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

CEPC

TPC Track Reconstruction in CEPCSW

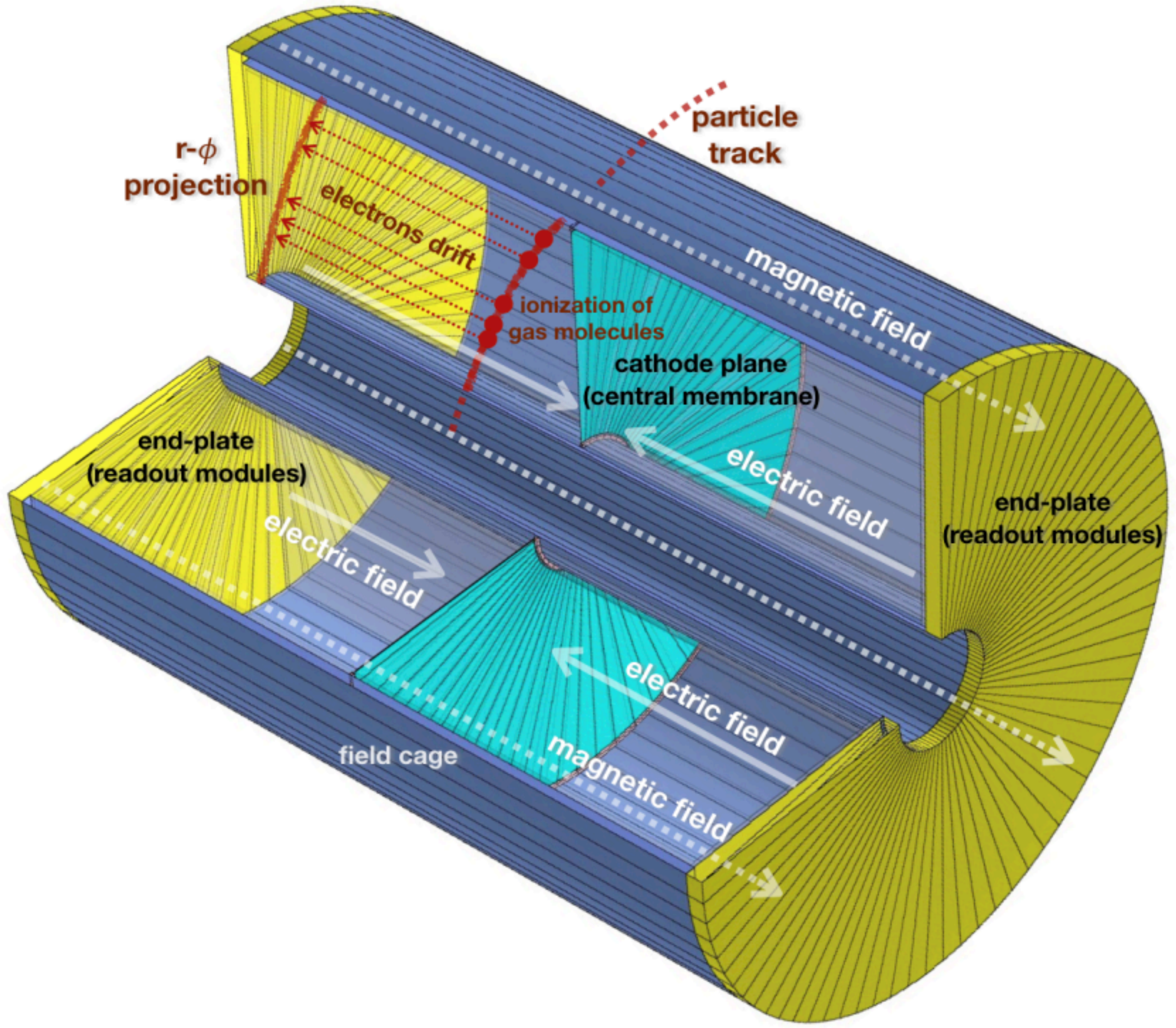
Chu Wang, On behalf of CEPC TPC Software group

Outline

* Time projection chamber in CEPC
* Pixelated TPC Updates
* TPC Simulation, Digitization and Reconstruction
* Machine learning based TPC hits merging
* Pixelated TPC Performance
* Summary

Time projection chamber in CEPC

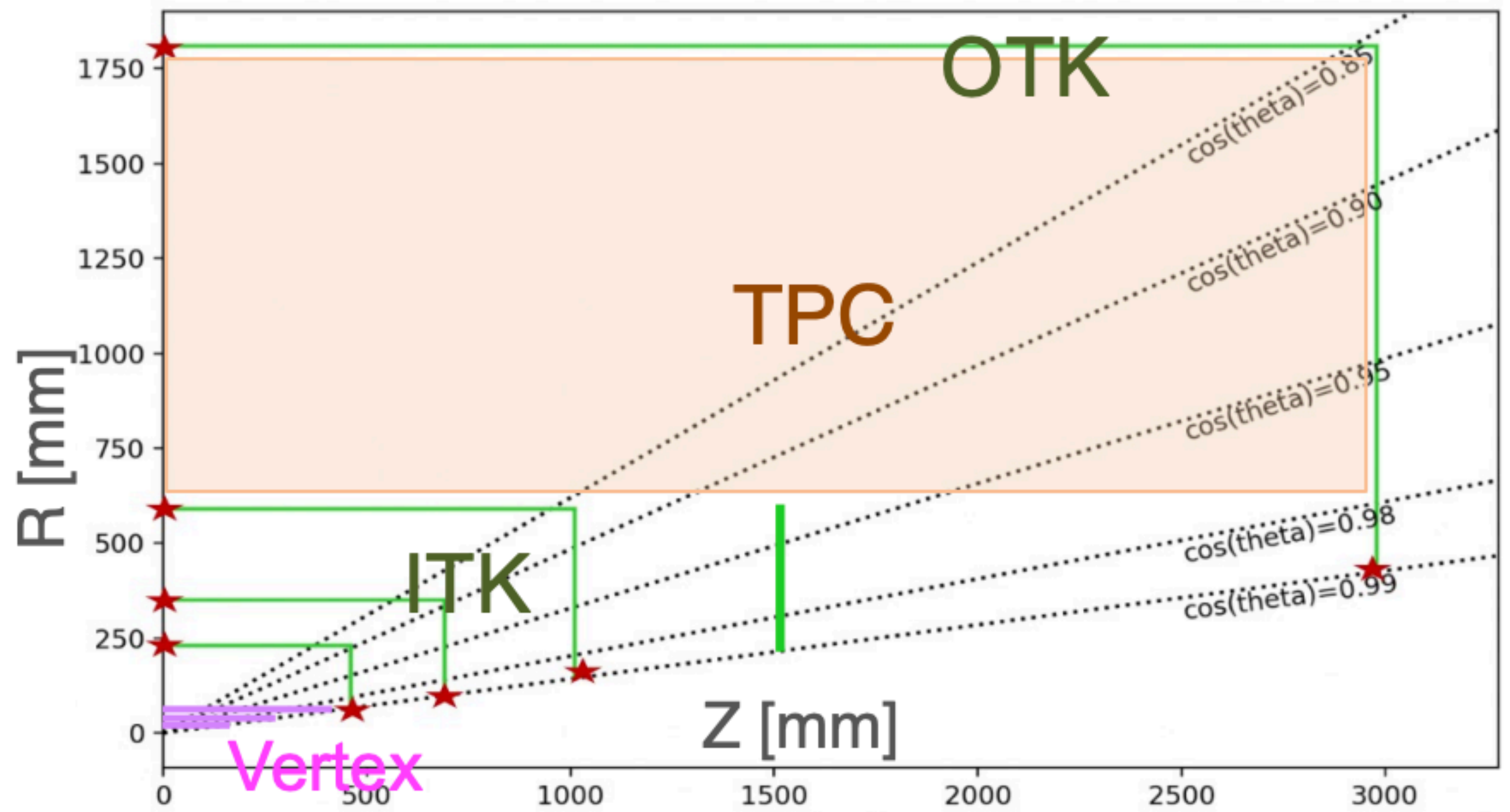
- **The TPC is the gas detector of CEPC for PID and tracking, especially improve track performance at low momentum**
  - TPC can provide large-volume high-precision 3D track measurement with stringent material budget

