Improving Particle Identification in the Belle II TOP Detector using Machine Learning

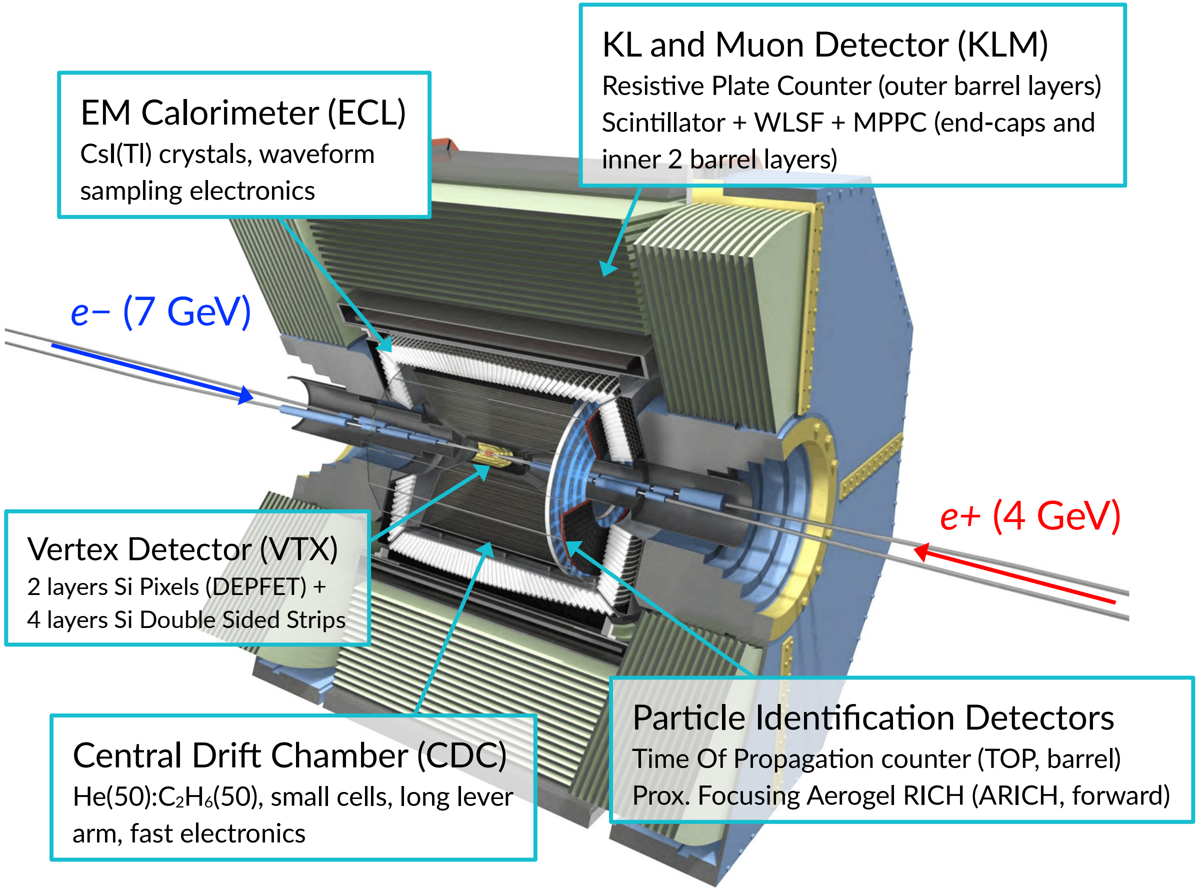
Cecilia Antonioli

on behalf of the Belle II TOP group

ML4Jet 2026

September 14 – 18, 2026, Vienna

The Belle II Detector



- Designed to collect data at the asymmetric *e+e−* collider SuperKEKB (Tsukuba, Japan)
- General purpose spectrometer
- Every sub-detector provides particle identification (PID) information:
  - dE/dx from CDC and VTX
  - Cherenkov signal from TOP and ARICH
  - Shower shapes from ECL
  - Penetration depth from KLM
- Six types of stable charged particles: e, μ, π, K, p, d

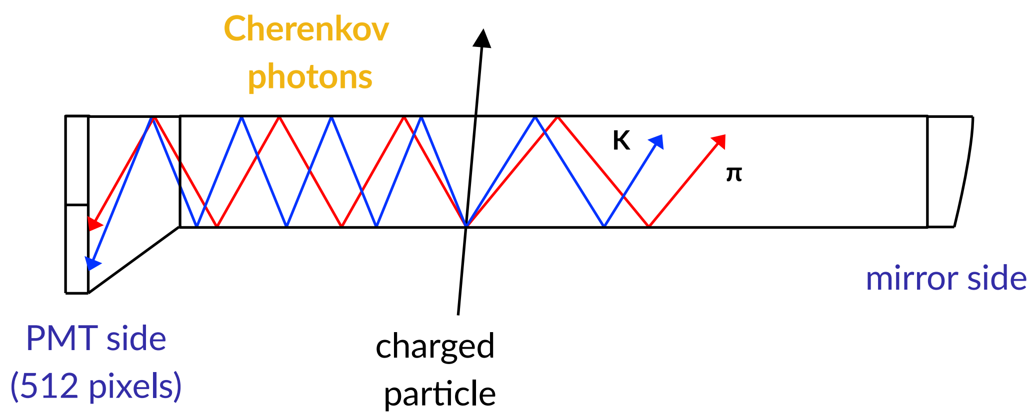
The Time Of Propagation Counter

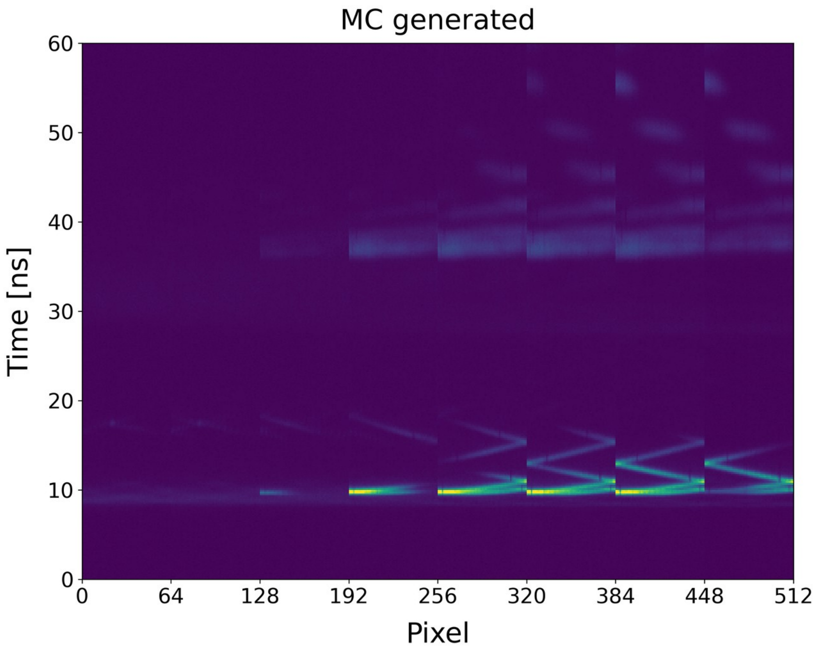
- The TOP counter is the main hadron ID device in the Belle II barrel region
- Principle of operation
  - Cherenkov photons propagate within a quartz bar (radiator) via total internal reflection
  - The 2D information of the Cherenkov ring-image is given by the **arrival time** and **hit position** of photons at the photon sensors

