CALORICH AI

An UBC MDS Capstone Project with TRIUMF, 2023

May 5, 2023

Students:
GENG, Crystal
MERIGO, Daniel
WONG, Kelvin
ZHANG, Peng

THE UNIVERSITY OF BRITISH COLUMBIA

TRIUMF

Agenda

1. Background and Scope
2. Data
3. ML Approach and Metrics
4. Timeline
5. Q&A

1. Background and Scope

NA62 Experiment

NA62 is an experiment conducted in CERN at Geneva, Switzerland.

It studies rare **kaon decays** to check some of the predictions the Standard Model makes about short-distance interactions.

Image Credit: CERN

Subatomic Particles

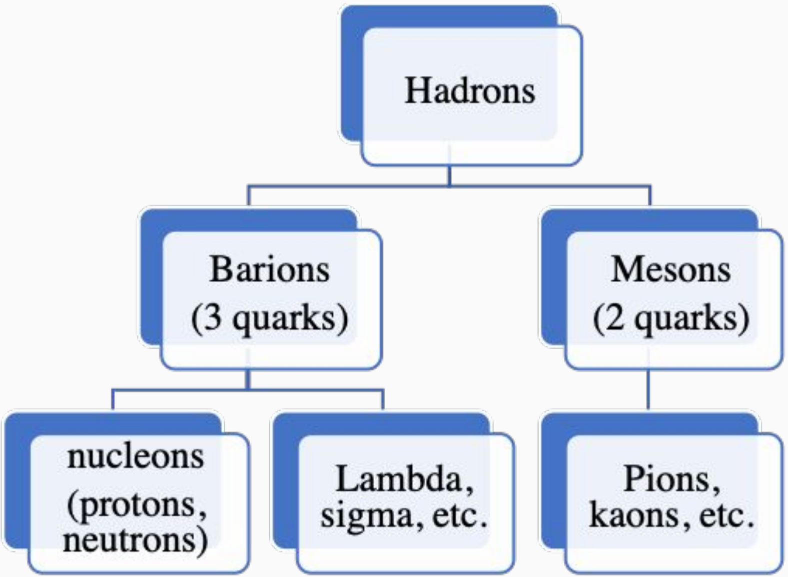


Image credit: INFN

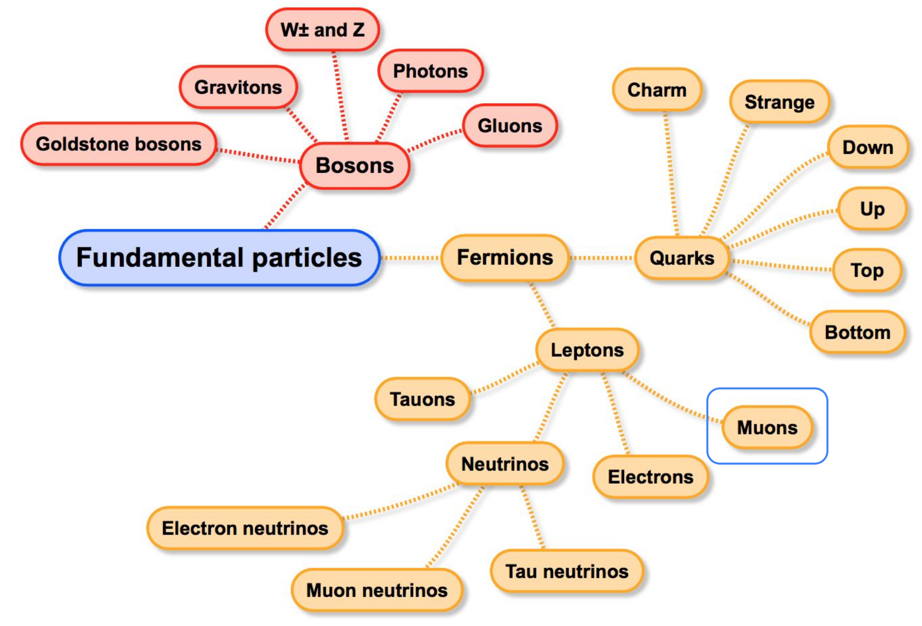
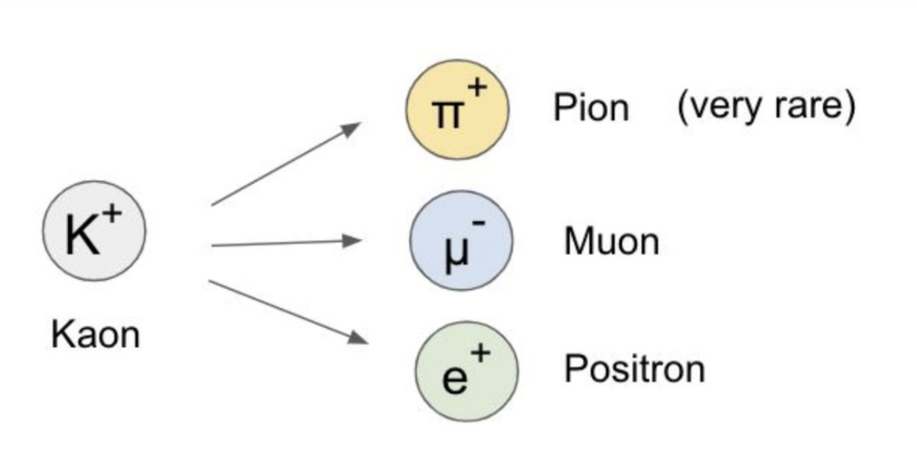


