Machine learning application for electron identification in CBM

Pavish Subramani
University of Wuppertal, Germany

FAIRNESS, Sep.2024

CBM

BERGISCHE
UNIVERSITÄT
WUPPERTAL

Di-electron probe :

- Di-electrons are produced in all stages of heavy-ion collisions $\iff$ Reflect history of collision.
- They are penetrating probe to understand the QCD medium produced in the initial stages of heavy-ion collisions $\iff$ No strong interaction.
- The low invariant mass regime of di-electrons spectrum is saturated by the low mass vector mesons.
- $\rho \to e^+ + e^-$ has B.R. of 4.7×10^{-5}, $\omega \to e^+ + e^-$ has B.R. of 7.28×10^{-5} $\iff$ **rare probe.**
- Hence clean sample of electrons is mandatory to reconstruct di-electrons with high signal purity.

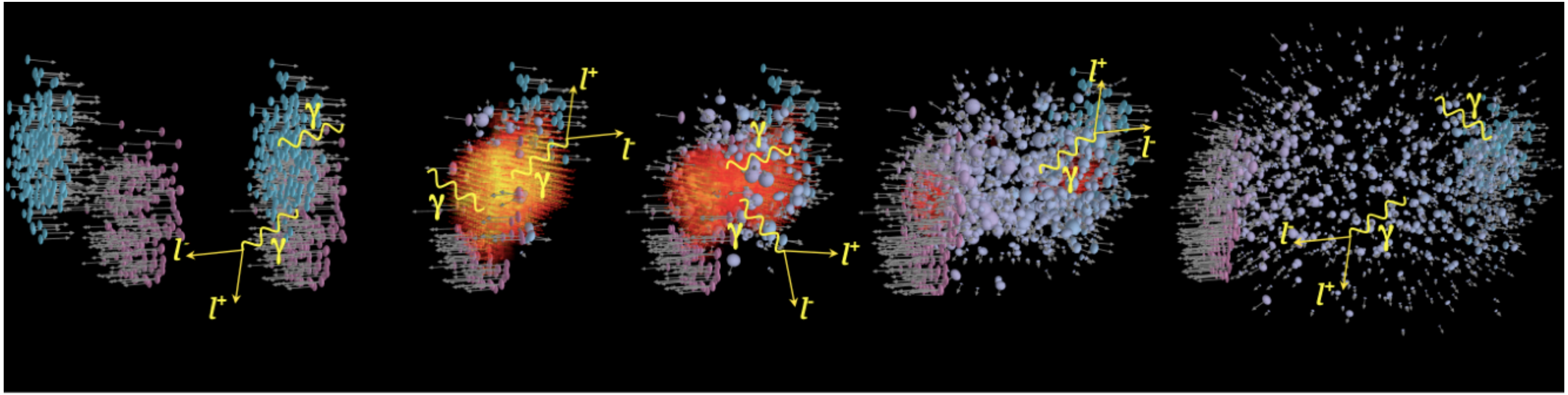
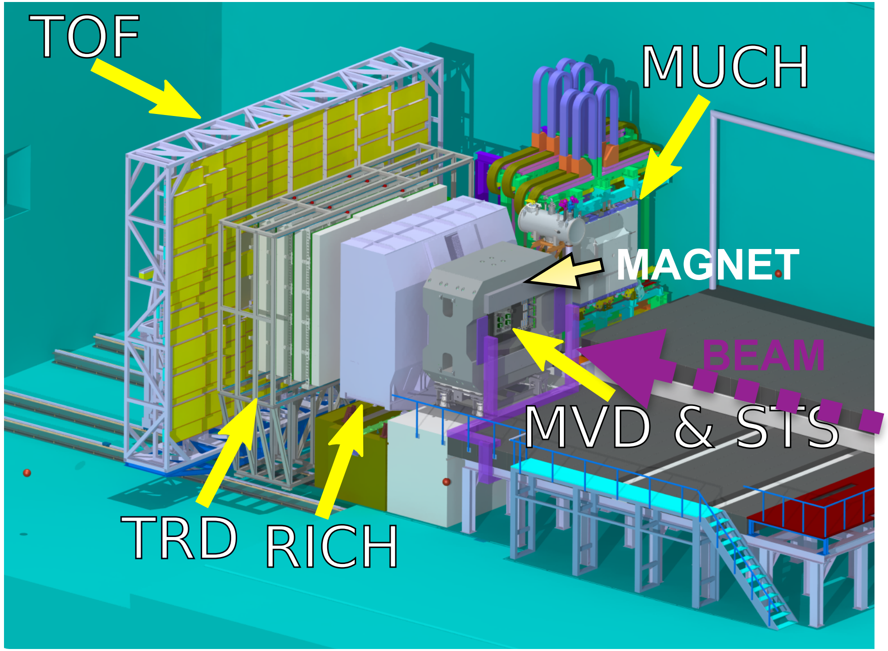


Figure: Florian Seck

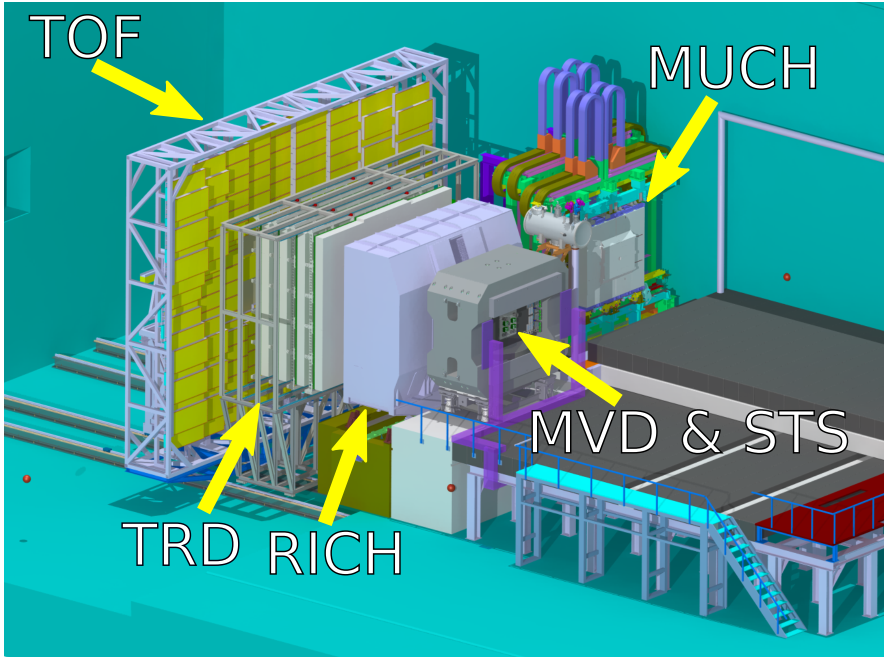
Overview of CBM



- Compressed baryonic matter (CBM) is a future fixed target experiment at Facility for anti-proton and ion research (FAIR), Darmstadt.
- Designed to explore QCD matter at high baryonic density and moderate temperatures.

MVD - Micro vertex detector
STS - Silicon tracking system
RICH - Ring Imaging Cherenkov detector
MUCH - Muon chamber
TRD - Transition radiation detector
TOF - Time of flight
FSD - Forward spectator detector

Overview of CBM



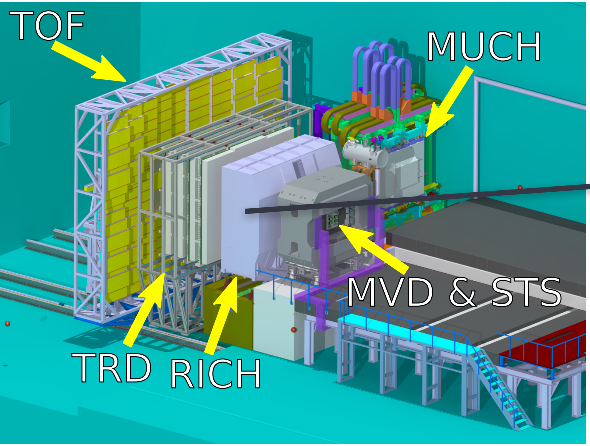
- Tracking acceptance : 2.5^0 - 25^0.
- High interaction rate : max. 10 MHz - Muon setup
  : few 100 kHz - Electron setup
- Free streaming readout.
- Online 4D reconstruction and event selection.

Electron setup :

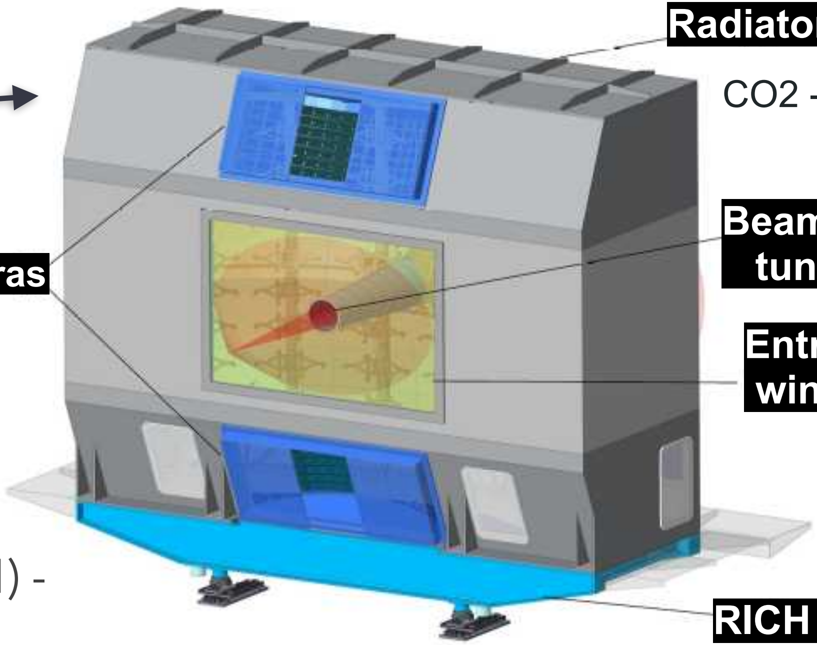
Tracking	- MVD + STS
electron/pion separation	- RICH + TRD
Other hadrons	- TOF

Muon setup :
STS + MUCH + TRD + TOF

CBM Ring Imaging Cherenkov detector



CBM detector

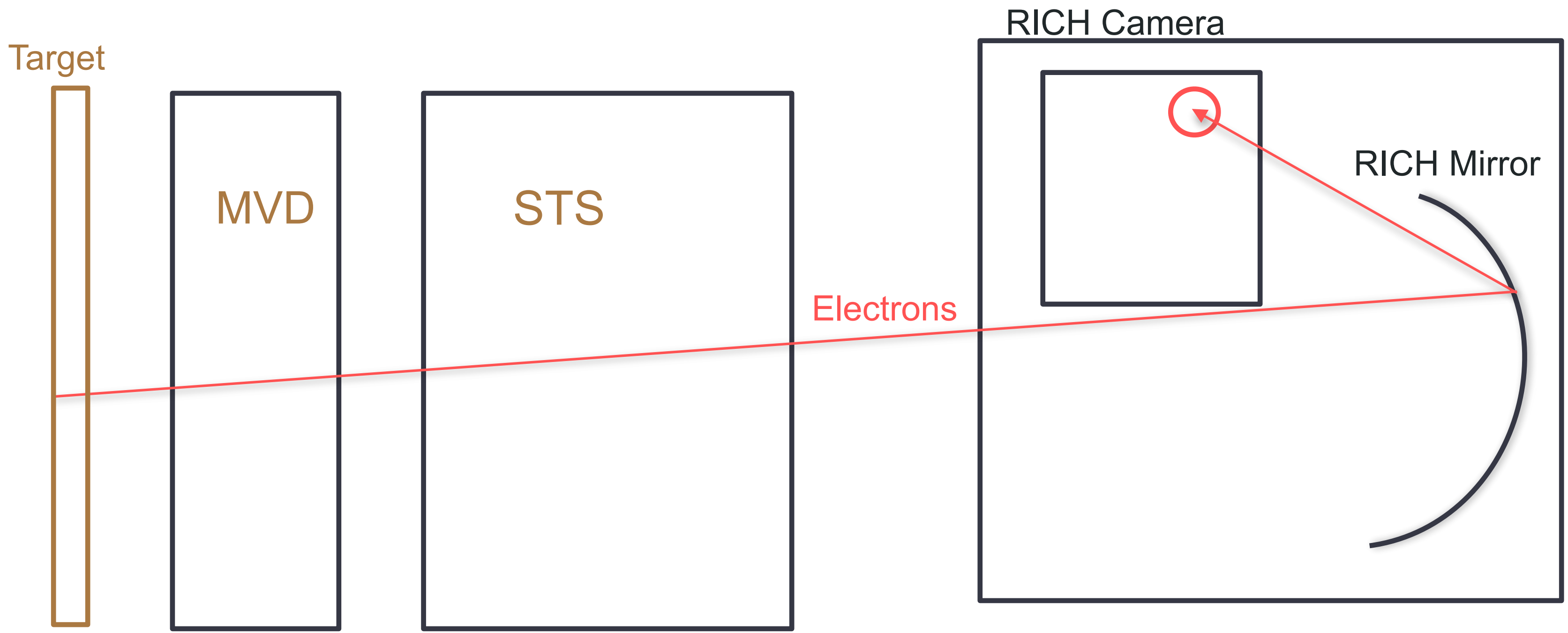


RICH Detector

Ring imaging Cherenkov detector (RICH) - identify e^- / e^+ for **p < 6 GeV/c.**

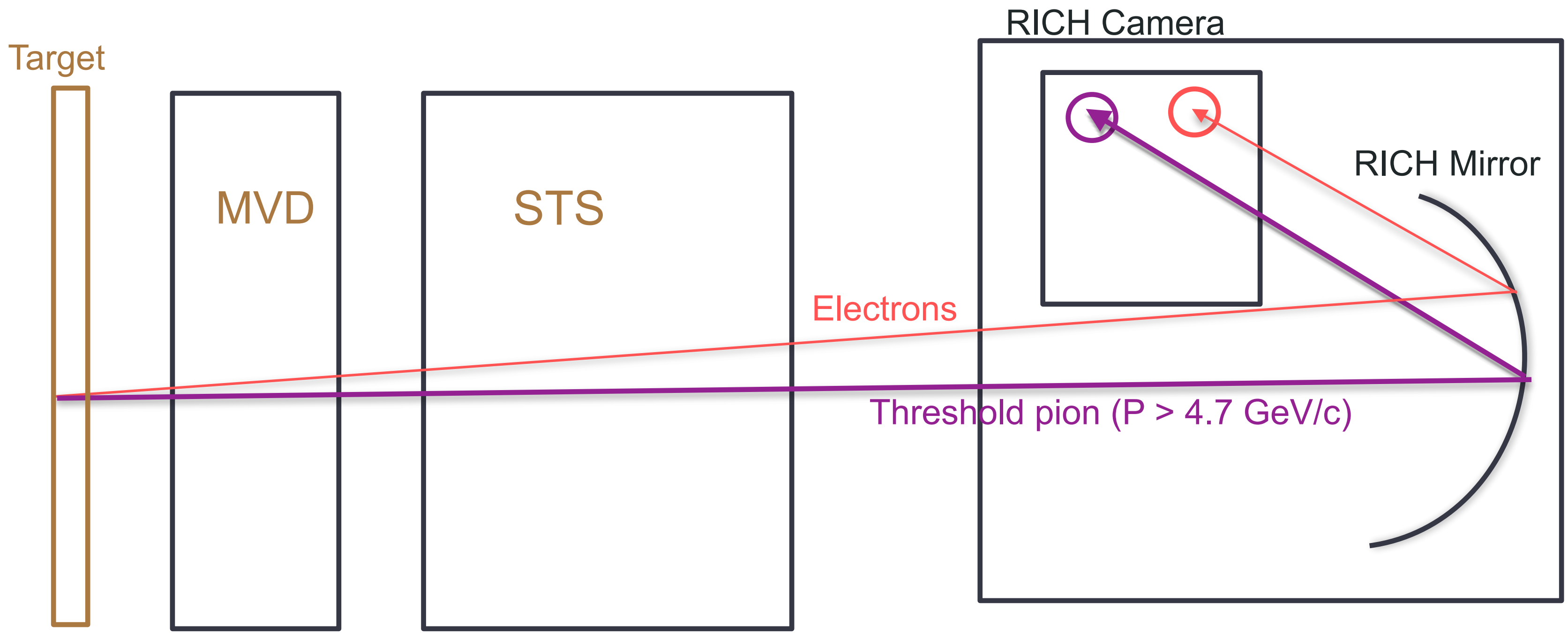
CO2 radiator -> Pion momentum threshold of 4.7 GeV/c.

