JLAB12 Activity report 2023

M. Mirazita (Resp.), A. Gyurjinyan (AdR), P. Rossi, S. Tomassini

1 Introduction

The JLAB12 group of LNF participates in the physics program carried on by the CLAS collaboration in the Hall B of the Jefferson Laboratory (JLab). The LNF group is involved in the data taking of the CLAS experiment and in the operation of the two modules of the Ring Imaging Cherenkov (RICH) detector of the CLAS12 spectrometer, that was completed in 2022.

2 CLAS data taking

During the 2023, two experiments took data in the Hall B of the Jefferson Laboratory with the CLAS12 spectrometer.

The first one, that ran from January to March and is called *Run Group C*, was started in June 2022 and ended in 2023 its first phase of data taking. It used a 11 GeV electron beam and longitudinally polarized NH_3 and ND_3 targets. A total of 5 different measurements dedicated to the study of the nucleon structure will be carried out with these data, including the Deeply Virtual Compton Scattering (DVCS) on proton and neutron, and the Semi-inclusive production of single hadron or di-hadrons in the Deep Inelastic Scattering region (SIDIS).

The second experiment, that run from September to December and is called *Run Group D*, used a 11 GeV electron beam and various different unpolarized nuclear targets. The goal of the experiment is the study of the Color Transparency in nuclei and the effect of the nuclear medium on the nucleon structure.

3 CLAS data analysis

One of the main milestones of the CLAS Collaboration during 2023 was the reprocessing of all the experimental data taken so far, to make them available for all the planned physics analysis. This project was carried out through a series of iterative steps that involved the evaluation of the performance of each individual component of the CLAS12 spectrometer (including the RICH) by developing new software tools and the improvement of the calibrations by reprocessing a small amount of data several times. The final step of the process was the presentation of the status of the various data sets in front of a review committee [1)] in charge to judge the results obtained and approve or defer the beginning of the reprocessing of the data. By the end of 2023, almost all the data collected since January 2017 and up to mid of 2020 have been reprocessed.

4 The RICH detector performance

The RICH detector is taking data together with the CLAS12 spectrometer in its final configuration with two modules in opposite sectors since June 2022, see Fig. 1.

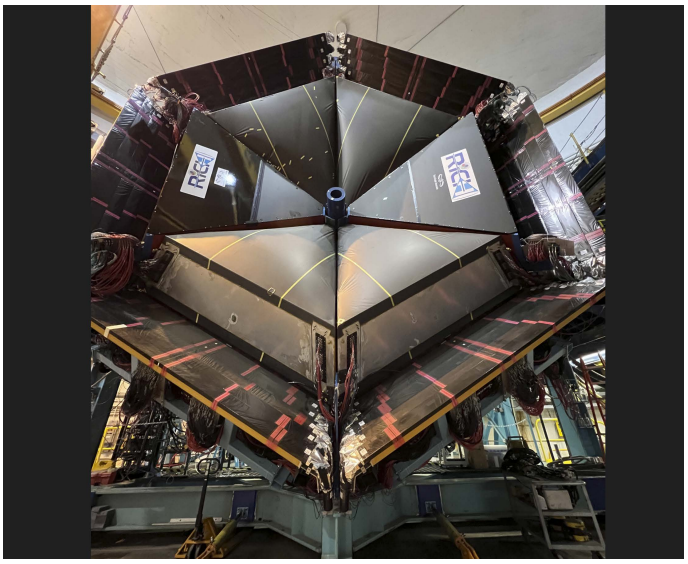

Figure 1: *The two RICH modules in the forward carriage of CLAS12.*

Each module is composed by an aerogel radiator, an array of multianode photomultiplier tubes (MAPMTs) for the Cherenkov light detection and a mirror system. All these elements are contained in a large trapezoidal box, of approximate height of 3.5 m, a large base of about 4 m and a depth of 1.2 m.

The radiator is composed by tiles with squared shape 20×20 cm^2 as well as smaller pentagonal, trapezoidal or triangular tiles to accomodate with the detector shape. The total number of tiles is 102, assembled in two sections: the forward angle one made by one layer with 2 cm thickness and the large angle one made by two layers with 3 cm thickness each.

The mirror system is composed by 10 carbon fiber spherical mirrors and 7 glass planar mirrors, for a total surface of about 10 m^2. The goal of the mirror system is to contain as much of the produced Cherenkov photons inside the detector and to direct them toward the photodetector array.

The photodector array uses 391 MAPMT Hamamatsu H12700, composed by a matrix of 8×8 matrix of pixel with about 6 mm pixel size, with a total of 25024 independent readout channels.

The readout electronics is based on the MAROC3 chip, a 64 channel microcircuit dedicated to MAPMT pulse processing. Each channel offers a low impedence adjustable gain preamplifier followed by a highly configurable shaping section, and produces prompt logic pulses from an adjustable threshold discriminator. The MAROC3 is configured and read out by a FPGA optically linked with the data acquisition node.

The running of the detector was quite smooth during the year and didn't require any particular intervention besides the normal maintenance operations.

The status of the detector is periodically checked by taking special calibration runs with the scaler readout system and measuring the pedestal and the dark count rate of each readout channel. An example of the results of these runs is shown in Fig. 2, where we show the dark count rate for the 64 pixels of one MAPMT from June 2022 to January 2024. No strong variations have been found, except for few (of the order of tenth out of 50000 channels) known hot channels.

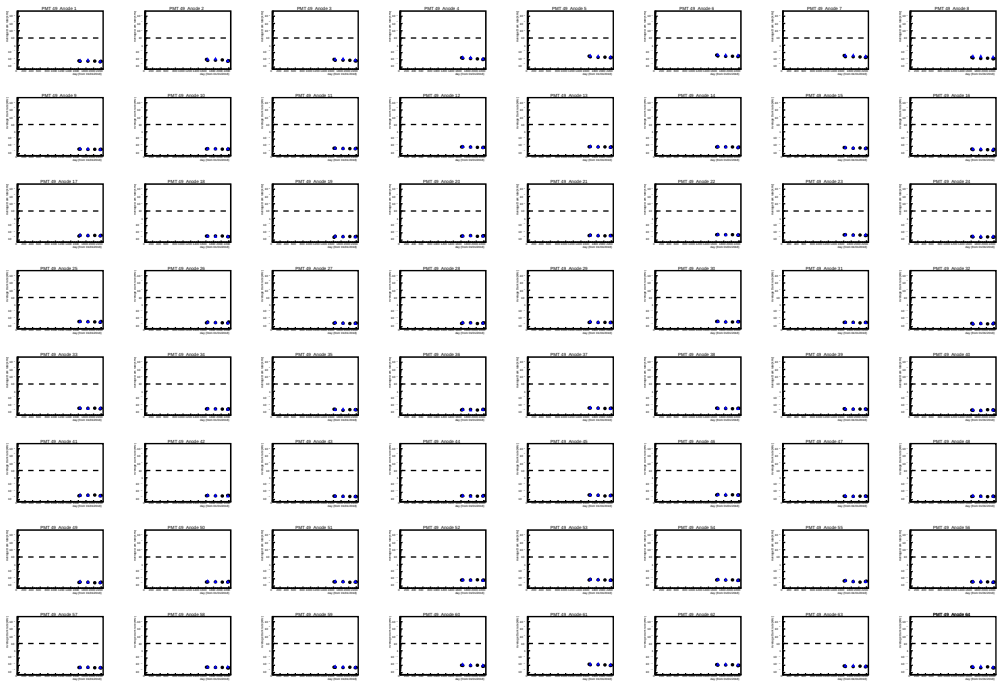

Figure 2: *The dark count rate measured in special calibration runs as a function of time for the 64 pixels of one MAPMT. The horizontal scale spans the time from June 2022 to January 2024 and the data points shown cover the year 2023, the vertical scale shows the rate in kHz. The dashed line indicates the 10 kHz limit set to identify hot channels.*

4.1 Software developments

One of the main activity during 2023 was the development of new software for the RICH in view of the final reprocessing of all the CLAS experimental data. These new tools included:

- the reorganization of the calibration database to include the second module;
- the validation of the RICH reconstruction software with data with two modules;
- the validation of the RICH calibration software with data with two RICH modules;
- the integration of the RICH data into the standard CLAS12 monitoring software.

We discuss in the next sections the results obtained by these new tools.

4.1.1 Time dependent monitoring of the RICH performance

The monitoring of the CLAS12 detector's performance is performed by using a java software that takes a small portion (less than few percent) of each experimental data run, produces plots of several key quantities and shows them as a function of the run number. For the RICH, we use SIDIS events $eN \to e'h^{\pm}X$ and we identified the following quantites:

1. Level 1: photon hit reconstruction
    - reconstructed hit time mean and RMS;

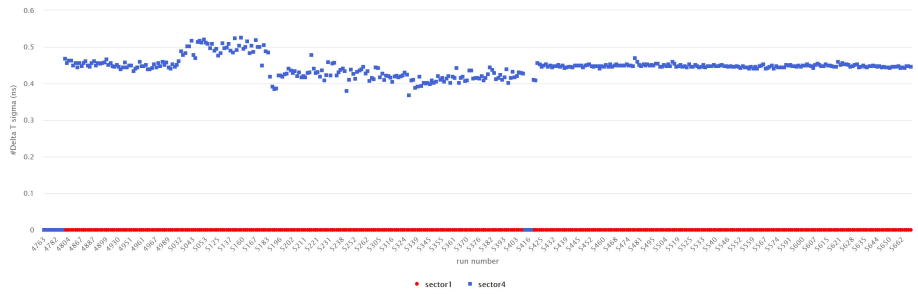

Figure 3: *The average RICH time resolution as a function of the run number.*

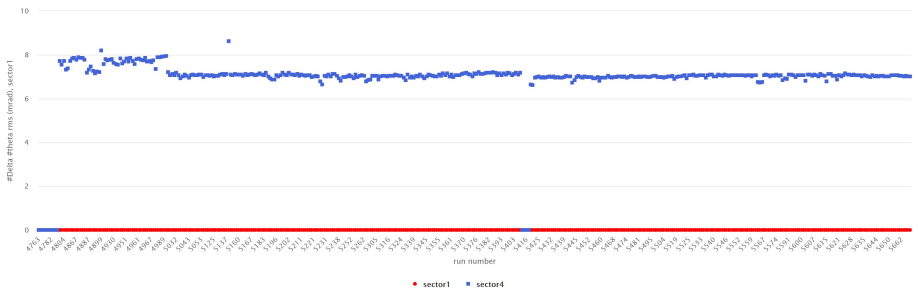

Figure 4: *The average Cherenkov angle resolution measured for photons detected without reflections as a function of the run number.*

- mean and sigma of the reconstructed Cherenkov angle for photons detected with zero, one or two reflections;
- mean number per track of Cherenkov photons with zero, one or two reflection;
- mean and sigma of the matching between the measured track cluster position and the projection of the track to the RICH photodetector plane;

2. Level 2: particle identification

   - mean number of pions, kaons and protons per electron trigger;

We show in the following few example of the monitoring plots obtained using the data taken during the fall 2018 data period with 11 GeV electron beam and an unpolirized liquid hydrogen target. Note that during this period only one module in the CLAS12 sector 4 was installed.

In Fig. 3 we show an example of the time resolution as a function of the run number. It is very stable over the whole period and around 0.5 ns, well below the desing value of 1 ns.

In Fig. 4 we show the resolution of the measured Cherenkov angle for photons detected with no reflections on the mirror system. In this plot, photons produced from all the 102 aerogel tiles are summed up. The average value, about 7 mrad, therefore includes contributions from the real detection resolution (of the order of 5 mrad, within specifications) folded with the small tile-to-tile variations (few additional mrad). The measured values are quite stable with the run number.

In Fig. 5 we show the average number of kaons identified by the RICH per trigger electron. The jump around run number 5420 corresponds to the reversal of the torus magnetic field, which implied an overall larger event rate (and therefore a reduction in the statistical fluctuations) but also a lower normalized production rate.

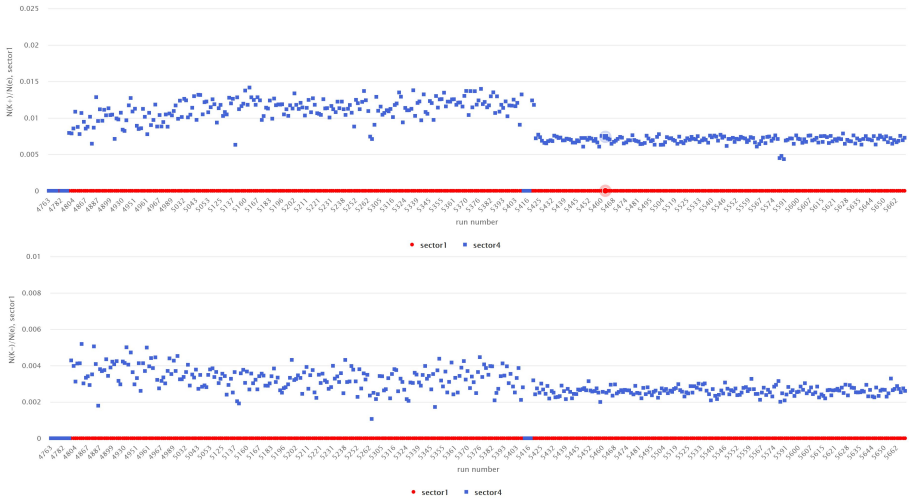

Figure 5: *The average number of K^+ (top plot) and K^- (bottom plot) per trigger electron as a function of the run number.*

4.1.2 First look at the second module data

The second RICH module was installed just before the start of the data taking with a longitudinally polarized target. The calibration of these data was not completed yet by the end of the year, however a first look at the basic performance of the RICH was still possible.

A preliminary rough alignment of the internal elements, radiator planes and mirrors, of the second RICH module was performed on the lower part of the detector, where most of the photons are detected without reflections or after only one reflection on the planar lateral mirrors. An example of the quality of the results obtained is shown in Fig. 6. Here, we show the measured Cherenkov angle for photons produced on aerogel block 5 of the more forward section. Plots of the top row are for the module 1, the ones of the bottom row for module 2. The blue histogram (always the same in the three plot of each row) is from photons detected with zero reflections. The three blue histograms are for photons detected after one reflection on the left, bottom and righ planar mirror, respectively. We see that position of the two histograms in each plot agrees within few mrad with similar widths and that the two modules have very comparable performance.

5 Machine learning development for the RICH software

At the end of 2022, two new projects were started, that aimed at the implementation of Machine Learning (ML) techniques in the RICH alignment and reconstruction software.

5.1 RICH alignment with ML technique

One of the main limiting factor in the current performance of the RICH detector is the quality of the alignment of the complicated radiator and mirror systems. The conventional alignment method we used is composed by two steps.

The first step is the alignment of the whole detector (i.e. the photodetector plane) to the CLAS12 tracking system. The procedure consists in generating a regular grid in the three coordinate of the position of the MAPMT plane and minimizing the distance between the track clusters measured by the RICH and the projection of the track on the plane.

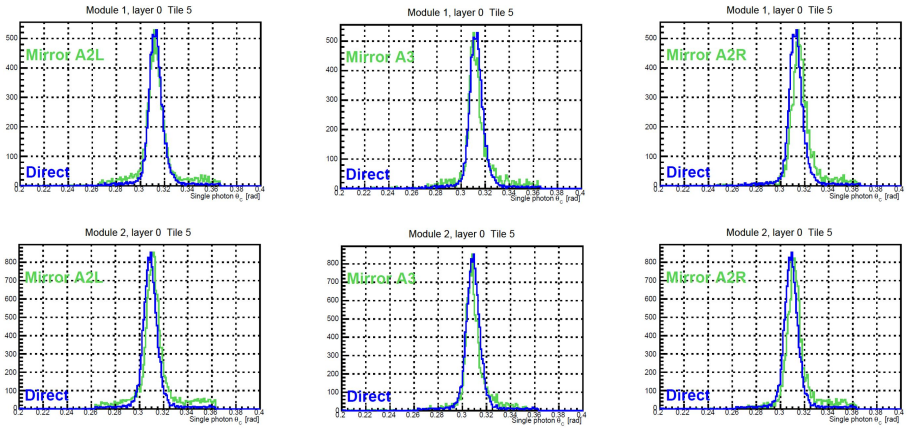

Figure 6: *The Cherenkov angle distribution measured for photons produced on one aerogel block of the more forward section of module 1 (top plot) and module 2 (bottom plot). The blue histogram is for photons with zero reflections, the green histograms are for photons with one reflection on the left (left plot), bottom (central plot) and right (right plot) planar mirror.*

The second step is the alignment of the internal element of the detector so that the reconstructed Cherenkov angle of the photons produced on one aerogel tile is the same regardless of its path before the detection. An iterative procedure has been set up, in which the radiator planes are aligned first using direct photons produced by electrons and protons, then one by one the lateral mirrors are aligned by using photons with one reflection on the mirror under study. Finally, the spherical mirrors are aligned, using photons with two reflections, the first one on one of the frontal mirrors ad the second one on the spherical mirror under study.

Two main issues have been found in this procedure. The first issue is that for the large angle radiator section the statistics of direct photons is very low. The second issue is that the iterative procedure can't take into account the correlations between the mirror and the radiator plane positions, therefore complicated photon detection topologies with more than 2 reflections remains not well reconstructed. As a results, the RICH data reconstructed so far miss all the large angle part of the detector.

To overcome these issues, a new alignment method based on ML algorithm has been implemented. In this method, a large set of random alignment parameters is generated and used to reconstruct a small selected portion of the real data. Then, these data are used to train a Neural Network (NN) model, which takes the alignment parameters as input and produces as output a number of quality parameters. These quality parameters are chi-square quantities calculated from the track matching (for the alignment of the whole RICH) or from the difference between the measured and the expected Cherenkov angle (for the internal element alignment). The data set is split in two parts, a training sample (typically 80% of the data) used to train the model and a validation sample (typically 20% of the data) used to validate the quality of the curremt iteration of the training. Once the training of the model has been completed, a genetic algorithm is used to find the best set of alignment parameters, defined as the ones that minimize the total chisquare.

The advantage of the method is that it automatically takes into account all the correlations among the alignment parameters and that it doesn't have a pre-defined shape, therefore it should be able to reproduce the complicated dependence of the measured Cherenkov angles on the radiator and mirror positions. On the other hand, one has to be very careful in the selection of the training data and also to be sure to not over-train the model. This is shown in Fig. 7, where we show the

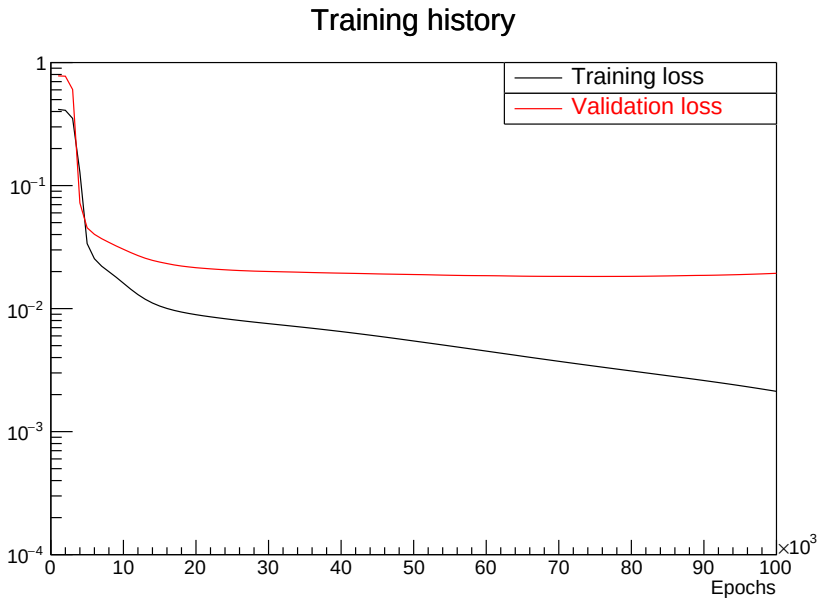

Figure 7: *Training and validation loss as a function of the epochs from the training of the ML model used to calculate the whole RICH alignment parameters.*

training and validation loss (which are basically the normalized total chisquare calculated from the training and validation samples) as a function of the training iterations (or epochs). At the beginning both losses start to decrease together, but at some point in the training (after about 50k iterations in the figure) the validation loss inverts the slope and starts to increase. This is because, as we go on with the training, the model is able to follow better and better the local statistical fluctuations in the quality parameters of the training sample, loosing the information on its real physical dependence on the input parameters.

This new approach has been tested to reproduce the already available alignment parameters of RICH module 1 obtained from old data, with promising results. Therefore it is now being used to align the new experimental data taken in 2023, and a sample of the preliminary results are shown in Fig. 6.

5.2 RICH particle identification with ML technique

To completely avoid the problems in finding the best alignment of the complicated optical system of the RICH, a completely new approach to the particle identification has been investigated. The idea is to apply a hit pattern recognition based on ML algorithm to identify pions, kaons and protons. The algorithm is a combination of a convolutional NN followed by a fully connected NN that takes as input the particle information and produces as output the best particle ID. The input information includes track momentum, position and entrance angles on the radiator plane, photon hit distribution on the RICH photodectors, number of Cherenkov photon hits, etc. The clear advantage is that a precise alignment of the RICH is not necessary anymore.

The critical point of this approach is to select a sample of known particles as clean as possible to train the model. An initial test of the π^+/K^+ separation capability has been done. The kaons have been selected by looking at $ep \to e'\Lambda K^+$ events, see the left plot in Fig. 8. The pions are selected by looking at $ep \to e'\rho^0 p$ events, see the right plot in Fig. 8. A cut on the momentum

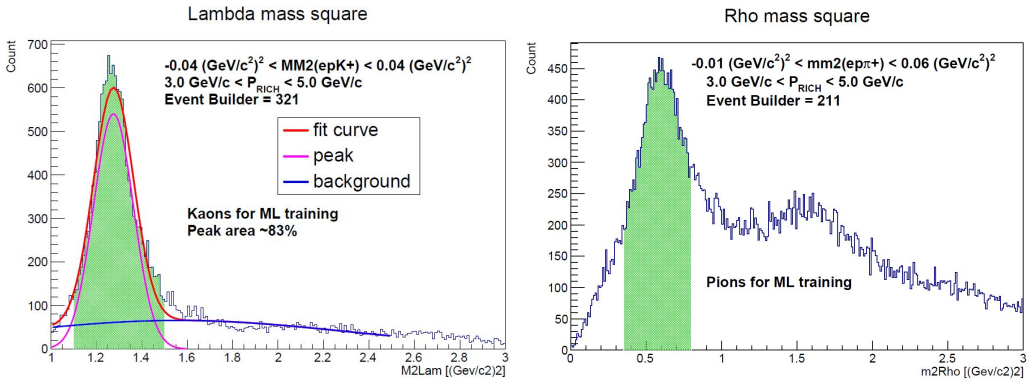

Figure 8: *Selection of the particle samples for the PID model training. Left plot: missing mass of the* $ep \to e'K^+X$*, the training kaon sample is selected in the green band. Right plot: missing mass of the* $ep \to e'pX$*, the training pion sample is selected in the green band.*

between 3 and 5 GeV/c is also applied.

The quality of RICH NN trained model is quantified by two parameter:

- precision: the probability the the model predicts the ID correctly, i.e. what fraction of identified kaons (or pions) were actually kaons (or pions);
- recall: the probability that the model prediction ID is correct, i.e. the fraction of the real kaons (or pions) that are correctly identified.

For each ID, a confidence level defined as

$$C = 1 - \frac{P_2}{P_1} \tag{1}$$

is provided, where P_1 and P_2 are the probabilities for the best and second best particle hypothesys. In Fig. 9 we show the confidence level obtained in this first test for the kaon and pion training sample. It peaks to 1 for clearly identified particles, while zero means equal probability for the best and second best hypothesys. The precision and recall for pions and kaons are both around 0.67, mostly limited by the contamination of misidentified kaons in the training sample.

