Line Segment Tracking GNN Optimization

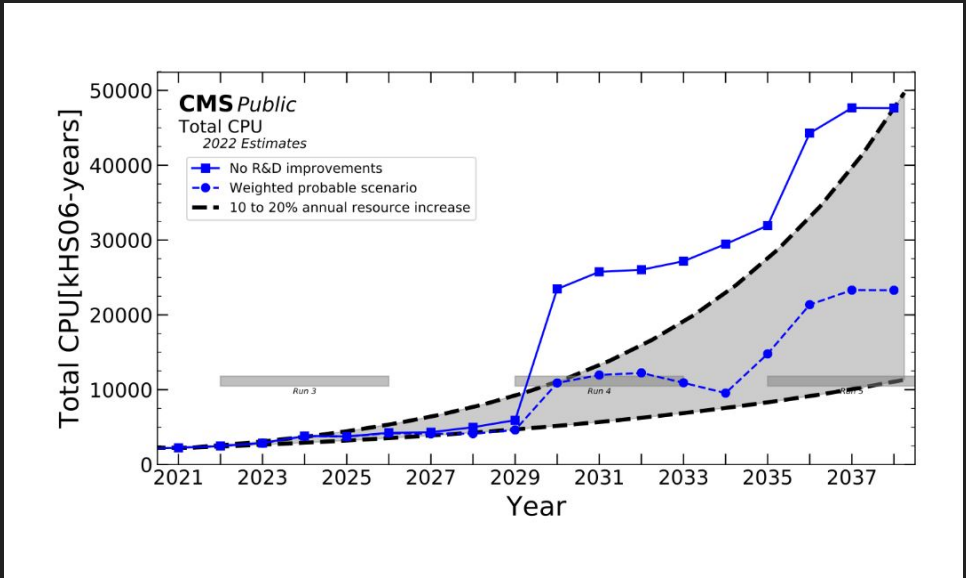
by Hubert Pugzlys, Dr. Chang

Introduction

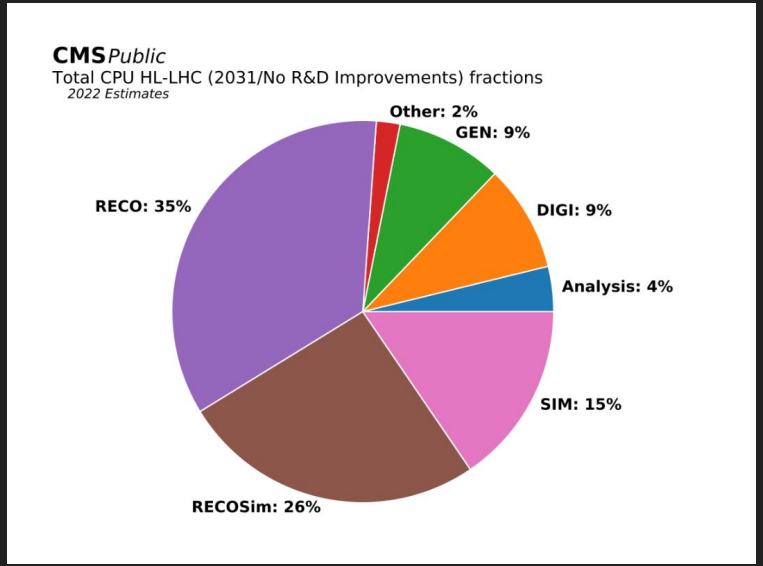
Outline

1. Background
2. Project Synopsis
3. Findings + Conclusion

