2026 International Workshop on CEPC

Shanghai, China | Oct 25, 2026

PID in drift chambers for future e^+e^- experiments

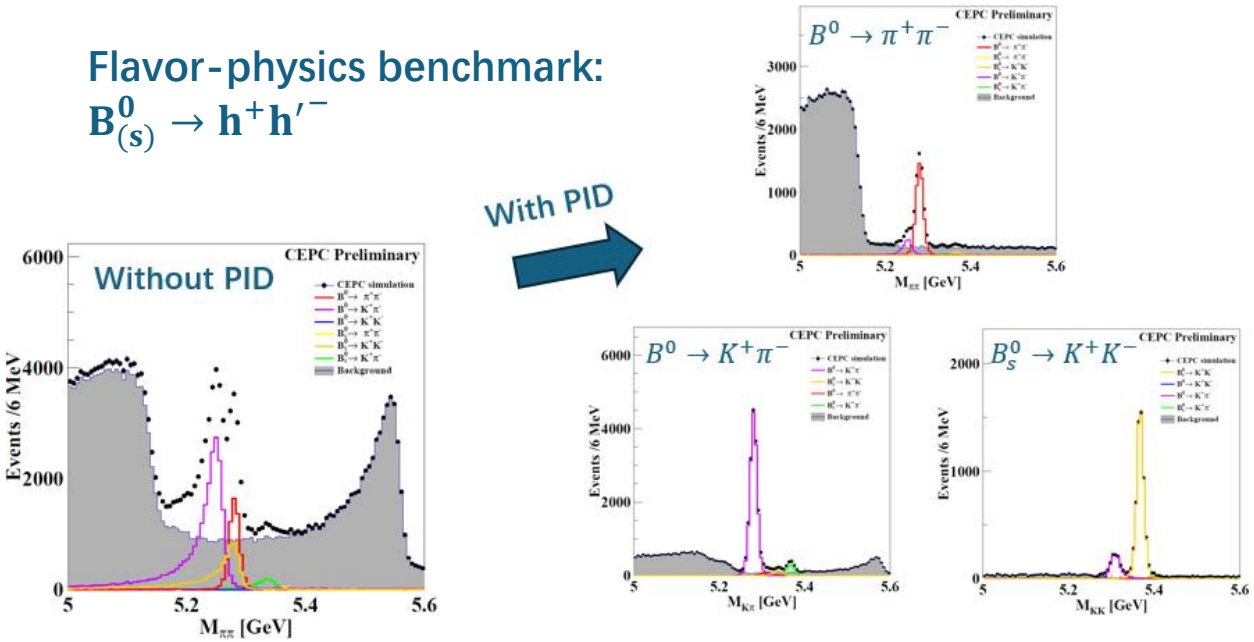
Guang Zhao zhaog@ihep.ac.cn

On behalf of CEPC drift chamber group

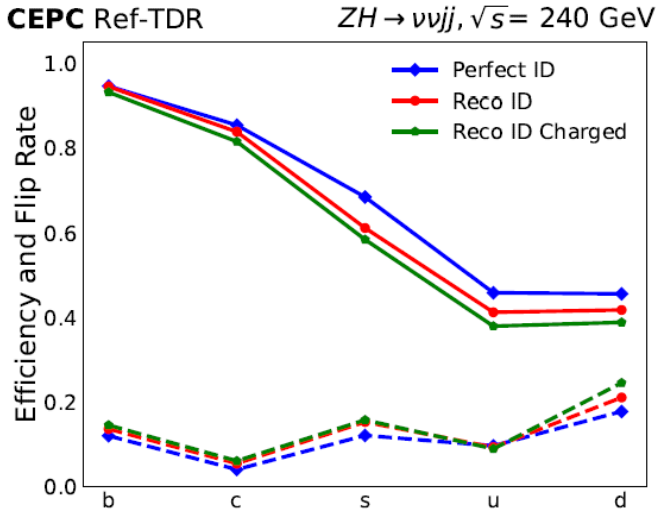
Motivation: Particle identification

- **PID is essential for high-energy physics experiments. In the context of future e^+e^- 'Higgs factory' studies, PID has garnered significant attention:**
  - The Tera-Z at CEPC/FCC-ee offers vast flavor-physics opportunities, necessitating robust hadron identification across a broad momentum range.
- **There is a growing recognition of the performance gains that PID provides for jet-flavor tagging.**

Flavor-physics benchmark:
$\mathbf{B^0_{(s)} \to h^+ h'^-}$

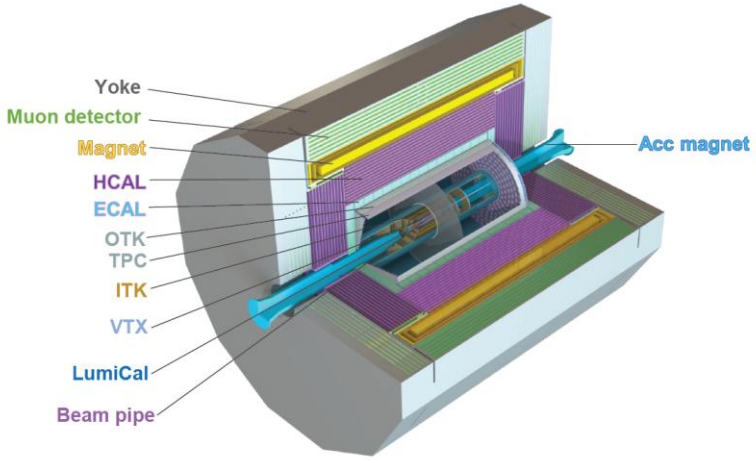


Jet tagging

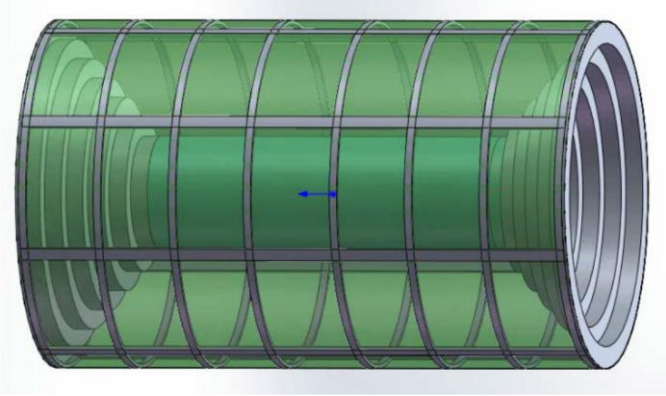


The CEPC drift chamber

AGORA detector for CEPC



CEPC drift chamber



A transparent drift chamber, placed in between silicon trackers, serves as an alternative gaseous detector for CEPC, providing:

- Precise low-momentum tracking
- **Excellent hadron ID**

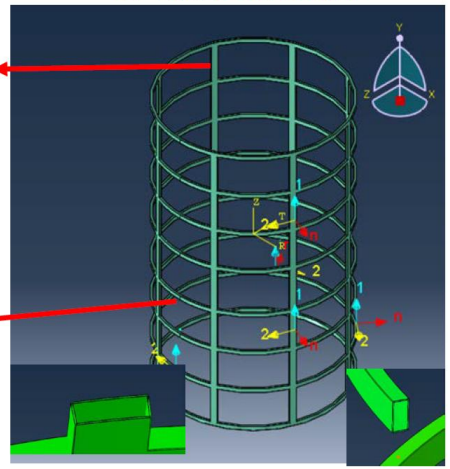
Design parameters:

- Layout:
  - R extension: 600-1800 mm
  - Length: 5800 mm
- Gas mixture:
  - He : iC_4H_{10} = 90 : 10
- Cell:
  - Cell size: 18 x 18 mm^2
  - 64 layers and 27623 cells
  - Field to sense wire ratio 3: 1

Preliminary mechanical design

Overall design

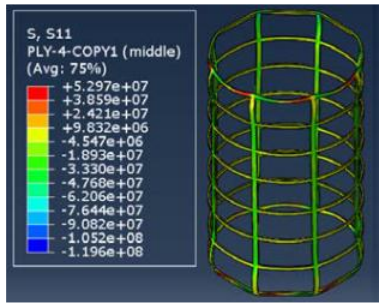
CF frame structure

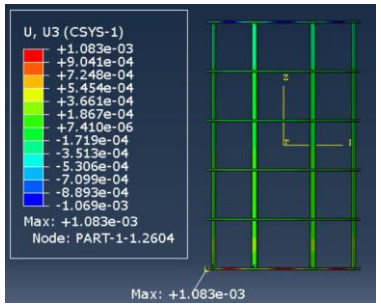


- **CF frame structure:**
  - 8 longitudinal + 8 annular hollow beams
  - inner and outer CF cylinder
- **6-step aluminum endplate:**
  - Thickness: 20 mm
  - Weight: 880 kg

Finite element analysis

	Max. stress	Max. deformation
CF frame	235 MPa	1.1 mm
Endplate	30 MPa	2.7 mm





Wire tension

- **Field wires:** 60um diameter, Al coated with Au
- **Sense wires:** 20um diameter, W coated with Au
- **Total tension:** ~ 10t, meet electrostatic stability condition

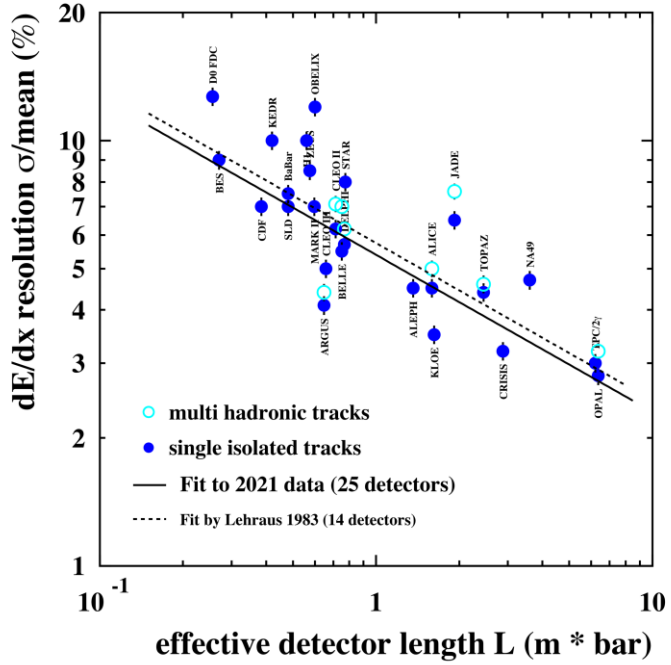
	Cell number/step	Length	Single sense wire tension (g)	Single field wire tension (g)	Total tension/step (kg)
	2684	4000	43.29	66.52	651.78
	3452	4360	51.43	79.03	995.95
	4220	4720	60.28	92.62	1426.88
	4988	5080	69.82	107.29	1953.63
	5756	5440	80.07	123.03	2585.27
	6524	5800	91.02	139.85	3330.85
Total	27623				10944

