Advanced machine-learning solutions in LHCb: *operations and data analysis*

Lucio Anderlini
on behalf of the LHCb Collaboration

Istituto Nazionale di Fisica Nucleare
SEZIONE DI FIRENZE

Flavour physics at the LHC: *high signal rate in the forward region*

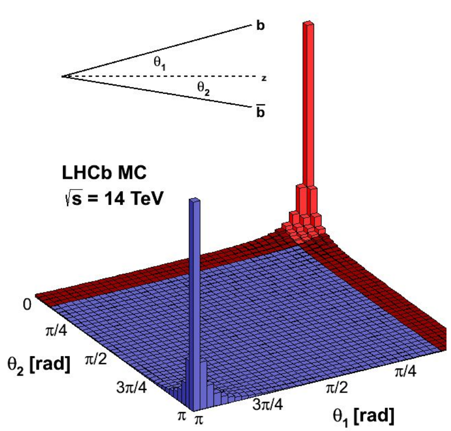
High-p_T physics:

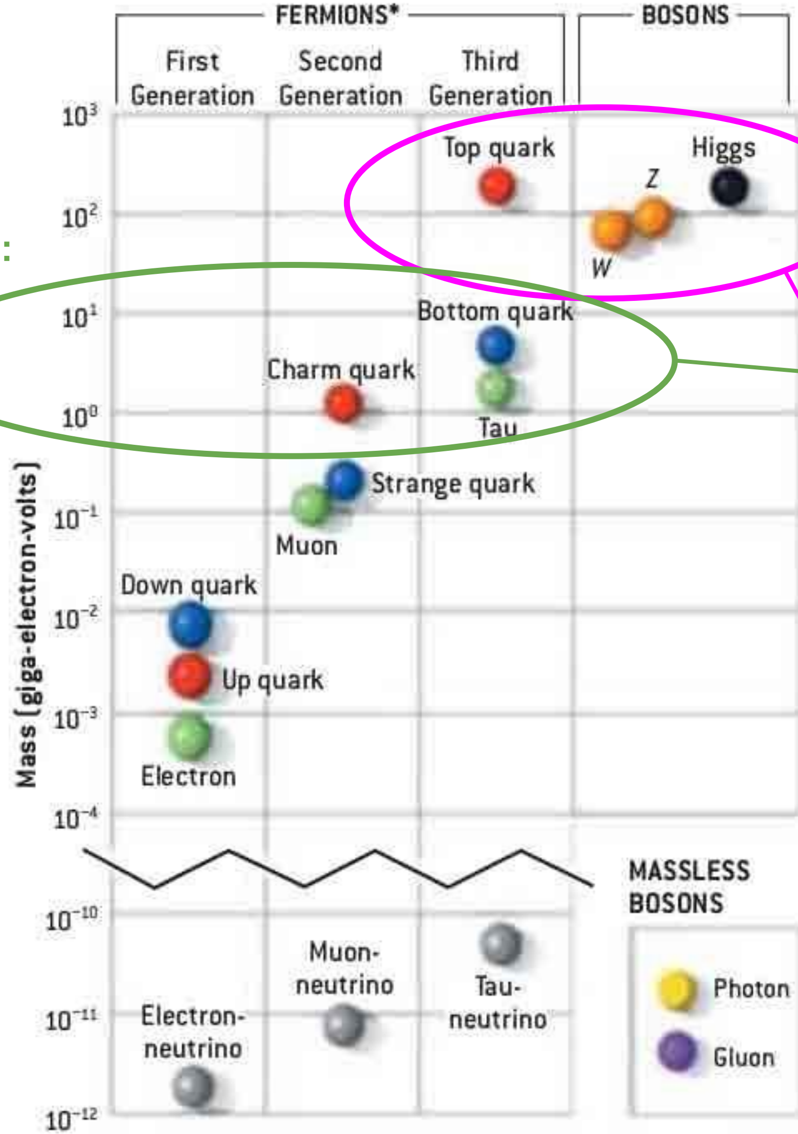
Focus of *ATLAS & CMS*

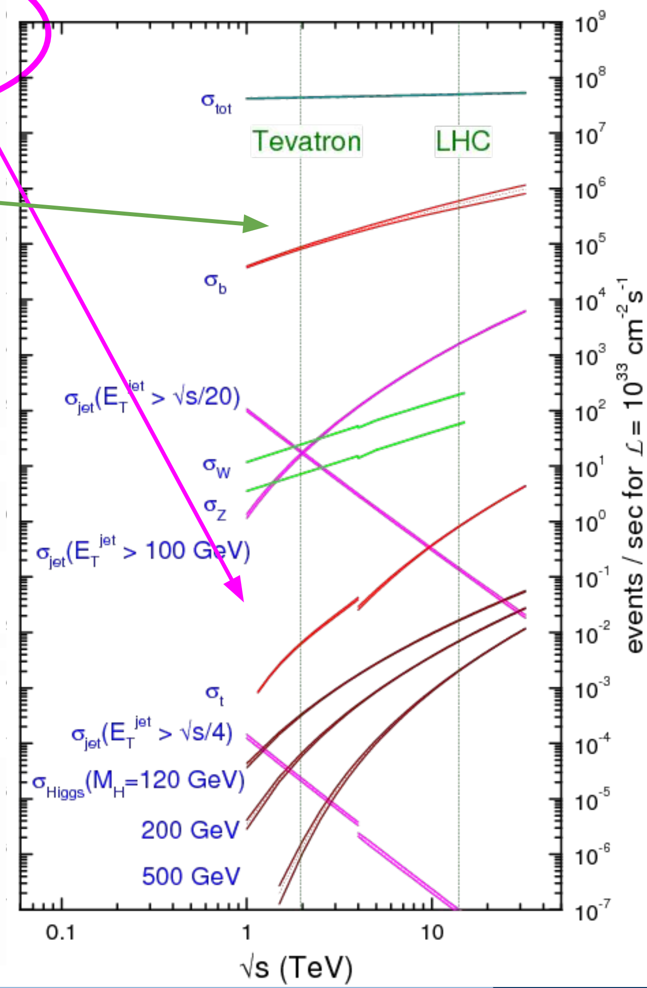
(see previous talks)

Heavy-Flavour physics:

Focus of *LHCb*



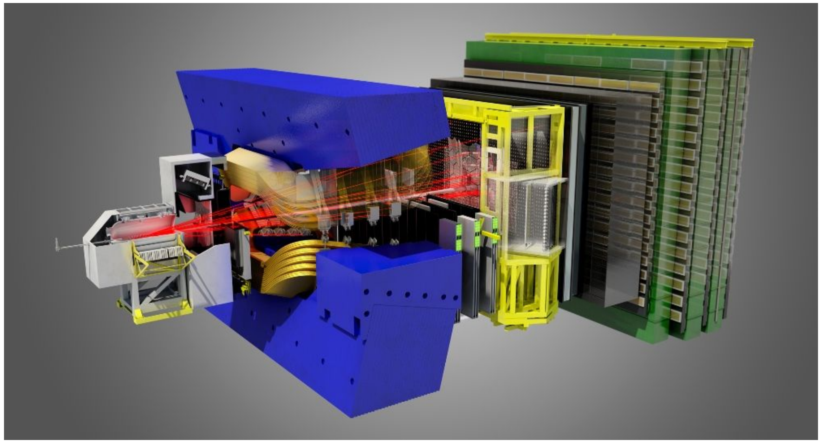




The LHCb experiment

Forward Region

Single-arm forward detector



High signal rate

Fast and reliable *real-time* data processing (**trigger**)

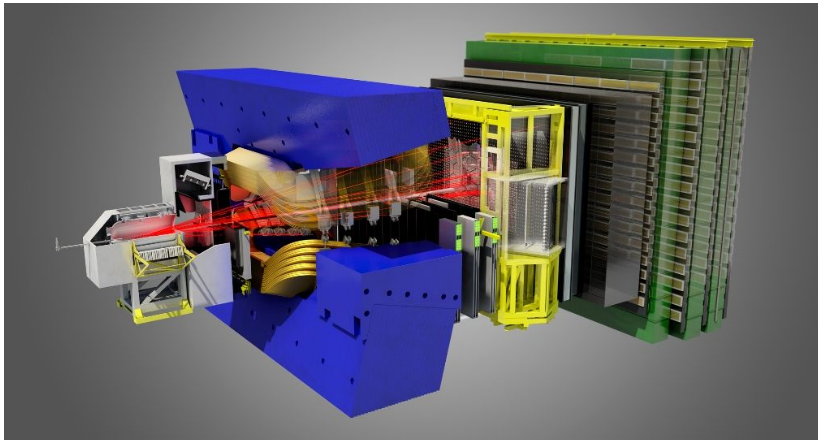
Good statistics for high precision physics

Outstanding control over the **systematic uncertainties** (*e.g.* dependences of the efficiencies)

Machine Learning for
The LHCb experiment

Forward Region

Single-arm forward detector



High signal rate

Fast and reliable *real-time* data processing (**trigger**)

- ✓ NN for fast reconstruction (tracking)
- ✓ BDT for High-level topological triggers

Good statistics for high precision physics

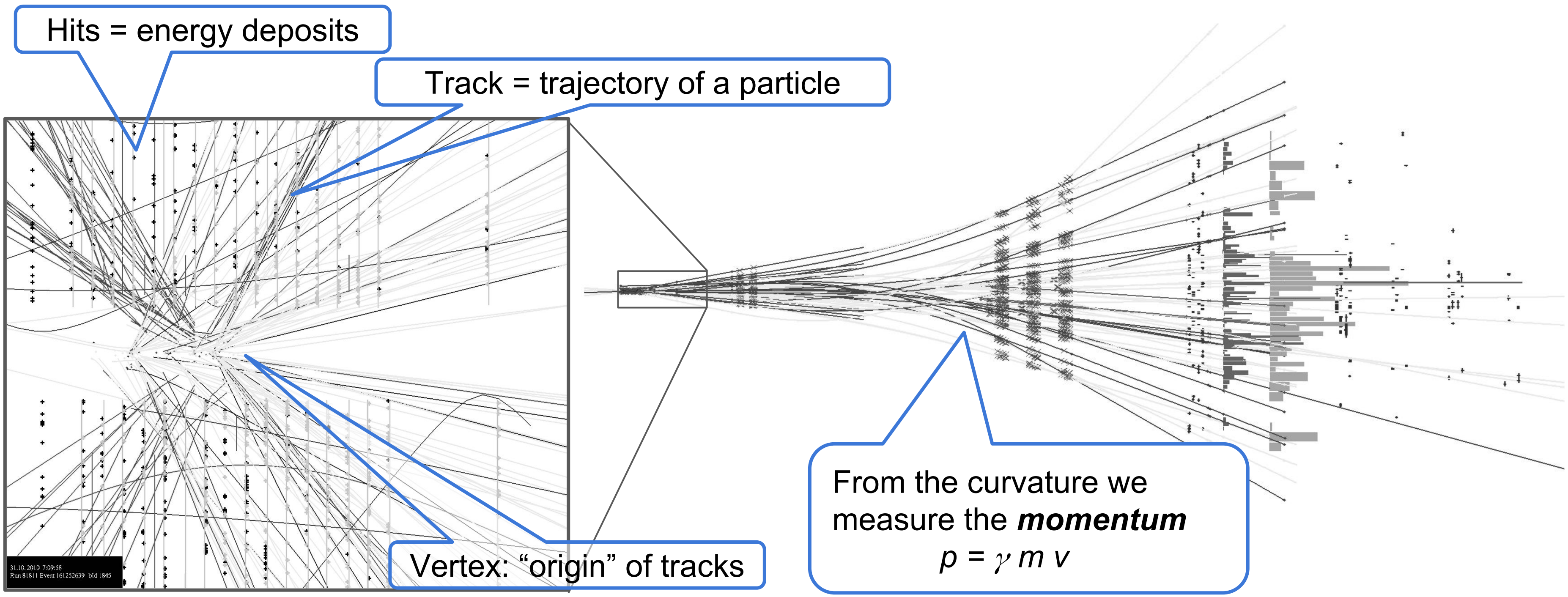
Outstanding control over the **systematic uncertainties** (*e.g.* dependences of the efficiencies)

- ✓ Combine detector responses for particle identification
- ✓ Gen. models for efficiency modeling

Conceptual steps

Hits in the detector → **Tracks** → Tracks with mass hypothesis (PID) → Tracks from heavy hadron decay → Fully reconstructed decays → Comparison with theory

