RG-E Lambda Analysis and ALERT AI-assisted PIDs Update

CLAS Collaboration Meeting

July 10th, 2025

Uditha Weerasinghe

Mississippi State University

Outline

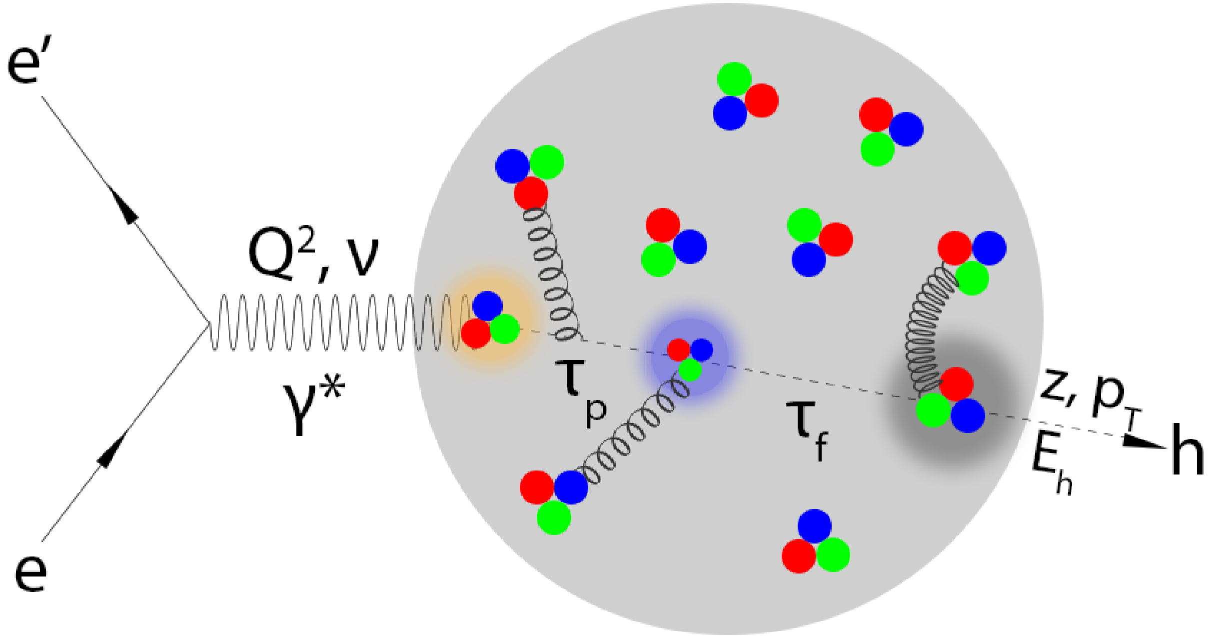
- RG-E Lambda Analysis
  - Semi-inclusive Deep Inelastic Scattering Production
  - SIDIS Kinematics Cuts
  - Particle Identification
  - Preliminary Vertex Cuts
  - Λ Production Channel
  - Event Mixing for Background Subtraction
- ALERT AI-assisted PID
  - AI/ML Alternative Solution
  - Data Sample Cleaning
  - Current Classification Accuracy

SIDIS Production

- Studying hadronization processes in SIDIS production helps improve our understanding of the confinement dynamics, a fascinating feature of Quantum Chromodynamics, the fundamental theory of strong interactions between quarks and gluons

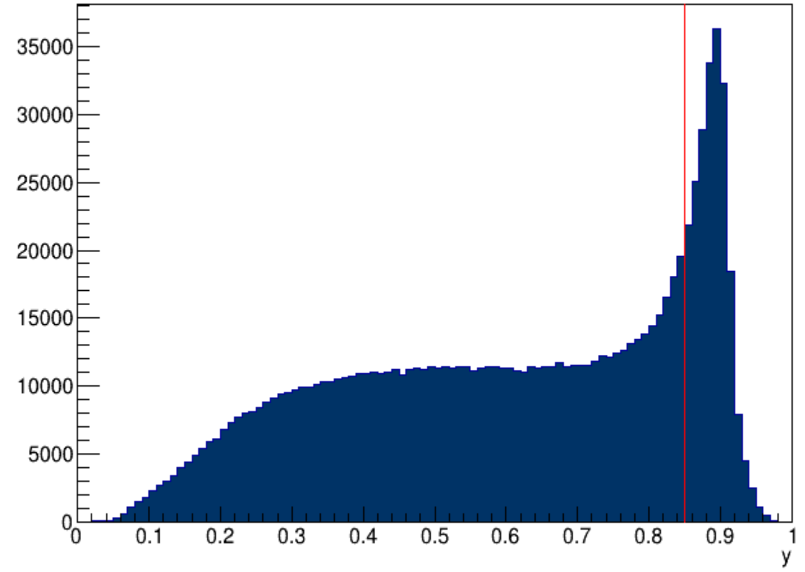
- Hadronization process is characterized by two time-distance scales:
  - **Production time (τ_p):** Struck quark propagates as a colored object during the color-neutralization stage
  - **Formation time (τ_f):** Time needed for the color-neutral prehadron to evolve into a fully dressed hadron with its gluonic field

