Application of the AI techniques to the reconstruction of the CLAS12 RICH data

Armen Gyurjinyan
INFN - Laboratori Nazionali di Frascati

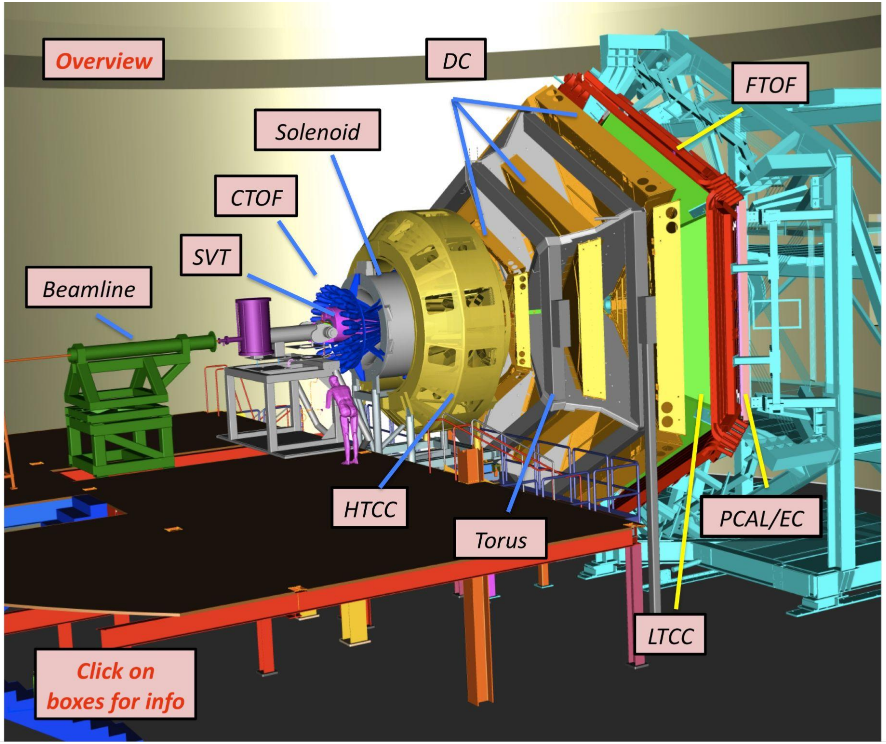
AANL, March 13, 2024

Outline

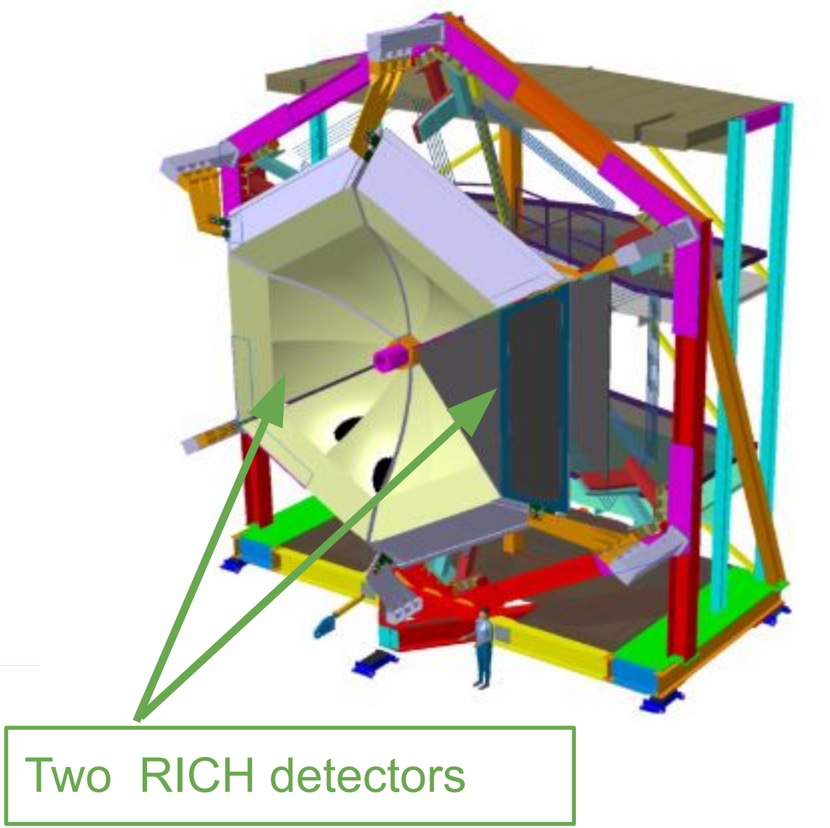
- CLAS12 RICH details and structure
- Current reconstruction with RayTracing
- Machine Learning application for alignment improvement
    - Alignment parameters from AI procedure
    - AI predictions vs current parameters
- Machine Learning application for PID
    - Data preprocessing for machine learning
    - Machine learning model
    - Preliminary results

CLAS12 RICH details and structure

CLAS12 RICH

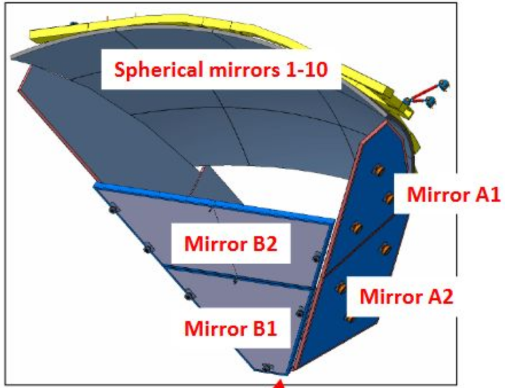


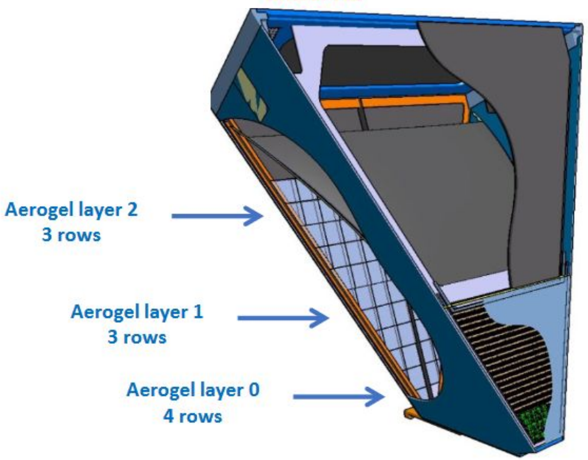
CLAS12 at JLab Hall B



RICH detector structure

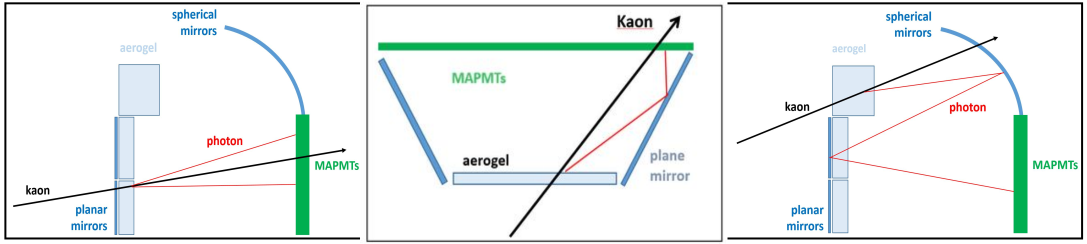
- 7 planar mirrors
  - Two on each side
  - Two frontal
  - One bottom
- 10 spherical mirrors
- 102 tiles of aerogels with nominal reflective index 1.05
- 391 MAPMTs total 25024 channels
- Time resolution 1 ns
- Cherenkov angle resolution 5 mrad