HL-LHC Resource Cost

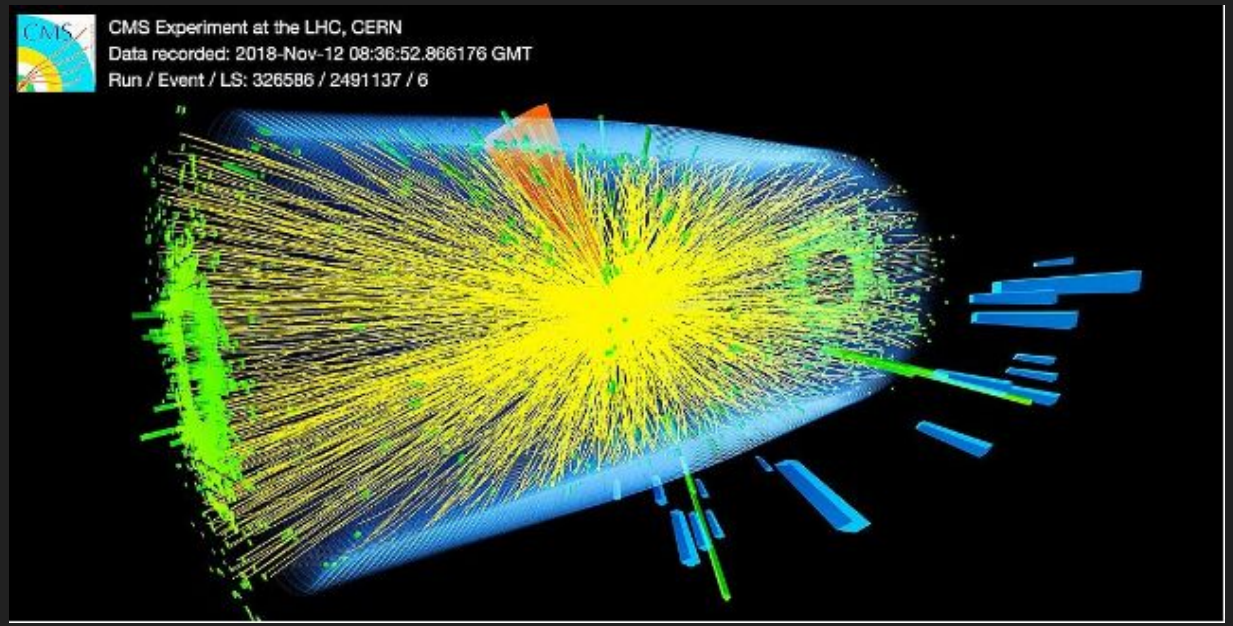


https://twiki.cern.ch/twiki/pub/CMSPublic/CMSOfflineComputingResults/cpu_cms2022.png



https://twiki.cern.ch/twiki/pub/CMSPublic/CMSOfflineComputingResults/cpu_pie_cms2022.png

Track Reconstruction

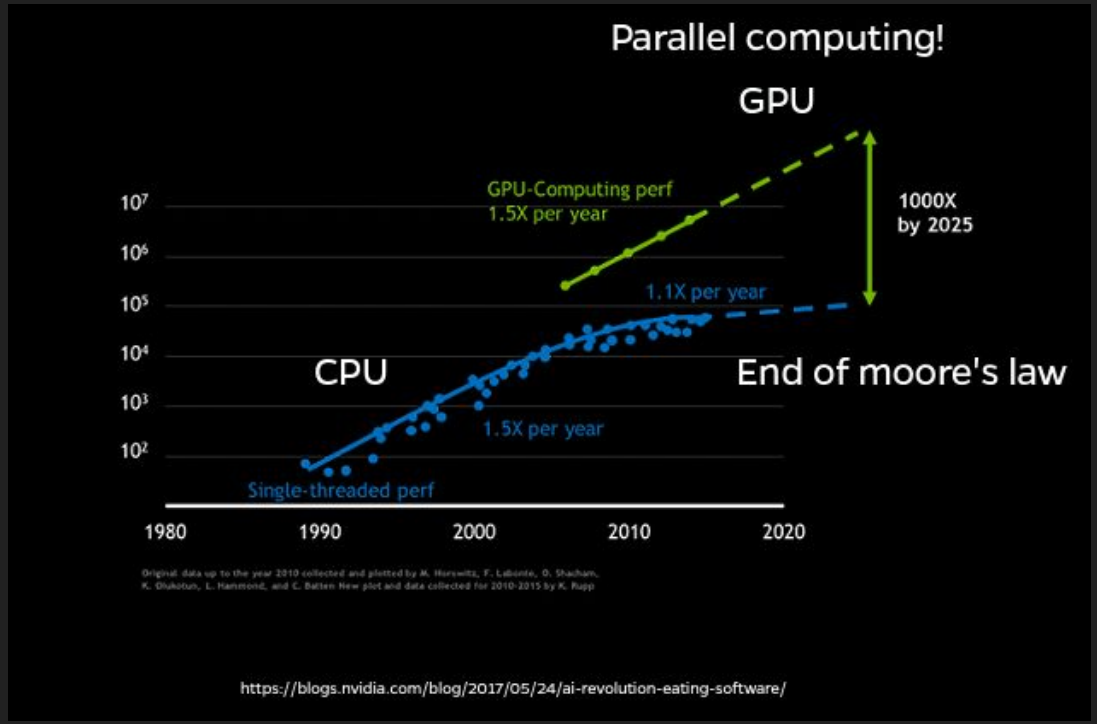


https://www.science-photo.de/bilder/12971111-Lead-ion-collision-event-in-CERN-s-CMS-detector

What do we do?

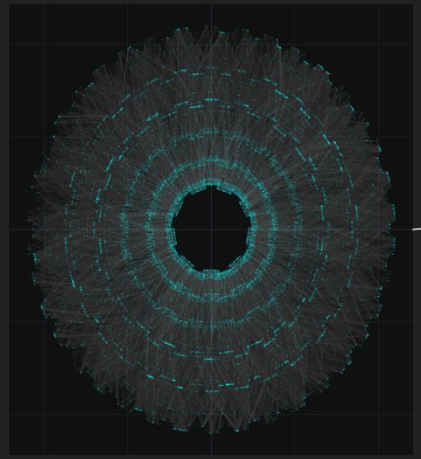
Parallelizable Solution

- Combinatorics increasing exponentially
- CPU computing power plateauing
- GPU parallelization favorable



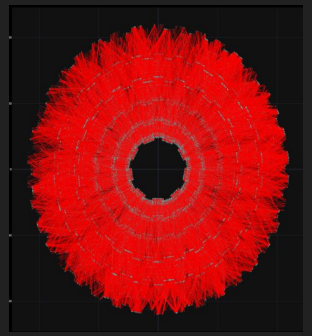
ML Parallelized Line-Segment Tracking

Possible Line Segments

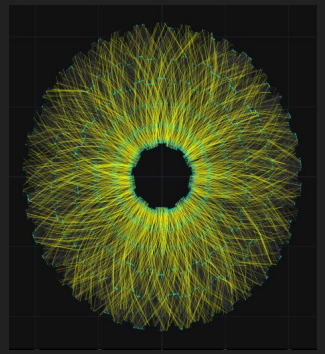


Graph Neural Network

Fake Segments

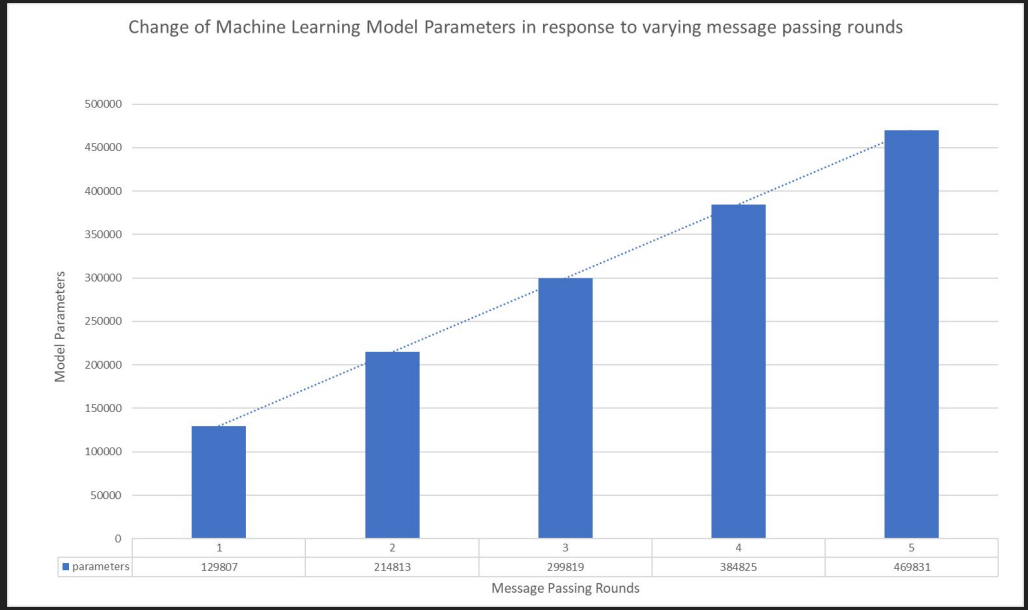


Real Segments



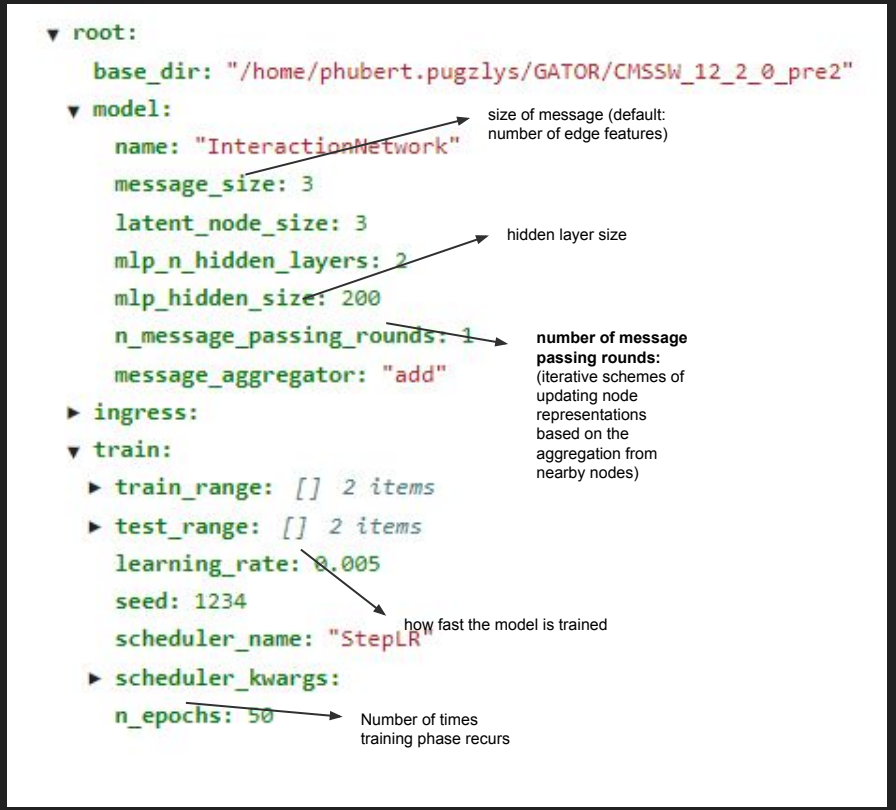
ML Optimization

- Parameters are the configuration variables that are internal to the model and whose value can be optimized by the machine learning model during the training phase of the algorithm



ML Hyperparameters

- hyperparameters, the explicit external parameters, such as learning rate and hidden neural network layers can have drastic effects on the performance of a machine learning algorithm
  - Goal: improve the accuracy of machine learning model
- can be managed and edited via JSON file

