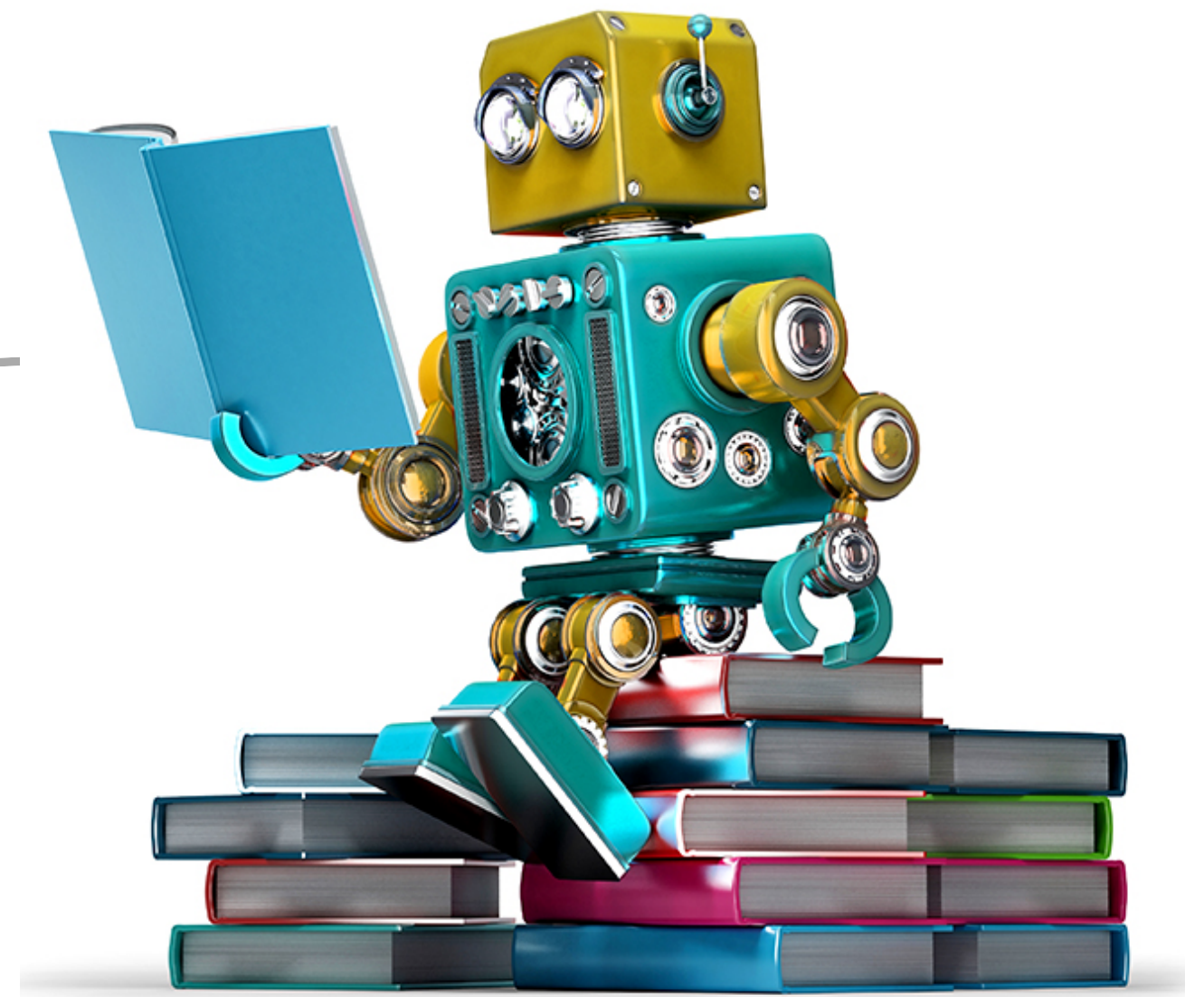
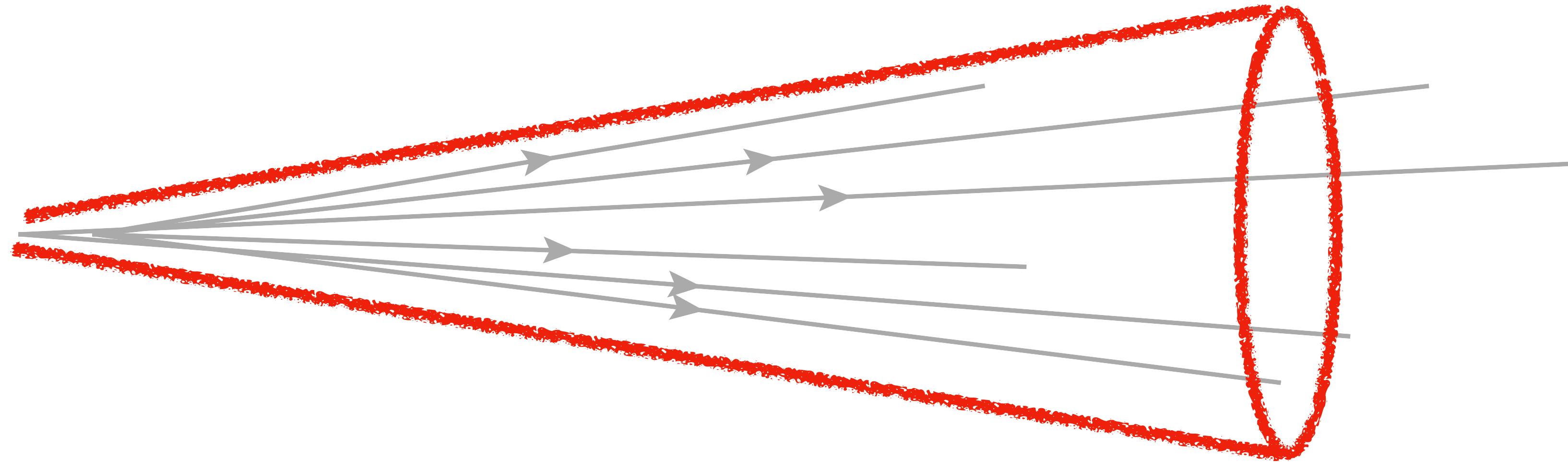




From Basics to Advanced

a short course on ML applications with jets in HE(N)P

AI/ML Workshop @ Lehigh July 20th, 2026

Rithya Kunnawalkam Elayavalli (she/they) [Vanderbilt University]

The basics

What is machine learning?

Why are these tools useful in high energy colliders?

How to quantify performance?

Physics with ML

Classifier - Identifying the jet flavor

Regressor - Correcting for detector effects, unfolding

Generator - Learn underlying phenomena

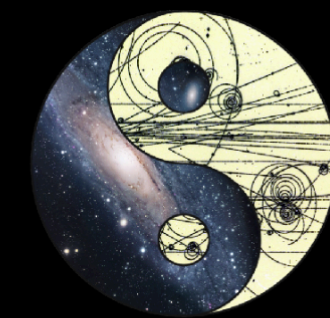
What is Machine Learning

Machine learning is the subfield of computer science that, according to Arthur Samuel, gives "computers the ability to learn without being explicitly programmed."

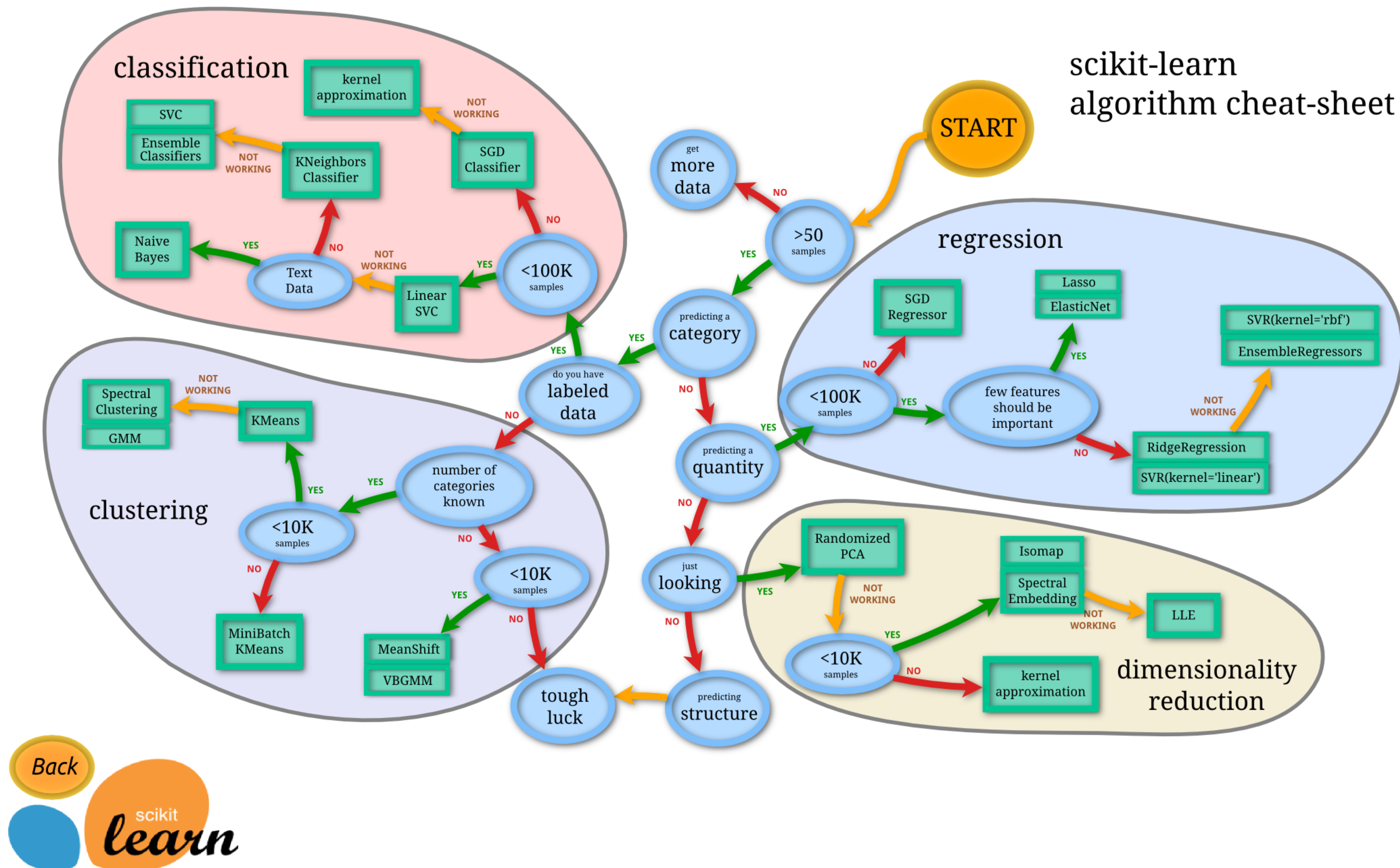
Machine learning - Wikipedia
https://en.wikipedia.org/wiki/Machine_learning

we will want AI to help us
debug our thinking by using all data from all
experiments optimally and “open our eyes” just as
AlphaGo opened the eyes of the professional go
players and enhance our intuition and creativity and
ability to break paradigms and boxes

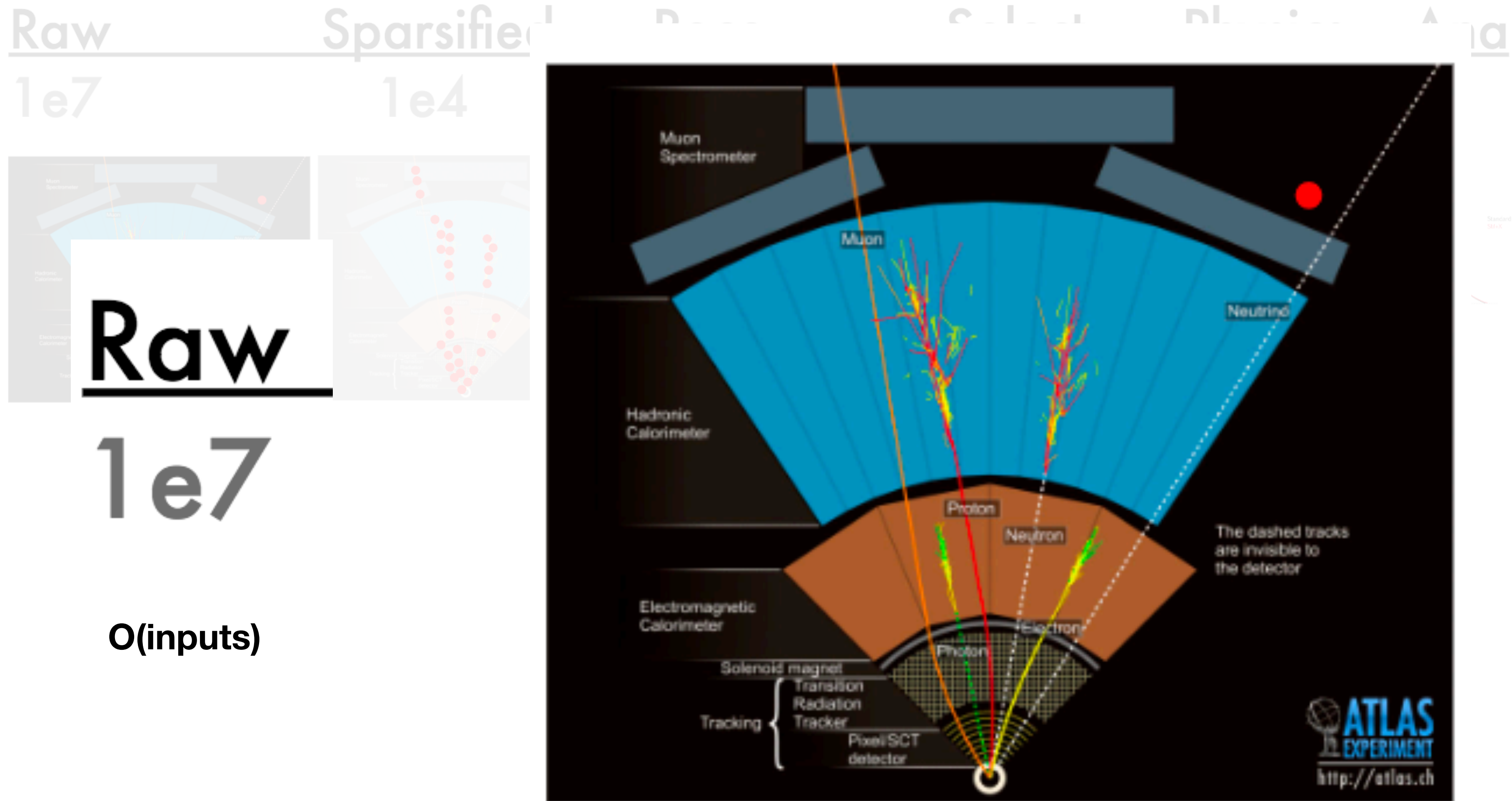
Maria Spiropulu (Caltech)



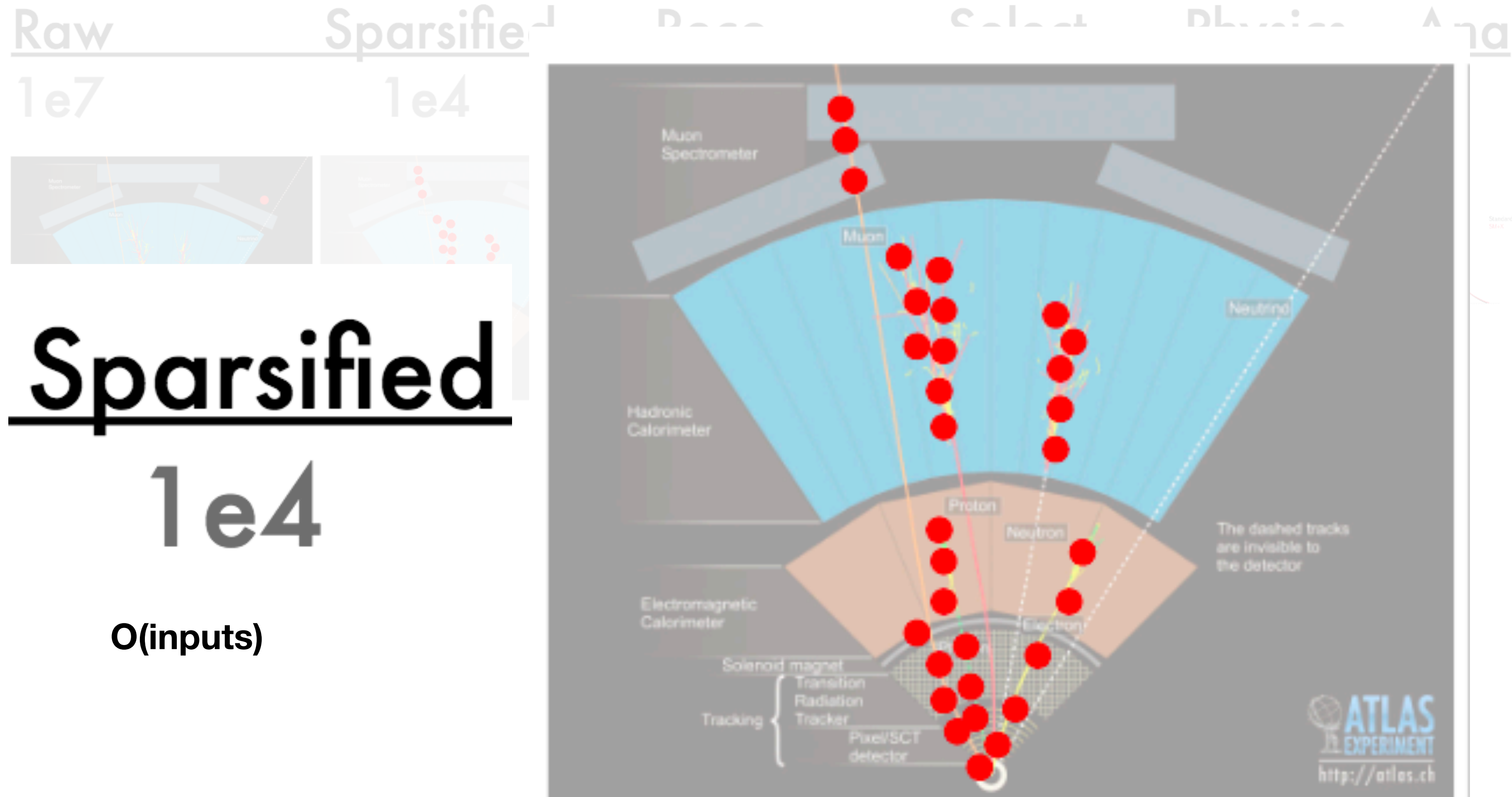
Kind of problems in ML



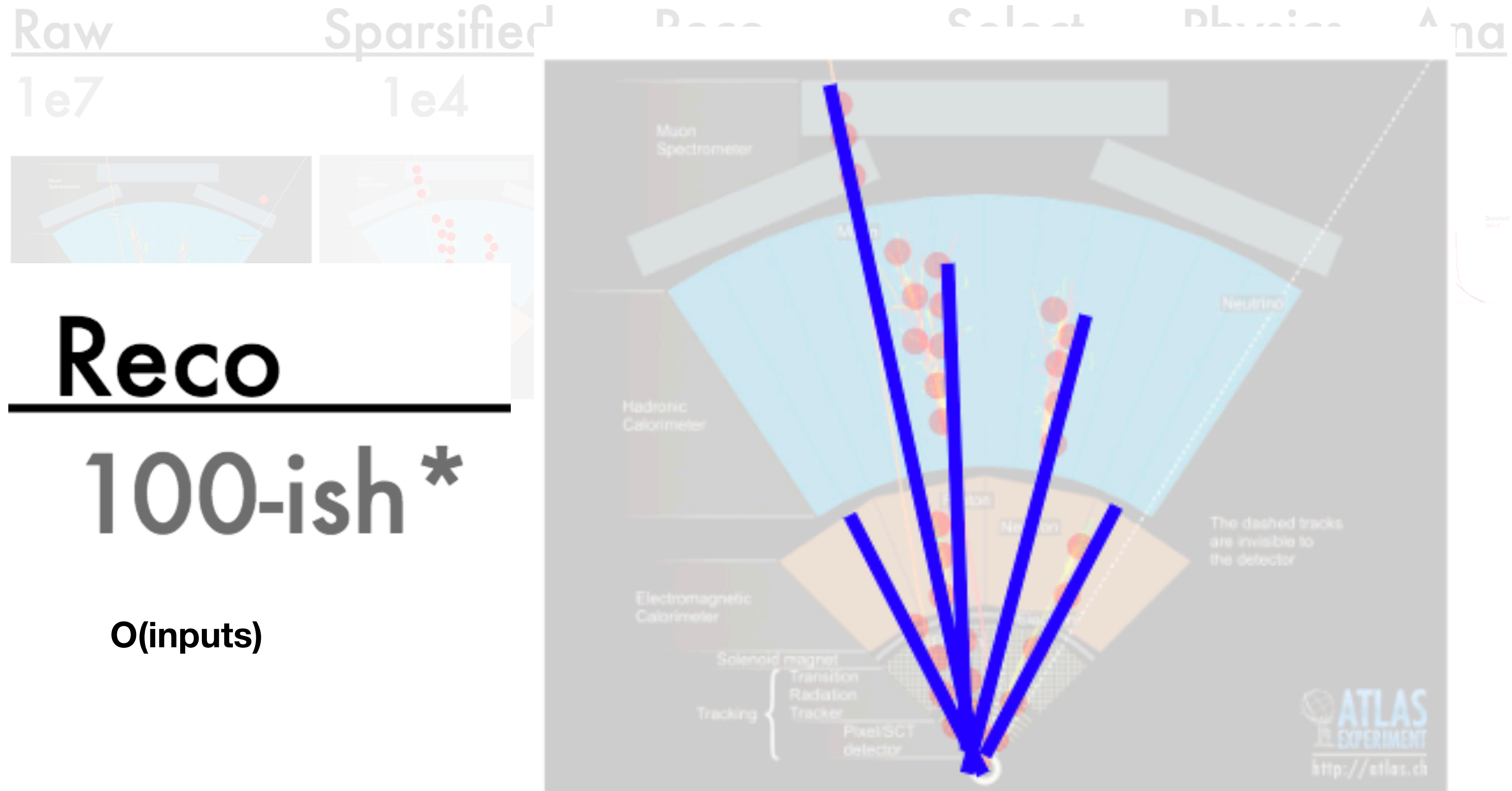
Snapshot of Experimental Workflow - Beam to Physics



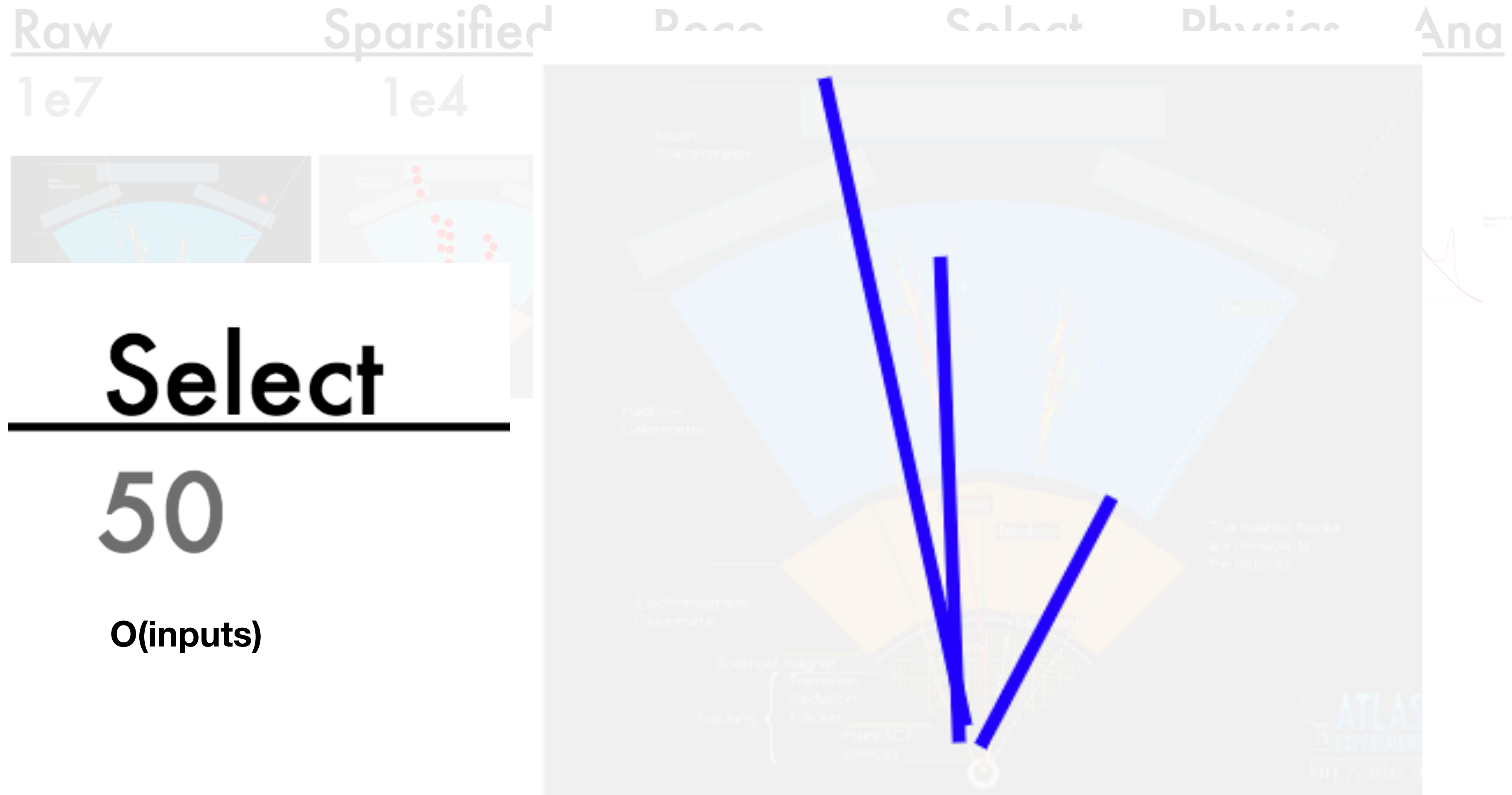
Snapshot of Experimental Workflow - Beam to Physics



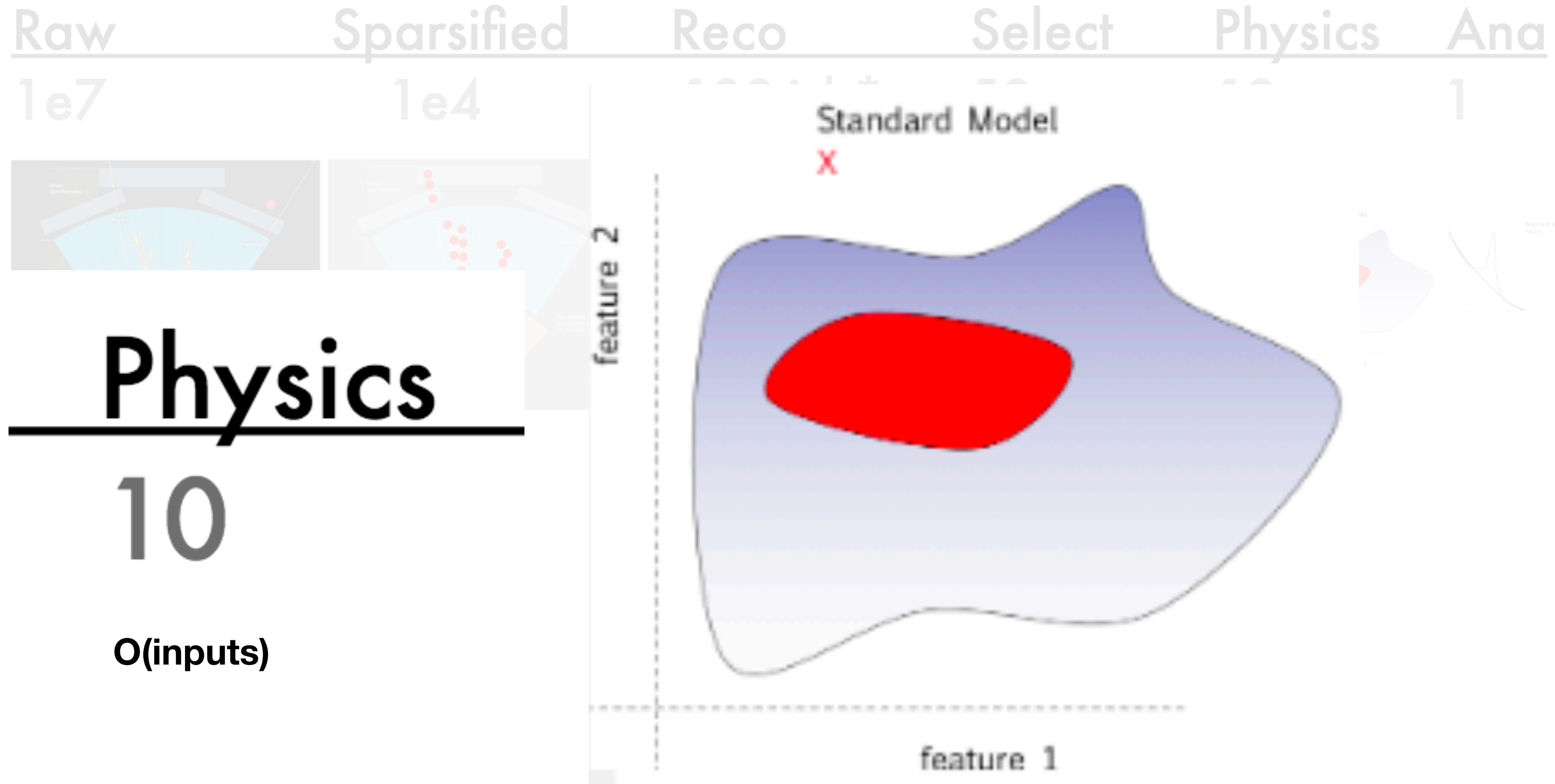
Snapshot of Experimental Workflow - Beam to Physics



