UParT: A unified approach for jet-based object identification in CMS in Run 3

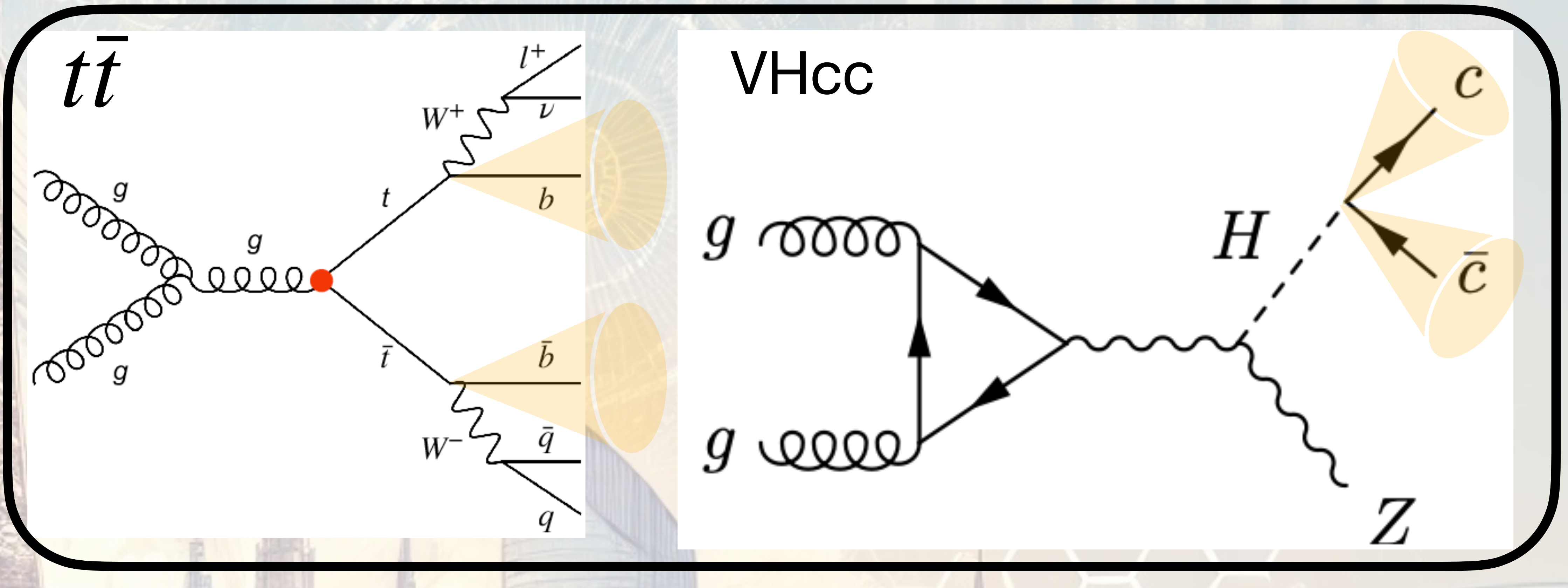
Uttiya Sarkar
on behalf of the CMS collaboration

Physics Institute III A
RWTH AACHEN UNIVERSITY

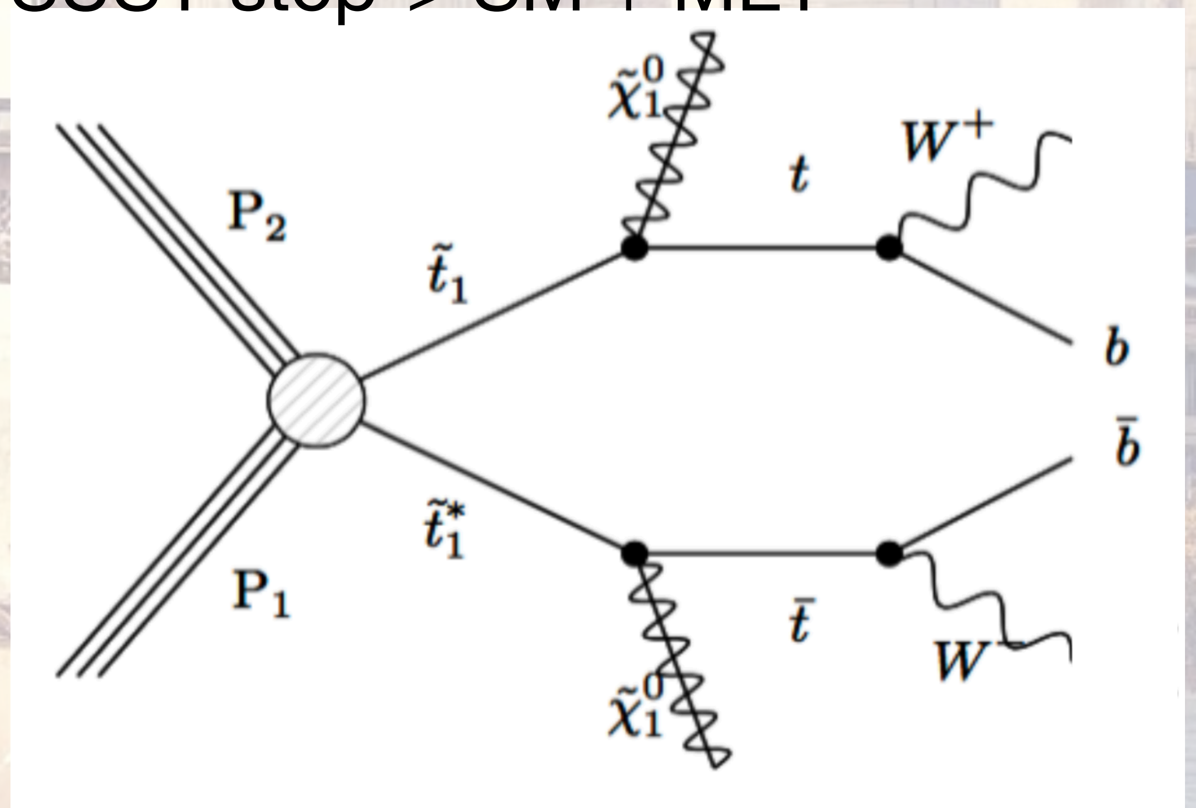
Physics Motivation

- Jet flavor identification is very crucial for standard model measurements and searches e.g.

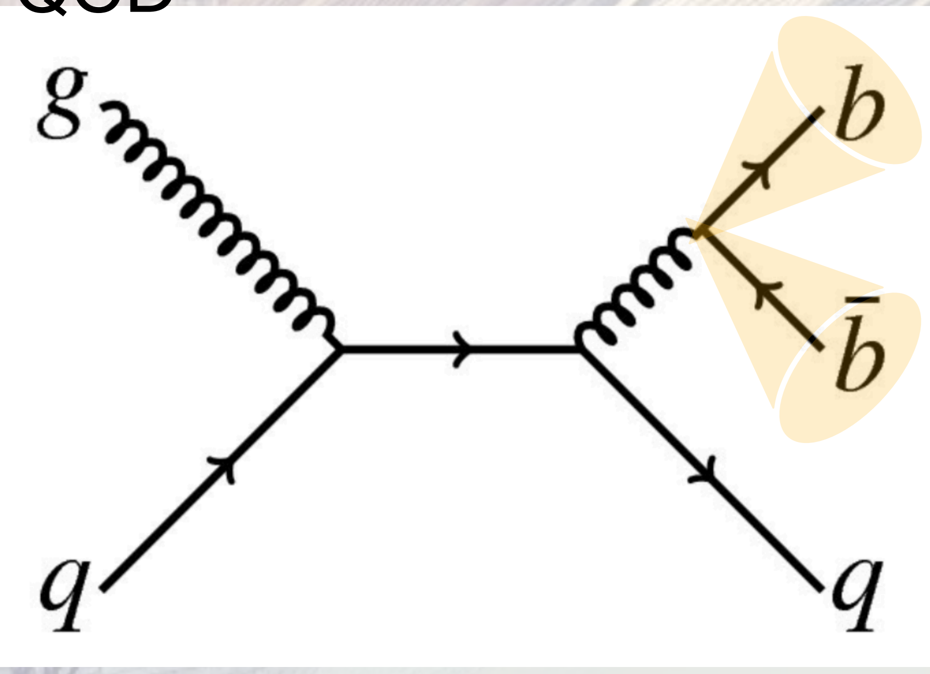
- Higgs sector: BR($H \to b\bar{b}$) ~ 60%
- Sensitivity for $H \to c\bar{c}$
- Top quark decay: BR(t → bW) ~ 100%
- New particles decaying to t, H, b or c quarks
- Understanding parton distributions in QCD



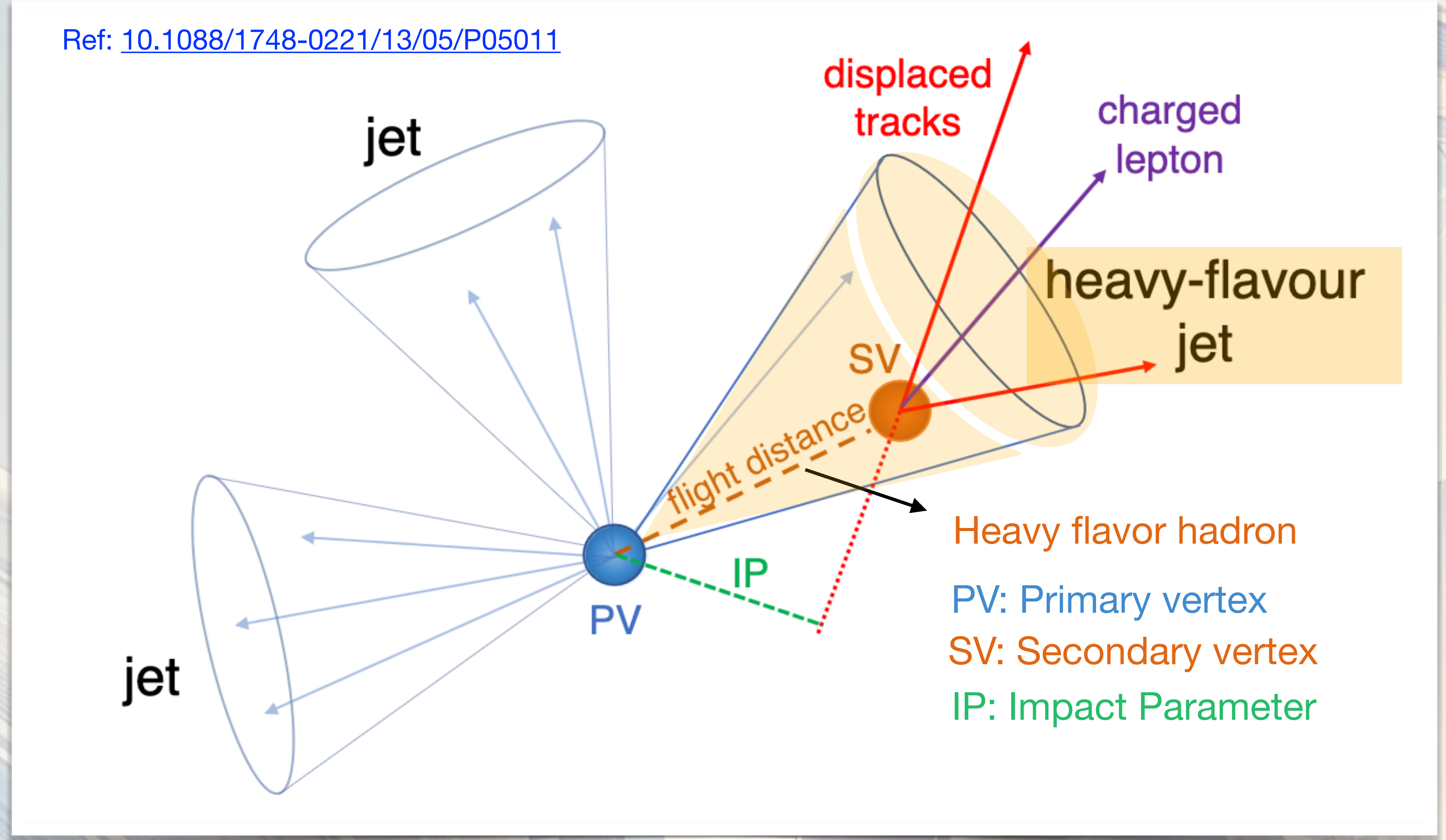
SUSY stop-> SM + MET



QCD



Heavy flavor jet tagging



- ***Heavy flavor jets:*** Jets originating from b (***b jets***) or c (***c jets***) quarks arising from the process of hadronization
- CMS tracker can detect unique features of heavy quarks
  - track impact parameter
  - reconstructed secondary vertex (flight distance, mass, energy fraction, multiplicity)
  - soft lepton

- **Discriminators:**

b jet

c jet

udg jet

$$\mathrm{BvsX} = \frac{\mathrm{P(B)}}{\mathrm{P(B) + P(X)}}$$

$P(B), P(x) =$ output nodes in a classifier

Historical evolution of particle taggers in CMS

Run 1 | LS 1 | Run 2 | LS 2 | Run 3

7 TeV | 8 TeV | 13 TeV | 13.6 TeV

2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024

combined secondary vertex (CSV)
Likelihood ratio

combined secondary vertex v2 (CSVv2)
MLP

DeepCSV
DNN

DeepJet
CNN + RNN

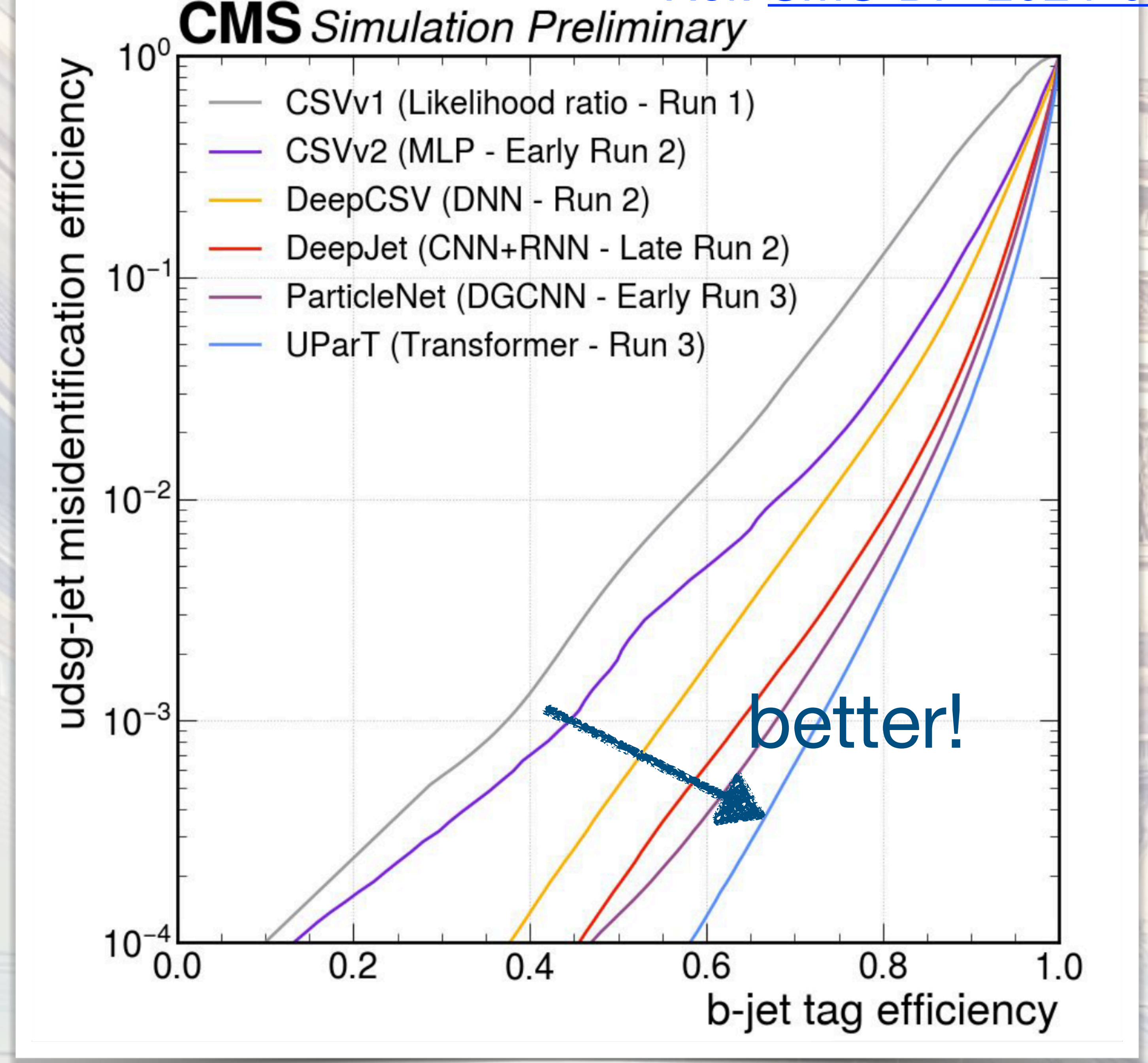
ParticleNet
DGCNN - Treat a jet as particle clouds - unordered set of its constituent particle

RobustParT
Multi Head Attention

UParT

Distortion of input features to enhance the robustness

Ref: CMS-DP-2024-066



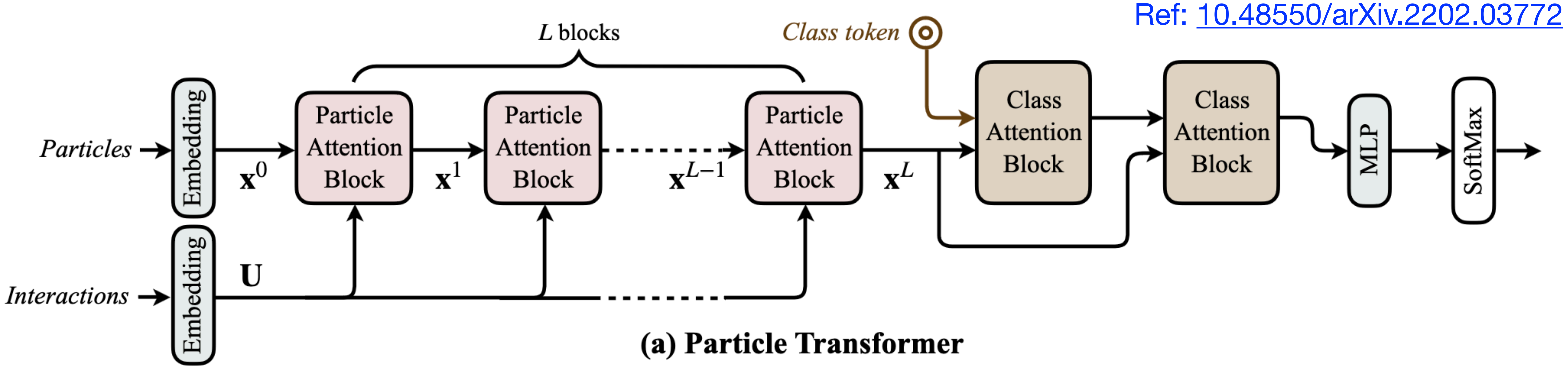
Significant improvement in performance over the last decade!