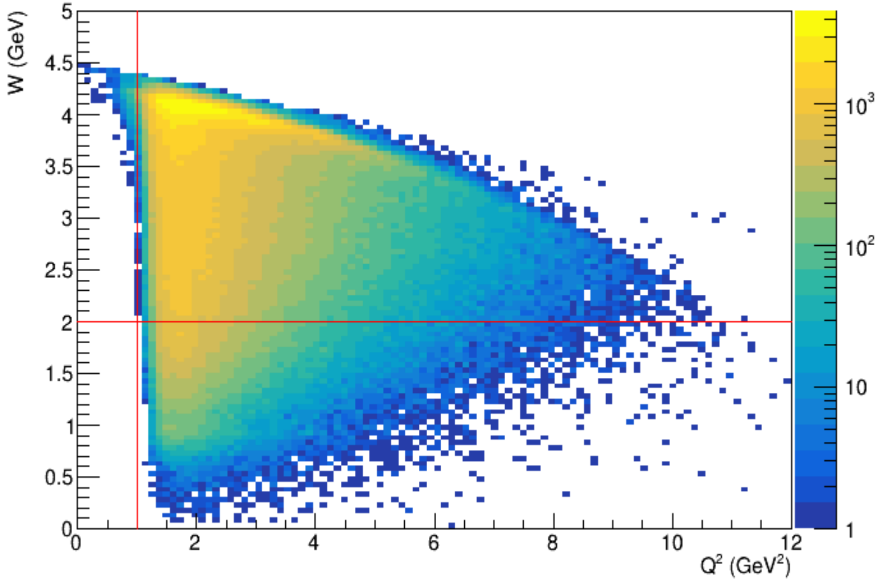
$$e + p \rightarrow e' + h + X$$



SIDIS Kinematics and Cuts

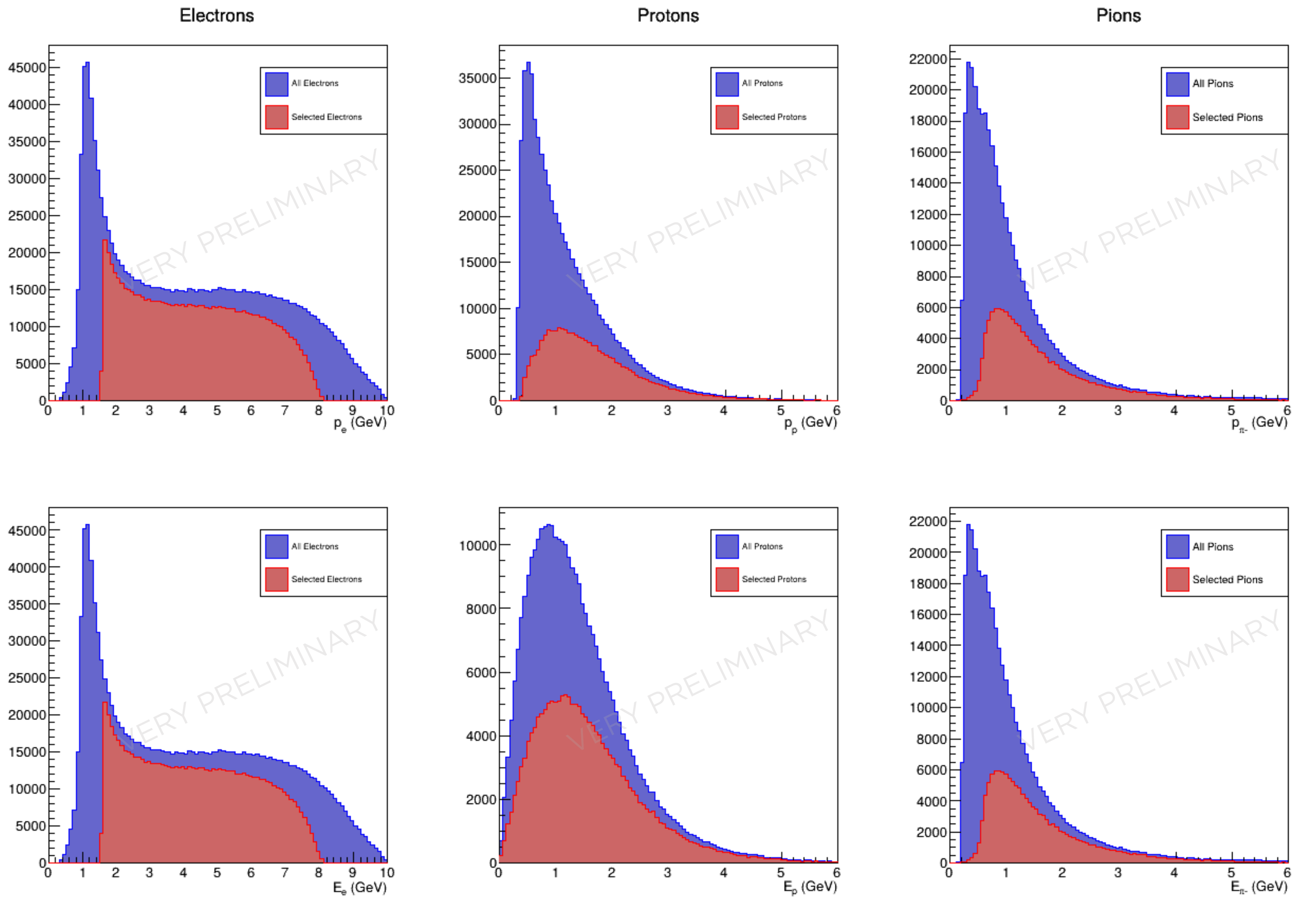
- The study of hadronization dynamics is probed in the SIDIS regime using this set of kinematics and cuts:
  - ν: electron energy loss or struck quark's initial energy
  - Q^2: four-momentum transferred squared
    - $\mathrm{Q}^2 > 1\ \mathrm{GeV}^2$: to probe the intrinsic structure of nucleons
  - $y = {}^{\nu}/_{E_{beam}}$: electron energy fraction transferred to a struck quark
    - $\mathrm{y} < 0.85$: to reduce radiative effects based on former HERMES studies
  - $W = \sqrt{M^2 + 2\nu M - Q^2}$: total mass of the hadronic final state, where M is the nucleon mass
    - $W > 2\ \mathrm{GeV}$: to avoid contamination from the resonance region
  - $z_h = {}^{E_h}/_{\nu}$: struck quark's initial energy fraction carried by the formed hadron
  - p_T: hadron transverse momentum measured relative to the virtual photon direction





