Energy correction in PandaX-III experiment using Machine Learning techniques

Andrii Lobasenko

CEA Saclay IRFU/DphN

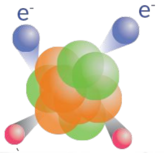
On behalf of PandaX-III collaboration

The 6th "International Workshop on Application of Noble Gas Xenon to Science and Technology"

08/06/2023

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 800945 — NUMERICS — H2020-MSCA-COFUND-2017

Searching for $0\nu2$

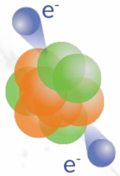


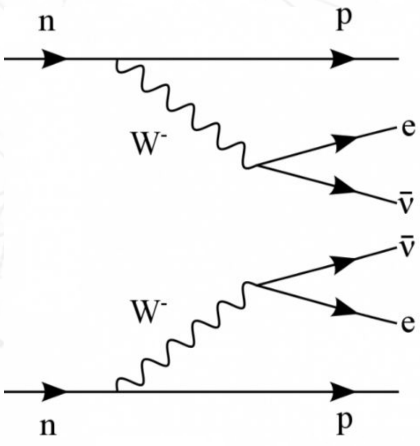
$2\nu\beta\beta$

hypothetical decay: $(A, Z) \rightarrow (A, Z+2) + 2e^-$

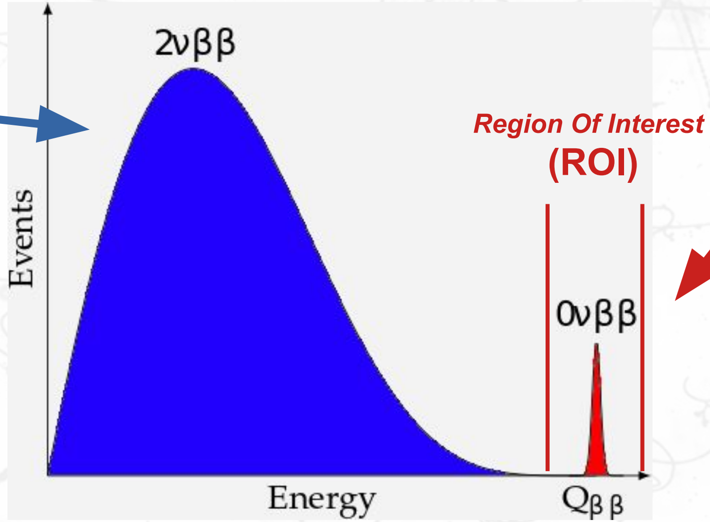
- **Majorana nature of the neutrino (own anti-particle)**
- Lepton number conservation violation
- Matter/antimatter asymmetry

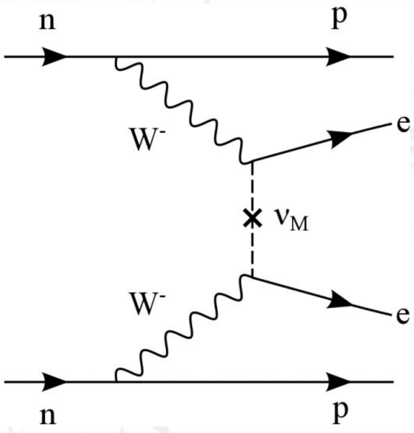
$0\nu\beta\beta$





Electron energy spectrum of ***ββ*** **decay**





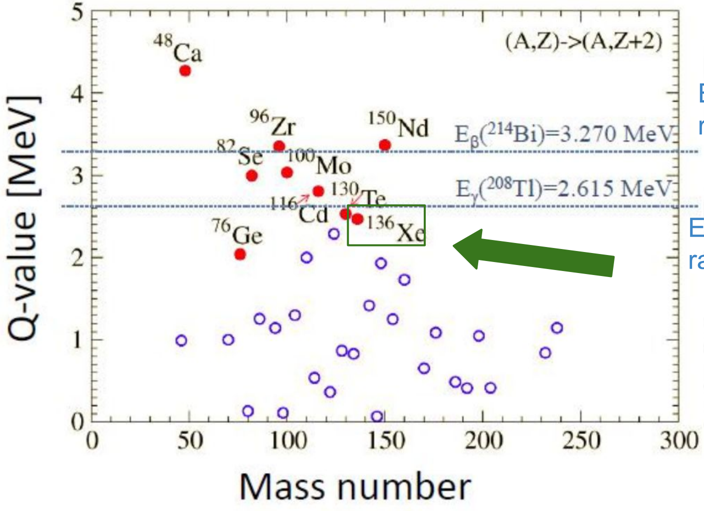
The rarest nuclear decay process

$$T_{1/2}(2\nu\beta\beta) \sim 10^{18} - 10^{24}\ yr$$

Expected half-life time of the 0ν decay

$$T_{1/2} > 10^{25} yr$$

Xe136 gaseous TPC

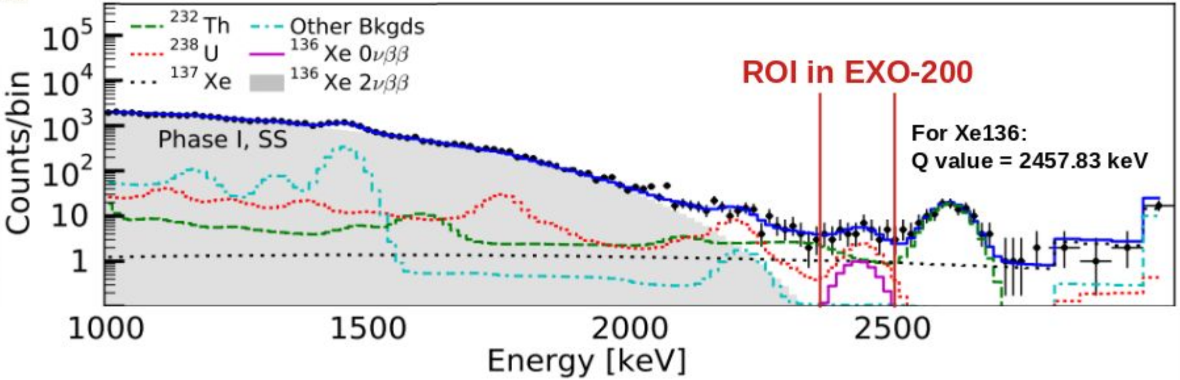


End of β radioactivity

End of γ radioactivity

Xe136 pros:
-Noble gas (**gaseous amplification**) => Can be used in TPCs
- good natural abundance ~9%
Cons:
- Low **Q-value** (**2457.83 keV**) => Higher probability of a bkg contamination

EXO 2019 results (liquid Xenon detector)

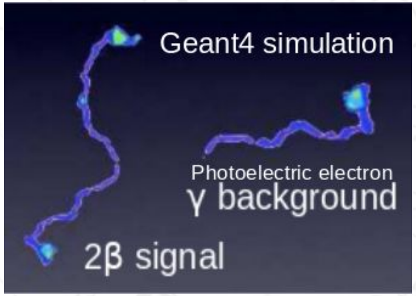


ROI is contaminated by the bkg:
U238 and Th232 decay chains:
- 2448 keV gamma from **Bi214**
- 2615 keV gamma from **Tl208**

PhysRevLett.123.161802, Search for $0\nu\beta\beta$ with the Complete EXO-200 Dataset (2019)

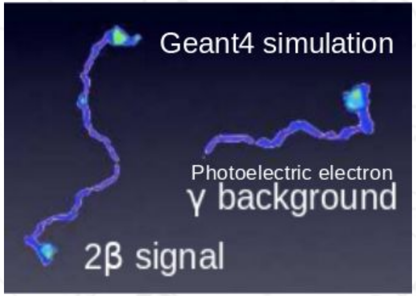
Xe136 gaseous TPC

Topological difference b/w bkg and signal



Xe136 gaseous TPC

Topological difference b/w bkg and signal



Characteristics of the double-beta decay events

Double-beta decay: 2 electrons → 2 Bragg peaks

Background gamma events: 1 electron → 1 Bragg peak

But very scattered tracks, recognition not always obvious

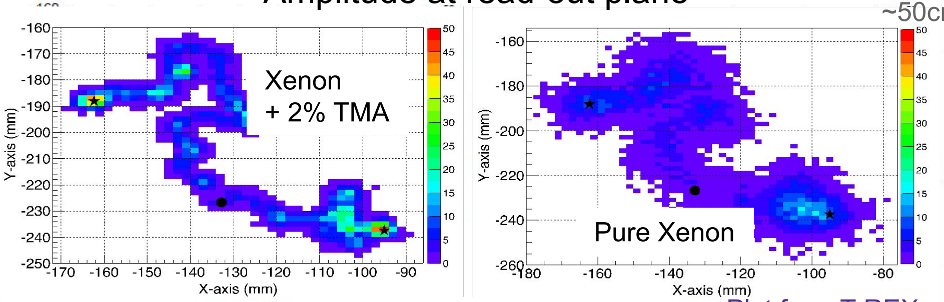
Also need to reconstruct precisely the deposited energy

1% tri-methyl amine (TMA) in gas mixture helps a lot:

- Lower diffusion
- Better energy resolution
- Quencher for the gaseous amplification
- Suppress scintillation

Amplitude at read-out plane

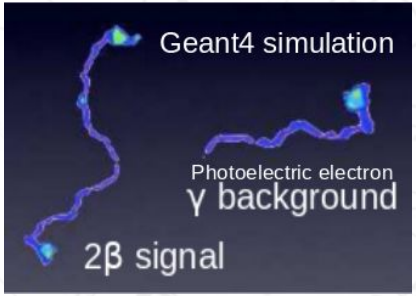
MC of double-beta decay at 10 bar ~50cm drift



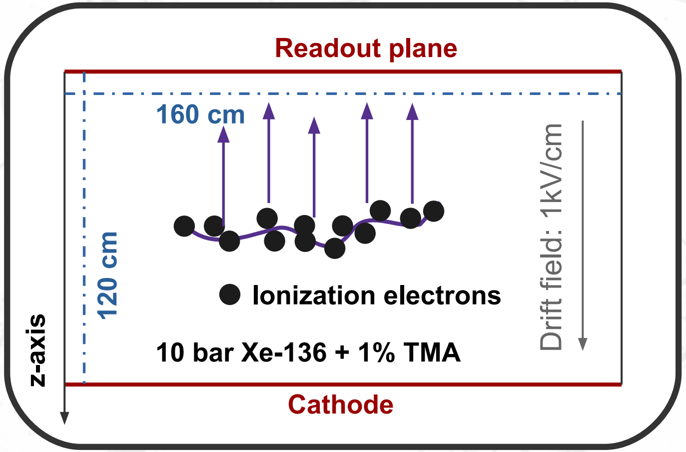
Plot from T-REX project

Xe136 gaseous TPC

Topological difference b/w bkg and signal



Schematics of the PandaX-III TPC geometry



Characteristics of the double-beta decay events

Double-beta decay: 2 electrons → 2 Bragg peaks

Background gamma events: 1 electron → 1 Bragg peak

But very scattered tracks, recognition not always obvious

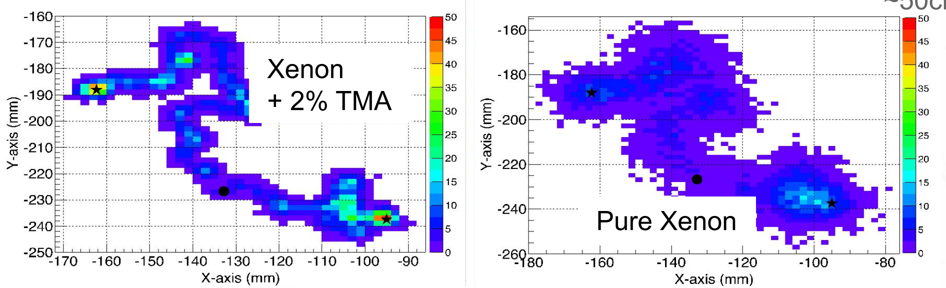
Also need to reconstruct precisely the deposited energy

1% tri-methyl amine (TMA) in gas mixture helps a lot:

- Lower diffusion
- Better energy resolution
- Quencher for the gaseous amplification
- Suppress scintillation

Amplitude at read-out plane

MC of double-beta decay at 10 bar ~50cm drift

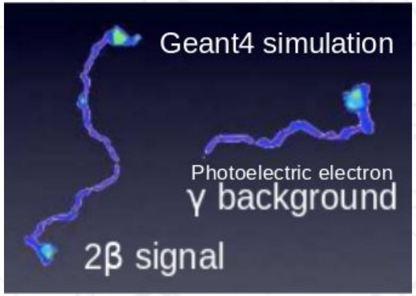


Plot from T-REX project

Tao Li et al. “Reconstruction of the event vertex in the PandaX-III experiment with convolution neural network”

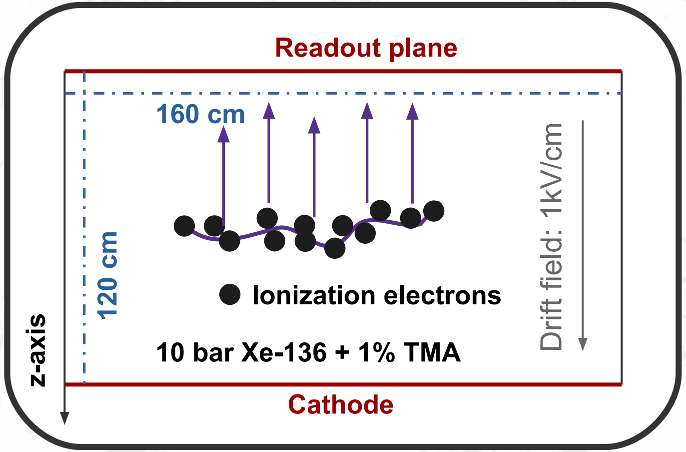
Xe136 gaseous TPC

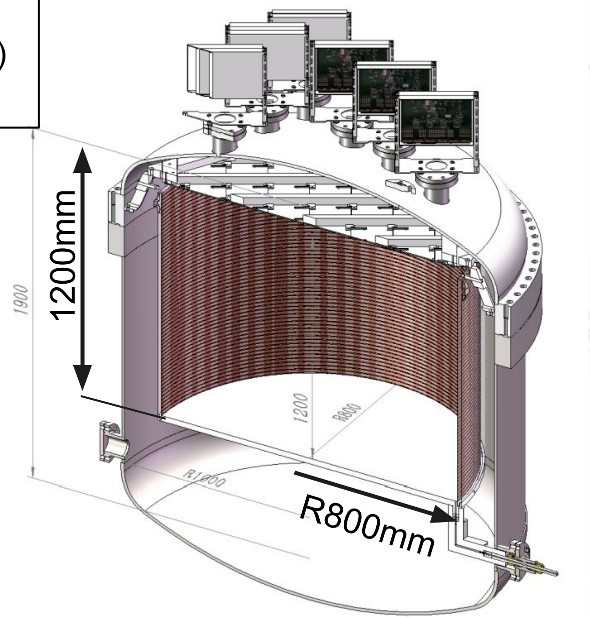
Topological difference b/w bkg and signal



90% enriched Xe136 gas +1% TMA
Pressure: **10 bar**
Total gas mass: **140 kg** (for the first module)
Goal : 5 x 200kg TPC modules in total

Schematics of the PandaX-III TPC geometry

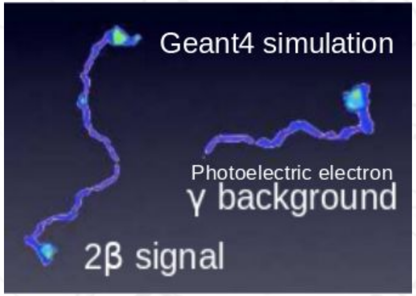




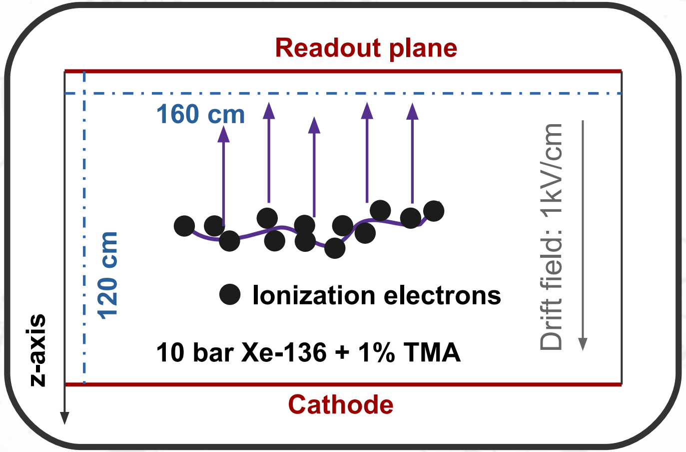
Ni Kaixiang's thesis: "Searching for Neutrinoless double beta decay of 136Xe in PandaX detectors"

Xe136 gaseous TPC

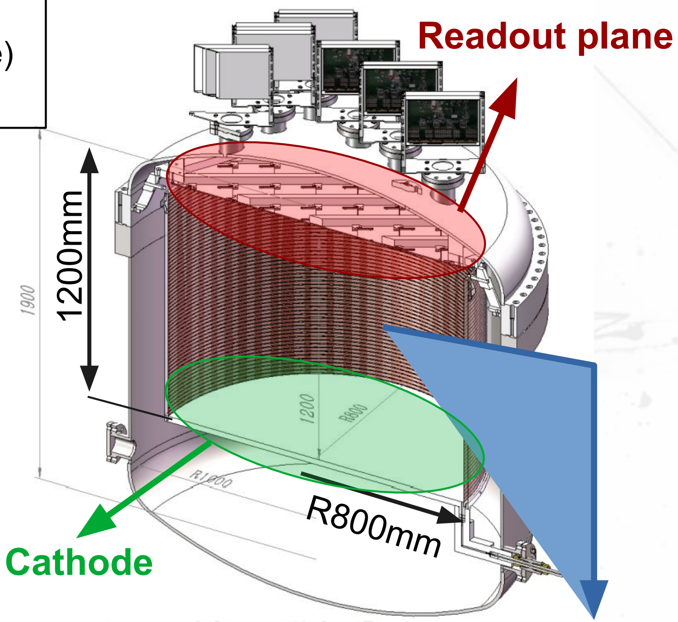
Topological difference b/w bkg and signal



