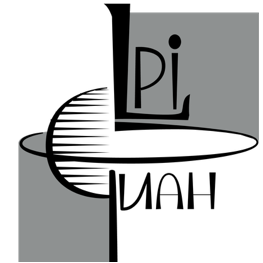


ICPPA 2026

Artificial Intelligence in particle physics and astrophysics

Event-Adaptive FARICH Reconstruction under Unknown Dark-Count Rates with a Track-Conditioned Transformer



P. Rogozhin^{1,2}

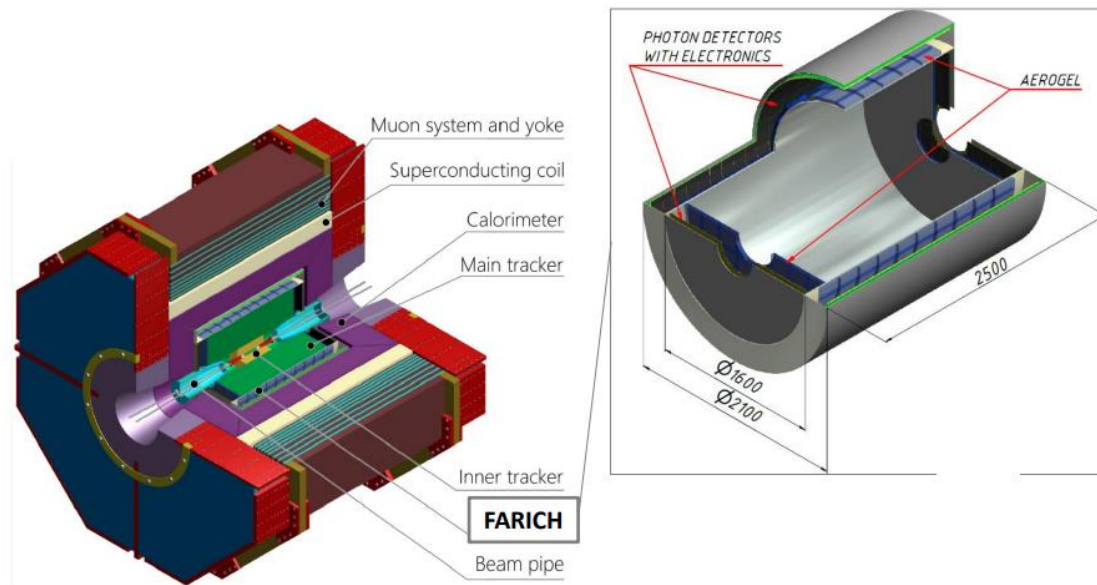
¹ HSE University, Moscow, Russia

² P.N. Lebedev Physical Institute, Russian Academy of Sciences

27–30 October 2026 Moscow

Super Charm-Tau Factory

Proposed high-luminosity e^+e^- collider



3–7 GeV

Centre-of-mass energy

$\approx 10^{35} \text{ cm}^{-2} \text{ s}^{-1}$

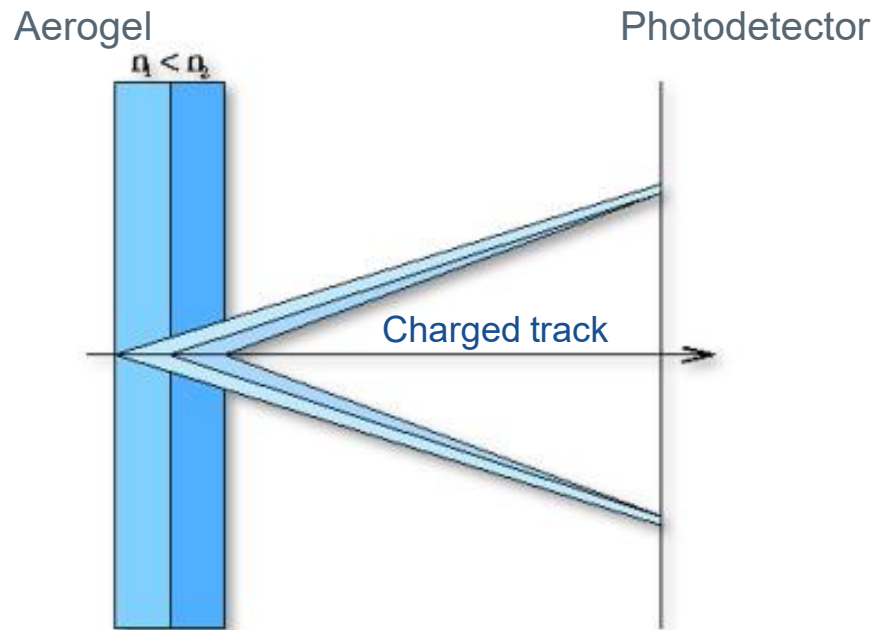
Design luminosity

Dedicated μ / π / K identification for charm-physics measurements

FARICH operating principle

Focusing Aerogel Ring-Imaging Cherenkov detector

Multilayer focusing



Different refractive indices reduce smearing from the photon emission depth.

Cherenkov light

A charged particle emits a light cone when its speed exceeds c/n .

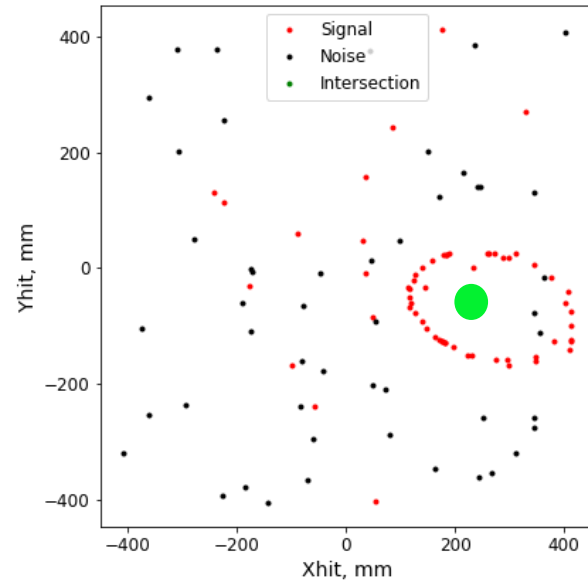
$$\cos \theta_C = \frac{1}{n\beta}, \quad \beta = \frac{v}{c}$$

Ring image and PID

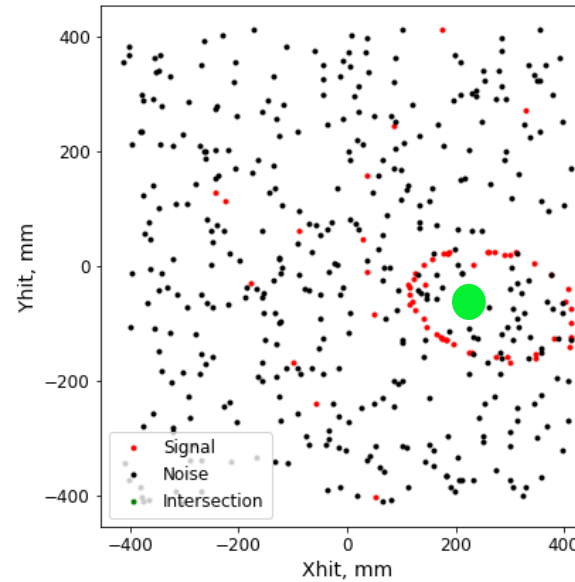
Pixelized photon hits constrain β . Together with track momentum, they distinguish $\mu / \pi / K$.

Photon patterns and dark counts

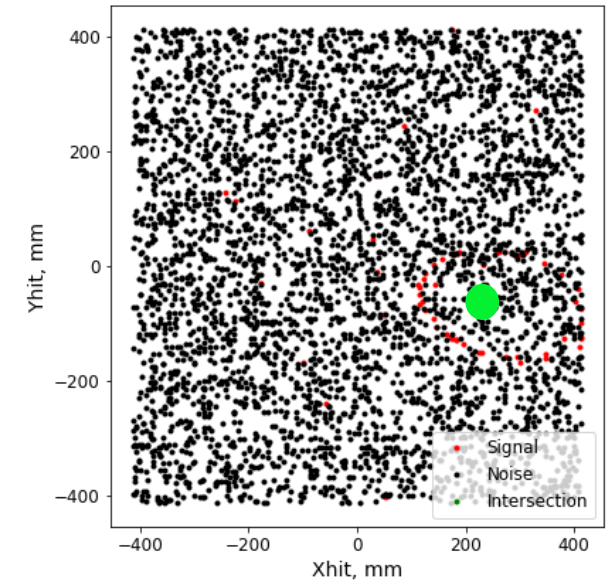
DCR: photosensor dark-count rate per unit area



10 kHz/mm²
cooled sensor reference



100 kHz/mm²
realistic no-cooling scenario



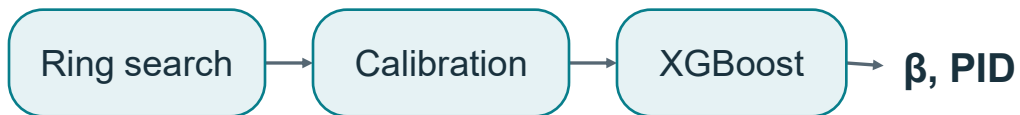
1 MHz/mm²
conservative radiation-damage scenario

Truth display: **signal**, dark counts, **track intersection**

Reconstruction objective

Reliable μ / π / K identification for charm-physics measurements

Earlier work [1]



