Identification of pions and muons with the Belle II calorimeter using CNN.

DPG-Frühjahrstagung, Heidelberg

Abtin Narimani Charan, Torben Ferber

abtin.narimani.charan@desy.de

23.03.2022

Belle II

HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

Karlsruher Institut für Technologie

DESY.

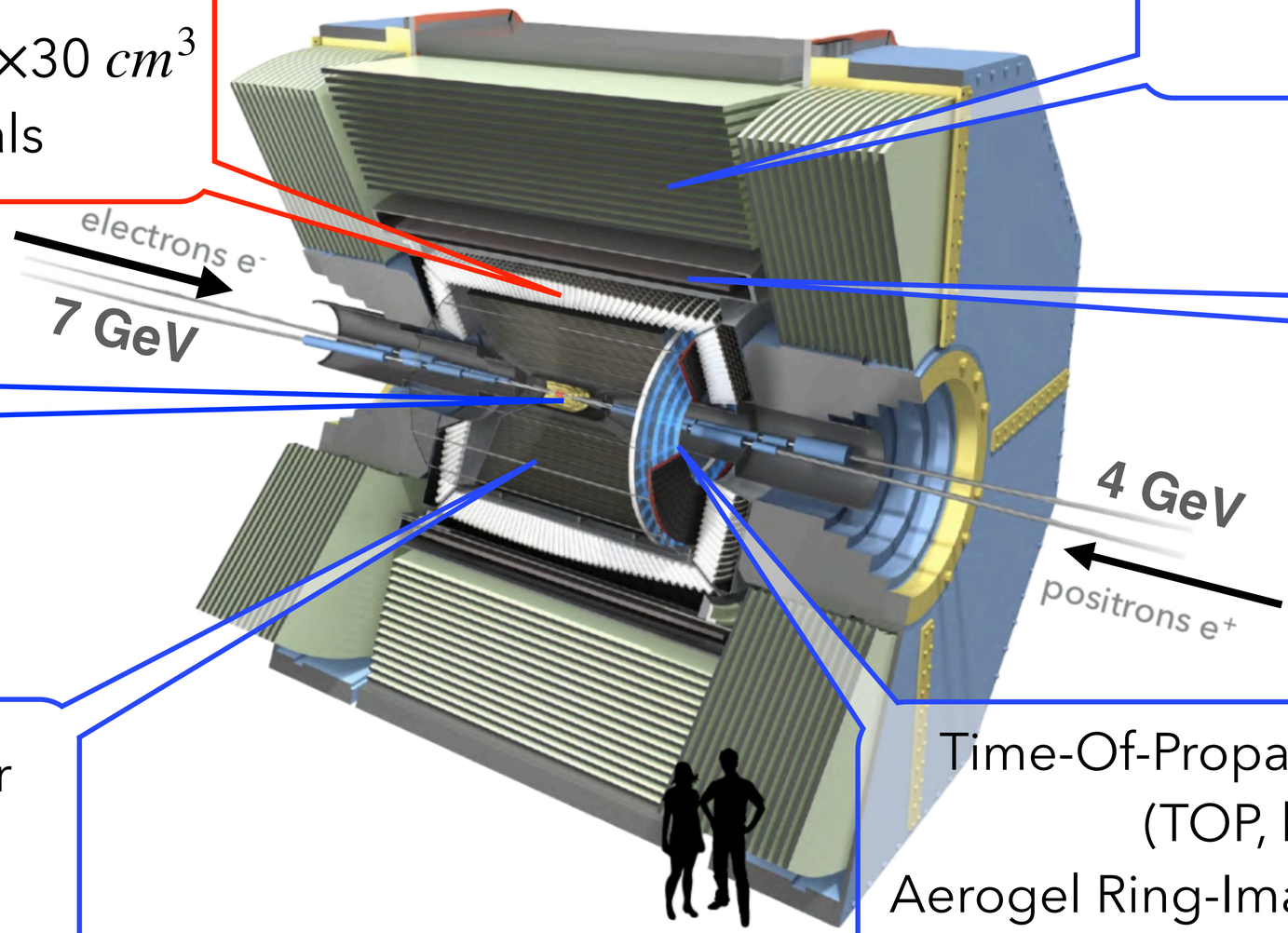
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Belle II & Particle Identification (PID)

$$e^+e^- \rightarrow \Upsilon(4S) \rightarrow B\bar{B}$$

Electromagnetic Calorimeter (ECL): 8736 CsI(Tl) 5×5×30 cm^3 scintillating crystals

K_L and μ detector (KLM)



Vertex Detectors (VXD)

Magnet

Central Drift Chamber (CDC) tracking system

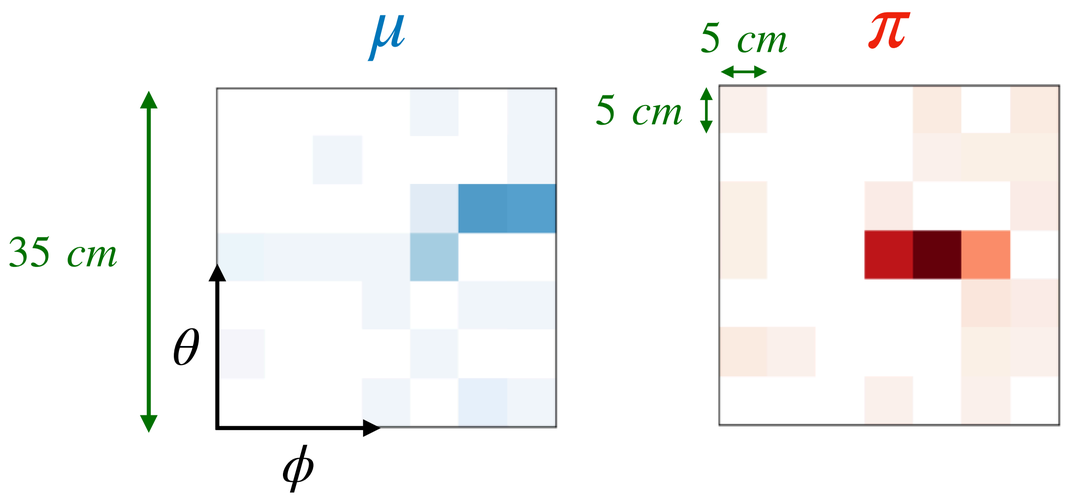
Time-Of-Propagation counter (TOP, barrel)
Aerogel Ring-Imaging Cherenkov counter (ARICH, end-caps)

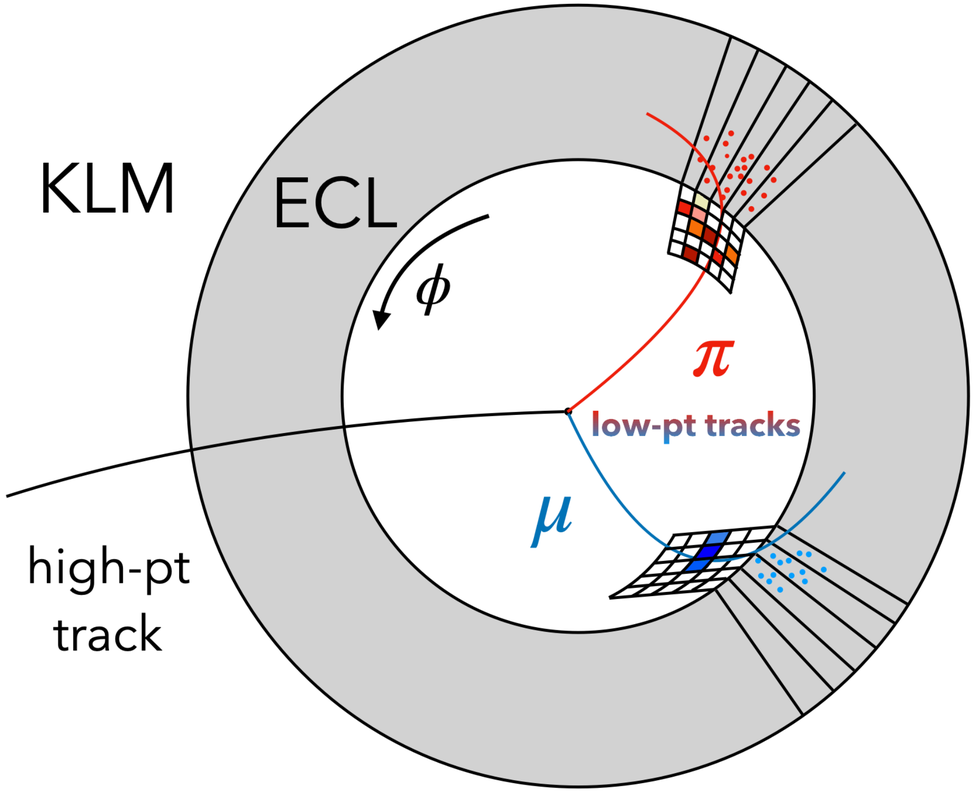
Muon-pion separation

- Identification of low-momenta muons relies on the calorimeter:

  $0.2 \lesssim p \lesssim 0.6$ GeV/c

- Low-momenta muons cannot reach KLM.
- Better muon-pion separation is advantageous for leptonic tau decays.