Although still at a very preliminary stage, it is interesting to see that the model is already able to provide reasonable kaon identification in a region of the RICH where the conventional reconstruction based on the ray tracing is not working yet due to the issues with the alignment. In Fig. 10 we show the missing mass of $ep \to e'K^+X$ events in the Λ mass region. The blue histogram is obtained using the kaon ID provided by the CLAS12 time-of-fligh system and without the RICH. The red histogram is the results of the RICH NN kaon ID, where the Λ peak has significantly higher statistics than the blue histogram with similar width and background level.

References

1. Marco Mirazita is member of the so-called pass-2 cooking review committee.

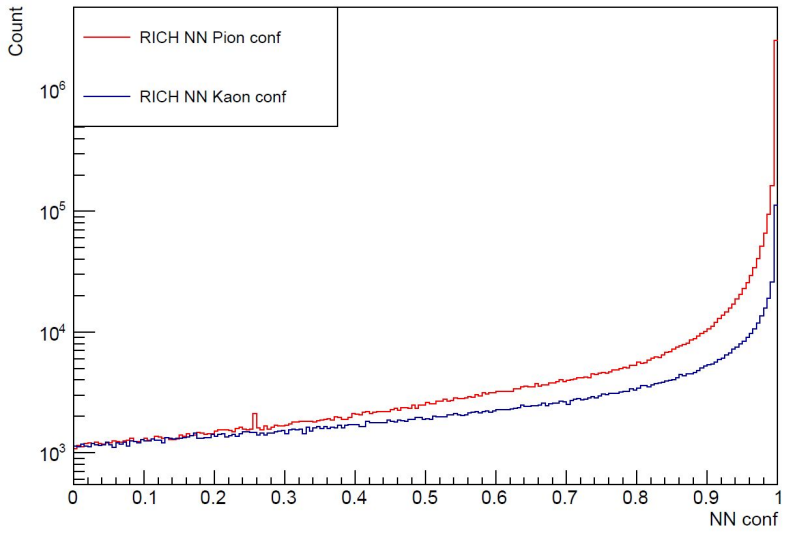

Figure 9: *Confidence level of the RICH NN particle ID model for pions (red) and kaons (blue).*

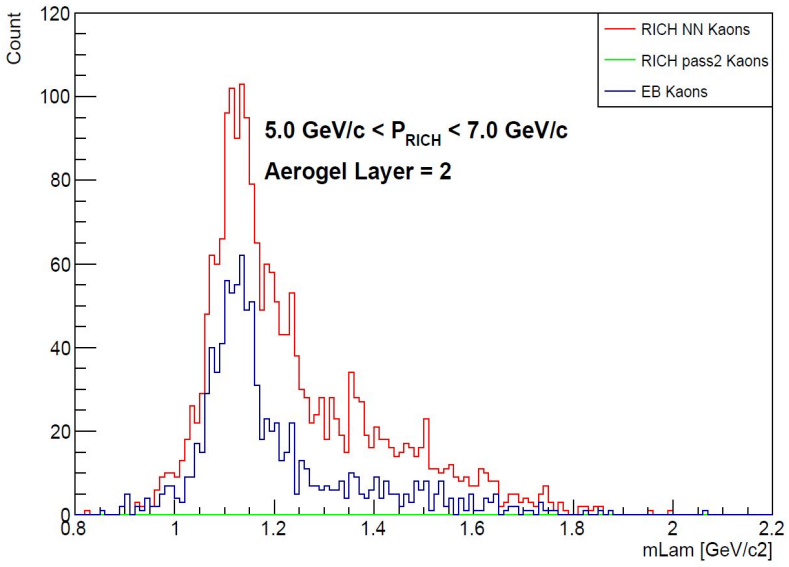

Figure 10: *Missing mass of* $ep \rightarrow e'K^+X$ *for kaons crossing the large angle aerogel radiator section (layer 2). Red histogram: RICH NN particle ID algorithm. Blue histogram: CLAS12 kaon ID without the RICH.*