Muon/pion separation in the ECAL and μID

Phase II Gaseous Argon TPC group Weekly Meeting
24.07.2023

Francisco Martínez López
f.martinezlopez@qmul.ac.uk

University of Southampton
Queen Mary University of London
Particle Physics STFC Rutherford Appleton Laboratory

Introduction

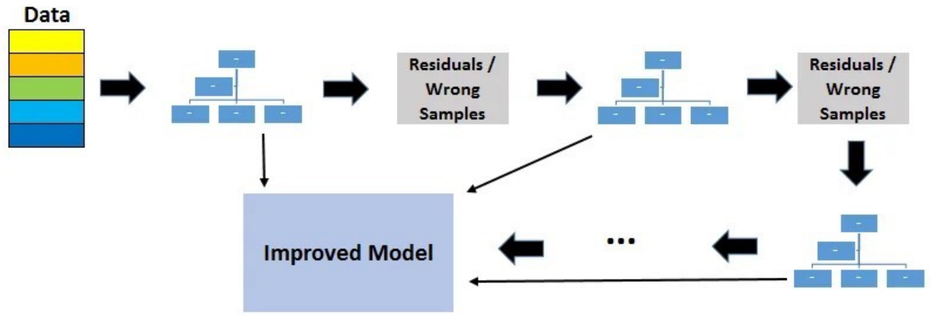
- Muons and pions have almost the same mass and a very similar energy loss per unit length.
  - Not possible to separate them in the TPC for momenta >= 300 MeV.
  - Misidentification of muons and pions leads to wrong estimations of the neutrino energy.
- In the case of ND-GAr the separation can be achieved using its high granularity ECAL and the MuID system embedded in the return yoke of the SPY magnet.
- For this kind of binary classification using a BDT can be a good option, as this are easier to interpret and can handle data without any pre-processing.
  - This is a fairly common problem, so we can also learn from what others did before (e.g. ND280 or ILD).

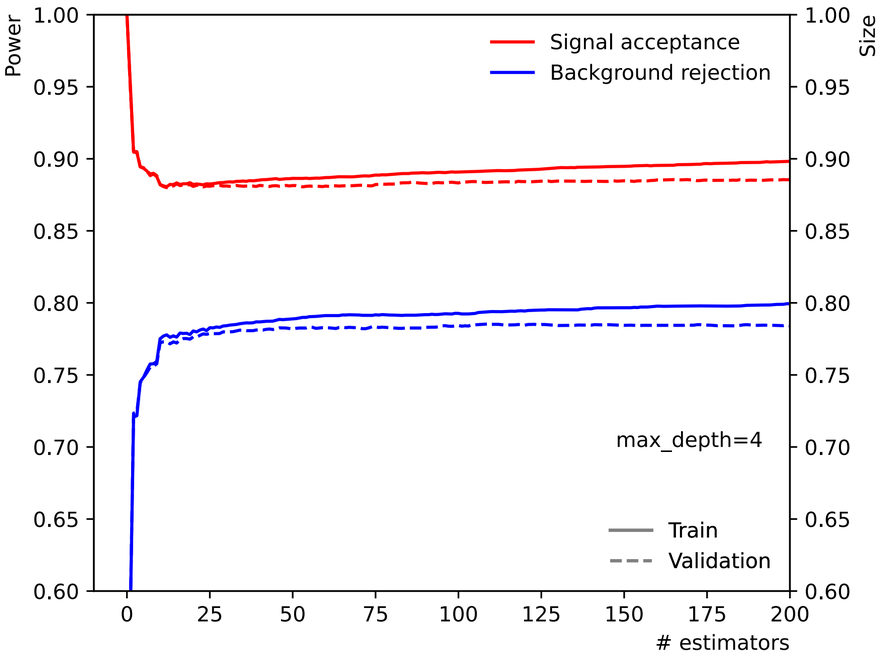
Software changes

- Fixed bug affecting position of MuID hits (PR already merged).
- Extended Track-Cluster association code to work with the MuID clusters.
  - So far it is uses the same logic as for the ECAL clusters, but calling the adequate methods by using the `fInstanceName` label (e.g. `PointInMuIDBarrel` instead of `PointInECALBarrel`).
  - I tuned the association parameters to achieve the optimal performance.
- Modified NN clustering algorithm to allow single hit clusters in the case a hit ends up alone.
  - I noticed this was fairly common for the muon system, where the hit multiplicity is also smaller.
  - We can associate the single hits to tracks so then the ECAL/MuID variables for the BDT can be computed with a more complete set of hits.

BDT setup

- In order to start prototyping the classifier I've used scikit-learn's `GradientBoostingClassifier`.
  - BDTs combine decision tress in series, which are trained from the residuals of the previous stage.
  - Once I have a clear way to approach the problem I'll move to ROOT's TMVA, so it could run within *GArSoft*.
- In the following, the Gradient BDT parameters used are:
  - `loss = 'log_loss'`
  - `n_estimators = 100`
  - `max_depth = 4`