PID in drift chambers: Classical dE/dx

- **"mean" energy loss (Bethe-Bloch):**
  - $< dE/dx > = \xi \times 1/\beta^2 \times Q^2 \times [K + \ln Q^2 + \ln \gamma^2 - \beta^2 - \delta(\beta)]$
  - PID principle: Fixed momentum → distinct $\beta\gamma$ → separable dE/dx

- **Limitations:**
  - Landau distributed due to secondary ionizations (loss 20-30% statistics by truncated mean)
  - Large fluctuations due to cluster size, gas gain …
  - **dE/dx resolution has not improved significantly over the last four decades**

Lehraus plot
(from Michael Hauschild)

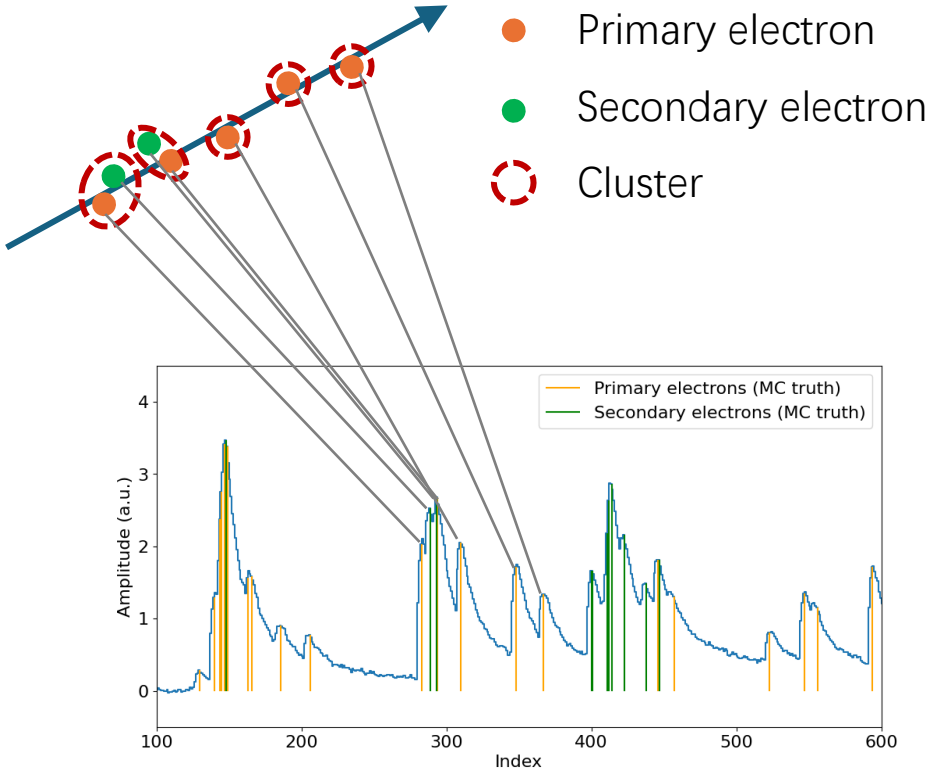


dE/dx resolution (L=1m):

- 1983: $5.7 \times L^{-0.37}$(%)
- 2021: $5.4 \times L^{-0.37}$(%)

PID in drift chambers: Cluster counting (dN/dx)

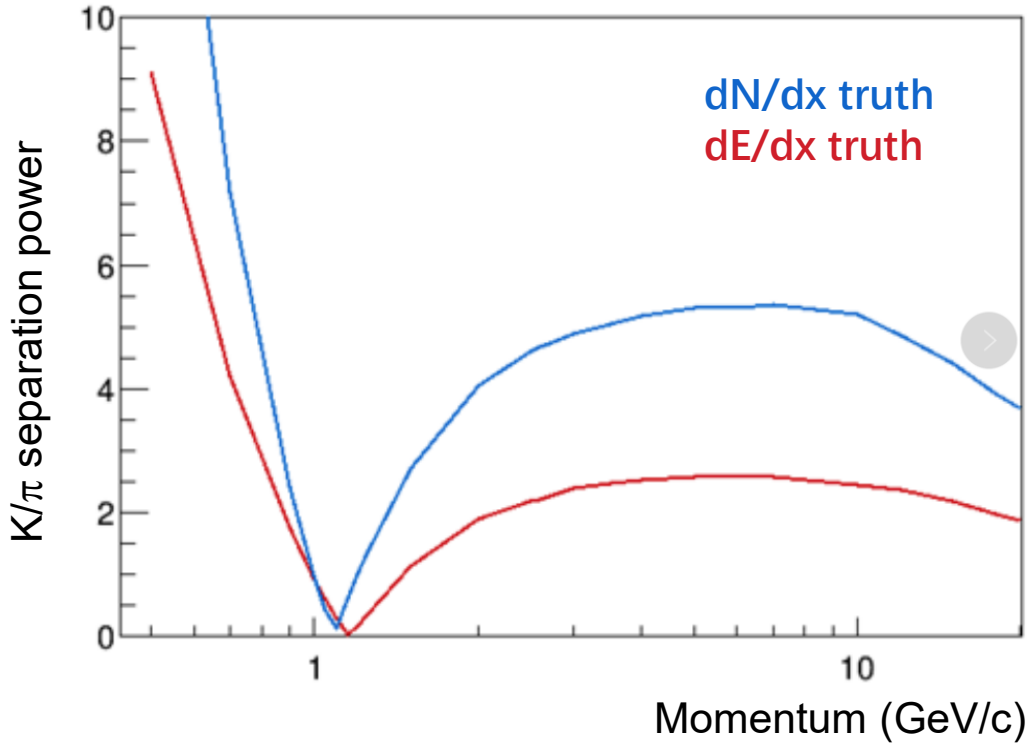
- **Count the number of primary ionizations:**
  - Poisson distributed
  - Avoid most fluctuations associated to dE/dx ➔ Significantly improve the resolution
- **Challenges:**
  - Hardware: High bandwidth and high sampling rate front-end electronics are required
  - **Software: Reconstruction is a main challenge**



A DC cell waveform

dN/dx vs. dE/dx

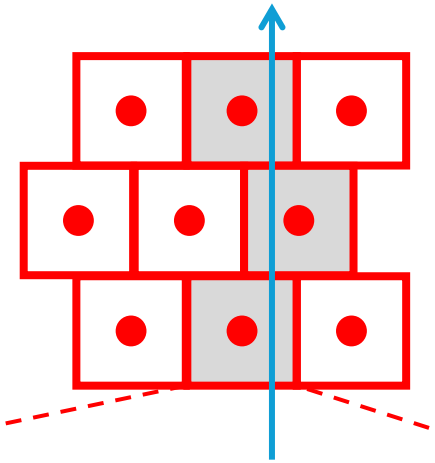
- **Particle separation power:**
  - Definition: $\frac{\text{separation}}{\text{resolution}} = \frac{|\mu_A - \mu_B|}{\sqrt{\sigma_A^2 + \sigma_B^2/2}}$
- **Typical K/π separation power:**
  - dE/dx: > 2σ up to 2...20 GeV/c
  - dN/dx: > 3σ up to 2...20 GeV/c



dN/dx has much better separation power than dE/dx
dN/dx is a potential breakthrough in PID

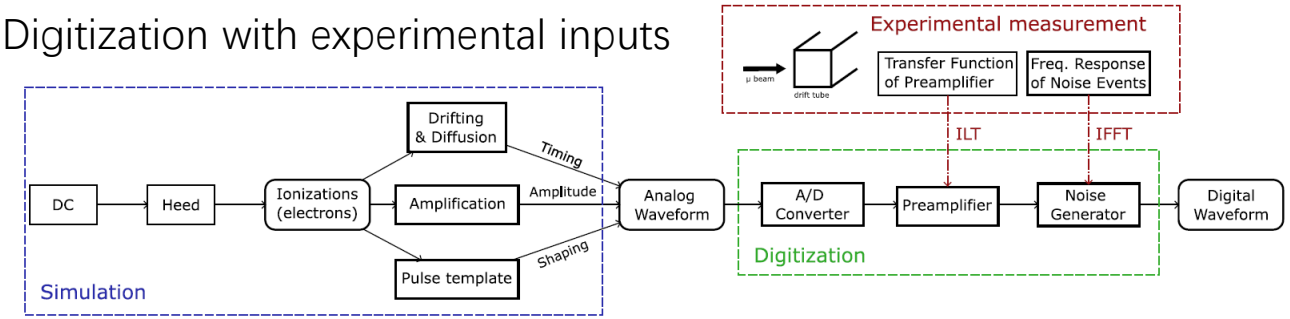
Simulation study of PID performance

Multi-cell full simulation package



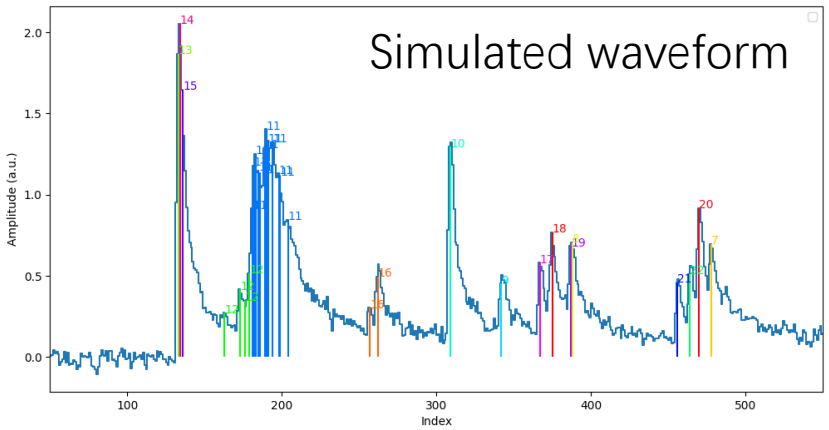
Single-cell waveform simulation:

- Garfield++-based ionization simulation
- Digitization with experimental inputs



Simulation setup:

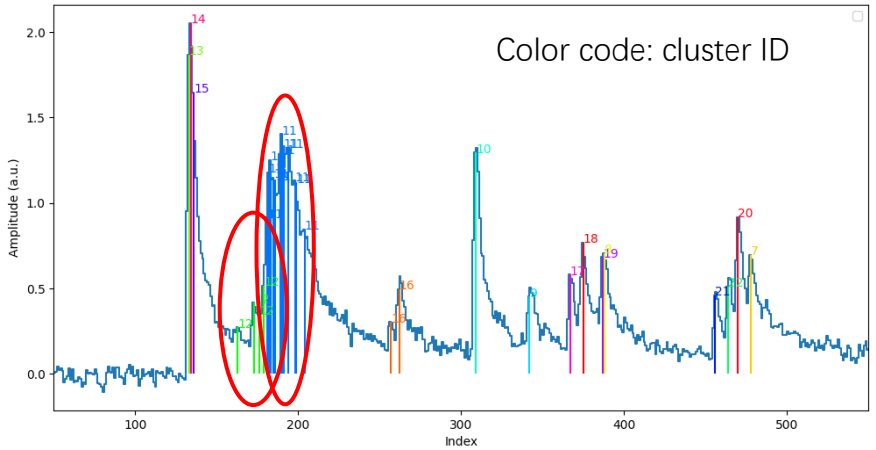
- Gas mixture: He/iC_4H_{10} (9:1)
- Sampling rate: 1.5 GHz
- High-freq. cutoff of preamplifier: ~0.9 GHz
- Noise ratio: ~5%



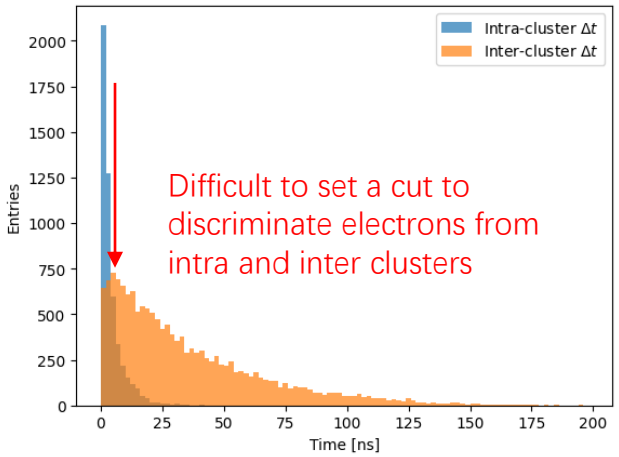
dN/dx reconstruction

- **2-step algorithm:**
  - Peak-finding: Detect peaks associated to all electron signals
  - Clustering: Identify primary clusters from detected electron peaks

- **Challenges:**
  - Electronic noises
  - Pile-up
    - Between electron pulses
    - Between clusters



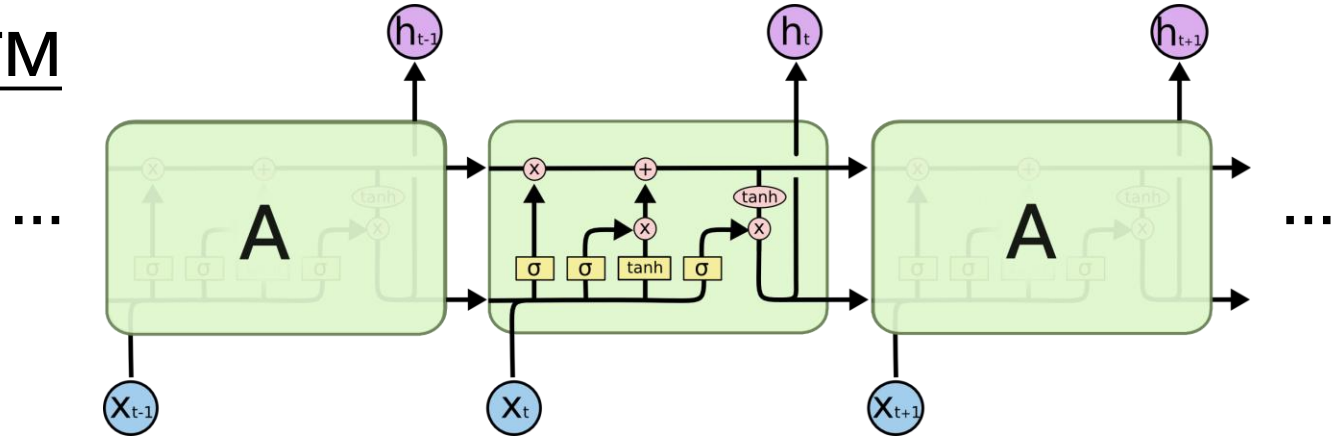
Cluster 11 & 12
are overlapped



Need deep learning for efficient reconstruction

Supervised learning for dN/dx reconstruction

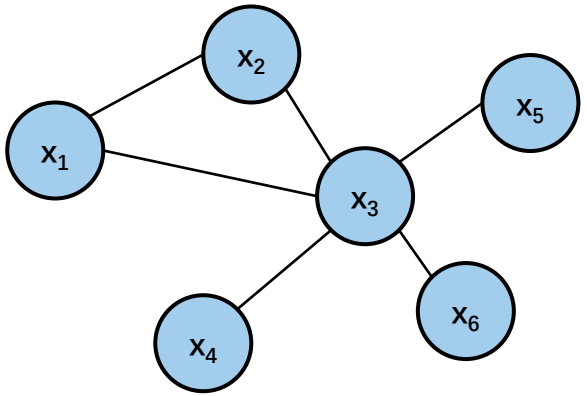
LSTM



LSTM-based peak finding:

- Can efficiently handle time-sequence
- Waveform slices as the LSTM input
- Binary classification of signals and noises

DGCNN



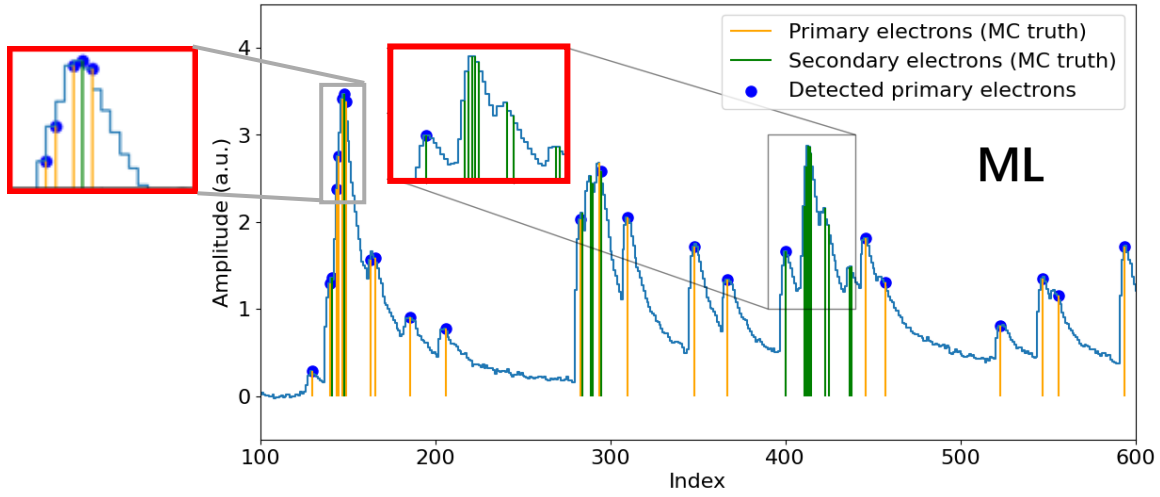
Dynamically connected Graph by k-NN

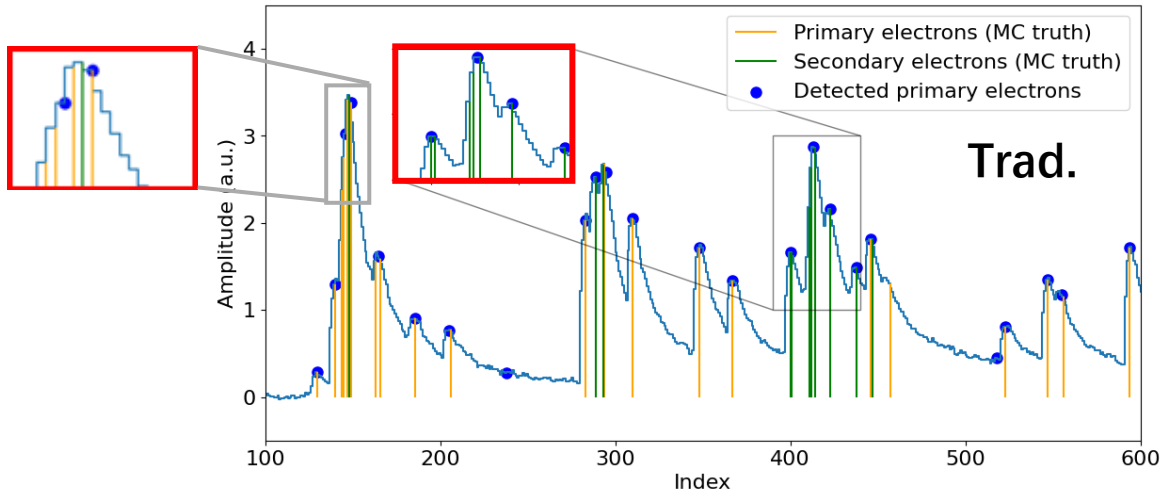
DGCNN-based clustering:

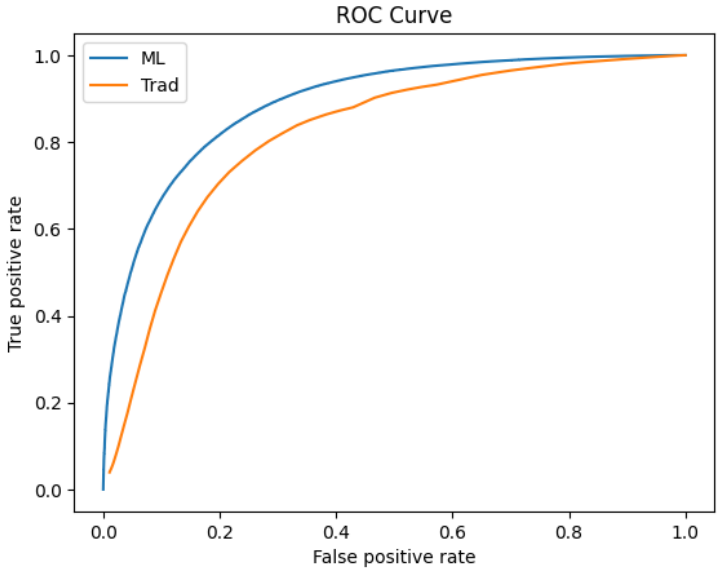
- Incorporate local information to learn global properties
- Detected timings from the peak-finding as the DGCNN input
- Binary node classification of primary and secondary electrons

Comp. Phys. Comm. 300 (2024) 109028
Nucl. Sci. and Tech. (2025) 36:113

Reconstruction results







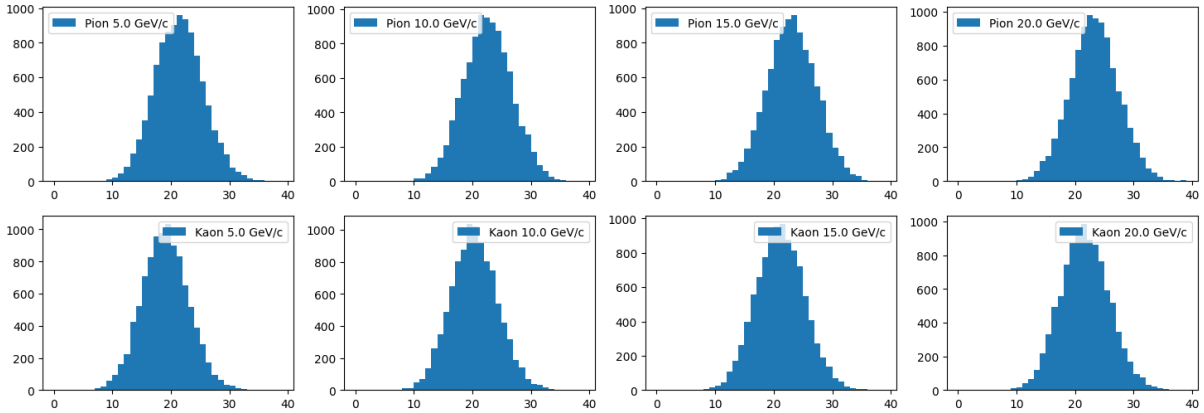
	Purity	Efficiency
ML	0.8986	0.8820
Trad.	0.8986	0.6827

The ML model is much more efficient for

- Pileup recovery
- Secondary ionization suppression

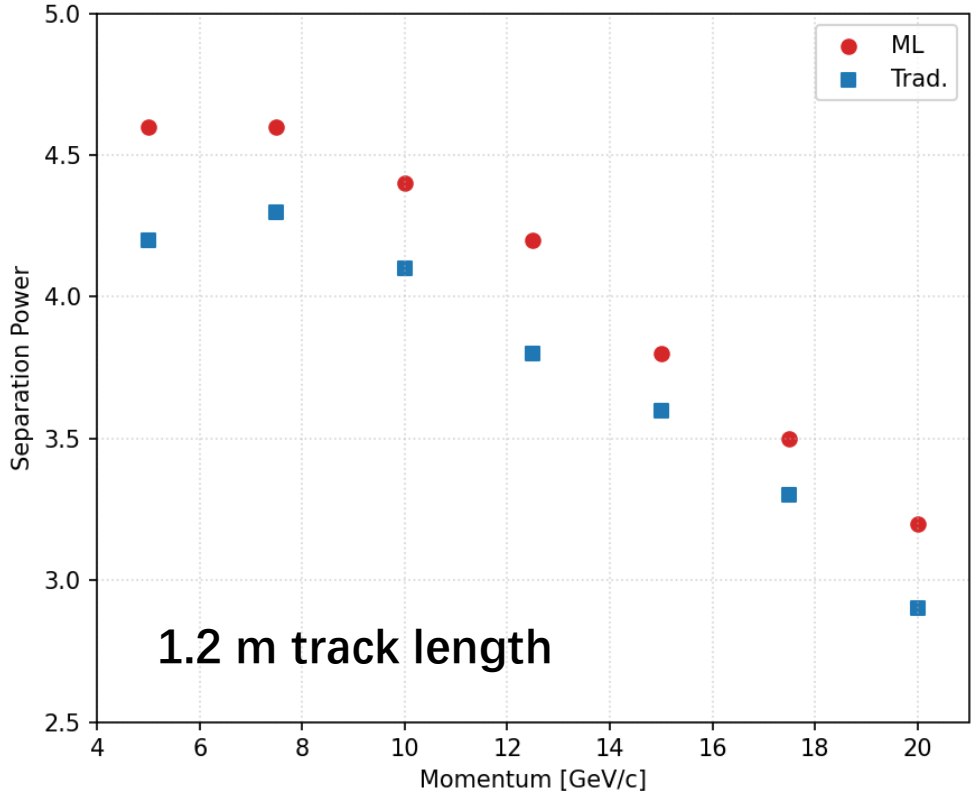
PID performances with supervised models

Reconstructed # of clusters distributions



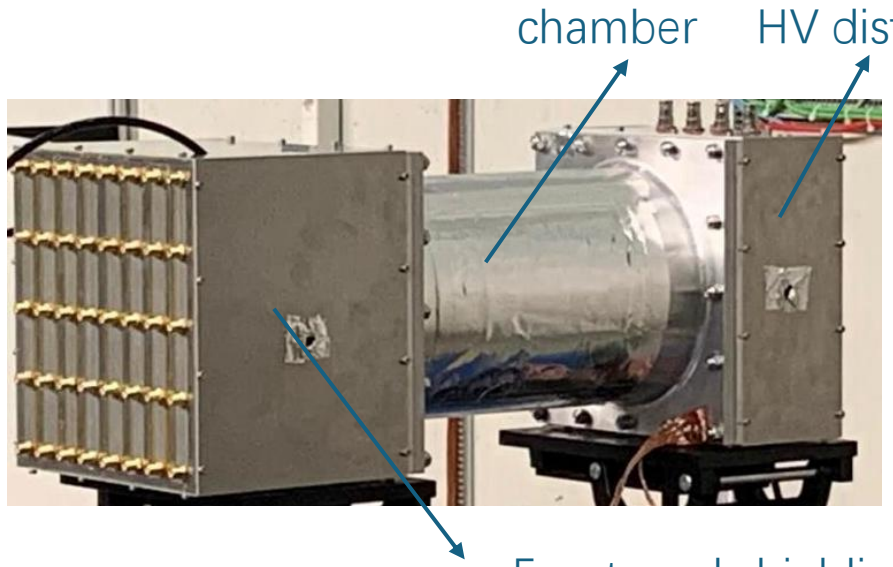
- Very good Gaussians ➔ Efficient secondary electron removal
- For 1.2m track length, dN/dx resolution < 3%, typical ~5% for dE/dx

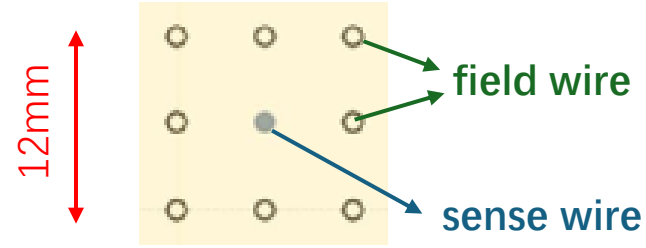
K/π separation power vs. momentum

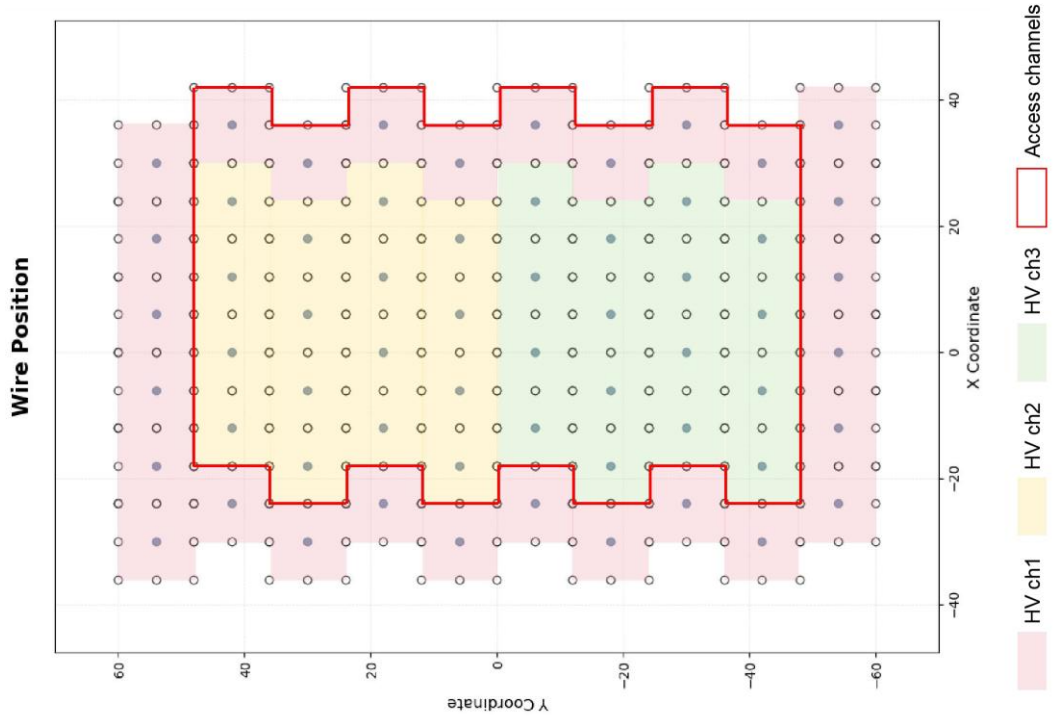


- Achieve K/π separation > 3σ at 20 GeV/c
- ML yields an approximately 10% improvement, equivalent to using a detector with a 20% larger radius with the traditional algorithm.

