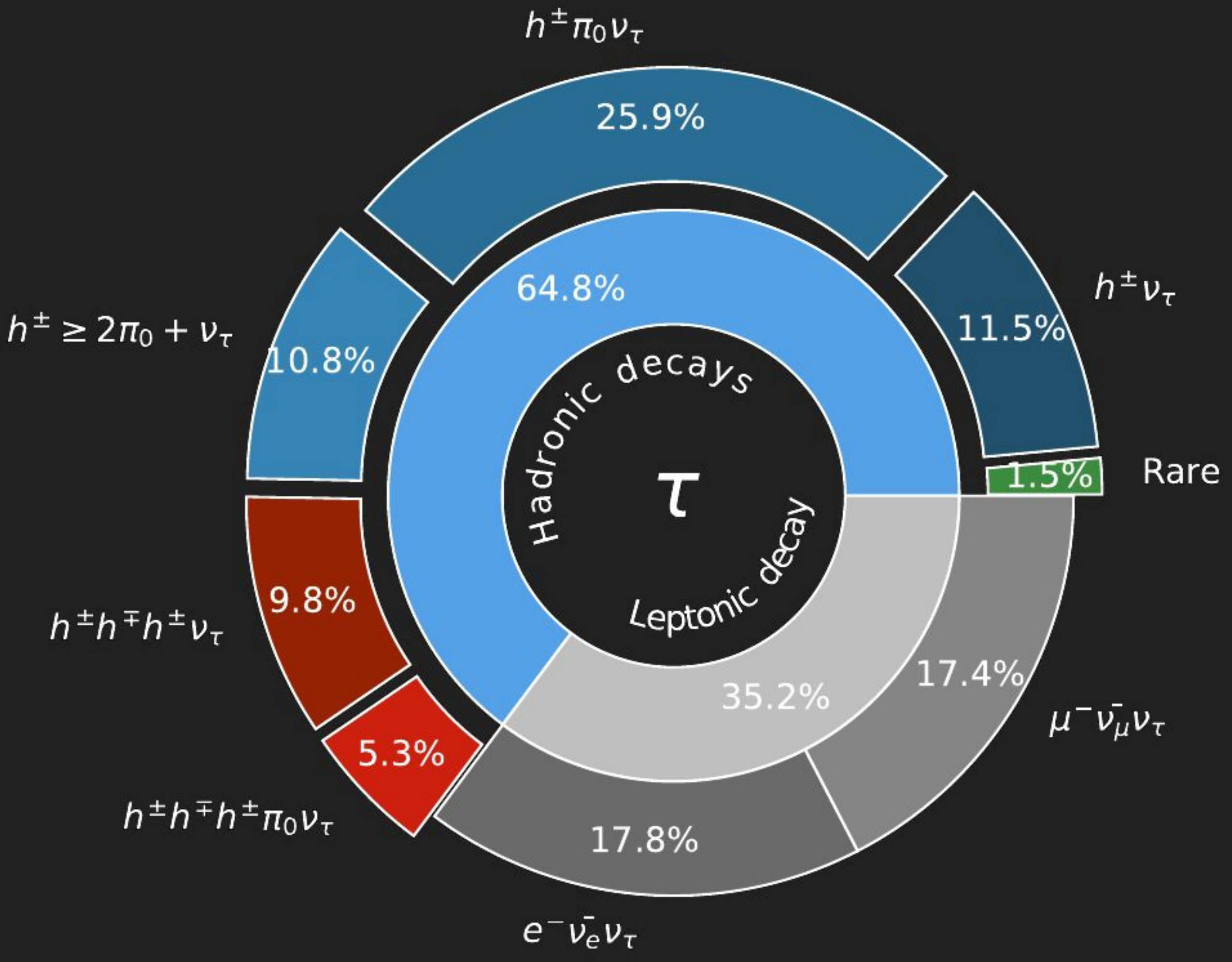
Improving Hadronically Decaying Tau Lepton Identification and Reconstruction with Unified End-to-End Machine Learning Methods

