Learning the language of QCD jets with transformers

Alexander Mück
RWTH Aachen University

together with

Thorben Finke, Michael Krämer, Jan Tönshoff

based on JHEP 06 (2023) 184, 2303.07364

ML4Jets2023

DESY, Hamburg, November 7, 2023

Institute for Theoretical Particle Physics and Cosmology

Outline

Learning the language of QCD jets with transformers

- Data
  - QCD and top jets
  - Turning particle and jets into words and sentences

Institute for Theoretical Particle Physics and Cosmology

Outline

Learning the language of QCD jets with transformers

- Data
  - QCD and top jets
  - Turning particle and jets into words and sentences
- Density estimation
  - on low level data
  - using transformers as in NLP

Institute for Theoretical Particle Physics and Cosmology

Outline

Learning the language of QCD jets with transformers

- Data
  - QCD and top jets
  - Turning particle and jets into words and sentences
- Density estimation
  - on low level data
  - using transformers as in NLP
- Quality assessment
  - use transformer as generative model
  - classifier tests

Data

standard benchmark data set:

- QCD and top jets: "Top Quark Tagging Reference Dataset"
  https://zenodo.org/records/2603256
- 600k jets each
- $p_T \in [550, 650]$ GeV $\rightarrow$ boosted top jets
- up to 200 constituent four-momenta $\Rightarrow$ low-level data

Data

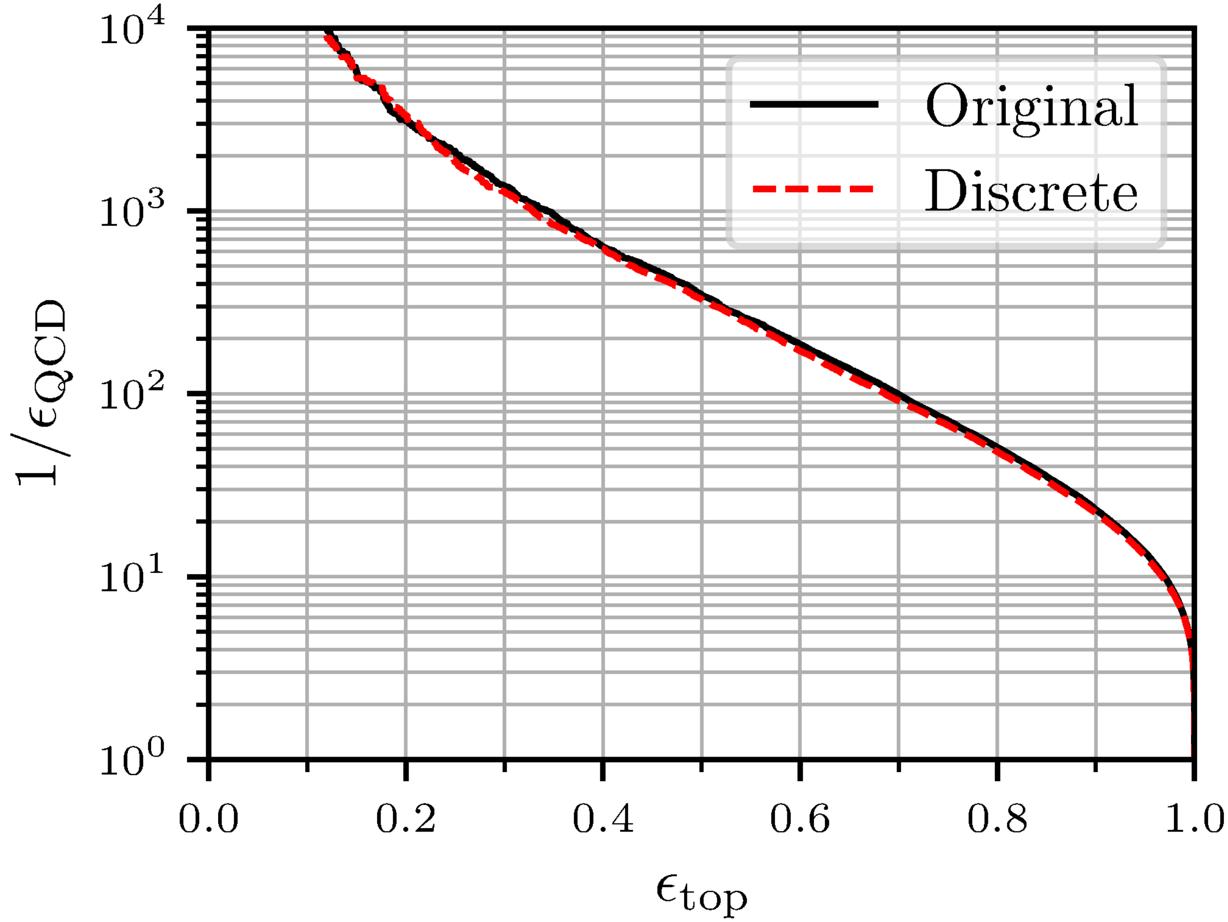
standard benchmark data set:

- QCD and top jets: "Top Quark Tagging Reference Dataset"
  https://zenodo.org/records/2603256
- 600k jets each
- $p_T \in [550, 650]$ GeV $\rightarrow$ boosted top jets
- up to 200 constituent four-momenta $\Rightarrow$ low-level data

special preprocessing:

- discretize constituent kinematics by binning:
  $\hat{p}_T \in [0, 40], \quad \Delta\hat{\eta} \in [0, 30], \quad \Delta\hat{\phi} \in [0, 30]$
- $\Rightarrow \sim$ 40k different constituents $\Leftrightarrow$ words in our language
- jets $\Leftrightarrow$ sentences in our language
- loss of information?

Data



⇒ almost no information loss for top tagging with ParticleNet

Institute for Theoretical Particle Physics and Cosmology

Density Estimation

- autoregressive approach:

$$p(\vec{x}) = p(\vec{x}_1)p(\vec{x}_2|\vec{x}_1)\dots p(\vec{x}_n|\vec{x}_1\dots\vec{x}_{n-1})$$

(see also JUNIPR by Andreassen et al., 1804.09720)

- standard transformer architecture as in NLP

(see also Trade by Fakoor et al., 2004.02441)

- simple grammar: order constituents by p_T

(more physics inspired grammar in 1804.09720)

