Nuclei identification using deep learning for AMS Tracker

Zhen Liu

INFN-Roma2

AMS-Italy in Bologna, 28 November 2024

Introduction

- In many experiments, deep learning has shown it can go beyond traditional methods, enabling the full exploitation of detector capabilities.

- In AMS, deep learning can be applied to many tasks, such as nuclei and charge identification, background estimation, charge reconstruction, rigidity reconstruction, unfolding, and more.

- In this study, we take a nuclei identification task for Silicon as an attempt to explore the application of deep learning to the AMS tracker detector using MC, providing an initial look at its potential.

- A future study using Data will build on this to address discrepancy between Data and MC.

Samples and Task

Samples:

- L1Inner MC: Mg, Al, Si, P, S
- from NAIA Version v1.1.1

Input variables:

- Tracker ADC (with Beta correction) and seed strip address:
  - the seed ADC of 10 clusters for each layer and side (**in Edep descending order**)
  - the 2 adjacent strip (left and right) ADC of the first 3 clusters for each layer and side
- TOF unbiased Beta ($\mathbf{1/\beta\hat{}2}$)

Task : Signal background discrimination (with MC samples only for now)

- **Signal**: primary Si survived below L2
- **Background 1**: primary Si survived above L2
- **Background 2**: other nuclei (Mg, Al, P, S)

Event selection

Event selection for Mg, Al, Si, P, S (L1Inner selection without tracker charge and fiducial volume cuts):

- Upper TOF selection
- TrTrack selection
- InnerNHitY>=5 && L2&(L3|L4)&(L5|L6)&(L7|L8)
- InnerNormChisqY < 10
- L1InnerNormChisqY < 10 && L1InnerChisqY-InnerChisqY < 10(for L1Inner analysis)

\+ **Primary particle survived on L1 (using MC truth information)**

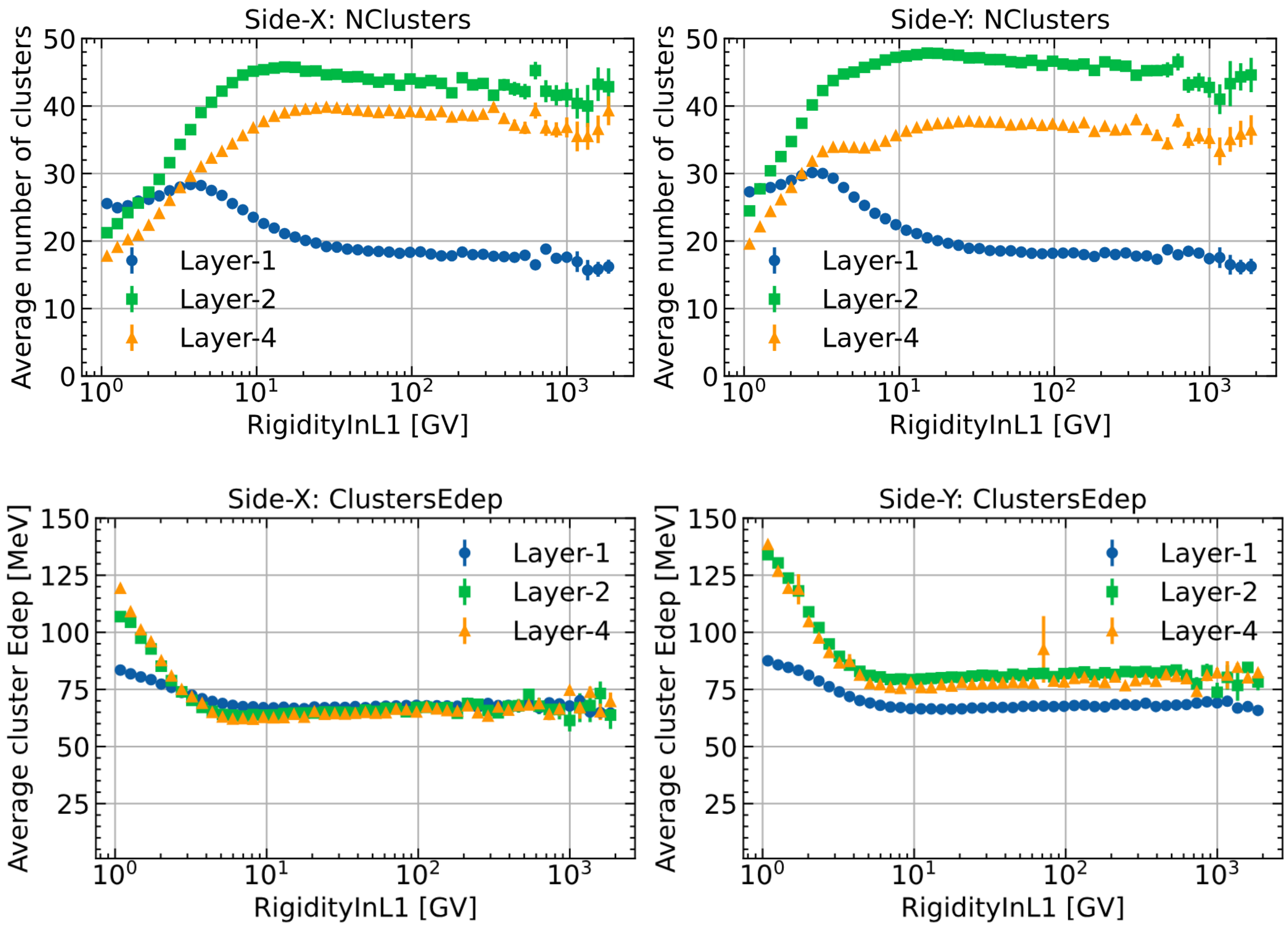
Selected samples (300k events for each categories):

- Signal: primary Si survived below L2
- Background 1: primary Si survived above L2
- Background 2: background nuclei (Mg, Al, P, S = 0.1 : 1 : 1 : 0.1)

Sample is divided into 80% for model training and 20% for testing

1. Check input variables

Clusters on each layer (Fe L1Inner selection with L1 and Inner Tracker Charge > 25.5)



(Plots from my slides about Event Size Assessment for Supporting L0 DAQ Scheme Development.)

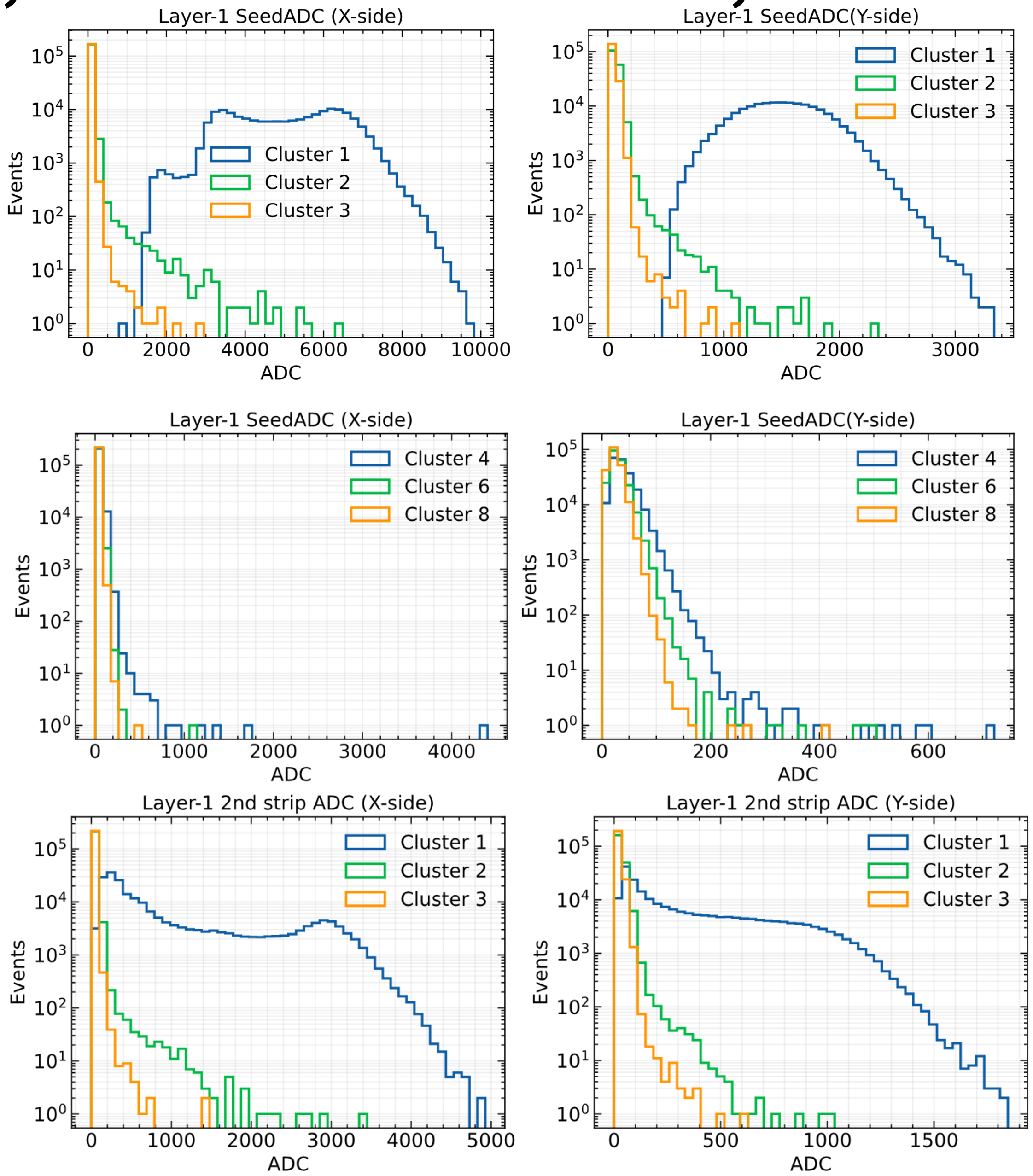
ADC Values (Sulfur MC, L1Inner selection, 200K events)

- **Clusters are arranged in Edep descending order.**
- **ADC: with Beta correction.**

☑ Seed strip of the max cluster has much higher ADC value than the others.
☑ For max ADC value: 2nd strip of cluster-1 ≈ seed strip of cluster-2
☑ Similar for the other layers (in backup slides)

For deep learning input:

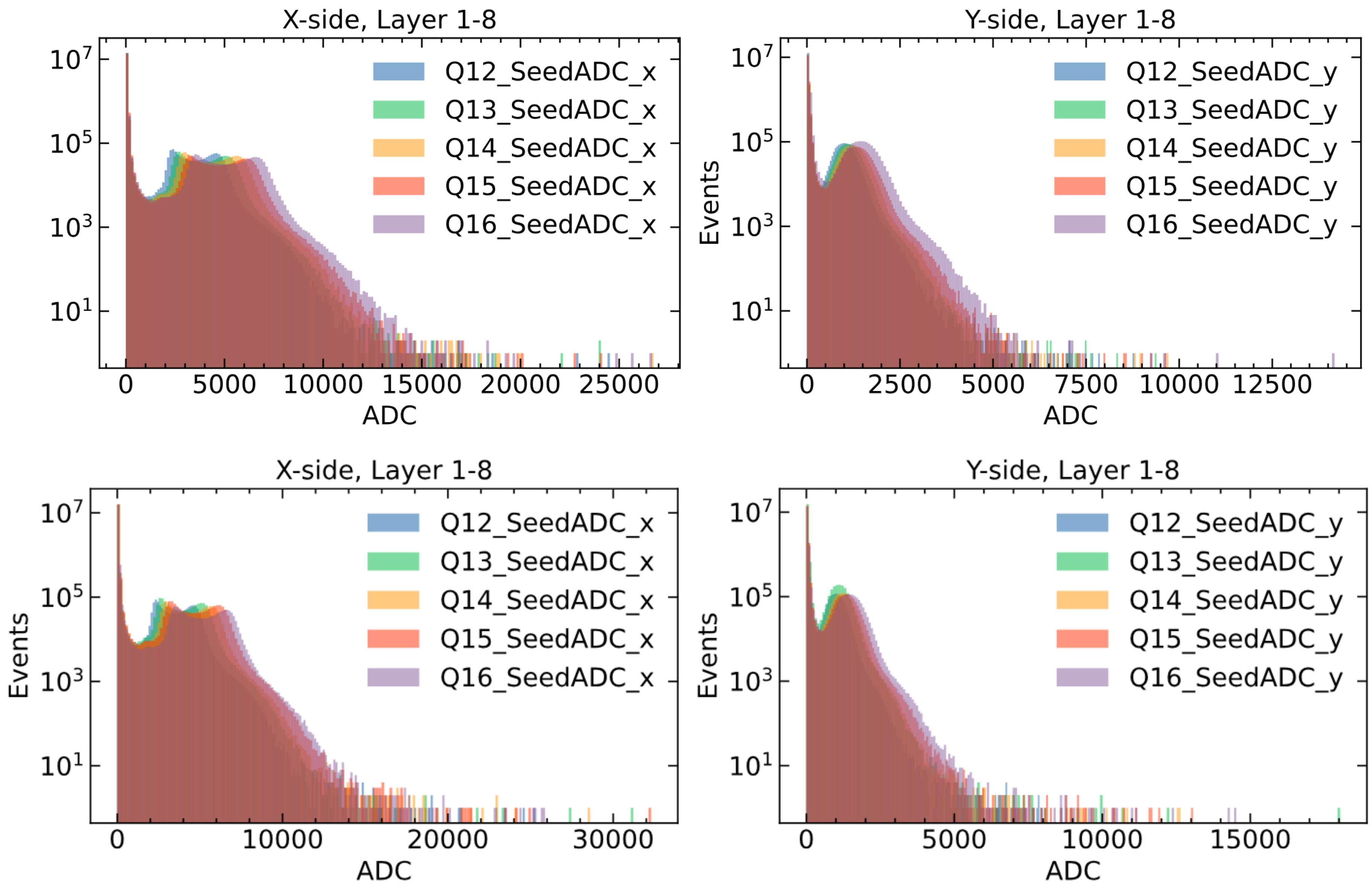
- the seed ADC of the first 10 clusters for each layer and side
- the 2 adjacent strip (left and right) ADC of the first 3 clusters for each layer and side



ADC distributions for Al, Mg, Si, P, S (MC, L1Inner selection)

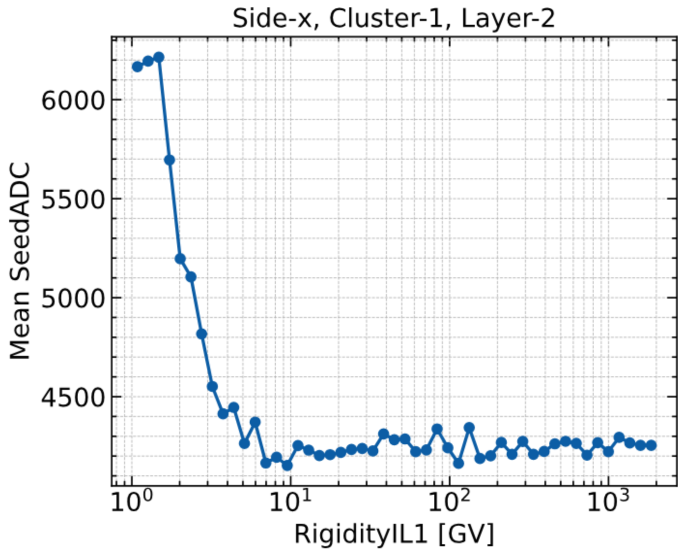
ADC without correction

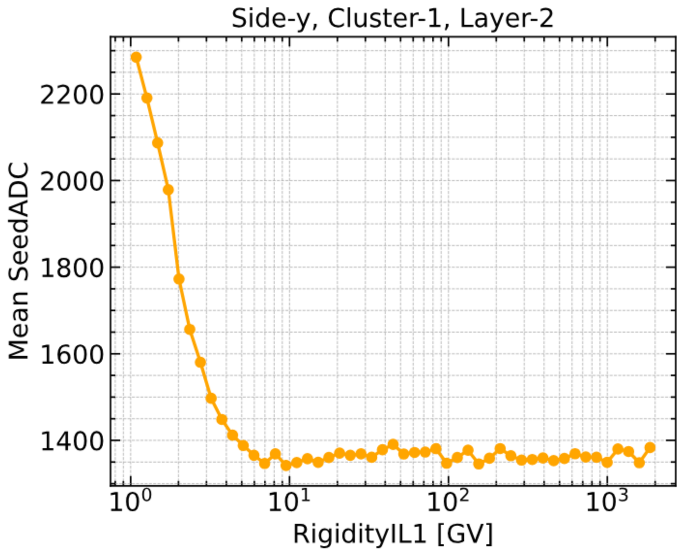
ADC with Beta correction (kBeta from AMS software)



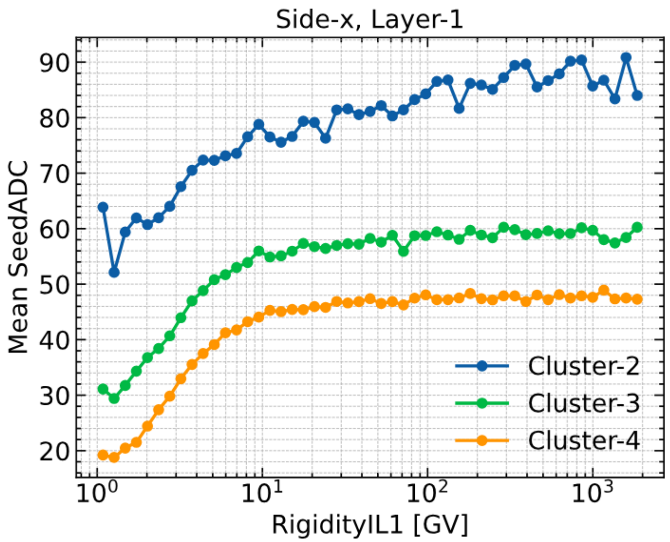
Rigidity dependence of ADC with Beta correction (Sulfur MC, L1Inner selection, 200K events)

