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

Agenda

1. Background and Scope
2. Data
3. ML Approach and Metrics
4. Timeline

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

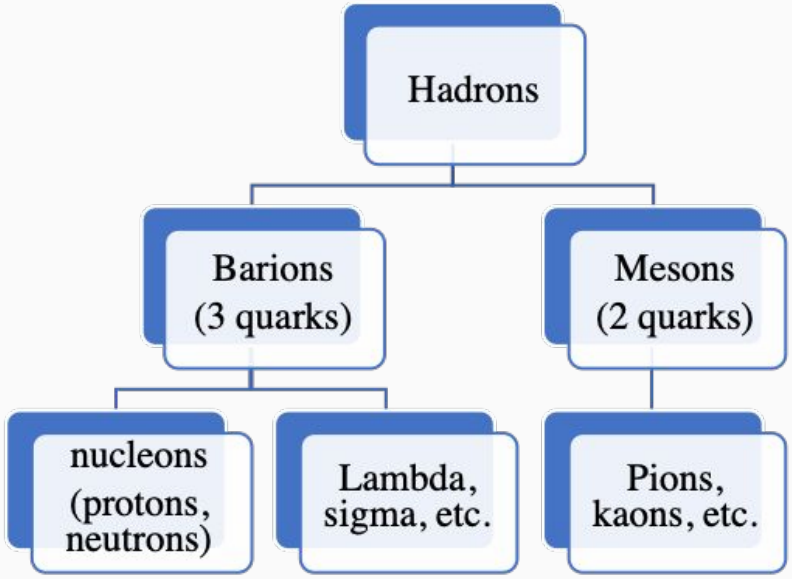


Image credit: INFN

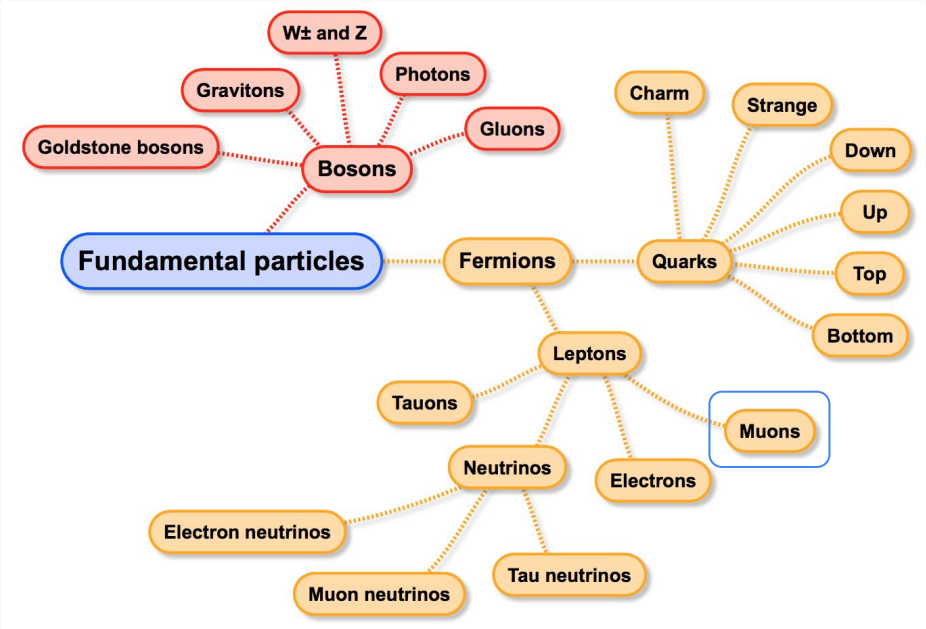
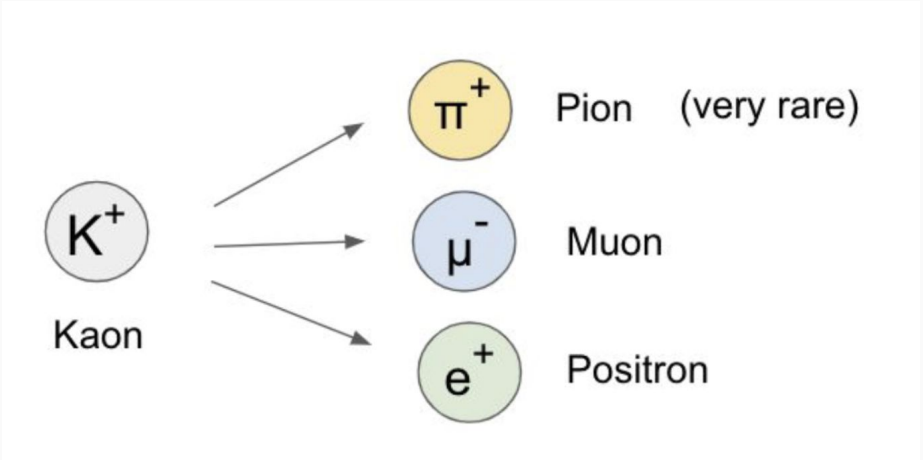


