Cluster Reconstruction Studies [Updated]

Victor Genty

Nevis Labs

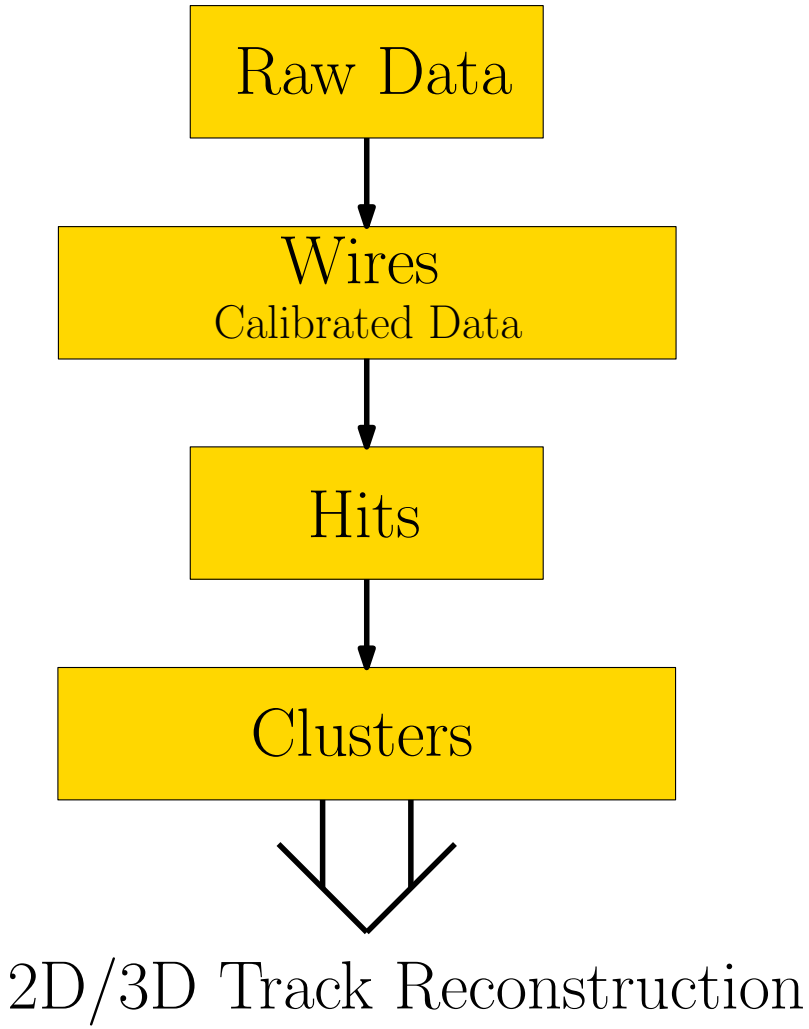
June 28, 2013

Outline

1. Clustering in LArSoft
2. Fuzzy Clustering
3. Purity and Efficiency Evaluation
   - Single Electrons
   - Single Muons
   - $1e^- + 1p$ Final states
   - Future

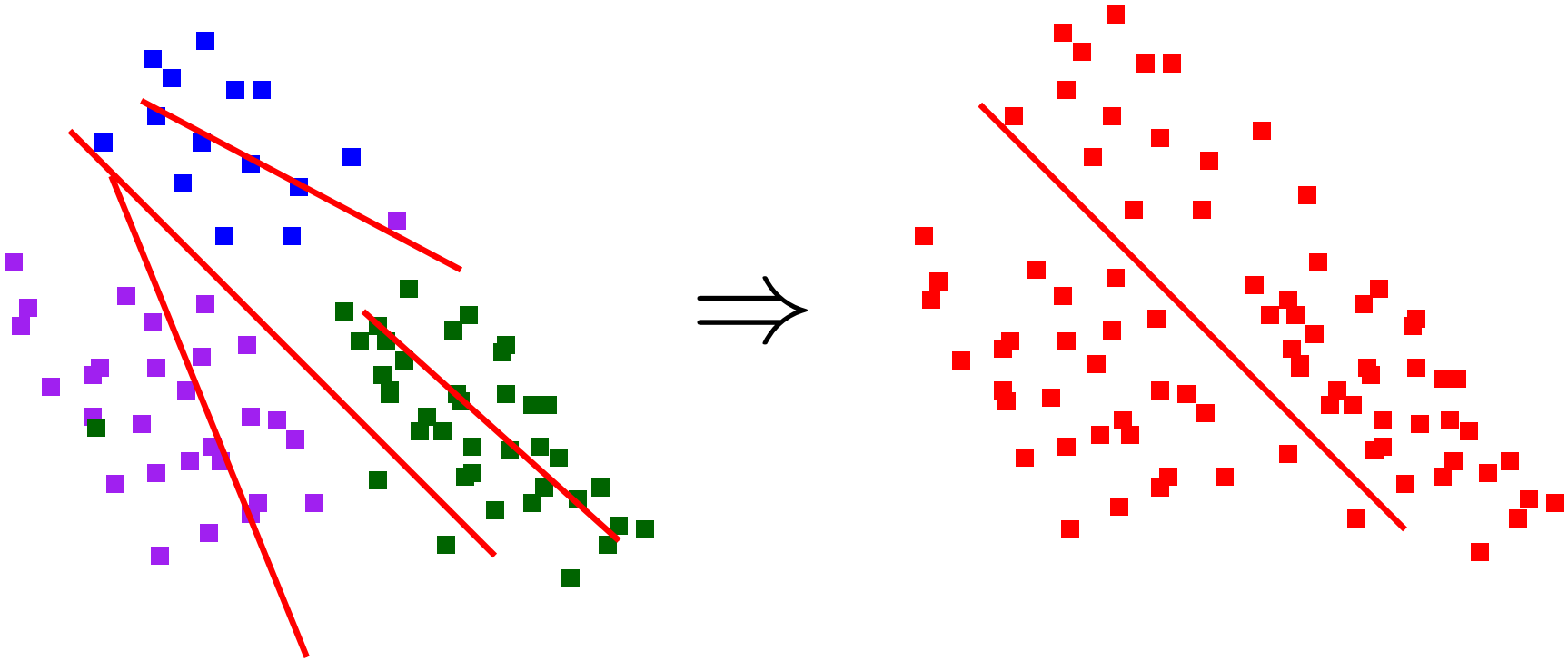
Clustering in LArSoft

- **Clustering** algorithms identify reconstructed wire hits which are correlated both spatially and temporally
- **Hits** are signal vs time information from a calibrated Wire object and looks for peaks that indicate real energy deposition occurred
- **DBscan** and **Fuzzy Clustering** are two such algorithms



Fuzzy Clustering

- Developed by Ben Carls at Fermilab
- Algorithm assigns "degrees of belonging" to reconstructed hits, instead of "definite belonging" as in DBscan $\rightarrow$ fed through an optimization criterion $\rightarrow$ clusters merged if necessary [1].
- Promising algorithm providing 2D physics-based clusters that track algorithms can utilize



My Evaluation I

1. Generate single electron, muon and uniform flux CC ν_e events with singles.fcl and GENIE. Filter for $1e^- + 1p$ final states
2. Reconstruct clusters with modified uboone offline .fcl script
3. Feed to a module I created to calculate purity and efficiency of reconstructed clusters
4. Compare DBscan, FuzzyCluster

$$\text{Purity} = \frac{\#\text{ of hits from trackID in cluster}}{\text{total }\#\text{ of hits in cluster}}$$

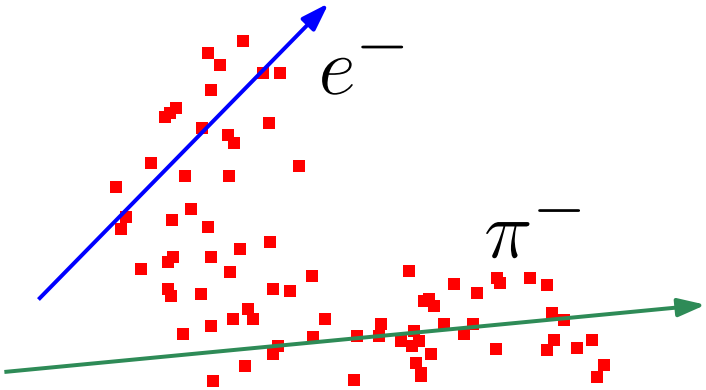
$$\text{Efficiency} = \frac{\#\text{ of hits from trackID in cluster}}{\text{total }\#\text{ of hits for that trackID}}$$

Purity

Formula

$$\text{Purity} = \frac{\#\text{ of hits from trackID in cluster}}{\text{total }\#\text{ of hits in cluster}}$$

Example



Hit Count

Recon: Total = 50

Truth: e^- = 15 $\rightarrow$ Purity=15/50=0.3

Truth: π^- = 35 $\rightarrow$ Purity=35/50=0.7

Measures

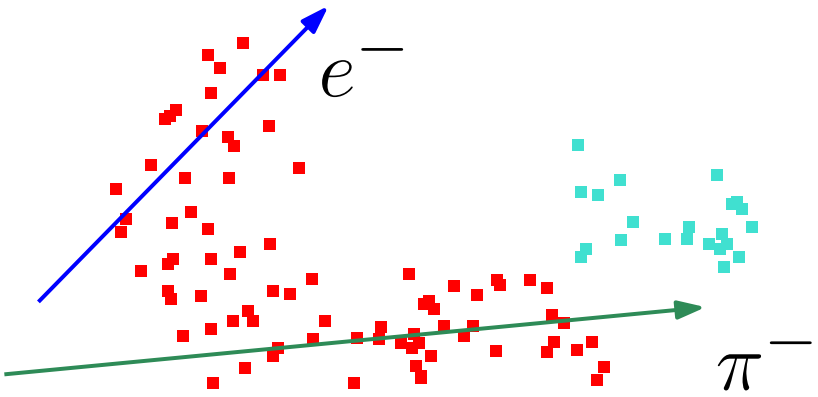
- How much of a cluster is composed of a each true particle
- If less than 1: either true tracks were on top of each other or clustering algorithm failed

Efficiency

Formula

$$\text{Efficiency} = \frac{\#\text{ of hits from trackID in cluster}}{\text{total }\#\text{ of hits for that trackID}}$$