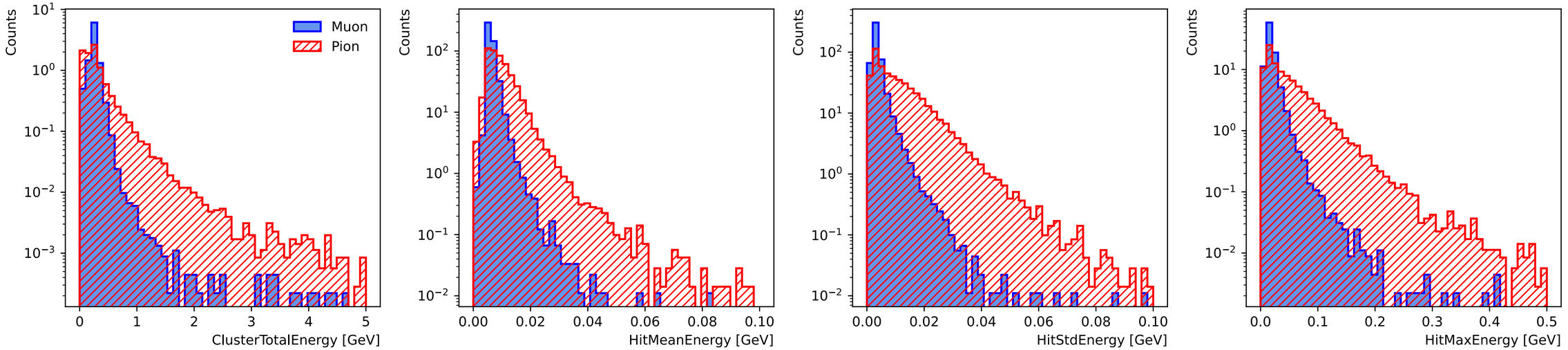




BDT features - ECAL

Energy-related ECAL

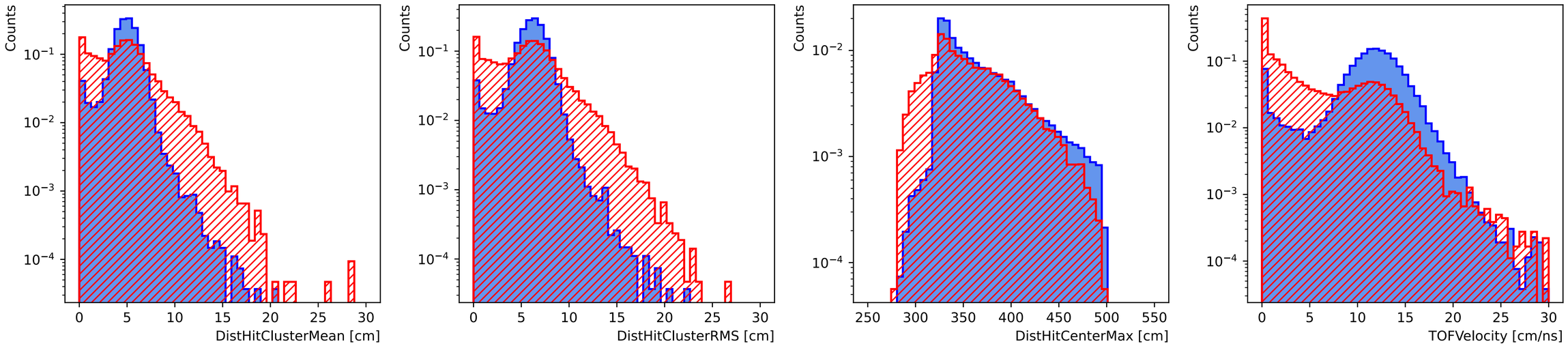
- ECAL total energy (ClusterTotalEnergy): sum of the energy of all the ECAL hits.
- Hit mean energy (HitMeanEnergy): mean of the hits energy distribution.
- Hit std energy (HitStdEnergy): standard deviation of the hits energy distribution.
- Hit max energy (HitMaxEnergy): maximum of the hits energy distribution.



BDT features - ECAL

Geometry-related ECAL

- Mean distance hit-to-cluster (DistHitClusterMean): mean of the distance distribution between the hits and the corresponding cluster's main axis.
- RMS distance hit-to-cluster (DistHitClusterRMS): root mean square of the distance distribution between the hits and the corresponding cluster's main axis.
- Maximum distance hit-to-centre (DistHitCenterMax): maximum of the distance distribution between the hits and the centre of the TPC.
- Time-of-Flight velocity (TOFVelocity): slope obtained when fitting a straight line to the hit time versus hit distance to the centre (i.e. d = v x t).



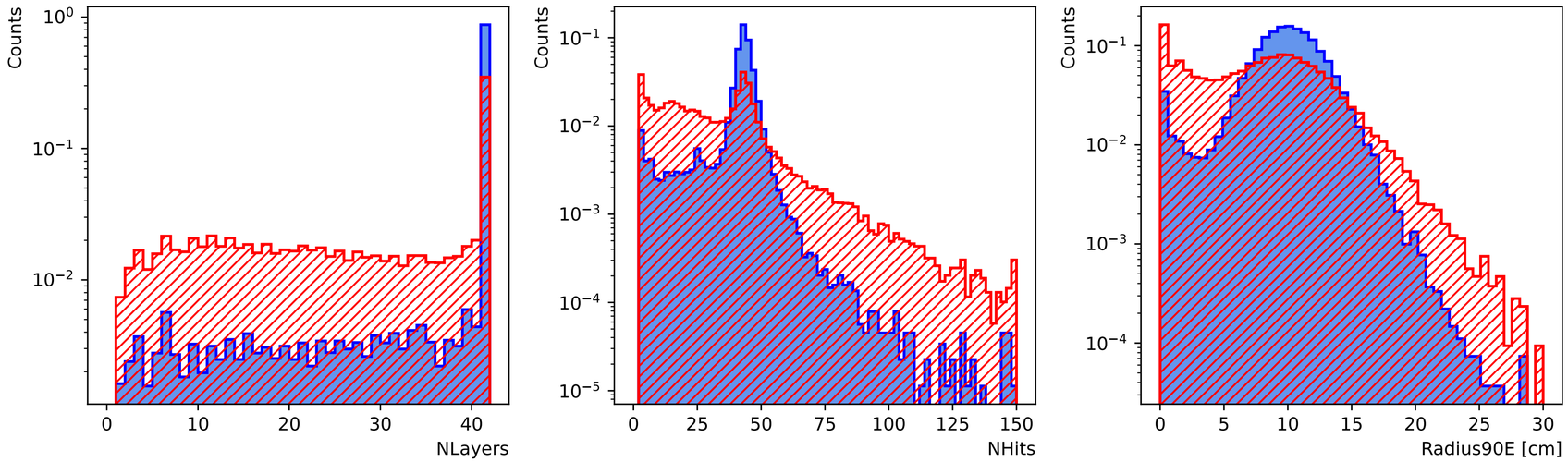
BDT features - ECAL

Statistical ECAL

- Number of hits (NHits): total number of hits associated to the track.
- Number of layers with hits (NLayers): not really a count of all layers with hits but the difference between the last and the first layer with hits.