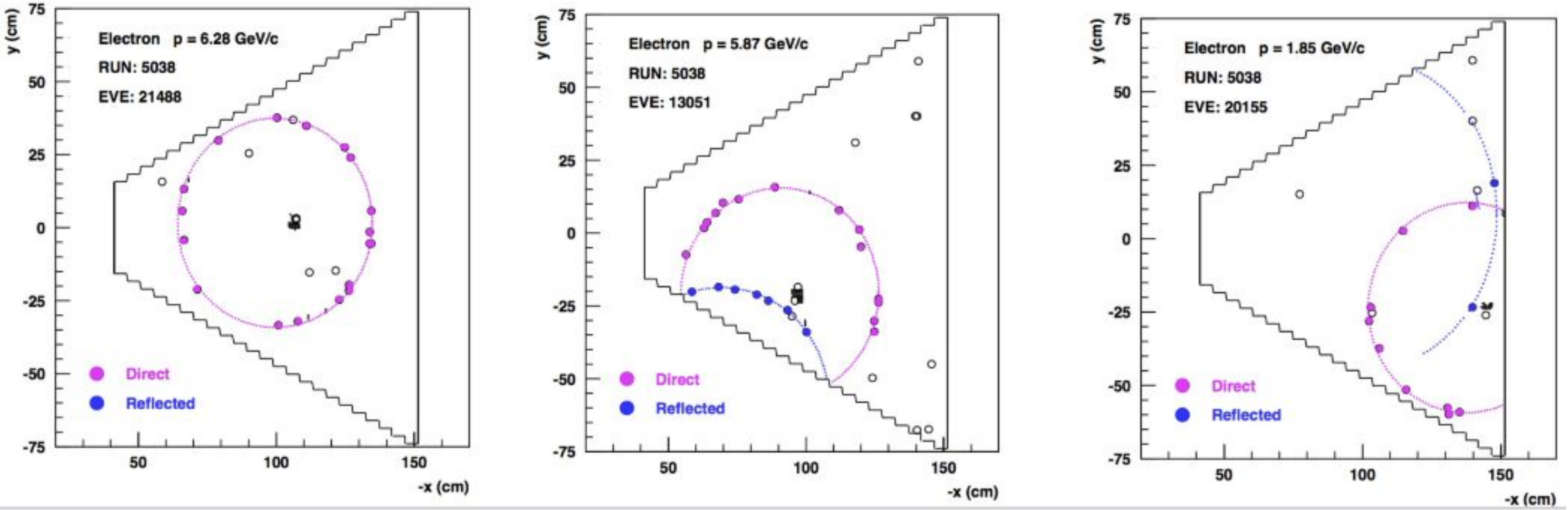
Current reconstruction with RayTracing

RICH Reconstruction



Photons detection topologies

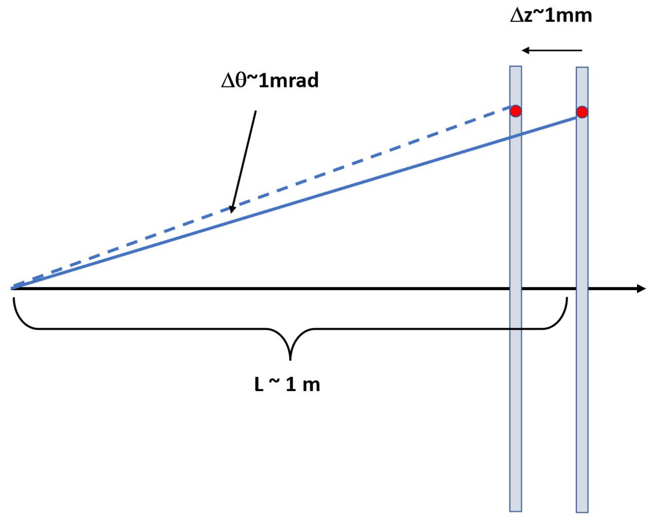
- Direct photons
- Single reflection from lateral mirrors
- Double reflections (spherical mirrors + frontal mirrors)
- And so on ….



RICH alignment

Each RICH component consist of 6 alignment parameters (Δx, Δy, Δz, $\theta_x, \theta_y, \theta_z$)

- Whole RICH with respect to CLAS12 (or MAPMT plane)
- 3 Aerogel planes
- 7 Planar mirrors
- 10 Spherical mirrors



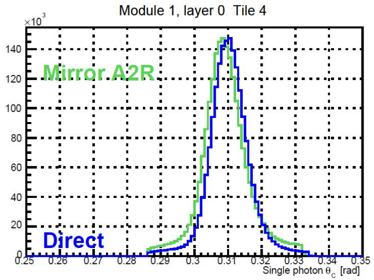
Total 126 RICH alignment parameters

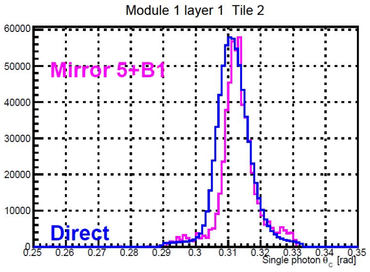
However sensitive parameters are Δz, θ_x, θ_y

Alignment strategy is:

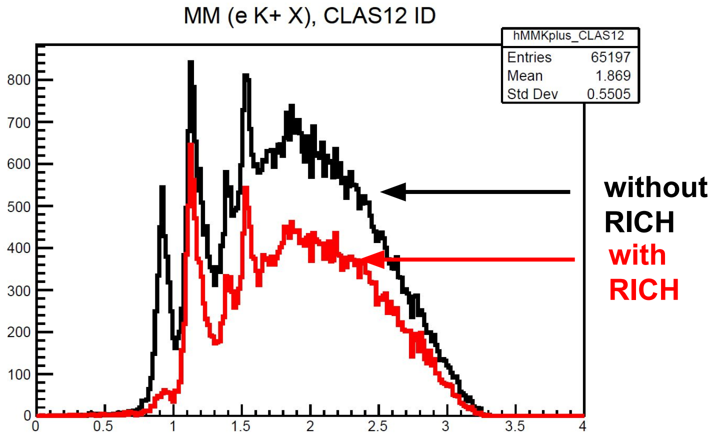
1. Align whole RICH with respect to CLAS12 with trajectory comparison.
2. Align the aerogel planes with direct photons and the known reflective index
3. Align the lateral mirrors with photons with 1 reflections assuming they have same Cherenkov angle as direct photons.
4. Align the spherical mirror with photons with 2 reflections, assuming they have same Cherenkov angle as direct photons

