EPS-HEP CONFERENCE
07-11 JULY, 2025
PALAIS DU PHARO
MARSEILLE, FRANCE

Muon identification with Deep Neural Network in the Belle II K-Long and Muon detector

Based on arXiv:2503.11351
Accepted by NIM-A

WANG Zihan 王子涵

The University of Tokyo

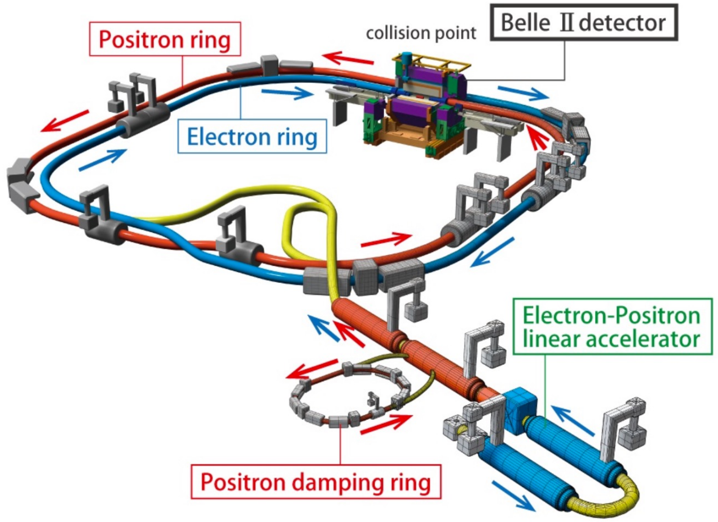
On behalf of authors

EPS-HEP 2025

2025.07.09

This work is supported by JSPS KAKENHI
Grant Number: JP24KJ0650 and JP22H00144

SuperKEKB and Belle II Experiment



SuperKEKB

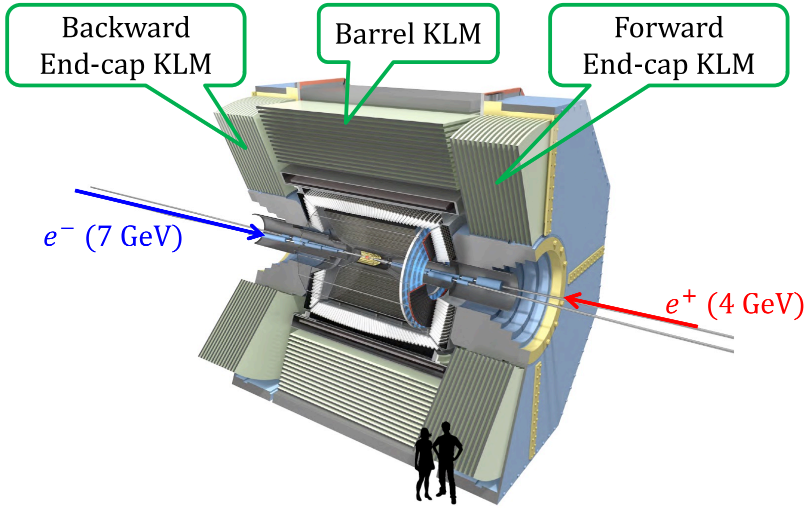
- Asymmetric e^+e^- collision at $\Upsilon(4S)$ resonance

	$\int \mathcal{L}\,dt$	$\mathcal{L}$
Target	50 ab^{-1}	6×10^{35} $cm^{-2}\,s^{-1}$
Achieved	575 fb^{-1}	5.1×10^{34} $cm^{-2}\,s^{-1}$

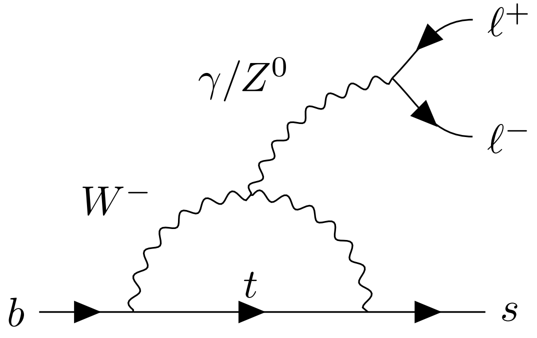
New world record!

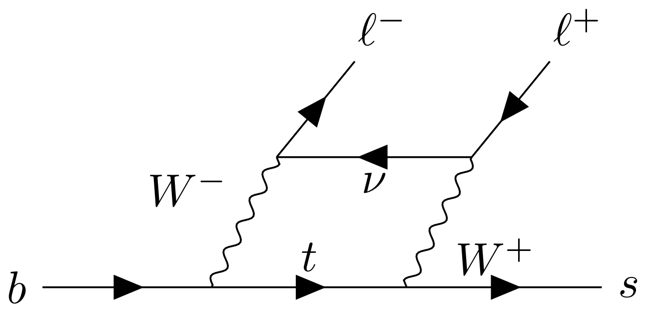
Belle II experiment

- New physics search at the luminosity frontier
- Rich muon-related physics topics like $b\to s\mu^+\mu^-$ transition, $b\to c\tau\nu$ with $\tau\to\mu\nu\bar{\nu}$
- Muon detection by K-Long and Muon (KLM) detector at the outer most of Belle II



Physics motivation



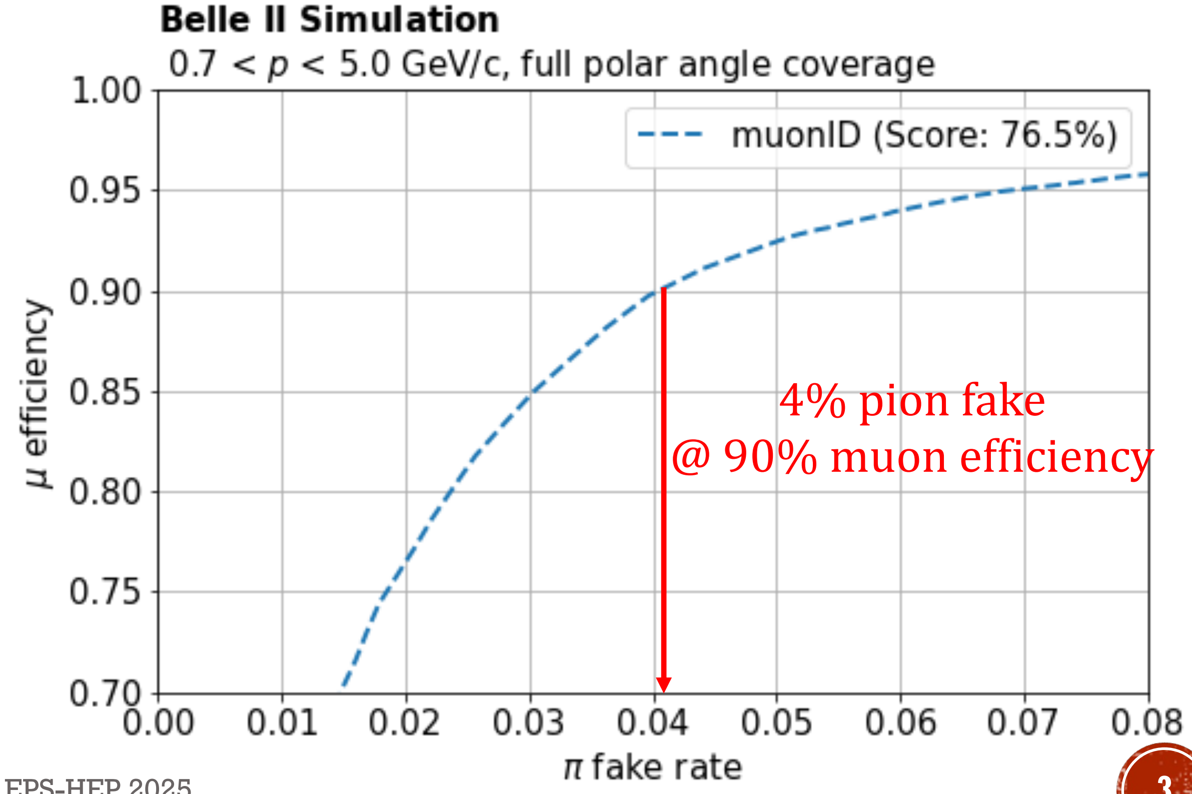


- Electro-weak penguin process $b \to s\mu^+\mu^-$ is suppressed in SM, and sensitive to new physics