CERN Baltic Conference

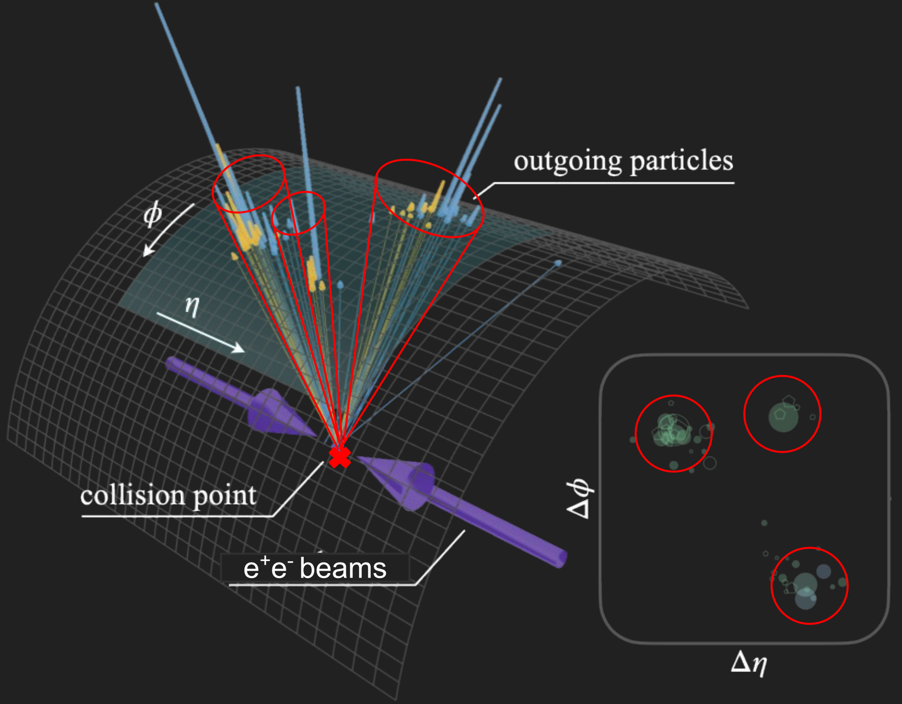
Laurits Tani



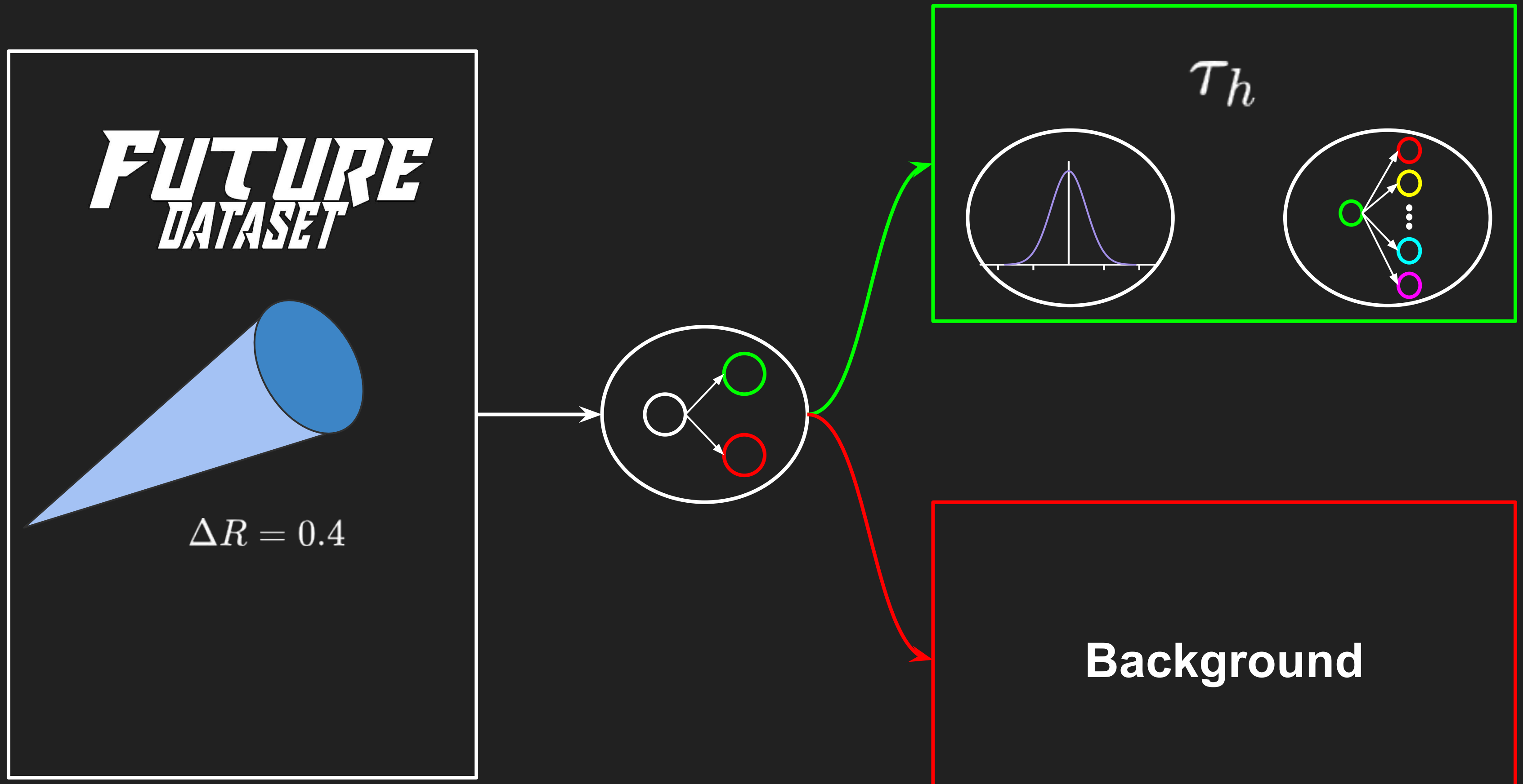
$h^{\pm}\pi_0\nu_\tau$
25.9%
$h^{\pm}\nu_\tau$
11.5%
Rare
1.5%
$h^{\pm} \geq 2\pi_0 + \nu_\tau$
10.8%
64.8%
Hadronic decays
τ
Leptonic decay
35.2%
$h^{\pm}h^{\mp}h^{\pm}\nu_\tau$
9.8%
$h^{\pm}h^{\mp}h^{\pm}\pi_0\nu_\tau$
5.3%
17.8%
$e^-\bar{\nu_e}\nu_\tau$
17.4%
$\mu^-\bar{\nu_\mu}\nu_\tau$

