Improving Muon-Pion & Electron-Pion Separation at Belle II with Machine Learning Using the Novel Pulse Shape Discrimination in CsI(Tl)

ALEXANDRE BEAUBIEN, PHD STUDENT

UNIVERSITY OF VICTORIA

THE BELLE II COLLABORATION

University of Victoria

Belle II

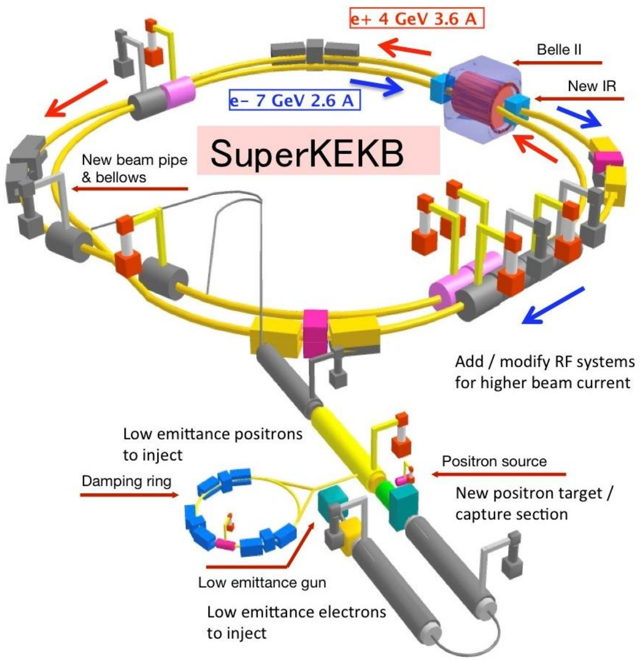
2022-06-08

Outline

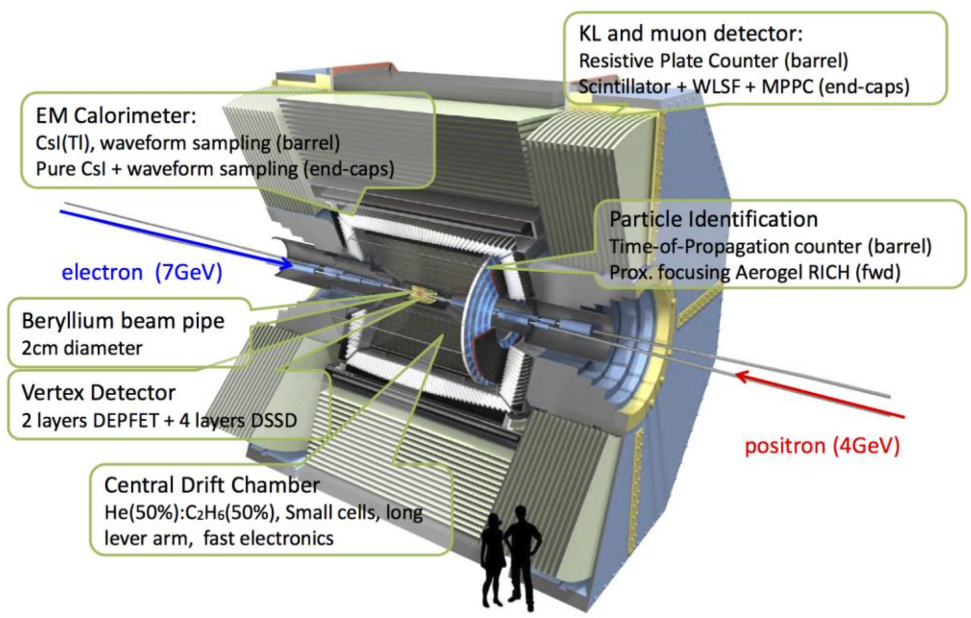
1. Intro to Belle II
2. Motivation for Pulse Shape Discrimination (PSD) based charged lepton-hadron Particle Identification (PID)
3. Intro to PSD
4. Results
5. Outlook and Conclusion

SuperKEKB – Belle II

~1150 collaborators
121 institutions



Belle II Detector



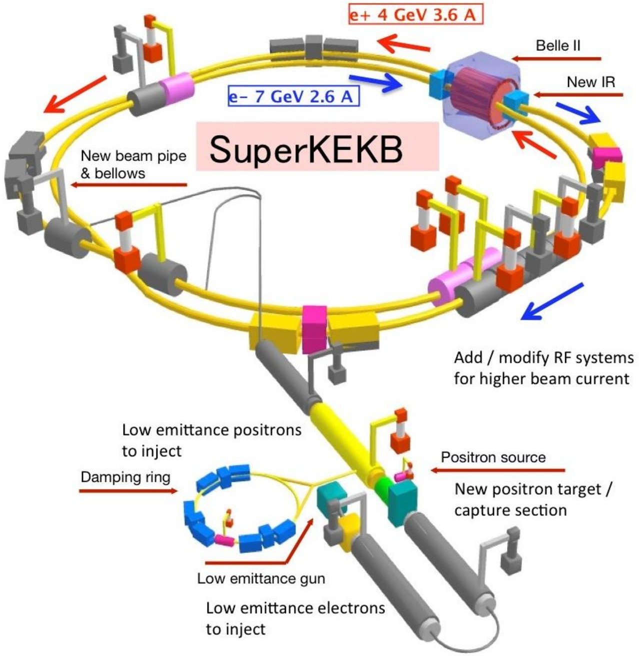
Belle II is a **multipurpose detector** operating at the **SuperKEKB** e^+e^- collider, ($\sqrt{s} = \mathbf{10.58}$ **GeV**). New generation of **B-Factories**, and is performing research at the **intensity frontier**.

