Missing data reconstruction using ML techniques with the gaseous TPC PandaX-III experiment

Andrii Lobasenko

CEA Saclay IRFU/DPhN

IN2P3/IRFU Machine Learning workshop

27/09/2022

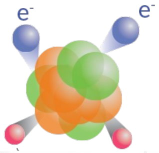
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 800945 — NUMERICS — H2020-MSCA-COFUND-2017

➤ Presentation Plan

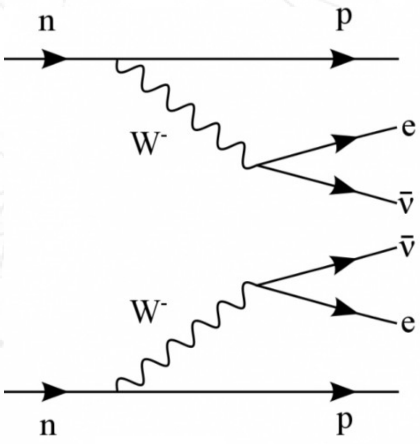
1. Introduction
   1.1 $0\nu 2\beta$ searches
   1.2 Xe136 gaseous TPC
2. Software environment REST
3. Simulations
4. Problems with Micromegas
5. Energy prediction with ML

➢ Searching for $0\nu2\beta$

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



$2\nu\beta\beta$

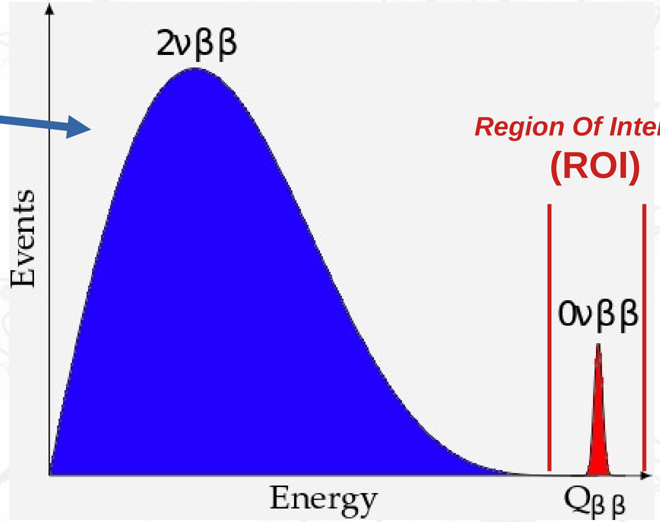


The rarest nuclear decay process

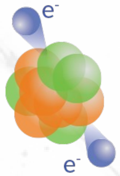
$$T_{1/2}(2\nu\beta\beta) \sim 10^{18} - 10^{24}\ yr$$

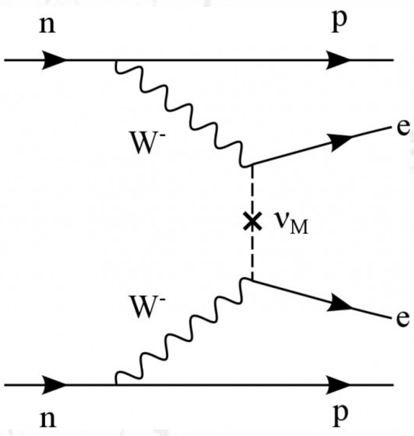
- Majorana nature of the neutrino (own anti-particle)
- Lepton number conservation violation
- Matter/antimatter asymmetry

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



$0\nu\beta\beta$

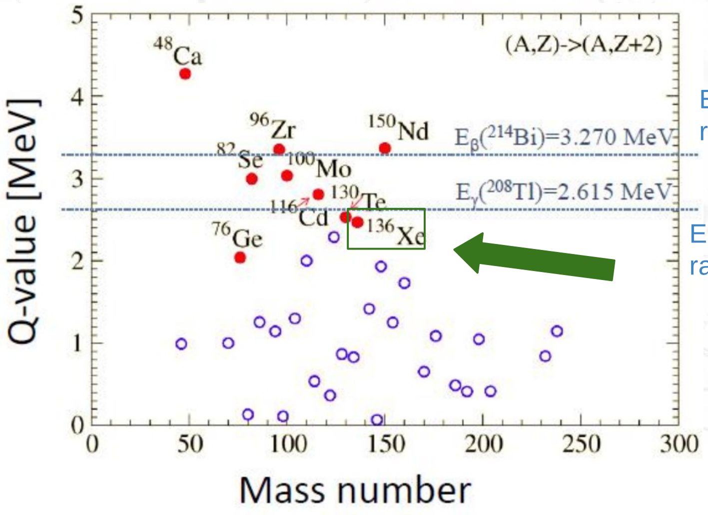




Expected half-life time of the 0ν decay

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

Xe136 gaseous TPC

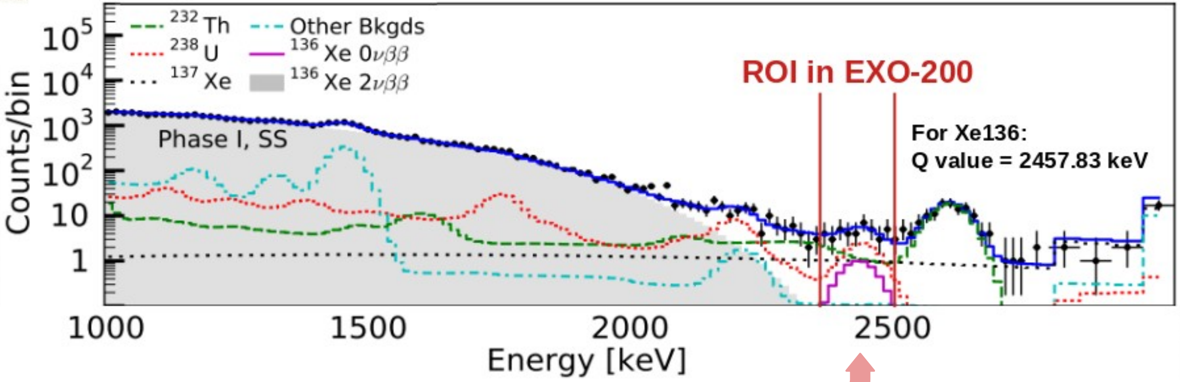


End of β radioactivity

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 results (liquid Xenon detector)

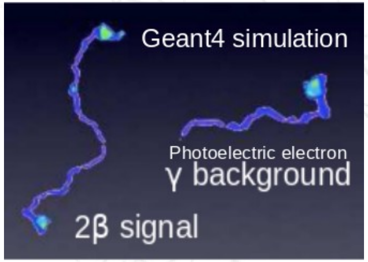


ROI is contaminated by the bkg:
U238 and Th232 decay chains:
- 2448 keV gamma from **Bi214**
- 2615 keV gamma from **Tl208**

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

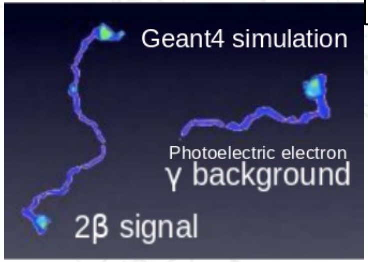
Xe136 gaseous TPC

Topological difference
b/w bkg and signal

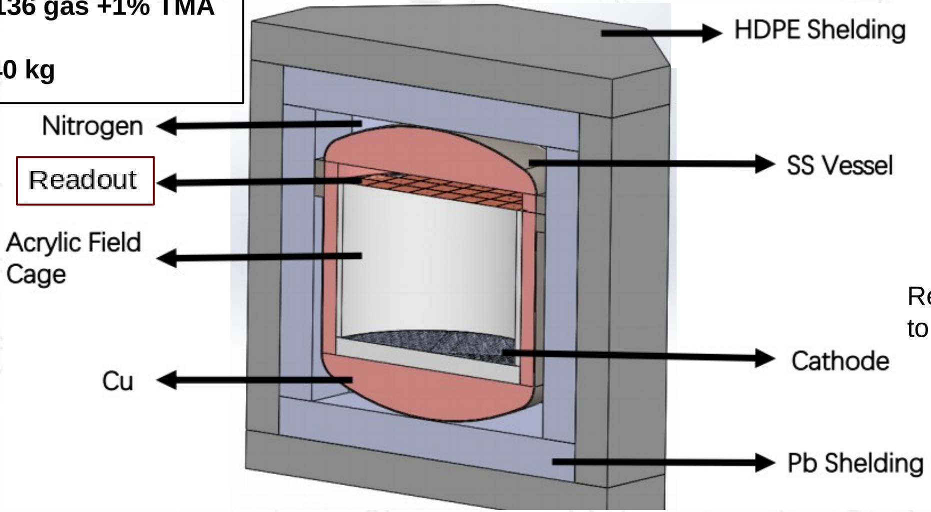


➢ Xe136 gaseous TPC

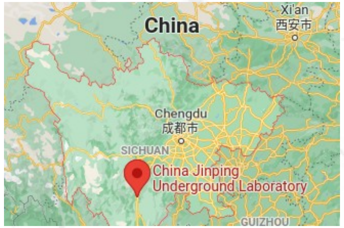
Topological difference b/w bkg and signal



