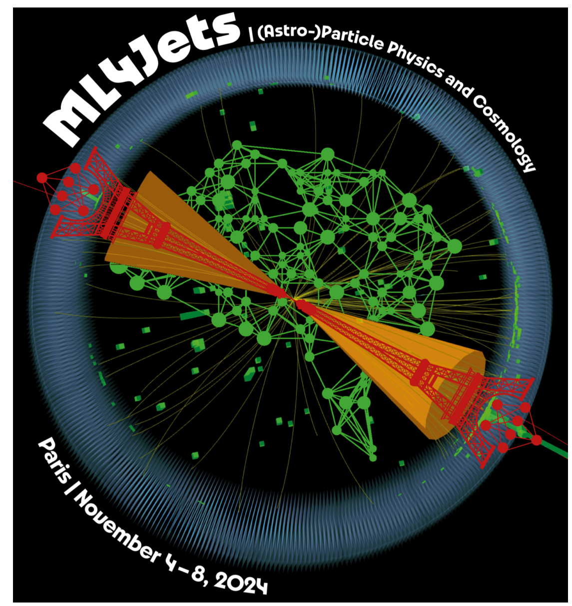
The Fundamental Limit of Jet Tagging

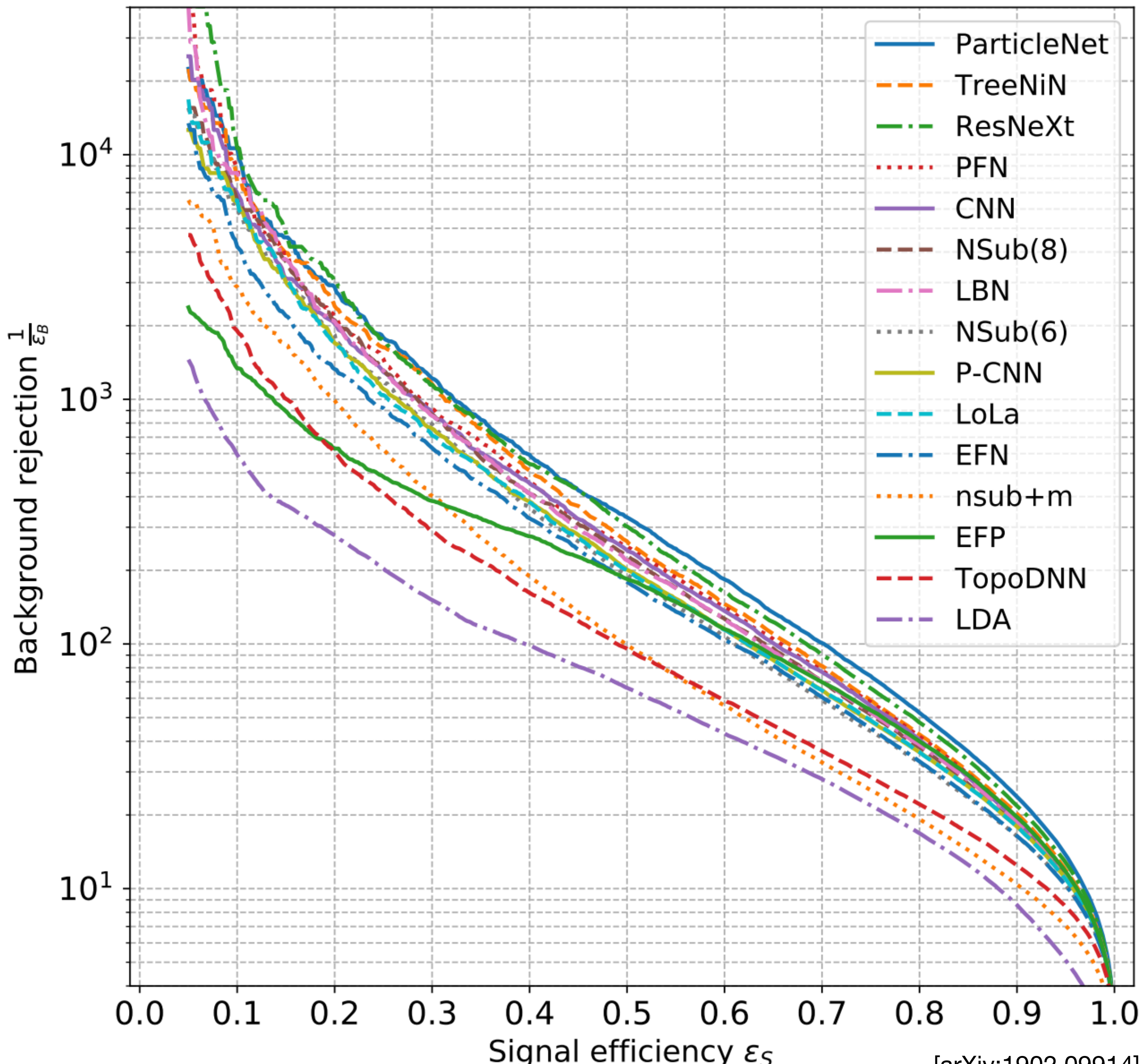
Nishank Gite (UC Berkeley)
Joep Geuskens (RWTH Aachen)
Michael Krämer (RWTH Aachen)
Vinicius Mikuni (NERSC)
Alexander Mück (RWTH Aachen)
Benjamin Nachman (LBNL/BIDS)
Humberto Reyes-González (RWTH Aachen)