FUTURE DATASET V.2

- dataset with full detector simulation for the development and training of τ_h reconstruction algorithms



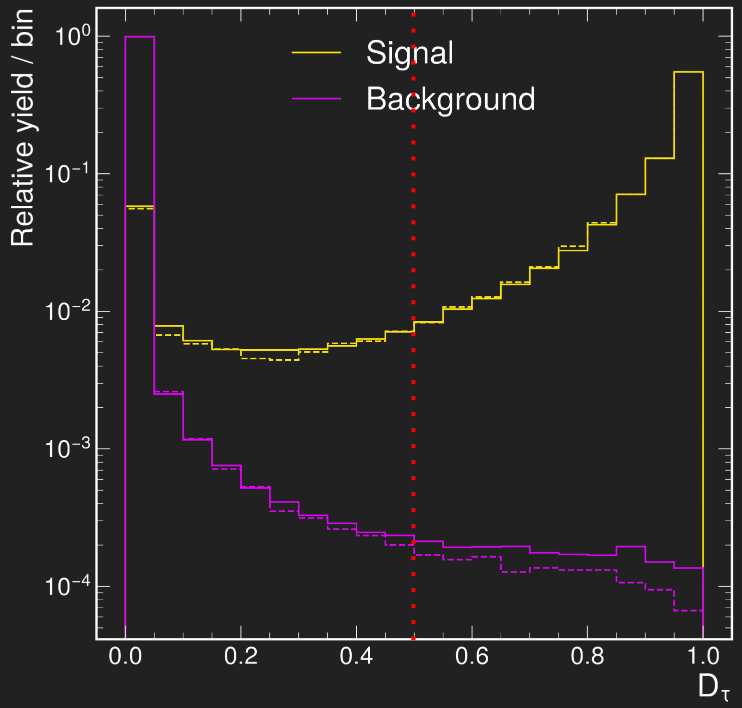
Collision event → Jet clustering → ML task

