Applications of Machine Learning at BESIII

Beijiang LIU[1], Xian Xiong[1], Guoyi Hou[1], Shiming Song[2]

(on behalf of the BESIII collaboration)

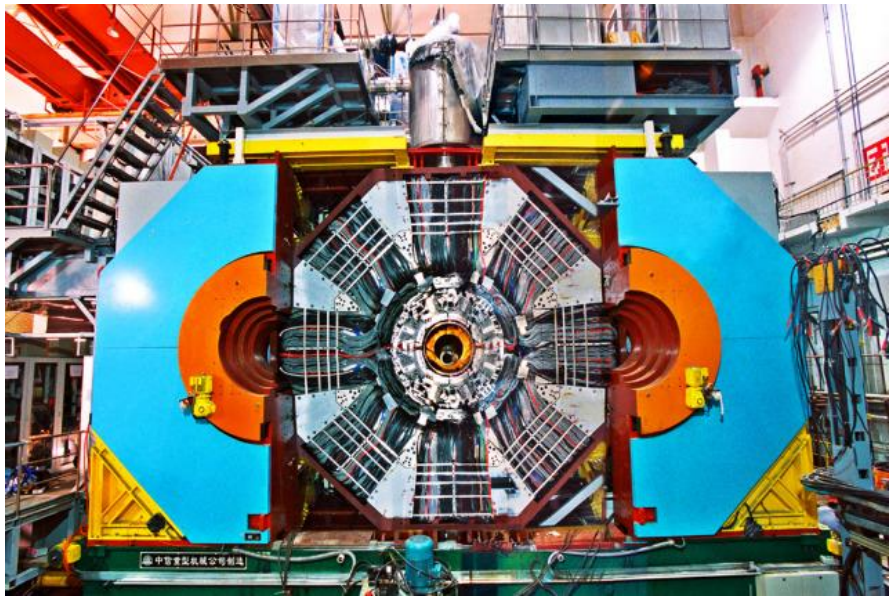
[1]Institute of High Energy Physics, CAS, [2]Sichuan University

CHEP 2018
Sofia, Bulgaria

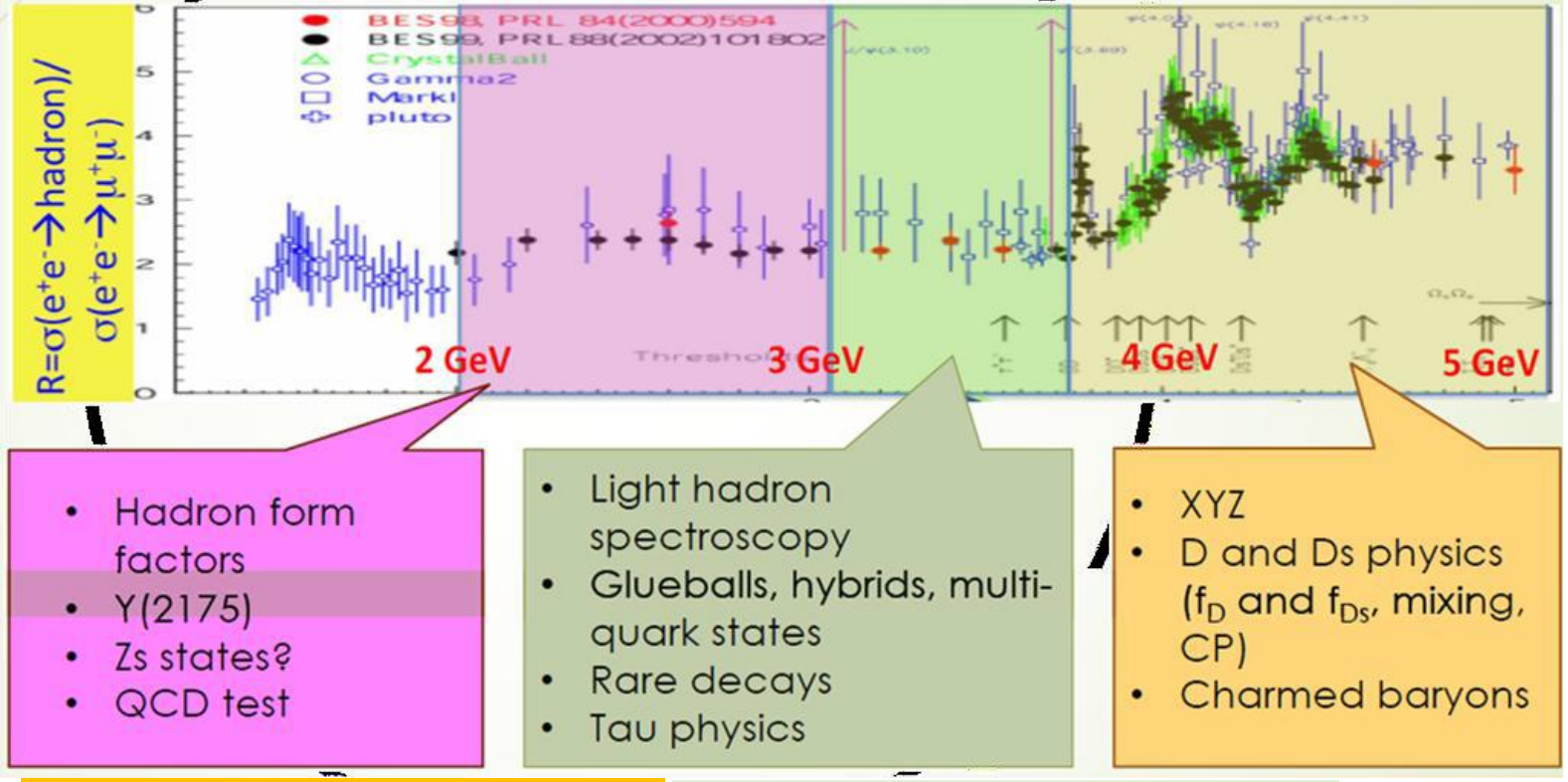
BESIII

BEijing Spectrometer experiment (BESIII) at Beijing Electron Positron Collider (BEPCII)

- **2004:start BEPCII construction**
- **2009-now:data taking**
- **Beam energy:1.0-2.3GeV**
- **Max luminosity: $10^{33} cm^{-2} s^{-1}$ (reached in 2016)**



- **A τ-charm factory for hadron physics and flavor physics**
- **~400 members**
- **14 countries**



World largest J/ψ, $\psi(3686)$, $\psi(3773)$, $\psi(4160)$, $Y(4260)$, … produced directly from e^+e^- annihilation

Drift Chamber

$\sigma_{r\phi}$ = 130 μm (single wire)

σ_{pt}/p_t = 0.5 % @ 1 GeV

$\sigma_{dE/dx}$= 6 %

Time of Flight

σ_t = 90 ps (barrel)

σ_t = 120 ps/60 ps (end caps)

