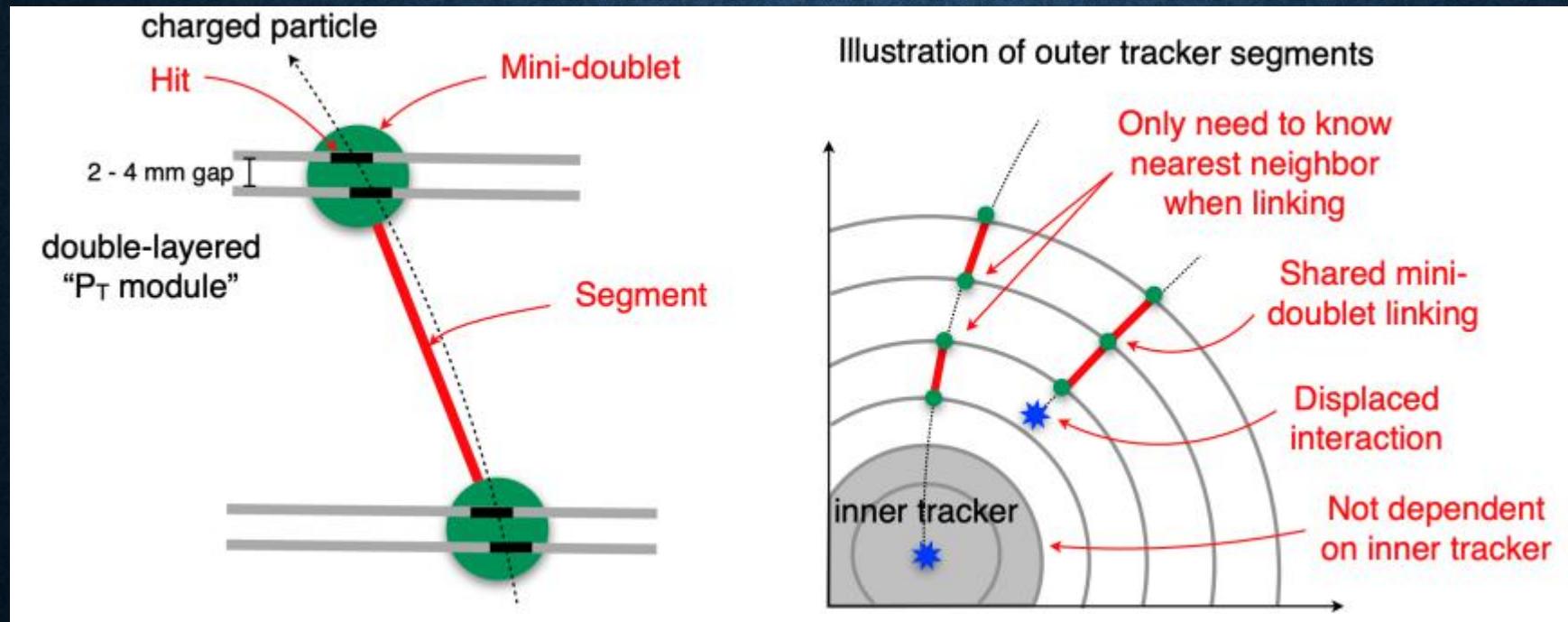


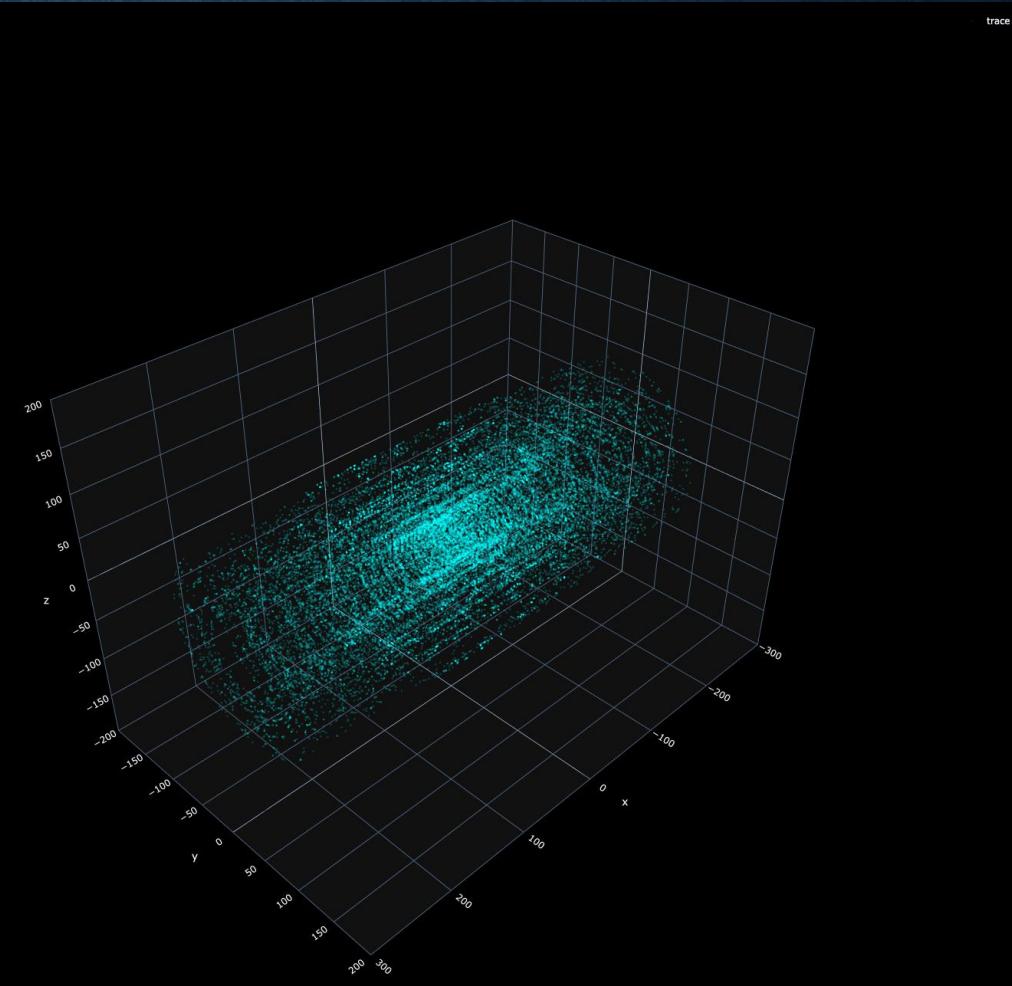
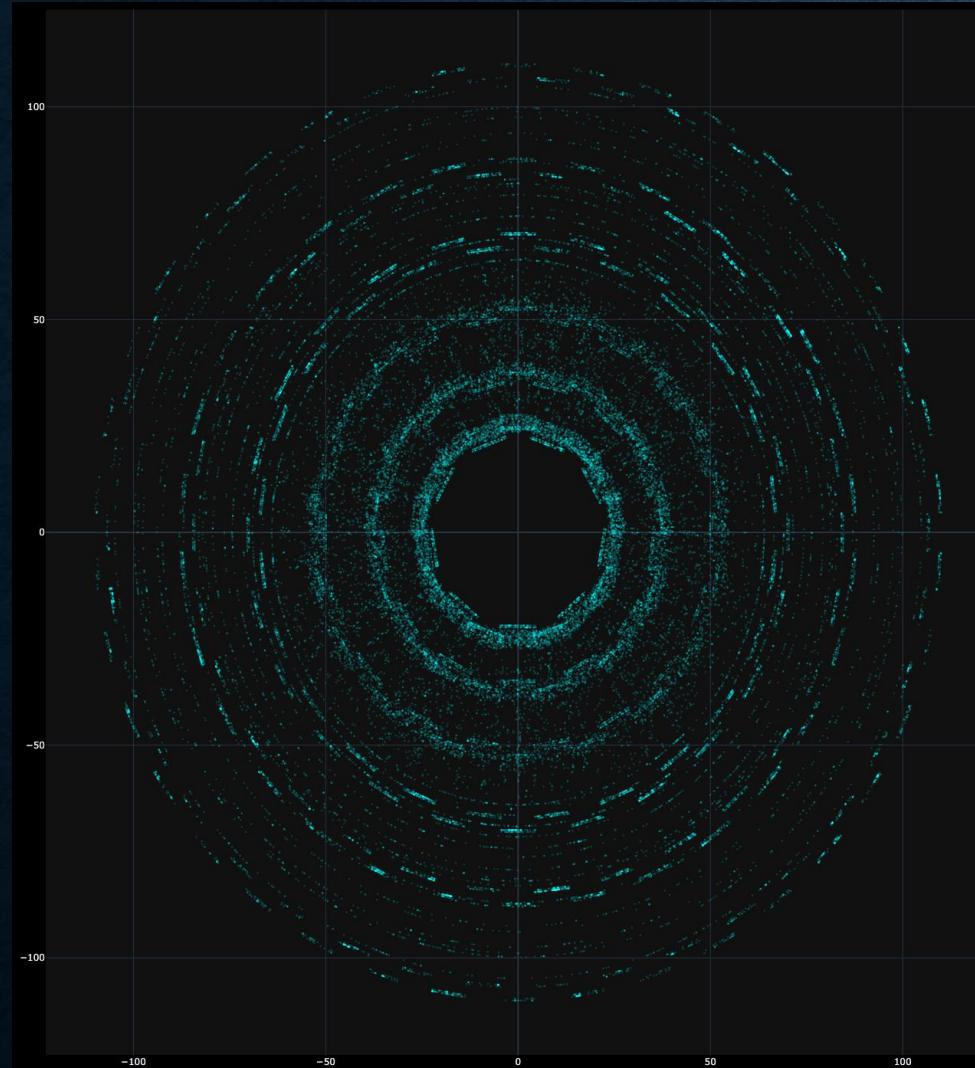
MATTHEW DITTRICH  
ESLAM ZENHOM



