Machine Learning at LHCb

Mikhail Hushchyn

National Research University Higher School of Economics

Moscow Institute of Physics and Technology

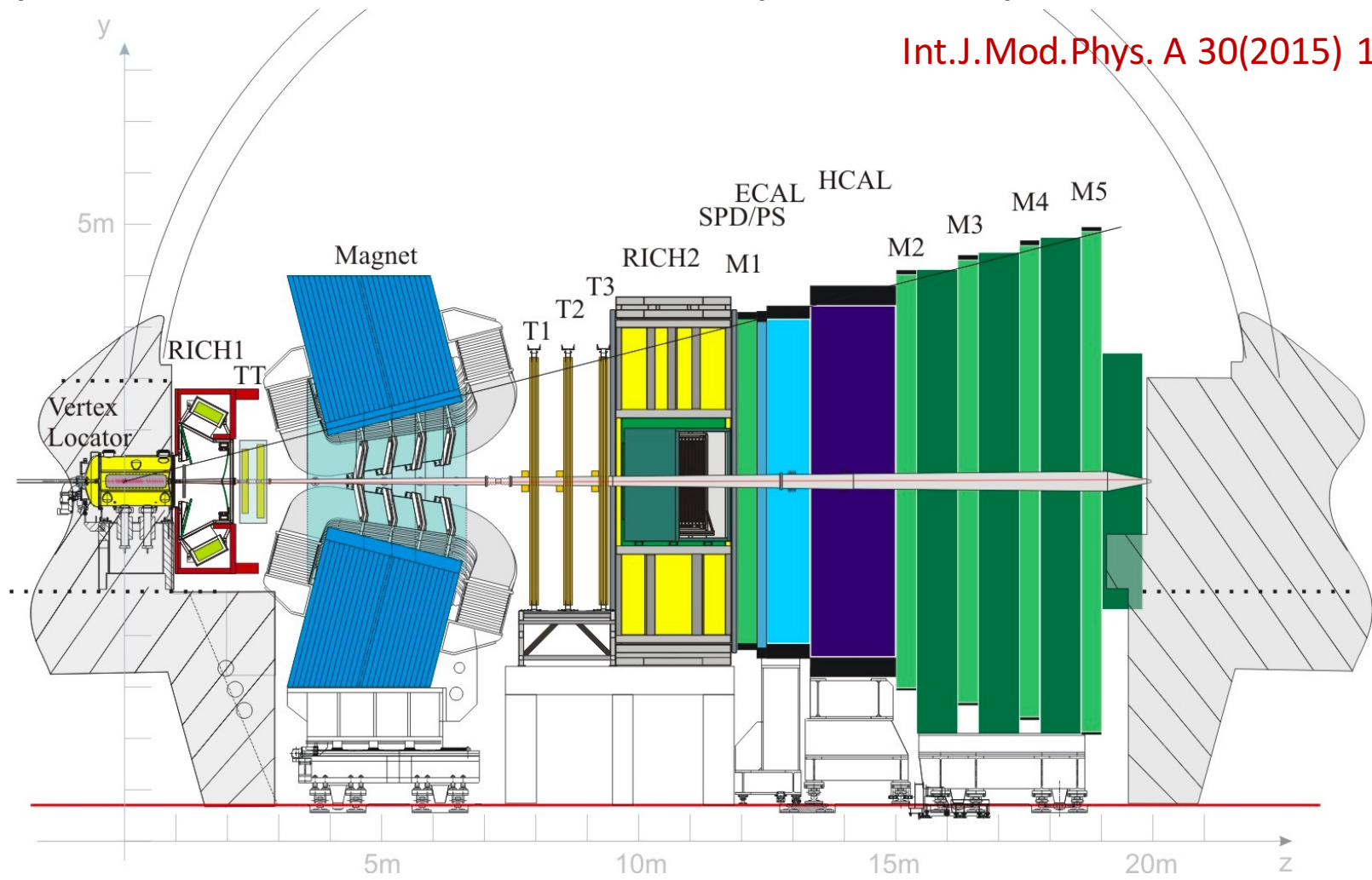Yandex School of Data Analysis

on behalf of the LHCb collaboration

ICPPA 2017, 3-6 October, Moscow

The LHCb Detector

- LHCb is a single-arm forward spectrometer.
- The main goal of the detector is to search for indirect evidence of new physics in CP violation and rare decays of beauty and charm hadrons.
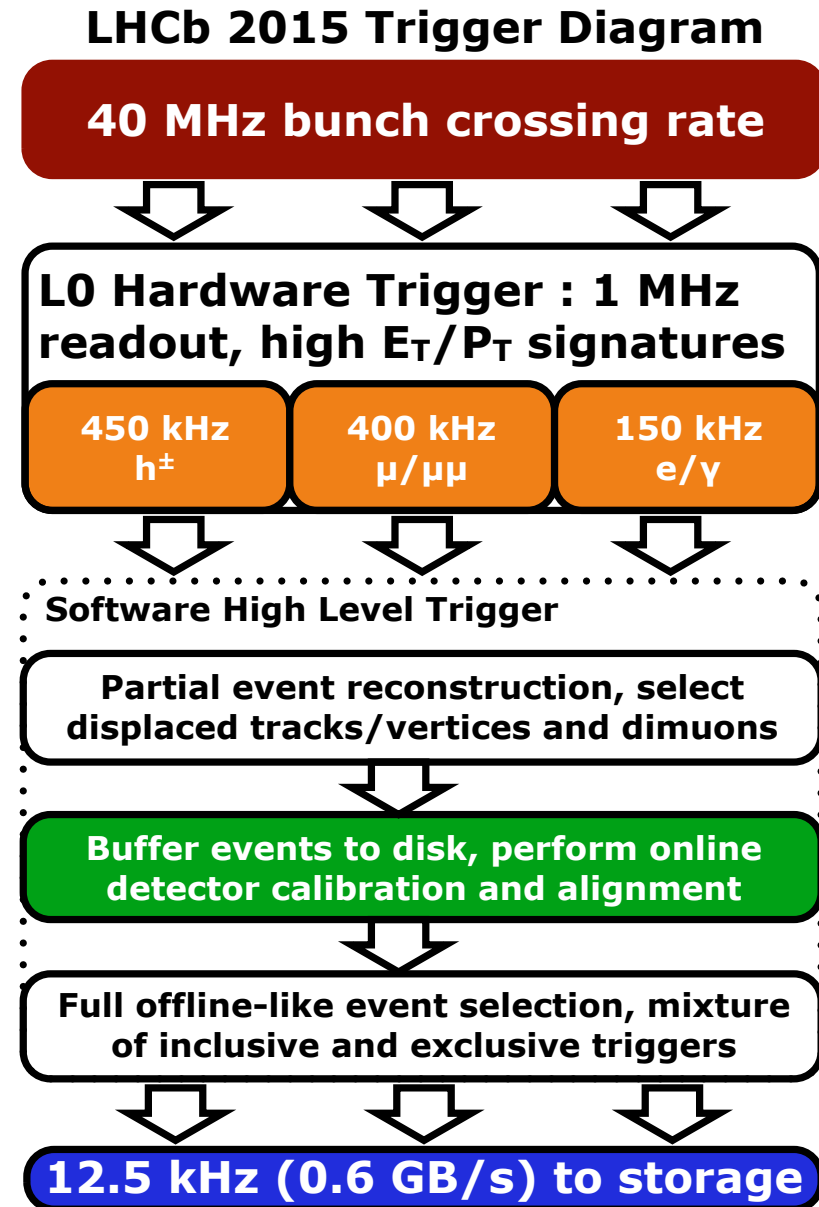
Why do we use ML?

The LHCb detector generates too much data to keep it all. Machine learning is needed for efficient selection of the most interesting events in Software High Level Trigger.

ML is also used in:

- Track pattern recognition
- Fake tracks rejection
- Particle identification
- Jets identification
- DQ monitoring
- Optimize use of storage capacity

LHCb 2015 Trigger Diagram

40 MHz bunch crossing rate

L0 Hardware Trigger : 1 MHz readout, high E_T/P_T signatures

| 450 kHz $h^{\pm}$ | 400 kHz $\mu/\mu\mu$ | 150 kHz e/γ |

Software High Level Trigger

Partial event reconstruction, select displaced tracks/vertices and dimuons

Buffer events to disk, perform online detector calibration and alignment

Full offline-like event selection, mixture of inclusive and exclusive triggers
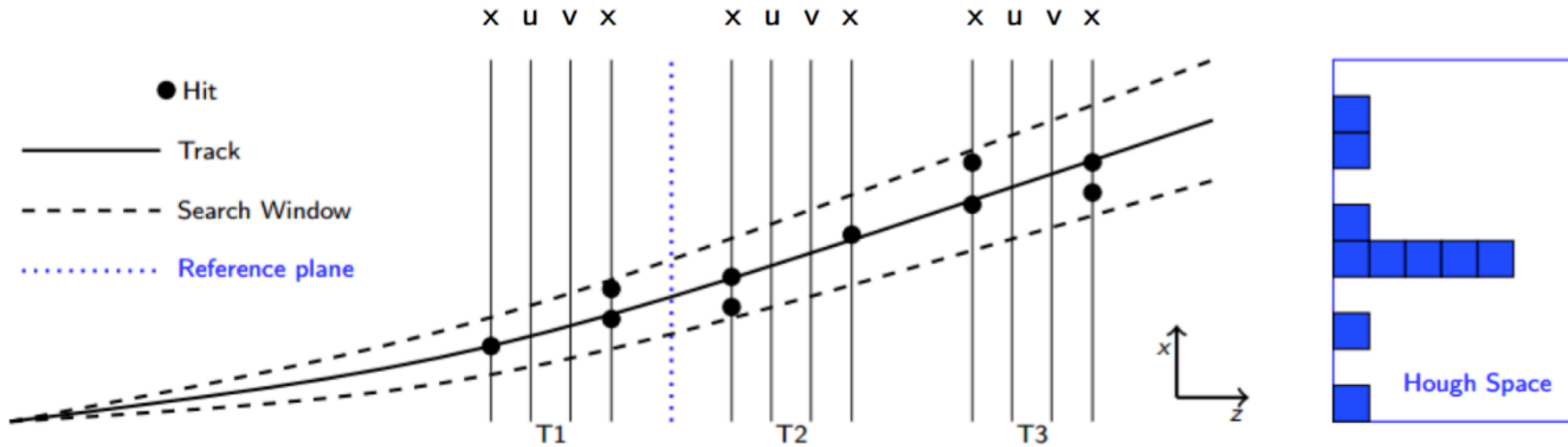
12.5 kHz (0.6 GB/s) to storage

