Toward Particle ID in Granular Hadron Calorimeters

Andrea De Vita, Tommaso Dorigo,
Abhishek, Joseph Willmore

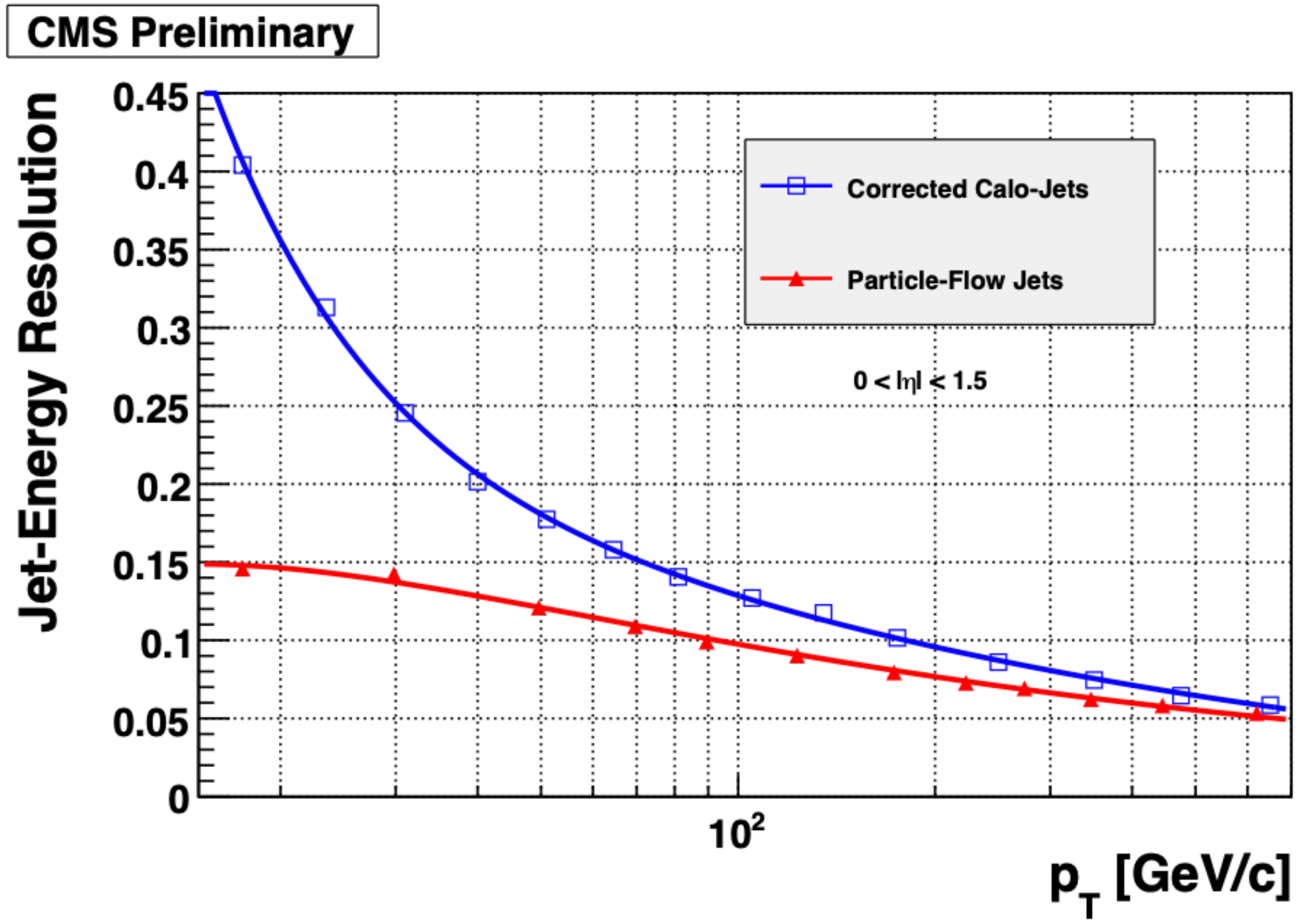
1st November 2024

Why do we need granular calorimeters?

Physics reach will include SM & Higgs, with searches for BSM including reactions initiated by **Vector Boson Fusion** (VBF) and **including highly-boosted objects**.

High-precision timing of particles mitigate the effects of pileup.

Particle Flow approach relies on high-precision tracking and finely segmented calorimeters for effective reconstruction.



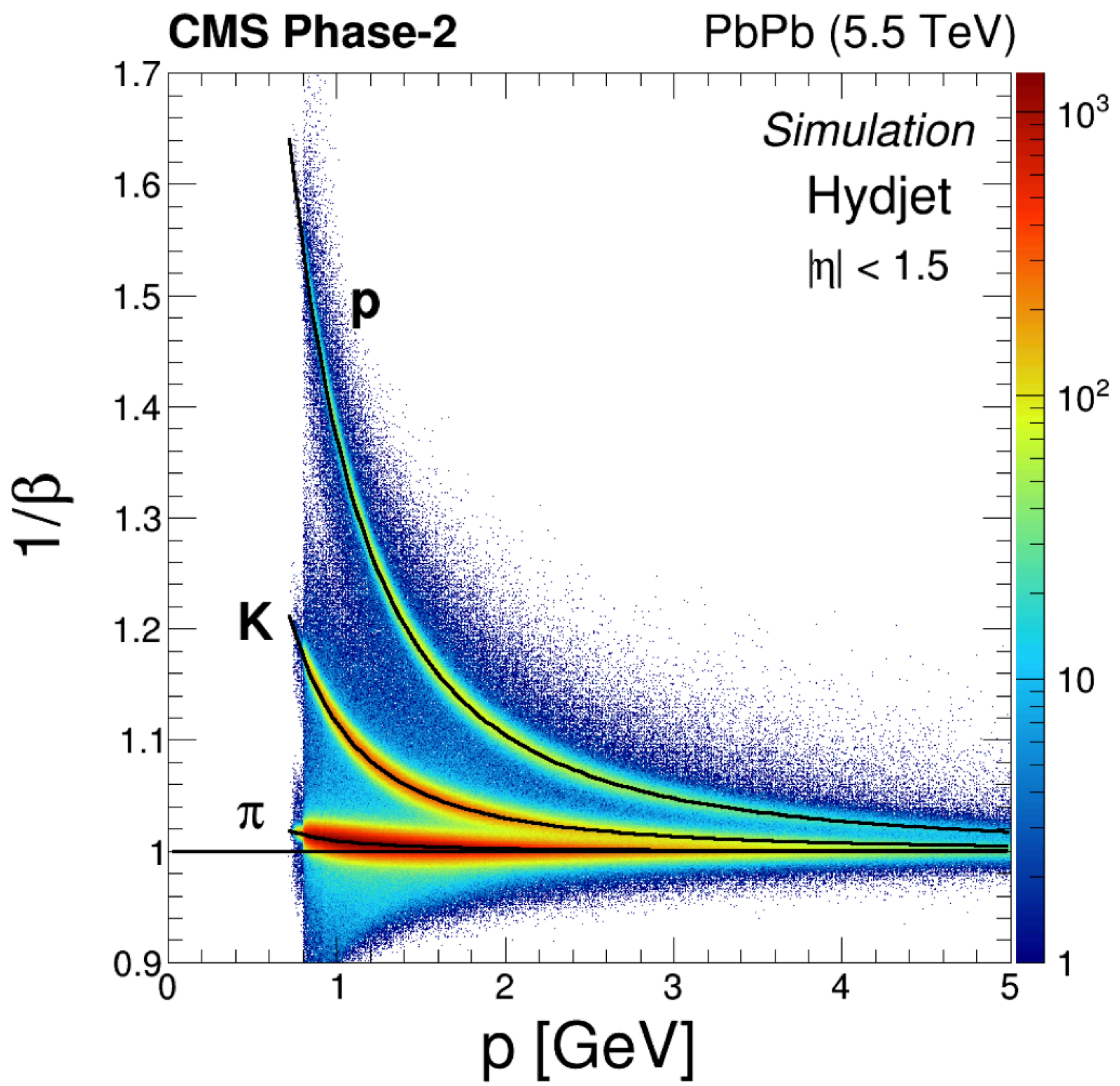
Can they also be useful for other purposes?

