Synergizing Physics

Deep Learning Techniques for Time-of-Flight Reconstruction and Jet Tagging in High Energy Physics

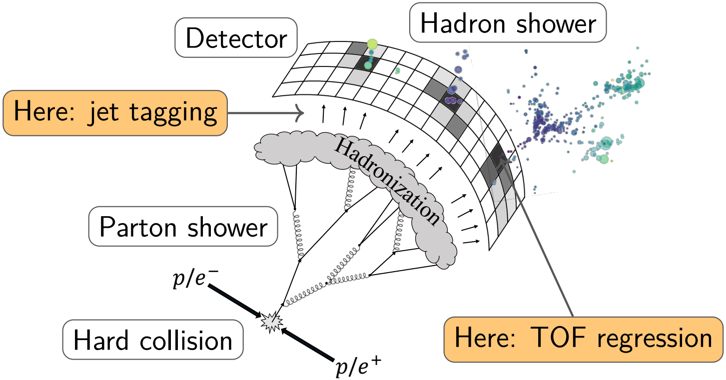
Konrad Helms

Georg-August-Universität Göttingen, Institut für Theoretische Physik

GEFÖRDERT VOM
Bundesministerium
für Bildung
und Forschung

Two Problems, One Solution

Two Problems



Jet Tagging

- classify sprays of particles, i.e. jets
- given a jet, infer whether it was quark or gluon initiated

Time-of-Flight (TOF) Regression

- given a shower in a calorimeter, infer TOF to the detector surface
- used for particle identification, example: ILD

One Solution

Q/G Jet Tagging

- binary **classification** task

Network Requirements

output: one number

Time-of-Flight Regression

- **regression** task

One Solution

Q/G Jet Tagging

- binary **classification** task
- data: jet, i.e. list of stable particles

Network Requirements

output: one number

input: tensor of variable length

Time-of-Flight Regression

- **regression** task
- data: shower in the calorimeter system, i.e. list of 5D hits

One Solution

Q/G Jet Tagging

- binary **classification** task
- data: jet, i.e. list of stable particles
- no natural ordering

Network Requirements

output: one number

input: tensor of variable length

permutation invariance

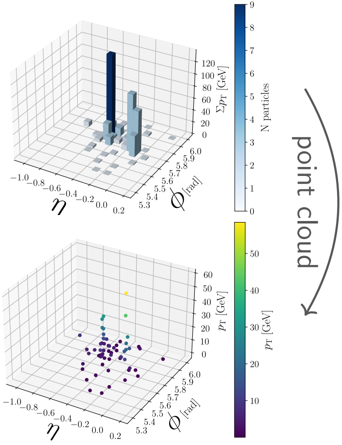
Time-of-Flight Regression

- **regression** task
- data: shower in the calorimeter system, i.e. list of 5D hits
- no ordering preserving shower evolution

$\Longrightarrow$ need permutation invariant network, relate points in a point cloud, predict one quantity

Q/G Jet Tagging

Jet Data



- in the past:

  classification based on hand crafted jet observables such as jet width, thrust, ...

  {jet} $\rightarrow$ **{list of observables}**

- late 2010s/early 2020s:

  point cloud-based approaches [arXiv:1902.08570, arXiv:2202.03772, ...], no hand crafted observables