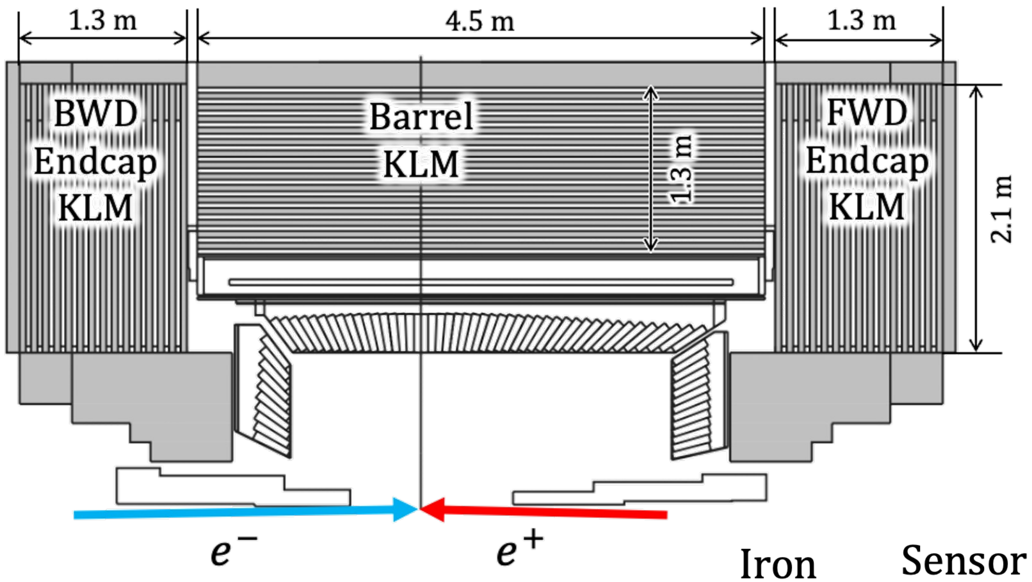
- $\pi^\pm \to \mu^\pm$mis-identification from $B \to X\pi^+\pi^-$ decay is one of the main background
  - $\mathcal{B}(B \to X\pi^+\pi^-)/\mathcal{B}(B \to X_s\mu^+\mu^-) = O(1000)$

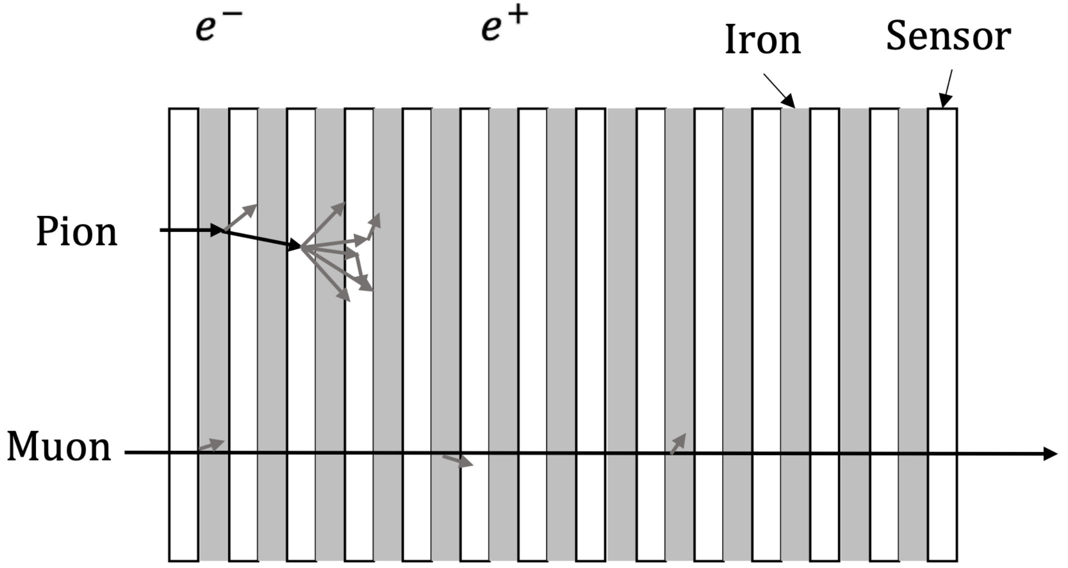
- π fake $< 2\%$ is required

- **Target：Suppress pion fake rate below 2% @ 90% muon efficiency**

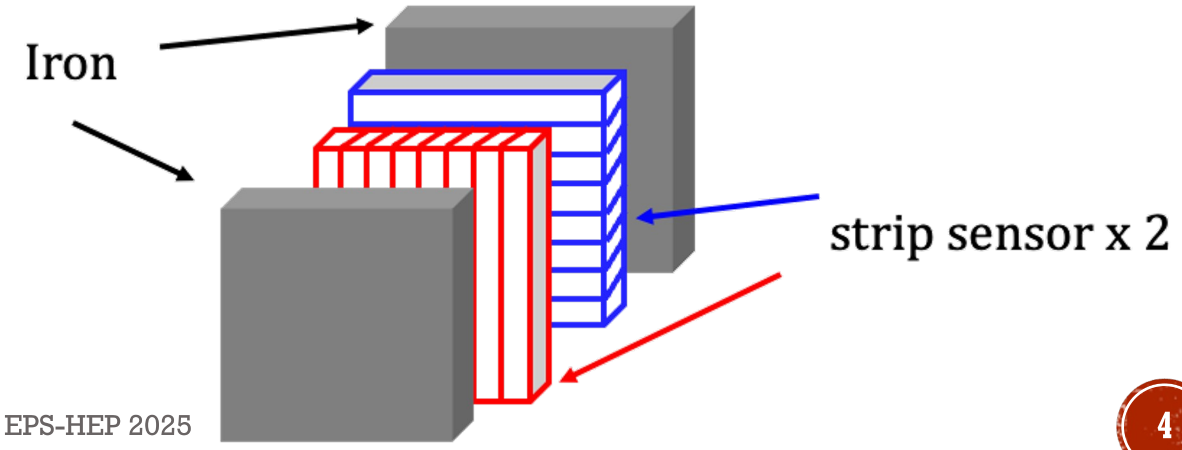


K-Long and Muon Detector (KLM)





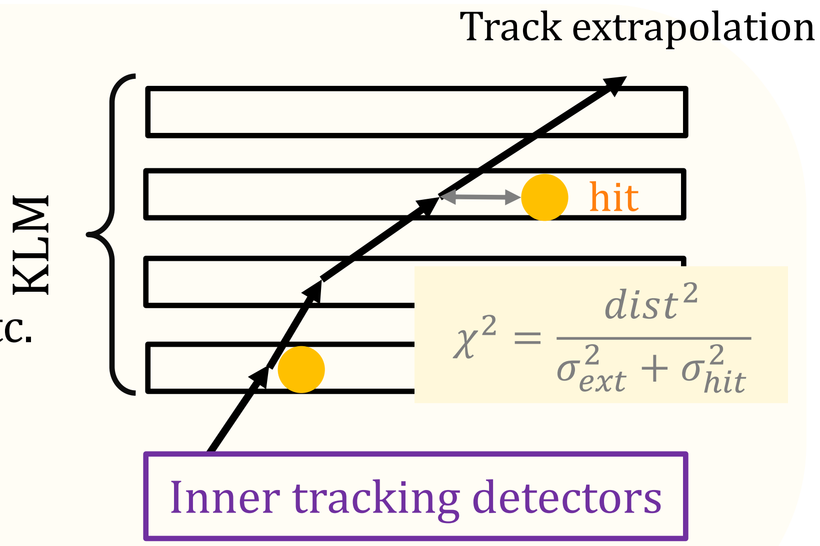
- Sandwich structure of iron plate and sensor
  - Barrel: 15 layers, Endcap: 14/12 layers for FWD/BWD
  - Sensor: Resistive Plate Chamber or Plastic scintillator
- Muon identification based on penetration ability
  - $>$ 3.9 interaction length
- Detect 2-dimentional spatial information with 2 orthogonally placed strip sensors in each layer



