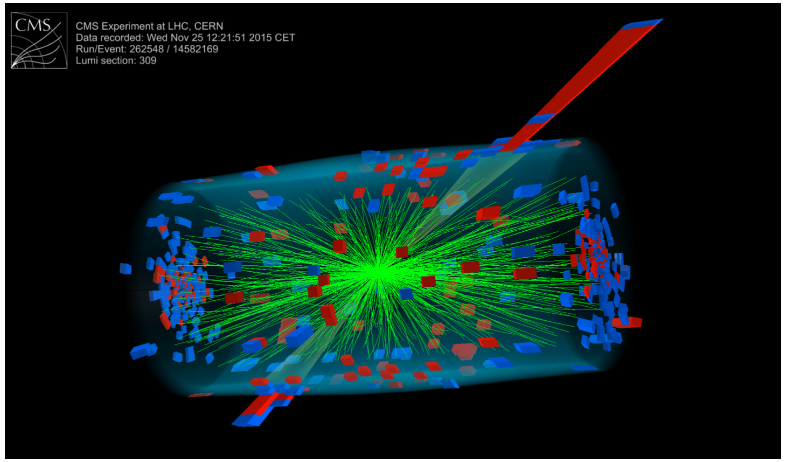
GOAL: search for rare/hidden new physics in high energy physics data.

PROBLEM: most analyses are model-dependent

→ heavily biased towards specific theoretical models.

Agnostic searches are hard to design:

large volumes of mutivariate, complex data.



To maximise the discovery potential at the LHC (and future experiments!),

it is crucial to develop hypothesis testing methodologies based on new paradigms!

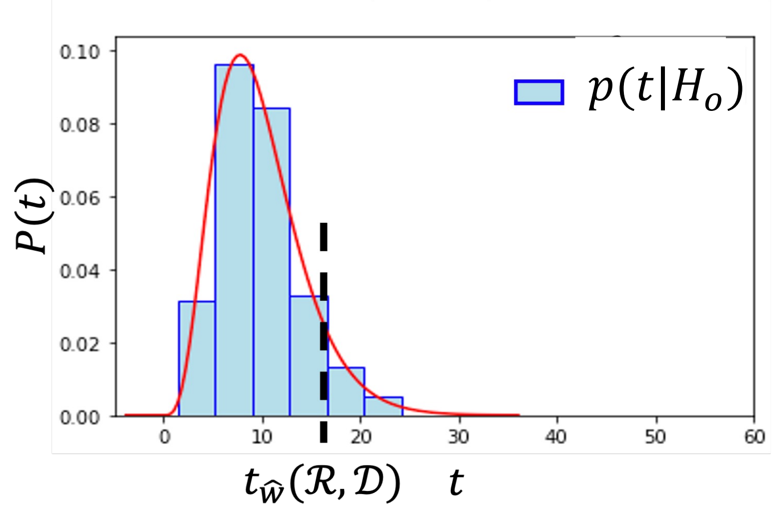
Learning new physics with a kernel machine

The New Physics Learning Machine

A likelihood-ratio test with a data-driven alternative hypothesis

$$n_w(x) = e^{f_w(x)} n(x|R) \qquad \rightarrow t_w(\mathcal{D}) = -2 \log \prod_{x \in D} \frac{\mathcal{L}_w(x)}{\mathcal{L}(x;R)}$$

- Unbinned
- Multivariate
- Signal-agnostic
- Efficient and robust machine learning
- Statistically sound
- Distribution and normalization shifts
- No data splitting



UniGe | MaLGa

Learning new physics with a kernel machine

Marco Letizia – Machine Learning Genoa Center and INFN
In collaboration with: P. Cappelli (UniPd), G. Grosso (IAIFI-MIT), N. Lai (UniPd), M. Pierini (CERN), L. Rosasco (UniGe-MaLGa), A. Wulzer (IFAE), M. Zanetti (UniPd).