Energy and Geometry ECAL

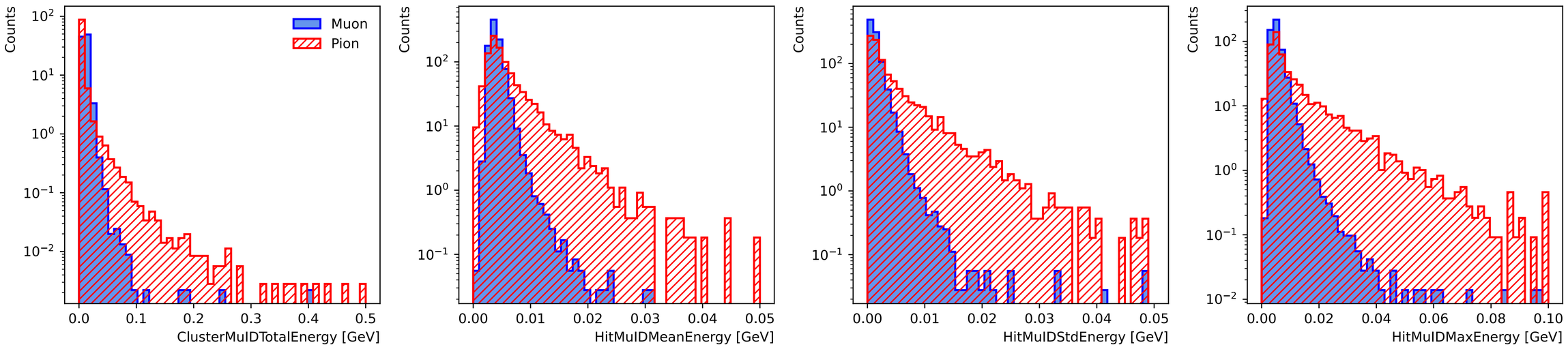
- Radius 90% energy (Radius90E): distance in the hit-to-cluster distribution for which 90% of the total energy is contained in the hits that are closer to the axis (i.e. radius that contains 90% of the energy).



BDT features - MuID

Energy-related MuID

- MuID total energy (ClusterMuIDTotalEnergy): sum of the energy of all the MuID hits.
- Hit MuID mean energy (HitMuIDMeanEnergy): mean of the MuID hits energy distribution.
- Hit MuID std energy (HitMuIDStdEnergy): standard deviation of the MuID hits energy distribution.
- Hit MuID max energy (HitMuIDMaxEnergy): maximum of the MuID hits energy distribution.



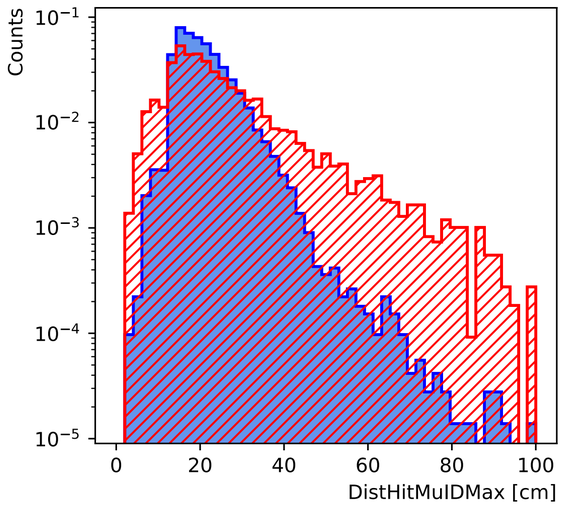
BDT features - MuID

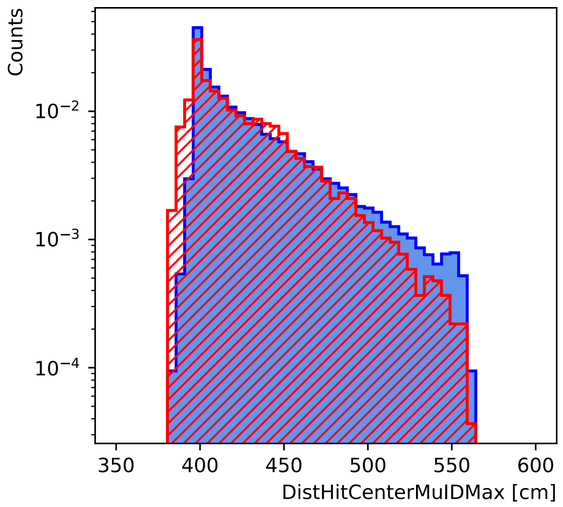
Geometry-related MuID

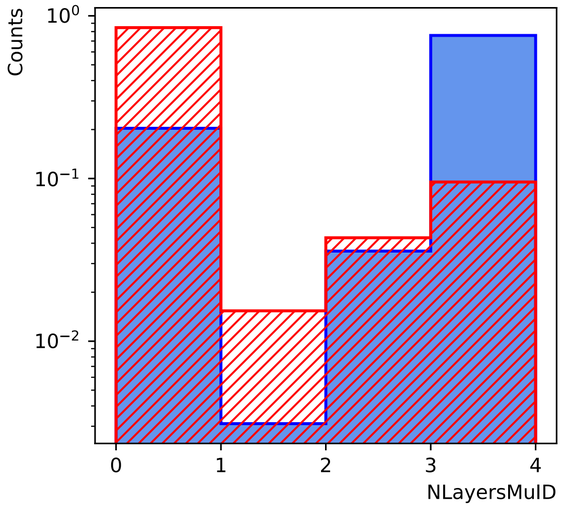
- Max distance MuID hits (DistHitMuIDMax): maximum distance between pairs of MuID hits.
- Maximum distance MuID hit-to-centre (DistHitCenterMuIDMax): maximum of the distance distribution between the MuID hits and the centre of the TPC.

