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

T. Jenegger[a], N. Hartman[a], R. Gernhäuser[a], L. Fabbietti[a], L. Heinrich[a]

[a] *TUM School of Natural Sciences, Technical University of Munich,Germany*

Abstract

The R3B experiment at FAIR studies nuclear reactions using high-energy radioactive beams. One key detector in R3B is the CALIFA calorimeter consisting of 2544 CsI(Tl) scintillator crystals designed to detect light charged particles and gamma rays with an energy resolution in the per cent range after Doppler correction. Precise cluster reconstruction from sparse hit patterns is a crucial requirement. Standard algorithms typically use fixed cluster sizes or geometric thresholds. To enhance performance, advanced machine learning techniques such as agglomerative clustering were implemented to use the full multi-dimensional parameter space including geometry, energy and time of individual interactions. An Edge Detection Neural Network exhibited significant differences. This study, based on Geant4 simulations, demonstrates improvements in cluster reconstruction efficiency of more than 30%, showcasing the potential of machine learning in nuclear physics experiments.

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 is expected for our understanding of exotic nuclei far from stability [1]. FAIR will provide high-intensity relativistic radioactive beams of rare isotopes with energies in the range of 1 A GeV, enabling investigations with full kinematic reconstruction [2]. A key experimental setup designed for this purpose is the **R**eactions with **R**elativistic **R**adioactive **B**eams (R3B) setup, providing access to high-resolution spectroscopic data. 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 2544 CsI(Tl) scintillator crystals that hermetically enclose the target area in the polar angular range of $7^\circ < \theta < 140^\circ$. This design enables the simultaneous measurement of gamma rays down to $E_\gamma \approx 100$ keV and light charged particles, such as protons and deuterons, up to several hundred A MeV [3]. To ensure optimal performance, extensive research has been conducted to refine the geometric design, minimize scattering and energy loss due to the mechanical structure [4], and develop a dead-time-free data acquisition system capable of handling high-rate experiments [5]. Furthermore, a seamless integration within the R3BRoot framework [6] has been achieved, enabling offline data analysis from the raw to the calibrated data level and ultimately to the cluster level, where individual hits are recombined for the final energy reconstruction.

This study presents the results of a hierarchical machine learning model to enhance the energy reconstruction of gamma rays in CALIFA. Using simulated Geant4 data, the performance of the geometrical R3B clustering algorithm is compared to an agglomerative clustering model [7] and a multi-layer perceptron architecture [8], demonstrating the potential of machine learning techniques in improving reconstruction efficiency and 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 which emerge from the reaction vertex presents significant challenges. These primarily arise from the inherently sparse and spatially distributed energy deposits resulting from the interaction mechanisms of photons with the scintillator material (see Fig. 1) [9].

At photon energies below approximately 300 keV, the photoelectric effect dominates the interaction cross-section in the CALIFA detector material (CsI(Tl)). As the photon energy increases, Compton scattering becomes the predominant process. For photon energies exceeding the pair production threshold ($E_\gamma > 2m_ec^2 \approx 1.022 MeV$), electron-positron pair creation becomes possible and is the dominant interaction mechanism above $E_\gamma \backsim 6$ 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 [11], leading to additional clusters in neighboring crystals.

At high photon energies, the dominant interaction mechanism

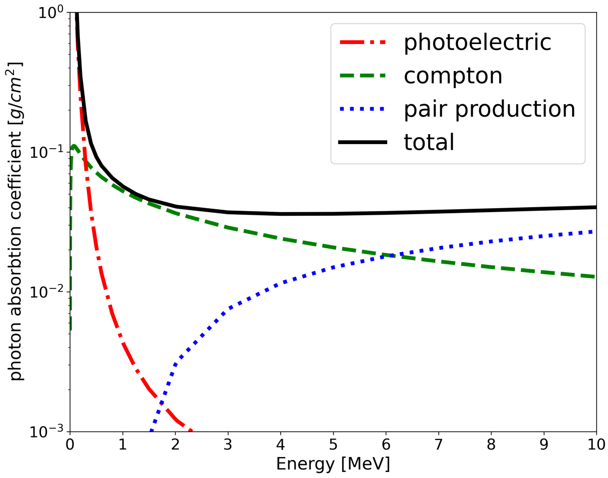


Figure 1: Photon absorbtion coefficients in CsI in the range from 100 keV to 10 MeV with data from XCOM database [10].