Detector R&D: Drift chamber prototype





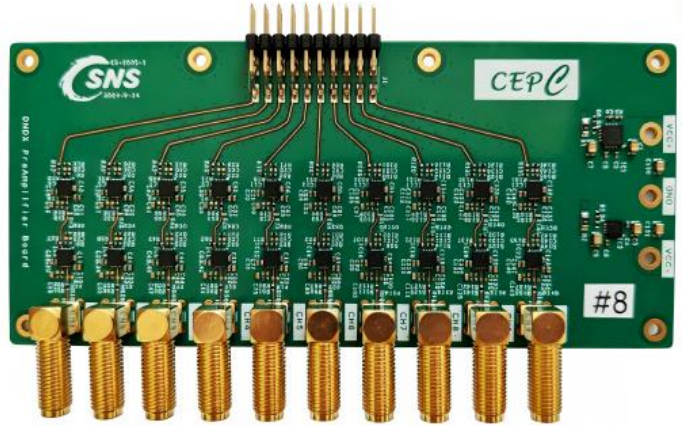


- **Prototype layout:**
  - Cell layout: 10 x 6 cells (8 x 5 connected to readout board)
  - Cell size: 12 mm x 12 mm
  - Length: 250 mm
- **Wires:**
  - 60 sense wires: 20 μm gold-plated tungsten
  - 222 field wires: 70 μm aluminum

Readout electronics & DAQ

- **Current-sensitive preamplifier**
  - Two-stage LMH6629 amplifiers for cascaded amplification
  - Equivalent transimpedance gain: ~1233 Ω
  - Bandwidth: ~601 MHz
  - Baseline noise: ~1.53 mVrms
- **Digitizer**
  - Sampling rate: 1.3 Gsps,
  - Resolution: 14 bits
  - Power consumption: 750 mW/channel
- **DAQ**
  - Transmit data by 10 Gbps UDP
  - Configure electronics by 1 Gbps TCP
  - Software with GUI for electronics configuration and event display

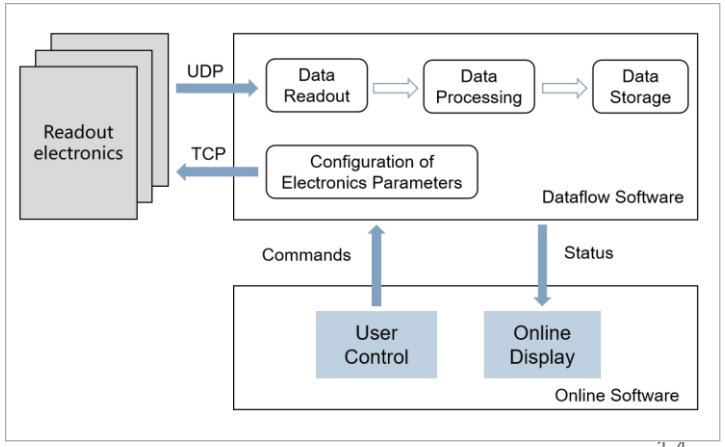
For more detail on the electronics, see Hongbin's upcoming talk



Preamplifier



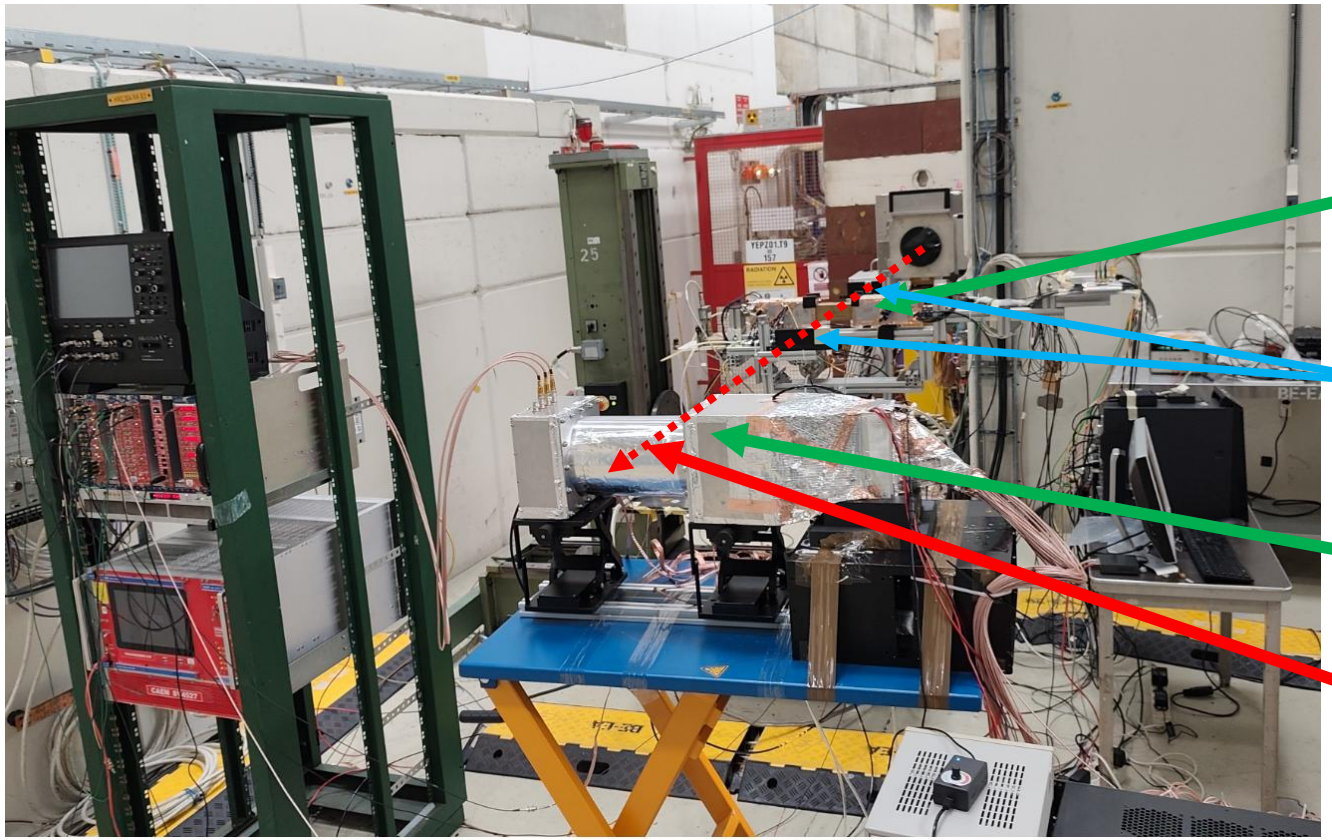
Digitization and FPGA board



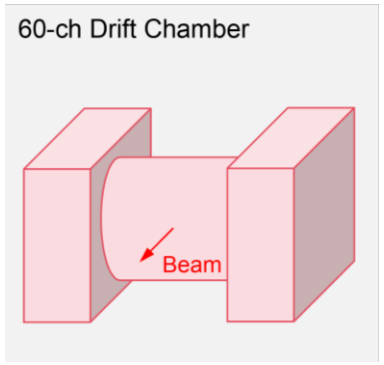
DAQ system

Beam test at CERN

- **Beamline:**
  - CERN PS T9
  - Low momentum particles
- **Time:**
  - May27-June10, 2026
- **Joint tests:**
  - INFN and IHEP groups
  - Two separate detector systems
  - Shared gas distribution and trigger systems



Analysis of 10 GeV/c pion data



beam direction is perpendicular to wires

10 GeV/c pions

Real data

Preprocessing

T0 calibration

X-T calibration

Tracking

L

dN/dx

Neural network

evaluation

N

tuning

Simulation

MC data

training

Tracking

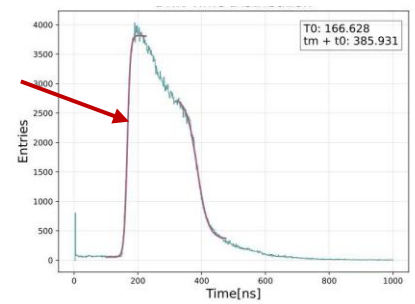
- **Determine T_0 by leading edge fit of raw time distribution**

- **Iterative tracking procedure:**
  - Determine drift time: $T = T - T_0$
  - Determine initial X-T by performing time-distribution integration method
  - Determine drift distance with X-T
  - Perform tracking:
    - Initial parameters: estimated by Hough transform
    - Final parameters: Chi square fit
  - Iteratively update X-T, and do track fitting