November 11, 2024



Why?

- Modern machine learning has led to rapid progress in jet tagging
- **Image:** Model performance on Top vs QCD jets [1]
- **No Limit?**



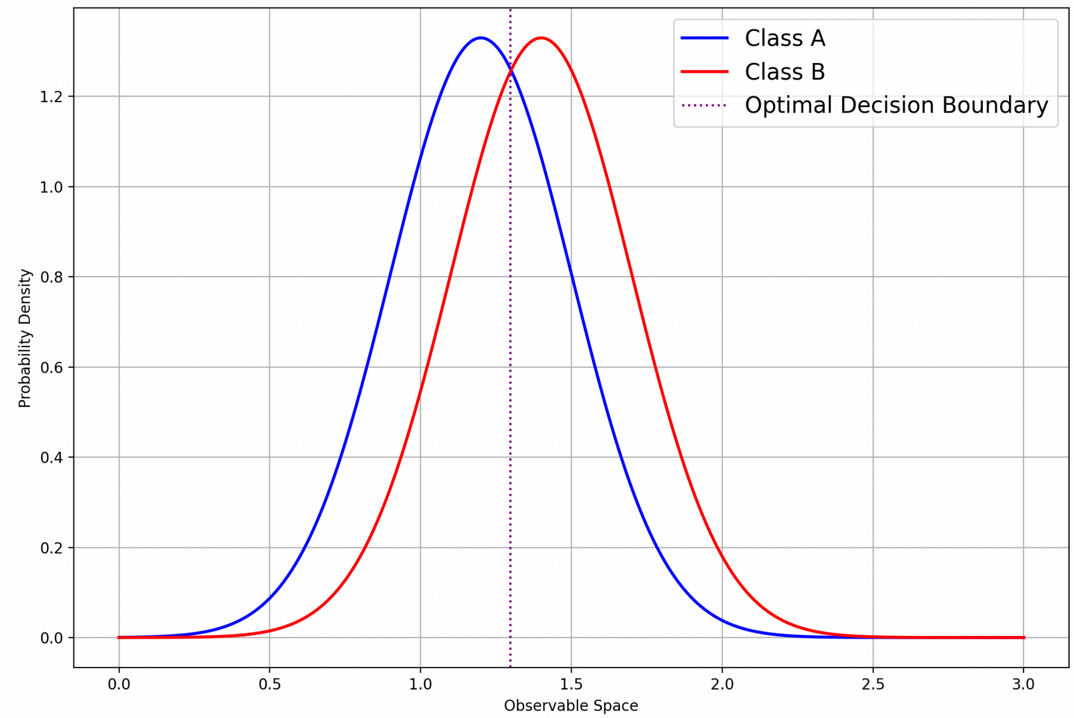
[arXiv:1902.09914]

Upper Bound?

