

Advancing Jet Physics in Heavy Ion Collisions with Generative AI

Yeonju Go

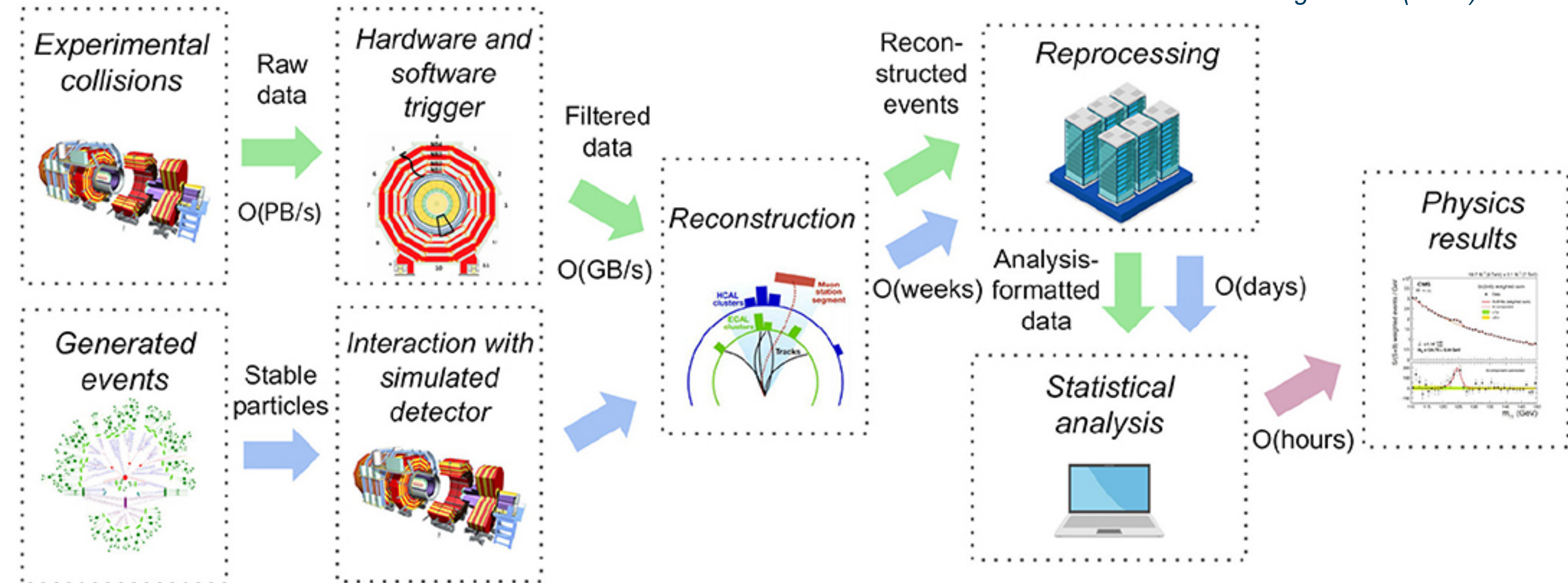
Goldhaber Distinguished Fellow
Brookhaven National Laboratory (BNL)

*BNL Physics Seminar
Nov. 18, 2025*



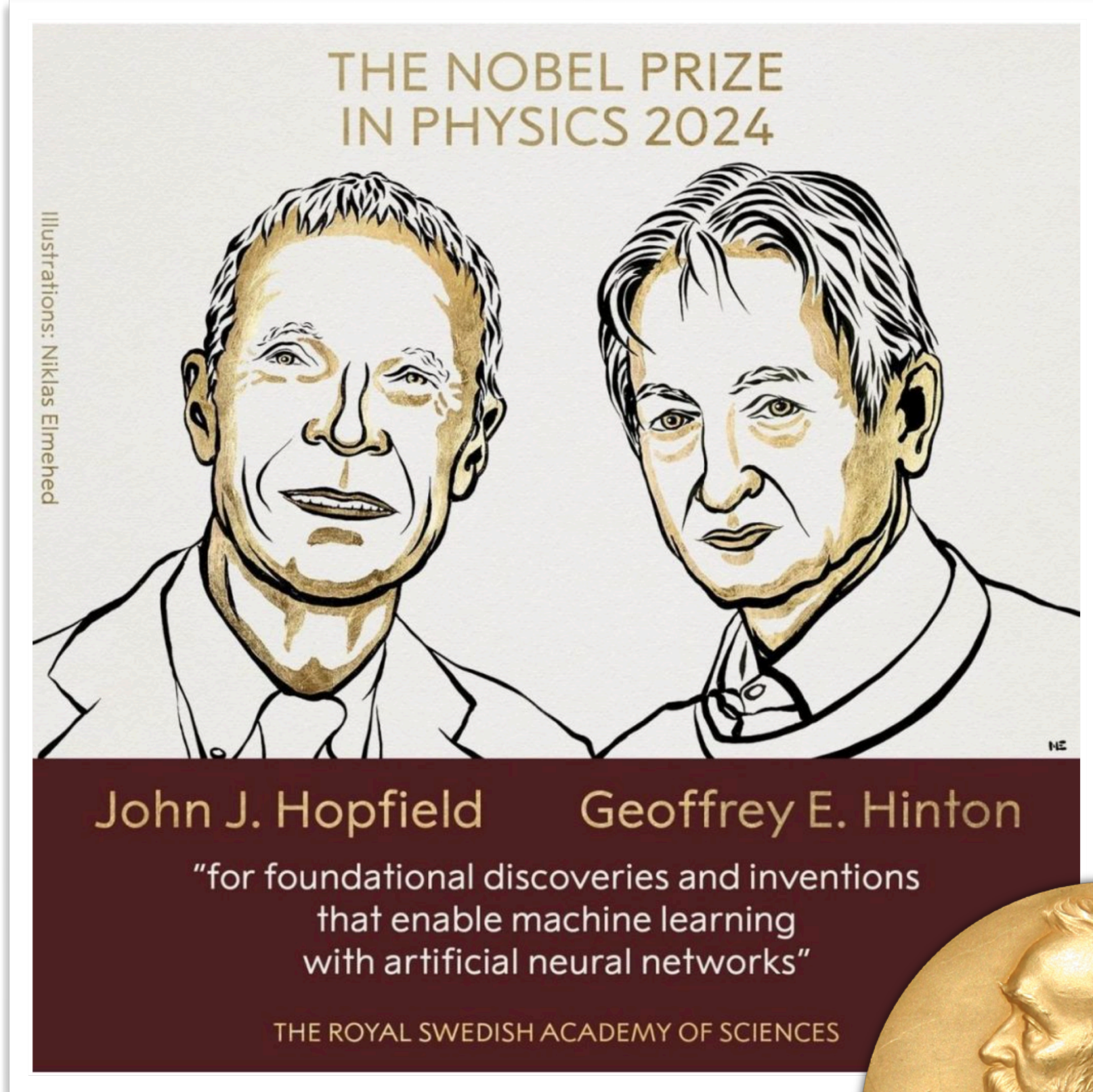
(High Energy) Nuclear Physics Analysis Workflows

Front.Big Data 4 (2021) 661501



- Each stage of analysis requires substantial human effort, computing resources, and time
→ **Leverage AI/ML to accelerate, automate, and improve these processes**

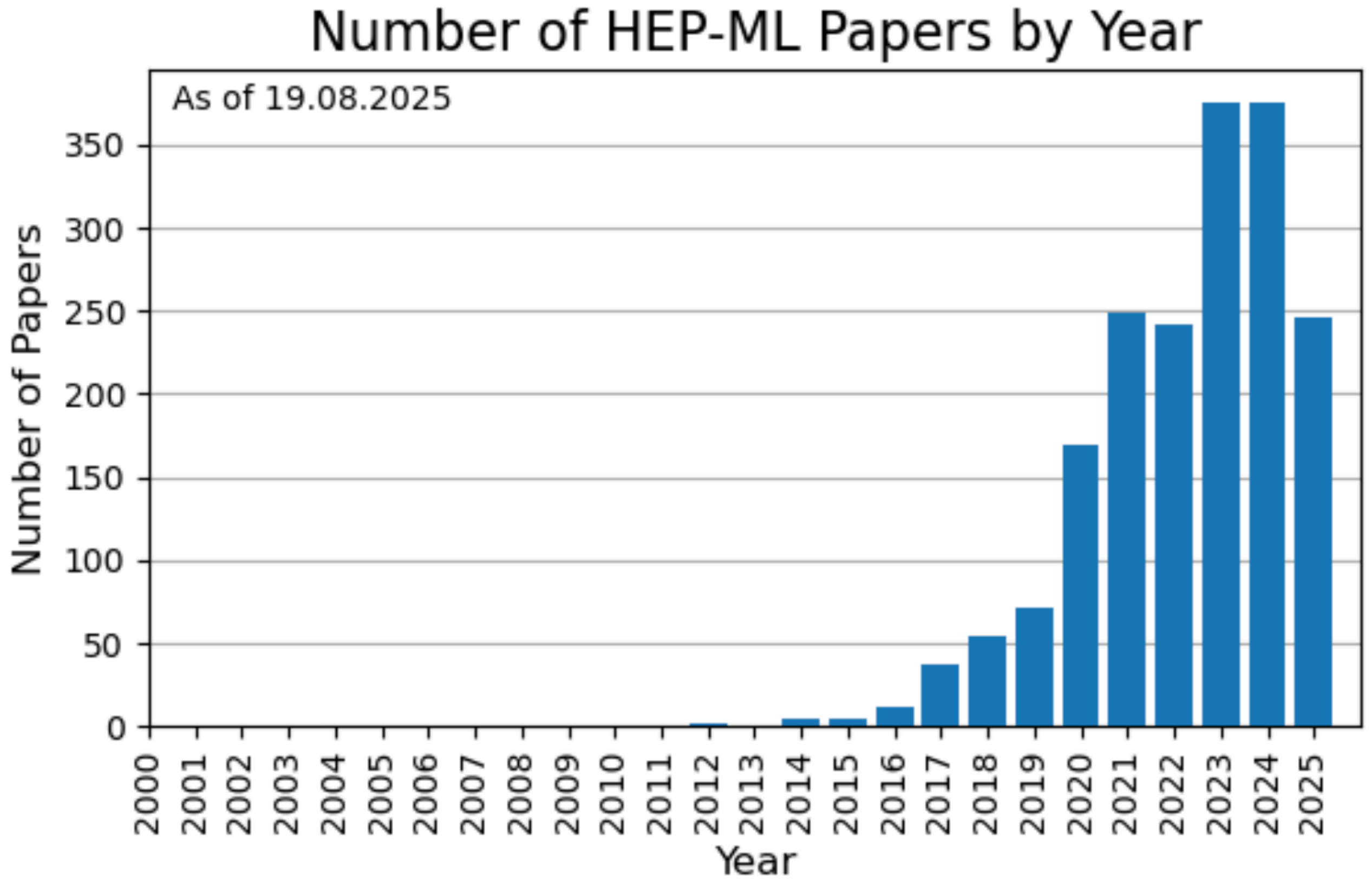
AI/ML in Physics



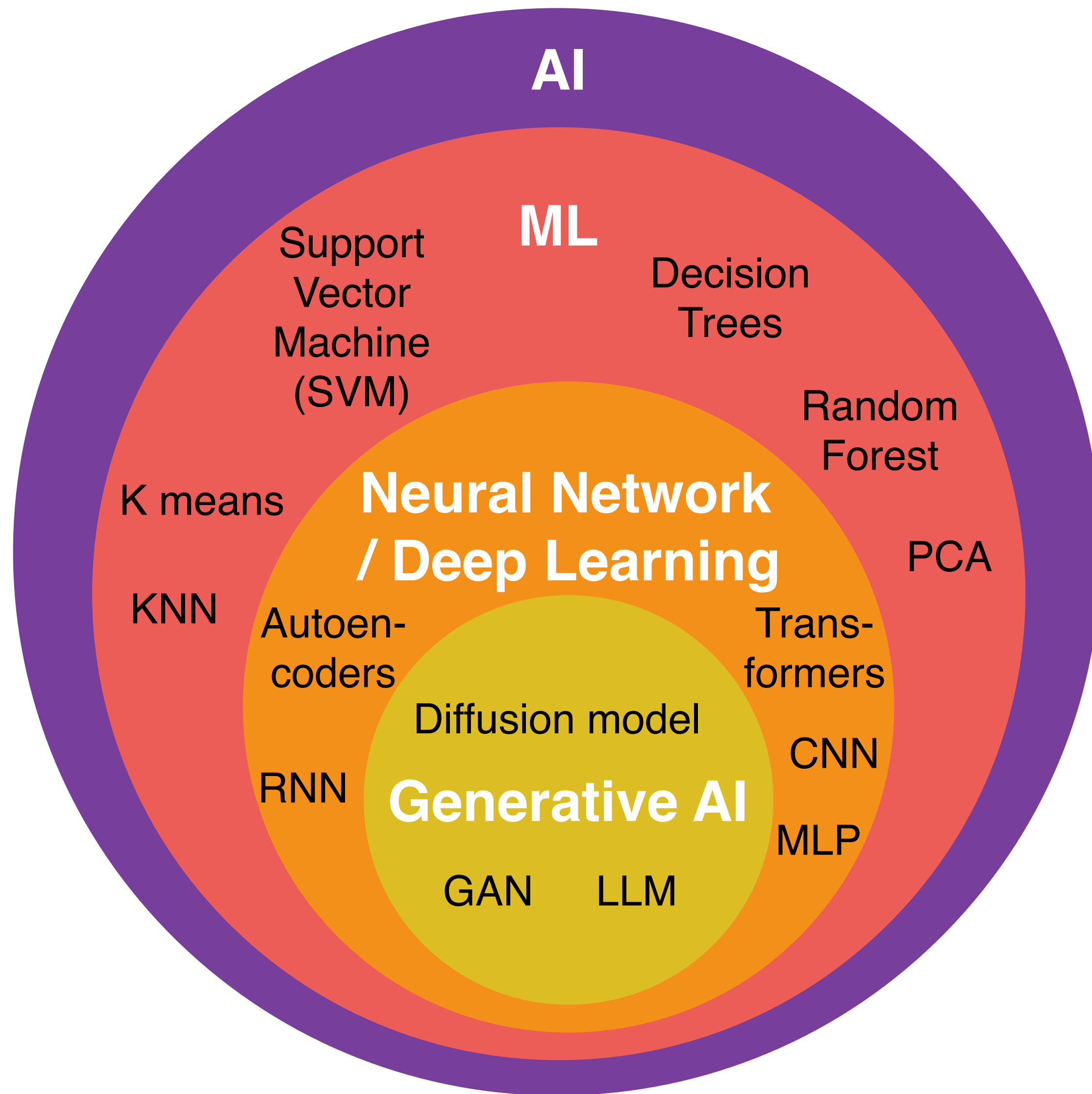
"for foundational discoveries and inventions that enable machine learning with artificial neural networks"



2024
Nobel Prize in Physics



Artificial Intelligence and Machine Learning



Artificial Intelligence (AI)

- **mimic** cognitive functions associated with **human intelligence**
e.g. see, understand, respond, analyze data, make recommendations, etc

Machine Learning (ML)

- **extract knowledge from data** and learn from it **autonomously**
- **no explicit programming**

Neural Network / Deep Learning

- inspired by human brain that uses “**neurons**”
- learn based on deep neural network

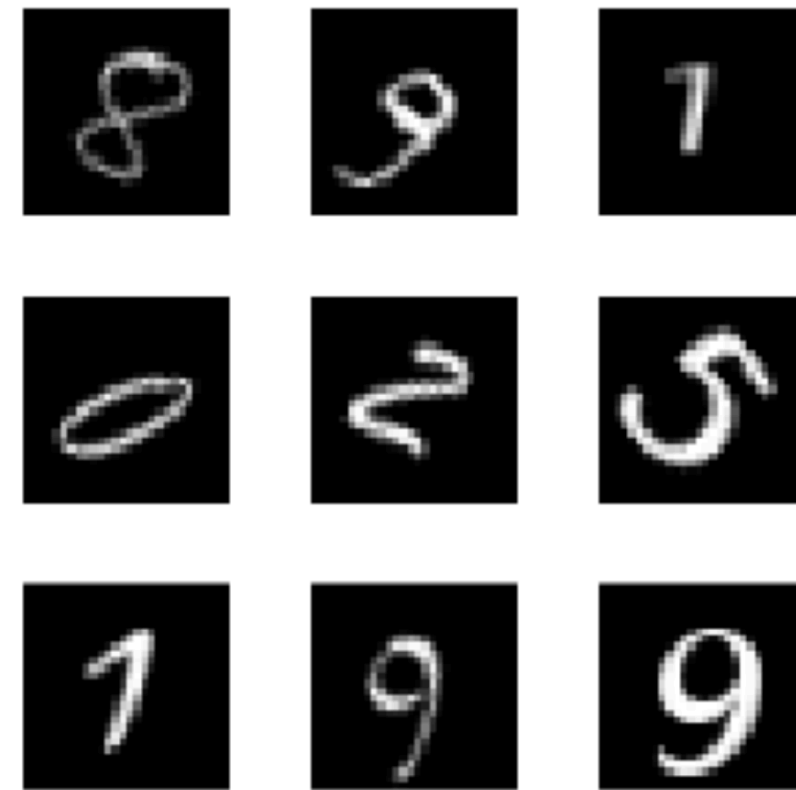
Generative AI

- **creates new content**, such as text, images, and audio, by learning patterns from existing data

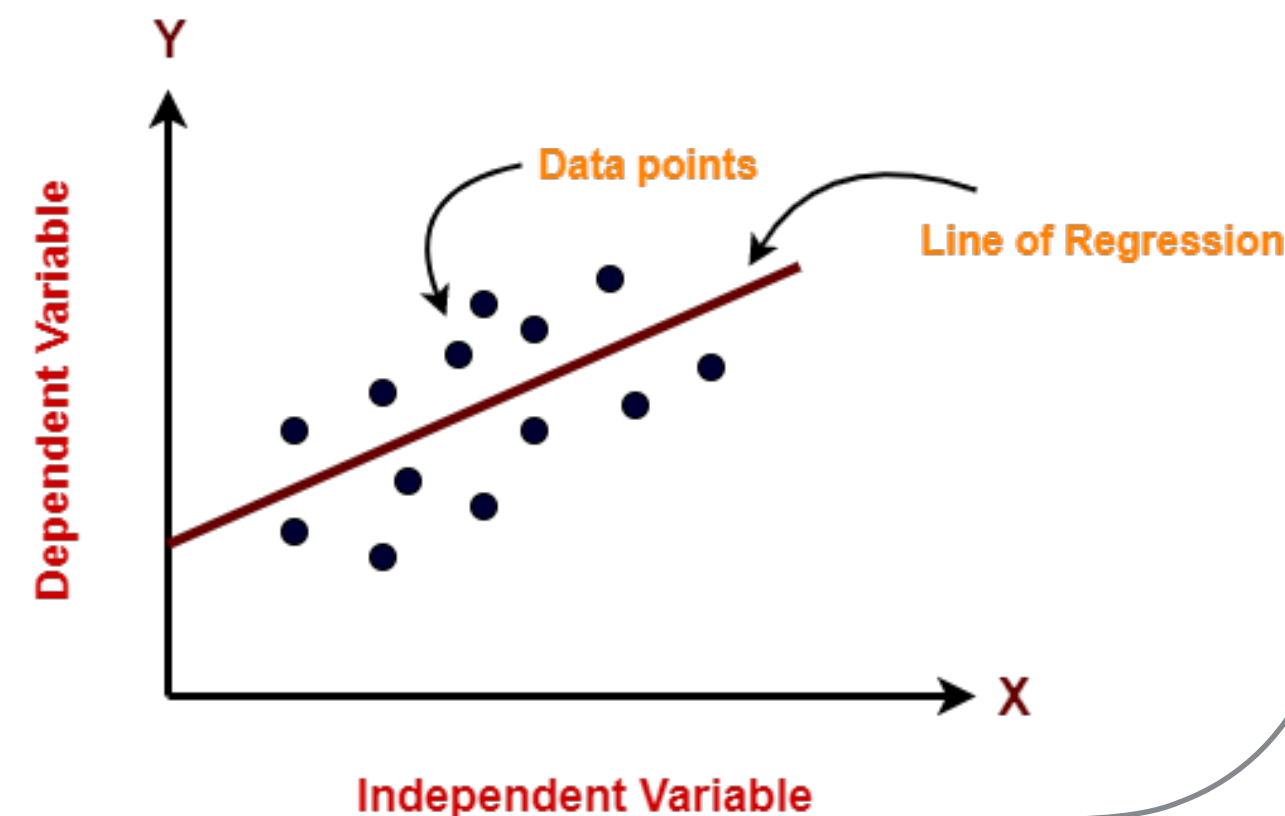
Types of Machine Learning

Supervised

- Use **labelled** data
- **Input-output values** are given by human
- Forecast outcomes
- Classification

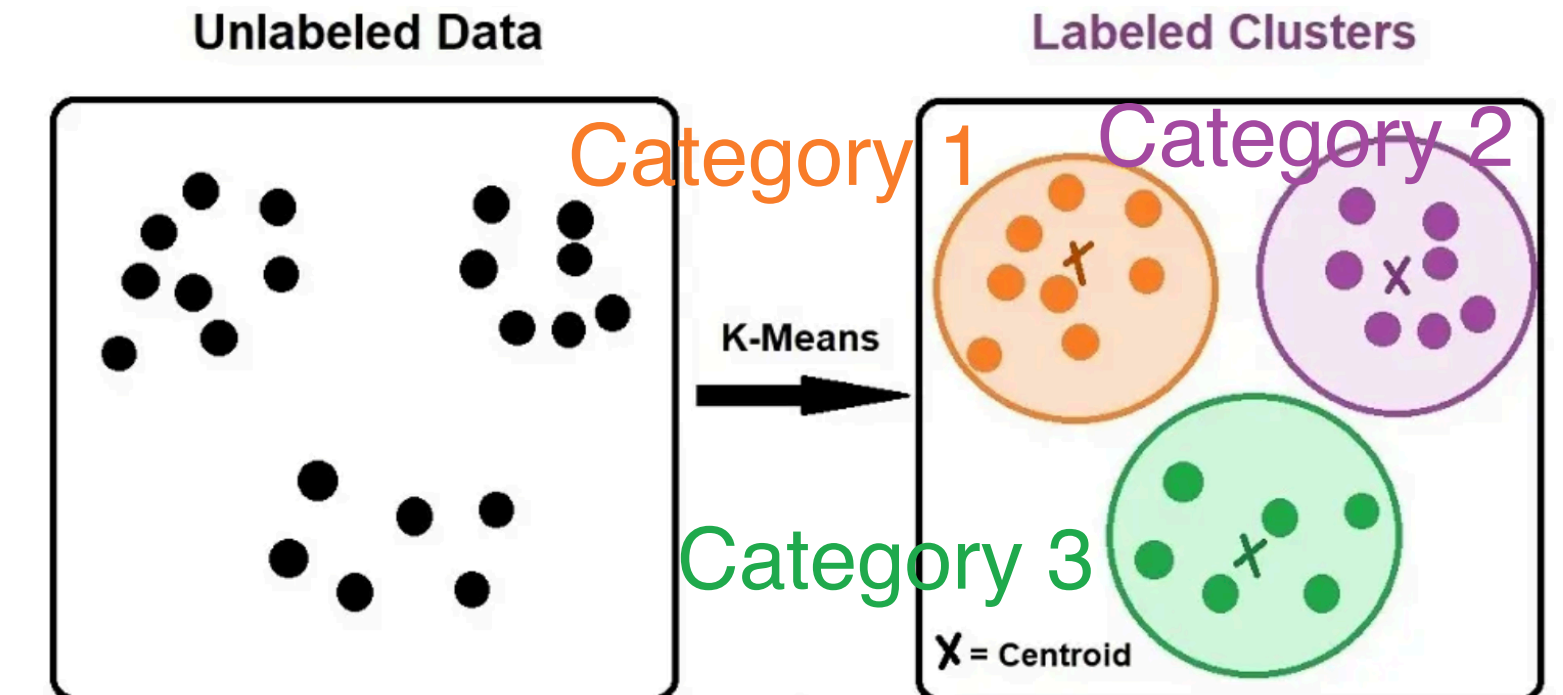


- Regression

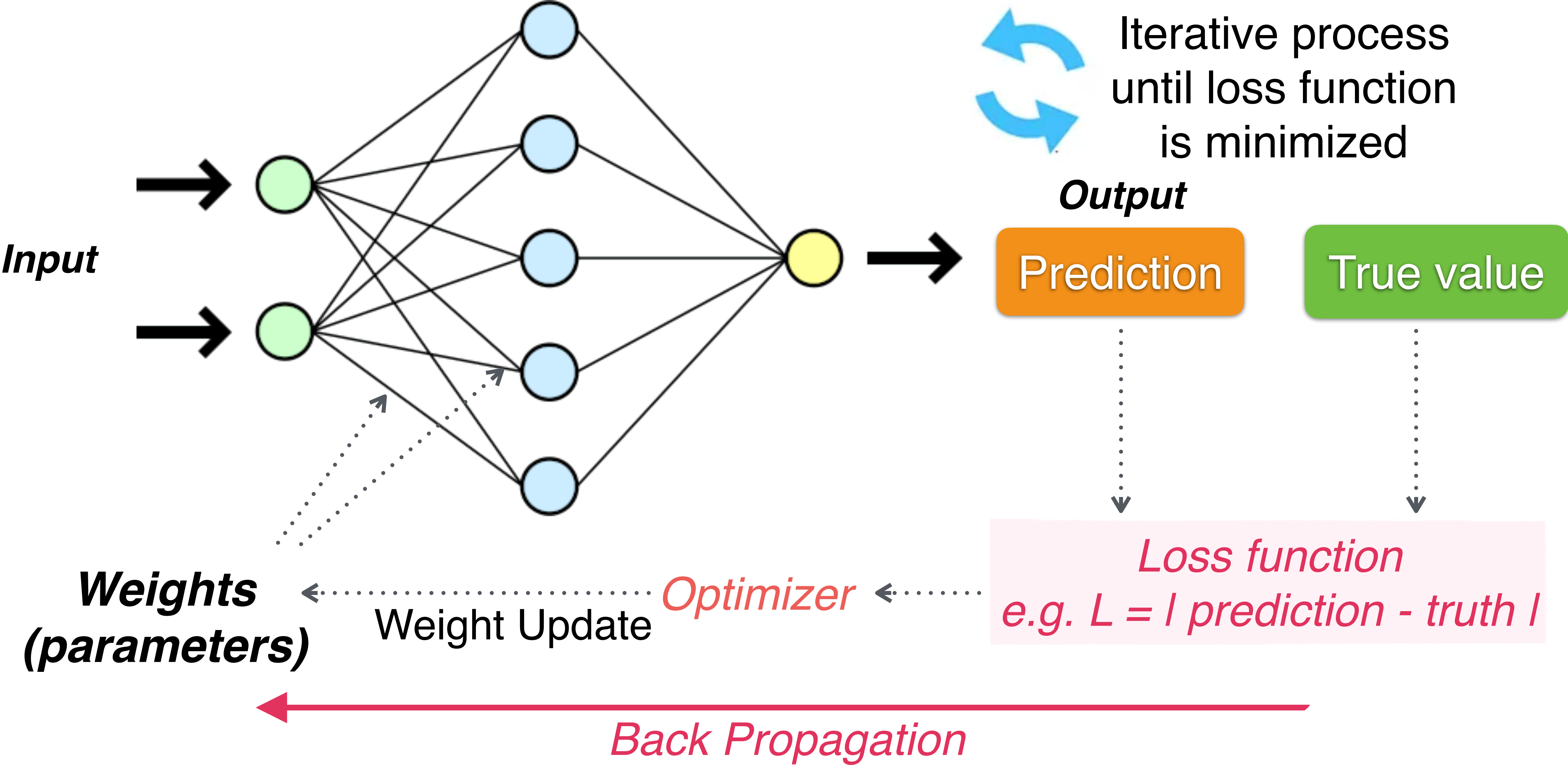


Unsupervised

- Use **unlabelled** data
- Find **hidden patterns** w/o human intervention
- Clustering
- Association
- Dimension Reduction
- Anomaly detection



Neural Network



Many layers: Multi Layer Perceptron (MLP), Deep Neural Network (DNN)

Various Neural Network Architecture

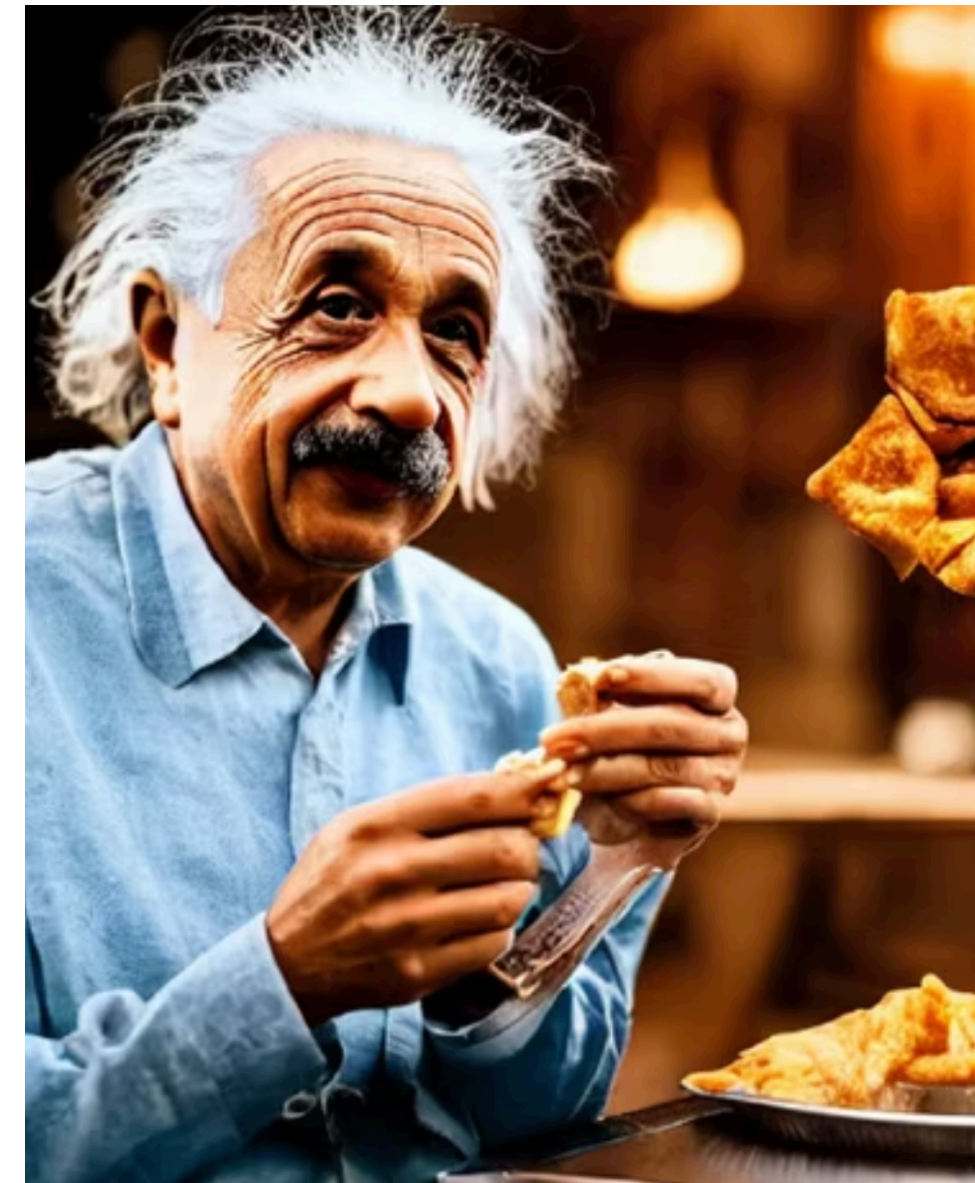
The different structure (architectures) based on how they organize, connect, and process inputs before or between MLPs

- **Convolutional Neural Network (CNN)** : good for image processing
- **Graph Neural Network (GNN)** : good for irregular, relational data
- **Recurrent Neural Network (RNN)** : good for sequential data
- **Transformer** : learns “attention”
- **Autoencoder** : encode $\rightarrow$ latent space $\rightarrow$ decode
- ...

Generative AI

- *Generates syntactic data e.g. text, image, audio, video, ...*
- *Generally more advanced and complex AI*
 - ➔ combines representation learning (understanding) with generation (creation)
- *Models*
 - ➔ Variational Autoencoder (VAE)
 - ➔ Generative Adversarial Network (GAN)
 - ➔ Diffusion models
 - ➔ Transformer-based Models, e.g. Large Language Model
 - ➔ ...

Images from Dalle-2 (Diffusion Model)



What Controls the Performance of AI?