UniGe | MaLGa INFN erc

EuCAIFCon 2024

Foundations

The ***New Physics Learning Machine***[1] is a methodology powered by machine learning to perform a likelihood-ratio two-sample test that is unbinned, multivariate, scalable, signal agnostic, sensitive to distribution shifts as well as normalization effects, and without data splitting. The goal is to compare an observed set of data with the prediction of a reference model. The NPLM model is trained as a supervised classifier on a *reference sample*

$$\mathcal{R} = \{x_i\}_{i=1}^{N_\mathcal{R}}, \quad x_i \sim p(x|R),$$

and a *data sample*

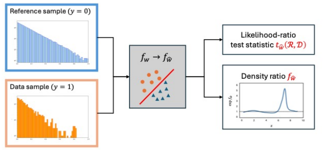
$$\mathcal{D} = \{x_i\}_{i=1}^{N_\mathcal{D}}, \quad x_i \sim p_{\text{true}}(x).$$

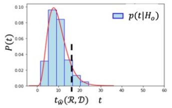
It learns the density ratio

$$f^*(x) = \log \frac{n_{\text{true}}(x)}{n(x|R)},$$

where $n(x|\cdot) = N(\cdot)\,p(x|\cdot)$ is the probability density normalized to the expected number of events, and outputs the extended likelihood-ratio test statistic

$$t_{\hat{w}}(\mathcal{R},\mathcal{D}) = -2\log\left[\sum_{x\in\mathcal{R}} \frac{N(R)}{N_\mathcal{R}}\left(e^{f_{\hat{w}}(x)}-1\right) - \sum_{x\in\mathcal{D}} f_{\hat{w}}(x)\right].$$





After estimating the distribution of the test statistic under the null hypothesis (e.g. with reference-distributed pseudo-experiments, bootstrap or permutations), the p-value and Z score associated with the observed data can be computed

$$p_{\text{value}} = \int_{t_{\hat{w}}(\mathcal{R},\mathcal{D})}^{\infty} dt\, p(t|H_0), \qquad Z = \Phi^{-1}(1-p_{\text{value}})$$

We focus here on the implementation based on kernel methods and the Falkon library, highly performant while extremely efficient.[2,3] It is based on a weighted logistic loss

$$\ell(y, f(x)) = \frac{N(R)}{N_\mathcal{R}}(1-y)\log(1+e^{f(x)}) + y\log(1+e^{-f(x)}),$$

with regularization term $\lambda\|f\|_\mathcal{F}^2$, and considers functions of the following form

$$f_w(x) = \sum_{i=1}^{M} w_i k(x,\tilde{x}_i), \qquad k(x,x') = \exp -\frac{\|x-x'\|^2}{2\sigma^2},$$

where $\{\tilde{x}_1,\dots,\tilde{x}_M\}$ is a set of points selected uniformly at random from the training set, known as *centers*.

Model-selection. The three main hyperparameters are tuned using only reference data to avoid any bias towards the data of interest. The criteria for their selection are[1,2]:

• The Gaussian width σ is selected as the 90th percentile of the pairwise distance among reference-distributed data points.

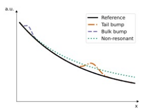
• To achieve optimal statistical bounds and preserve performance, the number of centres M must be at least be of order $\sqrt{N}$, with N the size of the dataset.[3]

• The L2 regularisation parameter λ is kept as small as possible while maintaining a stable training.

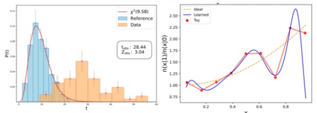
Model-independent searches

Experimental measurements are compared with a reference sample from the Standard Model without relying on any specific signal hypothesis.[2]

1D benchmark



Distribution of the test statistic and signal reconstruction for the non-resonant signal.



· $N_\mathcal{R} = 200k,\ N(R) = 2000$

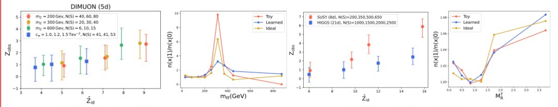
· $N(S) = 10$ (tail), 90 (bulk/non-res)

· $\bar{t}_{training} \approx 2$ sec

Median Z	Tail	Non-res	Bulk
Z_{id}	4.7	4.4	4.1
Z_{obs}	2.4	3.0	2.8

Multivariate benchmarks

DATASET: DIMUON (5D) , SUSY (9D), HIGGS (21D)



· $N_\mathcal{R} = 100k, N(R) = 20k$ · $N_\mathcal{R} = 500k, N(R) = 100k$

Average training times per single run with standard deviations (low level features and reference toys). Note that time measured in hours (for NN) and seconds (for Falkon).

