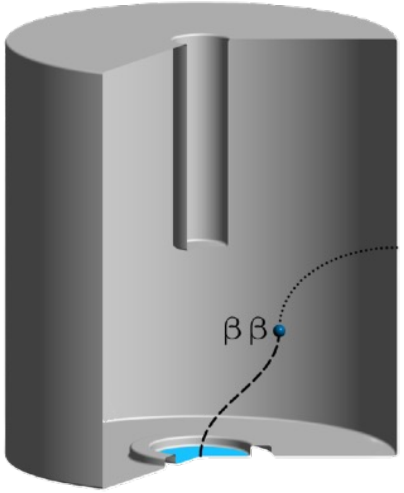
Enhancing event discrimination in high-purity Ge detectors with transformer

MARTA BABICZ
marta.babicz@physik.uzh.ch

Universität Zürich UZH

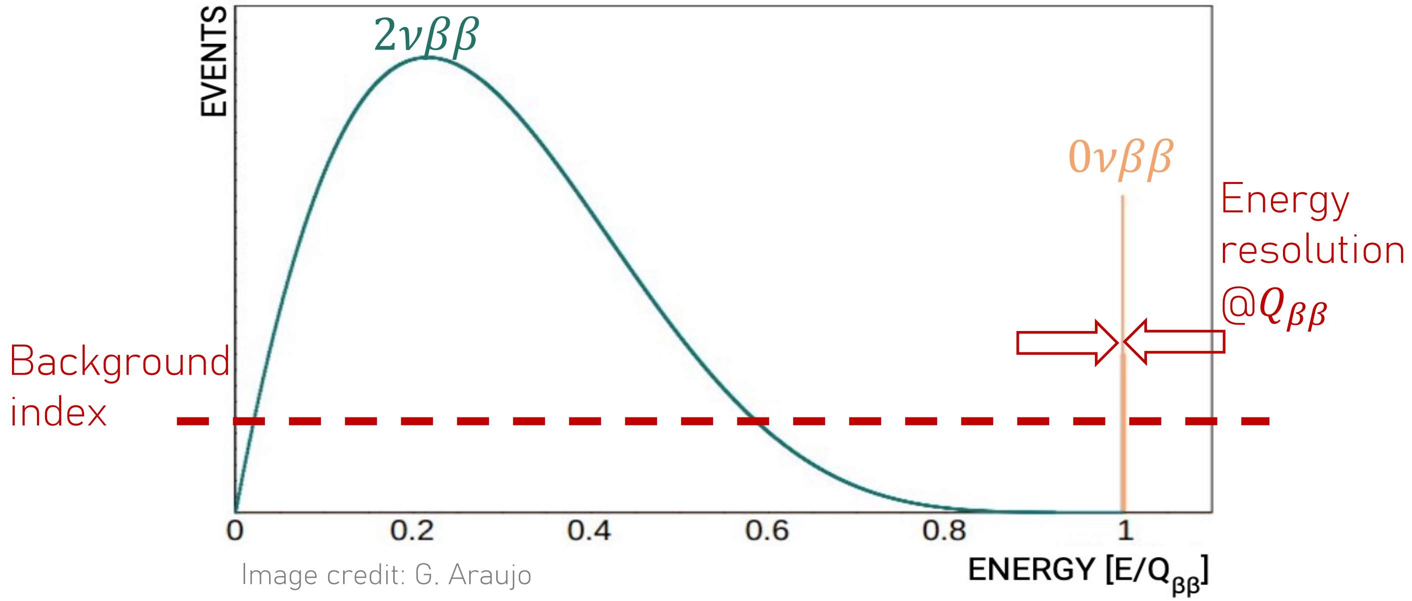
Searching for $0\nu\beta\beta$ with Germanium

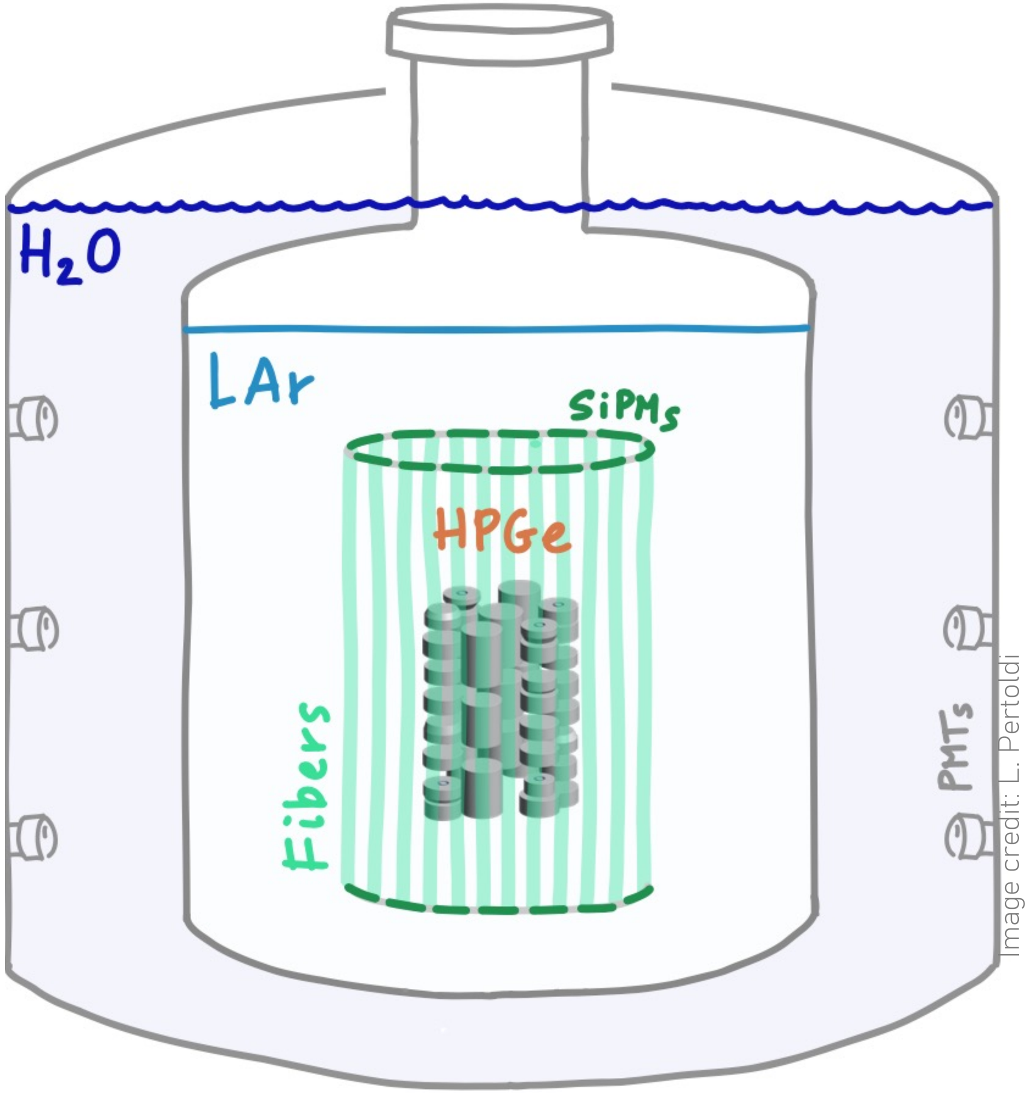


High-Purity Germanium detectors enriched in ^{76}Ge:

- source = detector → *high efficiency*
- High-purity → *low intrinsic background*
- Ge crystal → *outstanding energy resolution*
- Very good topological discrimination

$^{76}Ge \rightarrow {}^{76}Se + 2e^-$, $Q_{\beta\beta}$ = 2039 keV, $T_{1/2} > 10^{26}$ yr





GERDA + MAJORANA DEMONSTRATOR + new institutions

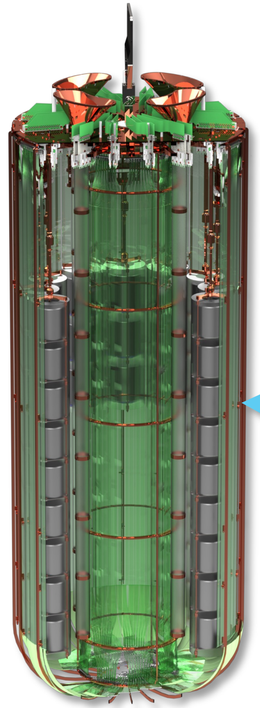
Lowest background index for 0νββ:
$5.2^{+1.6}_{-1.3} \cdot 10^{-4}$ cts/(keV kg yr)

Best energy resolution for 0νββ:
2.52 ± 0.08 keV (FWHM) at $Q_{\beta\beta}$

Gran Sasso Mountain: overburden (1.4 km)

LEGEND-200
- 200 kg of HPGe in GERDA cryostat
- Taking data since March 2023 with 142 kg of Ge
- $B \sim 2 \cdot 10^{-4}$ cts/(keV · kg · yr) → $T^{0\nu}_{1/2} > 10^{27}$

LEGEND-1000
- 1 ton of Ge, pending funding approval
- $B < 10^{-5}$ cts/(keV · kg · yr) → $T^{0\nu}_{1/2} > 10^{28}$
- Fully cover $m_{\beta\beta}$ inverted ordering region



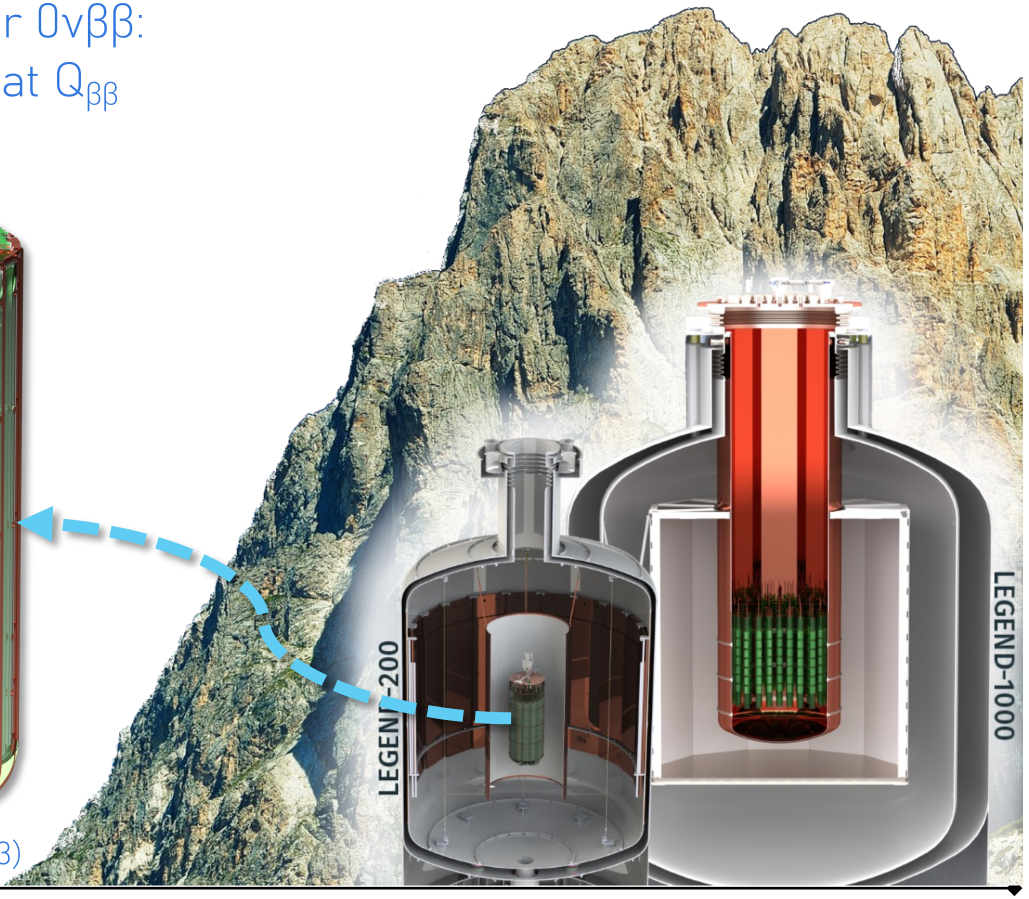
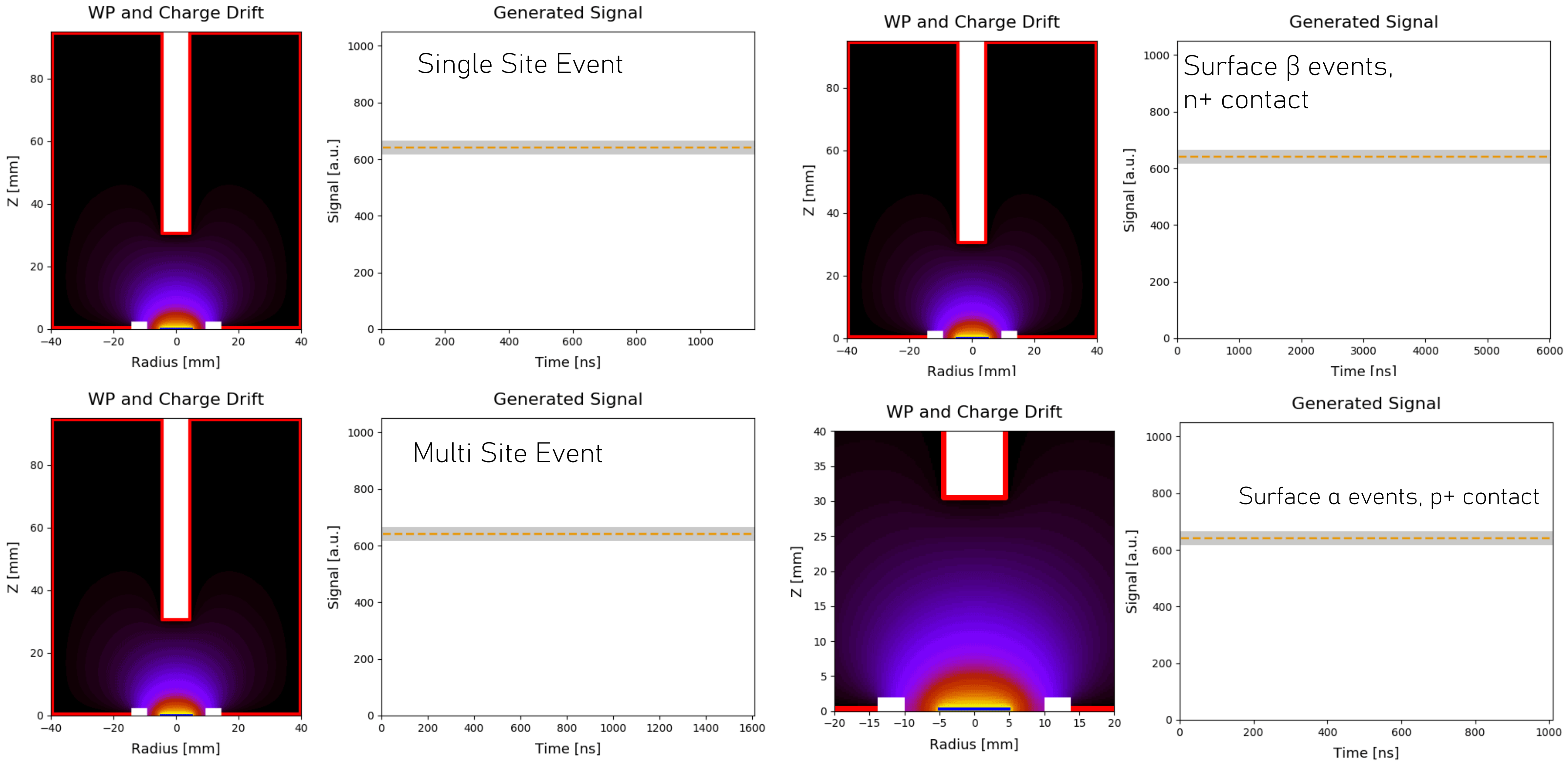


