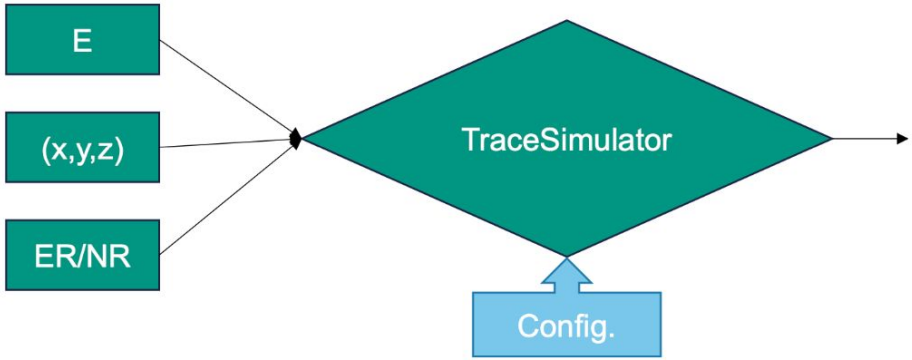
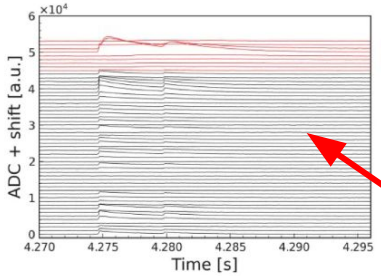
Reconstruction WG

Status Report

DELight Collaboration Meeting, Fri Sept 13, 2024

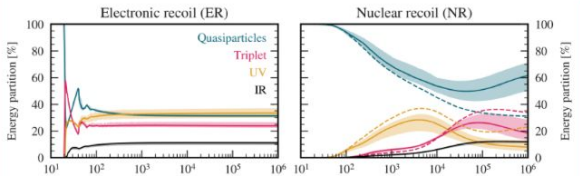
TraceSimulator: the newcomer



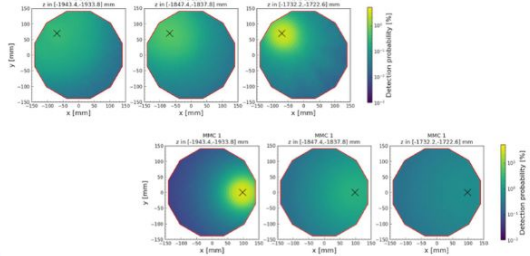


It's fantastic that we have this!

