Institute of High Energy Physics
Chinese Academy of Sciences

PID performance study with Drift Chamber at IHEP

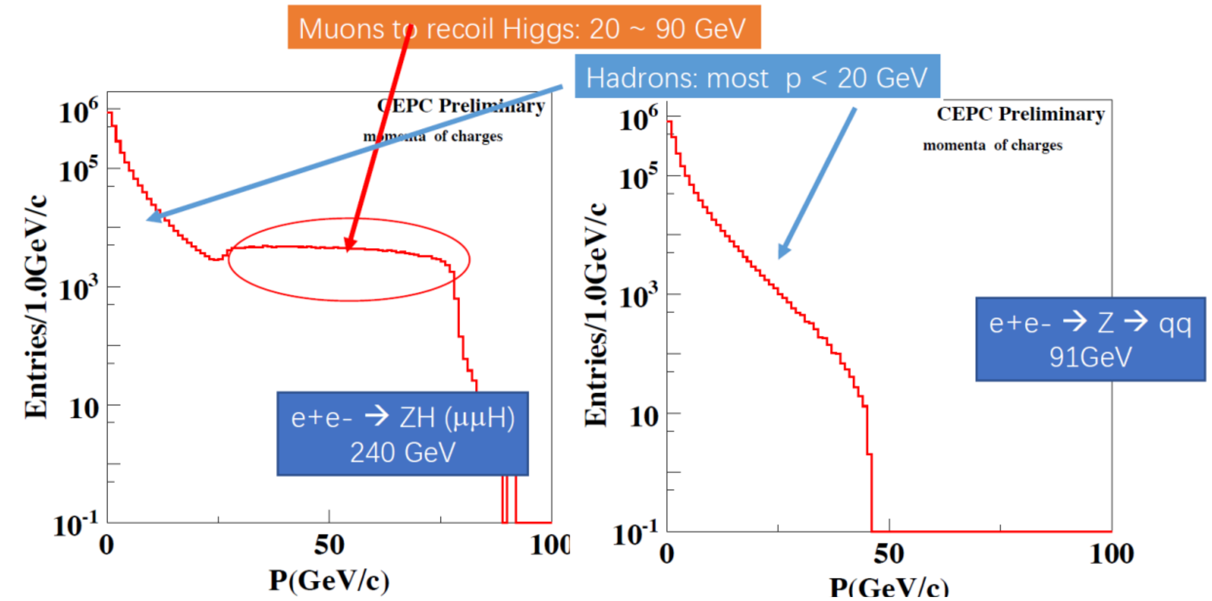
Shuiting Xin
For CEPC the 4th conceptual drift chamber working group

The 2022 International Workshop on the High Energy Circular Electron Positron Collider
24 - 28 October 2022

Motivations

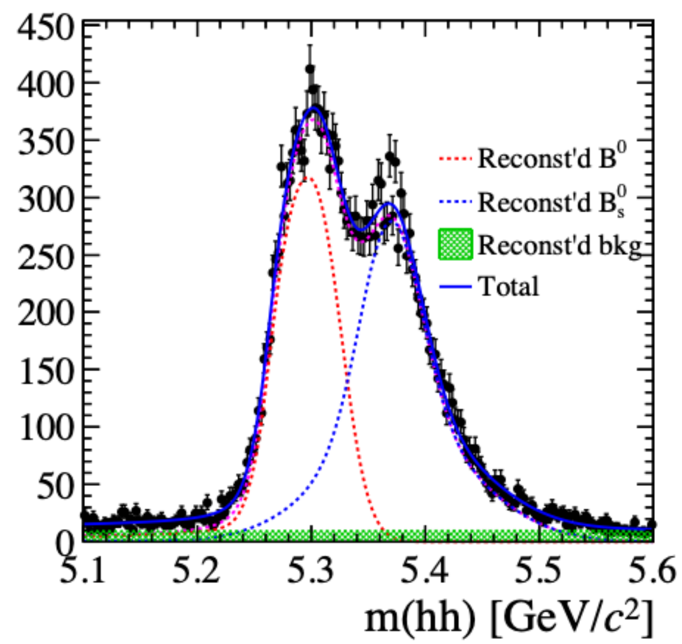
- CEPC is proposed as not only a Higgs, but also a W/Z factory.
- Especially at Z-pole, ~10^{12} Z bosons offer great opportunity for flavor physics and jet study etc.
- Giving the fact that produced charged hadrons are concentrated on low momentum region, particle identification (PID) would be critical.

Operation mode	Z factory	WW threshold scan	Higgs factory
$\sqrt{s}$ (GeV)	91.2	158 – 172	240
Running time (years)	2	1	7
L (10^{34} cm^{-2}s^{-1})	17 – 32	10	3
Integrated Luminosity (ab^{-1})	8 – 16	2.6	5.6
Higgs yield	–	–	10^6
W yield	–	10^7	10^8
Z yield	10^{11-12}	10^8	10^8

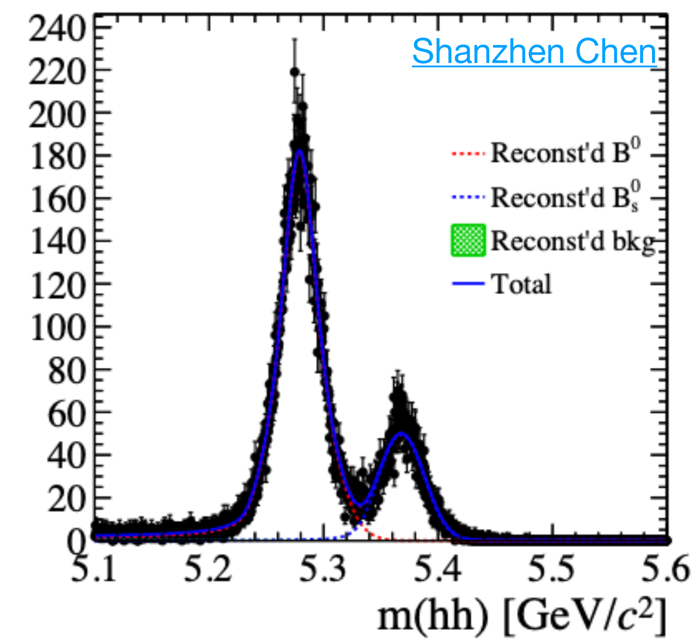


Example of the impact of PID in heavy flavor decay reconstruction

Disentangle the various $B_s^0(B^0) \to h^-h^+$ in same topology final-states.



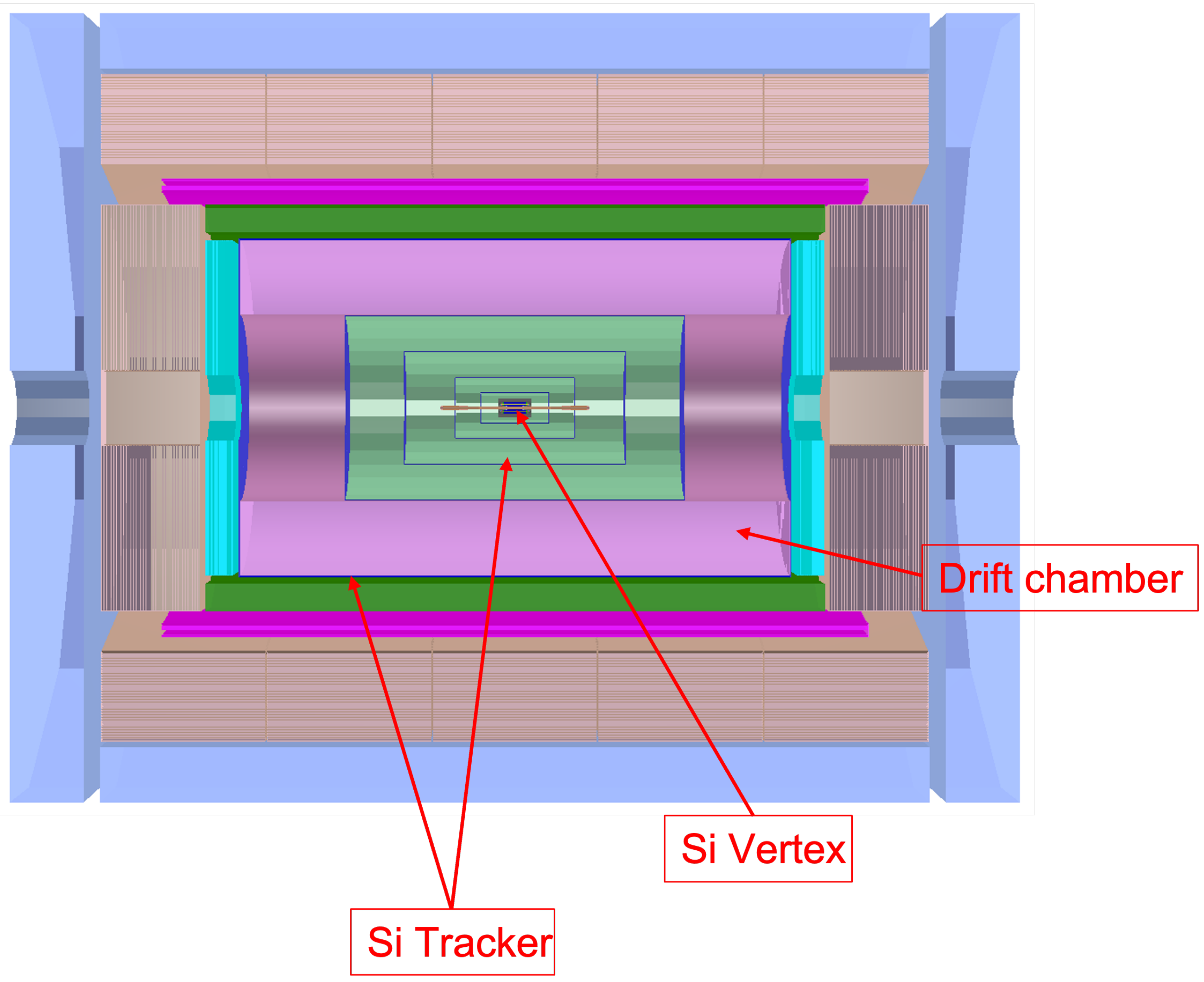
Assuming 50% PID efficiency



Assuming 100% PID efficiency

Motivations for Drift Chamber

- Drift Chamber is a design in 4th conceptual detector at CEPC.
  - between the Full Silicon Tracker layers for particle identification (PID).
  - optimized for PID by applying the cluster counting technique.



<u>cluster counting technique</u>: measuring the clusters directly allows a significant improvement with respect to dE/dx method

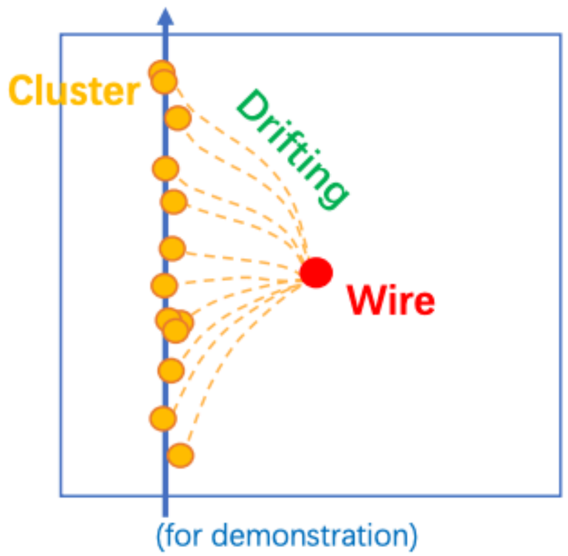
- Main goals are to provide
  - sufficient PID separation power for hadron momentum up to 20 GeV/c
  - good tracking performance combined with silicon tracker

Contents

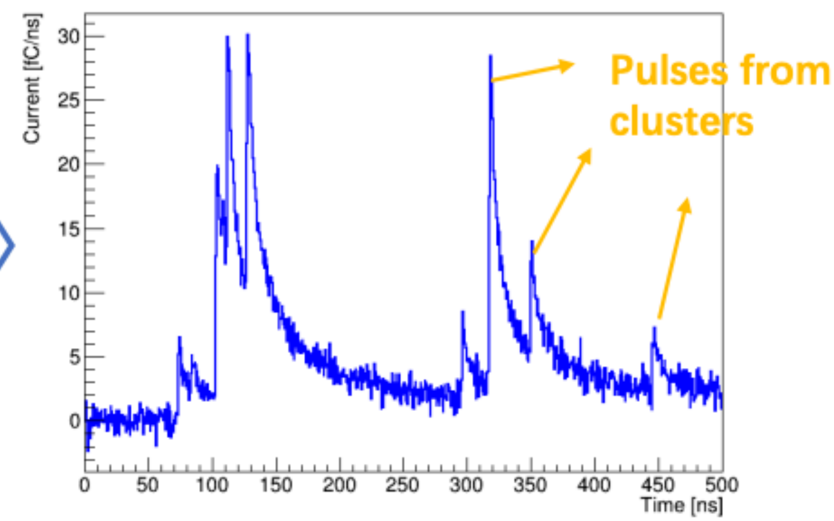
- Energy loss and cluster counting
- Simulation study and expected performance
- Counting algorithm with machine Learning

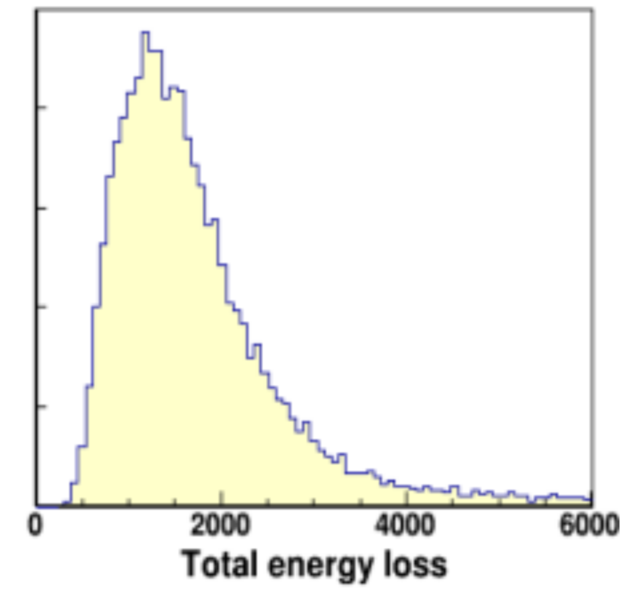
Ionization measurement: dE/dX and dN/dX

DC cell

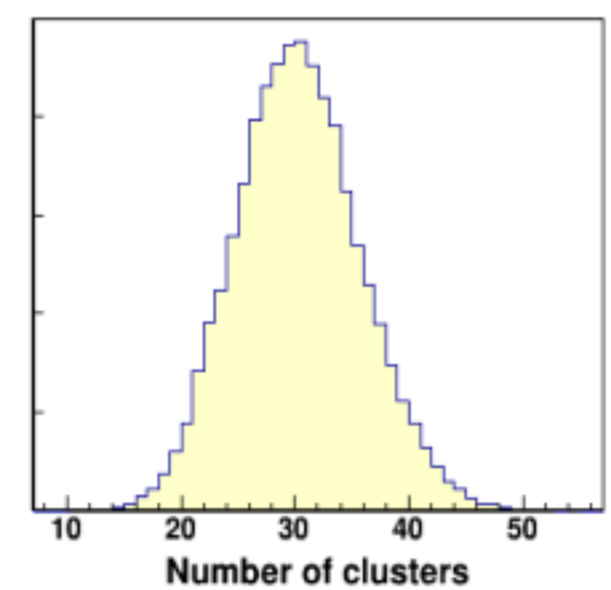


Induced current





dE/dx



dN/dx

For demonstration

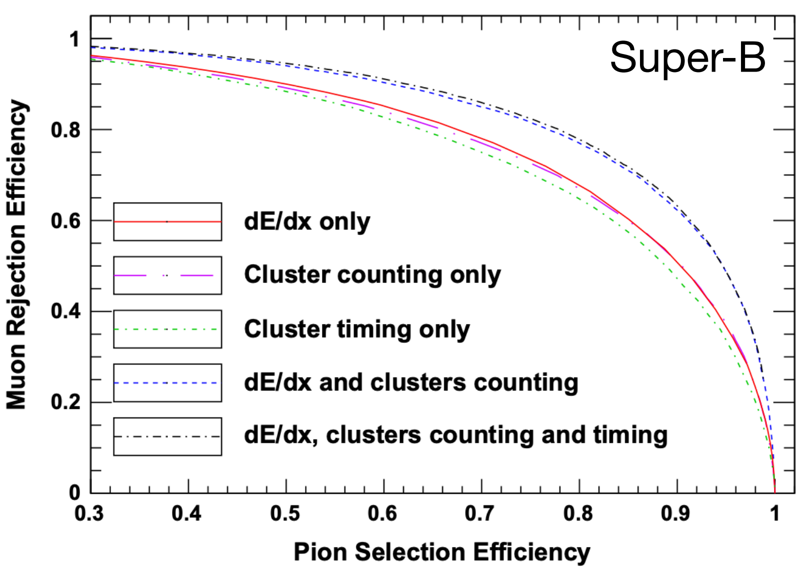
- Energy loss (dE/dx):
  - Fluctuations of number of clusters (N_p) + transferred energy result in Landau distribution.
  - Lost information due to mean truncation.
- Cluster counting (dN/dx): Measure number of clusters
  - Clean in statistics, only depends on N_p :