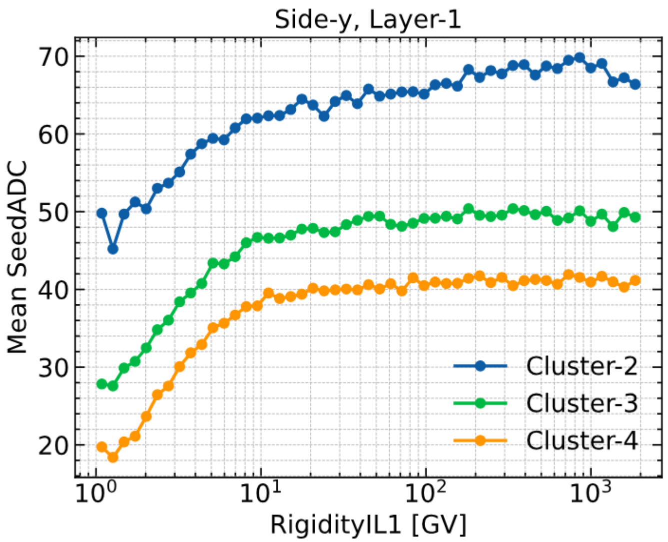
Mean seed ADC of max Edep clusters:





Mean seed ADC of other clusters:





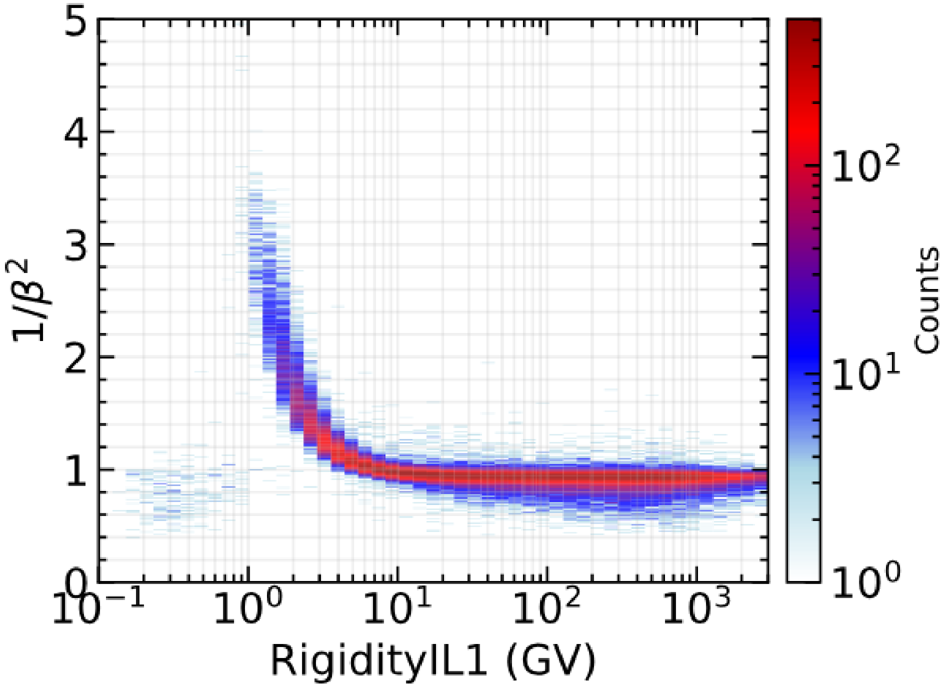
Further Beta correction is needed.
→ Add TOF Beta information to the deep learning models

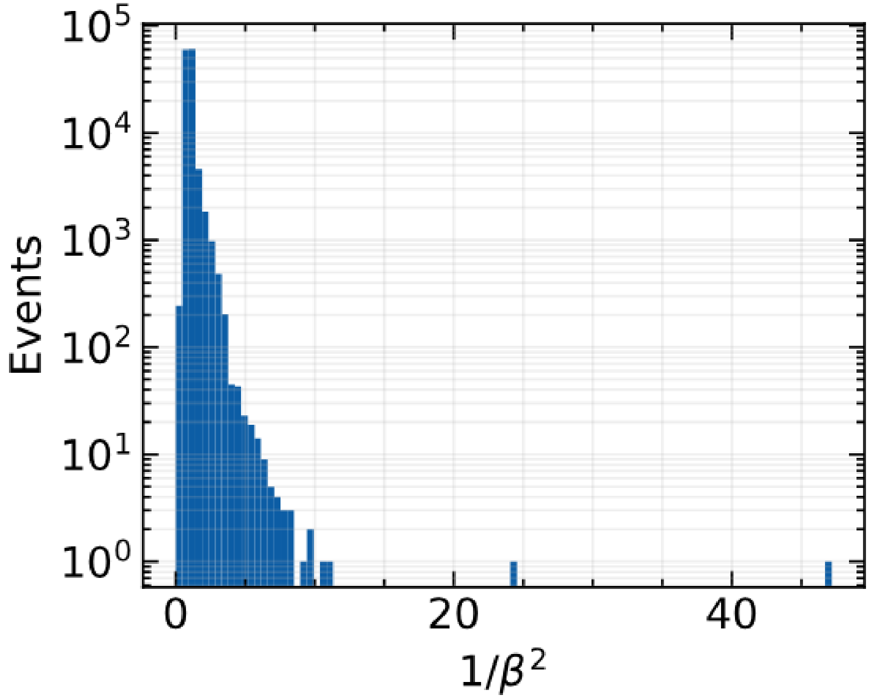
(Statistical error not added)
(Similar plots for layer-6 in Backup slides)

Unbiased TOF Beta (Sulfur MC, L1Inner selection, 200K events)

$\frac{1}{\beta^2}$, standalone reconstruction:

(Clamp Beta values below 0.1 to 0.1)





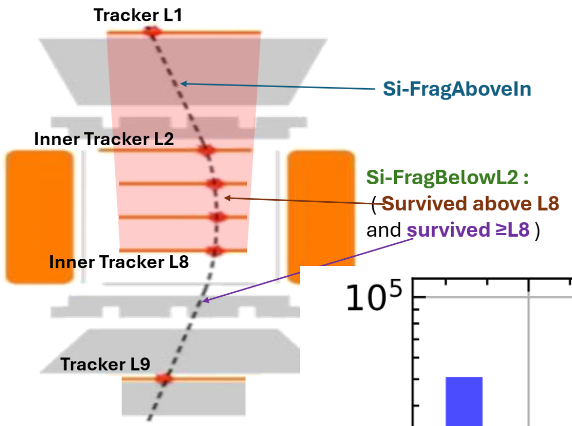
Besides Tracker hit cluster information,

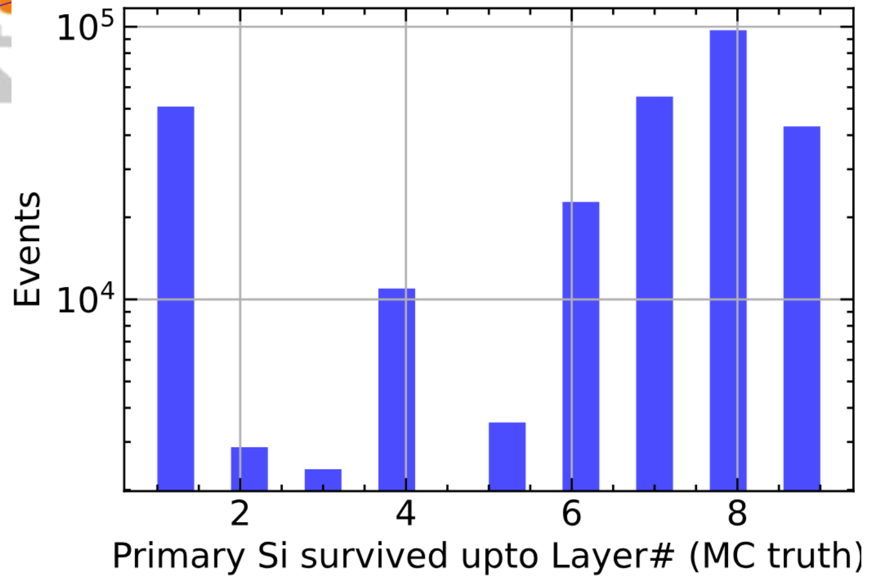
→ Add the unbiased TOF Beta information ($\frac{1}{\beta^2}$) to the deep learning model.

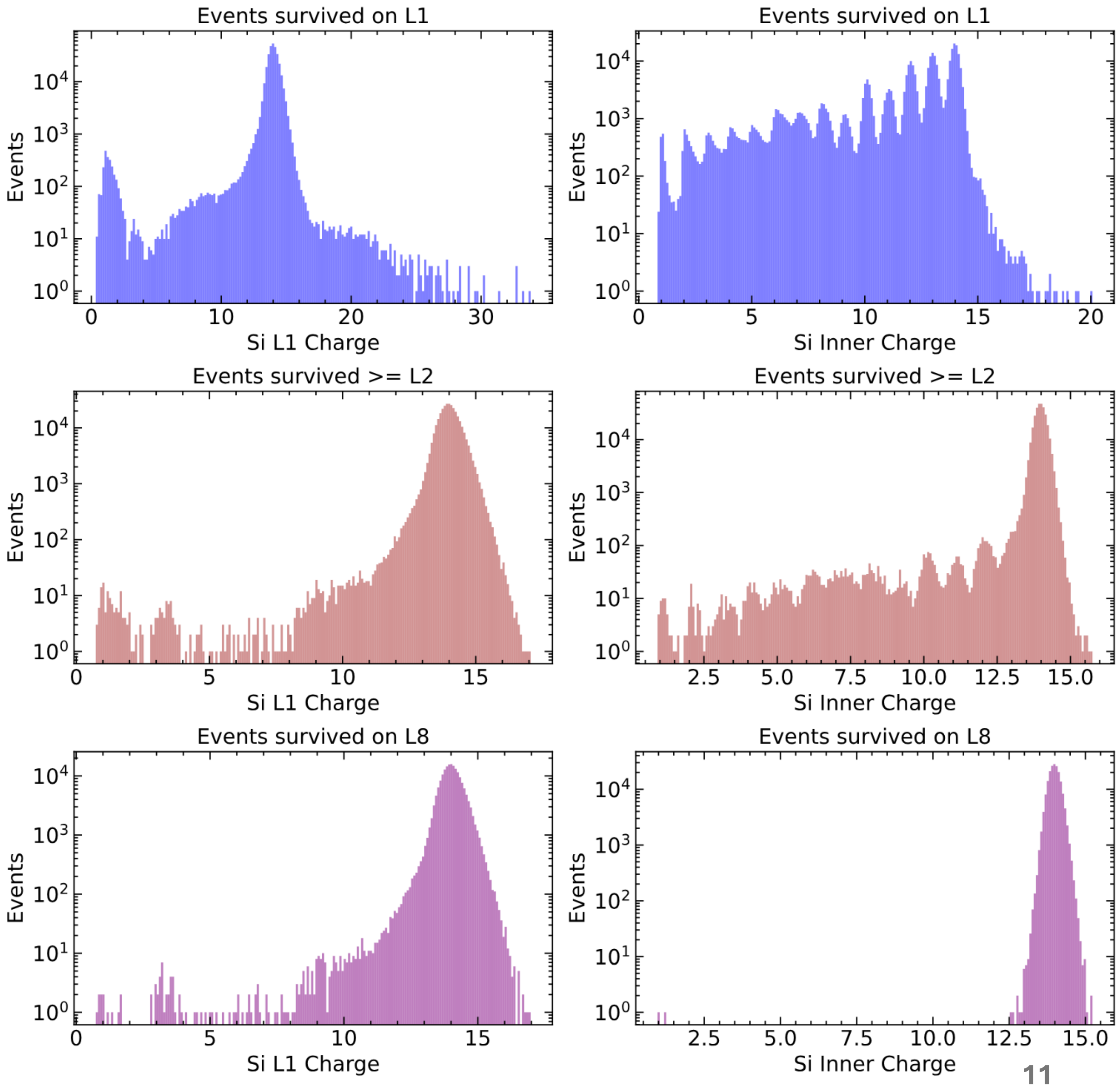
Survived primary Si particles (MC truth)

MC truth information is used to discriminate:

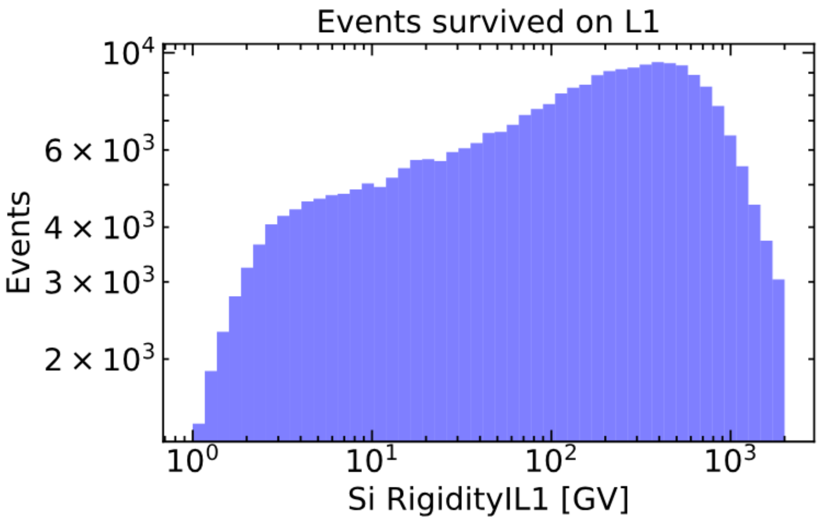
- **Background 1**: primary Si survived above Inner.
- **Signal**: primary Si survived below L2 (Survived in Inner or Survived on L8)

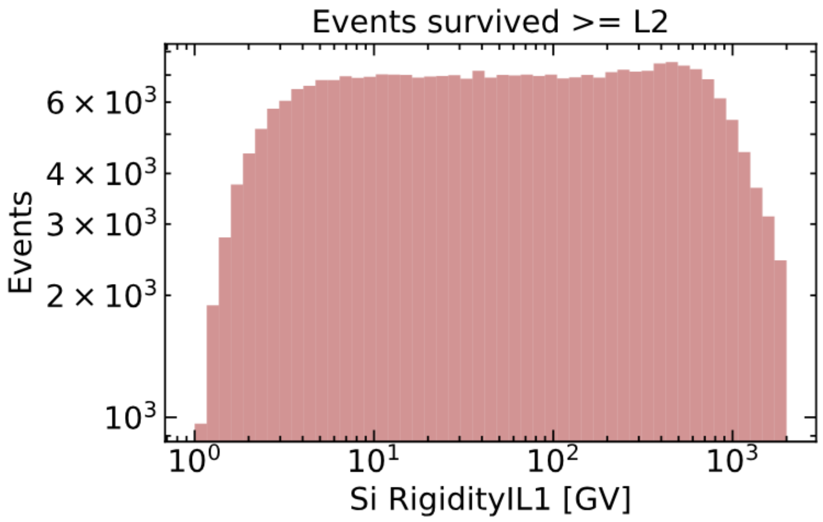


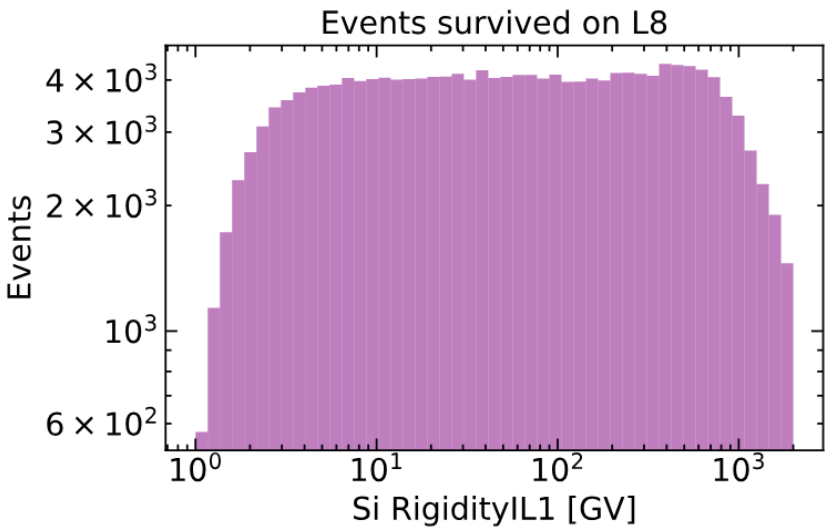




Survived primary Si particles (MC truth)





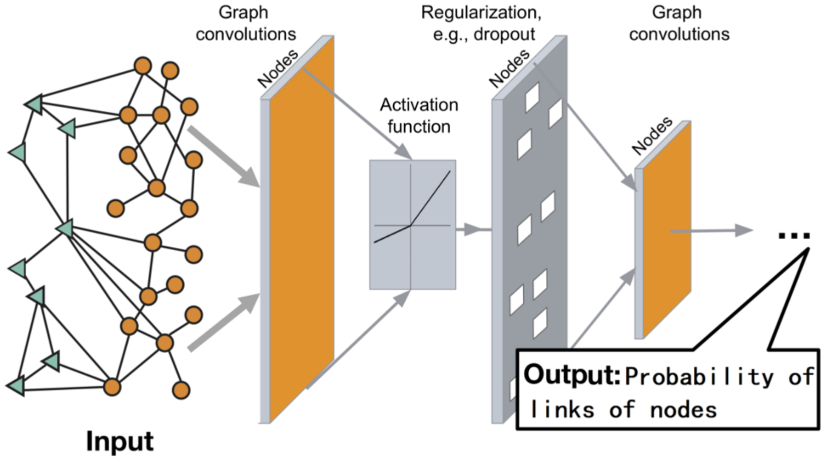


(For events fragmented above Inner, there are fewer events at low rigidity.)

2. GNN model

Deep learning model: GNN

- GNN is one of the deep learning models suitable for AMS studies.
- It is possible to develop a multipurpose model based on GNN for different AMS tasks.