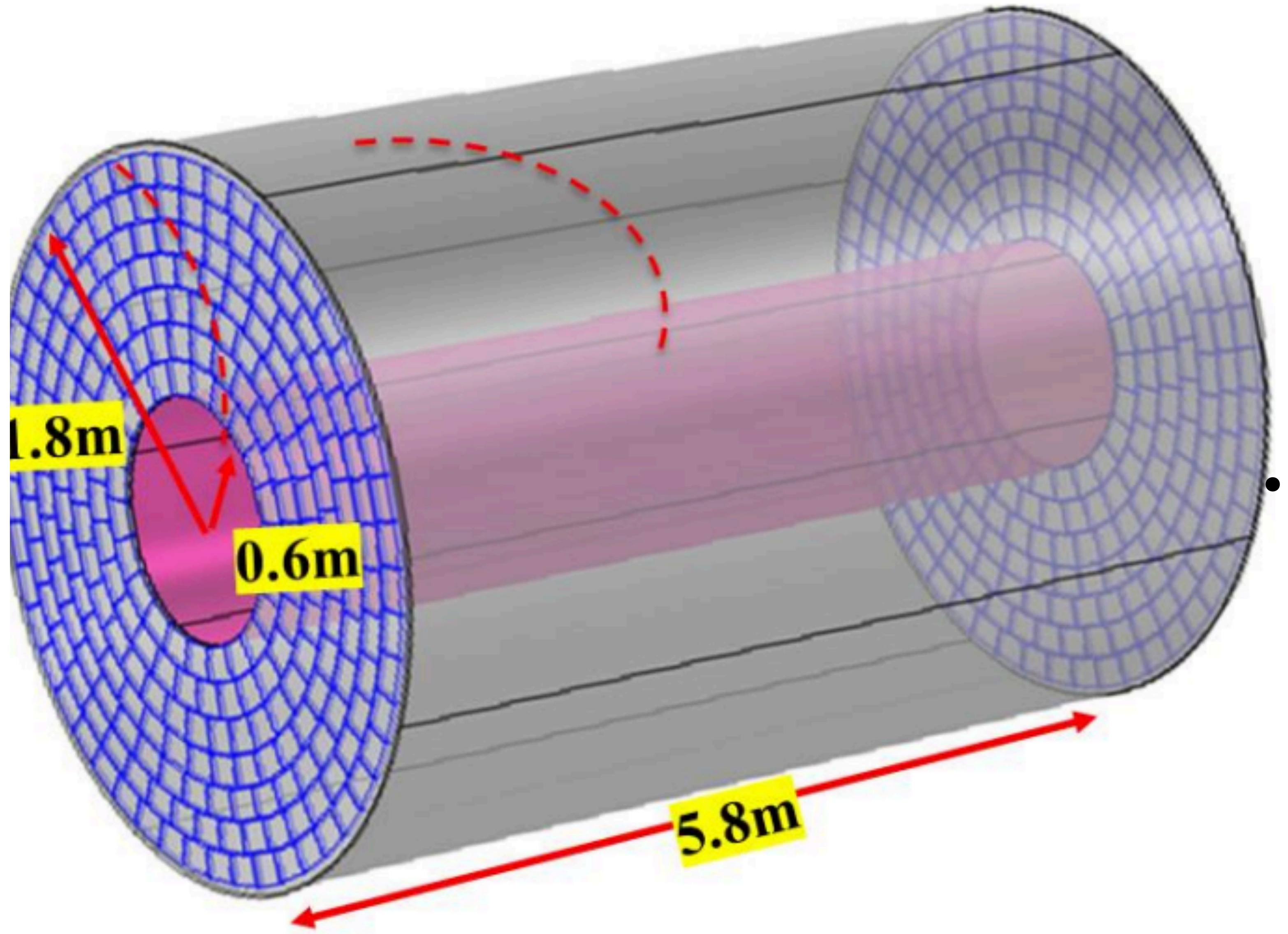


- Working principle
  - As the particles pass through the gas in the TPC, ionization will occur.
  - The ionized electrons will drift to the readout system in the presence of an electric field.
- TPC has many advantages:
  - Provided many hits in readout (can extrapolate the 3D positions of TPC hits)
  - Can reconstruct long tracks in TPC
  - High position and momentum resolutions
  - Can be used to identify different particles based on the density of the ionized electrons in readout
  - Good balance of physics performance and budget

Pixelated TPC Updates

▸ **To improve performance and reduce the difficulty at high luminosity Z pole run. A pixelated TPC has been designed.**



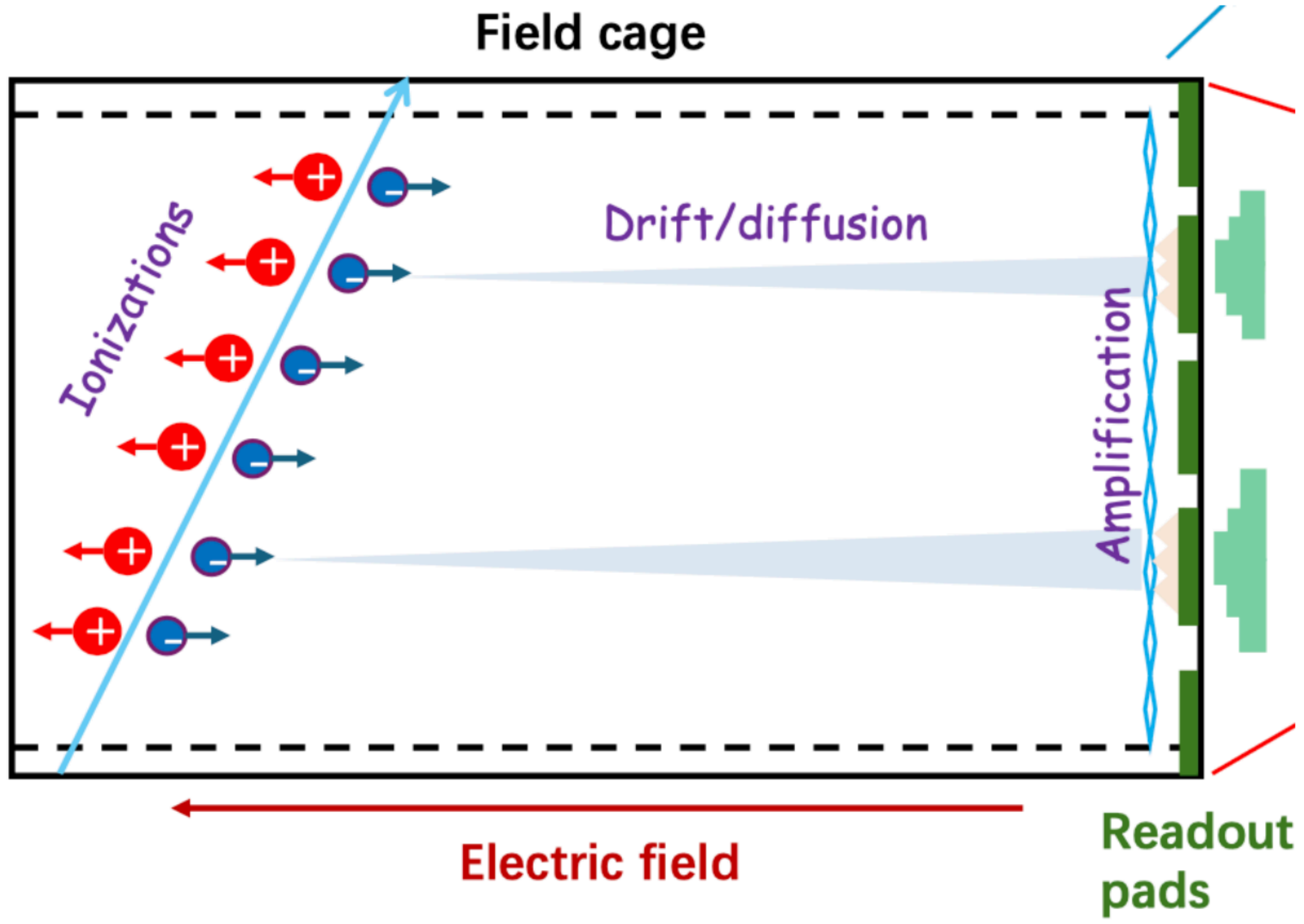


Updated Geometry of TPC in TDR

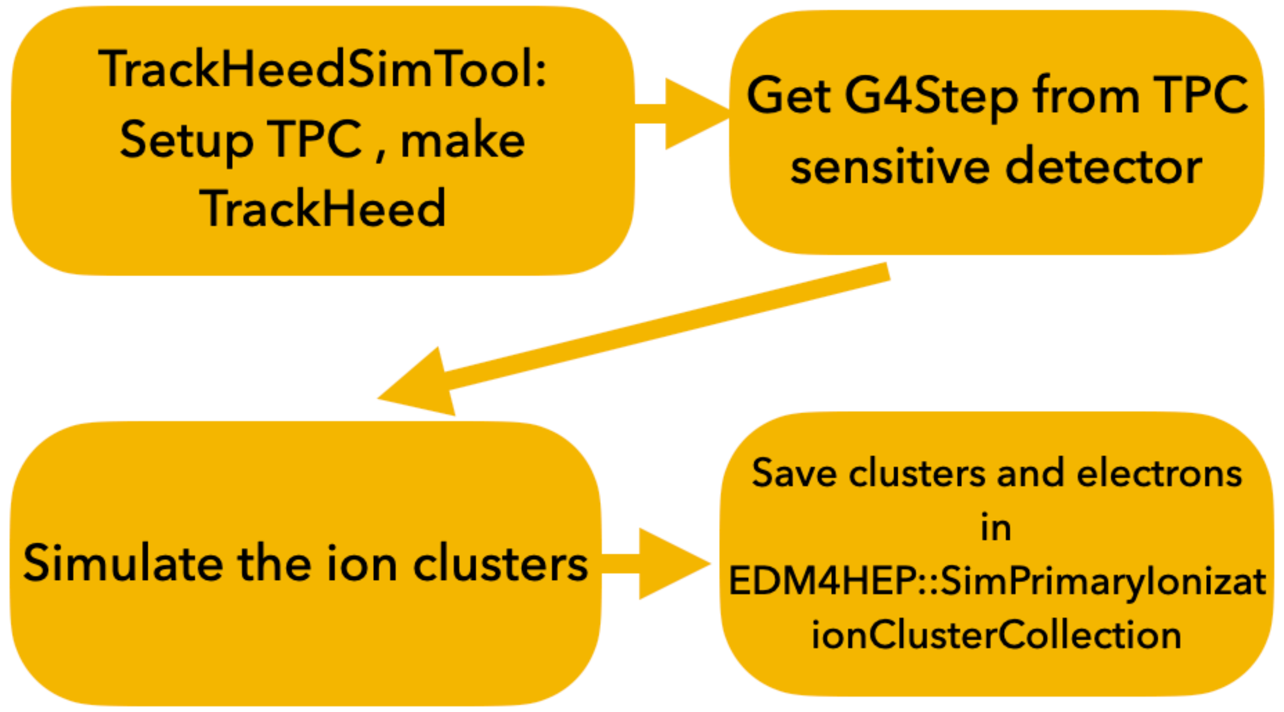
- Granularity has improved in readout
  - TDR: pad (1mm x 6mm) → pixel (0.5mm x 0.5mm).
  - $2x3x10^7$ channels in readout endplate.
  - Divided into 248 modules/endplate in 21x17cm.
- Updated TPC will have:
  - More hits (~200 hits in TDR pad readout→~2000 hits in pixel)
  - High density readout make particle ID possible in ID algorithm.
  - 5 ns drift time resolution

