

COMET Software and Analysis

Yao Zhang, Wu Chen, Zhang Zhaoke, Zheng Weihao

IHEP, China

Nanjing University

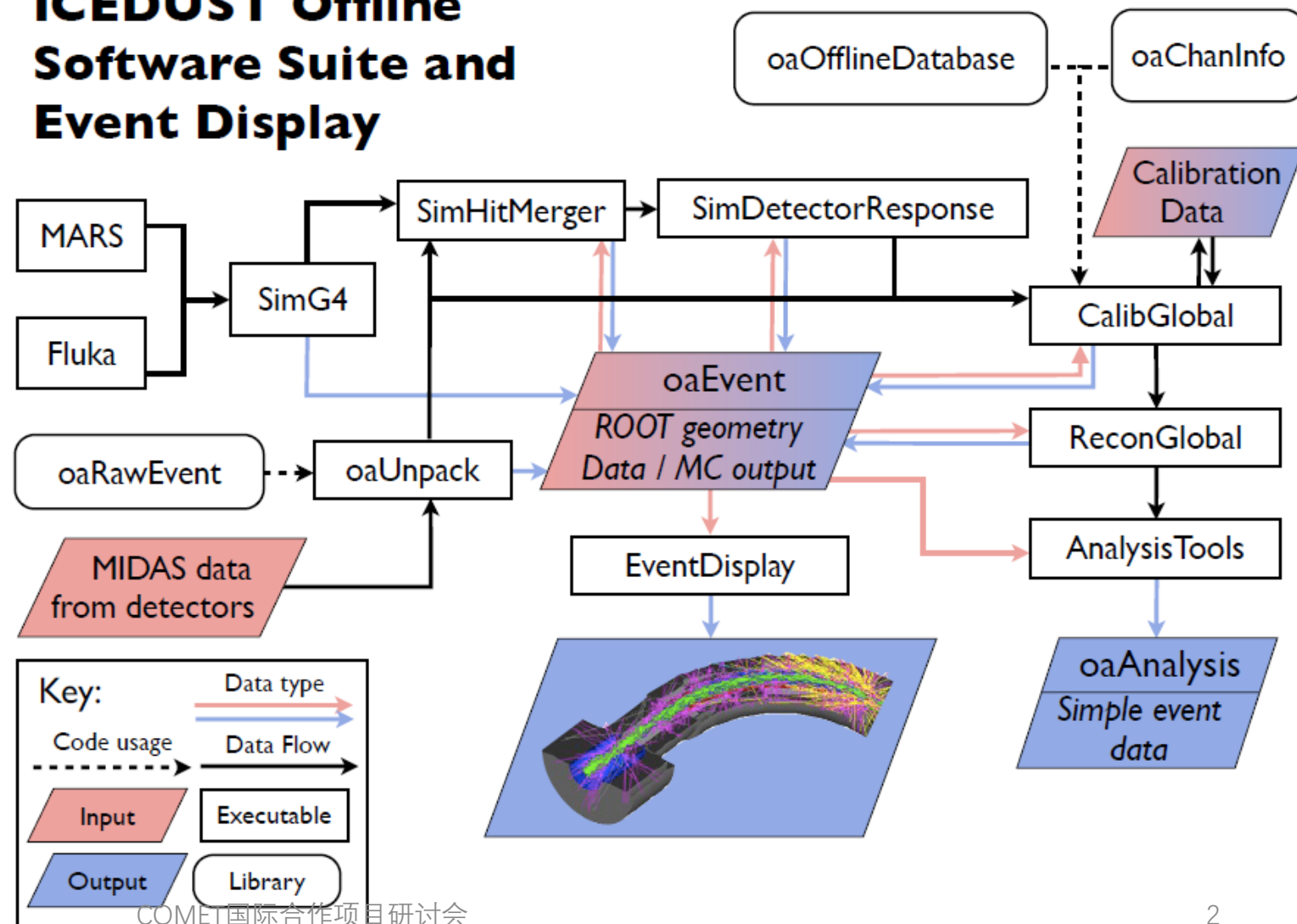
Oct. 10 2026

COMET Offline software framework---ICEDUST

(Integrated Comet Experimental Data User Software Toolkit)

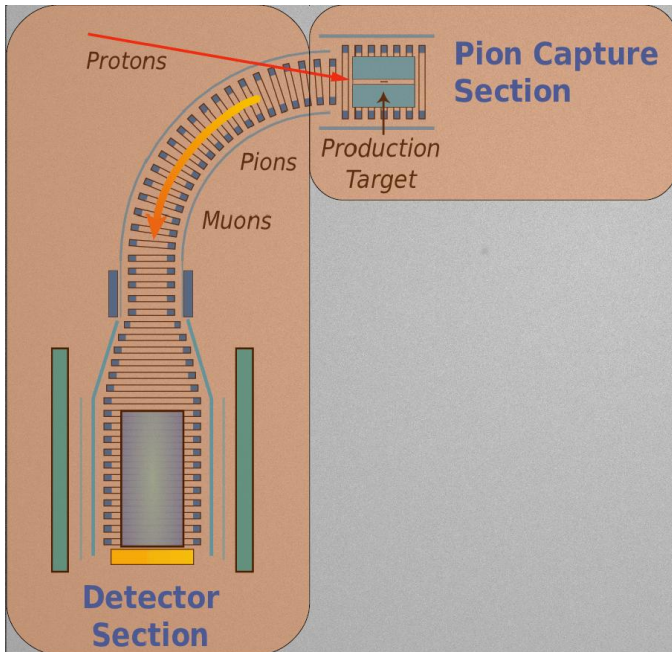
- Support full chain of calibration-reconstruction-analysis
- GEAN4,Fluka,MARS,PHITS and all the hardon model are supported

ICEDUST Offline Software Suite and Event Display

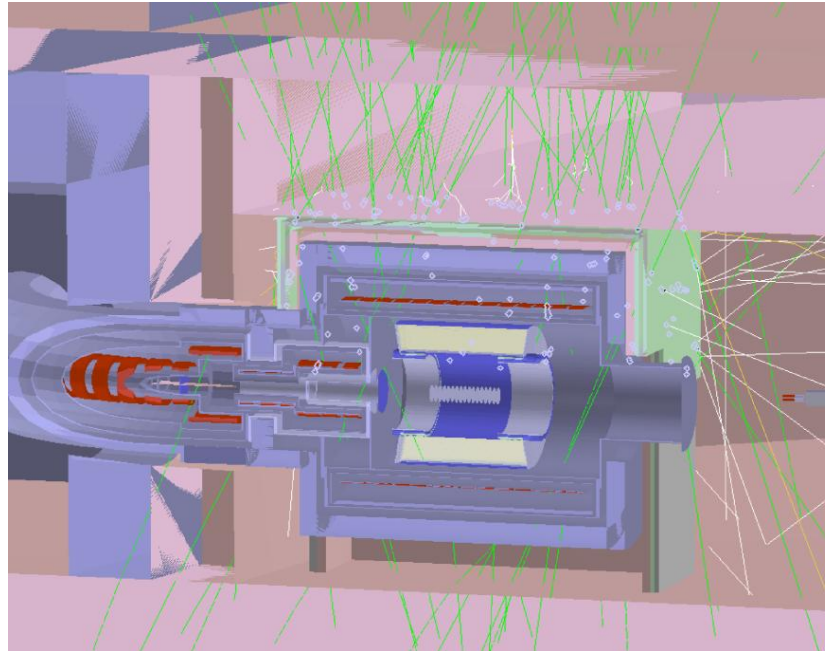


Simulation of COMET Phase-I

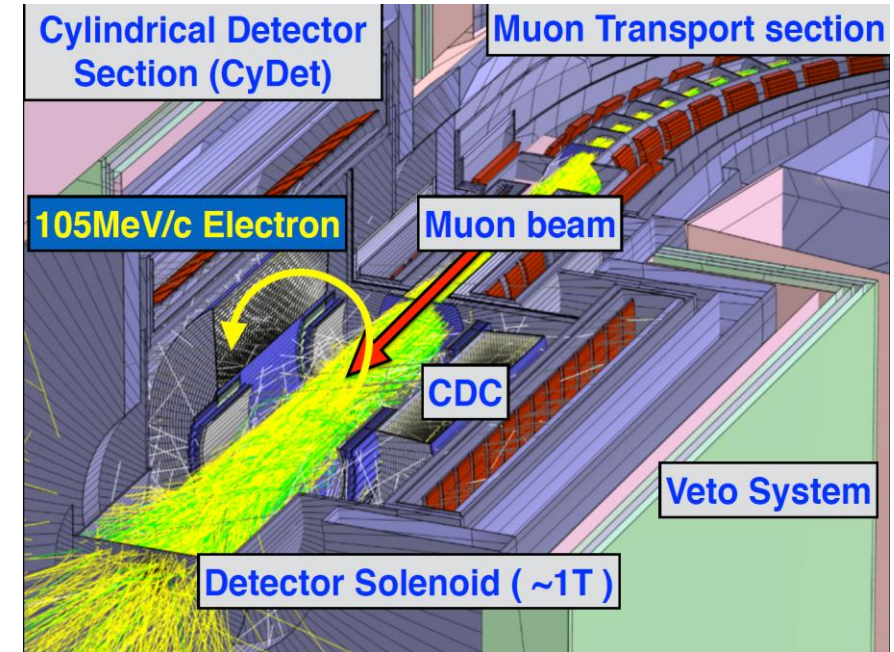
- The simulation of COMET has been sub-divided into smaller tasks



**Beam simulation
(for experiment optimization)**



**Cosmic ray simulation
(for cosmic ray background study)**



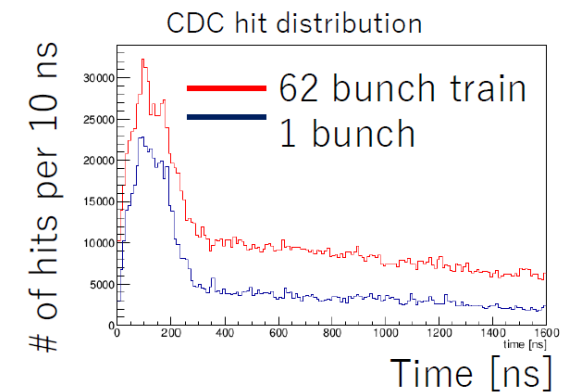
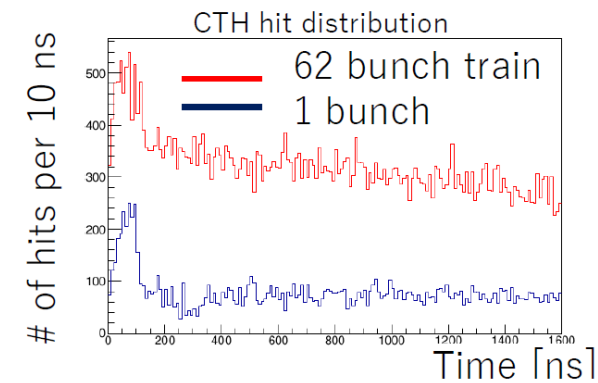
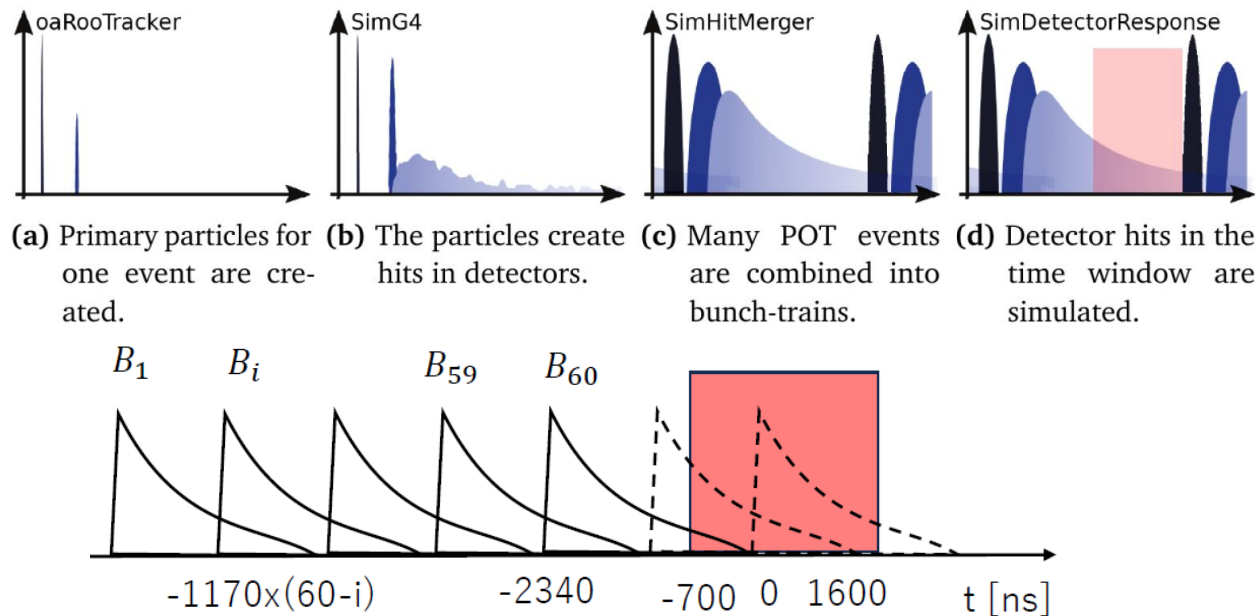
**Signal & background simulation
(for other background study)**