Particle Identification

Hadronic shower studies rely on high-purity data, but particle discrimination is difficult due to mixed test beams and limited calorimeter setups.

Fine segmentation offers observables for **calorimeter-based particle identification**.

MTD layers enable particle identification by measuring time-of-flight differences, but their **discrimination power decreases with 1/β**, thus requiring additional methods for high-energy particle identification.

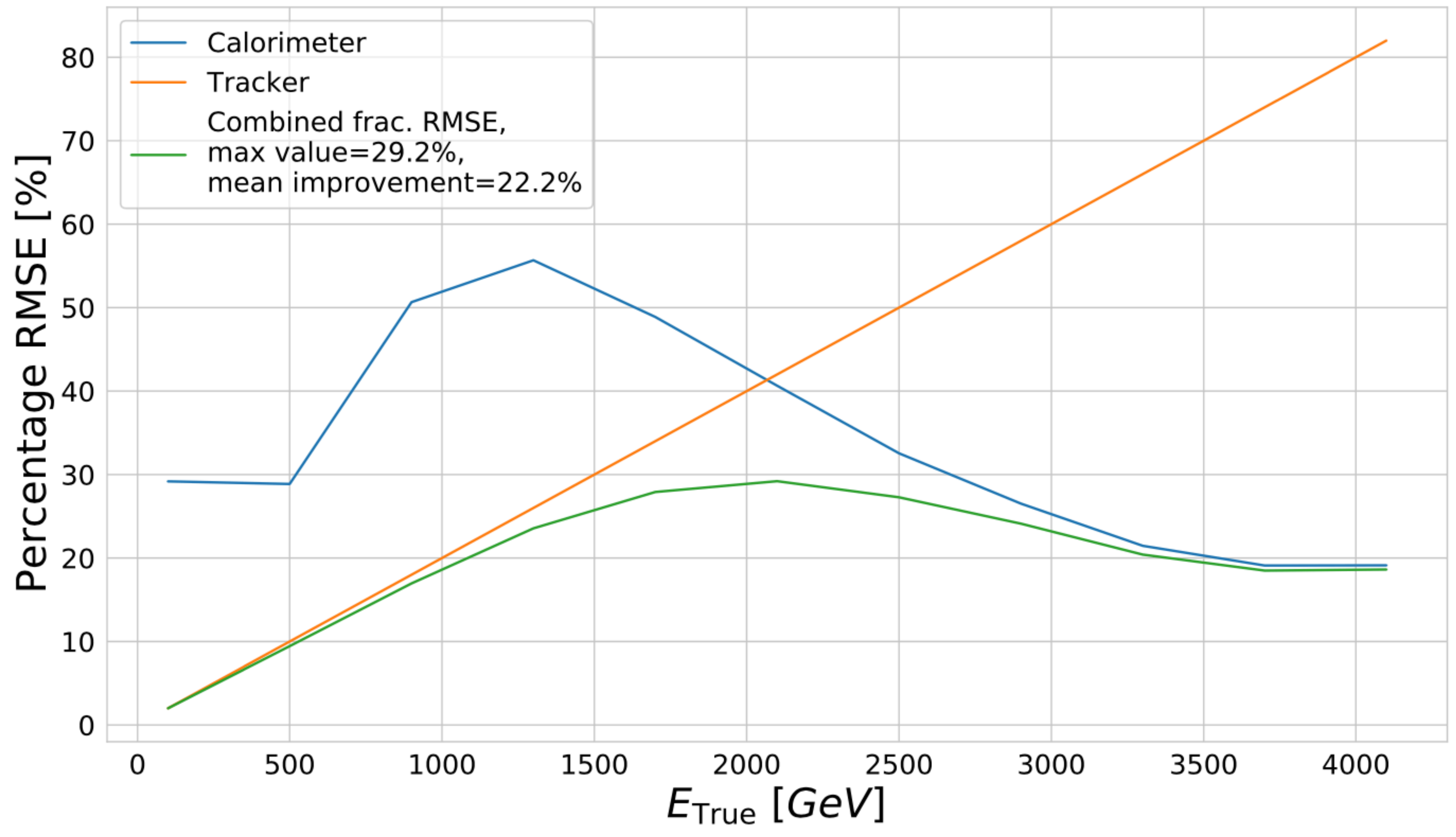


Combination of granular calorimeters and ML

Calorimetric Measurement of Multi-TeV Muons via Deep Regression

The use of granular calorimeters provides information not only on the intensity of the energy released in the calorimeter, but also on **the pattern of energy deposits detected in the calorimeter cells**.

It has been shown that combining momentum measurements from the tracker with signals from the calorimeter significantly **enhances the relative resolution of Multi-TeV muon energy**.



Jan Kieseler et al., Calorimetric Measurement of Multi-TeV Muons via Deep Regression (2022) : arXiv:2107.02119