Rings and tracks:



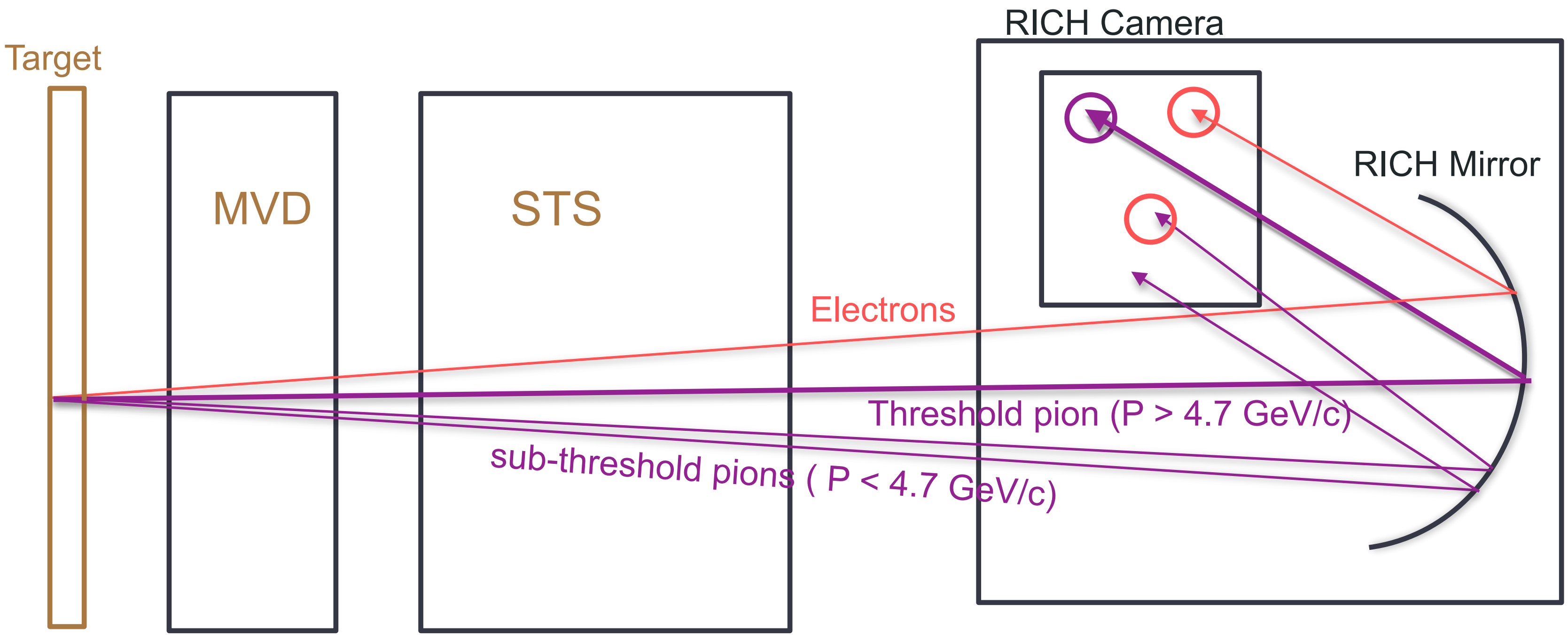
Schematic indicating electrons from target forming rings in RICH

Rings and tracks



Schematic indicating electrons and pions from target forming rings in RICH

Rings and tracks



Schematic indicating electrons and pions from target forming rings in RICH, and sub-threshold pions mismatching to electron rings

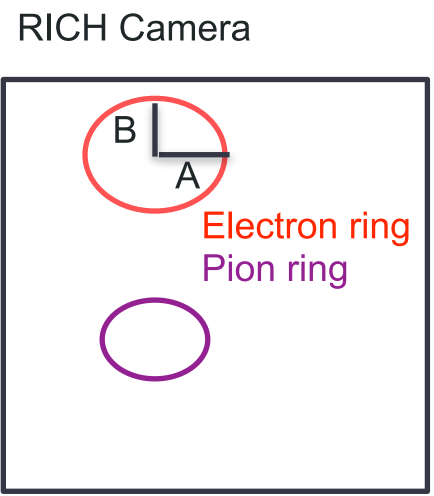
Simulation setup:

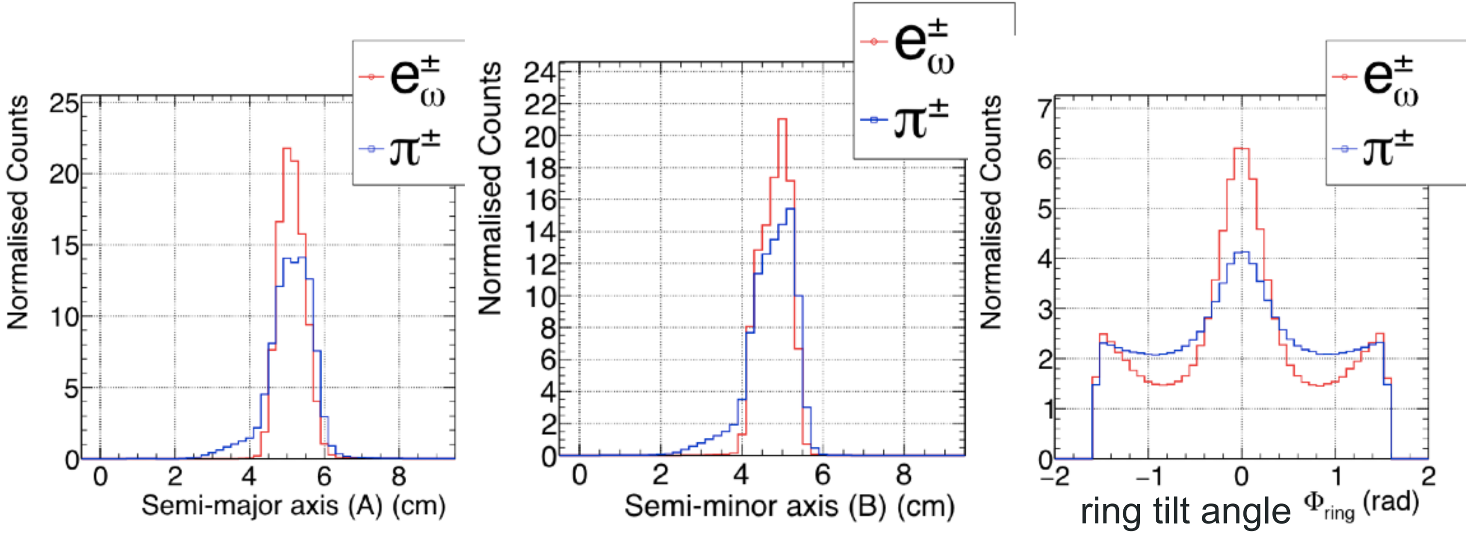
- Pluto : Omega -> e+ e- (1 per event)
- UrQMD : 8 AGeV/c, Central Au - Au collisions.
- Events : 1M
- Geometries : Standard FEB23 CBMROOT electron setup

Ring-track selection algorithm:

- All pairs with $\delta d = \sqrt{(x_{track} - x_{ring})^2 + (y_{track} - y_{ring})^2}$ < 10 cm are selected.
- The pair data set consists of electrons, threshold pions and mis-matched pions.
- Performance of algorithms are evaluated for Omega embedded UrQMD data set.

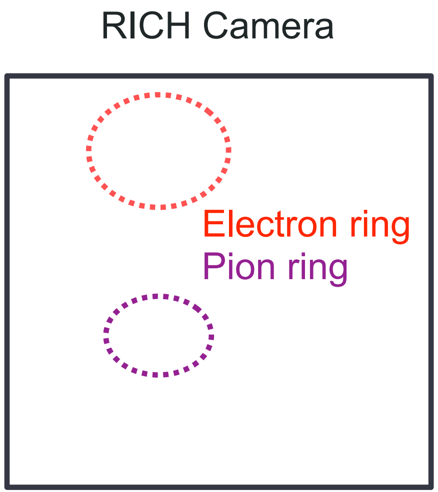
Ring-Track matching : Parameters

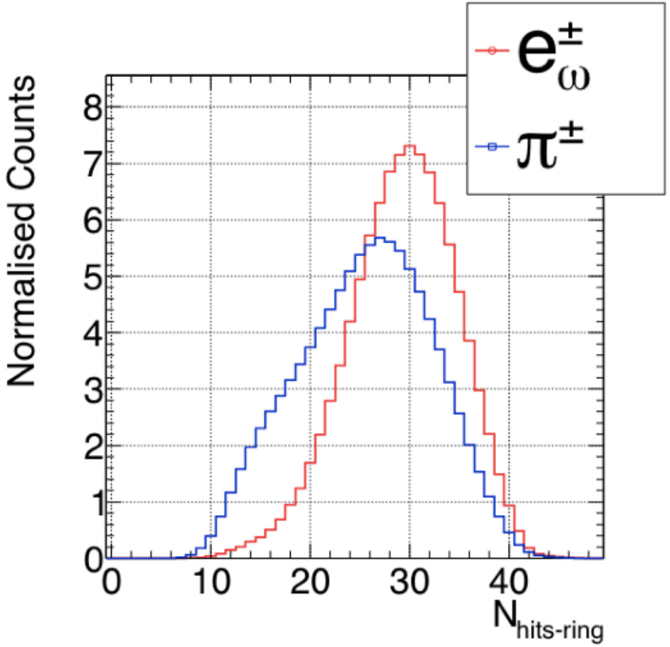


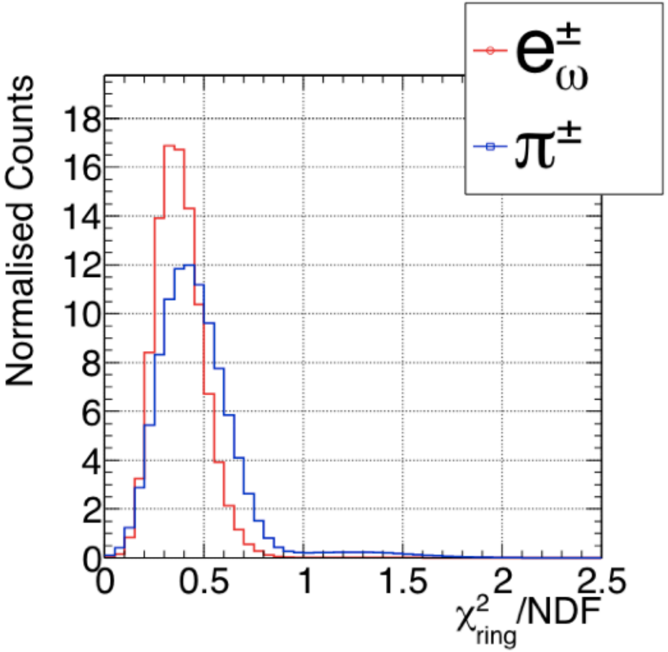


- The rings in RICH are partially elliptical.
- They are characterised by semi-major and minor axis and the ring tilt angle.
- Pion rings are typically smaller compared to electron rings.

Ring-Track matching : Parameters



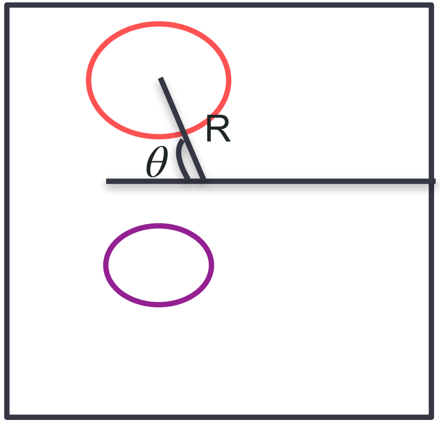


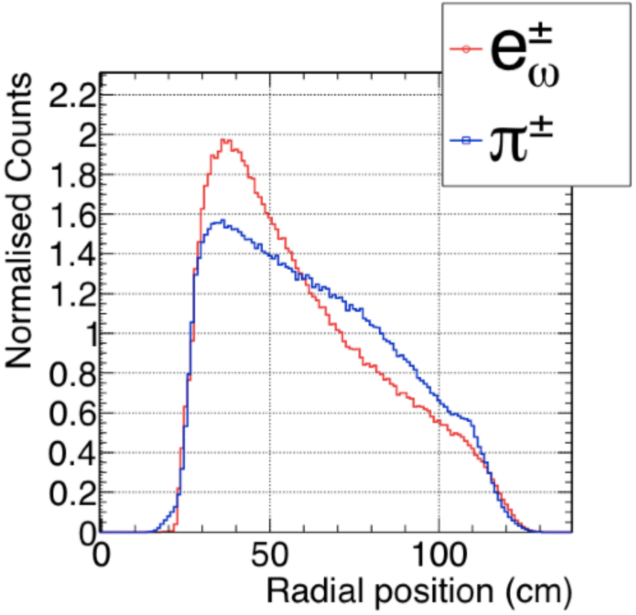


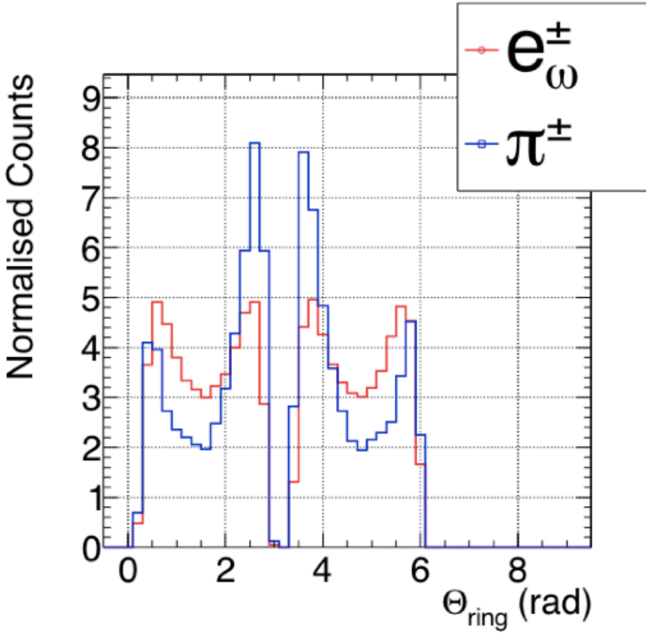
- Number of hits for electron rings are higher compared to pion rings (smaller rings).
- This translates to higher chi-square for the ring fitter.