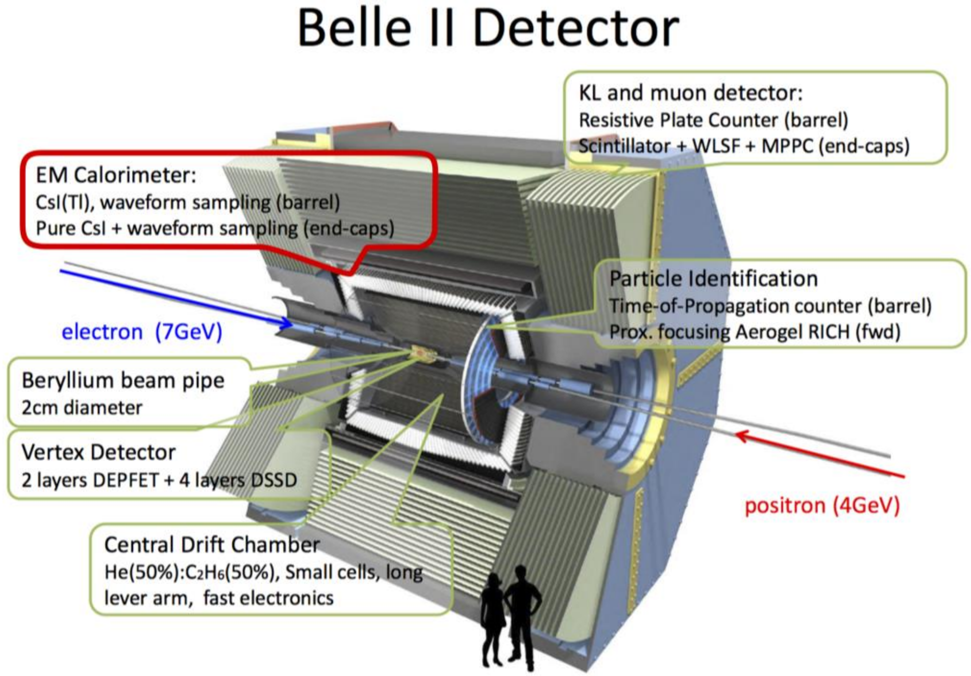
$$e^+e^- \to \Upsilon(4s) \to B\bar{B}$$

Target dataset size for Belle II is **50 ab^{-1}** which is around **30 times** that of the previous B-Factories, namely **BaBar & Belle**.

SuperKEKB – Belle II

~1150 collaborators
121 institutions





Belle II is a **multipurpose detector** operating at the **SuperKEKB** e^+e^- collider, ($\sqrt{s} = \mathbf{10.58}$ **GeV**). New generation of **B-Factories**, and is performing research at the **intensity frontier**.

$$e^+e^- \to \Upsilon(4s) \to B\bar{B}$$

Target dataset size for Belle II is **50 ab^{-1}** which is around **30 times** that of the previous B-Factories, namely **BaBar & Belle**.

PID Motivation

Problem:

Large systematic uncertainties stemming from **mis-identification** of **charged particles.**

An example: $\boldsymbol{\mu^{\pm}}, \boldsymbol{\pi^{\pm}}$ **mis-id.**

Similar particles -> regular PID is limited (e.g. Cherenkov).

Solution:

Use the novel **Pulse Shape Discrimination (PSD) tool.**

Discriminate between **hadrons** and **leptons** in the **Electromagnetic Calorimeter (ECL)** only.

Physics Motivation

Example physics measurements that would **benefit** from $\boldsymbol{\mu^\pm, \pi^\pm}$ **discrimination:**

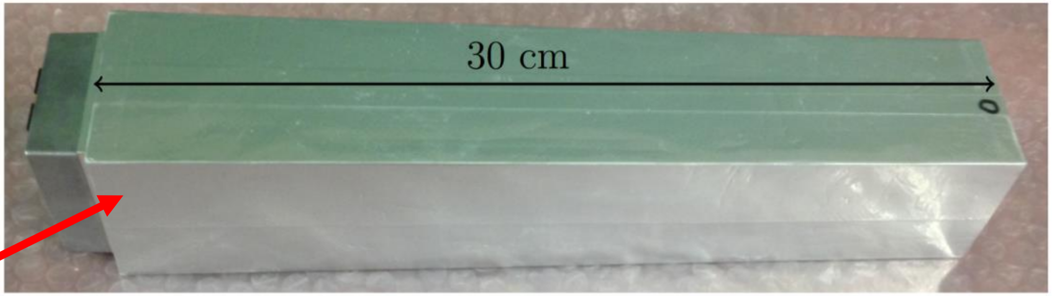
1. Measurement of A_{LR}: $\boxed{e^+e^- \rightarrow \tau^+\tau^- \rightarrow (\pi^+\overline{\nu_\tau})(a_1^-\nu_\tau).}$
   - Example mis-id: $\tau^+ \rightarrow \mu^+\nu_\mu\bar{\nu}_\tau$
2. Measurement of $\boxed{e^+e^- \rightarrow \pi^+\pi^-}$ cross section for vacuum polarization measurements.
   - Example mis-id: $e^+e^- \rightarrow \mu^+\mu^-$

Both have $\boldsymbol{\mu^\pm, \pi^\pm}$ **mis-identification**, where a pure $\pi^\pm$ sample is wanted.

Electromagnetic Calorimeter (ECL)

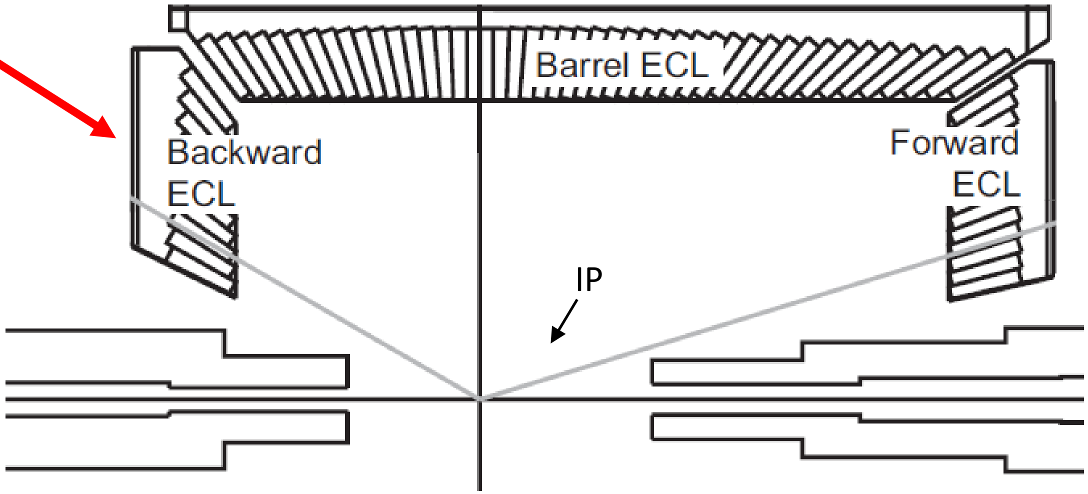
ECL is responsible for:

- **Measuring particle's energies** through energy depositions.



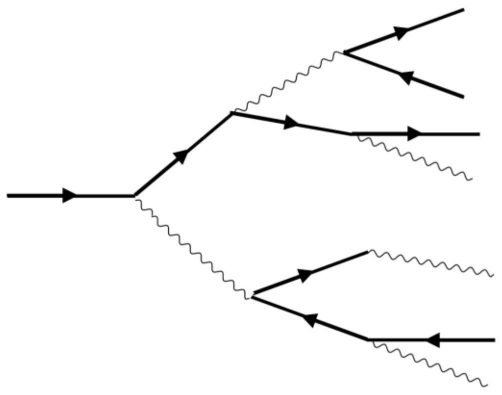
Composed of **8736 CsI(Tl) crystals** arranged in a **cylinder** around the IP.