Image credit: L. Pertoldi

Ref.: GERDA Collab., PRL 125, 252502 (2020)   Refs.: MAJORANA Collab., PRL130, 062501 (2023)

HPGe detectors event topology



Pulse Shape Discrimination (PSD)

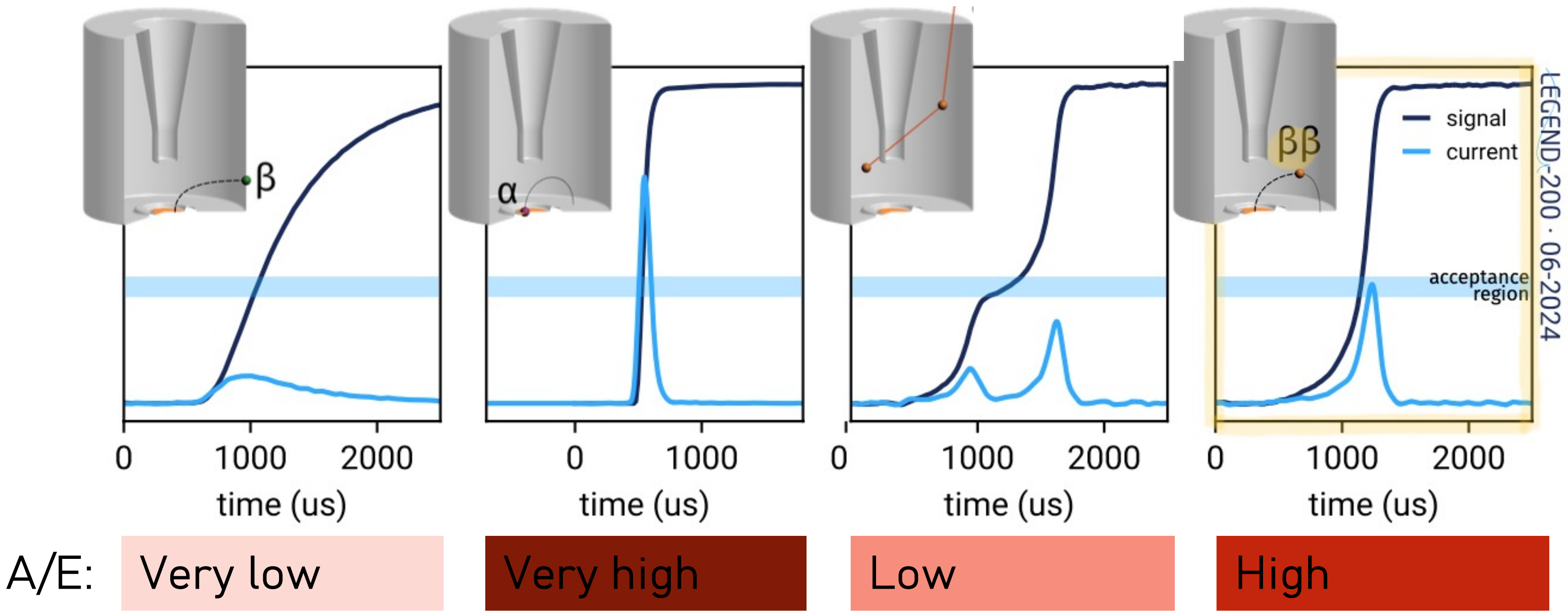
Pulse shape classifier :