90% enriched Xe136 gas +1% TMA
Pressure: **10 bar**
Total gas mass: **140 kg**



China Jinping Underground Laboratory (CJPL)

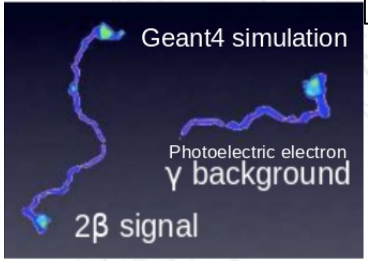


Reaching the cosmic **bkg level** to be **~1 cts/week/m^2**

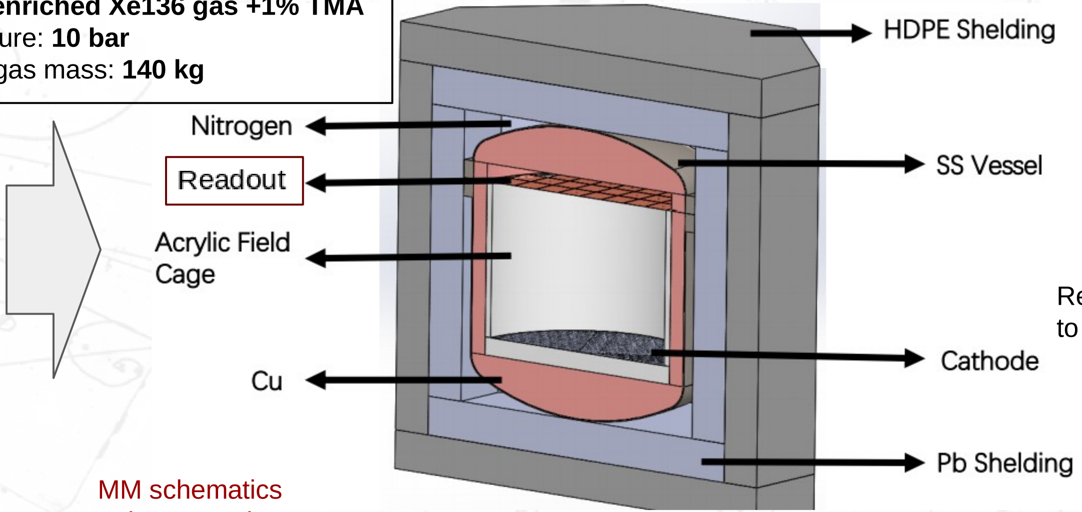
Bkg is suppressed from the surrounding contamination, leaving the irradiation from the electronics and Cu

➢ Xe136 gaseous TPC

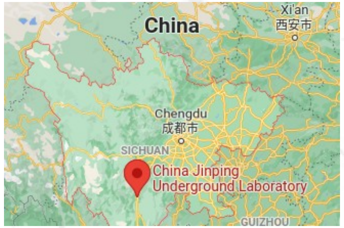
Topological difference b/w bkg and signal



90% enriched Xe136 gas +1% TMA
Pressure: **10 bar**
Total gas mass: **140 kg**



China Jinping Underground Laboratory (CJPL)

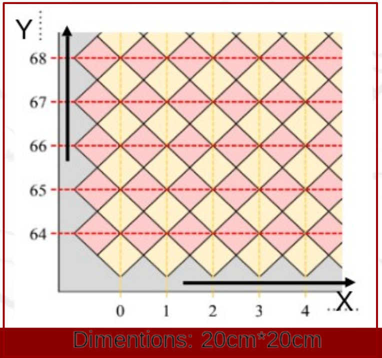


Reaching the cosmic **bkg level** to be **~1 cts/week/m^2**

Bkg is suppressed from the surrounding contamination, leaving the irradiation from the electronics and Cu

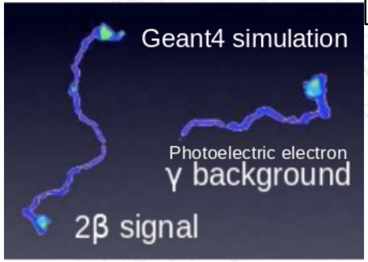
Readout:
52 Micromegas Modules
64+64 readout **channels** each

MM schematics (20x20 cm)

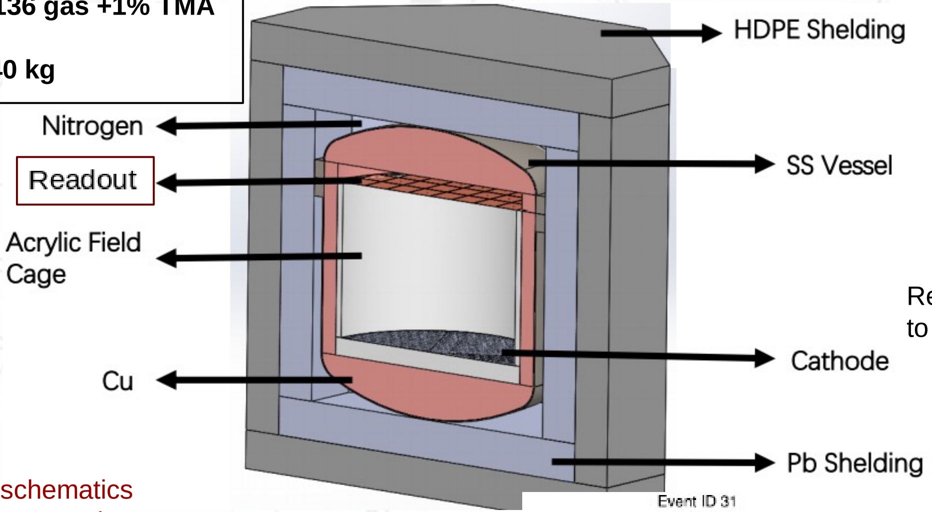


Xe136 gaseous TPC

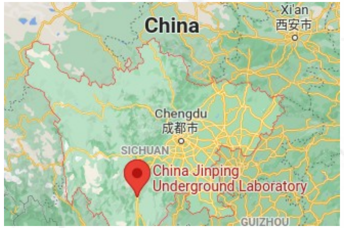
Topological difference
b/w bkg and signal



90% enriched Xe136 gas +1% TMA
Pressure: **10 bar**
Total gas mass: **140 kg**



China Jinping Underground
Laboratory (CJPL)

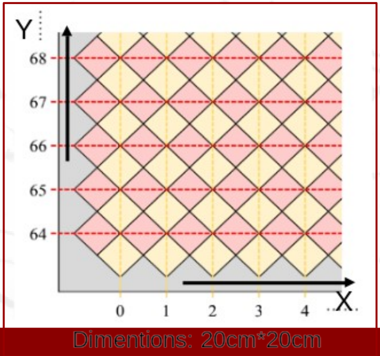


Reaching the cosmic **bkg level**
to be **~1 cts/week/m^2**

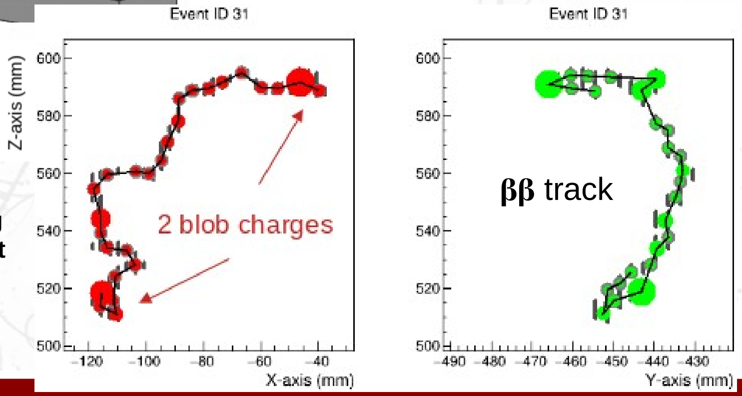
Bkg is suppressed from the surrounding contamination, leaving the irradiation from the electronics and Cu

Readout:
52 Micromegas Modules
64+64 readout **channels** each

MM schematics
(20x20 cm)