Image credit: wire

Kaon Decay

The rare kaon decay we are trying to capture involves **pion**. However, there are other pathways that involves **muon** instead!

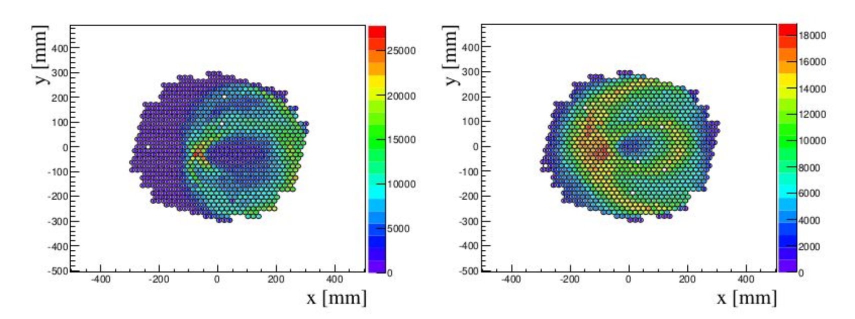
We would want to distinguish between these two particles from the observed data with a better accuracy.



RICH Detector

RICH detector from the NA62 experiment detects the hit pattern generated from the decay due to Cherenkov radiation.

We use this sensor data to do **classification**.