$$A/E = \frac{\max(\color{cyan}{current})}{Energy}$$

Strong suppression of surface α and β (^{42}K) events

~60% suppression of Compton multi-site events at $Q_{\beta\beta}$

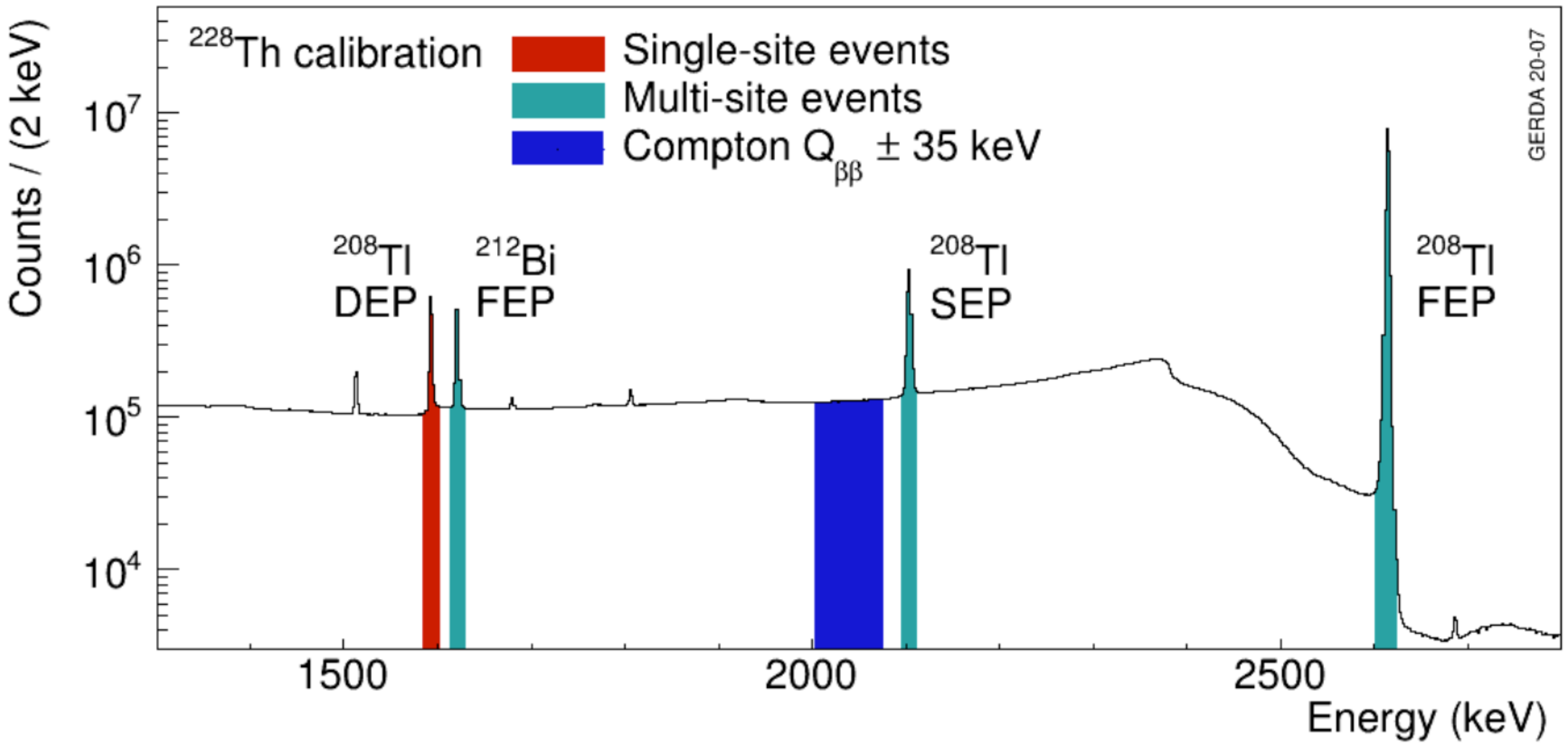
0νββ survival fraction of ~85%

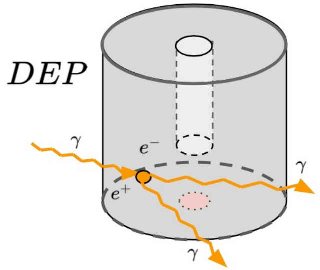


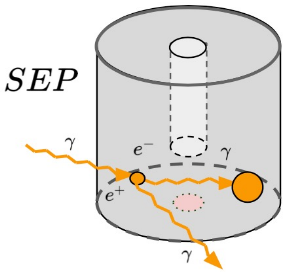
[$(A/E)_{MSE}$, $(A/E)_{n+}$]< $(A/E)_{SSE}$ < $(A/E)_{p+}$

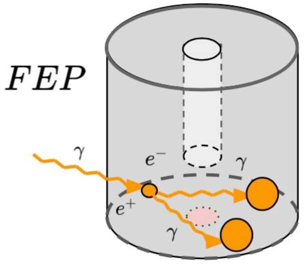
Calibration spectrum

Calibration data is crucial for determining precise upper and lower thresholds for optimal A/E cut.









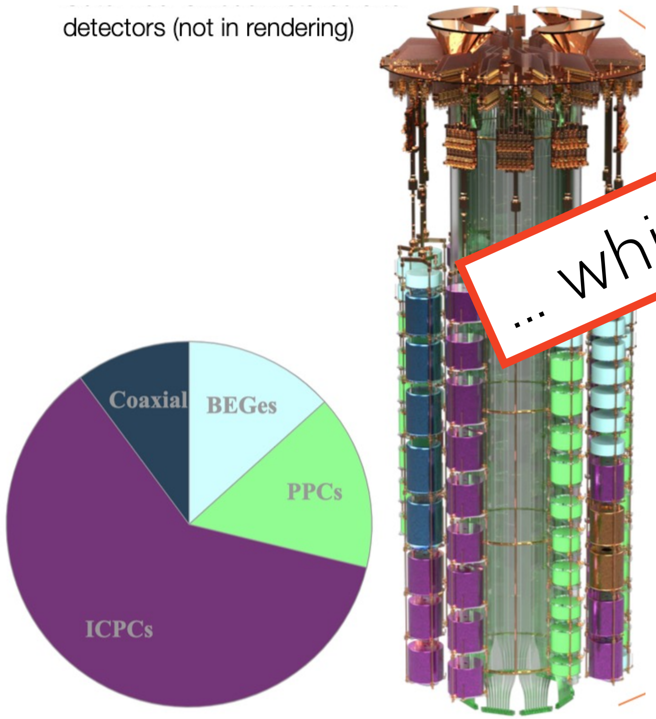
PSD labels

PSD Label	Description
psd_low_avse	multi-site background events
psd_high_avse	surface events near the point contact
psd_dcr	surface α events based on the decaying tail. DCR: Delayed Charge Recovery parameter
psd_lq	partial charge deposition events in the transition layer based on the turning region between rising edge and decaying tail. LQ: Late Charge parameter