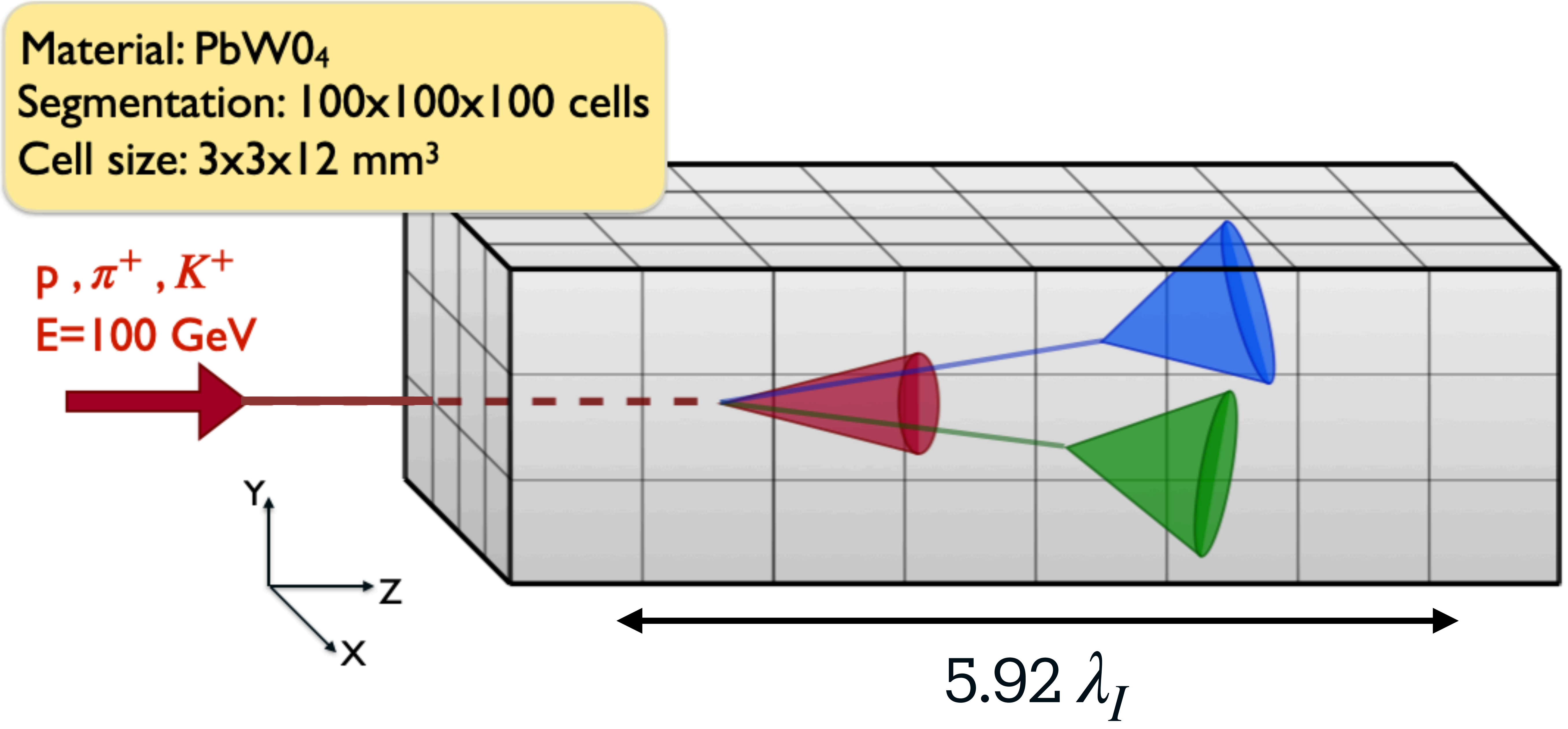
From the Simulation

Highly Segmented Calorimeter

Calorimeter composed of **1M cells**, where each cell is defined by three parameters: **position, total energy deposited, and a characteristic time**.

The characteristic time of each cell is defined as the **weighted average of the interaction times**.

$$\hat{t}_{cell} = \frac{\sum_i t_i \cdot E_i}{\sum E_i} = \frac{\sum_i t_i \cdot E_i}{E_{cell}}$$



From the simulation

Observable and Latent Spaces

The output of the simulation represents the **collection of all the steps** that deposited energy inside the calorimeter. For each of them, information belonging to the observable space is saved, but also information that is unknown in a real experiment.

- Tevent_id
- Tinteractions_in_event: number of entries
- Tpdg: particle id (from particle data group)
- Track: track id
- Tparent_id: parent track id
- Tmom: post Step momentum
- Tedep: Total Energy Deposit
- Tdeltae: post Step Kinetic energy - pre Step Kinetic energy
- Tglob_t: post Step global time
- Tcublet_idx: cubelet index within the whole calorimeter (0 to 999)
- Tcell_idx: cell index within a specific cubelet (0 to 999)

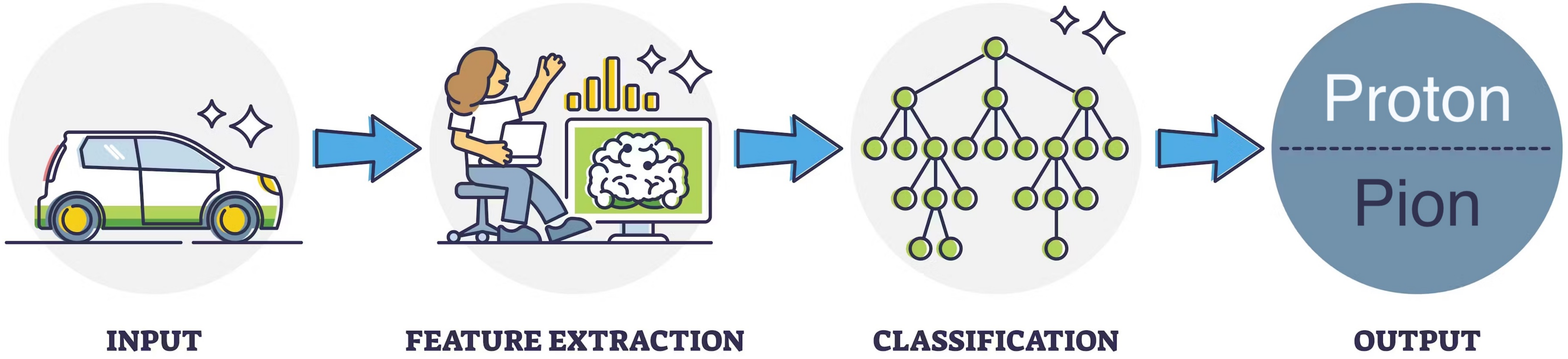
Step selection rules:
deltaE > 100 keV
edep > 1 keV

Project Setup

From Simulation to Physical Information + ML

From the simulation, it is possible to extract **a set of global and local variables** describing the showers within the calorimeter.

For each event, **50 features are generated and used as input to a machine learning model** (DNN or BDTs) to classify the primary particle.



High Level Features

Spatial distribution of energy deposits : radius, length, peak positions...

Characteristic shower times : vertex time, average event time...

Released energy magnitude : total energy, total energy close to the primary vertex...

Primary vertex identification

Introduction

The primary vertex is defined as the point where track 1 stops (momentum = 0)

How can we locate it?

In a real experiment we do not have access to the momentum of the particle inside the calorimeter.

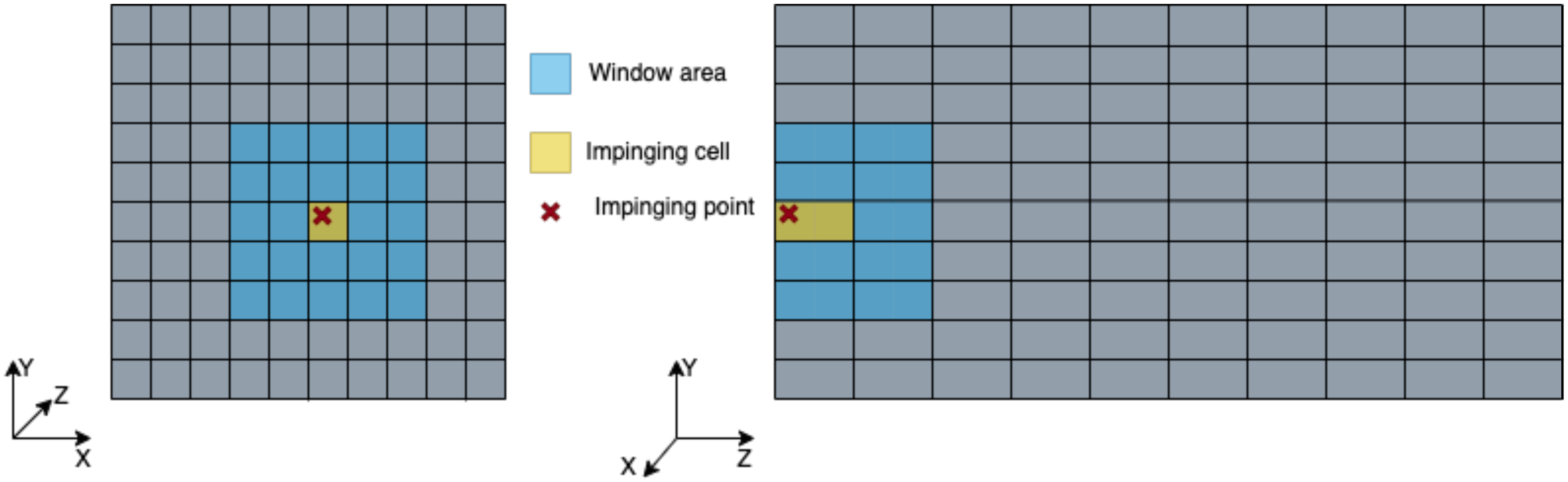
The primary vertex can be identified by an energy peak.