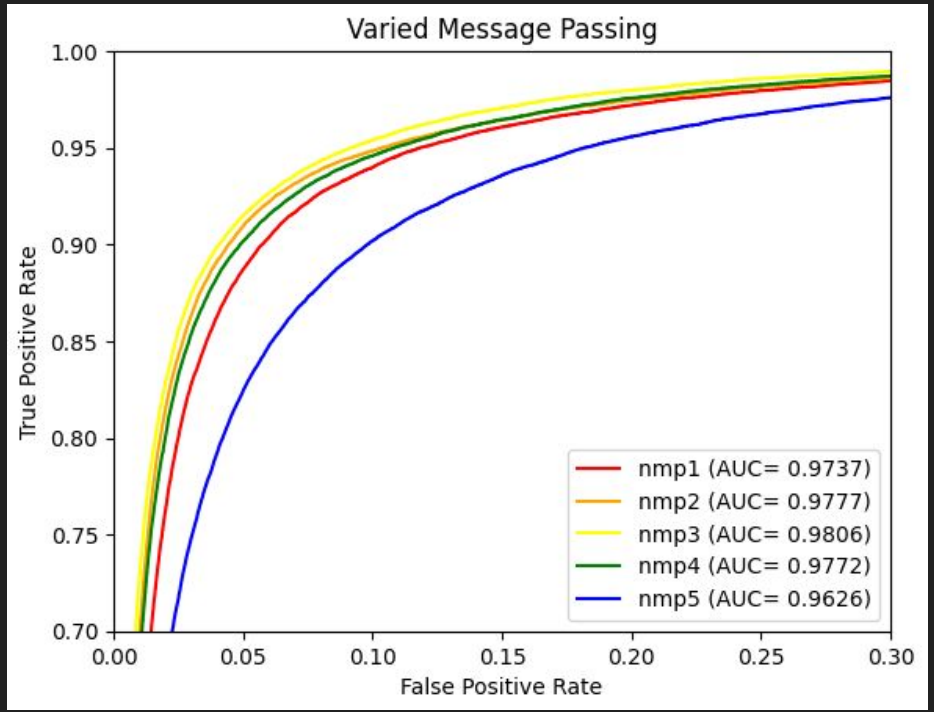


Findings

ROC Curve of Varying Message Passing Rounds

Receiver Operating Characteristic Curve demonstrates the performance of a machine learning model

NMP3 had the best performance while NMP5 had the worst.



Quantitative Performance

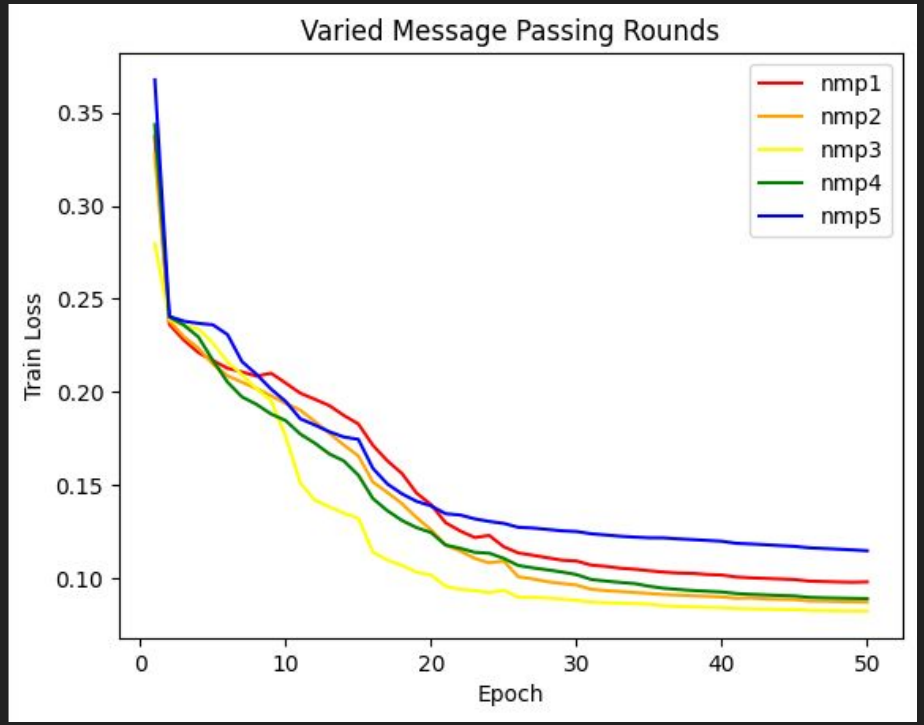
False Negative Rate	True Positive Rate (NMP1)	True Positive Rate (NMP3)
0.01	0.6286	0.7347
0.02	0.7661	0.8306
0.03	0.8292	0.8753
0.04	0.8639	0.8991
0.05	0.8875	0.915
0.06	0.904	0.922
0.07	0.916	0.931
0.08	0.926	0.938
0.09	0.933	0.943
0.1	0.939	0.948