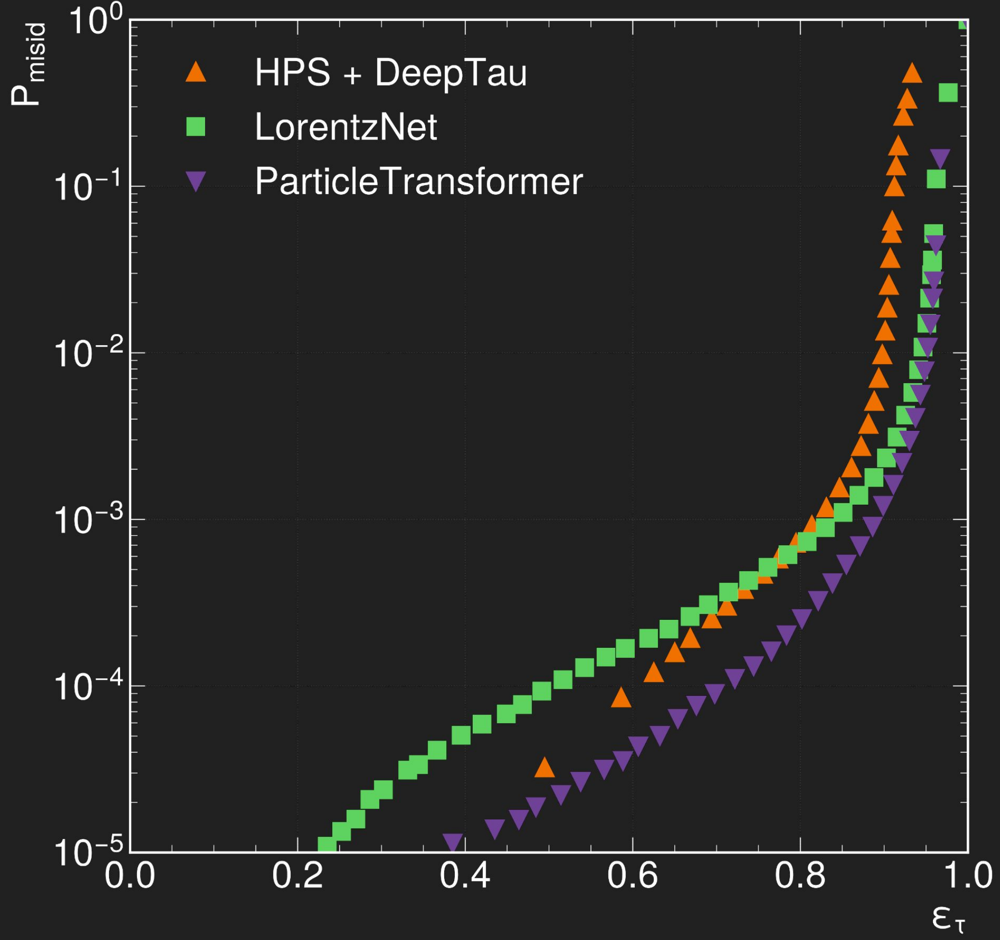


FUTURE
DATASET
$\Delta R = 0.4$
τ_h
Background

τ-tagging [binary classification]

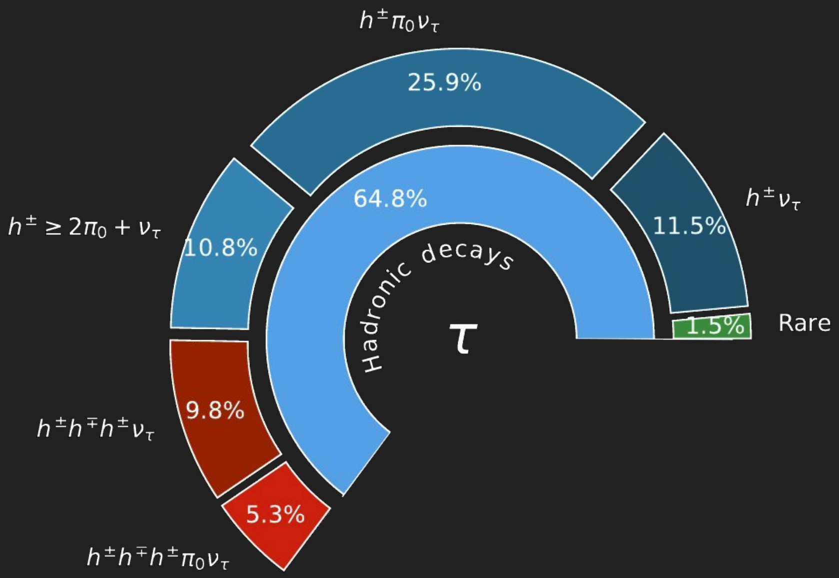
- Signal: τ-jets from $Z \to \tau\tau$
- Background: quark/gluon-jets from $Z \to qq$
- Binary classification