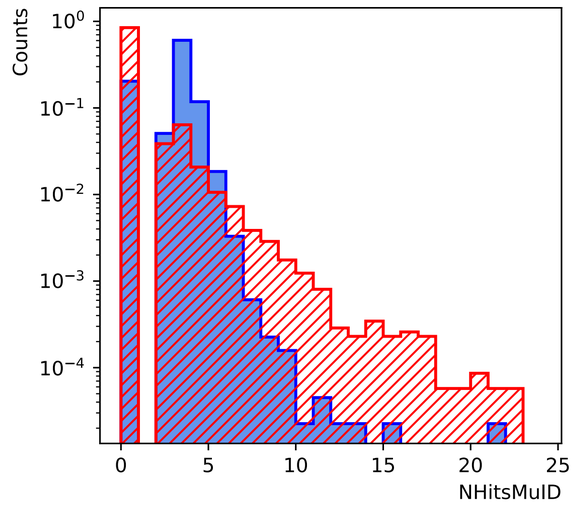
Statistical MuID

- Number of hits (NHitsMuID): total number of MuID hits associated to the track.



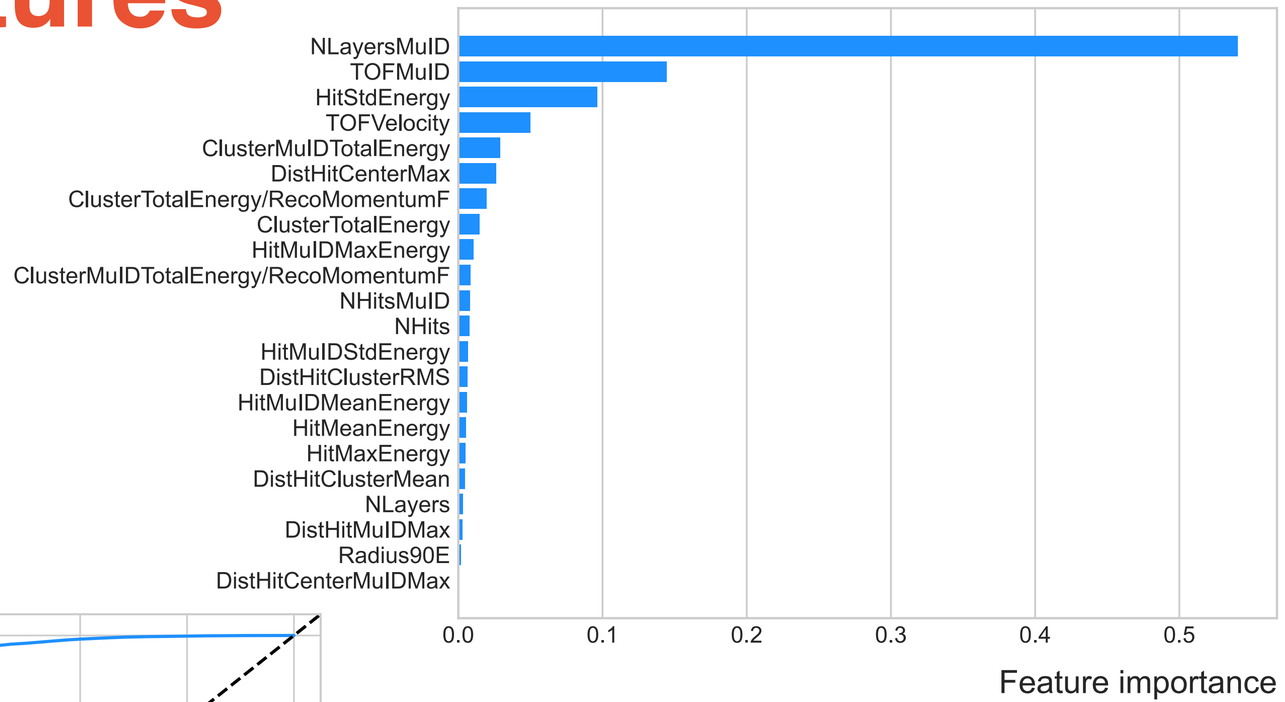


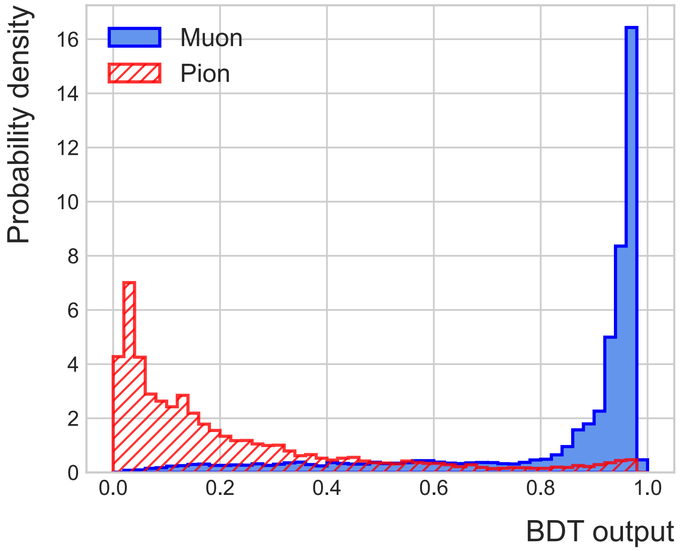


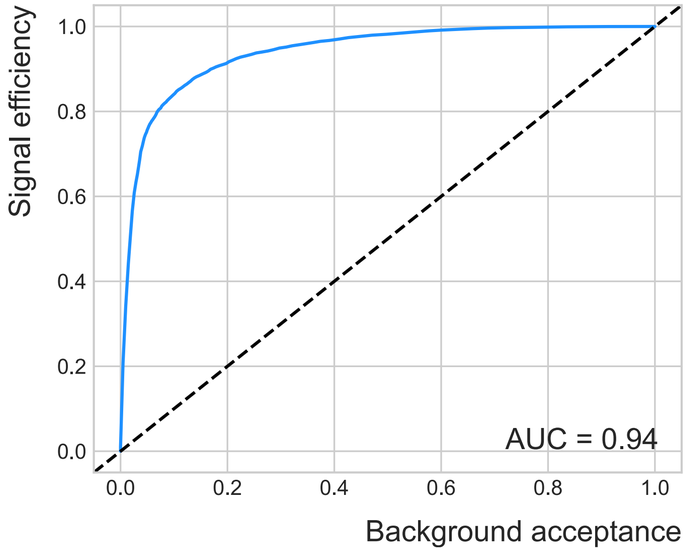


BDT output - All features

- I used a 100k FHC neutrino interaction sample, training the BDT with the ECAL and MuID variables.
  - Only requirement is having at least 1 ECAL hit associated to the track.