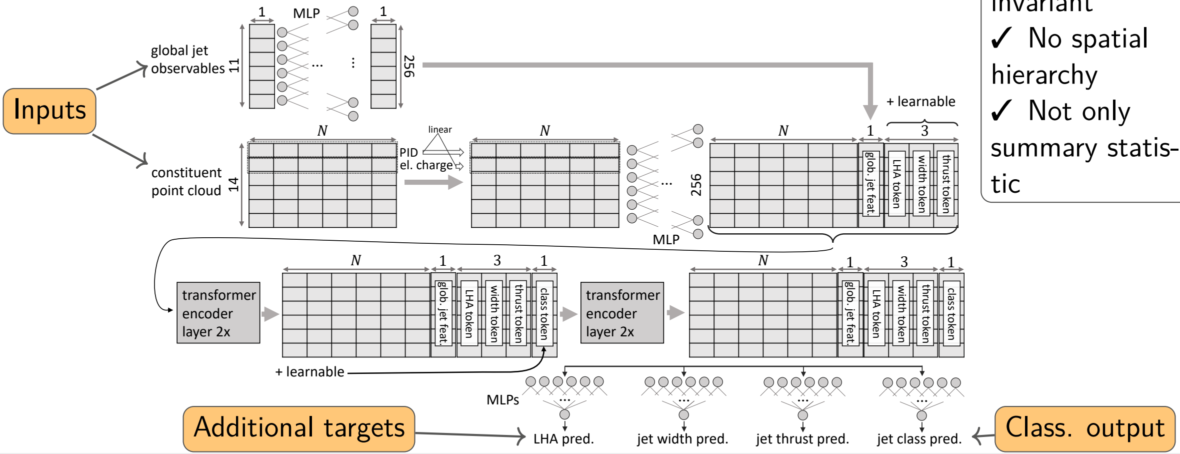
  {jet}

- why not combine them?

  {jet} $\rightarrow$ **{jet}** $\oplus$ **{list of observables}**

  point cloud $\oplus$ fixed size tensor

Jet Transformer

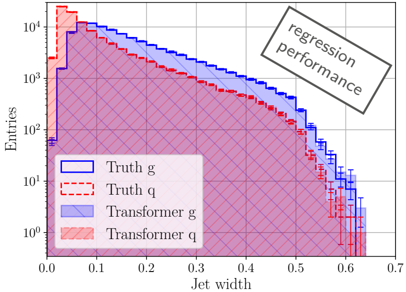


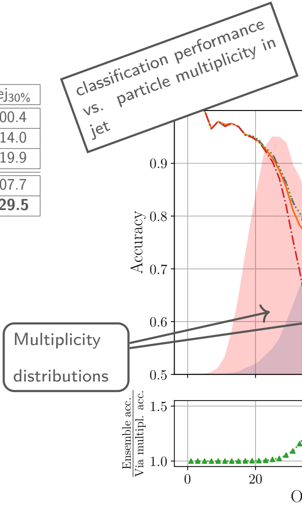
✓ Permutation invariant
✓ No spatial hierarchy
✓ Not only summary statistic

Jet Tagging Results

Classifier	Accuracy [%]	AUC	$\text{Rej}_{50\%}$	$\text{Rej}_{30\%}$
Benchmark CNN	82.8	0.902	35.1	100.4
Jet transformer	83.9	0.911	40.7	114.0
Ensemble (CNN & transformer)	84.1	0.913	41.8	119.9
$\text{OmniLearn}_{\text{full}}$ [arXiv:2404.16091]	84.4	0.916	43.7	107.7
$\text{ParT}_{\text{full}}$ [arXiv:2202.03772]	**84.9**	**0.920**	**47.9**	**129.5**

On LHC dataset:
Komiske, Metodiev, Thaler:
Pythia8 Quark and Gluon Jets for Energy Flow (2019)

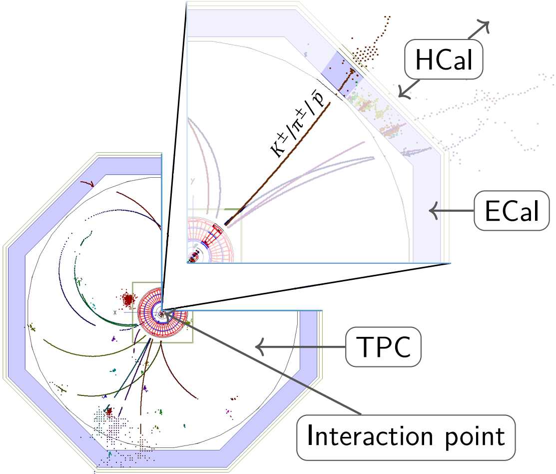




Time-of-Flight Regression

Hadronic Showers

- TOF estimation for hadron showers
  $\Longrightarrow$ particle identification
- for momenta $\lesssim \mathcal{O}(10\,\text{GeV})$:
  $\pi^\pm$ vs. $K^\pm$ vs. $\overset{(-)}{p}$
- particle identified by el. charge and
  $$m = \frac{p \cdot \text{TOF}}{\ell_\text{track}} \cdot \sqrt{1 - \left(\frac{\ell_\text{track}}{c \cdot \text{TOF}}\right)^2}$$
- for the TOF part of my talk:
  ILC @ 250 GeV & ILD