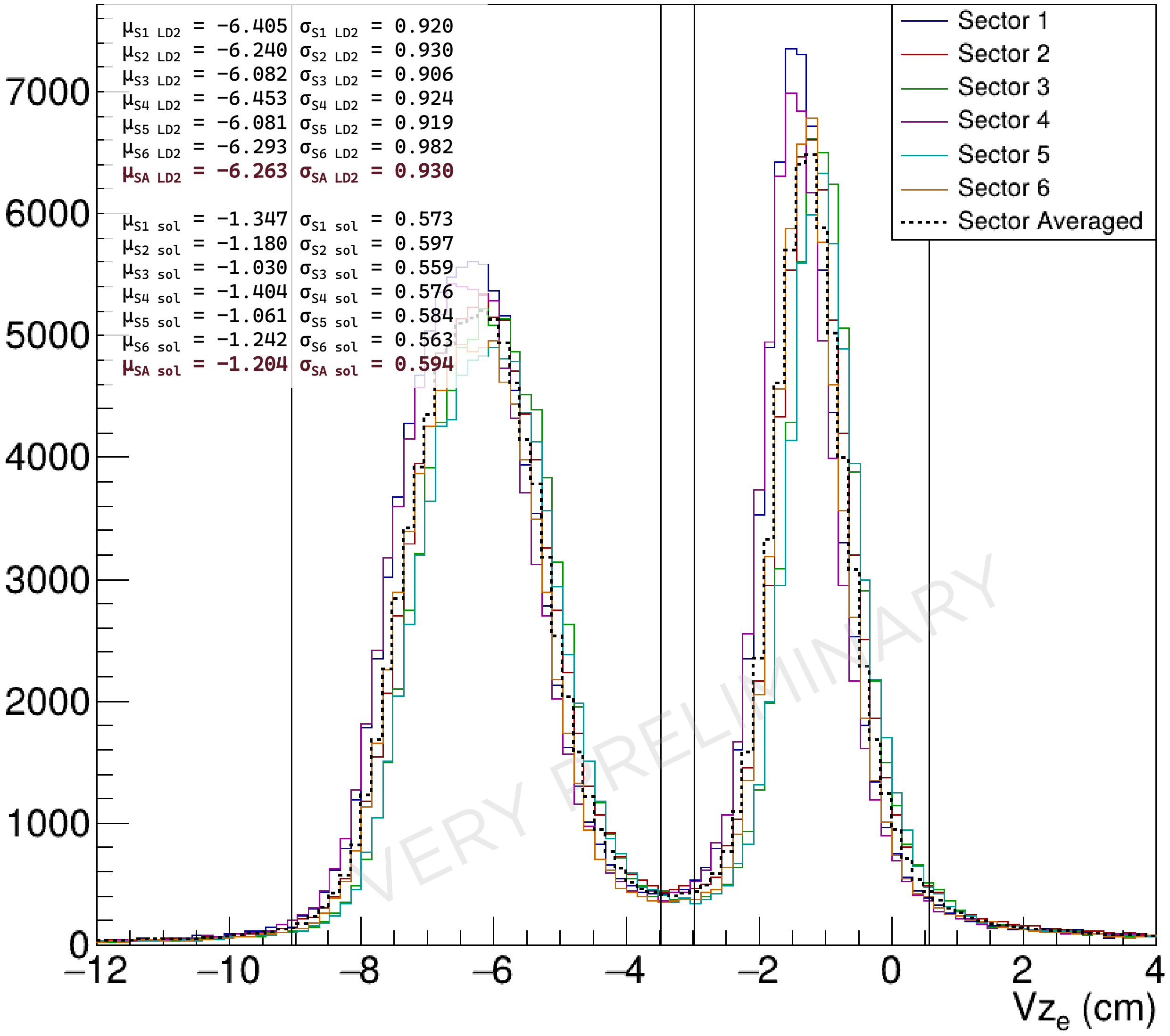
Particle Identification

- Particle ID:
  - Electron: +11
  - (+/-) pions: (+/-) 211
  - Proton: +2212
- Detectors cuts:
  - Electron should be in the forward region
  - Pions and protons are either in the forward or central region
- Fit quality cut:
  - 3 σ cuts on the χ^2 of reconstructed tracks