Current RICH performance



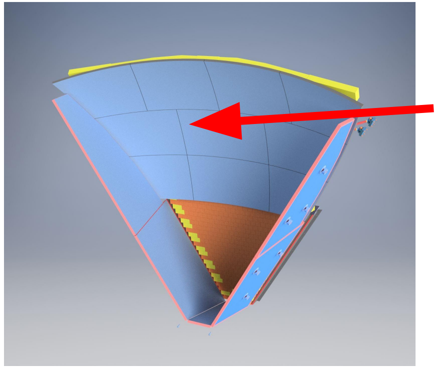


blue: zero reflections
green: 1 reflection
pink: 2 reflections



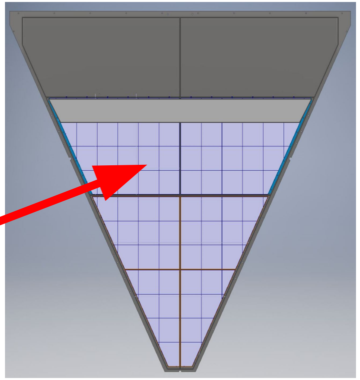
The alignment procedure does not take into account correlations among different mirrors

The upper part (large angles) of the detector is not properly aligned



Spherical mirrors upper and middle rows are not aligned

Aerogel Layer 2 not aligned



Machine Learning application for alignment improvement

RICH alignment parameters prediction with AI

Alignment parameters finding pipeline

Generate random alignment parameter combinations

Use generated parameter combination to reprocess real data

Train Neural Network model with alignment parameters and reprocessed data results

Genetic algorithm to find the best alignment parameters

Step 1: Alignment of the whole RICH to the CLAS12

Event selection:
- charged pions

Input parameters:
- the 6 alignment parameter: shifts and angles

Output parameters:
- the distance between the measured track cluster and the position of the projection of the tracks
- 391 values

Step 2: Alignment of the internal elements

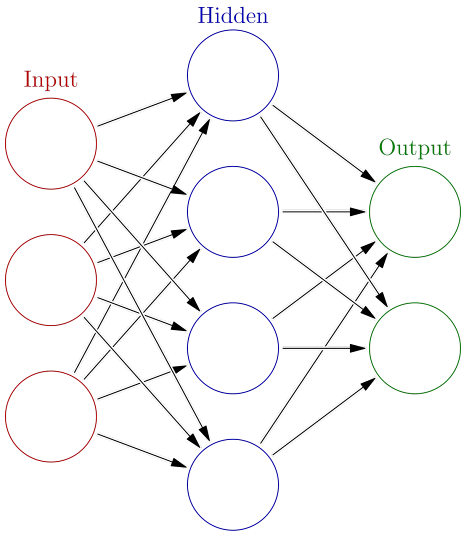
Event selection:
- electrons and protons

Input parameters:
- the alignment parameters of mirrors and aerogel planes (3 parameters each)

Output parameters:
- the difference between the measured Cherenkov angle and the expected one based on the known refractive index
- one value per aerogel tile and photon detection topology

AI model and genetic algorithm

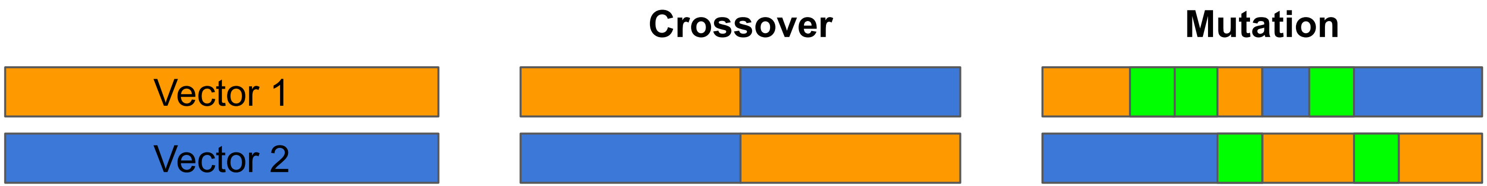
Fully Connected Neural Network



Genetic Algorithm procedure

1. Generate **n** input vectors
2. Calculate “goodness” of each input vector
3. Select best 20%
4. Crossover selected inputs to generate **n** new vectors
5. Mutation of crossovers
6. Calculate “goodness” of new **n** vectors
7. Select **n** best vectors from 2**n** vectors
8. Repeat from 3 to 7 steps until converges or max iteration number reached

“goodness” value = average of the neural network output. lower value better combination