$$P(N_P, k) = \frac{N_P^k}{k!} e^{-N_P}$$

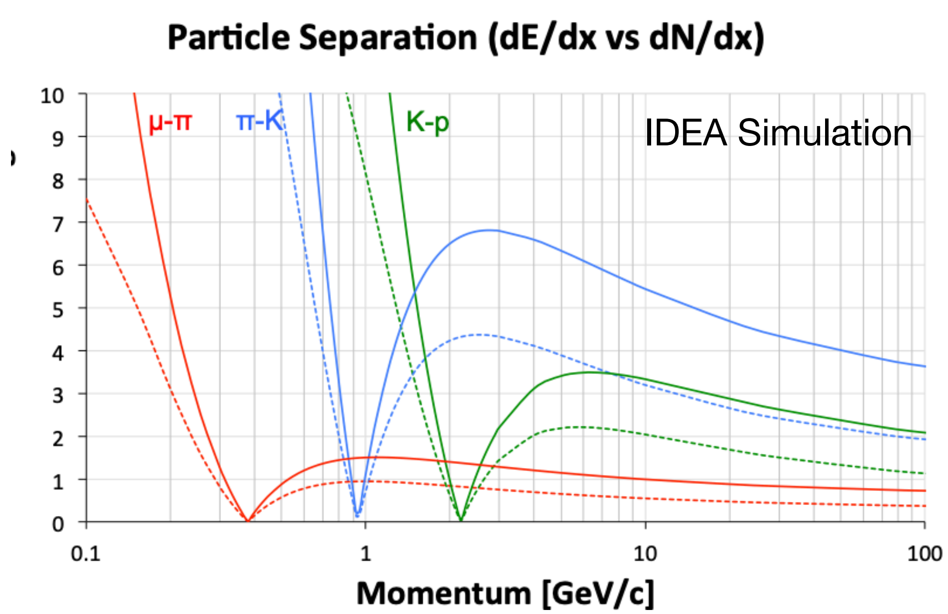
  - Theoretical resolution: $\frac{1}{\sqrt{N_P}}$ potential gain of a factor of 2

Researches on cluster counting

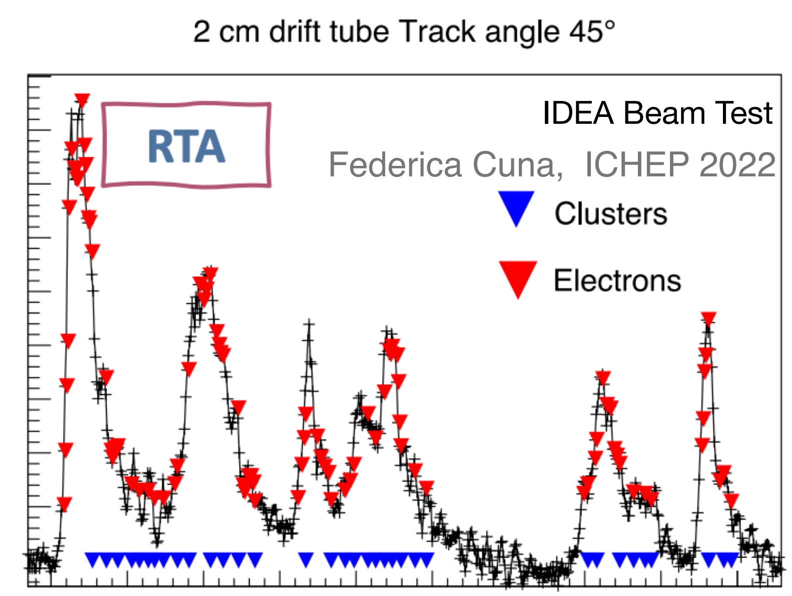
- Proposals for e^+e^- colliders:
  - Detector for Super-B
    - Done with full track length drift chamber prototype. 90% He 10% iC4H10 gas.
  - IDEA detector for FCC-ee or CEPC
    - Assuming 80% counting efficiency and 4.2% dE/dx resolution
    - Beam test work are undergoing.



J-F Caron,https://doi.org/10.1016/j.nima.2013.09.028

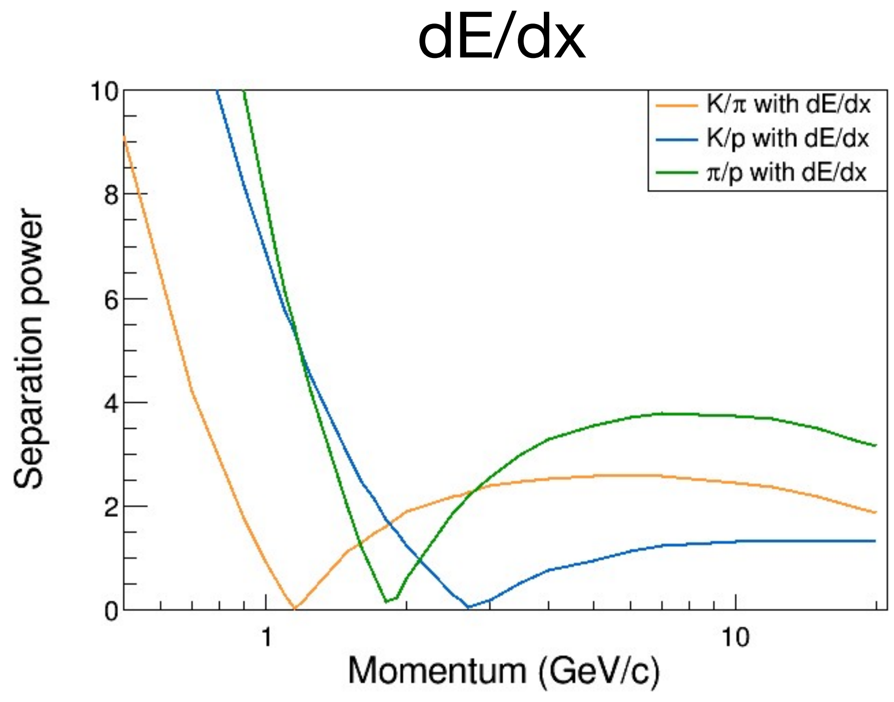


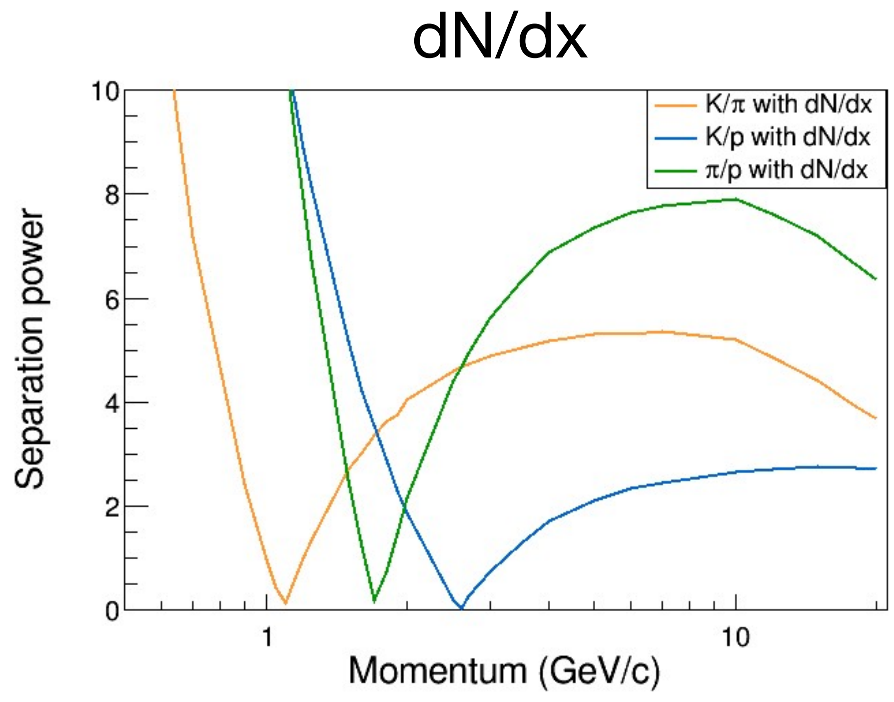
G. Chiarello et. al, NIM A 936 (2019) 503-504



PID with ionization simulation

- The theoretical results are obtained from Garfield++ simulation.
  - The dE/dx applies a truncated mean, the dN/dx counts all the primary ionizations.

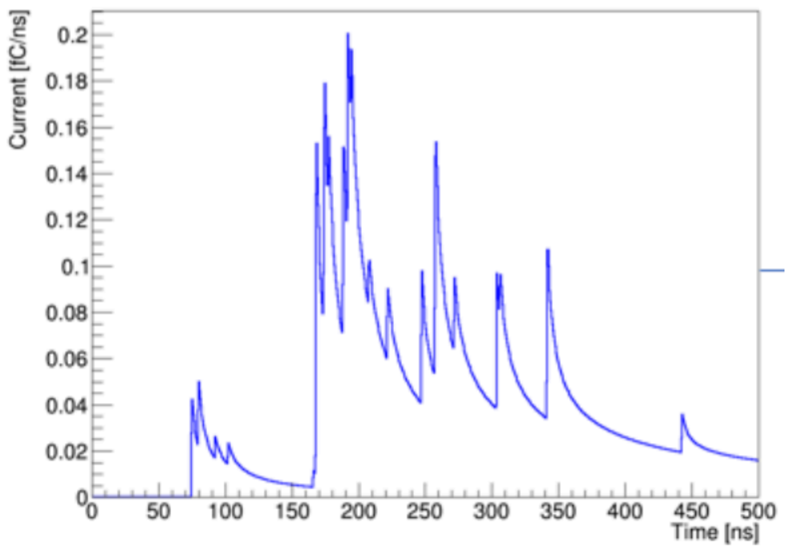


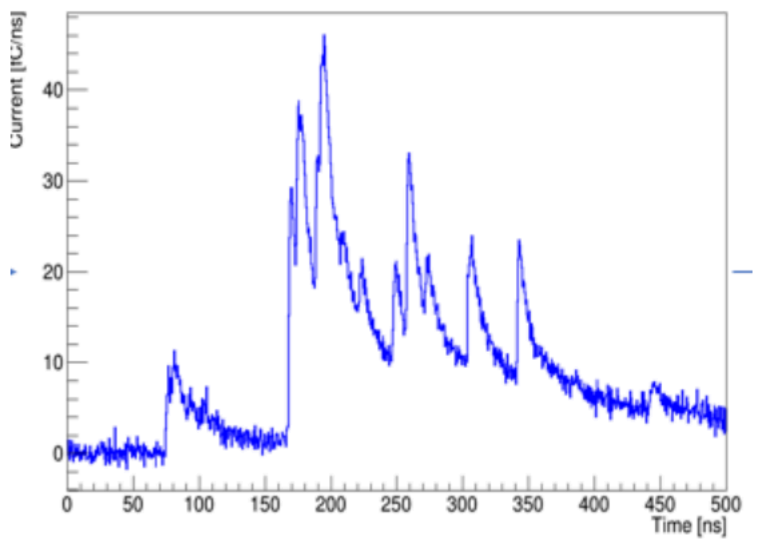


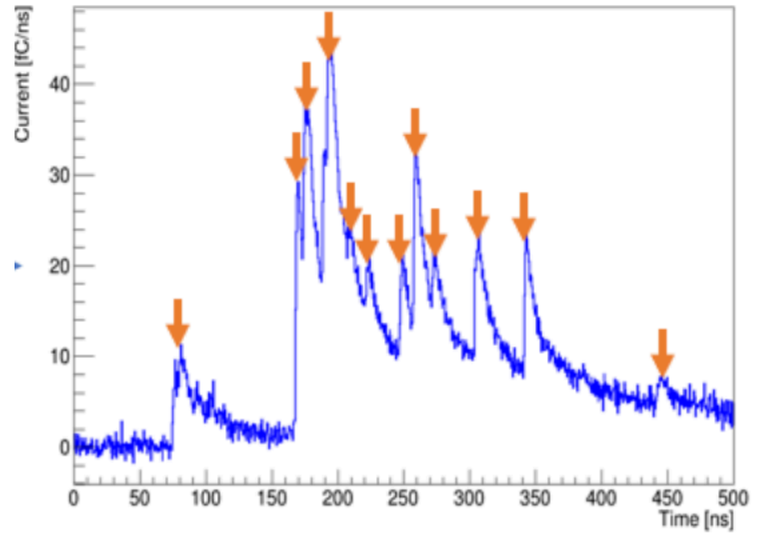
- Cluster counting delivers better PID performance than dE/dx.
- However, we need to consider the effect of electronics, noises etc. more importantly , the efficiency of counting algorithm.
- To simulate/control those factors, a waveform simulation is studied.

Waveform-based simulation

- The critical content of cluster counting is the waveform processing, an end-to-end workflow composes simulation, digitization and reconstruction steps.







Simulation

Signal generator (Garfield++)

- Heed: ionization process
- Magboltz: gas properties (drift/diffusion)

Digitization

Electronics

- Preamplifier
- Noise
- ADC: sampling rate

Reconstruction

Peak finding algorithm

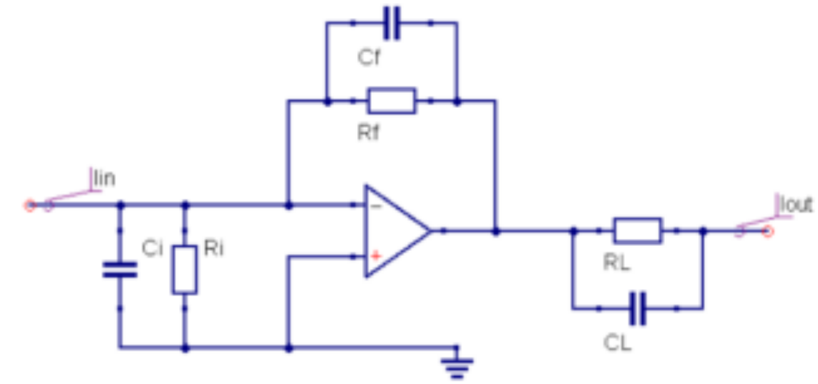
- Second derivative
- Machine Learning

Digitization

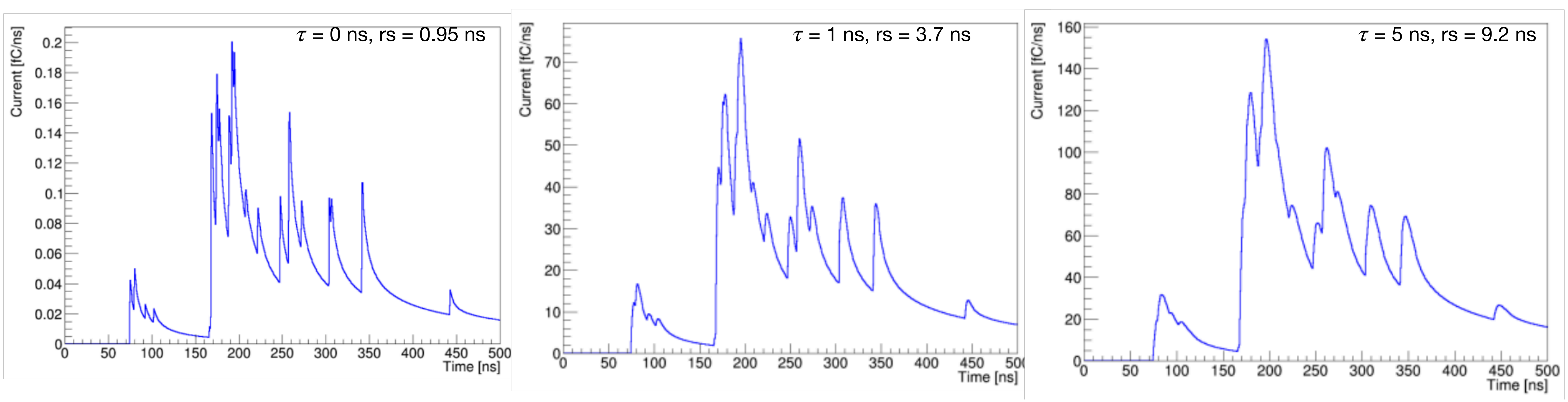
- A simple current-sensitive pre-amplifier is assumed.
  - A transform function is convoluted to raw signal:
  - Impulse response in time domain:

$$h(t) = 1/\sqrt{R_f C_f} \times e^{-t/R_f c_f}, R_f C_f = \tau \text{ (time const)}$$