The nmp hyperparameter optimization increase the performance of the machine learning model by 3% - 12%

Test Loss Curve of Varying Message Passing Rounds

Another visual performance metric for the machine learning model at each number of message passing rounds.

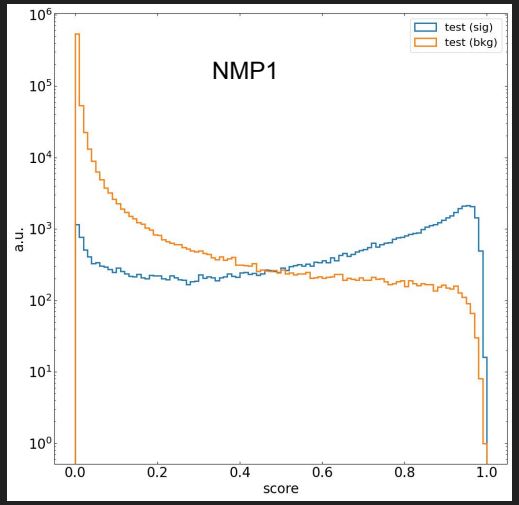
NMP3 had the best performance while NMP5 had the worst.

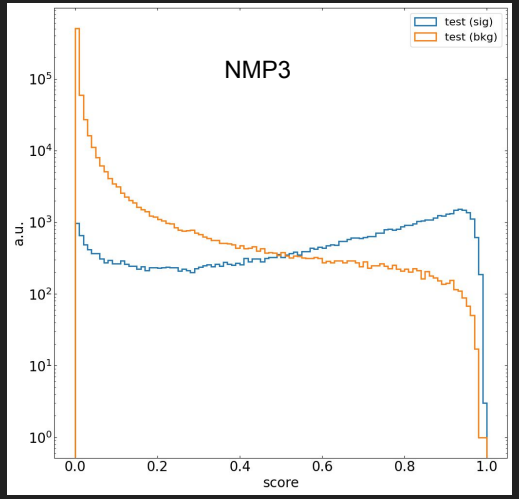


Conclusion

- N_message_passing_rounds set to 3 produces the best results for this GNN at the set hyperparameter configuration.