Ring-Track matching : Parameters

RICH Camera

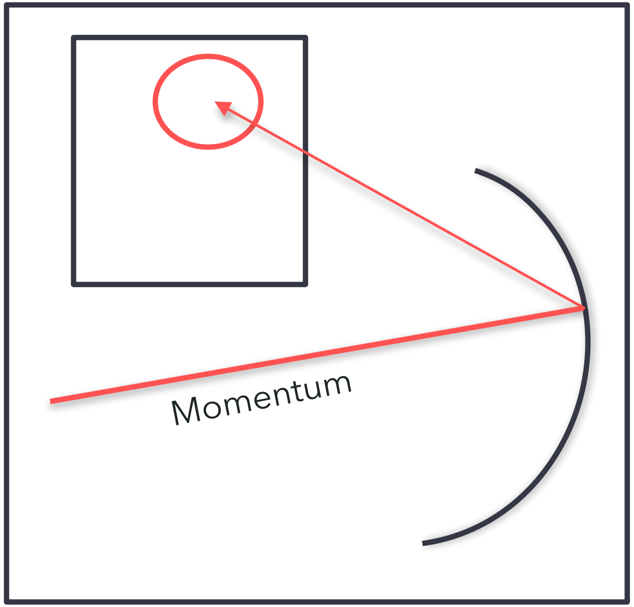




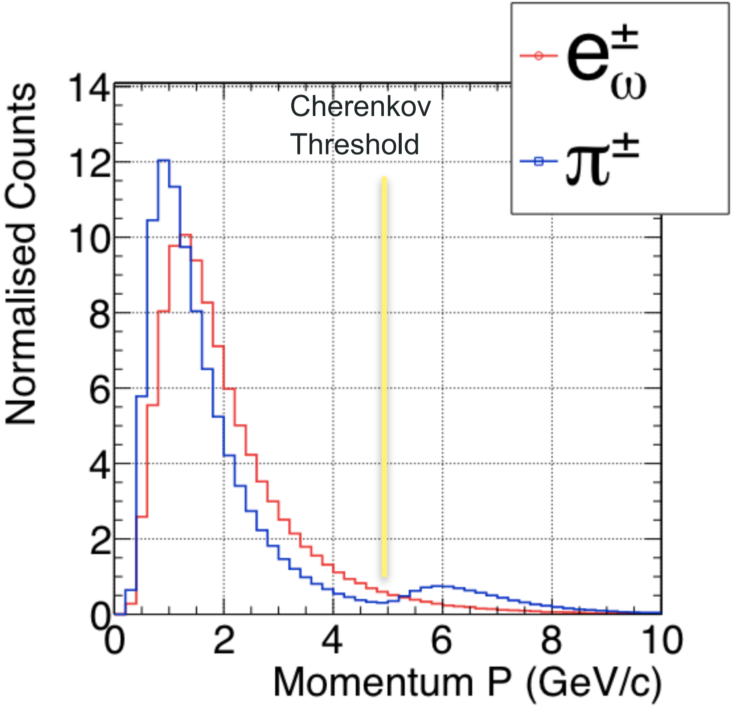


Absolute position of the rings in the RICH plane in polar coordinates

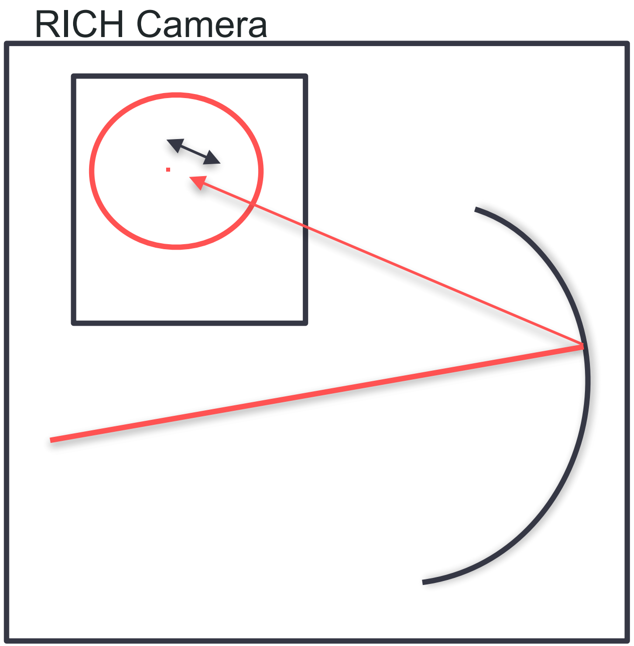
Ring-Track matching : Parameters

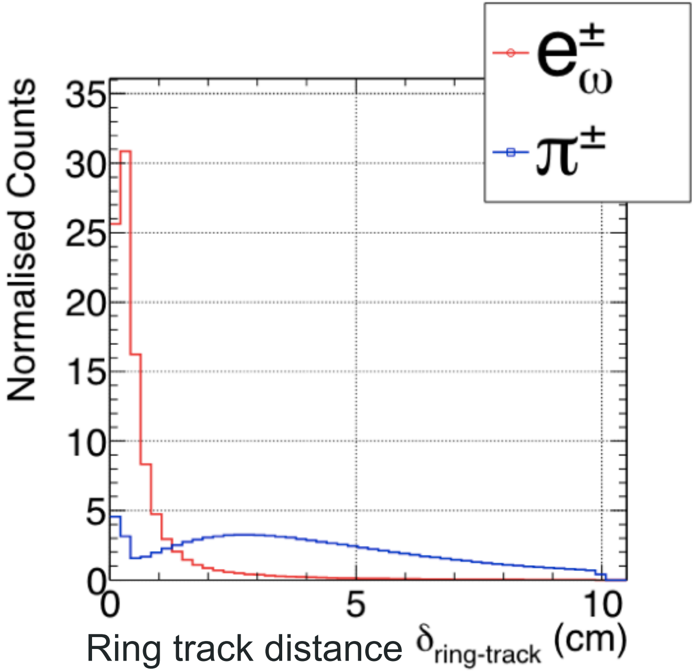


Track parameter : Momentum



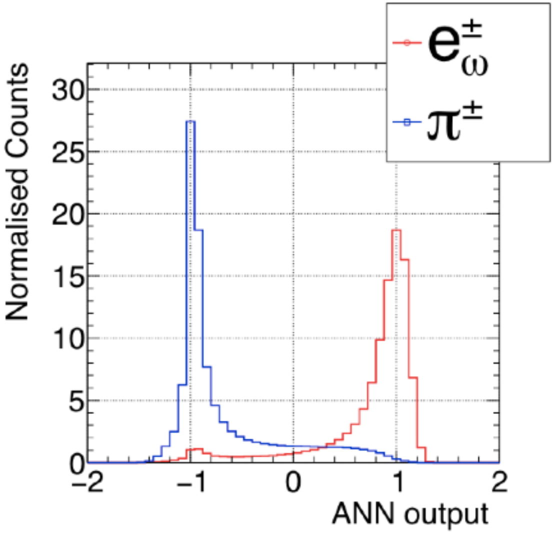
Ring-Track matching : Parameters



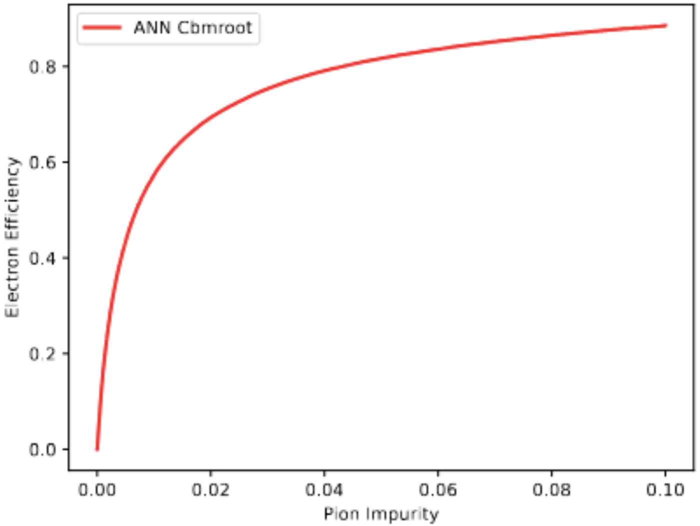


Clearly, best parameter with most separating power.

ANN in CBMROOT : Response



ANN response

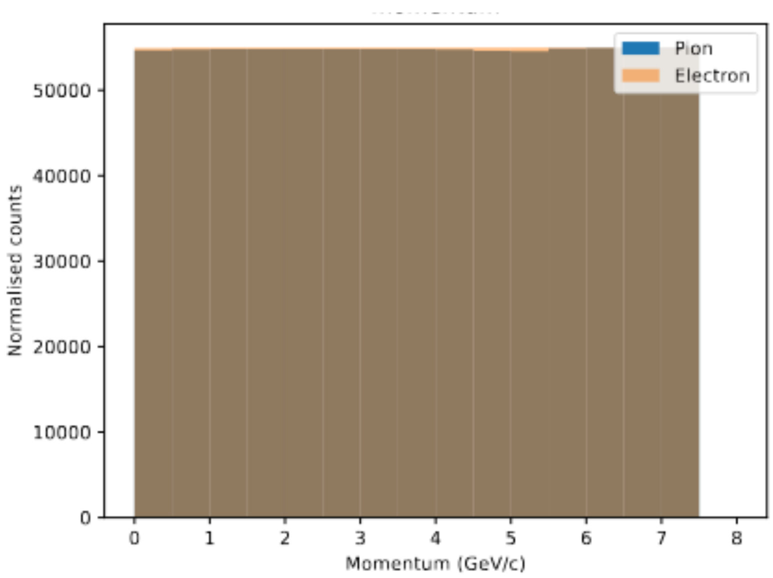


ROC curve for different threshold of the response

- 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.

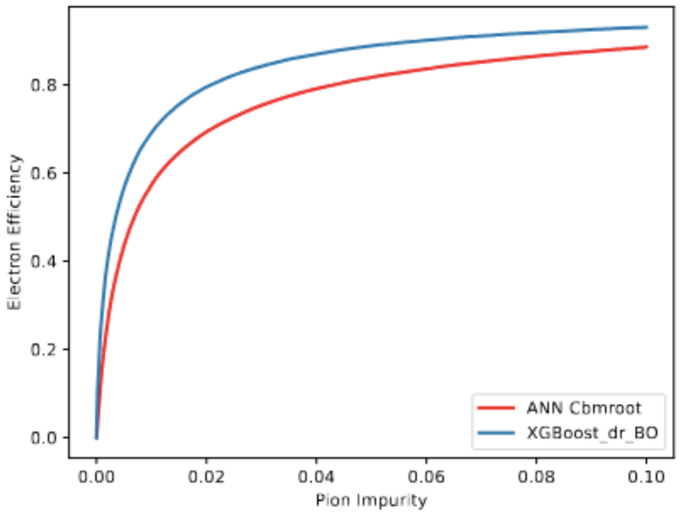
Training strategy :

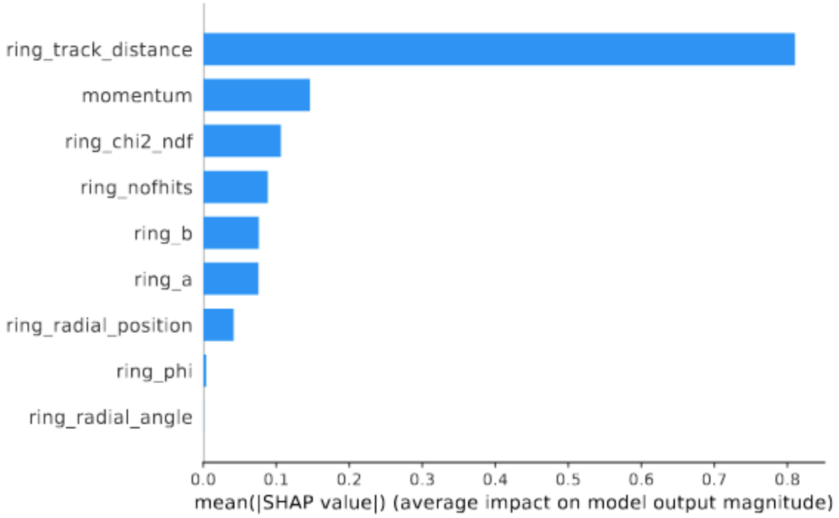
- ML model should be used for all energies -> Momentum independent learning.
- A set of 790×10^3 single electrons and 790×10^3 UrQMD pions are used for training.



- Hard cut on 7.5 GeV/c is by choice.
- All the hyper parameters are optimised using Bayesian approach.

XGBOOST : Response for Omega embedded sample



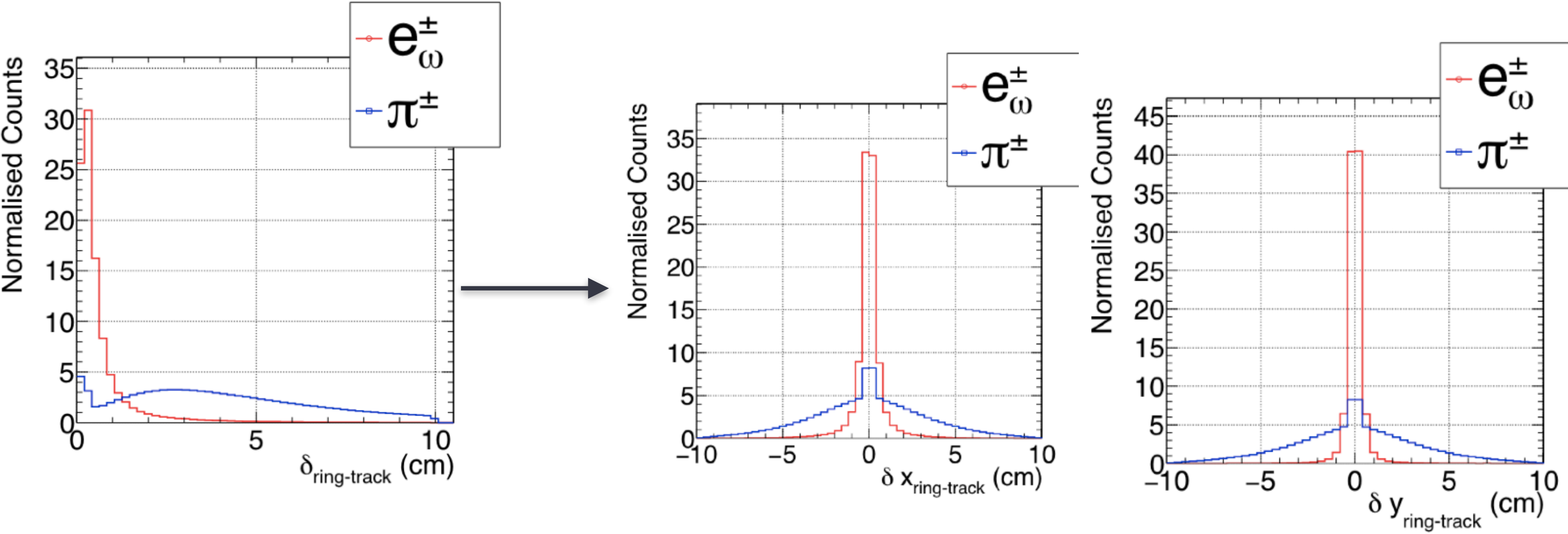


ROC curve for different threshold of the response

Variable importance (SHAP values)

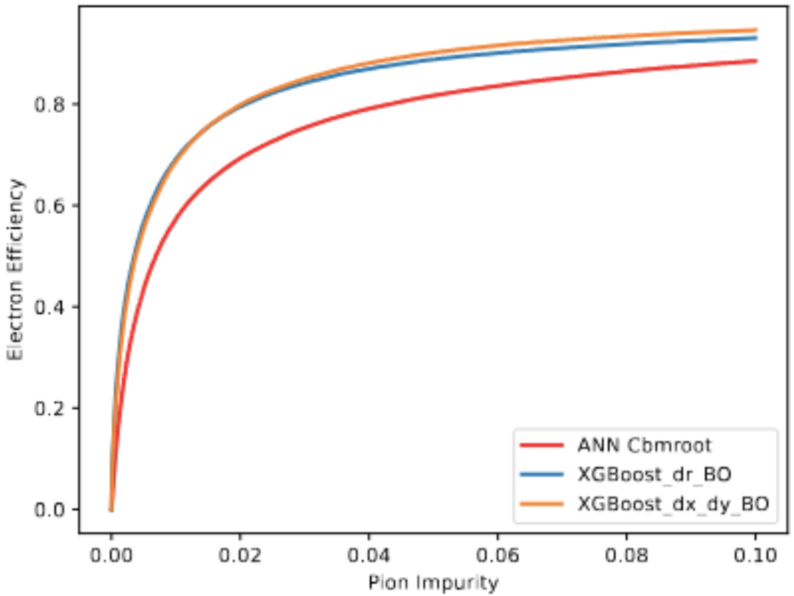
- Improved efficiency for omega electrons and improved suppression of pions.
- Some of ring parameters has less importance.

Ring track distance :

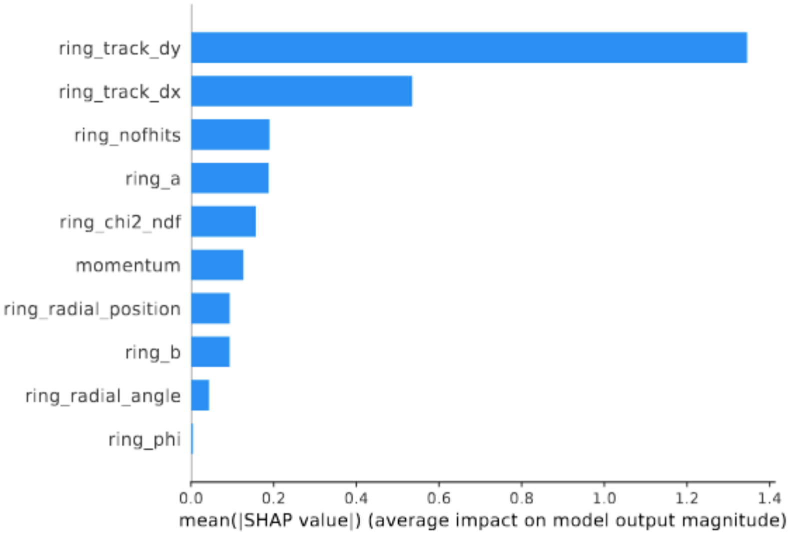


- The Y component of ring track distance is sharper compared to X.
- The increase is due to extrapolation error and the fit error in magnetic field (XZ).
- Model is trained with {dx, dy } in order to capture this differential width.