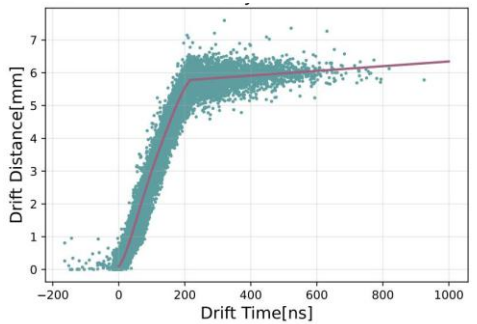
3 iters

Leading edge fit

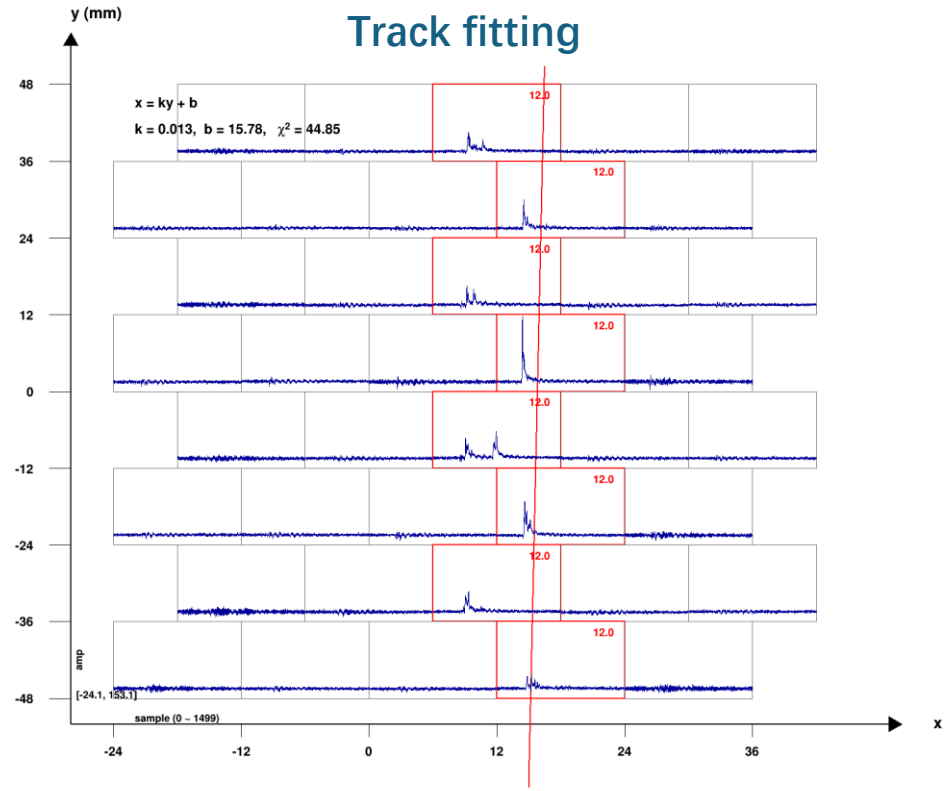
T_0 calibration



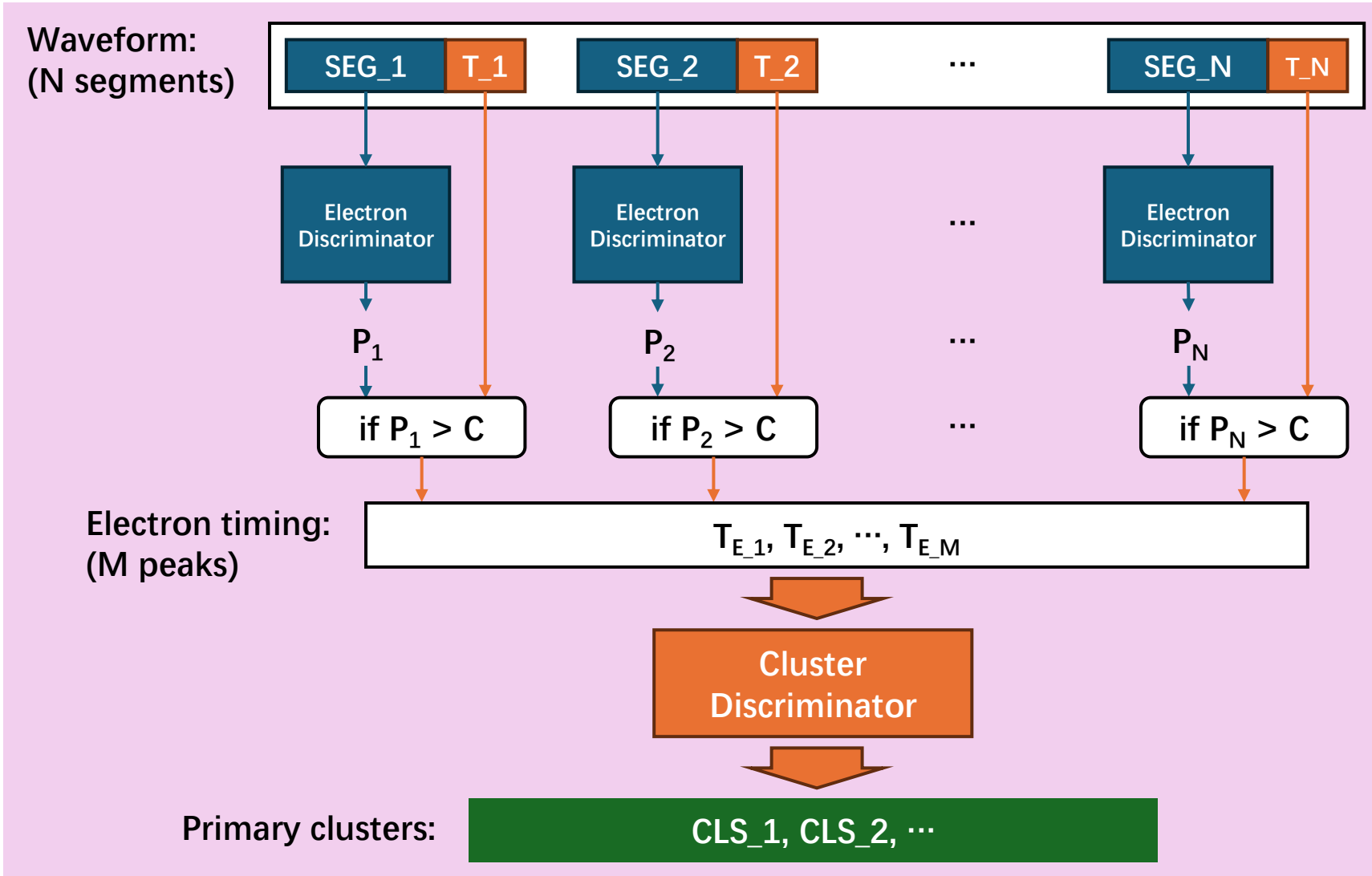
X-T relationship



Track fitting

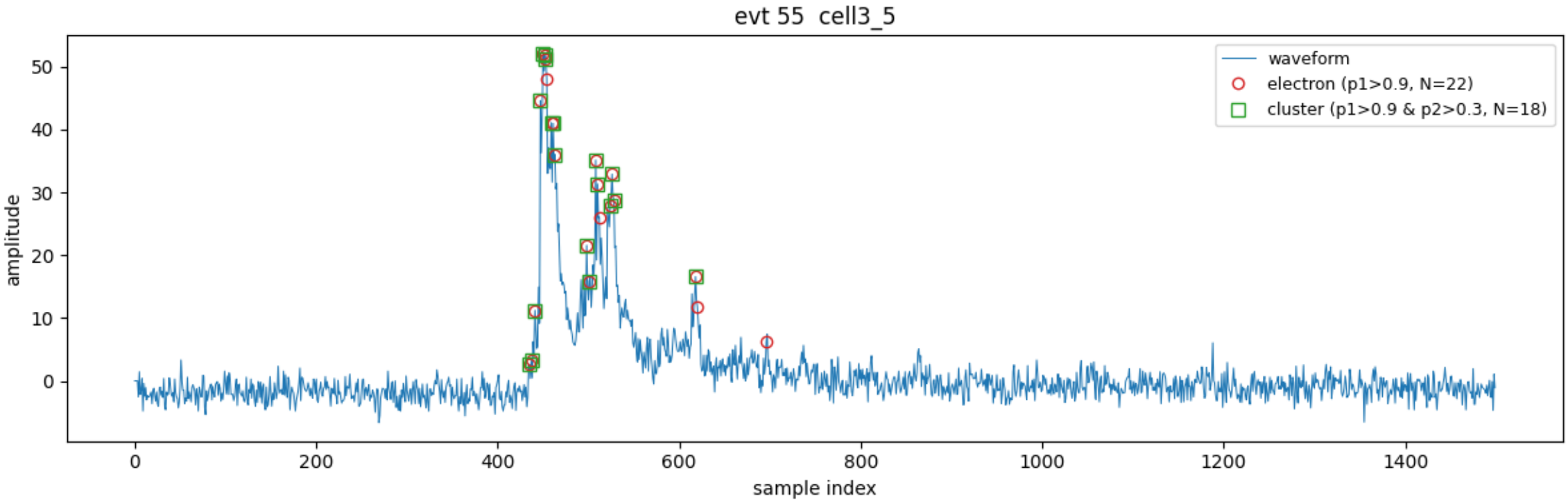


Transformer-based reconstruction



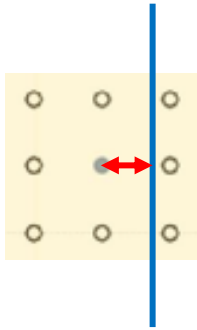
- **Peak-finding (Electron discriminator)**
  - Input:
    - Waveform segments
  - NN architecture:
    - Transformer + Attention Pooling + MLP
  - Output:
    - Signal/noise probability
- **Clustering (Cluster discriminator)**
  - Input:
    - Electron timing list
    - Embedded feature list
  - NN architecture:
    - Transformer (w/ Fourier PE for MLPs) + MLP
  - Output:
    - Primary/secondary electron probability

Apply transformer-based model to real data



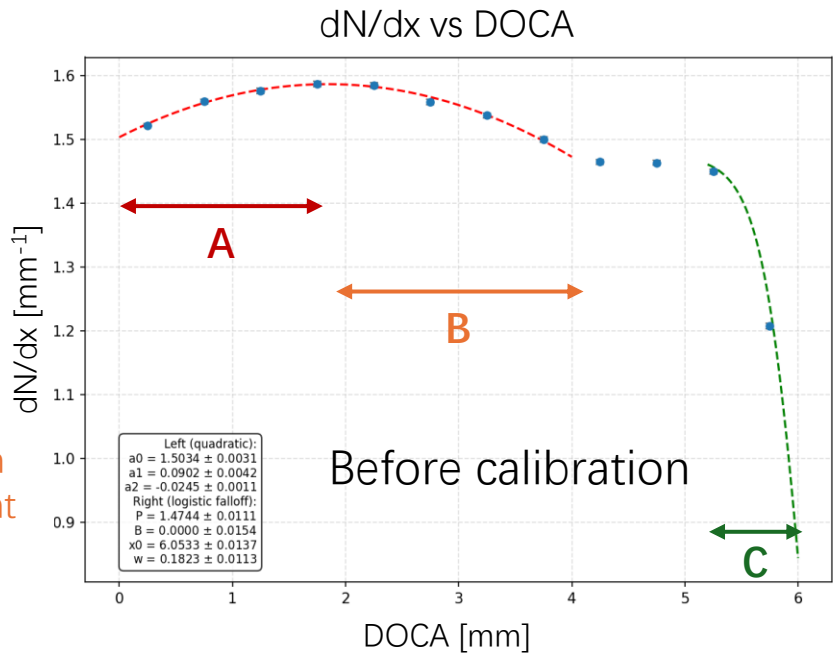
- Detected peaks appear reasonable
- Noise contamination is minimal