Unified Particle Transformer (UParT): focus of this talk

UParT Foundations: ParticleTransformer

$$P - MHA(K, Q, V) = \mathrm{Softmax}(QK^T/\sqrt{d_k} + U)V$$

Ref: 10.48550/arXiv.2202.03772



(a) Particle Transformer

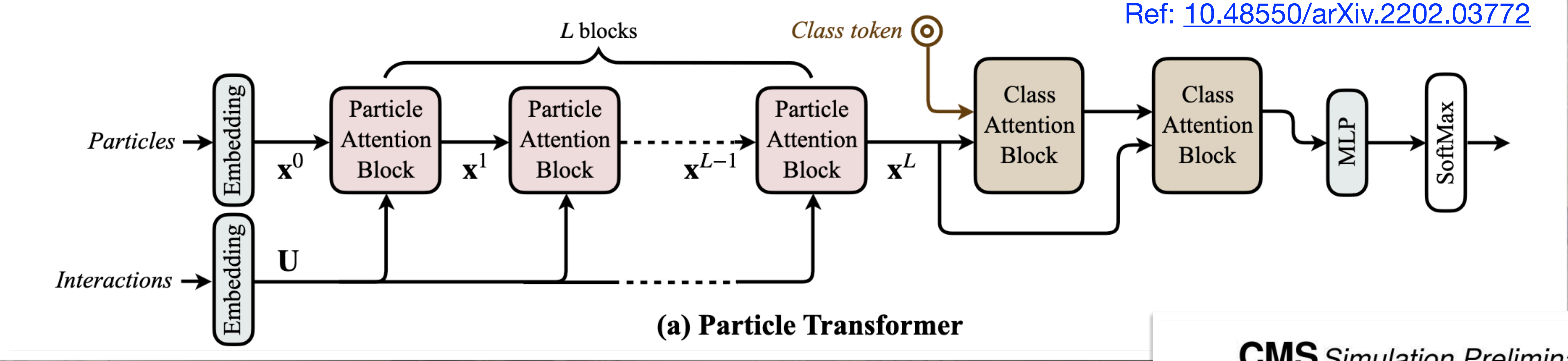
- Based on the "Attention" model designed for particles
- Input embedding:
  - Not only inject single particle information, but also include pair-wise features (interactions)
  - Multi-Head Attention (MHA)

- **Input features:**
  - **Particle features:**
    - **Kinematics**: 4-vector (E, px, py , pz)
    - **Particle identification:** charge, type (hadron{charged,neutral}, lepton {e,mu}, photon)
    - **Trajectory displacement:** longitudinal and transverse IP
  - **Interaction features:**
    - $\Delta = \sqrt{(y_a - y_b)^2 + (\phi_a - \phi_b)^2}$
    - $k_T = min(p_{T,a}, p_{T,b})\Delta$
    - $z = min(p_{T,a}, p_{T,b})/(p_{T,a} + p_{T,b})$
    - $m^2 = (E_a + E_b)^2 - ||\vec{p_a} + \vec{p_b}||^2$

UParT

$$P - MHA(K, Q, V) = \text{Softmax}(QK^T/\sqrt{d_k} + U)V$$

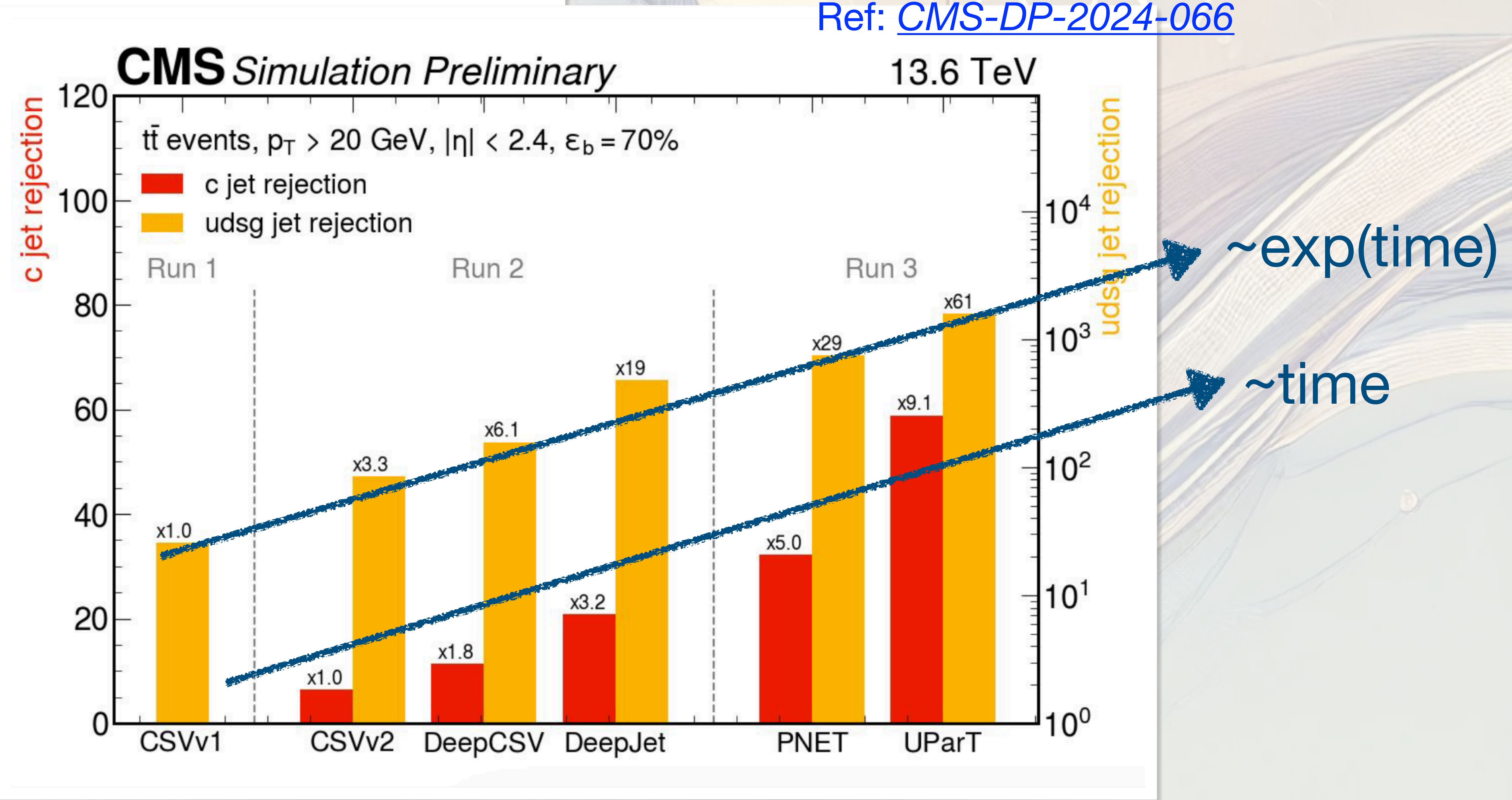
Ref: 10.48550/arXiv.2202.03772



(a) Particle Transformer

- Dataset size: 30M training / 5M validation
- 30 epochs with batch_size of 2048
- Complexity ~ 240M floating point operations in the forward path
- 6 Particle attention blocks + 2 class attention blocks

Ref: CMS-DP-2024-066



Significant improvement as compared to PNet!

UParT Modifications: I

- Extended class: extending from b and c jet identification to s and hadronic tau (one per final state) identification
- Extended regression: simultaneous flavor aware jet energy regression and resolution

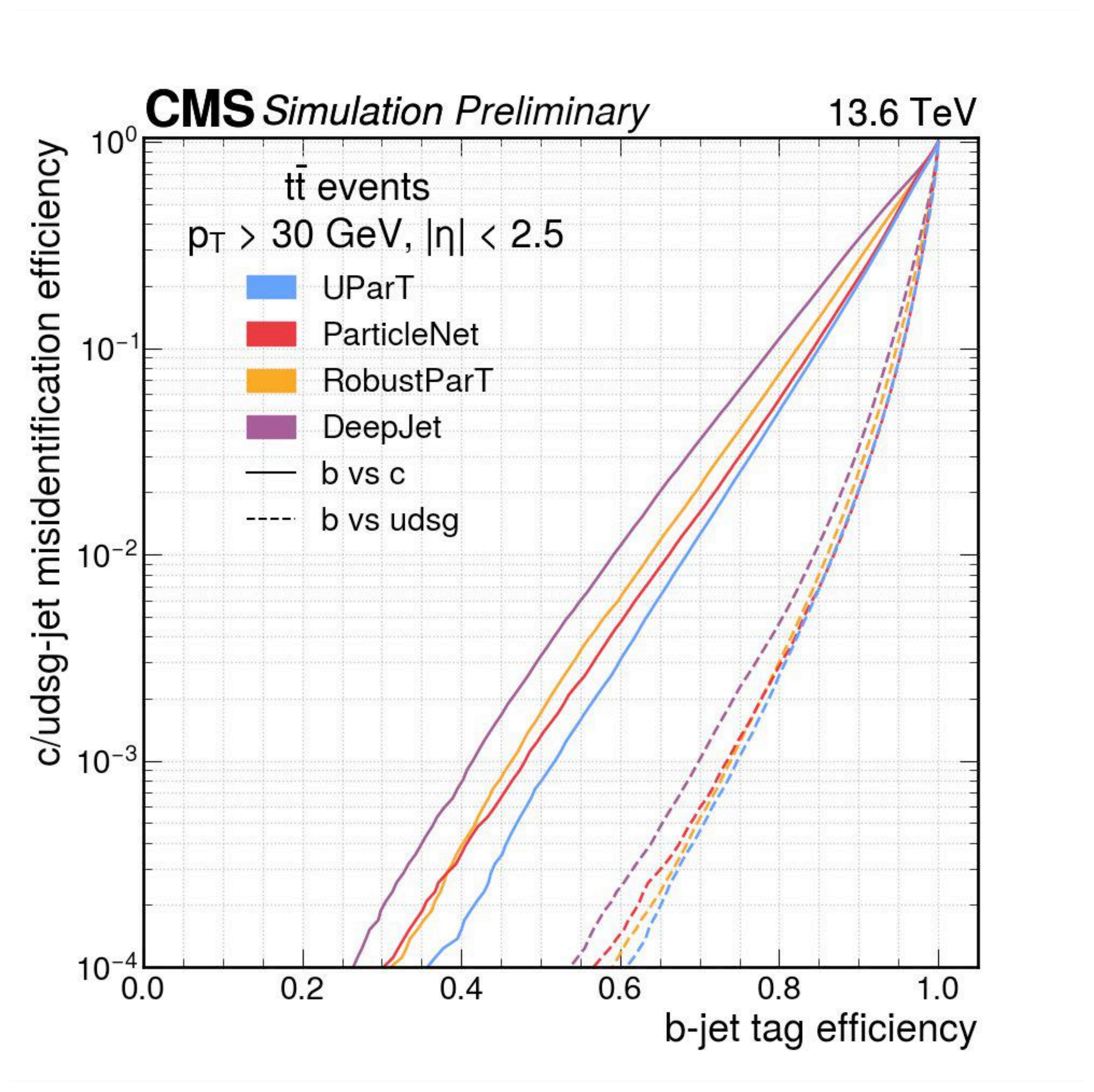
Tagging performance in pp collision

b-tagging

Ref: CMS-DP-2024-066

- Extended class: extending from **b** and c jet identification to s and hadronic tau (one per final state) identification
- Extended regression: simultaneous flavor aware jet energy regression and resolution

Significant improvement in b-tagging efficiency!

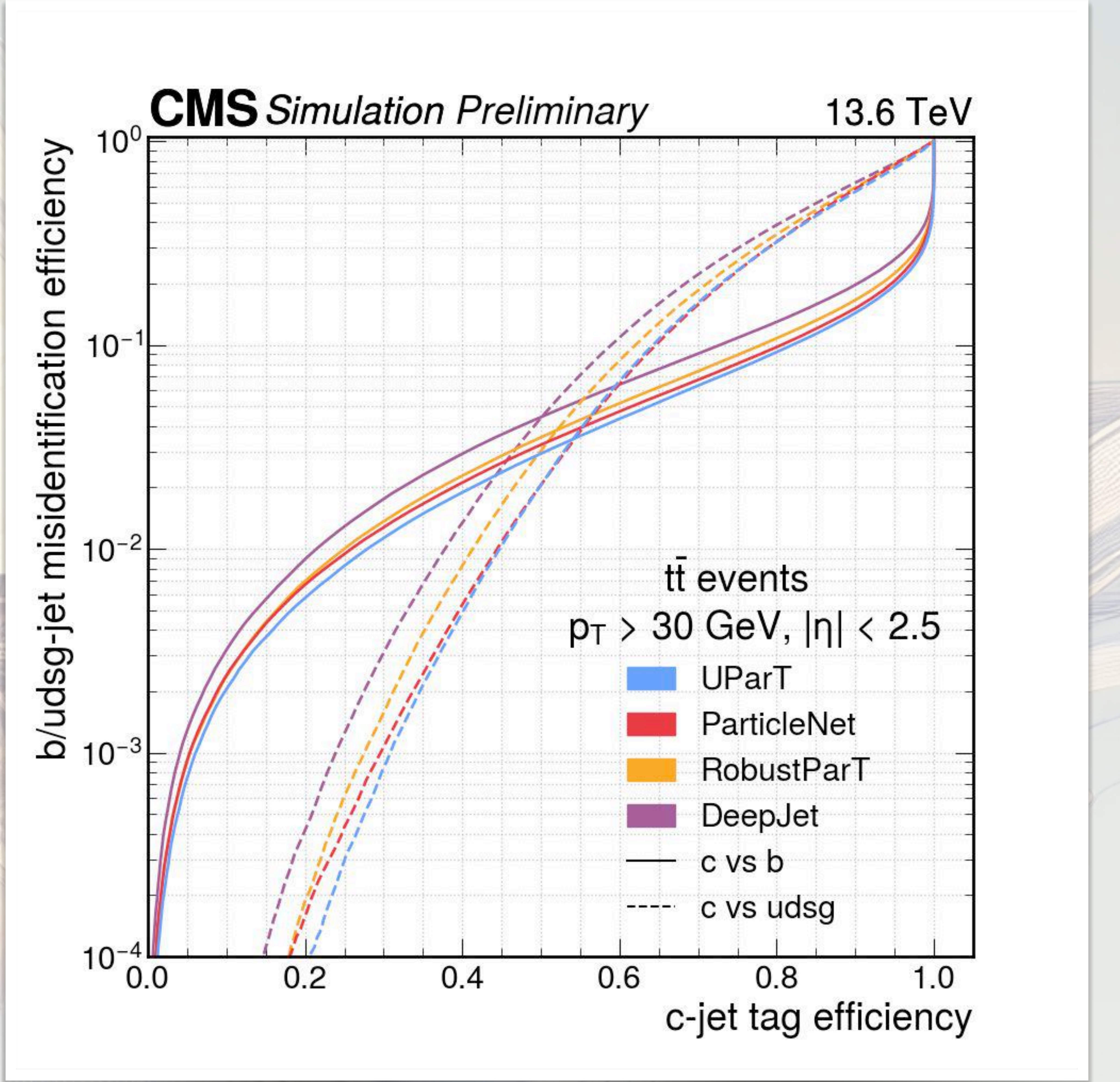


Tagging performance in pp collision

Ref: CMS-DP-2024-066

- Extended class: extending from b and **c** jet identification to s and hadronic tau (one per final state) identification
- Extended regression: simultaneous flavor aware jet energy regression and resolution

c-tagging



Improvement in c-tagging efficiency and c vs b discrimination

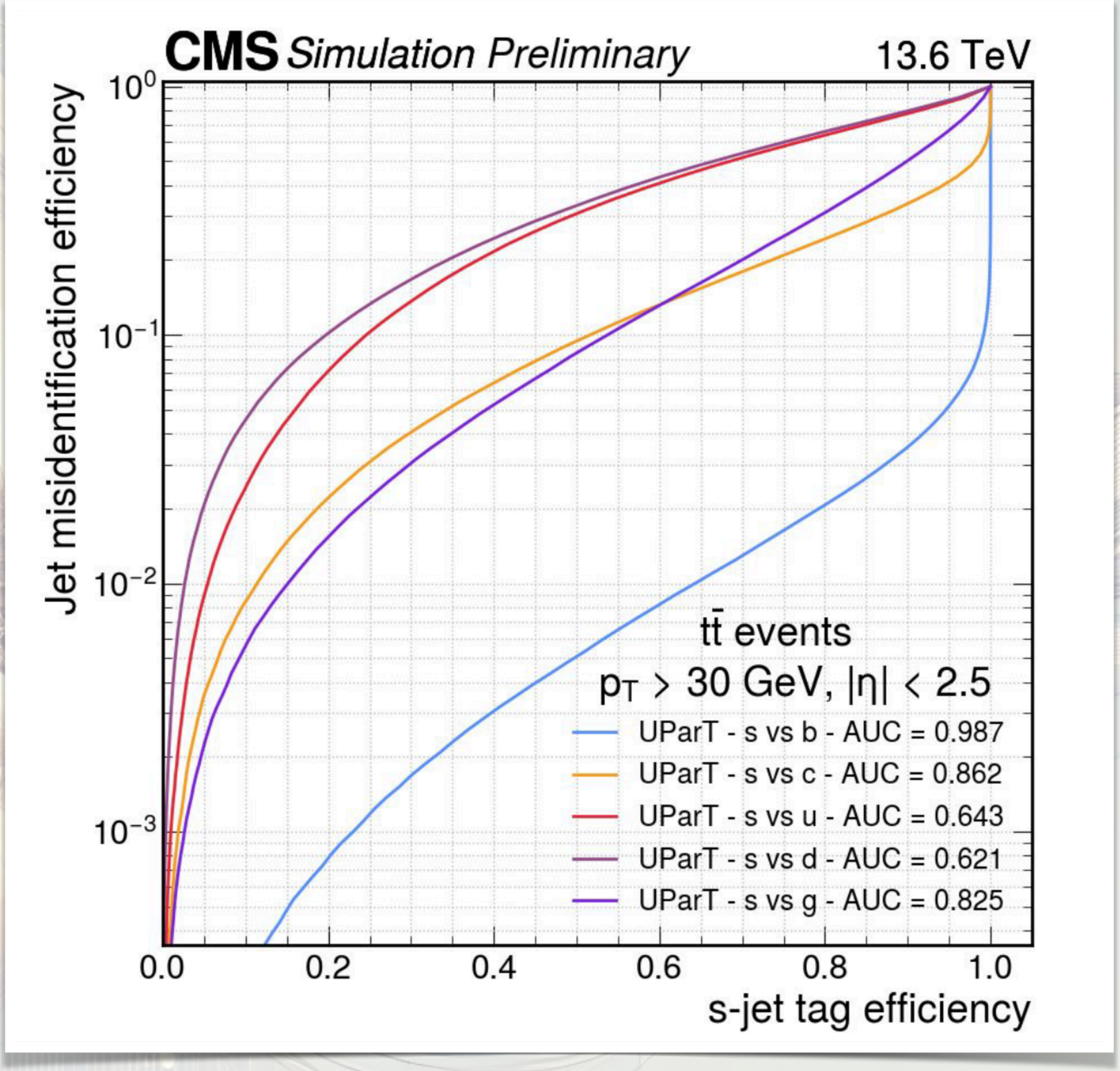
Tagging performance in pp collision

Ref: CMS-DP-2024-066

s-tagging

- Extended class: extending from b and c jet identification to **s** and hadronic tau (one per final state) identification
- Extended regression: simultaneous flavor aware jet energy regression and resolution

First attempt of s-tagging in CMS!