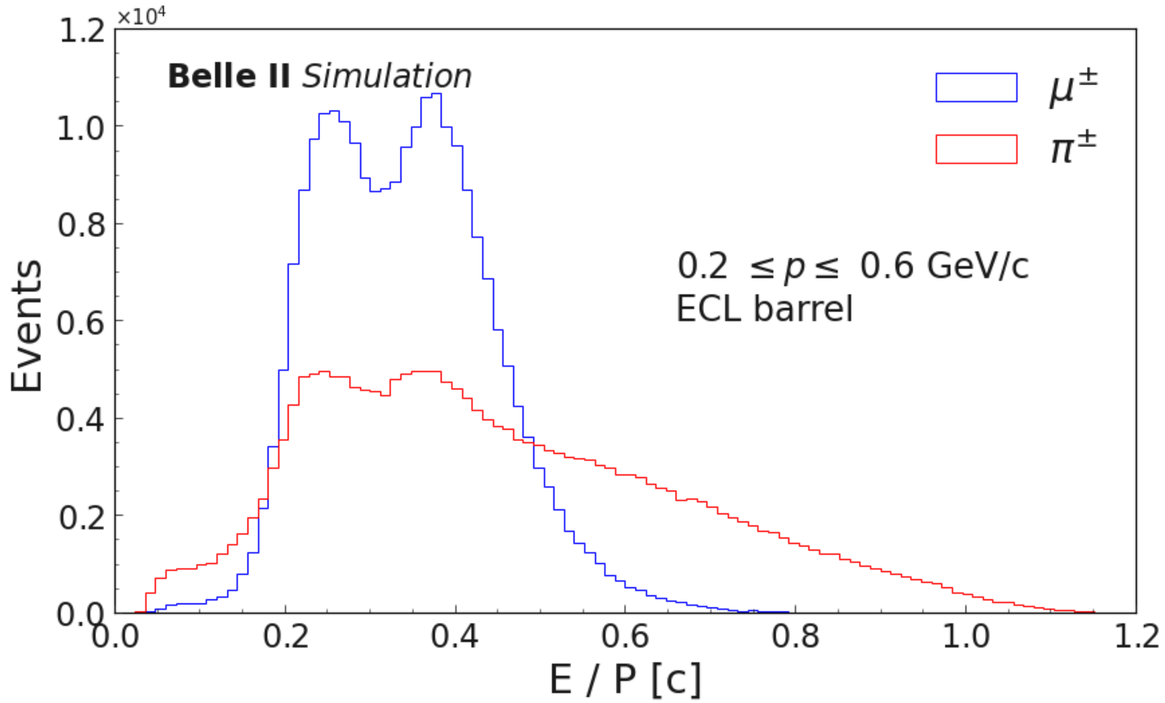


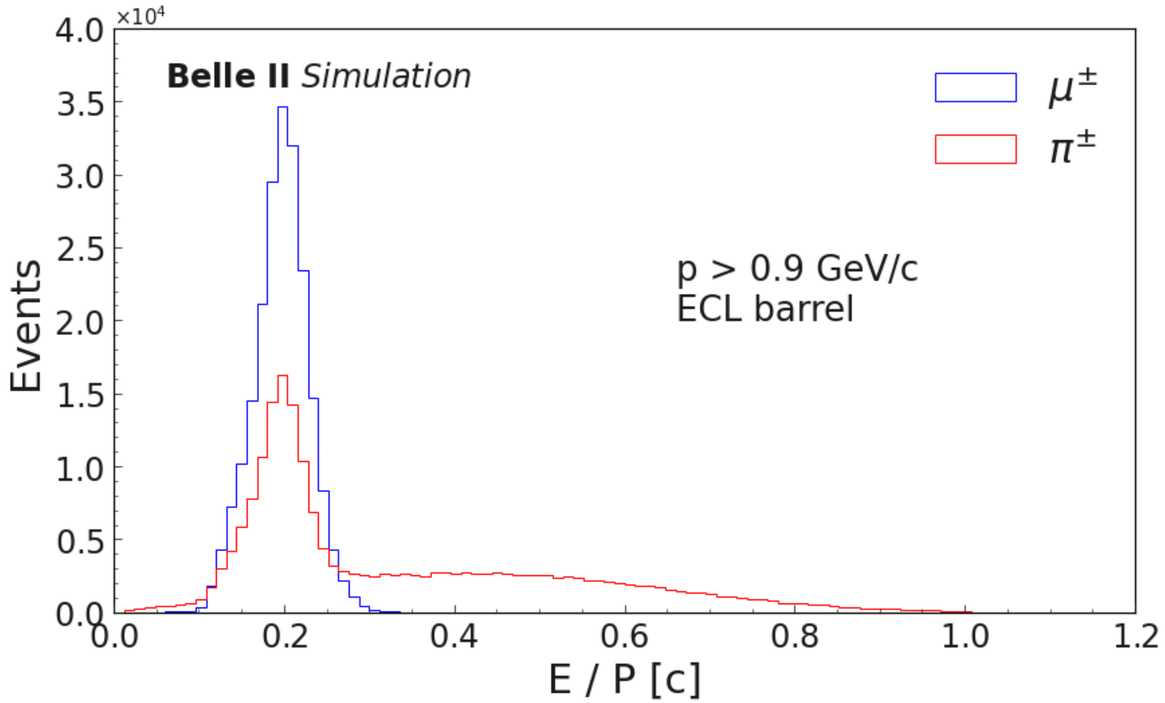


- The energy deposition patterns for pions are more dispersed than muons.
- Convolutional neural network (CNN) is useful to find patterns in crystal images.

Default PID at Belle II

- Default PID in the calorimeter defines a likelihood as a function of E/p.
- E/p is not very powerful for muon-pion separation in low momentum.





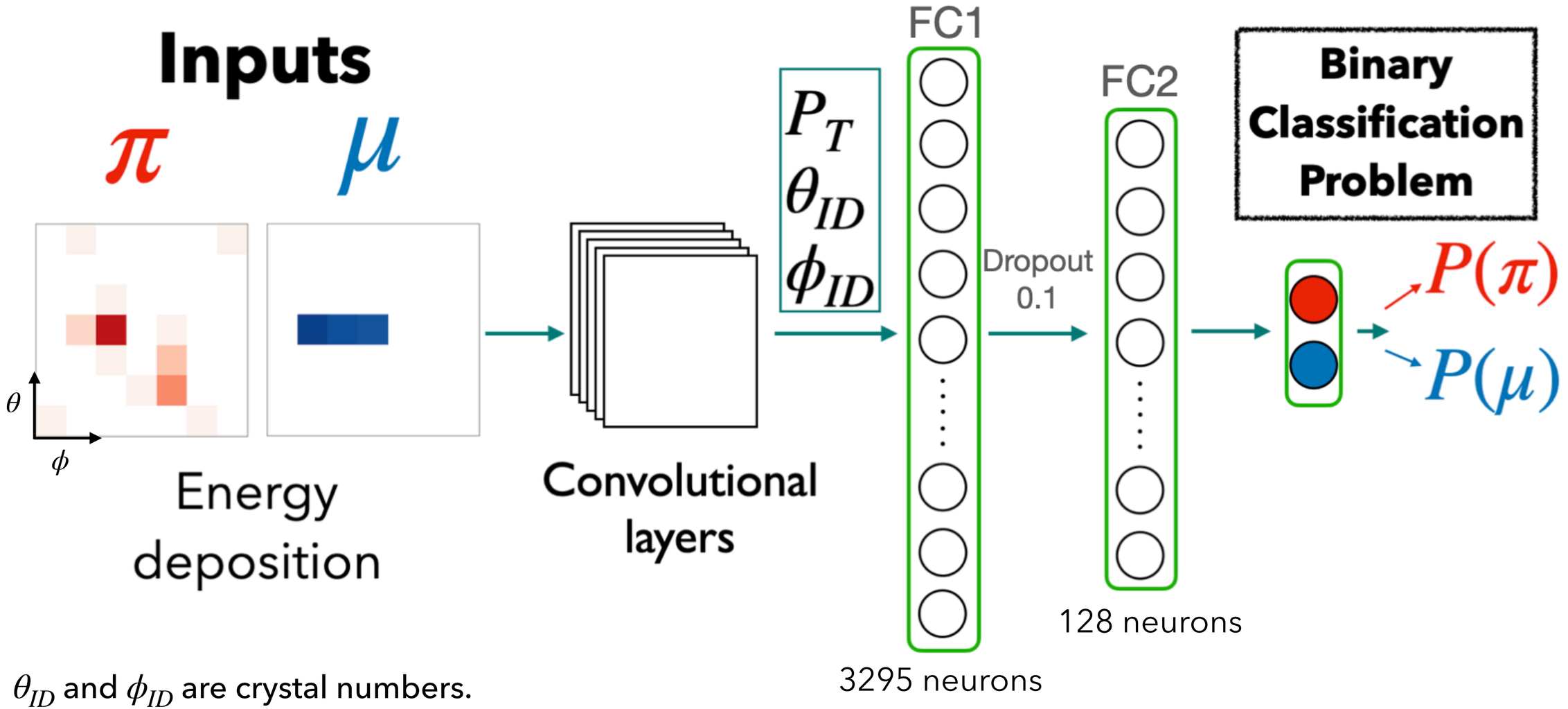
Boosted decision trees (BDT) for PID

- It is based on multivariate classification algorithm.
- BDTs are trained combining measurements from the ECL and the tracking system.
- There are observables that are defined to describe **lateral shower shape** development in the ECL which differs for e, μ, and π.

Variable	Range	Description
E/p [c]	–	Ratio of cluster energy over track momentum.
$E_{cluster}$ [GeV]	–	Cluster energy.
E_1/E_9	–	Ratio of the energy of the seed crystal over the energy sum of the 9 surrounding crystals.
E_9/E_{21}	–	Ratio of the energy sum of 9 crystals surrounding the seed over the energy sum of the 25 surrounding crystals (minus 4 corners).
$\|Z_{40}\|$	–	Zernike moment $n = 4$, $m = 0$, calculated in a plane orthogonal to the EM shower direction.
$\|Z_{51}\|$	–	Zernike moment $n = 5$, $m = 1$, calculated in a plane orthogonal to the EM shower direction.
Z_{MVA}	–	Score of BDT trained on 11 Zernike moments.
ΔL [mm]	–	Projection on the extrapolated track direction of the distance between the track entry point in the ECL and the cluster centroid.
$\Delta \log \mathcal{L}(\ell/\pi)_{CDC}$	–	Log-likelihood difference between $\ell - \pi$ hypothesis in the CDC.
$\Delta \log \mathcal{L}(\ell/\pi)_{TOP}$	ECL barrel	Log-likelihood difference between $\ell - \pi$ hypothesis in the TOP.
$\Delta \log \mathcal{L}(\ell/\pi)_{ARICH}$	ECL FWD endcap	Log-likelihood difference between $\ell - \pi$ hypothesis in the ARICH.
$\Delta \log \mathcal{L}(\mu/\pi)_{KLM}$	$p > 0.6$ GeV/c	Log-likelihood difference between $\mu - \pi$ hypothesis in the KLM.