  - Preliminary assumption, need further studies to align with experimental electronics.
  - Different τ will smooth waveform

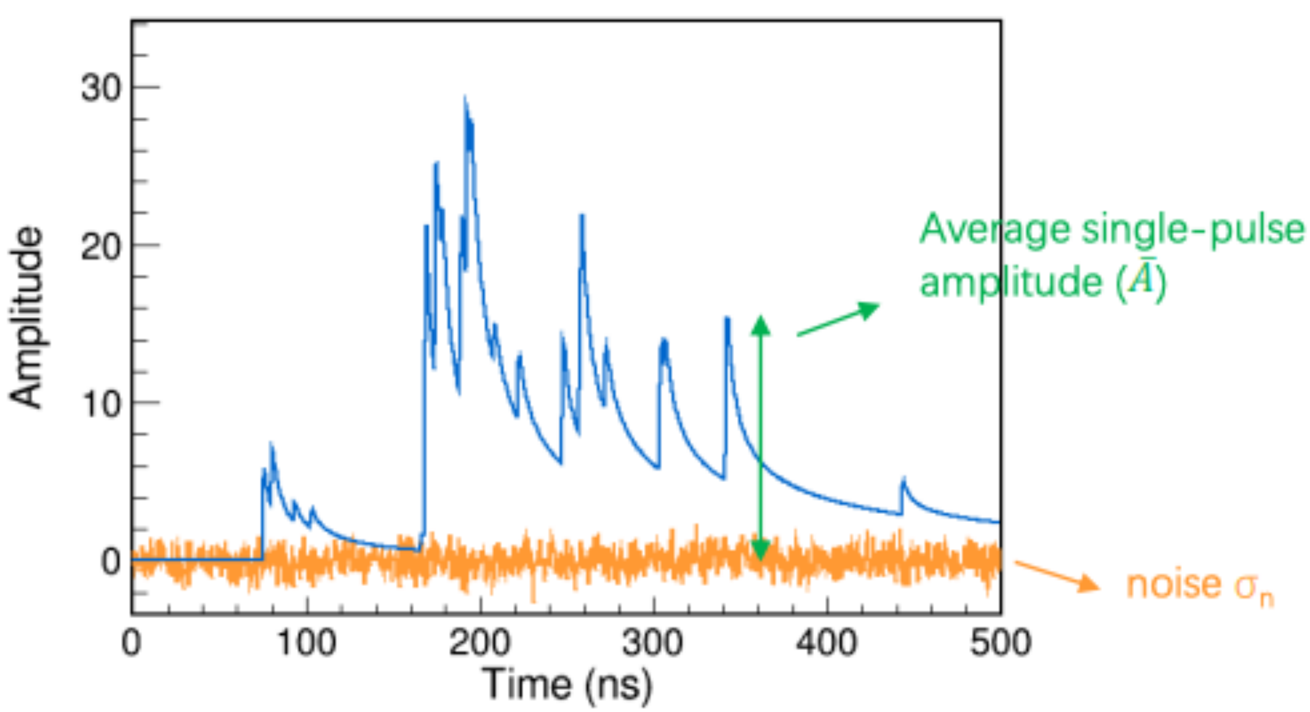


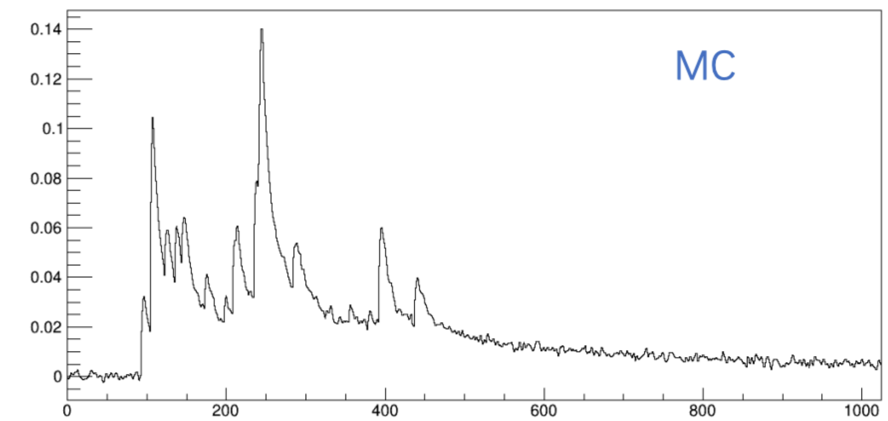
$Ri \parallel Ci \gg Rf \parallel Cf$ and $CL = 0$

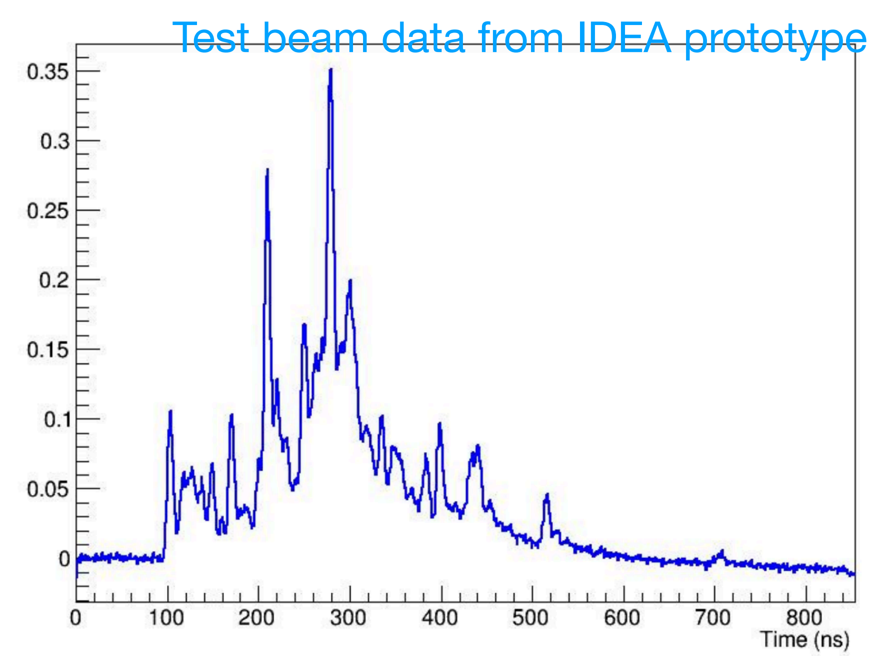


Digitization

- Noise simulation
  - Extract frequency spectra from experimental waveform using FFT.
  - Generate noise by performing iFFT with a random phase
  - Noise ratio definition: $\sigma_{Noise}/A_{signal}$

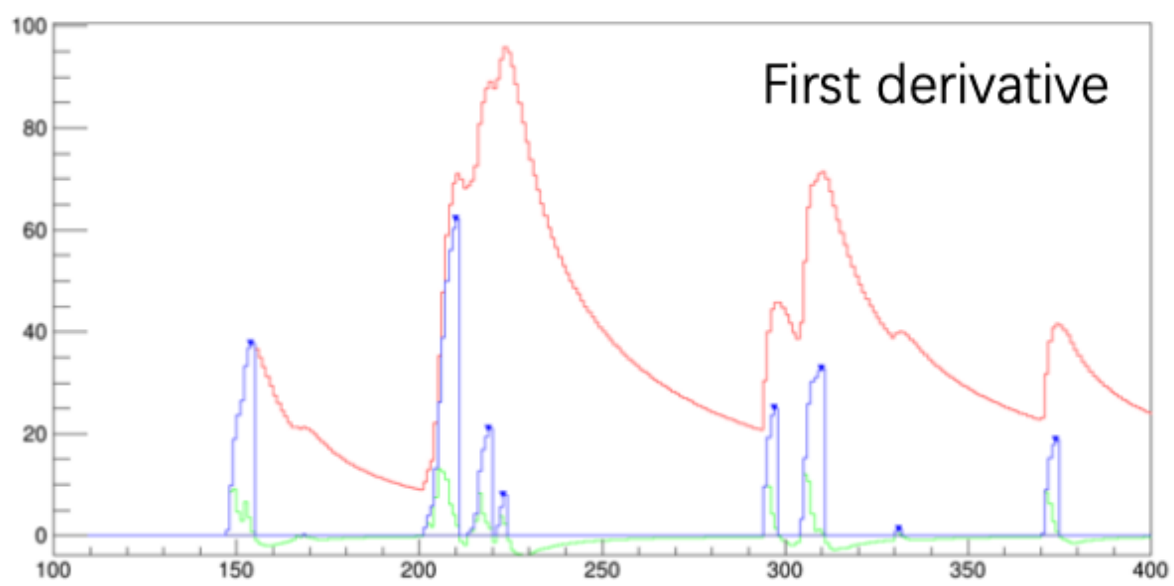


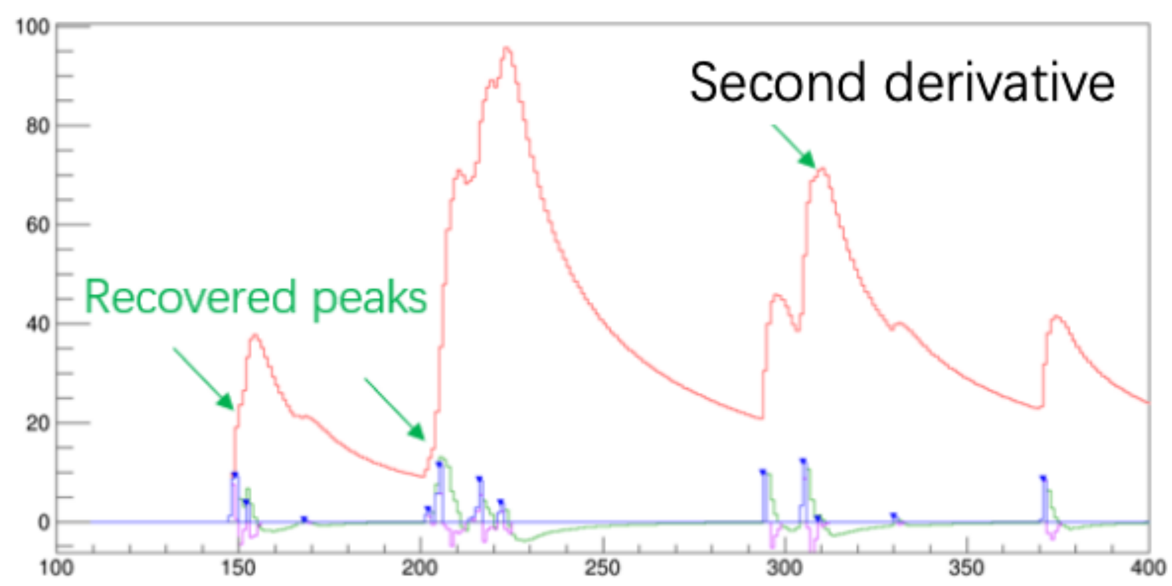




Reconstruction

- Looking for fast and efficient peak find algorithms
- Derivative: The start of a peak will be determined by a change in the derivative
  - Not sensitive to baseline
  - Can be easy to implement in hardware



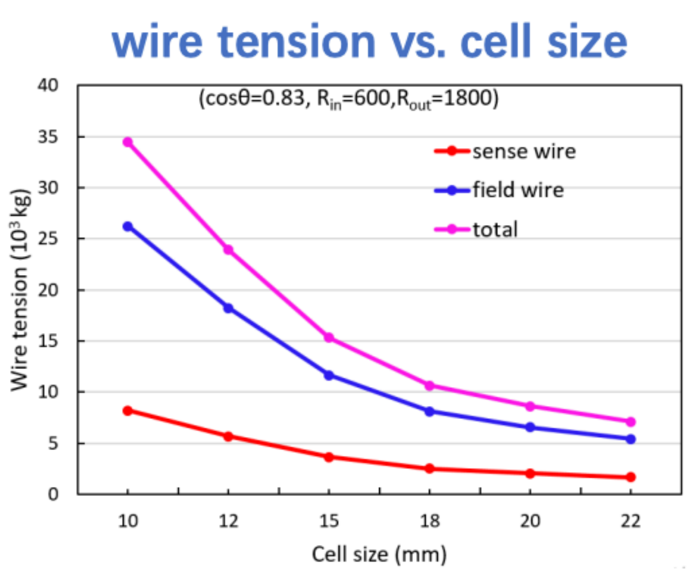


- Pile-up on the failing edge is easier to recover but not for the case on the rising edge
- Will show machine learning solution in a few coming slides

Detector preliminary design

- The separation depends on cluster density, track length, efficiency, etc.
- Gas mixture
  - Prefer smaller cluster density, slower drift velocity, smaller longitudinal diffusion
  - He/iC4H10 is selected from a branch of mixtures => Guang Zhao's talk.

- Cell size
  - In principle, the cell size would't affect too much PID.
  - Has impact on engineering. Prefer less wire tension, reduced number of wires => larger cell



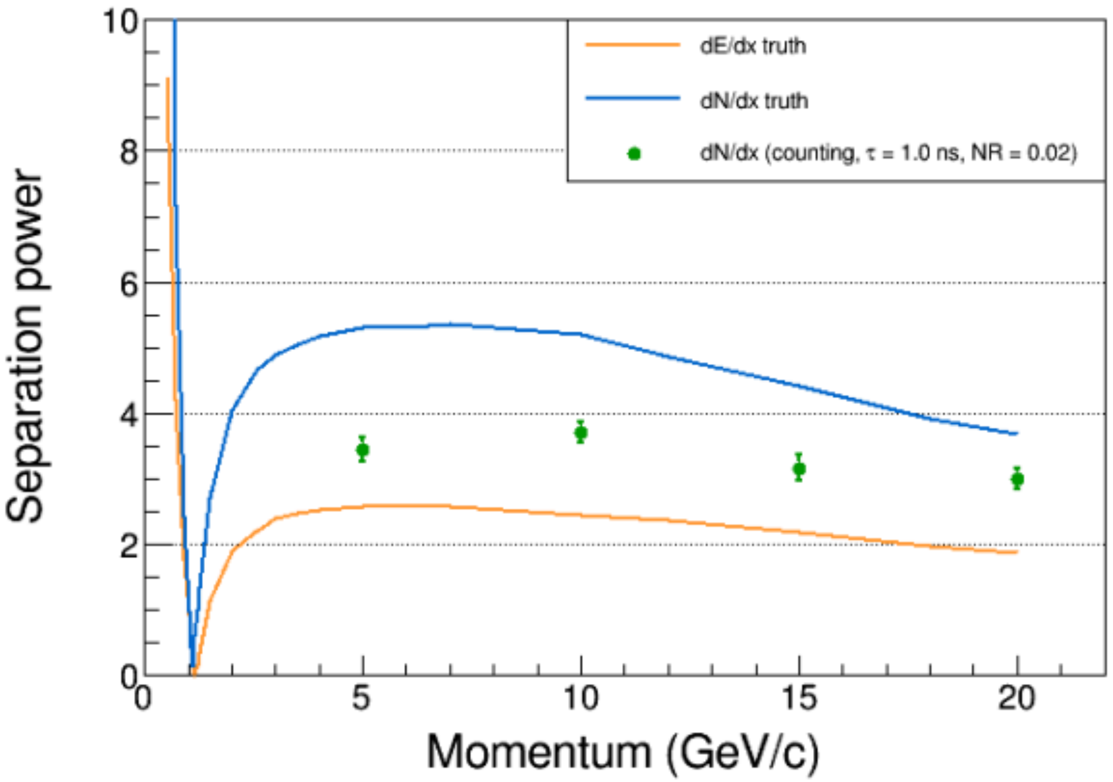
DC parameters

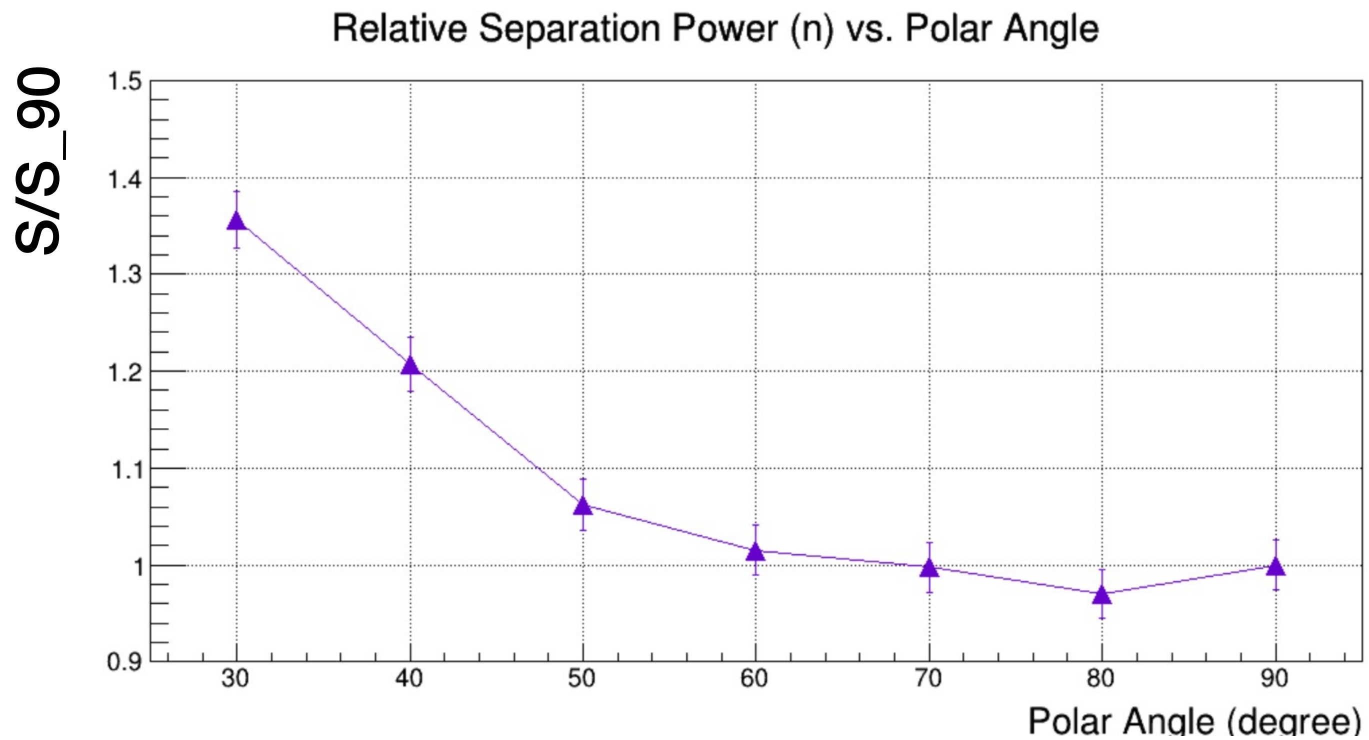
R extension		800 mm - 1800 mm
Length of outmost wires	$\cos\theta = 0.82$	5143 mm
Cell size		18 mm x 18 mm
Gas mixture		He/iC4H10 = 90:10