Primary vertex identification

How the algorithm works 1/2

The algorithm is based on a **moving 3D filter with a tuneable size**.

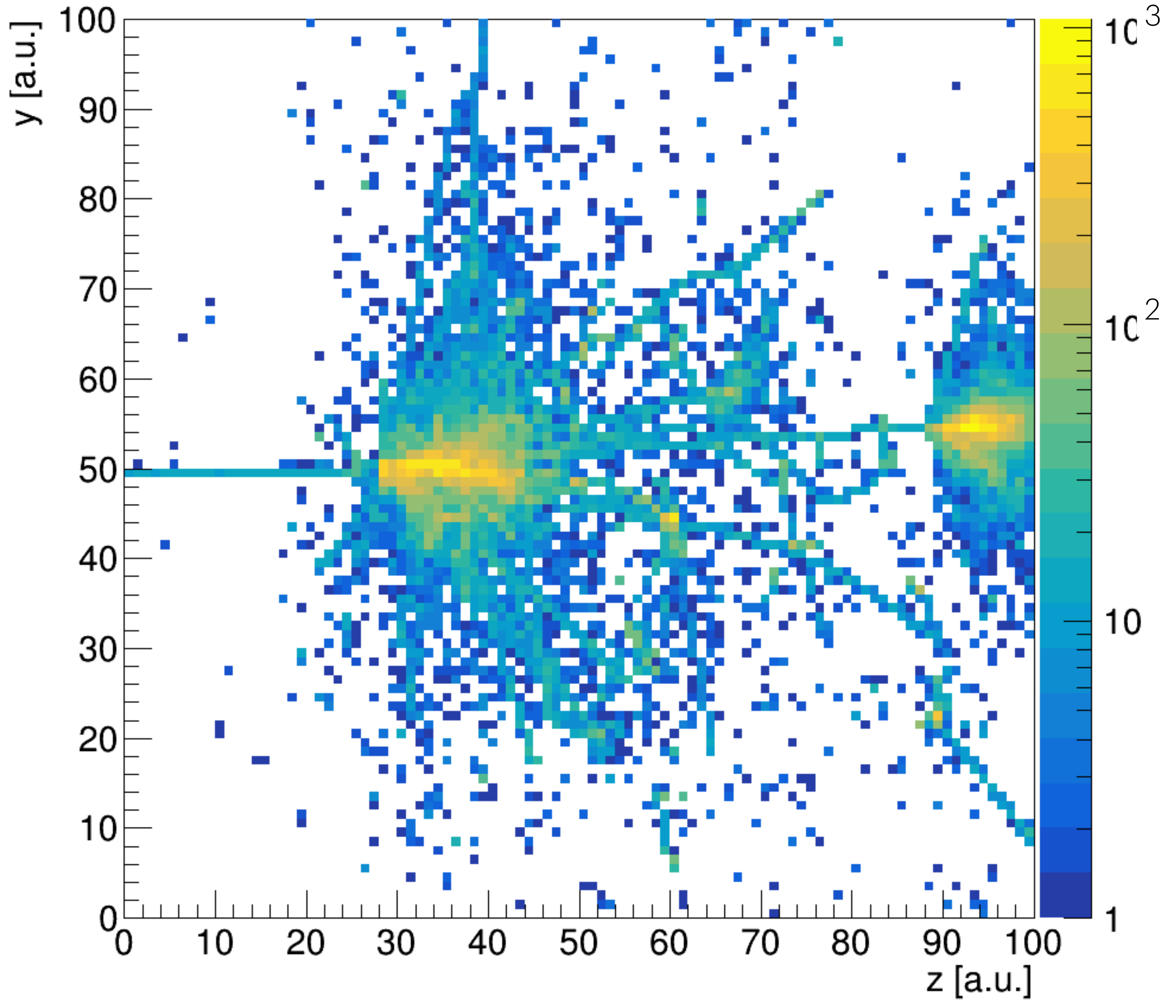
Assuming the beam impact position is known, I sum all the energy deposits of the cell around this point (within a certain user-defined window). The **moving filter works on the z-axis**, so I move the z cursor along the entire calorimeter and I collect a series of values that can be analysed.

