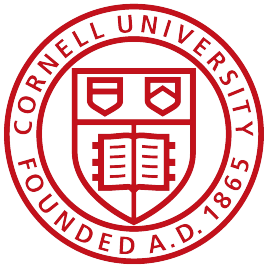


PR #170:

Track Embeddings for Duplicate Removal

Gavin Niendorf

Overview

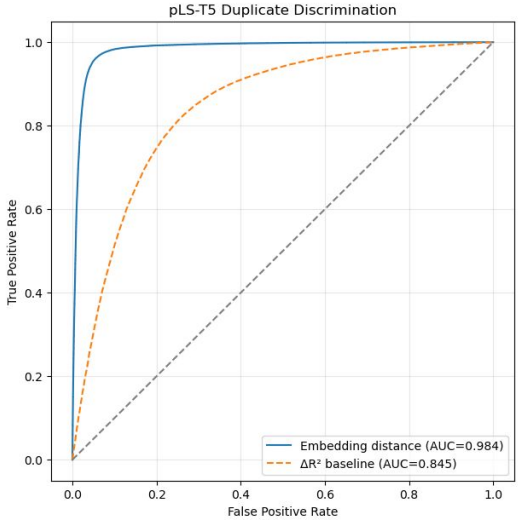
- Trained two DNN’s, one for pLS’s and one for T5’s, that map their unique input features to a shared six-dimensional embedding space.
- Placed cuts on the euclidean distance between T5-T5 embeddings and pLS-T5 embeddings to improve T5-T5, pT5-T5, and pLS-T5 duplicate cleaning.
- Increased displaced track efficiency through improved T5-T5 and pT5-T5 cleaning via embedding cuts, and reduced duplicate rate in the barrel with improved pLS-T5 cleaning.

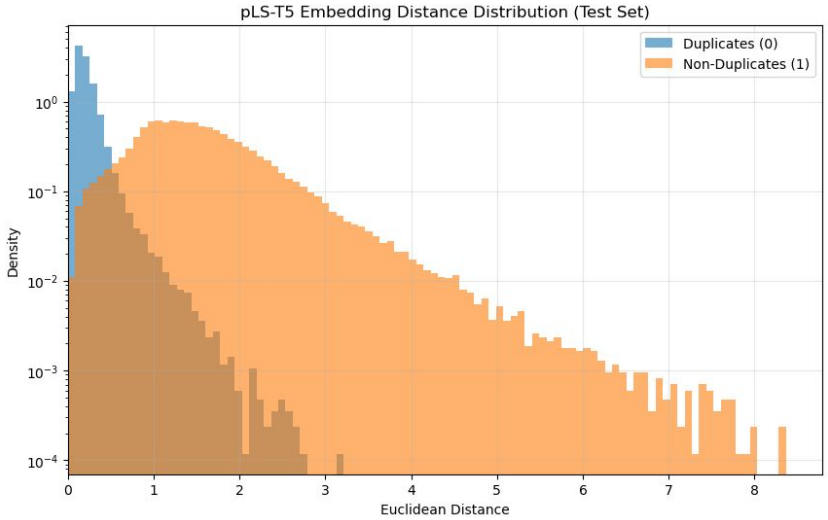
Embedding DNN Architectures

- Same architecture as other DNN's: 2 hidden layers with 32 nodes.
  - Both output a 6D vector representing a shared embedding space.
- **pLS Embedding DNN**
  - 10 input features - (eta, etaErr, cos(phi), sin(phi), 1 / pT, log10(ptErr), isQuad, log10(abs(circleCenterX)), log10(abs(circleCenterY)), log10(circleRadius)).
  - Still room to improve on normalizations / include other inputs.
- **T5 Embedding DNN**
  - 30 input features - Same as T5 fake rejection DNN, with small changes (abs(phi) → [cos(phi), sin(phi)], log10(radii) → 1 / radii)

Training Overview (pLS-T5 Plots)

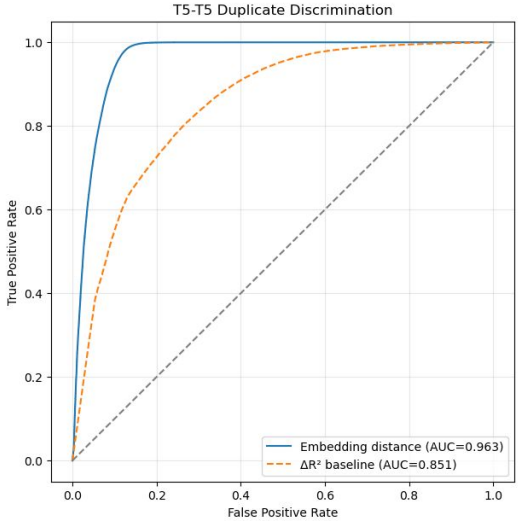
- The pLS and T5 embedding DNN's were trained simultaneously with the same Contrastive Loss function used previously.
- T5-T5 and pLS-T5 pairs are only included if at least one track is real; fake-fake pairs are excluded, and fake pLS's are not used. All pairs must satisfy $\Delta R^2 < 0.02$.
- Pairs with a displaced T5 are given a 5x weight in the loss function.

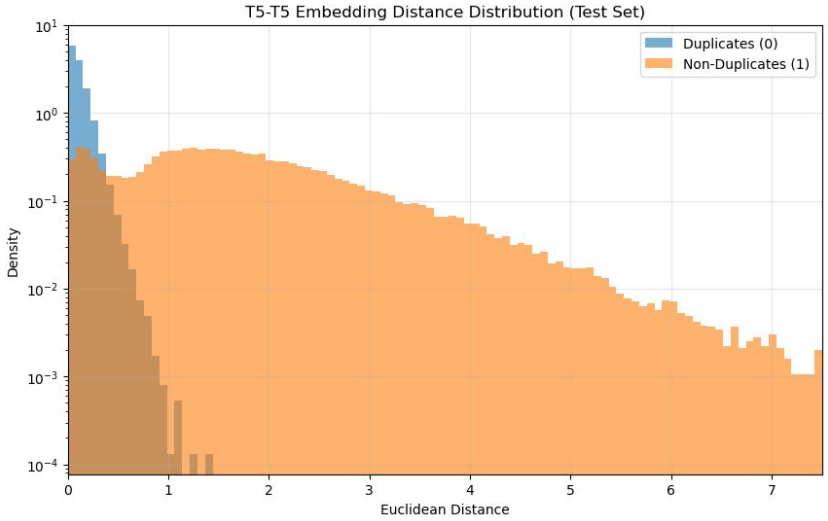




Training Overview (T5-T5 Plots)

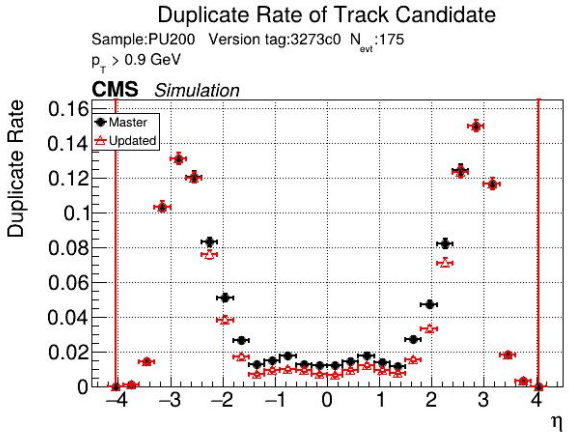
- The pLS and T5 embedding DNN's were trained simultaneously with the same Contrastive Loss function used previously.
- T5-T5 and pLS-T5 pairs are only included if at least one track is real; fake-fake pairs are excluded, and fake pLS's are not used. All pairs must satisfy $\Delta R^2 < 0.02$.
- Pairs with a displaced T5 are given a 5x weight in the loss function.

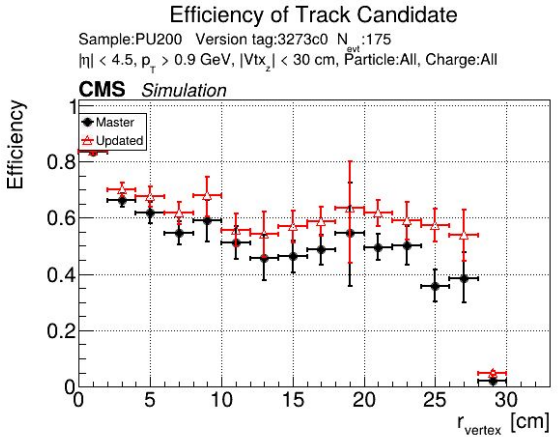


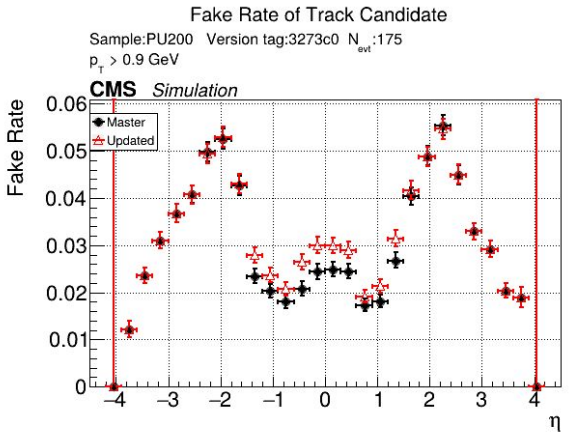


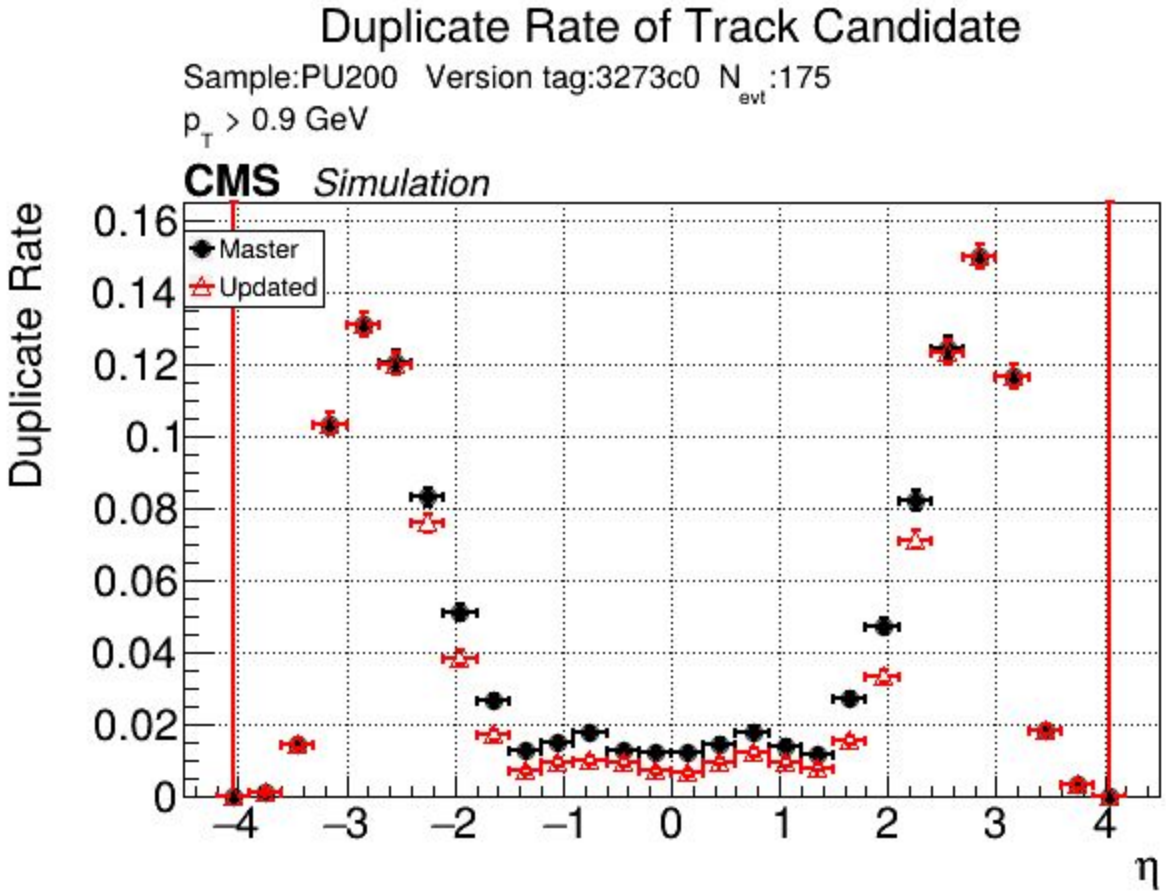
Initial Performance Results, Standalone

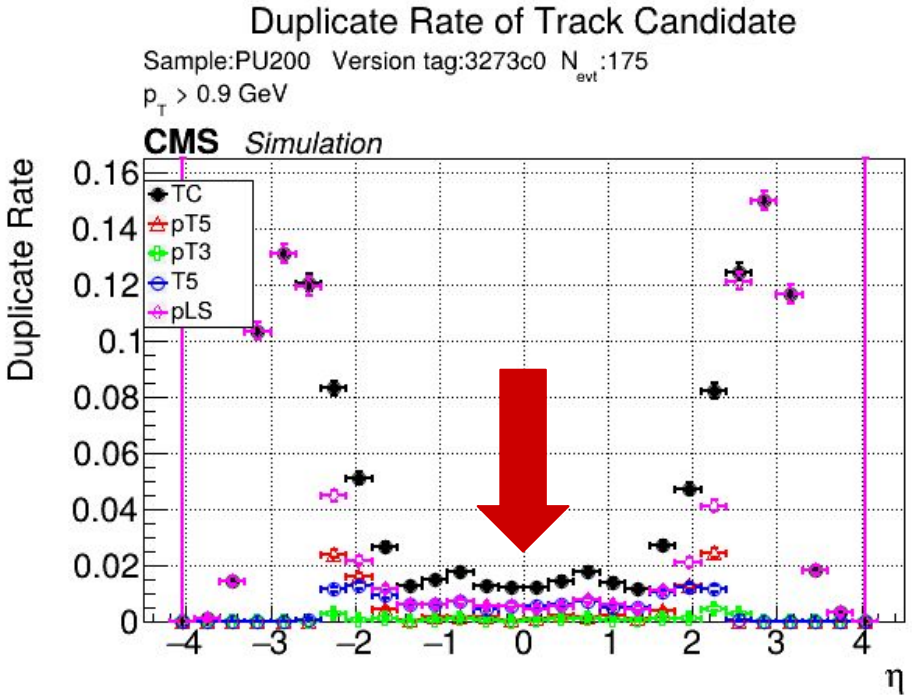
- Duplicate rate decreases from pLS-T5 embedding cuts, and displaced track efficiency increases from T5-pT5 and T5-T5 embedding cuts.
- The increase in T5 fake rate is due to fake tracks that were previously removed by duplicate cleaning. This can likely be mitigated by lowering the 75% real hit threshold for T5 matching during training.

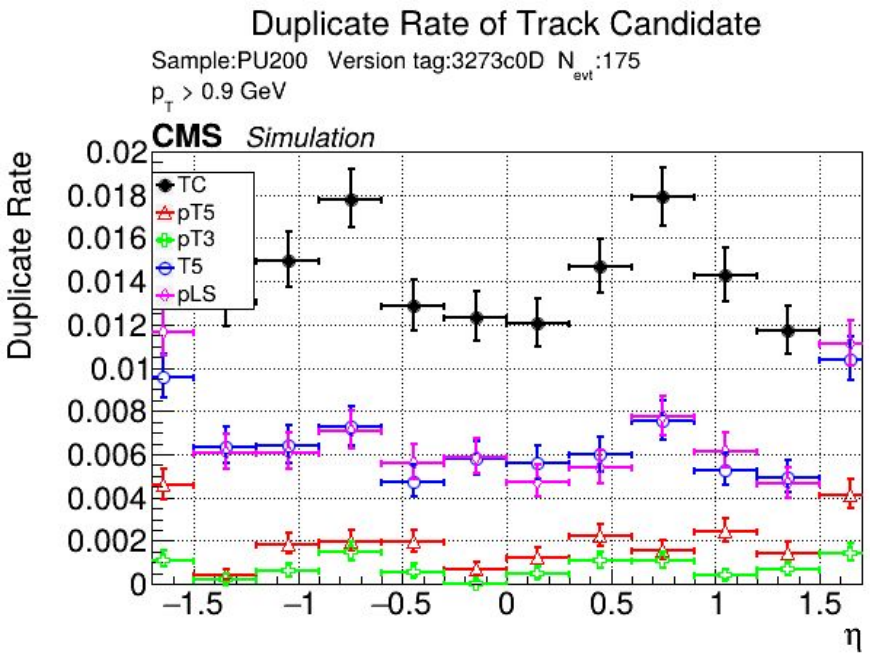




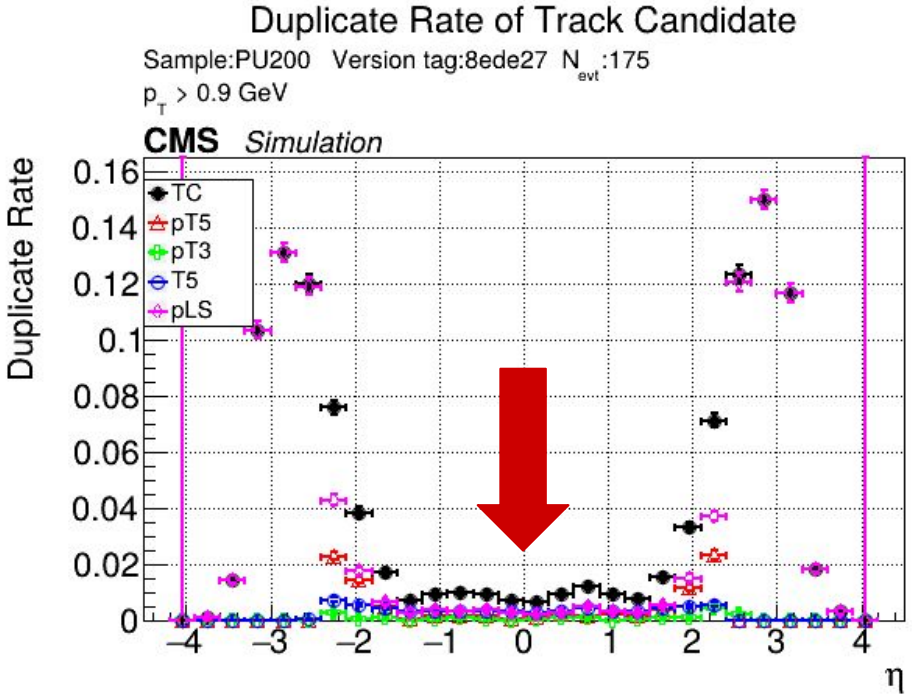


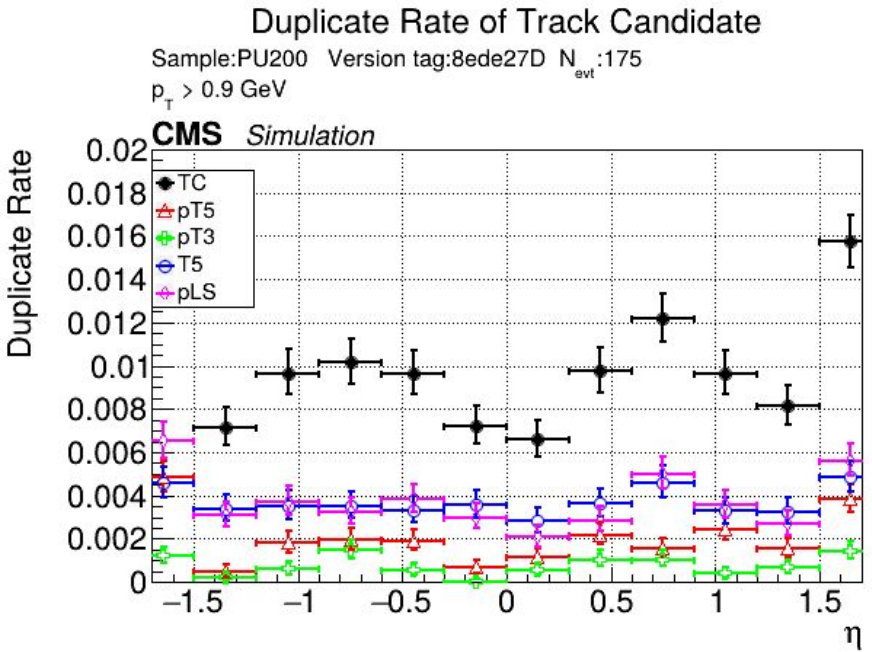

Duplicate Rate of Track Candidate
Sample:PU200 Version tag:3273c0 N_evt:175
p_T > 0.9 GeV
CMS Simulation
Duplicate Rate
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
Master
Updated
−4
−3
−2
−1
0
1
2
3
4
η





Barrel Region Zoom

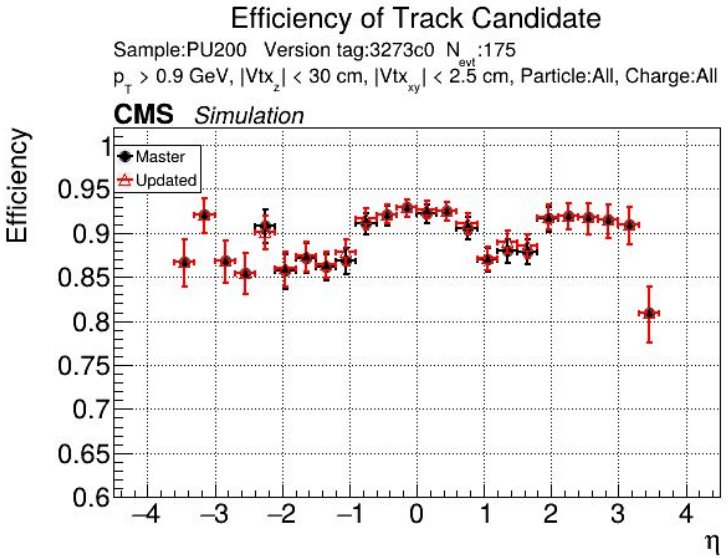


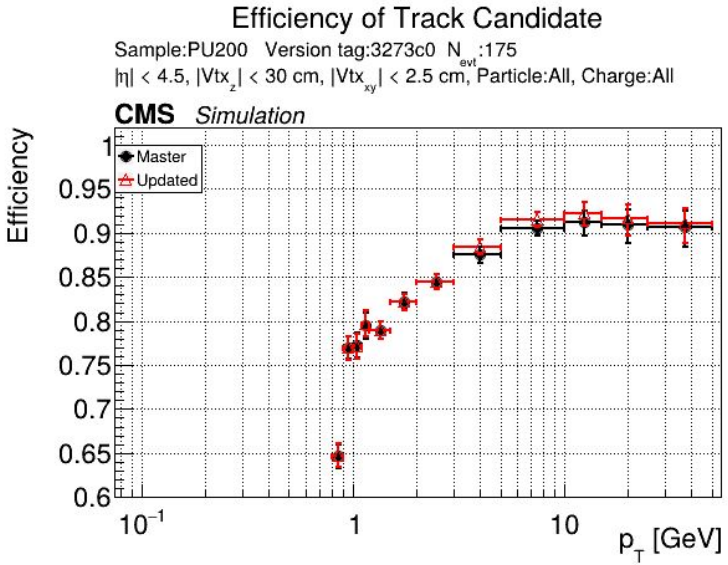


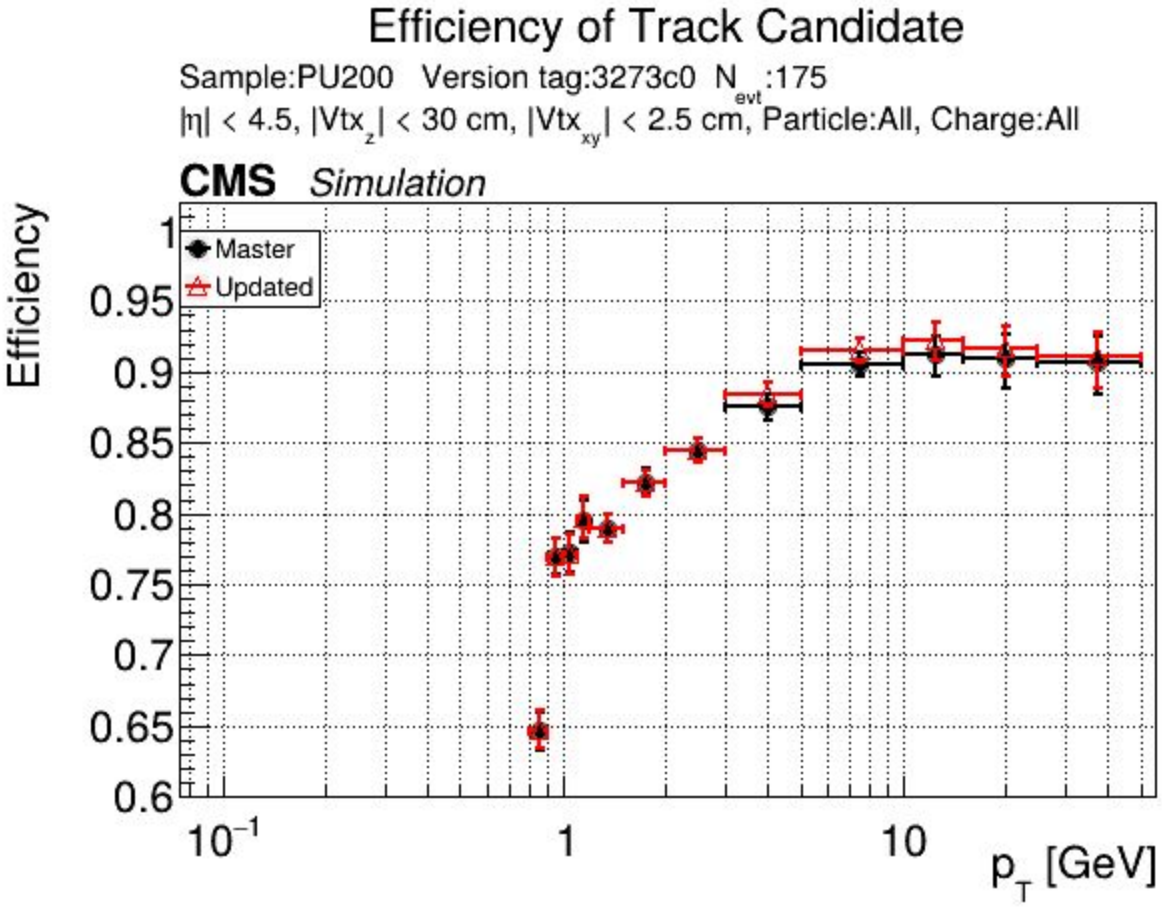
Barrel Region Zoom

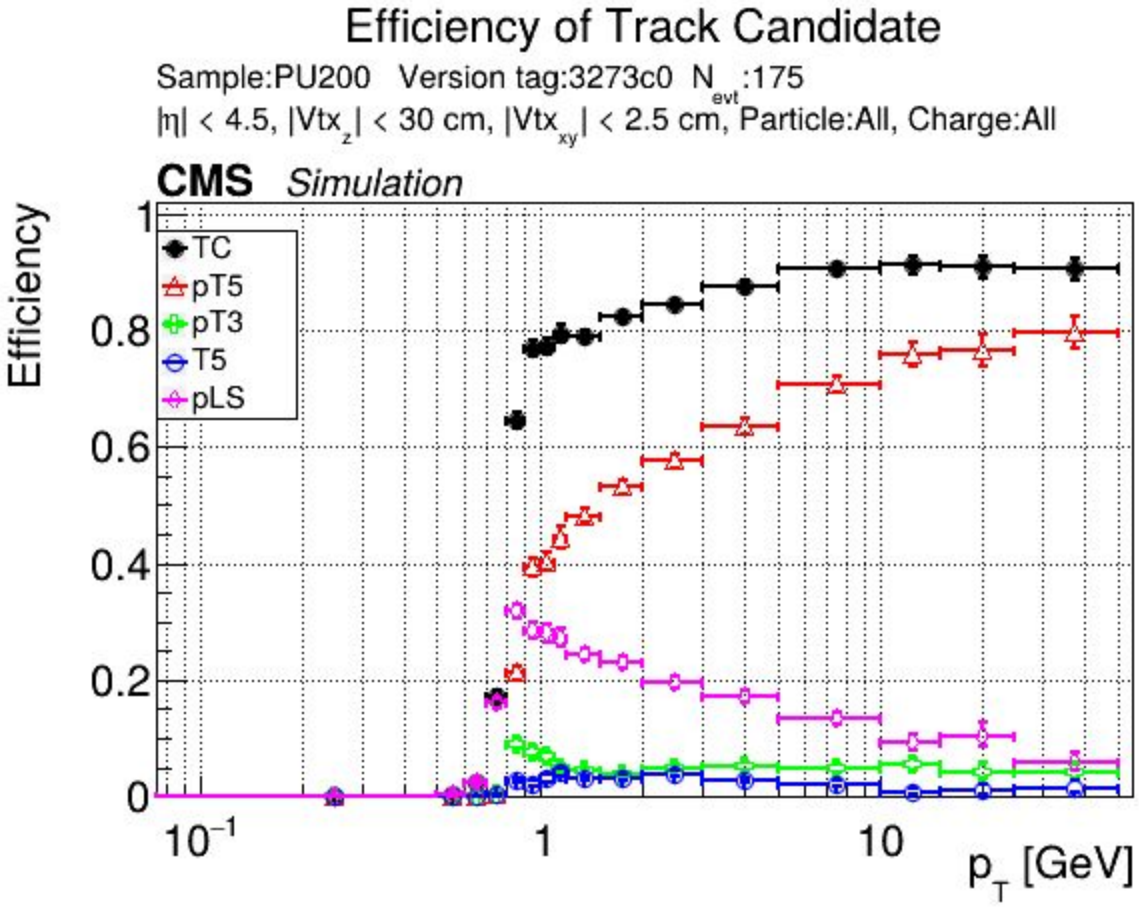
Initial Performance Results, Standalone

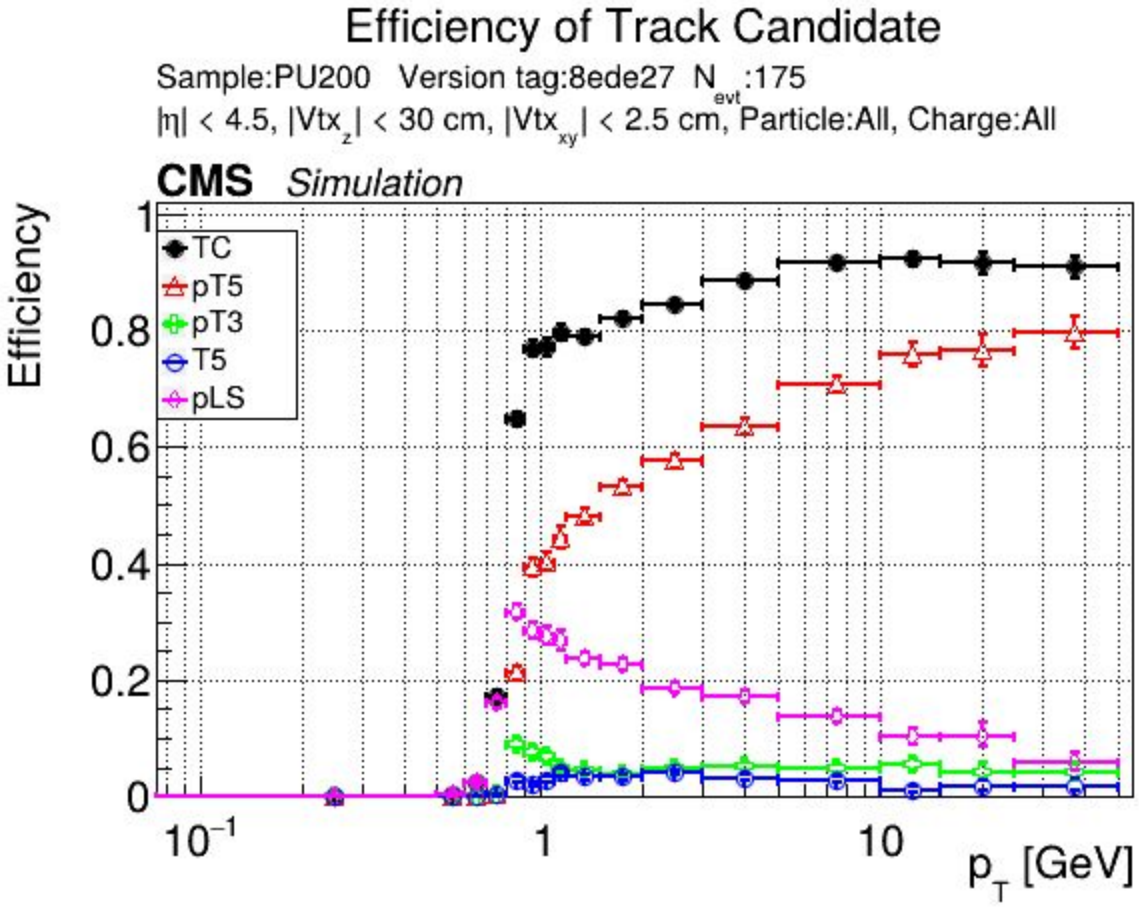
- Slight overall efficiency increase from improved duplicate cleaning via embedding cuts.





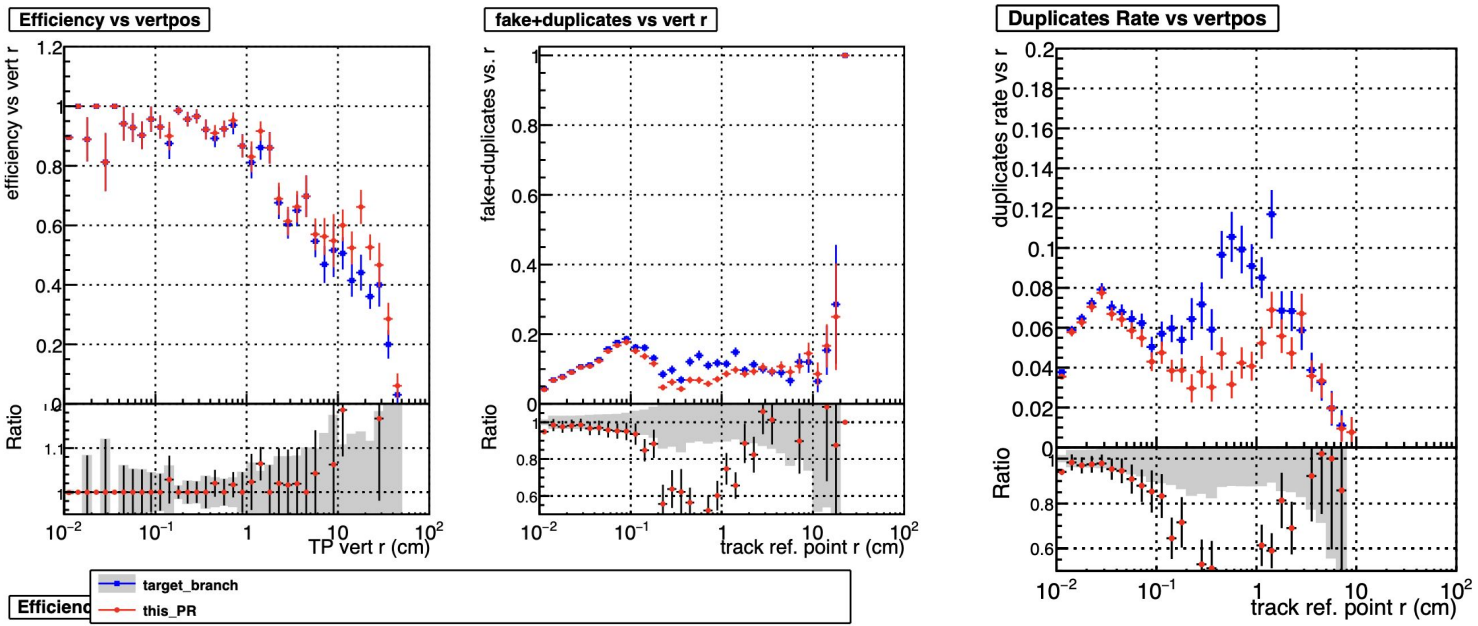

Efficiency of Track Candidate
Sample:PU200 Version tag:3273c0 N_evt:175
|η| < 4.5, |Vtx_z| < 30 cm, |Vtx_xy| < 2.5 cm, Particle:All, Charge:All
CMS Simulation
Master
Updated
Efficiency
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
10^-1
1
10
p_T [GeV]


Efficiency of Track Candidate
Sample:PU200 Version tag:3273c0 N_evt:175
|η| < 4.5, |Vtx_z| < 30 cm, |Vtx_xy| < 2.5 cm, Particle:All, Charge:All
CMS Simulation
Efficiency
TC
pT5
pT3
T5
pLS
1
0.8
0.6
0.4
0.2
0
10^-1
1
10
p_T [GeV]


Efficiency of Track Candidate
Sample:PU200 Version tag:8ede27 N_evt:175
|η| < 4.5, |Vtx_z| < 30 cm, |Vtx_xy| < 2.5 cm, Particle:All, Charge:All
CMS Simulation
Efficiency
TC
pT5
pT3
T5
pLS
1
0.8
0.6
0.4
0.2
0
10^-1
1
10
p_T [GeV]

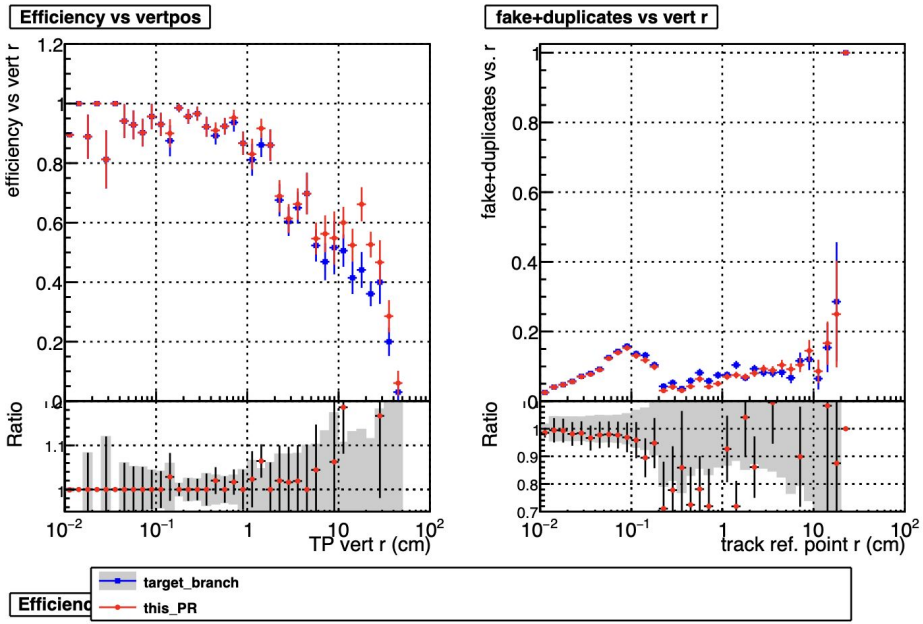
Initial Performance Results, Displaced (highPtTripletStep)

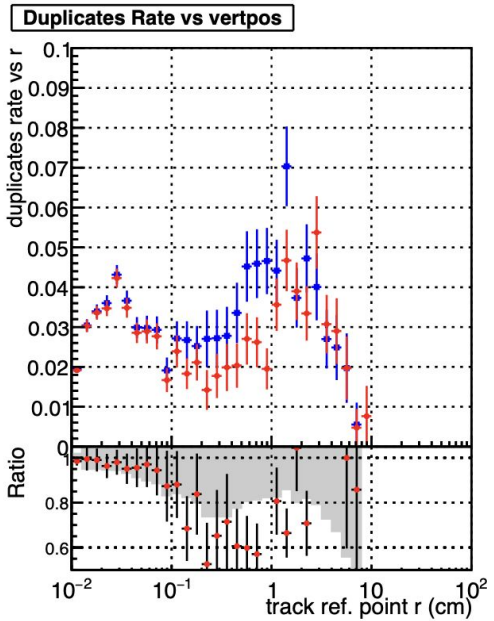
- Increase in displaced track efficiency, decrease in duplicate rate.



Initial Performance Results, Displaced (Loose ID)

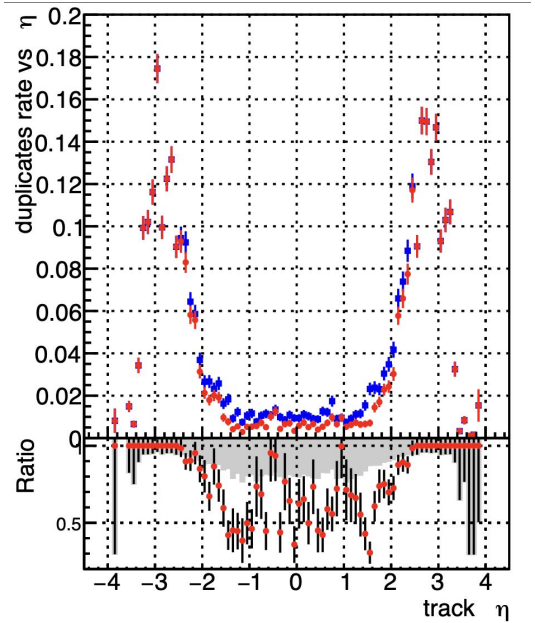
- Increase in displaced track efficiency, decrease in duplicate rate.



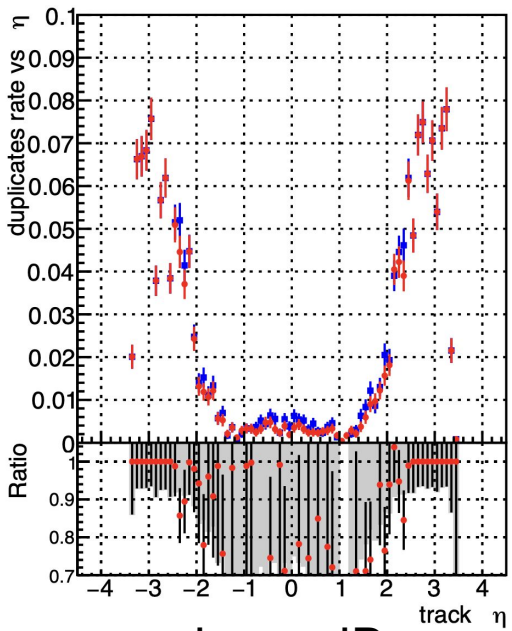


Initial Performance Results, Duplicate Rate

- Duplicate rate reduced by ~50% in barrel for highPtTripletStep, and reduced by ~30% in Loose ID configuration.



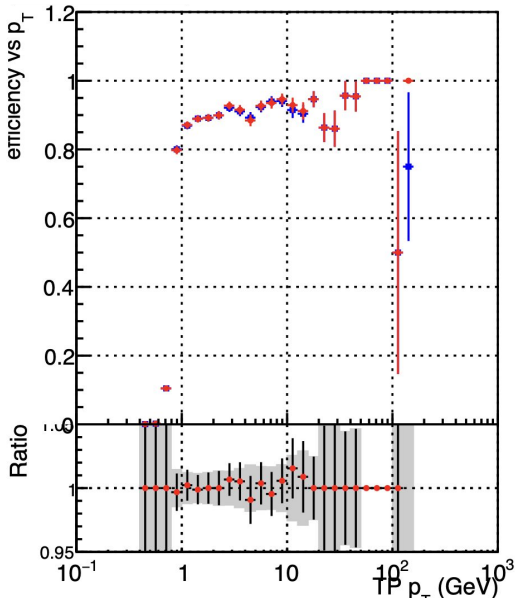
highPtTripletStep

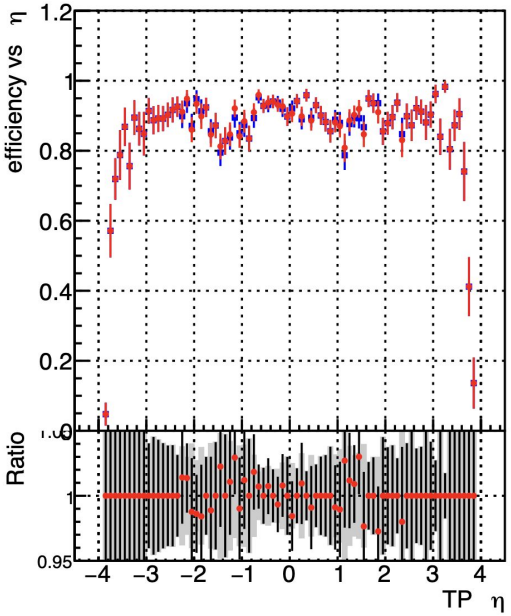


Loose ID

Initial Performance Results, Efficiency

- Minimal changes (slight increase?) in efficiency, similar to standalone plots.
- Plots below are identical for highPtTripletStep and Loose ID configurations.





Very Preliminary Timing

- Need to make some simple code improvements still, so timing is suboptimal, but even with the current implementation the differences are small.

Current Timing

```
Total Timing Summary
Average time for map loading = 327.175 ms
Average time for input loading = 6644.31 ms
Average time for lst::Event creation = 0.000613077 ms
   Evt    Hits       MD       LS       T3       T5       pLS      pT5      pT3      TC       Reset    Event    Short            Rate
   avg       0.6      0.3      0.2      0.4      0.4      0.2      0.5      0.2      0.6      0.0      3.4      2.7+/-  0.6      3.5     explicit[s=1]
   avg       0.7      0.5      0.3      0.6      0.5      0.2      0.7      0.3      0.8      0.0      4.5      3.6+/-  0.8      2.3     explicit[s=2]
   avg       1.0      0.7      0.5      0.8      0.8      0.3      1.1      0.5      1.2      0.0      6.8      5.5+/-  1.4      1.7     explicit[s=4]
   avg       1.3      0.9      0.8      1.1      1.1      0.4      1.6      0.6      1.6      0.0      9.5      7.8+/-  3.0      1.6     explicit[s=6]
   avg       1.6      1.2      1.0      1.5      1.5      0.5      2.2      0.9      2.0      0.0      12.5     10.3+/-  3.8      1.6     explicit[s=8]
```

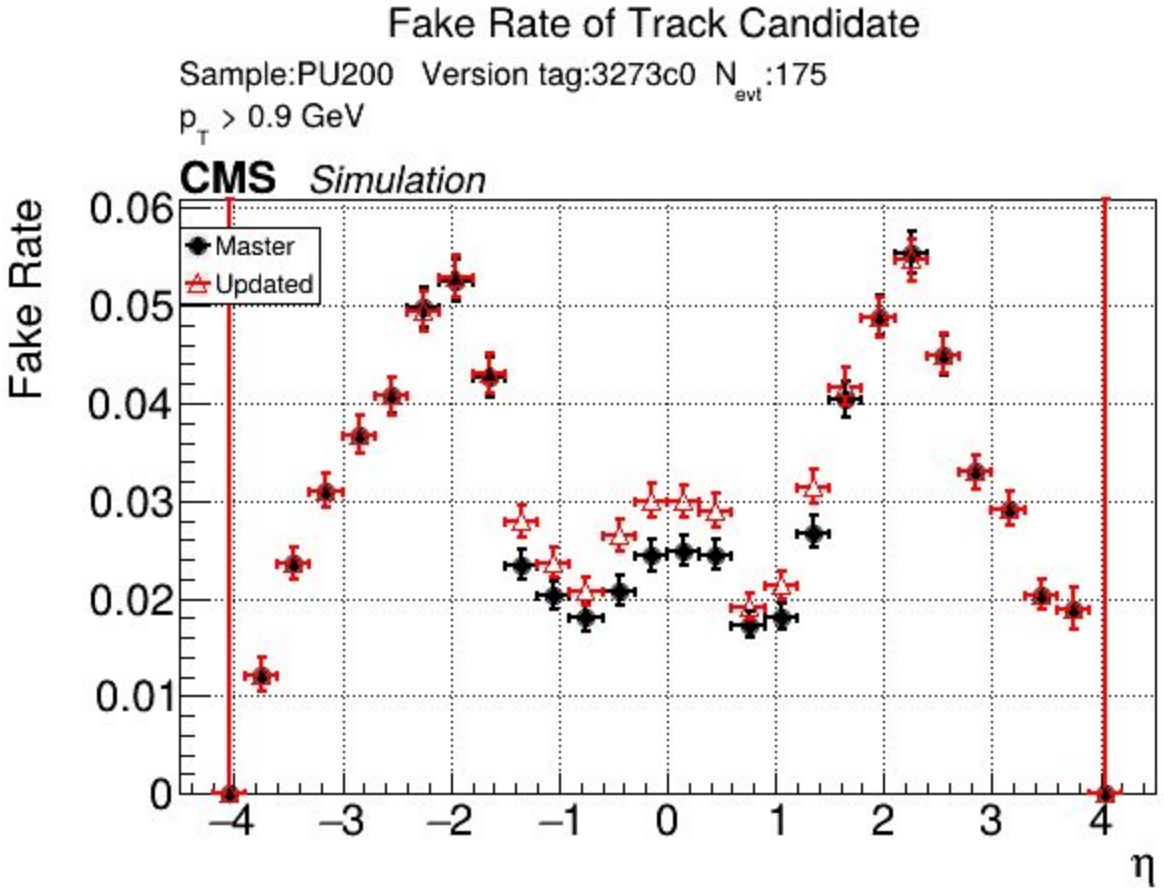
This PR Current Timing

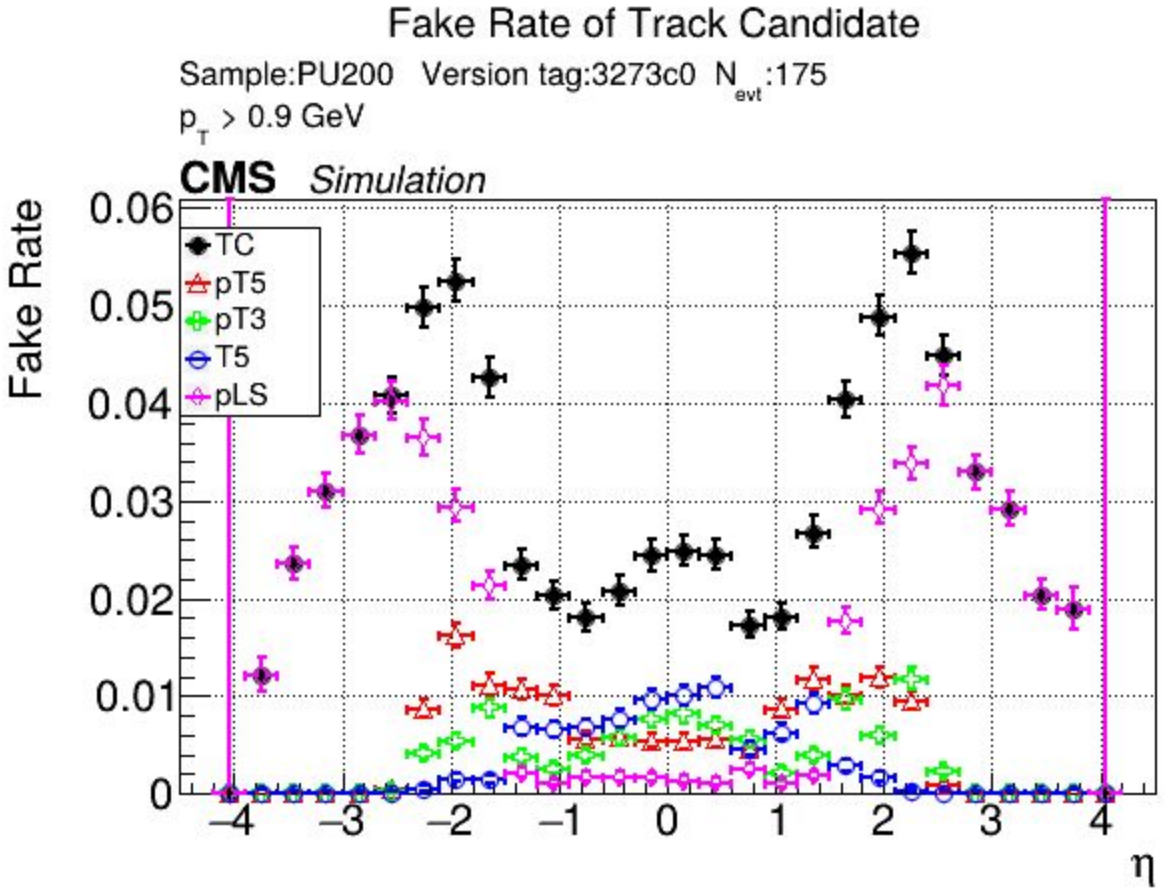
```
Total Timing Summary
Average time for map loading = 324.45 ms
Average time for input loading = 6661.93 ms
Average time for lst::Event creation = 0.000579016 ms
   Evt    Hits       MD       LS       T3       T5       pLS      pT5      pT3      TC       Reset    Event    Short            Rate
   avg       0.6      0.3      0.2      0.4      0.5      0.2      0.5      0.2      0.6      0.0      3.5      2.8+/-  0.6      3.6     explicit[s=1]
   avg       0.7      0.5      0.3      0.6      0.6      0.2      0.7      0.3      0.8      0.0      4.6      3.7+/-  0.8      2.3     explicit[s=2]
   avg       1.0      0.7      0.6      0.9      0.9      0.3      1.1      0.5      1.2      0.0      7.1      5.8+/-  1.5      1.8     explicit[s=4]
   avg       1.3      0.9      0.8      1.2      1.3      0.4      1.7      0.6      1.7      0.0      10.0     8.3+/-  2.9      1.7     explicit[s=6]
   avg       1.7      1.3      1.1      1.6      1.7      0.6      2.2      0.9      2.2      0.0      13.4     11.1+/-  4.2      1.7     explicit[s=8]
```

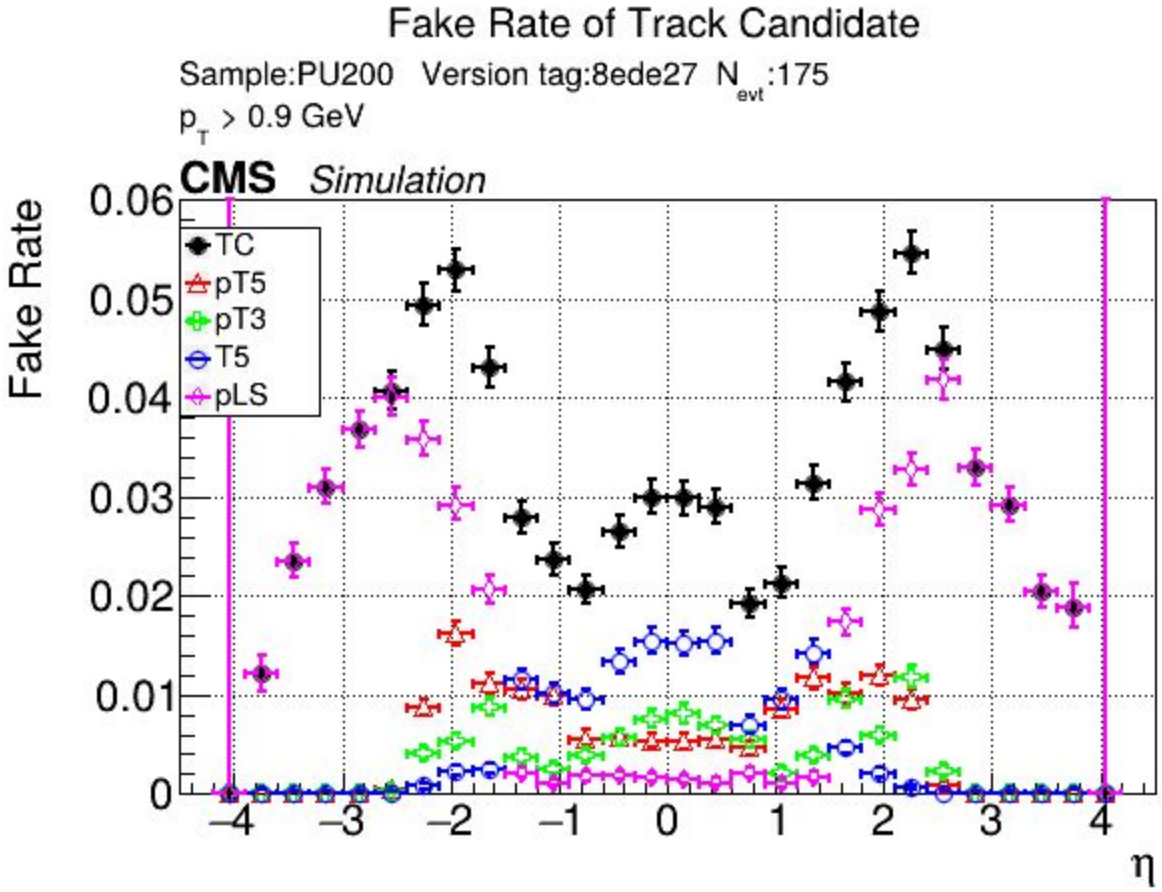
Summary + Next Steps

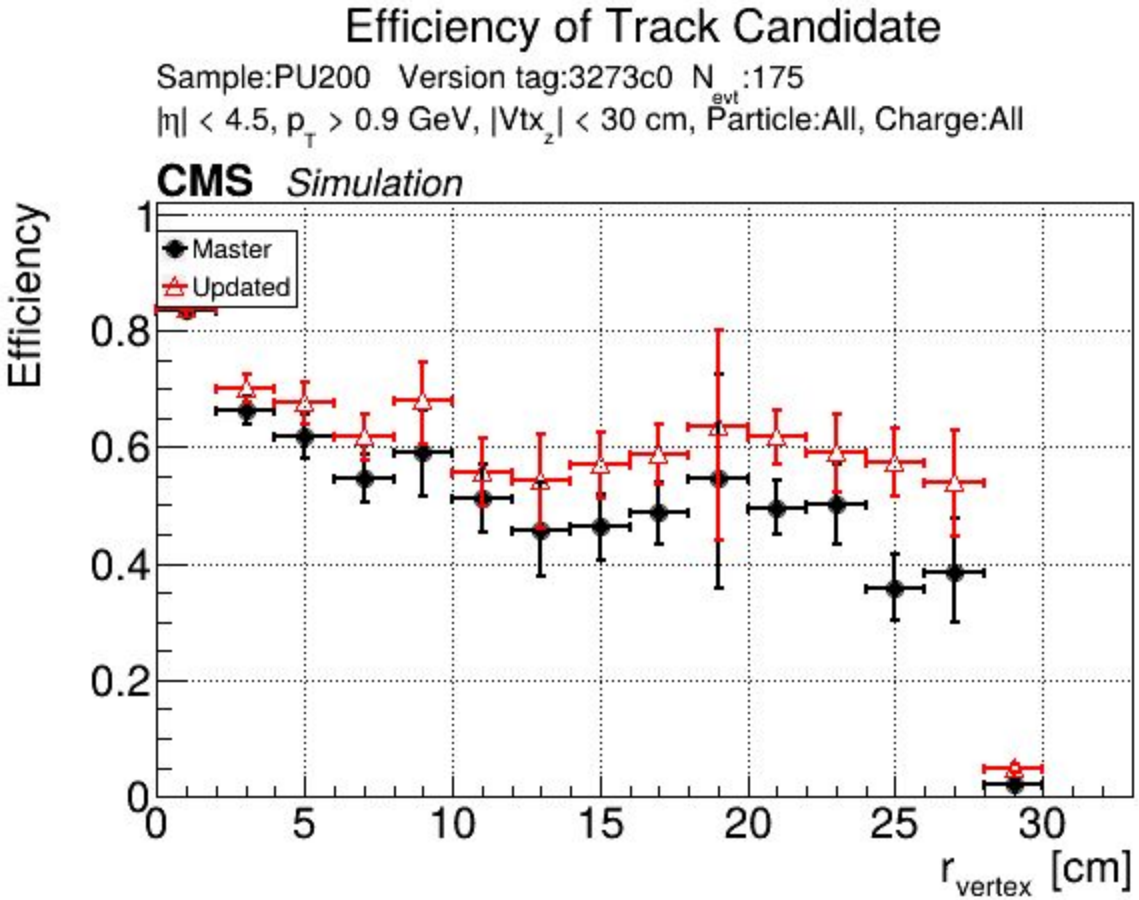
- Reduced the duplicate rate of LST in the barrel by up to 50% with a slight increase in overall efficiency by replacing the current duplicate removal for pLS-T5 pairs with a cut on the embedding distance.
- Increased the displaced track efficiency significantly from the additional embedding cuts on T5-T5 and pT5-T5 pairs, and reduced the duplicate rate of displaced tracks significantly with improved pLS-T5 cleaning.
- Plan to continue optimizing the current embedding DNN's and add pLS-pLS pairs to training set to replace remaining duplicate removal steps.

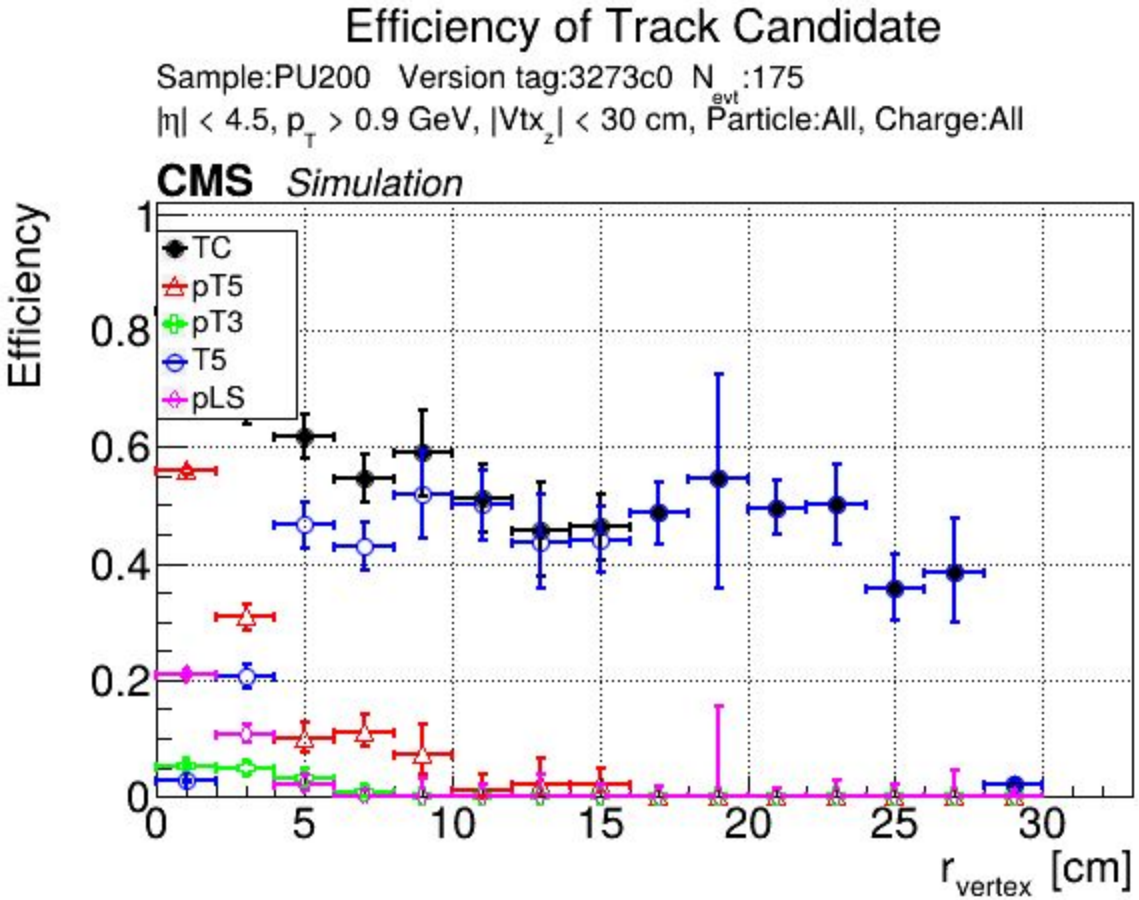
Backup

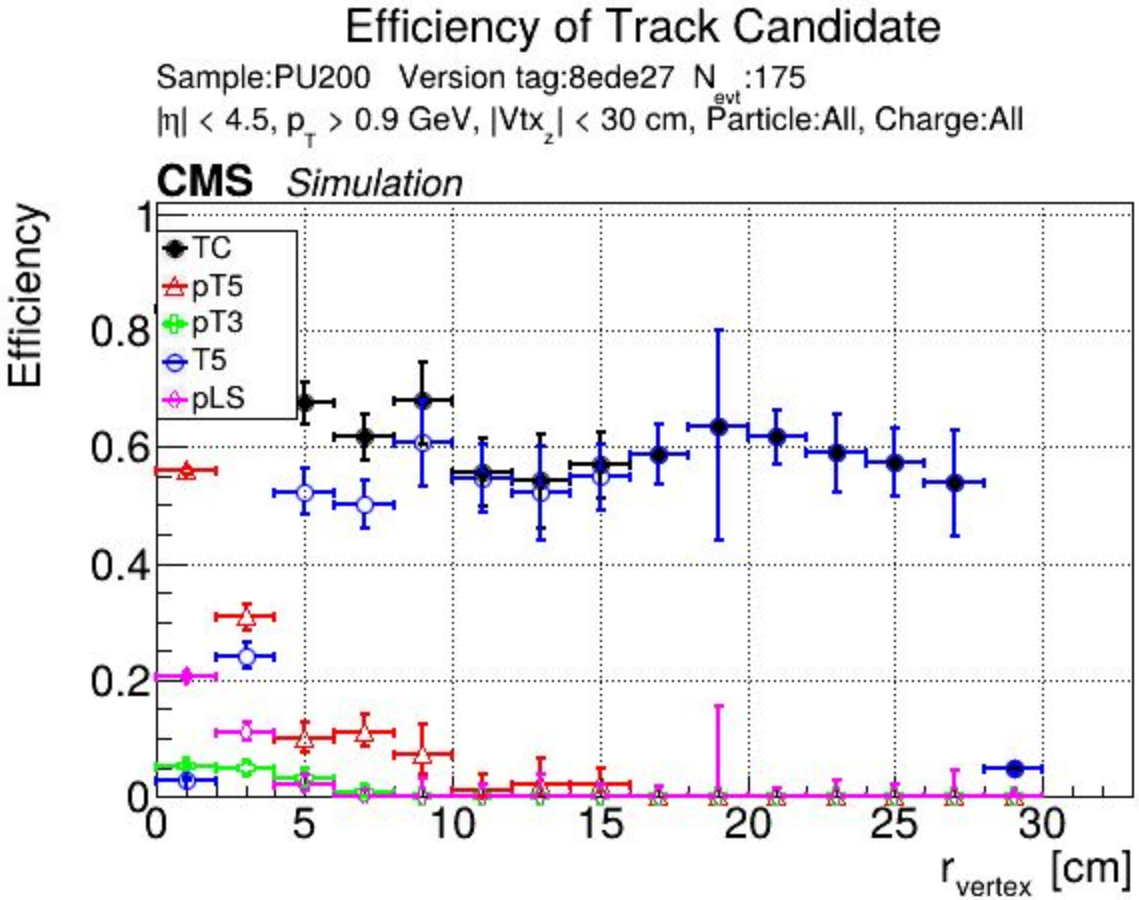

Fake Rate of Track Candidate
Sample:PU200 Version tag:3273c0 N_evt:175
p_T > 0.9 GeV
CMS Simulation
Fake Rate
Master
Updated
0.06
0.05
0.04
0.03
0.02
0.01
0
−4
−3
−2
−1
0
1
2
3
4
η

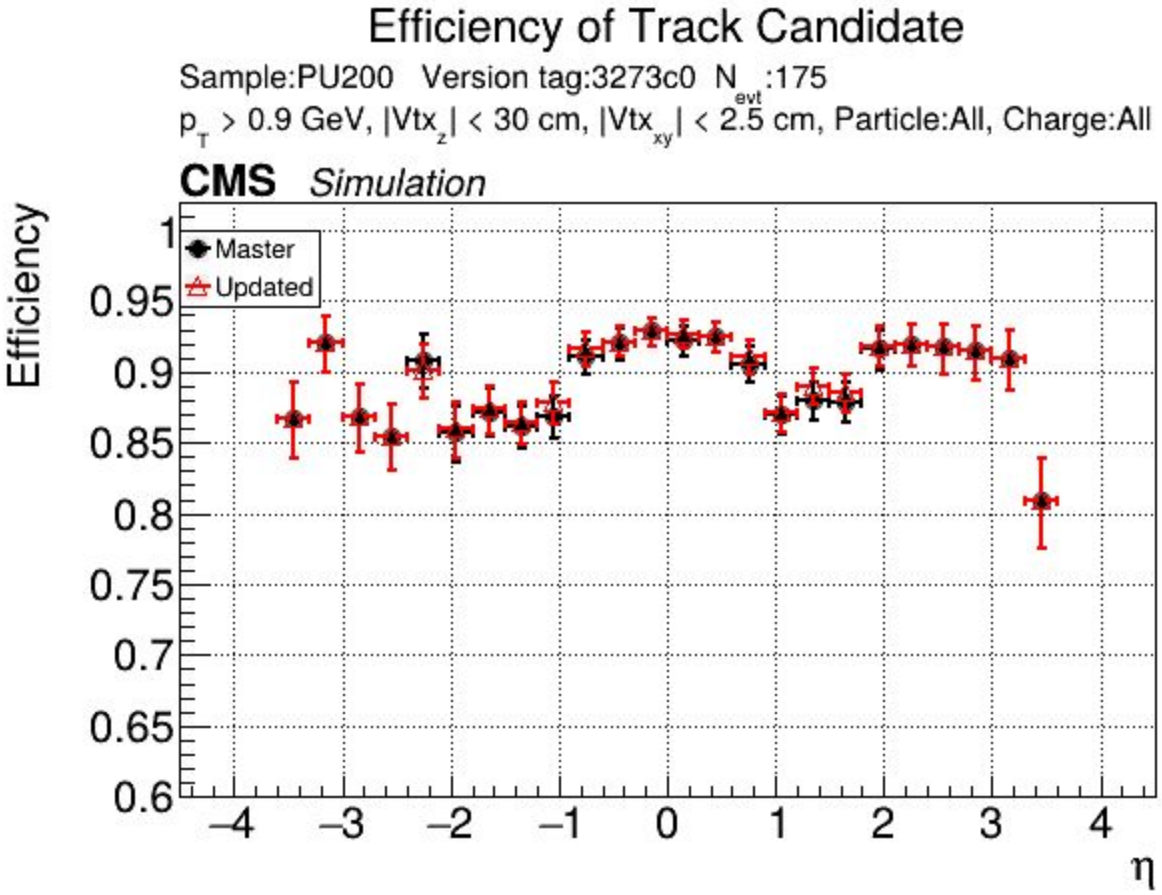

Fake Rate of Track Candidate
Sample:PU200 Version tag:3273c0 N_evt:175
p_T > 0.9 GeV
CMS Simulation
Fake Rate
0.06
0.05
0.04
0.03
0.02
0.01
0
TC
pT5
pT3
T5
pLS
−4
−3
−2
−1
0
1
2
3
4
η

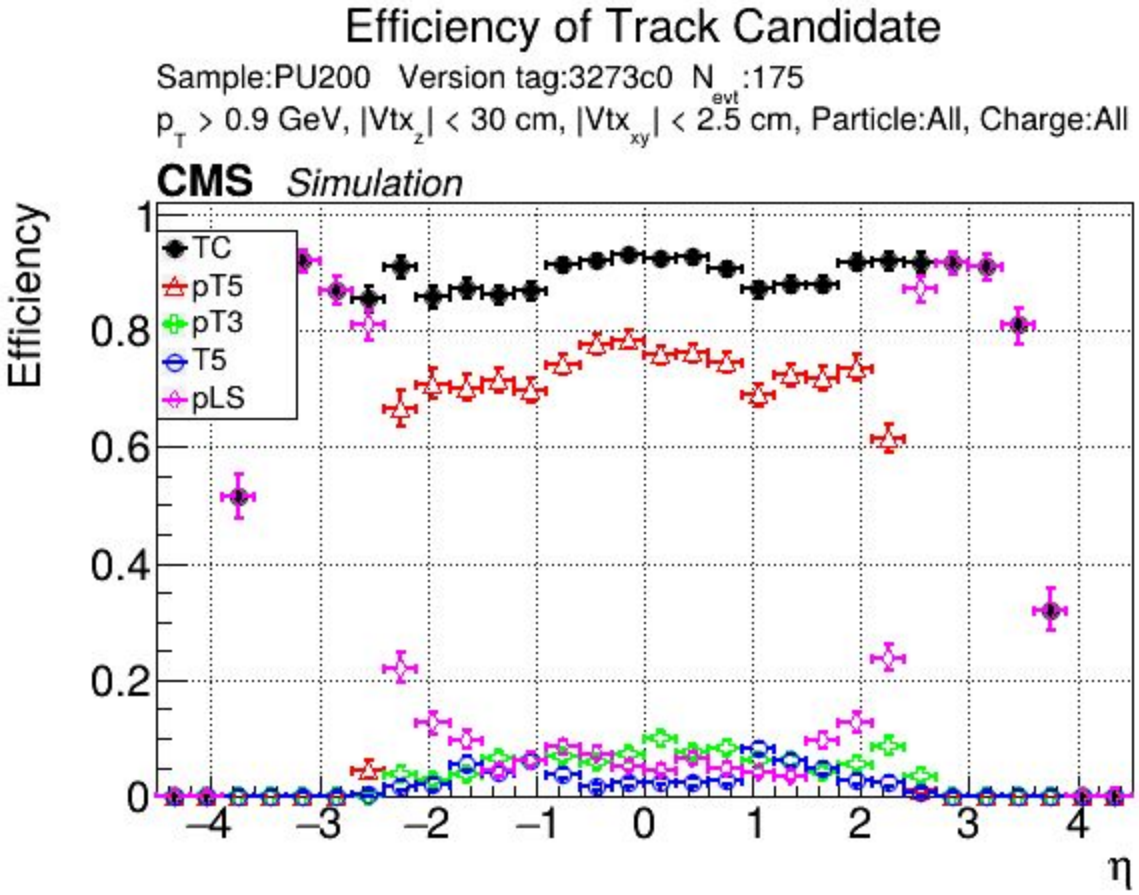

Fake Rate of Track Candidate
Sample:PU200 Version tag:8ede27 N_evt:175
p_T > 0.9 GeV
CMS Simulation
Fake Rate
0.06
0.05
0.04
0.03
0.02
0.01
0
TC
pT5
pT3
T5
pLS
−4
−3
−2
−1
0
1
2
3
4
η

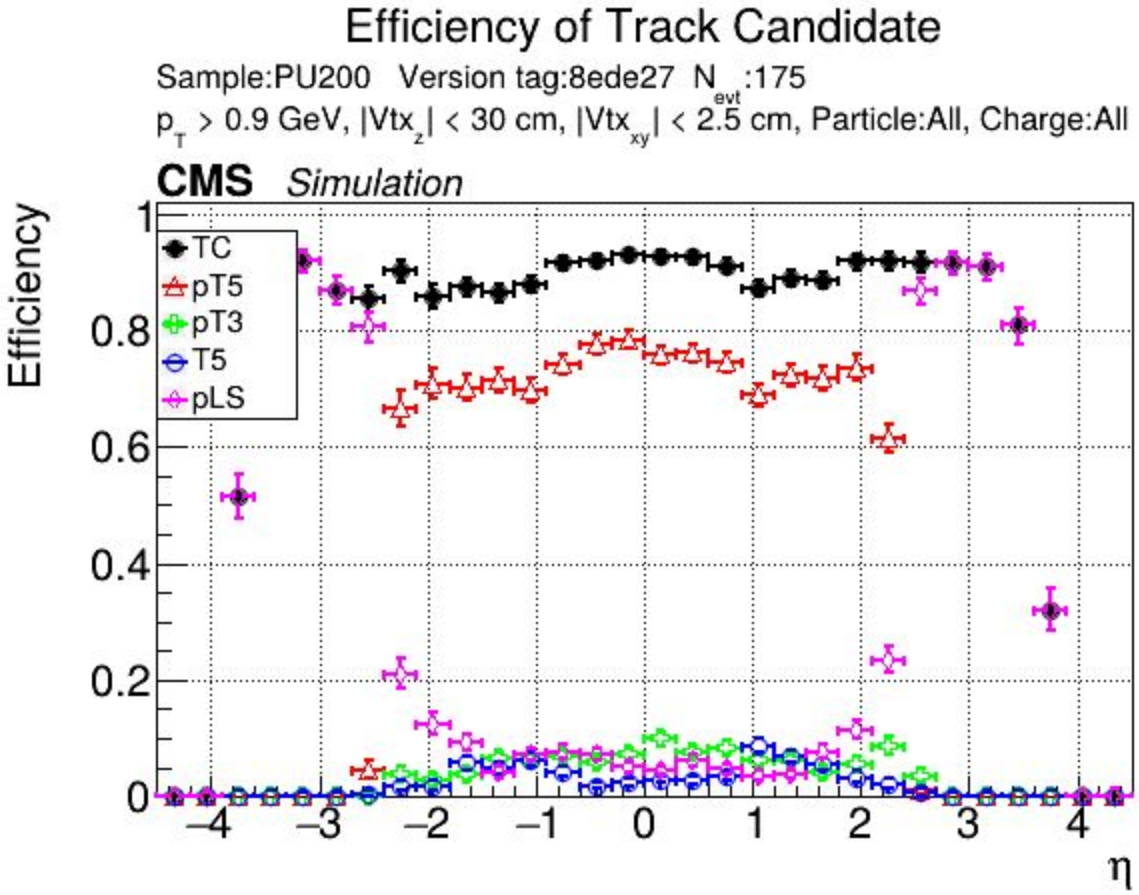

Efficiency of Track Candidate
Sample:PU200 Version tag:3273c0 N_evt:175
|η| < 4.5, p_T > 0.9 GeV, |Vtx_z| < 30 cm, Particle:All, Charge:All
CMS Simulation
Master
Updated
Efficiency
0
0.2
0.4
0.6
0.8
1
0
5
10
15
20
25
30
r_vertex [cm]


Efficiency of Track Candidate
Sample:PU200 Version tag:3273c0 N_evt:175
|η| < 4.5, p_T > 0.9 GeV, |Vtx_z| < 30 cm, Particle:All, Charge:All
CMS Simulation
Efficiency
TC
pT5
pT3
T5
pLS
0
0.2
0.4
0.6
0.8
1
0
5
10
15
20
25
30
r_vertex [cm]


Efficiency of Track Candidate
Sample:PU200 Version tag:8ede27 N_evt:175
|η| < 4.5, p_T > 0.9 GeV, |Vtx_z| < 30 cm, Particle:All, Charge:All
CMS Simulation
Efficiency
TC
pT5
pT3
T5
pLS
0
0.2
0.4
0.6
0.8
1
0
5
10
15
20
25
30
r_vertex [cm]


Efficiency of Track Candidate
Sample:PU200 Version tag:3273c0 N_evt:175
p_T > 0.9 GeV, |Vtx_z| < 30 cm, |Vtx_xy| < 2.5 cm, Particle:All, Charge:All
CMS Simulation
Master
Updated
Efficiency
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
−4
−3
−2
−1
0
1
2
3
4
η


Efficiency of Track Candidate
Sample:PU200 Version tag:3273c0 N_evt:175
p_T > 0.9 GeV, |Vtx_z| < 30 cm, |Vtx_xy| < 2.5 cm, Particle:All, Charge:All
CMS Simulation
Efficiency
TC
pT5
pT3
T5
pLS
1
0.8
0.6
0.4
0.2
0
−4
−3
−2
−1
0
1
2
3
4
η


Efficiency of Track Candidate
Sample:PU200 Version tag:8ede27 N_evt:175
p_T > 0.9 GeV, |Vtx_z| < 30 cm, |Vtx_xy| < 2.5 cm, Particle:All, Charge:All
CMS Simulation
Efficiency
TC
pT5
pT3
T5
pLS
1
0.8
0.6
0.4
0.2
0
−4
−3
−2
−1
0
1
2
3
4
η