Expected performance with cluster counting

$$S = \frac{(dN/dx)_\pi - (dN/dx)_K}{(\sigma_\pi + \sigma_K)/2}$$

K/π separation vs momentum at 90 degree

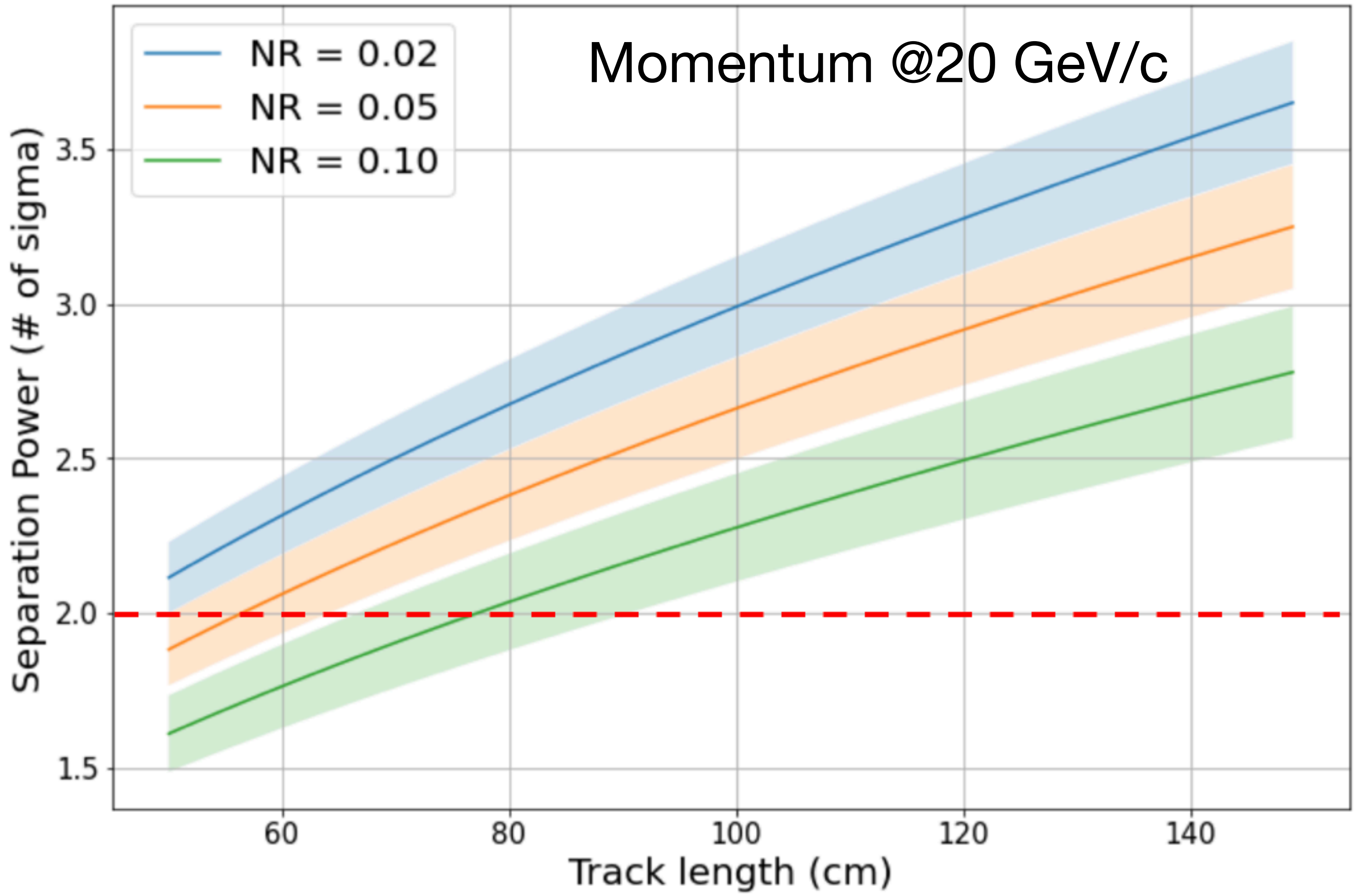




- 90 degree performance with 1m tracking length with noise: 1.5 better than dE/dx truth
- Polar angle scan: long track length allows better separation.

Expected performance with cluster counting

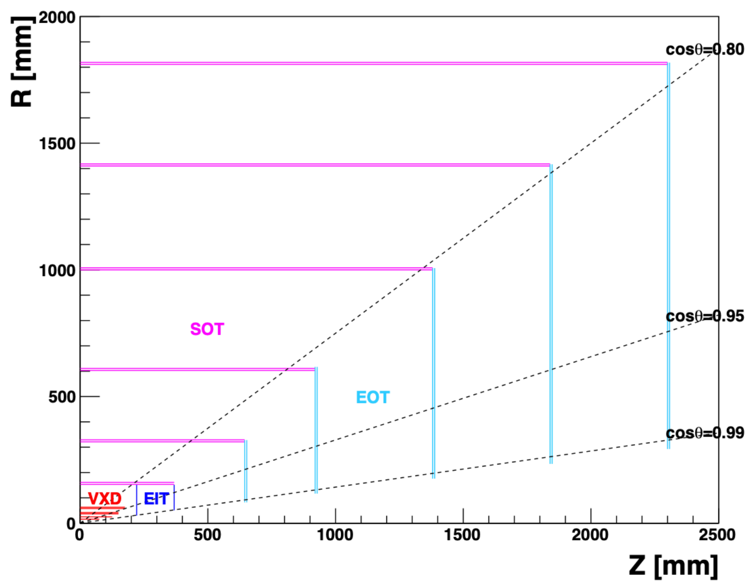
- A $2\sigma\ K/\pi$ separation at 20 GeV/c is achieved within 1m track length in 10% noise ratio.
- The requirement of separation power needs further studies with physics channels.
  - Some studies in physics channels considering dN/dx is ongoing.



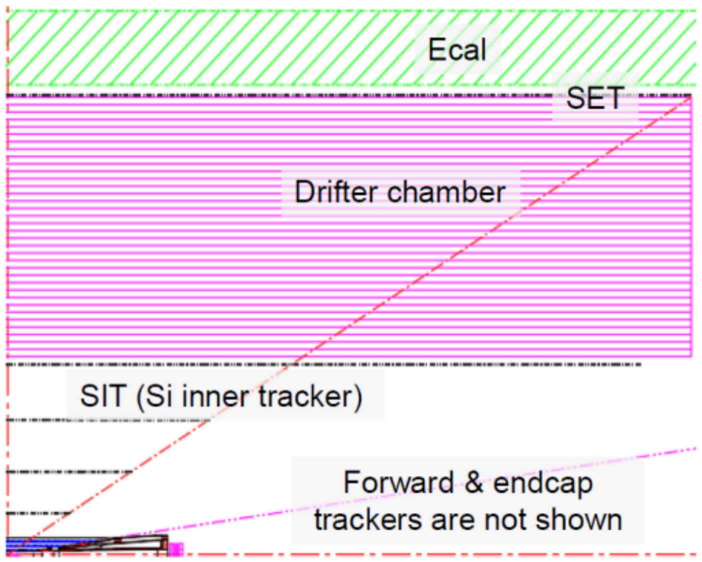
Resolution is scaled by $\sqrt{L}$

Momentum performance

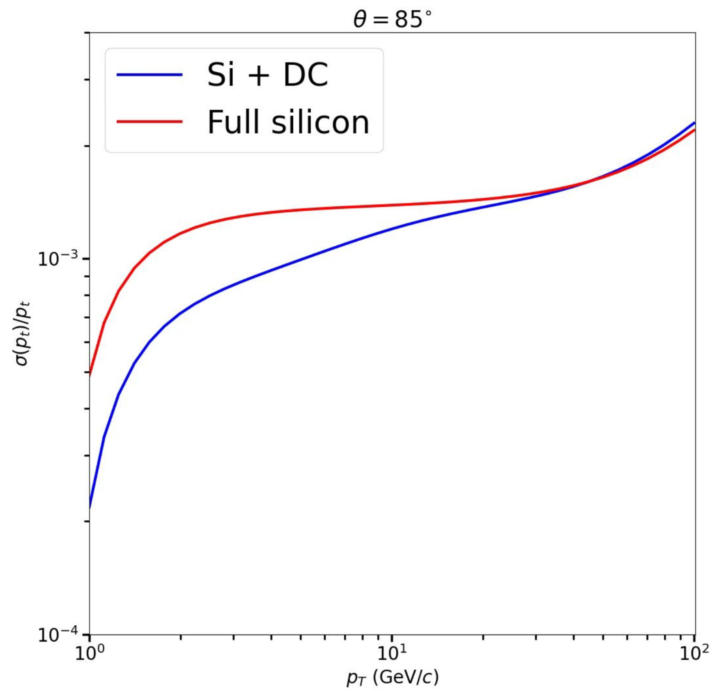
- High resolution and low budget Si system is required, such as silicon inner tracker, silicon external tracker for precise measurement.



Full Silicon Tracker (FST) from CDR



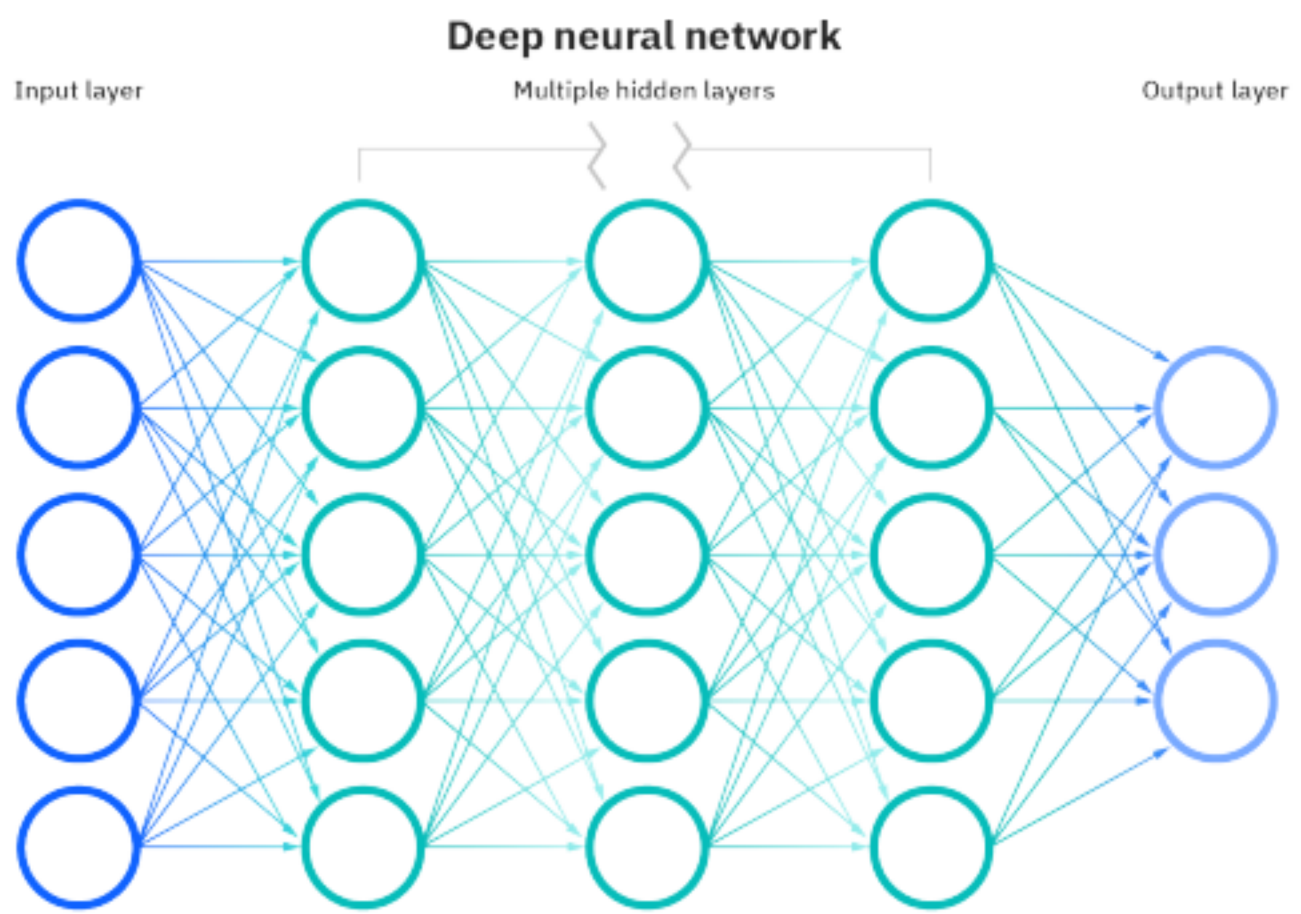
Si + DC



- The change of drift chamber is significant to low momentum measurement.
- Si + DC design gives better momentum resolution than FST design ~ Pt < 50 GeV/c.

counting algorithm based on Machine Learning

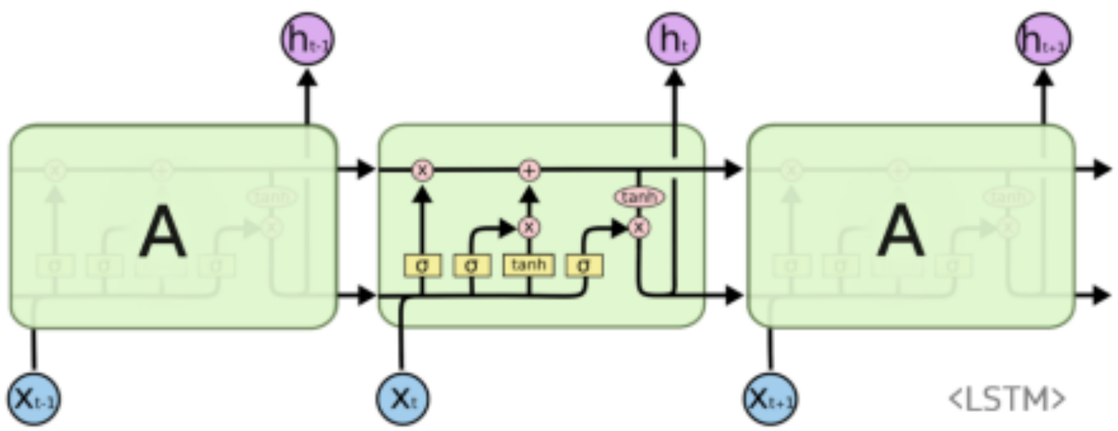
- The counting algorithm is the essential for the success of cluster counting
- Derivative peak finding algorithm recap:
  - Only use slope of the raising-edge information.
  - Not effective in high pile-up region and sensitive to noise.
- Machine learning: extract complex & nonlinear relationship of the input data
  - What architecture of ML can be a solution for cluster counting?