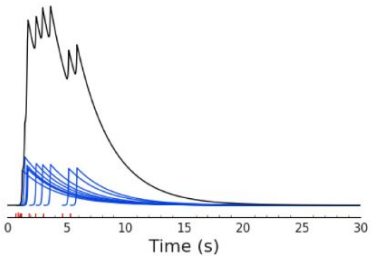
Energy partition



Detector effects

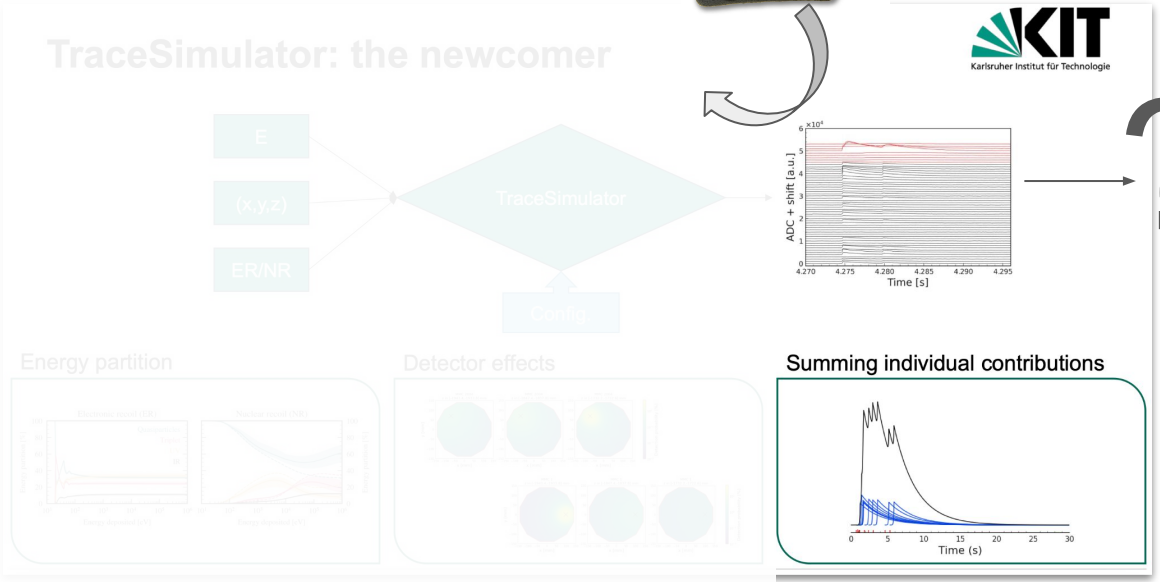


Summing individual contributions



Slide from Francesco

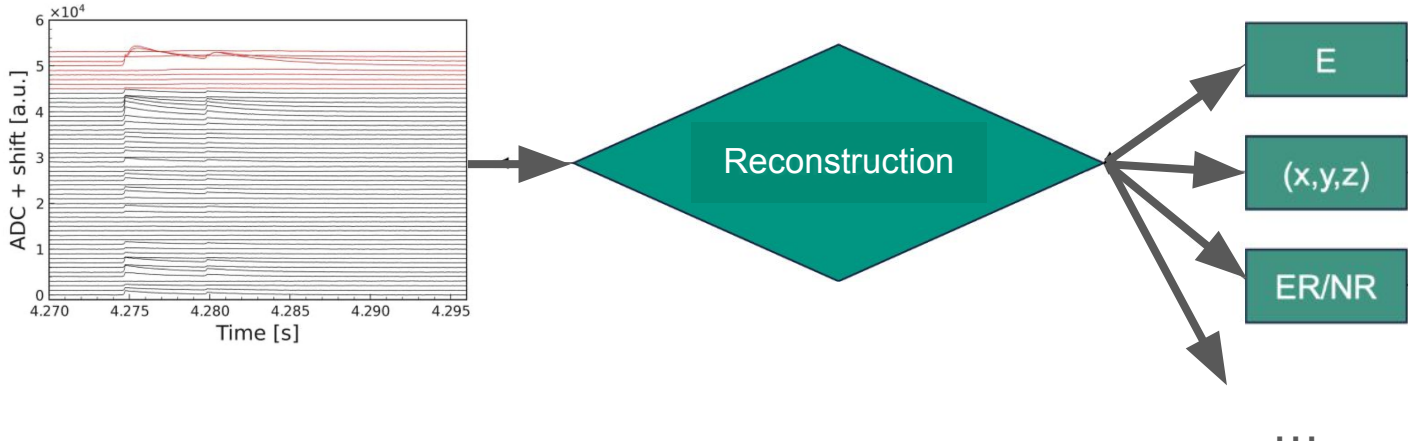
What do we want?



?

Crucial pieces missing here!
One of them: reconstruction!