Status of Monte Carlo simulation

- Constantly upgrade and updating
 - ICDUST framework: Geant4 etc.
 - Geometry, magnetic field map, detectors resolution
 - Proton Beam profile etc.
 - Validation and comparison on software and different physics models
- Mass production about every two years
 - 10^7 - 10^9 PoT each MC
 - The 7th mass production(MC7) done

Bunch train merging

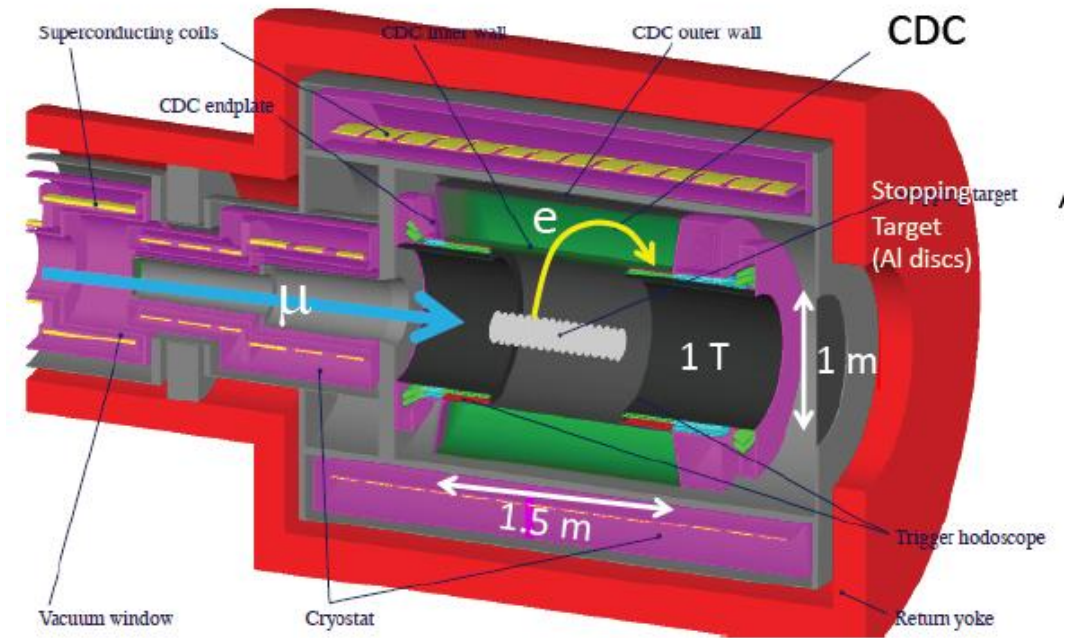
- Bunch train simulation is a unique challenge for COMET
 - Slow processes (neutron propagation/radioactive decay/etc.) from previous bunches are included.
- Bunch train events generated
 - Infinite bunches in front of each signal window is assumed



COMET Phase-I tracking system

---The Cylindrical Detector system (CyDet)

- Cylindrical trigger hodoscope (CTH):
 - Two layers of plastic scintillators
 - MPPC readout
 - 4-fold coincidence trigger
- Cylindrical drift chamber (CDC):
 - 18 sinse wire stereo layers
 - Helium based gas
 - Large inner bore
 - Momentum resolution: 200 keV/c (for $p=105$ MeV/c)
- Stopping target
 - Aluminum target with 17 disks
 - 100 mm radius, 0.2 mm thickness, 50 mm spacing.

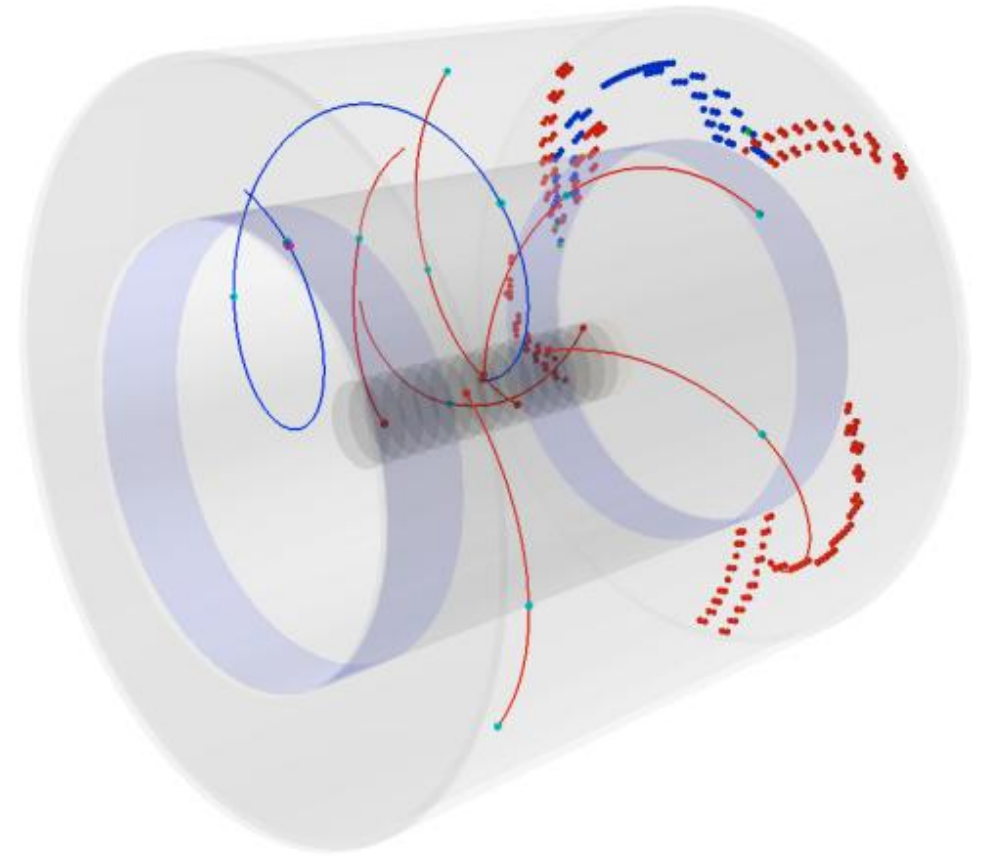


CDC tracking challenges

- No z measurement directly
- No vertex constraint
- No track seed
- All curling tracks which 1/3 are multiple turns
- Overlapping hits
- High hit occupancy

Requirements for tracking

- Tracking under high hit occupancy
- Estimate track parameters from enormous track segments
- Distinguish hits and tracks from different turns
- Suppress high momentum tail of reconstructed tracks



A Web-Based Event Display for COMET

By Weihao Zheng , Nanjing University

Interactive 3D inspection of CyDet reconstruction and beam data for data checks and algorithm debugging



CyDet geometry, hits and tracks in the browser

Status: Deployed at IHEP. Next: improve deployment and caching, and complete access for external collaborators.

Implementation and features

Phoenix 3D framework

Integrates Phase-I/II and CyDet geometries with visibility and transparency controls.

Event inspection

View hits, reconstructed tracks and beam trajectories; switch events, inspect metadata and filter beam particles by type, momentum and arrival time.

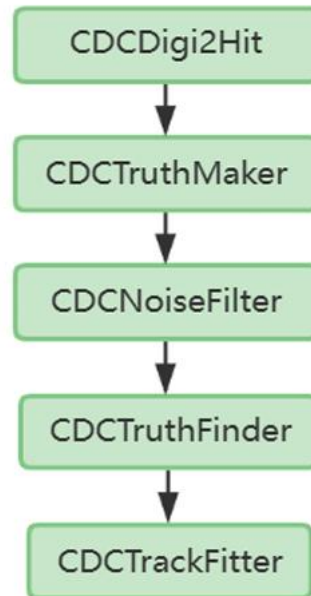
ROOT / JSON input

Load JSON directly or convert ROOT data through a Python / PyROOT backend.

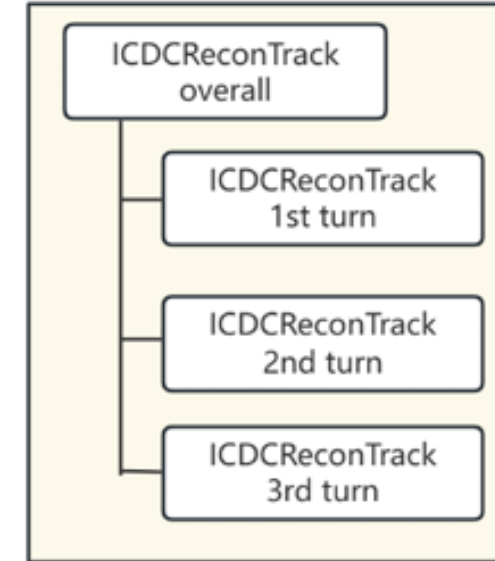
Tracking algorithms for CyDet

- De-nosing
 - GBDT
 - NN: FPN
- Track finding
 - RANSAC
 - Hough transform
 - GPU: Apollonius finding
- Track fitting
 - Genfit2
- Track quality selection
 - Machine learning