in the detector material is pair production (see Fig. 1), in which the incident photon converts into an electron–positron pair during the initial interaction. The subsequent annihilation of the positron results in the emission of two additional gamma photons, each with an energy of 511 keV. 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.

In experiments involving relativistic ion beams, such as those performed at R3B, Doppler broadening significantly affects the observed spectral features, including the single- and double-escape peaks. Moreover, for primary gamma rays with energies well above the pair production threshold ($E_\gamma > 2m_ec^2$), both the electron and the positron produced in the initial interaction are subject to substantial energy loss via bremsstrahlung. These effects contribute to a complex and highly non-trivial interaction pattern of gamma rays within the segmented detector system.

2.2. Data Structure and geometrical R3B Clustering Algorithm

The fundamental data entity in the analysis is a hit, defined as a discrete signal recorded by an individual detector segment at a specific time. To suppress contributions from low-energy background, only signals exceeding a predefined energy threshold are registered. In the present analysis, this threshold was set to 100 keV.

In the standard data acquisition (DAQ) configuration, all CALIFA detector hits occurring within a $\pm 4\,\mu$s time window are grouped into a single event. Each individual hit i in one of the detector crystals is represented by a data structure containing the calibrated energy deposit E_i, the polar angle θ_i, the azimuthal angle ϕ_i, and a time stamp t_i, which is synchronized using the White Rabbit Precision Time Protocol [12].

In the geometrical 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 cone emerging from the central target point with an aperture of 0.25 rad, is applied. This value has been found to provide an optimal compromise between compact high-energy clusters from light charged particles 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 to the current cluster if it sits within the specified cone 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 different clustering algorithms presented in this work, simulated datasets are employed to assess their performance. A geometrical model of the detector, closely matching the experimental setup, was implemented within the R3BROOT framework. The simulation employs a GEANT4-based Monte Carlo [13] backend, which accounts for all relevant secondary interaction processes. This approach enables realistic modeling of energy deposition and provides access to ground-truth labels for each individual interaction.

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 of signal and background, 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 $\pm 4\,\mu$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 using slow CsI(Tl) scintillator crystals.

Event selection is limited to cases in which all three gamma rays are—at least partially—detected within the geometrical acceptance of the CALIFA detector, which only partially encloses the target region. For gamma rays that deposit only a fraction of their energy in the detector volume, the corresponding true

energy is adjusted to reflect only the energy actually deposited in CALIFA.
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 incorrectly merged, 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.
Special attention must be given to the false negative rate, which is closely associated with the complex interaction pattern in the segmented detector. These processes produce widely spread hits that cannot be merged using the geometrical R3B clustering method, thereby motivating the development of a multi-layer perceptron architecture to improve clustering performance at the boundaries (see Subsection 2.6).

2.5. Agglomerative Clustering

To incorporate temporal information into the clustering process—unlike the geometrical R3B algorithm, which omits it—a generic, well-established method was adopted: agglomerative clustering [7] as implemented in the `SciPy` library [14]. This unsupervised learning algorithm enables flat clustering based on hierarchical linkage with a user-defined threshold.
Each hit was mapped into spherical coordinates (θ, ϕ, r), where the radial component r encodes time information. To ensure non-negative radii, the acquisition time window of $\pm 4\,\mu$s was shifted by $+4.5\,\mu$s. The Ward linkage criterion [15], which minimizes intra-cluster variance, was employed as the distance metric.
The threshold parameter was optimized to yield the best performance according to the custom-defined *true positive* (TP) and *well reconstructed* (WR) metrics.
As shown in Table 1, the agglomerative clustering algorithm demonstrates improved performance both on an event level (true positive rate) and on a cluster level (correctly reconstructed clusters) compared to the geometrical R3B clustering. However, this improvement is accompanied by an increased false negative rate, indicating that the algorithm tends to under-merge hits near the edges of clusters. This limitation motivated the development and application of an edge detection neural network, which is introduced in the following subsection.