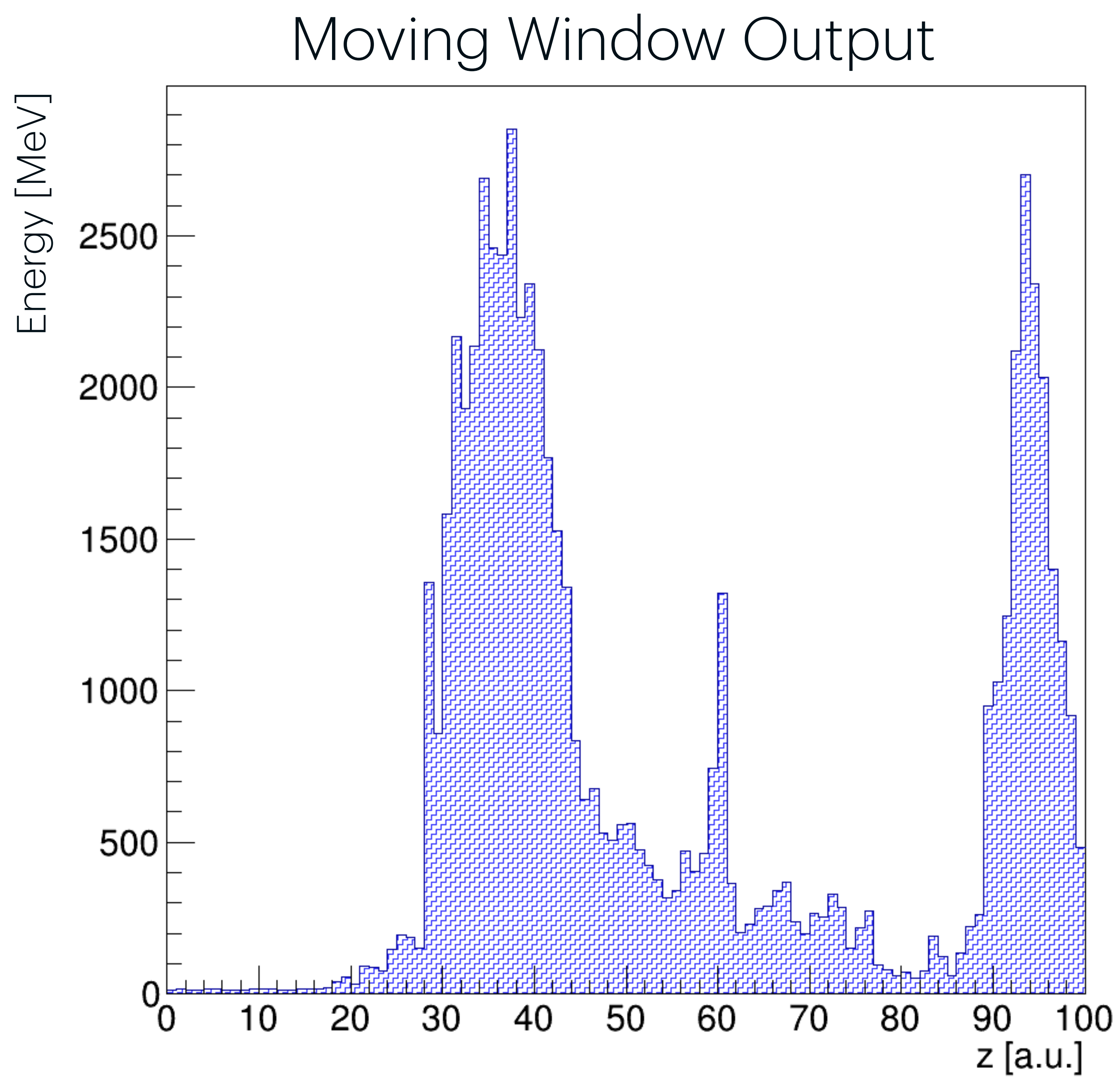


Primary vertex identification

How the algorithm works 2/2

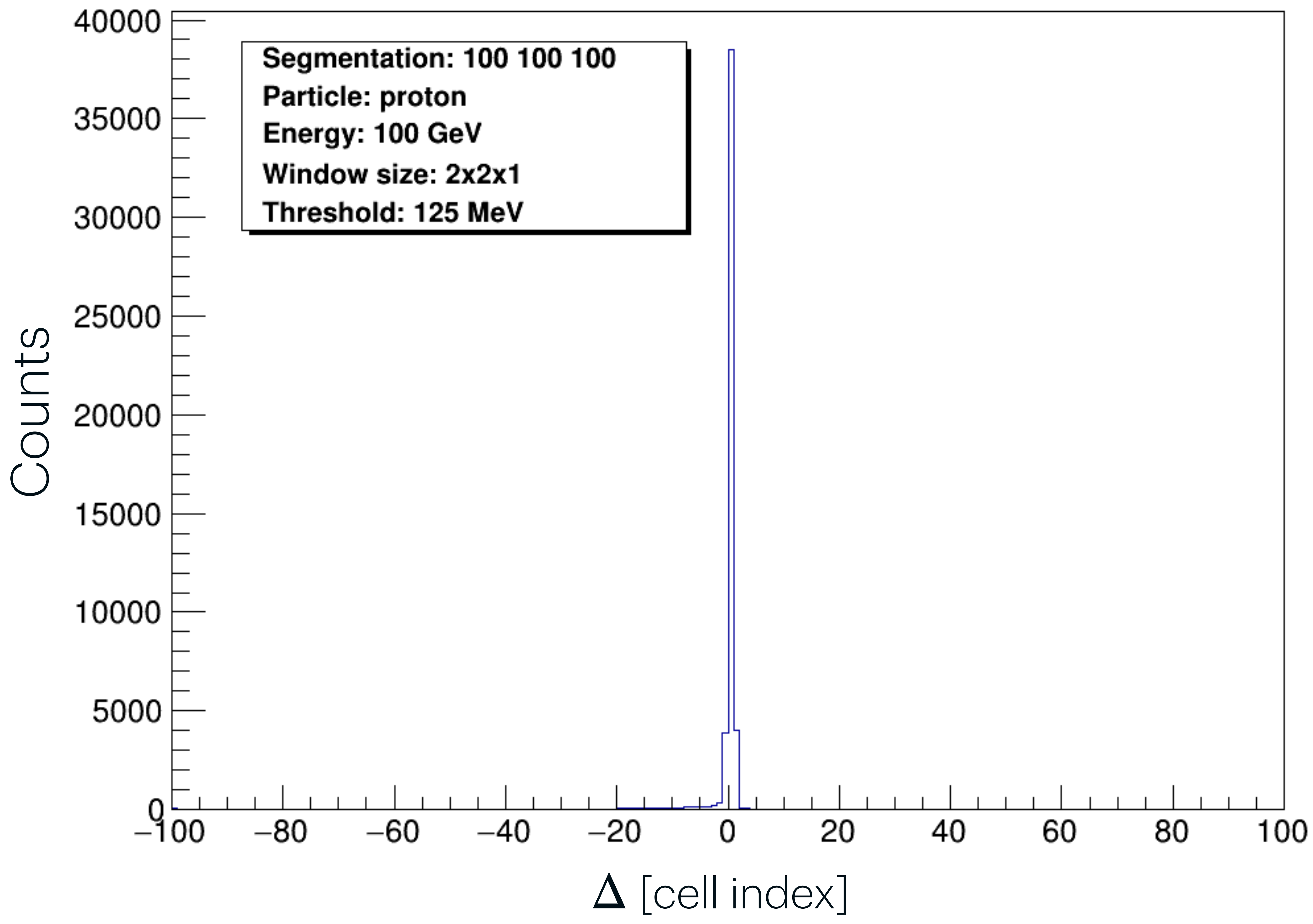
Given the series of energies from the moving filter step, the z layer corresponding to the primary vertex is the **first time when a certain threshold is crossed**.





