Drift Chamber with Cluster Counting for CEPC

Guang Zhao for DC PID Group

zhaog@ihep.ac.cn

Oct. 27th, 2023

2023 International Workshop on CEPC, Nanjing

中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

Istituto Nazionale di Fisica Nucleare

Outline

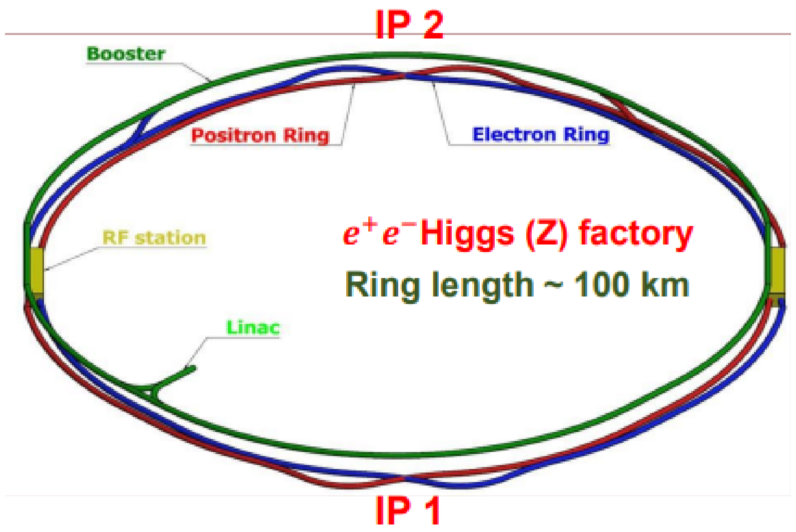
- **Introduction**
- **Reconstruction algorithm with deep learning**
  - Supervised model for simulated data
  - Semi-supervised domain adaption model for test beam data
- **Physics study with Delphes**
- **Prototype experiment**
- **Summary**

Physics programs at CEPC

- The CEPC aims to start operation in 2030's, as a Higgs (Z) factory in China. The plan is to operate
  - Above **ZH** threshold ($\sqrt{s}$ ~ 240 GeV) for 7 years.
  - Around and at the **Z** pole for 2 years.
  - Around and above **W^+W^-** threshold for 1 year.
  - It is upgradeable to run at the $t\bar{t}$ threshold.
- Possible *pp* collider (SppC) of $\sqrt{s}$ ~ 50–100 TeV in the future.

Particle	$E_{c.m.}$ (GeV)	Years	SR Power (MW)	Lumi. /IP ($10^{34}cm^{-2}s^{-1}$)	Integrated Lumi. /yr (ab^{-1}, 2 IPs)	Total Integrated L (ab^{-1}, 2 IPs)	Total no. of events
H*	240	10	50	8.3	2.2	21.6	4.3×10^6
			30	5	1.3	13	2.6×10^6
Z	91	2	50	192**	50	100	4.1×10^{12}
			30	115**	30	60	2.5×10^{12}
W	160	1	50	26.7	6.9	6.9	2.1×10^8
			30	16	4.2	4.2	1.3×10^8
$t\bar{t}$	360	5	50	0.8	0.2	1.0	0.6×10^6
			30	0.5	0.13	0.65	0.4×10^6

* Higgs is the top priority. The CEPC will commence its operation with a focus on Higgs.
** Detector solenoid field is 2 Tesla during Z operation, 3Tesla for all other energies.
*** Calculated using 3,600 hours per year for data collection.



- The large samples from 2 IPs: **10^6** Higgs, **10^{12}** Z, **10^8** W bosons, provide a unique opportunity for
  - High precision Higgs, EW measurements,
  - Study of flavor physics (b, c, tau) and QCD,
  - Probe physics beyond the standard model.
  - …

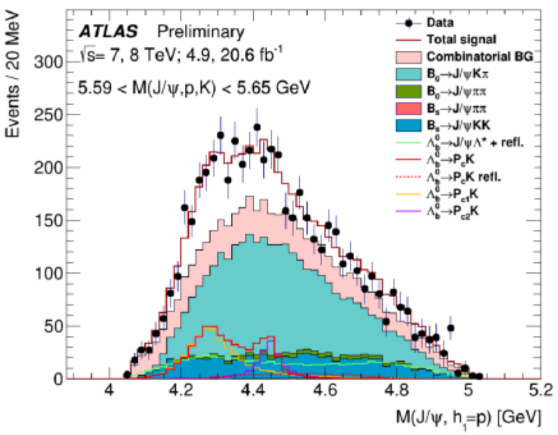
Particle identification (PID)

- **PID is essential for flavor physics**
  - Suppressing combinatorics
  - Distinguishing between same topology final-states
  - Adding valuable additional information for flavor tagging of jets
  - …

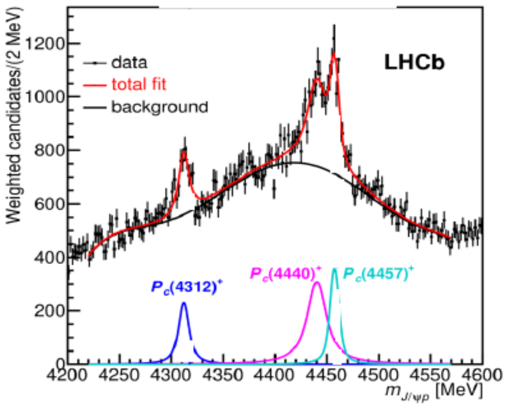
Pentaquark search in $\Lambda_b \to J/\Psi pK$

Without PID

[ATLAS-CONF-2019-048]

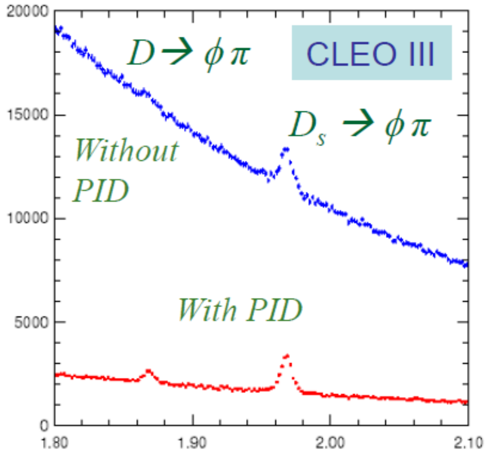


With PID ~ no background



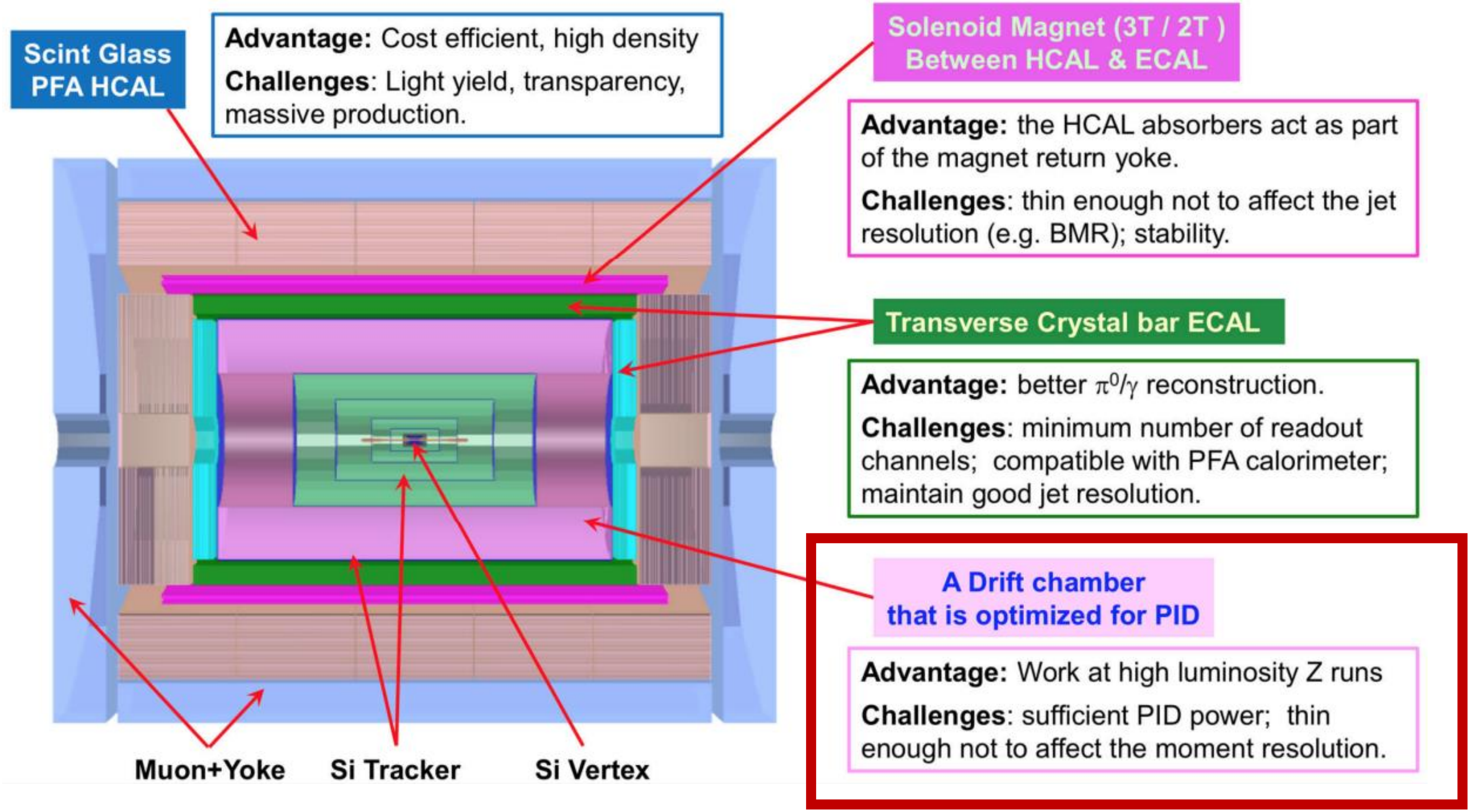
[LHCb, PRL 122 (2019) 222001]

$D_{(s)} \to \phi\pi$



* Figures from Guy Wilkinson's slides

CEPC 4th concept detector

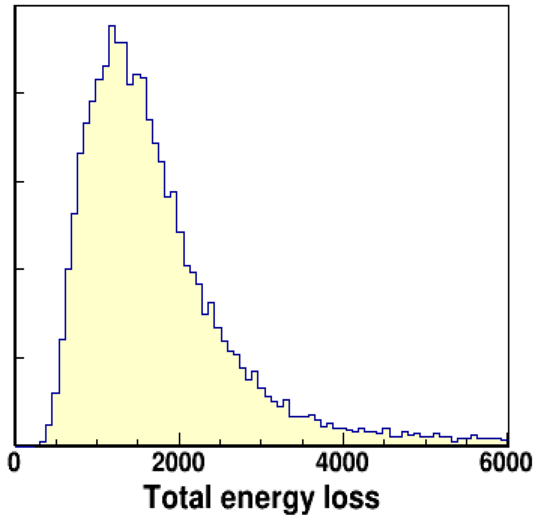


Preliminary PID requirement: >2σ K/π separation for 20 GeV/c tracks

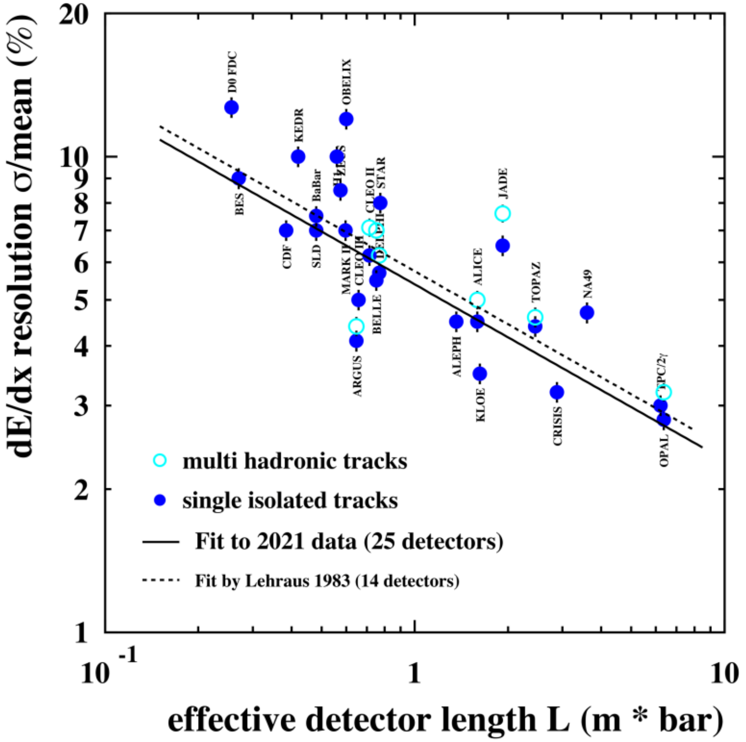
Energy loss measurement: dE/dx

- **Main mechanism: Ionization of charged tracks**
- **Traditional method: Total energy loss (dE/dx)**
  - Landau distribution due to secondary ionizations
  - Large fluctuation from many sources: energy loss, amplification …