Image credit: wire

Kaon Decay

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

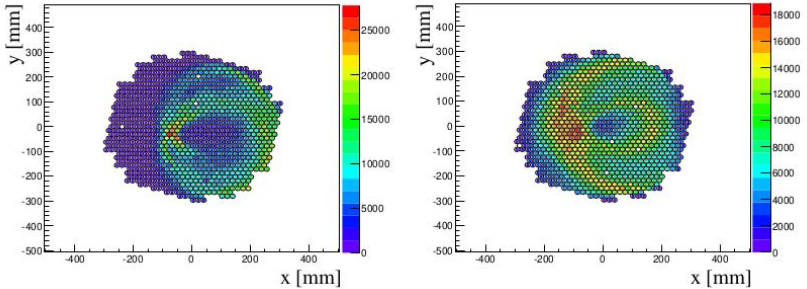
We would want to distinguish between these two particles from the observed data with a better accuracy.



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**.



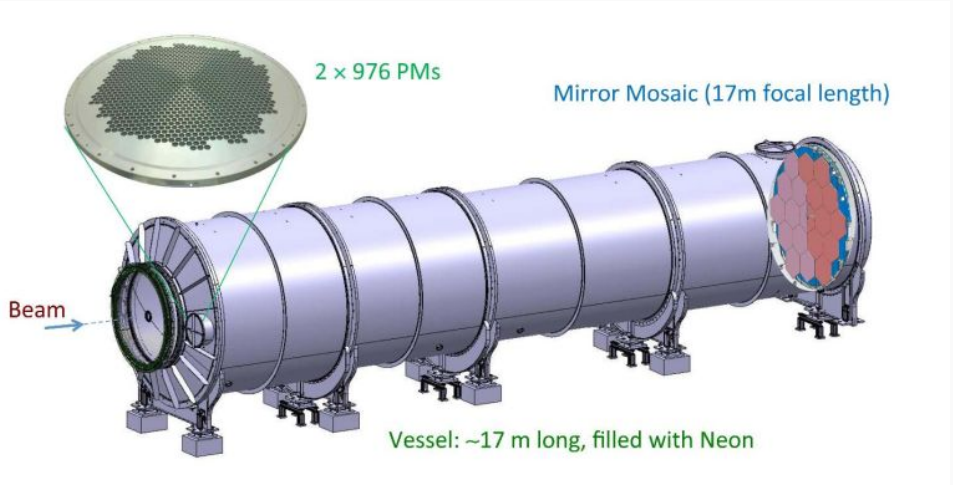
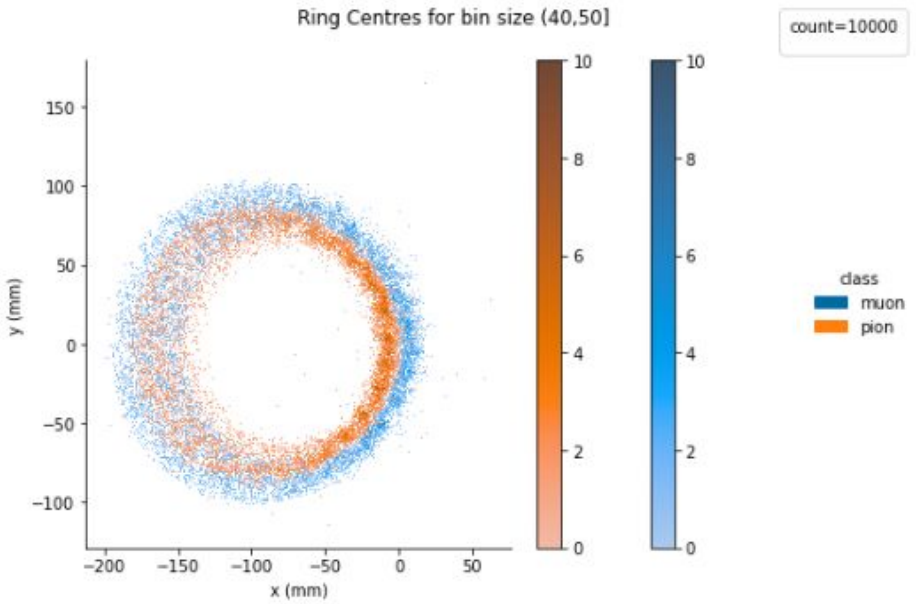