Current muonID algorithm

STEP 1 Track extrapolation & Hit association

- Measure track position/momentum with inner detectors
- Use Geant4E to estimate penetration depth inside KLM
  - Assuming tracks are muon, considering multiple-scattering effects etc.
- Associate one hit with smallest χ^2 to target track at each layer
  - Also require $\chi^2 < (3.5)^2$
- Use Kalman-Filter to correct extrapolation



$$\chi^2 = \frac{dist^2}{\sigma_{ext}^2 + \sigma_{hit}^2}$$

STEP 2 muonID calculation

- Use penetration ability and extrapolation quality to calculate muon/pion likelihood ($\mathcal{L}_\mu/\mathcal{L}_\pi$)
- muonID definition:

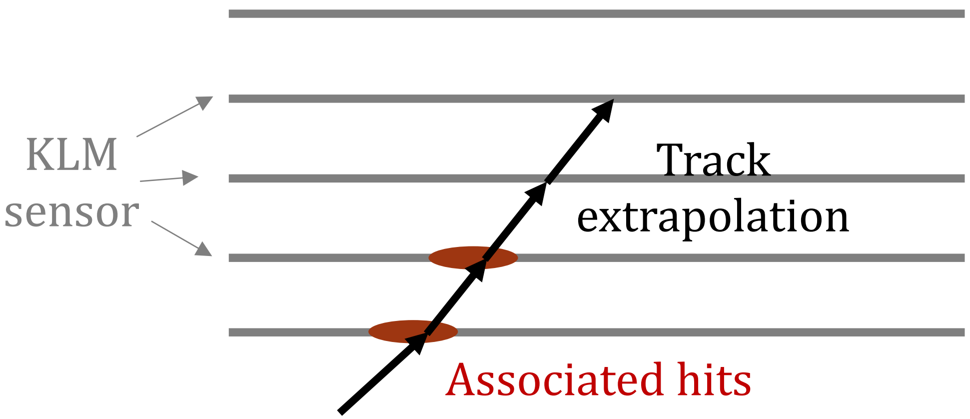
$$\text{muonID}_{\text{KLM}} = \frac{\mathcal{L}_\mu}{\mathcal{L}_\mu + \mathcal{L}_\pi}$$

Likelihood extraction on penetration ability

- Assign a probability to each layer depending on whether a hit is associated or not
- Calculate probability by direct product of layer probability

$$\mathcal{L}_t^{pene} = \prod_{n=1}^{N_{ext}} L_{t,n} \qquad t \in \{\mu, \pi\}$$

- For tracks extrapolated to stop at layer 4 ($N_{ext} = 4$) and only hits associated at the first two layers ($N_{hit} = 2$):



$$L_{t,4} = 1 - P_{t,4}$$
$$L_{t,3} = 1 - P_{t,3}$$
$$L_{t,2} = P_{t,2}$$
$$L_{t,1} = P_{t,1}$$

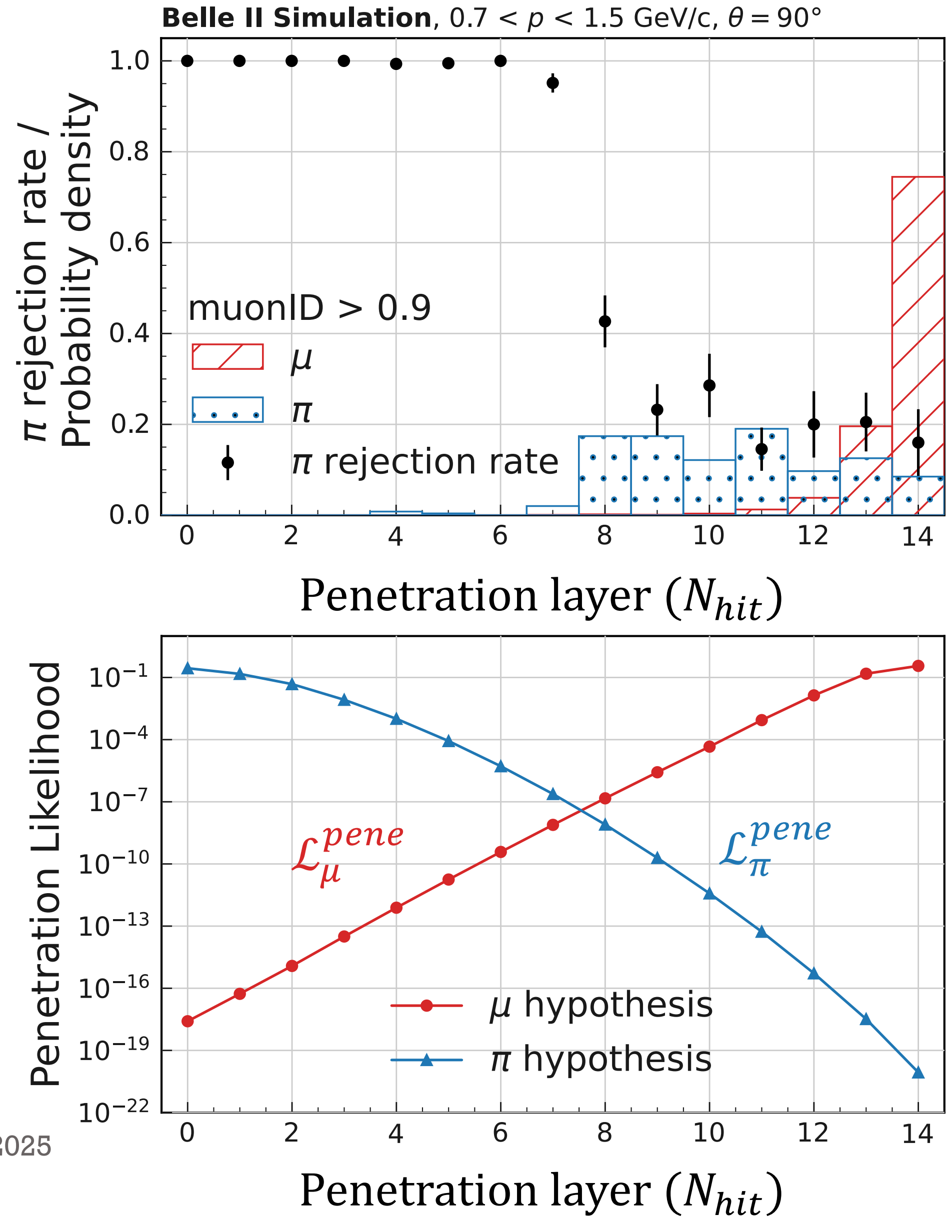
$$\mathcal{L}_t^{pene} = \prod_{n=1}^{N_{ext}=4} L_{t,n} = P_{t,1} P_{t,2} (1 - P_{t,3})(1 - P_{t,4})$$

Discussion on penetration likelihood

- For tracks extrapolated to layer 14 ($N_{ext} = 14$), muonID only rejects pions with penetration (stopping) layer $N_{hit} < 8$
- Still, quite a fraction of remaining pions exists in $8 \leq N_{hit} \leq 10$, which rarely happens for muon

- MuonID failed to reject pions in $8 \leq N_{hit} \leq 10$ due to imperfect modeling of penetration likelihood
  - $\mathcal{L}_{\mu}^{pene}$ larger than $\mathcal{L}_{\pi}^{pene}$ from $N_{hit} > 8$
  - Because **correlations between layers** are not properly considered

Better modeling of penetration information can improve performance



Strategy

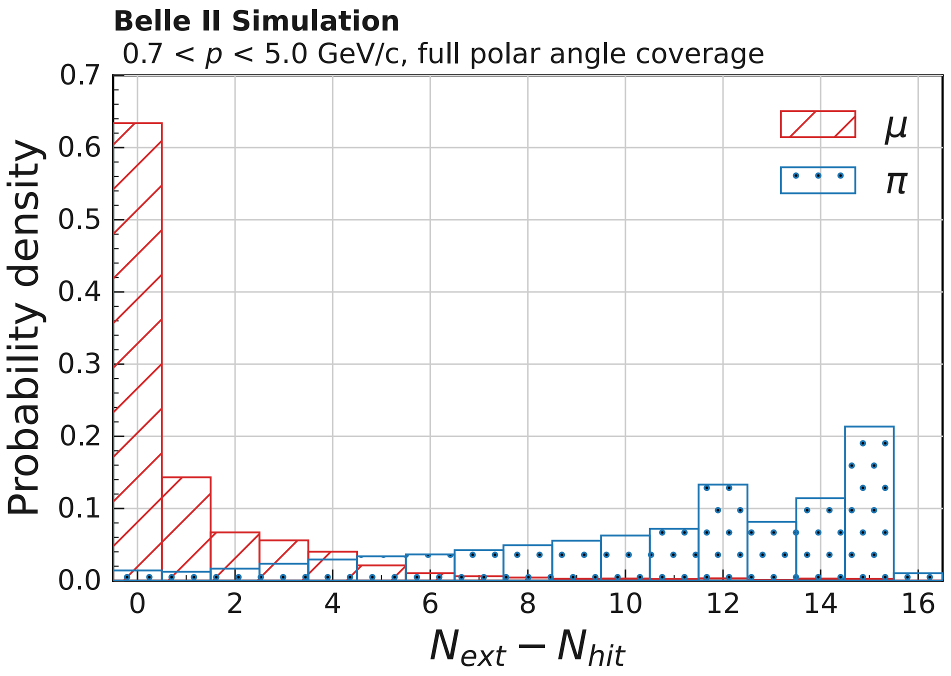
- New Deep Neural Network (DNN)-based algorithm
- Consider correlations between input variables

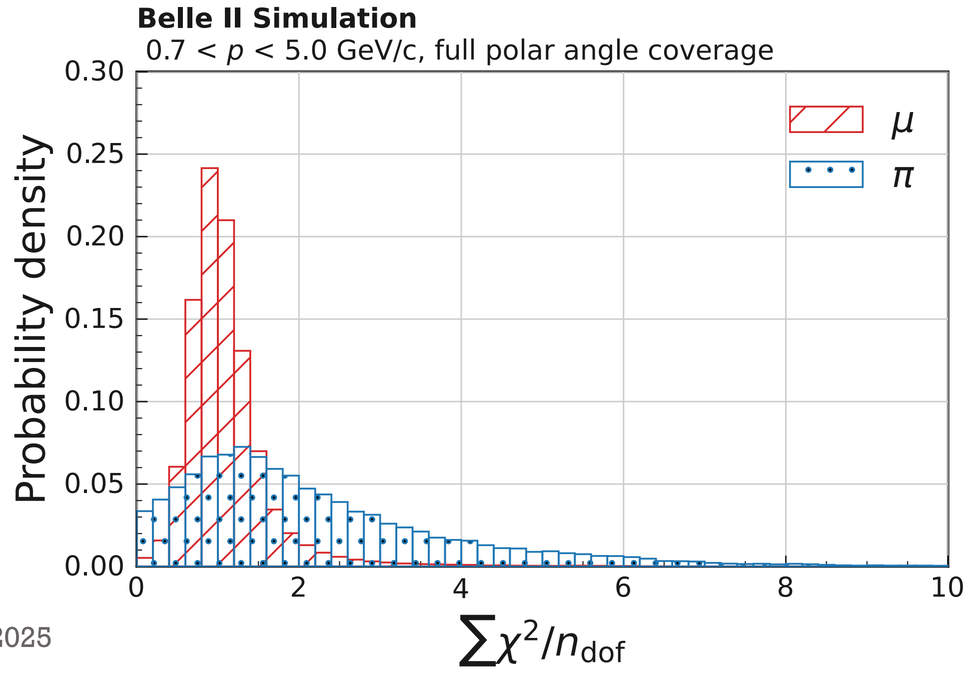
Input variables

1. Global variables
   - Track-level variables like penetration layers
2. Hit pattern variables
   - Detailed hit-level information
   - New for muon identification

Global variables

- Difference in extrapolation & penetration layer: $N_{ext} - N_{hit}$
- Extrapolation layer
  - Penetration layer distribution is different for different extrapolation layer
- Extrapolation quality($\sum\chi^2$ and n_{dof})
  - Muons extrapolated with muon hypothesis and thus peak at 1
  - Pion has larger value due to multiple scattering
- Transverse momentum



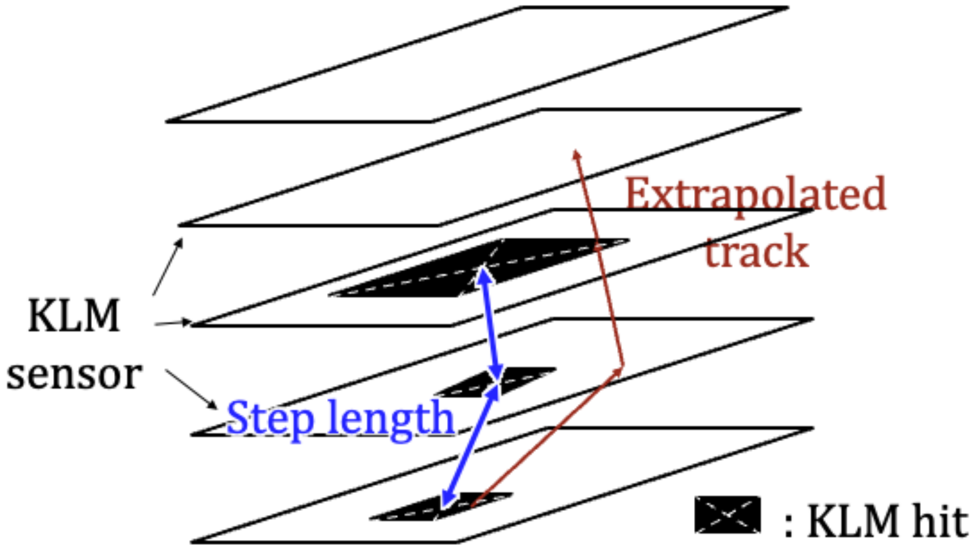


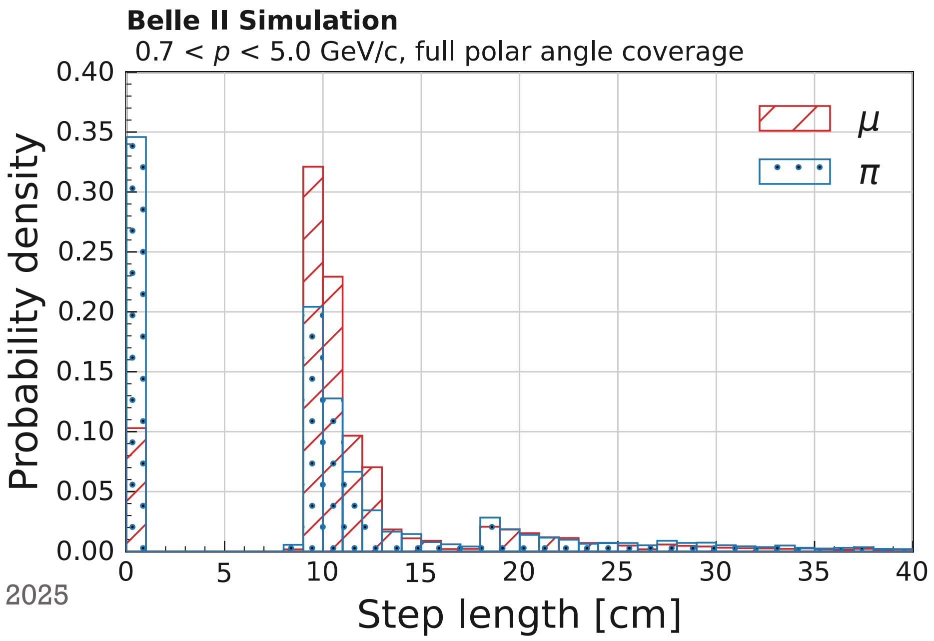
Hit pattern variable: Step length

- Step length: distance to hit in previous layer

Strength:

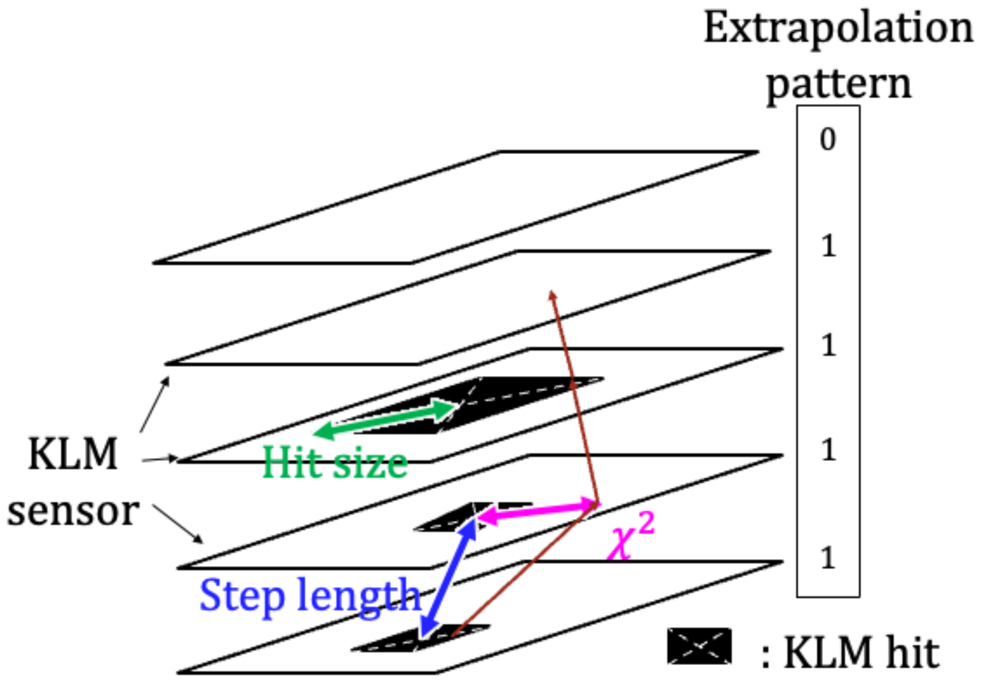
- Detailed 3D penetration information compared to penetration layer, which only reflects penetration projected in normal direction of detector plane
- Penetration depth (sum of step length) of muon is larger than pion

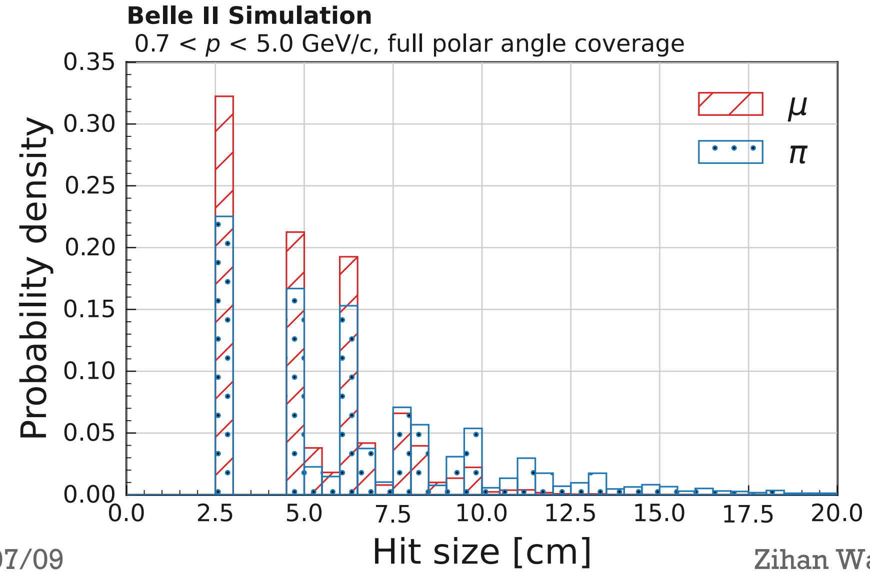


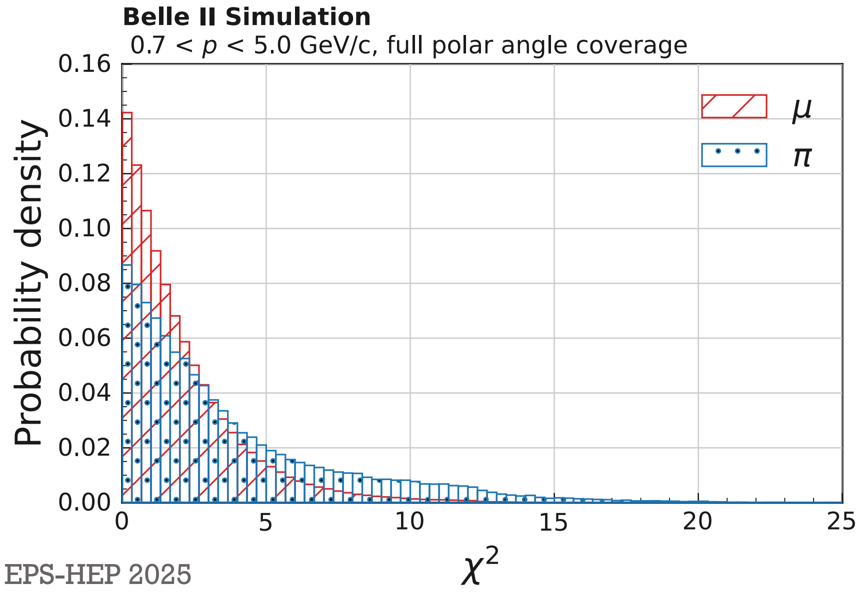


Other hit pattern variables

1. Hit size: Half of the diagonal length of the rectangular shape of a hit
2. χ^2: Deviation from hit to extrapolation position
3. Extrapolation pattern: whether extrapolation has reached this layer
   - To compare extrapolation and penetration pattern