Event reconstruction and classification

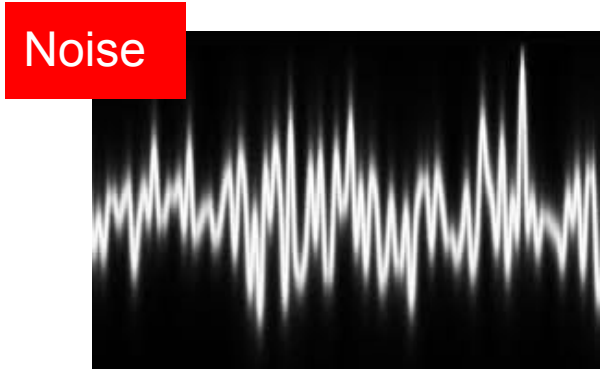


- Triggering technically part of reconstruction
  → Online and offline reconstruction
- We do have a prototype reconstruction setup based on OF and a framework Strax/Helix
  - Overall a good baseline, some bugs need to be fixed (e.g., some genuine events returning zero-amplitudes)

What’s currently missing to devise a reconstruction?

- ~~Realistic signal traces~~

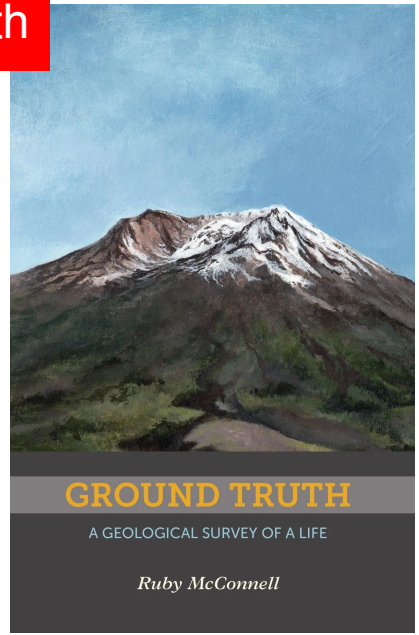
What's currently missing to devise a reconstruction?



- ~~Realistic signal traces~~
- Reasonable noise
  - Either use measured noise
  - Or better noise model with good scaling
  - Atm: randomly faked noise (white, pink, grey, …)
- First larger-scale sample generations starting ~now
- Missing noise not a problem: generate samples without noise until we are more sure of noise model → mix in & swap out noise later
- However: Question for Sebastian on what most realistic noise model for MMCs is
- Helix needs a new Plugin to get noise/baseline resolution via OF fit on noise events

What's currently missing to devise a reconstruction?

Ground truth



- ~~Realistic signal traces~~
- Reasonable noise

- Ground truth
  - Each simulated run gets a name and run meta JSON (length of simulation)
  - Right now, no plugin that matches injected events with triggered/reconstructed events
    - Would need to be done manually by matching inject & trigger times
  - Record is numpy array that corresponds to 5s of data
  - For first studies, we can inject == 1 event into the trace and truncate record, which makes matching trivial

What’s currently missing to devise a reconstruction?

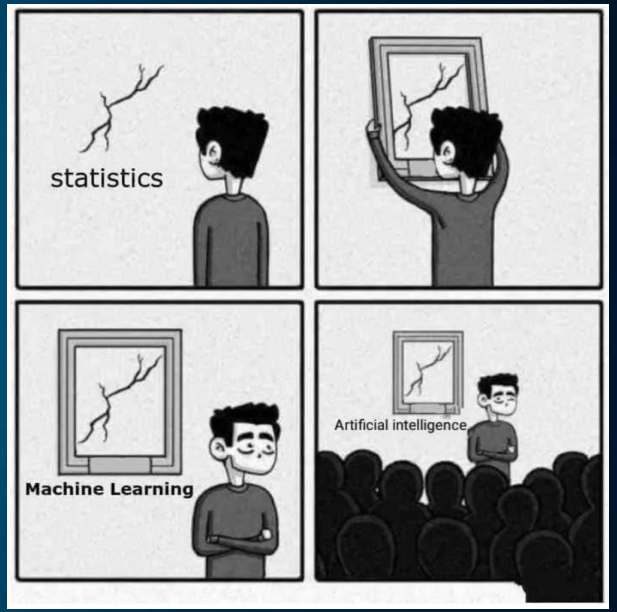
- ~~Realistic signal traces~~
- Reasonable noise 

- Ground truth 

- Speed up of simulating a timestream
  - Processing (triggering/noise gen/OF)
  - 10 sec containing 30 events with 45 submerged and 9 vacuum MMC takes O(1 min) at each step
  - Currently identifying bottlenecks, Bachelor thesis to speed up some of these steps with Generative AI

Ground truth

Can artificial intelligence help?