Cherenkov angle

$$\cos\theta_C = \frac{1}{n\beta}$$

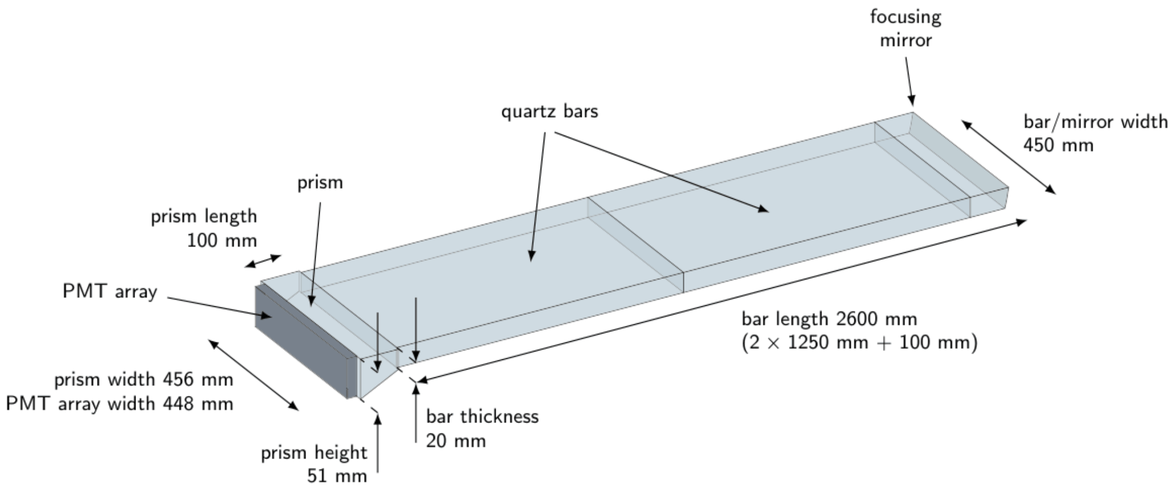
Dependence on particle mass and momentum



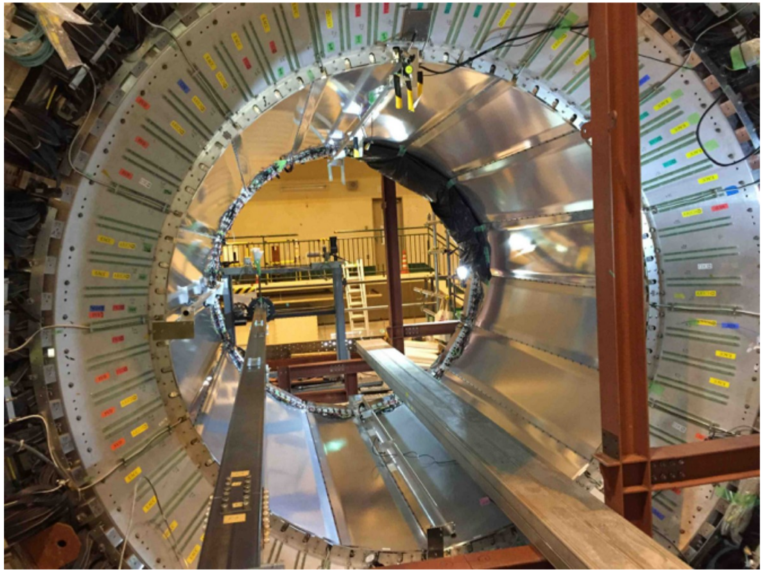


TOP Design Details

- 16 modules arranged around the interaction point (IP)
- Each module is made of two identical bars of quartz glued together (more than 2 m long, 2 cm × 45 cm in cross-section)
- Backward side: expansion prism, 32 PMTs and readout electronics
- Forward side: spherical mirror



[NIM A 1080 (2025) 170627]



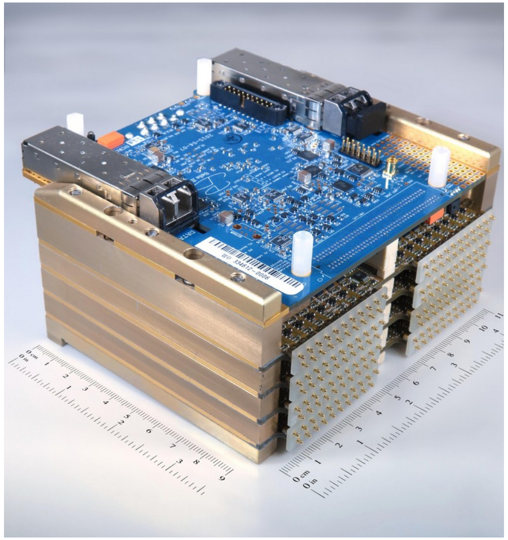
Installed TOP detector modules as viewed from the forward end of the Belle II barrel region

PMTs and Readout Electronics

- Hamamatsu Micro-Channel-Plate (MCP) PMTs
  - 4 × 4 anode channels, 5.5 mm pixel size
  - Single photon sensitivity
  - Excellent time resolution (< 100 ps)
  - Works in magnetic field
- 2 rows of 16 PMTs each per module (512 pixels)

- Readout electronics
  - 8 channel waveform digitizer (IRSX ASIC)
    - 2.7 GSa/s and 12 μs long storage ring buffer
  - 16 ASICs per boardstack (128 channels)
  - 4 boardstacks per module (512 channels)

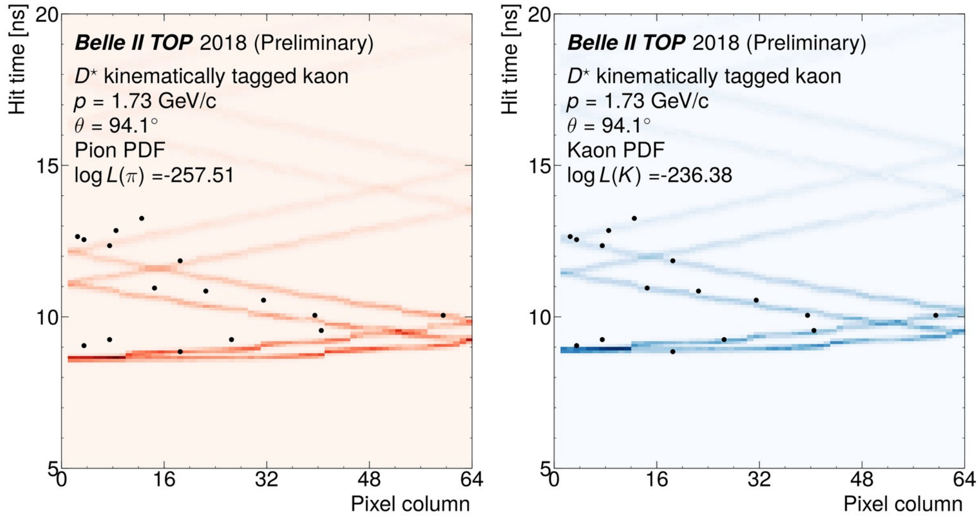
16-pixel Hamamatsu R10754-07-M16(N) MCP-PMT