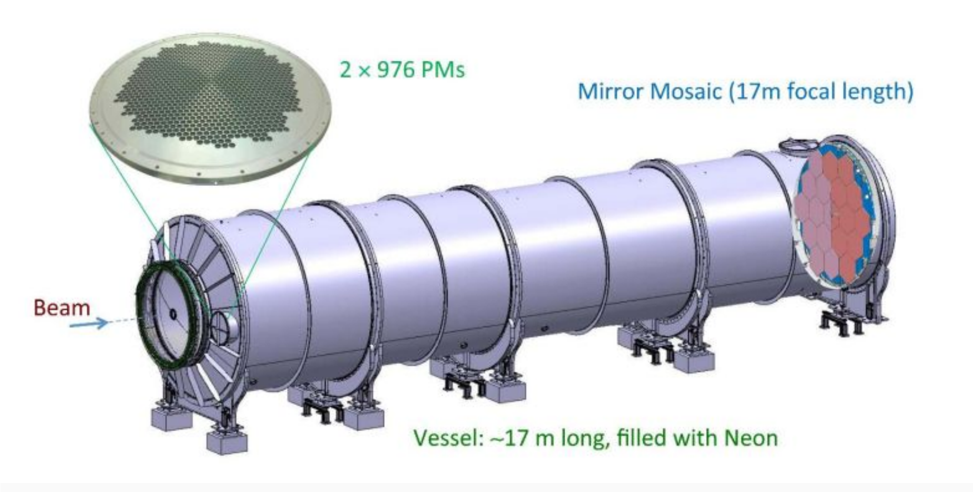
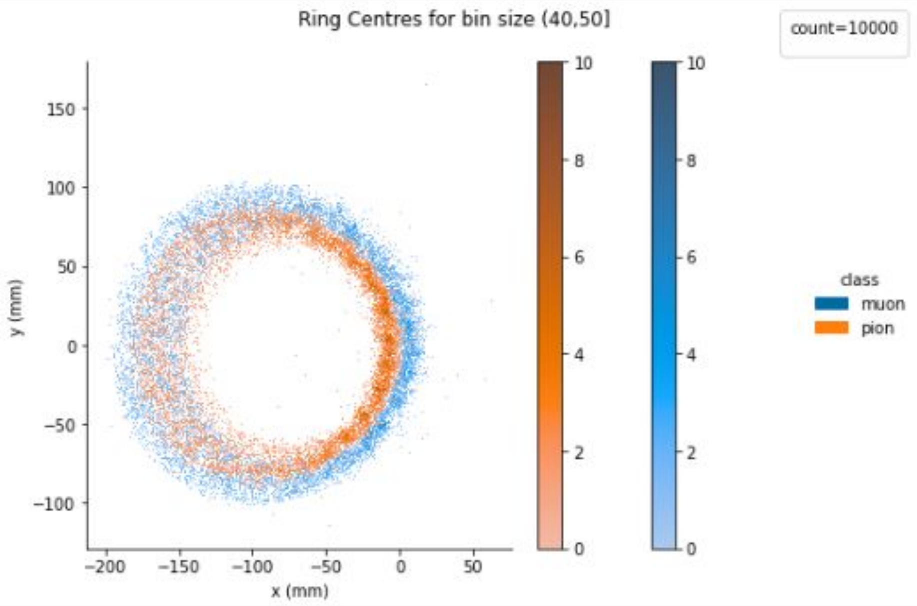


Image Credit: doi:10.1088/1748-0221/12/05/P05025

Deliverables

Our main deliverable is a trained model that **distinguishes** between **pion** and **muon**.

- It should make use of the fitted ring function and ring-momentum function



2. Data

Dataset

We are working on a small slice of data (~2.7GiB in HDF5 format).

It contains around **2.4M events**, labelled as pion or muon by another instrument (calorimeter).

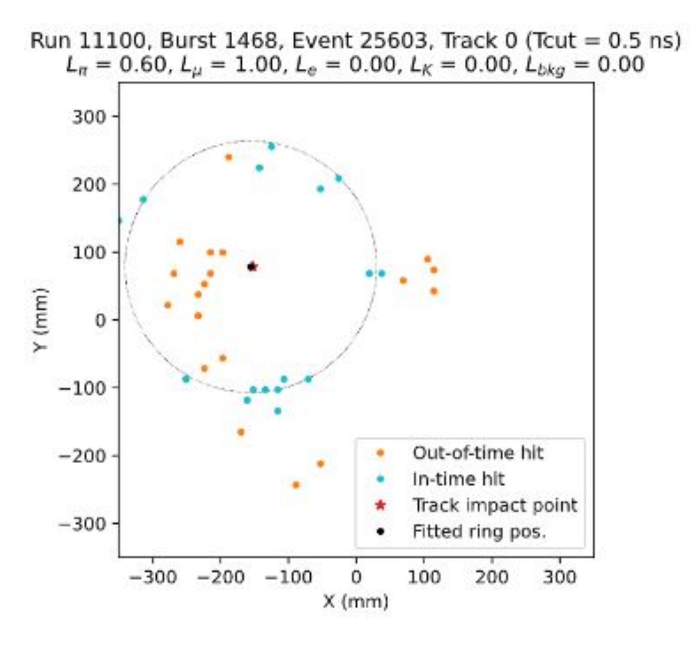
Each event can have a variable number of hits. There are around **101M hits** across all events.

