AI/ML Assisted Central Detector Denoising

CLAS Collaboration Meeting

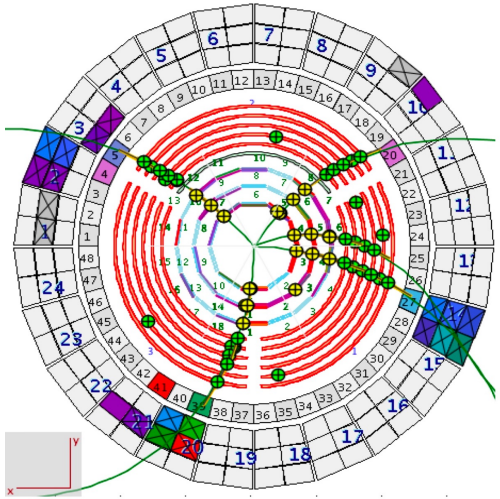
Mathieu Ronayette, Pierre Chatagnon, Tongtong Cao, Richard Tyson, Raffaella De Vita

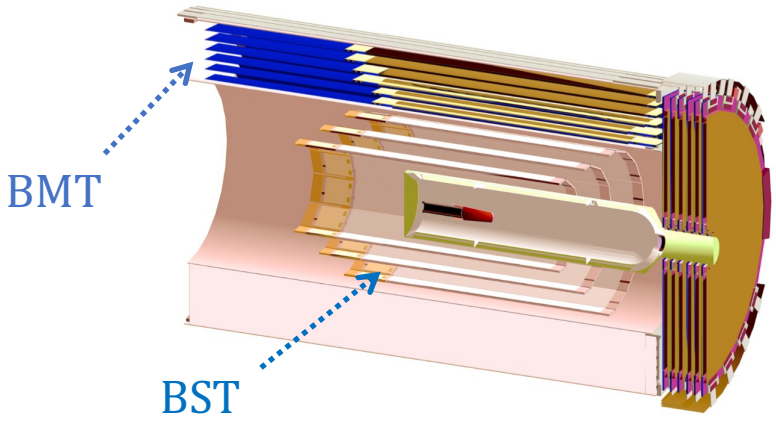
17 November 2026

Outline

1. The Central Detector of CLAS12
2. Motivations for using AI/ML methods to improve the tracking
3. Graph Neural Networks architectures
4. First results of denoising at the hit and track levels
5. Conclusion and outlook

Tracking in the CD





Central Tracking :

- Detectors : BST + BMT.
- Charged particles tracking like protons, π^{+-}, K^{+-}.
- Angular coverage : $35° < \theta < 125°$.
- Typical momentum range : $0.2 < P < 1.5$ GeV.
- 5T solenoid magnetic field.

- BST : 6 sillicons layers arranged in 3 stereo pairs.
- BMT : 6 Micromegas layers: Some strips along the beam line and some in circular arc. (CZZCZC).

Motivation for using IA/ML methods to improve tracking

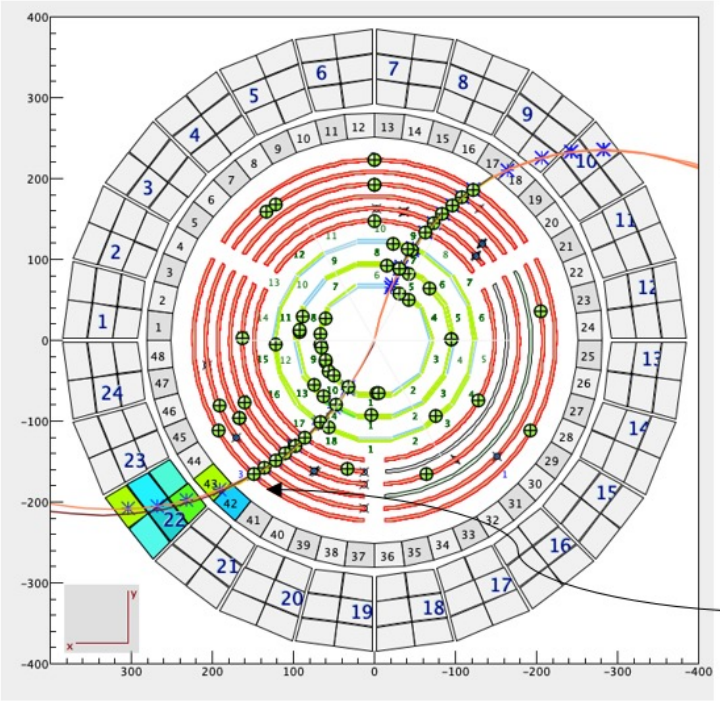
To reconstruct the particle tracks in the CD, hits in the BST and BMT are used.

But there is real hits which truely come from real particle and background hits. Background hits cause two main problems in track reconstruction :

- They **contaminate clusters**, degrade track reconstruction and can lead to lost tracks.
- They can also create **fake tracks**.

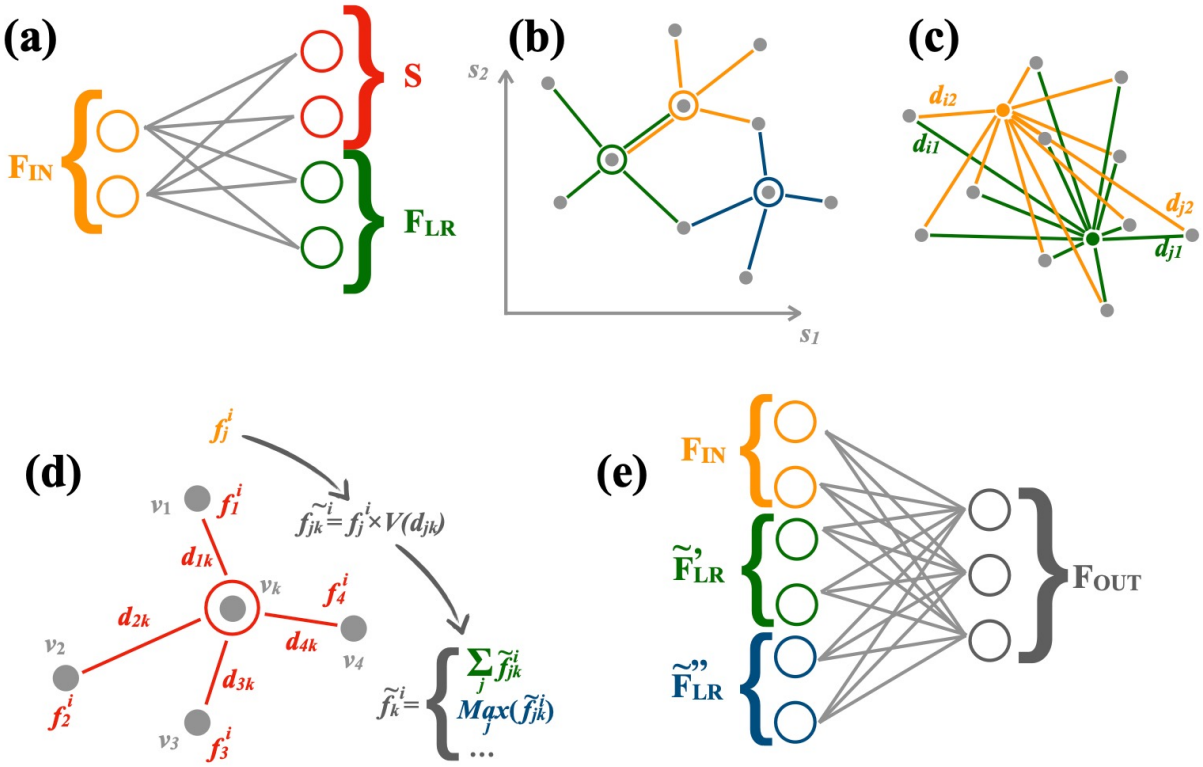
→ First **denoise** at hit level using ML methods to remove background and keep signal. (this talk)

→ Fit tracks in better conditions. (not in this talk)



Graph Neural Networks (GNNs) :

- GNNs are useful for this task. **Each hit is represented as a node**, and the network uses the relationships between hits to classify them.
- **GravNet** gathers information from neighboring nodes.
- **GarNet** uses aggregator nodes.



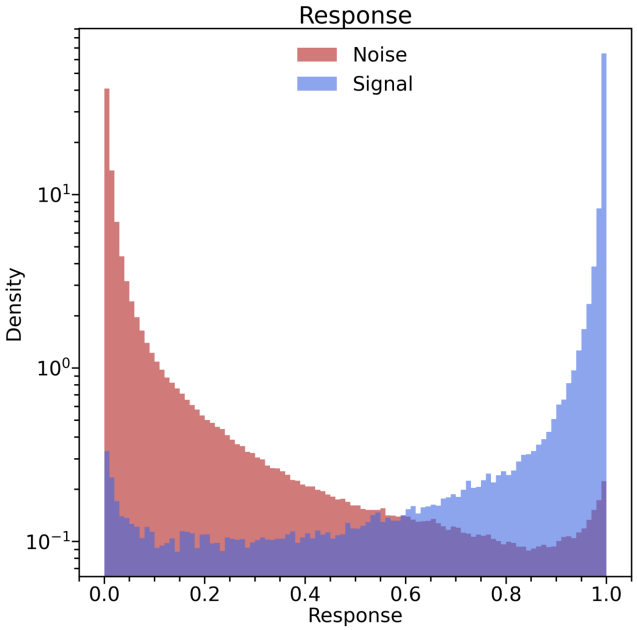
Ref : Qasim, S. R., Kieseler, J., Iiyama, Y., & Pierini, M. (2019). Learning representations of irregular particle-detector geometry with distance-weighted graph networks. arXiv. https://doi.org/10.48550/ARXIV.1902.07987

Training samples like :

- Single-muon events
- DVCS and DIS events
- **Flat generator** (p, π^{+-}, K^{+-}uniform in full angle and momemtum)

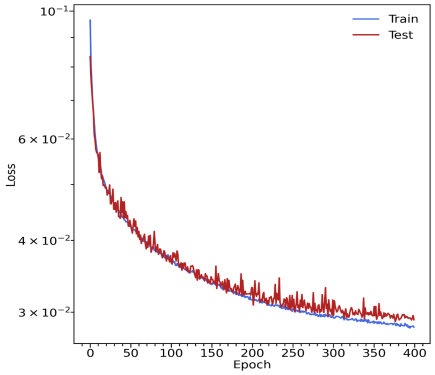
Background events are recorded using random triggers and merged with simulated events at the detector-response level, before reconstruction.