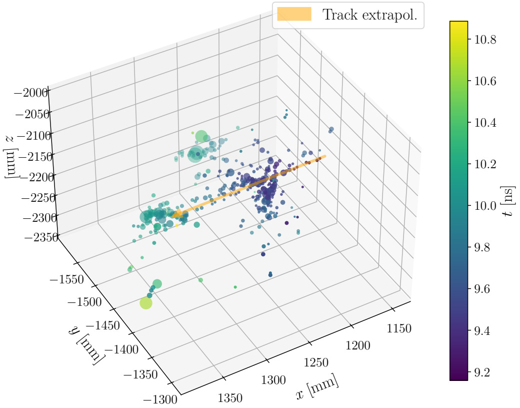


Calorimeter Data

Comparison to jet tagging:

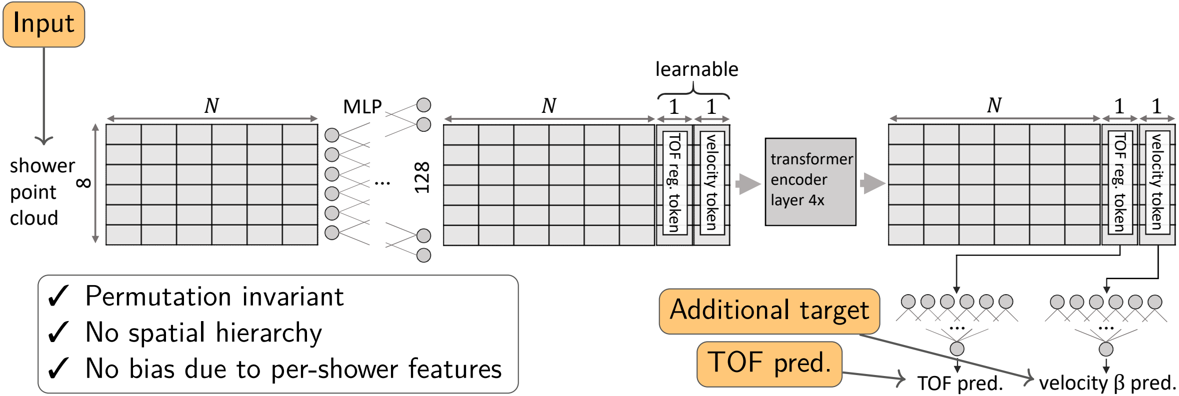
- $\{\textbf{jet}\} \oplus \{\textbf{list of observables}\}$
- $\{\textbf{shower}\} \oplus \{\ell_{\textbf{track}},\ p,\ p_{\textbf{T}},\ \ldots\}$
- network does (vague) PID internally
- predicts according to what it 'thinks' is the correct TOF
- solution:
- remove information that enables internal PID
- $\{\textbf{shower}\} \oplus \{\cancel{\ell_{\textbf{track}},\ p,\ p_{\textbf{T}},\ \ldots}\}$

see my ML4Jets talk '23



GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737

Time-of-Flight Transformer

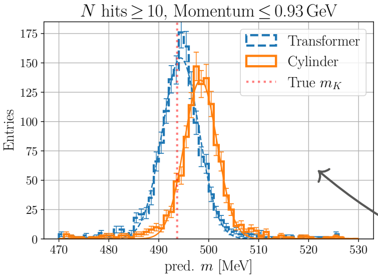


- ✓ Permutation invariant
- ✓ No spatial hierarchy
- ✓ No bias due to per-shower features

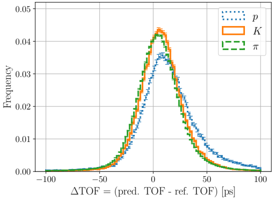
Time-of-Flight Regression Results

- comparison to current ILD benchmark: 'Cylinder' algorithm
- Transformer:

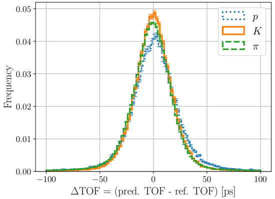
  narrower ΔTOF, less biased



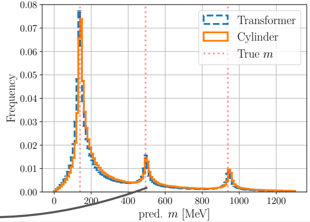
PRELIMINARY



Benchmark



TOF Transformer



focus on $K^{\pm}$ mass range

Summary

- two unrelated problems – one problem solving mechanism
- important: same data structures
  - unordered point clouds
  - variable sized point clouds