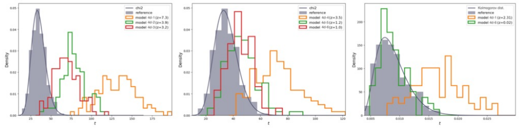
Model	DIMUON	SUSY	HIGGS
FLK	**(44.9 ± 3.4) s**	**(18.2 ±1.2) s**	**(22.7 ± 0.4) s**
NN	(4.23 ± 0.73) h	(73.1 ± 10) h	(112 ± 9) h

Bold values indicate the lowest for each column (lower is better)

Evaluation of generative models

The efficiency of the kernel-based model opens the door to several applications. We show here a first test on the evaluation of generative models.

- Data — correlated mixtures of three Gaussians in four dimensions.
- Model — normalizing flow: RealNVP.
- Architecture — models1,2 and 3: 3x64 hidden layers; models 2, 4 and 6: 3x128 hidden layers.
- Training — models 1 and 2 are trained with 100k samples, models 3 and 4 with 200k and models 5 and 6 with 500k.
- NPLM test — size of reference sample: 100k; size of data samples: 10k; average training time ≈ 1 sec.

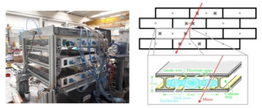


A good correlation between the average Z scores with the number of training examples and the model complexity can be observed.
On the right-side plot we show the results of a dimension-averaged KS test[4] on the best and worst models.

Data Quality Monitoring

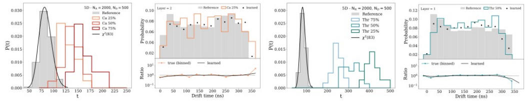
NPLM for monitoring particle detectors in real-time.[5]

- Reduced scale CSM drift tubes. • Data: 4 drift times, crossing angle.
- Anomalies: lowered cathodic strip voltages and front-end thresholds.
- $\bar{t}_{training} \approx 0.5$ sec



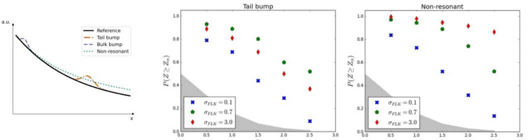
Median *p*-values.

Anomaly	NPLM (5D)	KS (t_1)	KS (t_2)	KS (t_3)	KS (t_4)	KS (ϕ)
Cathode 75%	1.1×10^{-6}	0.50	0.41	0.43	0.40	0.42
Cathode 50%	3.4×10^{-4}	0.47	0.27	0.47	0.37	0.41
Cathode 25%	0.0019	0.45	0.44	0.13	0.45	0.50
Threshold 75%	$<10^{-7}$	0.23	0.14	0.16	0.14	0.48
Threshold 50%	$<10^{-7}$	0.09	0.10	0.06	0.17	0.42
Threshold 25%	$<10^{-7}$	0.11	0.07	0.04	0.11	0.66



Multiple testing

Model selection can bias the test towards certain signal hypotheses.



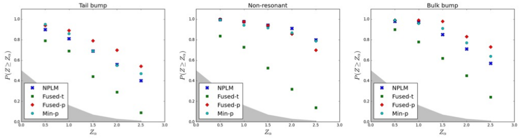
Multiple testing strategies can be leveraged to tame this effect and obtain a more uniform response. The following approaches are considered:

- Fused-t: $t_{\text{fused}} = \log n^{-1}\sum_{i=1}^{n}\exp t_i$
- Fused-p: $p_{\text{fused}} = -\log n^{-1}\sum_{i=1}^{n}\exp(-p_i)$
- Min-p: $p_{\text{min}} = -\log \min_{1\le i\le n} p_i$

In the first case, the new test statistic is obtained by combining the local variables $\{t_i\}$ with a smooth maximum while in the other two, the local p-values are used (smooth minimum and minimum).