- **Neymon-Pearson Lemma:** Given some features, x, and classes A and B, we determine the optimal decision boundary [2]:

$$\frac{p(x|A)}{p(x|B)} > \lambda$$

- $\frac{p(x|A)}{p(x|B)}$ is the likelihood ratio
- λ is your decision boundary

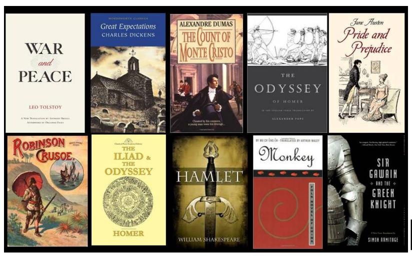


PROBLEM: We don't have these probability distributions

Learning Probabilities: Generative AI

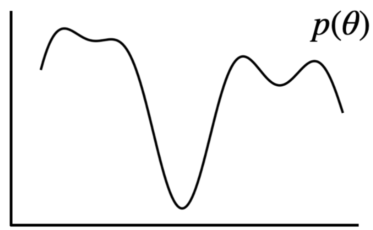
- Methods: Conditional Normalizing Flows, Diffusion, and **Autoregressive Models** → ChatGPT

Input Data



[3]

Autoregressive Model



Learns English Language!

Prompt this → In a distant ?

land - 40%
time - 25%
kingdom - 15%
past - 10%
memory - 5%
forest - 5%

$$p_{total} = p(In) \cdot p(a \mid In) \cdot p(distant \mid In, a) \cdot p(? \mid In, a, distant)$$

What does this look like for jets?

Raw Data

$p_T \in [0{,}500]$

$\Delta\eta_{jet} \in [-0.8{,}0.8]$

$\Delta\phi_{jet} \in [-0.8{,}0.8]$

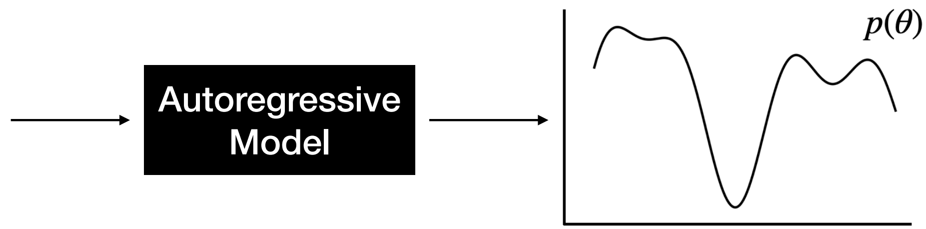
Discretize data by binning it, making it a QCD dictionary

