EPJ Web of Conferences **337**, 01010 (2025) https://doi.org/10.1051/epjconf/202533701010
CHEP 2024

Real-Time event reconstruction for Nuclear Physics Experiments using Artificial Intelligence

Gagik Gavalian [1,*]

[1]Thomas Jefferson National Accelerator Facility, Newport News, VA, USA

Abstract. Charged track reconstruction is a critical task in nuclear physics experiments, enabling the identification and analysis of particles produced in high-energy collisions. Machine learning (ML) has emerged as a powerful tool for this purpose, addressing the challenges posed by complex detector geometries, high event multiplicities, and noisy data. Traditional methods rely on pattern recognition algorithms like the Kalman filter, but ML techniques, such as neural networks, graph neural networks (GNNs), and recurrent neural networks (RNNs), offer improved accuracy and scalability. By learning from simulated and real detector data, ML models can identify and classify tracks, predict trajectories, and handle ambiguities caused by overlapping or missing hits. Moreover, ML-based approaches can process data in near-real-time, enhancing the efficiency of experiments at large-scale facilities like the Large Hadron Collider (LHC) and Jefferson Lab (JLAB). As detector technologies and computational resources evolve, ML-driven charged track reconstruction continues to push the boundaries of precision and discovery in nuclear physics.

In these proceedings, we highlight advancements in charged track identification leveraging Artificial Intelligence within the CLAS12 detector, achieving a notable enhancement in experimental statistics compared to traditional methods. Additionally, we showcase real-time event reconstruction capabilities, including the inference of charged particle properties, such as momentum, direction, and species identification, at speeds matching data acquisition rates. These innovations enable the extraction of physics observables directly from the experiment in real-time.

1 Introduction

Nuclear physics experiments have grown increasingly complex over recent decades, featuring more sophisticated detector systems and higher luminosities. In new experiments with elevated detector occupancies, innovative approaches to data processing are essential to enhance both the accuracy and speed of data reconstruction. Advances in Artificial Intelligence (AI) offer promising alternatives to traditional algorithms, with Machine Learning (ML) techniques being utilized at various stages of experimental data processing, including detector reconstruction, particle identification, detector simulations, and physics analysis.

This paper explores the integration of machine learning models into the CLAS12 charged-particle track reconstruction software. It provides a comprehensive analysis of the reconstruc-

*e-mail: gavalian@jlab.org

tion performance, highlighting improvements in track reconstruction efficiency and processing speed compared to conventional methods.

2 CLAS12 Detector

The CLAS12 (CEBAF Large Acceptance Spectrometer for 12 GeV) detector [1] is a state-of-the-art experimental apparatus used in nuclear physics research. It is located at the Thomas Jefferson National Accelerator Facility (Jefferson Lab) in Newport News, Virginia. The detector is part of an upgrade to the Continuous Electron Beam Accelerator Facility (CEBAF), which increased the maximum energy of the electron beam from 6 GeV to 12 GeV. This upgrade allows for a more in-depth exploration of the structure and properties of nucleons (protons and neutrons) and the nature of the strong force that binds them together in the atomic nucleus.

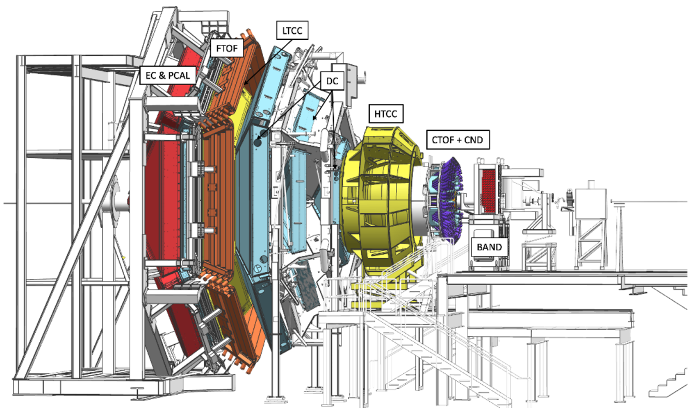

Figure 1. The CLAS12 detector in the Hall B beamline. The electron beam enters from the right and impinges on the production target located in the center of the solenoid magnet shown at the right (upstream) end of CLAS12, where other detector components are also visible. Scattered electrons and forward-going particles are detected in the Forward Detector (FD), consisting of the High Threshold Cherenkov Counter (HTCC) (yellow), followed by the torus magnet (gray), the drift chamber tracking system (light blue), and time-of-flight scintillation counters (brown), and electromagnetic calorimeters (red).

Key features and capabilities of the CLAS12 detector include:

- **Large Acceptance:** As its name suggests, CLAS12 has a large angular and momentum acceptance. This feature is crucial for detecting particles over a wide range of angles and energies, allowing comprehensive analysis of nuclear interactions.
- **Electron Beam Experiments:** CLAS12 is designed to investigate the interactions of high-energy electrons with nucleons and nuclei. By scattering electrons off target materials, scientists can probe the internal structure of nucleons and the dynamics of the strong force.
- **High Luminosity:** The detector operates at high luminosities, enabling it to collect a vast amount of data from electron scattering experiments. This high data rate is essential for studying rare processes and achieving statistically significant results.
- **Sophisticated Detection Systems:** CLAS12 consists of various subsystems designed to detect different types of particles and measure their properties. These include drift chambers for tracking charged particles, time-of-flight counters for particle identification, calorimeters for measuring energy, and Cherenkov detectors for identifying electrons.

- **Versatility:** The detector is versatile and can be used for a wide range of experiments, from studying the quark-gluon structure of nucleons to investigating the properties of nuclei under extreme conditions.
- **Data Analysis and Simulation:** Advanced software and computational tools are used to analyze the data collected by CLAS12. These tools include simulation packages that model the detector's response and data analysis frameworks for extracting physical quantities from the experimental data.

In summary, CLAS12 is a critical tool in modern nuclear physics, enabling researchers to delve deeper into the quantum world of nucleons and nuclei. Its advanced technology and capabilities contribute significantly to our understanding of fundamental physics, particularly in the realm of quantum chromodynamics (QCD), the theory describing the strong interaction.

3 Drift Chamber Particle Tracking

The CLAS12 forward detector is built around a six-coil toroidal magnet which divides the active detection area into six azimuthal regions, called "sectors". Each sector is equipped with three regions of drift chambers [2] designed to detect charged particles produced by the interaction of an electron beam with a target. Each region consists of two chambers (called super-layers), each of them having six layers of wires. Each layer in a super-layer contains 112 signal wires, making a super-layer a 6×112 cell matrix. The schematic view of all sectors and super-layers is shown in Figure 2.

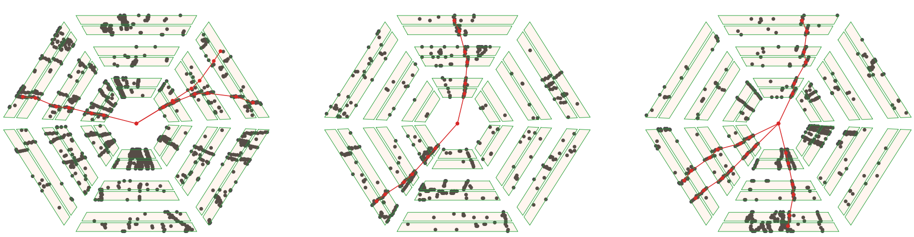

Figure 2. Example views of the six sectors of the Drift Chambers of CLAS12. The points show wire hits for each of the layers in the drift chambers and the lines indicate reconstructed tracks by the conventional CLAS12 tracking algorithm.

Particles that originate from the interaction vertex travel through the magnetic field and pass through all three regions of the drift chambers in a given sector and are reconstructed by tracking algorithms. First, in each super-layer adjacent wires with a signal are grouped into segments. Track candidates are constructed by connecting segments in each super-layer to form a track trajectory. Then each track candidate is fitted through a magnetic field to calculate the quality of the track (χ^2), and the track candidates with χ^2 below the cut value are saved for further refinement using Kalman-filter [3].

The positions of these segments in each super-layer are used to fit the track trajectory to derive initial parameters, such as momentum and direction (called "hit-based" tracking). After the initial selection, good track candidates (shown in Figure 2 with lines) are passed through a Kalman-filter to refine measured parameters further (called "time-based" tracking).

The CLAS12 track reconstruction process is already using AI at different stages of data reconstruction. First, a Convolutional Autoencoder is used to de-noise the drift chamber

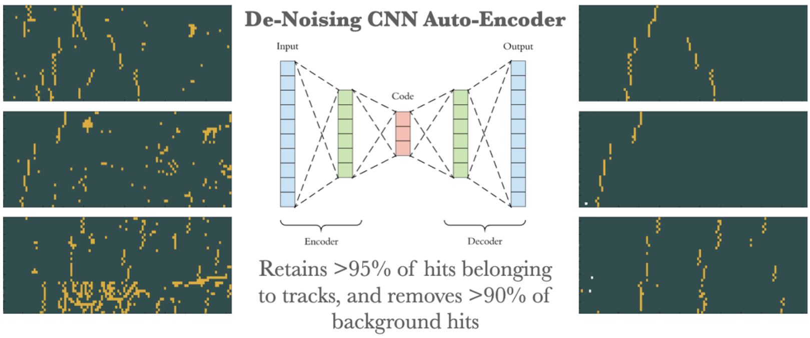

Figure 3. Convolutional Auto-Encoder (CAE) trained on real data can efficiently remove noise hits from raw Drift Chmaber hits, leaving only hits potentially associated with tracks.

hits [6], which leads to an improved cluster identification in each super-layer. Examples of raw and denoised drift chamber hist are shown in Figure 3, where the network was trained on experimental data, providing the raw hits in drift chambers as an image of 36x112 as input and the hits belonging to a reconstructed track as an output image. The Autoencoder learns to remove the noise hits, leaving only hits that can potentially form a track. After the denoising, a clustering algorithm finds segments in each super-layer to pass to the track-finding algorithm.

Then after the segment finding, the track candidate list is composed from all sensible combinations of segments in each super-layer. These track candidates are then identified using a Multi-Layer Perceptron (MLP) classifier, which identifies 6-super-layer and 5-super-layer track candidates [4]. Shown in Figure 4 is the architecture of the network and tracks selected by the network from possible track candidates in one sector.

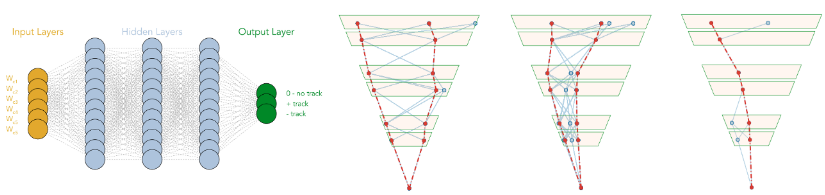

Figure 4. Multi-Layer Perceptron (MLP) network is used to identify tracks from combinations of segments measured in each super-layer of Drift Chambers. The input is a vector of 6 numbers (average segment position in each super-layer, and the output is a vector of probabilities (3 labels), of false track, negatively charged track, and positively charged track.

The combination of de-noising and AI-assisted track candidate identification in CLAS12 results in a 60% – 75% increase in statistics for multi-particle final states, depending on the kinematics and the number of detected particles in the final state [5]. Figure 5(a) illustrates the analysis of the interaction $ep \rightarrow e^{-}\pi^{+}\pi^{-}(X)$, comparing three different reconstruction workflows. The missing mass of the three detected particles is shown, with the missing proton peak clearly visible.

The conventional workflow represents the standard reconstruction process without AI assistance. In the AI-assisted workflow, the conventional procedure is augmented with track candidate predictions provided by the AI track classifier, yielding an approximate 30% improvement in statistics. The denoised/AI-assisted workflow utilizes denoised drift chamber data to identify segments, which are then processed by the MLP-based track candidate identification to suggest track candidates for the conventional track fitting algorithm. This workflow achieves an increase in statistics of approximately 60% in the missing mass distribution (for this particular interaction).

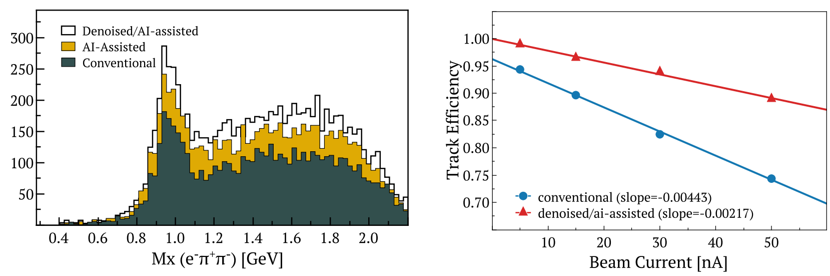

Figure 5. The increase in statistics in the interaction $ep \to e^-\pi^+\pi^-(X)$ (left). The missing mass of three particles is plotted, showing a peak for the missing proton for three different workflows of reconstruction, "Conventional" is the tracking without interference from AI, "AI-Assisted" is track reconstruction where suggestions on track candidates are provided by AI classifier, and "Denoised/AI-Assisted" is where the raw data was first denoised by CAE and then tracks classifier finds track candidates and passed them to reconstruction procedure. The dependence of the single-track reconstruction efficiency as a function of luminosity is compared for "Conventional" and "Denoised/AI-Assisted" workflows (right).

In Figure 5, the dependence of single particle reconstruction efficiency is shown as a function of luminosity (beam current), calculated from data taken at different beam currents. It is evident from the figure that using AI (Denoising and track-finding) significantly improves the efficiency loss as a function of luminosity, making it possible to run experiments at higher than design luminosities, leading to a significant increase in collected data for the experiments.

4 Track Parameter Finding

An Artificial Intelligence (AI) approach for track reconstruction was developed to estimate track parameters, such as momentum and direction, based on cluster positions along the track [7]. The AI-estimated track parameters were shown to align more closely with those reconstructed using the Kalman filter compared to conventional hit-based tracking methods. This capability enables the identification of complete physics event topologies directly from raw hits in drift chambers, leveraging AI to achieve significantly faster processing than traditional tracking algorithms.

Figure 6 presents the results of the track parameter predictions, demonstrating the accuracy of AI-reconstructed track parameters relative to the Kalman filter-based time-based tracking.

Using track parameters predicted by AI, physics events can be analyzed to reconstruct particle final states. In Figure 7, reconstructed particles are shown for two inclusive interactions, $ep \to e^- pK^-(K+)$ and $ep \to e^-\pi^+\pi^-(p)$, where cut on the missing mass of the system is placed around the mass of the missing particle. The peaks corresponding to $\Lambda^\circ(1520)$ and $\rho^\circ(770)$ can be seen in the figure.

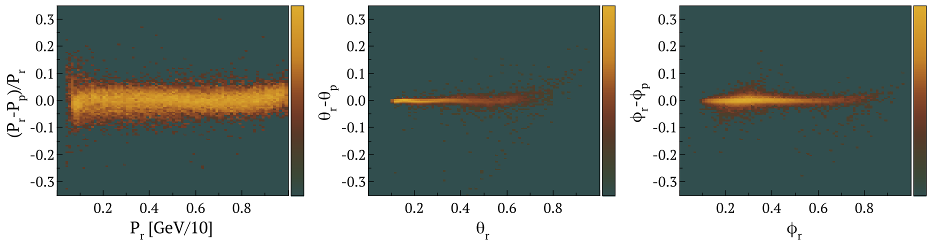

Figure 6. Results of track parameter estimating network (MLP) showing the difference between real track parameters and inferred parameters, for momentum, polar angle (θ), and azimuthal angle (ϕ) (shown in the normalized range from 0 to 1).

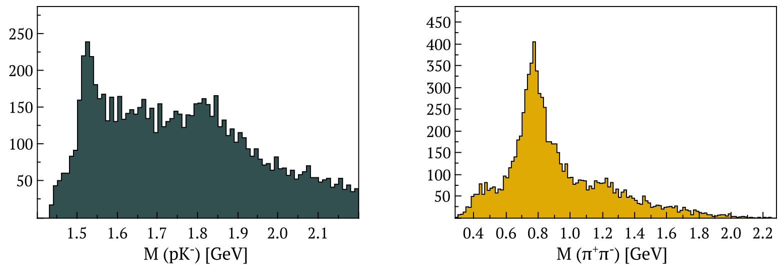

Figure 7. Particle reconstruction using the invariant mass of detected decay products for interactions $ep \to e^- pK^-(K+)$ and $ep \to e^-\pi^+\pi^-(p)$ respectively. The particle parameters used in calculating the invariant masses are derived by a neural network. A cut on missing mass was used to select appropriate exclusive interactions.

The particle parameters, such as momentum and direction, in the physics distributions are derived from AI-based track reconstruction. However, the data files had to first undergo conventional reconstruction to select events containing the required particles, as particle species identification has not yet been integrated into the AI workflow. While our group has developed an electron identification system, it has not yet been incorporated into the AI tracking process. An ongoing project aims to enable full particle identification using AI. Once this capability is developed and implemented, it will be possible to generate such distributions in real-time during the data acquisition stage.

5 Discussion

In CLAS12, we have implemented robust AI support for track reconstruction, which is now actively used in production. This innovation has resulted in a significant improvement in statistics, with increases of $60\% - 70\%$ depending on the physics interaction and event topology. Additionally, the AI-assisted track finding has enhanced the performance of conventional tracking code, delivering a 30% speedup.

A comprehensive track reconstruction framework was developed to identify tracks directly from raw hits in drift chambers and predict track parameters with greater accuracy than traditional hit-based Kalman-filter fits. Benchmark testing demonstrated that the AI reconstruction operates at 8 kHz in a single-threaded mode, enabling multi-threaded deployment for online use. This approach can process physics events faster than the current CLAS12 data acquisition rate of 16–20 kHz.

The fast AI reconstruction will also be employed online to classify events by topology and organize them within the output data stream, significantly accelerating calibration and physics validation workflows. With the integration of electron identification neural networks, this system can function as a level-3 software trigger, enhancing the level-1 trigger purity and potentially reducing the data volume by up to 40%.

This development is a critical step toward enabling CLAS12's transition to streaming readout systems, where software-driven triggers are essential for reducing streaming data. Real-time tracking reconstruction and particle identification will be indispensable solutions in this new paradigm.

6 Acknowledgments

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics under contract DE-AC05-06OR23177, and NSF grant no. CCF-1439079 and the Richard T. Cheng Endowment.

References

[1] Burkert, V.D. and others, The CLAS12 Spectrometer at Jefferson Laboratory, Nucl. Instrum. Meth. A **959**,163419 (2020)

[2] "Mestayer, M.D. and others", The CLAS12 drift chamber system, Nucl. Instrum. Meth. A",**959** 163518 (2020)

[3] Kalman, R. E., A New Approach to Linear Filtering and Prediction Problems, Journal of Basic Engineering, **82**, 35-45 (1960)

[4] Gavalian, Gagik and Thomadakis, Polykarpos and Angelopoulos, Angelos and Ziegler, Veronique and Chrisochoides, Nikos, Using Artificial Intelligence for Particle Track Identification in CLAS12 Detector, **2008.12860**, (2020)

[5] Gavalian Gagik,Auto-encoders for Track Reconstruction in Drift Chambers for CLAS12, **2009.05144**, (2020)

[6] Thomadakis, Polykarpos and Angelopoulos, Angelos and Gavalian, Gagik and Chrisochoides, Nikos, De-noising drift chambers in CLAS12 using convolutional autoencoders, **271**, 108201 (2022)

[7] P. Thomadakis, K. Garner, G. Gavalian and N. Chrisochoides, "Charged particle reconstruction in CLAS12 using Machine Learning," Comput. Phys. Commun. **287**, 108694 (2023) doi:10.1016/j.cpc.2023.108694

[8] G. Gavalian, High-Performance Data Format for Scientific Data Storage and Analysis, [arXiv:2501.07666 [physics.data-an]].