Snapshot of Experimental Workflow - Beam to Physics

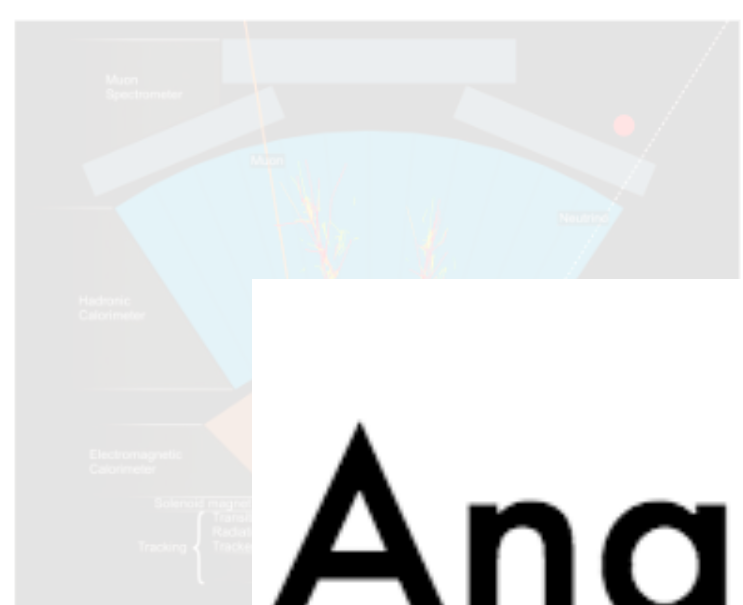


Snapshot of Experimental Workflow - Beam to Physics

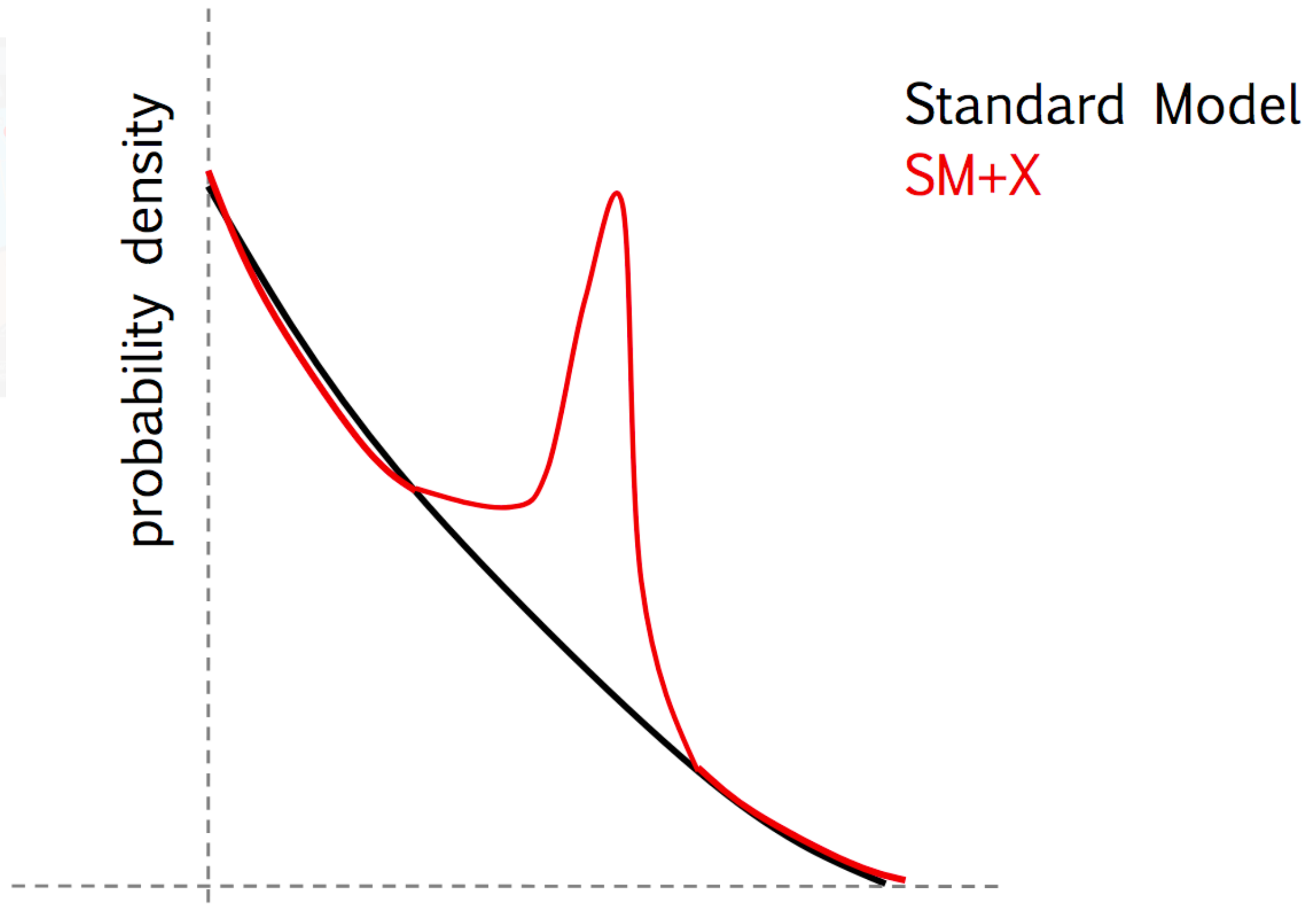


Snapshot of Experimental Workflow - Beam to Physics

Raw	Sparsified	Reco	Select	Physics	Ana
1e7	1e4	100 : 1e6 *	50	10	1

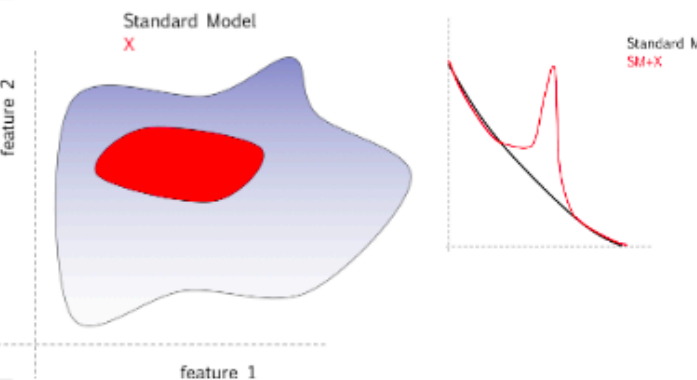
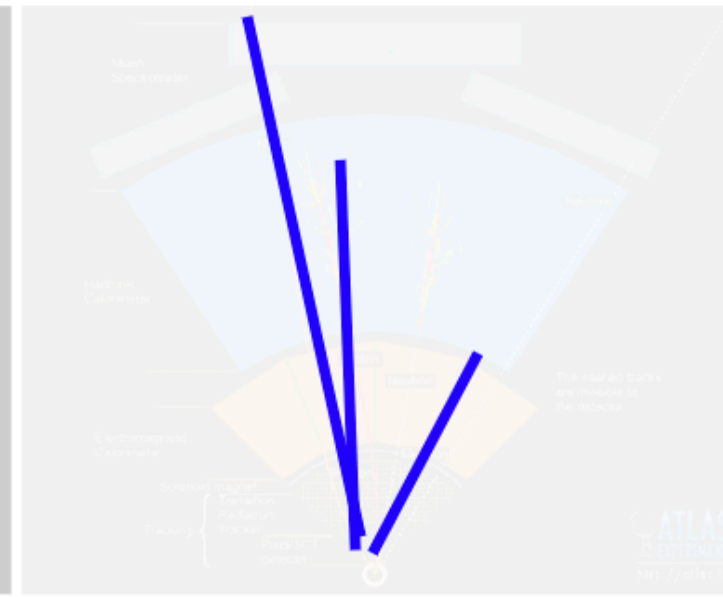
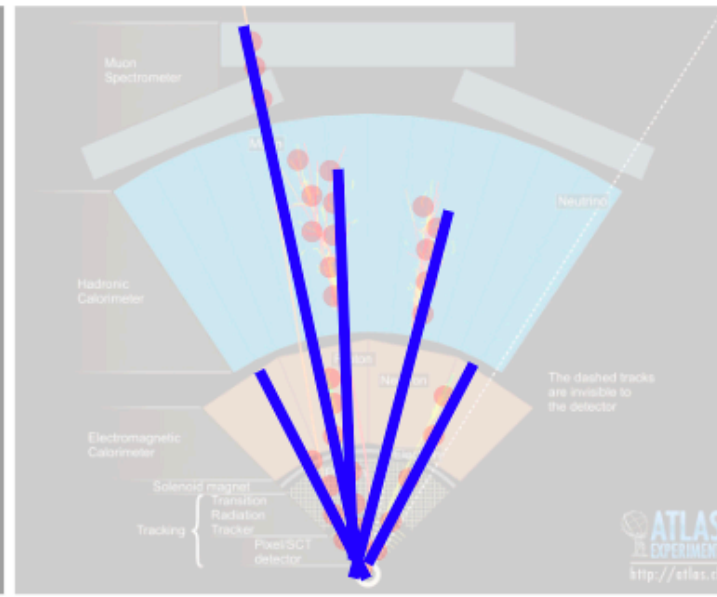
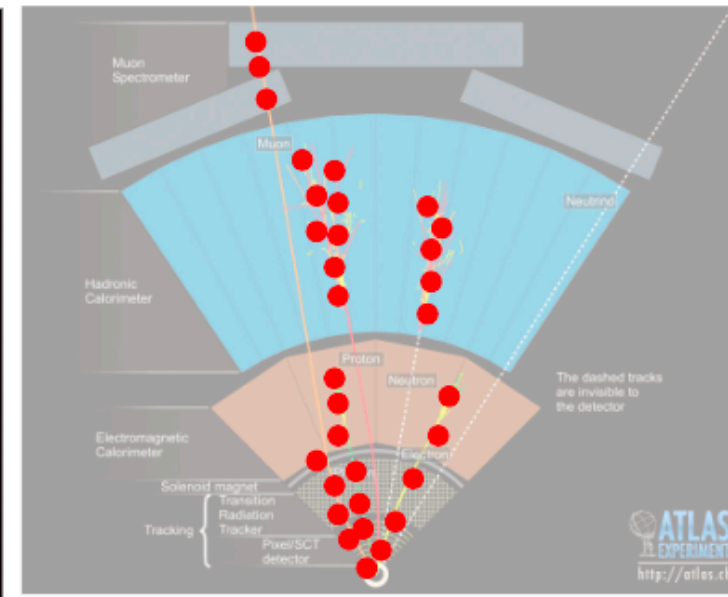
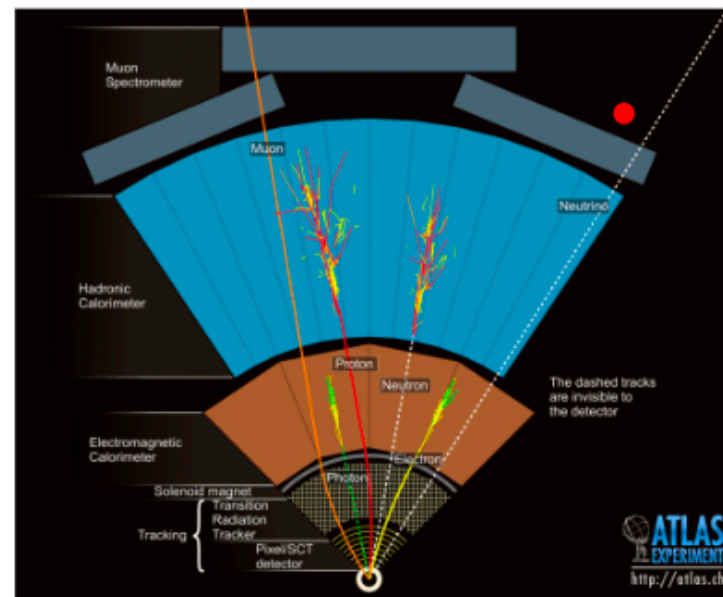


Ana
1
O(inputs)



Snapshot of Experimental Workflow - Beam to Physics

Raw	Sparsified	Reco	Select	Physics	Ana
1e7	1e4	100-ish*	50	10	1



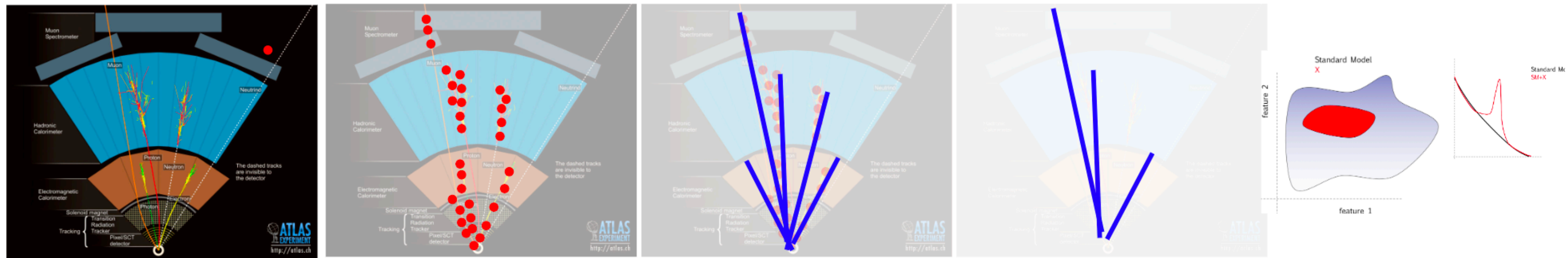
Particle Physics	Zero suppression	Reconstruction	Quality Selection	Observables
Machine Learning	Down-Sampling	Dimensionality Reduction	Low Level Features	High Level Features



- Who does this work?

Snapshot of Experimental Workflow - Beam to Physics

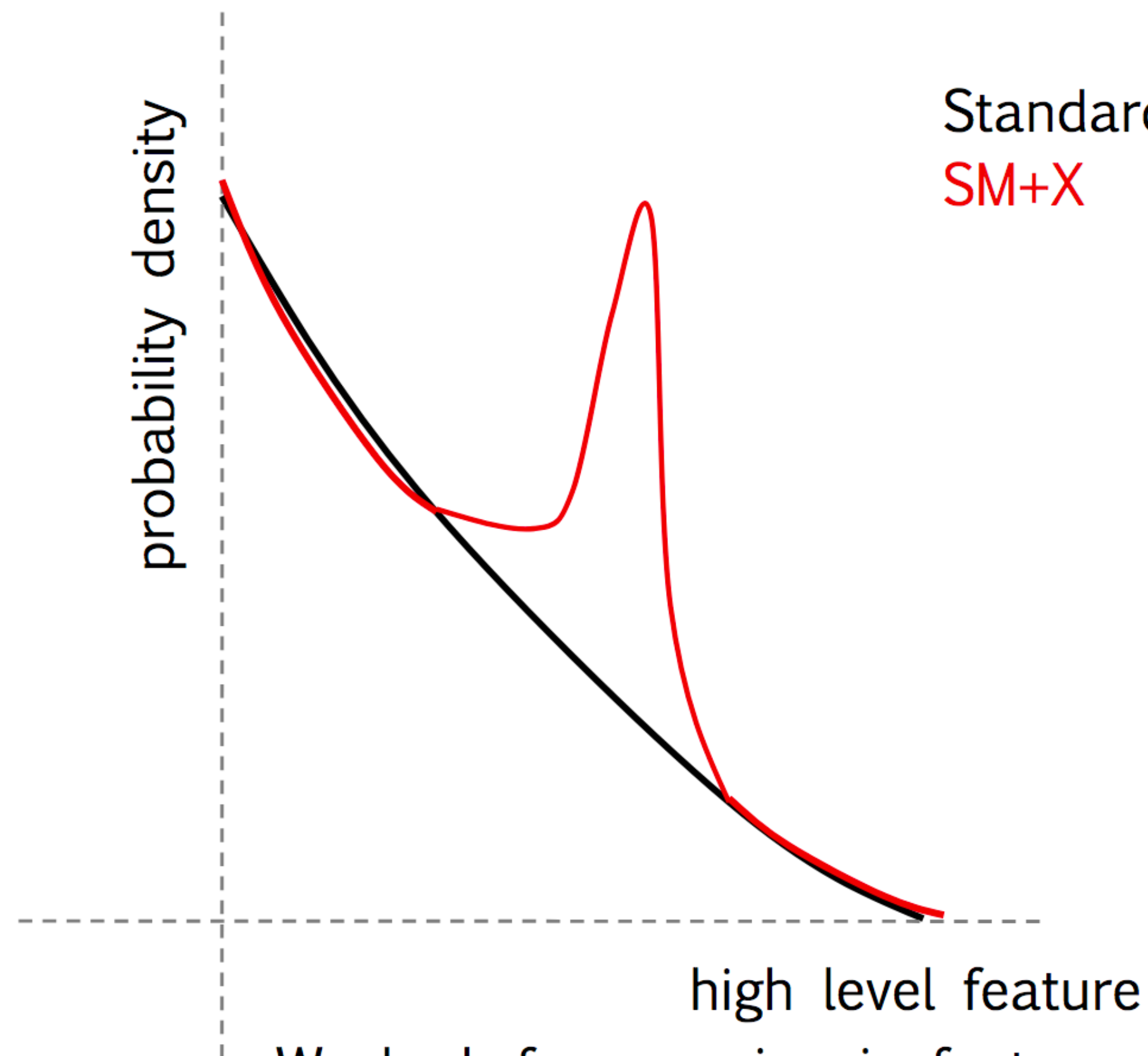
Raw	Sparsified	Reco	Select	Physics	Ana
1e7	1e4	100-ish*	50	10	1



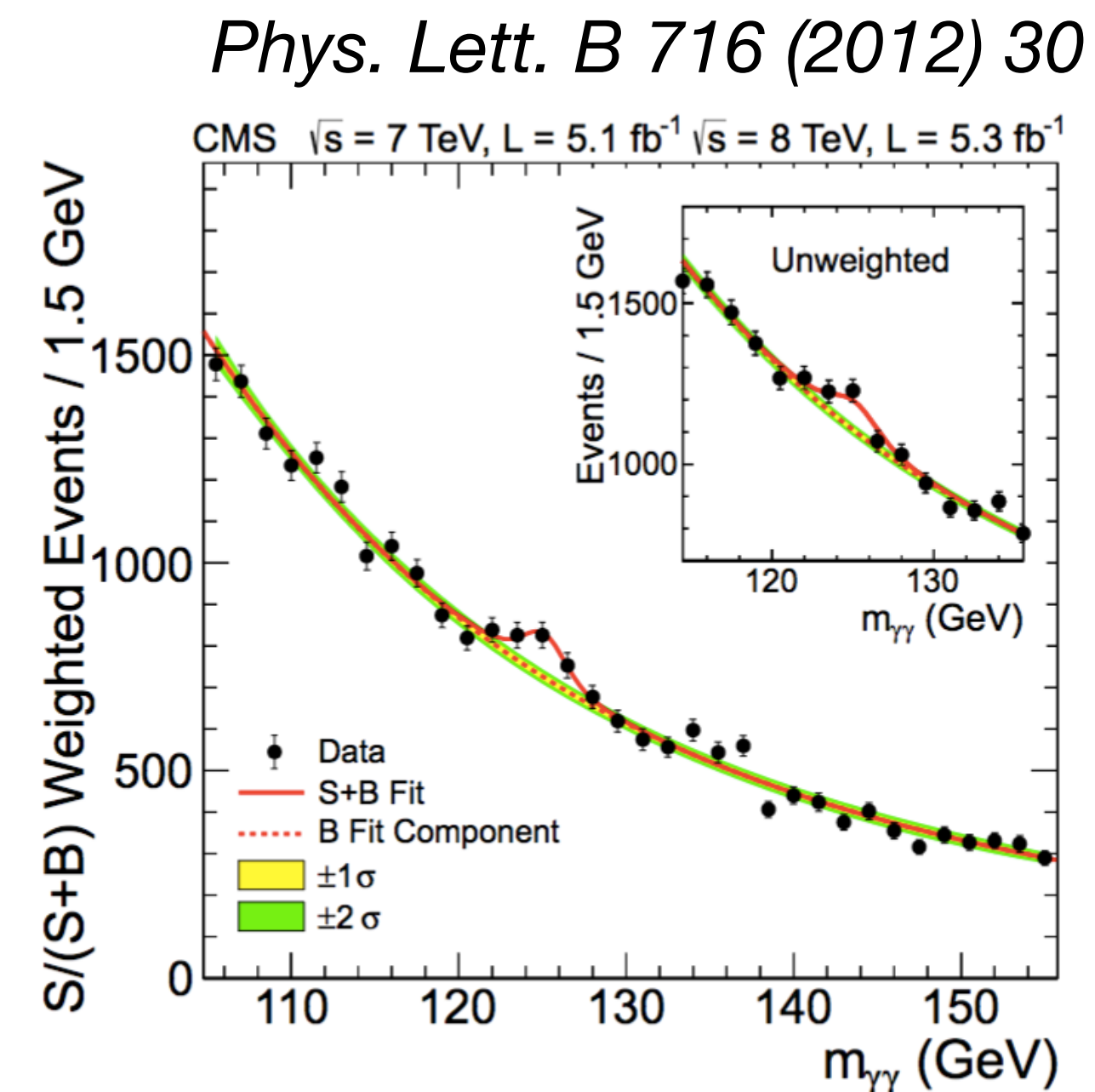
Particle Physics	Zero suppression	Reconstruction	Quality Selection	Observables
Machine Learning	Down-Sampling	Dimensionality Reduction	Low Level Features	High Level Features

- Active human input in going from 'Raw' to 'High Level' feature(s)
- Algorithms run (extremely quickly) to select events with useful features for further human analysis

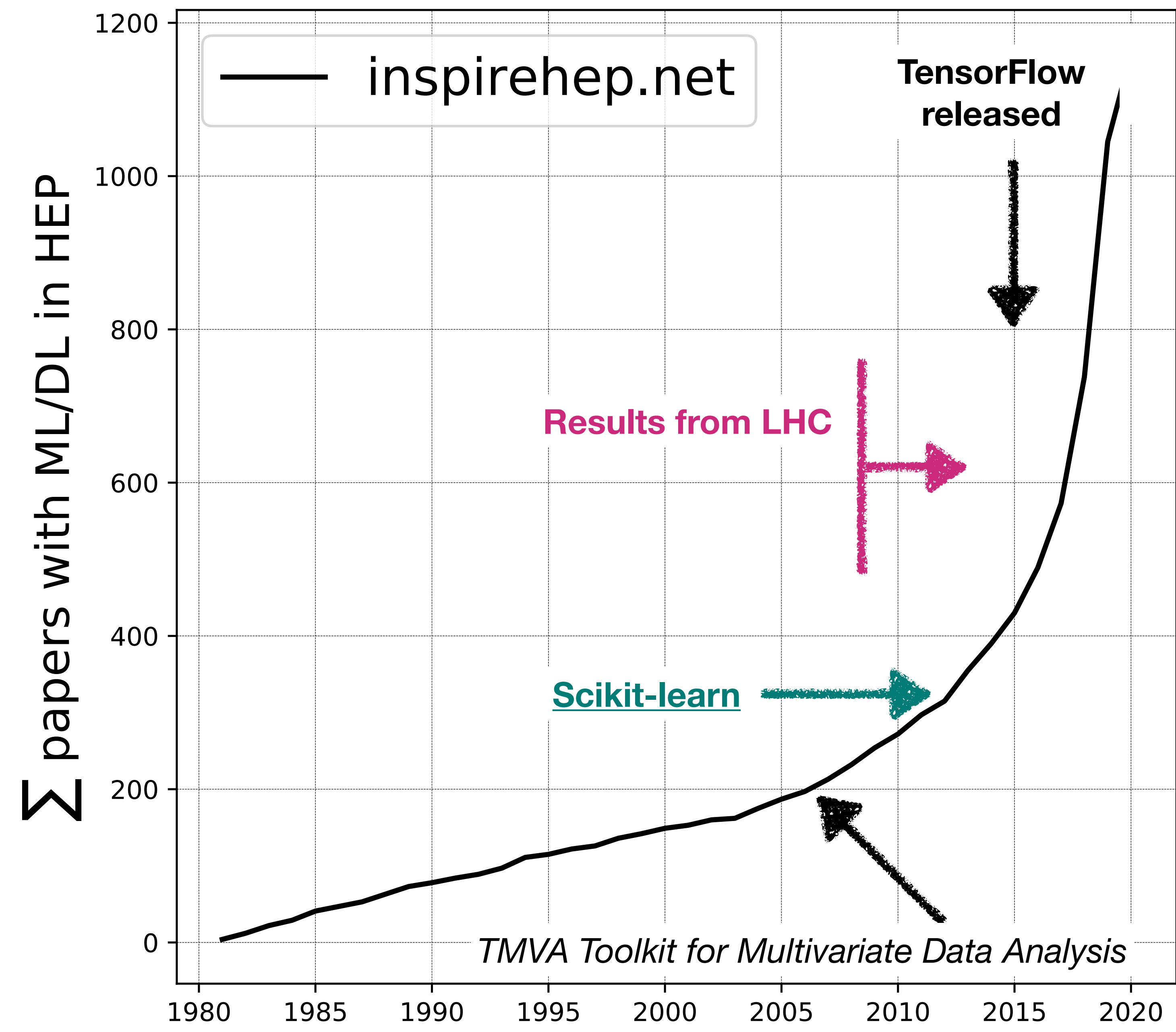
Successful analysis!

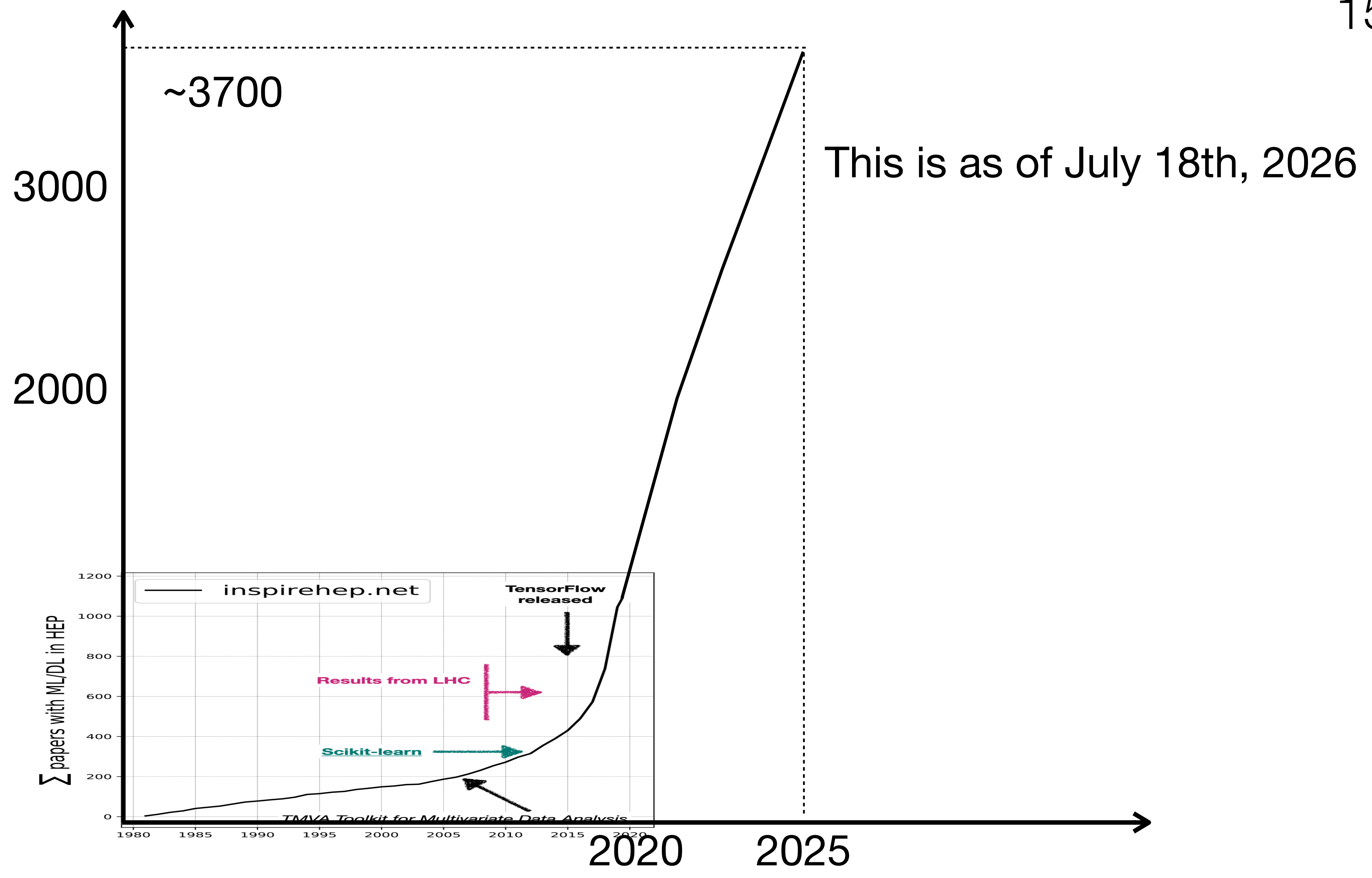


We look for a region in feature space where the two hypotheses have large differences in their predictions.

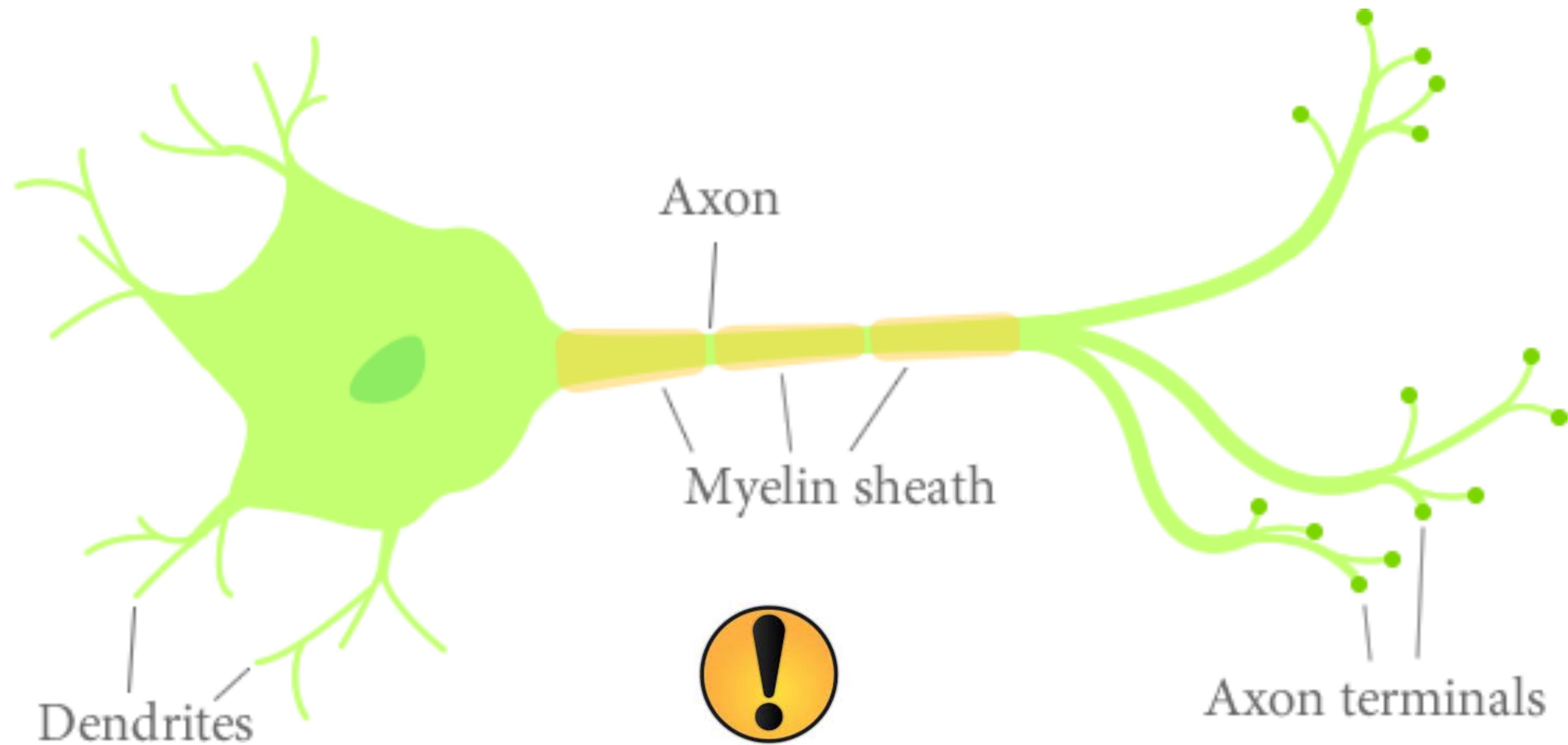


The Nobel Prize in Physics 2013
François Englert, Peter Higgs





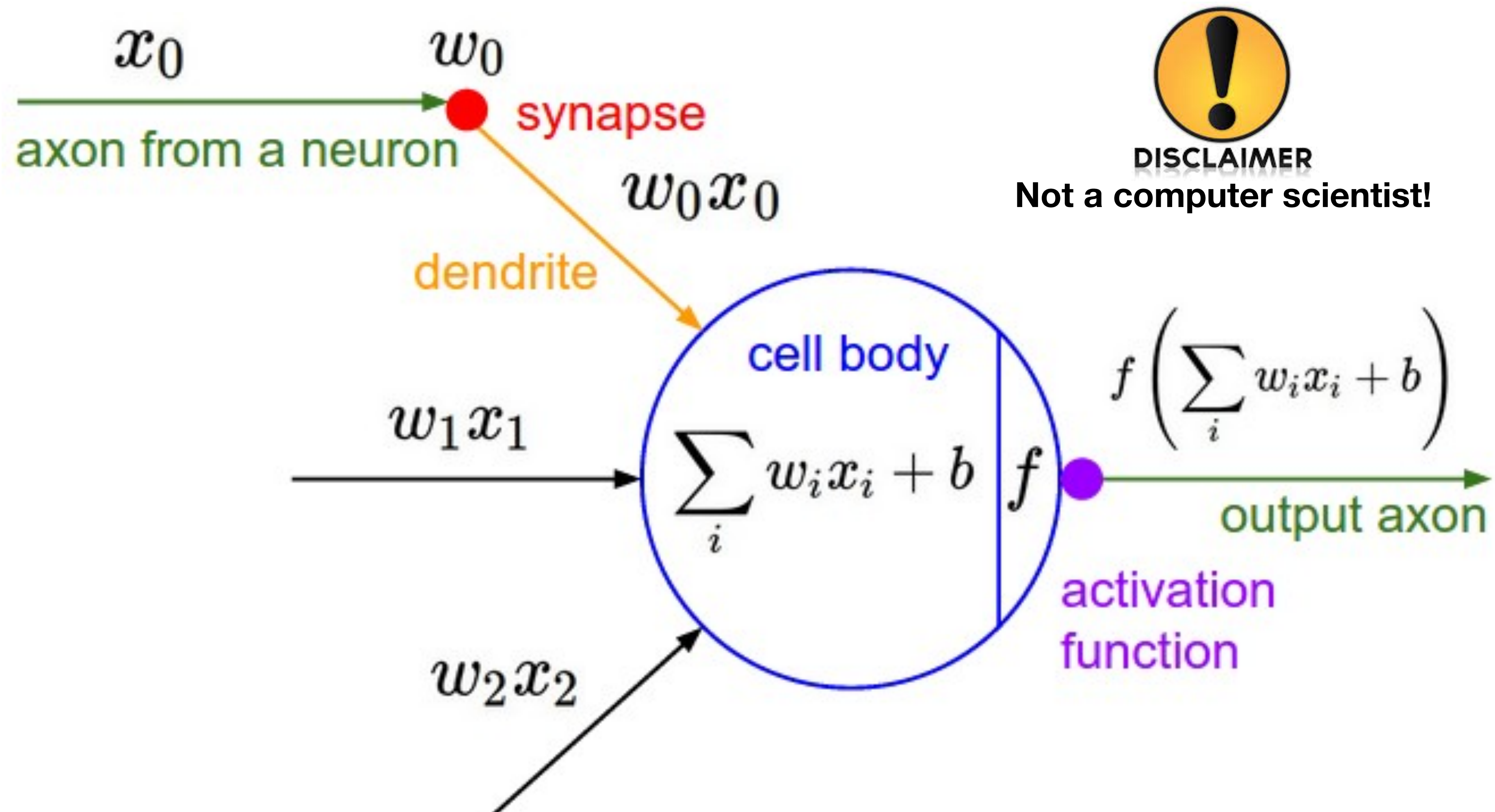
Real Neuron



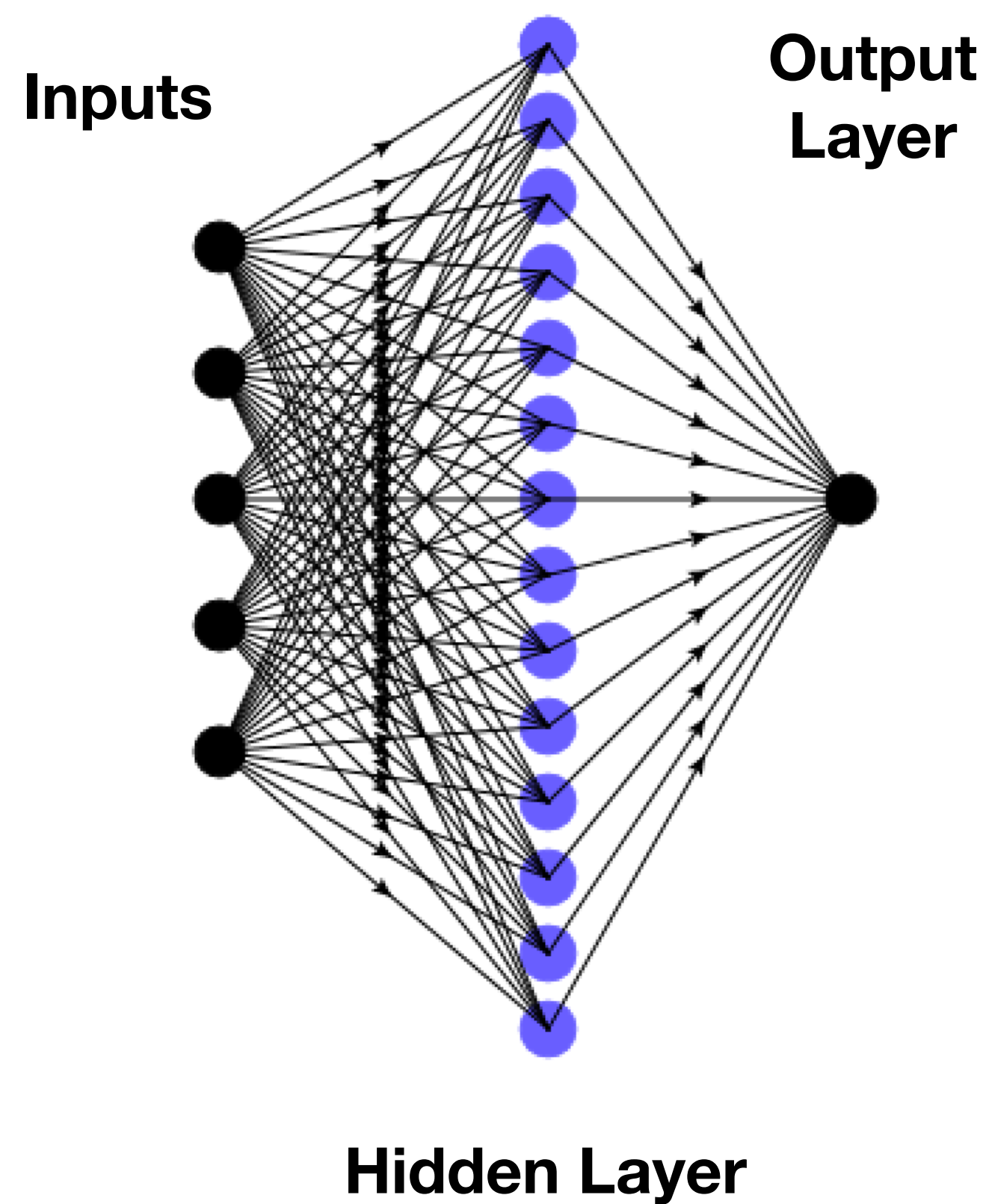
DISCLAIMER
Not a biologist!