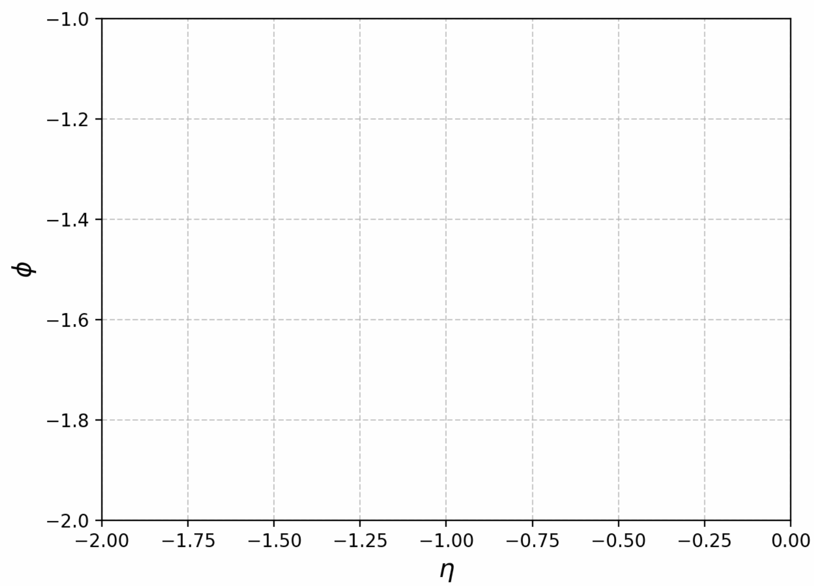
Input Data

$p_T \in [0{,}40]$

$\Delta\eta_{jet} \in [0{,}30]$

$\Delta\phi_{jet} \in [0{,}30]$





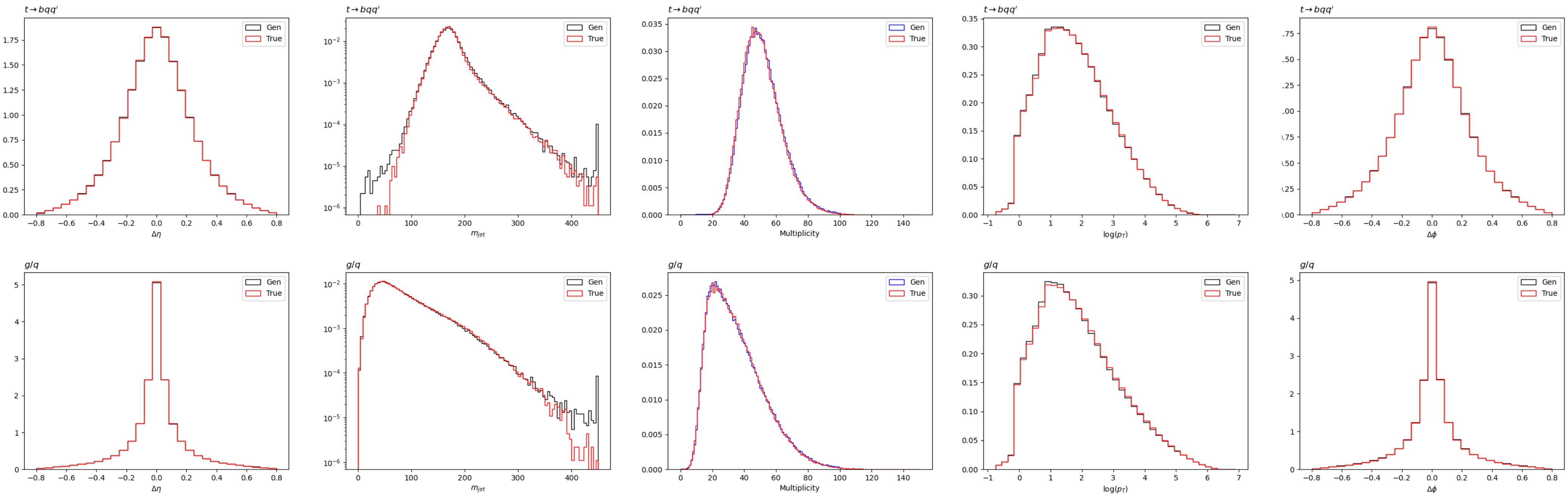
Now to sample from this probability

$P(\text{jet}) = 1$

Initializing...

GOAL: Create synthetic jet data that mirrors the real BUT we know the likelihood ratio

Generated Data Matches Real Data?