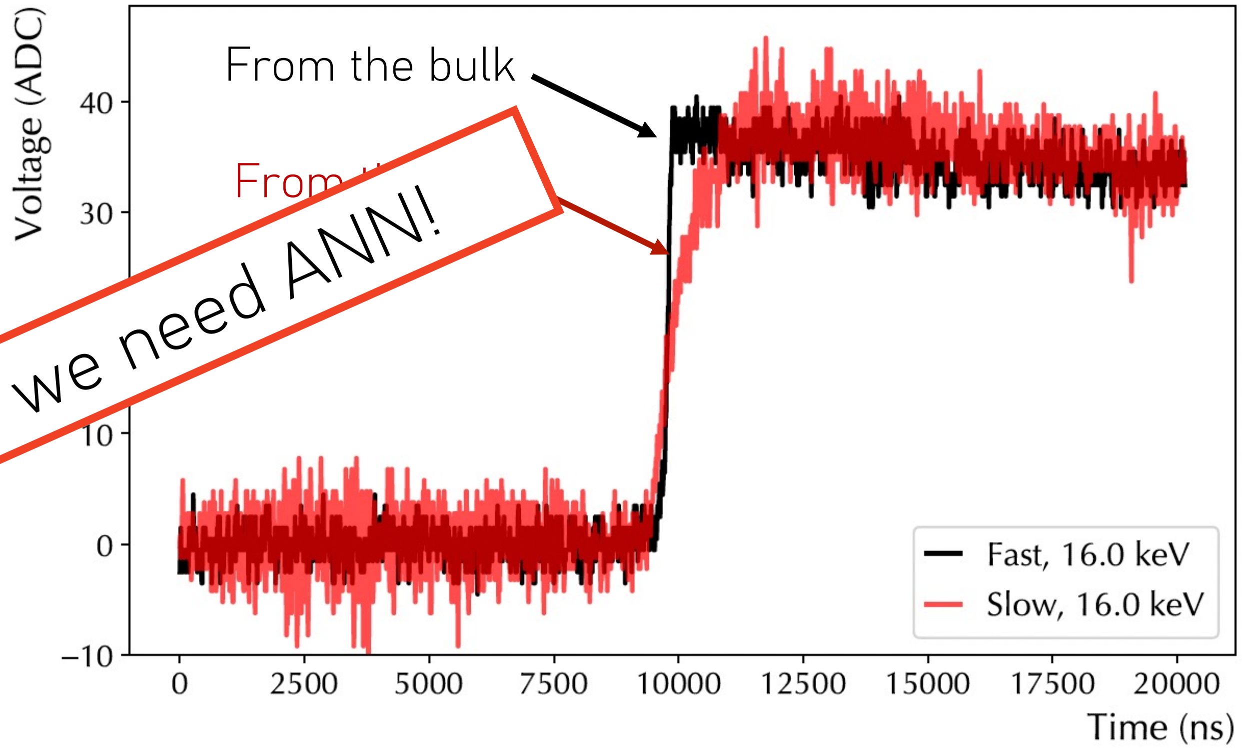
Note: Each PSD label is individually calibrated and corrected per detector, with uncertainty evaluations to ensure accuracy.

Coax detectors and "slow pulses"

For coaxial detectors, the A/E approach does not yield good results: similar maximal current amplitude (A) is possible for both SSEs and MSEs, even if the interaction sites happen to be at different radii.



"slow pulses" at low energies

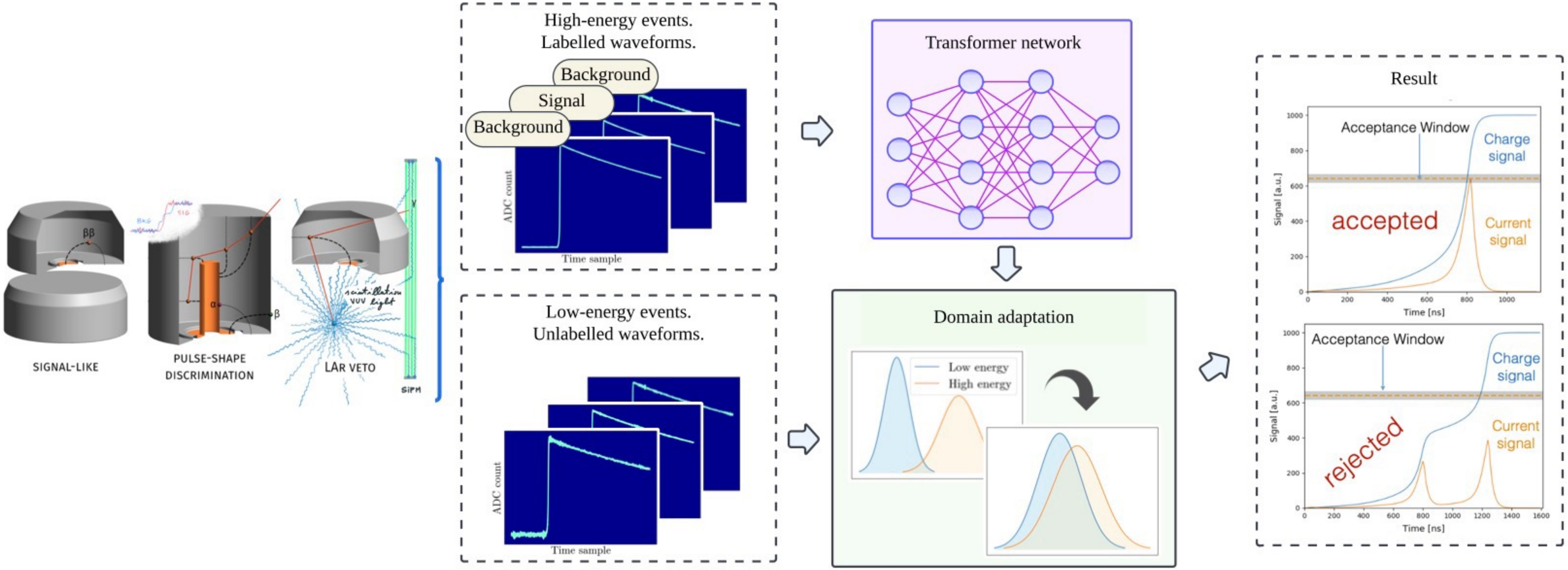


... which is why we need ANN!

