PID Techniques and Performance at LHCb in Run 2

14th Pisa Meeting on Advanced Detectors

Mikhail Hushchyn[1,2] **and Victoria Chekalina**[1,2]
on behalf of the LHCb collaboration

[1]National Research University Higher School of Economics
[2]Yandex School of Data Analysis

Introduction

Particle identification (PID) plays a crucial role in LHCb analyses. The LHCb PID system is composed of two ring-imaging Cherenkov detectors, a series of muon chambers and a calorimeter system. Combining information from these subdetectors allows one to distinguish between various species of long-lived charged and neutral particles. Advanced multivariate techniques are employed to obtain the best PID performance and control systematic uncertainties in a data-driven way. A novel strategy has been introduced in Run 2, where the selection of PID calibration data is implemented in the LHCb software trigger, with further processing achieved through a centralised production that makes highly efficient use of computing resources. This poster covers the major steps of the implementation, and highlights the PID performance achieved in Run 2.

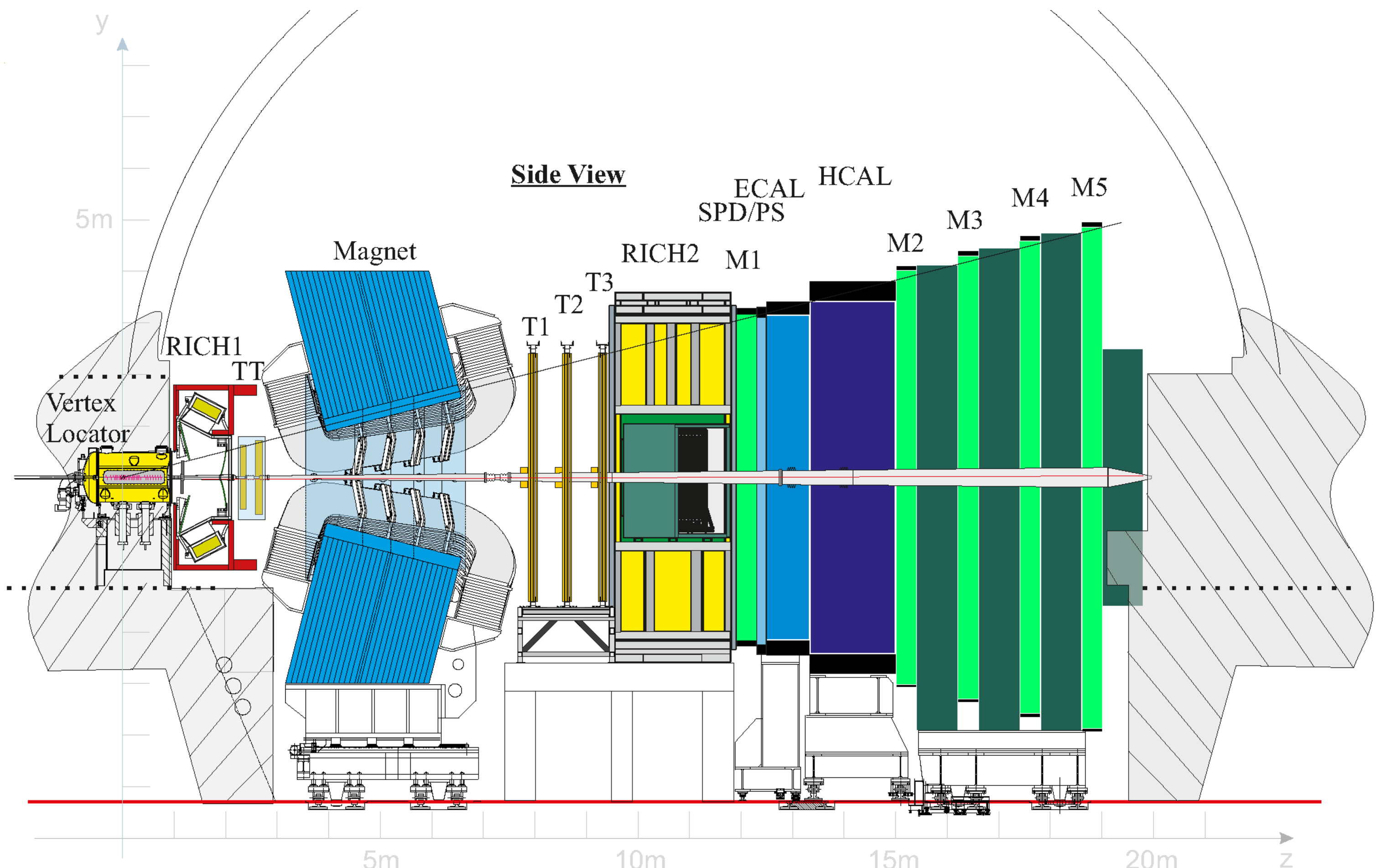

Figure 1: LHCb detector layout. The interaction point is on the left, inside the VELO detector.

Global Charged PID

Particle identification (PID) is a crucial part in high-energy physics analysis. The global charged PID consists in identifying the charged particle type associated with a given track. There are five particle types: electron, muon, pion, kaon, proton, and ghost track (charged tracks that do not correspond to a real particle which passed through the detector). This is a multiclassification problem in machine learning. Global PID uses the information from all LHCb systems: tracking system, RICHs, the electromagnetic and hadronic calorimeters and muon chambers are used as inputs for the following classifiers:

- **ProbNN** [1] (baseline): One hidden layer neural network of TMVA library.
- **Deep NN**: Deep neural network of Keras library.
- **CatBoost** [2] : Gradient boosting over oblivious decision trees classifier.
- **Flat 4d**: Boosted decision trees classifier with flat dependency of signal efficiency from particle P, P_T, η and $nTracks$ (event multiplicity) observables.

The classifiers were trained on a MC sample containing all of the different charged particle types. Calibration samples, containing particles that can be identified purely based on only kinematic properties, were used to estimate the classifier performance on real data. The samples contain decays that allow to identify particle types based on only kinematic properties. The PID performance of each classifier is shown in Fig 2.