A boardstack

Current TOP Reconstruction Software

- **Different particle species** (with same momentum, impact position, and impact angle on the TOP module) give rise to **distinct photon patterns**
- We compare event-by-event the **measured photon pattern** to **analytical PDFs** calculated for each particle hypothesis and for the track parameters measured by the tracking system
  - We compute a TOP likelihood value for each particle hypothesis
- The ratios of the TOP likelihoods are used to assign TOP PID probabilities



Space–time distribution of the photon hits associated to a kaon candidate track compared with the expected PDF for the pion hypothesis (left) and for the kaon hypothesis (right)

Global TOP Likelihood Ratio (LR)

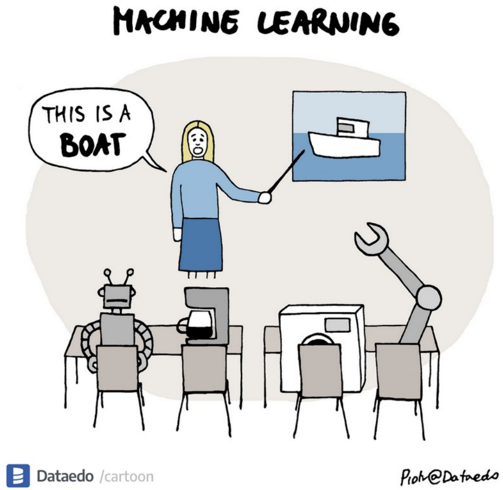
$$P_h = \frac{\mathcal{L}_h}{\sum_\gamma \mathcal{L}_\gamma}$$

Binary TOP LR for K/π separation

$$\mathcal{R} = \frac{\mathcal{L}_K}{\mathcal{L}_K + \mathcal{L}_\pi}$$

TOP PID with Machine Learning

- The analytical PDF-based approach requires detailed knowledge of the geometry and optical properties of all TOP components
- Machine learning (ML) techniques can offer an alternative data-driven strategy for PID in the TOP detector
- **Convolution neural networks (CNNs) are suitable for pattern recognition tasks**
  - CNNs apply convolutional filters to scan input images and extract local sub-features
  - CNNs can analyze images representing TOP photon patterns to obtain the PID information encoded within them



https://dataedo.com

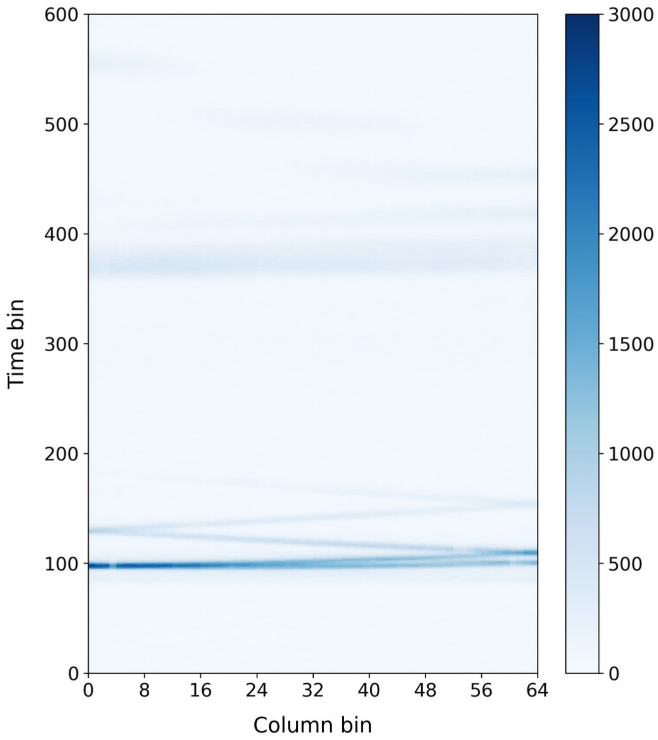
K/π separation using a Convolutional Neural Network

Photon Pattern

- Photons (**Cherenkov + background**) detected in the TOP module hit by the particle
- Photons within the time window **[0.0, 60.0] ns** (calibrated times)
- **Time resolution of 100 ps**
- PMT pixels arranged in **64 columns** (and 8 rows)
- **Image size of (64 x 600)**
  - Very large
  - Information is rare and sparse

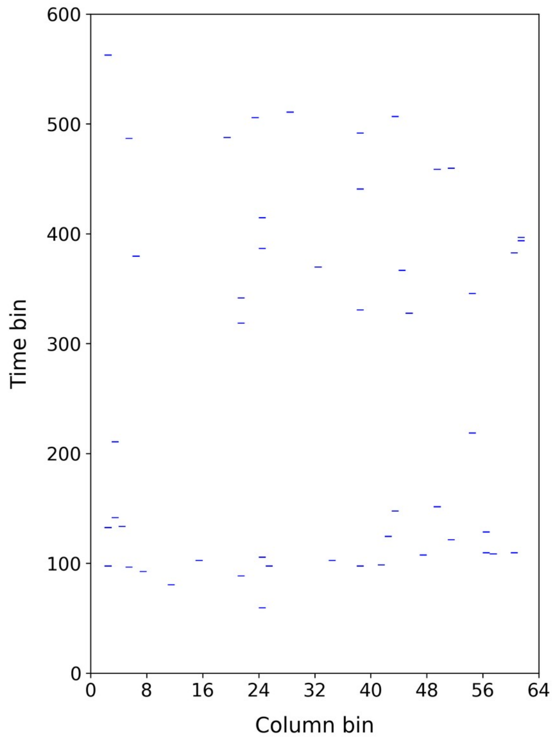
Example of Cherenkov patter for **K+** (MC)

- p in [1.75, 2.25] GeV/c
- θ in [88.0, 92.0] °
- x_{TOP}* in [−4.5, 4.5] cm



Example of image for **K+** (MC)

- p = 2.06 GeV/c
- θ = 89.73°
- φ = 19.76° (TOP mod. 1)



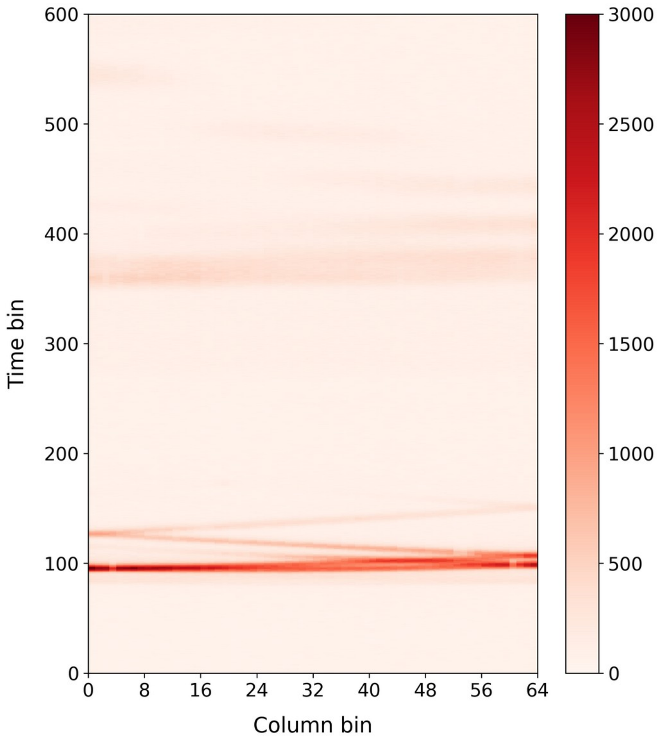
* x_{TOP} is the track impact position expressed in the local reference frame, related to the azimuthal angle φ through geometrical transformatios