Noise studies up to 1 MHz/mm²
Single-particle and D⁰-decay simulations

This study



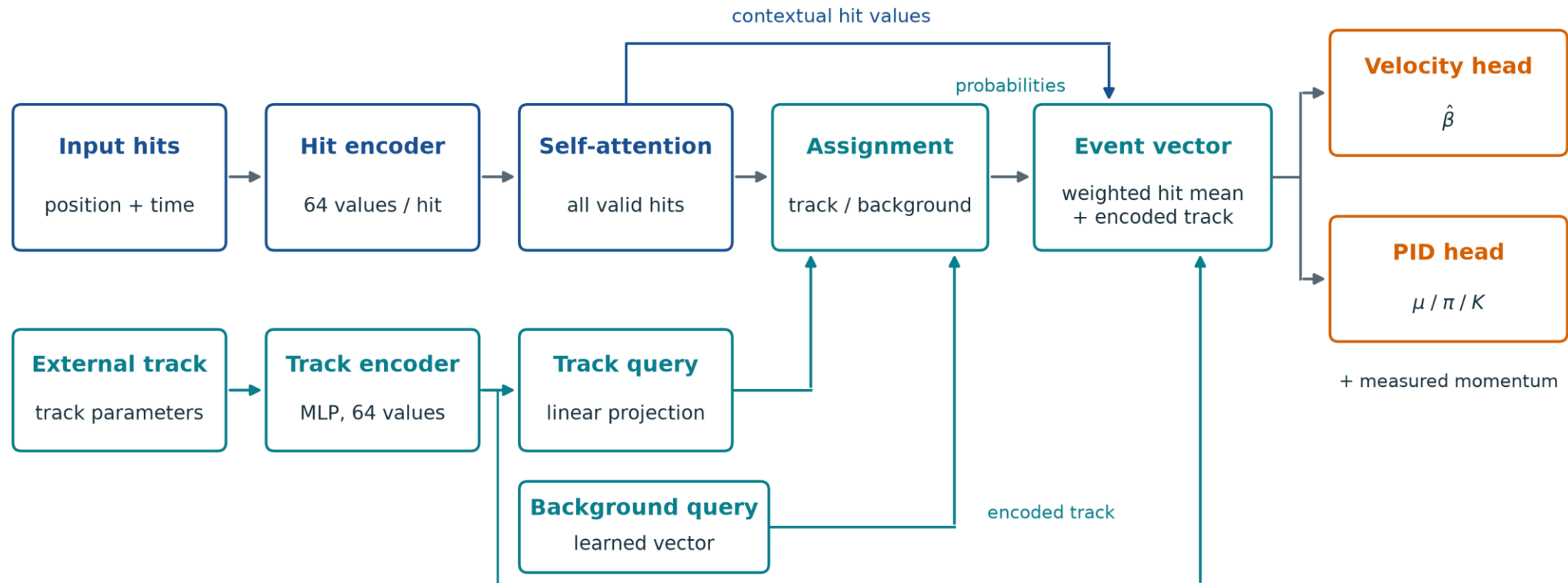
One model for 0–1 MHz/mm²
No dark-count rate (DCR) input:
it adapts to the hits of each event

Goal: joint hit assignment, β reconstruction and PID without a supplied DCR

[1] M. Chadeeva, P. Rogozhin, T. Uglov, Nucl. Sci. Tech. 37, 124 (2026)

TCRAT architecture

Track-Conditioned Ring Assignment Transformer

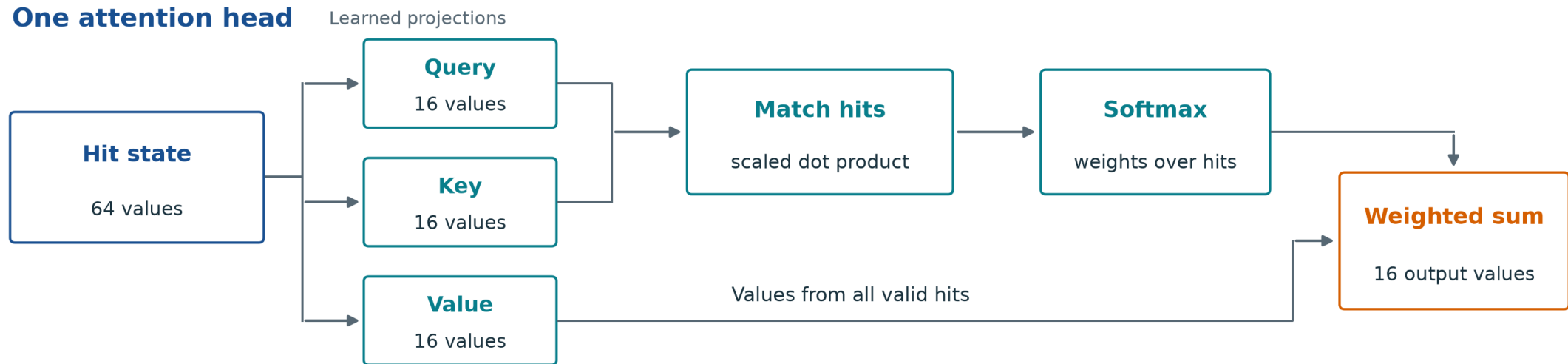
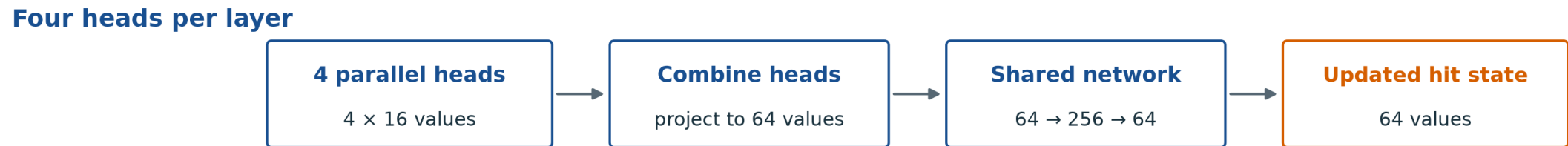


Preprocessing: candidate window around the track, then relative hit coordinates

Track/background scores sum to one for each hit. Scores do not sum to one across all hits.

Inside a transformer layer

Self-attention combines information from all valid hits



Two layers, four heads per layer, 64 values per hit

Simulation and training

Simulated data

600k

single-particle events, μ / π / K
balanced

420k train / 90k validation / 90k test
Geant4 optical simulation
Track direction smeared by 5 mrad
1/p smeared by 0.5%
3.36 mm pixels

Noise augmentation

0–1 MHz/mm²

dark-count rate during training

70% fixed, 30% continuous levels
Same event with two noise draws
Consistency loss keeps β stable
between the draws

Evaluation

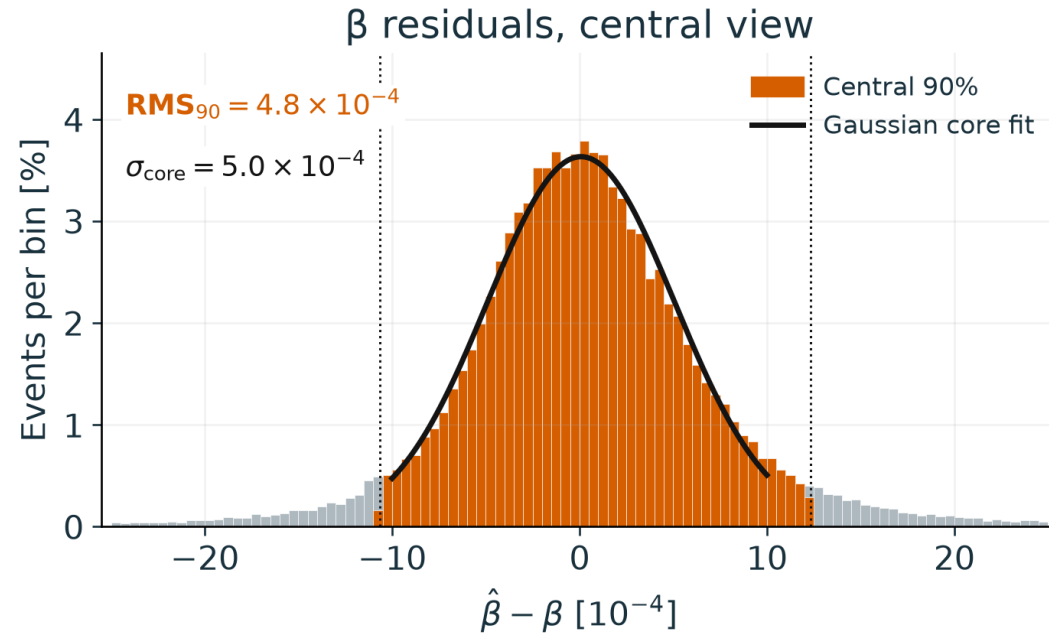
90k

test events, 30k per species

Labels from simulation: signal or dark
for each hit, true β and species for
each event
XGBoost comparison: 15k-event
subset with shared noise

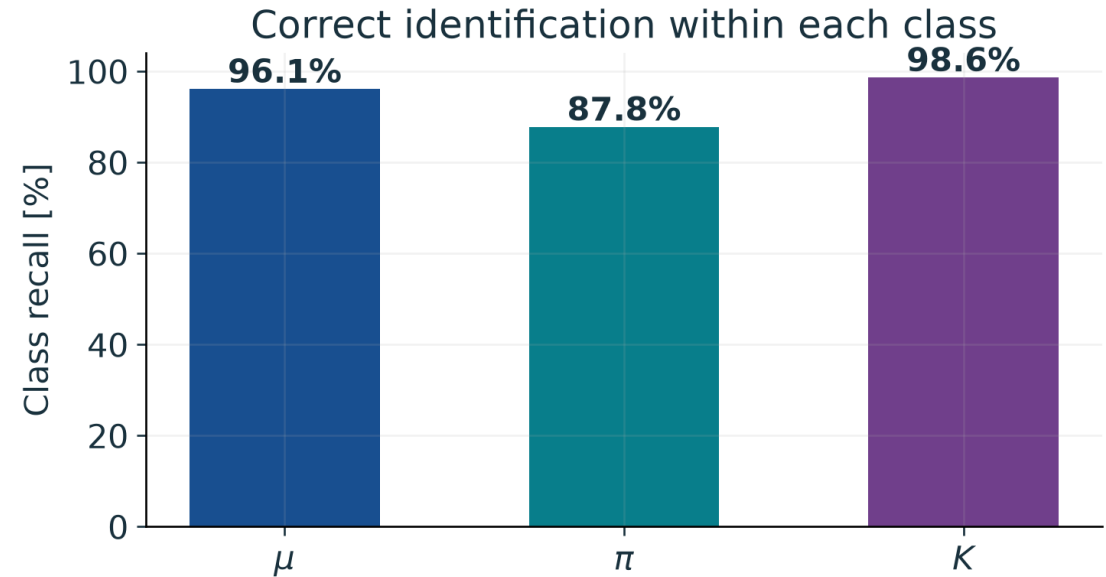
Reconstruction metrics

Comparison with earlier FARICH studies. Example: TCRAT at 1 MHz/mm².



$$\text{RMS}_{90} = \sqrt{\frac{1}{N_{90}} \sum_{i \in C_{90}} (\Delta\beta_i - \overline{\Delta\beta}_{90})^2}$$

C_{90} : central 90% of $\Delta\beta = \hat{\beta} - \beta$; N_{90} : its size.