Integrated charge



"Lehraus" Plot 2021*



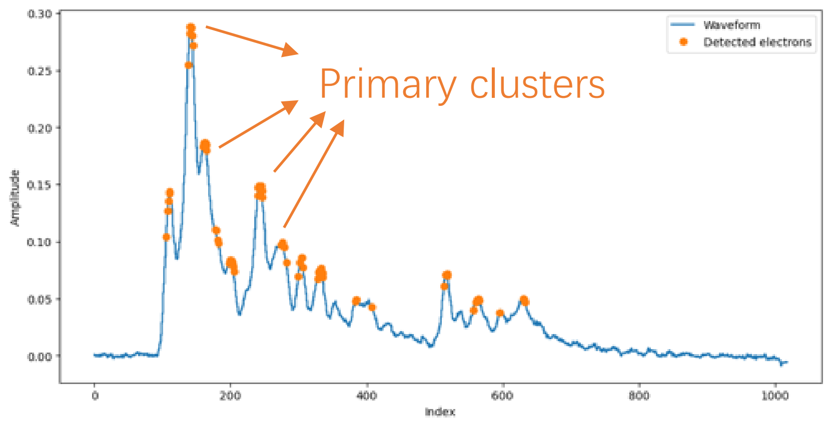
- **Fit by Lehraus 1983:**
  - dE/dx res. = **5.7** ∗ $L^{-0.37}$ (%)
- **Fit in 2021:**
  - dE/dx res. = **5.4** ∗ $L^{-0.37}$ (%)
- **No significant improvement in the past 40 years**

* From Michael Hauschild's talk @ RD51 workshop

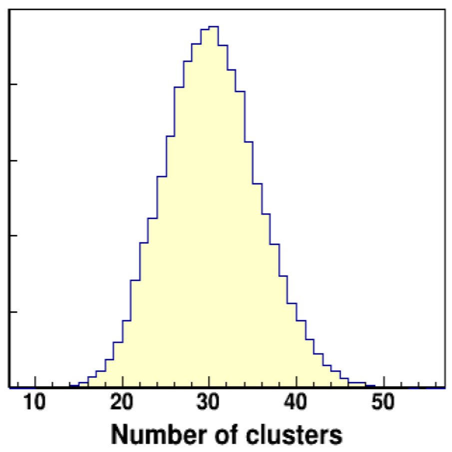
Cluster counting measurement: dN/dx

- **Alternative method: Counting primary clusters (dN/dx)**
  - Poisson distribution ➔ Get rid of the secondary ionizations
  - **Small fluctuation ➔ Potentially, a factor of 2 better resolution than dE/dx**

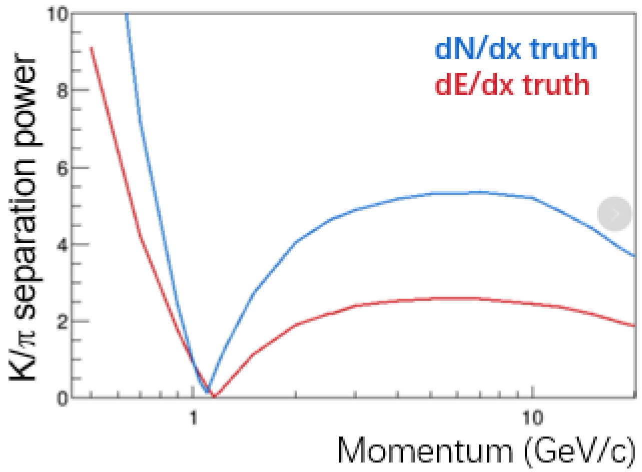
Waveform



Counting clusters



K/π separation power
dN/dx vs dE/dx



Require fast electronics and sophisticated counting algorithm

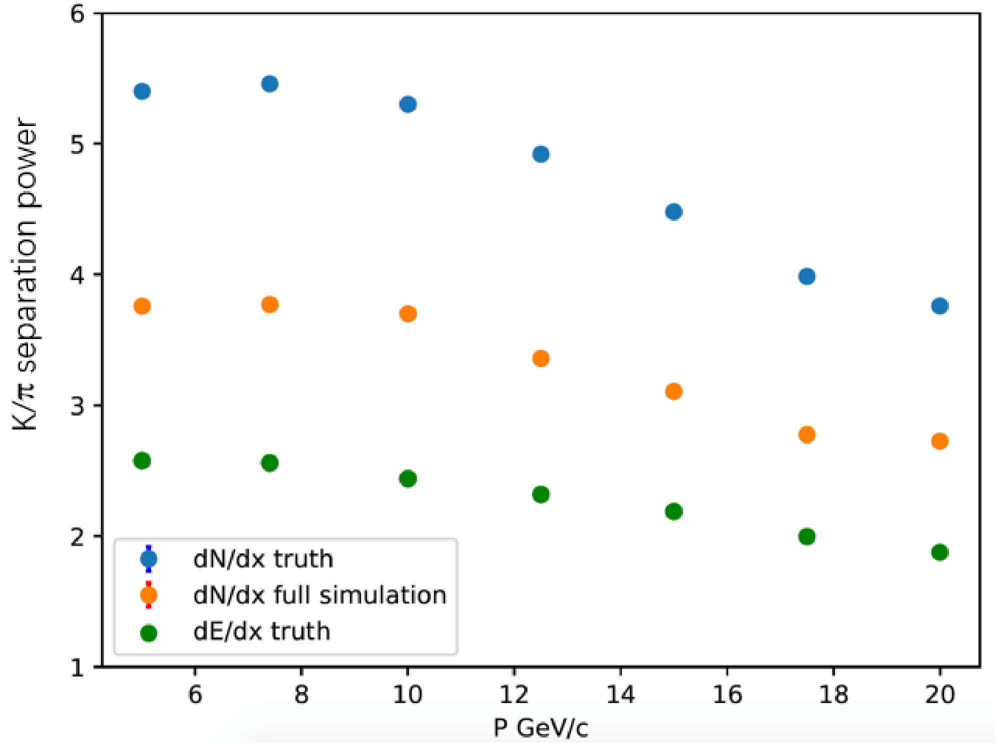
dN/dx is extremely powerful, proposed in ILC, FCC-ee, CEPC

Preliminary DC design according to previous simulation studies

Optimized DC Parameters

DC Parameters	
Radius extension	800-1800 mm
Length of outermost wires（cosθ=0.82)	5143 mm
Thickness of inner CF cylinder	200 μm
Outer CF frame structure	Equivalent CF thickness: 1.63 mm
Thickness of end Al plate	35 mm
Cell size	18 mm × 18 mm
# of cells	24766
Ratio of field wires to sense wires	3:1
Gas mixture	He/iC_4H_{10}=90:10

K/π separation power vs P (1m track length, cosθ=0)



Separation power

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

2σ K/π separation for 20 GeV/c tracks could be achieved (preliminary)

Further efforts on cluster counting

Major challenges	Efforts

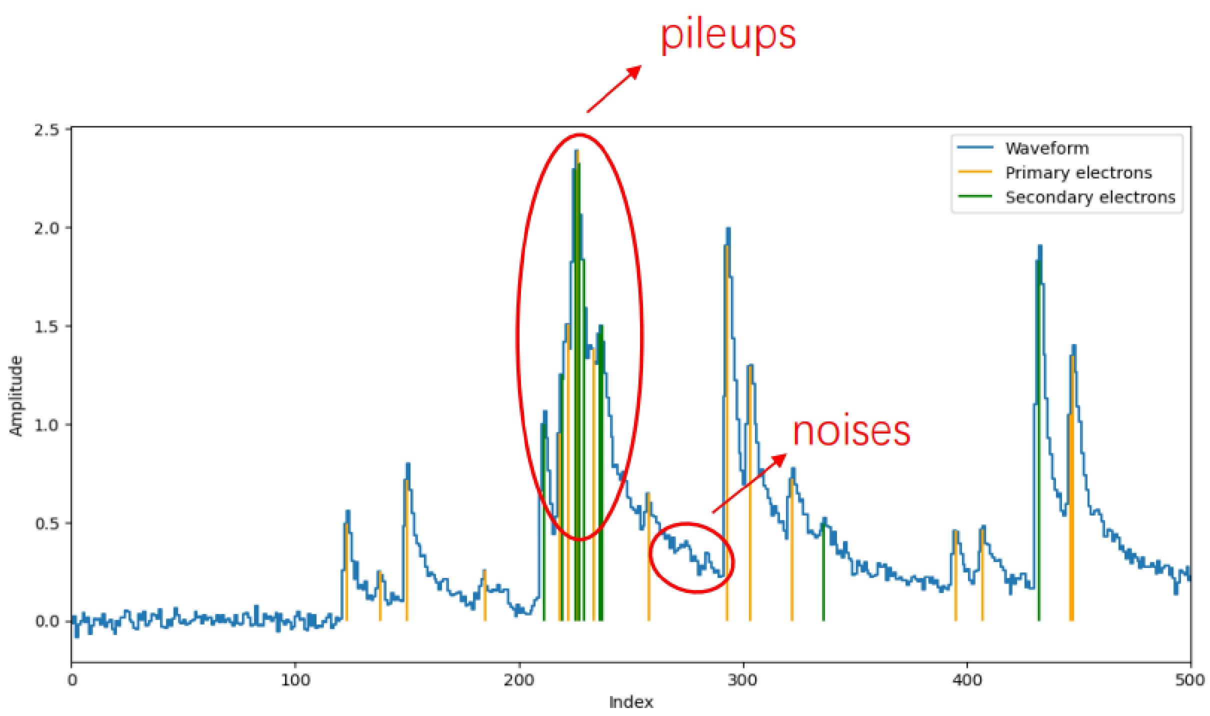
Hardware:
- Detector design
- Working gas
- Fast electronics
  - Bandwidth > 1 GHz
  - Amplification > 10
  - Sampling rate > 1.5 GS/s
  - ADC resolution > 12 bit
More separatable and less overlapped pulses between clusters (Detector design, Working gas)
Good enough pulse shape resolution (electronics requirements)
	Prototype experiment
	- Test beam
	- Radioactive source

Software: - Pileup recovery - Noise discrimination - Secondary electrons removal Efficient counting	**Deep learning algorithm**

Reconstruction with Deep Learning

1. DL algorithm for full simulation study
2. DL algorithm for test beam data

Reconstruction algorithm



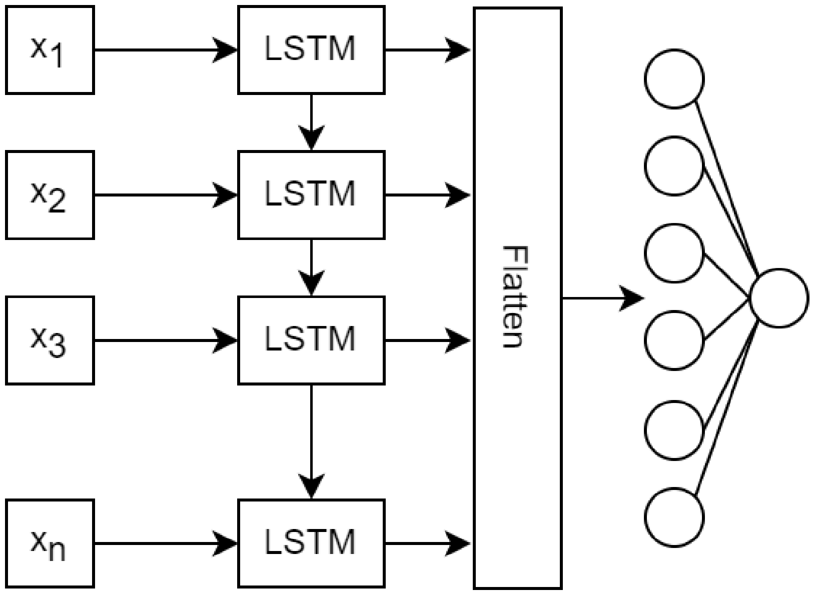
Simulated waveform of a DC cell. Orange lines are primary electrons. Green lines are secondary electrons.