XGBOOST with dr->dx,dy : Response for Omega embedded sample



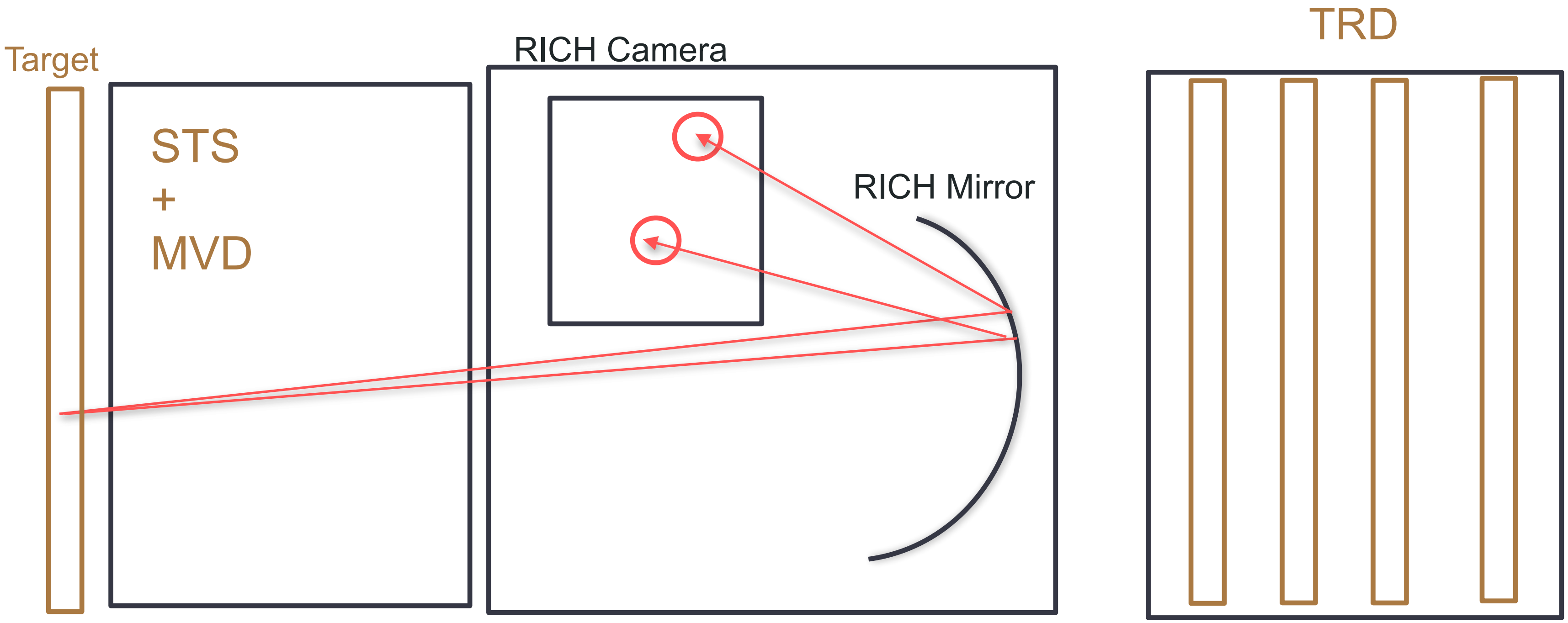
ROC curve for different threshold of the response



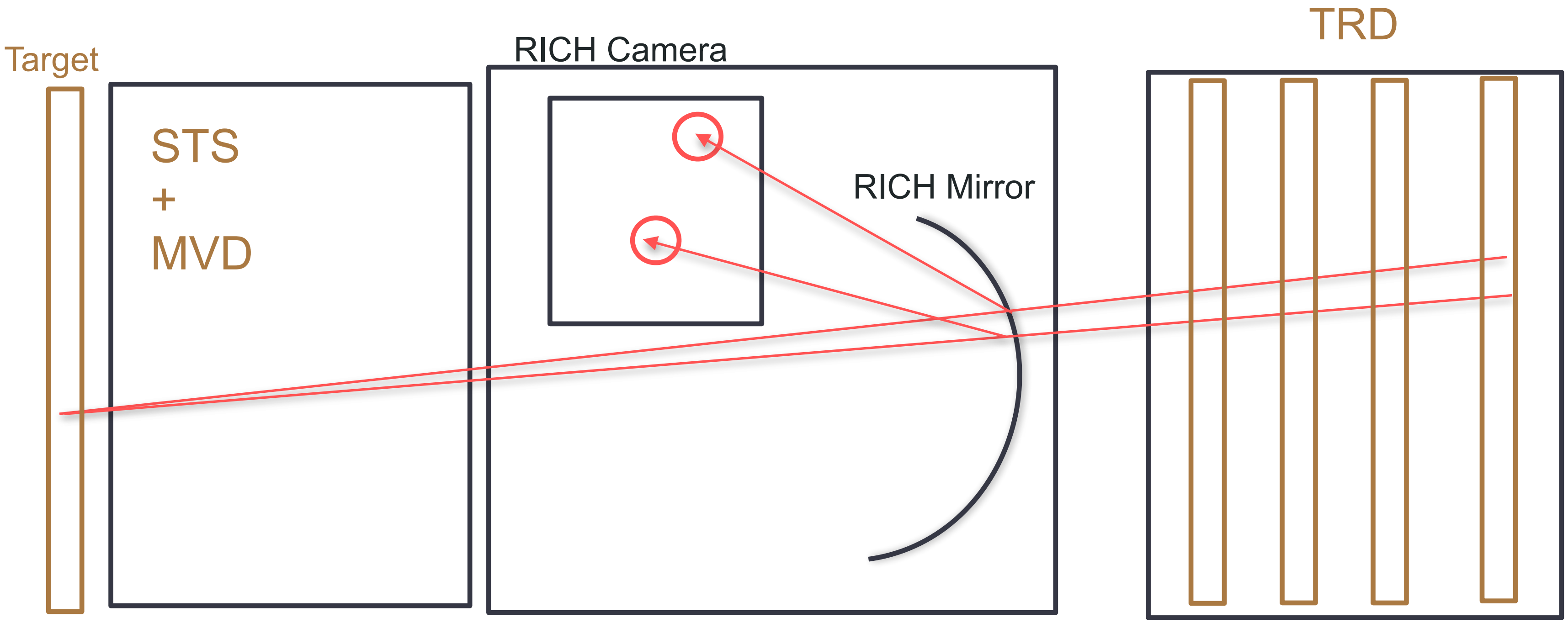
Variable importance (SHAP values)

- The differential distance improves the efficiency in large impurities.
- dy has large impact compared to dx.

Ring-Track matching

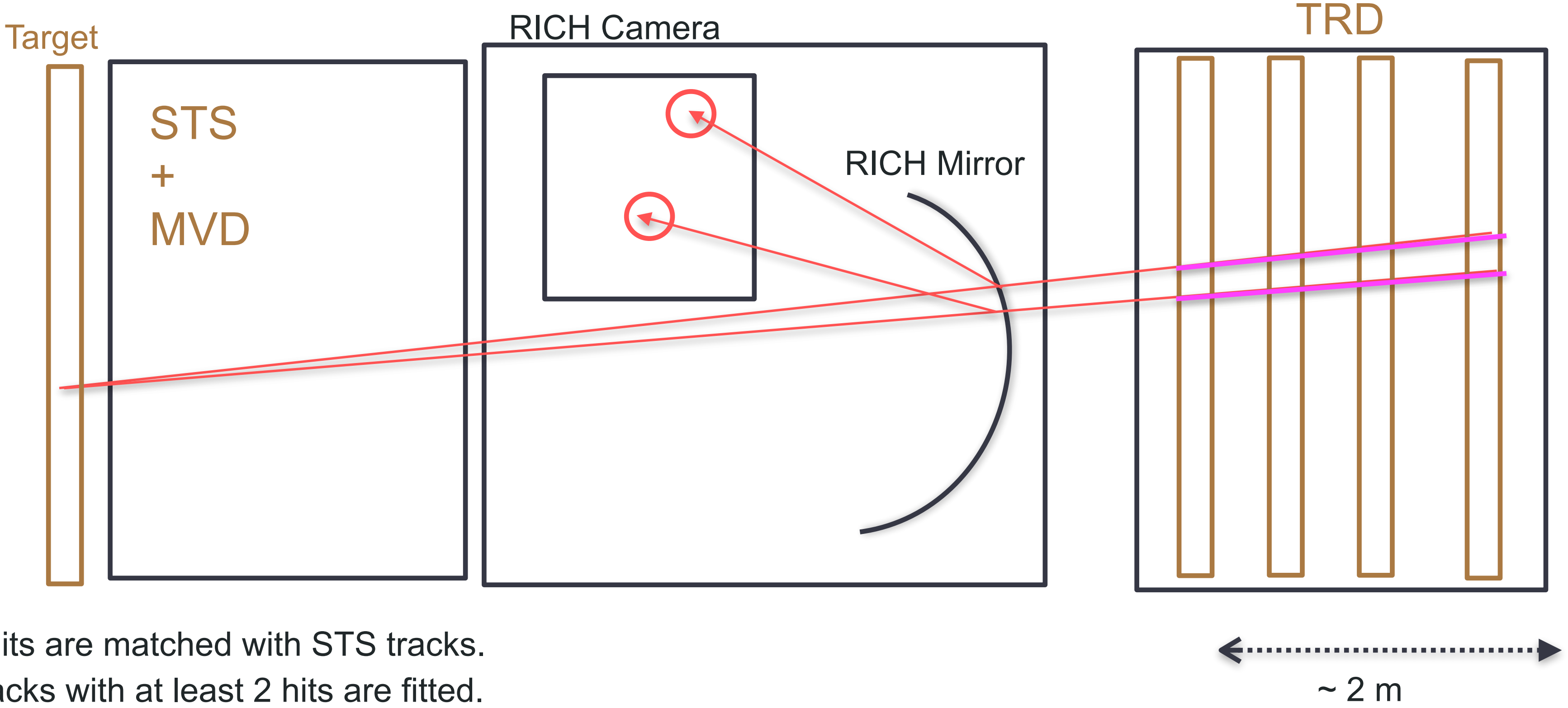


Ring-Track matching



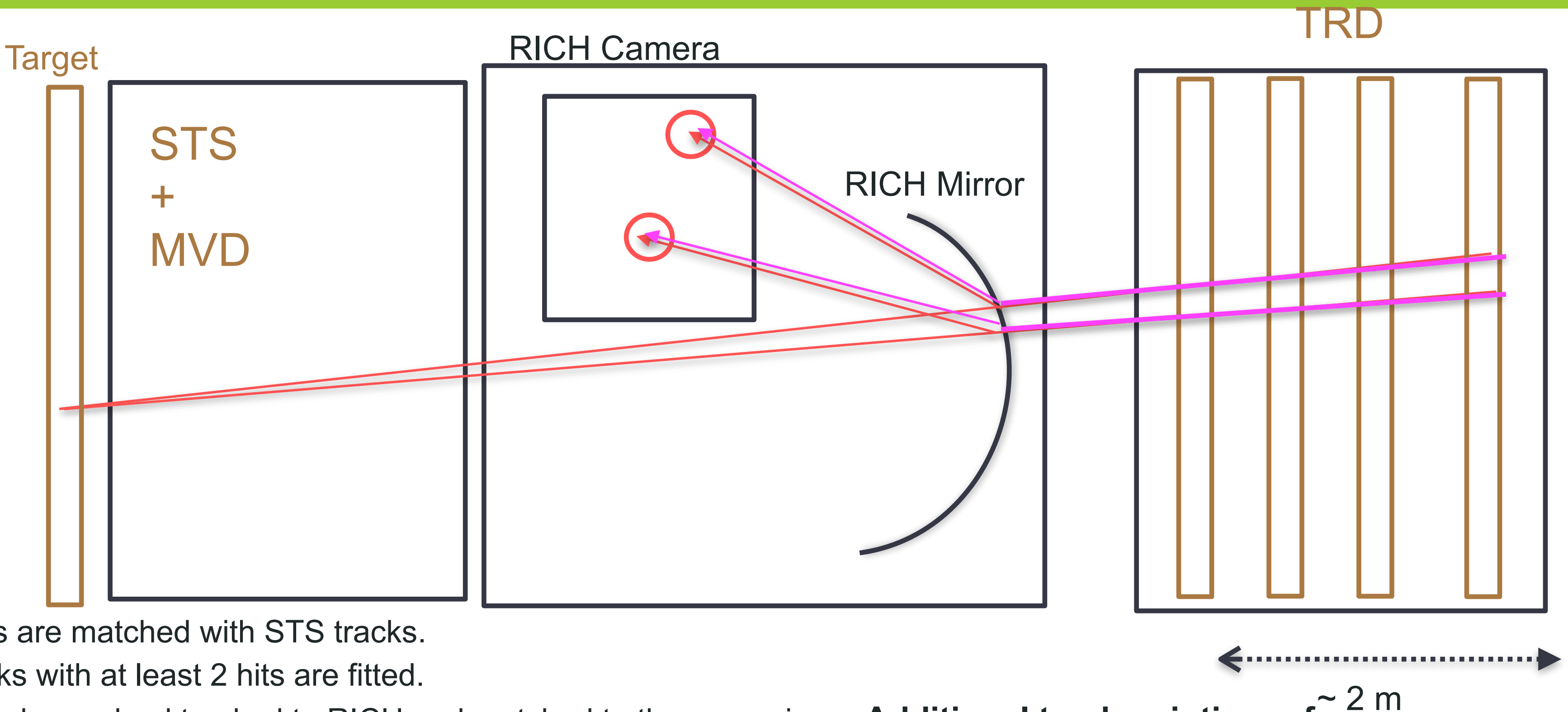
- TRD Hits are matched with STS tracks.

Ring-Track matching



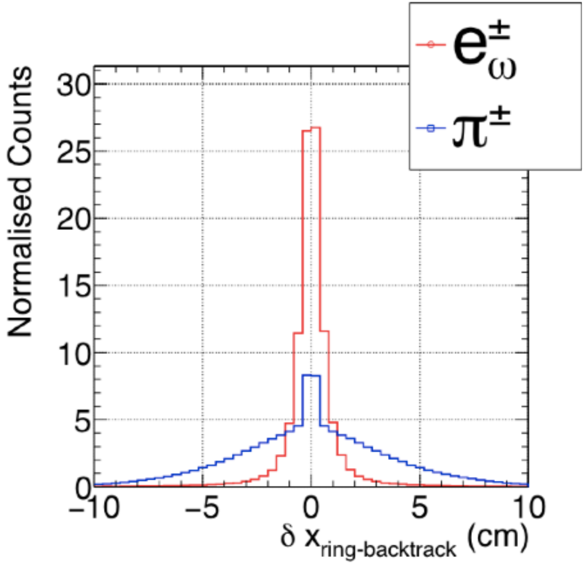
- TRD Hits are matched with STS tracks.
- The tracks with at least 2 hits are fitted.

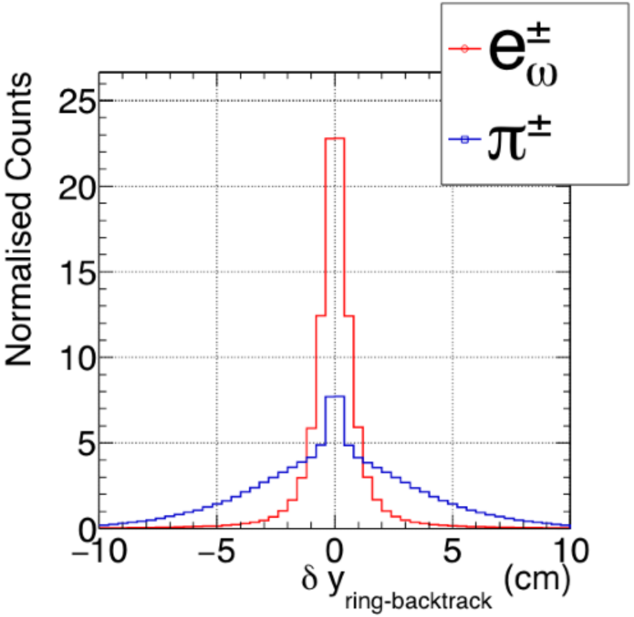
Ring-Track matching

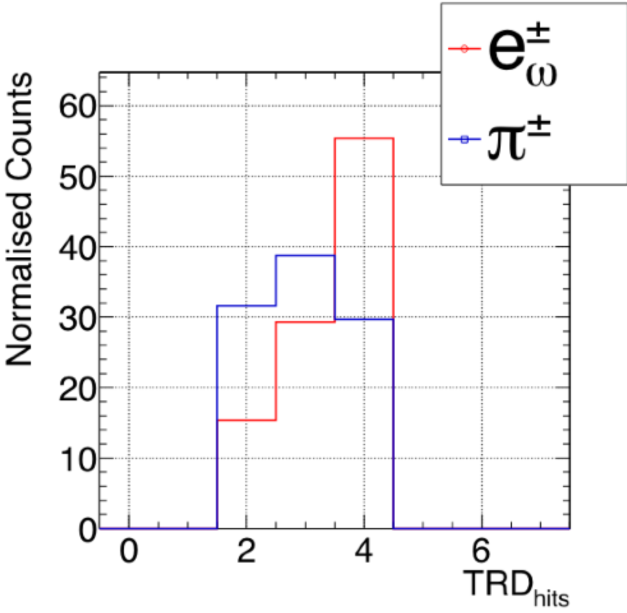


- TRD Hits are matched with STS tracks.
- The tracks with at least 2 hits are fitted.
- Fitted tracks are backtracked to RICH and matched to the same ring-> **Additional track pointing reference**

Additional parameters:

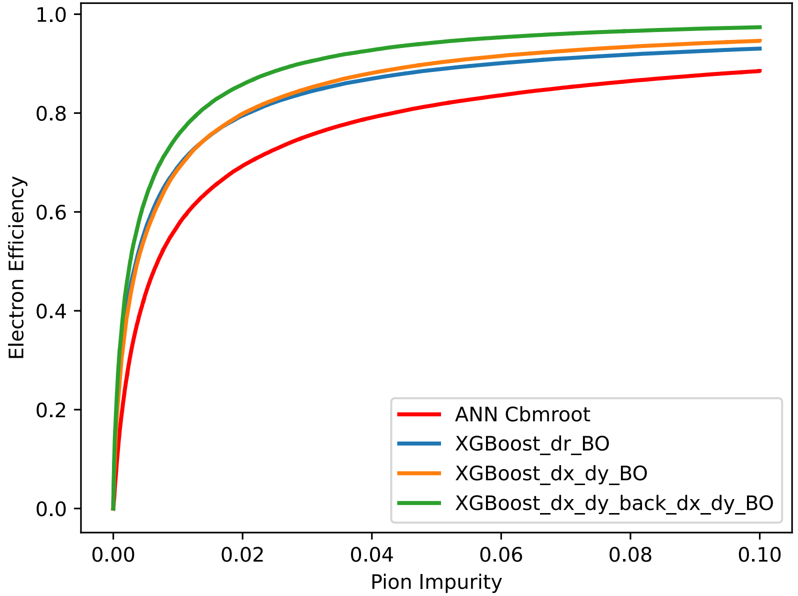




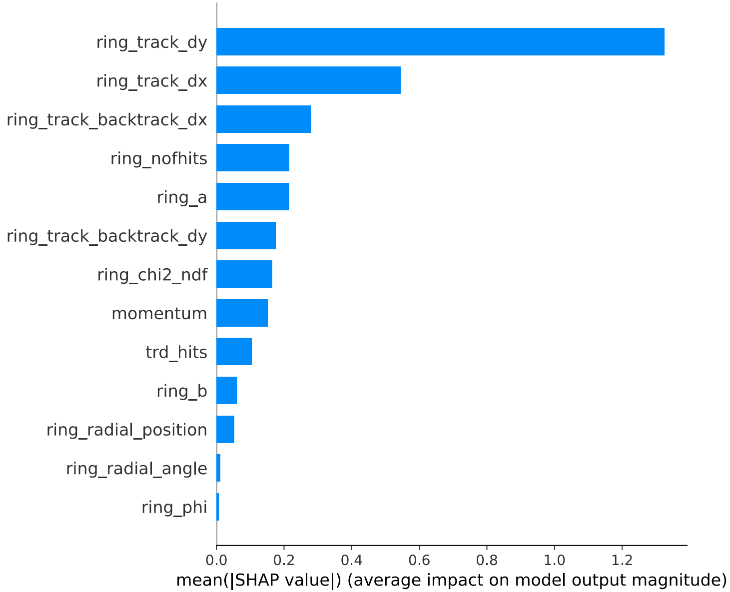


- Number of TRD hits and backtracked distances are added as additional parameters to the ML model.