Track Pattern Recognition

- **VELO tracking:** vertex reconstruction
- **Long tracks:** used in majority of analyses (B/D decays)
- **Downstream tracks:** daughters of long lived particles
- Average tracking efficiency > 96%
- Momentum resolution varies from 0.5% at low momentum to 1.0% at 200 GeV

Int.J.Mod.Phys. A 30(2015) 1530022
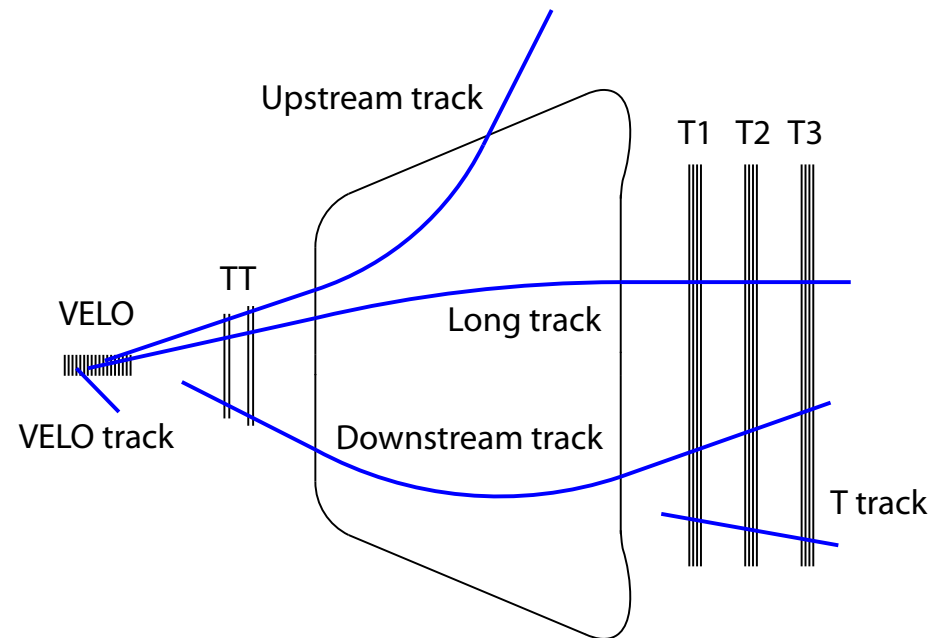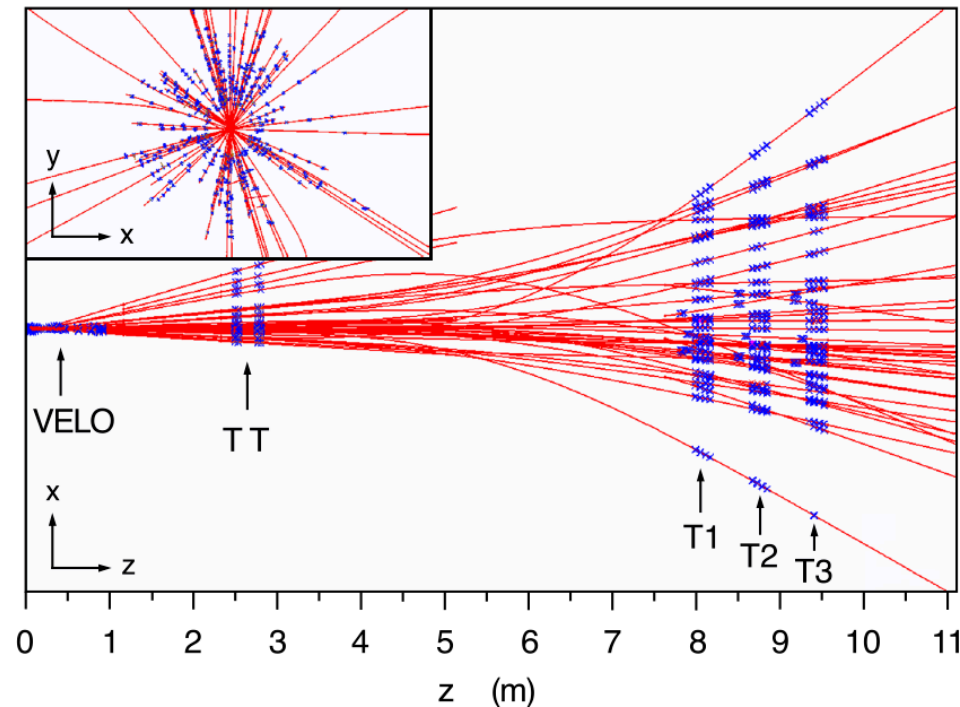
Long Tracks Reconstruction

Starting from seeds in the VELO, tracks are searched in T stations:
- Search window in T stations defined by VELO track.
- Project x-hit into reference plane.
- Fit **4-layer-x-cluster** and remove outliers.
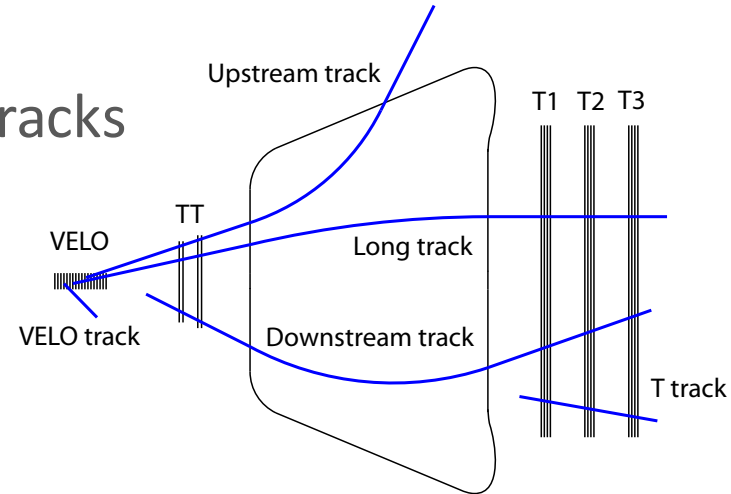- Add and fit track with stereo hits.

LHCb-PROC-2017-013

Two Deep Neural Networks:
- First of them is tuned for **rejection of bad 4-layer-x-clusters.**
- Second one is trained for **candidates selection** after stereo fit.

5

Downstream Tracks Reconstruction
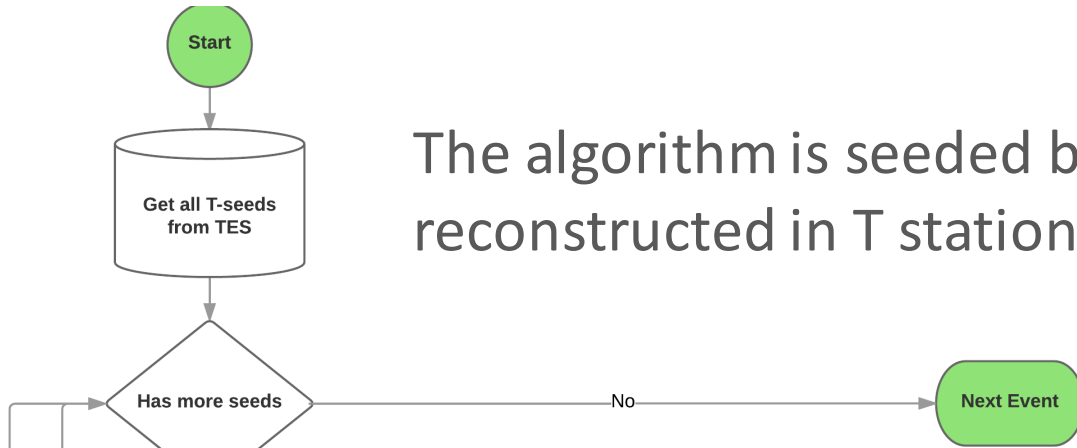
The algorithm is seeded by tracks reconstructed in T stations.