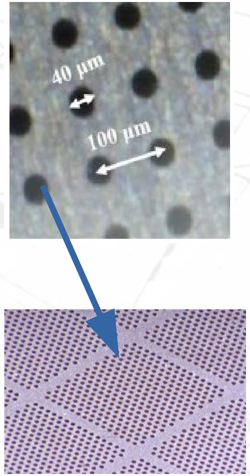


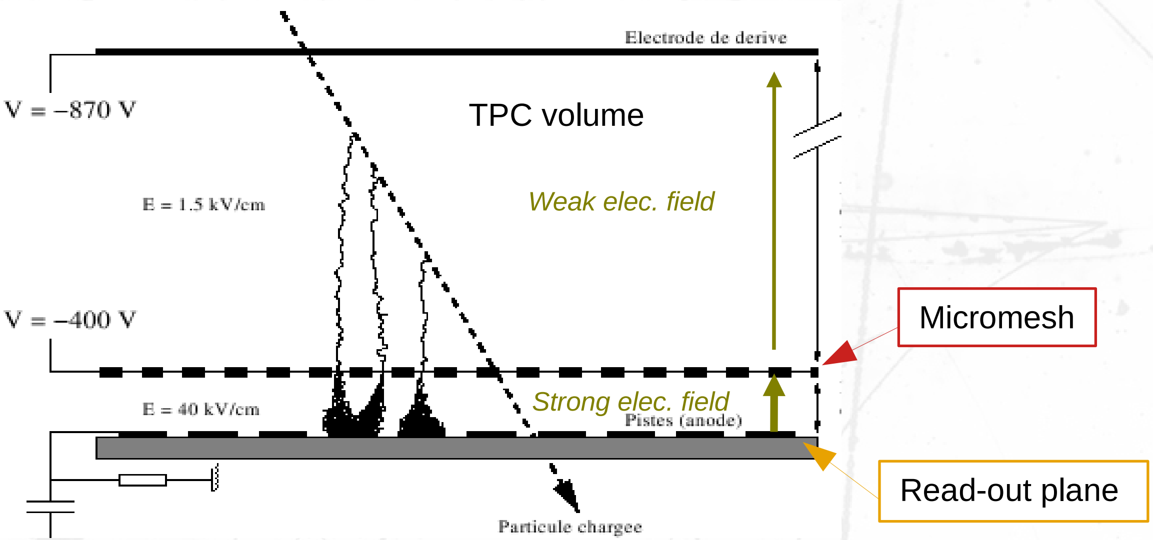
XZ&YZ projections after data processing of the detector output



Xe136 gaseous TPC

Micromegas basic principle

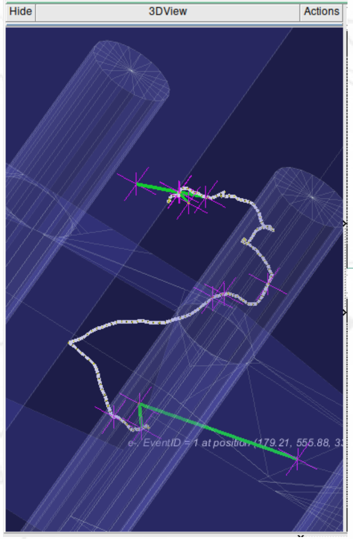




Software environment REST

Monte-Carlo simulation
Done with **REST** environment

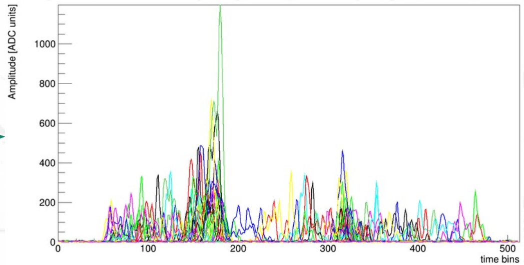
CAST
IAXO
T-REX



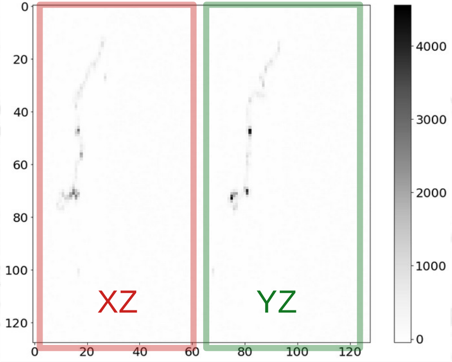
Simulation of the detector response

Electron diffusion, smearing, Signal shaping, etc.

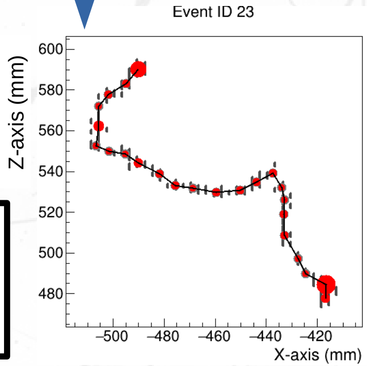
Signal from the readout system

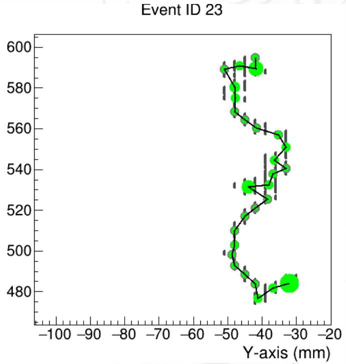


128 x 128 matrix



Track Reconstruction processing chain

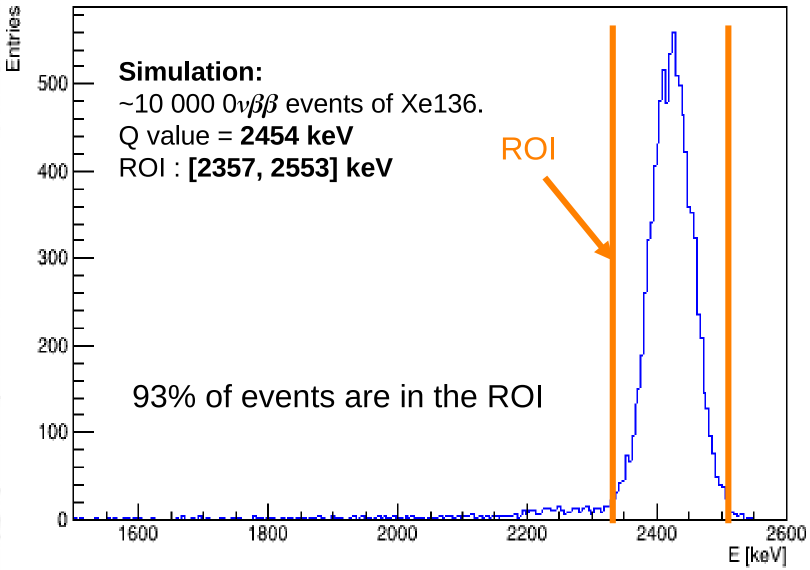




Output:
- Total reconstructed energy of the track
- Track topology

➢ Software environment REST

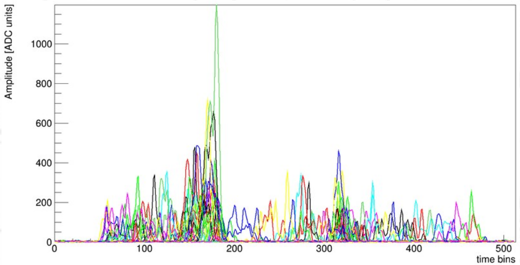
Energy spectrum of the reconstructed 0 events



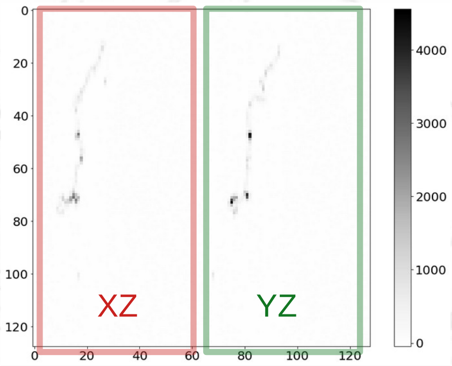
Simulation:
~10 000 $0\nu\beta\beta$ events of Xe136.
Q value = **2454 keV**
ROI : **[2357, 2553] keV**

93% of events are in the ROI

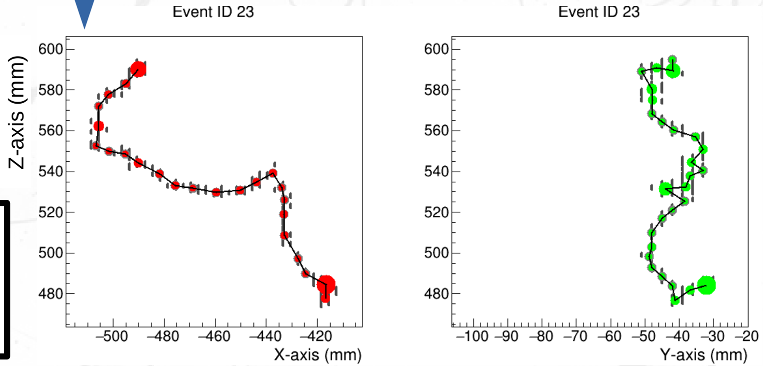
Signal from the readout system



128 x 128 matrix



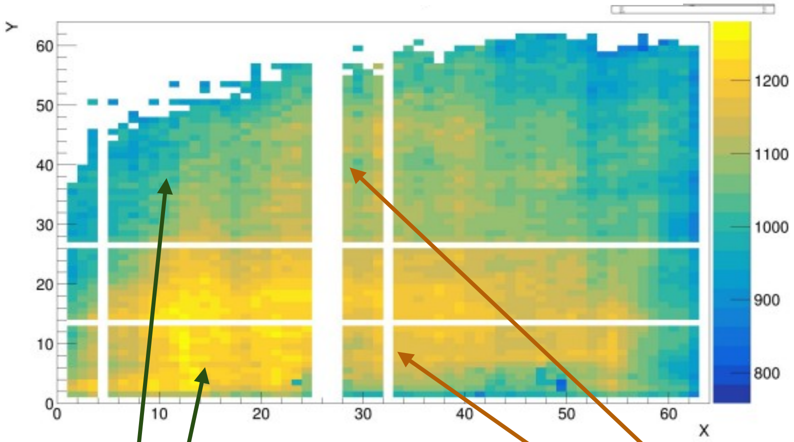
Track Reconstruction processing chain



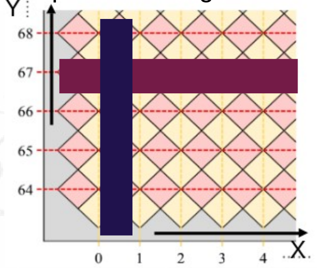
Output:
- Total reconstructed energy of the track
- Track topology