Figure credit: Google Images

Artificial Neuron



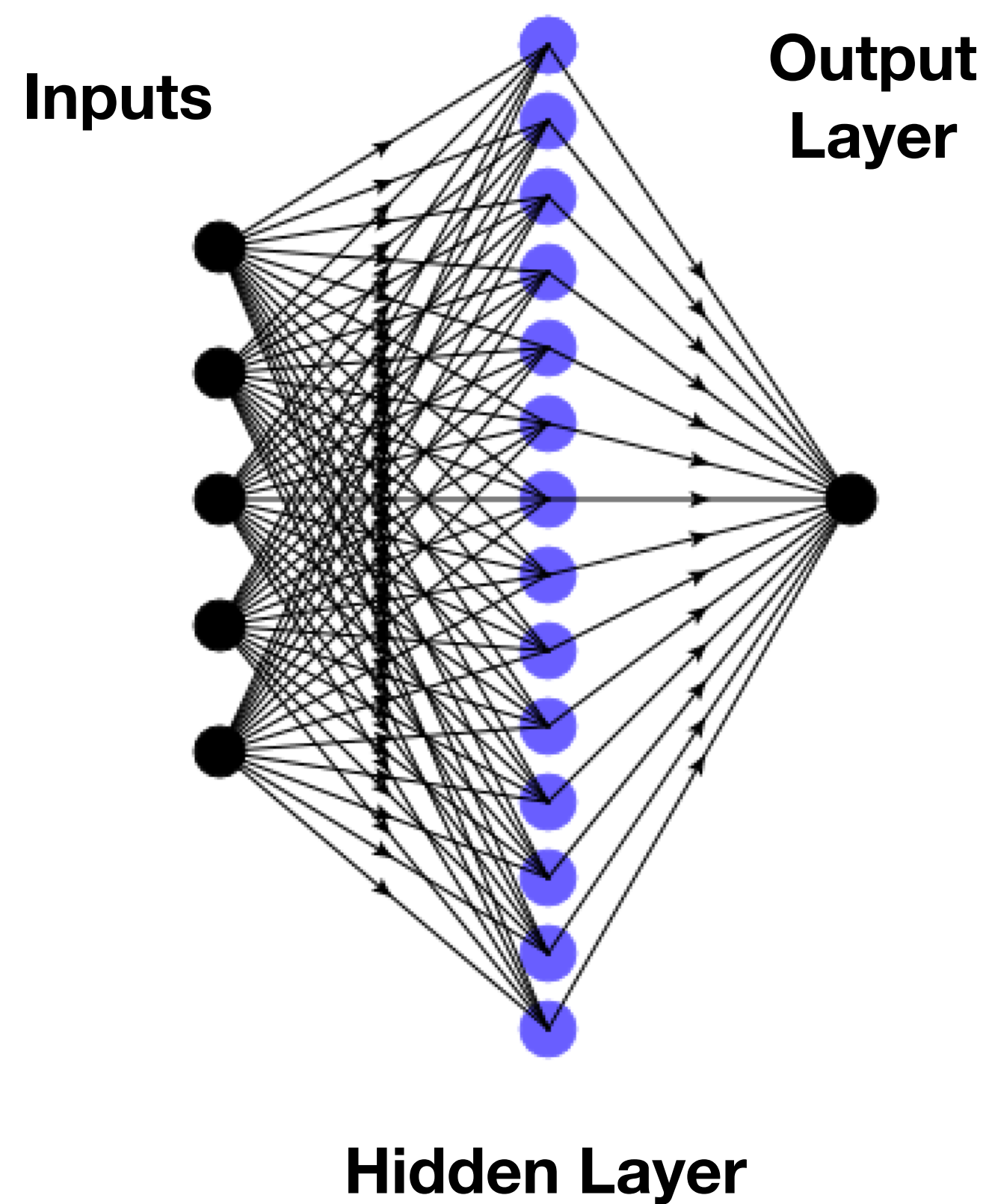
MultiLayer Perceptron (MLP)



- Hidden node $\bullet = h_i + \text{AF}(\text{Sum } w_i \bullet)$
- AF : Activation Functions (non-Linear)
- Output Layer -> For binary classification problems [0,1]

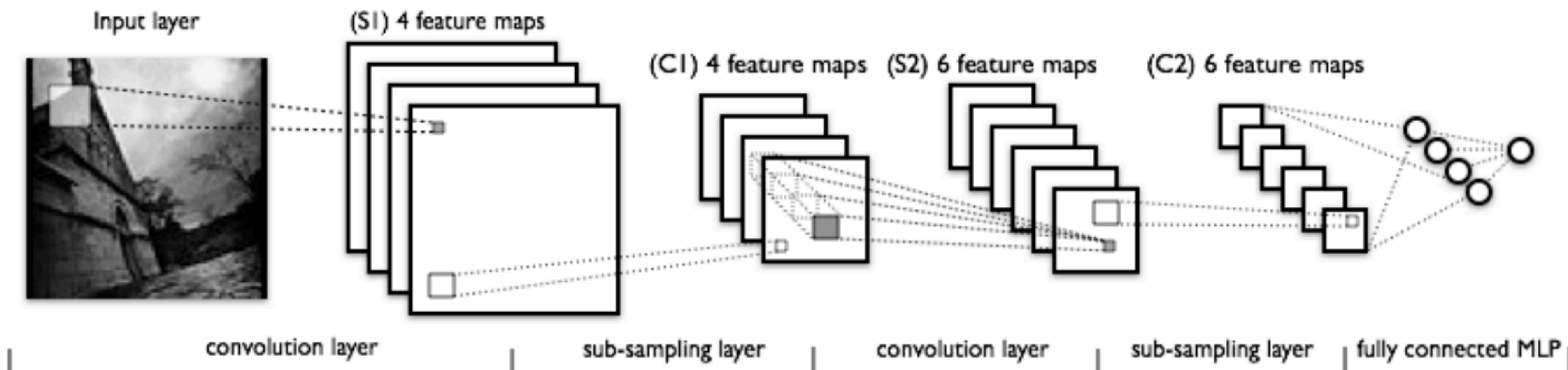


MultiLayer Perceptron (MLP)



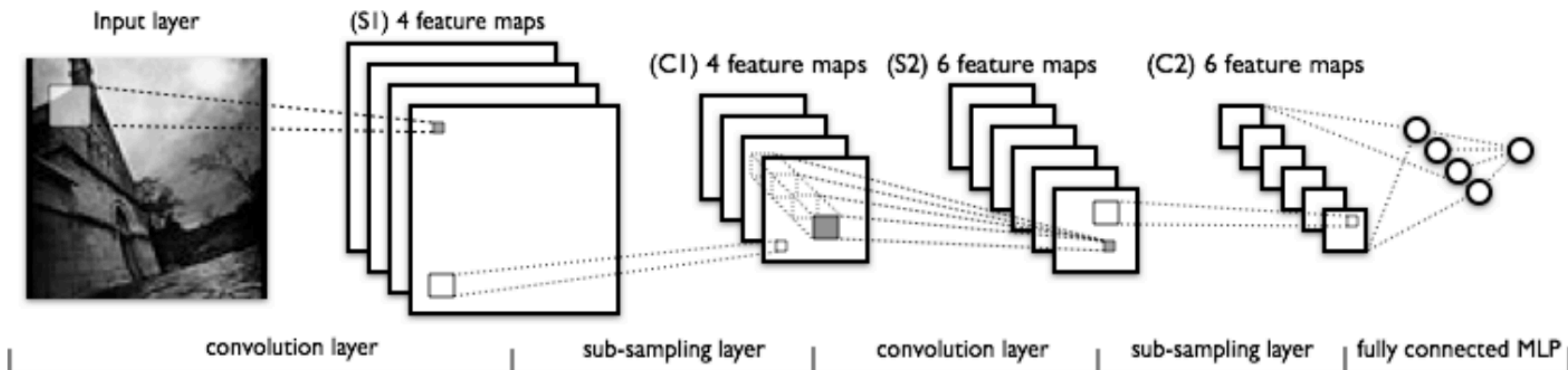
- Hidden node $\bullet = h_i + \text{AF}(\text{Sum } w_i \bullet)$
- AF : Activation Functions (non-Linear)
- Output Layer -> For binary classification problems [0,1]
- For non-linear behavior - increase number of hidden layers and vary activation functions
- This is a fully connected network

Convolutional Neural Networks



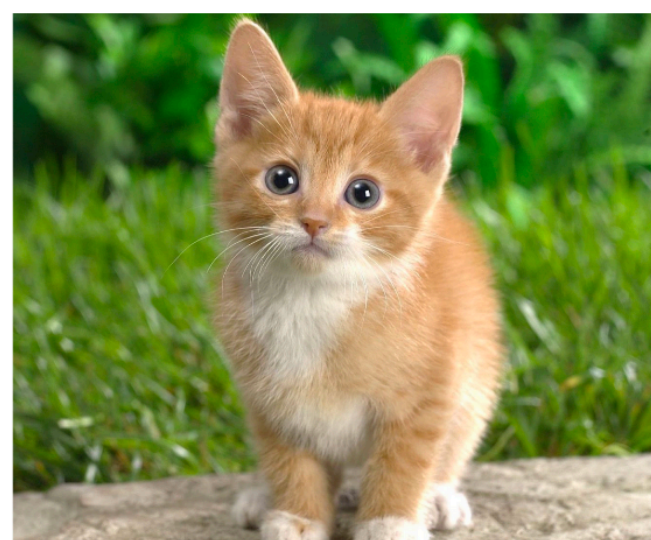
<https://skymind.ai/wiki/convolutional-network>

Convolutional Neural Networks

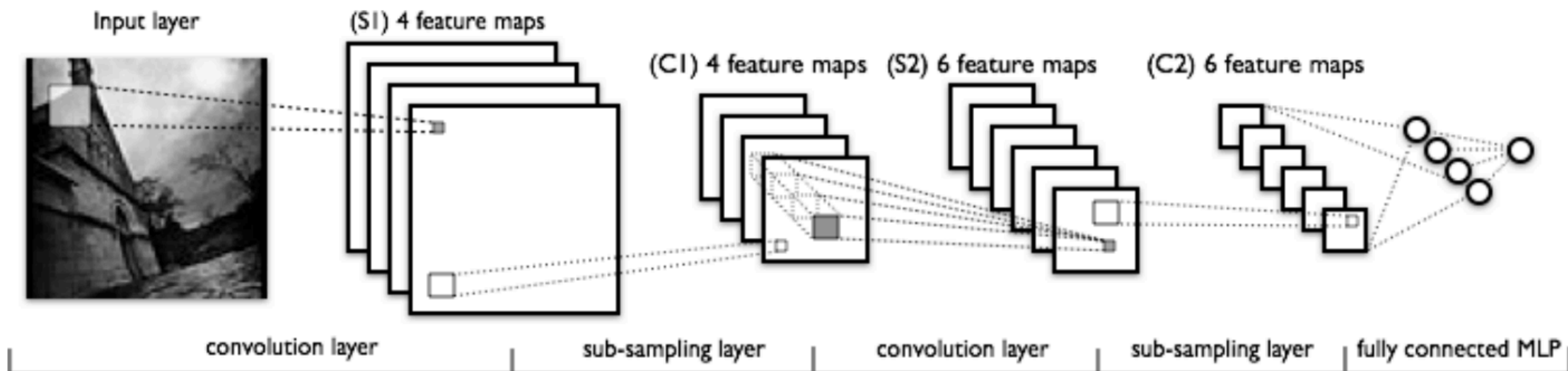


<https://skymind.ai/wiki/convolutional-network>

Classification

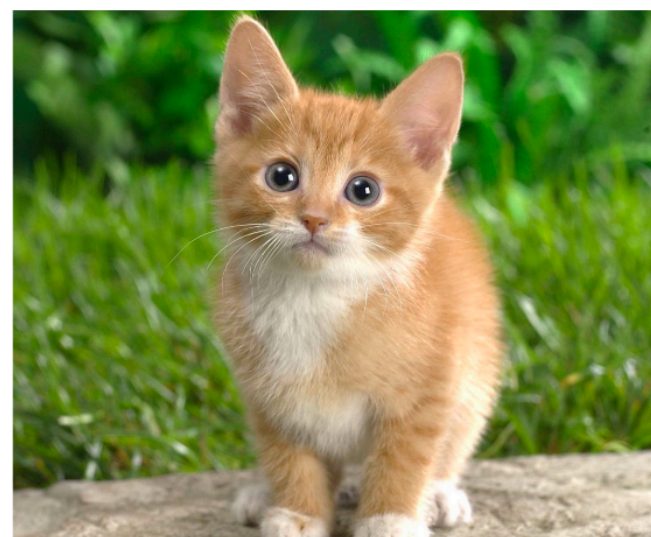


Convolutional Neural Networks

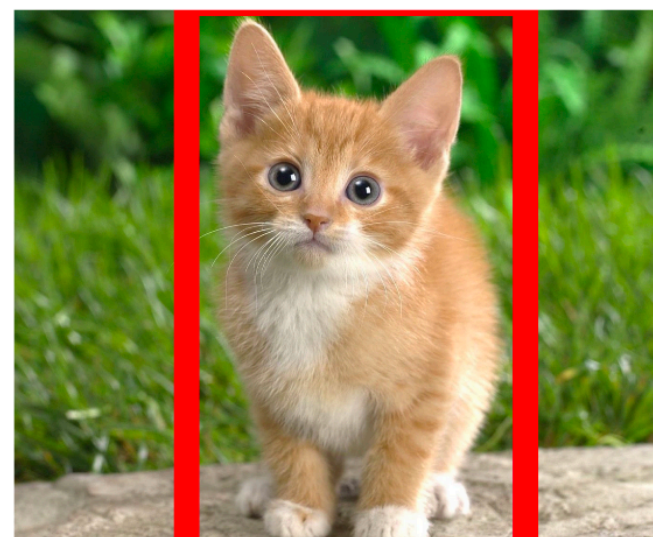


<https://skymind.ai/wiki/convolutional-network>

Classification

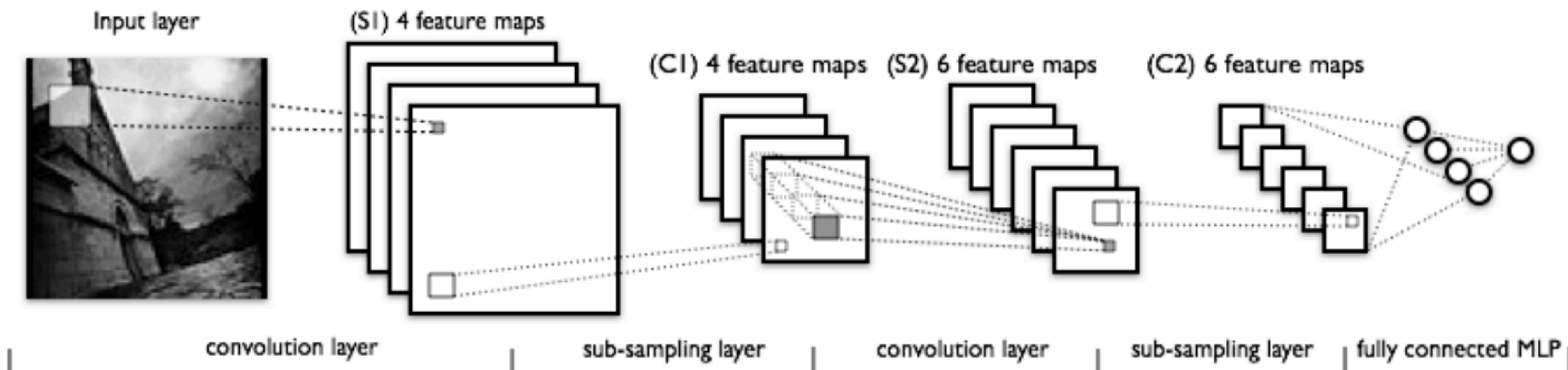


Classification +
Localization



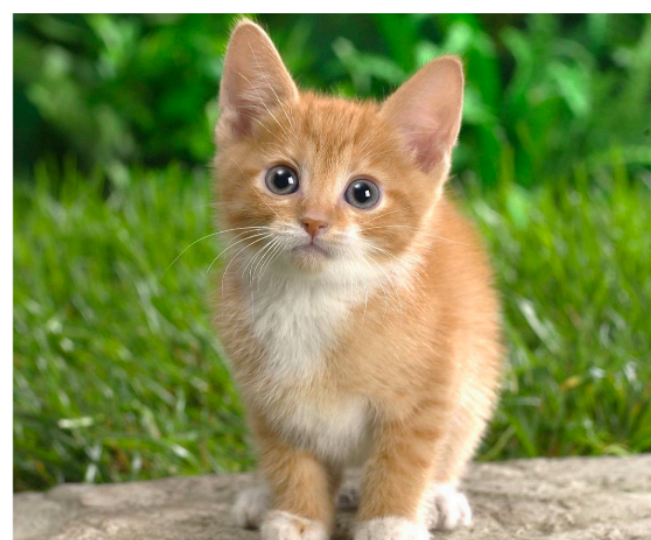
Cat

Convolutional Neural Networks

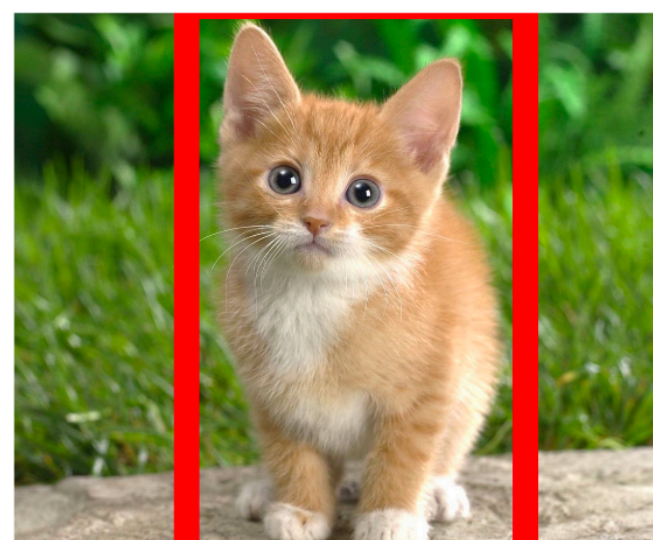


<https://skymind.ai/wiki/convolutional-network>

Classification

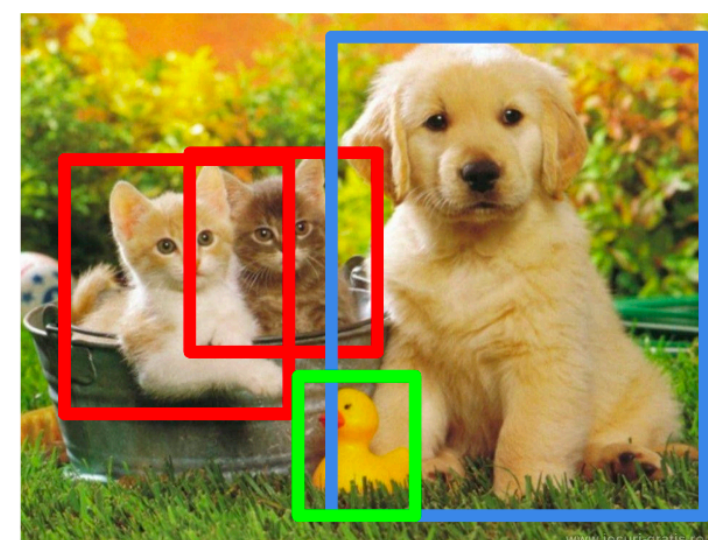


Classification +
Localization



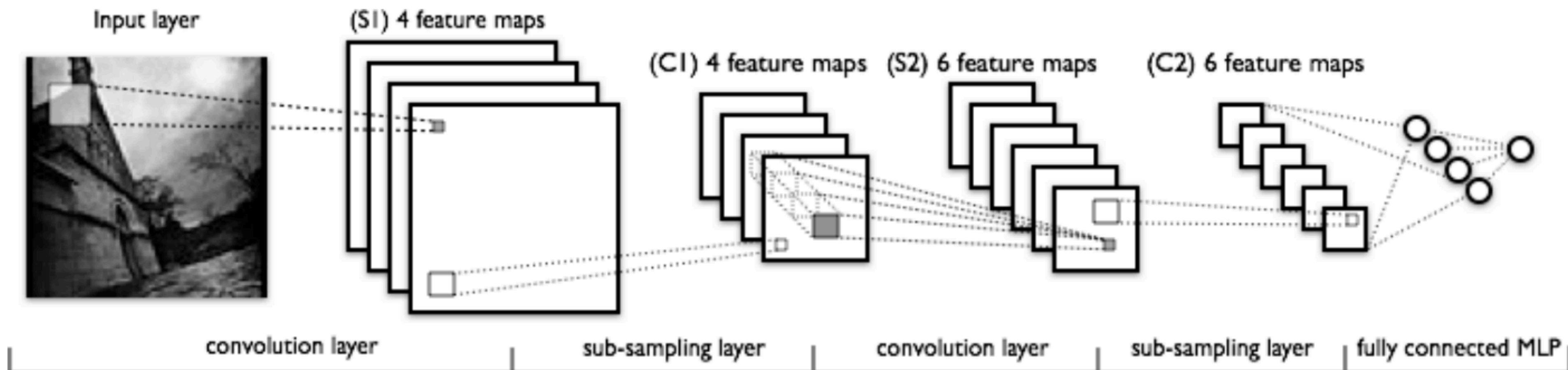
Cat

Object detection



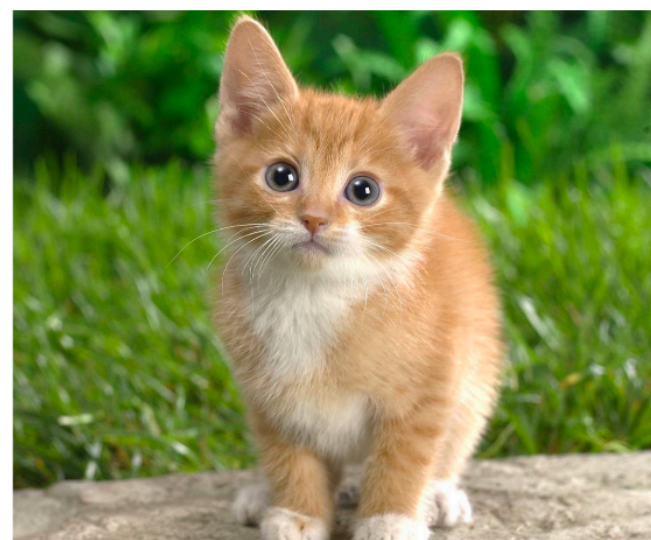
Cat, Dog, Duck

Convolutional Neural Networks

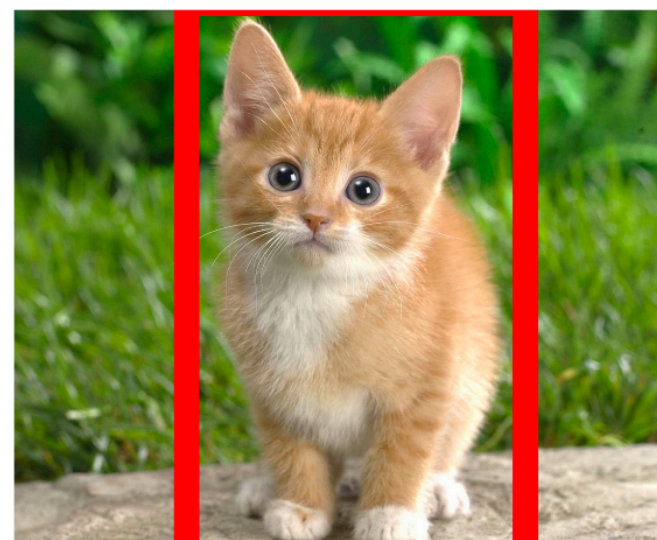


<https://skymind.ai/wiki/convolutional-network>

Classification

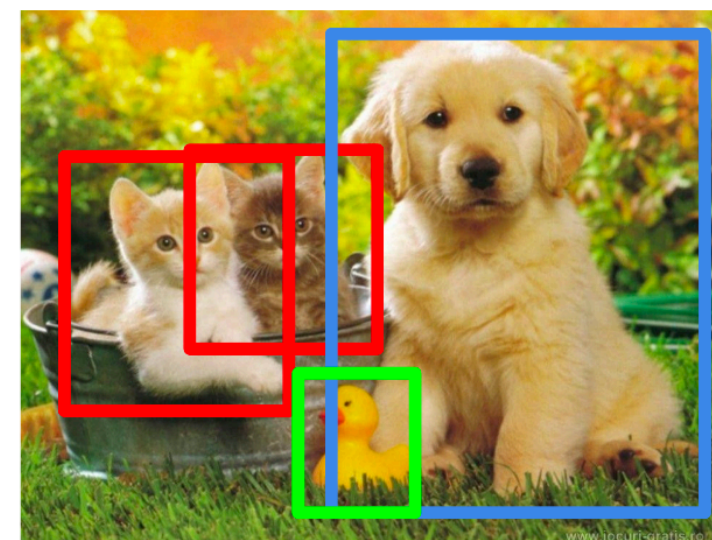


Classification +
Localization



Cat

Object detection



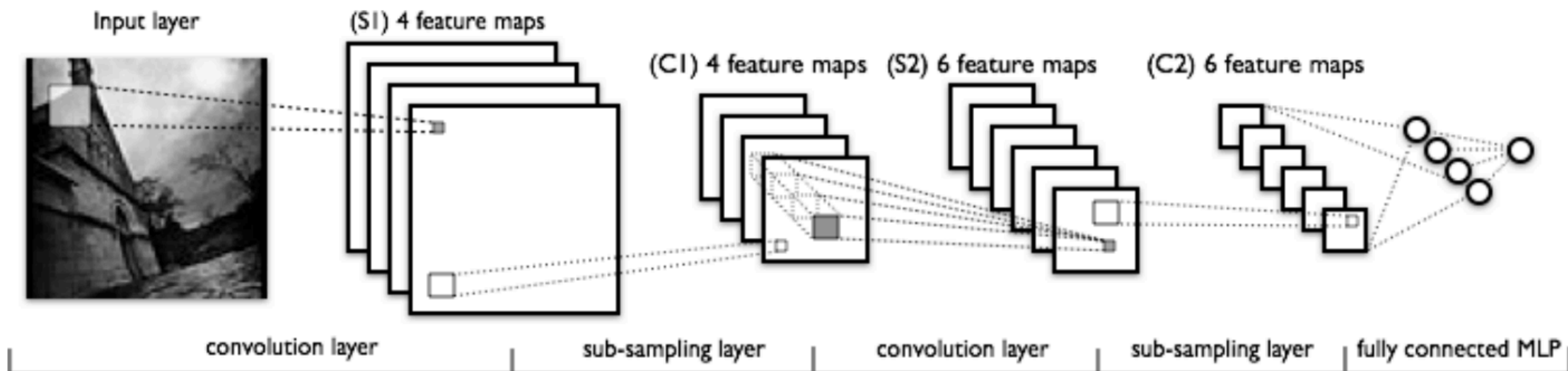
Cat, Dog, Duck

Instance
segmentation



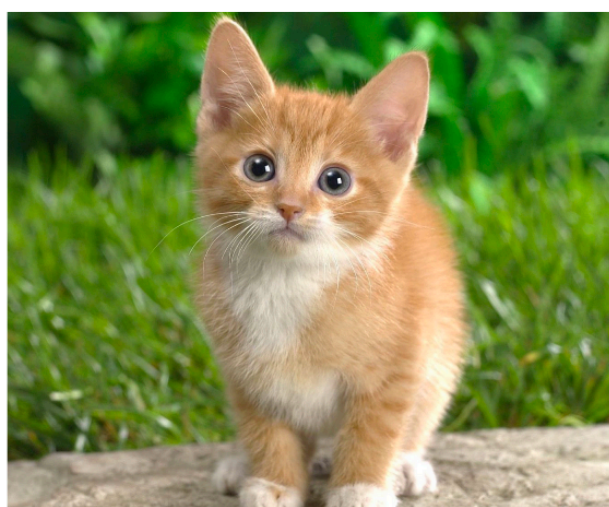
Cat, Dog, Duck

Convolutional Neural Networks

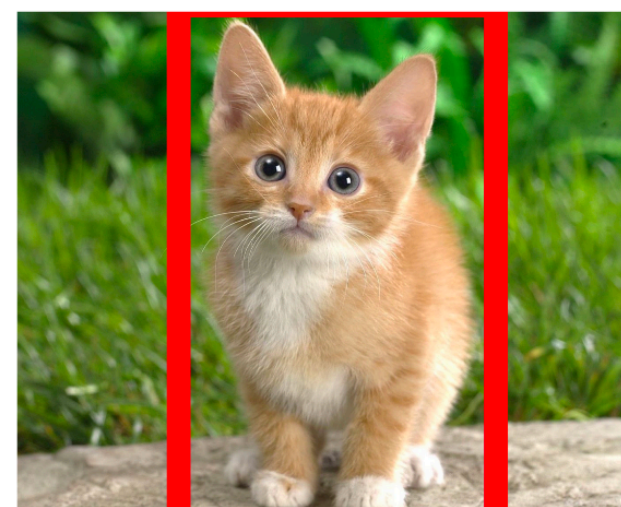


<https://skymind.ai/wiki/convolutional-network>

Classification

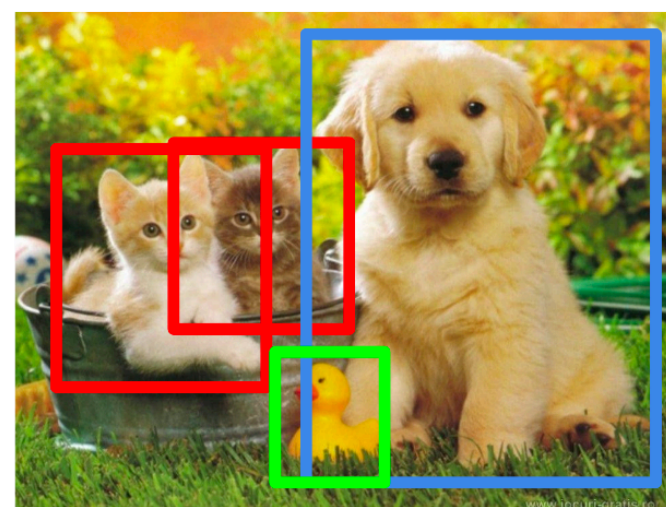


Classification +
Localization



Cat

Object detection



Cat, Dog, Duck

Instance
segmentation



Cat, Dog, Duck

Image read-out



“There is a dog
and two cats
sitting down”

Tools of the trade



<https://www.tensorflow.org/>



Keras

<https://keras.io/>



PyTorch

<http://pytorch.org/>

- TensorFlow - Google developed backend for multi-dimensional vectors and their manipulations
- Keras - External Package good for beginners
- PyTorch - more state of the art with potential for multiple variations

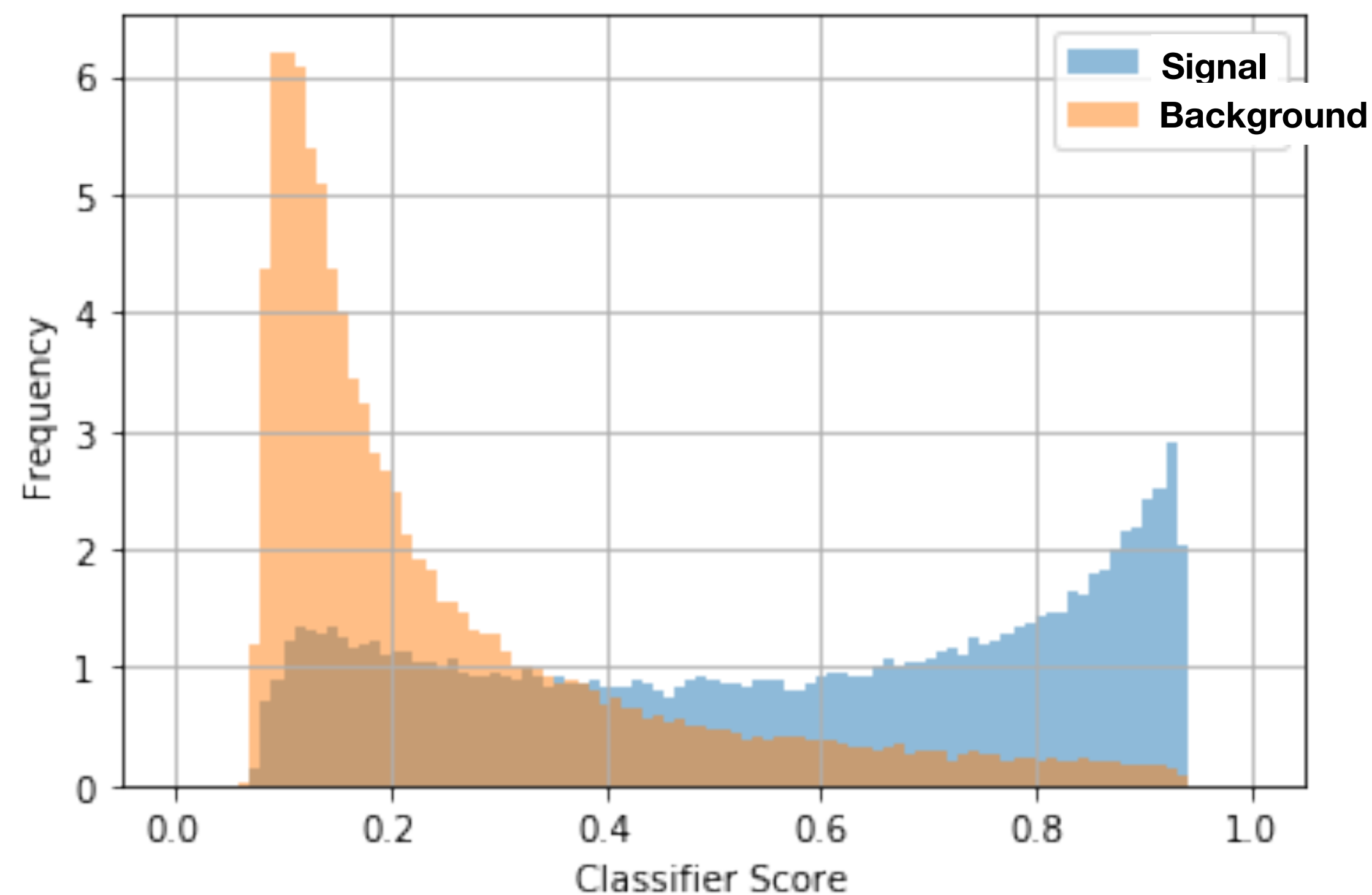
How to quantify or grade a model at classification