2.6. Edge Detection Neural Network

To enhance the clustering performance, particularly at the boundaries of hit distributions, a multi-layer perceptron architecture was developed using the Pytorch library [16] to perform pairwise classification of detector hits. This model is applied either to individual raw hits or to hits pre-clustered via agglomerative clustering, on an event-by-event basis.

The model takes 12 input features for each hit pair (i, j): absolute values of energy (E_i, E_j), polar angle (θ_i, θ_j), azimuthal angle (ϕ_i, ϕ_j), and time (t_i, t_j). Additionally, four differential features are computed: $\Delta E = |E_i - E_j|$, $\Delta\theta = |\theta_i - \theta_j|$, $\Delta\phi = |\phi_i - \phi_j|$, and $\Delta t = |t_i - t_j|$. These differential inputs are helpful for training stability and convergence with our limited model sizes tested. In particular, $\Delta\phi$ resolves the discontinuities caused by the periodicity of the azimuthal angle (e.g., distinguishing between $\phi = 355^\circ$ and $\phi = 5^\circ$), which would otherwise introduce large erroneous differences in angular comparisons.

Of the 12 features, only the hit time is normalized to the $[0, 1]$ interval; all other values are used in their native physical units. The neural network architecture takes the 12-dimensional input vector and passes it through a fully connected feed-forward network with one hidden layer of 10^3 nodes, followed by a rectifier linear unit activation function (ReLU) [17]. Two additional hidden layers, each with 10^2 nodes, are applied sequentially. The output layer consists of a single node with a sigmoid activation, yielding a score in the interval $[0, 1]$, where values close to 1 indicate that the hits (or clusters) are likely to originate from the same event cluster.

Training is performed using the binary cross-entropy loss function [18, 19] and stochastic gradient descent (SGD) [20] with a fixed learning rate of 5×10^{-3}. Given the moderate size of the training dataset, full-batch training is employed without mini-batching. The model is trained for 8×10^4 epochs. After training, a threshold is applied to the prediction scores to classify hit pairs. This threshold is tuned to optimize the performance across all defined metrics, as described in Subsection 2.4. Final clusters are then formed by grouping all connected hit pairs based on the predicted associations.

The edge detection NN was implemented and tested in three configurations:

- **Plain Edge NN:** The model is applied directly to individual hits without any pre-clustering. All clustering is performed based solely on the NN predictions.
- **R3B + Edge NN:** The data are first clustered using the geometrical R3B clustering algorithm as an initial clean-up

step. For each resulting cluster, an energy-weighted center of mass is calculated, replacing individual hits. The NN is then trained exclusively on false negative cases, i.e., events where reconstructed clusters exhibit detached hits. In application, the geometrical R3B clustering is first applied to the test data, followed by the NN to refine cluster boundaries and reduce the false negative rate as clean-up step.

- **Agglo + Edge NN:** This strategy mirrors the R3B+Edge approach, with the key difference that time information is incorporated. As in the R3B+Edge model, the NN is trained on false negative cases to perform a final clean-up step after pre-clustering the hits using the agglomerative clustering algorithm described in the previous subsection. The significant reduction of the false negative rate achieved by the clean-up step in the Agglo+Edge implementation is demonstrated in Figure 2, which compares the reconstructed energy spectra from simulations of monoenergetic 2.1 MeV gamma events using the geometrical R3B clustering and the Agglo+Edge method.