- **Task:**
  - Both **primary electrons** and **secondary electrons** contribute peaks on the waveform
  - Find the number of peaks from **primary electrons**
- **Traditional algorithm:**
  - Use partial information of the raw waveform
  - Require prior knowledge
- **Deep learning could be more powerful** because
  - make full use of the waveform information
  - automatically learn characteristics of signals and noises from large labeled samples

Supervised model for simulated samples

Peak finding with LSTM

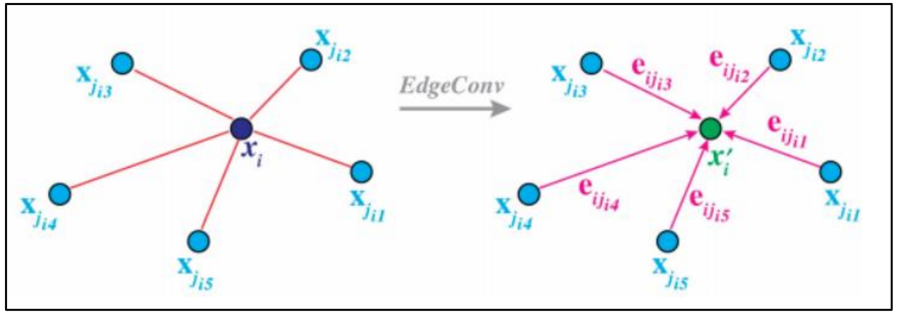
Why LSTM? ➔ Waveforms are time series



- Architecture: LSTM (RNN-based)
- Method: Binary classification of signals and noises on slide windows of peak candidates

Clusterization with DGCNN

Why DGCNN? ➔ Locality of the electrons in the same primary cluster, perform massage passing through neighbor nodes in GNN

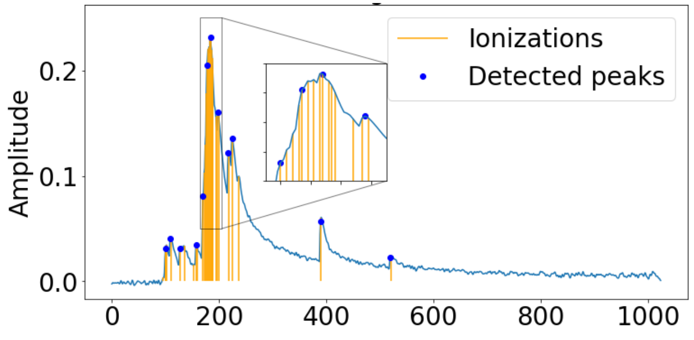


arXiv: 1801.07829

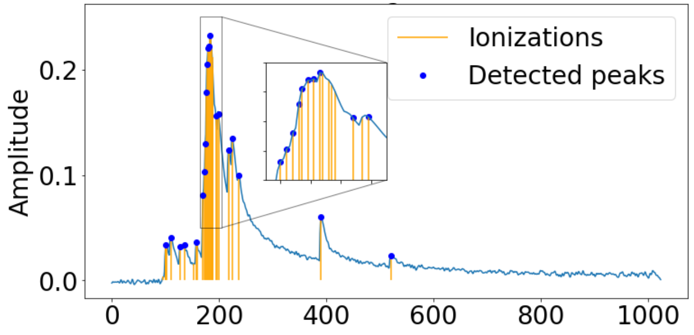
- Architecture: DGCNN (GNN-based)
- Method: Binary classification of primary and secondary electrons

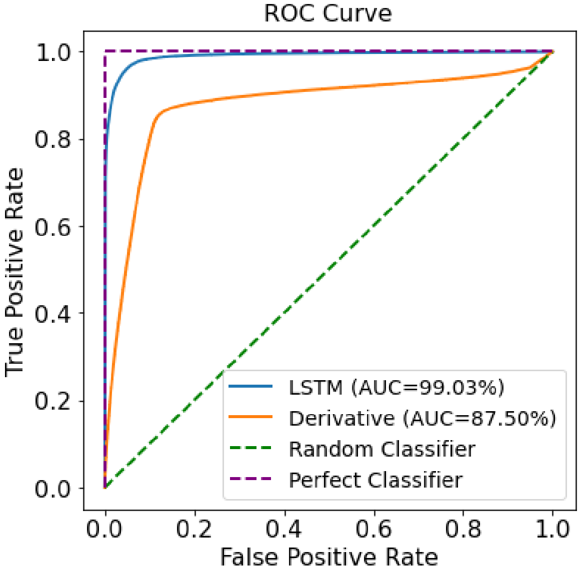
Results for simulated samples

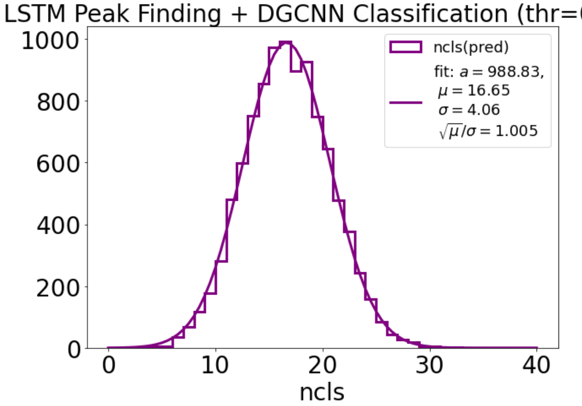
Derivative-based method

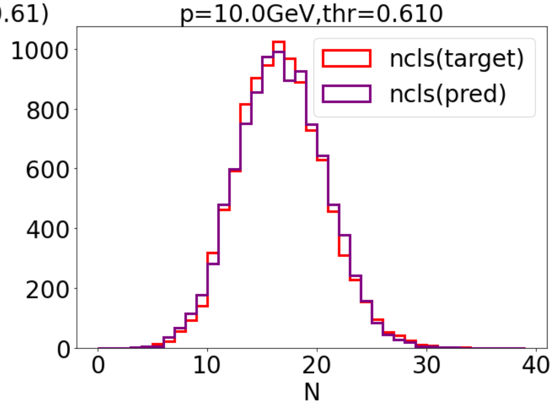


LSTM







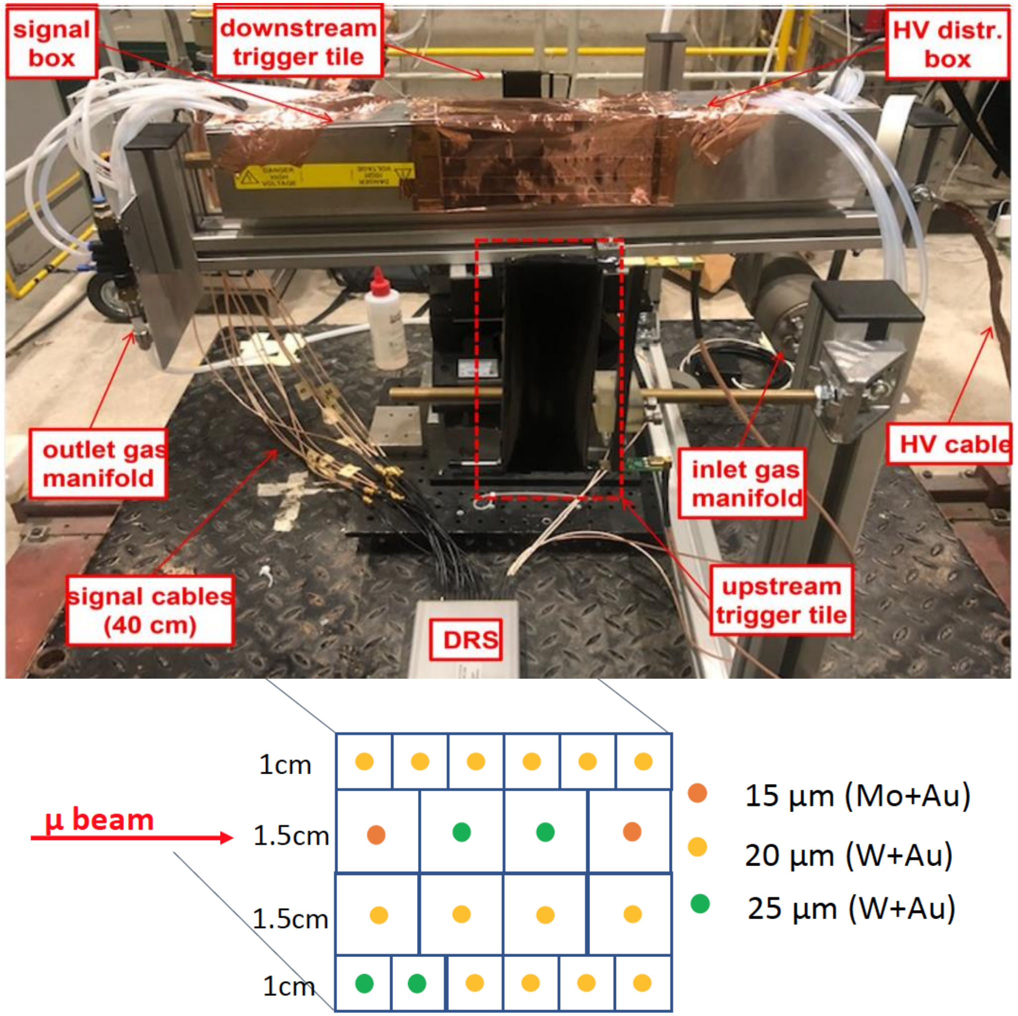


Model	μ	σ
MC truth (input)	**16.53**	**3.93**
Traditional alg.	18.67	4.60
Deep learning alg.	**16.65**	**4.06**

LSTM model is a better classifier compared to derivative-based model

Results from deep learning is much closer to MC truth than traditional algorithm ($10\ \mathrm{GeV}/c$ pions)