Primary vertex identification

Performance

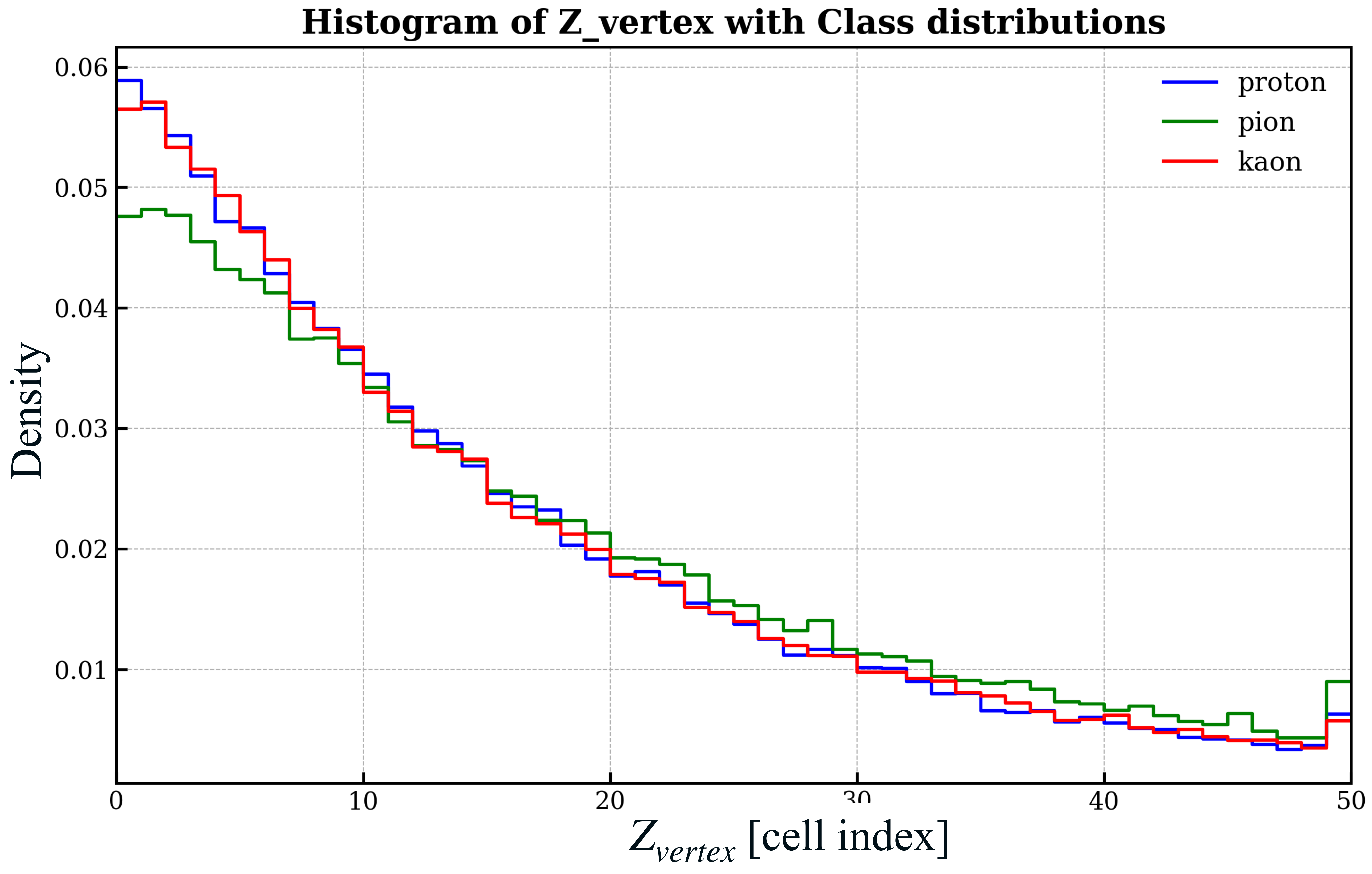


Using a segmentation of 100x100x100, a perfect **accuracy of 85%** can be achieved, while if we consider a **difference less than or equal to 1, we reach 94%**.

$$\text{Accuracy}_n = \frac{\text{number of events with } \Delta < n}{\text{total number of events}}$$

Primary vertex identification

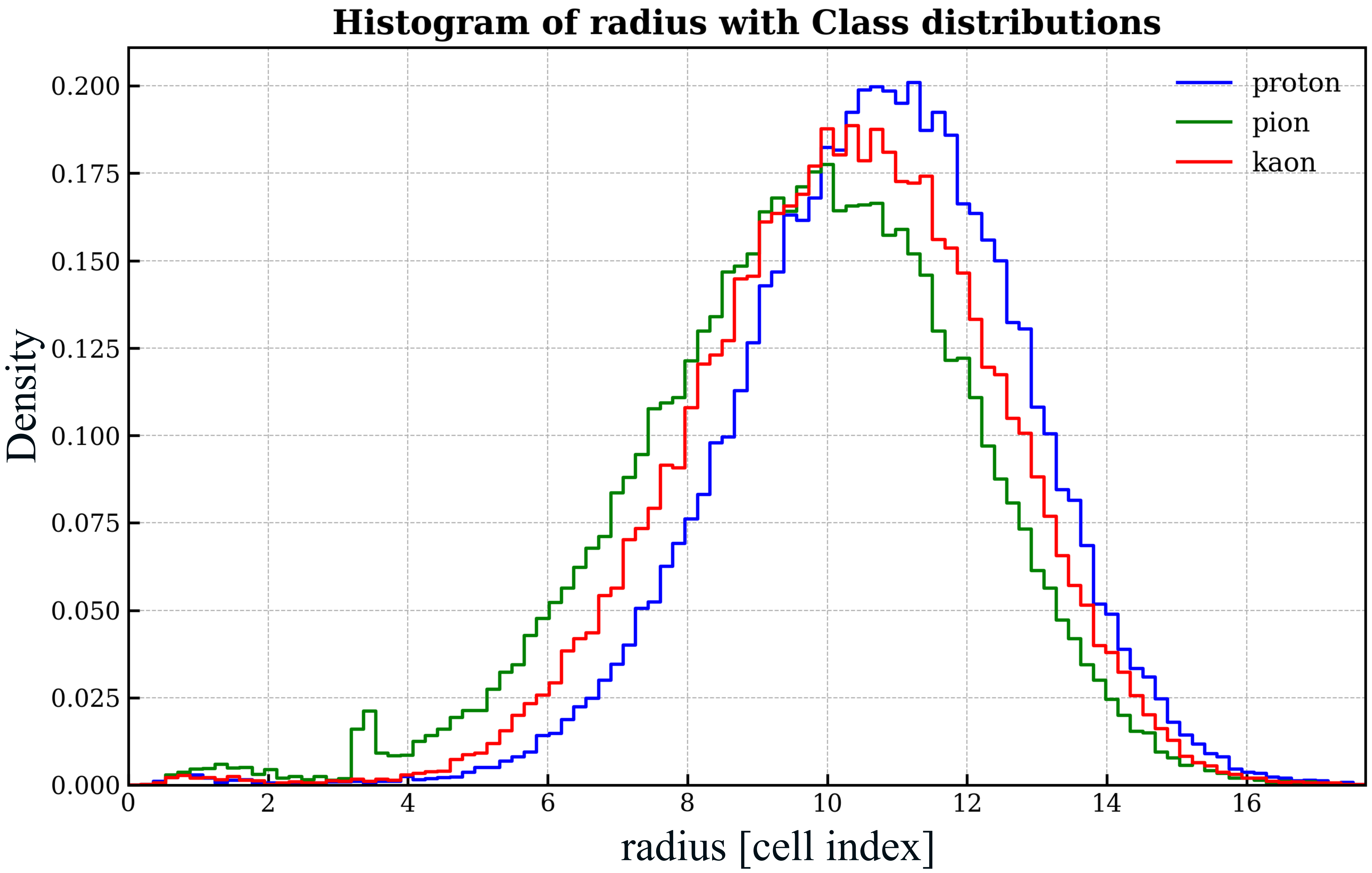
Feature Distributions for Protons, Pions and Kaons.



The pion distribution shows a difference compared to the other two particles in the region where Z_{vertex} is close to 0.