Photon Pattern

- Photons (**Cherenkov + background**) detected in the TOP module hit by the particle
- Photons within the time window **[0.0, 60.0] ns** (calibrated times)
- **Time resolution of 100 ps**
- PMT pixels arranged in **64 columns** (and 8 rows)
- **Image size of (64 x 600)**
  - Very large
  - Information is rare and sparse

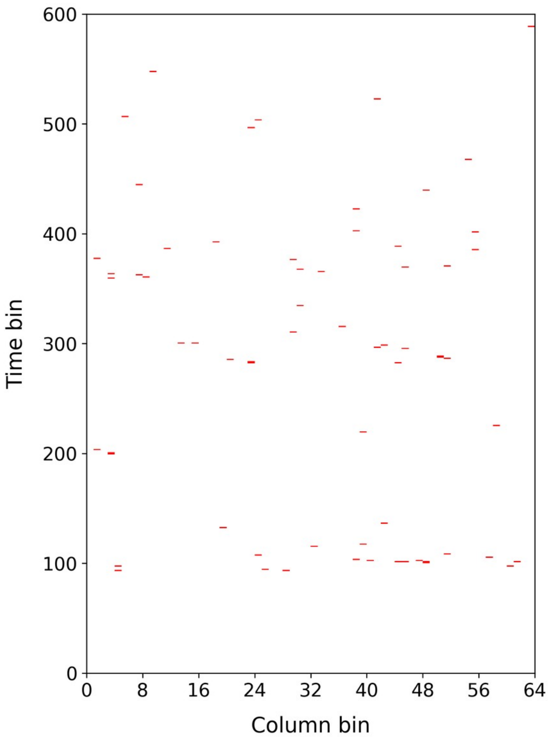
Example of Cherenkov patter for **π+** (MC)

- p in [1.75, 2.25] GeV/c
- θ in [88.0, 92.0] °
- x_{TOP}* in [−4.5, 4.5] cm



Example of image for **π+** (MC)

- p = 2.02 GeV/c
- θ = 91.08°
- φ = 108.80° (TOP mod. 5)

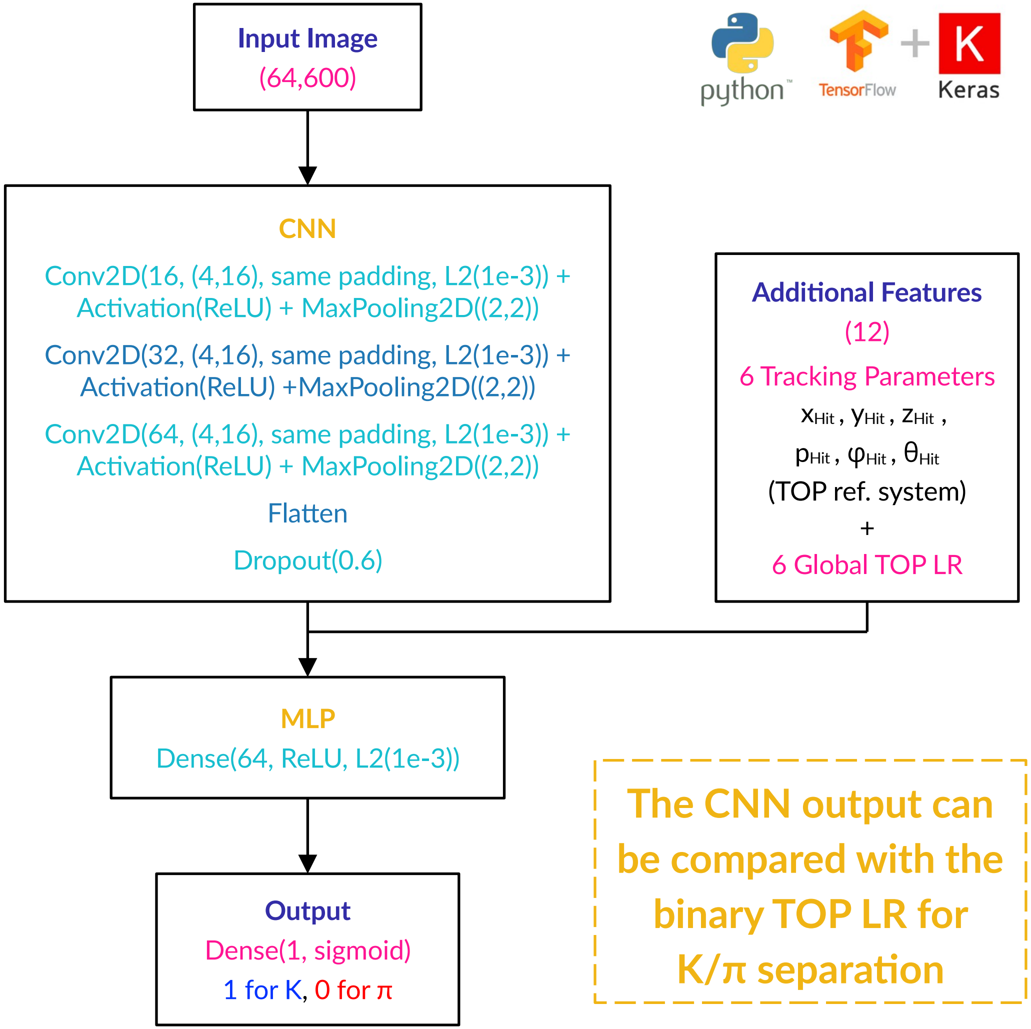


* x_{TOP} is the track impact position expressed in the local reference frame, related to the azimuthal angle φ through geometrical transformatios

CNN Architecture

- Training Configuration
  - Loss: Binary Crossentropy
  - Optimizer: Adam (learning rate = 5e-4)
  - Metrics: Binary Accuracy, AUC
  - Adaptive learning rate strategy

- Training Parameters
  - Epochs: 100
  - Batch size: 128
  - Hardware:
    - 1× NVIDIA Tesla V100 32GB (CloudVeneto)
    - 4× NVIDIA A100 40GB (NERSC)



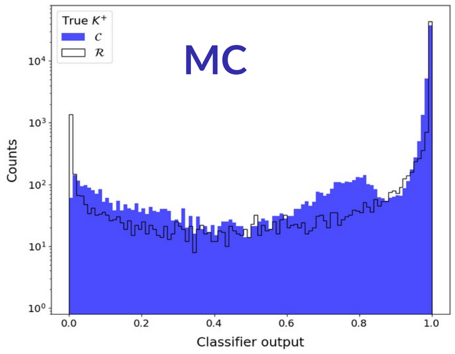
Total parameters: 2,623,473 (10.01 MB)