Neural Networks for Track Reconstruction

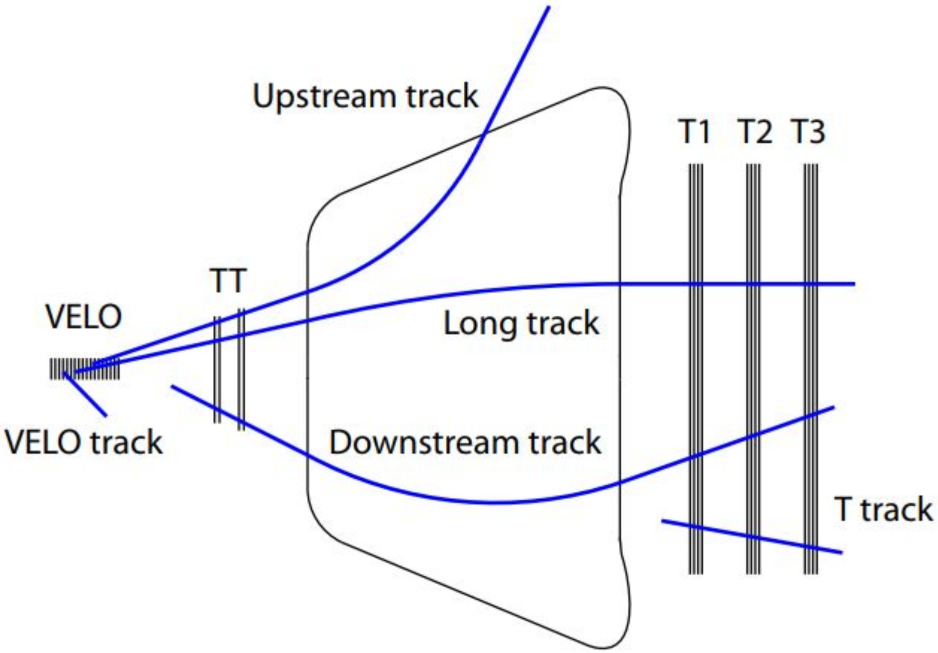


Tracking

Machine Learning used in several steps of the track reconstruction.

For instance, in creation of:
- Long tracks,
- Downstream tracks

At the end of the "tracking sequence", fake tracks are rejected using a ***deep neural network.***



Profits:

- efficiency gain,
- fake tracks reduction,
- faster execution in the trigger.

More about tracking at LHCb

- LHCB-PROC-2017-013
- LHCb-PUB-2017-001

Fake track rejection

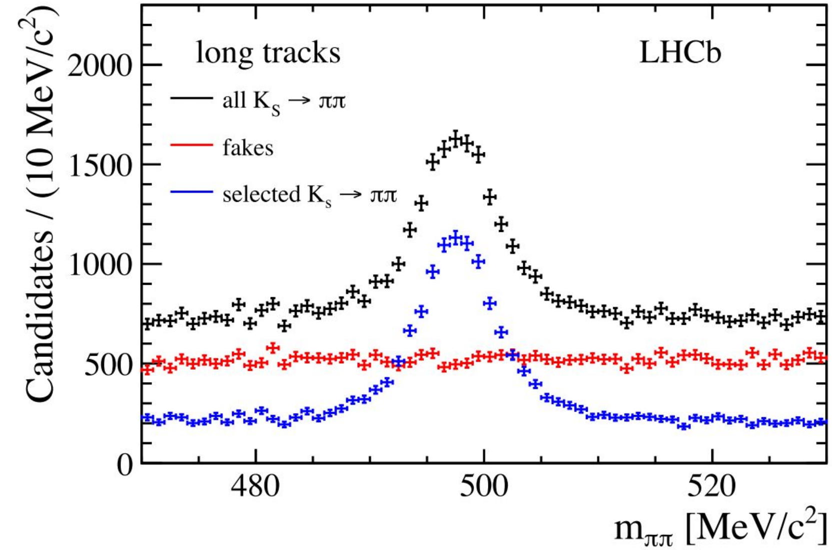
Fake tracks produced:

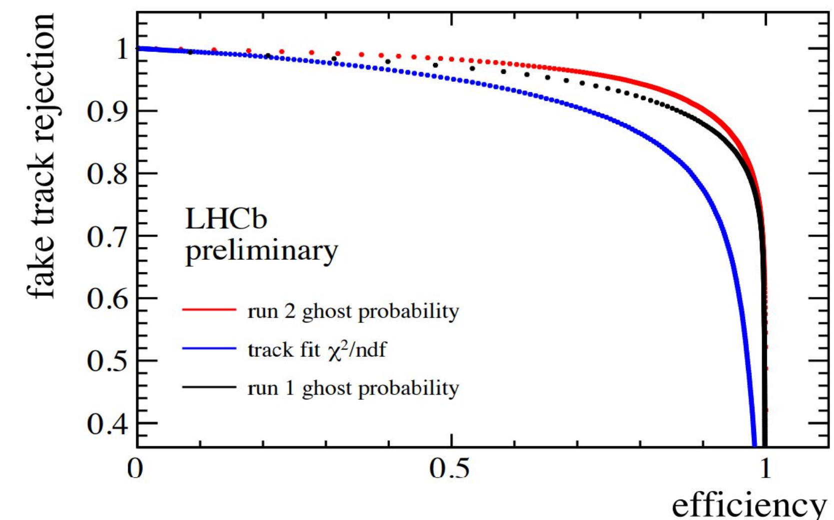
- ➪ in the matching between the VELO and the upstream tracking stations (step 2)
- ➪ In the Kalman-Fit procedure (step 4)

Rejecting fake tracks at an early stage is crucial to **reduce the CPU cost** of the upcoming Particle Identification and event reconstruction.