Upper profile of the ECL

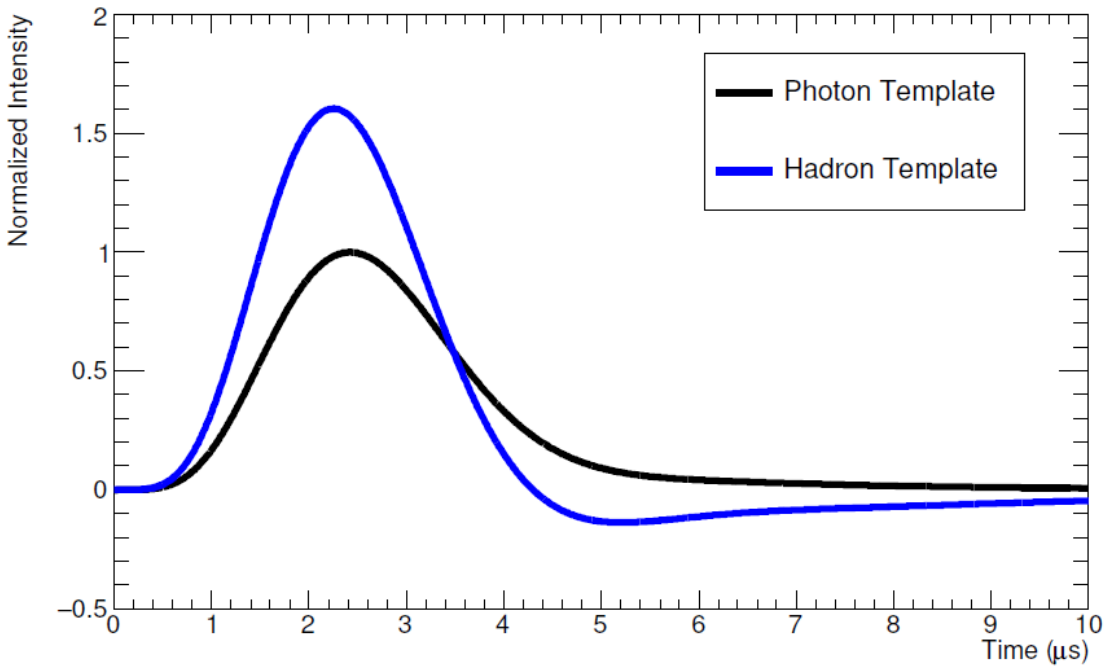


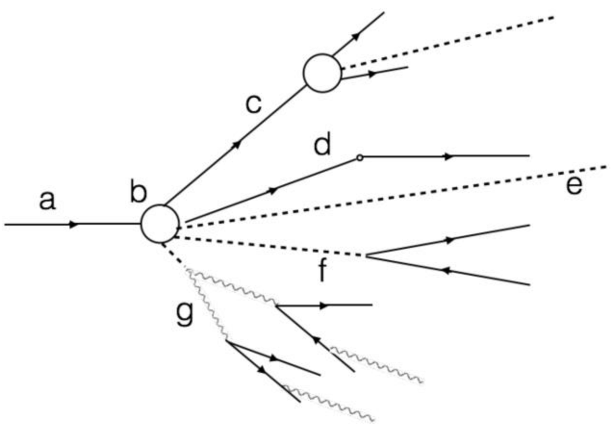
Crystal measurements are **digitized** in **31-length waveforms**. The waveforms are **fit** to obtain energies, and times.

Pulse Shape Discrimination (PSD)



Electromagnetic shower





Hadronic shower

Energy absorption in the ECL CsI(Tl) crystals has different **pulse shapes** based on the **type** of energy absorption.

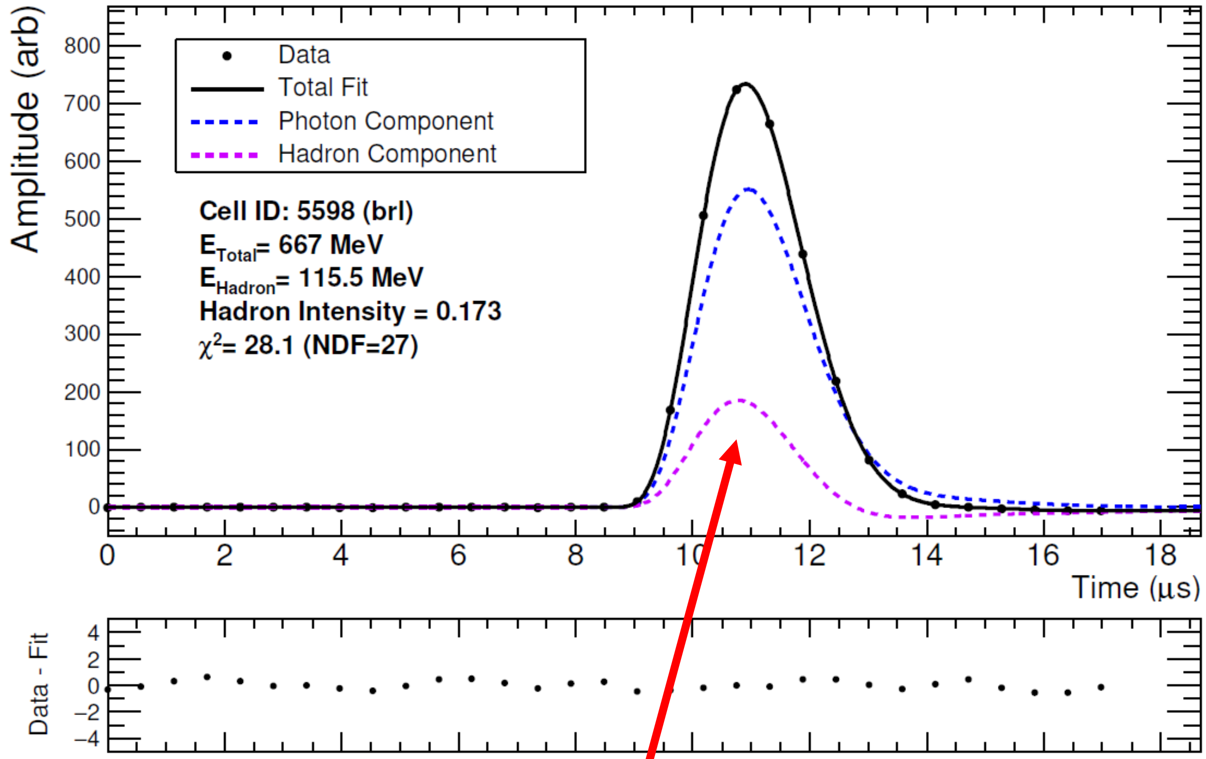
Mainly: **electromagnetic** vs **hadronic**.