Q/G Jet Tagging

- ensemble (CNN & Transformer) classification performance comparable to OmniLearn
- classification performance gain (w.r.t. class. based on multiplicity) in range 20-60, i.e. $\sim$ 80% of the data

Time-of-Flight Regression

- TOF Transformer outperforms current ILD benchmark
  - ΔTOF distribution less biased
  - ΔTOF distribution narrower
- comparable TOF regression performance for all particle species
- prospect of resolving the $K^{\pm}$ double mass peaks

Thank you for your attention!

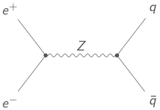
Georg-August-Universität Göttingen
Institut für Theoretische Physik
E-Mail: konrad.helms@theorie.physik.uni-goettingen.de

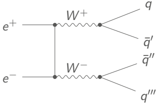
Backup

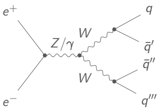
Hadron Shower Dataset

- training/validation/testing = 1.5/0.05/1 million showers
- ILC @ $\sqrt{s} = 250\,\text{GeV}$, ILD v02
- Whizard 2.8.5 generated processes:

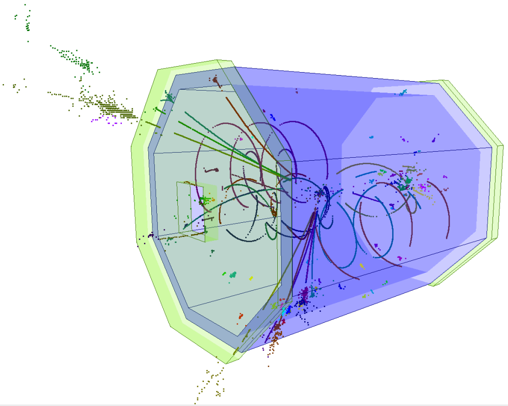
$$e^+e^- \to Z \to q\bar{q},$$
$$e^+e^- \to W^+W^- \to 4q$$







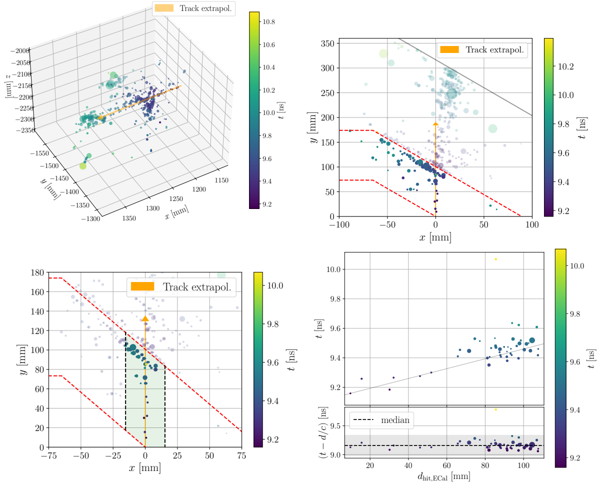
- hadronization done using Pythia 6.4
- passage of particles through matter: Geant4 10.04



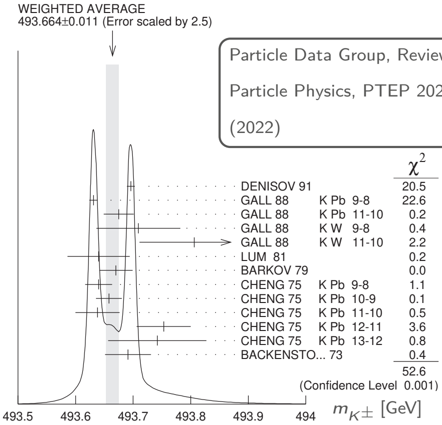
Benchmark Algorithm

- developed by Bohdan Dudar (DESY, University of Hamburg)
- consider only first 10 ECal layers
- focus on $\pi^{\pm}$ vs. $K^{\pm}$ vs. $\overset{(-)}{p}$
- cylindrical and temporal hit selection
- cyliner radius and temporal cut optimized on training data

$$\mathrm{TOF} = \frac{1}{N}\sum_{i=1}^{N}\left(t_i - \frac{d_{\text{hit } i,\text{ ECal}}}{c}\right)$$