➢ Problems with Micromegas

Gain map for one Micromegas module



Example of missing channels

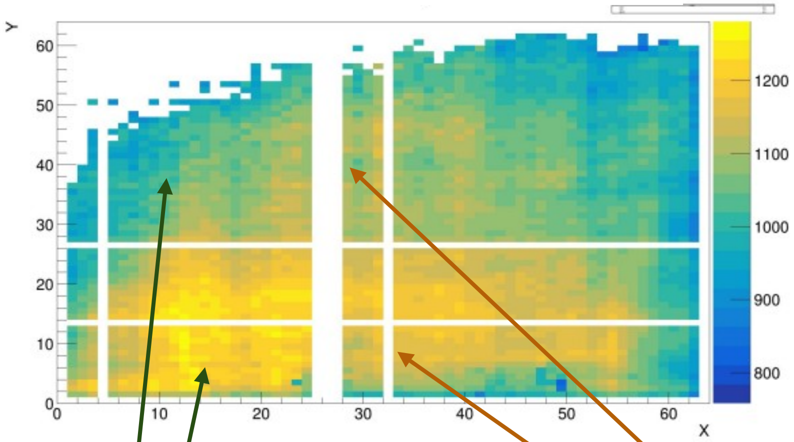


Inhomogeneity of the Gain causes an incorrect energy reconstruction

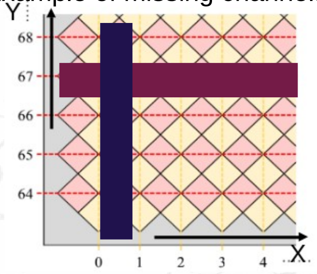
Missing channels cause loss of:
- Topology of the track
- Energy reconstruction info

➤ Problems with Micromegas

Gain map for one Micromegas module



Example of missing channels

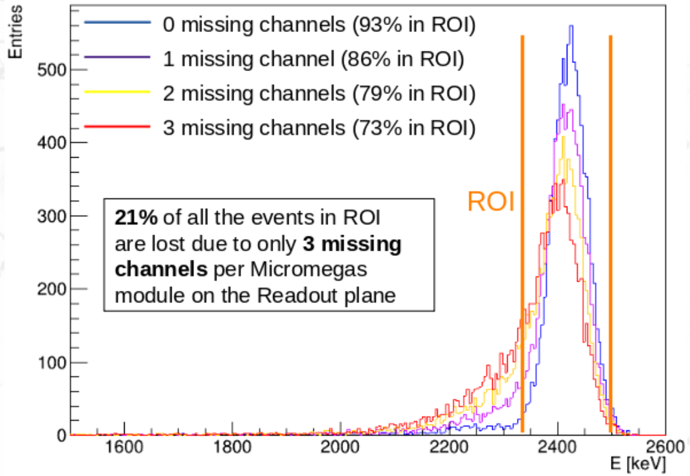


Inhomogeneity of the Gain causes an incorrect energy reconstruction

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

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

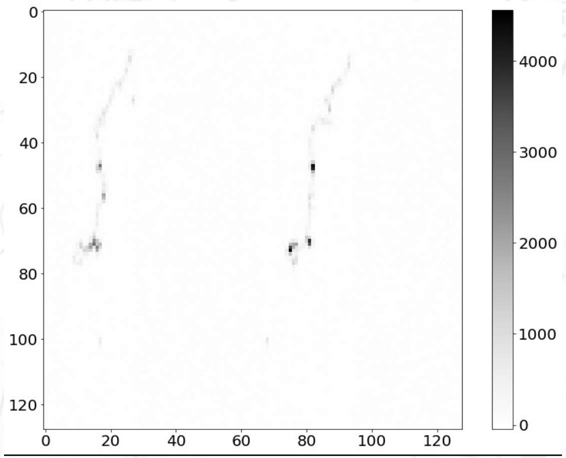
Energy spectrum of the reconstructed $0\nu\beta\beta$ events



Energy prediction with ML

Simulated data:
- 80k Signals
- one e- events
- registered by 1 MM
- primary energy range [1, 3] keV

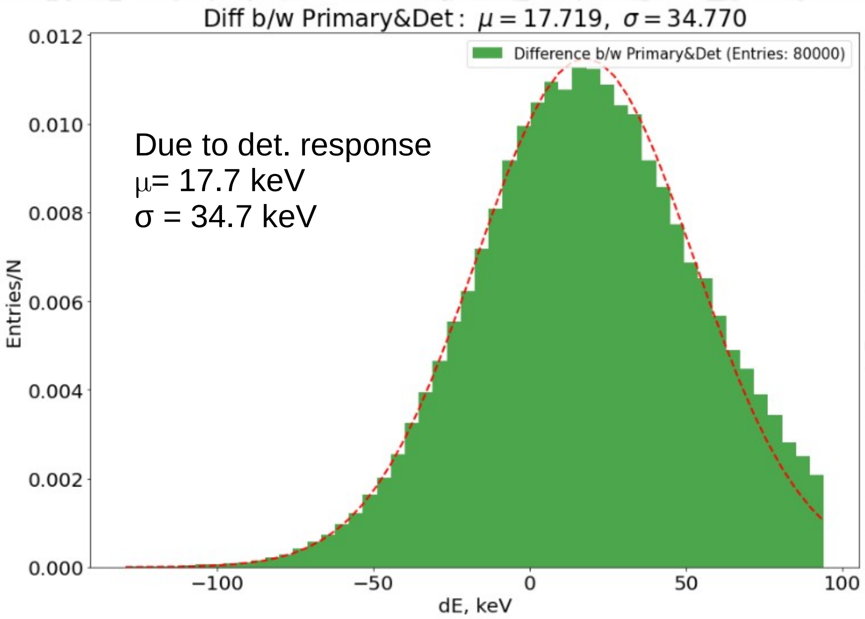
128 x 128 matrix



(Simulated Signal from the detector)

Primary energy: initial energy of the generated e- by the Geant4 *(labels)*

Detected energy: energy “registered” by the detector in REST

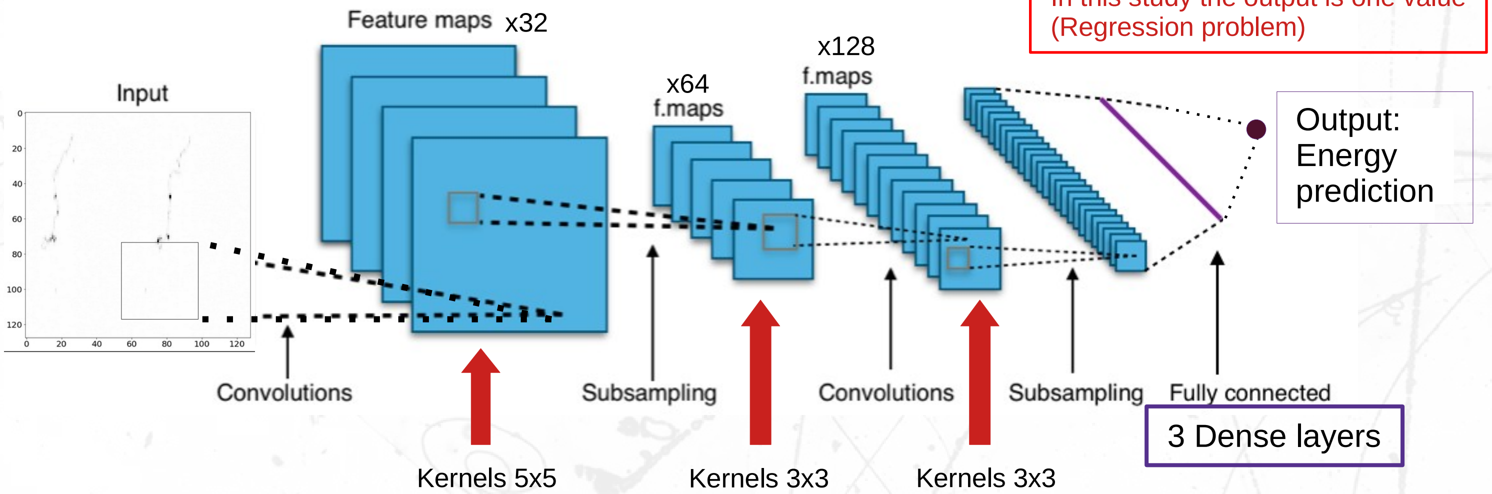


Due to det. response
μ= 17.7 keV
σ = 34.7 keV

➢ Energy prediction with ML

Before starting the energy reconstruction, a Convolutional Neural Network (CNN) model is being tested to predict the primary energy of the event from the simulated detector data without missing information