Image Credit: doi:10.1088/1748-0221/12/05/P05025

Deliverables

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

- Making 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 **2M events**, labelled as pion or muon by another instrument (calorimeter).

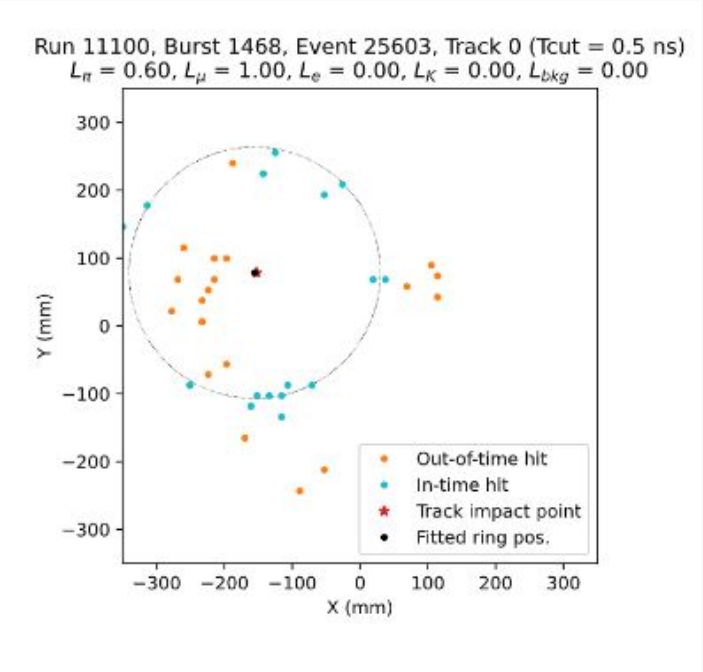
Each event can have a variable number of **hits**. There are around **100M** record of them across all events.

An Event

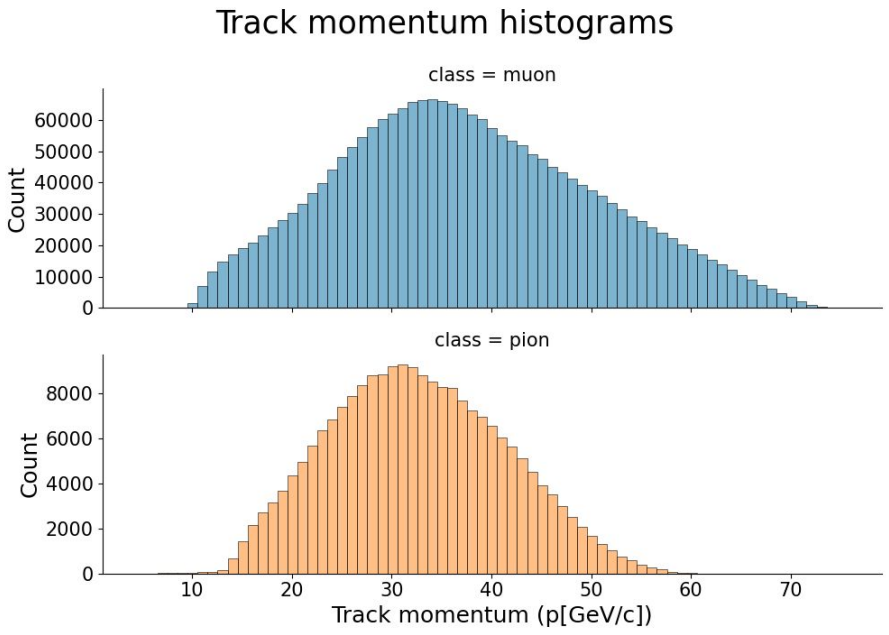
An **event** refers to a decay (could be pion or muon), and the **hit** refers to the sensor excited.