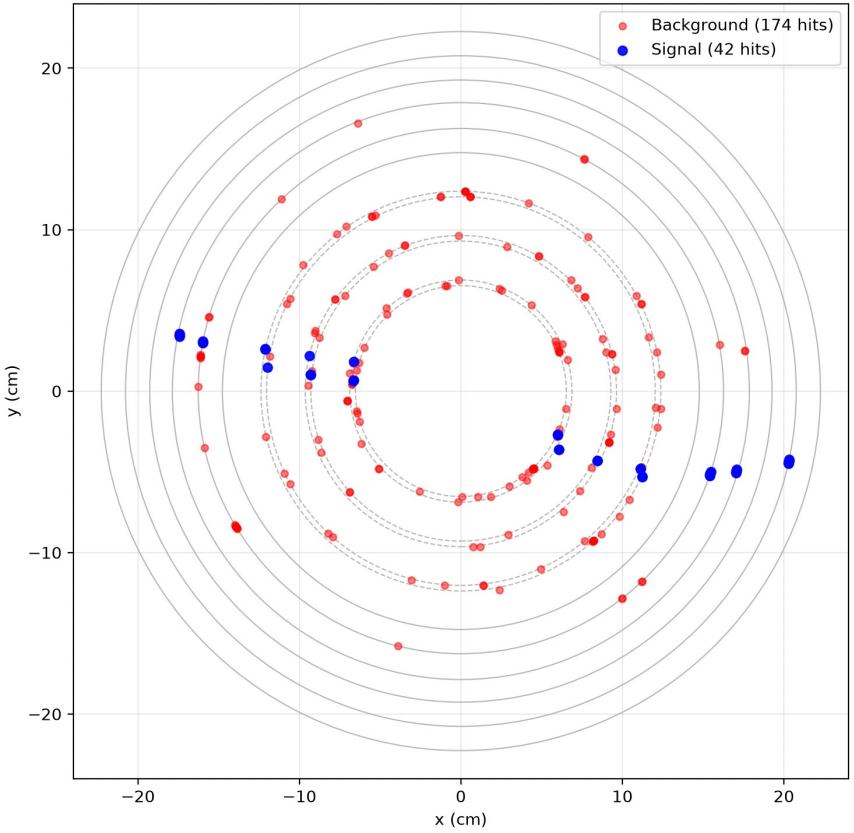


Inputs features :

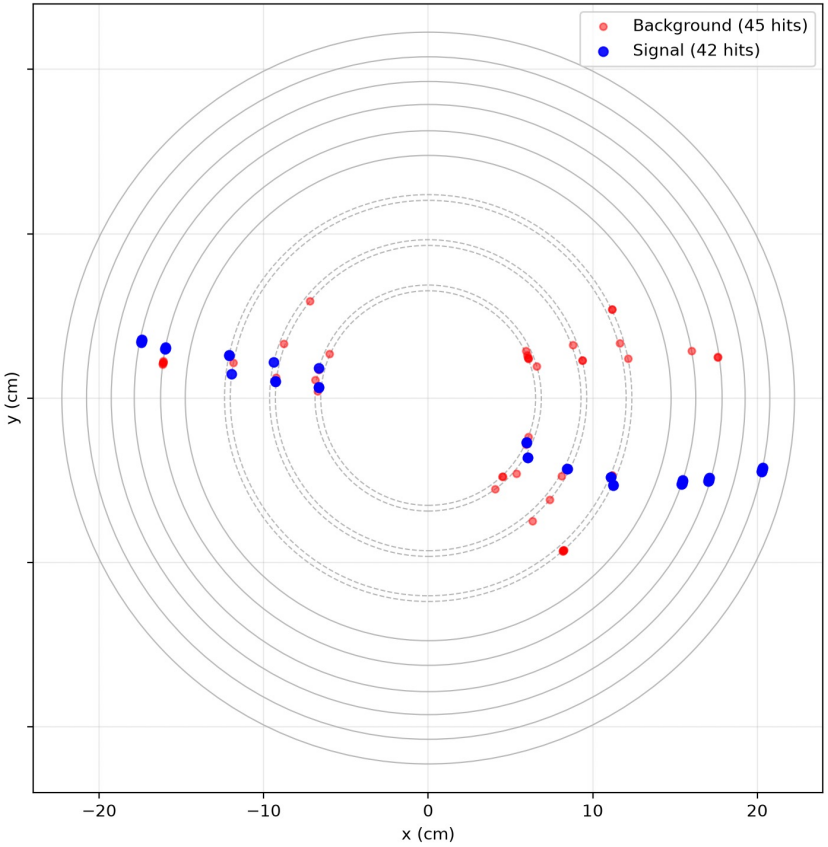
- Strip coordinates x1,x2 y1,y2 z1,z2 and strip ID
- Layers ID



