GNN LST

Propagating the “cherry-on-top” GNN to LST efficiency plots

April 4th, 2023

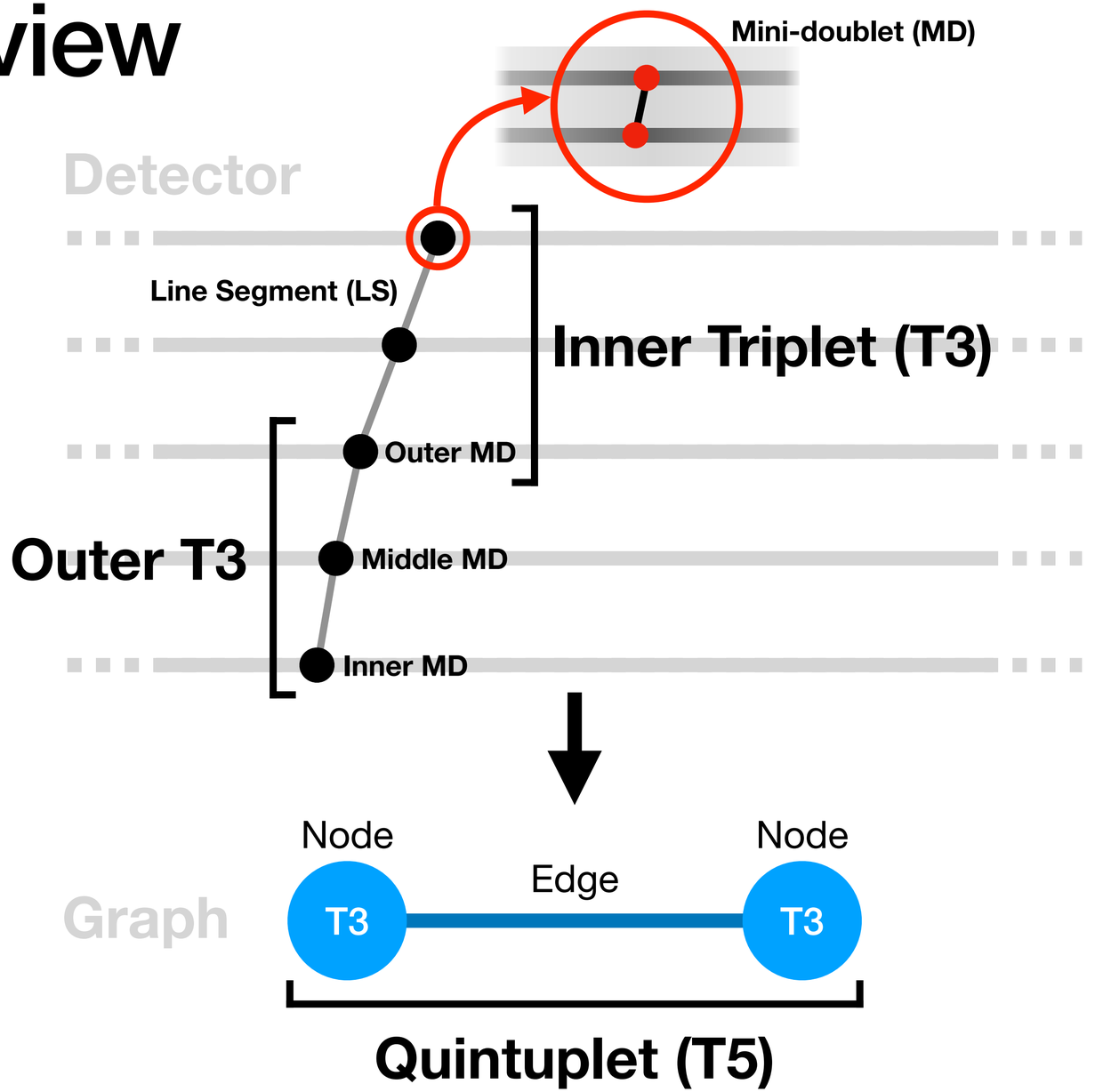
P. Chang, ***J. Guiang***

UC San Diego

UF UNIVERSITY of FLORIDA

Overview

- Exploring GNN for T5 classification
  - T5s have highest fake rate
  - Maybe GNN can improve over Balaji's χ^2 "magic" cuts
- In the graph:
  - T3s become "nodes"
  - T5 candidates become "edges"

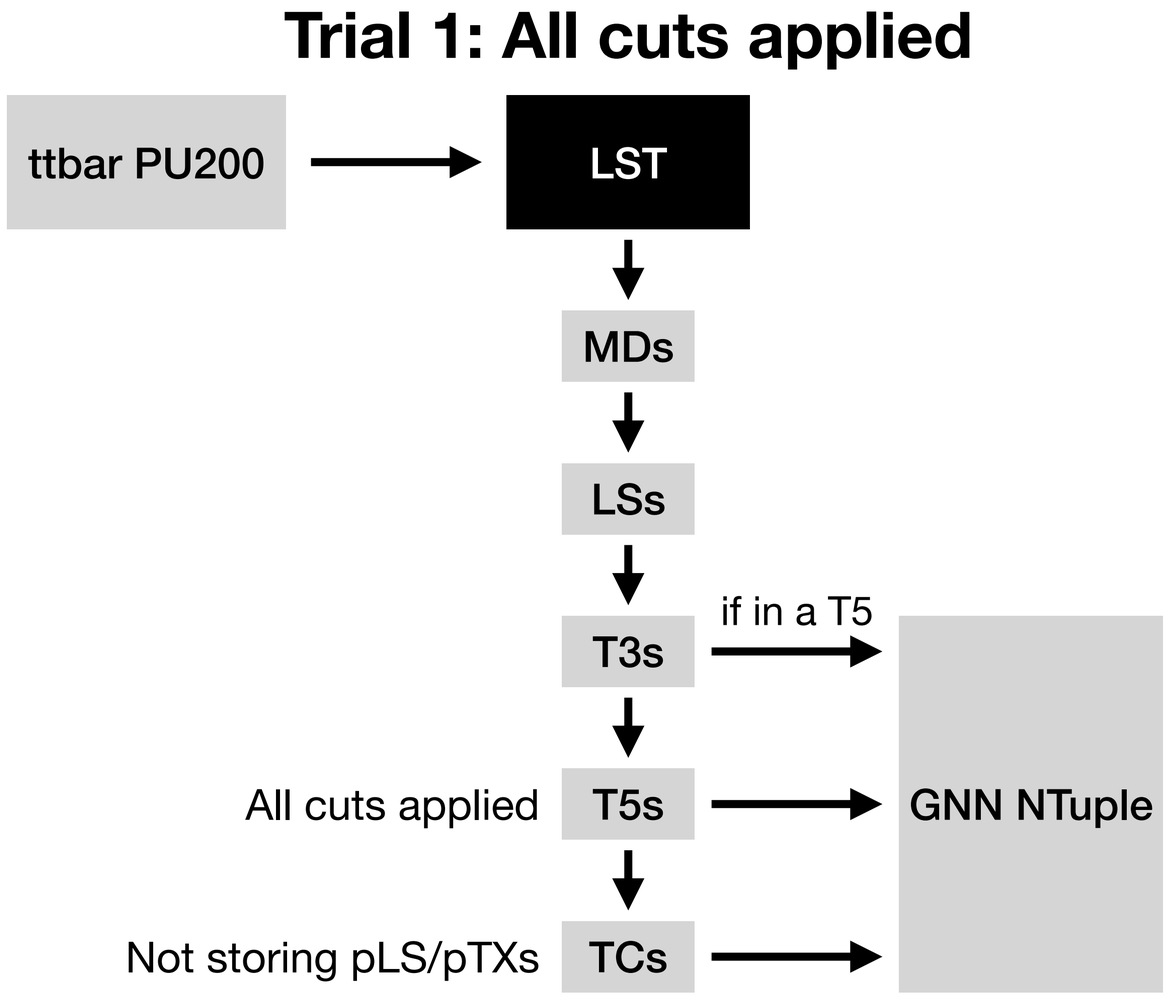


UC San Diego

Trial 1: GNN NTuple

- Training on T5s/T3s without duplicate removal
- pLS/pTX objects are excluded from the TCs
  - i.e. only T5s in TCs
  - Duplicate removal is applied here
- **All cuts applied in T5 selection algo.**
- Target question:

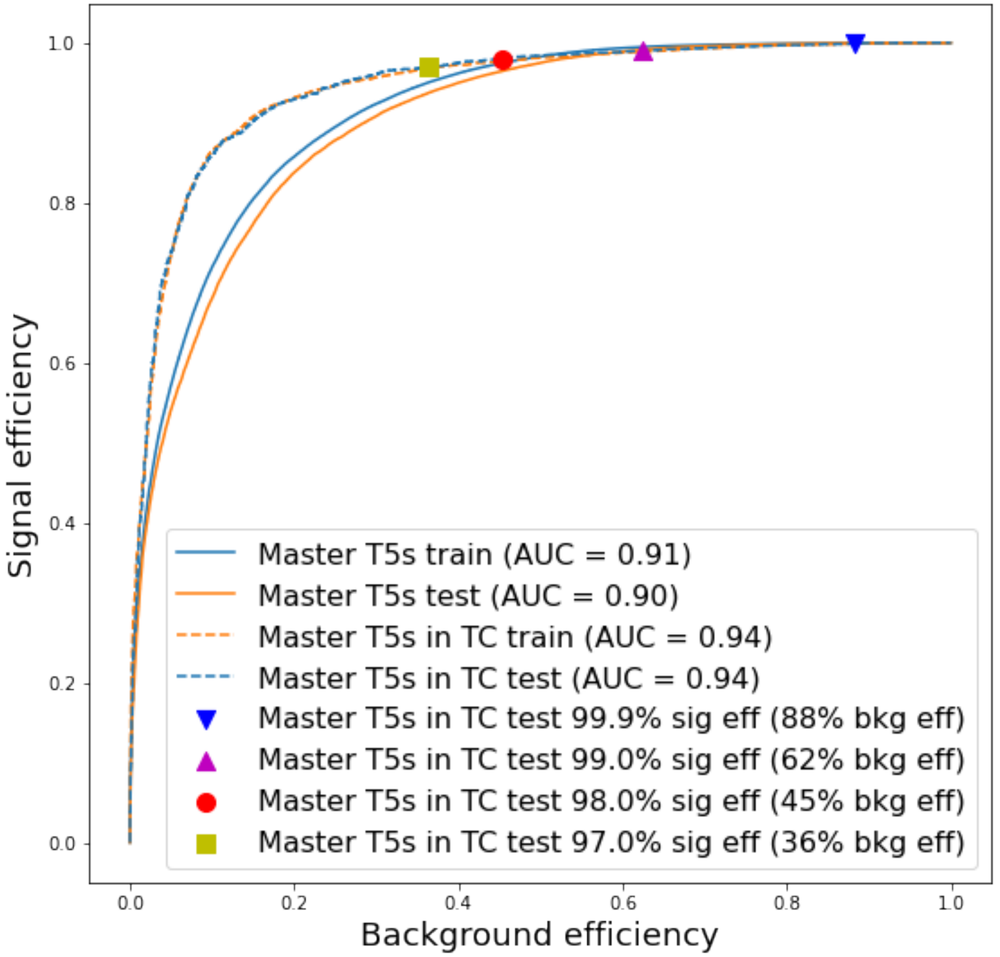
Can the GNN give us anything for free?



Trial 1: GNN Performance

ROC for LST master T5s, T5s in TC → possibly large fake reduction!