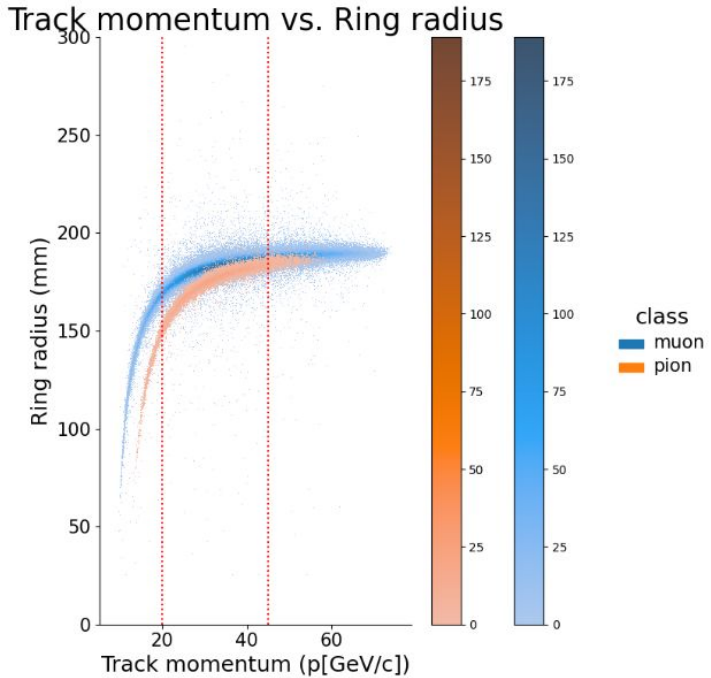
Each blue point represents an **in-time hit**, that we will use to train our model.

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

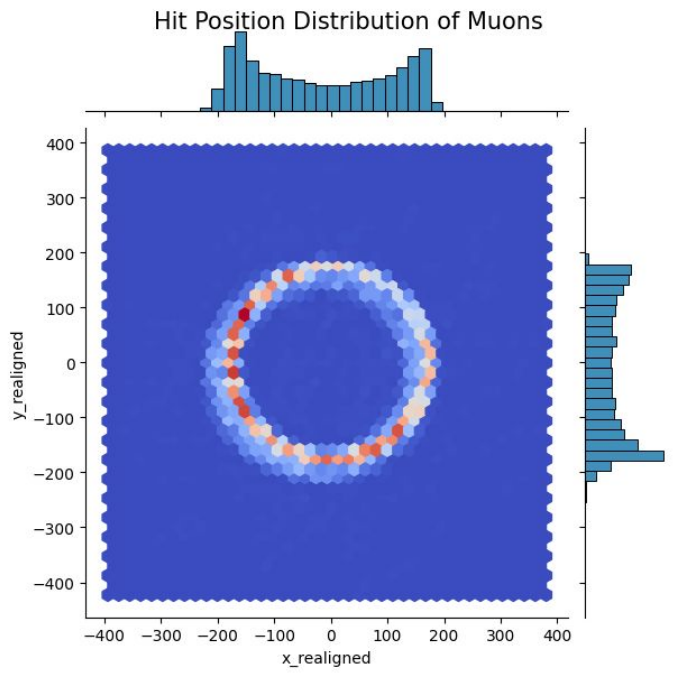


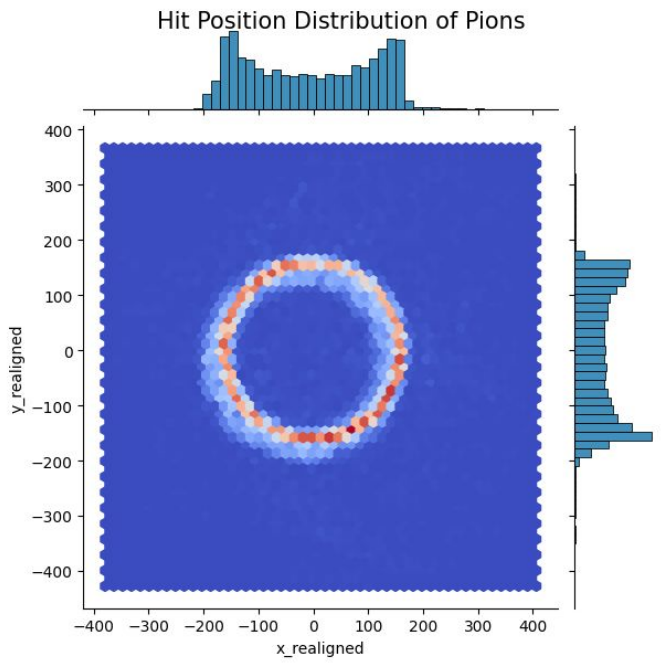
EDA on Events Data





EDA on Hits Data

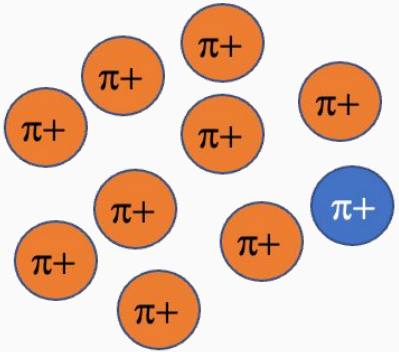




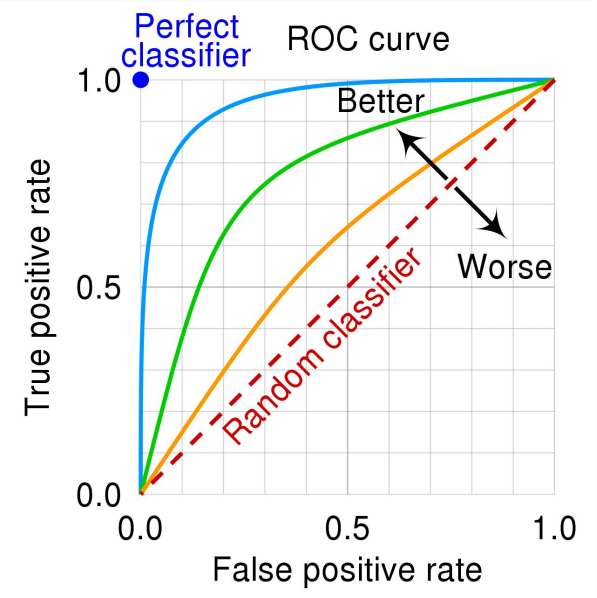
3. ML Approach and Metrics

Approach 1: Classification

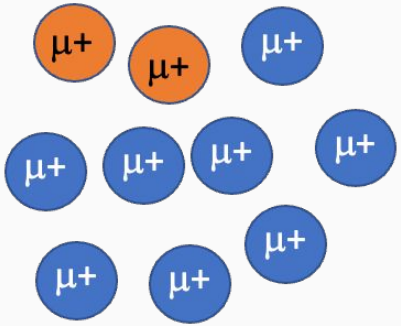
$$\text{Pion Efficiency} = \frac{TP}{TP + FN}$$



Metrics: Accuracy (Efficiency)