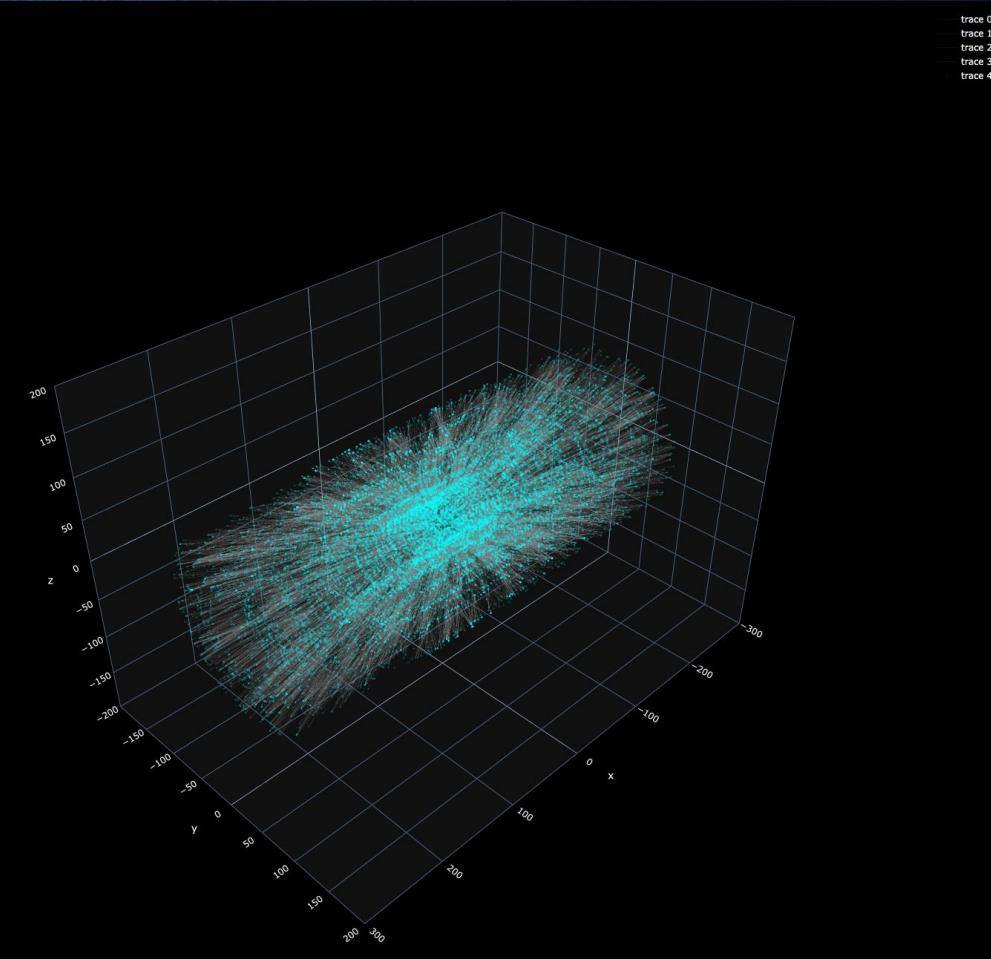
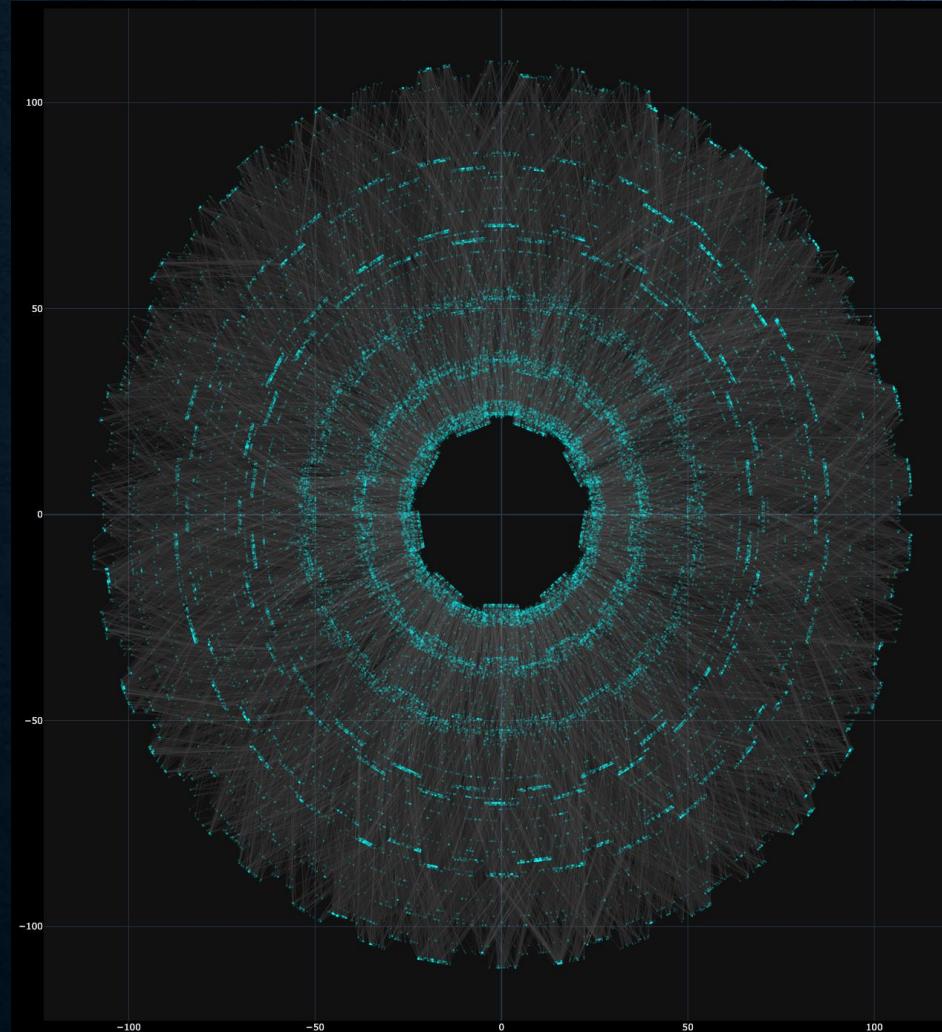
# Charged particle trajectory in a tracker