CDC Tracking Flow

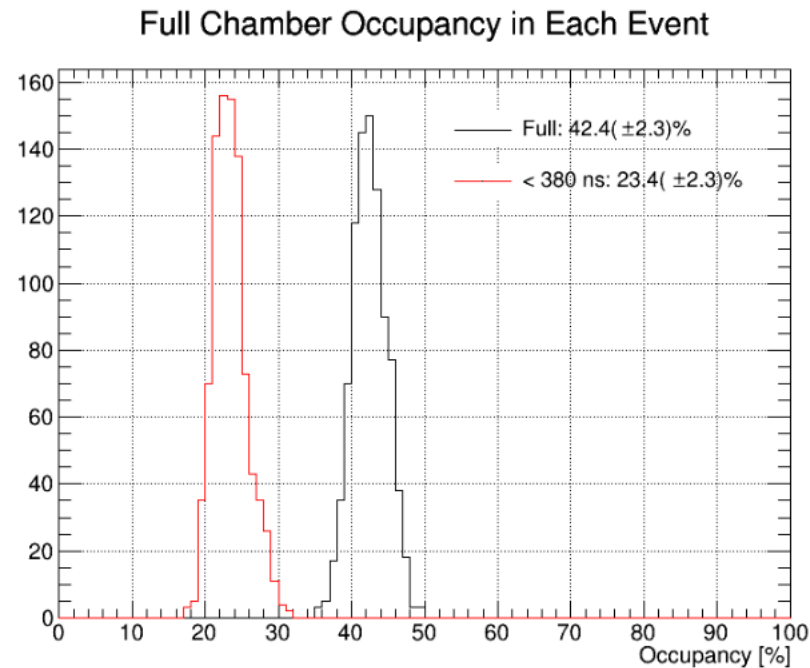


CDC Rec. Track model

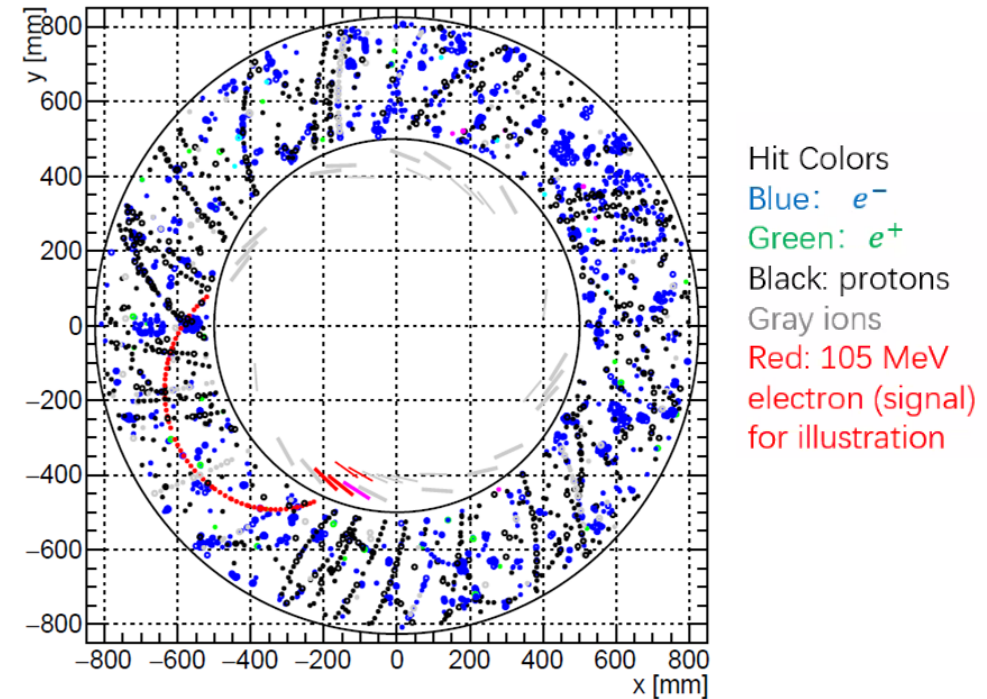


High occupancy in the tracking detector

40% occupancy @ 1.6×10^7 POT/bunch

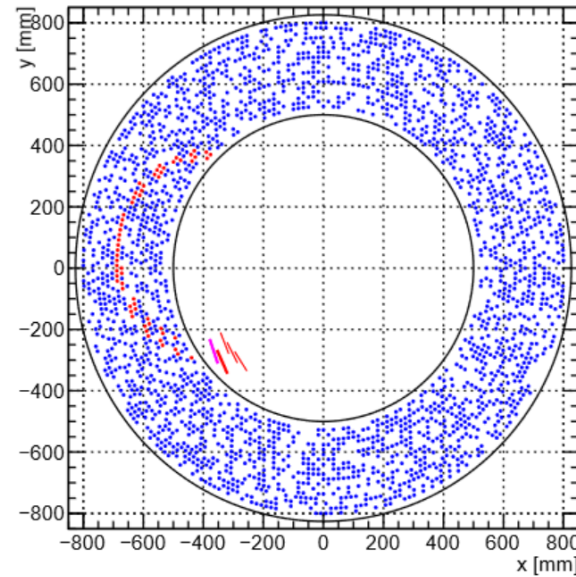
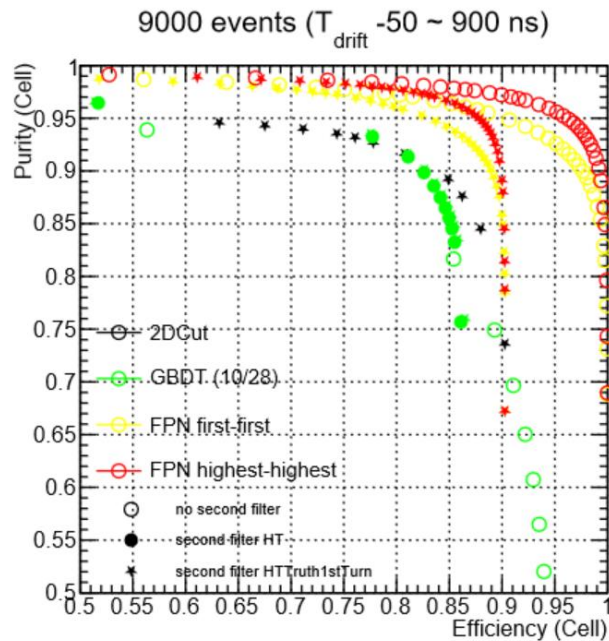


A typical event display with 40% occ.

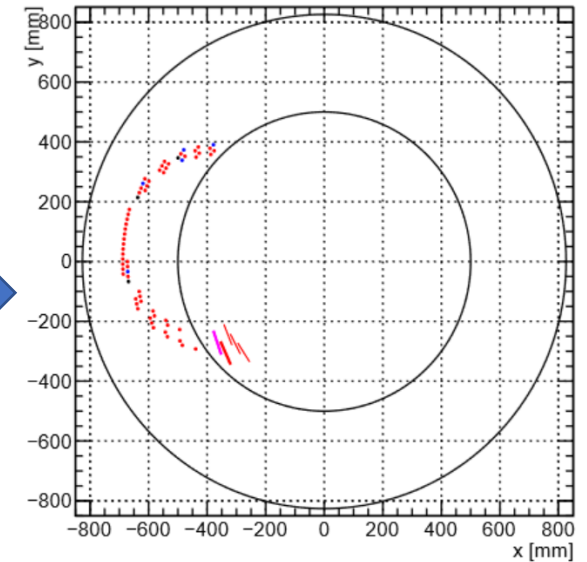


De-noising performance

- 2D Cut, GBDT, FPN have been tested
- FPN gives best efficiency and purity performance



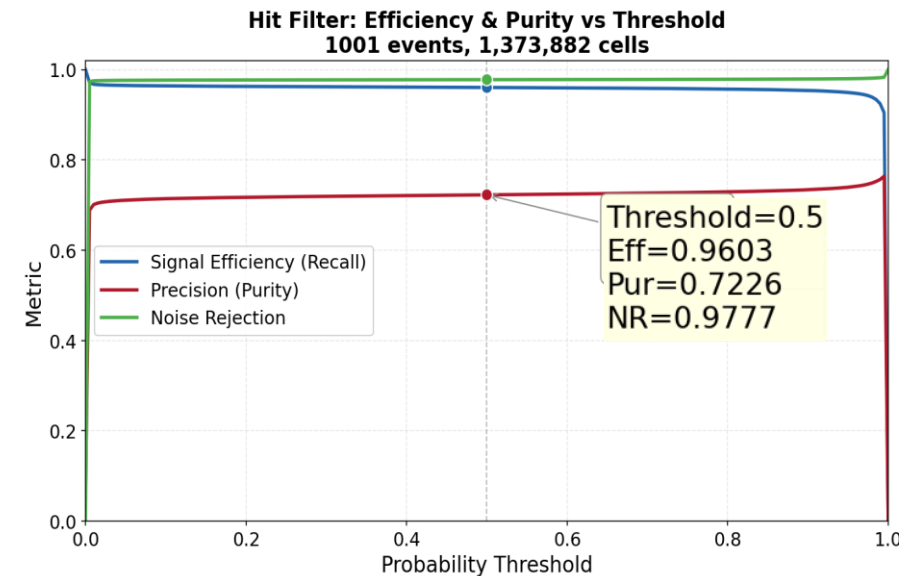
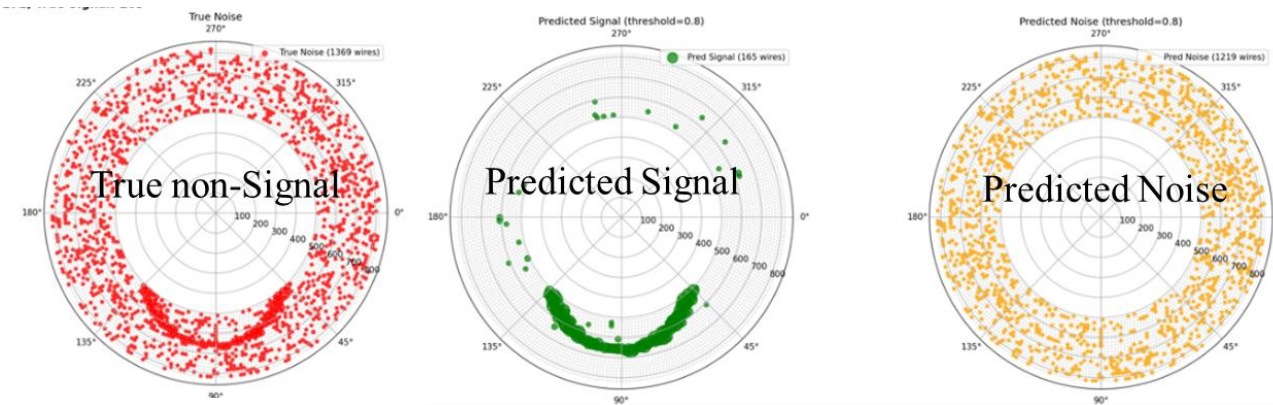
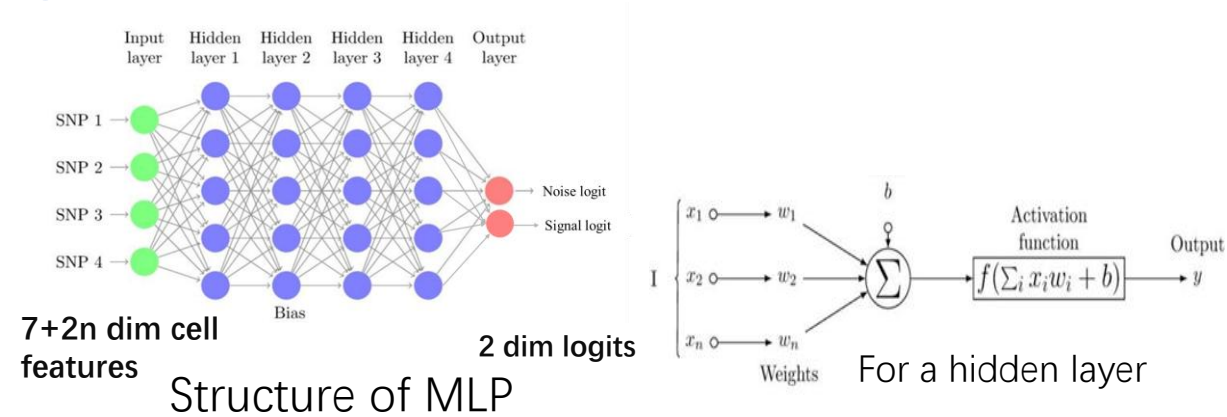
FPN