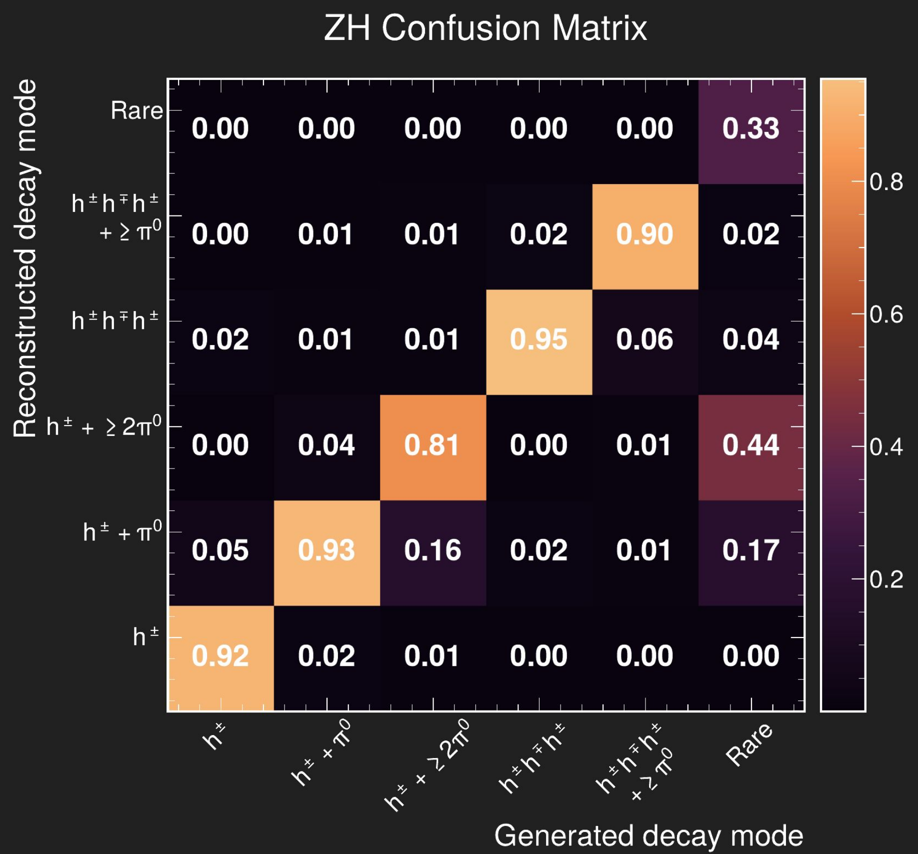


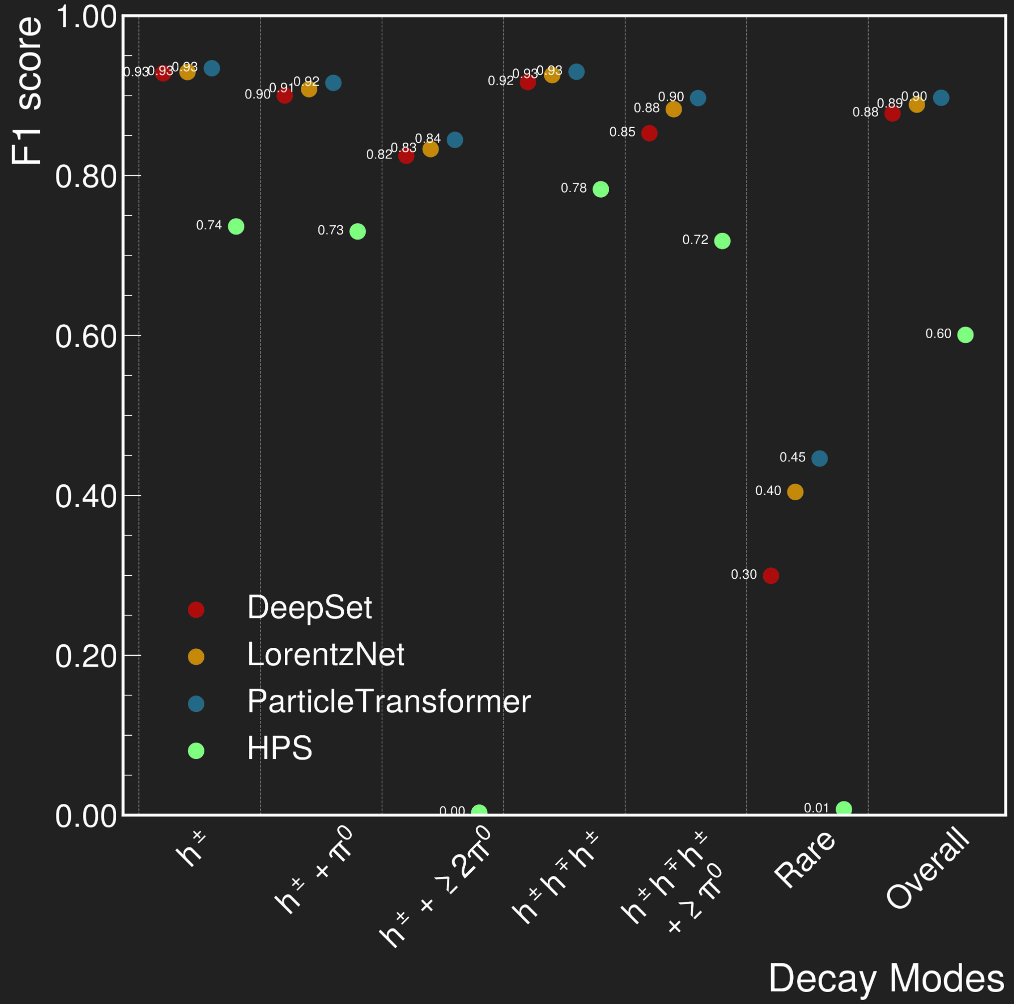


HPS + DeepTau
LorentzNet
ParticleTransformer
P_{misid}
ε_τ
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
0.0
0.2
0.4
0.6
0.8
1.0