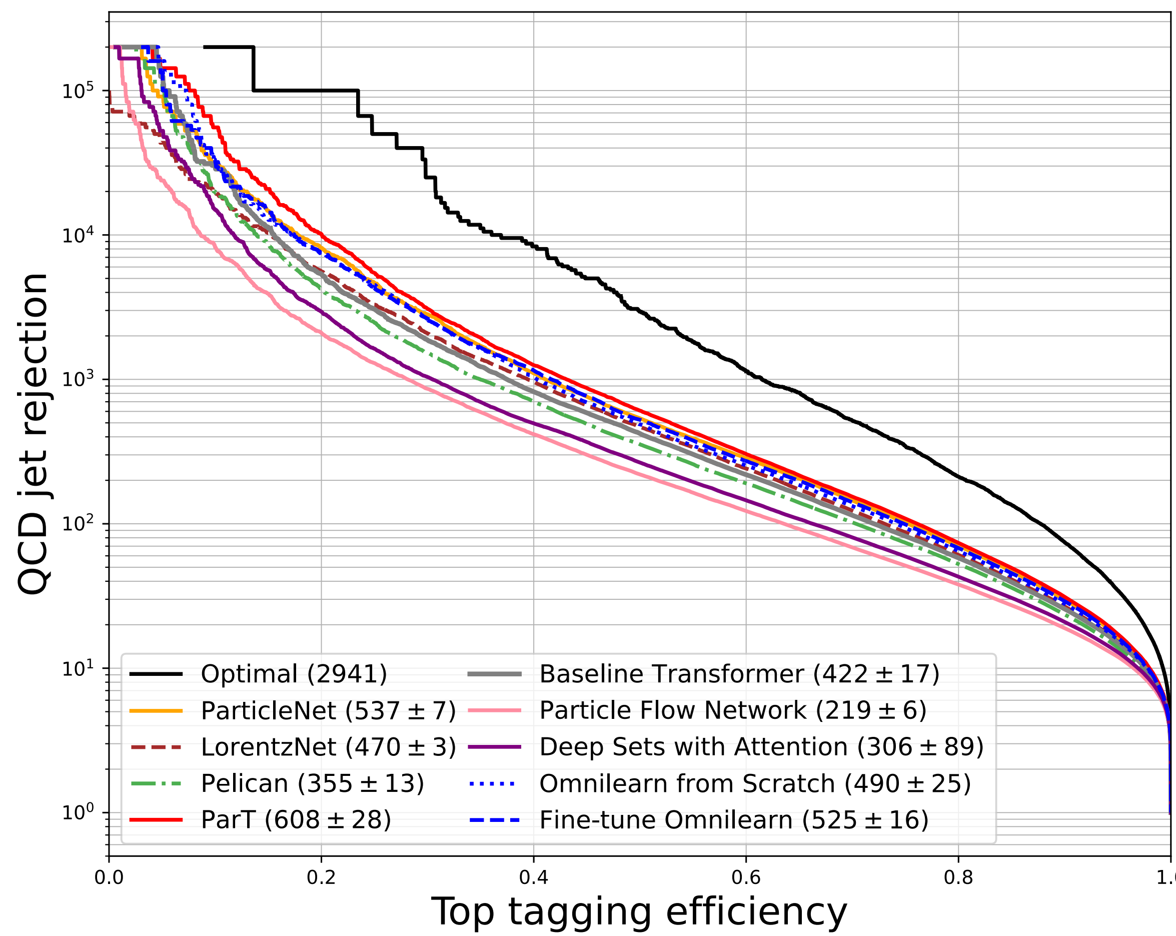
- Model Learned the underlying physics [4]
- We now have the probabilities!