An Event

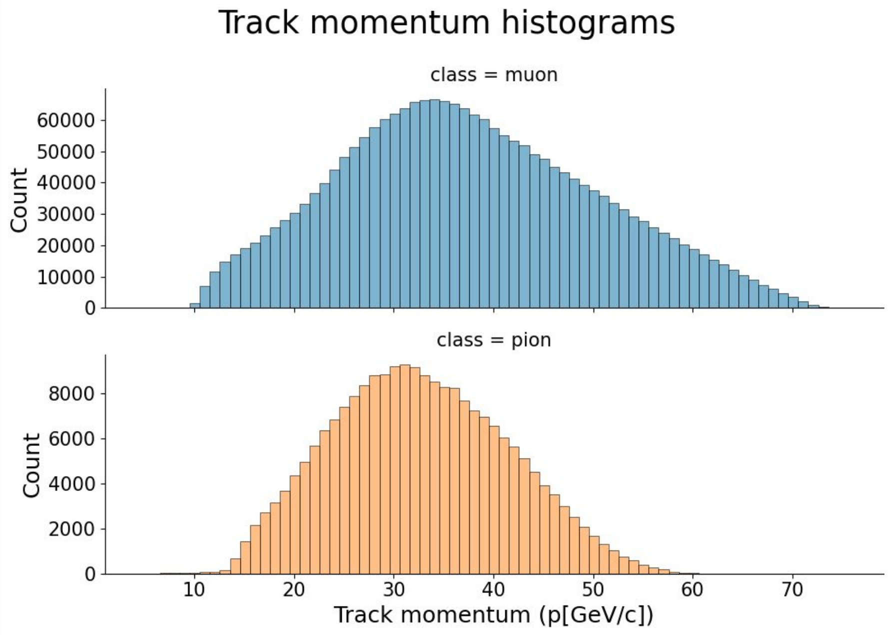
An **event** refers to a decay (could be pion or muon), and the **hit** refers to the sensor excited.

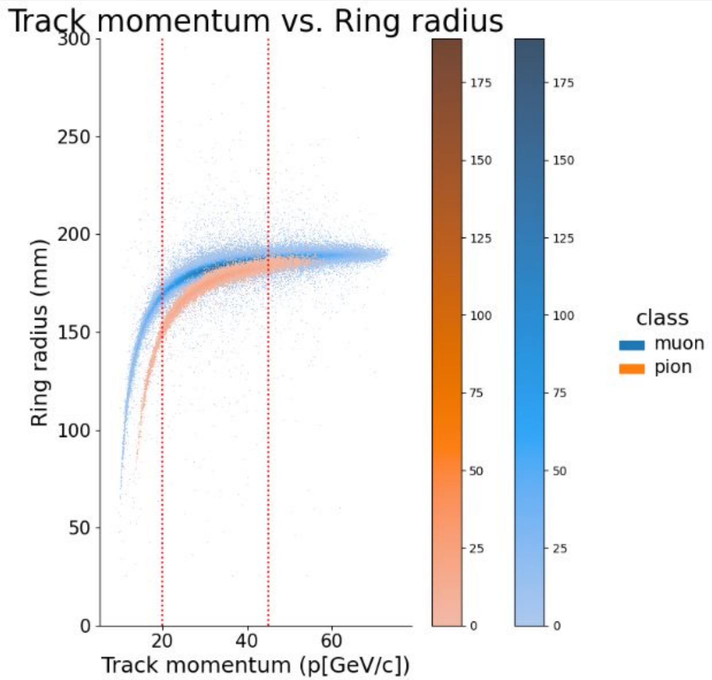
Each blue point represents an **in-time hit**, that we will use to train our model.

We will ignore the out-of-time hits (orange), as they are background noise.