Schematics of the PandaX-III TPC geometry



90% enriched Xe136 gas +1% TMA
Pressure: **10 bar**
Total gas mass: **140 kg** (for the first module)
Goal : 5 x 200kg TPC modules in total



Tiled kapton flexible PCB
Low radioactive material
Built by TangChen (JUNO vendor)
Tested successfully at 120kV voltage

Ni Kaixiang's thesis: "Searching for Neutrinoless double beta decay of 136Xe in PandaX detectors"

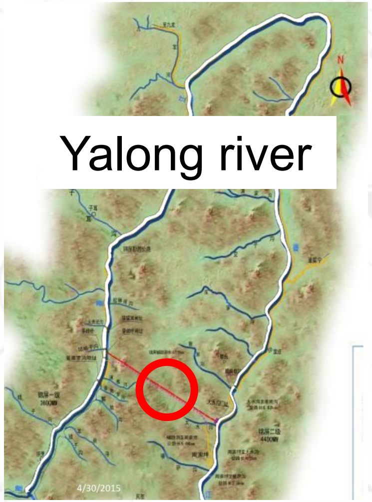
Xe136 gaseous TPC

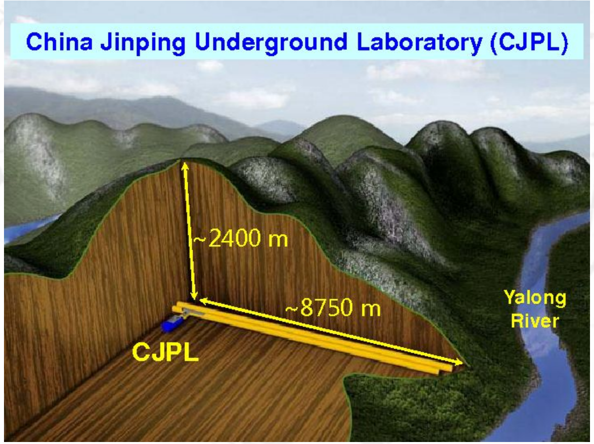
International collaboration:

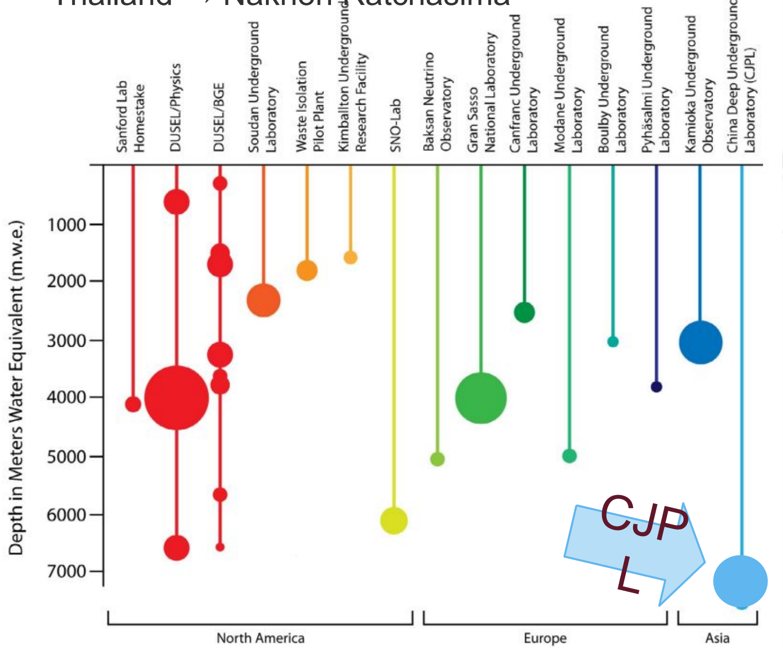
- China → 7 institutes (lead by SJTU)
- France → CEA Saclay
- Spain → Zaragoza
- USA → BNL + Maryland University
- Thailand → Nakhon Ratchasima

Laboratory

- Jinping CJPL-II underground laboratory (Sichuan, China)
- One of the worldwide lowest muon flux
- Large caverns, easy access to trucks





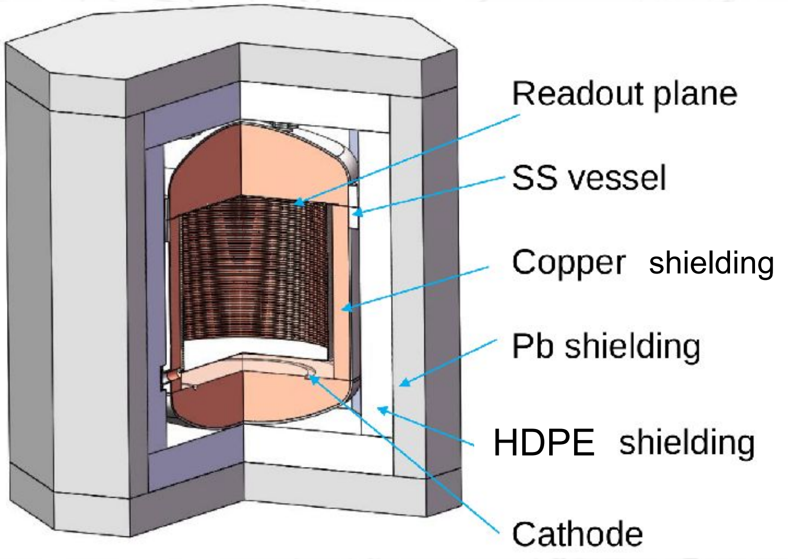


An Assessment of the Deep Underground Science and Engineering Laboratory (DUSEL)

Reaching the cosmic **bkg level** of **~1 cts/week/m^2**

Xe136 gaseous TPC

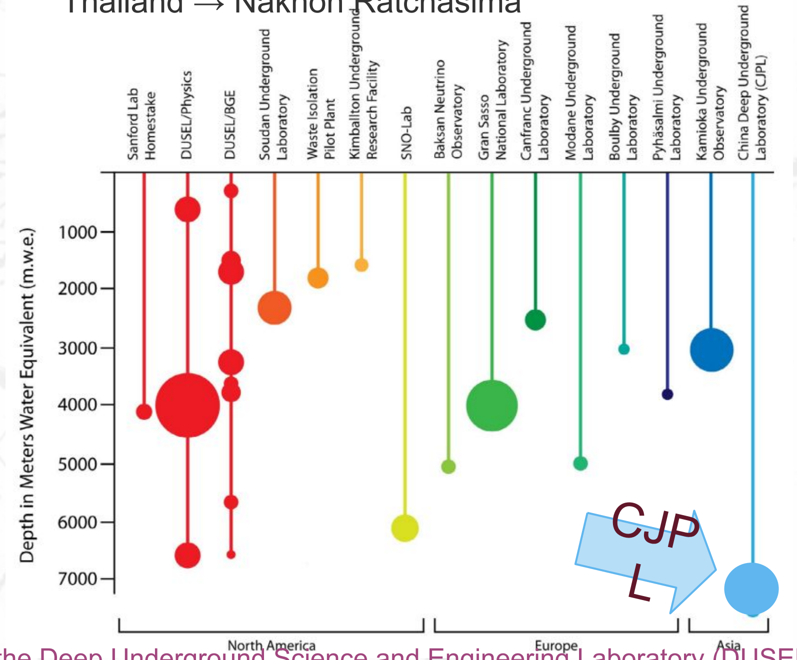
Shielding to reduce U (^{214}Bi) and Th (^{208}Tl) contamination, gammas, neutrons, muons



Reaching the cosmic **bkg level** of **~1 cts/week/m^2**

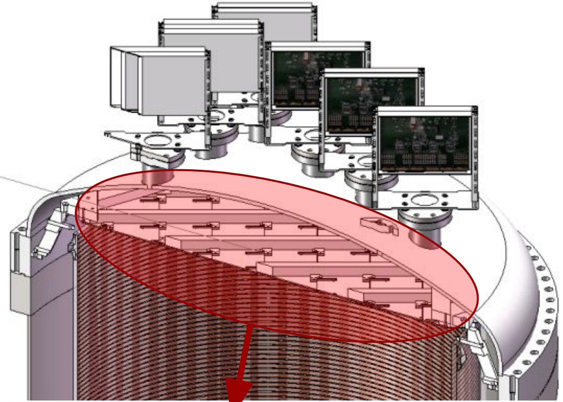
International collaboration:

- China → 7 institutes (lead by SJTU)
- France → CEA Saclay
- Espagne → Zaragoza
- USA → BNL + Maryland University
- Thailand → Nakhon Ratchasima



An Assessment of the Deep Underground Science and Engineering Laboratory (DUSEL)

Xe136 gaseous TPC

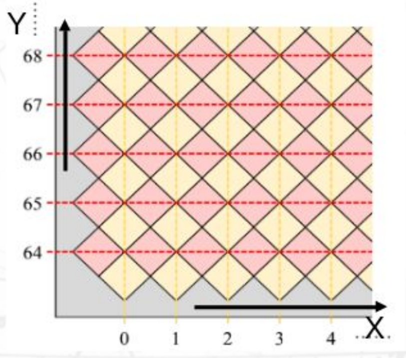


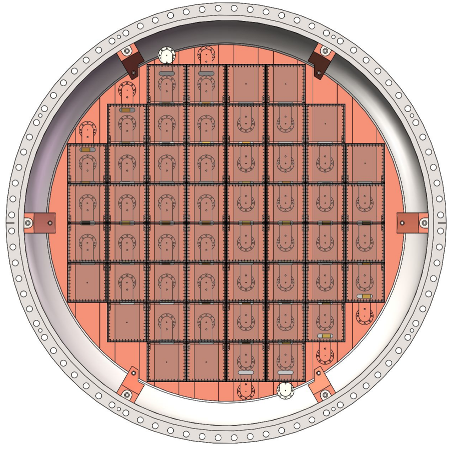
Readout plane

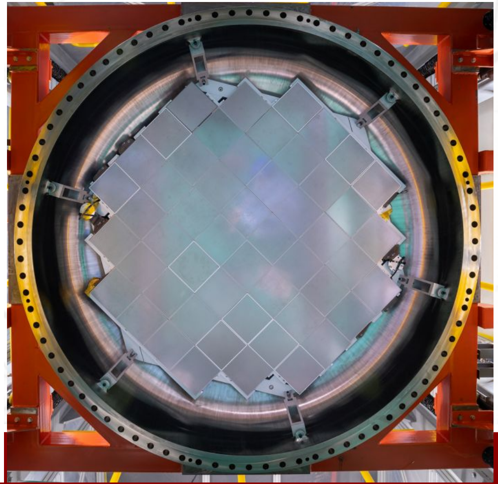
2 x 2D readout plane

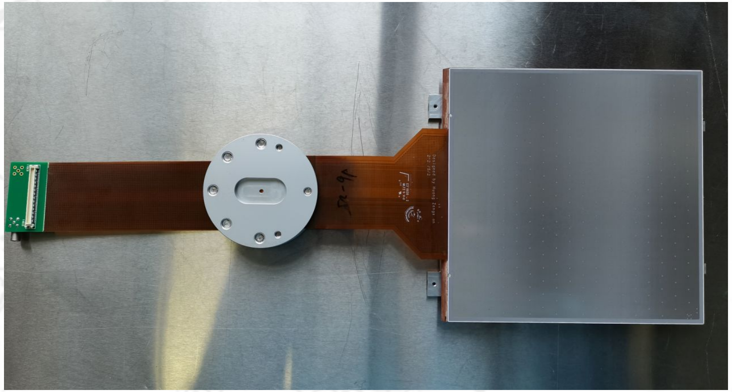
- 52 20x20 cm large Micromegas, 3mm pitch
- *X and Y readout on same board*, 64 channels each
- But not 3D, XZ and YZ read independently

MM schematics



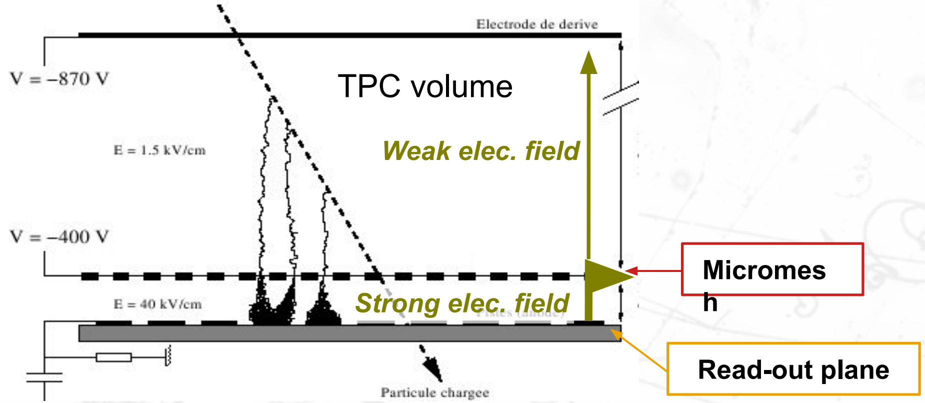






Ni Kaixiang’s thesis: “Searching for Neutrinoless double beta decay of 136Xe in PandaX detectors”

Xe136 gaseous TPC



Charge readout with Micromegas

- Fast gaseous detector
- Ionization and amplification decoupled
- Able to work in high pressure Xenon

Principle and advantages

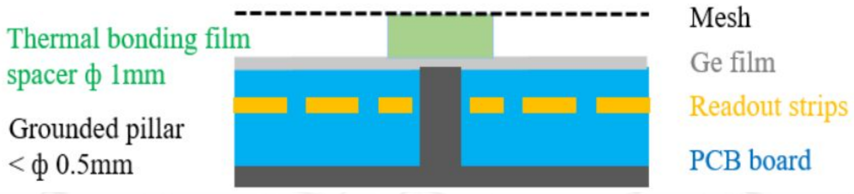
Regular Micromegas with resistive Germanium layer

Mesh spacing by thermo-bonded polyester layer, placed manually

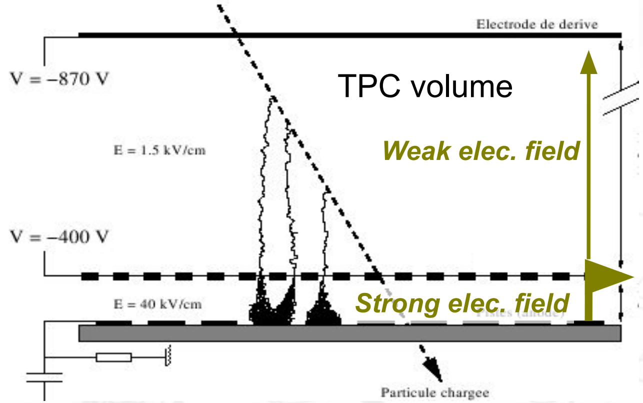
- robust
- low radioactive material
- sparks protection with resistive layer
- good energy resolution expected **~15% at 5.9keV**

Developed and built at USTC (Hefei, China), local chinese production (not linked to CERN)