- Slow pulses: energy-degraded pulses that occurred in the dead layer of the detector, e.g. Ar-39 events.
- These events distort the spectrum due to the lost energy, so need to be removed

Methodology

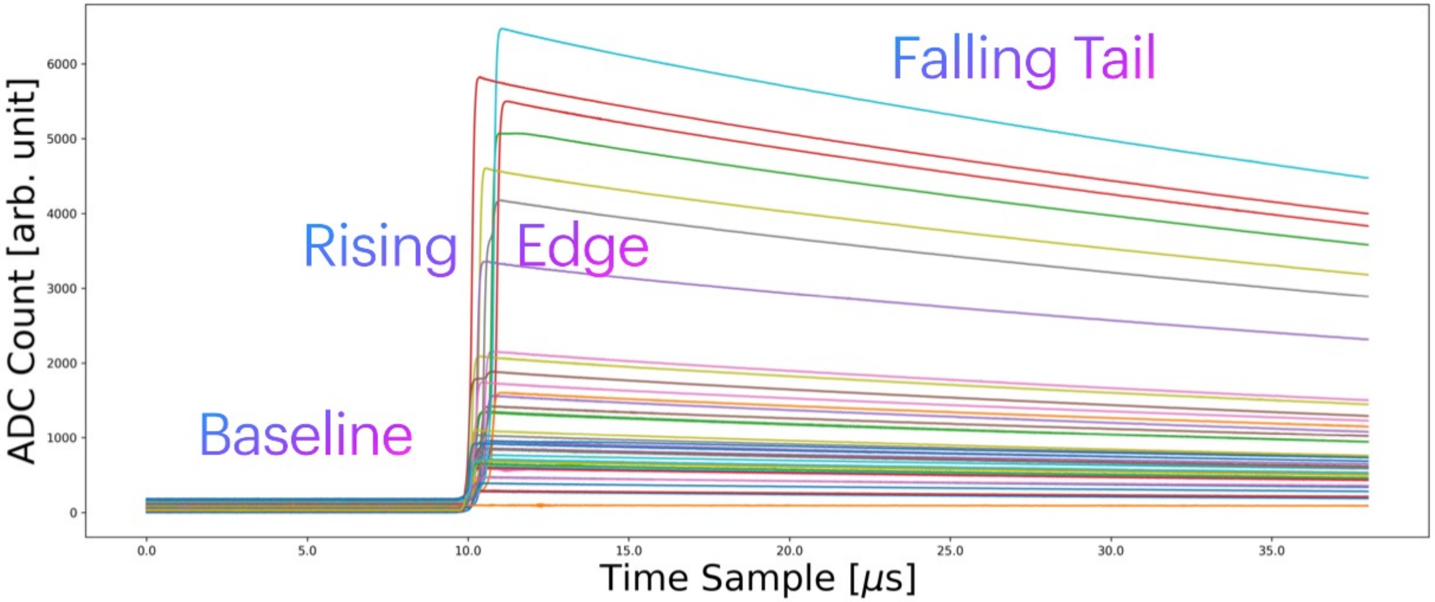
- Challenge: obtaining labelled training data, especially in low-energy regions where signal events are sparse.
- Solution: Domain adaptation mechanisms



NPML 2023 Machine Learning Challenge

https://doi.org/10.48550/arXiv.2308.10856

- Using the ^{228}Th calibration data, develop:
  - **PSD ML Model**: Classify events by predicting four PSD labels from raw waveform data, identifying clean events passing all PSD cuts.
  - **Energy ML Model**: Construct a model predicting event energy from waveform characteristics.
- Dataset:
  - Includes Germanium detector waveforms, PSD cuts, and calibrated energies in HDF5 format.
  - Serves for training and testing ML models.



- HPGe detector waveform, simple but information rich
- Pulse shape contains physics information (up to many corrections):
  - Waveform amplitude is proportional to particle energy
  - Shape of the rising edge reflects interaction type
  - Falling tail reflect electronic response

NPML Challenge

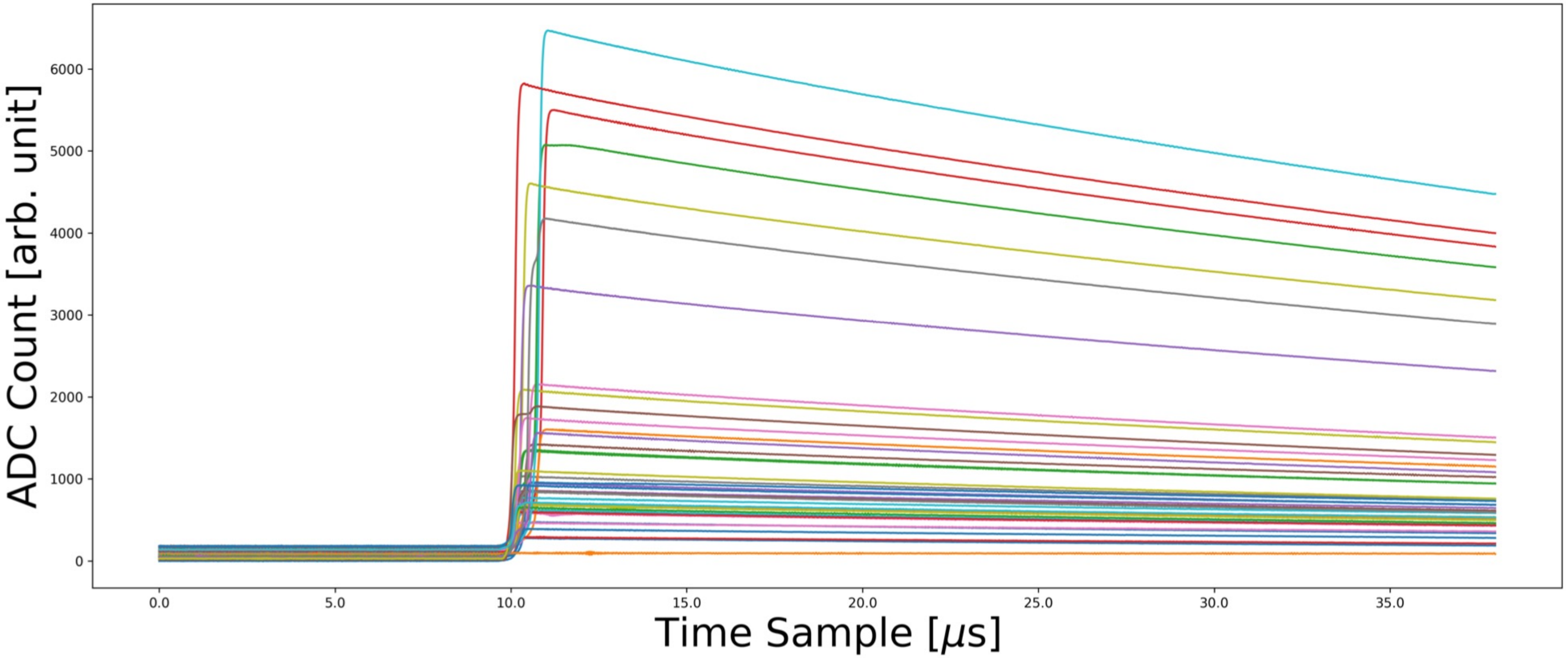
NPML Dataset

- Subset of ^{228}Th calibration data from 56 HPGE detectors with raw waveforms, PSD cuts, and calibrated energies. Divided into training, test, and NPML challenge subsets.
- Contains 3,193,486 events (after removing the background noise), :
  - Raw Waveform: Time series signals from Germanium detectors reflecting particle interactions.
  - Detector: Identifier for the recording Germanium detector.
  - Run Number: Identifies the experimental run for data collection.
  - tp0: Timestamp or calibration parameter for data acquisition.
  - Energy: Energy values derived from waveform data.
  - Analysis Labels: Summarizes event characteristics based on physics analysis.

psd_label_low_avse	multi-site backgrounds
psd_label_high_avse	surface events near the point contact
psd_label_dcr	surface α events based on the decaying tail
psd_label_lq	partial charge deposition events in the transition layer based on the turning region between the rising edge and decaying tail

Preparing the data

- Training data: 2.395.098
- Test data: 638.691
- Due to the size of the waveforms (3800 of length), models are trained with sequences representing either 128 or 256 values after the baseline



- The initial part of the waveform (rise time) contains the most information about the event, while the latter part (decay phase) is less informative due to less variation, higher noise influence, and completion of charge collection. So, using a window of 128 or 256 samples from the rise time for analysis provides sufficient information for event classification.