Results at hits level

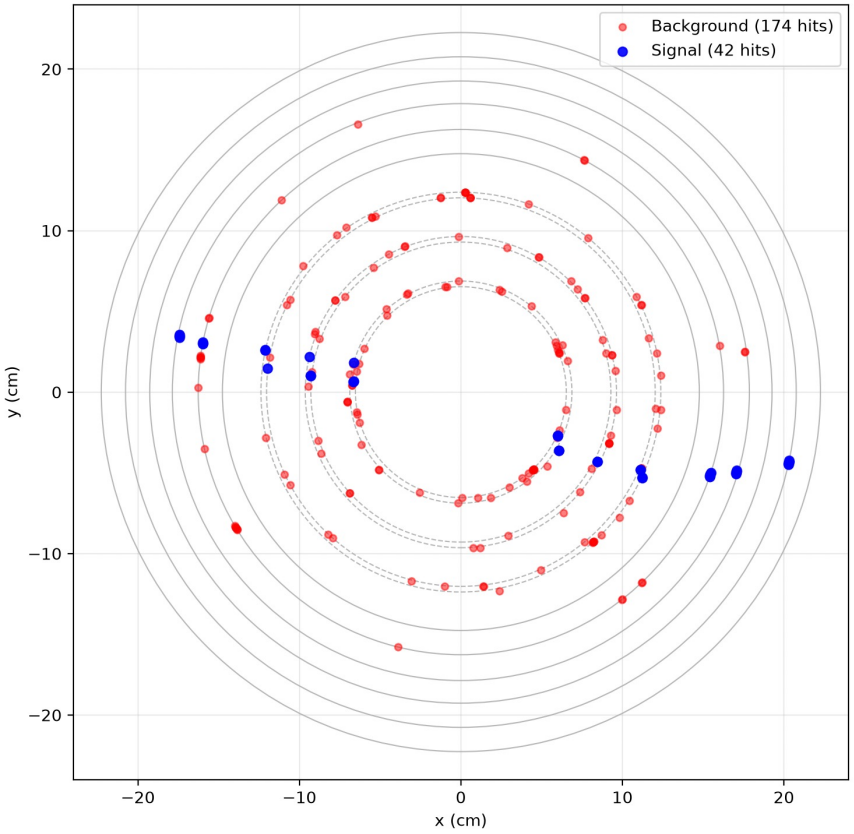


Hits projected on (X,Y) plan **before** denoising

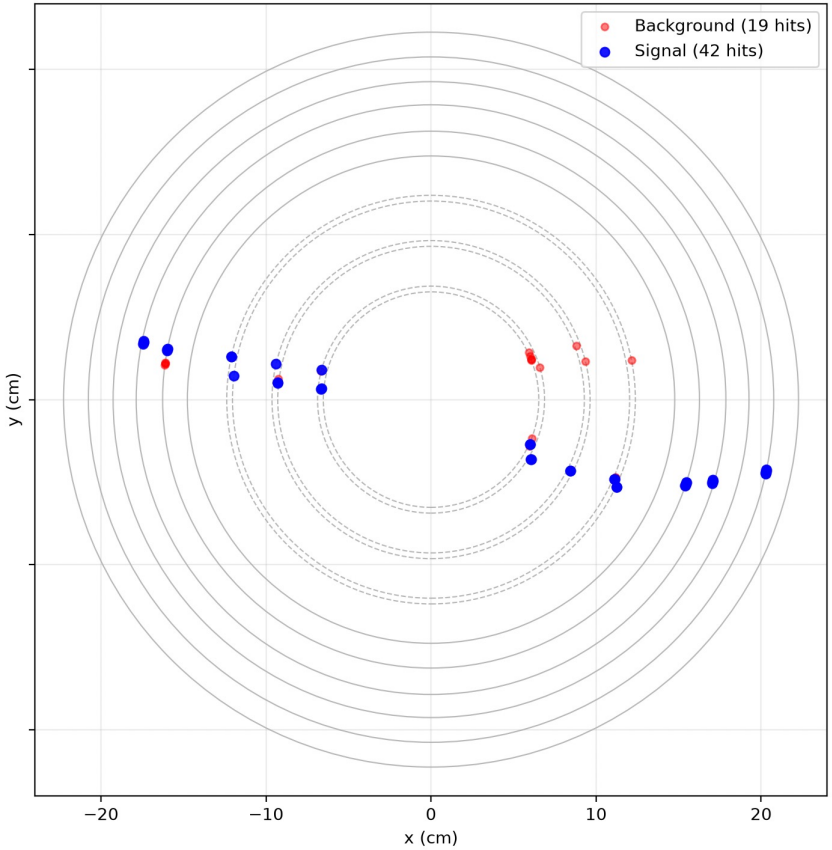


After denoising with **threshold fixed at 0.1**

Results at hits level

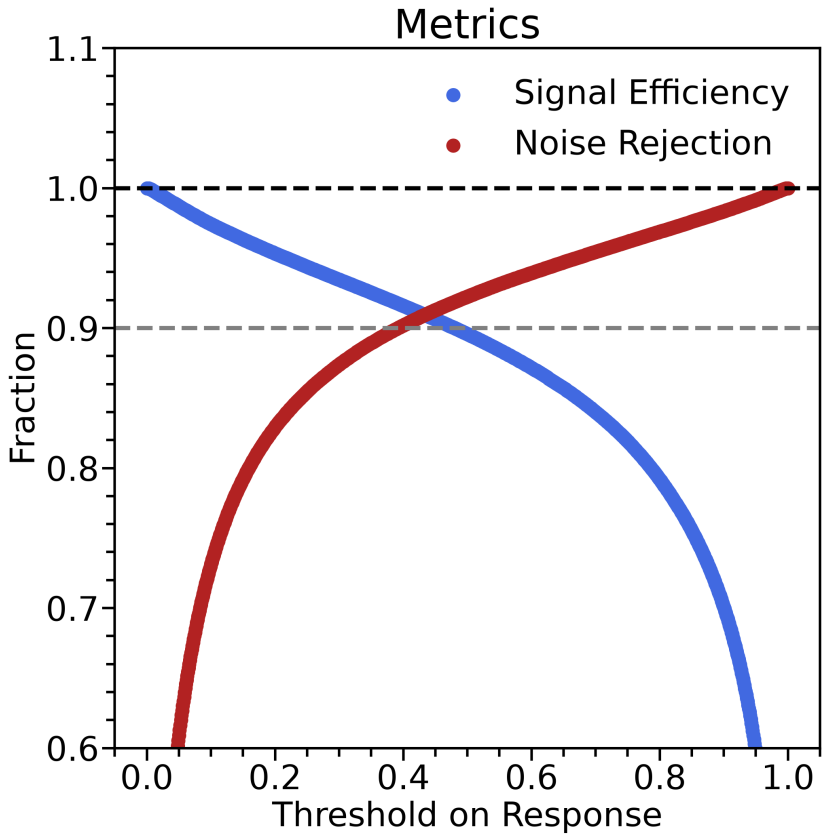


Hits projected on (X,Y) plan **before** denoising



After denoising with **threshold fixed at 0.3**

Results at hits level



$$\epsilon_{\text{signal}} = \frac{N_{\text{signal, kept}}}{N_{\text{signal, total}}}$$

$$R_{\text{background}} = 1 - \frac{N_{\text{background, kept}}}{N_{\text{background, total}}}$$

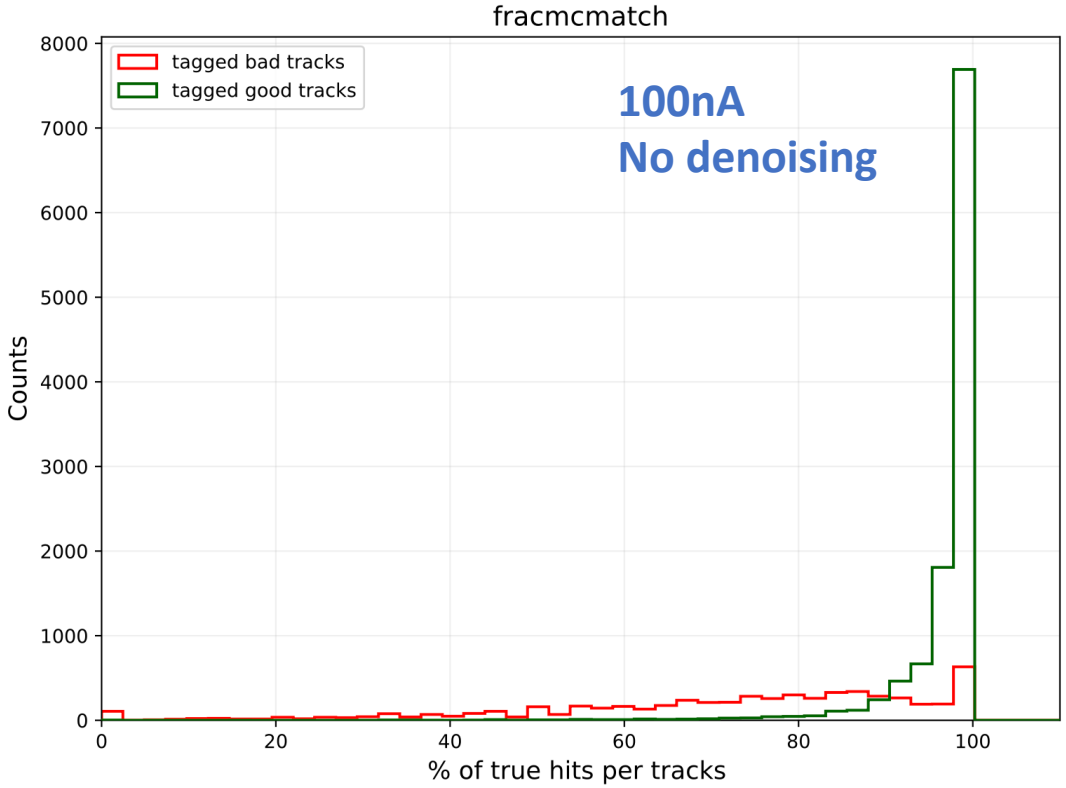
Strategy :

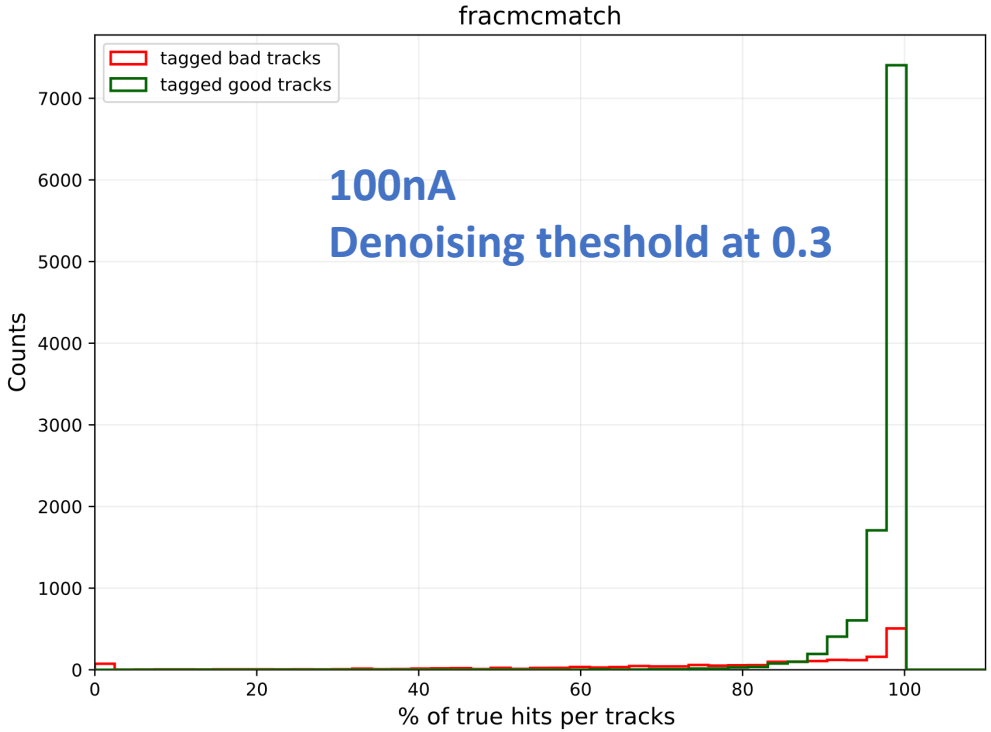
- Use again the flat generator to create sample for tests (so test on MC with models train on MC).
- Tag good tracks and bad tracks with cuts.
- Apply different threshold and look the effect of denoising.