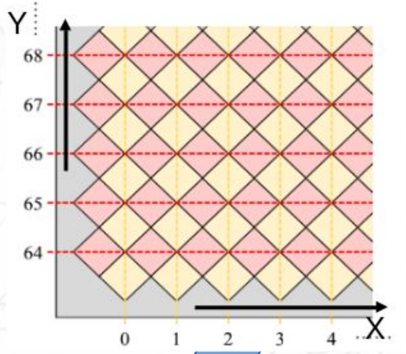
Thermal-bonded Micromegas lateral view



Xe136 gaseous TPC

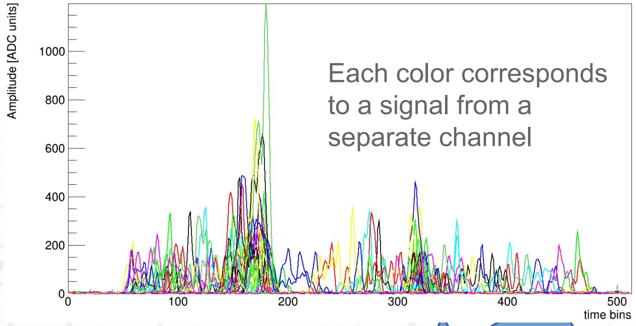


MM schematics

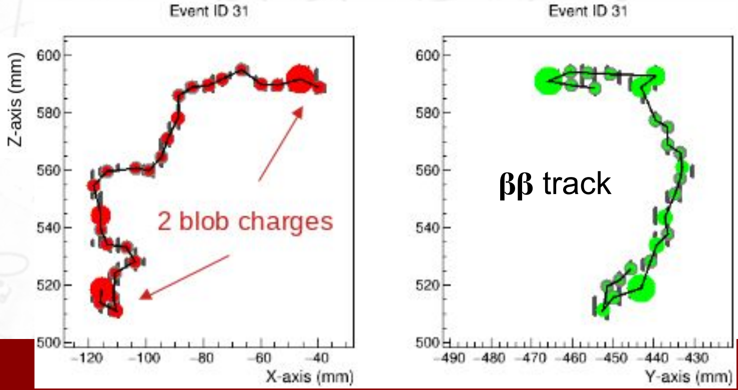


Dimentions: 20cm*20cm

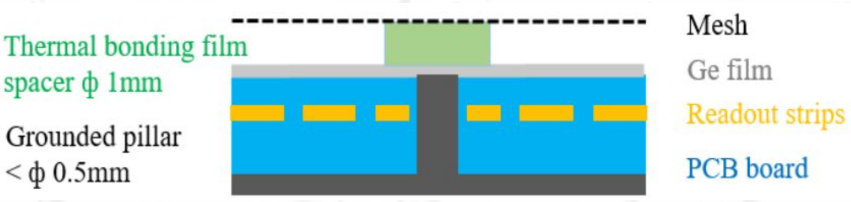
Signal output from the readout system



XZ&YZ projections after data processing of the detector output

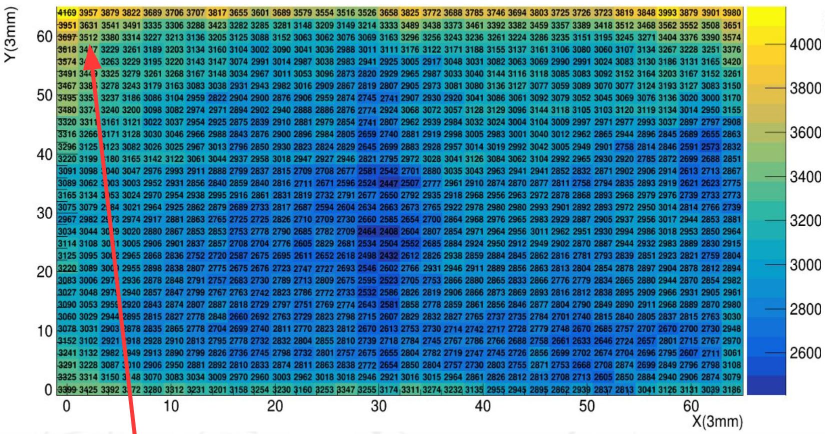


Thermal-bonded Micromegas lateral view

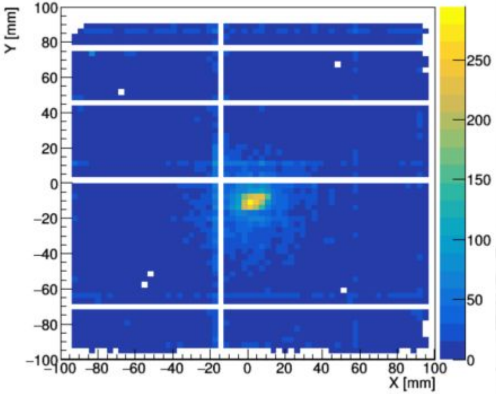


Problems with Micromegas

Gain map for one Micromegas module



Example of missing channels

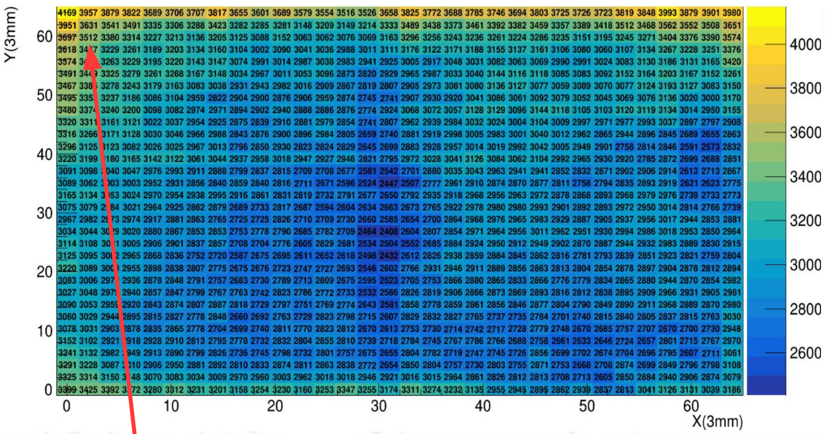


Inhomogeneity of the Gain causes an incorrect energy reconstruction;
PandaX-III prototype showing an inhomogeneity at a level of 6%

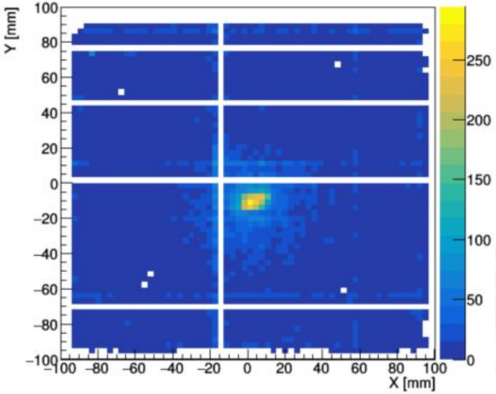
Missing channels cause loss of:
- Topology of the track
- Energy reconstruction info

Problems with Micromegas

Gain map for one Micromegas module

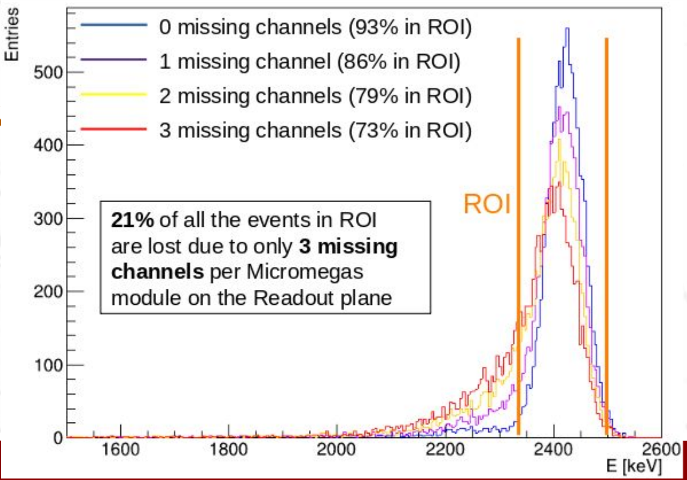


Example of missing channels



Simulation:
~10 000 0nbb events of Xe136.
Q value = **2454 keV**
ROI : **[2357, 2553] keV**

Energy spectrum of the reconstructed $0\nu\beta\beta$ events



Inhomogeneity of the Gain causes an incorrect energy reconstruction;
PandaX-III prototype showing an inhomogeneity at a level of 6%

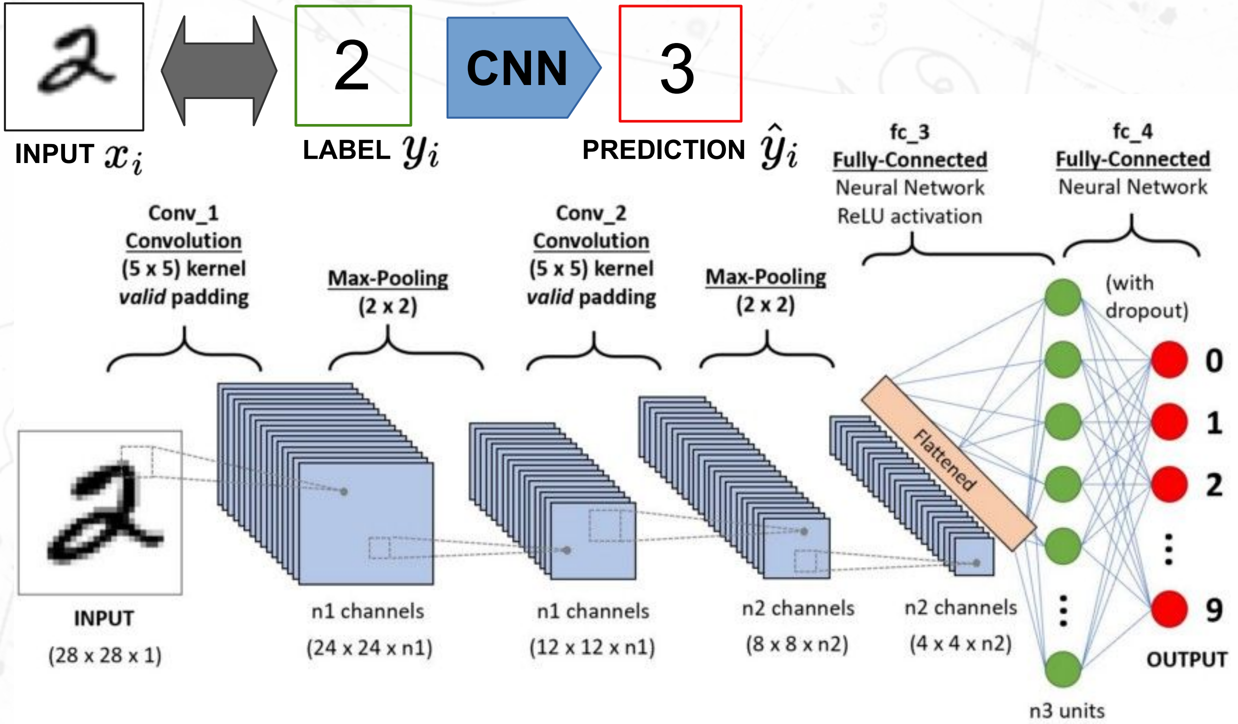
Missing channels cause loss of:
- Topology of the track
- Energy reconstruction info

Introduction

PyTorch

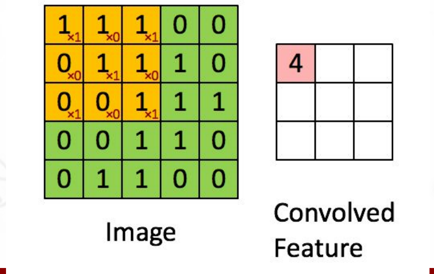
The problem of missing channels should be addressed:

- Proposition to use Machine Learning techniques
- (Convolutional Neural Network in particular) to reconstruct missing energy



Loss: estimate of an error b/w label and predicted $f(y_i, \hat{y}_i)$

Backpropagation algorithm (computing gradient of the loss wrt weights and biases) → updating weights and biases → minimizing loss



Preliminary event selection

Monte-Carlo simulations in **REST** software:

- **Electrons (1 e- per event)** generated in Gas volume of the PandaX-III 140kg setup
- Energy range: [500 ; 3500] keV
- Isotropic angular distribution

Primary: initial energy of the electron

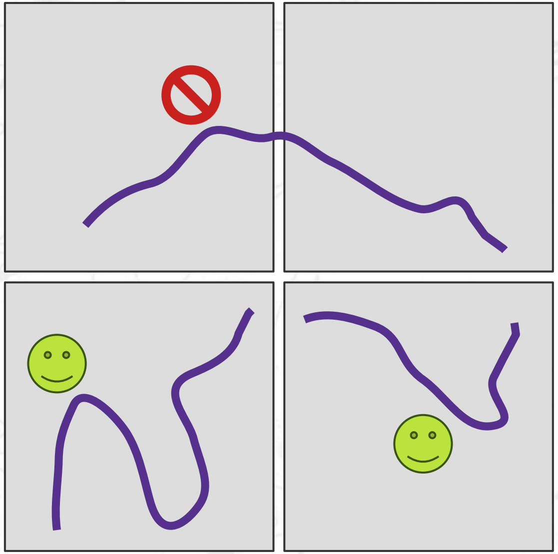
Detected: total registered energy of the event by the readout system without missing channels including detection effects (*gain fluctuations*)

Z_distance: perpendicular distance from the electron generation vertex in the gas volume to the readout plane

Data selection:

- Electron events registered only by one MM module
- (spatial selection)
- Energy cut for residual b/w primary & detected E
- ($E_{PRIMARY}$ - $E_{DETECTED}$= [-142, 178] keV)

Energy cut to remove events with gamma generation during electron propagation(Bremsstrahlung, Excitation)

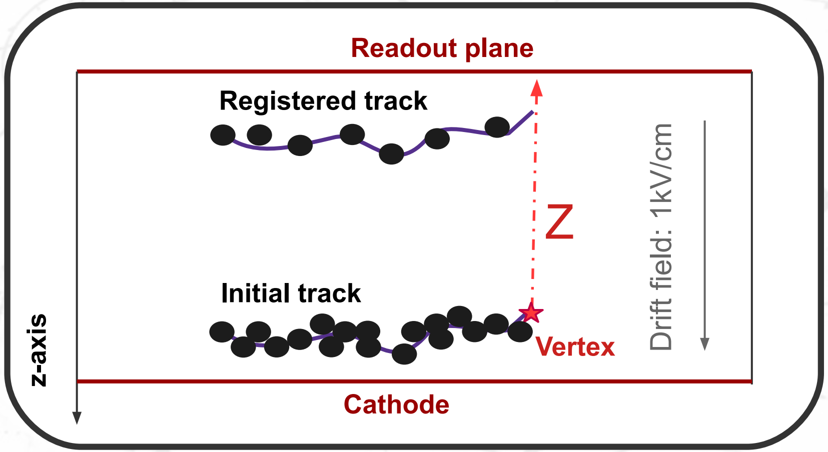


Preliminary event selection

Monte-Carlo simulations in **REST** software:

- **Electrons (1 e- per event)** generated in Gas volume of the PandaX-III 140kg setup
- Energy range: [500 ; 3500] keV
- Isotropic angular distribution

Primary: initial energy of the electron
Detected: total registered energy of the event by the readout system (without missing channels)
Z_distance: perpendicular distance from the electron generation vertex in the gas volume to the readout plane

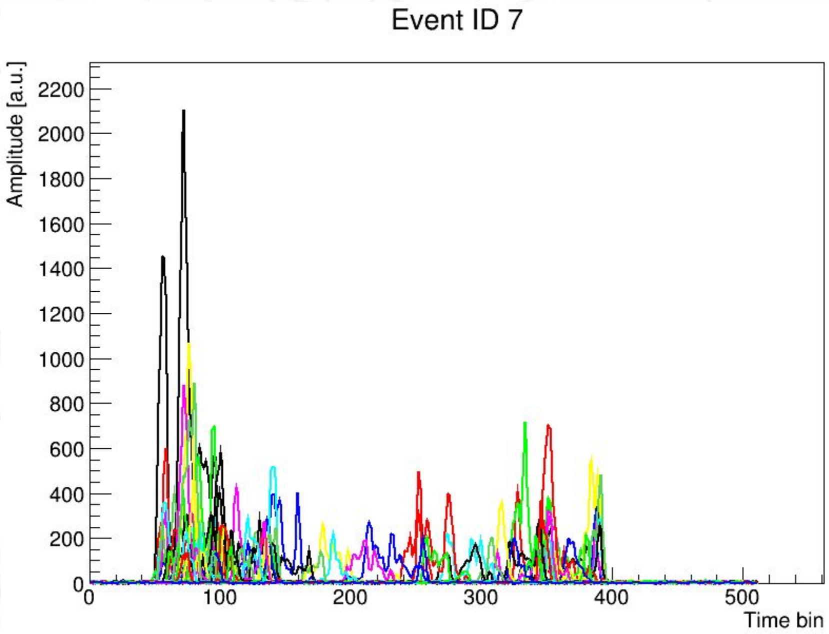


A portion of the ionization electrons will be reabsorbed due to the gas impurities during the drifting
→ “electron attachment effect” characterized by the electron lifetime
→ worsens energy resolution

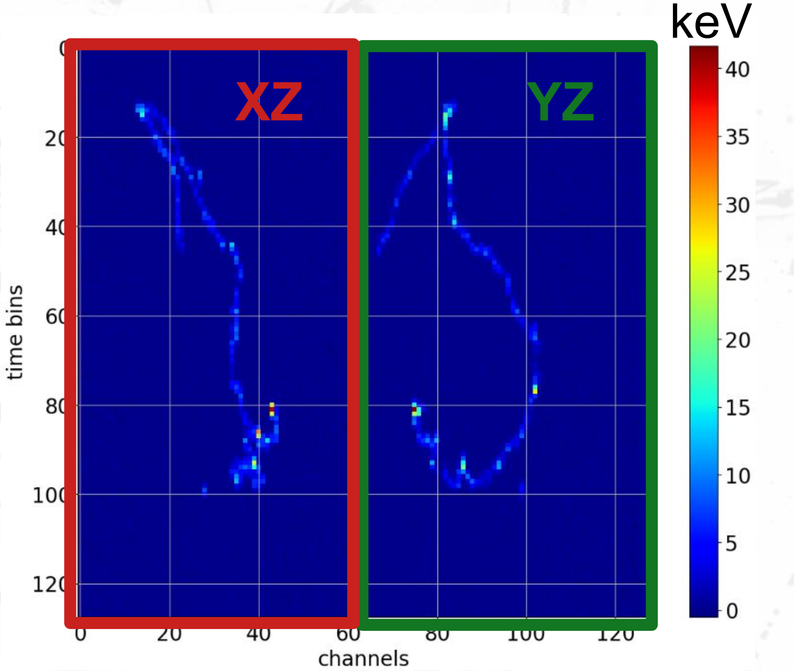
Estimation of Z from electron diffusion

INPUT DATA

Signal from the readout system



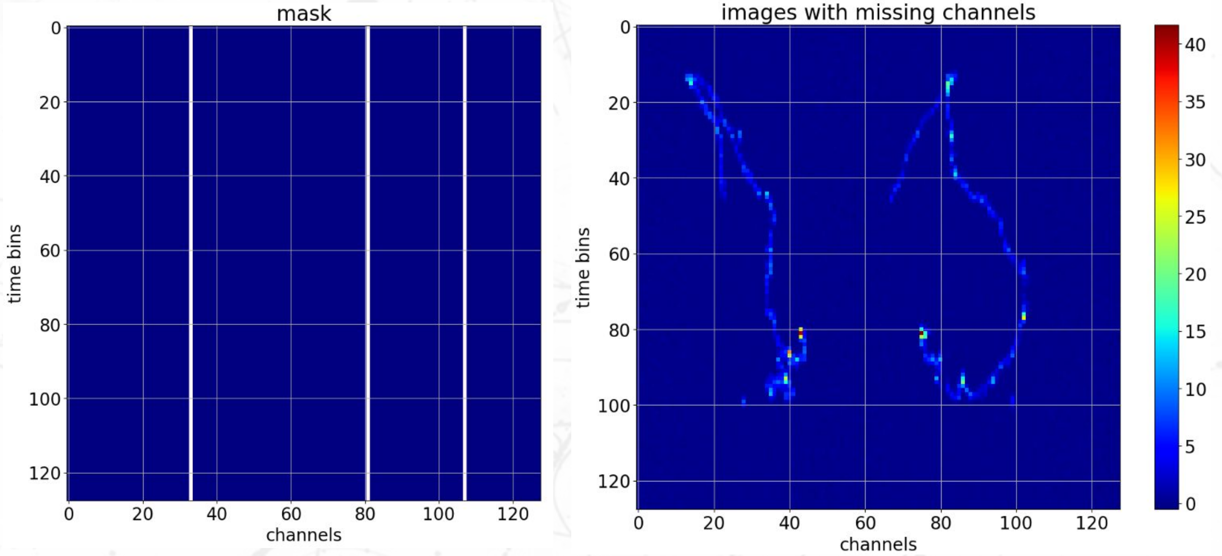
- Converted into XZ and YZ 2D arrays
- Time bins rebinned by the factor of 4
- Concatenated into one array of shape (128, 128)