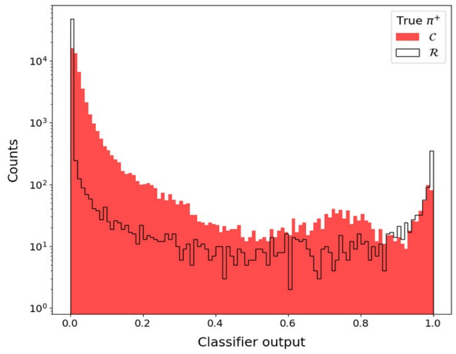
The CNN output can be compared with the binary TOP LR for K/π separation

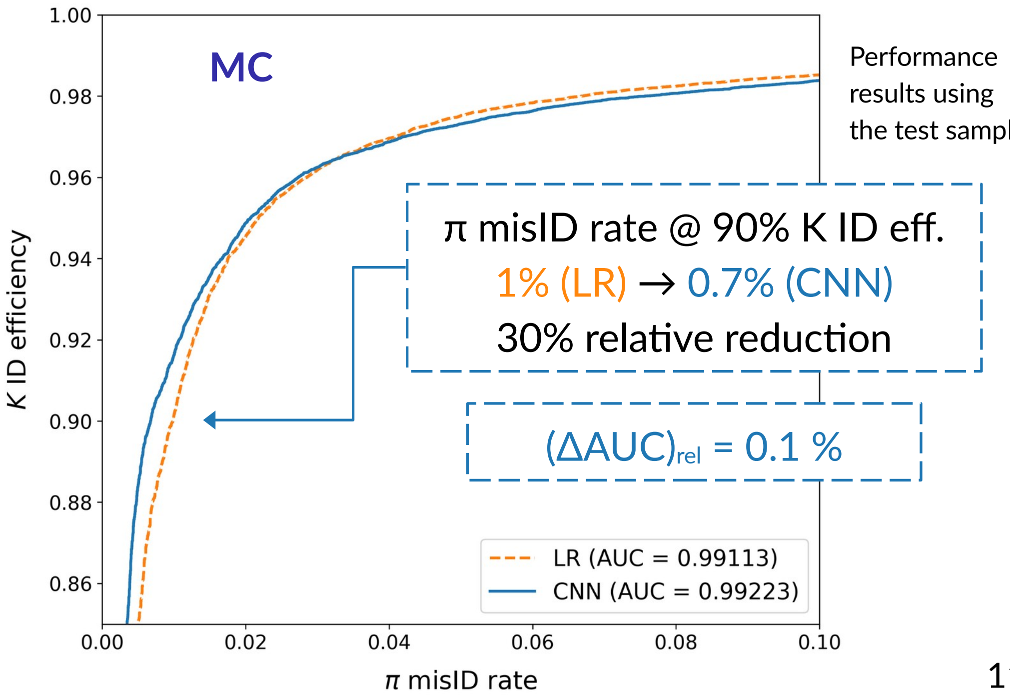
TOP PID Performance with CNN

- Training a CNN in a **limited region of phase space** covered by the TOP detector (baseline bin)
  - p in [1.75, 2.25] GeV/c
  - θ in [88.0, 92.0] °
  - x_{TOP} in [−4.5, 4.5] cm
  - q > 0
- Train sample: **200,000 MC tracks** (100,000 K + 100,000 π)
- 90:10 Training-to-Validation ratio
- Test sample: **100,000 MC tracks** (50,000 K + 50,000 π)

$$(\Delta\mathrm{AUC})_{\mathrm{rel}} = \frac{\mathrm{AUC}_{\mathrm{CNN}} - \mathrm{AUC}_{\mathrm{LR}}}{\mathrm{AUC}_{\mathrm{LR}}} \times 100$$







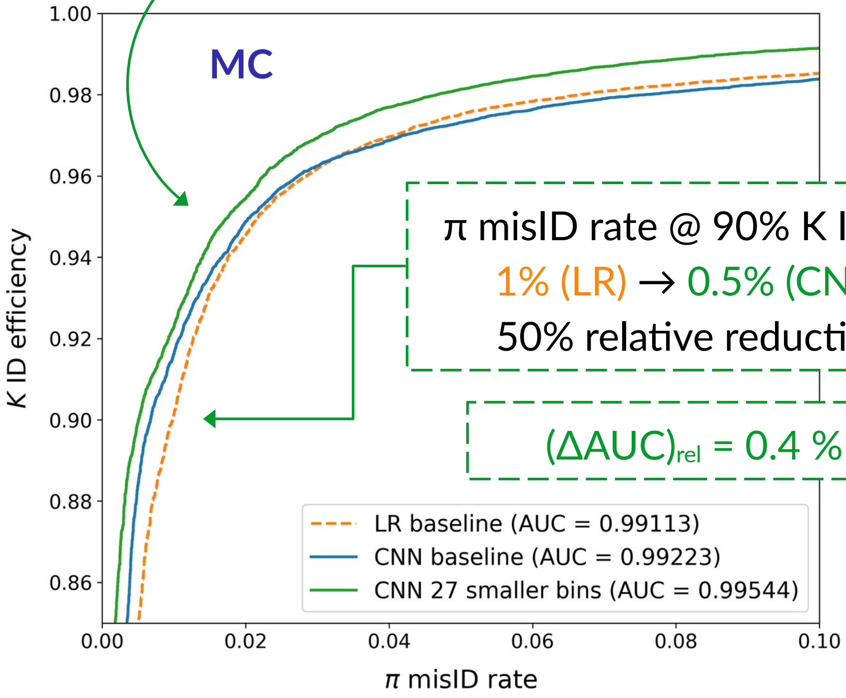
Performance results using the test sample

TOP PID Performance with CNN (finer binning)

- Training independent CNNs by dividing the baseline region into **27 smaller bins**
  - p in [1.75, 1.90], [1.90, 2.10], [2.10, 2.25] GeV/c
  - θ in [88.0, 89.0], [89.0, 91.0], [91.0, 92.0] °
  - x_{TOP} in [−4.5, −1.0], [−1.0, 1.0], [1.0, 4.5] cm
  - q > 0
- Train sample for each bin: **200,000 MC tracks** (100,000 K + 100,000 π)
- 90:10 Training-to-Validation ratio
- Test sample (baseline): **100,000 MC tracks** (50,000 K + 50,000 π)

The finer the binning, the greater the improvement in PID performance with CNNs

ROC curve obtained by merging the performance results of the CNNs trained individually in each bin



