# Machine Learning Particle Flow

Ana Palavandishvili *supervised by Maria Dolores Garcia Marti*

August 2023

## 1 Introduction

Particle Flow refers to the intricate process of identifying and reconstructing individual particles generated during collision events. Reconstruction of the entire event is essential for measurement of particles, also the reconstruction of jets of particles originating from the fragmentation and hadronization of hard scattering partons[1]. One of the most problematic aspect is to differentiate particles of various nature when they are close to or overlap each other. Utilizing Machine Learning (ML) techniques offers a range of advantages, including enhanced pattern recognition, effective feature extraction, event simulation refinement, and anomaly detection, among others. The first applications of Neural Networks date back to the 1980s [3]. In recent years many studies demonstrated contemporary applications of Deep Learning (DL), in collider research have paved the way for accelerated processes and heightened algorithm performance. Modern Machine Learning approaches, particularly those employing Graph Neural Networks (GNN), have found widespread utility across various tasks such as particle flow reconstruction, jet identification, and tackling the challenges posed by pileup interference. In recent years, GNNs have emerged as the defacto toolkit for graph-based data analysis and learning. Their application has significantly advanced our capabilities in comprehending and extracting insights from intricate graph-structured data. The paper also discusses essential adjustments needed to pre-process a raw root file, ensuring its compatibility with the Machine Learning model. It involves extracting critical information from the root file and transforming the data set into a structured format of nodes and edges. This format effectively captures particle details and hit types. Additionally, the paper covers the extraction of vital features and emphasizes the normalization of the data set to bring it to a consistent scale.

## 2 The Data set

The raw root data set file which is used in this paper contains 28 branches, 20 entries, 1000 event numbers, and up to 3000 hits. To prepare the data set for the modifications, we use Pytorch and Deep Graph Library (DGL), which allows us to perform various ML tasks on the graph, this takes this batched graph and splits it to individual graphs, allowing us to process them separately. The input parameters are the number of hits, particle number, hit types, energy of the particles, and their coordinates. In this data set, after each iteration number of nodes changes from 100 to 1000, with the average number 758, maximum reached value is 998.The number of edges ranges from 805 to 7000, with the average number 5075, maximum reached value 6986. Afterward, we are iterating with those number of nodes with particle number, hit type, and coordinates of the particles. Within this data set, we identify the categories of particle interactions, each characterized by specific types of hits. Calorimeter Hits: These instances correspond to measurements captured by calorimeters. The calorimeter gauges the collective energy of particles by absorbing them, thereby generating signals proportionate to the energy absorbed. Particularly, electrons and photons energy measurements are carried out using Electromagnetic Calorimeters. Vertex Hits: In the context of particle collisions, certain points serve as the particle's origin. The accurate recording of vertex information holds significant importance, as it facilitates a comprehensive comprehension of particle interactions. At the left of the 3D plot it displays information on first 10 events, with particle and hit types. Also it is noteworthy that it is not normalized. The shape defines hit types, color -particle ID. Figure 1 shows 10 events with non-normalized data, which allowing us to see which event displays in Hcal or Ecal. Figure 2 is displayed only one event, with normalized re-scaled data.

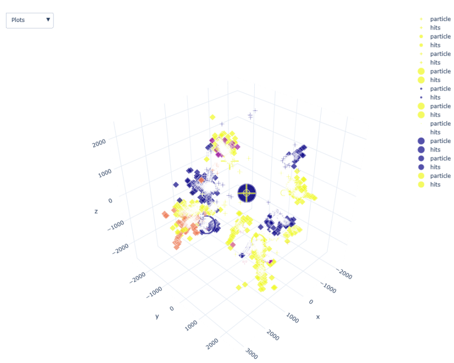

Figure 1: First 10 events with not normalized data

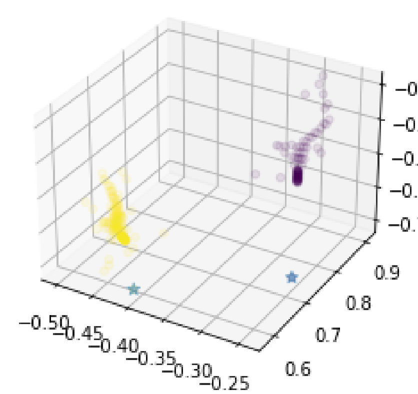


Figure 2: One event with normalized data

In the Figure below it is described all the necessary steps for modification of raw root file to graph data.

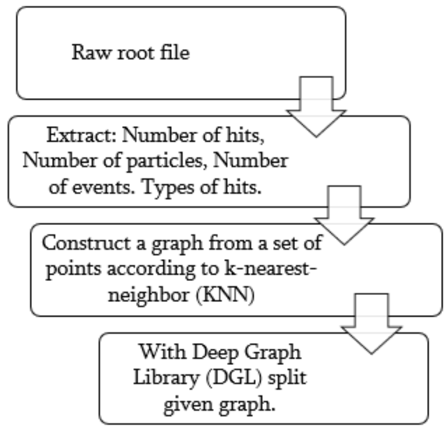


Figure 3: Processing the data set

# 3 Particle Detector Geometry

In This part it is briefly described the Clic detector geometry[2]. Geometry of the detector is written in XML file. Of course this is a simplified visualisation barrels of Calorimeters, otherwise we should include dimensions of Yoke, Anti-Solenoid, Lumical, Beamcal and Support tube, which is not necessary in our case.