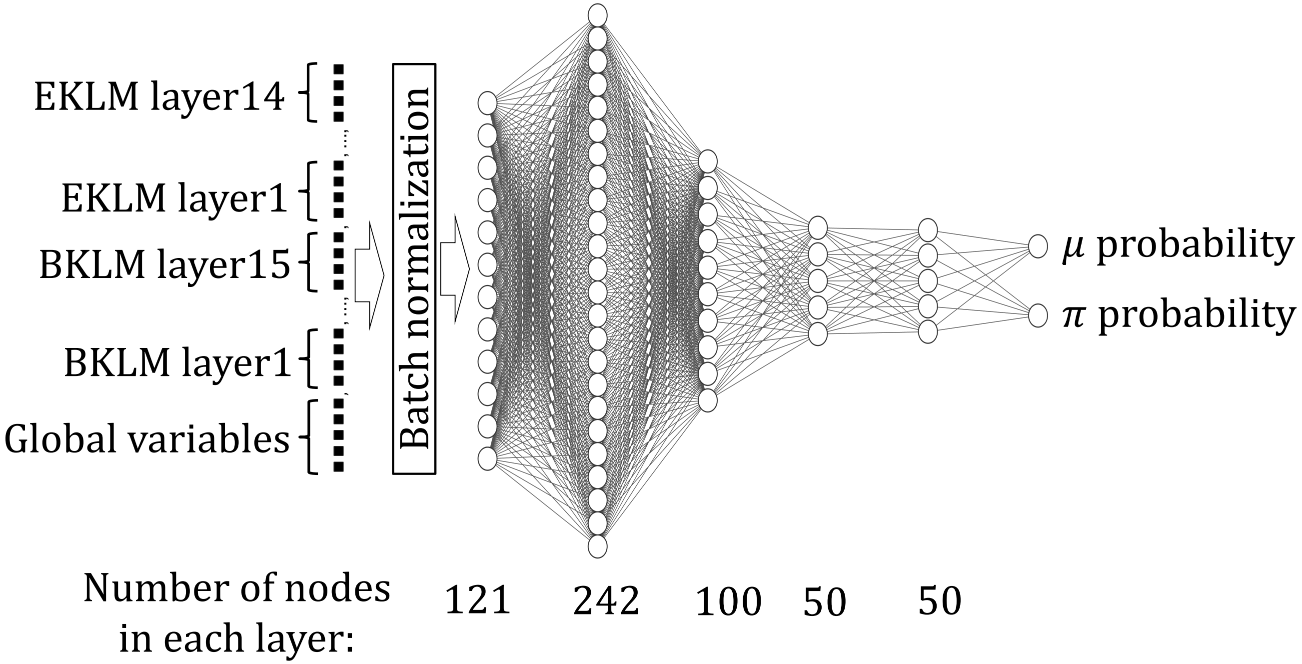




Network structure & training

Network structure

- 121 input variables
  - 5 global variables
  - 4 hit pattern variables per layer, 29 layers in total
- Fully connected Network
  - 64k trainable parameters
- LeakyReLU for activation function
- Softmax for probability output



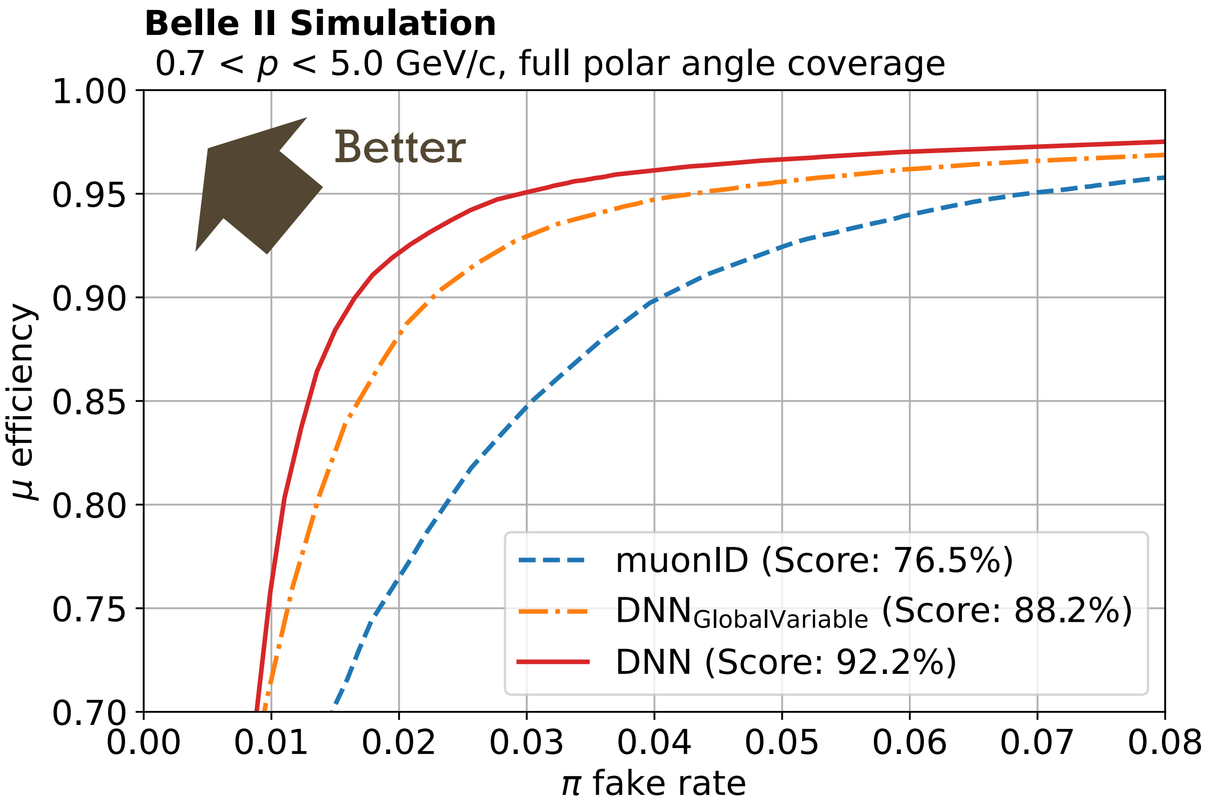
Training

- Event generated with 4 to 16 tracks for simulating different event multiplicity
  - Simulated beam-induced background overlayed
- Uniform phase space distribution covers full KLM geometric acceptance; 0.5M pions and muons
- No overtraining

Performance

- Also trained DNN with global variables only, showing the importance of hit pattern variables
  - Almost same network structure

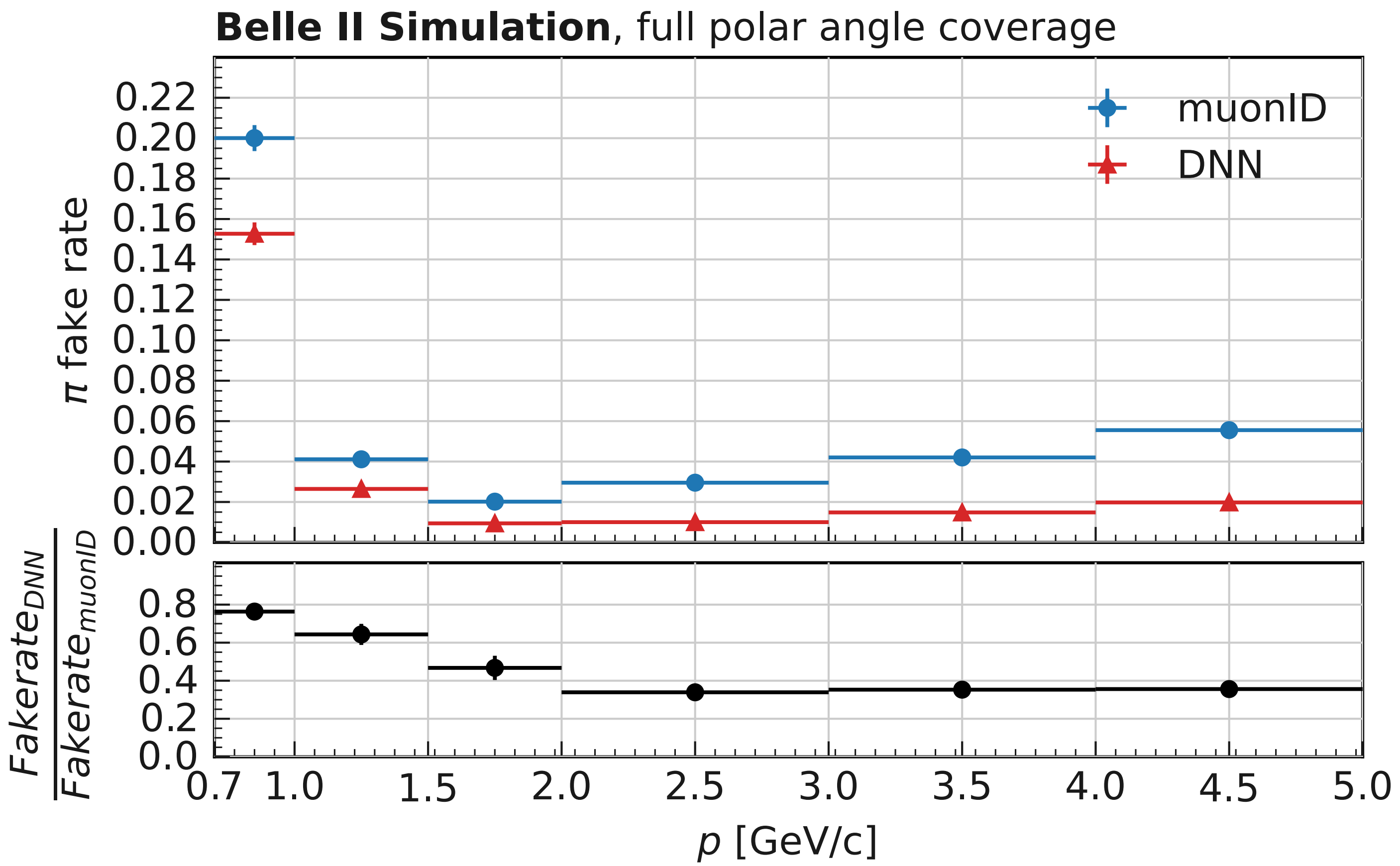
	Fake rate @90%muon effi	Muon effi @ 2% fake rate
muonID	4.1%	76.5%
$DNN_{GlobalVariable}$	2.3%	88.2%
DNN	1.6%	92.4%