Shower Radius

Another Example from the Features Table

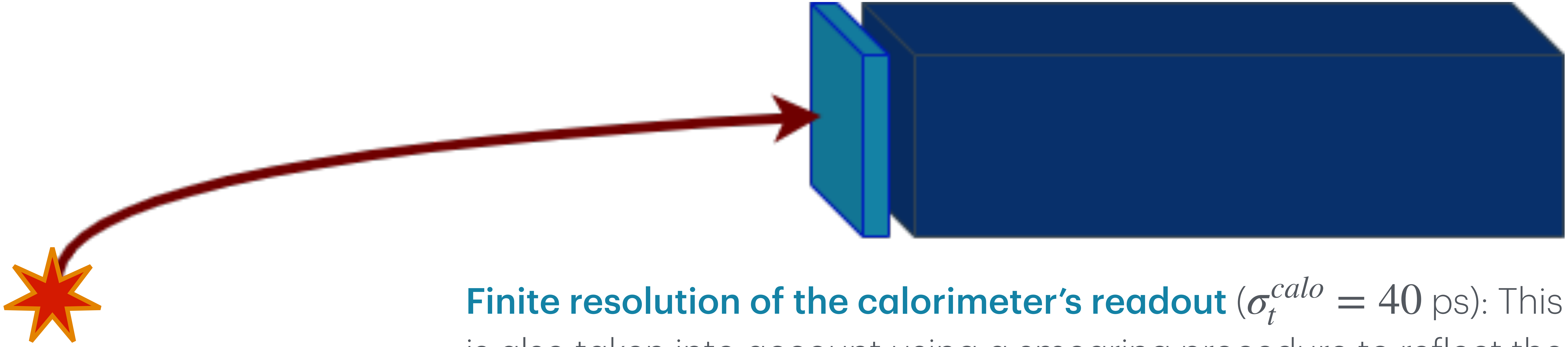


$$R = \frac{\sum r_i e_i}{\sum e_i} \qquad r_i = \sqrt{(x_i - x_c)^2 + (y_i + y_c)^2}$$

Timing Features

Towards a reliable experimental approach

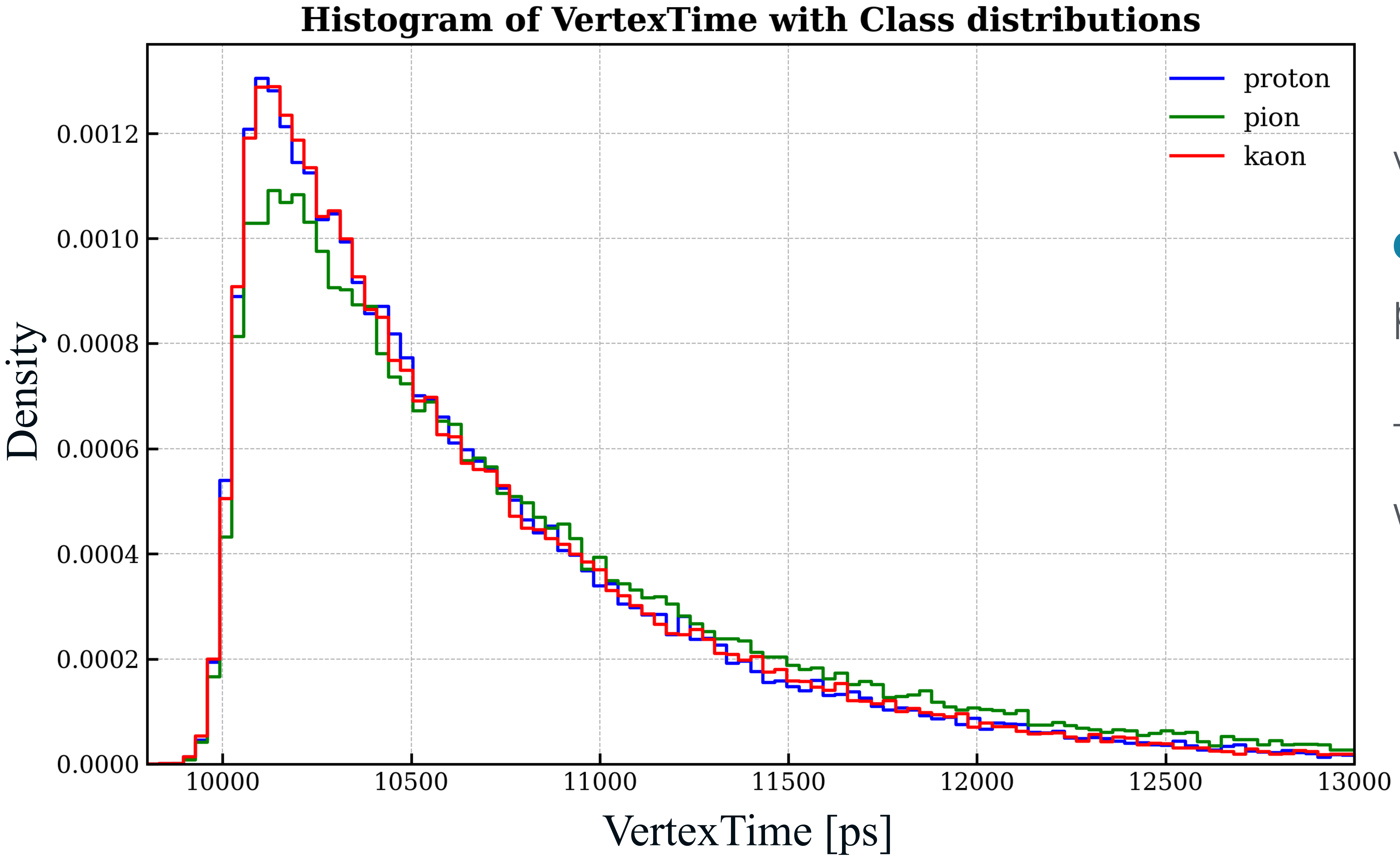
Propagation time of the particle through the 3 m long tracker before reaching the calorimeter: It is measured by a timing layer located just before the calorimeter and it has a finite resolution ($\sigma_t^{TL} = 40$ ps).



Finite resolution of the calorimeter's readout ($\sigma_t^{calo} = 40$ ps): This is also taken into account using a smearing procedure to reflect the limitations of the electronics.