NMP1

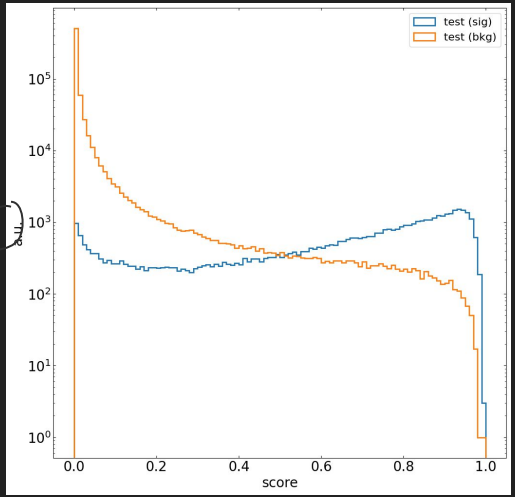




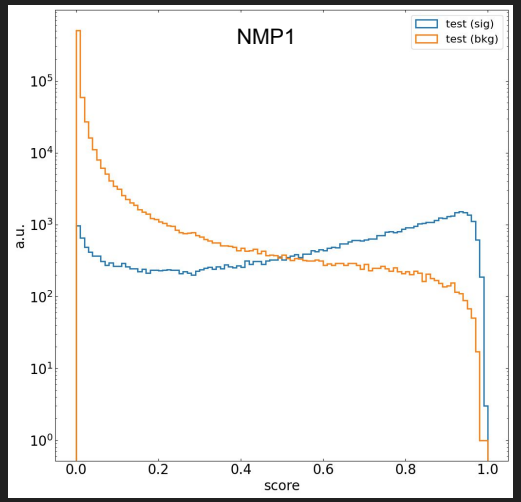
Future Research

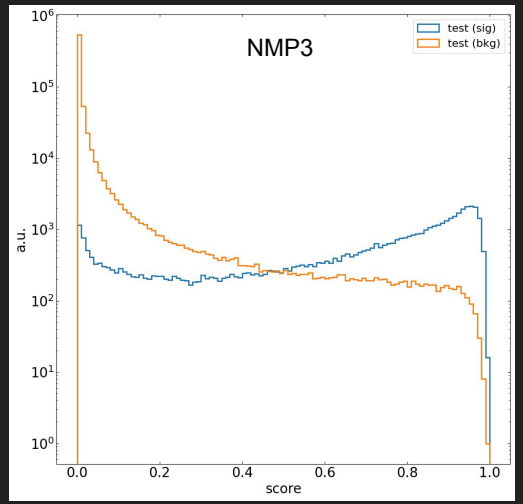
Future Research

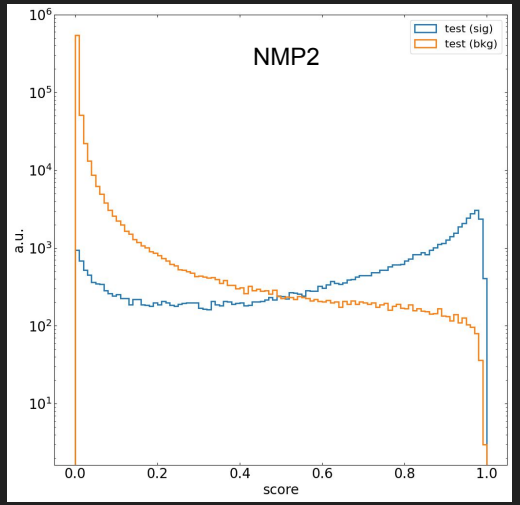
- optimization of other hyperparameters such as learning rate and number of hidden layers
- Identify and explain model anomalies (prev. slide)

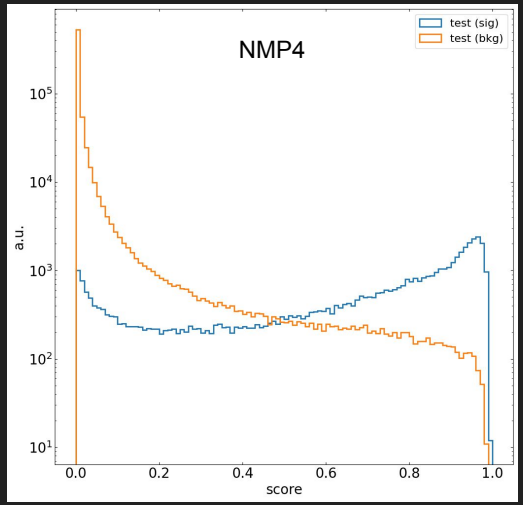


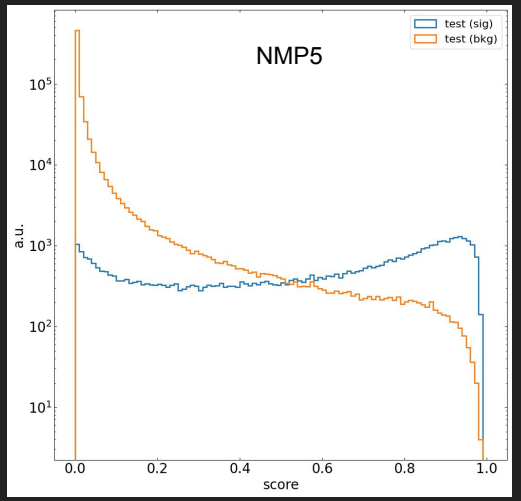
Score Distributions











Expanded Roc Curve