Decay mode reconstruction [multiclass classification]





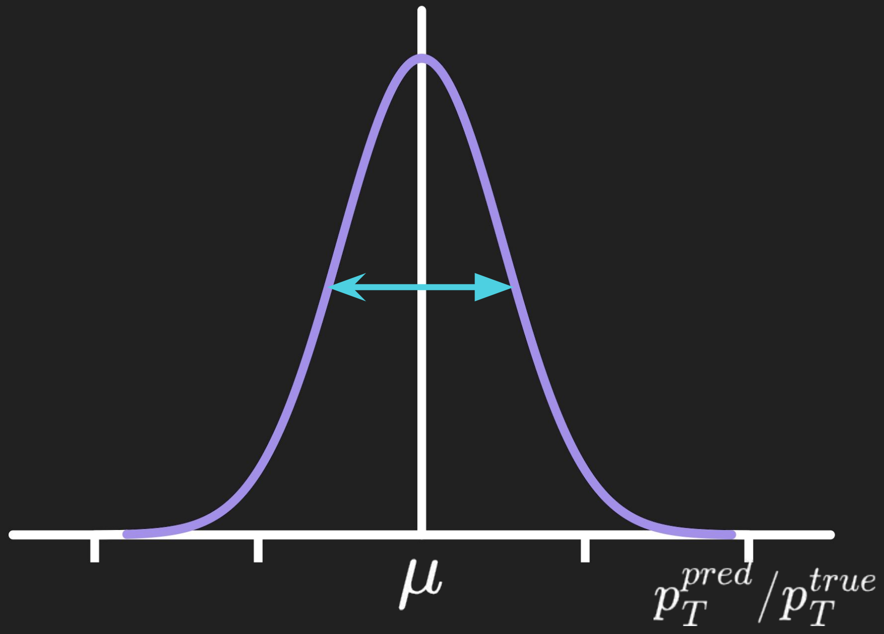


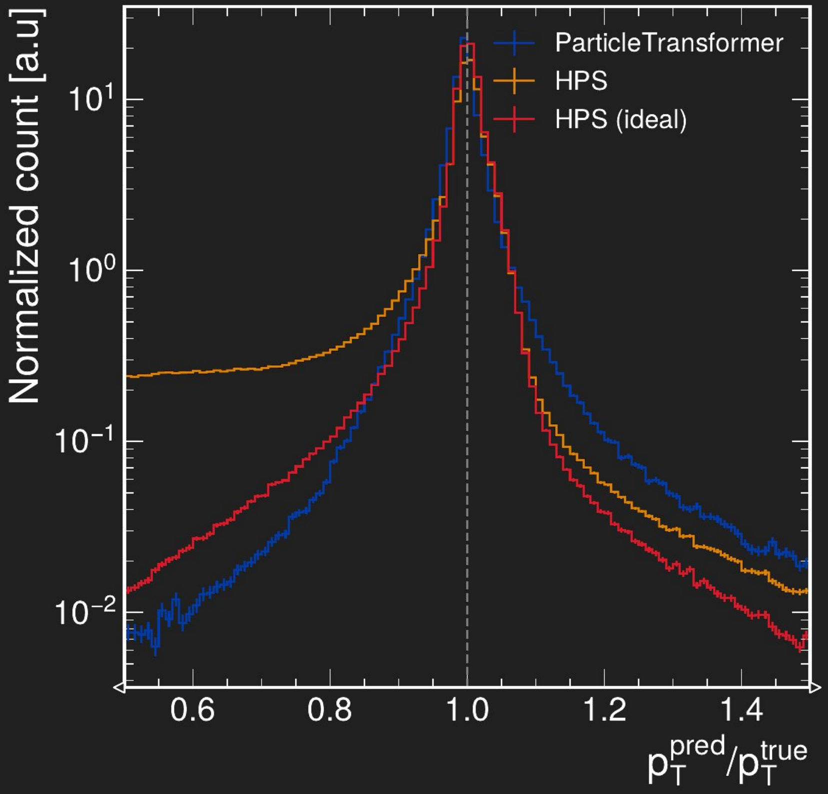
F1 score
1.00
0.80
0.60
0.40
0.20
0.00
0.93
0.93
0.93
0.90
0.91
0.92
0.82
0.83
0.84
0.92
0.93
0.93
0.85
0.88
0.90
0.88
0.89
0.90
0.74
0.73
0.78
0.72
0.60
0.45
0.40
0.30
0.00
0.01
DeepSet
LorentzNet
ParticleTransformer
HPS
$h^{\pm}$
$h^{\pm} + \pi^{0}$
$h^{\pm} + \geq 2\pi^{0}$
$h^{\pm}h^{\mp}h^{\pm}$
$h^{\pm}h^{\mp}h^{\pm} + \geq \pi^{0}$
Rare
Overall
Decay Modes