Simulation algorithm:

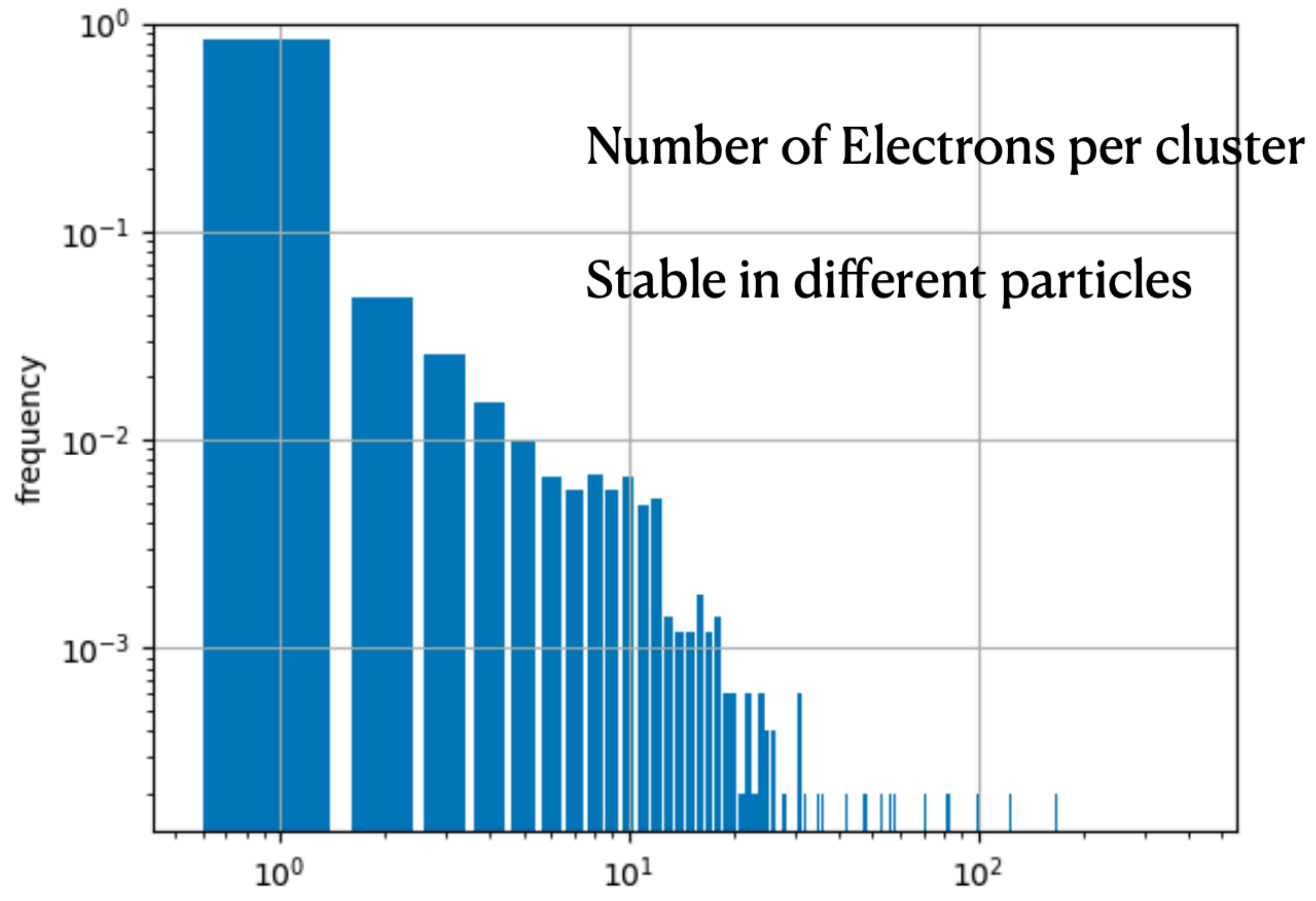
- A new simulating process is implemented to simulate it by Garfield/Heed in sensitive detector stage
  - A SimPrimaryIonizationClusterCollection is created to save the ionized electrons.



Schema of TPC ionization



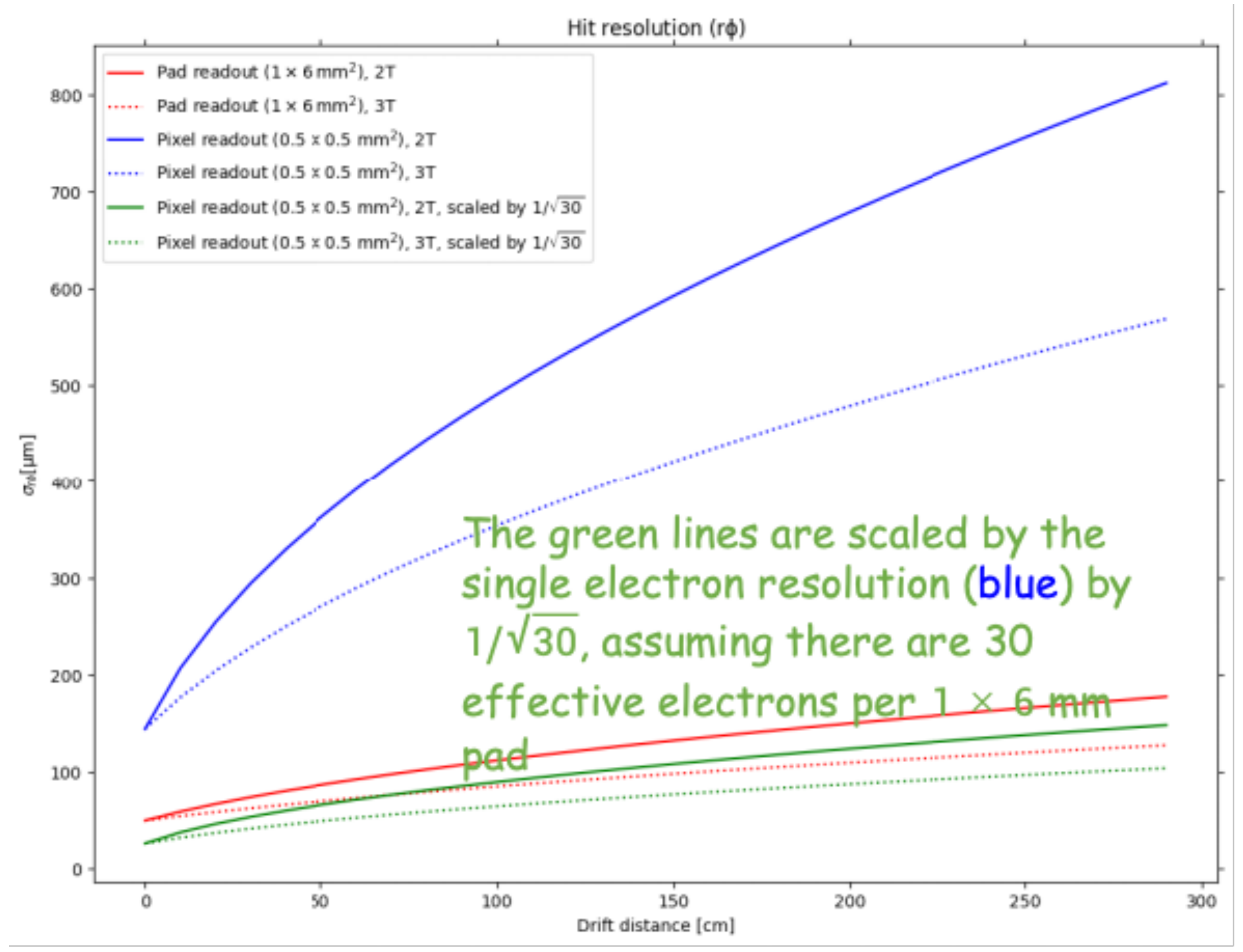
Workflow of TPC simulation



The number of Electrons per cluster is expected

Pad-like Digitization

- **Simple digitization in CDR (pad)**
  - Parameterized pad spatial resolution
  - Smearing resolution on MC truth
- **Simple digitization in current TDR (pixel) → pad-like**
  - Parameterized single electron resolution according to Garfield simulation
  - Change resolution according average electrons number in multi pixel layer
    - 12 layers of 0.5mm pixels→same size with 6mm pad
  - Smearing resolution on MC truth
  - Equivalent to ideal (find out the exact center of cluster) pixel readout