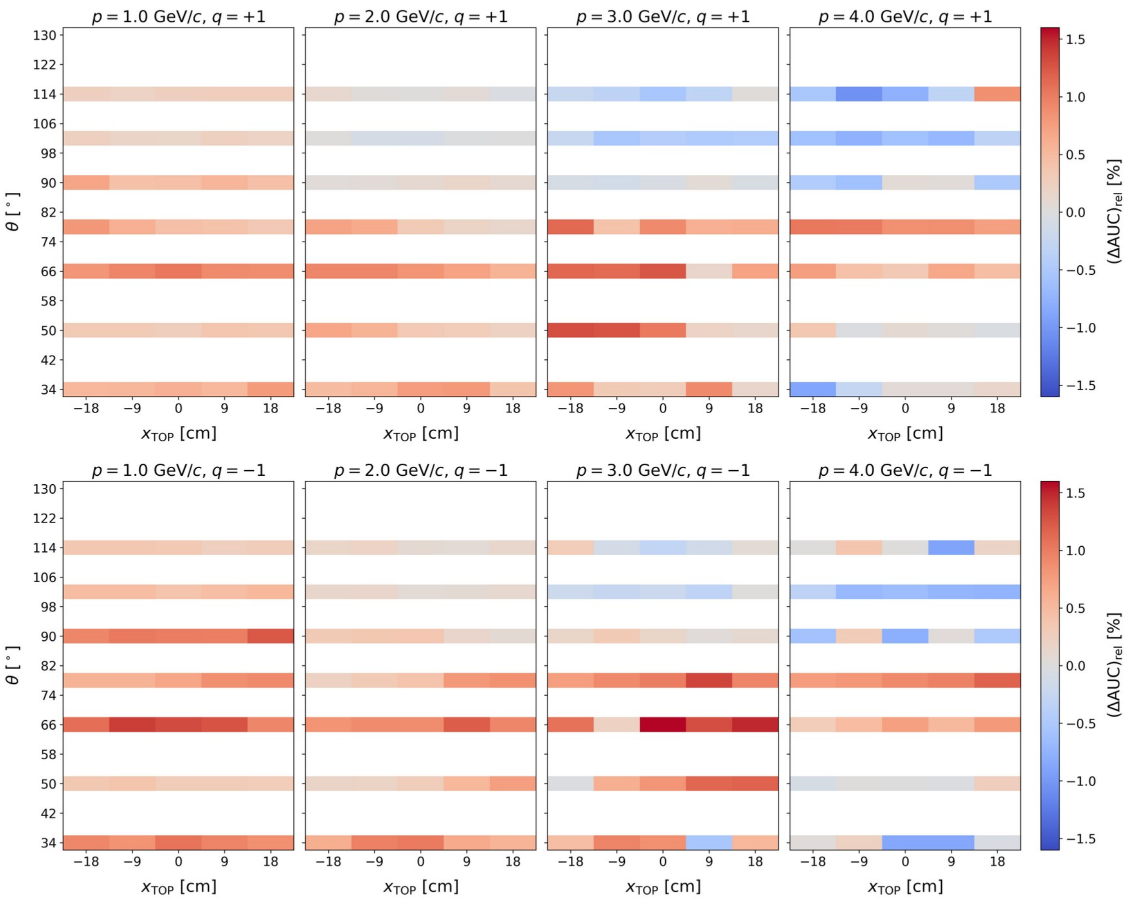
Performance results using the test sample

π misID rate @ 90% K ID eff.
1% (LR) → 0.5% (CNN)
50% relative reduction

$(\Delta AUC)_{rel}$ = 0.4 %

TOP PID Performance with CNN in the TOP Phase Space

- Training independent CNNs in several bins of the TOP phase space
  - p in [0.3, 4.5] GeV/c
  - θ in [31.0, 128.0] °
  - x_{TOP} in [−22.5, 22.5] cm
  - q > 0, q < 0
- Each bin has the same size as the baseline bin
- Train and test sample sizes of **MC tracks** as in the baseline configuration