Deep learning algorithm for test beam data

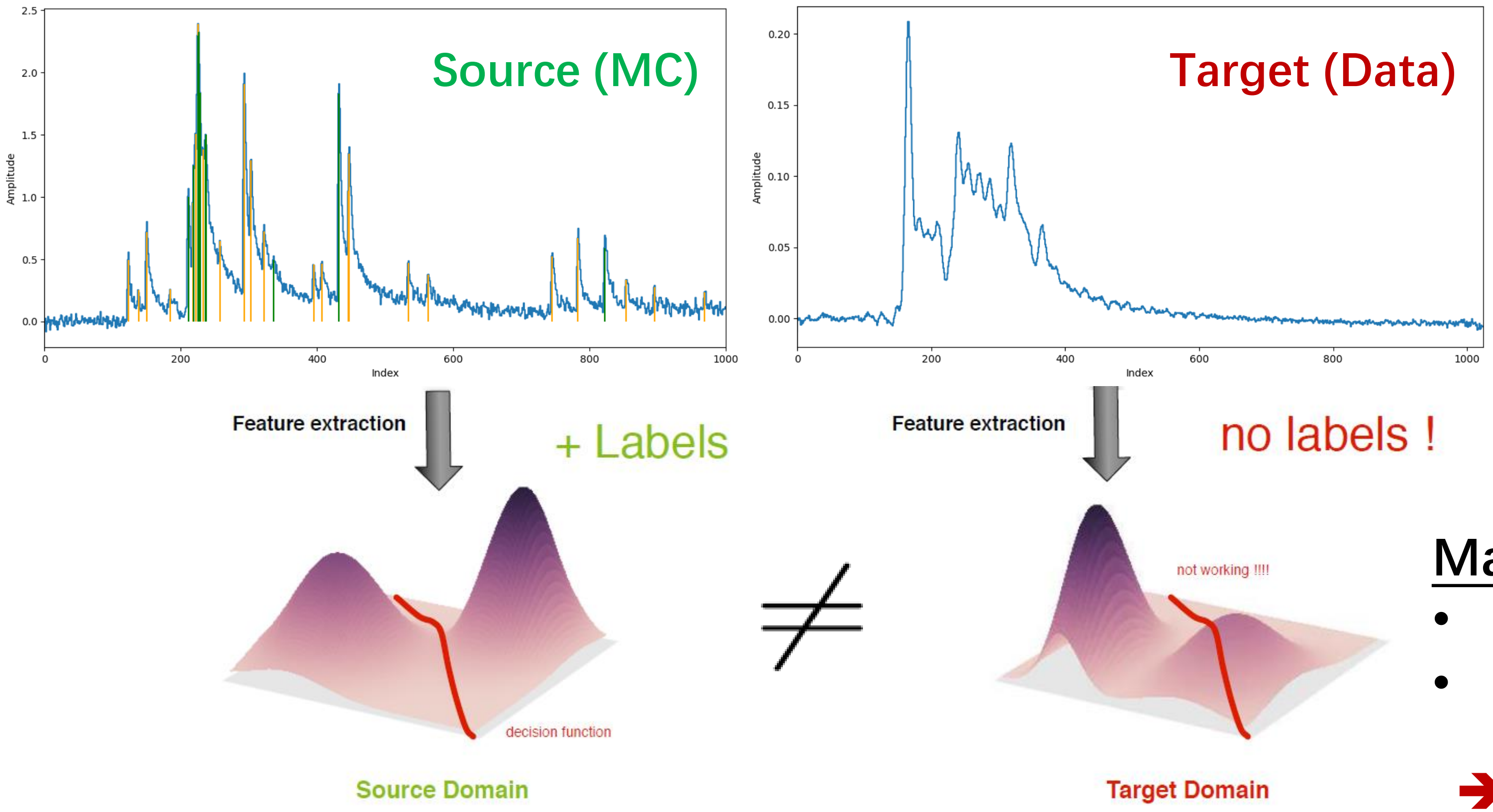


Since 2021, three rounds of test beam experiment @ CERN organized by INFN colleagues:

- Participate data taking and collaboratively analyze the test beam data
- **Develop the deep learning algorithm for peak finding**

→ See Nikola De Filippis's talk for details of test beam

Challenge of deep learning algorithm on experimental data

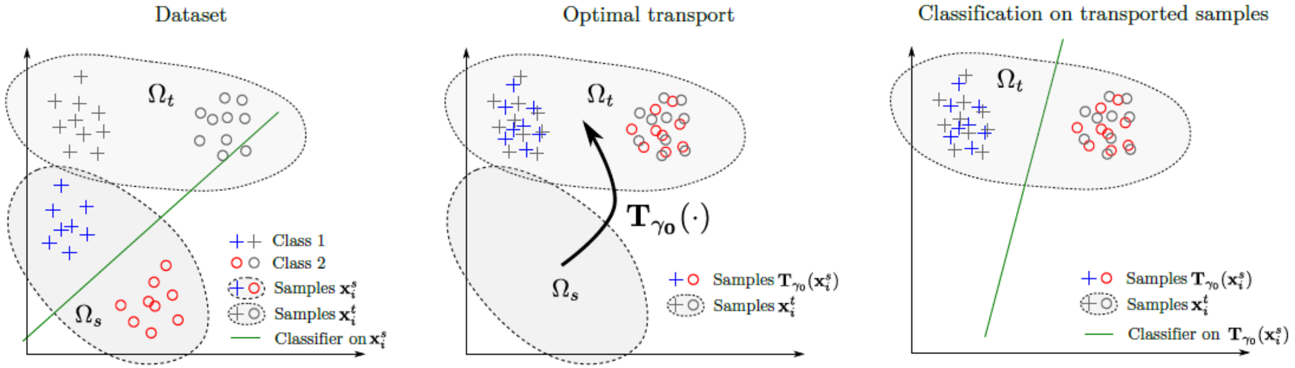


Main challenges:

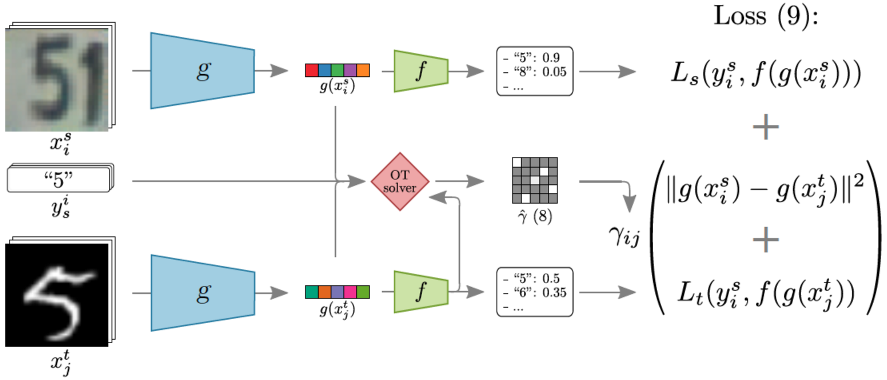
- Discrepancies between data and MC
- Lack of labels in experimental data

➔ **Cannot directly apply the supervised model trained by simulated samples**

Semi-supervised domain adaptation



Align data/MC samples with Optimal Transport



Original work by Damodaran, et al. (arXiv: 1803.10081)

Loss for labeled samples in source domain

$$\min_{f,g}\left[\sum_{i=1}^{m} L_s\left(y_i^s, f(g(x_i^s))\right) + \boxed{\frac{1}{m_l}\sum_{i=1}^{m_l} L_t\left(y_i^{t,l}, f\left(g(x_i^{t,l})\right)\right)} + \min_{\gamma\in\Delta}\sum_{i,j}^{m}\gamma_{ij}\left(\alpha\|g(x_i^s)-g(x_j^t)\|^2 + \lambda_t L_t\left(y_i^s, f\left(g(x_j^t)\right)\right)\right)\right]$$

Loss for labeled samples in target domain (THIS WORK)

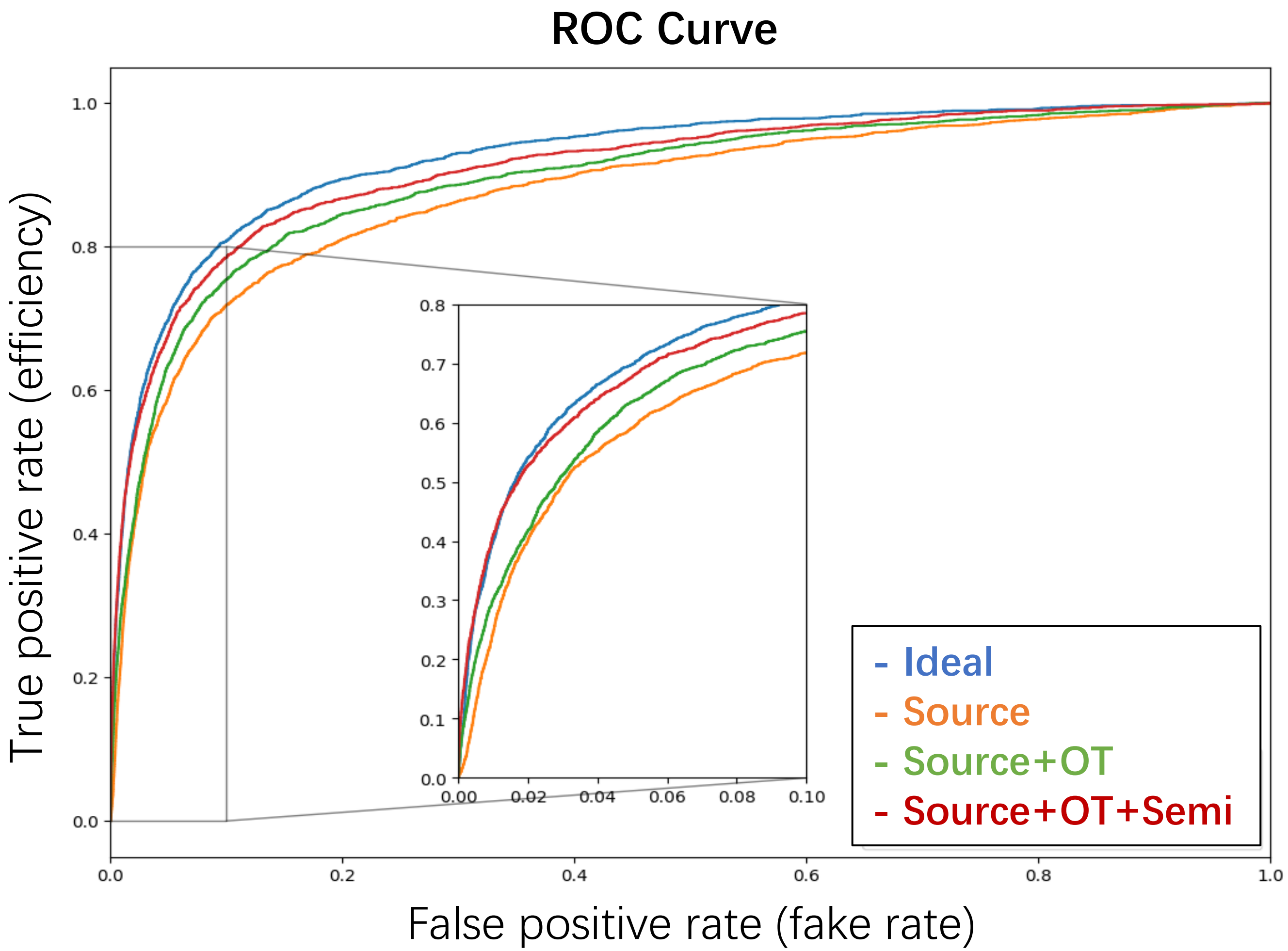
Cost of feature differences between source and target

Cost of ‘label’ differences between source and target

Cost of joint feature-label distribution for OT

Semi-supervised implementation in our work

Numeric experiment



Numeric experiment with pseudo data:

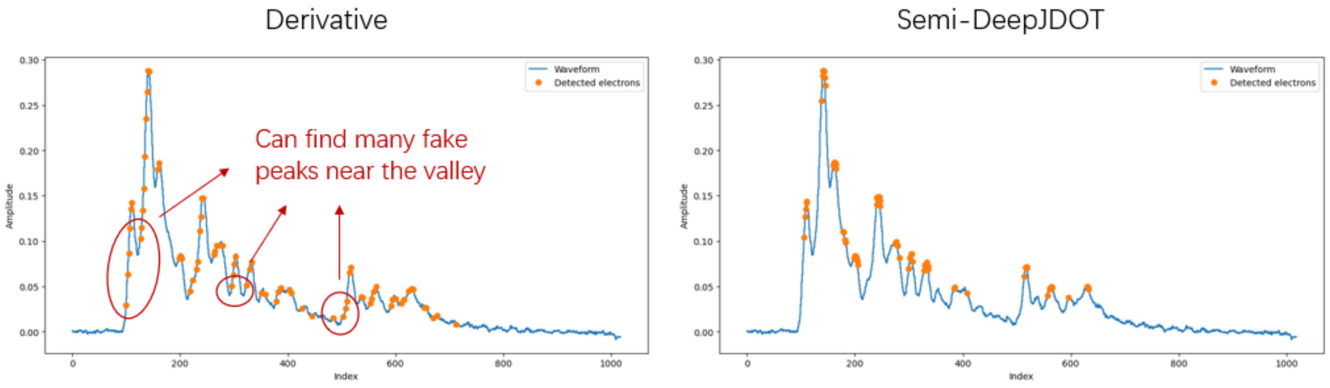
- Use labels in pseudo data to evaluate

Model	AUC	pAUC (FPR<0.1)
Ideal (supervised)	0.926	0.812
Source (baseline)	0.878	0.749
Source + OT	0.895	0.769
Source + OT + Semi (Semi-supervised DeepJDOT)	0.912	0.793