Literature search – what's out there?

Anomaly detection

CNN Autoencoder

A semi-supervised approach to dark matter searches in direct detection data with machine learning

Juan Herrero-Garcia (Valencia U., IFIC), Riley Patrick (ARC, CoEPP, Australia), Andre Scaffidi (INFN, Turin)

Oct 23, 2021

20 pages
Published in: *JCAP* 02 (2022) 02, 039
Published: Feb 28, 2022
e-Print: 2110.12248 [hep-ph]
DOI: 10.1088/1475-7516/2022/02/039
View in: ADS Abstract Service

pdf cite claim

Machine Learning @ SuperCDMS

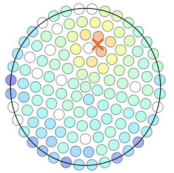
To Chin YU (ytc@stanford)
on behalf of SuperCDMS collaboration

PAPER • OPEN ACCESS

Neural-network-based level-1 trigger upgrade for the SuperCDMS experiment at SNOLAB

H. Meyer zu Theenhausen[1], B. von Krosigk[1] and J.S. Wilson[2]

RNN



Sig vs Bkg

CNN Autoencoder

Convolutional Neural Networks for Dire

Charanjit K. Khosa (Sussex U.), Lucy Mars (Sussex U.), Joel Richards (Sussex U.), Veronica Sanz (Sussex U. and Valencia U., IFIC)

Nov 20, 2019

20 pages
Published in: *J.Phys.G* 47 (2020) 9, 095201
Published: Jul 22, 2020
e-Print: 1911.09210 [hep-ph]
DOI: 10.1088/1361-6471/ab8e94 (publication)
View in: ADS Abstract Service

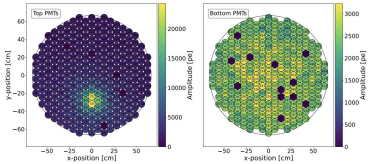
pdf cite claim reference search 20 citations

A Deep Learning-Based 3D Position Reconstruction for XENONnT

BACHELOR THESIS

Philip Siddhartha Thielges

Westfälische Wilhelms-Universität Münster



Position regression

CNN Autoencoder

Suppression of accidental backgrounds with deep neural networks in the PandaX-II experiment