GNN applications for AMS tracker:

- Edge/node classification → tracking;
- global graph classification → nuclei, and isotope identification;
- global graph regression → charge and rigidity reconstruction, etc.
- ~~graph pooling~~

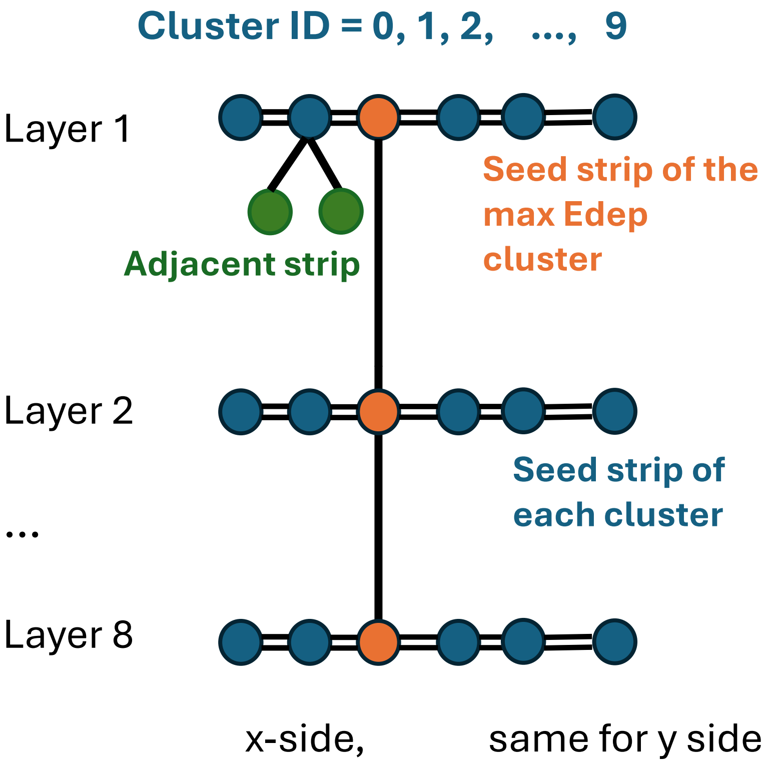
- Start with a hybrid GNN model as an initial exploration.

```
=================================================================
Layer (type:depth-idx)                   Param #
=================================================================
├─Sequential: 1-1                        --
│    └─Linear: 2-1                       640
│    └─LayerNorm: 2-2                    256
│    └─GELU: 2-3                         --
├─GCNConv: 1-2                           --
│    └─SumAggregation: 2-4               --
│    └─Linear: 2-5                       16,384
├─GCNConv: 1-3                           --
│    └─SumAggregation: 2-6               --
│    └─Linear: 2-7                       16,384
├─LayerNorm: 1-4                         256
├─LayerNorm: 1-5                         256
├─SAGEConv: 1-6                          --
│    └─MeanAggregation: 2-8              --
│    └─Linear: 2-9                       16,512
│    └─Linear: 2-10                      16,384
├─SAGEConv: 1-7                          --
│    └─MaxAggregation: 2-11              --
│    └─Linear: 2-12                      16,512
│    └─Linear: 2-13                      16,384
├─LayerNorm: 1-8                         512
├─GATConv: 1-9                           --
│    └─SumAggregation: 2-14              --
│    └─Linear: 2-15                      32,768
├─LayerNorm: 1-10                        256
├─Sequential: 1-11                       --
│    └─Linear: 2-16                      33,024
│    └─LayerNorm: 2-17                   256
│    └─GELU: 2-18                        --
│    └─Dropout: 2-19                     --
├─ModuleList: 1-12                       --
│    └─Linear: 2-20                      8,256
│    └─LayerNorm: 2-21                   128
│    └─GELU: 2-22                        --
│    └─Dropout: 2-23                     --
│    └─Linear: 2-24                      2,080
│    └─LayerNorm: 2-25                   64
│    └─GELU: 2-26                        --
│    └─Dropout: 2-27                     --
│    └─Linear: 2-28                      99
=================================================================
Total params: 177,411
Trainable params: 177,411
Non-trainable params: 0
=================================================================
```

Model architecture used in this work

Graph construction for GNN input

- **Graph nodes:** 10 seed strips and 3 adjacent strips(left and right) for each layer and side.
- **Graph Node features**: ADC value, strip id, side, layer;
- **Graph Edge connection**:
  1. connect the seed strip of the max Edep cluster to other seed strips within the same layer and side, (bidirectional connections);
  2. connect each seed strip to its adjacent strips;
  3. connects the seed strips of the largest clusters between the X- and Y- sides within the same layer, (bidirectional connections) connect the seed strips of the max Edep clusters in adjacent layers for the same side.
- **Event level features: Unbiased TOF Beta ($1/\beta^2$)**



Number of nodes and edges (including strips with ADC=0):

- 366 edges
- 256 nodes
- 4 node features
- 1 event level feature

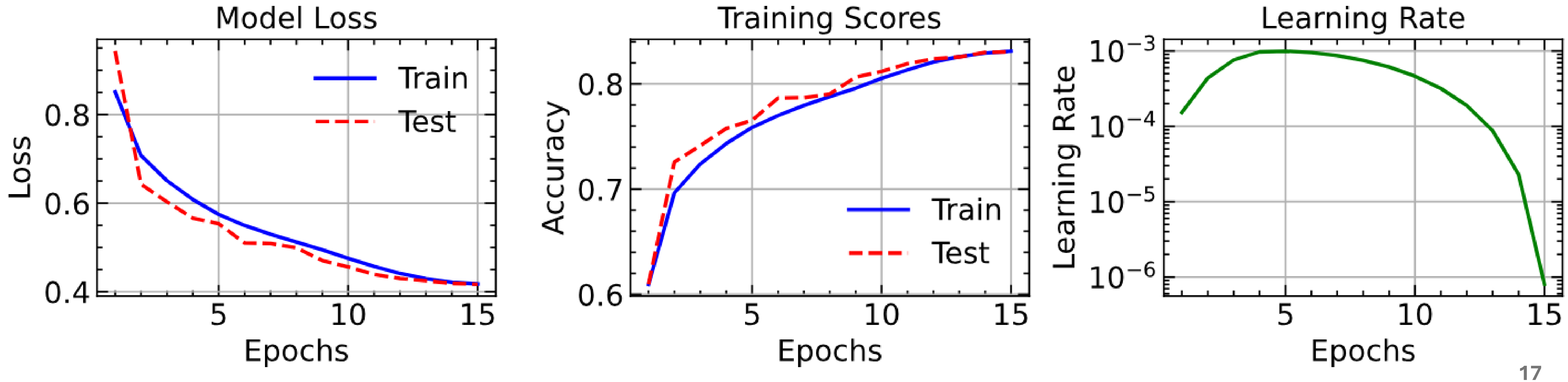
3. Results

GNN model training result

Task: 3-label classification (input sample: 3 category, 300k events each)

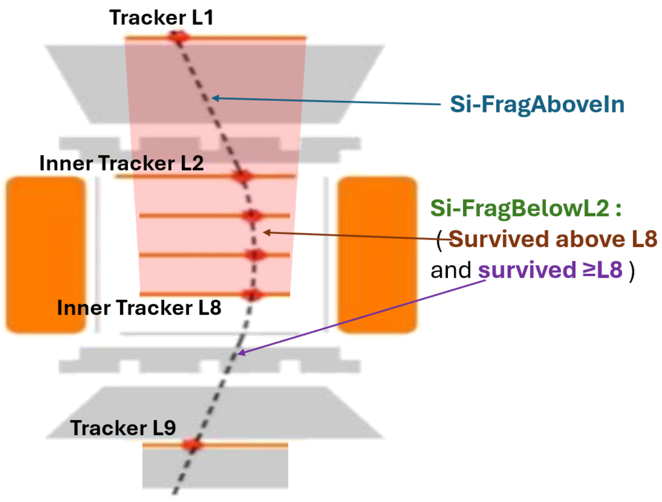
1. Si fragmented above Inner
2. Si fragmented below L2
3. Background nuclei (Mg, Al, P, S = 0.1 : 1 : 1 : 0.1)

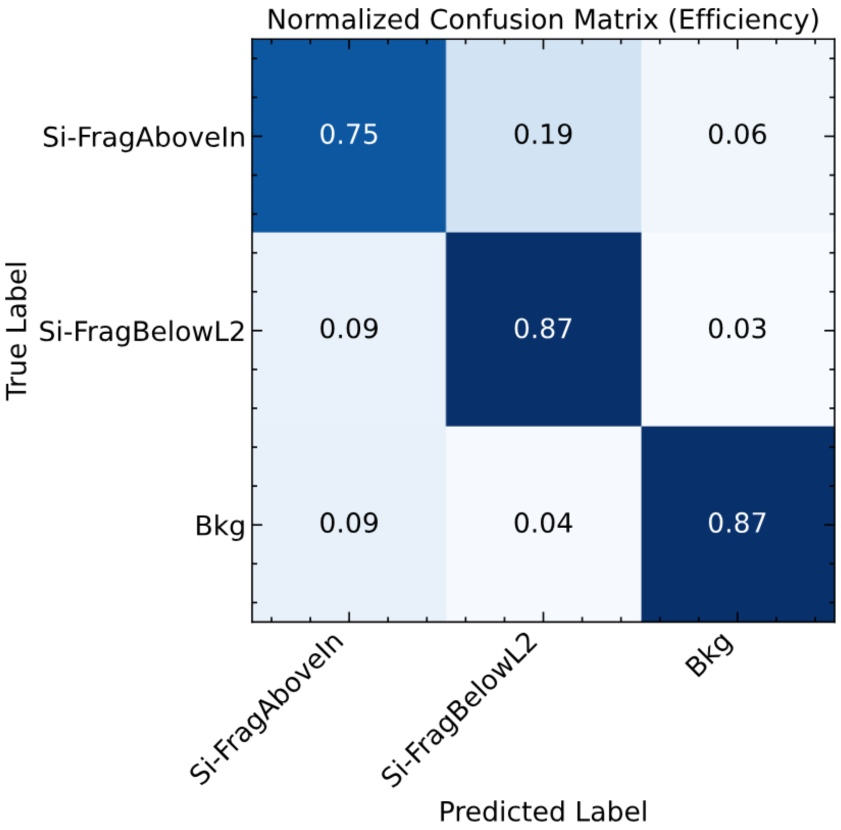
Model training and validation performance:

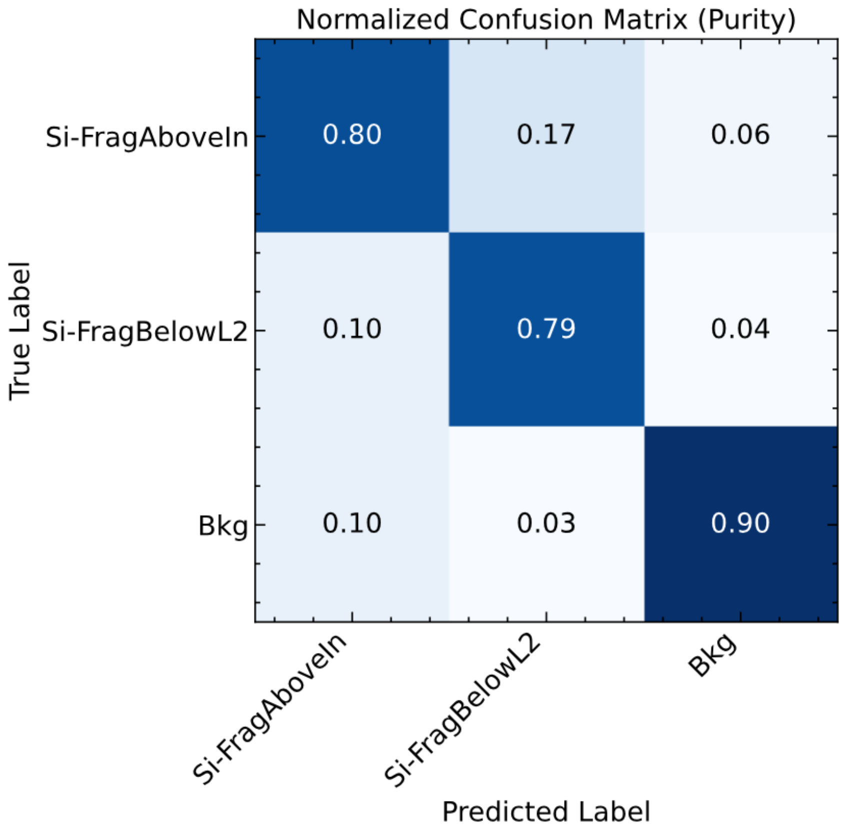


Model overall performance: model validation result

Based on validation sample for each nuclei (~200k events in total):

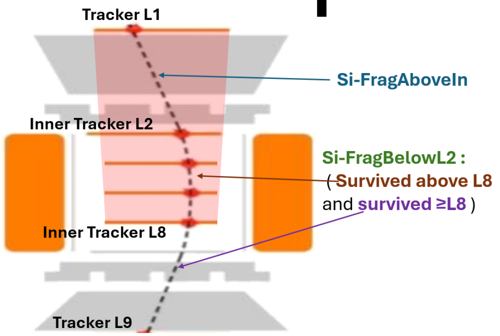






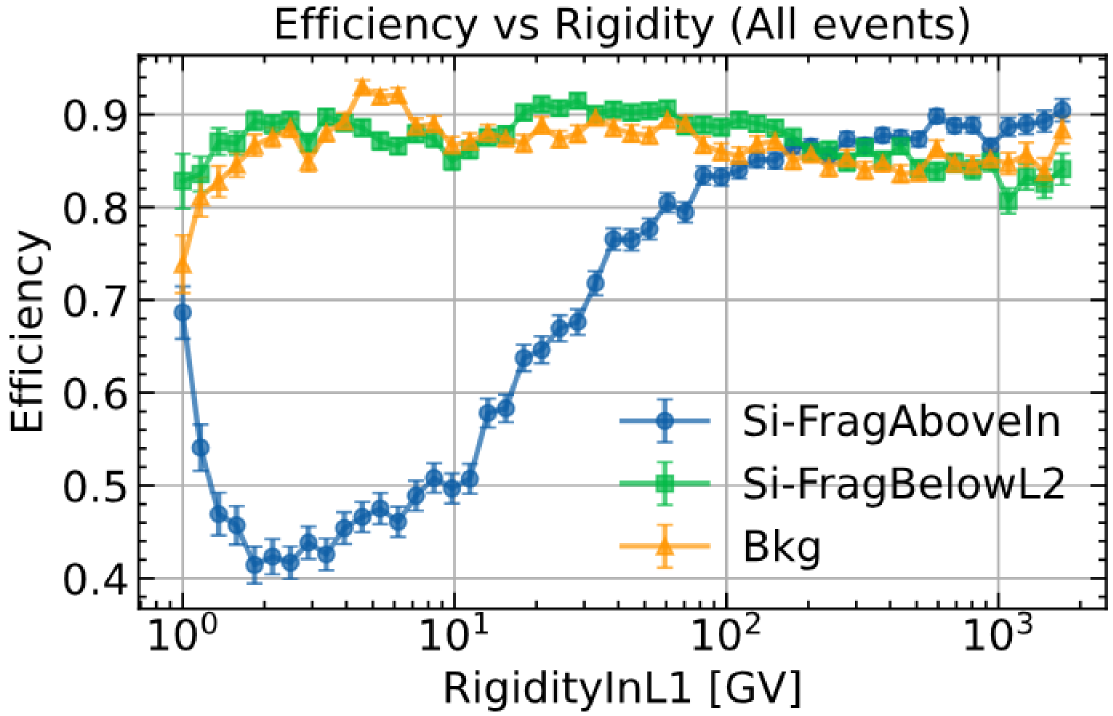
- Given a true label, Efficiency is the probability of each prediction;
- Given a prediction, Purity is the probability of each true label.

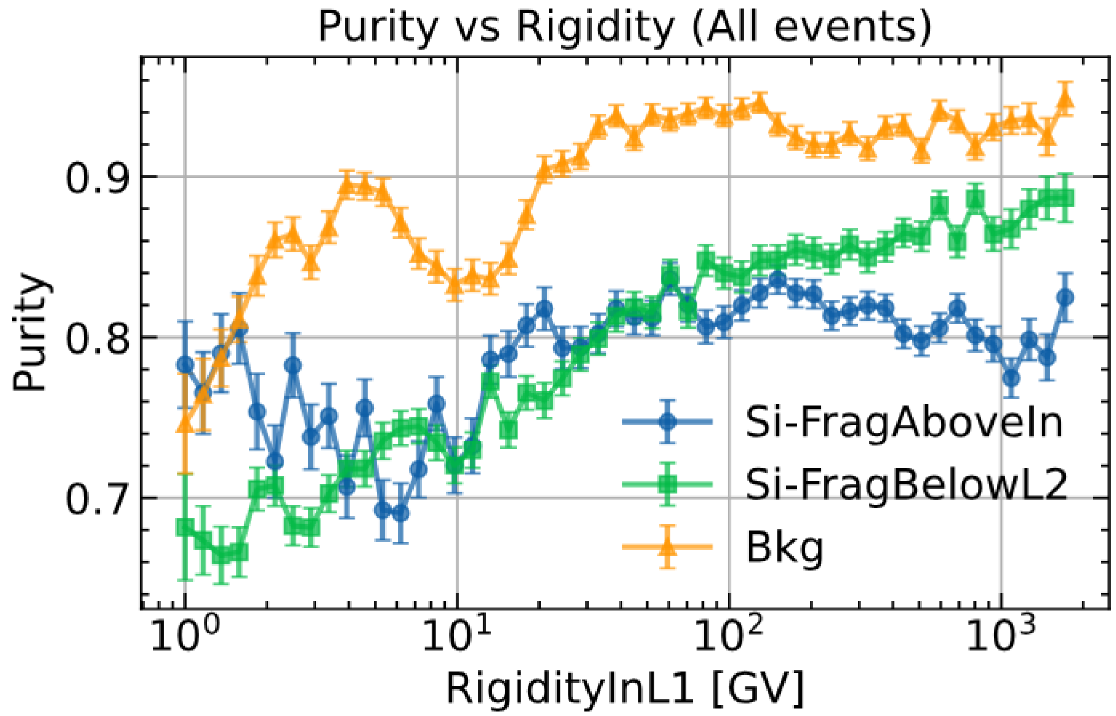
Model performance: model validation result



Purity calculated based on events ratio:

- Si-FragAboveIn : Si-FragBelowL2 : Bkg = 1 : 1 : 1
- Bkg: Mg, Al, P, S = 0.1 : 1 : 1 : 0.1