A ***deep neural network*** is trained on Simulation to improve the fake track rejection. Track features are (22 features):

- ✓ quality of the Kalman-Filter fit (**χ^2**) and number of hits for each sub-detector
- ✓ the reconstructed momentum (**p_T and η**)
- ✓ average **occupancy** of each sub-detector





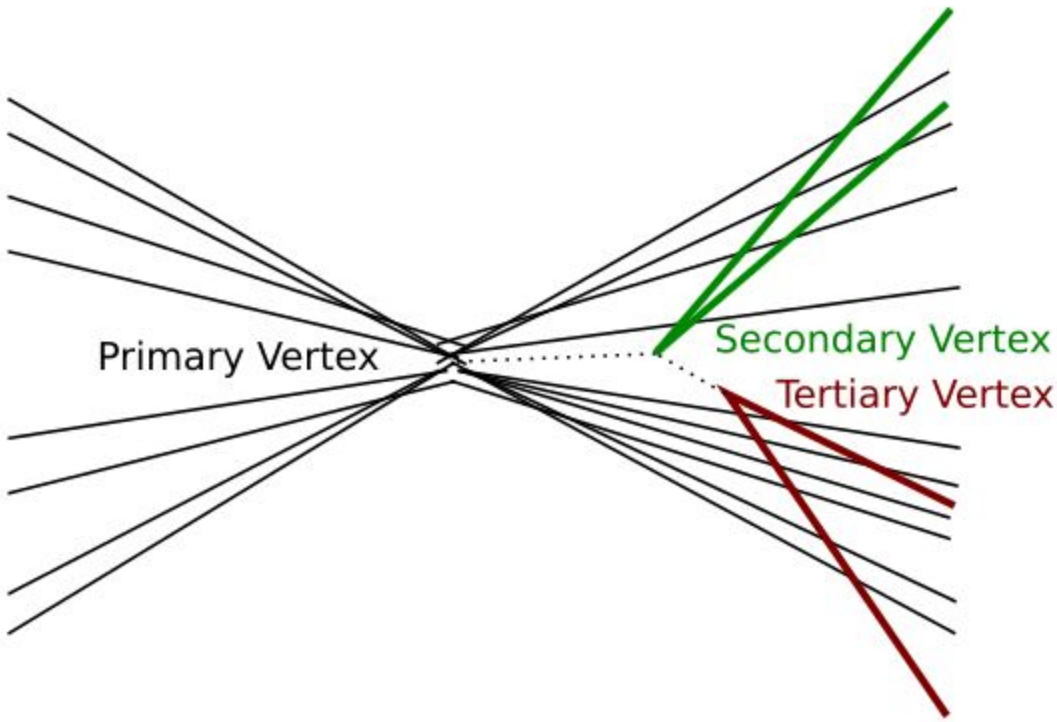
Overall gain of 16% of CPU resources in the real-time processing (trigger) PC farm. CPU

Output-rate reduced by 36%. TAPE & DISK

Conceptual steps

Hits in the detector → **Tracks** → Tracks with mass hypothesis (PID) → **Tracks from heavy hadron decay** → Fully reconstructed decays → Comparison with theory

Topological trigger selection



How to trigger on *b*-hadron *decays*

HIGH RATE — GENERIC and SIMPLE criteria

Selecting events with at least *n* tracks inconsistent with the PV

Irreducible background from strange, charm and fake tracks.
Used at an early stage with n = 1

Trigger on partially reconstructed *b*-hadron decays

Topological trigger: cannot rely on hadron identification.
Requires ≥ 3-body decays

Preselect tracks inconsistent with all PVs and $p > 5$ GeV/c and $p_T > 0.5$ GeV/c

For each **pair of tracks** with *distance of closest approach (DOCA)* < 0.2 mm

Add to the 2-track **combination** a 3rd track with *DOCA* < 0.2 mm

Add to the 3-track **combination** a 4th track with *DOCA* < 0.2 mm

Veto on prompt charm

Binary classification of the combination

LOW RATE — TIME-CONSUMING criteria

Fully reconstruct the decay through an exclusive decay mode

*Used for 2-body decays since **2010***
*Since **2015**: **"Turbo" triggers** available, but suboptimal for many-body decays: too many PID combinations; too likely to miss a track*

Make classification *fast* and *stable* with a Bonsai-BDT

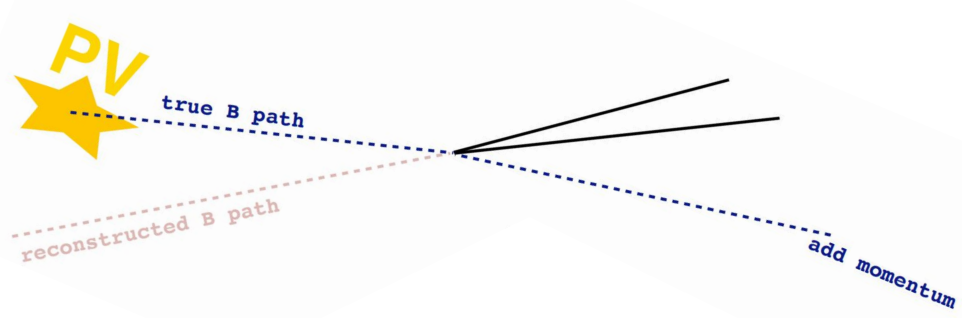
Train a *Gradient-Boosted Decision Tree* on **discretized features** and convert the decision rule into a **1D array look-up problem.**

***discrete features* ⇒ insensitive to fluctuations of the resolution functions;**
1D-array look-up ⇒ virtually zero evaluation time.

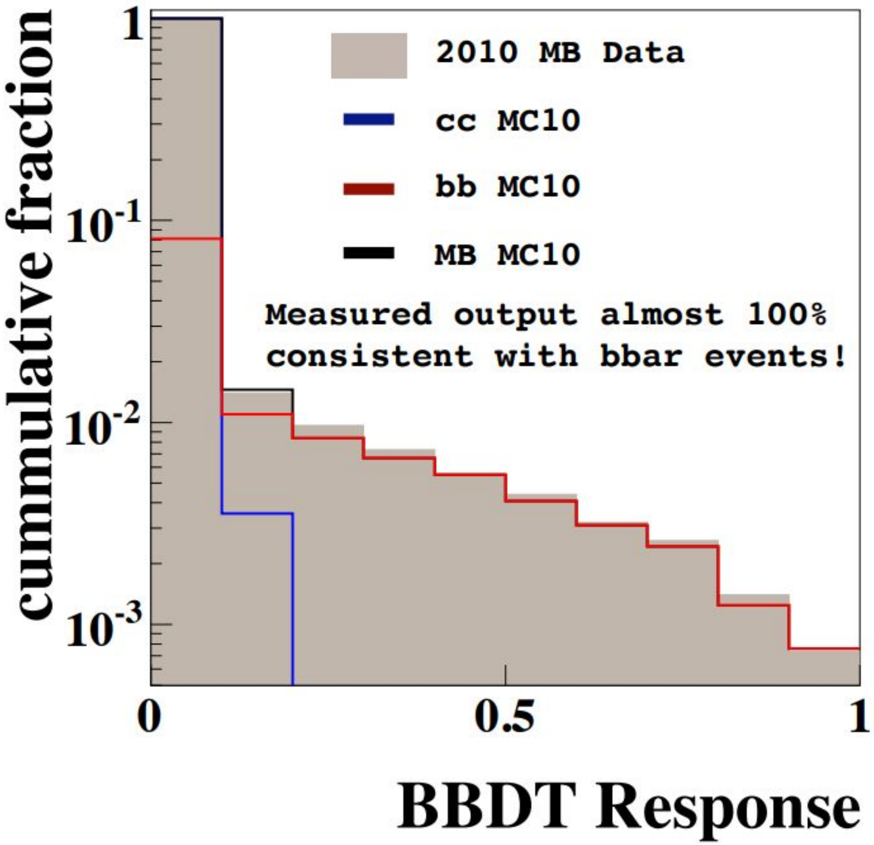
Features:

- Sum of the p_T of the tracks (12 bins) and minimum (15 bins)
- invariant mass of the combination (3 bins)
- Distance of closest approach (4 bins)
- Consistency of tracks (2 bins) and secondary vertex (13 bins) with any PV
- Corrected mass (11 bins):

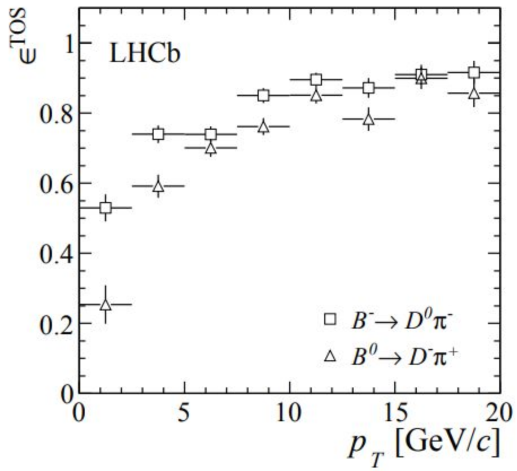
$$m_{\rm corr} = \sqrt{m^2 + |p'_{T\rm miss}|^2} + |p'_{T\rm miss}|$$



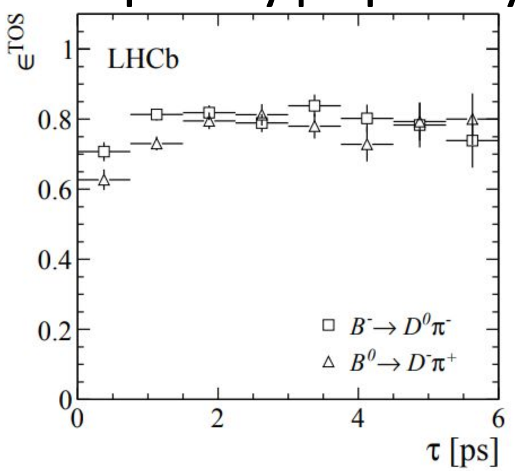
Performance of the BBDT-based triggers



Topo2Body | Topo3Body



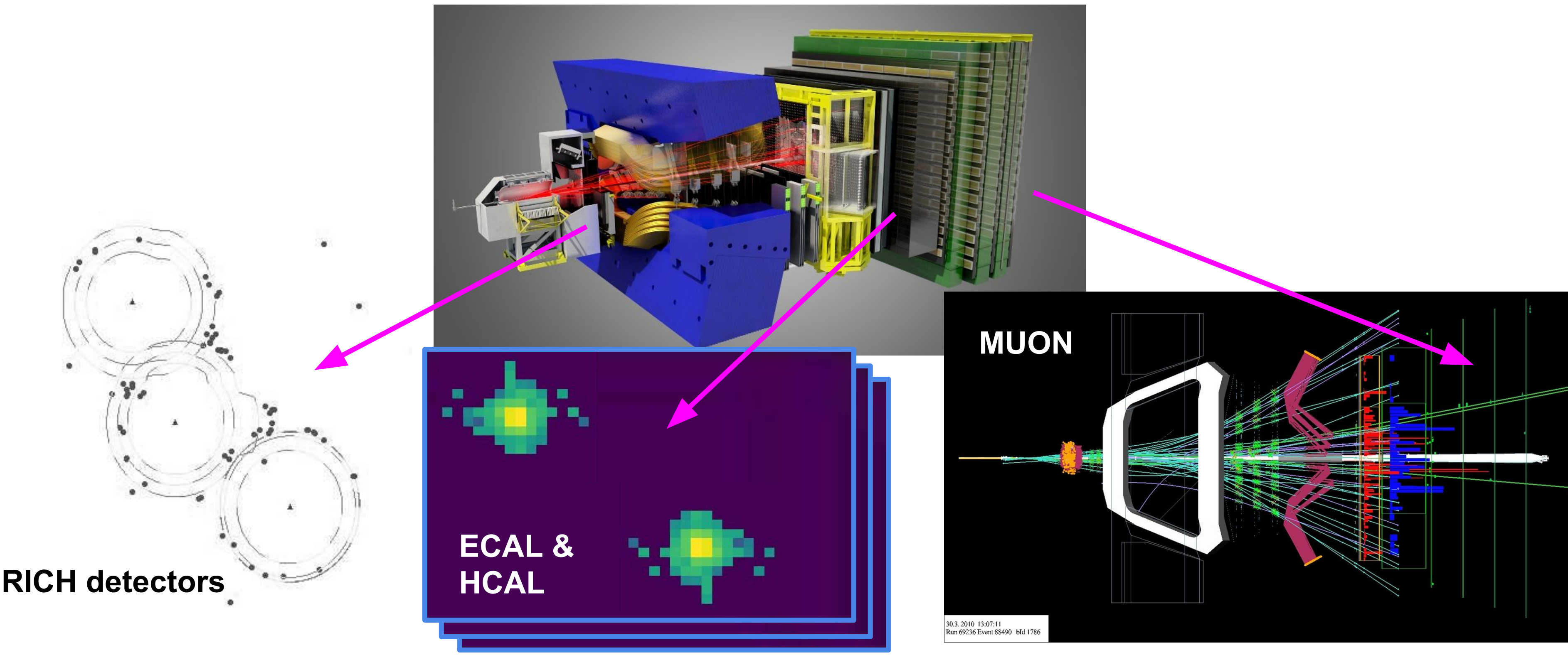
Topo2Body | Topo3Body



Conceptual steps

Hits in the detector → **Tracks** → **Tracks with mass hypothesis (PID)** → Tracks from heavy hadron decay → Fully reconstructed decays → Comparison with theory