Name in Physics	Name in ML/CV	Definition
Efficiency	True Positive Rate/Recall	$TPR = \frac{TP}{P} = \frac{TP}{TP + FN}$
Misid. Probability	False Positive Rate	$FPR = \frac{FP}{N} = \frac{FP}{FP + TN}$
Rejection	-----	$Rej = \frac{1}{FPR}$
Purity	Precision	$PREC = \frac{TP}{TP + FP}$

- T: True, F: False, P: Positive, N: Negative

ROC Curves

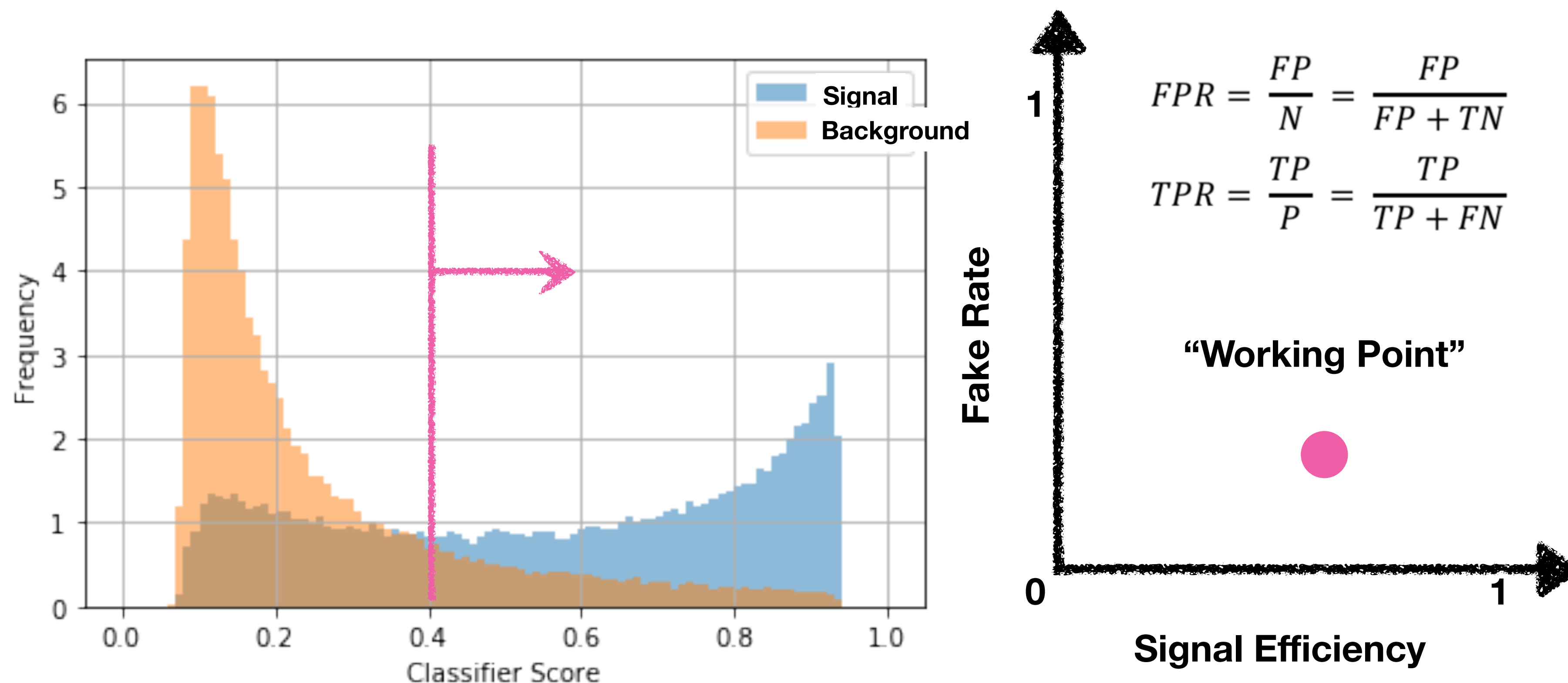
Receiver Operator Characteristic



- Classifier score - Values from the output layer

ROC Curves

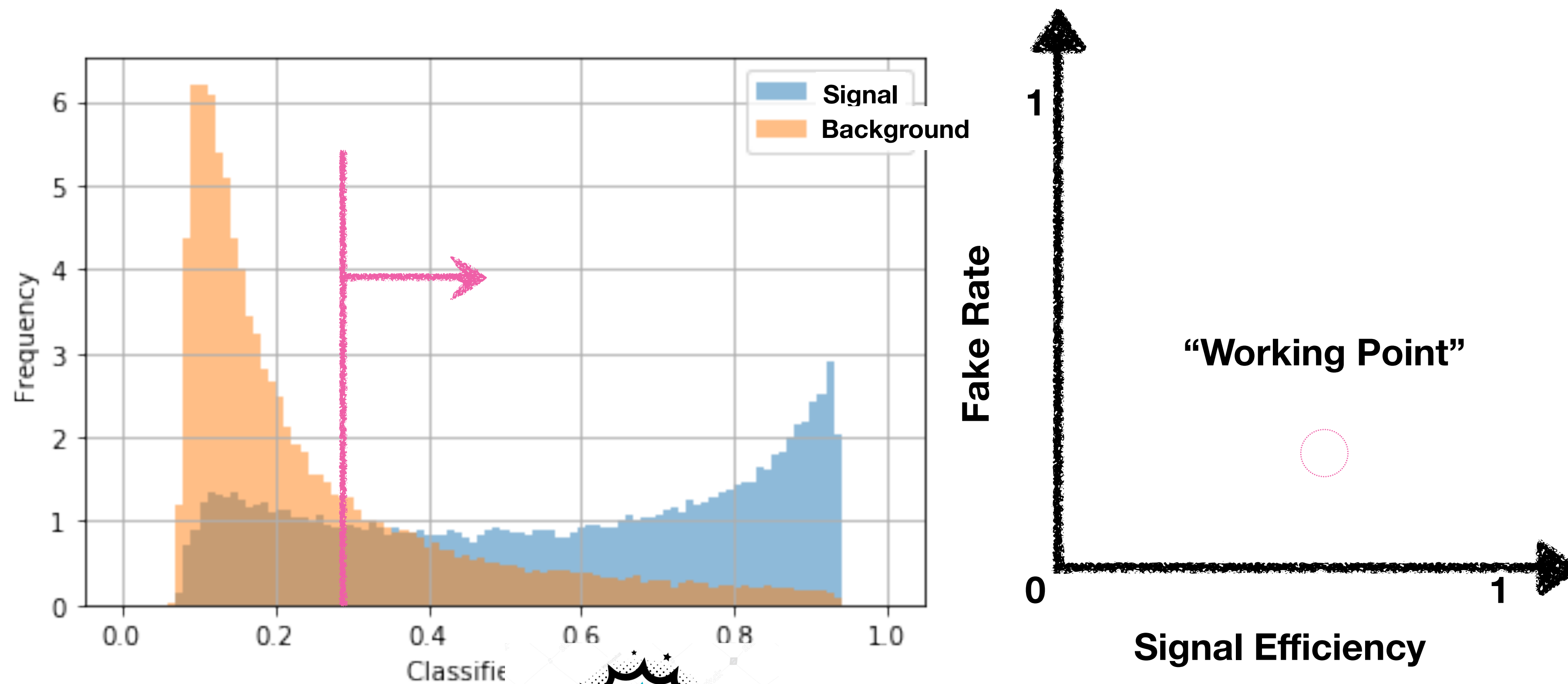
Receiver Operator Characteristic



- Choose a particular value and estimate Efficiency/Fake-Rate
- Efficiency = Selected Blue / Total Blue
- Fake Rate = Selected Yellow / Total Selected

ROC Curves

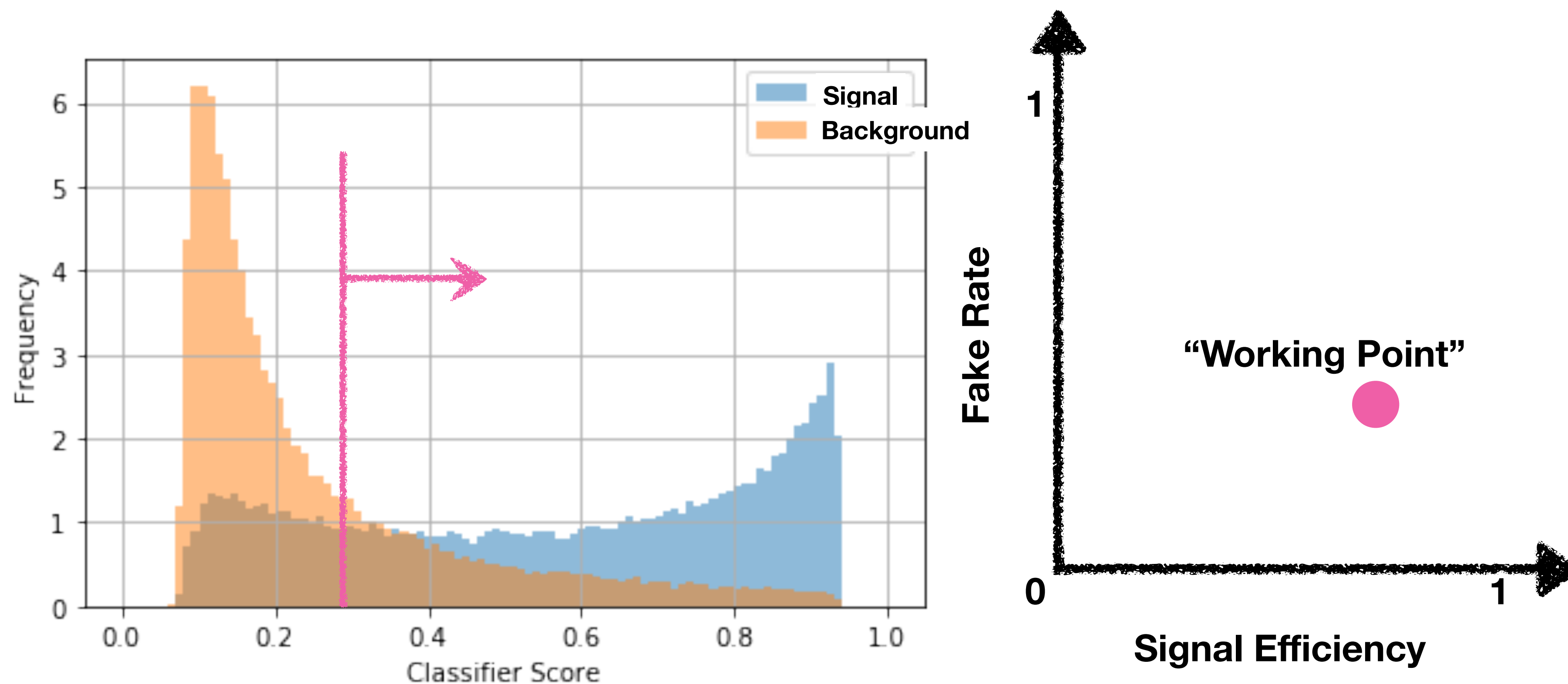
Receiver Operator Characteristic



Where is the new working point?

ROC Curves

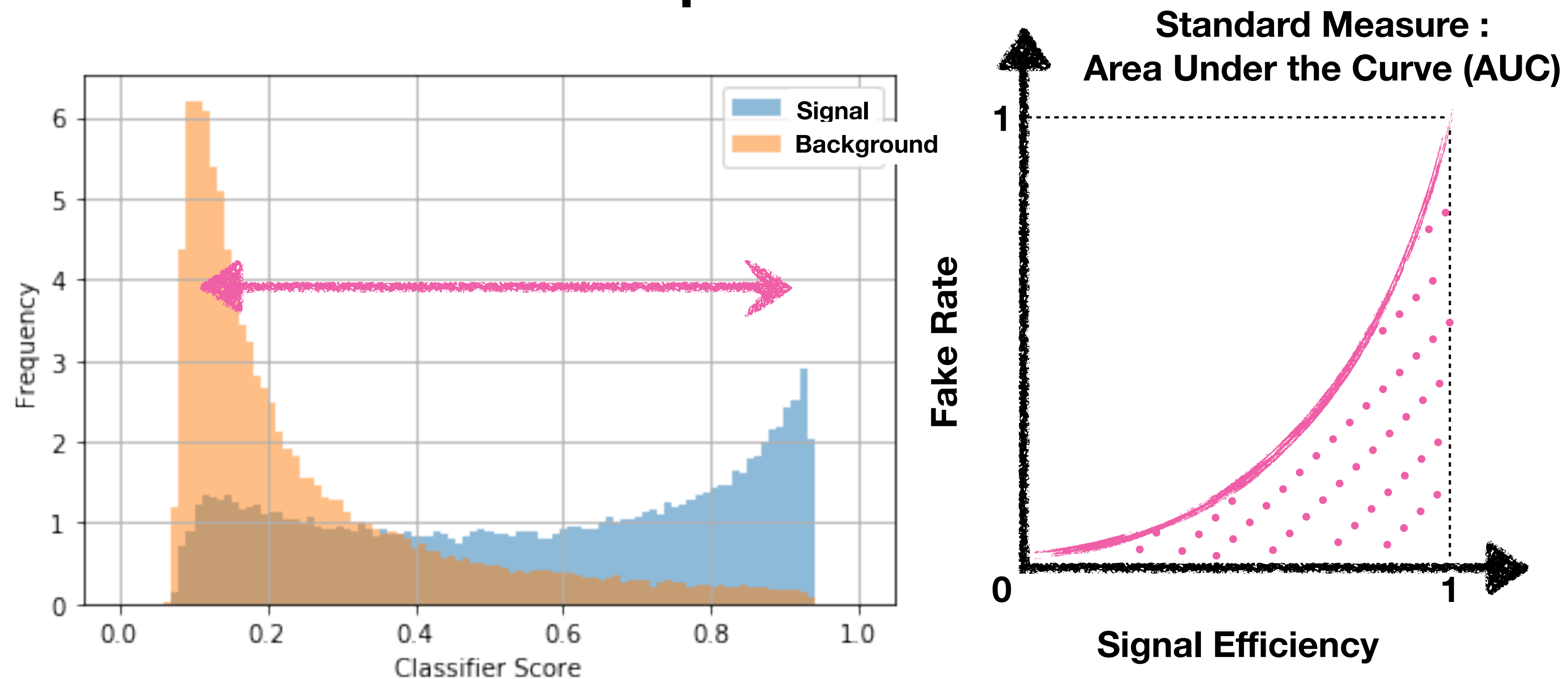
Receiver Operator Characteristic



- Scanning across the classifier score, builds up a ROC

ROC Curves

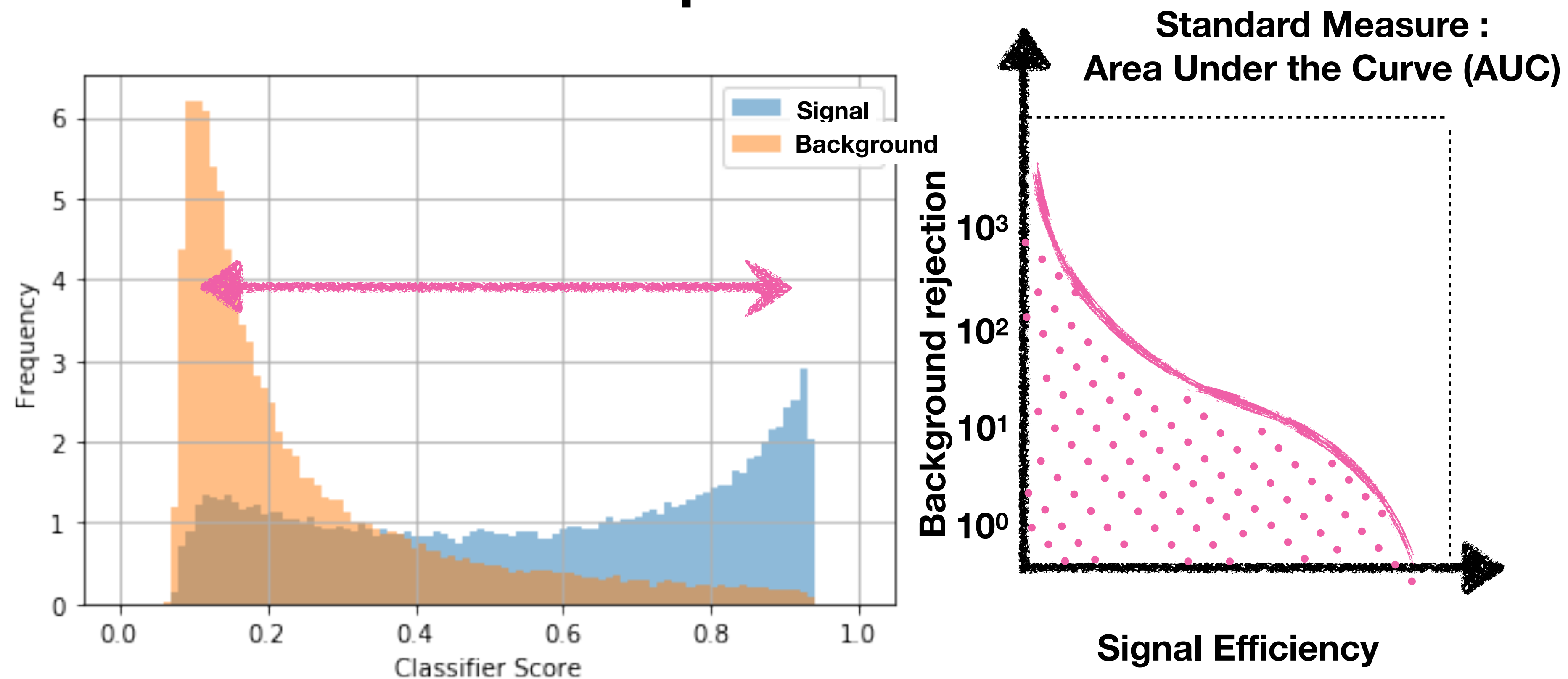
Receiver Operator Characteristic



- AUC useful measure for comparing models
- NOTE: not an ultimate measure

ROC Curves

Receiver Operator Characteristic



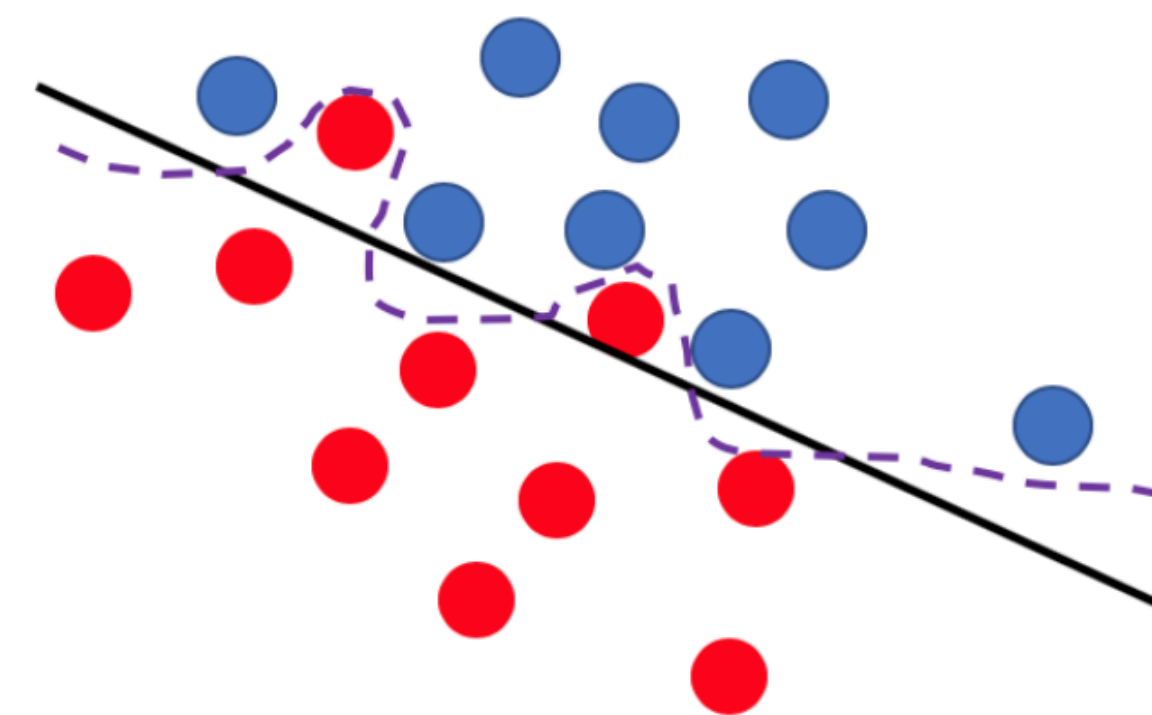
- Sometimes, the ROC curve is represented as efficiency vs background rejection (in log scale highlights differences between similar curves)

Model validation

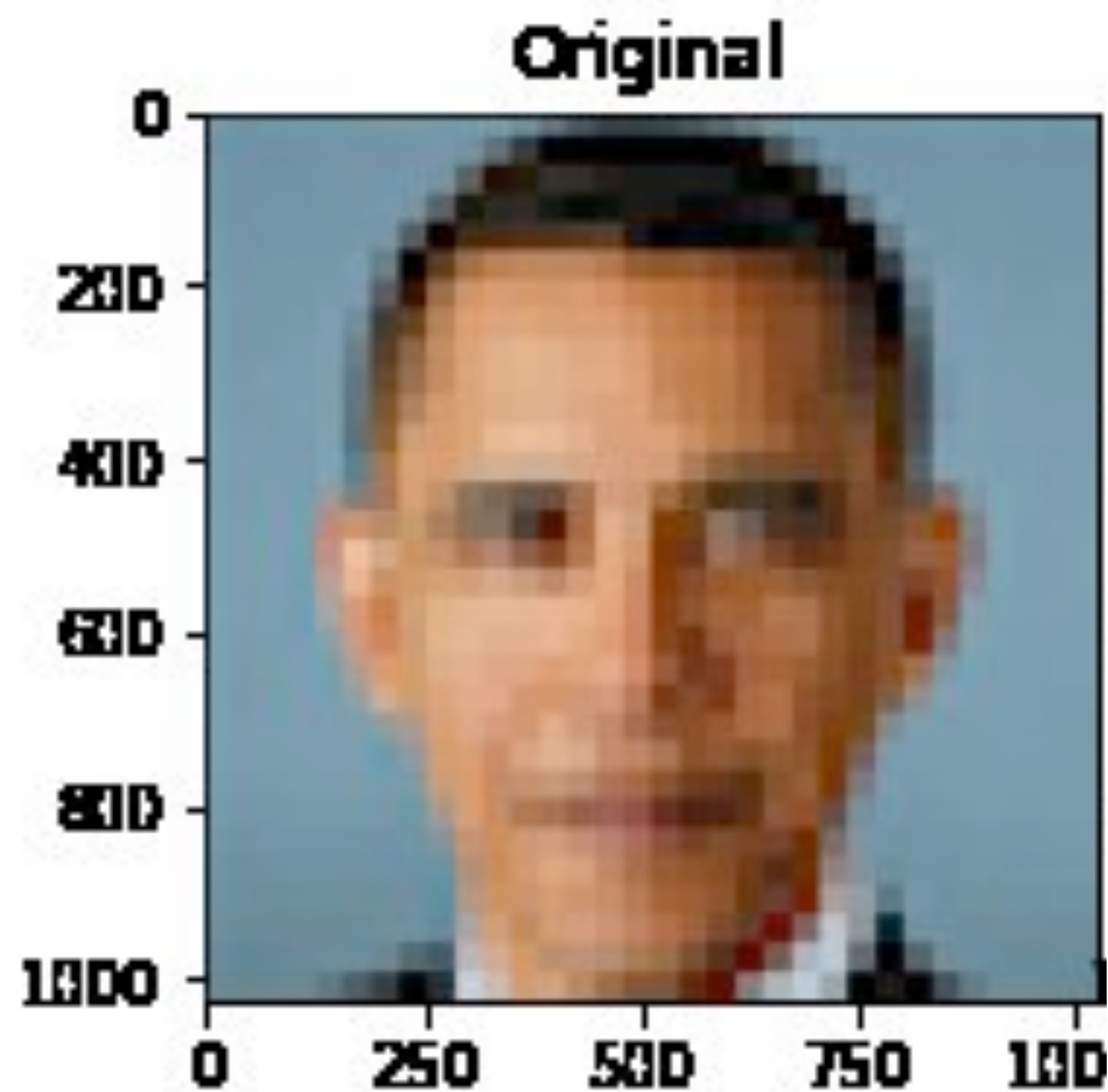
ML is not a magic fix!

- *Put garbage in, get garbage out!*
- ML cannot replace domain knowledge.
- ML is not a causation tool.
- Model should be generalizable (i.e. should perform well on unseen data).

Hannah Bossi (Yale) RHIC/AUM 2021

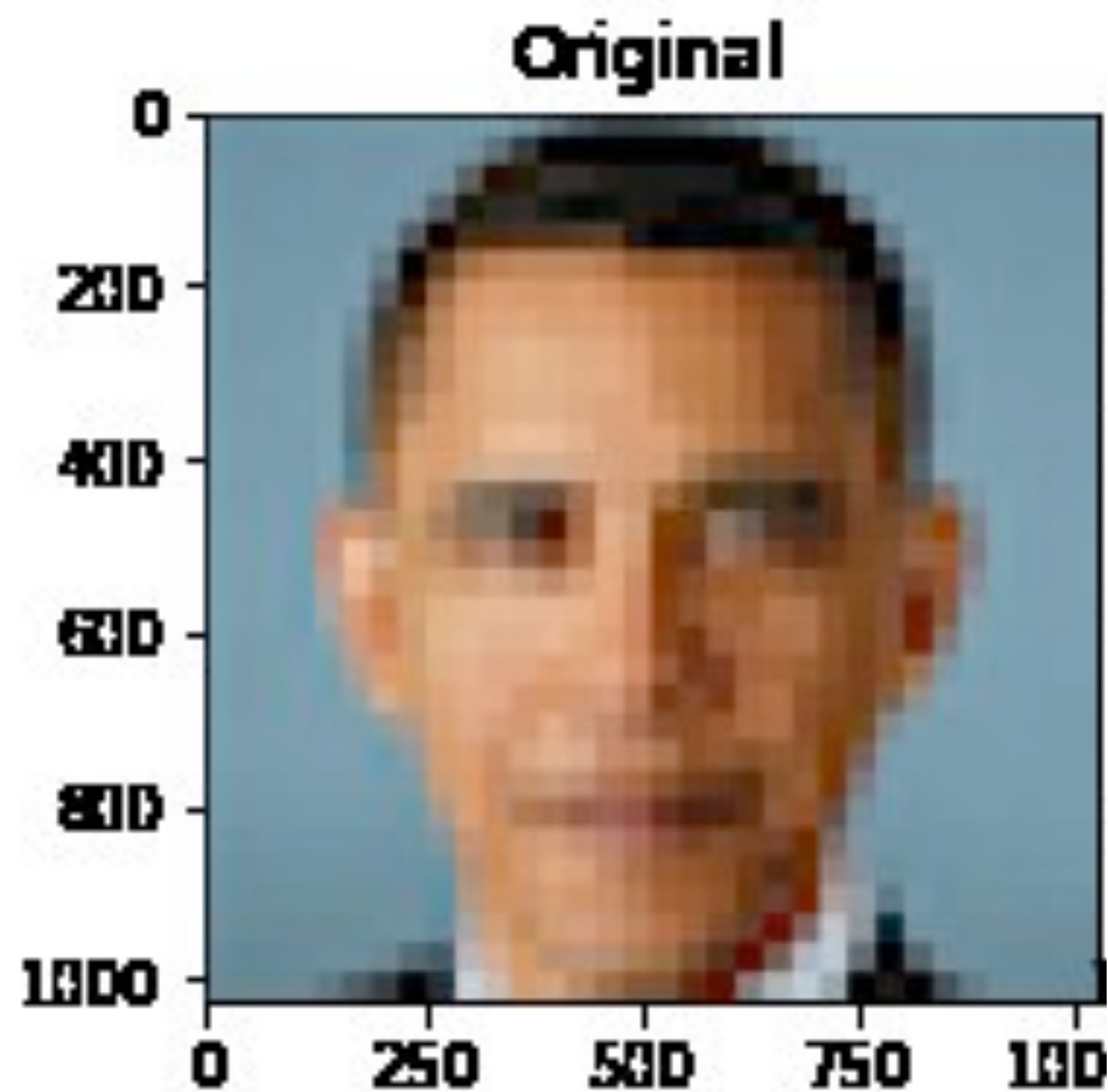


Quick example



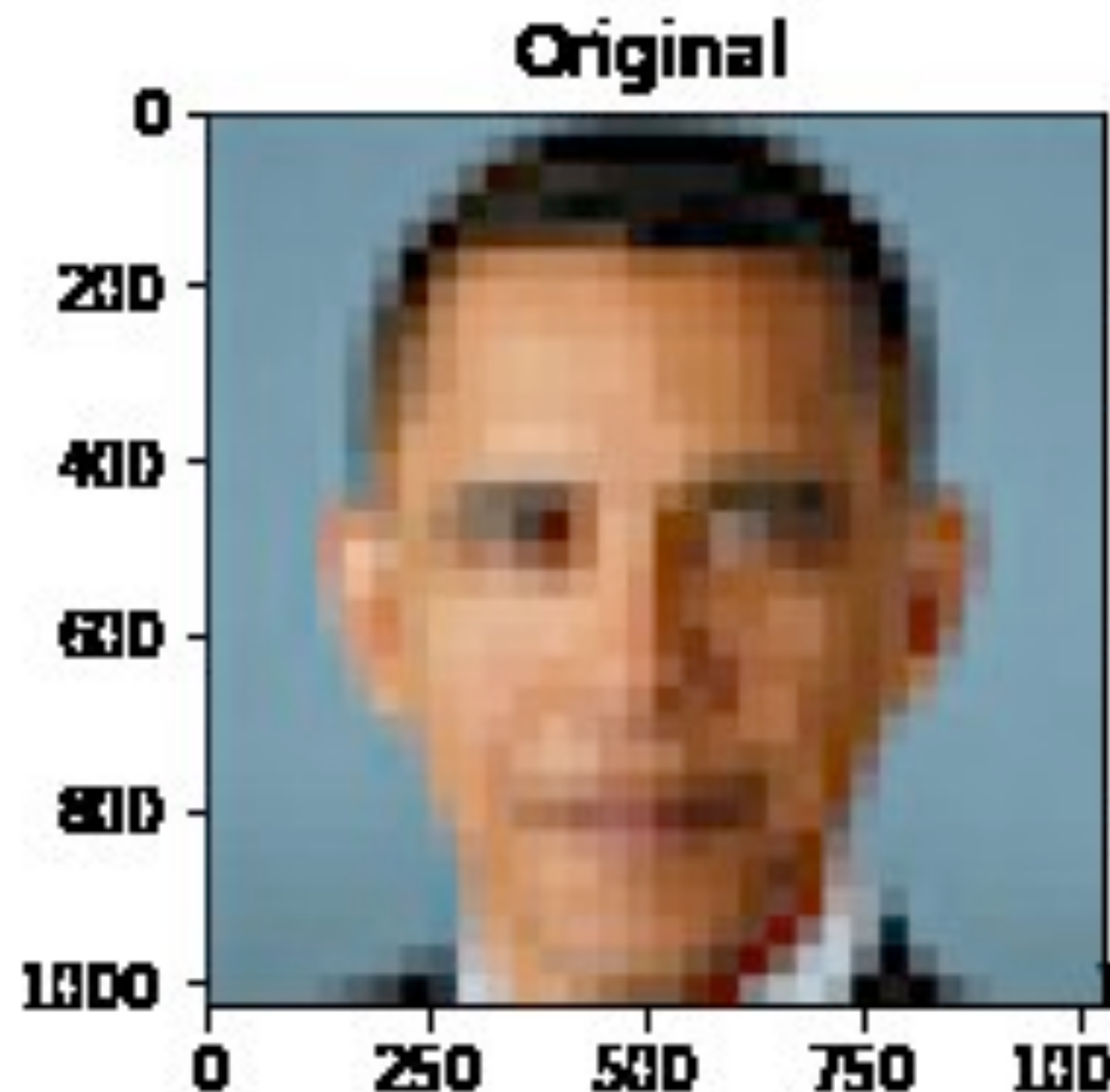
https://colab.research.google.com/github/tg-bomze/Face-Depixelizer/blob/master/Face_Depixelizer_Eng.ipynb

Quick example



https://colab.research.google.com/github/tg-bomze/Face-Depixelizer/blob/master/Face_Depixelizer_Eng.ipynb

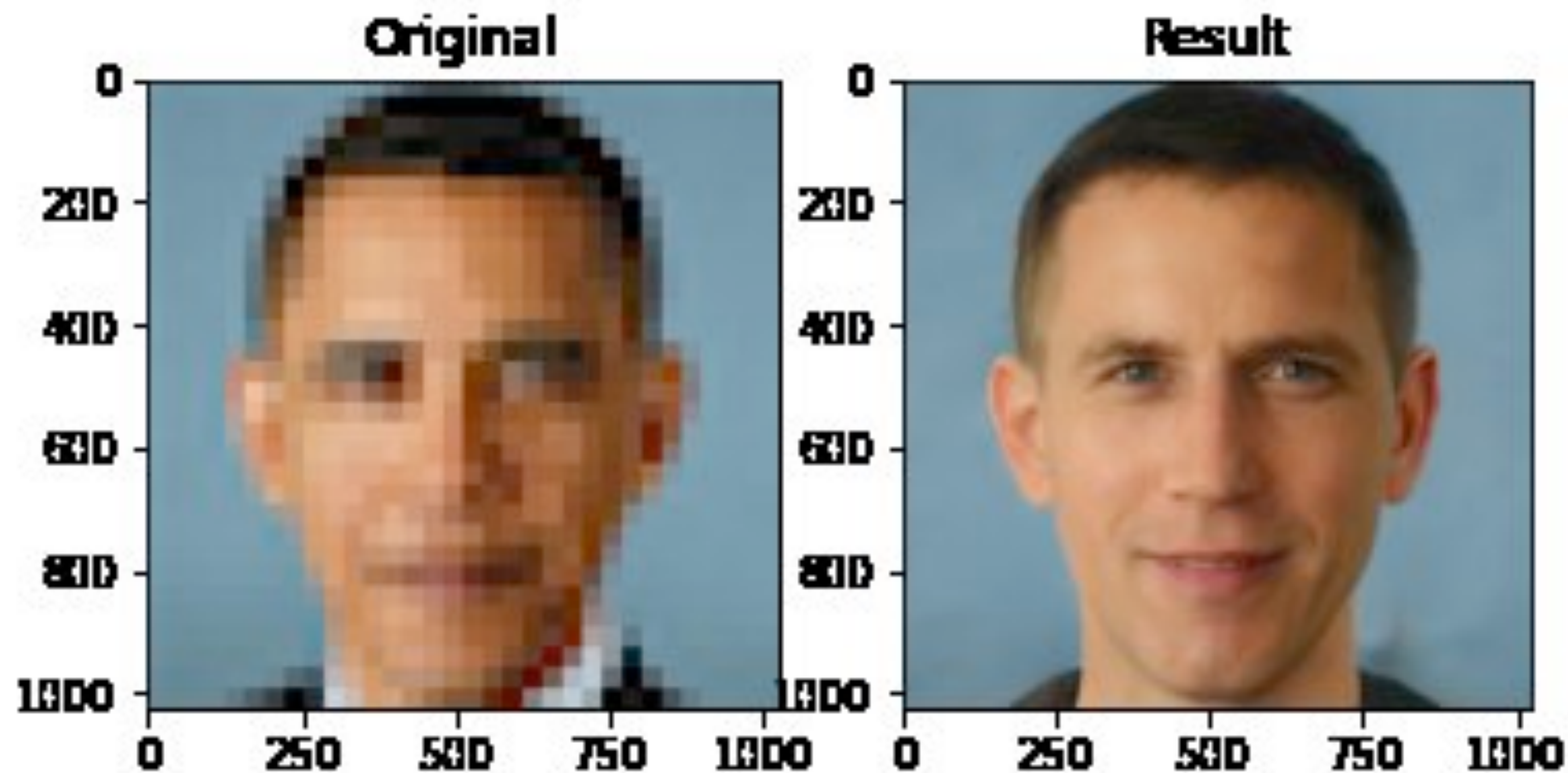
Quick example



What do you think is going to happen?