INPUT DATA

3 missing channels per MM module

Total N events: 360 000

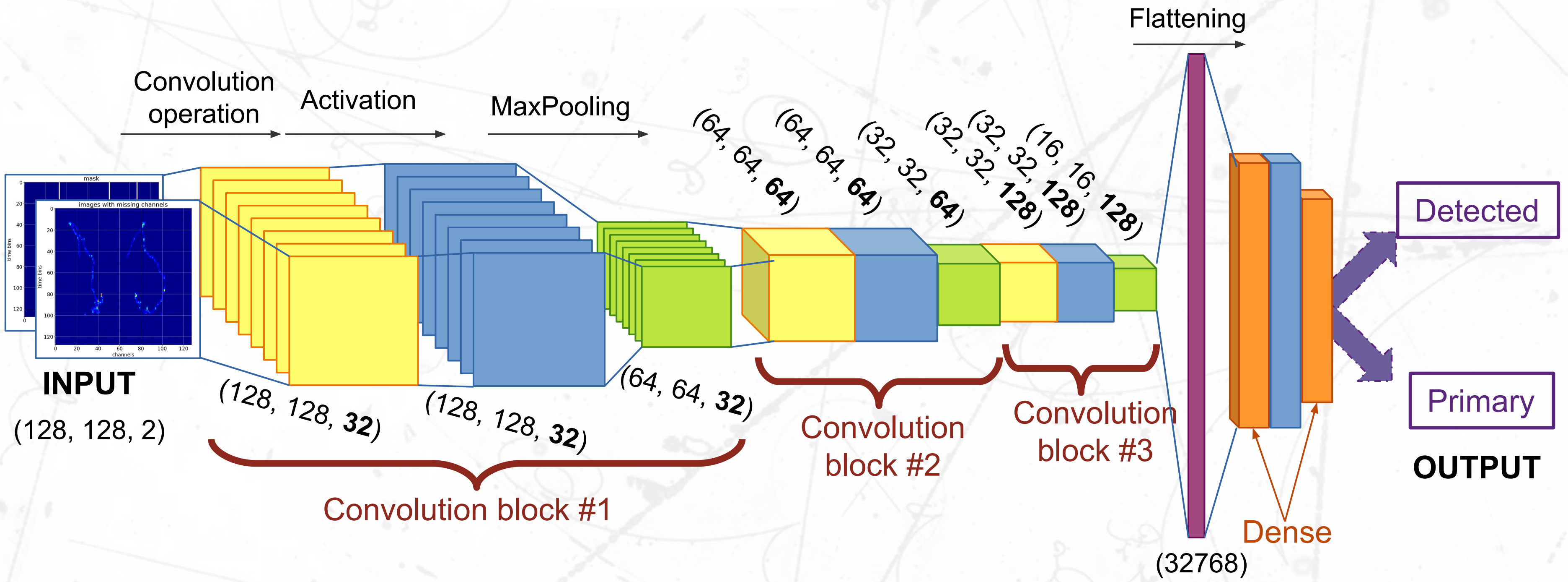


60% - Training
20% - Validation
20% - Test

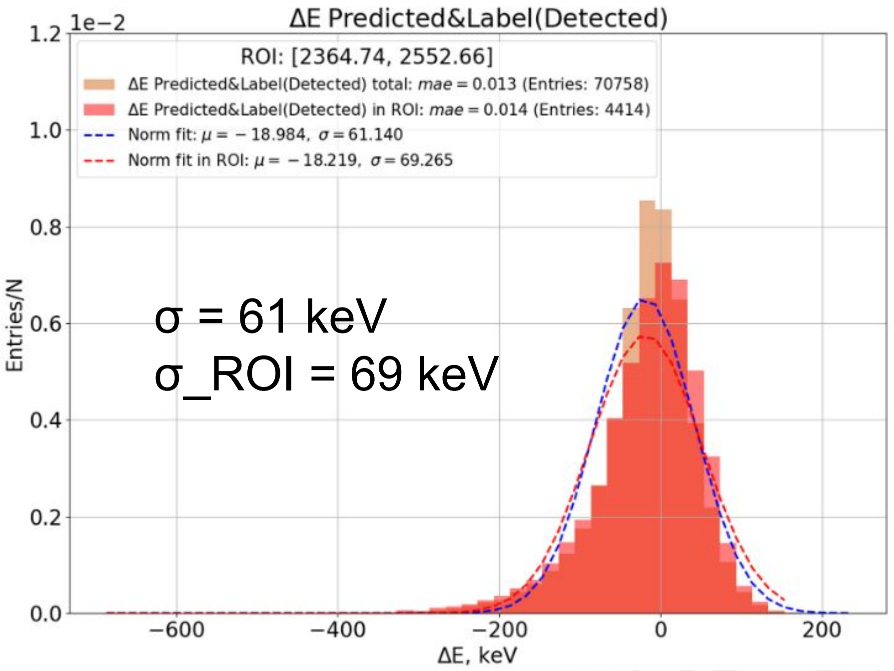
Concatenated: array (128, 128, 2)

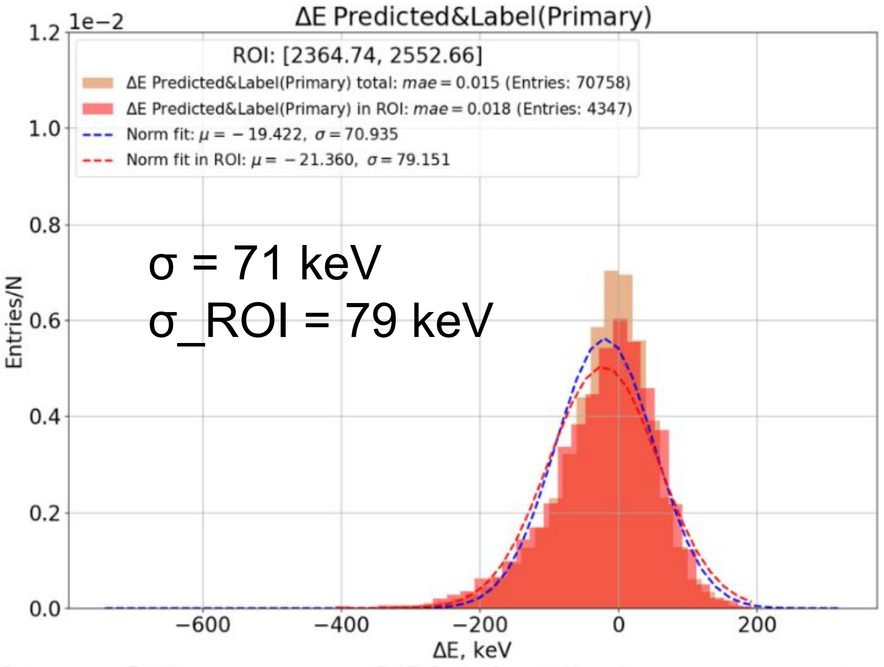
Visual Geometry Group (VGG)

Loss function: Mean Squared Error $\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$



Visual Geometry Group (VGG)

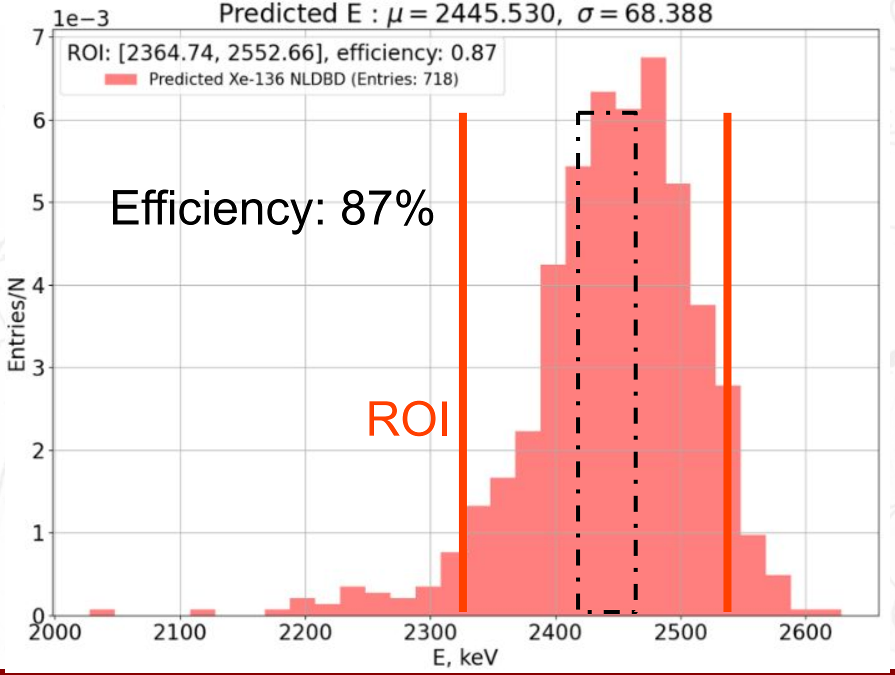




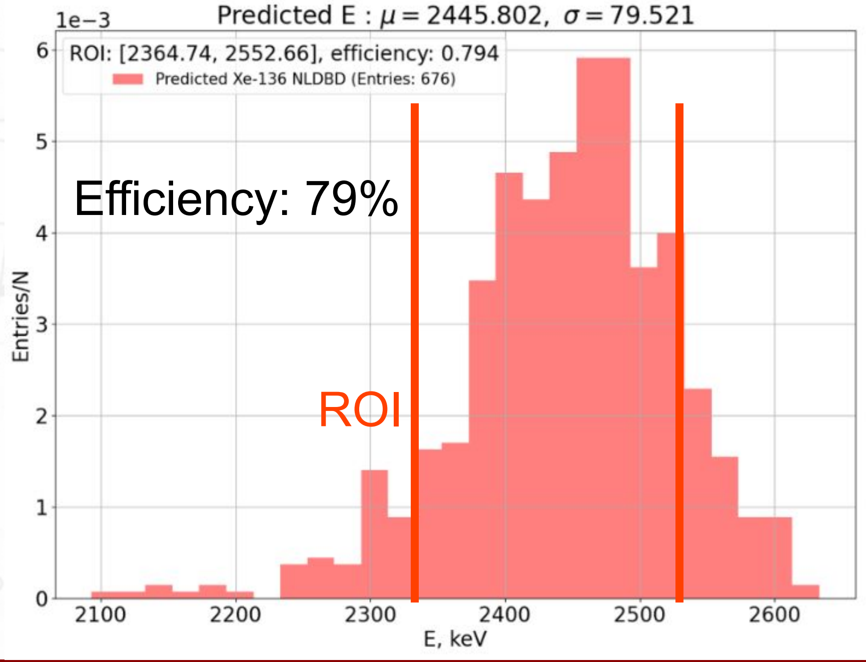
Predictions for Detected and Primary energies in ROI

Events with labels in the ±15 keV region around Qbb for Xe136 – to imitate 0NDBD
Efficiency computed regarding these events: Predicted in ROI / Labels in ROI * 100%

Prediction: Detected

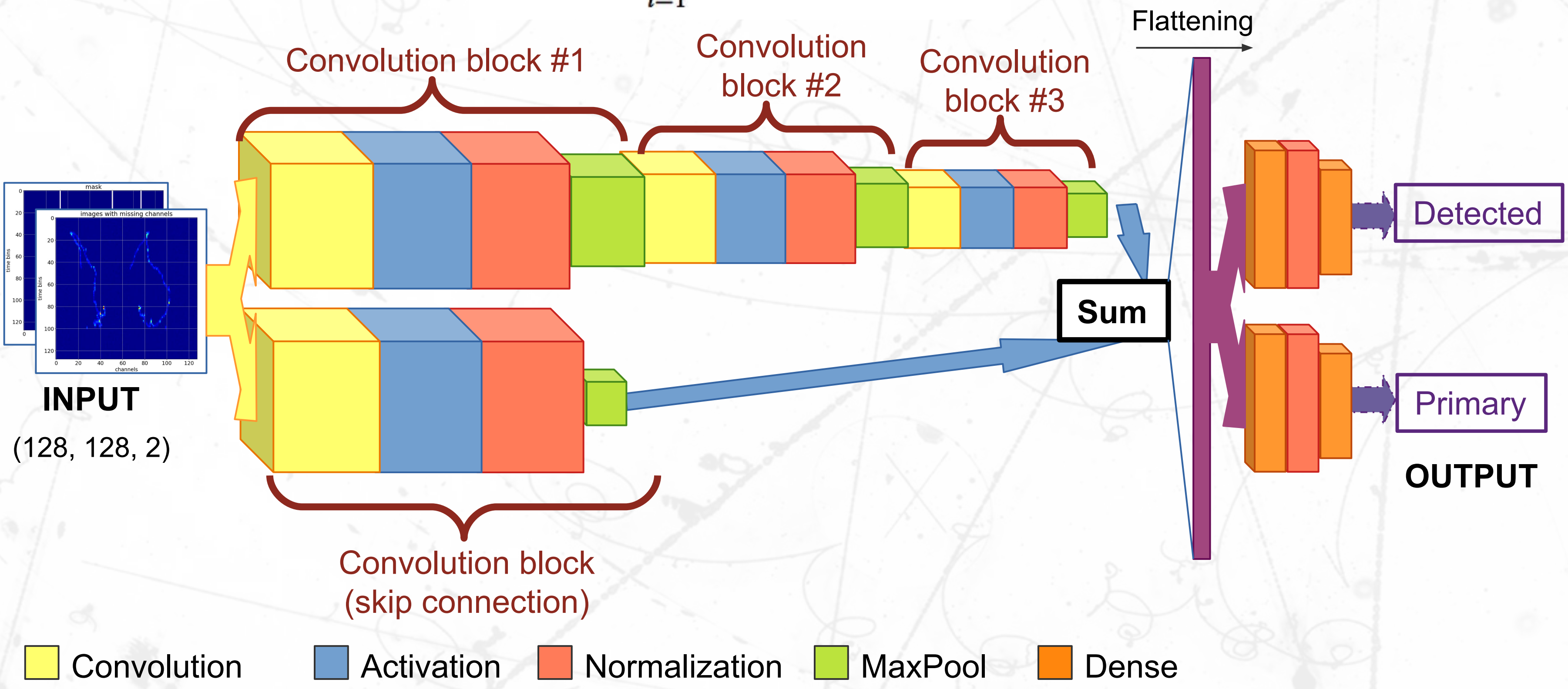


Prediction: Primary

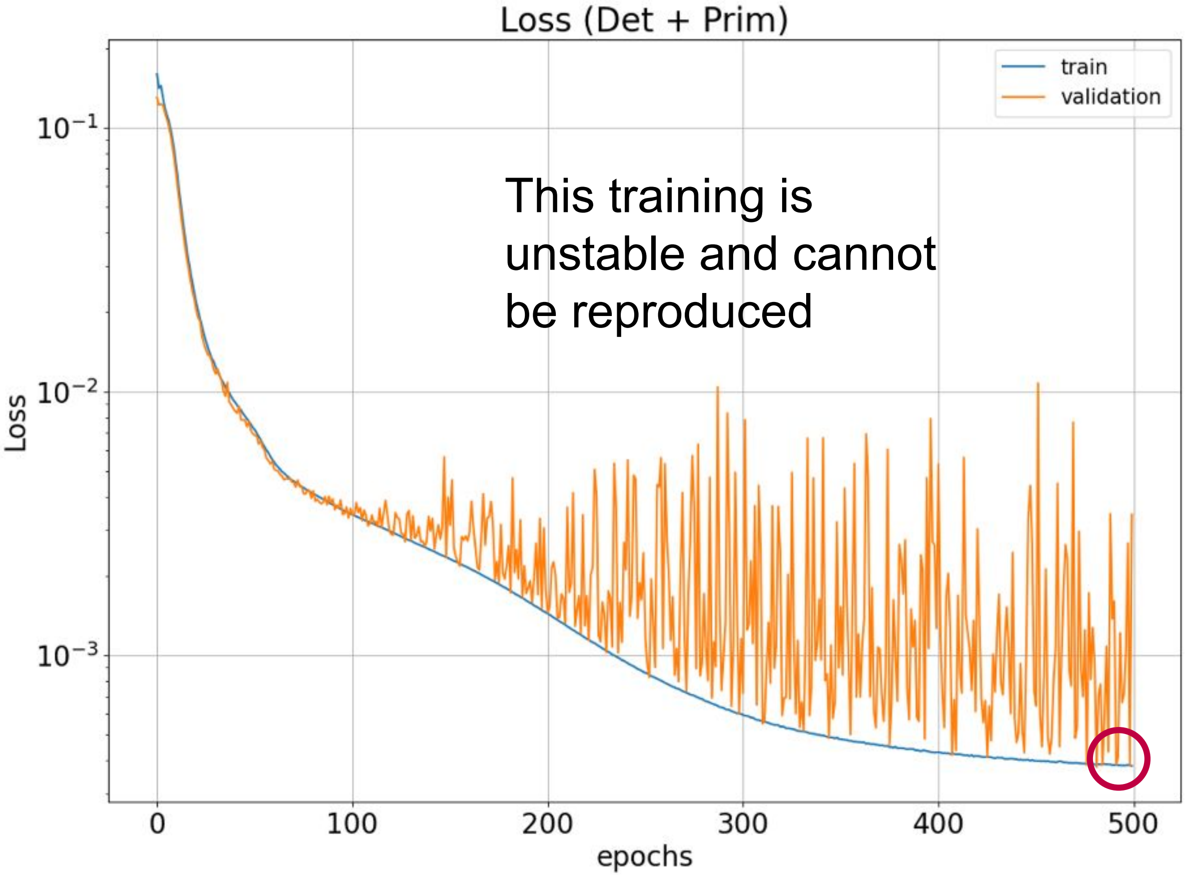


CNN ResNet with Convolution block as skip connection

Loss function: LogCosh loss $L(y, y^p) = \sum_{i=1}^{n} \log(\cosh(y_i^p - y_i))$