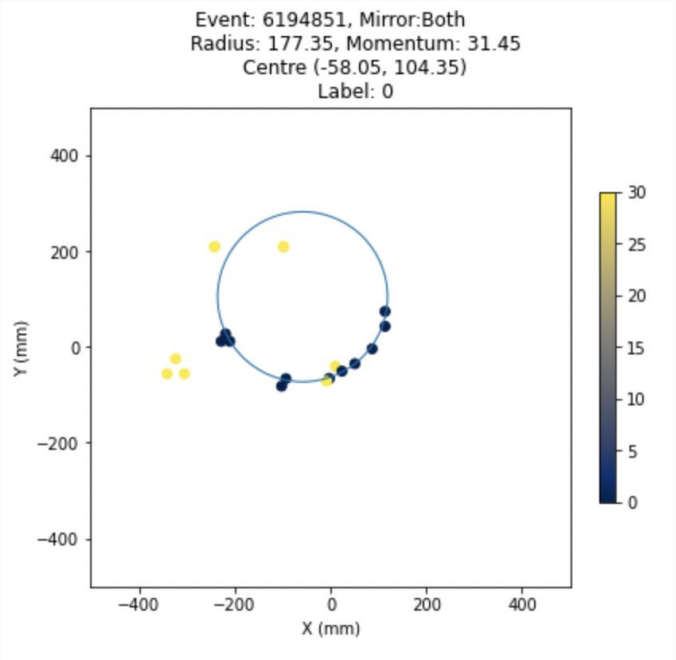
Source: Wikipedia



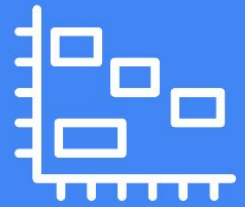
$$\text{Muon Efficiency} = \frac{FP}{FP + TN}$$

Approach 2: Regression

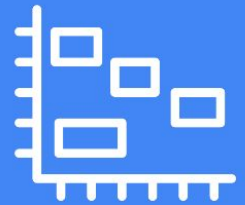
- From the **hit location** (x and y), we can find x_max, x_min, y_max, y_min, and the largest distance between two points
- Using these features, fit a regression model to predict the ring radius
- Compare the predicted ring radius with the theoretical value
- Metrics: RMSE, MSE, R^2

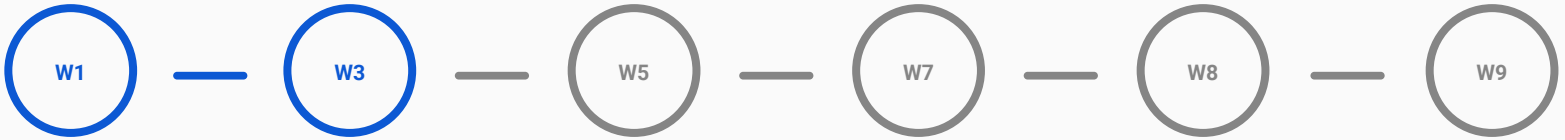


4. Timeline



Timeline





Proposal

Hackathon

Proposal Presentation

ML Approach

Brainstorm and Fit ML Models

Prepare and contrast results

Brainstorm Features to optimize

Prepare Draft Report

Prepare data for submission

Model Training

Modularize Code

Final Presentations

Prepare Report

Analyze Results

Feedback

Analyze and Apply Feedback

Data Product Delivery

Prepare Final Tool

Organize Repository

Present to Partner

Thank you!