Example



Track Hit Count

Truth: Total $\pi^- = 100$

Truth: Cluster 1 $= 75 \rightarrow$ Eff=0.75

Truth: Cluster 2 $= 25 \rightarrow$ Eff=0.25

Recon: Grouping of hits

Measures

- How many of all hits the particle generated are in a specific cluster
- If less than 1: Algorithm failed to group the hits created by the particle into a single cluster

My Evaluation II

Parameters

- 100 single electron events at 1.5 GeV
- 100 single muon events at 1.5 GeV
- 5000 CC events with $1e^- + 1p$ final state

Cuts

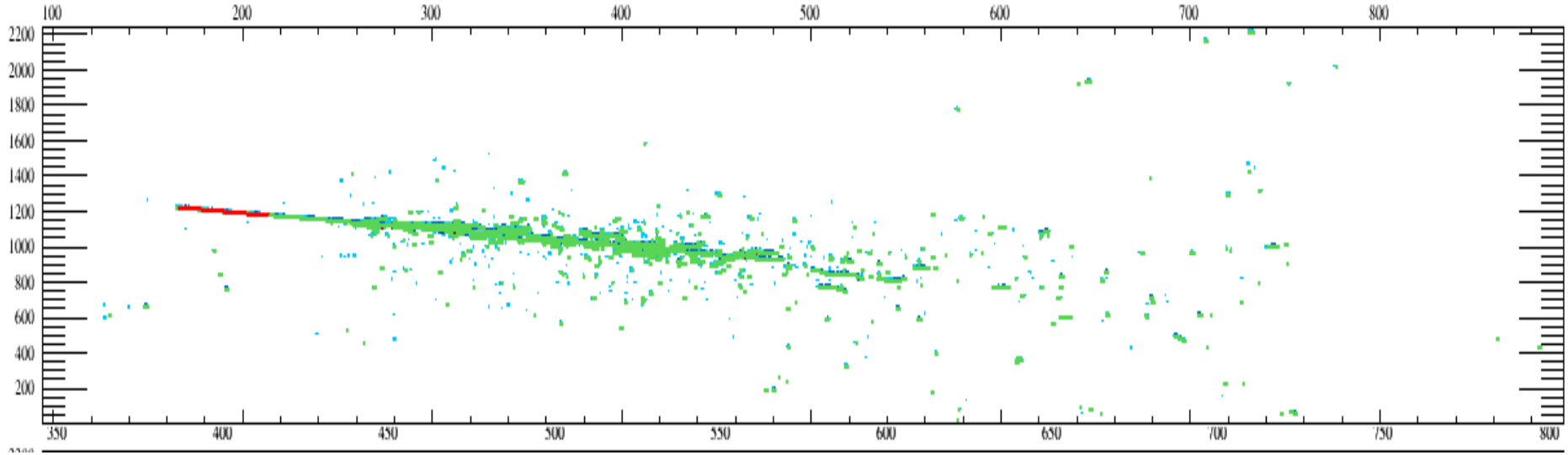
- Cut on efficiencies greater than 0.1 to remove clusters which contain low hit count.

Organization

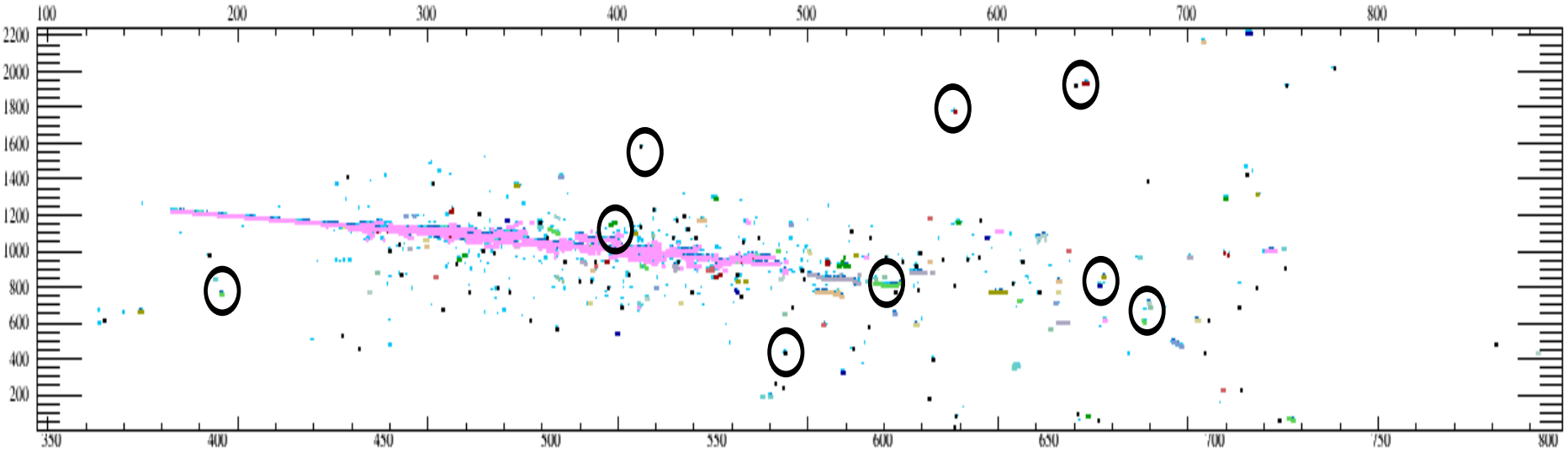
- In the slides that follow Fuzzy clustered events are on the left and DBscanned on the right