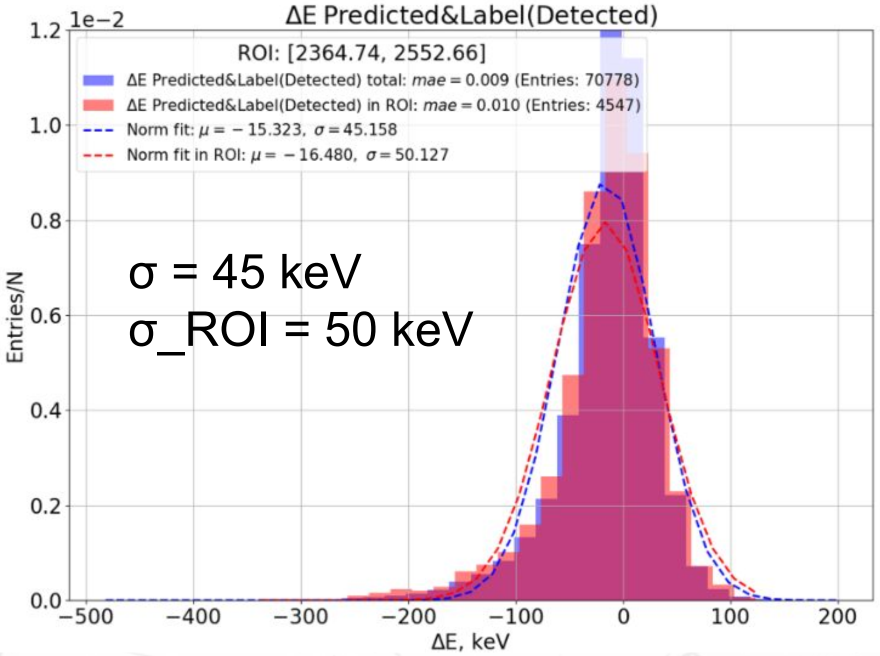
CNN ResNet with Convolution block as skip connection

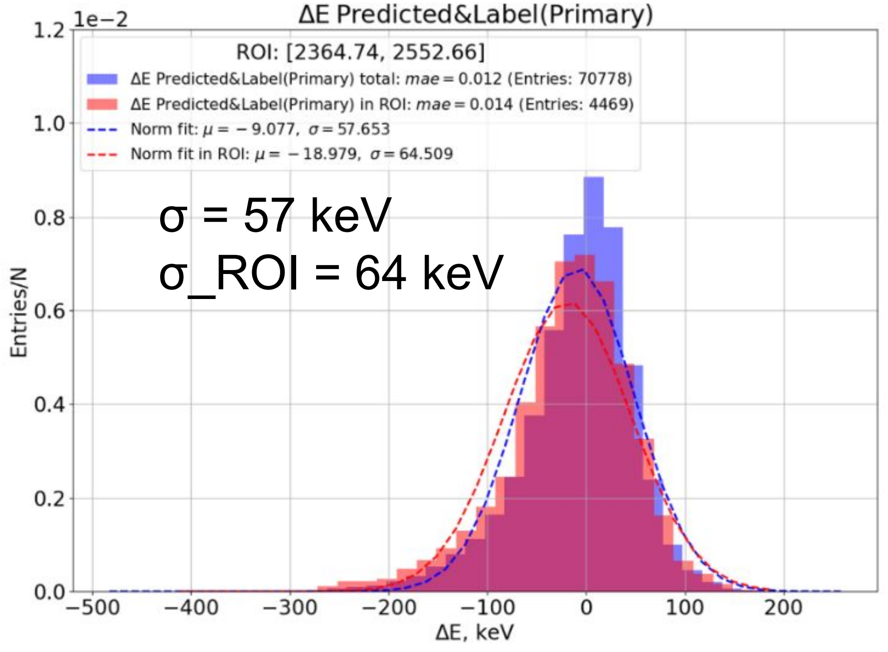


This training is unstable and cannot be reproduced

Epoch #499 was considered for prediction

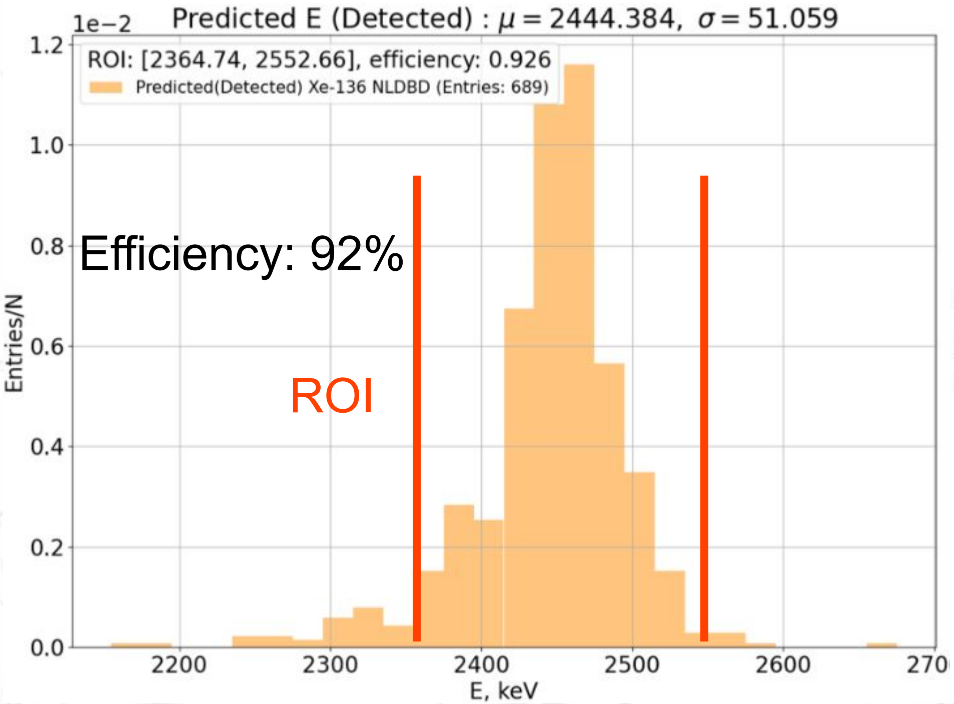
CNN ResNet with Convolution block as skip connection



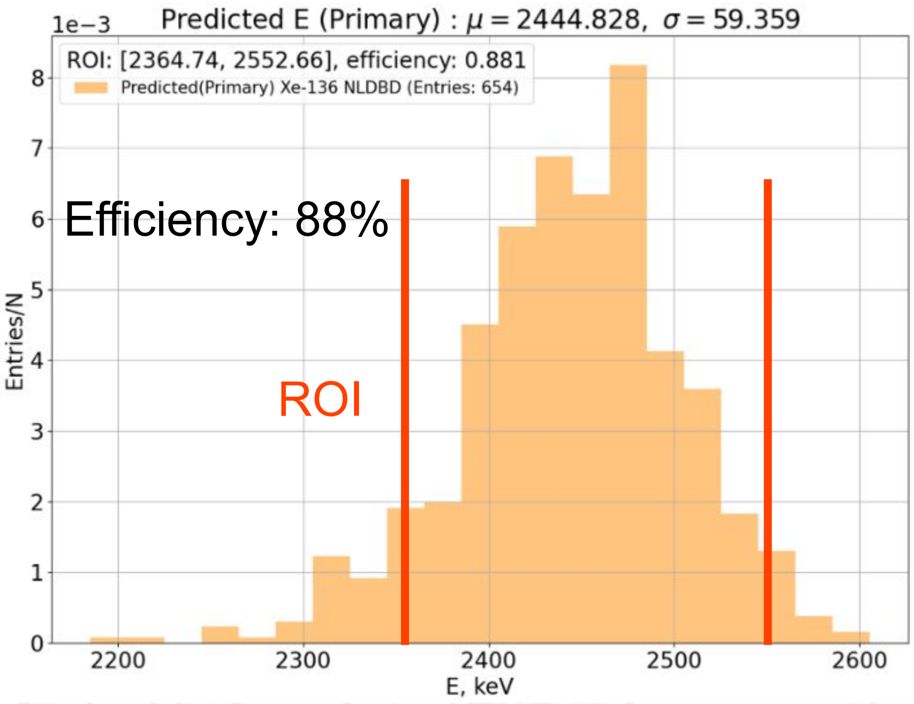


CNN ResNet with Convolution block as skip connection

Prediction: Detected

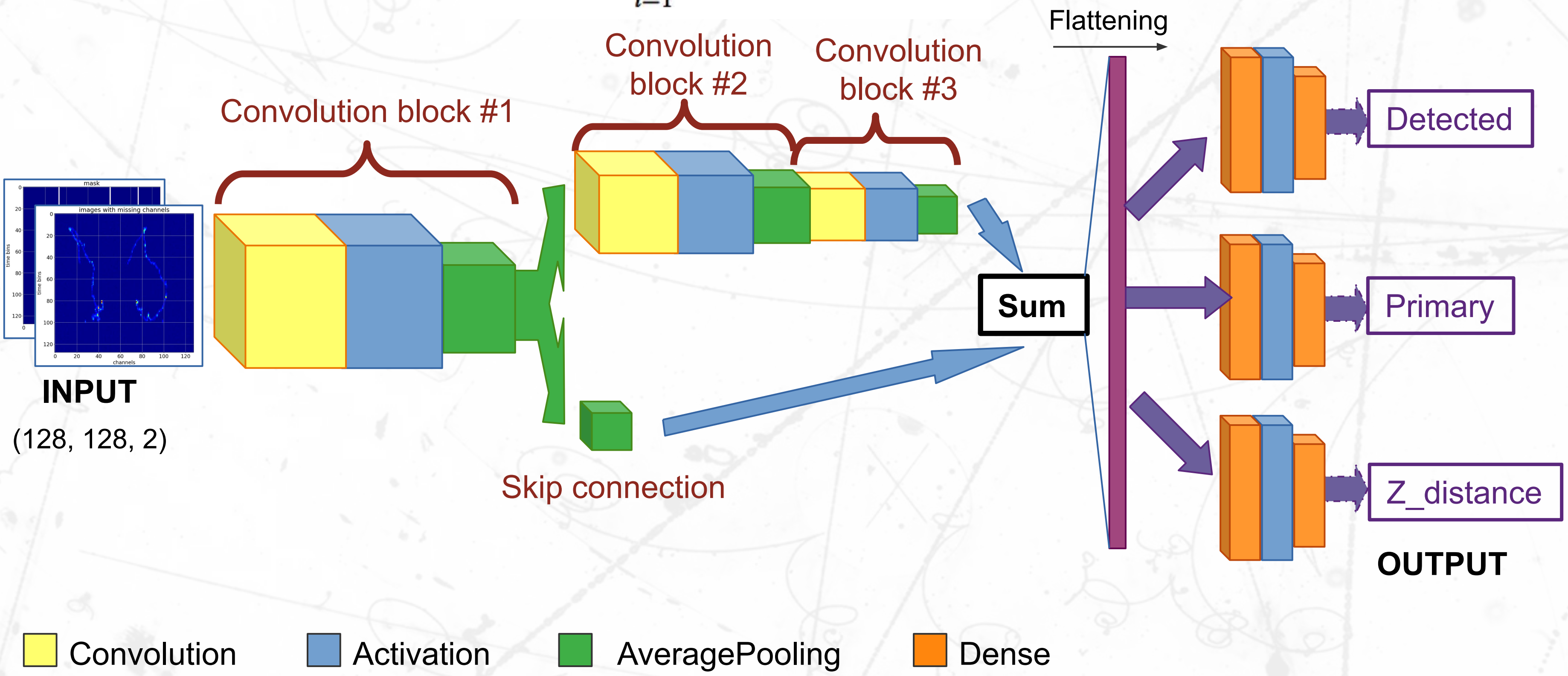


Prediction: Primary

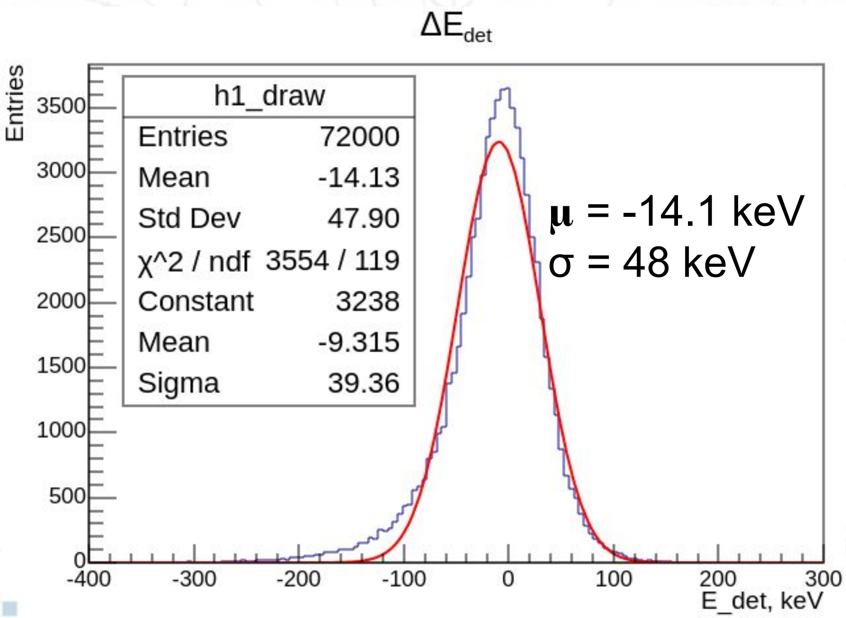


CNN ResNet with Z distance prediction

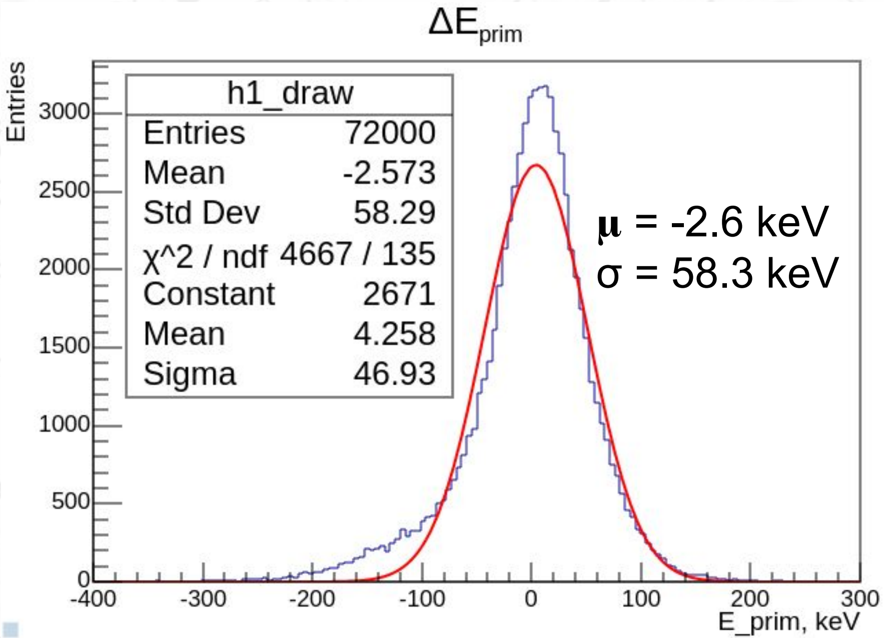
Loss function: LogCosh loss $L(y, y^p) = \sum_{i=1}^{n} \log(\cosh(y_i^p - y_i))$