The different tests are characterized by the following kernel widths:

$$\sigma \in [0.1, 0.3, 0.7, 1.4, 2.4, 3.0].$$



The *fused-p* NPLM combined test shows an advantaged when compared to the other methods, including the standard NPLM ($\sigma = 2.4$). These results suggest that higher sensitivity can be achieved with this strategy while reducing the dependence on hyper-parameters.

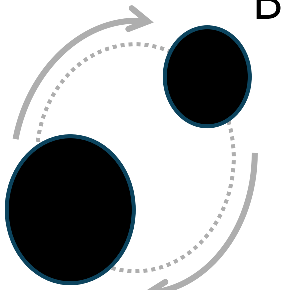
[1] Grosso et al, *Goodness of fit by Neyman–Pearson testing*, SciPost Physics 2024, 2305.14137; [2] Letizia et al, *Learning new physics efficiently with nonparametric methods*, EPJC 2022, 2204.02317; [3] Meanti et al, *Kernel methods through the roof: handling billions of points efficiently*, NeurIPS 2020, 2006.10350;
[4] Coccaro et al, *Comparative Study of Coupling and Autoregressive Flows through Robust Statistical Tests*, 2302.12024; [5] Grosso et al, *Fast kernel methods for data quality monitoring as a goodness-of-fit test*, MLST 2023, 2303.05413.