XGBOOST with TRD parameters : Response for Omega embedded sample



ROC curve for different threshold of the response

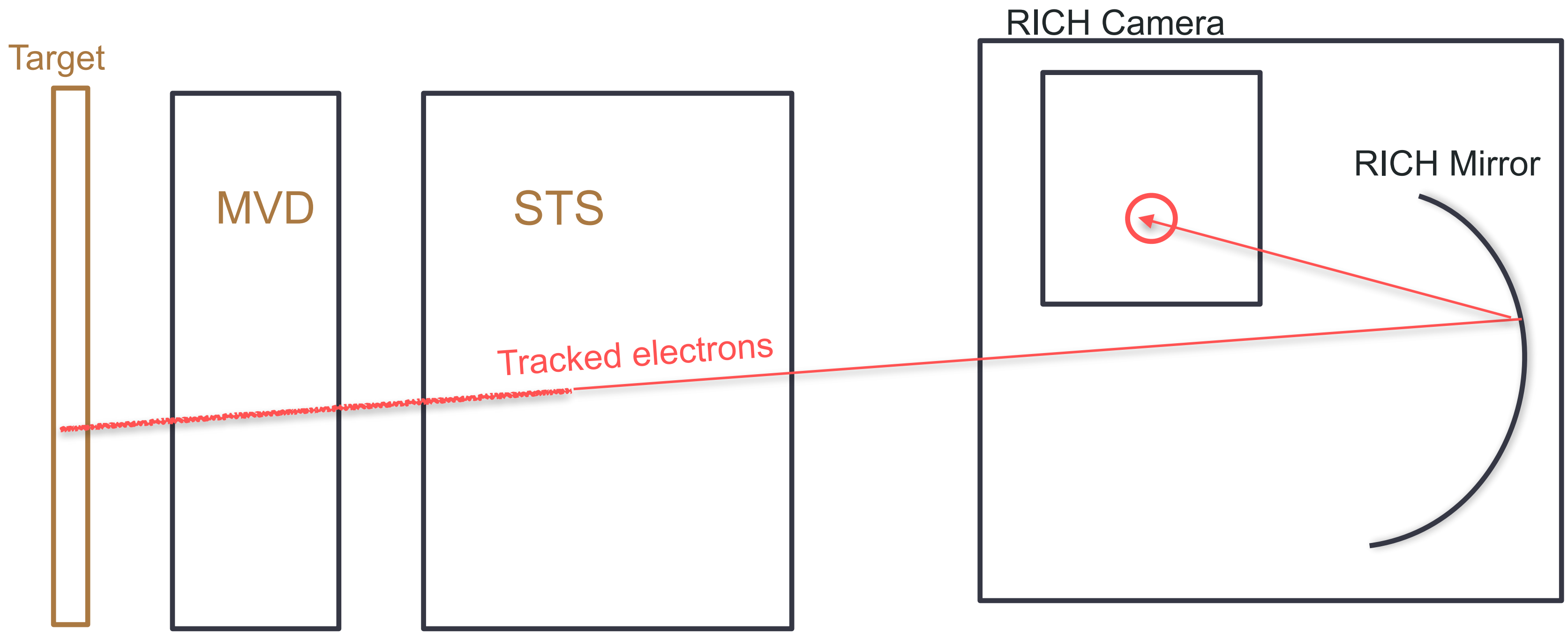


Variable importance (SHAP values)

- 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.

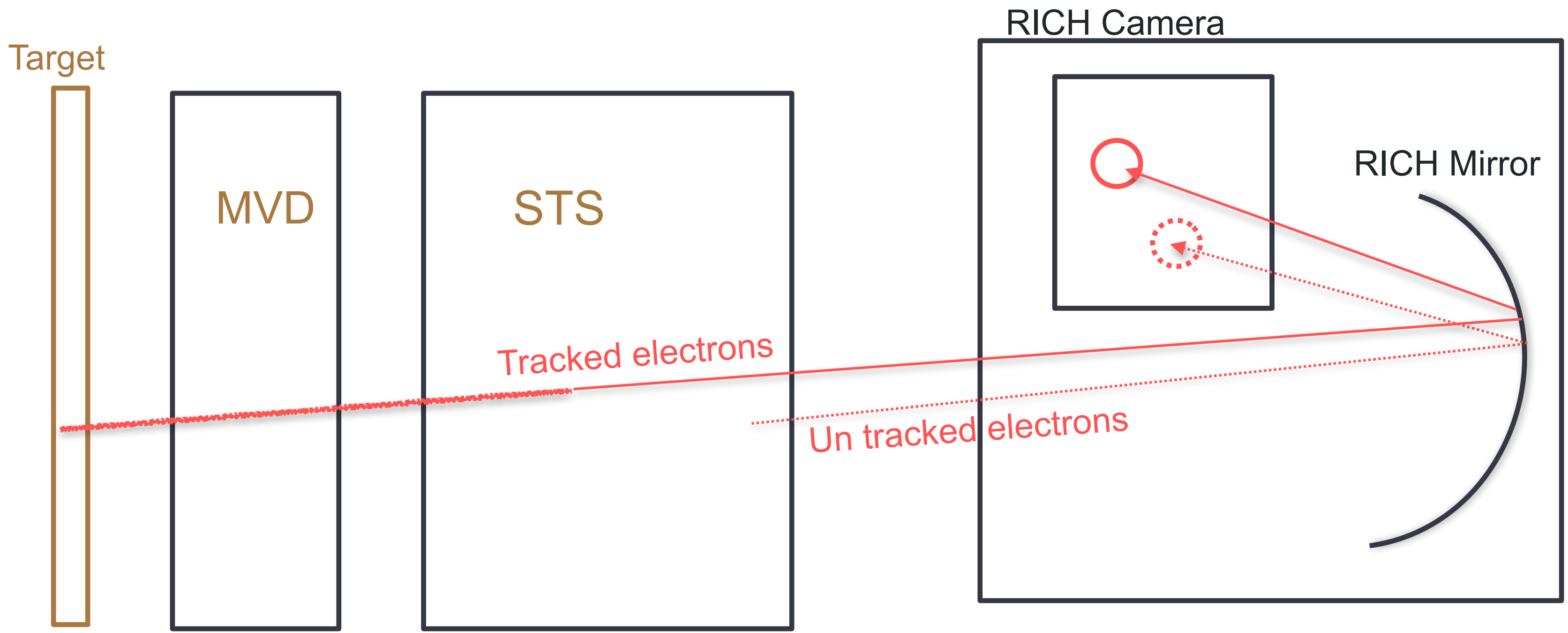
Conversion electrons : Feasibility study

Mis-matched pions:



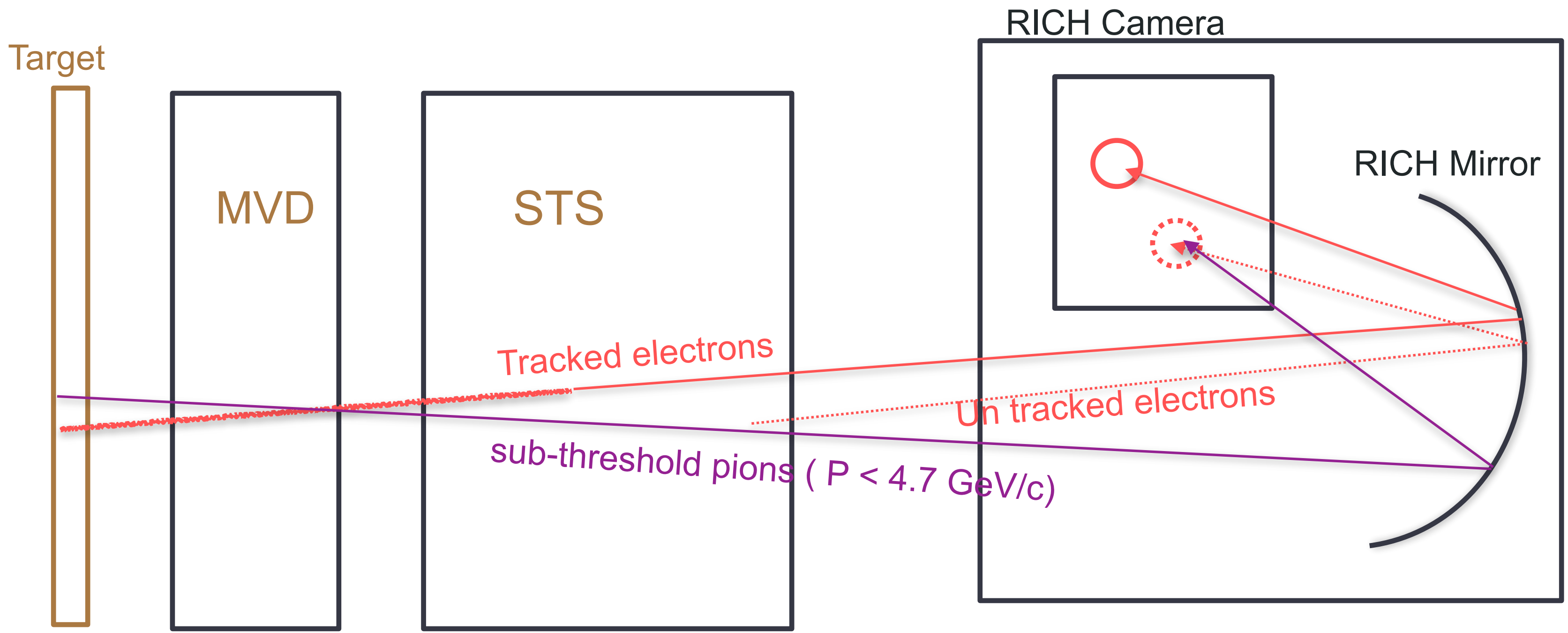
Schematic indicating electrons tracked in STS forming rings in RICH

Mis-matched pions:



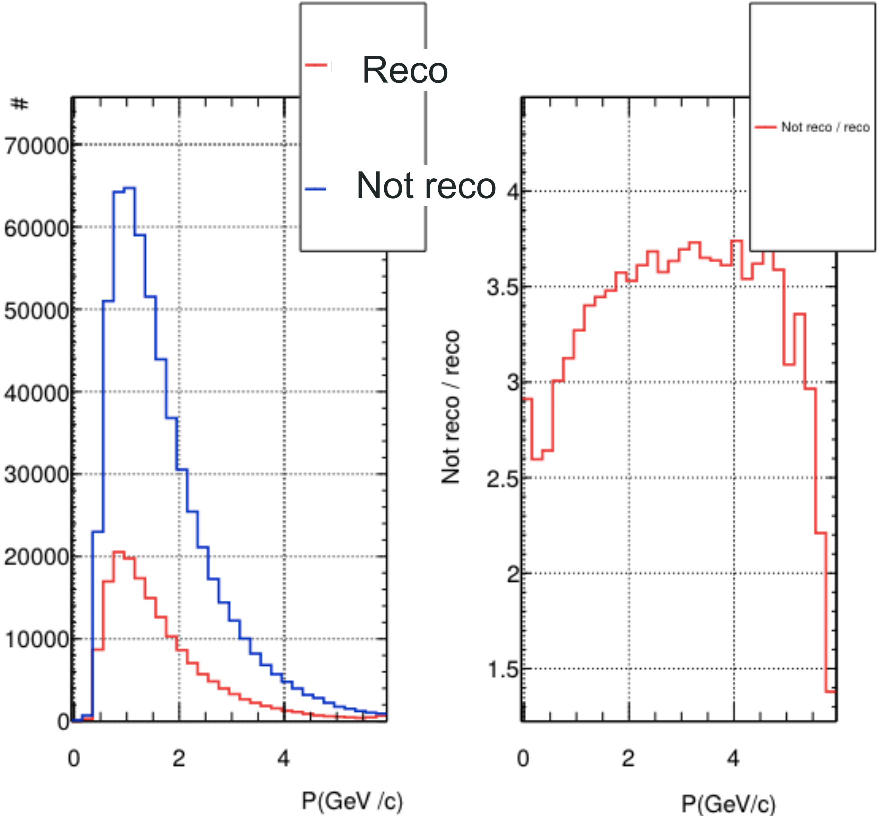
Schematic indicating electrons tracked in STS and untracked electrons forming rings in RICH

Mis-matched pions:



Schematic indicating electrons from target forming rings in RICH

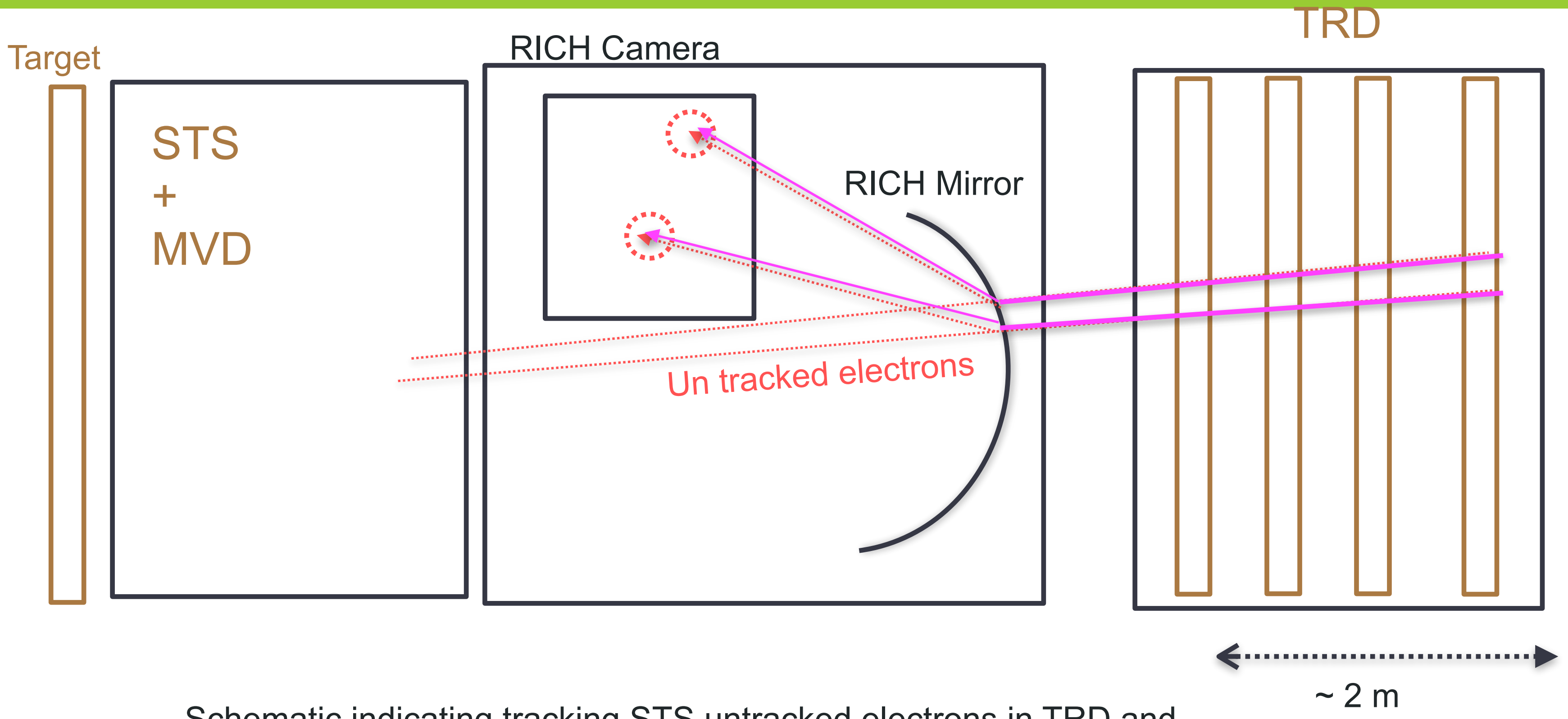
Mis-matched pions:



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 :