Rejection of about 40% of fake T-Seeds using Bosai BDT [*JINST* **8** P02013]

Find matching TT hits

Results: 3 - 5% improvement in fake track rejection and increase in signal efficiency

Fake Track Rejection

NNs are trained for background rejection at given (97 to 99 %) efficiency. Fake track (ghost) probability based on the DNN output allows to reduce fake rate. Results:

- Increased efficiency
- Reduced fake rate (22% → 14%)
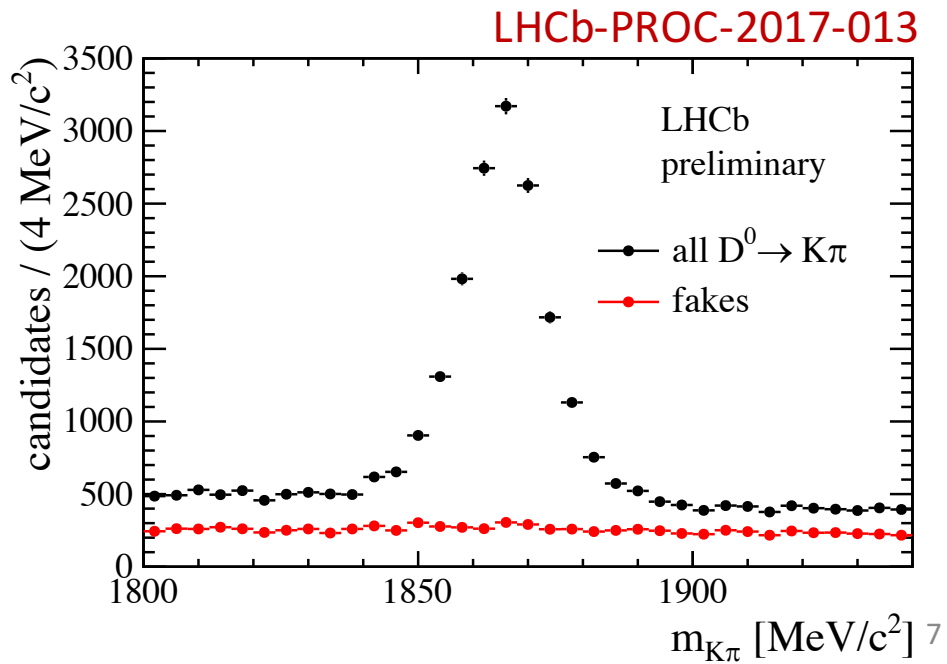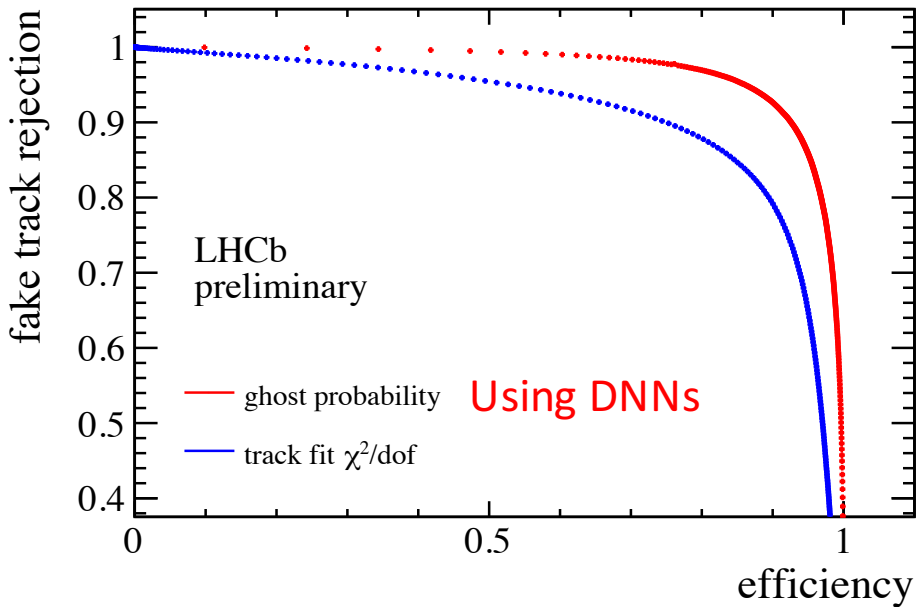
Particle Identification

- **Problem:** identify particle type associated with a track.
- **Particle types:** Ghost, Electron, Muon, Pion, Kaon, Proton.
- **LHCb subdetectors:** RICH, ECAL, HCAL, Muon Chambers and Track observables
- Different particle types has different responses in the subdetectors.
- The problem can be considered as multiclass classification problem in machine learning.
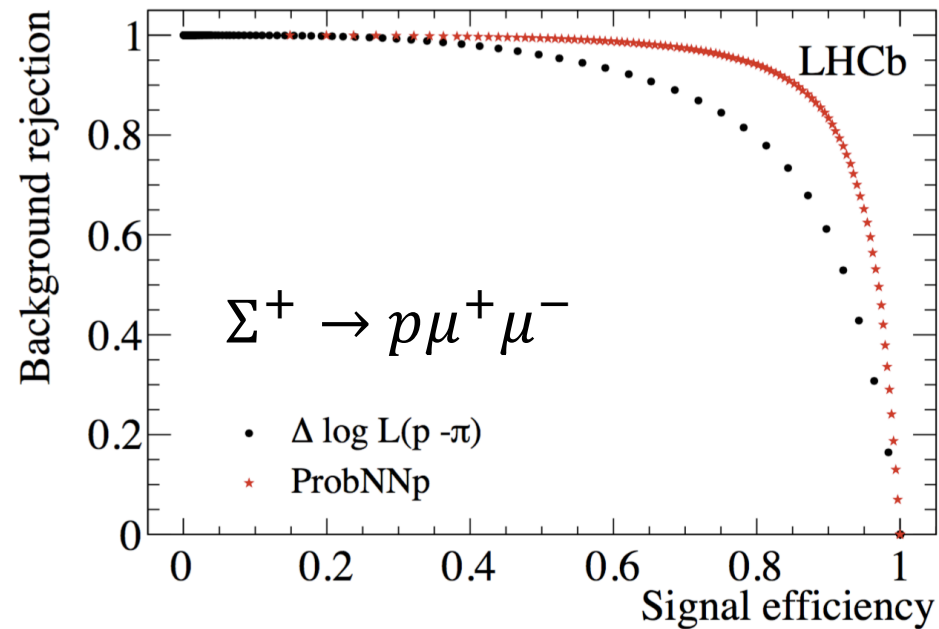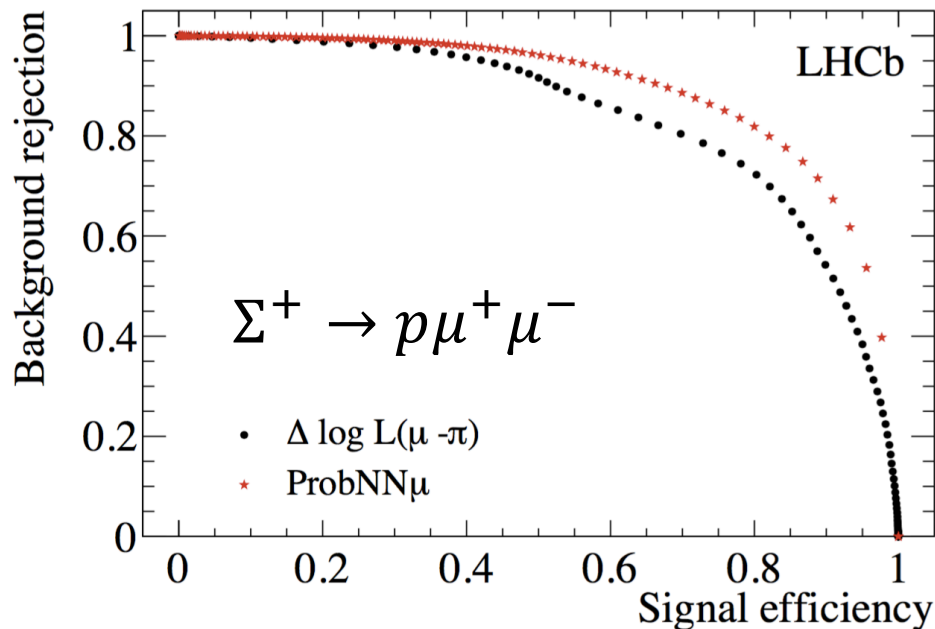