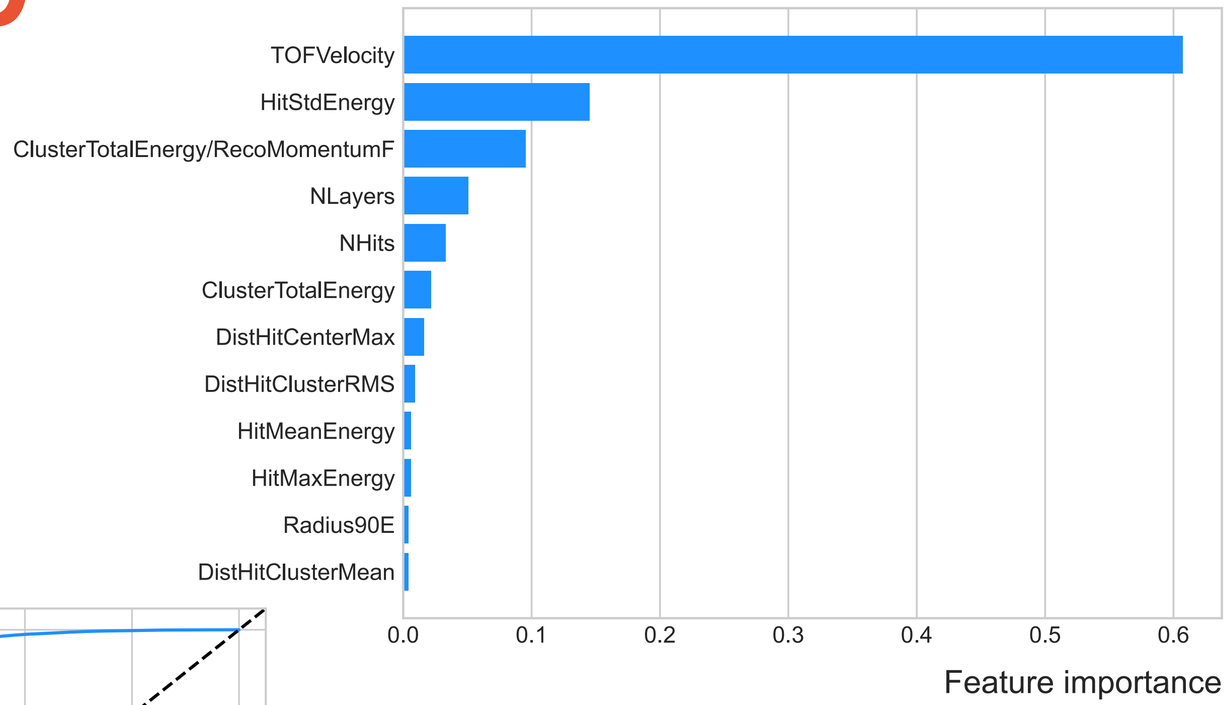


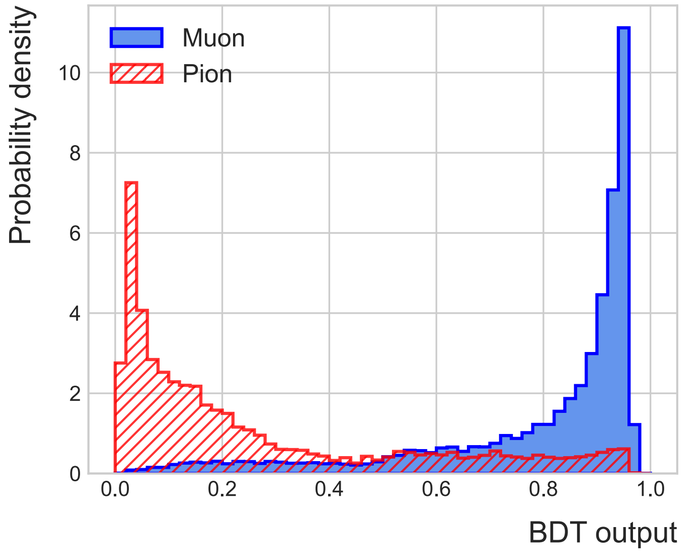


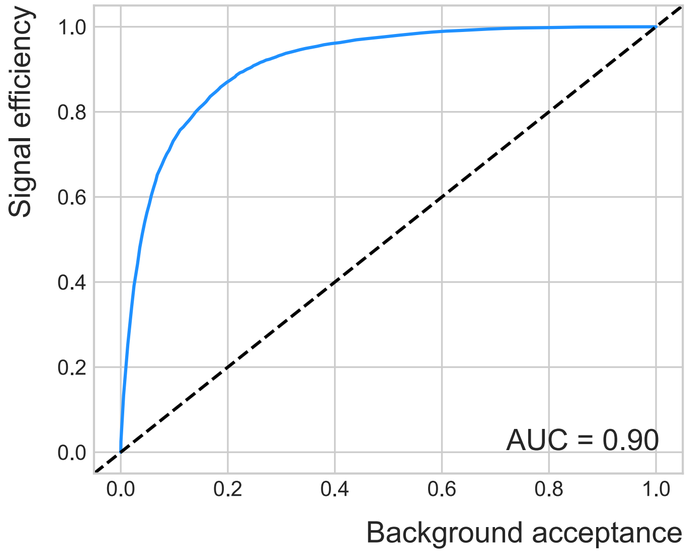

```
Default  detection threshold: 0.5

Signal efficiency:     86.98%

Background rejection:  87.07%
```

BDT output - No MuID

- I repeated the training with the full sample of muons and pions, now only using the ECAL variables.



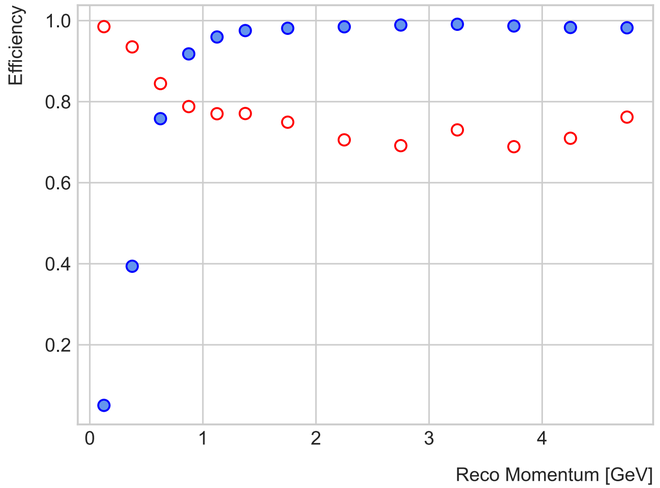


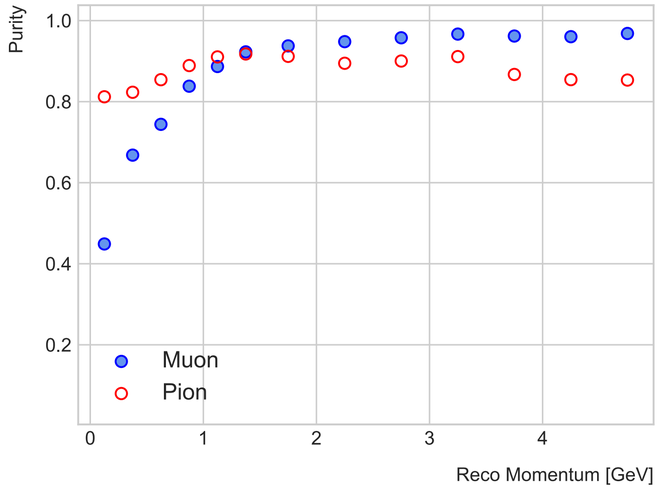



```
Default  detection threshold: 0.5

Signal efficiency:     88.47%

Background rejection:  78.55%
```