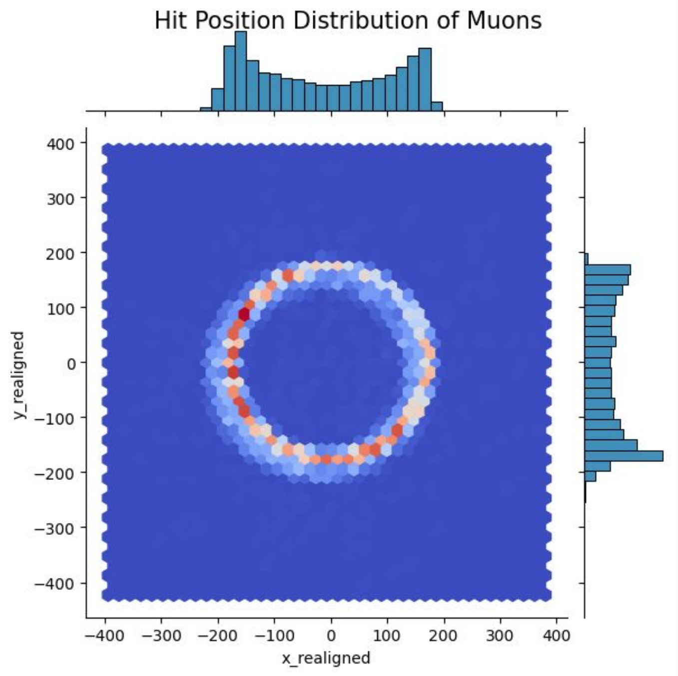


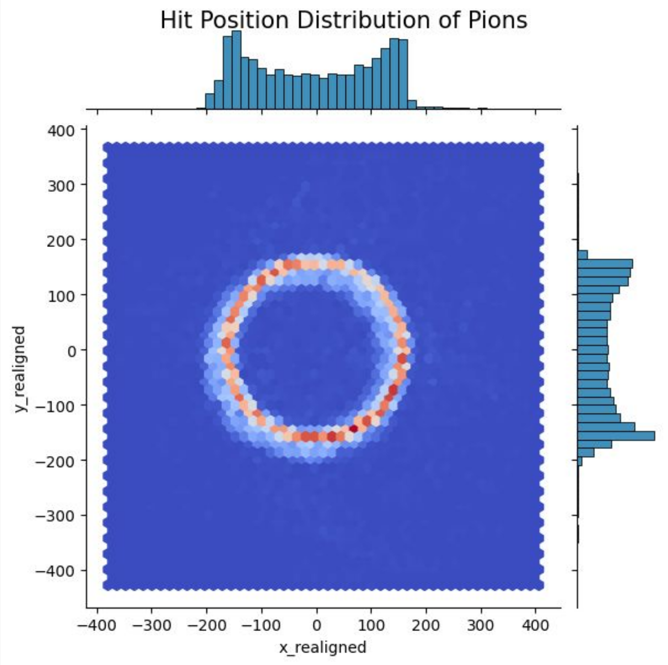
EDA on Events Data





EDA on Hits Data

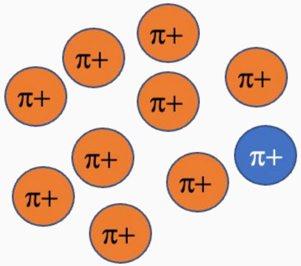


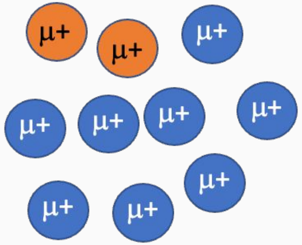


3. ML Approach and Metrics

Approach 1: Classification

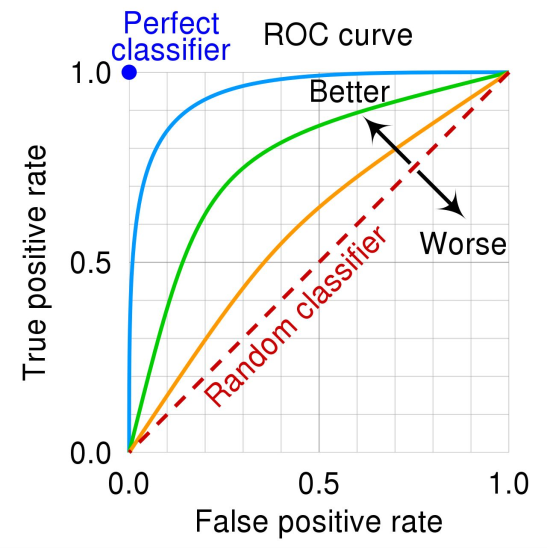
- Neural networks (prior work)
  - PointNet
  - Dynamic Graph CNN
- Metrics: “Efficiency” (TPR/FPR)