Event Display - Electron

Fuzzy

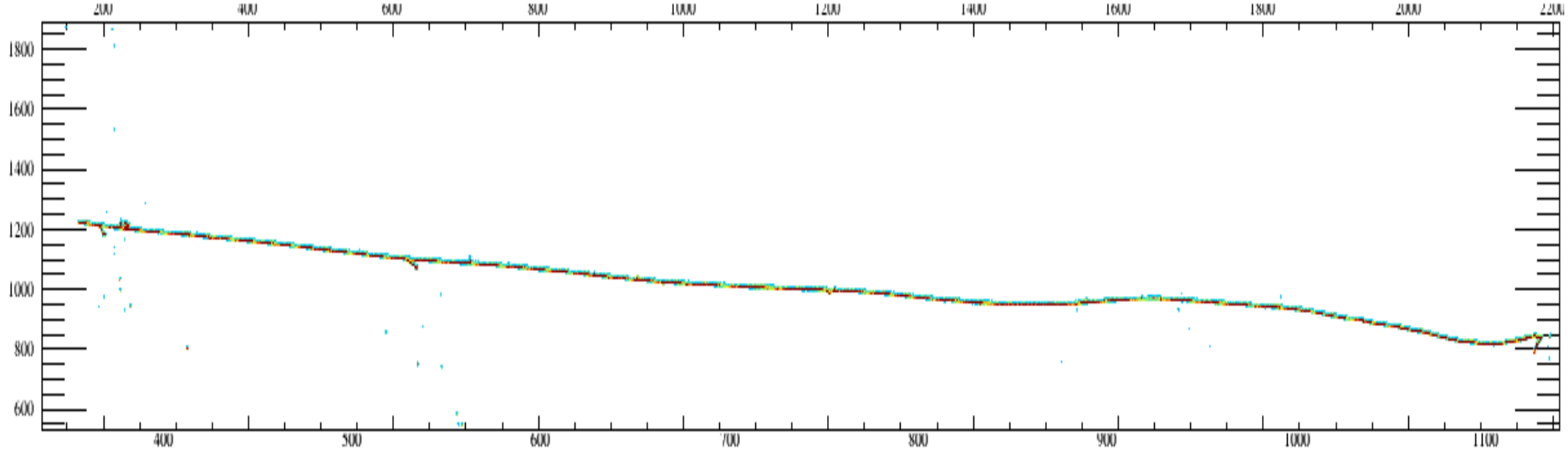


DB

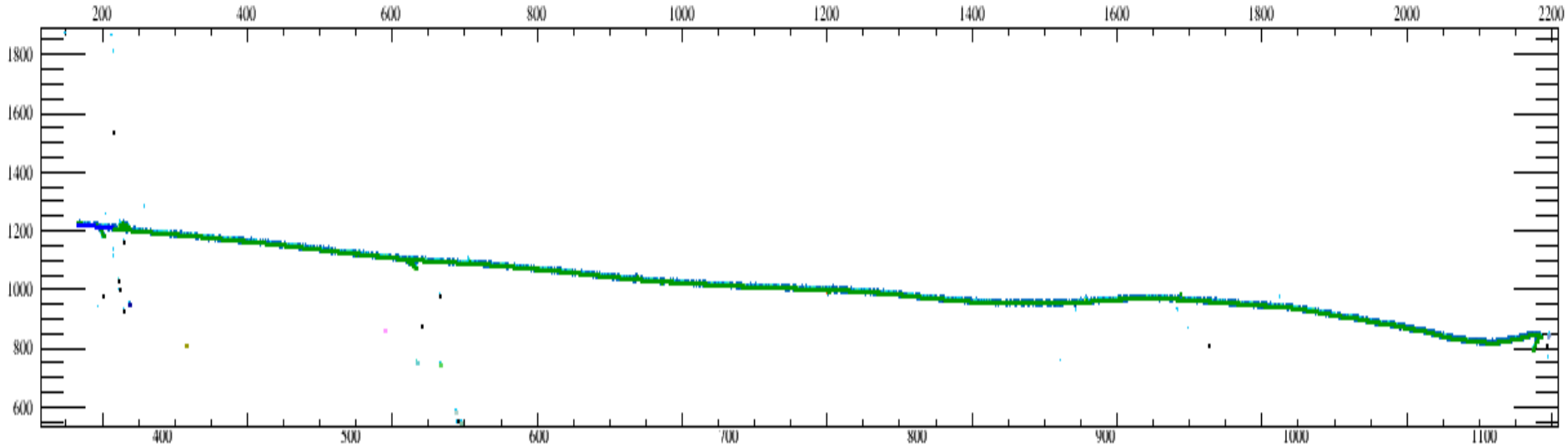


Event Display - Muon

Fuzzy

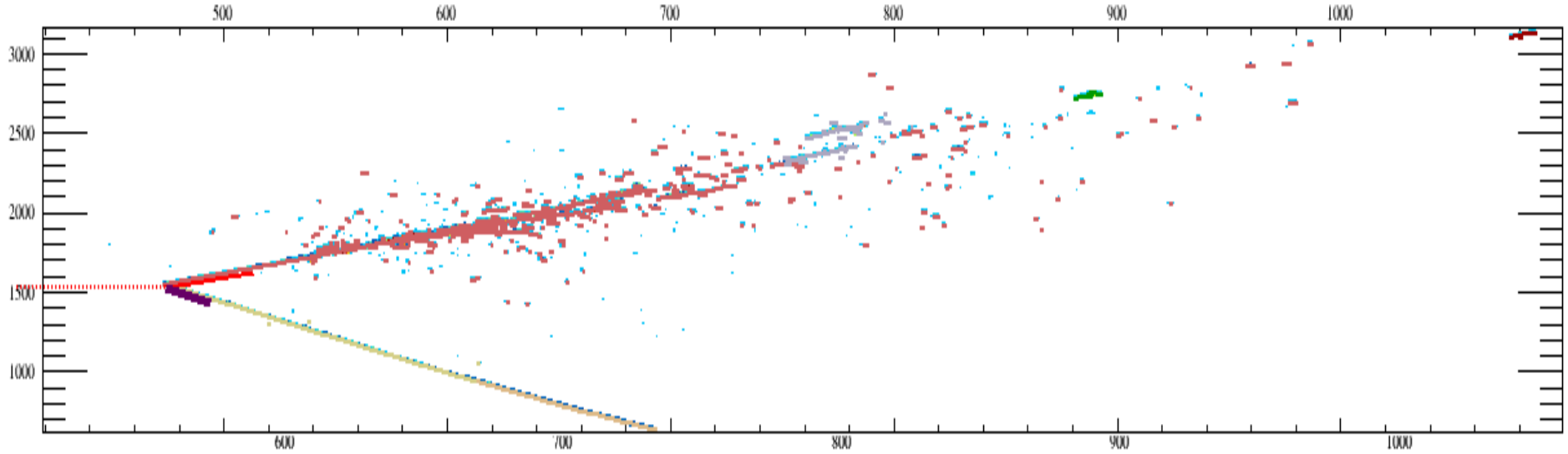


DB

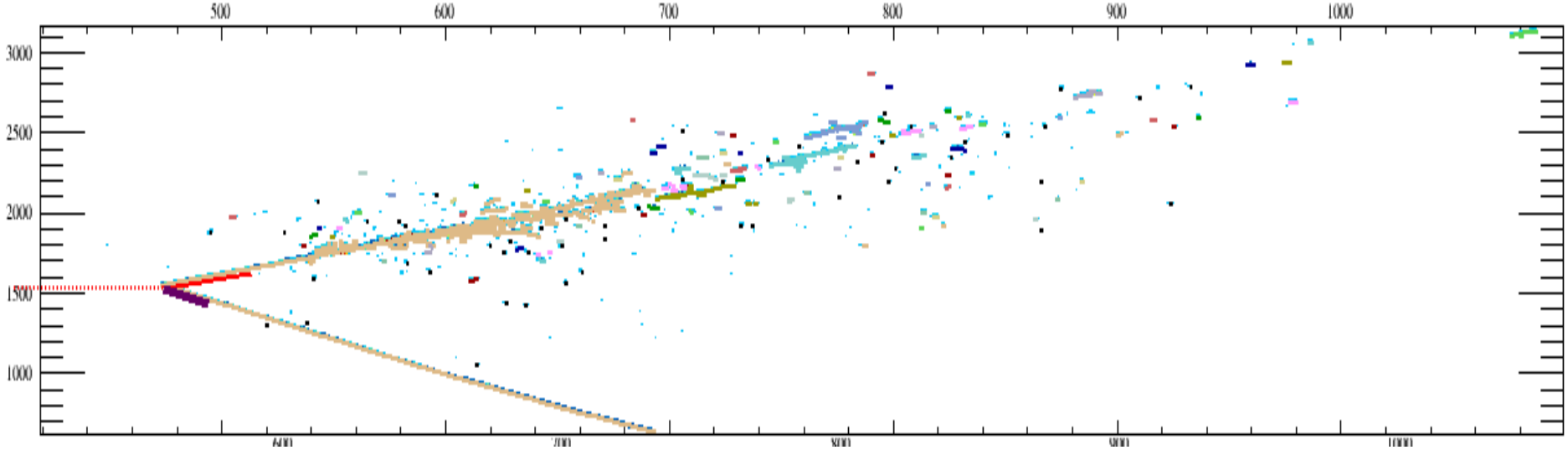


Event Display - $1e^- + 1p$

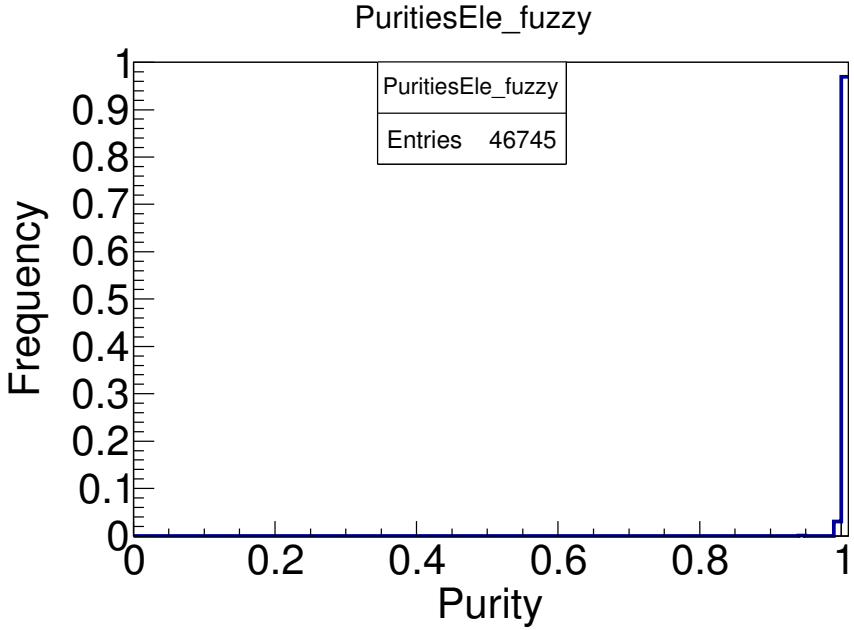
Fuzzy

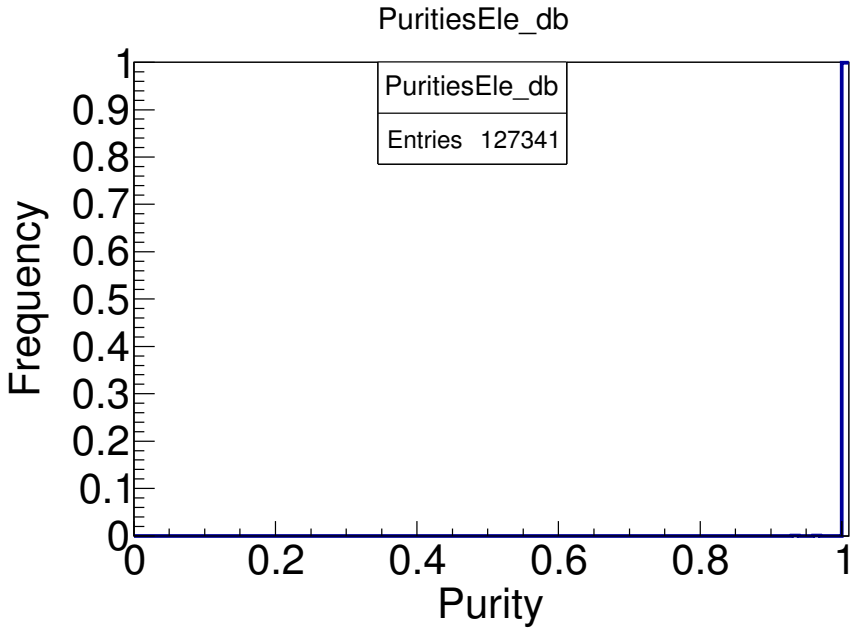


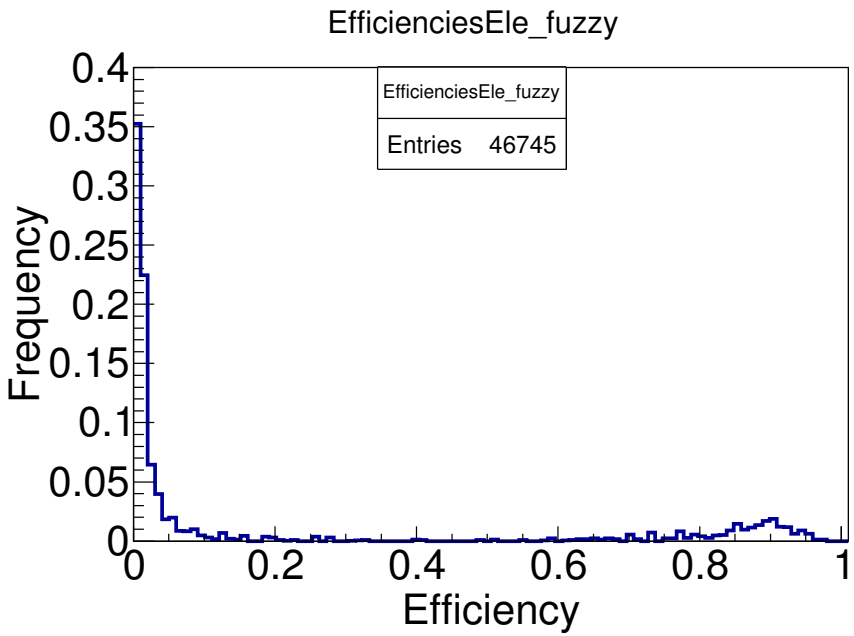
DB

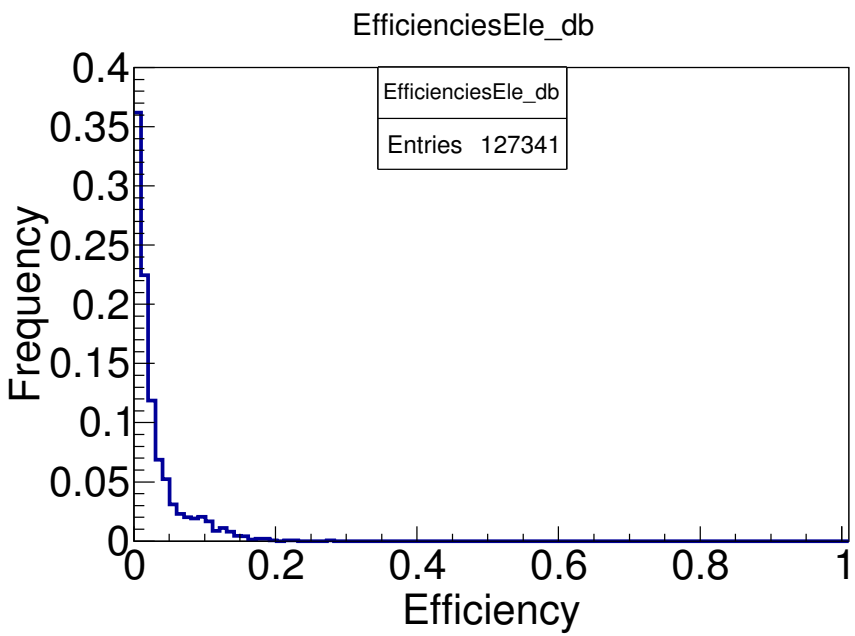