- loss function: categorical cross entropy between

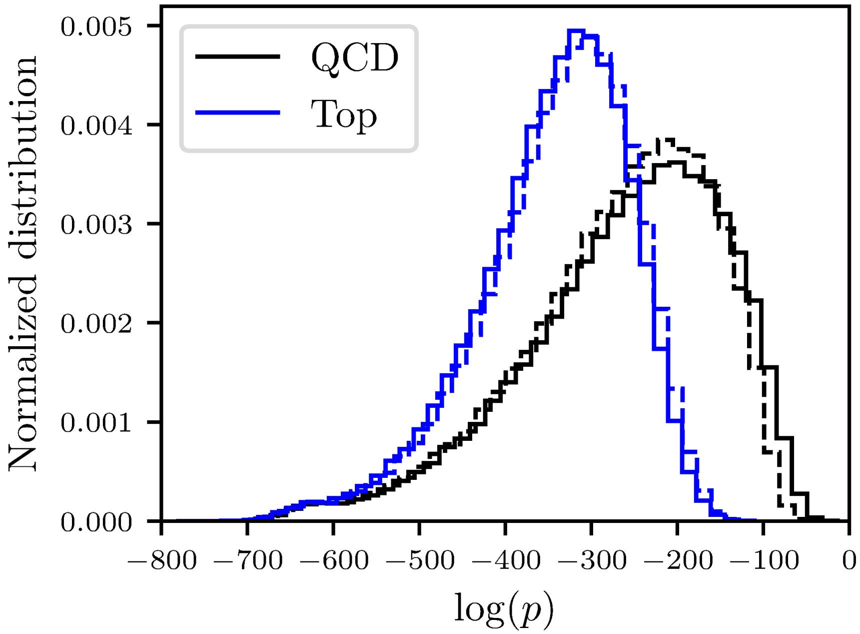
predicted and actual next particle/word

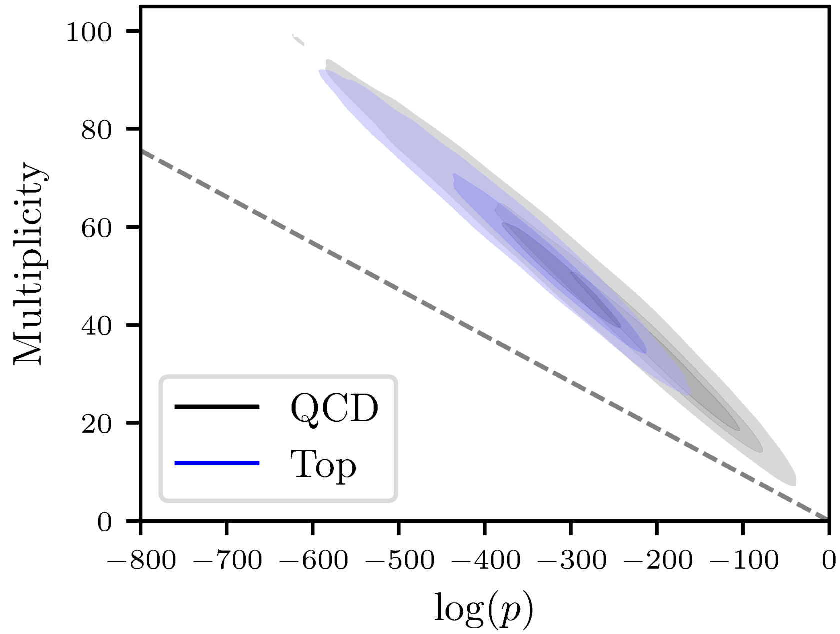
- jet multiplicity: predict stop token to terminate jet

$\Rightarrow$ variable multiplicity

Estimated Density

- probability for QCD and top jets:





solid: trained on QCD jets $\Rightarrow p_{\mathrm{QCD}}(x)$

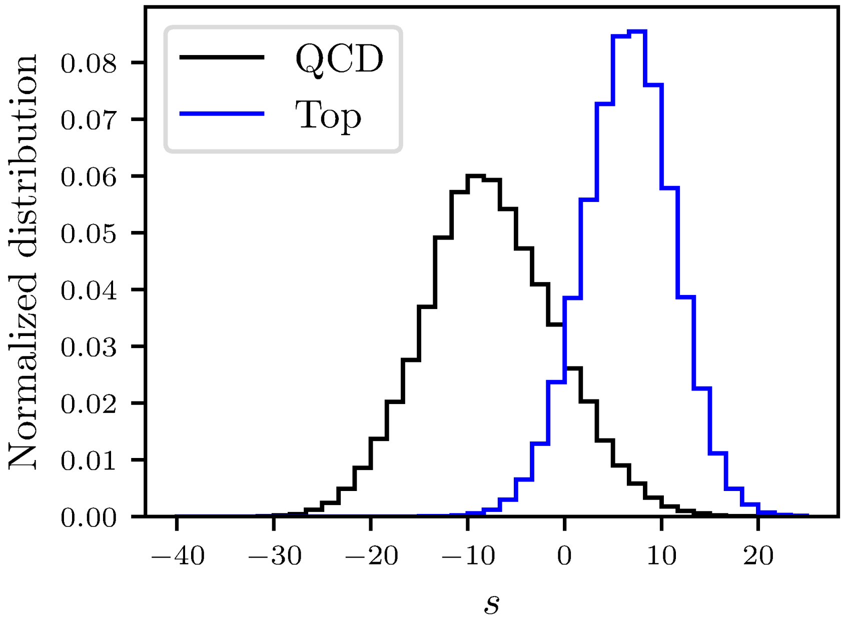
dashed: trained on top jets $\Rightarrow p_{\mathrm{top}}(x)$