Smart Particle Identification

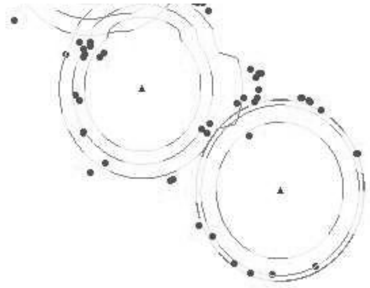


Particle Identification at LHCb

Particle Identification (PID) is a **crucial step in the** reconstruction pipeline that allows:

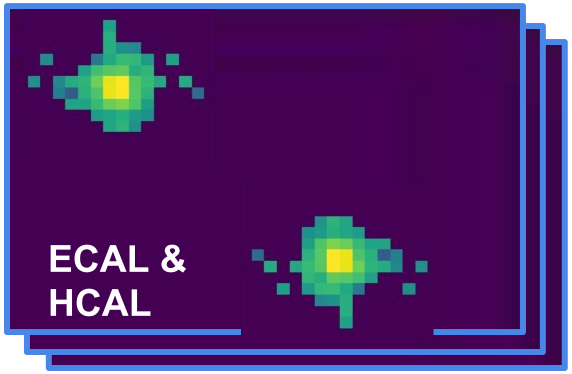
- To attach a **mass hypothesis** to the reconstructed charged tracks
- To filter out abundant particles when looking for **rare signatures**
  - ✓ **N(pions) > N(kaons) > N(protons) > N(electrons) > N (muons)**
- *To observe photons (and reconstruct $\pi^0 \to \gamma\gamma$ decays)*

Four different sub-detectors contribute to form the mass hypothesis, with totally different **principle, read-out, reconstruction strategy**:

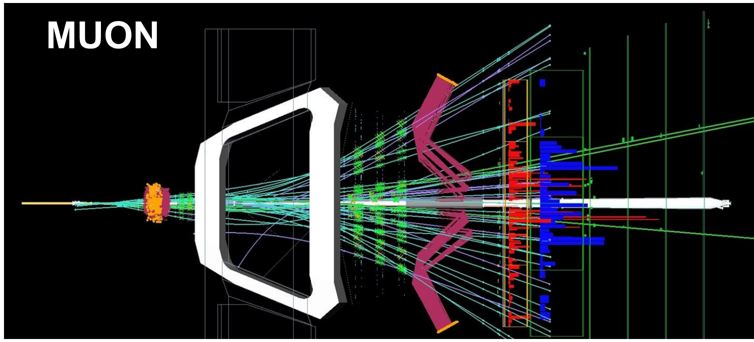


RICH detectors

(Better) identifies:
pions, kaons, protons
Rings expected from the track parameters are compared to hits



(Better) identifies electrons.
Check consistency of clusters of hits w/ tracks.



(Better) identifies muons.
Check consistency of tracks of hits w/ tracks.

How to combine the response from the detectors?

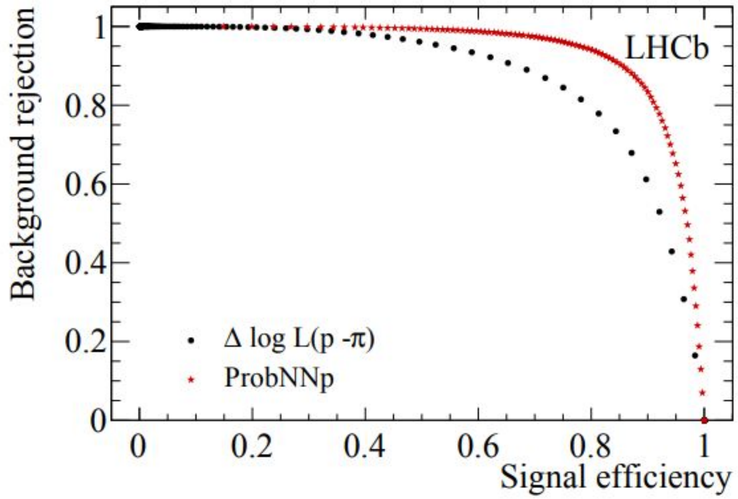
Combined Likelihood Approach

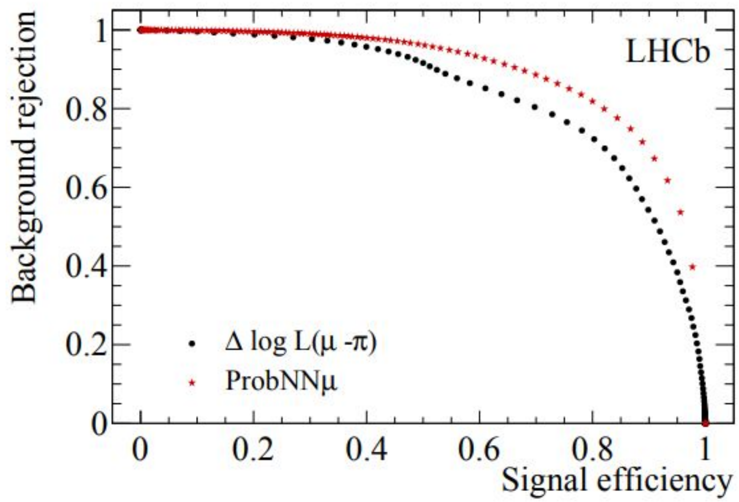
Compute separately the likelihood ratio $\mathcal{L}_X/\mathcal{L}_\pi$ for the various detector and arithmetically sum up the $\log \mathcal{L}_X/\mathcal{L}_\pi$

Theoretically the most powerful test,
in practice there are parameters of the detector response that cannot be easily included in a likelihood computation (e.g. the number of hits shared with neighbour tracks)

Machine Learning Approach

Feed a Multi-Label classifier with all the features associated to a track in the reconstruction of each detector, and train it on a large simulated dataset.