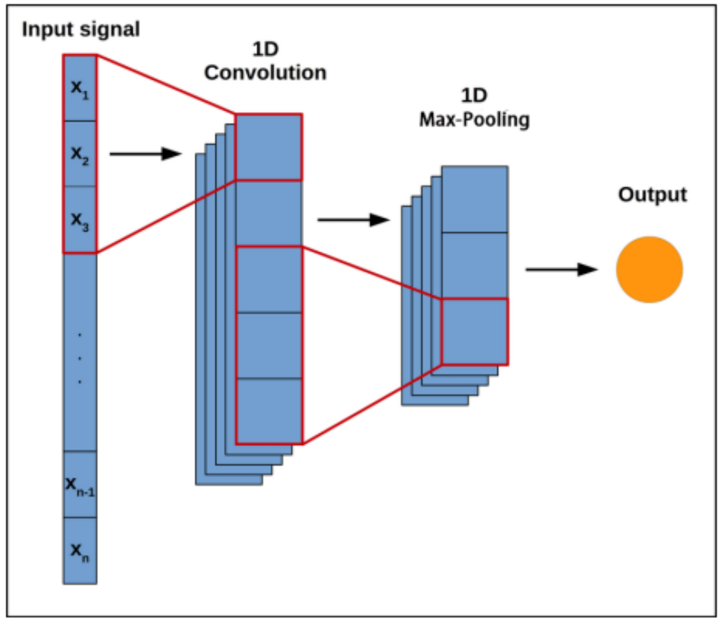


[Zhefei Tian's poster]

Machine Learning counting algorithm

- In practice, we treat the cluster counting as a classification and a regression problem in two steps:
  - **Peak finding**: find primary and secondary ionization peaks with the sequence information
  - N_{cluster} **determination**: group electron peaks to clusters

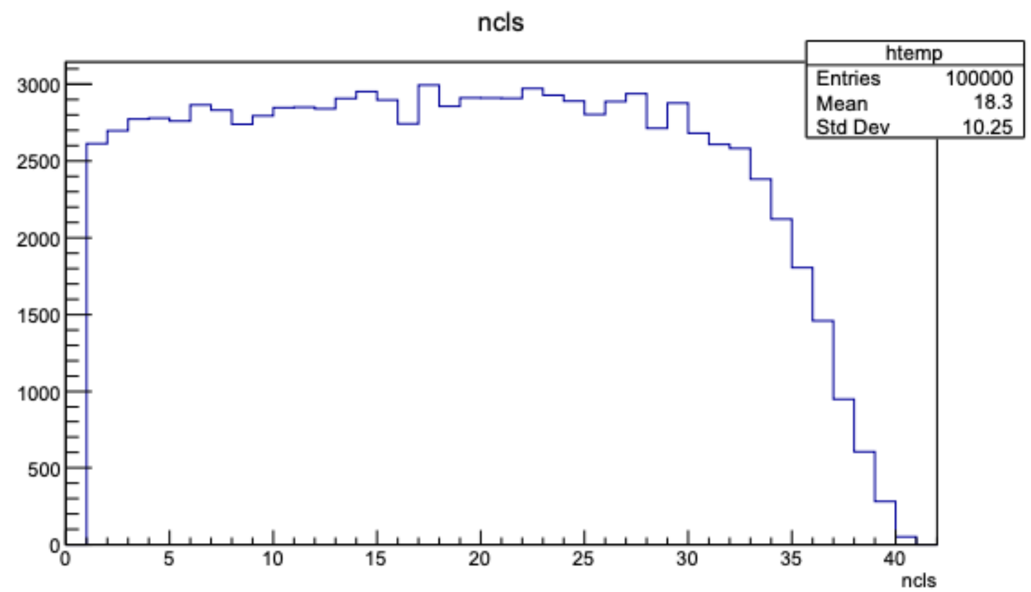


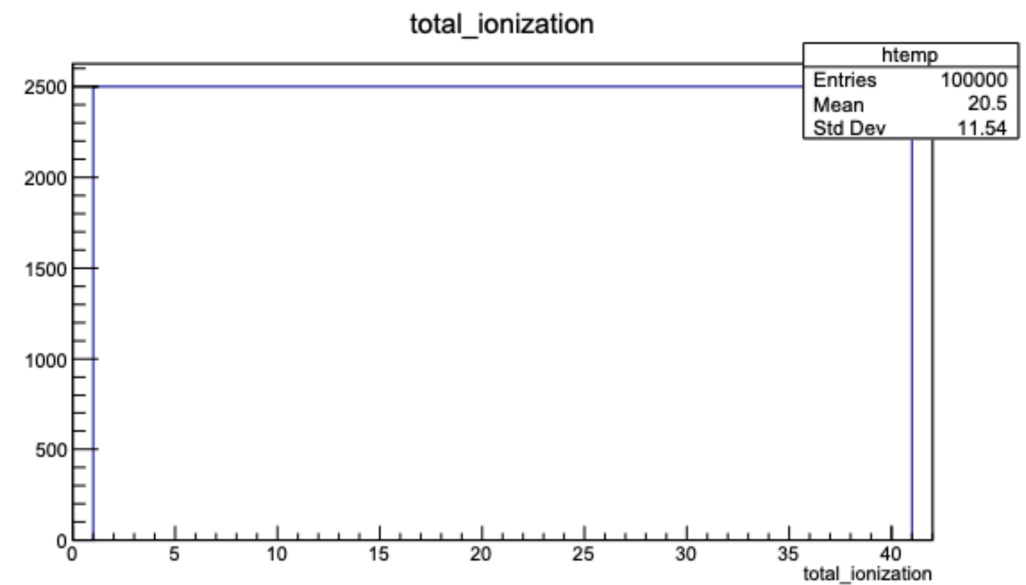


- "LSTM" = long short-term memory
  - Can retain contextual information from previous samples
  - Good for time series datasets for which the order of the events matters
- CNN:
  - Detect distinct features without any human intervention.
  - Provide an efficient dense network for identification

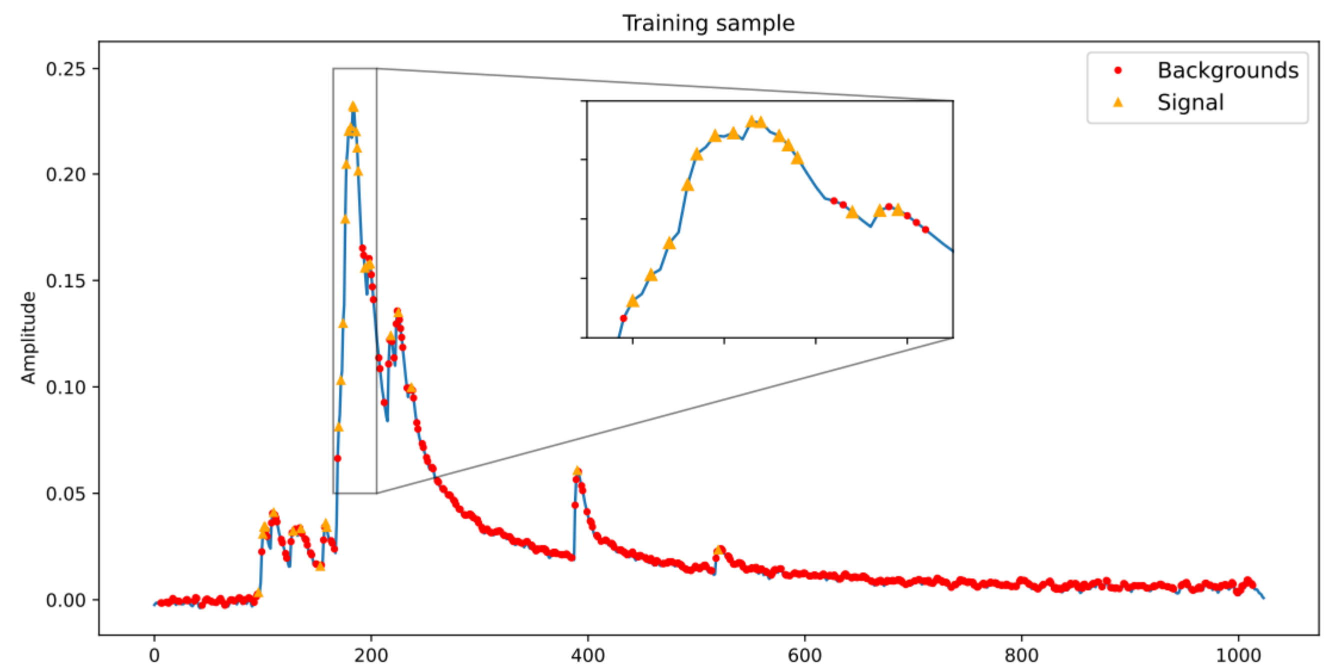
Data preparation

- A set of Monte-Carlo sample was generated as training sample
- The number of total ionization is 1 to 40, where the number of clusters are sampled from certain distribution.





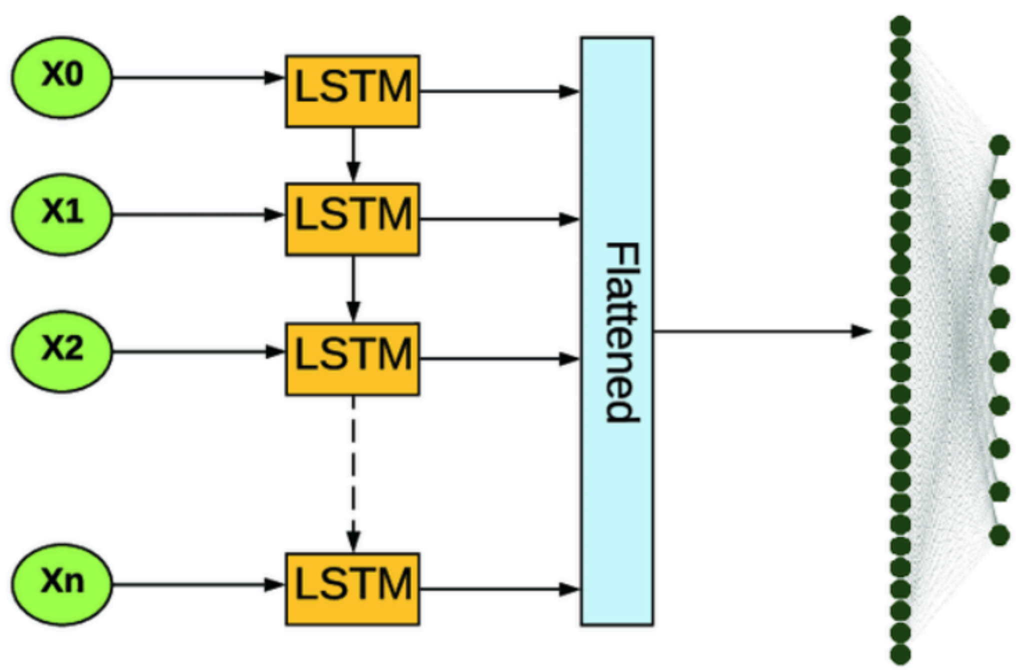
- Use slice windows of the waveform as the inputs
  - Label is added to match the signal/background
  - Balancing signal/background sample

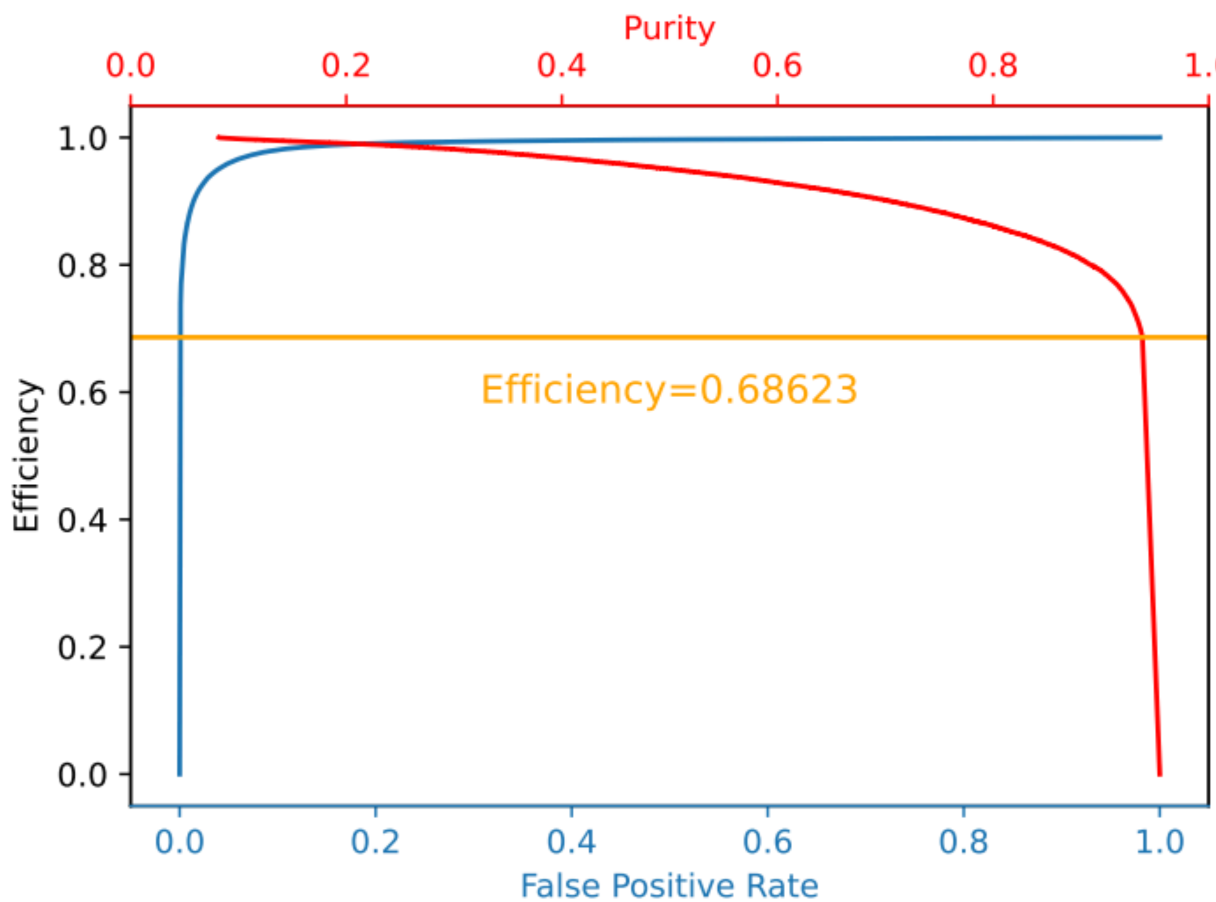


Training model- LSTM

- **Step1: peak finding: find all ionization peaks from current waveform.**
- Purity of detected signal peaks, efficiency of detected signal peaks and other metrics are used to evaluate the model:
  - Purity (Precision) = TP / (TP + FP) = 0.892
  - Efficiency (Recall) = TP / (TP + FN) = 0.686
  - False Positive Rate = FP/(FP+TN) = 0.0005