https://colab.research.google.com/github/tg-bomze/Face-Depixelizer/blob/master/Face_Depixelizer_Eng.ipynb

Quick example



https://colab.research.google.com/github/tg-bomze/Face-Depixelizer/blob/master/Face_Depixelizer_Eng.ipynb

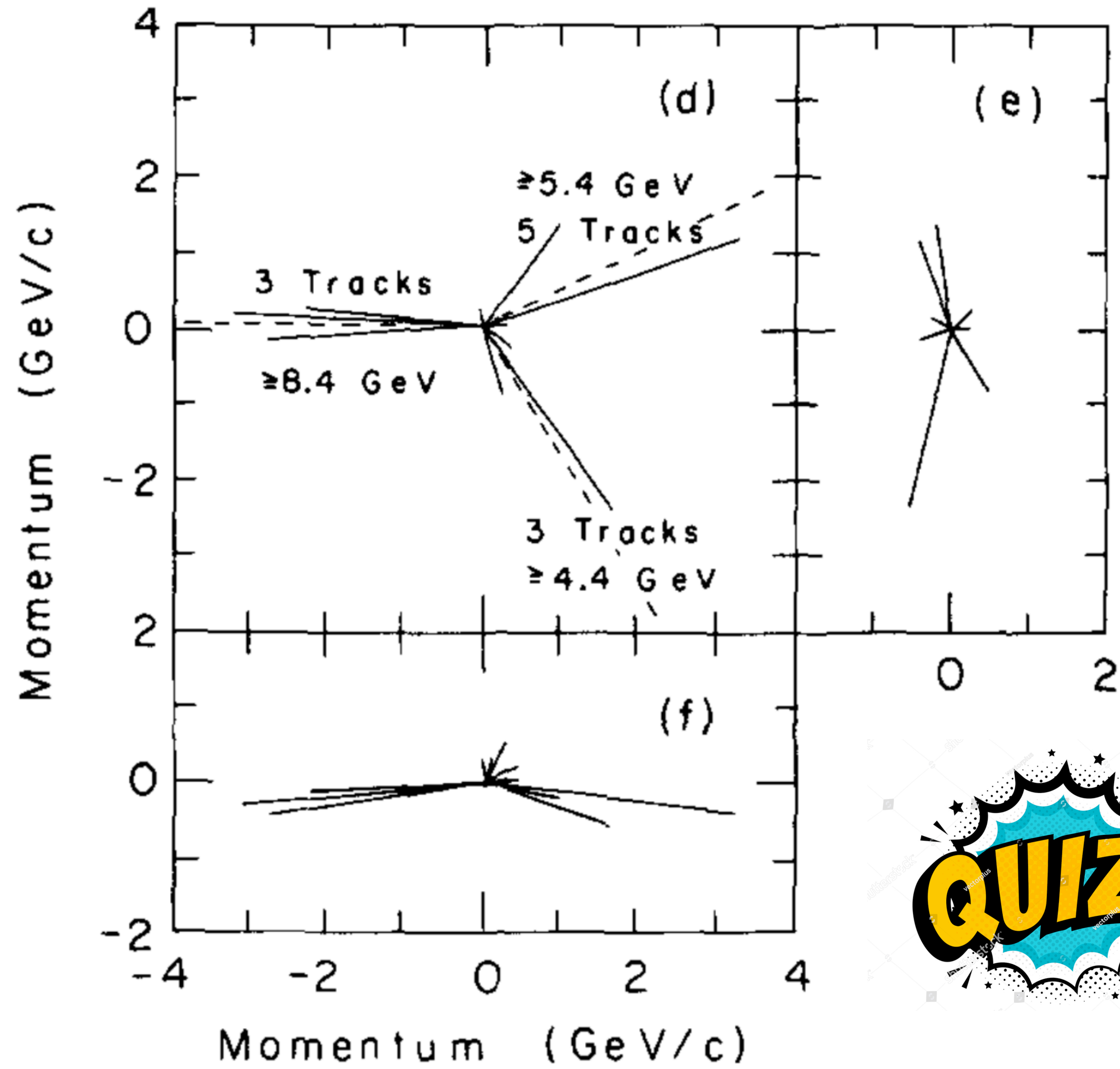
Recap - 1

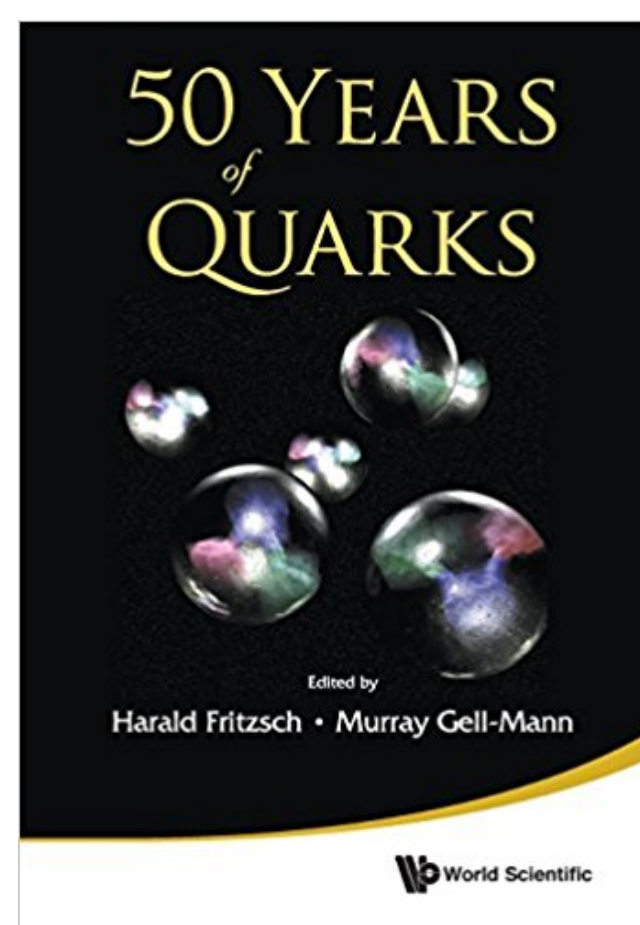
- Direct connection between analysis tools in physics and modern ML techniques
- The ML community has advanced and well calibrated classification (and regression, as we shall see) tools for us to exploit and adapt to our purposes
- Garbage in... Garbage out. Biased data will result in unreliable biased learning

Onto some classification!

Input representation vs Model Choice

- Input variables - BDTs, MLPs,
- Images - CNNs
- Sequential info - Recursive Neural Networks, Long Short-Term Memory units
- Jets - can go for any of these
- Even graphs, transformers, vectors (whats in our tutorial notebook)

**QUIZ!**



The Discovery of the Gluon

<https://arxiv.org/abs/1409.4232>

John Ellis

*Theoretical Particle Physics and Cosmology Group, Department of Physics,
King's College London, London WC2R 2LS, United Kingdom
& Theory Division, CERN, CH-1211 Geneva 23, Switzerland,
John.Ellis@cern.ch*

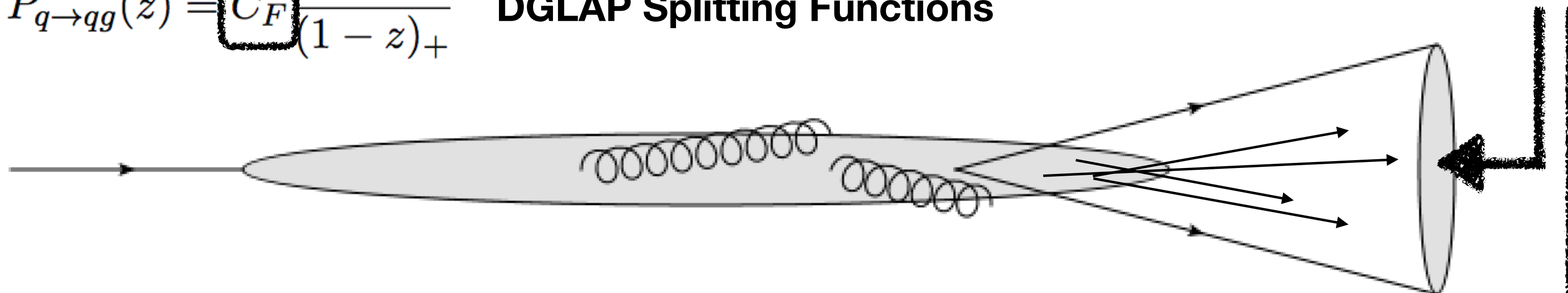
The public announcement of the gluon discovery came at the Lepton/Photon Symposium held at Fermilab in August 1979. All four PETRA experiments showed evidence: JADE and PLUTO followed TASSO in presenting evidence for jet broadening and three-jet events as suggested in our 1976 paper, while the Mark J collaboration led by Sam Ting presented an analysis of antenna patterns along the lines of our 1978 paper. There was a press conference at which one of the three-jet events was presented, and a journalist asked which jet was the gluon. He was told that the smart money was on the jet on the left (or was it the right?). Refereed publications by TASSO [37] and the other PETRA collaborations [38] soon appeared, and the gluon finally joined the Pantheon of established particles as the second gauge boson to be discovered, joining the photon.



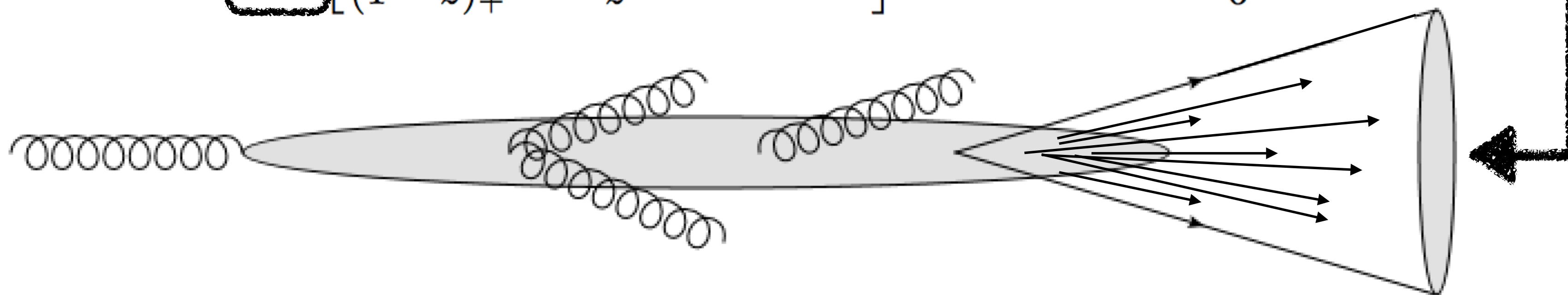
Jet Algorithm

$$P_{q \rightarrow qg}(z) = \boxed{C_F} \frac{1+z^2}{(1-z)_+}$$

DGLAP Splitting Functions



$$P_{g \rightarrow gg}(z) = \boxed{2C_A} \left[\frac{z}{(1-z)_+} + \frac{1-z}{z} + z(1-z) \right] + \delta(1-z) \frac{11C_A - 4n_f T_R}{6}$$

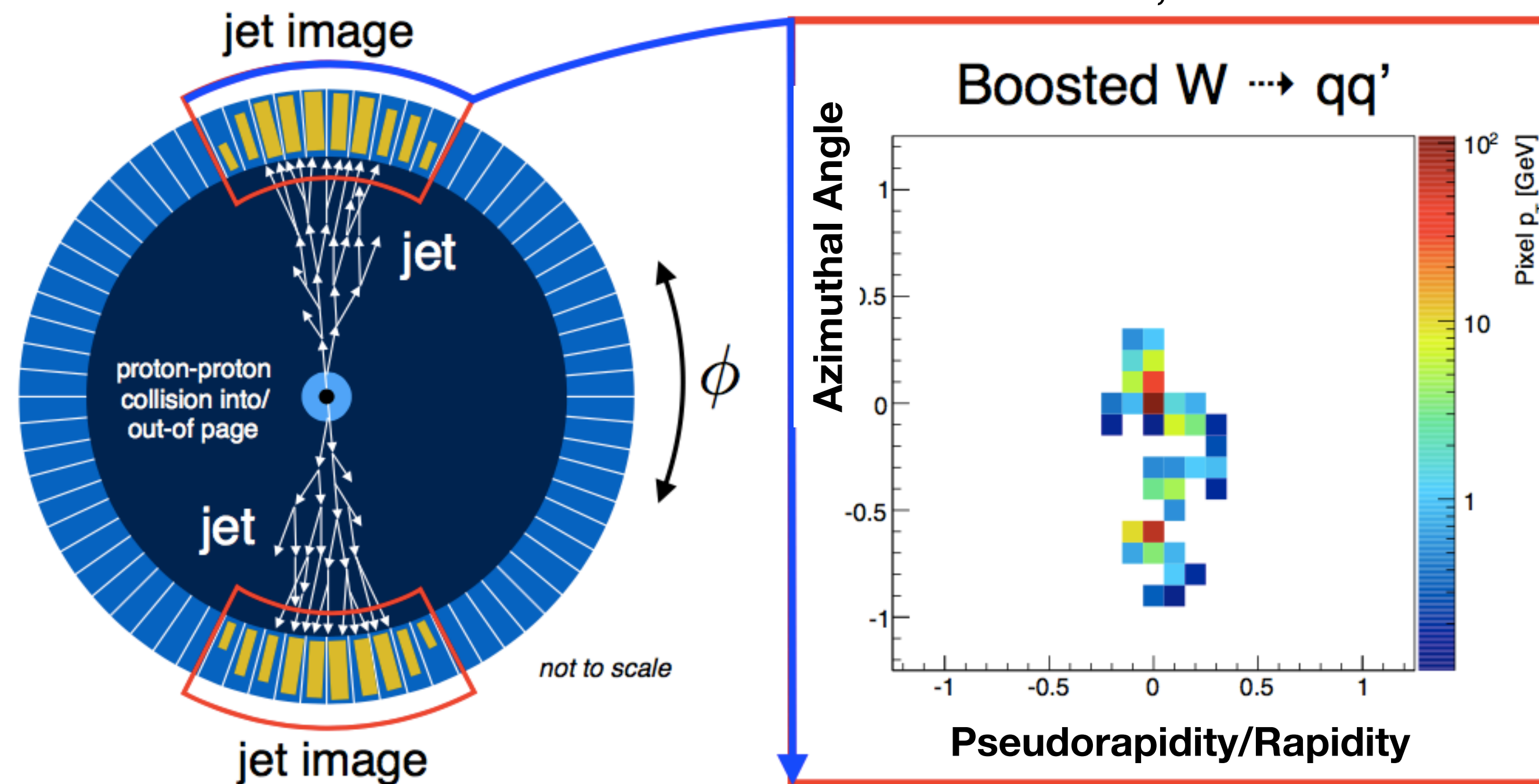


C_A vs C_F primary difference could potentially lead to larger multiplicity and overall width for gluon jets

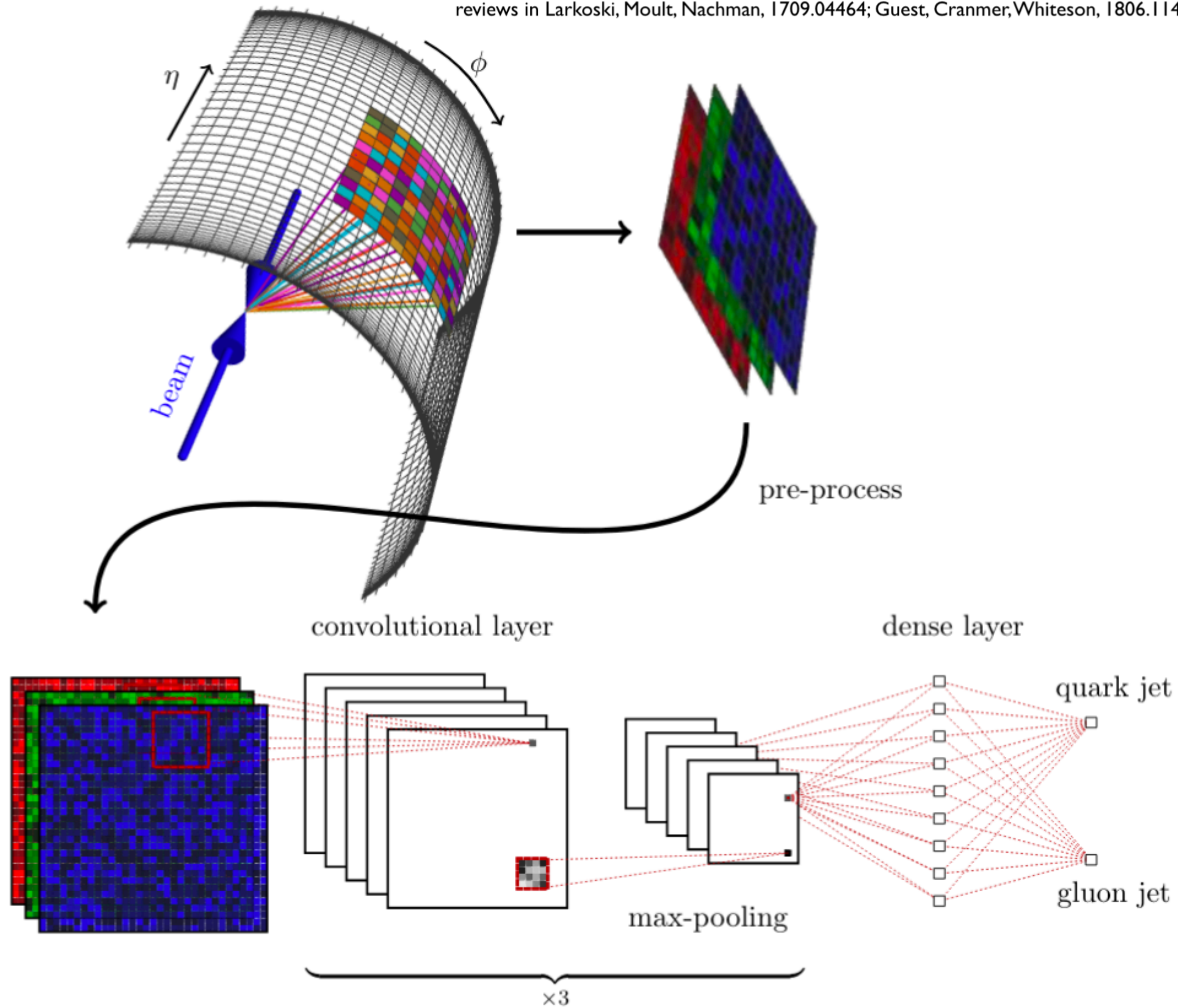
What is a Jet Image?

- Jet Image: A two-dimensional fixed representation of radiation pattern inside a jet

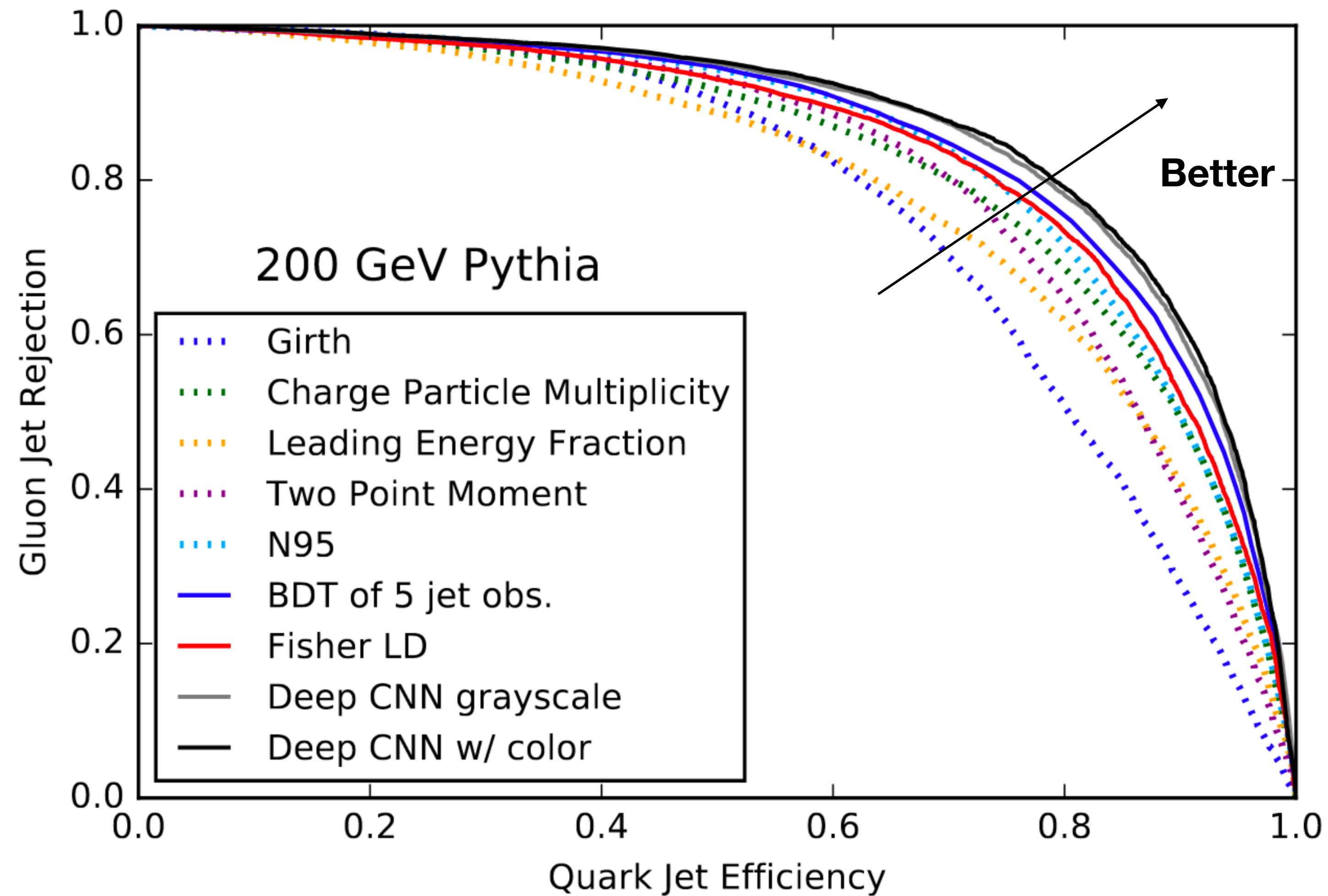
Ben Nachman, DS@HEP 2017



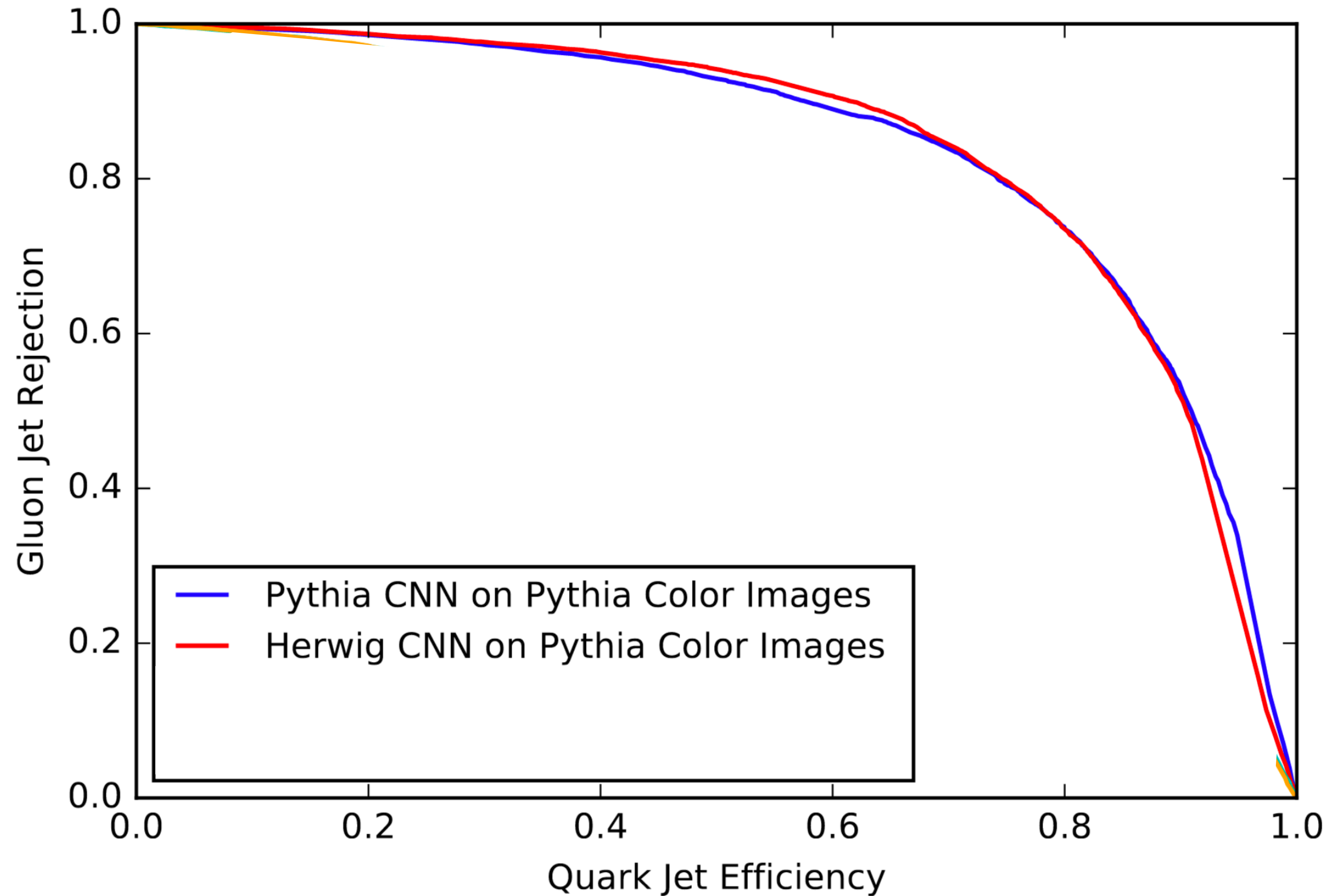
- Including pre-processing to fix center at (0,0), rotating image ($\Delta\eta$ - $\Delta\phi$) so that secondary peak lies along 6-o clock direction



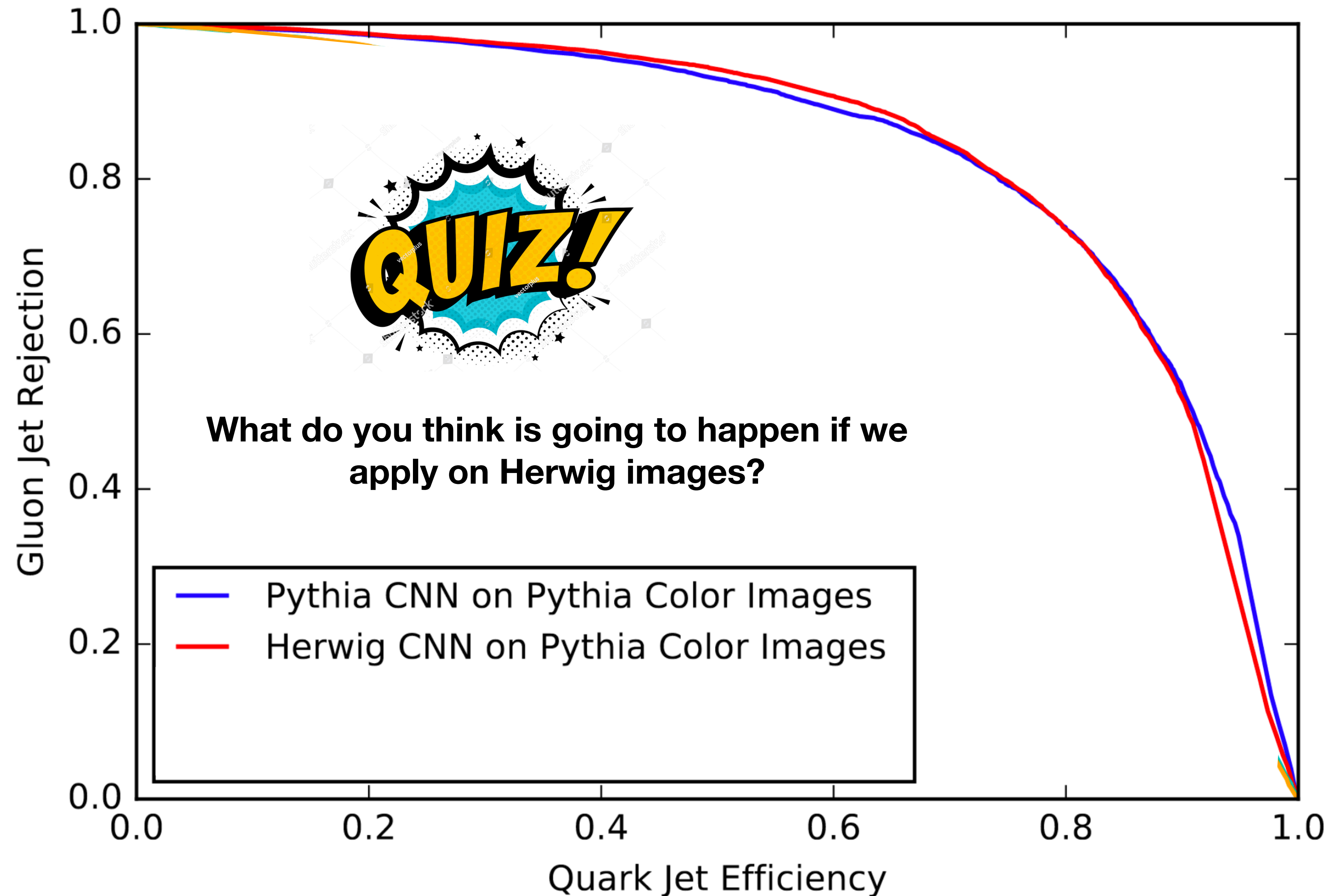
Comparing a variety of methods



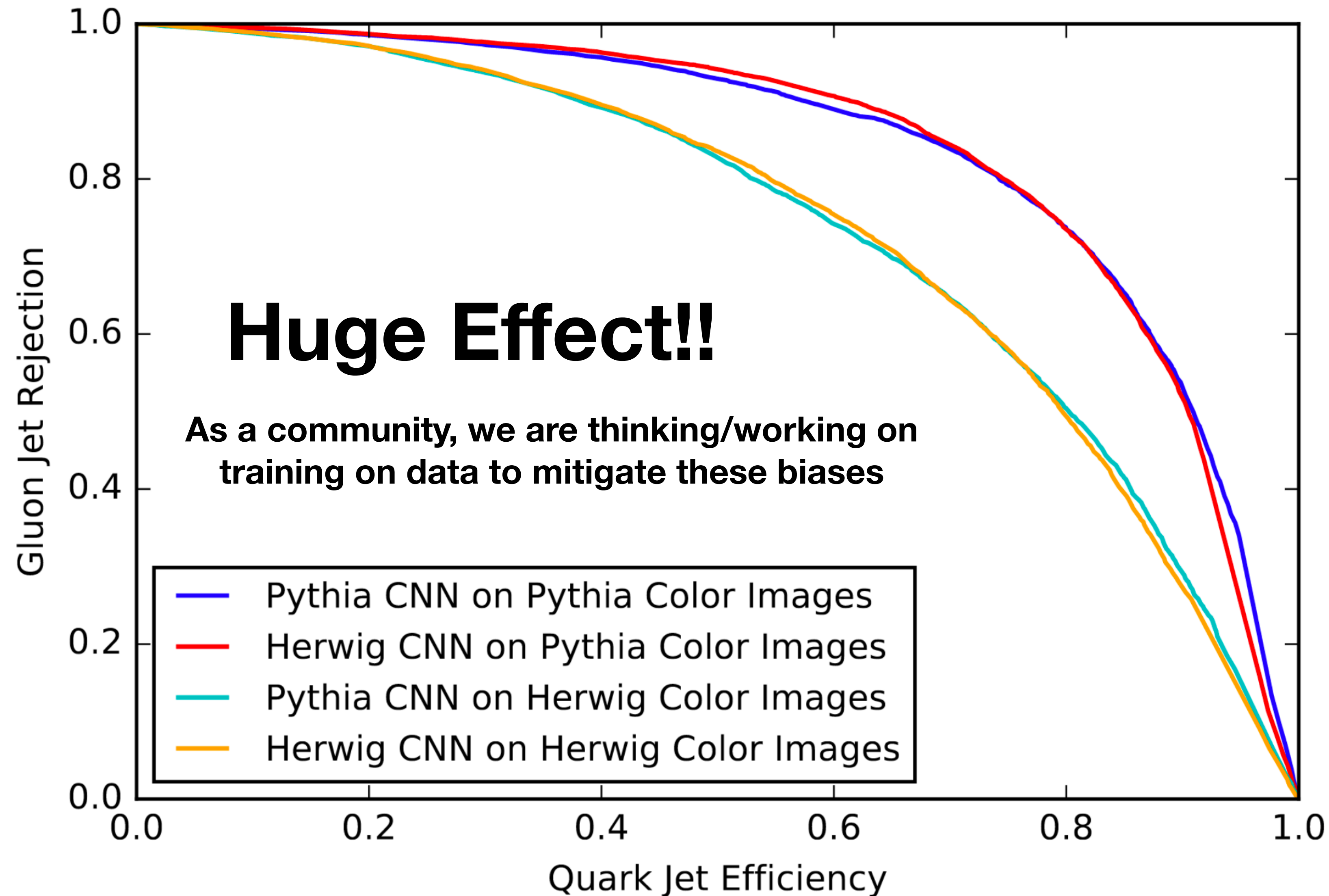
Effect of the MC training bias



Effect of the MC training bias



Effect of the MC training bias



Recap - 2

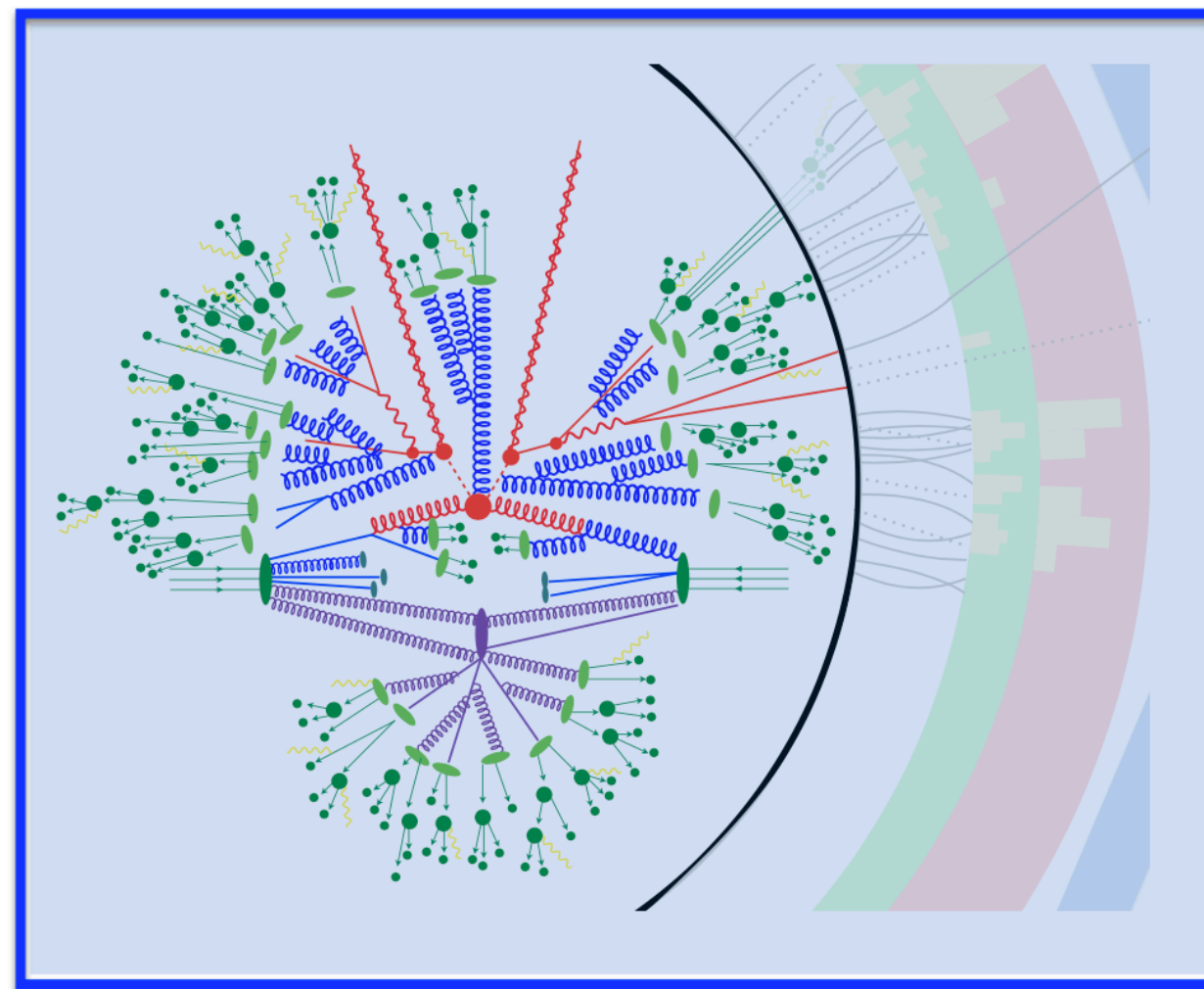
- Classifiers in our field are mostly supervised - with a potential built-in bias (utilize it!)
- There are many different ways to represent jets - information content is available to be exploited
- They are deployed in pp, but how about it in heavy ions?