ML Model = Transformer

- Input: waveform sequences
- Applied linear transformation to the input for embedding
- Positional encoding to preserve sequence order
- Transformer encoding:
  - uses transformer encoder layers to capture contextual information
  - self-attention mechanism for sequence processing
- Adds a classification token (for each label) at the beginning of the sequence.
  - differentiates the classification token from the main sequence.
- Output Layer: Uses a linear layer on Transformer encoder output.
- Training: CrossEntropyLoss for training (for PSD class.) and ~~MSE~~/MAE (for energy regression), Adam optimizer.

- Model Hyperparameters (tr. encoder):
  - Layers: 5
  - Attention heads: 8
  - Model (embedding) size: 64.
  - Optimiser: Adam with weight decay of 0.0001.
  - Learning rate: 0.002 (warmup during the first epoch, cosine annealing afterwards).
  - Batch size: 64.
  - It's a multi-task network, so 5 outputs (4 PSD labels + energy)

PSD ML Mode

- Classify events by predicting four PSD labels from raw waveform data **256 samples** after the baseline

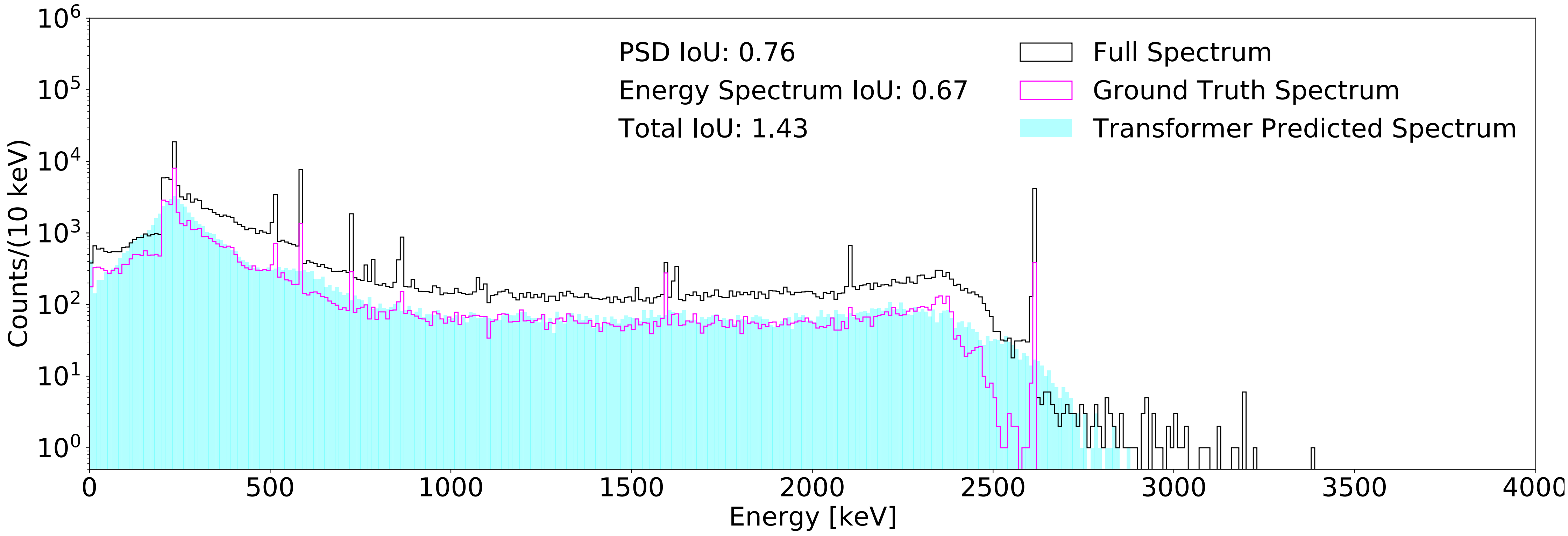
Label Type	Correct (%)	Misidentified (%)
psd_label_low_avse	93.73	6.27
psd_label_high_avse	99.29	0.71
psd_label_dcr	98.25	1.75
psd_label_lq	87.04	12.96
final_label	88.91	11.09

Energy reconstruction

total area covered by both the predicted and actual distr. combined

IoU=Area of Intersection/Area of Union

shared area where the predicted and actual distr. overlap.



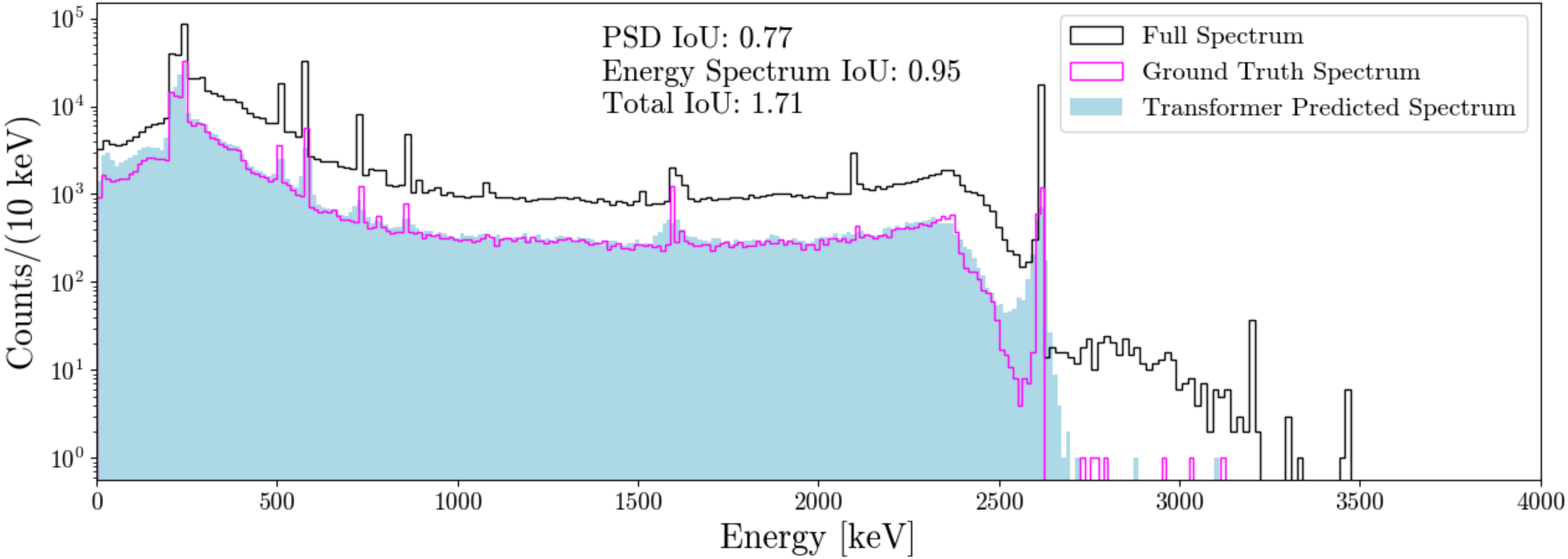
Mean squared error (MSE) as loss function. Clearly smoothing the energy distribution