Validation: Performance of Semi-DeepJDOT model is very close to the ideal model (supervised model)

Peak finding for test beam data

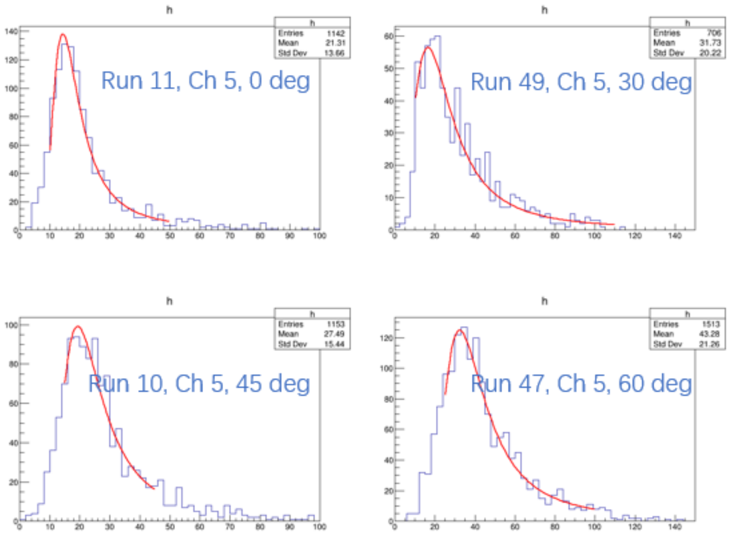
Single-waveform results between derivative alg. and DL alg.

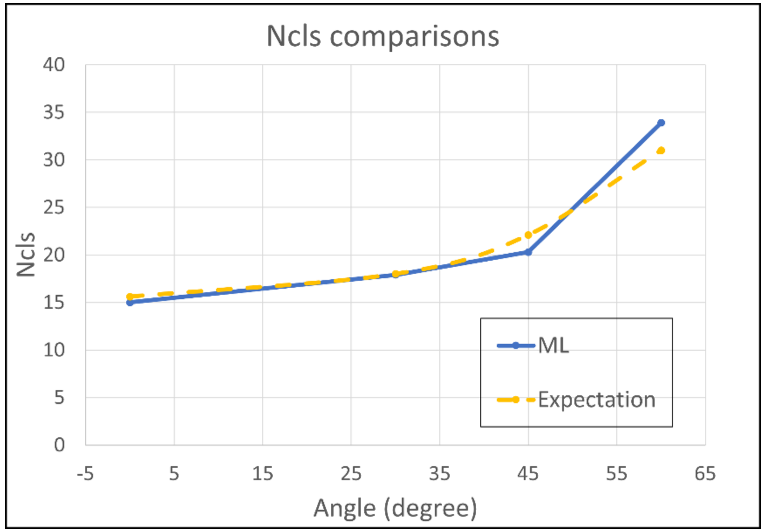


Note: Require similar efficiency for both cases

DL algorithm is more powerful to discriminate signals and noises

Multi-waveform results for samples in different angles





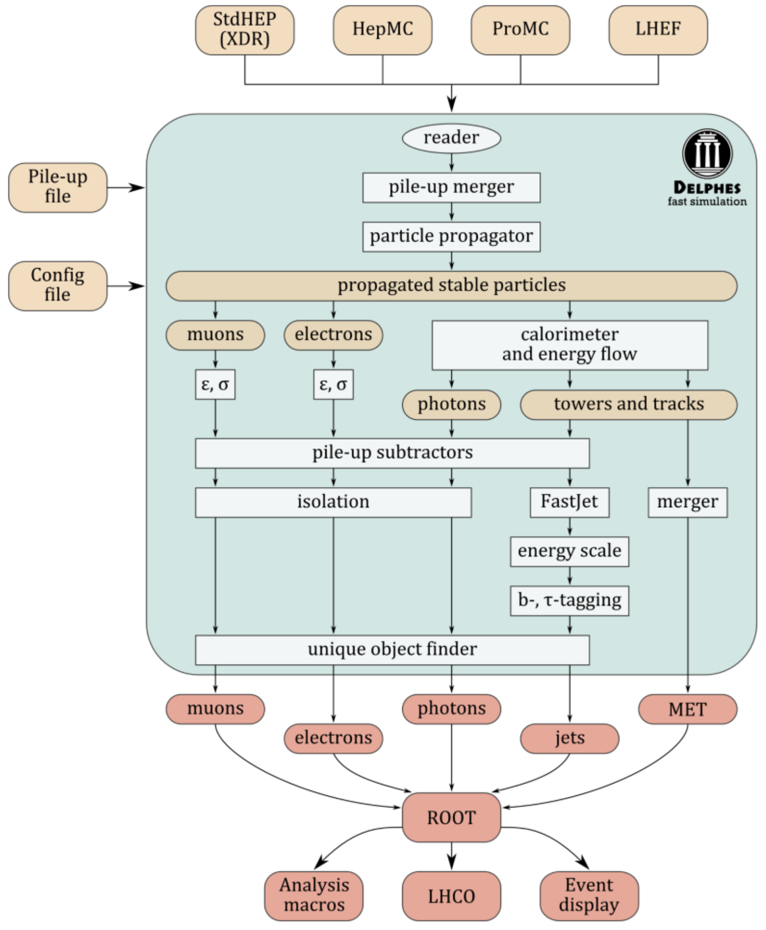
The trend of the results w.r.t. angle is consistent with theoretical estimation

Physics with Delphes

Fast simulation with Delphes

- **Delphes: A C++ framework, performing a fast multipurpose detector response simulation**
  - $10^2 \sim 10^3$ faster than the fully GEANT-based simulations
  - Sufficient and widely used for phenomenological studies

- **Develop dedicated PID modules (dN/dx and TOF) and perform quick physics studies**

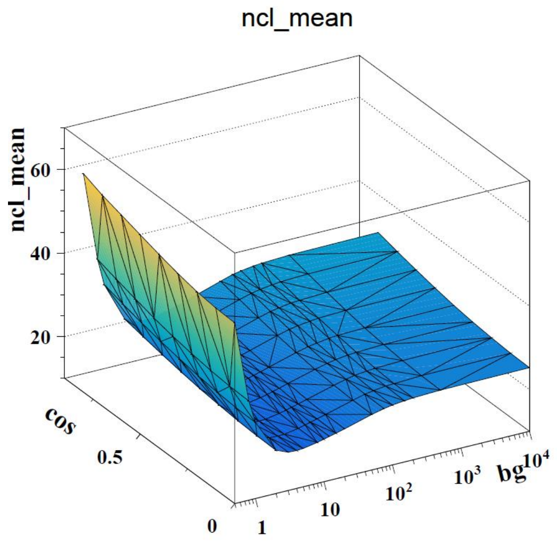


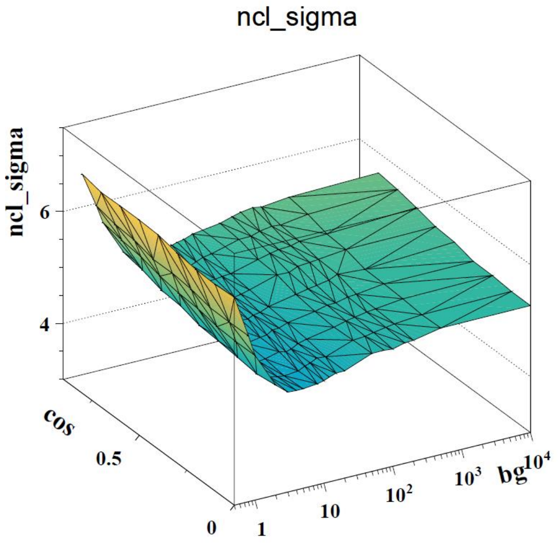
J. High Energ. Phys. **2014**, 57 (2014)

→ See Xu Gao's talk for details.

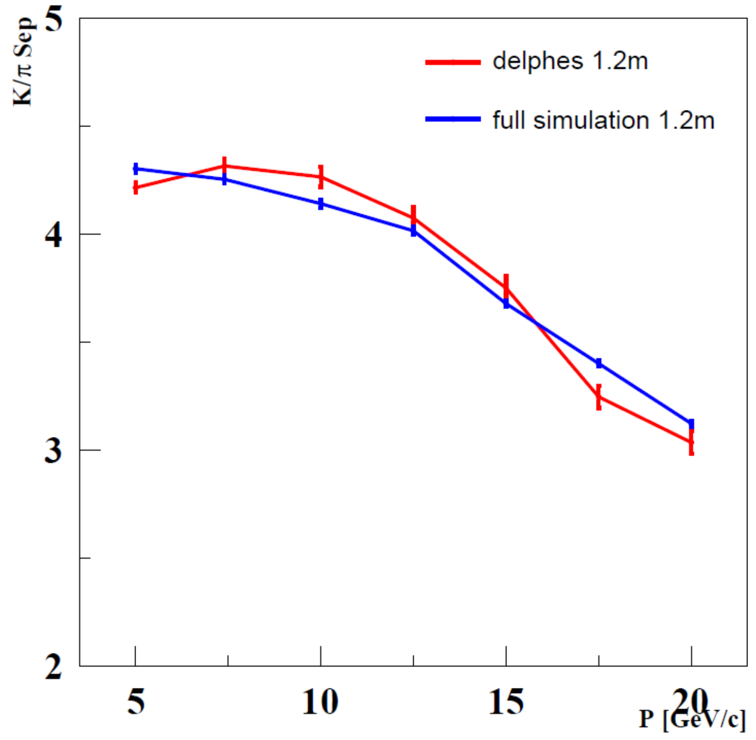
PID modules implementation

- **dN/dx parameterization from full simulation**
  - dN/dx_{mean} vs. βγ and cosθ
  - dN/dx_{sigma} vs. βγ and cosθ
- **TOF parameterization by assuming a resolution of 30 ps**





K/π separation power

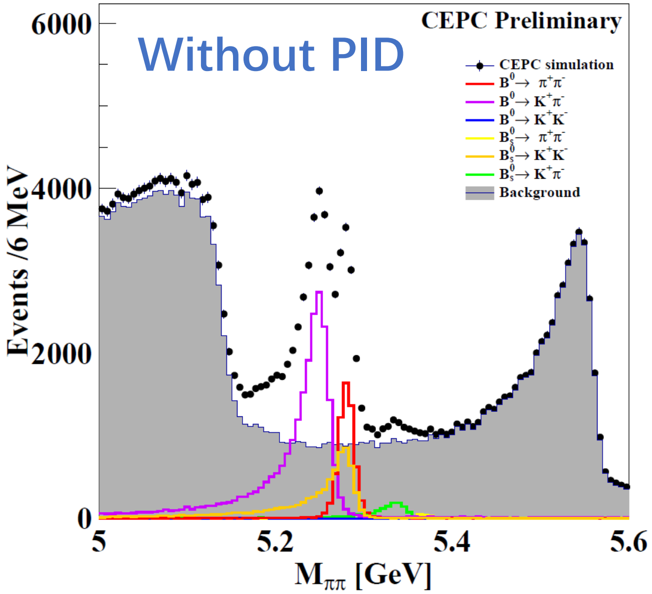


Good consistent to full simulation

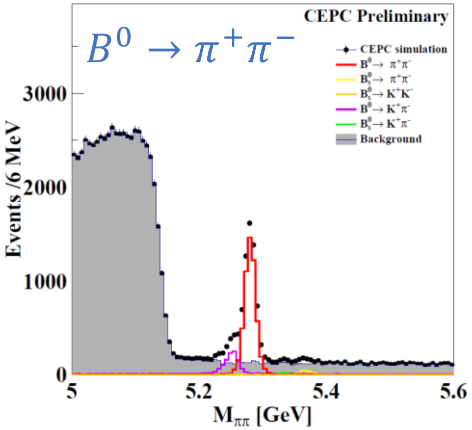
Study of $\mathbf{B^0_{(s)} \to h^+h'^-}$

