

Machine Learning for the Cluster Reconstruction in the CALIFA Calorimeter at R³B

Author name^a

^aorganization=University of the Moon,addressline=, city=Earth, postcode=, state=, country=

Abstract

ABSTRACT... work in progress ...

Keywords: keyword 1, keyword 2, keyword 3, keyword 4

1. Introduction

With the advancements in facilities dedicated to the production of radioactive beams at relativistic energies, such as the Facility for Antiproton and Ion Research (FAIR) at GSI, significant progress has been made in the study of exotic nuclei far from stability. FAIR will provide high-intensity relativistic radioactive beams of rare isotopes with energies reaching up to 1 AGeV, enabling investigations in inverse kinematics with full kinematic reconstruction. A key experimental setup designed for this purpose is the Reactions with Relativistic Radioactive Beams (R³B) Setup, which allows for high-resolution particle spectroscopy. This setup serves as a unique tool for unveiling the structure of nuclei and their reaction dynamics with unprecedented precision.

At the core of the R³B Setup is the CALIFA calorimeter (Calorimeter for the In-Flight Detection of Gamma Rays and Light Charged Particles), a highly segmented detection system composed of more than 2500 CsI(Tl) scintillator crystals that hermetically enclose the target area. This design enables the simultaneous measurement of gamma rays down to 100 keV and light charged particles, such as protons and deuterons, up to several hundred A MeV. To ensure optimal performance, extensive research has been conducted to refine the geometric design, minimize scattering and energy losses due to the holding structure, and develop a dead-time-free data acquisition system capable of handling high-rate experiments. Furthermore, a seamless integration within the R3BRoot framework has been achieved, enabling offline data analysis from the raw data level to the calibrated (cal) level and ultimately to the cluster level, where individual hits are recombined for the final energy reconstruction.

This study presents the results of applying machine learning-based graph networks to enhance the energy reconstruction of gamma rays in CALIFA. Using simulated Geant4 data, the performance of standard R³B clustering algorithms is compared to an agglomerative clustering model implemented with SciPy and a generic neural network (NN) architecture, demonstrating the potential of machine learning techniques in improving reconstruction accuracy.

2. Methodology

2.1. Challenges in Relativistic Gamma Spectroscopy

While the detection of light charged particles such as protons typically yields well-localized energy deposits in segmented detector arrays, the detection of gamma rays emitted from reaction products moving at relativistic velocities ($\beta \approx 0.8$) presents significant challenges. These difficulties primarily arise from the inherently sparse and spatially distributed energy deposits resulting from the interaction mechanisms of photons with the scintillator material (see Fig. X).

At photon energies below approximately 300 keV, the photoelectric effect dominates the interaction cross-section. As the photon energy increases, Compton scattering becomes the predominant process. For photon energies exceeding the pair production threshold ($E_\gamma > 2m_e c^2 \approx 1.022 \text{ MeV}$), electron-positron pair creation becomes possible and is the dominant interaction mechanism above $\sim 6 \text{ MeV}$.

Compton scattering broadens the clustering by the deflection of the incident gamma ray. According to the Klein–Nishina formula, the scattering is predominantly forward-focused for moderate to high photon energies, leading to clusters in neighboring crystals.

In the case of pair production, which occurs above the $2m_e c^2$ threshold, the resulting annihilation of the positron yields two 511 keV gamma photons. These secondary photons often escape the initial interaction site, leading to a significant fraction of the incident photon's energy being deposited in multiple detector elements.

For gamma rays emitted by nuclei at rest, this behavior gives rise to well-defined single- and double-escape peaks in the recorded energy spectra – corresponding to the escape of one or both 511 keV photons, respectively – if these photons exit the cluster volume without interaction (see Fig. X).

In experiments involving relativistic ions, such as those conducted at R³B, Doppler broadening significantly distorts spectral features, including single- and double-escape peaks. This effect hinders accurate reconstruction of the photon energy and complicates the extraction of absolute gamma-ray yields and reaction cross sections.

2.2. Data Structure and Standard R3B Clustering Algorithm

In the standard data acquisition (DAQ) configuration, all CALIFA detector hits occurring within a $\pm 4\mu\text{s}$ time window are grouped into a single event. Each individual hit i in CALIFA is represented by a data structure containing the following calibrated information:

- Energy deposit E_i
- Polar angle θ_i
- Azimuthal angle ϕ_i
- Time stamp t_i (White Rabbit time)

In the standard R3B clustering approach, the time information t_i is not utilized during the spatial reconstruction of clusters. The initial stage of the clustering algorithm begins by sorting all hits in descending order of energy. A user-defined geometric condition, typically a conical cluster shape with a default aperture of 0.5 rad, is applied. This value has been found to provide an optimal compromise between compact high-energy clusters and more diffuse gamma-ray showers. The hit with the highest energy defines the seed or center of the first cluster. The algorithm then iterates through the remaining hits and includes each hit in the current cluster if it lies within the specified cone aperture relative to the seed direction. Once the list is fully processed for the current cluster, the next highest-energy unassigned hit becomes the seed of a new cluster. This procedure repeats until no unassigned hits remain.

2.3. Simulation Setup

To evaluate and compare the clustering algorithms presented in this work, simulated data are required, as supervised machine learning approaches rely on access to ground truth labels. For this purpose, the R3BROOT framework with a Geant4-based Monte Carlo backend was employed.

The CALIFA detector geometry used in the simulation corresponds to the configuration implemented in early 2024. At that time, the iPhos region (polar angles 19° – 43°) was fully instrumented, while only the forward half of the Barrel region (43° – 87°) was active. The forward-most CEPA region (7° – 19°) was not yet equipped.

Gamma-ray energies were sampled from a uniform distribution between 0.3 MeV and 10 MeV. The interaction of the primary gamma rays with the CsI(Tl) scintillation material was modeled using Geant4 transport physics.

To emulate realistic event topologies, three gamma rays were generated per event, resulting in multiple detector hits. Timing information was coarsely simulated by assigning each primary gamma a random emission time within the $\pm 4\mu\text{s}$ event window. The corresponding hit times were then Gaussian-smeared with a standard deviation of 200 ns to reflect typical electronic channel timing variations.

The resulting dataset was split into training and test subsets, comprising 13,000 and 7,000 events, respectively.

2.4. Performance Metrics

To quantitatively assess the performance of the clustering algorithms presented in this work, a set of four custom metrics was defined. Three of these are event-based, while an optional fourth metric evaluates clustering quality on a per-cluster basis:

- **True Positive (TP)**: All hits in an event are correctly assigned to their respective clusters.
- **False Positive (FP)**: At least one hit in an event is incorrectly merged into a cluster it does not belong to.
- **False Negative (FN)**: At least one hit is not merged into its true cluster and instead forms a spurious cluster.
- **False Mixed (FM)**: An event is classified as false mixed if it contains both FP and FN characteristics—i.e., at least one hit is wrongly assigned, and at least one true cluster is partially reconstructed.

In addition, a cluster-based metric is defined:

- **Well Reconstructed (WR)**: The ratio of correctly reconstructed clusters to the total number of true clusters in the dataset.

These metrics allow a comprehensive evaluation of clustering accuracy, robustness, and failure modes.

2.5. Agglomerative Clustering

2.6. Implementation of Edge Detection NN

3. Results

4. Summary and Discussion

Acknowledgements

Thanks to ...

Appendix A. Appendix title 1

Appendix B. Appendix title 2