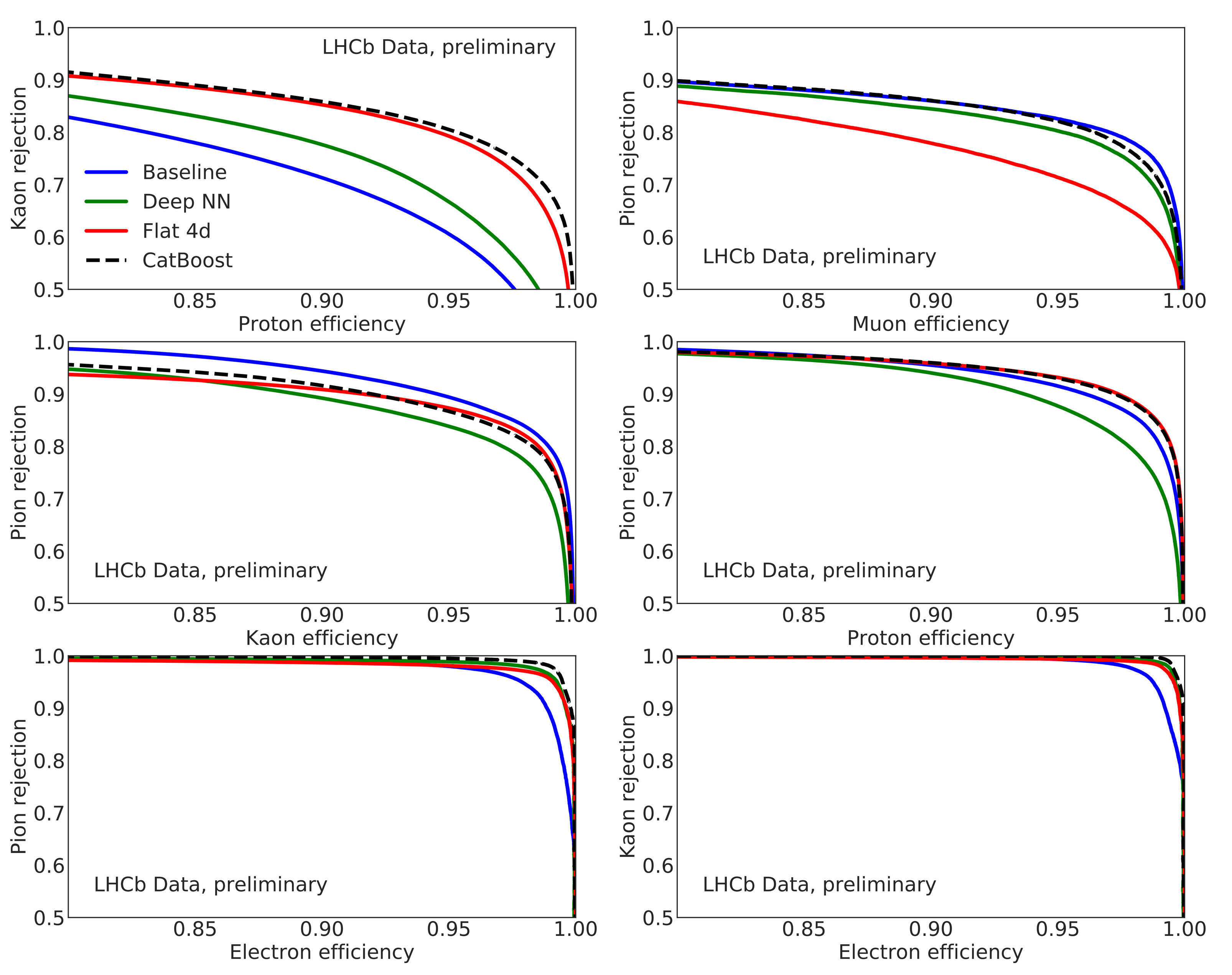

Figure 2: Dependences between background rejection and signal efficiency for six particle pairs.

The PID information strongly depends on the kinematic variables. This relationship leads to strong dependency between PID efficiency and kinematic variables as shown in Fig 3. Relative to the baseline model, the **Flat 4d** model has a flatter PID efficiency as a function of the kinematic variables. The classifier achieves this flatness using a modified loss function [3].