EPJ Web Conf., 245 (2020) 06023

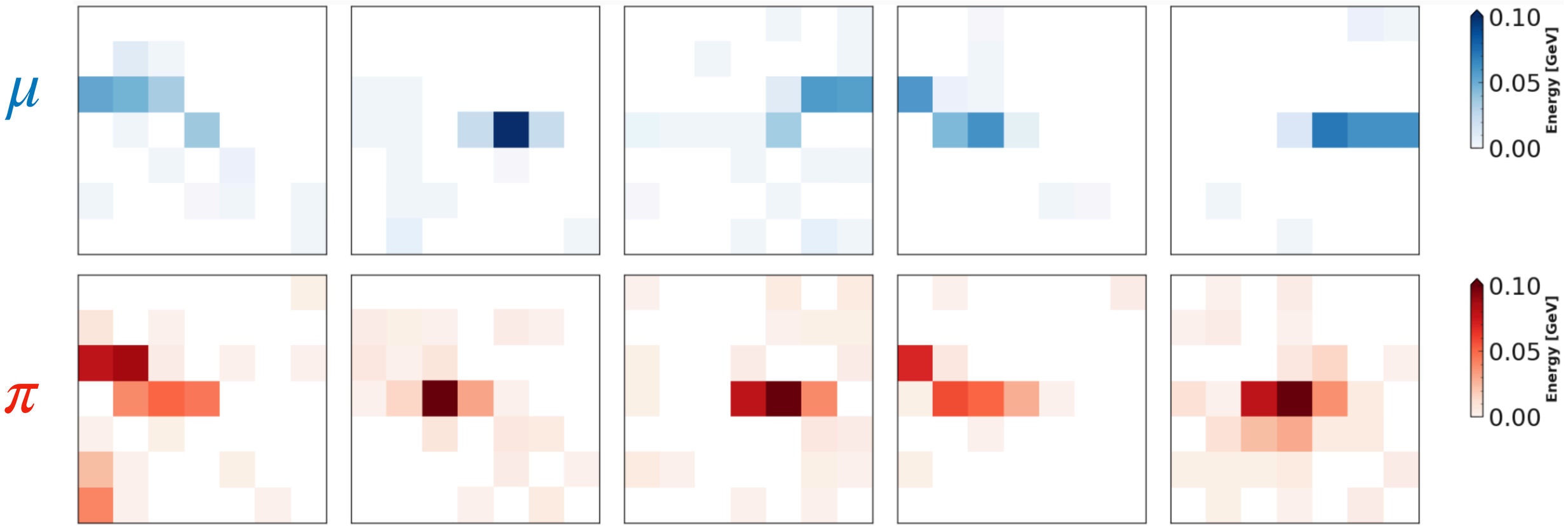
Convolutional Neural Network (CNN)



θ_{ID} and ϕ_{ID} are crystal numbers.

Due to the geometry of the ECL, two separate CNNs are trained with positive and negative charged tracks

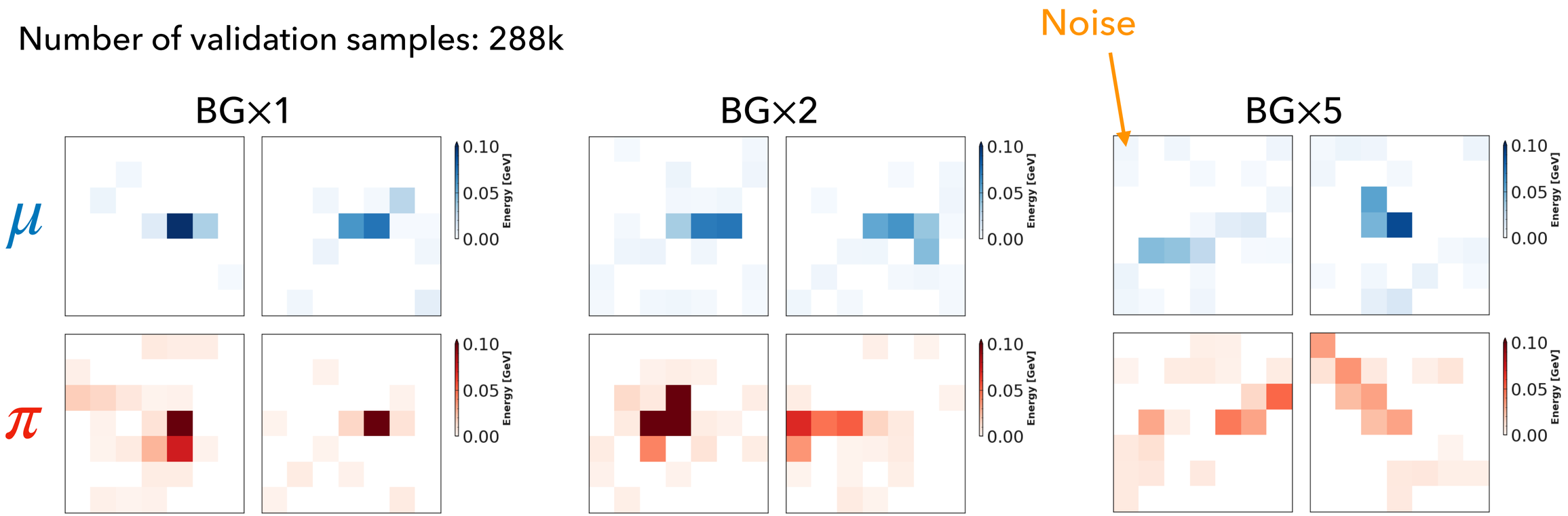
Energy deposition in crystals



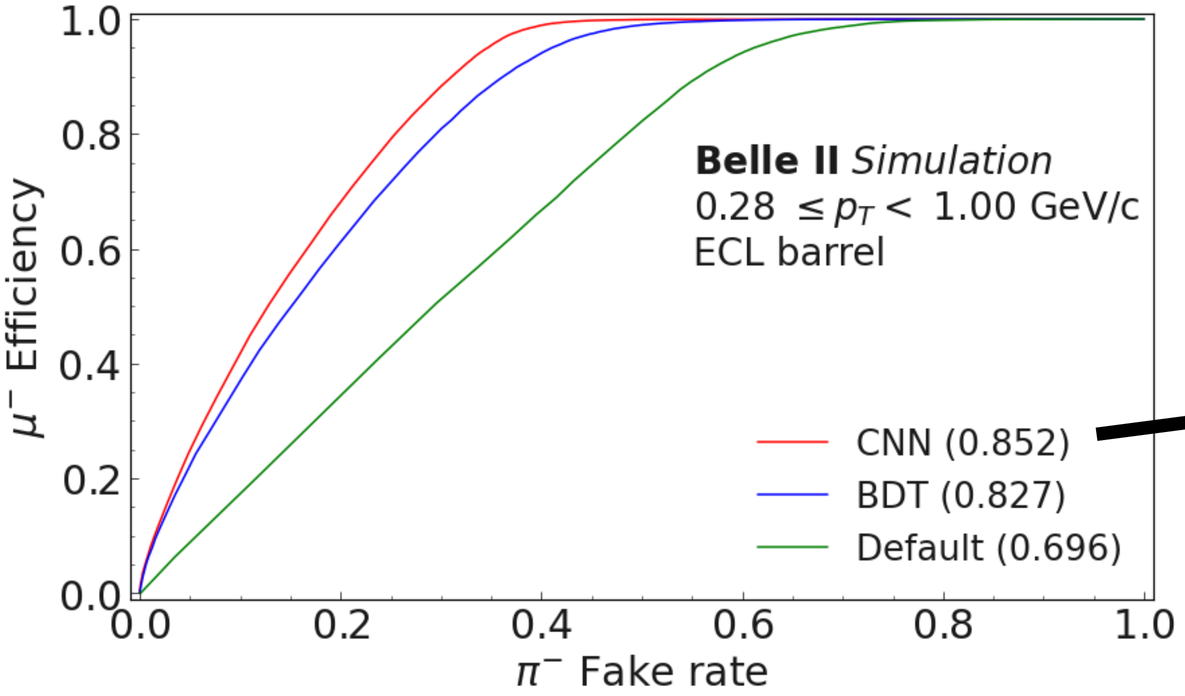
- The central crystal shows the entering point of the track into the ECL barrel.
- Pions undergo ionization and hadronic interactions.