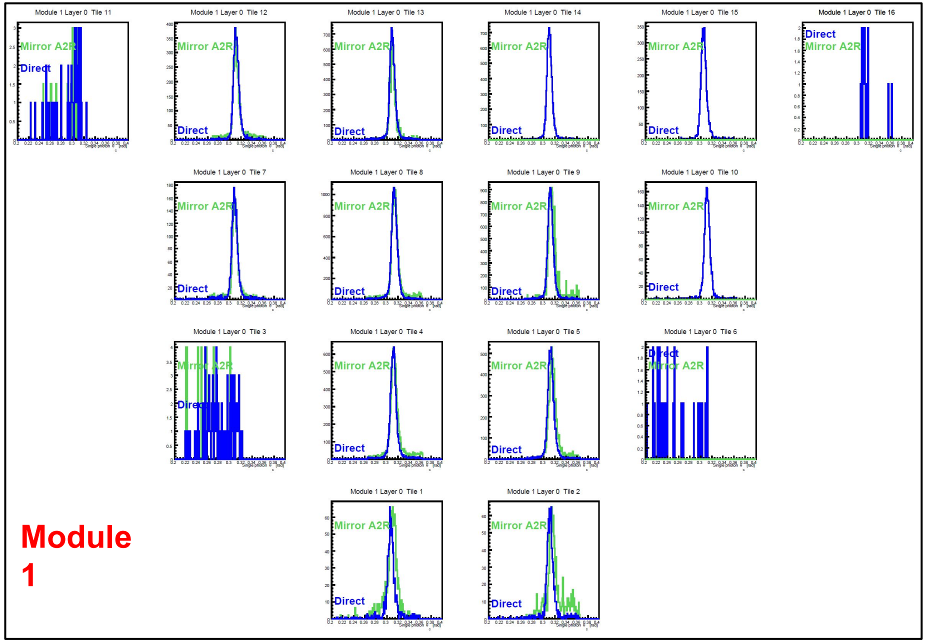
Preliminary alignment results with AI

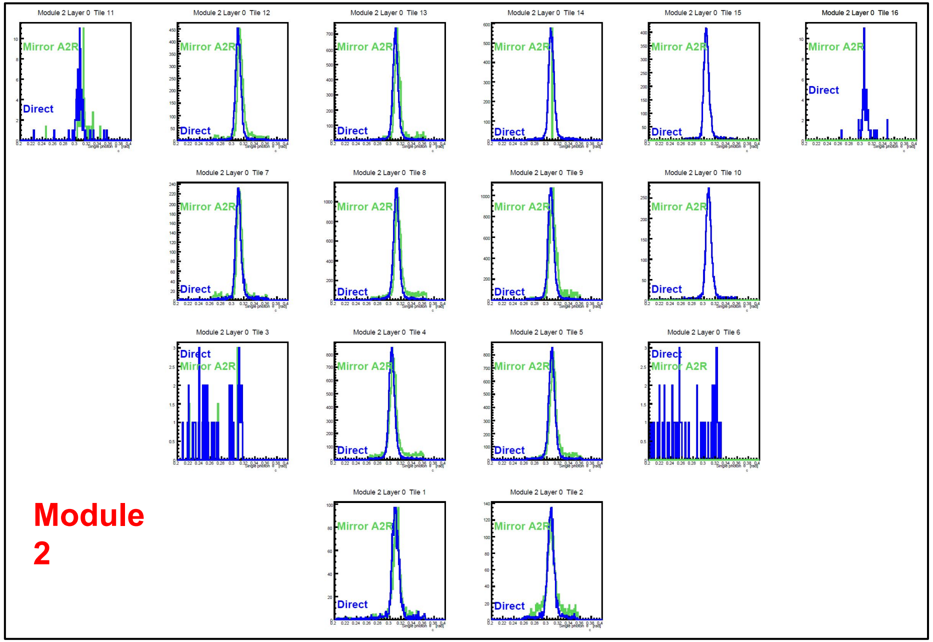
First AI algorithm tests with simple photon topologies: zero and 1 reflection

- results comparable with conventional alignment

Next step: apply to more complicated topologies where the conventional alignment fails

blue: zero reflections
green: 1 reflection





Machine Learning application for PID

Particle ID with ML algorithm

Photon hit pattern recognition to separate pi/K/protons based on machine learning

Advantages:
- real geometry knowledge not needed
- no need of accurate alignment

Critical points
- selection of known tracks with low misidentification to train the model
- low statistics in the edges (high momentum, large angles) of the detector that might be relevant for the physics
- develop tools to validate the results

Data filtering

ep -> eph+ (π-)

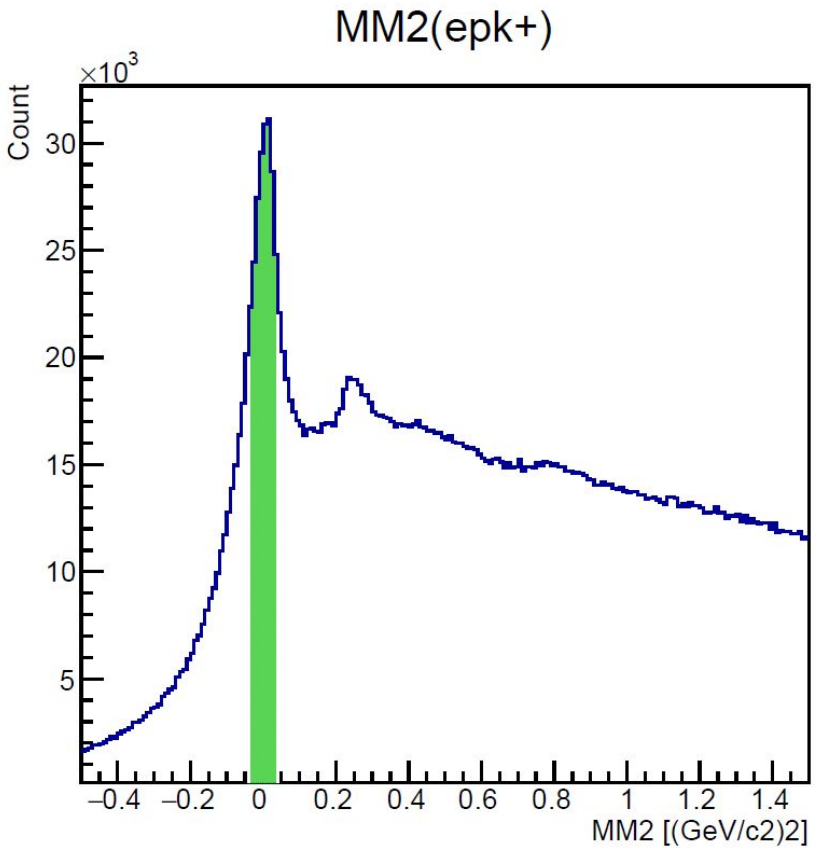
Track based filters

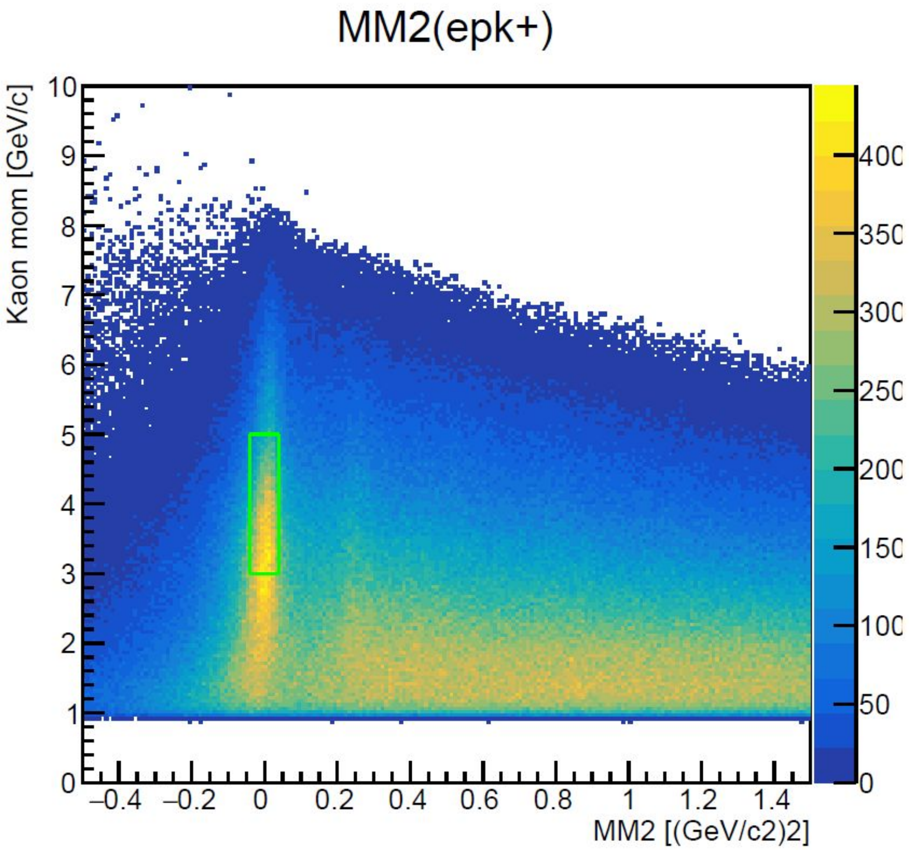
- 1.5 GeV < E(e) < 8 GeV
- **One** charged particle in the RICH
- At least one hit on MAPMT
- CLAS12 EB identifies as kaon or pion
- Missing **π-** cut for reactions **h+** kaon or pion