Hit Filter : to reduce the input for GNN

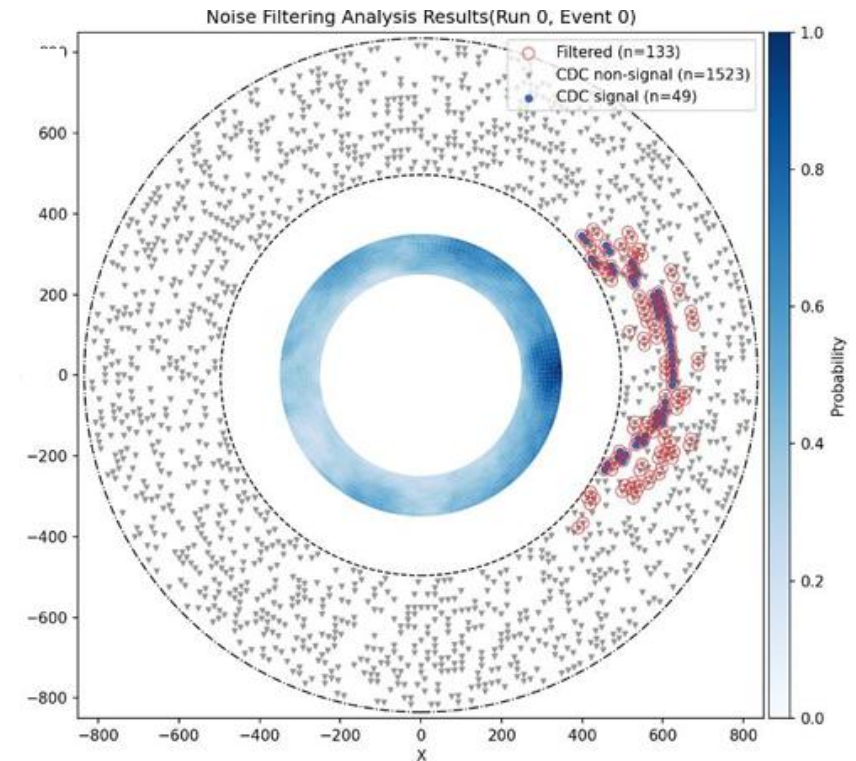
(by Weihao Zheng Nanjing University)

- Aim: reduces noise hits for GNN by $> 95\%$
- Suppresses background hits while retaining signal
- **Model: 4-hidden-layer MLP (Multi-Layer Perceptron)**
- **Output: signal / noise probability**

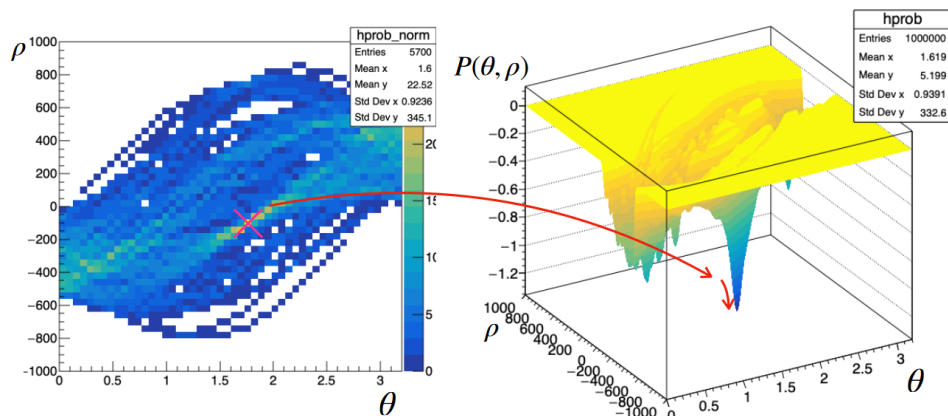


Traditional de-noise algorithm (By Zhaoke Zhang IHEP)

- Occupancy of sample : random background
- Goal: reject background before track reconstruction
- Algorithm:
 - Noise hit → no hit in left/right adjacent cells (same layer)
 - Reverse Hough transform → evaluate hit track probability
 - Circle-intersection method → common intersections reveal hits lying on track circle



CDC Track Finding



Apollonius's problem : construct circles that are tangent to three given circles in a plane
 Greek Apollonius of Perga (200 BC)
 Compass constructions by French François Viète (1600)
 Algebraic solution (cyclographic model) :

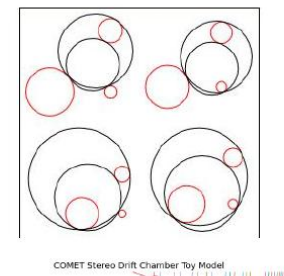
$$(-X_1 + X_s)^2 + (-Y_1 + Y_s)^2 = (r_s - r_1 s_1)^2$$

$$(-X_2 + X_s)^2 + (-Y_2 + Y_s)^2 = (r_s - r_2 s_2)^2$$

$$(-X_3 + X_s)^2 + (-Y_3 + Y_s)^2 = (r_s - r_3 s_3)^2$$

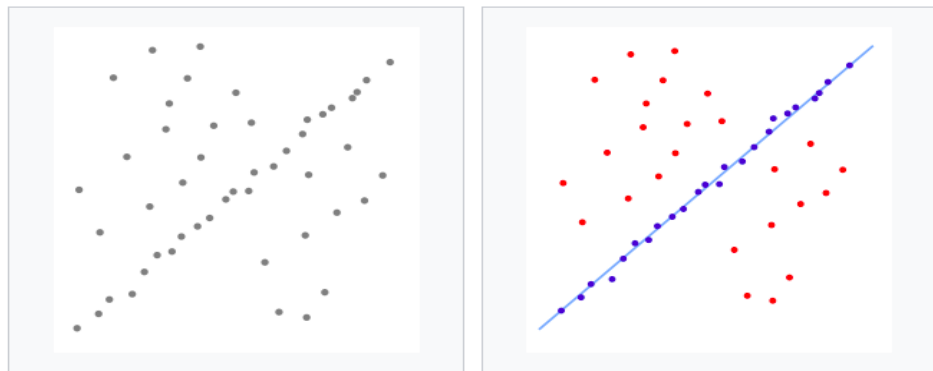
with $s_{1,2,3} = \pm 1$

toy model: 46 hits of electron signal (red)
 use $d_i^{\text{St.}}$ true signed stereo drift distance

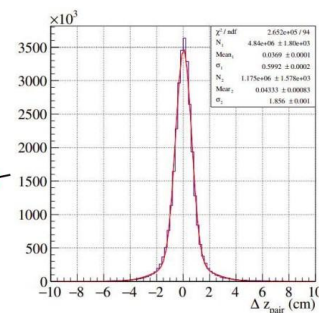
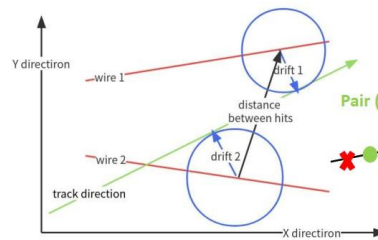


Track finding based on Hough Transform

The Apollonius Problem



- Calculate Z of pairs using track direction and other parameters
- Calculate Z of hits by linear fitting
- Performance:
 - Resolution of pair z: 9.8 mm
 - Resolution of hit z: 10.1 mm



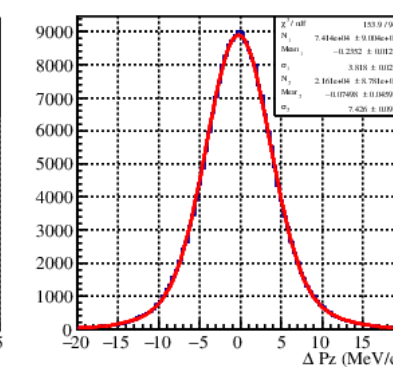
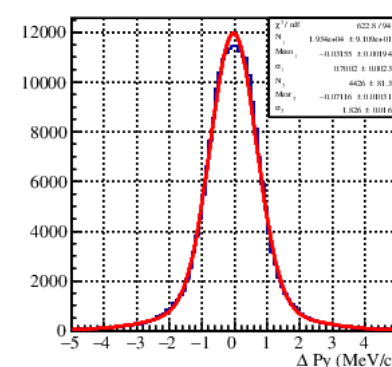
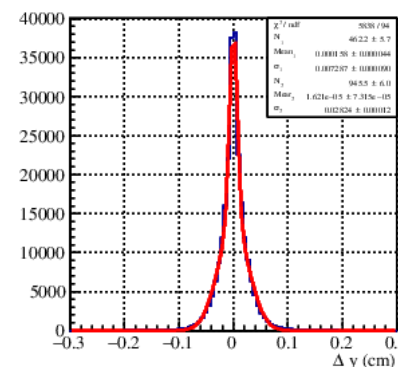
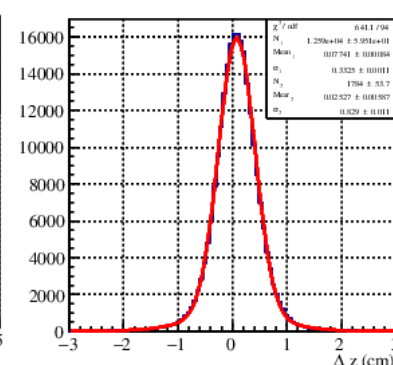
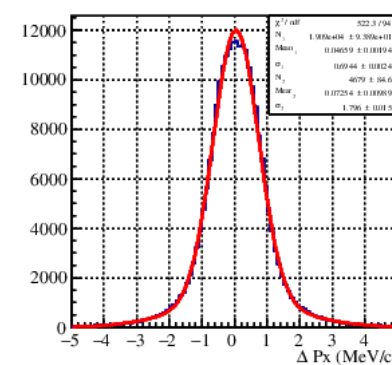
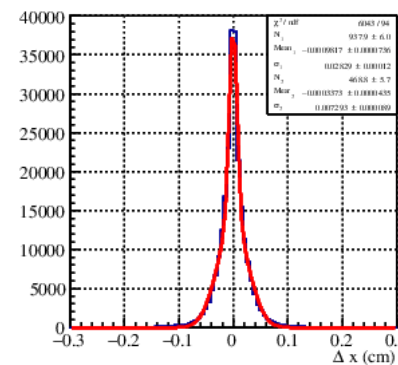
Random Sample Consensus (RANSAC)

Performance of Ransac Finding for Single Turn Event

- Track segments efficiency: 96.2%
- Peak finding efficiency: 99.99%
- Track finding efficiency: 96.2%
- Efficiency after track fitting: 95.6%
- Resolution:

$$\sigma_x = 0.24 \text{ mm}, \sigma_y = 0.24 \text{ mm}, \sigma_z = 4.3 \text{ mm},$$

$$\sigma_{px} = 1.0 \text{ MeV/c}, \sigma_{py} = 1.0 \text{ MeV/c}, \sigma_{pz} = 4.9 \text{ MeV/c}$$