PSD in CsI(Tl)

Specifically, we **fit** for the **contribution** of each **template.**

Get the **proportion** from **hadronic** and **electromagnetic sources.**

$\pi^{\pm}$ have ~**50%** probability to **hadronically shower.**



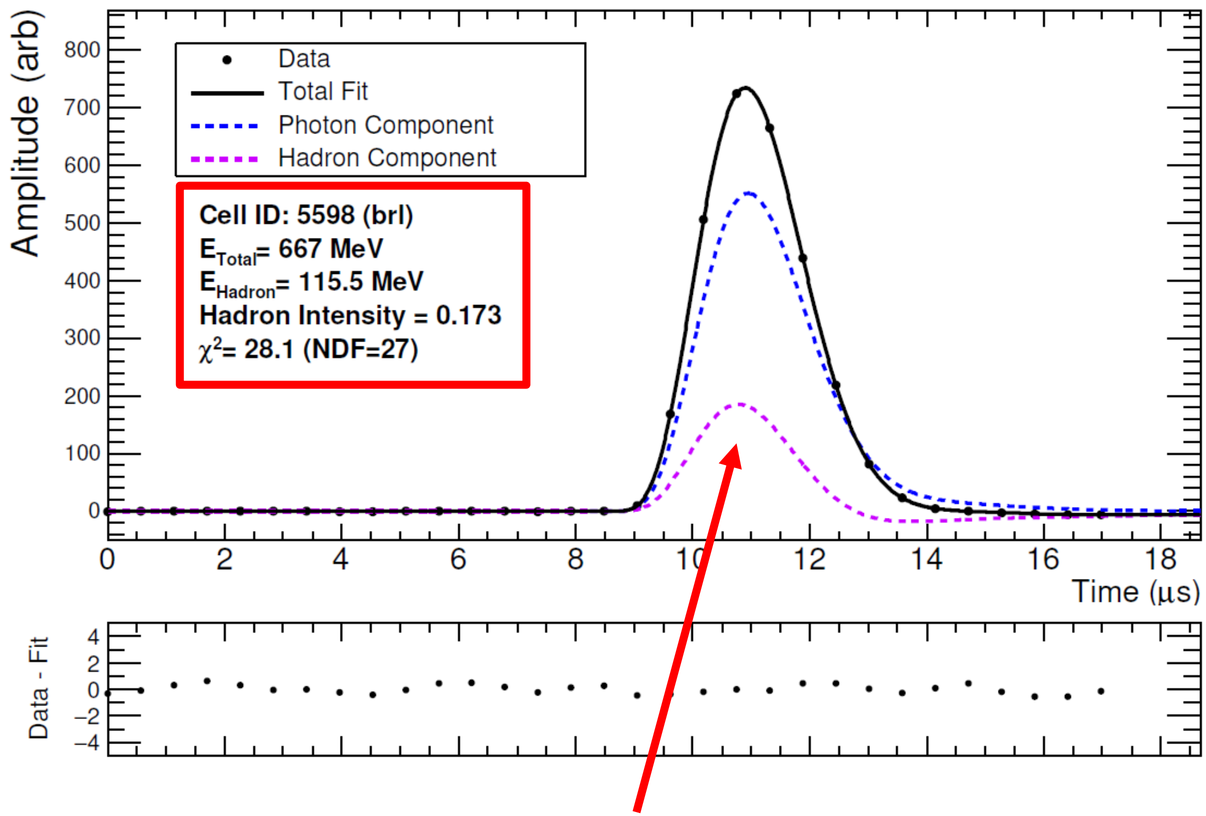
Energy contribution from each type of interaction.

PSD in CsI(Tl)

Specifically, we **fit** for the **contribution** of each **template.**

Get the **proportion** from **hadronic** and **electromagnetic sources.**

$\pi^{\pm}$ have ~**50%** probability to **hadronically shower.**



Energy contribution from each type of interaction.

Training a Boosted Decision Tree

GBDT uses the **crystal information** of **clusters** for **20 highest energy crystals:**

1. $F_{crystal}$: The fit type used on the crystal waveform to obtain the PSD information. This variable take the value 0 if one gamma and one hadron templates were used, 1 if two gamma and one hadron templates were used and 2 if diode crossing templates were used.
2. $E^{total}_{crystal}$: The total energy of the crystals.
3. $E^{hadron}_{crystal}$: The hadron energy contribution to the crystals.
4. $R_{crystal}$: The distance of crystals to the center of the cluster.
5. $W_{crystal}$: The weight of the crystal. This variable is obtained from the clustering algorithm and describes how strongly a particle relates to its associated cluster.

If there are not 20 crystals, fill with empty crystals.

Cluster Variables:

1. $E_{cluster}$: The total energy of the ECLCluster.
2. $E1E9$: The ratio of energies of the central crystal, E1, and 3x3 crystals, E9, around the central crystal.

Training a Boosted Decision Tree

GBDT uses the **crystal information** of **clusters** for **20 highest energy crystals:**

1. $F_{crystal}$: The fit type used on the crystal waveform to obtain the PSD information. This variable take the value 0 if one gamma and one hadron templates were used, 1 if two gamma and one hadron templates were used and 2 if diode crossing templates were used.
2. $E^{total}_{crystal}$: The total energy of the crystals.
3. $E^{hadron}_{crystal}$: The hadron energy contribution to the crystals.
4. $R_{crystal}$: The distance of crystals to the center of the cluster.
5. $W_{crystal}$: The weight of the crystal. This variable is obtained from the clustering algorithm and describes how strongly a particle relates to its associated cluster.

PSD contribution

If there are not 20 crystals, fill with empty crystals.

Cluster Variables:

Multi-crystal variables

1. $E_{cluster}$: The total energy of the ECLCluster.
2. $E1E9$: The ratio of energies of the central crystal, E1, and 3x3 crystals, E9, around the central crystal.

Evaluation Metrics

Metric used to evaluate the models is the **ROC curve**.