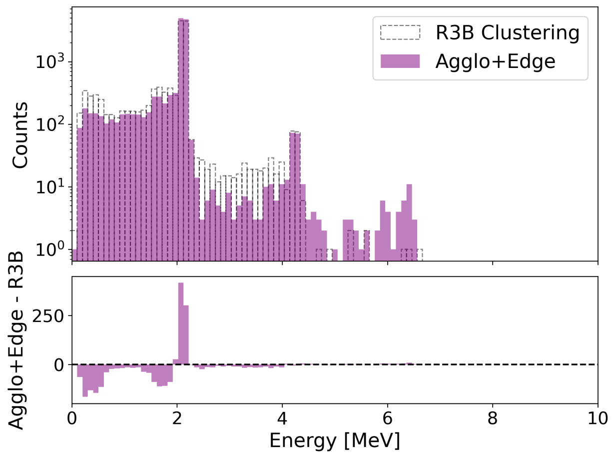


Figure 2: Reconstructed gamma energy spectrum from simulated events, each consisting of three 2.1 MeV gamma photons emitted from the target point. The upper panel shows the comparison between the geometrical R3B clustering and the Agglo+Edge method. The lower panel displays the bin-by-bin count difference between the two approaches. The Agglo+Edge model demonstrates a significant improvement by successfully reattaching escaped hits, notably in cases where sparse energy deposits around 1.6 MeV and 0.5 MeV result from pair production and subsequent annihilation processes of the original gamma photons. This clean-up step leads to a marked reduction in false negatives compared to the geometrical R3B clustering.

3. Discussion

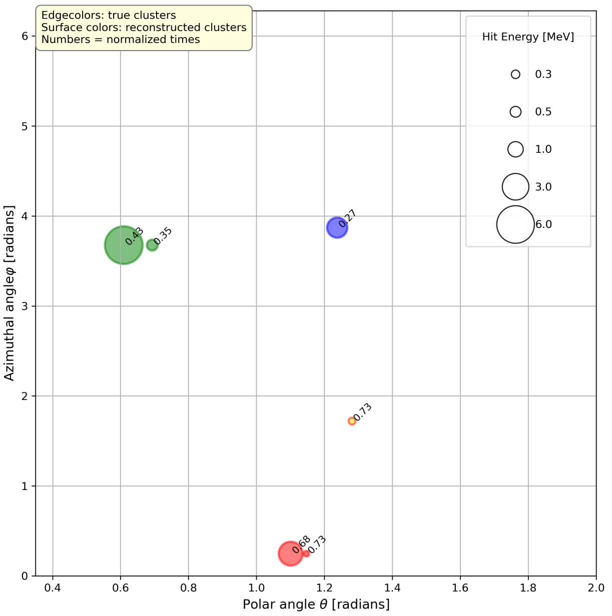
The results of this study are summarized in Table 1, organized according to increasing levels of reconstruction complexity. For completeness, the previously obtained results from the comparison between the "baseline"geometrical R3B clustering algorithm and the agglomerative model are also included.
The agglomerative model shows improved performance over the R3B baseline in terms of both event-level true positives

Clustering Model	TP(↑)	FP(↓)	FN(↓)	FM(↓)	WR(↑)
Geometrical R3B Clustering	60.6	5.3	25.2	8.9	80.4
Agglomerative Clustering	62.8	**3.3**	32.0	1.9	84.1
Edge Clustering (no time)	63.4±0.3	7.2±0.3	24.8±0.7	4.6±0.1	82.4±0.1
Edge Clustering (with time)	74.7±0.5	3.4±0.6	20.5±1.3	1.4±0.1	89.2±0.1
R3B + Edge (no time)	67.4±0.3	8.5±0.3	16.0±0.4	8.0±0.3	82.2±0.1
Agglo + Edge (with time)	**81.3±0.3**	5.1±0.0	**12.2±0.3**	**1.5±0.1**	**91.0±0.1**