Poster Session A - Wednesday 12:00 - 15:00

UniGe | MaLGa

BBH gravitational waves and TianQin

Binary

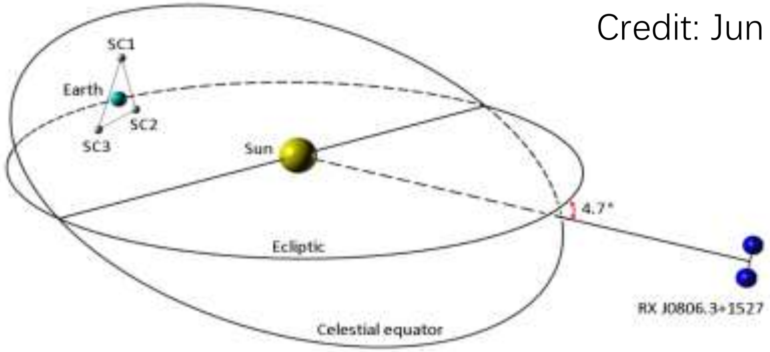


Signal: $s(\theta)$

GW data analysis

Signal Detection
Parameter Estimation

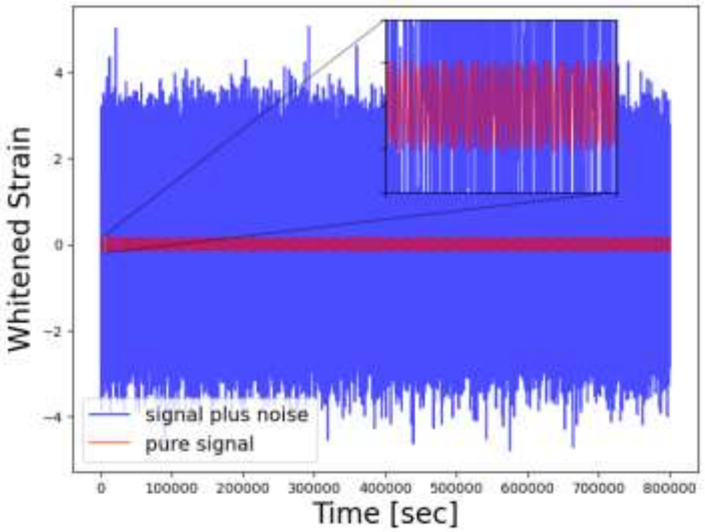
Credit: Jun Luo, 2015

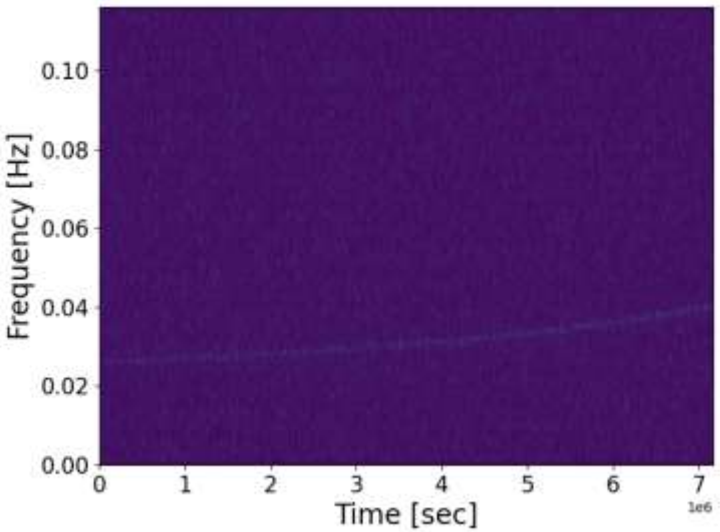


Observable: $d = s(\theta) + n$

Signal-to-noise ratio 200

$m_1 = 85\ M_{\odot}$,
$m_2 = 66\ M_{\odot}$.





$$f = \left(\frac{5}{256}\right)^{\frac{3}{8}} M_c^{-\frac{5}{8}} (t_{\rm ref})^{-3/8}$$

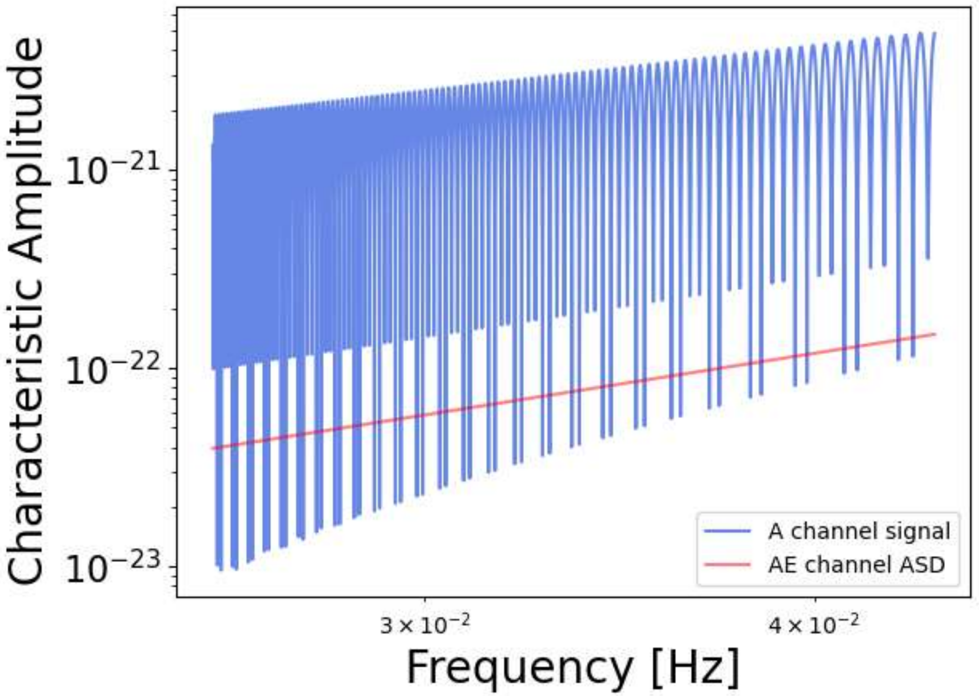
There are lots of gravitational wave cycles in band.
The use of the match-filtering method is computationally prohibitive due to the immense template bank(over 10^{31})

Challenges

- Signal morphology:

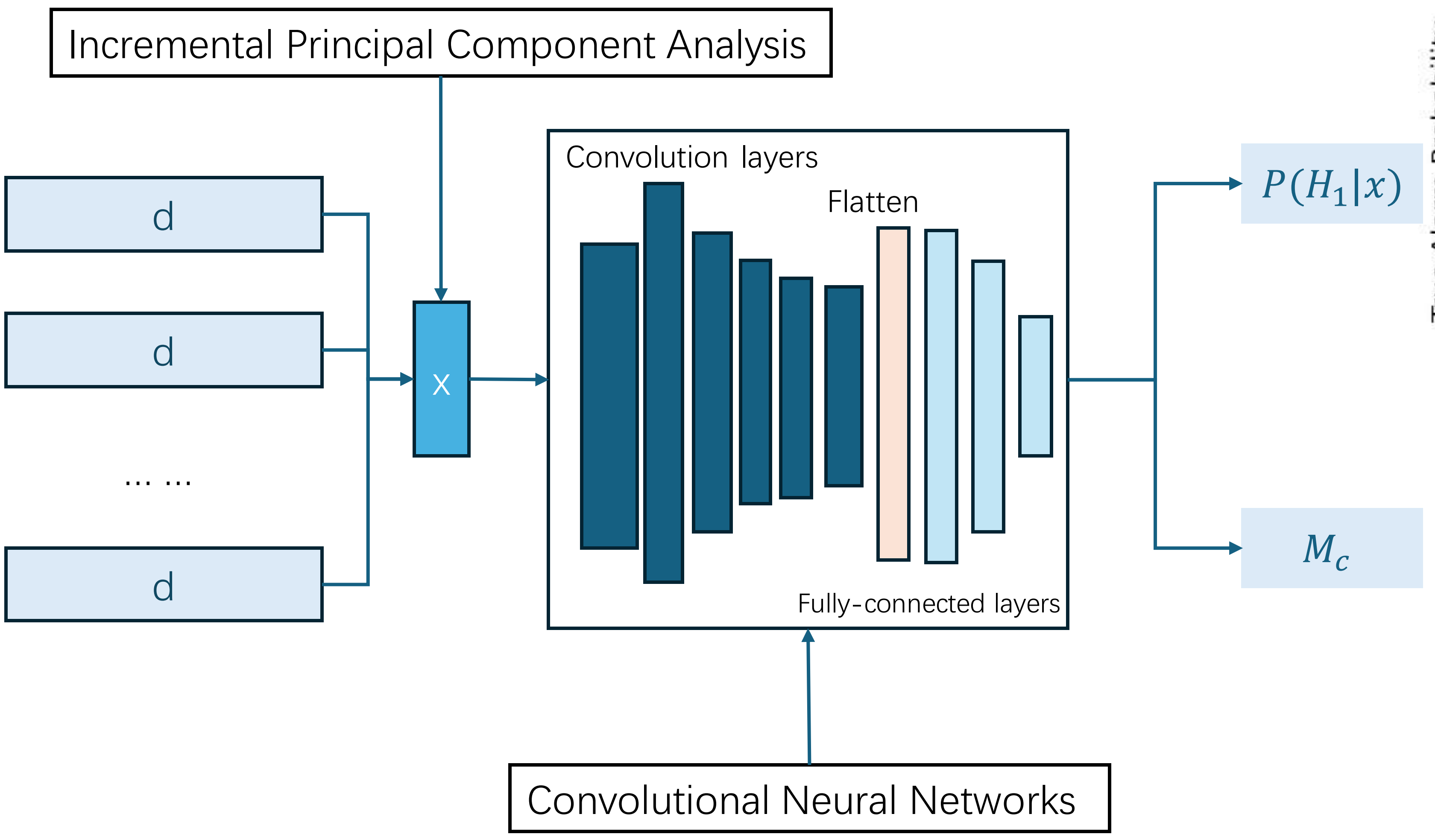
1. Many data point: 10^6 with a duration spanning three months.
2. Intense oscillations observed due to the detector's modulation.