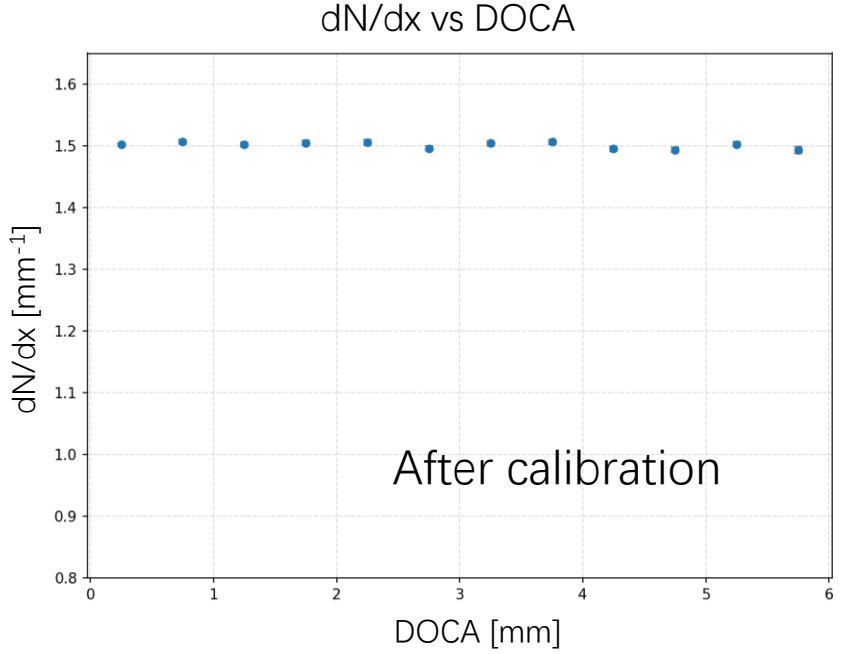
DOCA calibration



DOCA: Distance of Closest Approach

A. Space charge
B. Recombination and attachment
C. Edge effects





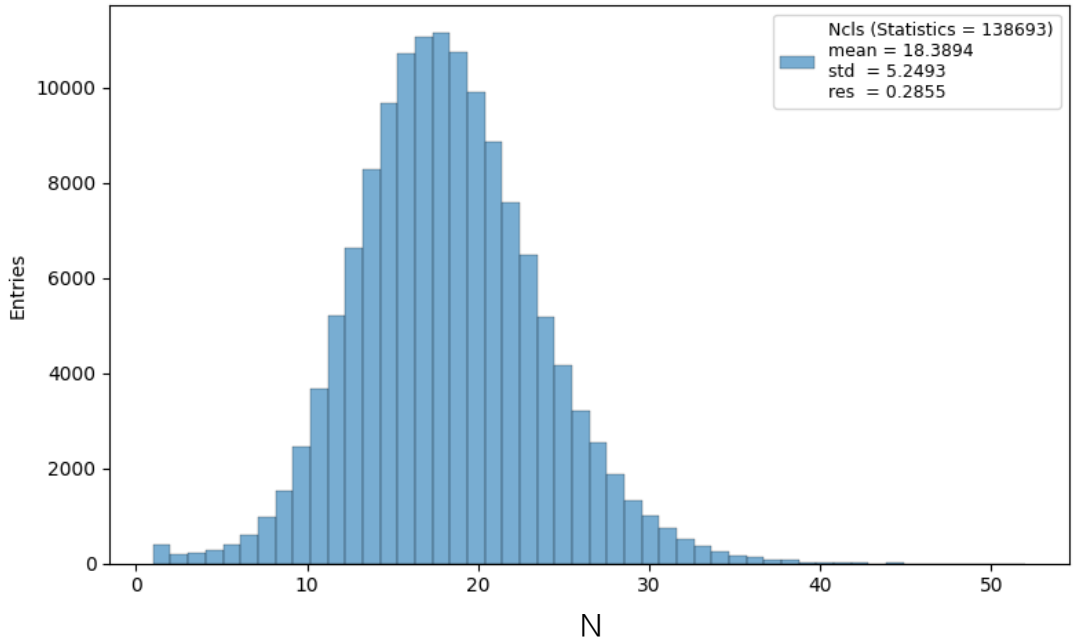
$$f(x) = \begin{cases} a_0 + a_1 x + a_2 x^2, & x < 4.0 \\ a_0 + 4a_1 + 16a_2, & 4.0 \le x < 5.2 \\ B + \dfrac{P - B}{1 + \exp(\frac{x - x_0}{w})}, & x \ge 5.2 \end{cases}$$

DOCA calibration:

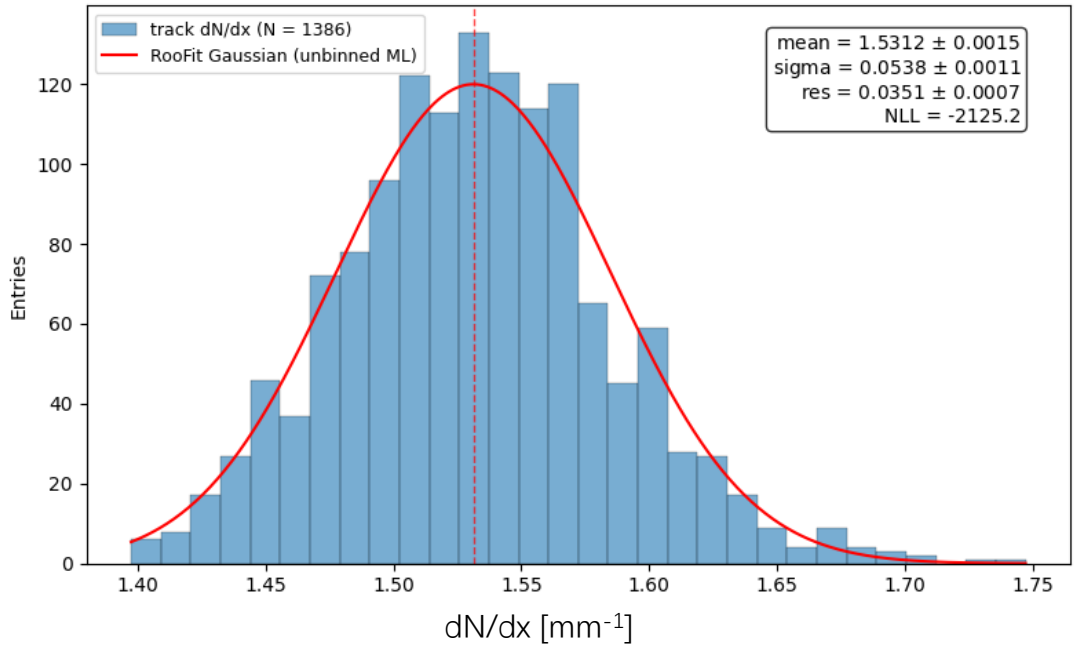
- $dN/dx^{corr} = dN/dx * f(0.0) / f(DOCA)$
- Also require DOCA < 5.2

10 GeV pion dN/dx

of clusters (L=12mm hit)



dN/dx (L=1.2m track)



Resolution results:

- dN/dx hit resolution (L = 12mm): 28.6%
- dN/dx track resolution (L = 1.2m): 3.51%
  - ➢ dE/dx track resolution (L = 1.2m): 3.96%

	Simulation	Beam test
Resolution	**2.6%**	**3.5%**
Space charge, Recombination & attachment	No	Yes
Bandwidth	0.9 GHz	0.6 GHz
Sampling rate	1.5 GHz	1.3 GHz
Noise ratio	5%	>10%

Note: Due to realistic detector responses and differences in electronics parameters, resolution is not as good as simulation

Summary

- **The cluster-counting (dN/dx) technique in drift chambers offers superior PID performance, making it a promising candidate for future e^+e^- collider experiments.**

- **A beam test at CERN T9 was carried out with a small drift chamber prototype. Preliminary data analysis shows good dN/dx resolution.**

- **Next to do:**
  - Improve the detector design and electronics system
  - Perform more beam tests (at BEPCII…)
  - Conduct more detailed data analysis and cross-check the results with the INFN detector data

Acknowledgement

Testbeam @ CERN



We are especially grateful to Prof. Nicola De Filippis and his team for organizing the beam test.

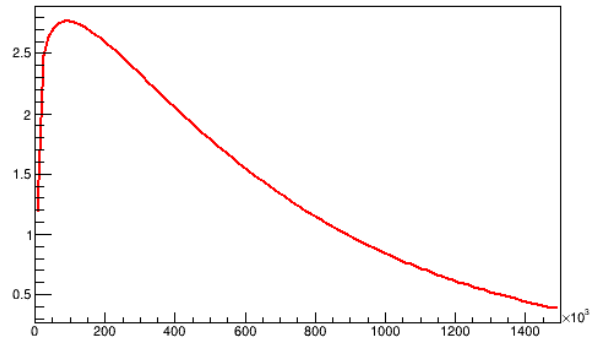
Thank you for your attention!