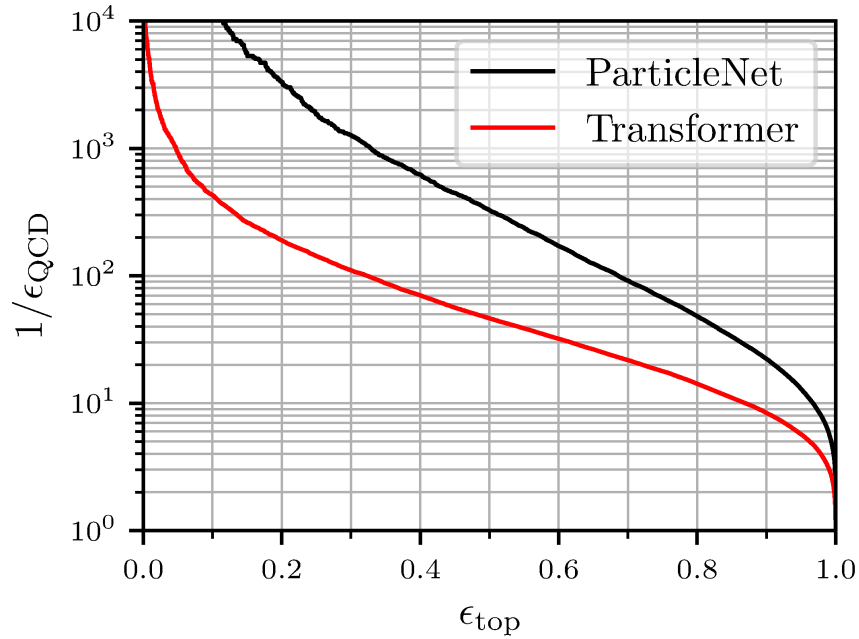
- dominated by particle multiplicity
- indeed depends on specific training data

Quality assessment

- Use density ratio as classifier score:

$$s = \log(p_{\text{top}}(x)) - \log(p_{\text{QCD}}(x))$$





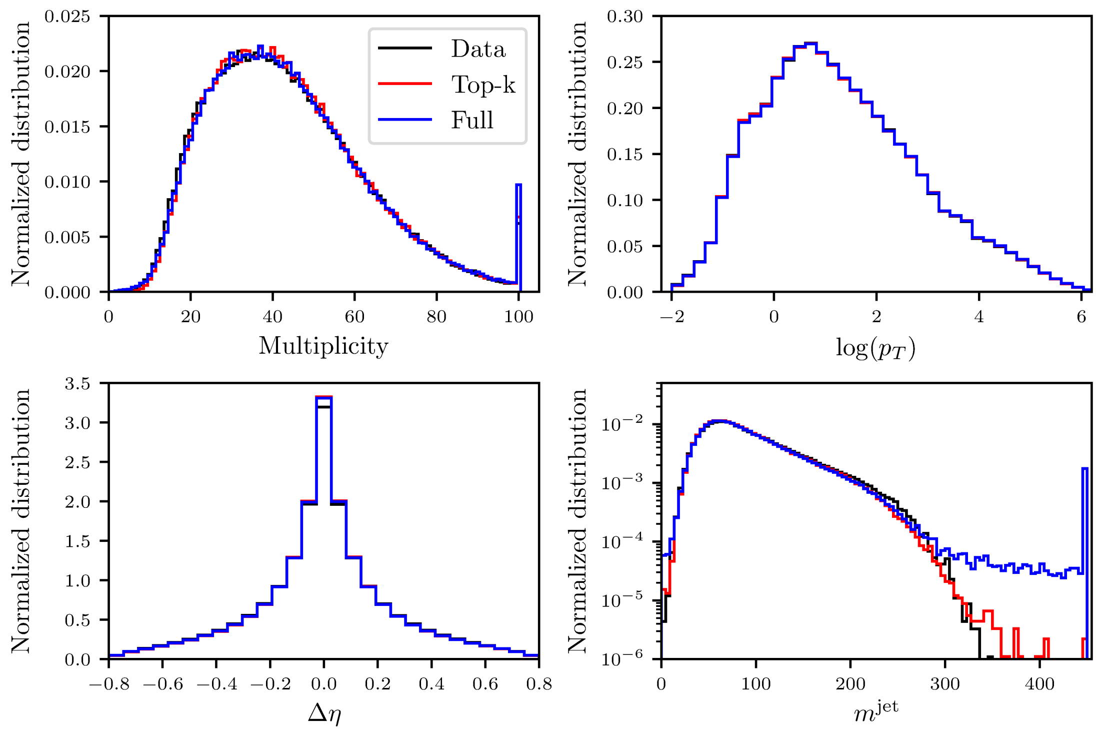
- density provides discrimination power
- room for improvement
- strong overfitting observed