Int.J.Mod.Phys. A 30(2015) 1530022

8
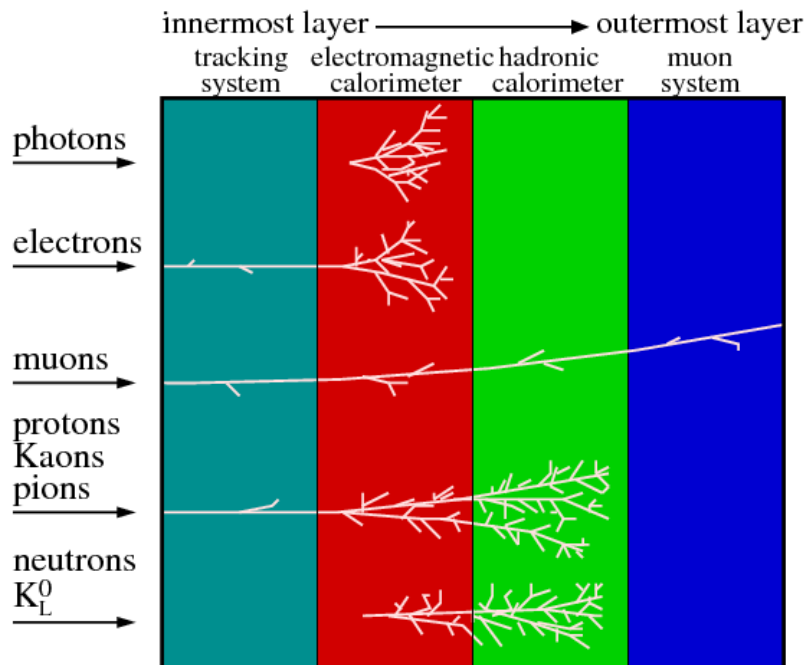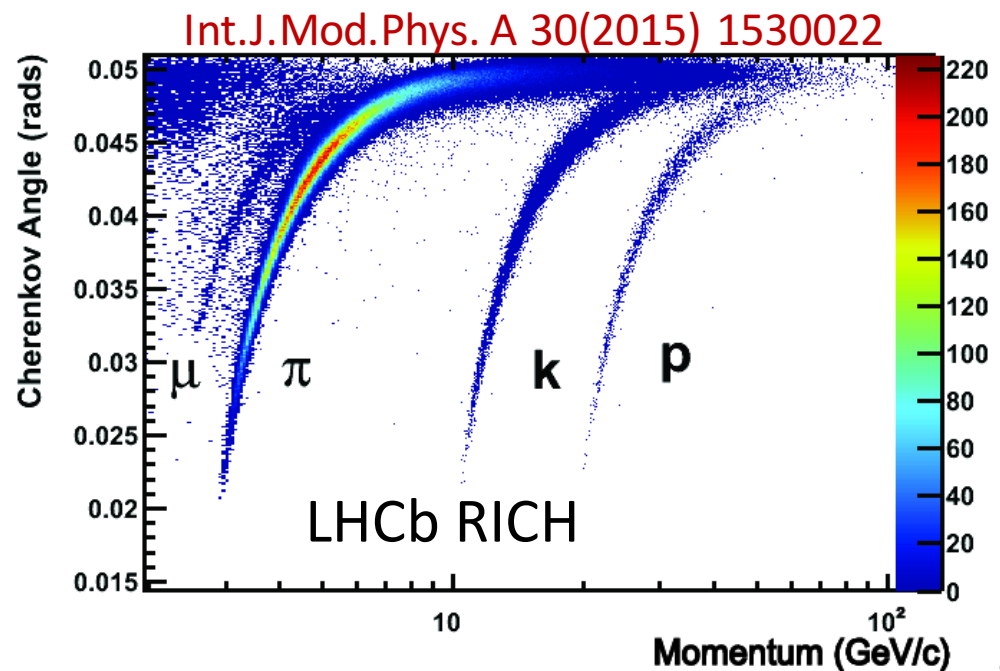
Particle Identification

- The first machine learning algorithms used for the PID in LHCb is one-hidden-layer neural network (TMVA MLP).
- Each particle type has its own binary NN trained in one-particle-vs-rest mode.

Plots: using data sidebands for backgrounds and Monte Carlo simulation for the signal

Particle Identification

Further PID performance improvement is done using different multiclass models: deep neural networks (**DNN**) and BDTs (**XGBoost** and **CatBoost**):

- One model for all particle types.
- ROC AUCs $\approx 0.91 - 0.99$ for different particle types.

6xProbNNs trained in one-vs-rest mode are considered as **baseline**.
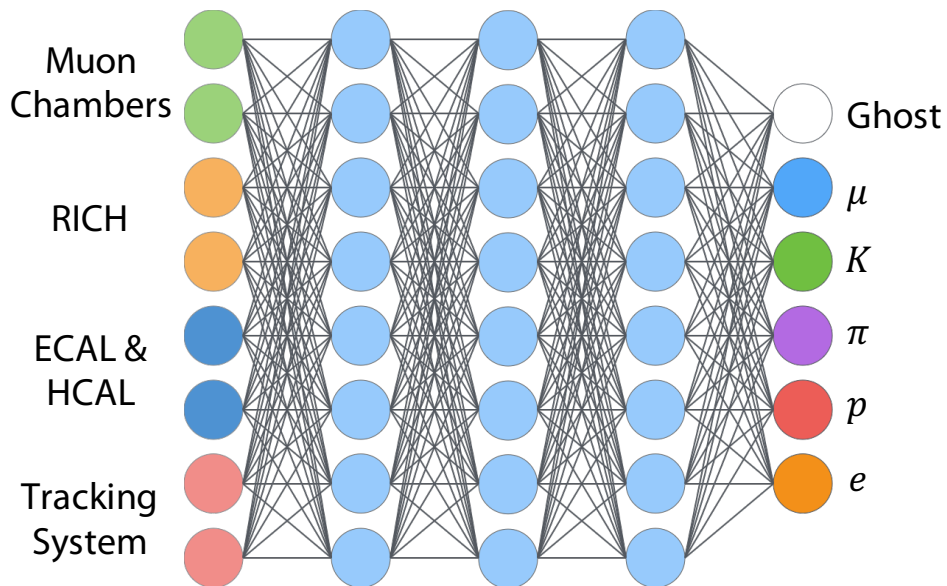
LHCb Simulation, preliminary

$(1-AUC)/(1-AUC_{baseline})$

	Ghost	Electron	Muon	Pion	Kaon	Proton
baseline	1	1	1	1	1	1
deep NN	-29 %	-41 %	-52 %	-37 %	-20 %	-17 %
XGBoost	-24 %	-37 %	-50 %	-34 %	-18 %	-15 %
CatBoost	-30 %	-43 %	-54 %	-37 %	-20 %	-18 %

Particle Identification

Several DBT models with flat efficiencies along P, Pt, η and $nTracks$ are provided. The models are trained with special loss function described in [JINST 10 (2015) T03002].
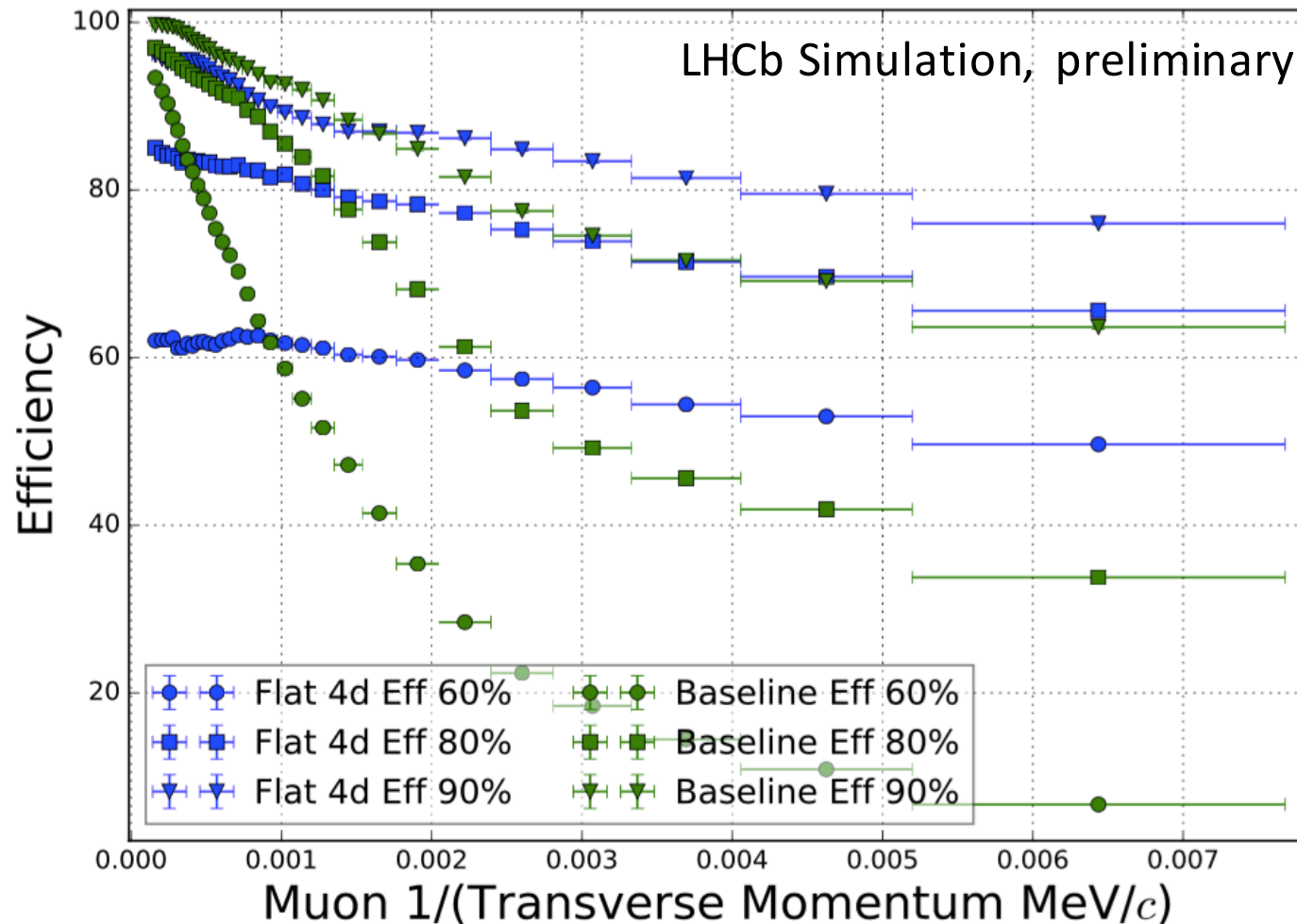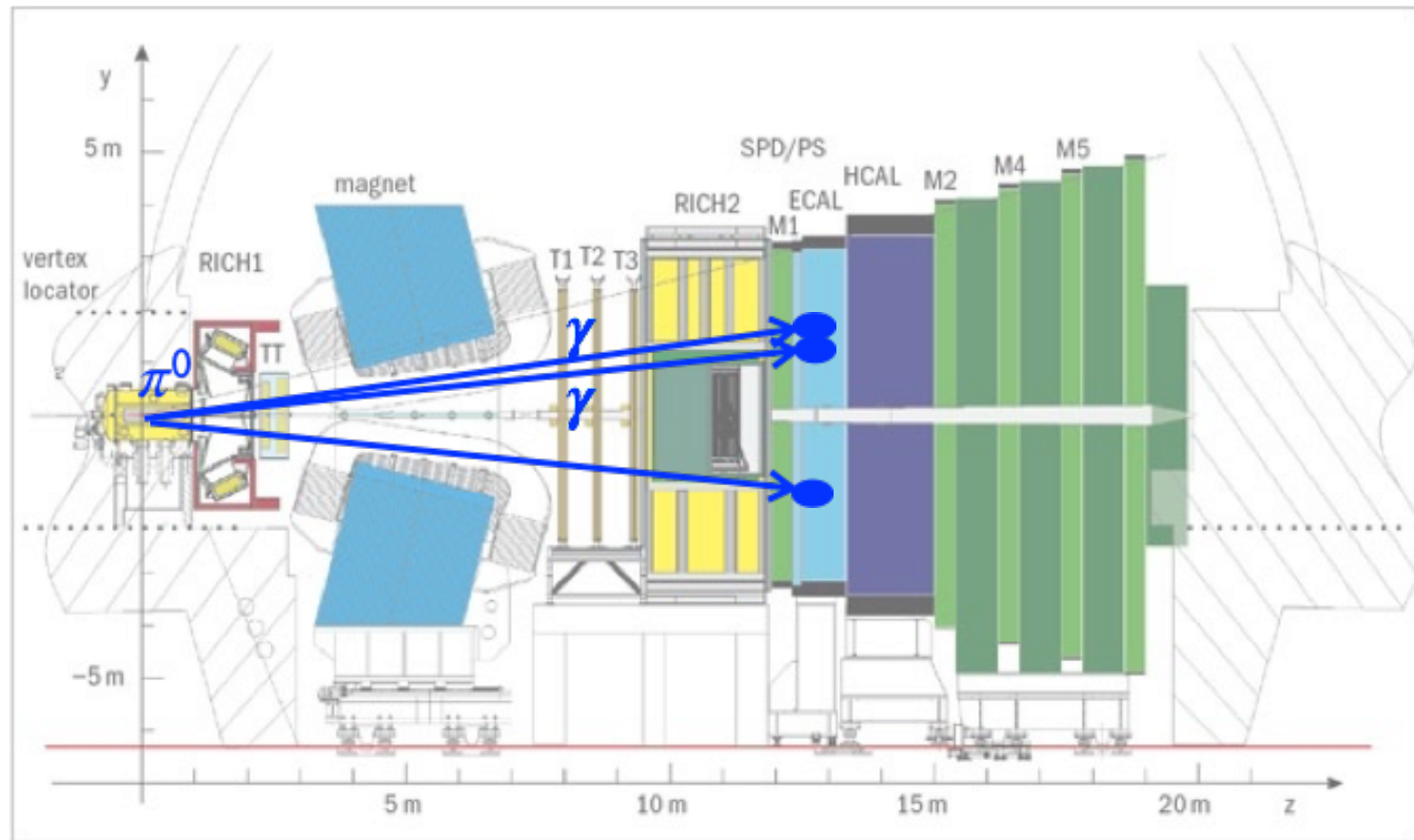
$\pi^0 - \gamma$ separation