Cut to select good tracks :

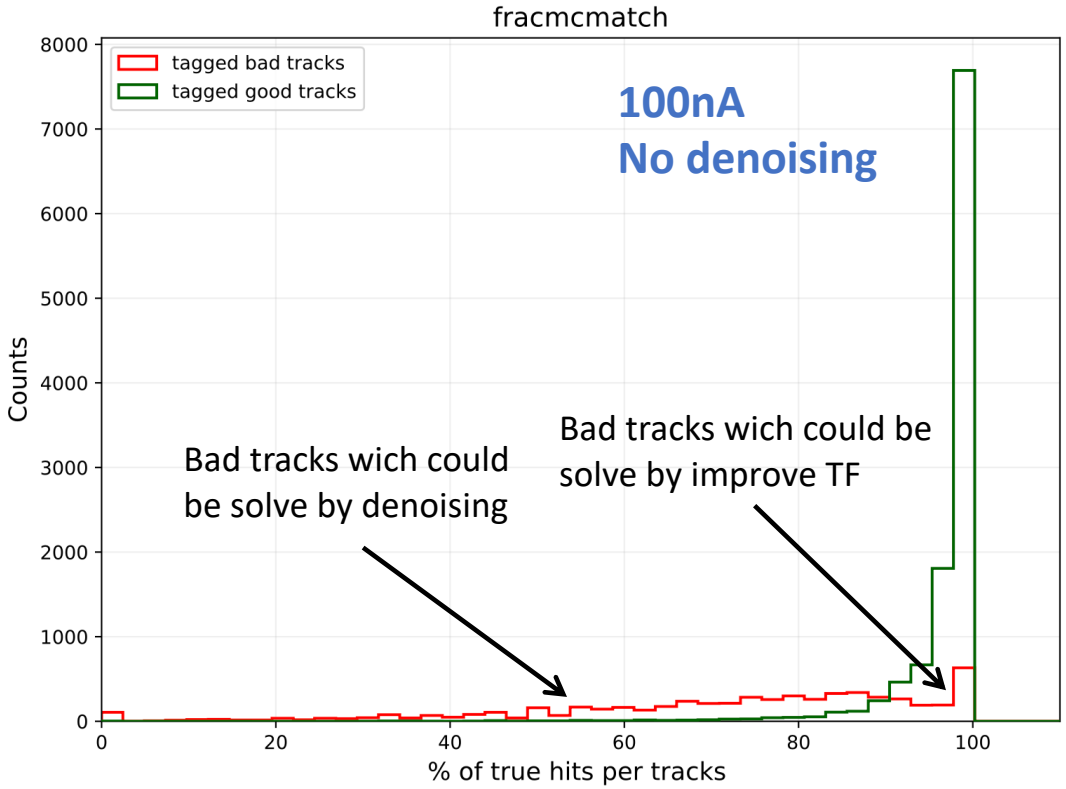
- Full angle REC/GEN $< 7°$
- $\Delta P/P < 20\ \%$
- Same charge
- GEN associated send in the CD