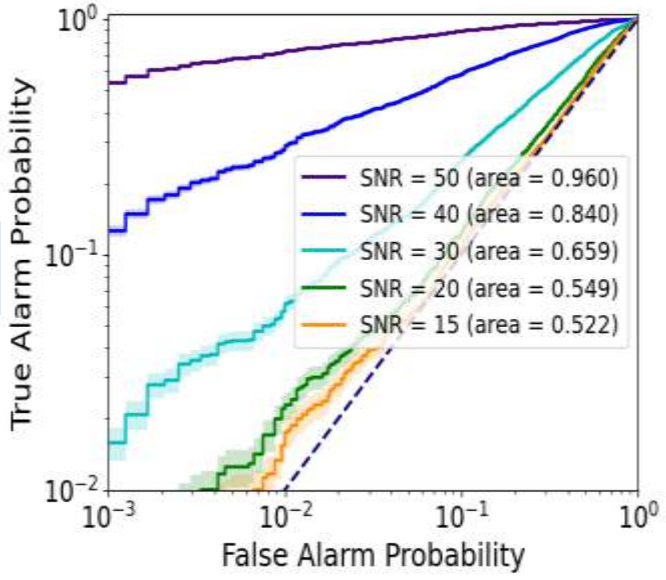
... ...

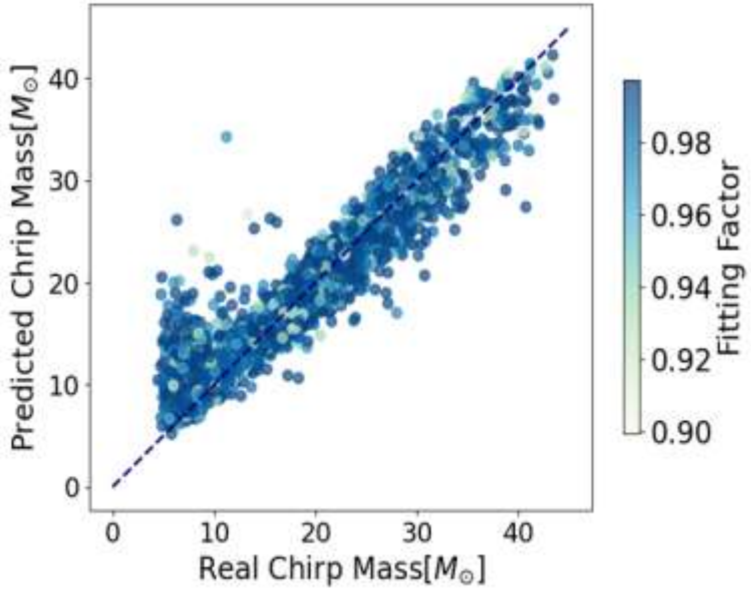


The signal from a GW190521-like binary black hole system. The observation period spans three months, with a sample rate of 0.25 Hz. The number of data point is 972001.

Compression and Search







Limited by the GPU memory

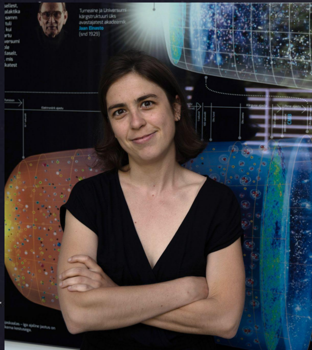
Searching for Dark Matter Subhalos in Astronomical Data using Deep Learning