Training samples and inputs

- $\mu^{\pm}$ and $\pi^{\pm}$ samples (1 track per event)
- The samples include different beam background levels or noise (BG×1, BG×2, BG×5)
  - BG×1: nominal beam background
- Number of training samples: 865k
- Number of validation samples: 288k



CNN - different beam background levels

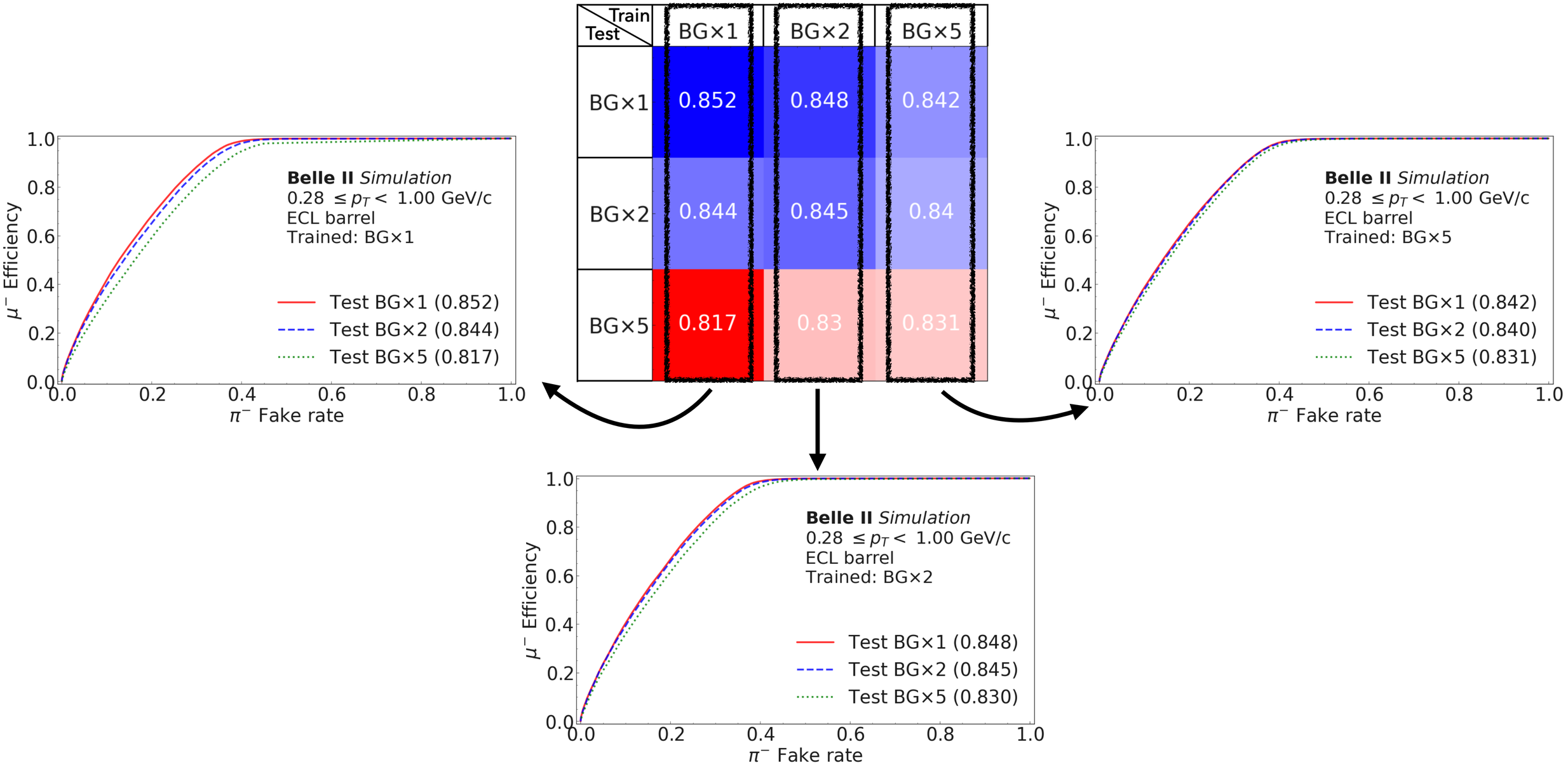


Train \ Test	BG×1	BG×2	BG×5
BG×1	0.852	0.848	0.842
BG×2	0.844	0.845	0.84
BG×5	0.817	0.83	0.831

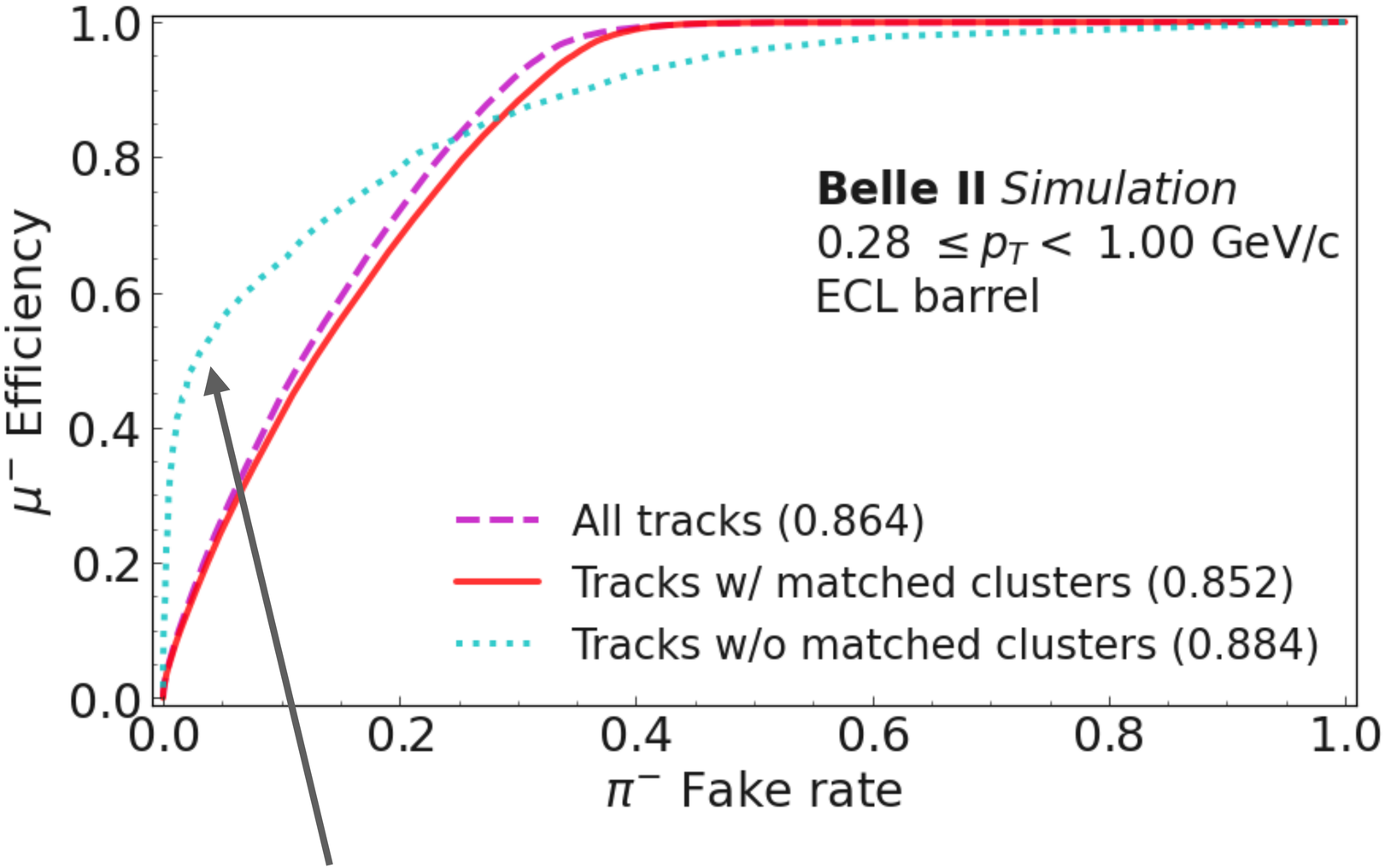
$$\mu \text{ Efficiency} = \frac{\text{Number of muons identified as muons}}{\text{Total number of muons}}$$

$$\pi \text{ fake rate} = \frac{\text{Number of pions identified as muons}}{\text{Total number of pions}}$$