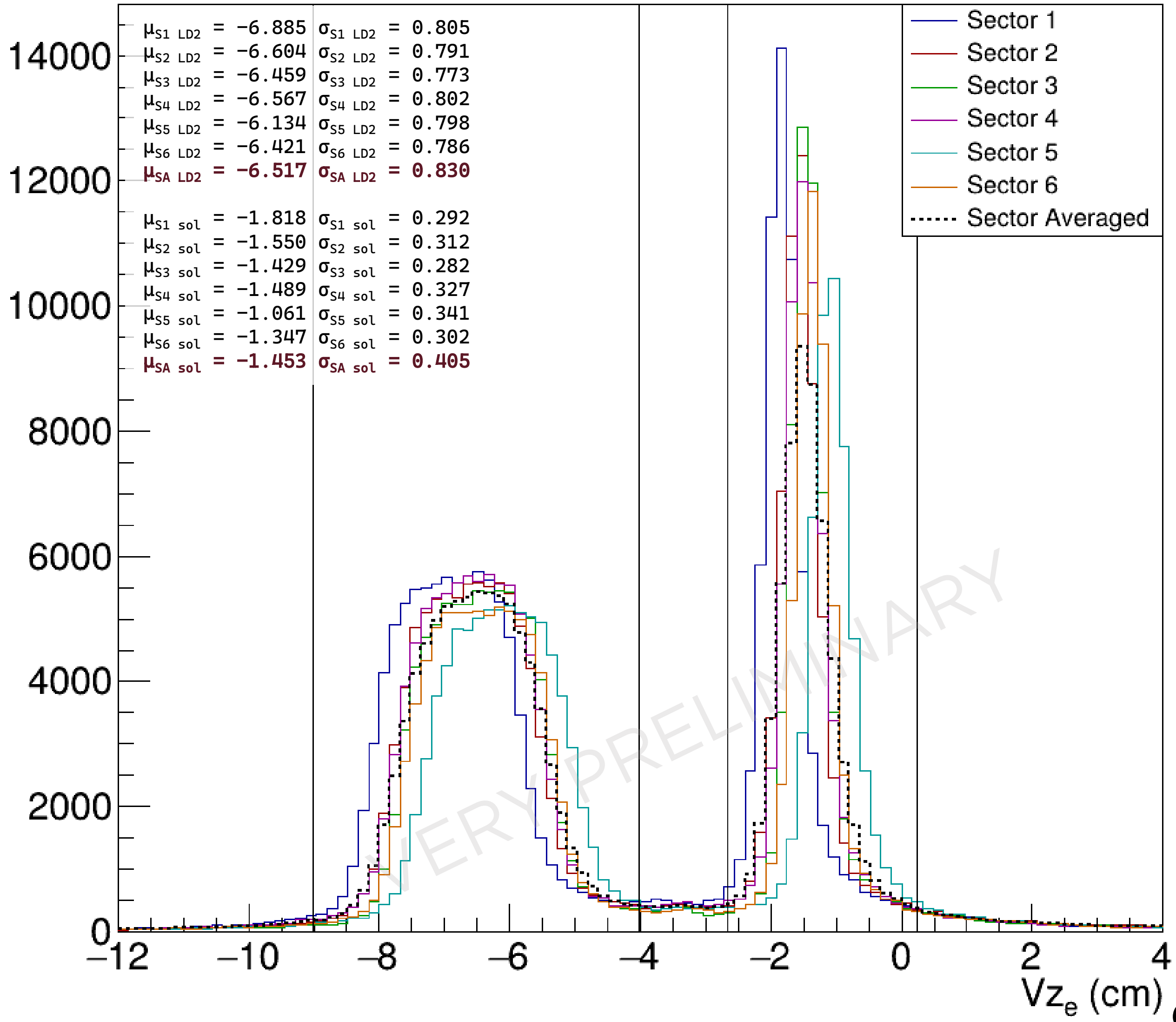


Preliminary Vertex Cuts

DC Vertex Distributions

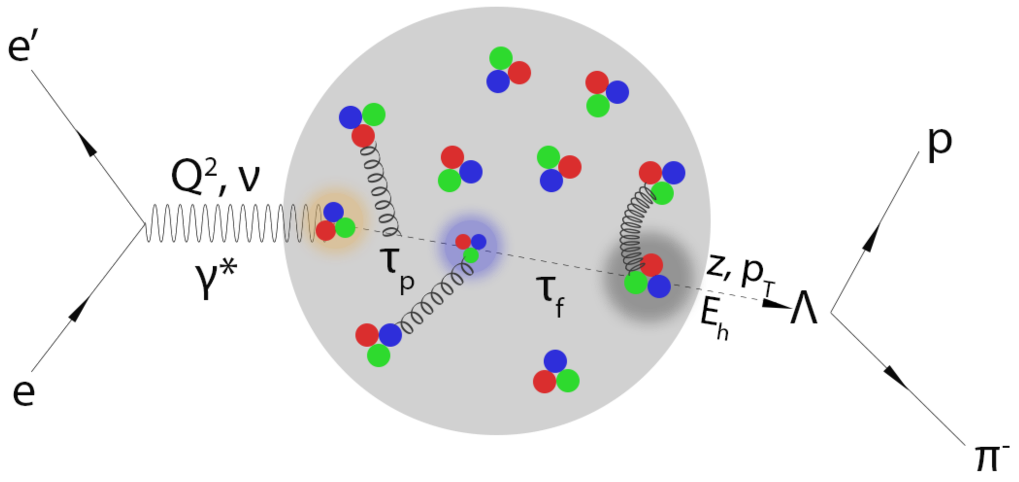


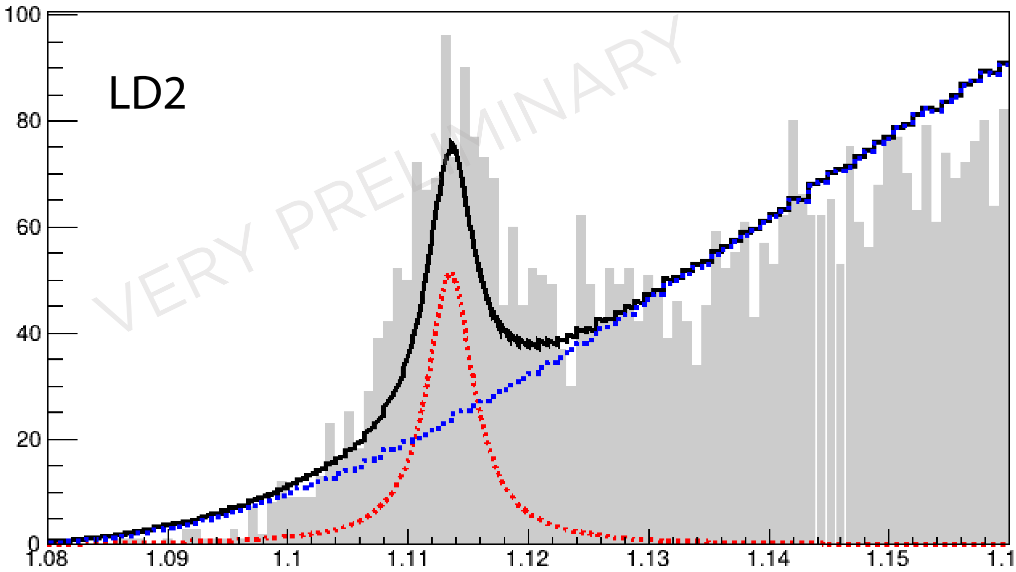
FMT Vertex Distributions

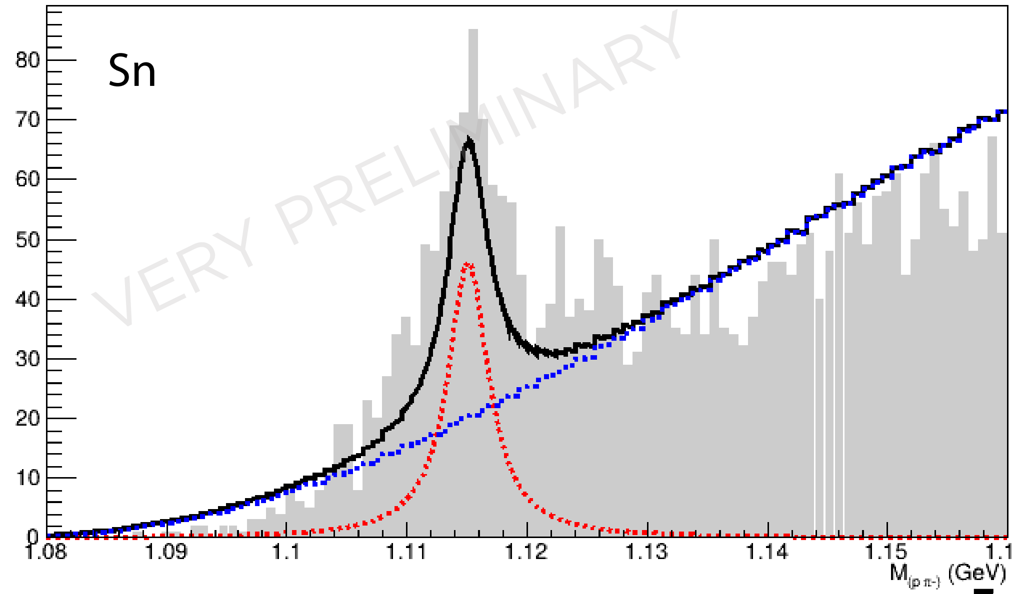


Λ Production Channel

- Our channel of interest is Λ SIDIS production off nuclei
- Λ is identified through its decay daughter particles, proton and π- detected in coincidence with the scattered electron
- Cuts applied on secondary vertex to refine the Λ signal
  - Distance between the electron and secondary Λ vertex
  - Opening angle between protons and π-s