# Mini-Doublets are Nodes (MD)

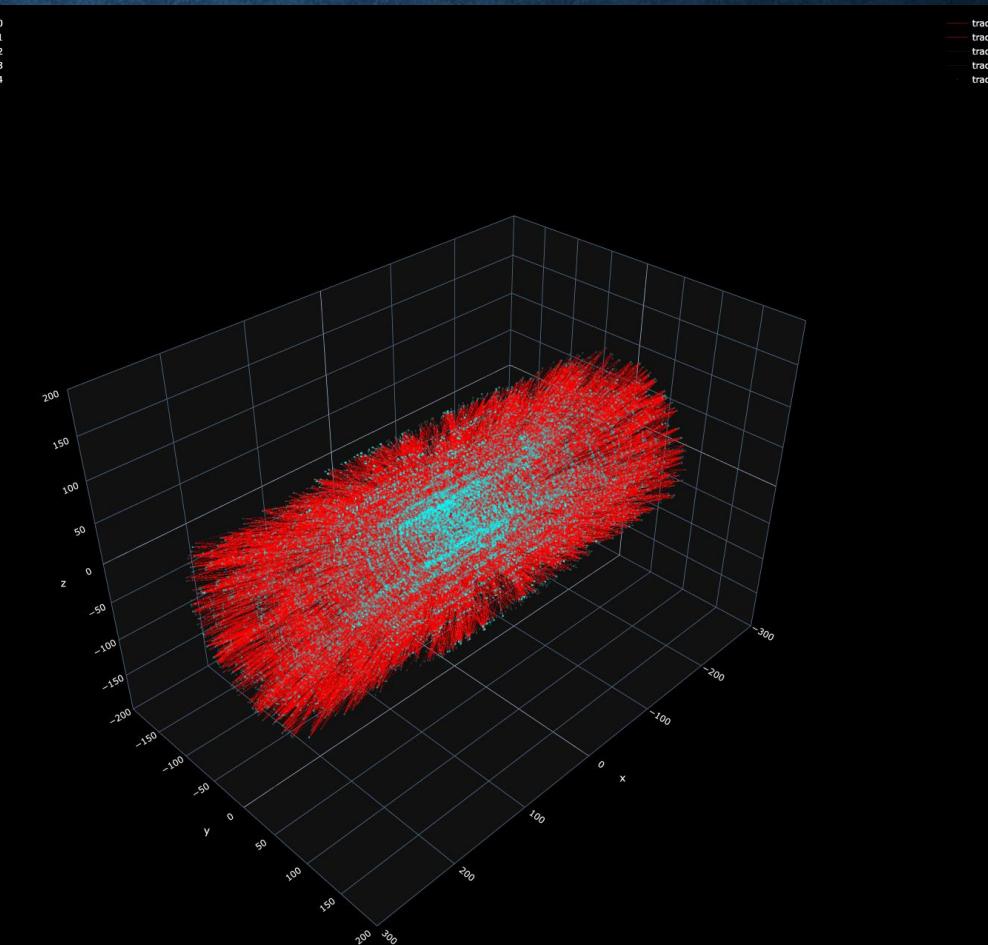
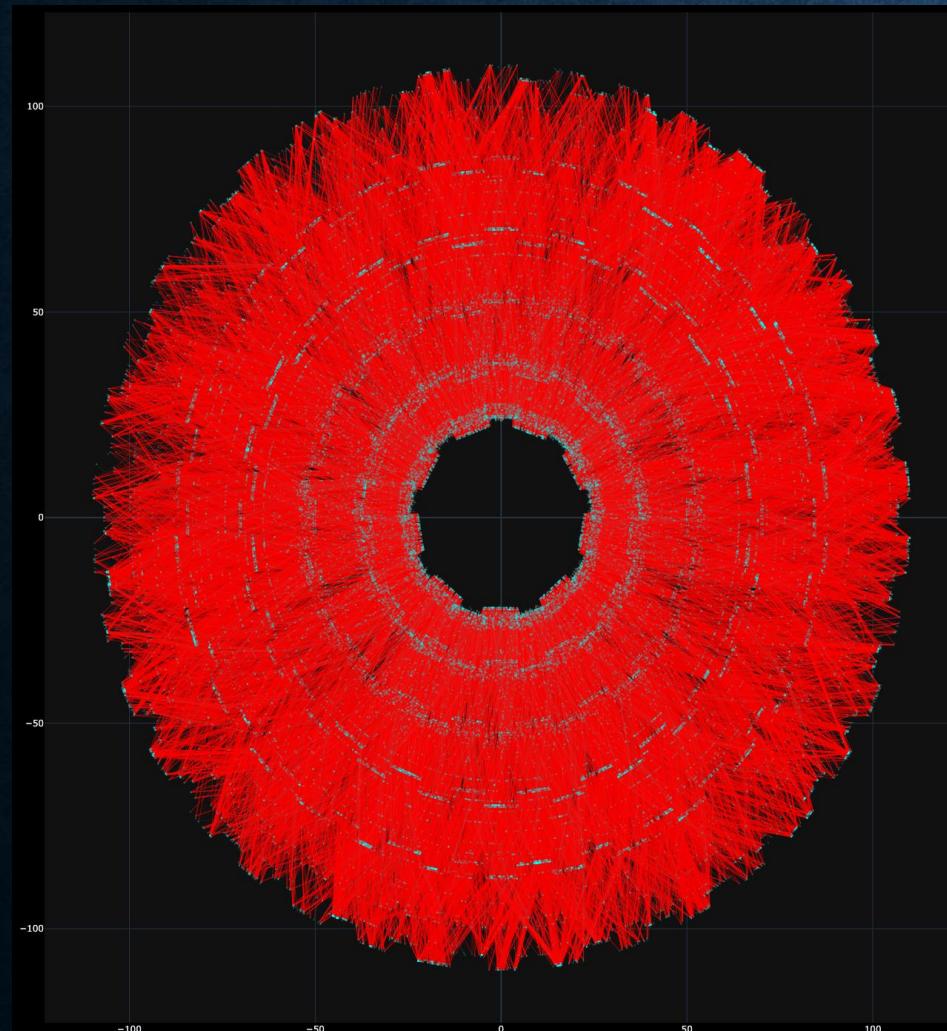


# Line Segments are Edges (LS)



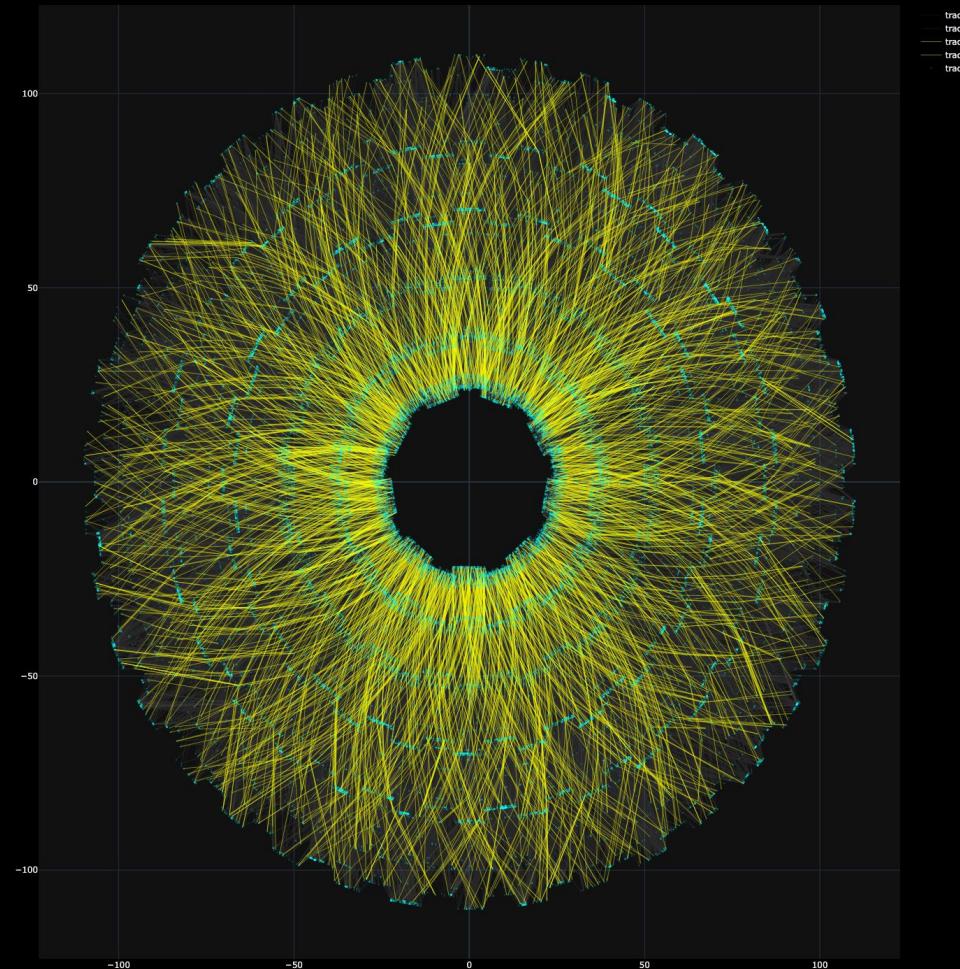
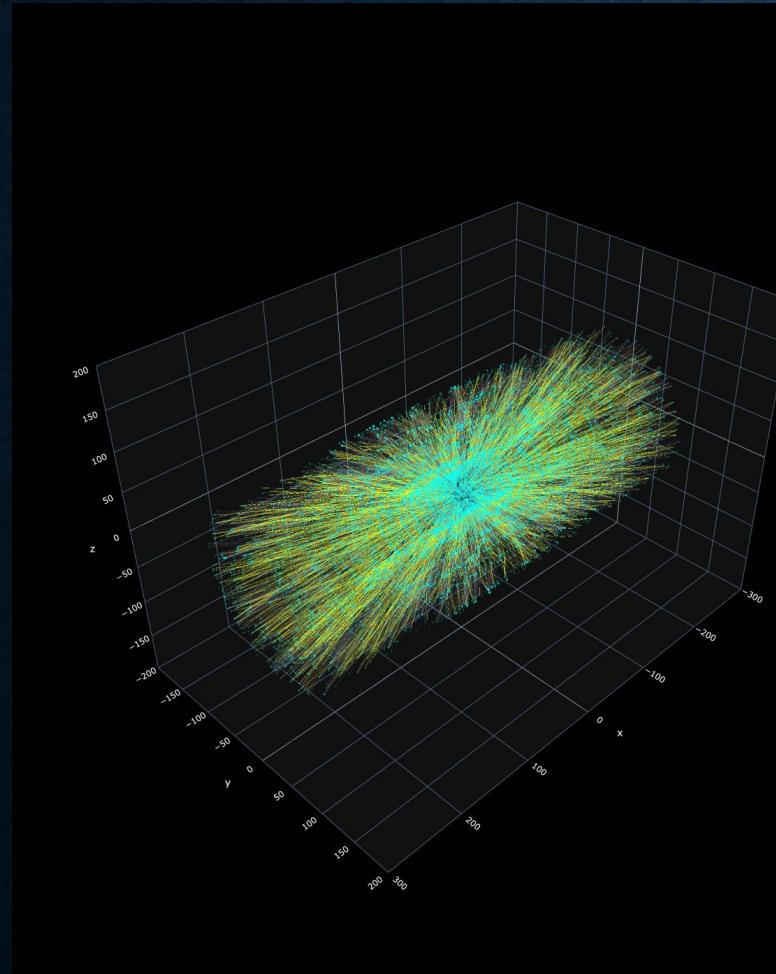
# Fake Edges