Signal: single photon γ.
Background: photons from $\pi^0 \to \gamma\gamma$ decay.
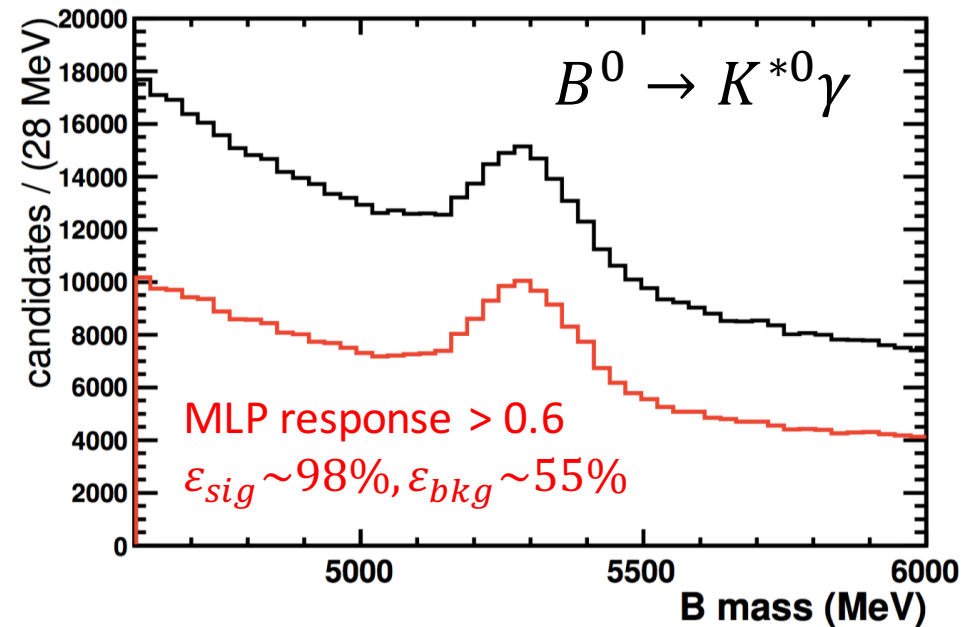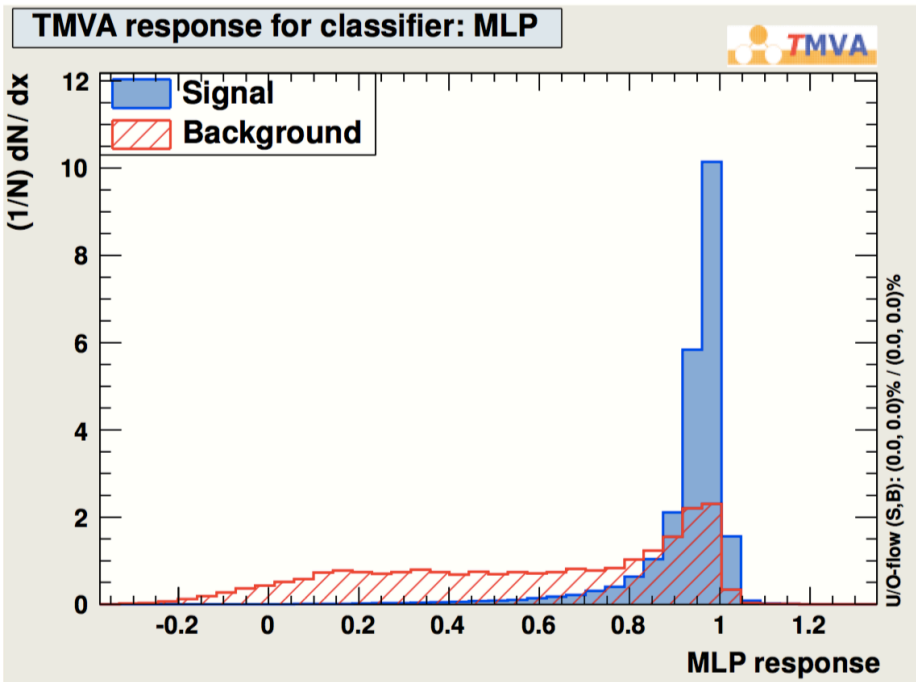**Problem:** separate signal and background clusters in the electromagnetic calorimeter.

$\pi^0 - \gamma$ separation

Baseline solution:

- Clusters shape and symetry are described by set of features.
- 2-layers MLP is trained to separate signal and background clusters.

$B^0 \to K^{*0}\gamma$

MLP response > 0.6
$\varepsilon_{sig} \sim 98\%, \varepsilon_{bkg} \sim 55\%$

# $\pi^0 - \gamma$ separation

New approach:

- Responses in 5x5 cell clusters for ECAL and pre-shower detectors are considered as new features.
- Several NN and BDT models are trained on these 2x25 input features.
- BDT model shows better performance.
- Promising possibility of aggressive background suppression is demonstrated on simulated data.
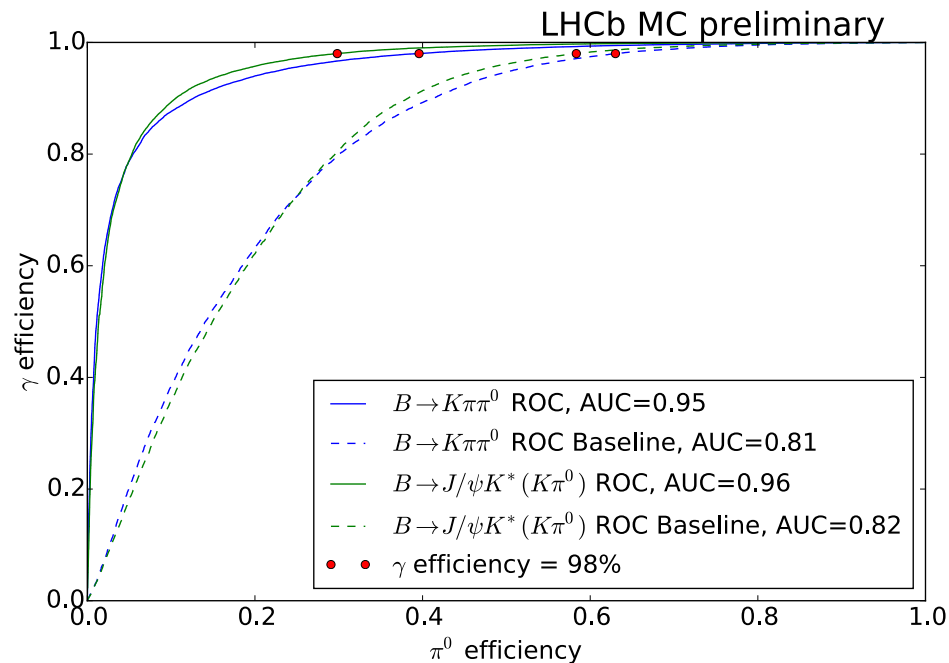
Jet Tagging