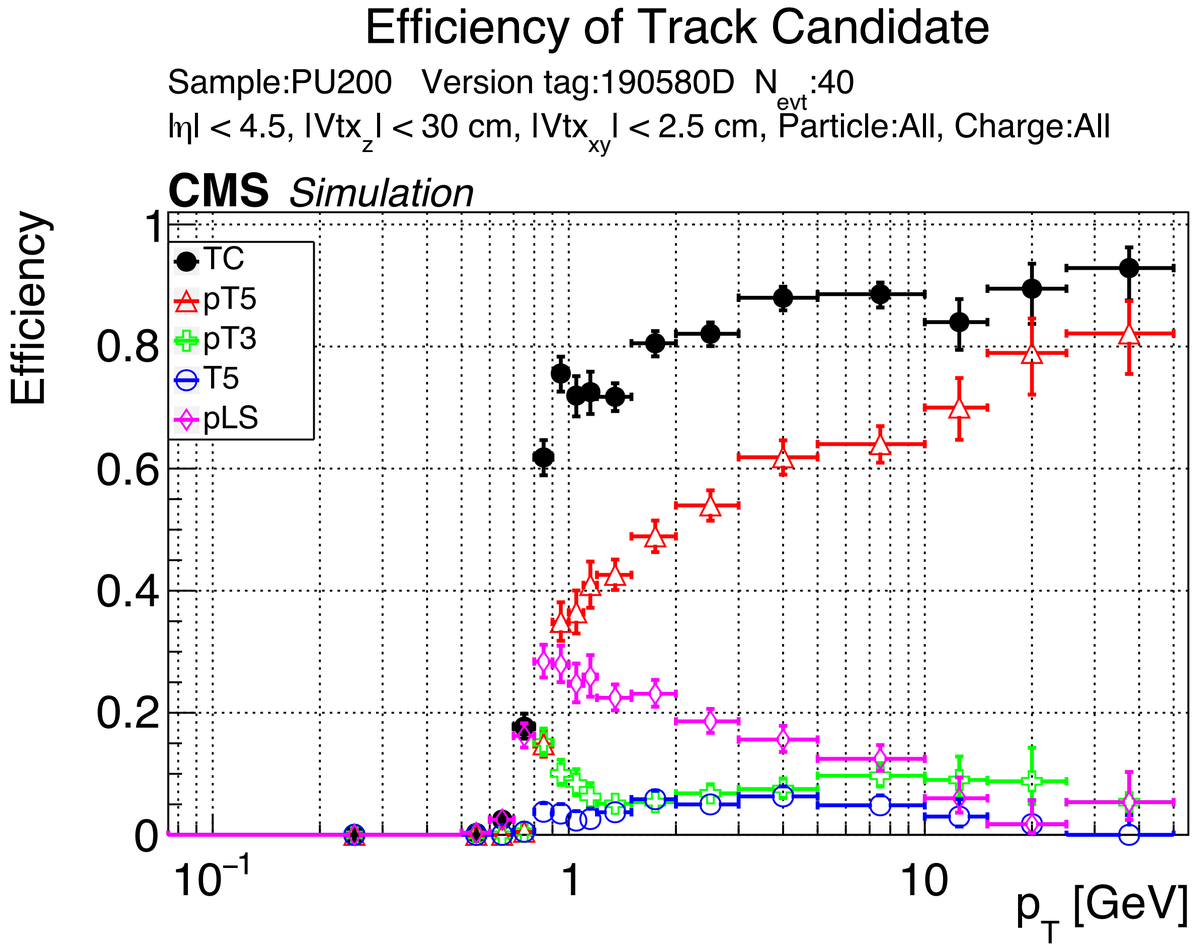
- Solid curves show performance immediately after making χ^2 cuts
- Dashed curves show performance on only T5s in TC collection (i.e. after duplicate removal)
  - Q: If we run inference on only final T5s, **how many fakes can we remove** (while keeping X% of reals)?
  - A: We can **reduce fake T5s by > x2** while keeping 98% of reals
- **To-do:** propagate GNN score to LST NTuple and make DP-like efficiency plots

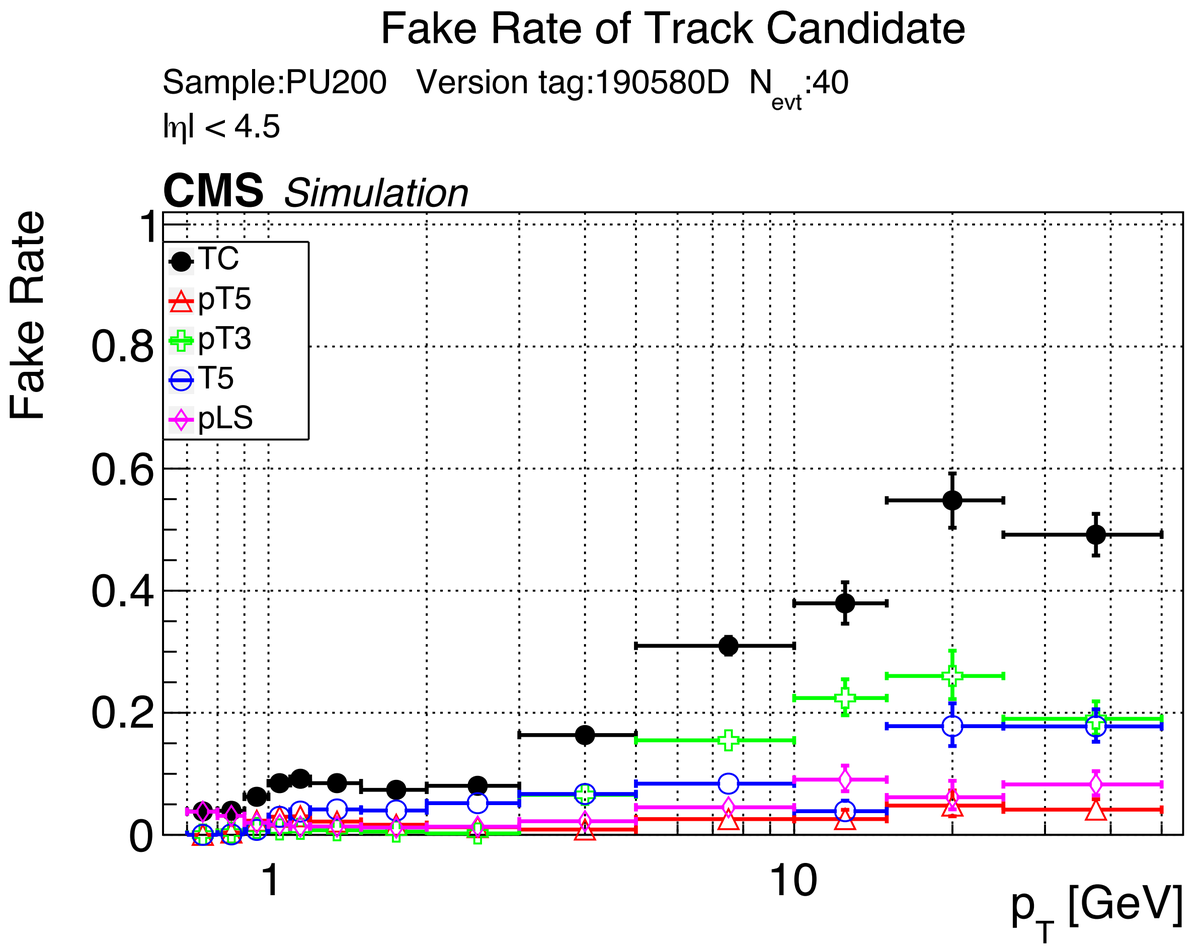


UC San Diego

LST Master Efficiency Plots

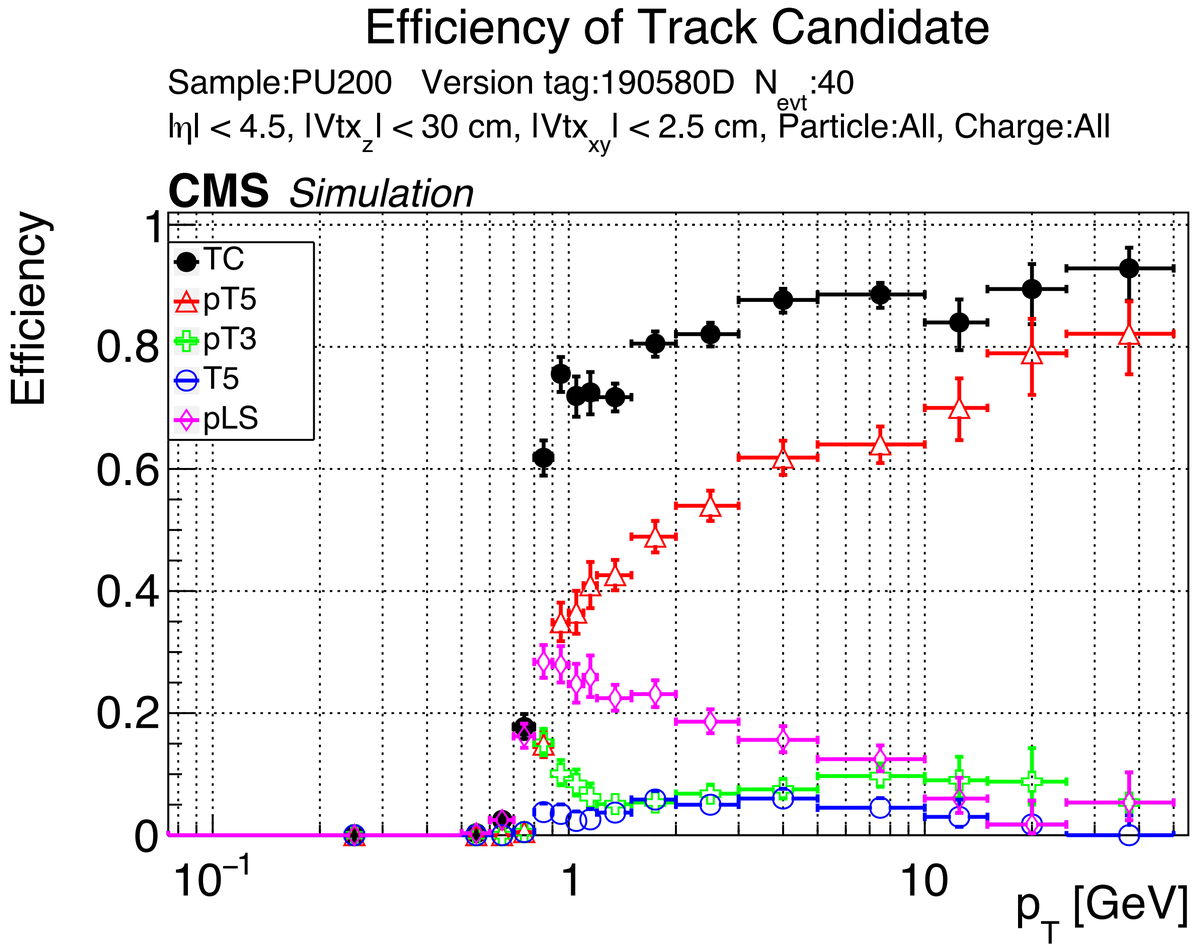
Showing p_T-binned eff. & FR for the **40 testing events**