- Fake: when not all hits in the line segment matched to a unique simulated track



# True Edges

- True: when all hits in the line segment matched to a unique simulated track



# Numbers for a specific events

All Mini-Doublets	41459
Total Line Segments	144245
True Line Segments	10549
Fake Line Segments	133696

96th event in the PU200 ttbar (CMSSW\_12\_2\_0\_pre2)

# Graph Neural Network training

- Used PU 200 ttbar sample (first 95 events in CMSSW\_12\_2\_0\_pre2)
- The 96th event is used as a test (the ones shown in previous slide)

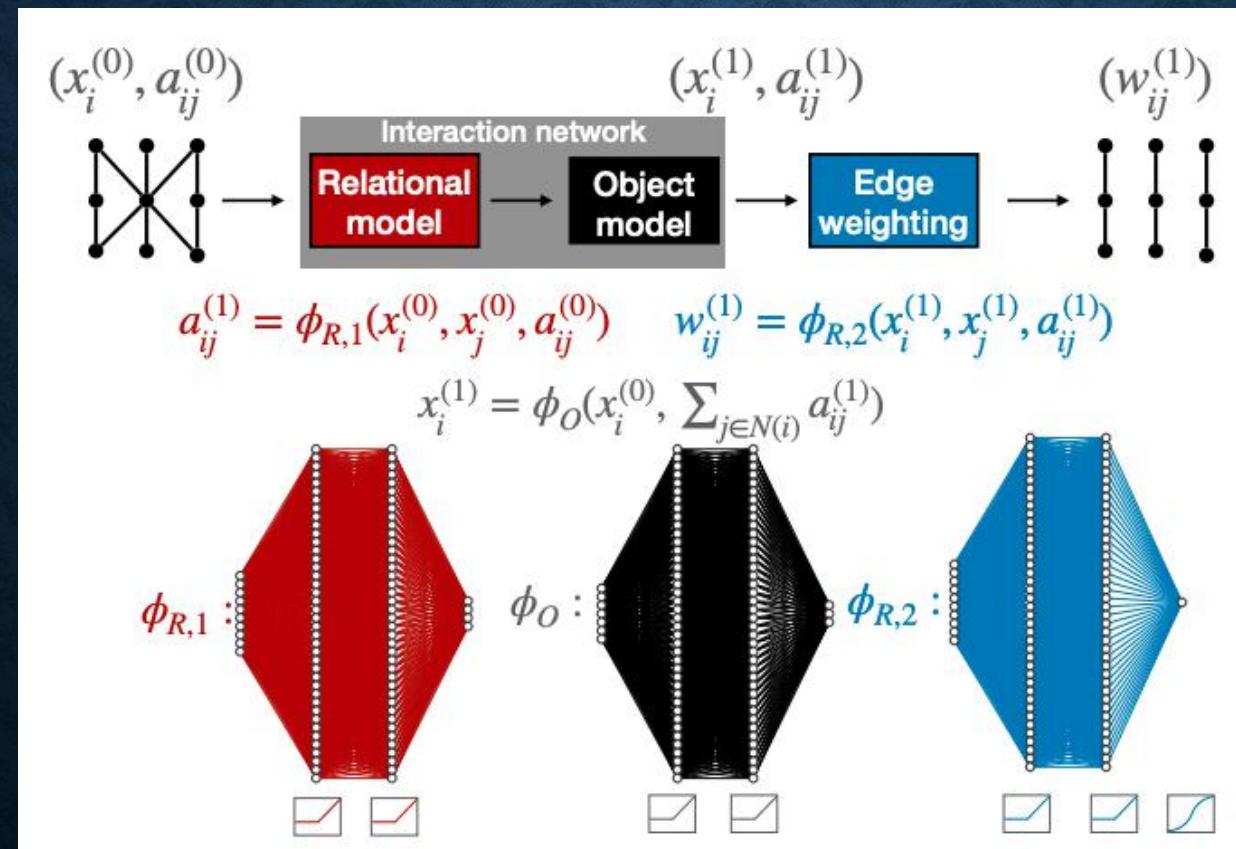
# Training inputs

- Node features:
  - xyz positions of mini-doublets
  - $\Delta\phi$  change (our MD pt estimator)
- Edge features :
  - pt, eta, phi estimates based on two anchor MD points

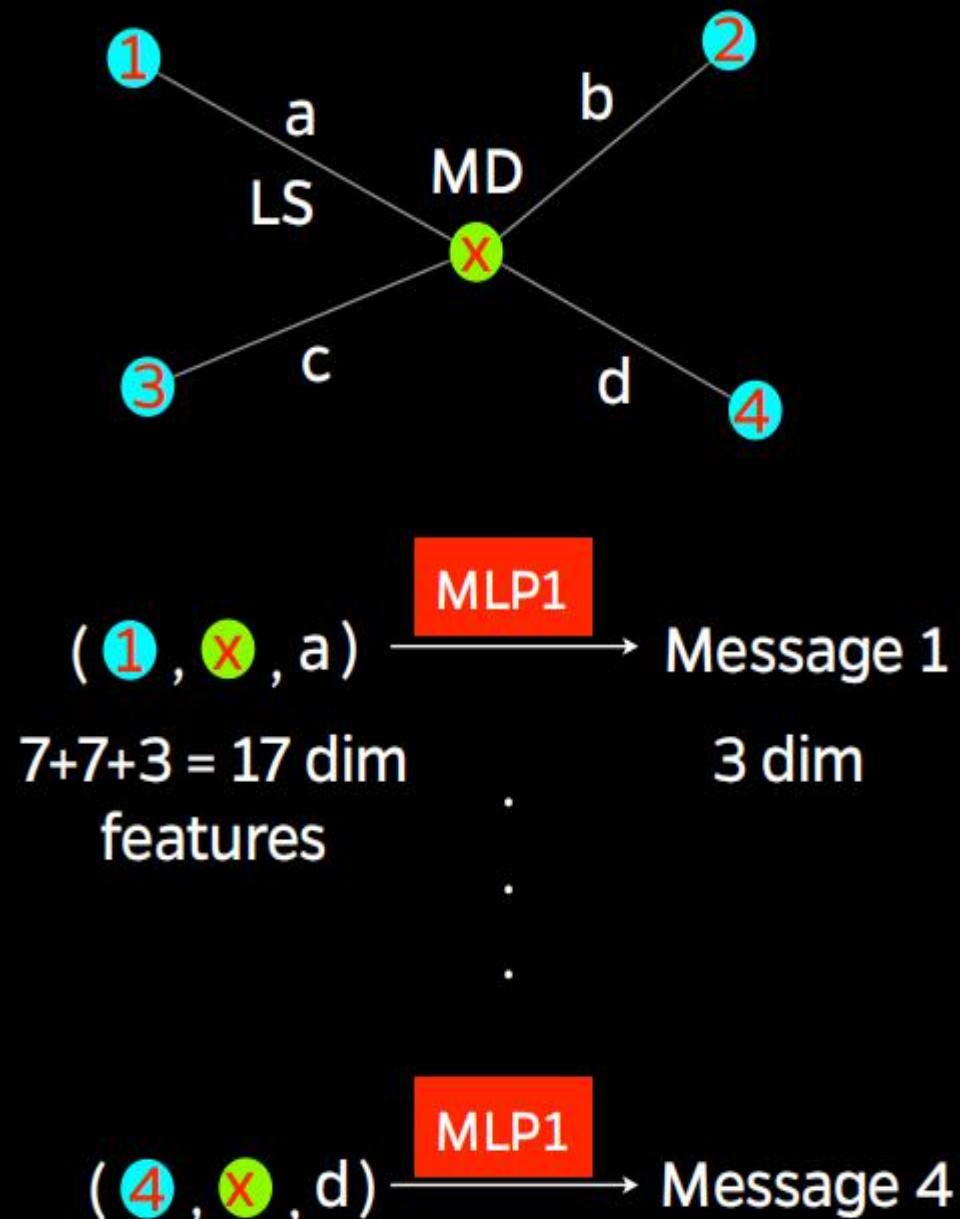
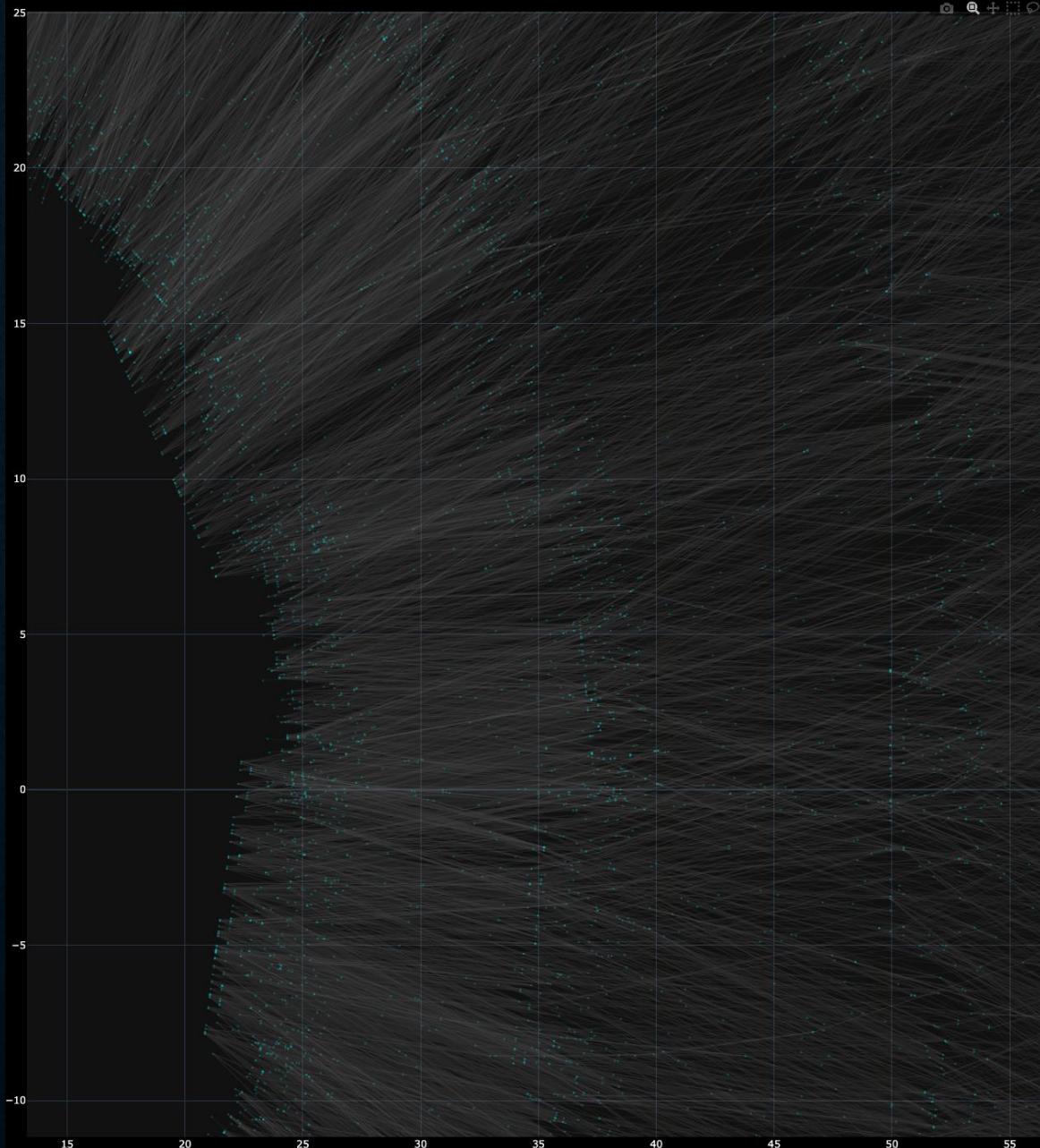
# Training model

“Interaction Network”

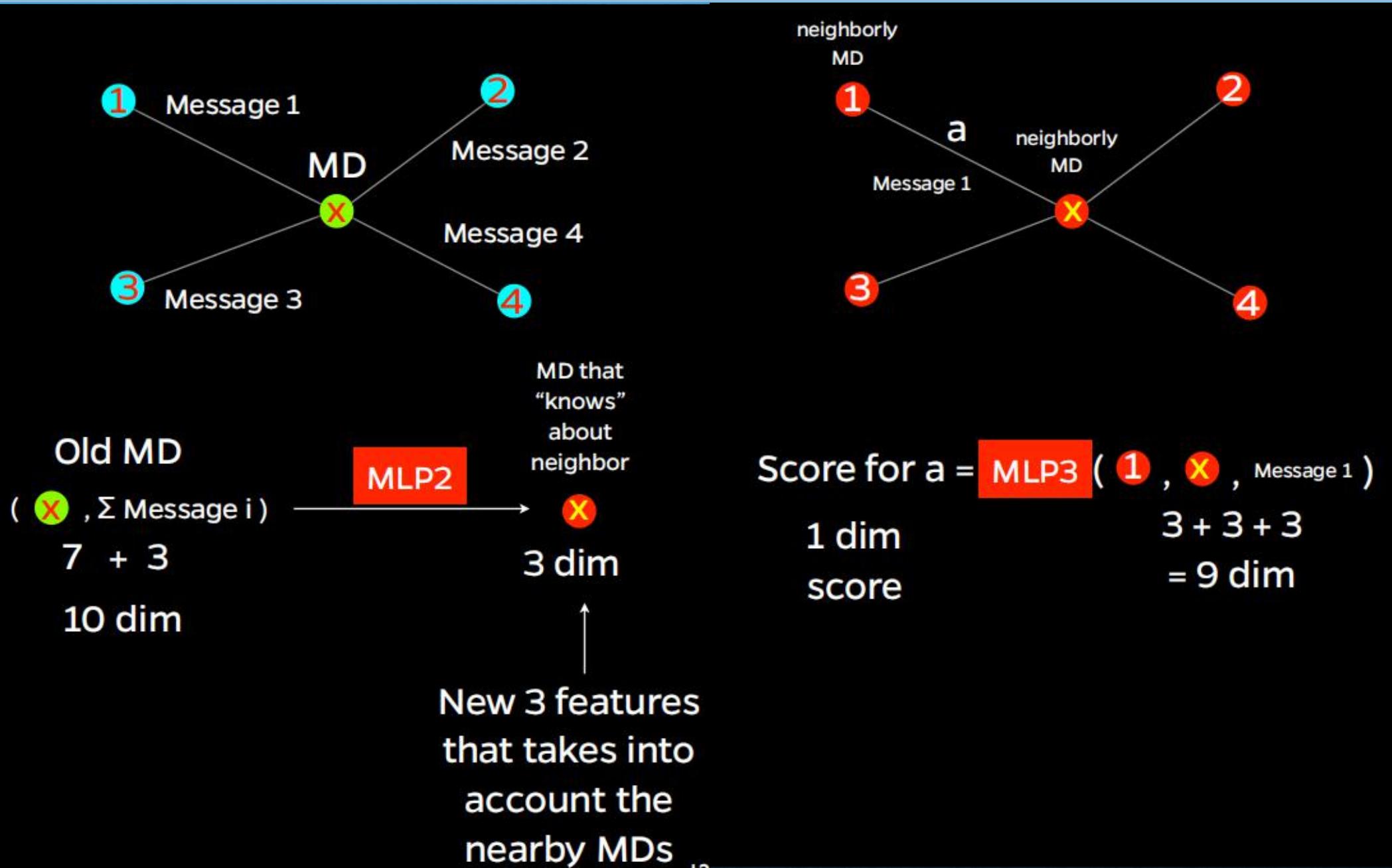
<https://arxiv.org/pdf/2103.16701.pdf>



# Training Model



# Training Model



# The MLPs to be trained

There are 3 MLPs being trained:

1. MLP to compute message based on two MDs, and the LS
2. MLP to compute aggregate messages and output “neighborly” MD
3. MLP to take “neighborly” MDs and message to classify the LS