Pad readout：

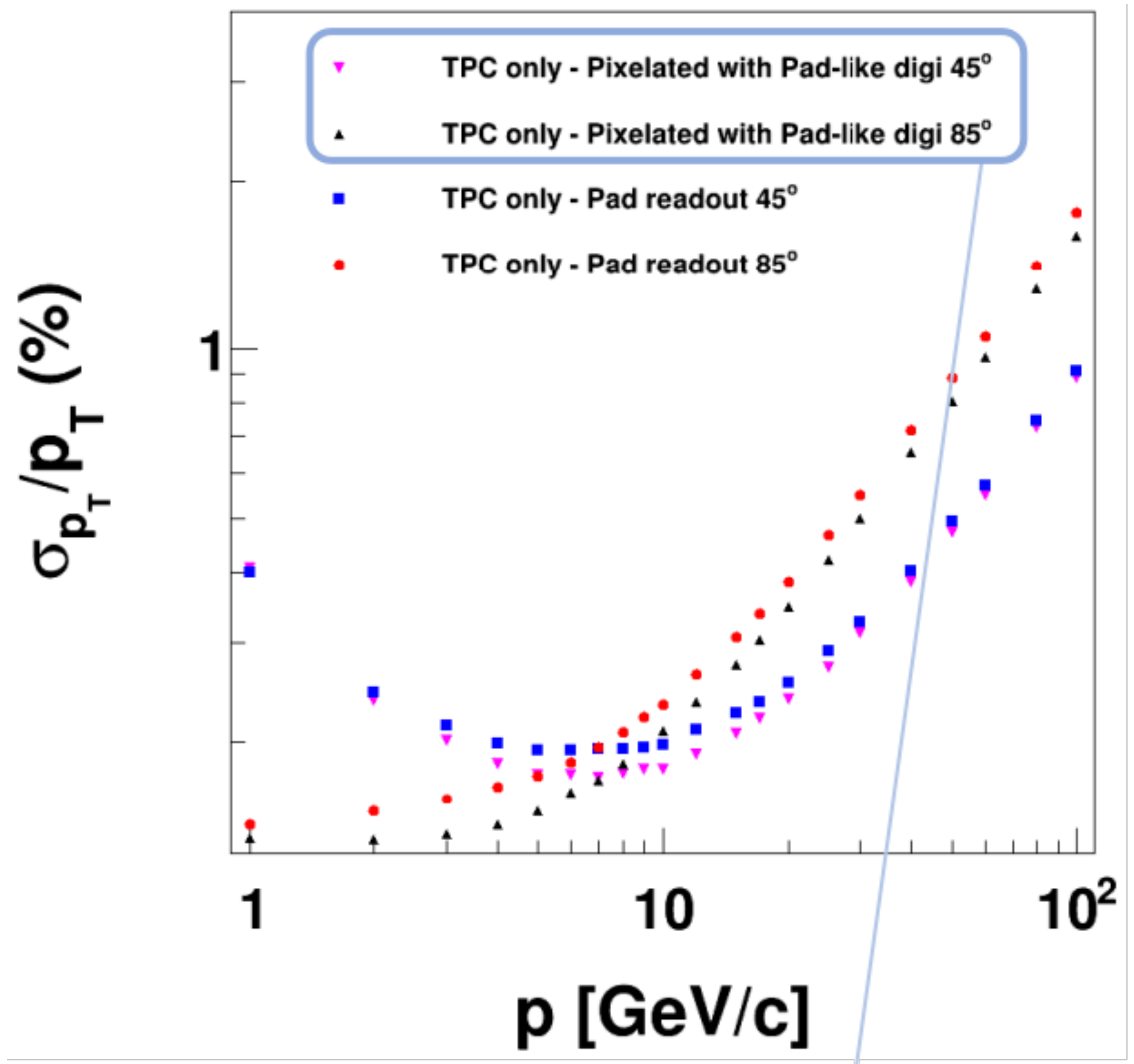
$$\sigma_{r\phi}^{pad} = \sqrt{(\sigma_{r\phi 0}^{pad})^2 + \sigma_{\phi 0}^2 \sin^2(\phi_{track}) + L\frac{D_{r\phi}^2}{N_{eff}} sin(\theta_{trk}) (\frac{6mm}{h_{pad}})}$$

- $\sigma_{r\phi 0} = 50\mu m$
- $N_{eff} = 22$
- $D_{r\phi} = 46.9\mu m/\sqrt{cm}(2T)$， $32.3\mu m/\sqrt{cm}(3T)$

Pixel readout：

$$\sigma_{r}^{pixel} = \sigma_{r\phi}^{pixel} = \sqrt{(\sigma_{r\phi 0}^{pixel})^2 + LD_{r\phi}^2}$$

- $\sigma_{r\phi 0} = \frac{500}{\sqrt{12}} = 144\mu m$
- $D_{r\phi} = 46.9\mu m/\sqrt{cm}(2T)$， $32.3\mu m/\sqrt{cm}(3T)$



Ideal performance limit, can be regarded as goal of pixelated TPC (0.5mm)

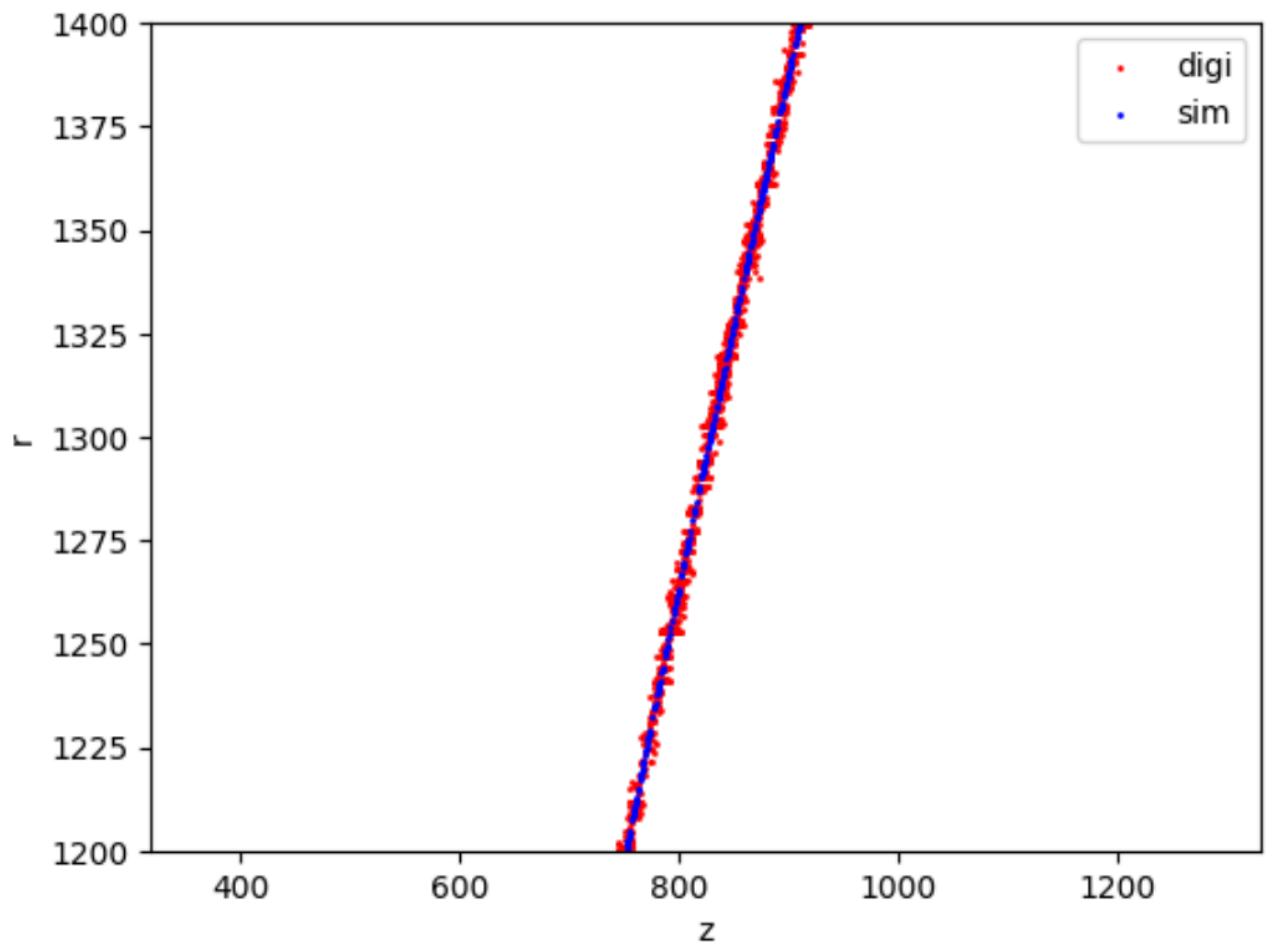
Digitization on Ionization PrimaryIonCluster

Digitization Algorithm :