Let's see this likelihood ratio curve compared to current models

Current State of Jet Tagging

- Optimal (LLR) in black
- 5 run averaged ROC curves for various models trained and tested on **MODEL** generated data
- Rejection at 50% Top tagging efficiency in parentheses

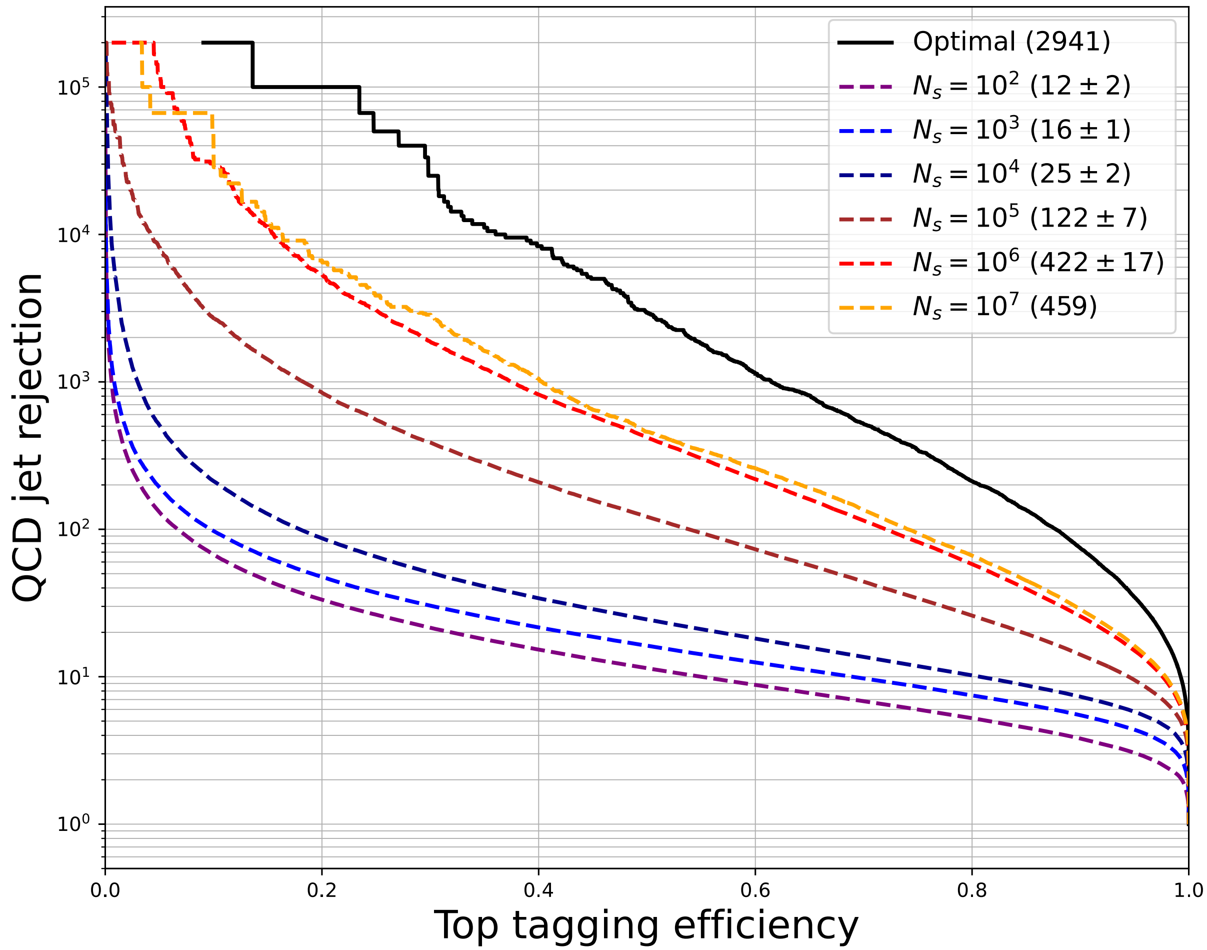
Now let's focus on Aachen's Baseline Transformer and see what happens as we vary sample size