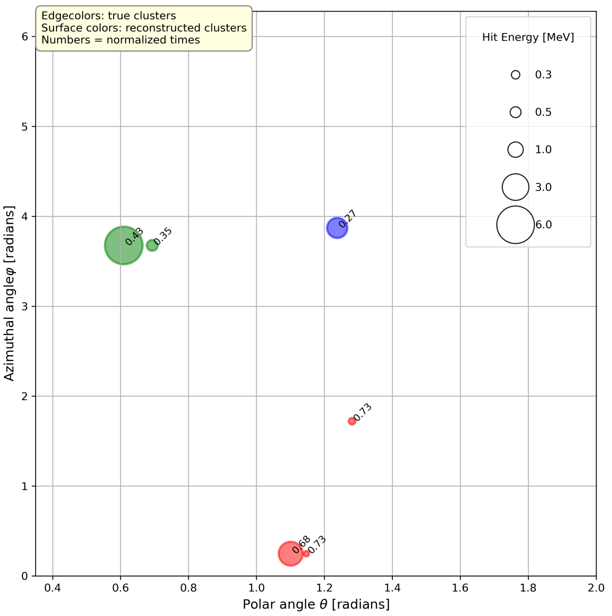
Table 1: Summary of performance metrics as defined in Subsection 2.4, evaluated for the different clustering algorithms. The models *Geometrical R3B Clustering*, *Edge Clustering (no time)*, and *R3B + Edge (no time)* utilize only angular and energy information on a per-hit basis for cluster reconstruction. In contrast, *Agglomerative Clustering*, *Edge Clustering (with time)*, and *Agglo + Edge (with time)* additionally incorporate time-of-hit information into the clustering process. Uncertainties reported for the four edge detection neural network variants correspond to the standard deviation of the results obtained from ten independent training runs.

(TP) and cluster-level (WR) values. However, it exhibits inferior performance with respect to the false negative (FN) rate, indicating a tendency to miss relevant hits during reconstruction. This limitation motivated the development of an Edge Detection Neural Network, initially evaluated as a standalone clustering algorithm and subsequently integrated into the agglomerative framework, yielding the combined model denoted as *Agglo + Edge*.
The *Agglo + Edge* model demonstrates superior performance across all evaluated metrics, achieving an overall correct reconstruction rate of 81.3%, significantly outperforming the *Geometrical R3B Clustering* algorithm, which reaches 60.6%.
A visual representation of an example event, contrasting the incorrectly merged hits from the geometrical R3B clustering with the correctly reconstructed clustering using the Agglo+Edge model, is shown in Figure 3.

To further explore the capabilities of neural network-based clustering approaches, two additional models were evaluated: a standalone *Edge Detection Neural Network* and a hybrid approach combining *Geometrical R3B Clustering* with edge-based postprocessing (*R3B + Edge*). Notably, both of these models operate without incorporating time-of-hit information, similar to the R3B baseline. Nonetheless, both outperform the *Geometrical R3B Clustering*, underscoring the potential of edge-based neural network models for improving cluster reconstruction in high-granularity detector systems.
The edge detection NNs presented here represent a special case of Graph Neural Networks (GNNs) [21], which, along with the more sophisticated transformer models [22, 23], have seen widespread adoption in particle physics over the past five years [24, 25, 26]. Interestingly, for this application, using an unsupervised learning algorithm (agglomerative clustering) to first define a graph structure presented a powerful inductive bias for our application which much improved our results over the standalone edge-NN. One limitation of this approach is it cannot improve an overly aggressive pre-clustering (e.g, a false positive rate can never be decreased in the clean-up step by the edge-NN). However, the necessity of reattaching edge hits to address the high false negative rate—exceeding the false positive rate by a factor of five in the baseline reconstruction



(a) Geometrical R3B clustering: the red-edged cluster with sparse hits is incorrectly split up into two clusters.



(b) Agglo+Edge clustering: successful reattaching of sparse hits.

Figure 3: Exemplary event with three simulated gamma photons in the energy range of 0.3–10 MeV, comparing the reconstruction performance of the geometrical R3B clustering and the Agglo+Edge method.