Hit filtering

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

↓

Track finder

↓

Least square linear fitter

↓

Extrapolation to RICH

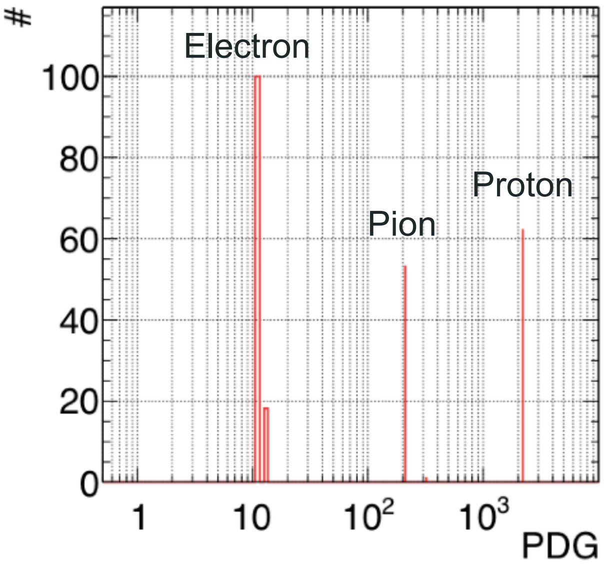
↓

Find closest ring

↓

Apply conversion probability for matched ring

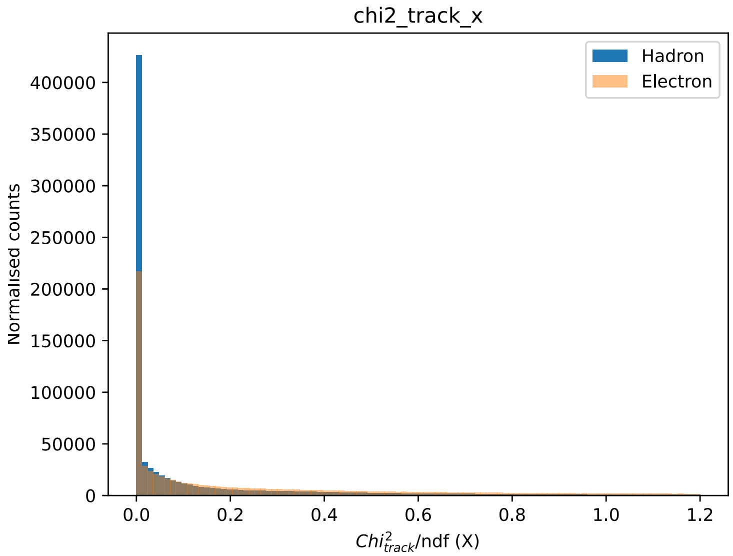
TRD tracks:

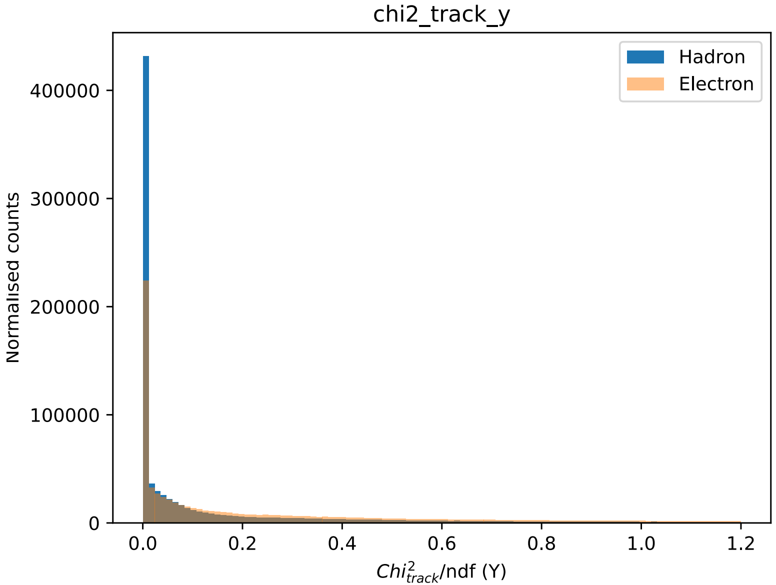


- “Signal” - Electrons not tracked in the STS, forms the rings in RICH.
- “Background” - Other particles

- 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.

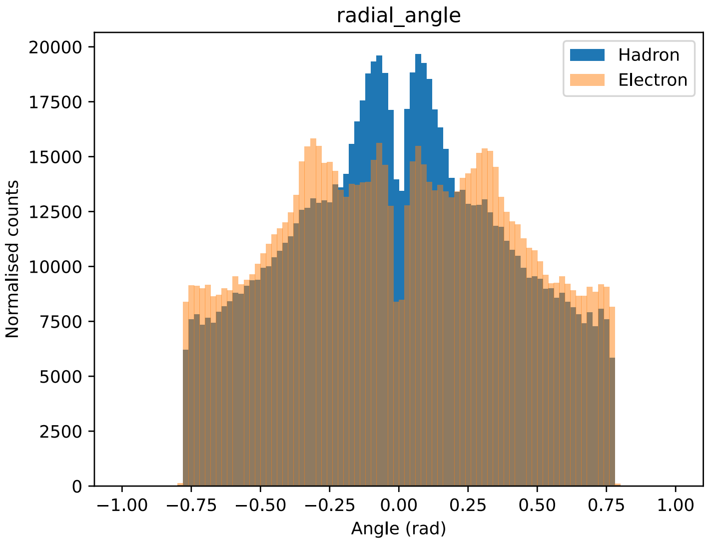
Conversion probability : Parameters

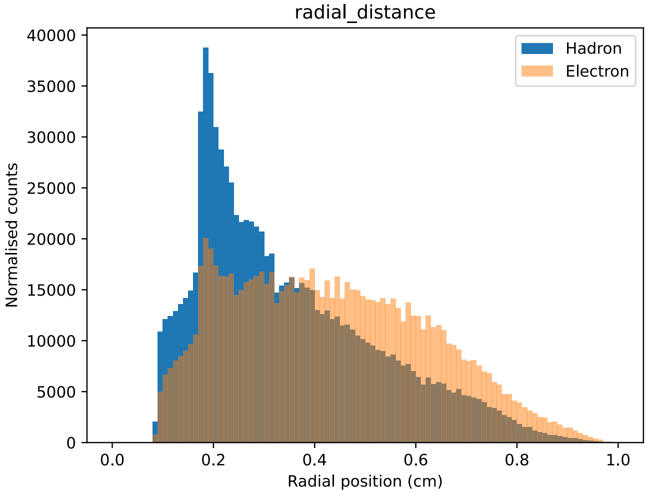




TRD track fit parameters

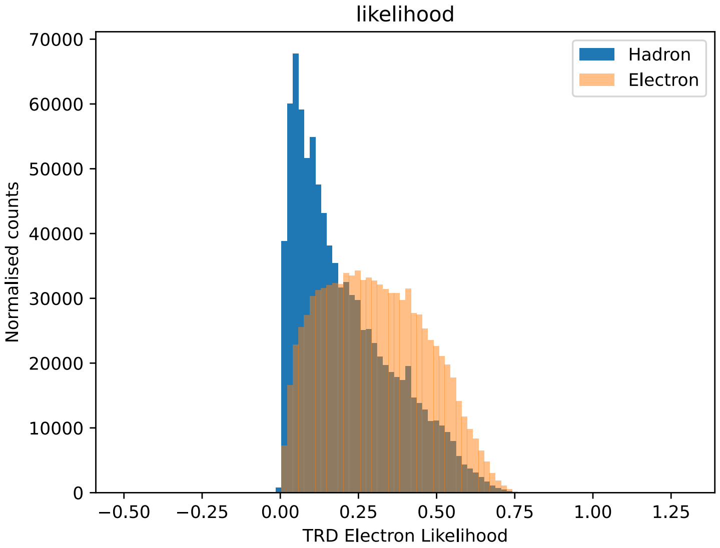
Conversion probability : Parameters



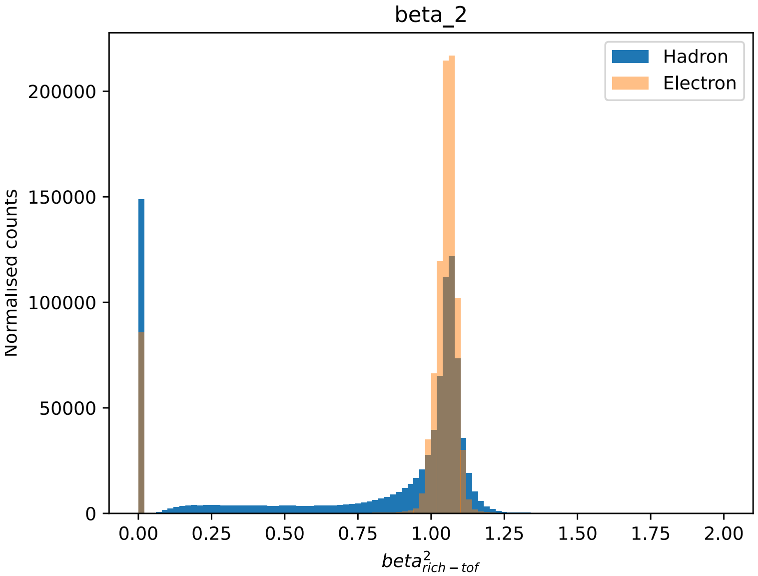


Absolute position of the first hit of the TRD track

Conversion probability : Parameters

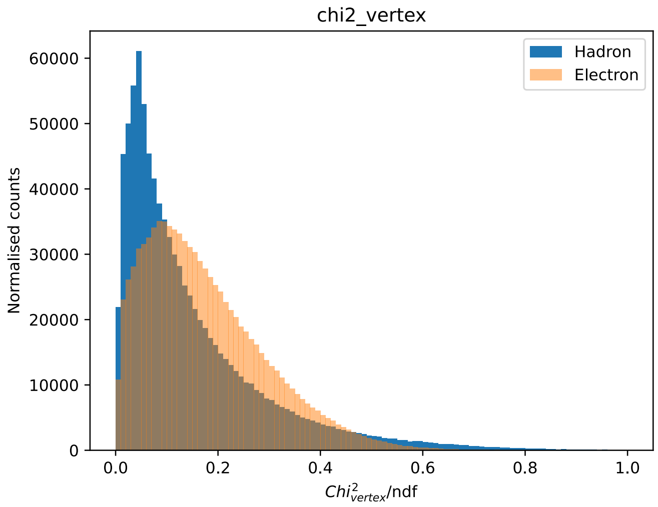


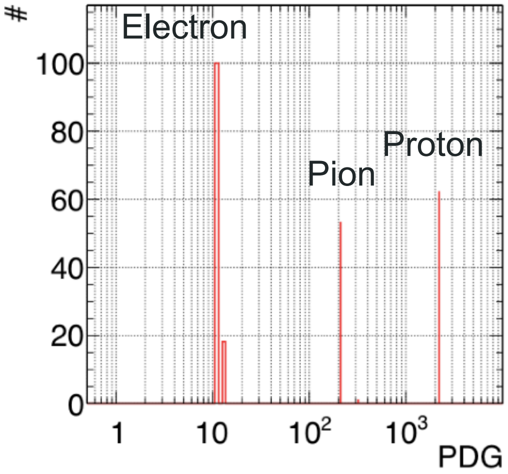
TRD electron likelihood for the fitted tracks

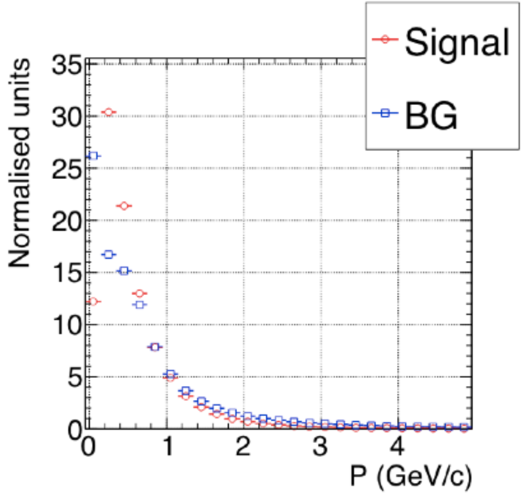


- Extrapolated the fitted track towards TOF.
- Using ring time as the start time and TOF time as stop time, velocity is derived.

Conversion probability : Parameters

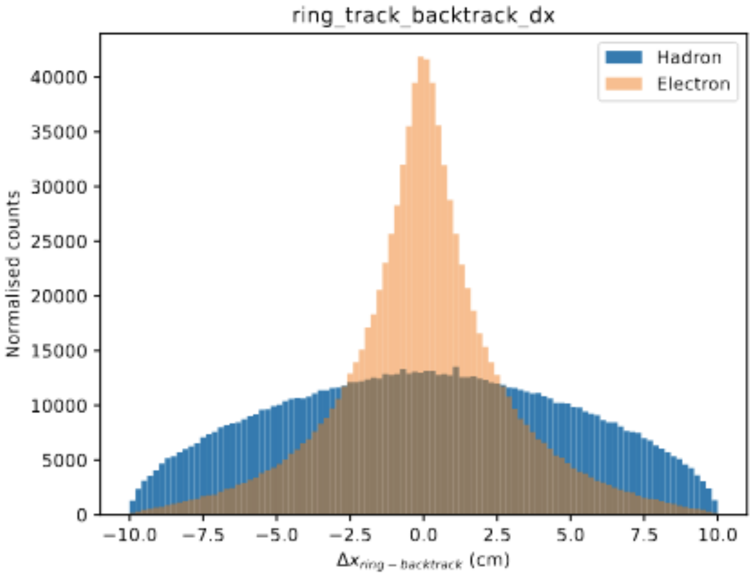


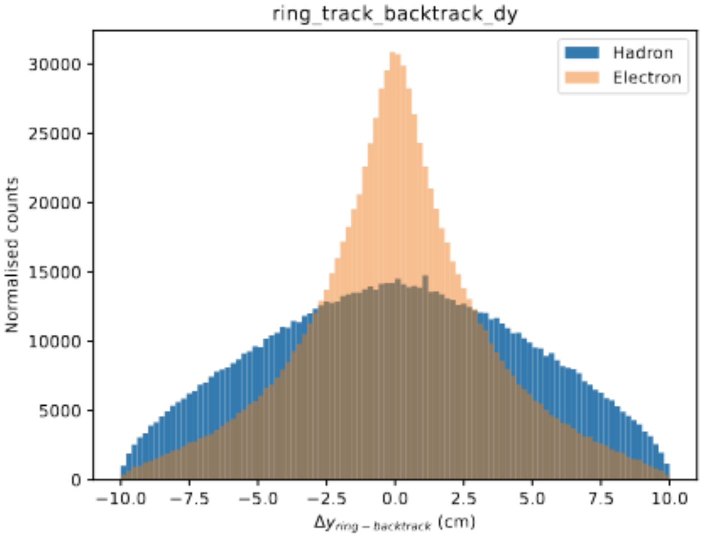




- Some of the hadron tracks are originating from vertex.
- TRD fitted track is extrapolated to the origin assuming proton hypothesis and 0.6 GeV/c as momentum.

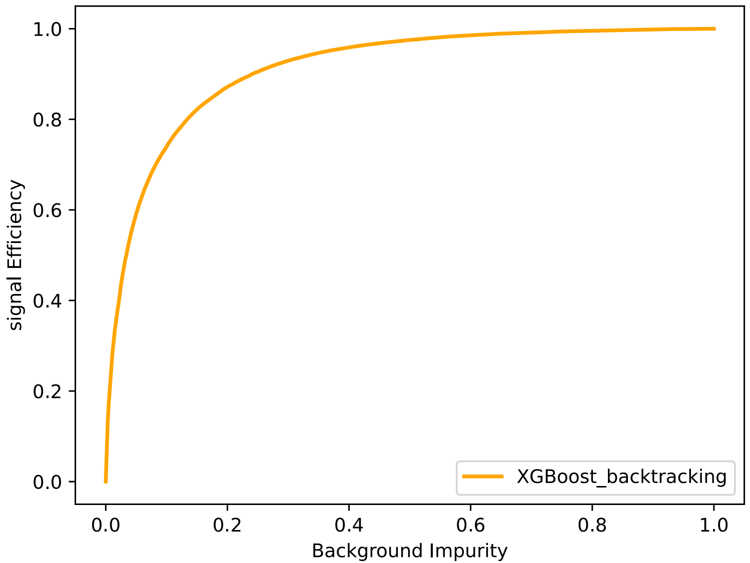
Conversion probability : Parameters



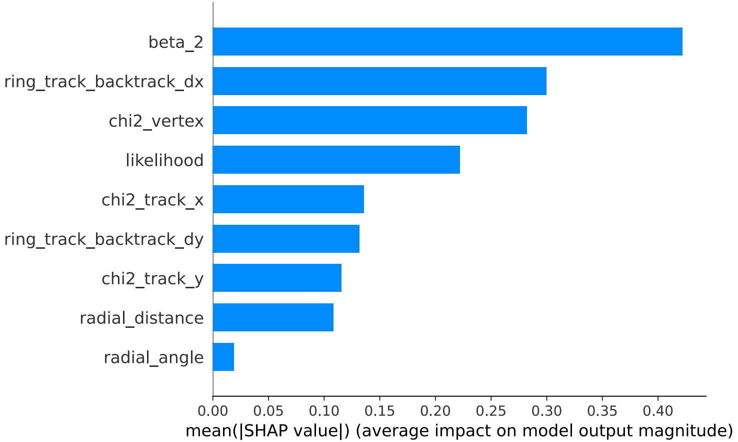


Backtrack - ring matching distance

XGBOOST for deriving conversion probability :

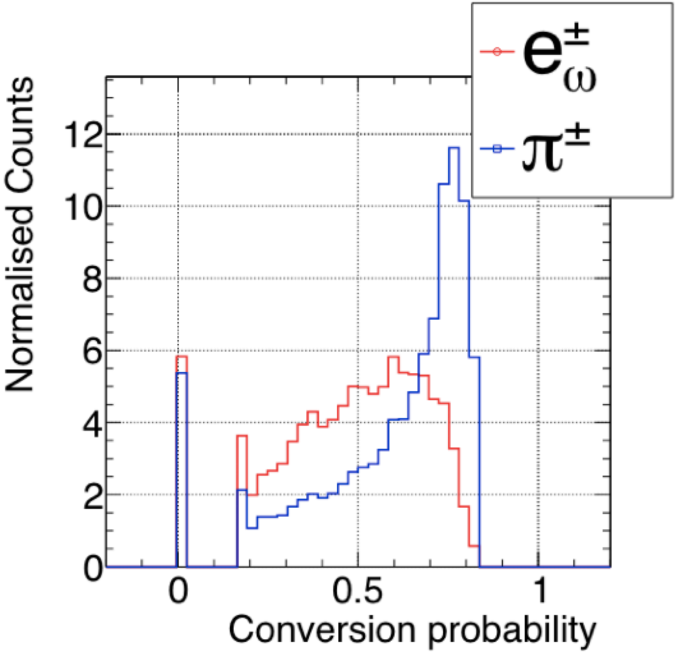


ROC curve for different threshold of the response



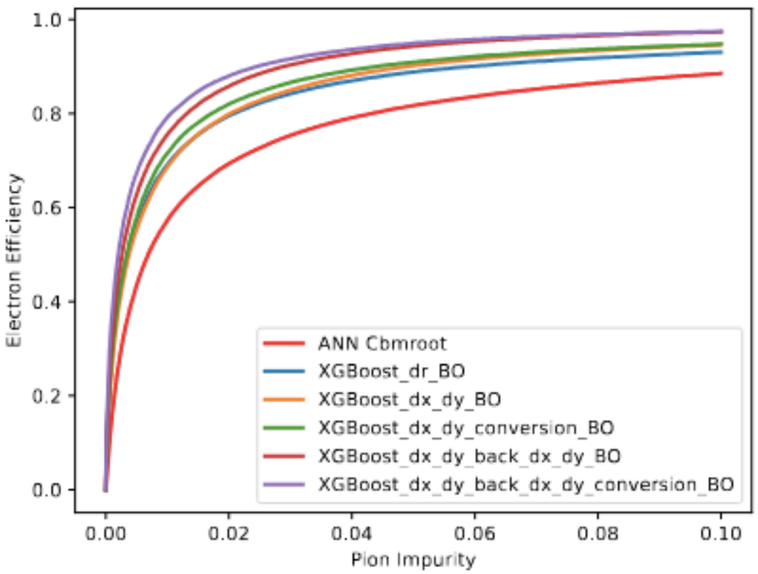
Variable importance (SHAP values)

Inclusion of the conversion probability :

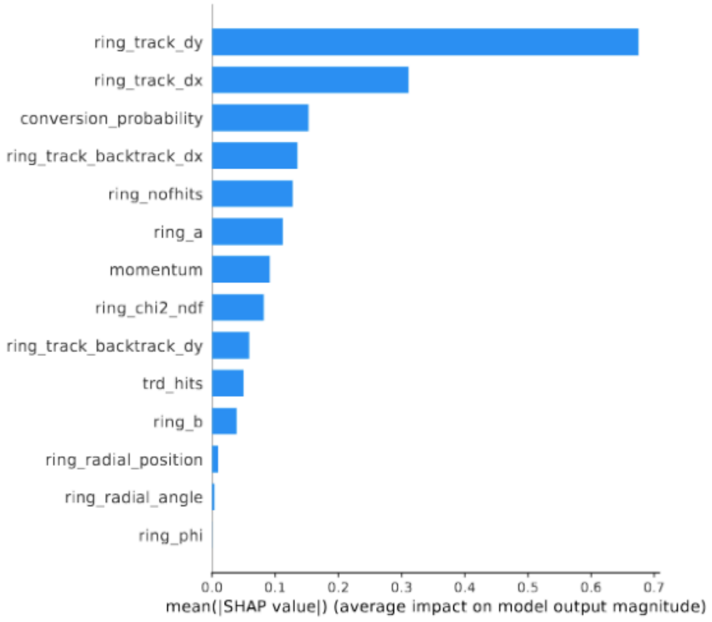


- Each ring is tagged with the probability that it originates from the photon conversion.
- This additional parameter is used to train the forward matching ML model.

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

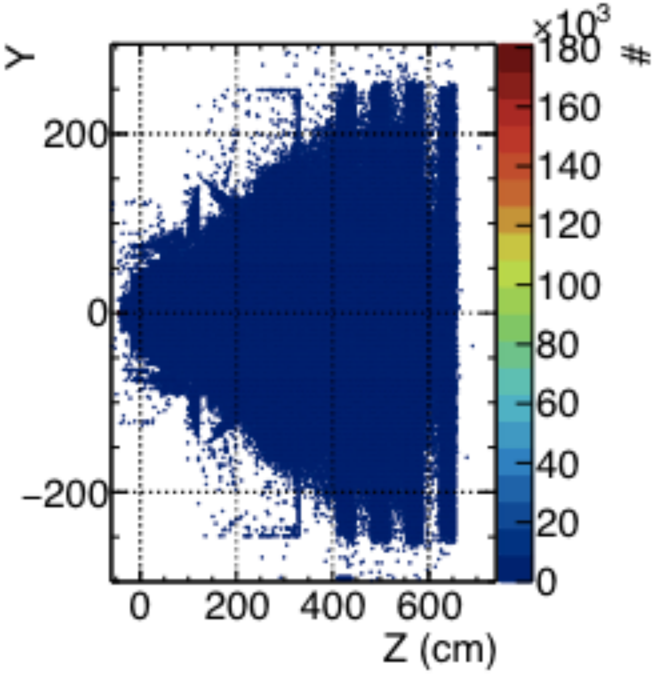
- XGBOOST is used as replacement for ANN for RICH electron ID - Improved identification performance was achieved for the omega electrons.
- Differential ring track distance is added to the model - enhancement of efficiency at higher pion impurity is observed.
- The STS tracks are re-fitted in TRD and 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.

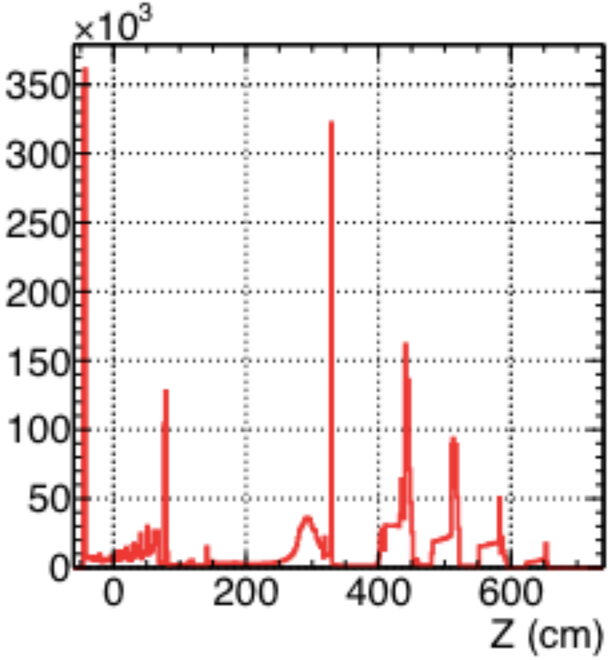
Outlook

- 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.
- Backtracking procedure can be extended till STS : Missing track of conversion electron pair from target could be searched in STS -> reduce combinatorial background.

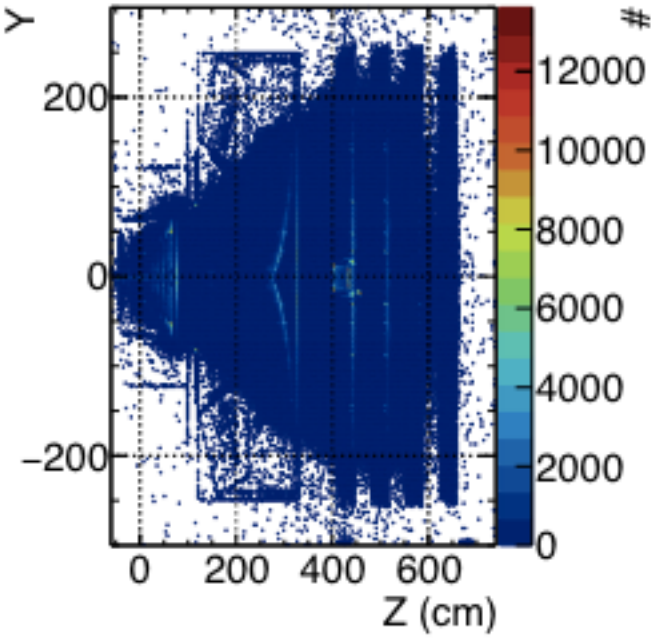
Thank you

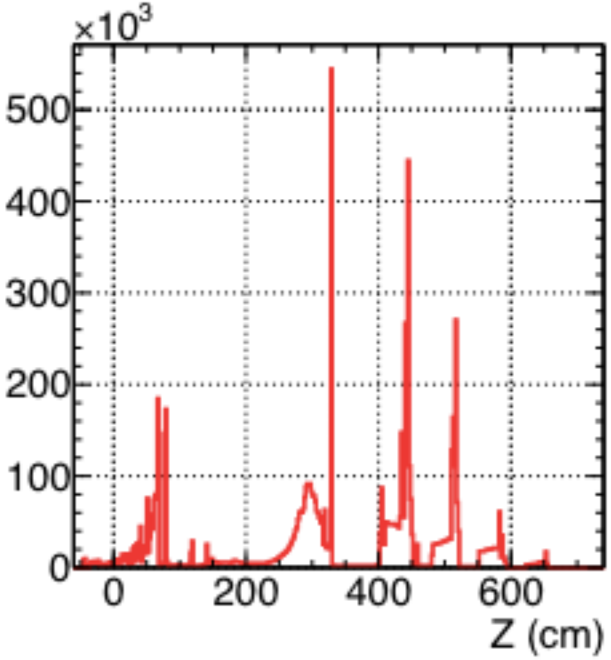
Vertex of fitted tracks protons



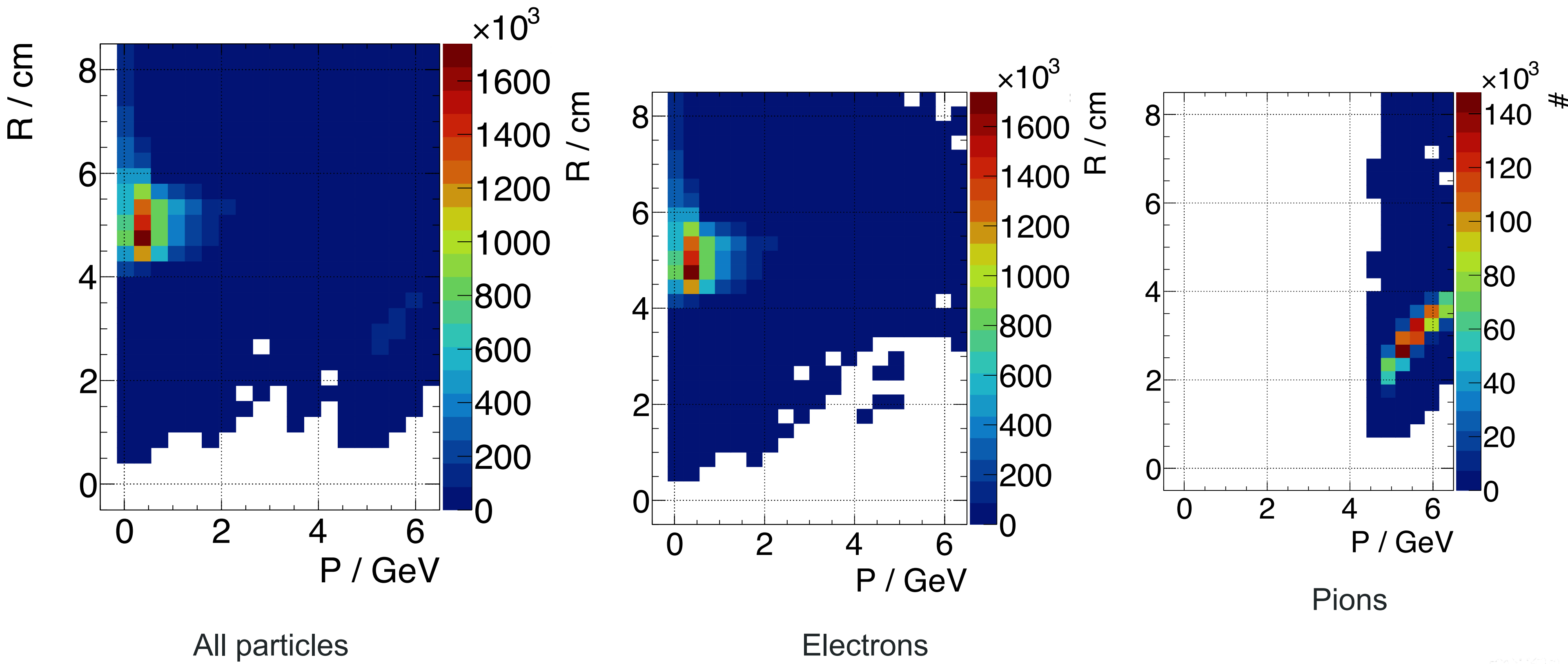


Vertex of fitted tracks electron

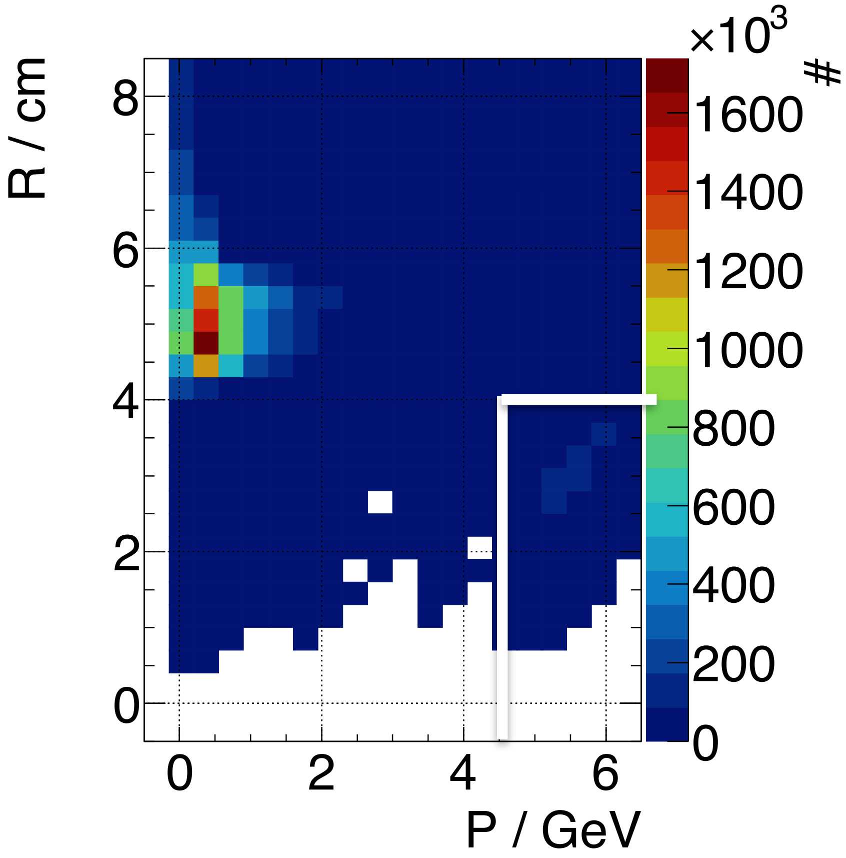




RICH electron ID :



RICH electron ID :



In best match scenario :

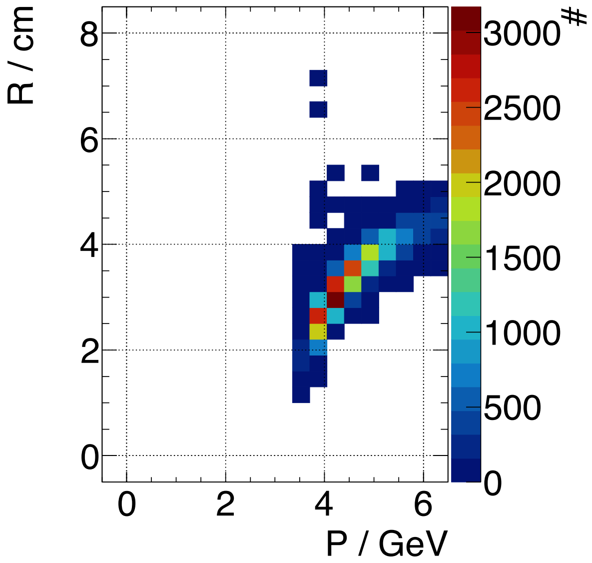
Pion cuts : p > 4.5 GeV / c and R < 4 cm

Electron efficiency ~ 99.9 %

$$\text{Pion suppression} = \frac{\text{Total number of pions}}{\text{Pions mis} - \text{id as electrons}}$$

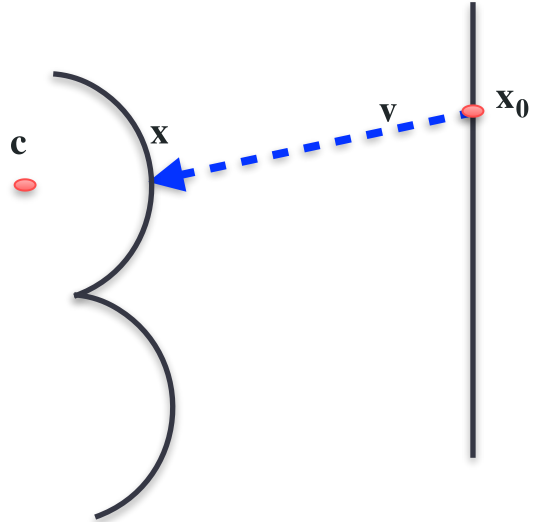
~ 15000 (p > 4.5 GeV)

RICH electron ID :



Muons

Track Extrapolation to mirror :



$$\mathbf{x} = \mathbf{x_0} + d\ \mathbf{v} \quad \text{(Track)}$$

$$||\mathbf{x} - \mathbf{c}||^2 = r_m^2 \quad \text{(Mirror)}$$

$$d = -\mathbf{v}\,.\,(\mathbf{x_0} - \mathbf{c}) - \sqrt{|\mathbf{v}\,.\,(\mathbf{x_0} - \mathbf{c})|^2 - |\mathbf{x_0}|^2 - |\mathbf{c}|^2} \quad \text{(Closest distance)}$$

$$\mathbf{x_{cross}} = \mathbf{x}$$

Track Extrapolation to mirror :

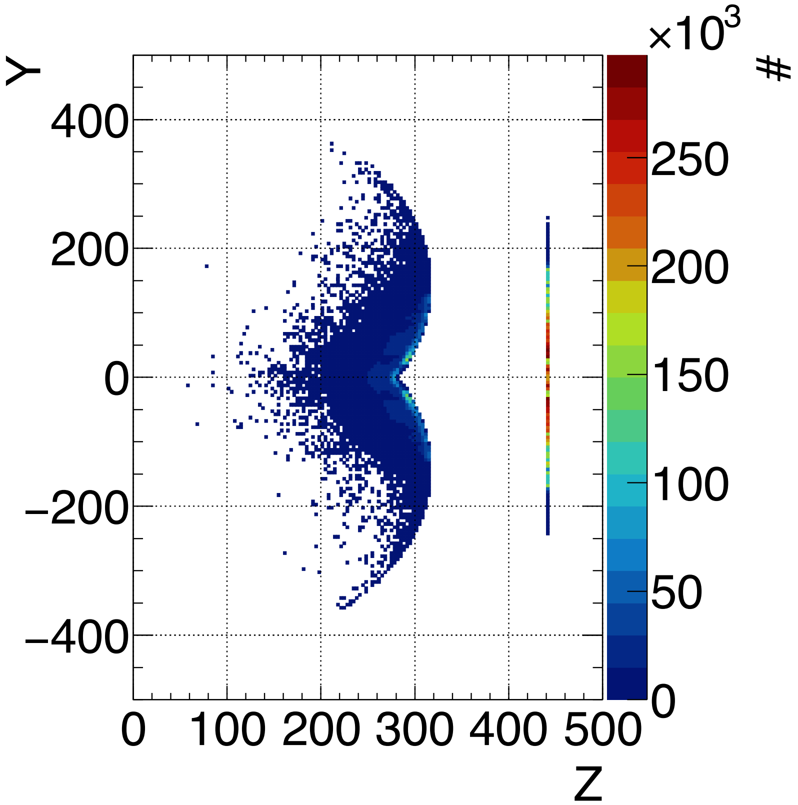
$$\mathbf{x} = \mathbf{x_0} + d\ \mathbf{v}$$ (Track)

$$||\mathbf{x} - \mathbf{c}||^2 = r_m^2$$ (Mirror)

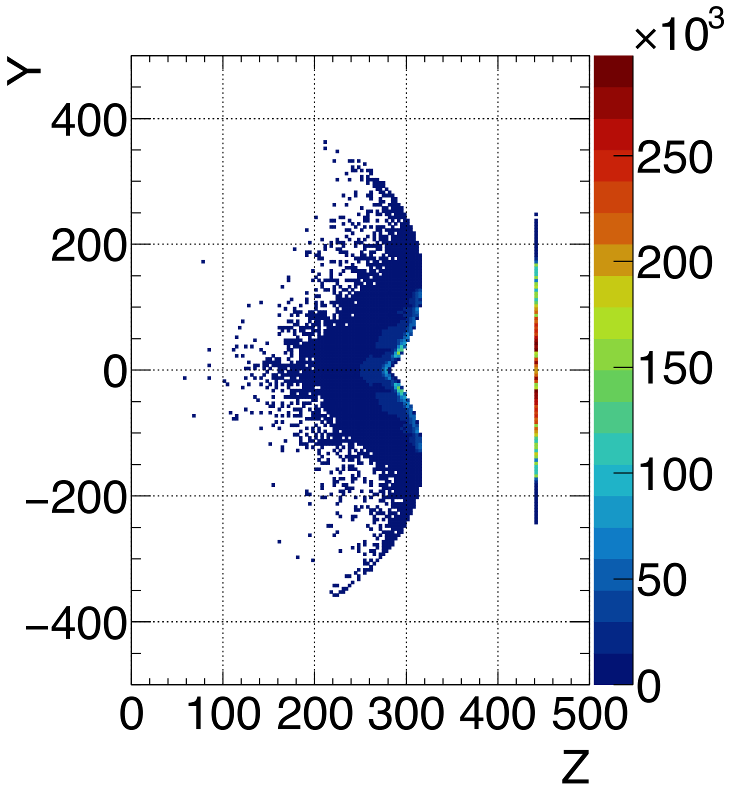
$$d = -\mathbf{v}.(\mathbf{x_0} - \mathbf{c}) - \sqrt{|\mathbf{v}.(\mathbf{x_0} - \mathbf{c})|^2 - |\mathbf{x_0}|^2 - |\mathbf{c}|^2}$$

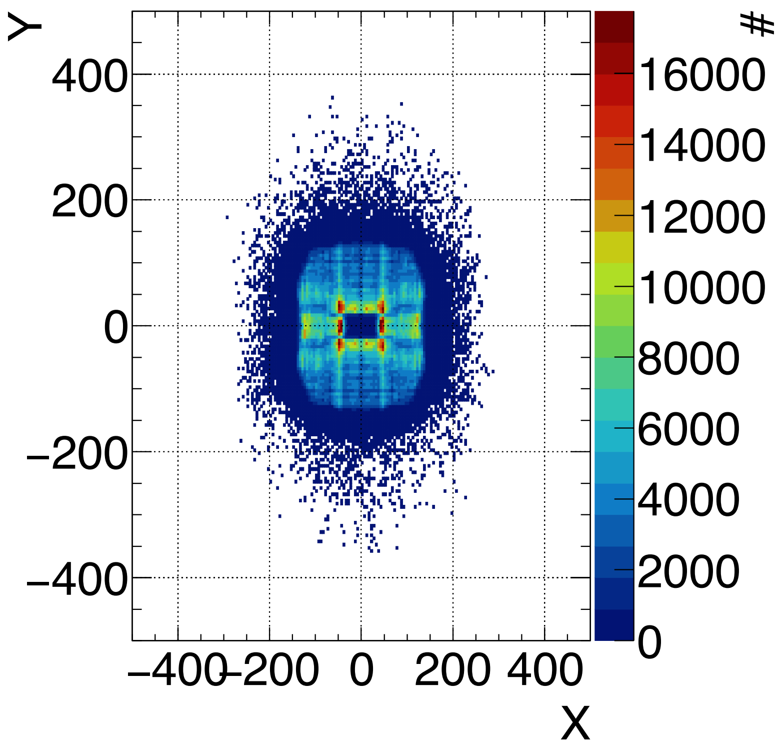
(Closest distance)

$$\mathbf{x_{cross}} = \mathbf{x}$$

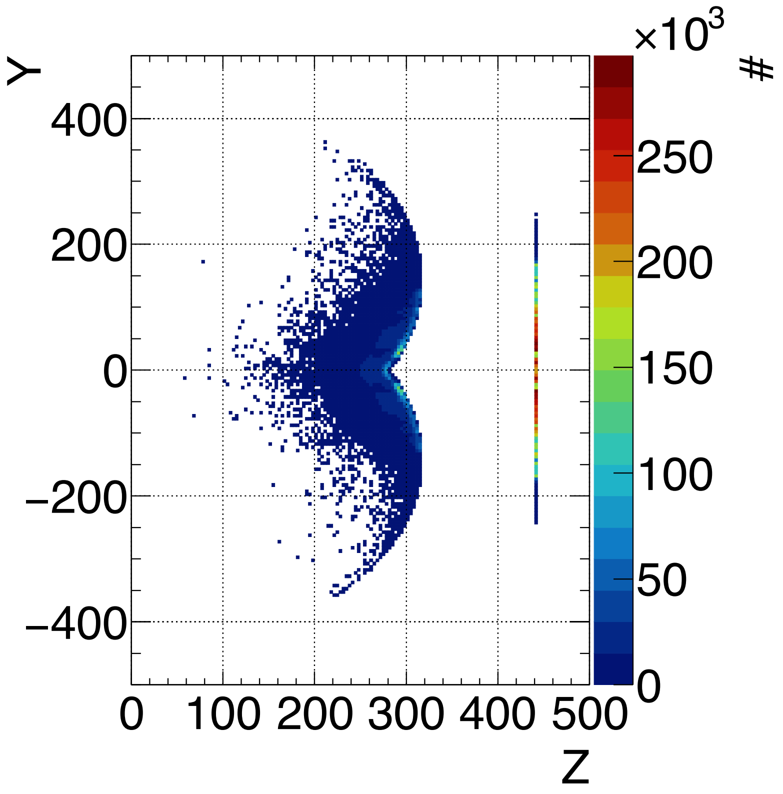


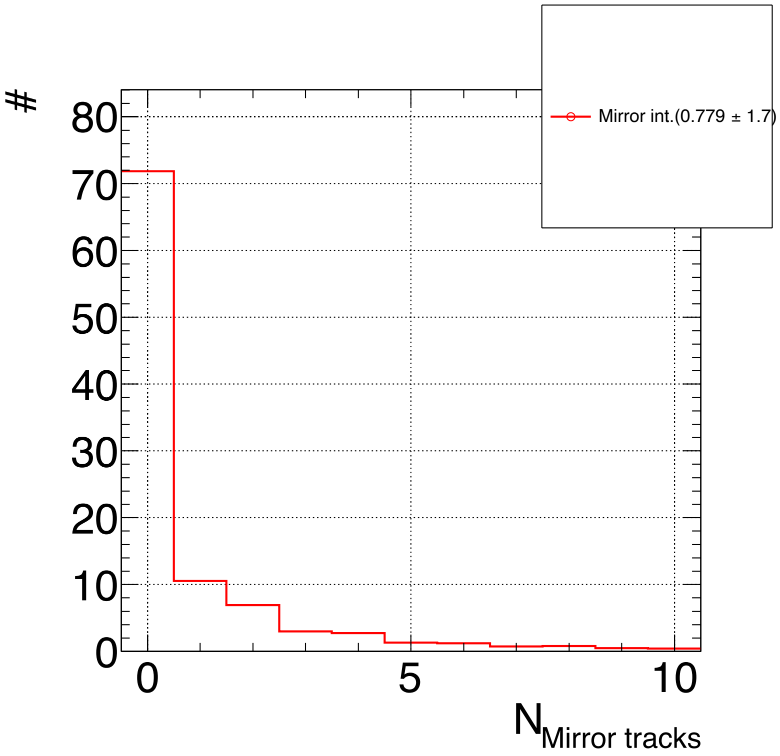
Track Extrapolation to mirror :



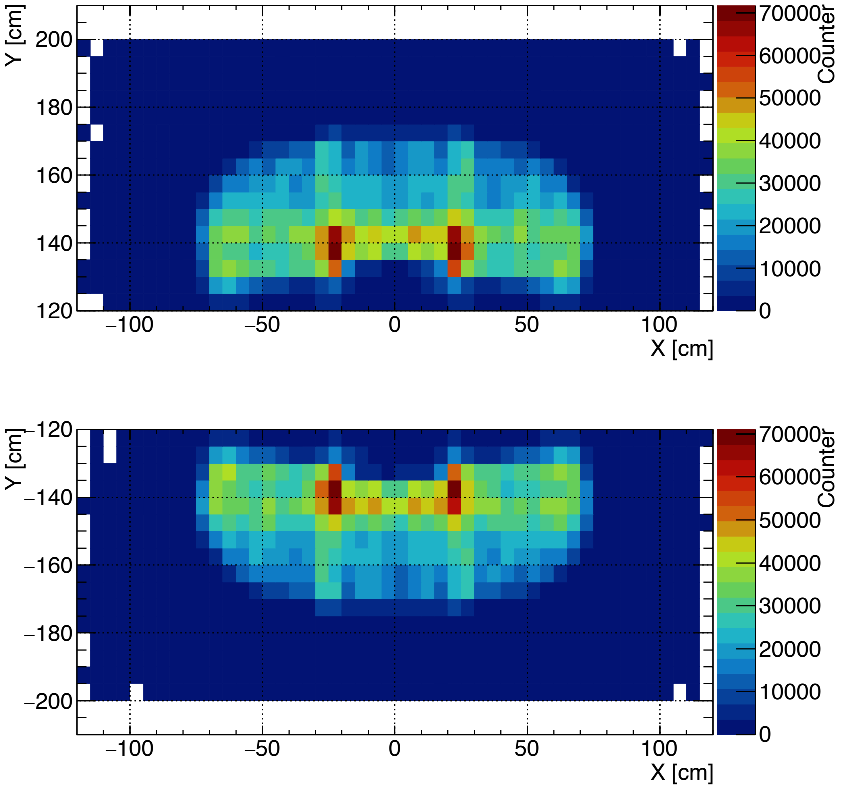


Track Extrapolation to mirror :

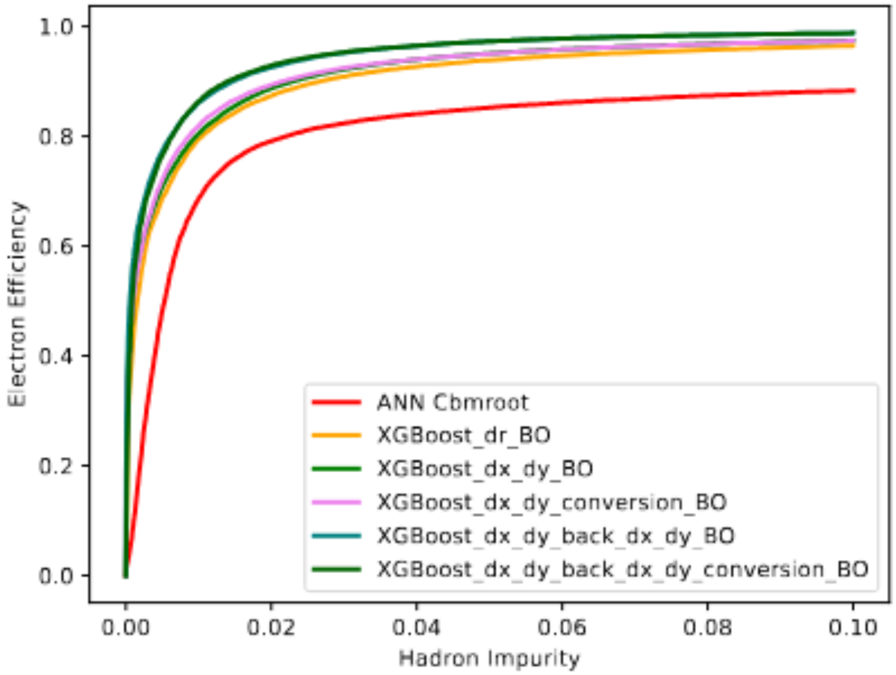




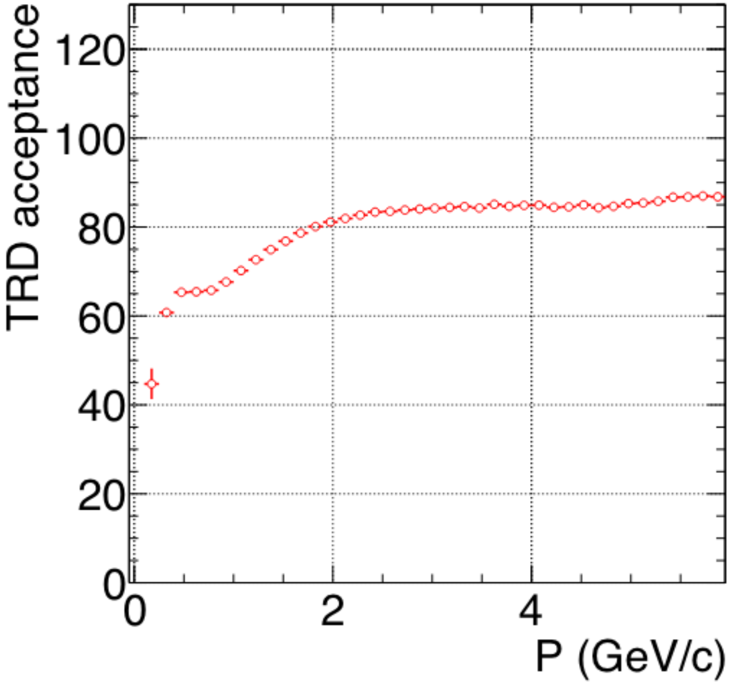
Track Extrapolation to RICH Plane :



Training sample :



TRD acceptance :



TRD acceptance :