Defining the physics problem and ML task

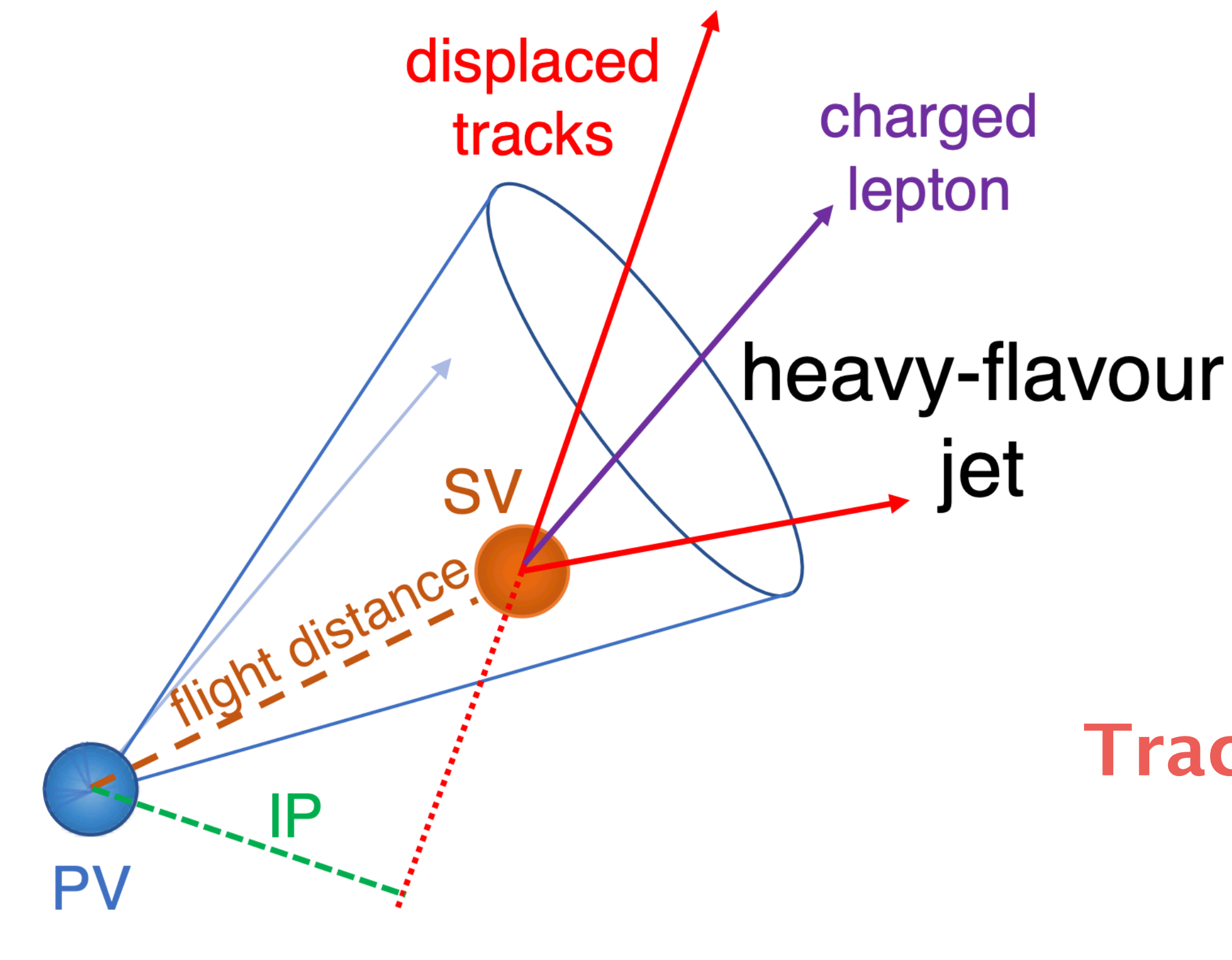
Choosing the model architecture

Curating and preprocessing the data

Defining loss, optimizer, and evaluation metrics

*What kind of **machine learning** developments/applications
are there for **jet physics in heavy ion collisions**?*

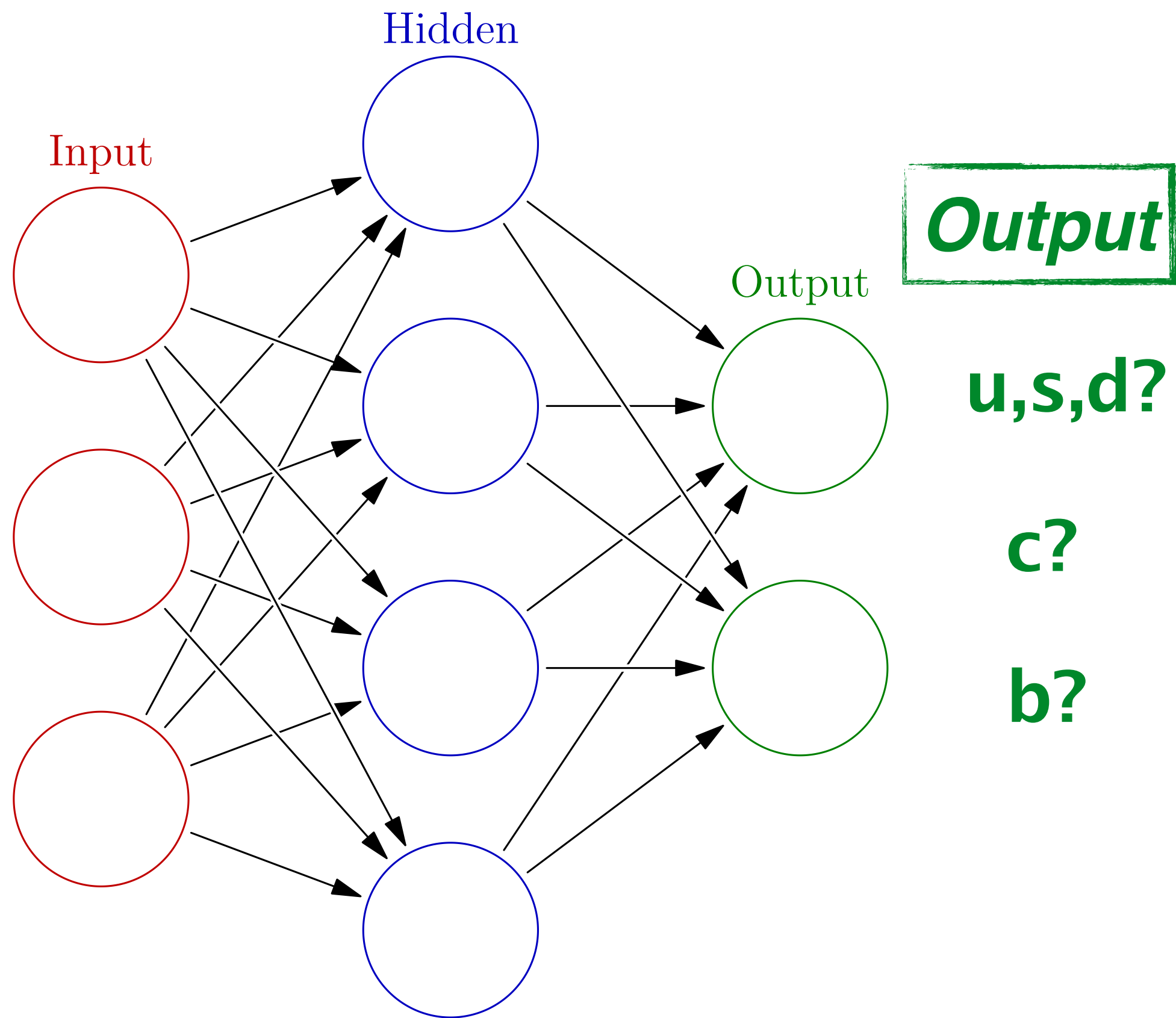
Classification: Heavy Flavor Jet Tagging



[arXiv:1712.07158](https://arxiv.org/abs/1712.07158)

Input

Track kinematics,
IP,
SV,
Neutral particle info.,
...



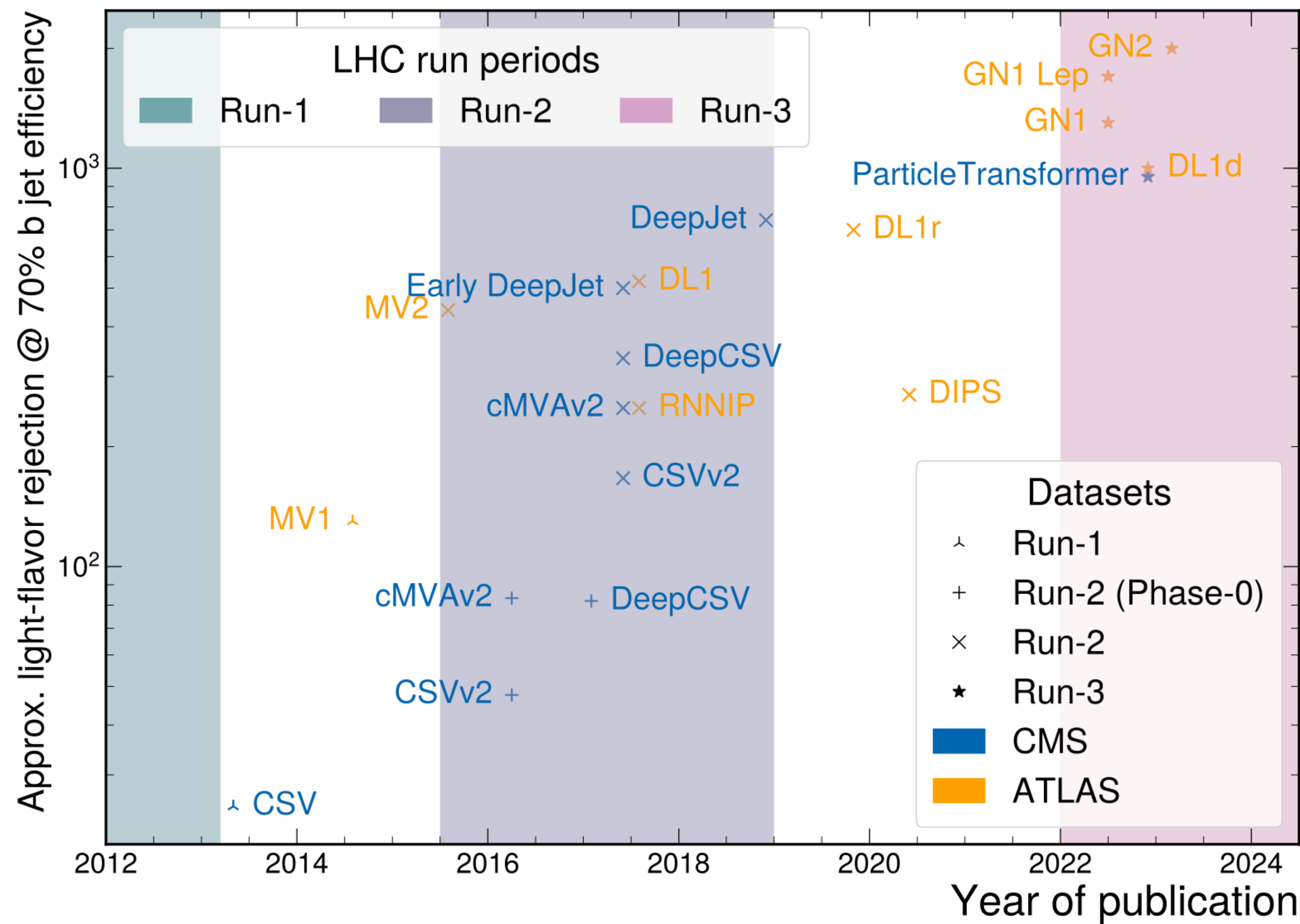
Output

u,s,d?
c?
b?

Classification: Heavy Flavor Jet Tagging

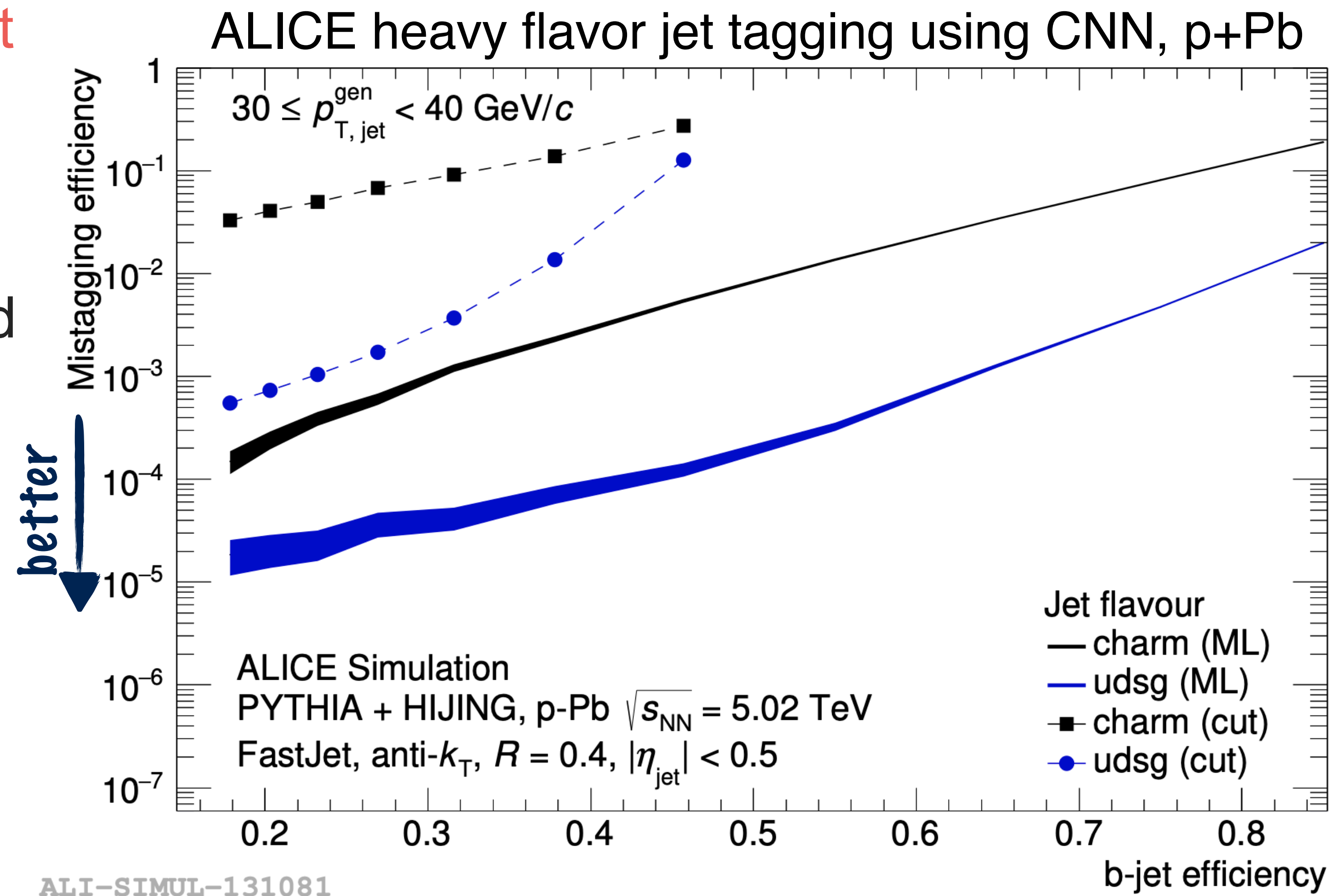
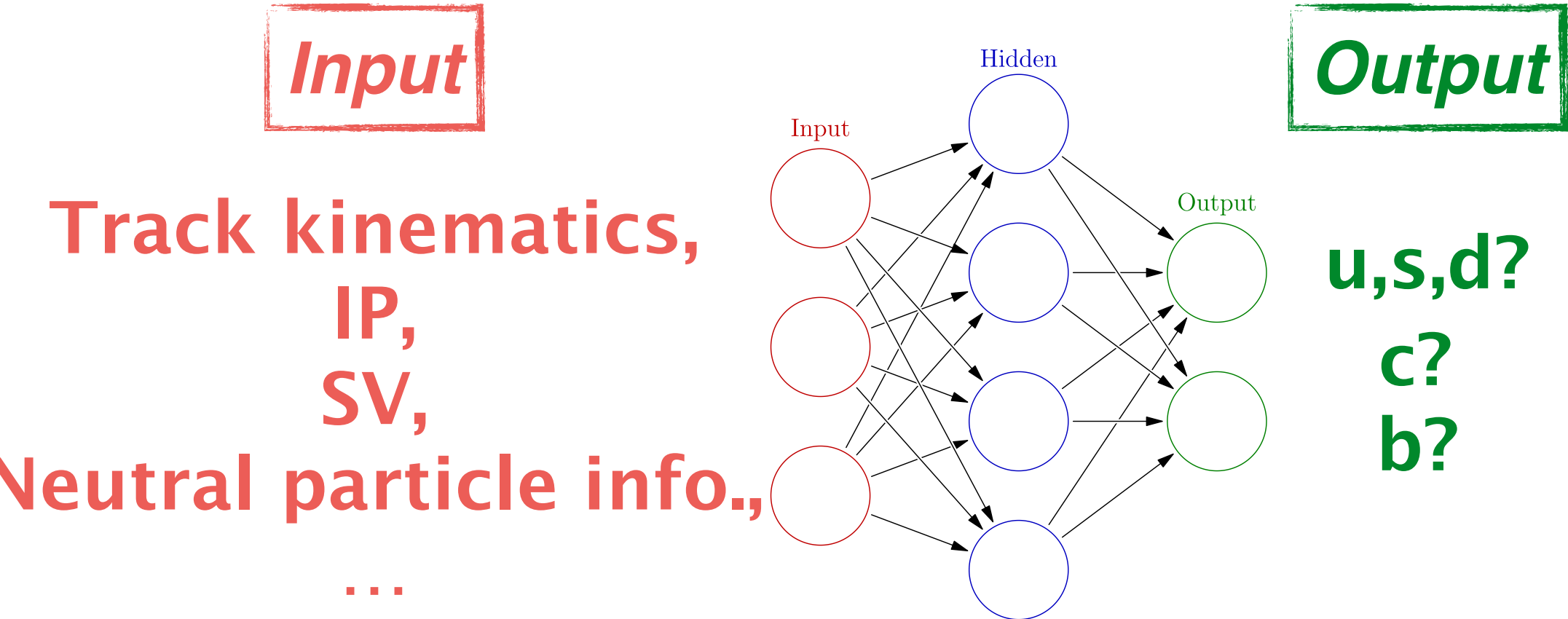
Eur.Phys.J.ST 233 (2024) 15-16, 2657-2686

- Significant evolution on **flavor tagging** using ML over the decade
 - ➔ Multi Layer Perceptron
 - ➔ Convolutional Neural Networks
 - ➔ Recurrent Neural Networks
 - Long Short-Term Memory network
 - ➔ Transformer
 - ➔ Deep Sets
 - ➔ Graph Neural Network
 - ➔ ...



Classification: Heavy Flavor Jet Tagging in HI

- Much better performance than conventional method !
- However, in *heavy ion collisions*, the jet structure and its constituents are modified
- Most existing studies use *unquenched jet* samples (e.g. PYTHIA+HIJING) without explicit in medium modification of heavy flavor jets. [CMS-DP-2025-088](#)
- **Careful systematic studies** are required to understand *how quenching and the heavy ion environment affect tagging performance and mistag rates*



[PoS\(EPS-HEP2017\)498](#)

Classification: Quenched vs Unquenched Jets

- Classify **quenched-like** vs **vacuum-like** jets
- Supervised Learning
 - ➔ Sequential Attention-based NN → iteratively pass weights
 - ➔ Trained with multiple theoretical models (**JEWEL**+**CoLBT** and **PYTHIA**+**HERWIG**) to reduce model-dependence

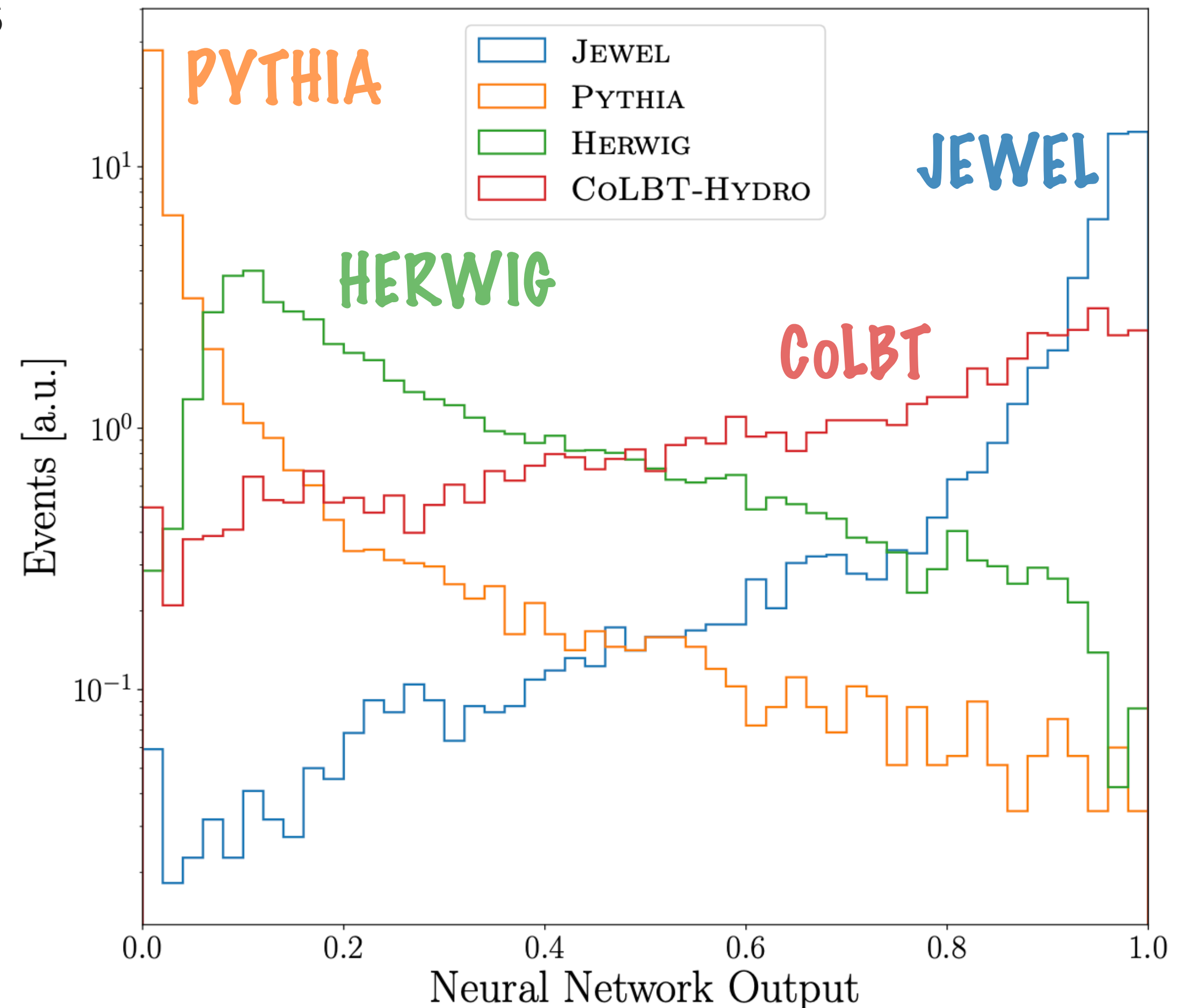
Trained with jet substructures as input features,
and with output of

JEWEL **CoLBT** = 1

PYTHIA **HERWIG** = 0

Therefore, output close to 1 → quenched-like jet
output close to 0 → unquenched-like jet

[arXiv:2411.19389](https://arxiv.org/abs/2411.19389)



Classification: Quenched vs Unquenched Jets

- Many similar works
 - ➔ *Liliana Apolinário et al. JHEP 11 (2021) 219*
 - ➔ *Lihan Liu et al. JHEP 04 (2023) 140*
 - ➔ *Yue Shi Lai et al. JHEP 10 (2022) 011*
 - ➔ *Miguel Romão et al. SciPost Phys. 16 (2024) 015*
 - ➔ *Yilun Wu et al. arXiv:2511.04005*
- **Usecase**
 - ➔ test how much **quenching information** is present in jet substructure
 - ➔ identify which representations and observables are most sensitive
- *Careful studies of model dependence* are needed before applying quenching scores to data

Regression for Jet Energy Loss

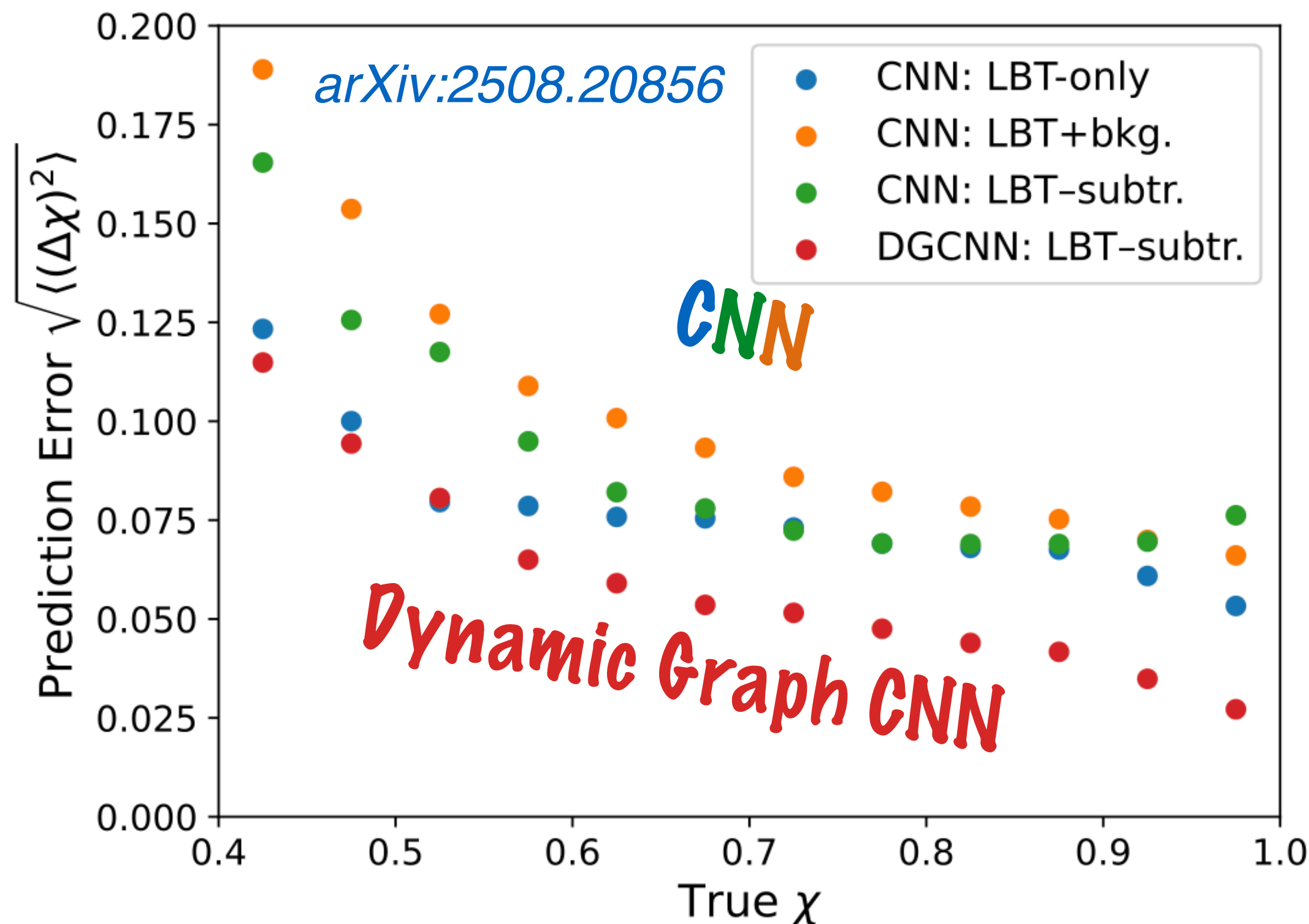
- Predict **amount of jet energy loss** on a jet-by-jet basis

$$\chi = p_T^{\text{quenched}} / p_T^{\text{before quenching}}$$

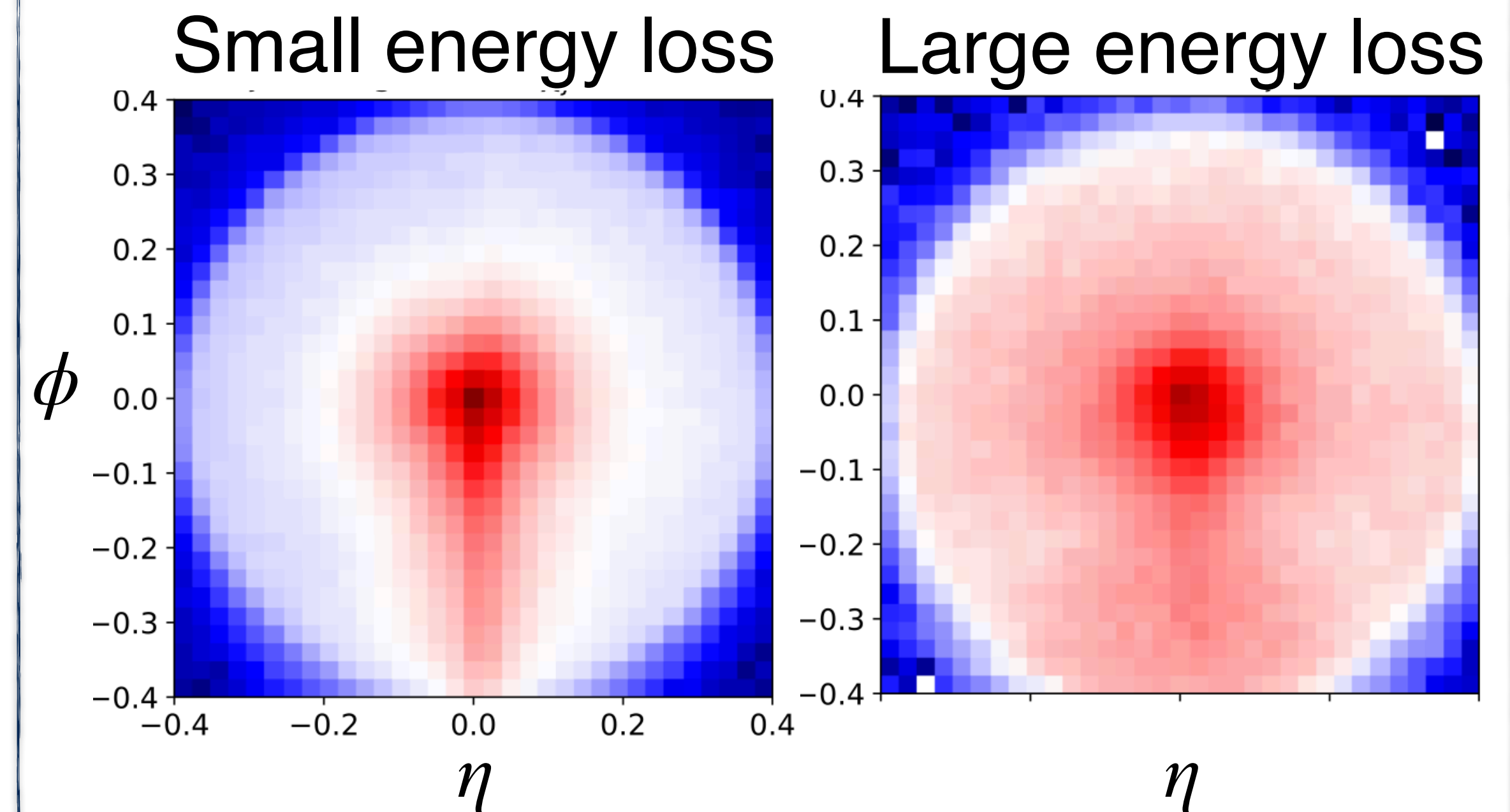
- **Supervised Learning**

- ➔ Convolutional Neural Network (CNN) → good for image processing

- ➔ Trained with *Hybrid* [JHEP 03 (2021) 206] or *LBT* [arXiv:2508.20856] theoretical models



Jet Image (on a jet-by-jet basis)



JHEP 03 (2021) 206

Regression for Jet Energy Loss

- Predict **amount of jet energy loss** on a jet-by-jet basis

$$\chi = p_{\text{T}}^{\text{quenched}} / p_{\text{T}}^{\text{before quenching}}$$

- **Supervised Learning**

- ➔ Convolutional Neural Network (CNN) → good for image processing

- ➔ Trained with *Hybrid* [JHEP 03 (2021) 206]
or *LBT* [arXiv:2508.20856] theoretical models

- **Caution to apply to real experimental data**

- ➔ simple thermal background, **no full hadronization, no detector simulation**
→ need careful validation under realistic conditions

- ➔ Trained only on PYTHIA+LBT **model-dependent bias**
→ robustness across **different quenching and medium models** should be studied carefully

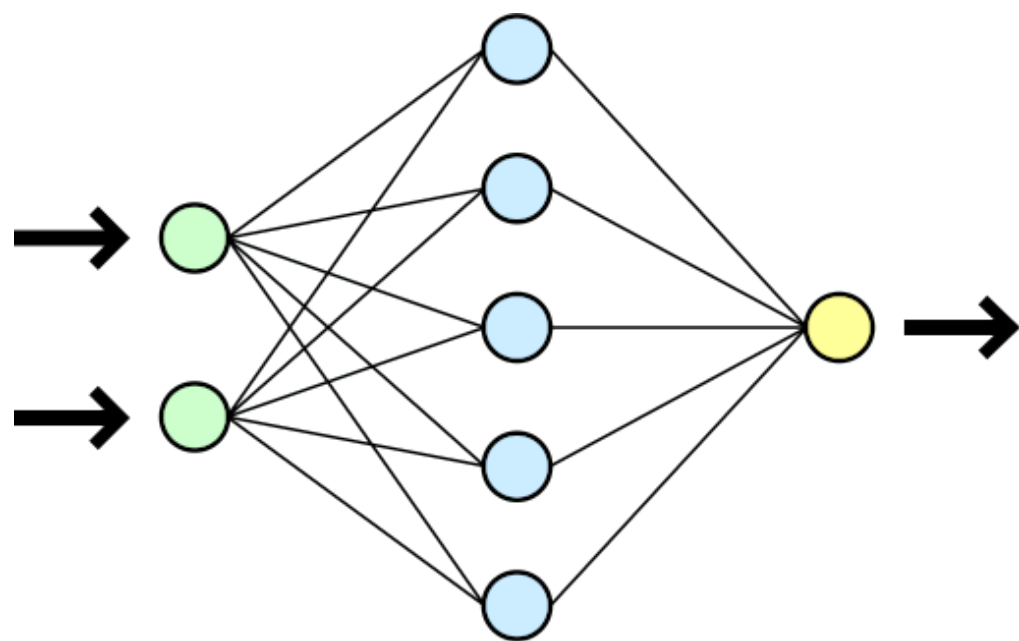
Regression for Jet Momentum

- Given the *jet constituents* and *substructure* information, what is the *background-subtracted truth-level jet momentum*?

Related works:
[PRC 108 \(2023\) 6, 021901](#)

- Supervised Learning MLP**

- Input features: jet p_T after area-based correction, 10 highest constituent's p_T , jet mass etc

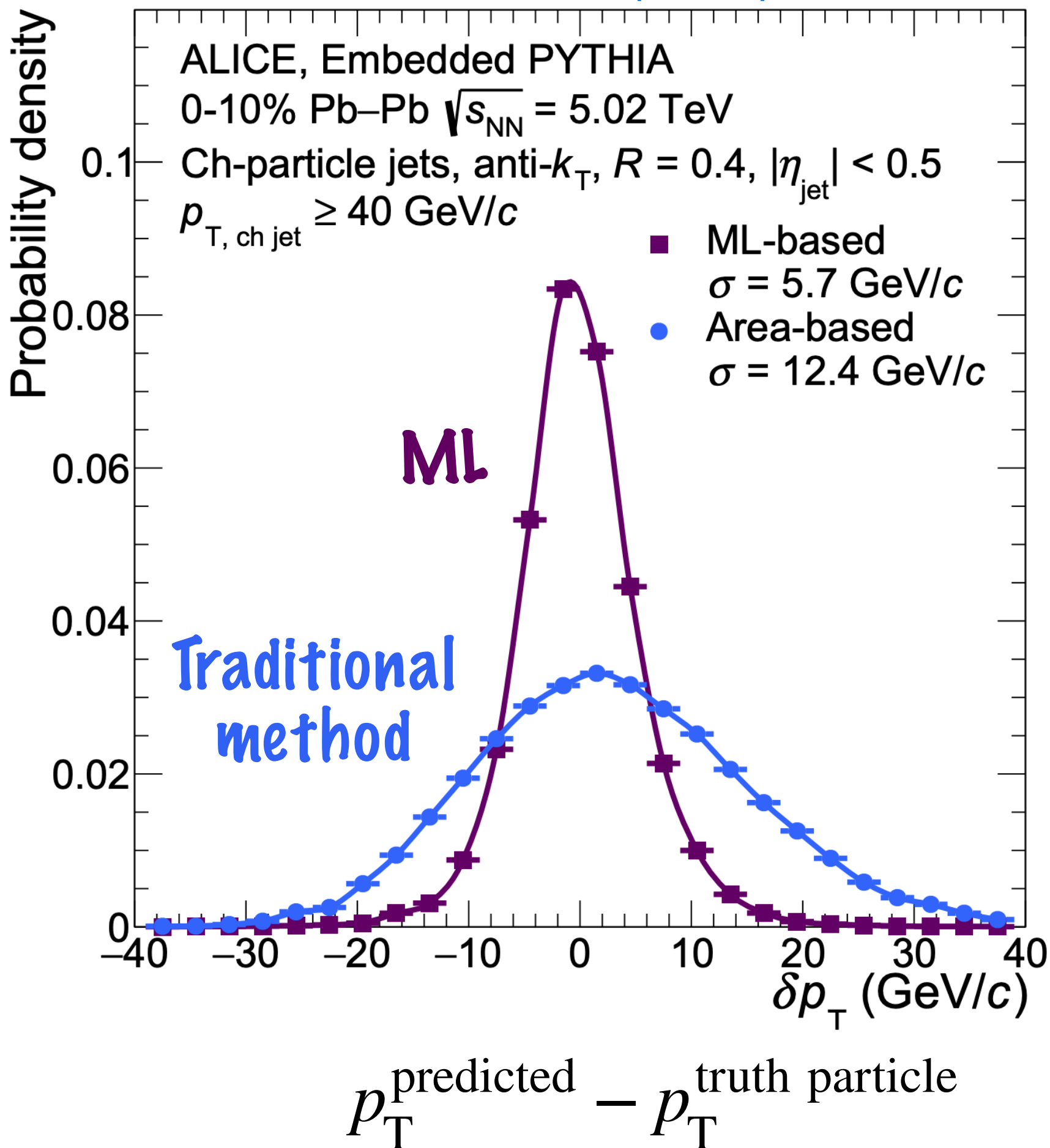


- 3 layers with 100 and 50 nodes

[PRC 99 \(2019\) 6, 064904](#)

Feature	Score	Feature	Score
Jet p_T (no corr.)	0.1355	$p_{T, \text{const}}^1$	0.0012
Jet mass	0.0007	$p_{T, \text{const}}^2$	0.0039
Jet area	0.0005	$p_{T, \text{const}}^3$	0.0015
Jet p_T (area-based corr.)	0.7876	$p_{T, \text{const}}^4$	0.0011
LeSub	0.0004	$p_{T, \text{const}}^5$	0.0009
Radial moment	0.0005	$p_{T, \text{const}}^6$	0.0009
Momentum dispersion	0.0007	$p_{T, \text{const}}^7$	0.0008
Number of constituents	0.0008	$p_{T, \text{const}}^8$	0.0007
Mean of const. p_T	0.0585	$p_{T, \text{const}}^9$	0.0006
Median of const. p_T	0.0023	$p_{T, \text{const}}^{10}$	0.0007

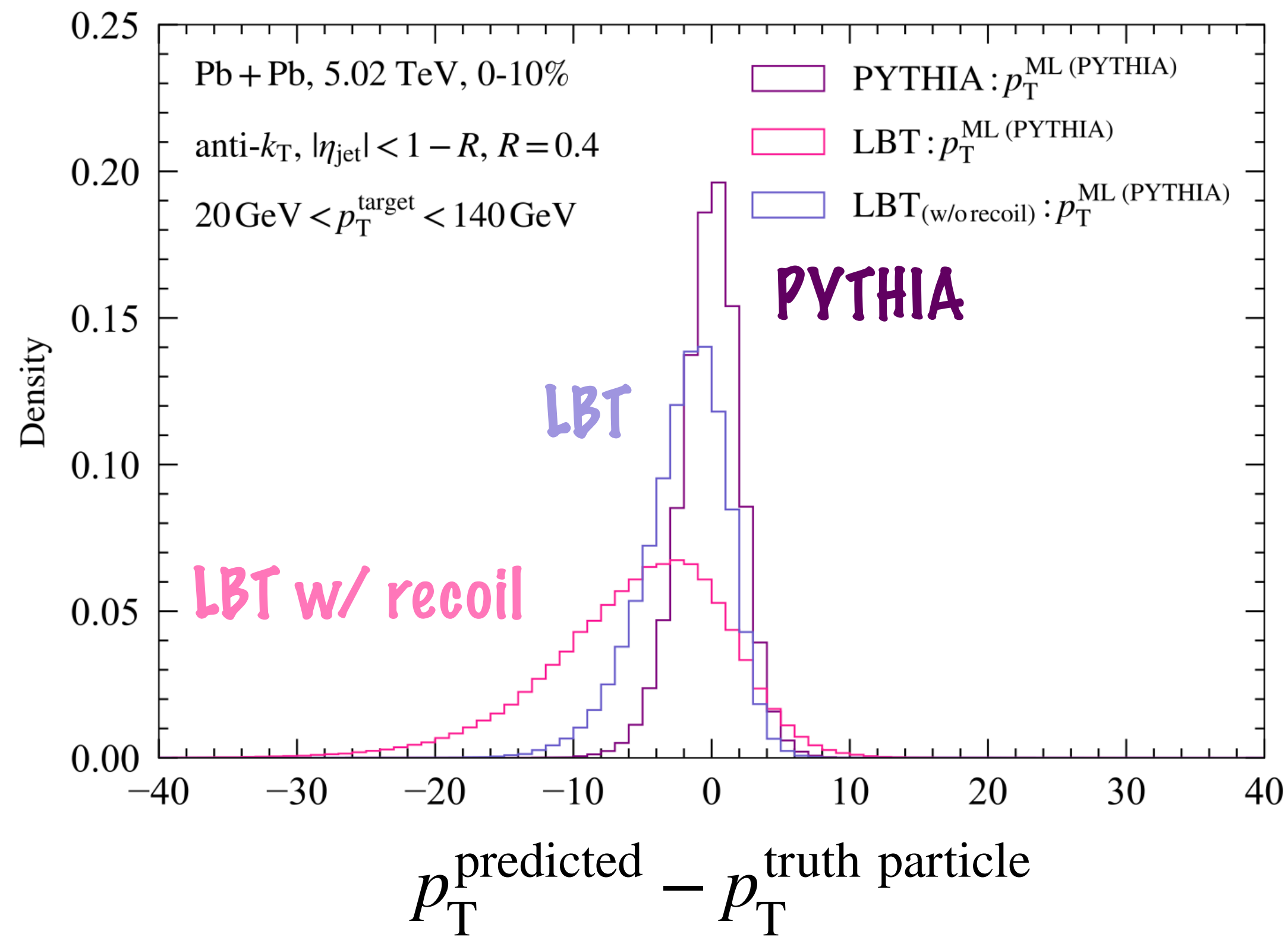
[PLB 849 \(2024\) 138412](#)



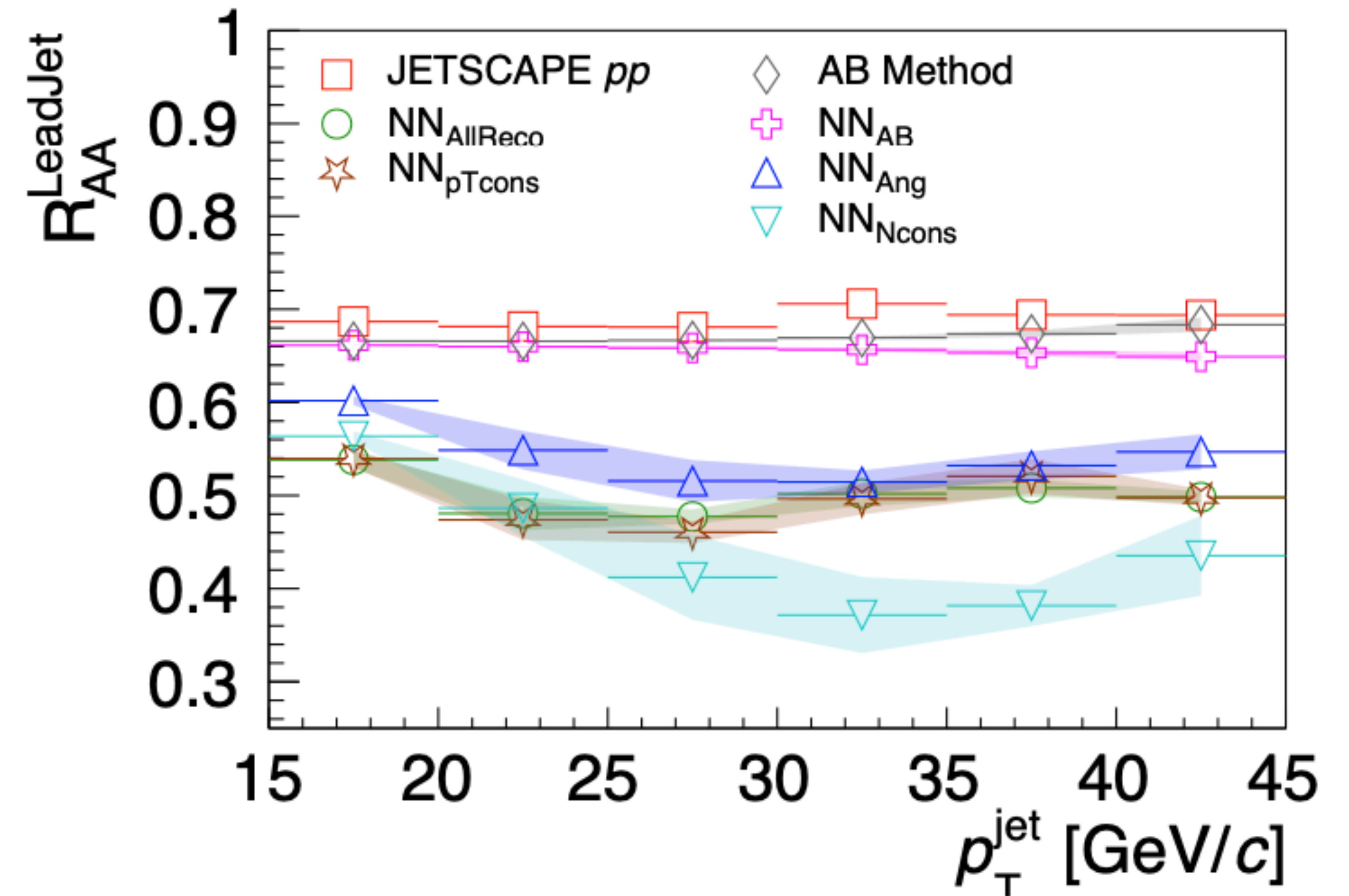
Regression for Jet Momentum

- Limitation

➔ trained on *PYTHIA* (unquenched jets)+uncorrelated underlying events and applied to quenched jets → **model-dependent bias**



[arXiv:2412.06466](https://arxiv.org/abs/2412.06466)

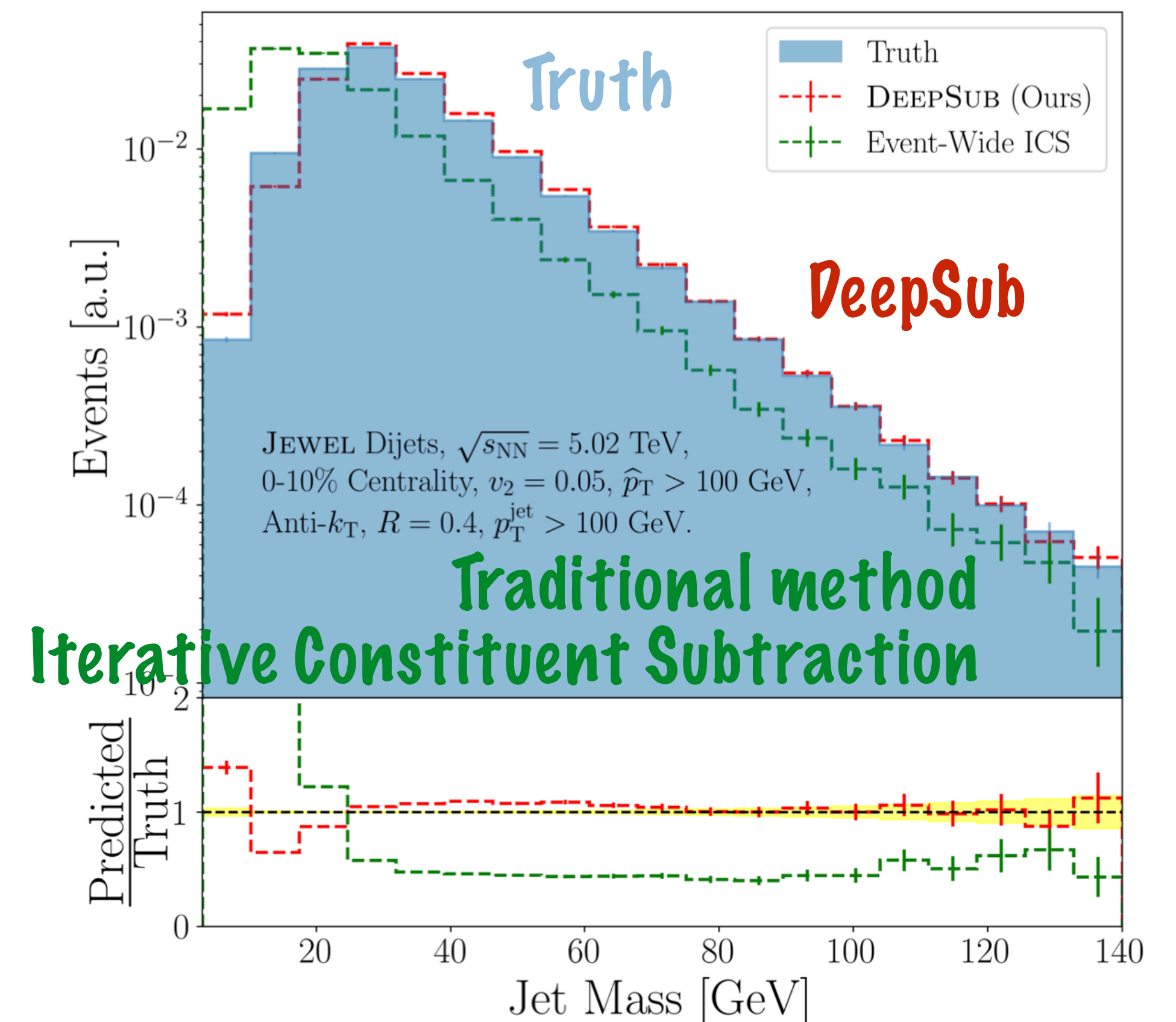
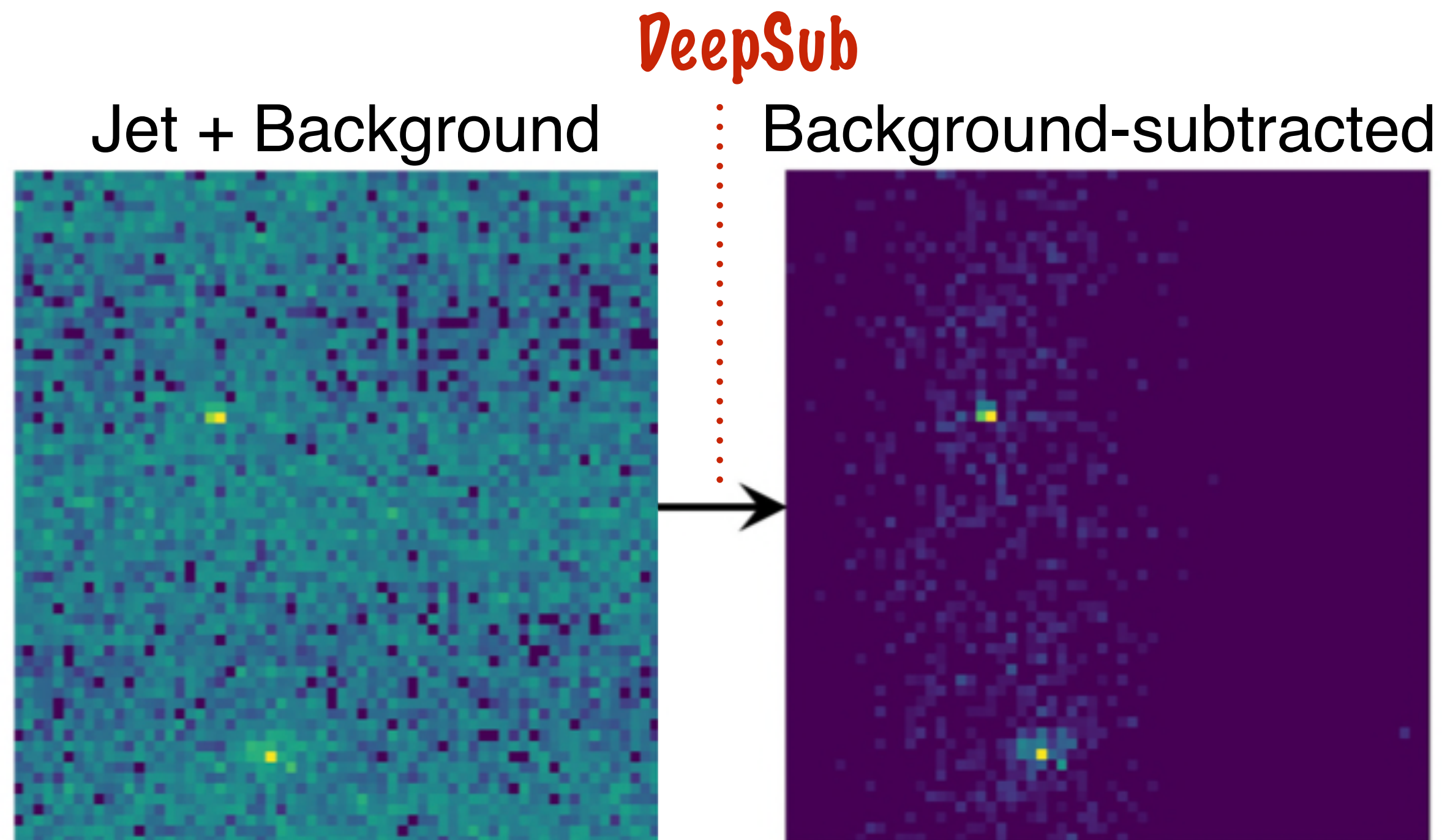


[arXiv:2412.15440](https://arxiv.org/abs/2412.15440)

Generative AI for Jet Background Subtraction

- Jet background subtraction as an “image” using Swin Transformer in a *supervised way*
 - ➔ trained with JEWEL **DeepSub model**
- Rather than a single value, “image” preserves full dimensional information
- **Limitation:** supervised learning → *model-dependent bias*

arXiv:2507.14036



Generative AI for Jet Background Subtraction

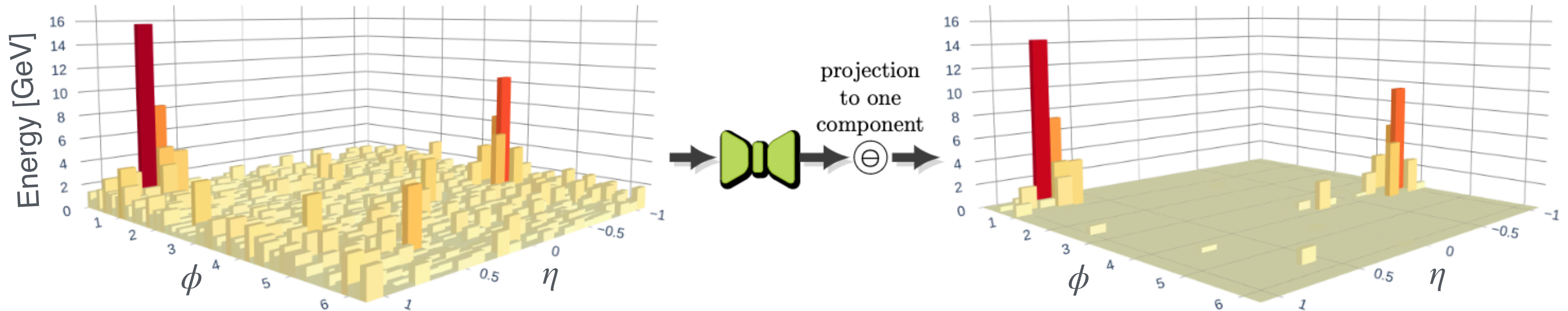
- Jet background subtraction as an “image” using cycle-consistent GAN

↓
Unsupervised learning, **unpaired**
image-to-image translation

- **Unsupervised** generative AI

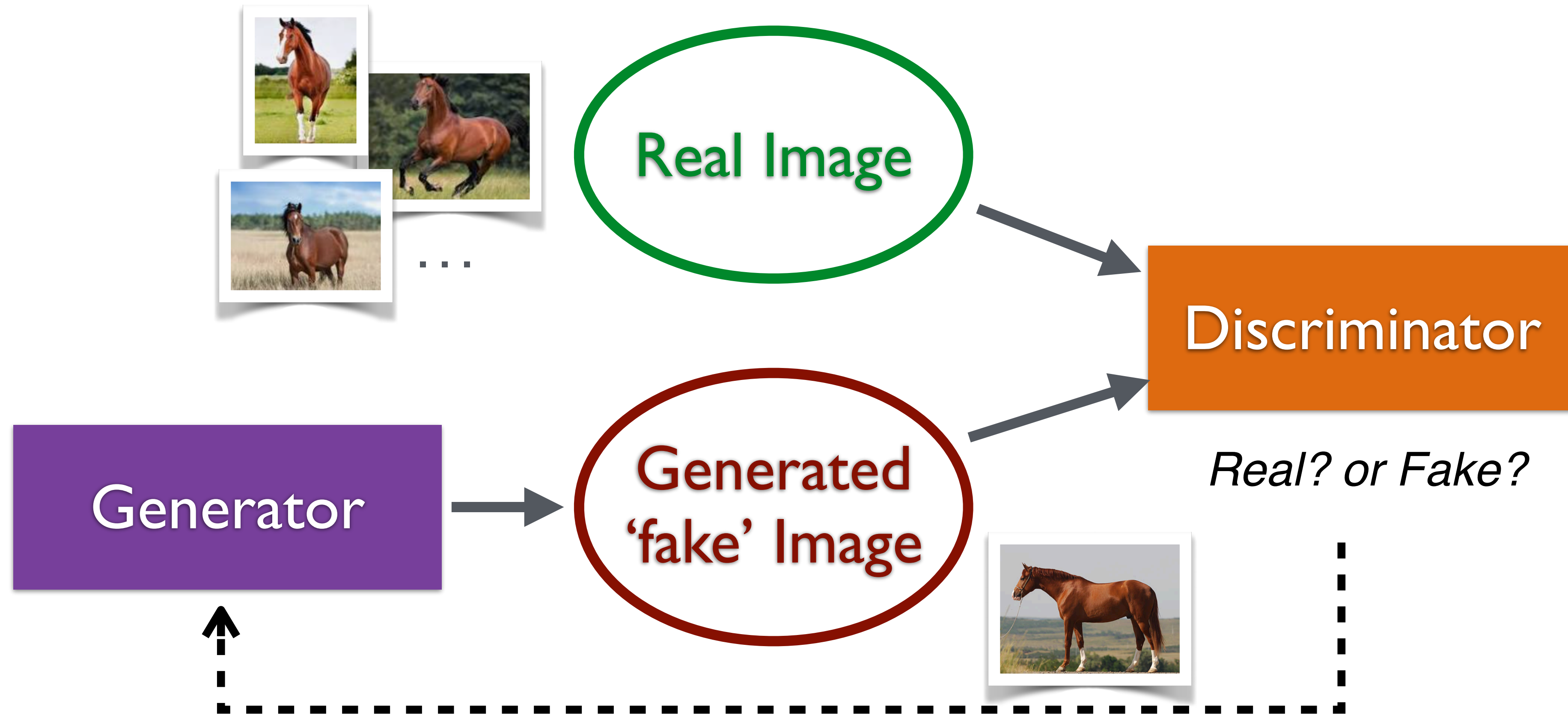
→ *preserves the higher dimensional information*

→ **can use experimental data (without label) for training**
: mitigate model-dependence



[arXiv:2510.23717](https://arxiv.org/abs/2510.23717)

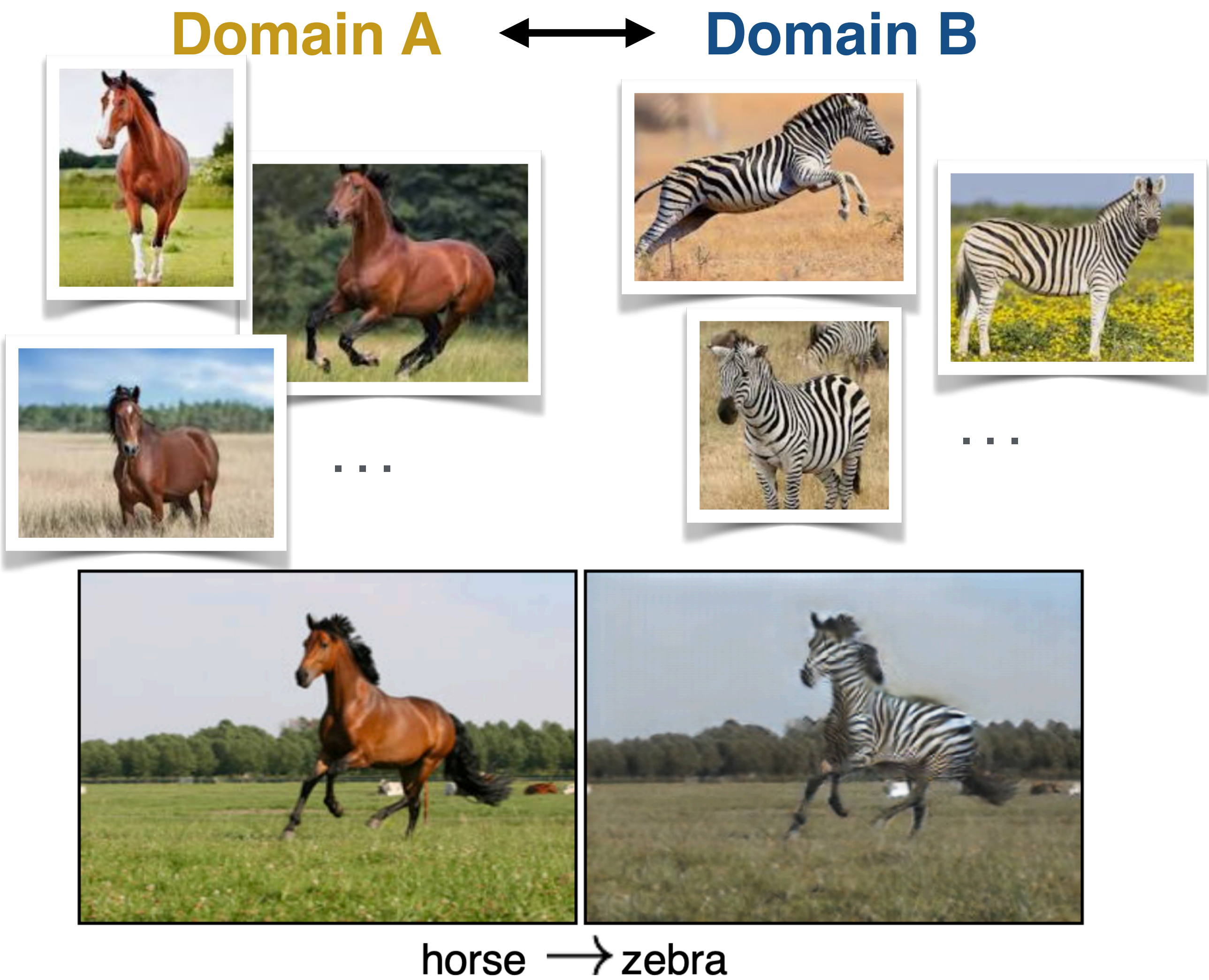
Generative Adversarial Network (GAN)



- Two neural networks (**Generator** and **Discriminator**) *compete* against each other
 - ➔ **feedback** to each other to improve
 - ➔ the process is repeated until **Discriminator** can no longer distinguish fake generated by **Generator** from real

Cycle-Consistent GAN

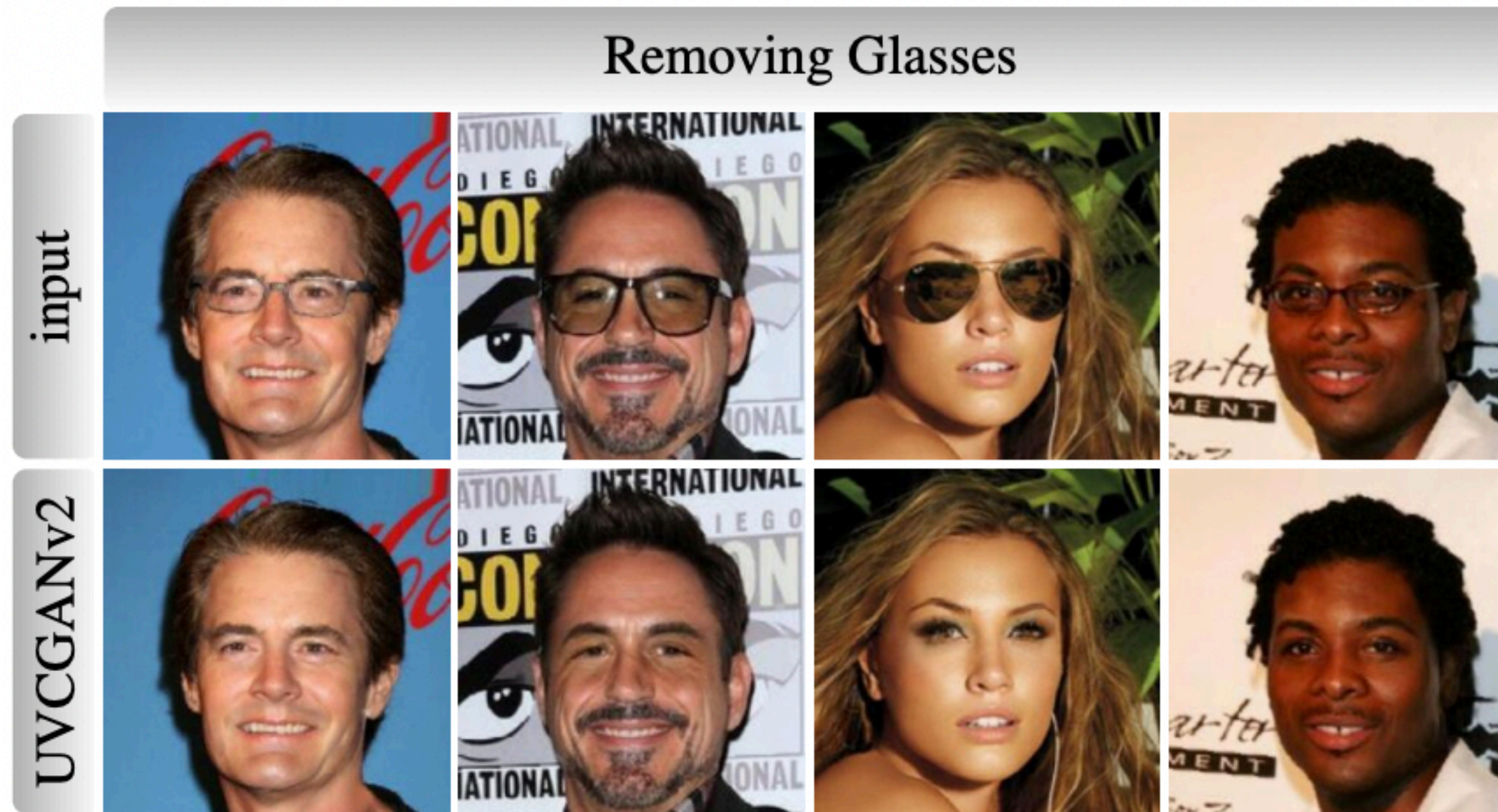
- **Unsupervised** learning, **unpaired** image-to-image translation



arXiv:1703.10593

UVCGAN

- **UVCGAN** (U-net Vision transformer Cycle-consistent Generative Adversarial Network)
 - ➔ **unpaired** image-to-image (I2I) translation; bridging gap between simulation and data reference

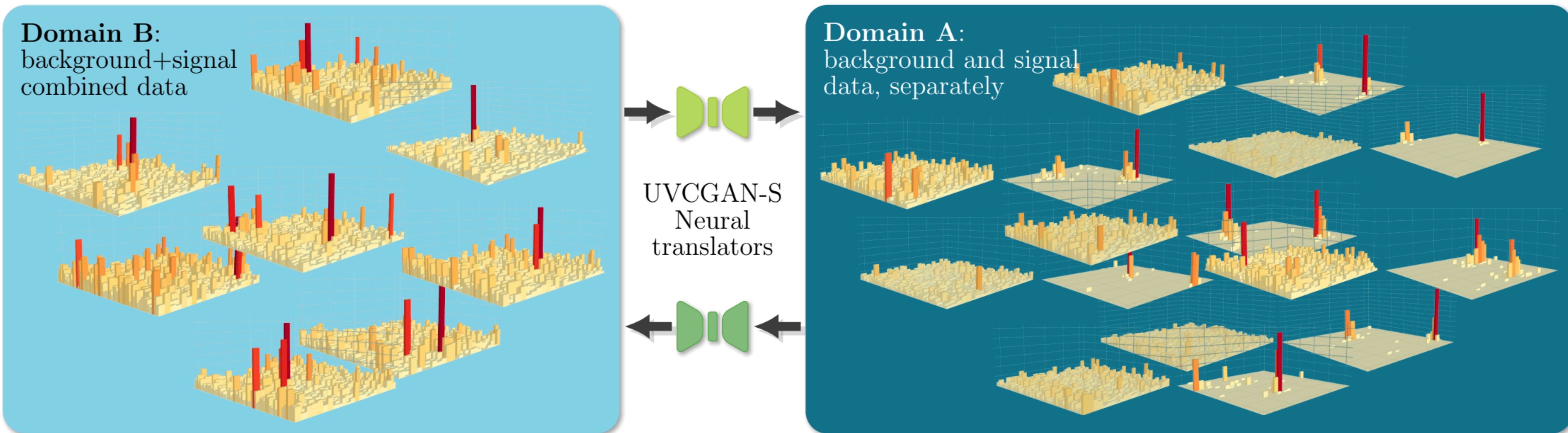


[arXiv:2303.16280](https://arxiv.org/abs/2303.16280)

UVCGAN-S: Unpaired Unsupervised Training

- As a proof-of-principle, UVCGAN-S model was trained on images of *EMCal + HCAL combined* calorimeter energies in (η, ϕ) with three samples
 - **PYTHIA**: Inclusive Jet $\hat{p}_T = 30$ GeV
 - **HIJING**: 0-10% Centrality HIJING
 - **PYTHIA+HIJING**: sum of calorimeter energies of PYTHIA and HIJING

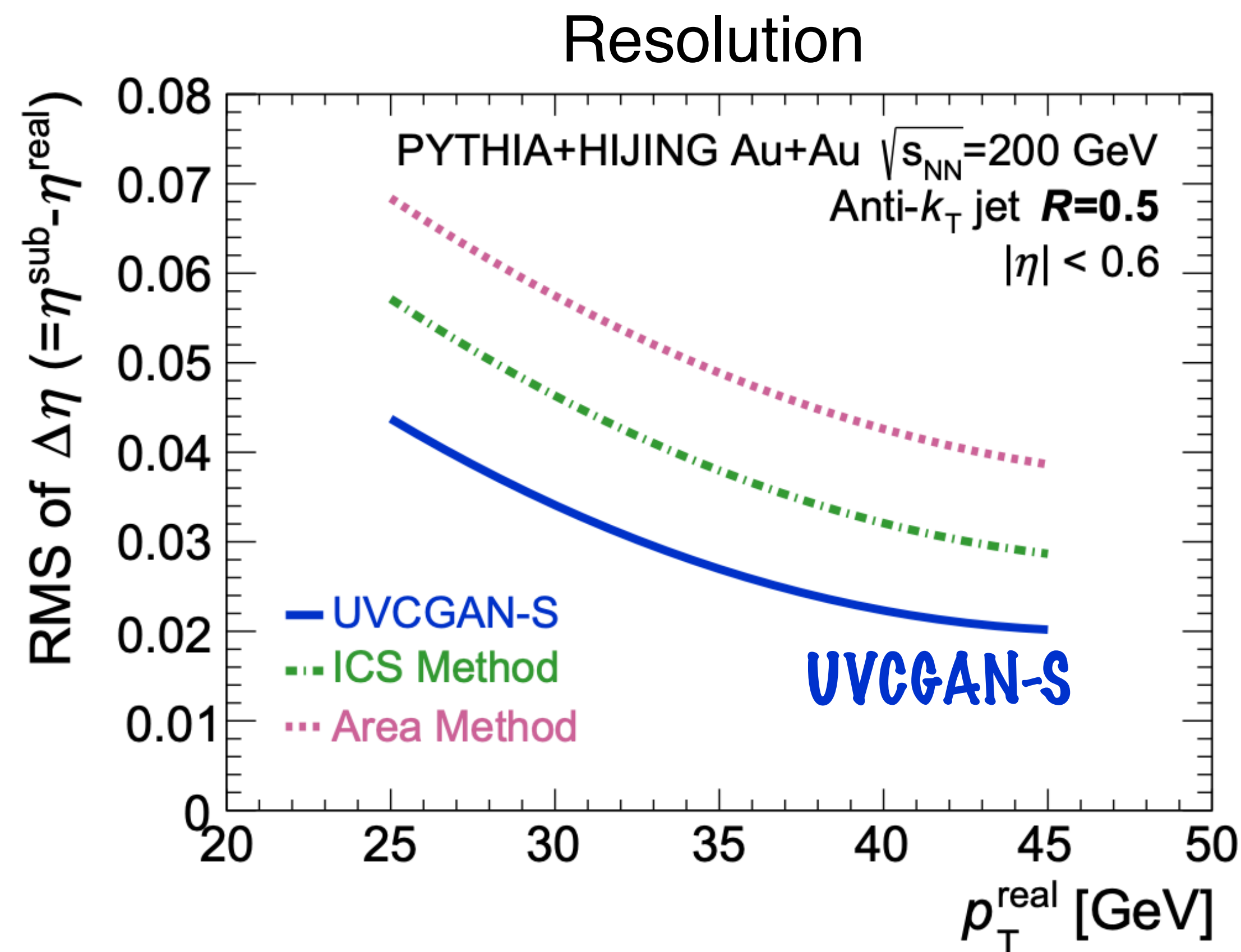
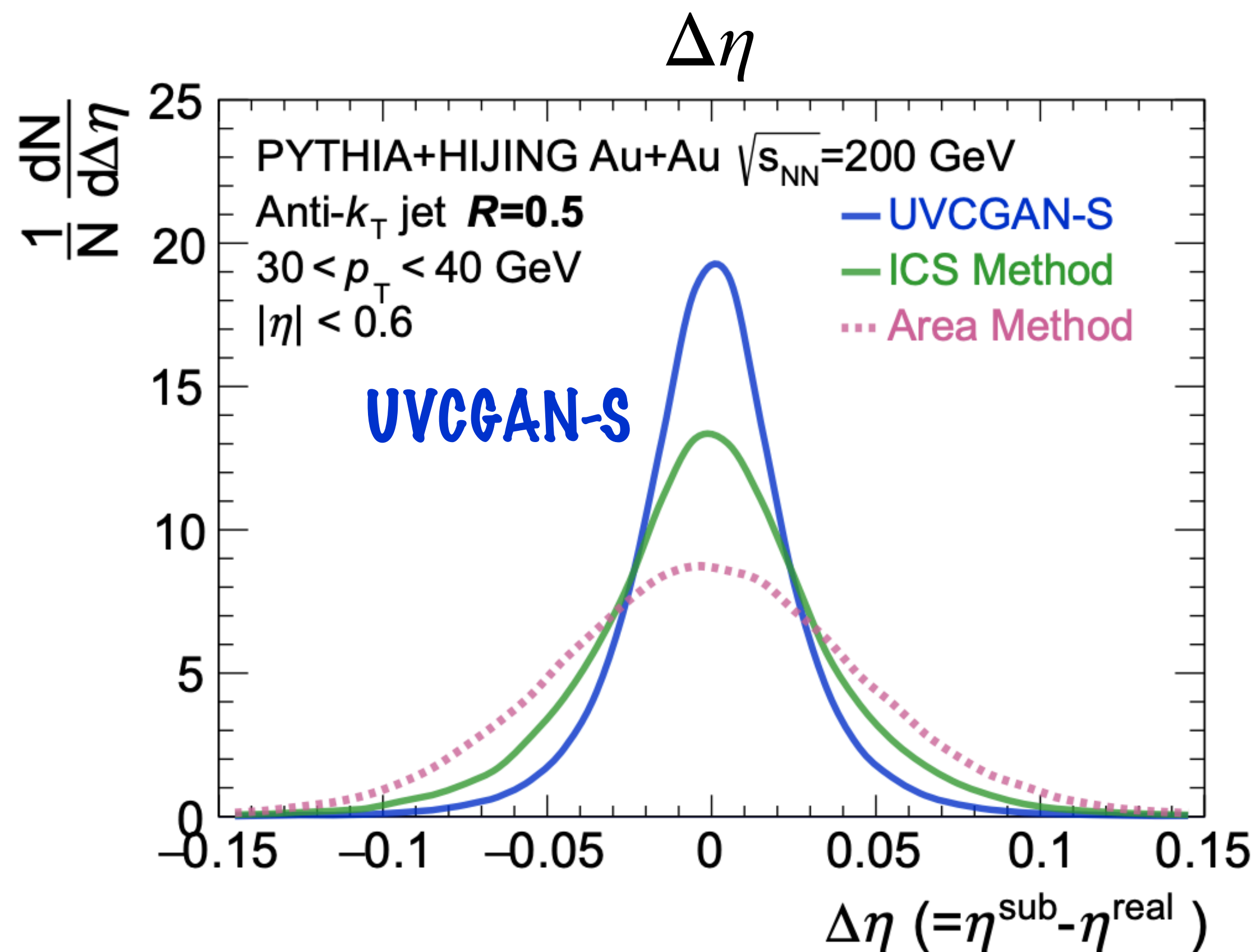
[arXiv:2510.23717](https://arxiv.org/abs/2510.23717)



UVCGAN-S: Position Response

sub: bkg-subtracted
real: response in $p+p$

- Compared to **Iterative Constituent Subtraction (ICS)** and **Area-based Method**
- **UVCGAN-S** achieves up to 40% better position resolution than traditional methods

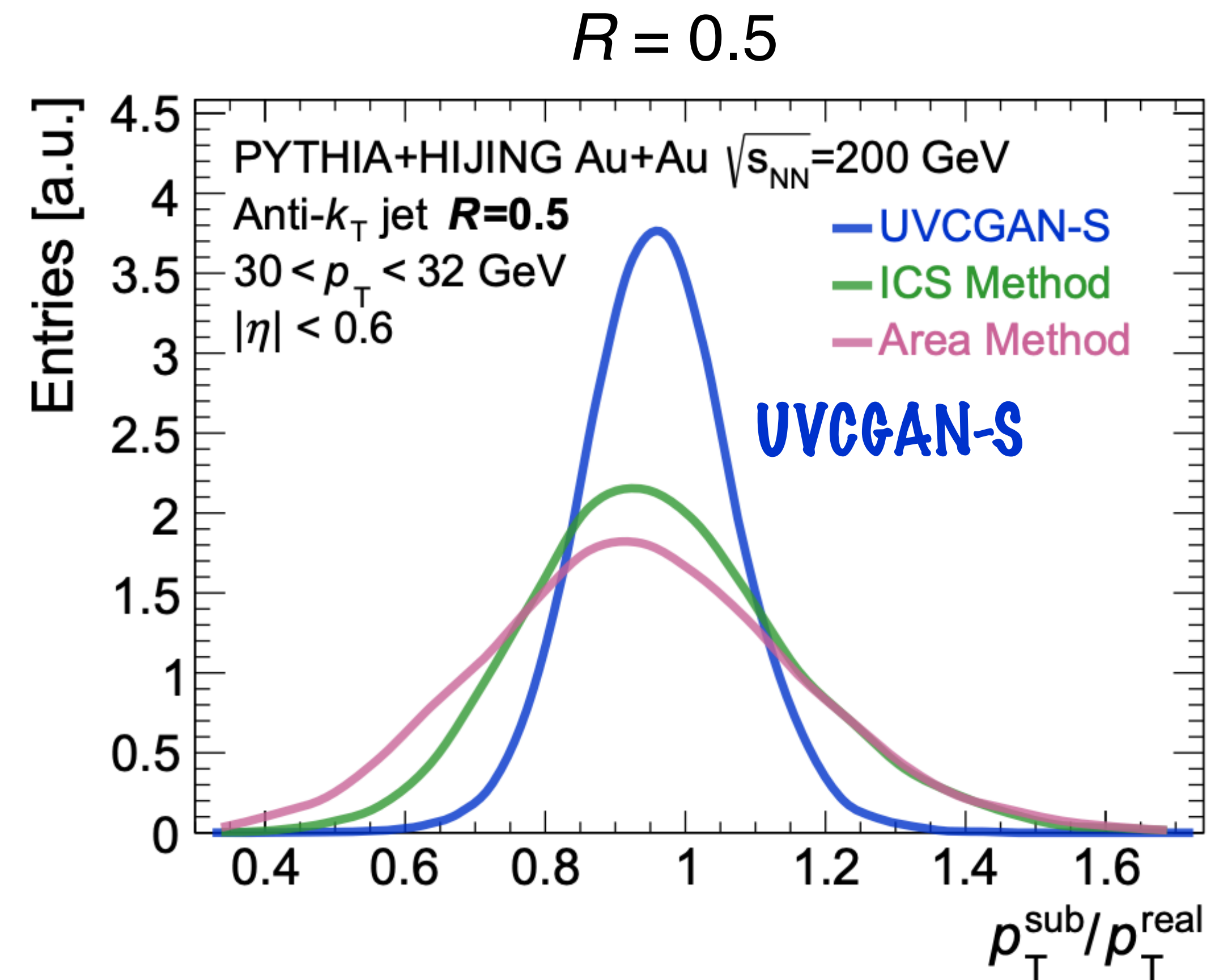
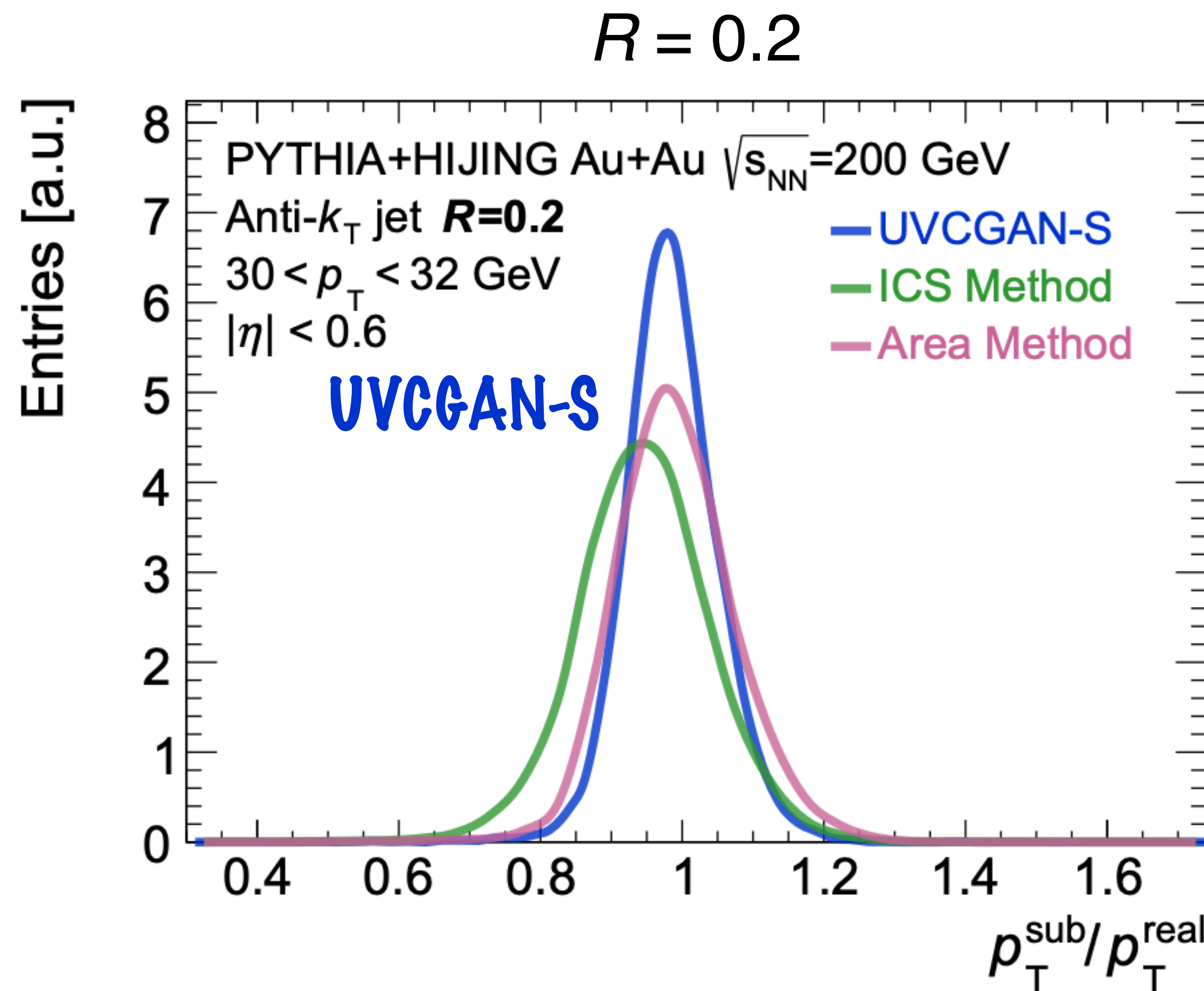


[arXiv:2510.23717](https://arxiv.org/abs/2510.23717)

UVCGAN-S: Momentum Response

sub: bkg-subtracted
real: response in $p+p$

- Compared to **Iterative Constituent Subtraction (ICS)** and **Area-based Method**
- Dramatic improvement in resolution, especially at large R



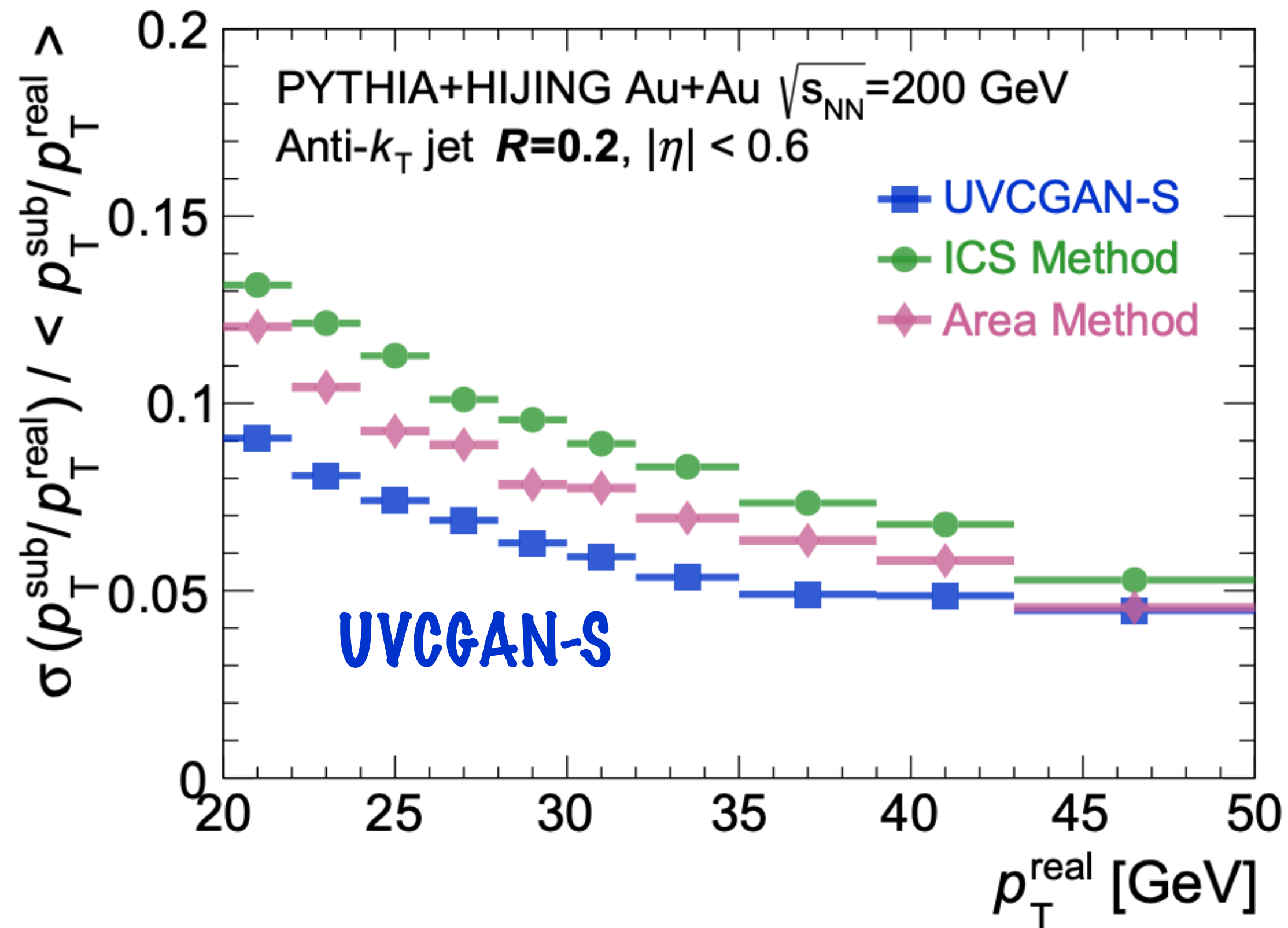
[arXiv:2510.23717](https://arxiv.org/abs/2510.23717)

UVCGAN-S: Momentum Response

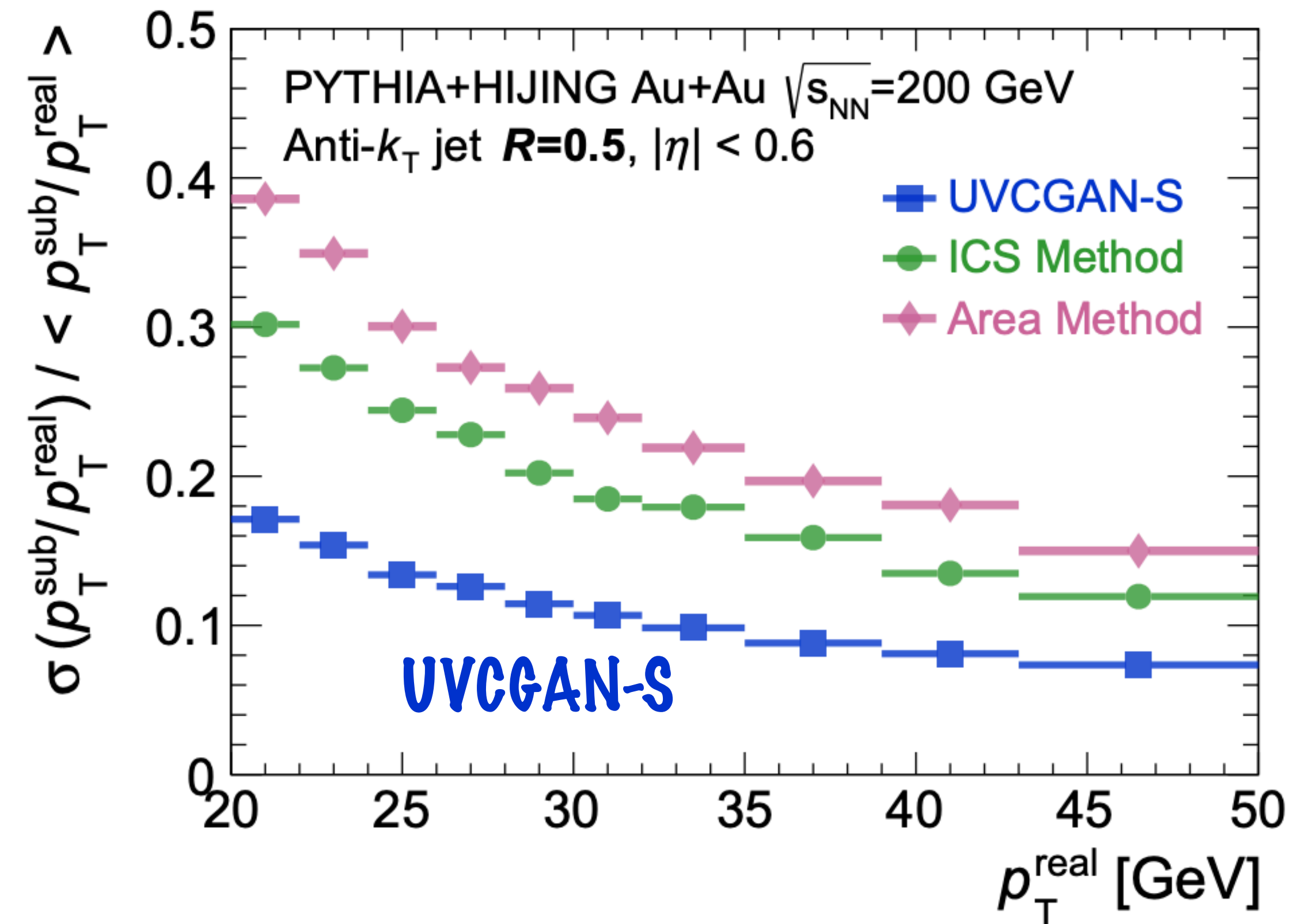
sub: bkg-subtracted
real: response in $p+p$

- Compared to **Iterative Constituent Subtraction (ICS)** and **Area-based Method**
- **Dramatic improvement in resolution, especially at large R**
 - ➔ 100% better momentum resolution for $R=0.5$

$R = 0.2$



$R = 0.5$

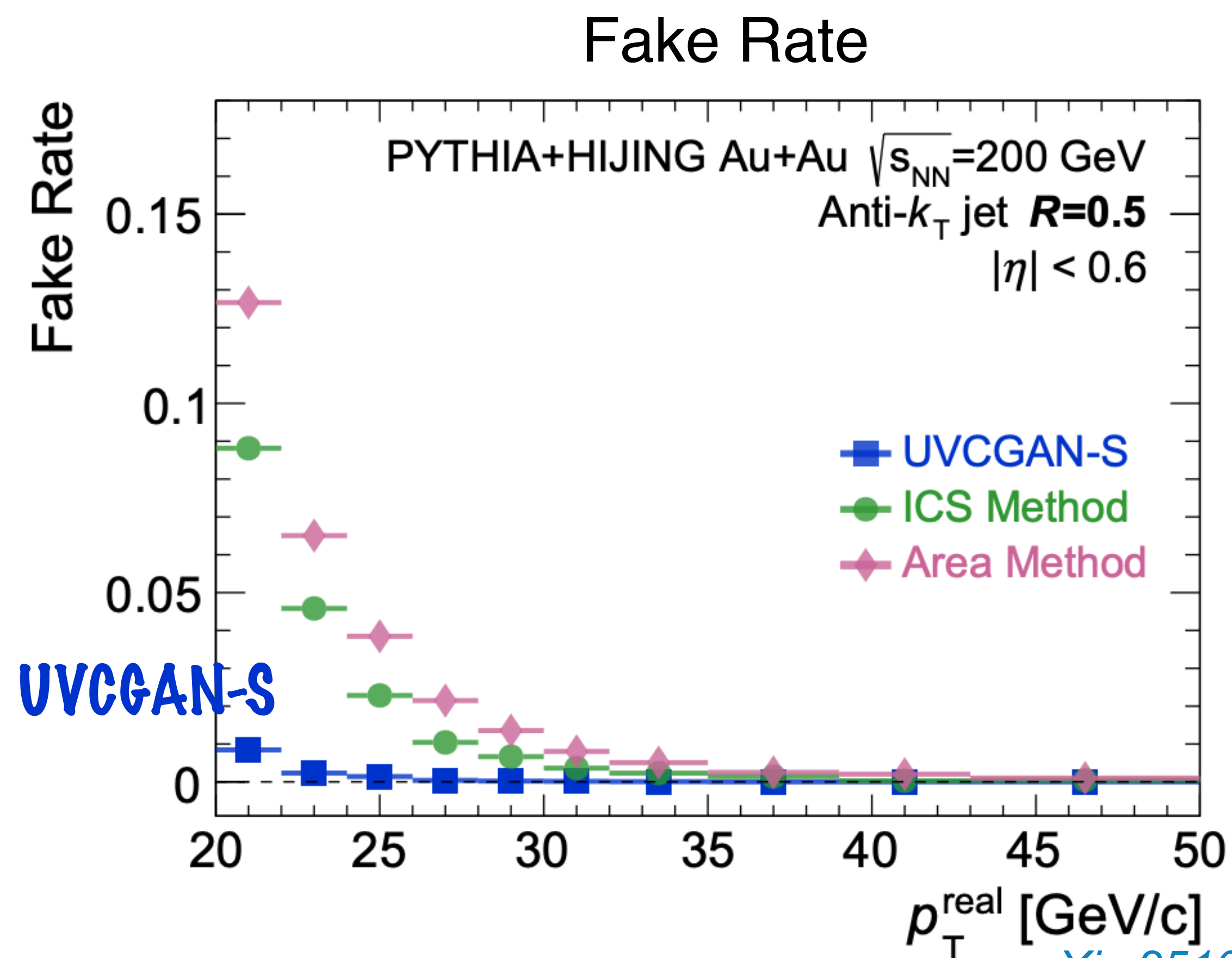
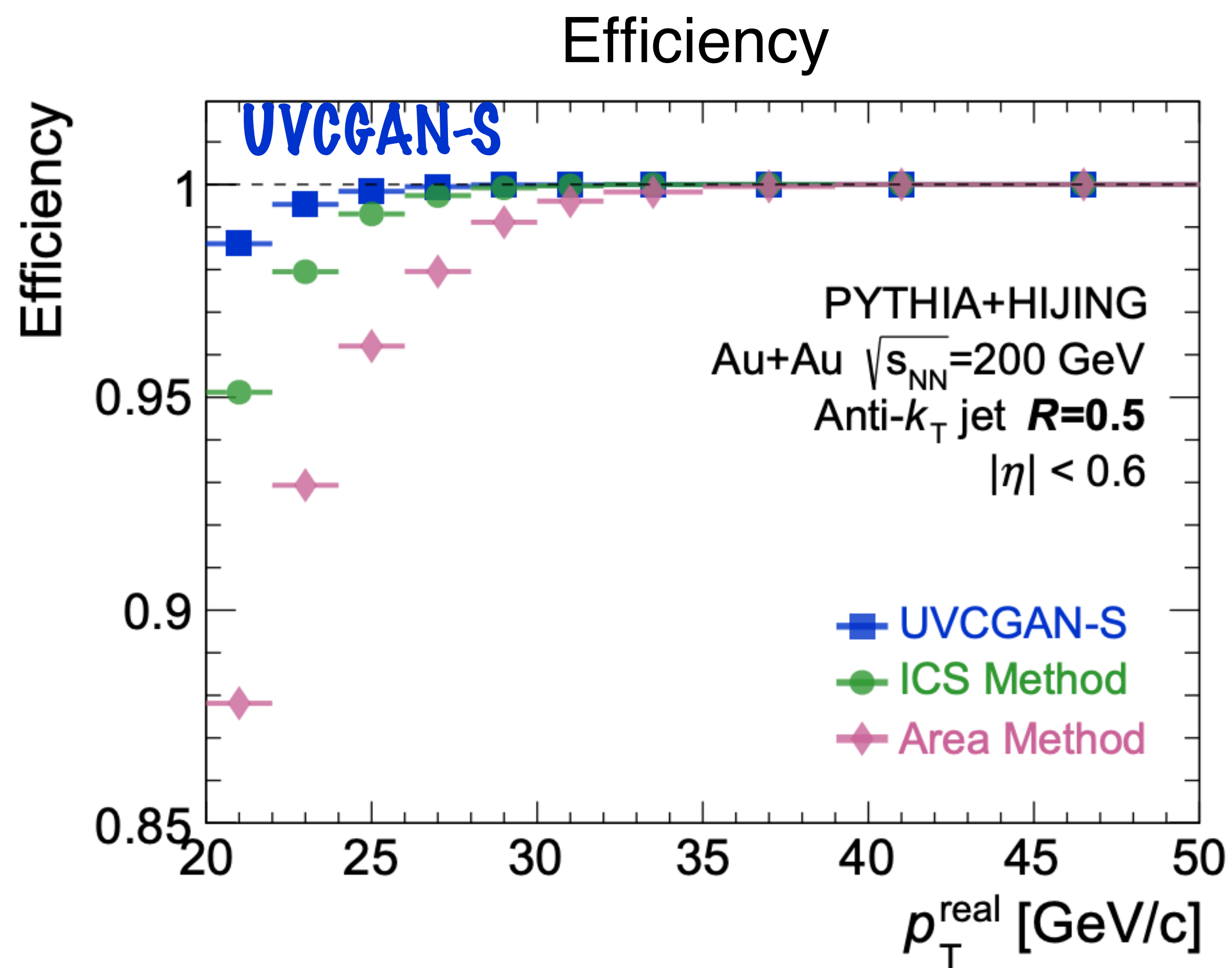


[arXiv:2510.23717](https://arxiv.org/abs/2510.23717)

UVCGAN-S: Efficiency and Fake Rate

sub: bkg-subtracted
real: response in $p+p$

- Compared to **Iterative Constituent Subtraction (ICS)** and **Area-based Method**
- **Dramatic improvement in efficiency and fake rate, especially at large R**
➔ efficiency 88% vs. 99%, fake rate 13% vs. 1% for $R=0.5$

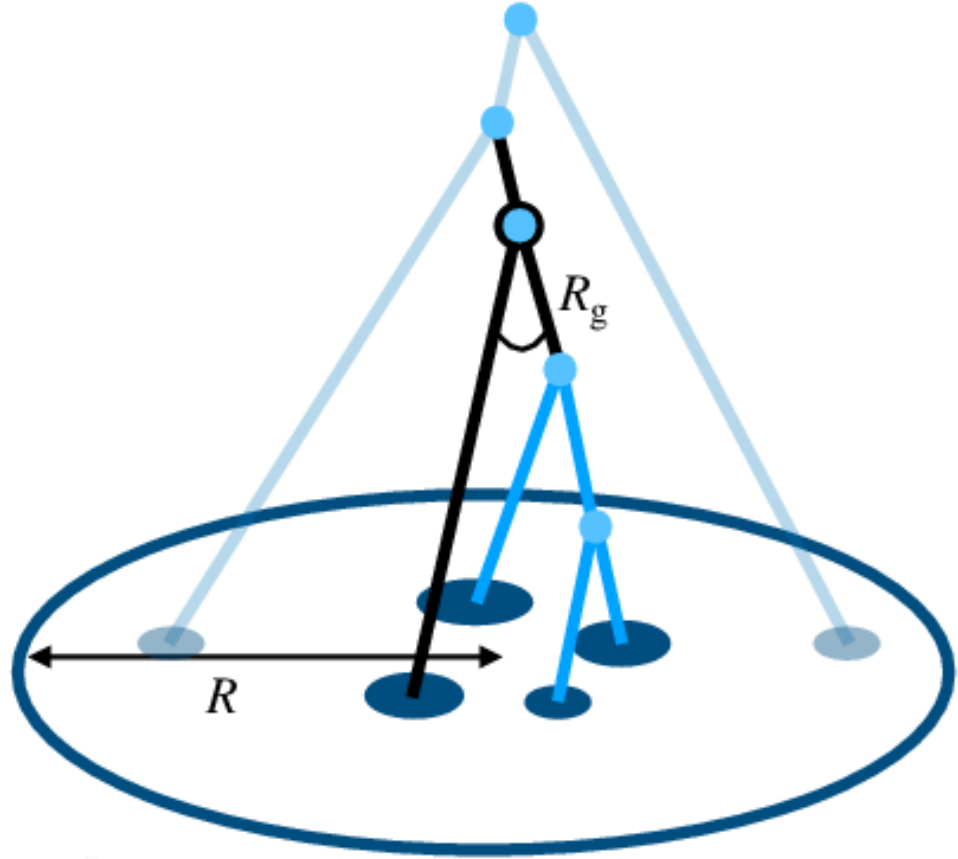


[arXiv:2510.23717](https://arxiv.org/abs/2510.23717)