DNN with hit pattern variables

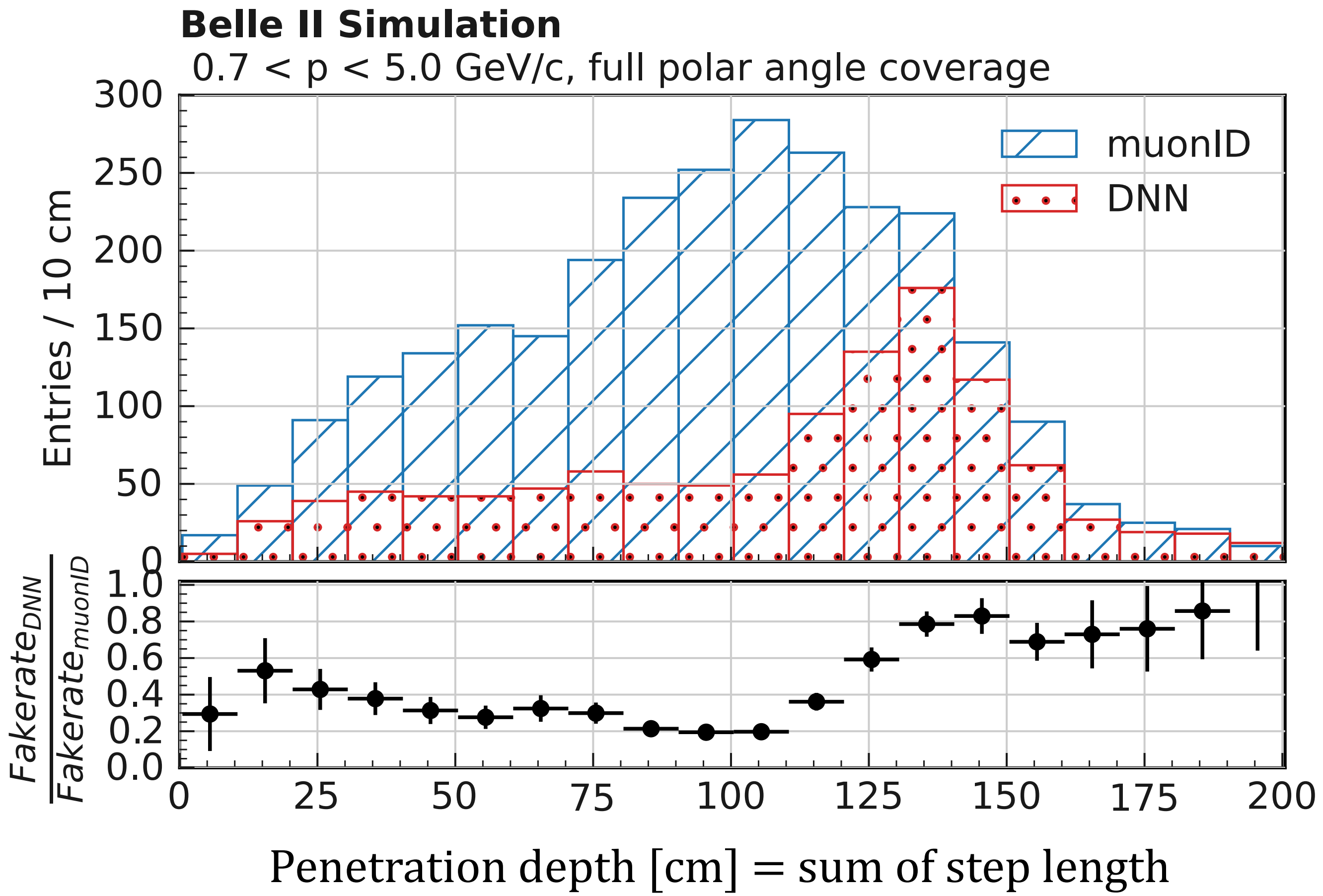
suppresses pion fake rate below 2% @ 90% muon efficiency

Performance



- Fix muon efficiency to 90% at each momentum bin.
- Pion fake rate improved over 60% at high momentum range ($p > 2.0$ GeV/c).
- Improvement less significant at low momentum range where muons cannot traverse KLM entirely.

Performance



- Above plot shows penetration depth distributions of pions faking muons at 90% muon efficiency
- DNN rejected about 60% of deeply penetrated pions up to a depth of 125 cm
  - Rise of fake rate at 125 cm penetration depth similar KLM detector thickness of 130 cm
  - For tracks escaping from KLM, μ/π separation by penetration ability is impossible

Summary

- Muon identification algorithm with high performance is necessary for precise measurement of flavor physics at Belle II
- Current muonID cannot reject deeply penetrating pion
- We developed a new algorithm using DNN, employing additional hit-pattern variables
- We learned that this new algorithm suppresses pion fake rate below 2% at 90% muon efficiency in a simulation sample
  - Satisfied the requirement in $b \to s\mu^+\mu^-$ measurement
- This algorithm is implemented into Belle II software framework and validations on full dataset is expected in the future
- Check arXiv:2503.11351 for more details!

BACK UP

Likelihood calculations for muonID

$$\mathcal{L}_\mu = \mathcal{L}_\mu^{pene} \times \mathcal{L}_\mu^{in-plane}$$

- Extract the probability of track t leaving hits in layer n, $P_{t,n}$ from simulation, and assign probability to each layer:

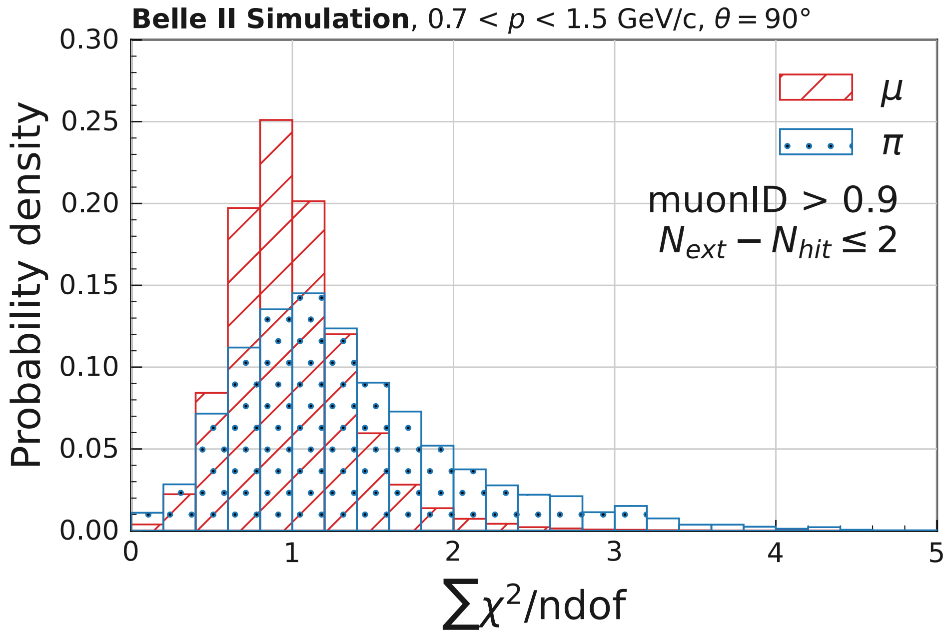
$$L_{t,n} = \begin{cases} P_{t,n} \cdot \varepsilon_n, & with\ hit \\ 1 - P_{t,n} \cdot \varepsilon_n, & without\ hit \end{cases}, \qquad t \in \{\mu, \pi\}$$