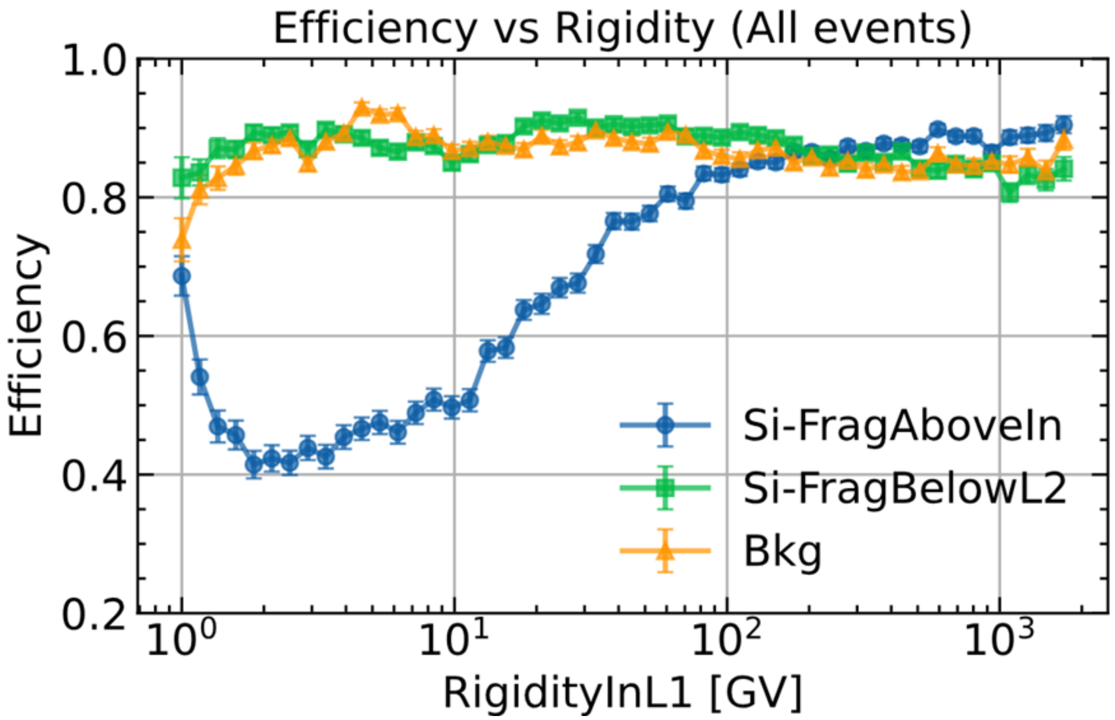




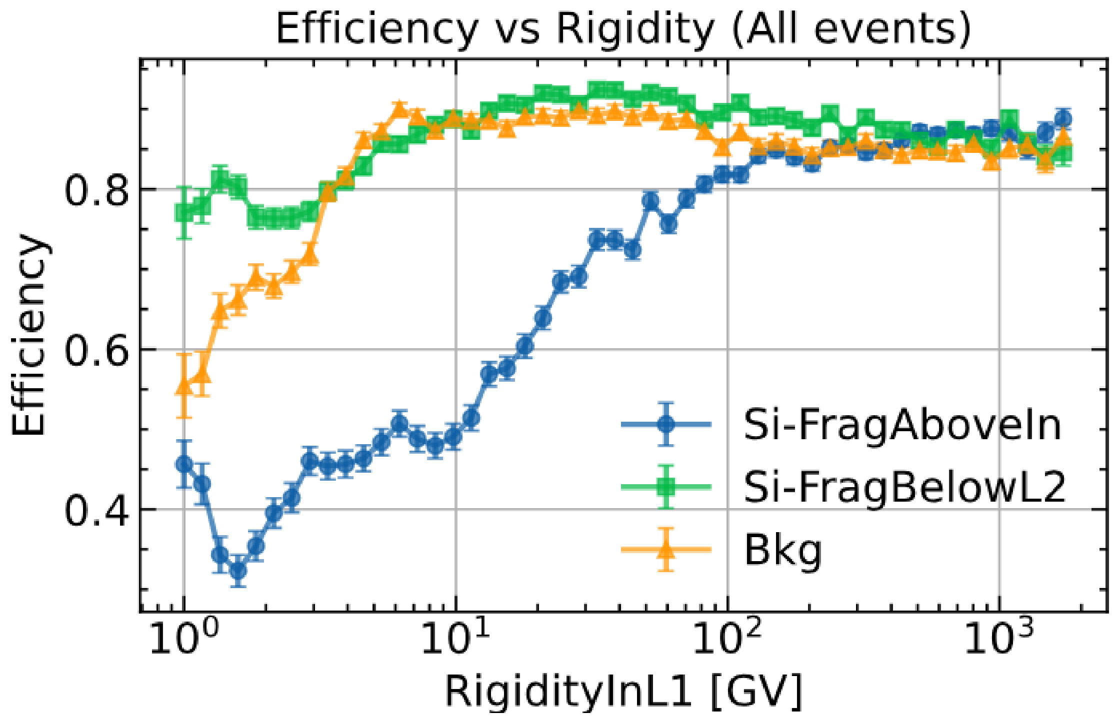
In the sample, for events fragmented above Inner, there are fewer events at low rigidity (slide#12),
→ larger statistical error for Si-FragAboveIn at low rigidity.

Model performance: model validation result

With TOF Beta information (this result)

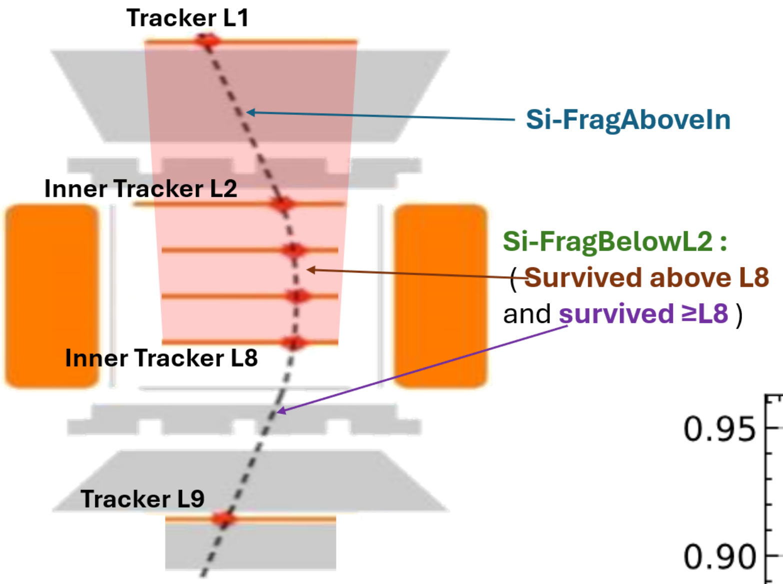


Without TOF Beta information

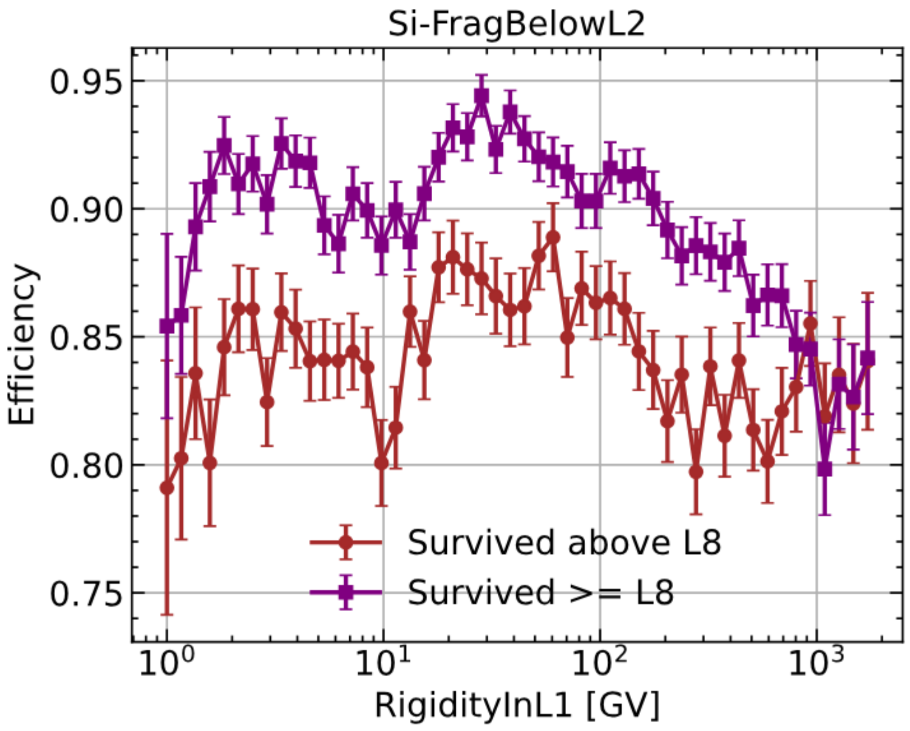


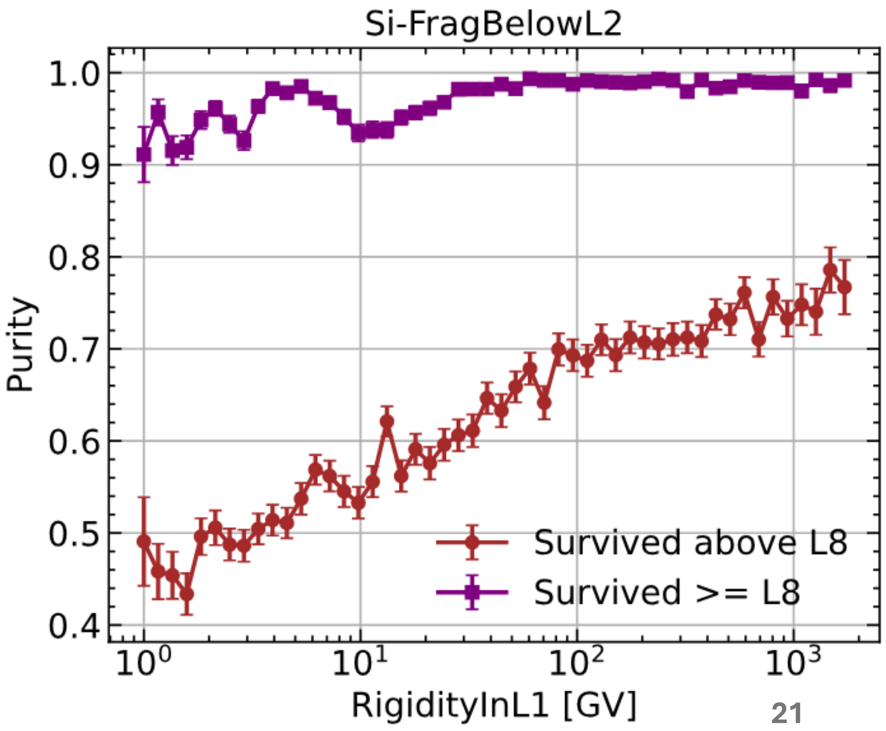
Adding TOF information:
overall performance improves, and the rigidity dependence at low rigidity is reduced.

Model performance: model validation result for Si fragmented below L2

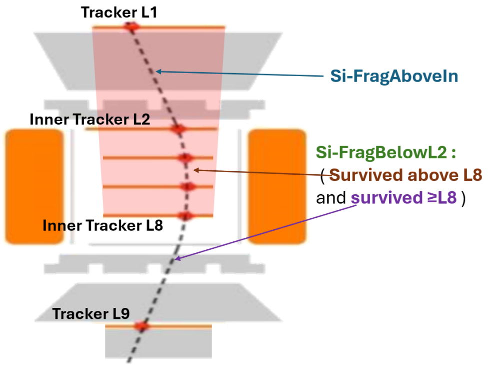


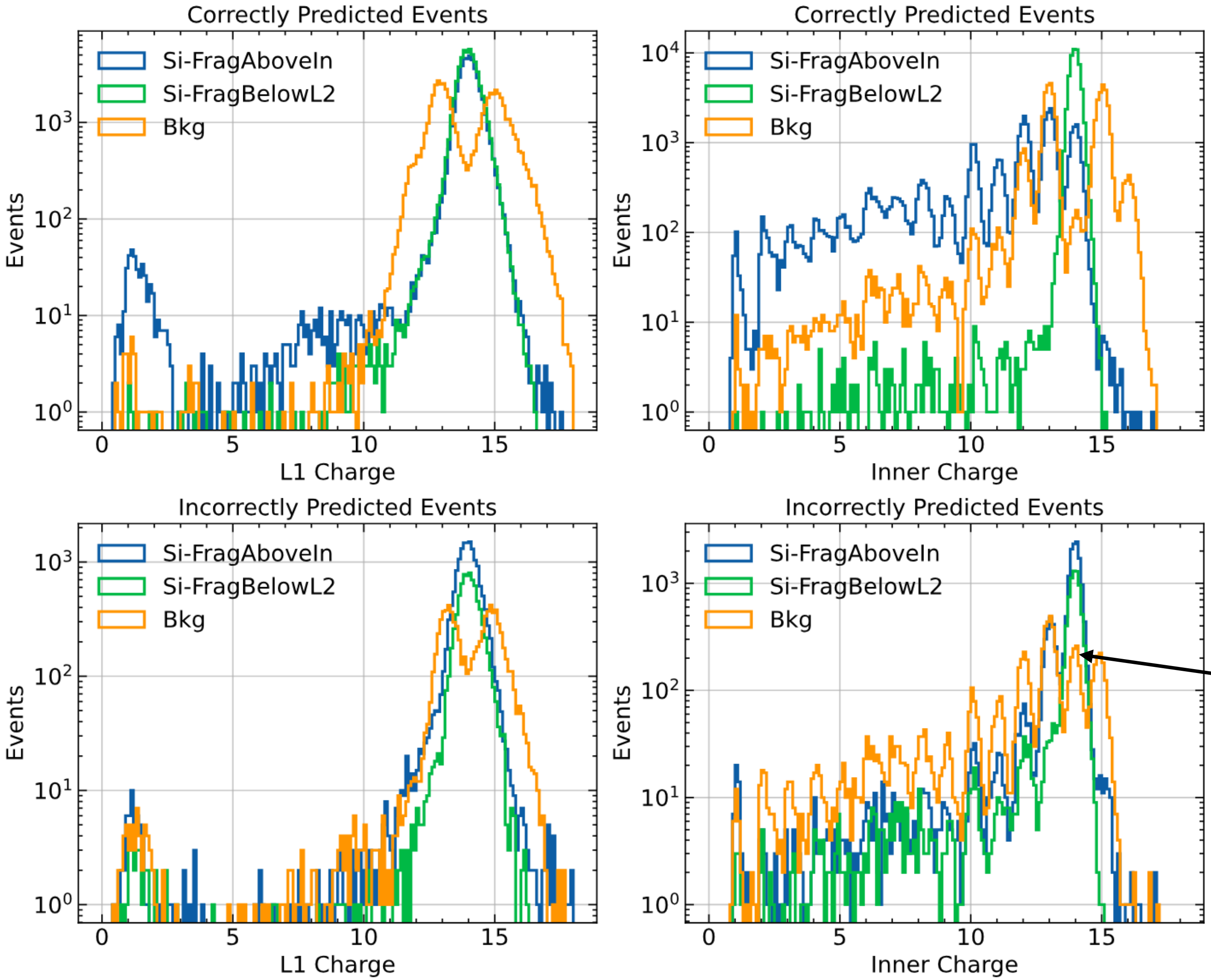
Model testing result for **Interacting Si events in Inner Tracker** and **Non-Interacting Si events** :





Model performance: comparison to standard charge reconstruction results



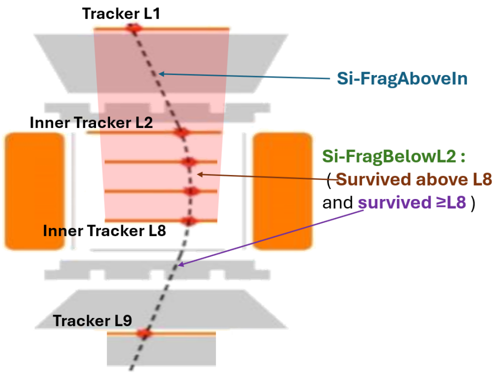


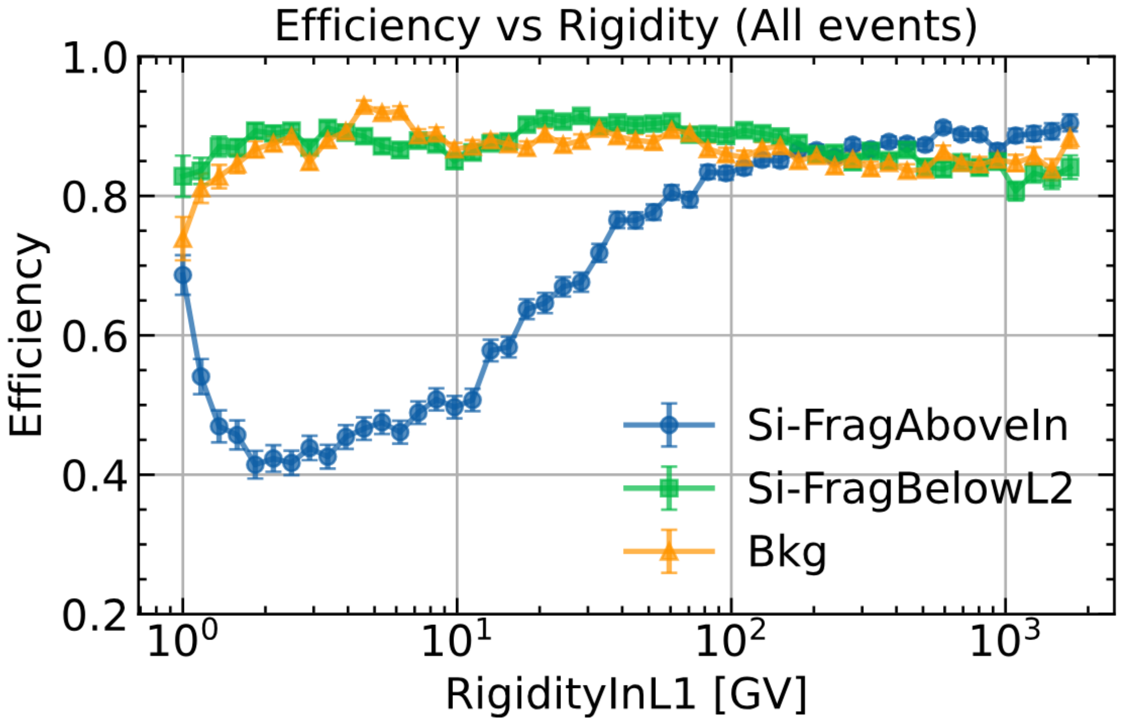
(
Low rigidity events:
- Small charge
- Large ADC
- Small number of clusters