Hit based filters

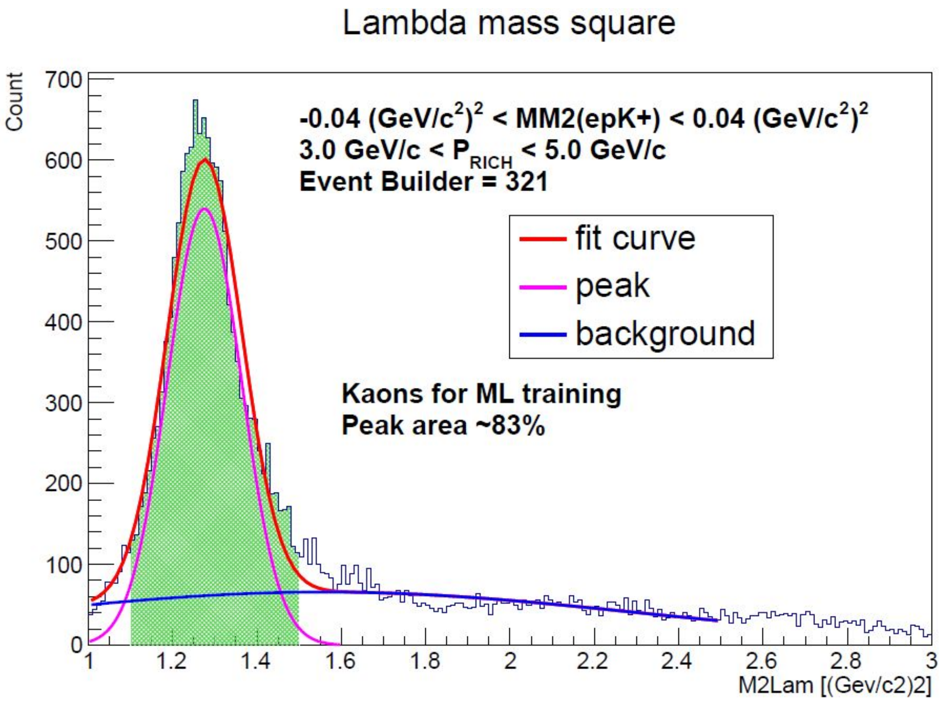
- Remove noisy anode hits
- Remove background hits based on timing

