Global ~~electron~~ particle identifier(s) for CBM

Pavish Subramani
University of Wuppertal, Germany

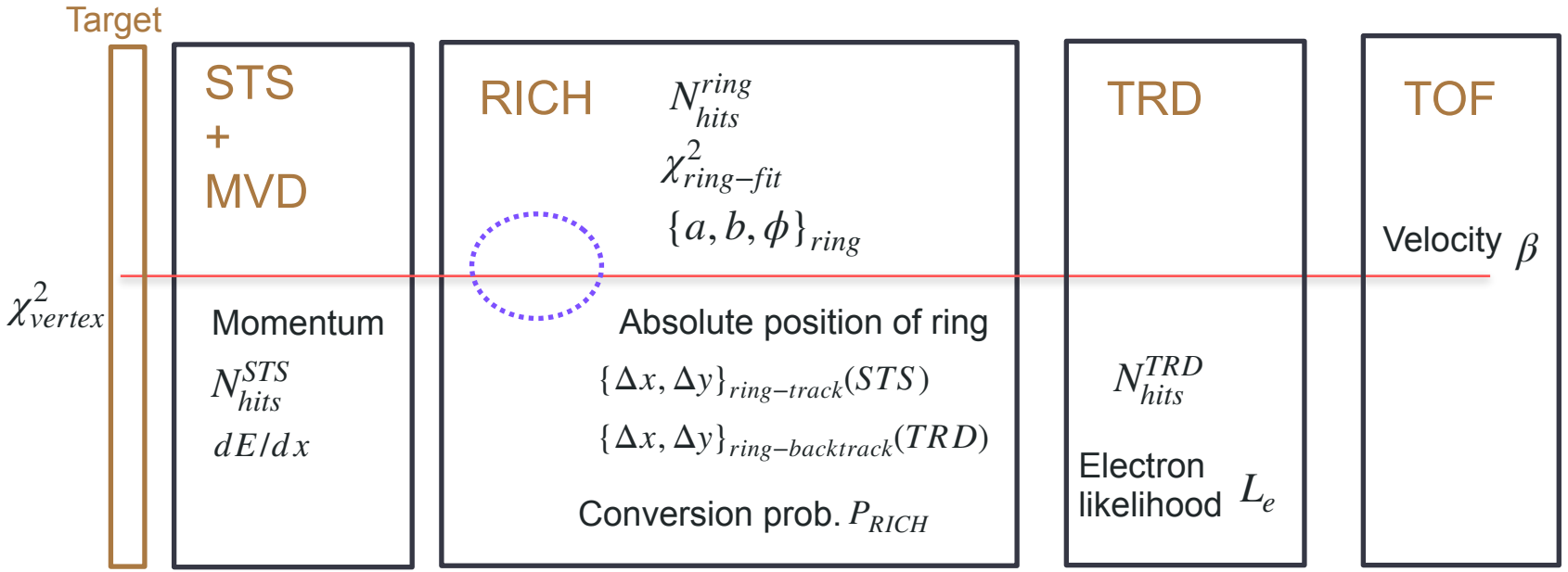
CBM Collaboration meeting, Oct.2025

BERGISCHE UNIVERSITÄT WUPPERTAL

Electrons

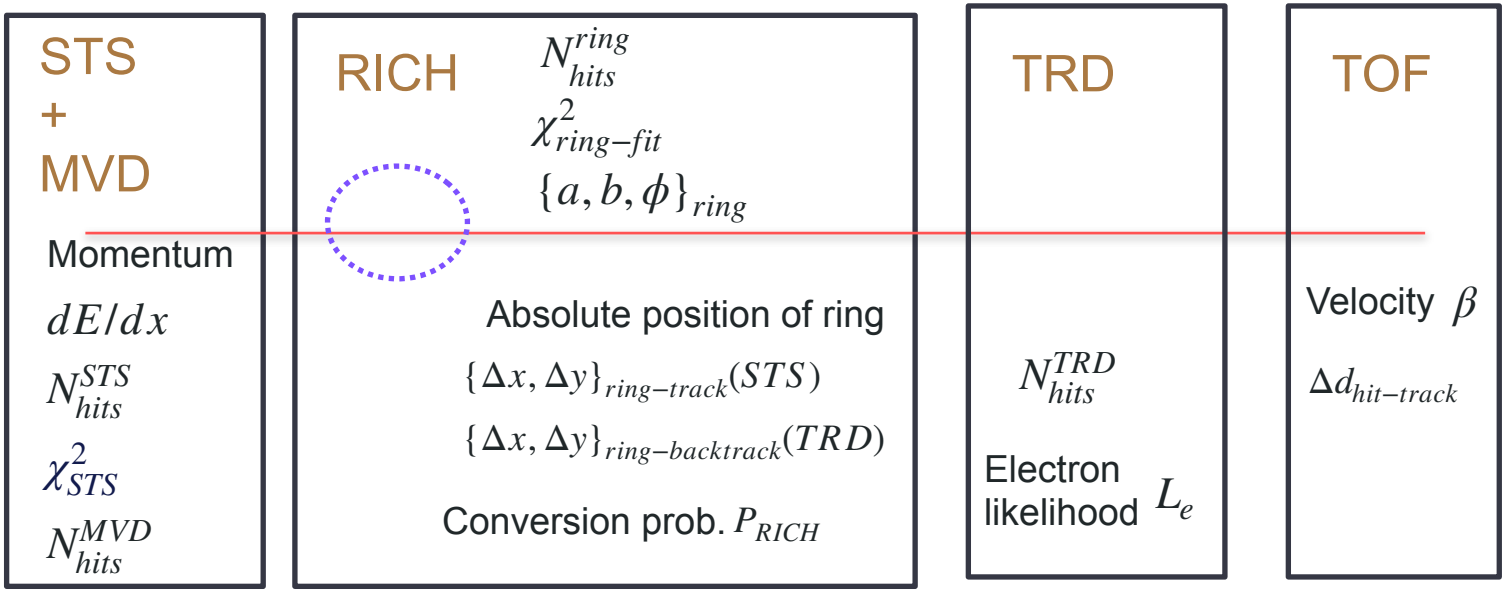
Status last CM: Global electron identifier

- Combining the electron identification performance of the RICH+TRD+TOF+STS.
- Objective: **Primary electron vs others**



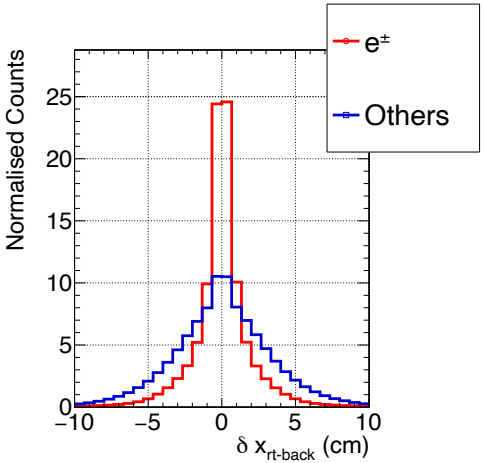
Global electron identifier:

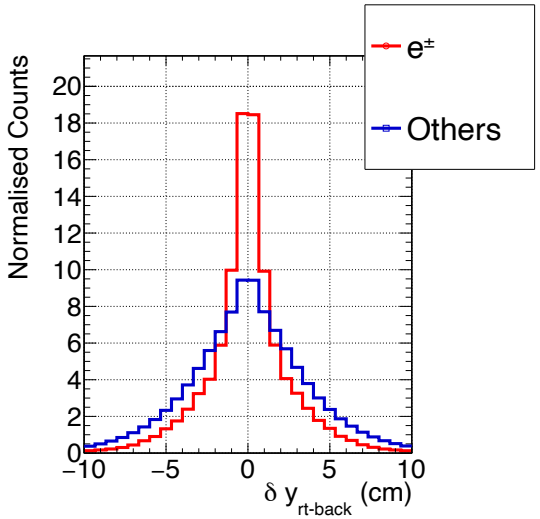
- Combining the electron identification performance of the RICH+TRD+TOF+STS.
- Objective: **Electron vs others** -> ID secondary electrons, and then to reject them using topological cuts.

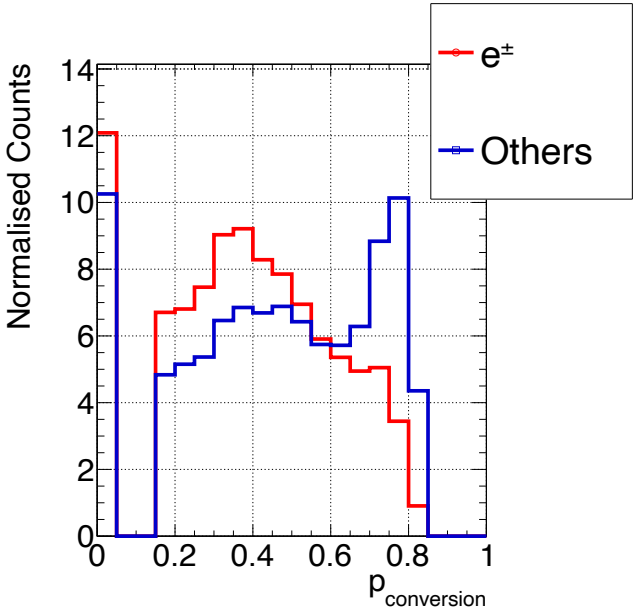


Realisation:

- **Geometry: Day 1 except TRD -> Use of TRD 2D (v23b)**







Ring-backtrack distances:

Global track fitted in TRD and extrapolated back to RICH

Each ring is tagged with a probability that it is a ring from untracked conversion electron tracked in TRD by searching TRD hits which are not matched to any global track.

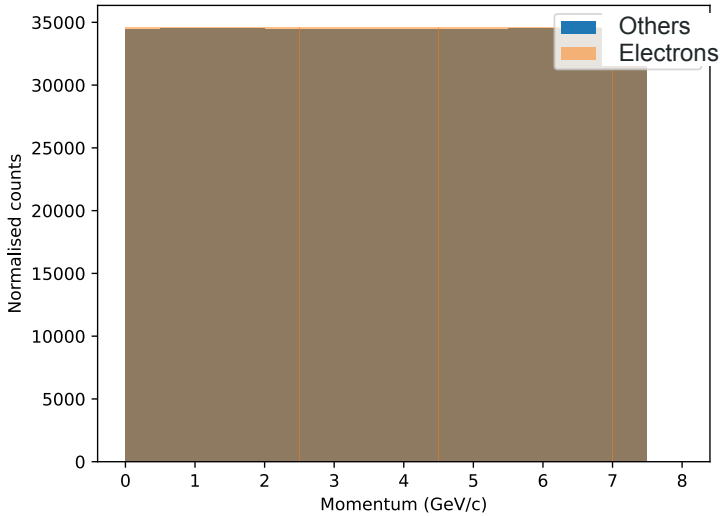
https://indico.gsi.de/event/19534/contributions/82142/attachments/48620/80235/electron_id.pdf

Realisation:

- Geometry: Day 1 except TRD -> Use of TRD 2D.
- Testing: $\{inmed - \rho, \omega, \phi, qgp\} \rightarrow e^+e^-$ + UrQMD central Au-Au @ 8A GeV/c beam momentum.
- Training: Single electrons (increased statistics for low momenta) + UrQMD Au-Au @ 8A GeV/c beam momentum.

Realisation:

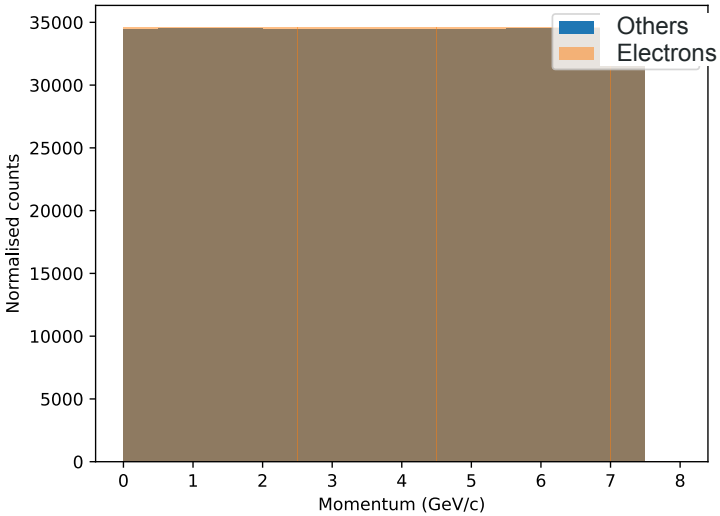
- Geometry: Day 1 except TRD -> Use of TRD 2D.
- Testing: $\{inmed - \rho, \omega, \phi, qgp\} \rightarrow e^{+}e^{-}$ + UrQMD central Au-Au @ 8A GeV/c beam momentum.
- Training: Single electrons + UrQMD central Au-Au @ 8A GeV/c beam momentum.
- ML model should be used for all energies and centralities -> Momentum independent learning.



Momentum distribution of training sample

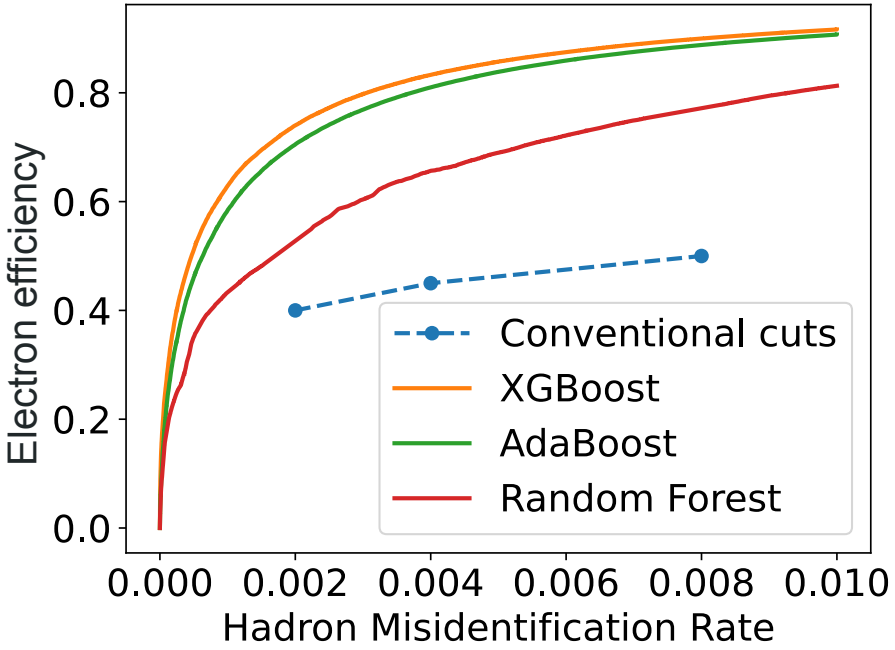
Realisation:

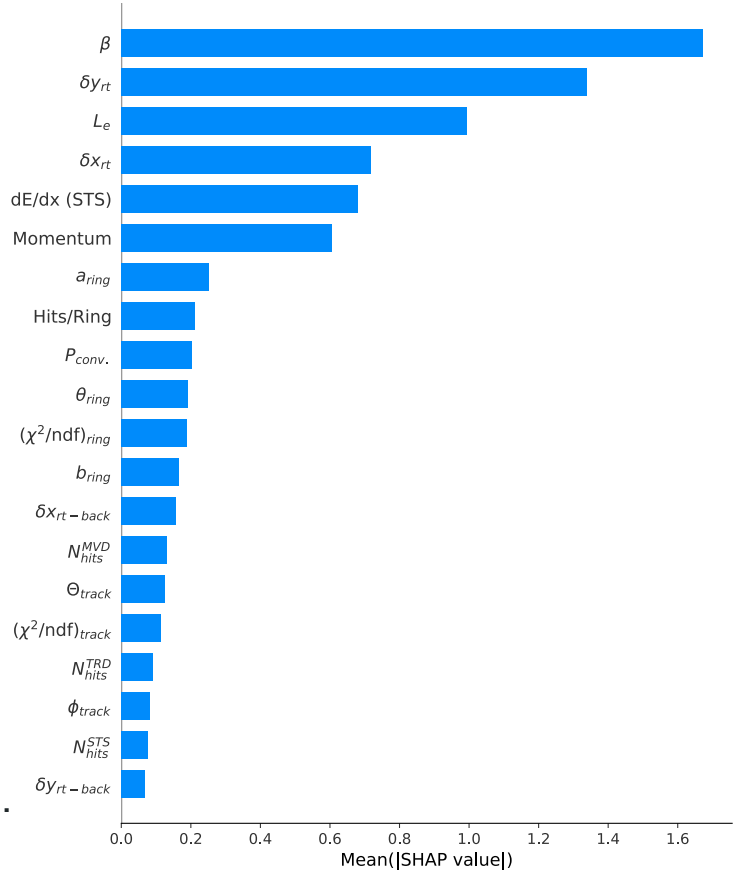
- Geometry: Day 1 except TRD -> Use of TRD 2D.
- Testing: $\{inmed - \rho, \omega, \phi, qgp\} \rightarrow e^+e^-$ + UrQMD central Au-Au @ 8A GeV/c beam momentum.
- Training: Single electrons + UrQMD central Au-Au @ 8A GeV/c beam momentum.
- ML model should be used for all energies and centralities -> Momentum independent learning.
- 3 models: Ensemble of decision trees
  - Bagging: Random forest
  - Gradient boosting: xGBoost
  - Adaptive boosting: AdaBoost
- 500×10^3 electrons, 500×10^3 others.
- Hyper-parameters are optimised by a combination of grid and random search methods.