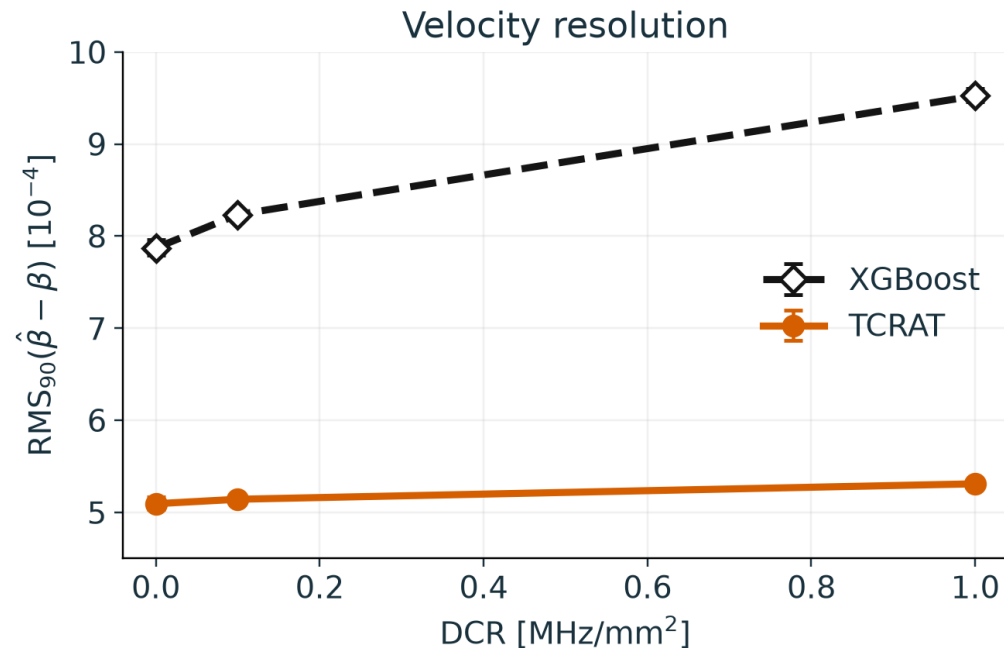


$$\text{BA} = \frac{R_{\mu} + R_{\pi} + R_K}{3}$$

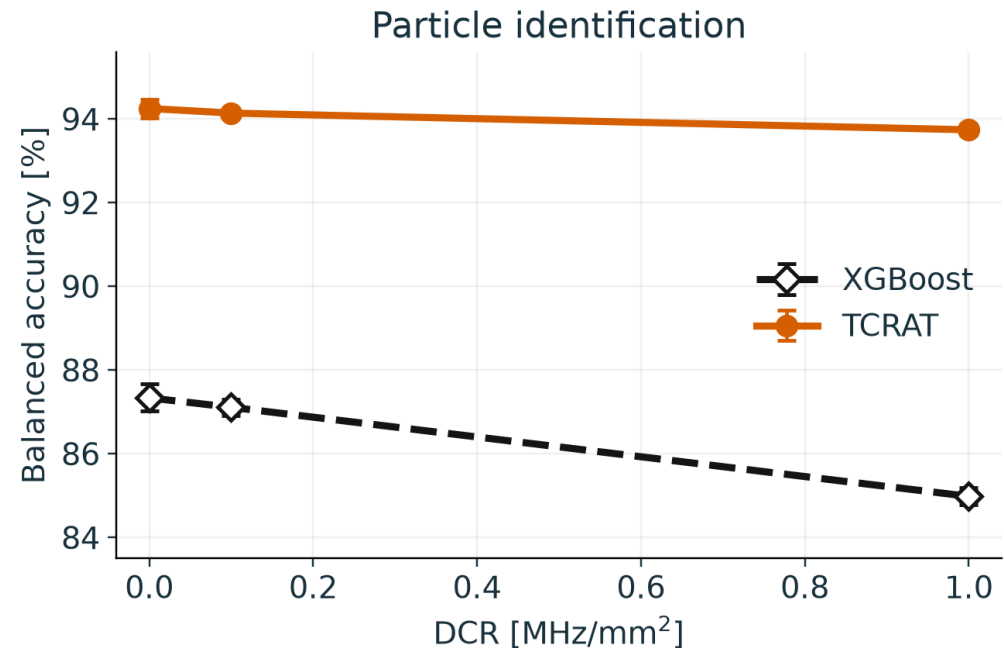
Recall: correct labels / all true particles of that species.
BA gives μ , π and K equal weight.

Reconstruction performance

XGBoost and TCRAT, matched events and identical noise realizations



35–44%
lower β RMS₉₀

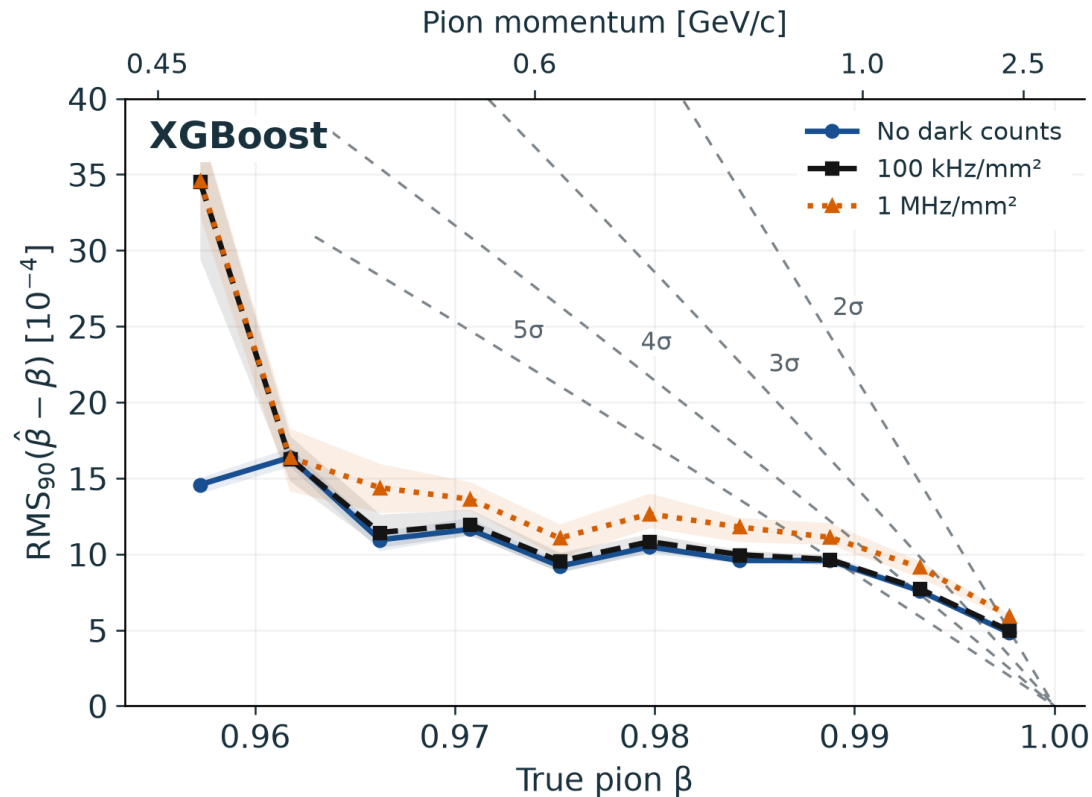
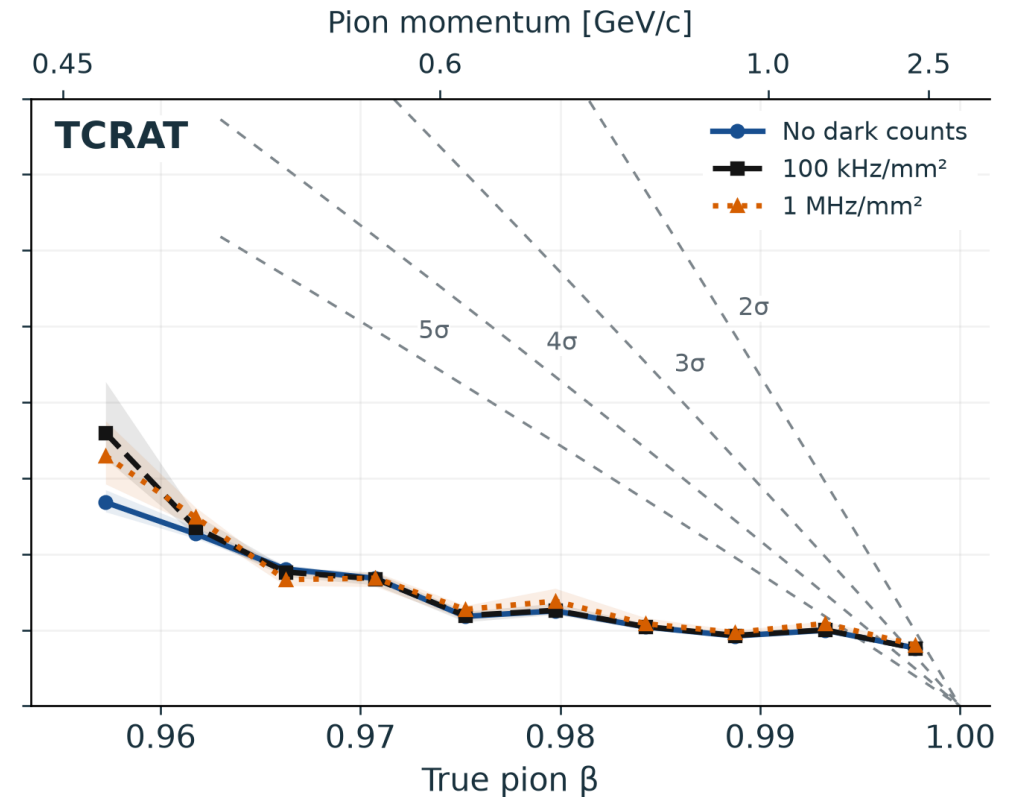