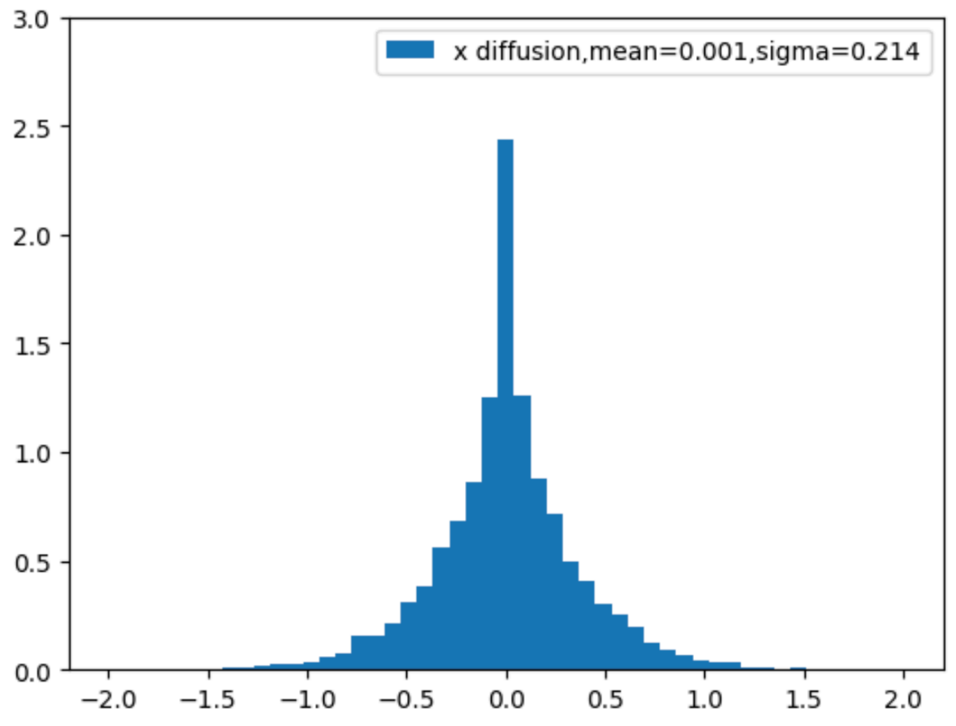
- Read the electron clusters in SimPrimaryIonizationClusterCollection
- Loop all electrons in each cluster, gaussian smear the electrons from initial positions.
  - Based on R-ϕ and Z direction resolutions from standalone Garfield simulation
  - Create the output hits based on the smeared position (More than 2200 hits will be created)
  - If multiple hits hit the same pixel, it now will be treated as only one hit (Will update in future)
- Match the sim tracker hits and output digitized hit for performance study.
- Save the digitized hits in TPCTrackerHit collection

Hits after Digitization

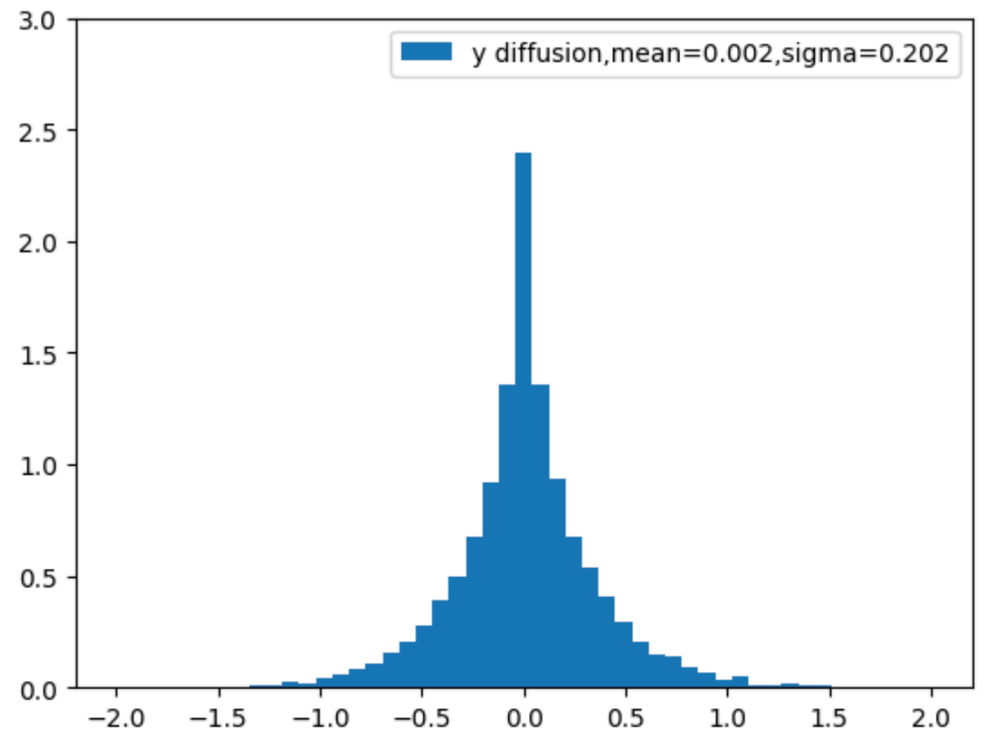
- **Digitization output:**



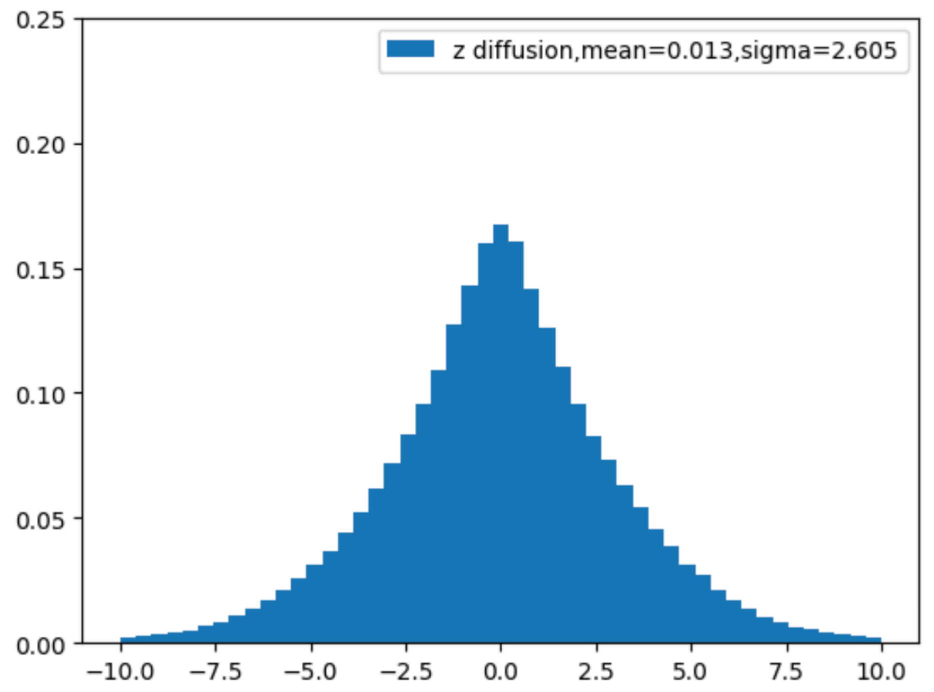
The digitized hits and sim tracker hits, large diffusion observed, theta=60, pt=1GeV, drift length:1480~2380.



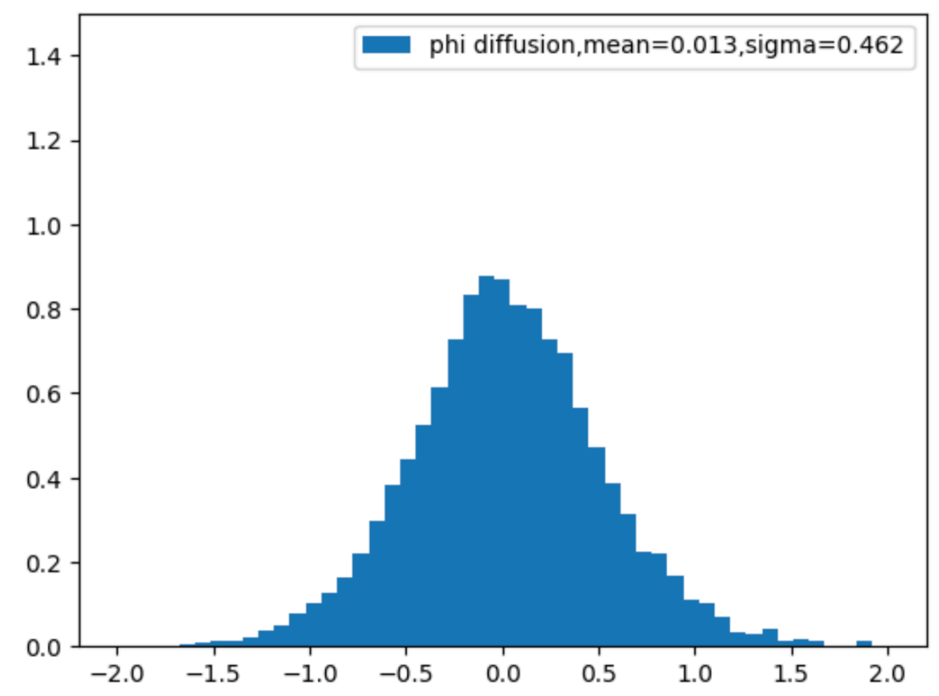
x diffusion w.r.t sim tracker hits in same layer