Timing Features

VertexTime



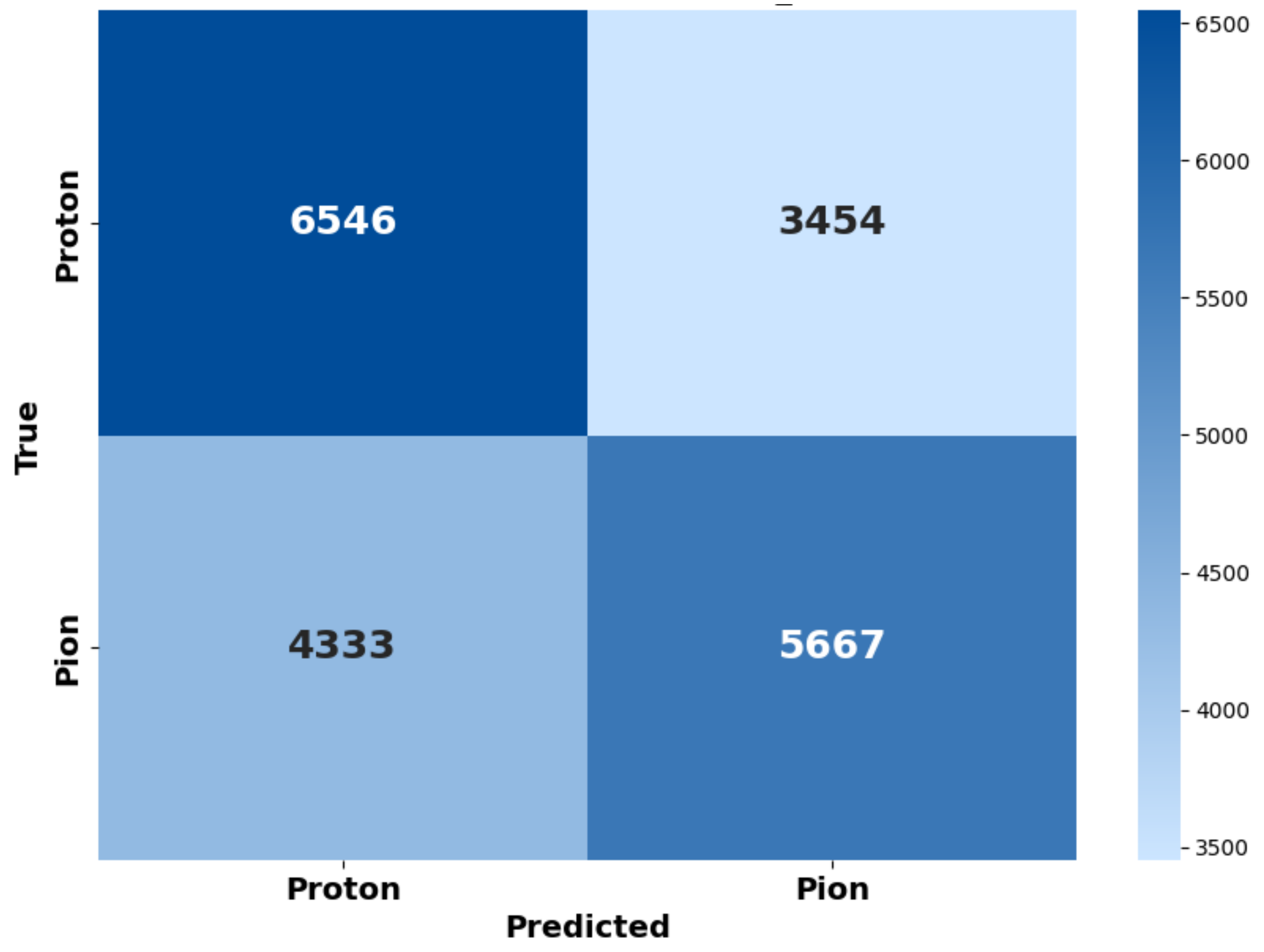
VertexTime is defined as the **characteristic time of the cell** which the primary vertex has been assigned to.

The pion distribution shows a difference with respect to protons and kaons.

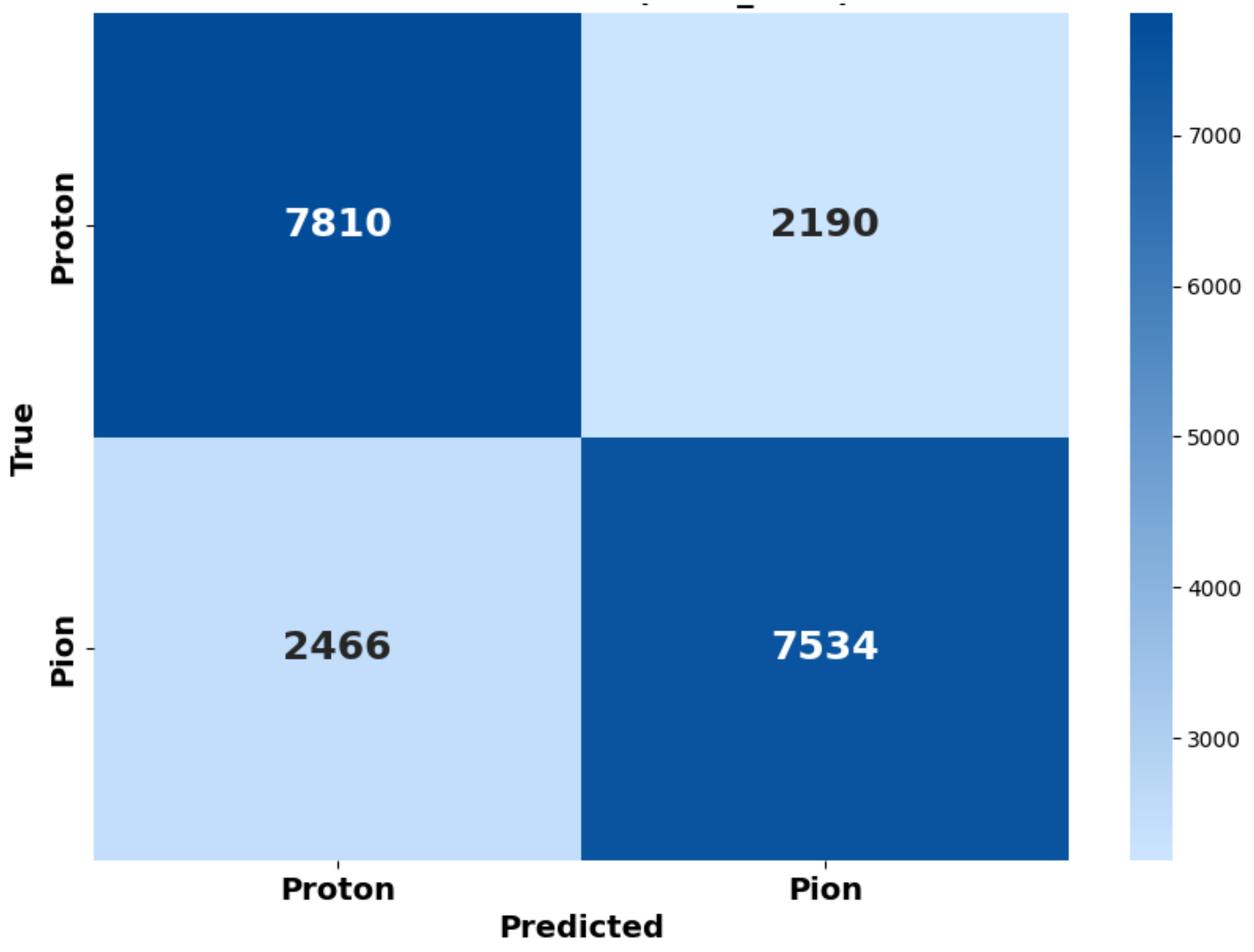
Results for proton-pion classification

XGBoost with 100x100x100

Without time-of-flight -> 61.1% accuracy



With time-of-flight -> 76.7% accuracy



What's next?

Studies on **how cell size affects particle identification performance** (already in progress)

Analysis of behaviour at **different energy levels**

Combination of DNNs (or BDTs) with CNNs in order to exploit the 3D shower pattern

Exploit **dual-readout technique** in order to compare different hadrons depending on their electromagnetic fraction

Study the power of identification in the case of **proton-kaons and pion-kaons**