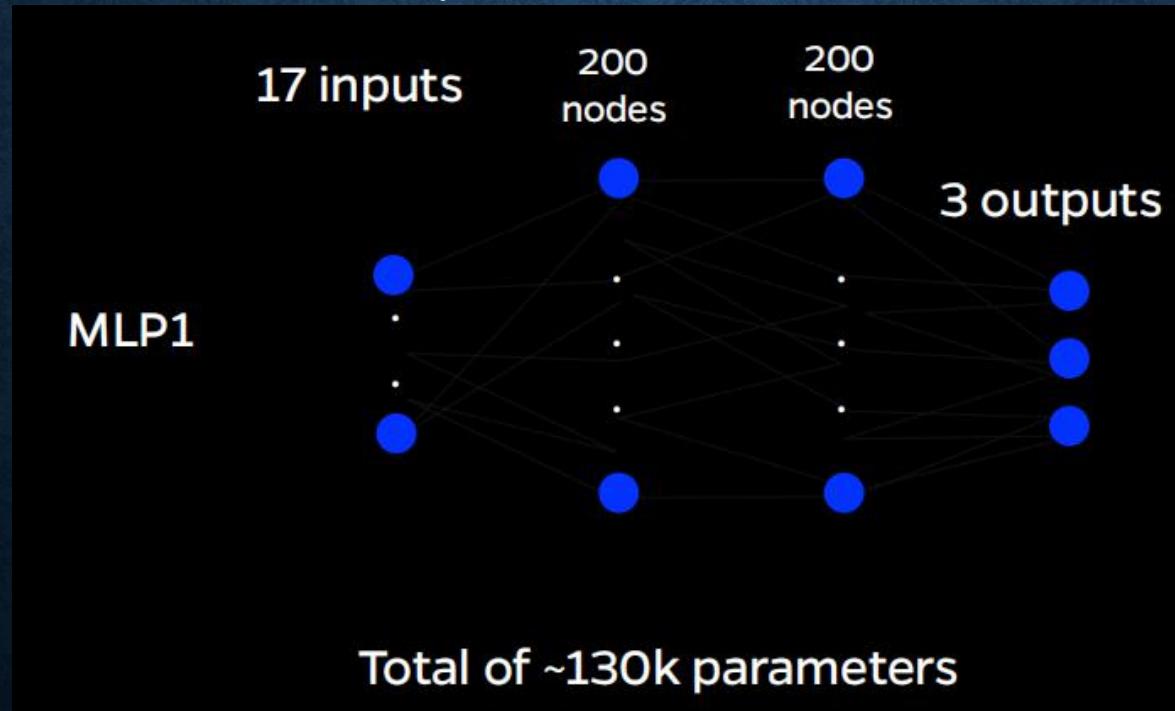
# Hidden layers in the MLPs

MLP1 had 17 inputs, 3 outputs

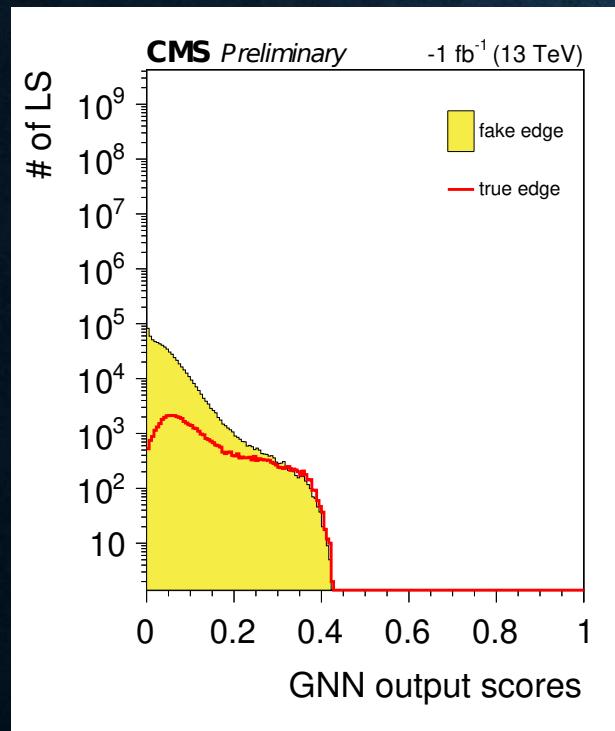
MLP2 had 10 inputs, 3 outputs

MLP3 had 9 inputs, 1 output

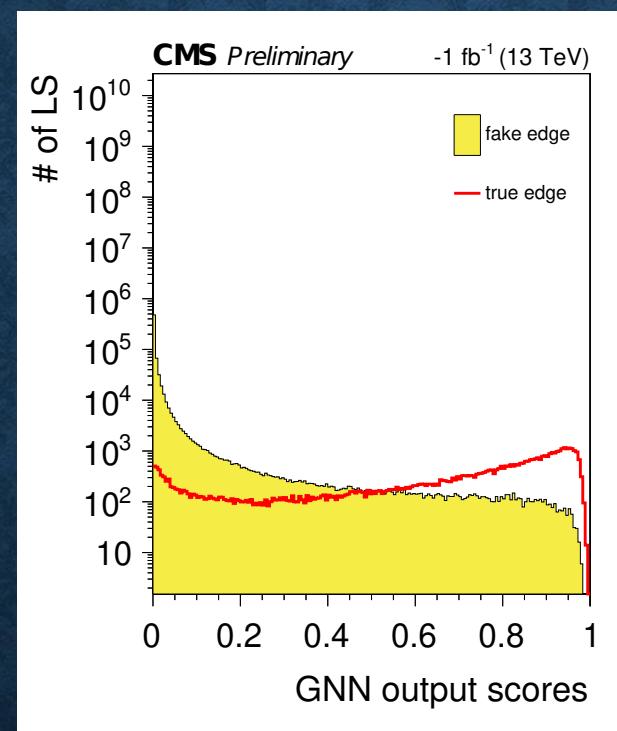
Each had 2 layers with hidden nodes of 200