(for diffusion based results see 2306.03933, Vinicius' talk today at 14.15h)

Quality assessment

- Transformer as generative model: draw jets from $p(\vec{x})$

(see also 2305.10475, Jonas' talk today at 14.30h)



- good agreement for 1D distributions
- multiplicity extrapolation works well (trained only with 50 consituents)
- top-k sampling to suppress low probability bins

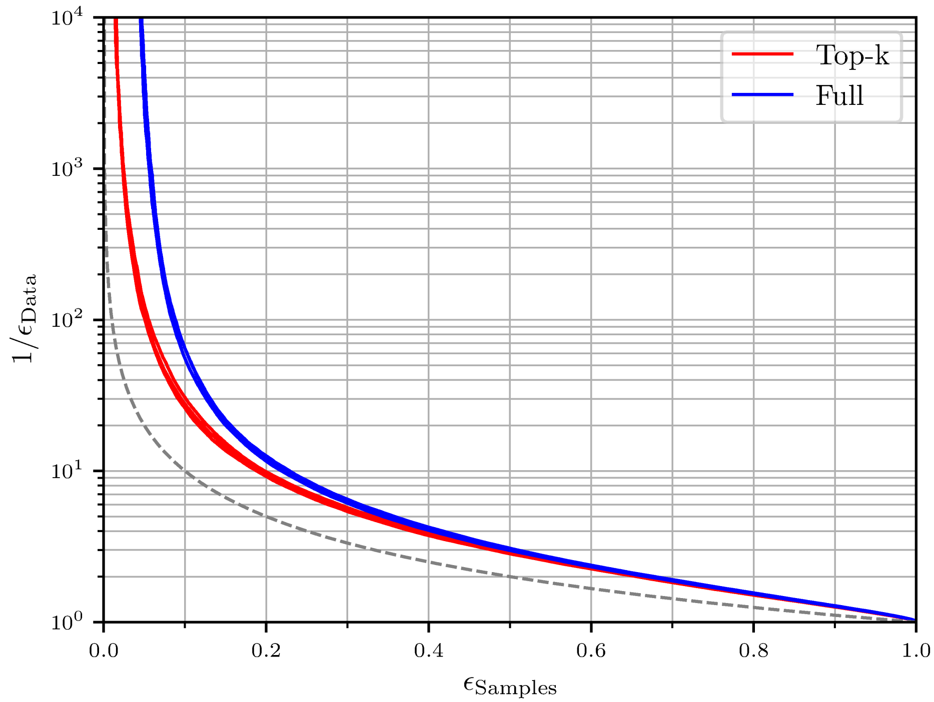
(sample from k=5000 particles with highest probability)

Quality assessment

- Transformer as generative model: draw jets from $p(\vec{x})$

  $\Rightarrow$ use classifier to discriminate samples from data

  (see also 2305.16774, Luigi's talk Thu 11.45h)



- poor classification $\Rightarrow$ good sampling $\Rightarrow$ good density estimate
- Top-k sampling $\Rightarrow$ fewer poor samples

Institute for Theoretical Particle Physics and Cosmology

Conclusion

Density estimation for low-level jet data

- following Natural Language Processing
- works for flexible number of constituents

Promising results

- the transformer can speak QCD with a slight accent
- use classifier to assess quality