- **Threshold independent** metric

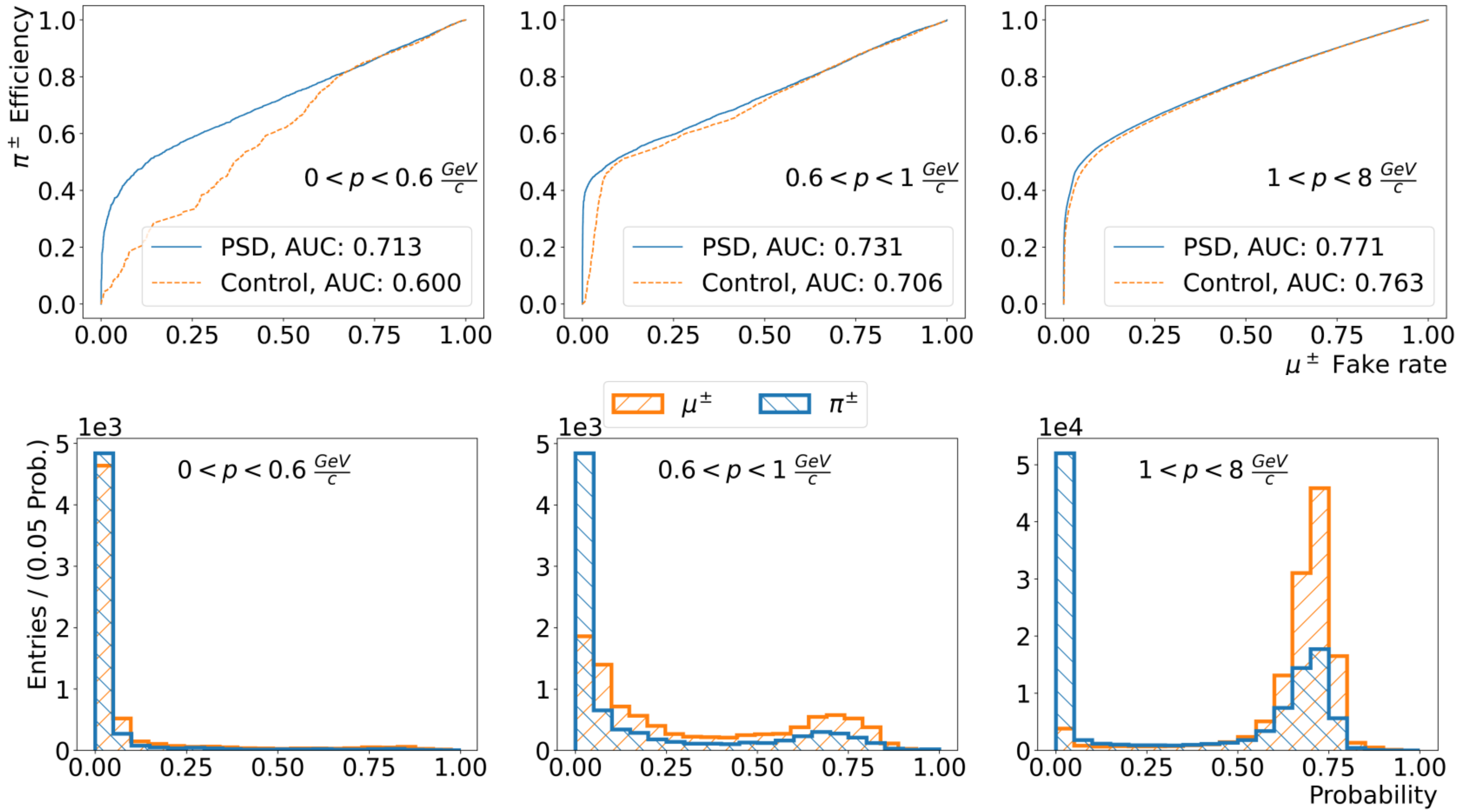
$$\pi^{\pm}\ \text{Efficiency} = \frac{\#\ \text{of real pions predicted as pions}}{\text{total}\ \#\ \text{of pions}}$$

$$l^{\pm}\ \text{Fake Rate} = \frac{\#\ \text{of real leptons predicted as pions}}{\text{total}\ \#\ \text{of leptons}}$$

The **area under the ROC curve (AUC)** is obtained to quantify the model quality.