(see Table 1)—motivated this strategy.
Subsequent work could consider also adding a subsequent cluster splitting step in an end-to-end optimizable algorithm.
Although, in principle, transformers could learn the graph structure directly from hit distributions, initial tests showed limited performance, highlighting an opportunity for the community to further develop combined machine learning-based reconstruction methods.

4. Outlook

The results presented in the previous section clearly demonstrate that high-level machine learning approaches, such as the Edge Detection NN, can significantly enhance the accuracy of cluster reconstruction. These models not only reduce distortions in the measurement process but also exhibit increased sensitivity to low-statistics reactions—an important feature for experiments targeting rare processes.
It is noteworthy that even the models, which do not utilize time-of-hit information (similarly to the *Geometrical R3B Clustering*), outperform the baseline method. This underscores the general effectiveness of neural network-based methods in extracting structural features from detector data.
The inclusion of time-of-hit information proves to be a critical factor for enhancing clustering performance. As this observable is typically available for CALIFA at R3B, the results of this study support the recommendation to incorporate it into the reconstruction pipeline wherever possible.
Furthermore, these findings are intended to encourage broader adoption of advanced machine learning techniques by experimental groups, particularly in setups involving highly granular detectors. Such tools offer substantial performance benefits and can support more precise event reconstruction.
However, a consideration of computational complexity is essential. Both the *Geometrical R3B Clustering* and *Agglomerative Clustering* methods exhibit a time complexity of $\mathcal{O}(N^2)$ [27], where N is the number of hits. The proposed *Edge Detection Neural Network*, whether used in standalone mode or as an auxiliary module (*R3B + Edge*, *Agglo + Edge*), requires additional computational resources. In particular, the neural network involves large matrix operations—for instance, the second fully connected layer with 10^3 input and 10^2 output features—whose computational cost dominates for events with $N \sim \mathcal{O}(10^1)$. Consequently, the current implementation is not suitable for online processing and is better suited for detailed offline analyses, where accuracy is prioritized over execution speed.
Future work should focus on optimizing the Edge Detection Neural Network for speed and scalability. Additionally, transformer-based models [22]—capable of analyzing full event topologies—may offer further improvements in clustering accuracy by capturing complex, global features.

Acknowledgements

The work was supported by BMBF 05P24WO2 and Excellence Cluster ORIGINS from the DFG (Excellence Strategy

EXC-2094-390783311). It was made possible through the close collaboration of experts from different disciplines within the Cluster of Excellence ORIGINS [28].

References

[1] Experiments at the Interface of Nuclear, Atomic, and Hadron Physics with FRS at GSI and Super-FRS at FAIR.

[2] Y. Leifels, Status and physics perspectives of FAIR, IL NUOVO CIMENTO 100 (59) (2025) 48.

[3] D. Cortina-Gil, H. Alvarez-Pol, T. Aumann, V. Avdeichikov, M. Bendel, J. Benlliure, D. Bertini, A. Bezbakh, T. Bloch, M. Böhmer, et al., CALIFA, a Dedicated Calorimeter for the R3B/FAIR, Nuclear Data Sheets 120 (2014) 99–101.

[4] H. Alvarez-Pol, N. Ashwood, T. Aumann, D. Bertini, P. Cabanelas, E. Casarejos, J. Cederkall, D. Cortina-Gil, P. D. Fernández, I. Durán, et al., Performance analysis for the CALIFA Barrel calorimeter of the R3B experiment, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 767 (2014) 453–466.

[5] T. Le Bleis, M. Bendel, R. Gernhäuser, W. Henning, P. Klenze, R. Krücken, M. Winkel, A Digital Readout for CALIFA.

[6] D. Bertini, R3BRoot, simulation and analysis framework for the R3B experiment at FAIR, in: Journal of Physics: Conference Series, Vol. 331, IOP Publishing, 2011, p. 032036.

[7] F. Nielsen, Hierarchical Clustering, Springer International Publishing, Cham, 2016, pp. 195–211.

[8] M.-C. Popescu, V. E. Balas, L. Perescu-Popescu, N. Mastorakis, Multilayer perceptron and neural networks, WSEAS Transactions on Circuits and Systems 8 (7) (2009) 579–588.