In this study the output is one value (Regression problem)

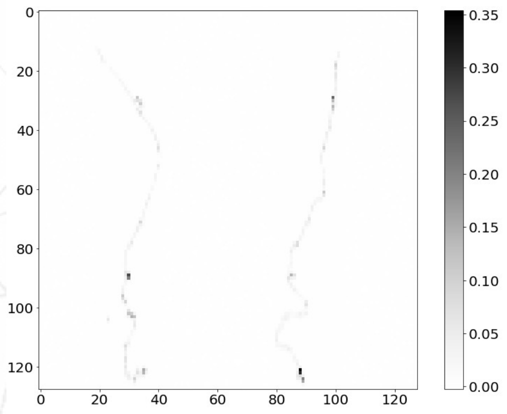


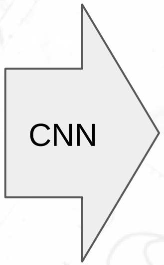
Loss function for the back-propagation: Mean Squared Error (MSE)

Energy prediction with ML

Preliminary

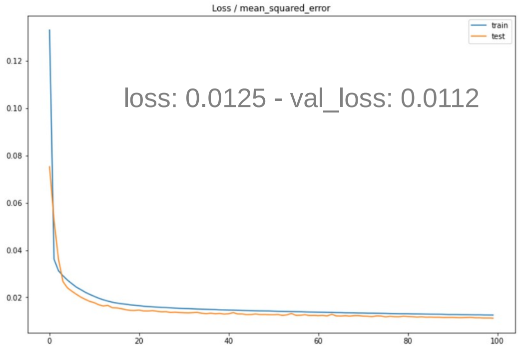
Global event normalization by the max amplitude among all the events

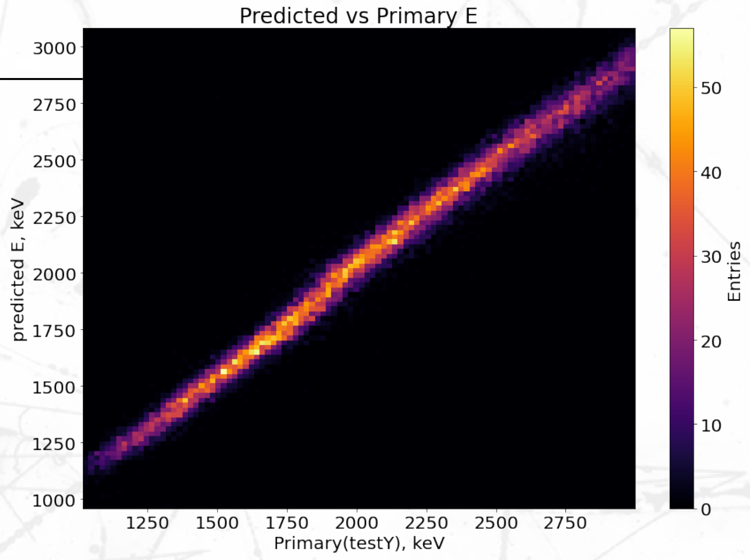


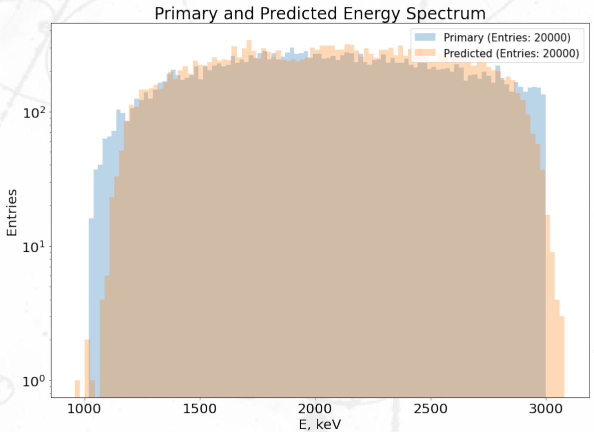


L2regularization = 0.00003
SGD(learning_rate=2e-2, momentum = 0.005)
loss="mean_squared_error"

epochs = 100
batch size = 128



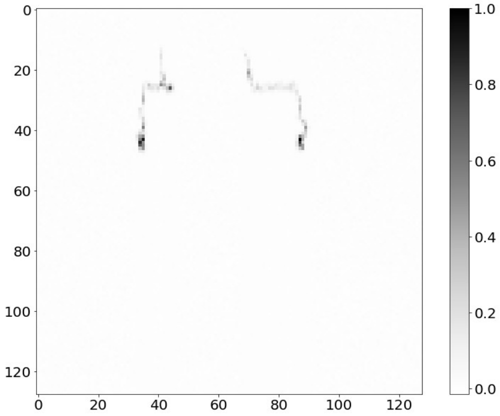




Energy prediction with ML

Preliminary

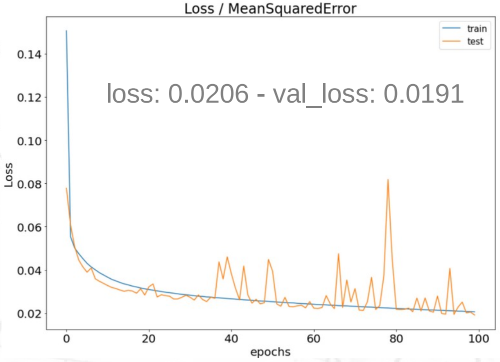
Individual event normalization by the max amplitude in each event
(only the topological information is stored)

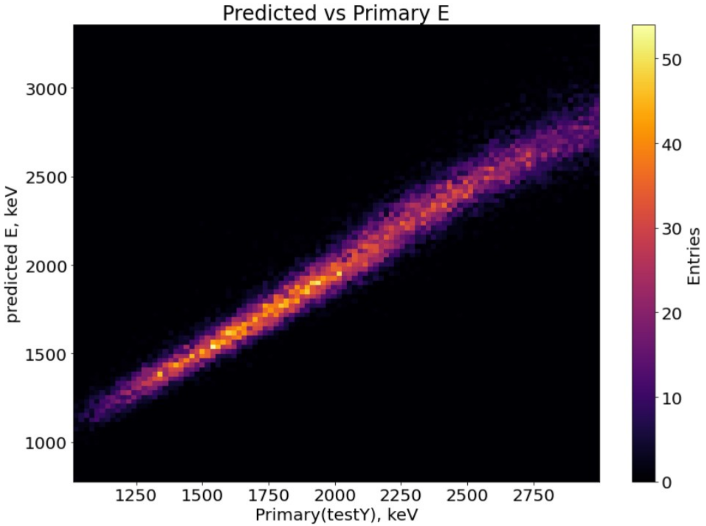


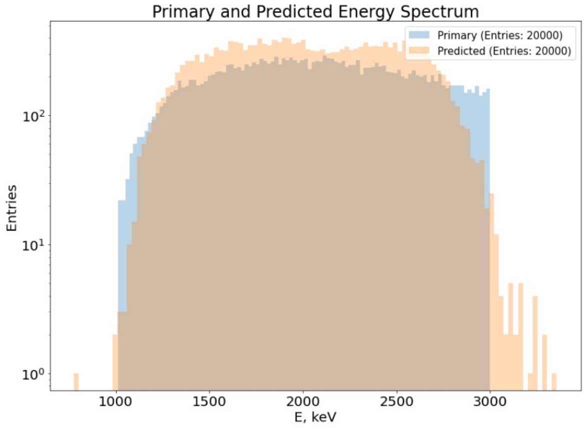
CNN

L2regularization = 0.0002
SGD(learning_rate=2e-2,
momentum = 0.005)
loss="mean_squared_error"