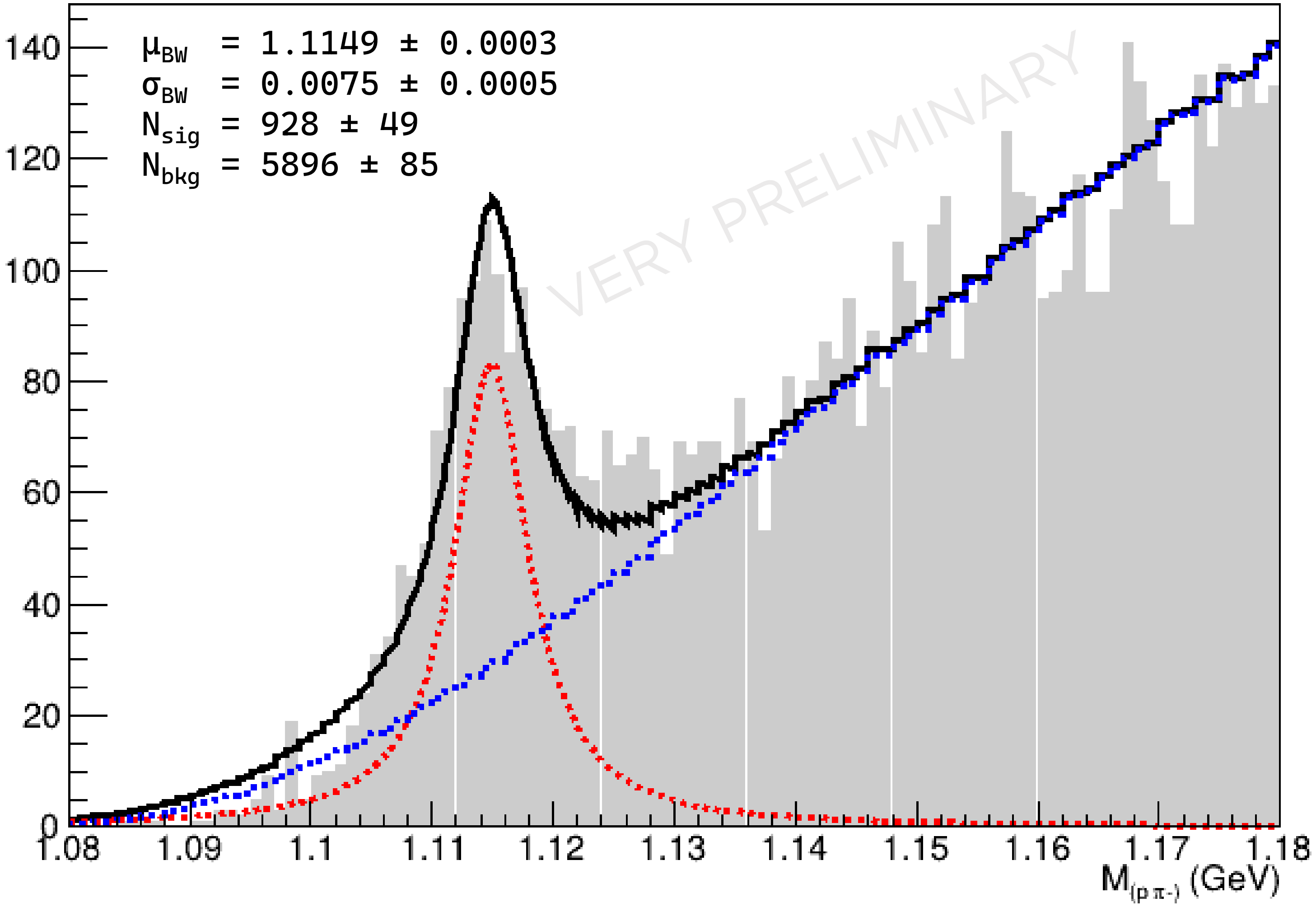






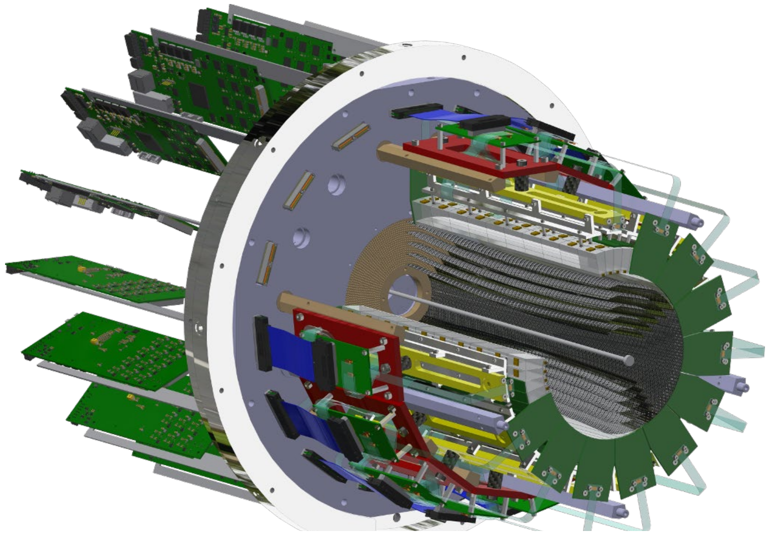
Event Mixing for Background Subtraction

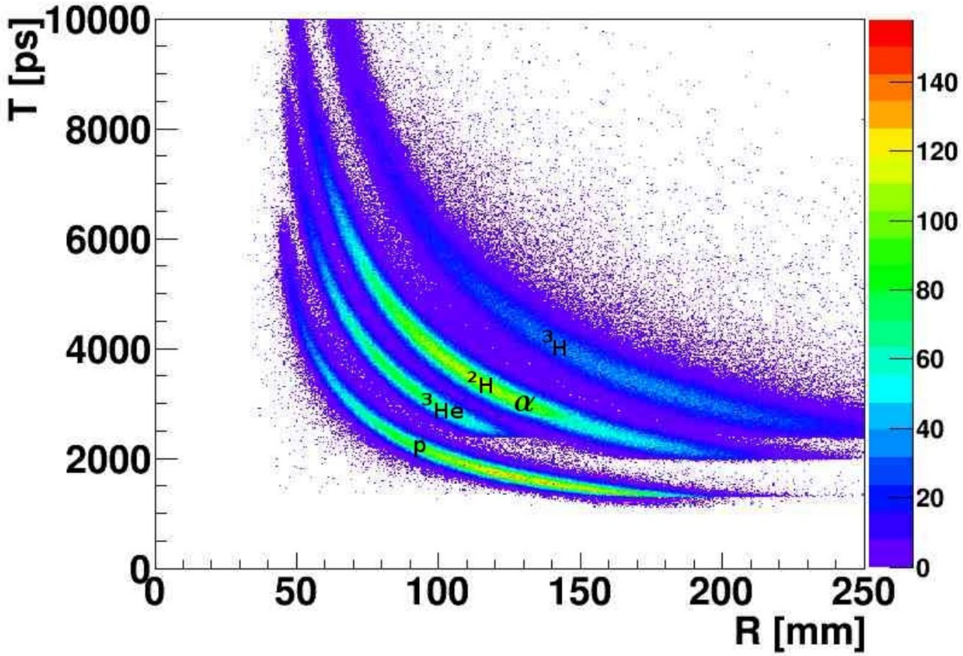
- Event mixing technique proved to be effective in modeling the background in the Λ invariant mass for CLAS6 EG2 data
- Each correlated event protons and pions pairs are mixed, respectively, with pions and protons from uncorrelated events to model the combinatorial background underneath the Λ peak
- The RooFit library was used to assign probability distribution functions- Breit-Wigner for the signal and the mixed events sample for the background, to fit the invariant mass distribution and obtain the Λ yield