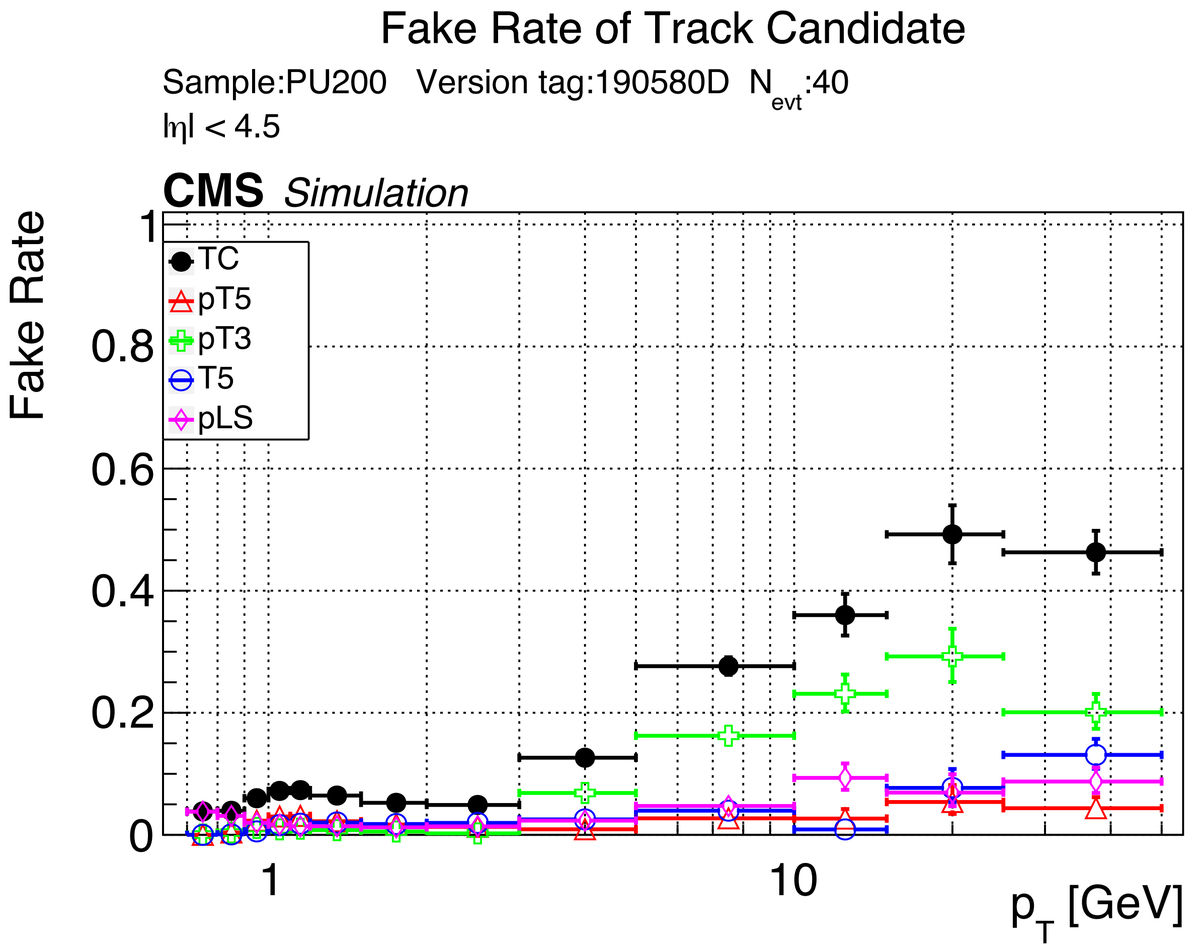




LST GNN Efficiency Plots

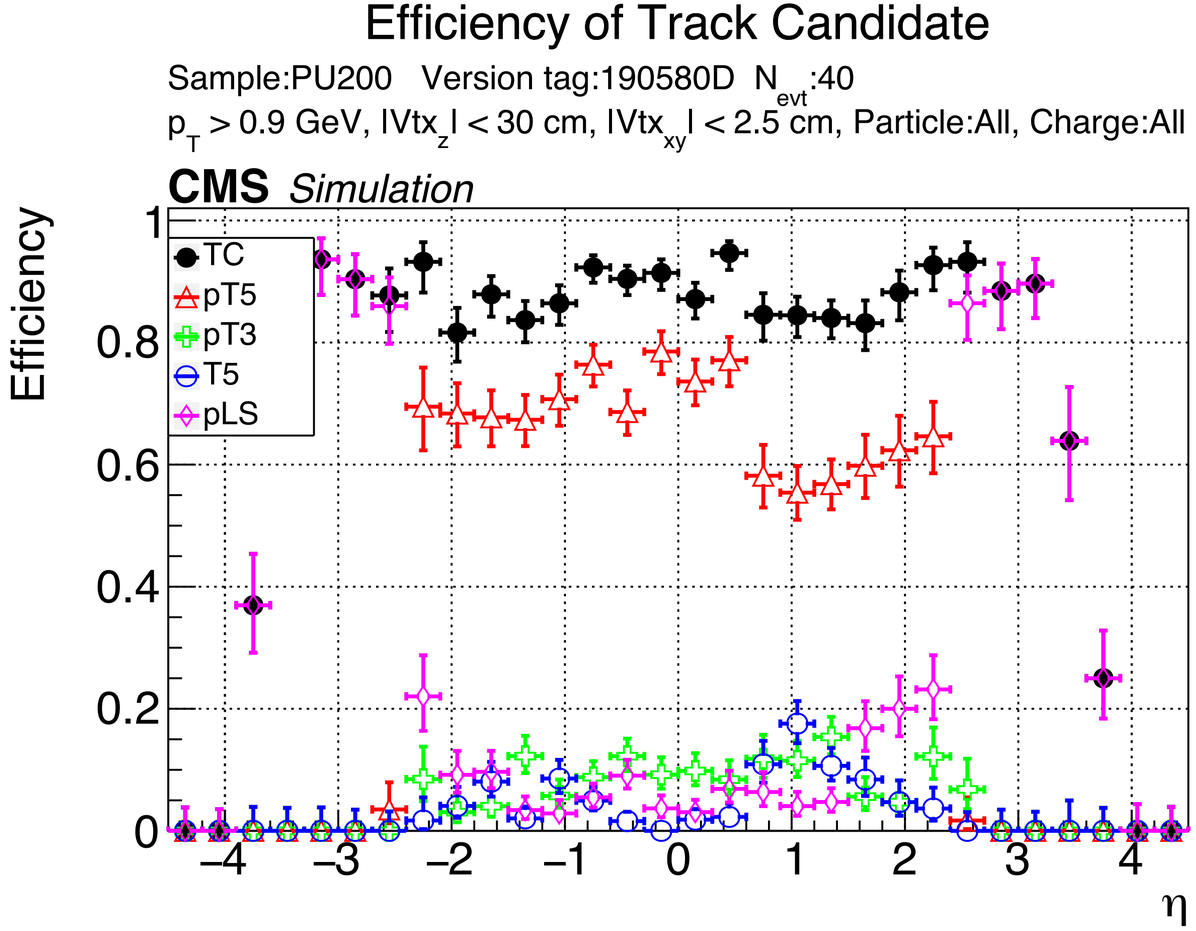
Showing p_T-binned eff & FR for the **40 testing events (98% sig eff)**