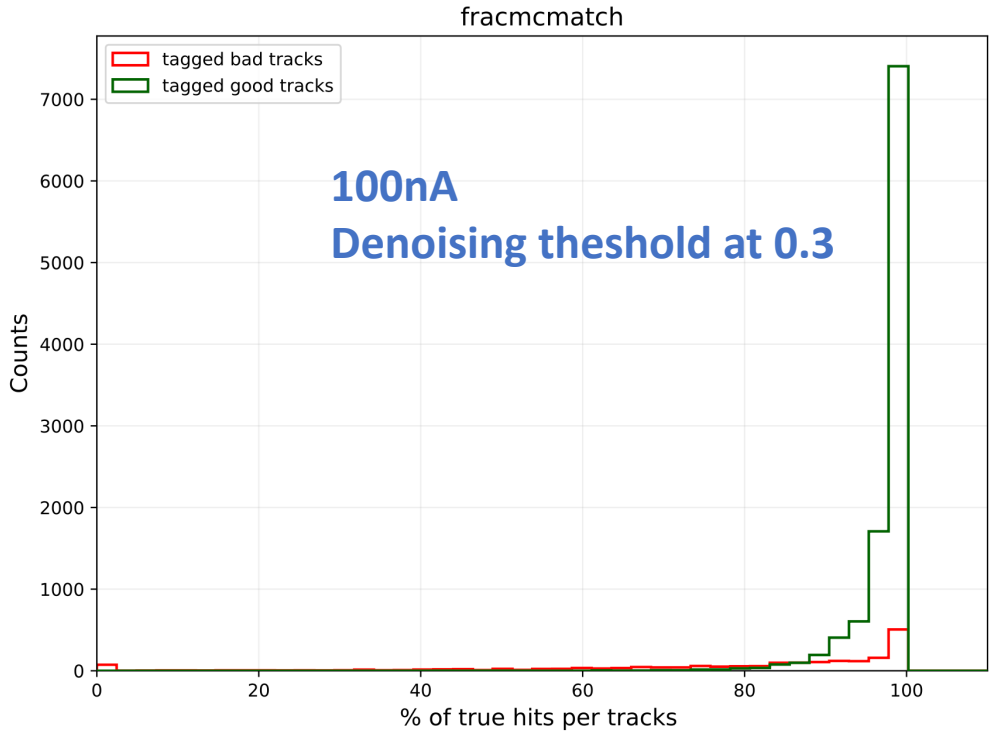
Results at tracks level



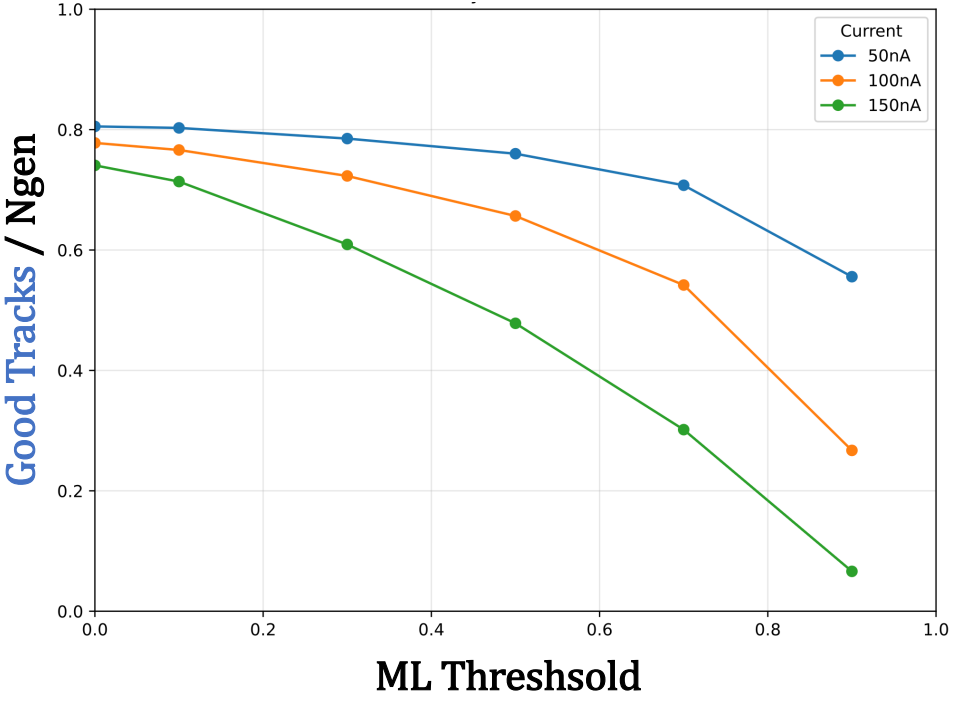


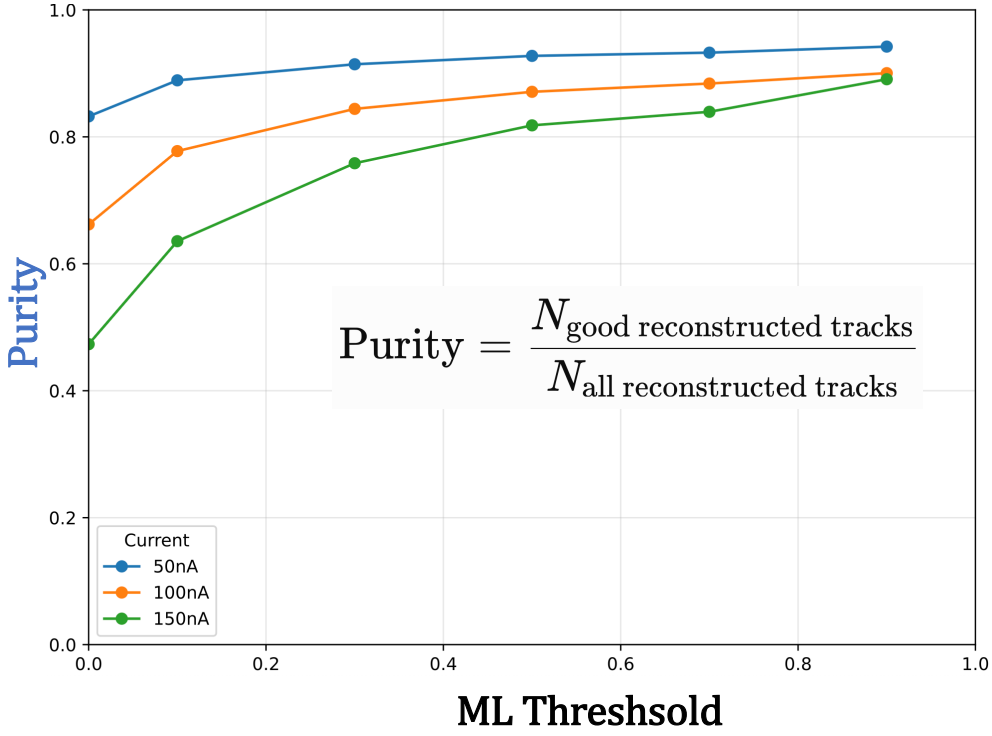
Results at tracks level





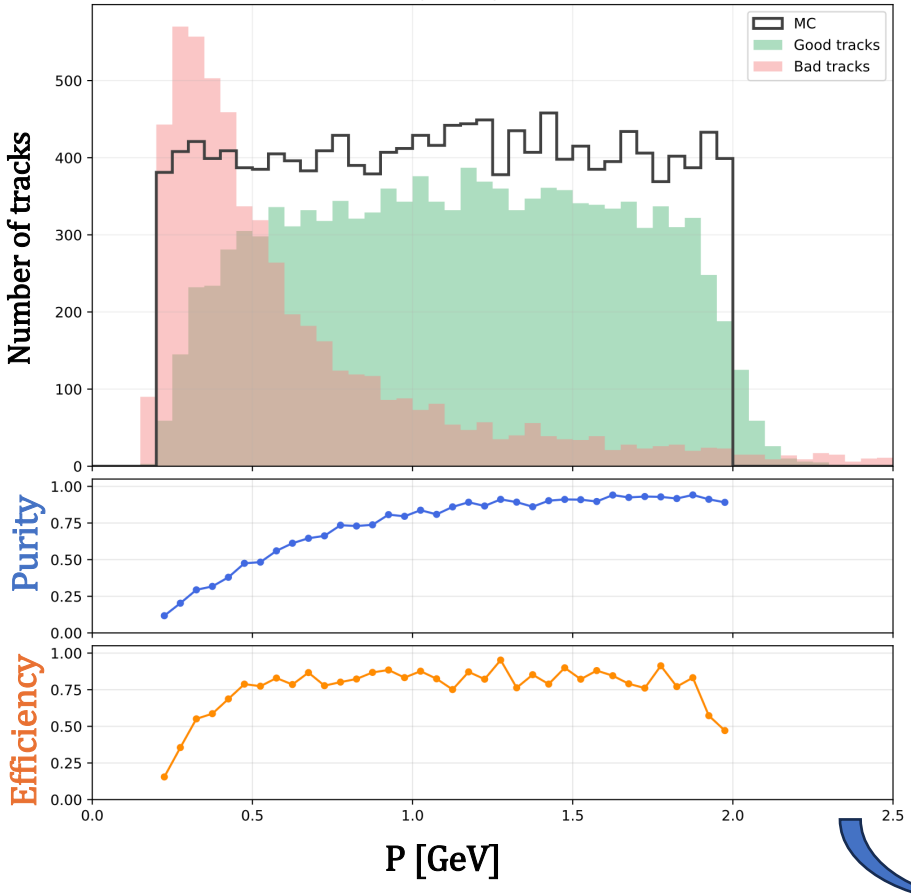
Results at tracks level

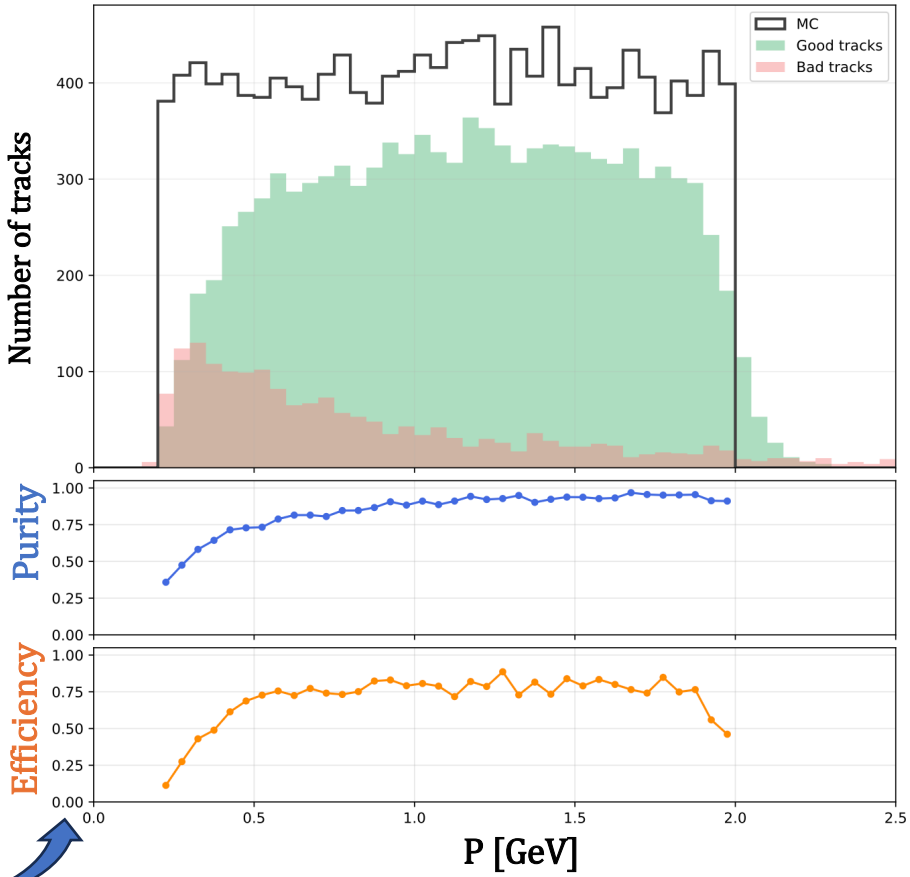




Results at tracks level

Current :100nA

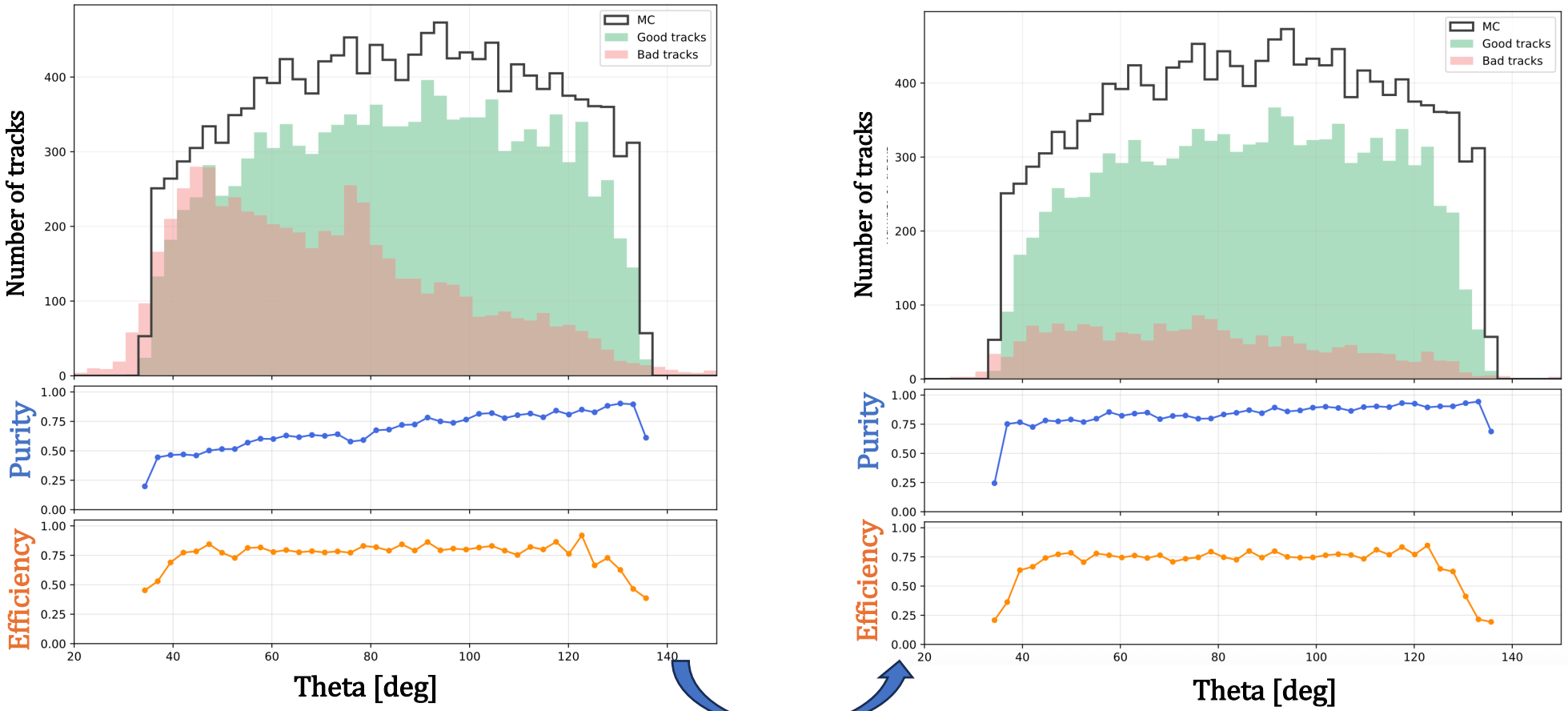




Denoising at threshold 0.3

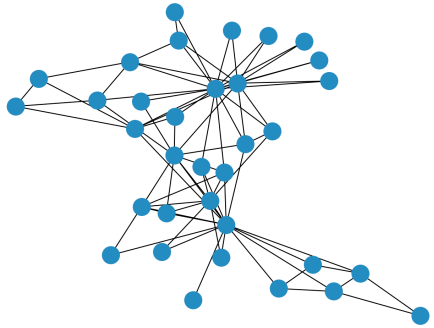
Results at tracks level

Current :100nA



Denoising at threshold 0.3

Work already done :



- Tests with differents data sample for training/testing.
- Tests with differents GNNs/Hyperparameters.
- Proof of principle of the denoising at hits level and tracks levels with differents currents.