Sample Size vs. Performance

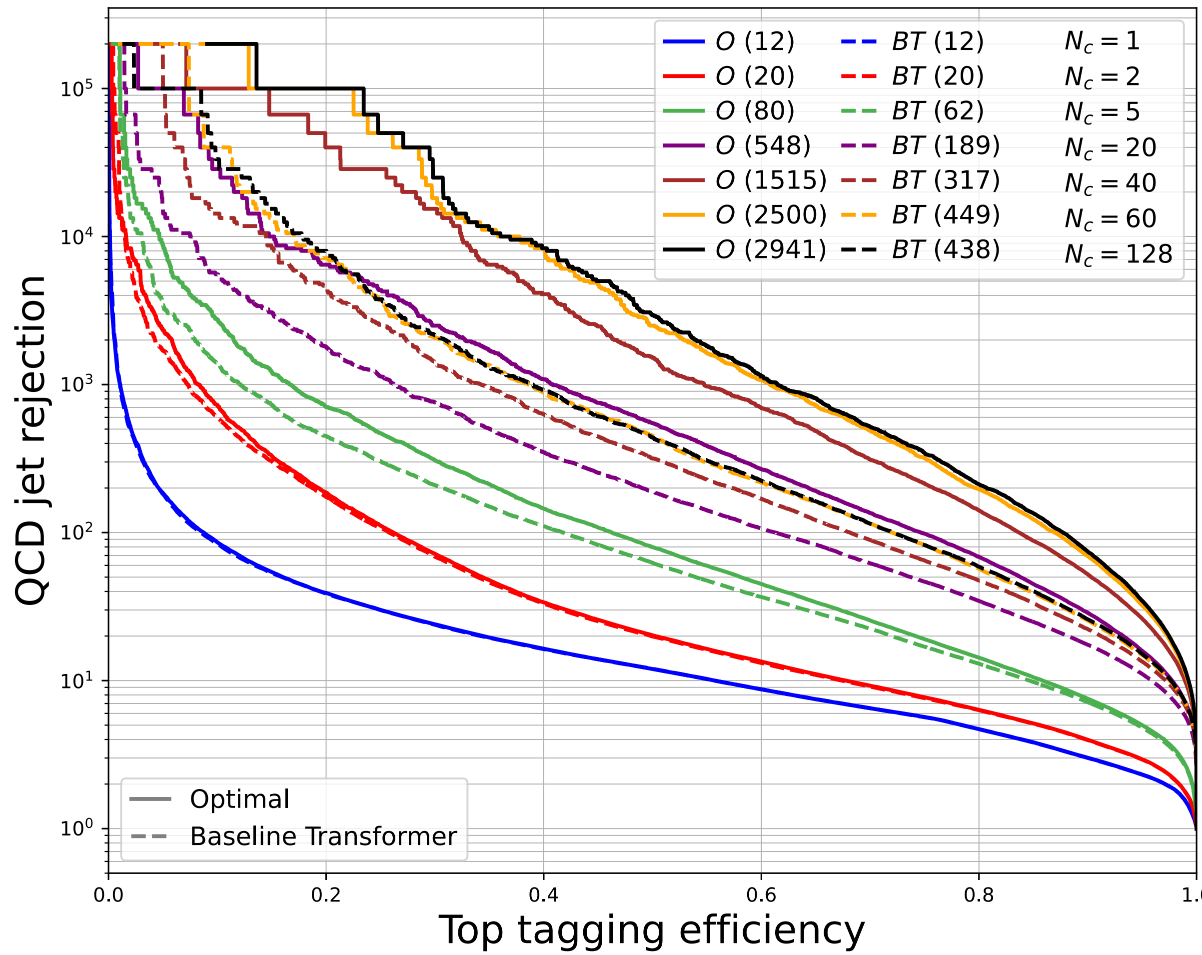
- Averaged Aachen Baseline transformer (5 runs) trained on various sample sizes and tested on the same data to compare performance
- Rejection at 50% Top tagging efficiency in parentheses

Do we achieve optimal performance for jets with lower particle multiplicity?