epochs = 100
batch size = 256

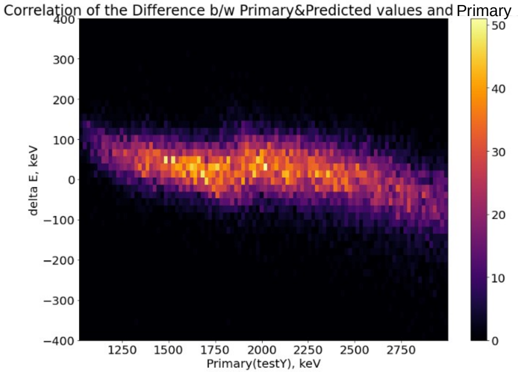


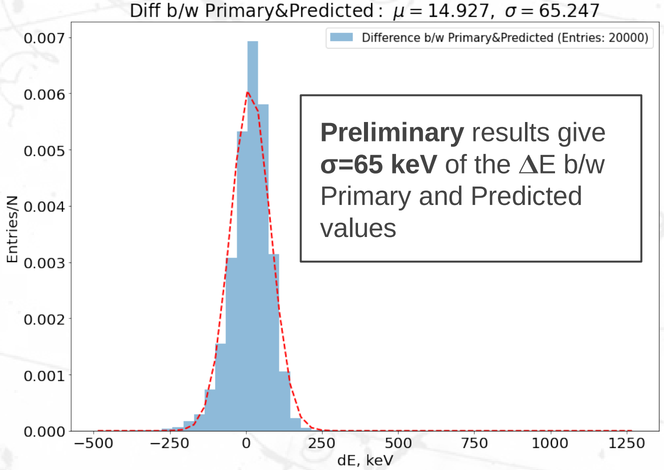




➢ Energy prediction with ML

Global event normalization

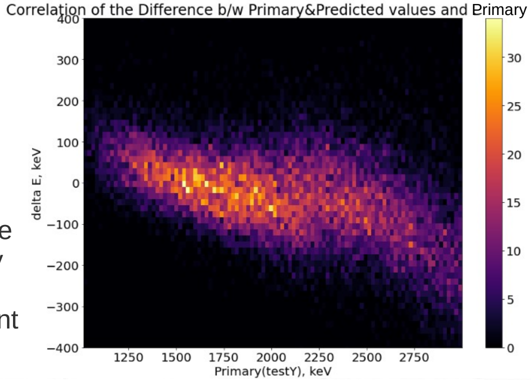


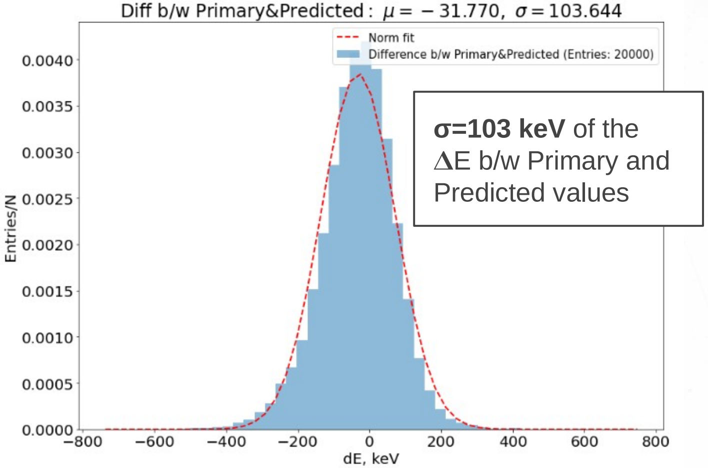


Preliminary results give **σ=65 keV** of the ΔE b/w Primary and Predicted values

Individual event normalization

Slight dependency of the diff. Predicted – Primary wrt Primary value
=> room for improvement





σ=103 keV of the ΔE b/w Primary and Predicted values

Conclusions

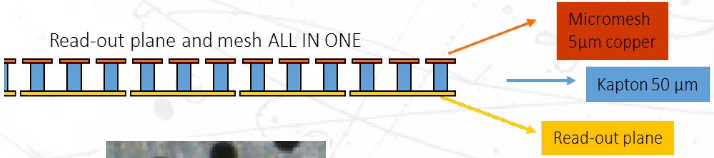
Promising preliminary results on ML techniques on the energy prediction
Individually normalized data prediction will be completed

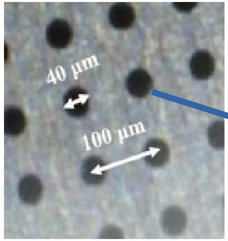
Next steps:
Energy prediction with events **with missing channels**
Update of the models architecture for **bkg/signal discrimination**
Energy prediction based on **real detector gain**

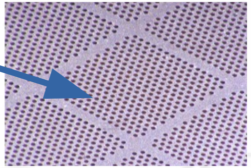
Xe136 gaseous TPC

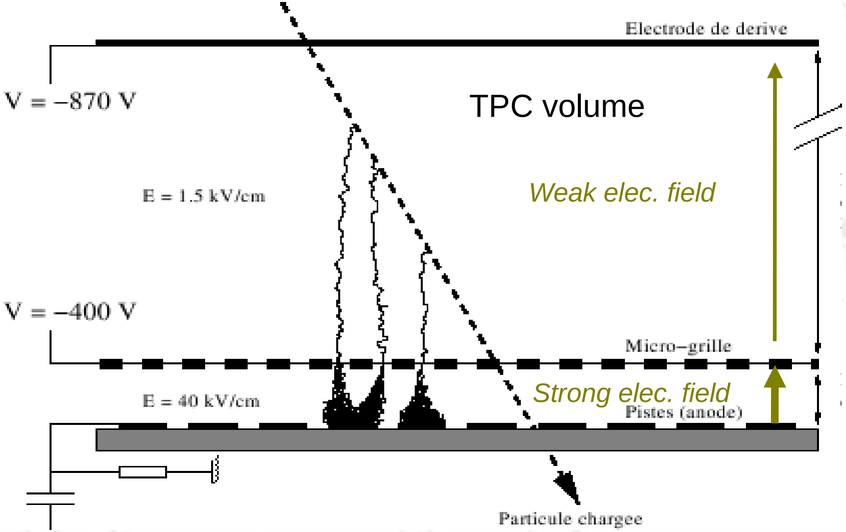
Micromegas basic principle

(Microbulk example)



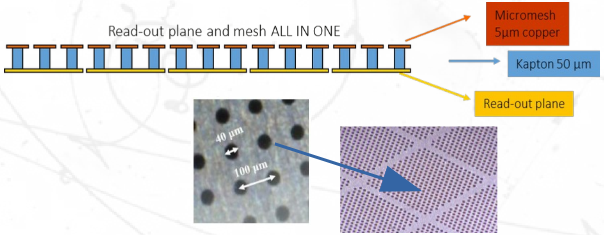






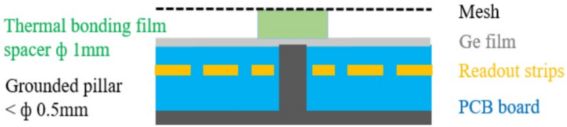
➢ Xe136 gaseous TPC

Microbulks



Micromegas based on a copper clad 50µm-thick kapton foil; 40µm diameter holes
Top face → mesh
Bottom face → read-out plane
Constant kapton foil thickness
→ very good gain homogeneity
→ best energy resolution among MPGDs
Only kapton and copper → excellent radiopurity
~0.1 µBq/cm² for ^{214}Bi and ^{208}Tl
Studied by Zaragoza, IRFU and SJTU
Built at CERN, used at CAST, n_TOF

Thermo-bonded MM



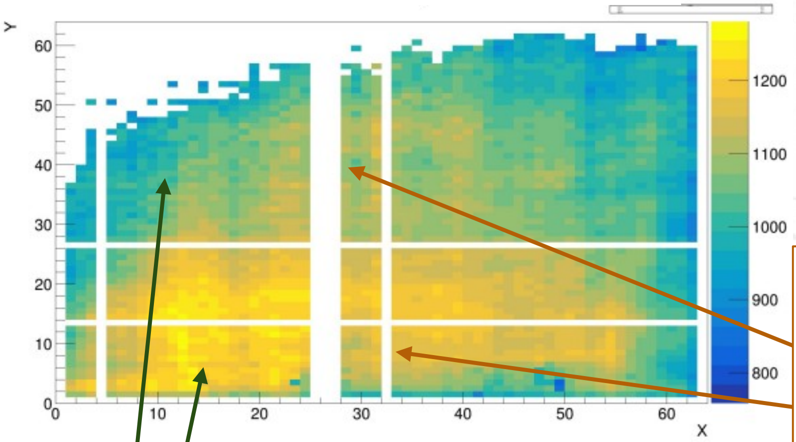
Regular Micromegas with resistive Germanium layer
Mesh spacing by thermo-bonded polyester layer
Comparison with Microbulk:

- more robust
- low radioactive material
- sparks protection with resistive layer
- larger energy resolution expected compared to Microbulks

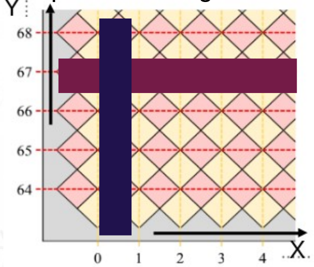
Developed and built at USTC (Hefei, China)

Problems with Micromegas

Gain map for one Micromegas module



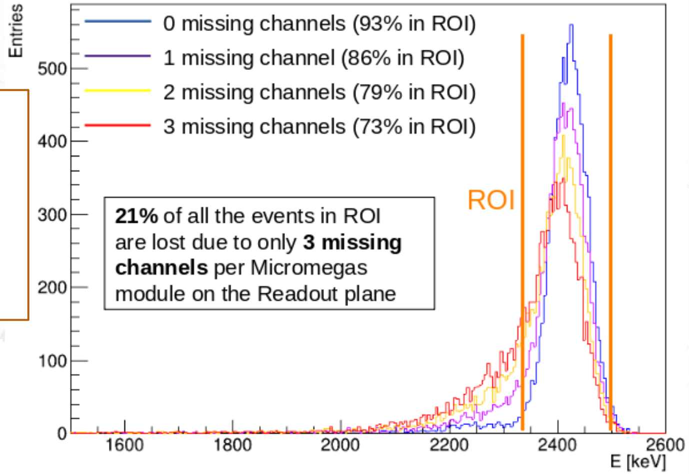
Example of missing channels



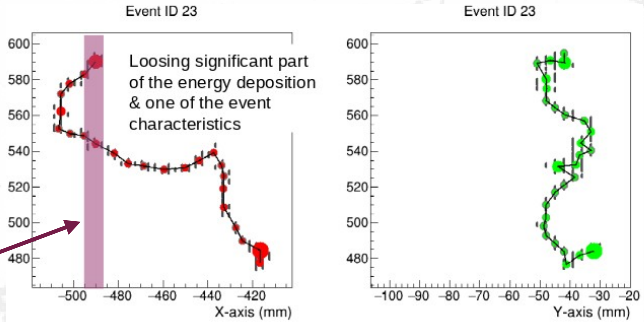
Simulation:
~10 000 $0\nu\beta\beta$ events of Xe136.
Q value = **2454 keV**
ROI : **[2357, 2553] keV**