Lets regress the truth!

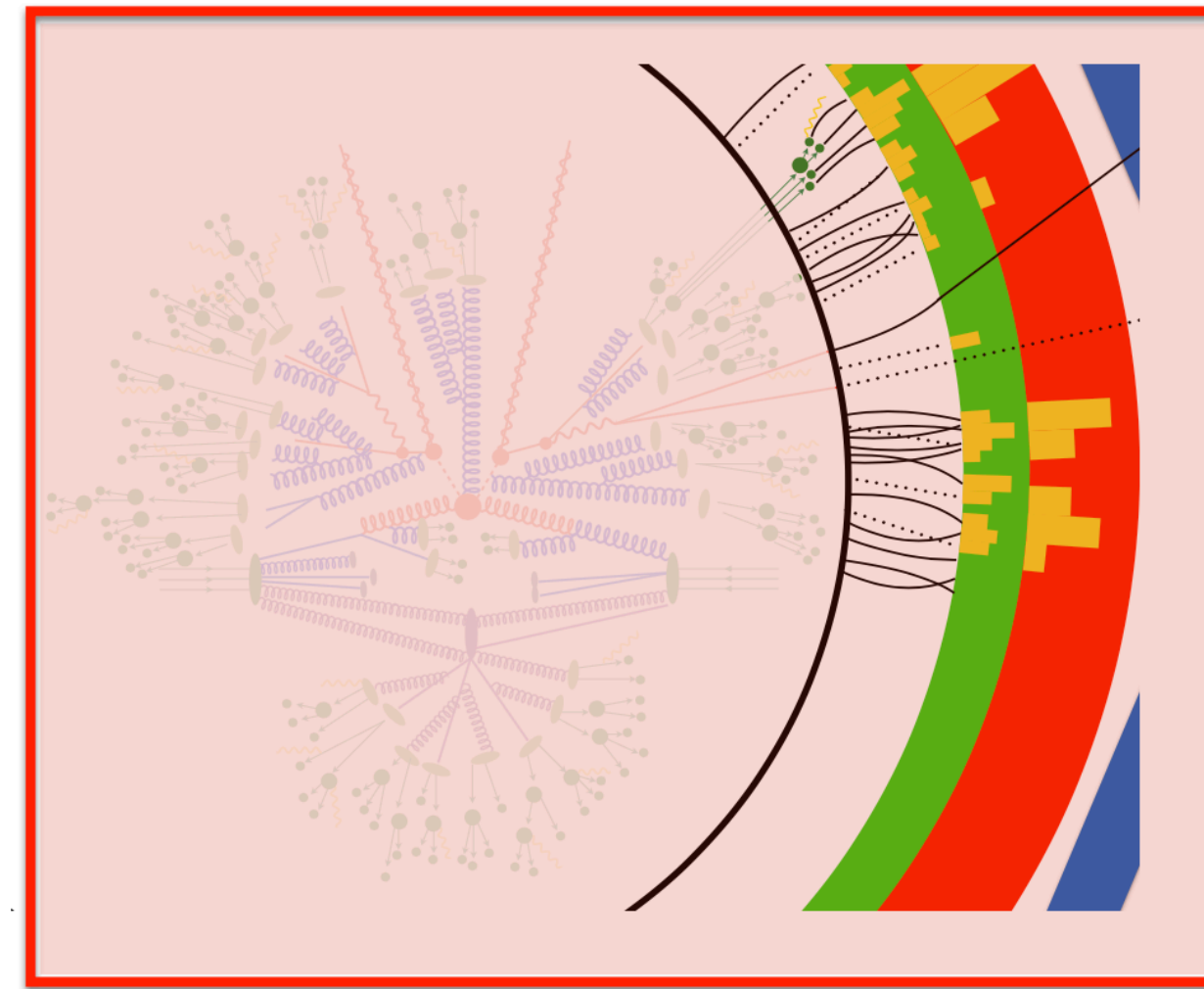
Regression in HEP

- Correction procedures for energy scales and resolutions
- Multi-dimensional unfolding techniques

Want this

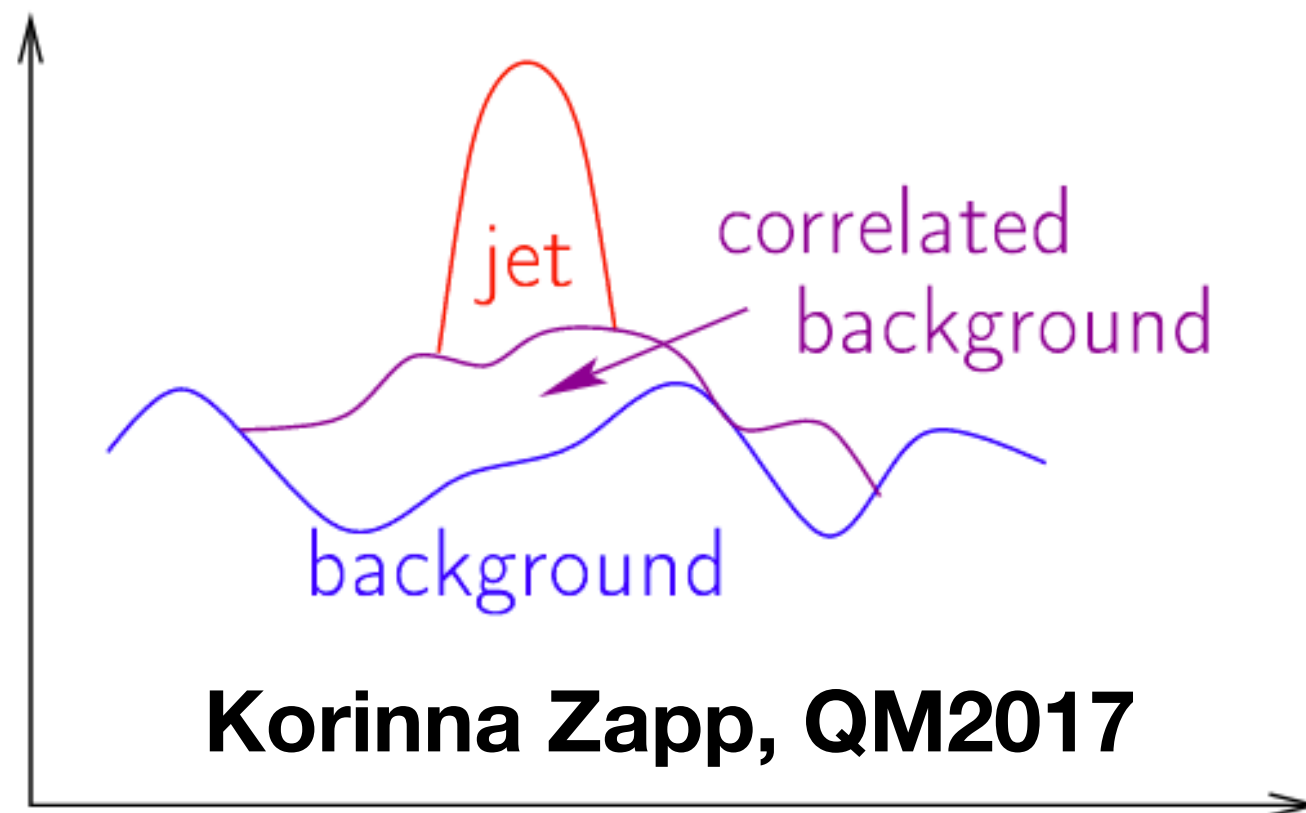


Measure this

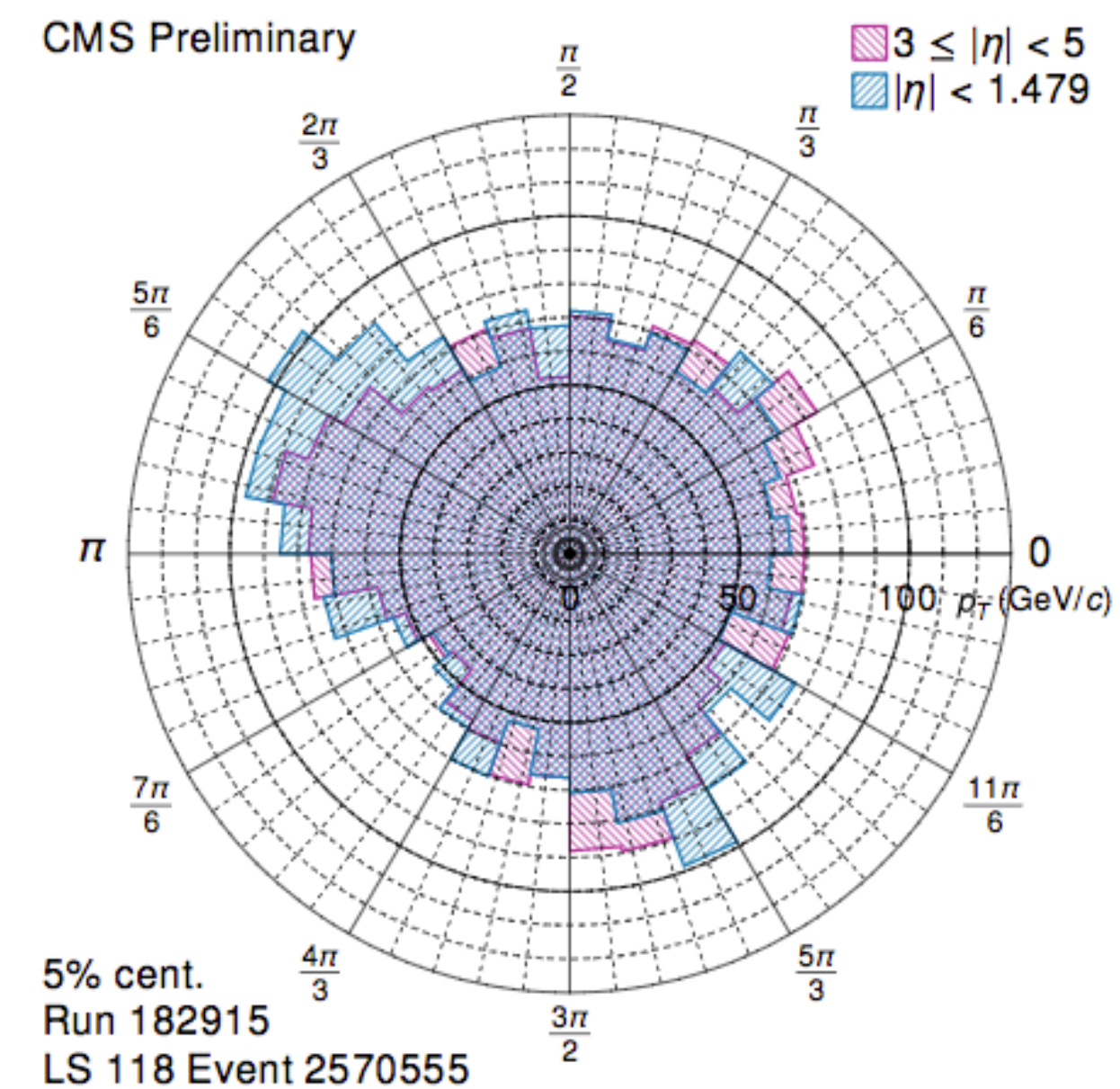
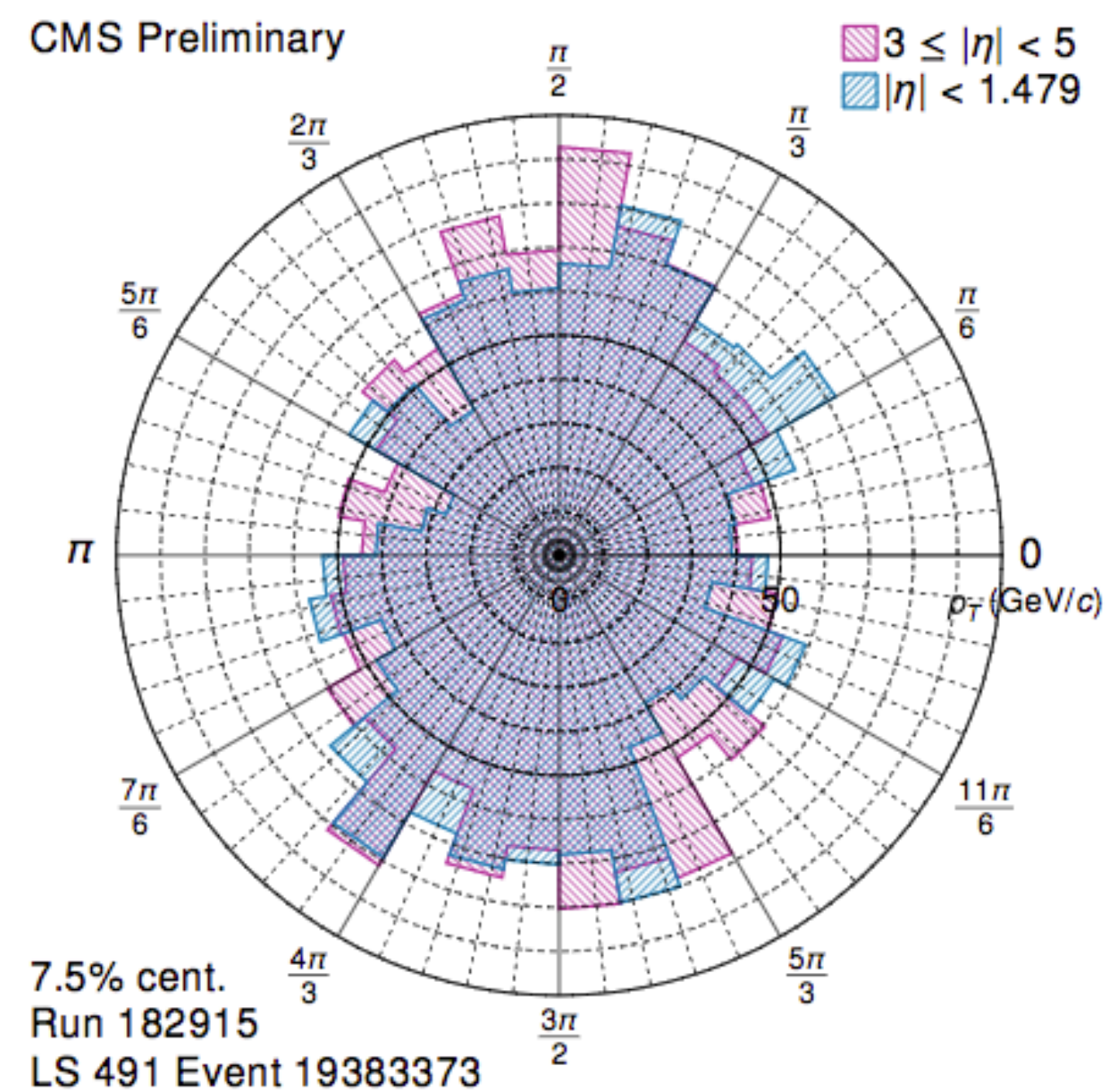


*Image credit:
Ben Nachman*

Impact of the heavy ion background

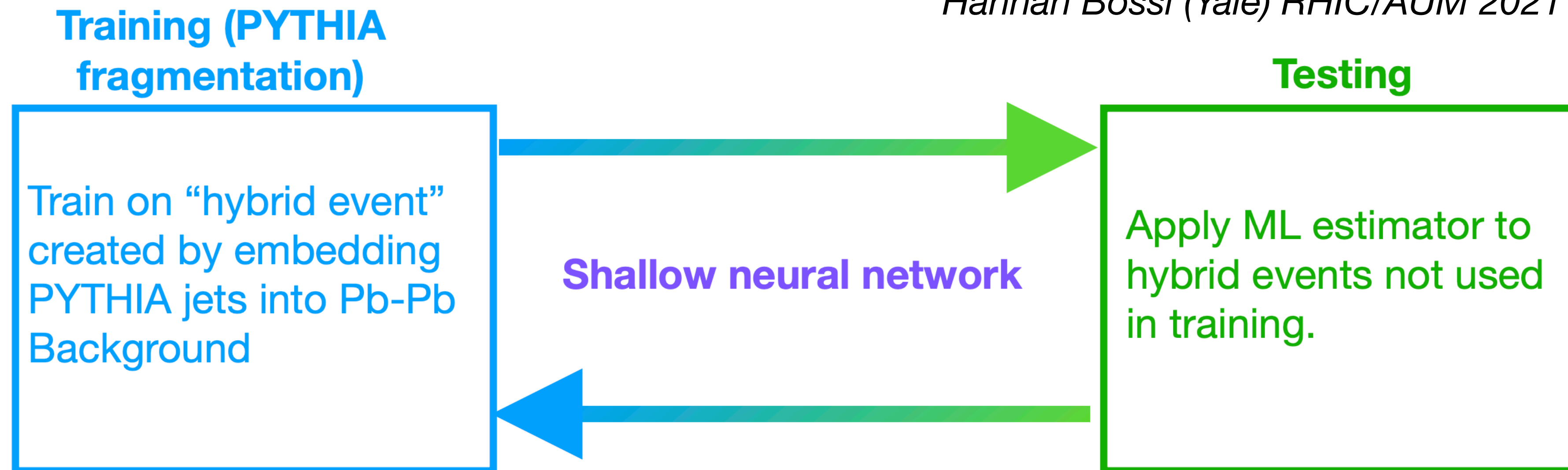


Underlying event has flow, fluctuations and is correlated with the jet (like a wake)



ALICE method of ML based subtraction

Hannah Bossi (Yale) RHIC/AUM 2021

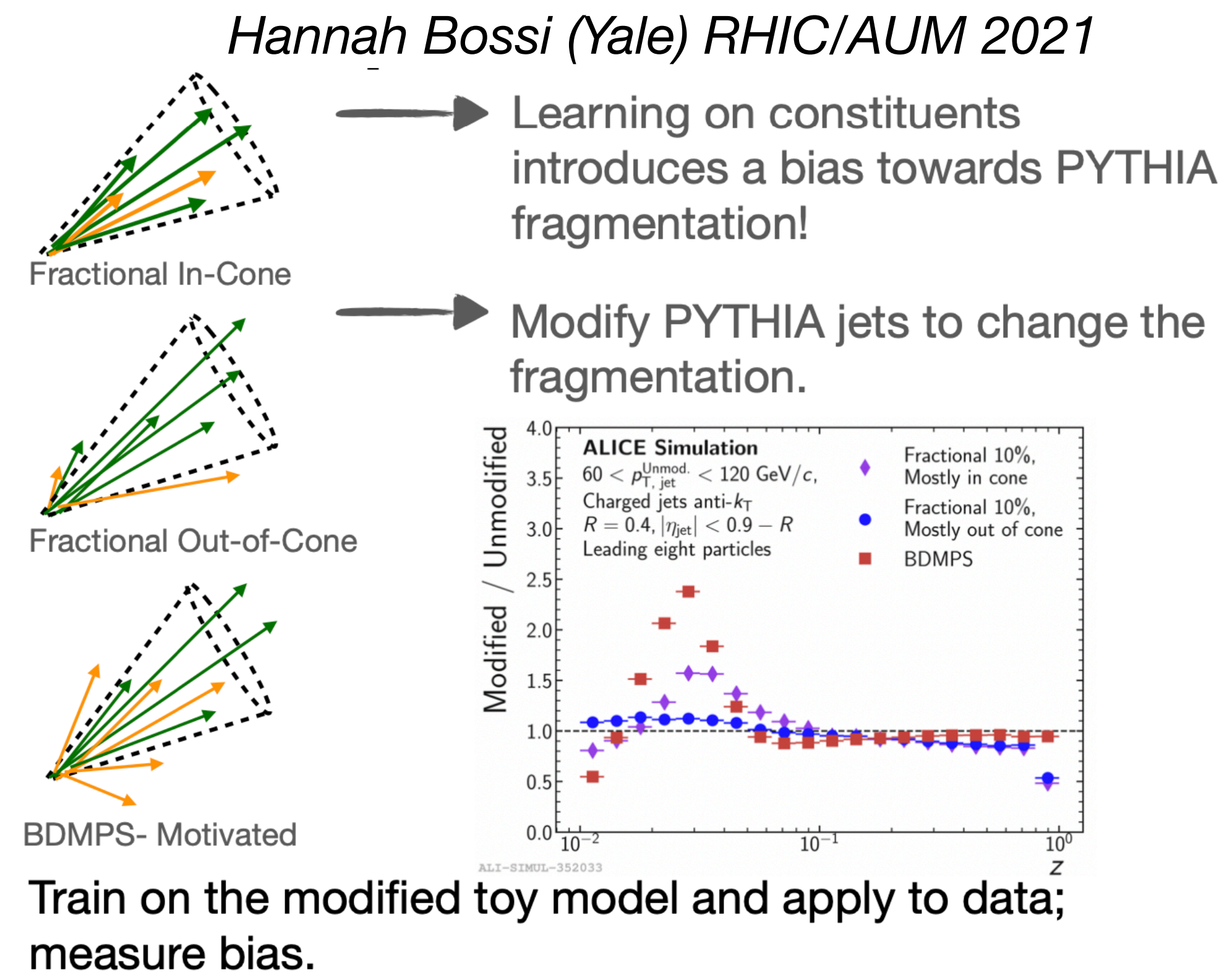
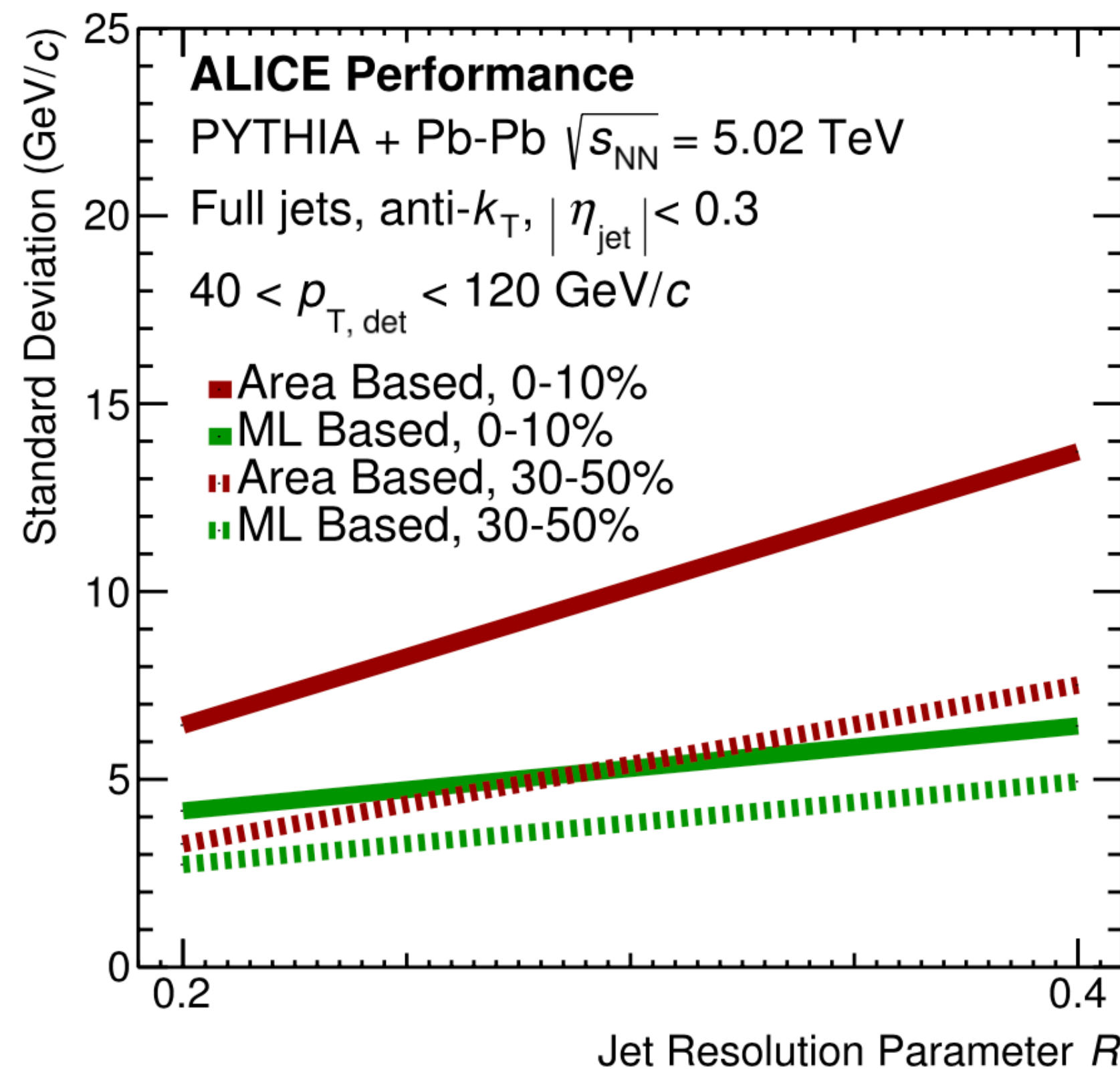


Key is that this background is *realistic*.

Simple question,
relatively simple network
can get a short clear answer!

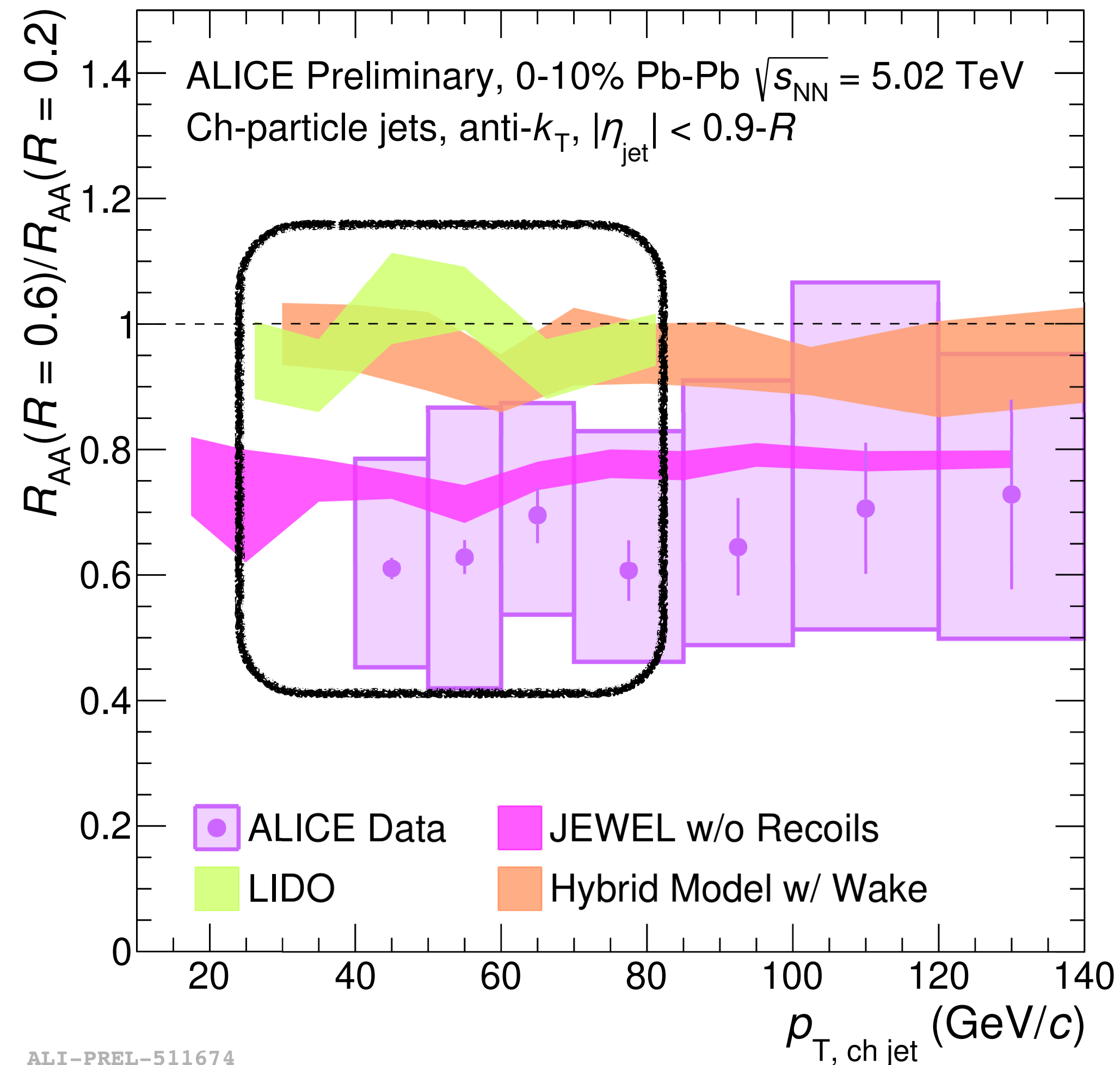
Do we get back the
signal we put in?

ML corrector in action



- Significantly less jet energy resolution with the ML based method along with first ever estimate of impact of truth shape ‘bias’ in correction

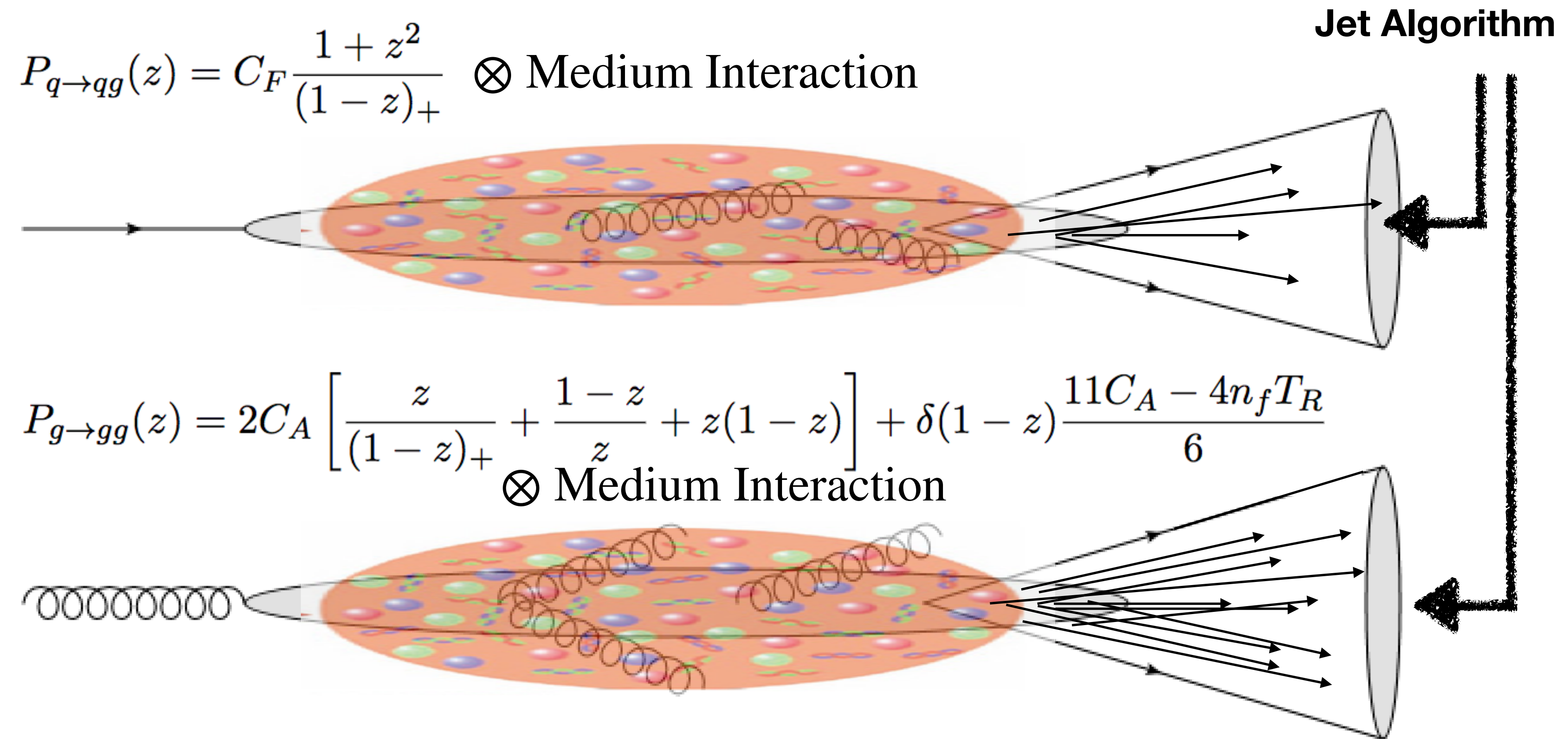
What did it enable us to do?