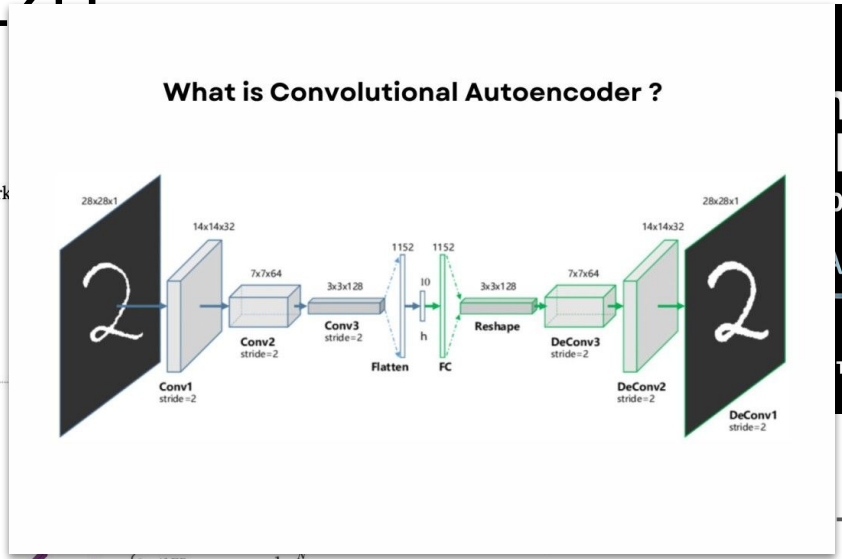
Sig vs Bkg

MLP

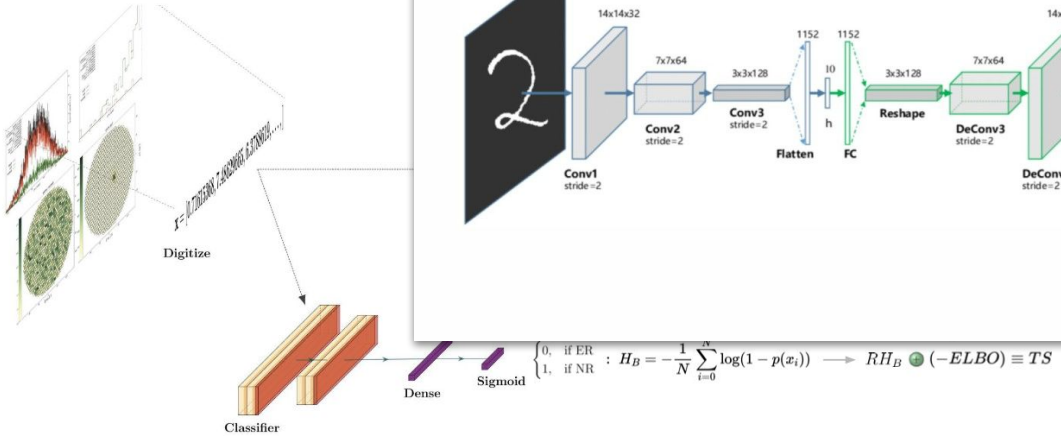
Nasir Shaheed[a] Xun Chen,[b,c,1] Meng Wang,[a,2]

→ Lots of room for improvement

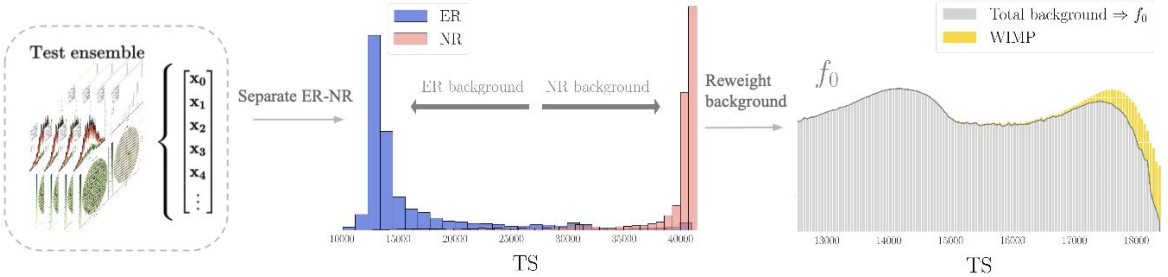
DARWIN / XLZD



1 Extraction of anomaly function from neural network



2 Extraction of NR and ER background pdf from TS distribution to determine presence of anomalous (non=background) events



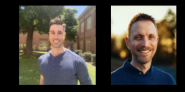
Train autoencoder only on ER!