y diffusion w.r.t sim tracker hits in same layer



z diffusion w.r.t sim tracker hits in same layer



phi diffusion w.r.t sim tracker hits in same layer

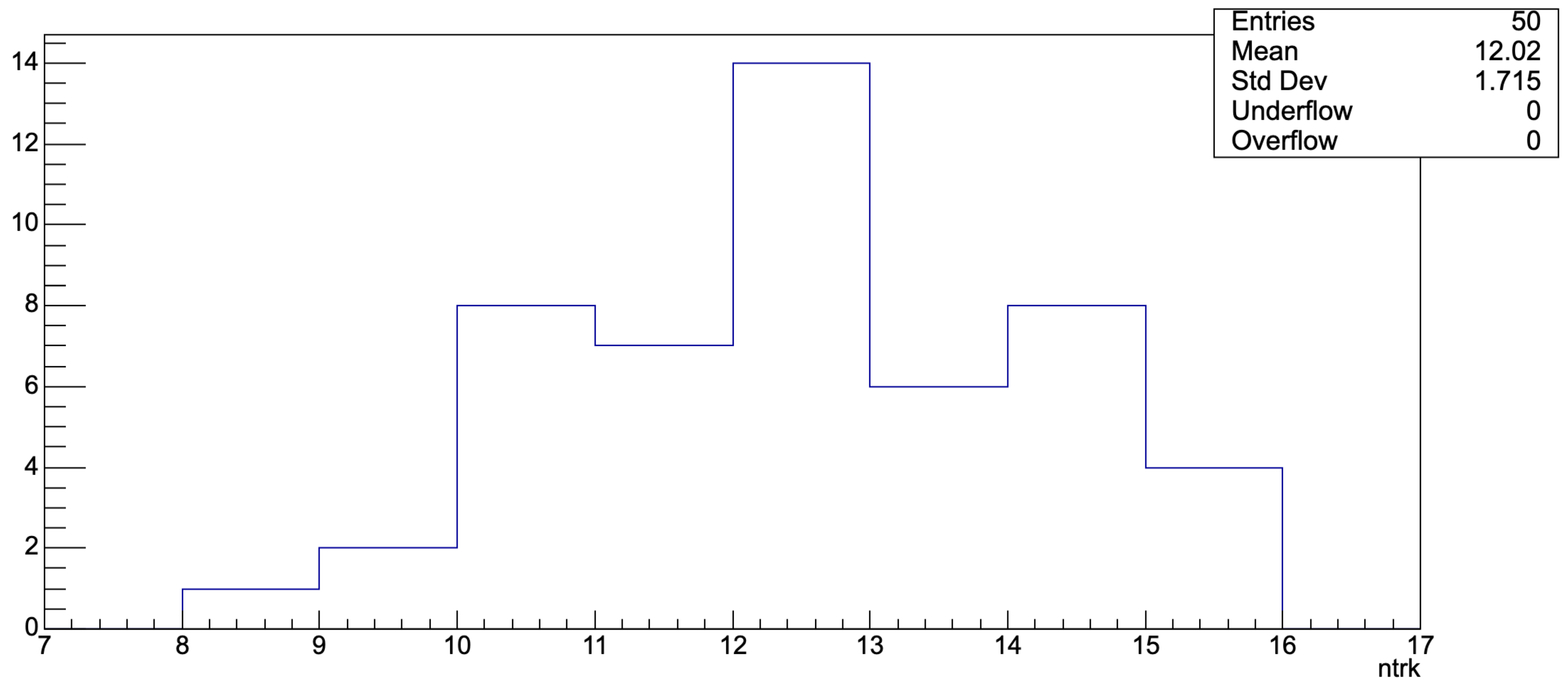
TPC Reconstruction

Reconstruction Algorithm (clupatra):

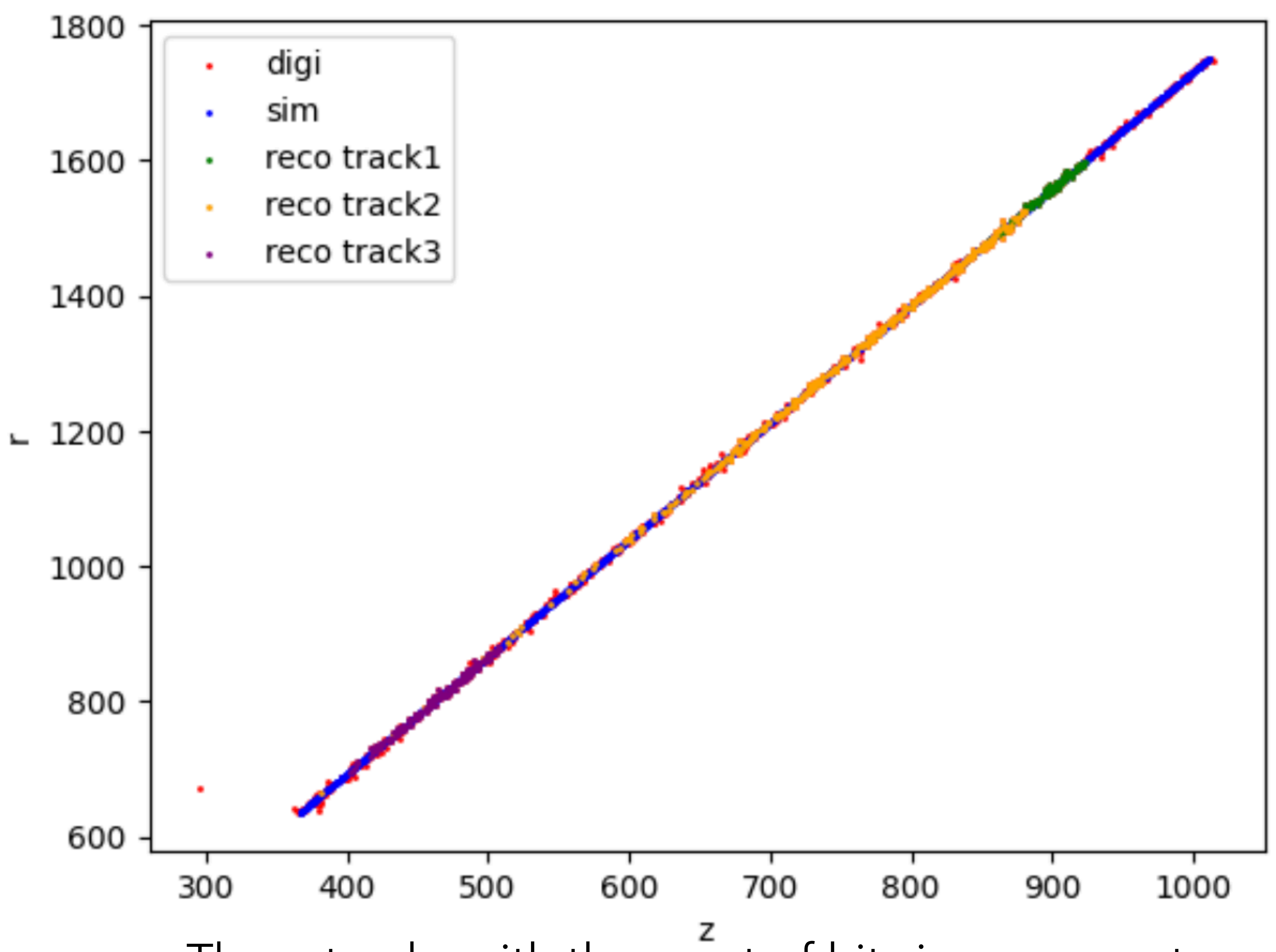
- Clupatra is an algorithm developed by ILC, it works as follows:
  - Initial Clustering by Near-Neighbor-Clustering:
    - It will loop all hits, and merge the hist into one cluster with distance less than some criteria (eg. distance and angle)
  - Extend the cluster forward and backward using Kalman-fitter
    - Create track segments
  - Merge track segments based on criterion for R, tan(lambda)
- In high granularity and high diffusion hits, the algo may:
  - Find multiple tracks, because of the diffusion.
  - Need to merge hits before clustering.

Tracking on Raw Hits

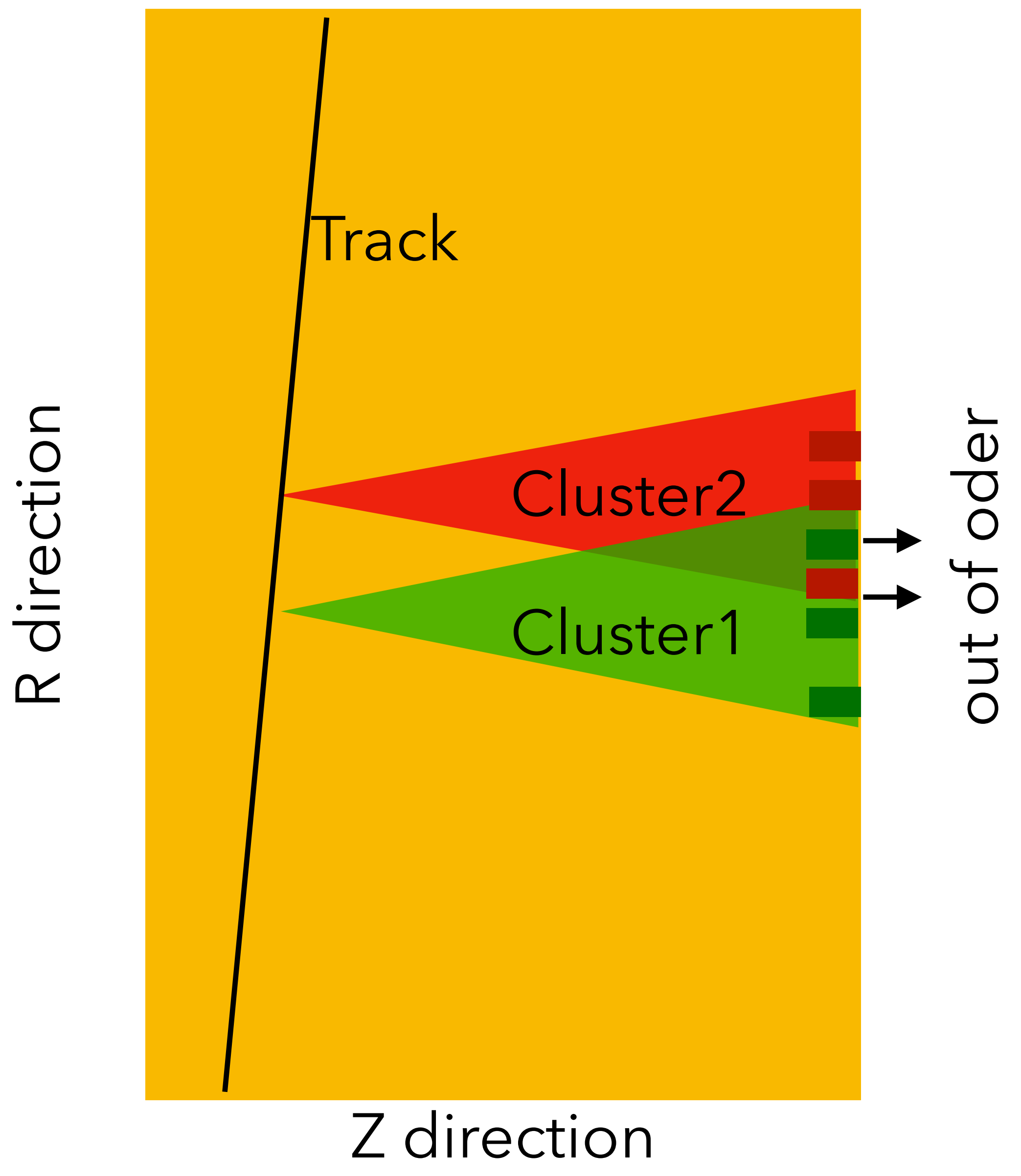
- **To avoid the splitting of the tracks in TPC, Hits merging are needed.**
  - If no merging, there will be multiple tracks reconstructed by clupatra algo.
  - Possible to cause hits out of order because of diffusion