ALI-PREL-511674

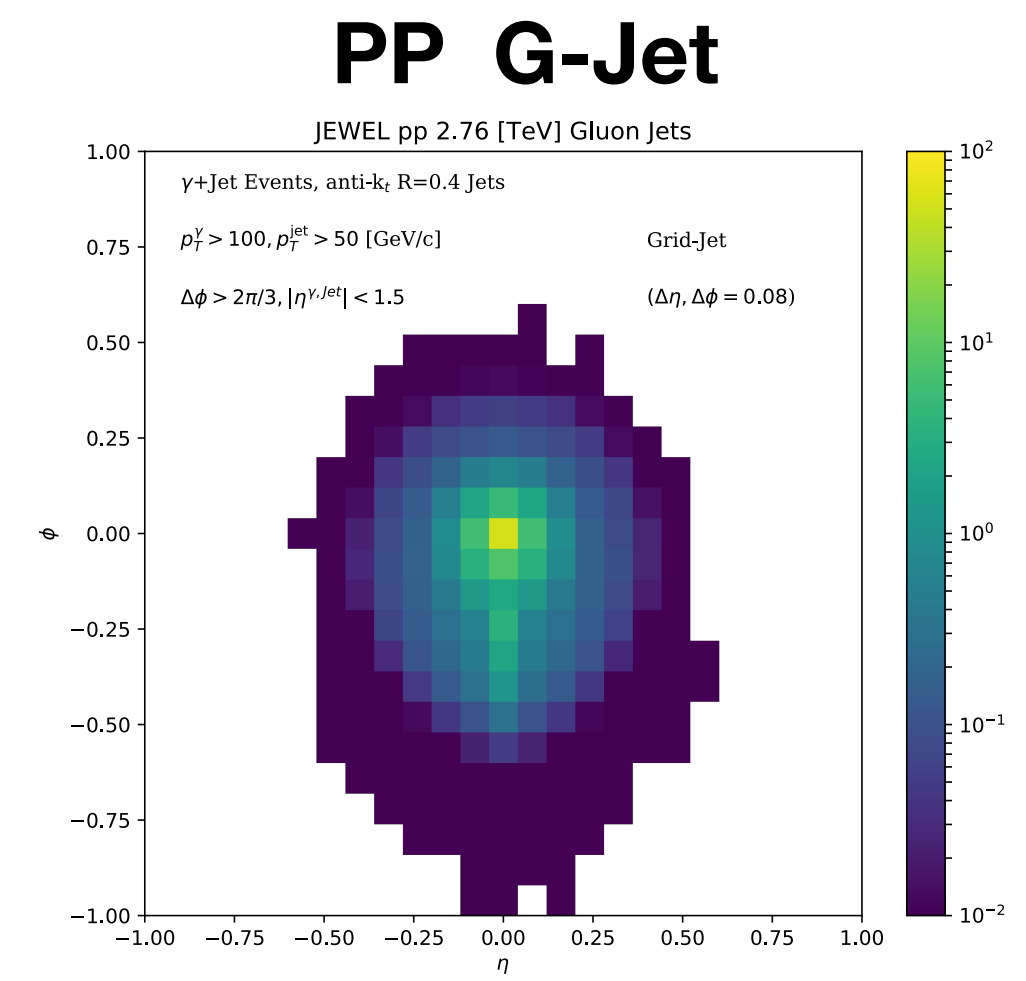
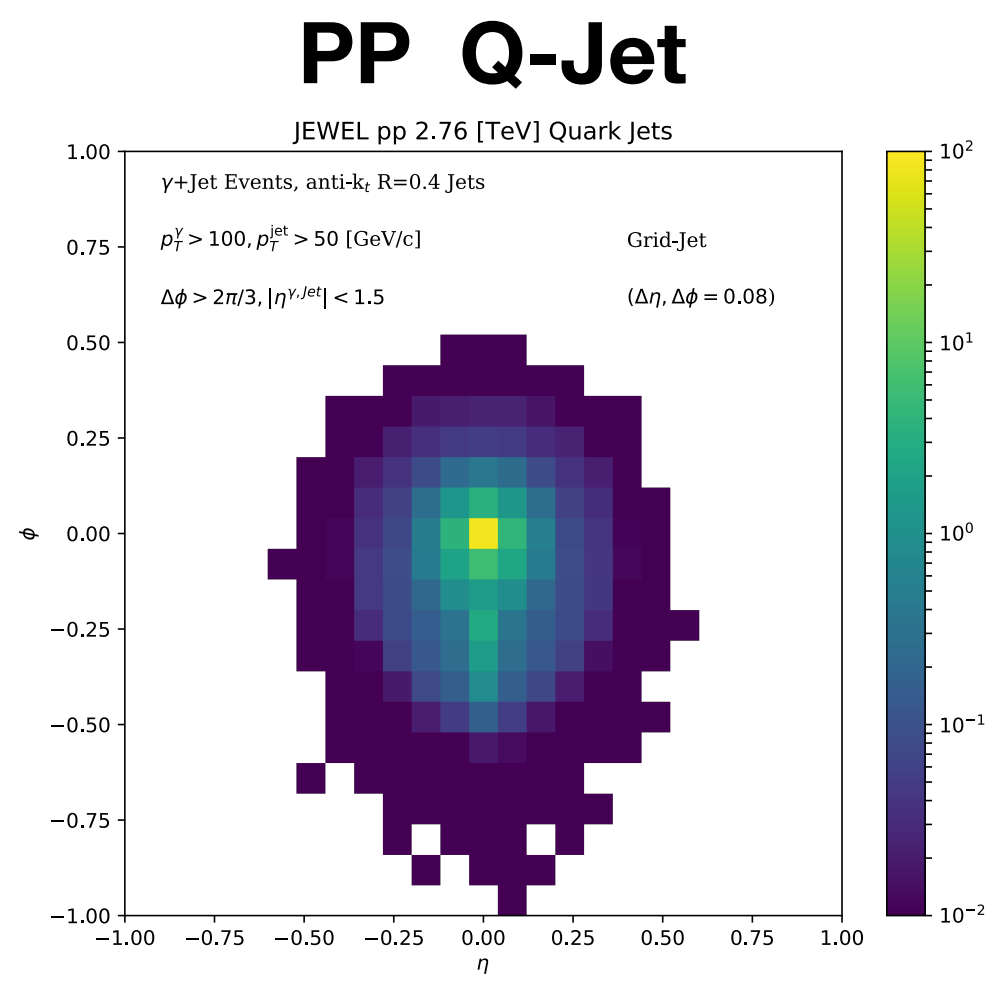
- Extend our measurements to lower momentum range where the impact of the background is large
- Reduced uncertainties key to making a potentially tantalizing statement about radial dependence of energy loss

Quark and Gluon Jets



**Effects of the QGP on jet propagation manifests
via modifications to jet energy and jet sub-structure**

Jet Images



QCD Color factor
G-Jets are broader

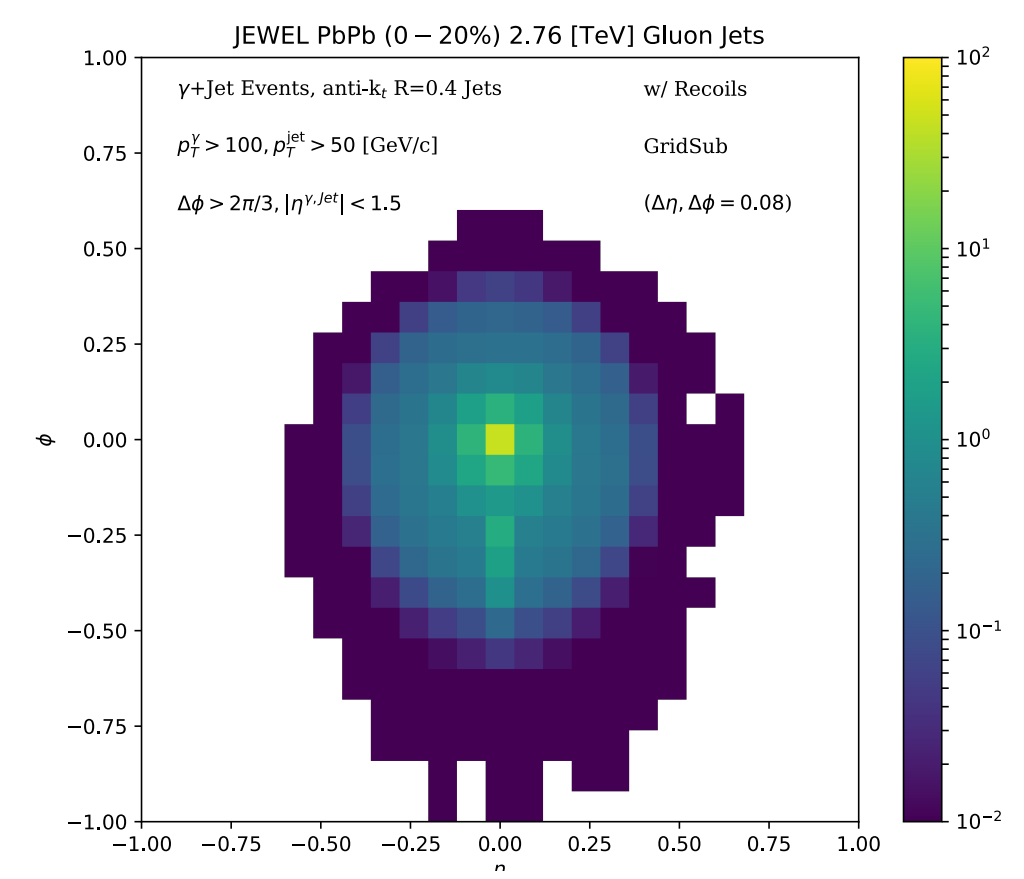
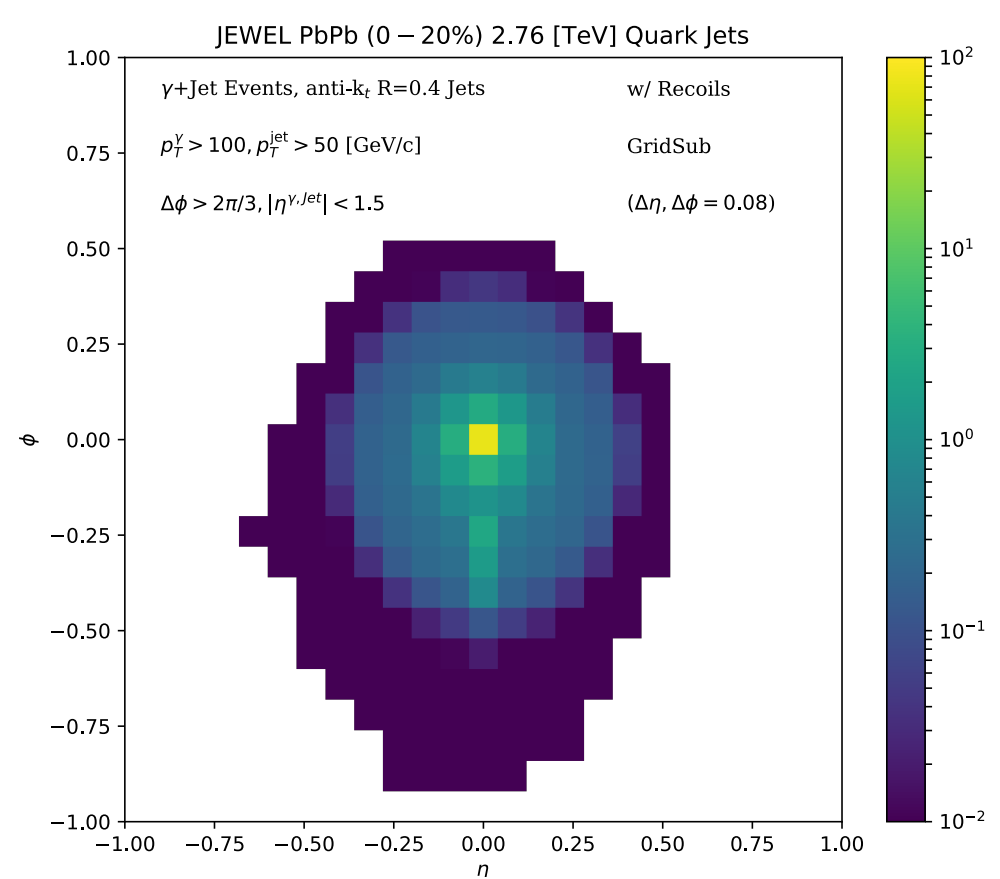
JEWEL

JEWEL

- Pros
- Image representation should contain all info.
 - Current State of the art - easy to implement
- Cons
- classification in non-physics basis
 - Best case scenario - no fluctuating background!

Quenched Q-Jet

Quenched G-Jet



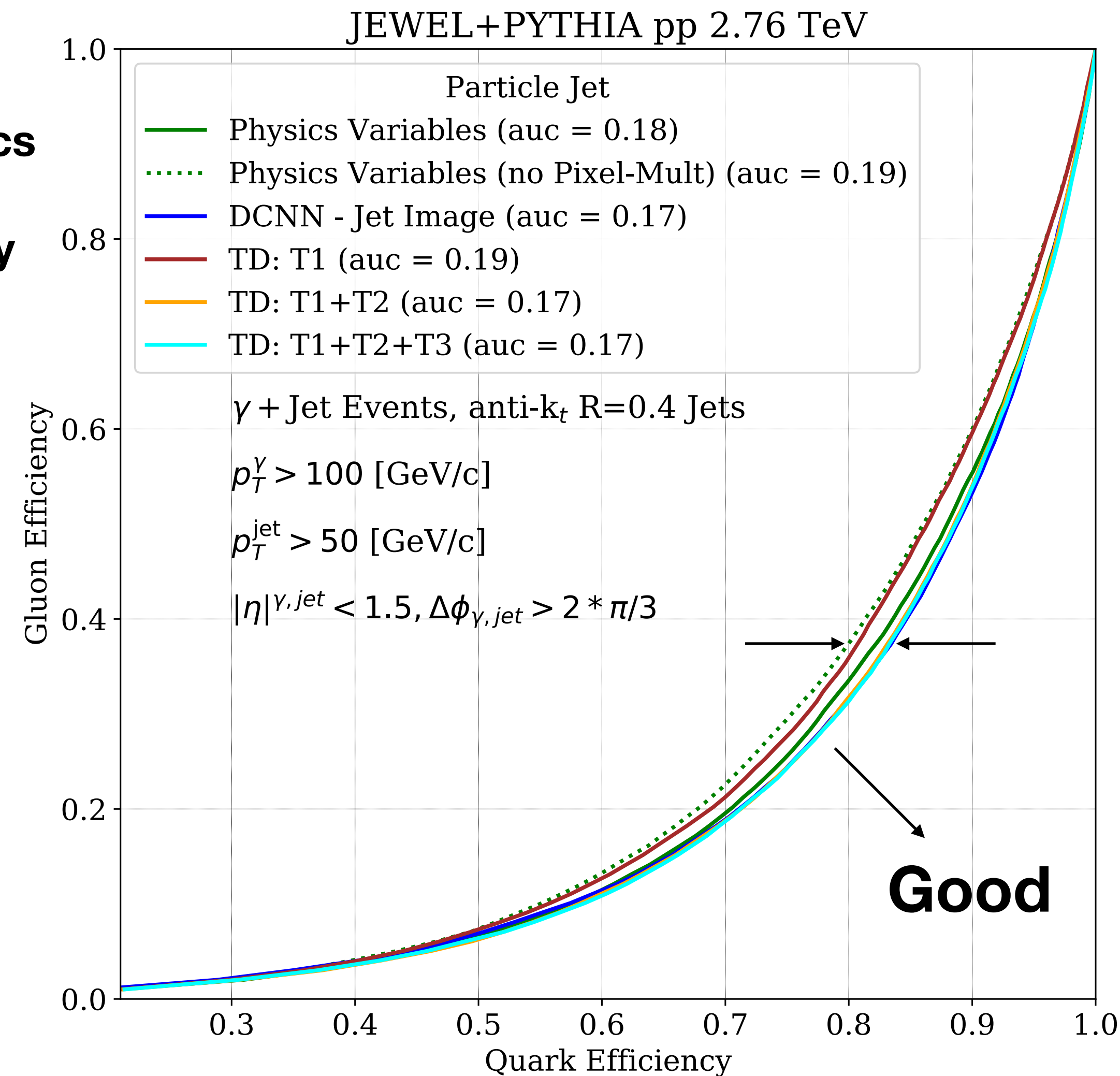
Quenched Q-Jet looks
similar to G-Jet

G-Jet Still broader

ROC curve for pp Particle Jets

**Traditional Physics
variables are
also consistently
good**

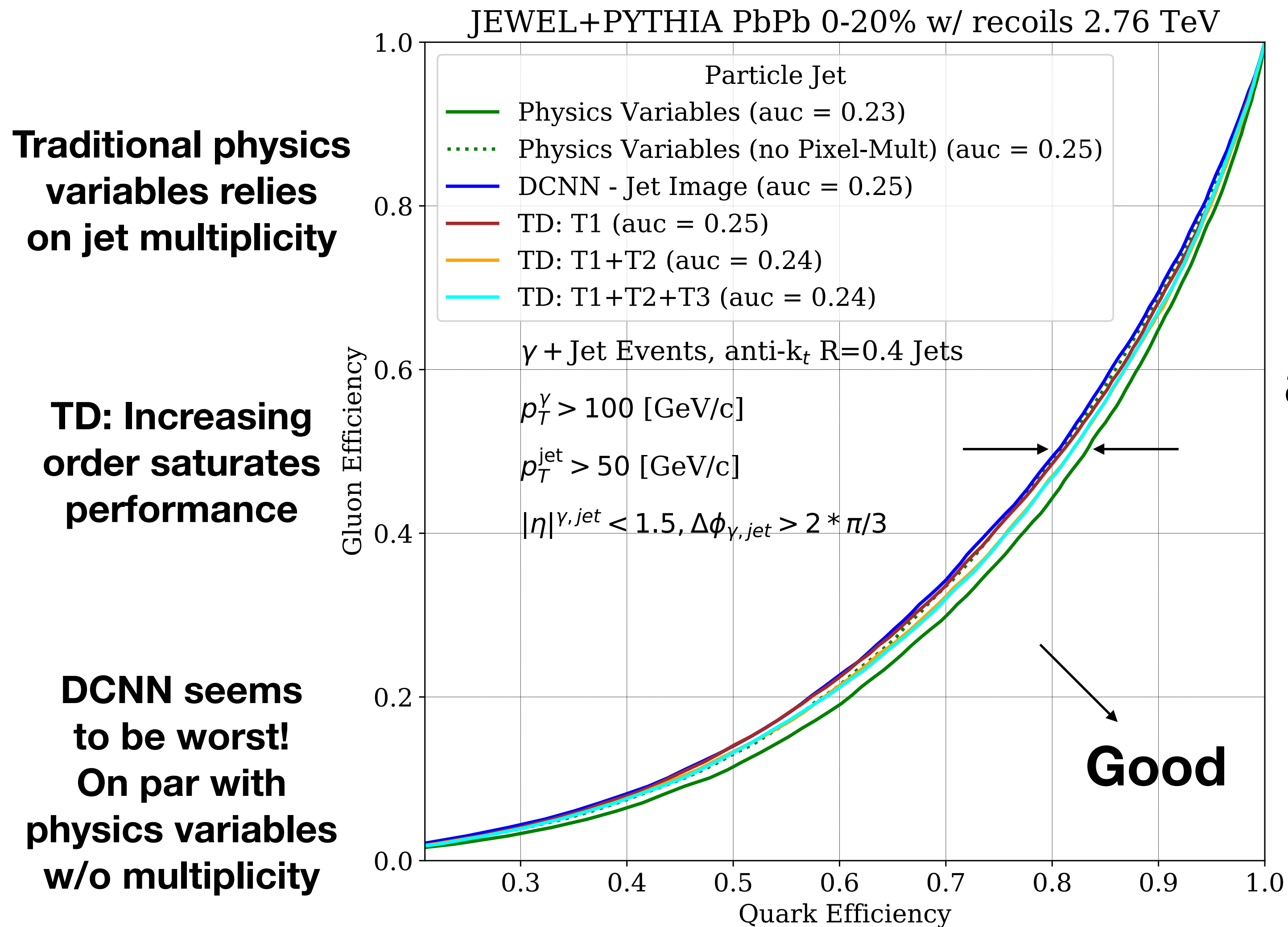
**TD: 3rd order
and DCNN
gives the best
performance!**



**All methods
are relatively
close to
each other**

How about Quenched jets?

ROC curve for Quenched PbPb Particle Jets



Traditional physics
variables relies
on jet multiplicity

TD: Increasing
order saturates
performance

DCNN seems
to be worst!
On par with
physics variables
w/o multiplicity

All methods
are relatively
close to
each other

Performance reduces!

Transformers

Provided proper attribution is provided, Google hereby grants permission to reproduce the tables and figures in this paper solely for use in journalistic or scholarly works.

Attention Is All You Need

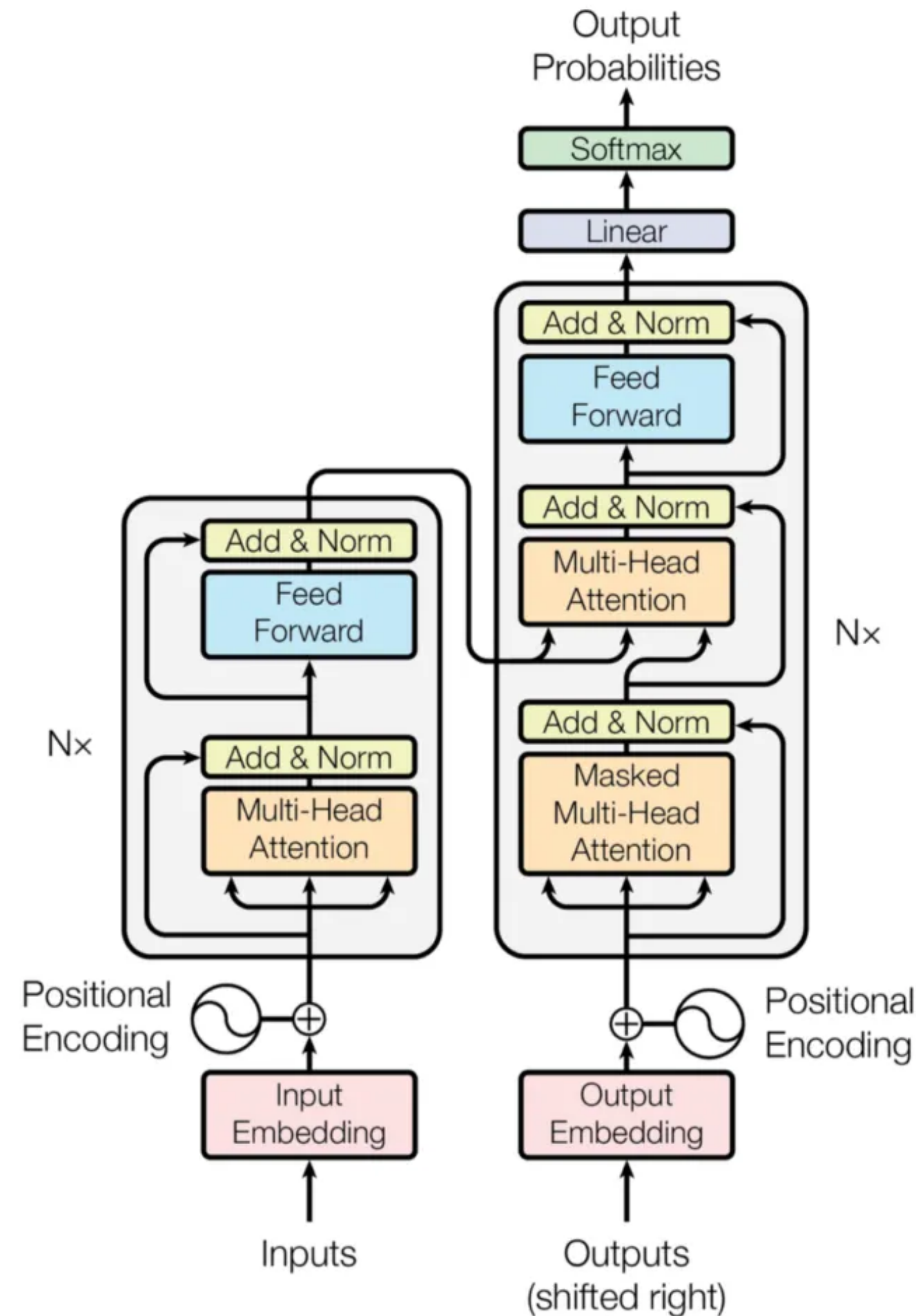


Figure 1: The Transformer - model architecture.

arXiv:1706.03762v7 [cs.CL] 2 Aug 2023

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usz@google.com

Llion Jones*
Google Research
llion@google.com

Aidan N. Gomez*[†]
University of Toronto
aidan@cs.toronto.edu

Łukasz Kaiser*
Google Brain
lukaszkaiser@google.com

Illia Polosukhin*[‡]
illia.polosukhin@gmail.com

Abstract

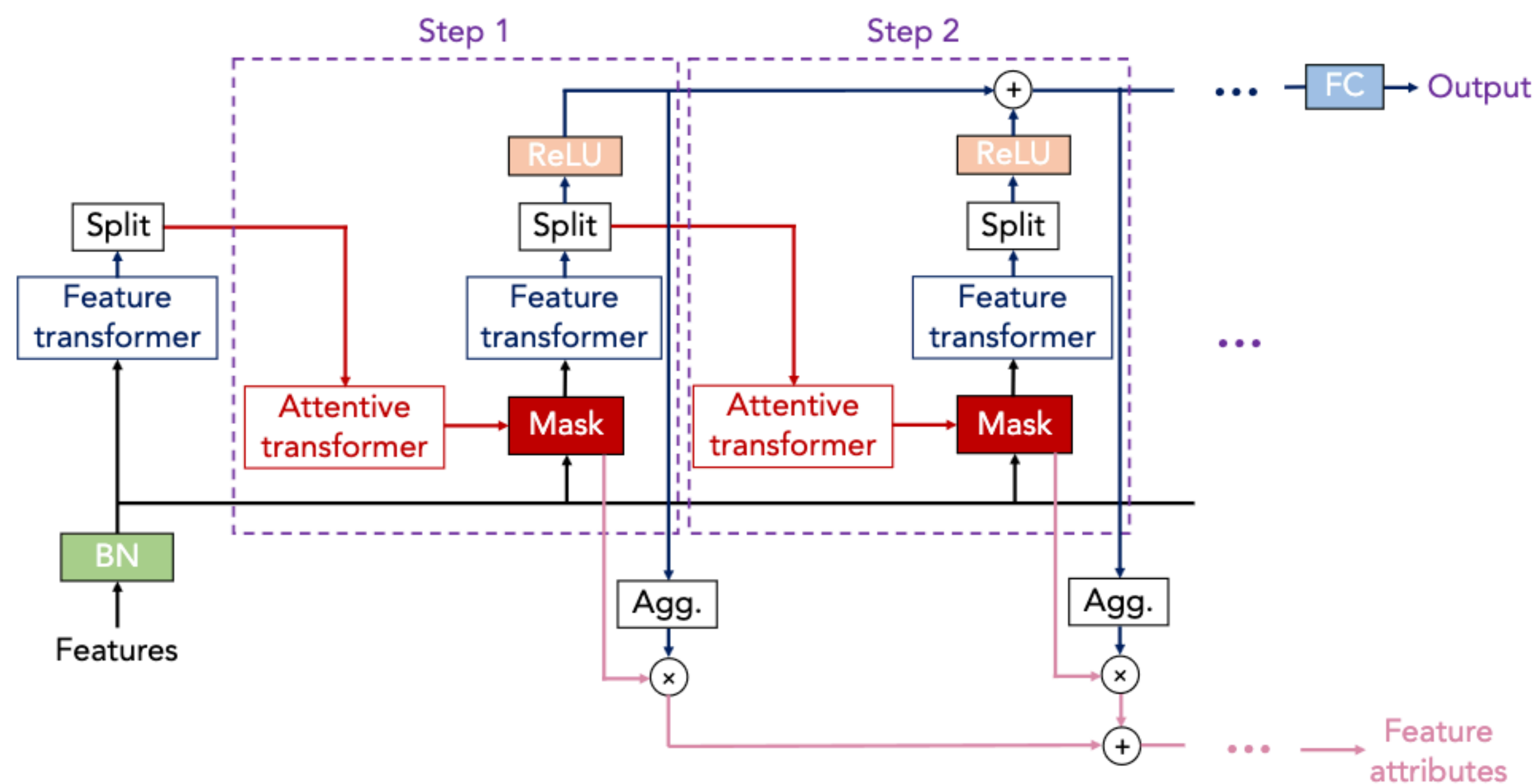
The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

Key features of transformers

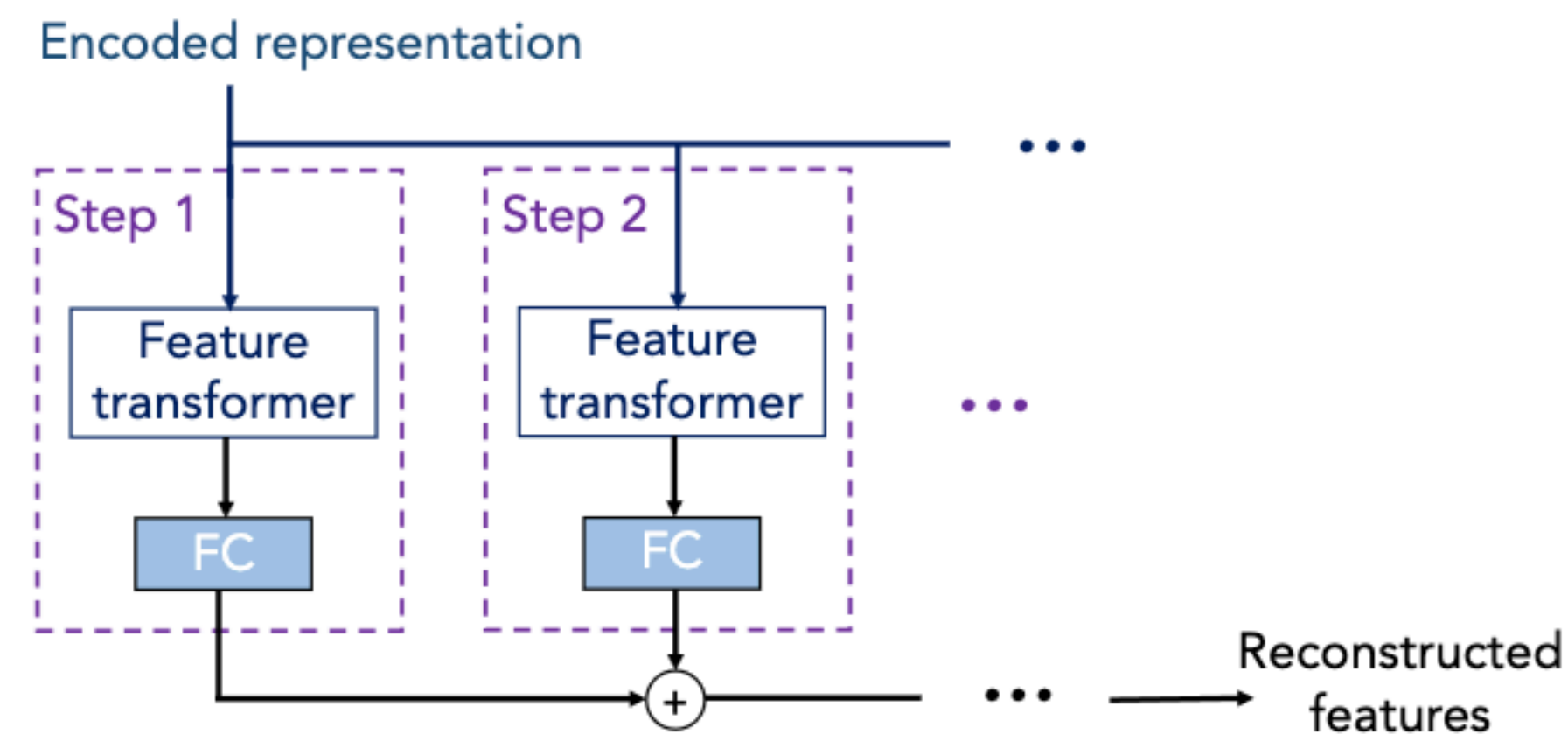
- CNNs see local patterns (nearby elements) vs Transformers see all pairs simultaneously and learn which pairs matter
- Input → LayerNorm → Multi-Head Attention → ResNet → LayerNorm → FFN → ResNet → Output
- Implements both ResNets vs FeedForward Networks
- Applicable when individual features spaces is aggregated across many events in high energy physics
- The key idea is to head to a latent space as an intermediary

<https://medium.com/h7w/an-introduction-to-transformers-in-machine-learning-50c8a53af576>

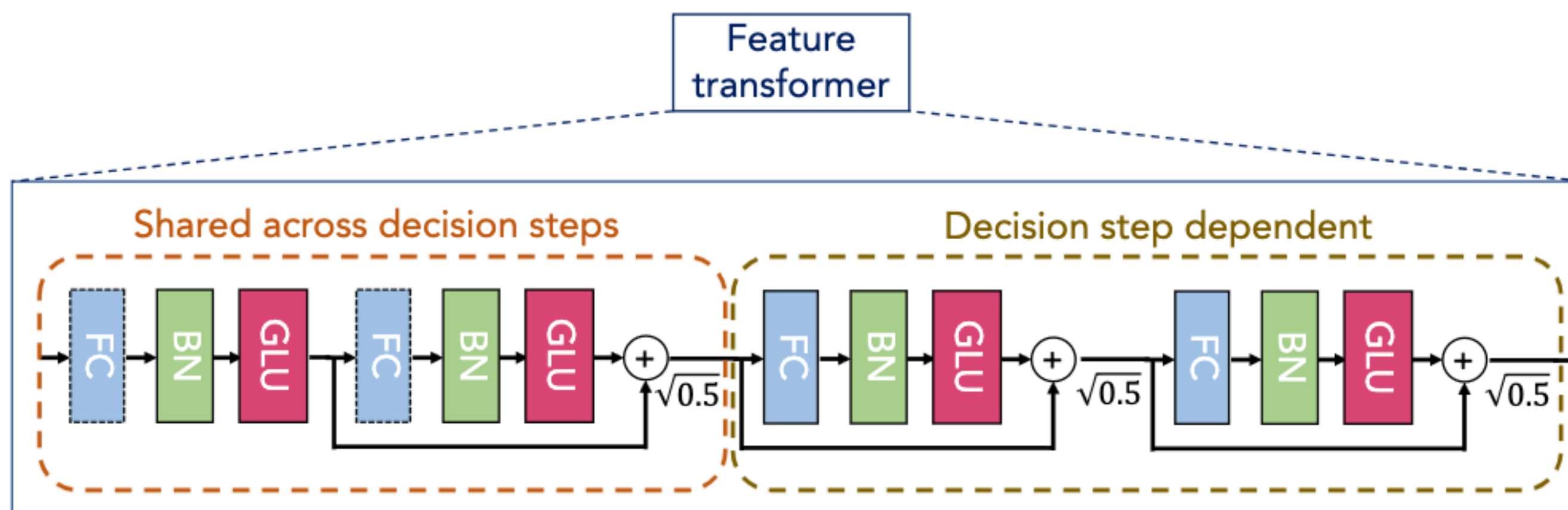
<https://arxiv.org/pdf/2304.10557>



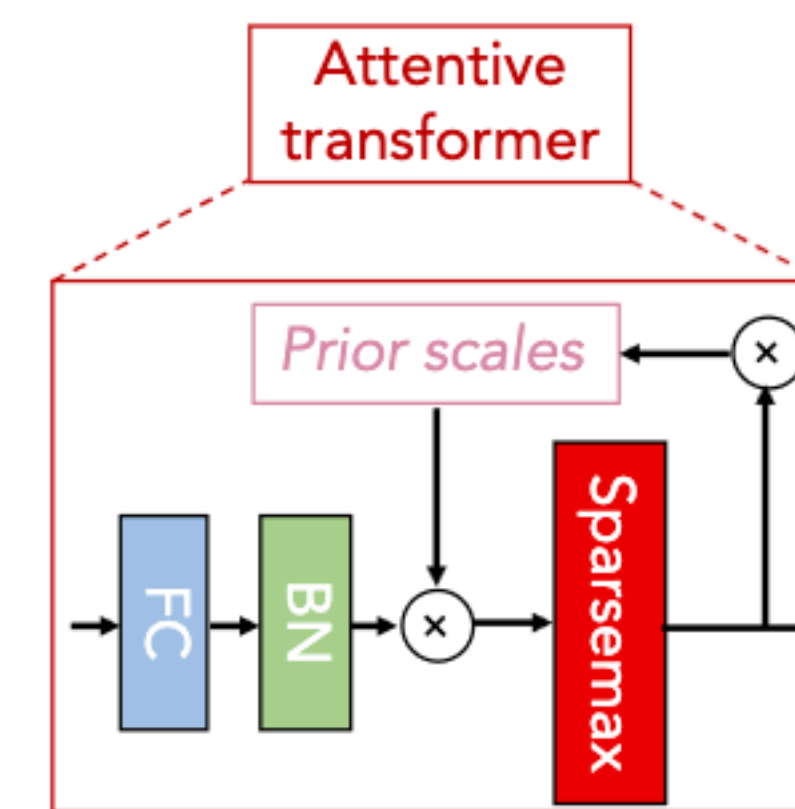
(a) TabNet encoder architecture



(b) TabNet decoder architecture



(c)



(d)

The basics

What is machine learning?