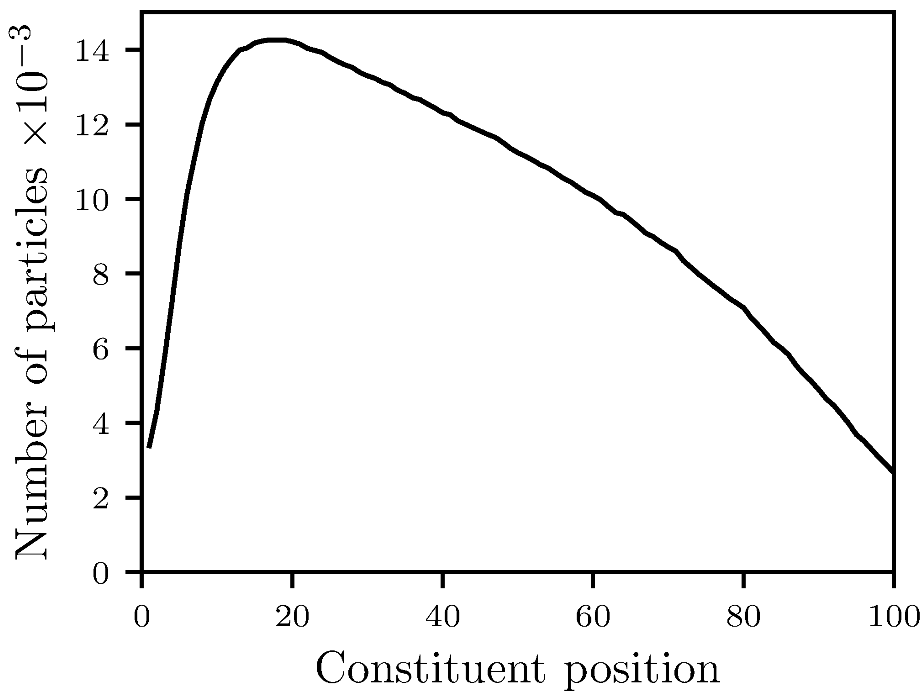
Outlook

- improvements on larger datasets expected
- use in the context of Anode and Cathode

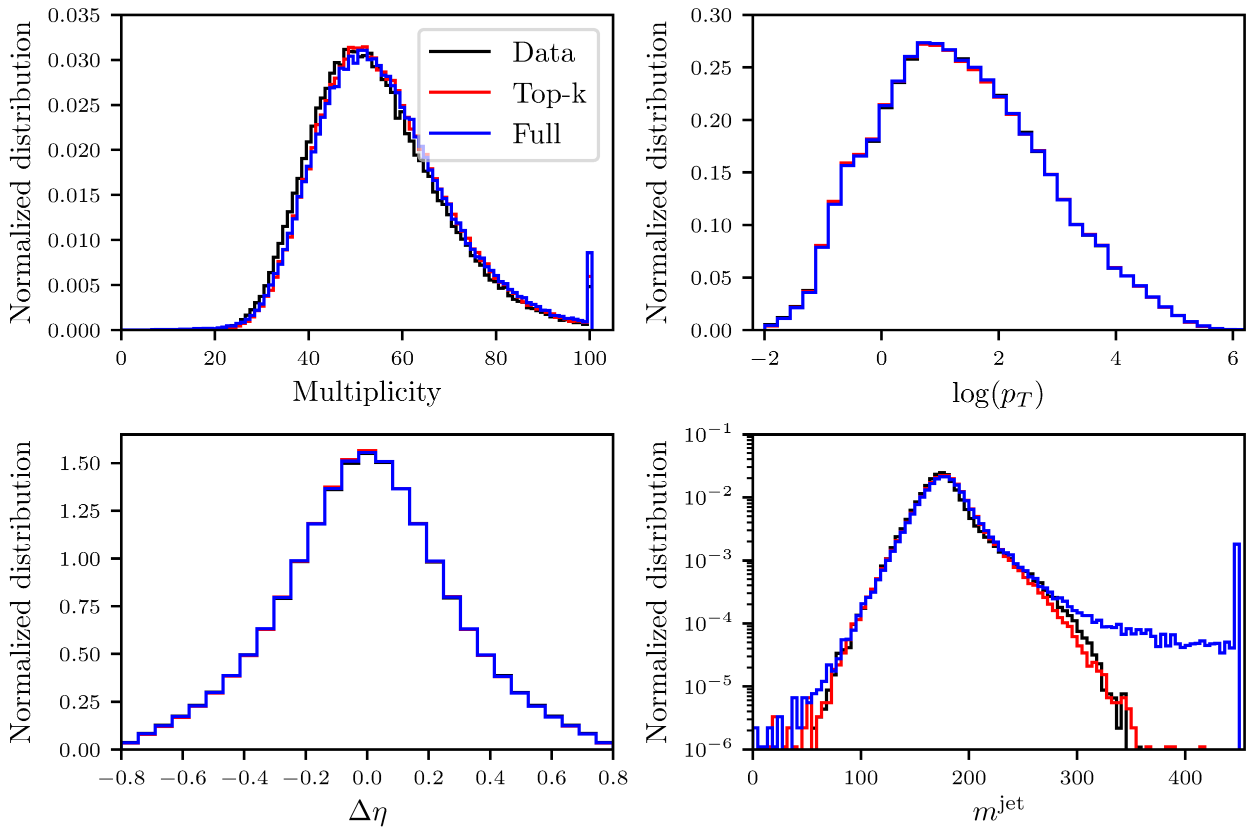
  (see also 2310.06897, talk by Cedric et al. Wed 17.15h)
- general pretraining (see also talk by Matthew)