UVCGAN-S: Jet Substructure

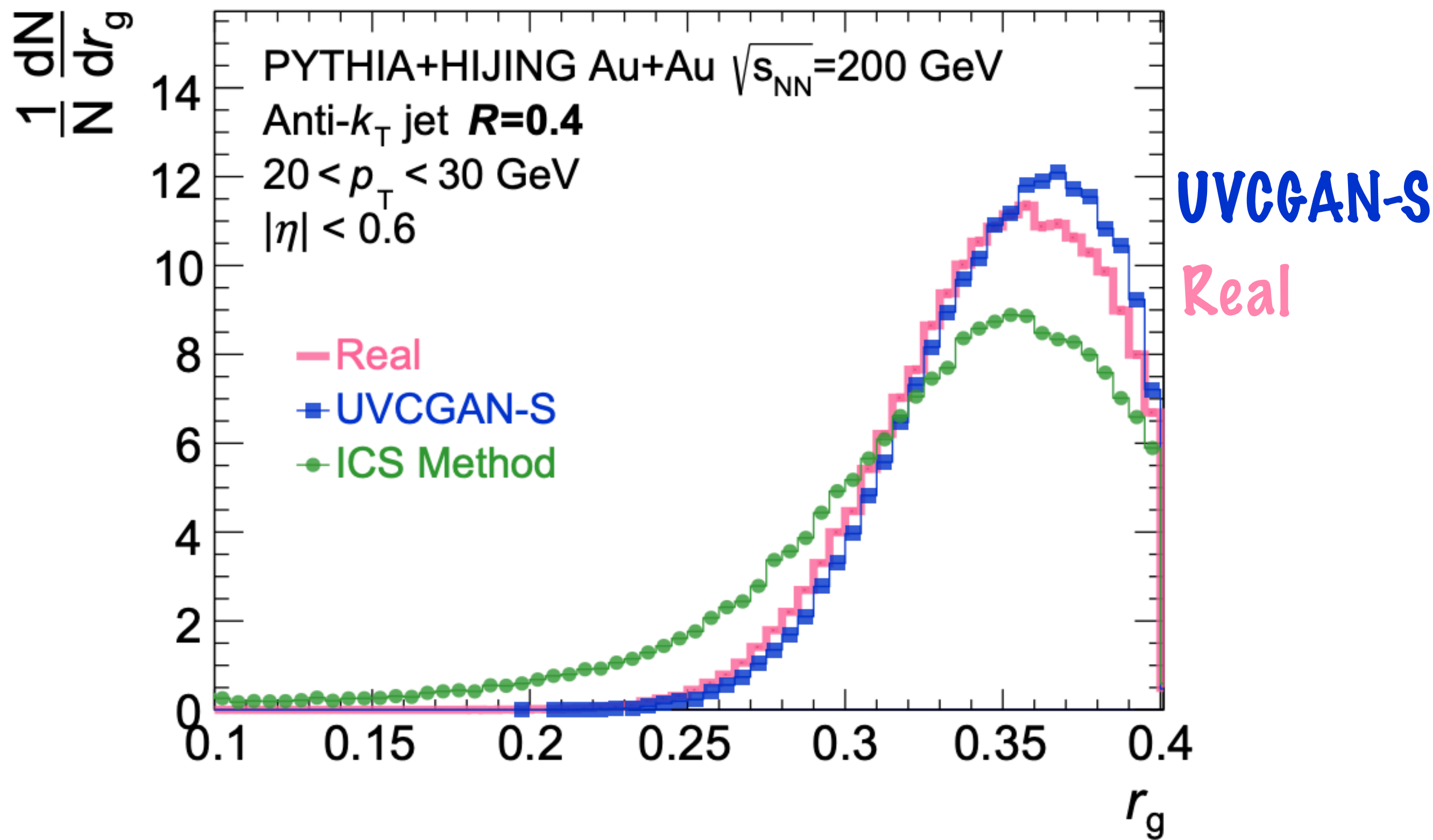
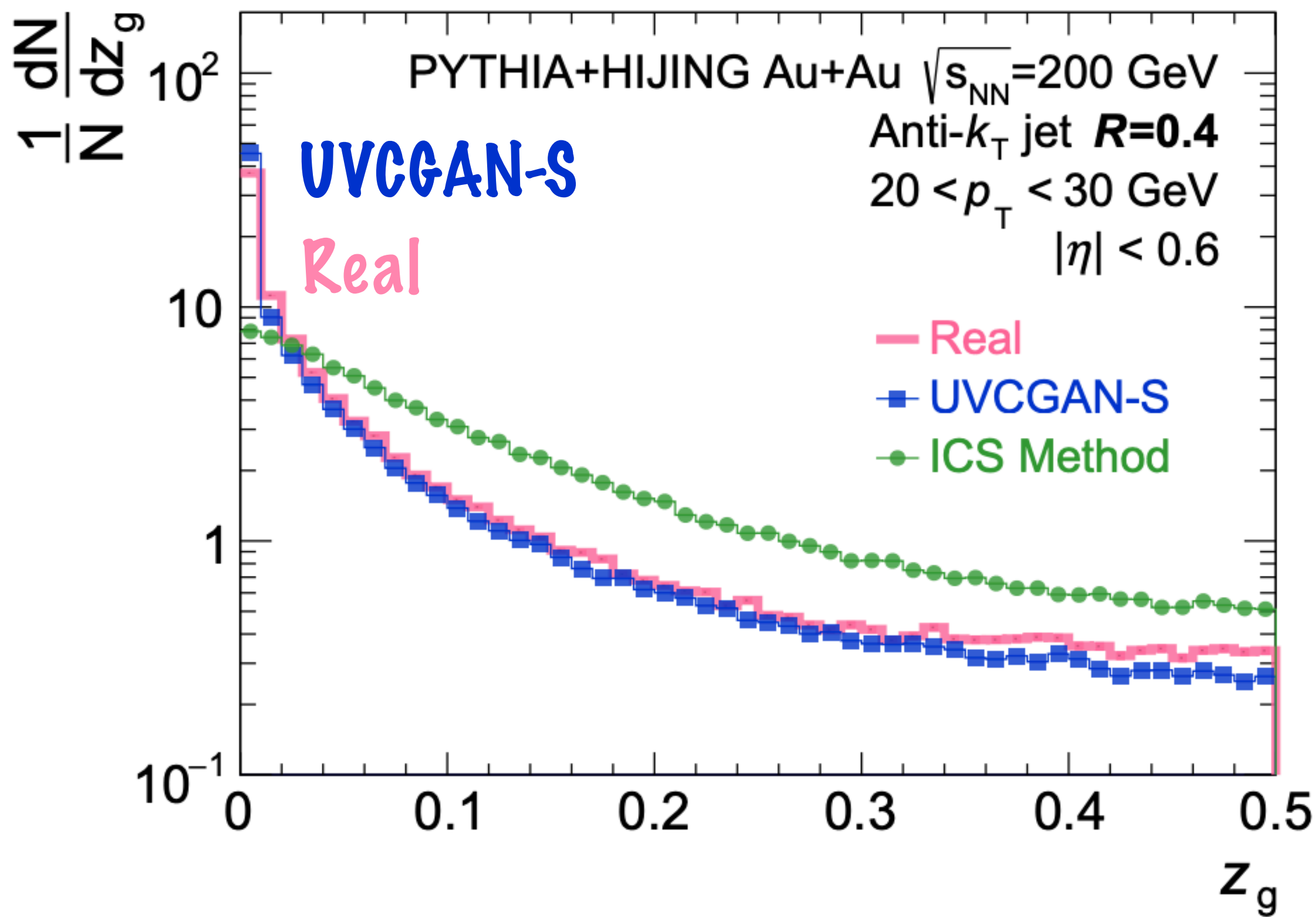
- Compared to **Iterative Constituent Subtraction (ICS)**
- Superior jet substructure reproduction of UVCGAN-S

sub: bkg-subtracted
real: response in p+p



$$z_g \equiv \frac{p_{T,\text{subleading}}}{p_{T,\text{leading}} + p_{T,\text{subleading}}}$$

$$R_g = \sqrt{\Delta y^2 + \Delta \phi^2}$$



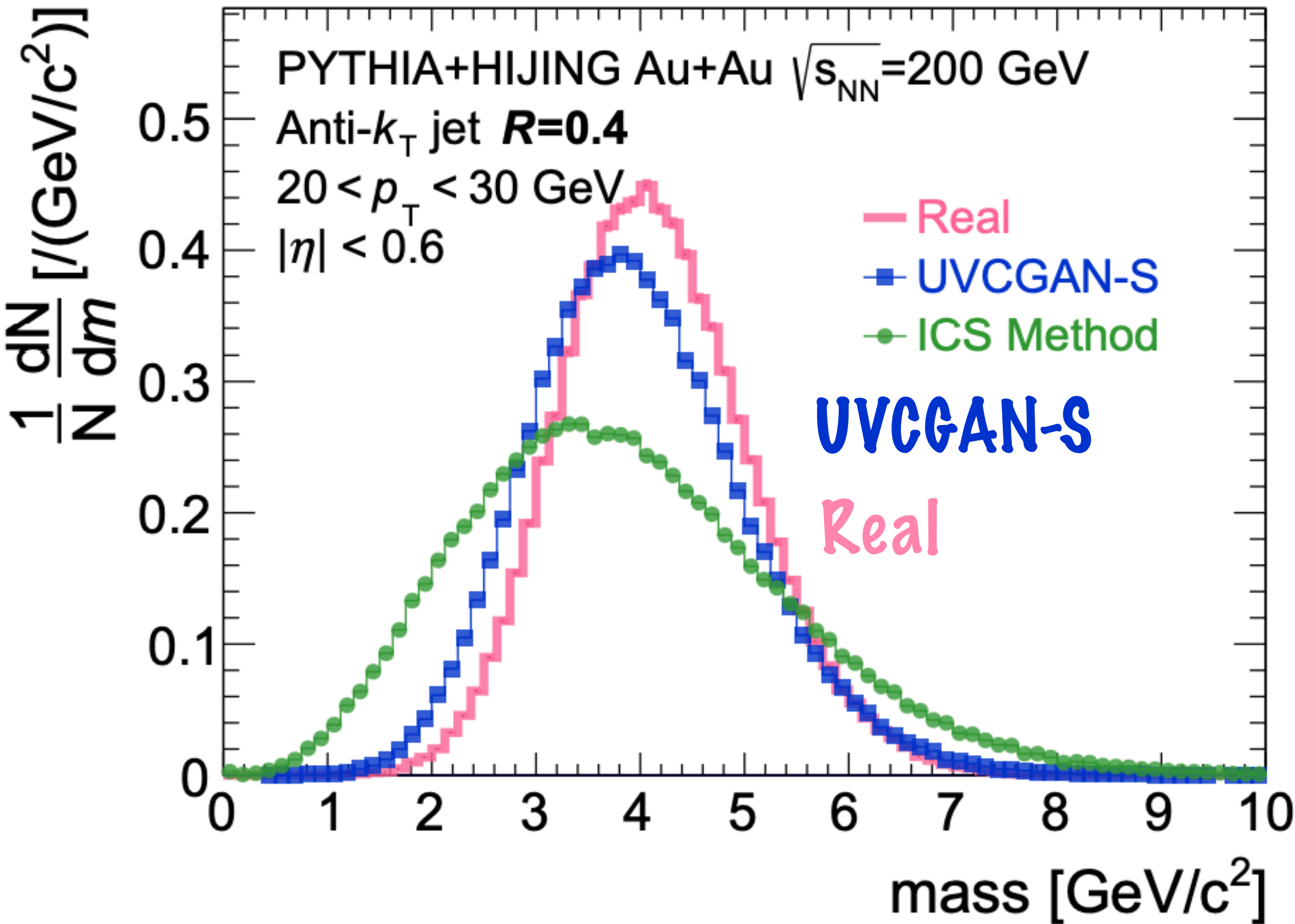
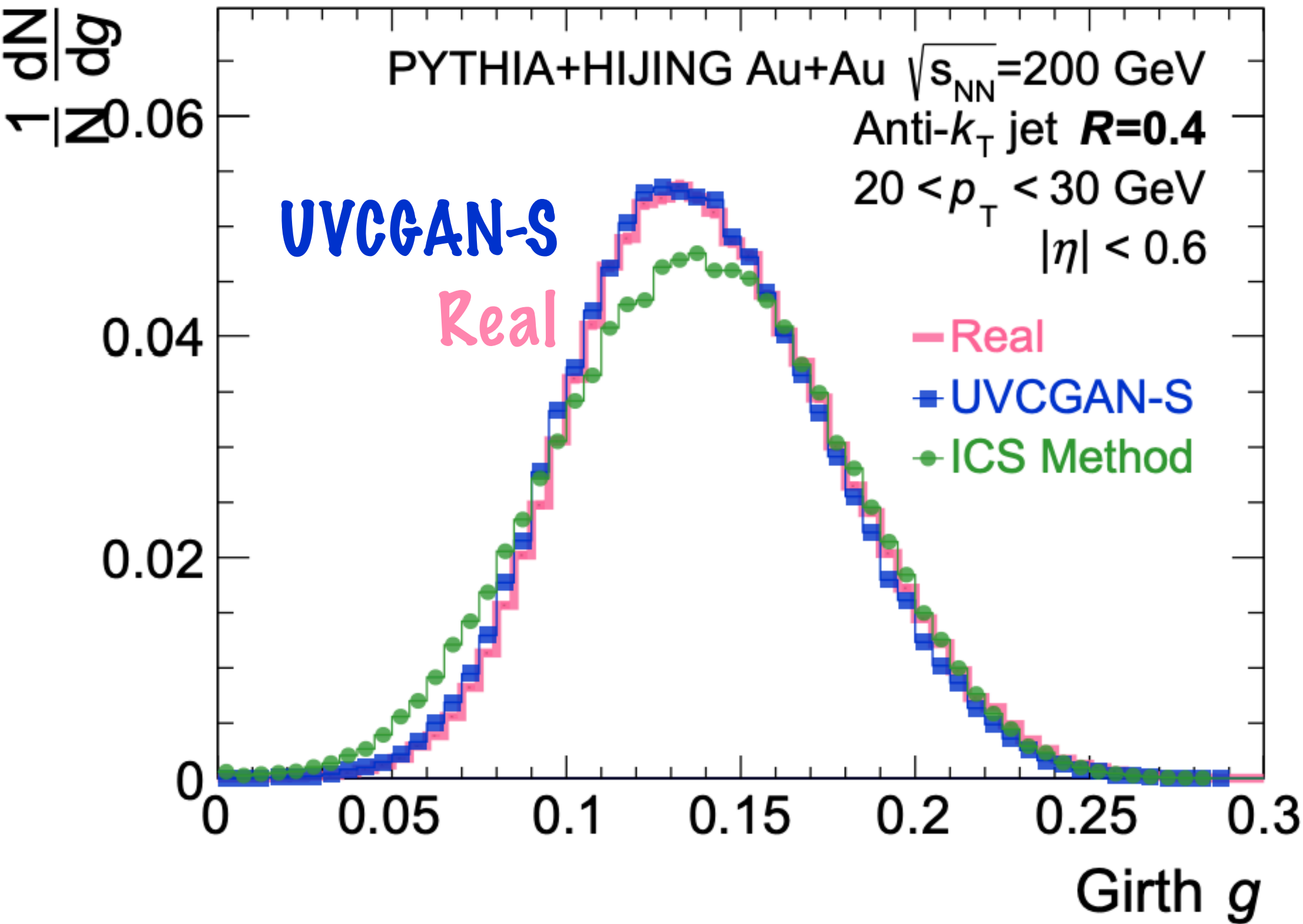
UVCGAN-S: Jet Substructure

sub: bkg-subtracted
real: response in p+p

- Compared to Iterative Constituent Subtraction (ICS)
- Superior jet substructure reproduction of UVCGAN-S

$$g = \frac{1}{p_{T,jet}} \sum_{i \in jet} p_{T,i} \Delta R_{i,axis}$$

Jet mass

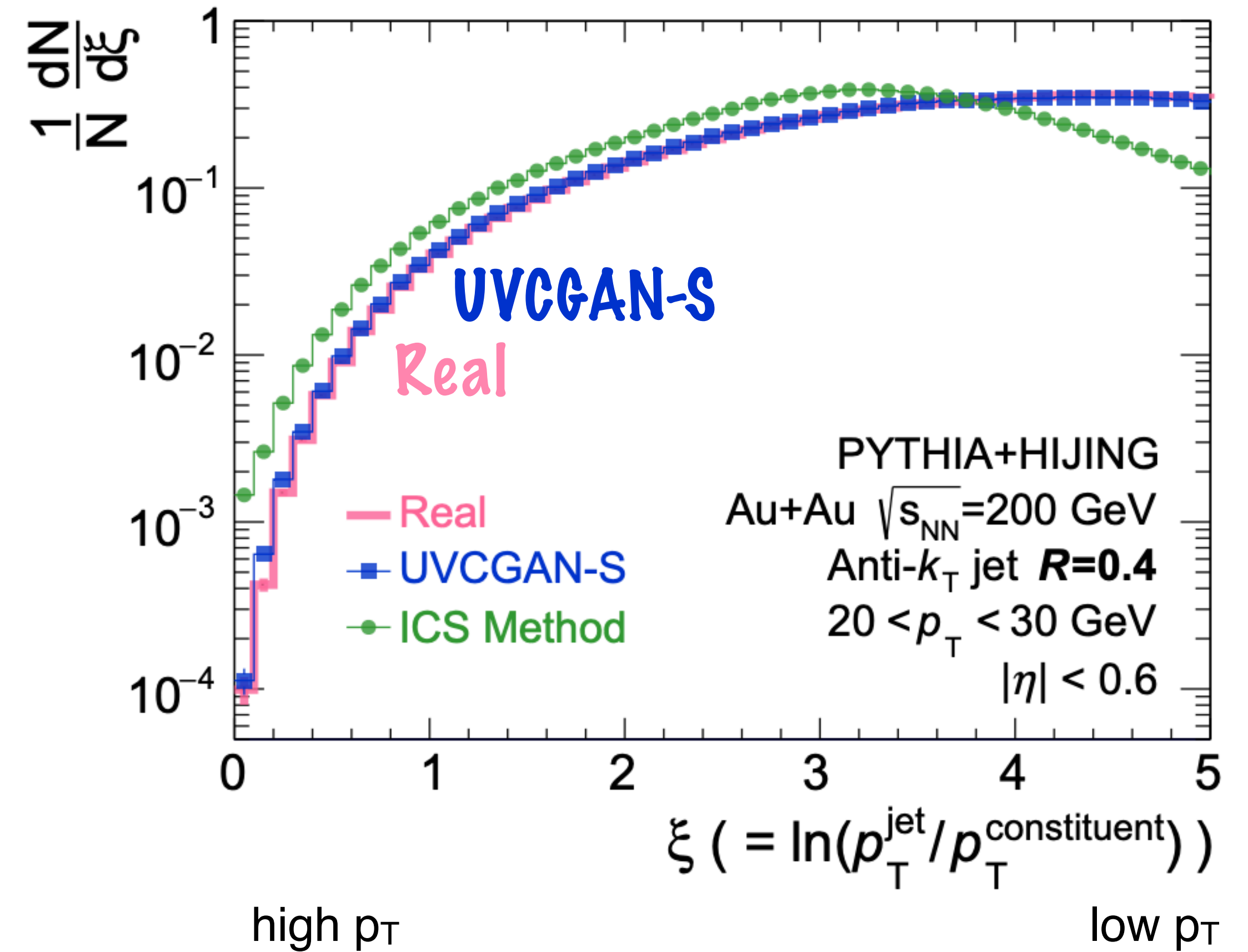
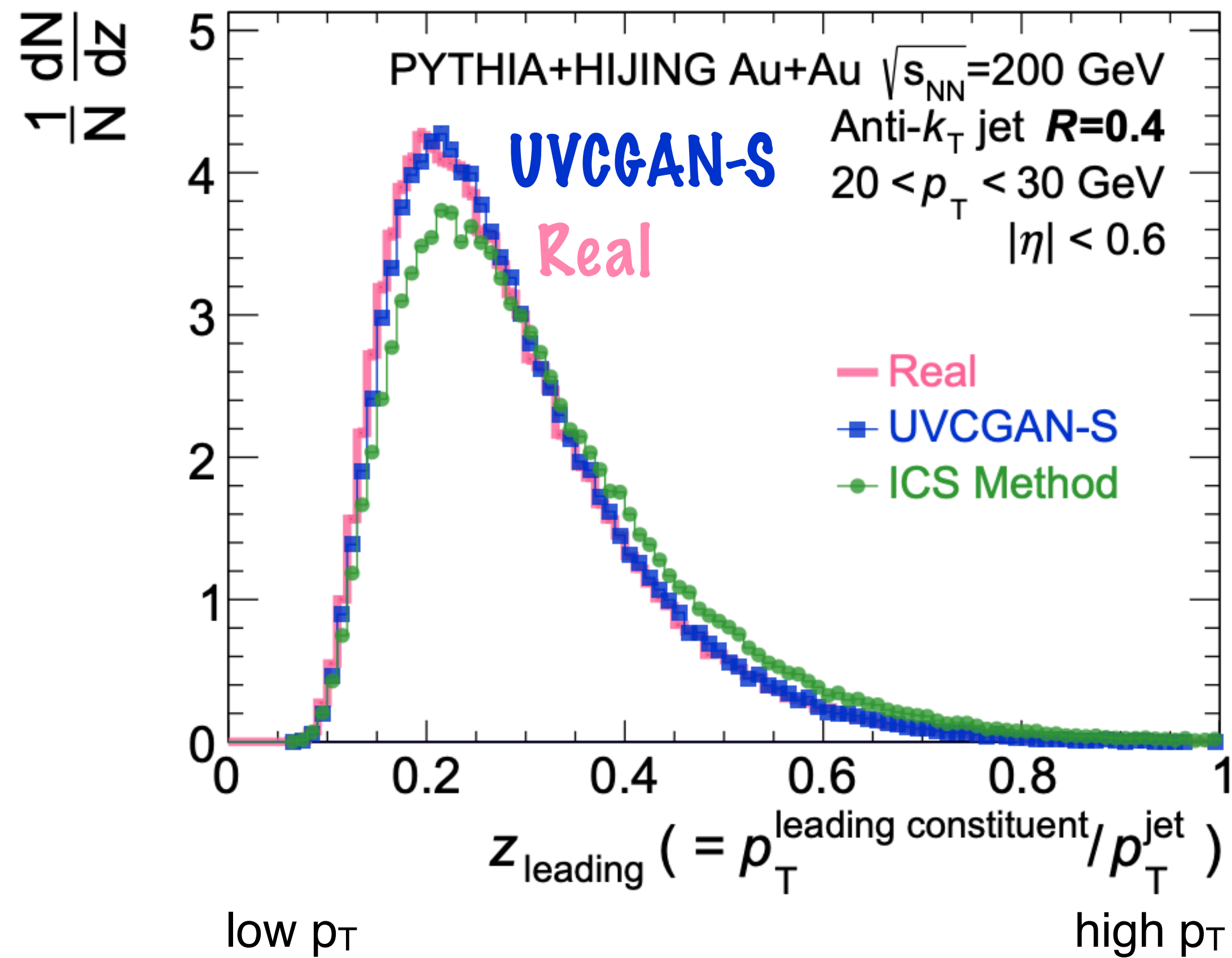


[arXiv:2510.23717](https://arxiv.org/abs/2510.23717)

UVCGAN-S: Jet Substructure

sub: bkg-subtracted
real: response in p+p

- Compared to Iterative Constituent Subtraction (ICS)
- Superior jet substructure reproduction of UVCGAN-S



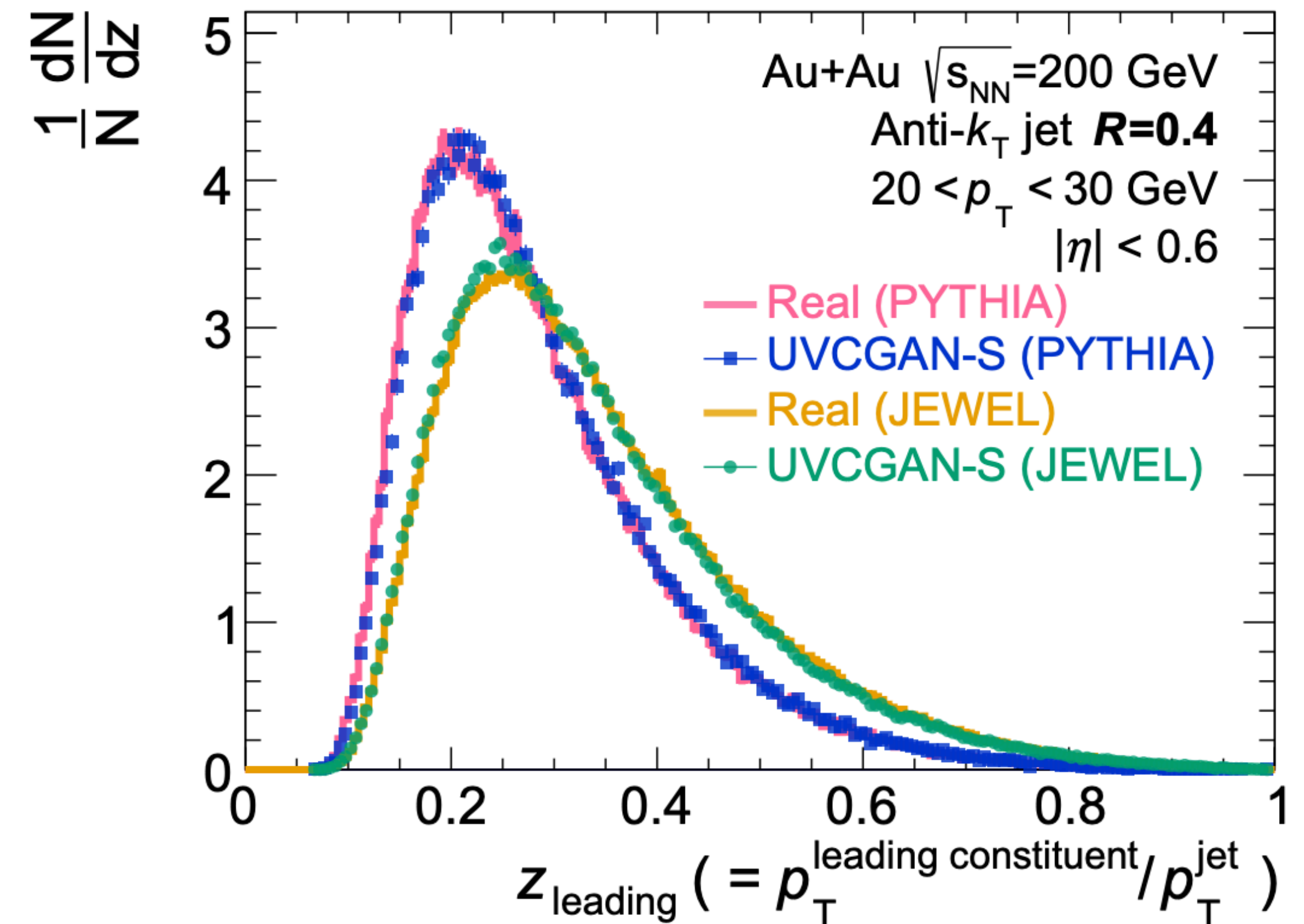
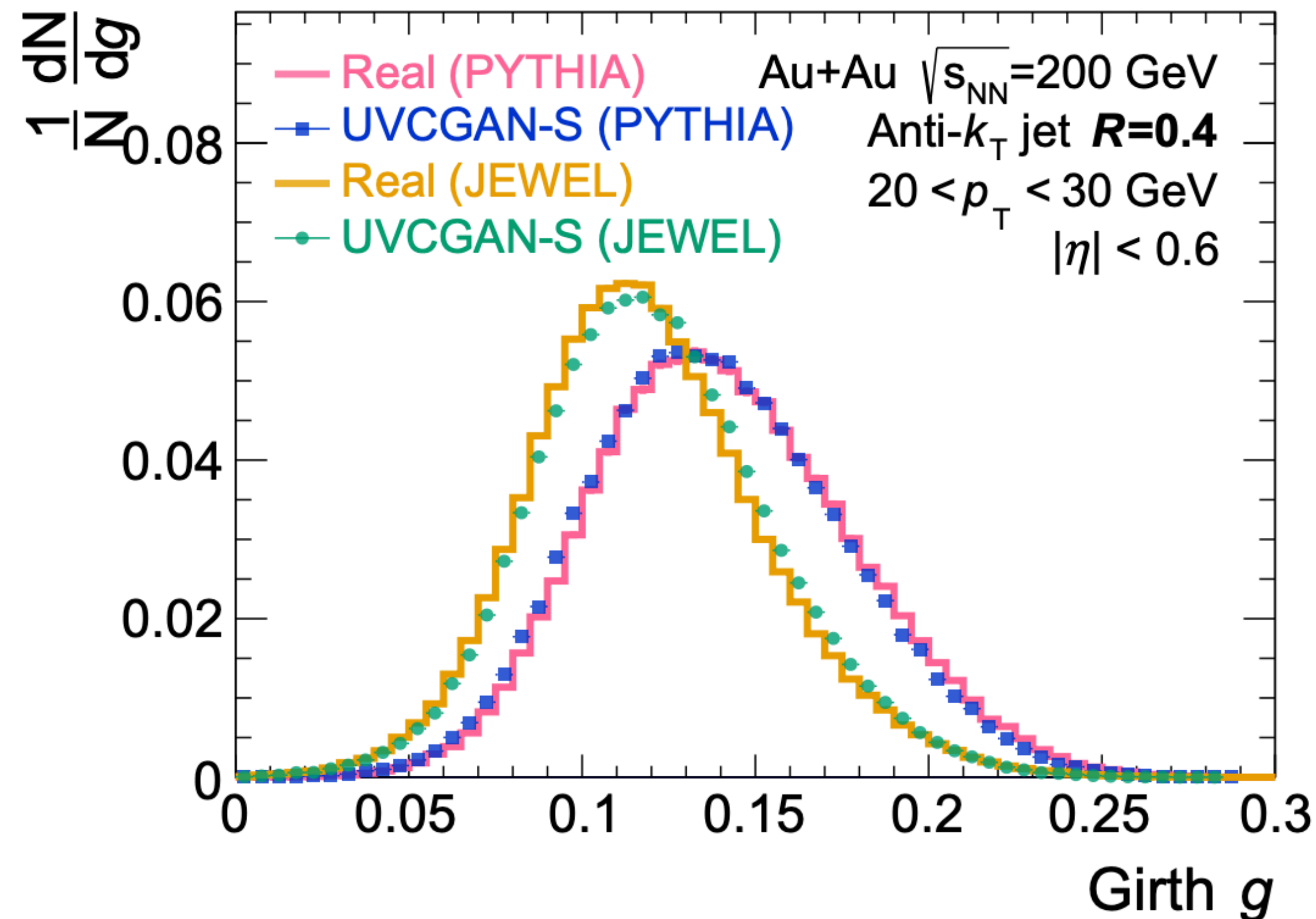
[arXiv:2510.23717](https://arxiv.org/abs/2510.23717)

UVCGAN-S: Application on Quenched Jets

- UVCGAN-S trained on **PYTHIA** (unquenched jets) and tested on **JEWEL** (quenched jets)
 - ➔ *significant difference in substructure distributions for quenched jets*
 - ➔ **UVCGAN-S** successfully reproduce the **JEWEL**'s distribution

**Robust Generalizability of the
UVCGAN-S for out-of-distribution handling!**

[arXiv:2510.23717](https://arxiv.org/abs/2510.23717)

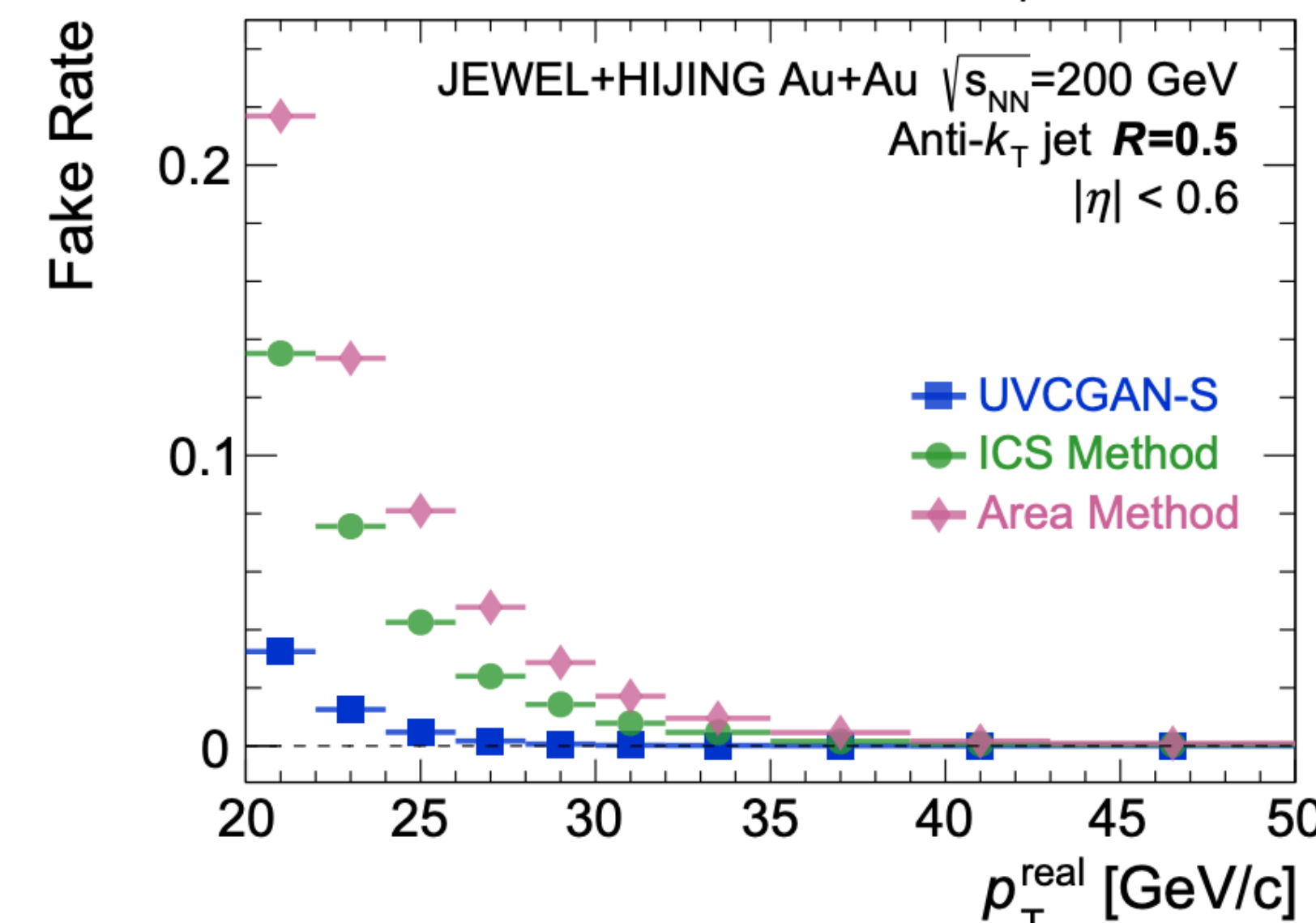
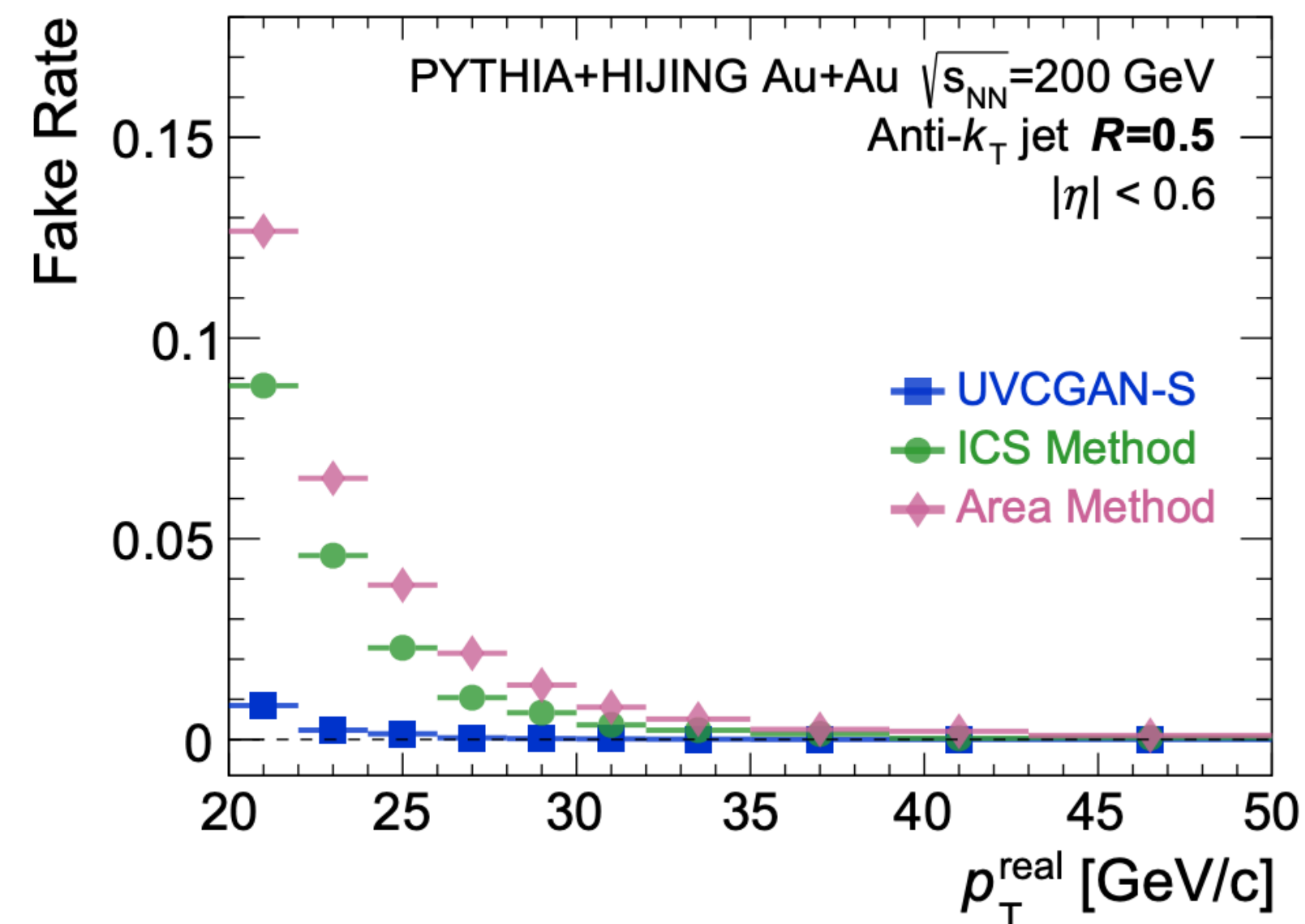
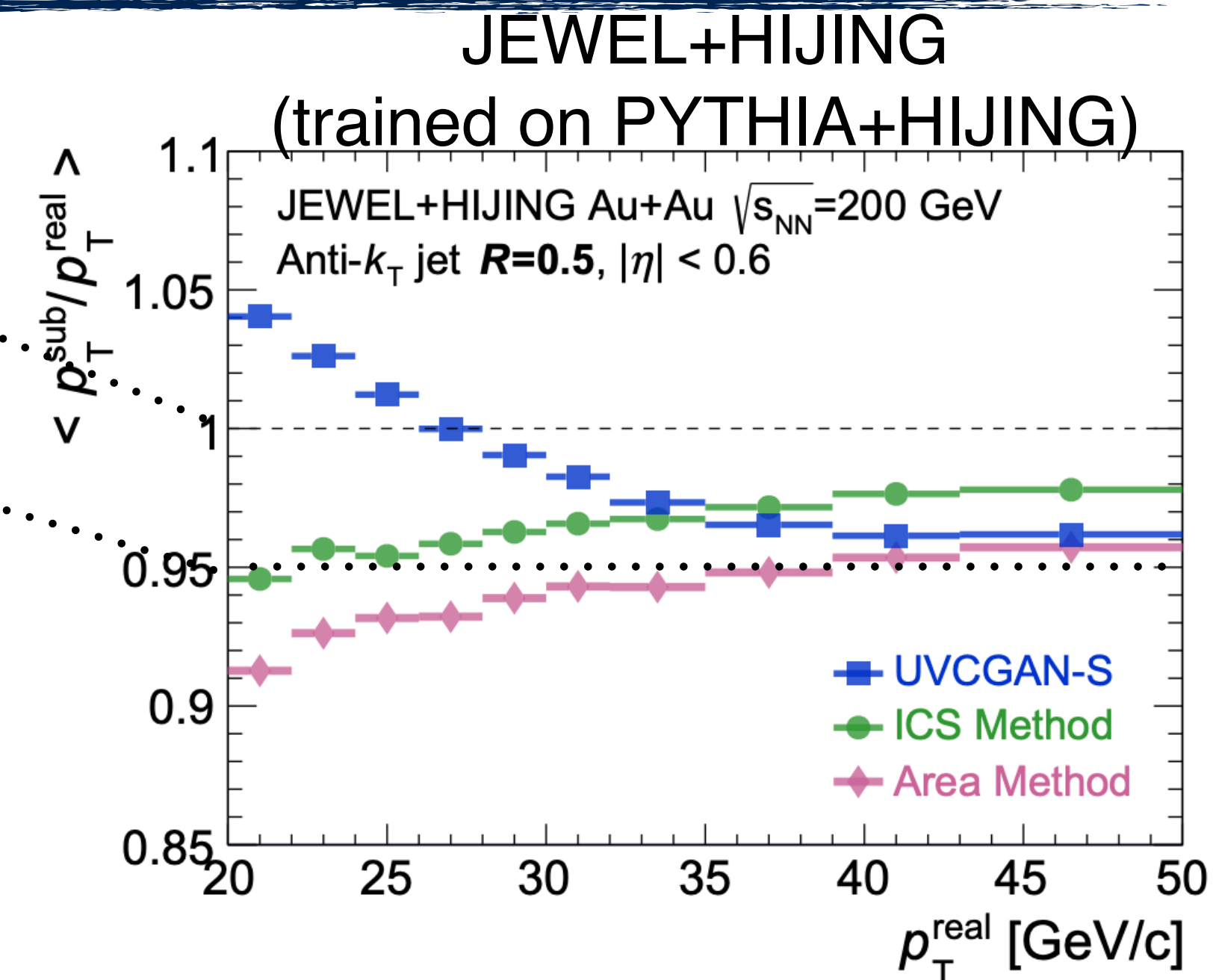
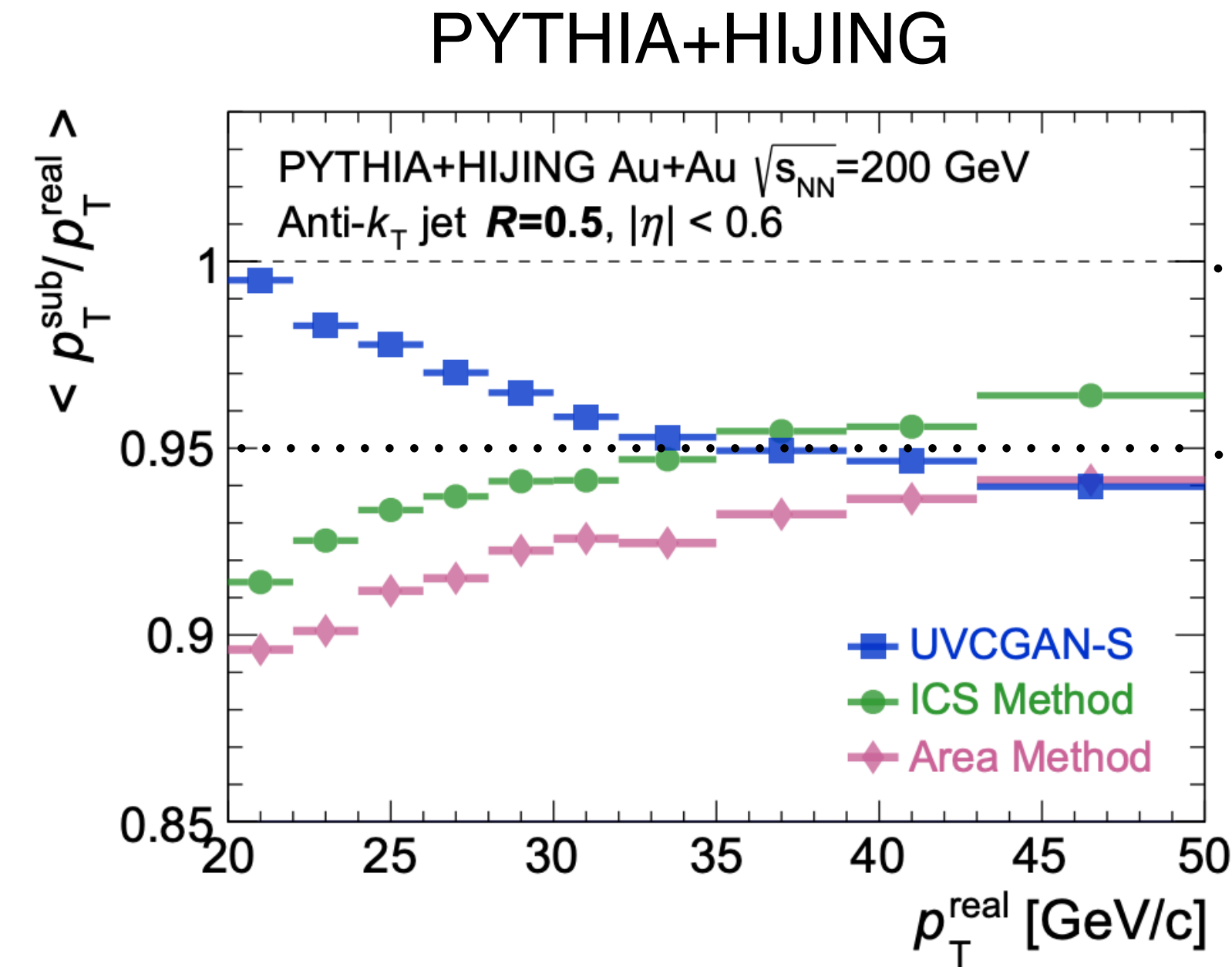