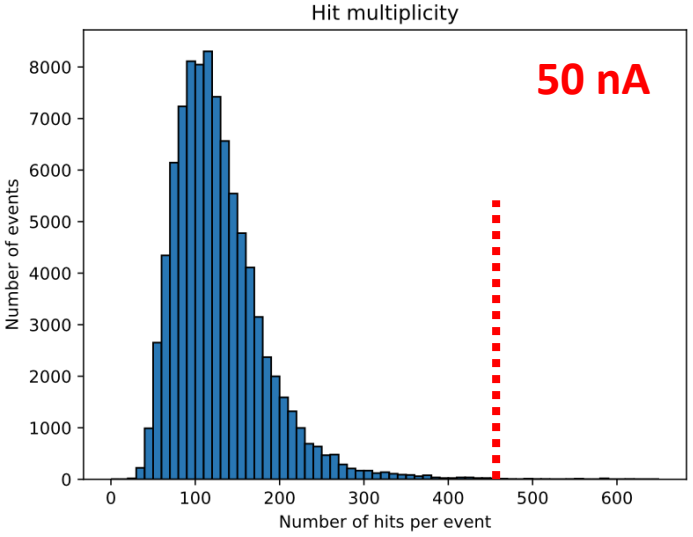
Future work planned :

- We can still improve the model (size of data sample for training, better choice of hyperparametters).
- Validate the method on experimental data using known reactions and evaluate tracking efficiency.

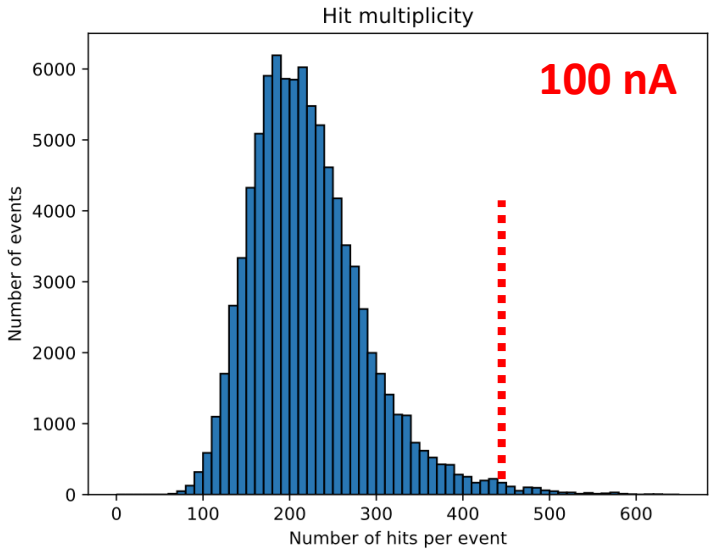
Backup slides

Cut events with number of hits > 450:

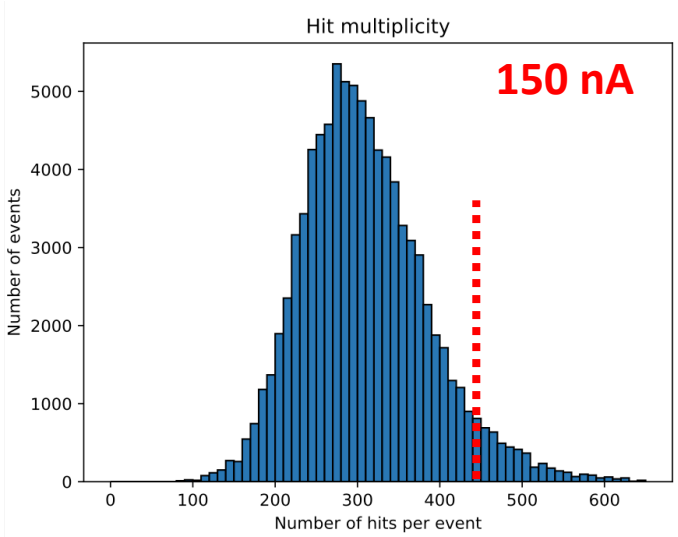
It's plan to increase the size of the model in the future!