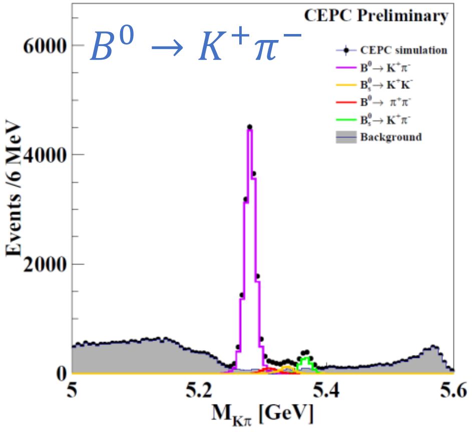
- **Motivation**
  - Rich physics programs in $B^0_{(s)} \to h^+h'^-$ decays
    - Time-dependent asymmetry, direct CP violation, lifetime measurement, …
  - Good test bed to study impact of PID in flavor physics
  - Explore physics potential of Tera-Z

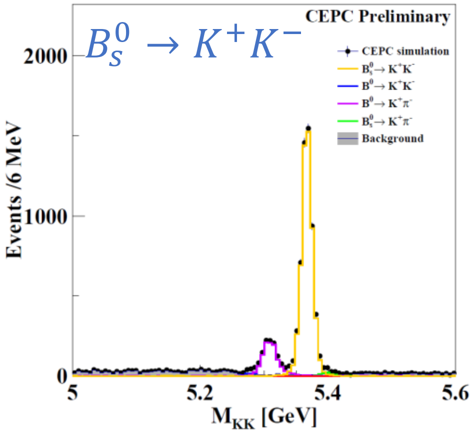
- **Significantly improved SNR with PID**
- **More detailed studies ongoing**



With PID



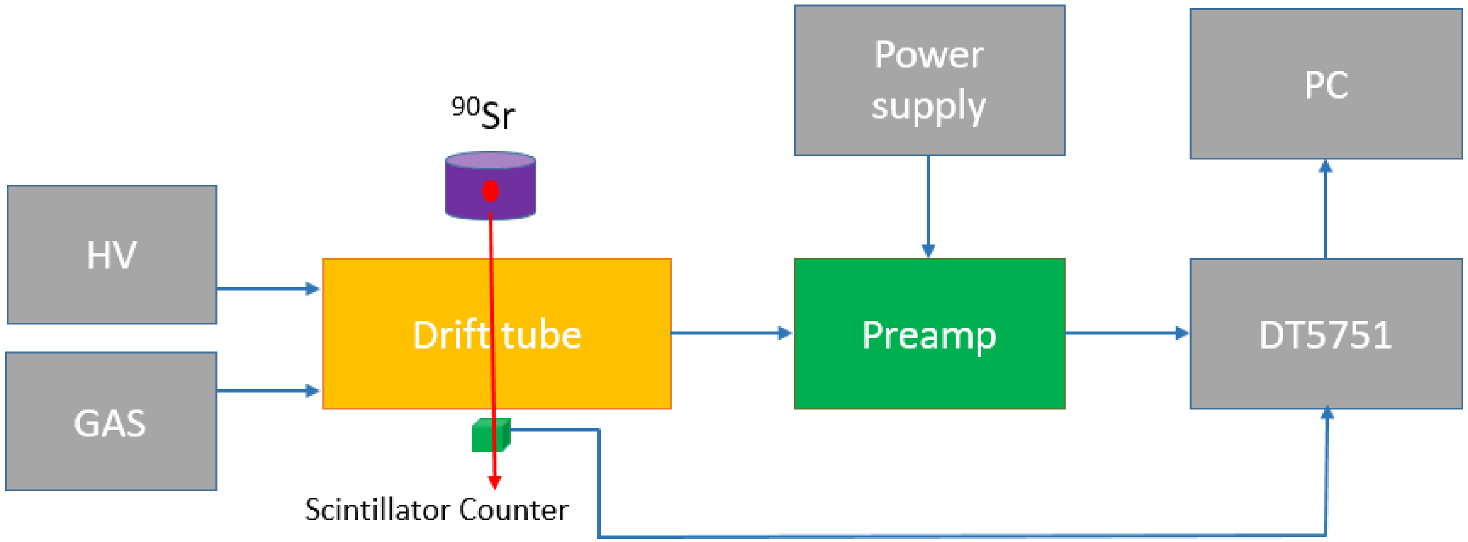




Prototype Experiment

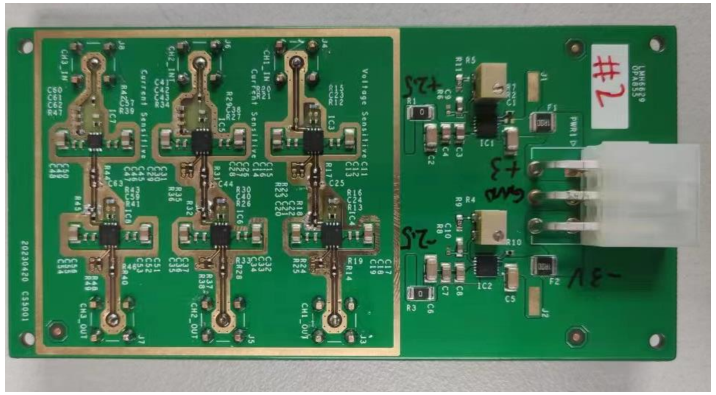
Drift tube experiment on radioactive source

Experiment layout



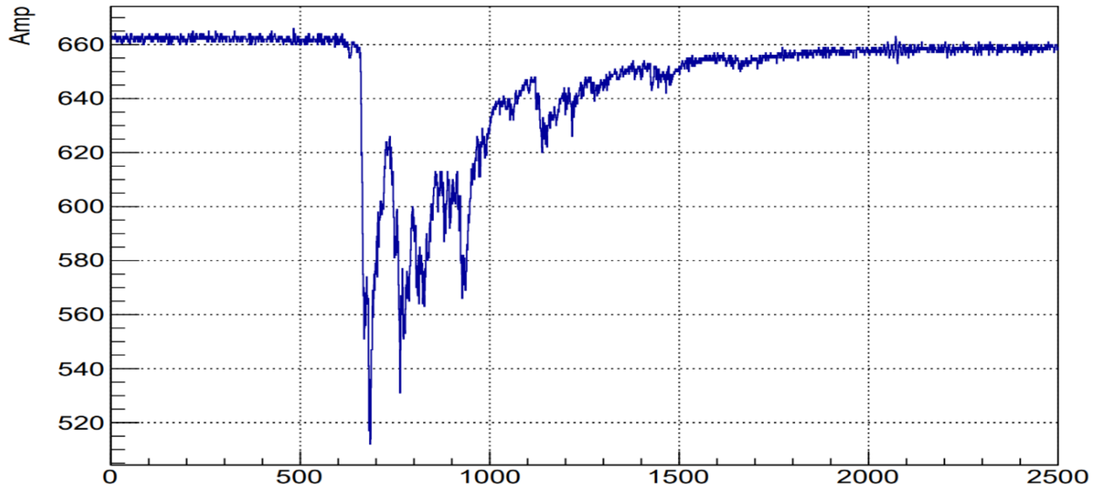
- Diameter of the tube: 30 mm
- Working gas: He/iC_4H_{10}=90:10

Preamplifier

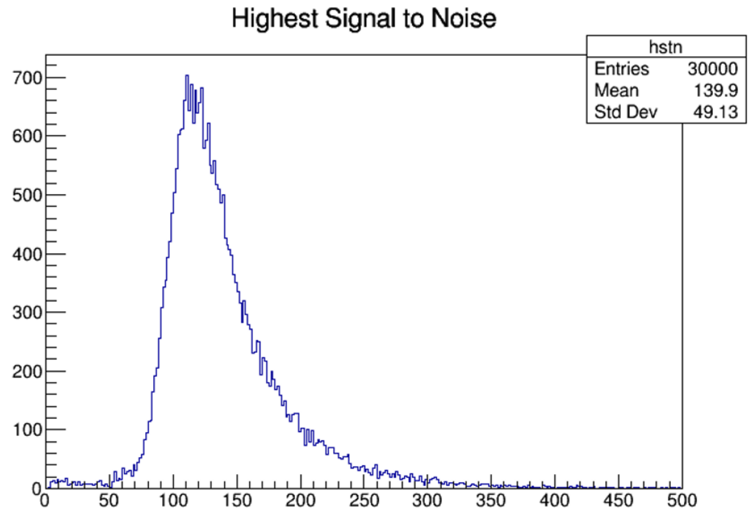


- Bandwidth ~1GHz
- ADC sampling rate 1GHz

Preliminary results

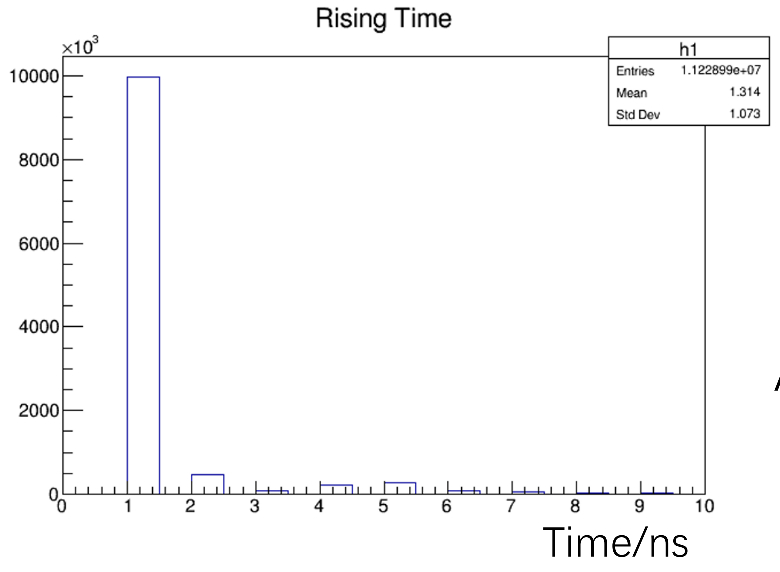


Typical recorded waveform



$$SNR = \frac{Highest\ Signal}{RMS_{noise}}$$

Average value: ~140



Average risetime: ~1.3ns

- **Preliminary tests show**
  - Low noise (SNR~140)
  - High bandwidth
  - Fast risetime peaks (~1.3 ns)

Summary

- **Preliminary full simulation study shows 2σ K/π separation can be achieved at 20 GeV/c for CEPC DC with 1m thickness**

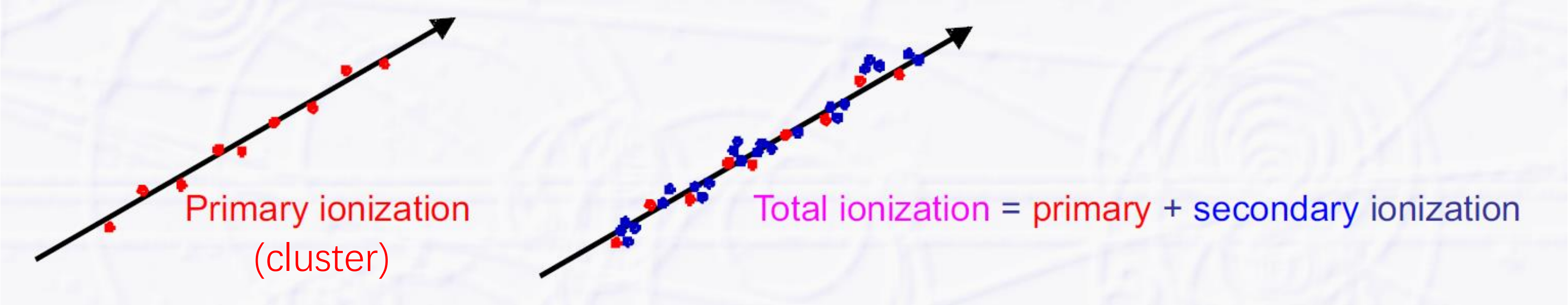
- **Recent efforts on cluster counting feasibility**
  - Deep learning algorithms: models for both simulated and experimental samples, outperform traditional algorithms
  - Delphes fast simulation: PID performance consistent to full simulation, improved signal sensitivity with PID in physics channels
  - Prototype experiment: improved setup and new preamplifier, show potential for cluster counting