CNN ResNet with Z distance prediction

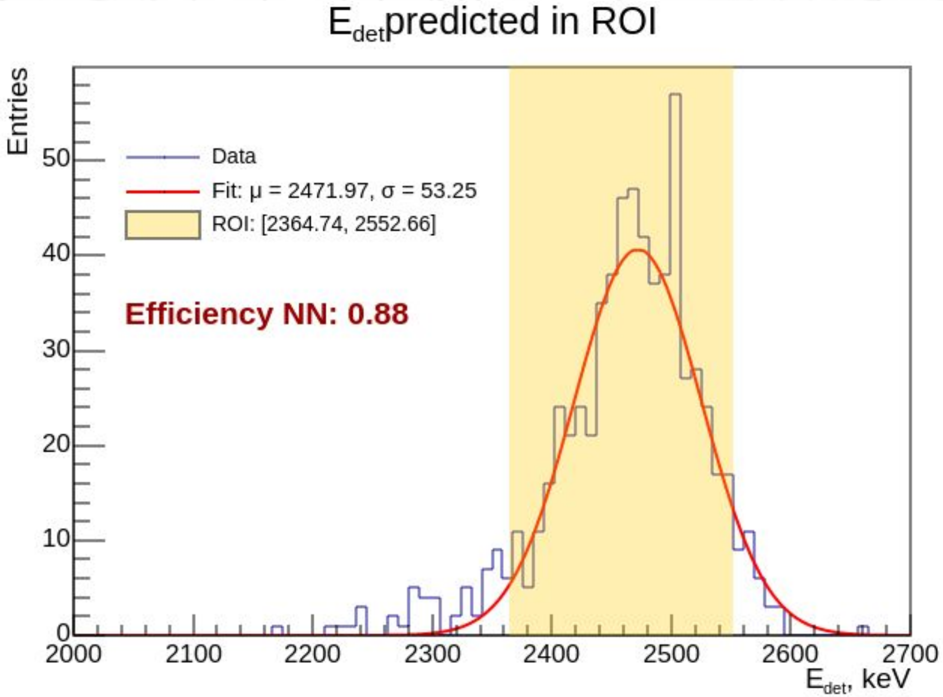


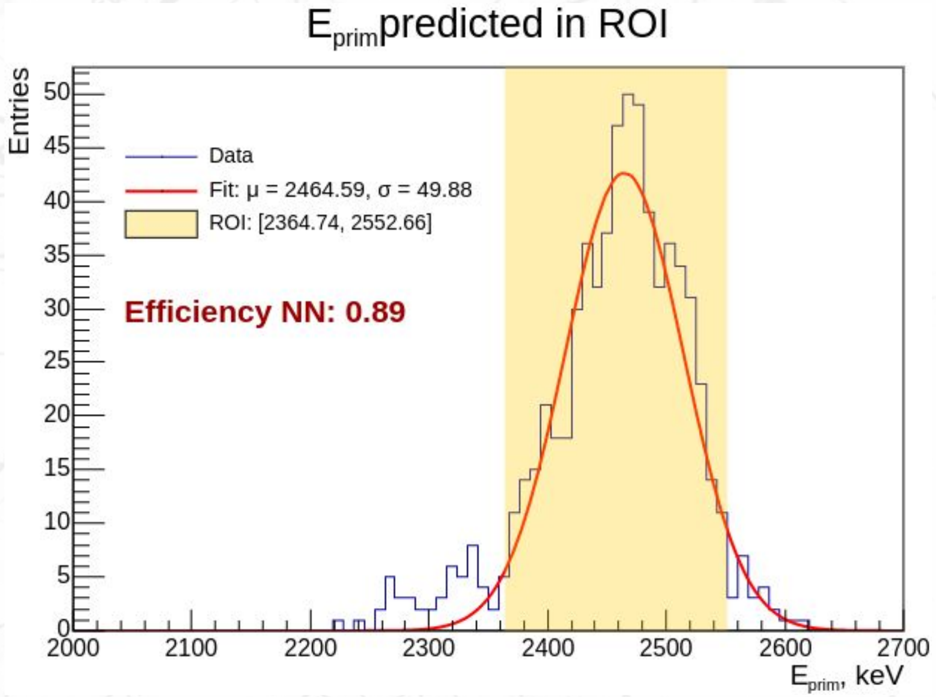
μ = -14.1 keV
σ = 48 keV



μ = -2.6 keV
σ = 58.3 keV

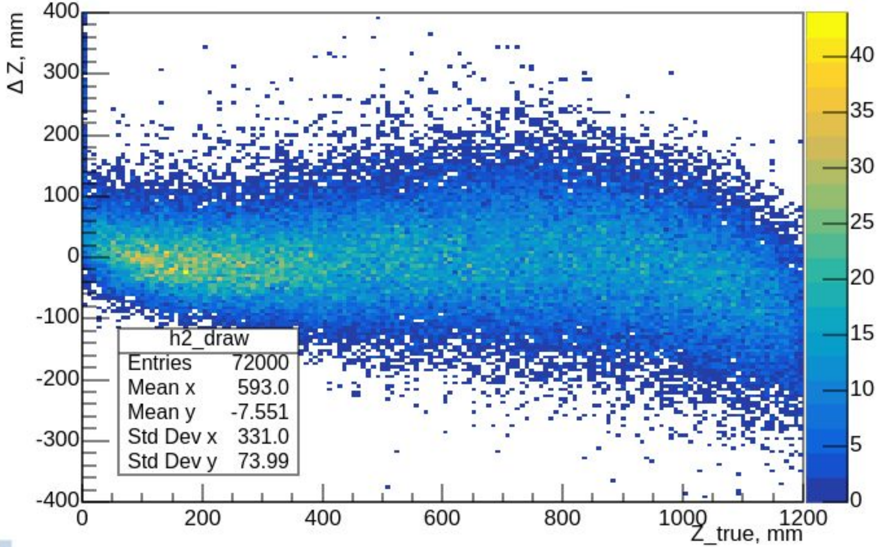
CNN ResNet with Z distance prediction



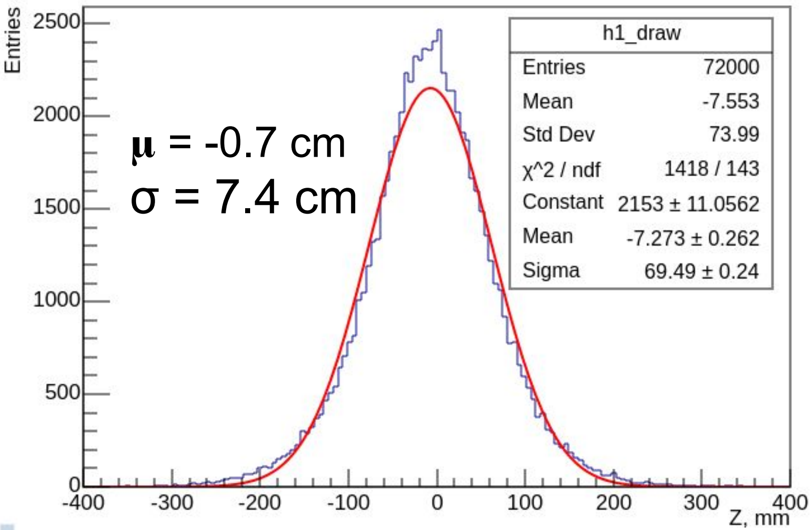


CNN ResNet with Z distance prediction

ΔZ correlation map



ΔZ



μ = -0.7 cm
σ = 7.4 cm

Conclusion and Prospects

	$\sigma_{Detected}$ $\sigma_{Primary}$	$\sigma_{ROI(Detected)}$ $\sigma_{ROI(Primary)}$	$\text{Efficiency}_{Detected}$ $\text{Efficiency}_{Primary}$
CNN VGG	61 keV 71 keV	69 keV 79 keV	87% 79%
CNN + ResNet (UNSTABLE)	45 keV 57 keV	50 keV 64 keV	92% 88%
CNN + ResNet with Z	48 keV 58 keV	- -	88% 89%

Conclusion and Prospects

The work is still in progress and results are preliminary

- Ongoing process to find the most optimal architecture and appropriate configuration of hyper-parameters
- Classification of the event type (0NDBD/gamma) to improve bkg/signal discrimination is under way
- Reconstruction of the absolute Z position is ongoing, yet already promising
- Events registered by multiple Micromegas will be added as inputs
- Inhomogeneous gain of the micromegas will be simulated

For a single weight (w_jk)^l, the gradient is:

$$\frac{\partial C}{\partial w_{jk}^{l}} = \frac{\partial C}{\partial z_j^l} \frac{\partial z_j^l}{\partial w_{jk}^l} \qquad \textit{chain rule}$$

$$z_j^l = \sum_{k=1}^{m} w_{jk}^l a_k^{l-1} + b_j^l \qquad \textit{by definition}$$

$$m \;-\; \textit{number of neurons in } l-1 \textit{ layer}$$

$$\frac{\partial z_j^l}{\partial w_{jk}^l} = a_k^{l-1} \qquad \textit{by differentiation (calculating derivative)}$$

$$\frac{\partial C}{\partial w_{jk}^l} = \frac{\partial C}{\partial z_j^l} a_k^{l-1} \qquad \textit{final value}$$

Similar set of equations can be applied to *(b_j)^l*:

$$\frac{\partial C}{\partial b_j^l} = \frac{\partial C}{\partial z_j^l} \frac{\partial z_j^l}{\partial b_j^l} \qquad \textit{chain rule}$$

$$\frac{\partial z_j^l}{\partial b_j^l} = 1 \qquad \textit{by differentiation (calculating derivative)}$$

$$\frac{\partial C}{\partial b_j^l} = \frac{\partial C}{\partial z_j^l} 1 \qquad \textit{final value}$$

Equations for derivative of C in a single bias (b_j)^l

Adam optimizer

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$$

where **m** is the decaying average of past gradients and **v** is the decaying average of past squared gradients.

Thus we work on these biases by computing bias-corrected terms:

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t}$$

$$\hat{v}_t = \frac{v_t}{1-\beta_2^t}$$

Adam (Adaptive Moment Estimation):

$$\theta_t = \theta_{t-1} - \frac{\eta}{\sqrt{\hat{v}_t + \epsilon}} \hat{m}_t$$

Unscaled input variables can result in a slow or unstable learning process, whereas unscaled target variables on regression problems can result in exploding gradients causing the learning process to fail.

Input arrays and labels are normalized

the distribution of the inputs to layers deep in the network may change after each mini-batch when the weights are updated. This can cause the learning algorithm to forever chase a moving target. This change in the distribution of inputs to layers in the network is referred to the technical name “internal covariate shift.”

Batch normalization is a technique used in neural networks to standardize the input data of each layer. It involves adjusting and scaling the activations of the previous layer by subtracting the mean and dividing by the standard deviation of the batch of data being processed. This helps to alleviate issues with internal covariate shift and allows the model to converge more quickly during training. Additionally, it has been shown to improve the generalization of the model by reducing overfitting. Batch normalization is commonly used in deep neural networks and is often applied after convolutional or fully connected layers.

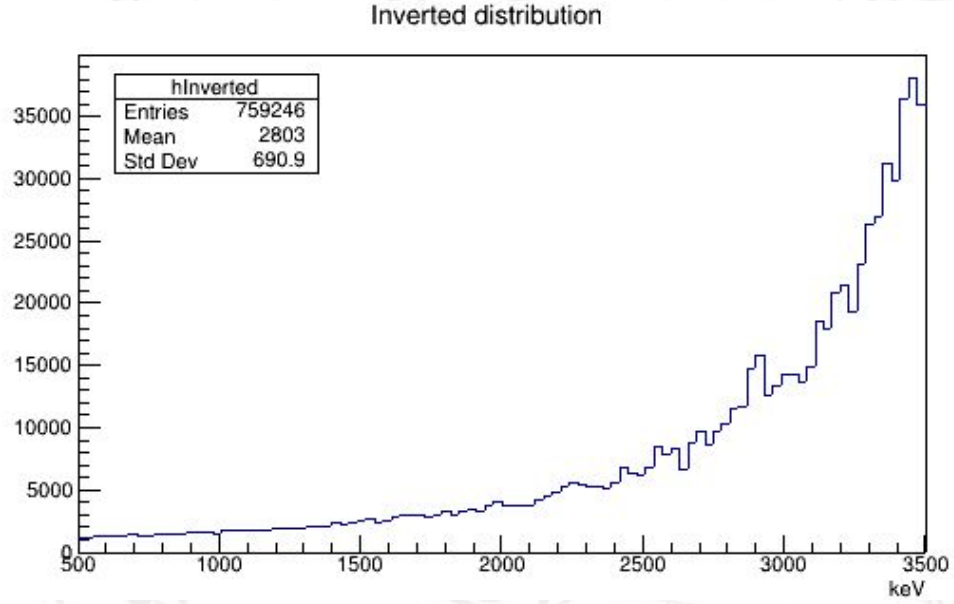
Preliminary event selection

Simulations in REST (v2.3.12):

- **Electrons** generated in Gas volume of the PandaX-III 140kg setup
- Isotropic angular distribution

Primary: initial energy of the electron in restG4

Primary energy distribution [500, 3500] keV



Data selection:

- Electron events registered only by one MM module
- (spatial selection)
- Energy cut for residual b/w primary & detected E
- (ΔE = [-142, 178] keV)

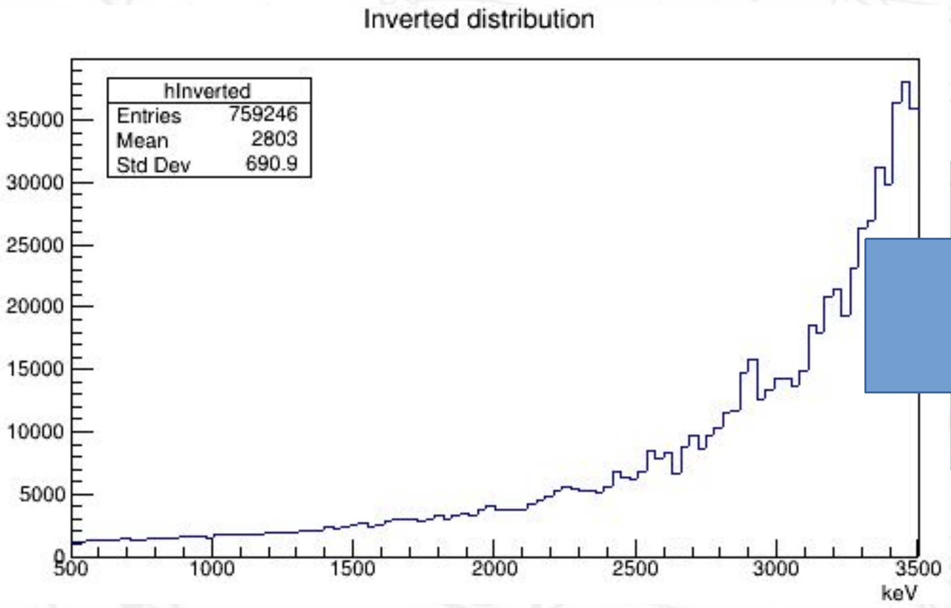
Energy cut to remove events with gamma generation during electron propagation(Bremsstrahlung, Excitation)

Preliminary event selection

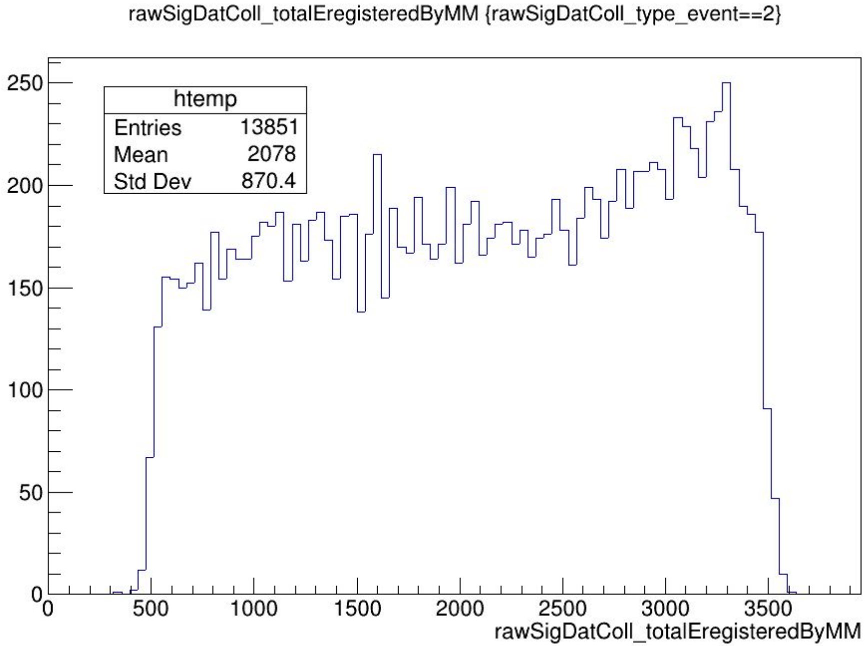
Simulations in REST (v2.3.12):

- Electrons generated in Gas volume of the PandaX-III 140kg setup
- Isotropic angular distribution

