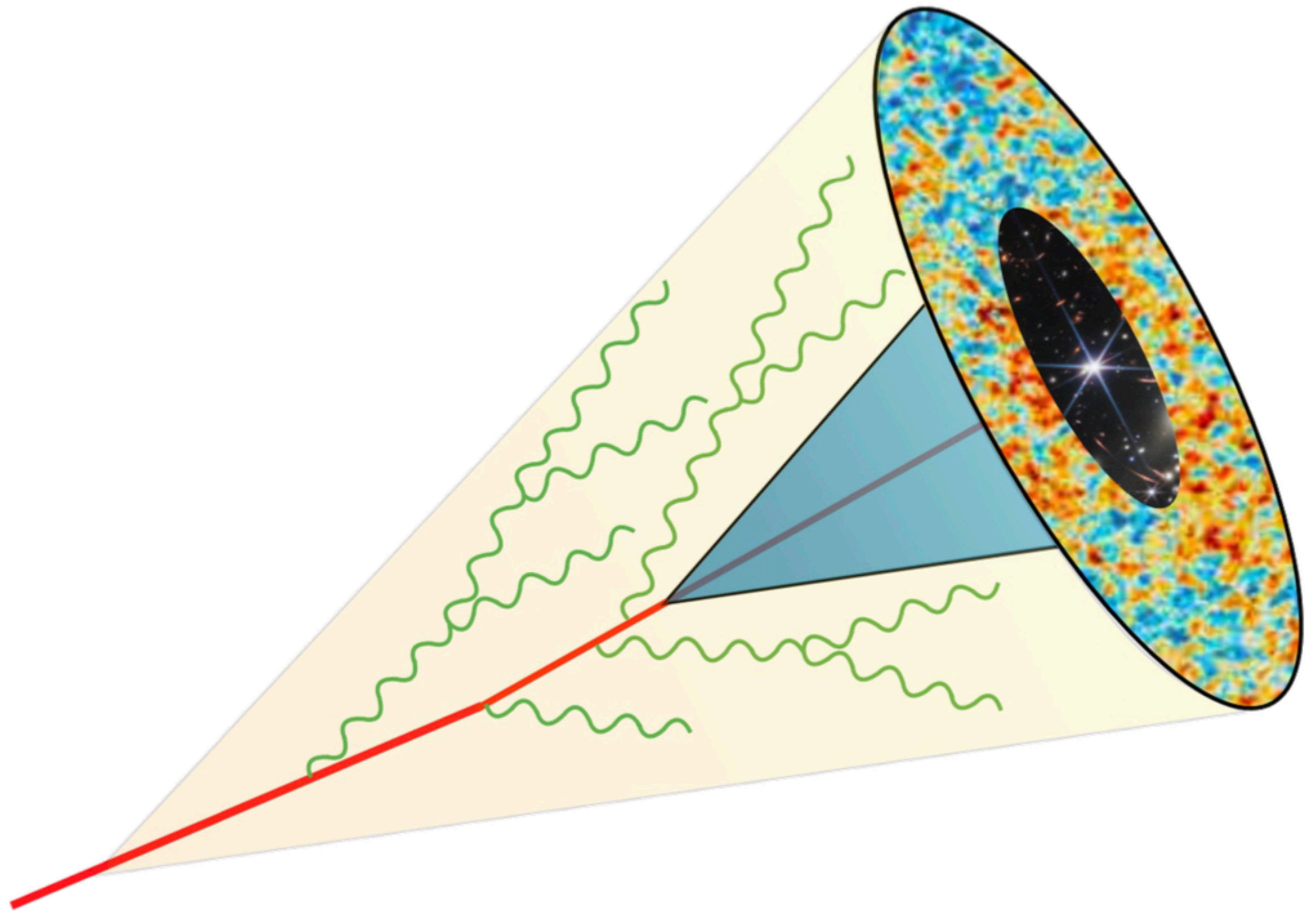
NATIONAL SCIENCE FOUNDATION

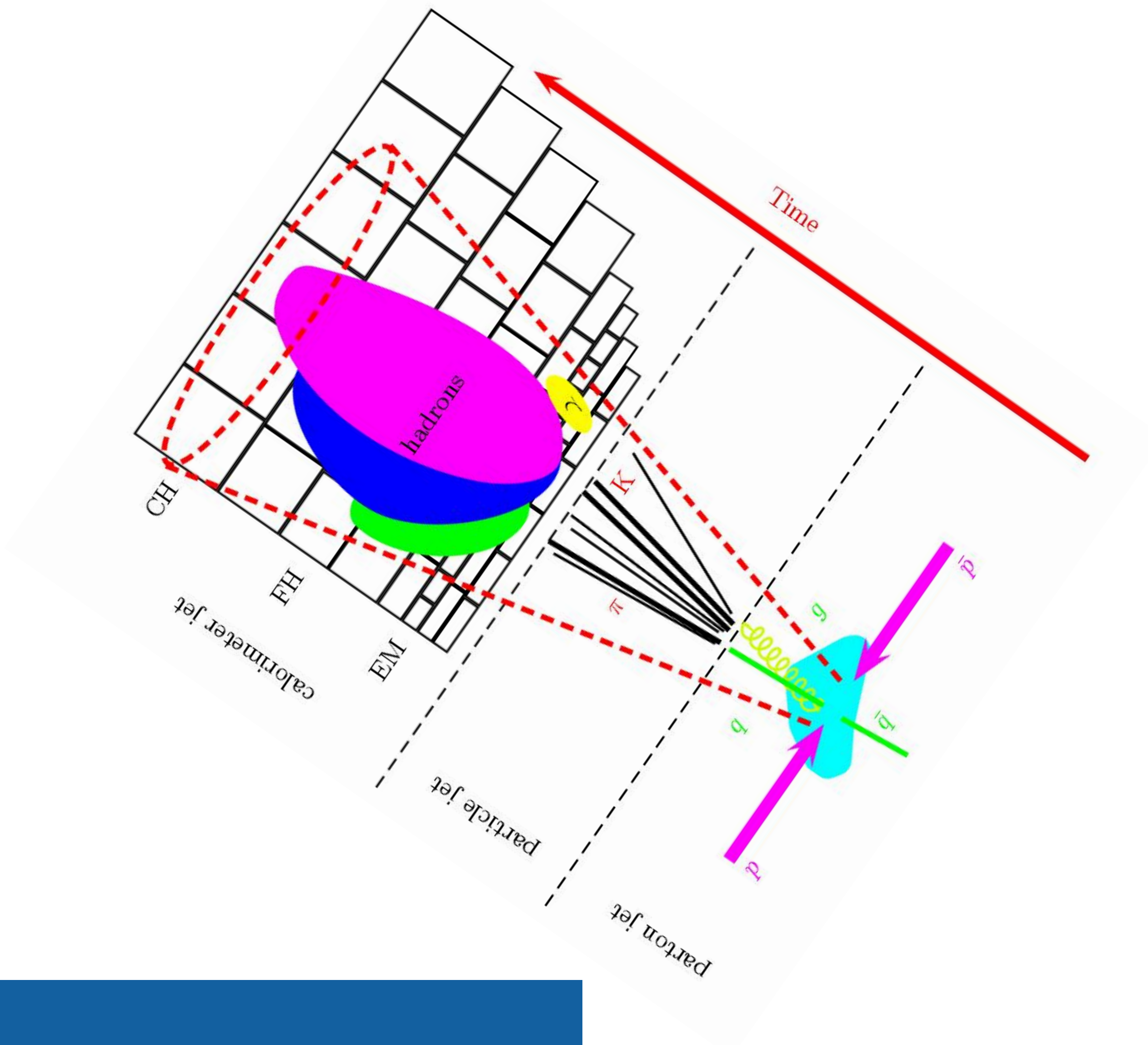
U.S. DEPARTMENT OF ENERGY | Office of Science

VANDERBILT UNIVERSITY

Preparing the Jet-AI/ML landscape for the EIC



Raghav (**Rithya**)
Kunnawalkam Elayavalli (she/they)
Vanderbilt University
Data Science Institute
raghavke.me



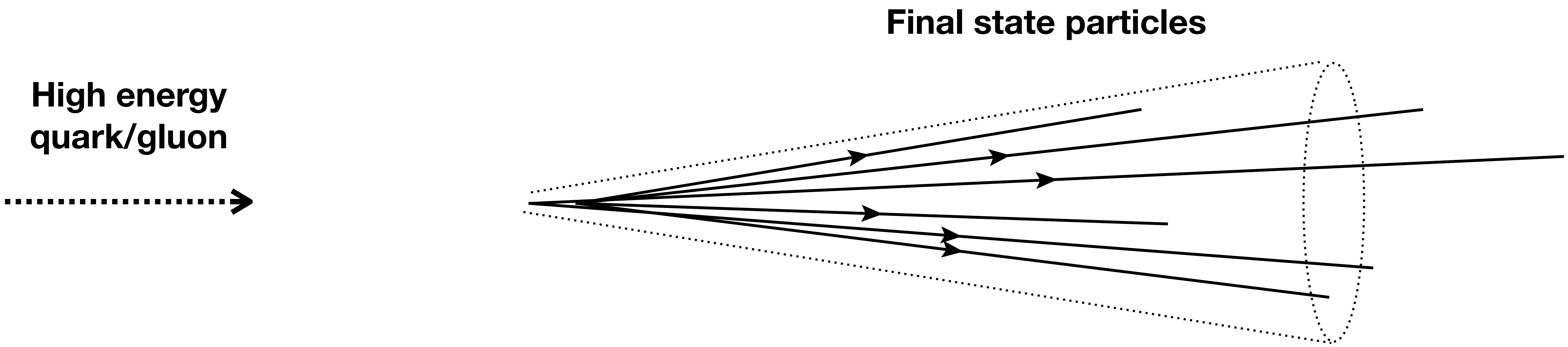
Probing the frontiers of nuclear physics with AI at the EIC (II)

March 19-21 2025, CFNS @ Stony Brook University

Why Jets and Why AI/ML

Basics - what are Jets?

How we observe quarks/gluons in nature

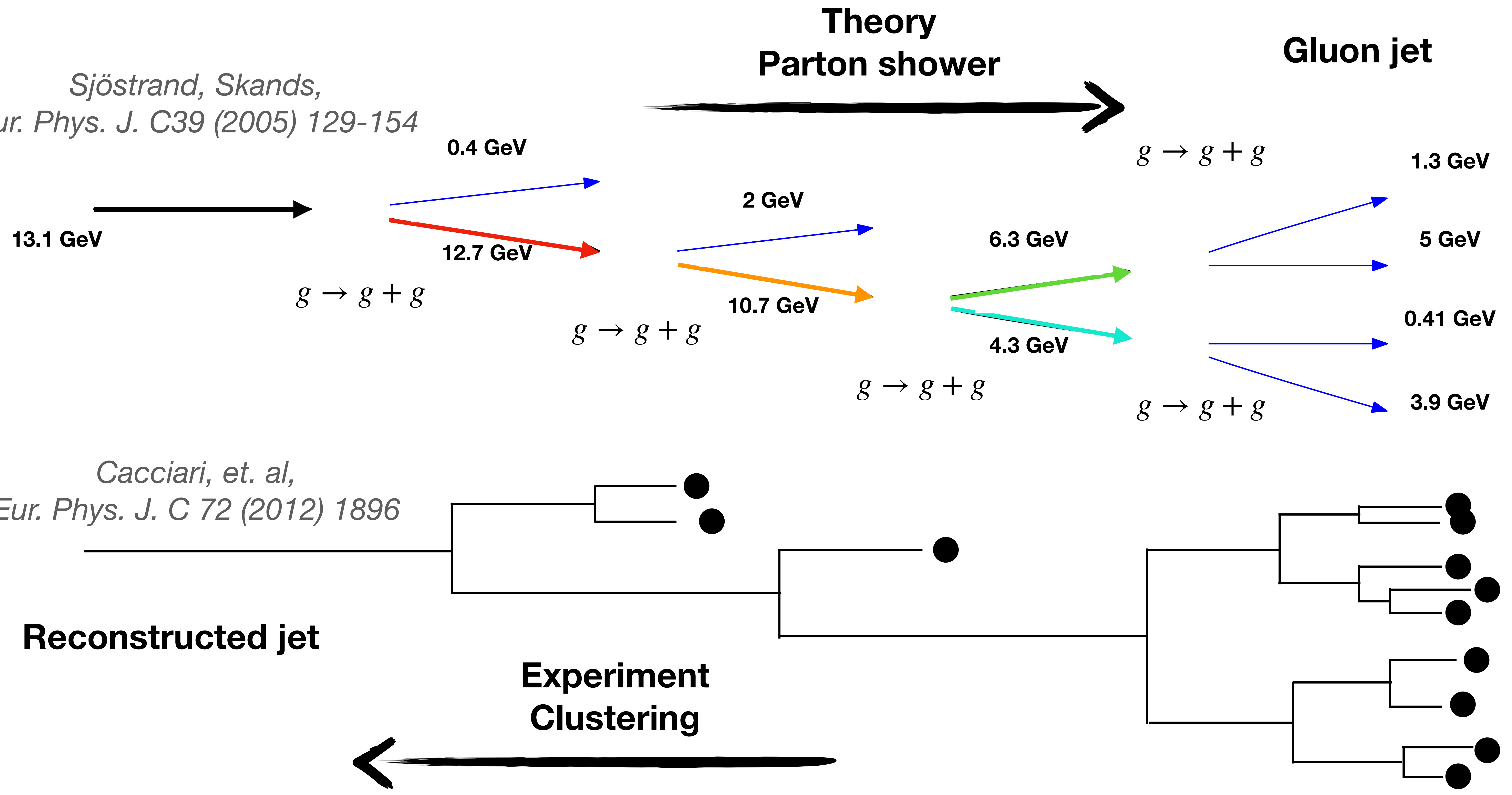


Collimated collection of hadrons resulting from the ‘***metamorphosis***’ of partons due to fragmentation and hadronization

Gaillard et. al, Nucl. Phys. B111 (1976) 253-271

Jet correspondence

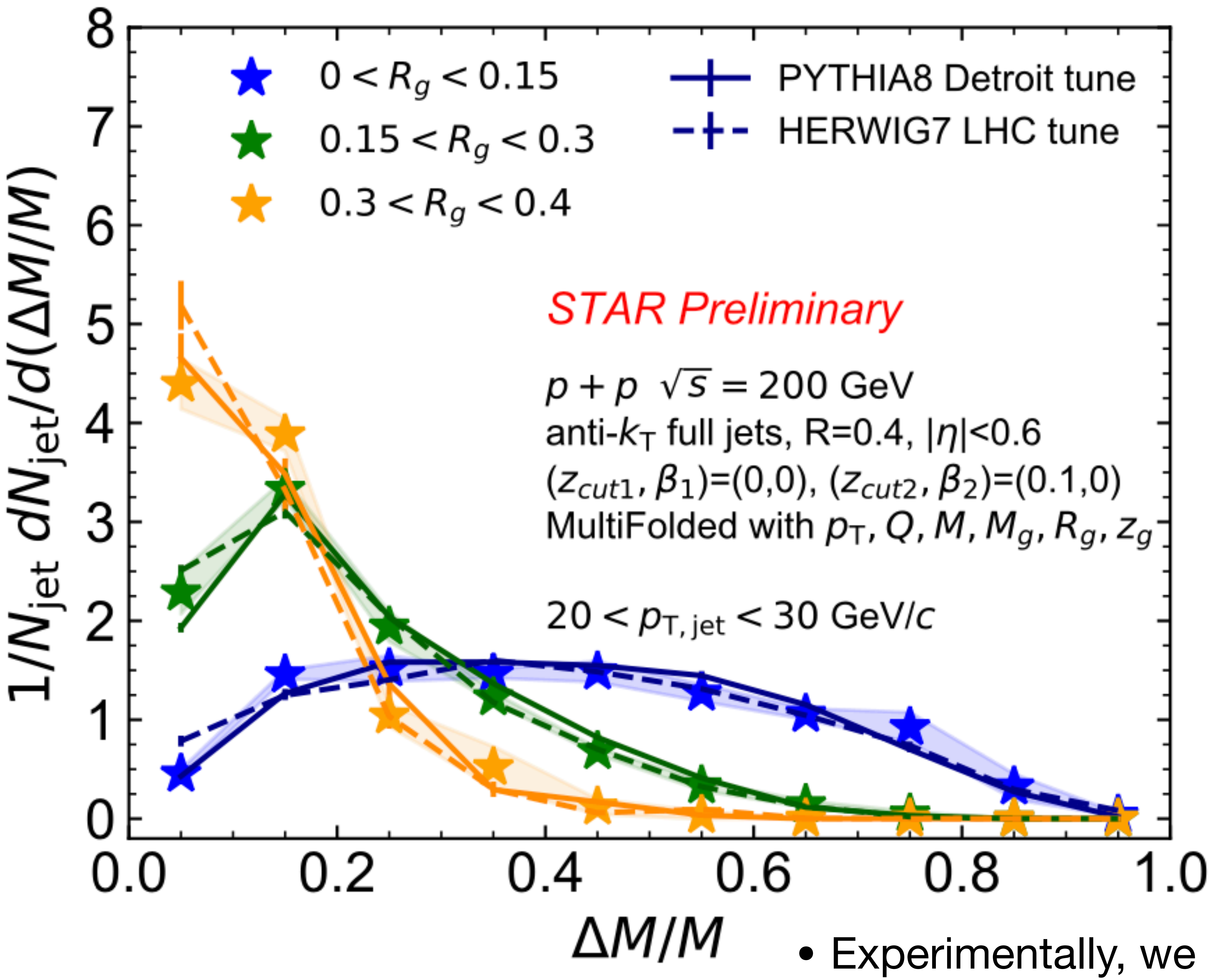
Sjöstrand, Skands,
Eur. Phys. J. C39 (2005) 129-154



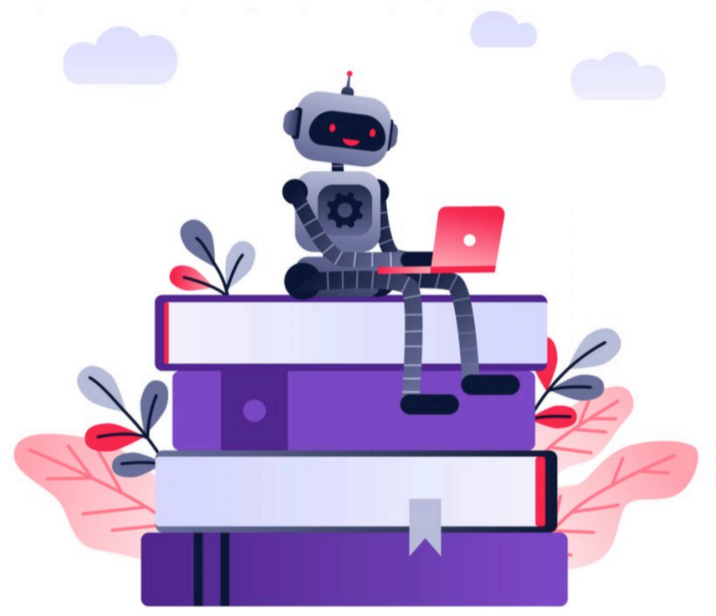
Cacciari, et. al,
Eur. Phys. J. C 72 (2012) 1896

Fundamental question - why AI/ML?

- Jets are user defined objects - *varied* representation phase-space
- Multi-scale objects, in both its energy and angle



Multifold allowed us to measure this!



p_T vs Q vs M vs z_g vs R_g vs M_g

Youqi Song (Yale) @ DIS 2023

Andreassen et.al
Phys. Rev. Lett. 124, 182001 (2020)

6D unfolded simultaneously via MultiFold machine learning technique

- Experimentally, we have shown virtuality loss along the direction of the jet - AI/ML unfolding made possible

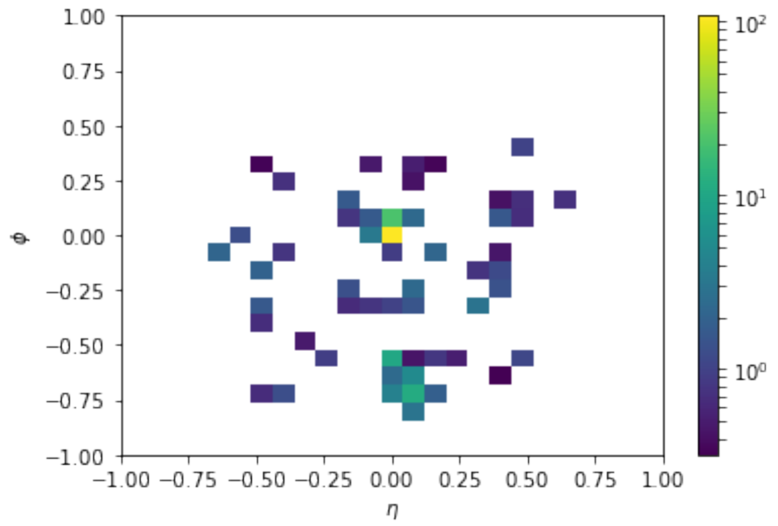
Fundamental question - why AI/ML?

- Jets are user defined objects - *varied* representation phase-space
- Multi-scale objects, in both its energy and angle
- Every single jet goes through a perturbative parton shower followed by a non-perturbative process of hadronization which results in fragmentation
- Basic assertion - the information content within jets is multi-dimensional
- We have specific questions - lets use specific models to answer those

Birth of a jet

Can we tag the flavor of the jets?

- What do we learn from this - flavor dependent fragmentation
- Proton's PDF and possibly extending all the way to GPDs
- Can we find the mother q/g?



Vertexing	Tracking	Fragmentation
DCAXY, DCAZ	p_T, η, ϕ	z, ΔR, $z\Delta R^2$

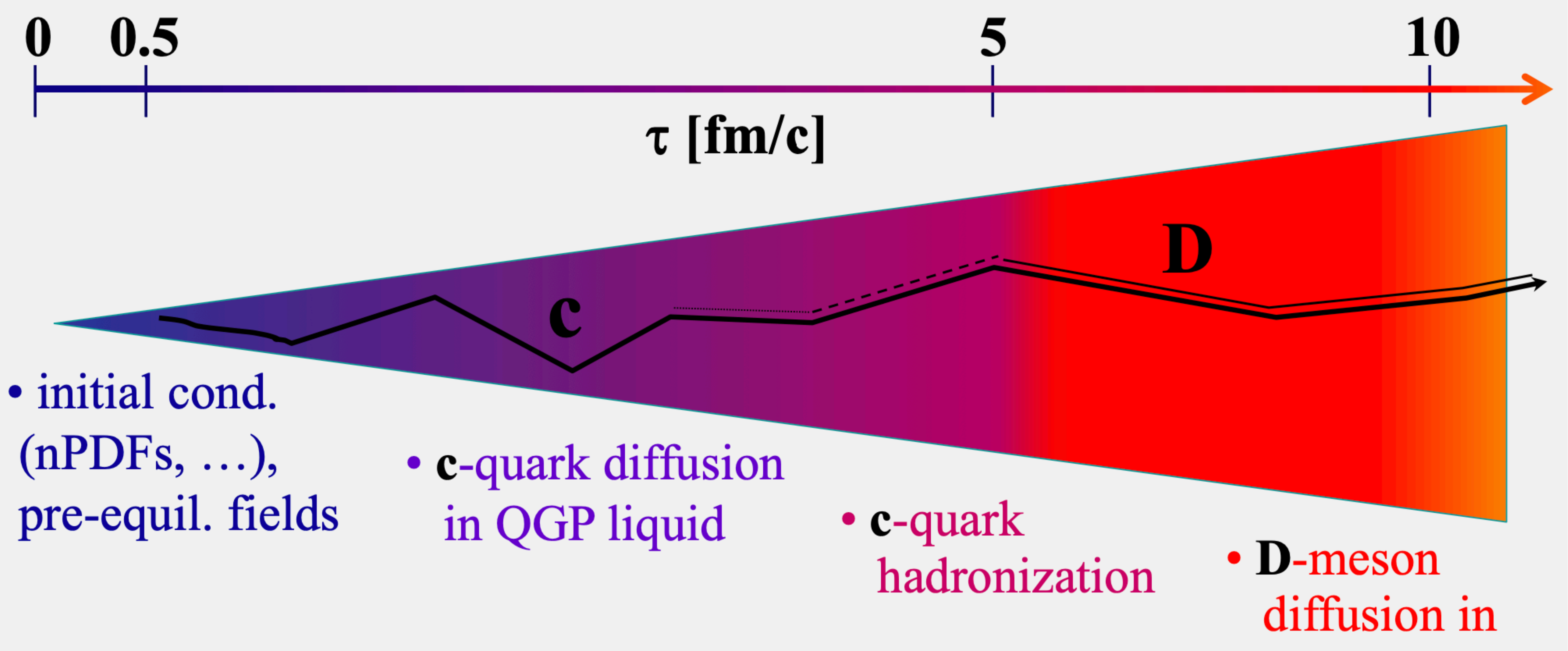


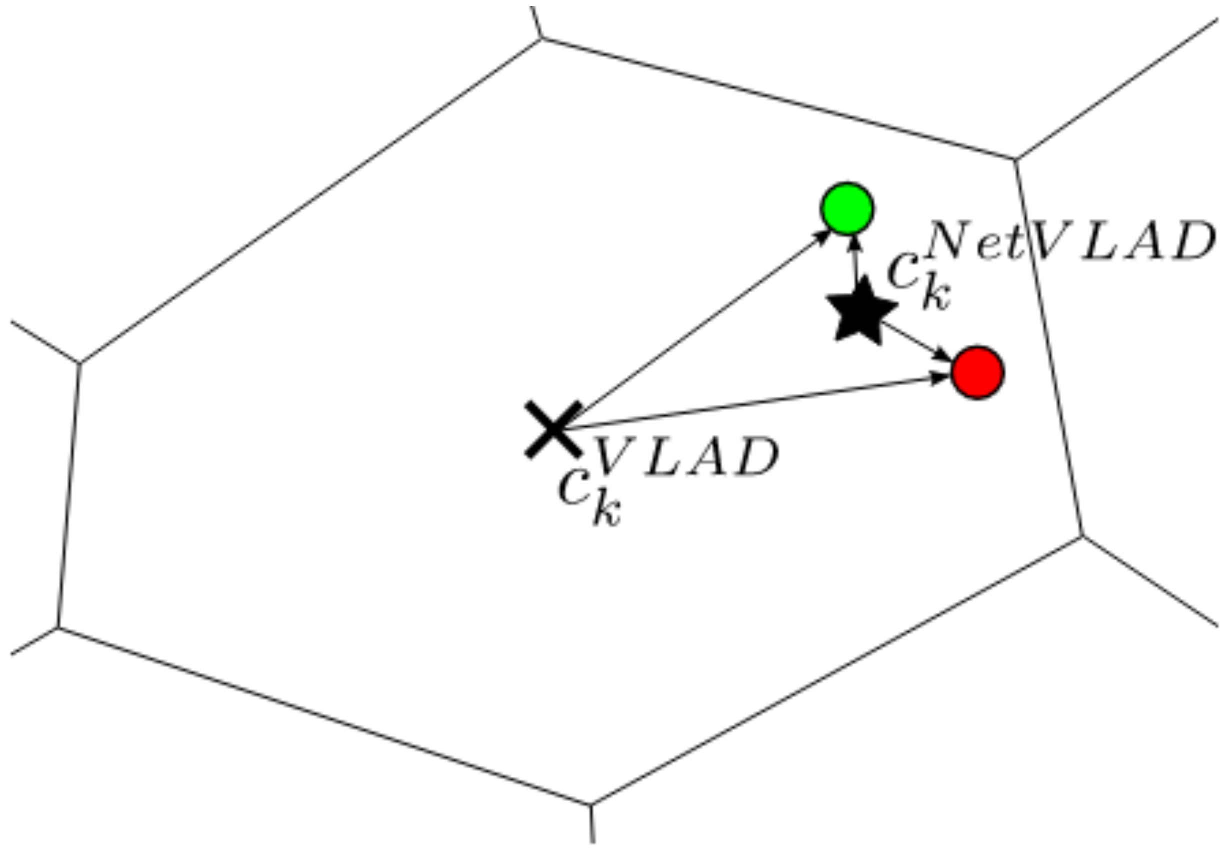
Figure courtesy - Ralf Rapp

NetVLAD

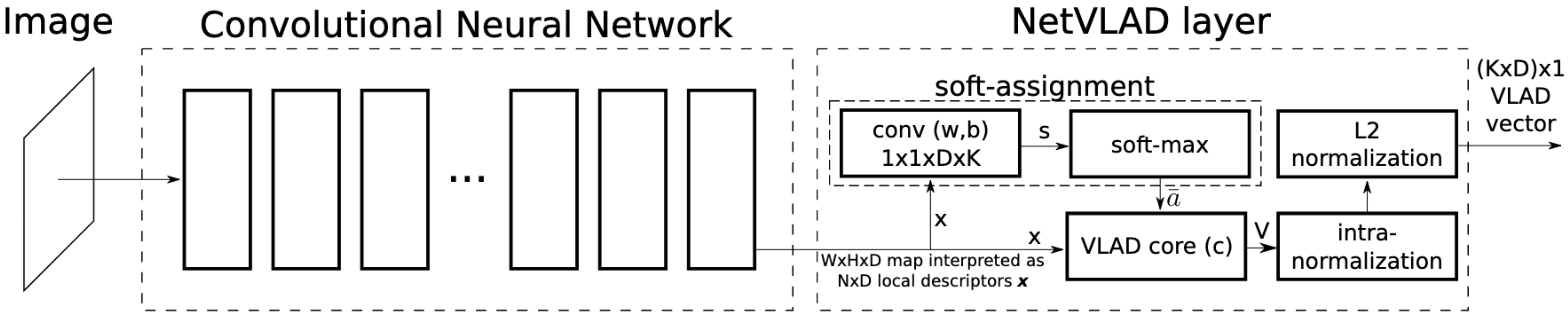
Arandjelović et. al 1511.07247

(a) Mobile phone query

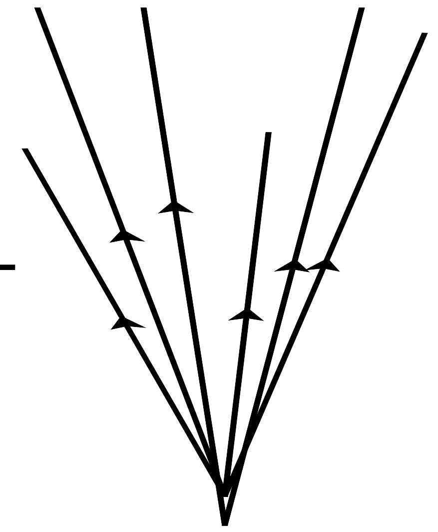
(b) Retrieved image of same place



- CNN architecture for weakly supervised place recognition



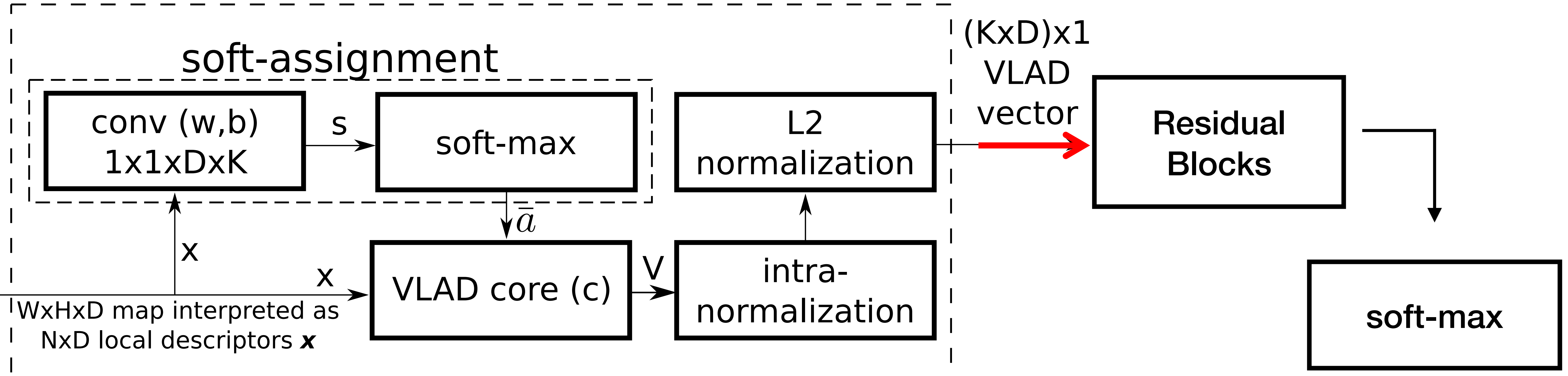
Tagging Heavy-Flavor JetVLAD



$$\mathcal{J} = \{(p_{T,i}, \eta_i, \phi_i, \ldots)\}_{i=1}^{n},$$

$$V(:,k) = \sum_{i=1}^{N} \frac{e^{w_k^T x_i + b_k}}{\sum_{k'} e^{w_{k'}^T x_i + b_{k'}}} (x_i - c_k)$$

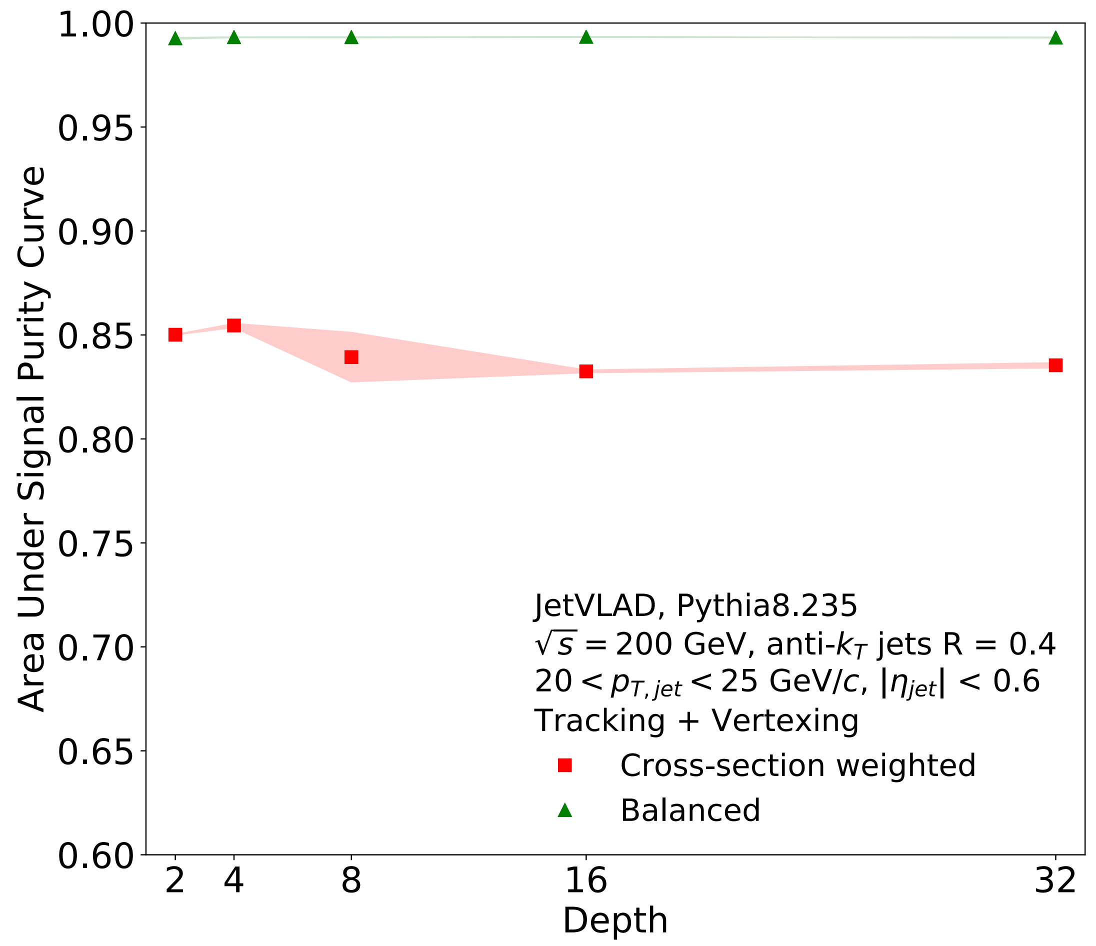
NetVLAD layer

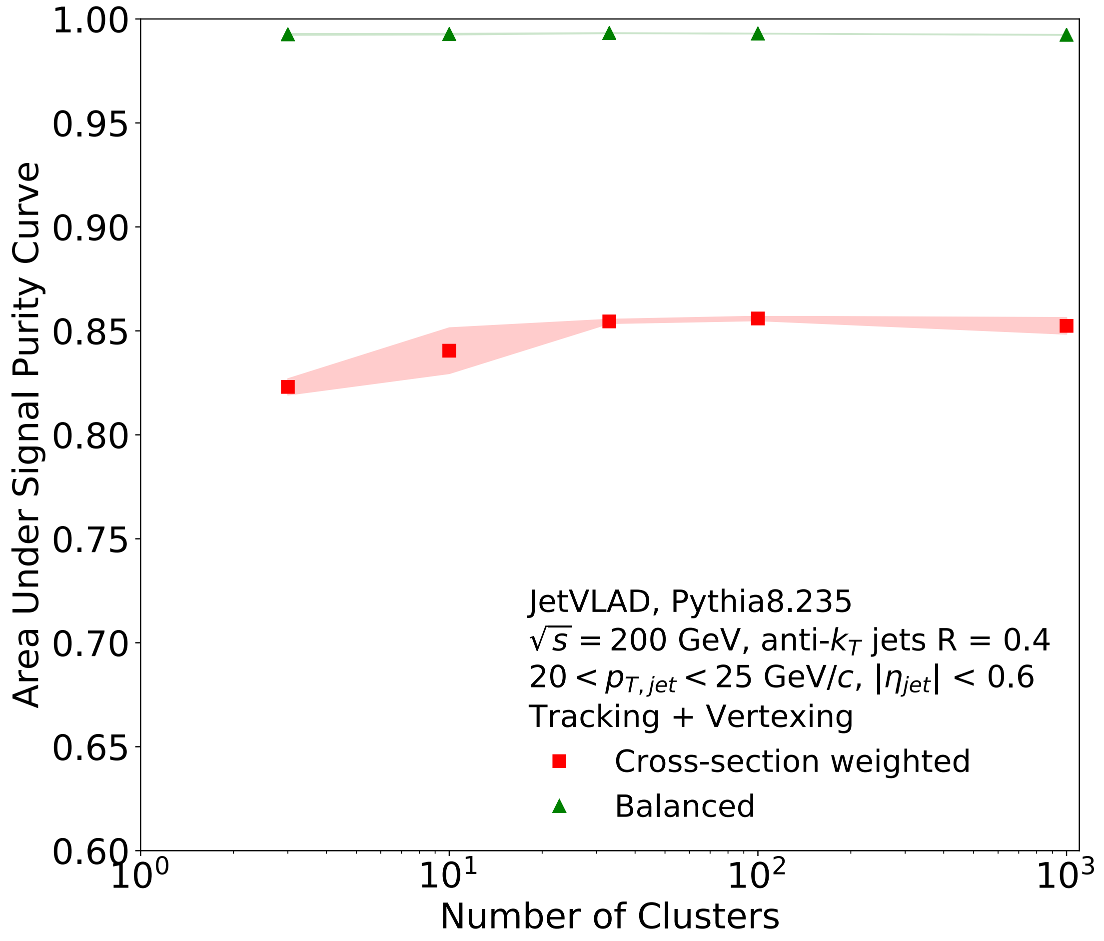


Vector of Locally Aggregated Descriptors

Jitka Mrazkova
PhD student @ NPI

Georgy Ponimatkin
PhD. In ML @ Ecole des Ponts ParisTech





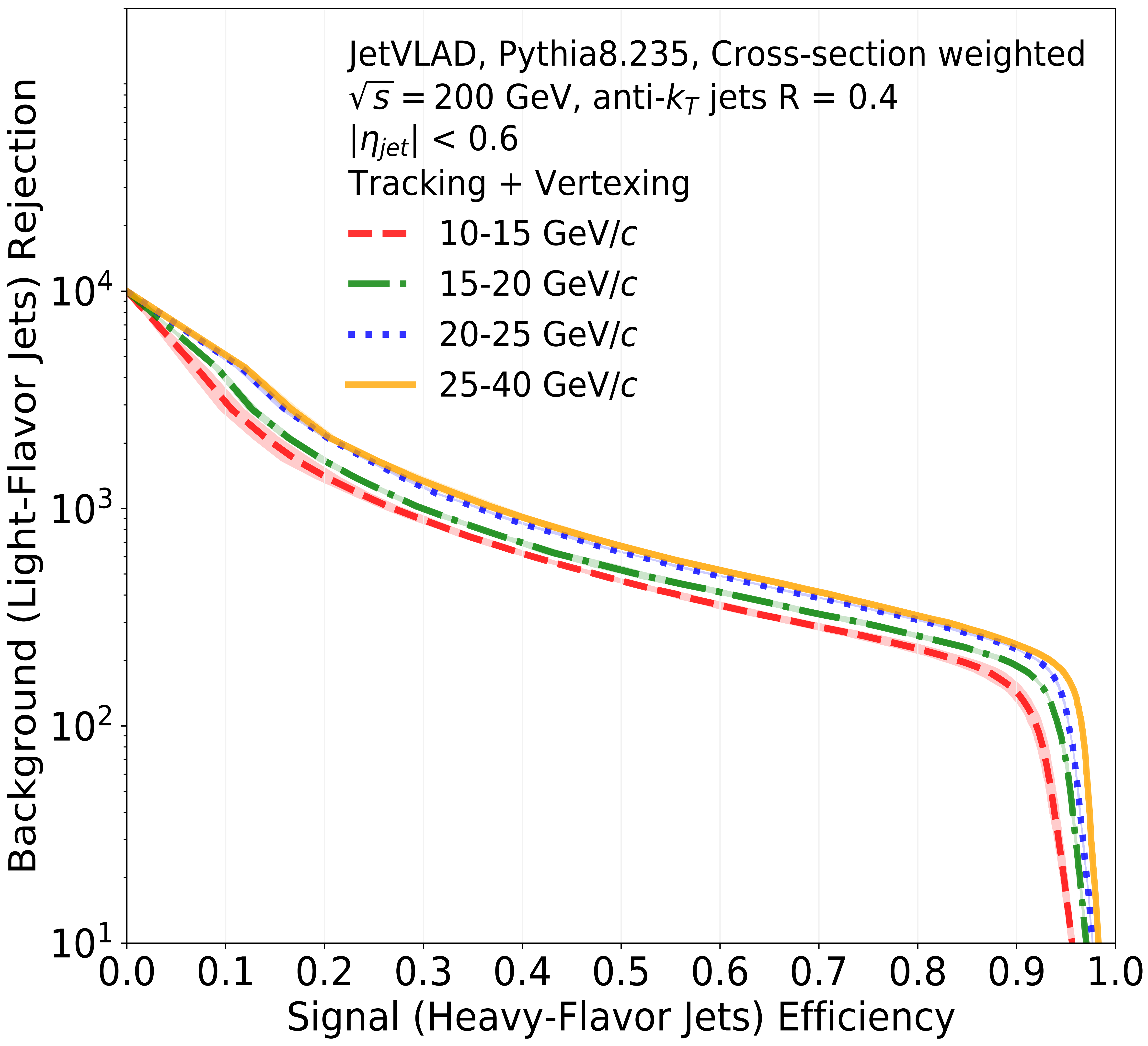
Ponimatkin, et. al JINST 2005.01842

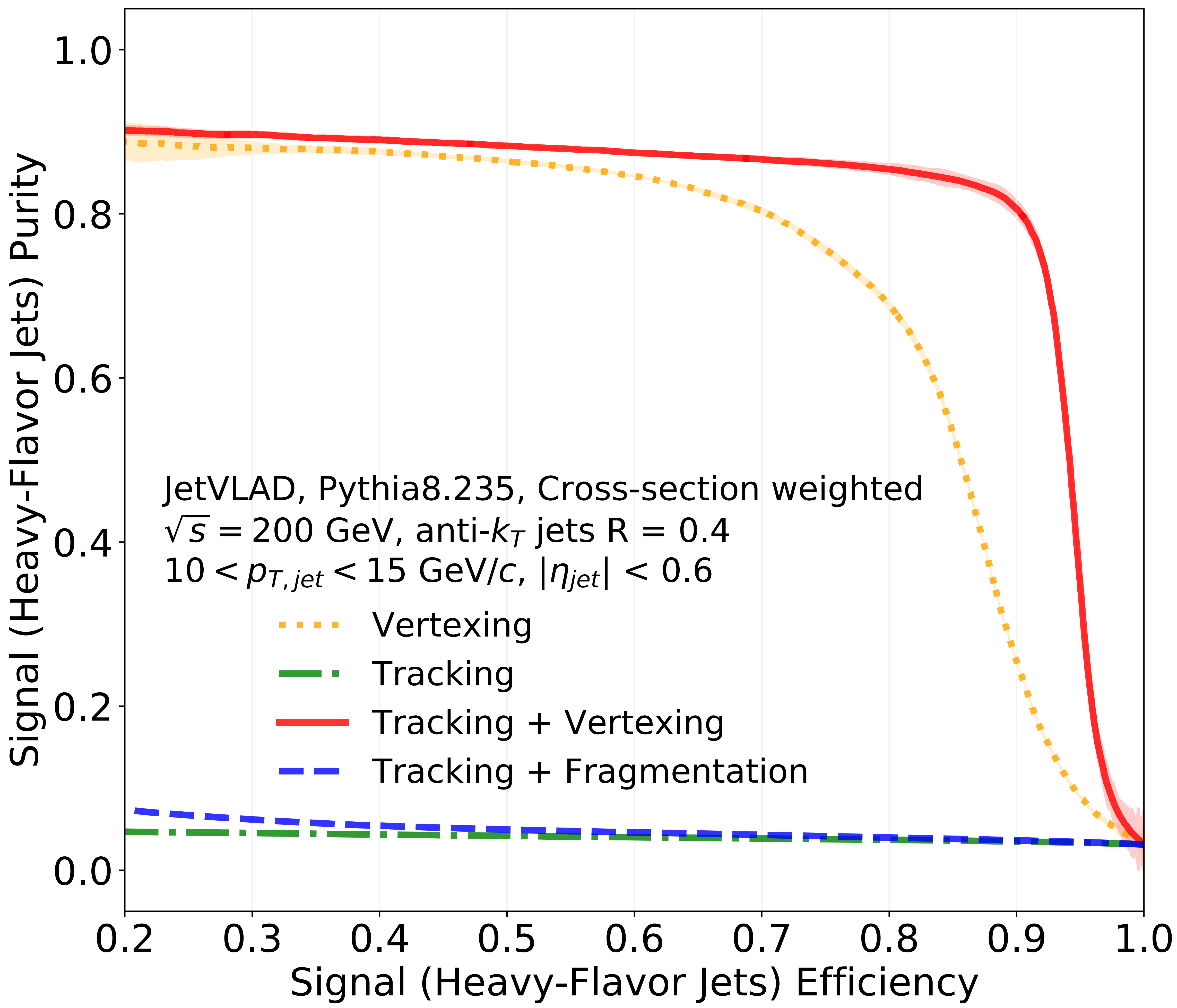
D - Depth
K - # Clusters

Total of 111608 trainable parameters

- impact of varying the random weights for each trainable parameters - probes the inbuilt uncertainty for your optimized model

Performance benchmarks





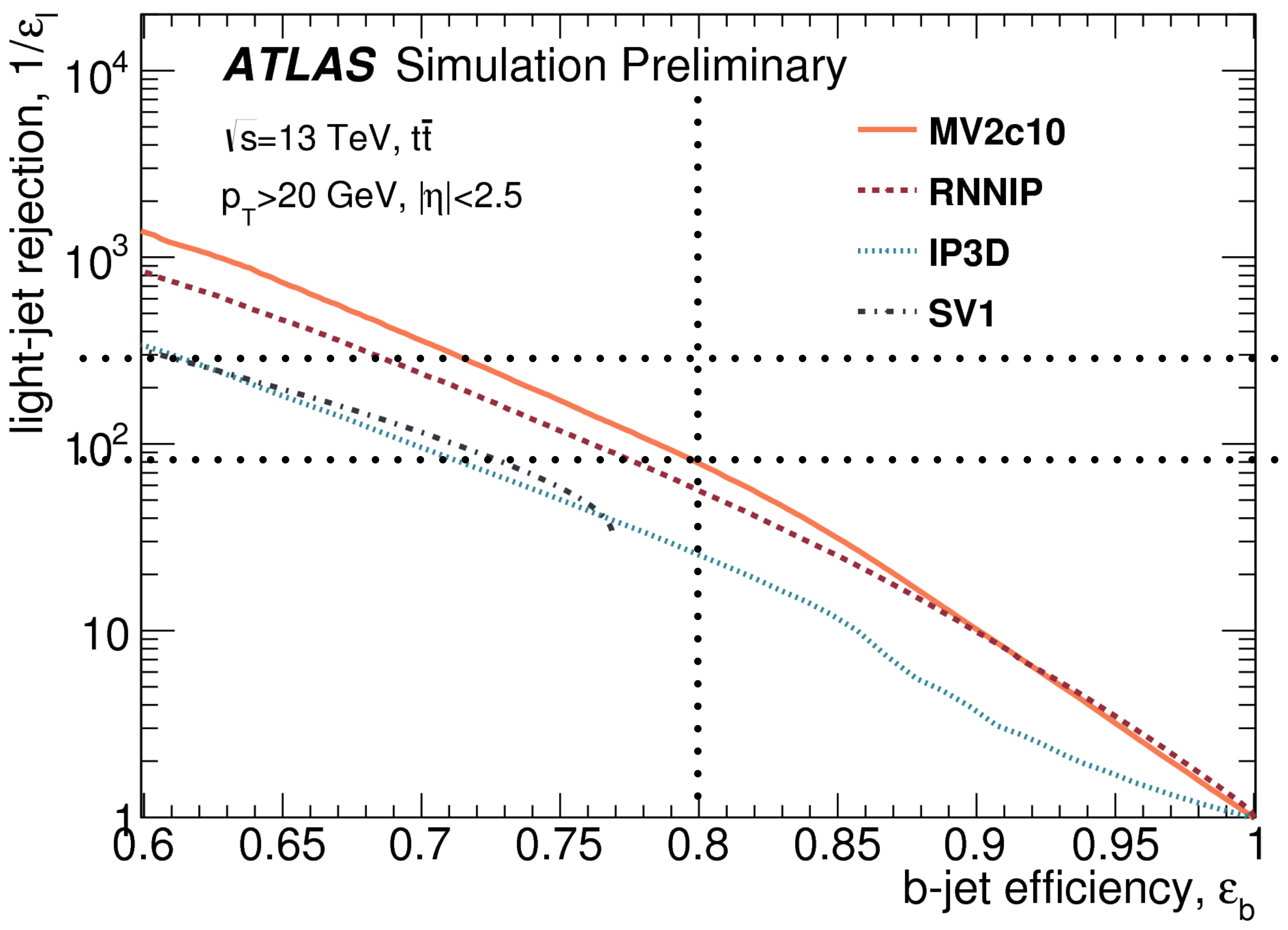
- Improvement (factor of 2) attributed to algorithmic differences primarily in comparison to RNN (which are quite hard to train)

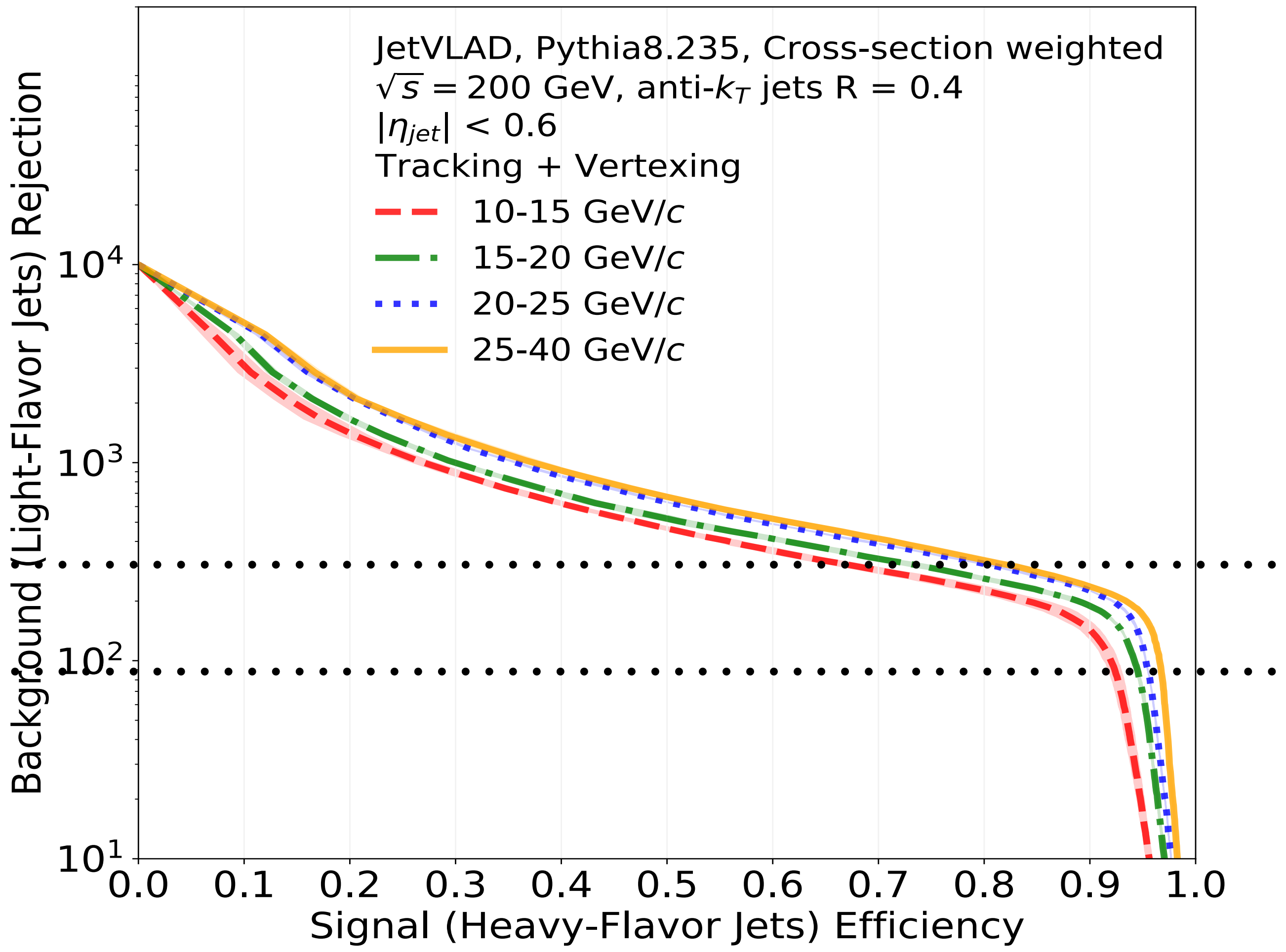
Performance benchmarks

Comparison w/ ATLAS and CMS

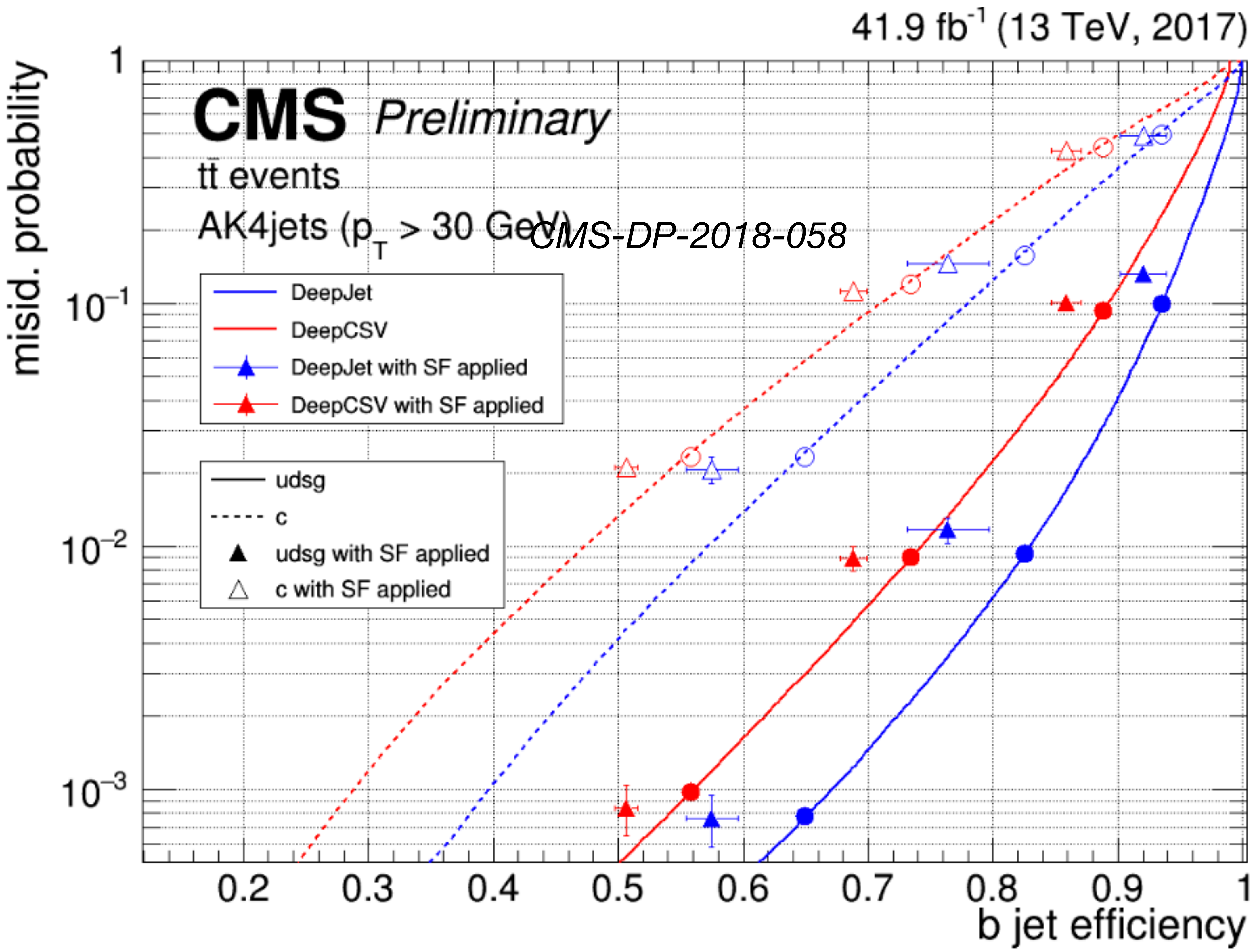
- Recurrent neural network along with IP3

ATL-PHYS-PUB-2017-003



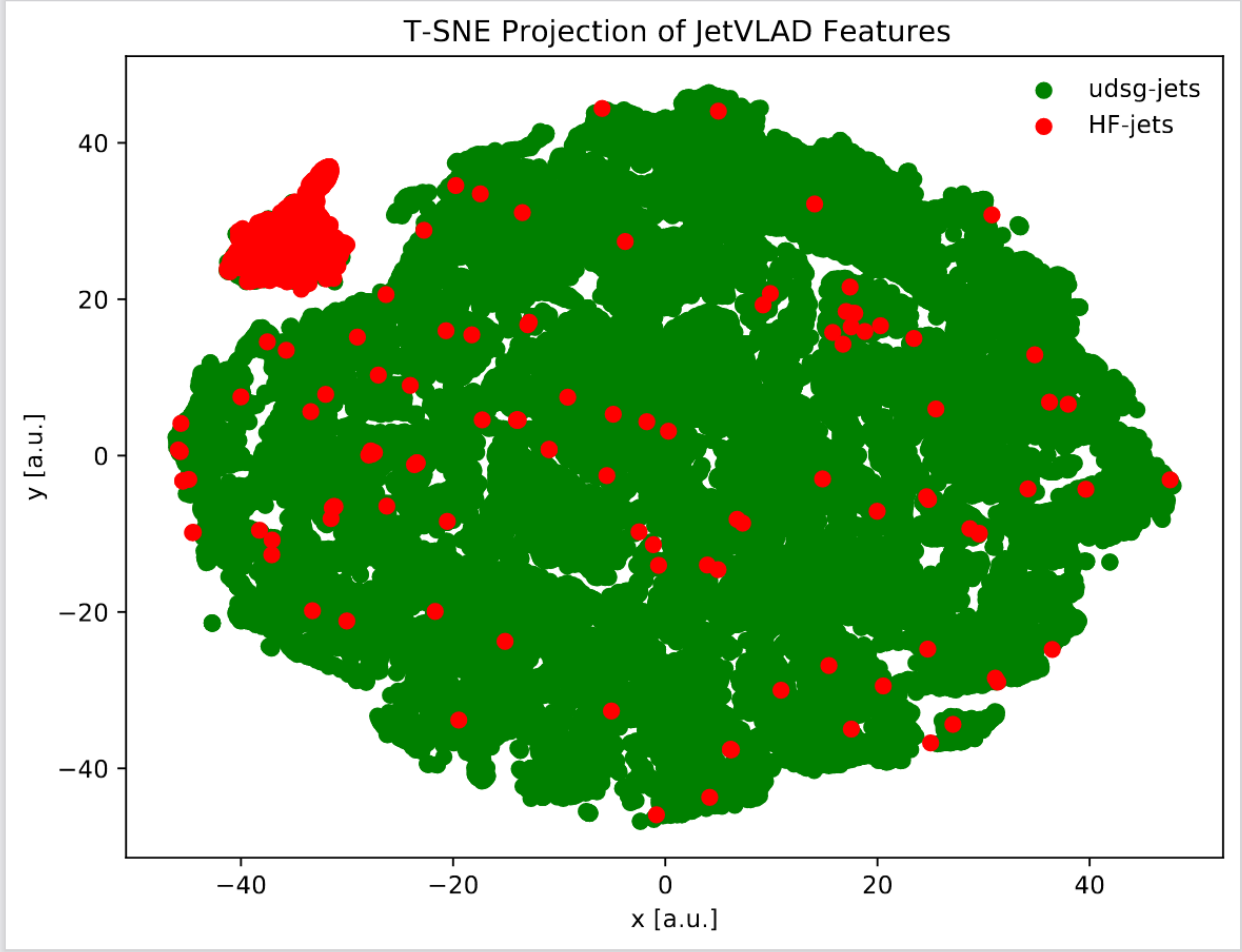


- Improvement of a factor 2-3 in background rejection compared to current leading models!



- DeepJet includes all secondary vertex info and particles along with global event features
- Improvement (factor of 2) attributed to algorithmic differences primarily in comparison to RNN (which are quite hard to train)

Performance benchmarks

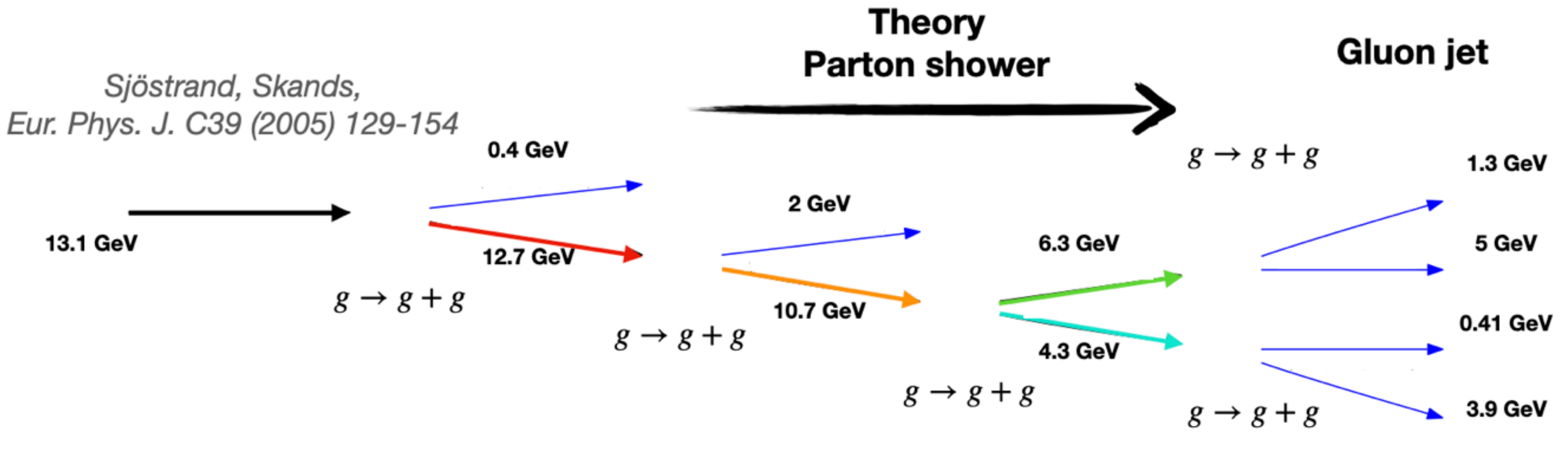


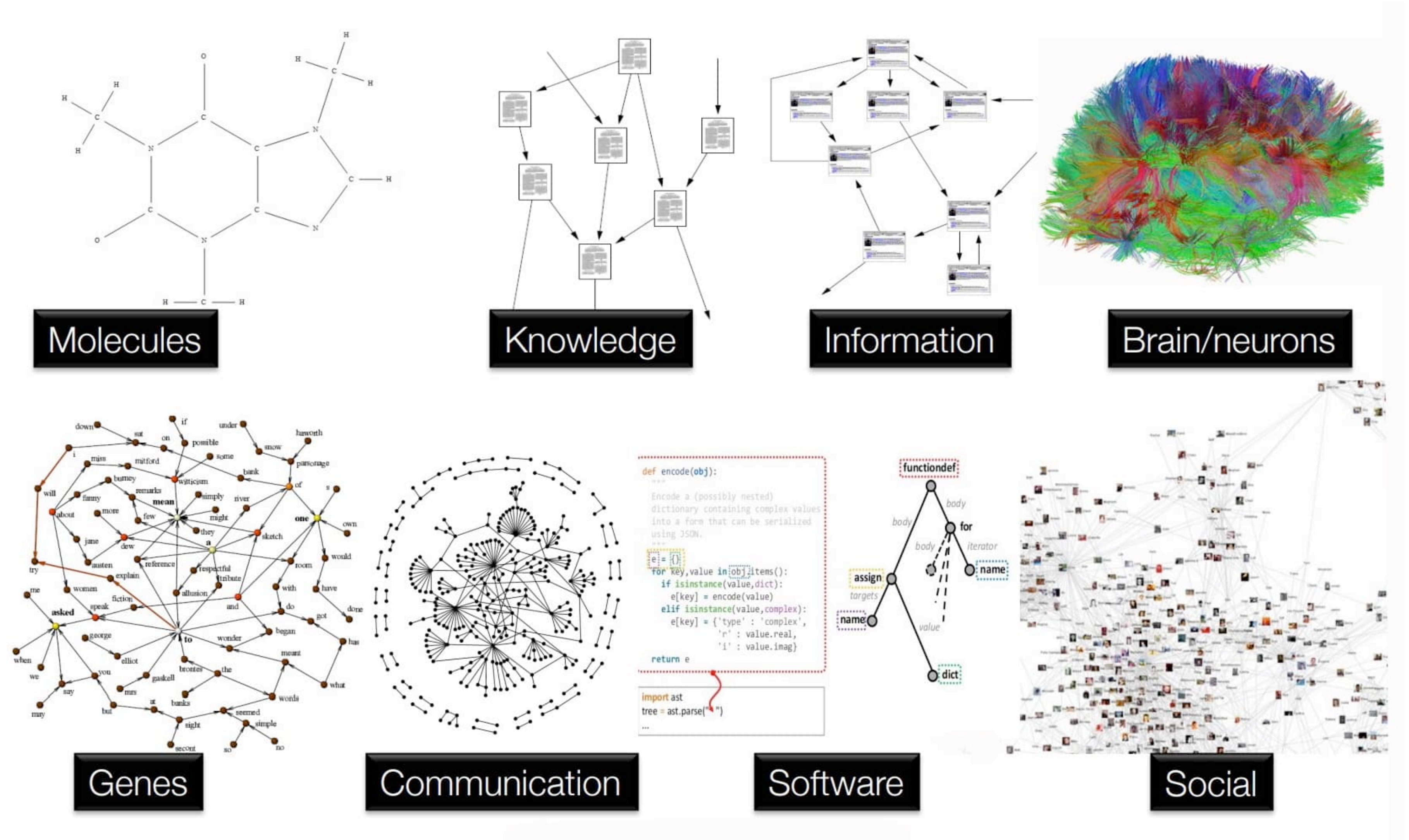
T-SNE projection

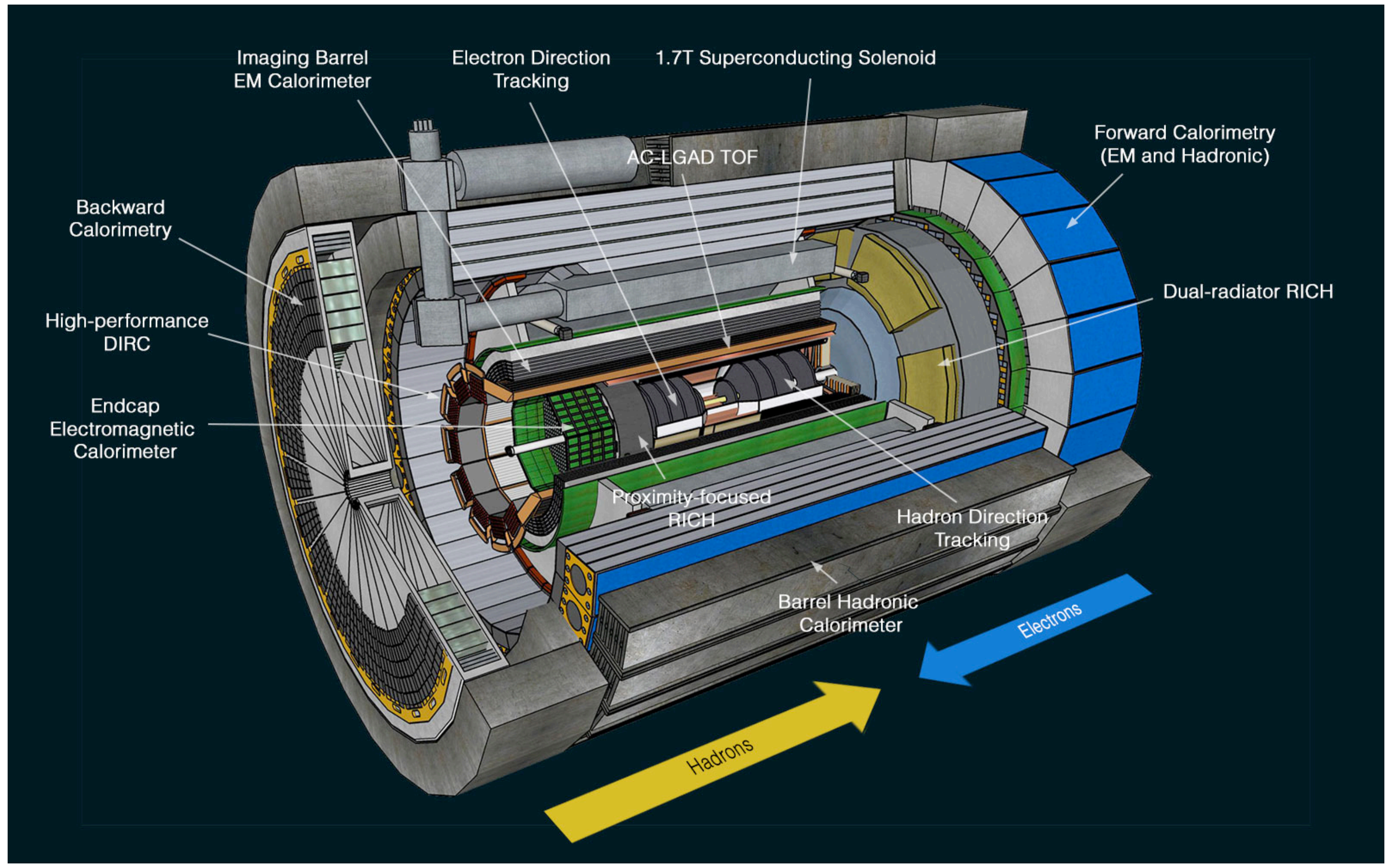
- Arbitrary projection from multi-dimensional phase-space to a 2D
- Isolated regions of overlap
- Further exploration in progress!

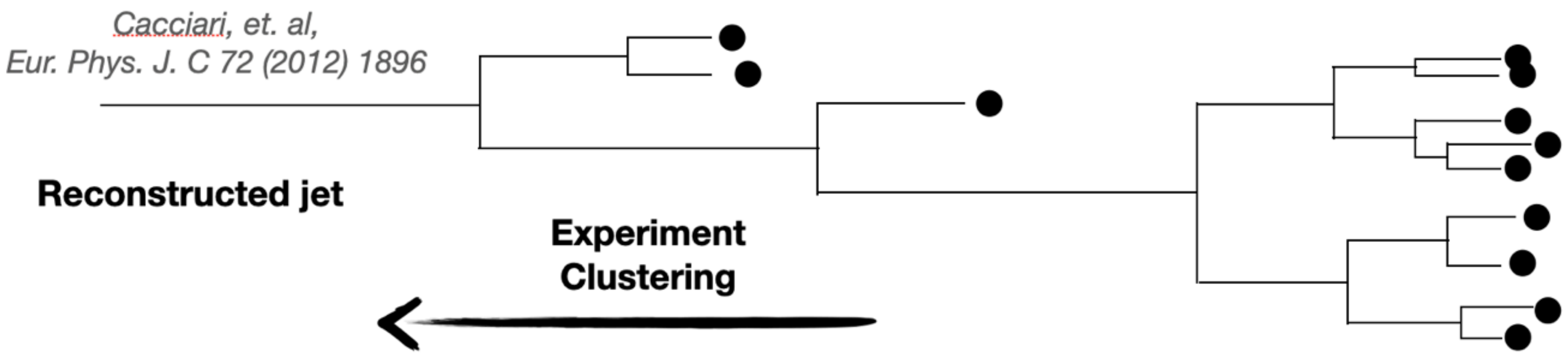
Evolution of a jet

Can we pick out fragmentation patterns?





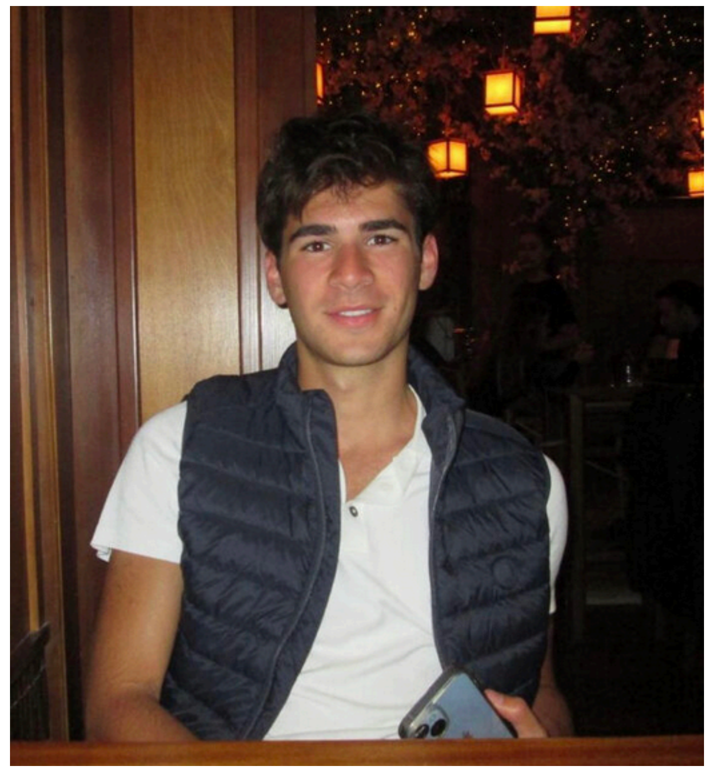




Multi dimensional pattern recognition

- if there are changes to a jet’s fragmentation, is that represented in the information space of a jet?
- Can we identify what those changes or ‘effects’ are specifically compared to a well understood baseline (read *pp* or ideally *ep*)?
- Can we translate those effects to a ‘cause’?
- Once we build up a library of possible causes, can we isolate specific sub-population of jets for future differential studies?

Jets as connected graphs



Michael Taleb,
Vandy Class of 2026

Umar Soheil Qureshi,
Vandy Class of 2025

Event Generation (PYTHIA 8.312)

- pp beams with $\sqrt{s} = 14$ TeV.
- Photon-tagged events $qg \to q\gamma$.
- $\hat{p}_T > 1000$ GeV.
- Anti-k_t $R = 0.8$ parton-level and hadron-level jets.
- Visible final-state particles.
- $1000 <$ Jet $p_\perp < 2000$ GeV.
- 100K events to ensure sufficient statistics.

pythia.org/latest-manual/welcome.html

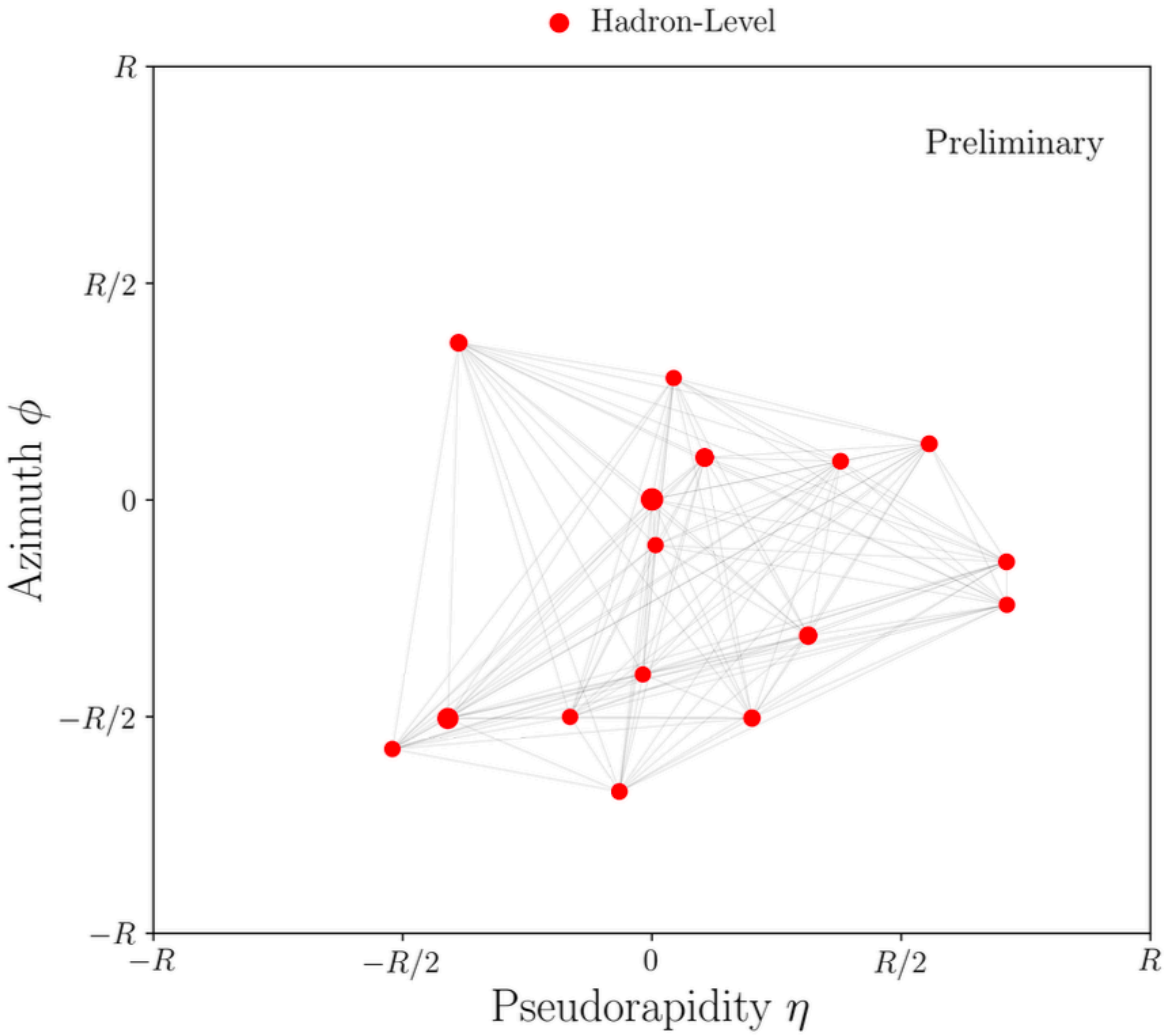
Graph Representation of Pythia Quark Jets

Jets represented as graphs, connected by ΔR:

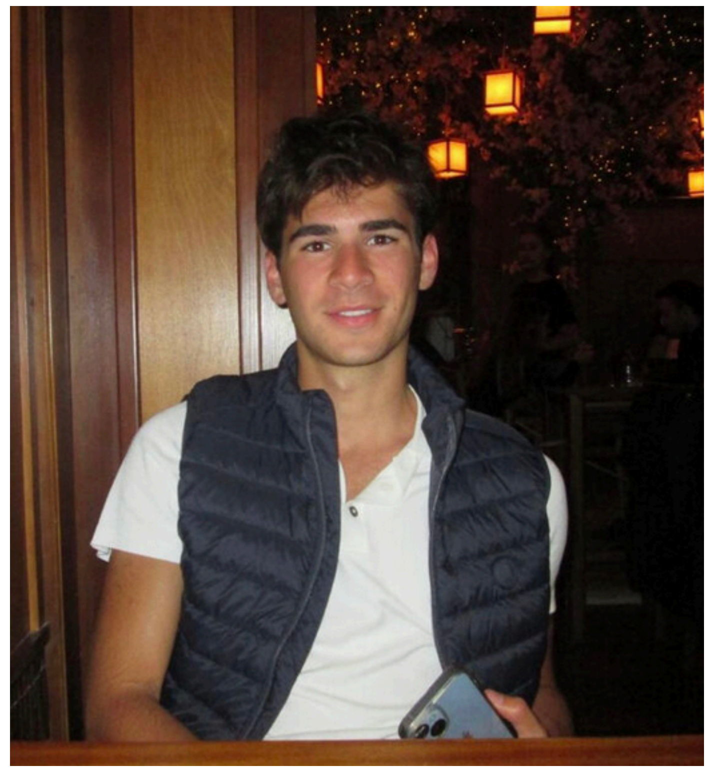
$$\text{Vertices}: \mathcal{J} = \left\{ (p_\perp^i, \eta^i, \phi^i)_{i=1}^n \right\}$$

$$\text{Edges}: E = \left\{ \Delta R(i,j)_{i,j=1}^n, i \neq j \right\}$$

Fully connected graphs, no self-loops.



Jets as connected graphs



Michael Taleb,
Vandy Class of 2026

Umar Soheil Qureshi,
Vandy Class of 2025

Event Generation (PYTHIA 8.312)

- pp beams with $\sqrt{s} = 14$ TeV.
- Photon-tagged events $qg \to q\gamma$.
- $\hat{p}_T > 1000$ GeV.
- Anti-k_t $R = 0.8$ parton-level and hadron-level jets.
- Visible final-state particles.
- $1000 <$ Jet $p_\perp < 2000$ GeV.
- 100K events to ensure sufficient statistics.

pythia.org/latest-manual/welcome.html

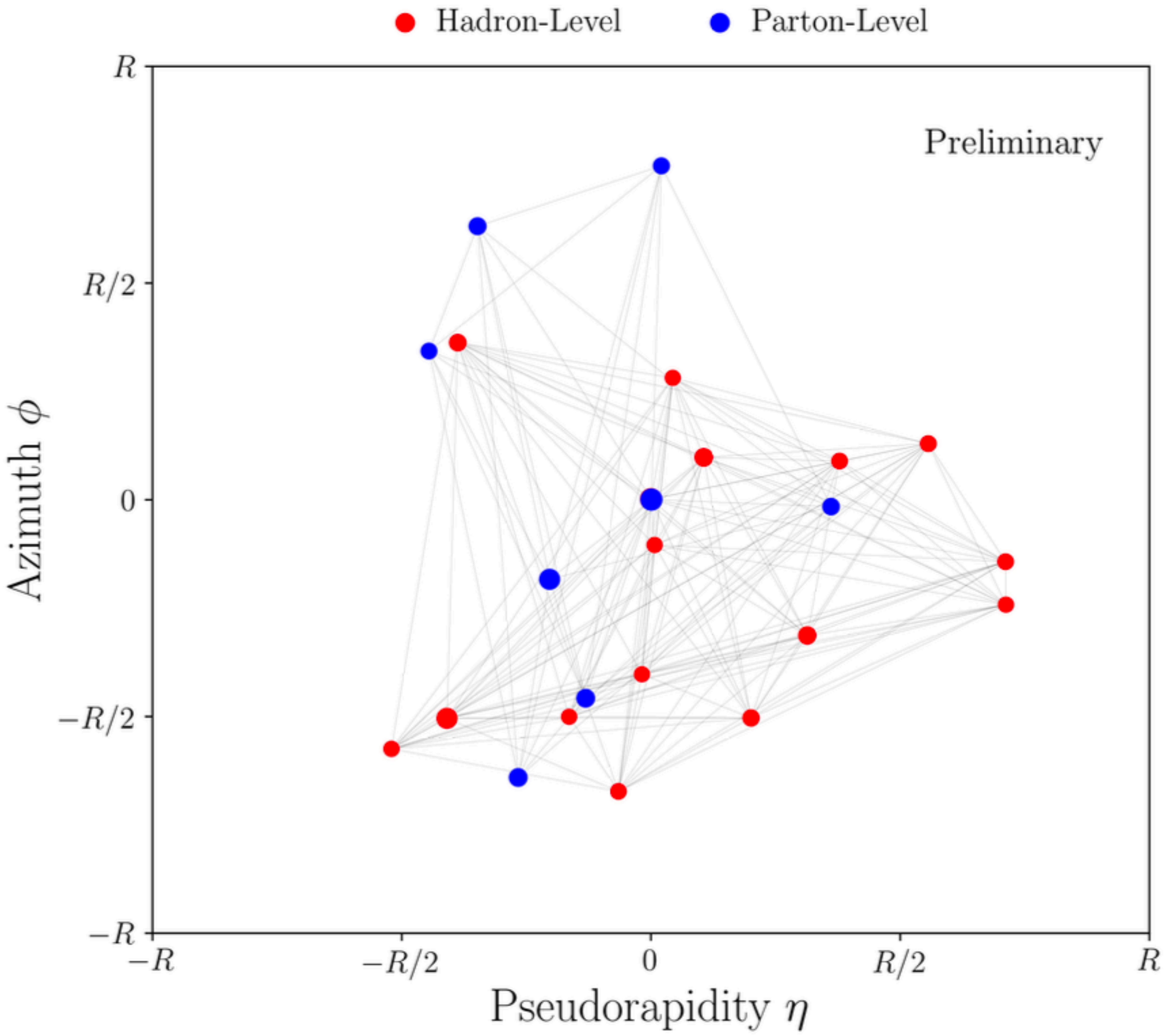
Graph Representation of Pythia Quark Jets

Jets represented as graphs, connected by ΔR:

$$\text{Vertices} : \mathcal{J} = \left\{ (p_\perp^i, \eta^i, \phi^i)_{i=1}^n \right\}$$

$$\text{Edges} : E = \left\{ \Delta R(i,j)_{i,j=1}^n, i \neq j \right\}$$

Fully connected graphs, no self-loops.



Mapping one graph to another graph via latent space representation

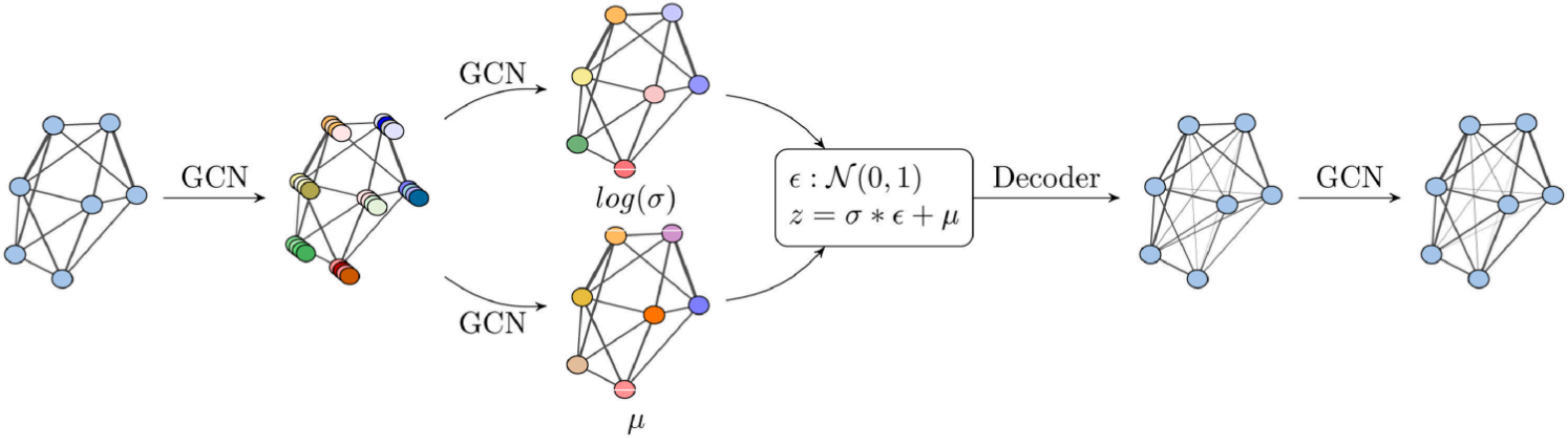


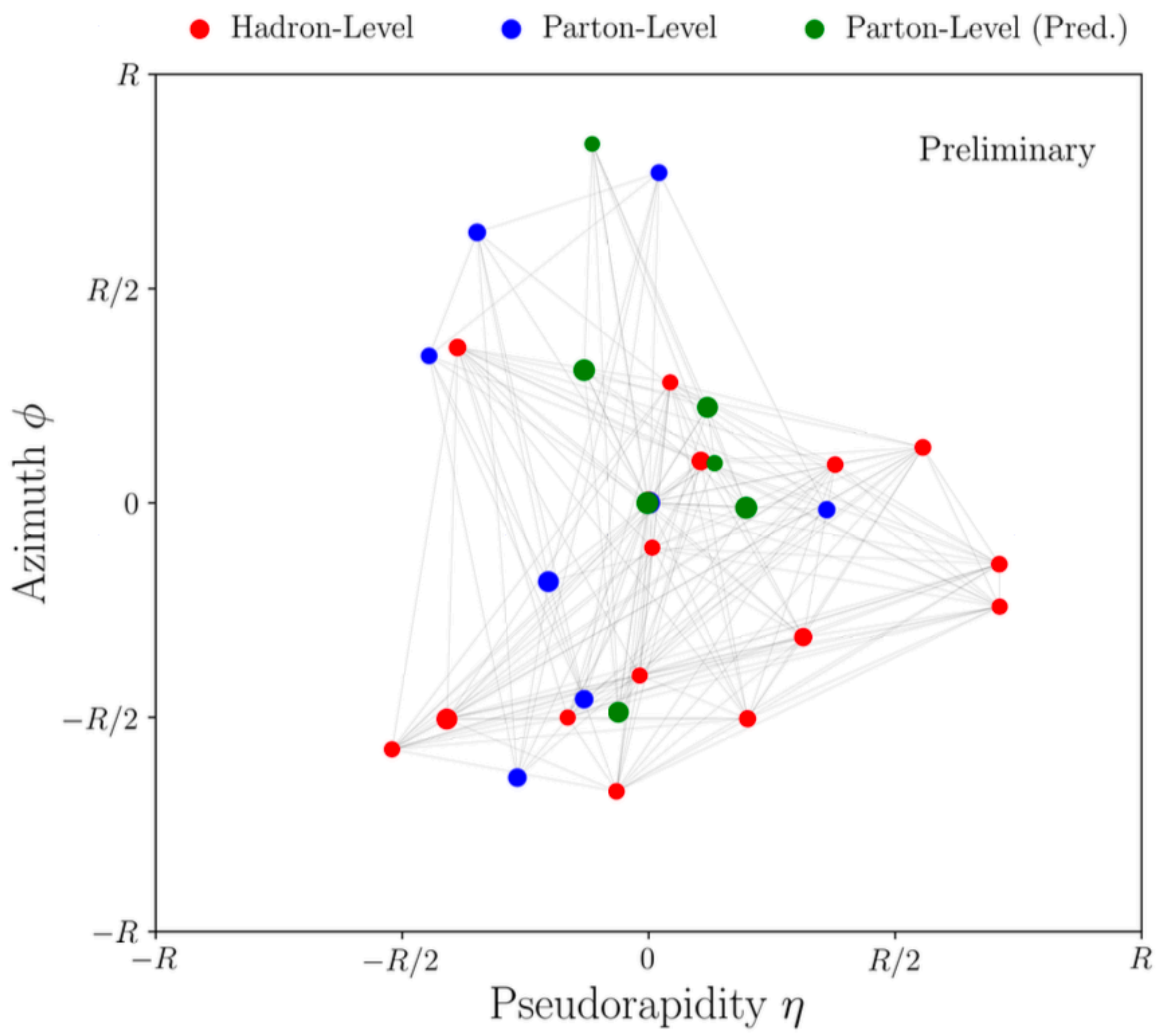
Image Credit: Tina Behrouzi et. al.

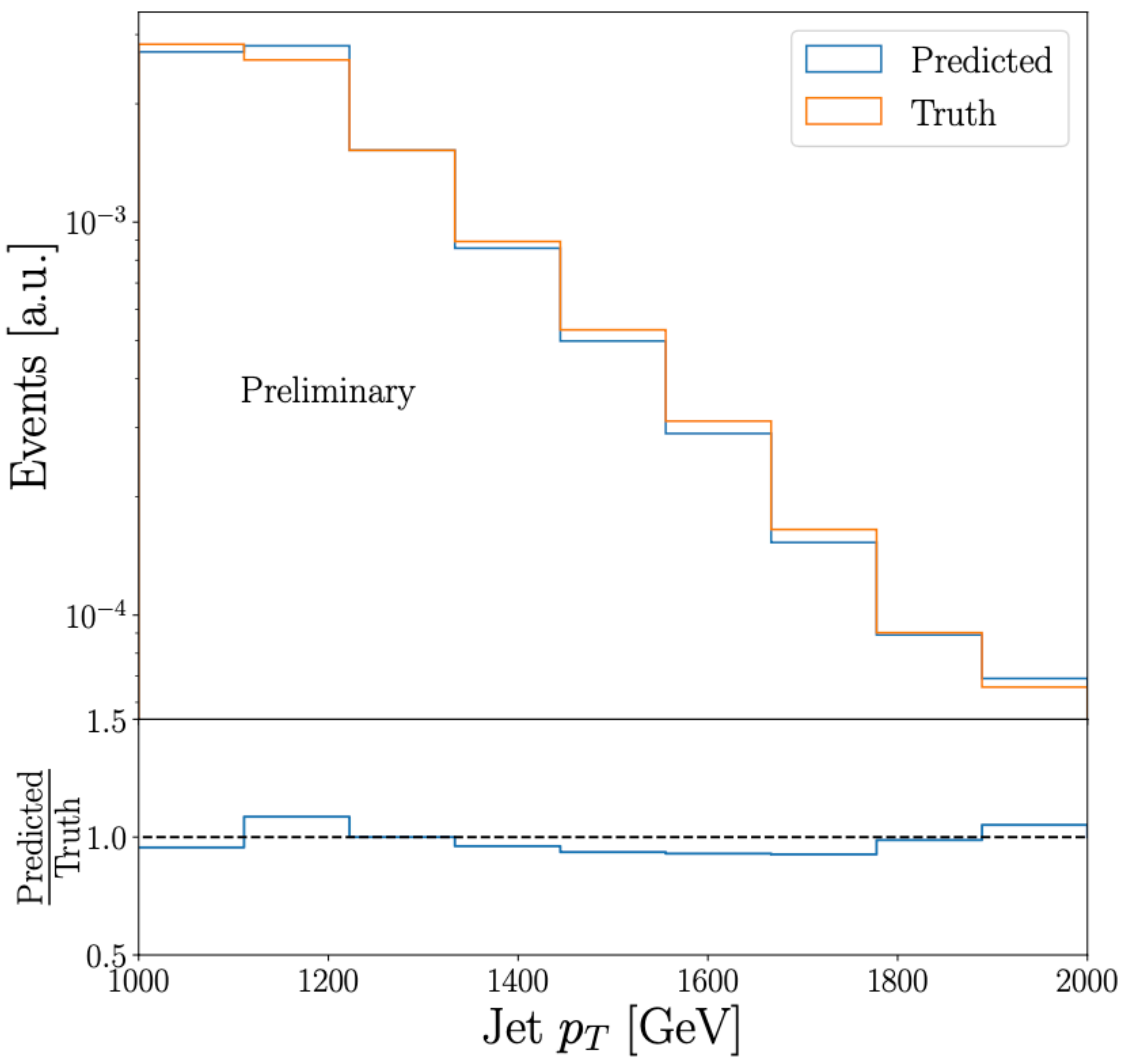
Variational Graph Autoencoder (VGAE)

- Input hadron-level jets $\mathcal{H}$.
- Output parton-level jets $\mathcal{P}$.

- Encoder: learns an embedding (z, μ) for $\mathcal{H}$ in latent space.
- Decoder: learns reconstructing parton-level jets $\mathcal{P}$ from embedding.

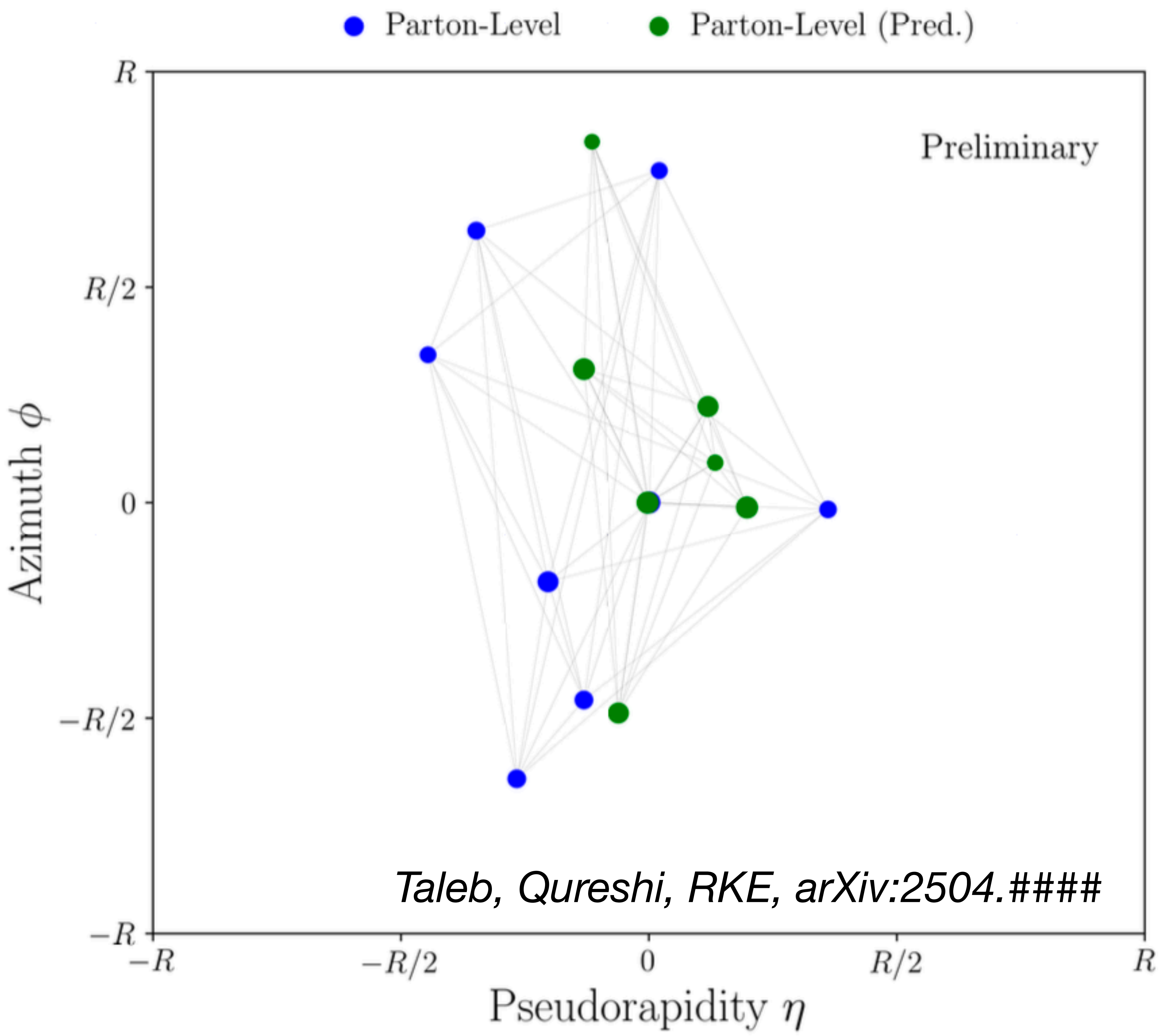
Predicted Parton level Jet





- We can more/less get the scalar jet momenta, but…

How similar are these?



EMD Metric (PRL 123.041801)

- Quantifies the distance between two jets.
- The minimum "energy" required to rearrange a jet $\mathcal{G}$ to $\mathcal{G}'$.

$$\mathcal{E}(\mathcal{G},\mathcal{G}') = \min_{\{f_{ij}\geq 0\}} \sum_{i=1}^{M}\sum_{j=1}^{M'} f_{ij}\left(\frac{\Delta R_{ij}}{R}\right) + \left|\sum_{i=1}^{M} E_i - \sum_{j=1}^{M'} E'_j\right|,$$

$$\sum_{j=1}^{M'} f_{ij} \leq E_i, \quad \sum_{i=1}^{M} f_{ij} \leq E'_j, \quad \sum_{i=1}^{M}\sum_{j=1}^{M'} f_{ij} = E_{\min},$$

$\mathcal{E}(\widehat{\mathcal{P}},\mathcal{P})$ gives a discrepancy measure between reconstructed graphs $\widehat{\mathcal{P}}$ and the ground truth $\mathcal{P}$.

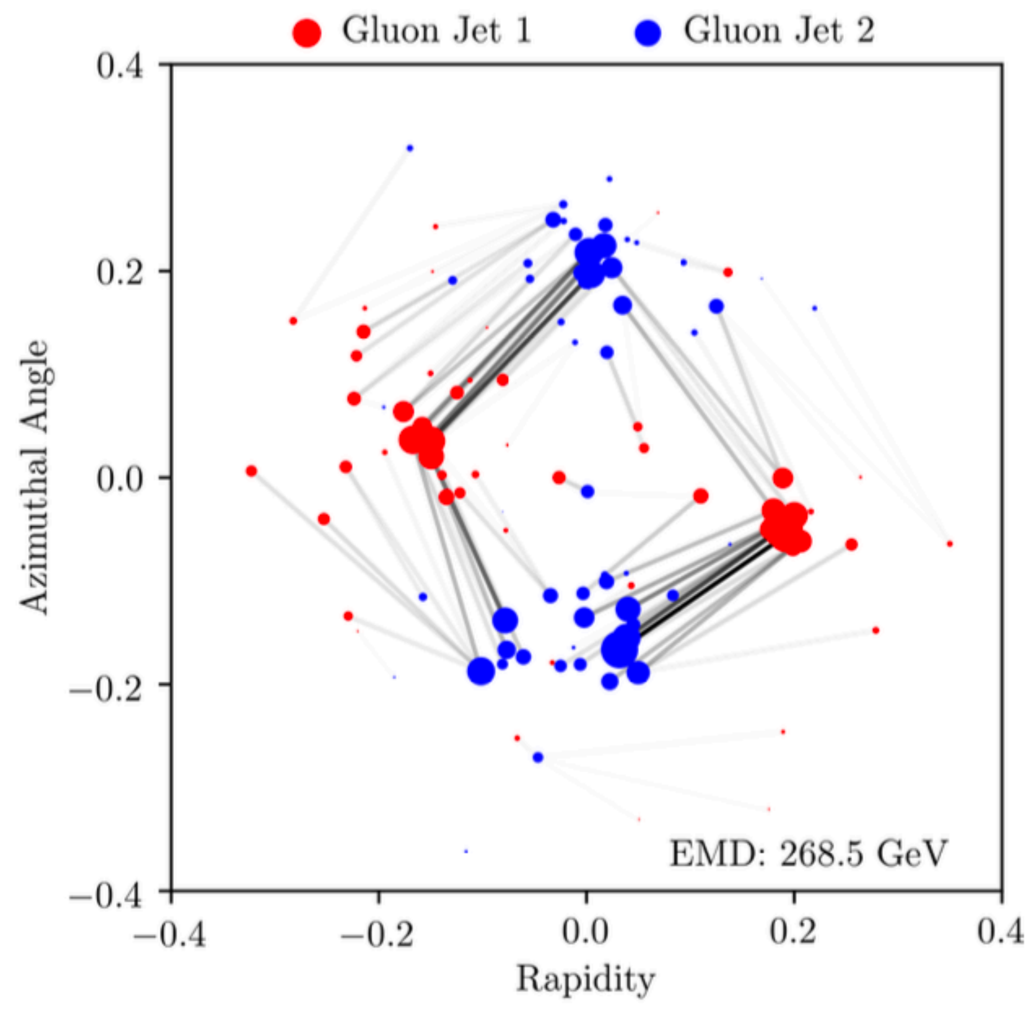
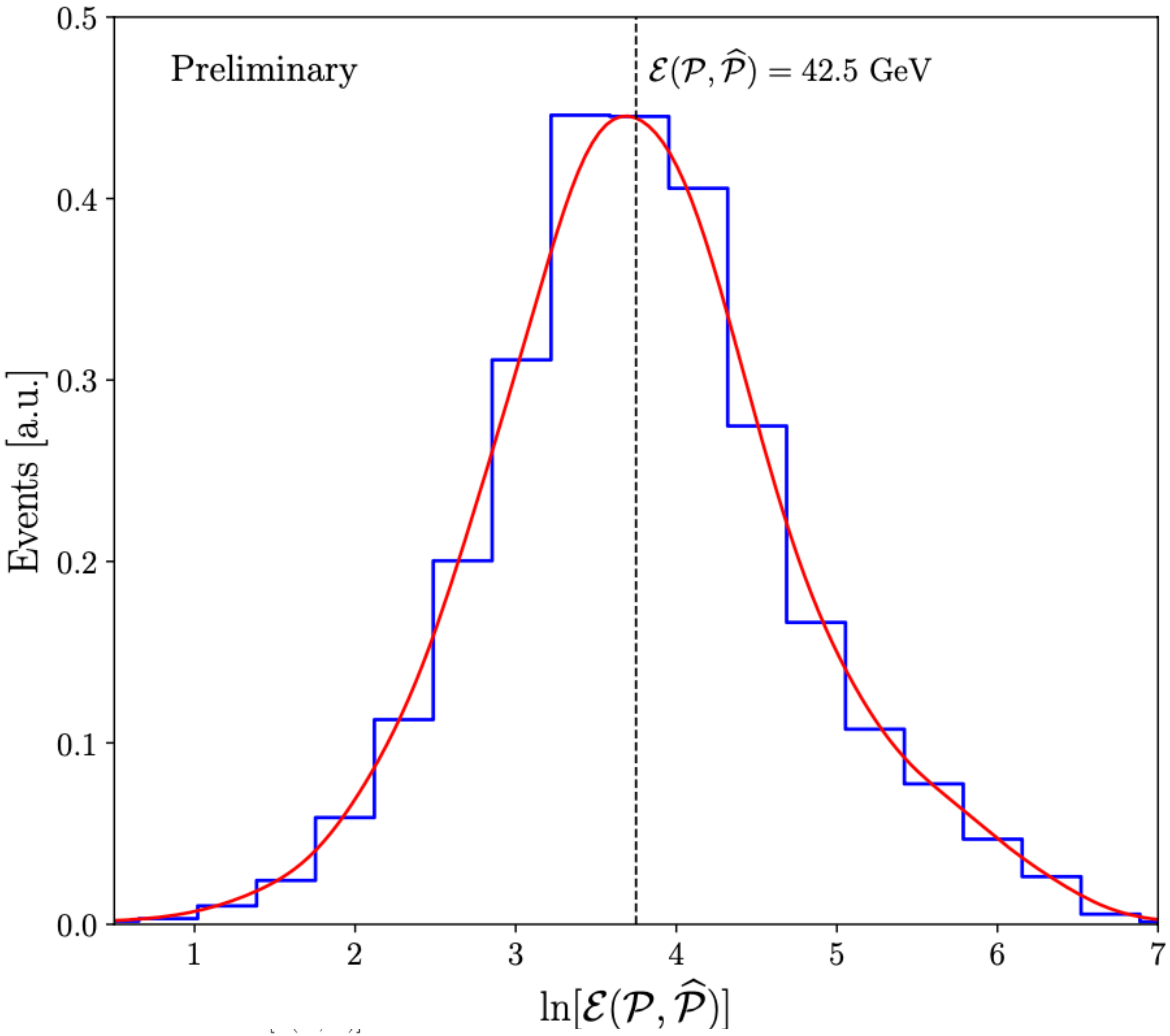


Figure 5: EMD between two gluon jets.

- EMD essentially estimates how much 'work' you need to move one to another

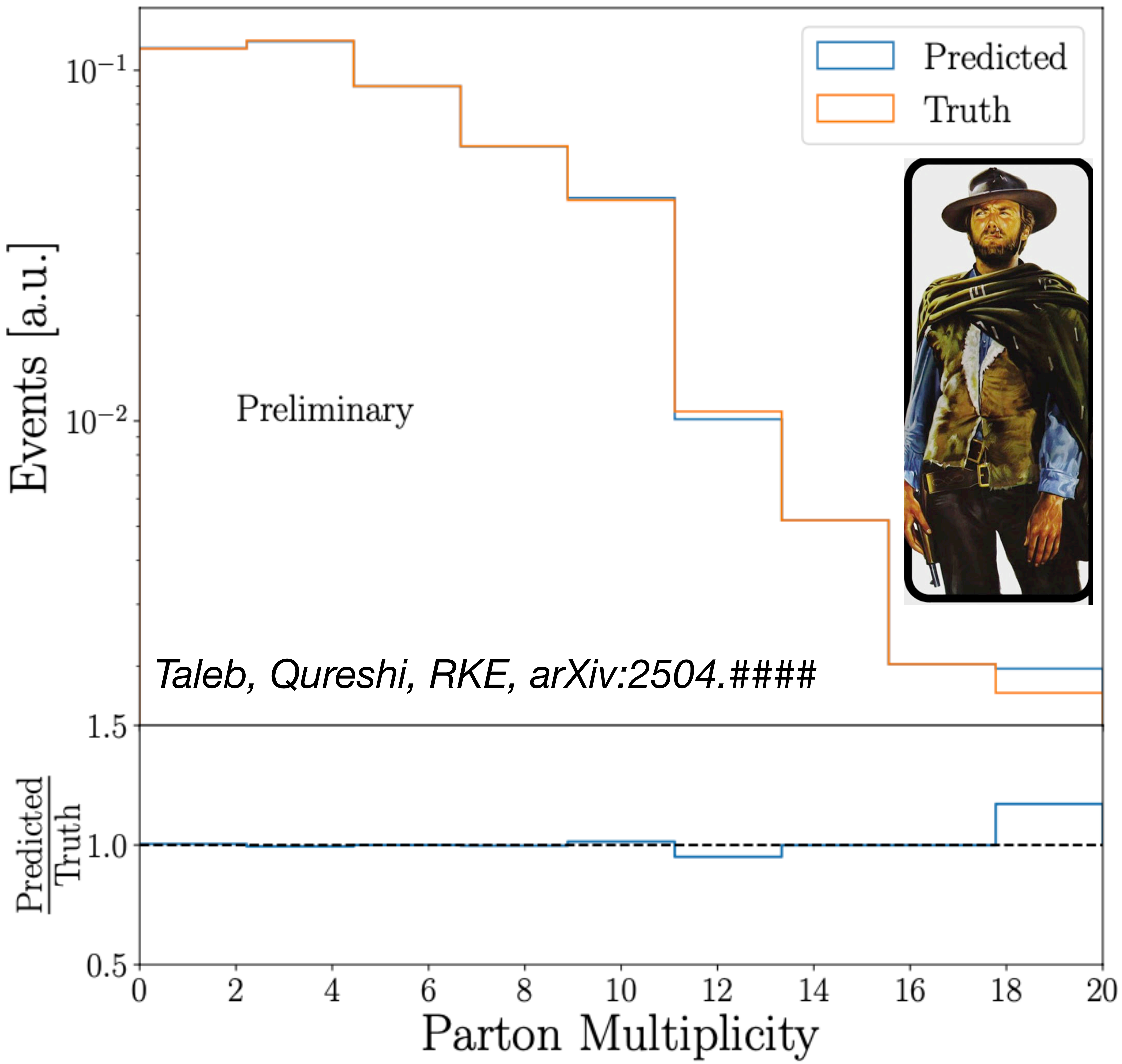


Predicted jets close to ground truth (Pythia)!

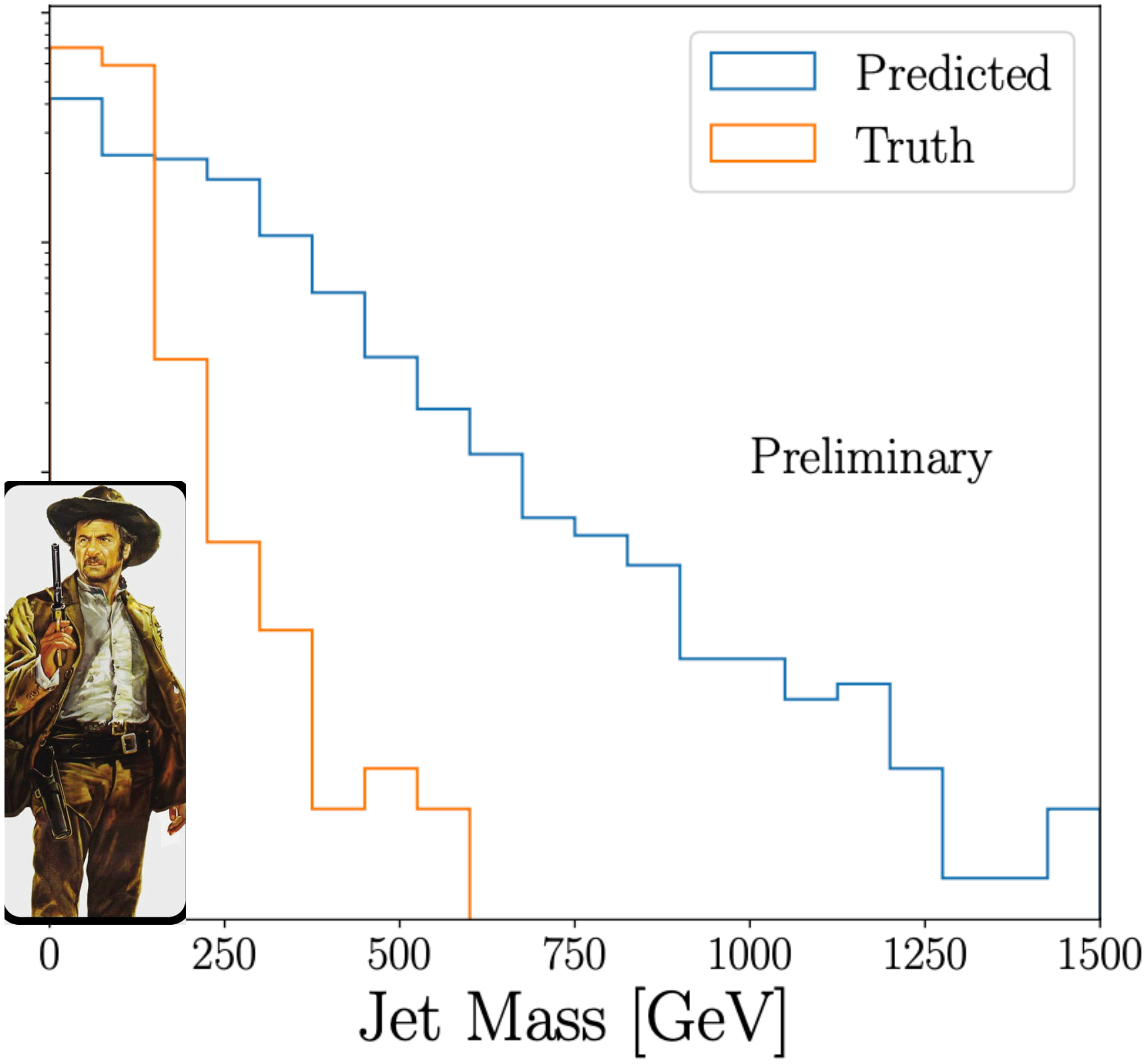
Benchmark EMDs:

- Good: $\ln \mathcal{E} \leq 4$
  - Jets are similar.
- Fair: $4 \leq \ln \mathcal{E} \leq 5.5$
  - Jets are fairly similar.
- Bad: $\ln \mathcal{E} \geq 5.5$
  - Jets are disparate.

Good


Predicted
Truth
Events [a.u.]
10^{-1}
10^{-2}
Preliminary
Taleb, Qureshi, RKE, arXiv:2504.####
1.5
1.0
0.5
$\frac{\text{Predicted}}{\text{Truth}}$
0 2 4 6 8 10 12 14 16 18 20
Parton Multiplicity

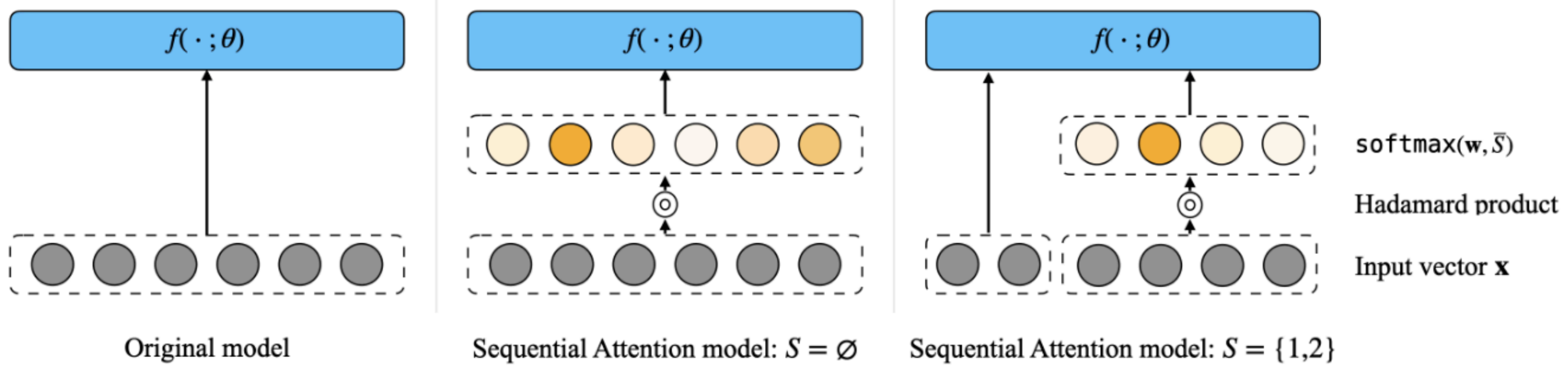
Ugly


Predicted
Truth
Preliminary
0 250 500 750 1000 1250 1500
Jet Mass [GeV]

Detour - Can we identify jets that have modified fragmentation

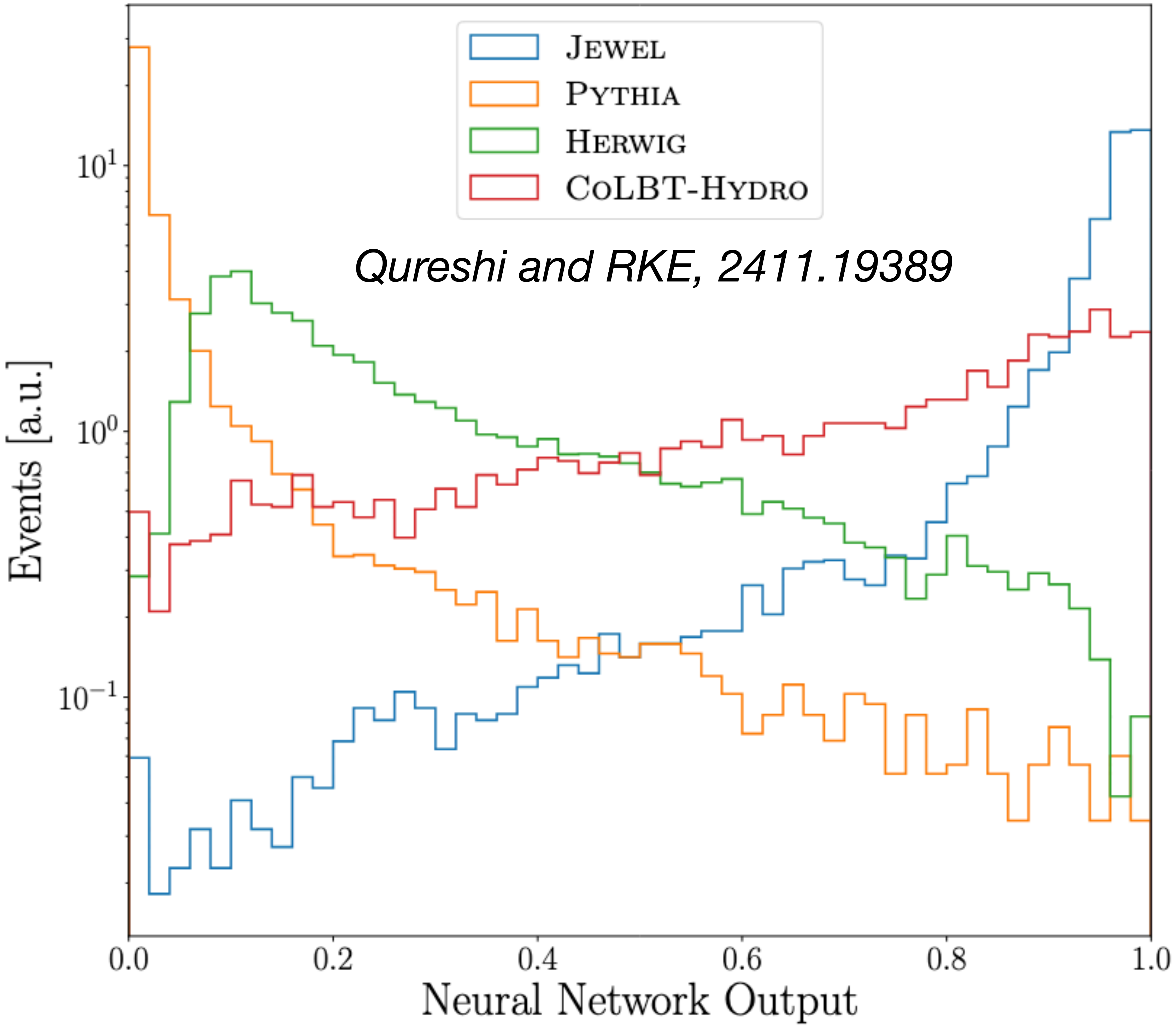
Umar Soheil Qureshi, Vandy Class of 2025

Yasuda, T et. al 2209.14881



- greedy forward selection algorithm, which repeatedly selects the feature with the largest marginal improvement
- introducing a new set of trainable variables w ∊ Rd that represent feature importance

Model	Thermal Background	Detector Effects	Pileup	Performance (AUC)	Reference
Energy Flow Network	×	×	×	0.67	[32]
Particle Flow Network	×	×	×	0.86	[32]
Particle Flow Network	✓	×	×	0.75	[32]
Long-Short Term Memory	✓	×	×	0.76	[30]
Long-Short Term Memory	✓	×	×	0.74	[31]
Multi-Layer Perceptron	✓	×	×	0.73	[31]
Autoencoder + Decision Tree	✓	×	×	0.70	[33]
Convolutional NN	✓	×	×	0.75	[31]
Sequential Attention	✓	✓	✓	**0.95**	Our Study



Detour - Can we identify jets that have modified fragmentation

Umar Soheil Qureshi, Vandy Class of 2025

- Not all inputs are made the same
- Motivate selective observables to go and measure!

$$\mathcal{J} = \{n, m, p_\perp, z_g, R_g, k_\perp, m_g\} \bigoplus_{i=1}^{n} \{p_\perp^i, \eta^i, \phi^i, \mathrm{PID}^i\}$$

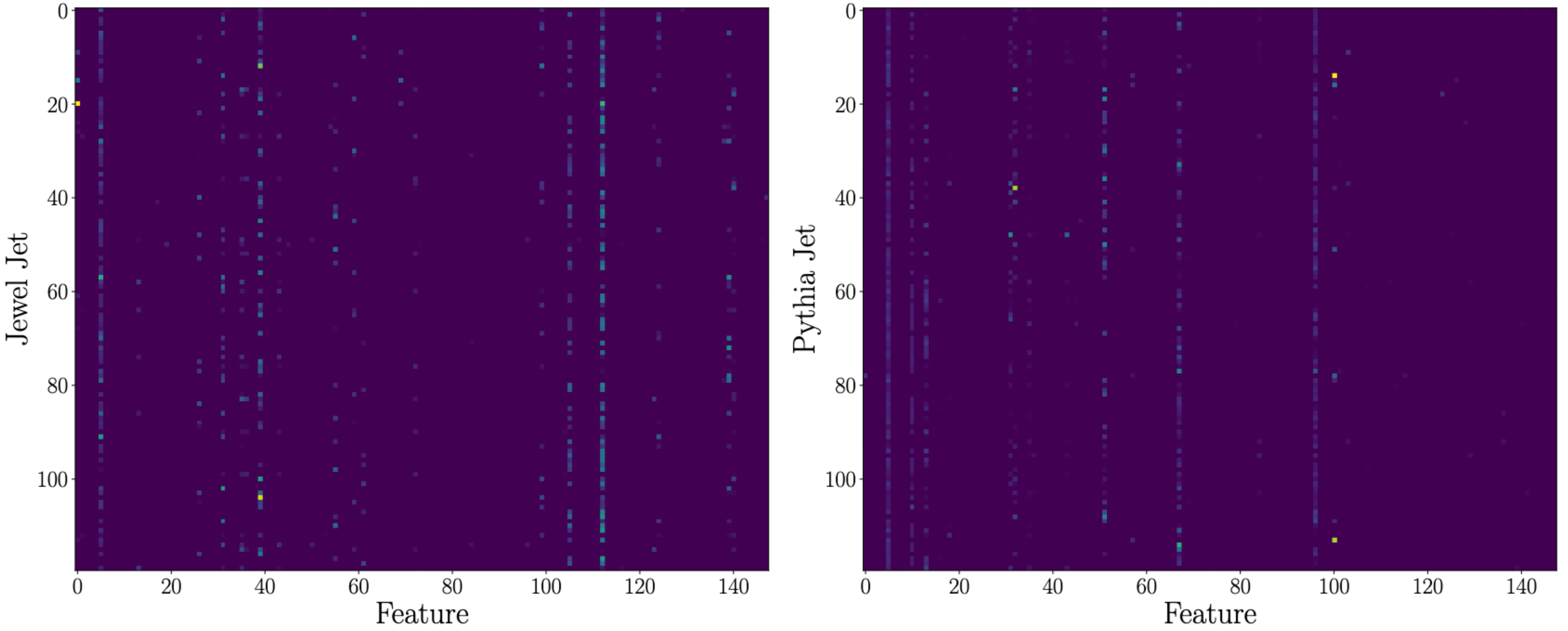


FIG. 4. Heatmap illustrations of the aggregate feature mask (Eq. 8) for the first 125 JEWEL (left) and PYTHIA (right) truth jets. The sparsity in feature activation highlights the attention-based mechanism's focus on relevant features for classification.

Qureshi and RKE, 2411.19389

Jet in a background

Signal vs background

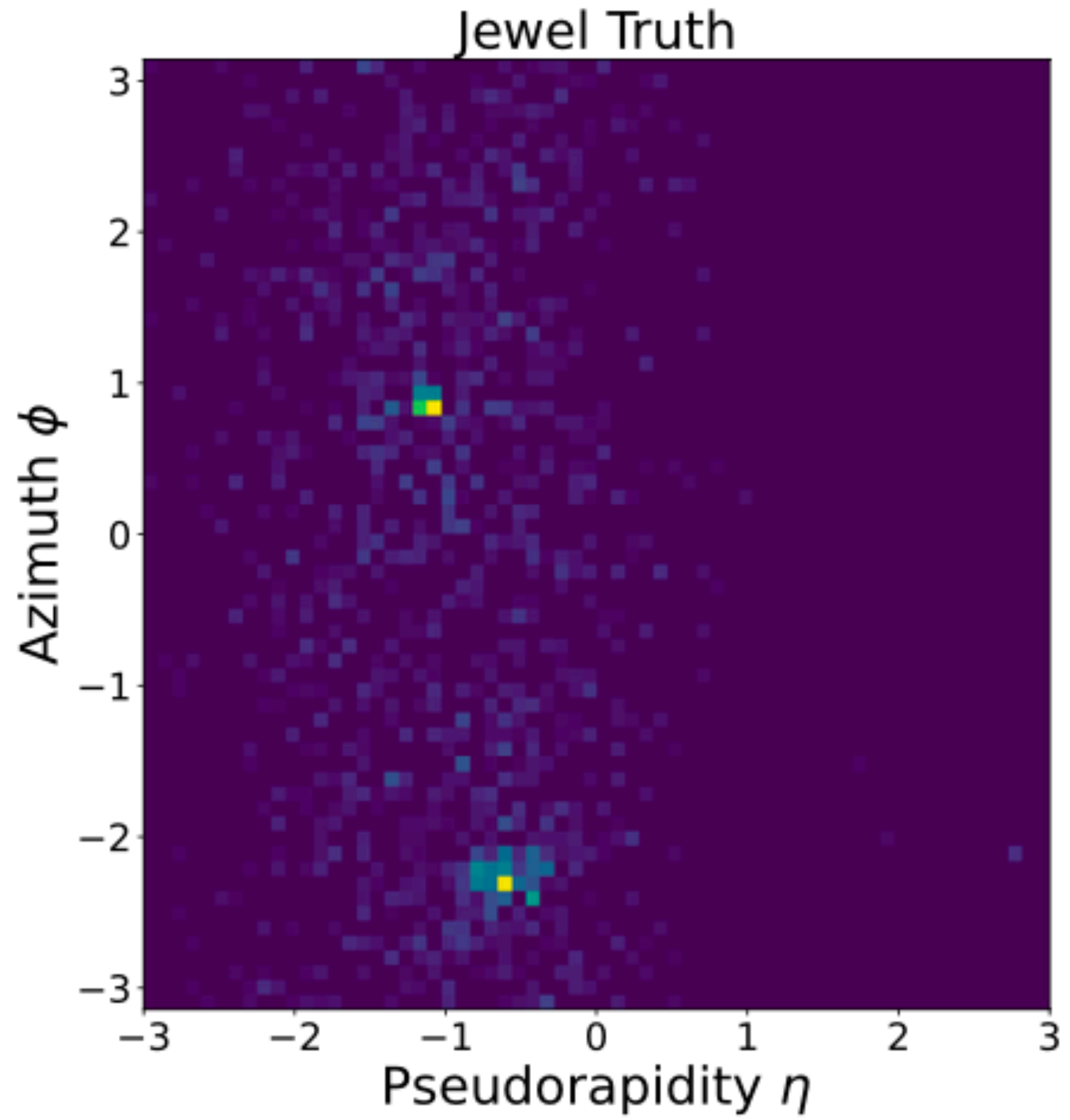
Events as noisy images

Event Generation (JEWEL)

- $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV PbPb beams.
- Dijets at 0-10% centrality.
- $\hat{p}_{\mathrm{T}} > 100$ GeV.
- 100K events to ensure sufficient statistics for ML training.

Thermal Background

- 15k particles uniform over $|\eta| < 3$.
- ϕ-Modulation with $v_2 = 0.05$.
- Boltzmann distribution in p_{T} with $\langle p_{\mathrm{T}} \rangle = 1.2$ GeV.



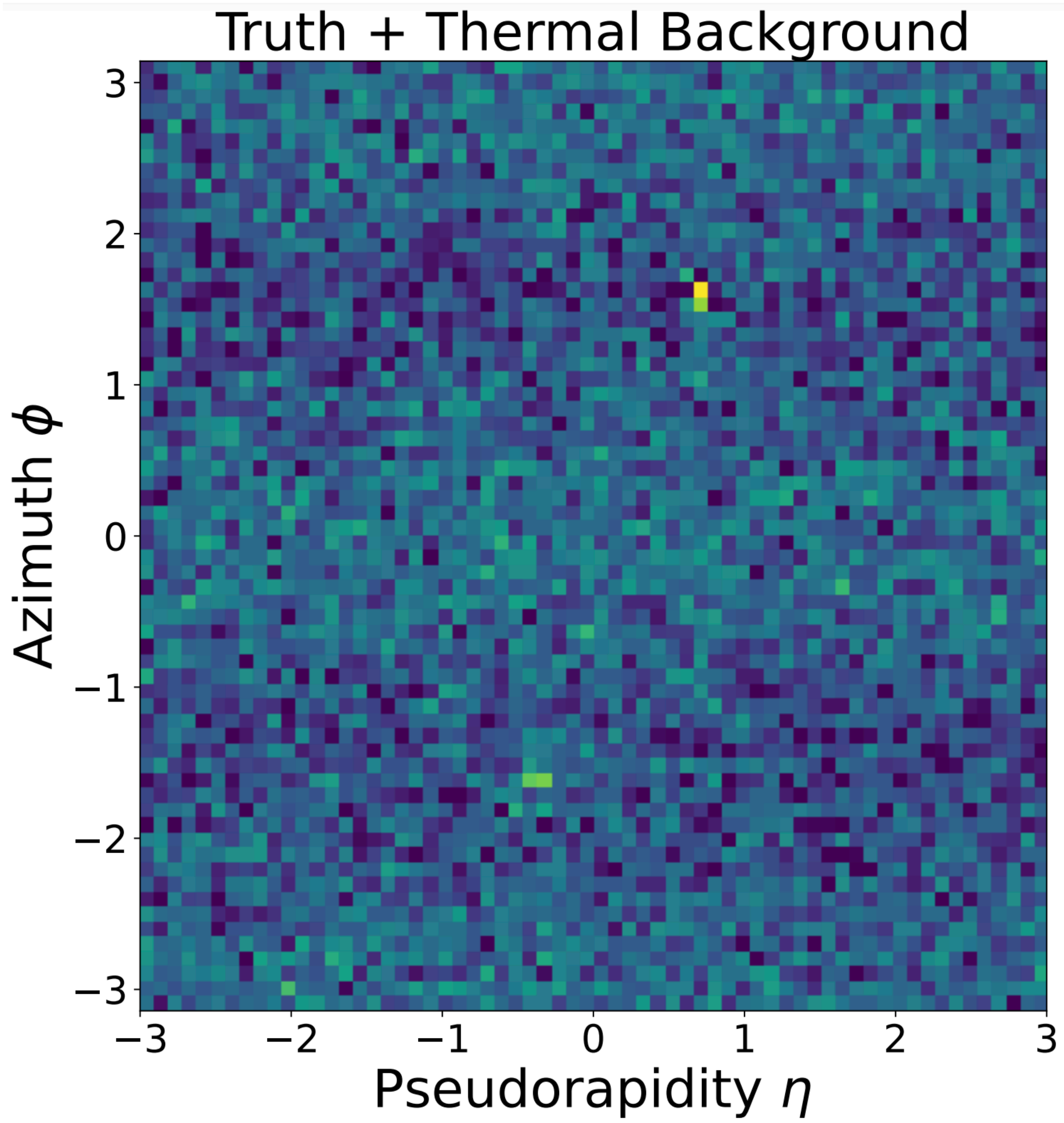


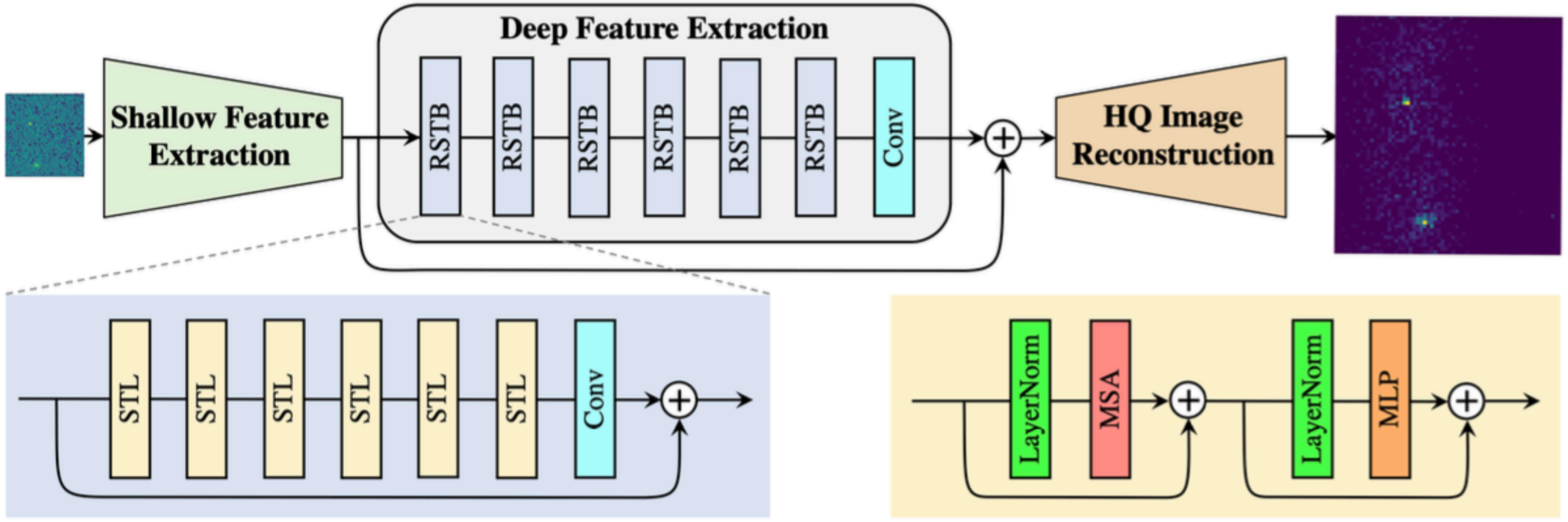
Image Representations

- Dijet events as images in the $(\eta, \phi) \in [-3, 3] \times [-\pi, \pi]$ plane.

Umar Soheil Qureshi, Vandy Class of 2025

Hierarchical Vision Transformer using Shifted Windows

Liu, Ze et.al 2103.14030



(a) Residual Swin Transformer Block (RSTB)

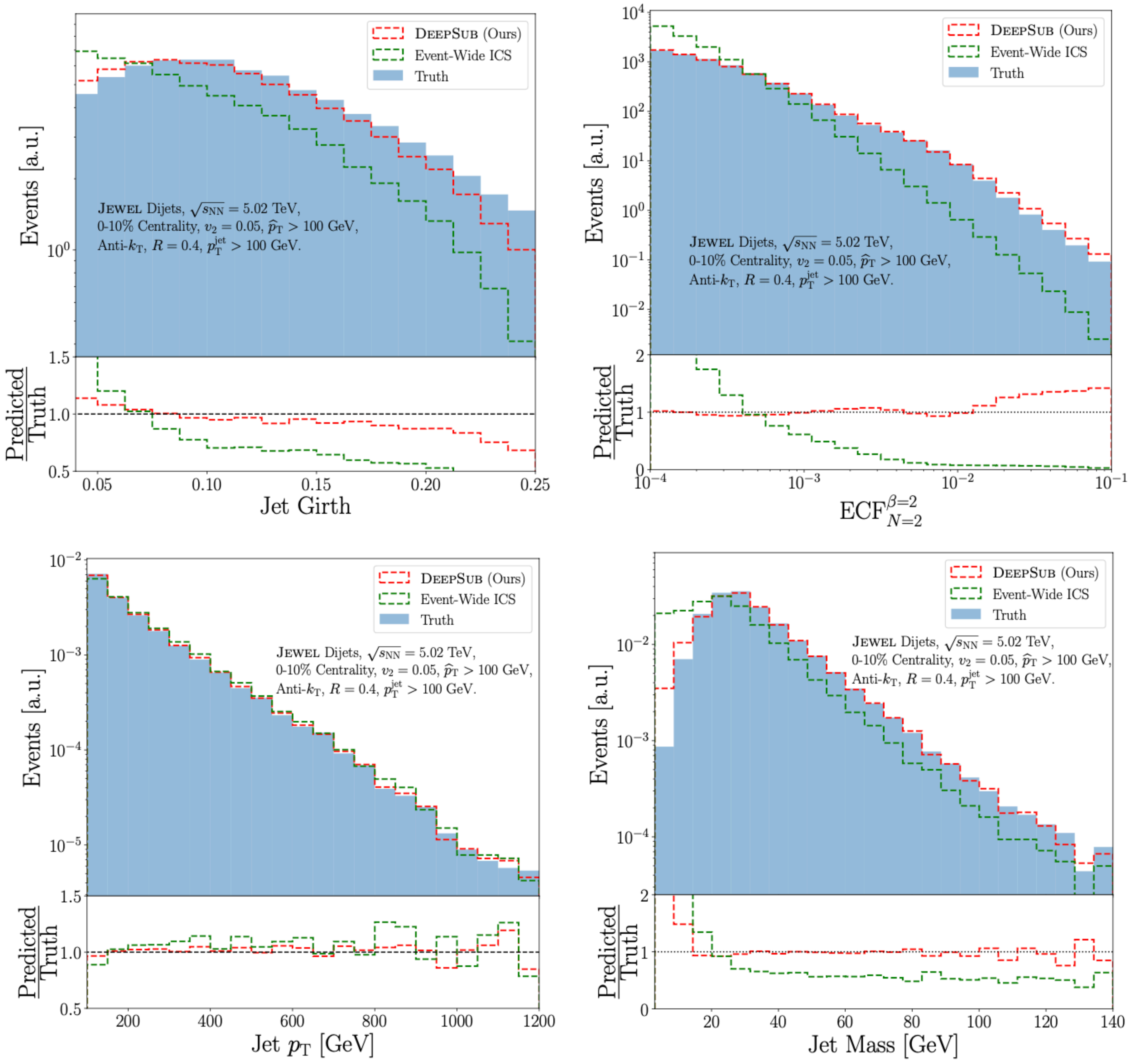
(b) Swin Transformer Layer (STL)

- The shifted windowing scheme brings greater efficiency by limiting self-attention computation to non-overlapping local windows while also allowing for cross-window connection.

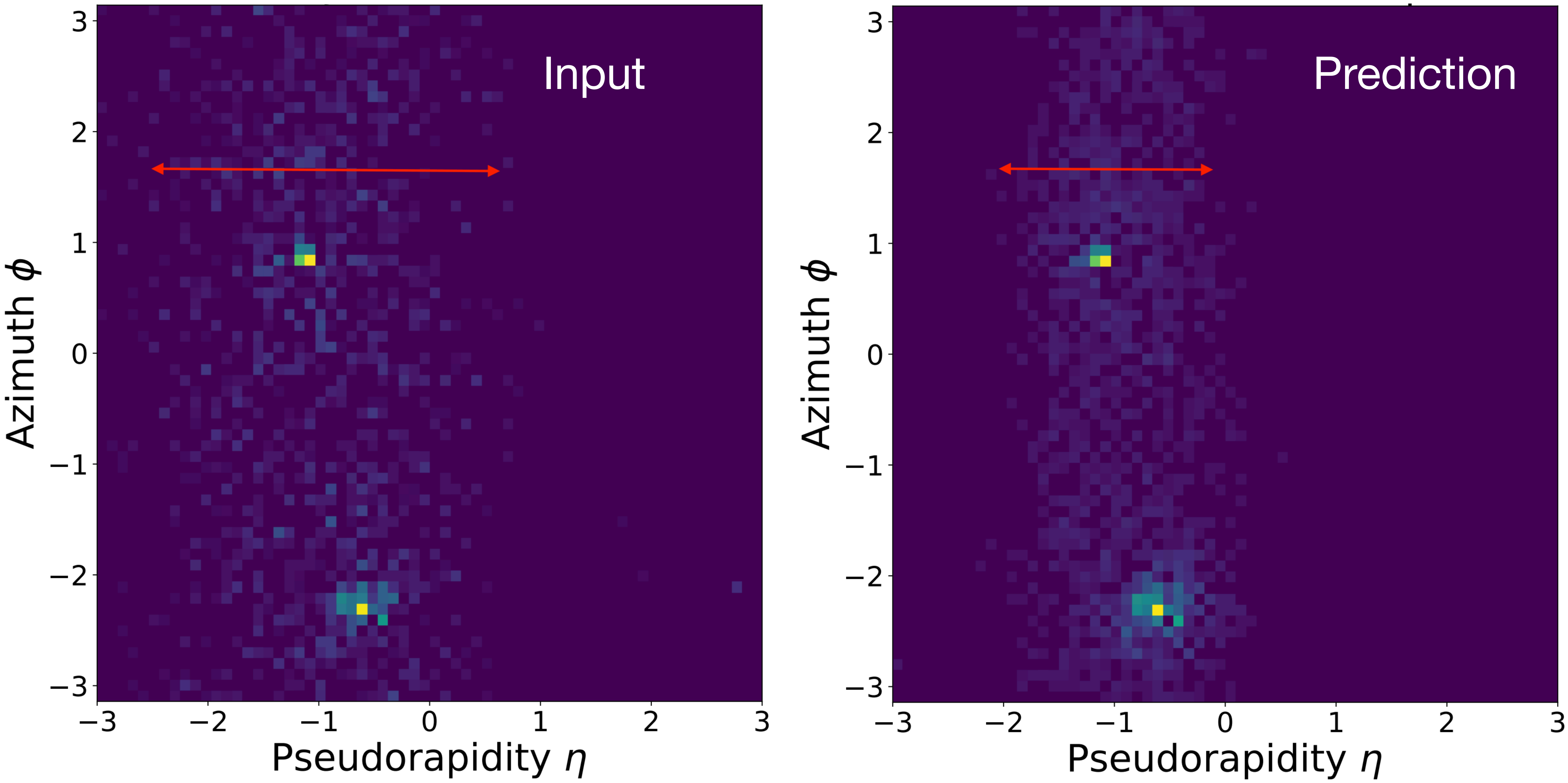
- This hierarchical architecture has the flexibility to model at various scales and has linear computational complexity with respect to image size.

Comparing to existing methods

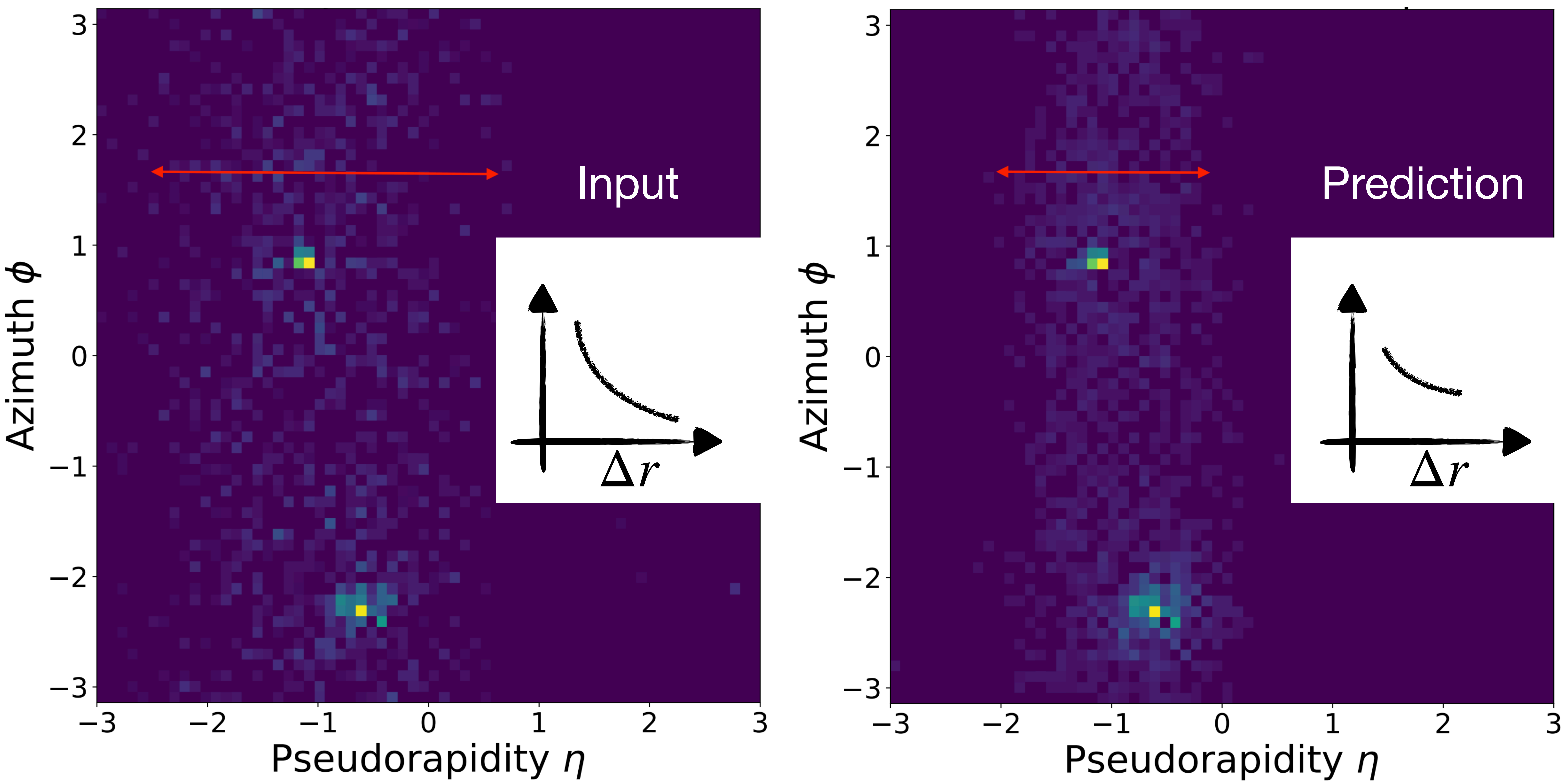
- Jets almost always have steeply falling distributions which make it hard for model predictions to get right
- Scalar quantities as always are *very* good, BUT 4-momentum distributions are difficult since they are sensitive to low p_T objects



Whats wrong here?



Need physics-motivated loss functions



Conclusion

- We are on the roadmap towards discovery with the EIC
- We are building systems now that will enable fast physics extraction with specific models that answer specific questions
- **Very few questions are solved out of the box**
- Jets are multi-scale, multi-dimensional, information (n) sparse but dimensionally dense and are a good laboratory for study these questions
- **Different jets are different - we need physics motivated models**
- EIC will teach us a lot of physics - but it will also be a very pure baseline for comparison with current pp or pA or AA jets!