$$\text{Pion Efficiency} = \frac{TP}{TP + FN}$$

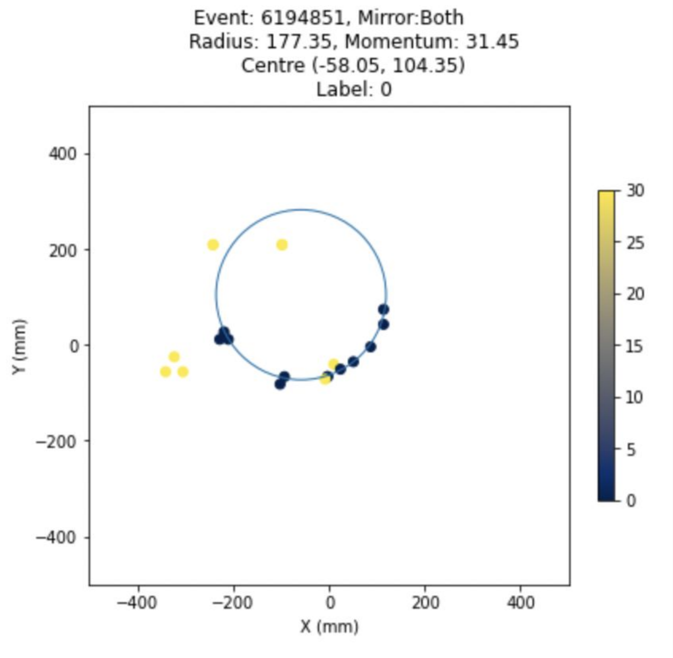
$$\text{Muon Efficiency} = \frac{FP}{FP + TN}$$



Source: Wikipedia

Approach 2: Regression

- From the **hit location** (x and y), we can find x_max, x_min, y_max, y_min, and the largest distance between two points
- Using these features, fit a regression model to predict the ring radius
- Compare the predicted ring radius with the theoretical value
- Metrics: RMSE, MSE, R^2



4. Timeline

Timeline

W1	W3	W5	W7	W8	W9
Proposal	**ML Approach**	**Prepare Draft Report**	**Final Presentations**	**Feedback**	**Data Product Delivery**
Hackathon	Brainstorm and fit ML models	Prepare data for submission	Prepare report	Analyze and apply feedback	Prepare final tool
Proposal presentation	Prepare and contrast results	Model training	Analyze results		Organize repository
	Brainstorm features to optimize	Modularize code			Present to partner

Thank you!