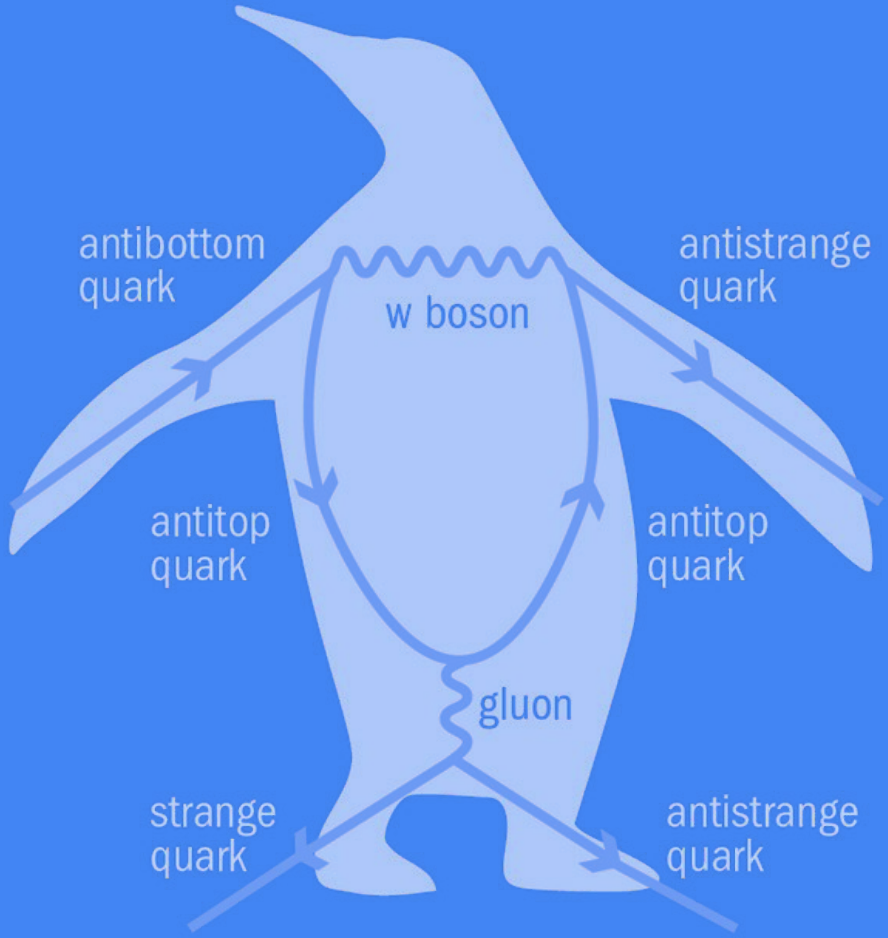
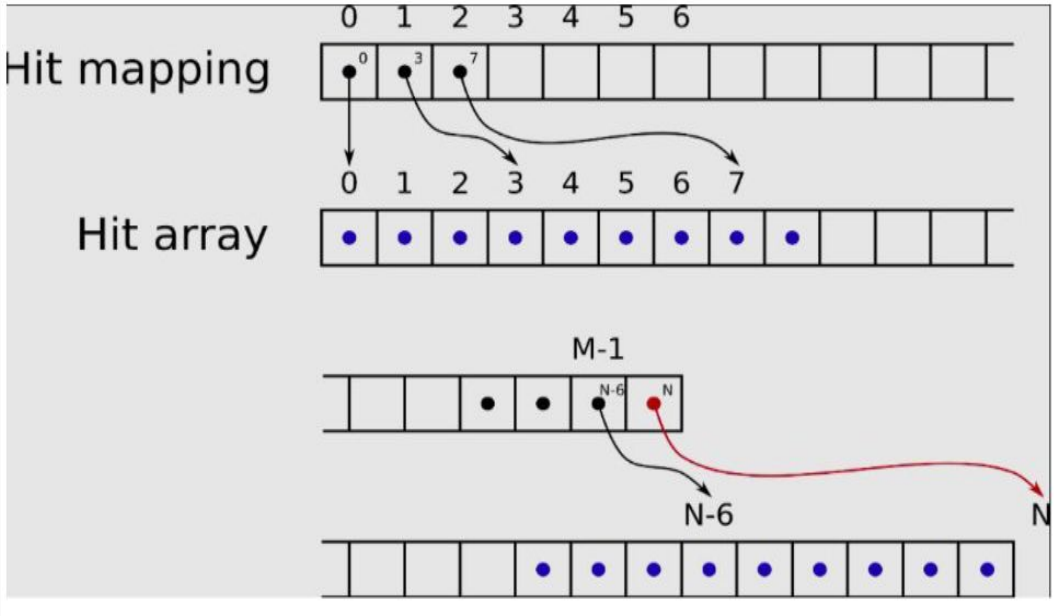


Image Credit: CERN

Appendix 1
Data Structure



Source: RICH AI