An example of the number of tracks in pt=50GeV, and theta=60^o



Three tracks with the most of hits in one event



A schema of the hits out of order

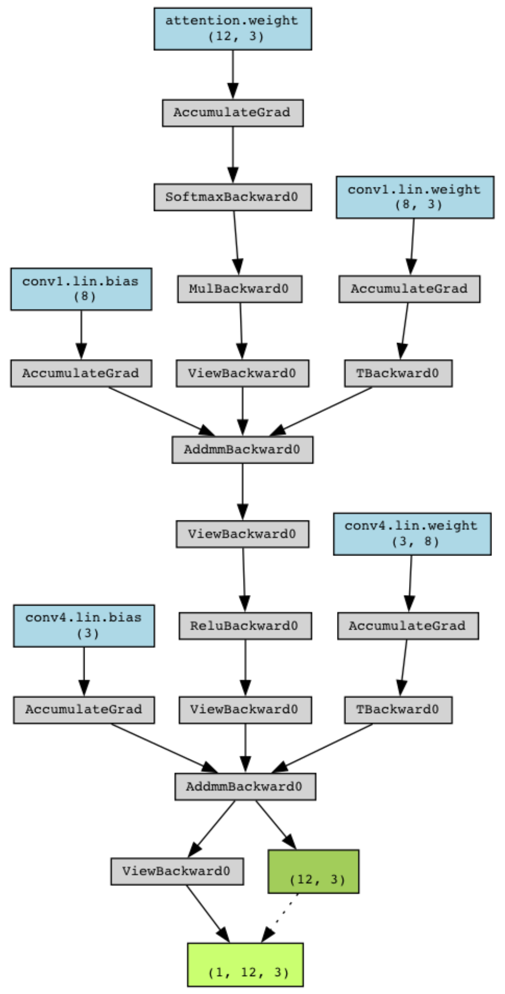
Implementation of ML-based clustering

- **As the Pixelated TPC will have more than 2200 hits in all "layers"**
  - Large diffusions are observed in x-y plane.
  - To reduce the inefficiencies of track hits and improve the resolution of single hits. Need to merge hits, there are two possible ways to do so:
    - Calculate the mean value of the hits in multiple layers, where layers are defined as the ring with height=0.5mm in readout. If the curvature is large, the mean value will be biased.
    - Develop a Machine-learning based method to merge hits.
  - For the ML merging, we chosen a simple GCN model to collect informations from hits in multiple layers
    - To match the position of sim tracker hits

Machine learning based hits merging

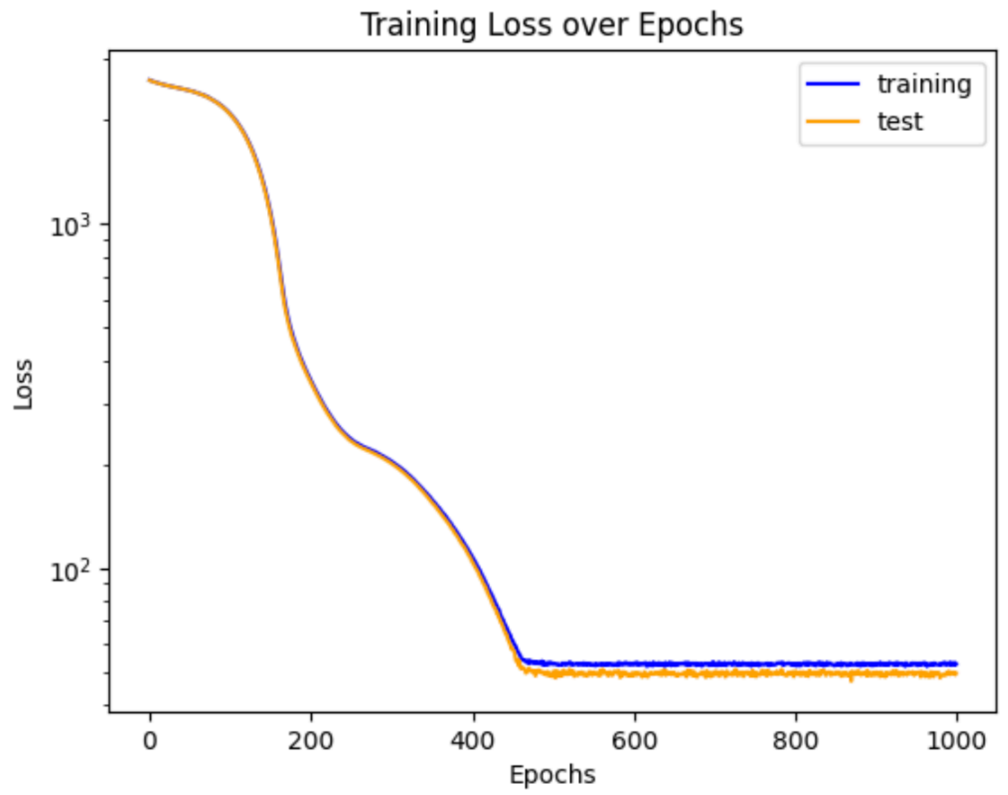
- **Architecture of GCN model**
  - The graph was created by connect 11 nodes (hits) to one output node (hit).
  - Each hit(node) has 3 features (r,phi,z)

- **Training**
  - Trained with simulated samples in different:
    - pt: 1,3,15,100GeV
    - Theta: 60,85
    - Phi: randomized
  - Loss: the distance between "output-node" and SimTrackerHit.



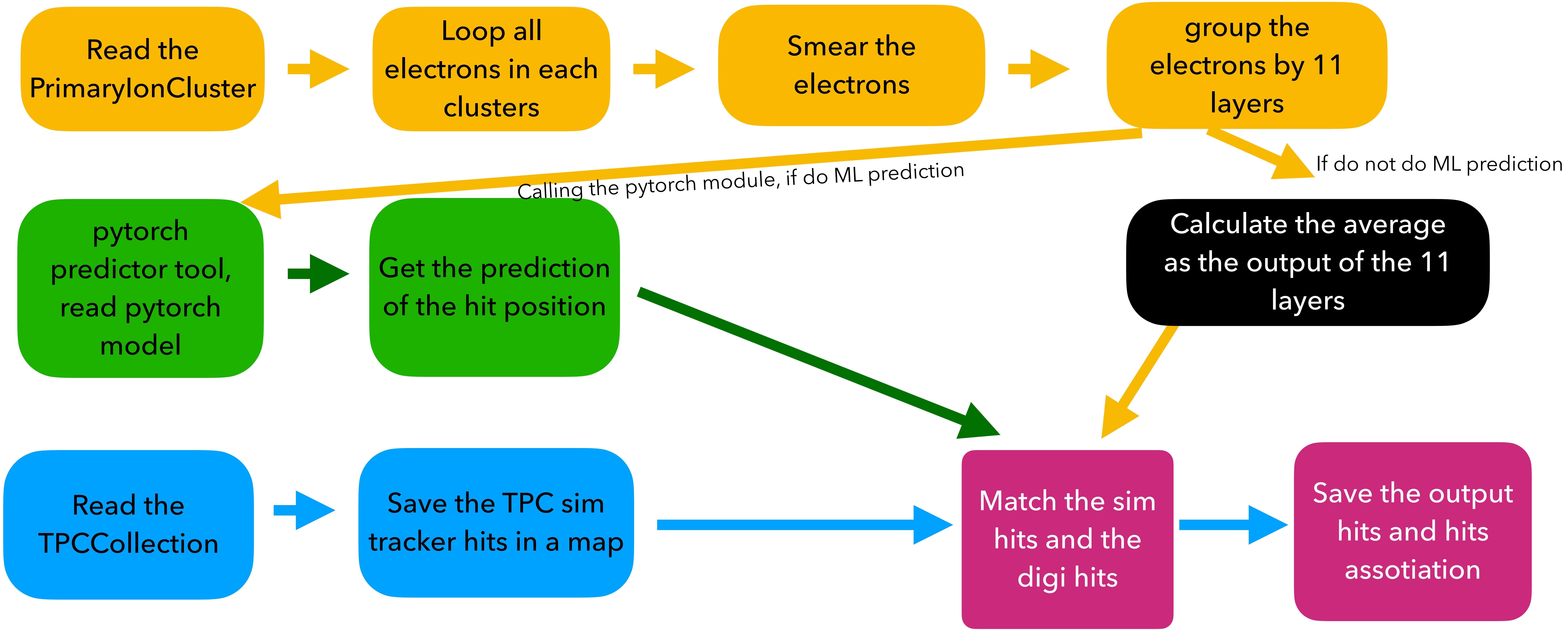
```
----------------------------------------------------------------
        Layer (type)               Output Shape         Param #
================================================================
            Linear-1            [-1, 1, 12, 8]              32
           GCNConv-2            [-1, 1, 12, 8]               0
            Linear-3            [-1, 1, 12, 3]              27
           GCNConv-4            [-1, 1, 12, 3]               0
================================================================
Total params: 59
Trainable params: 59
Non-trainable params: 0
----------------------------------------------------------------
Input size (MB): 0.00
Forward/backward pass size (MB): 0.00
Params size (MB): 0.00
Estimated Total Size (MB): 0.00
----------------------------------------------------------------
```