)

Background nuclei (Mg, Al, P, S) with Inner charge == 14

Model performance: comparison to standard selection efficiency

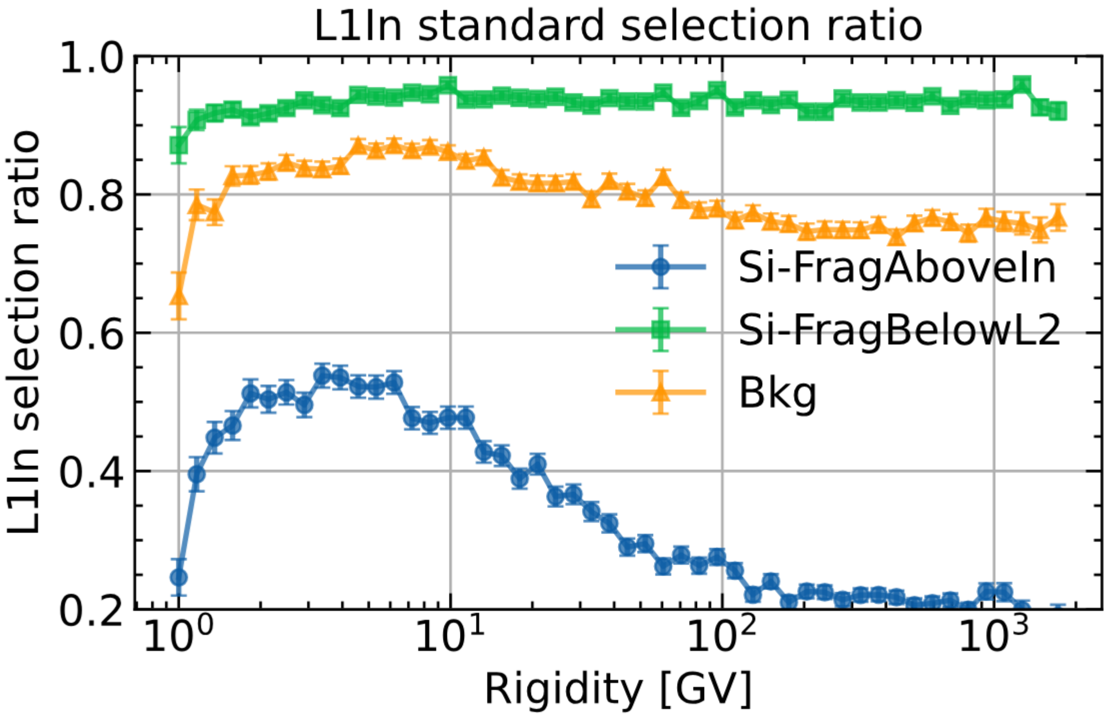




Standard selection results for reference:

(Standard Si selection for Si-FragAboveIn and Si-FragBelowL2)

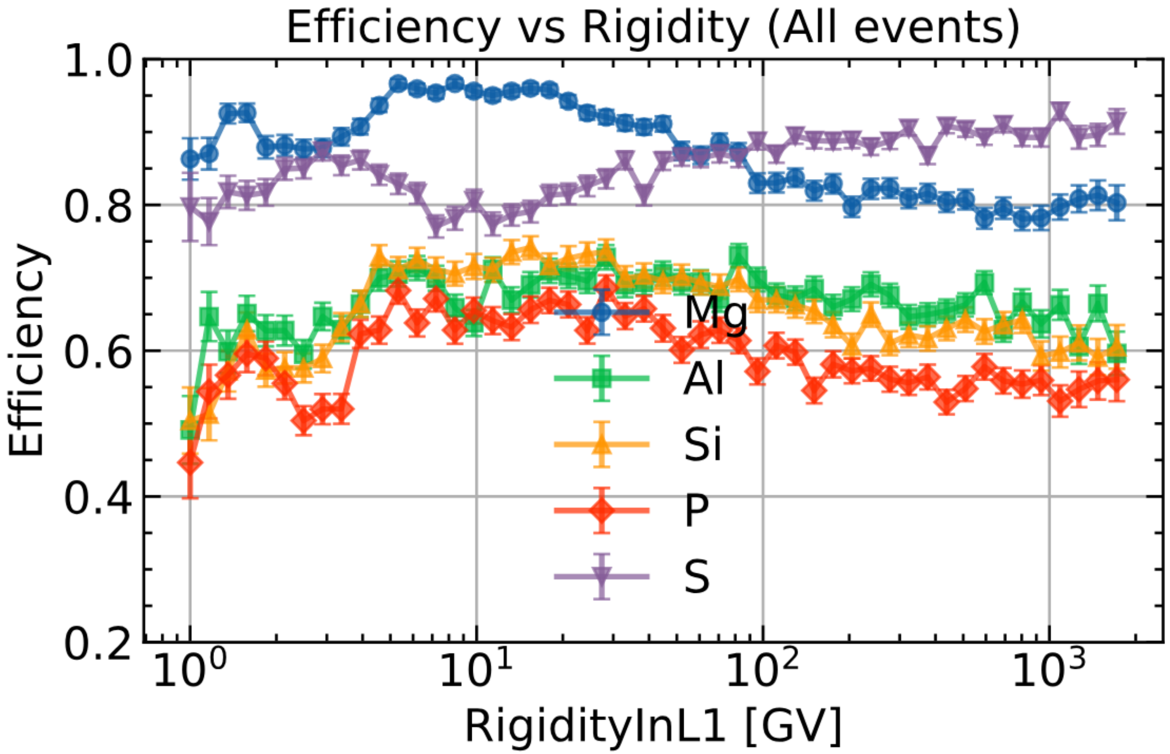
(Standard Mg, Al, P, S selections for Mg, Al, P, S)



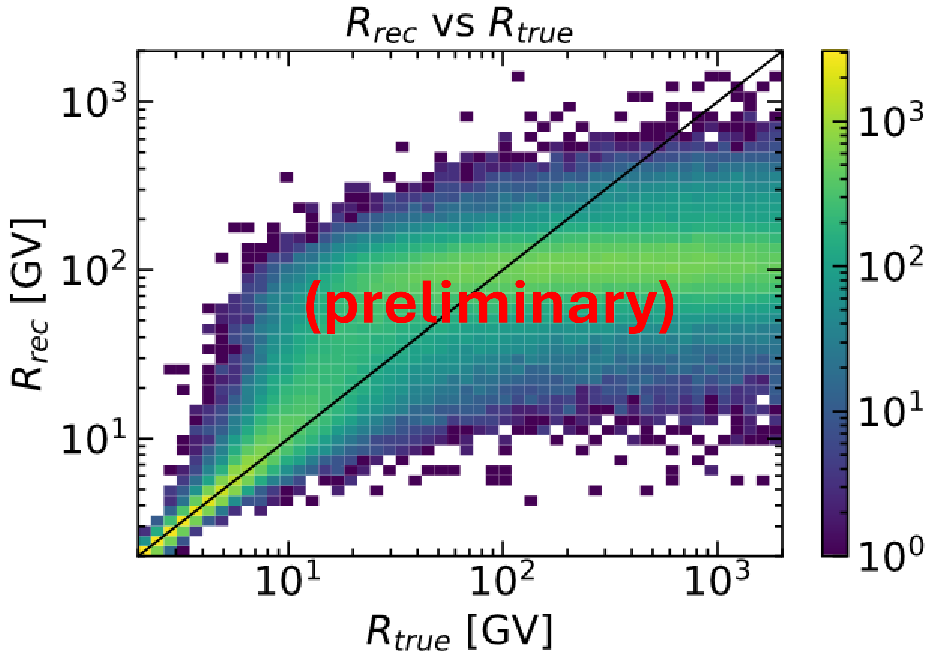
Extending the model for other tasks:

Based on the method, the model can be extended for other tasks:

On-going:
nuclei identification for Mg, Al, Si, P, S



On-going:
true rigidity reconstruction



Summary and To-do

A preliminary study using MC for nuclei identification with deep learning method shows promising potential compared to traditional method.

To-do:

- **Lots of optimizations needed:**
  1. Optimizing graph construction method (variable number of nodes/edge, group by distance, find Edep threshold value, etc.)
  2. Implementing stat-of-the-art models
  3. Model hyperparameter optimization
  4. Further improve Low rigidity performance with Rigidity Piecewise training
- **Implementing MC reweight method for Data–MC consistency**

Backup

ADC correction options in AMS software

◆ CorrectionOptions

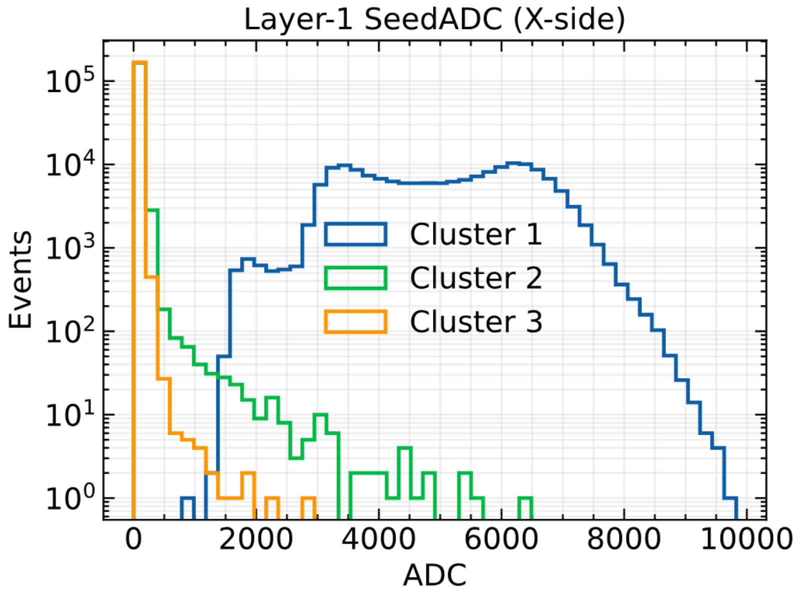
enum **TrClusterR::CorrectionOptions**

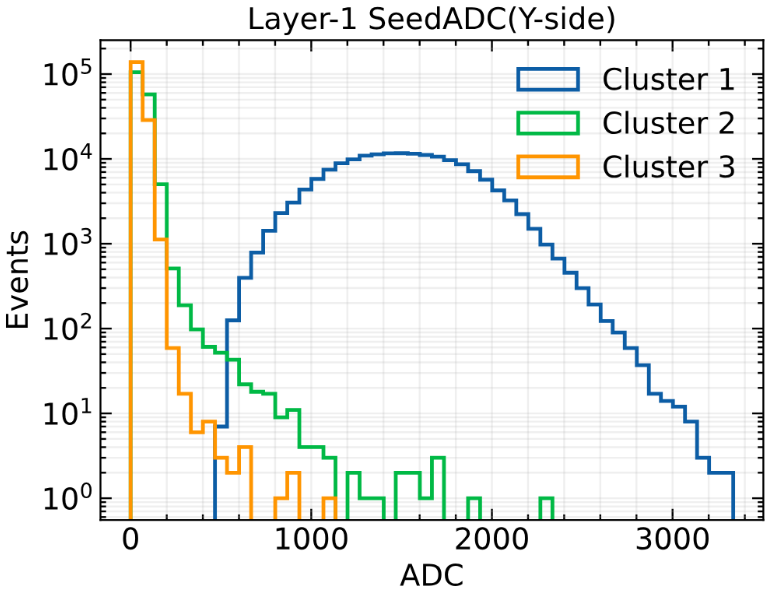
Enumerator	
kNoCorr	No Correction Applied.
kAsym	Signal Corr.: Cluster Asymmetry Correction (left/right)
kPStrip	Signal Corr.: P-Strip Correction.
kAngle	Total Signal Corr.: Energy Loss Normalization at 300 um [cos(Theta)^-1].
kGain	Total Signal Corr.: Gain Correction.
kLoss	Total Signal Corr.: Charge Loss Correction.
kLoss2	Total Signal Corr.: Charge Loss Correction (alternative to kLoss)
kPN	Total Signal Corr.: Normalization to P-Side (probably not working, however not really needed)
kMIP	Total Signal Corr.: Normalization to number of MIP.
kMeV	Total Signal Corr.: Multiply by 300 um MIP energy deposition (estimated to be 81 keV)
kBeta	Total Signal Corr.: Beta correction.
kRigidity	Total Signal Corr.: Rigidity correction.
kCoupl	Coordinate Corr.: Correct for the charge coupling (4%)
kBelau	Coordinate Corr.: Belau correction.
kOld	Use old charge calibration.
kAsymEta	Signal Corr.: New Cluster Asymmetry Correction (left/right)
kQ2Eta	ADC->Q2 Correction: ADC to Q2(Z gain)
kTotSign2017	Charge calibration 2017 (for now used only for MC)
kSimAsym	Signal Corr.: Cluster asymmetry correction using TRMCFFKEY.Asymmetry.
kSimSignal	Total Signal Corr.: raw ADC to MIP scale for every VA.
kOverflow	ADC Overflow Corr.

Definition at line **54** of file **TrCluster.h**.

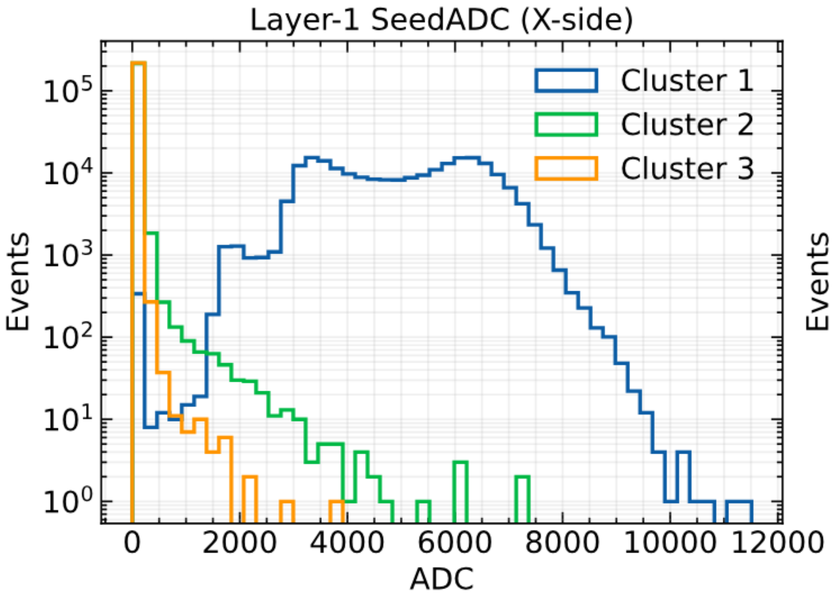
ADC Values (Sulfur MC, L1Inner selection, 200K events)

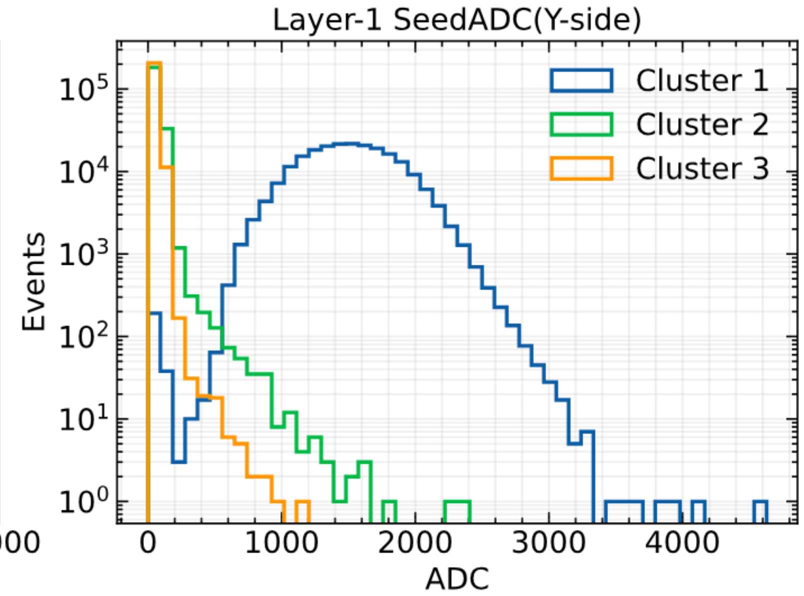
ADC without correction



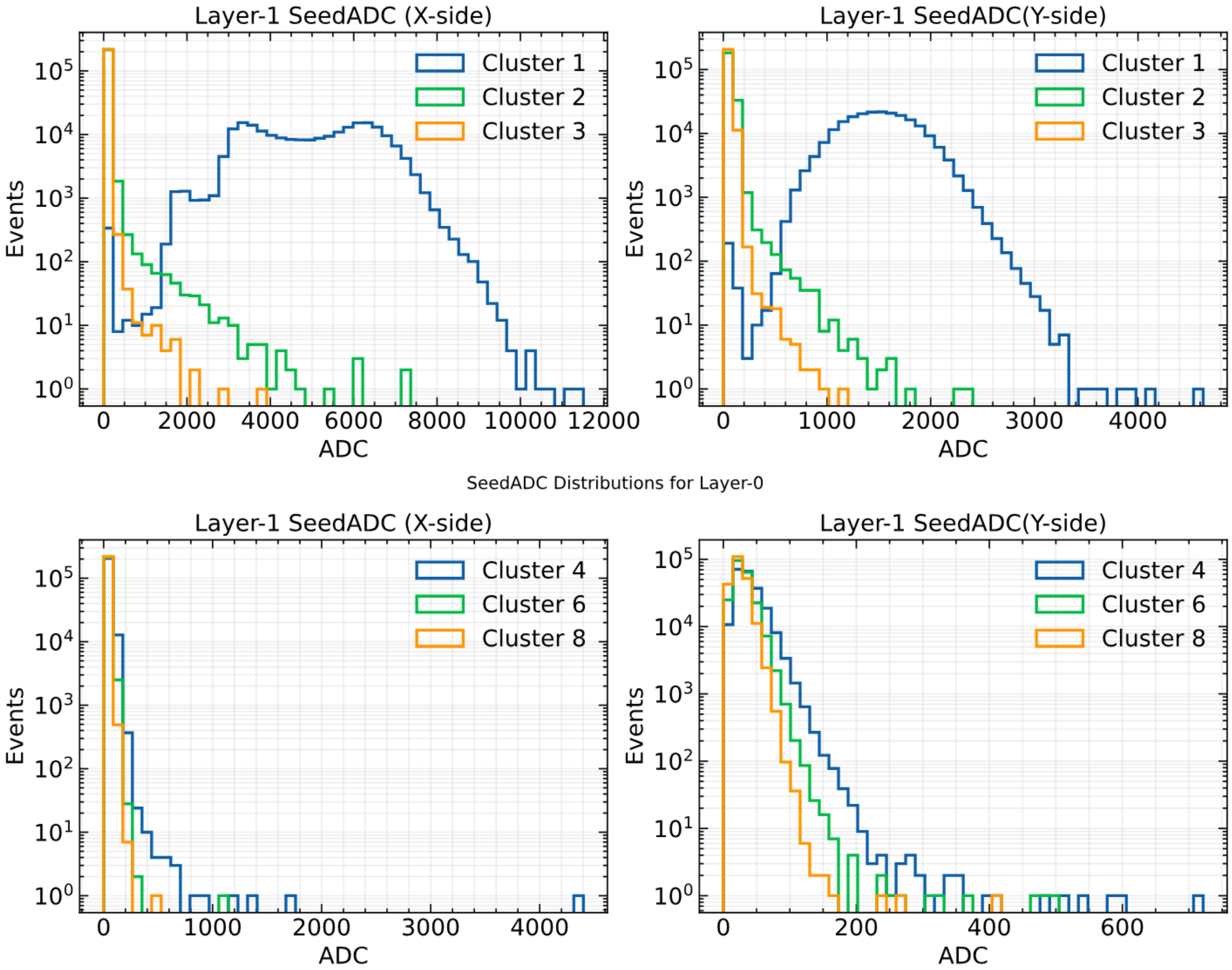


ADC with Beta correction (kBeta)