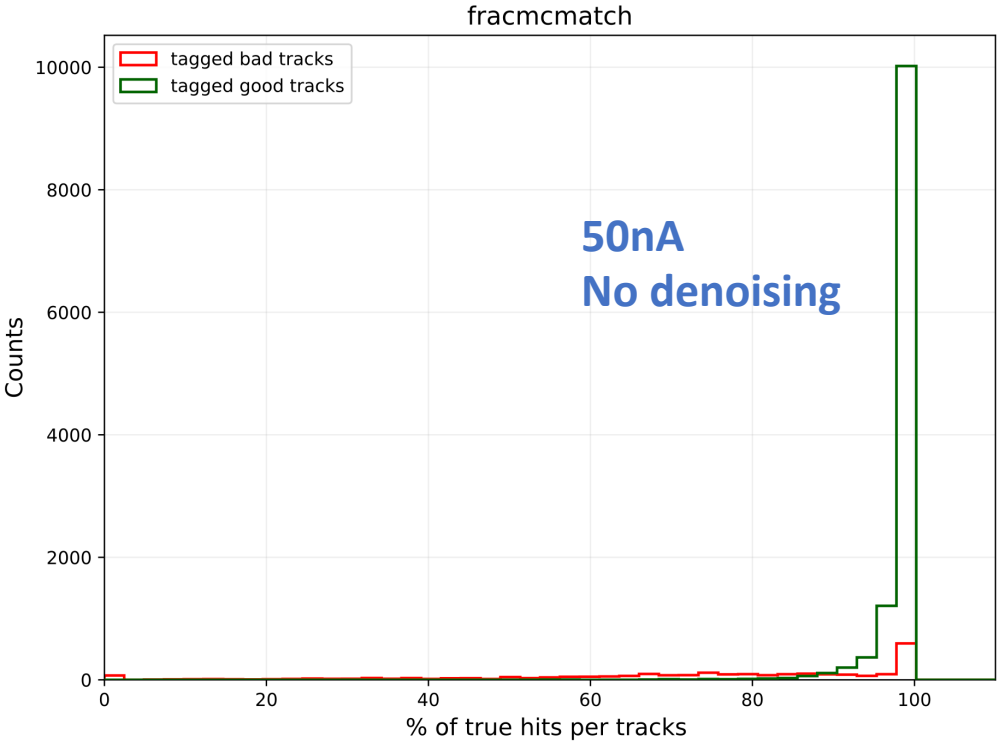
Negligeable cut

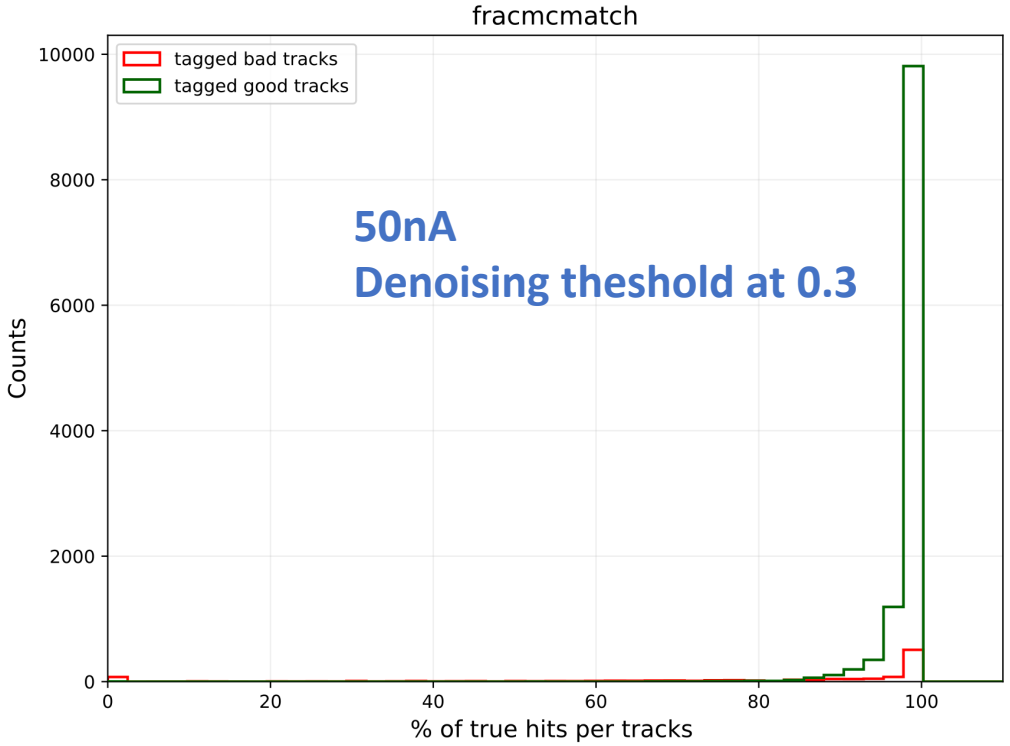


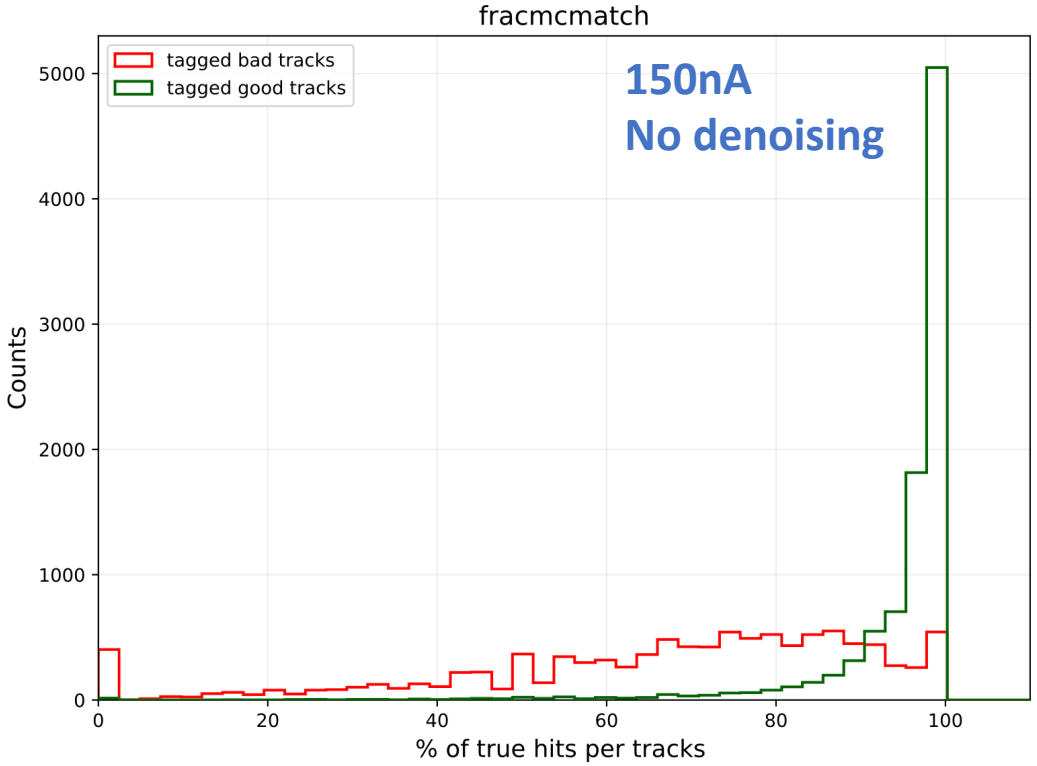
Remove ~ 2 %

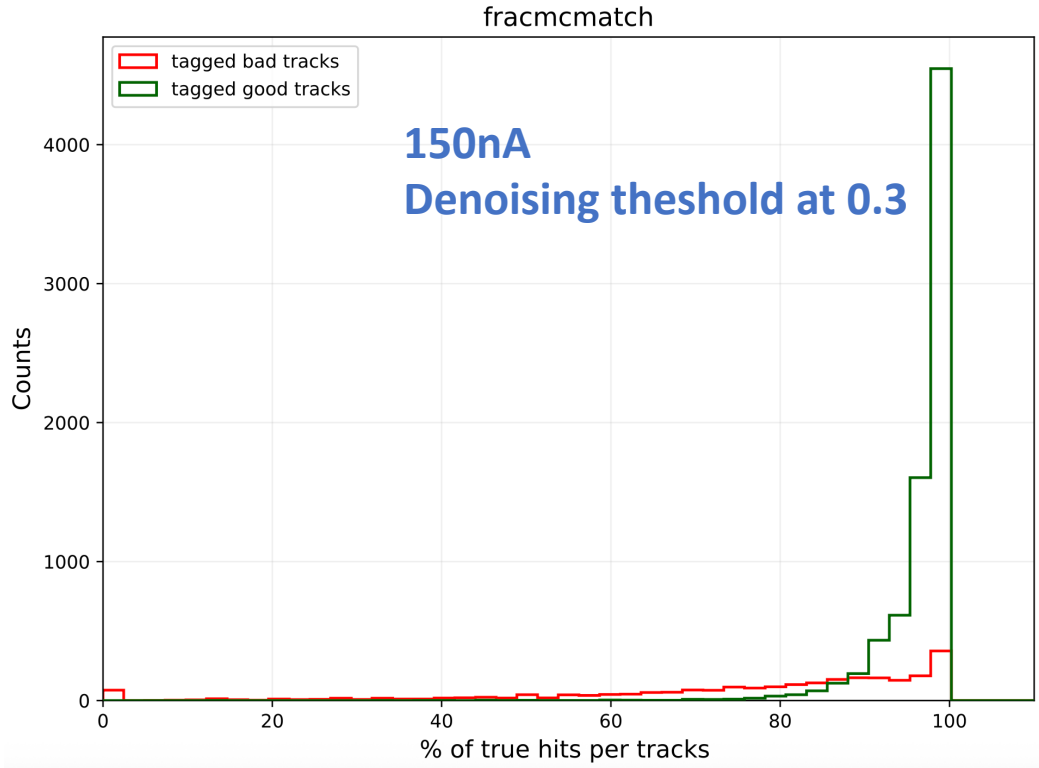


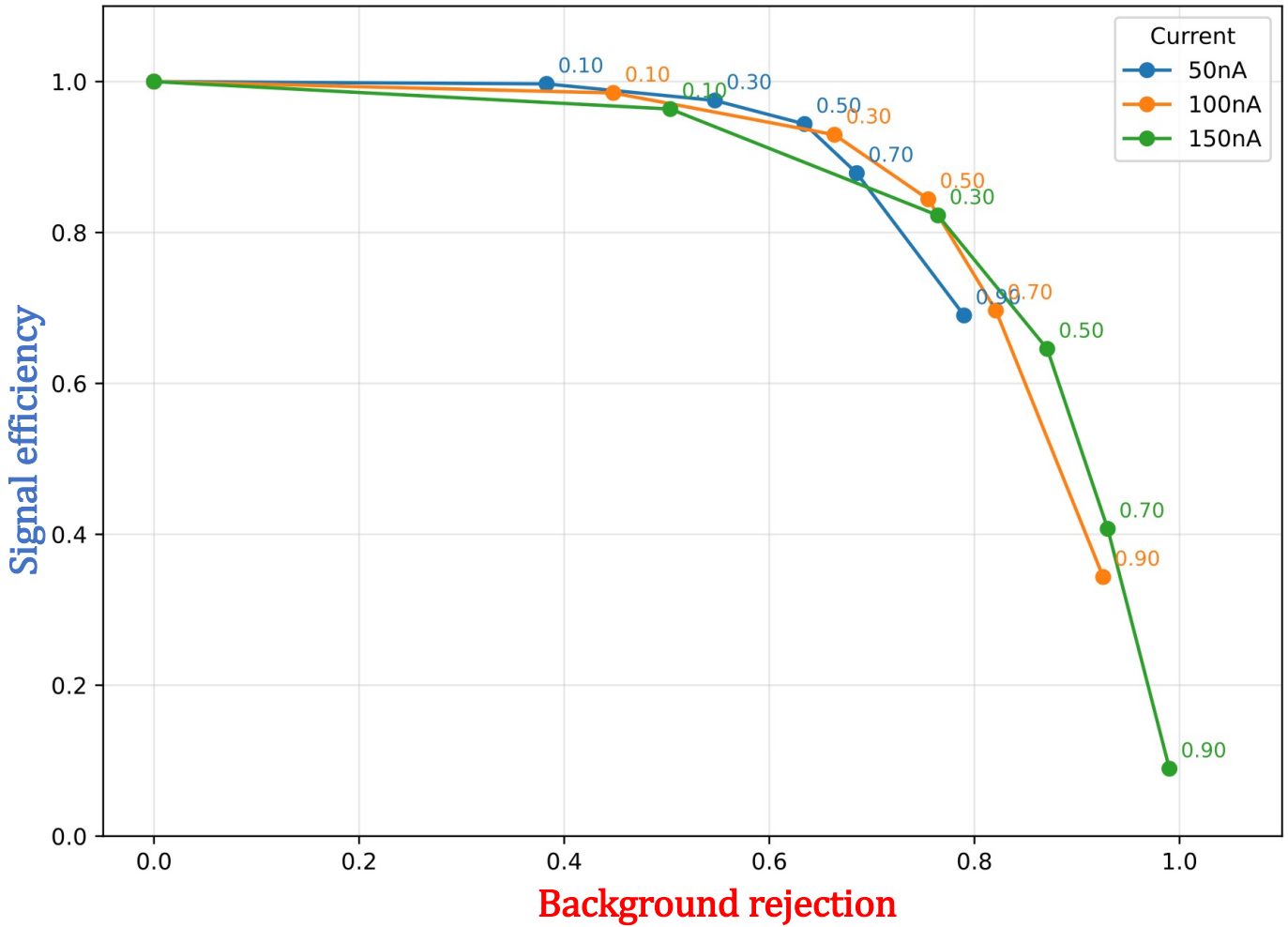
Remove ~ 13 %


fracmcmatch
tagged bad tracks
tagged good tracks
50nA
No denoising
Counts
% of true hits per tracks
0
2000
4000
6000
8000
10000
0
20
40
60
80
100


fracmcmatch
tagged bad tracks
tagged good tracks
50nA
Denoising theshold at 0.3
Counts
% of true hits per tracks
0
2000
4000
6000
8000
10000
0
20
40
60
80
100


fracmcmatch
tagged bad tracks
tagged good tracks
150nA
No denoising
Counts
% of true hits per tracks
0
1000
2000
3000
4000
5000
0
20
40
60
80
100


fracmcmatch
tagged bad tracks
tagged good tracks
150nA
Denoising theshold at 0.3
Counts
% of true hits per tracks
0
1000
2000
3000
4000
0
20
40
60
80
100