+6.9–8.8 percentage points
higher balanced PID accuracy

The β -resolution gain grows with the studied dark-count rate.

Pion β resolution

Ten disjoint true- β bins. Top axis: corresponding pion momentum.

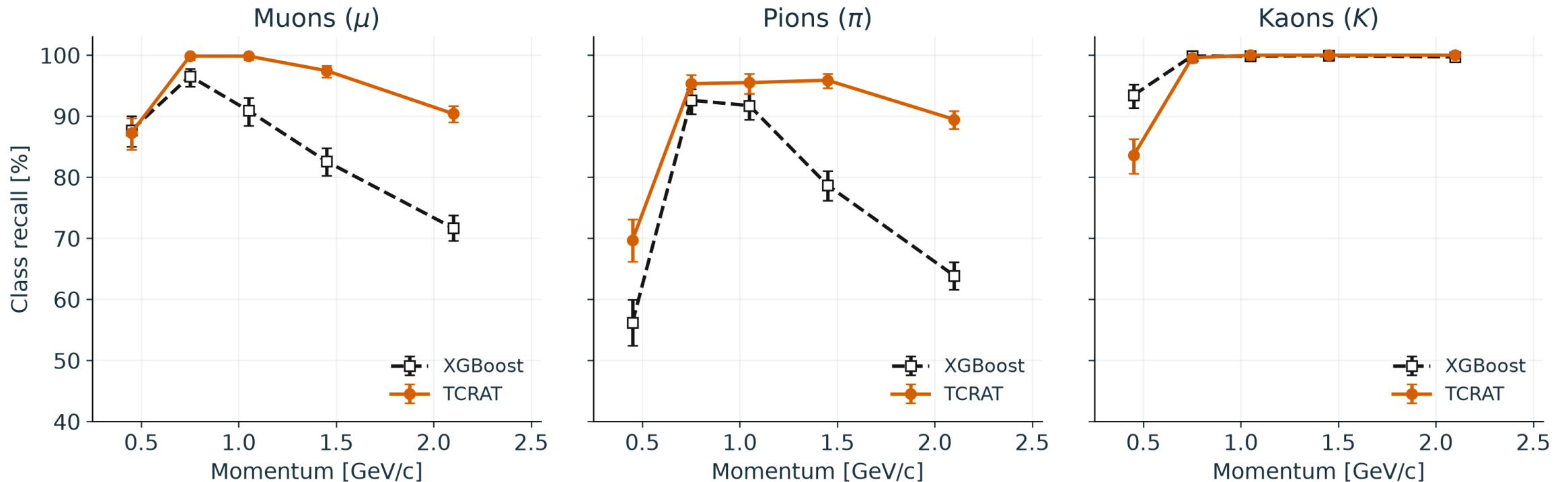


Gray curves: required β spread for 2–5 σ μ/π separation at equal momentum

Shading: realization min–max.

PID recall versus momentum

1 MHz/mm², matched events. Efficiency for each true particle species.



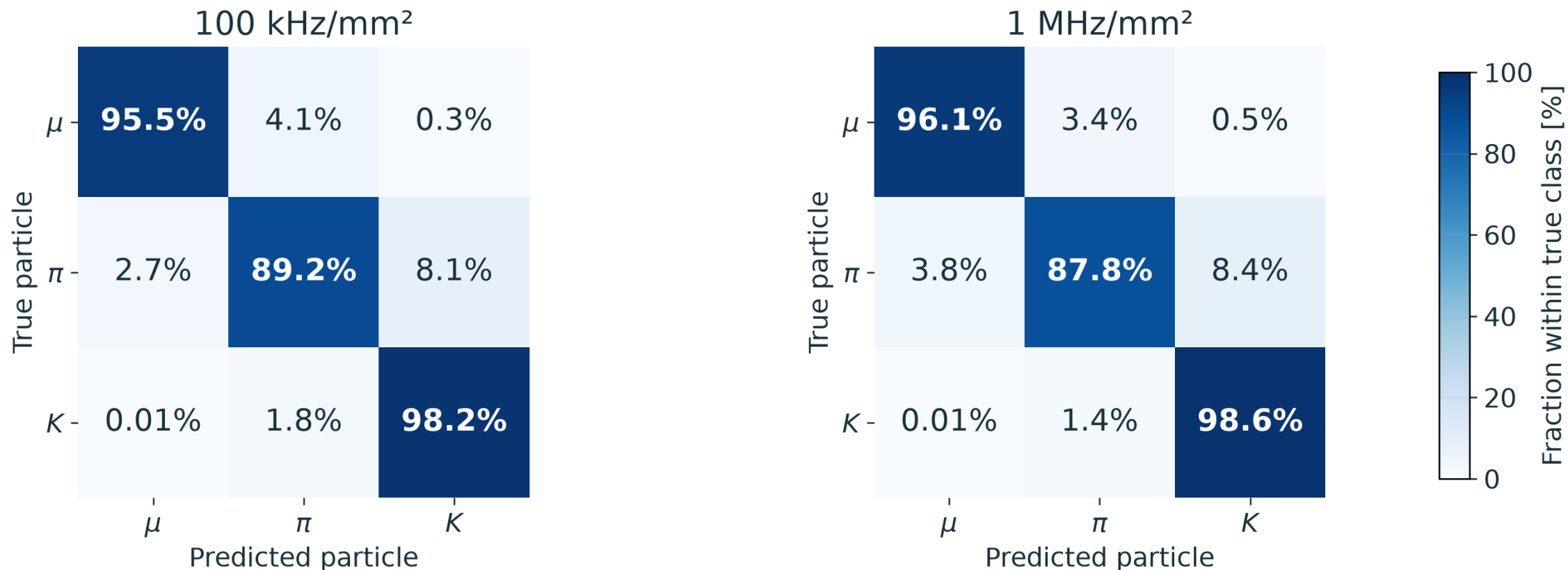
$$R_c(p) = \frac{N_{\text{correct}, c(p)}}{N_{\text{true}, c(p)}}$$

Denominator: all true particles of that species in the bin, including events without a ring.

Three-class argmax. Bars: conditional 95% count intervals.

PID confusion matrices

Rows: true species; columns: predicted.