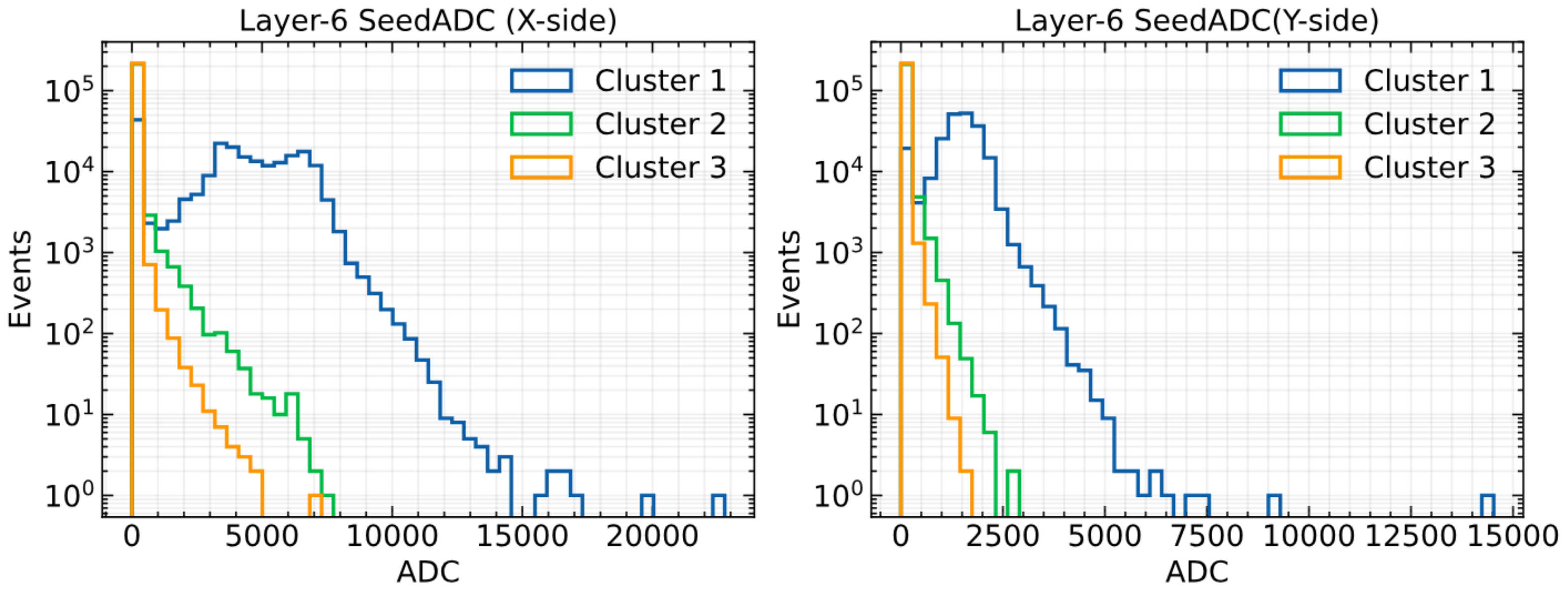




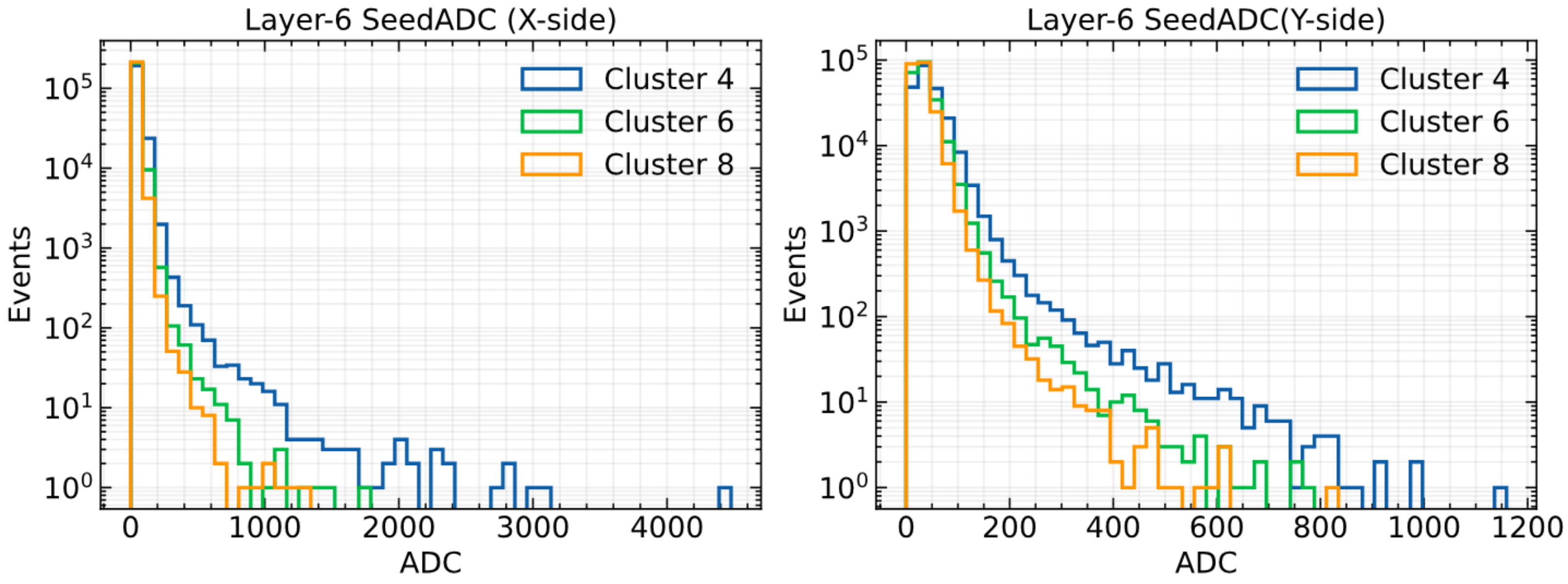
ADC Values (S MC L1Inner selection, 200K events)



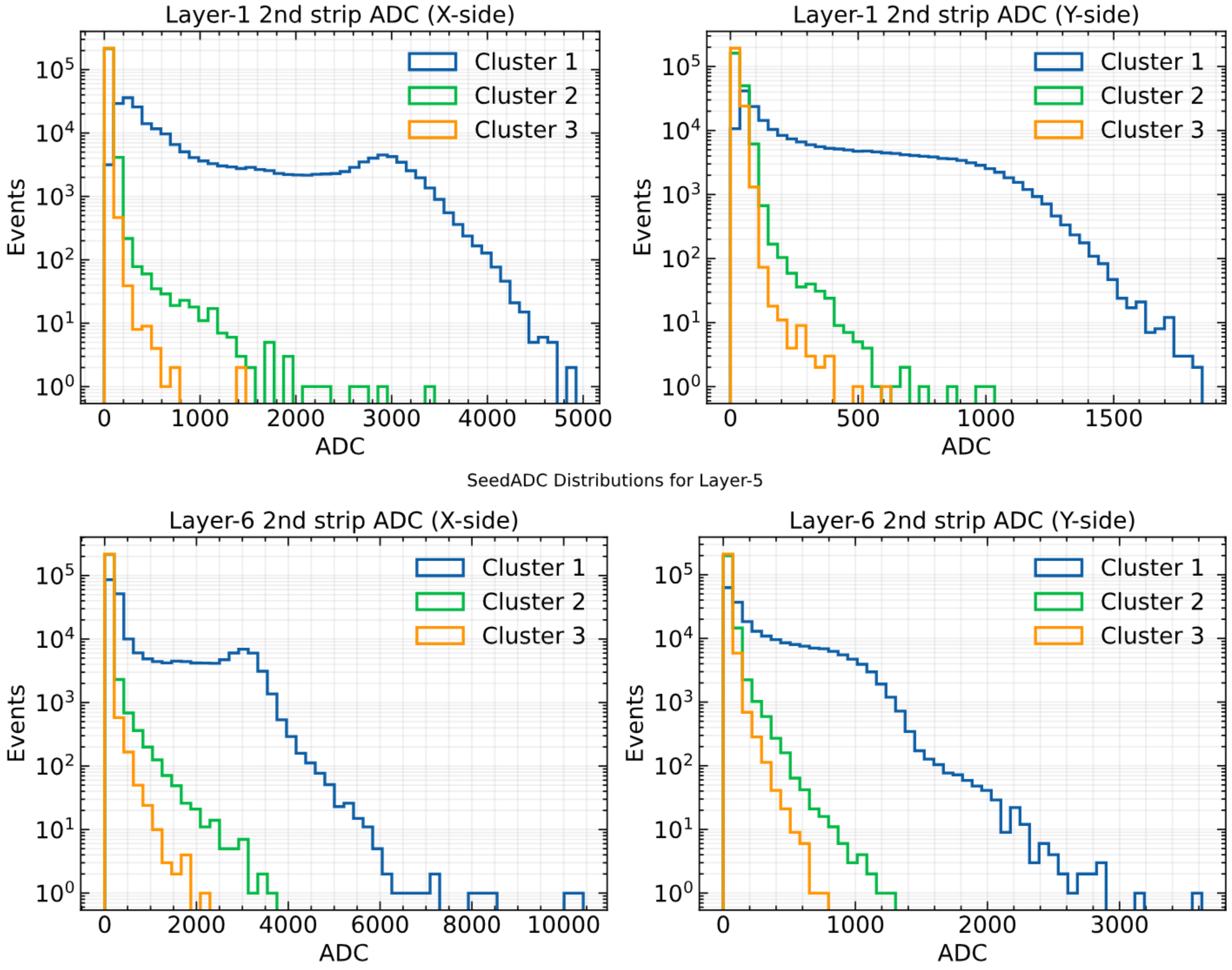
ADC Values (S MC L1Inner selection, 200K events)



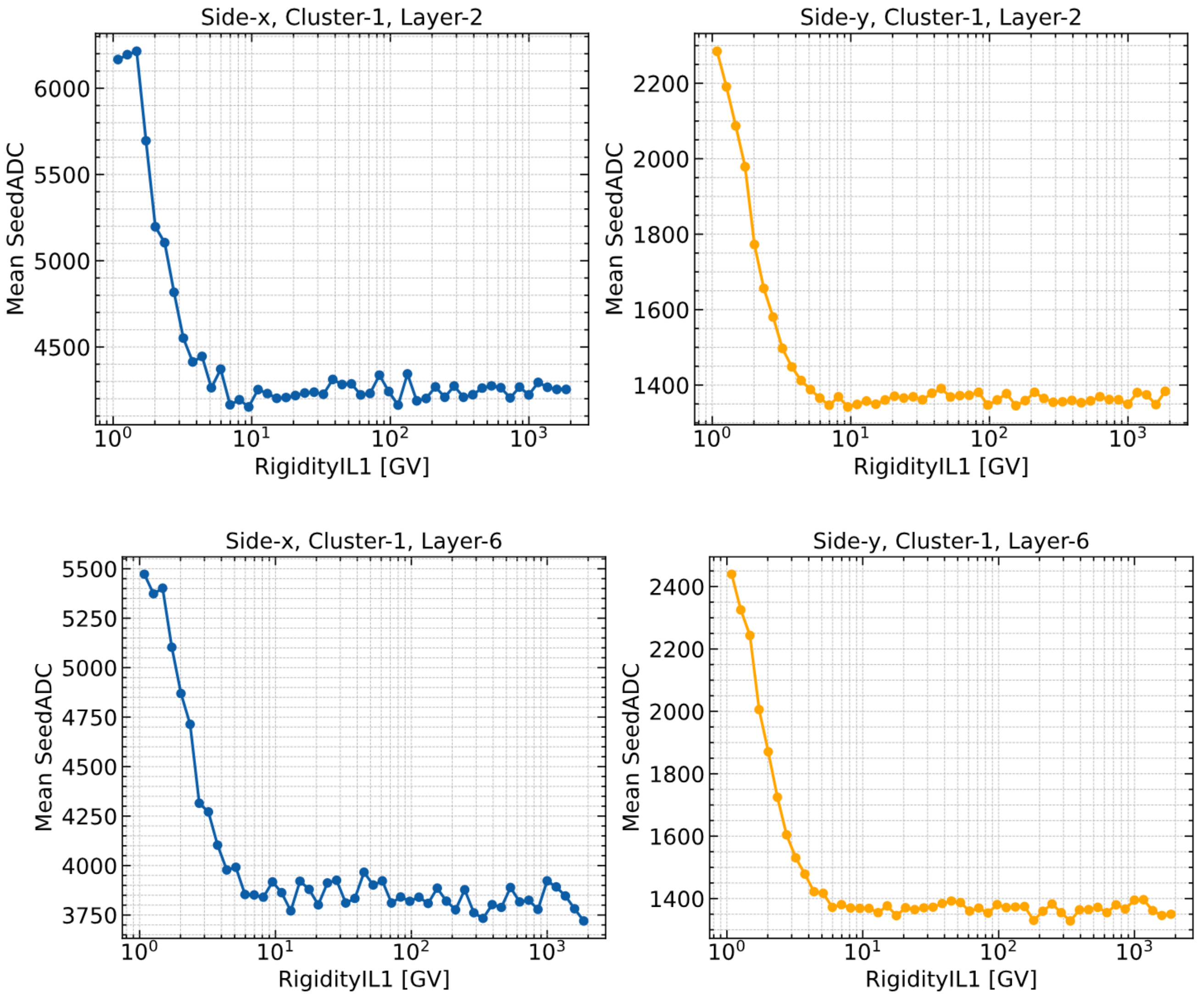
SeedADC Distributions for Layer-5



ADC Values (S MC L1Inner selection, 200K events)



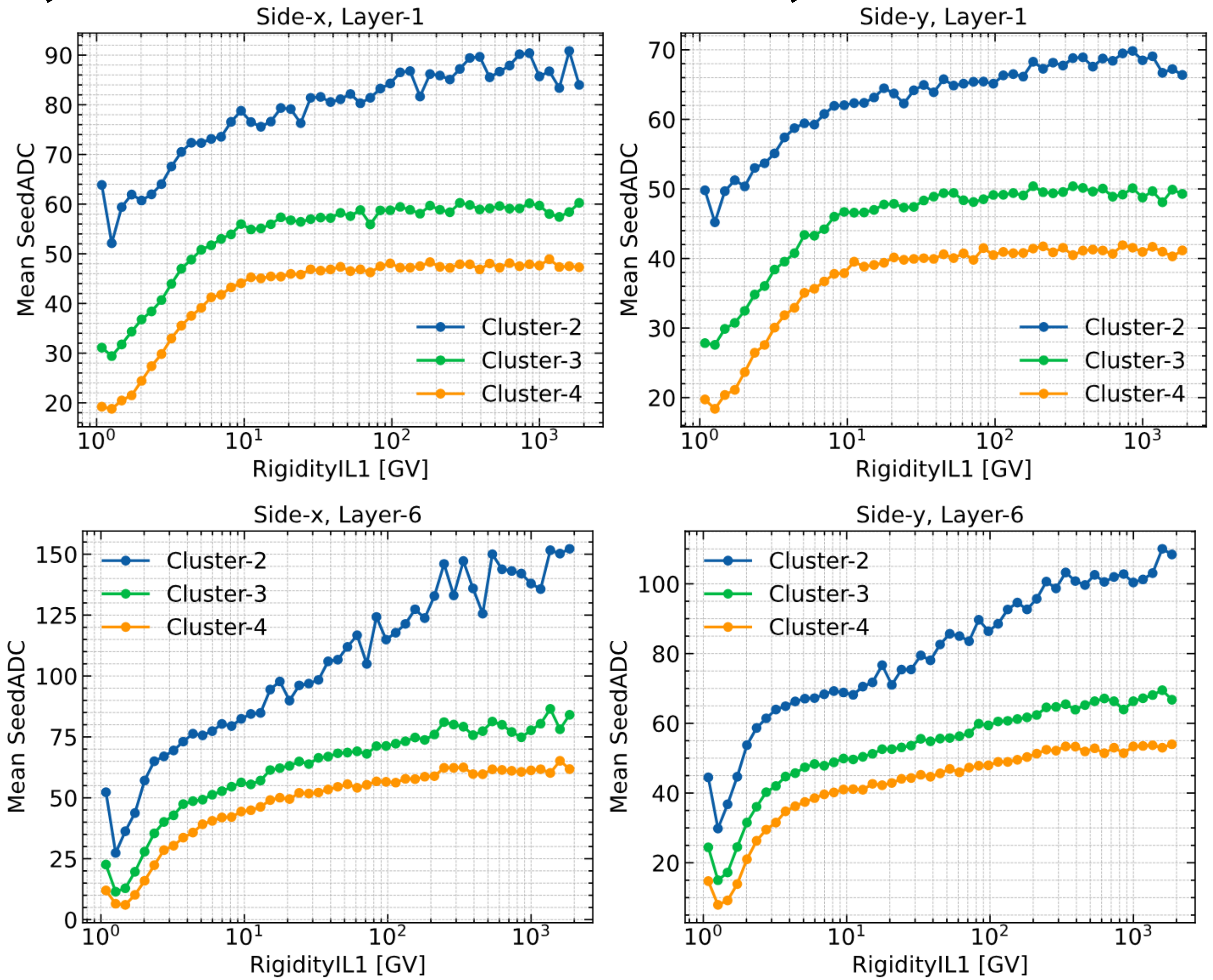
Rigidity dependence of ADC with Beta correction (Sulfur MC, L1Inner selection, 200K events)