CNN - different beam background levels



CNN comparison with tracks that have no cluster

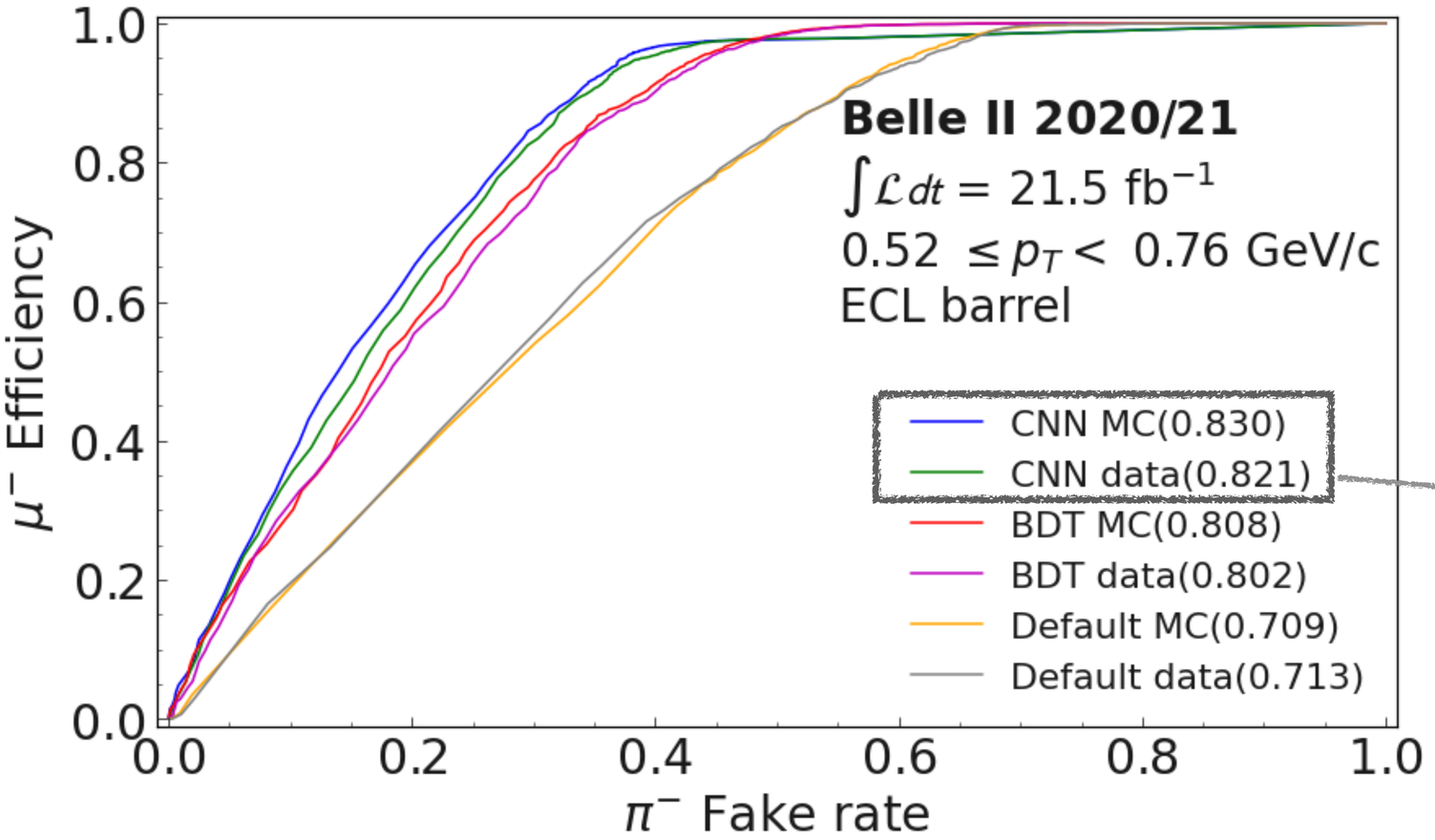


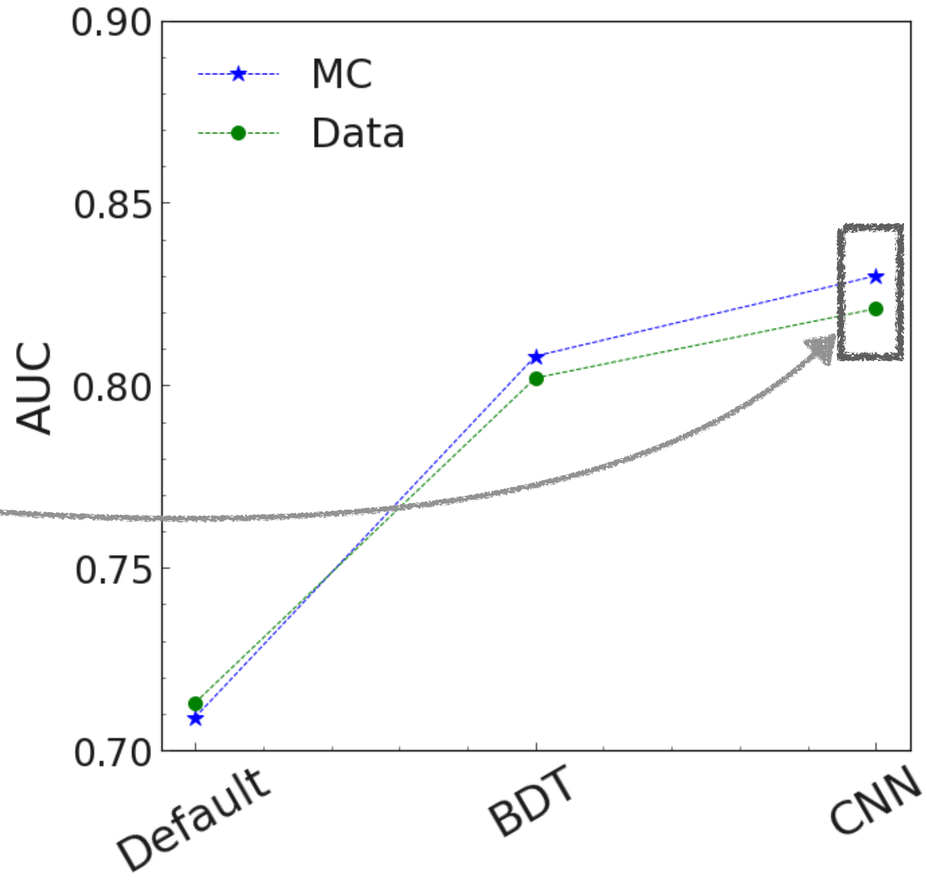
About **5.5%** of tracks do not have a matched cluster.

Test on data

- $e^+e^- \rightarrow \mu^+\mu^-\gamma$
- $D^* \rightarrow D^0\,[\rightarrow K\,\pi]\,\pi$

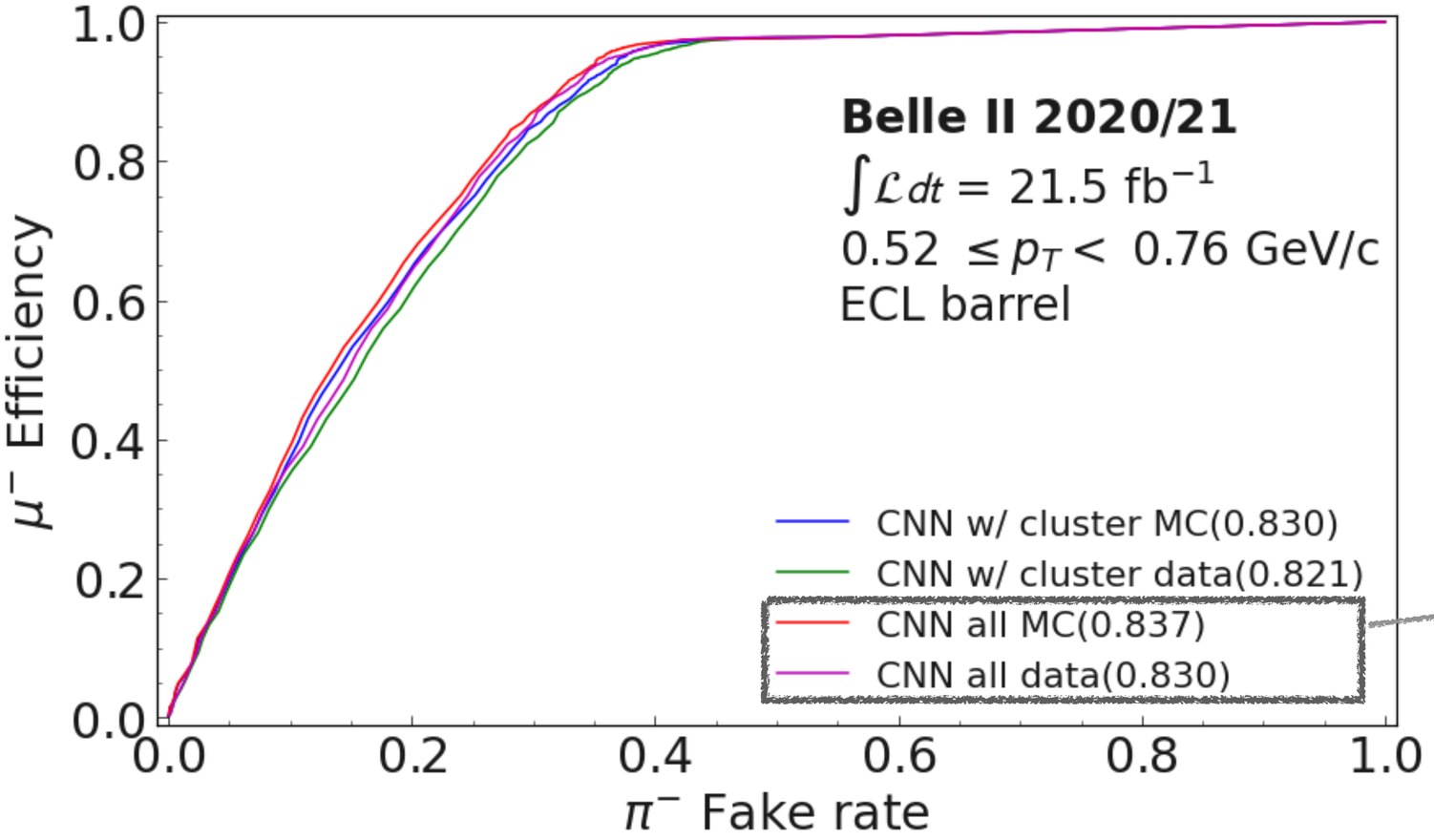
CNN comparison with different PID methods