- **Future works**
  - Optimize deep learning algorithm, prepare papers
  - Test beam data analysis
  - More prototype experiments

Backup

PID by ionization

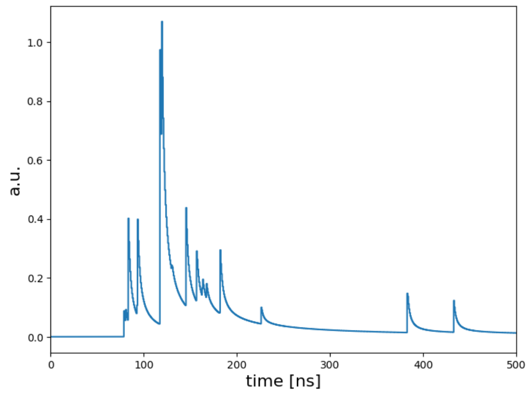
- **Main mechanism: Ionization of matter by charged particles**



- Number of clusters per unit length is Poisson-distributed
- Primary electrons sometimes get large energies
  - Can make secondary ionization
  - Can even create visible secondary track (“delta-electron”)

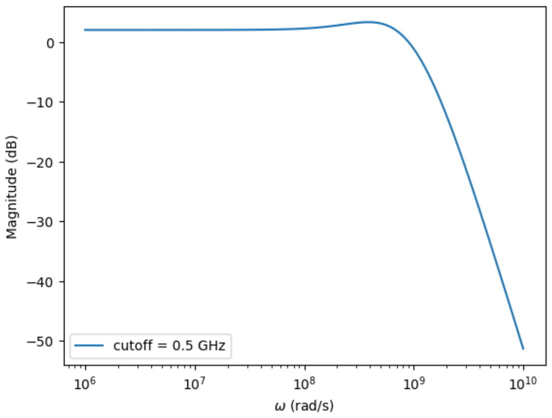
Waveform simulation

Induced signal



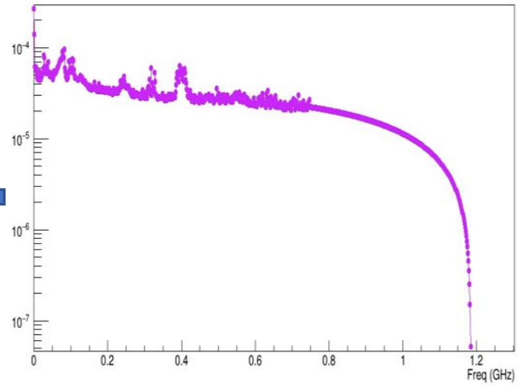
Fast simulation based on Garfield++ study

Electronics response



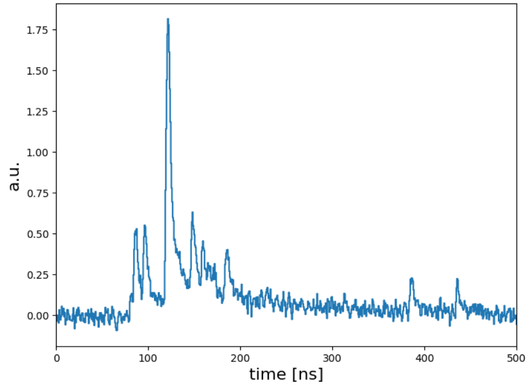
From electronics using in the beam test

Noise



From the beam test

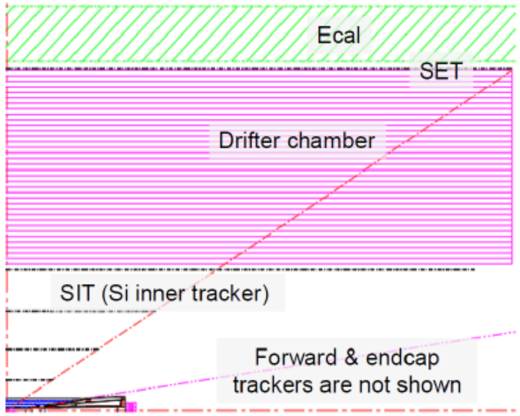
Waveform

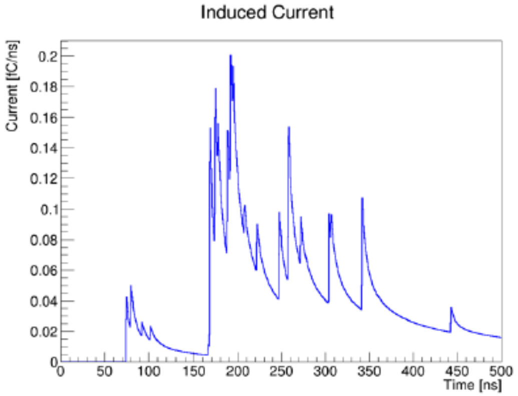


Data size estimation

Drift Chamber

Mingyi Dong





- **A Drift Chamber optimized for PID**
  - better than 2σ K/π separation for P < 20GeV/c
- **Signal measurement**
  - dN/dx for cluster counting method
- **Signal characteristics**

Parameters	Value	Parameters	Value
Rising	0.5~1ns	Falling	~tens ns
Pulse width	Hundreds ns	Pulse spacing (overlapping)	few~dozen ns
Amplitude	Dozen~hundred nA	Pulse charge	Ten~dozen fC

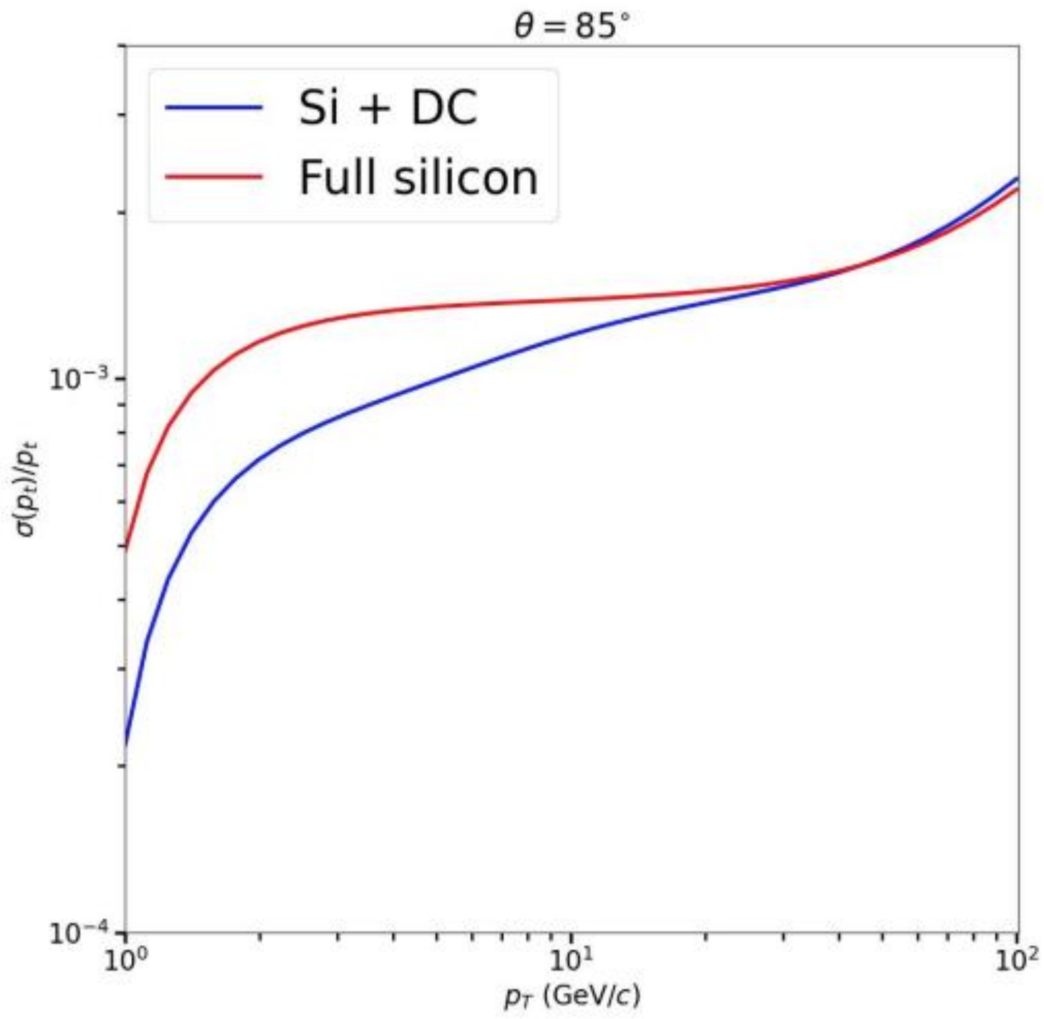
- **Detector channels and data rate**

		Higgs	Z
Trigger-less		256 Gbps	6.4 Tbps
Trigger	Trigger rate	1 kHz	100 kHz
	Max #wires/event	25k	10k
	bandwidth	20 Gbps	800 Gbps

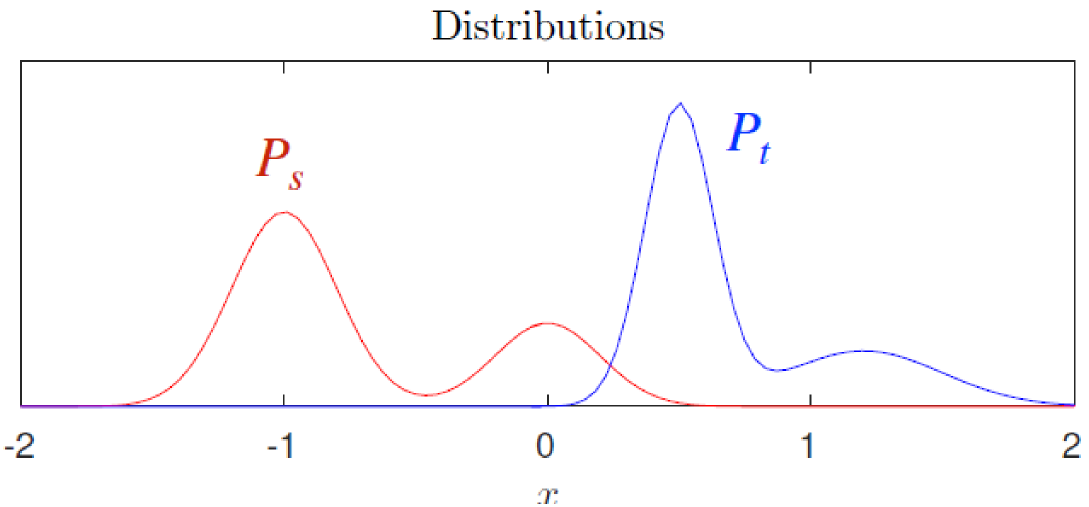
50 peaks/wire*, 16bit/peak from F.Grancagnolo

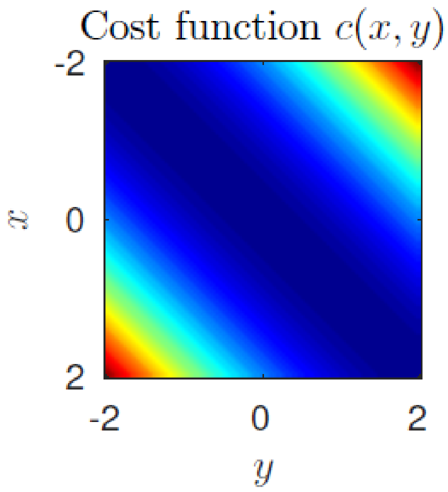
FY Guo, CEPC workshop 2021

Momentum resolution



Optimal transport (OT)