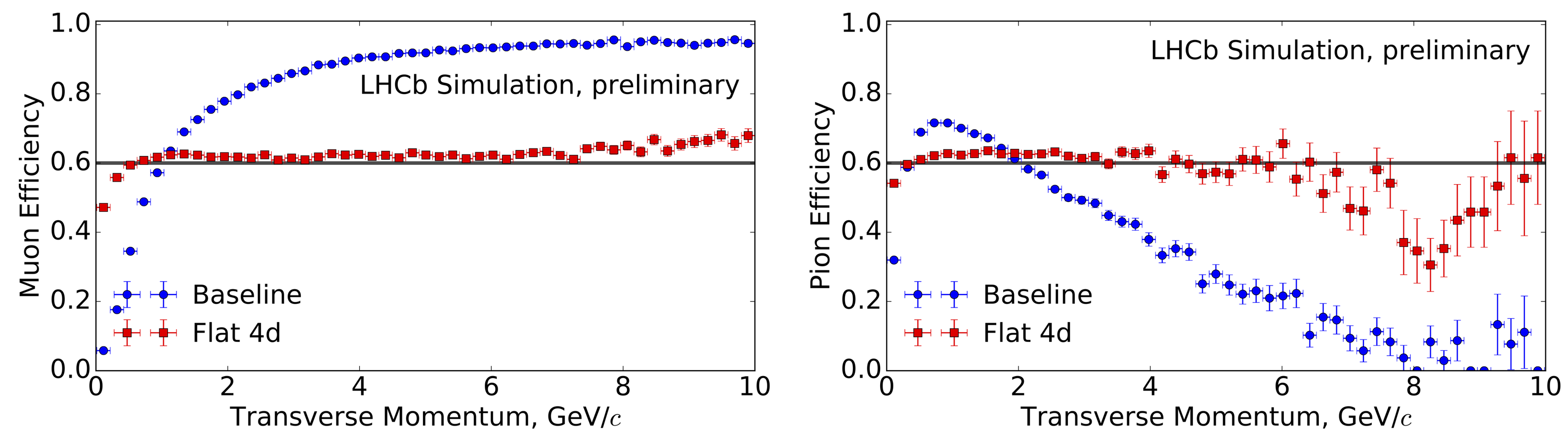

Figure 3: Dependence between Flat 4d model efficiencies and particle transverse momentum for each particle type. The curves correspond to the same global signal efficiency of 60%.

Neutral PID

Identification of neutral particles is based on responses in the calorimeter system. The particular problem is to separate prompt photons produced in collisions and photons produced by π^0 which decay into two photons. The most confusing case is when two photons from the π^0 decay are highly collinear, which look similar to single clusters produced by prompt photons. ECAL energy deposits for a prompt photon (left) and photons from a π^0 decay (right) are shown in Fig 4.

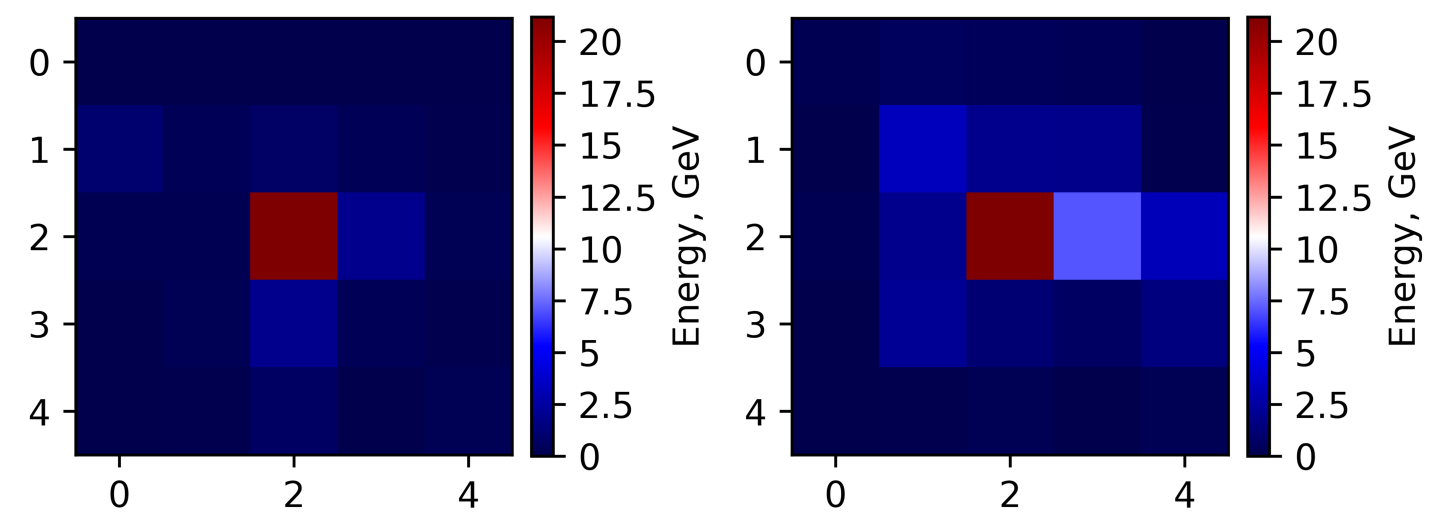

Figure 4: Energies in 5×5 ECAL cells around the cell seed for single photon (left) and for π^0 (right).