Will be

DARWIN / XLZD

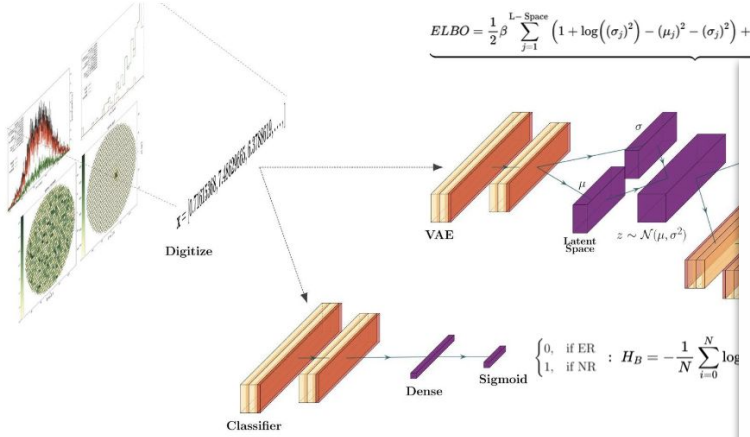
Anomaly aware machine learning for dark matter direct detection at DARWIN

PHYSTAT 10/11/2024, London

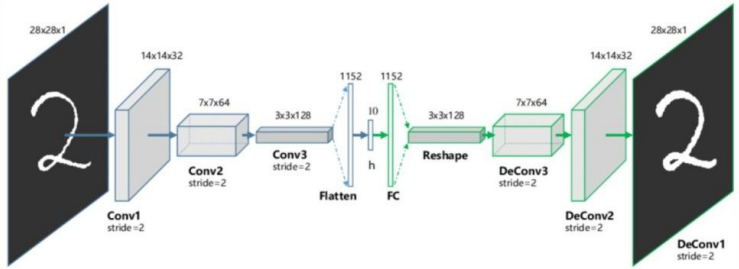


1 Extraction of anomaly function from neural networks

$$ELBO = \frac{1}{2}\beta \sum_{j=1}^{\text{L-Space}} \left(1 + \log\left((\sigma_j)^2\right) - (\mu_j)^2 - (\sigma_j)^2\right) + \frac{1}{L}\sum_{i=1}^{\text{Batch}} \log \mathcal{N}\left(\mathbf{x}^{(l)} \mid \boldsymbol{\mu}_D{}^{(l)}, \log \boldsymbol{\sigma}_D^{2(l)}\right).$$



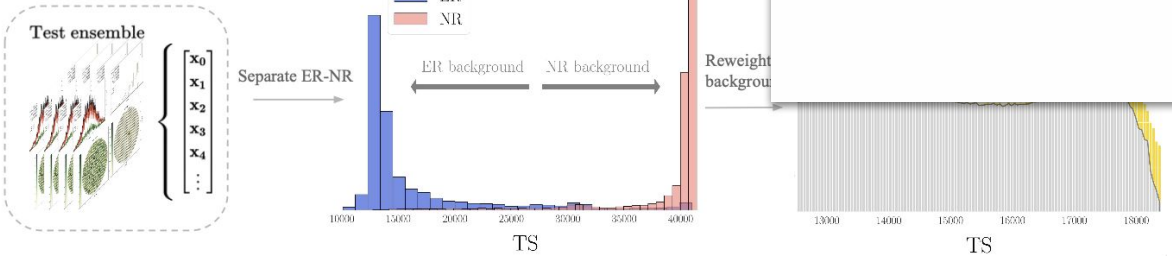
What is Convolutional Autoencoder ?