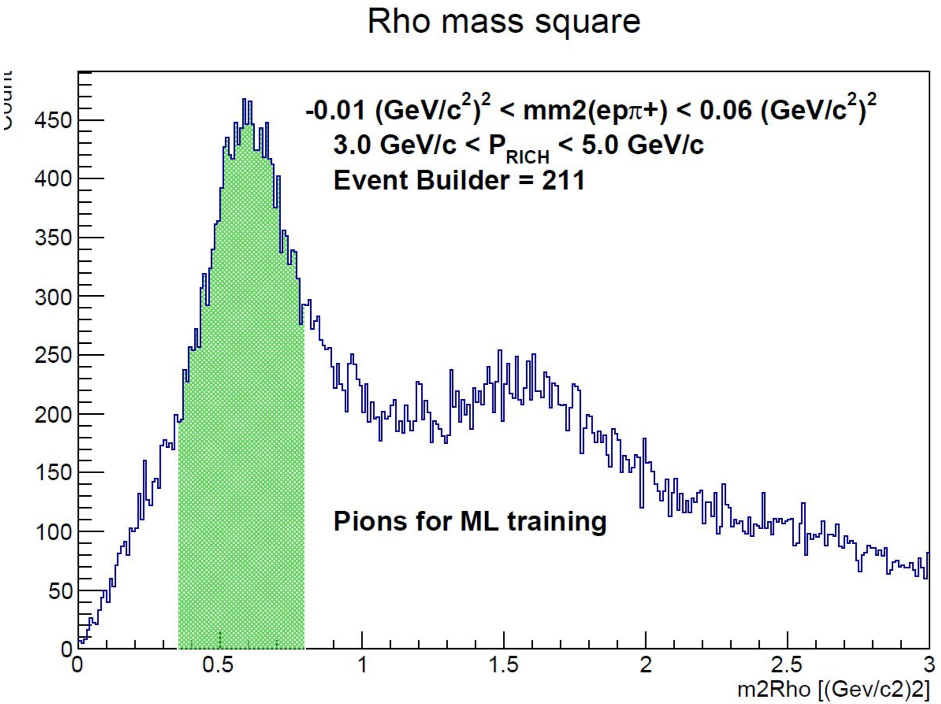
Kaon/Pion training data selection



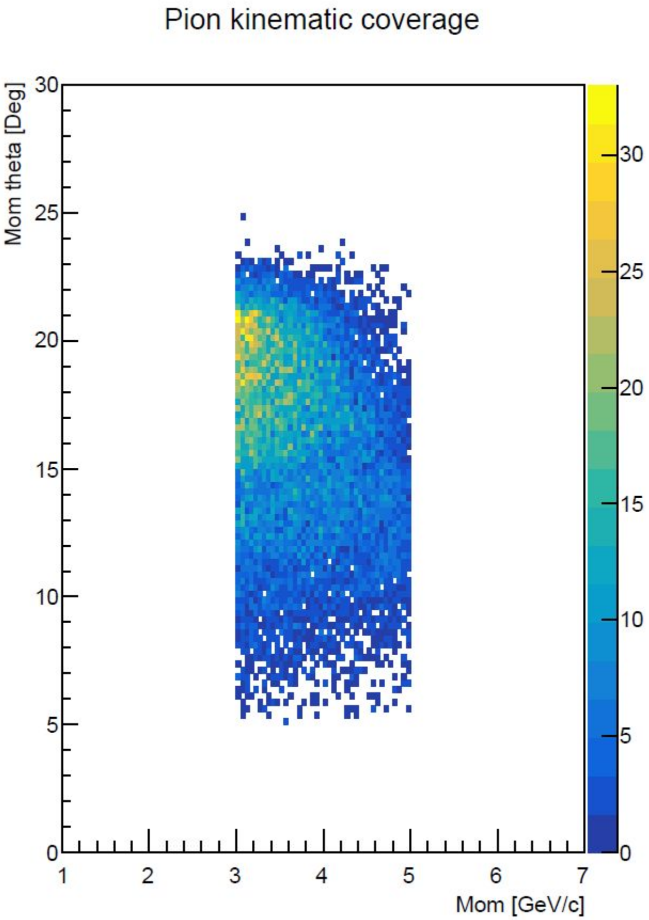


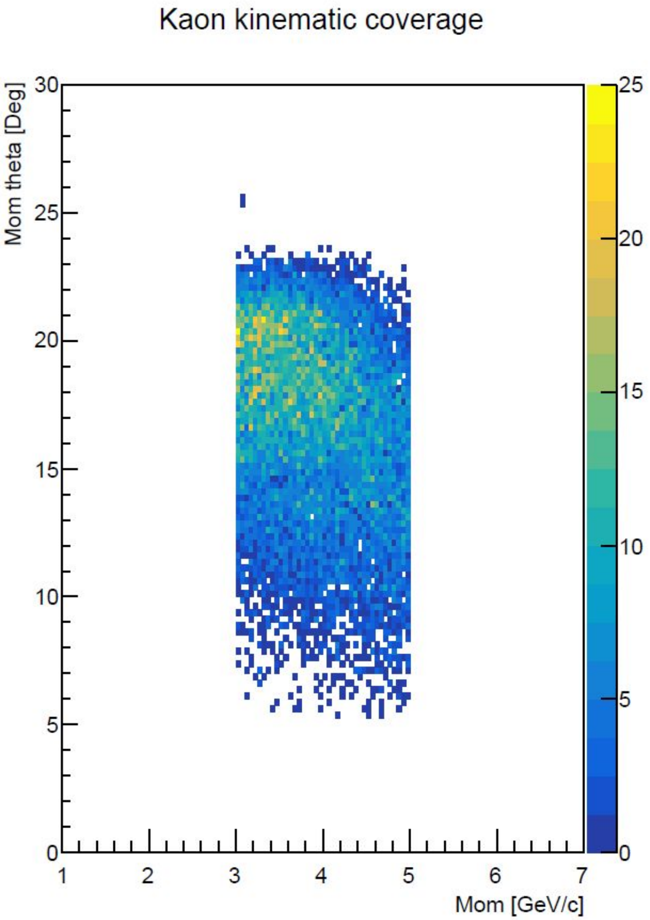
Kaon/Pion training data selection



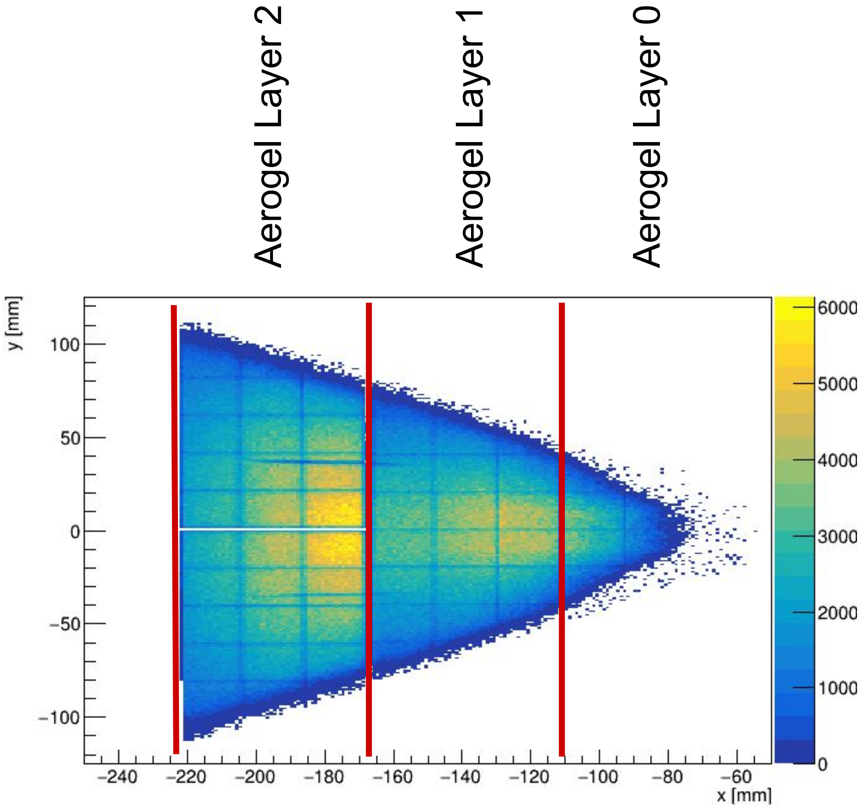


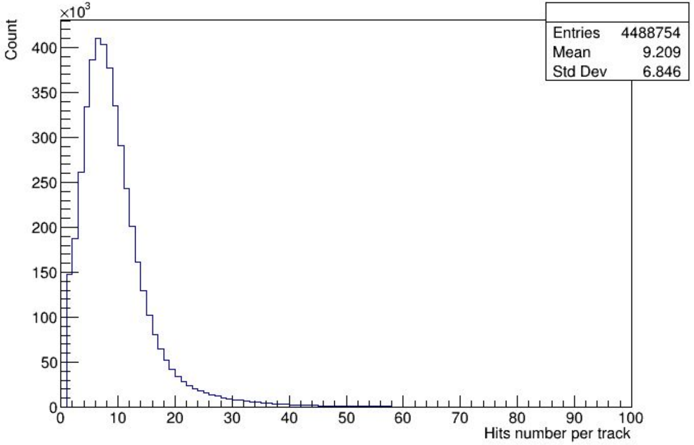
Kinematic coverage in training sample

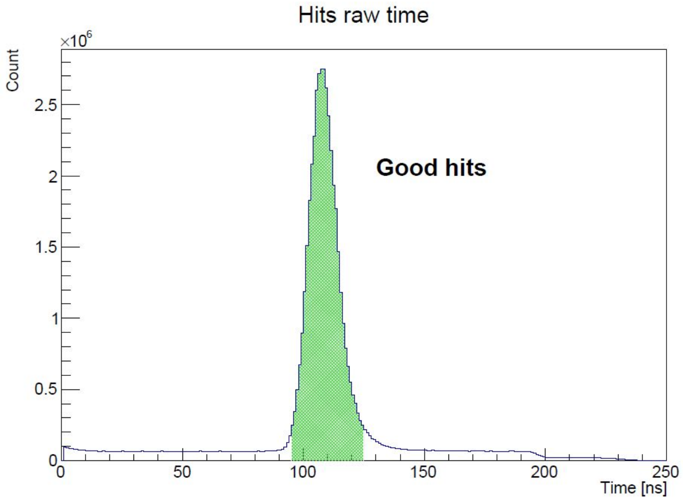




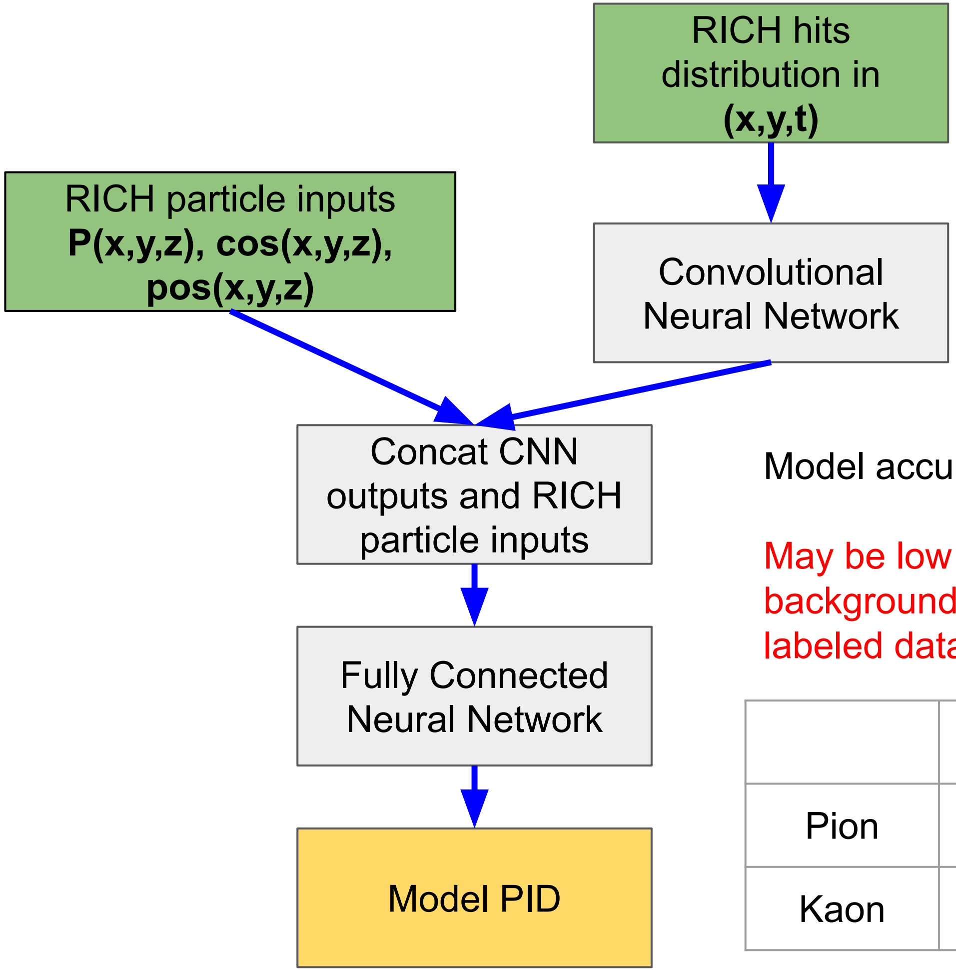
Hits distribution