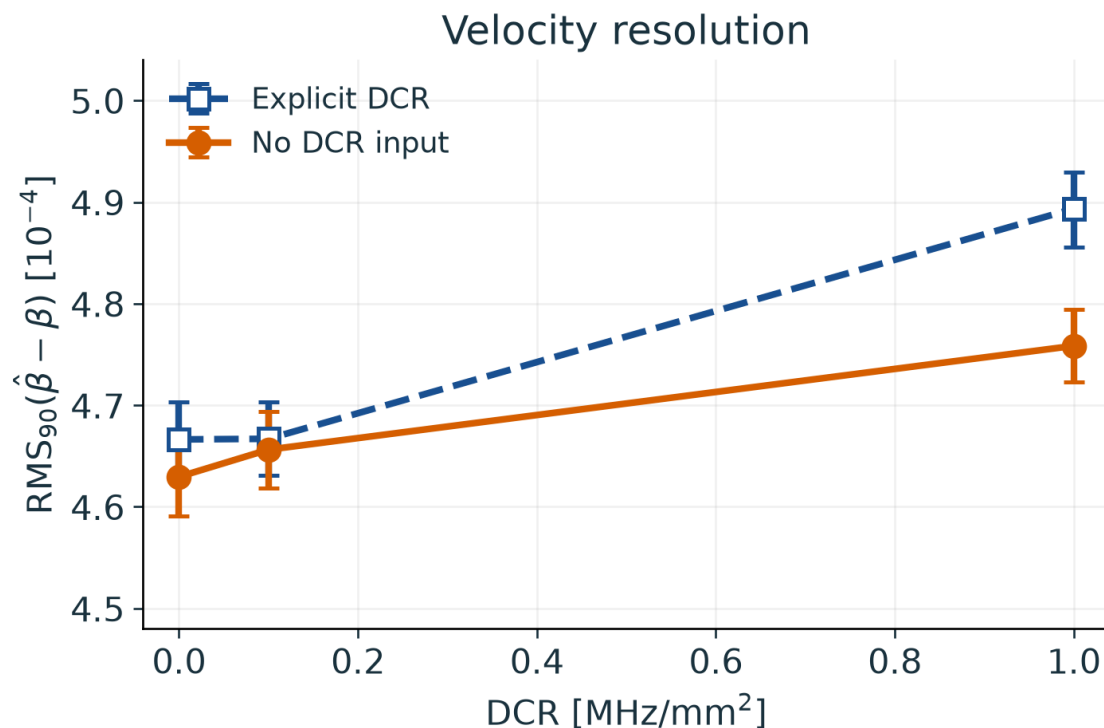


Signal-free candidate windows dominate $\pi \rightarrow K$ confusion

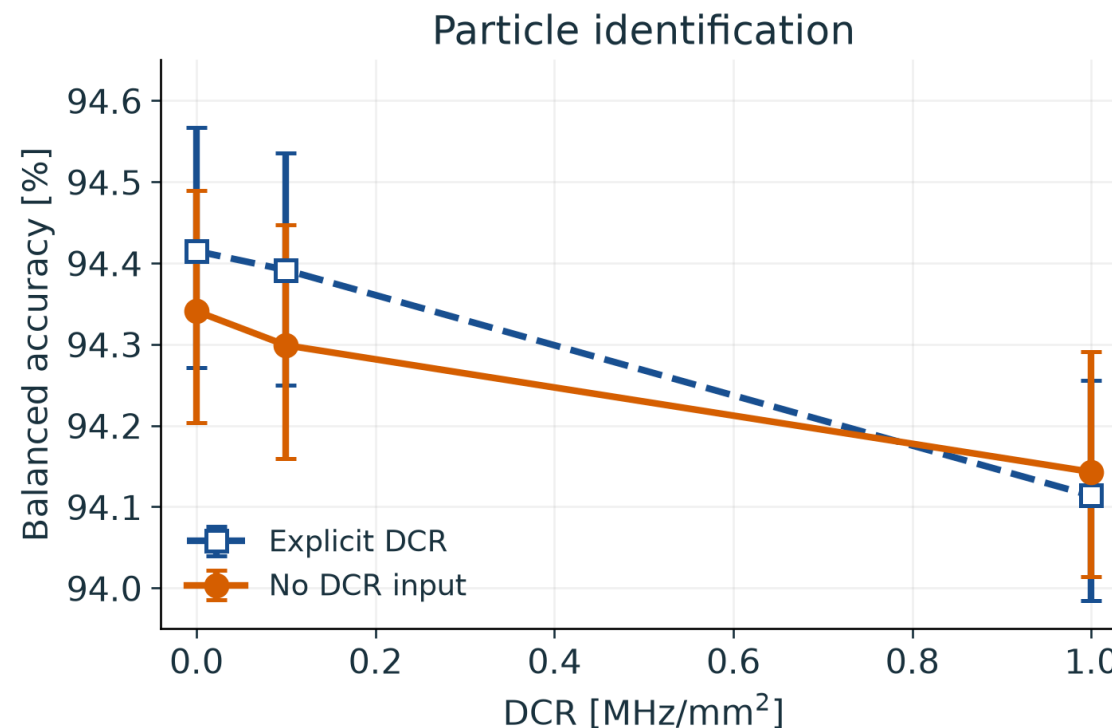
Opposite errors condition on different true species and need not be equal.

Reconstruction without a DCR input

Full event test set, mean of three paired training runs



$$\text{RMS}_{90} = (4.63\text{--}4.76) \times 10^{-4}$$



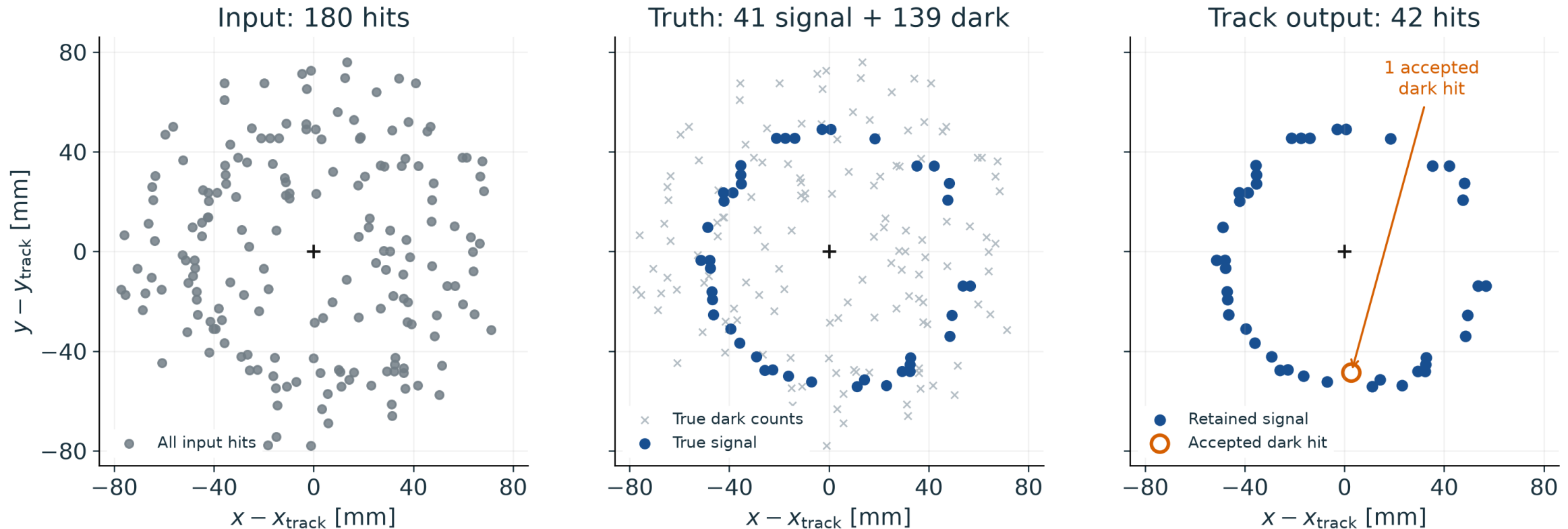
$$\text{Balanced accuracy} = 94.14\text{--}94.34\%$$

At 1 MHz/mm²: PID agrees within uncertainty, β RMS₉₀ is $\approx 2.8\%$ lower without DCR.

Bars: 95% event-bootstrap intervals.

Hit assignment: an event example

π , measured $p = 0.957$ GeV/c, 1 MHz/mm². Input: unlabelled hits + track.



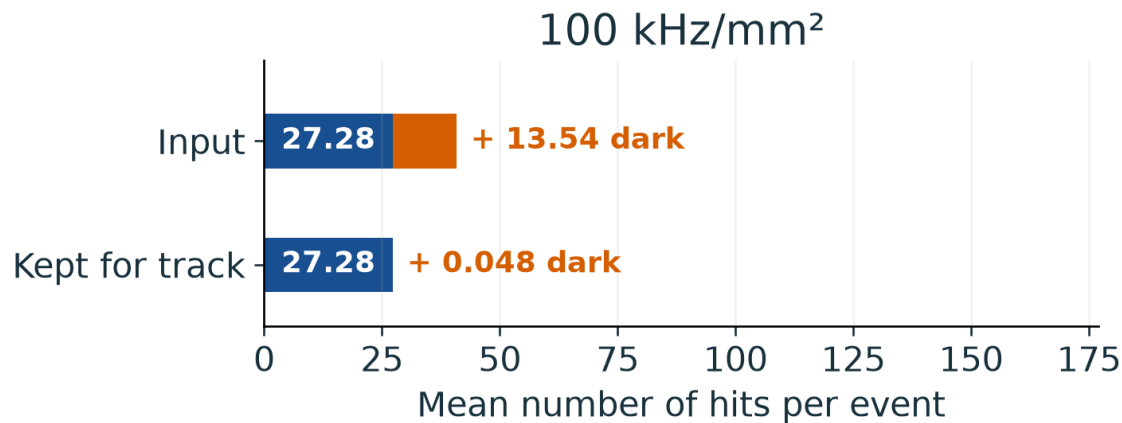
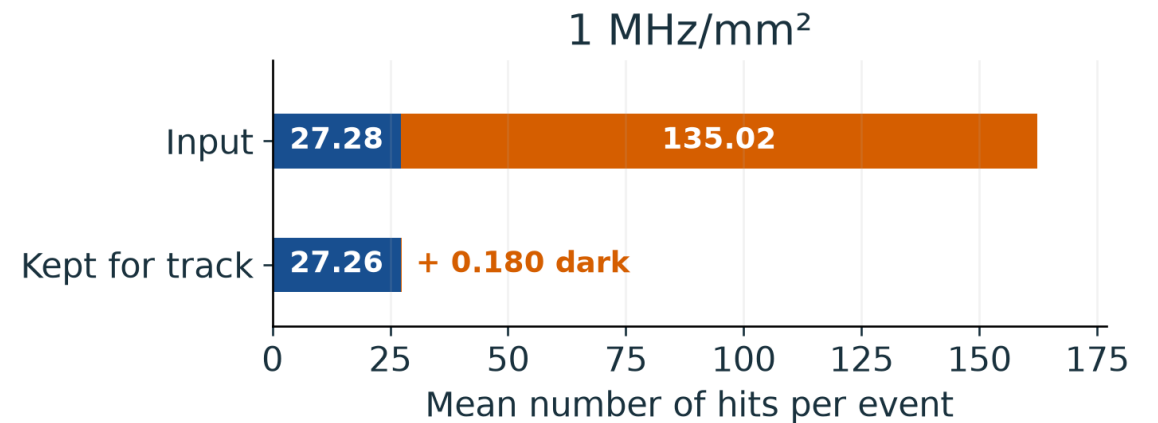
Track score ≥ 0.5 : keep the hit. Otherwise: assign it to background.

41 signal hits retained, 1 dark hit accepted, 0 signal hits lost

Hit-selection efficiency and purity

No-DCR model, threshold 0.5. Metrics below: 1 MHz/mm².