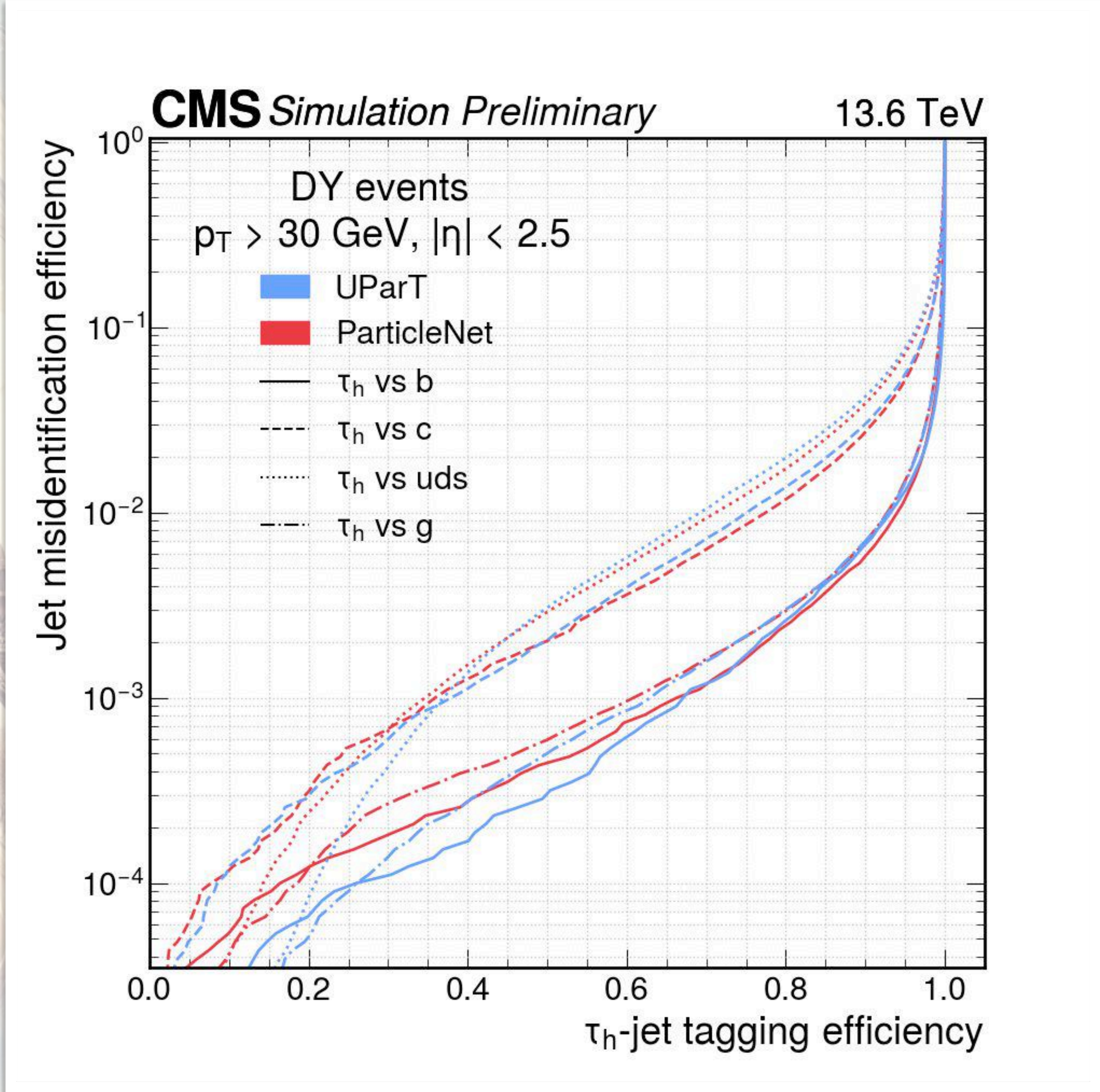
Tagging performance in pp collision

- Extended class: extending from b and c jet identification to s and hadronic **tau** (one per final state) identification
- Extended regression: simultaneous flavor aware jet energy regression and resolution

Improvement in τ-tagging performance

Ref: CMS-DP-2024-066

τ-tagging



UParT Modifications: II

- Extended class: extending from b and c jet identification to s and hadronic tau (one per final state) identification
- Extended regression: simultaneous flavor aware jet energy regression and resolution

A modified Loss function

Classification | Regression | Quantile regression (resolution estimation)

$$L = \underbrace{\mathrm{CatEntropy}(\mathrm{x}, \mathrm{x}_{\mathrm{truth}})}_{\text{Classification}} + \underbrace{\lambda \times \log(\cosh(y - y_{truth})}_{\text{Regression}} + \underbrace{\gamma \times [\rho_{0.16}(\mathrm{z} - \mathrm{z}_{\mathrm{truth}}) + \rho_{0.84}(\mathrm{z} - \mathrm{z}_{\mathrm{truth}})]}_{\text{Quantile regression (resolution estimation)}}$$

Jet Energy Regression

Ref: CMS-DP-2024-066

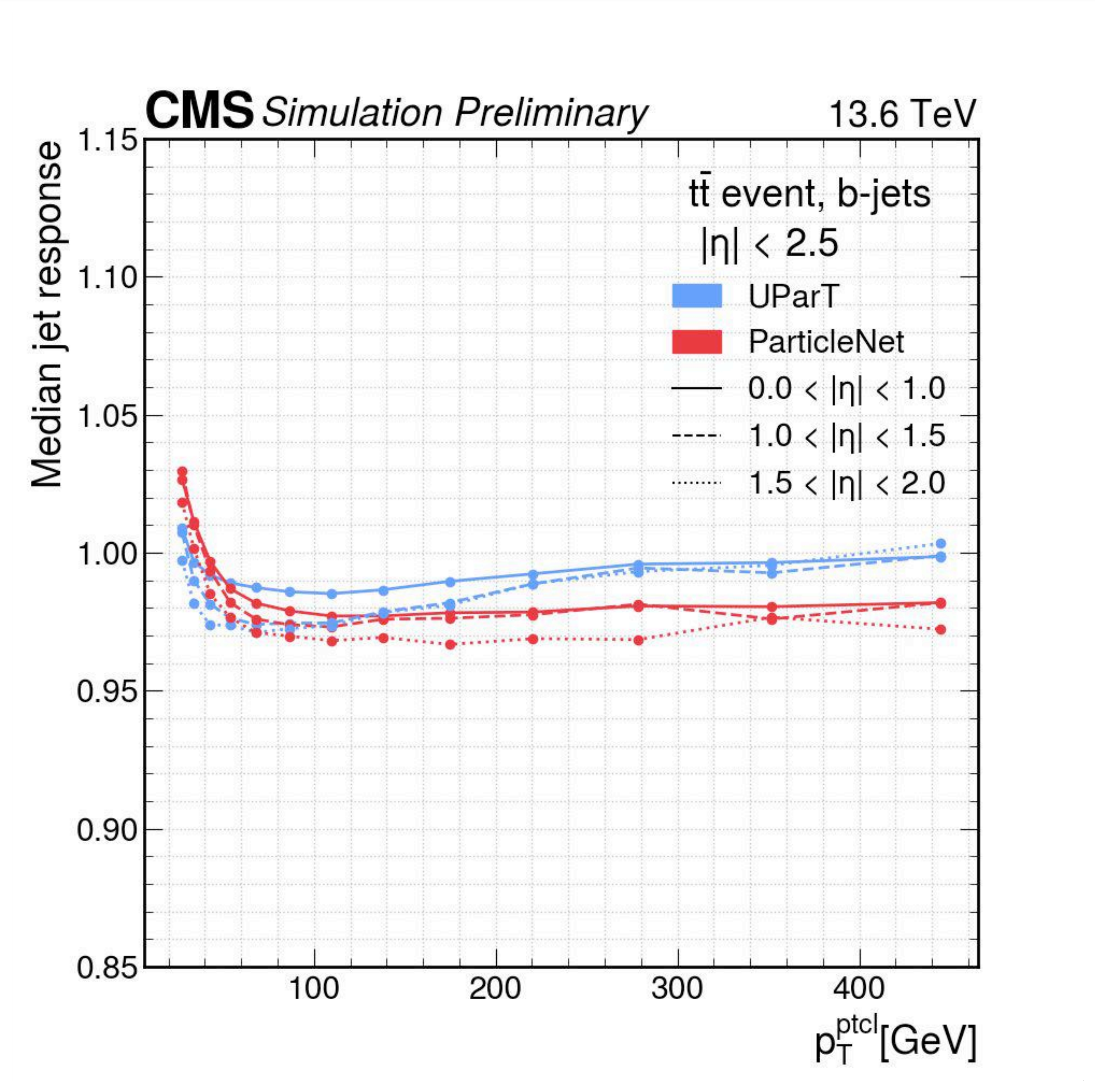
- Extended class: extending from b and c jet identification to s and hadronic tau (one per final state) identification
- Extended regression: simultaneous flavor aware jet energy regression and resolution

A modified Loss function

Classification	Regression	Quantile regression (resolution estimation)

$$L = \mathrm{CatEntropy}(x, x_{\mathrm{truth}}) + \lambda \times \log(\cosh(y - y_{truth})) + \gamma \times [\rho_{0.16}(z - z_{\mathrm{truth}}) + \rho_{0.84}(z - z_{\mathrm{truth}})]$$

- *UParT trained in full Run3 sample*
- *Shows improvement over ParticleNet*
  - *Effect is more in the most extreme |η| bin*



Jet Energy Resolution

Ref: CMS-DP-2024-066

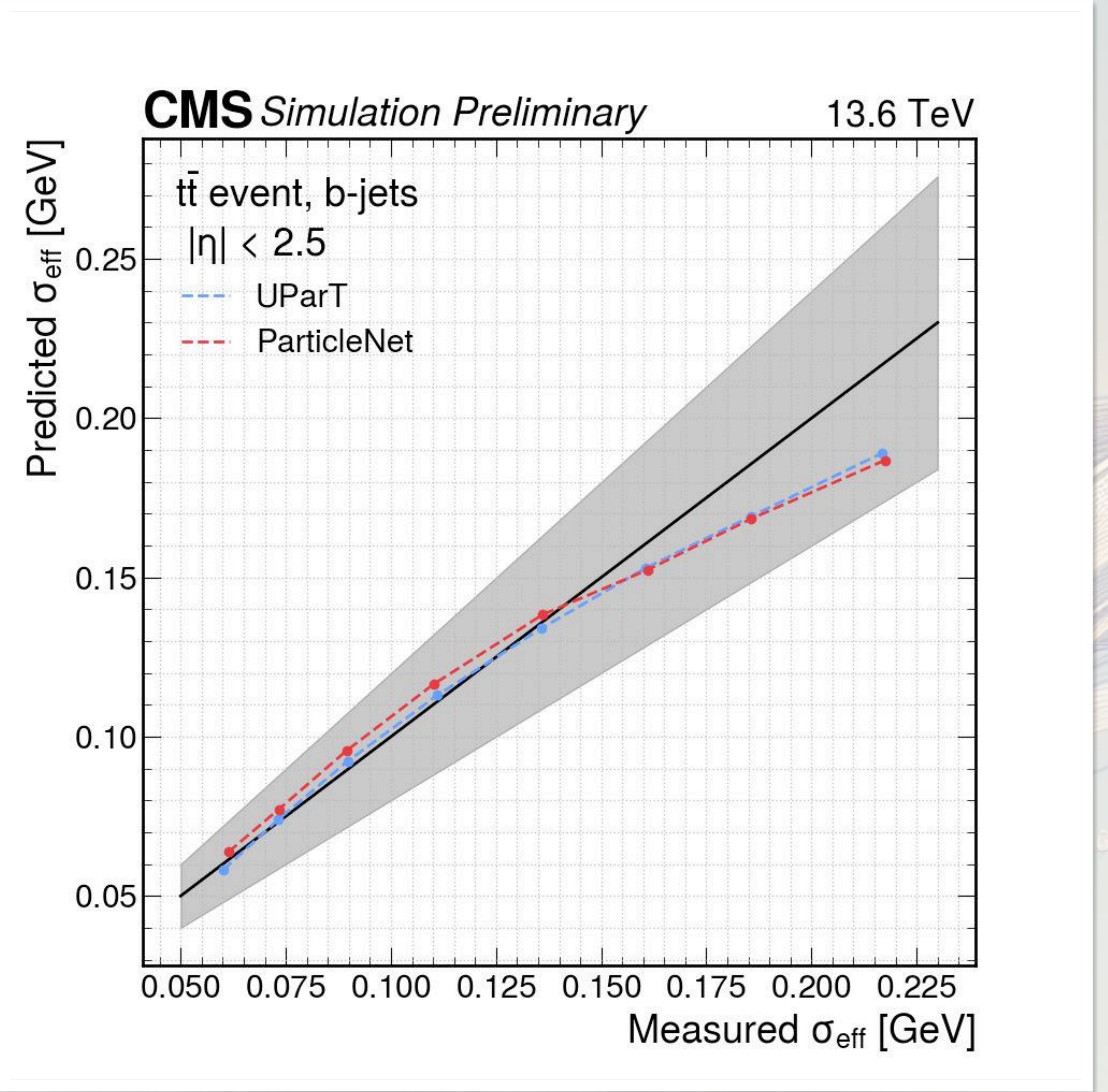
- Extended class: extending from b and c jet identification to s and hadronic tau (one per final state) identification
- Extended regression: simultaneous flavor aware jet energy regression and resolution

A modified Loss function

Classification	Regression	Quantile regression (resolution estimation)

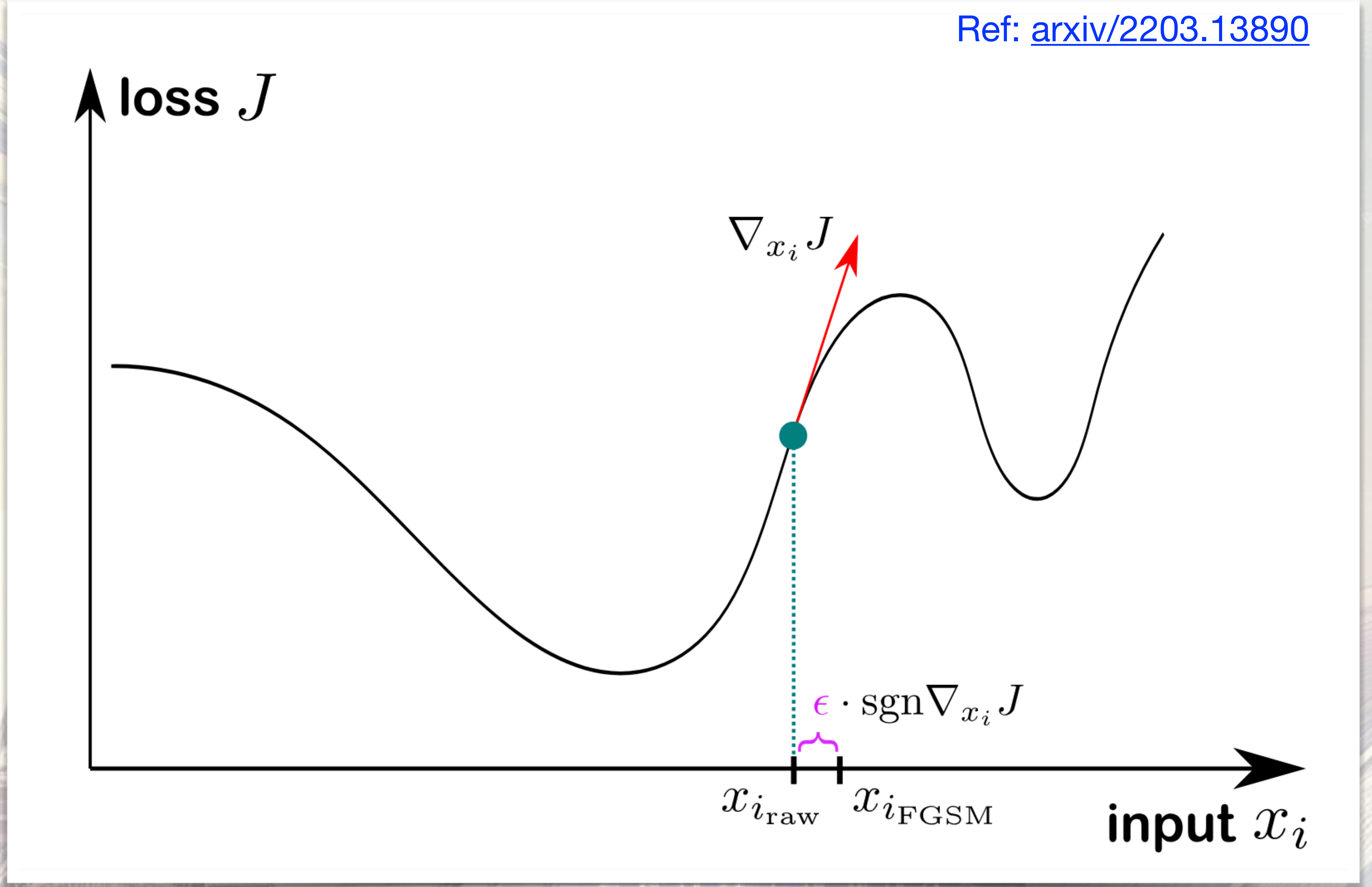
$$L = \mathrm{CatEntropy}(x, x_{\mathrm{truth}}) + \lambda \times \log(\cosh(y - y_{truth})) + \gamma \times [\rho_{0.16}(z - z_{\mathrm{truth}}) + \rho_{0.84}(z - z_{\mathrm{truth}})]$$

- *UParT and ParticleNet show similar jet resolution response*
- *Capable of estimating a per-jet resolution from the quantile regression*



UParT Modifications: II

- Adversarial training strategy:
  - Introduce perturbations in loss function against the gradient
  - Scrutinize robustness of the model
  - Reduce the observed differences prior to any calibration



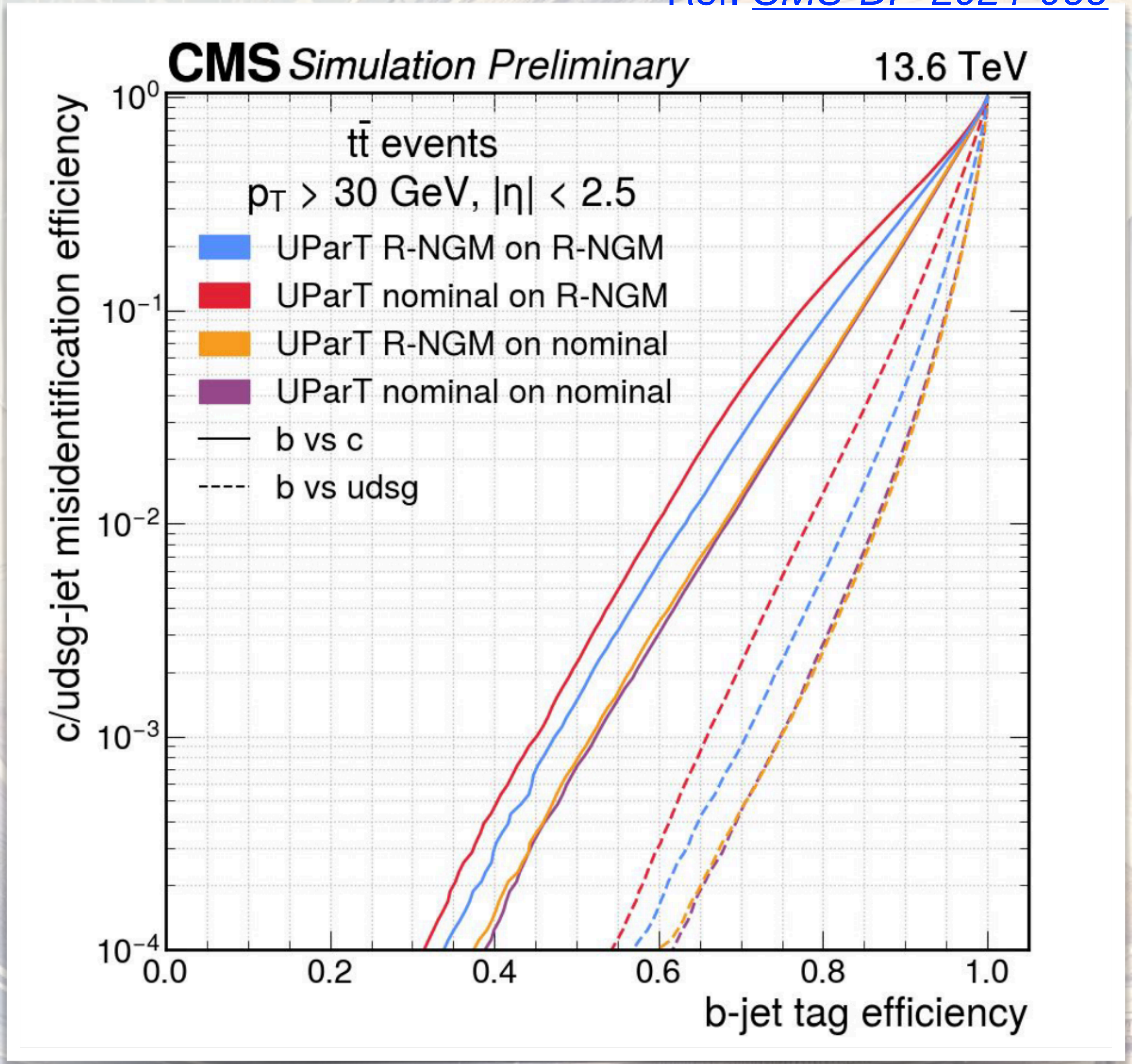
- Adversarial Attack - Rectified Normed Gradient Method (RNGM):
  - $x_{i,adv} = x_i + \epsilon |\nabla CE(x_i, \theta)|$

➡Preserving the Particle Cloud representation and the feature importance mapping

Robustness of the Attack: pp collision

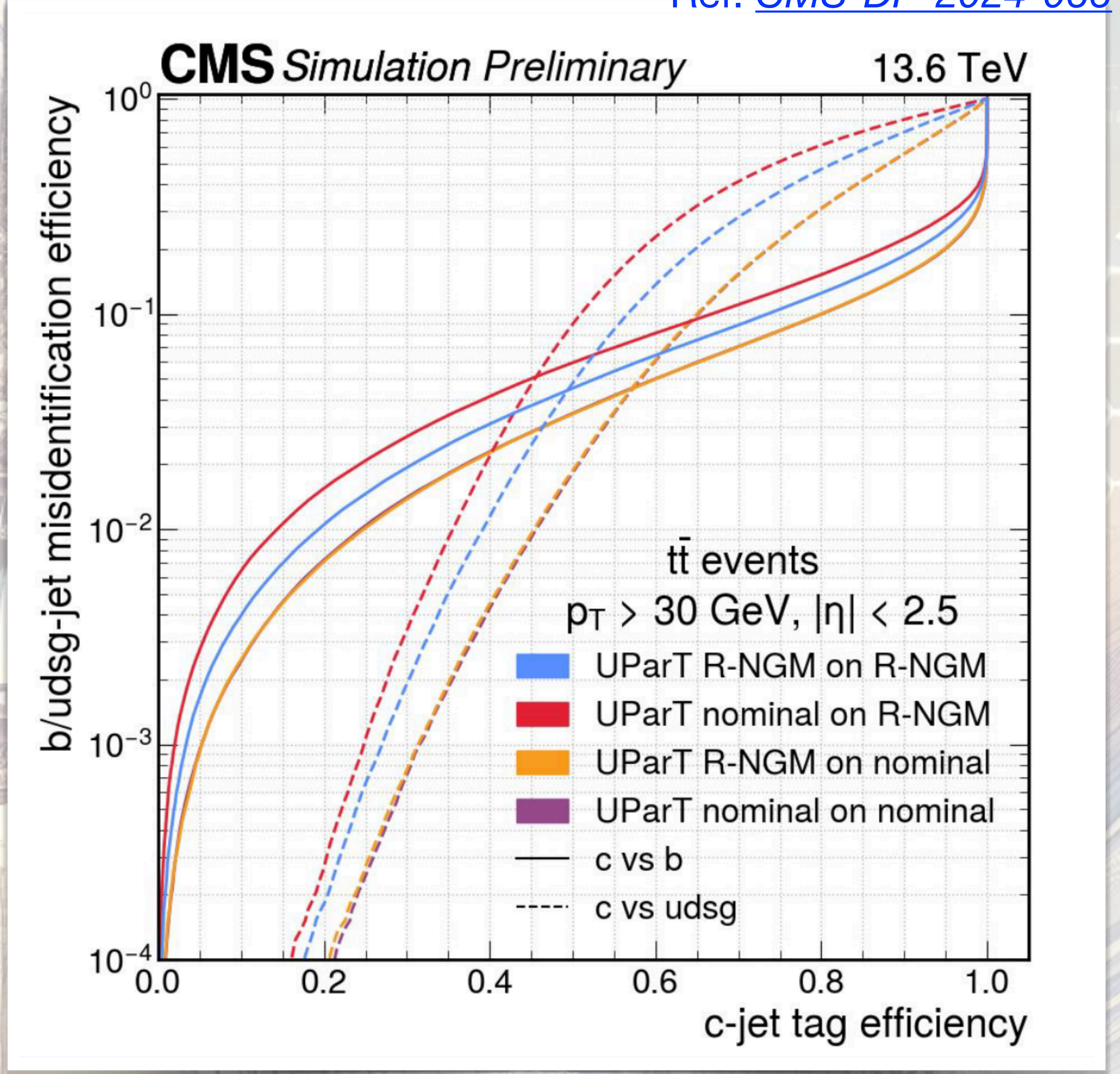
b-tagging

Ref: CMS-DP-2024-066



c-tagging

Ref: CMS-DP-2024-066

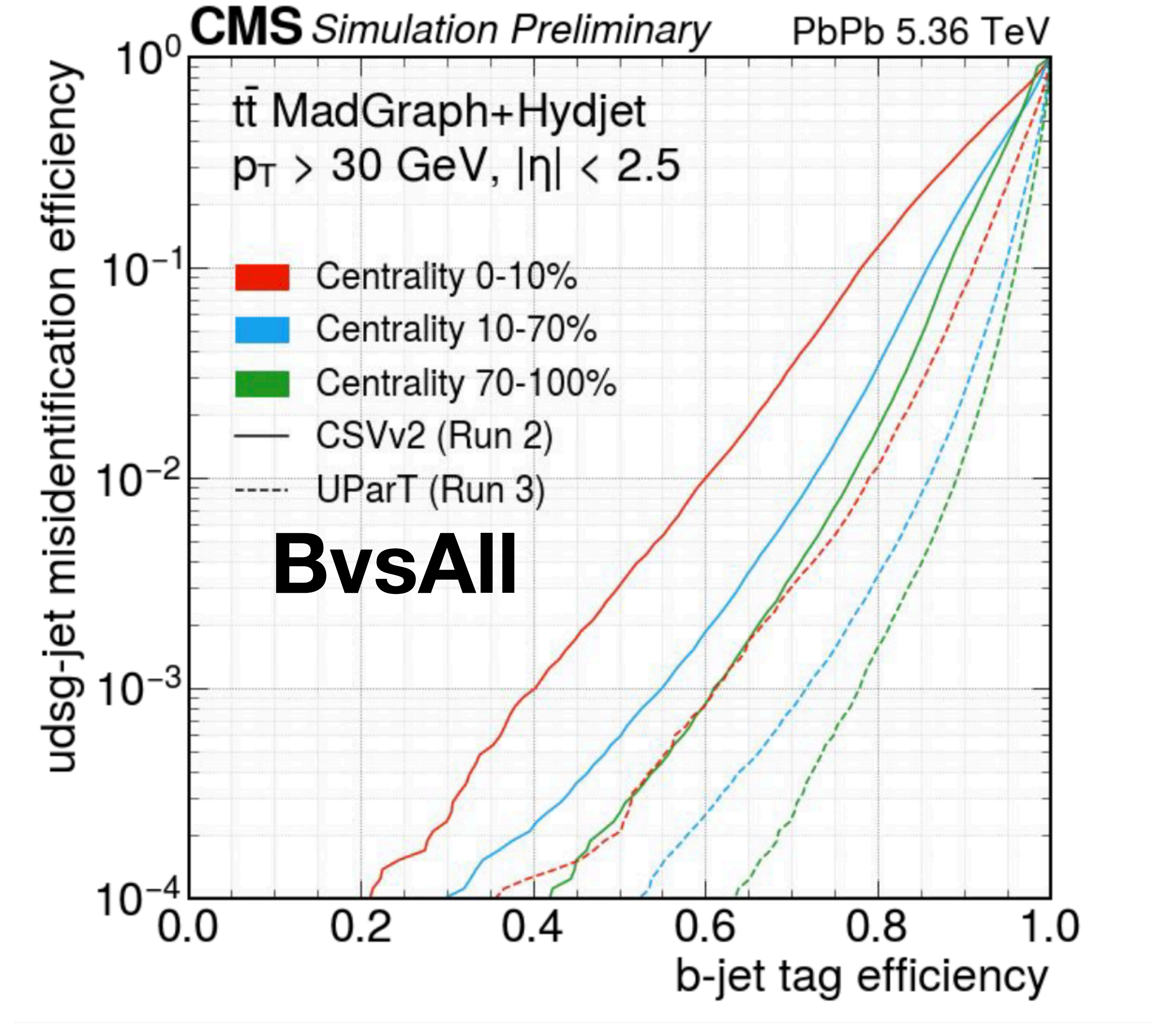


R-NGM adversarial strategy shows substantial improvement in robustness compared to the nominal training!

UParT in PbPb collision

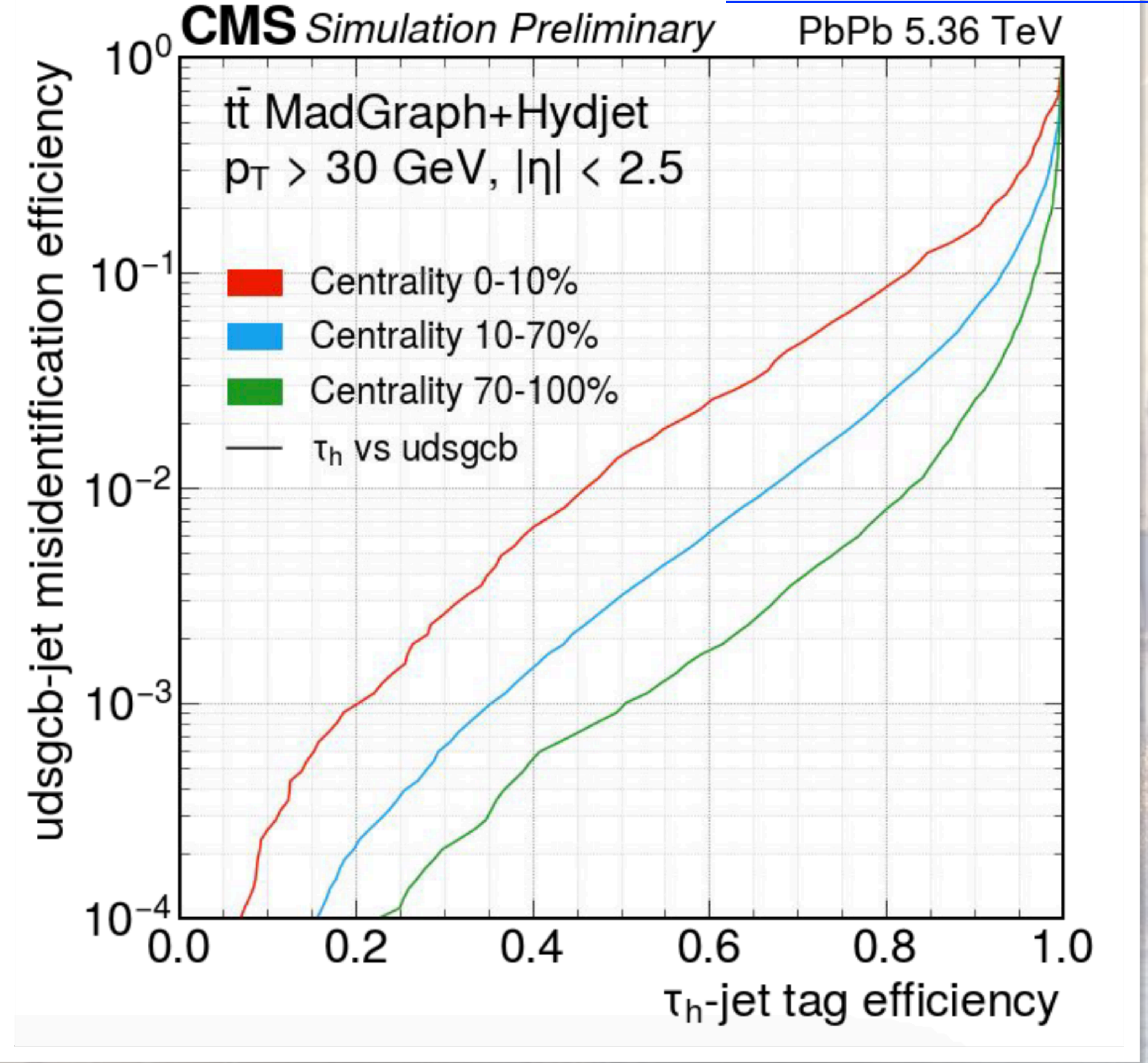
- During Run2 (2015-2018): Lead-lead (PbPb) collision relied on MLP - room for performance improvement
- UParT - ParT model and RNGM attack enables a more robust flavor tagging in PbPb
- UParT retrained specifically with $t\bar{t}$ and embedded in PbPb Hydrodynamics+JET underlying events samples - first time in Run3 data

Ref: *CMS-DP-2024-088*



Significant improvement in performance (light-jet rejection by a factor of 10)

Ref: *CMS-DP-2024-088*

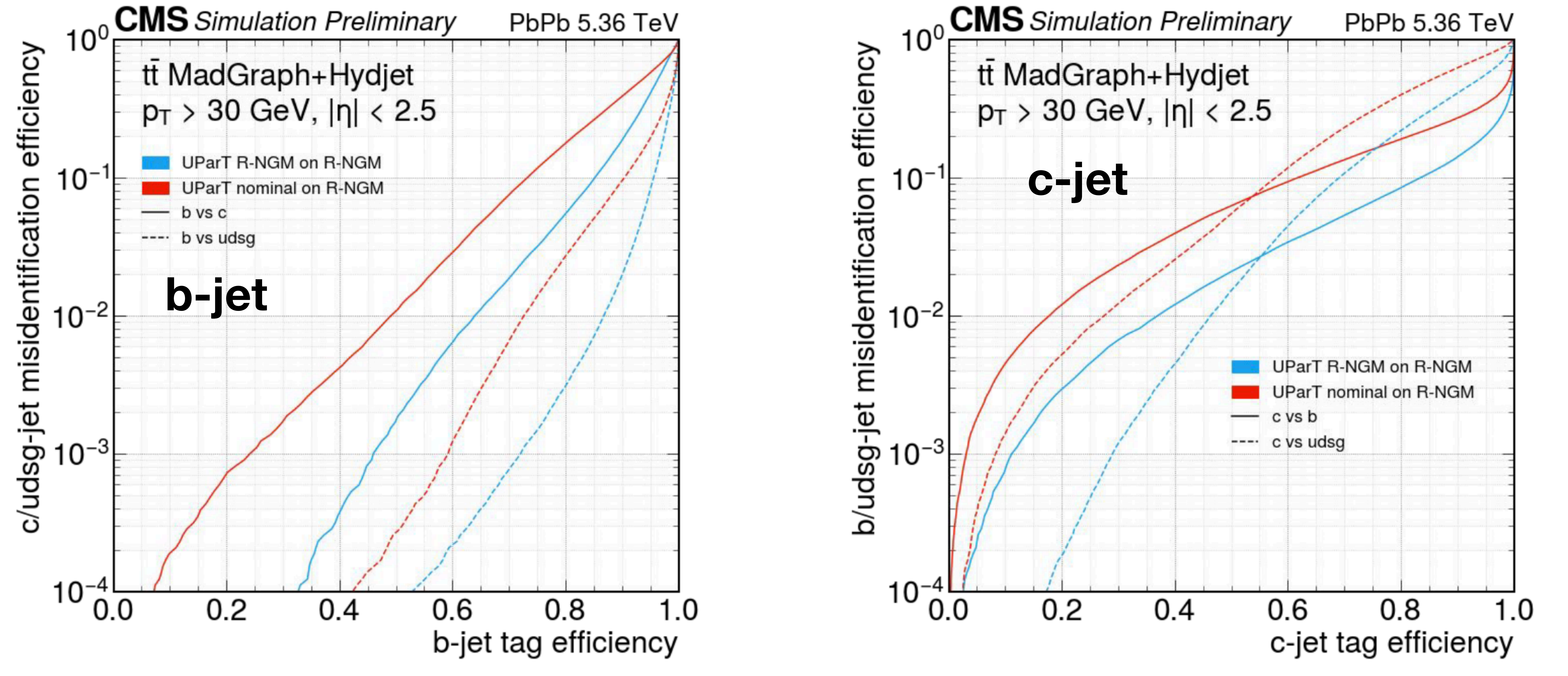


First tau identification in heavy ion

Centrality: degree of overlap of the two colliding nuclei

Robustness of the Attack: PbPb collision

Ref: CMS-DP-2024-088

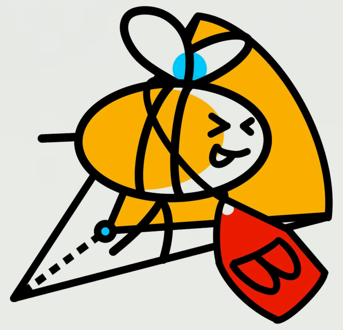


R-NGM adversarial strategy shows substantial improvement in robustness compared to the nominal training!

Summary

- **Rapid Development of Machine Learning Architectures over the past decade**:
  - From simple discriminator to ***UParT* -** increasing complexity and capability of models
- A unified approach to jet tagging based on a Transformer model named UParT is explored including features like
  - Extended flavor classification
  - Include jet based τ identification
  - Flavor aware jet energy regression and resolutions
  - Achieves best performance in heavy flavor (b,c) tagging
  - First attempt to tag s-jets
- Outstanding performance in heavy-ion collisions across different PbPb centralities

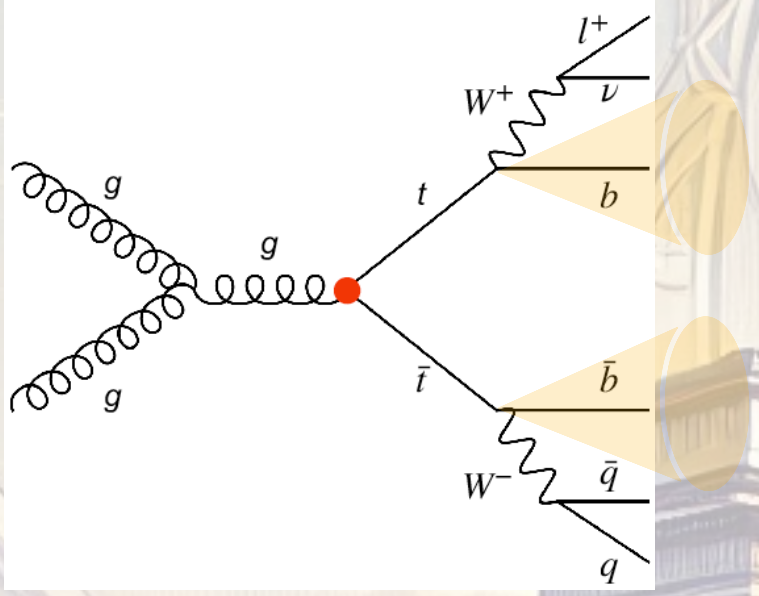
Thank you for listening!



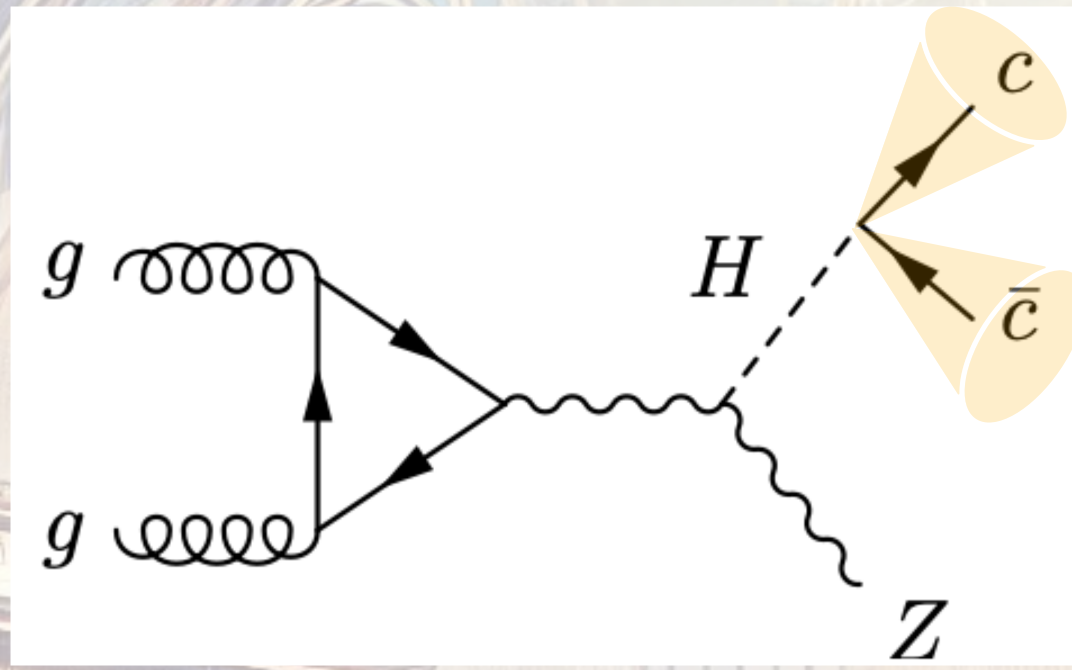
Spare

Theory

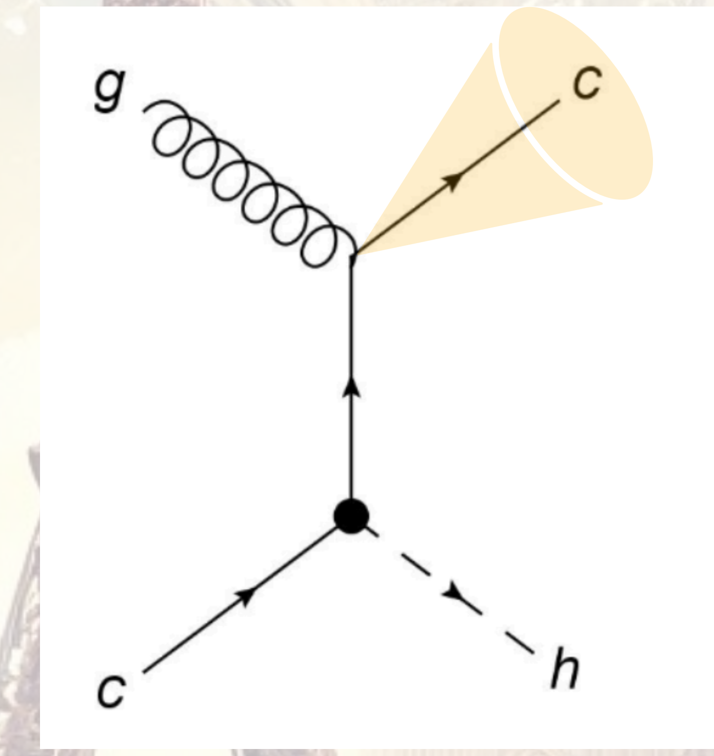
- **Heavy flavor jets =** jets originating from b (***b jets***) or c (***c jets***) quarks arising from the process of hadronization
- Important in Standard Model (SM), Top, Higgs(H->bb,cc), BSM and SUSY processes



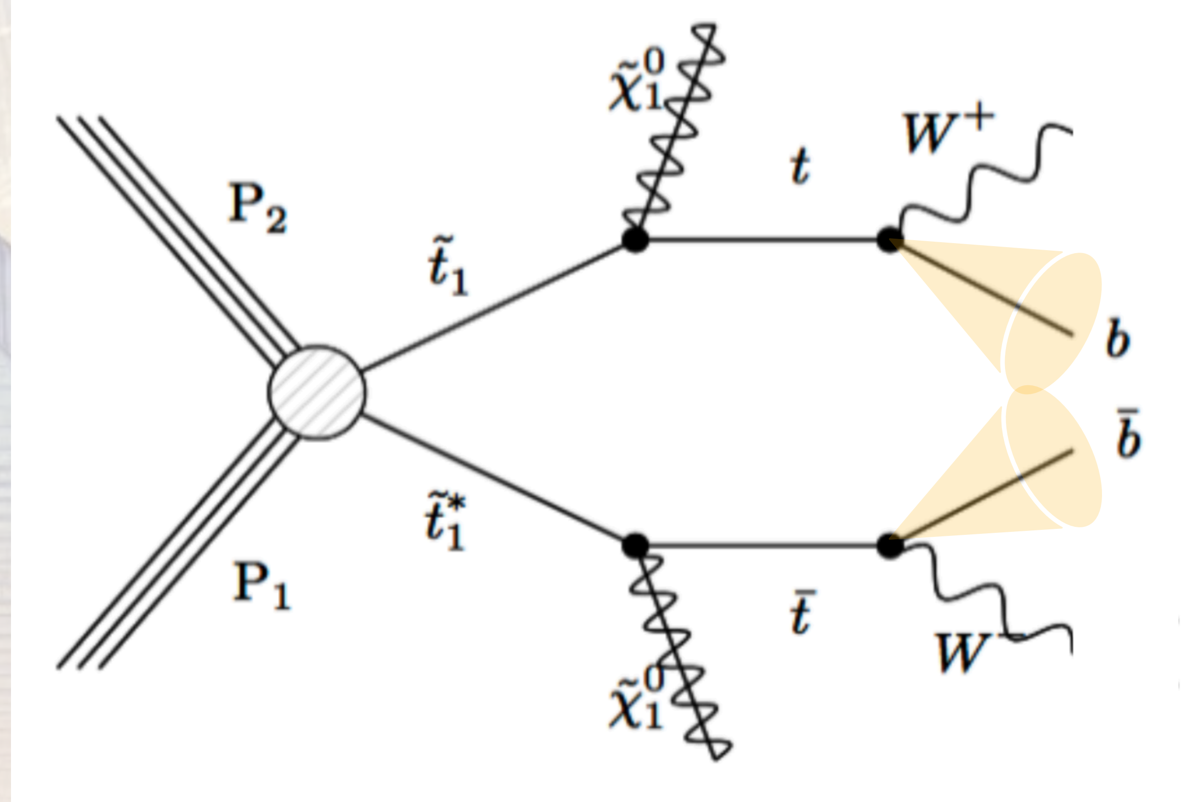
$t\bar{t}$



VHcc[1]

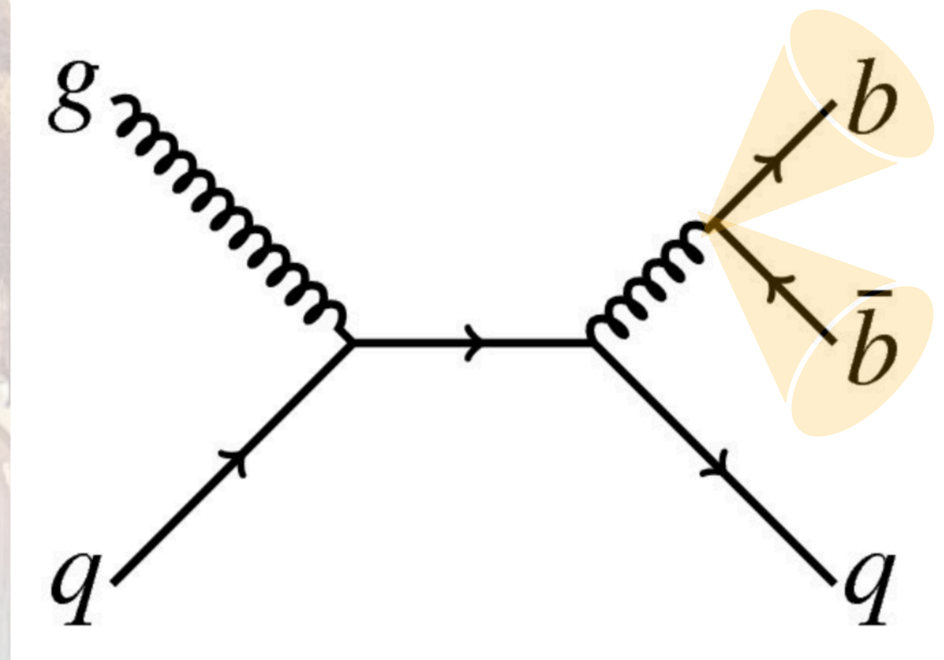


Higgs+c[2]