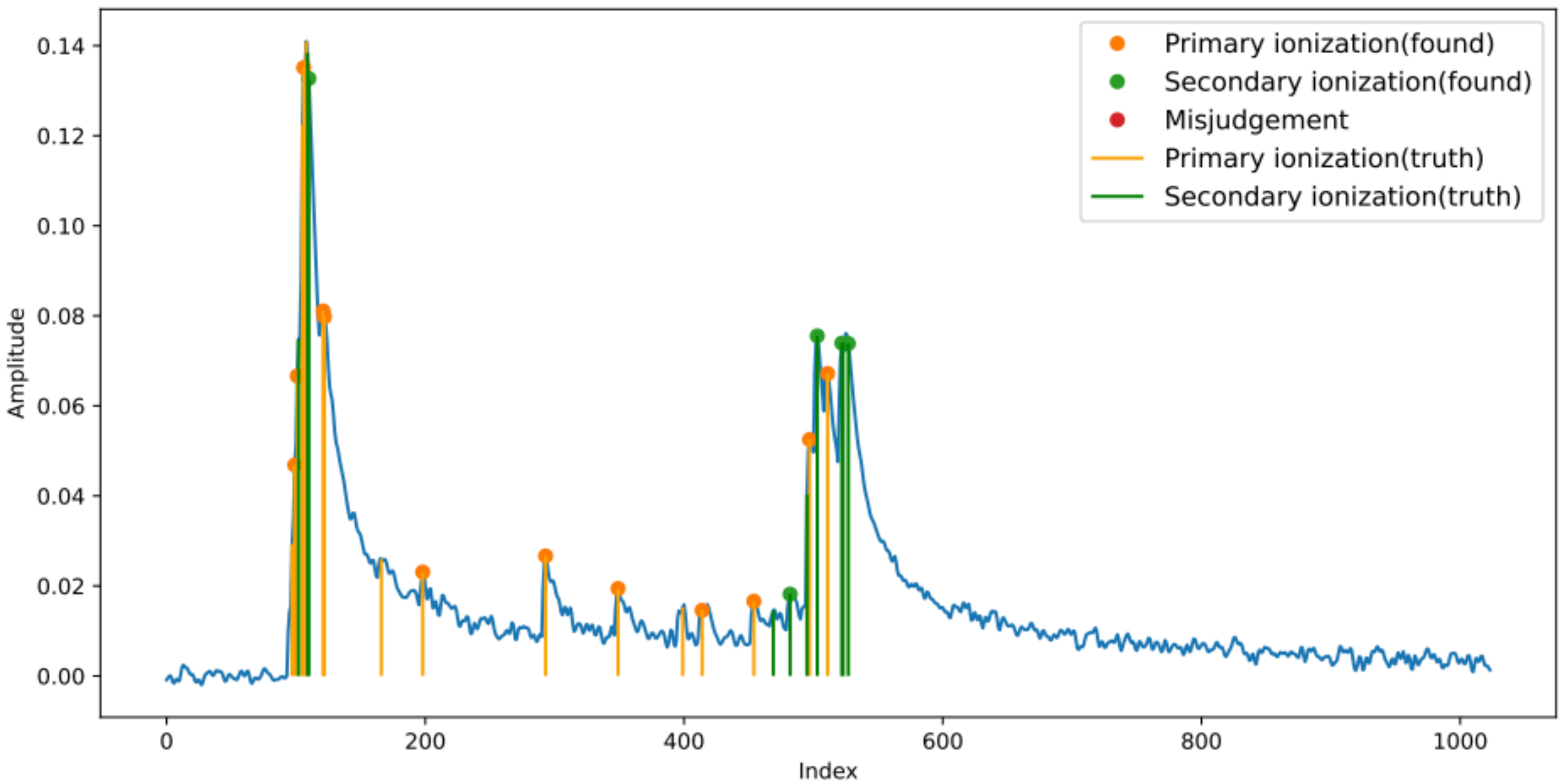
		Prediction	
		Signal	Noise
Truth	Signal	TP	FN
	Noise	FP	TN

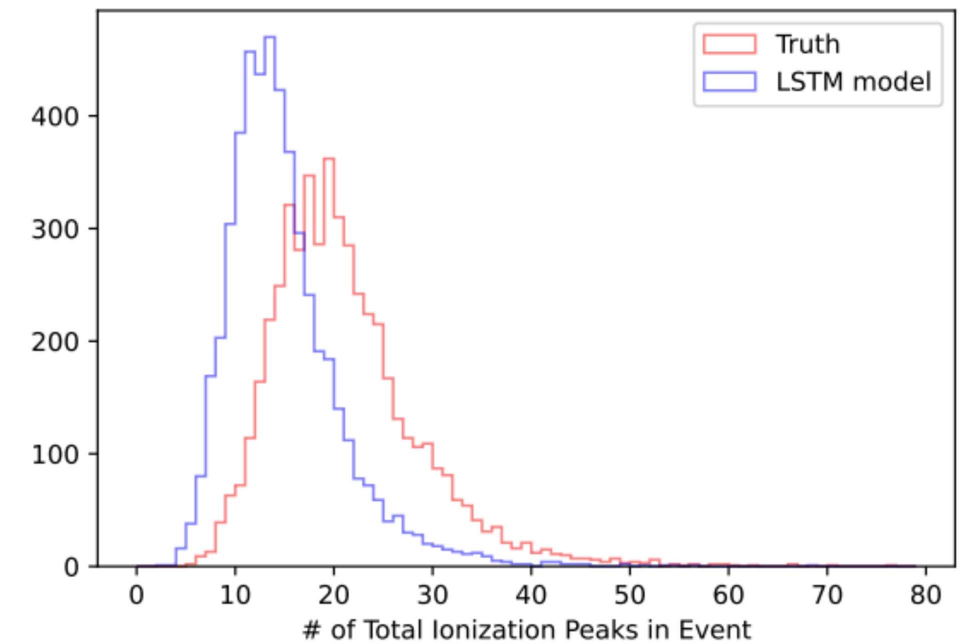




Peak finding model testing

- Test the training model on a separate datasets
- Peak finding result:
  - Moderate counting efficiency and tiny fake rate.

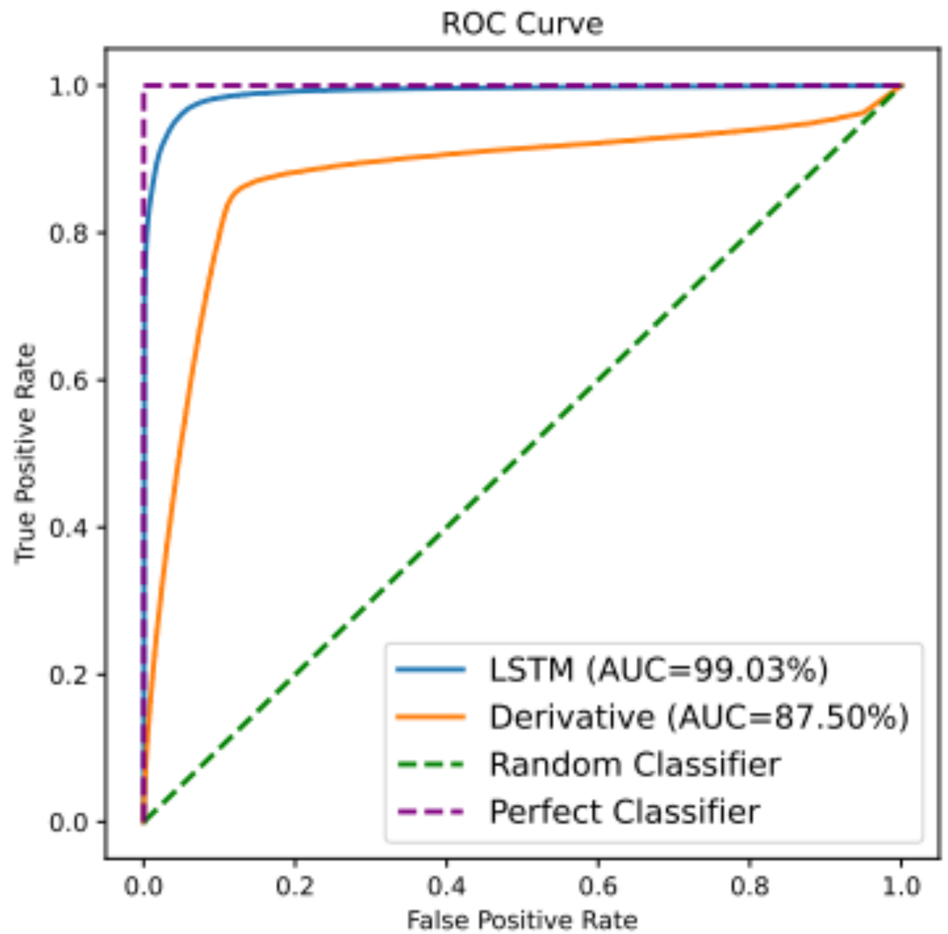


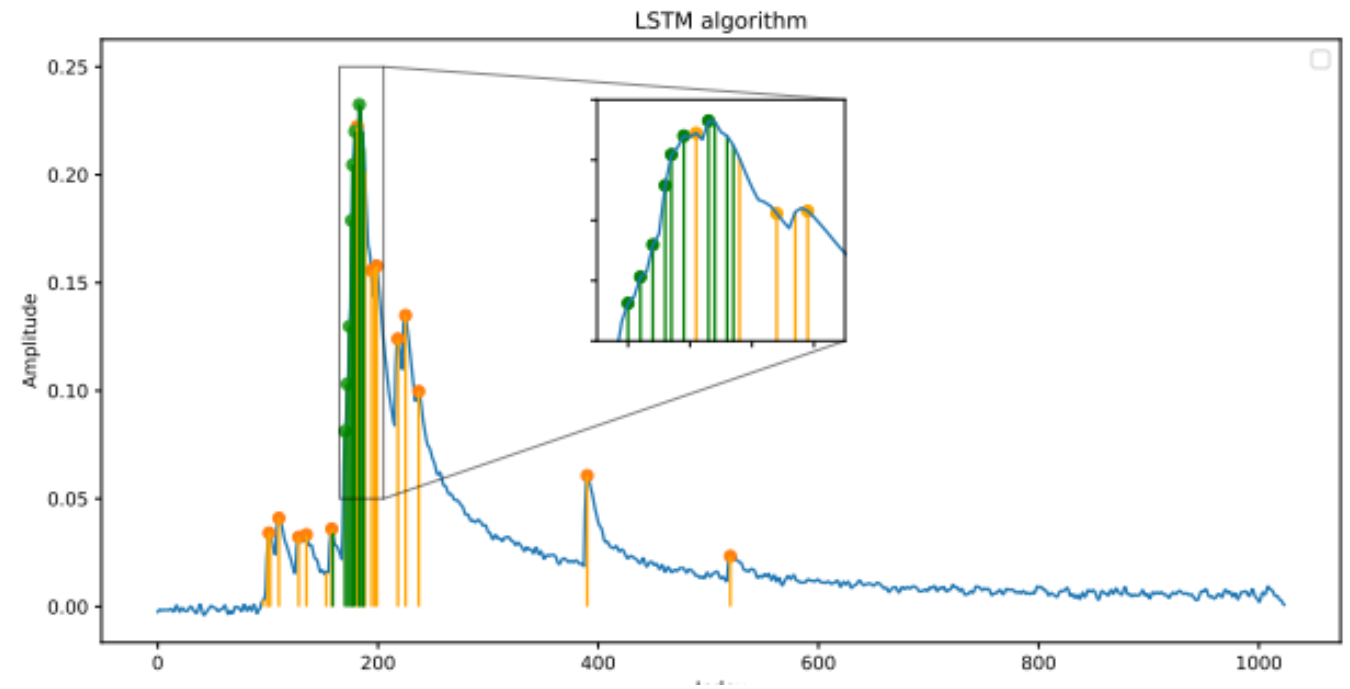


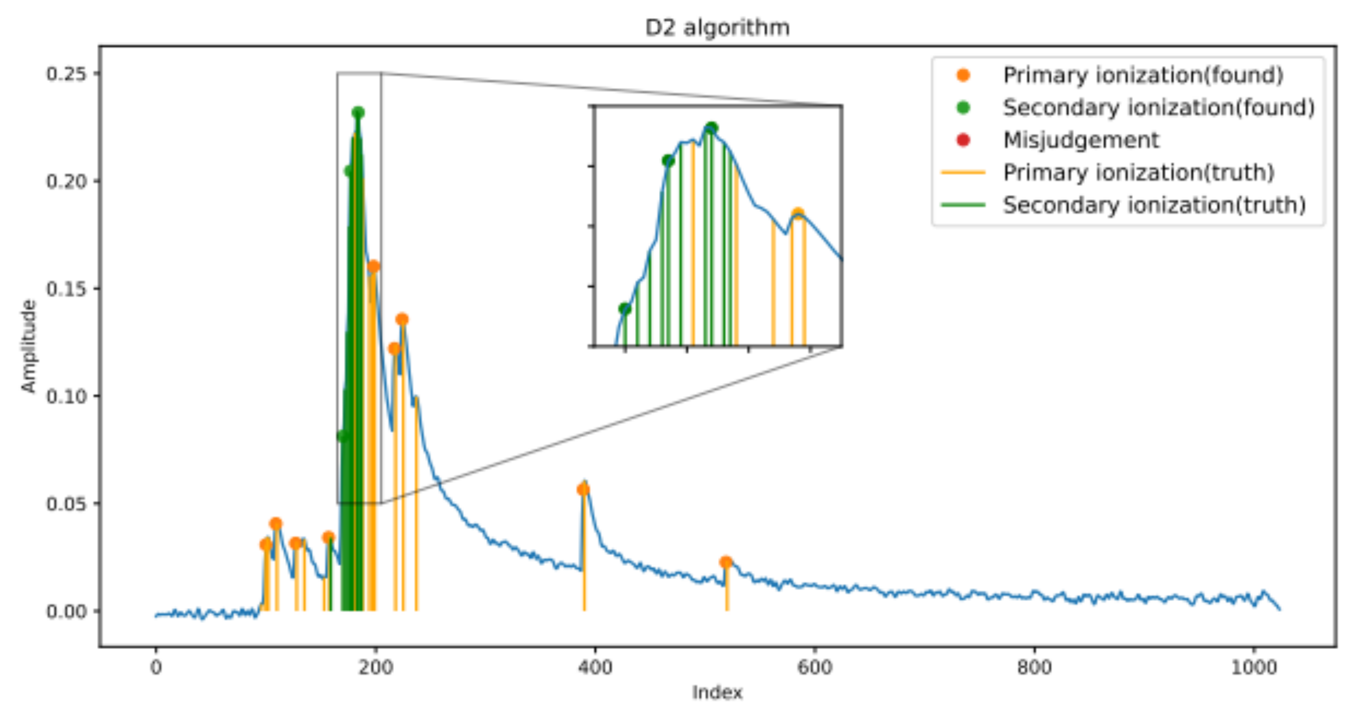
  - The lost peaks mostly come from pileup regions.

Comparison of two algorithms

- ROC curve is frequently used for evaluating the performance of binary classification algorithms. ROC curve with lager area under-curve (AUC) is better.