Backup

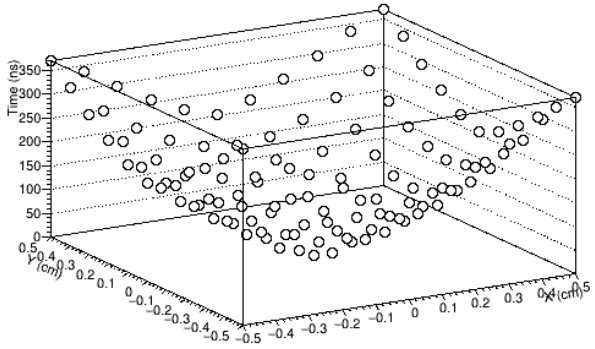
Parameterization updates

- **Setup beam test configuration in Garfield++:**
  - Cell size: 1.2 x 1.2 cm^2
  - Layout: 8 rows, 5 columns
  - High voltage: 1775 V
- **Extract parameters from full Garfield++ simulation**
  - Amplitude: N_{aval} distribution, amplitude = f(N_{aval})
  - Timing: drift time = g(x, y), drift time error = h(x, y)
- **Extract noise from beam test noise waveforms**
  - FFT mag. from data + random phases
  - Perform inverse FFT

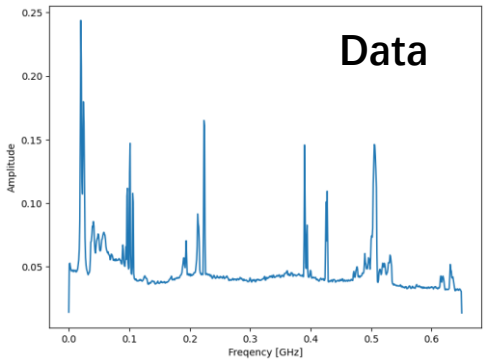
N_{aval}

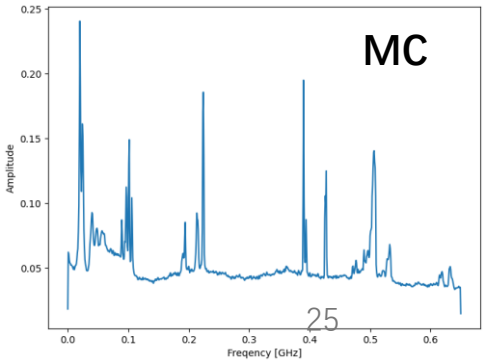


Drift time = g(x, y)



Noise FFT mag.



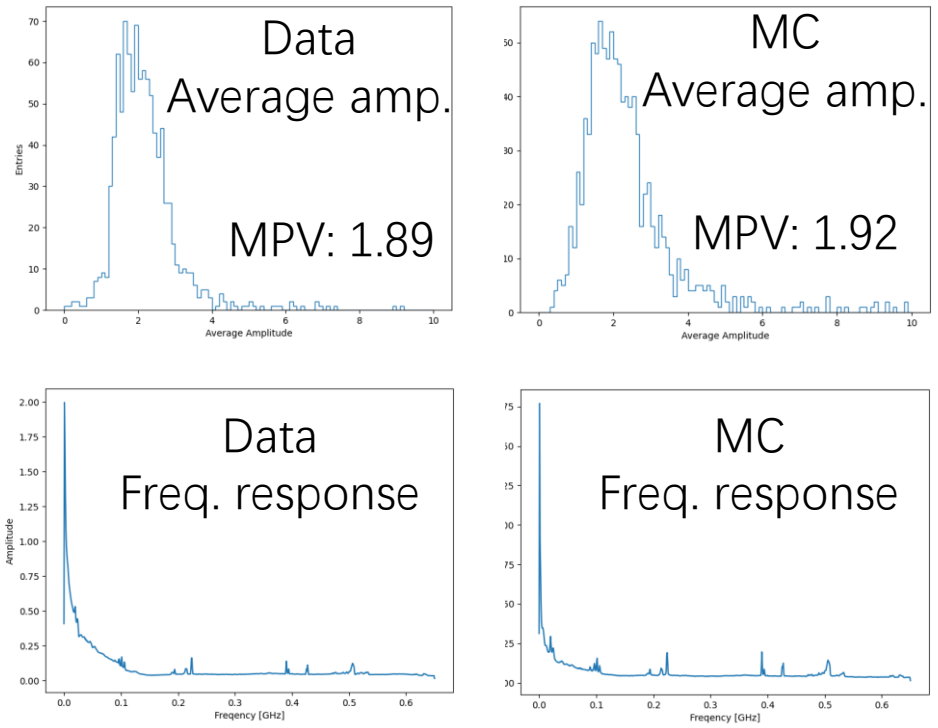


Simulation tuning

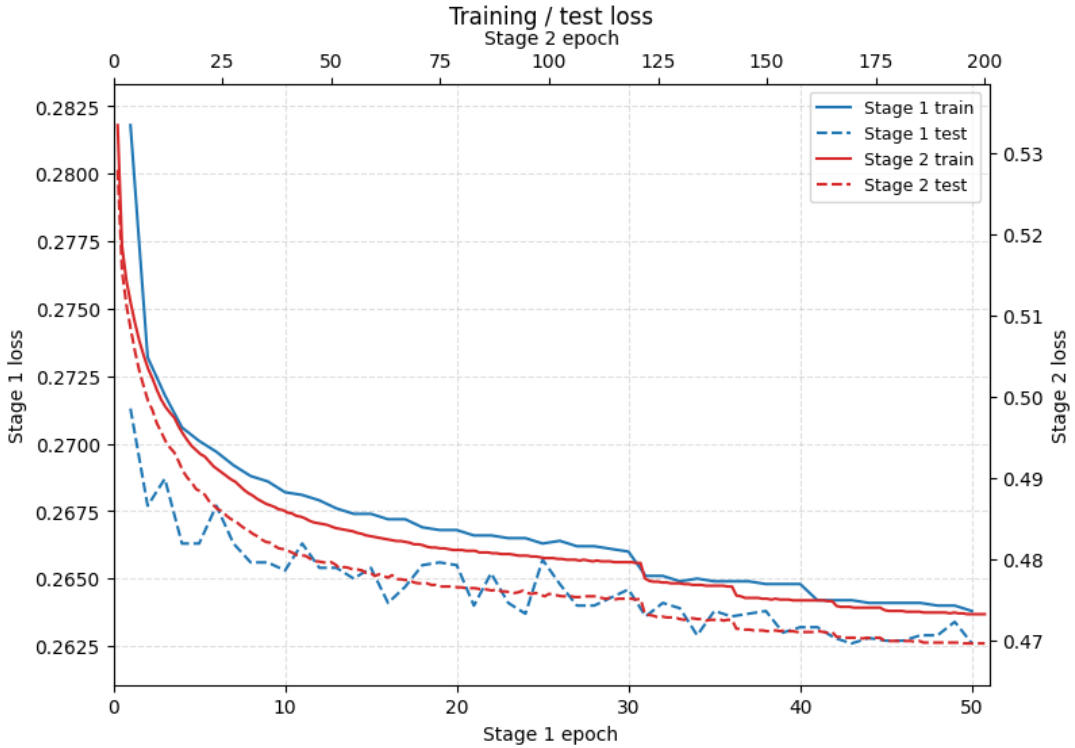
- **Metrics:**
  - Ratio of average amplitude in time domain:
    - $R_{amp} = A_{data}/A_{MC}$
  - Pearson correlation coefficient between frequency responses:
    - $P_{freq} = \frac{\Sigma(x-m_x)(y-m_y)}{\sqrt{\Sigma(x-m_x)^2\Sigma(y-m_y)^2}}$, where x, y are frequency responses from data and MC, m_x and m_y are their mean values

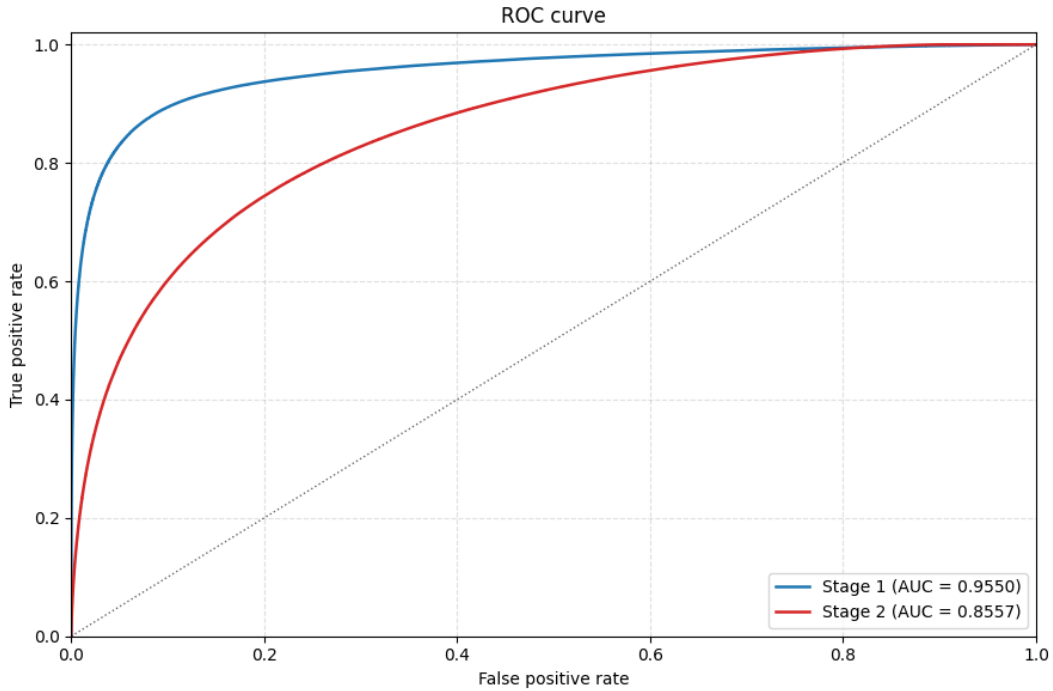
- **Optimization procedure:**
  - Iteratively scale the waveform amplitude and transfer function to obtain the best P_{freq}

After 3 iterations



NN training





- Training is stable with the AdamW optimizer and ReduceLROnPlateau scheduler
- The stage2 AUC is smaller than stage1 AUC because clustering is more challenging. However, they are not as good as previous CEPC study because the NR is 3 times higher.