Speaker: **Sven Põder**

TAL TECH

UNIVERSITY OF TARTU
Tartu Observatory
1632

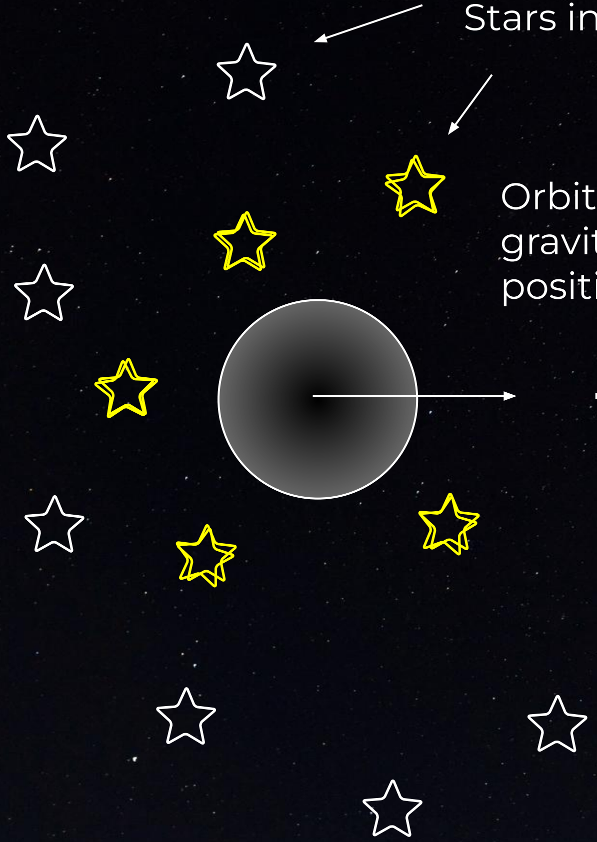
Also go check out poster of
María Benito Castaño



EuCAIFCon 2024

The signal we are looking for

Stars in the Galaxy



Orbiting subhalo imprints a gravitational signature in the position and velocity of stars

Can we detect these disturbances from the data?

The abundance of subhalos is a **function of the DM model** -> constraining the subhalo mass function a way to study the nature of the DM particle

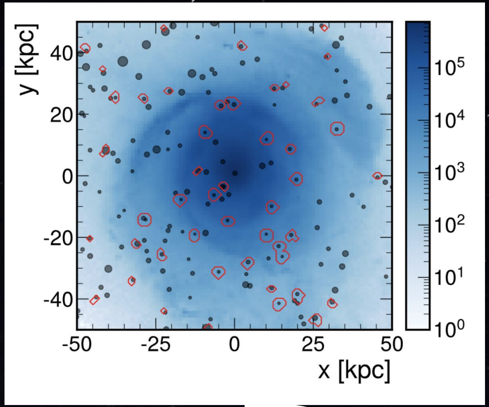
Sven Põder

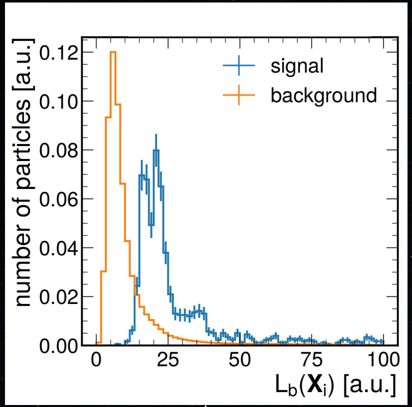
Ongoing work

On constraining the subhalo mass function..

doi:10.1016/j.ascom.2022.100667

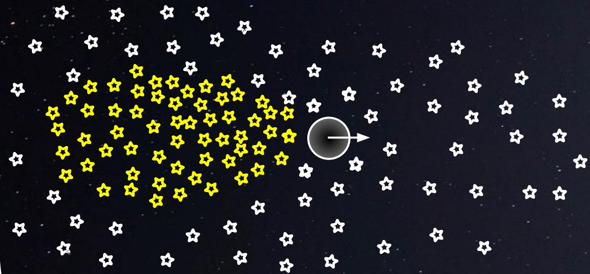
Star-by-star anomaly detection in galaxy simulations

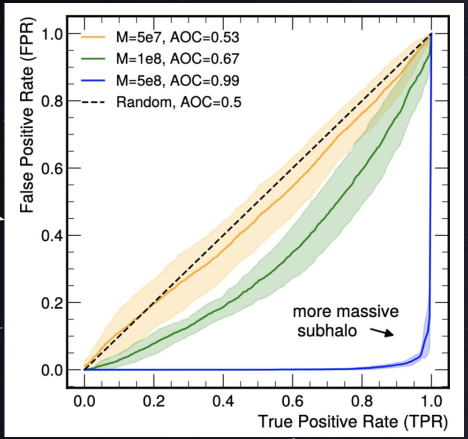




Difference between background and stars near subhalos

Studying signal detectability in N-body simulations





Sven Põder