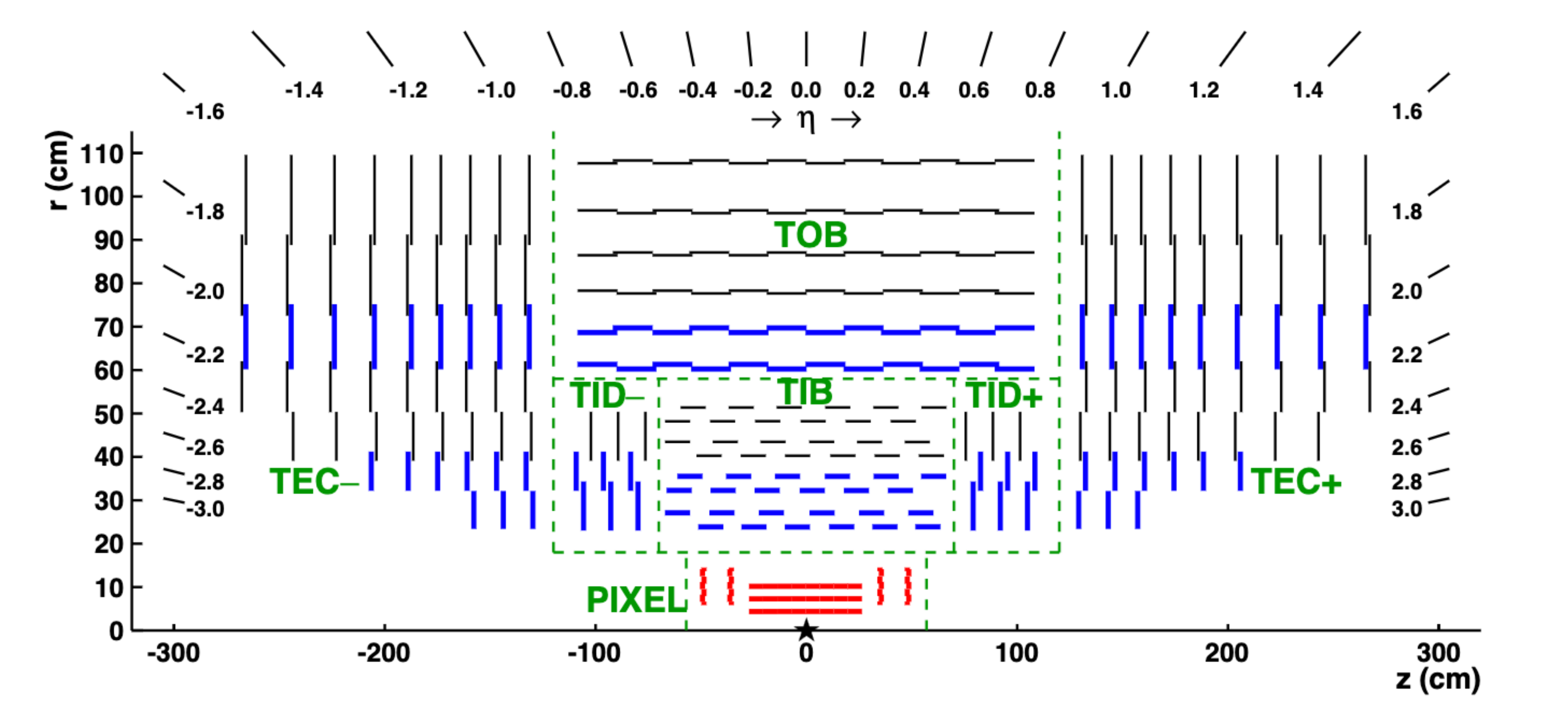
SUSY stop-> SM + MET

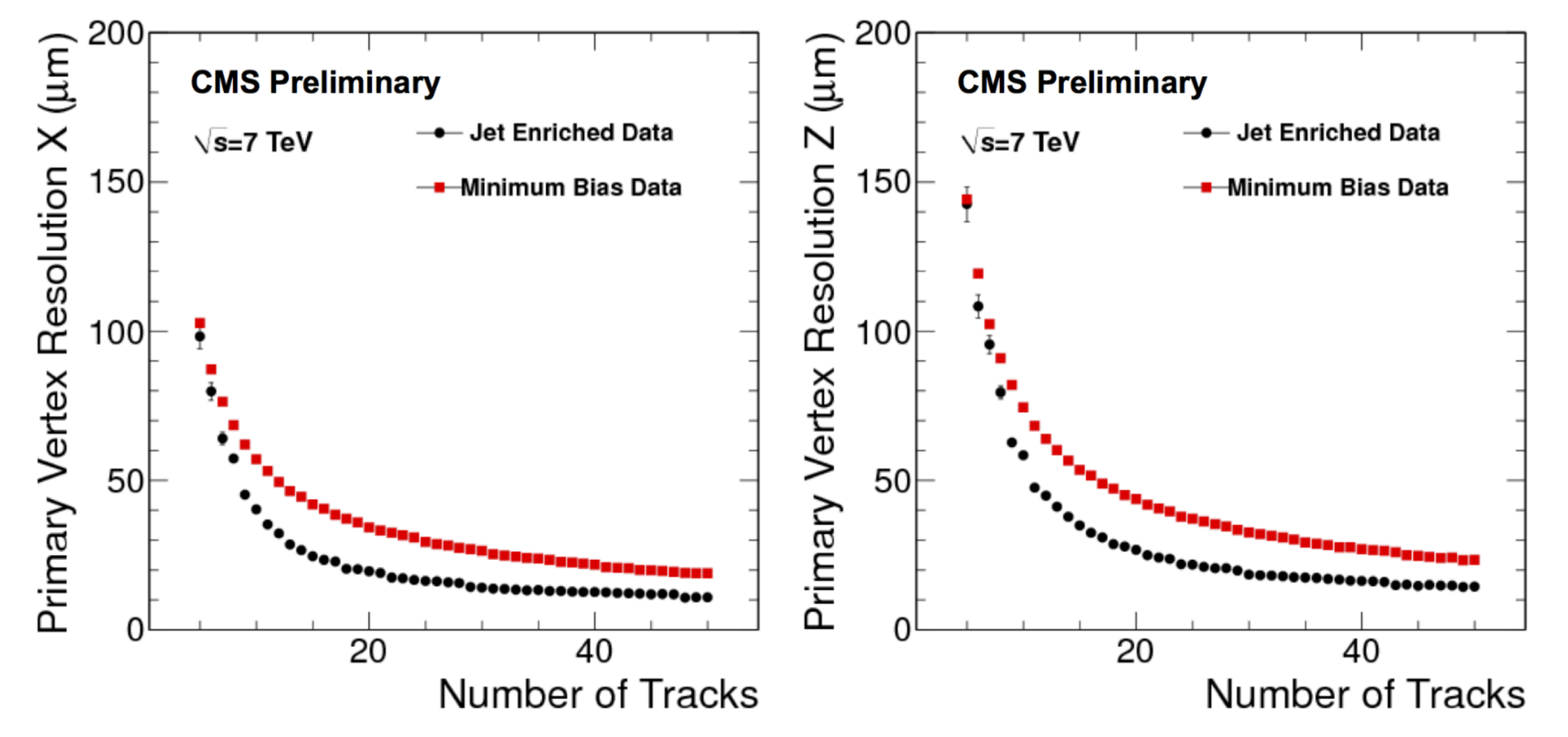
- **QCD:** understanding heavy-parton effects



[1] https://www.dpg-verhandlungen.de/year/2024/conference/karlsruhe/part/t/session/20/contribution/6
[2] https://www.dpg-verhandlungen.de/year/2024/conference/karlsruhe/part/t/session/71/contribution/7

CMS Tracking





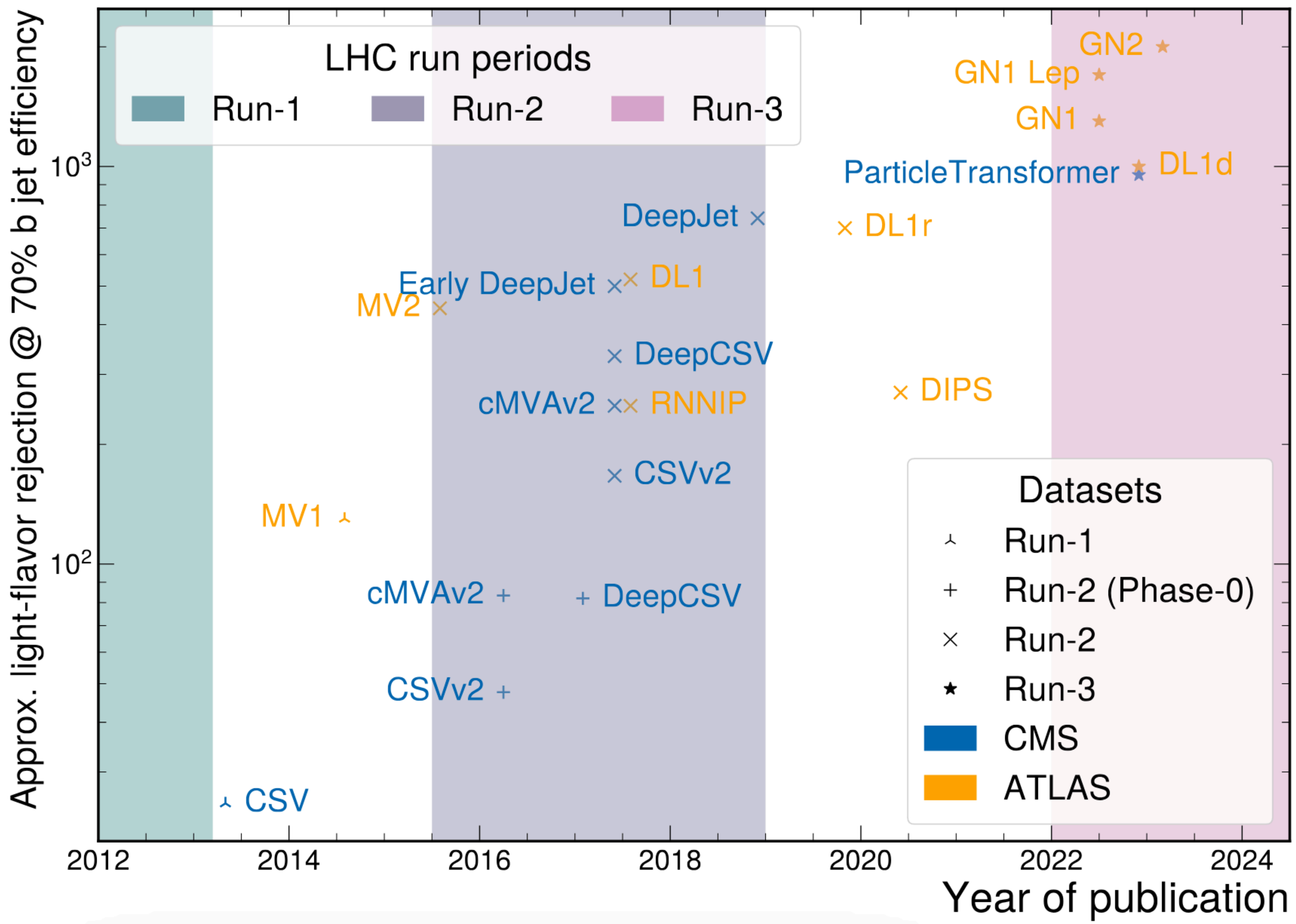
Subdetector	Radius [cm]	Sensor size [μm]	Resolution [μm]	<Hits on track>
Pixel	4.4-10.2	100×150	Rϕ:10 z:20	3
Strip tracker	25.5-110	$\sim$ 100	$\sim$ 15– $\sim$ 45	13

https://pos.sissa.it/190/041/pdf

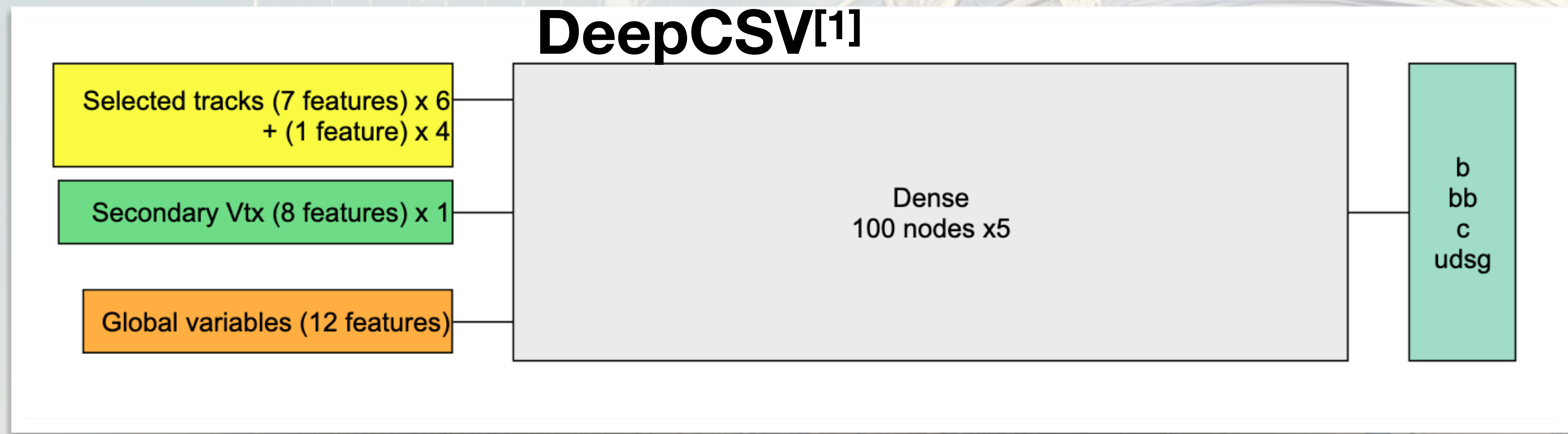
Ideal to observe in CMS, though challenging!

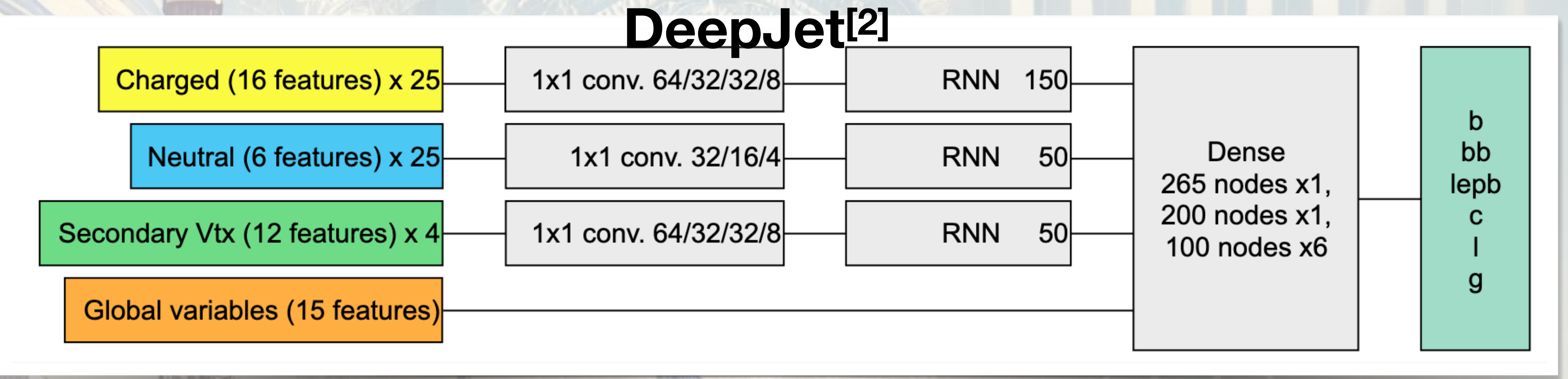
[1] http://cds.cern.ch/record/1279383/files/TRK-10-005-pas.pdf

Historical evolution of particle taggers in CMS



DeepCSV and DeepJet

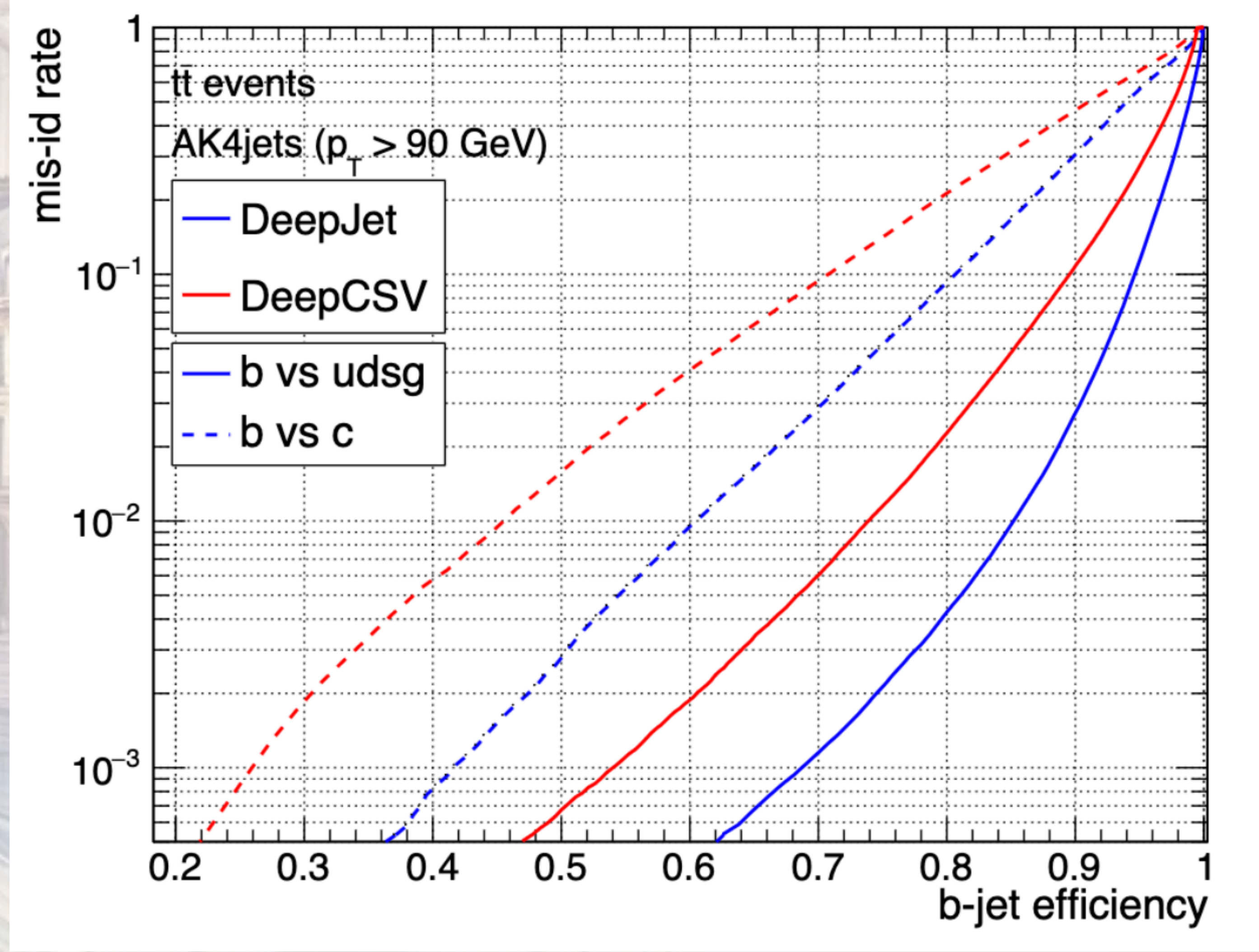




- **Fully connected neural network (dense)**
- Combines properties from **selected** tracks, secondary vertices and global variables directly (66 features)
- Only a small subset of the charged jet constituents pass stringent quality criteria
  - clean and simple environment for the classifier
  - **information loss - potential performance degradation**

- **Convolution, RNN,** and **Dense** layers
- **Does not rely on a selection** of the jet constituents
  - better purity, more number of inputs
- Full information of all **jet constituents, charged and neutral particles, secondary vertices, and global event variables simultaneously**

~20% gain in efficiency at 10^{-3} misidentification probability[2]

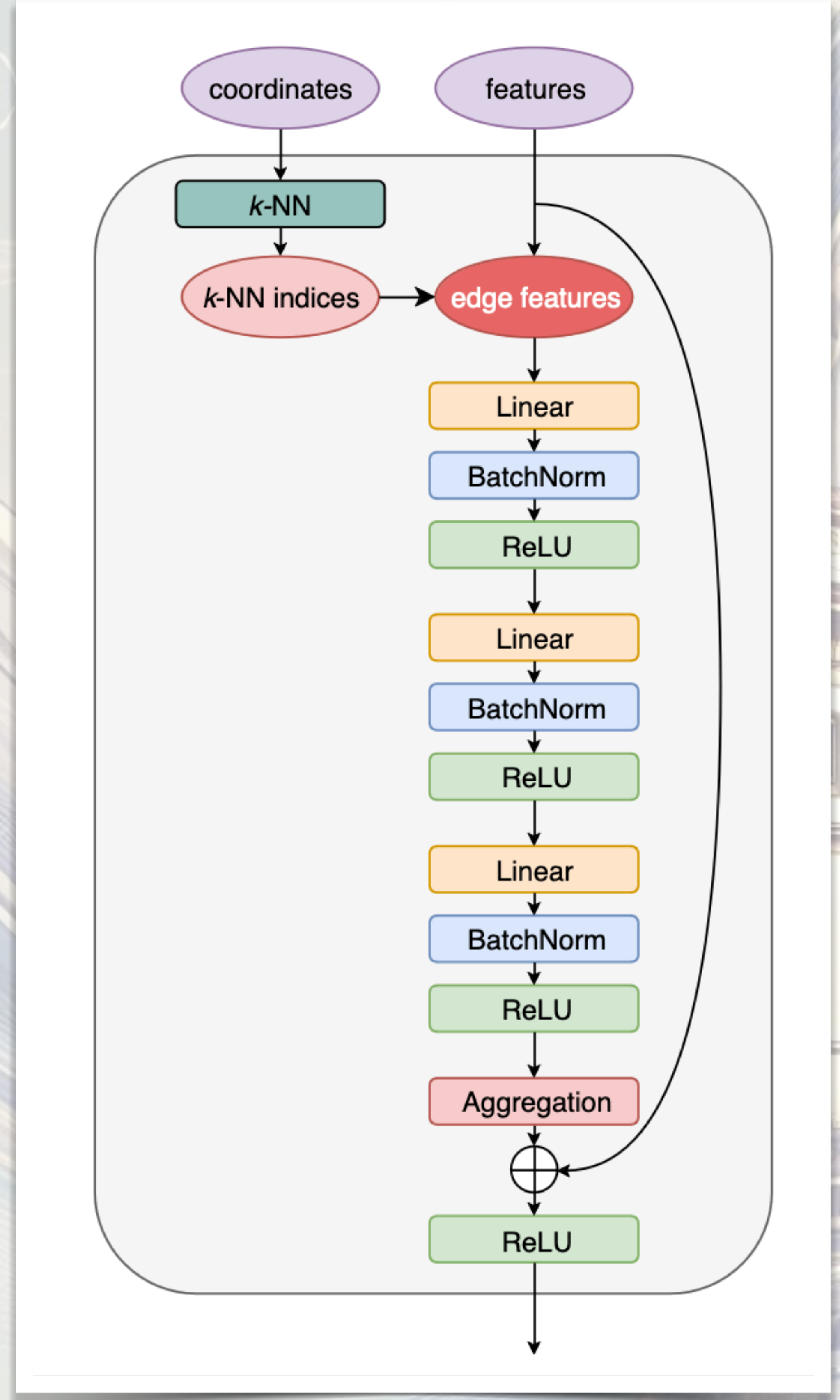


[1] 10.1103/PhysRevD.94.112002
[2] 10.1088/1748-0221/15/12/P12012

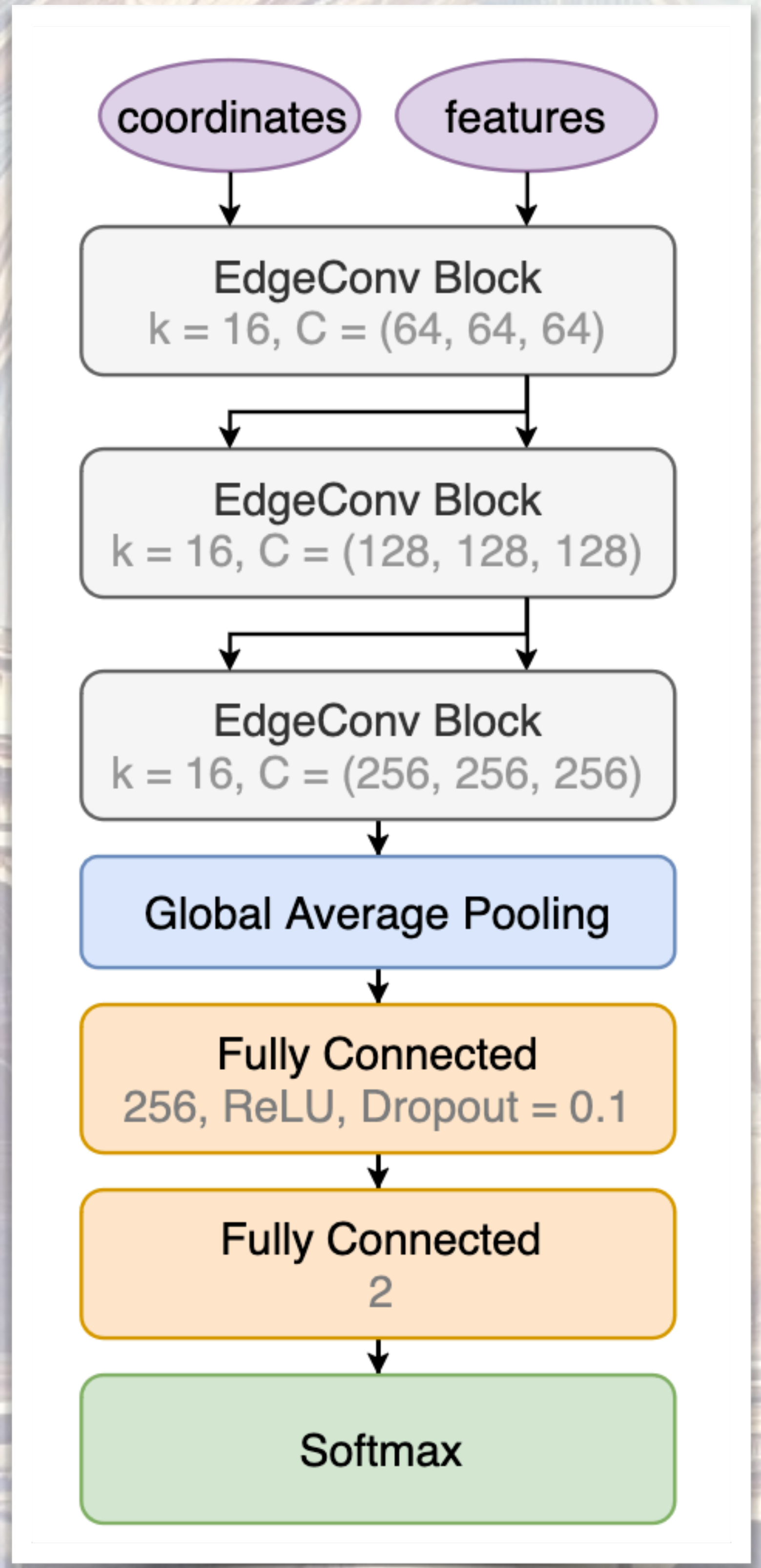
ParticleNet

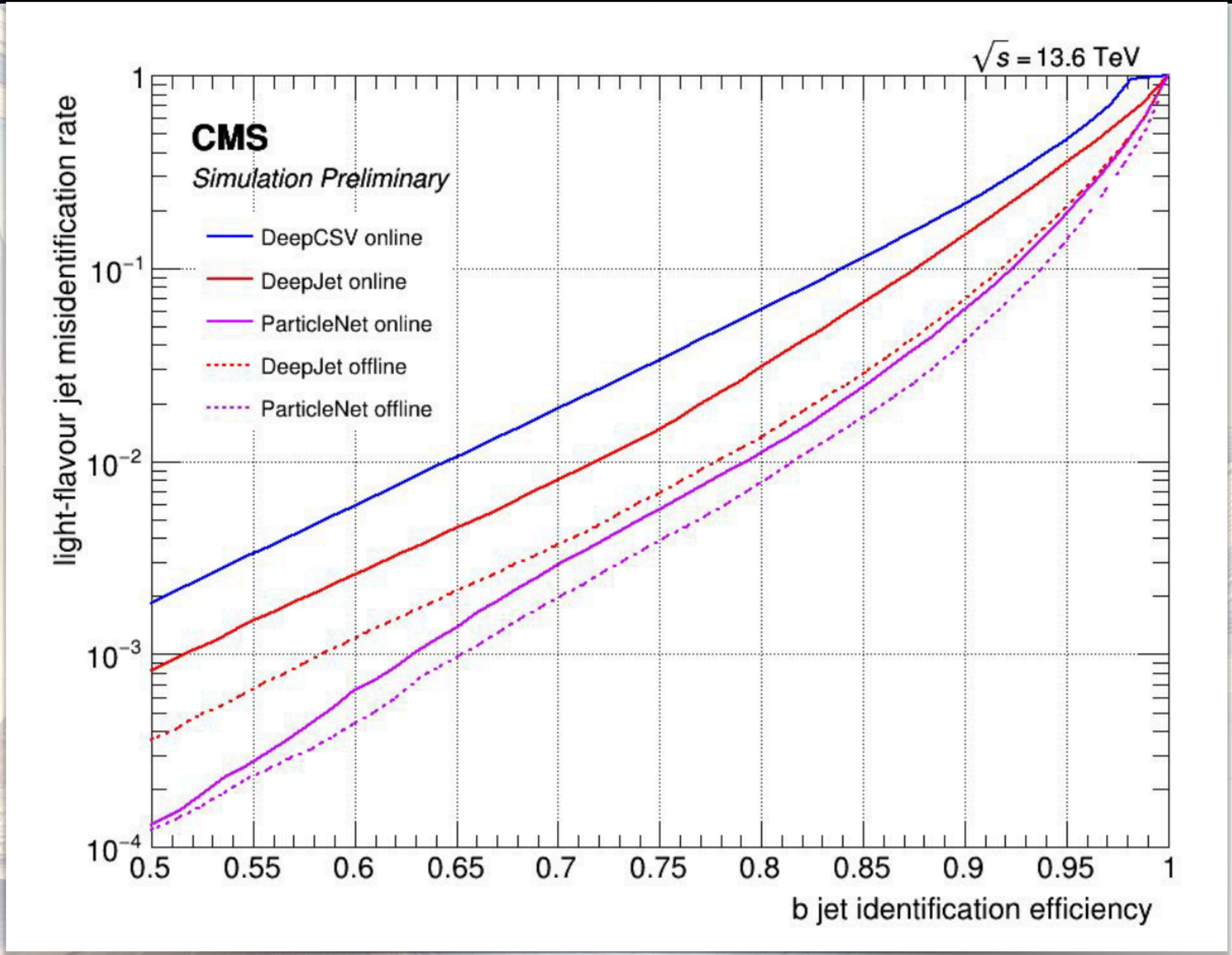
Treat a jet as unordered sets of constituent particles

Perform Edge-convolution and Dynamic Graph Convolutional Neural Network (DGCNN)



One edge-convolution block





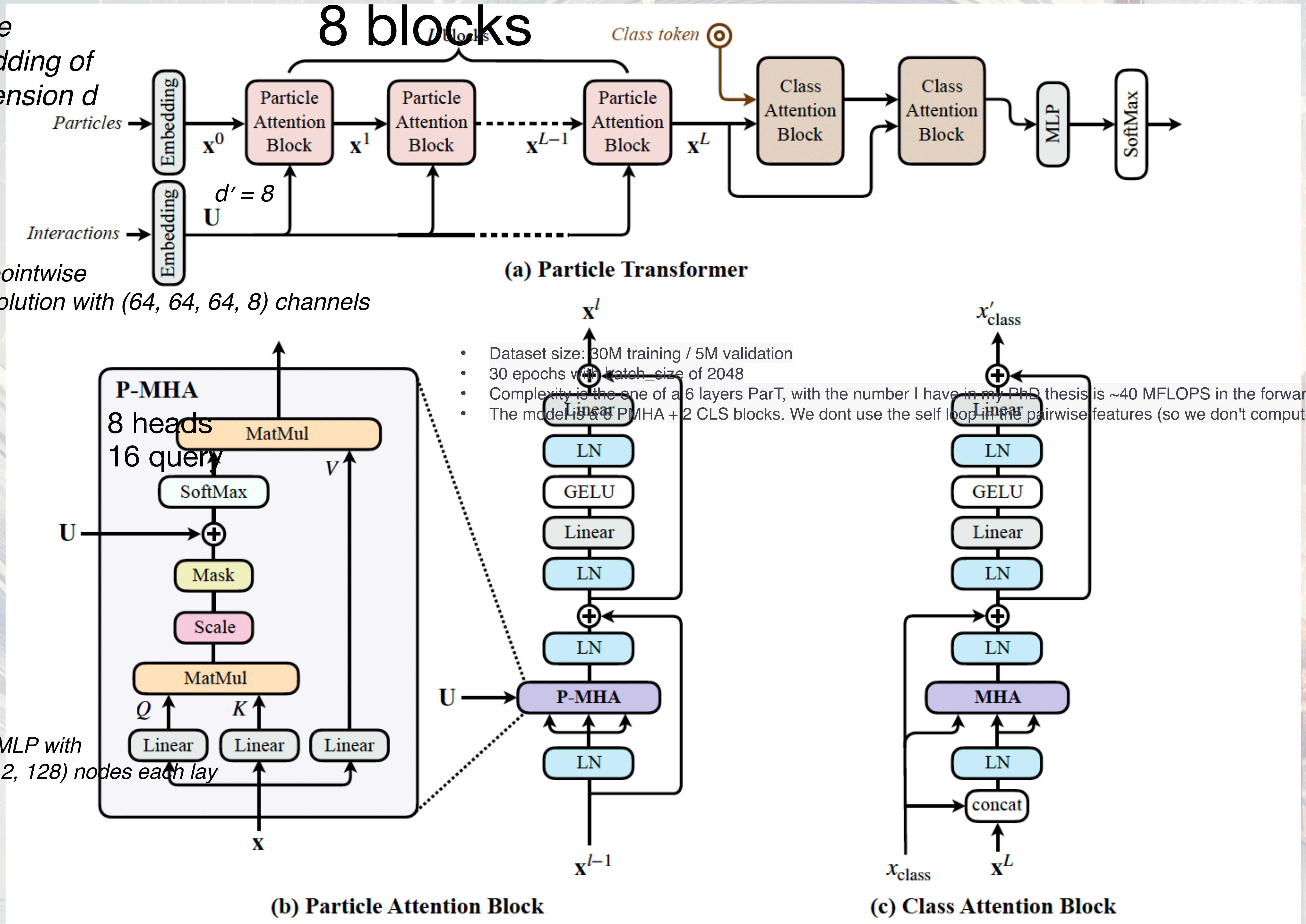
~10% gain in efficiency at 10^{-3} misidentification probability[2]

[1] arxiv1902.08570

[2] CMS_DP2023_021

ParT: Technicals

particle embedding of a dimension d = 128



4-layer pointwise 1D convolution with (64, 64, 64, 8) channels

3-layer MLP with (128, 512, 128) nodes each lay

- Dataset size: 30M training / 5M validation
- 30 epochs with batch_size of 2048
- Complexity is the one of a 6 layers ParT, with the number I have in my PhD thesis is ~40 MFLOPS in the forward path
- The model is a 8 PMHA + 2 CLS blocks. We dont use the self loop in the pairwise features (so we don't compute the pairwise kin of x_ii which makes sense.

- JETCLASS dataset
- 10 classes
- training is performed on the full training set of 100M jets
- Lookahead optimizer for cross entropy loss
- train for a total of1M iterations, 5 epochs.
  - LR remains constant for the first 70% of the iterations, and then decays exponentially, at an interval of every 20 k iterations, down to 1% of the initial value at the end of the training.
- Performance of the model is evaluated every 20 k iterations on the validation set and a model checkpoint is saved. checkpoint with the highest accuracy on the validation set is used to evaluate the final performance on the test set.
- *500 k jets per class (in total 5M) is intended for model validation.*
- *a separate test set with 2M jets in each class (in total 20 M) for performance evaluation*

Model complexity

	Accuracy	# params	FLOPs
PFN	0.772	86.1 k	4.62 M
P-CNN	0.809	354 k	15.5 M
ParticleNet	0.844	370 k	540 M
ParT	**0.861**	2.14 M	340 M
ParT (plain)	0.849	2.13 M	260 M

ParT: Performance benchmarks

Comparison with previous taggers

	All classes Accuracy	All classes AUC	$H \to b\bar{b}$ Rej$_{50\%}$	$H \to c\bar{c}$ Rej$_{50\%}$	$H \to gg$ Rej$_{50\%}$	$H \to 4q$ Rej$_{50\%}$	$H \to \ell\nu qq'$ Rej$_{99\%}$	$t \to bqq'$ Rej$_{50\%}$	$t \to b\ell\nu$ Rej$_{99.5\%}$	$W \to qq'$ Rej$_{50\%}$	$Z \to q\bar{q}$ Rej$_{50\%}$
PFN	0.772	0.9714	2924	841	75	198	265	797	721	189	159
P-CNN	0.809	0.9789	4890	1276	88	474	947	2907	2304	241	204
ParticleNet	0.844	0.9849	7634	2475	104	954	3339	10526	11173	347	283
ParT	**0.861**	**0.9877**	**10638**	**4149**	**123**	**1864**	**5479**	**32787**	**15873**	**543**	**402**
ParT (plain)	0.849	0.9859	9569	2911	112	1185	3868	17699	12987	384	311