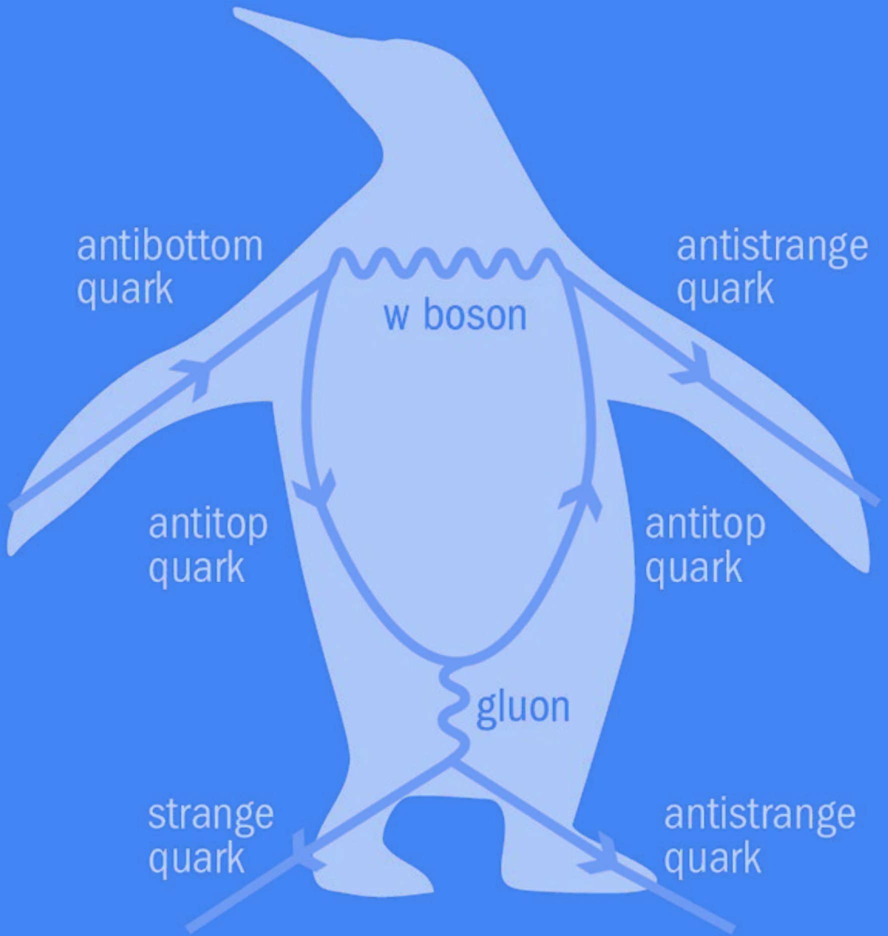
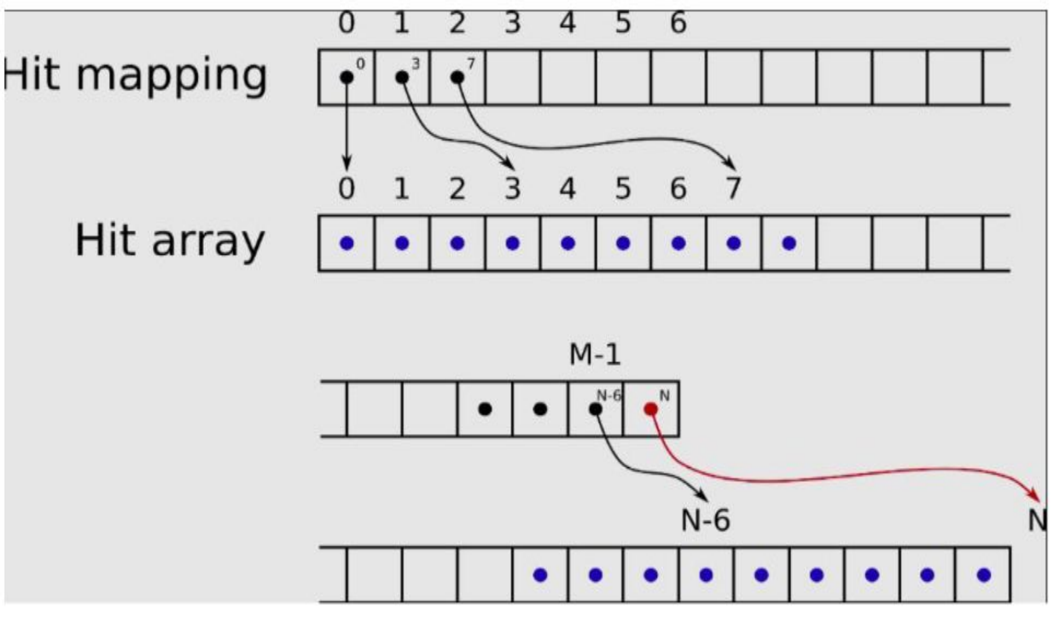


Image Credit: CERN

Appendix 1
Data Structure



Source: RICH AI