Kinematic reconstruction [regression]

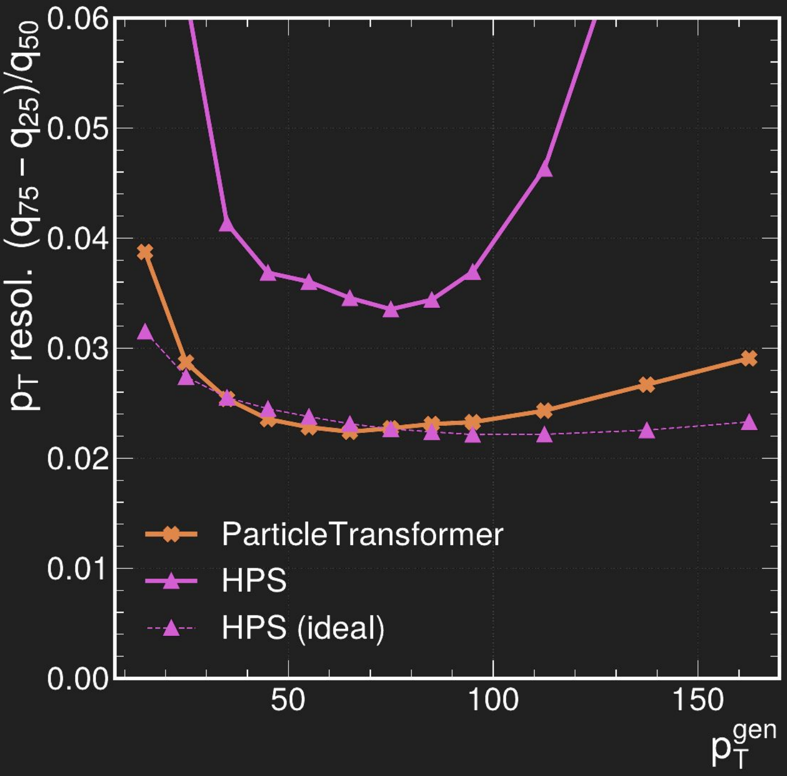
- Predict tau p_T
- Measures of interest:
  - scale/response: μ
  - resolution: IQR