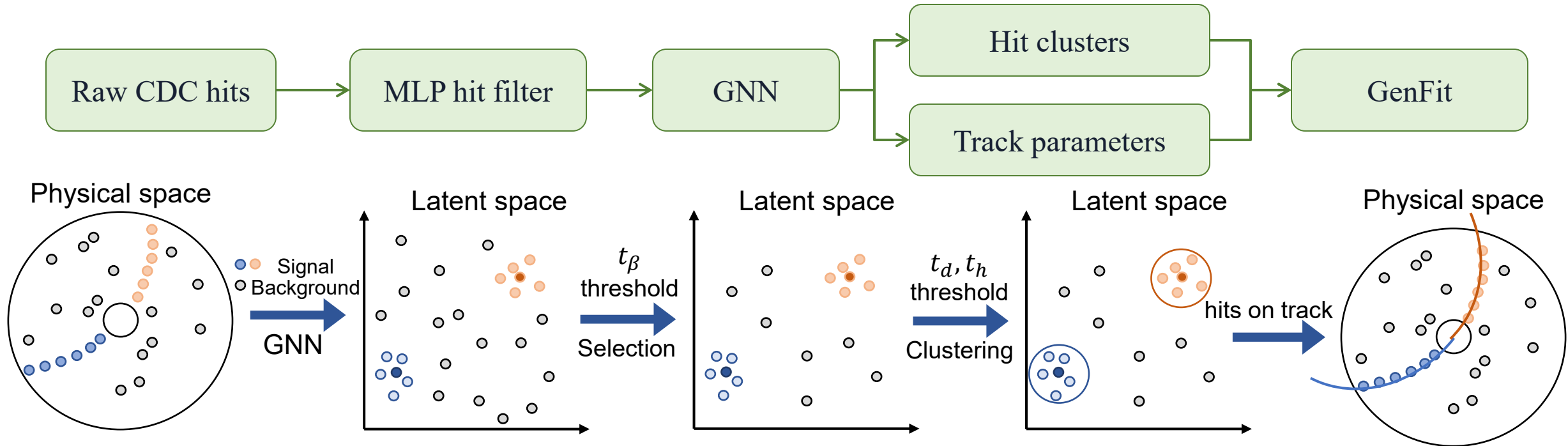


	Geometrical Acceptance	Tracking		Totally	Tail	Momentum resolution (core/tail)
		finding	fitting			
		96.2%	99.4%			
Single Turn	14.0%	95.6%		12.7%	1.6%	214keV/533keV

GNN-based end-to-end Machine Learning tracking

By Weihao Zheng Nanjing University

- Learns extract hits association and track parameter directly from data **instead of relying on hand-crafted algorithm and manual parameter tuning**



GNN-based end-to-end Machine Learning tracking

By Weihao Zheng @ Nanjing University

Reconstruction approach

MLP Reject background hits and reduce the graph size.

GNN Associate hits with GravNet and Object Condensation and predict momentum, position and hit-level quantities.

GenFit Fit GNN candidates from predicted initial states and apply quality selections.

Key results

MLP

Signal hit recall $\sim 96\%$

Noise hit reduction $\sim 97\%$

GNN

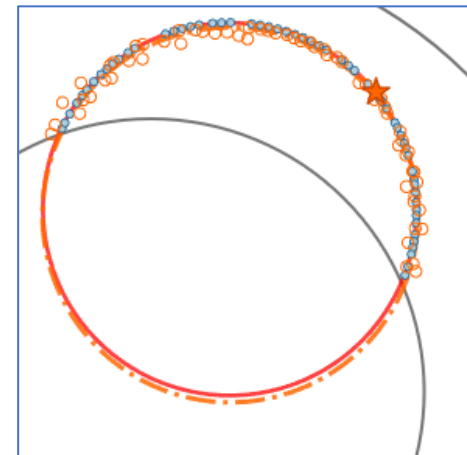
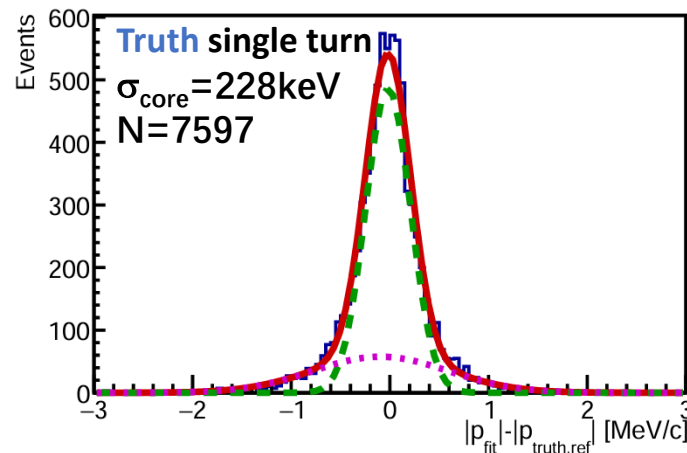
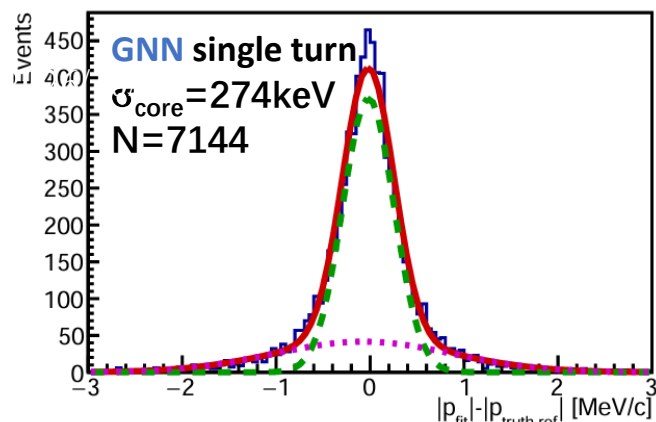
Single-/multi-turn hit efficiency: $\sim 96\%/95\%$

Hit purity $\sim 89\%$

GNN + GenFit

Final acceptance $\sim 68\%$

Single-turn $|p|$ residual: core $\sigma \approx 0.274$ keV/c



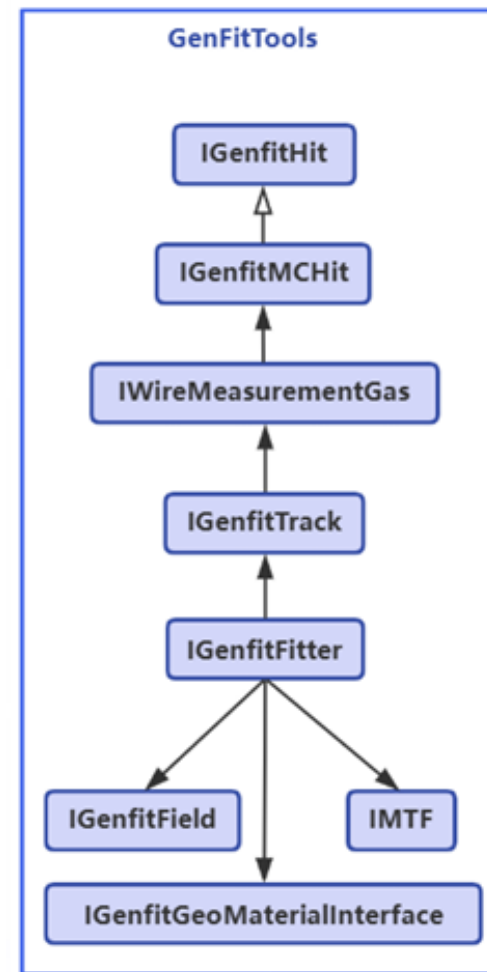
CDC Track Fitting

- GenFitTool

- An experiment-independent **generic track fitting** framework
- Open sourced, active development and large user community
- Official track fitting for BelleII, also used by PANDA, CEPC, BESIII etc.

- ICDCTrackFitter

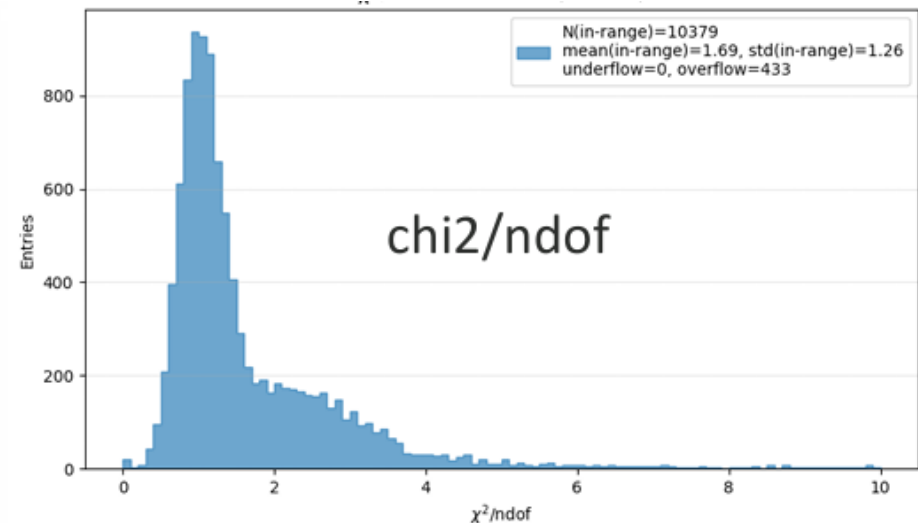
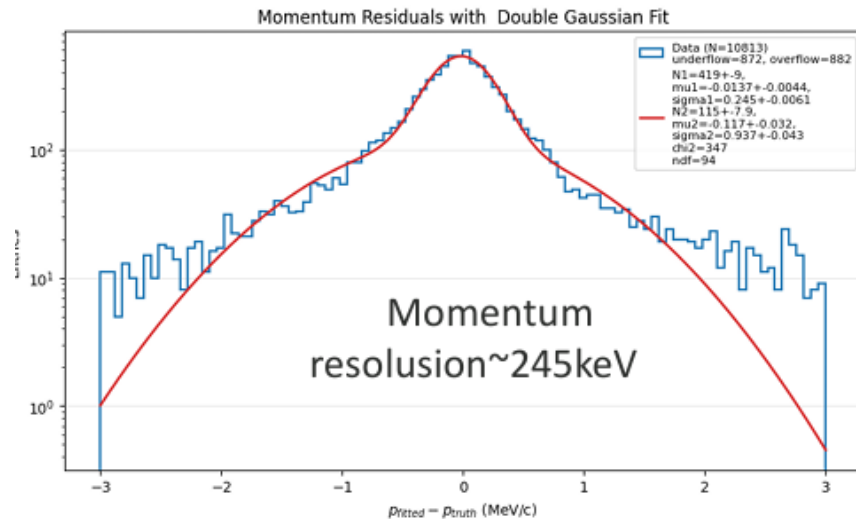
- Wrapper for GenFit-based CDC track fitting
- Input: ICDCTruthTrack and hits
- Output: ICDCReconTrack and hits
- An example usage of this tool will be provided soo



CDC Track Fitting

Sample: Pure signal single turn events

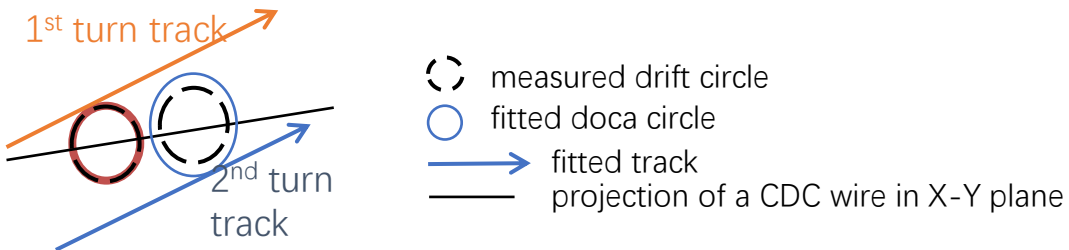
- Momentum resolution & fit quality → as expected