Backup

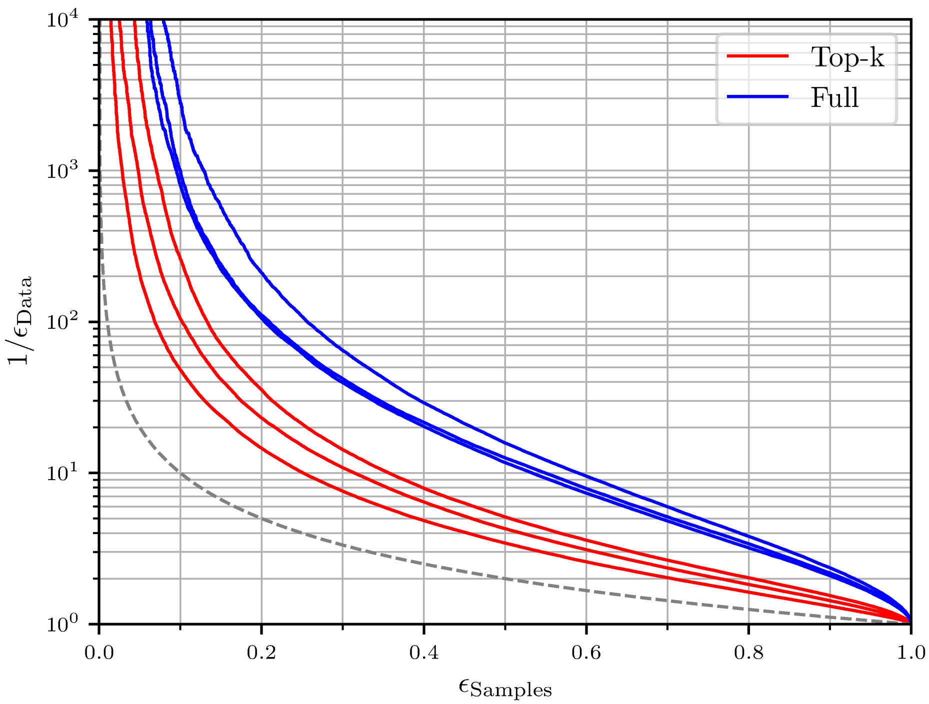
How many different particle types are in the data set?



Training on and sampling top jets: 1D distributions



Training on and sampling top jets: Classifier test



Institute for Theoretical Particle Physics and Cosmology

Sampling speed: