Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

A FIRST APPLICATION OF MACHINE AND DEEP LEARNING FOR BACKGROUND REJECTION IN THE ALPS II TES DETECTOR

Manuel Meyer[1], Katharina Isleif[2,*], Friederike Januschek[2], Axel Lindner[2], Gulden Othman[1], José Rubiera Gimeno[2], Christina Schwemmbauer[2], Matthias Schott[3], Rikhav Shah[3]
for the ALPS Collaboration

[1]Institut für Experimentalphysik, Universität Hamburg, Luruper Chaussee 149, 22761, Hamburg, Germany
[2]Deutsches Elektronen-Synchrotron DESY, Notkestr. 85, 22607 Hamburg, Germany
[3]Johannes Gutenberg-Universität Mainz, Staudingerweg 7, 55128 Mainz, Germany
[*]Now at Helmut-Schmidt-Universität, Holstenhofweg 85, 22043 Hamburg, Germany

INTRODUCTION

- Axions and axion-like particles (ALPs) are hypothetical particles predicted in extensions of the Standard Model [e.g., 1]
- Axions could explain the strong CP problem in QCD [2] and axions and ALPs are candidates for cold dark matter [3]
- The ALPS II experiment [e.g., 4] aims to produce and subsequently detect these particles using the "light-shining-through-a-wall" technique (see Fig. 1, see G. Othman's overview talk on Tuesday 11.30am)
- ALPS II will use a transition edge sensor (TES), a single photon detector that can achieve high quantum efficiencies and energy resolution [5], see Fig. 2
- To significantly detect an expected signal rate of $\sim 1\,\mathrm{photon\ day}^{-1}$ in a 20 day measurement, we need to achieve extremely low background rates $\lesssim 10^{-5}$ Hz
- Here, we investigate the performance of machine learning (ML) and deep learning (DL) classification algorithms to discriminate signal and background events recorded with a test setup of the TES detector.

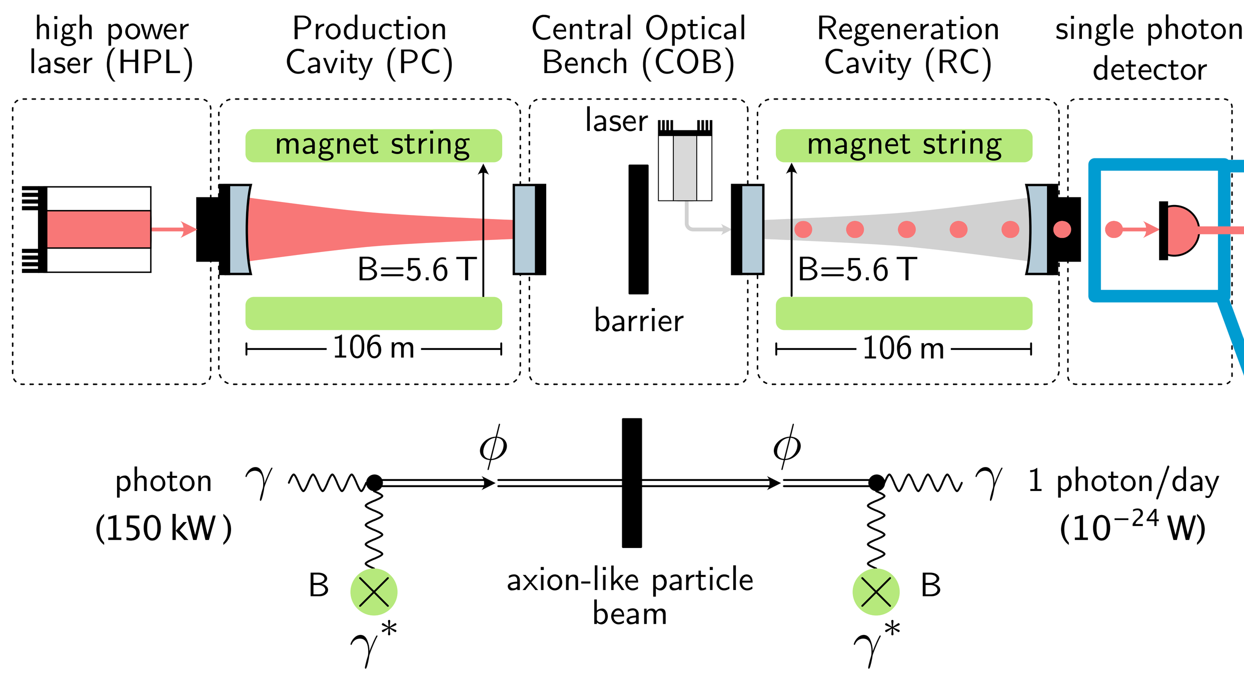

Fig. 1: schematic layout of the ALPS II detector (top) and the Feynman diagram for photon-ALP conversion (bottom). Taken from [4].

TRANSITION EDGE SENSOR FOR THE ALPS II EXPERIMENT

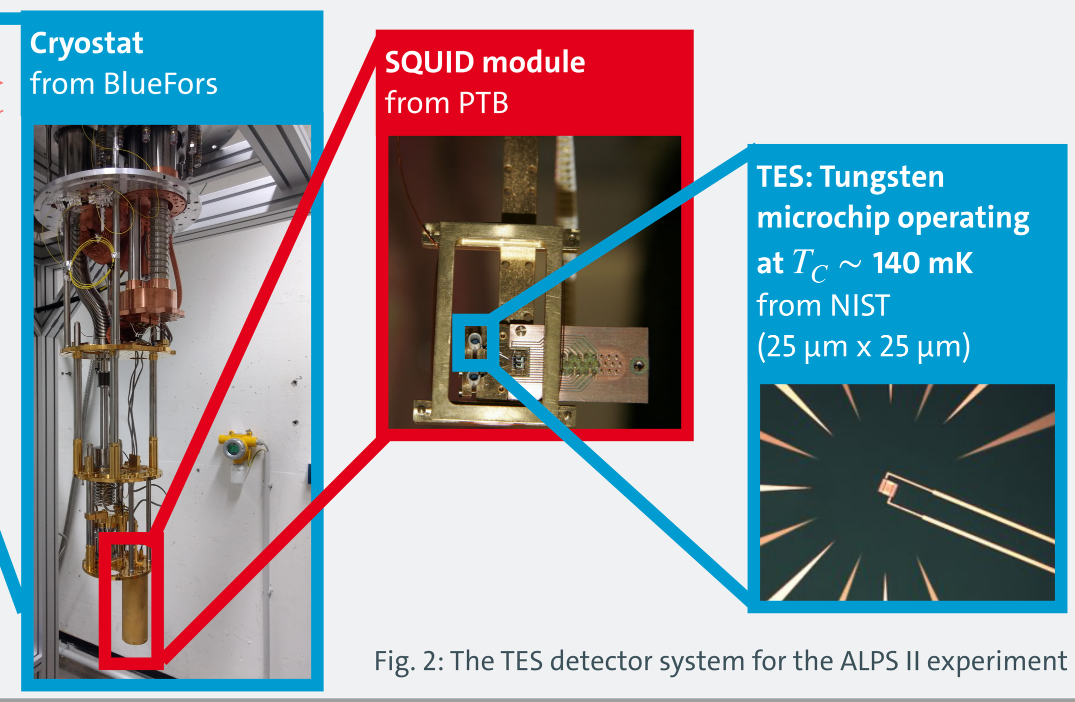

Fig. 2: The TES detector system for the ALPS II experiment

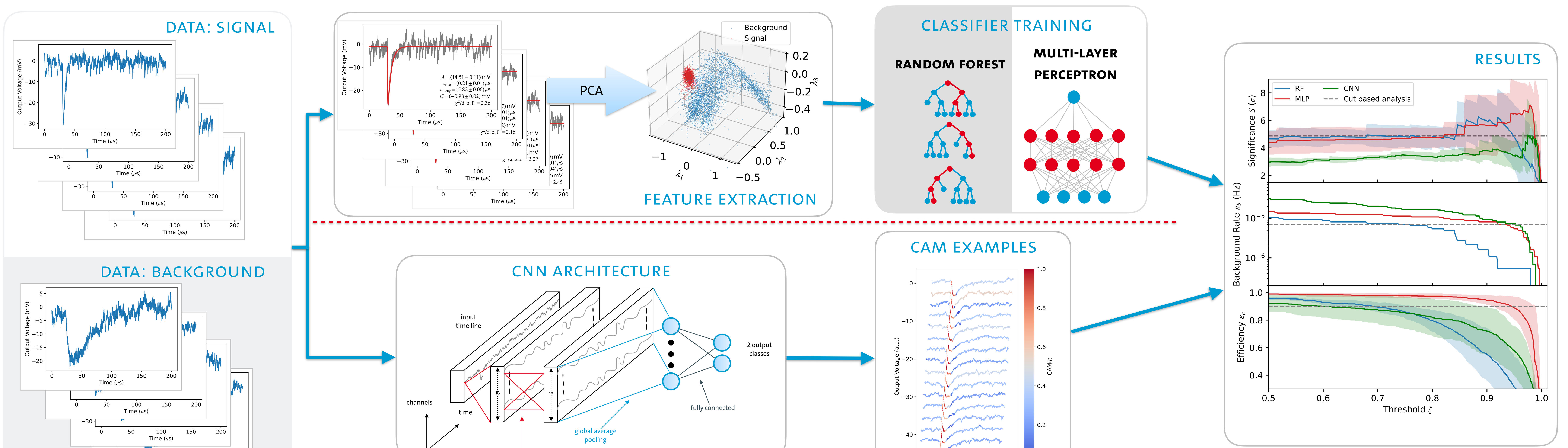

Fig 3: Schematic view of the performed analysis. In the central part the top row corresponds to ML using extracted features, whereas in the bottom row, the raw data is used. In the latter case we can generate class activation maps (CAMs) to visualize which parts of the time series are most important for correct classification.

TRAINING DATA SET

- Data collected in a test setup of the TES detector: background events were collected in $T \sim 518$ hours while no optical fiber was connected to the TES [5]
- Fake signals generated by connecting a 1064nm laser to the TES
- Each event consists of a voltage vs current time line measured with the TES and SQUID setup. Events were triggered and recorded once the amplitude reached < -20 mV. Each trigger window is 200μs long (see Fig. 3, left)
- After initial data cleaning: 39,580 background and 1,066 signal events

TRAINING OF CLASSIFIERS ON EXTRACTED FEATURES

- Each time line fit with an exponential rise and decay function [5]
- Best-fit parameters (amplitude, rise and decay times, constant offset), pulse integral and χ^2 value of the fit are recorded in a feature vector X (see Fig. 3 upper central panels)
- X is transformed using principal component analysis (PCA) and split into 80% training and 20% test data
- Two ML classifiers are tested: random forests (RF) and a multilayer perceptron (MLP)
- Hyper-parameters of the classifiers are optimized over parameter grids using K-fold cross validation with $K=5$ on the training set using again a 80%-20% split
- Best hyper-parameters selected that optimize the detection significance S [6]:

$$S/\sqrt{T} = 2\left(\sqrt{\epsilon_d\,\epsilon_a\,n_s + n_b} - \sqrt{n_b}\right)$$

- $\epsilon_d = 0.5$ is the detector efficiency, ϵ_a is the analysis efficiency to correctly classify signal evens, n_b is the background rate from mis-identified background events, and n_s is the signal rate that depends on the photon-ALP coupling
- Classifier with best hyper-parameters re-trained on entire training set

PRELIMINARY TRAINING OF CNN ON TIME SERIES DATA

- We also test the performance of convolutional neural networks (CNN) trained on the time series data itself, which eliminates the need for feature extraction
- CNNs have been found to perform very well for time-series classification [7]
- Only pre-processing step: time series data is z transformed before training and downsampled by factor 4
- CNN architecture [following 7]: 2 convolutions with kernel size 11 and 16 filters, followed each by batch normalization and ReLU activation. After convolution, global average pooling performed.
- CNN also provides Class Activation Map (CAM) which shows which parts of time series are most important for classification (see Fig. 3 and [7])

PRELIMINARY RESULTS

- Classifiers used here have advantage over cut-based analysis: each event is assigned probability of being a true light signal
- One can tune a threshold ξ for this probability above which events are classified as true light signals to achieve best significance
- Table 1 shows mean values and standard deviations for ϵ_a, n_b, S from K-fold cross validation for example values of ξ

	Threshold ξ	Signal efficiency	Background rate (µHz)	Detection significance (σ)
Cut based analysis [5]	—	0.898	6.9	4.88
RF	0.862	0.66 ± 0.15	2.16 ± 2.02	6.04 ± 1.50
MLP	0.944	0.90 ± 0.07	5.93 ± 5.23	6.51 ± 2.47
CNN	0.974	0.42 ± 0.18	< 8.54	4.94 ± 2.56

DISCUSSION & OUTLOOK

- Both feature-based classifiers and CNNs achieve a detection significance and background rate comparable or better than a simple cut based analysis used in [5]
- Heavily imbalanced data set with ratio ~40:1 of background vs light events makes training of classifiers challenging, more data taking with updated experimental setup in progress
- Larger training data set should also improve errors on performance metrics
- CNN on time lines performs worst in our test, might suffer most from imbalanced data set and electrical noise in time lines
- Results are encouraging: promising to use ML and DL algorithms for signal and background discrimination when fiber is connected to TES

REFERENCES

(1) I. Irastorza and J. Redondo, *Prog.Part.Nucl.Phys.* 102 (2018) 89-159
(2) R.D. Peccei and H. Quinn, *Phys.Rev.D* 16 (1977) 1791-1797; S. Weinberg, *Phys.Rev.Lett.* 40 (1978) 223-226; F. Wilczek *Phys.Rev.Lett.* 40 (1978) 279-282
(3) M. Dine and W. Fischler, *Phys.Lett.B* 120 (1983) 137-141; L.F. Abbott and P. Sikivie *Phys.Lett.B* 120 (1983) 133-136
(4) Isleif, K. for the ALPS Collaboration (2022), arXiv:2202.07306
(5) Shah, R et al., *PoS* EPS-HEP2021 (2022) 801
(6) S. I. Bityukov and N.V. Krasnikov, Nucl. Instrum. Methods Phys. Res., Sect. A 452, 518 (2000)
(7) H. Ismail Fawaz, *Data Mining and Knowledge Discovery* 33, 917-963 (2019)

M. M. acknowledges support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program Grant agreement No. 948689 (AxionDM).

I'm happy to discuss with you!
manuel.meyer@uni-hamburg.de

More information:
https://axion-alp-dm.github.io/