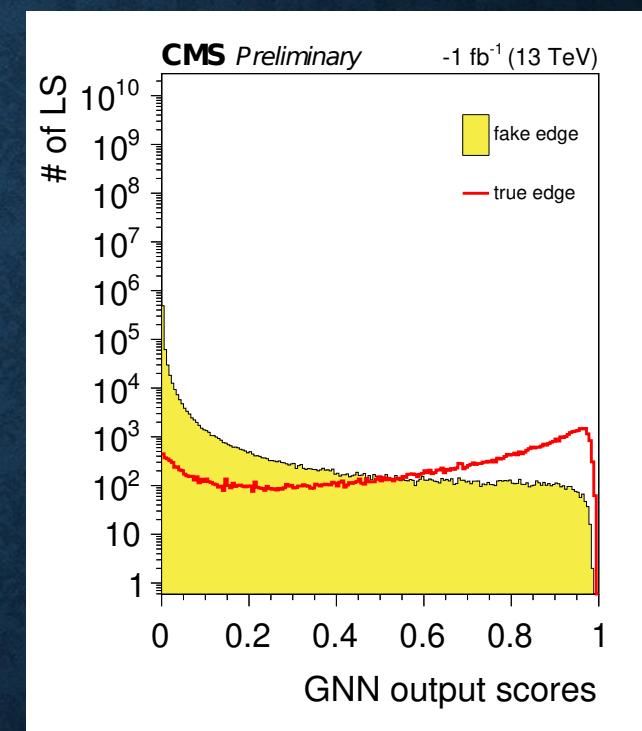
# Edge Output Distribution



Epoch 1



Epoch 25

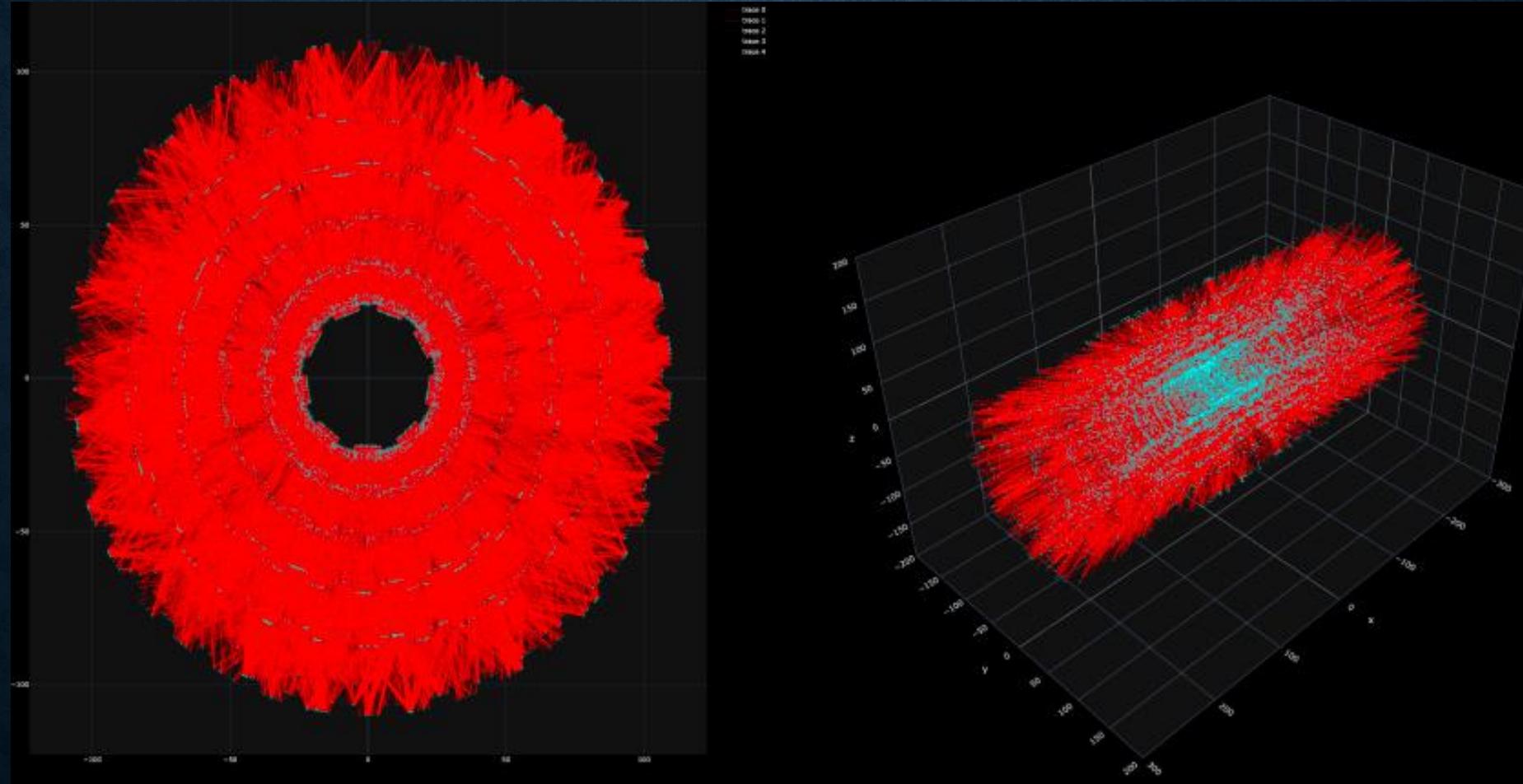


Epoch 50

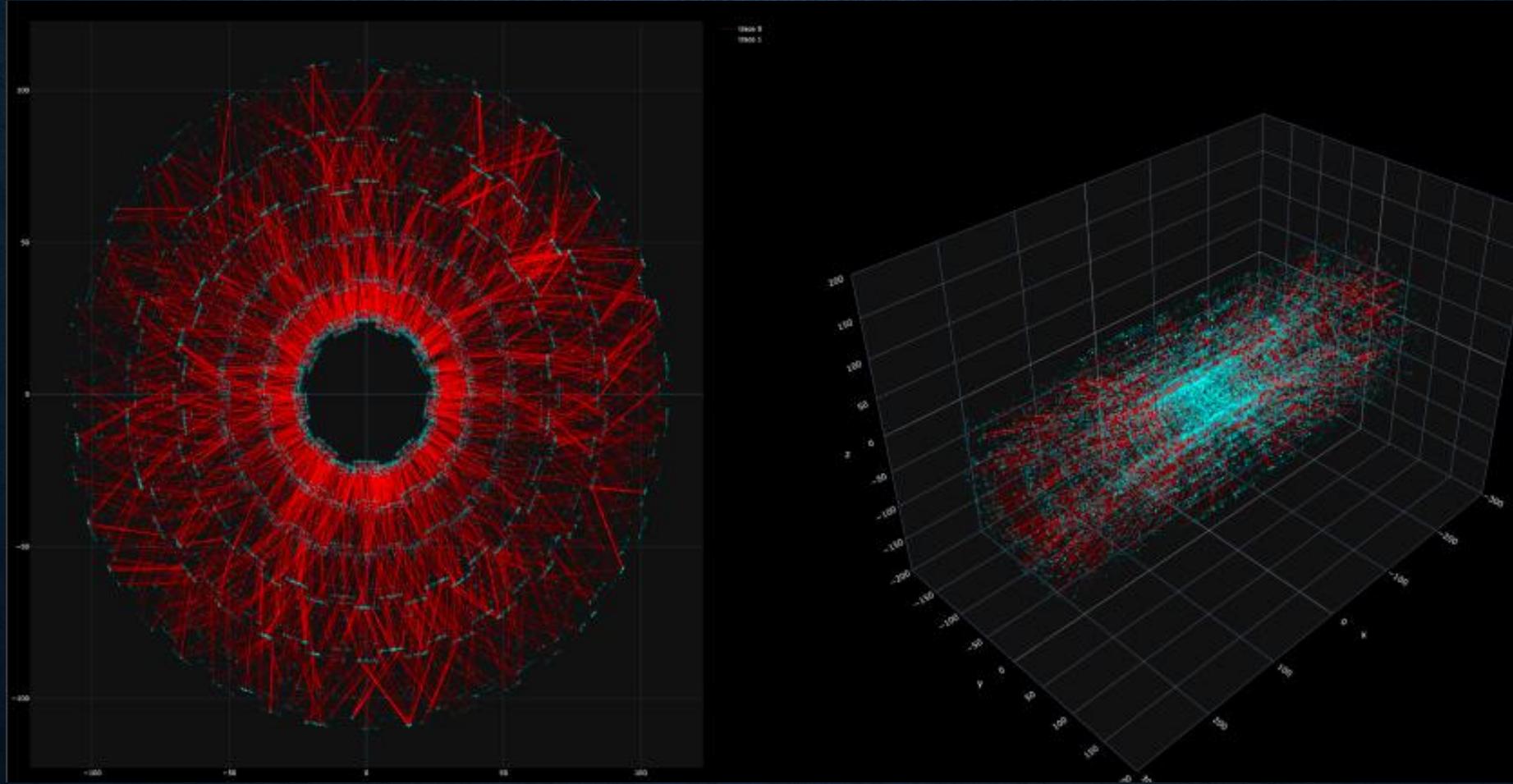
# Inference Testing

- Ran inference on separate Event (still 200 PU)
- Made cut at 0.0774
  - True positive rate = True negative rate
  - Future work will involve optimizing this cut value