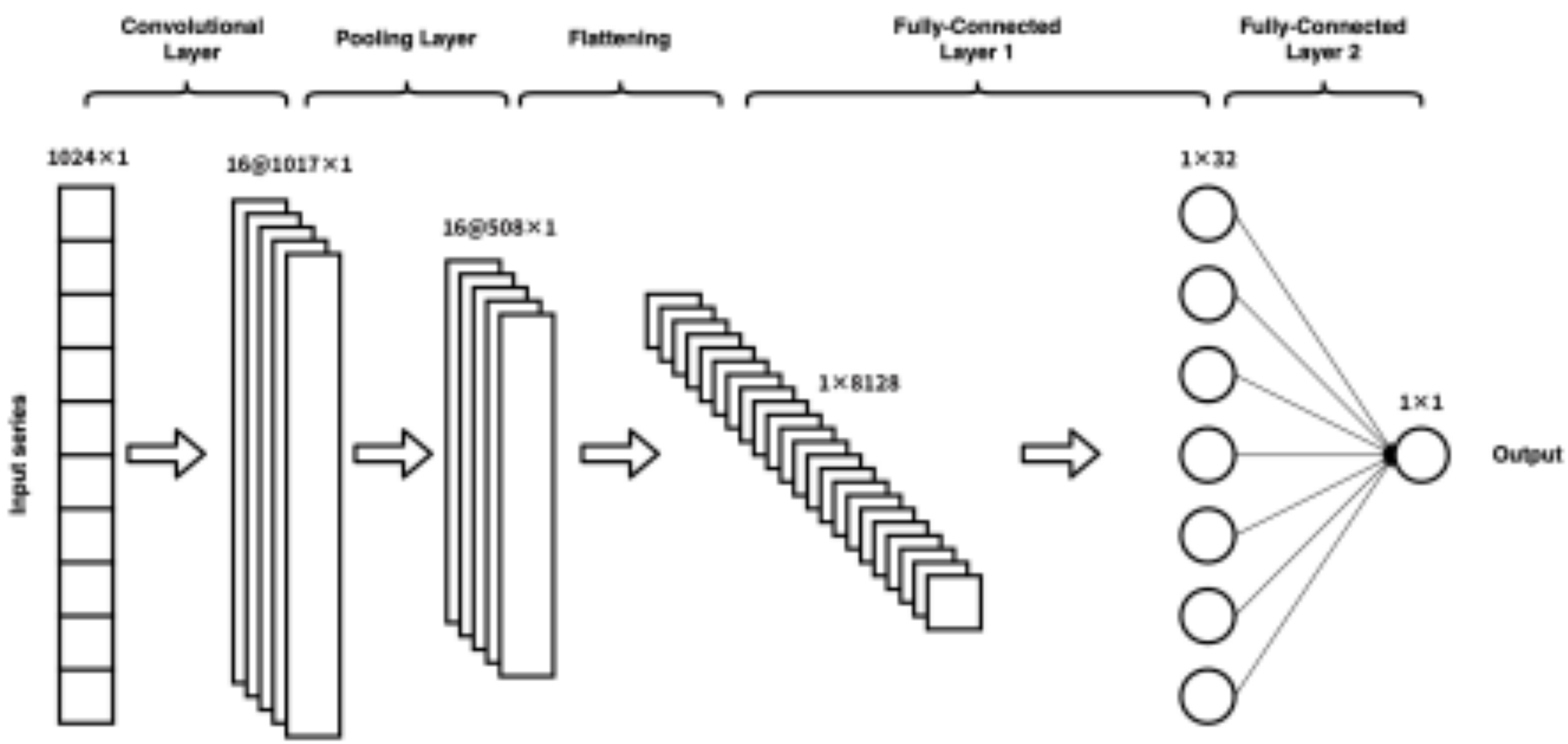






N_{cluster} determination training

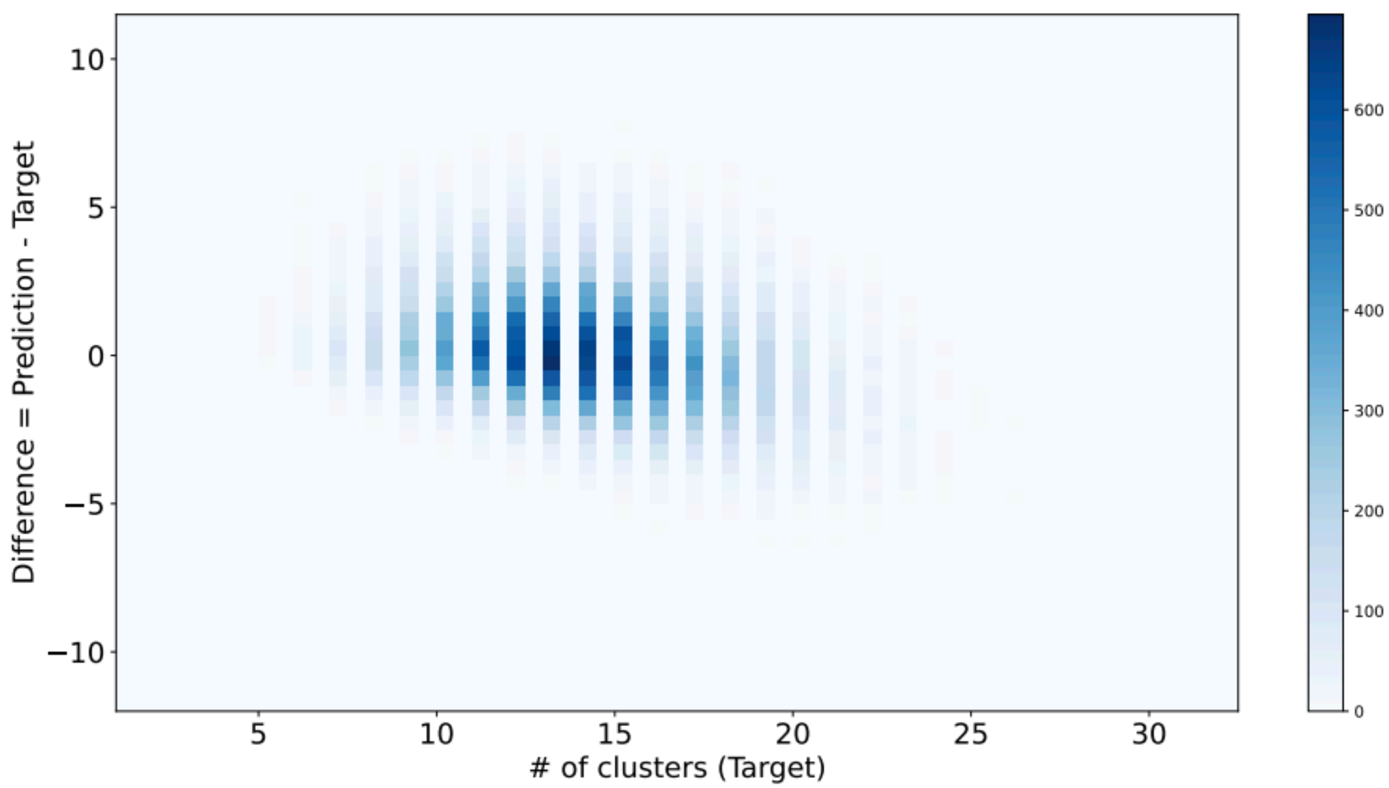
- **Step2: Determine the number of clusters according to the ionization peaks found by peak finding algorithm.**
- Regression problem to predict number of clusters
- The features are detected time (from first step), labels are number of clusters.

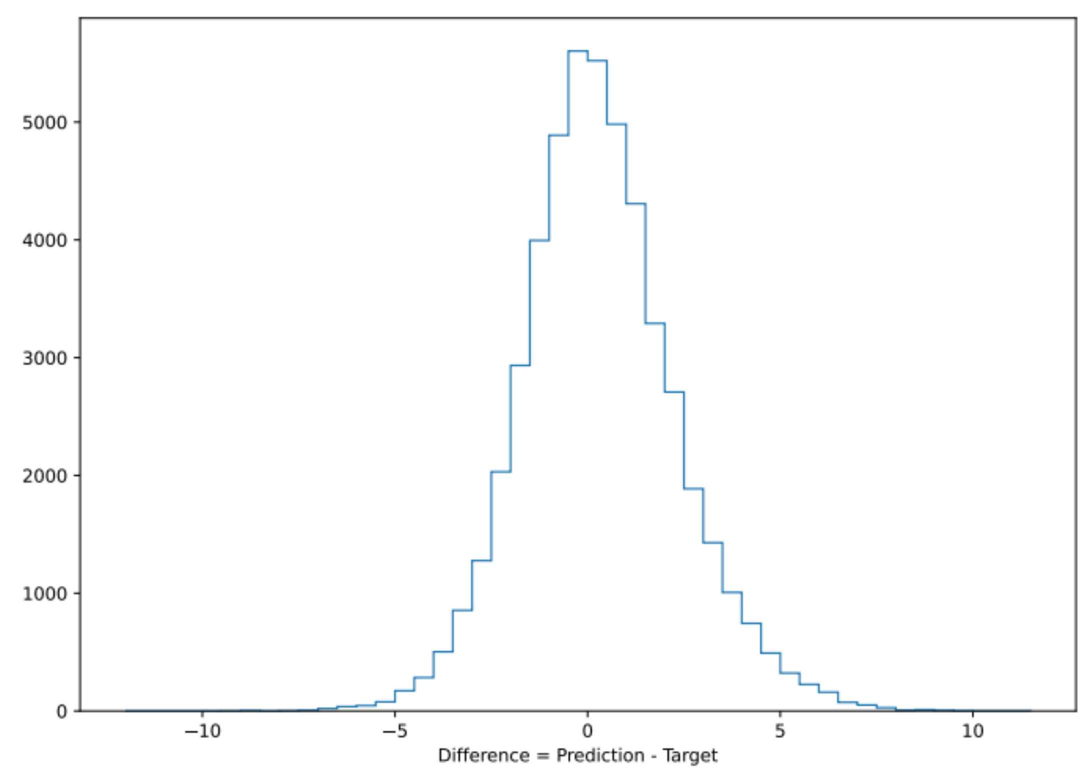


- 1D CNN to handle sequence data

N_{cluster} determination testing

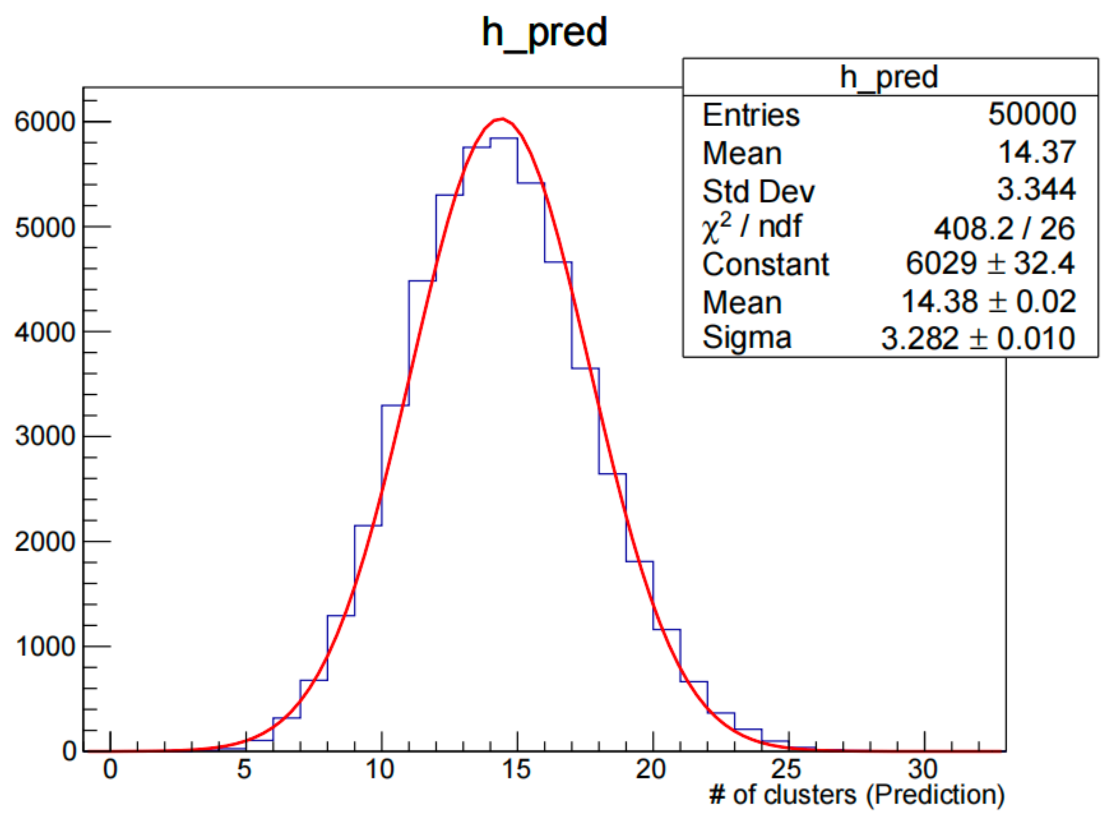
- Test samples with Poisson distribution (μ = 20) processed by peak finding model was used to test the Cluster determination model.
- The difference between predicted and target values are small and reminds stable.

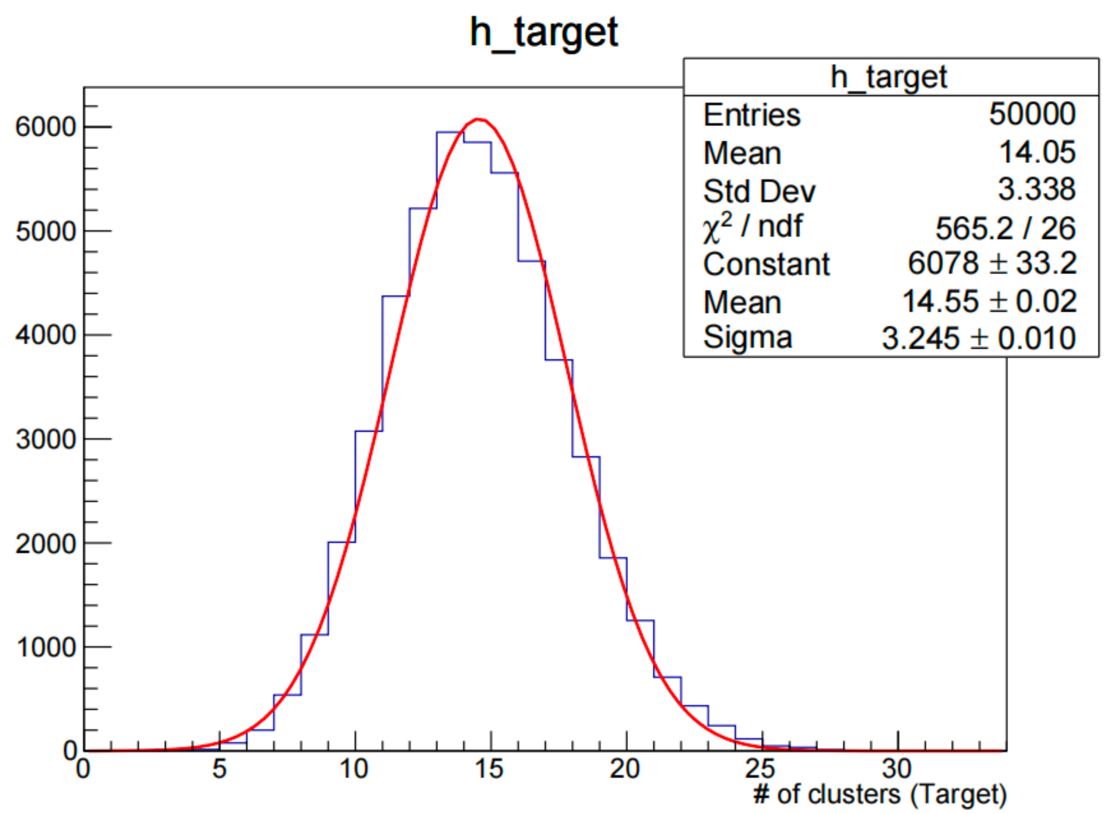




N_{cluster} determination testing

- Single cell resolution (σ/μ) ~ 22.8% (22.3 % in truth)
- Very good Gaussian distribution
- The relative error is quite similar to the truth value. Which implies stable efficiency.





Summary

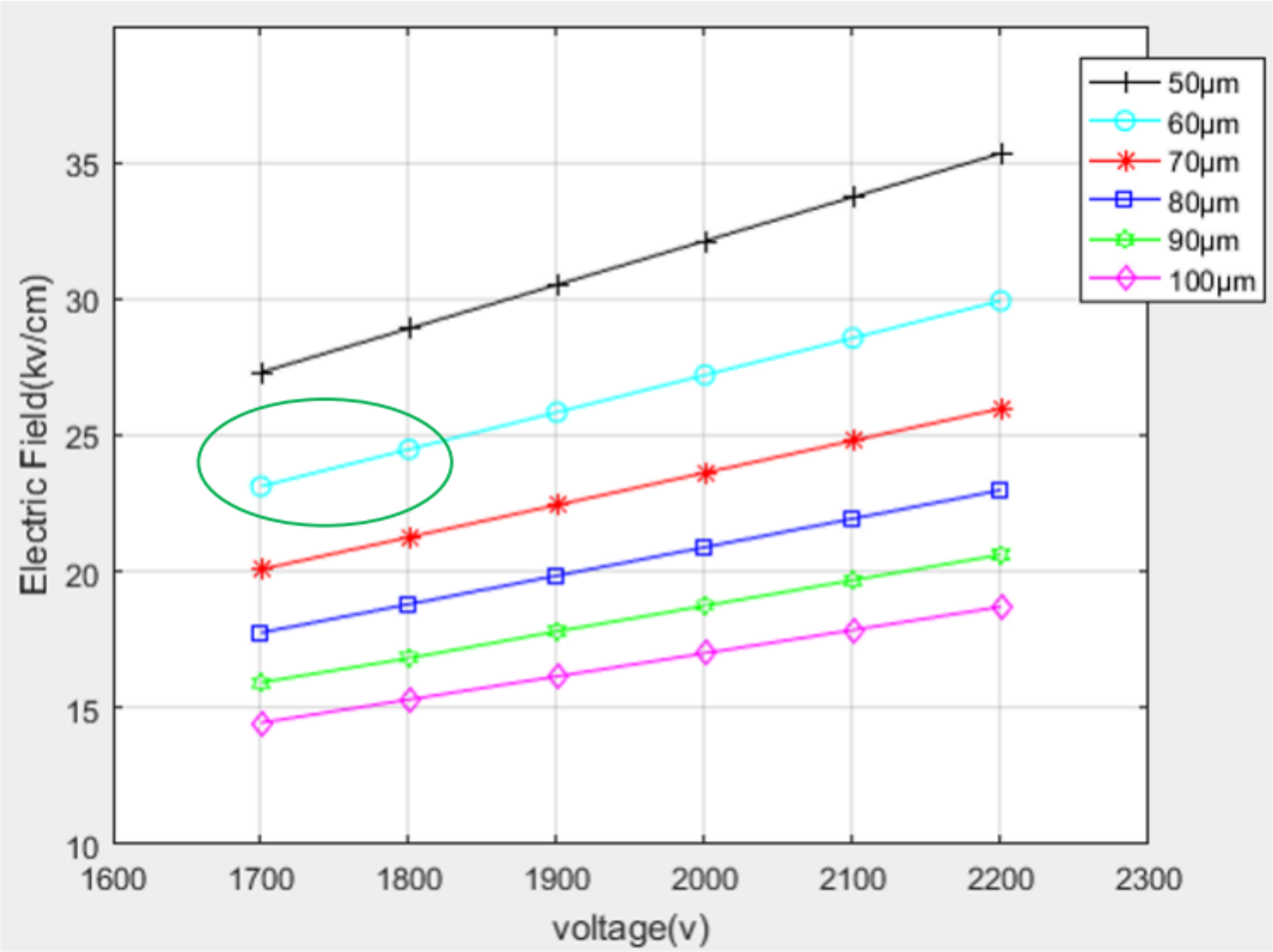
- A tracker with silicon layers and a drift chamber is proposed for CEPC, simulation study and expected PID performance are presented.
  - K/π separation reaches to 2σ up to 20 GeV/c for 1m track length under certain noises.
  - The momentum resolution of low momentum is benefited.
- A machine learning based cluster counting algorithm is developed.
- Result with the ML algorithms shows promising performance for MC samples.