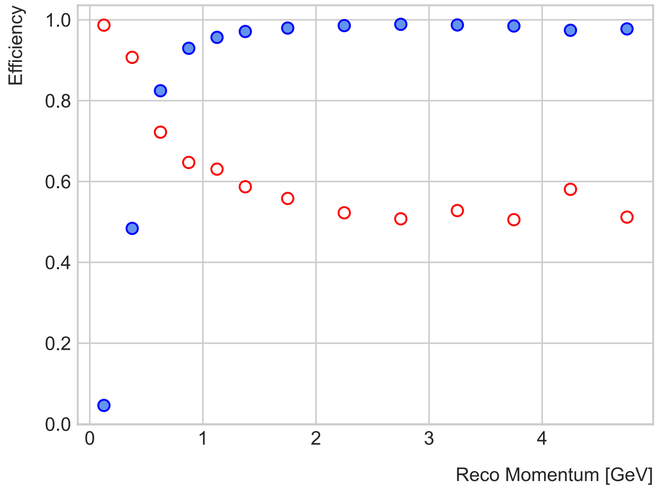
BDT performance - Features used

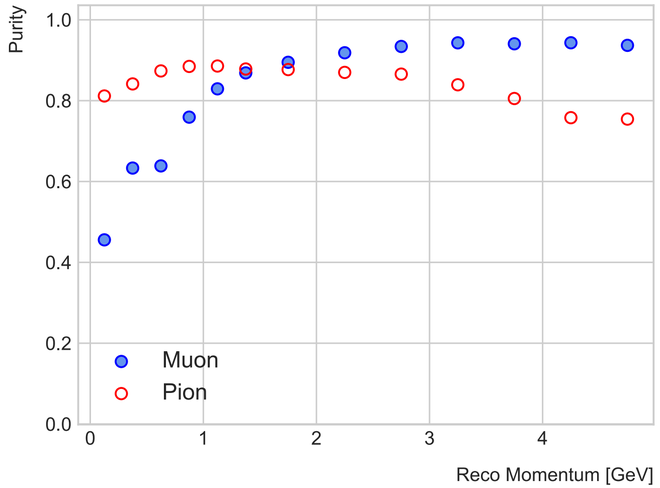
ECAL+MuID





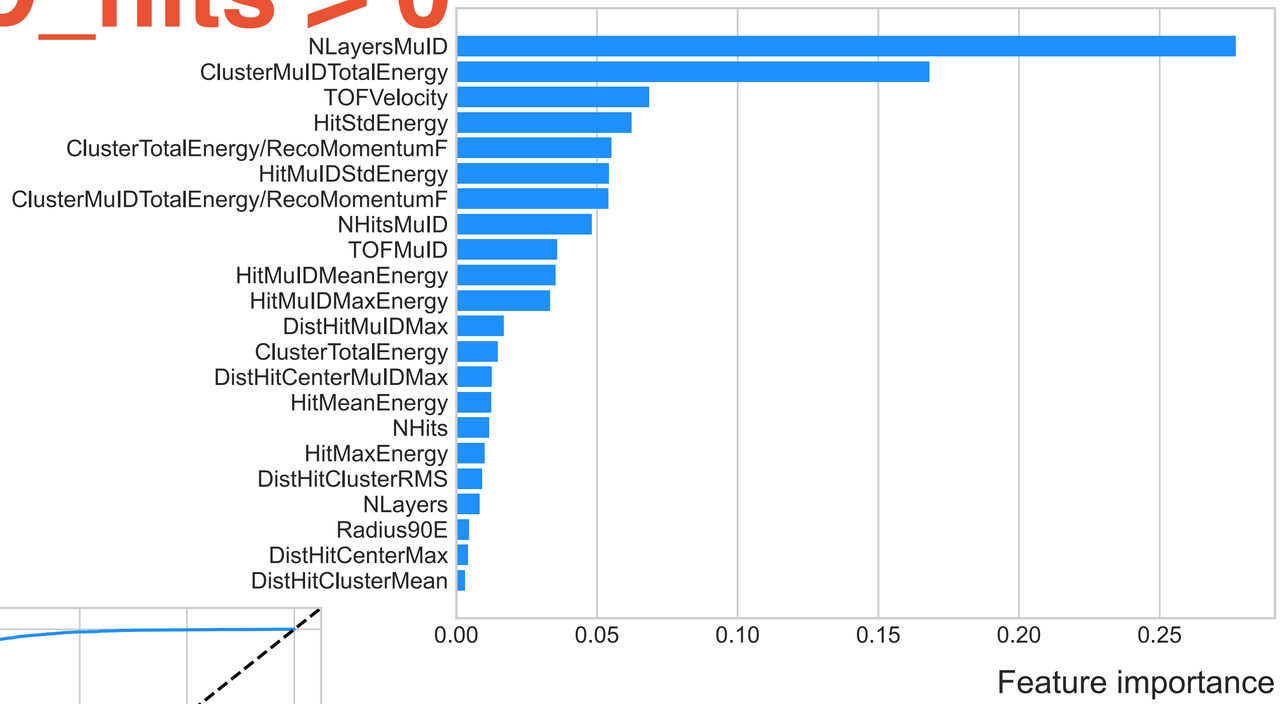
ECAL only

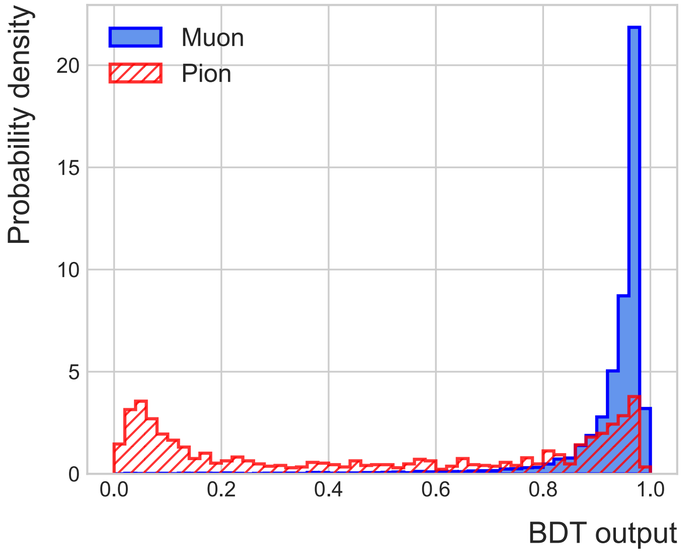