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

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



Inhomogeneity of the Gain causes an incorrect energy reconstruction



Due to consecutively missing channels

Proper reconstruction of the missing information must be addressed
Starting from **Blob charge correlation** study

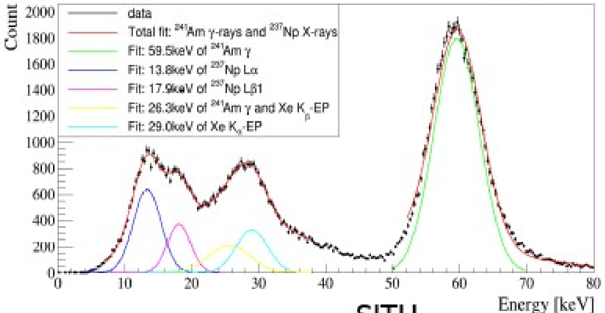
Microbulks

Fragile detectors: cut channels, dark currents

Gain inhomogeneity for some detectors (not all) linked to production problems

Rather good resolution but not as good as expected

^{241}Am source, 5 bar Xe+1%TMA



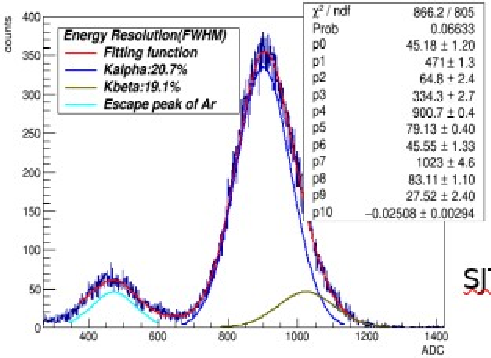
SITU

Thermo-bonded MM

Good energy resolution of 15% at 6 keV (Ar + 5% isobutane 1 bar)

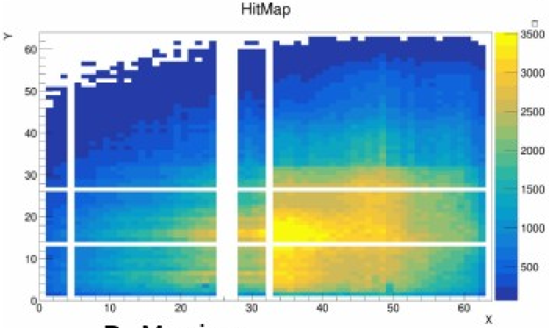
Some non-uniformity of the gain a priori due to production methods, improved performance with new methods

Unstable dark current at high pressure

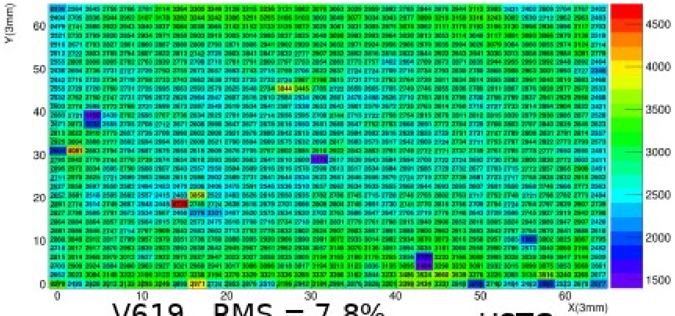


SITU

Work in progress



B. Manier

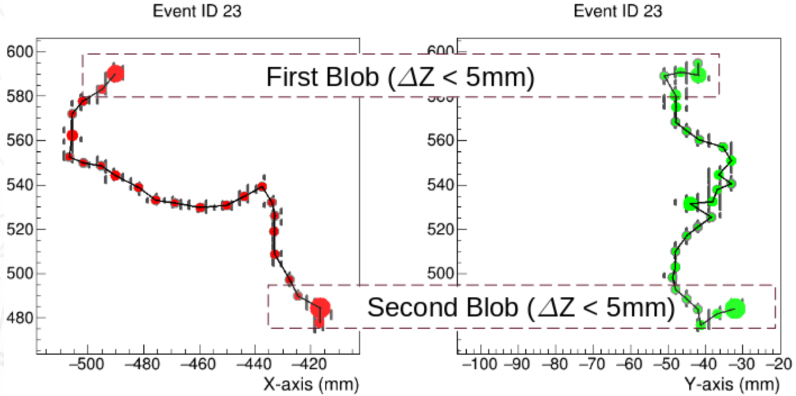


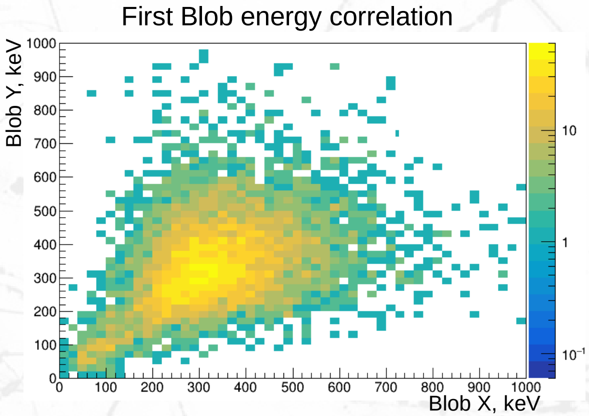
V619 RMS = 7.8%

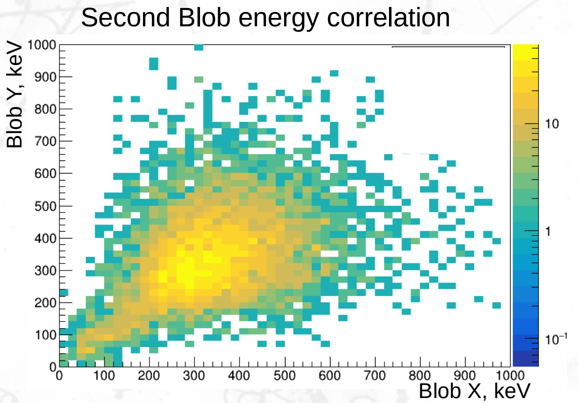
USTC

➢ Blob charge correlation

Approximation of the lost Blob charge information from the second projection → Study of the Blob charge energy correlation



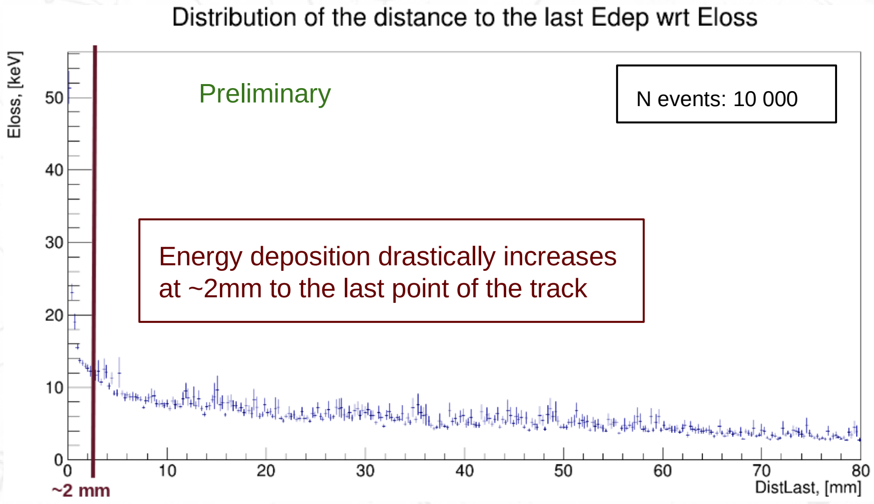


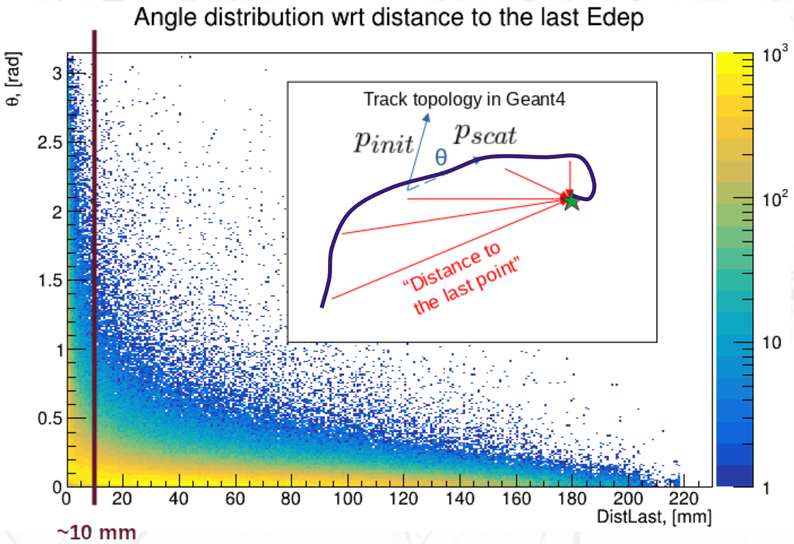


The correlation b/w Blob Charges is not evident, however from the first approximation we may notice that both **Blobs contain ~700 keV** of the total reconstructed energy (E_x + E_y)