■ True signal ■ True dark counts



Signal efficiency

99.930%

95% CI: 99.925–99.935%

kept signal / input signal

Dark-hit acceptance

0.133%

95% CI: 0.131–0.135%

kept dark / input dark

Selected-hit purity

99.346%

95% CI: 99.335–99.357%

kept signal / all kept hits

Conclusions

vs XGBoost

35–44% lower β RMS90, +6.9–8.8 pp balanced accuracy
Matched events and identical noise realizations

Unknown DCR

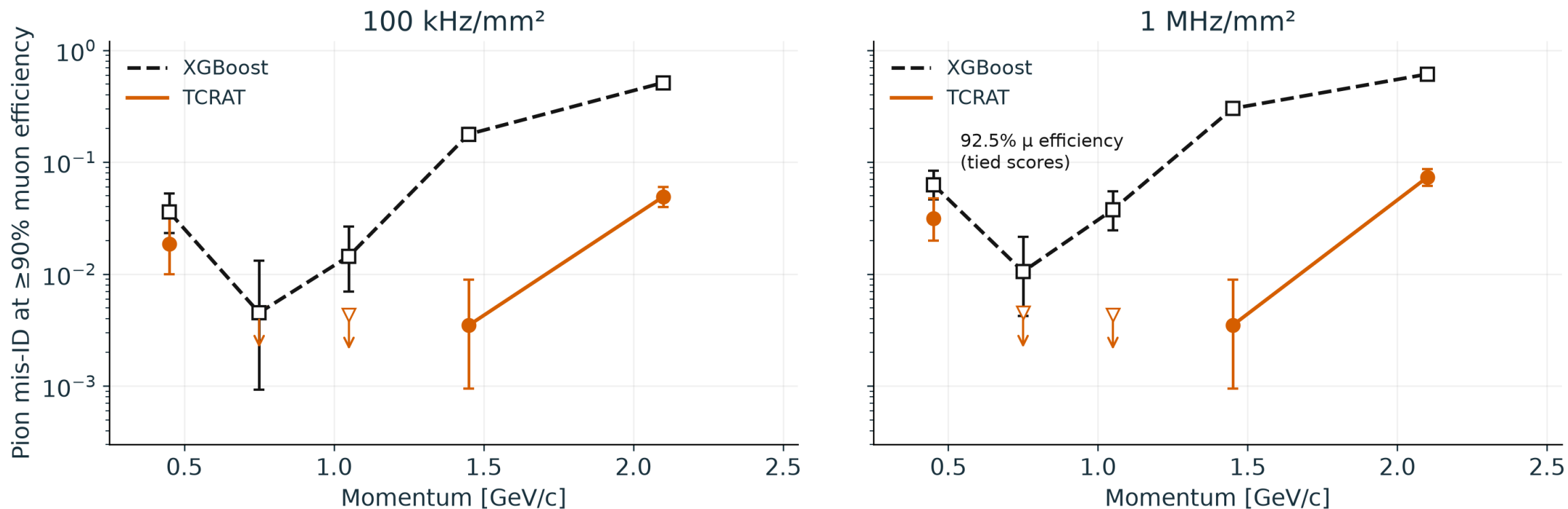
0–1 MHz/mm² with no DCR input: RMS90 $\approx 4.7 \times 10^{-4}$, PID $\approx 94\%$
Explicit DCR gives no demonstrated reconstruction advantage

Next steps

Measured sensor timing and realistic dark-noise maps
Validation beyond simulated single-particle events

Pion misidentification at $\geq 90\%$ muon efficiency

Matched XGBoost benchmark, one measurement replica and shared noise bank

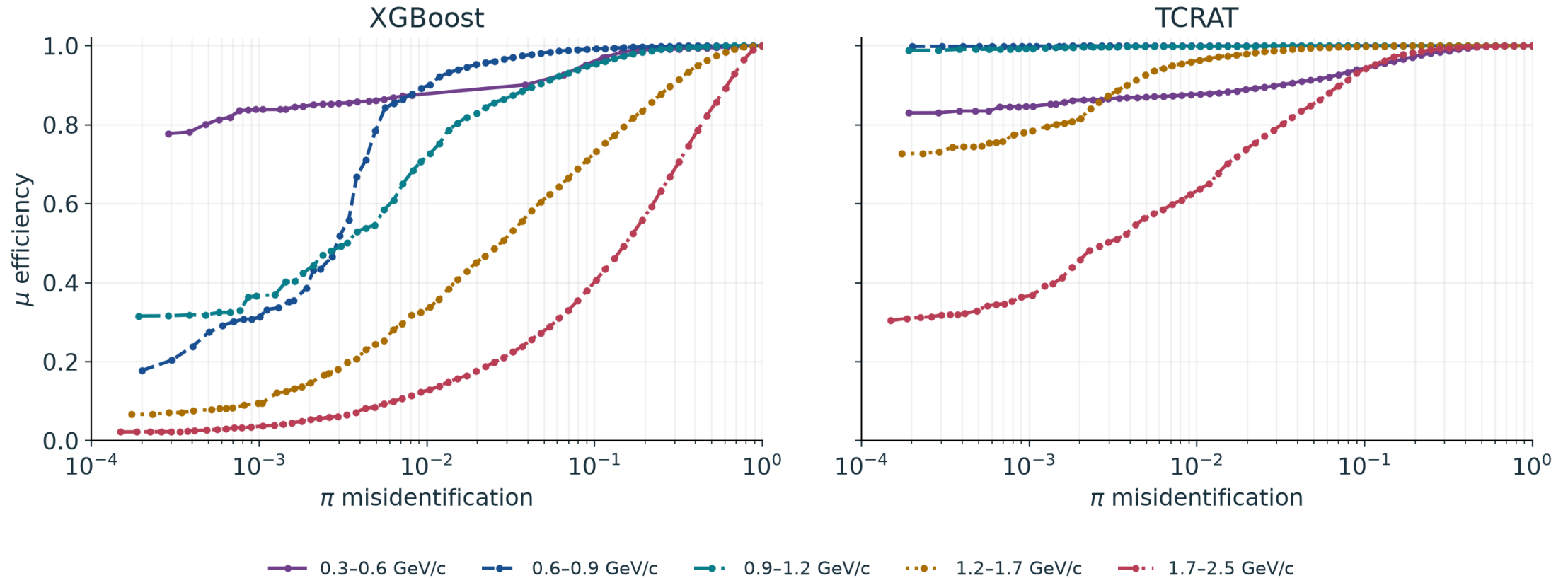


Open downward triangles: zero observed errors, shown as 95% upper limits.

Bars: conditional 95% count intervals at the observed threshold. Threshold uncertainty is not included.

Muon-pion separation at 1 MHz/mm²

Muon efficiency versus pion misidentification, in momentum bins



Better performance: higher efficiency at the same misidentification rate

Same benchmark as slide 10. Actual score thresholds, averaged over measurement/noise realizations.

Query, key and value projections

Within hit self-attention

$$a_{ij} = \text{softmax}_j \left(\frac{q_i^\top k_j}{\sqrt{16}} \right), \quad z_i = \sum_j a_{ij} v_j$$

Each query, key and value is a learned linear projection of a hit state.
Query/key dot products determine weights. Values carry the weighted information.
Four heads use 16 values each. Two layers update every hit representation.

Track assignment uses another set of projections:
track state gives the query, contextual hit states give 64-dimensional keys and values.

In the first formula, i labels the receiving hit and j the source hit. Attention sums to one over j .

Assignment and pooling equations

Track query and background query score each contextual hit

$$s_i^{\text{trk}} = q^T k_i / \sqrt{64} + b_i^{\text{geom}}, \quad s_i^{\text{bg}} = q_{\text{bg}}^T k_i / \sqrt{64} + b_i^{\text{bg}}$$

$$p_i = \frac{\exp(s_i^{\text{trk}})}{\exp(s_i^{\text{trk}}) + \exp(s_i^{\text{bg}})}$$

p_i : probability that hit i belongs to the track
 $1 - p_i$: probability of background
No normalization across different hits

$$c = \frac{\sum_i p_i v_i}{\max(\sum_i p_i, 10^{-6})}, \quad r = [u_{\text{track}}, c]$$

v_i : hit value vector
 c : weighted mean
 u : encoded track state

q : track query. Background query: learned vector. k_i : hit key. Biases use learned hit / relative geometry features.