CDC Multi-turn fitting By Zhaoke Zhang, IHEP

- A multi-turn fitting algorithm developed for COMET CDC
 - Parallel fitting, hit competition, annealing fitting
- Significantly improves the fitting quality and reduce tail of multi-turn track momentum

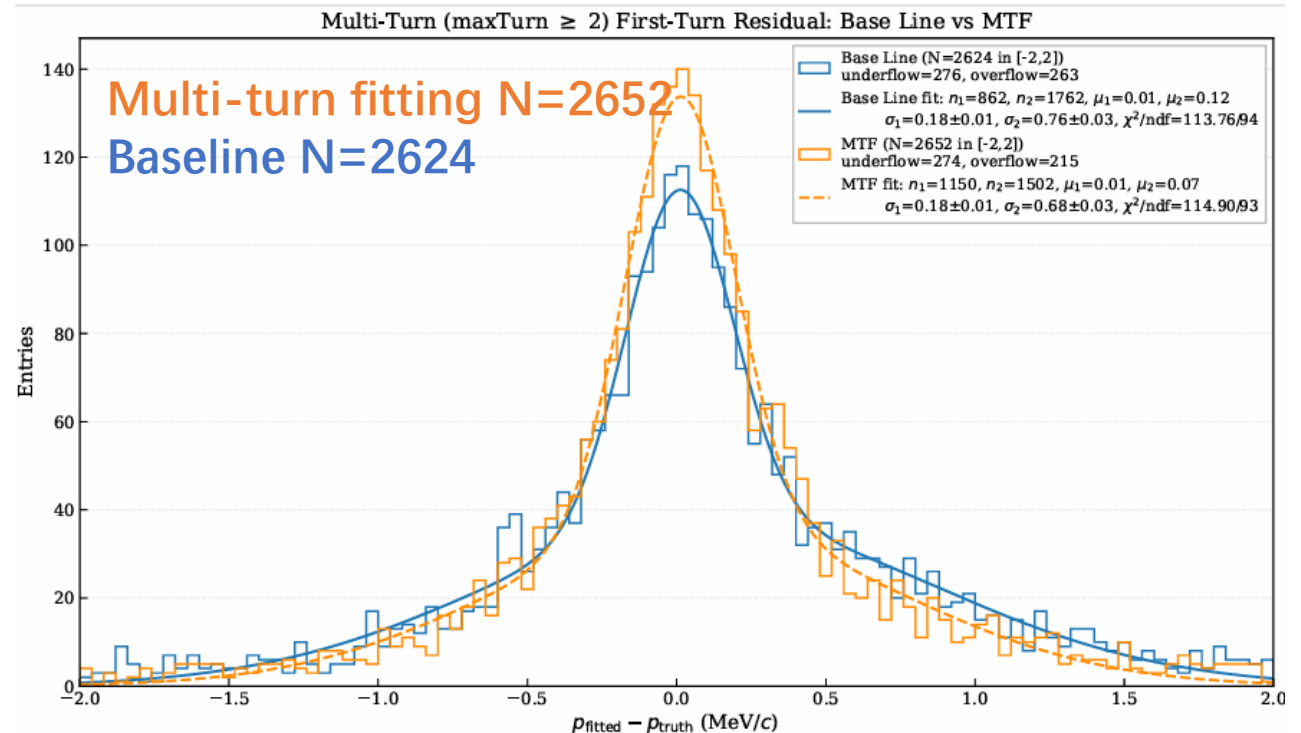
one hit associated with two tracks



Weighted mean of measurement and it's covariance matrices in Kalman filter :

$$(\tilde{\mathbf{V}}_k)^{-1} \equiv \tilde{\mathbf{G}}_k = \sum_i p_k^i \mathbf{G}_k^i \quad \tilde{\mathbf{m}}_k = \tilde{\mathbf{V}}_k \cdot \left(\sum_i p_k^i \mathbf{G}_k^i \cdot \vec{m}_k^i \right)$$

p : assignment probability



Track quality cut by Machine Learning

- Gradient Boosted Decision Trees (GBDT) is a ML method for classification
- Tail events with bad quality are suppressed using GBDT

- Parameters used by GBDT

- Number of hit
- NDF
- χ^2
- χ^2/NDF
- Maximum layer reached by the electron track (after fitting)
- Genfit error matrix Longitudinal position
- Genfit error matrix Longitudinal momentum
- Fitted longitudinal momentum

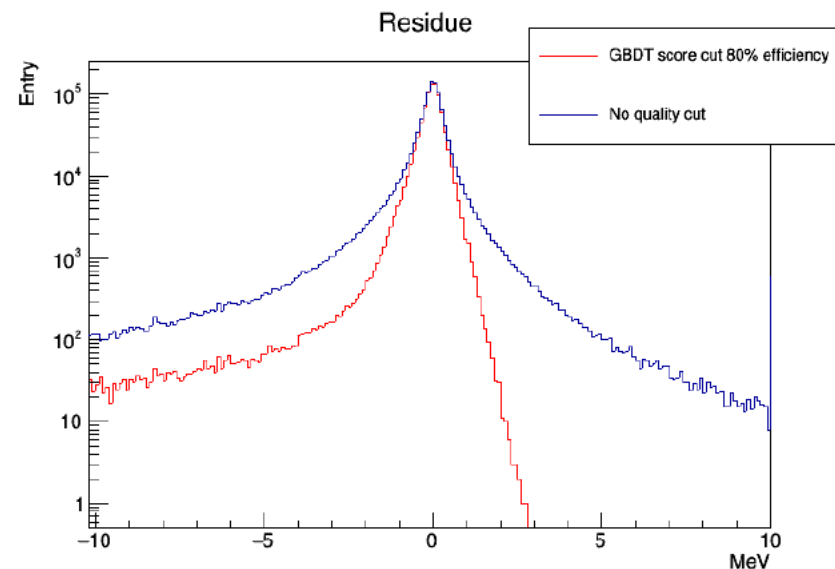
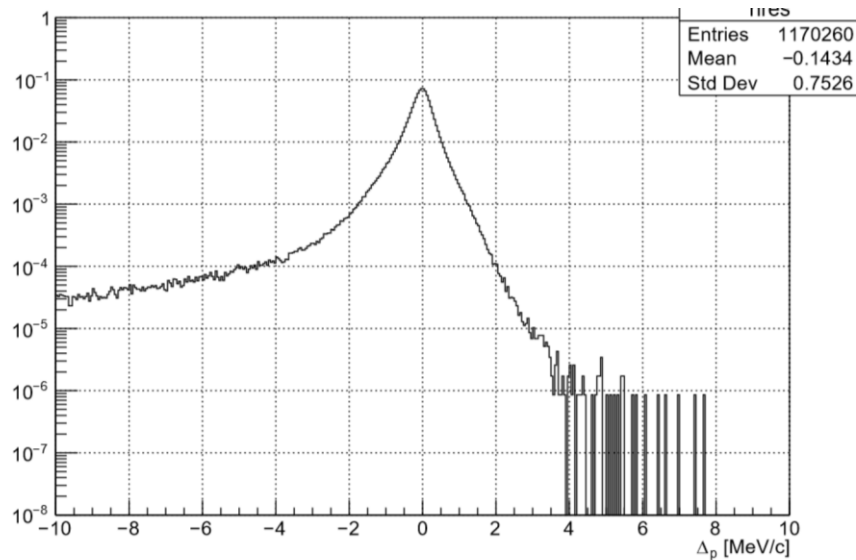


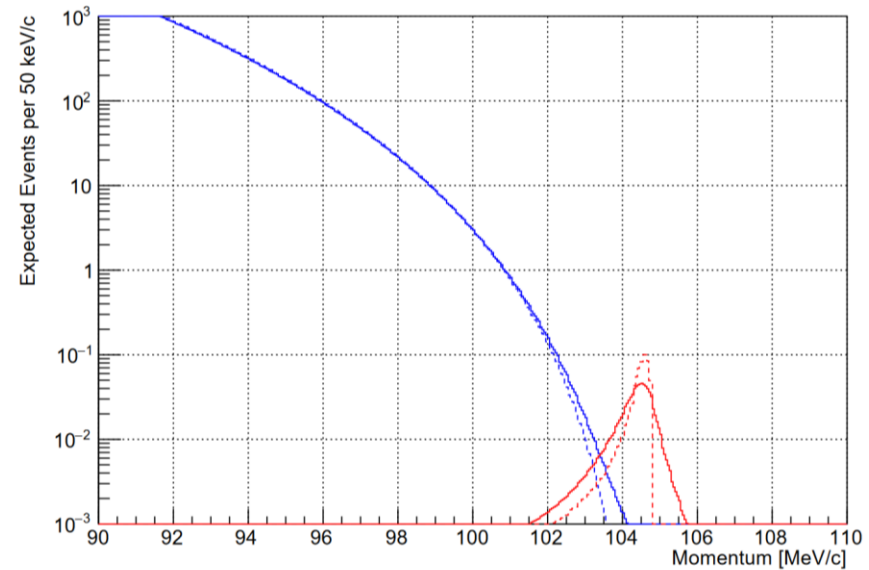
Figure B.8: Residue distribution by taking the best 80% fitted events according to the GBDT score and without cut

Performance at high occupancy

The total efficiency is 50% when DIO contribution is 0.05 @Hit occupancy of 50%
(w/o track finding, failed on multiple turn)



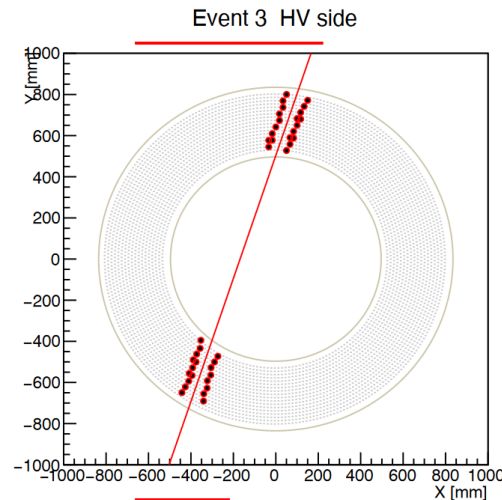
Momentum distribution



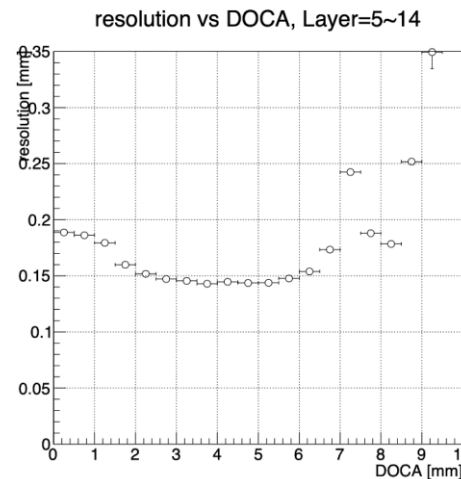
DIO and signal distributing

Cosmic ray test analysis with ICEDUST

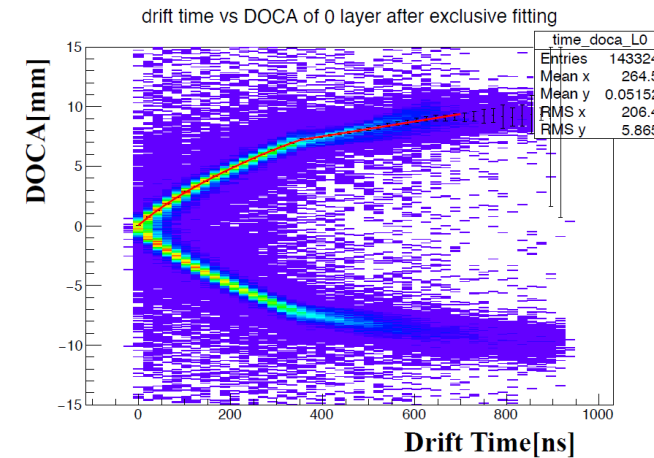
- CDC cosmic ray test data study is done under ICEDUST
- X-T relationship , T0 and resolution calibration is done