Single Electrons - Fuzzy left, DBscan right

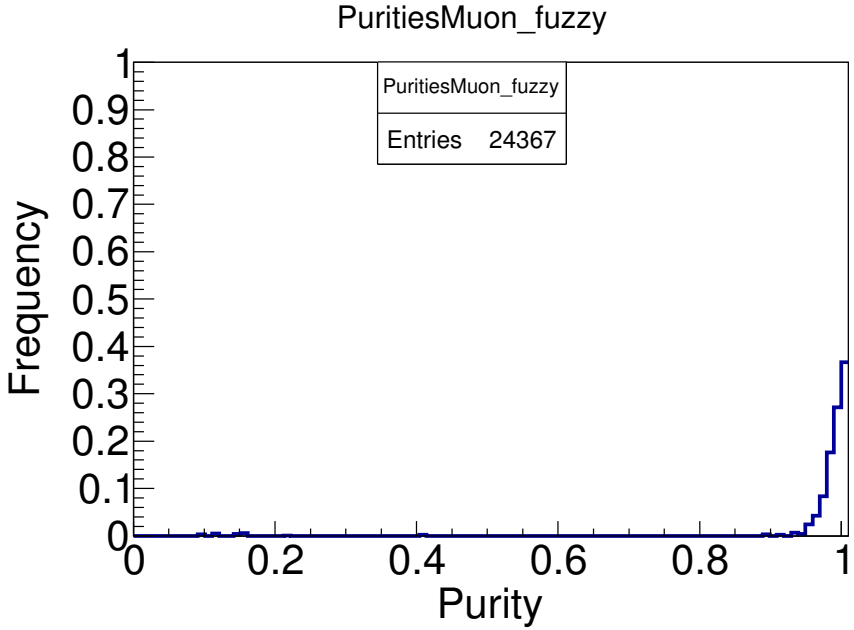


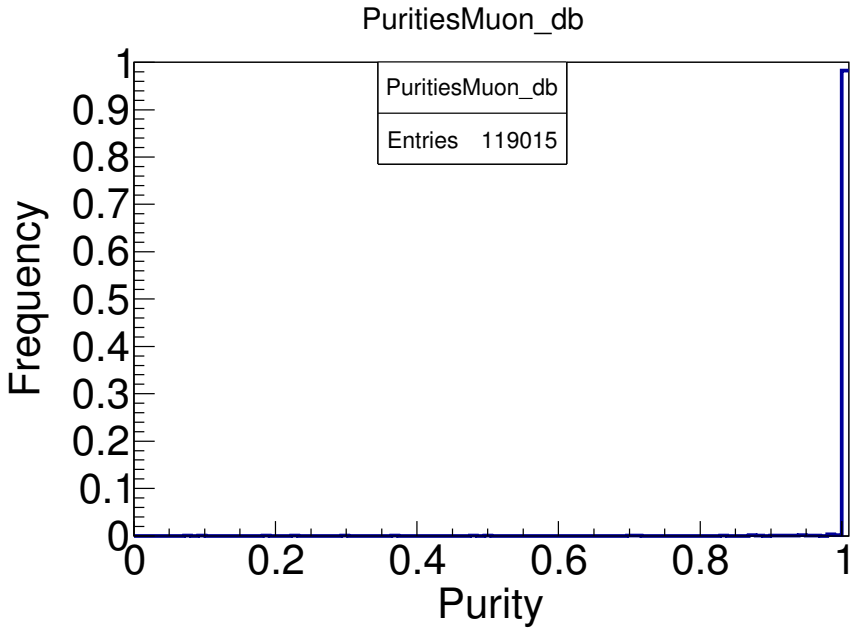


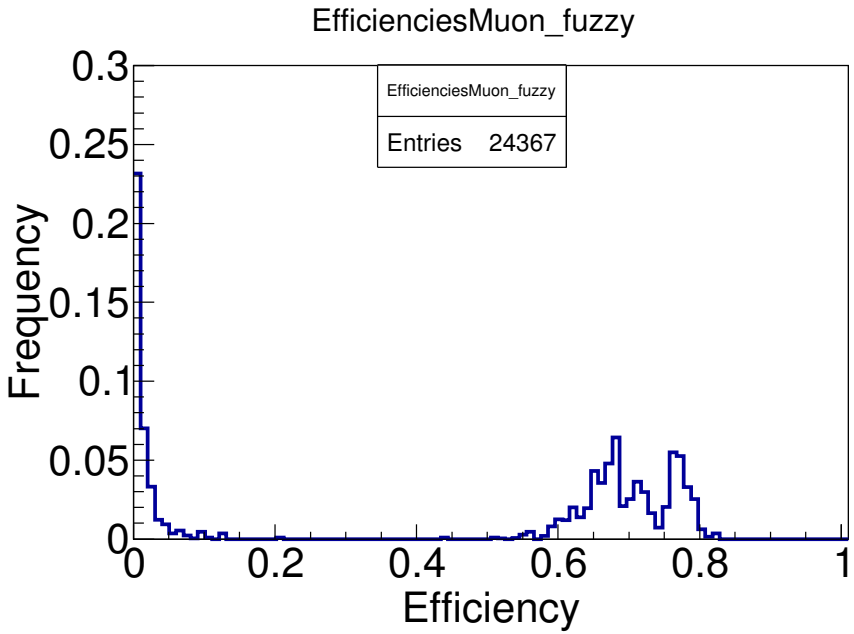


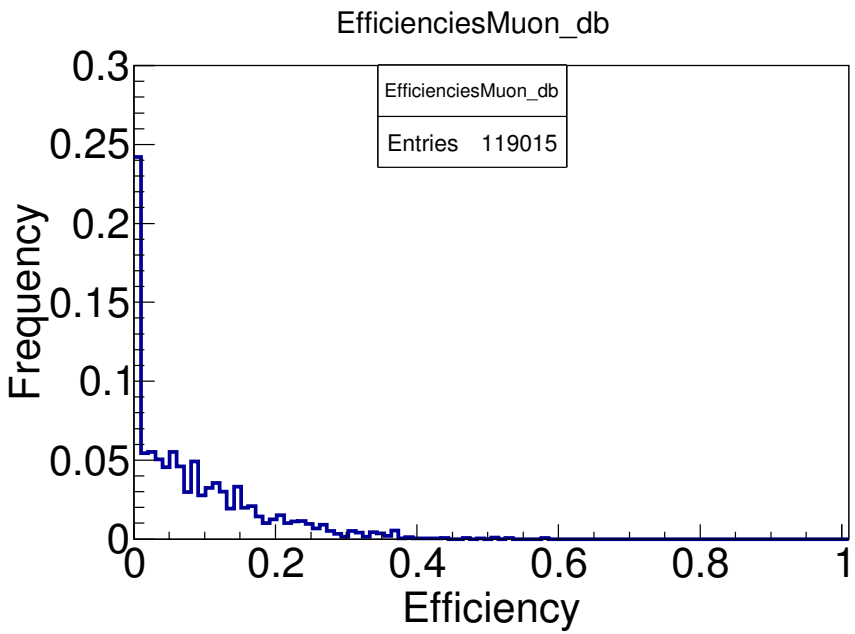


Single Muons

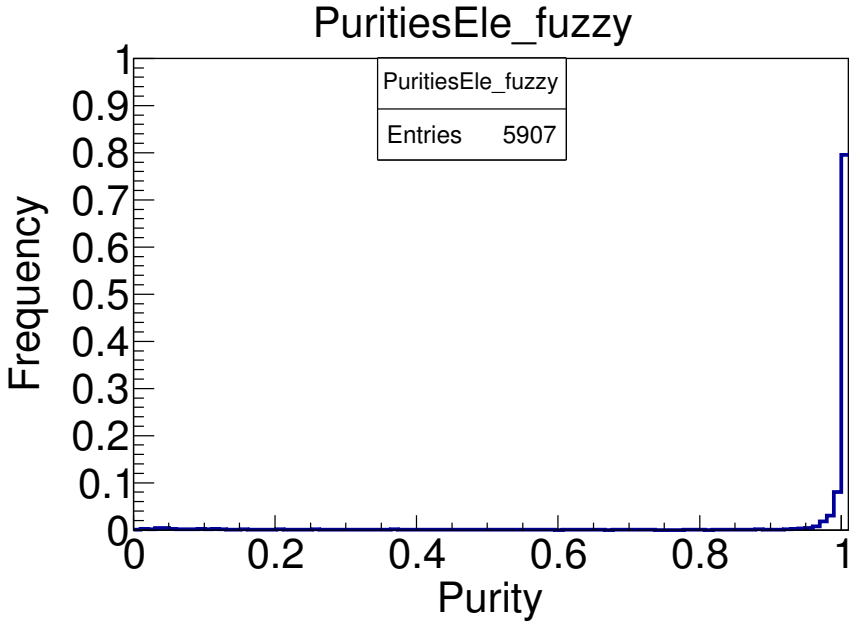


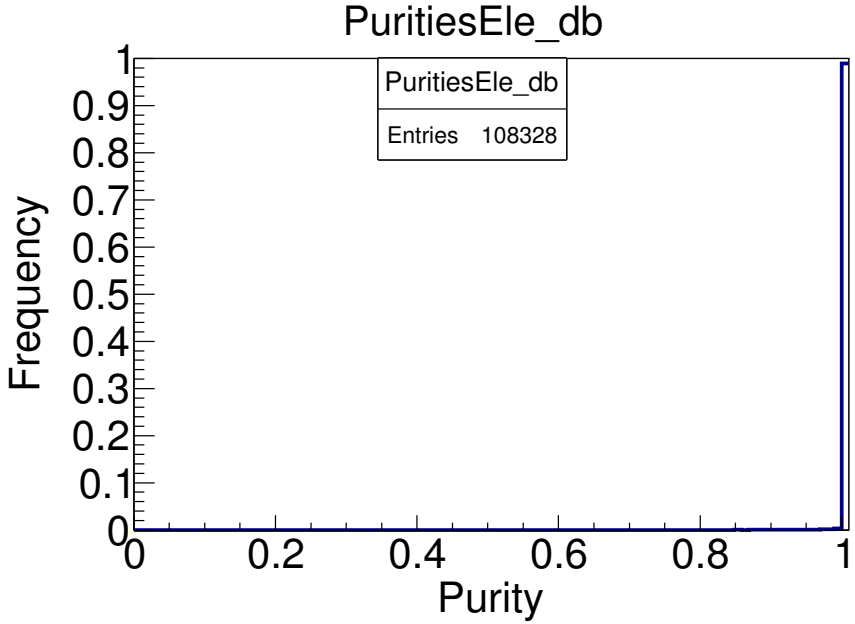


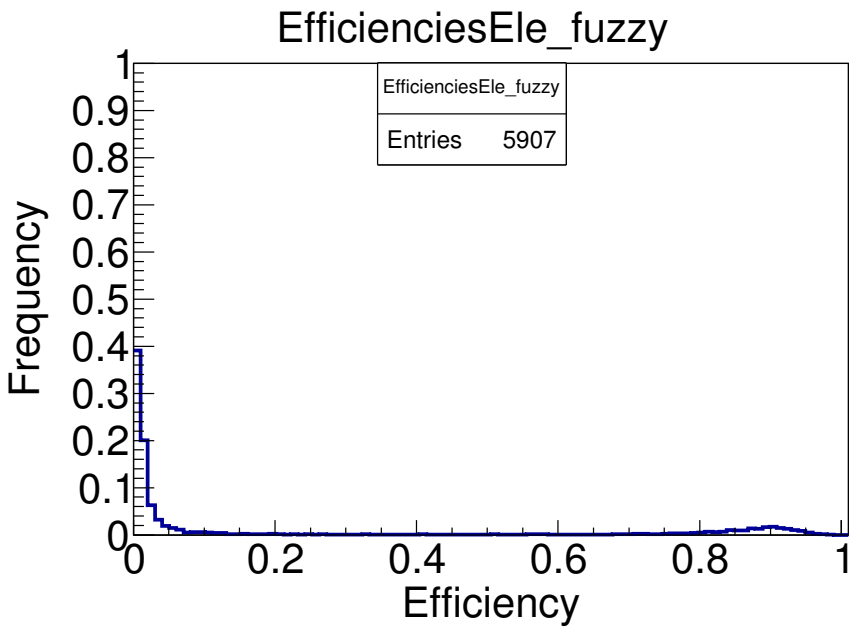


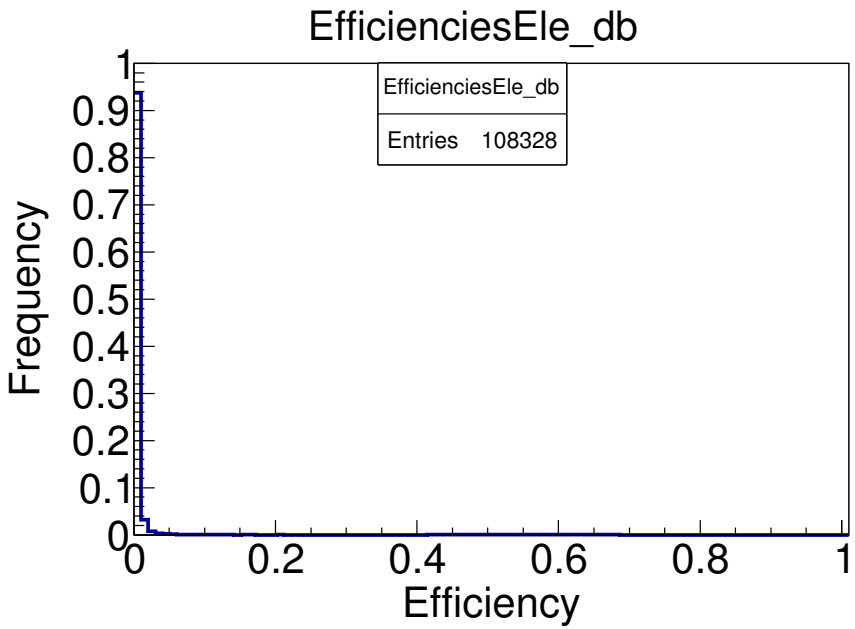