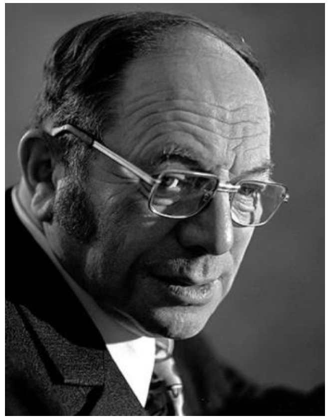
Monge

Optimal transport [Monge 1781, Kantorovich 1942]

Given $x \sim P_s$, $y \sim P_t$ and a cost function $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}^+$. Find a **probablistic coupling** $\gamma \in \Pi(\mathcal{X}_s \times \mathcal{X}_t)$

$$\gamma = \operatorname*{argmin}_{\gamma} \int_{\mathcal{X}_s \times \mathcal{X}_t} c(x,y)\gamma(x,y)dxdy = \operatorname*{argmin}_{\gamma} \mathbb{E}_{x,y\sim\gamma}[c(x,y)]$$

$$s.t.\gamma \in \Pi = \left\{\gamma \geq 0, \int_{\mathcal{X}_s} \gamma(x,y)dy = P_t, \int_{\mathcal{X}_t} \gamma(x,y)dx = P_s\right\}$$

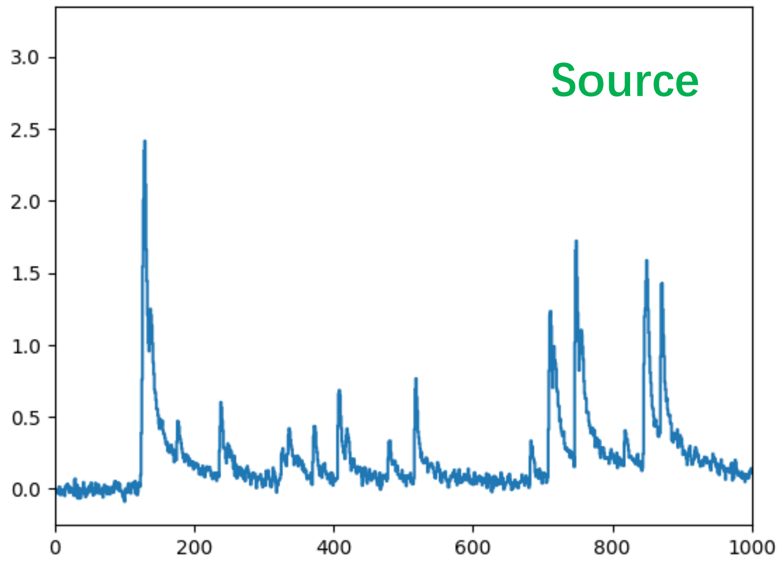


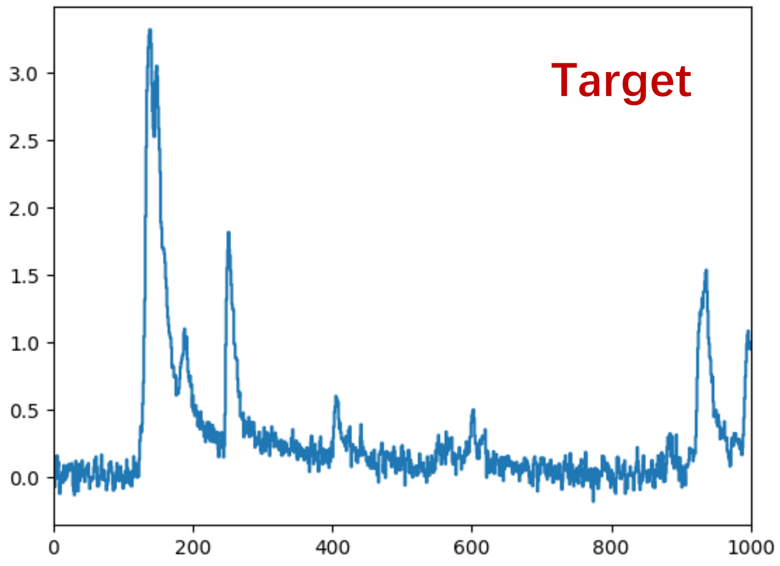
Kantorovich
(Nobel Prize 1975)

Kantorovich, L. (1942). On the translocation of masses. Acad. Sci. URSS (N.S.), 37:199-201.

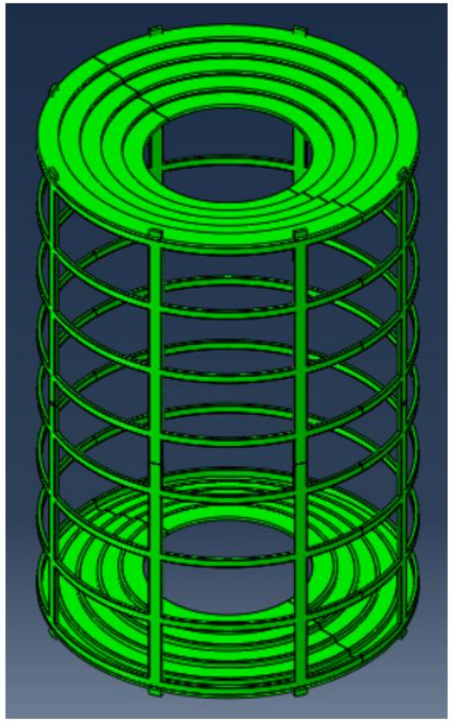
Pseudo data samples

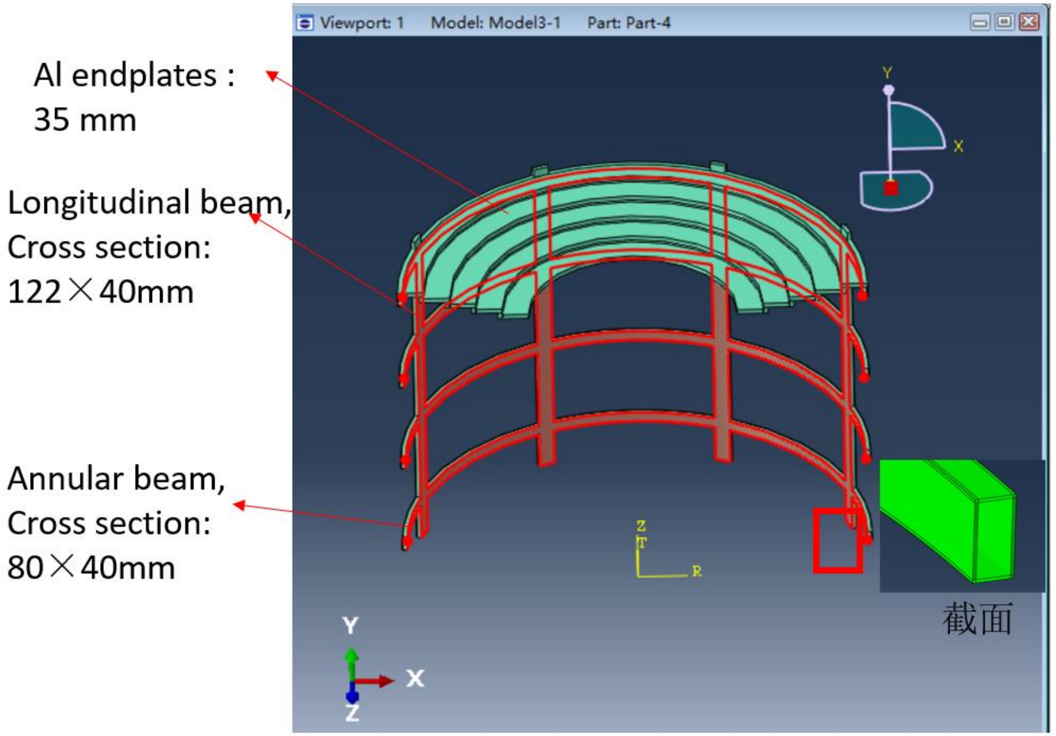
- **Samples**
  - Source domain: ~2ns risetime, ~10% noise
  - Target domain: ~4ns risetime, ~20% noise





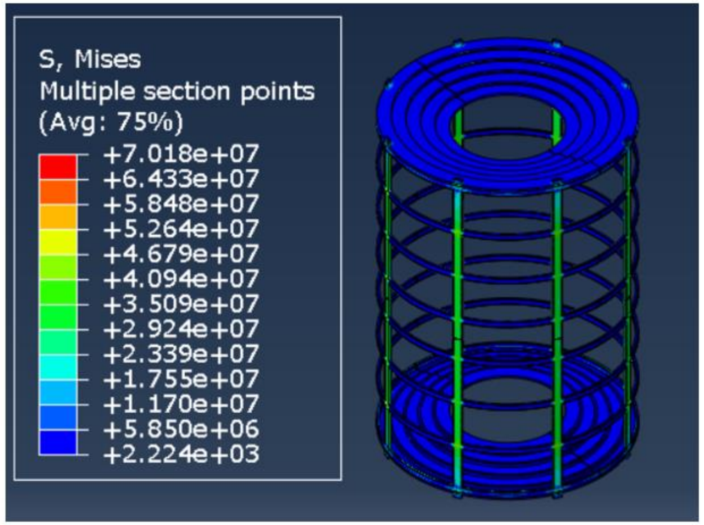
Mechanical study: support structures



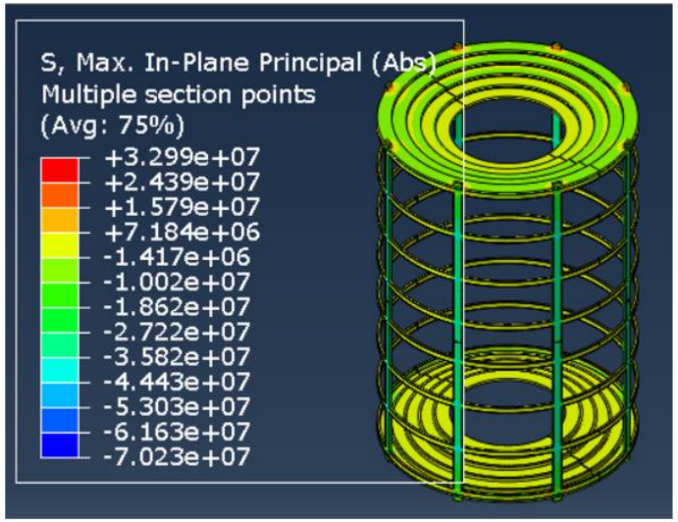


- Carbon fiber frame structure, including 8 longitudinal hollow beams and 8 annular hollow beams
- Thickness of inner CF cylinder: 200 μm/layer
- Effective outer CF frame structure: 1.63 mm
- Thickness of end Al plate: 35 mm

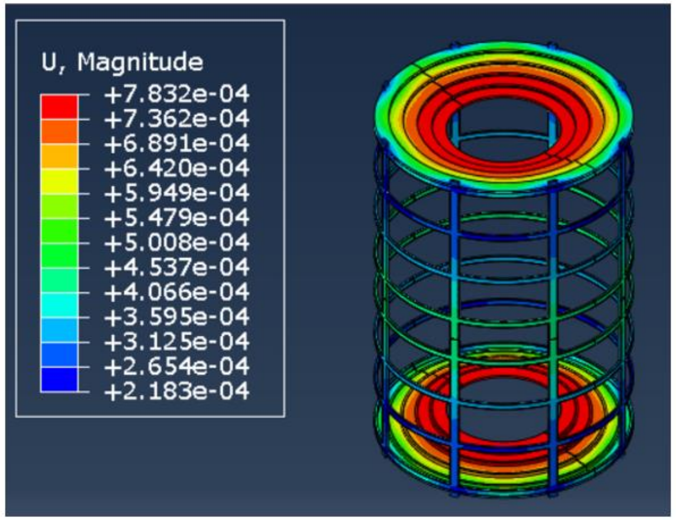
Mechanical study: stability



Mises stress



Principle stress



Deformation

Finite element model——wire tension + weight loads (supported by eight blocks at each endplate)

Mises stress: 70MPa
Principal stress : 33MPa
Deformation: 0.8mm
Buckling coefficient: 17.2 , it is safe

The support structure is stable, and the deformation is acceptable