Machine learning model



Training sample

Kaons : 14283

Pions : 16131

Average 10 hits per event

Model accuracy 66.7 %.

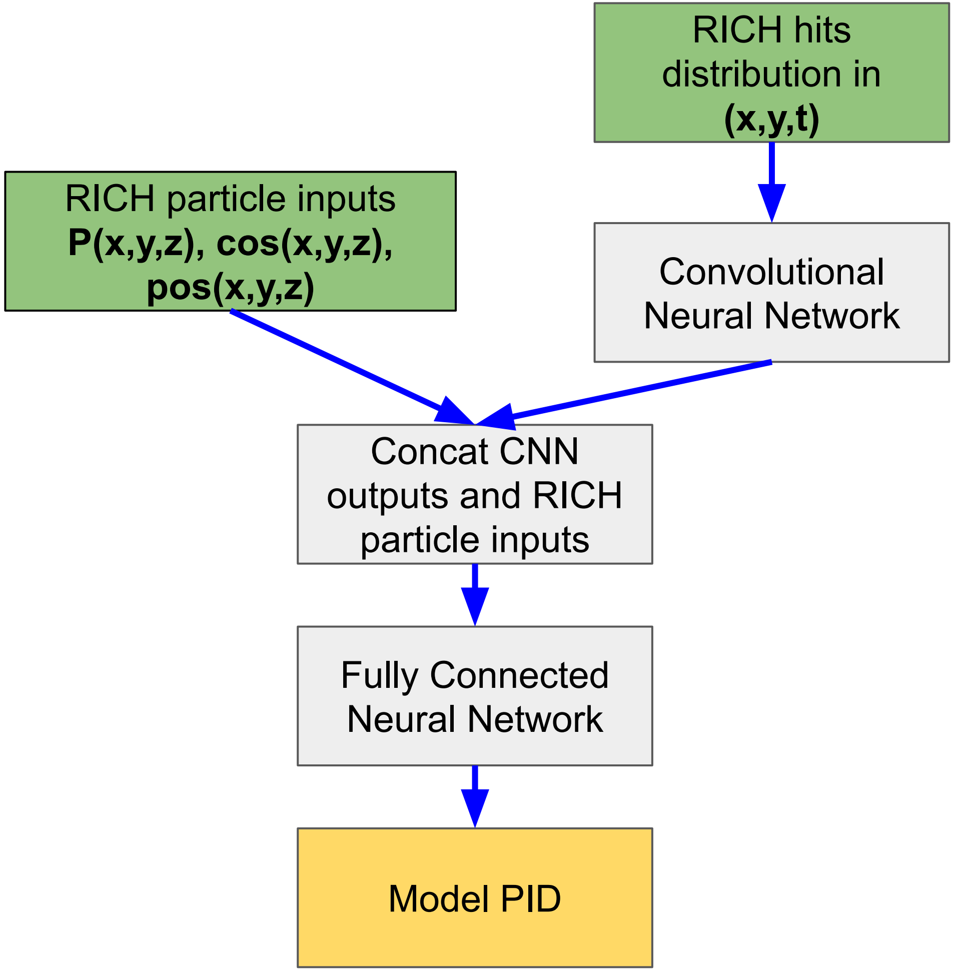
May be low because of background and wrong labeled data.

	Precision	Recall
Pion	67.8 %	66.7 %
Kaon	65.6 %	66.6 %

Precision - What is the probability that the model will predict label correctly.

Recall - What percentage of actual labels were predicted correctly.