Primary energy distribution [500, 3500] keV

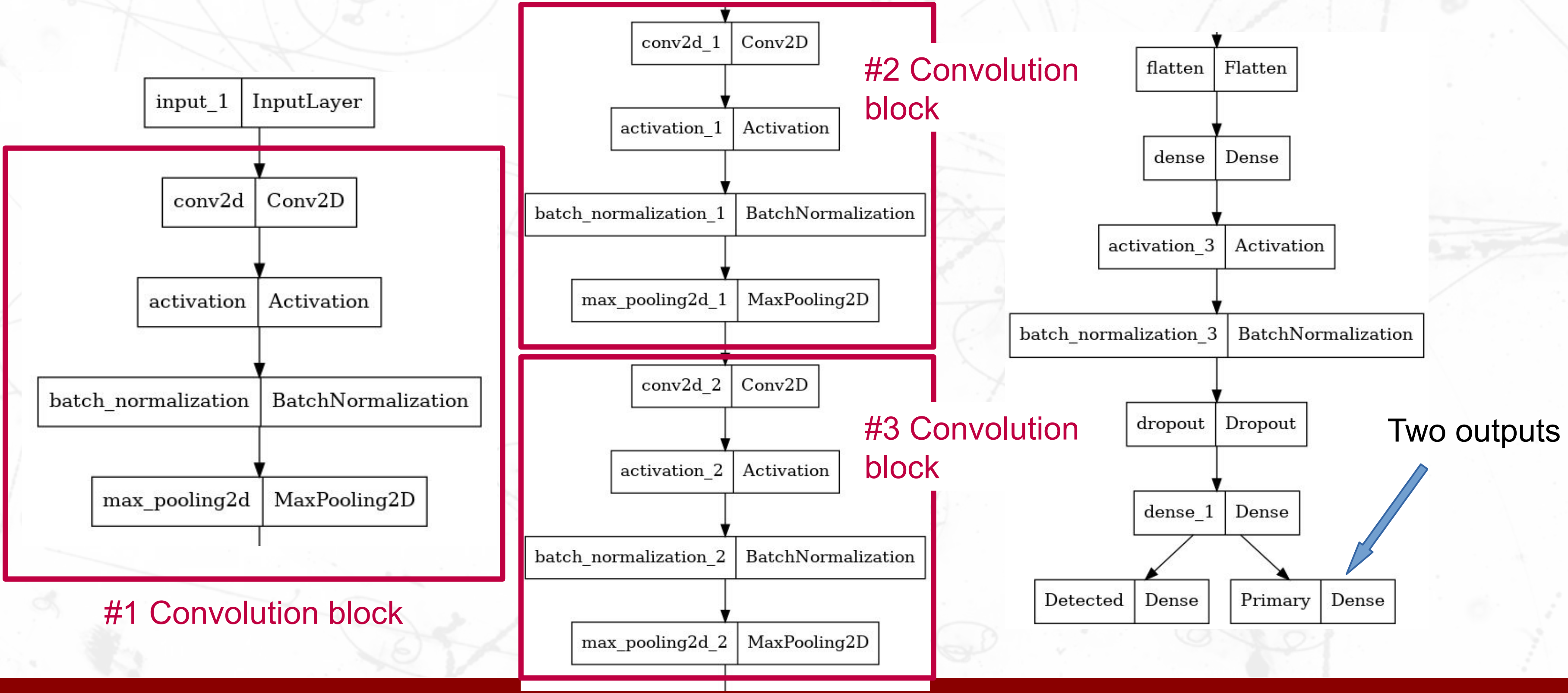


Energy spectrum after selection



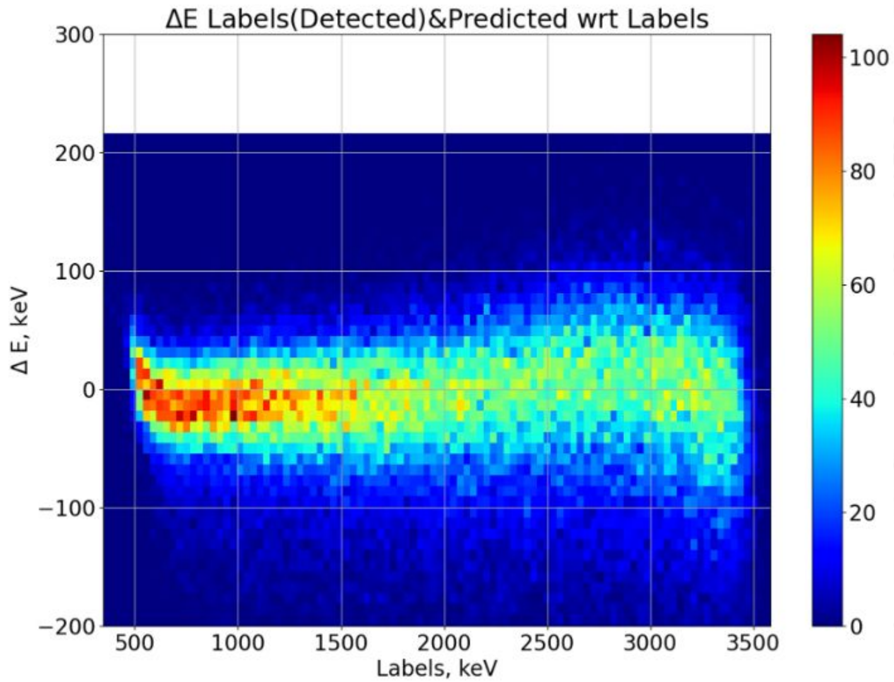
Visual Geometry Group (VGG)

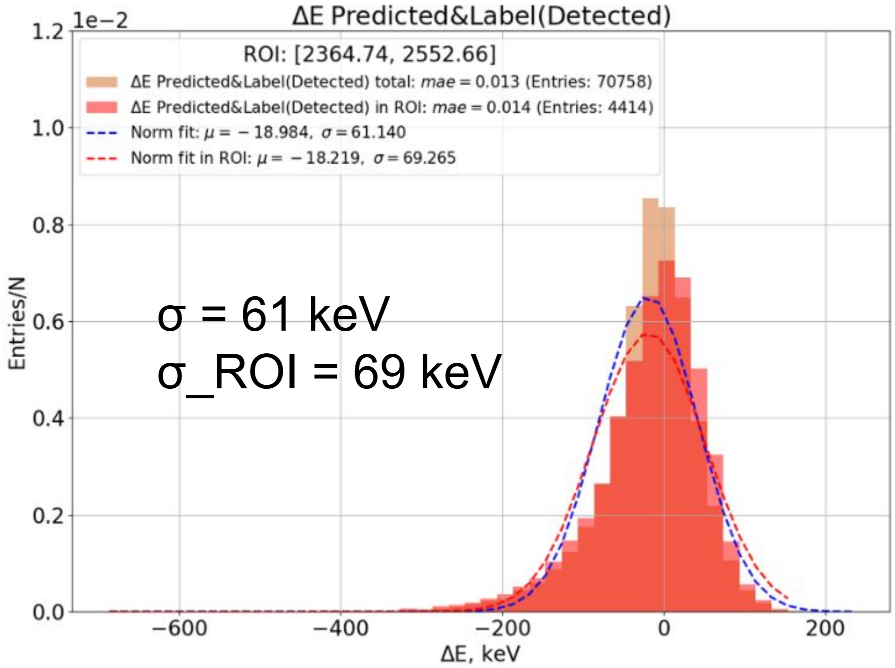
Loss function: Mean Squared Error $\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$



Visual Geometry Group (VGG)

Label: Detected energies



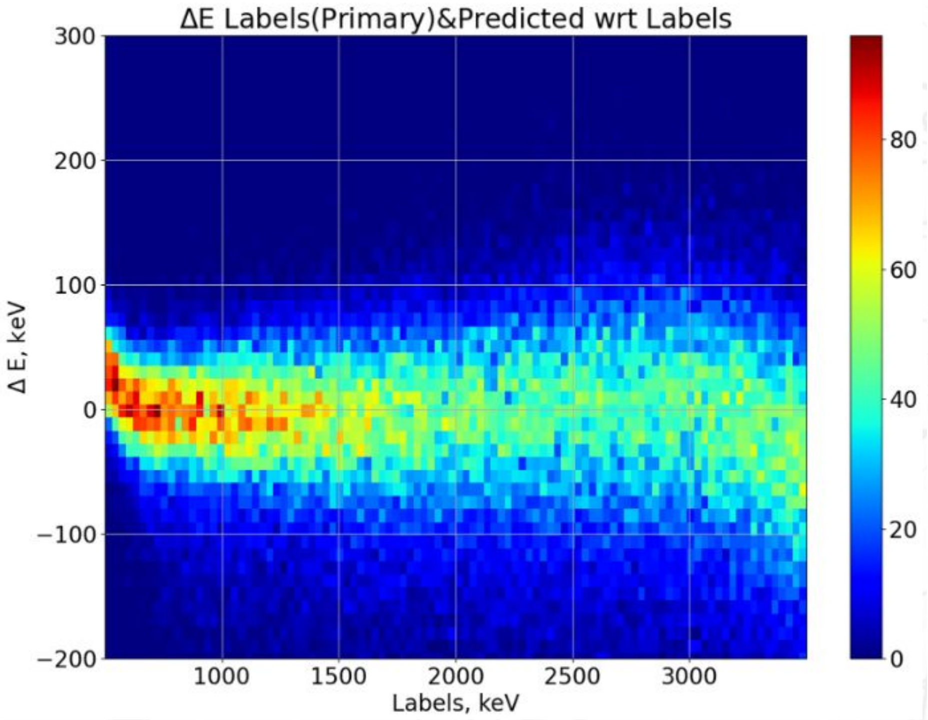


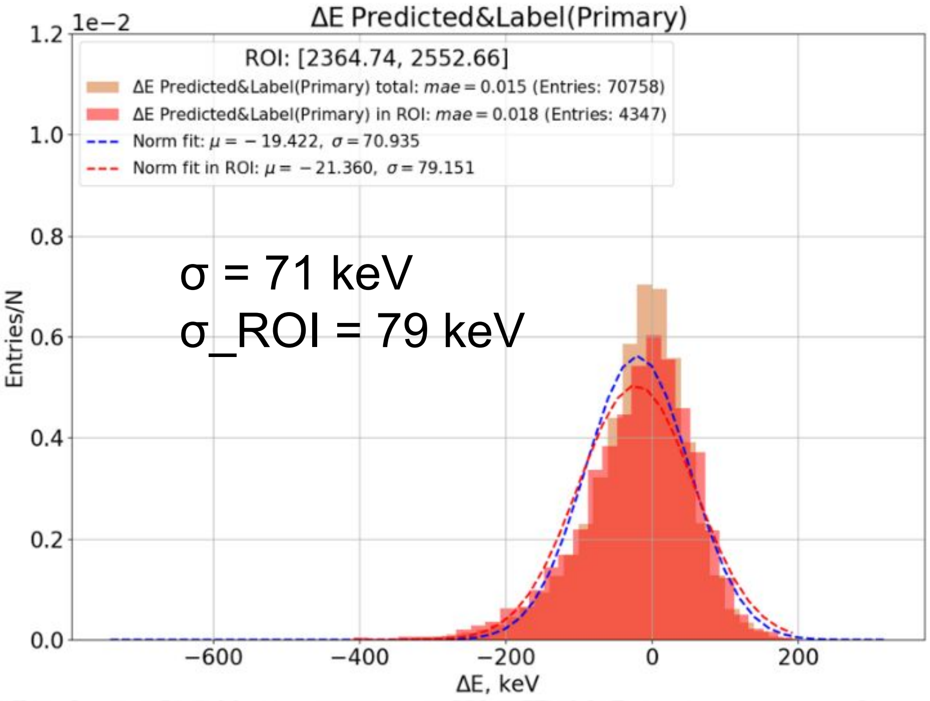
σ = 61 keV

σ_ROI = 69 keV

Visual Geometry Group (VGG)

Label: Primary energies



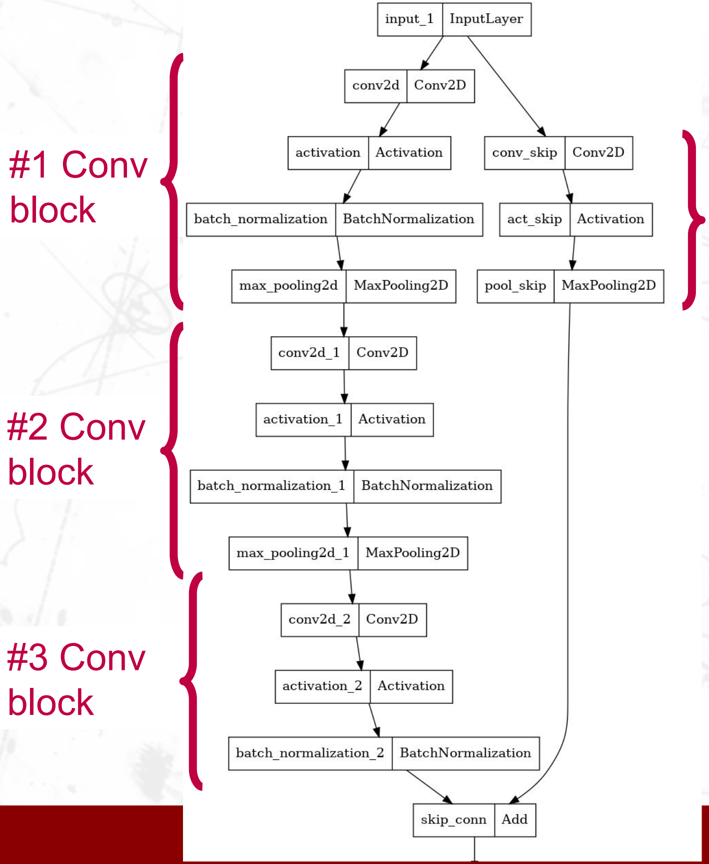


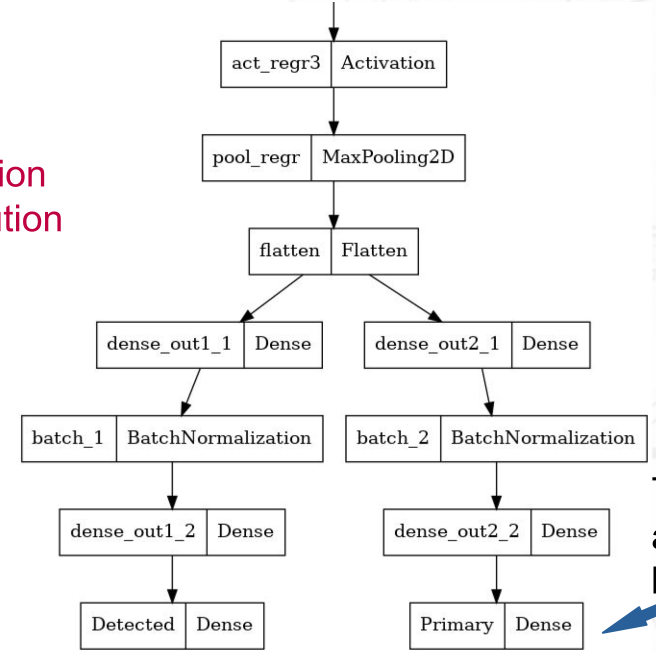
σ = 71 keV

σ_ROI = 79 keV

CNN ResNet with Convolution block as skip connection

Loss function: LogCosh loss $L(y, y^p) = \sum_{i=1}^{n} \log(\cosh(y_i^p - y_i))$