Optimization details

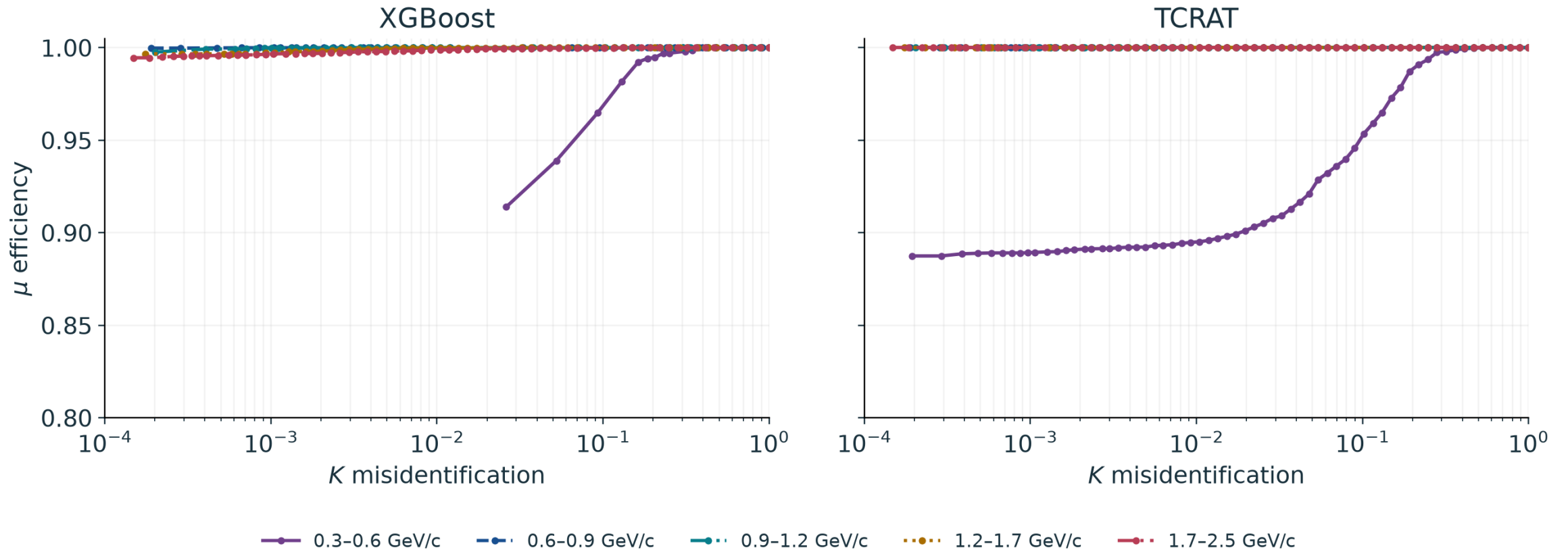
$$\mathcal{L} = \mathcal{L}_{\text{assign}} + 2\mathcal{L}_{\beta} + 0.5\mathcal{L}_{\text{PID}} + 0.20\mathcal{L}_{\text{consistency}}$$

Assignment	Hit-label cross-entropy, background weight 0.35
Velocity	Smooth-L1 on standardized β , available candidates
PID	Class-weighted cross-entropy, all events
Consistency	Two noise versions of the same event, weight 0.20
Optimizer	AdamW, learning rate 5×10^{-5} , weight decay 10^{-4}
Fine-tuning	4 epochs, batch 256, dropout 0.1, patience 2

35% of batches use noise pairs. PID gradients stop at the shared reconstruction representation.

Muon-kaon ROC at 1 MHz/mm²

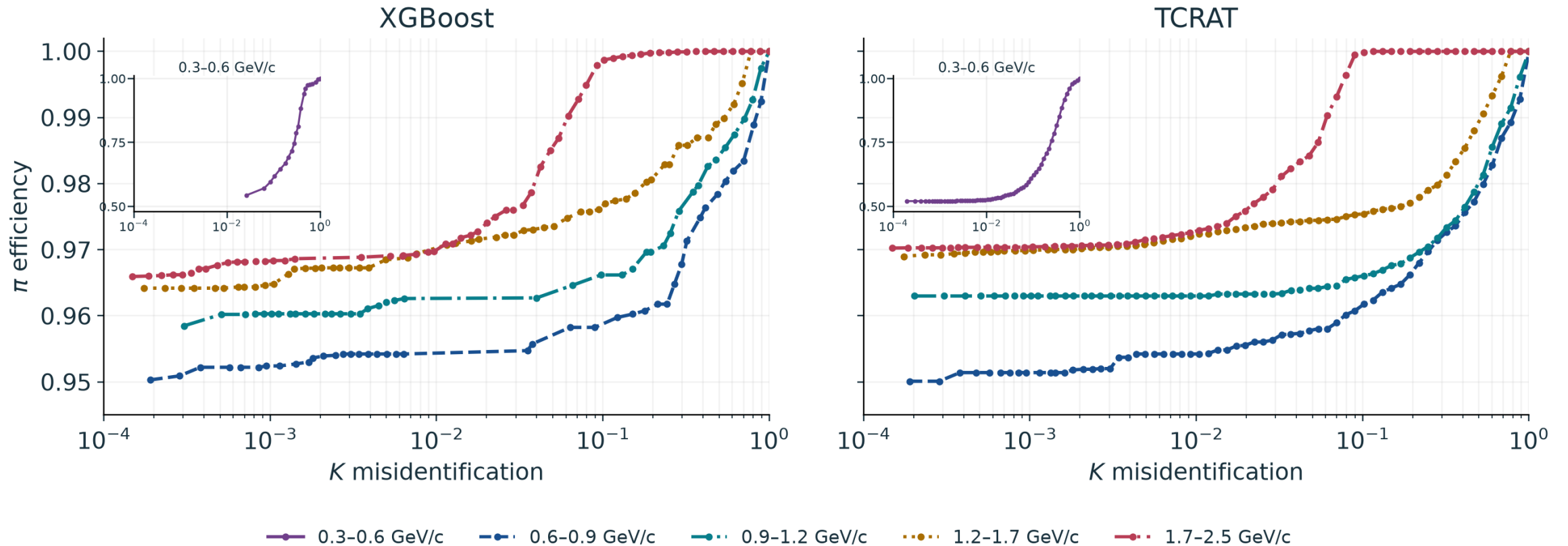
Muon efficiency versus kaon misidentification



Actual threshold points. Vertical scale starts at 80% efficiency.

Pion-kaon ROC at 1 MHz/mm²

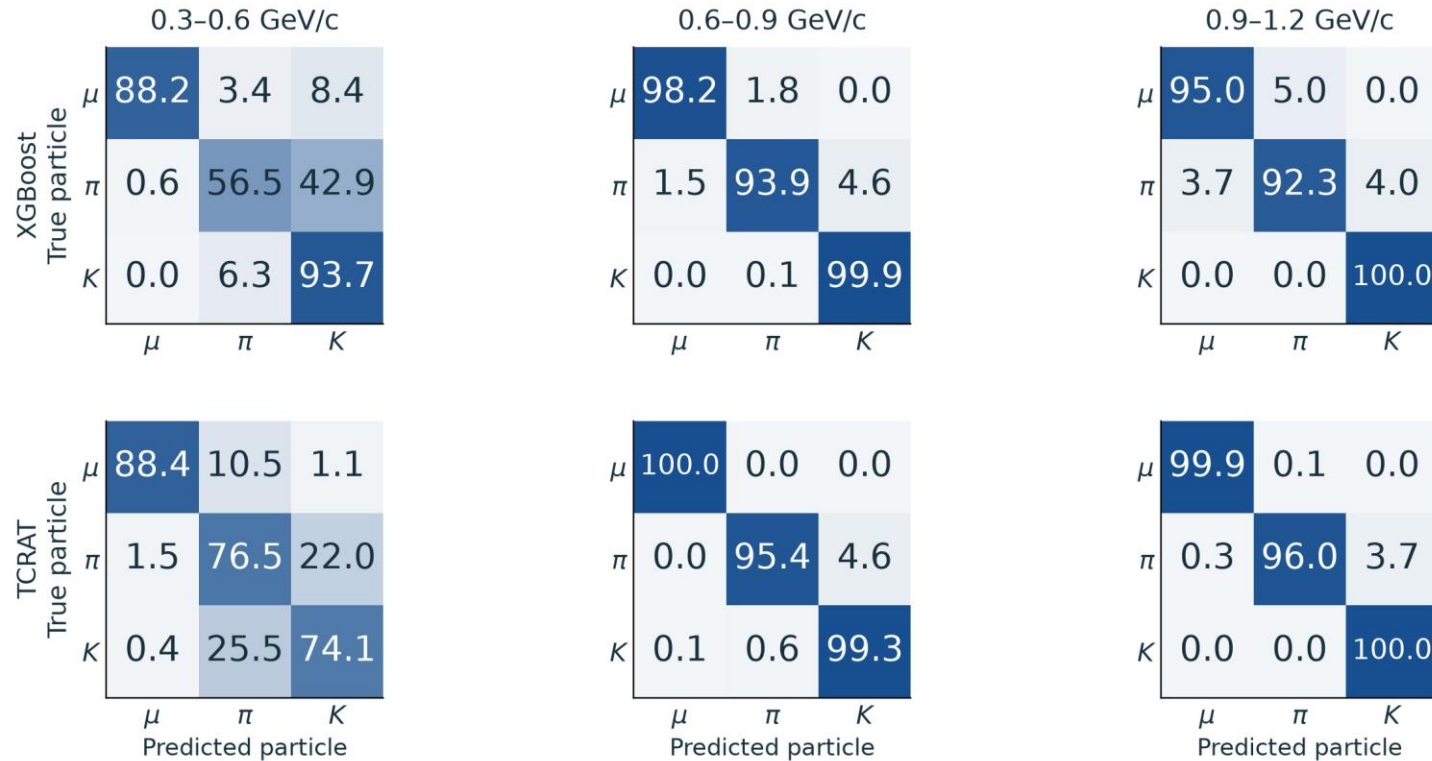
Pion efficiency versus kaon misidentification



Zoomed vertical axes. Insets retain the 0.3–0.6 GeV/c curves.

Confusion matrices: 100 kHz/mm²

0.3–1.2 GeV/c, row-normalized percentages



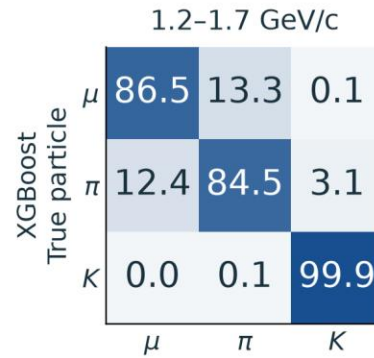
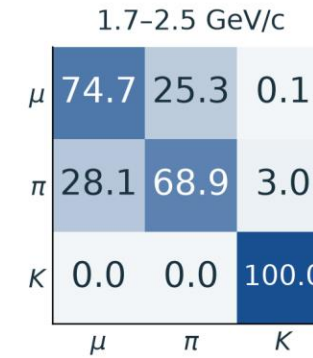
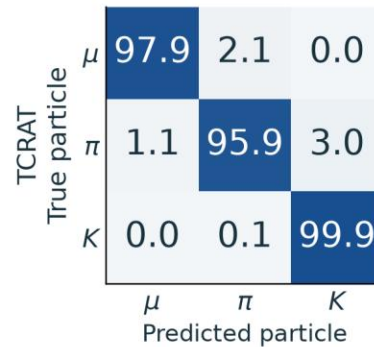
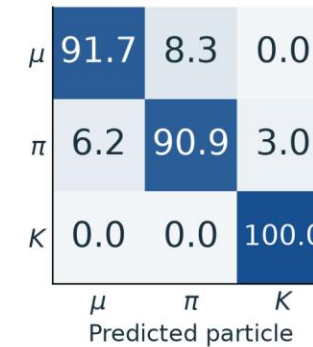
653–705 events per true class / bin (range across measurement replicas)

K recall at 0.3–0.6 GeV/c: XGBoost 93.7%, TCRAT 74.1%

Rounded cells: 0.0% does not establish zero misidentification.