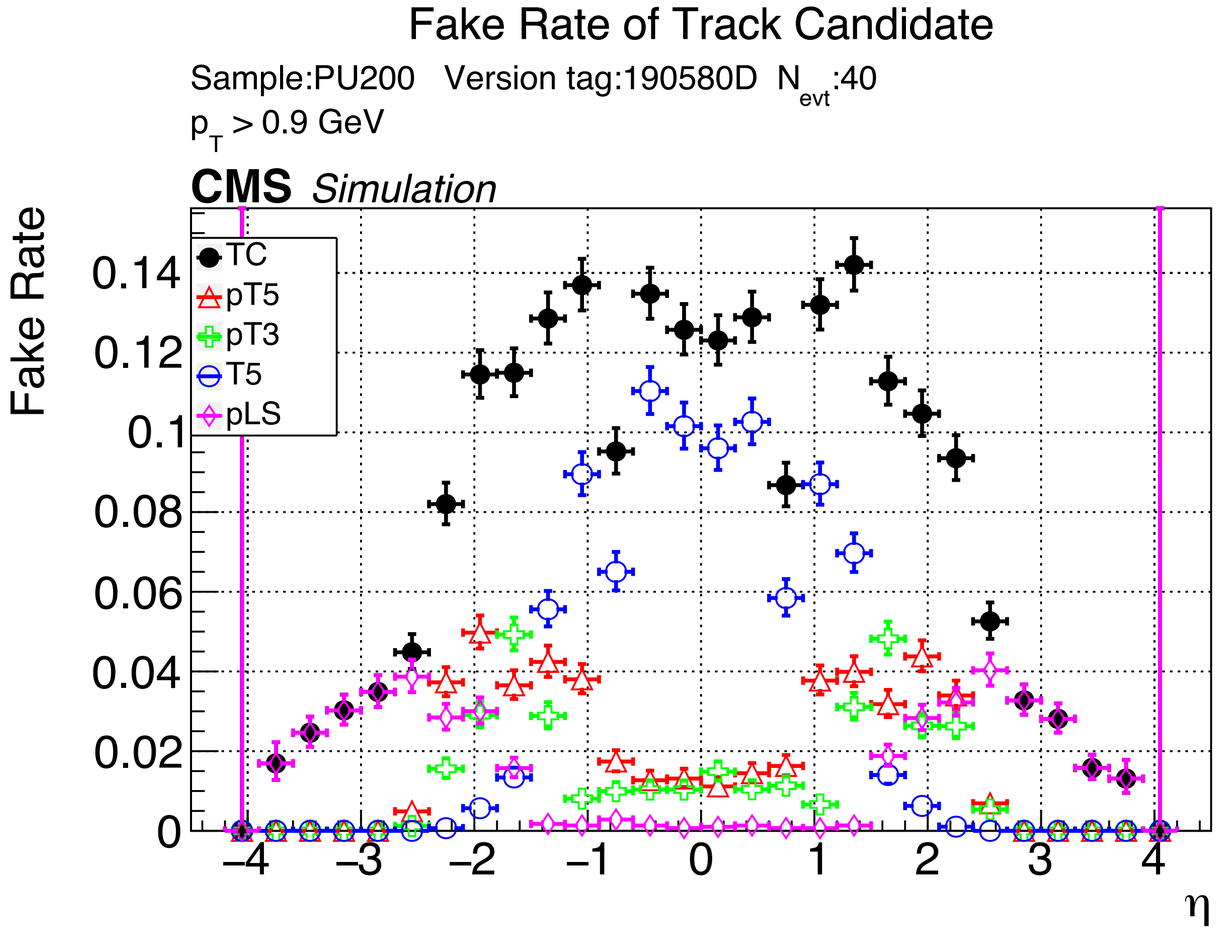




LST Master Efficiency Plots

Showing η-binned eff & FR for the **40 testing events**

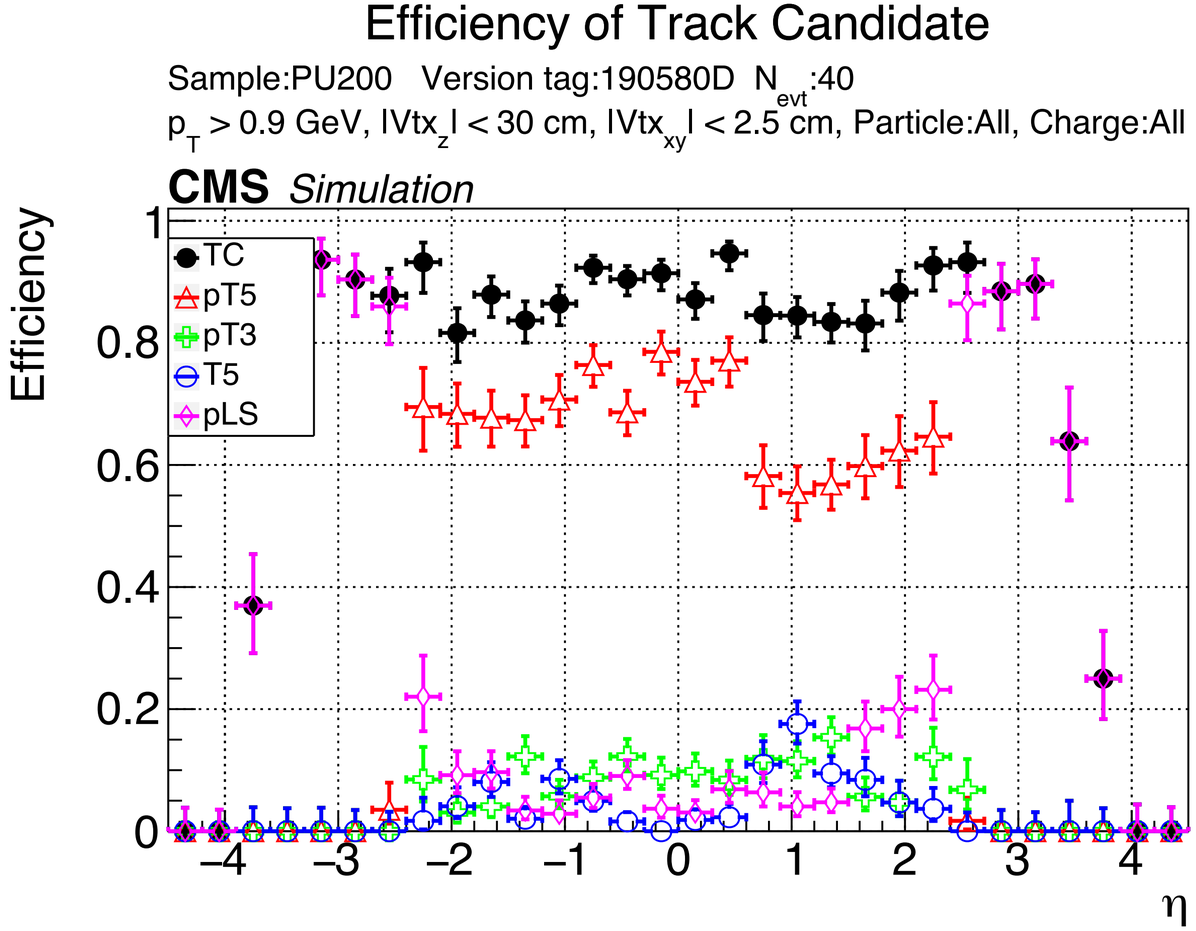


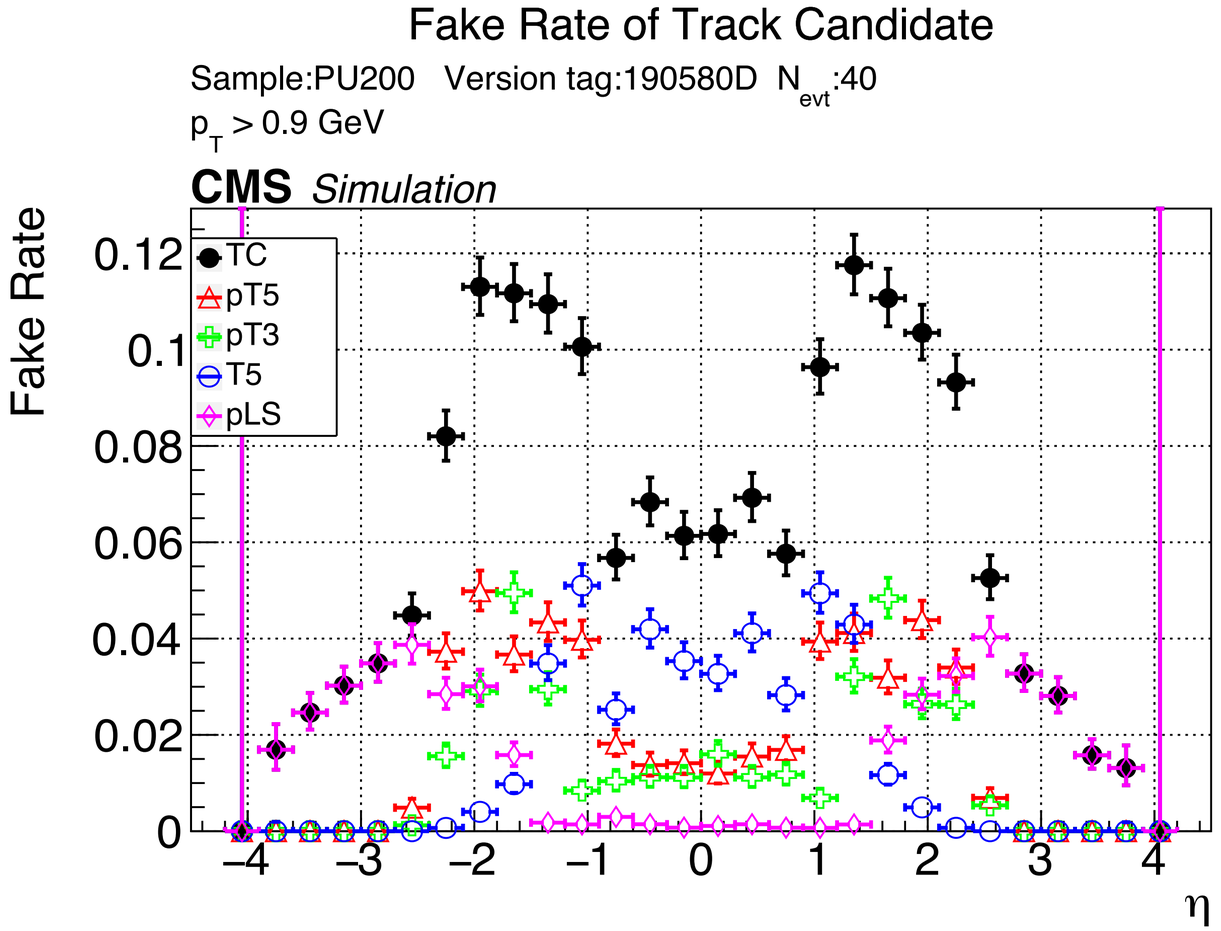


UC San Diego

LST GNN Efficiency Plots

Showing η-binned eff & FR for the **40 testing events (98% sig eff)**





UC San Diego

Summary

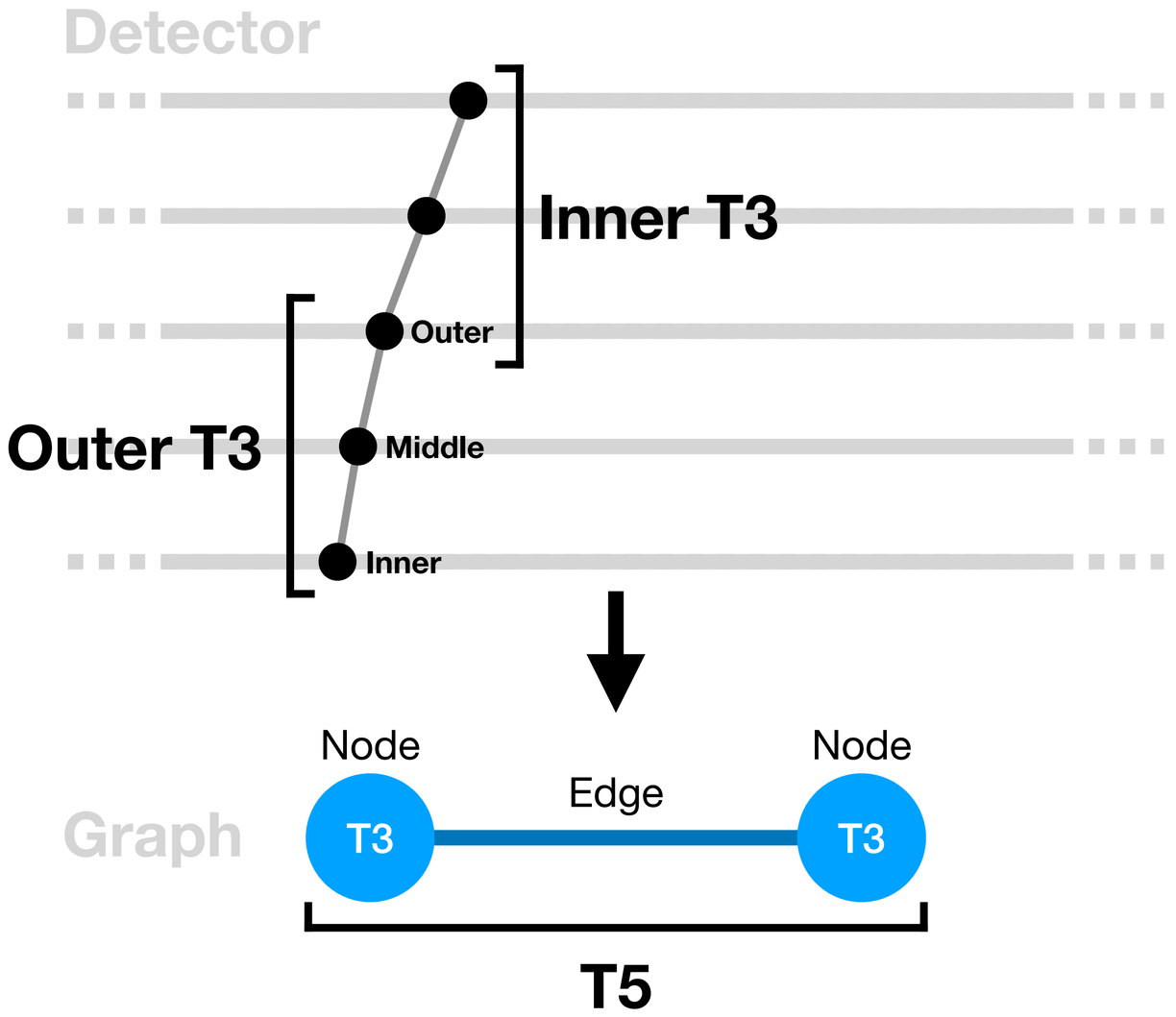
- The GNN is working:
  - “Trial 1” GNN from last week (i.e. “cherry-on-top”) indeed reduces fake T5 TCs
  - Almost no efficiency loss
- All plots available here: GNN, Master, GNN vs. Master
- Next steps:
  - Add anchor hit ϕ’s as T3 features
  - Edit efficiency code to allow us to look at the “Trial 2” efficiency

Backup

GNN Inputs

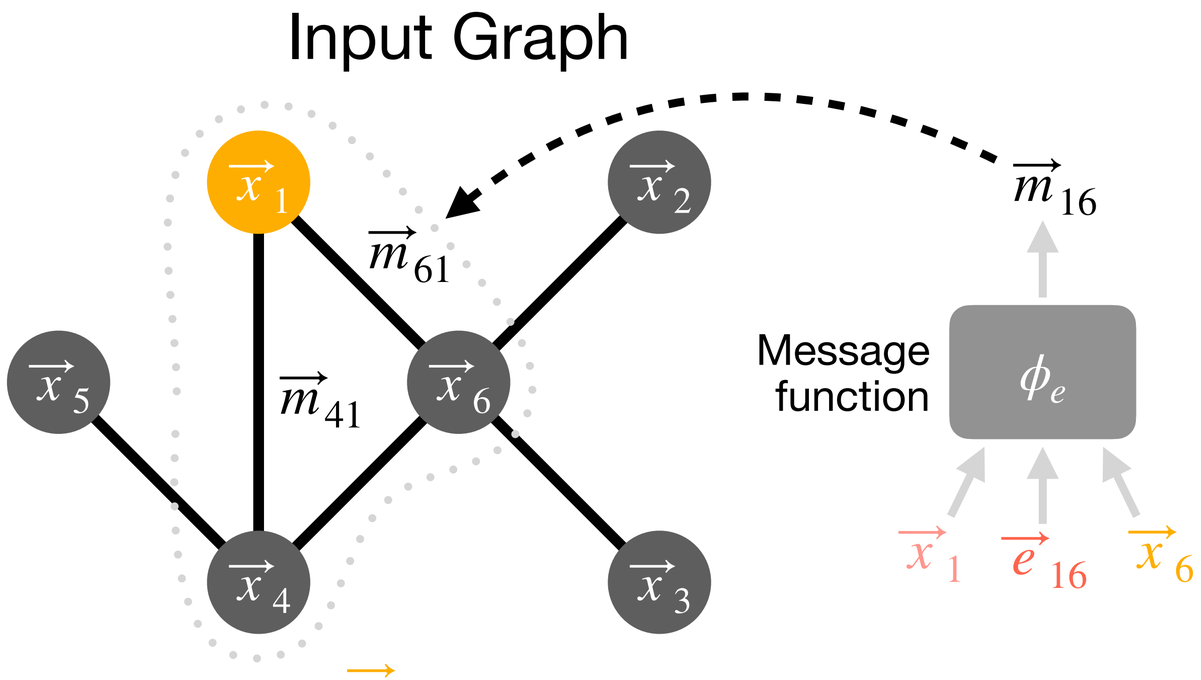
Object	Feature
Node (T3)	p_T, η, ϕ
	Inner anchor hit r, z, δr, layer
	Middle anchor hit r, z, δr, layer
	Outer anchor hit r, z, δr, layer
Edge (T5)	p_T, η, ϕ
	χ^2
	Inner T3 radius
	Bridge T3 radius
	Outer T3 radius

Scaled such that all features ∈ [0, 1]

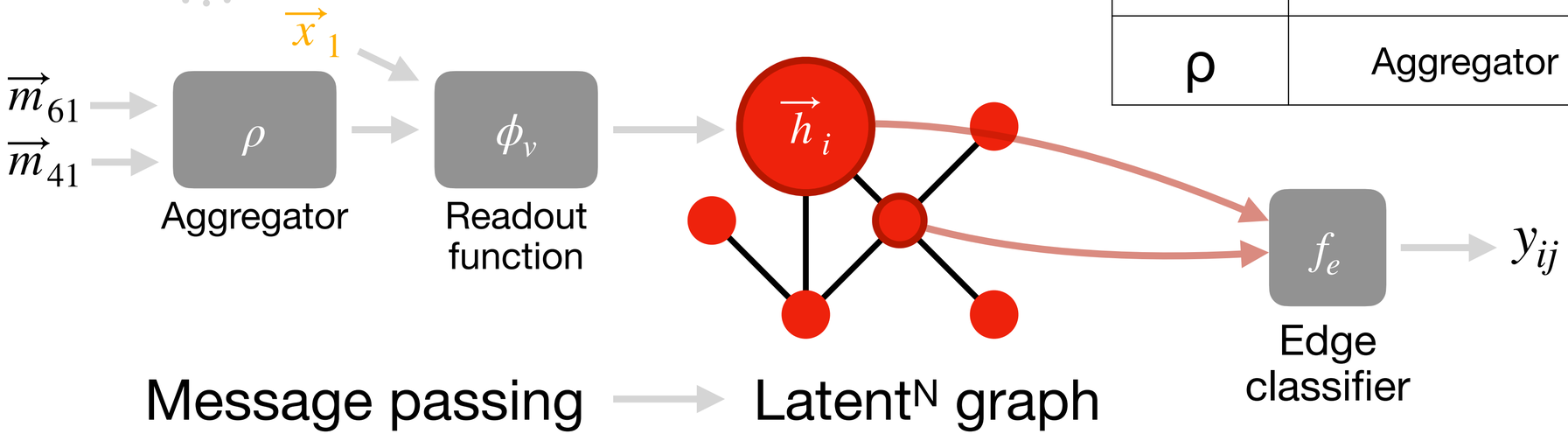


UC San Diego

GNN Configuration

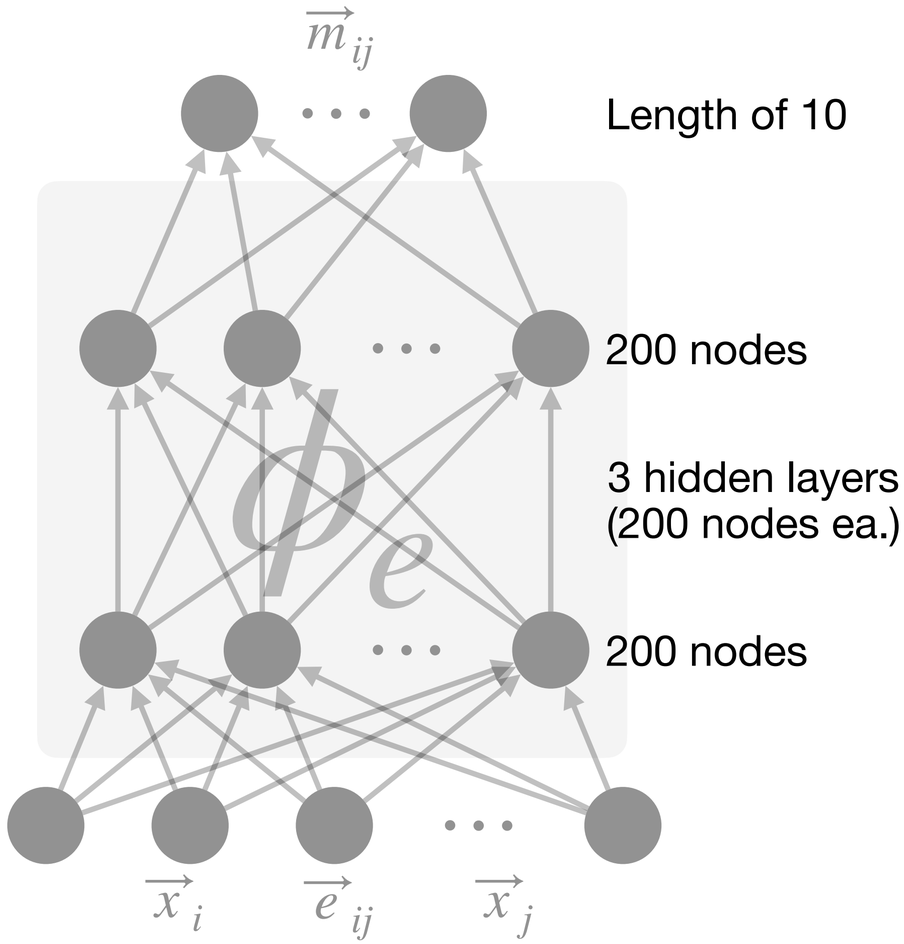


Symbol	Name	Description
ϕ_e	Message function	• Neural network • 3 hidden layers • 200 nodes per layer
ϕ_v	Readout function	• Neural network • 3 hidden layers • 200 nodes per layer
f_e	Edge classifier	• Neural network • 3 hidden layers • 200 nodes per layer
ρ	Aggregator	Sum