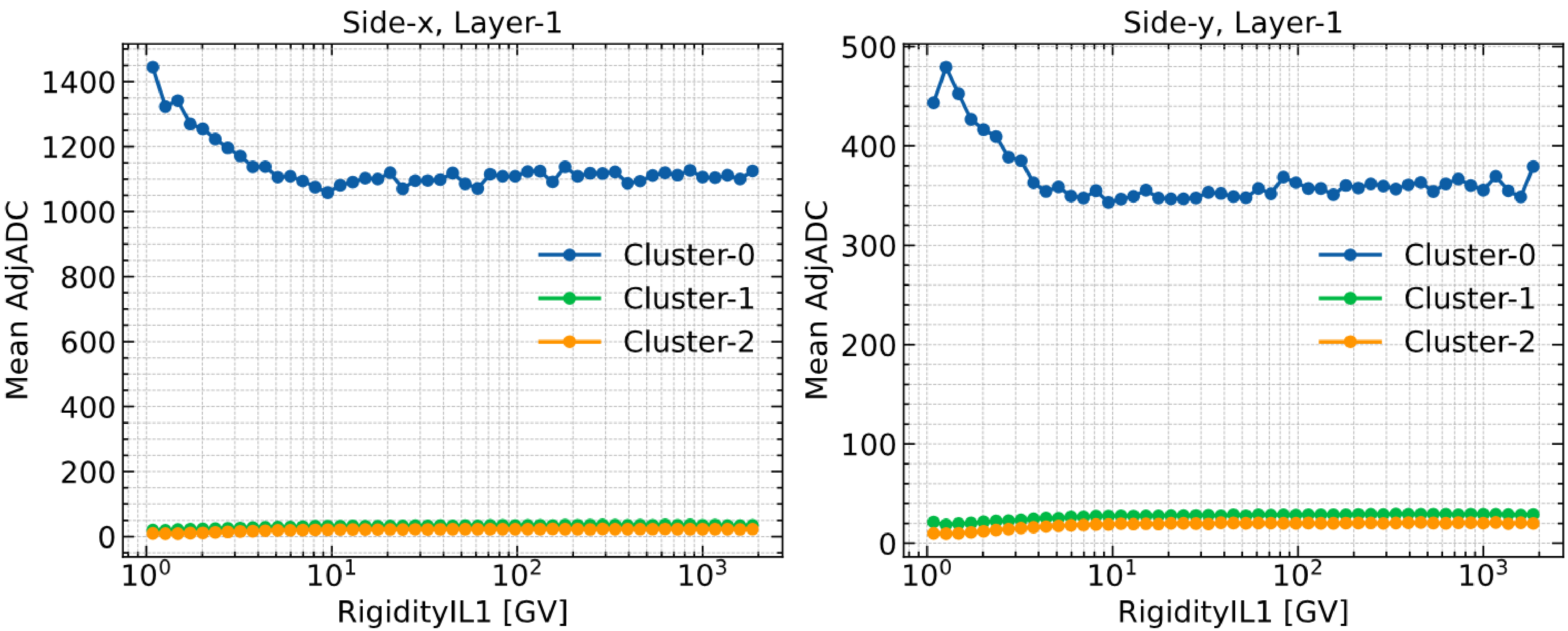
Further Beta correction is needed.
→ Add TOF Beta information to the deep learning models

(Statistical error not added)

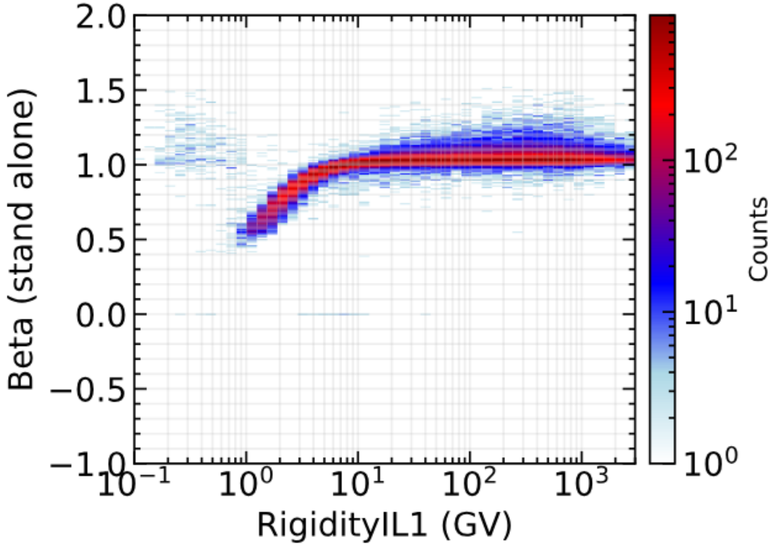
Rigidity dependence of ADC with Beta correction (Sulfur MC, L1Inner selection, 200K events)



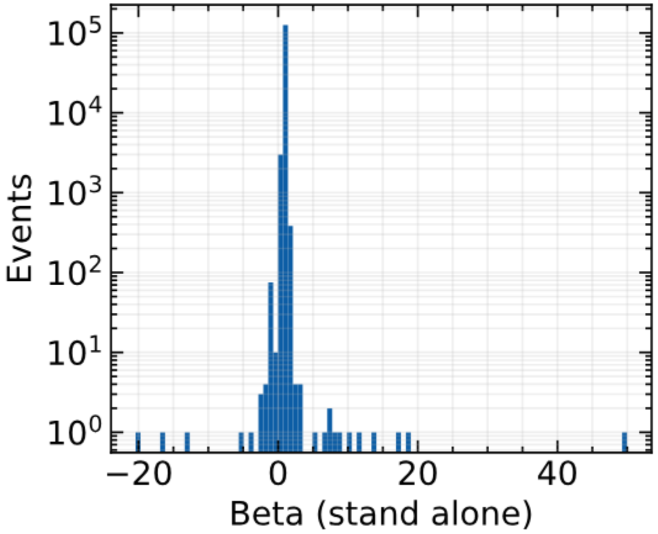
Rigidity dependence (ADC with kBeta)



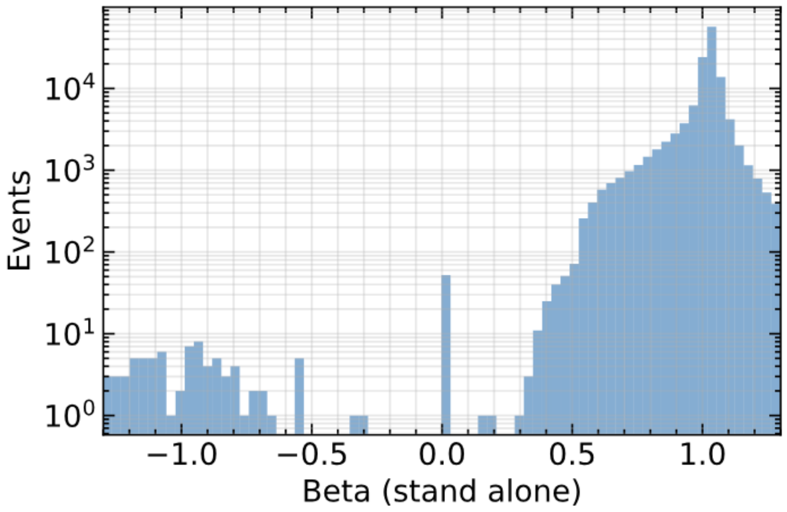
TOF Beta (standalone reconstruction)



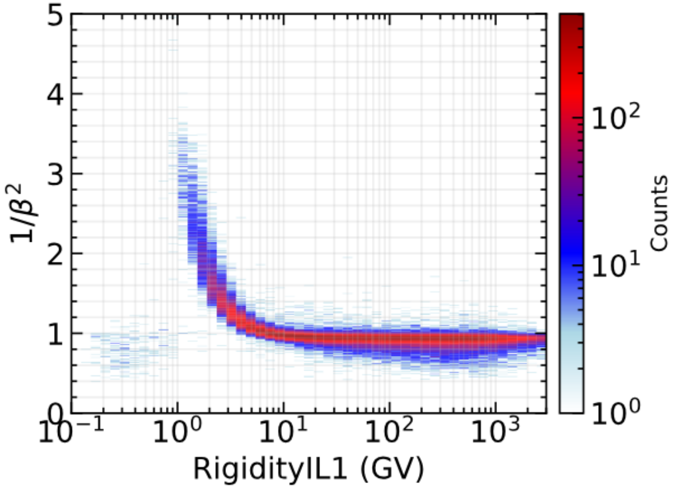
Beta (full range)

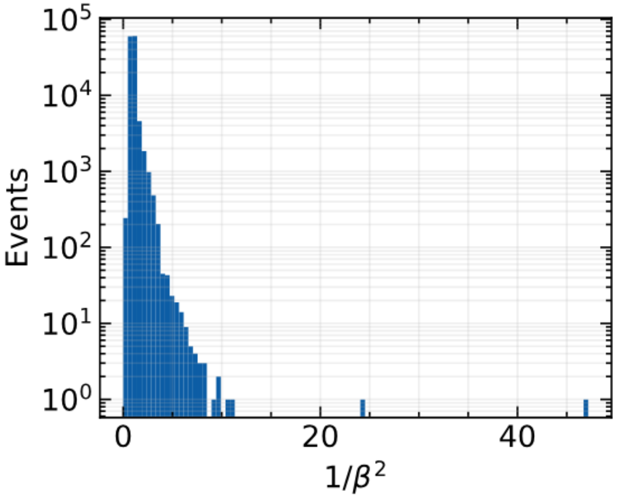


Beta (zoom-in)

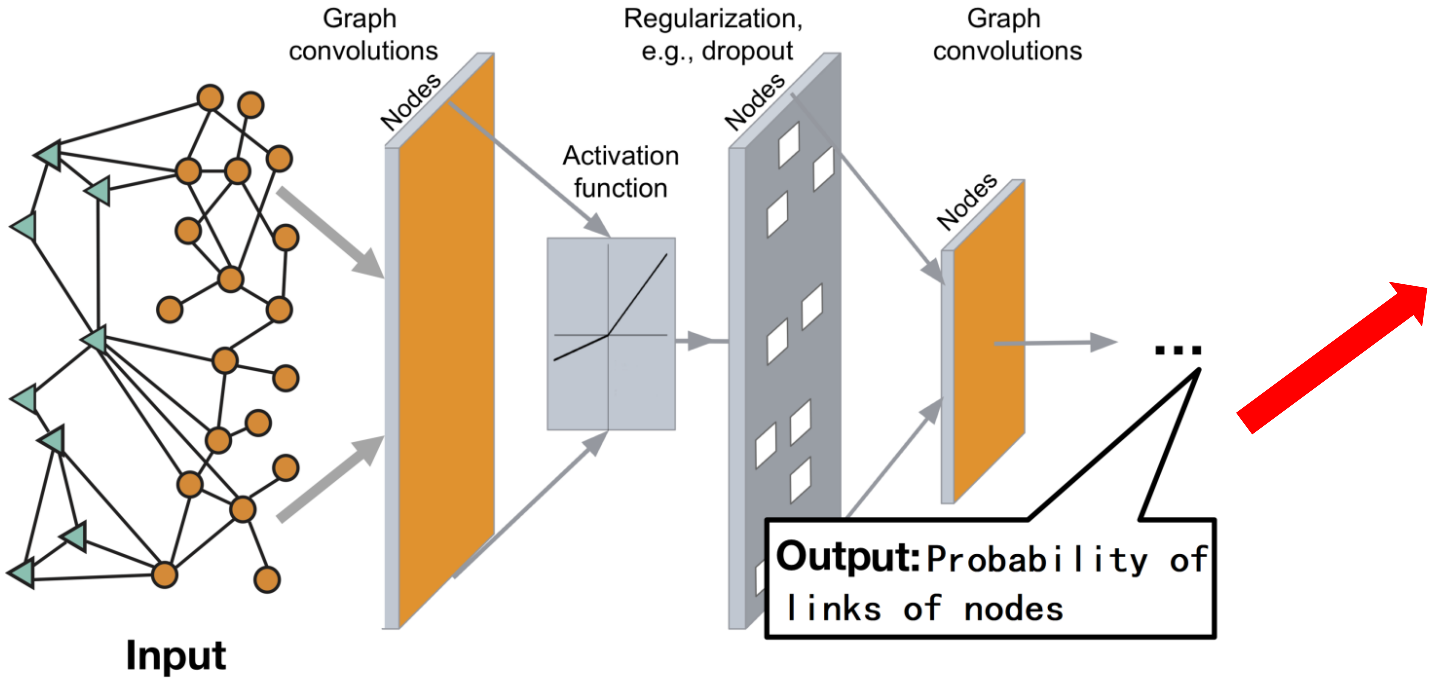


Clamp Beta values below 0.1 to 0.1:





Deep learning models: GNN



GNN applications for AMS tracker:

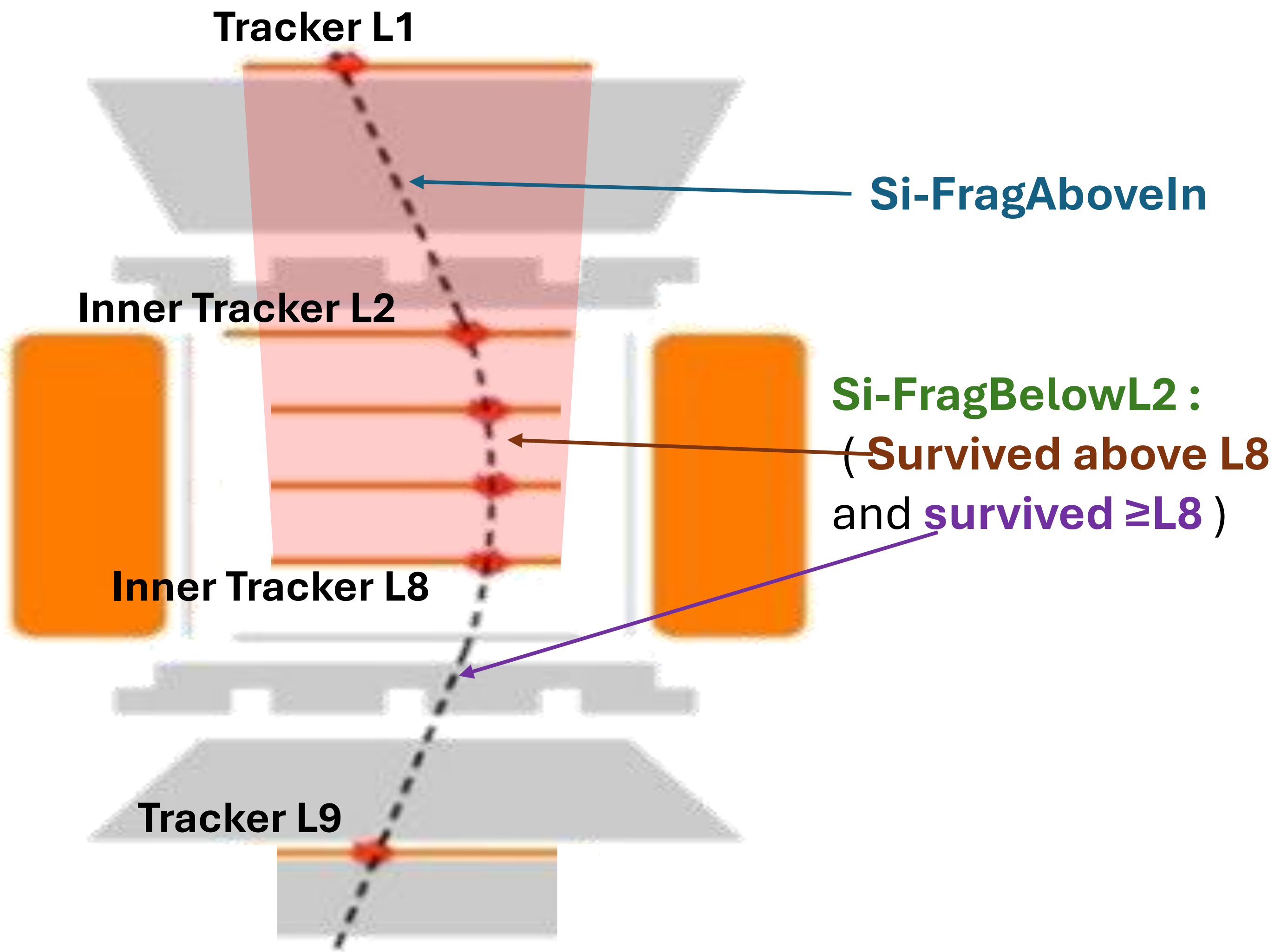
- Edge/node classification → tracking;
- global graph classification → nuclei, and isotope identification;
- global graph regression → charge and rigidity reconstruction, etc.
- ~~graph pooling~~

Pros: suitable for sparse signal; suitable track finding tasks.