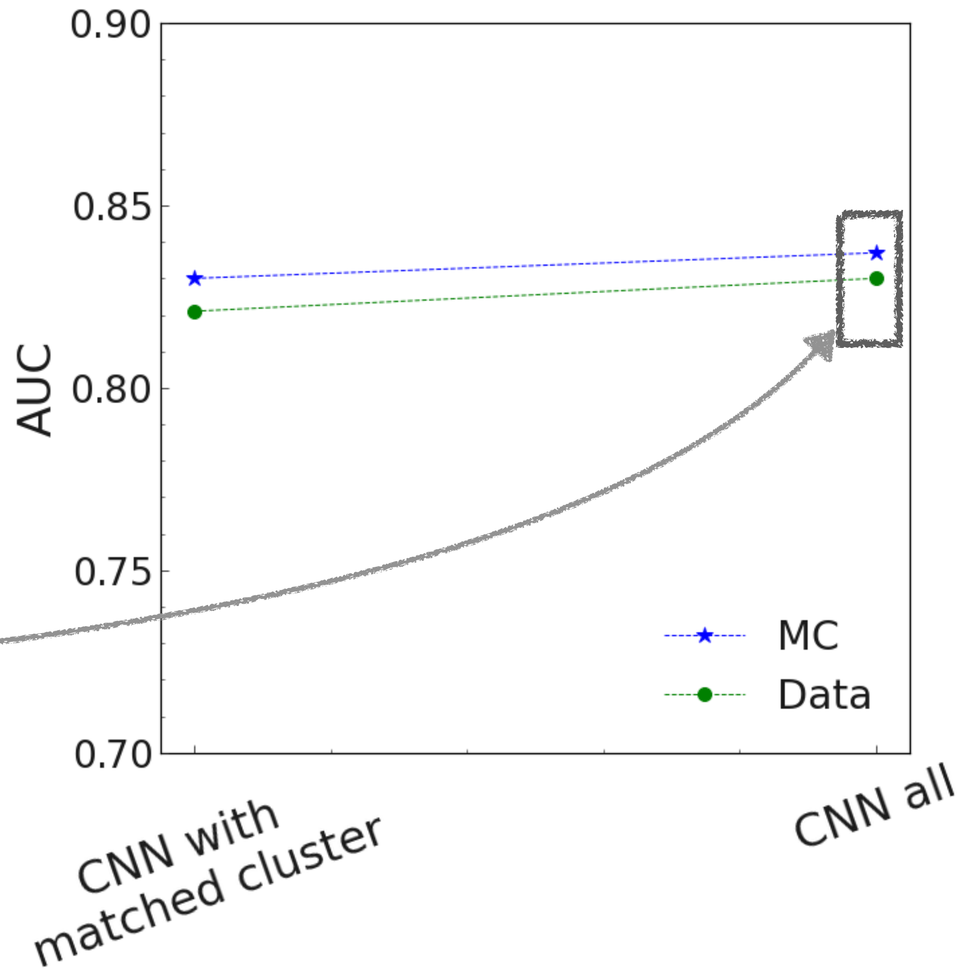




- Model is trained with BG×1.

CNN comparison with all tracks

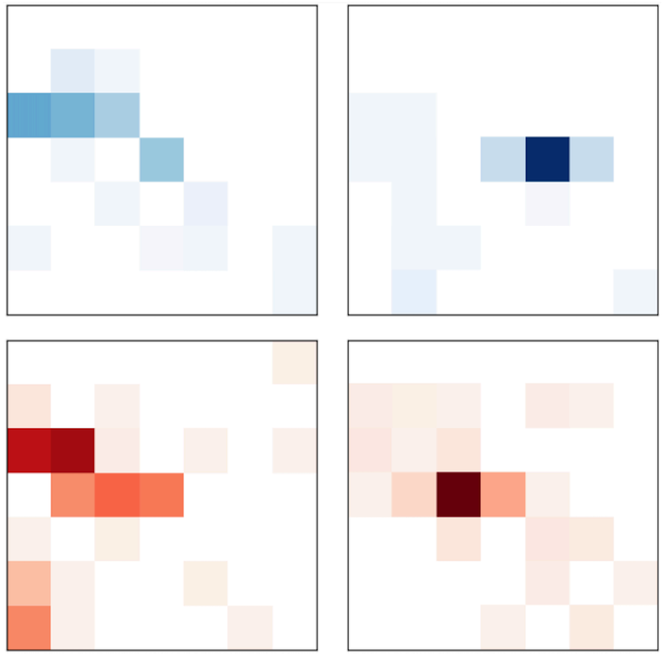




- Model is trained with BG×1.

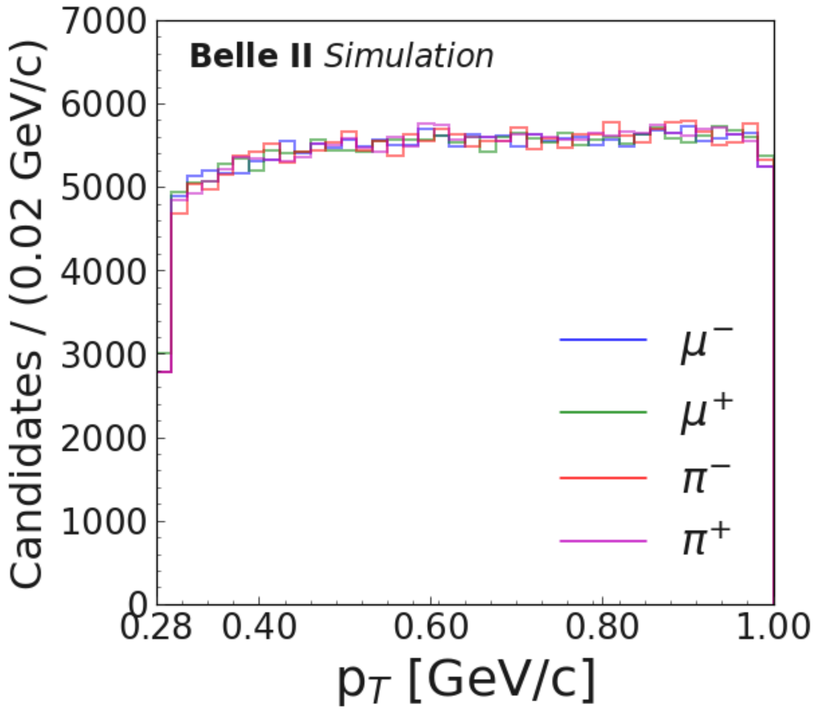
Summary

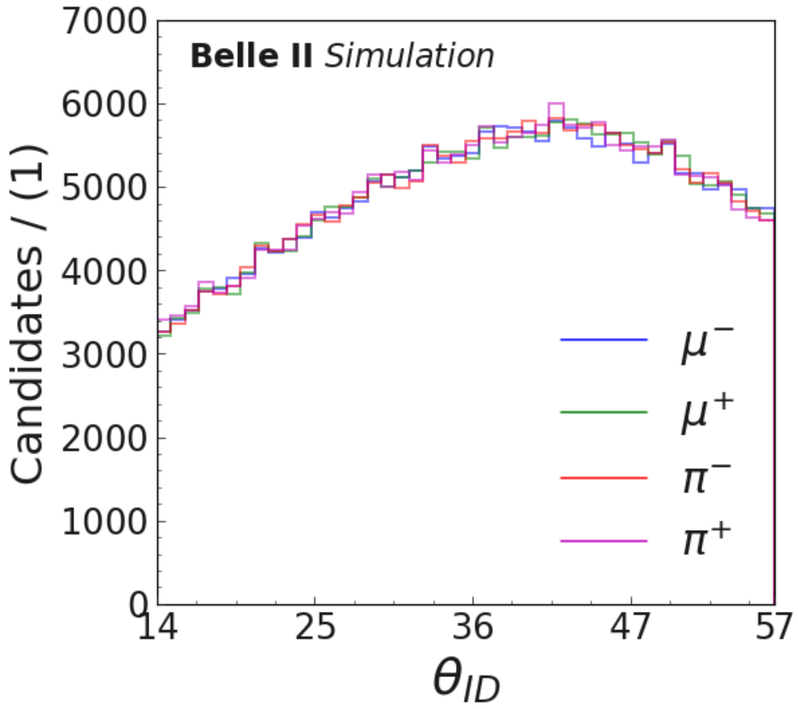
- Using 7x7 crystal images in the ECL, muon-pion separation can be improved for low-momenta tracks.
- Two separate CNNs are trained for positive and negative charged tracks.
- CNN method does not rely on clustering or shower-shape information.
- CNN outperforms default and BDT method:
  - For a π fake rate of 0.2, μ efficiency is:
    - Default: 38.21% –> BDT: 61.78% –> CNN: **65.26%** (+)
    - Default: 34.40% –> BDT: 61.16% –> CNN: **68.12%** (–)