Momentum distribution of training sample

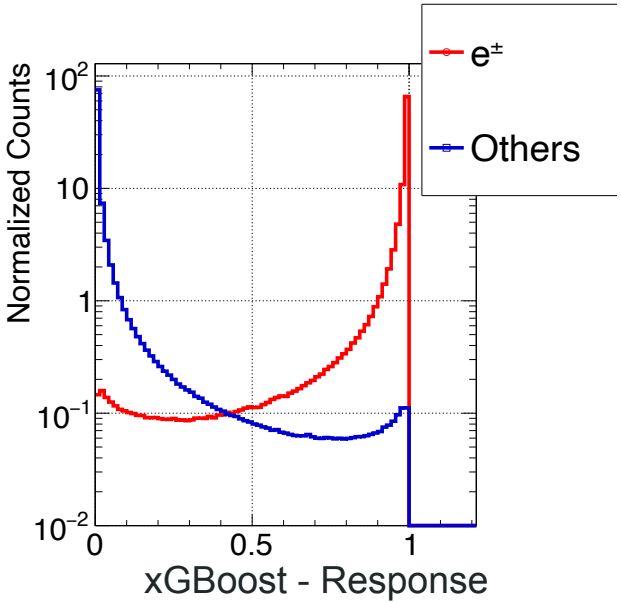
Results:

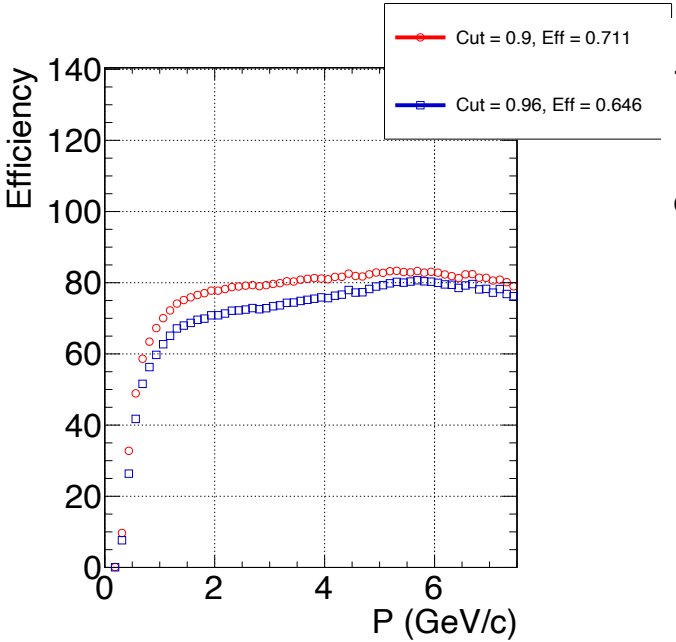


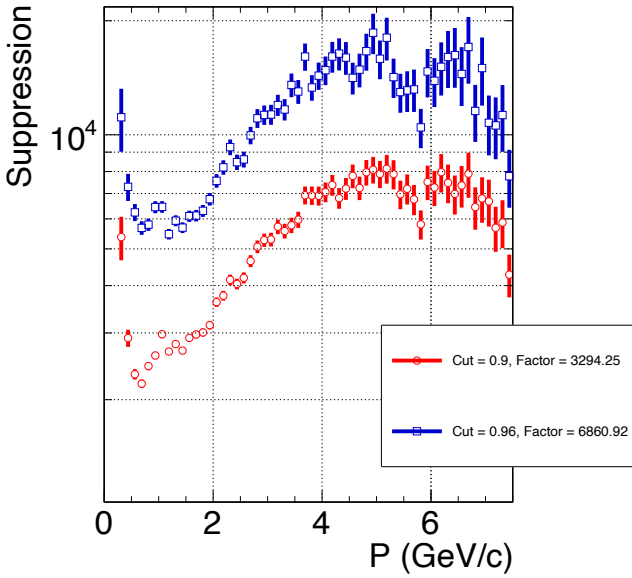


- xGBoost performs better for all efficiencies.
- All variables have influence on the model performance.

Results:



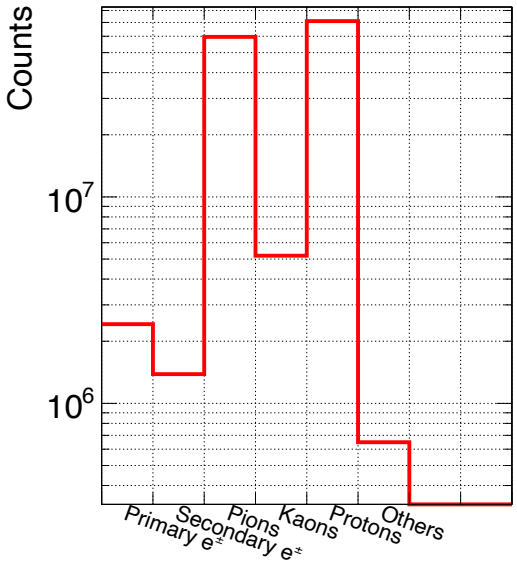




- xGBoost model is imported into ONNX format and integrated into CBMROOT simulation.
- Good response -> can be used by cutting on the response values.

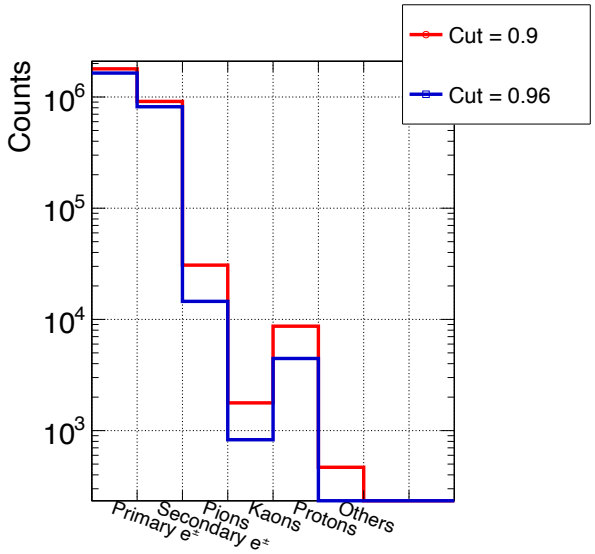
Results:

Accepted tracks



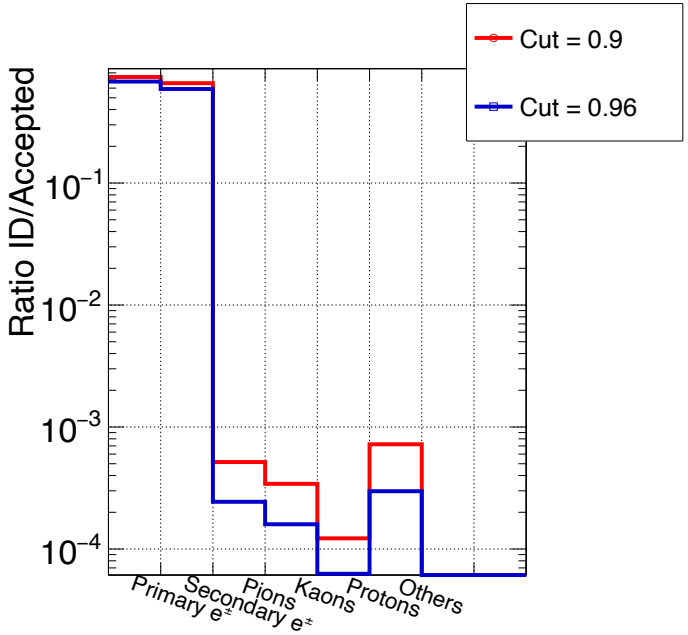
Number of events = 1×10^6

Identified tracks



Primary $e^{\pm}$:
LMVM + QGP (unweighted)

Ratio



Proton suppression >10^4
Pion suppression > 6×10^3

Accepted:
Reconstructed track + 2 TRD hits + 1 TOF hits

Outlook:

- Understand the response in different momentum regime -> Scrutinise input features.
- Test the model for different collision setups and energies.
- Possibly test on “Anything can go wrong” data sample.

  - Broader ring track distance distribution.

  - Lower momentum resolution.

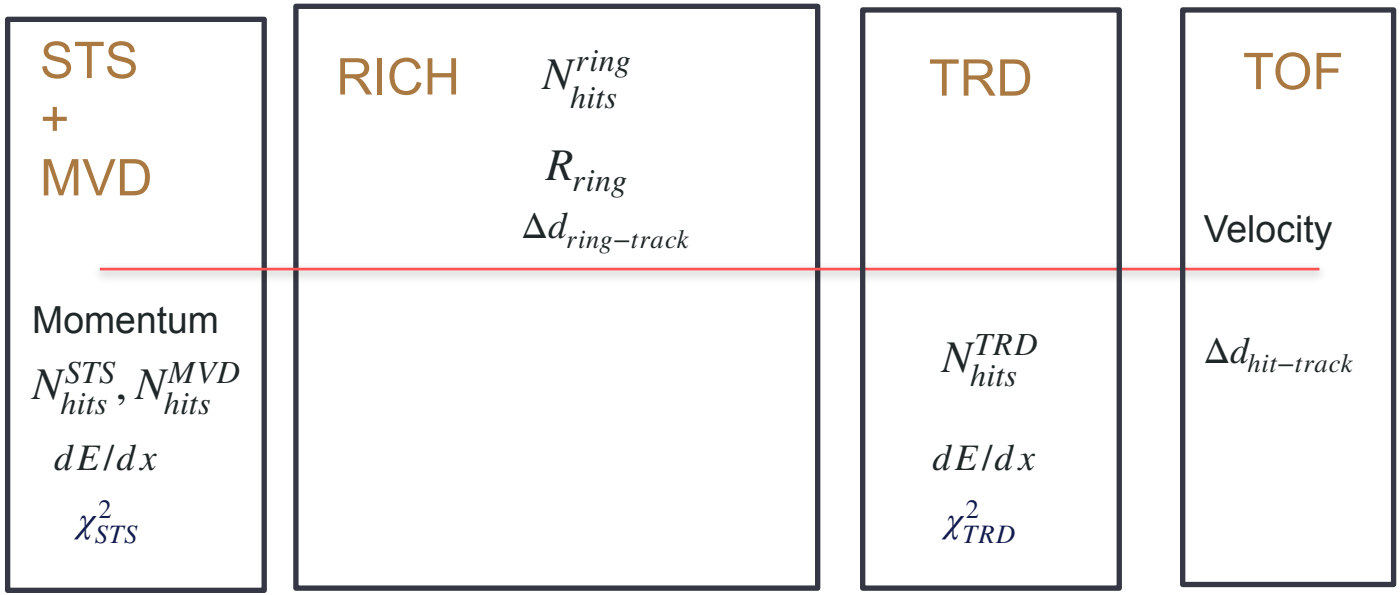
- Need for a portable tracker for TRD -> Standalone tracks in TRD

  - Important for backtracking, and help in reducing mismatches in TOF.

Hadrons

Global hadron identification scheme:

- No standard hadron identification -> especially combining all detectors.
- In some analysis, 1D or even 2D cuts are employed.
- Need for a global hadron identification scheme, similar to the one we developed for electrons.



Global hadron identification scheme:

- Application of machine learning is foreseen.
- Unlike simple electron vs others, implementation could be complicated due to multi-class classification.
- Preliminary Strategy:
  1. Using RICH for pion ID for tracks with momenta above 4.7 GeV/c and for electron veto.
  2. Different strategy for kaons (?).
- Ideas:
  1. DNN -> Multiple outputs, easy to build, usually model-heavy, slow.
  2. Decision tree ensembles: Multiple models. Light, fast, and better efficiency (inspired by LHCb).
- System and energy independent training.

Work in progress!

Thank you

Conventional cuts:

MVD hits: 0-4 (integer)

STS hits: 0-8 (integer)

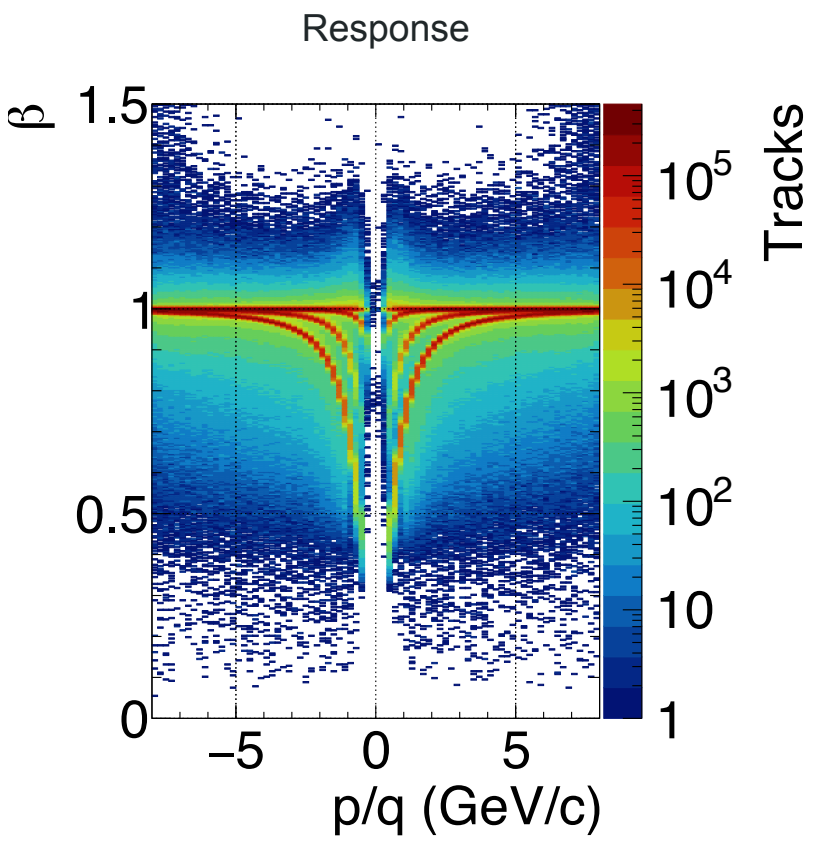
STS dE/dx = 20-40 (60 bins)

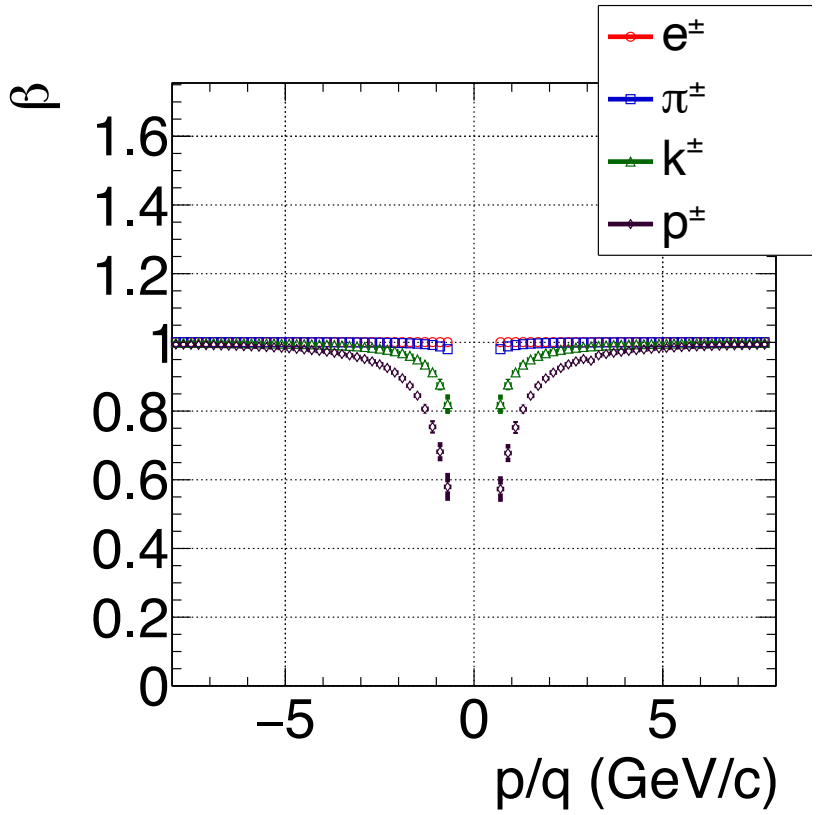
Mass squared : 0-0.1 (100 bins).