Increasing training dataset size

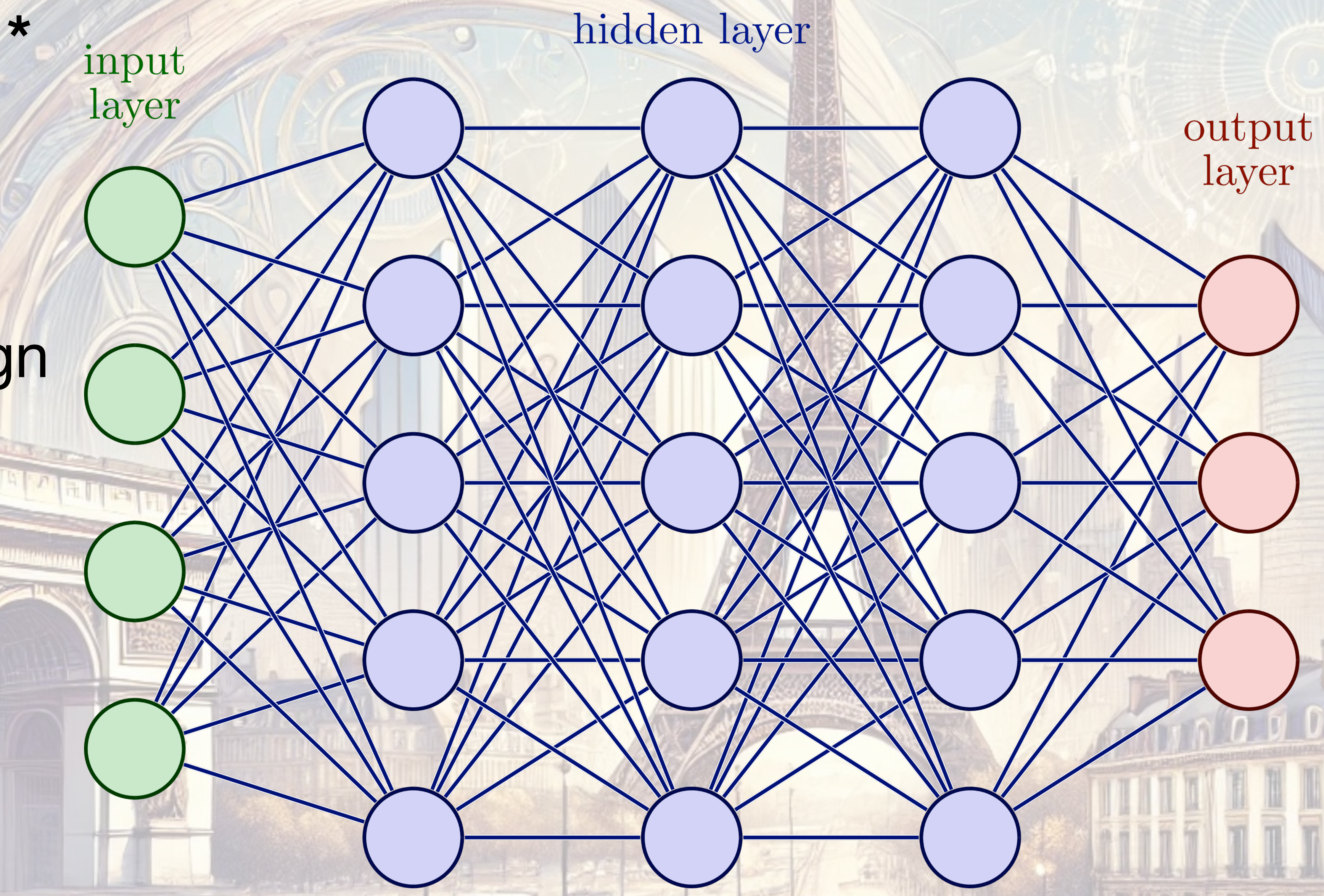
	All classes Accuracy	All classes AUC	$H \to b\bar{b}$ Rej$_{50\%}$	$H \to c\bar{c}$ Rej$_{50\%}$	$H \to gg$ Rej$_{50\%}$	$H \to 4q$ Rej$_{50\%}$	$H \to \ell\nu qq'$ Rej$_{99\%}$	$t \to bqq'$ Rej$_{50\%}$	$t \to b\ell\nu$ Rej$_{99.5\%}$	$W \to qq'$ Rej$_{50\%}$	$Z \to q\bar{q}$ Rej$_{50\%}$
ParticleNet (2 M)	0.828	0.9820	5540	1681	90	662	1654	4049	4673	260	215
ParticleNet (10 M)	0.837	0.9837	5848	2070	96	770	2350	5495	6803	307	253
ParticleNet (100 M)	0.844	0.9849	7634	2475	104	954	3339	10526	11173	347	283
ParT (2 M)	0.836	0.9834	5587	1982	93	761	1609	6061	4474	307	236
ParT (10 M)	0.850	0.9860	8734	3040	110	1274	3257	12579	8969	431	324
ParT (100 M)	0.861	0.9877	10638	4149	123	1864	5479	32787	15873	543	402

UParT: Technicals

- Dataset size: 30M training / 5M validation
- 30 epochs with batch_size of 2048
- Complexity is the one of a 6 layers ParT, ~40M FLOPS in the forward path
- 6 PMHA + 2 CLS blocks

UParT: In a nutshell

The Fast Gradient Sign Method (FGSM) on input features



Addition of flavor classes

Addition of flavor aware Jet energy regression and resolution

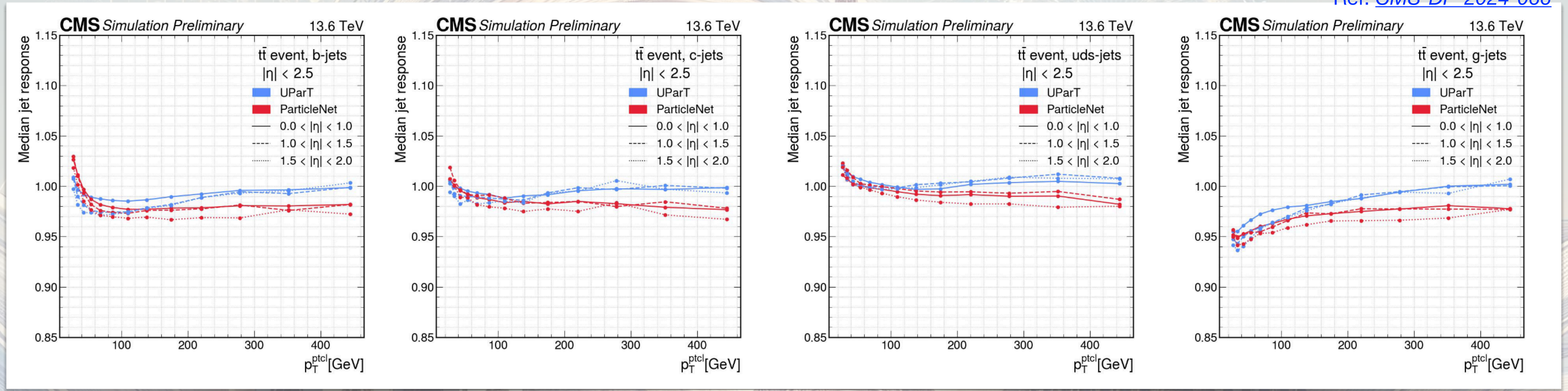
P-MHA block from ParT

*illustrative

Jet Energy Regression

- **Flavor aware jet energy regression**

Ref: [CMS-DP-2024-066]



*ParticleNet being trained with Run 2 MC samples

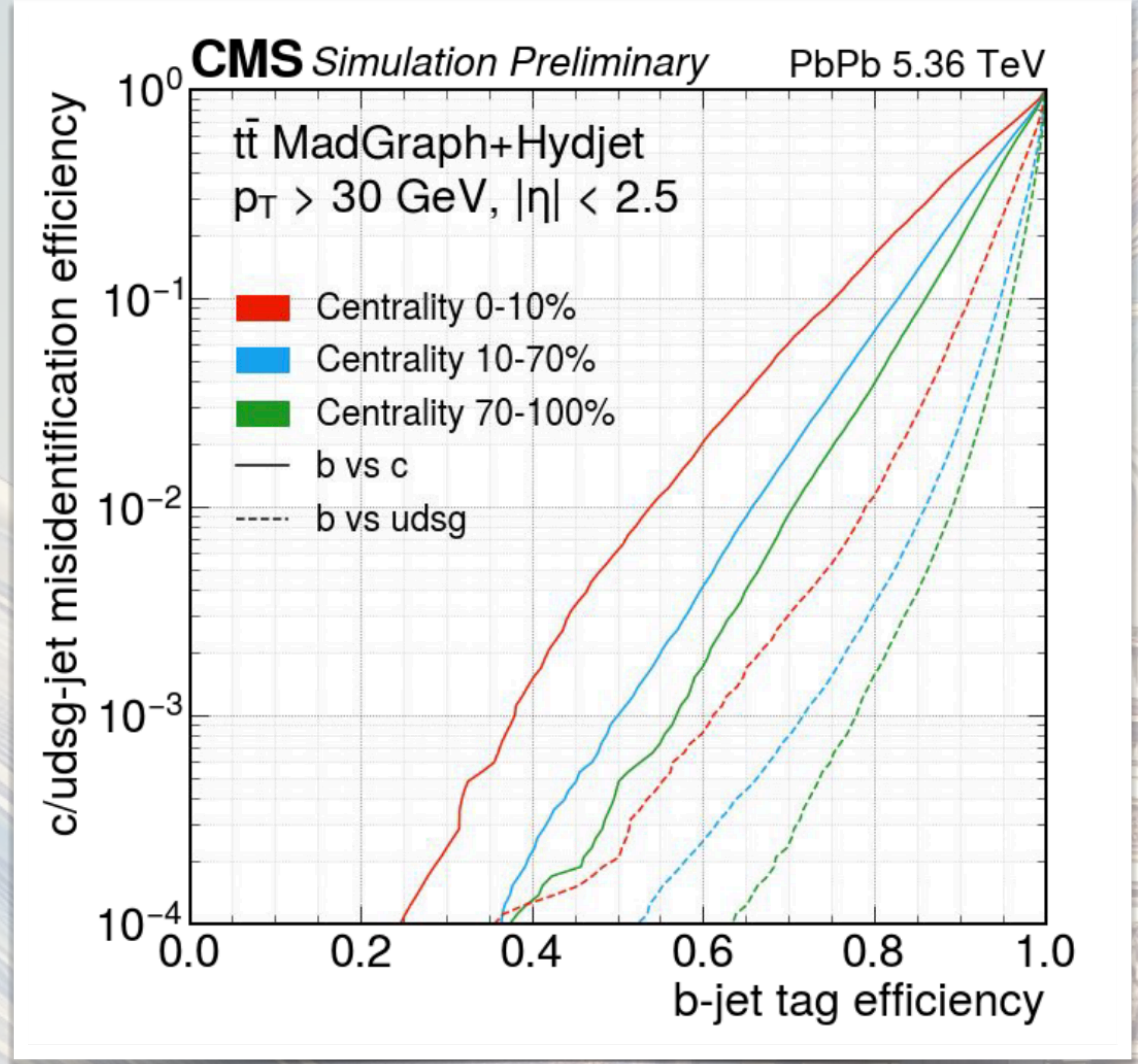
Jet Energy Resolution

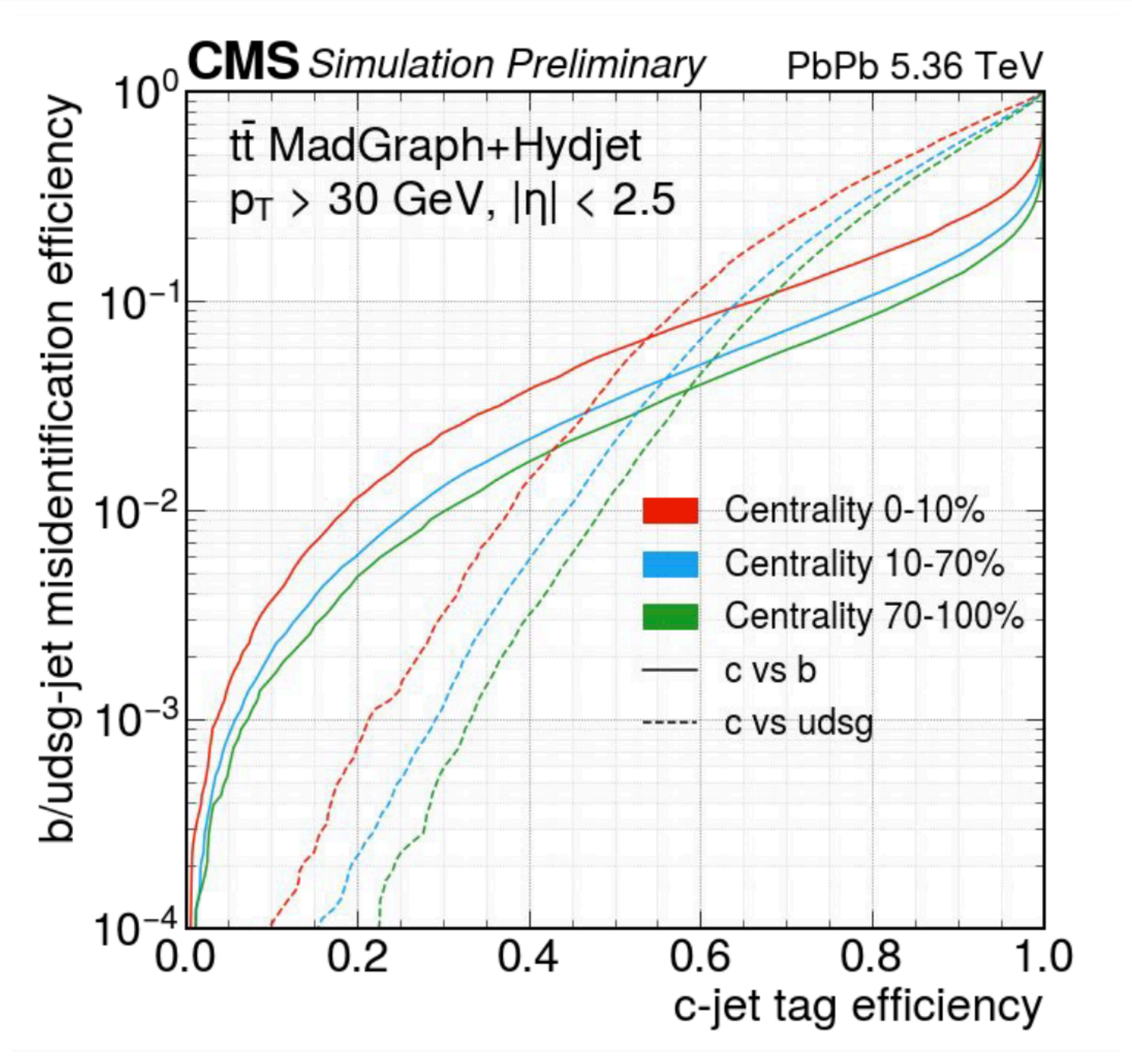
Ref: CMS-DP-2024-066

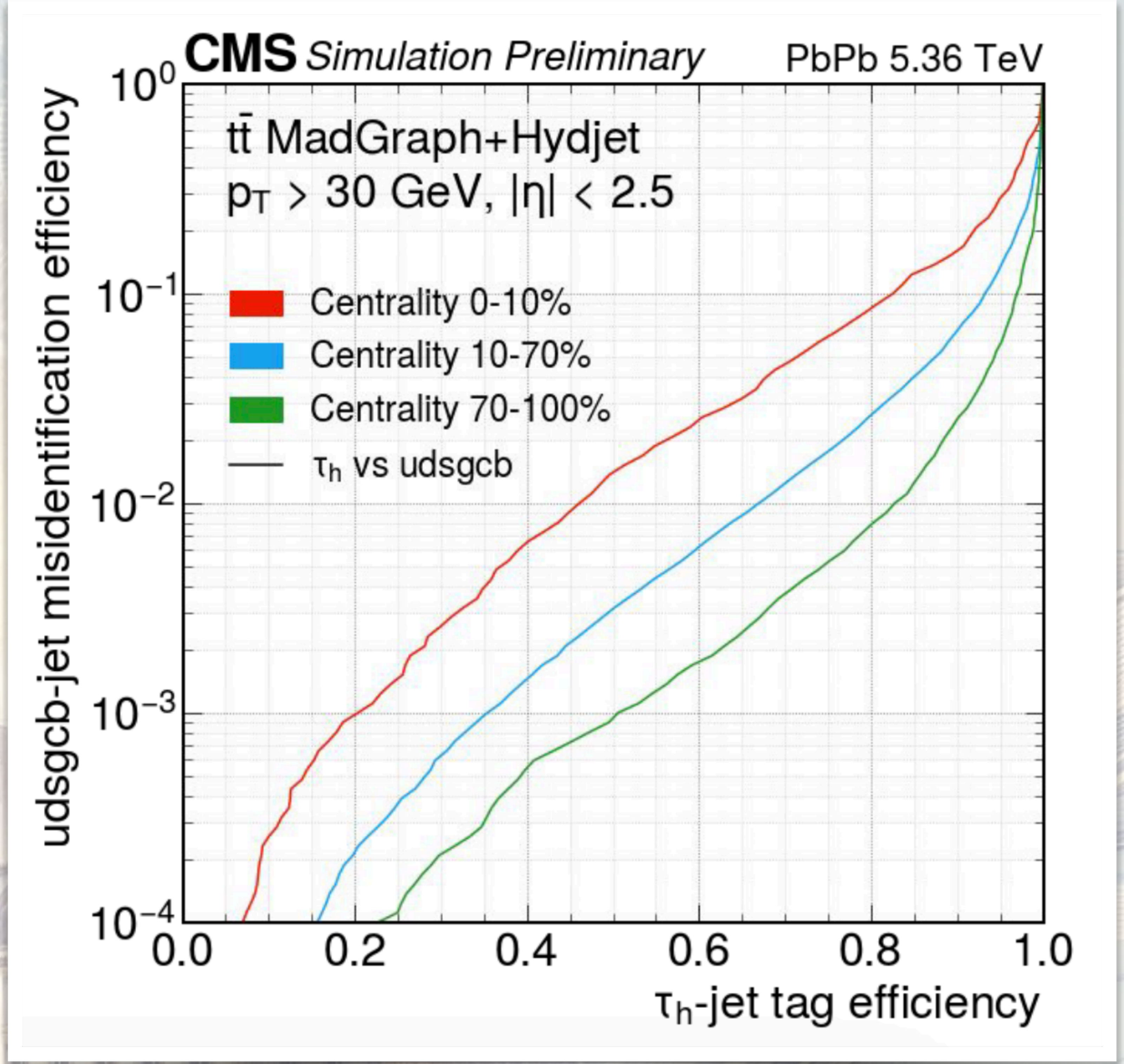
- UParT and ParticleNet show similar jet resolution response
- Capable of estimating a per-jet resolution from the quantile regression

UParT performance: PbPb collision

Ref: [BTV-24-088]







In different centrality - improvement in performance in the more central collisions