Structure of the Model



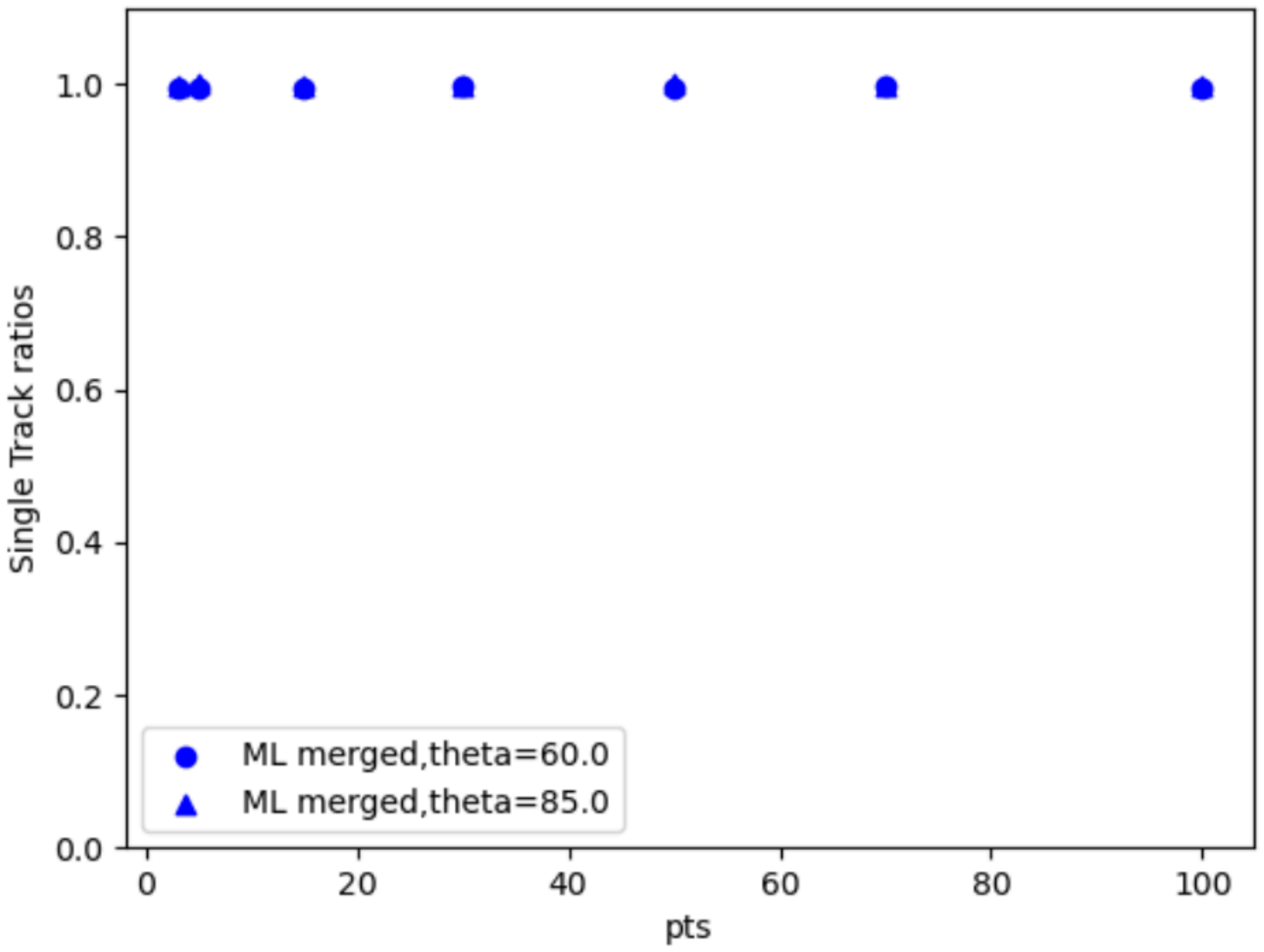
Loss vs Epoches

Workflow of the Digitization

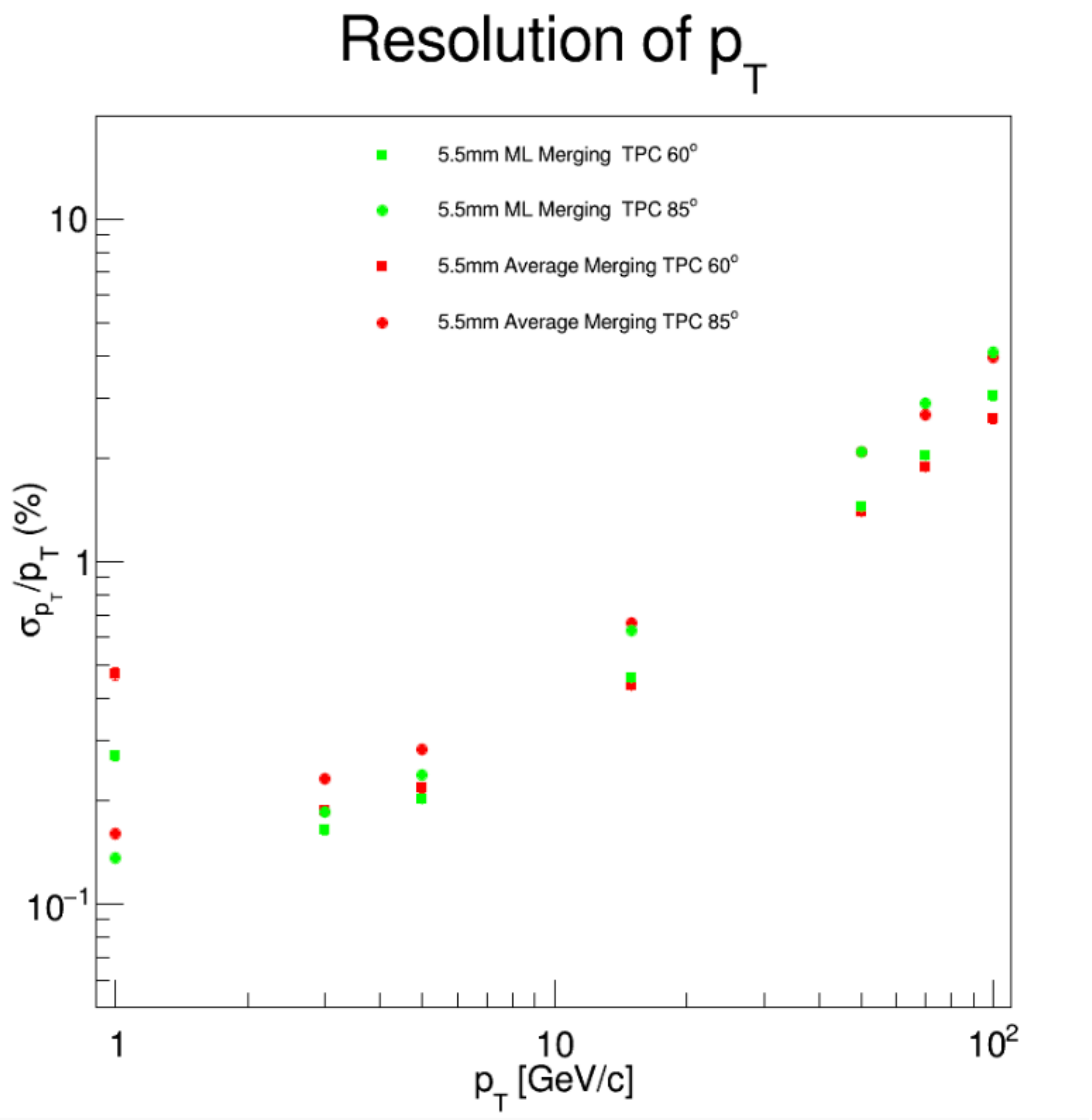


Performance of ML based merging

- **After deployed ML model, the track splitting has been fixed, and the pt resolution has improved in low pT region**



Single track ratios



pT resolution has improved in low pT region

Summary

- **The upgrade of the pixelated TPC need new algorithms for the simulation, digitization**
  - We implemented the new algorithms for the pixelated TPC.
- **New issues raised from the higher density of the TPC pixels**
  - More pixels may misleading the reconstruction algorithm. Result in track splitting.
  - We proposed to merge the hits before reconstruction.
  - A ML based merging has been implemented, it has better performance than the simple average of the hits at low pT region.
  - There will be a further improve the performance by optimizing the algorithm.
- **The new digitization can provide an input for future's particle ID study.**