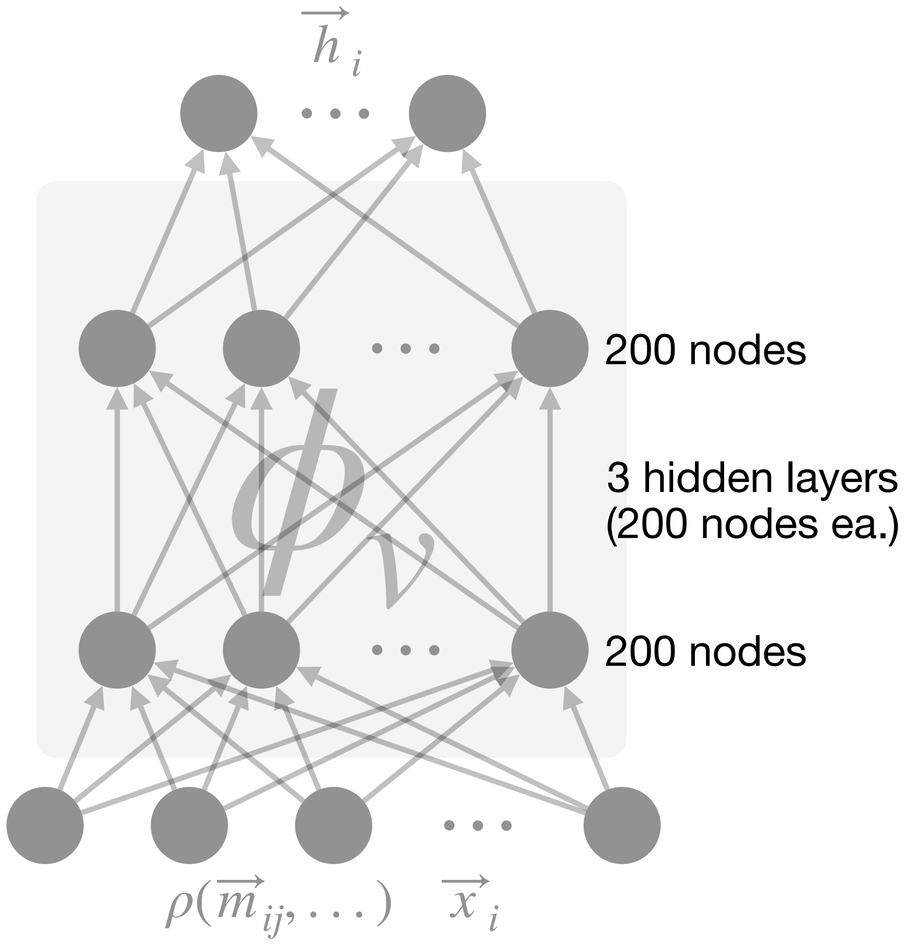


UC San Diego

GNN Internal NN Configurations



Message function

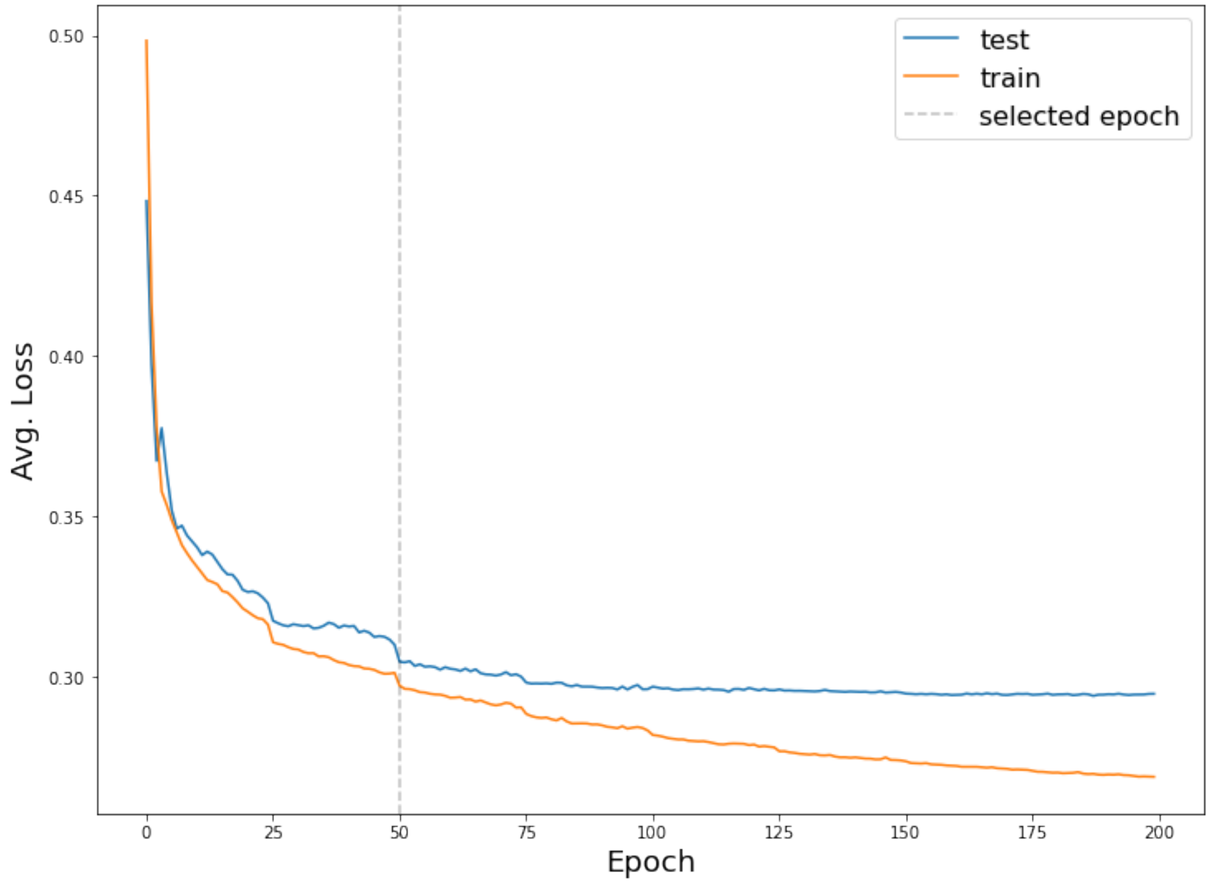


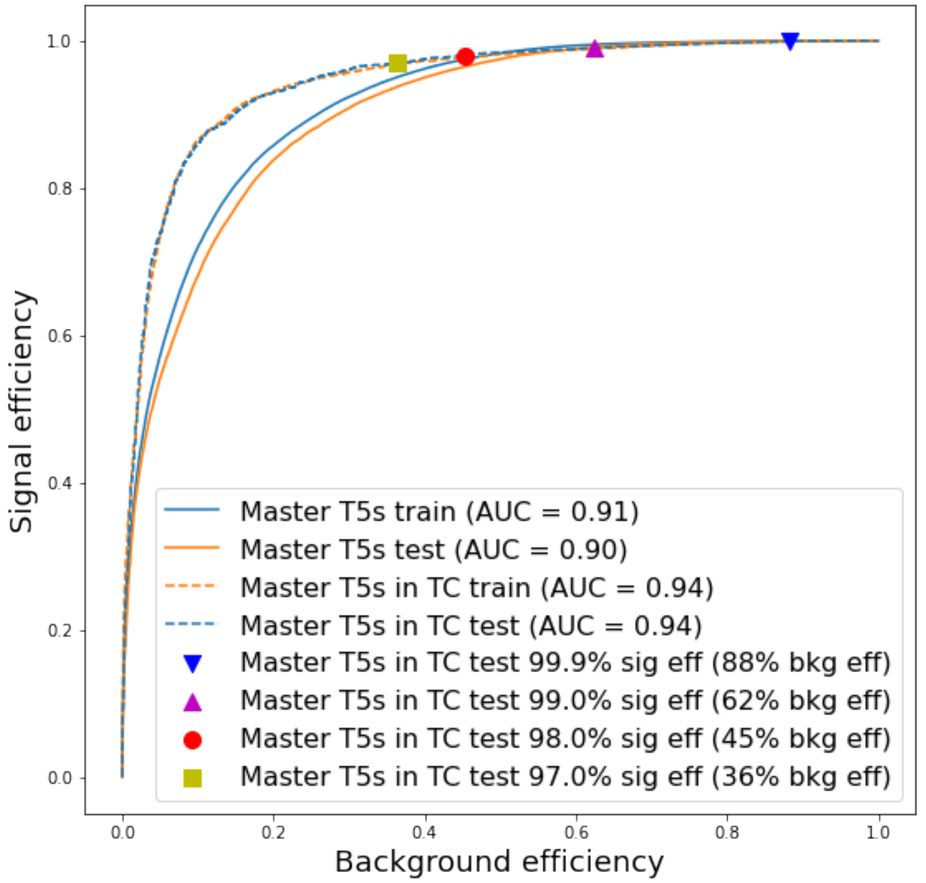
Readout function

UC San Diego

Trial 1: GNN Performance

Showing inference ROC curves before/after duplicate removal (DR)