The most widely adopted ML solution: *ProbNN*

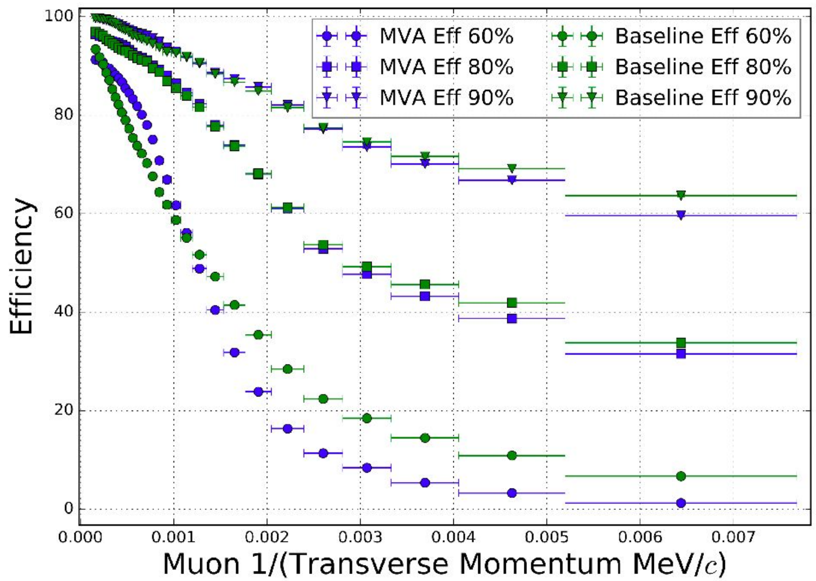
The most widely adopted ML solution: ProbNN

- Shallow Neural Network (TMVA)
- **Sigmoid** activation function
- Loss function: ***Bernoulli Cross-Entropy***

Input features:

- ✓ ***Tracking***: momentum and track quality
- ✓ ***RICH:*** likelihood ratios; geometrical and kinematical acceptance flags
- ✓ ***Calorimeters:*** likelihood ratios, quality of the track−cluster matching
- ✓ ***Muon system:*** geometrical acceptance, binary response used at trigger level, likelihood ratio based on muon-track quality.

New algorithm based on Deep Neural Nets (keras) in multiclassification mode.



Both **implicit ($\mathcal{L}_X/\mathcal{L}_\pi$)** and **explicit** (features) dependence on kinematic variables: **careful modeling is required**.

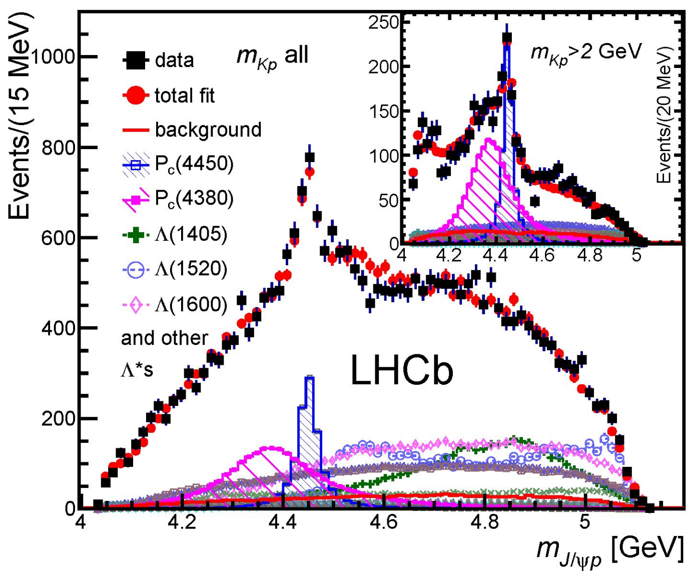
(1-AUC)/(1-AUC$_{\text{baseline}}$) LHCb Simulation, preliminary

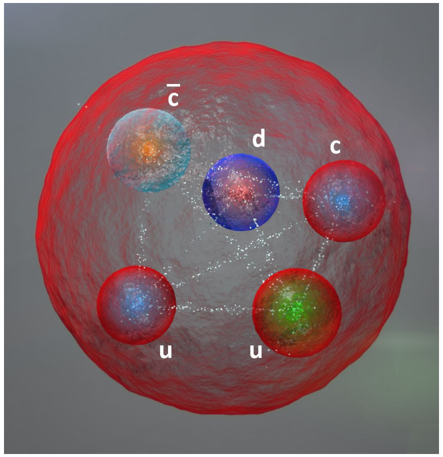
	Ghost	Electron	Muon	Pion	Kaon	Proton
baseline	1	1	1	1	1	1
deep NN	-29 %	-41 %	-52 %	-37 %	-20 %	-17 %

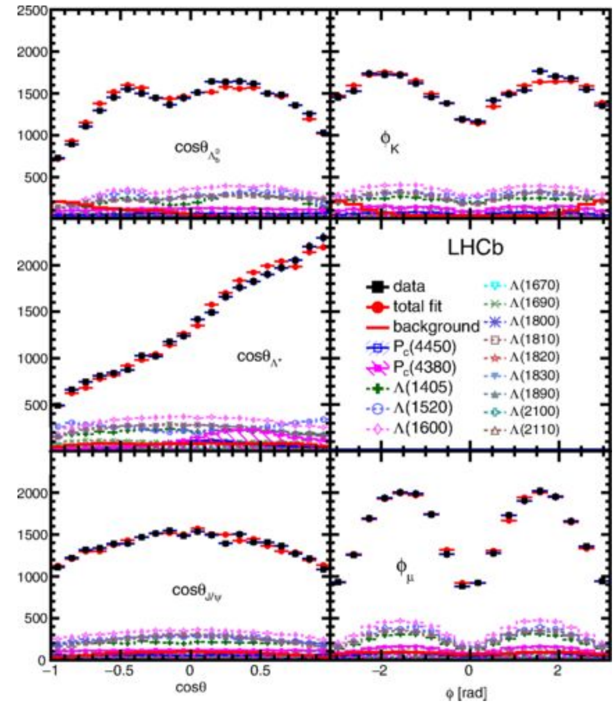
Conceptual steps

Hits in the detector → Tracks → Tracks with mass hypothesis (PID) → Tracks from heavy hadron decay → **Fully reconstructed decays** → **Comparison with theory**

Machine Learning in the comparison of DATA with THEORY







Modelling of steep variation of a complicated efficiency

1. To measure spectra, one needs good modelling of **relative efficiency variations.**
2. A **perfect simulation** of all the detectors concurring to PID is a **very challenging task**