Problem: identify b and c jets with a small misidentification probability of light-parton jets. The identification of (b, c) jets is performed using SVs from the decays of (b, c) hadrons.
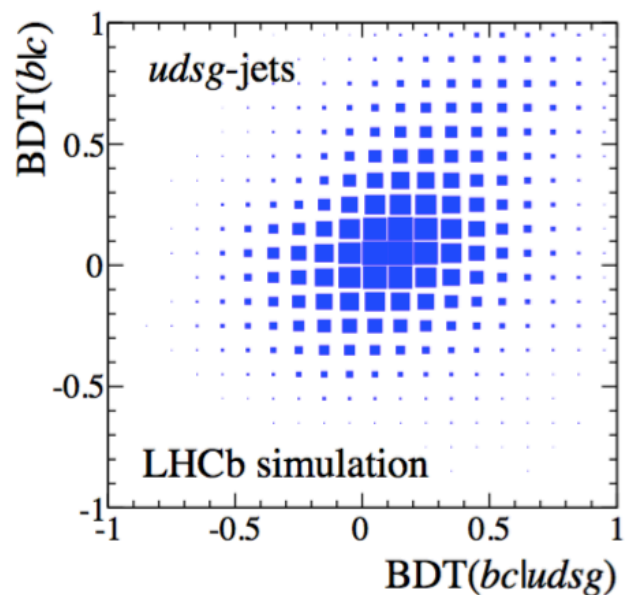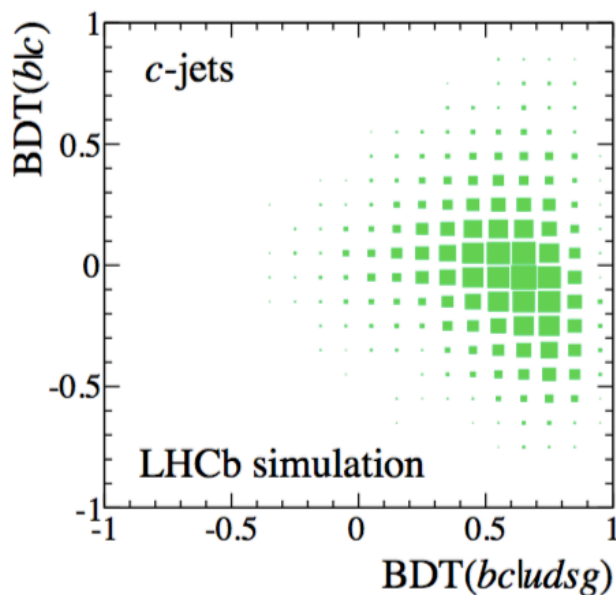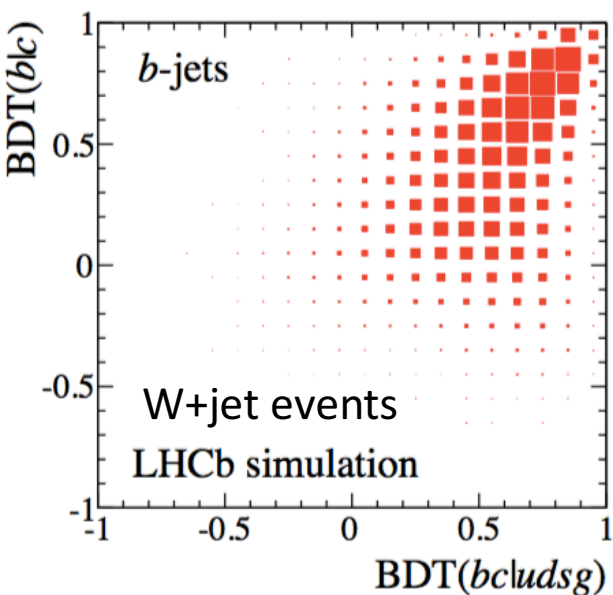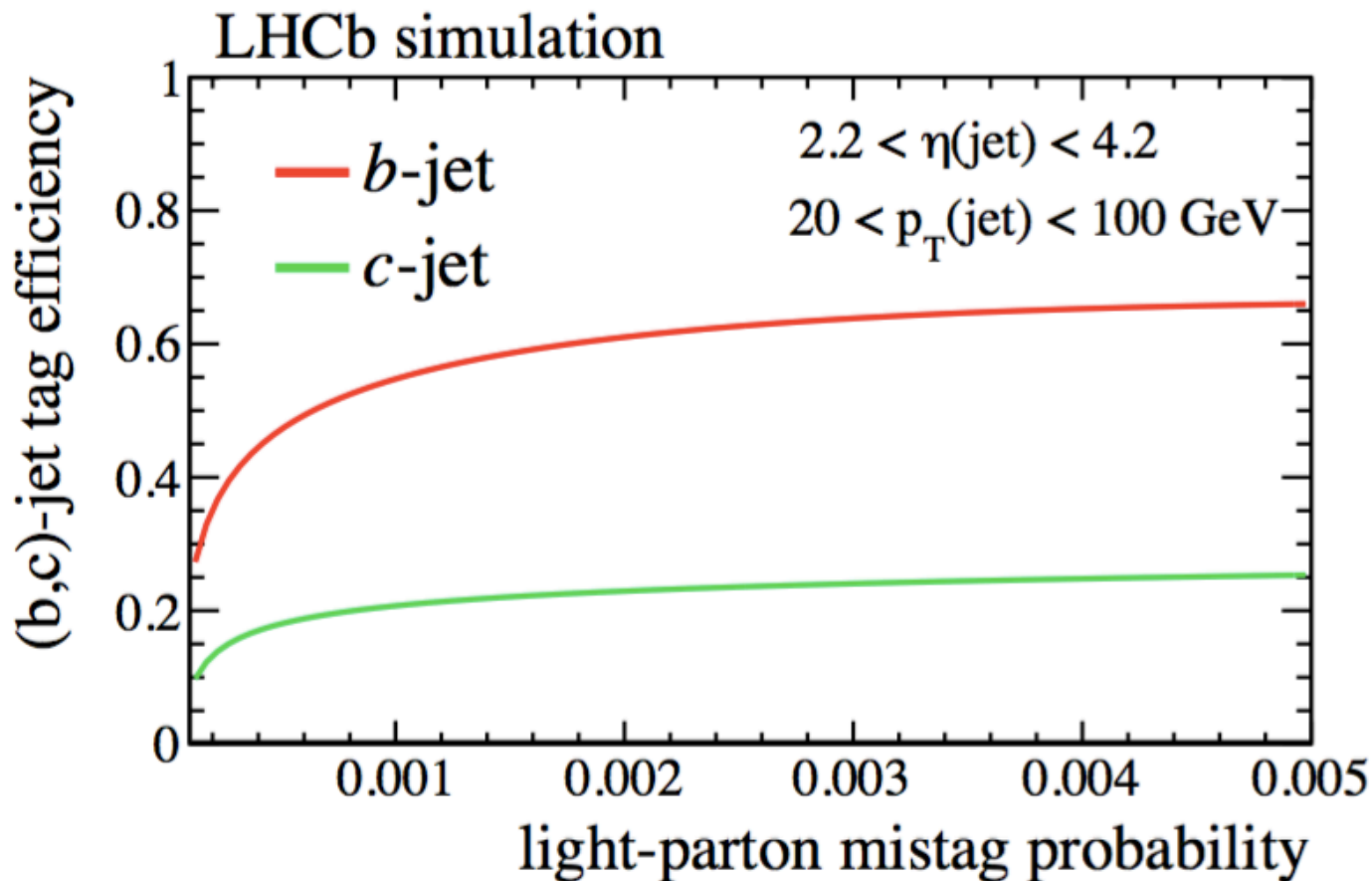
LHCb-PAPER-2015-016
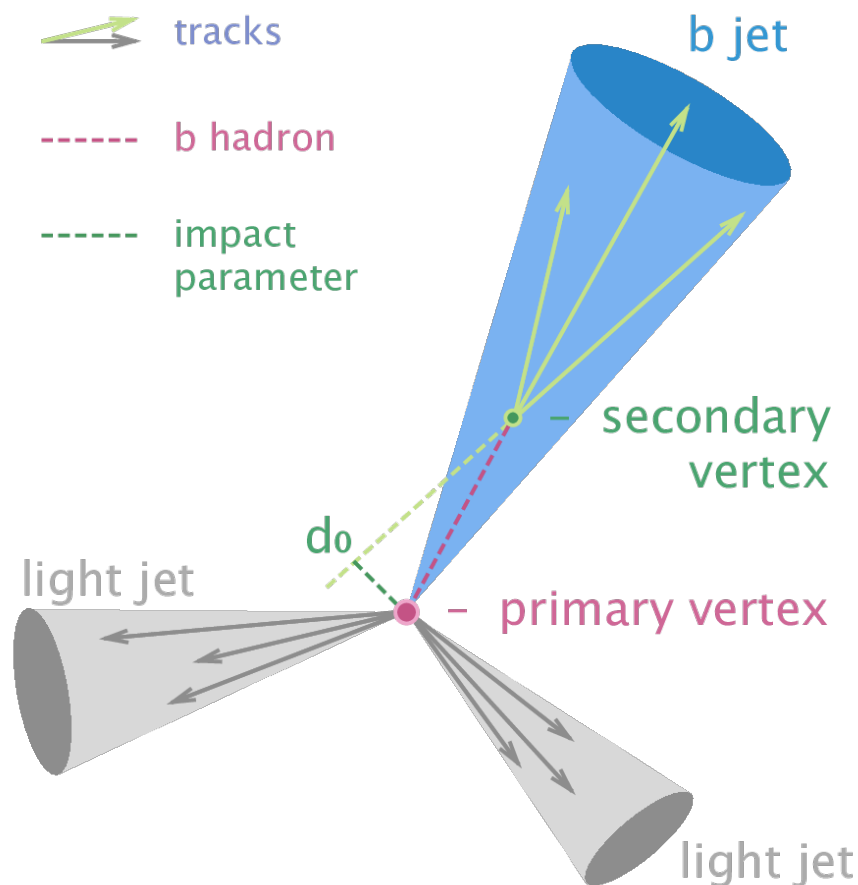
Jet Tagging