ALERT AI-Assisted PID

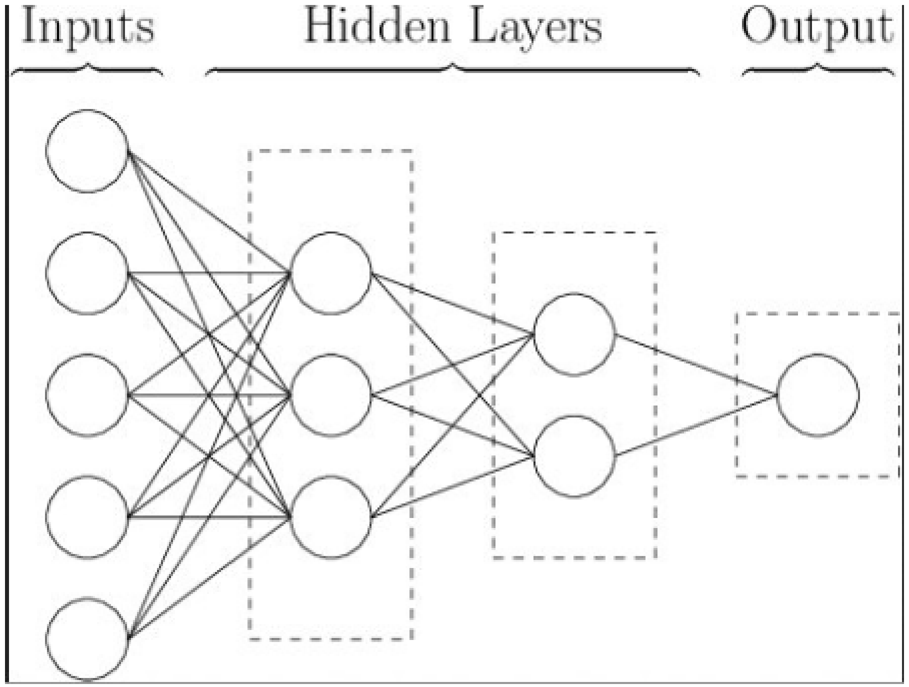
- Ongoing efforts to deploy AI and ML techniques to improve track reconstruction and particle identification.
- In this context, a Multi Layer Perception (MLP) model has been used to classify recoil nuclear-target fragments that are detected in the ALERT detector
- The CLAS12 ALERT detector consists of
  - A Hyperbolic Drift Chamber (AHDC)
  - A Time-Of-Flight detector (ATOF)
- Relying on conventional tracking is insufficient to distinguish nuclear-recoil fragments, such as helium-4 and deuterium, from each other





AI/ML Alternative Solution

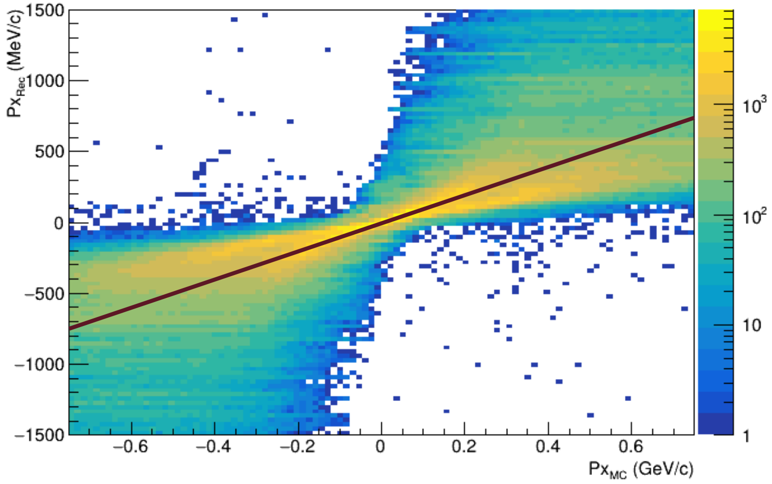
- A MLP neural network with an input layer, several hidden layers, and an output layer which determines PIDs
- A set of 27 features to include the following inputs:
  - momentum, energy deposited, supercluster position, and the AHDC residuals;
  - cluster position, time, and the ATOF path length
- Currently, the MLP model is trained with a sample of 5 M events generated using GEMC, and reconstructed using the CoatJava
- As the ALERT AI track finding is further trained with the collected data, the PID MLP model will be tweaked accordingly before being integrated in CoatJava

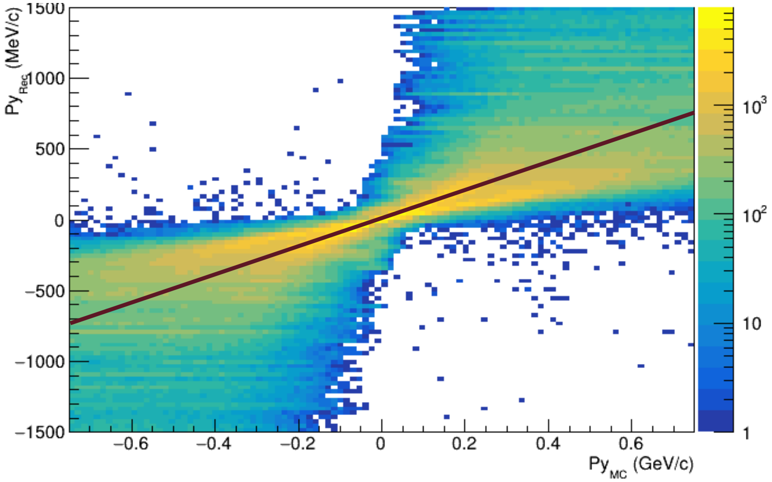


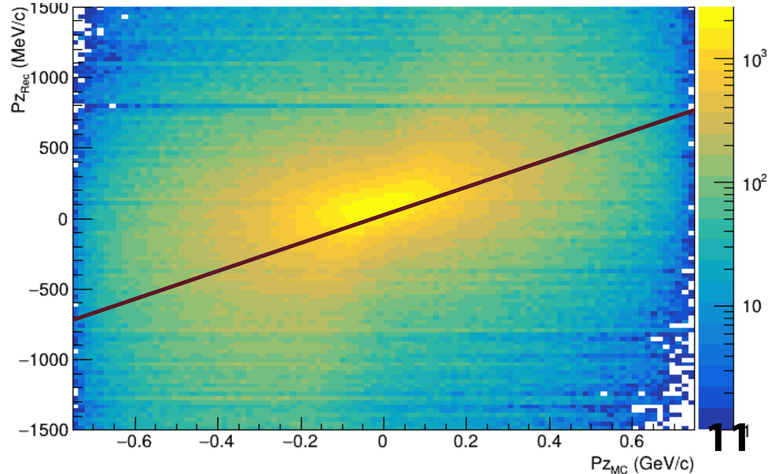
Data Sample Cleaning

❖ To clear the large outliers in the reconstructed training sample from simulation, the following cuts are used:

Variable	Selected Range	Percentage of Sample
Vx (cm)	[-0.2,0.2]	99.99%
Vy (cm)	[-0.2,0.2]	99.98%
Vz (mm)	[-600,-100]	90.29%
Px (MeV)	[-5000,5000]	94.98%
Py (Mev)	[-5000,5000]	95.12%
Pz (MeV)	[-5000,5000]	92.98%
Sum of ADC (arb.)	[0,45000]	99.99%
Sum of residuals (arb.)	[-25,15]	98.28%
chi2	[0,200]	98.33%
Number of hits	[0,16]	100%
Energy (MeV)	[0,250]	99.99%
Time (ns)	[124,132]	99.51%
X (cm)	[-100,100]	100%
Y (cm)	[-100,100]	100%
Z (cm)	[-50,170]	99.62%
Path length (cm)	[75,200]	99.87%
In path length (cm)	[1.4,25]	100%

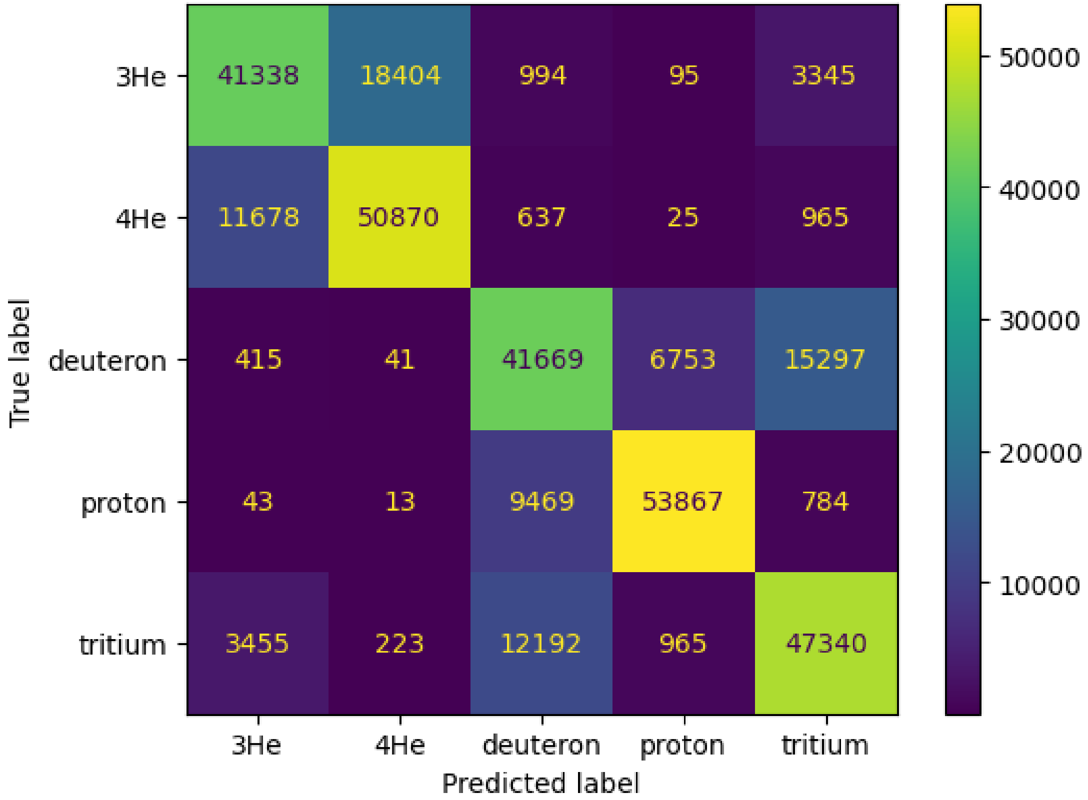






Current Classification Accuracy

For Reconstructed Momentum



Protons: 83.9 %

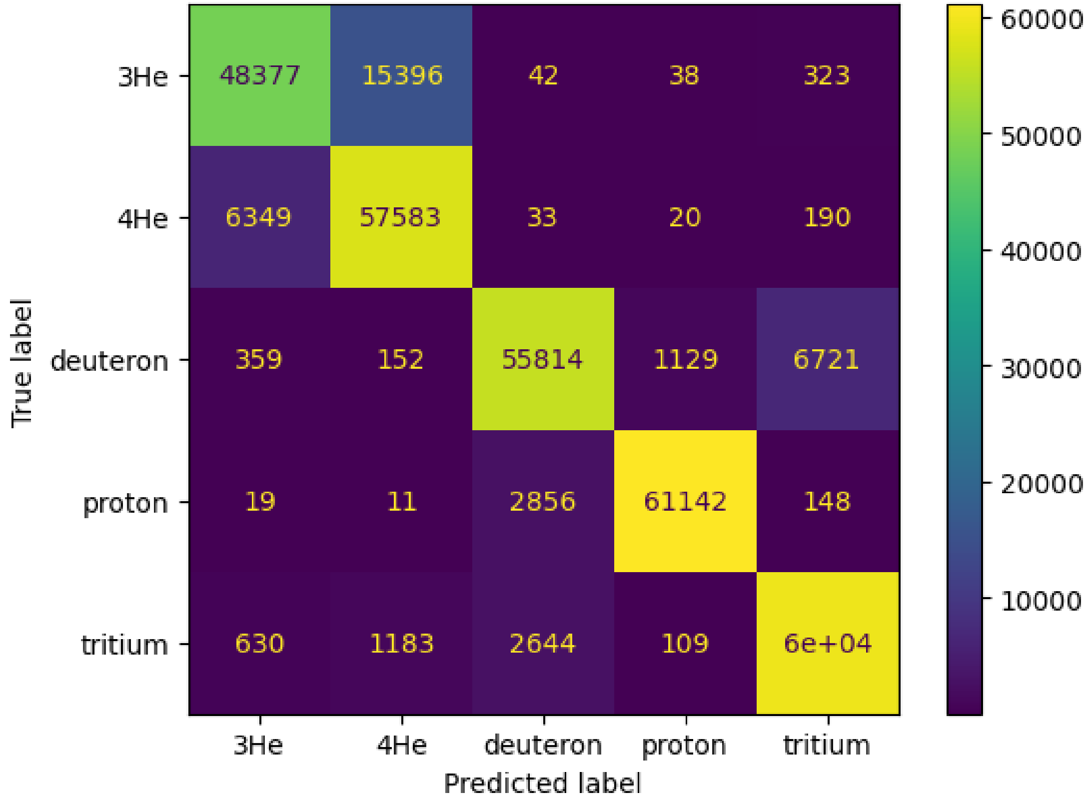
Deuterons: 64.9 %

Tritium: 73.7 %

Helium-3: 64.4%

Helium-4: 79.3%

For Simulated Momentum



Protons: 95.3 %

Deuterons: 87.0 %

Tritium: 92.9 %

Helium-3: 75.4%

Helium-4: 89.7%

Summary and Outlook

- Efforts to calibrate the collected CLAS12 RG-E dataset are underway
- Analysis codes are under development to
  - improve particle identification, vertex cuts and corrections
  - polish the Lambda signal using secondary vertex cuts
  - improve the event mixing algorithm for the background subtraction underneath the Lambda peak
  - extract the Lambda preliminary results for multiplicity ratios and transverse momentum broadening
- Ongoing studies to improve the AI-assisted identification of the nuclear-recoil fragments, such as ^{1}H, ^{2}H, ^{3}H, ^{3}He, as well as ^{4}He via a multi-class classifier for the current ALERT experiment
- Efforts are being made to
  - Cross-check and compare with the collected real data
  - Implement the model using the Deep Java Library making it compatible with CoatJava

Thank You!

Backup Slides

Physics Observables

Multiplicity Ratio

$$R_A^h = \frac{N_{SIDIS}^{h\,(A)} \Big/ N_{DIS}^{e\,(A)}}{N_{SIDIS}^{h\,(LD_2)} \Big/ N_{DIS}^{e\,(LD_2)}}$$

R_A^h describes the attenuation of formed hadrons, h, in the medium

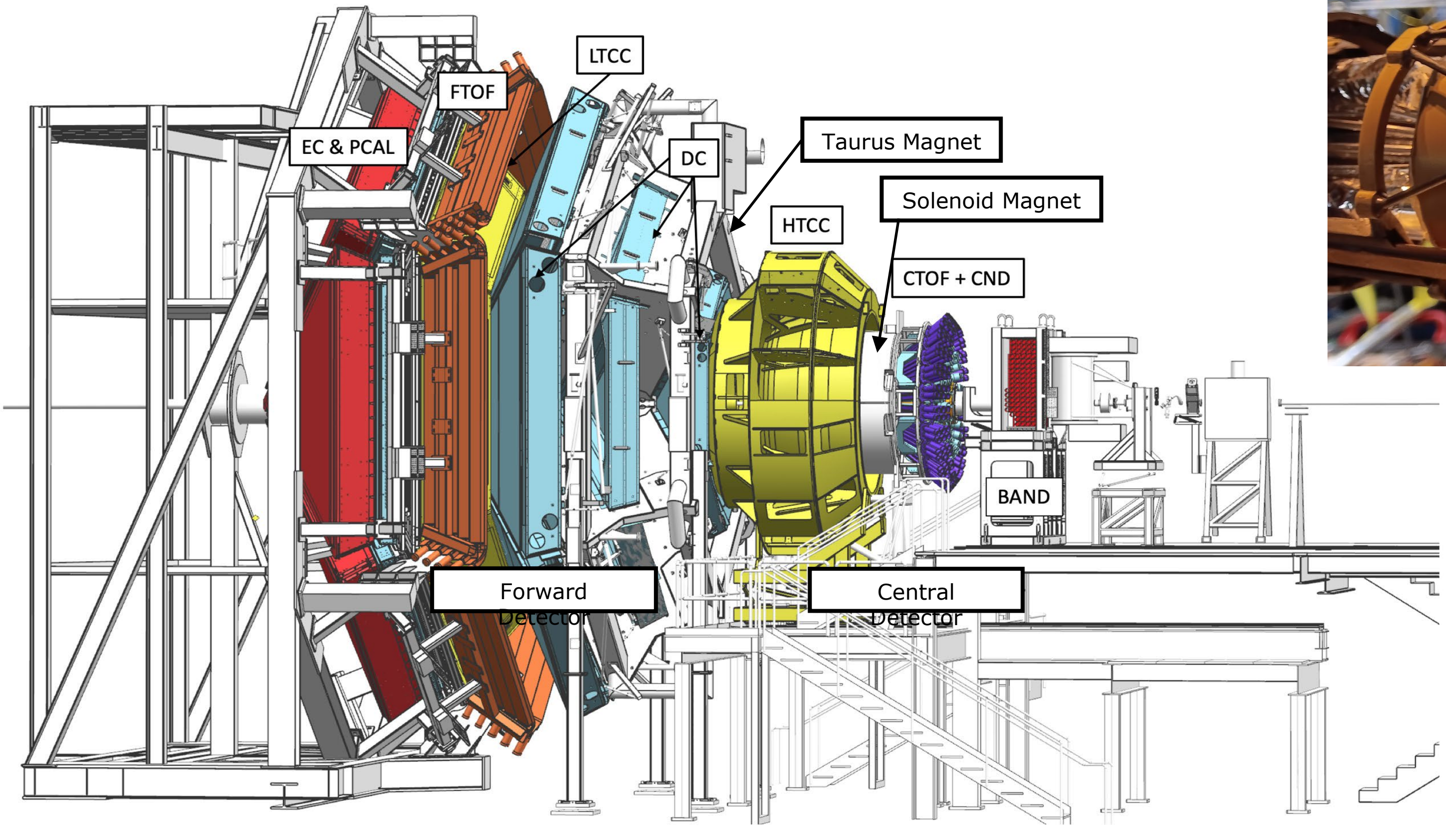
Transverse Momentum Broadening

$$\Delta p_T^2 = \langle p_T^2 \rangle_A - \langle p_T^2 \rangle_{LD_2}$$

Δp_T^2 is due to the energy loss of the propagating struck quark(s) and/or the elastic and inelastic scattering of prehadrons and hadrons

RG-E Experiment Setup

- RG-E experiments collected data during the spring of 2024 using the standard CLAS12 detectors with FT-OFF
- A double target assembly consisting of liquid deuterium (LD2) and solid foil targets placed inside the solenoid magnet



Electron Beam

Solid foils:
- C
- Al
- Cu
- Sn
- Pb

z-Vertex Comparison between DC and FMT