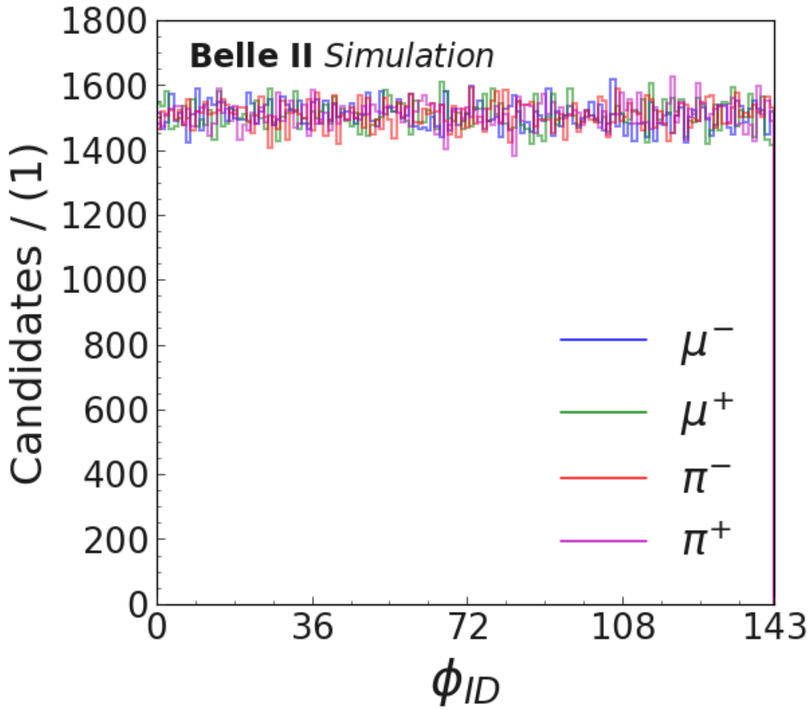


Back-up

Training samples and inputs



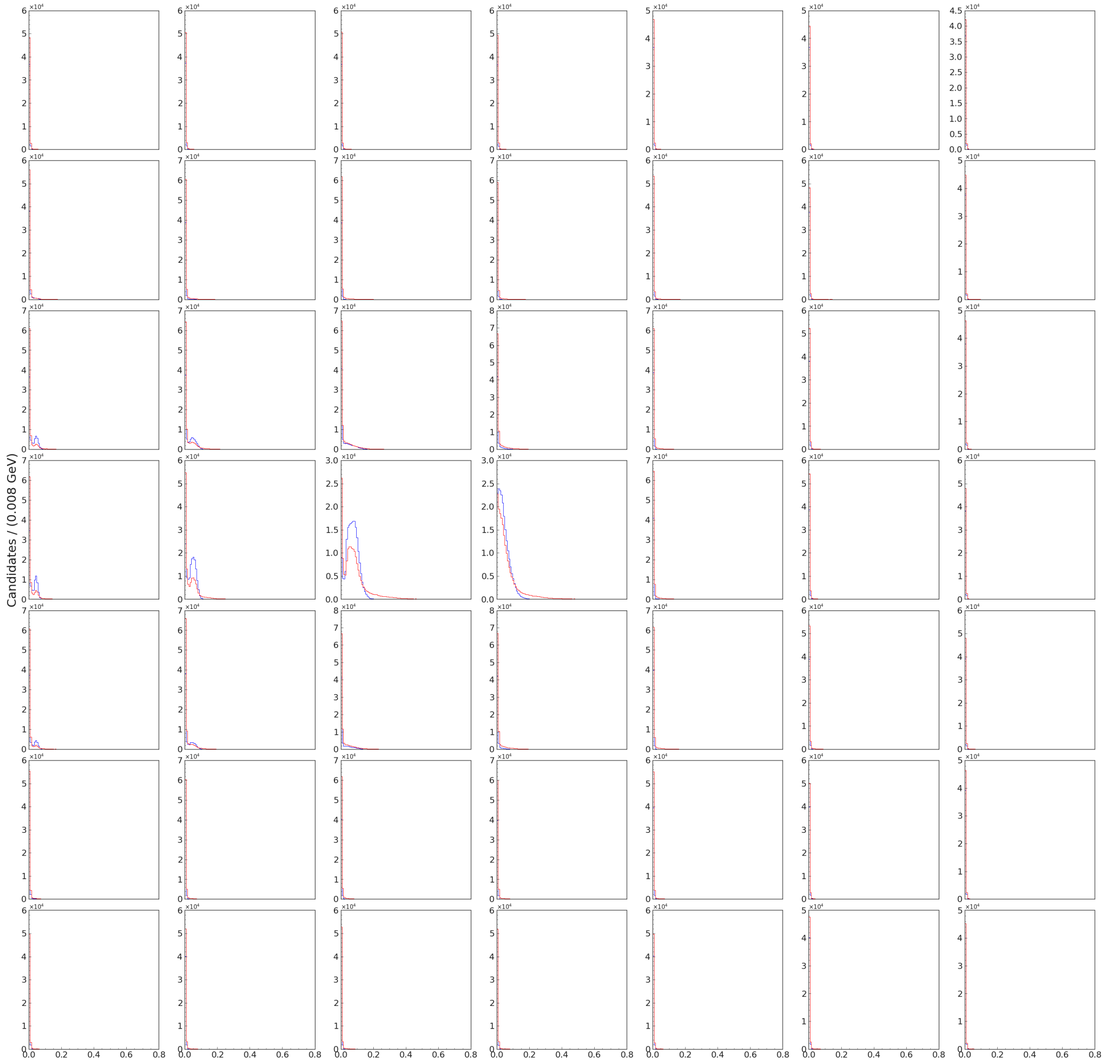




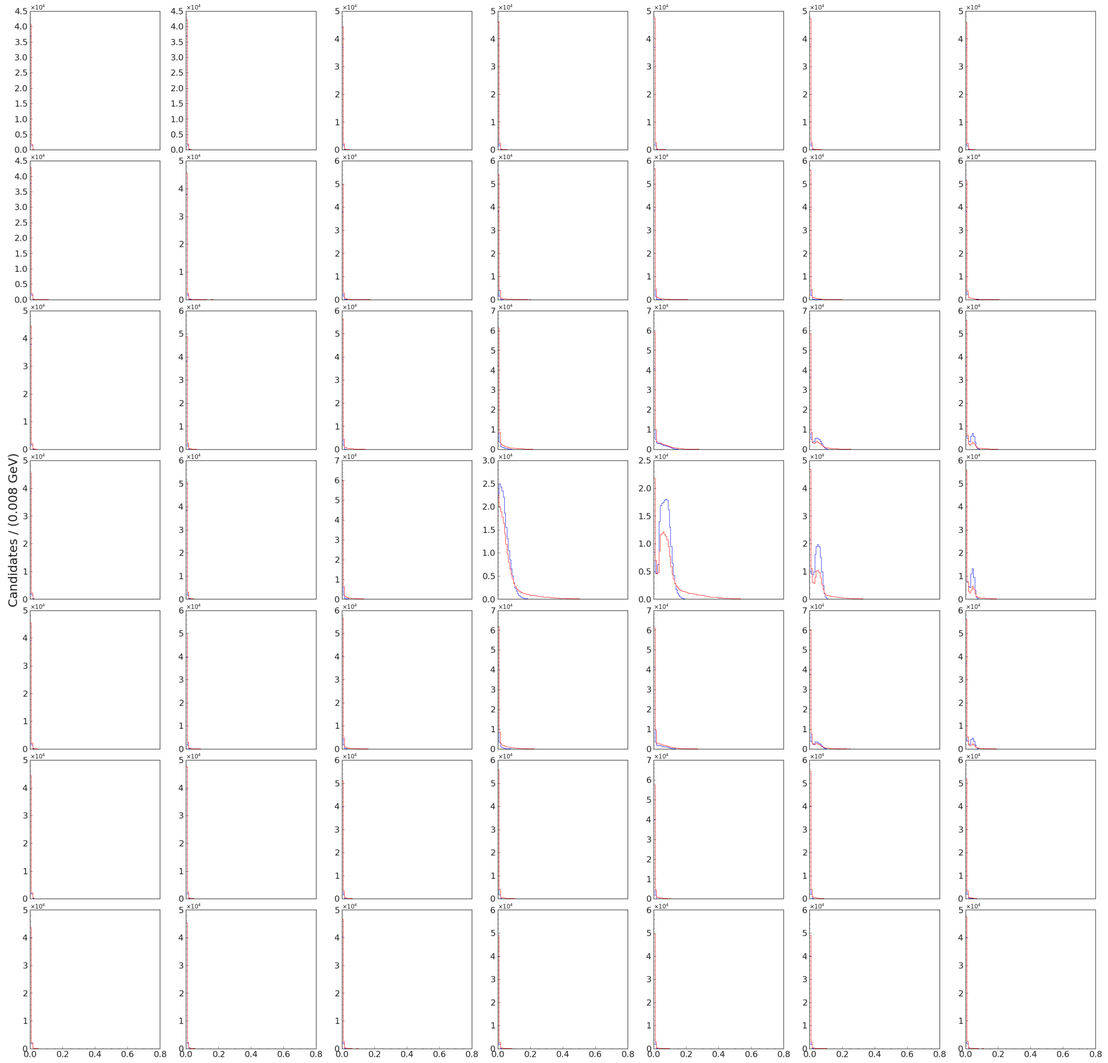
Energy deposition inputs

Muons
Pions

Negative charged tracks



Positive charged tracks



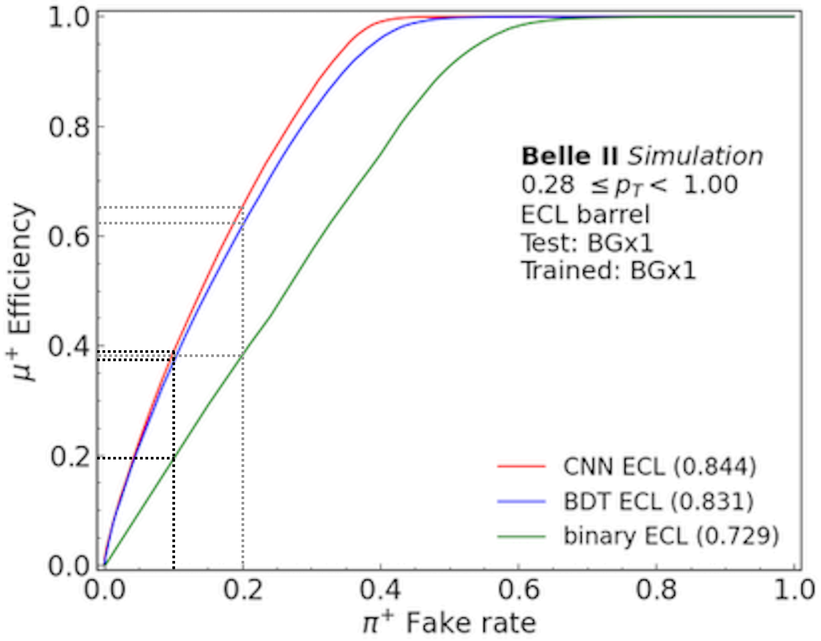
Inputs & CNN parameters

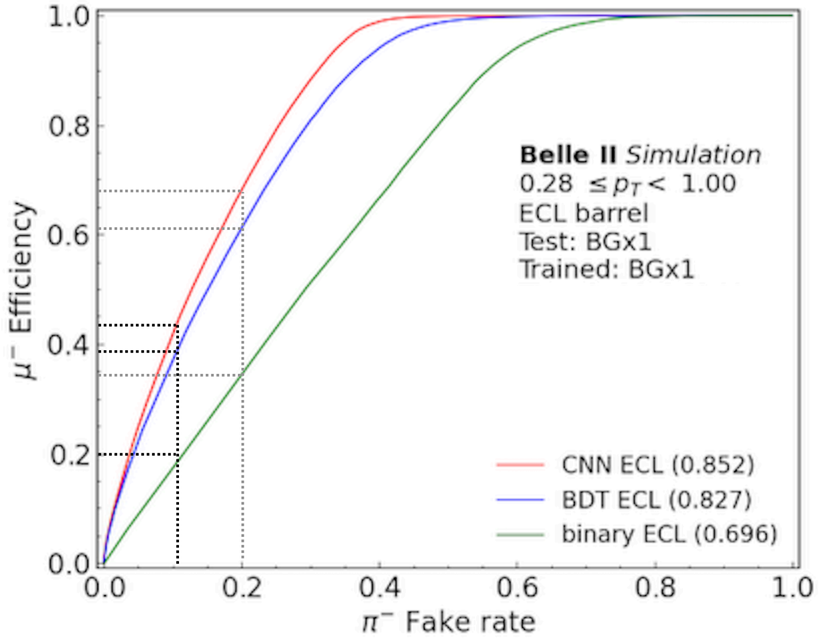
- Extrapolated energy (Reconstructed track-seeded cluster energy):
  - Threshold: 1 MeV (If energy < 1 MeV: energy = 0)
  - If energy > 1GeV: energy = 1
  - Since energy values are already between 0 and 1, I do not apply any scaling.
- Transverse momentum:
  - No scaling is applied.
- θ_{ID} and ϕ_{ID}:
  - They are given as categorical inputs (embedding layer).