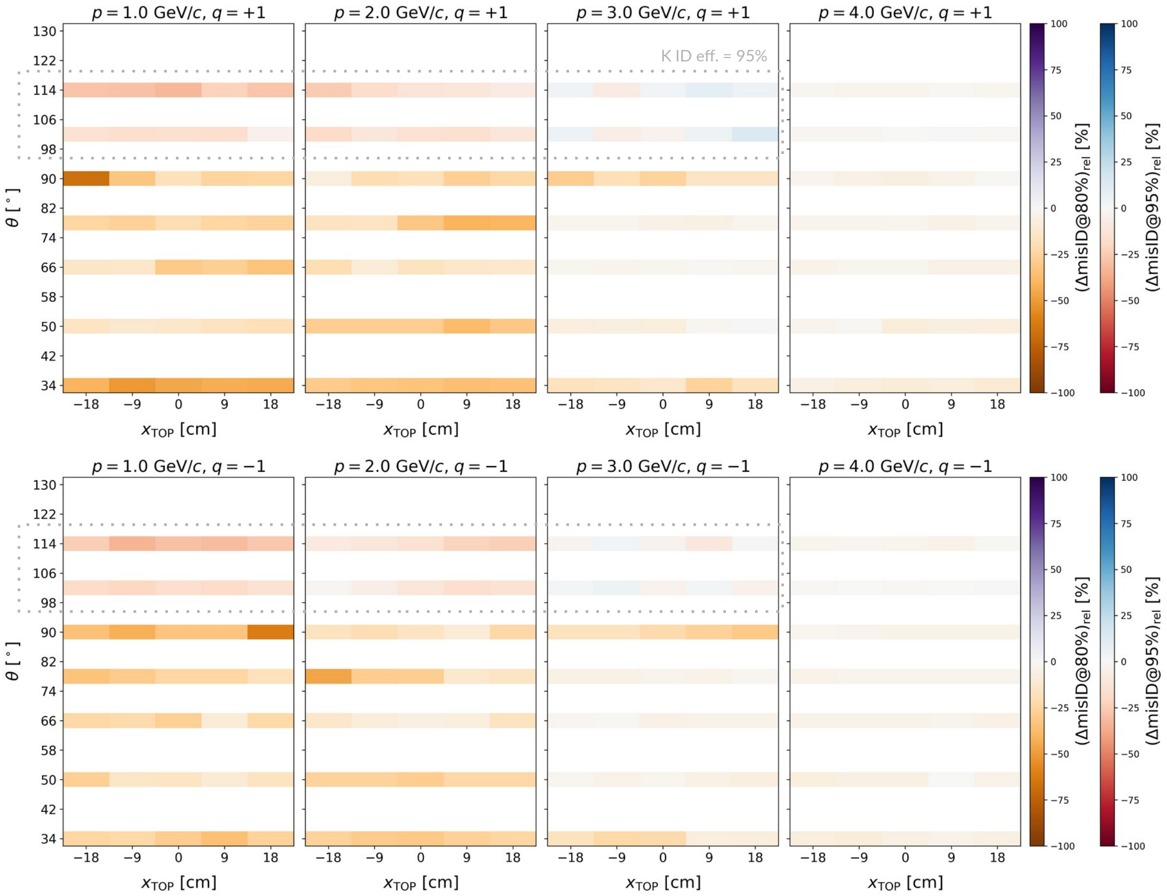


Only in some bins the CNN provides a gain, $(\Delta AUC)_{rel}$ > 0 …

TOP PID Performance with CNN in the TOP Phase Space

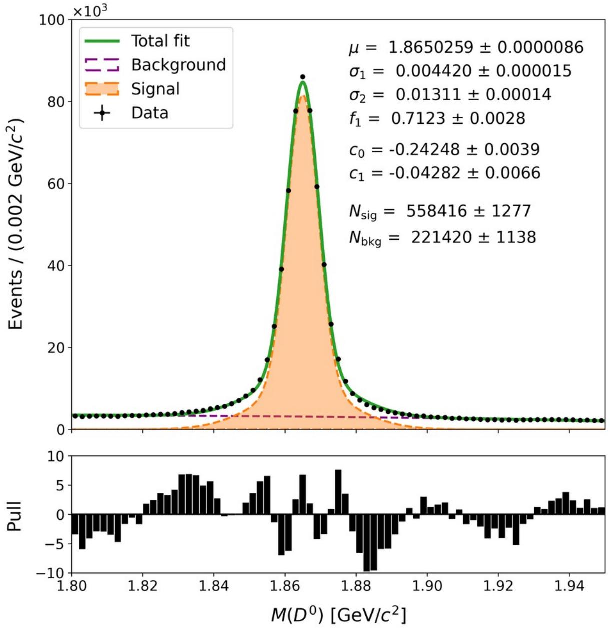
- Training independent CNNs in several bins of the TOP phase space
  - p in [0.3, 4.5] GeV/c
  - θ in [31.0, 128.0] °
  - x_{TOP} in [−22.5, 22.5] cm
  - q > 0, q < 0
- Each bin has the same size as the baseline bin
- Train and test sample sizes of **MC tracks** as in the baseline configuration

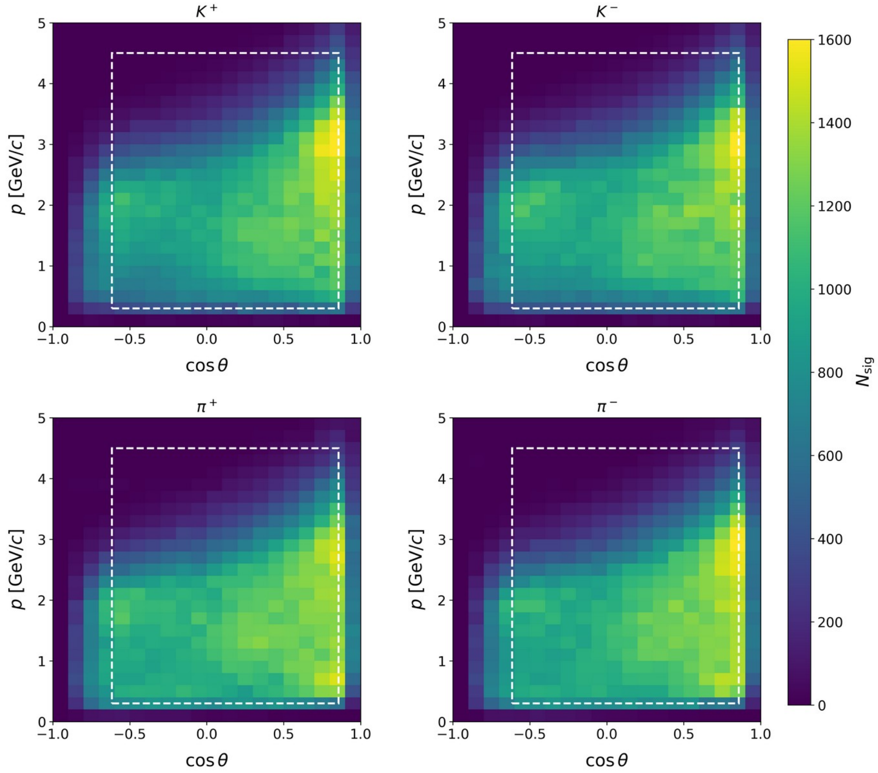
… but the π misID rate @ fixed K ID efficiency either decreases or remains almost unchanged!



Global TOP PID Performance with CNN

- Reconstruction of K and π tracks from $D^{*+} \to D^0 [\to K^- \pi^+] \pi^+$ using a fraction of **Run 1 data** (114 fb^{-1}*)
- Fit to the D^0 invariant mass to extract sWeights
- Selection of tracks within the analyzed bins

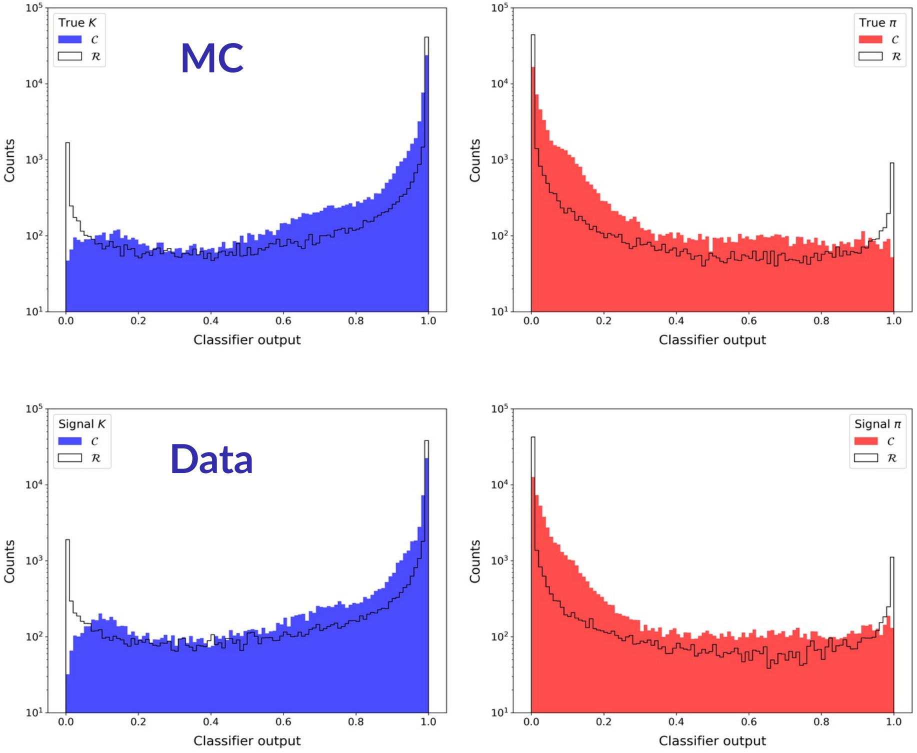




* The full Run 1 dataset (426 fb^{-1}) could not be used, as raw TOP detector information is available only for the subset of data used for the Belle II detector calibration