  - Typically, a 96% hit detection efficiency per layer (ε_n) is achieved in operations
- Calculate probability by direct product of layer probability

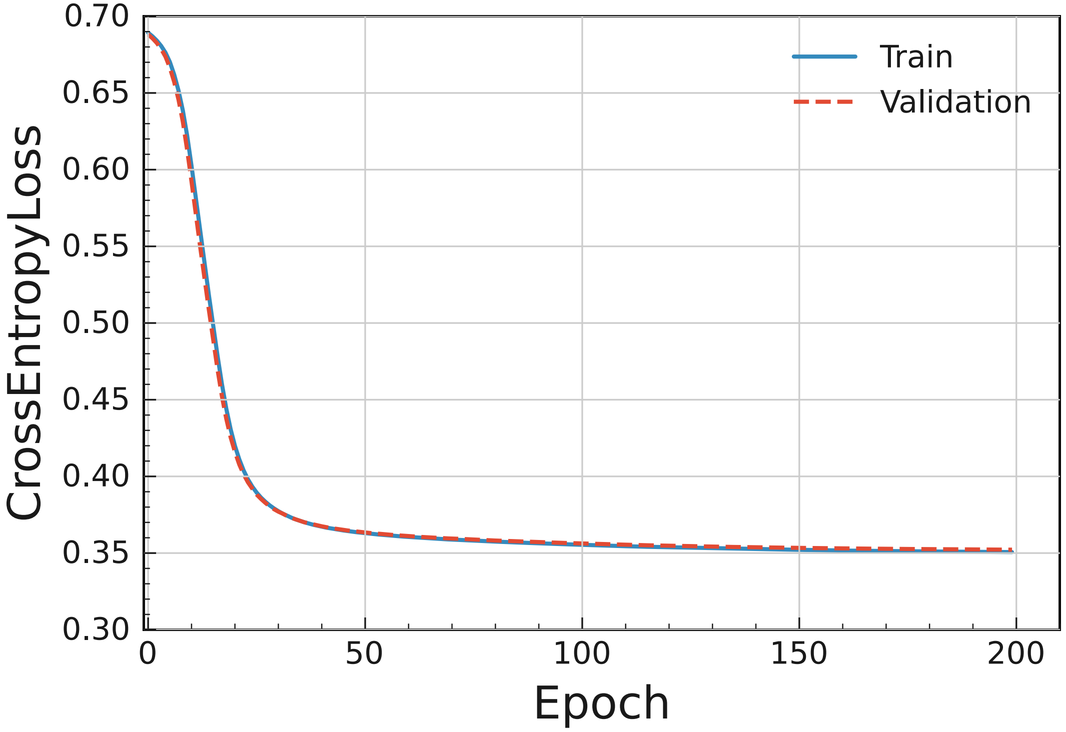
$$\mathcal{L}_t^{pene} = \prod_{n=1}^{N_{ext}} L_{t,n}$$

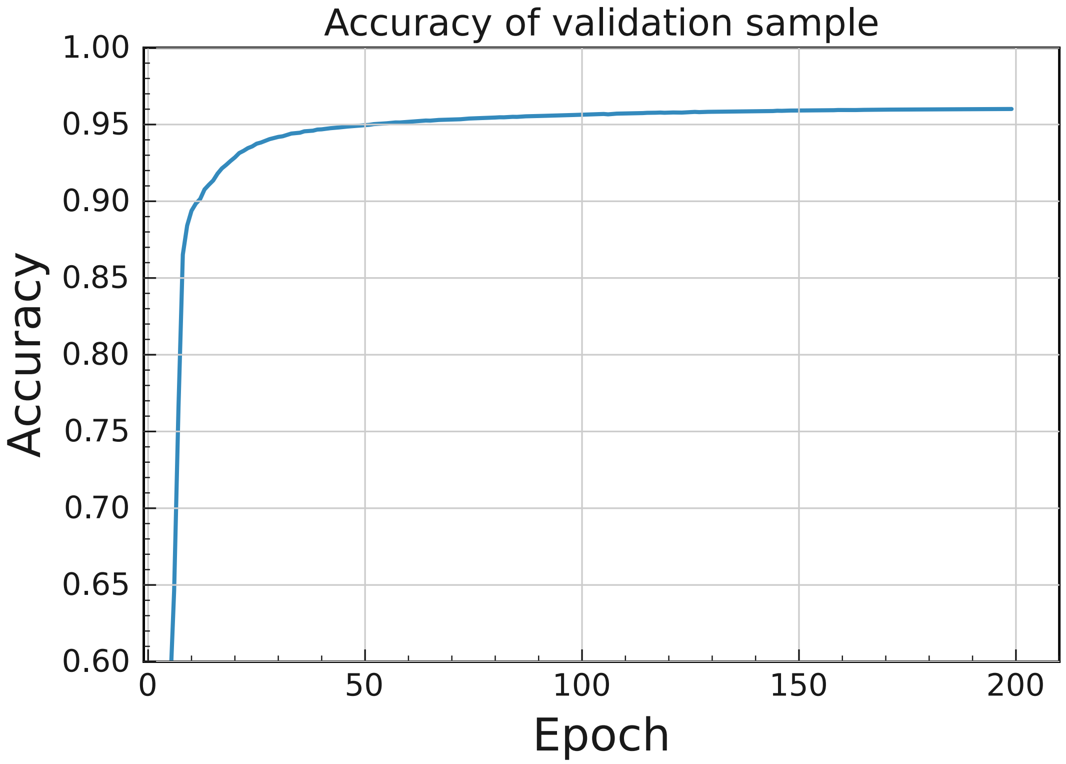
In-plane information

- In-plane likelihood derived from $\sum\chi^2/n_{dof}$ distribution
  - Extrapolated with muon hypothesis and thus peak at 1
  - Pion has larger value due to multiple scattering
- Still, some remaining pions can be rejected by requiring $\sum\chi^2/n_{dof} < 2$
- Failed to do so because of different scale of $\mathcal{L}^{pene}$ and $\mathcal{L}^{in-plane}$.
- Specifically, $\mathcal{L}_\mu^{in-plane}/\mathcal{L}_\pi^{in-plane} \approx 10^{-1}$
- , $\mathcal{L}_\mu^{pene}/\mathcal{L}_\pi^{pene} \approx 10^{13}$
- The relationship muonID>0.9 can hardly be influenced by the in-plane likelihood.



Training details





- Adam optimizer; learning rate of 10^{-5}; batch size $= 10000$; CrossEntropy loss
- Early stop when accuracy does not increase for 10 epochs
- Device: RTX 3090 for 30 min training.

Robustness test and implementation

Robustness test

- Background:
  - Sample overlayed with beam background simulated at Belle II nominal luminosity (x3 of training sample)
  - Pion fake rate degrades to 2.4%, probably because of changes in hit pattern
  - Still DNN performs better than muonID, and re-training can be expected when machine status evolves
- Detector inefficiency
  - Set hit detection efficiency 10% below typical efficiency
  - Pion fake rate slightly degrades to 2.0%
- Extrapolation uncertainty from magnetic field modeling
  - Verified B field by 10%, no significant performance degradation

Implementation into Belle II Analysis Software Framework (basf2)

- Tested inference latency is 1.81 ± 0.78 ms per $e^+e^- \to B\bar{B}$ event on KEK CPU cluster (KEKCC)
- Implemented in offline reconstruction phase of Belle II data processing