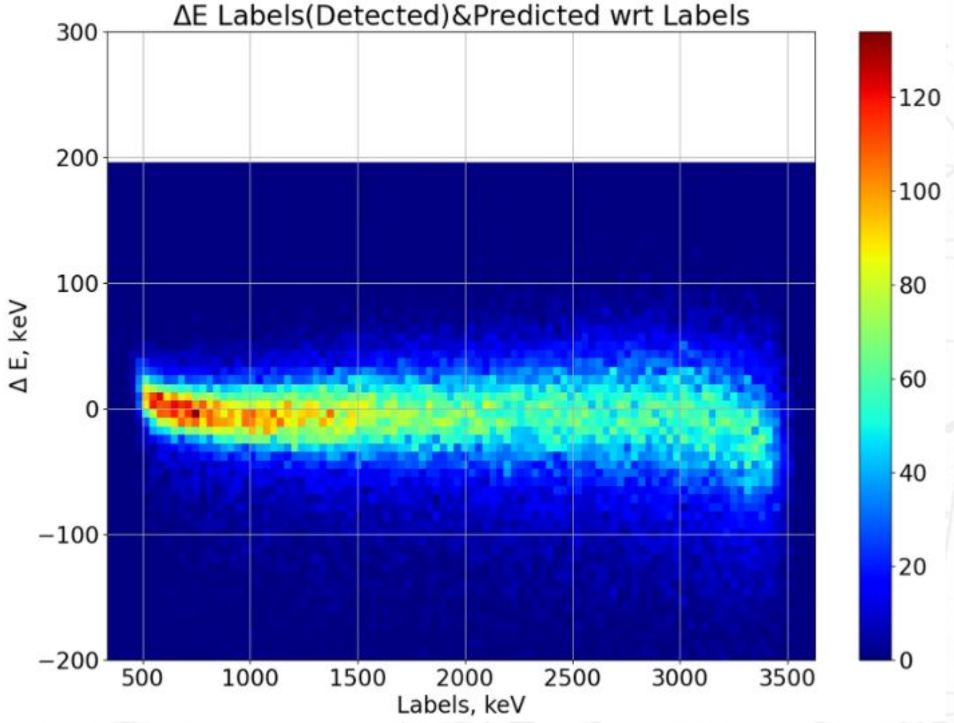


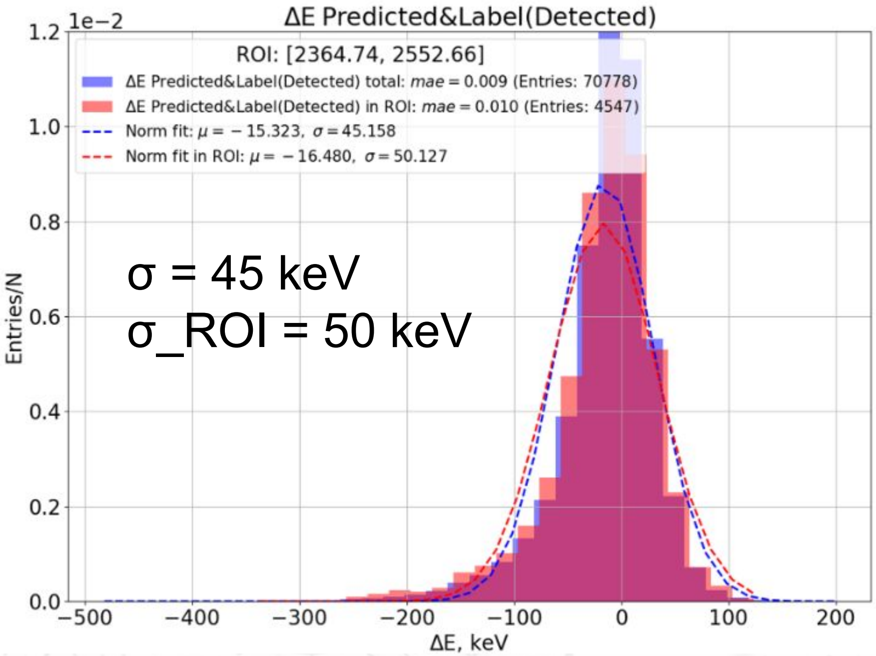


Two outputs with additional dense layers

CNN ResNet with Convolution block as skip connection

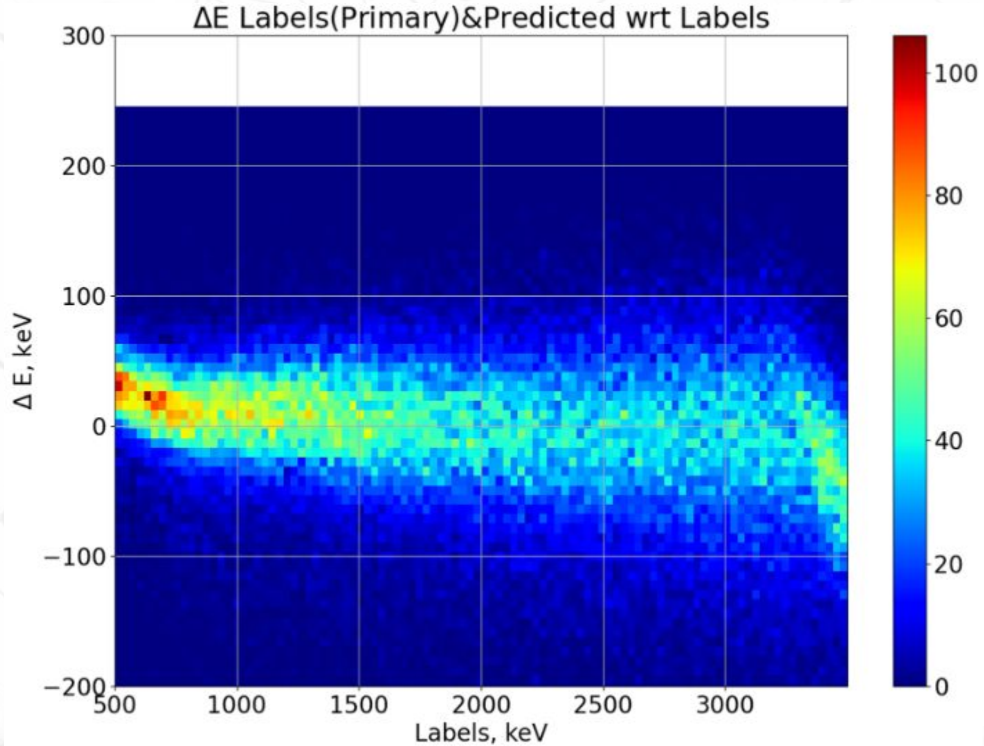
Label: Detected energies

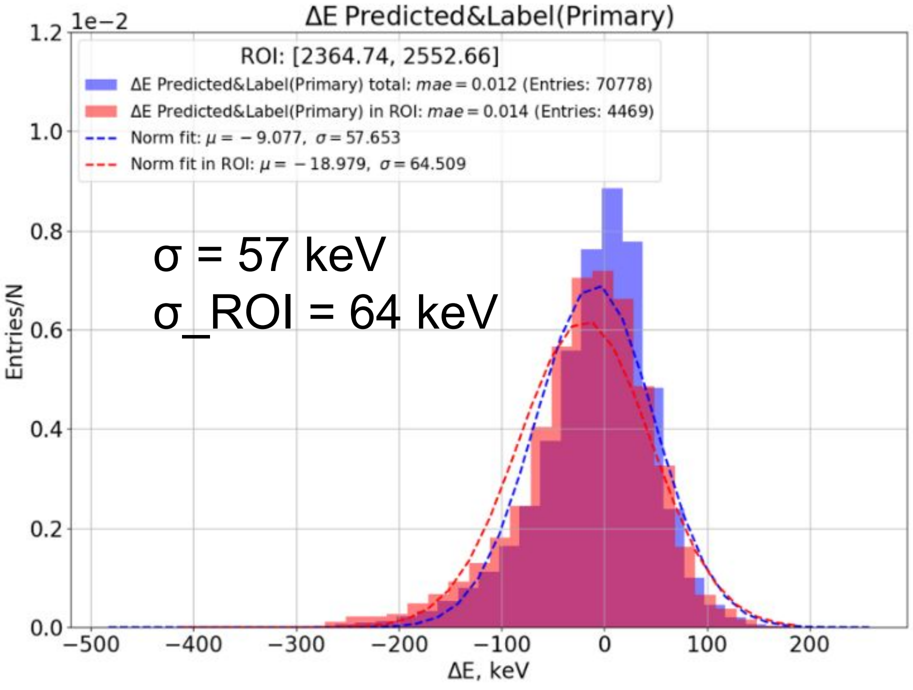




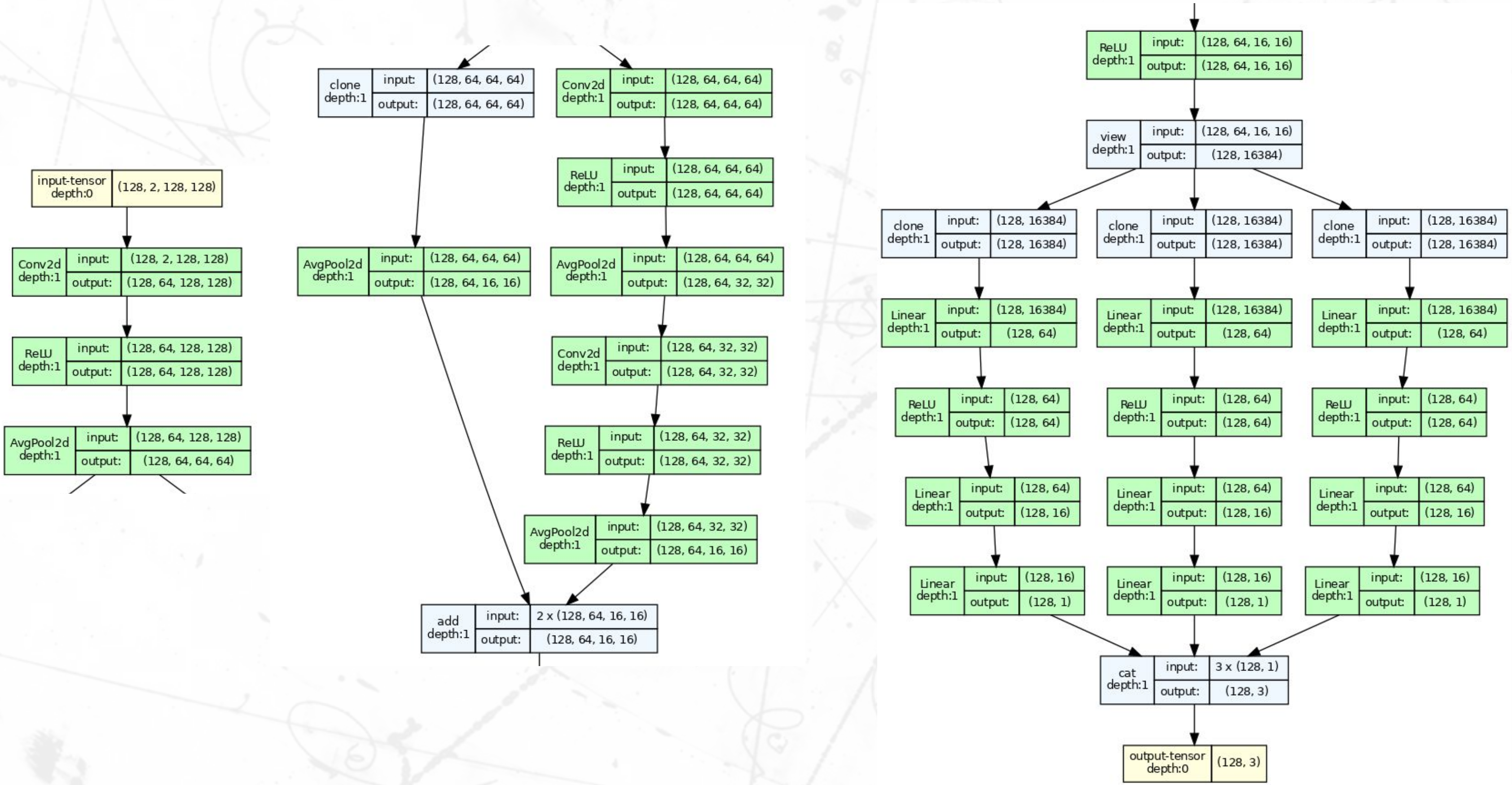
CNN ResNet with Convolution block as skip connection

Label: Primary energies

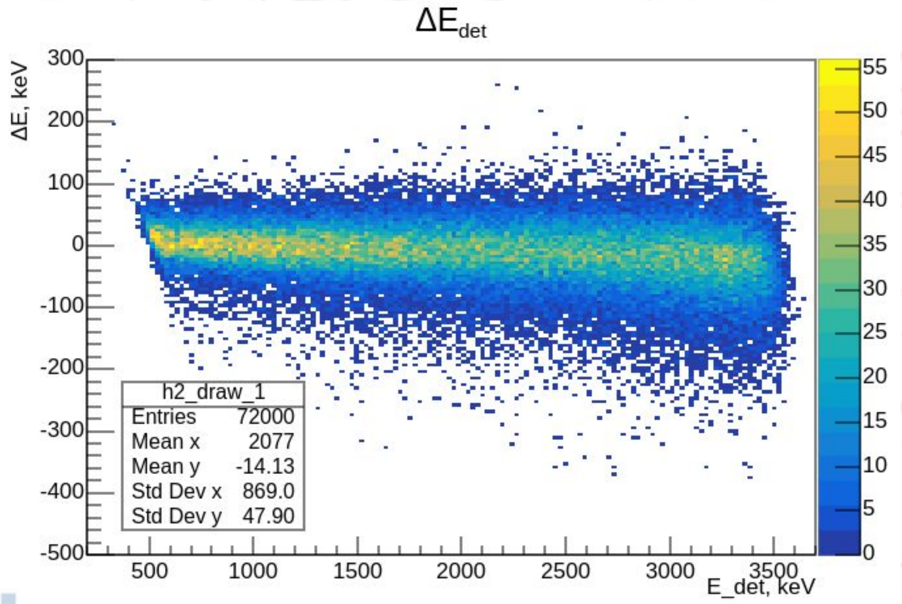


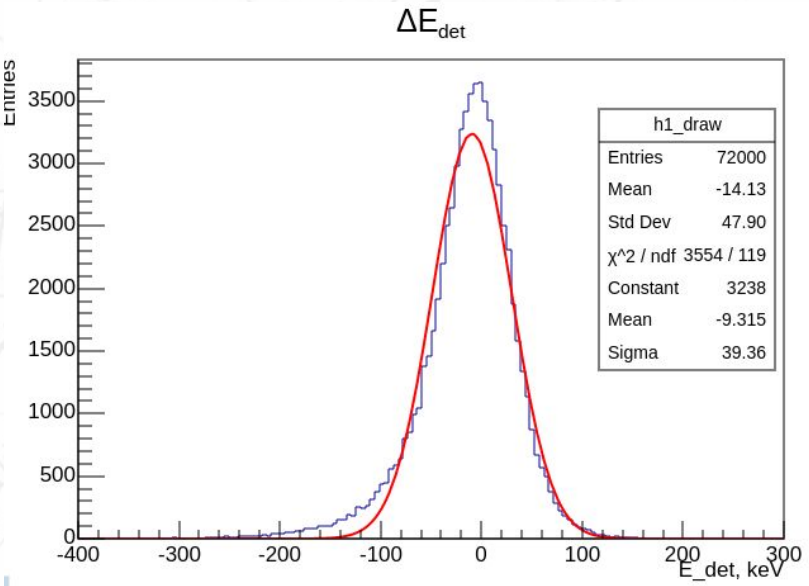


CNN ResNet with Average Pooling

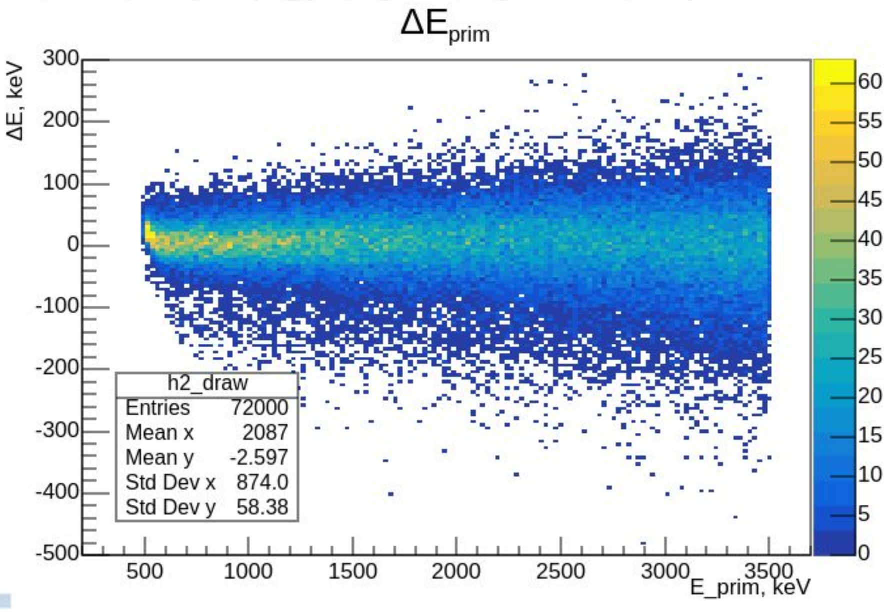


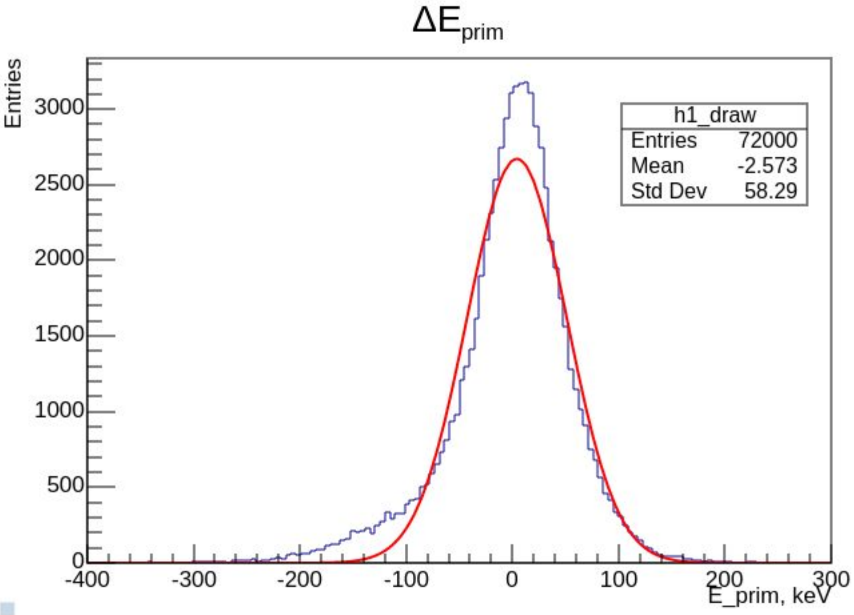
CNN ResNet with Average Pooling



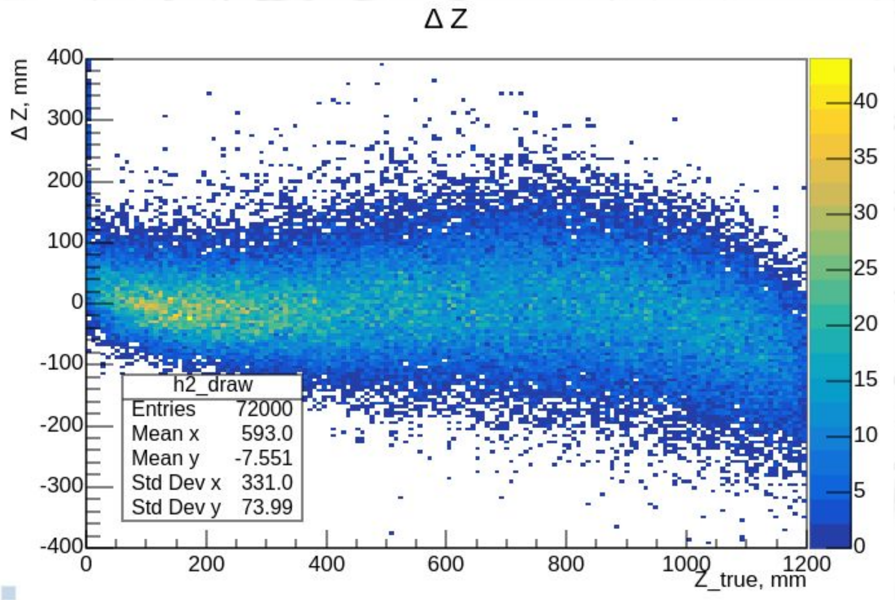


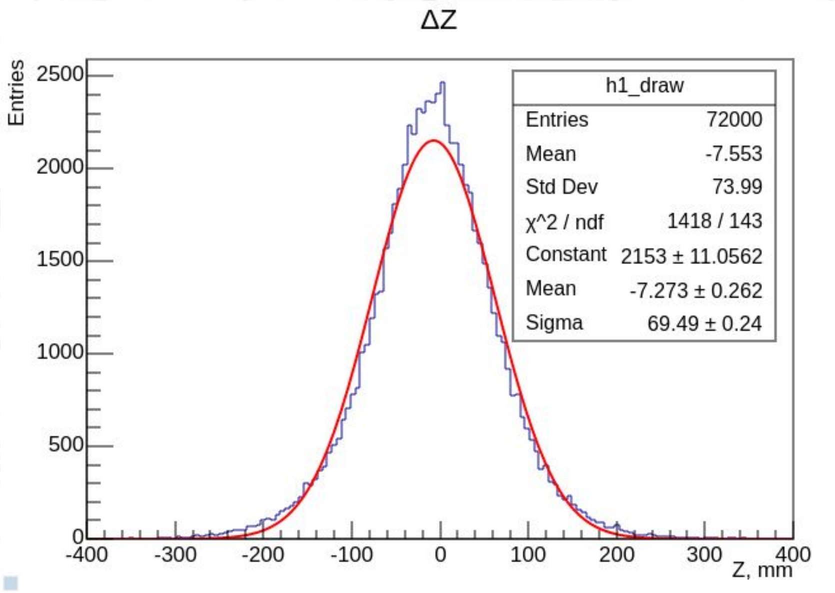
CNN ResNet with Average Pooling





CNN ResNet with Average Pooling





CNN ResNet with Average Pooling