[9] H. Kolanoski, N. Wermes, Teilchendetektoren, Springer, 2016, Ch. 13.5.2.

[10] S. Seltzer, XCOM-Photon Cross Sections Database, NIST Standard Reference Database 8, `http://www.nist.gov/pml/data/xcom/index.cfm`, accessed: 2025-04-10 (2010).

[11] O. Klein, Y. Nishina, Über die Streuung von Strahlung durch freie Elektronen nach der neuen relativistischen Quantendynamik von Dirac, Zeitschrift für Physik 52 (11) (1929) 853–868.

[12] M. Lipiński, T. Włostowski, J. Serrano, P. Alvarez, White rabbit: A PTP application for robust sub-nanosecond synchronization, in: 2011 IEEE International Symposium on Precision Clock Synchronization for Measurement, Control and Communication, IEEE, 2011, pp. 25–30.

[13] S. Agostinelli, J. Allison, K. a. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. Barrand, et al., GEANT4—a simulation toolkit, Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment 506 (3) (2003) 250–303.

[14] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, et al., SciPy 1.0: fundamental algorithms for scientific computing in Python, Nature methods 17 (3) (2020) 261–272.

[15] F. Nielsen, F. Nielsen, Hierarchical clustering, Introduction to HPC with MPI for Data Science (2016) 195–211.

[16] S. Imambi, K. B. Prakash, G. Kanagachidambaresan, Pytorch, Programming with TensorFlow: solution for edge computing applications (2021) 87–104.

[17] A. F. Agarap, Deep learning using rectified linear units (relu), arXiv preprint arXiv:1803.08375 (2018).

[18] S. Mannor, D. Peleg, R. Rubinstein, The cross entropy method for classification, in: Proceedings of the 22nd international conference on Machine learning, 2005, pp. 561–568.

[19] P.-T. De Boer, D. P. Kroese, S. Mannor, R. Y. Rubinstein, A tutorial on the cross-entropy method, Annals of operations research 134 (2005) 19–67.

[20] D. Newton, R. Pasupathy, F. Yousefian, Recent trends in stochastic gradient descent for machine learning and Big Data, in: 2018 Winter Simulation Conference (WSC), IEEE, 2018, pp. 366–380.

[21] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi, M. Malinowski, A. Tacchetti, D. Raposo, A. Santoro, R. Faulkner, et al., Relational inductive biases, deep learning, and graph networks, arXiv preprint arXiv:1806.01261 (2018).

[22] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Advances in neural information processing systems 30 (2017).

[23] X. Amatriain, A. Sankar, J. Bing, P. K. Bodigutla, T. J. Hazen, M. Kazi, Transformer models: an introduction and catalog, arXiv preprint arXiv:2302.07730 (2023).

[24] G. DeZoort, S. Thais, J. Duarte, V. Razavimaleki, M. Atkinson, I. Ojalvo, M. Neubauer, P. Elmer, Charged particle tracking via edge-classifying interaction networks, Computing and Software for Big Science 5 (2021) 1–13.

[25] X. Ju, D. Murnane, P. Calafiura, N. Choma, S. Conlon, S. Farrell, Y. Xu, M. Spiropulu, J.-R. Vlimant, A. Aurisano, et al., Performance of a geometric deep learning pipeline for HL-LHC particle tracking, The European Physical Journal C 81 (2021) 1–14.

[26] S. Van Stroud, P. Duckett, M. Hart, N. Pond, S. Rettie, G. Facini, T. Scanlon, Transformers for Charged Particle Track Reconstruction in High Energy Physics, arXiv preprint arXiv:2411.07149 (2024).

[27] S. Sieranoja, P. Fränti, Fast agglomerative clustering using approximate traveling salesman solutions, Journal of Big Data 12 (1) (2025) 21.

[28] Cluster of Excellence ORIGINS, Cluster of excellence origins, `https://www.origins-cluster.de`, accessed: 2025-05-03 (2025).