Two machine learning based models are used to separate these cases:

- **Baseline** [4]: A 2-layer neural network of TMVA library. The NN inputs are features that describe a cluster shape in 3×3 ECAL cells area around the cluster seed. Example of the features: center of gravity of a cluster energy, second moments of the cluster shape S_{xx}, S_{yy}, S_{xy} and their functions.
- **New method**: XGBoost classifier which is a Gradient Boosting over Decision Trees classifier. Inputs are raw energy values in 5×5 ECAL and PS cells around the cell seed as shown in Fig 4. There are no any additional input features.

The models were trained using MC samples. Photons and π^0 are generated from $B^0 \to K\pi\gamma$ and $B^0 \to K\pi\pi^0$ decays respectively. Two additional decays $B^0 \to J/\psi K^*$ and $K^* \to K\pi^0$ are used as an additional source of π^0 to check the models stability. The performance of γ-π^0 separation is demonstrated in Fig 5.

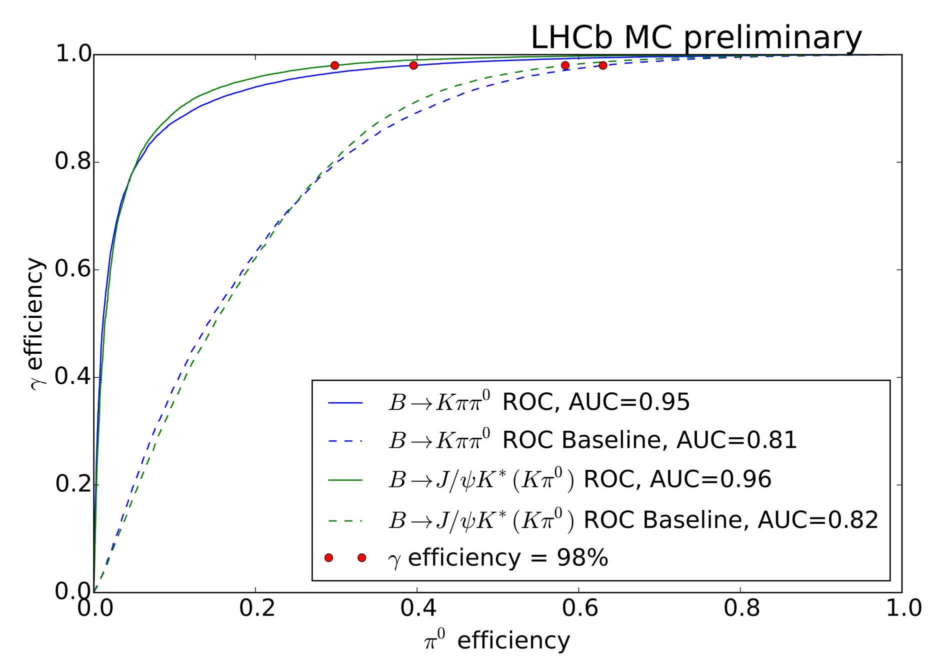

Figure 5: Classification ROC curves for the baseline (dashed line) and new method (solid line). Different colors refer to different test samples.

Conclusions

Combining information from the LHCb tracking system, ring-imaging Cherenkov detectors, electromagnetic and hadron calorimeters, and muon chambers using advanced machine learning techniques allows to achieve high background rejection in high signal efficiency regions for charged particle identification. Energy deposits in the calorimeter system provide good prospects for background suppression for neutral particle identification.

References

[1] LHCb collaboration. Lhcb detector performance. *Int. J. Mod. Phys.*, A(30):1530022, 2015.

[2] A. Veronika Dorogush, A. Gulin, G. Gusev, N. Kazeev, L. Ostroumova Prokhorenkova, and A. Vorobev. Fighting biases with dynamic boosting. *ArXiv e-prints*, June 2017.

[3] Alex Rogozhnikov, Aleksandar Bukva, Vladimir Gligorov, Andrey Ustyuzhanin, and Mike Williams. New approaches for boosting to uniformity. *Journal of Instrumentation*, 10(03):T03002, 2015.

[4] O. Deschamps M. Hoballah M. Calvo, E. Cogneras. A tool for γ π^0 separation at high energies. *LHCb-PUB-2015-016*, 2015.