Energy reconstruction

total area covered by both the predicted and actual distr. combined

IoU=Area of Intersection/ Area of Union

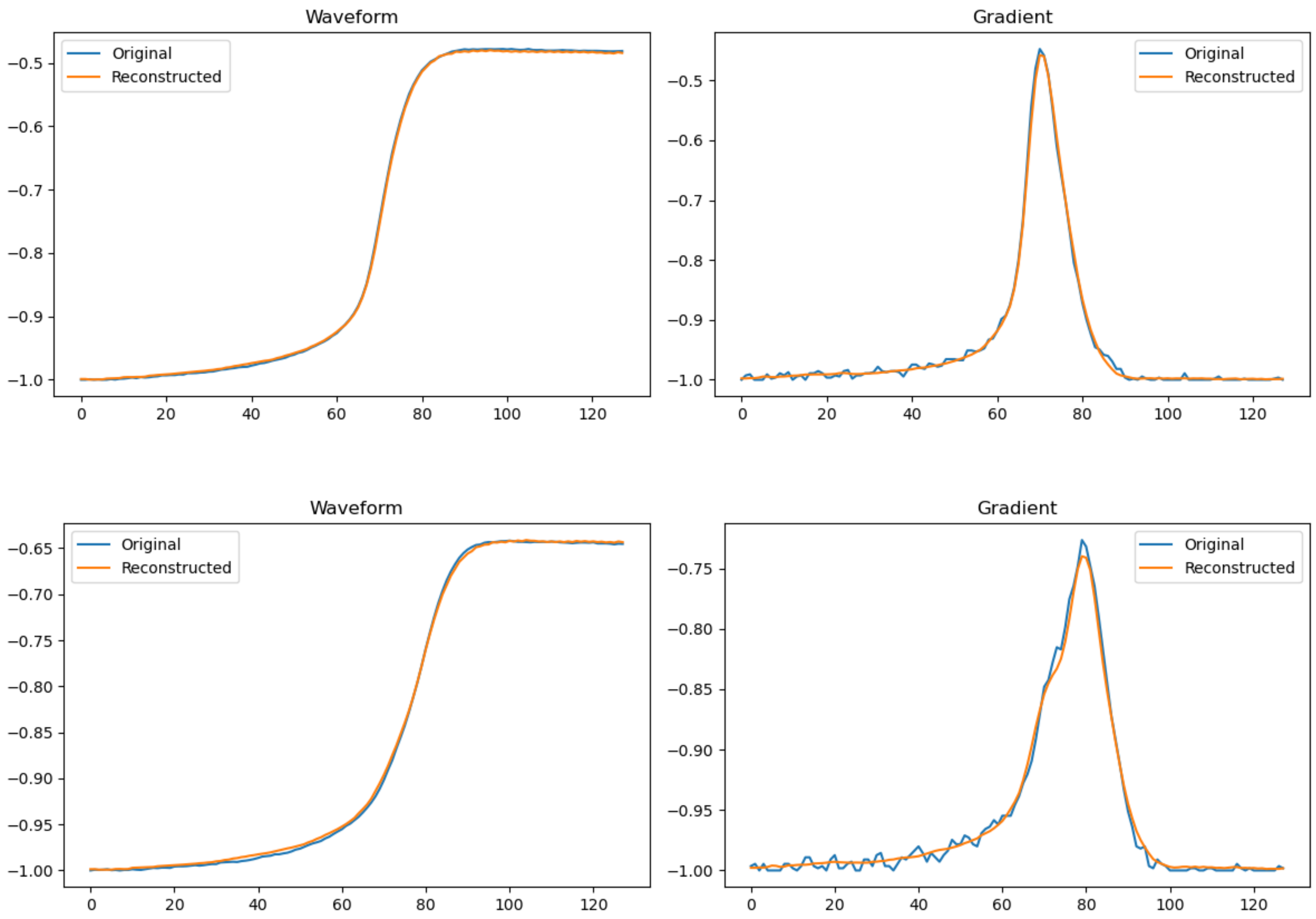
shared area where the predicted and actual distr. overlap.



Improvement: Mean absolute error (MAE) catches very well the energy distribution!

Denoising the waveforms

Using a Variational AutoEncoder (VAE) to reconstruct the waveforms (and thus, the reconstructed waveform looks denoised)



The denoised waveforms retain the essential information about the particle interactions while reducing the noise that can interfere with the analysis.

Summary and Outlook

- The Transformer model's performance achieved high classification accuracy for various label types, with the highest accuracy of 99.29% for psd_label_high_avse and the lowest accuracy of 87.04% for psd_label_lq.
- The overall accuracy for the final_label was 88.91%. Better that the ANN used in GERDA.
- MSE emphasizes larger errors and smooths the energy distribution by averaging out variations, while MAE treats all errors equally, preserving sharp peaks and important details in the energy distribution.
- The use of an autoencoder can effectively denoising the waveforms, but so far it doesn't improve the classification task.
- Next step: prepare a clean labelled dataset from the LEGEND calibration spectrum and train the model independently of the PSD labels.

Thank you

PSD ML Mode

- Classify events by predicting four PSD labels from raw waveform data **128 samples** after the baseline

Label Type	Correct Identification (%)	Misidentification (%)
psd_label_low_avse	93.76	6.24
psd_label_high_avse	99.29	0.71
psd_label_dcr	98.09	1.91
psd_label_lq	85.45	14.55

Label Type	Correct Identification (%)	Misidentification (%)
final_label	87.78	12.22