Machine learning model

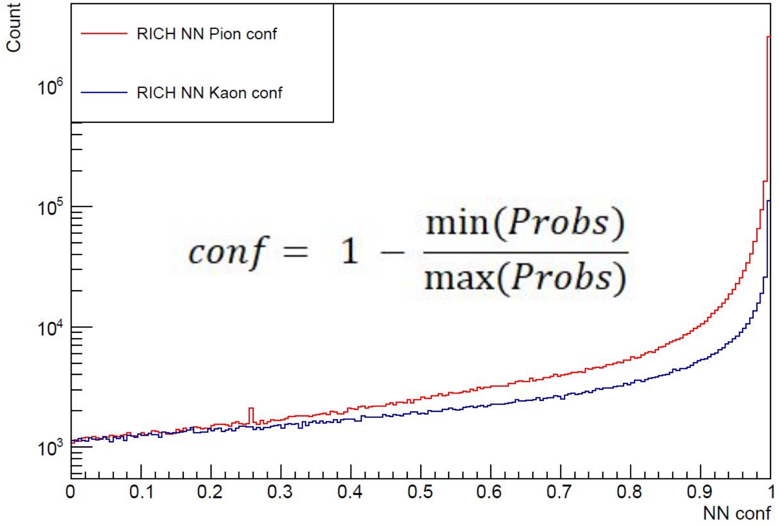


Training sample

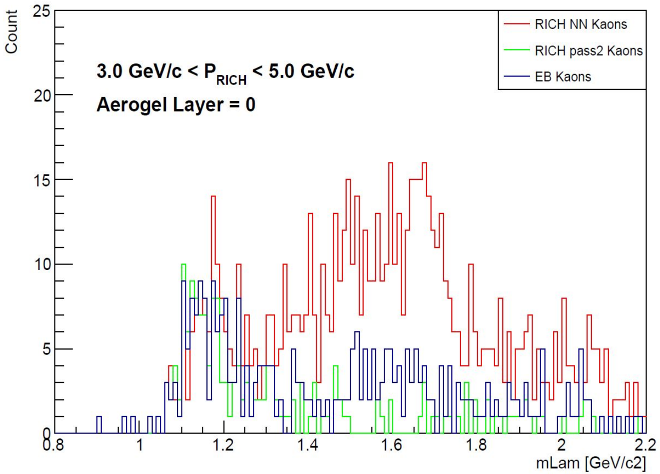
Kaons : 14283

Pions : 16131

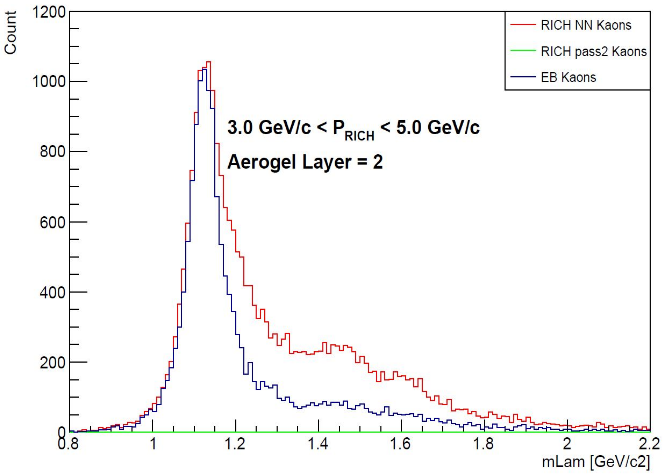
Average 10 hits per event

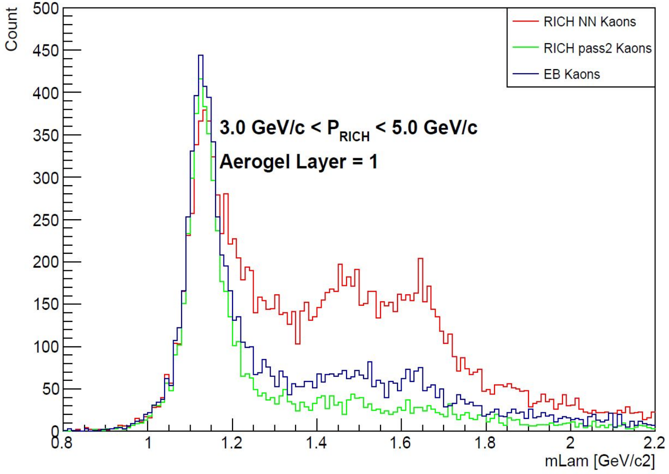


Preliminary results



Training region

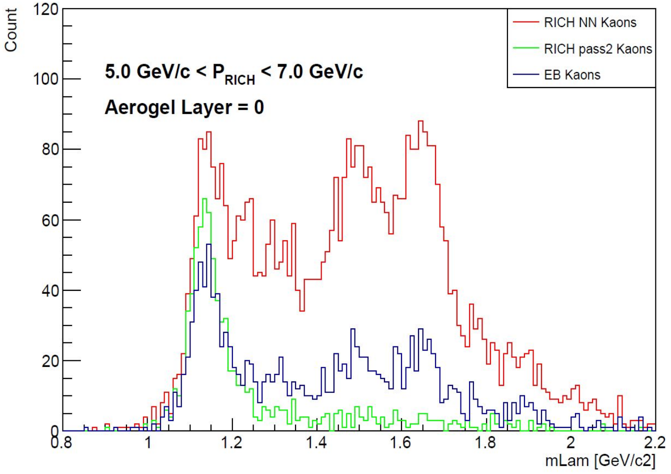


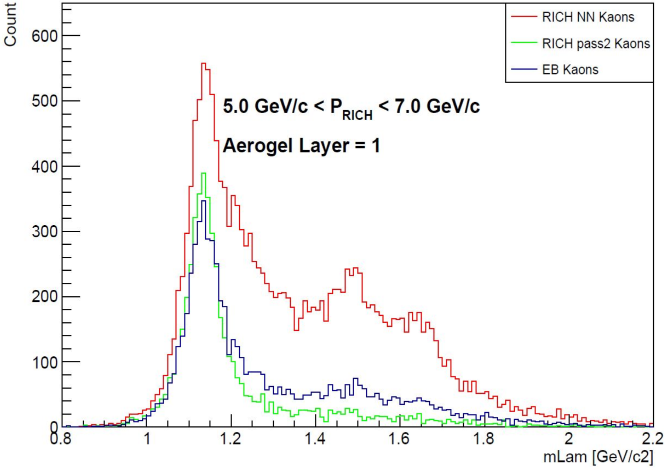


RICH NN model applied blindly on the CLAS12 data

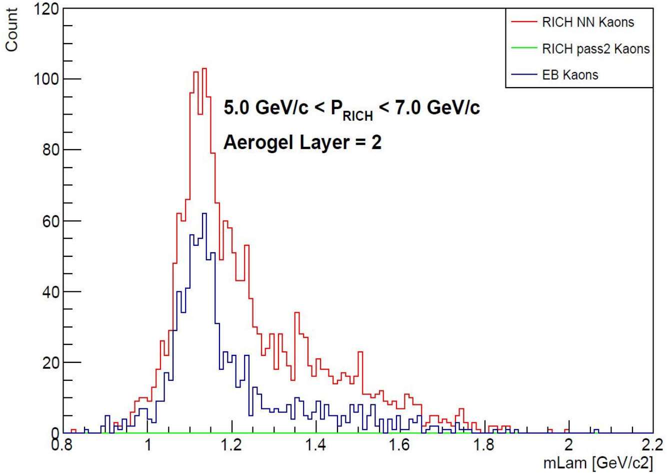
- RICH NN was able to correctly identify a large fraction of Kaon
- The level of misidentified kaons is however still large: need to improve the training sample quality

Preliminary results





Outside training region based on momentum



RICH NN model applied blindly on the CLAS12 data

- RICH NN was able to correctly identify a large fraction of Kaon
- The level of misidentified kaons is however still large: need to improve the training sample quality
- The quality of the results is slightly, but not dramatically, worse than in the training region

Conclusion

A new generation of tools based on ML algorithm is being developed and tested on the CLAS12 RICH data

- Alignment software
  - A neural network algorithm is trained on real data to model the complicated dependence of the reconstructed Cherenkov angle on the alignment parameters
  - Preliminary results on simple cases produced results comparable to the ones obtained with the conventional procedure
  - The AI approach hopefully will be able to take into account the correlations among the alignment parameters and allow to extend the coverage of the detector to the full kinematic range

- Particle identification
  - A hit pattern recognition ML algorithm is under development to perform particle identification
  - The new approach would allow to skip the complicated alignment process
  - The preliminary results are encouraging, but also emphasized the need of a highly pure selection of kaon, pion and proton samples to train the model

Thank you for you attention!
Questions?

Aerogel Layer 0 | **Aerogel Layer 1** | **Aerogel Layer 2**

P < 2 GeV/c

2 GeV/c < P < 3 GeV/c

3 GeV/c < P < 4 GeV/c

4 GeV/c < P < 5 GeV/c

5 GeV/c < P < 6 GeV/c

6 GeV/c < P