Discussion: Generalizability of UVCGAN-S

- **Supervised** predictors trained on a *single domain* risk **inheriting biases** when the internal structure differs between training and application
- Background subtraction is an **emergent property** of UVCGAN-S:
 - ➔ the network *does not simply learn the average statistical properties* of the training sample but rather successfully separates the signal from the background regardless of the jet's properties.
 - ➔ the reconstruction relies on *correlations learned from the surrounding image features, minimizing the bias applied to the reconstructed jet core.*

Discussion: Correction and Systematic Uncertainty

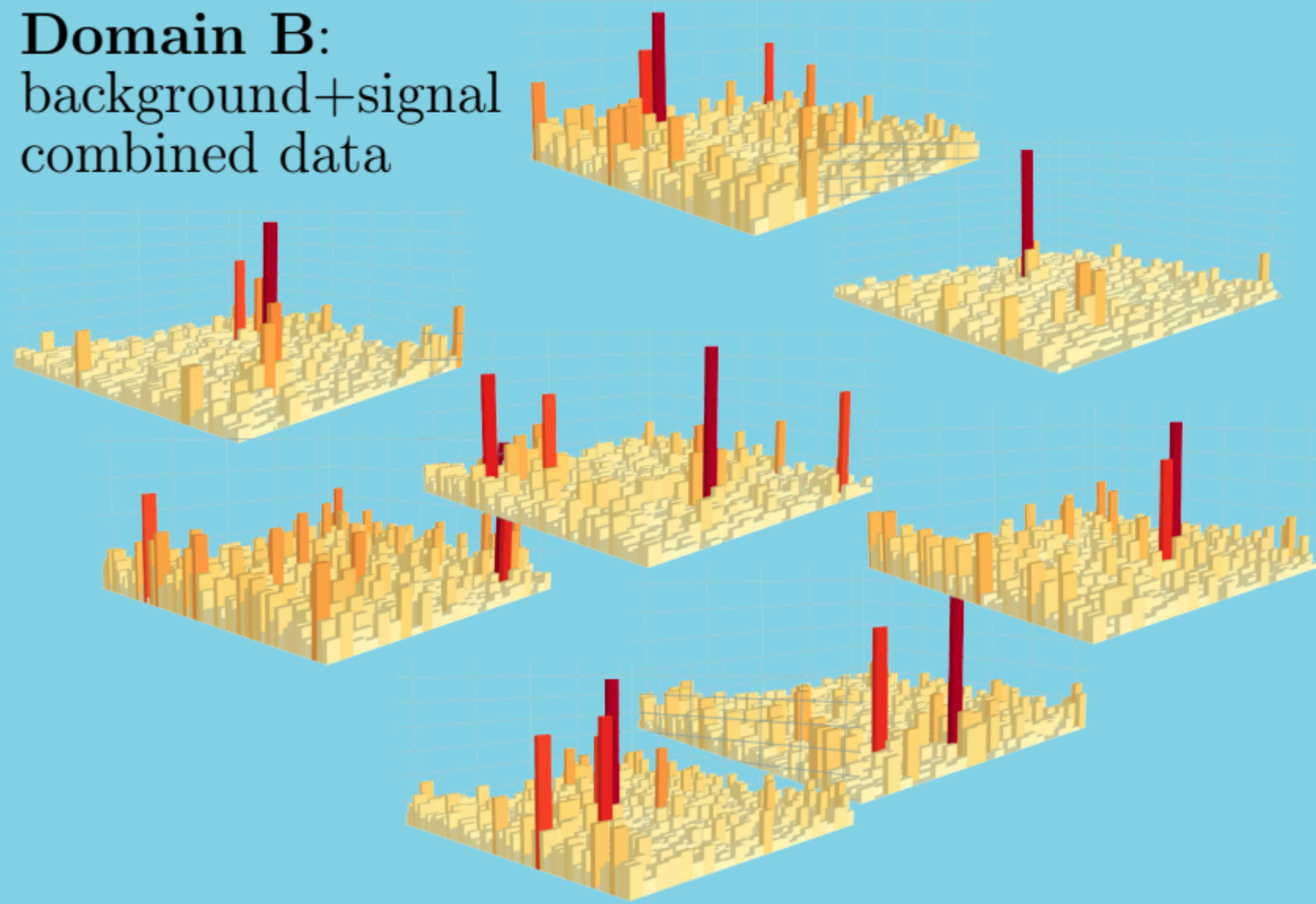
- Remaining systematic non-closure (uncertainties/correction) should be quantified and corrected
- ➔ test on labeled simulation samples (e.g. both quenched and unquenched) to quantify residual effects



UVCGAN-S: Even More Realistic and No Bias

- Ultimately, use **experimental data** to train the model !
 - ➔ PYTHIA → jet events in p+p collisions
 - ➔ HIJING → minimum-bias events in A+A collisions
 - ➔ PYTHIA+HIJING → jet events in A+A collisions

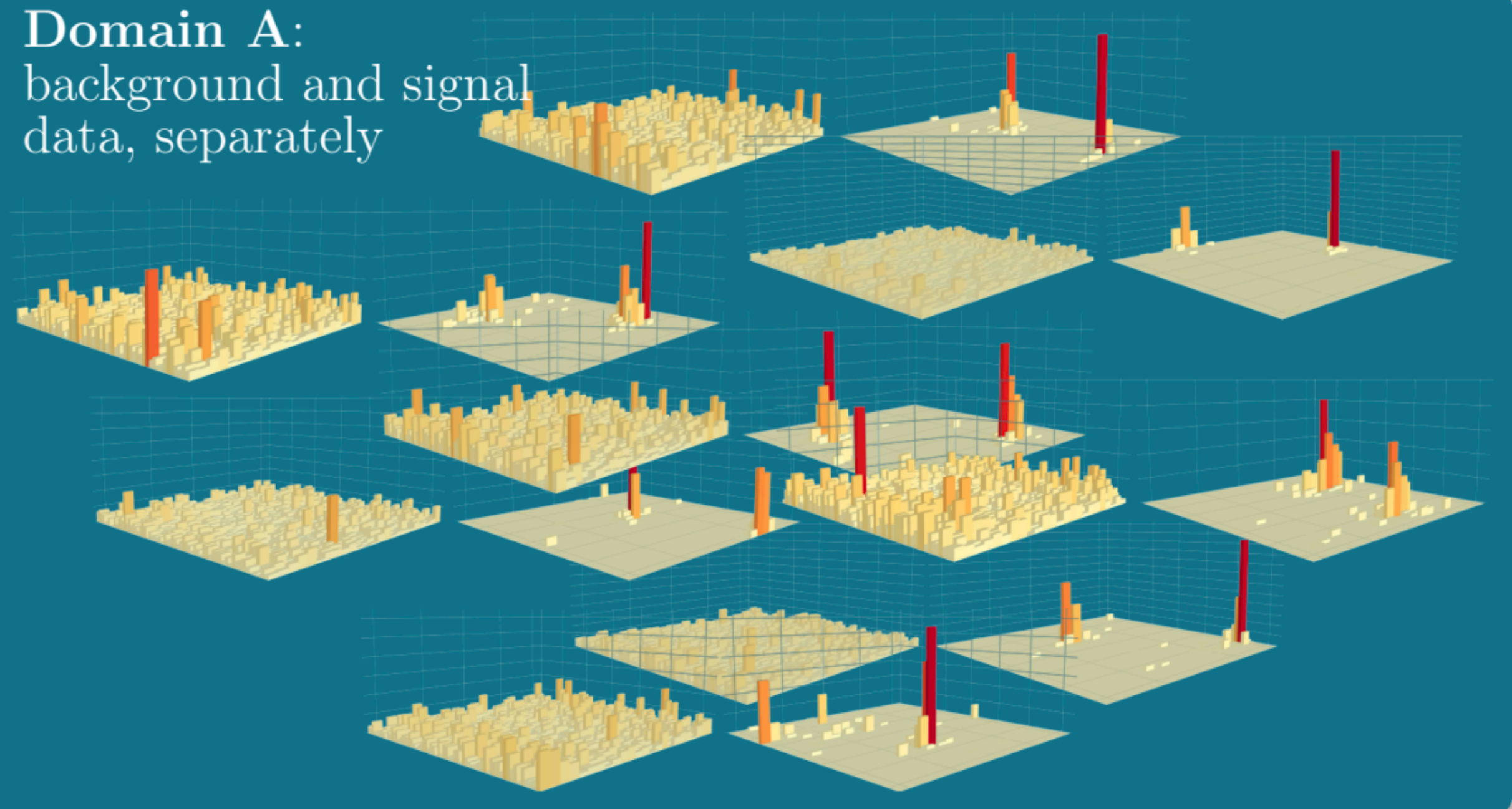
Domain B:
background+signal
combined data



UVCGAN-S
Neural
translators

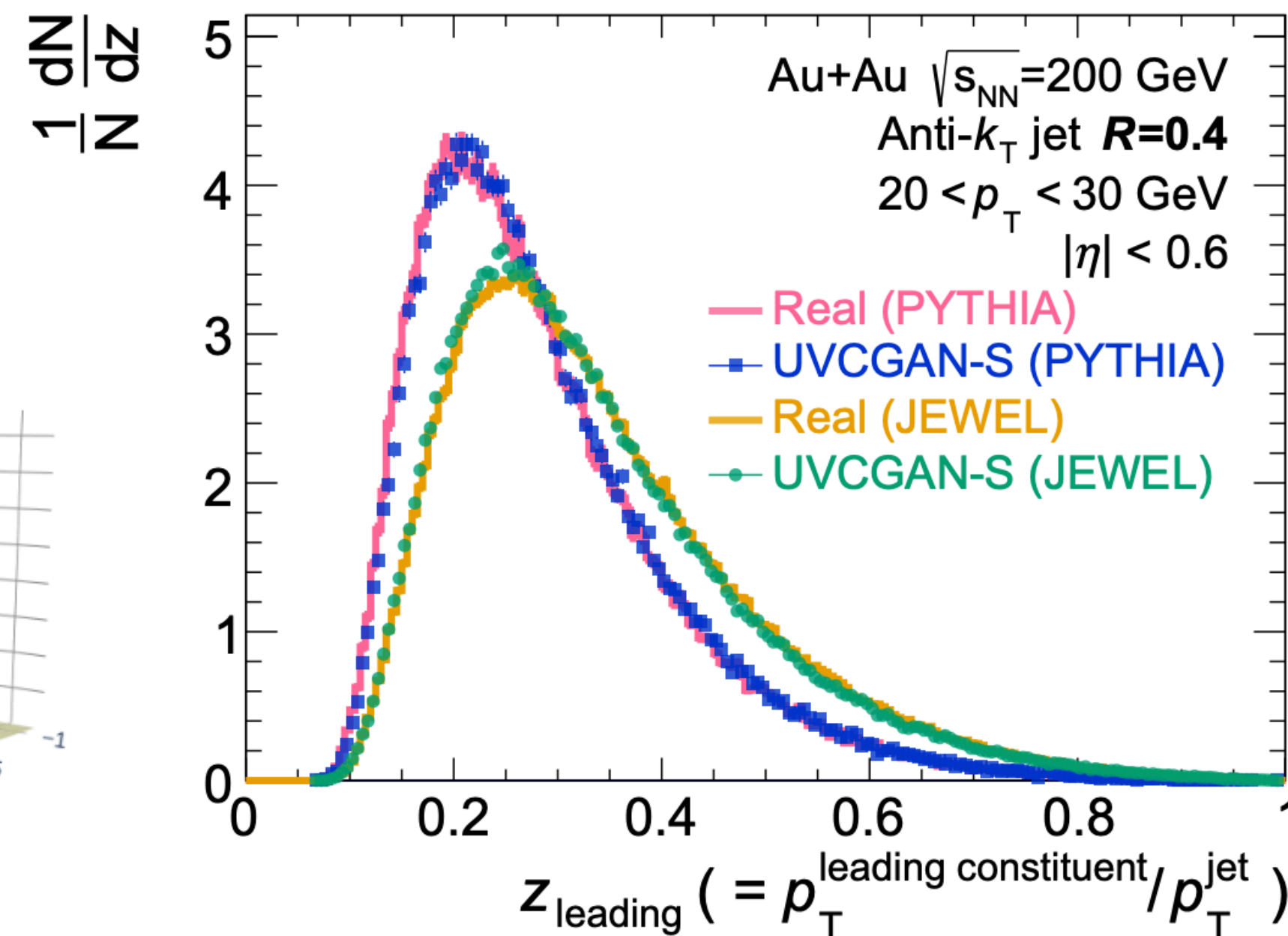
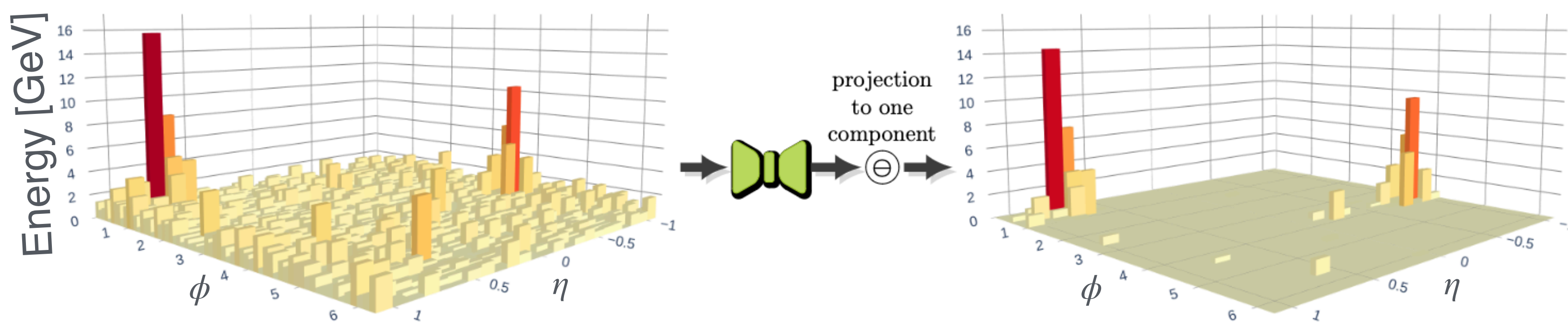


Domain A:
background and signal
data, separately



Summary

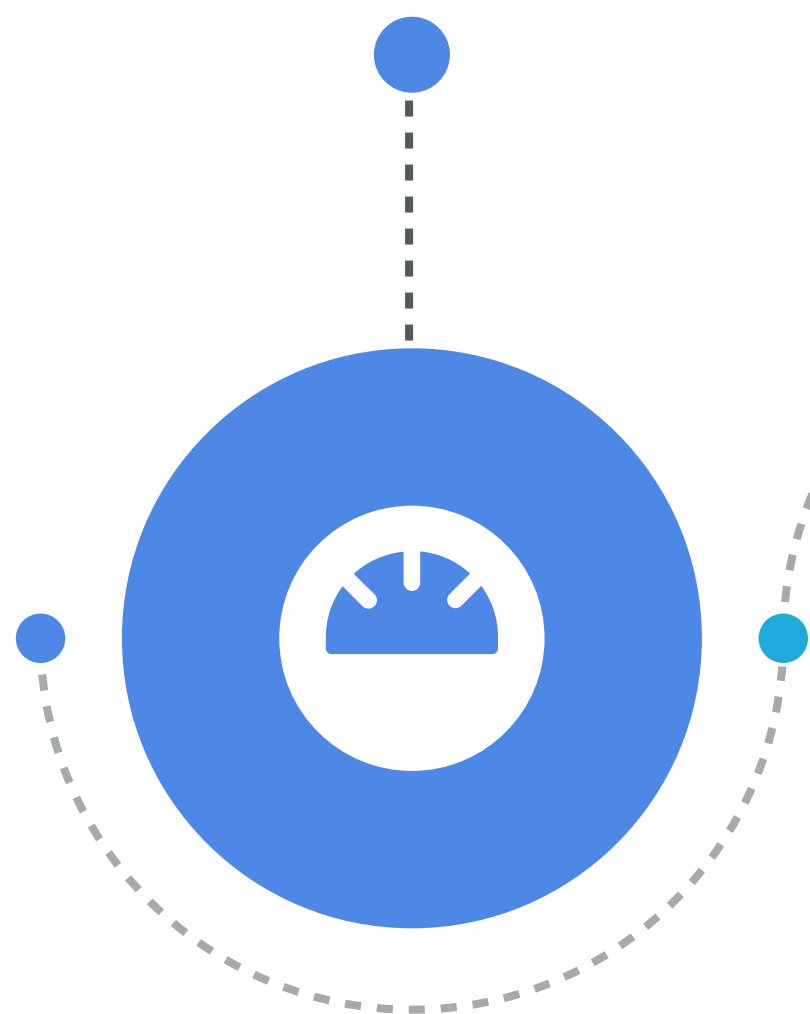
- Active developments in progress using machine learning for jet physics
 - ➔ *More challenging in heavy ion collisions and caution needed* [arXiv:2510.23717](https://arxiv.org/abs/2510.23717)
- **UVCGAN-S**, *unsupervised and unpaired generative AI learning*, paves a new way for precision measurements for jet background subtraction in heavy ion collisions
 - ➔ provides **superior accuracy** compared to conventional subtraction methods (Iterative Constituent Subtraction and Area-based Method), *especially for large jet radii ($R = 0.5$) and lower jet p_T* , the region experimentally difficult to reach
 - ➔ **robust generalizability** for out-of-distribution handling!
 - ➔ can be applied whenever signal and background superpose in an image-like representation



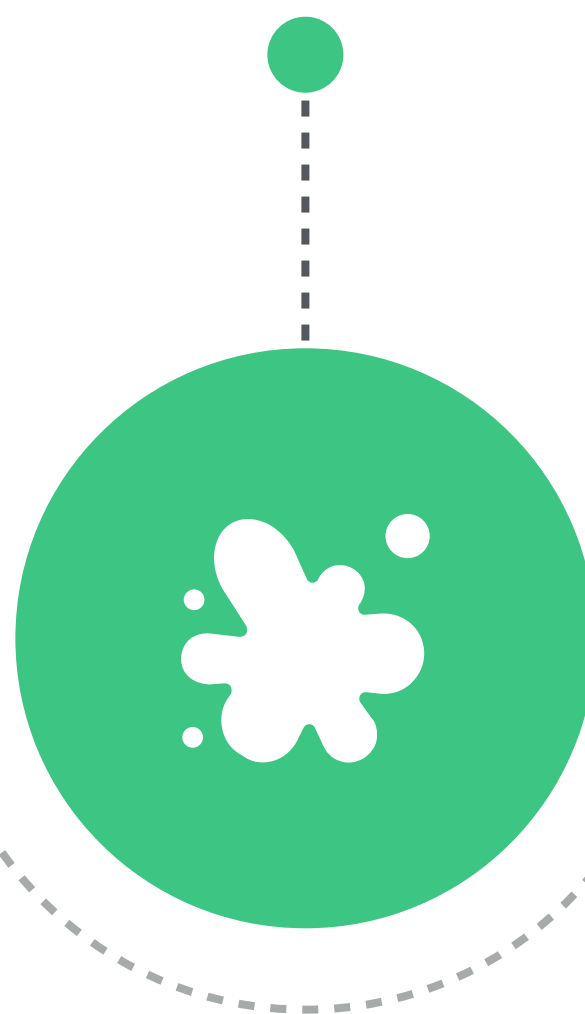
Backup

AI/ML Applications in the QCD Field

Detector Control,
Calibration



Event Selection,
Triggering



Reconstruction,
Analysis



Data Compression

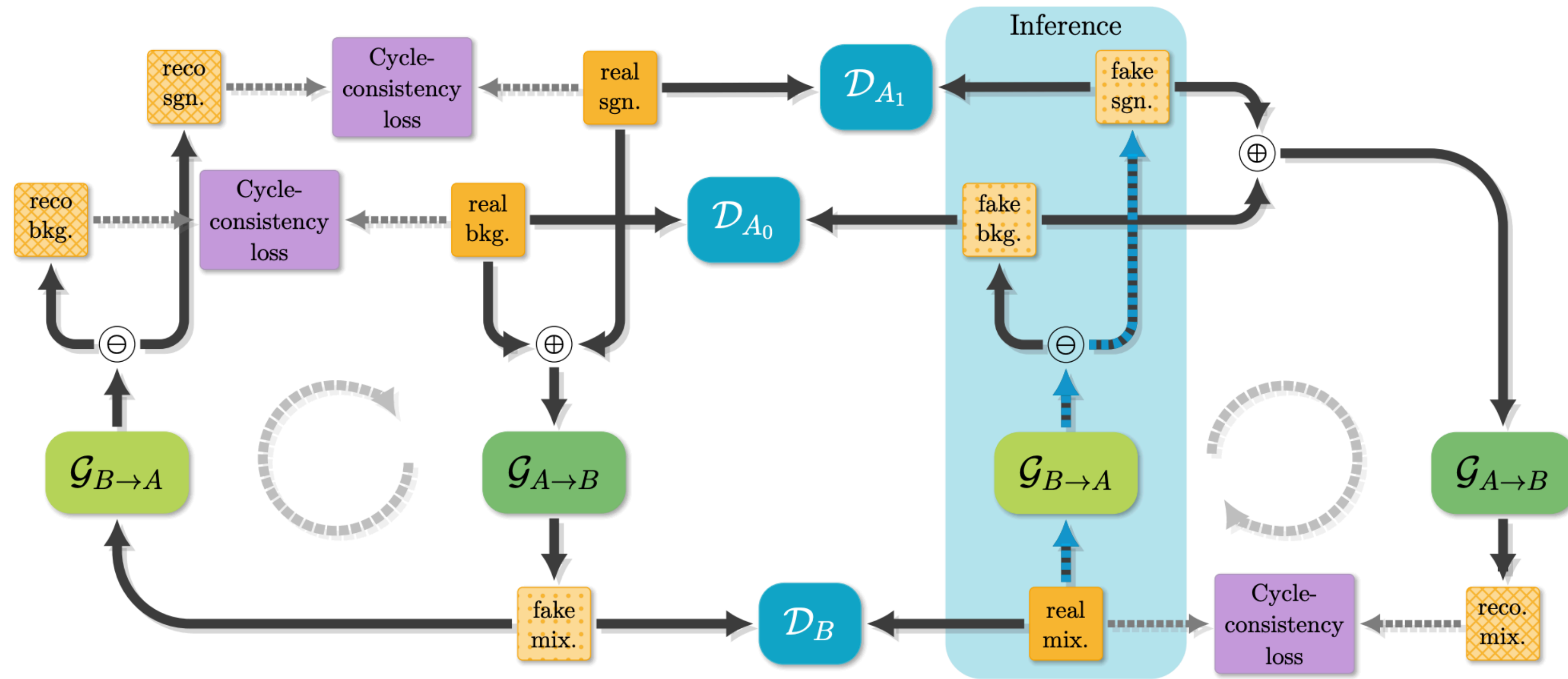


Simulation

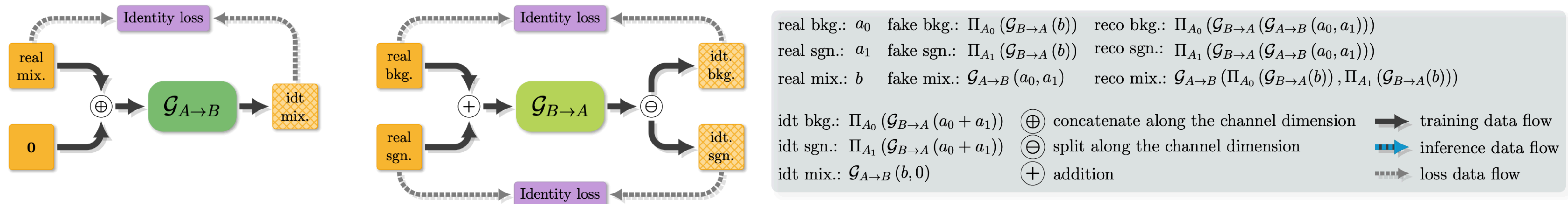


UVCGAN-S Architecture

(a) UVCGAN-S Architecture (identity losses shown separately below)

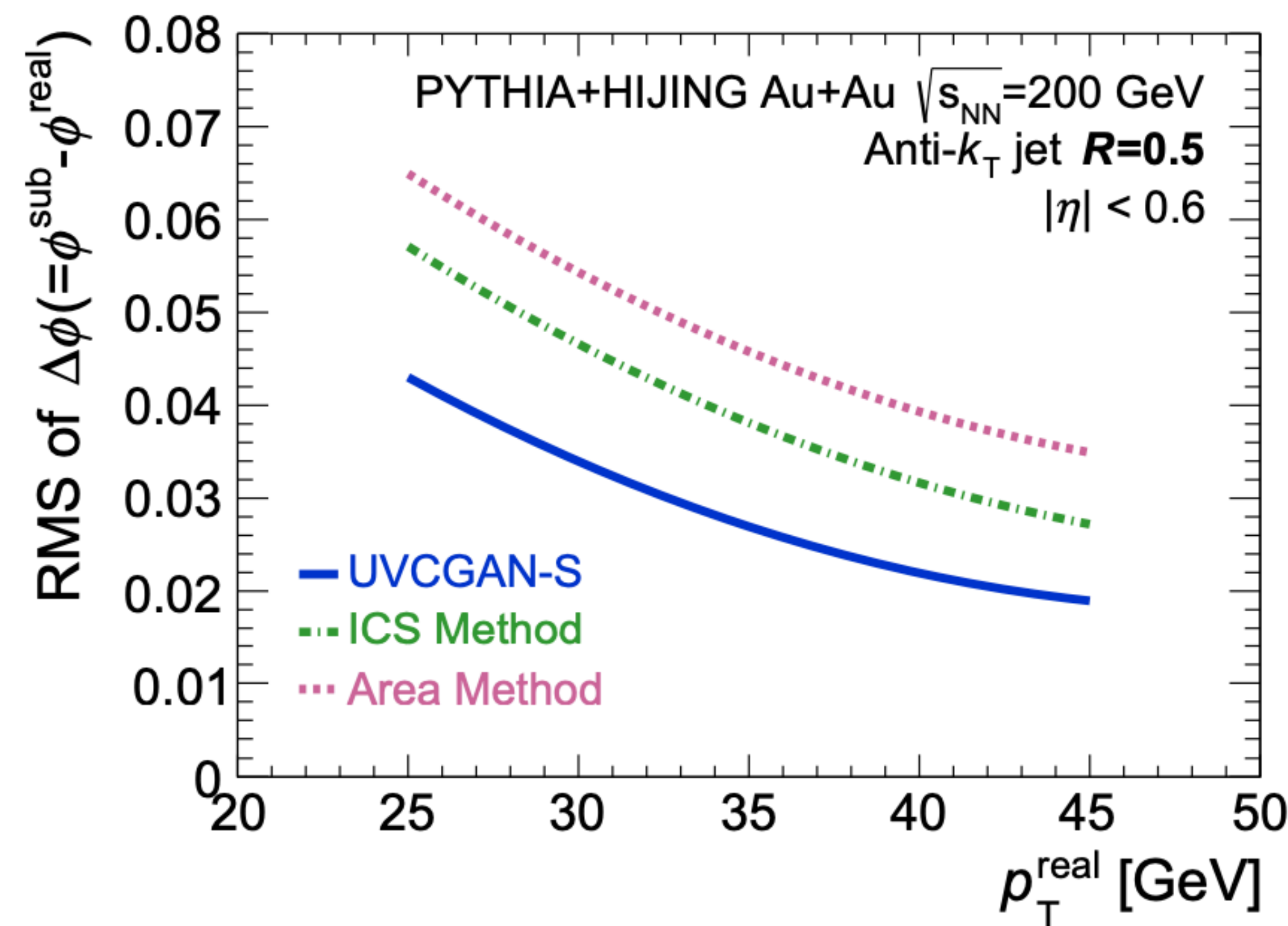
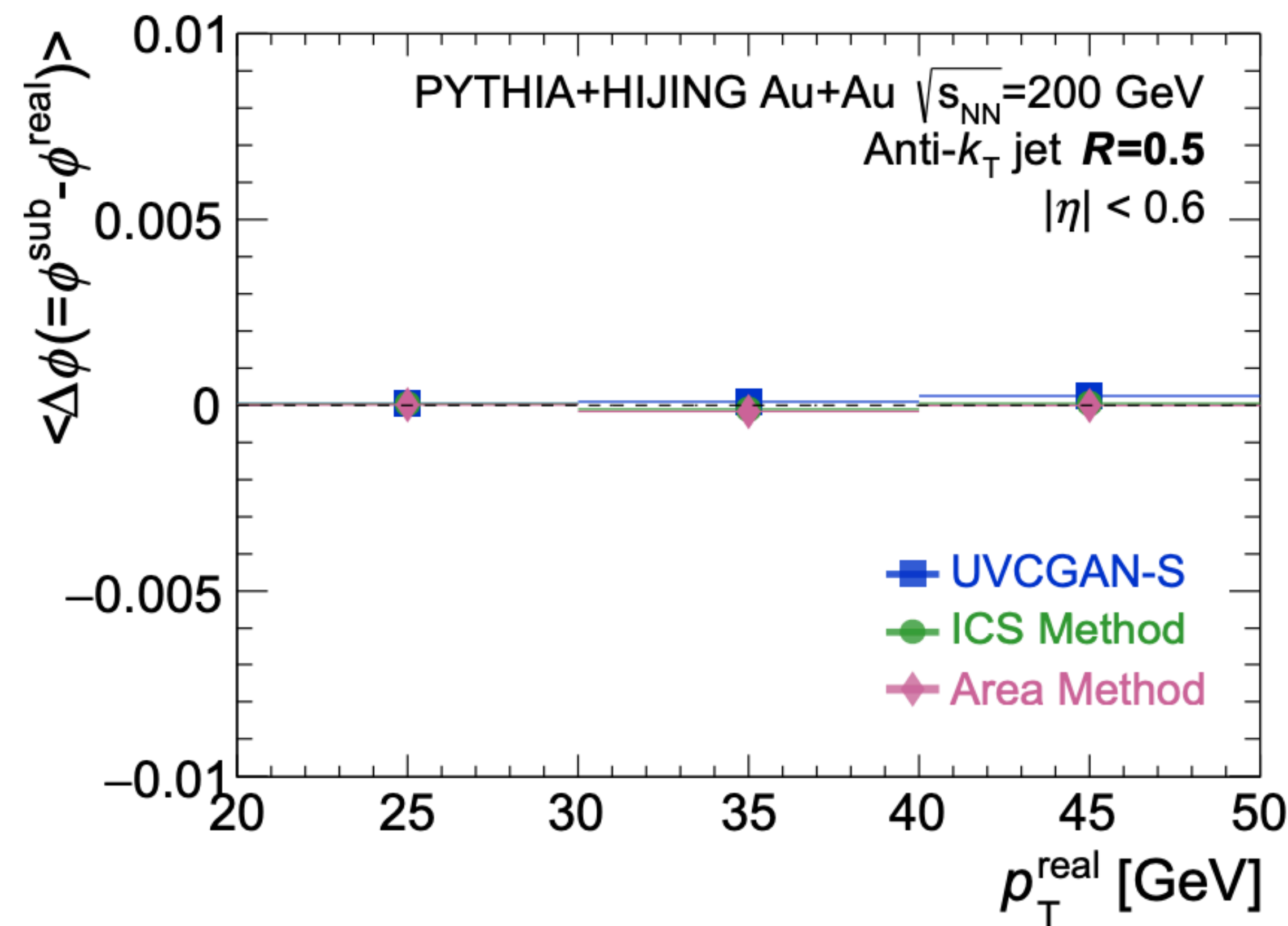


(b) Identity Losses



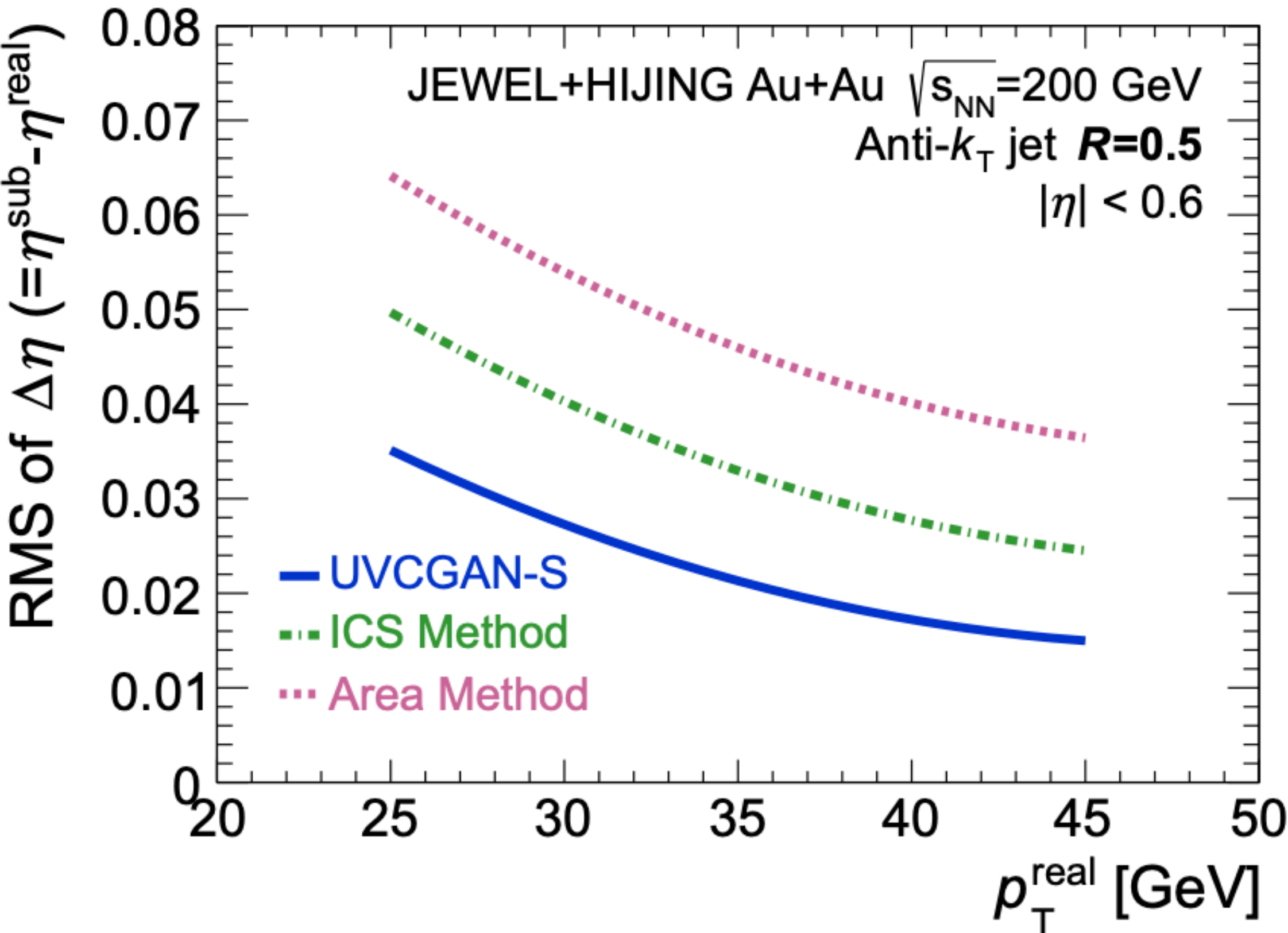
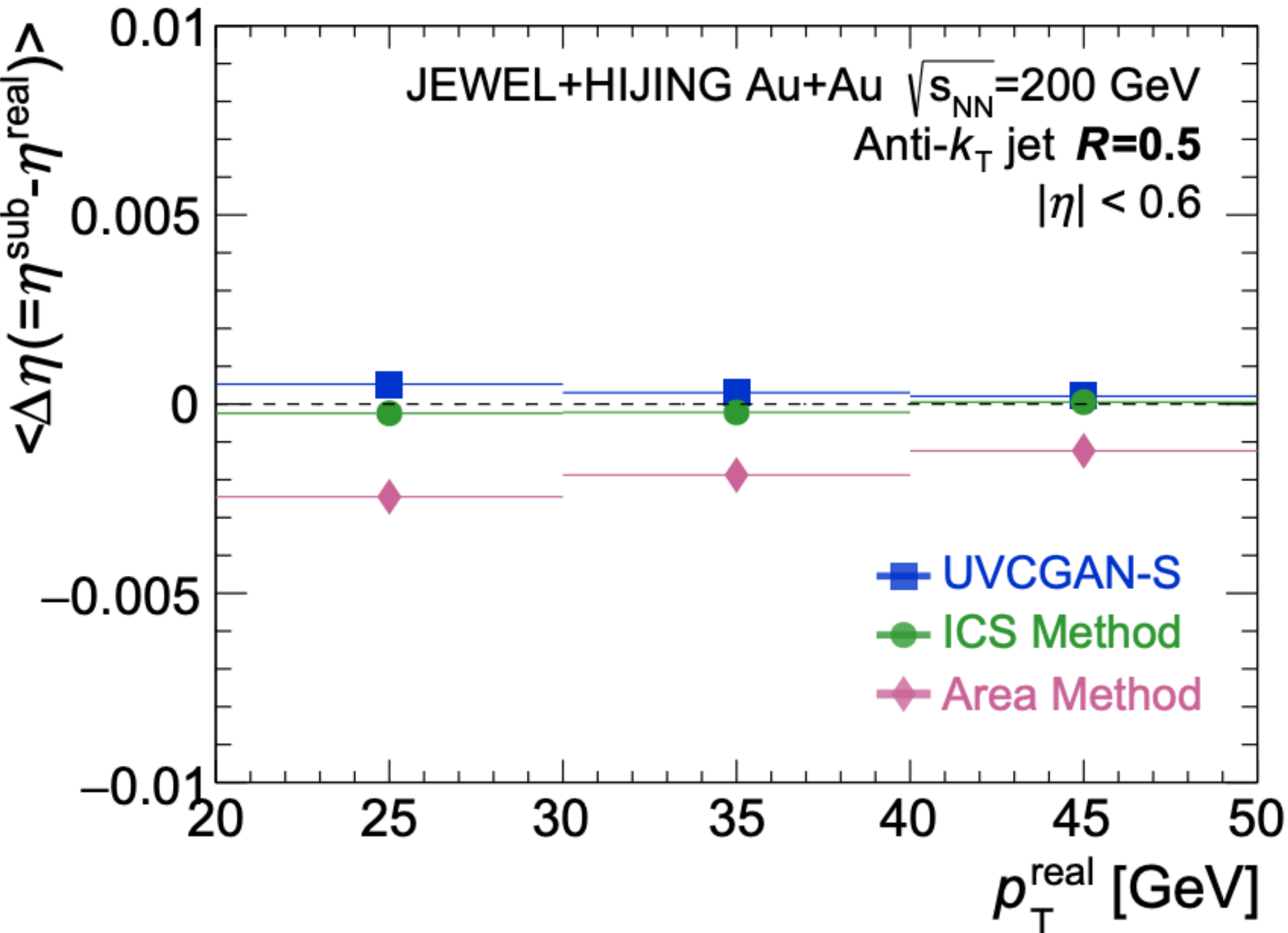
UVCGAN-S: $\Delta\phi$ (PYTHIA)

- PYTHIA+HIJING Anti- k_T $R=0.5$ jets
- $\Delta\phi$



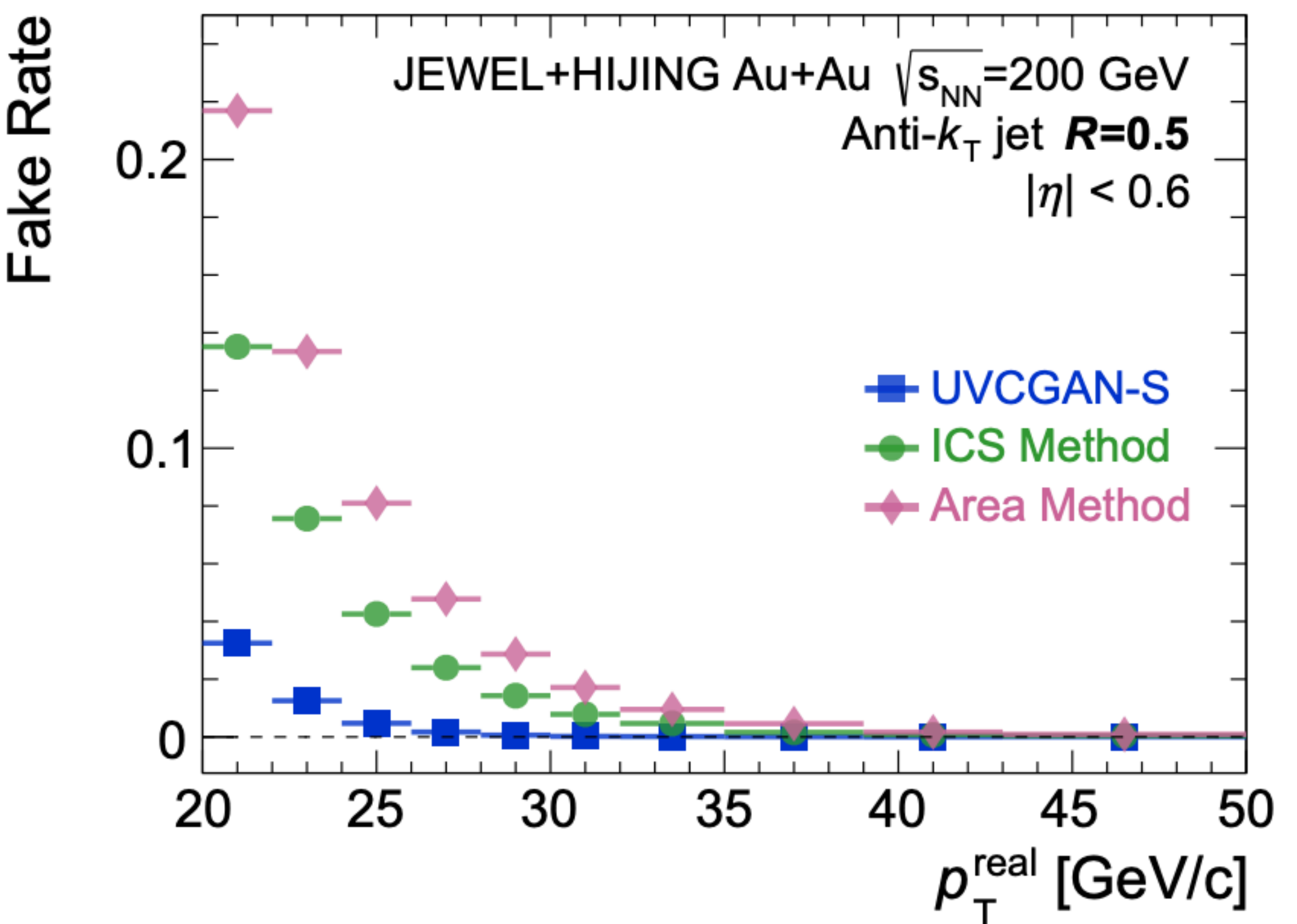
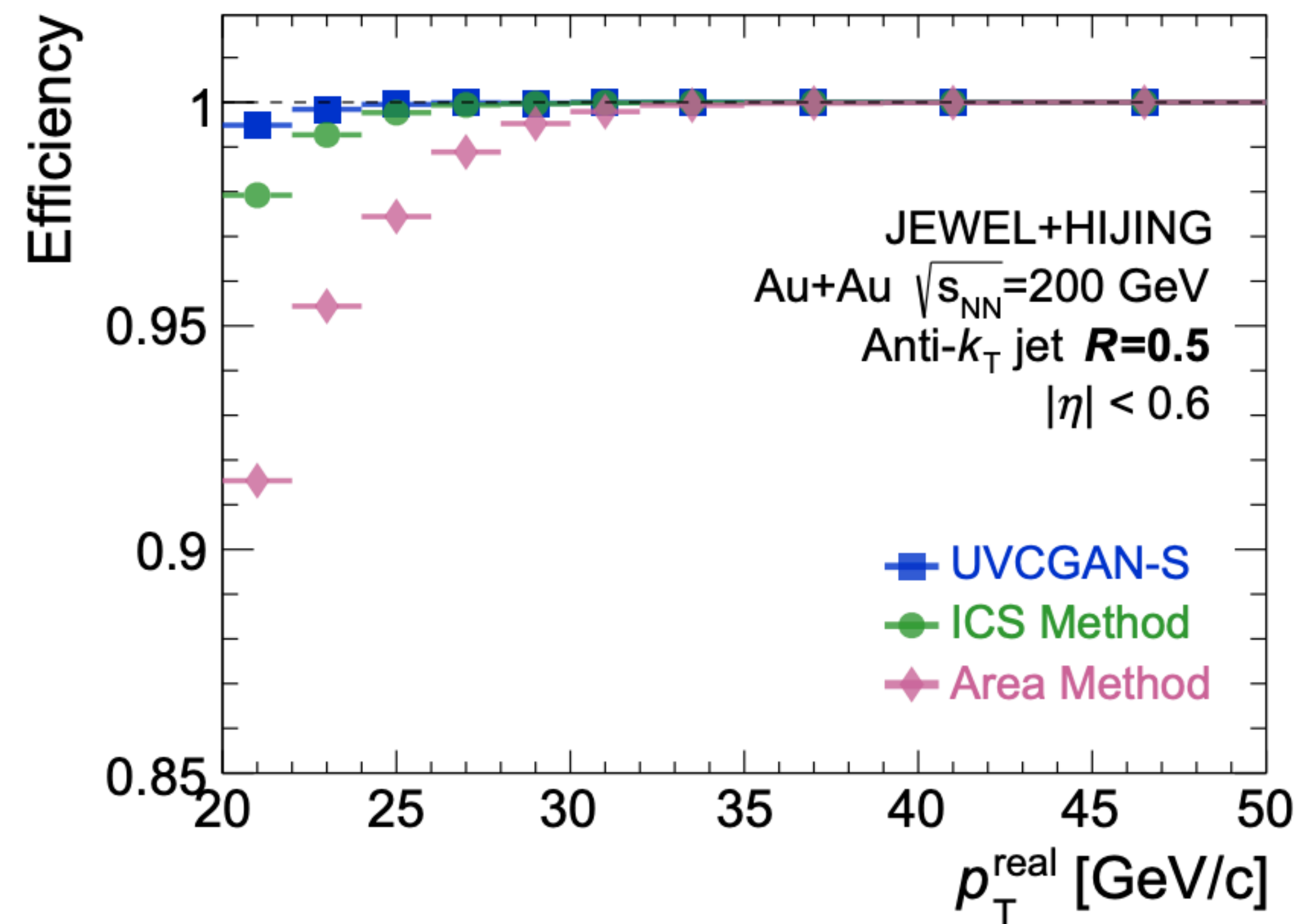
UVCGAN-S: $\Delta\eta$ (JEWEL)

- JEWEL+HIJING Anti- k_T $R=0.5$ jets
- $\Delta\eta$



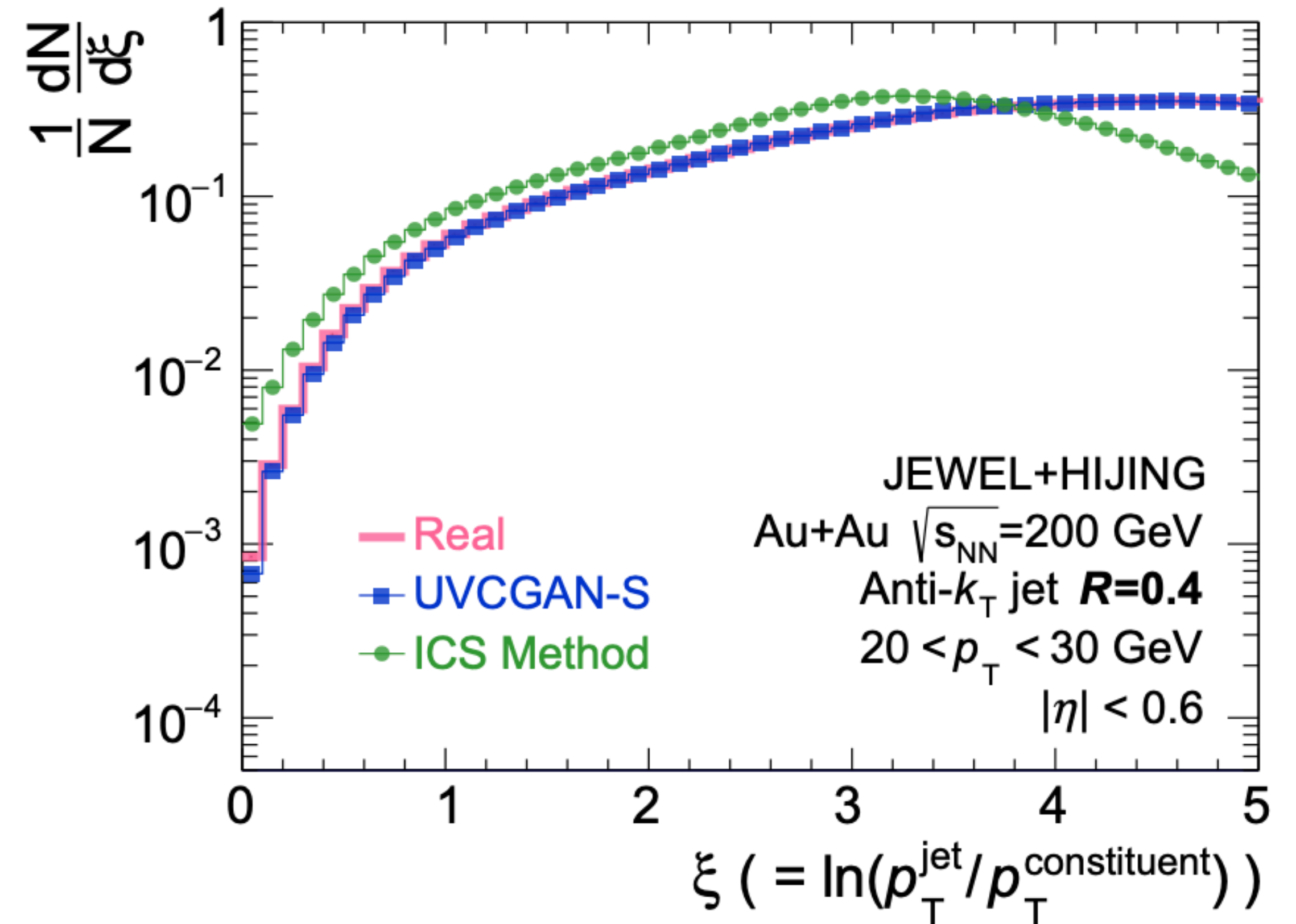
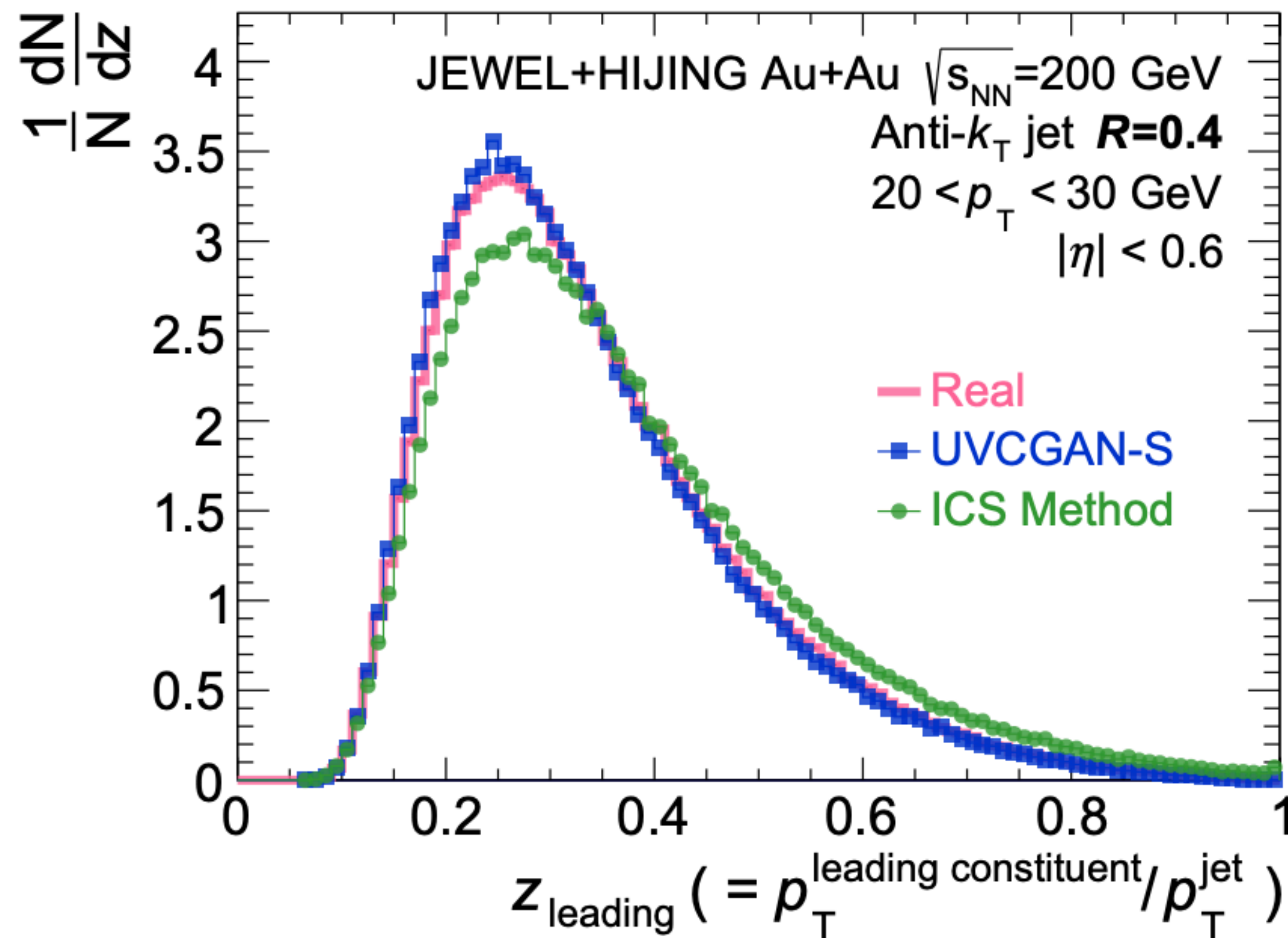
UVCGAN-S: Efficiency and Fake Rate (JEWEL)

- JEWEL+HIJING Anti- k_T $R=0.5$ jets
- Efficiency and Fake Rate



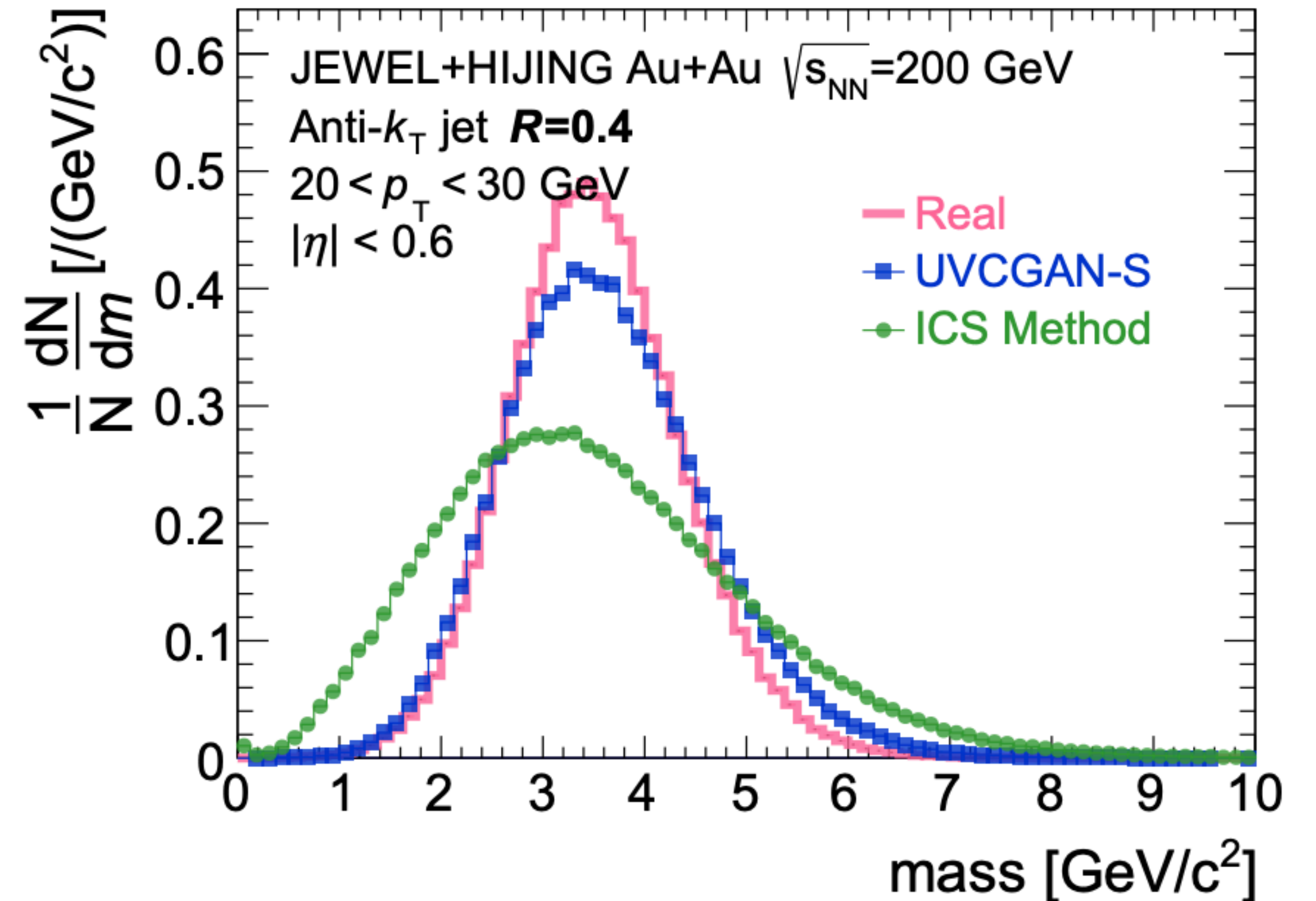
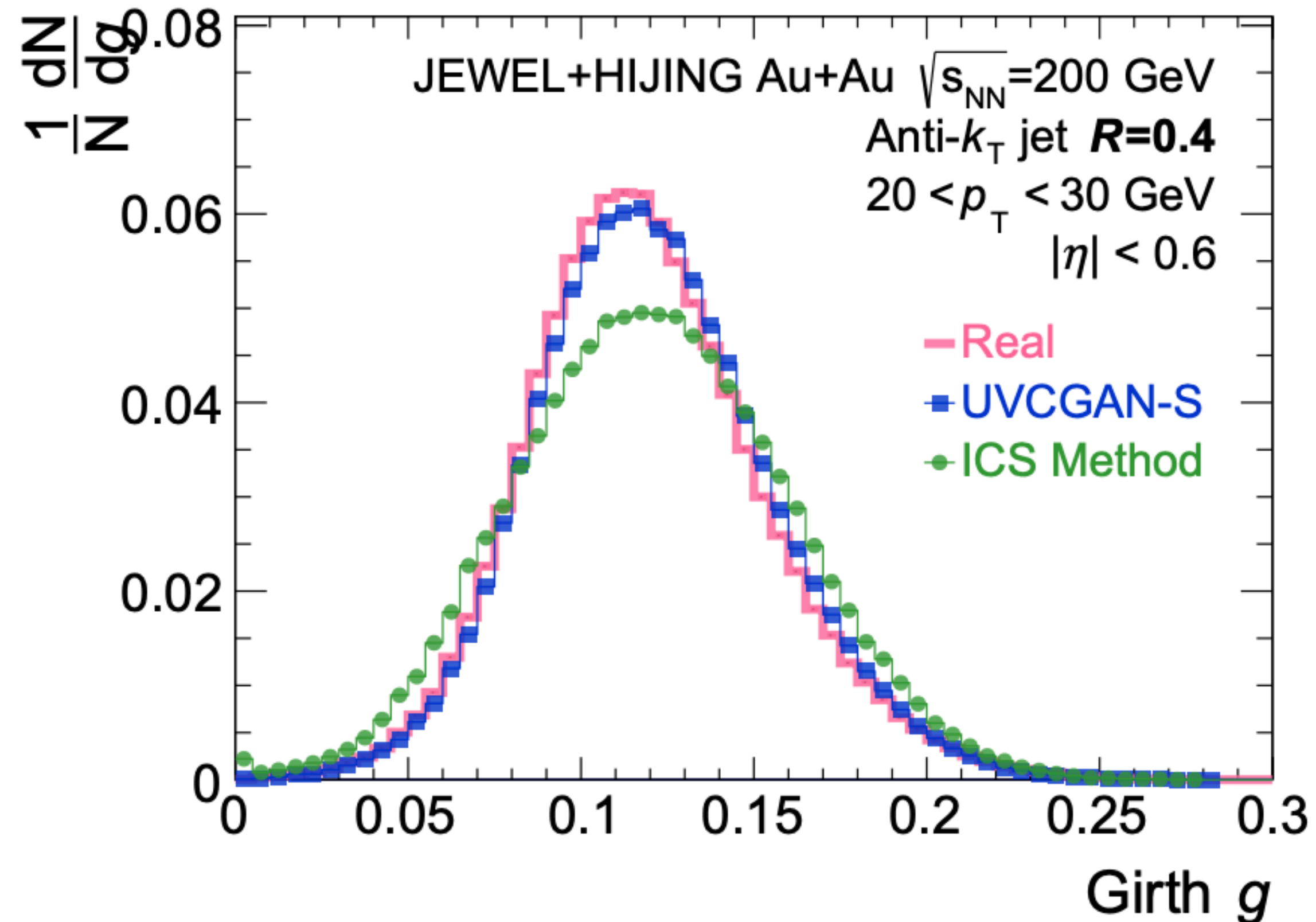
UVCGAN-S: Jet substructure (JEWEL)

- JEWEL+HIJING Anti- k_T $R=0.5$ jets
- jet substructure



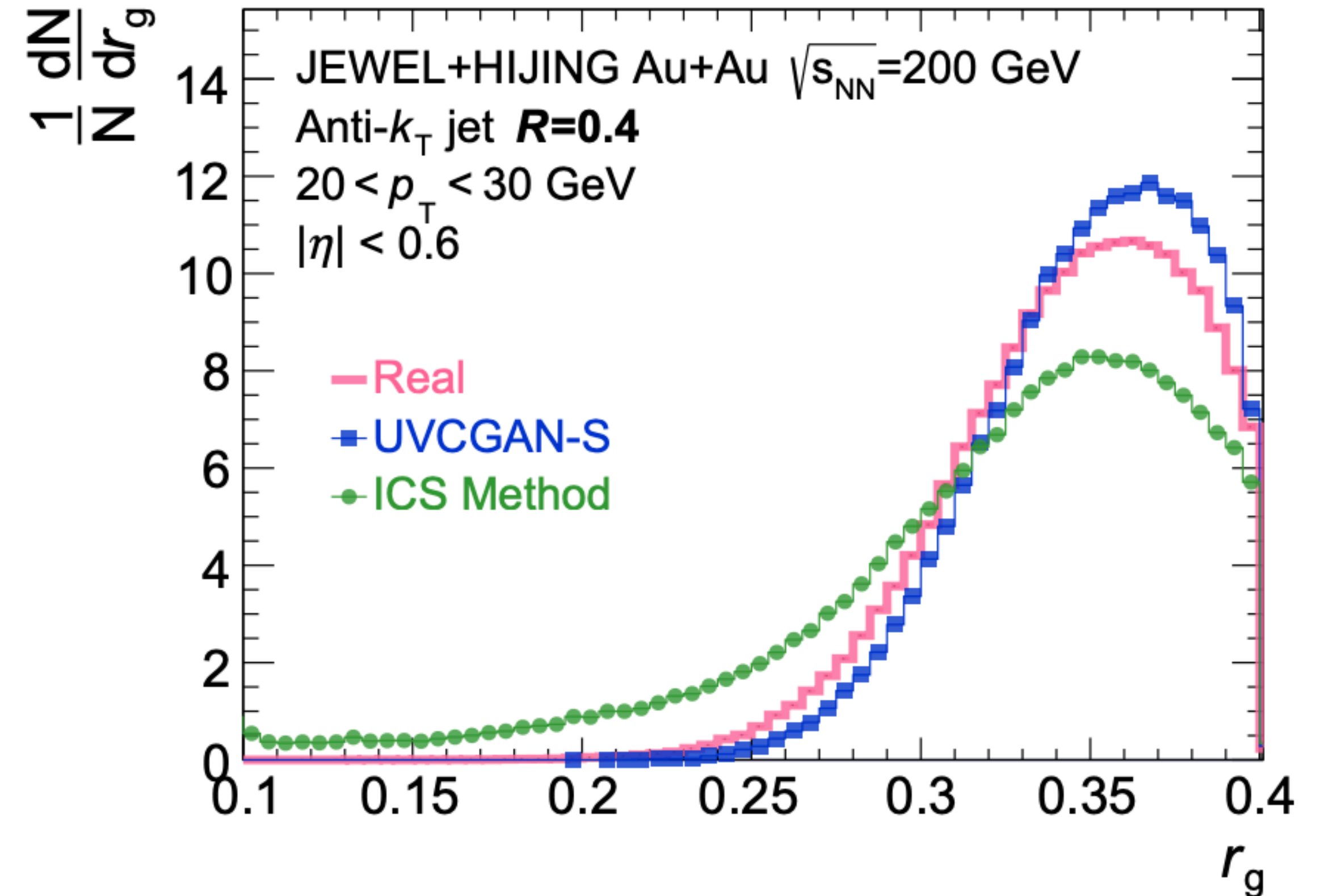
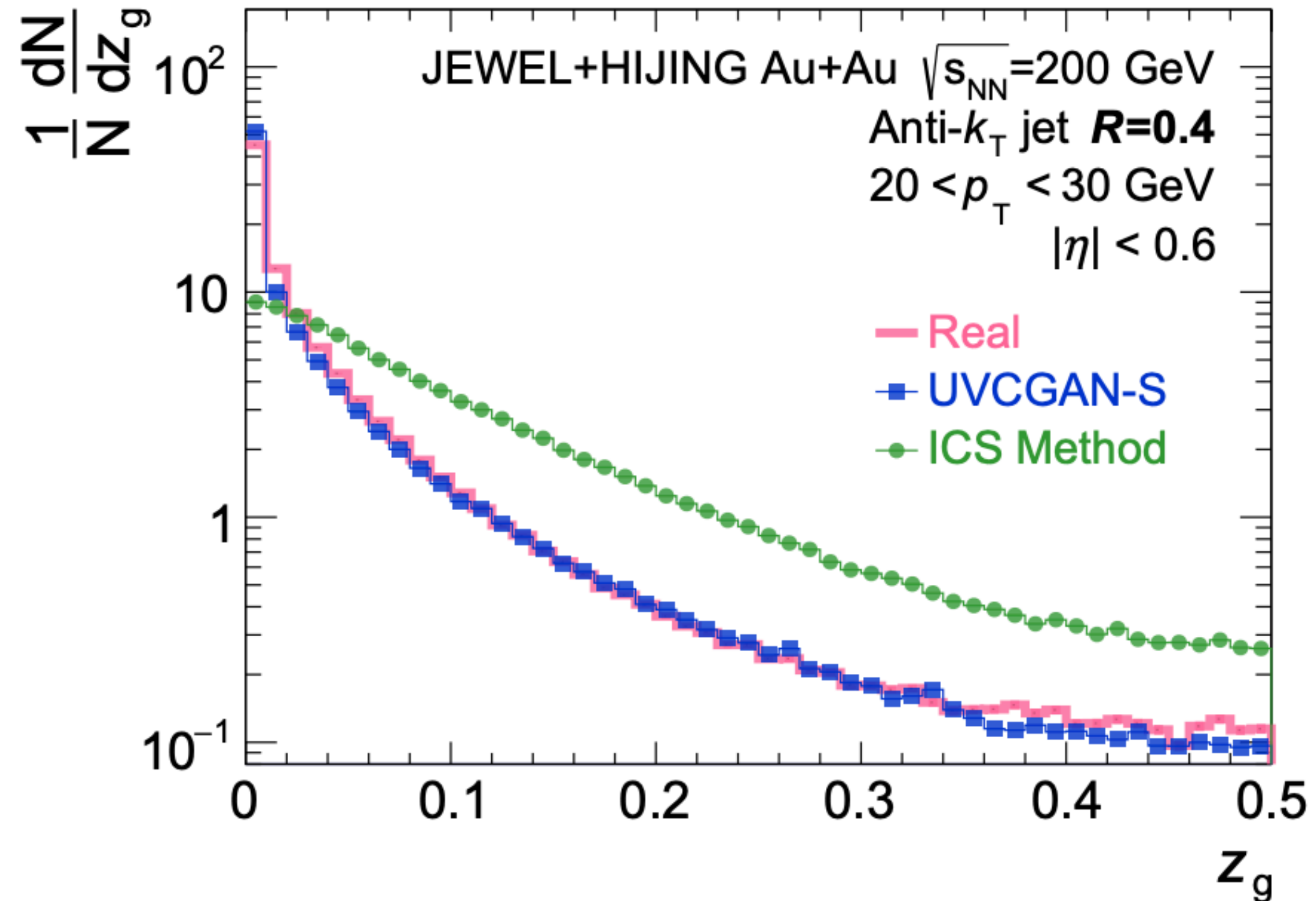
UVCGAN-S: Jet substructure (JEWEL)

- JEWEL+HIJING Anti- k_T $R=0.5$ jets
- jet substructure



UVCGAN-S: Jet substructure (JEWEL)

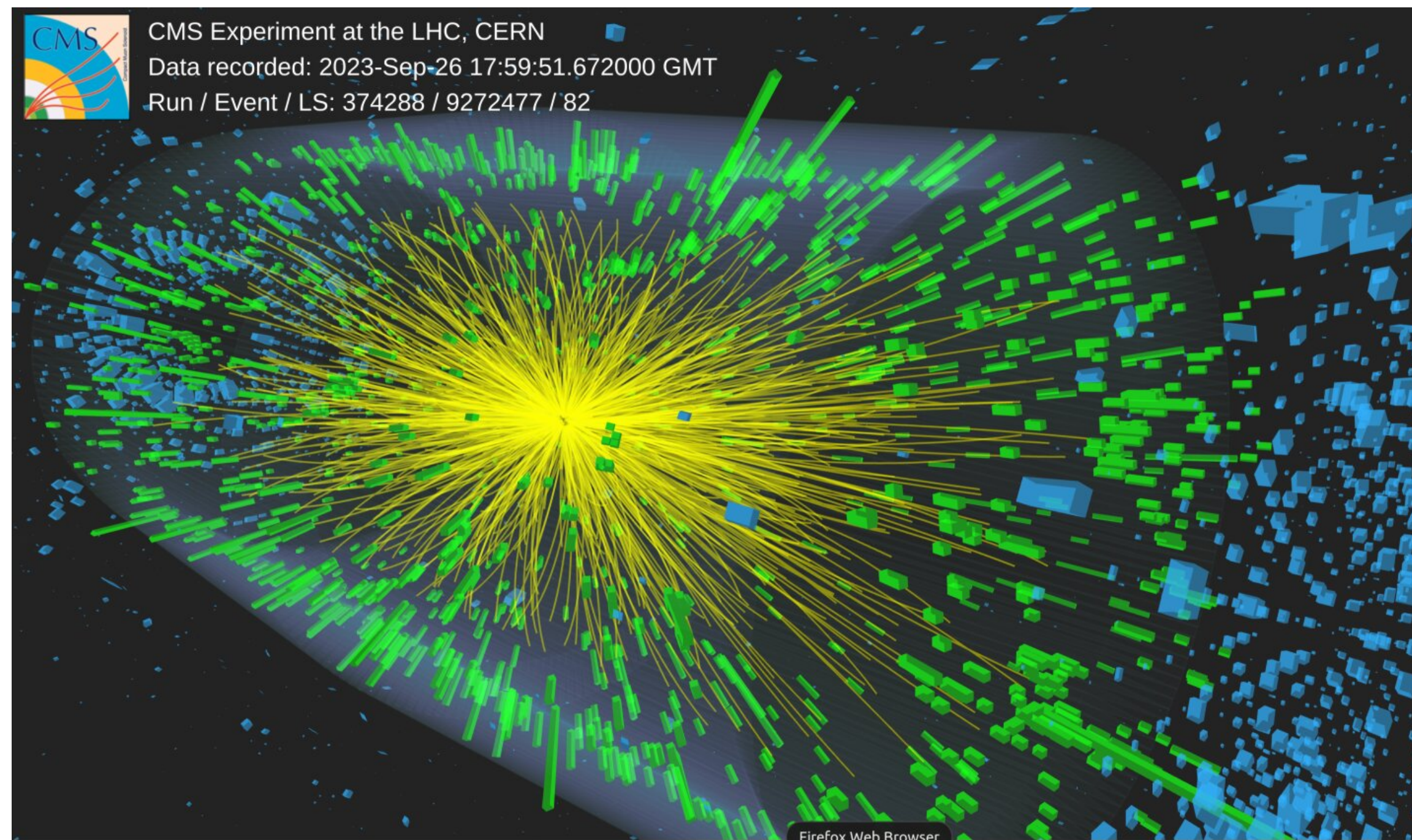
- JEWEL+HIJING Anti- k_T $R=0.5$ jets
- jet substructure



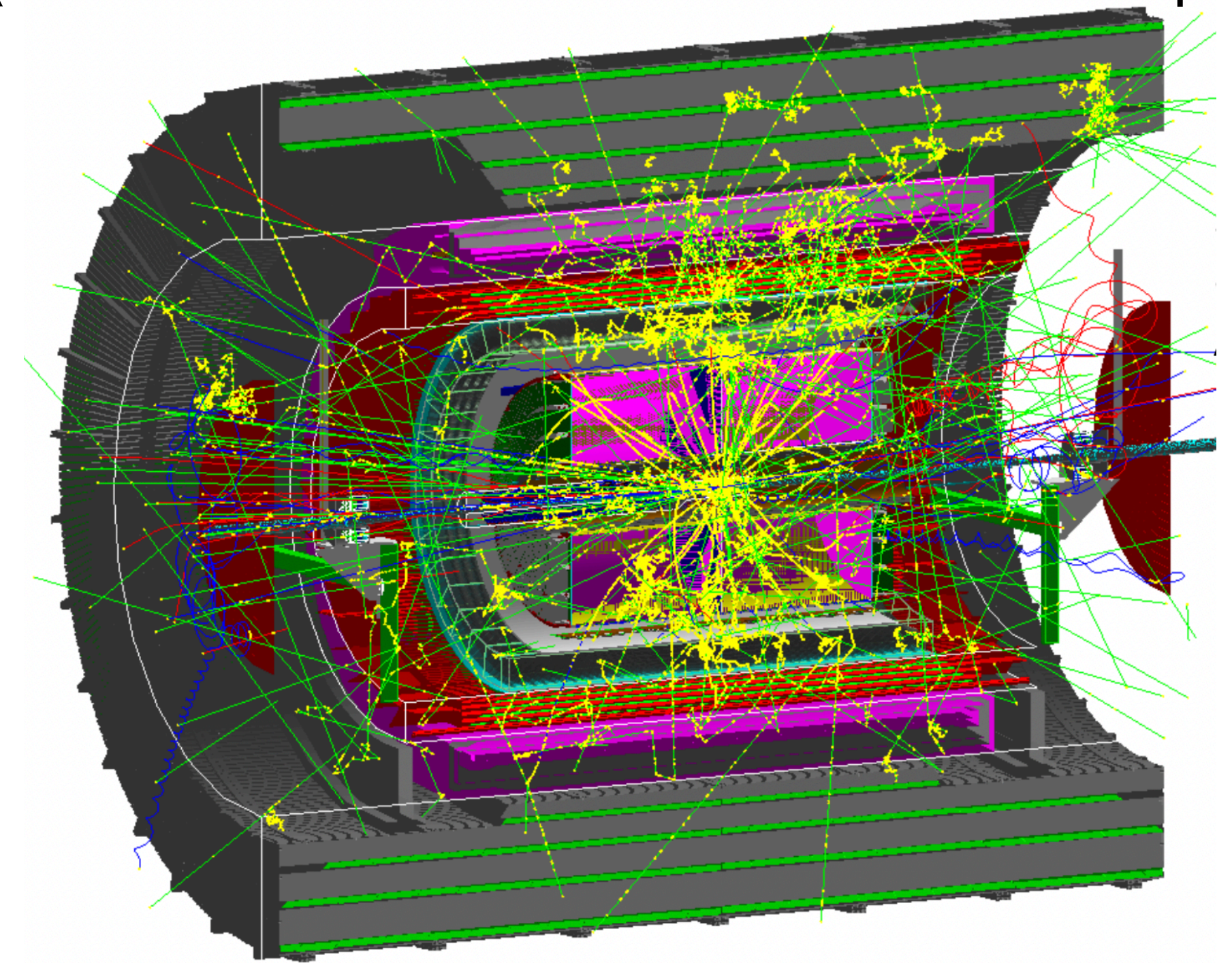
High Energy Heavy Ion Collisions

- $O(1000)$ particles in one nuclear collision event + *thousands shower steps* per particle
→ Simulation of particle interaction with detectors is **highly complex and computationally intensive**

CMS Pb+Pb collision experimental data event display



sPHENIX Geant4 detector simulation
for proton-proton collisions
(Au+Au central collisions have x1000 more particles)

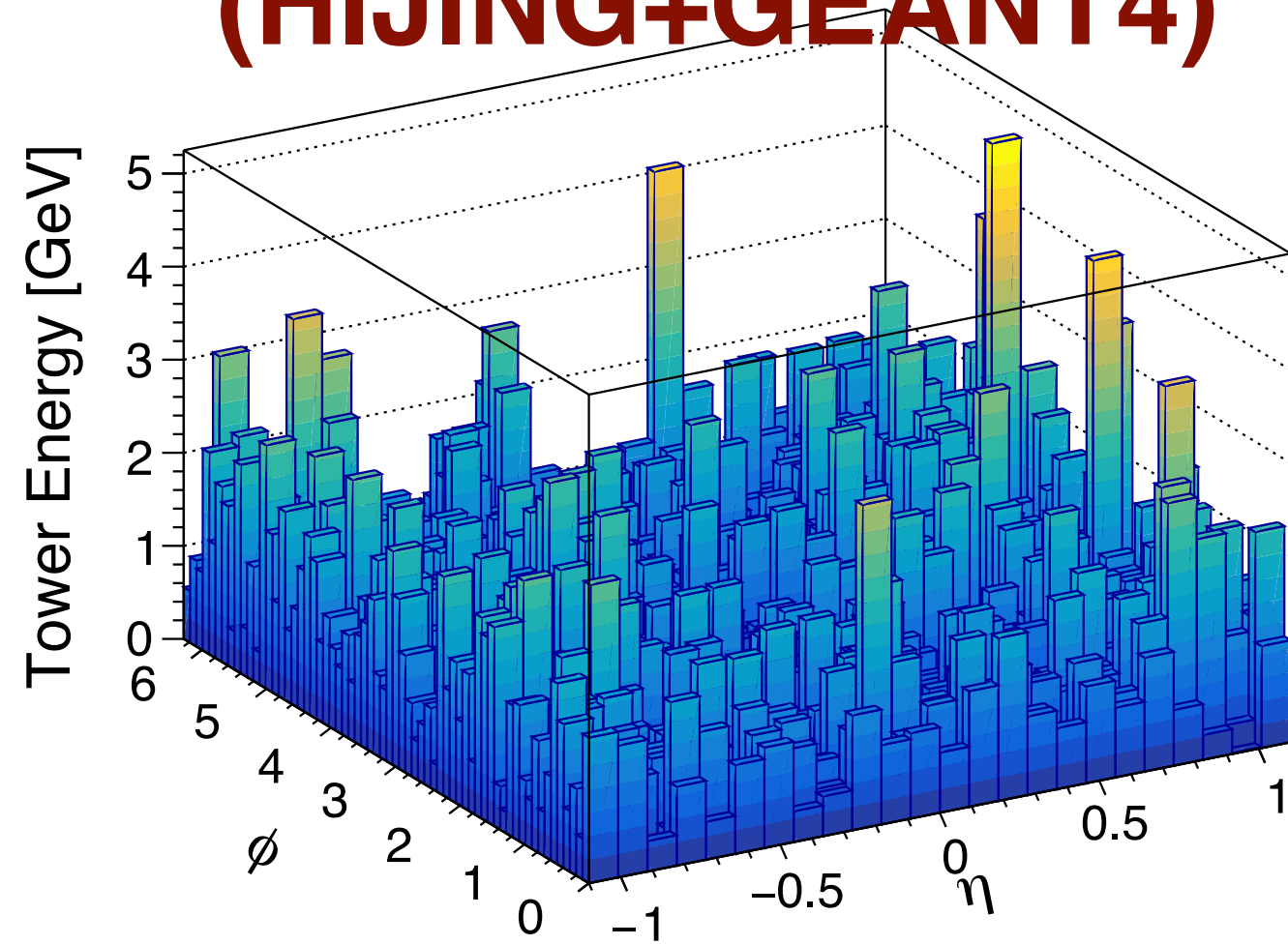


ML can speed up and produce large amount of the heavy ion event simulations!

Fast Simulation: Full detector, Whole Event

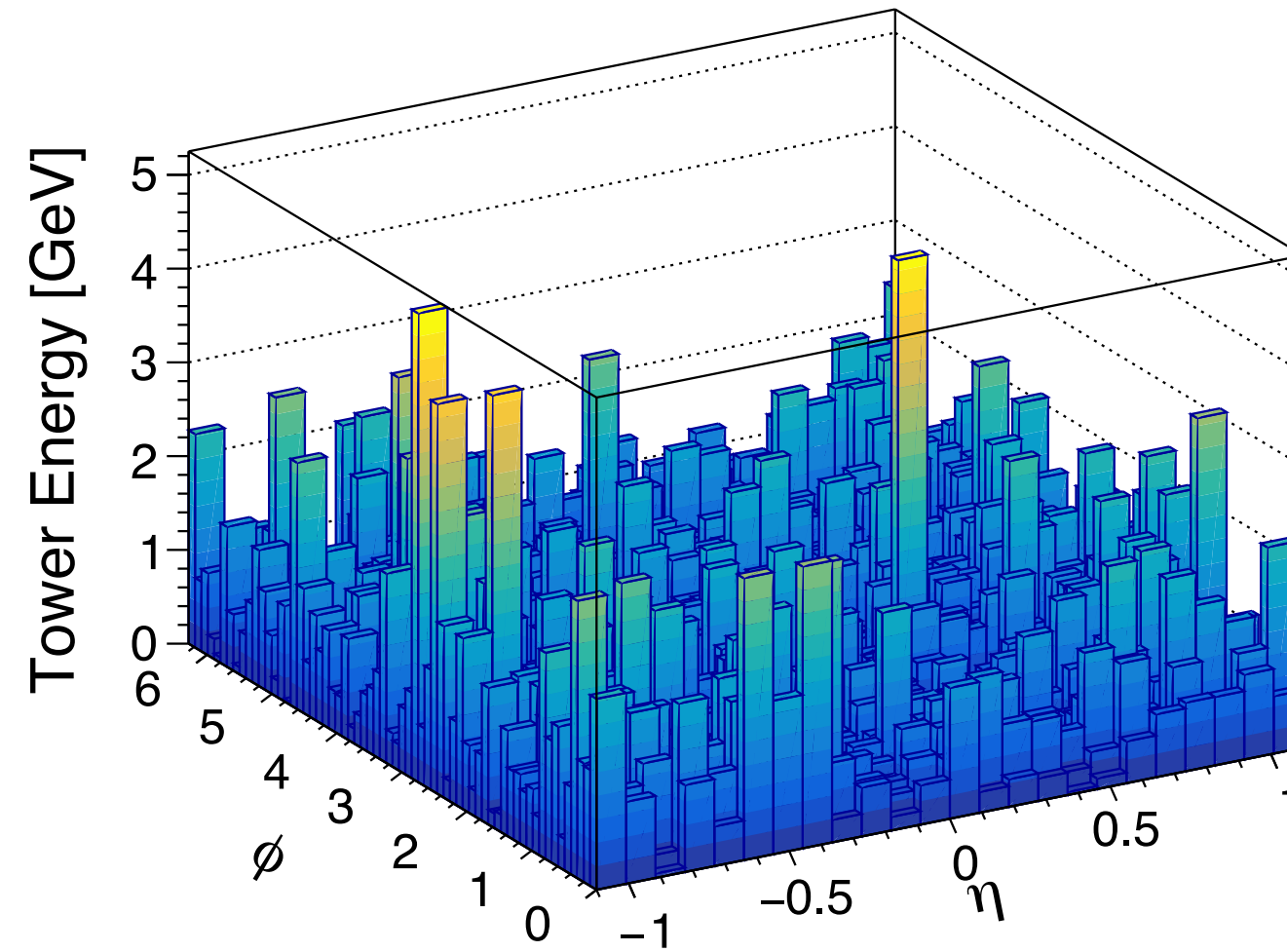
- Generate high fidelity “particle production + full detector simulation” data in *central heavy ion collisions* using Denoting Diffusion Probabilistic Model (**DDPM**)

**Training sample
(HIJING+GEANT4)**

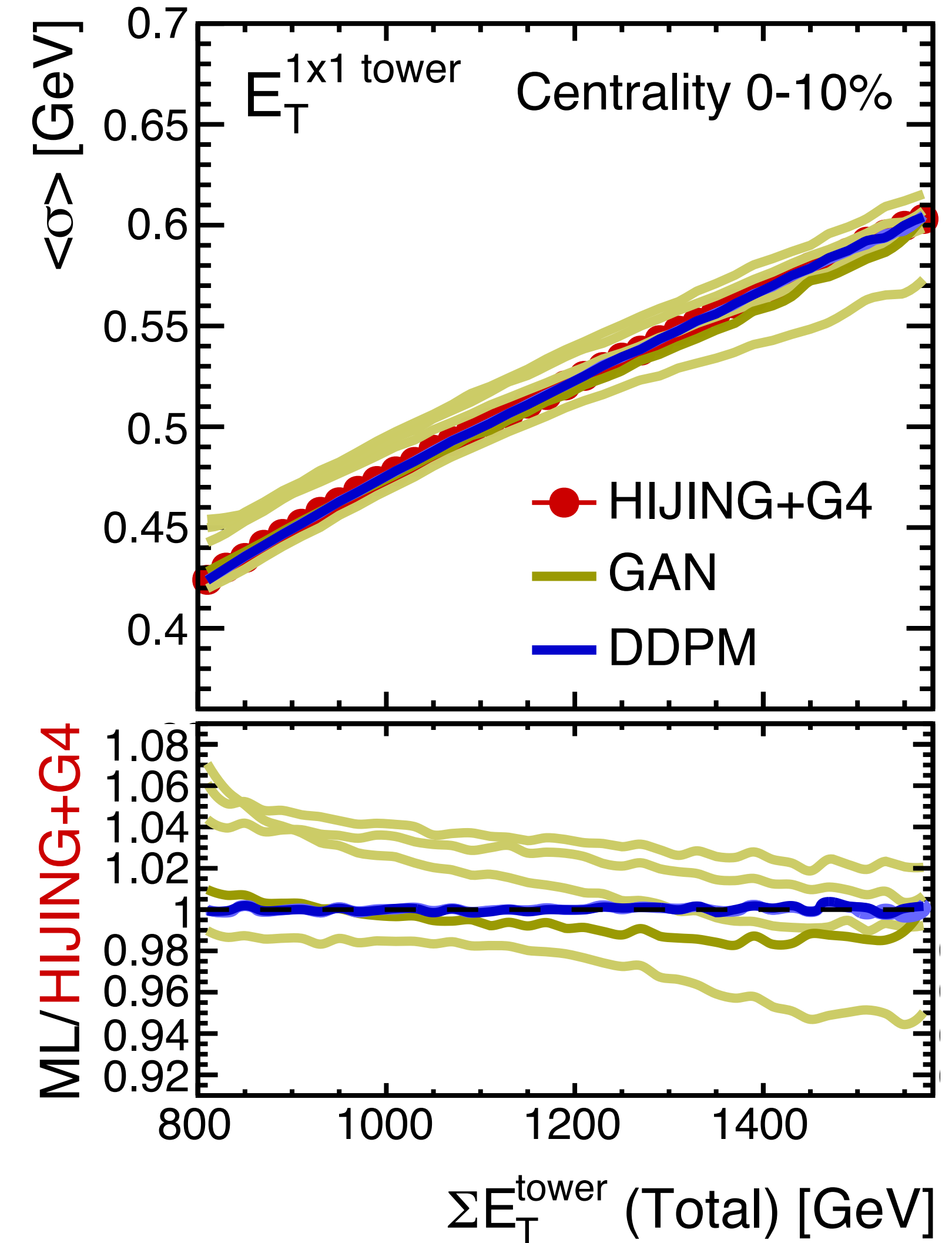


40 minutes per event
on CPU

Generated (DDPM)



1.34 seconds per event
on GPU

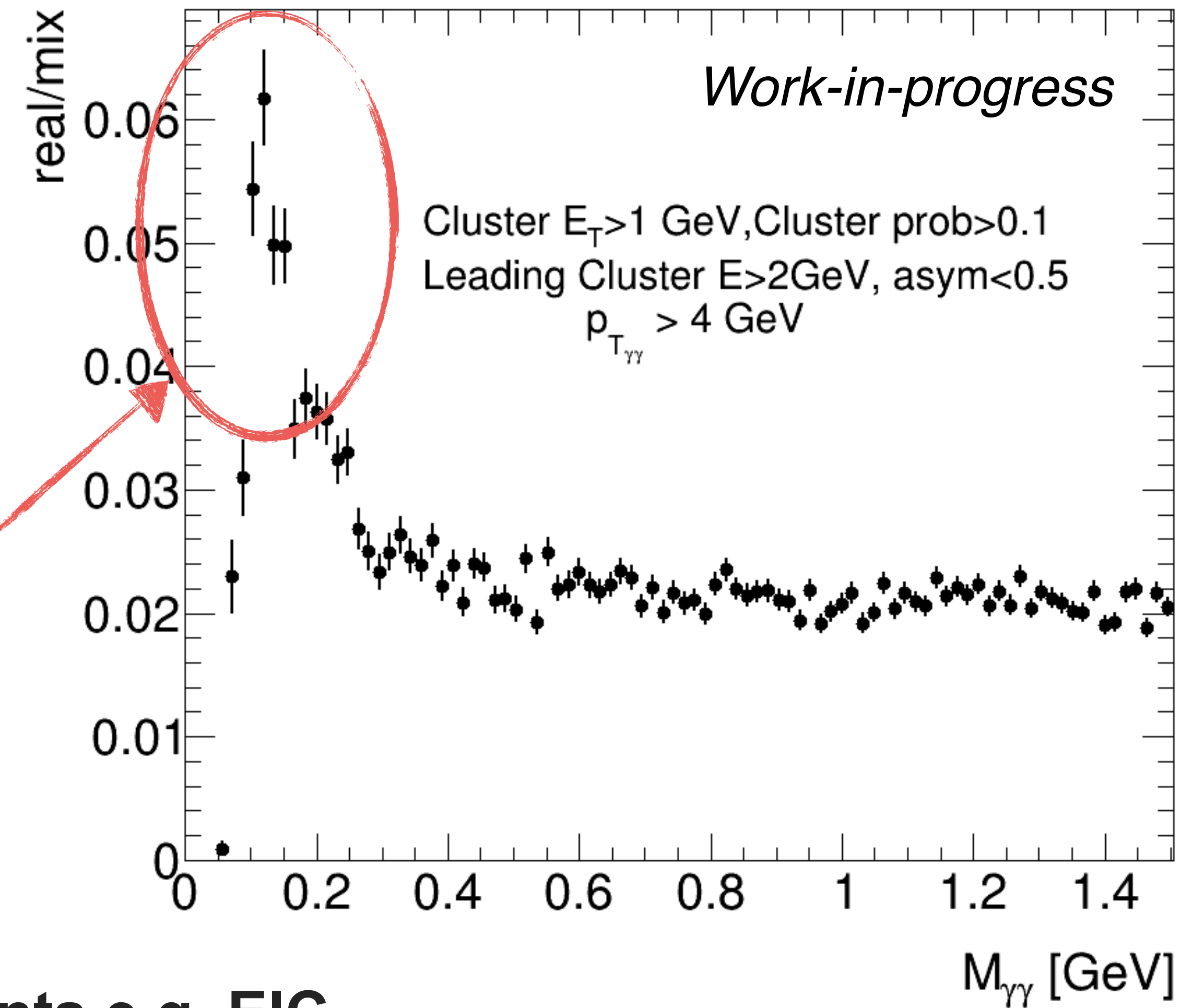


- **DDPM** provide a **speedup of 1800 times!** (on GPU)
➔ at least 100 times faster, considering a 32-core CPU equivalent to a GPU

Emerging Phenomena Captured by AI

- Can DDPM describe more complex features of heavy ion collisions?
➔ **Resonance, flow** can be reproduced by DDPM!

π^0 peak reconstructed using
simulation samples generated by DDPM

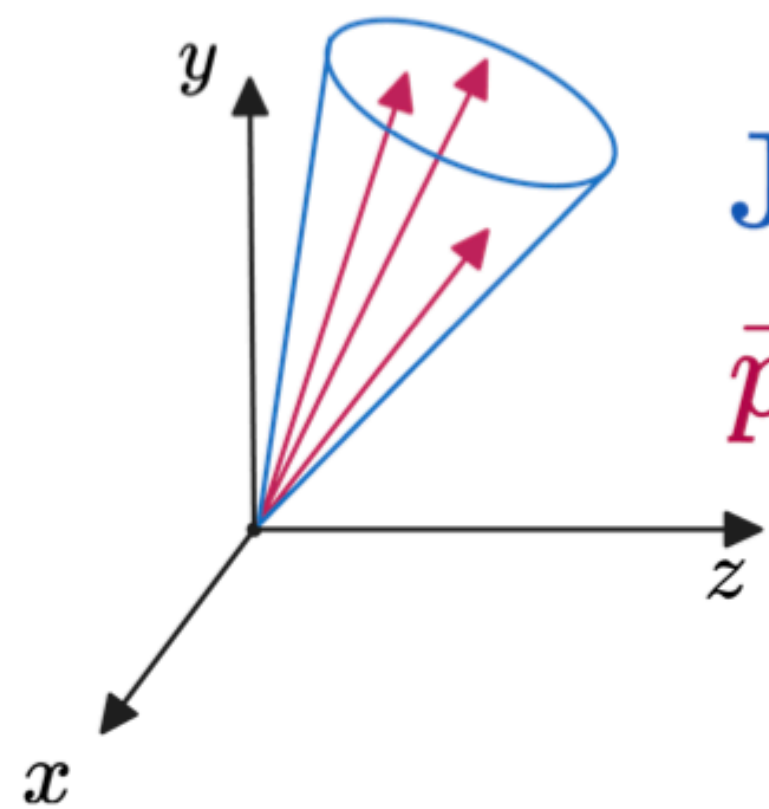


- Similar approach applicable to other experiments e.g. EIC

Foundation Model for Jet Analysis

- **OmniJet- α**

- ➔ Transformer-based, next token prediction
- ➔ **Input: jet constituents kinematics**
- ➔ pre-train with **6.7 million** parameters
- ➔ downstream task: hadronic top quark tagging



$$\text{Jet} = \{\vec{p}_1, \vec{p}_2, \dots, \vec{p}_n\}$$

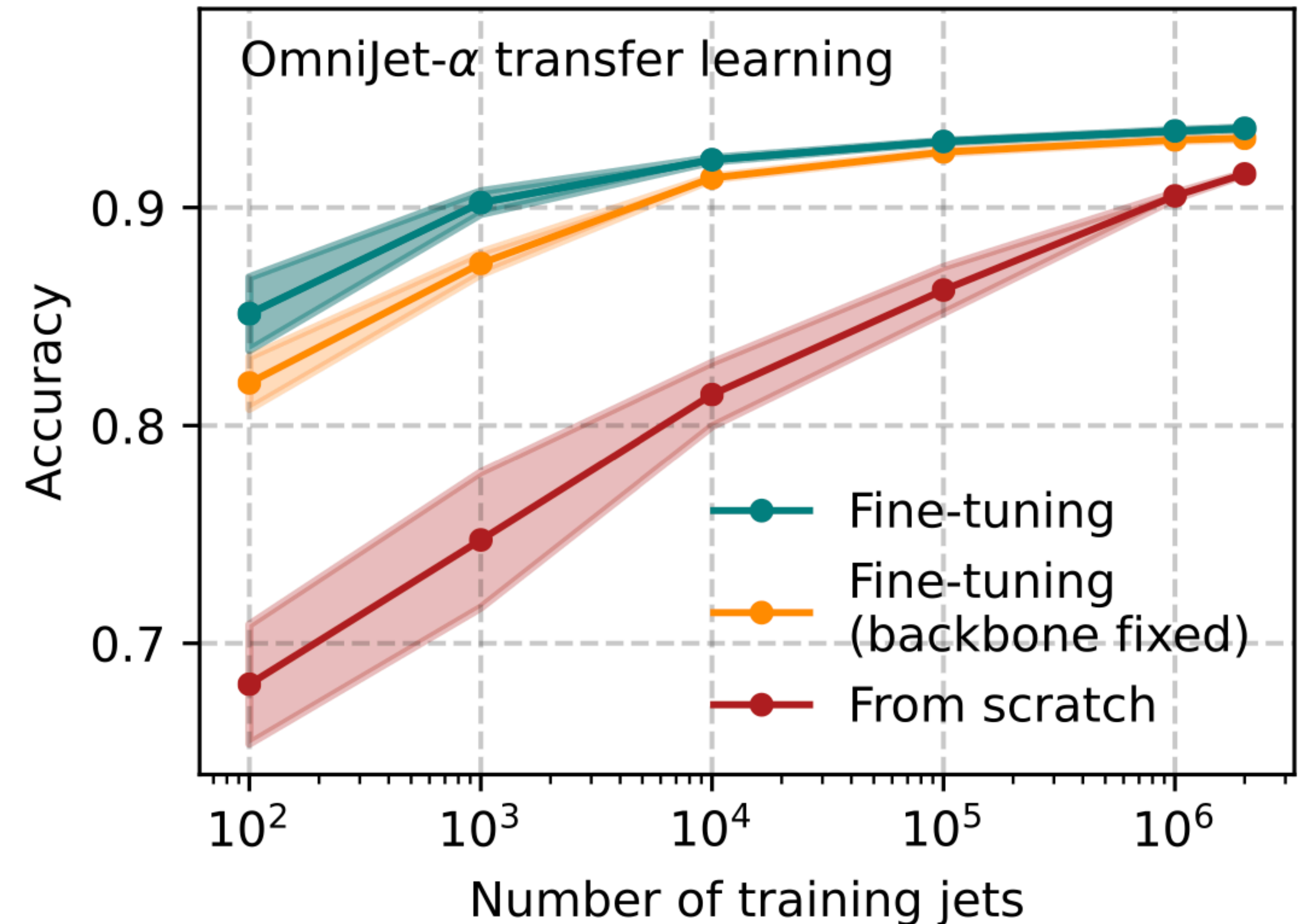
$$\vec{p}_i = (p_T, \Delta\eta, \Delta\phi)$$

Other works:

[ParT arXiv:2202.03772](#)

[MPM arXiv:2401.13537](#)

[Mach.Learn.Sci.Tech. 5 \(2024\) 3, 035031](#)



Classification: Heavy Flavor Jet Tagging

Generation	Main architecture	Single-pronged (AK4)		Multi-pronged		Novelty
		CMS	ATLAS	CMS (AK8/AK15)	ATLAS (AK10)	
First (Run-1 and early Run-2, shallow architectures)	Shallow NN	CSVv2	MV1	CSVv2	MV1 using subjets	Switch from likelihood methods to first ML-based methods. Incorporate larger number of inputs.
	BDT	c-tagger, cMVAv2	MV2	double-b	MV2 using subjets	
Second (Late Run-2, early deep NNs)	Deep dense NN	DeepCSV	DL1	DeepAK8	DL1r using subjets, D_{Xbb}	Use information from multiple selected tracks. Multiclassification.
	CNN					Use a large number of particles per jet.
	RNN		RNNIP, DL1r			Accommodate an arbitrary number of tracks per jet.
	CNN+RNN	DeepJet		DeepDoubleX		Avoid applying selection criteria; use all jet constituents of jets including neutral particles.
Third (Early Run-3, sets and particle cloud representations)	Deep Set		DIPS, DL1d		DeXTer	Avoid sorting of jet constituents: permutation invariant.
	GNN		GN1			Use large number of track features, including pixel hits. Assign auxiliary track origin and vertex finding tasks.
	DGCNN			ParticleNet		Leverage local features of point clouds using convolutions.
	Transformer	ParticleTrans-formerAK4	GN2		GN2X	Leverage pairwise features, attention mechanisms, and larger training datasets.

Table 2: Flavor tagging algorithms developed in the CMS and ATLAS experiments, classified by generations and main ML architecture used.

[Eur.Phys.J.ST 233 \(2024\) 15-16, 2657-2686](#)

sPHENIX Detector at RHIC



Tracking System

TPC
INTT
MVTX

Calorimeters

Electromagnetic
Inner Hadronic
Outer Hadronic

- Hermetic Electromagnetic & Hadronic **calorimeters**
- 3 layers: EMCAL + inner HCAL + outer HCAL