$1e^- + 1p$ - Electrons

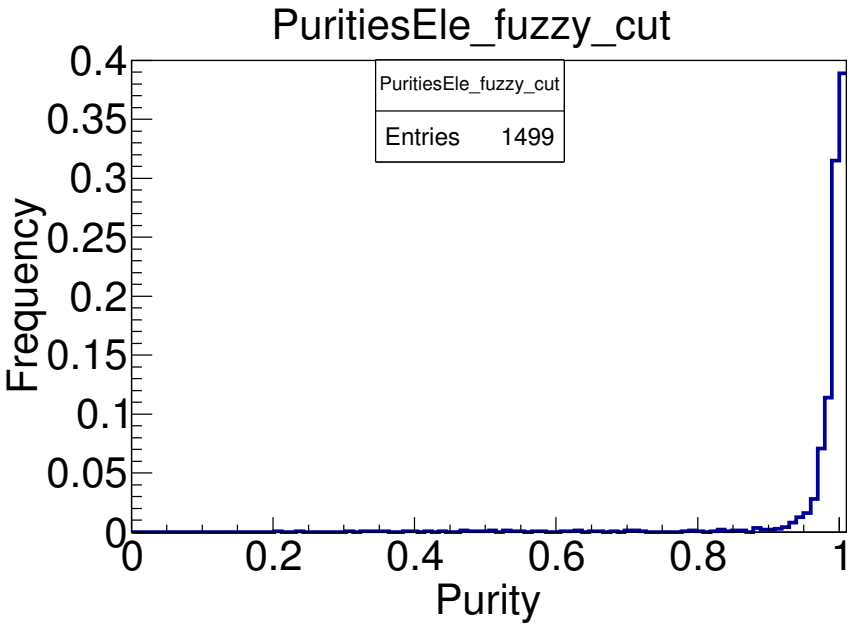


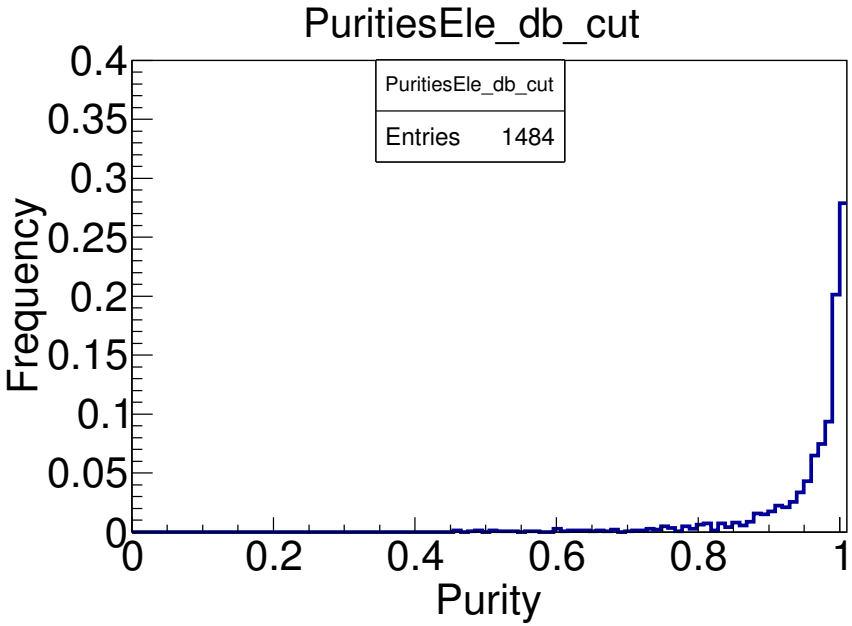


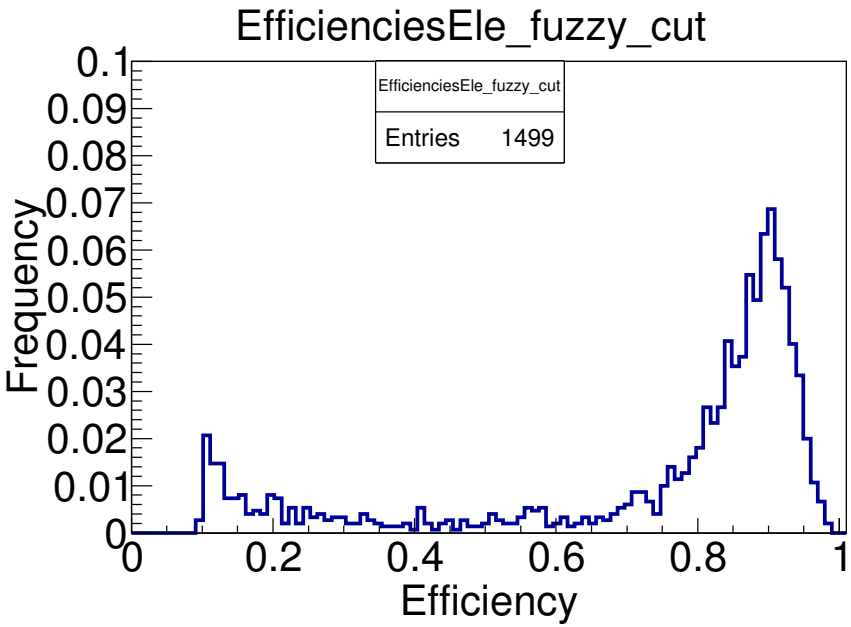


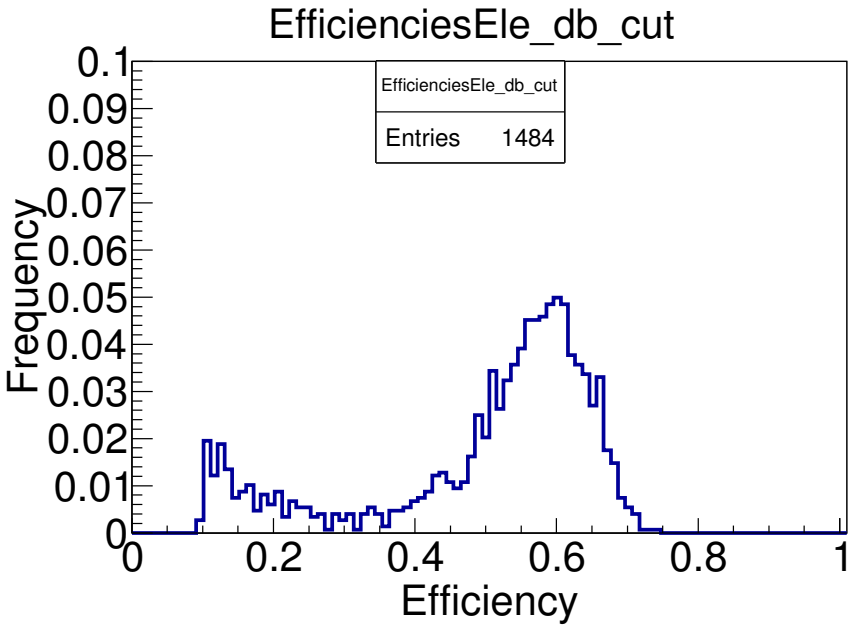


$1e^- + 1p$ - Electrons Cut

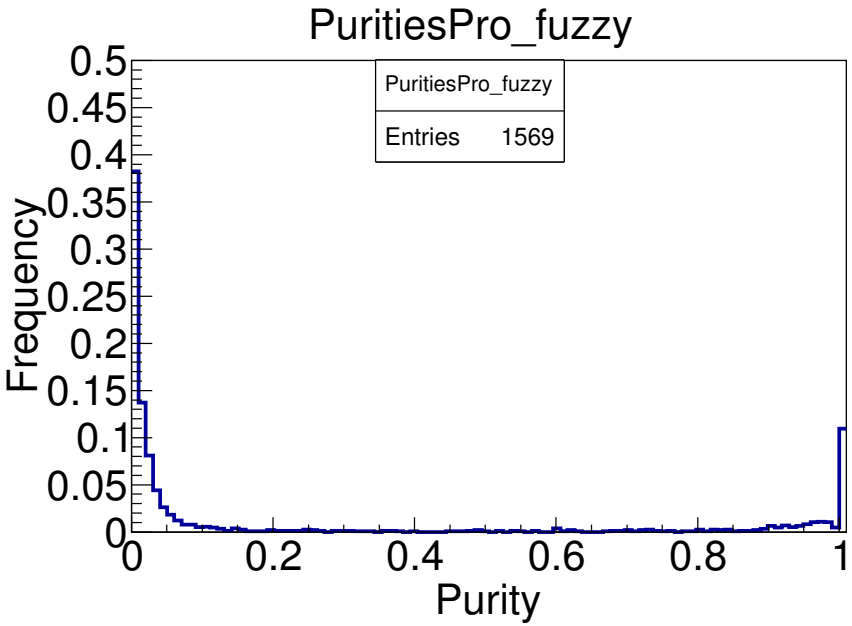


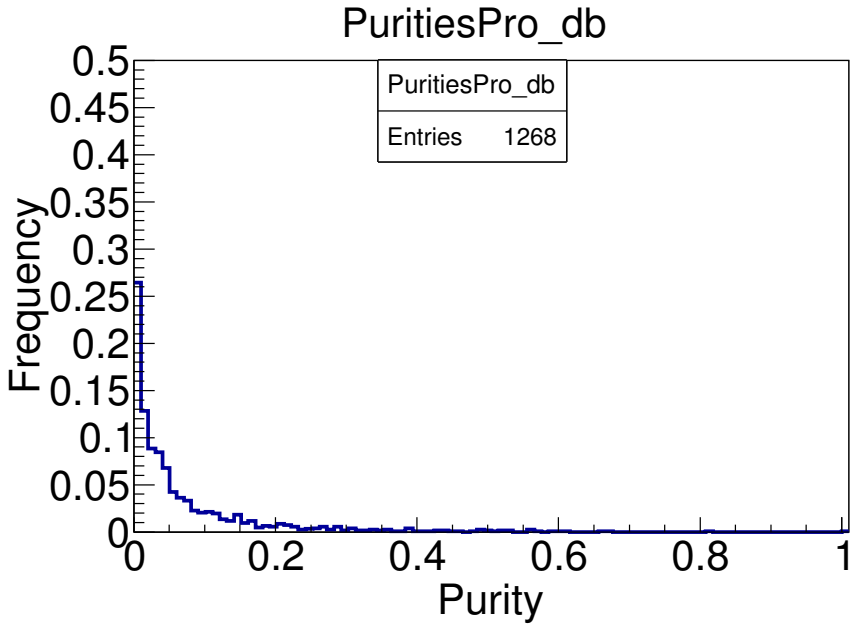


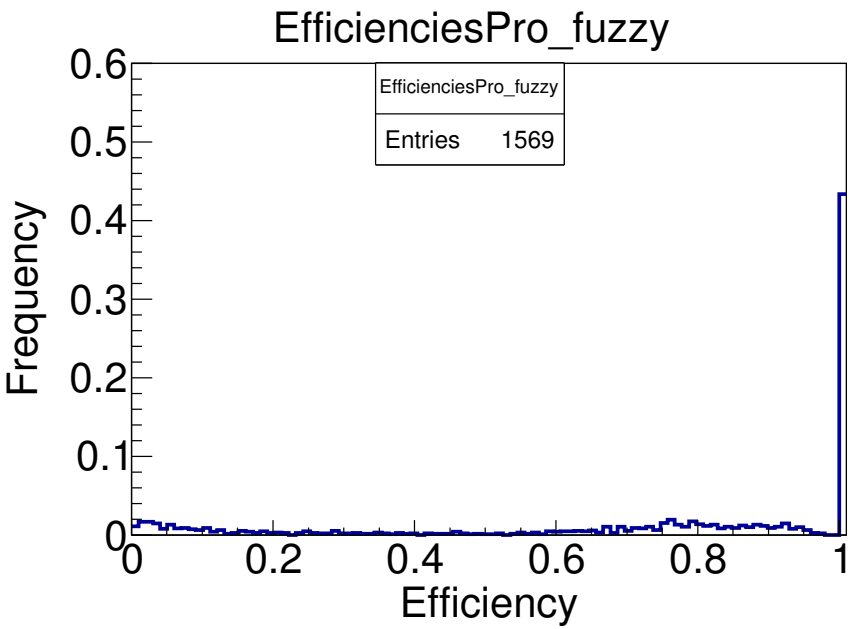


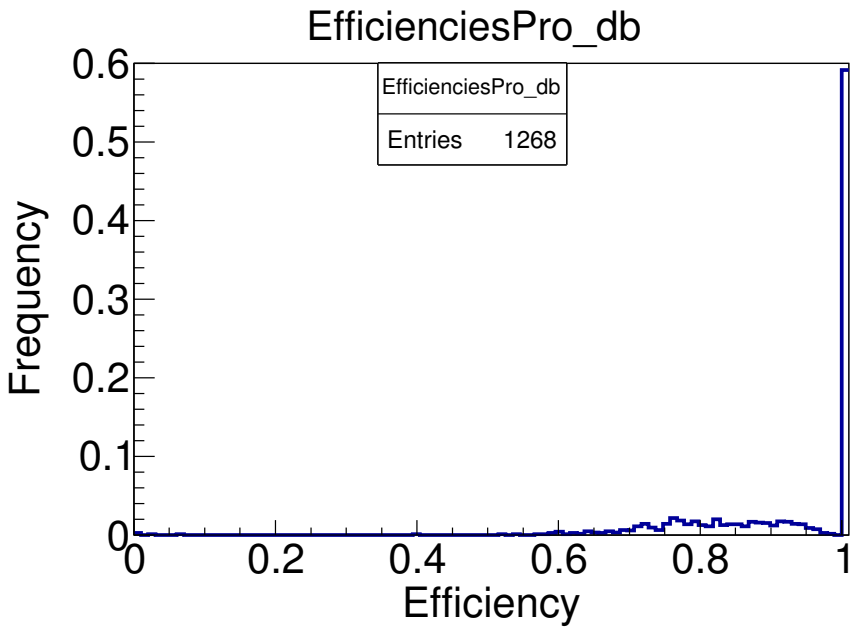


$1e^- + 1p$ - Protons