Two approaches: Flatten the efficiency response | Model efficiency variation w/ data

Modelling of steep variation of a complicated efficiency

1. To measure spectra, one needs good modelling of **relative efficiency variations.**
2. A **perfect simulation** of all the detectors concurring to PID is a **very challenging task**

Two approaches: Flatten the efficiency response | Model efficiency variation w/ data

Train a classifier with a loss function including a "FLAT Cramer−von Mises" term

$$\mathcal{L}_{FL_X} = \left\langle \int \left(F_{global}(s) - F^{(X)}_{local}(s)\right)^2 ds \right\rangle$$

Classifier response

Average on X bins

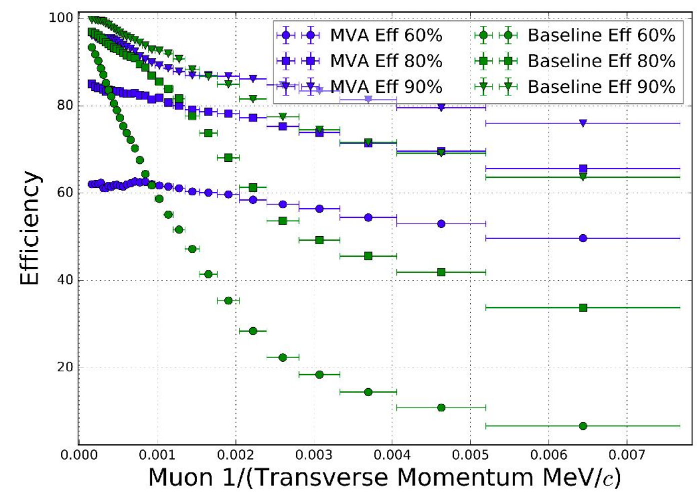
Cumulative distribution of s

Cumulative distribution of s in *intervals of variable X*

Loss function for $X = p, p_T$, multiplicity and pseudorapidity are summed up:

$$\mathcal{L}_{FL_{4d}} = \mathcal{L}_{FL_p} + \mathcal{L}_{FL_{p_T}} + \mathcal{L}_{FL_{nTracks}} + \mathcal{L}_{\eta}$$

And used to train *oblivious decision trees to obtain "Flat PID models"*



… but to introduce ***flatness,*** some **discrimination** power is lost…

Modelling of steep variation of a complicated efficiency

1. To measure spectra, one needs good modelling of **relative efficiency variations.**
2. A **perfect simulation** of all the detectors concurring to PID is a **very challenging task**

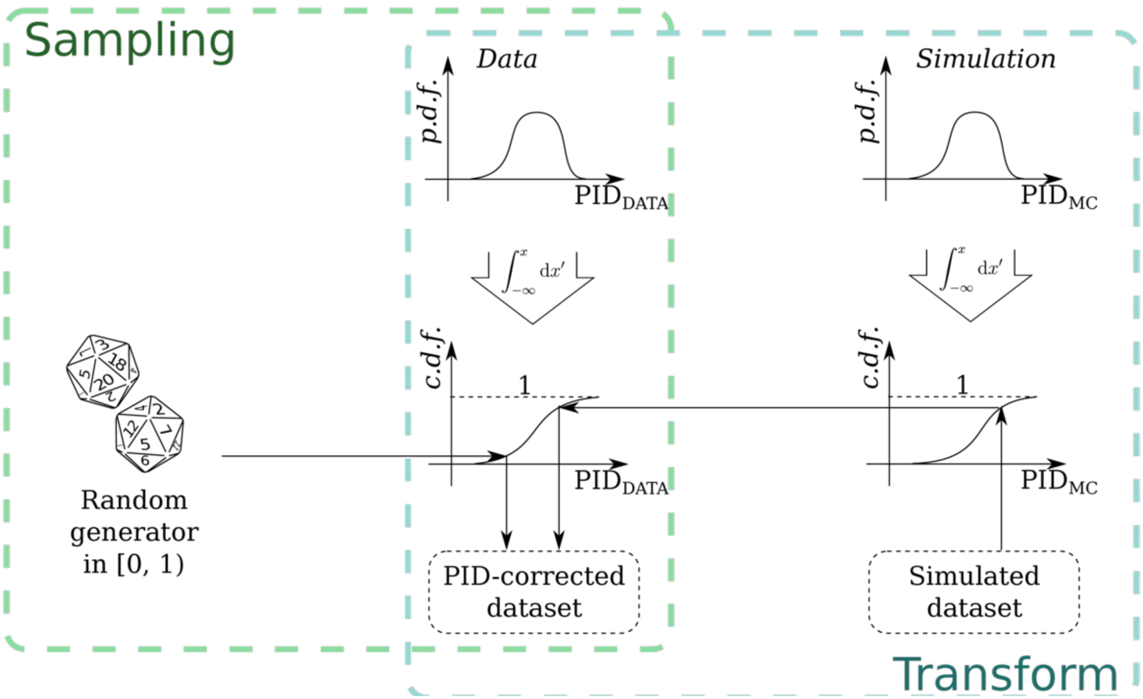
Two approaches: **Flatten the efficiency response** | **Model efficiency variation w/ data**

Non-parametric density estimation in the space (PID, η, p_T, multiplicity) based on the **Meerkat** algorithm.

Training on ***real*** and ***simulated*** calibration samples (unambiguous PID from kinematic constraints).

Build a ***generative model*** on top of data PDEs (sampling);

or ***correct*** the simulated response for signal sample.



$$\mathrm{PID_{corr}} = P^{-1}_{\mathrm{exp}}(P_{\mathrm{MC}}(\mathrm{PID_{MC}}|p_{\mathrm{T}}, \eta, N_{\mathrm{evt}})|p_{\mathrm{T}}, \eta, N_{\mathrm{evt}}),$$

Conclusion

Conclusion and summary

The LHCb software is being complemented end-to-end with Machine Learning solutions.

- ➪ Reconstruction
- ➪ High-Level Trigger Selection
- ➪ Combination of PID detector responses
- ➪ Random generation (or correction of full simulation) from what learnt from data

The challenge for the future upgrade is to further increase the Machine-Learning solutions to

- ➪ speed-up the reconstruction
- ➪ drastically reduce the background yield on disk
- ➪ replace (the most expensive) parts of detector simulation.