- Outlook:
  - Evaluate the full performance of the ML algorithm and apply to the experimental data
  - Study the PID requirement in physics channels using Delphes fast simulation:

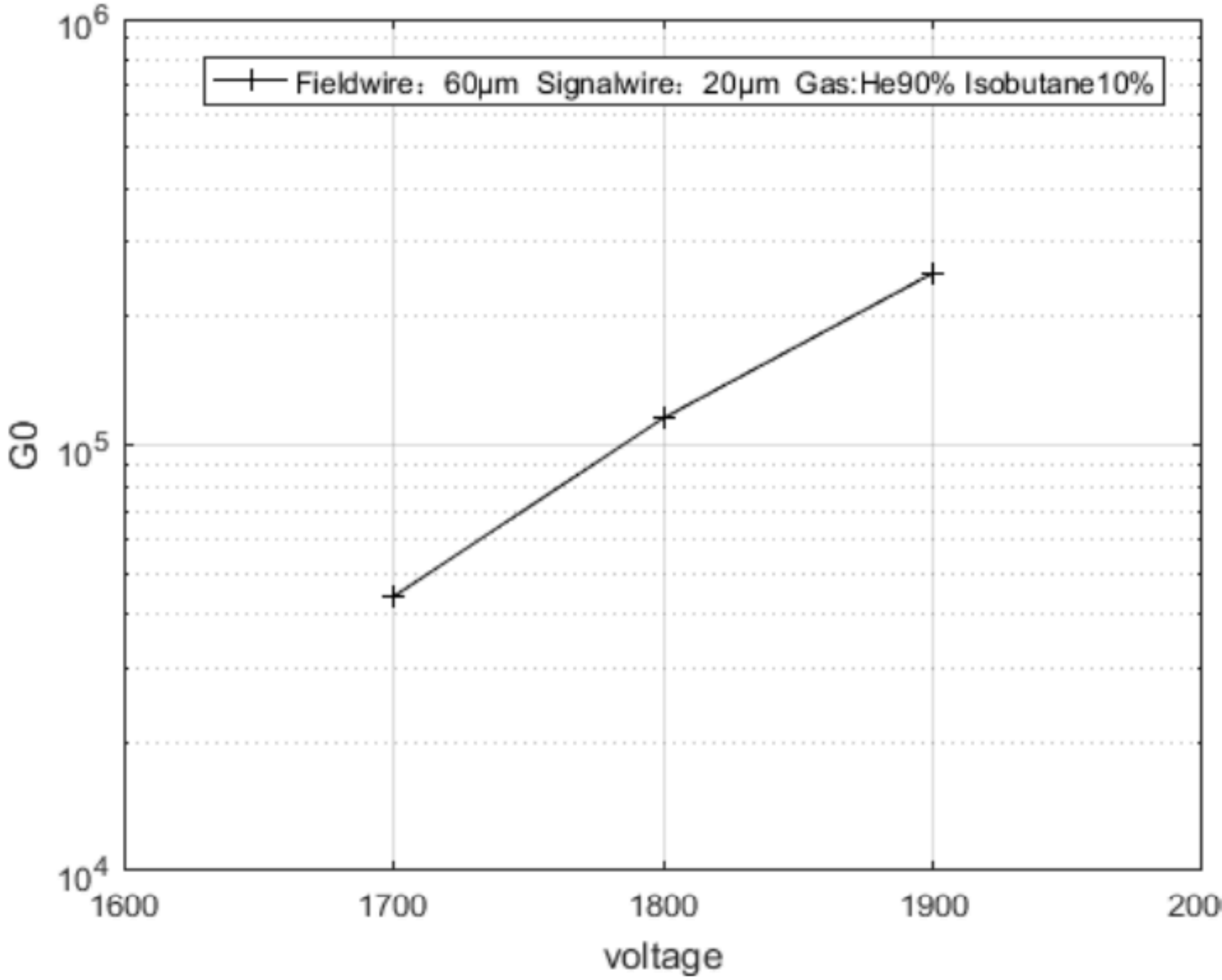
Many thanks to INFN, Jilin University, Wuhan University, Shandong University, Nanjing University and ZhengZhou University for their contributions !

Backup

Field and gain

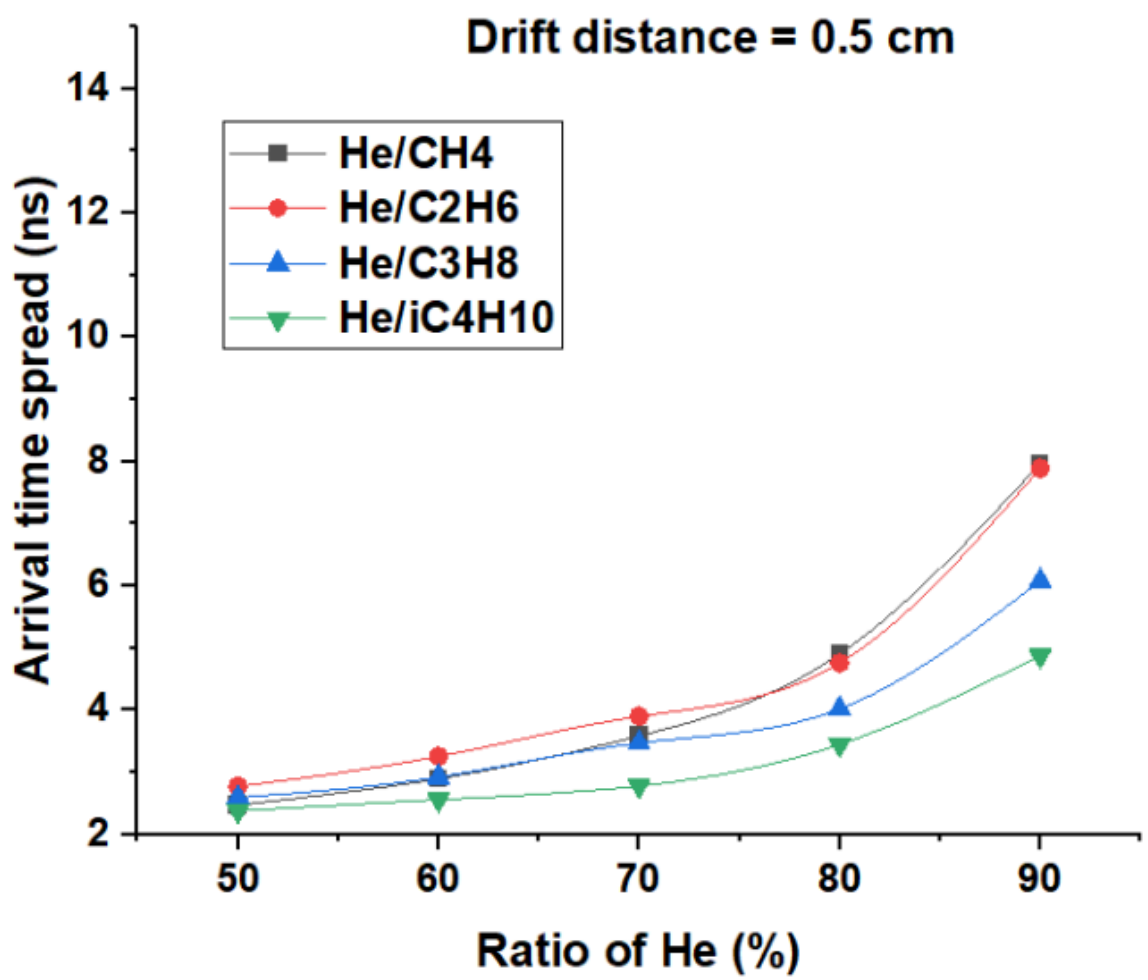


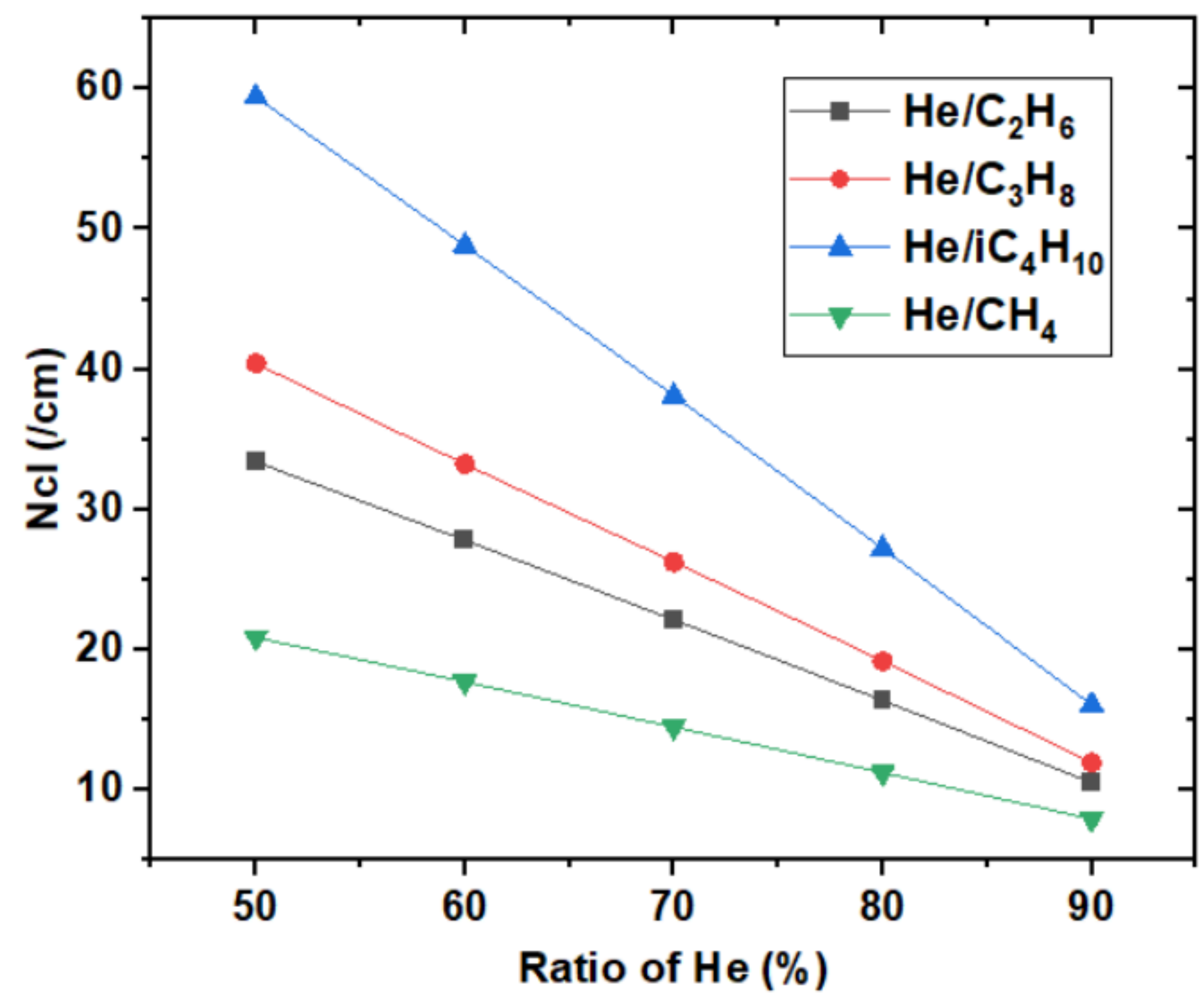
Electric field of field wires



Gain

Property for more gas mixtures



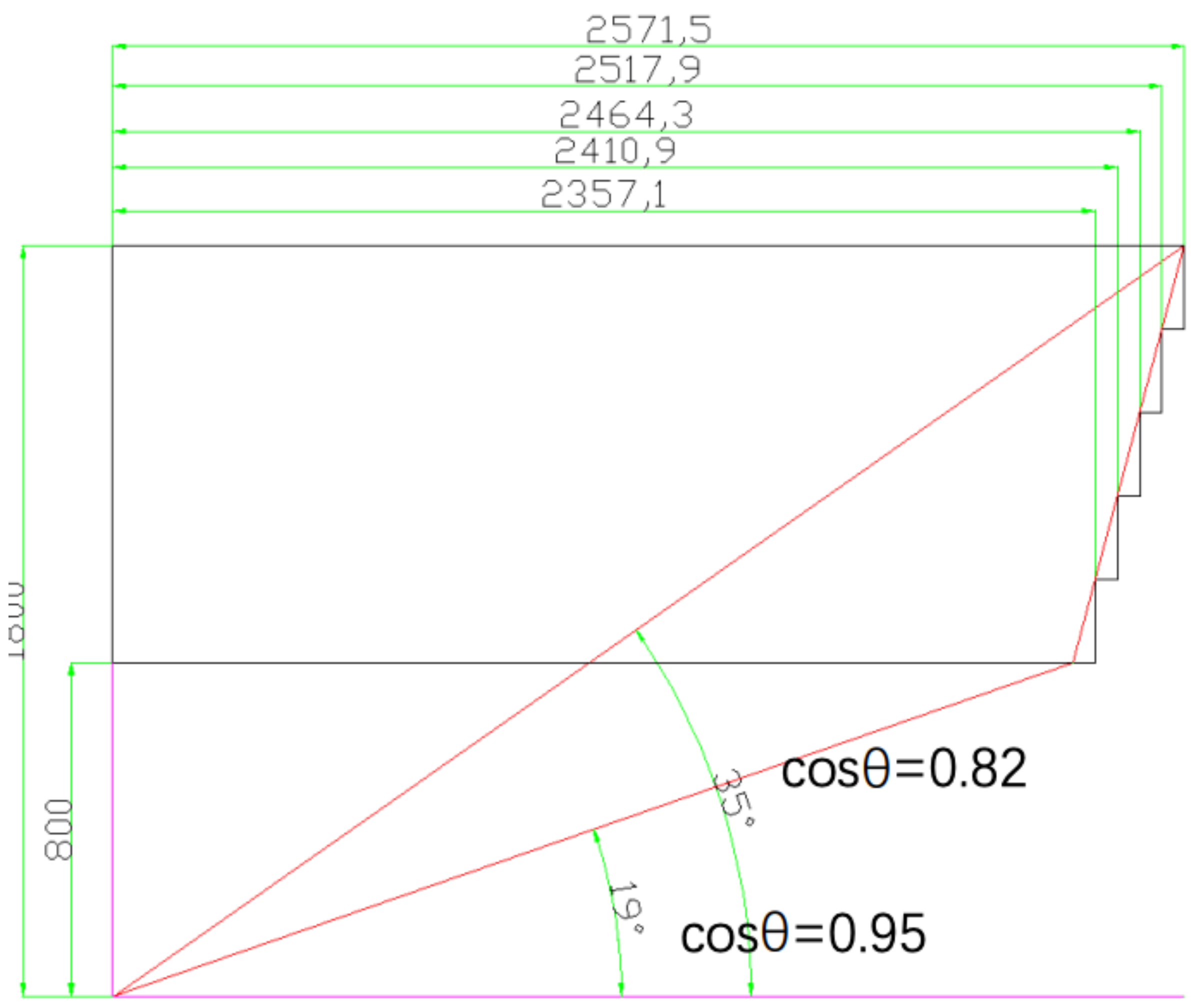


Tracker parameters

Gang Li's talk: https://indico.ihep.ac.cn/event/16011/session/2/contribution/12/material/slides/0.pdf

Layers	Radius(mm)	$\sigma_{R\phi}$(mu)	σ_Z(mu)	Thickness(1%/X_0)
Beam Pipe	14.5	-	-	0.15
VTX	16/18/37/39/58/60	2.8/6/4/4/4/4	2.8/6/4/4/4/4	0.15
VTX-shell	65.0	-	-	0.15
SITs	80/253/600	7.2/7.2/7.2	86.6/86.6/86.6	0.65
DC inner shell	798	-	-	0.104
DC wires (20*20mm) and gas	800 -- 1800	100	2828	0.0108+0.0031
DC outer shell	1803.0	-	-	1.346
SET (HV-CMOS 25x300 μm^2)	1811.0	11.5	138.5	0.65

Drift chamber design with mechanical structures



Total tension: ~9200 kg

DC Parameters	
R extension	800-1800mm
Length of outermost wires （cosθ=0.82）	5143mm
Thickness of inner CF cylinder	200μm
Outer CF frame structure	Equivalent CF thickness: 1.63mm
Thickness of end Al plate	35mm
Cell size	18 mm × 18 mm
Diameter of field wire (Al coated with Au)	60μm
Diameter of sense wire (W coated with Au)	20μm
Ratio of field wires to sense wires	3:1
Gas mixture	He/iC4H10=90:10