Blob charge correlation

Better determination of the Blob charge deposition → Study of the correlation b/w E deposition and Scattering angle θ (Pure Geant4)



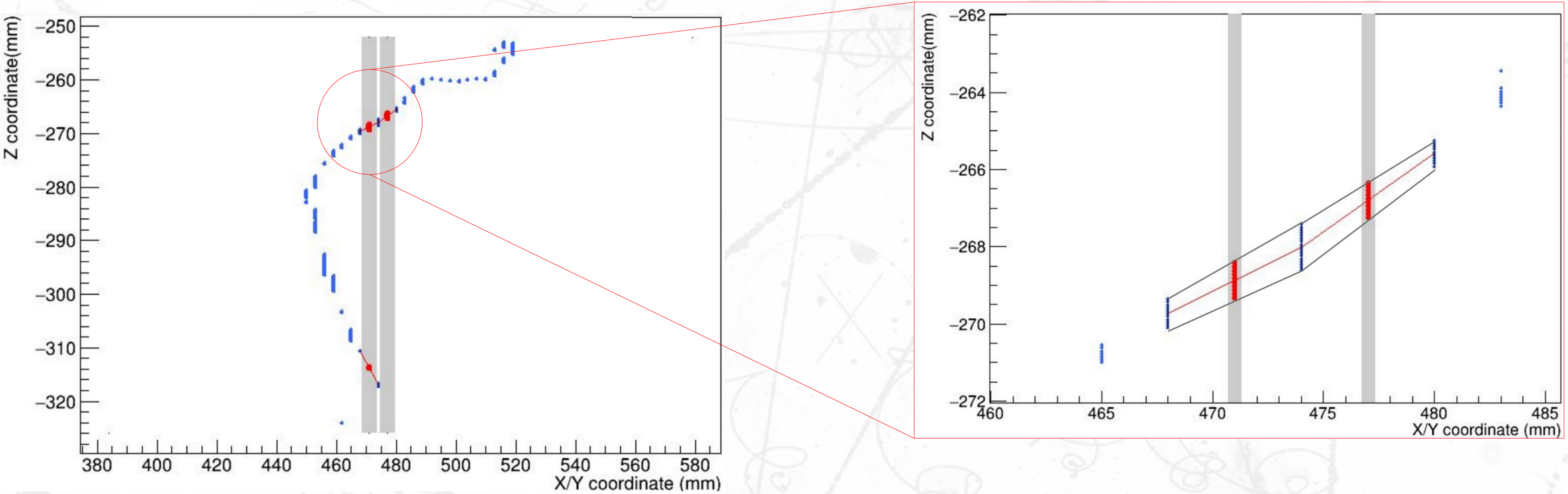


Most of the deposited energy is in the "tail"
Scattering angle could be used as a feature for NN application
Ongoing study of the real topology of the event determination

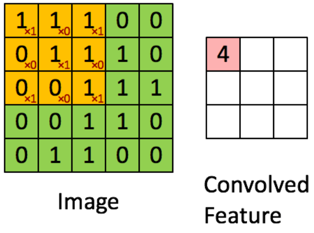
Missing channel repairing with linear interpolation (Benjamin Manier)

- Added hits are based on the side segments on the cut strip
- Energy is interpolated linearly from the side segment

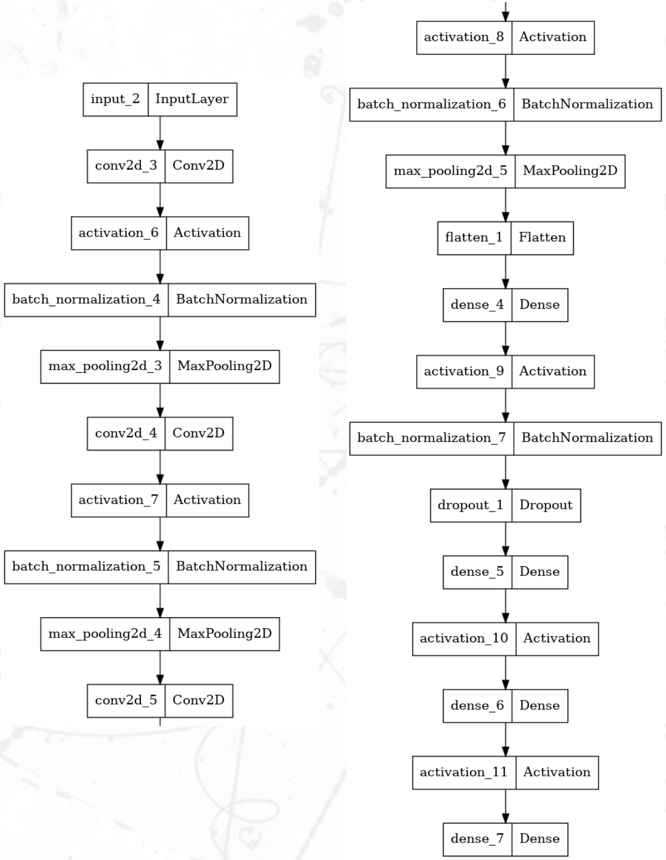
Hits Event



Energy prediction with ML

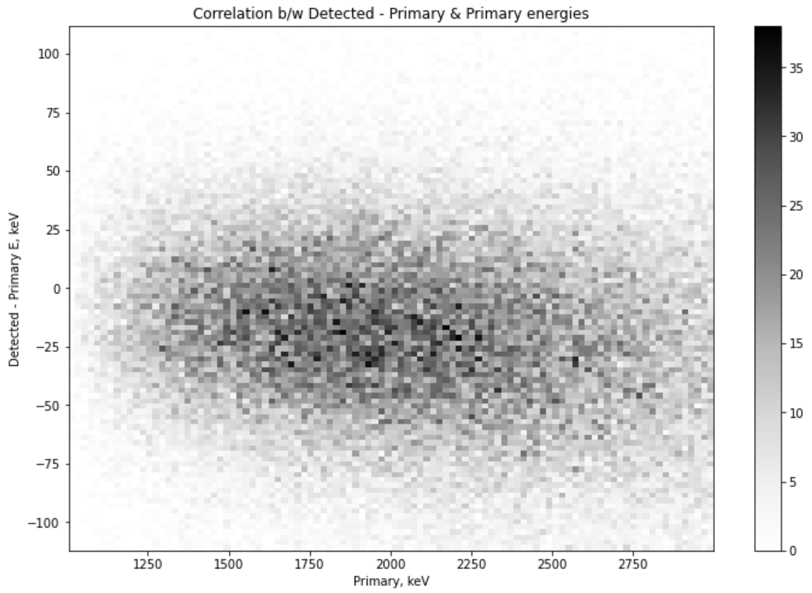


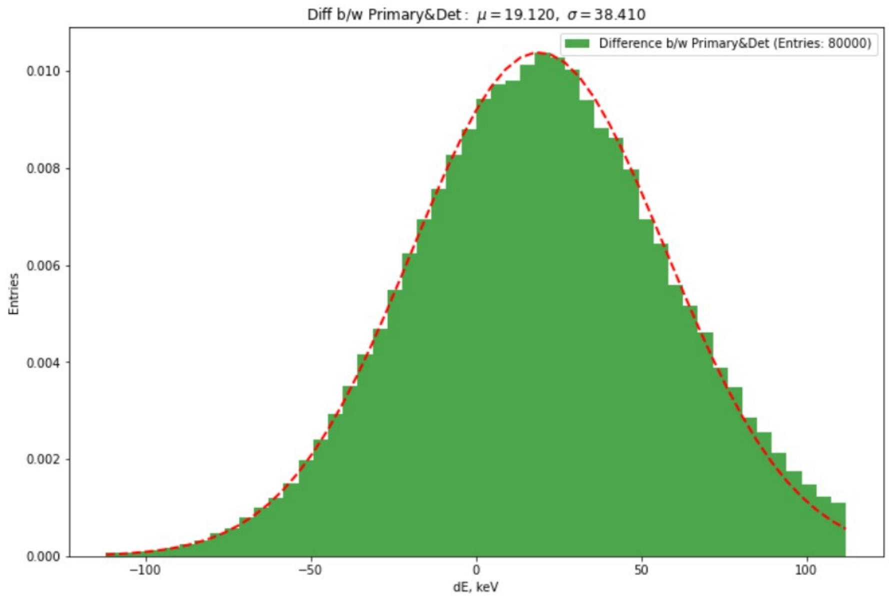
$$y_{jk}(x) = f\left(\sum_{i=1}^{n_H} w_{jk}x_i + w_{j0}\right)$$



➢ Energy prediction with ML

Data coherence





Expect similar distribution for the difference b/w Primary and Predicted values