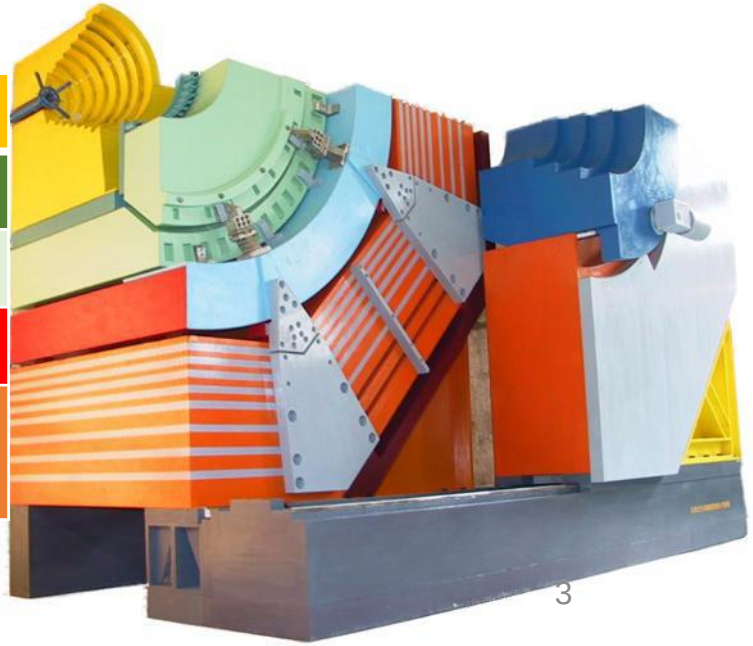
EM Calorimeter

σ_E/E < 2.5% @1 GeV (barrel)

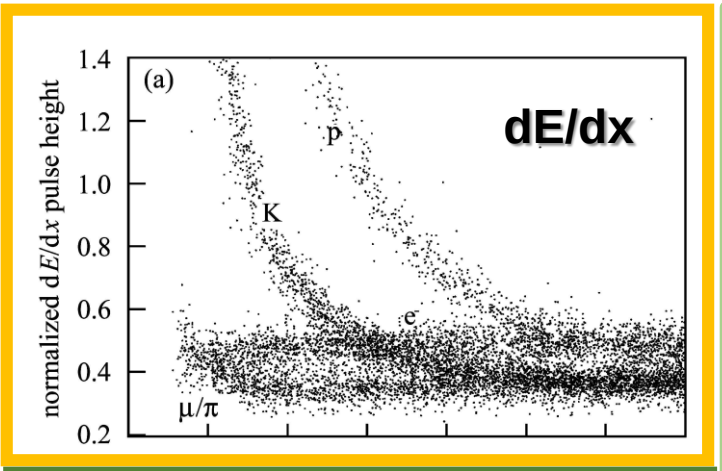
σ_E/E < 5% @1 GeV (end caps)

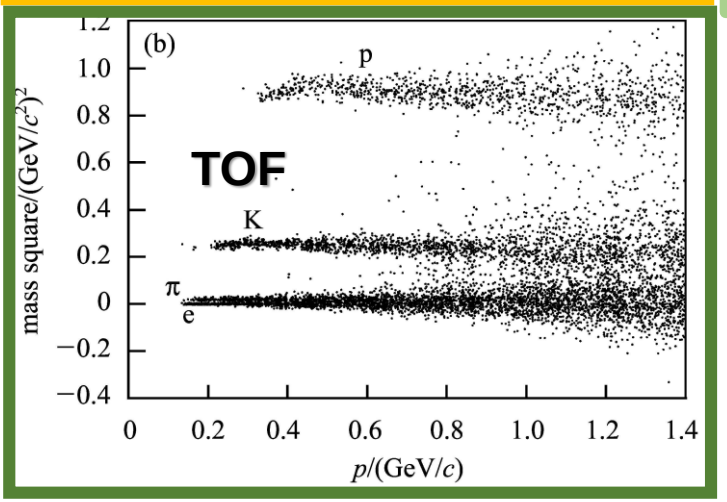
σ_{xy} = (6 mm)/$E^{1/2}$ @1 GeV

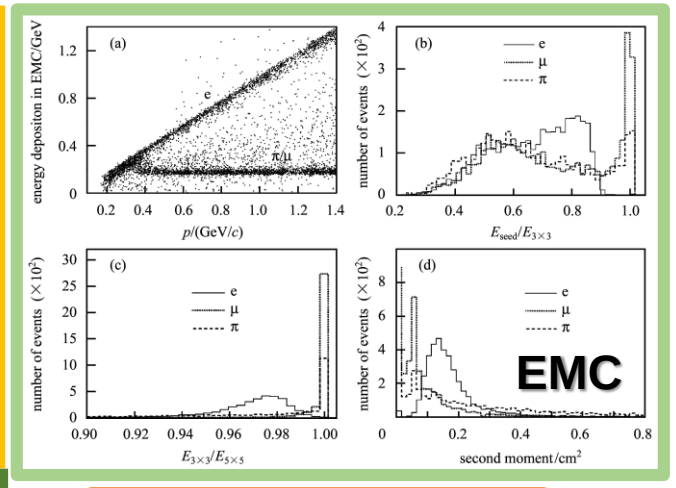
- Drift Chamber
- Time of Flight
- EM Calorimeter
- SC magnet
- RPC Muon chamber

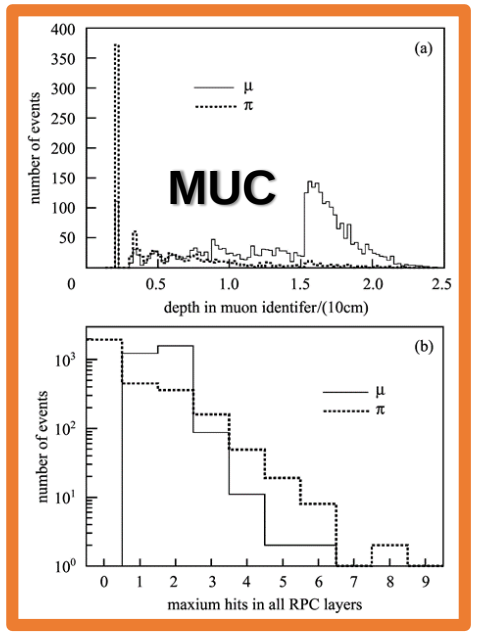


Particle Identification at BESIII









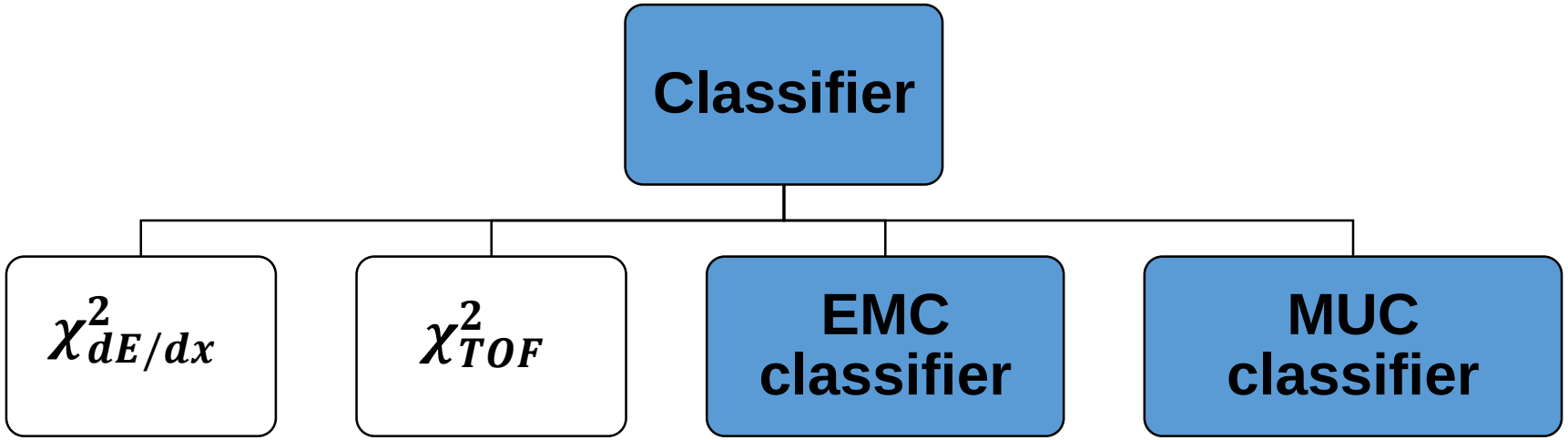
$$Likelihood = PDF_{dE/dx} \cdot PDF_{TOF} \cdot PDF_{EMC} \cdot PDF_{MUC}$$