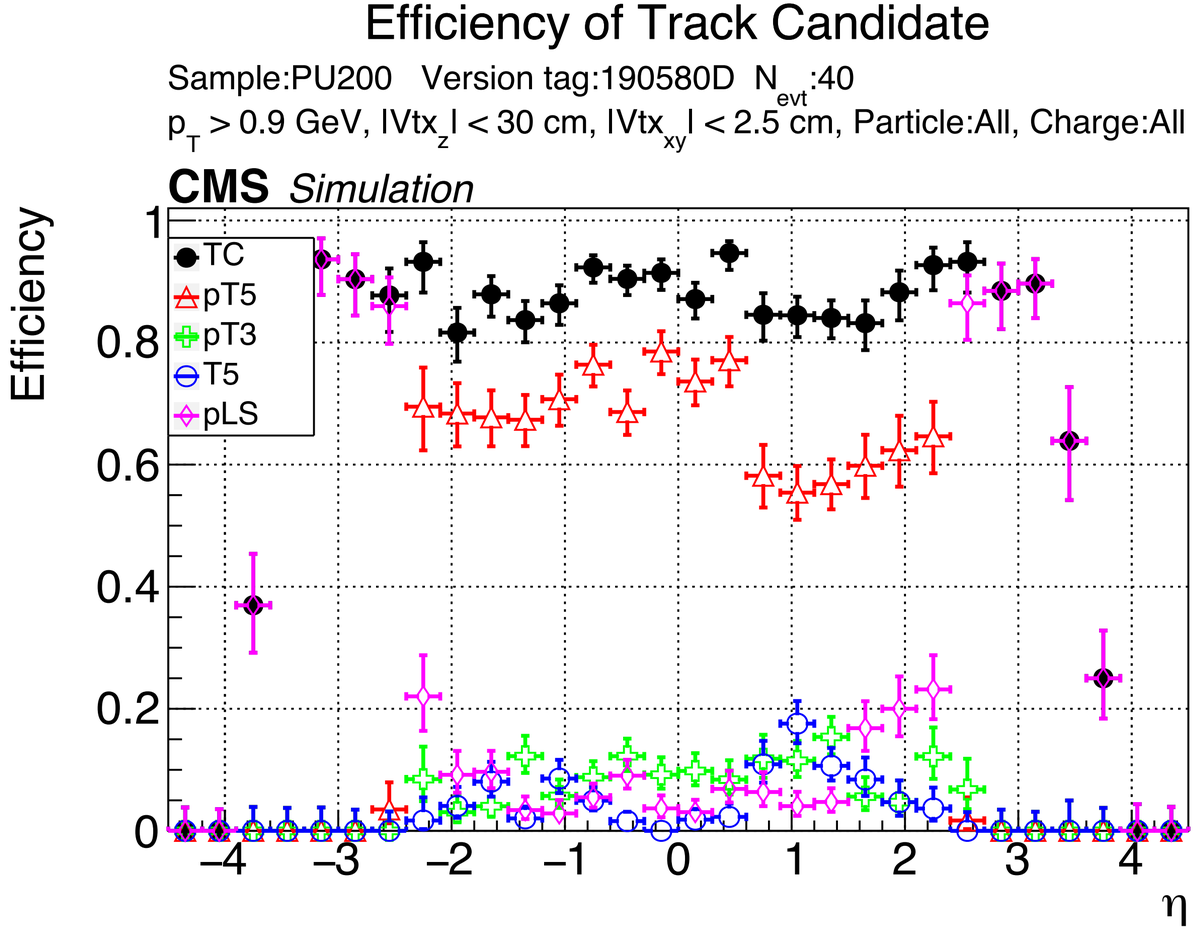


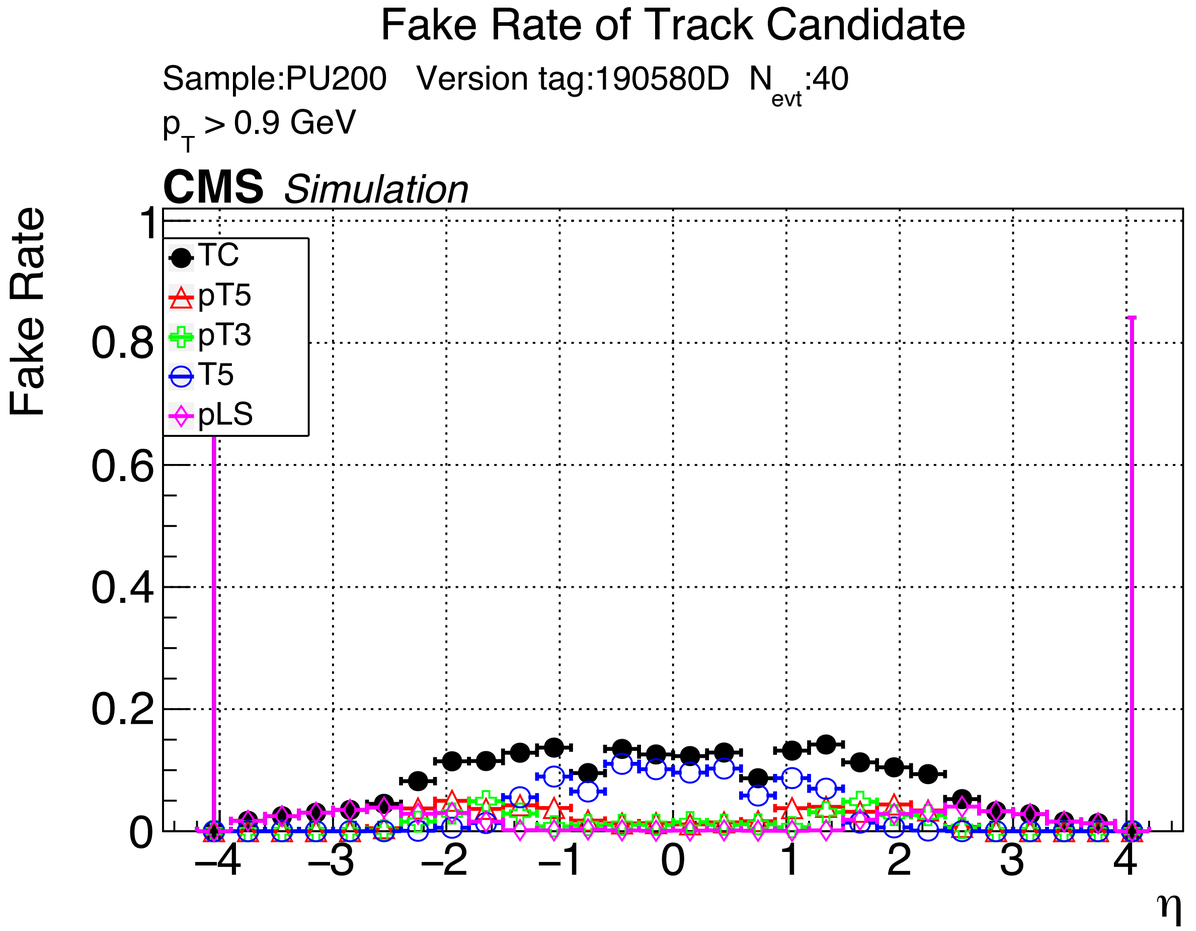


UC San Diego

LST Master Efficiency Plots

Showing η-binned eff & FR for the **40 testing events**

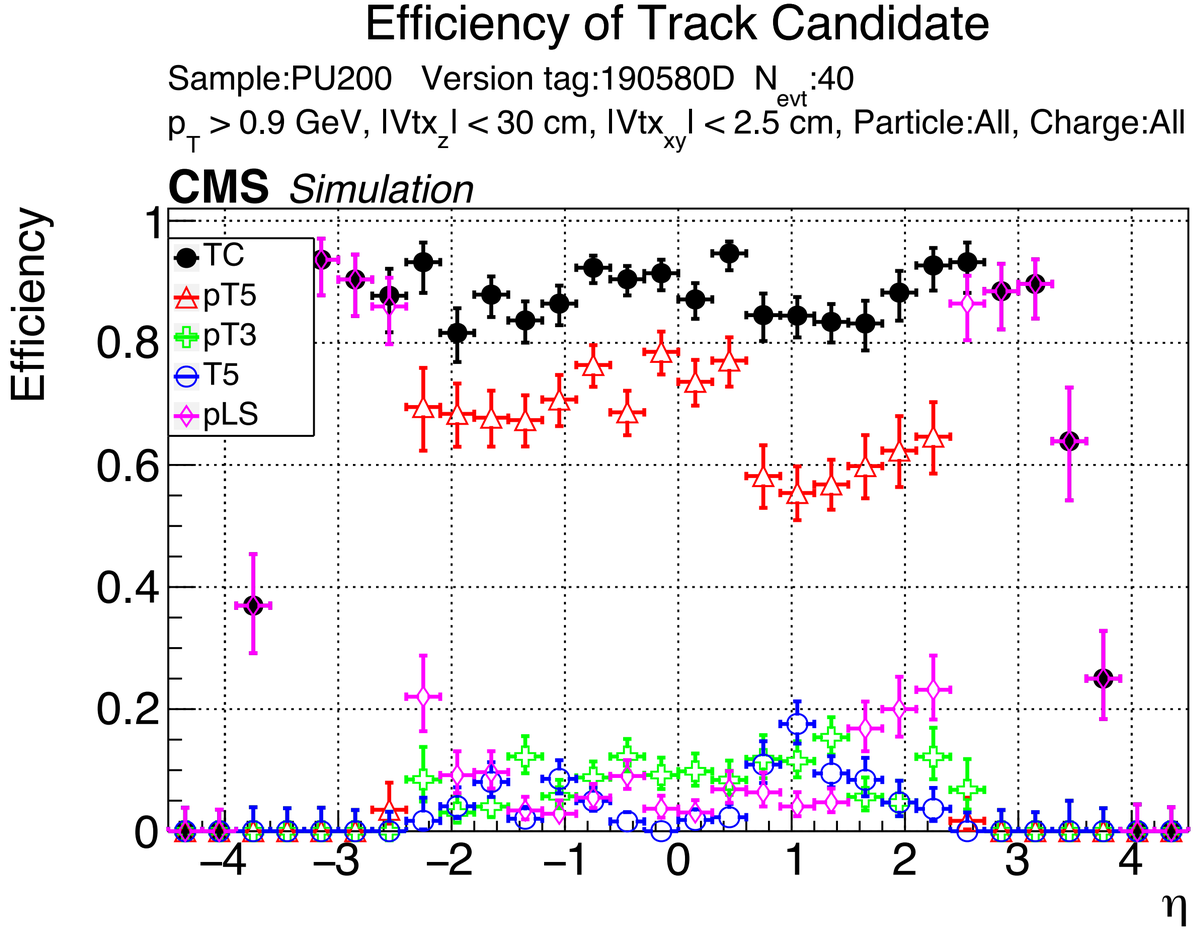


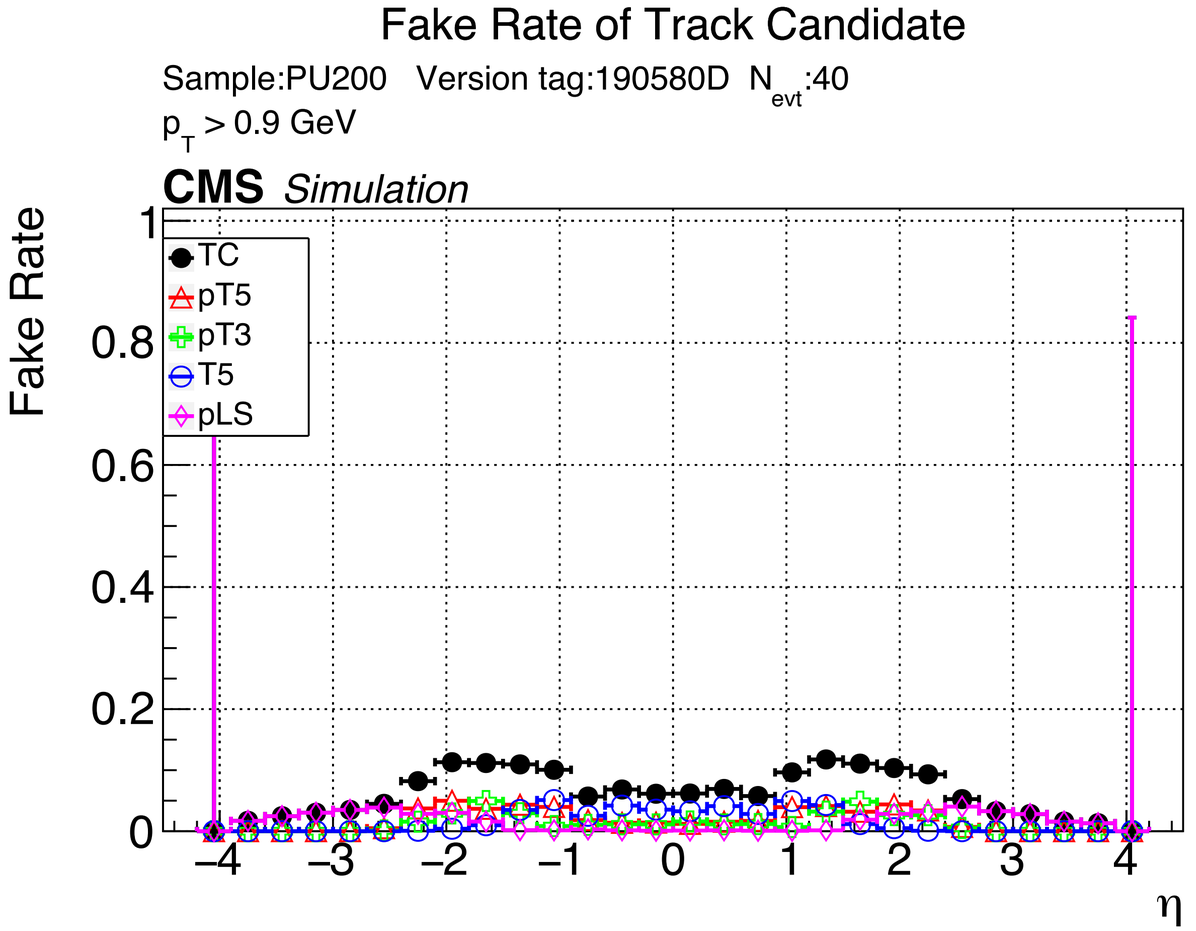


UC San Diego

LST GNN Efficiency Plots

Showing η-binned eff & FR for the **40 testing events (98% sig eff)**

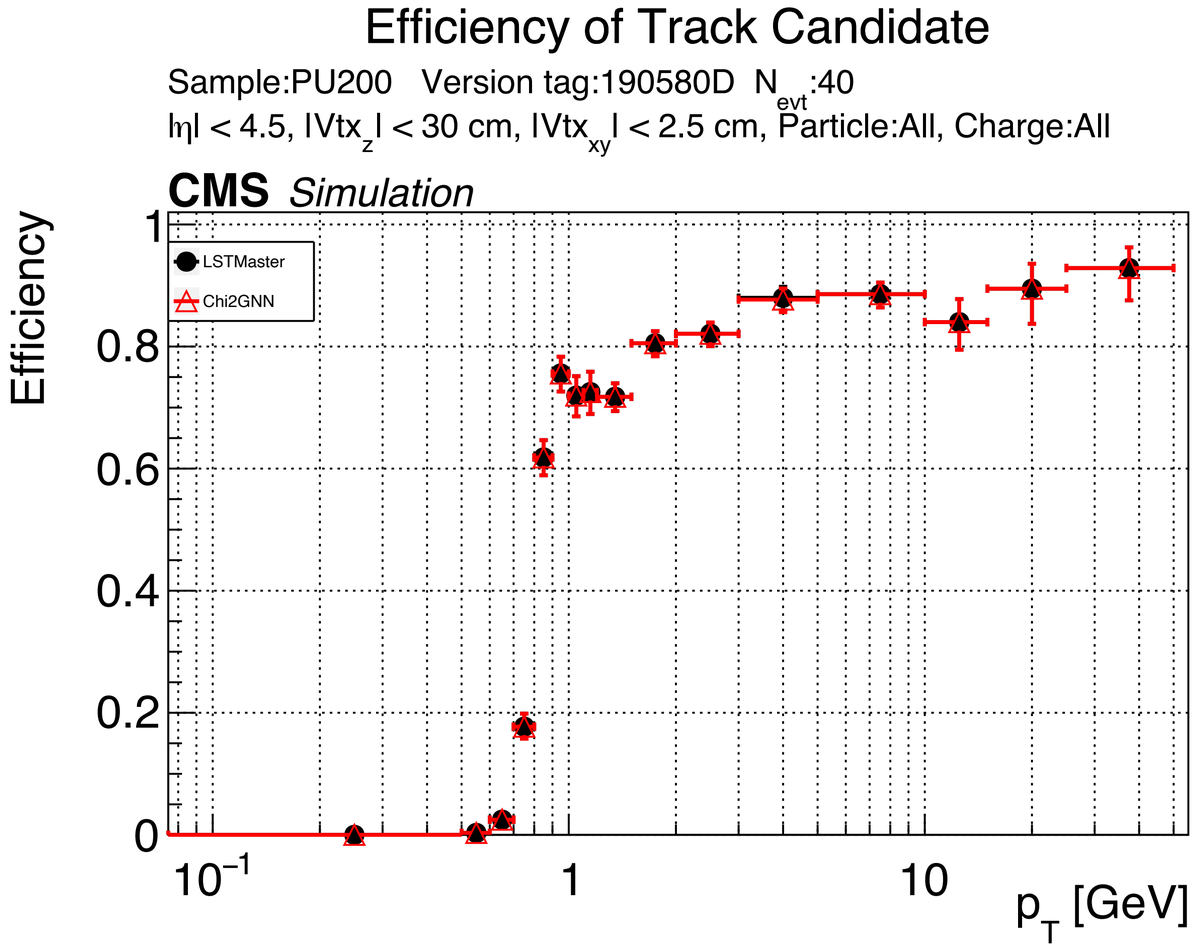


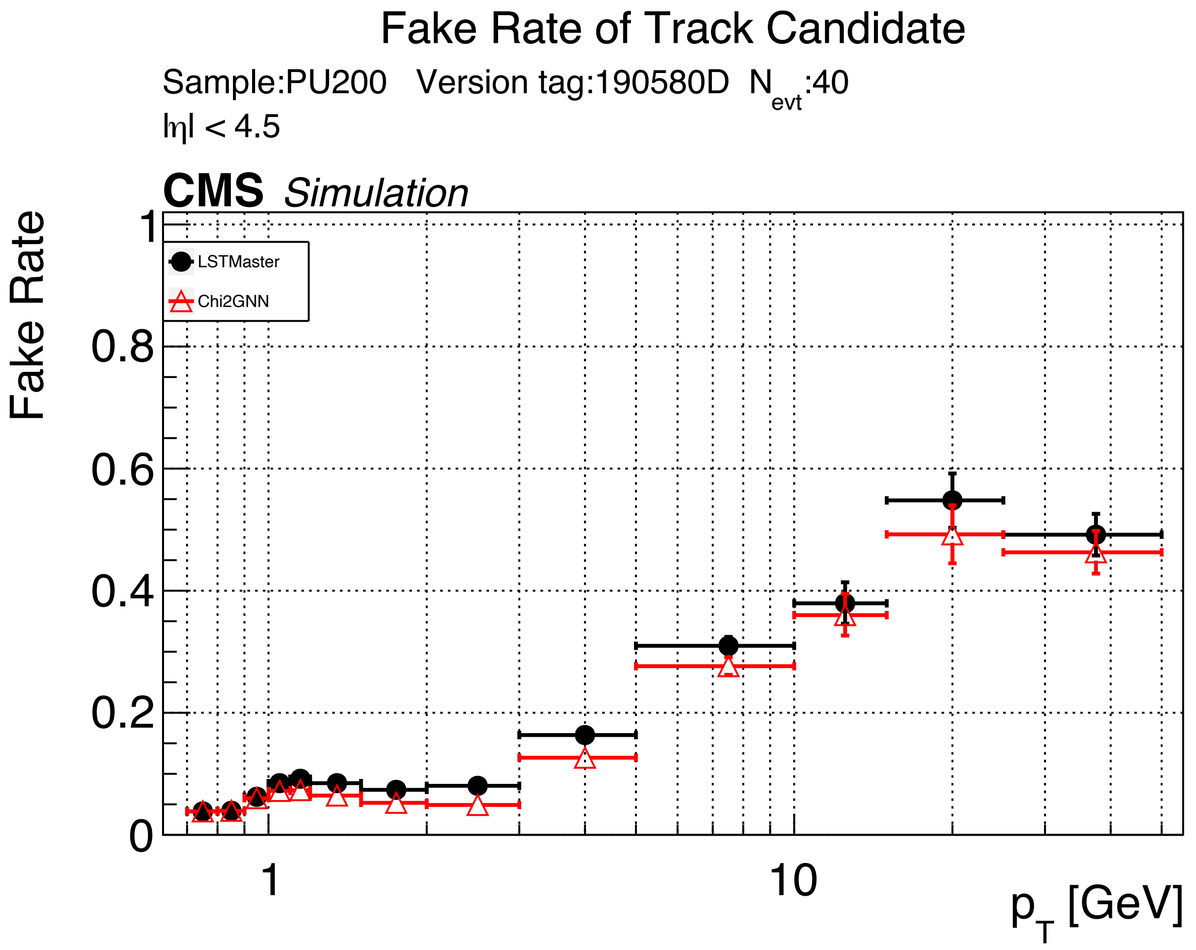


UC San Diego

LST GNN vs. Master Plots

Showing pT-binned eff & FR for master (**black**) vs. GNN (**red, 98% sig eff**)

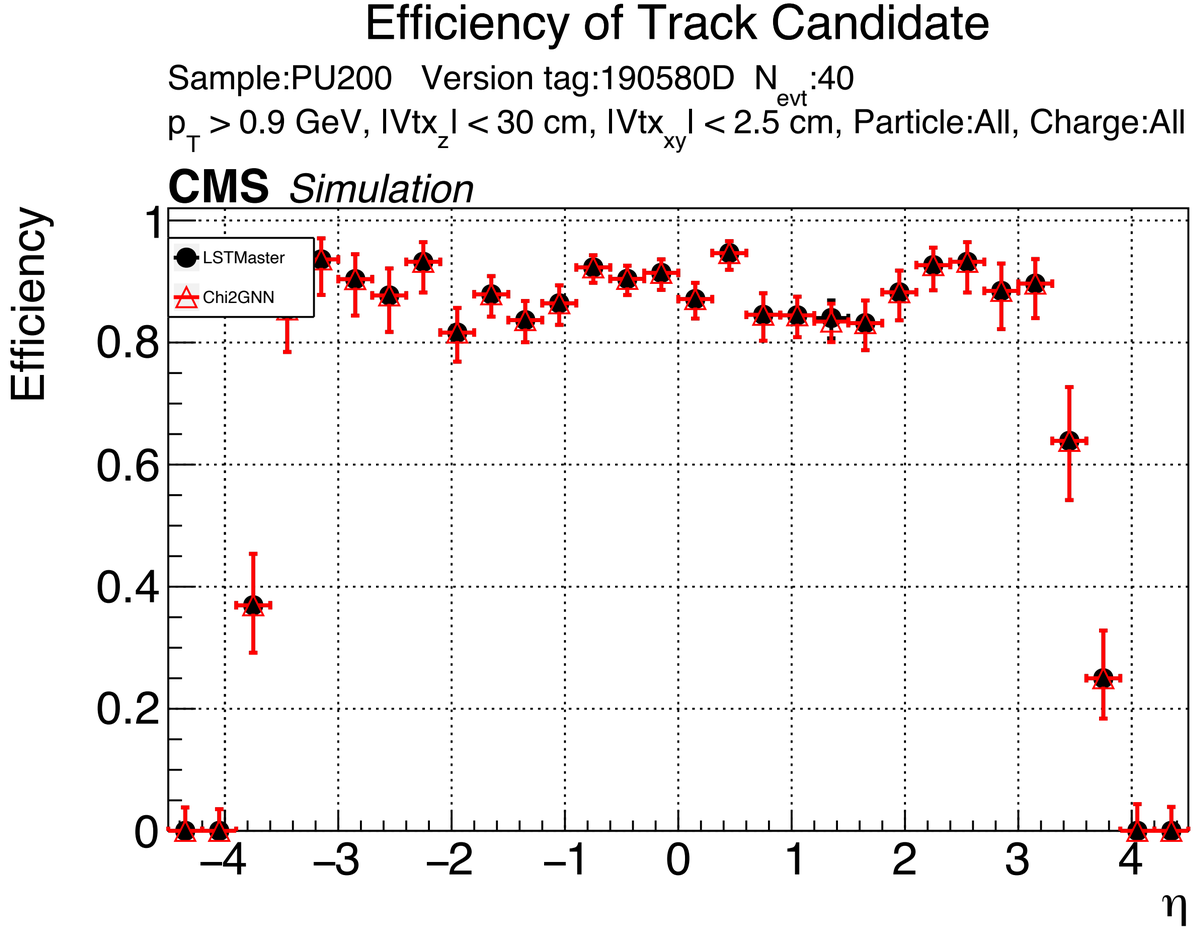


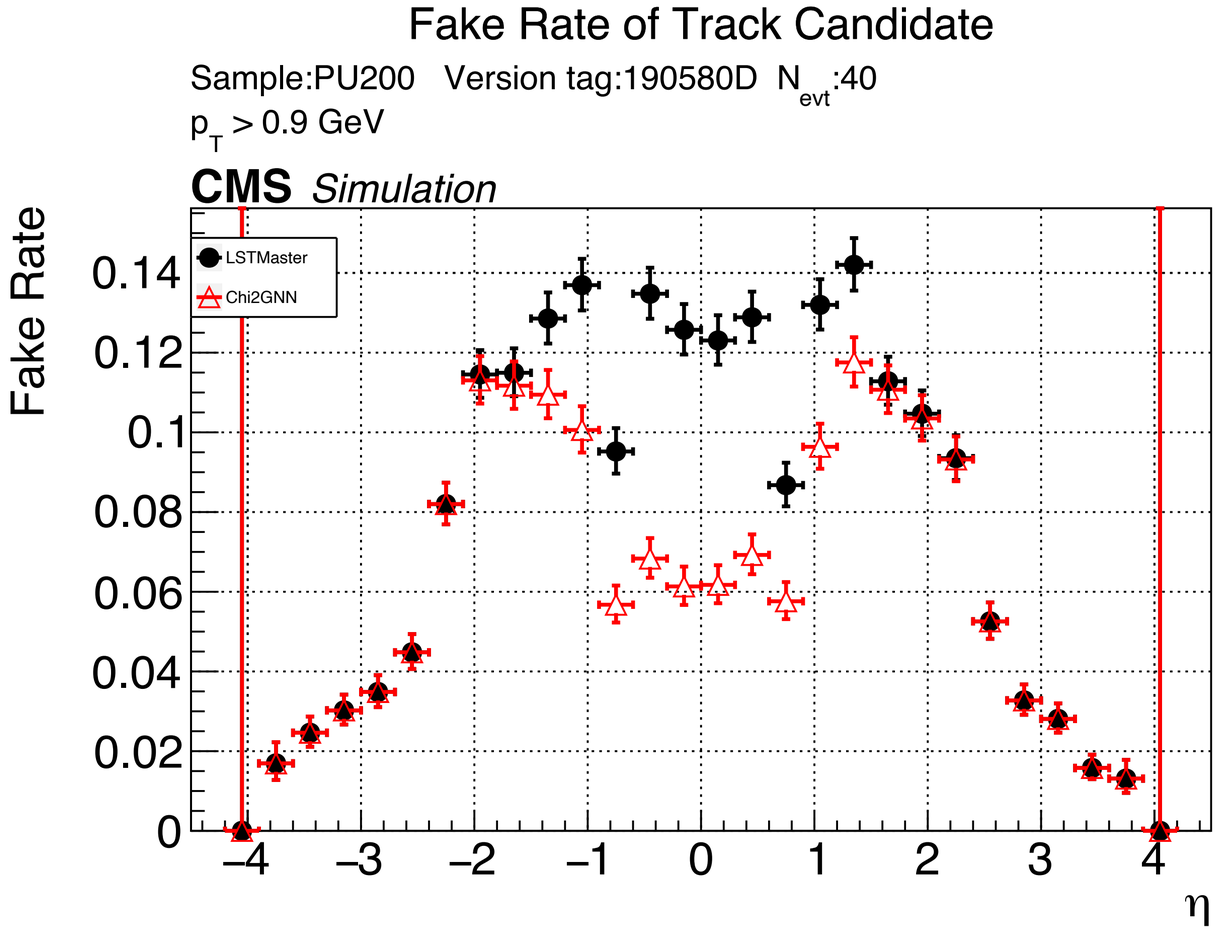


UC San Diego

LST Master vs. GNN Plots

Showing pT-binned eff & FR for master (**black**) vs. GNN (**red, 98% sig eff**)

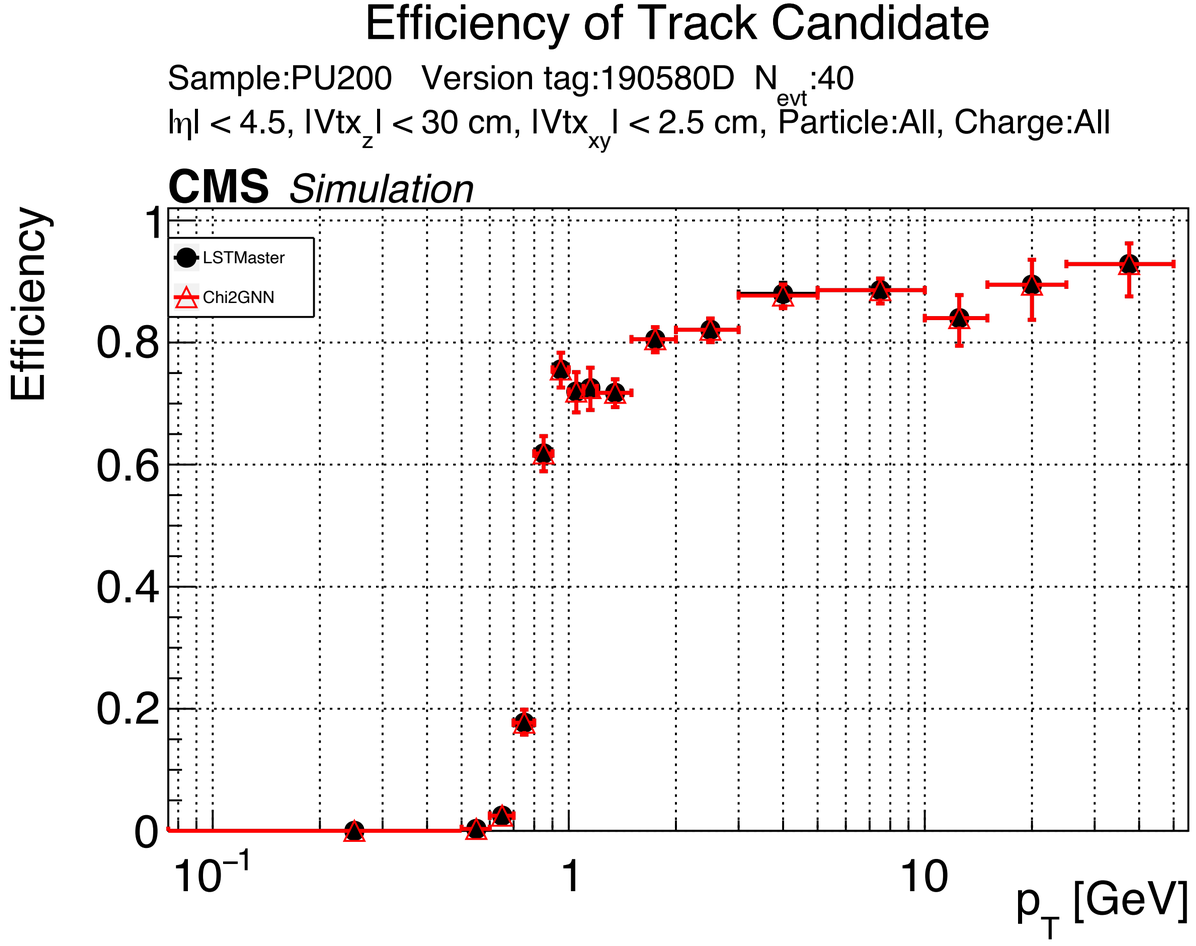


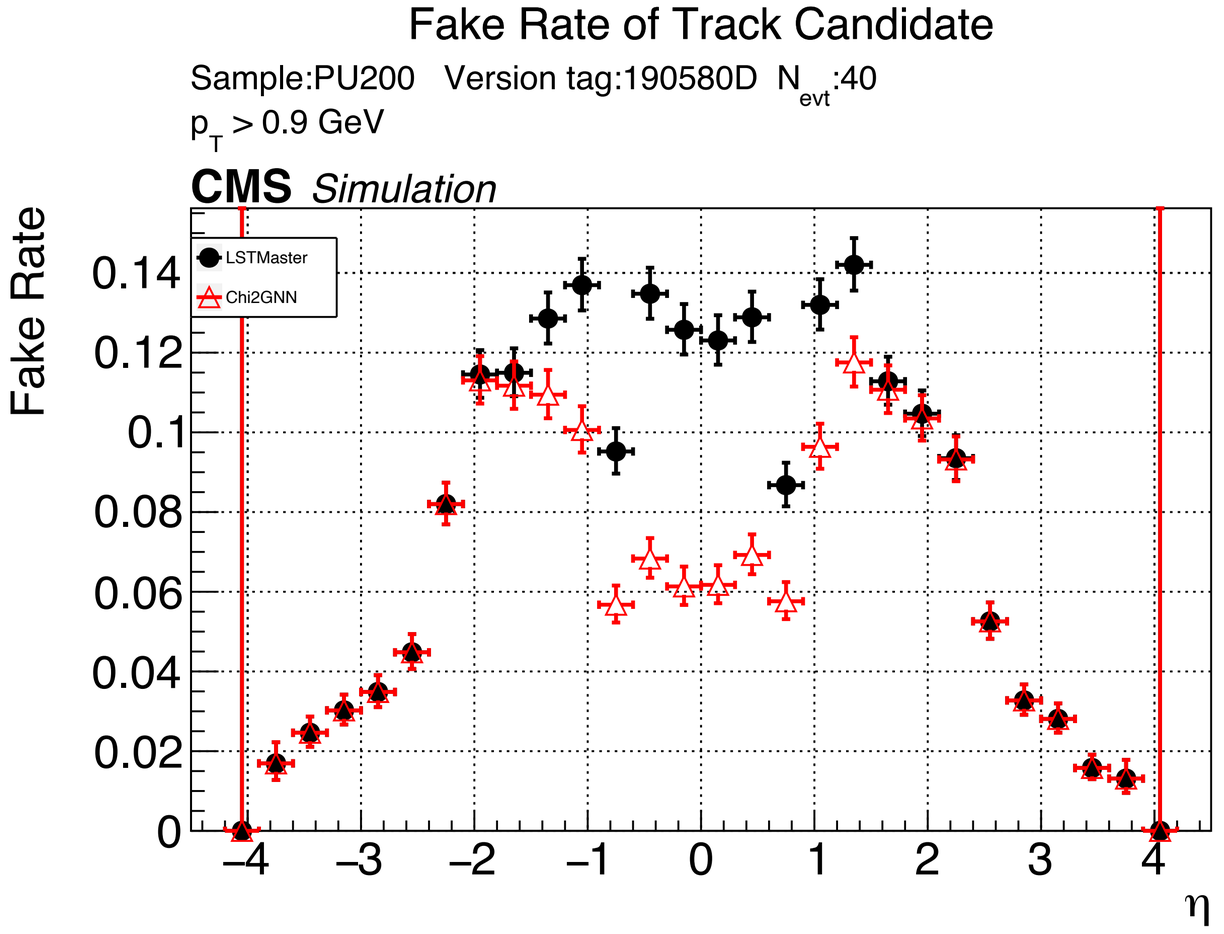


UC San Diego

LST Master vs. GNN Plots

Showing pT-binned eff & FR for master (**black**) vs. GNN (**red, 98% sig eff**)

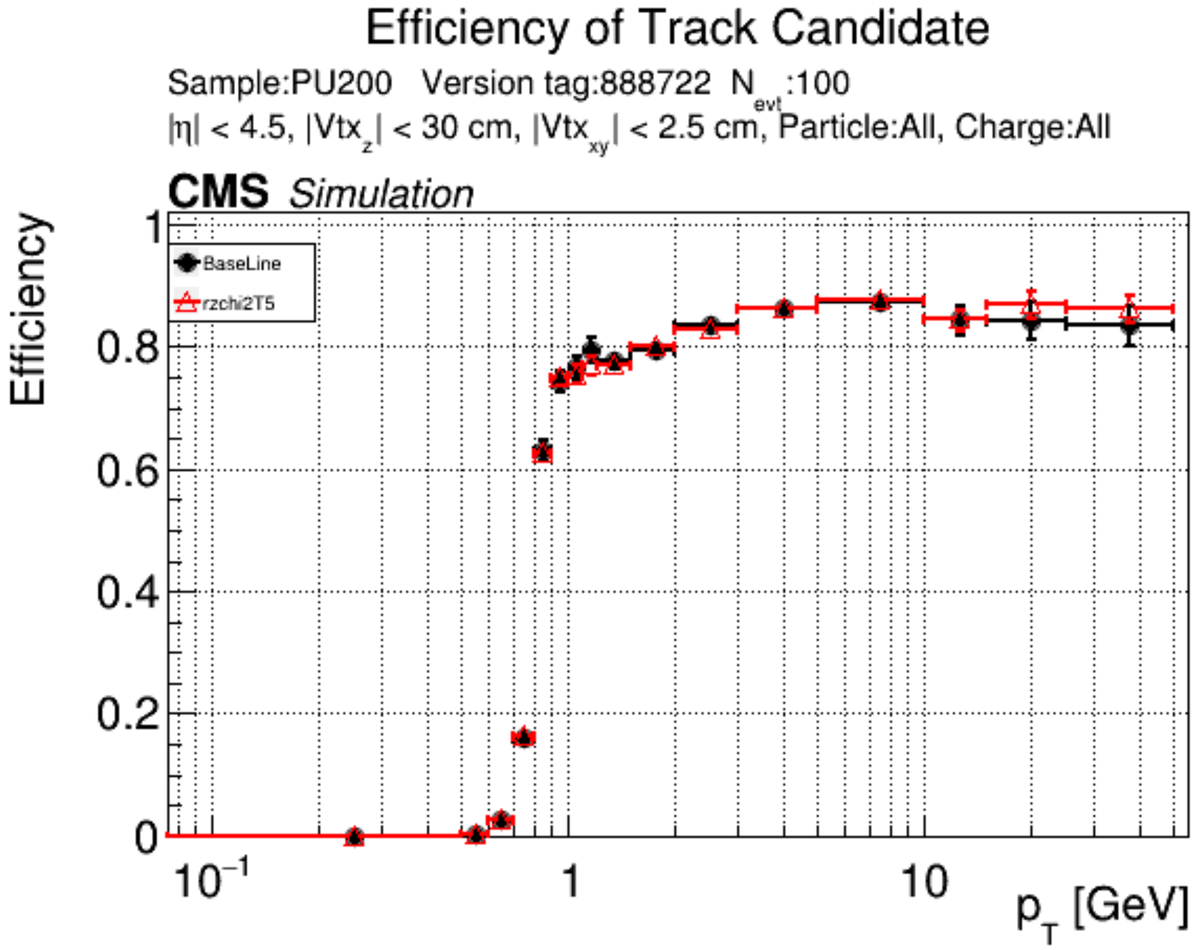


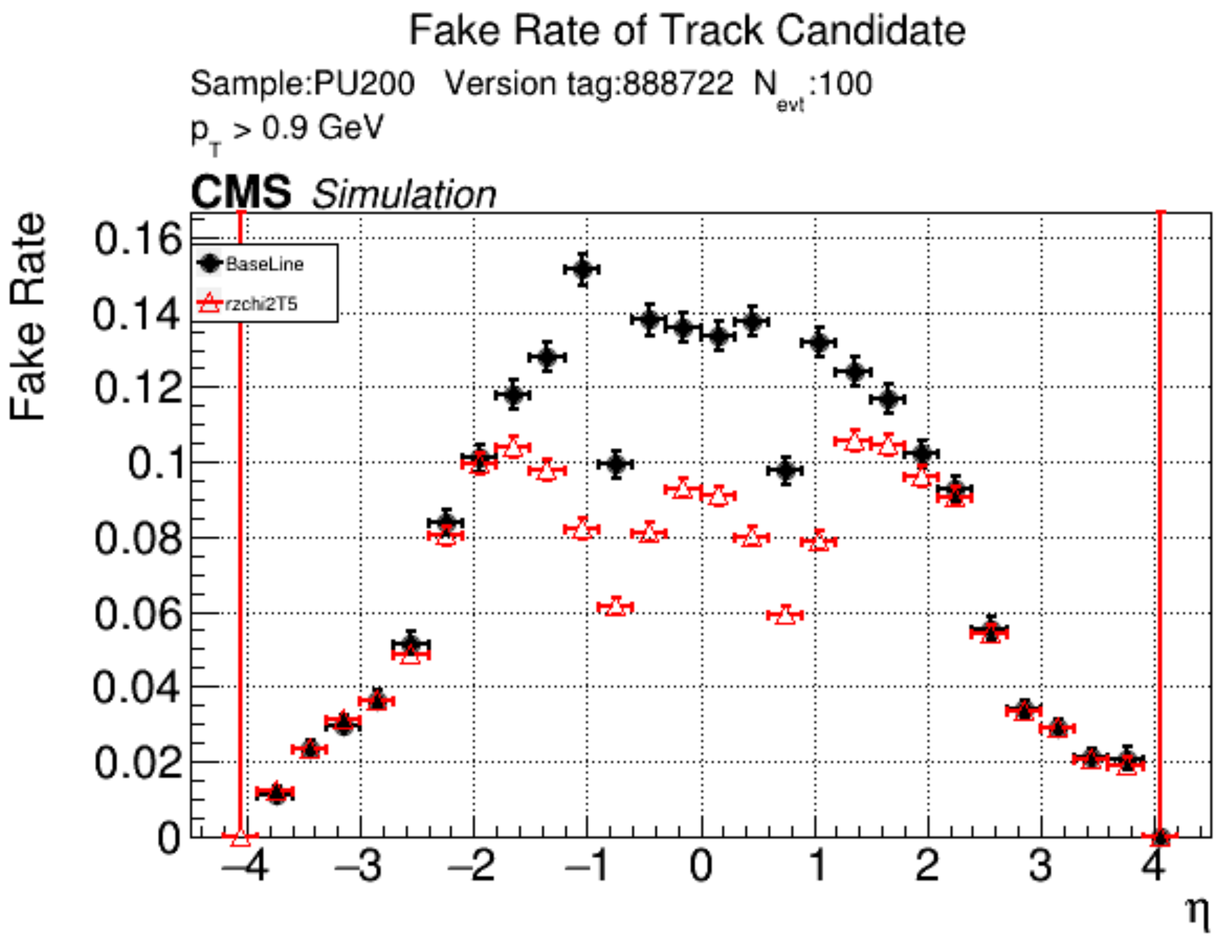


UC San Diego

LST Master vs. r-z χ^2 Plots

Showing pT-binned eff & FR for master (**black**) vs. GNN (**red, 98% sig eff**)





UC San Diego