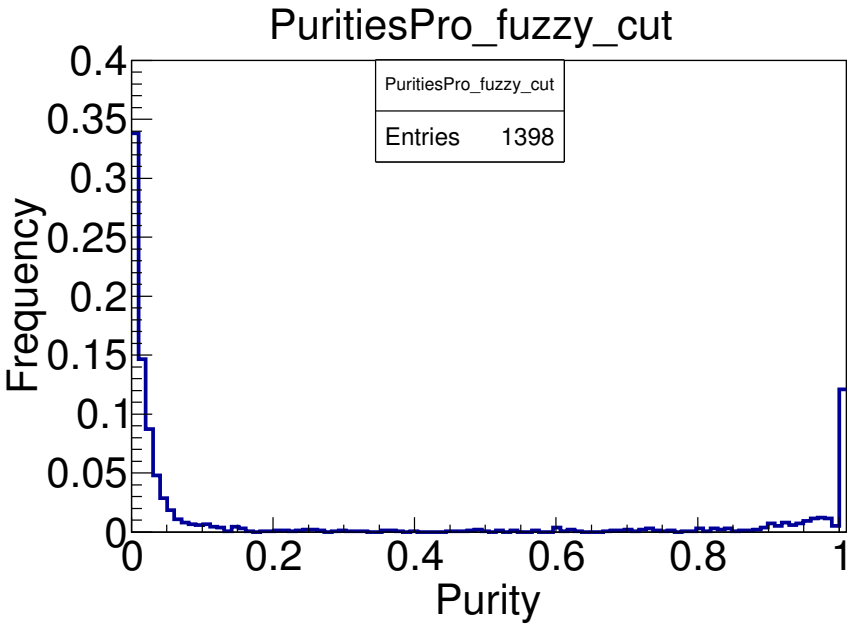


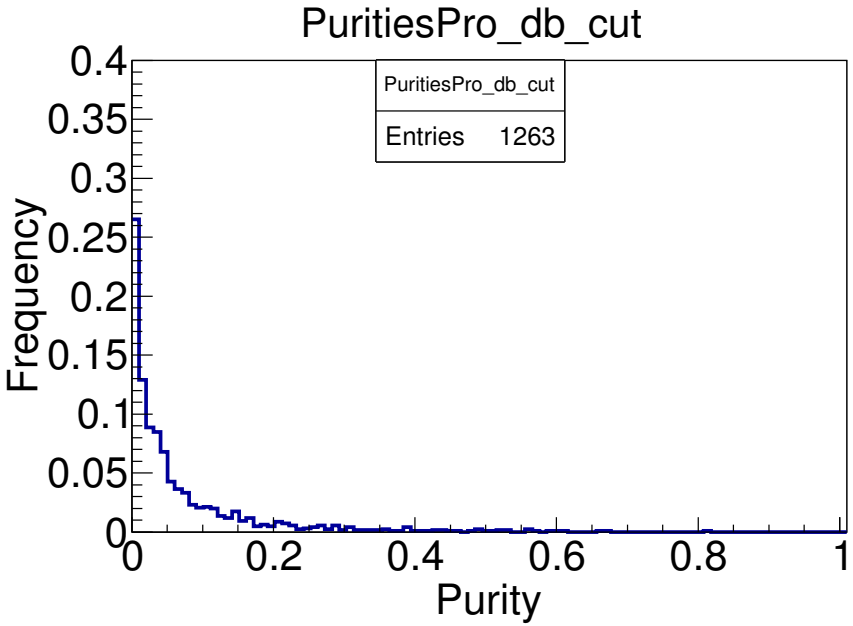


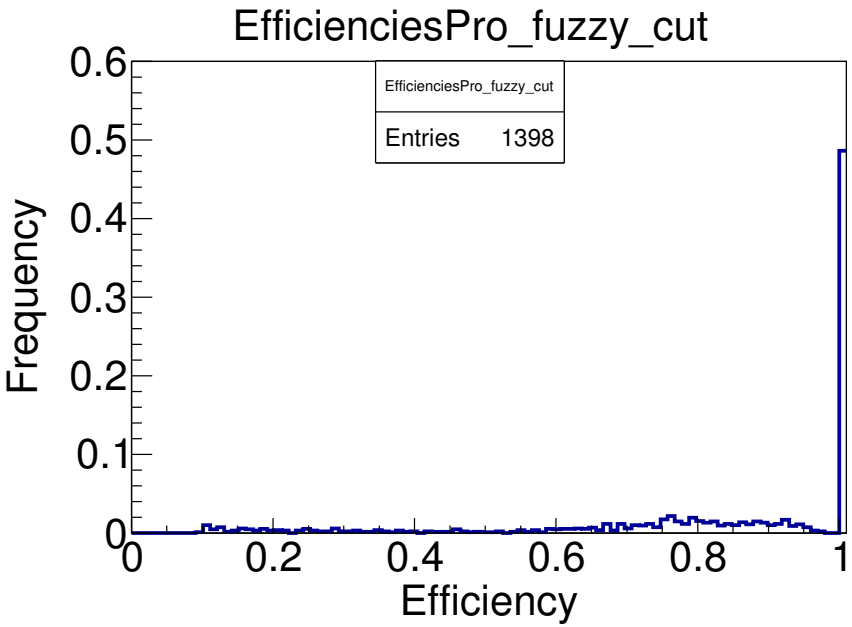


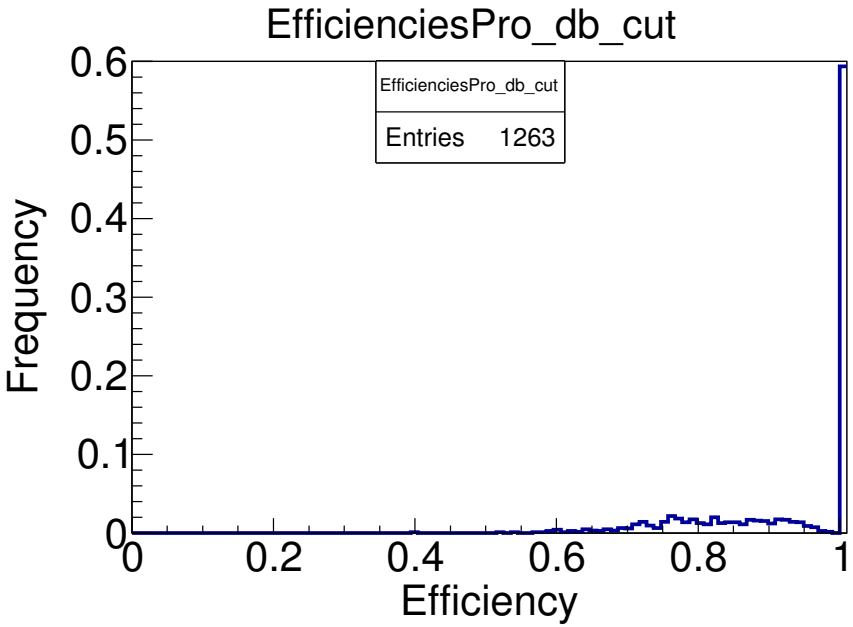


$1e^- + 1p$ - Protons Cut

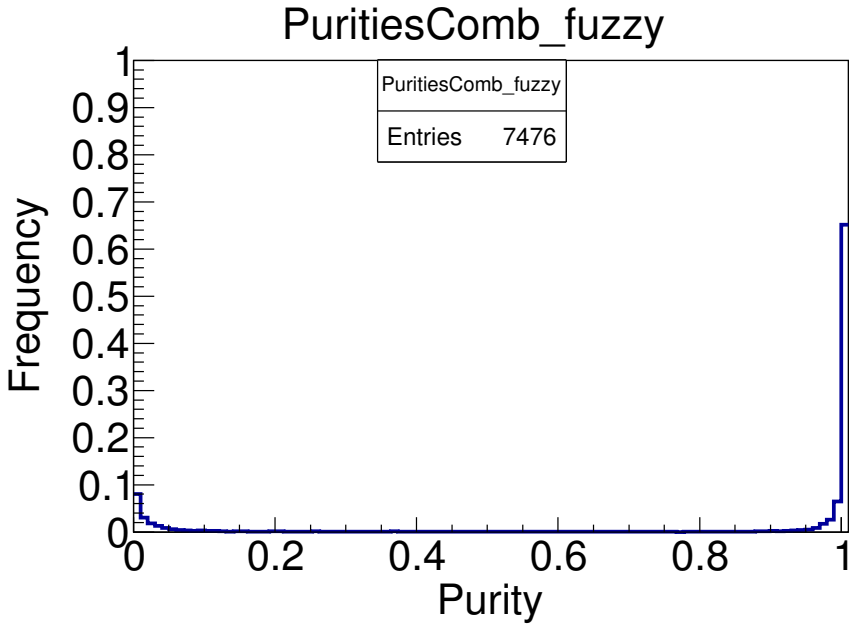


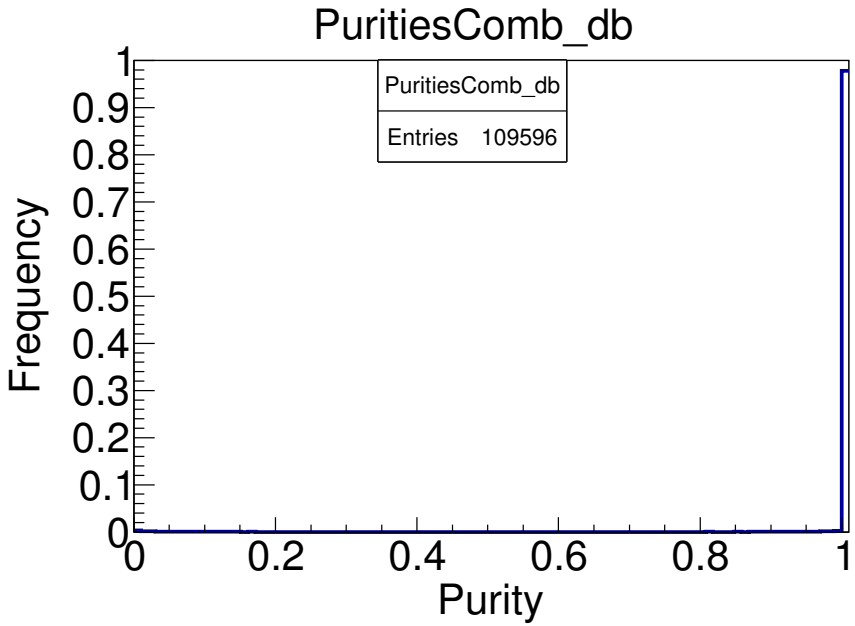


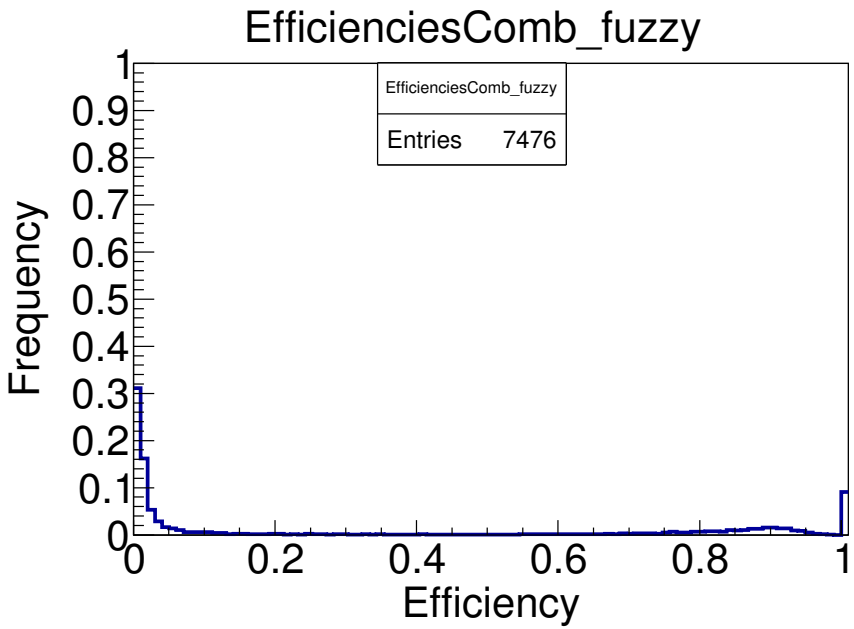


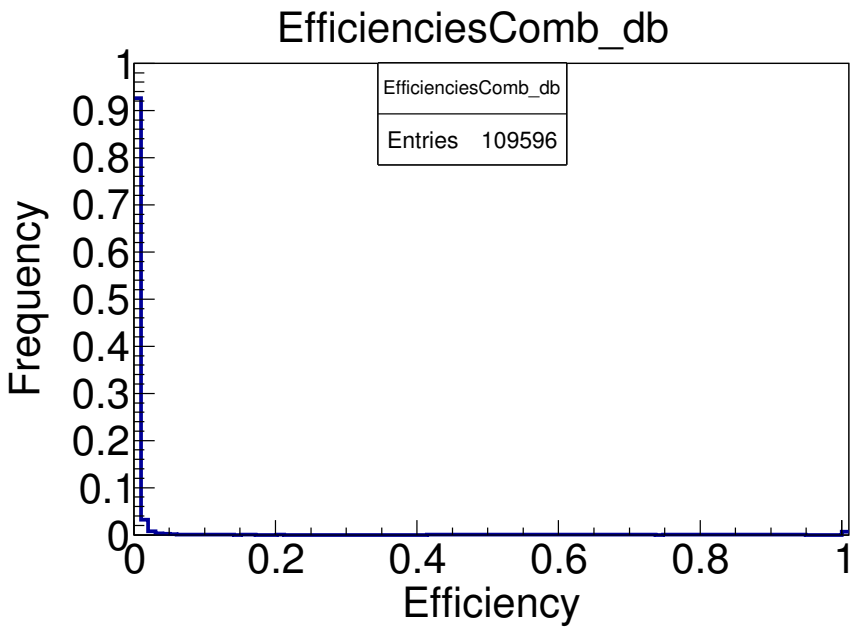


$1e^- + 1p$ - Combined

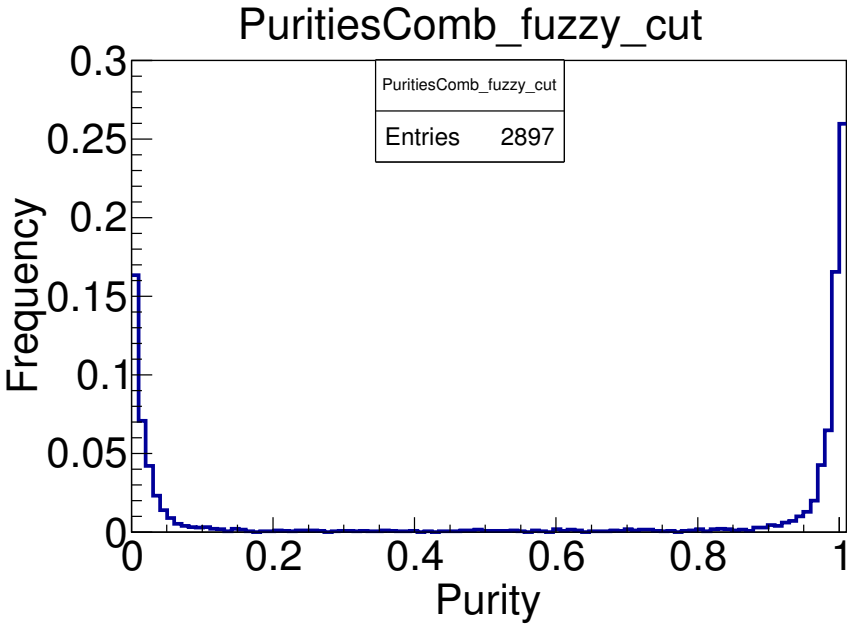


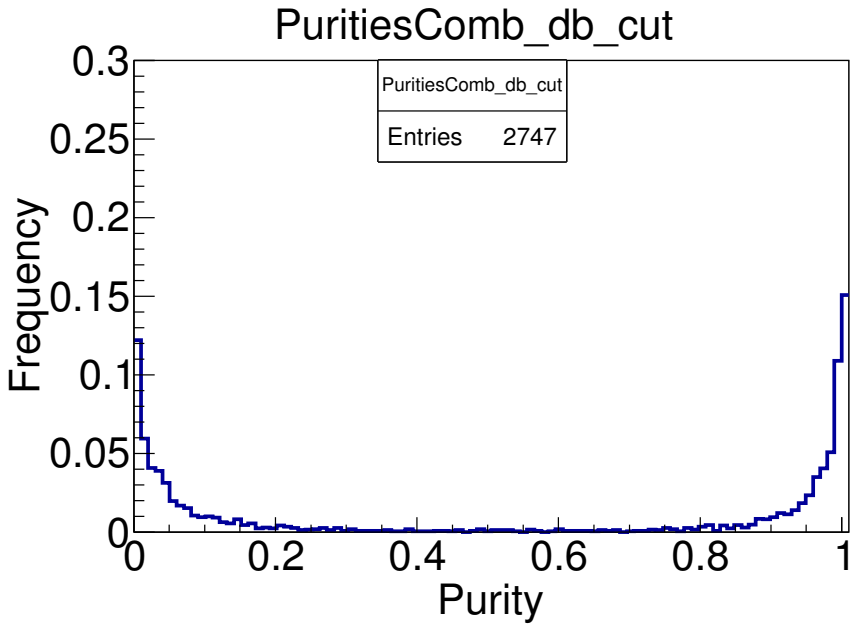


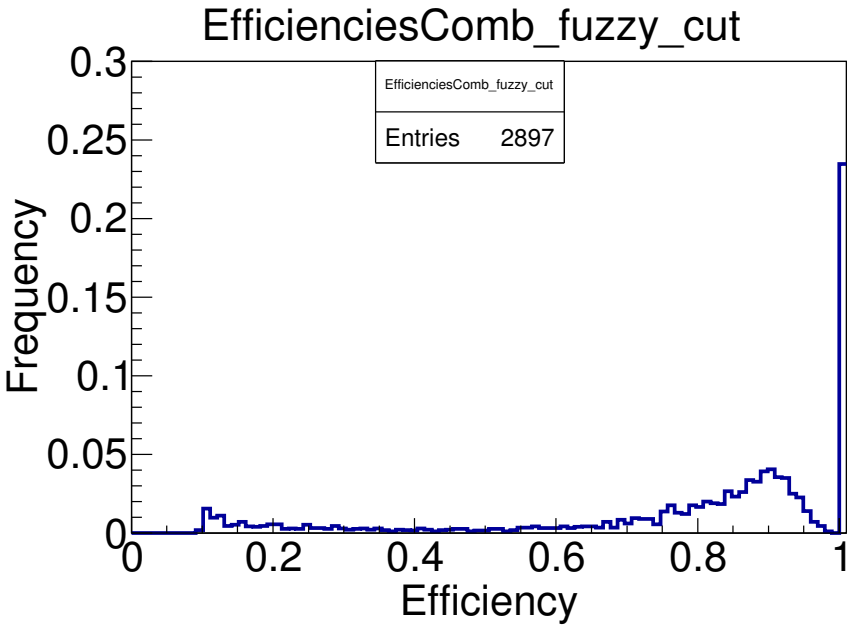


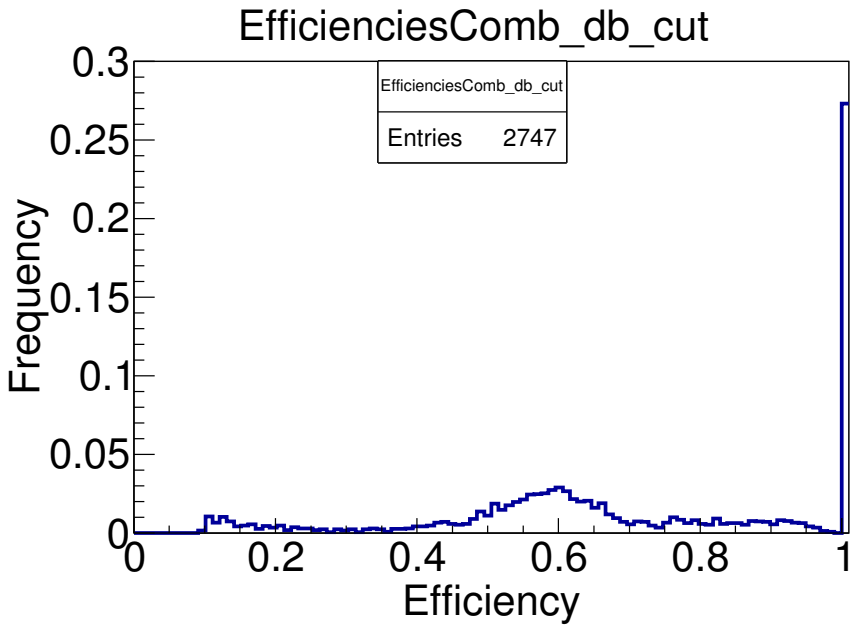


$1e^- + 1p$ - Combined Cut









Conclusions

1. Fuzzy cluster shows improved purity and efficiency values for filtered $1e^- + 1p$ events
2. Clustering also looks improved based on event display output
3. The main issue I've come accross is that the algorithm will fail to cluster proton hits separate from electron showers, especially when proton track is underneith the electron shower
4. Low efficiency clusters with high purity containing 1-5 hits still plague the algorithm

Future & References

Future

- Get neutrino flux files working
- Test 2/3D track finding algorithm

References

[1] "*Quick overview on fuzzy clustering for the March collaboration meeting*"
http://microboone-docdb.fnal.gov:8080/cgi-bin/ShowDocument?docid=2439