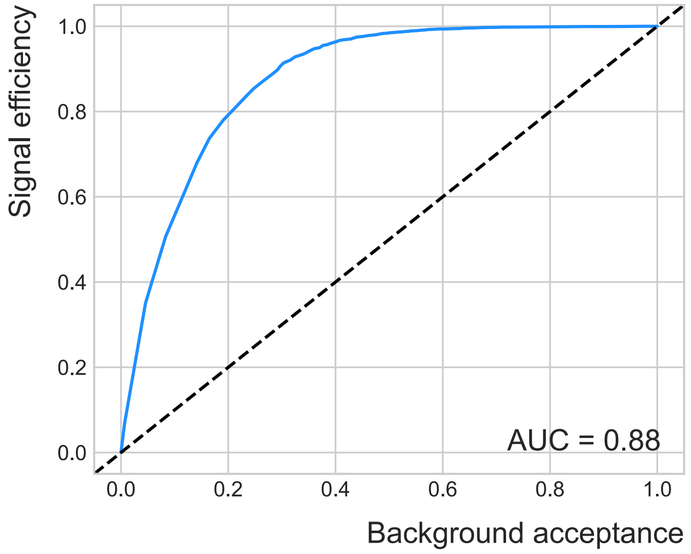


BDT output - n_MuID_hits > 0

- As a check, I trained and tested the BDT using only events with hits in the muon system.
  - That is about 80% of muons but only 15% of pions.



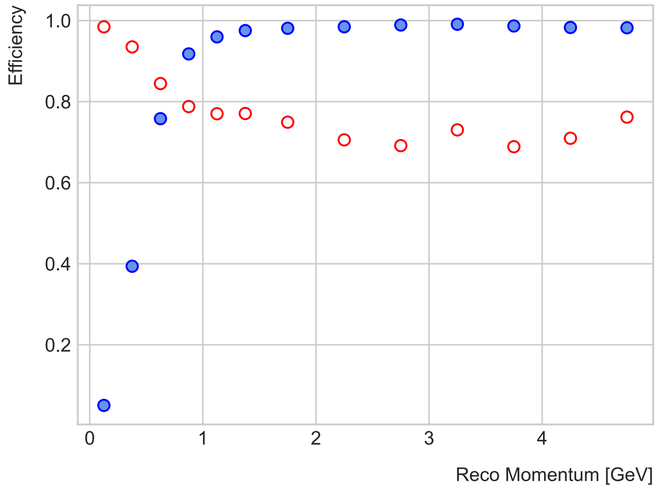


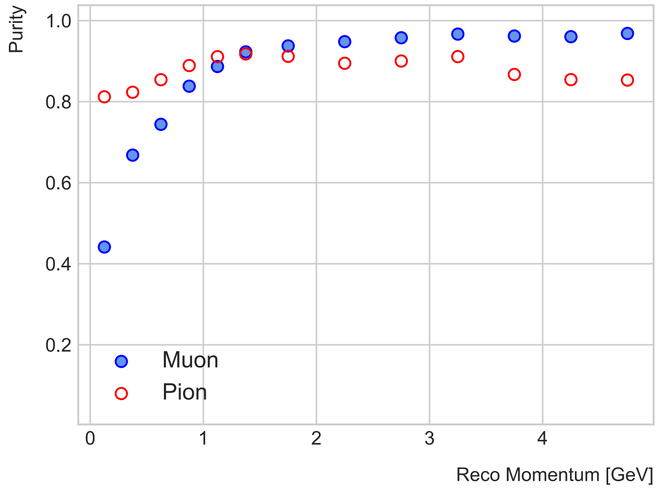



```
Default  detection threshold: 0.5

Signal efficiency:     98.35%

Background rejection:  50.86%
```

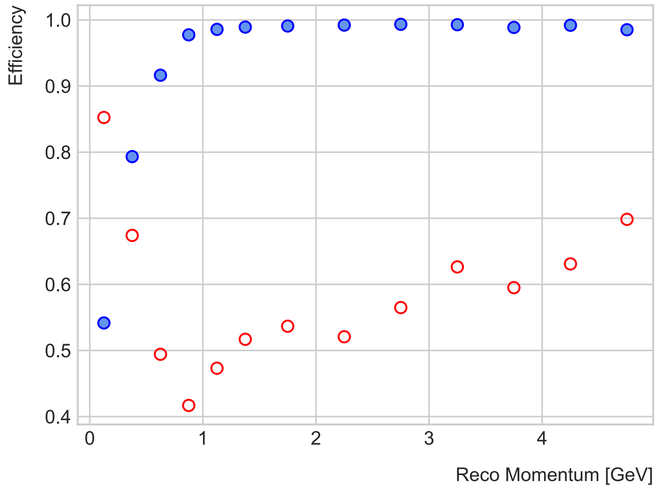
BDT performance - Samples used

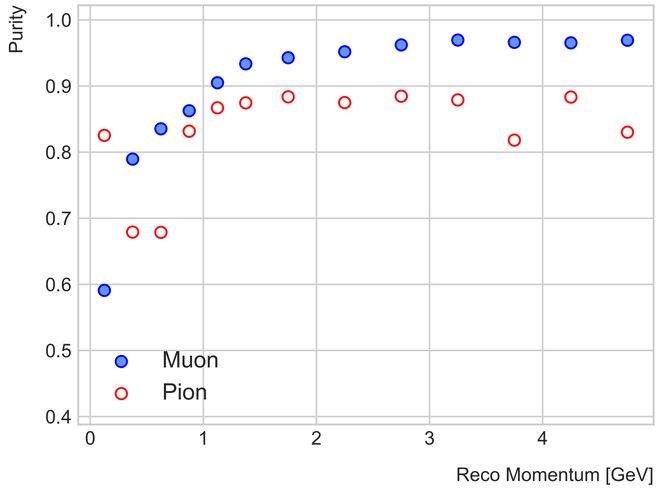
All events





n_MuID_hits > 0





Conclusions

- Separating muons from pions is challenging, using the TPC information is not possible for a good part of the momentum range of interest.
- Using the ECAL and MuID systems we can achieve a reasonably good separation.
  - How good is good enough?
- The muon system proved to be especially important for high energy pions, with some residual improvement for muons too.
- All training and testing have been done with FHC samples.
  - Maybe worth trying to train the BDT with single particle events?

Backup slides

Single hit cluster example

1000 single pion sample	w/o single hist clusters	w/ single hit clusters
ECAL clusters	3481	8773
ECAL clusters w/ assns	1404	1992
MuID clusters	295	1321
MuID clusters w/ assns	136	416