Kaon Mass Distribution

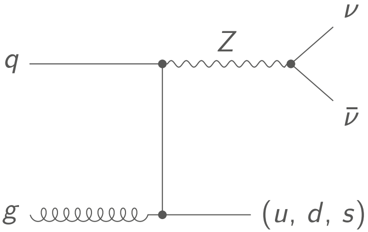


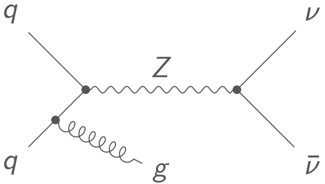
$m_{K^\pm}$ [GeV]

Particle Data Group, Review of Particle Physics, PTEP 2022, 083C01 (2022)

Q/G Dataset #1

- 2 million q/g initiated jets
- generated with Pythia 8.226
- $Z(\to \nu\bar{\nu}) + (u, d, s)$, $Z(\to \nu\bar{\nu}) + g$
- $\sqrt{s} = 14$ TeV
- clustering with anti-k_t algorithm, $R = 0.4$ in FastJet 3.3.0
- cuts: $p_{\mathrm{T,\,jet}} \in [500, 550]$ GeV, $|\eta_{\mathrm{jet}}| < 1.7$





Dataset: Komiske, Metodiev, Thaler, (2019), Pythia8 Quark and Gluon Jets for Energy Flow, available at Zenodo

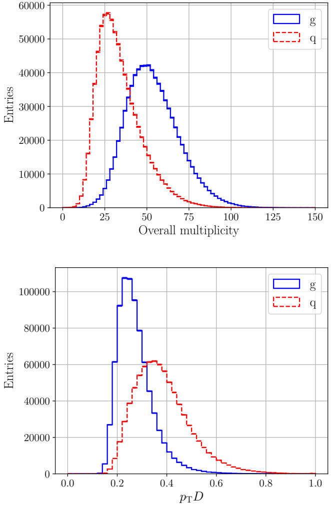
Q/G Dataset #2

- training/validation/testing = 1.6/0.2/0.2 million jets
- used **full** PID information
- jet point-cloud:

  $J_k = \{\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_{N_k}\}$ with

  $\vec{c}_i = (p_\mathrm{T}, \eta, \phi, \mathrm{PID})_i$
- particle multiplicity folded into many jet observables, for example

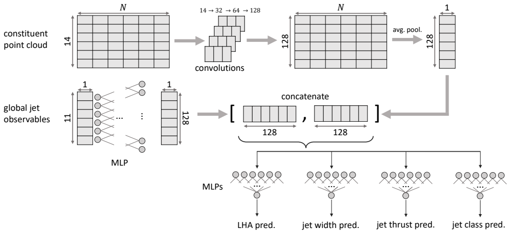
$$p_\mathrm{T} D = \frac{\sqrt{\sum_{i \in \mathrm{jet}} p_{\mathrm{T},i}^2}}{\sum_{i \in \mathrm{jet}} p_{\mathrm{T},i}}$$



Q/G Benchmark CNN & Ensemble

- CNN serves as a benchmark
- same input as Jet Transformer
- need to order the jet point cloud

  $\Longrightarrow$ use (approximate) roatation invariance around jet axis
- locality bias
- spatial hierarchy
- ensemble of Jet Transformer and benchmark CNN



- ensemble weights optimized on validation dataset
- $w_{\text{best}} = 0.69$
- $w \cdot (\text{Transformer class pred.}) + (1 - w) \cdot (\text{CNN class pred.})$

Classification Metrics

- $\mathrm{Rej}_{X\%} = \frac{1}{\mathrm{FPR}}$ at $\mathrm{TPR} = X\%$

Classifier	Acc. [%]	TPR [%]	FPR [%]	TNR [%]	FNR [%]
Based on multiplicity	76.61	72.74	19.50	80.50	27.26
Transformer	83.91	**81.47**	13.64	86.36	**18.53**
Benchmark CNN	82.75	80.35	14.83	85.17	19.65
Ensemble	**84.11**	81.37	**13.14**	**86.86**	18.63

- positive class = quarks
- negative class = gluons