Event display



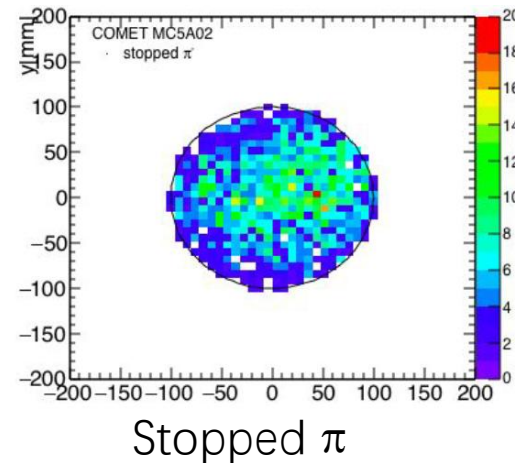
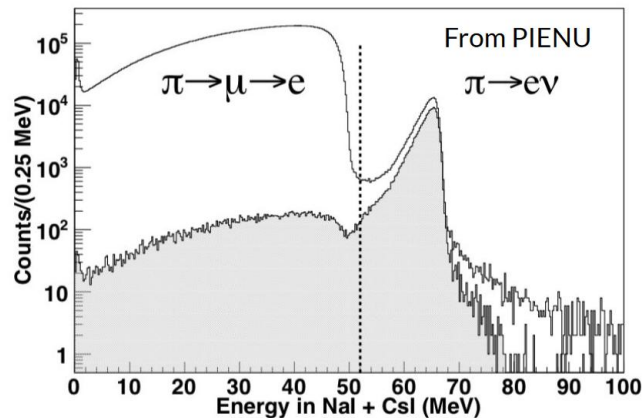
Spatial resolution v.s. doca



X-T relationship

Momentum calibration

- $\pi^+ \rightarrow e^+ \nu$
 - A stopped π^+ decays into a monoenergetic e^+ at 69.8 MeV ($r=23.5\text{cm}@1\text{Tesla}$)
 - Used by SINDRUM II and the PIENU for momentum calibration
- Data taking time and background have been studied preliminary



Data management

- Raw data
 - Raw data is stored at storage near the detector.
 - Data rate: ~300MB/s
 - Data size(150 days): ~9PB of raw data
- Metadata -- information of data file
 - Run number
 - Software library which processes data
 - Location of file and so on
- Database
 - Calibration information
 - Detector condition and so on.

Data distribution

- Data sites and distribution
 - Main data site is at KEKCC
 - Shared with Osaka University, in2p3-cc and Dongjiang Lab. China
 - Phase- α data distribution has been proved
- Analysis
 - By using DIRAC, detailed analysis is realized by ROOT
- Software distribution
 - CVMFS (CERN Virtual Machine File System), ready at IHEP

Computing estimation

- Cosmic ray simulation (150 days)
 - Sneaking cosmic ray muon: ~0.5M CPU hours
 - Atmosphere cosmic ray muon: ~3.5M CPU hours
- Beam simulation
 - 0.5M CPU hours / year
- Event reconstruction (data)
 - Data trimming (trigger rate 13kHz)
 - ~0.05M CPU hours
 - Reconstruction
 - ~ 2M CPU hours
- 3 years computing budget
 - ~8M CPU hours
 - 14k\$/M CPU hours and ~110k\$ in total for 3 years

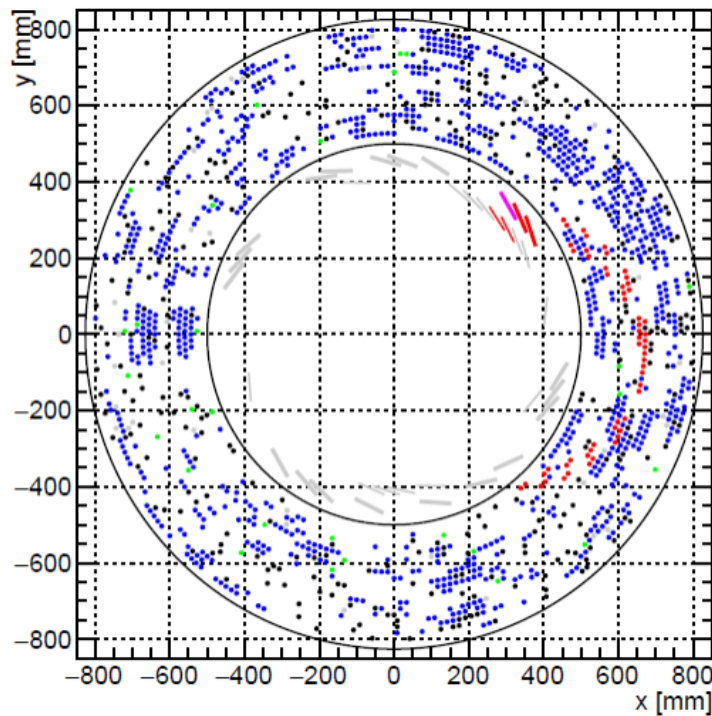
Summary

- ICEDUST as the offline software framework
 - Simulation
 - Reconstruction
 - Calibration
 - Real data (cosmic ray test data)
 - Mass production
- Tracking algorithms for COMET Phase-I are ready
 - Algorithms:
 - De-noising and pre-fitting
 - Track finding and fitting for single turn, multi-turn is under development
 - GBDT track quality cut
- Data management and computing estimation have been made

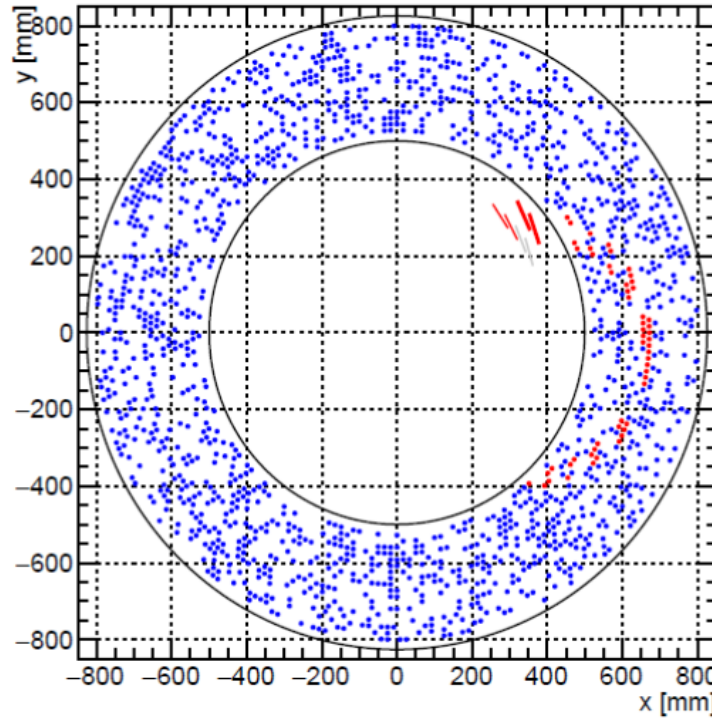
backup

Background hits simulation

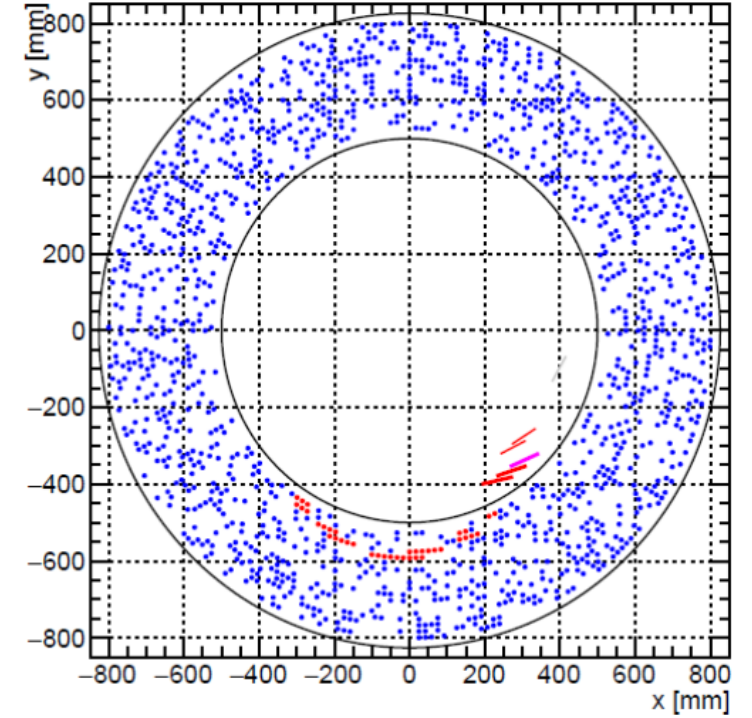
From MC (only noise-like)



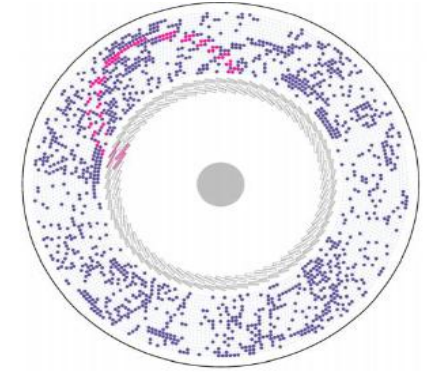
From GAN (half bunch)



From random noise (30%)



Tracking at high hit occupancy



Can be done
using GBDT or NN

Denoising



Can be done
using either
iterative fitting, or
parameter
scanning, or NN

Track Finding



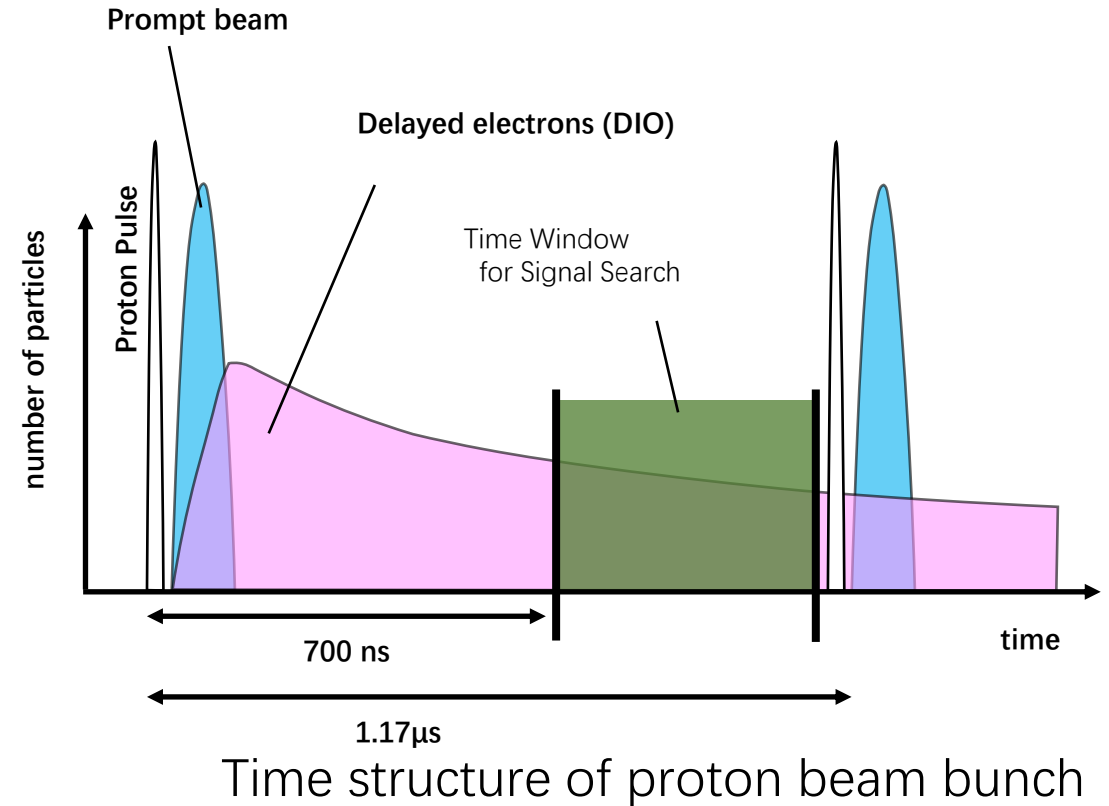
Kalman filter is
eventually needed
for the final fitting
step. Genfit2 is
the default choice.