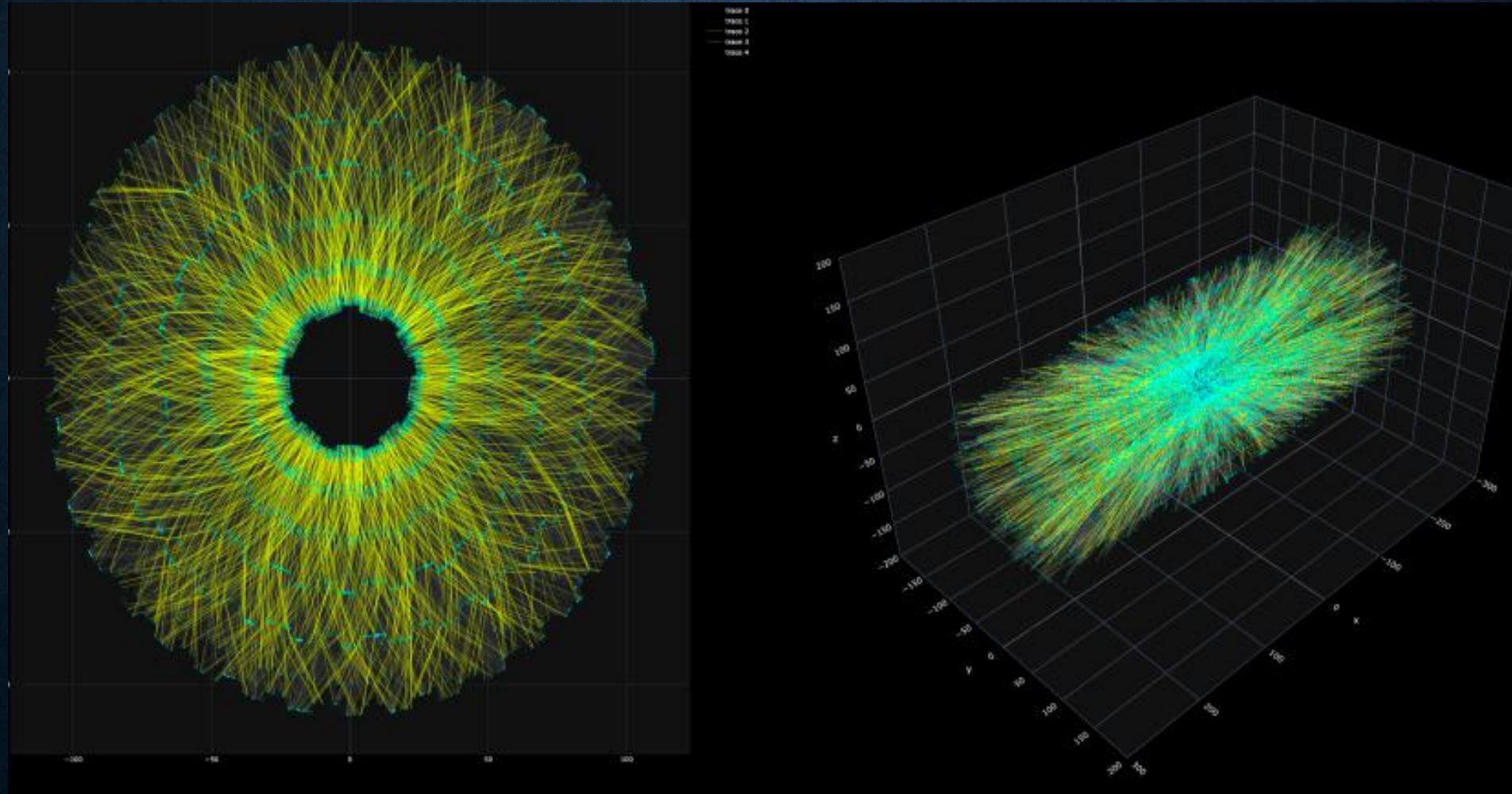
# Fake Edge Reduction - Before



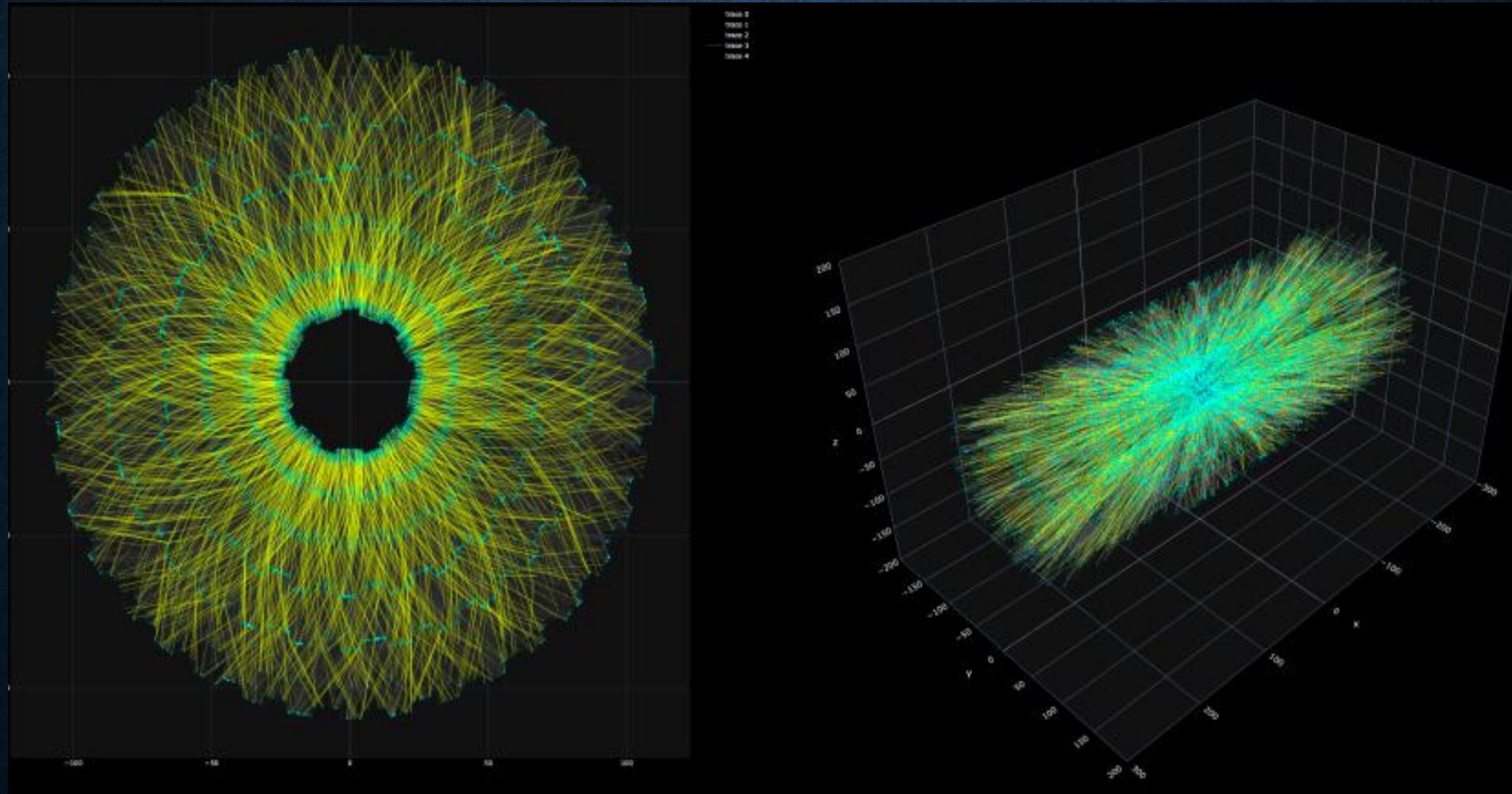
# Fake Edge Reduction - After



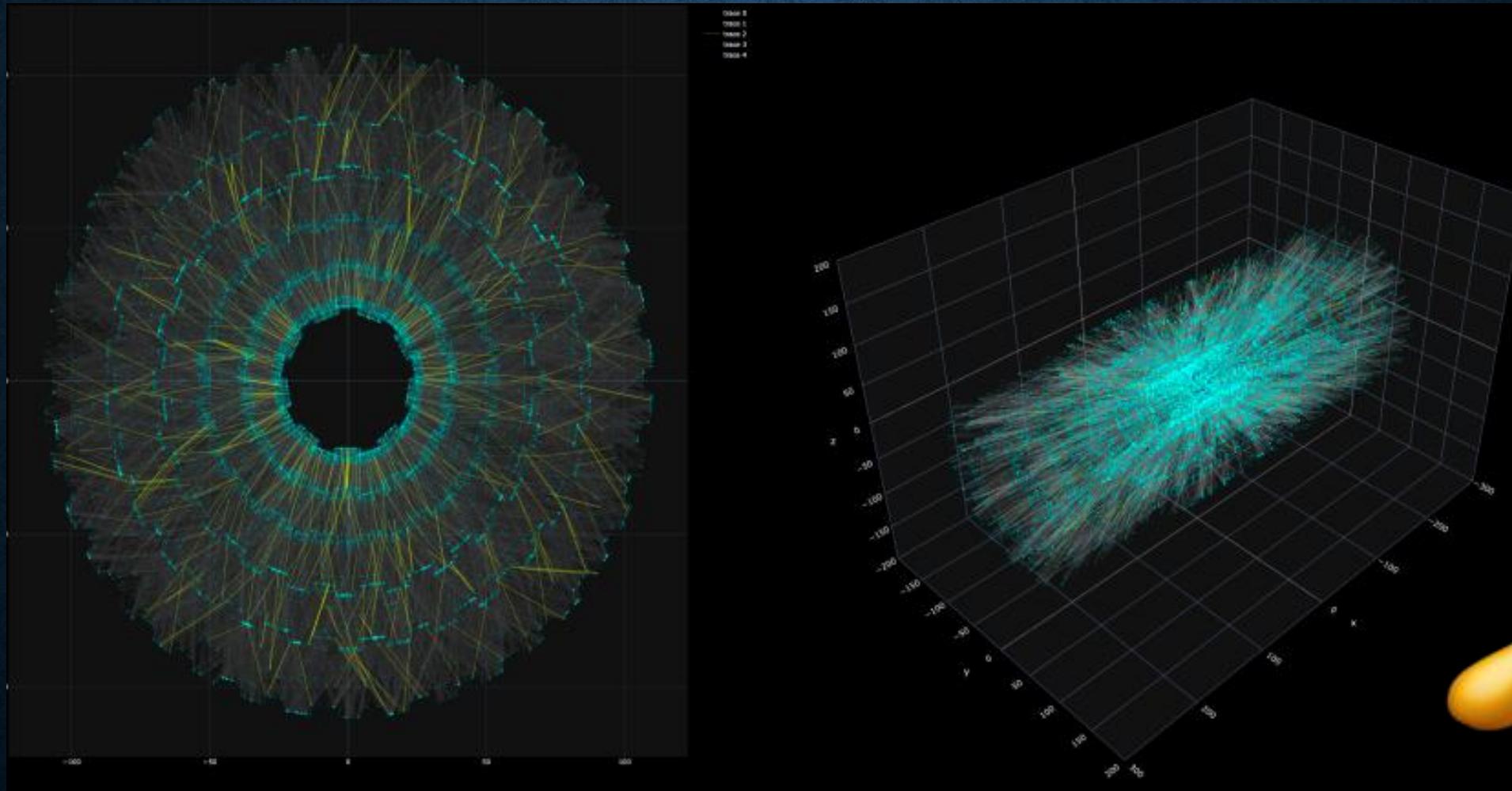
# True Edge Retention - Before



# True Edge Retention - After



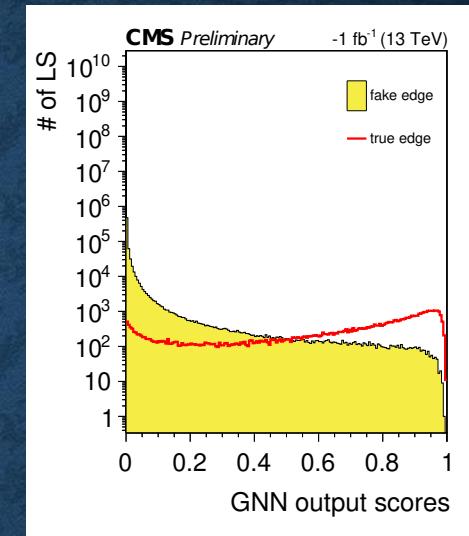
# True Edge Removal



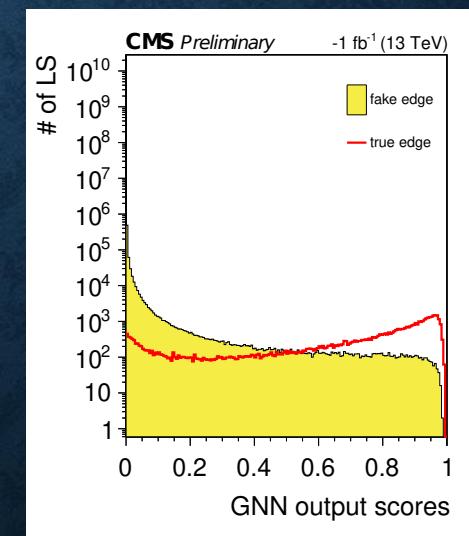
Cut threshold 0.0774	All	Passing GNN cut	Failing GNN cut		
All Mini-Doublets	41459				
Total Line Segments	144245	19705	124540		
True Line Segments	10549	9844	705	7%	loss
Fake Line Segments	133696	9861	123835	93%	reduction

# Future Work

- Optimizing the cut value
- Loss function is still quite high
  - MLP layer messages?
  - Normalizing data?
  - Learning Rate?
- Do we need every LS for reconstruction? – NO
  - LST is getting >99% with ~ 4000 tracks
  - Are we getting the *right* ones?



- 100 Epoch
- LR 0.001



- 50 Epoch
- LR 0.005

# Outline