Two BDT models are considered: DBT($bc|udsg$) is trained to separate bc and $light$ jets, BDT($b|c$) is trained to separate b and c jets. Both BDTs are trained on simulated samples of b, c and light-parton jets. 10 kinematic observables of SVs are used as inputs.

LHCb-PAPER-2015-016

Jet Tagging

The (b, c)-jet efficiencies versus the mistag probability of light-parton jets obtained by increasing the DBT($bc|udsg$) cut.

Topological Trigger

- The goal of HLT2 topological trigger is efficient selection of any B (and D) decay with at least 2 charged daughters.
- It is designed to handle the possible omission of child particles.
- In Run 1, a simple BDT was used to define interesting SVs.
- In Run 2, the algorithm is reoptimized using several ML models.

J. Phys.: Conf. Ser. 664 (2015) 082025
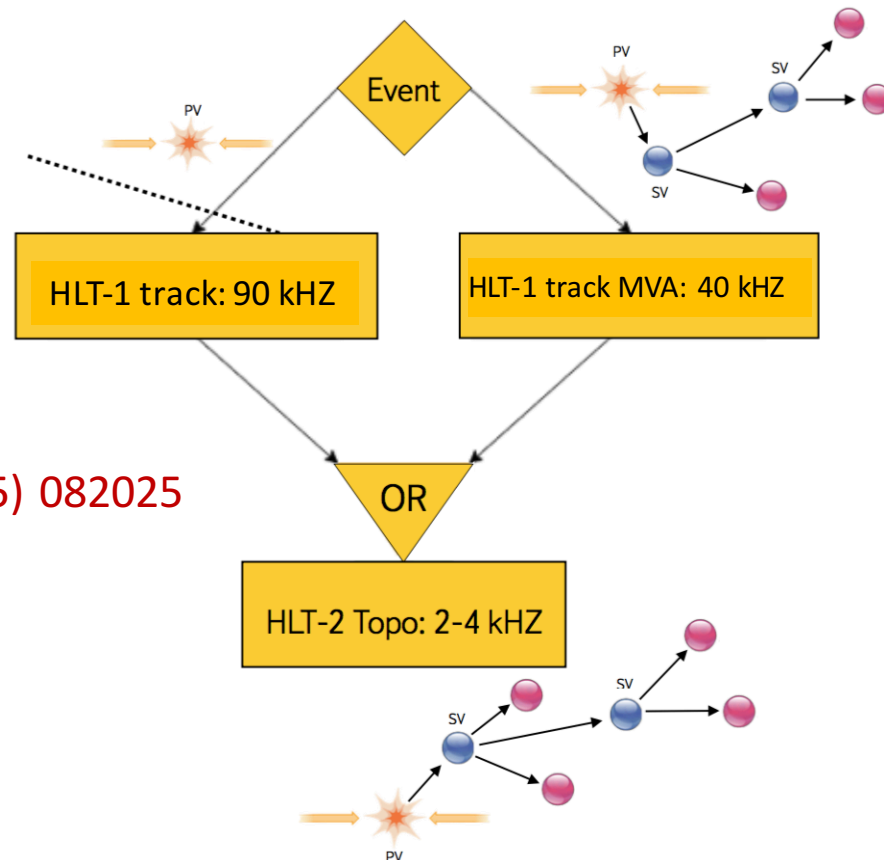
LHCb-PUB-2011-016
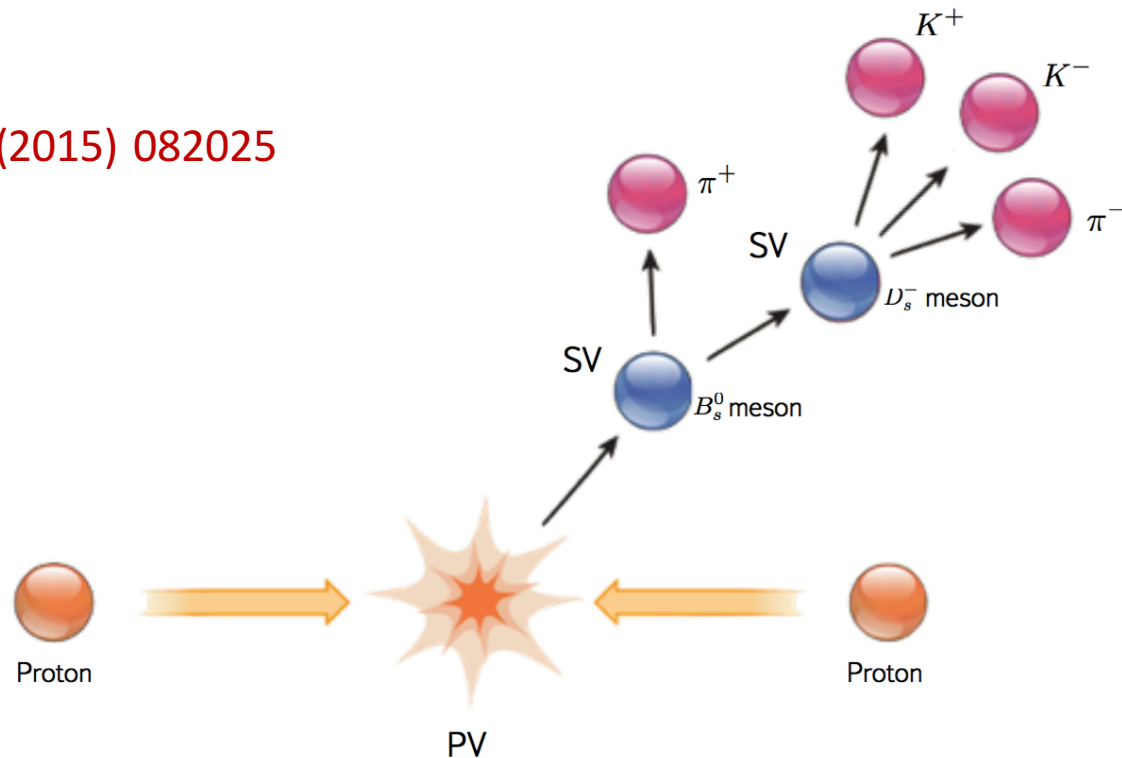
Topological Trigger