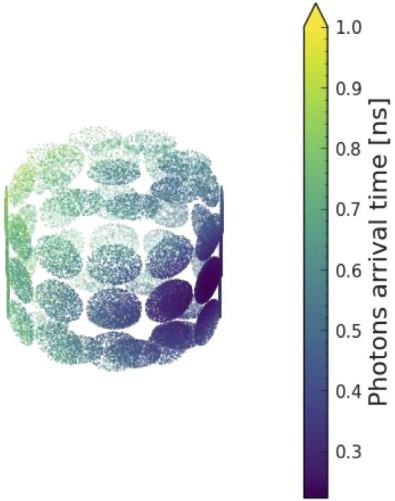
der only on

tag NR

2 Extraction of NR and ER background pdf from TS distribution to determine presence of anom



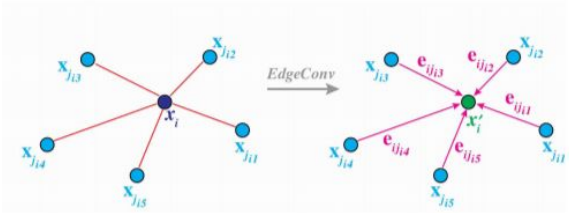
Event reconstruction and classification

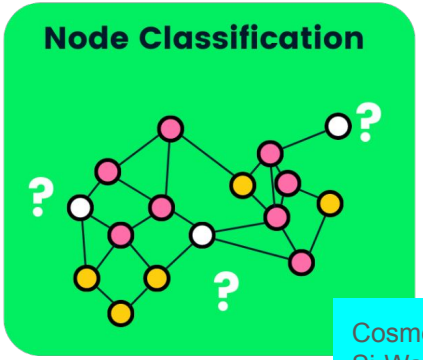


Web (“point cloud”) of 50-100 MMCs surrounding cell

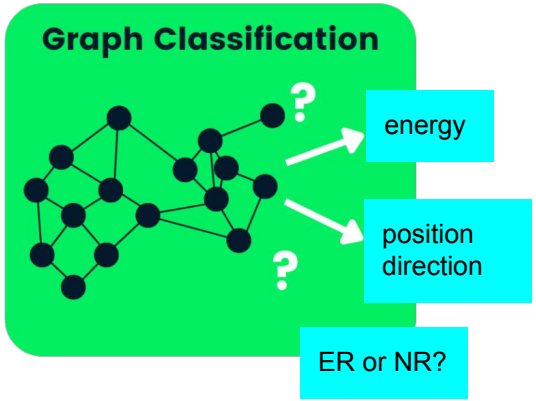
Lesson from Machine Learning in Particle Physics:

- Best performance if structure of your data matches neural network architecture

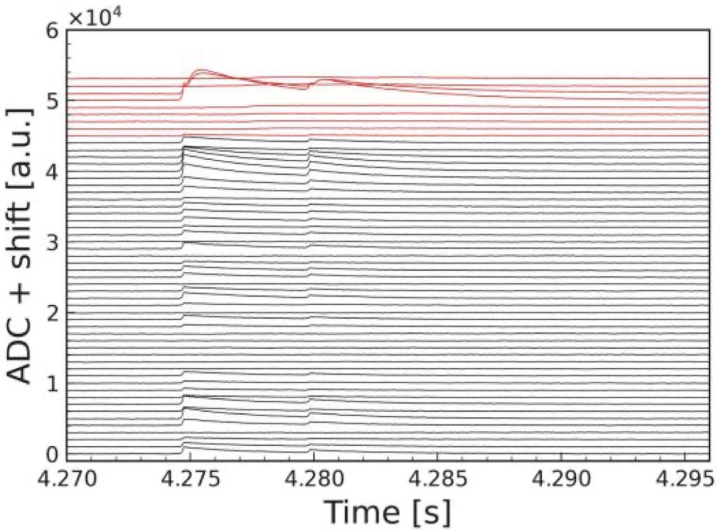




Cosmogenics in certain Si-Wafer?



Event reconstruction and classification



- 50 channels
- Most information contained in correlation of multiple channels
- Poor scaling behaviour (n^2) of many traditional algorithms when correlating channels
- Can explore algorithms known from Natural Language Processing (ChatGPT), i.e., Transformers:
  - Ideal for sequence modelling and modelling of correlations
  - Can model longe-range dependencies efficiently (e.g. for Helium triplet states arriving much later)

Conclusions

- Breakthrough achieved in availability of signal trace generator
- Students starting their theses on several aspects of simulation/reconstruction chain → Speeding up, reconstructing (x,y,z), NR vs ER …
- Some use ML, but traditional reco essential
- Talking to Henrique Araujo, for him the critical point about using ML is the robustness/quality of the simulation at threshold
- Exciting times ahead