Results PID – $\mu^{\pm}, \pi^{\pm}$

Target:
0: $\pi^{\pm}$ 1: $\mu^{\pm}$

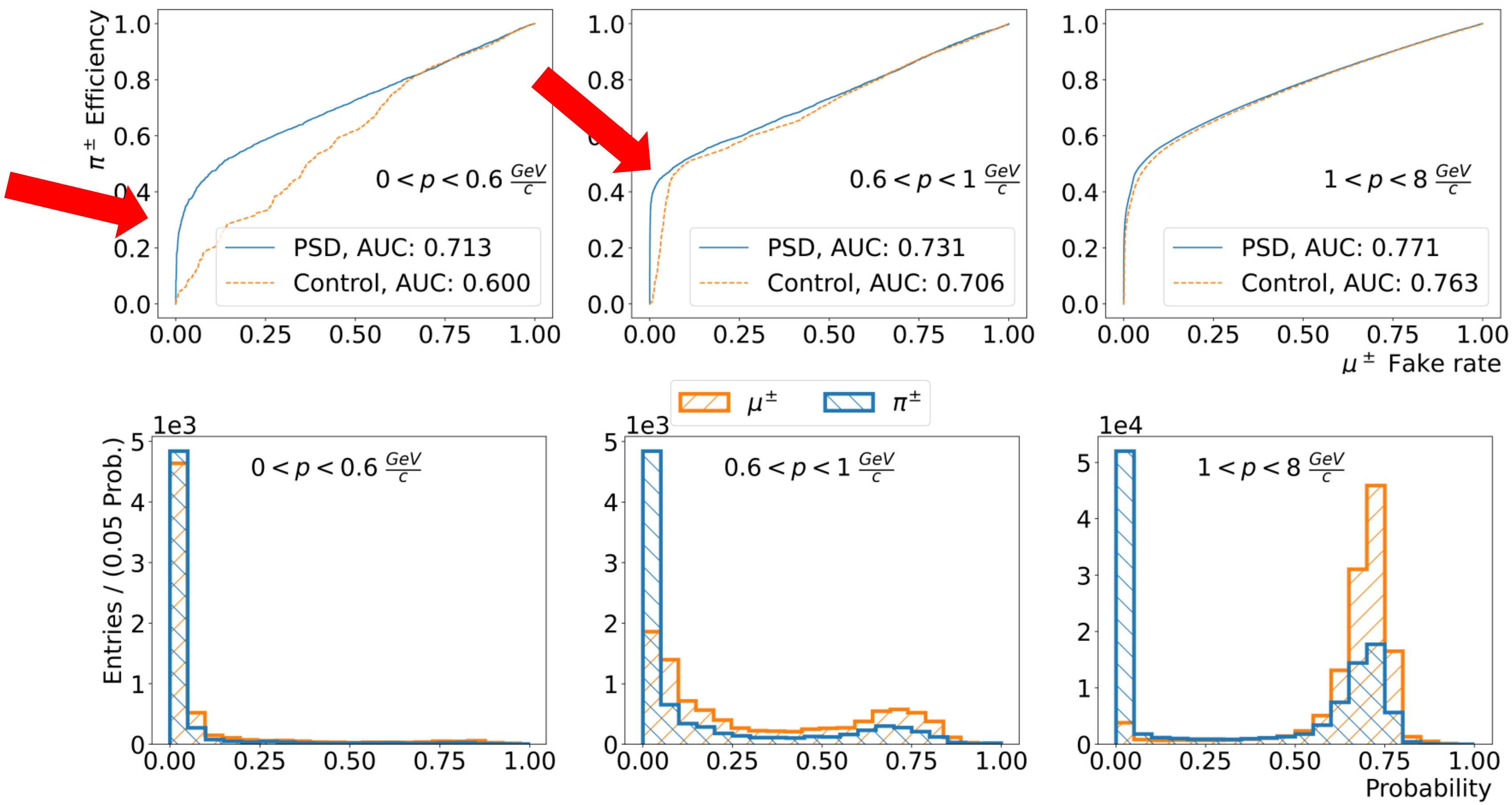


The metrics are shown against a **control model** training which contains the same variables **minus** the **PSD variables**. An **improvement** to the ROC curve and its AUC are observed when **adding PSD** information.

Results PID – $\mu^{\pm}, \pi^{\pm}$

Target:

0: $\pi^{\pm}$ 1: $\mu^{\pm}$

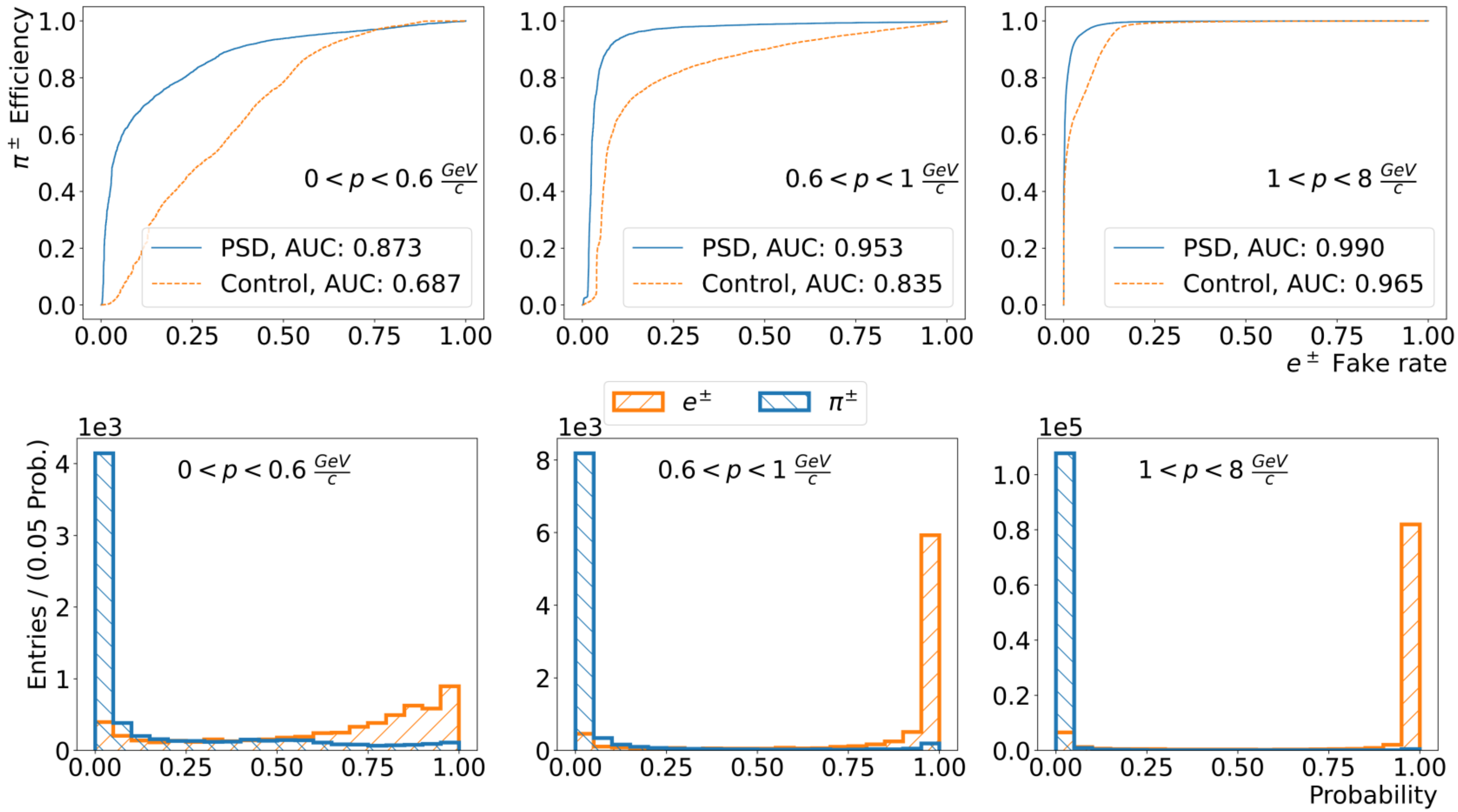


The metrics are shown against a **control model** training which contains the same variables **minus** the **PSD variables**. An **improvement** to the ROC curve and its AUC are observed when **adding PSD** information.

Results PID – $e^{\pm}, \pi^{\pm}$

Target:

0: $\pi^{\pm}$ 1: $e^{\pm}$

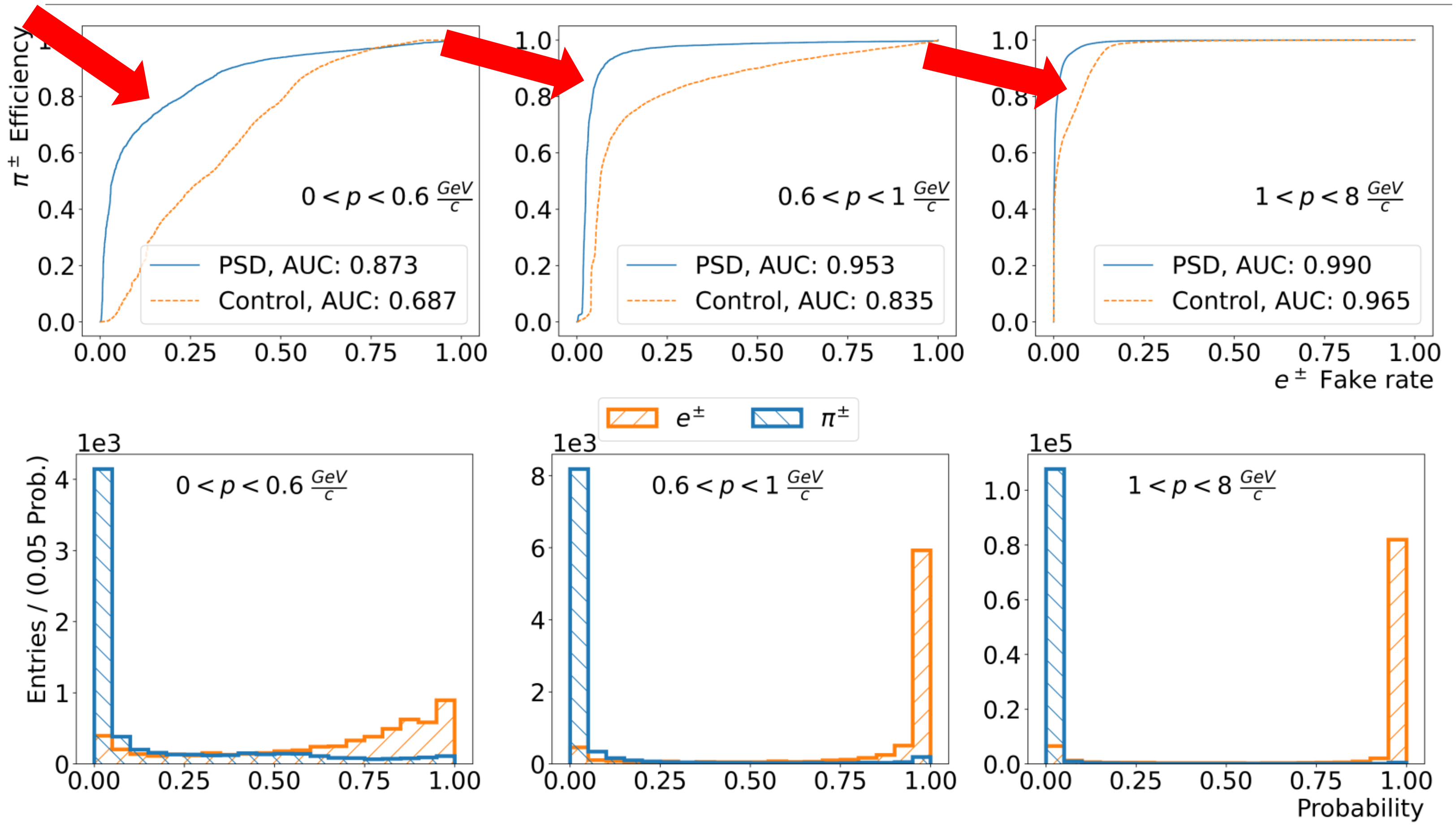


The metrics are shown against a **control model** training which contains the same variables **minus** the **PSD variables**. An **improvement** to the ROC curve and its AUC are observed when **adding PSD** information.

Target:

0: $\pi^{\pm}$ 1: $e^{\pm}$

Results PID – $e^{\pm}, \pi^{\pm}$



The metrics are shown against a **control model** training which contains the same variables **minus** the **PSD variables**. An **improvement** to the ROC curve and its AUC are observed when **adding PSD** information.

Summary

This work is part of a larger effort to improve the Belle II particle identification system, which **improves our physics results.**

PSD is used to **discriminate charged particles** in the **ECL** using some elementary **ML methods**.

Shows **great discrimination power** especially for **low momentum particles,** for both $\mu^{\pm}$ and $e^{\pm}$ vs $\pi^{\pm}$.

PSD and **PSD based MVA** can and should be considered in all existing and future experiments as an integral part of detectors PID systems.

Thank you!

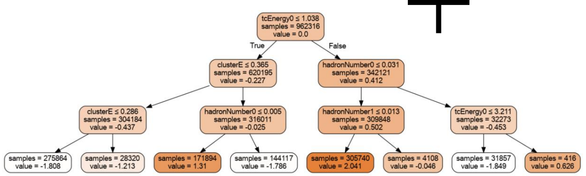
Backup

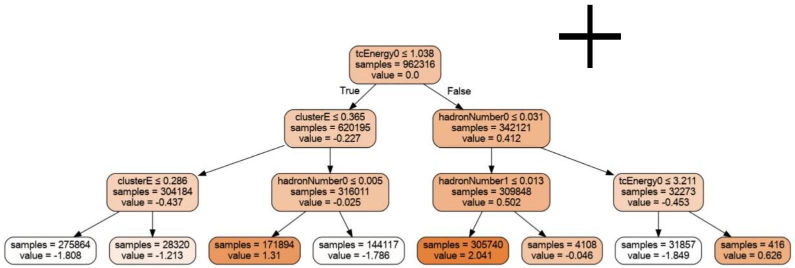
Building a Boosted Decision Tree

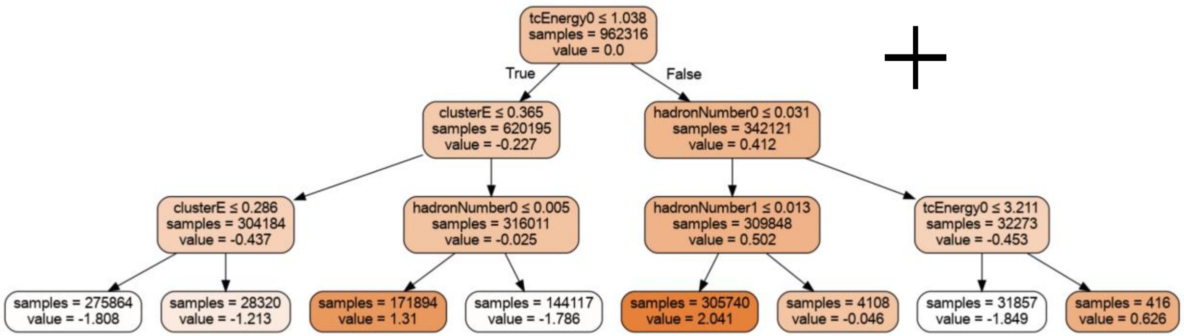
Gradient Boosted Decision Trees (GBDT) use a series of decision trees.

New trees « fix » the previous trees' mistakes.

...