$$IQR = \frac{q_{75} - q_{25}}{q_{50}}$$

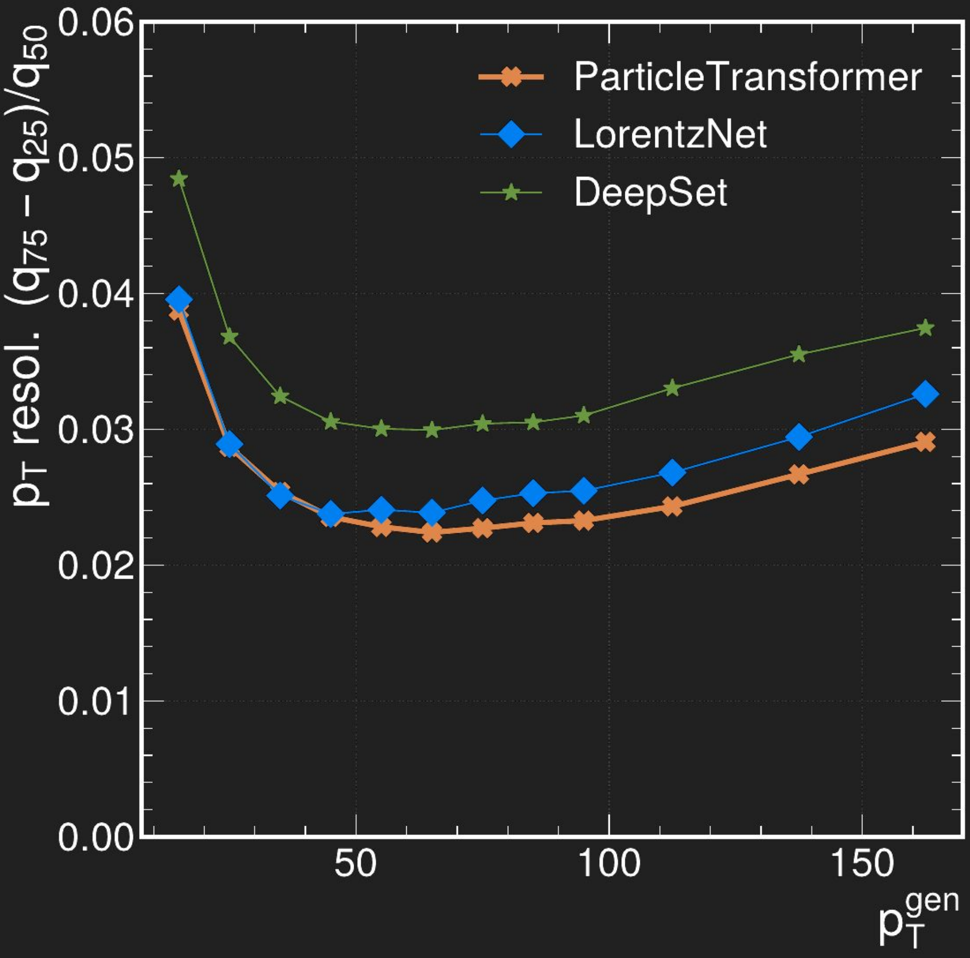




ParticleTransformer
HPS
HPS (ideal)
Normalized count [a.u]
10^1
10^0
10^{-1}
10^{-2}
0.6
0.8
1.0
1.2
1.4
p_T^{pred}/p_T^{true}



p_T resol. $(q_{75} - q_{25})/q_{50}$
0.06
0.05
0.04
0.03
0.02
0.01
0.00
50
100
150
p_T^{gen}
ParticleTransformer
HPS
HPS (ideal)


ParticleTransformer
LorentzNet
DeepSet
p_T resol. $(q_{75} - q_{25})/q_{50}$
0.06
0.05
0.04
0.03
0.02
0.01
0.00
50
100
150
p_T^{gen}

Conclusion

- Created a new full-sim dataset for studying τ_{h} reconstruction at future colliders
- ParticleTransformer outperforms current state-of-the-art methods in all scenarios
- Can be used for future detector development studies

References

- T. Lange et al. “Tau lepton identification and reconstruction: A new frontier for jet-tagging ML algorithms”. In: Comput. Phys. Commun. 298 (2024), p. 109095. doi: 10.1016/j.cpc.2024.109095. arXiv: 2307.07747 [hep-ex]

- L. Tani et al. “A unified machine learning approach for reconstructing hadronically decaying tau leptons”. In: Comput. Phys. Commun. 307 (2025), p. 109399. doi: 10.1016/j.cpc.2024.109399. arXiv: 2407.06788 [hep-ex]

- L. Tani et al., FuTure - dataset for studies, development, and training of algorithms for reconstructing and identifying hadronically decaying tau leptons. [Online]. Available: https://doi.org/10.5281/zenodo.12664634

Fin

Backup

FUTURE DATASET V.2

$Z \to \tau\tau$	$ZH, H \to \tau\tau$	$Z \to qq$
4.35M jets 0.9 GB	5.34M jets 1 GB	31.83M jets 7 GB

- Full detector simulation: Geant4
- CLIC like detector (CLICdet) setup
- Event reconstruction: Marlin reconstruction
- ParticleFlow candidates: PandoraPF
- Jet clustering: generalized k_t algorithm for ee $(p = -1; \Delta R = 0.4)$
- No $\gamma\gamma \to \texttt{hadrons}$ overlay

Cuts for efficiency calculation

$$\mathcal{P} = \frac{p_{\mathrm{T}}^{\mathrm{rec}} > 20\ \mathrm{GeV}\ \&\ 10 < \theta_{\mathrm{rec}} < 170^{\circ}\ \&\ \mathcal{D}_{\tau} > \mathcal{T}}{p_{\mathrm{T}}^{\text{gen-X}} > 20\ \mathrm{GeV}\ \&\ 10 < \theta_{\text{gen-X}} < 170^{\circ}}$$

$$gen\text{–}X = \begin{cases} gen\text{–}tau, & \text{if } signal \\ gen\text{–}jet, & \text{otherwise} \end{cases}$$

$$\mathcal{P} = \begin{cases} \varepsilon_{\tau}, & \text{if } signal \\ P_{misid}, & \text{otherwise} \end{cases}$$

F1 score definition

$$F_1 = \frac{2 \times \texttt{precision} \times \texttt{recall}}{\texttt{precision} + \texttt{recall}}$$

$$\texttt{precision} = \frac{tp}{tp + fp}$$

$$\texttt{recall} = \frac{tp}{tp + fn}$$