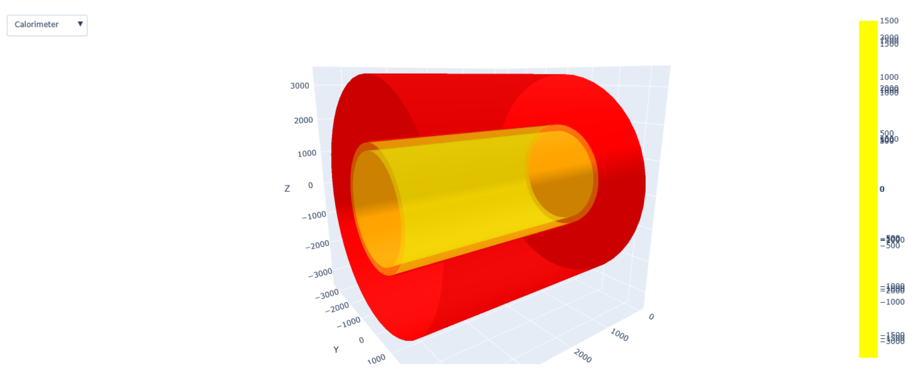


Figure 4: Visualization of Hadronic and Electormagnetic Calorimeters

Hadronic and Elecromagnetic Calorimeters

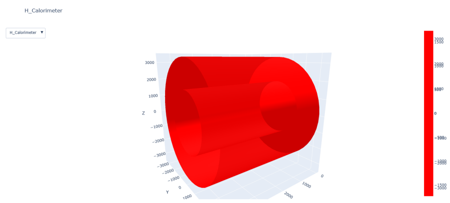

Figure 5: Visualization of Hadronic Calorimeter

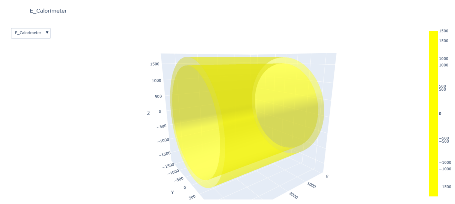

Figure 6: Visualization of Electormagnetic Calorimeter

Figure 7 illustrates a dataset of non-normalized values. The objective is to determine the appropriate calorimeter for each hit type and particle association.

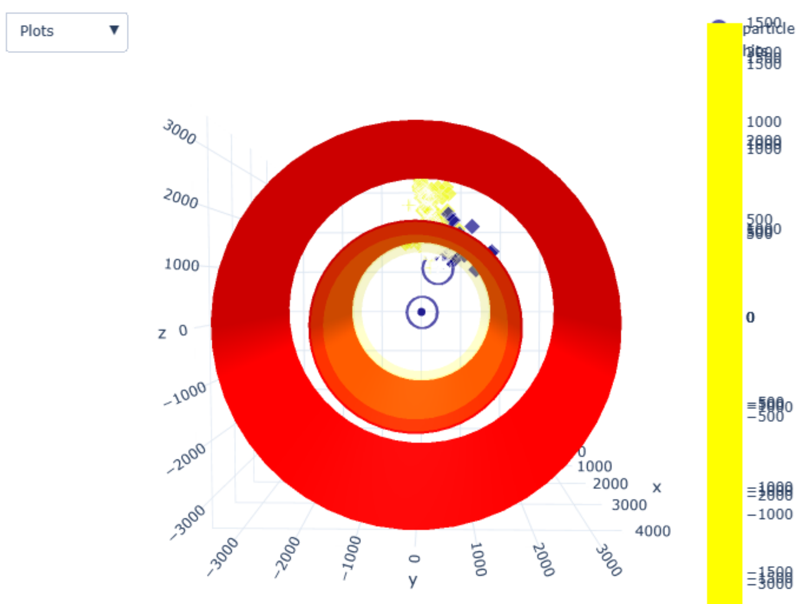


Figure 7: Non normalized data set with Hcal and Ecal

# 4 Summary

To prepare the data set for our Machine Learning model, we must first extract essential values from the raw root file. These values include particle coordinates, particle count, hit types, and particle energy. Leveraging DGL (Deep Graph Library), we can conveniently process individual graphs separately. On average, each graph contains around 758 nodes and 5075 edges. To differentiate between hit types, specifically Hcal and Ecal, we use the non-normalized data set for improved distinction.

# 5 References

[1] J.Pata, J.Duarte, J.R. Vlimant, M.Pierini, M.Spiropulu. MLPF: Efficient machine-learned particle-flow reconstruction using graph neural networks. (2021)
[2] D. Arominski, A. Latina, A. Sailer, Beam - included backgrounds in CLICdet.
[3] Particle-flow reconstruction and global event description with the CMS detector, Journal of Instrumentation, Volume 12, October 2017. [4]ATLAS Collaboration, Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment," Tech. Rep. ATL-PHYS-PUB-2017-003, 2017. [5]A. Radovic et al. Machine learning at the energy and intensity frontiers of particle physics," Nature, vol. 560, no. 7716, 2018. 7. V. Khachatryan
[6] F. Mokhar, R.kansal, D. Diaz, J.Duarte, J. Pata, M. Pierini, J.R. Vlimant. Explaining machine learned particle flow reconstruction. (2021)