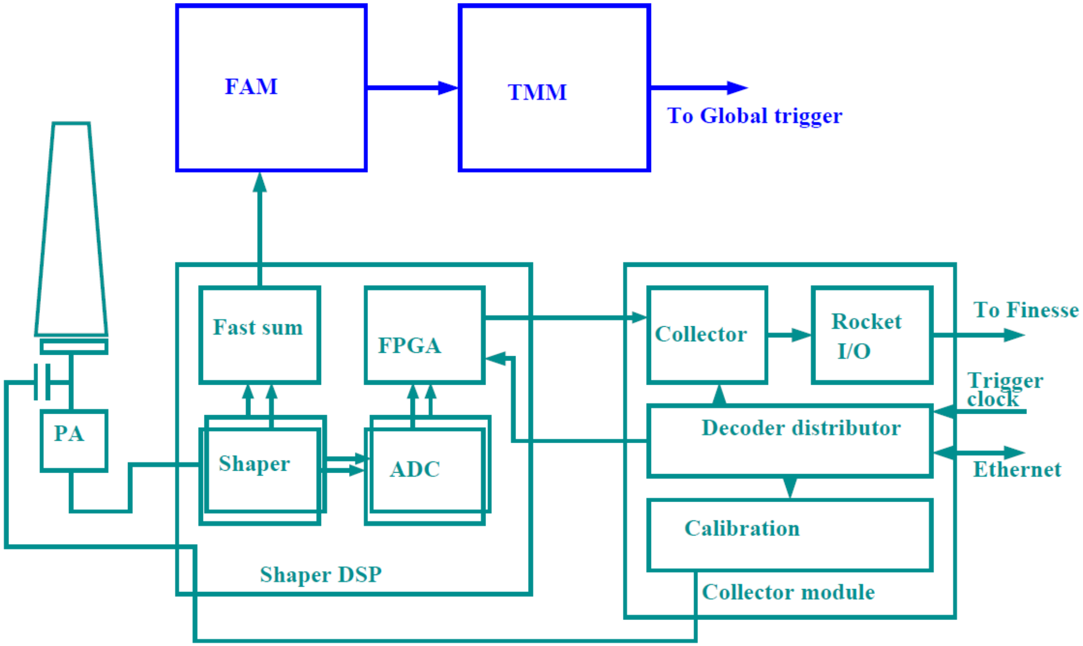




Resilient to **overtraining** and the **curse of dimensionality**!

Electronics & Triggers

Each ECL crystal is connected to a ShaperDSP circuit to perform the waveform shaping and digitization. A fast shaping is done to obtain trigger information.



A collector module sends the digitized waveforms to the data acquisition (DAQ) framework.

Training Data

The data set is **simulated** with **GEANT4**. **1 000 000** particles are simulated and reconstructed for $\mu^{\pm}, \pi^{\pm}$ and $e^{\pm}$. Model is applied on **ECLClusters** (multi-crystal particle interaction).

Cut	$\pi^{\pm}$	$\mu^{\pm}$	$e^{\pm}$
Number of clusters reconstructed	1063743	1010974	1009004
Dropped due to requirements 1 or 2	309934	239257	211601
Dropped due to requirement 3	108642	36878	229826
Dropped due to all crystals failing 4 or 5	9458	19610	390
Number of clusters added to the training set	635709	715229	567187

ECLClusters are **used** in **training** if:

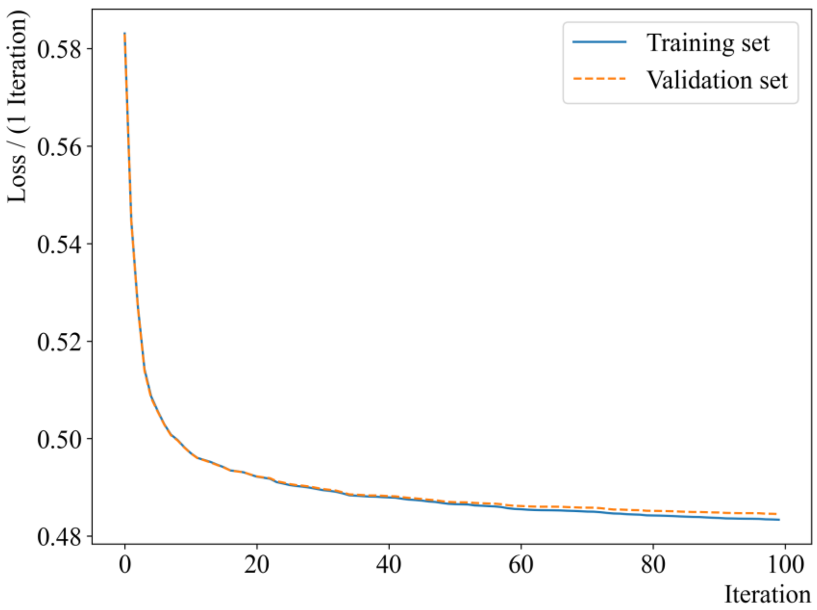
1. ECLCluster is in the barrel.
2. ECLCluster is clearly matched to a simulated particle.
3. Has at least one crystal containing a waveform that can be used for PSD.

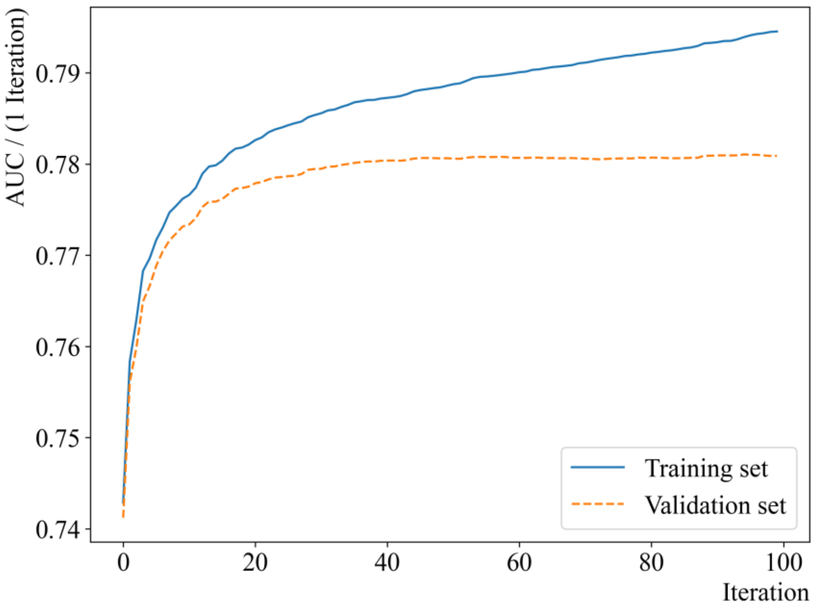
Crystals are **kept (up to 20)** if:

4. PSD fit has $\chi^2 < 60$.
5. PSD fit does not use the diode crossing template (showers is in the diode).

Training Monitoring

Train the model and record for overtraining by looking at the validation and training loss and AUC after adding every tree.





The metrics plateau (no overtraining).