- Kernel: (3, 3), Stride: (1, 1), Padding: (1, 1)
- Feature maps: 64
- # neurons in FC1: 3295
- # neurons in FC2: 128
- Dropout between FC1 and FC2: 0.1
- Optimizer: Adam
- Learning rate: 0.001
- Loss: CrossEntropy
- Batch size: 512

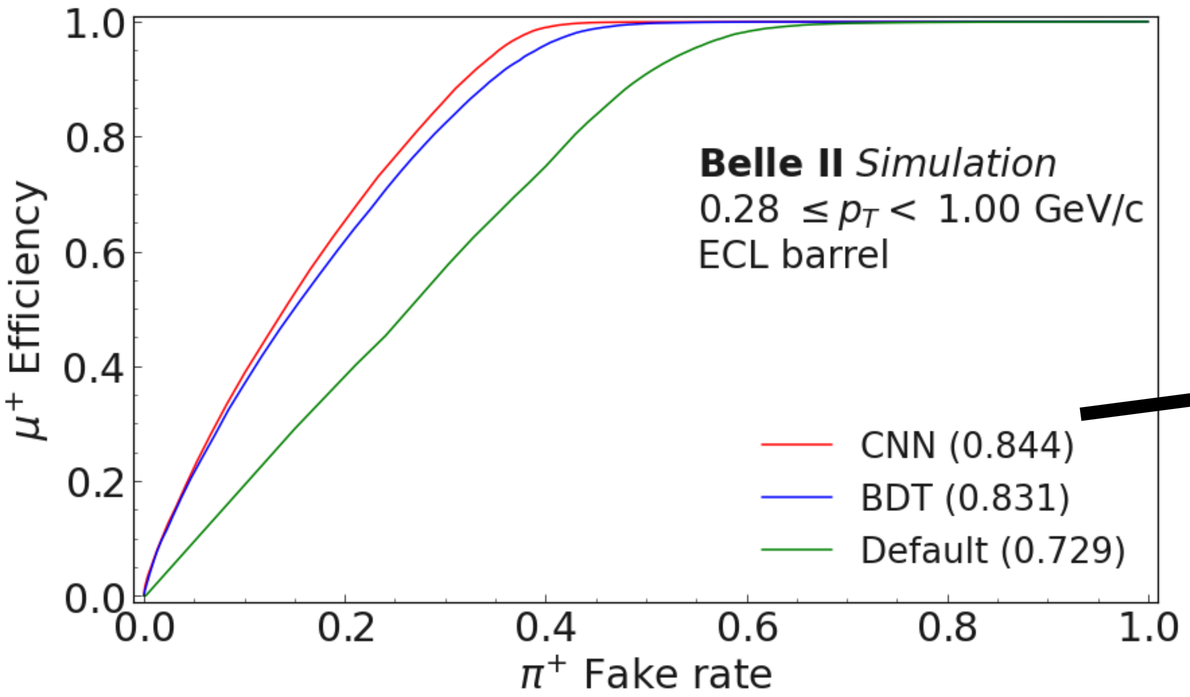
CNN and other PID methods

Method	10% (positive)	20% (positive)	10% (negative)	20% (negative)
CNN	38.71	65.26	41.75	68.12
BDT	36.89	61.78	36.97	61.16
Default	19.25	38.21	17.32	34.40



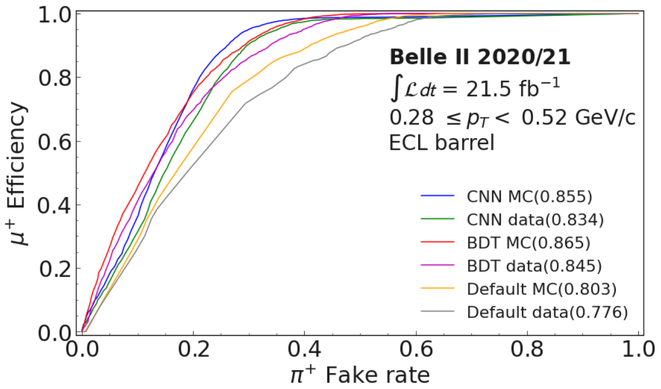


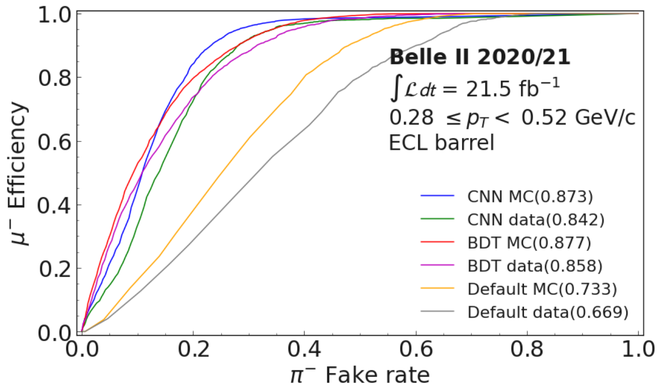
CNN - different beam background levels (+)

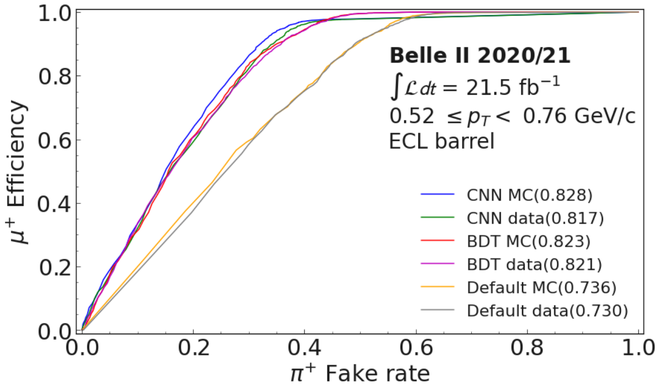


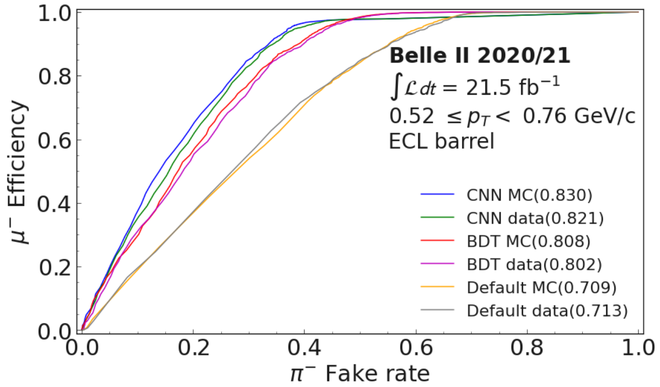
Test \ Train	BG×1	BG×2	BG×5
BG×1	0.844	0.841	0.838
BG×2	0.825	0.837	0.834
BG×5	0.818	0.83	0.833

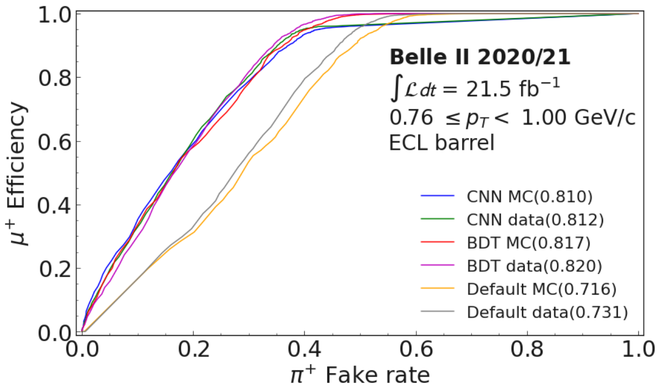
CNN comparison with different PID methods for all p_T regions

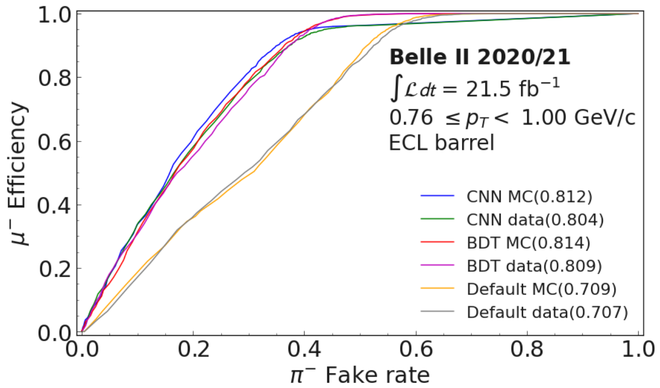












CNN comparison with all tracks for all p_T ranges (data and MC)