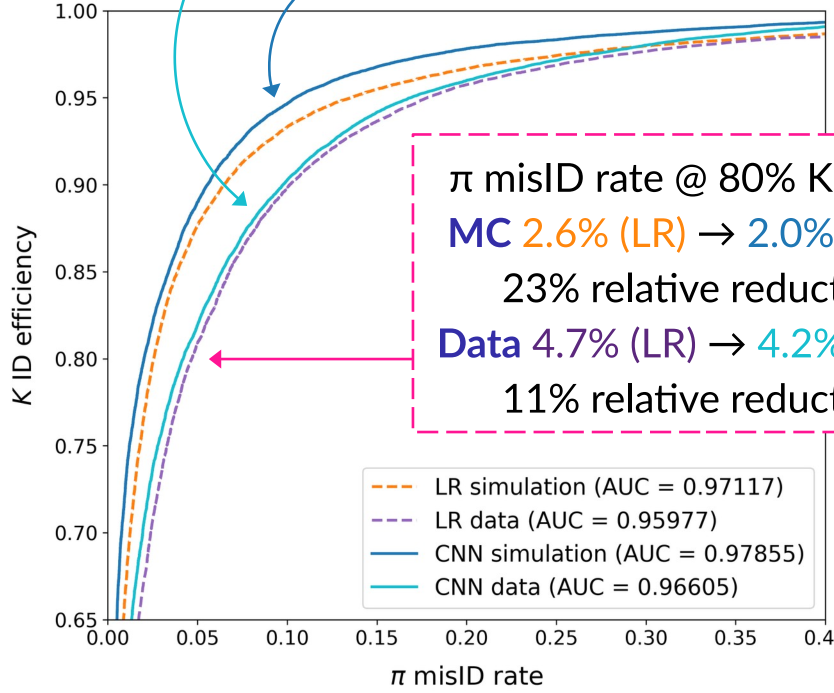
Global TOP PID Performance with CNN

- Combination of the performance results from CNNs trained individually in each bin



Data: performance results for sWeight-selected signal tracks

MC: performance results with track distributions matched to data



π misID rate @ 80% K ID eff.

MC 2.6% (LR) → 2.0% (CNN)

23% relative reduction

Data 4.7% (LR) → 4.2% (CNN)

11% relative reduction

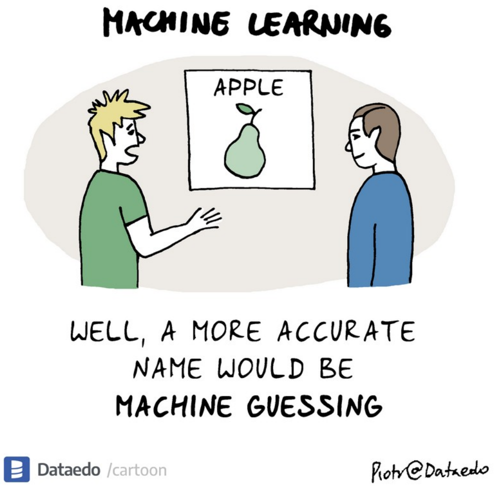
MC $(\Delta \mathrm{AUC})_{\mathrm{rel}}$ = 0.75 %

Data $(\Delta \mathrm{AUC})_{\mathrm{rel}}$ = 0.65 %

Summary

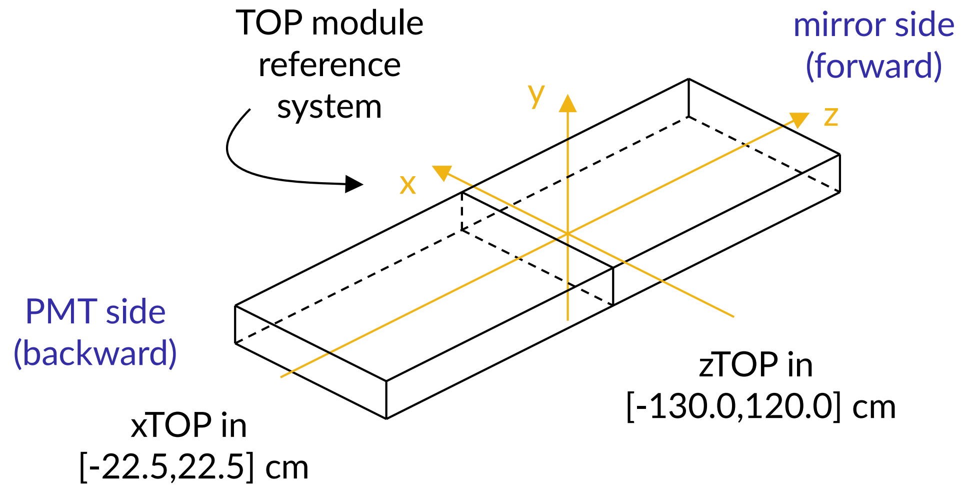
- CNNs trained in TOP **phase space bins** using MC tracks improve TOP PID performance for K/π separation when the results are combined
- **Performance improvements are observed in both simulation and experimental data**
- In MC, the gain is also observed when evaluating individual bins
- The CNN-based approach looks promising, for the future
  - Adaptive phase space binning
  - Optimization of the CNN architecture for each phase space region
- **Large computational resources are required for both MC generation and training campaigns**

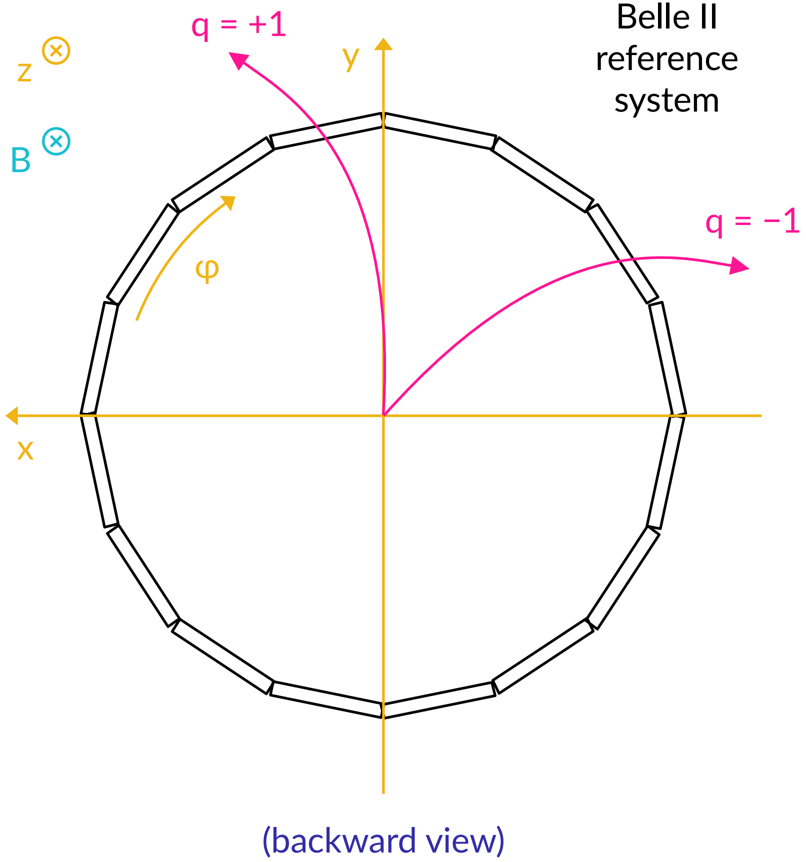
Backup



https://dataedo.com

TOP Module Ref. System and Belle II Ref. System





TOP PID Performance with CNN (no global TOP LR)

- Training a CNN without including the 6 global TOP LR as additional features
  - p in [1.75, 2.25] GeV/c
  - θ in [88.0, 92.0] °
  - xTOP in [−4.5, 4.5] cm
  - q > 0
- Train sample: **200,000 MC tracks** (100,000 K + 100,000 π)
- 90:10 Training-to-Validation ratio
- Test sample: **100,000 MC tracks** (50,000 K + 50,000 π)

The goal is to improve the PID performance compared to the LR