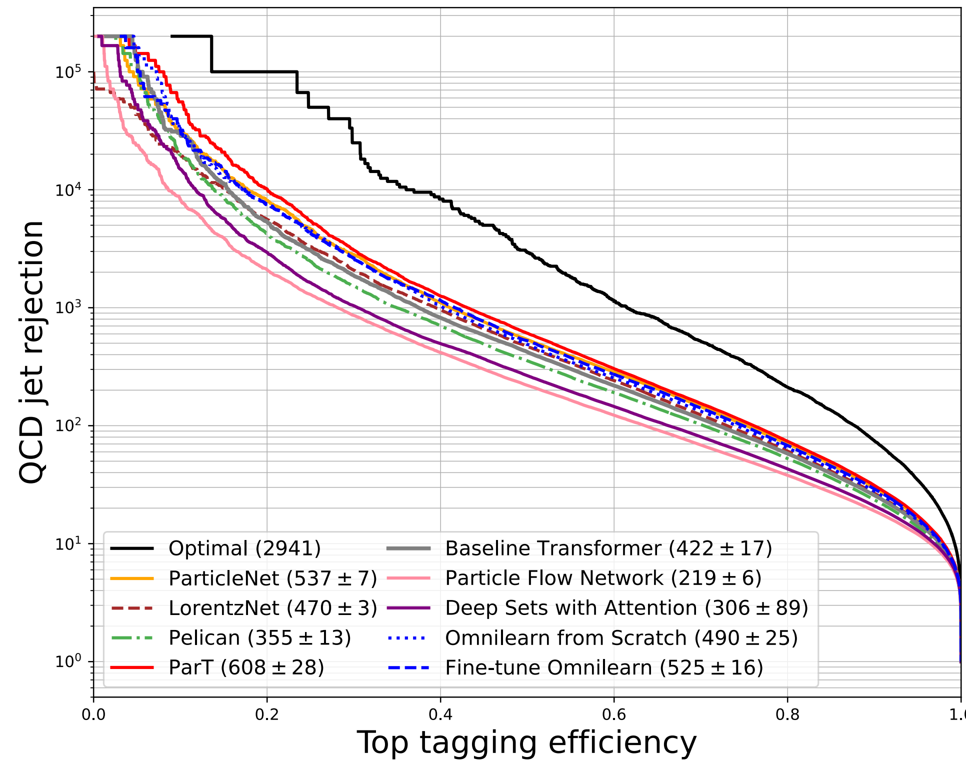
Number of Particles vs. Performance

- Optimal Curves and color associated Aachen Baseline Transformer for various numbers of particle constituents
- Rejection at 50% Top tagging efficiency in parentheses



Conclusion

- Used autoregressive models to create synthetic highly realistic jet data
- Found a significant gap between current performance and the theoretical optimum
- Determined the gap cant be closed by increasing train size
- We should continue developing stronger taggers!
- Data is Public: https://zenodo.org/records/14023638
- Fundamental Limit Paper: https://arxiv.org/abs/2411.02628
- Autoregressive Model Paper: https://arxiv.org/abs/2303.07364



References

[1] G. Kasieczka, et. al., "The Machine Learning Landscape of Top Taggers," SciPost Phys. 7, 014 (2019). DOI: 10.21468/SciPostPhys.7.1.014

[2] J. Neyman and E. S. Pearson, Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 231, 289 (1933).

[3] Classical Poets. (2017, January 1). 10 greatest novels ever written. Classical Poets. https://classicalpoets.org/2017/01/01/10-greatest-novels-ever-written/

[4] T. Finke, M. Krämer, A. Mück, and J. Tönshoff, "Learning the language of QCD jets with transformers," JHEP (2023). DOI: 10.1007/JHEP06%282023%29184