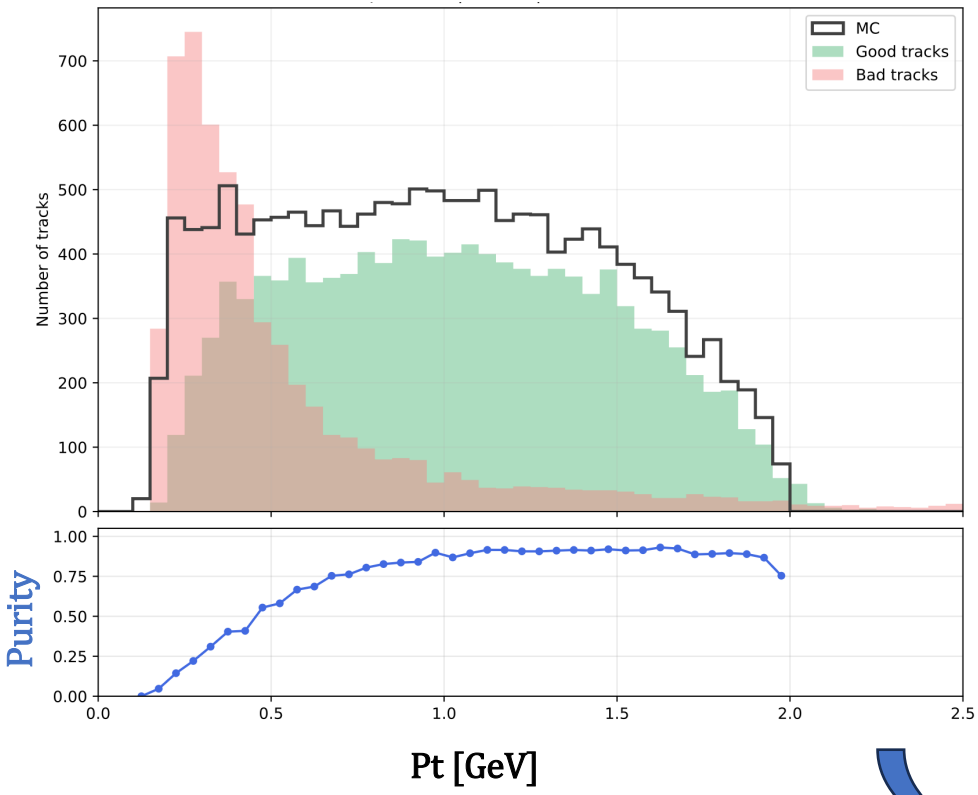


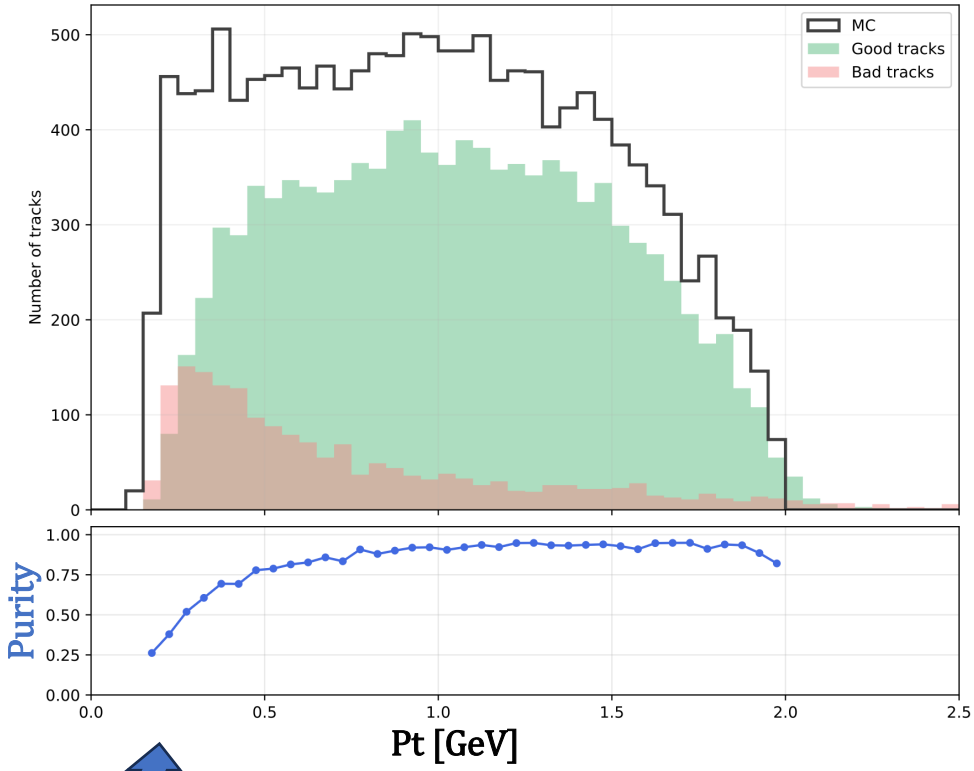
$$\epsilon_{signal} = \frac{N_{good\ tracks\ kept}}{N_{good\ tracks\ initialy}}$$

$$Bg_{rej} = 1 - \frac{N_{bad\ tracks\ kept}}{N_{bad\ tracks\ initialy}}$$

Results at tracks level

Current :100nA

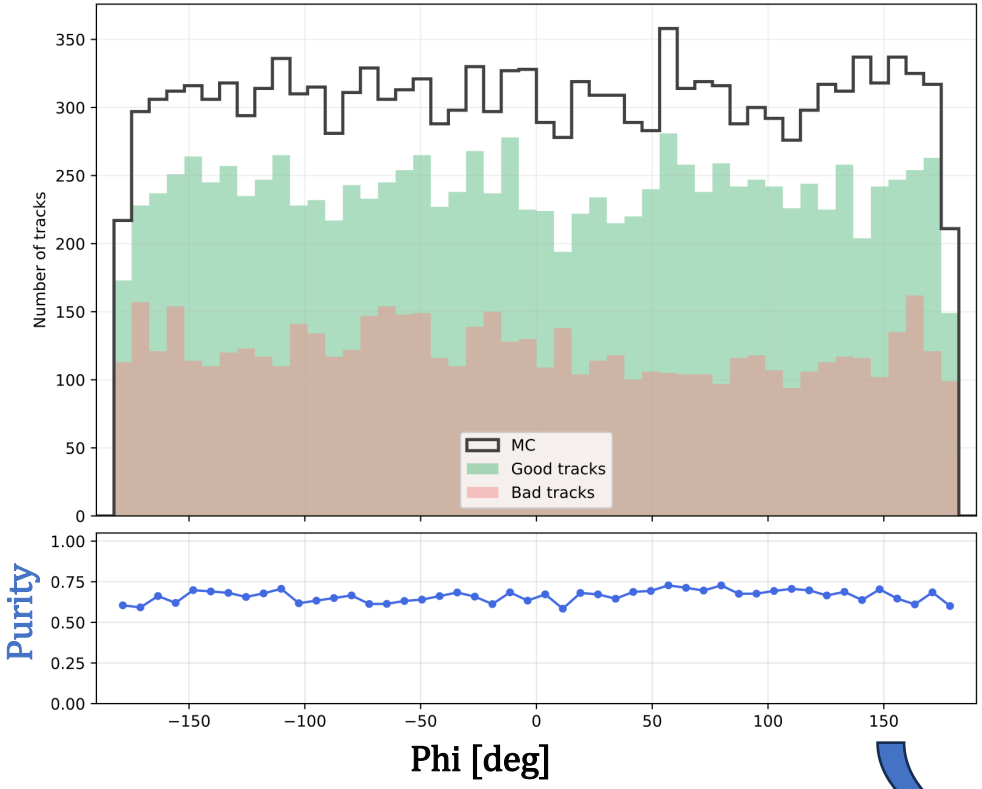


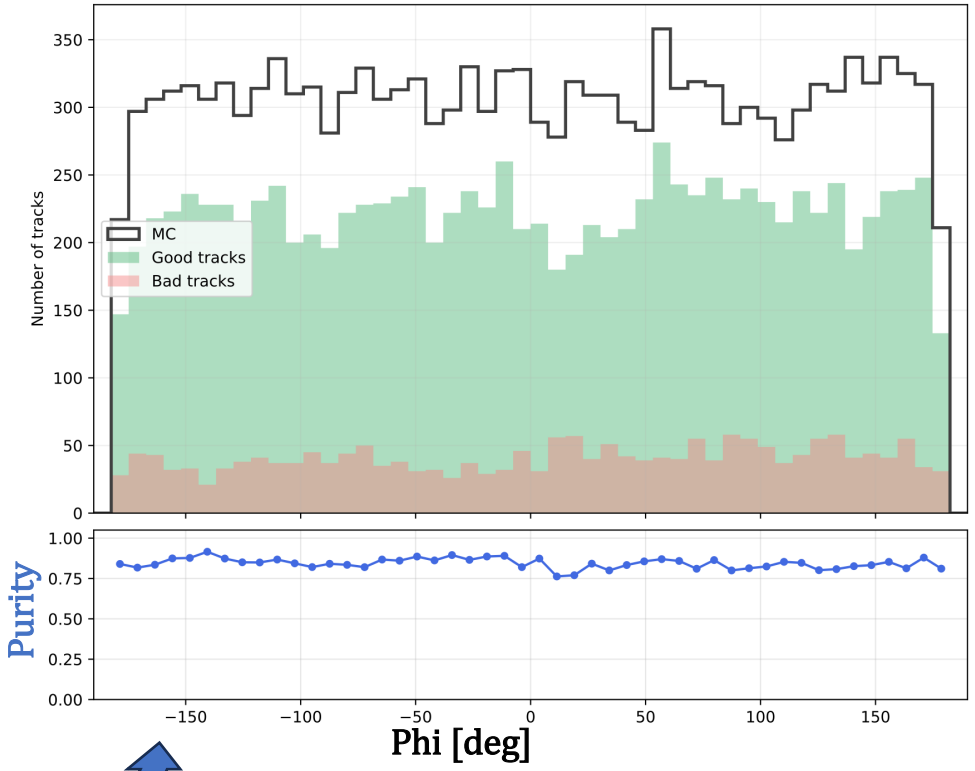


Denoising at threshold 0.3

Results at tracks level

Current :100nA





Denoising at threshold 0.3