HLT-1 track is looking for one super high PT or high displaced track.
HLT-1 track MVA classifier is looking for two tracks making a vertex.
HLT-2 topological classifier uses full reconstructed event to look for 2, 3, 4 and more tracks making a vertex.
Kinematic observables of SVs are used as the classifiers inputs.
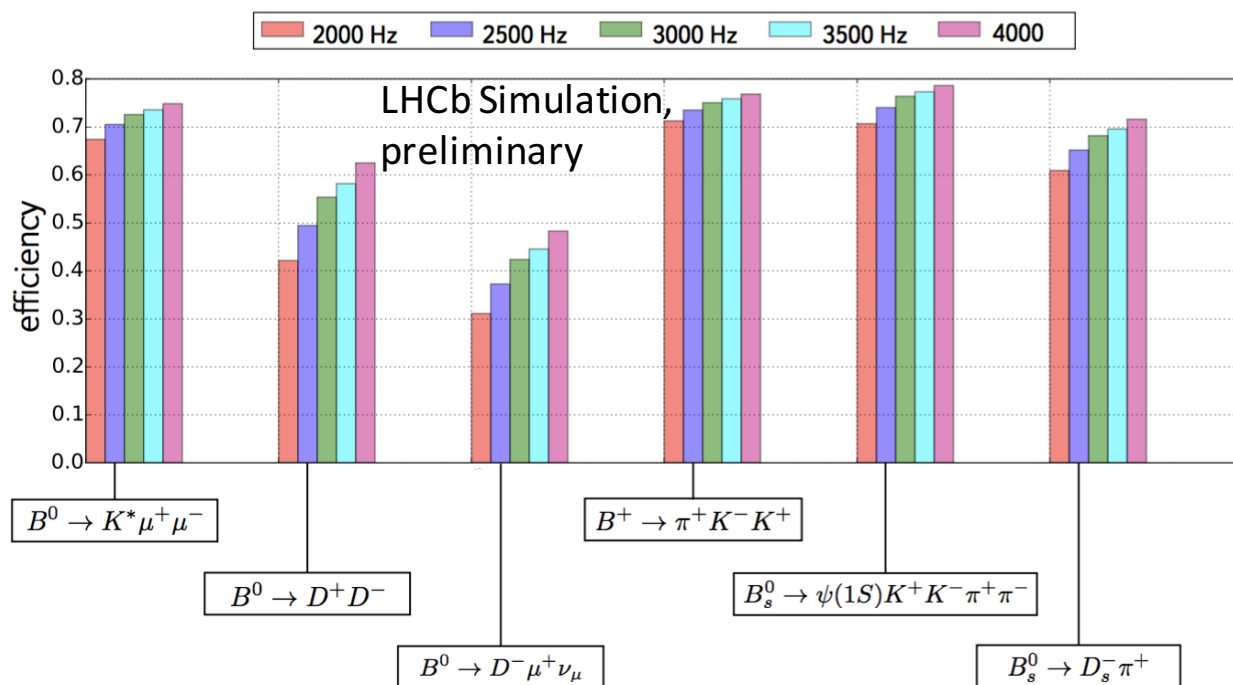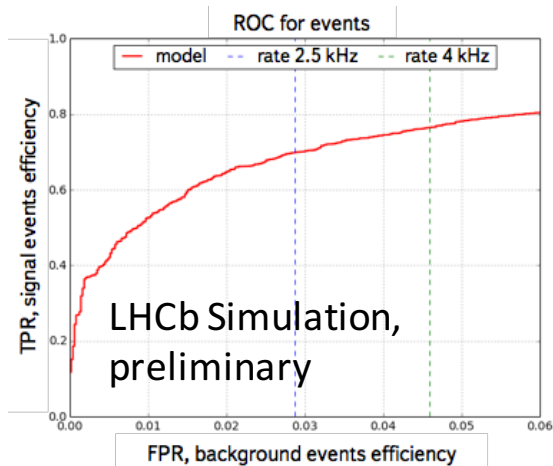

J. Phys.: Conf. Ser. 664 (2015) 082025

Topological Trigger

- Several ML models are considered during the trigger reoptimization: BDTs (MatrixNet), Neural Networks, Logistic Regression.
- ROC curve in a region with small False Positive Rate is optimized.

J. Phys.: Conf. Ser. 664 (2015) 082025

LHCb Simulation, preliminary

LHCb Simulation, preliminary

Topological Trigger

- Most n-body hadronic B decays (n ≥ 3) are only triggered on efficiently in LHCb by the topological trigger.
- Gain 50%..80% efficiency for different channels.

$$\varepsilon_{HLT}(Run\ 2)/\varepsilon_{HLT}(Run\ 1)$$

mode	4. kHz
$B^0 \to K^*[K^+\pi^-]\mu^+\mu^-$	1.72
$B^+ \to \pi^+ K^- K^+$	1.65
$B_s^0 \to D_s^-[K^+ K^- \pi^-]\mu^+\nu_\mu$	1.47
$B_s^0 \to \psi(1S)[\mu^+\mu^-]K^+ K^- \pi^+\pi^-$	1.71
$B_s^0 \to D_s^-[K^+ K^- \pi^-]\pi^+$	1.52
$B^0 \to D^+[K^-\pi^+\pi^+]D^-[K^+\pi^-\pi^-]$	1.86

Jet Tagging & TOPO Trigger

- The topological trigger algorithm uses SVs that satisfy similar criteria to those used in the SV-tagger algorithm to build two-, three- and four-track SVs.
- The SV used by the TOPO to trigger recording of the event can also be used to tag a b jet.
- The BDT used in the TOPO algorithm uses similar inputs as jet-tagger BDT models.

LHCb-PAPER-2015-016
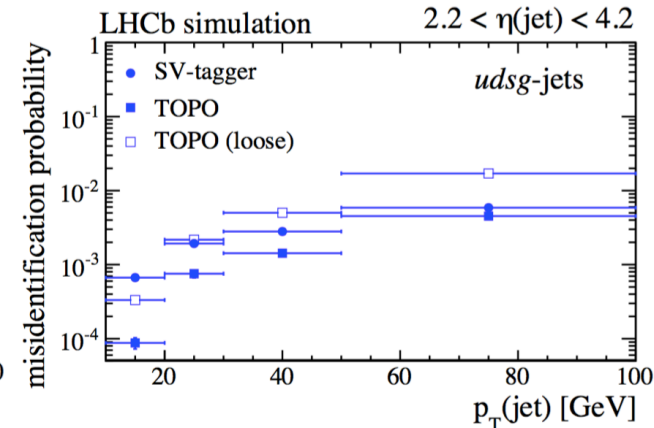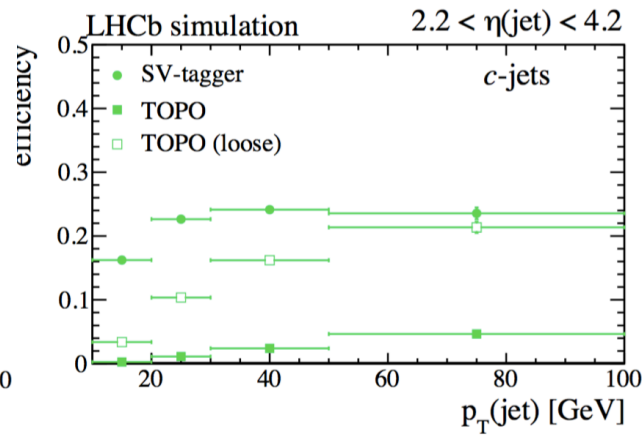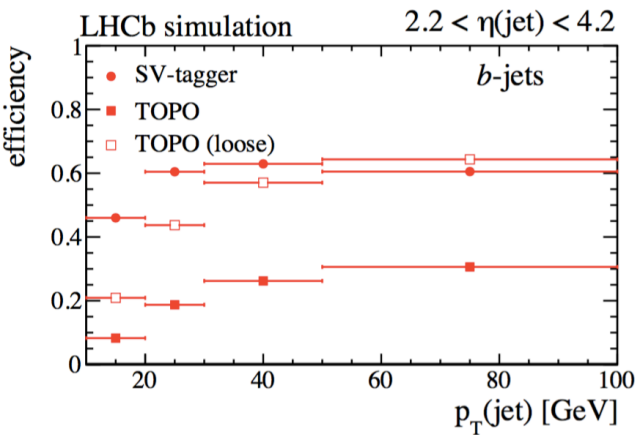
The "loose" label for the TOPO refers to the BDT requirement used in the trigger for SVs that contain muon candidates.

DQ Monitoring Robo-Shifter

- Robo-shifter is machine-learning based system designed to assists the DQ shifter.
- Given run data it can predict probability of run being good or bad.
- Provides potential problem sources extracted from decision trees.
- The first version of robo-shifter is currently being tested by the DQ shifters.

A Robo-shifter   Run: - + 183156   G

Robo-shifter

The prediction for this run is **0.47**

Please judge by distribution of predictions:
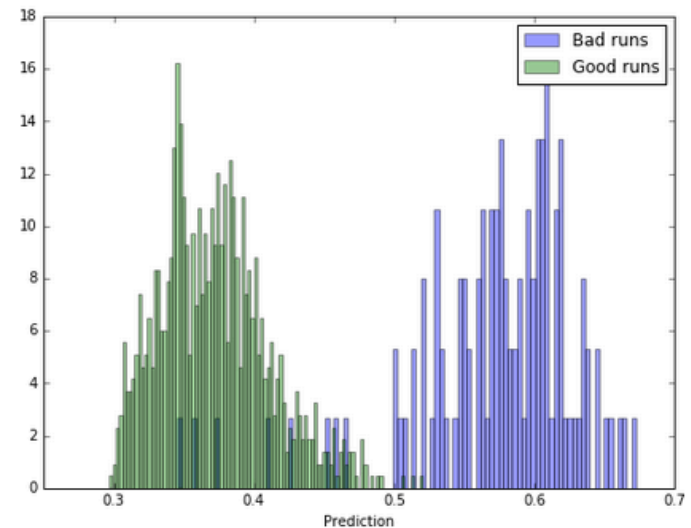
Suspicious histograms:

- **/OfflineDataQuality/ALIGNMENT: page 06: IT overlap residuals:** histogram *IT1TopBox dx*
- **/OfflineDataQuality/TESLA-BRUNEL: page 01: Tesla Brunel monitor:** histogram *TeslaBrunelMonitor*
- **/OfflineDataQuality/CALO: page 1: Photon and Electrons Reconstruction:** histogram *(gg) mass Rec/Calo/Photons*
- **/OfflineDataQuality/TESLA-BRUNEL: page 01: Tesla Brunel monitor:** histogram *TeslaBrunelMonitor*
- **/OfflineDataQuality/RICH: page 8: PID Monitoring with J-Psi:** histogram *Mass of J/psi(1S)_all*
- **/OfflineDataQuality/ALIGNMENT: page 04: RICH HPD Panel Alignment:** histogram *dTheta v phi CSide-right*

Machine learning is everywhere at LHCb helping to improve the detector operation and data processing.