Likelihood: 0.3-1 (70 bins)

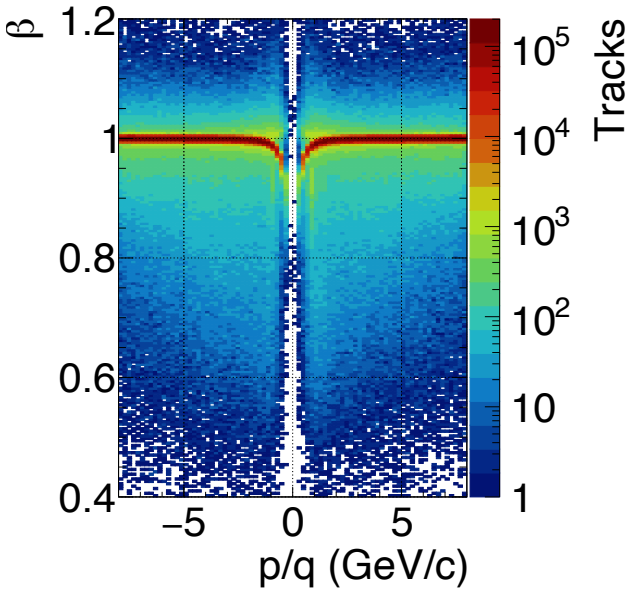
RICH xGBOOST response : 0.3-1 (70 bins)

TOF response:

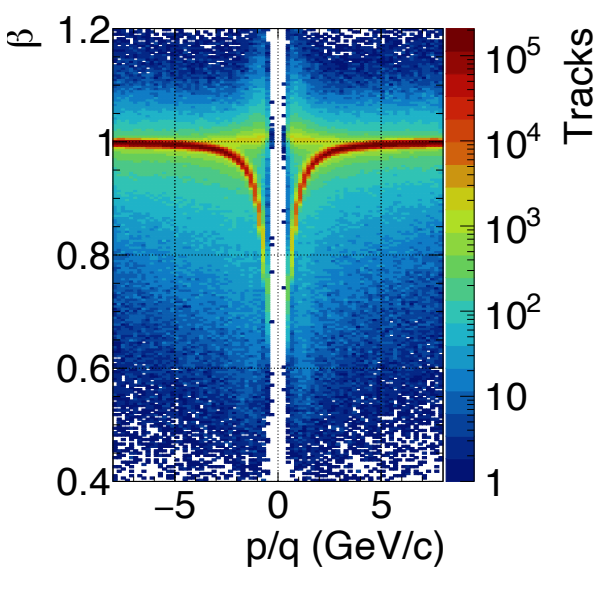




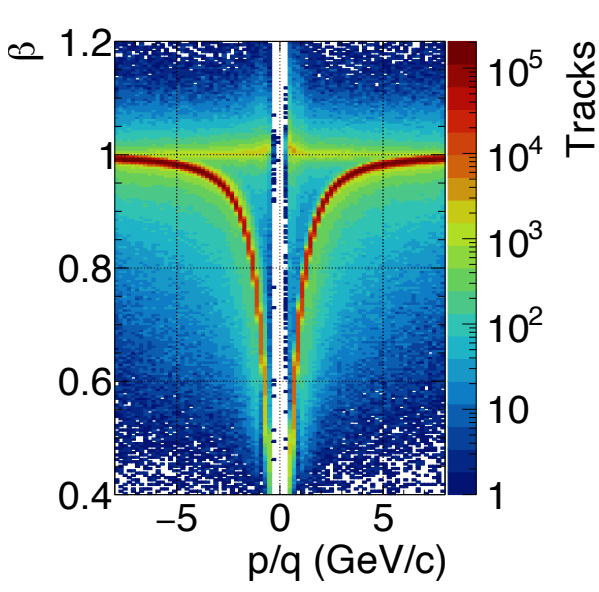
Individual response:



Pions

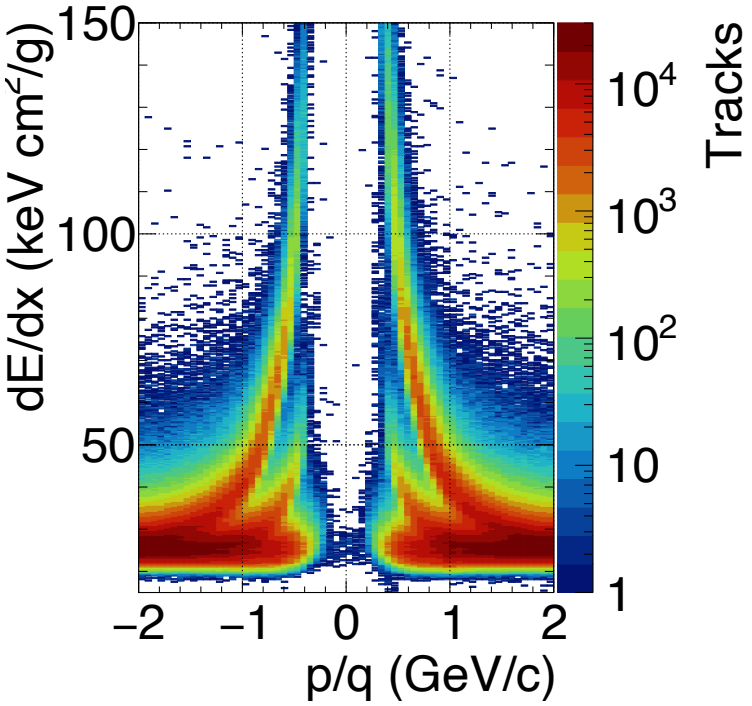


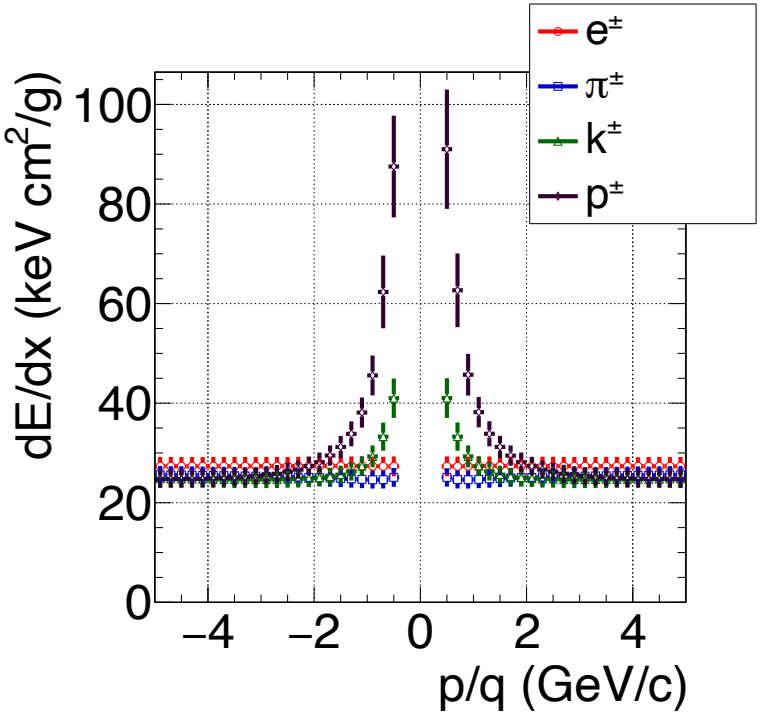
Kaons



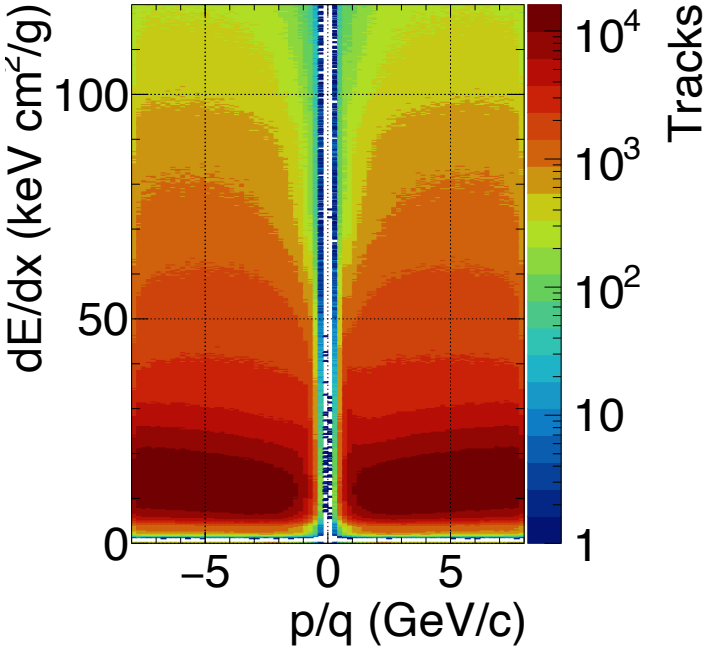
Protons

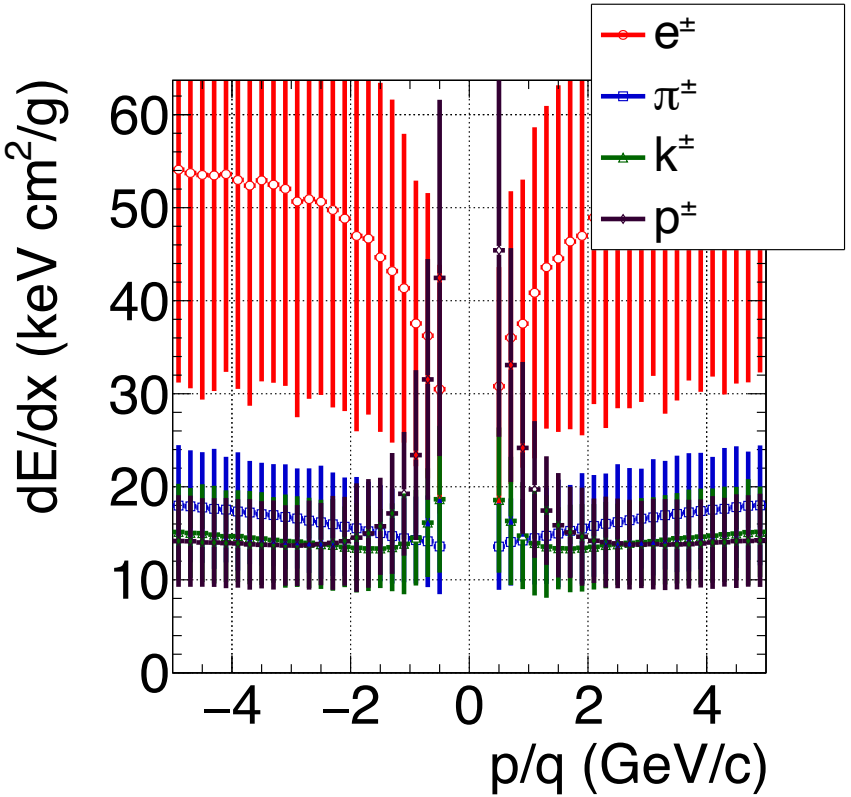
STS response:



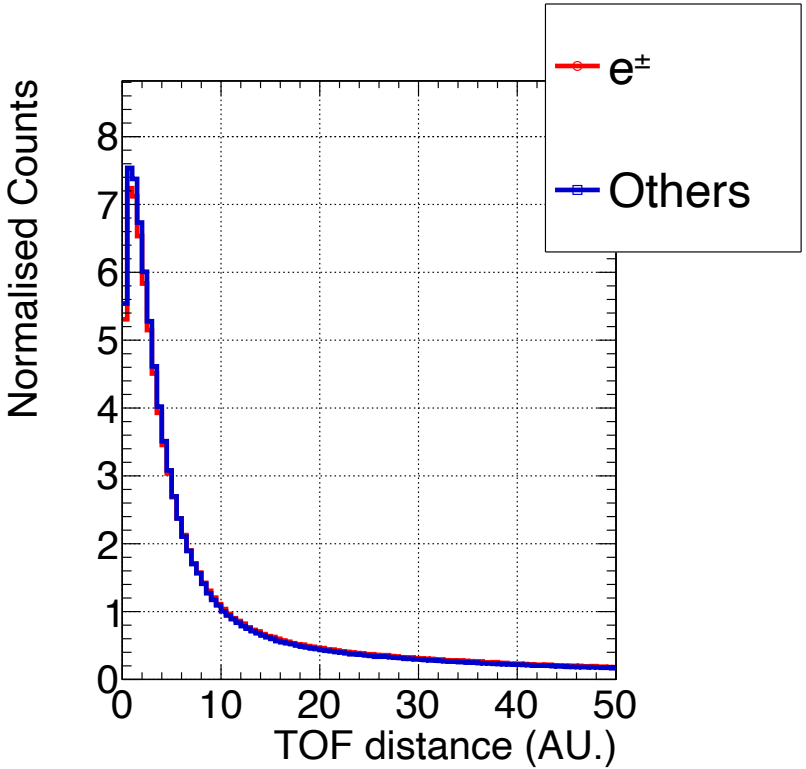


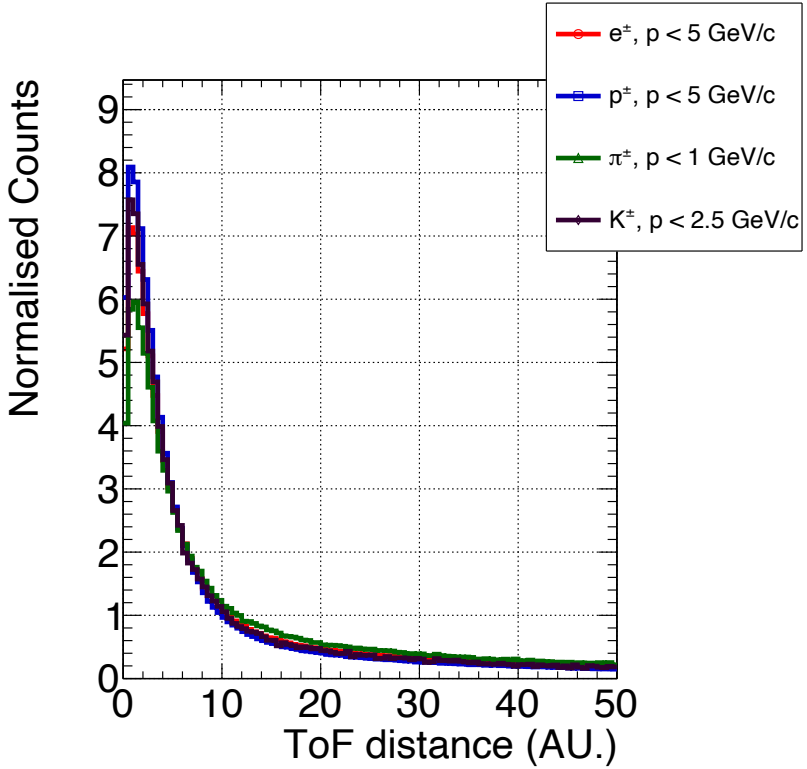
TRD response:



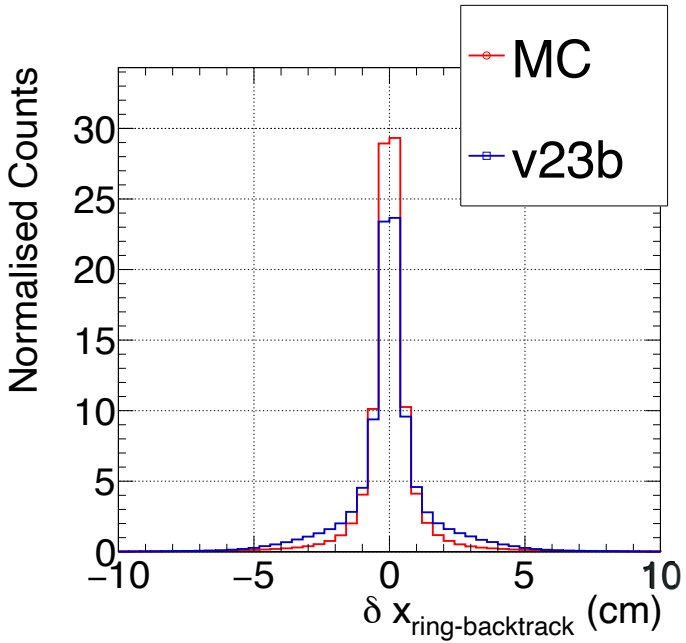


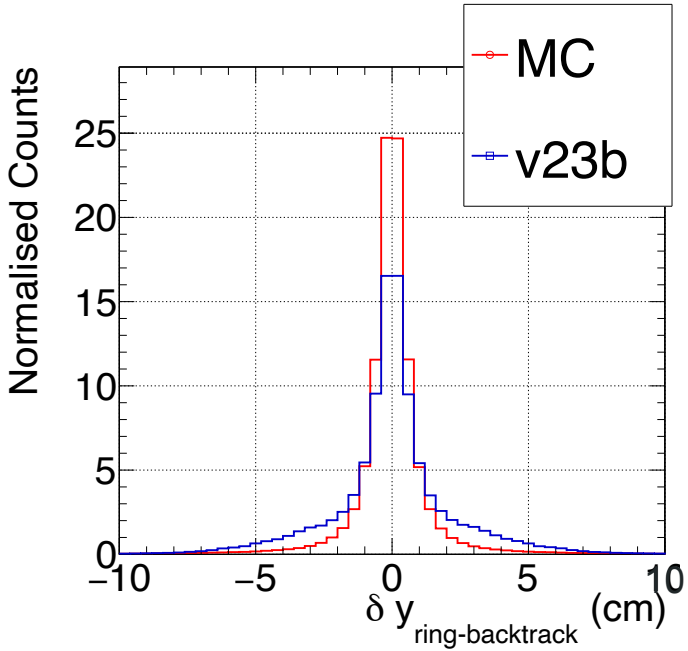
TOF distance:





TRD23 geometry: Ring track matching distances



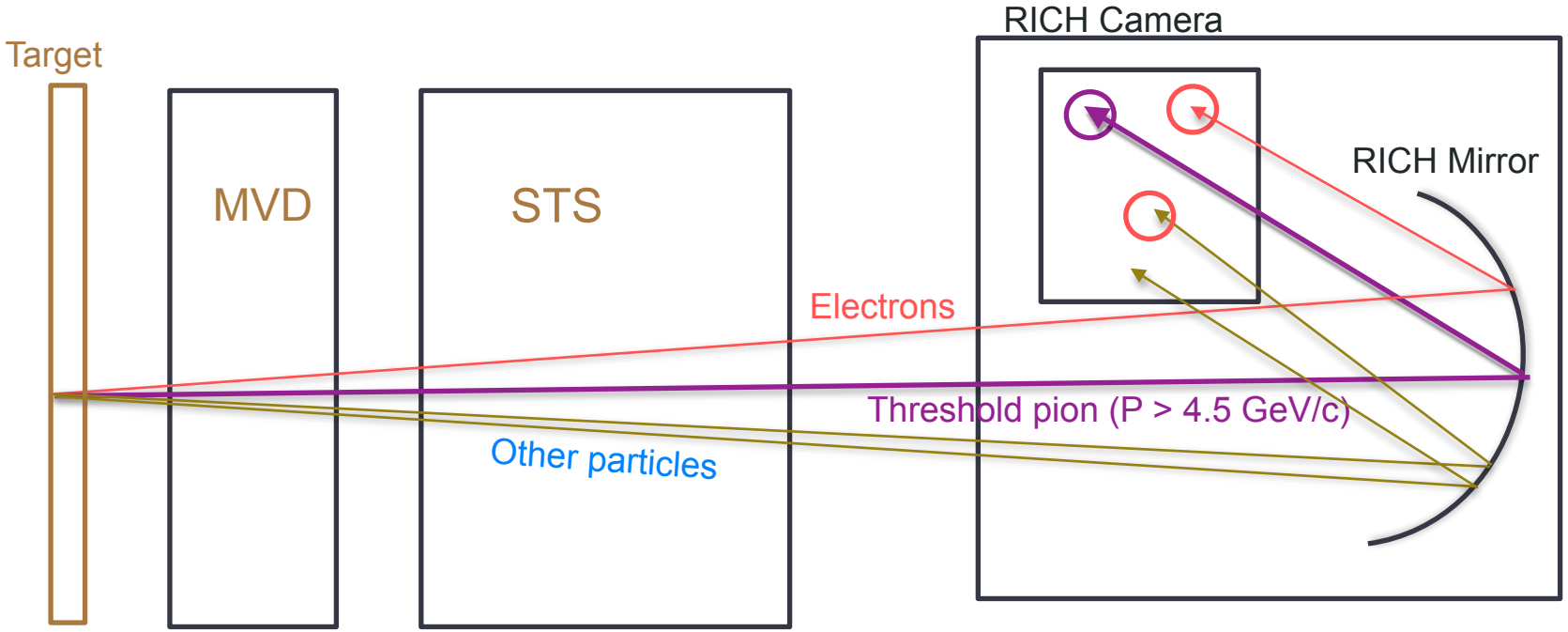


Primary electrons : $\{inmed - \rho, \omega, \phi, qgp\} \rightarrow e^+e^-$

MC: TRD hits smeared with anticipated TRD2D resolution

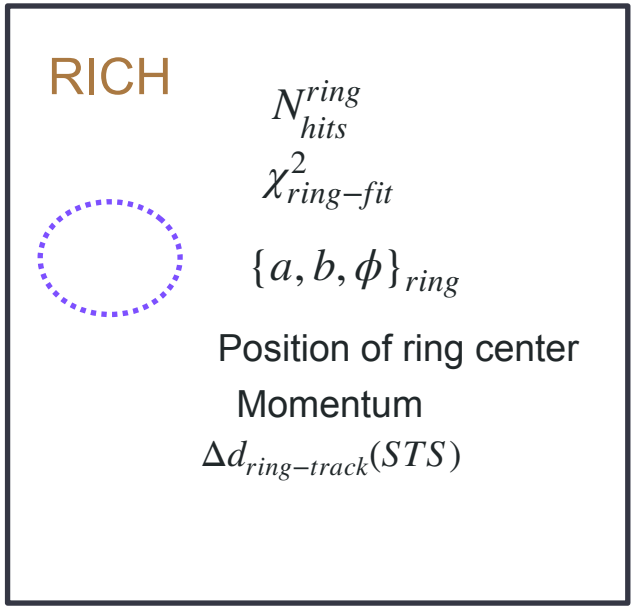
Reduced but acceptable ring-backtrack distances

Rings and tracks

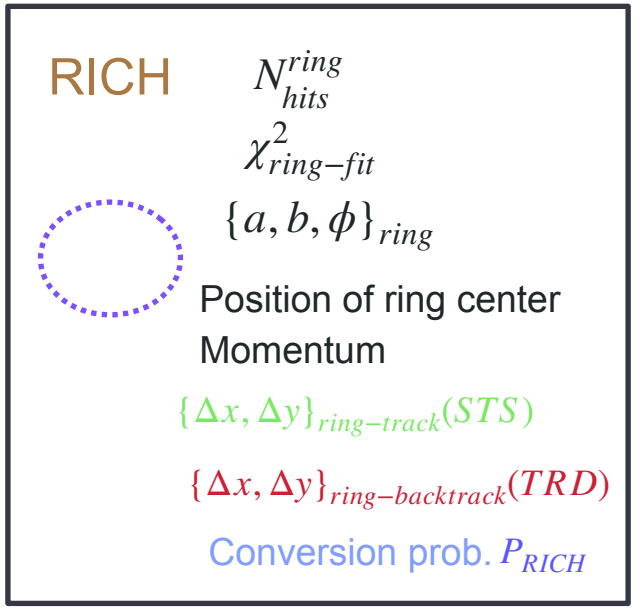


Schematic indicating electrons and pions from target forming rings in RICH, and random mismatching of other particles to electron rings.

Recent upgrades in RICH:

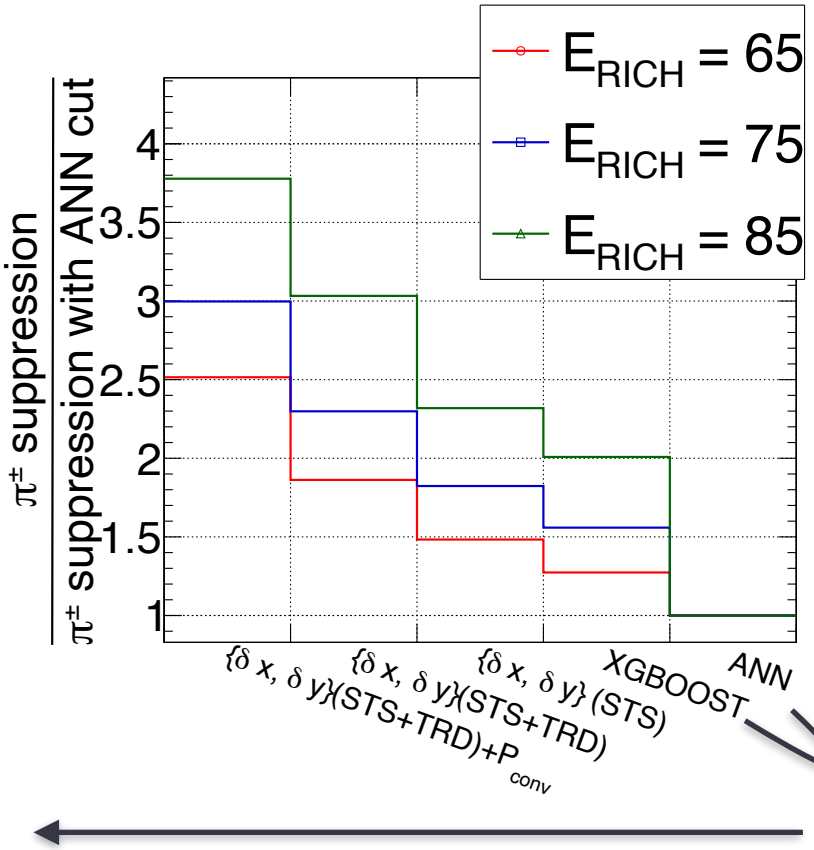


Status-quo input features



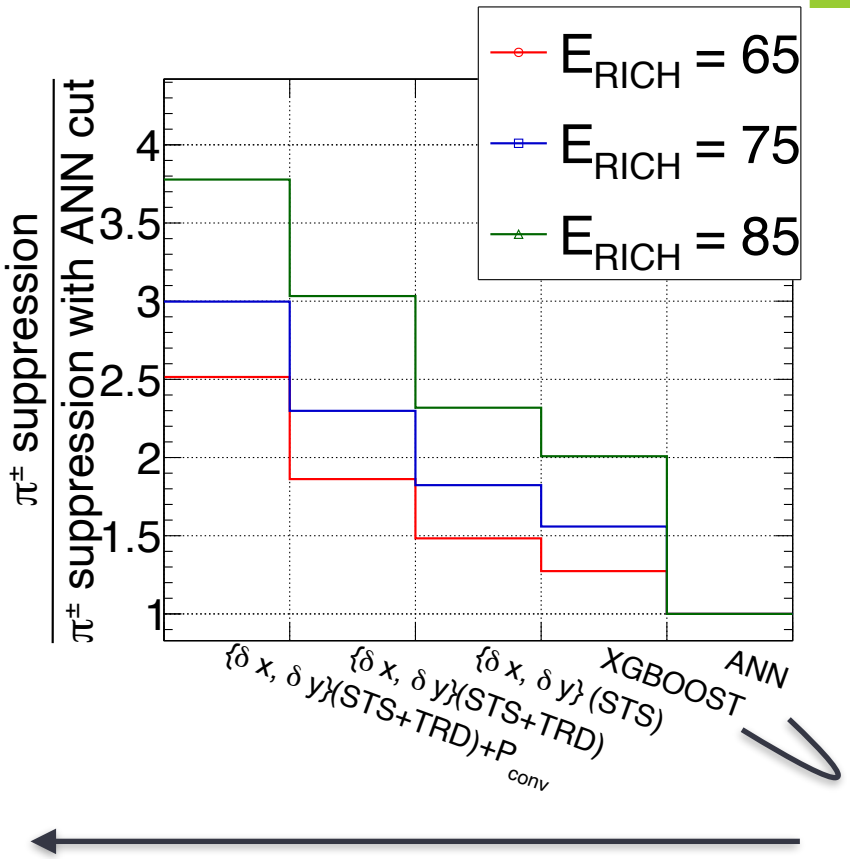
New input features

Recent upgrades in RICH:



- Each modification resulted in an enhancement of the performance of the model.
- TRD backtracked variables and conversion probability for the ring has the largest impact.

Recent upgrades in RICH:



Status-quo input features

- Each modification resulted in an enhancement of the performance of the model.
- TRD backtracked variables and conversion probability for the ring has the largest impact.

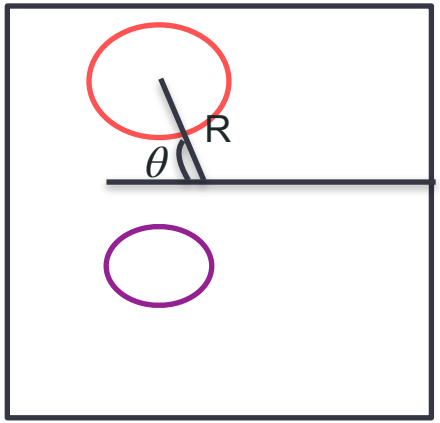
Pinch of salt:

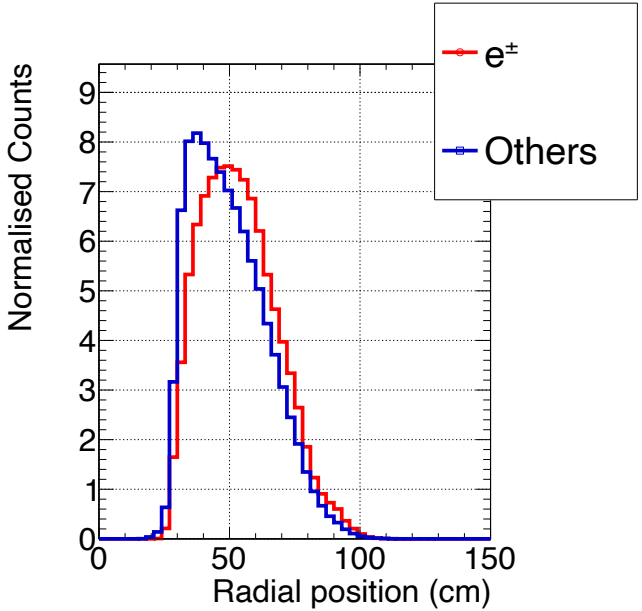
1. Deriving conversion probability involve ideal TRD tracks -> TRD (+TOF) seeded tracks are necessary.
2. TRD hits are re-smeared to TRD2D spatial resolution

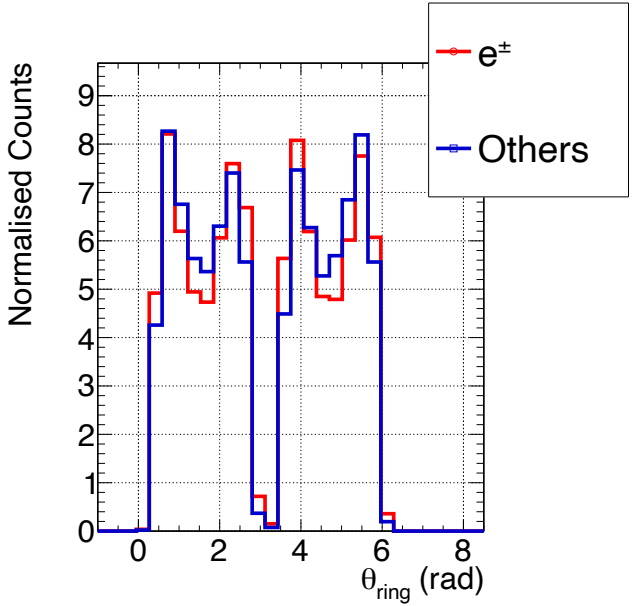
$\sigma_x \sim 100\mu m, \sigma_y \sim 800\mu m$

Ring-Track matching : Parameters

RICH Camera

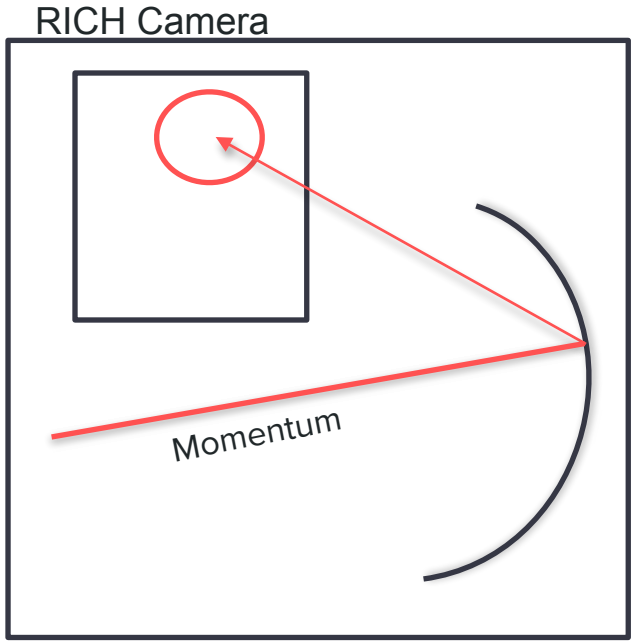




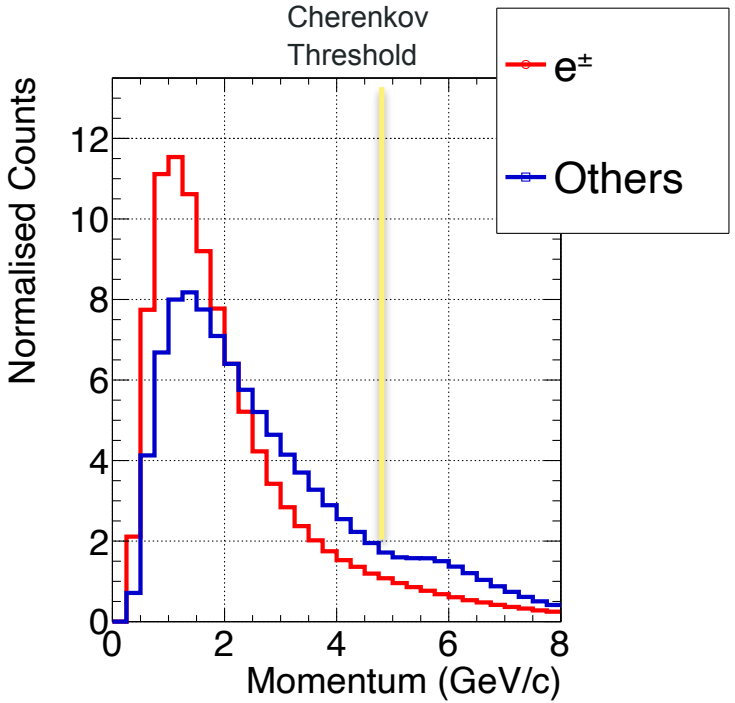


Absolute position of the rings in the RICH plane in polar coordinates

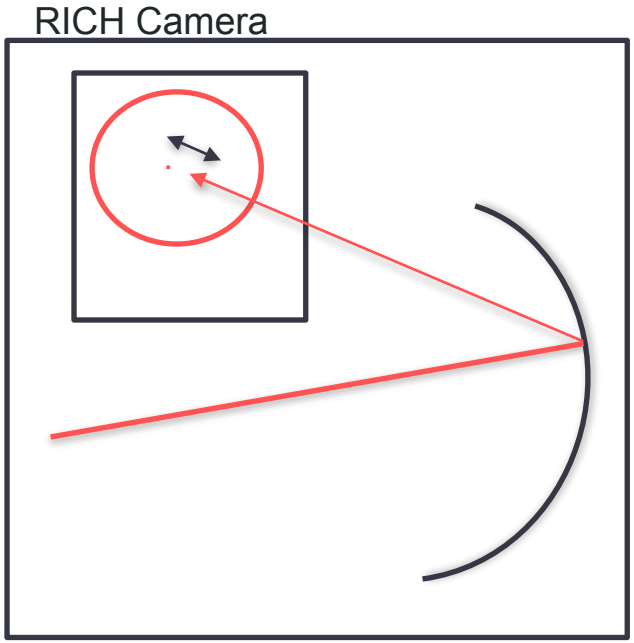
Ring-Track matching : Parameters

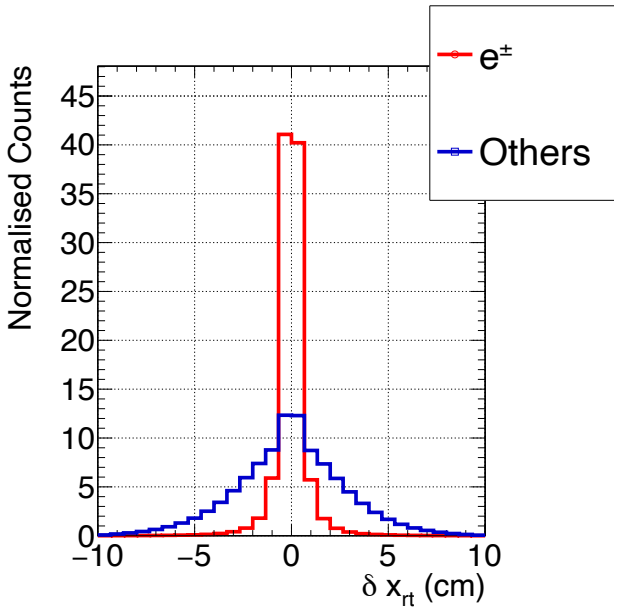


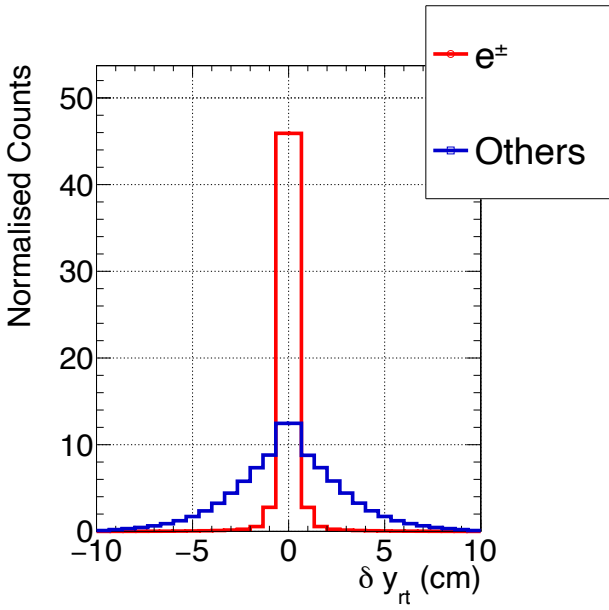
Track parameter : Momentum



Ring-Track matching : Parameters



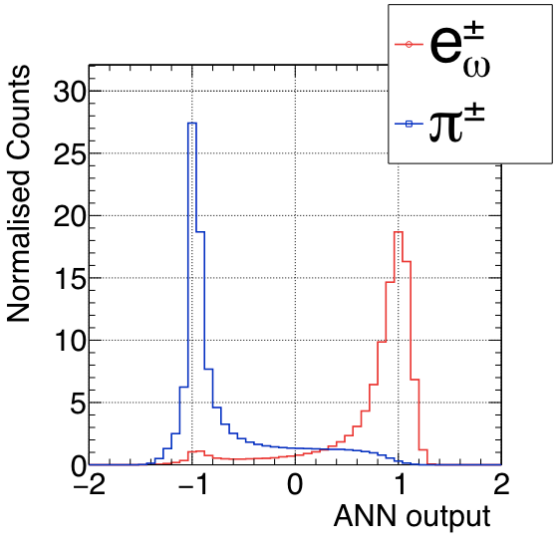




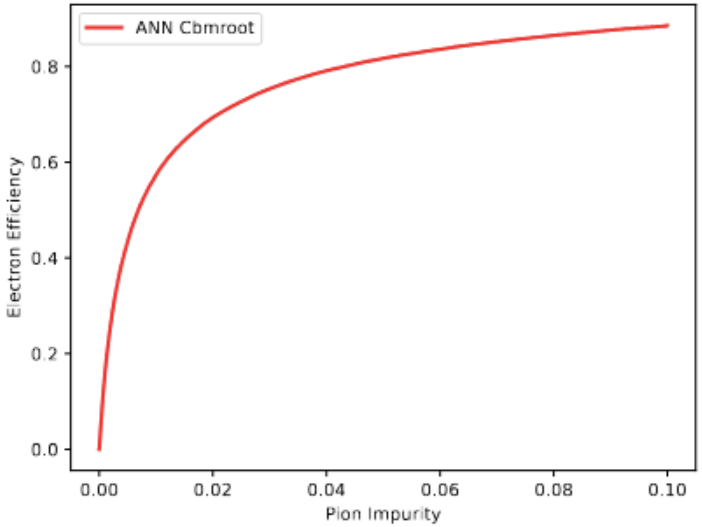
Ring track distances

Clearly, best parameter with most separating power.

ANN in CBMROOT : Response



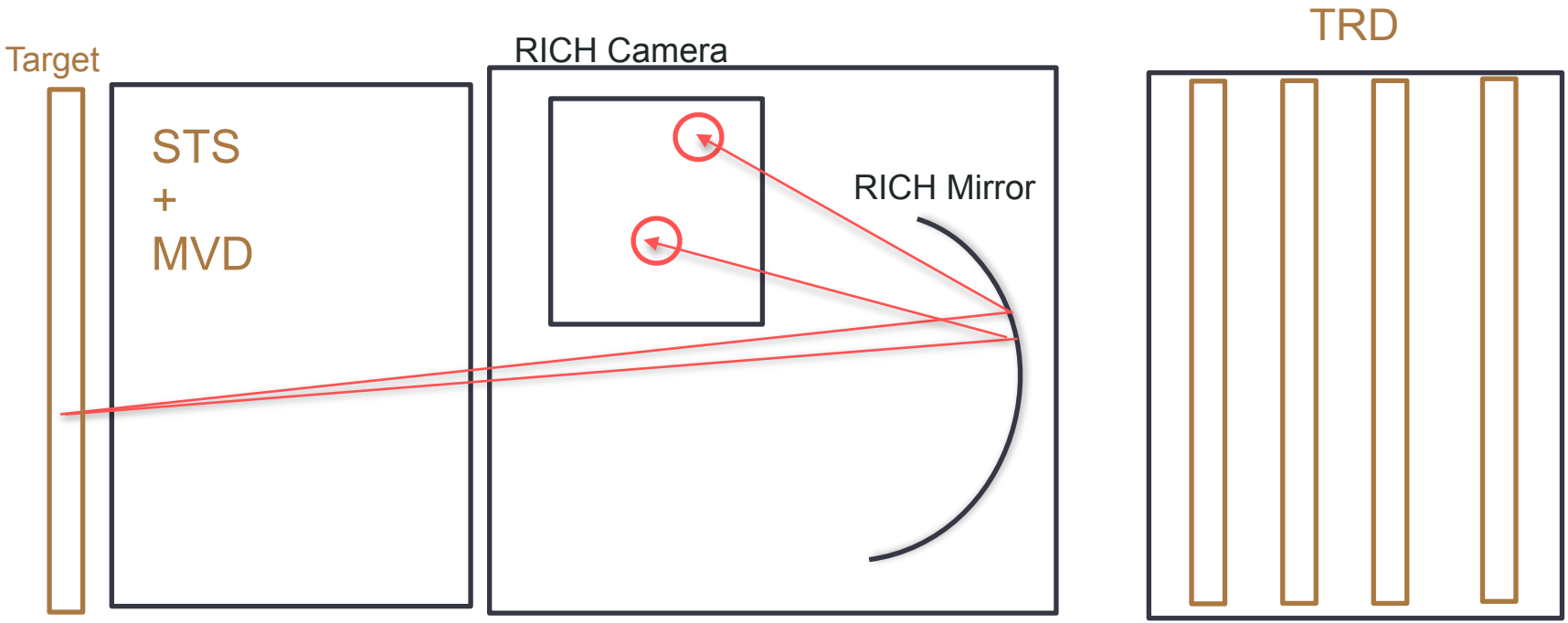
ANN response



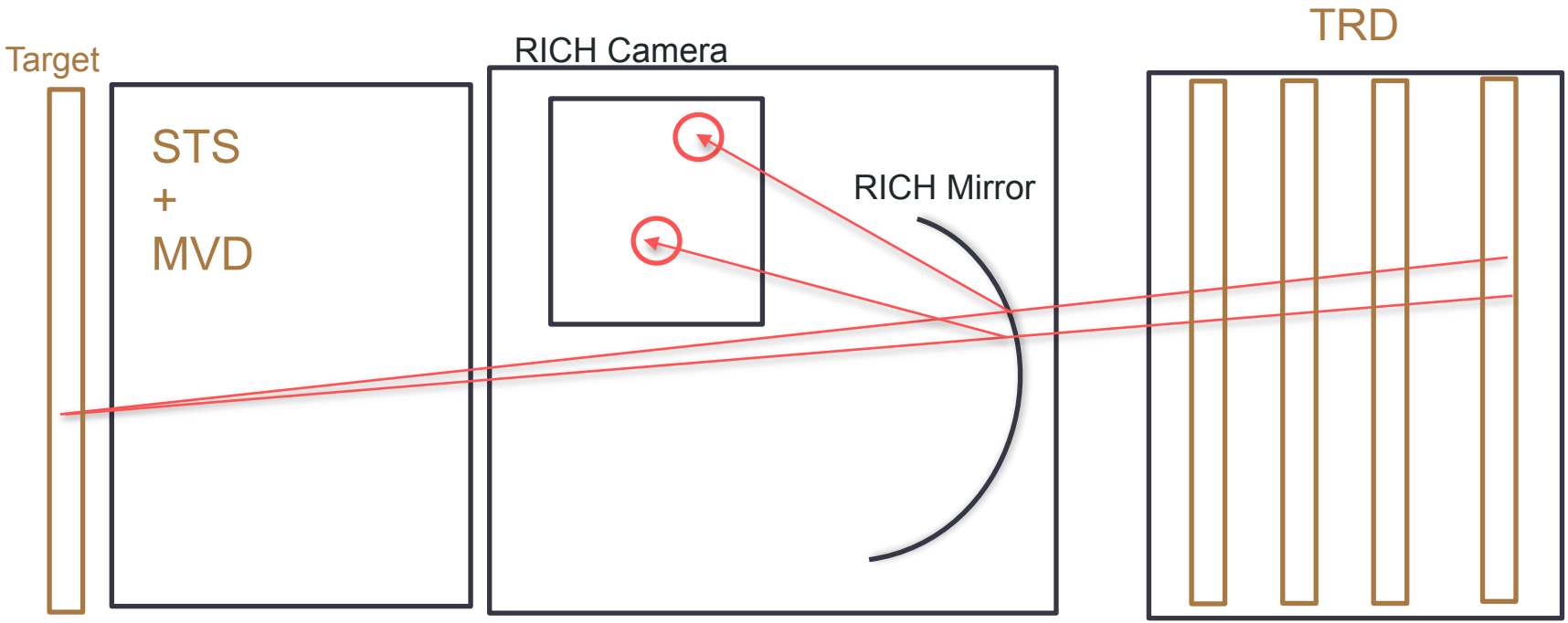
ROC curve for different threshold of the response

- The ANN is CBMROOT is trained on the ring-track matching parameters.
- **Outdated - last optimised on 2017.**
- XGBOOST: extreme gradient boosting (decision trees) -> New and advanced approach.
- Fast, uses sparse matrices, ensemble of weak learned decision trees.

Ring-Track matching

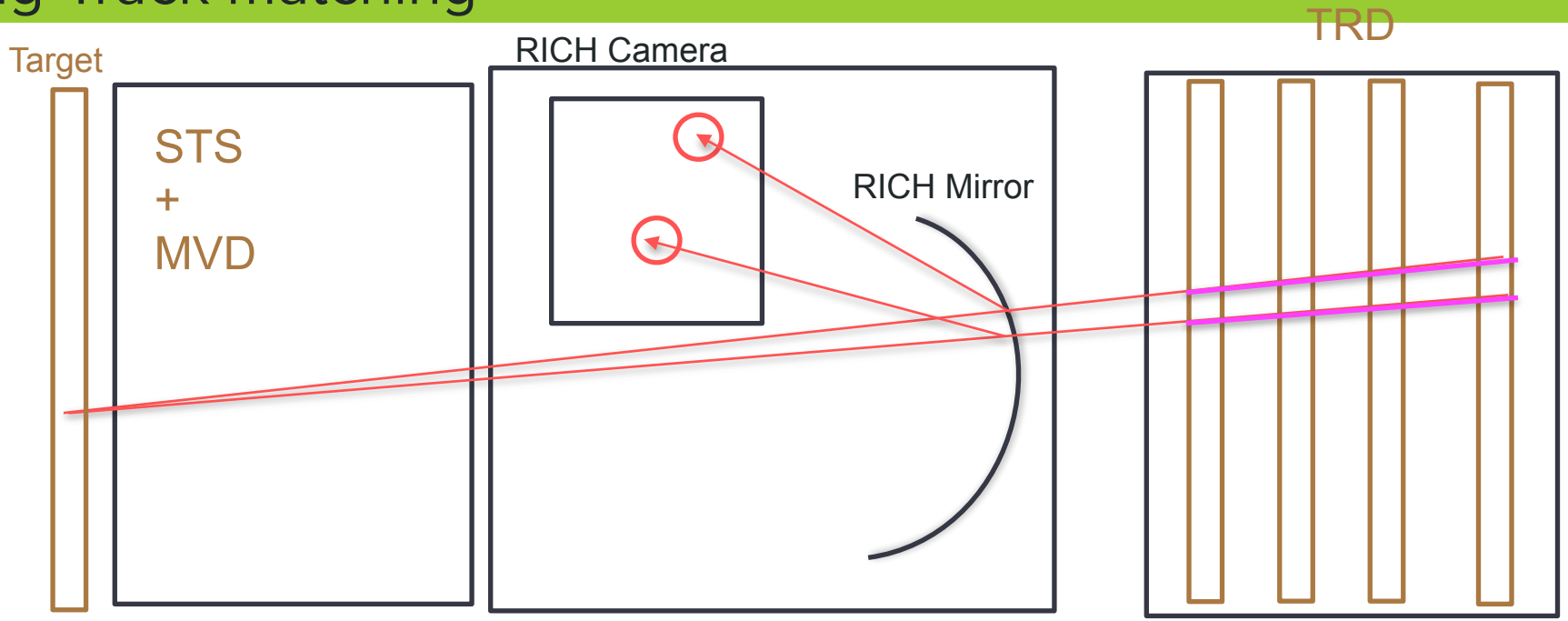


Ring-Track matching



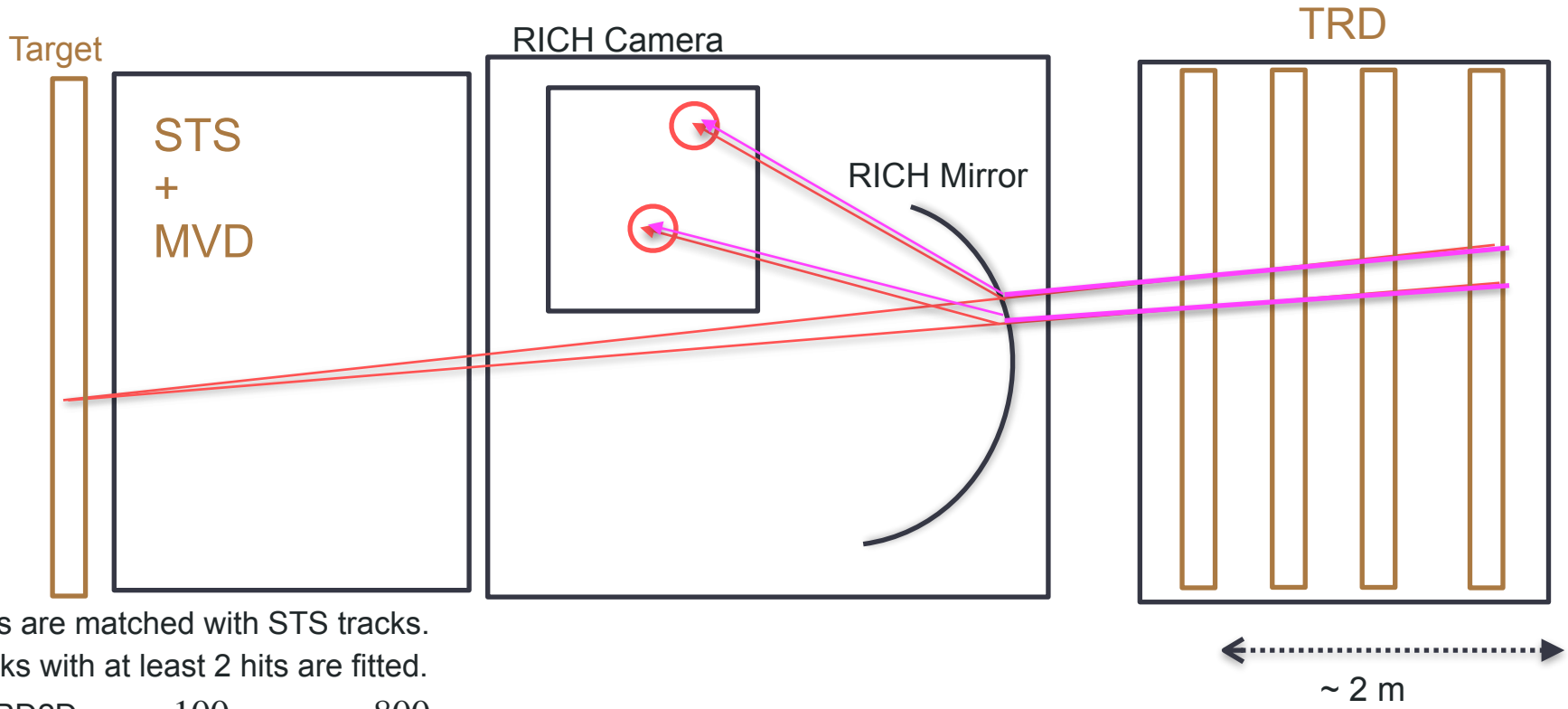
- TRD Hits are matched with STS tracks.

Ring-Track matching



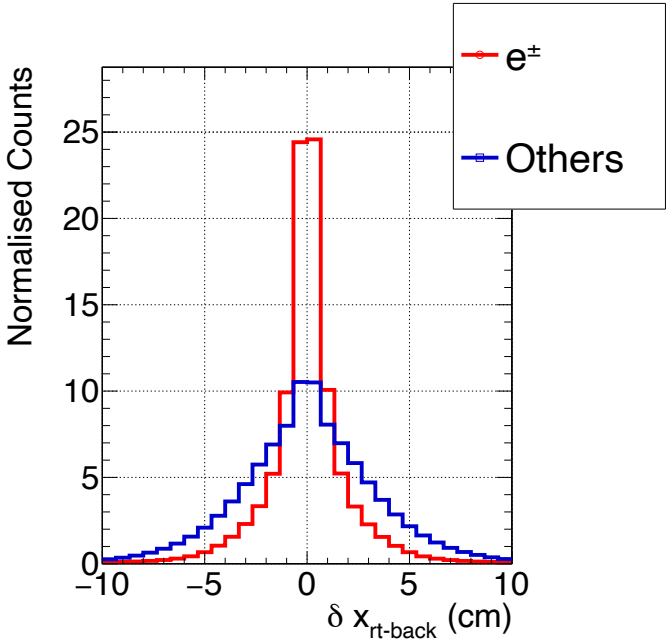
- TRD Hits are matched with STS tracks.
- The tracks with at least 2 hits are fitted.
- Here hits are **re-smeared to TRD2D spatial resolution** $\sigma_x \sim 100\mu m, \sigma_y \sim 800\mu m$

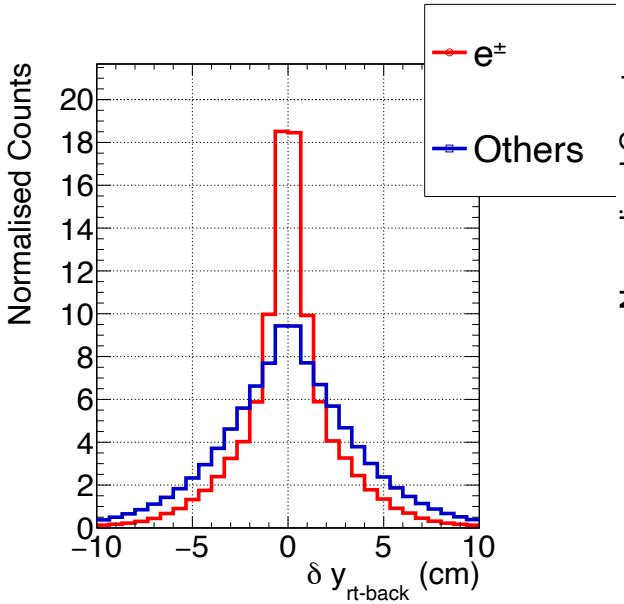
Ring-Track matching

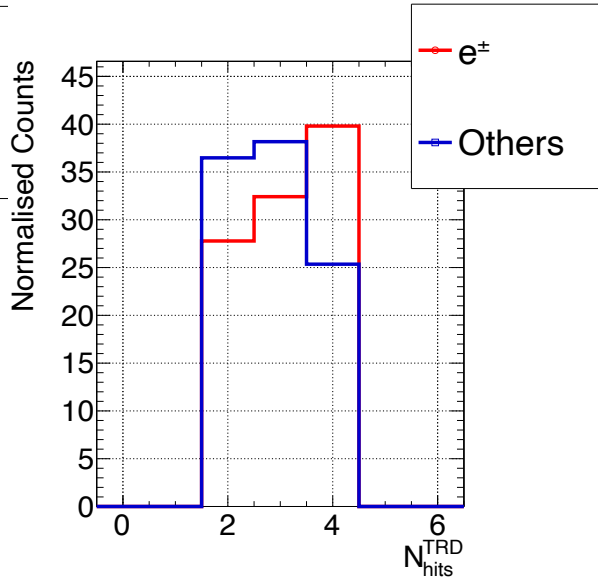


- TRD Hits are matched with STS tracks.
- The tracks with at least 2 hits are fitted.
- Using TRD2D, $\sigma_x \sim 100\mu m, \sigma_y \sim 800\mu m$
- Fitted tracks are backtracked to RICH and matched to the same ring-> **Additional track pointing reference**

Additional parameters:

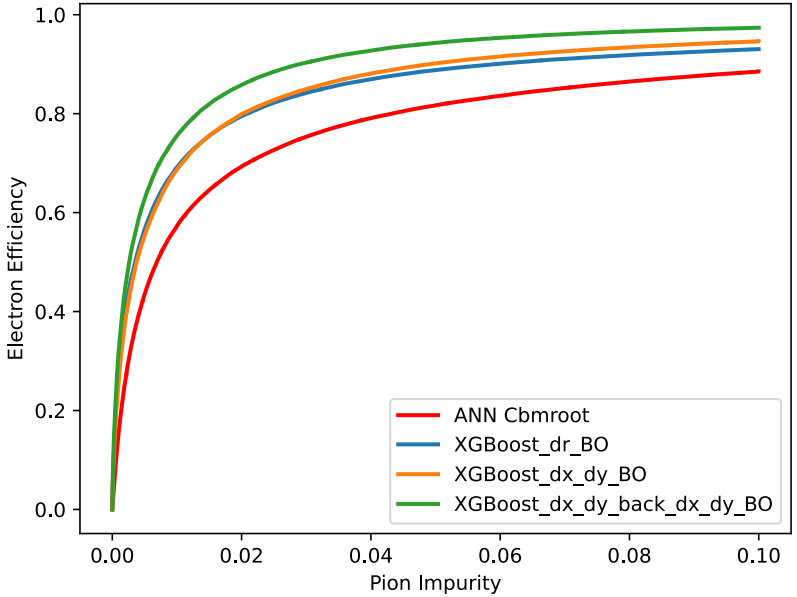




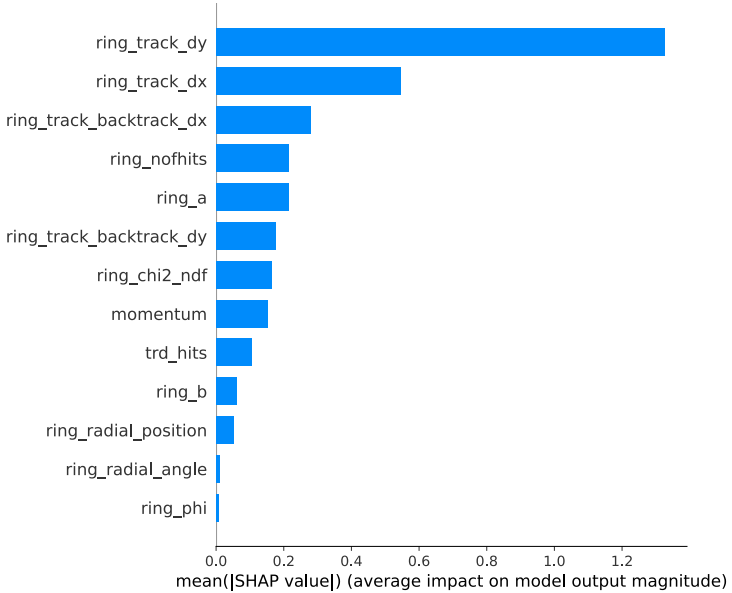


- Number of TRD hits and backtracked distances are added as additional parameters to the ML model.

XGBOOST with TRD parameters : Response for Omega embedded sample



ROC curve for different threshold of the response

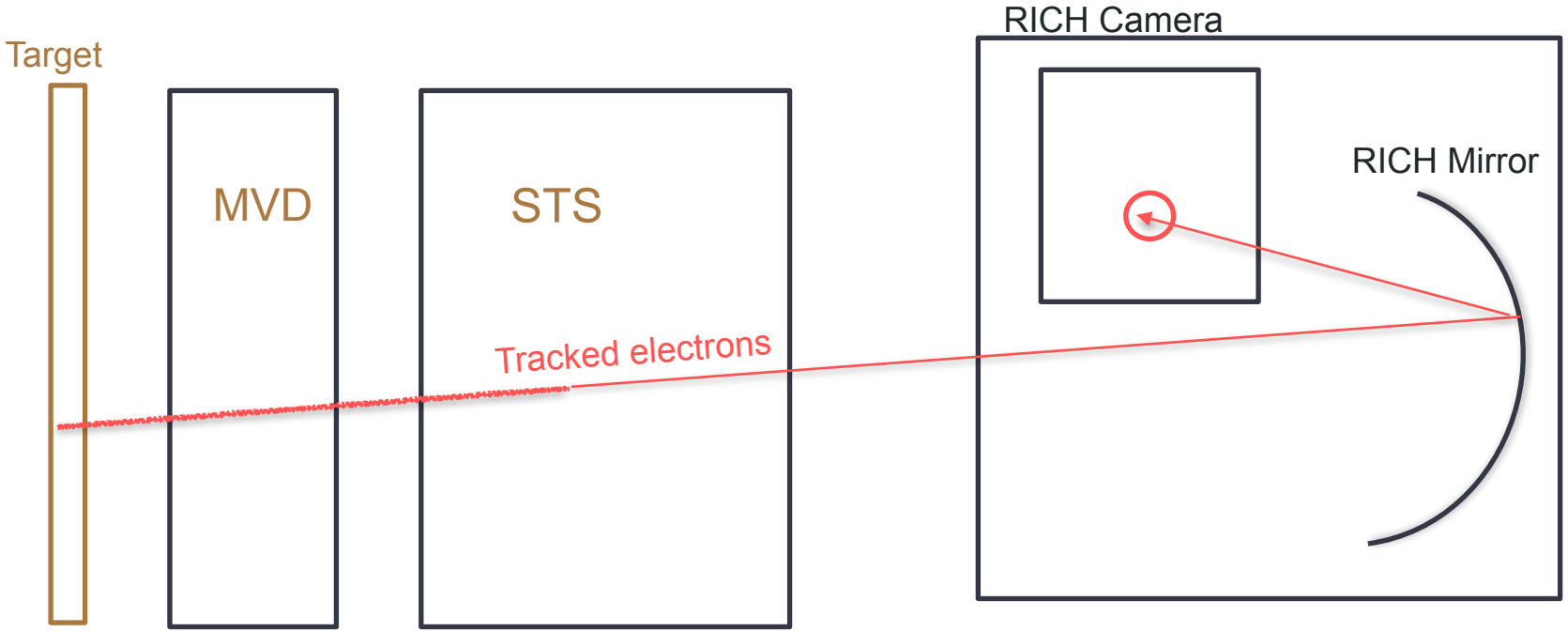


Variable importance (SHAP values)

- Inclusion of the TRD variables improved the model performance.
- Note : TRD has 80 % acceptance for STS tracks with RICH projection.
- Hence two ML models, with and without TRD parameters are used for analysis.

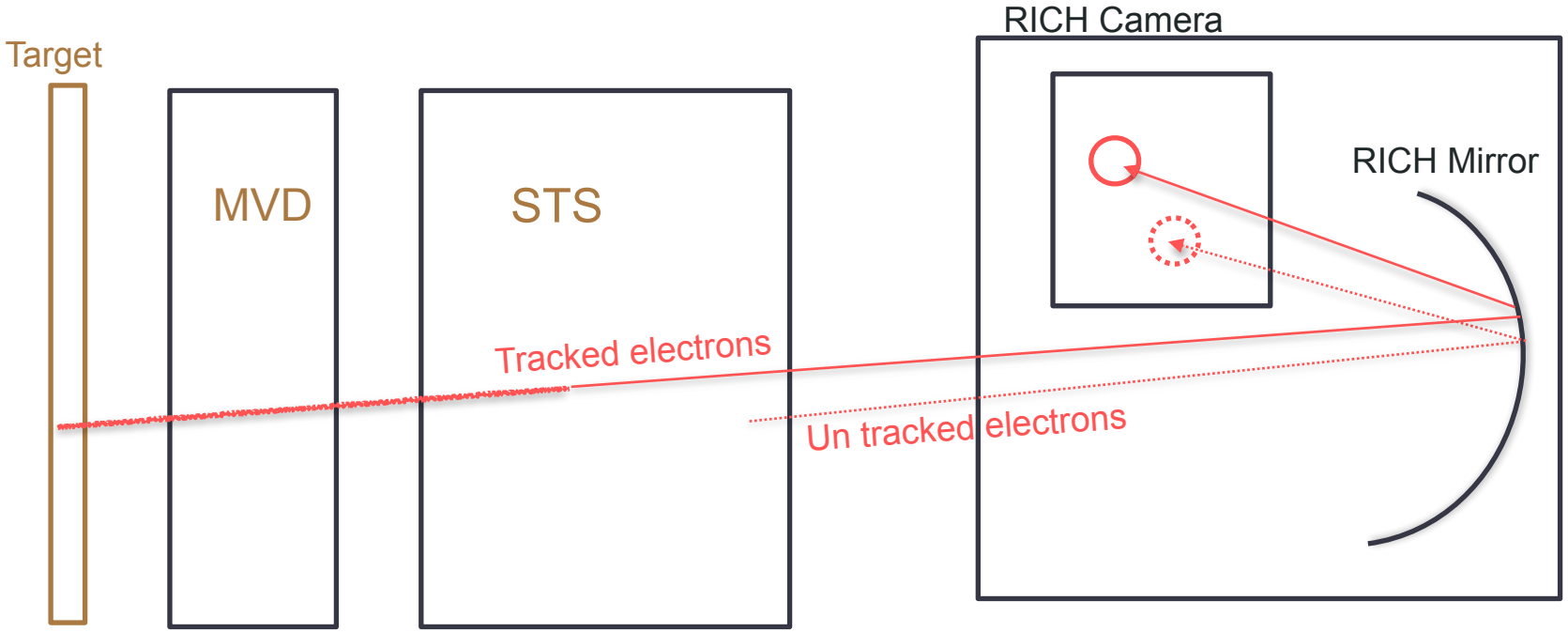
Conversion electrons :

Mis-matched pions:



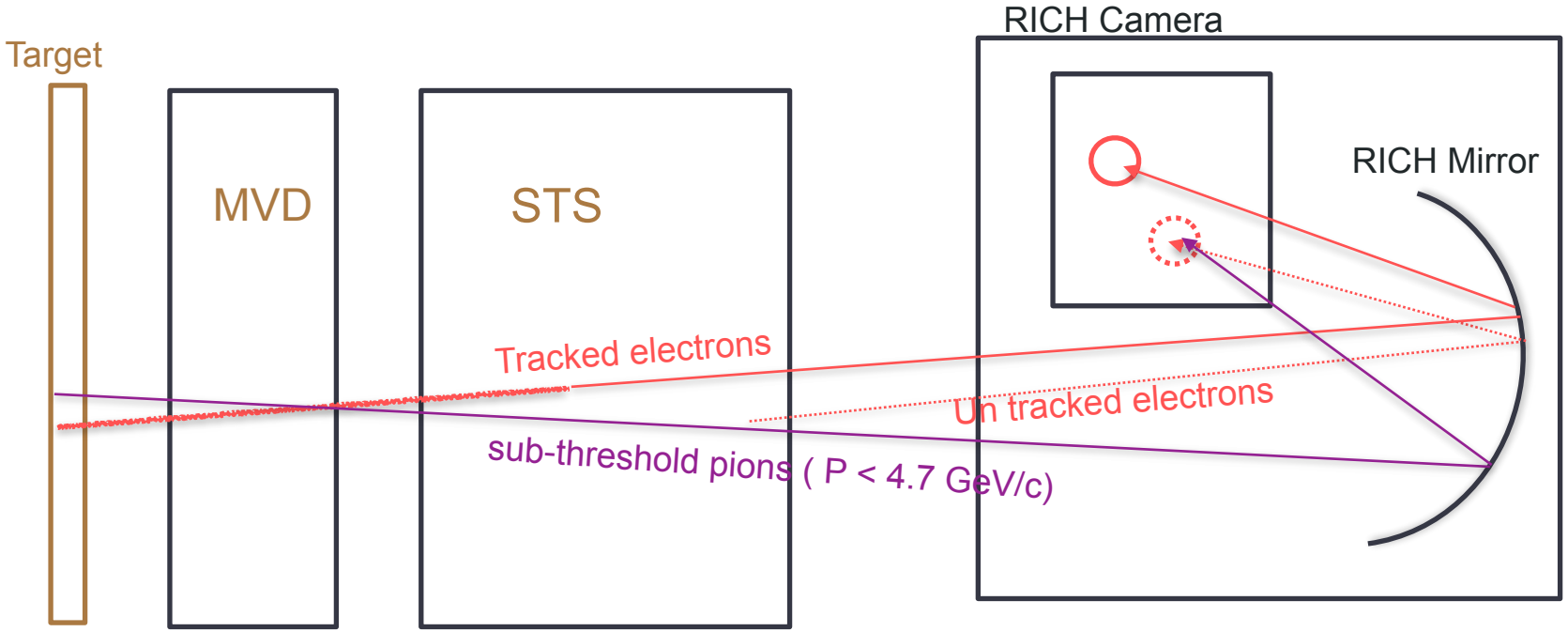
Schematic indicating electrons tracked in STS forming rings in RICH

Mis-matched pions:



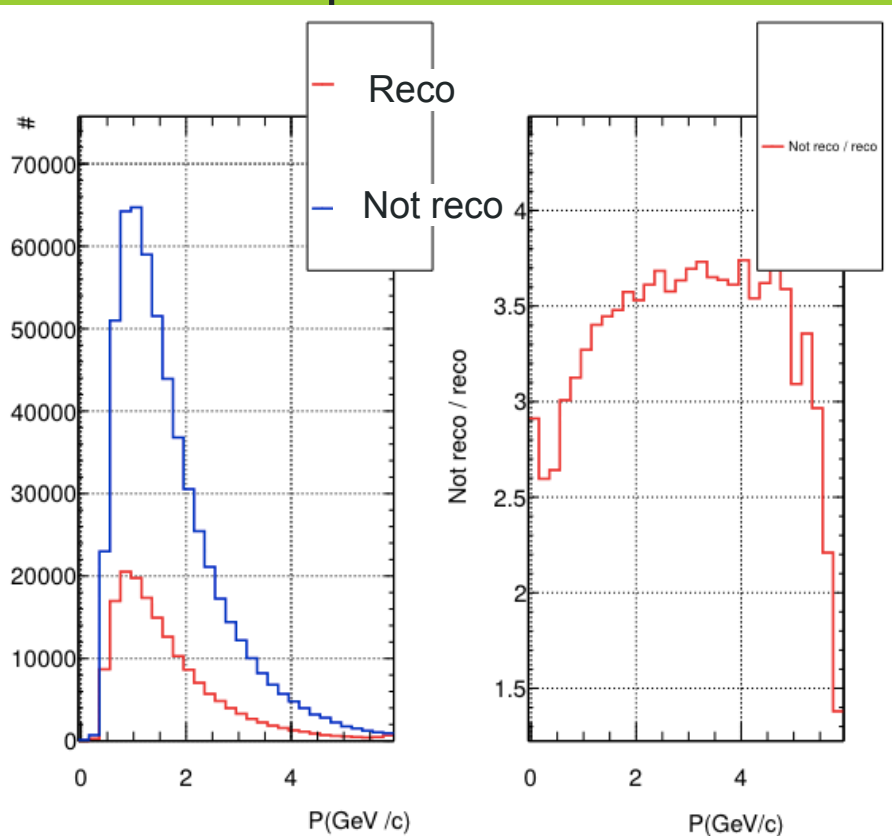
Schematic indicating electrons tracked in STS and untracked electrons forming rings in RICH

Mis-matched pions:



Schematic indicating electrons from target forming rings in RICH

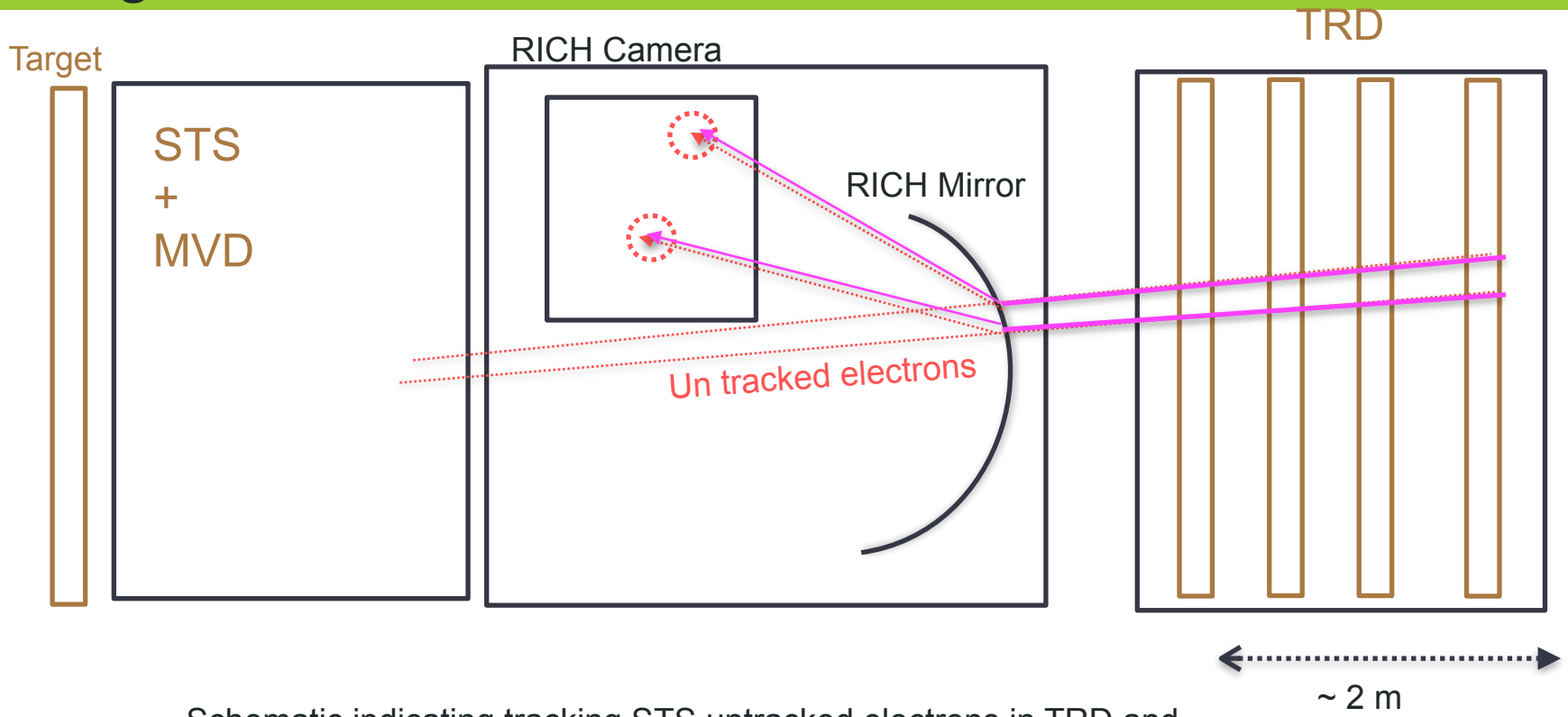
Mis-matched pions:



Pions mismatched to tracked and un tracked electron rings.

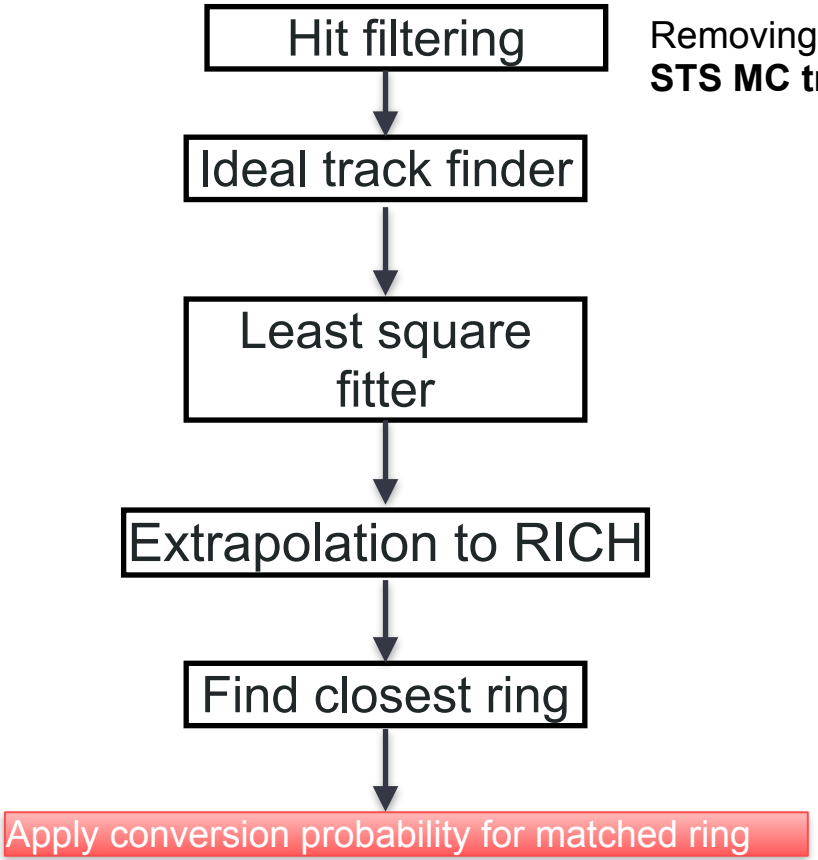
- More than 70% of mis-matched pions are to untracked electron rings

Tracking STS untracked electrons in TRD :



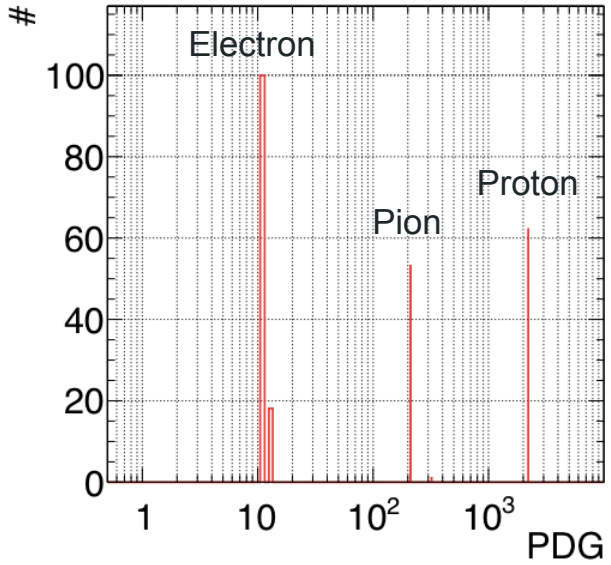
Schematic indicating tracking STS untracked electrons in TRD and extrapolating back to RICH and matching to the nearest ring.

Secondary (TRD) tracking flowchart :



Removing TRD hits associated with **STS MC track**.

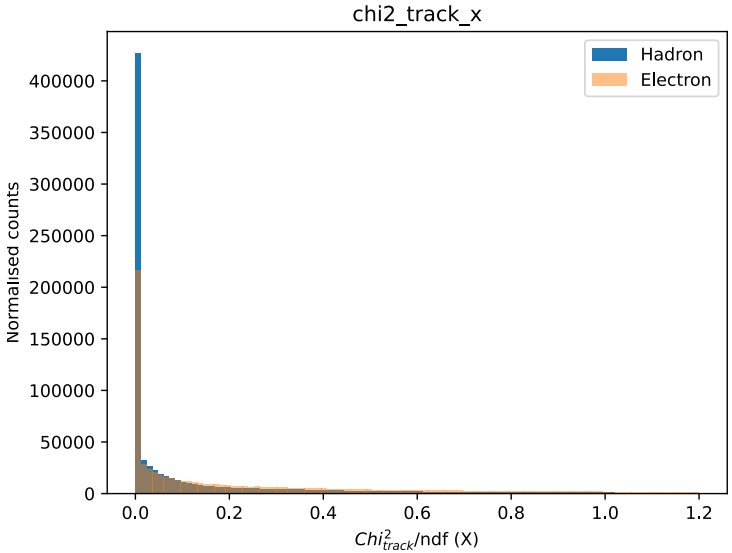
TRD tracks:

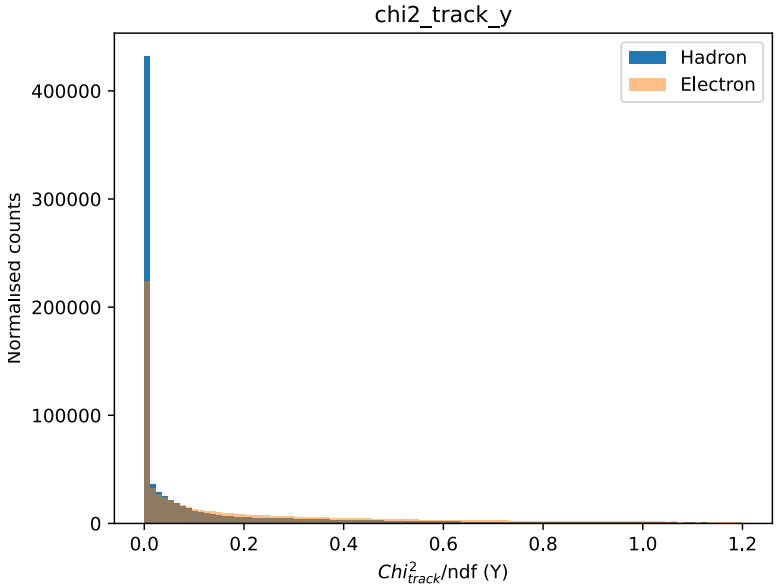


- “Signal” - Electrons not tracked in the STS, forms the rings in RICH.
- “Background” - Other particles

- Along with the signal electrons, large background of hadrons (partially from target) are tracked in TRD after the hit filter.
- This scenario leads to BG tracks being matched to omega electrons - which needs to be avoided.
- A ML model is trained with TRD fitted track parameters is used to reduce the hadron BG in TRD backtracks.

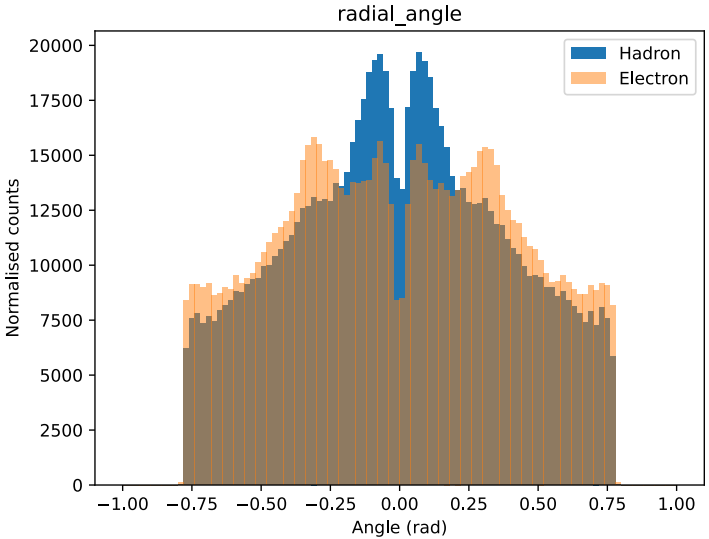
Conversion probability : Parameters

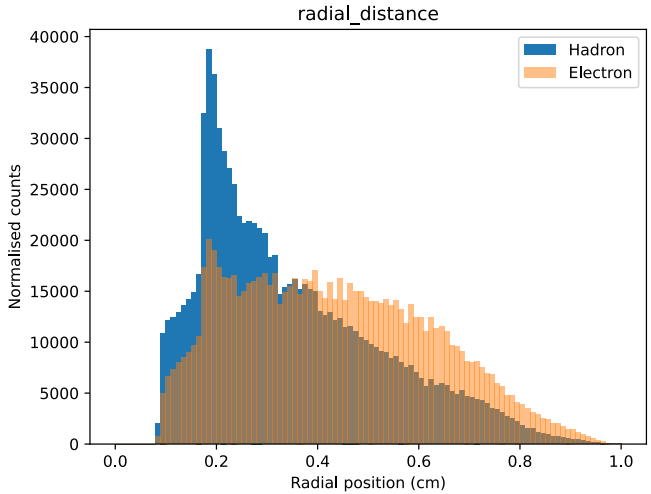




TRD track fit parameters

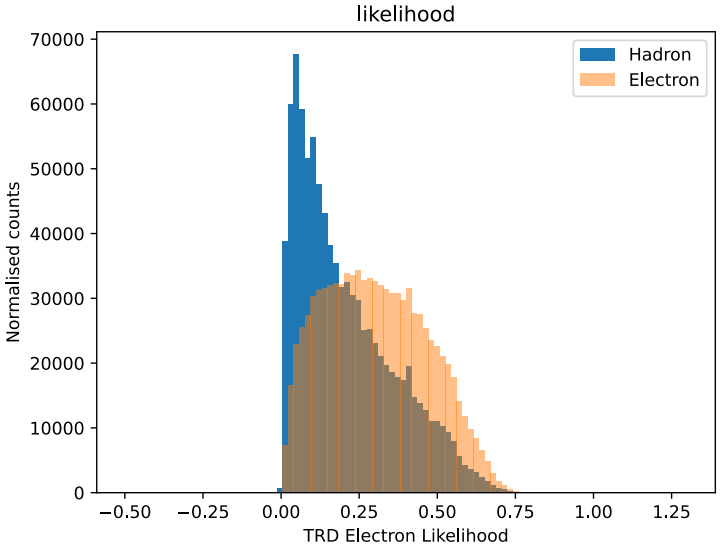
Conversion probability : Parameters



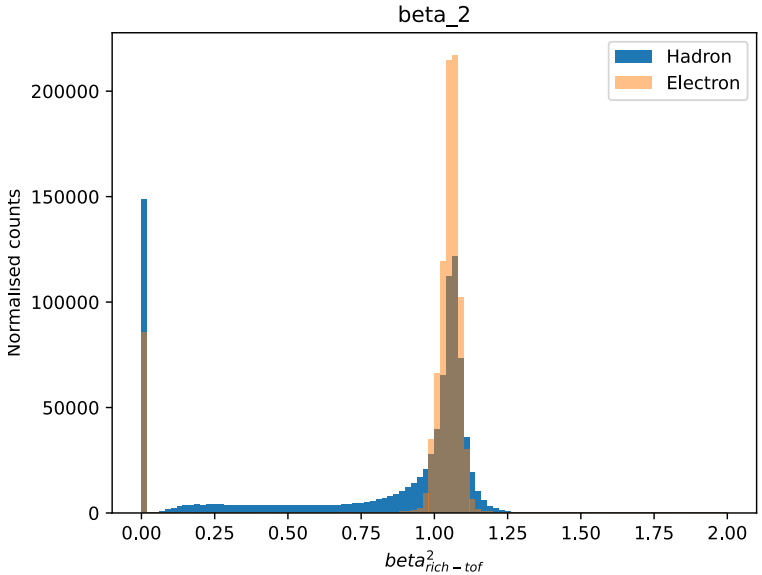


Absolute position of the first hit of the TRD track

Conversion probability : Parameters

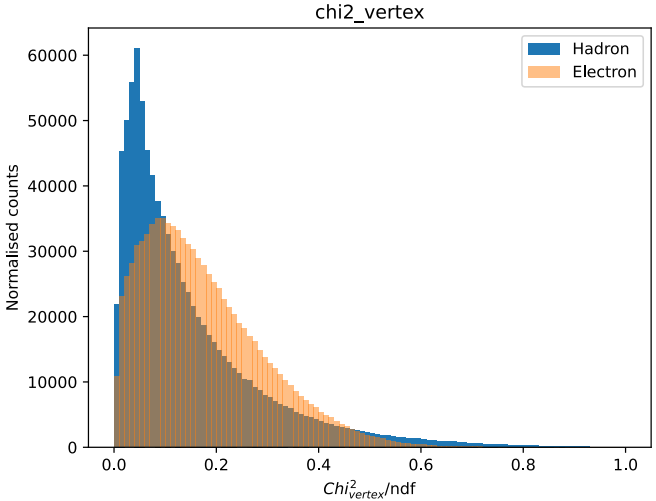


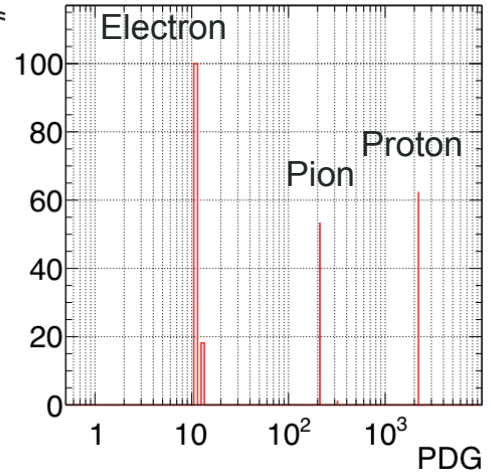
TRD electron likelihood for the fitted tracks

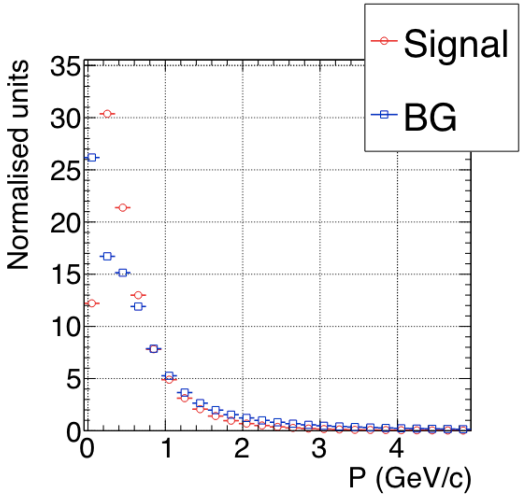


- Extrapolated the fitted track towards TOF.
- Using ring time as the start time and TOF time as stop time, velocity is derived.

Conversion probability : Parameters

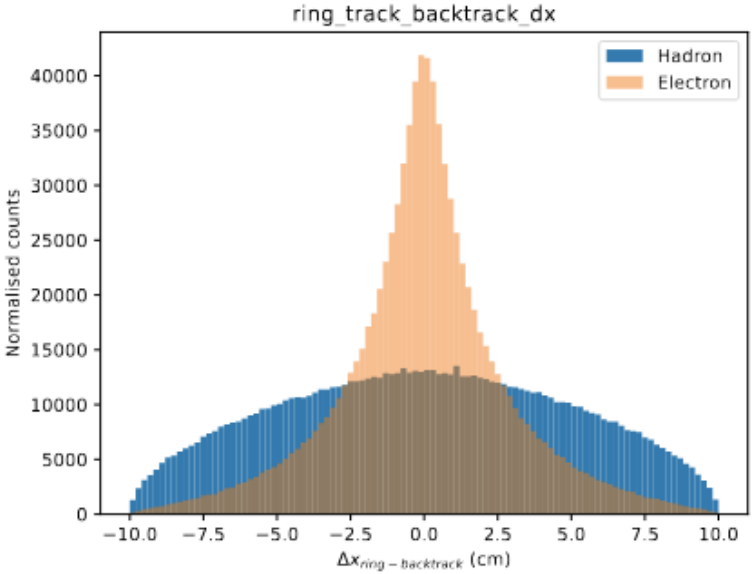


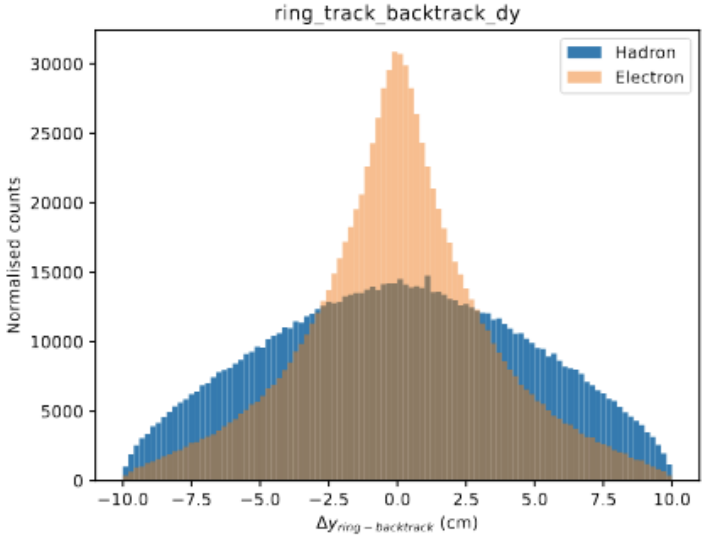




- Some of the hadron tracks are originating from vertex.
- TRD fitted track is extrapolated to the origin assuming proton hypothesis and 0.6 GeV/c as momentum.

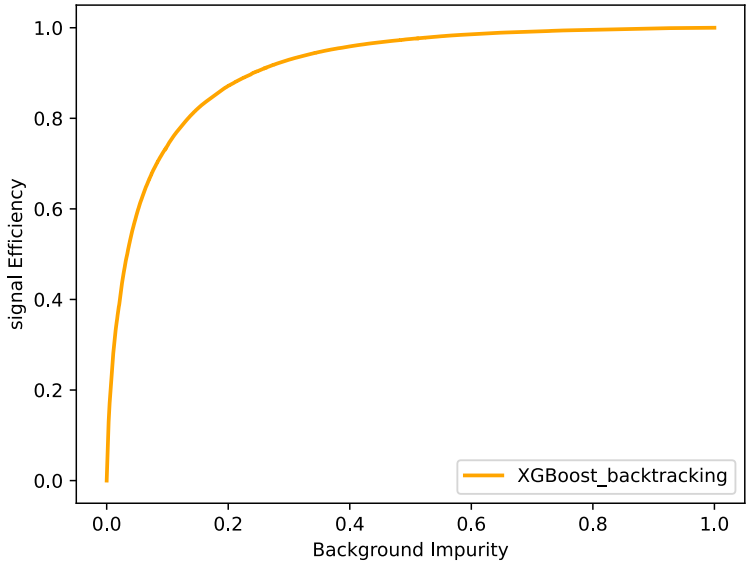
Conversion probability : Parameters



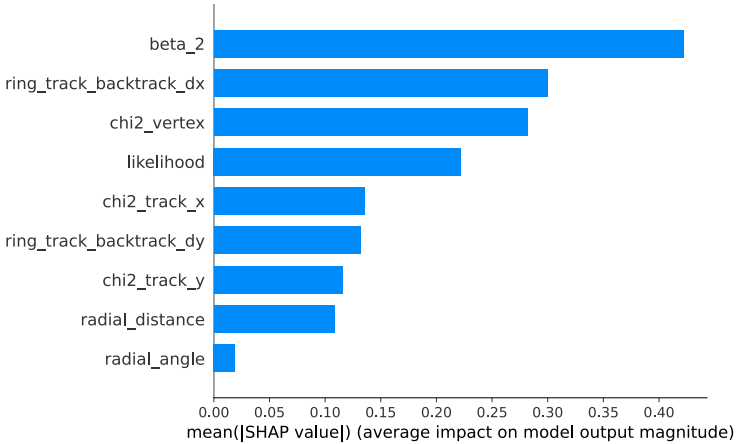


Backtrack - ring matching distance

XGBOOST for deriving conversion probability :

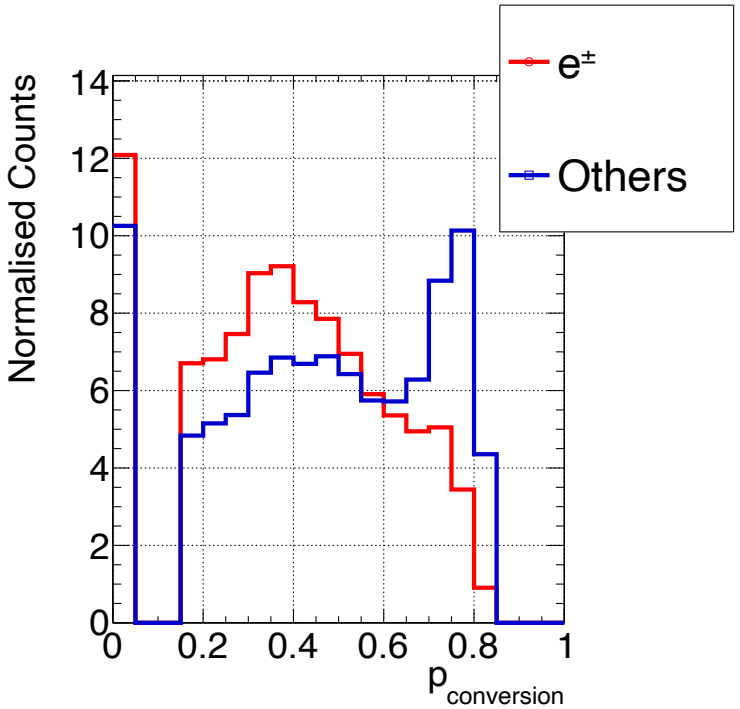


ROC curve for different threshold of the response



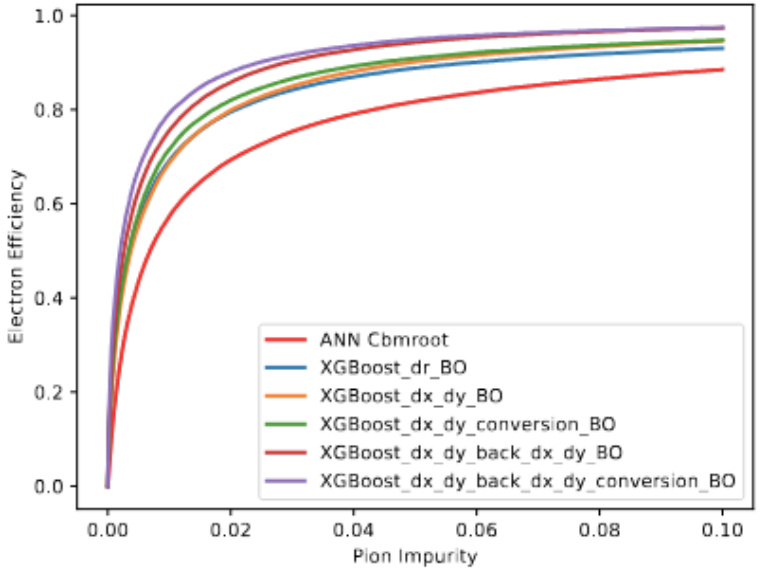
Variable importance (SHAP values)

Inclusion of the conversion probability :

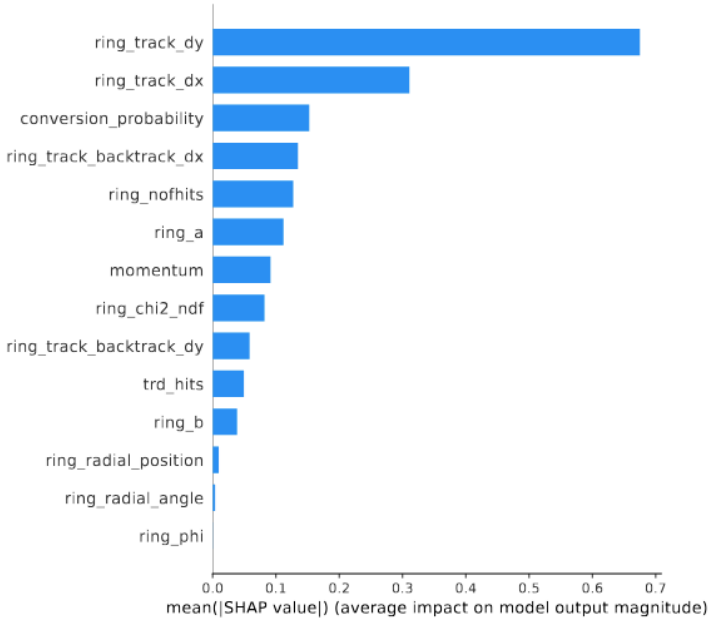


- Each ring is tagged with the probability that it originates from the photon conversion.
- This additional parameter is used to train the forward matching ML model.

XGBOOST with conversion probability : Response for Omega embedded sample



ROC curve for different threshold of the response



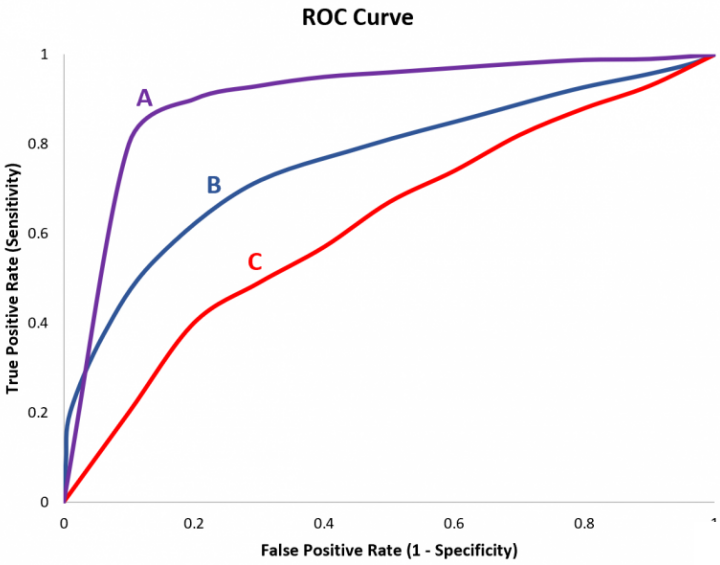
Variable importance (SHAP values)

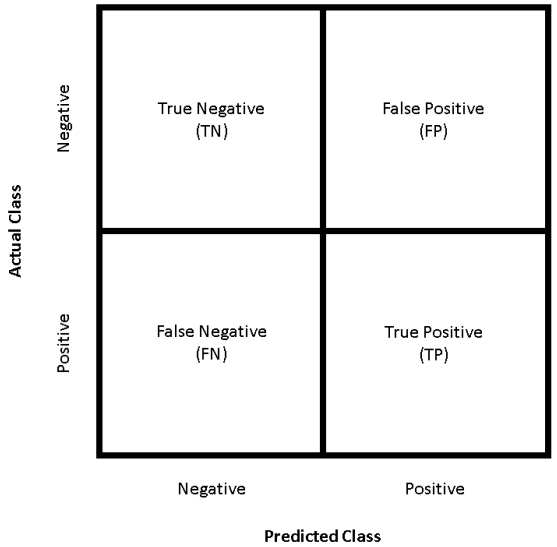
- Inclusion of the conversion probability enhanced the model performance.

Summary

- XGBOOST is used as replacement for ANN for RICH electron ID - Improved identification performance was achieved for the omega electrons.
- Differential ring track distance is added to the model - enhancement of efficiency at higher pion impurity is observed.
- The STS tracks are re-fitted in TRD and extrapolated back to RICH for additional ring-track distance reference.
- Complimenting the forward track, the TRD backtrack information increased the model performance.
- The STS untracked electrons are tracked in TRD and projected back to RICH.
- Using the TRD backtrack and ring-backtrack distance and TOF time, a conversion probability factor is assigned for each ring.
- The use of conversion probability as additional parameter in the RICH matching network improved the model performance even at low pion impurity.

ROC





$$\text{True Positive Rate (TPR)} = \frac{TP}{P} = \frac{TP}{TP + FN}$$

also called sensitivity/recall/hit rate

$$\text{False Positive Rate (FPR)} = \frac{FP}{N} = \frac{FP}{FP + TN}$$

also called fall out

Outlook

- Global electron identification network including all the parameters for RICH , TRD, and TOF is foreseen.
- Priority matching schemes : “ A good electron track downstream is a good electron in RICH”, hence should be prioritised in ring matching and corresponding ring should be removed or conversely a “downstream electron probability” can be set for each ring.
- Complete re-do of the backtracking procedure with TRD23 geometry with TRD2D pads.
- Backtracking procedure can be extended till STS : Missing track of conversion electron pair from target could be searched in STS -> reduce combinatorial background.