Confusion matrices: 100 kHz/mm²

1.2–2.5 GeV/c, row-normalized percentages



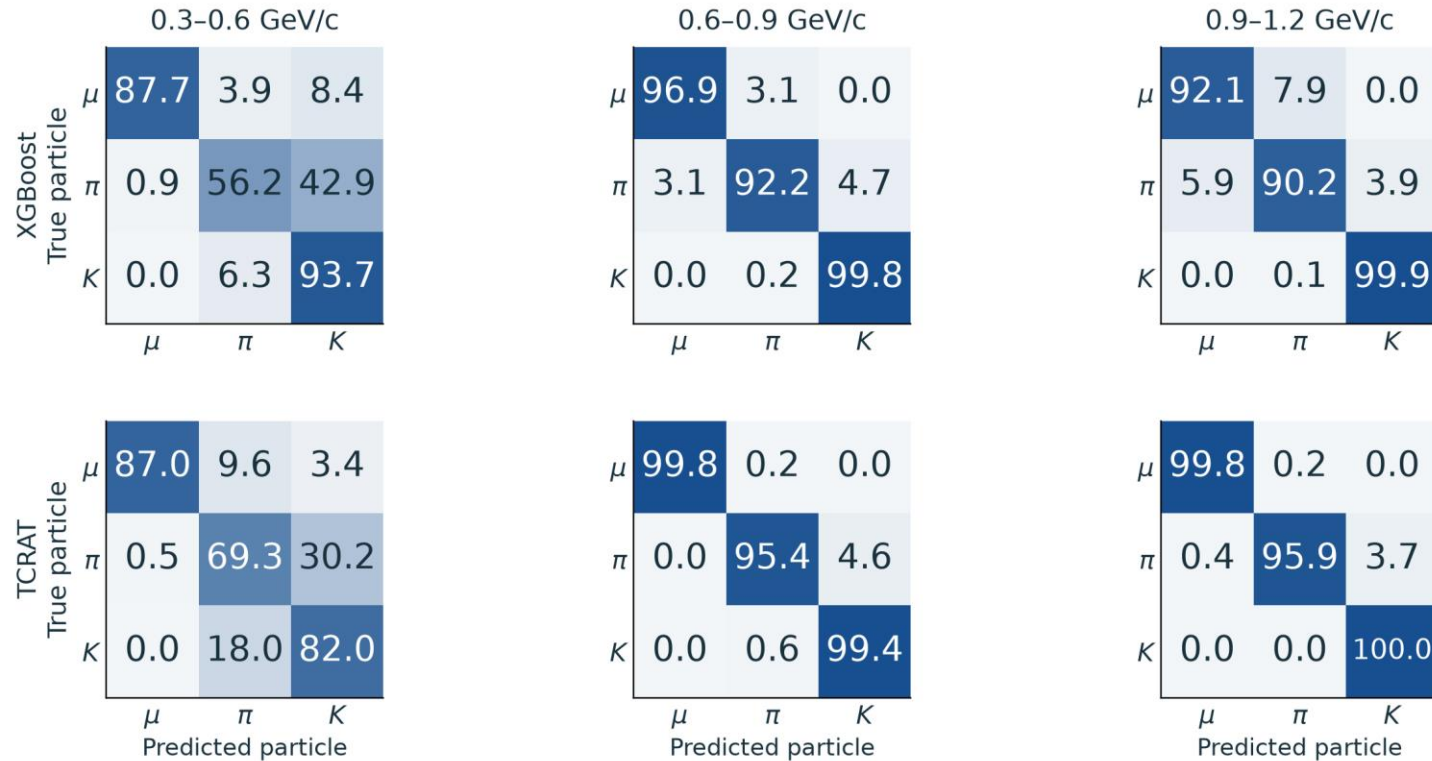
1104–1861 events per true class / bin (range across measurement replicas)

The aggregate PID gain includes particle-dependent tradeoffs.

Rounded cells: 0.0% does not establish zero misidentification.

Confusion matrices: 1 MHz/mm²

0.3–1.2 GeV/c, row-normalized percentages



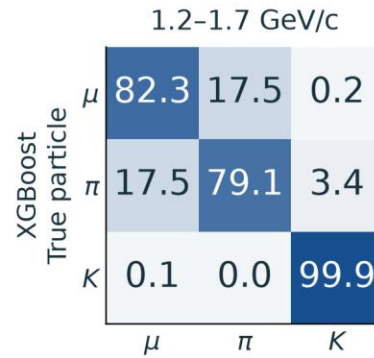
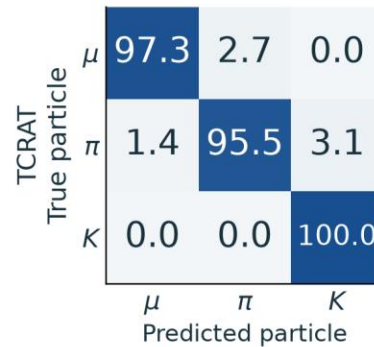
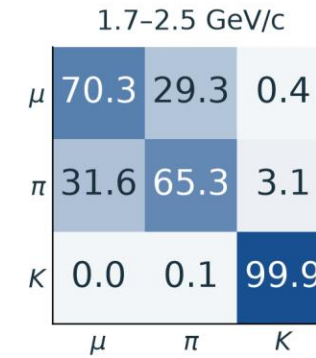
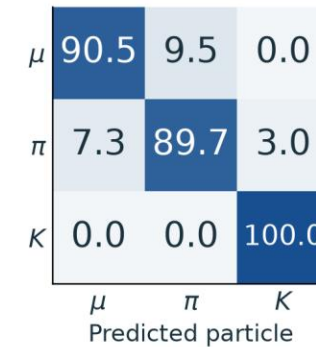
653–705 events per true class / bin (range across measurement replicas)

K recall at 0.3–0.6 GeV/c: XGBoost 93.7%, TCRAT 82.0%

Rounded cells: 0.0% does not establish zero misidentification.

Confusion matrices: 1 MHz/mm²

1.2–2.5 GeV/c, row-normalized percentages



1104–1861 events per true class / bin (range across measurement replicas)

The aggregate PID gain includes particle-dependent tradeoffs.

Rounded cells: 0.0% does not establish zero misidentification.

References

M. Chadeeva, P. Rogozhin, T. Uglov

Performance of the FARICH-based particle identification at charm superfactories using machine learning

Nuclear Science and Techniques 37, 124 (2026). doi:10.1007/s41365-026-01943-0

A. Yu. Barnyakov et al.

Calibration and Reconstruction Algorithm for FARICH System of the Detector at Super Charm-Tau Factory

Bulletin of the Lebedev Physics Institute 50, 534–539 (2023). doi:10.3103/S1068335623120047

M. Belozyorova et al.

Software framework for the Super Charm-Tau factory detector project

EPJ Web of Conferences 251, 03017 (2021). doi:10.1051/epjconf/202125103017

A. Vaswani et al.

Attention Is All You Need

Advances in Neural Information Processing Systems 30 (2017). arXiv:1706.03762

Training loss: terms and weights

$$\mathcal{L} = \langle \mathcal{L}_{\text{assign}} + 2\mathcal{L}_{\beta} + 0.5\mathcal{L}_{\text{PID}} \rangle_{\text{views}} + 0.20 \mathbf{1}_{\text{pair}} \mathcal{L}_{\text{cons}}$$

Supervised terms average over views. Paired batches: two noise versions of the same event.

Hit assignment

$$\mathcal{L}_{\text{assign}} = - \frac{\sum_i w_{y_i} \ln p_i(y_i)}{\sum_i w_{y_i}}$$

Cross-entropy for true hit labels.

Signal / dark weights: 1 / 0.35; exclude padding.

Particle identification

$$\mathcal{L}_{\text{PID}} = - \frac{\sum_e a_{c_e} \ln q_e(c_e)}{\sum_e a_{c_e}}$$

All events; inverse-frequency species weights.

PID gradients do not update reconstruction.

p: hit scores; q: PID scores; y: true hit label; c: true species; i: hit; e: event.

$$z = \frac{\beta - \mu_{\text{train}}}{s_{\text{train}}}, \quad h(r) = \frac{1}{2}r^2 \quad (|r| < 1), \quad h(r) = |r| - \frac{1}{2} \quad (|r| \geq 1)$$

Velocity reconstruction

$$\mathcal{L}_{\beta} = \frac{1}{|\mathcal{V}|} \sum_{e \in \mathcal{V}} h(\hat{z}_e - z_e)$$

Only events with an available β candidate.

Standardize β using the training mean and spread.

Noise consistency

$$\mathcal{L}_{\text{cons}} = \frac{1}{|\mathcal{V}_{12}|} \sum_{e \in \mathcal{V}_{12}} h(\hat{z}_e^{(1)} - \hat{z}_e^{(2)})$$

Paired views in 35% of training batches.

Require an available β candidate in both views.