Cons: need more study on the structure of GNN input; Event pre-selection needed.

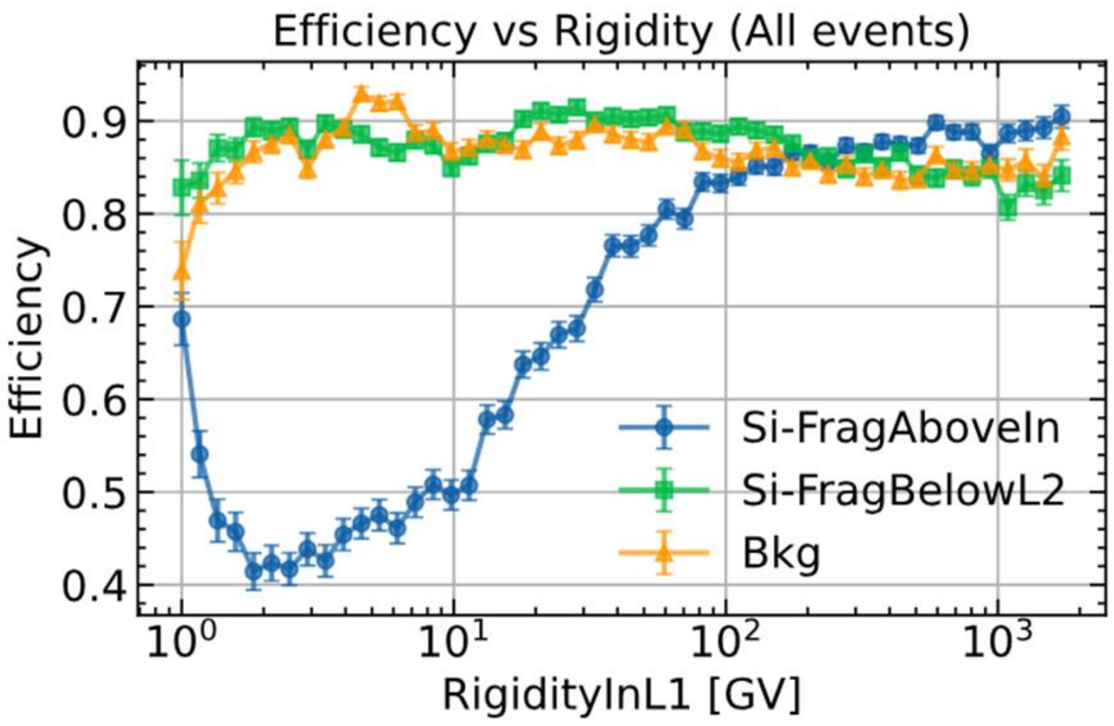
Task : Signal background discrimination (with MC samples only for now)

Signal (Si-FragAboveIn): primary Si survived below L2
Background 1 (Si-FragBelowL2): primary Si survived above L2
Background 2 (Bkg): other nuclei (Mg, Al, P, S)

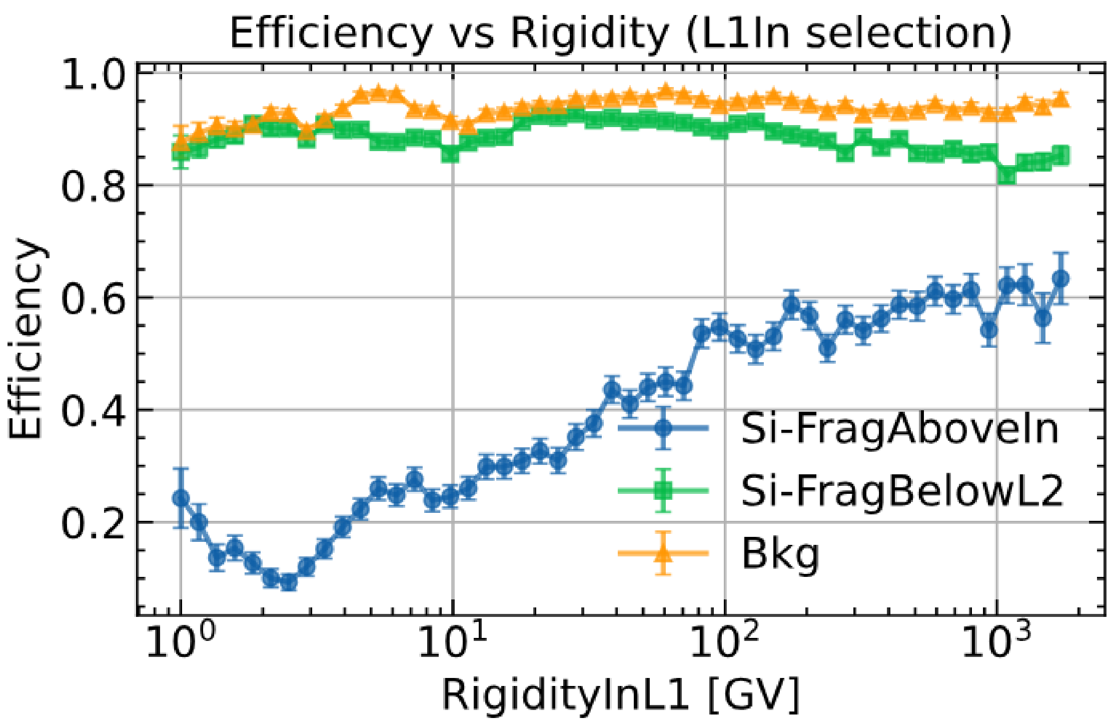


Model performance: model testing result

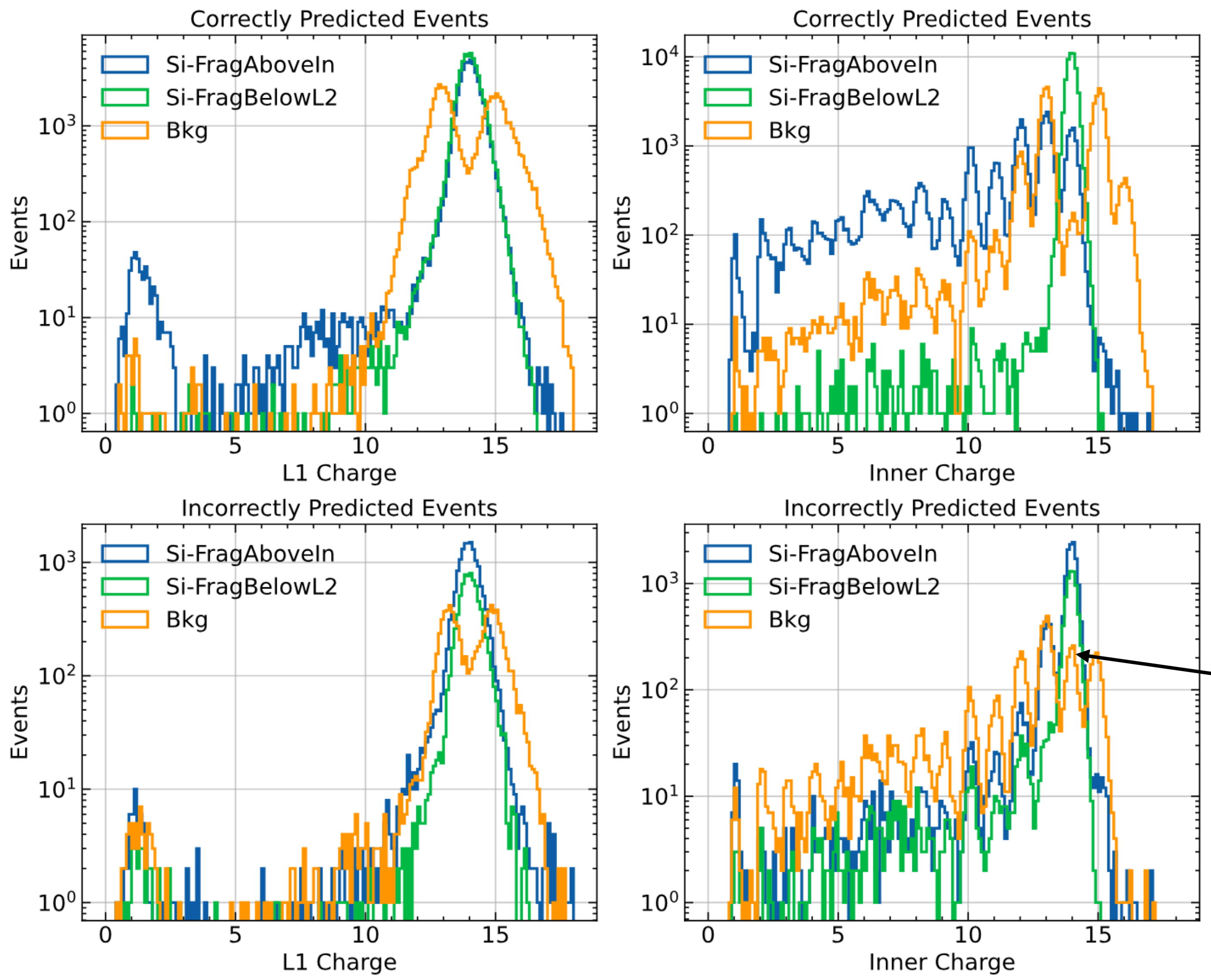
Performance results from all events (i.e. events with ML selection)



Performance results from sample selected with standard nuclei event selection



Model performance: comparison to standard charge reconstruction results



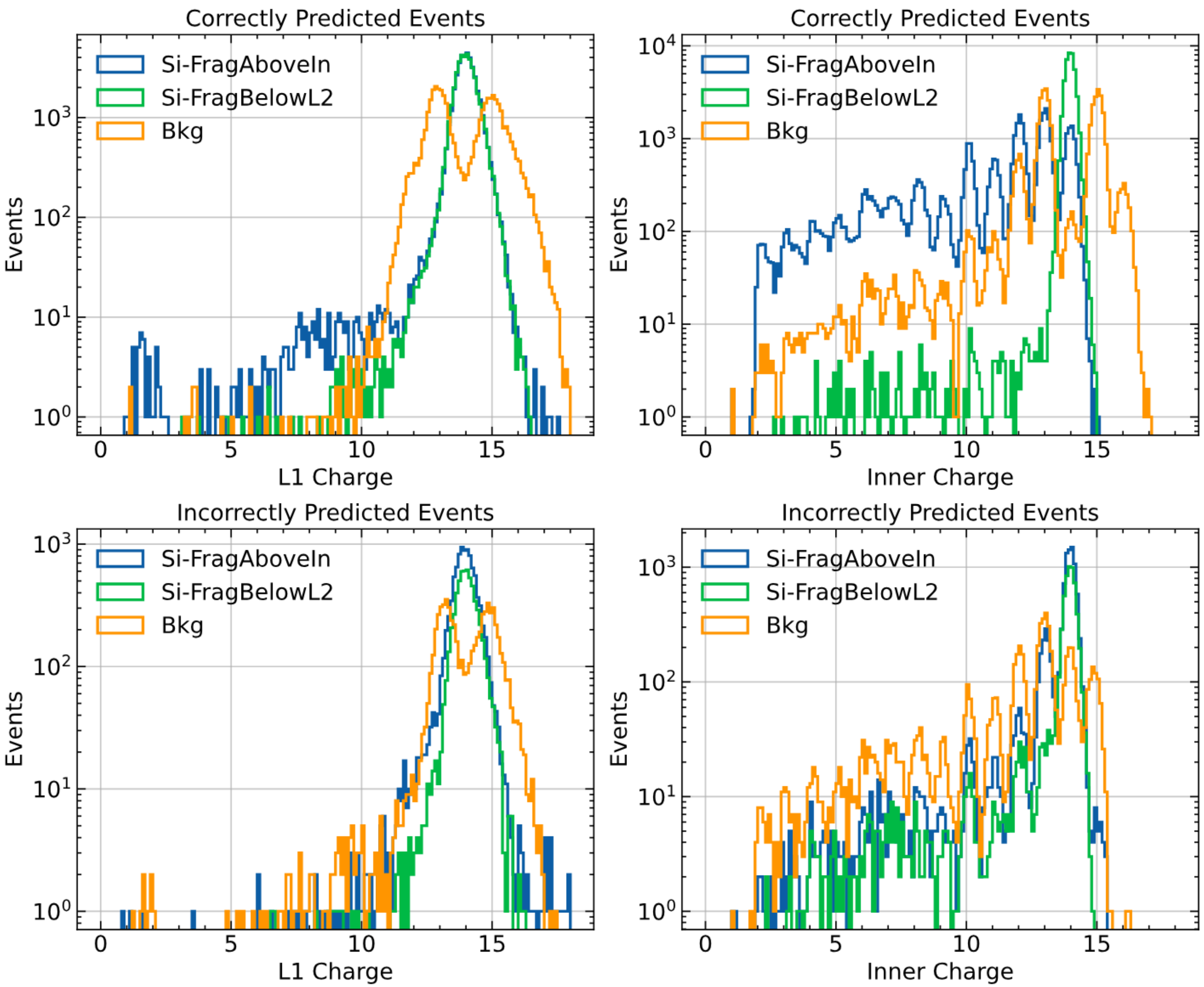
(
Low rigidity events:
- Small charge
- Large ADC
- Small number of clusters

)

Background nuclei (Mg, Al, P, S) with Inner charge == 14

Model performance: comparison to standard charge reconstruction results

Test sample selected with L1Inner Rigidity>10GV

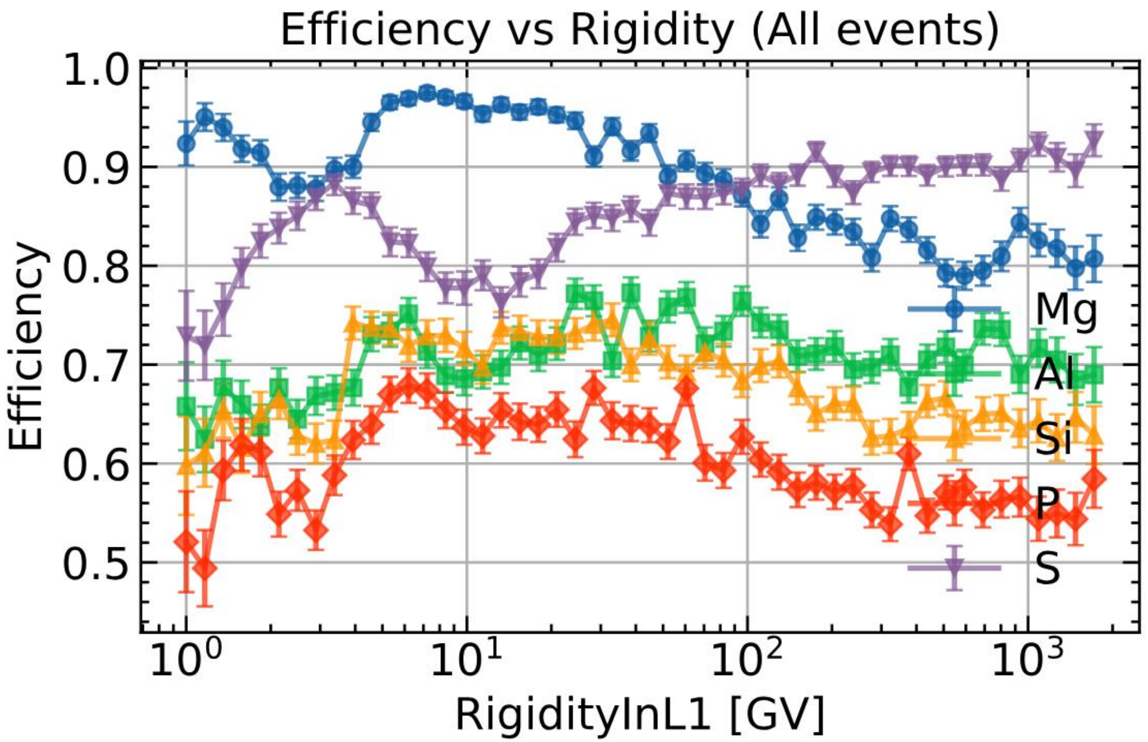


Optimization plan for Graph construction with ADC information:

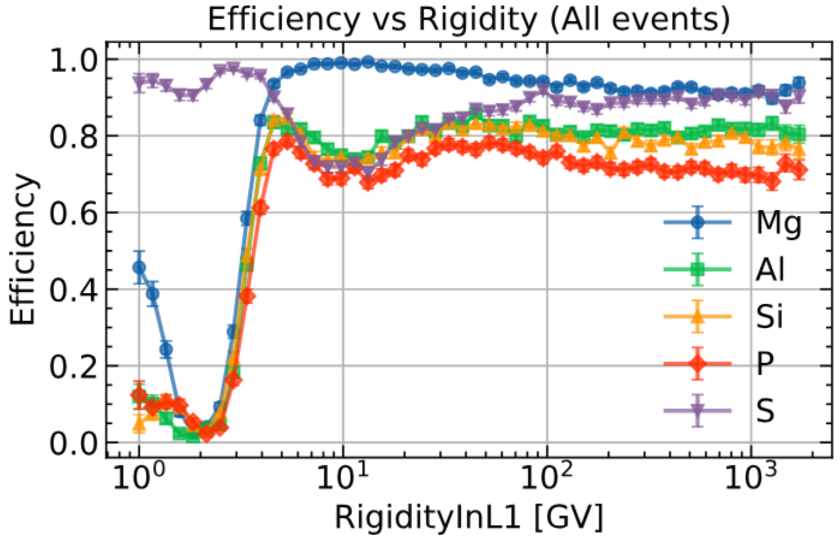
- **Optimization on number of clusters, adjacent strips and edges**
- Instead of connecting nodes based on cluster Edep ranking, → **connect nodes that are physically adjacent based on strip distance**
- Variable number of nodes/edge → **set a threshold based on Edep ratio or distance range**

Nuclei identification for Mg, Al, Si, P, S

Results (with clusters and TOF Beta)



Result (without TOF Beta)



Standard selection for reference