Track fitting

- Denoising
 - To filter out the majority of noise in the chamber
- Track finding
 - To find out track candidates:
 - Group hits
 - Estimate track parameters as seed for further fitting
- Track fitting
 - To provide final estimation of the momentum of the studied trajectory upon its entrance in the drift chamber.

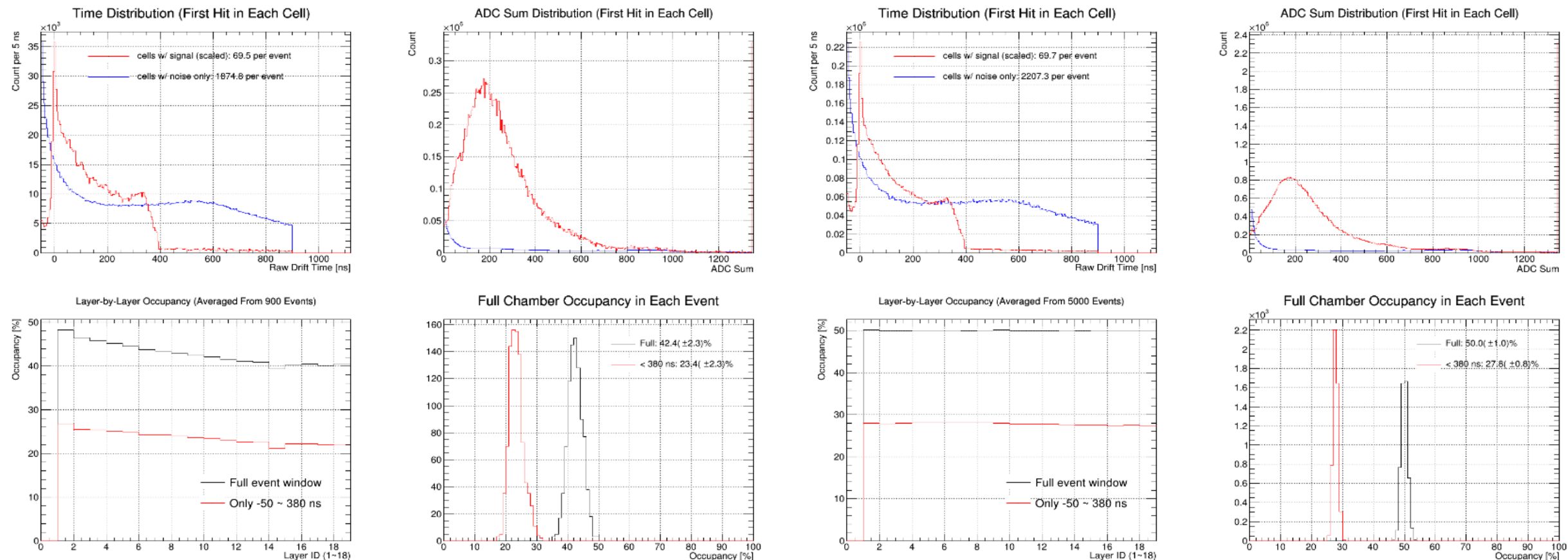
Background suppression with beam window

- 3kW proton prompt beam:
 - $\sim 10^{19}$ protons on pion targets (in 145 days running time)
- Bunch structure of proton beam
 - Bunch size $\sim 10^7$ POT
 - Bunch spill/width $\sim 100\text{ns}$
 - Extinction factor 3×10^{-11}
 - Bunch separation time = 1170ns



Background simulation

Resampling from beam simulation

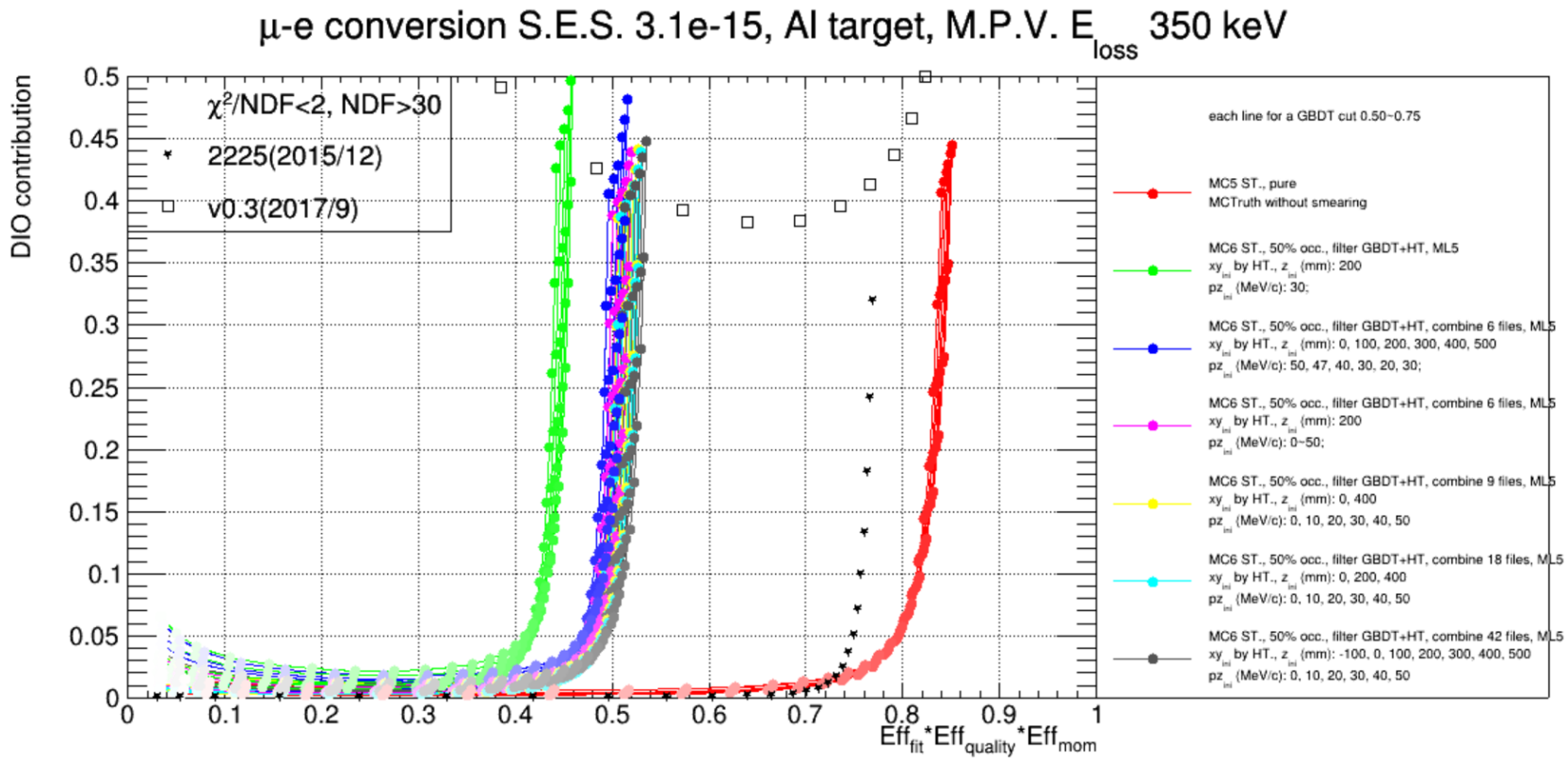


MC sample

Random noise sample (50% occ.)

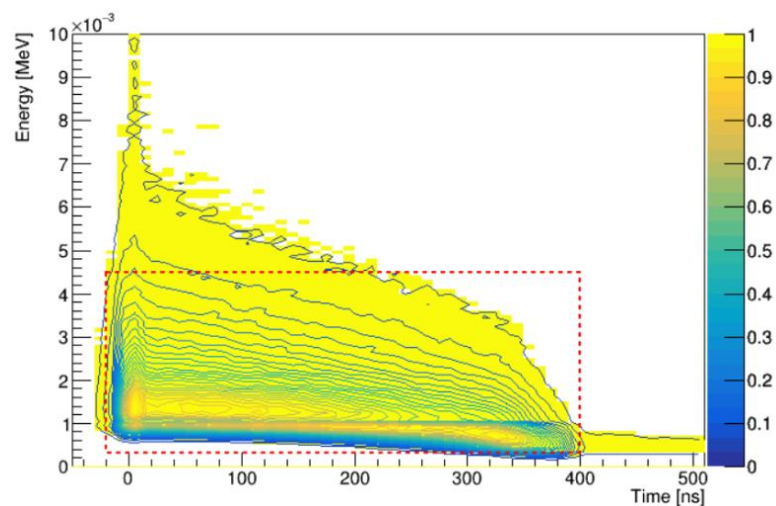
15

DIO contribution vs efficiency



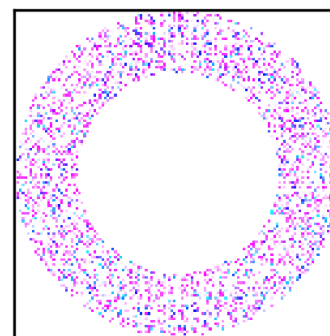
De-noising

- ADC v.s. TDC as 2D box cut



- Feature Pyramid Network (FPN)
 - Network for object detection

input



FPN prediction



Track fitting quality study

- Tracking fit quality → key to COMET physics goals
- Tail events at high momentum → background source

Momentum resolution & tail vs. mass cut

	nEvent	Mom res(keV/c)	Tail($\Delta > 2$)
No cut	12554	250	1406(11.2%)
Fitstatus == 3or7	10813	245	1116 (10.321%)
&chi2<=100	9286	233	787(8.475%)
&chi2<=75	8159	225	666(8.163%)
&chi2/ndof<=4	9882	240	796(8.055%)
&chi2/ndof<=2	7636	207	317 (4.151%)

Preliminary Results:

- Momentum resolution: 250 → 207 keV/c(17% improvement)
- Tail ($\Delta > 2$): 11.2% → 4.15% (63% reduction)

Momentum resolution vs corner cut