$PDF_{dE/dx(TOF)}$, calculated from χ^2 for different hypothesis of particle types

$PDF_{EMC(MUC)}$, histogram PDF of a shallow NN output

- Drift Chamber
- Time of Flight
- EM Calorimeter
- SC magnet
- RPC Muon chamber

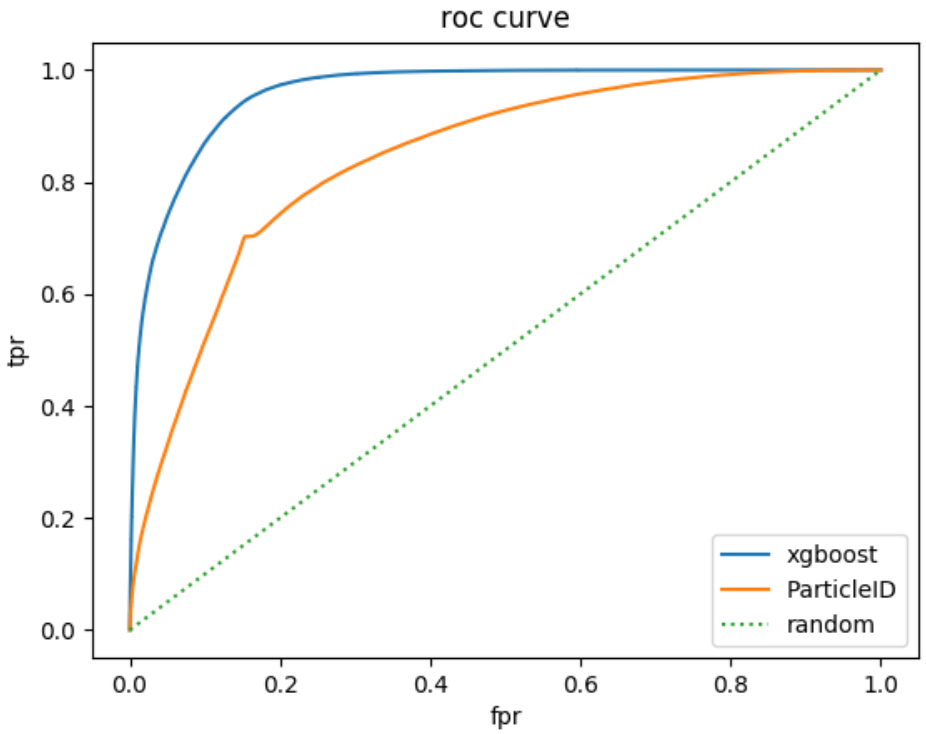
A new approach for muon identification

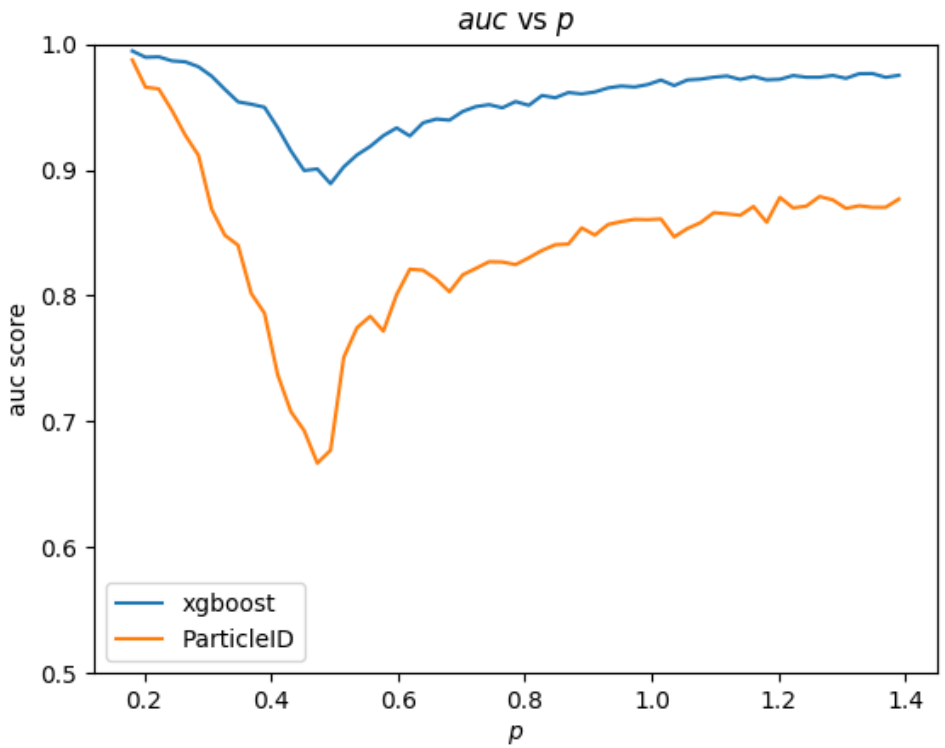


- DNN/CNN/XGBoost : choose to use XGBoost
- All reconstructed sub-detector information are used as inputs for EMC and MUC classifier
- A classifier to combine the PID results of dE/dx, TOF, EMC and MUC

A new approach for muon identification

Data sets: full simulation of μ and π, p∈[0.1 , 1.4] GeV, cos θ ∈[-0.8, 0.8]





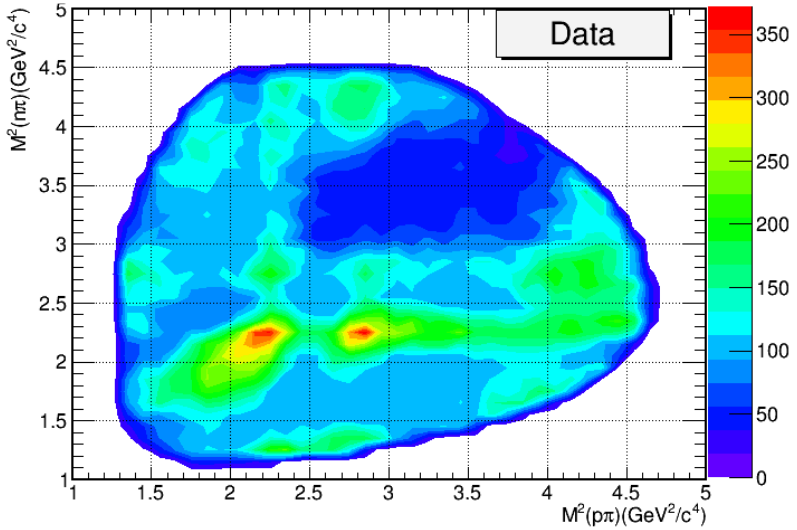
Comparison between the new approach with XGBoost and the default PID seems promising

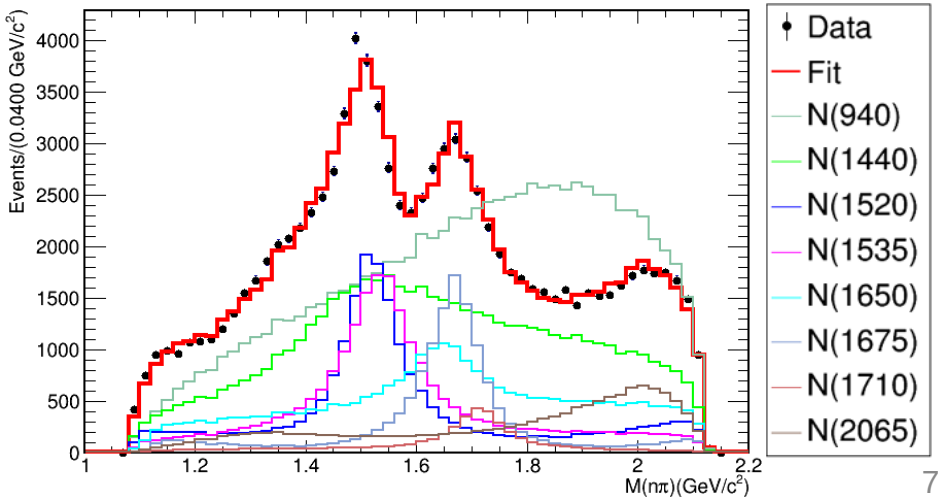
Reweighting: to create data-like MC

- Rich structures in data of hadron spectroscopy cannot be described by generic MC
- Amplitude analysis is mandatory to extract the resonances

- Data-like MC is useful to calculate efficiency or to estimated backgrounds

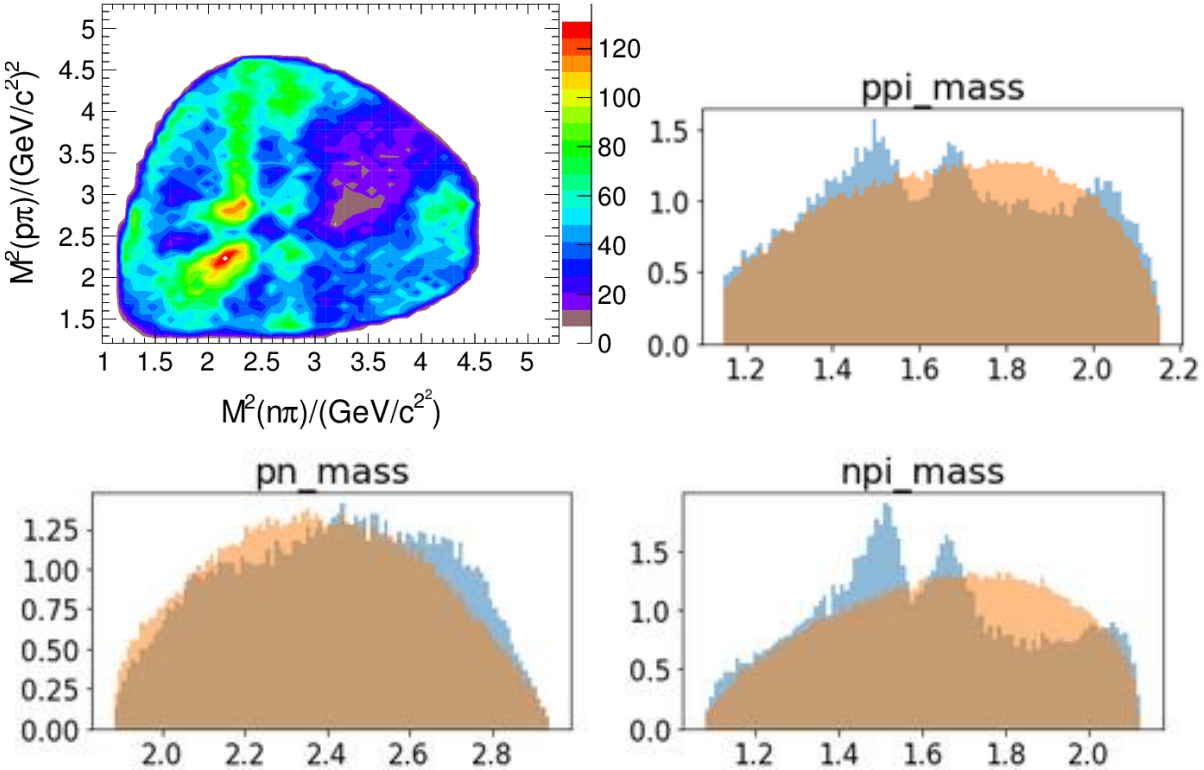
$J/\psi \to \bar{p}n\pi^+$



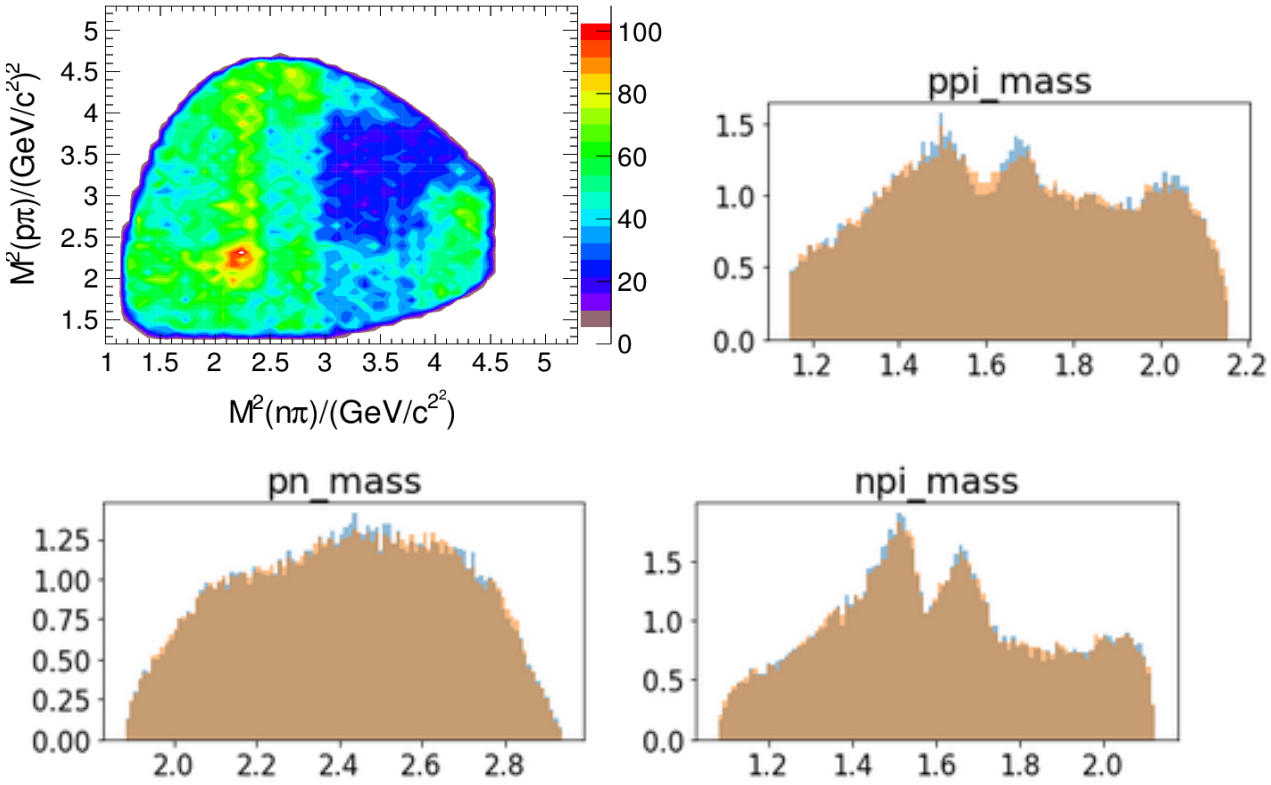


Reweighting with XGBoost

Original distribution



After reweighting



- A classifier to distinguish Data and MC provide probabilities $p_{RD}(x)$ and $p_{MC}(x)$[1]

$$weight\ factor\ w(x) = \frac{f_{RD}(x)}{f_{MC}(x)} \sim \frac{p_{RD}(x)}{p_{MC}(x)}$$

- We utilize this approach with XGBoost algorithm

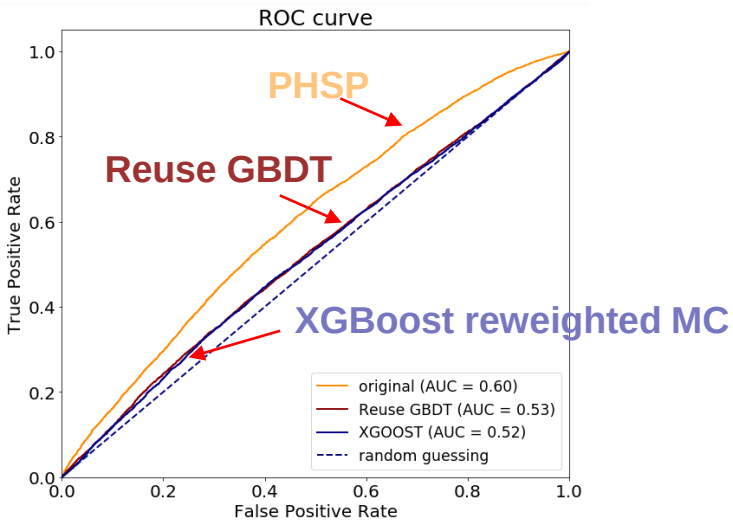
[1] Martschei, D., Feindt, M., Honc, S., & Wagner-Kuhr, J. (2012). Advanced event reweighting using multivariate analysis. *Journal of Physics: Conference Series*, *368*(1), 012028.

Reweighting with XGBoost: Performance

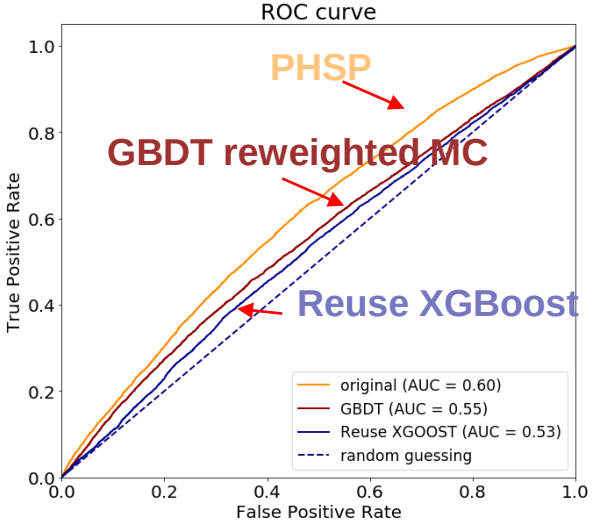
- The discrepancy of PHSP MC and data can be significantly reduced by a reweighter of a XGBoost classifier
- For comparison, the symmetrized χ^2 approach with GBDT [2] is also performed
  - It seems that the results of GBDT can be slightly improved by the our implementation

	Detection Efficiency
‘Data’(DIY MC)	(68.42±0.16)%
PHSP MC	(75.71±0.08)%
Reweighted MC	67.94%

Reuse GBDT

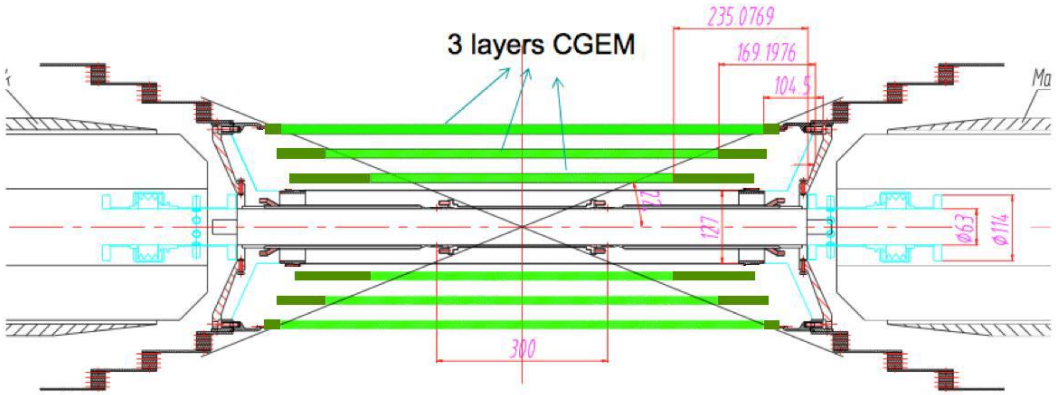


Reuse XGBoost

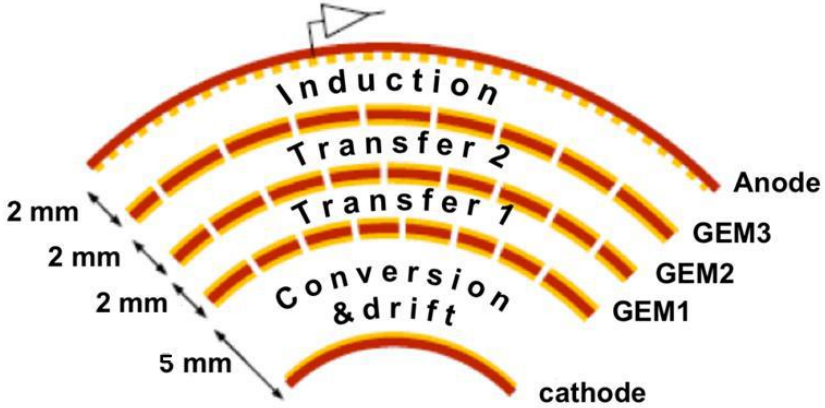


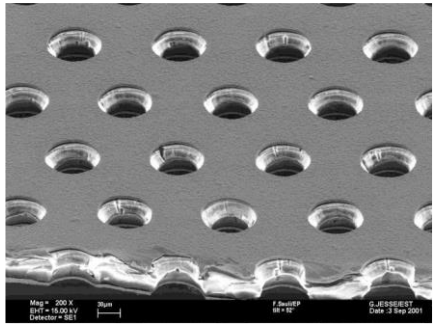
[2]Rogozhnikov, A. (2016). Reweighting with Boosted Decision Trees. *Journal of Physics: Conference Series*, *762*(1), 012036.

BESIII CGEM-IT



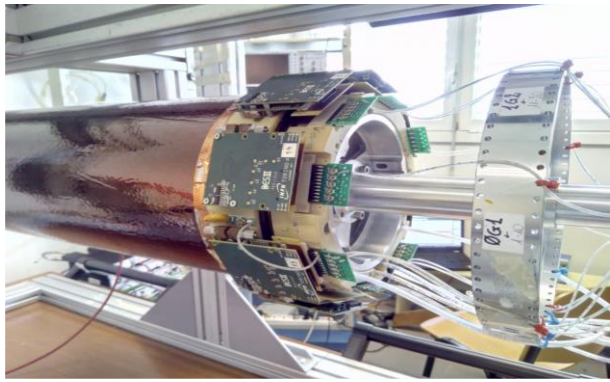
Cylindrical
Gas
Electron
Multiplier
Inner
Tracker

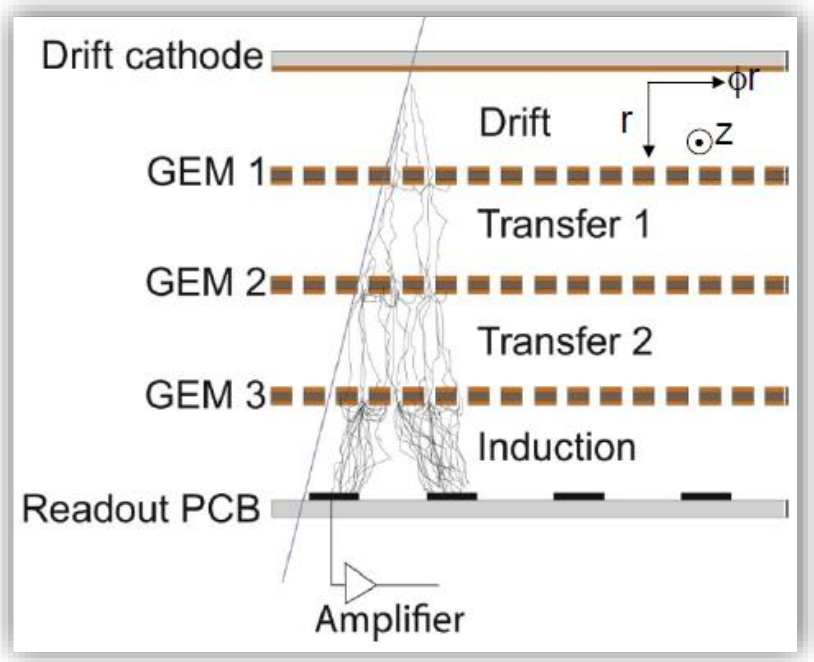


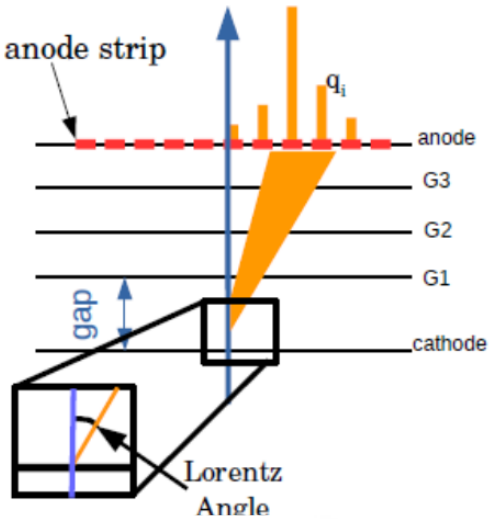


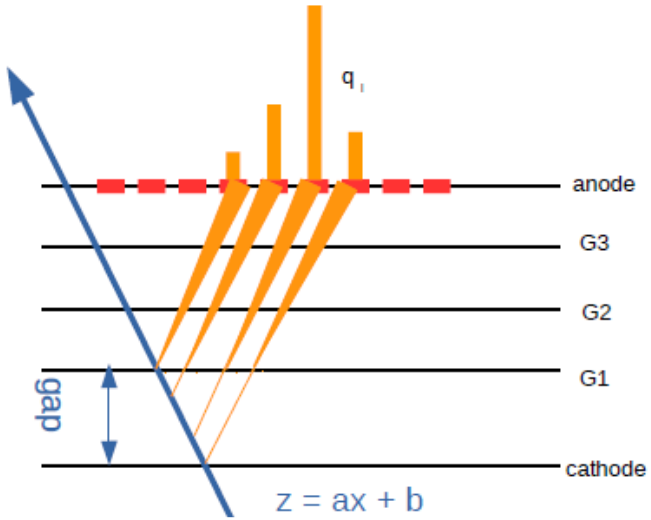
- Inner drift chamber shows aging effects [Ming-Yi Dong et al., Chinese Phys. C40(2016)0160010]
- CGEM is efficient, radiation hard and has high resolution, low aging
- BESIII will upgrade its inner tracker with 3 layers of cylindrical triple-GEMs in 2019

Cluster reconstruction of CGEM-IT









With Q: Charge Centroid

$$< x > = \frac{\sum_i x_i q_i}{\sum_i q_i}$$

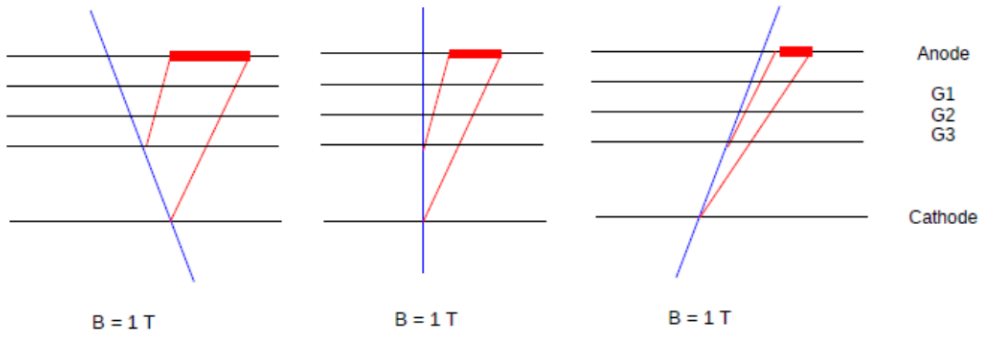
With T: Micro TPC

[Alexopoulos et al, NIM A617 (2010) 161]

$$x = \frac{\frac{gap}{2} - b}{a}$$

Cluster reconstruction of CGEM-IT

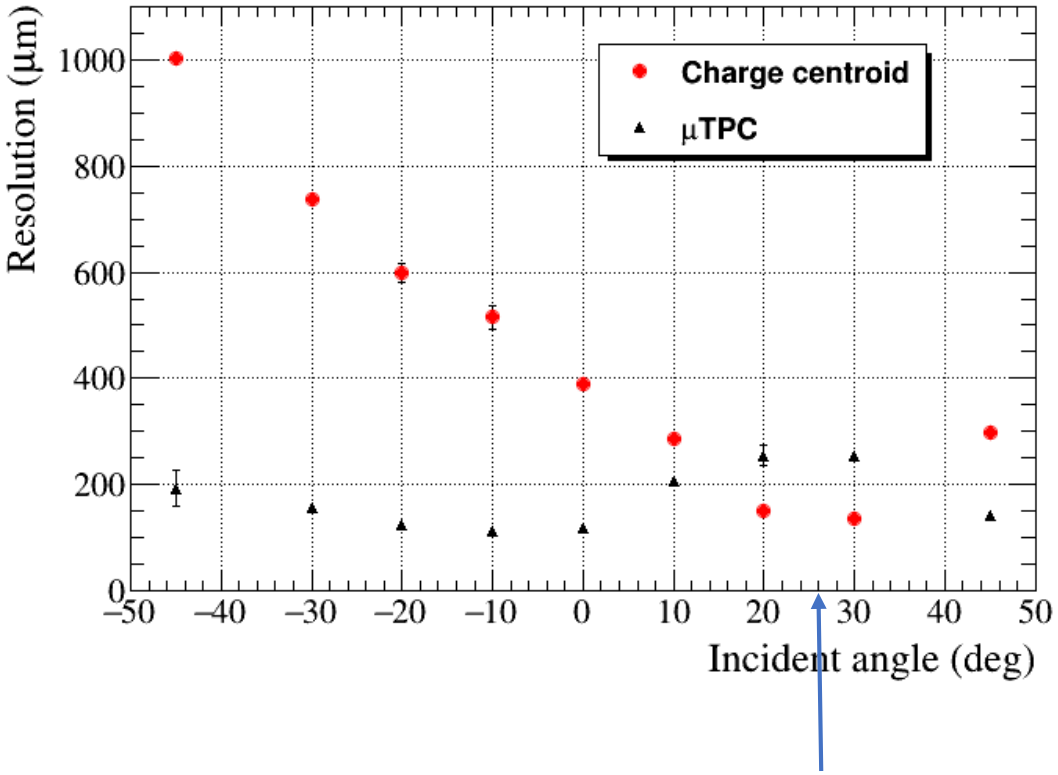
- Correction of Lorentz angle
- Combine the results of charge centroid and micro TPC according to their resolutions
- Dependency of incident angle



defocusing

focusing

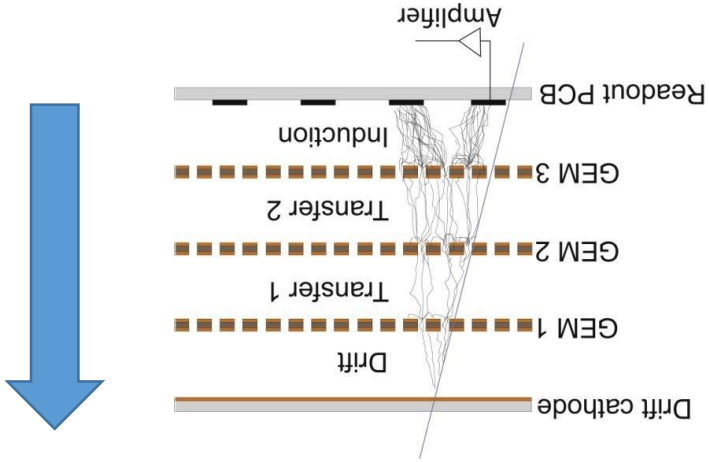
Beam test @ B=1T [arXiv:1807.00500]



Lorentz Angle ≈ 26°

Cluster reconstruction of CGEM-IT with ML

- Use Q and T of the fired strips to measure the initial ionizing particle position X

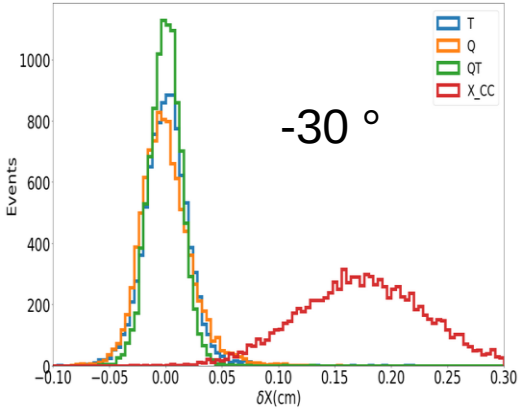


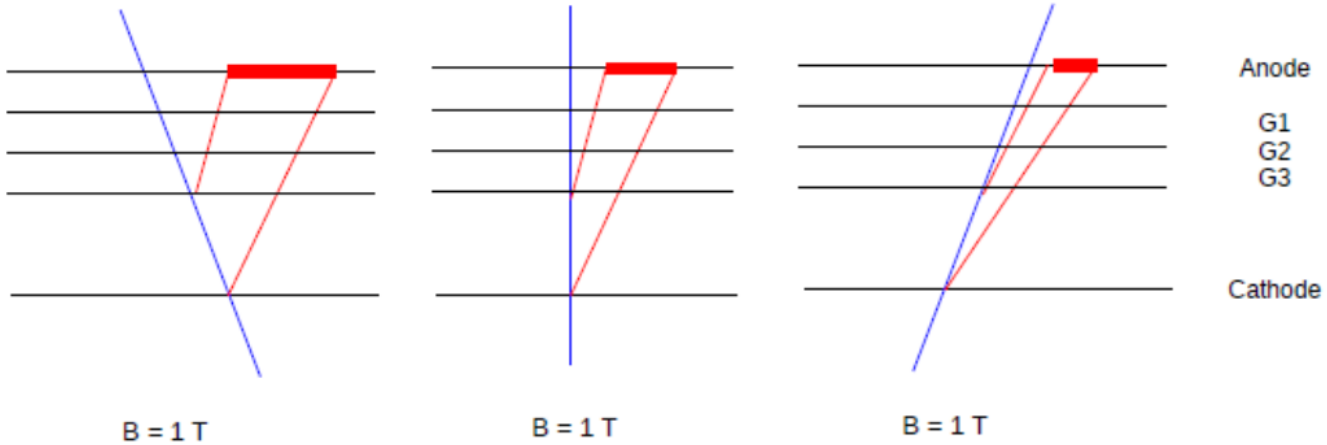
- A regression problem

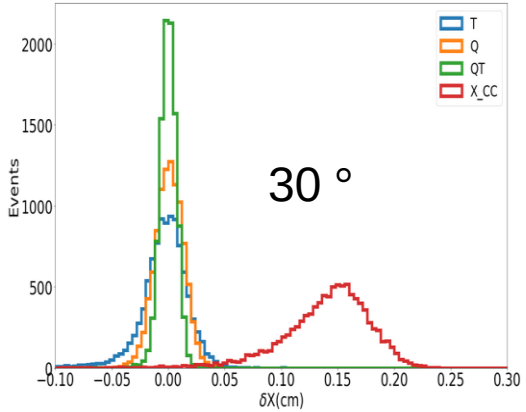
Set up

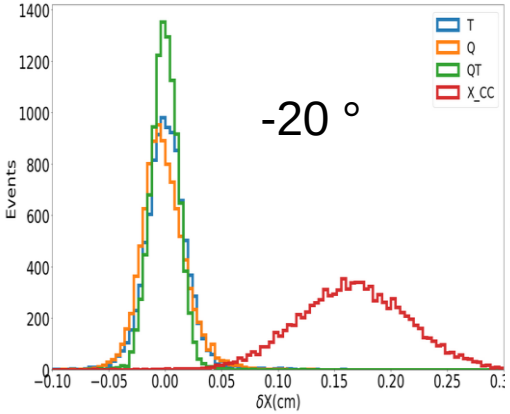
- Simulation with a standalone digitization code, based on GARFIELD results [R.Farinelli, L.Lavezzi, etc. arXiv:1807.01210]
  - B=1T
  - Incident angle [-30 °, 30 °]
  - 1 layer of planar Triple-GEM
- XGBoost regressor

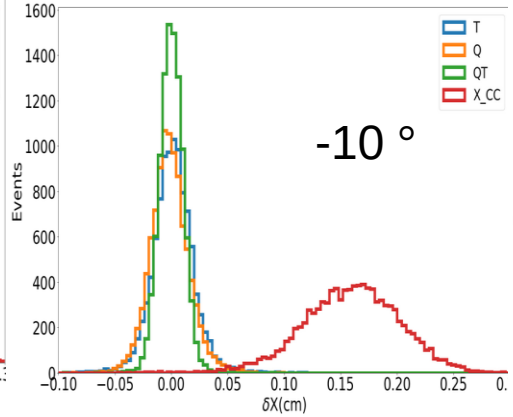
Cluster reconstruction of CGEM-IT with ML

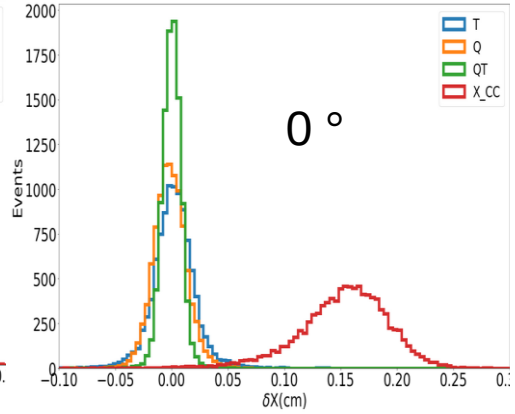


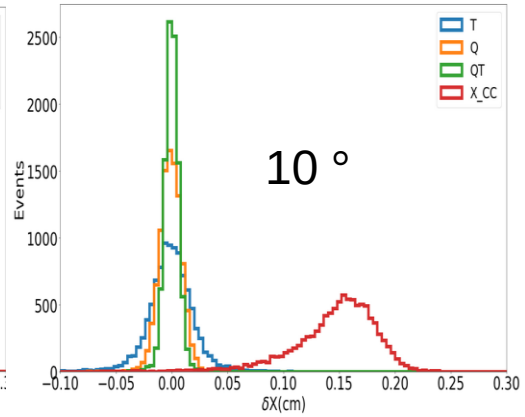


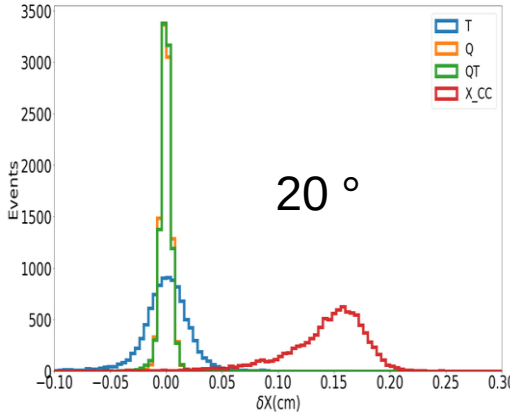




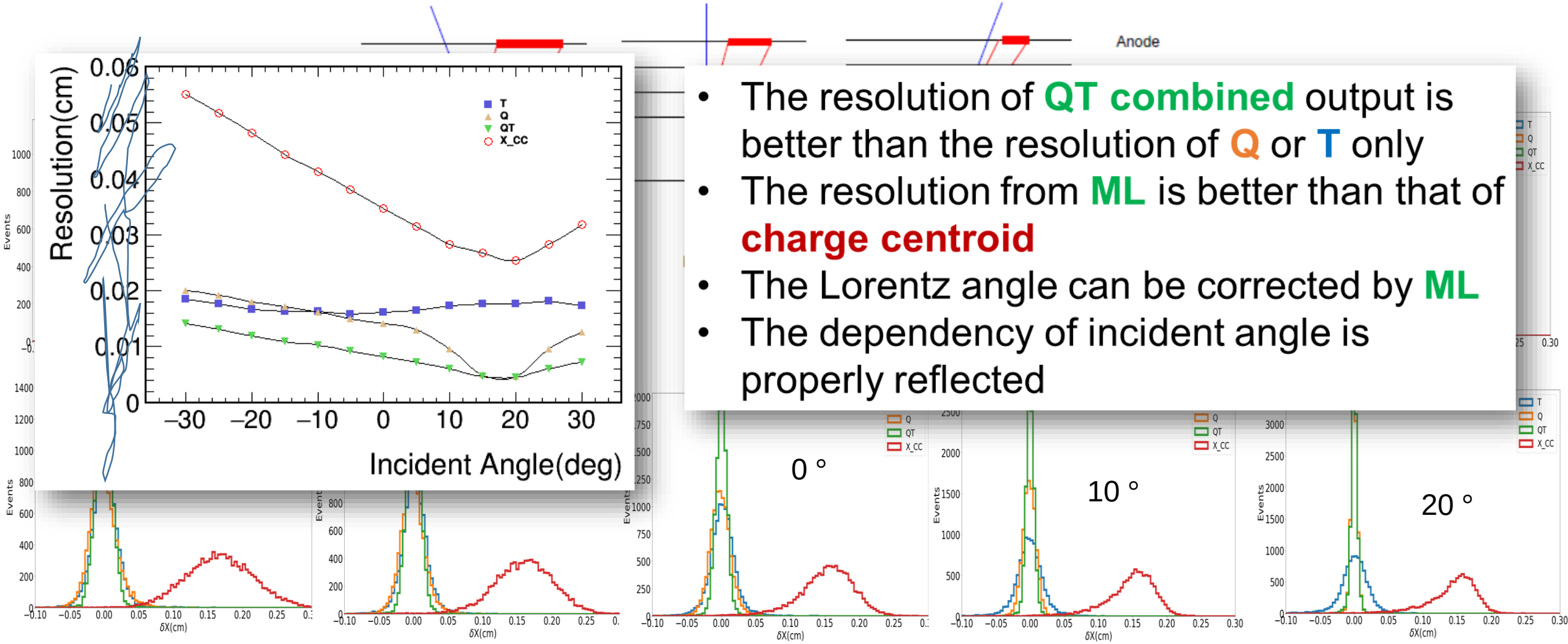








Cluster reconstruction of CGEM-IT with ML



- The resolution of **QT combined** output is better than the resolution of **Q** or **T** only
- The resolution from **ML** is better than that of **charge centroid**
- The Lorentz angle can be corrected by **ML**
- The dependency of incident angle is properly reflected

Work in progress

- e and mu Particle identification
  - Performance measured with control samples of real data (using reweighting to avoid p and cosθ dependency)
  - Estimation of systematic uncertainty (data/MC difference)

- CGEM tracking
  - Cluster reconstruction: position and <u>direction</u>
  - Background suppression in track finding: track segment building with 3 layers

Conclusions

- ML techniques work fine in
  - Reweighting of multi-dimensional distributions
  - Muon particle identification
  - Cluster reconstruction of CGEM inner tracker

- The results look very promising

Thank you for your attention