Why are these tools useful in high energy colliders?

How to quantify performance?

Physics with ML

Classifier - Can select Heavy-Flavor or Quark vs Gluons

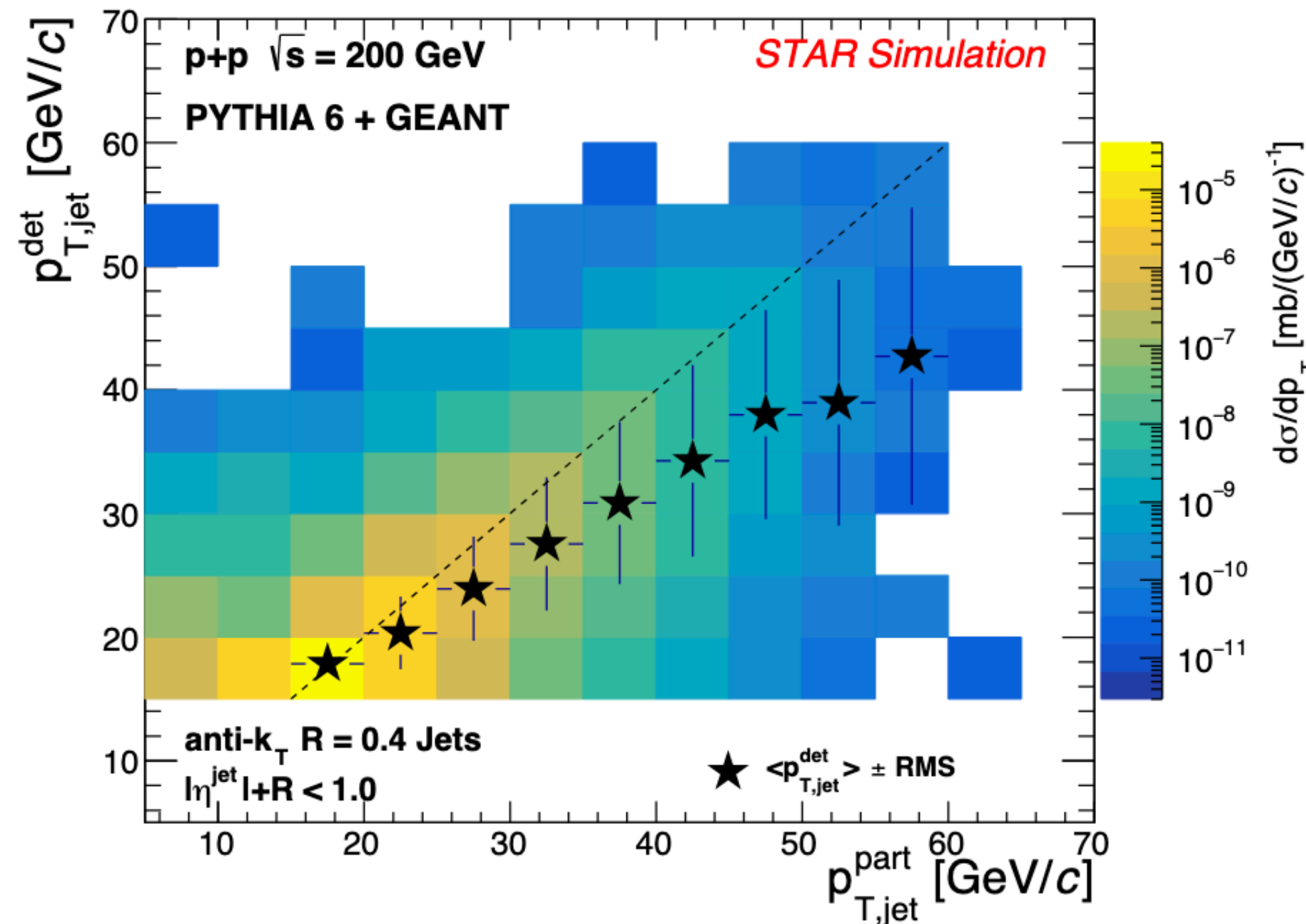
Regressor - multi-dimensional correction and unfolding

Generator - learn underlying physics of MC generators

Unfolding - a quick primer

Corrections for Detector Resolution

Response Matrix



For a given generator jet p_T - the probability get reconstructed at a certain p_T

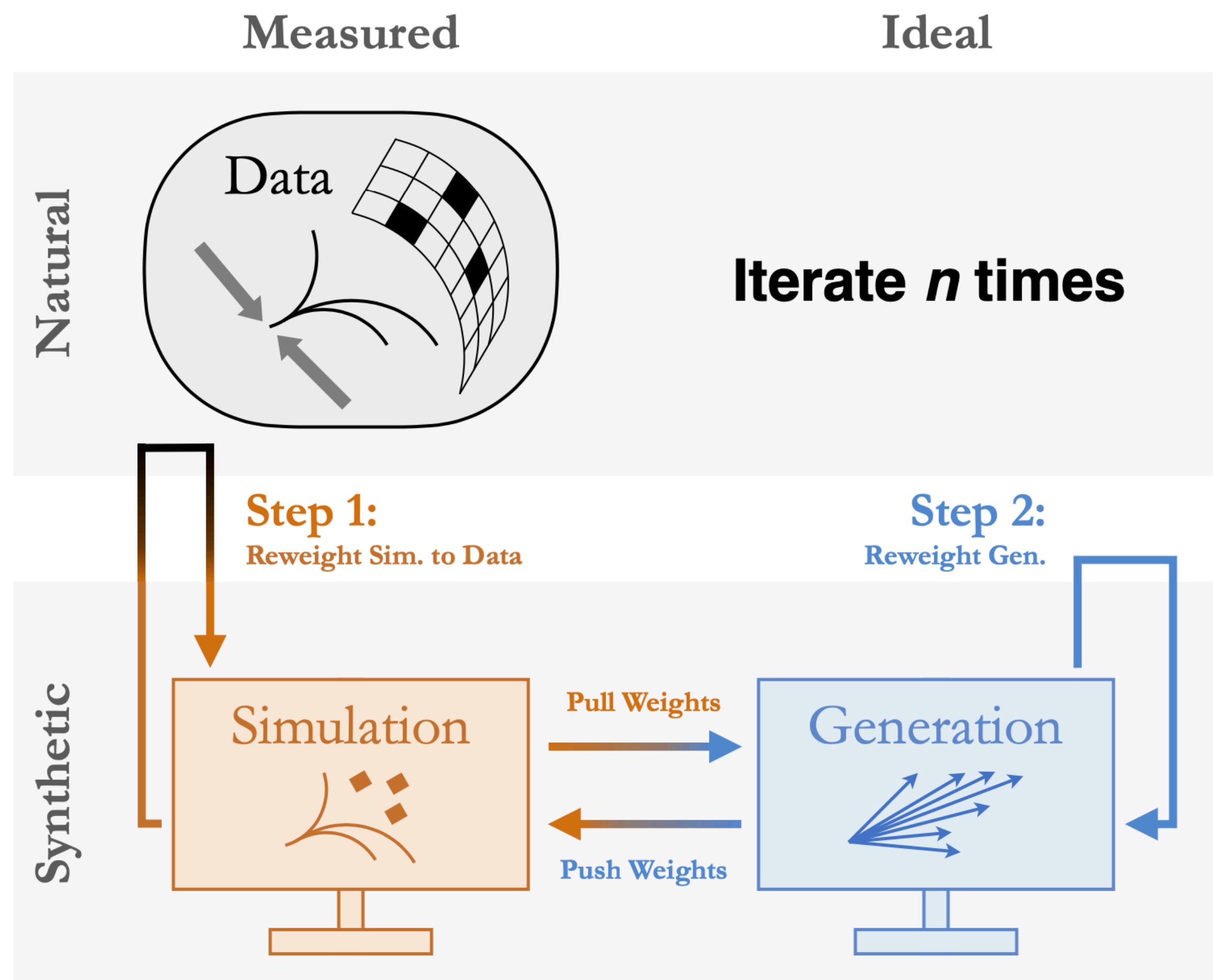
Two separate methods

- Bayesian
- Single Value Decomposition

Based on RooUnfold Package

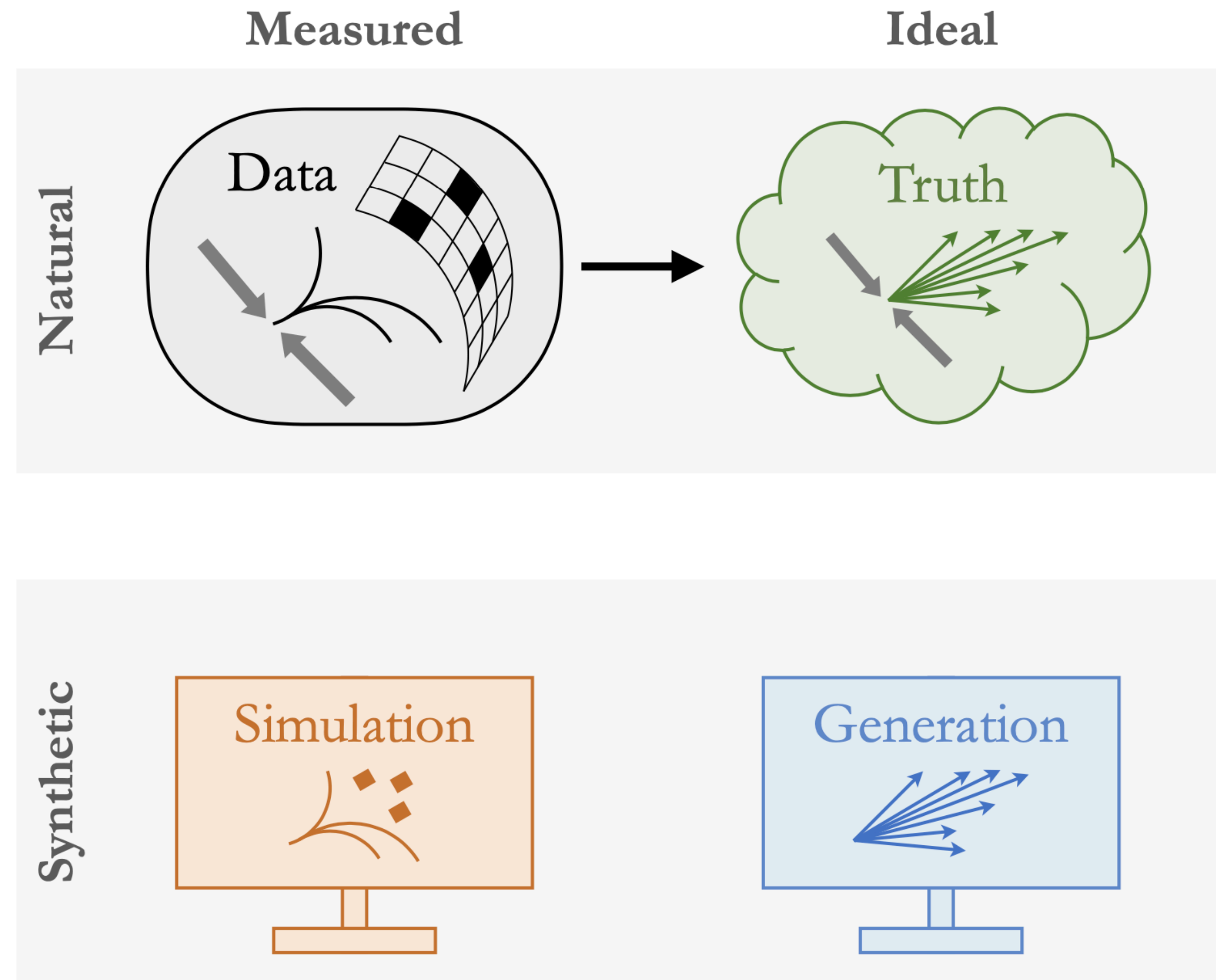
After unfolding - can directly compare with theory calculations

MultiFold (Omnifold)



Ben Nachman (LBL)

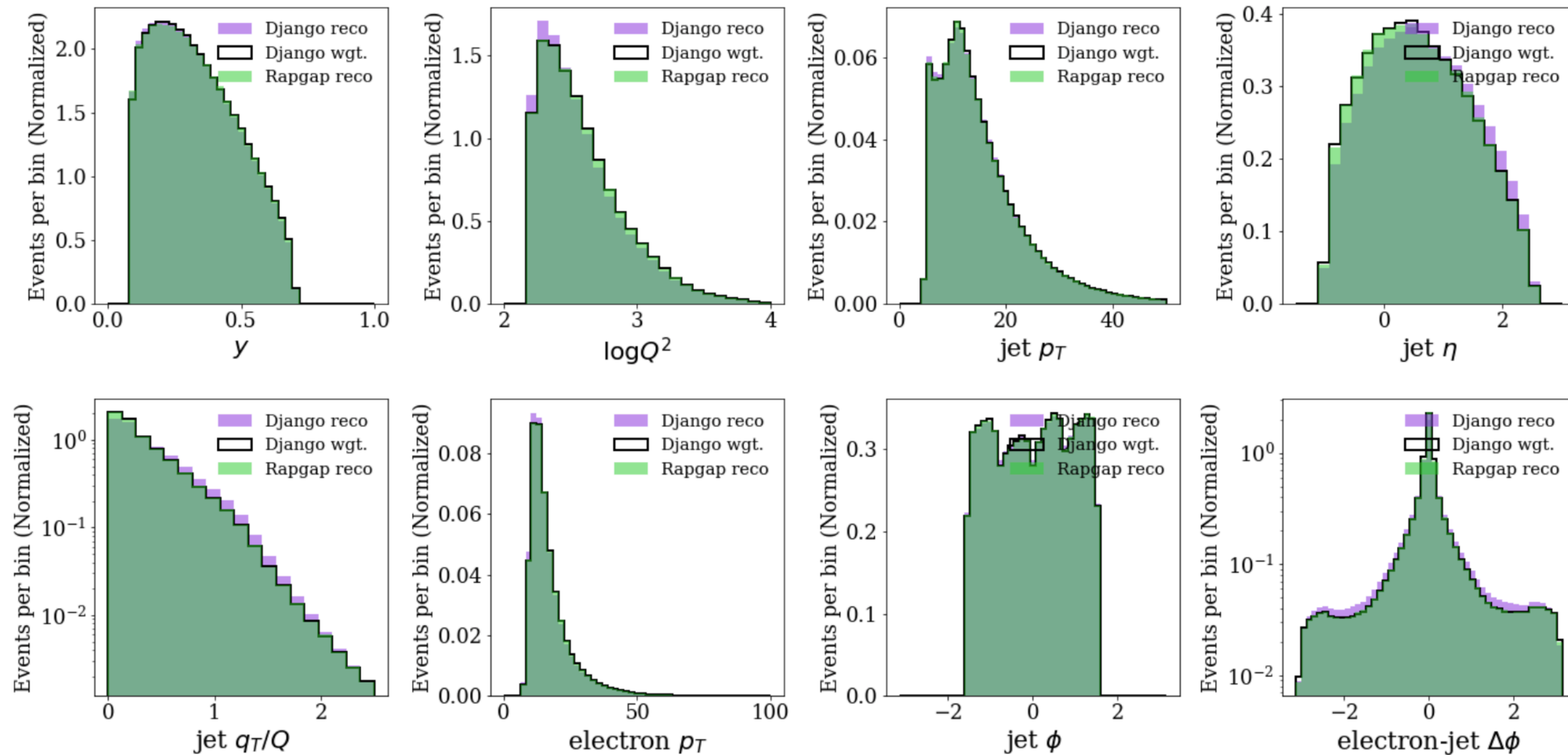
MultiFold (Omnifold)



Ben Nachman (LBL)

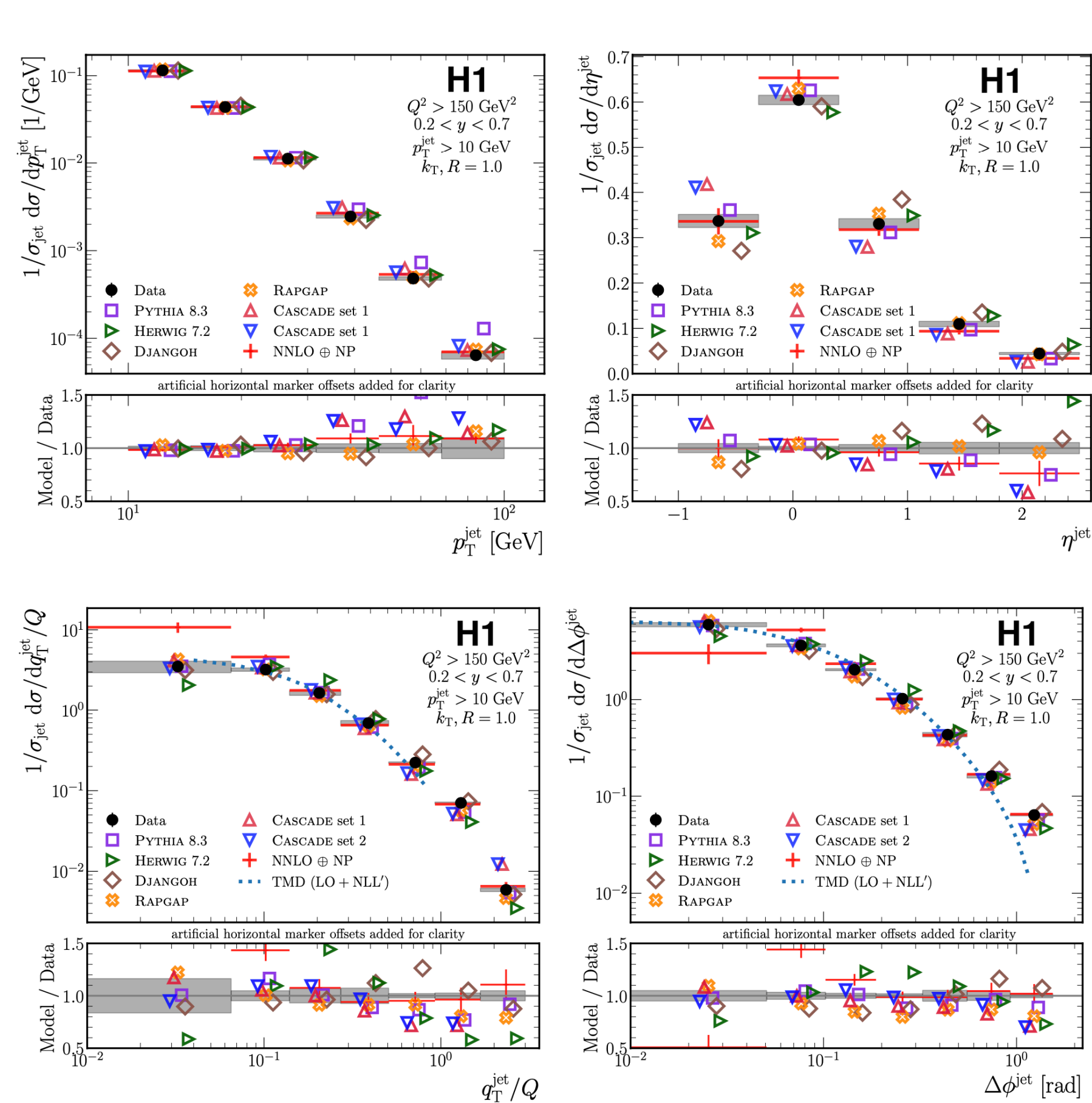
Unfolding closure tests using two different MC samples

All of these distributions are simultaneously reweighted!

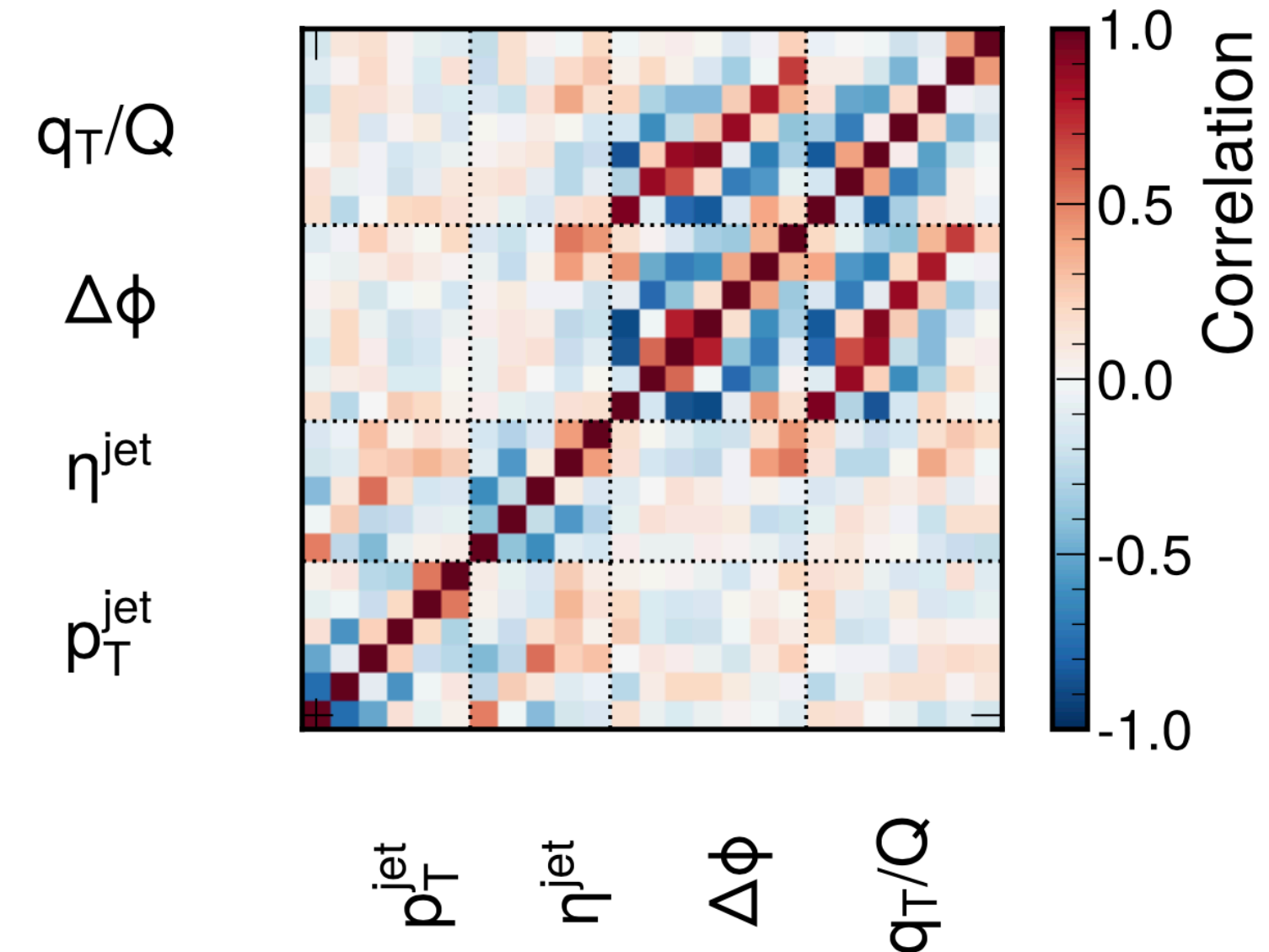


Ben Nachman (LBL)

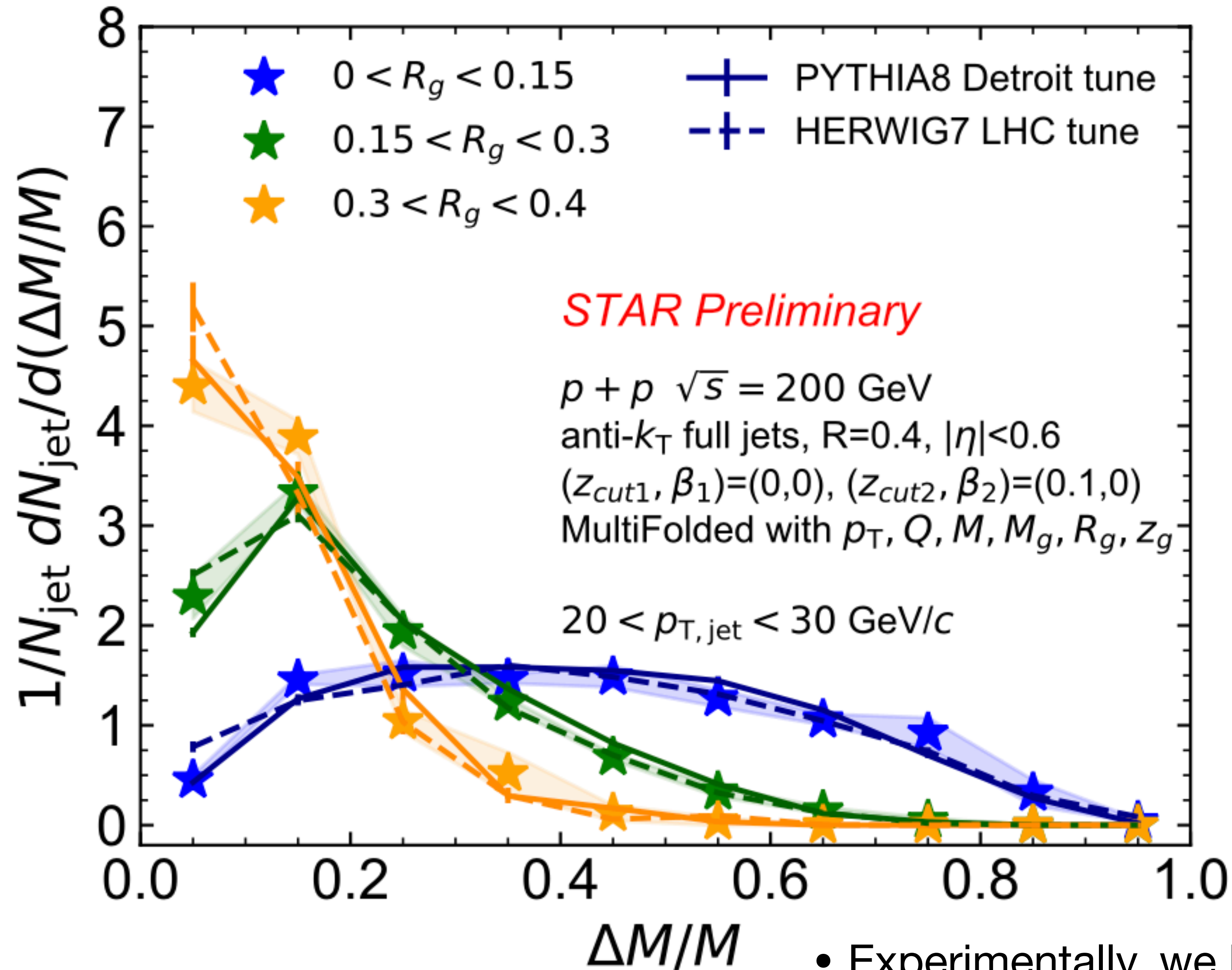
What you get at the end?



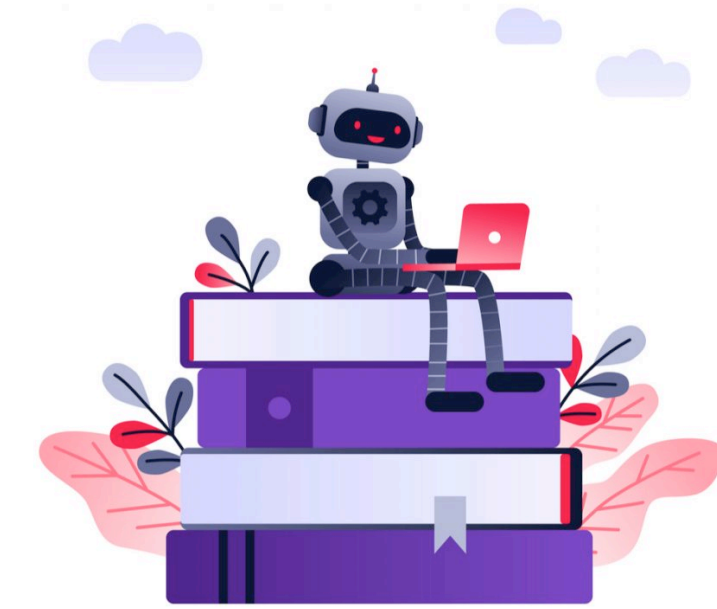
Ben Nachman (LBL)



- Multi-dimensional measurements ‘un-binned’
- You get the correlations for ‘free’



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

The basics

What is machine learning?

Why are these tools useful in high energy colliders?

How to quantify performance?

Physics with ML

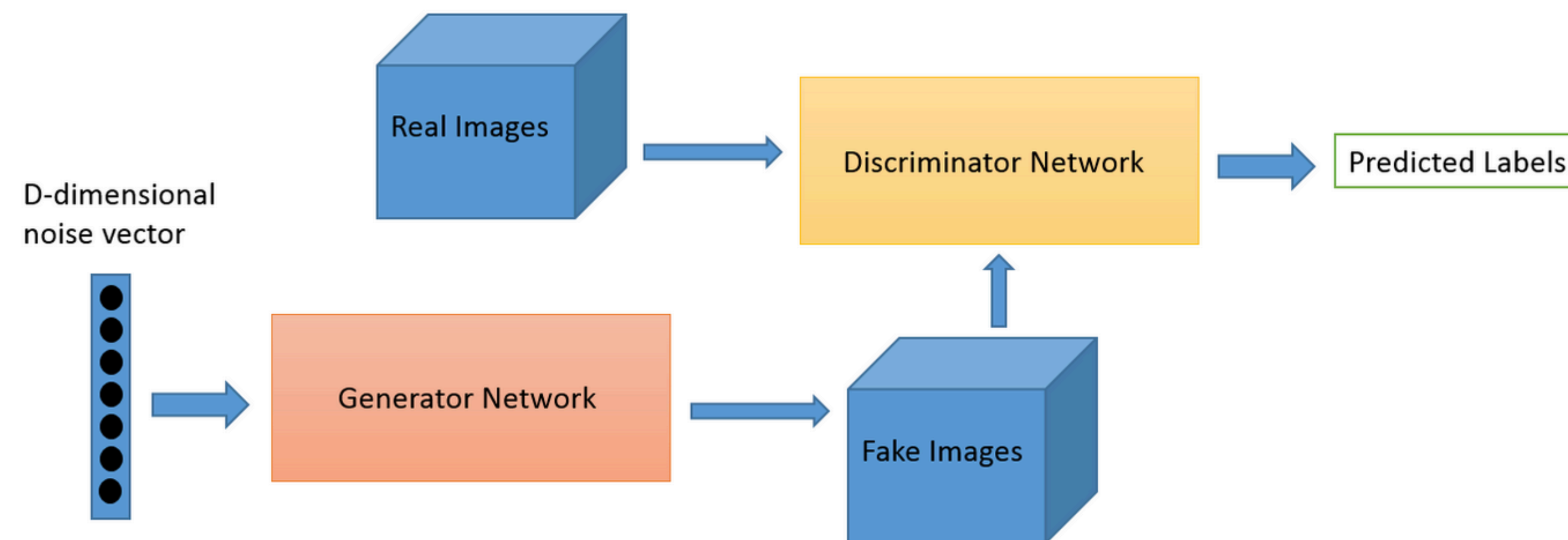
Classifier - Can select Heavy-Flavor or Quark vs Gluons

Regressor - multi-dimensional correction and unfolding

Generator - learn underlying physics of MC generators

Lets ask the AI to learn physics (or something..?)

- Given a particle-by-particle, event-by-event distribution of quantities - can a model learn the intricacies of the generation?
- Enter Generative-Adversarial-Networks (GAN) - playing one network vs another



Credit: O'Reilly

<https://skymind.ai/wiki/generative-adversarial-network-gan>

A few things GANs can do!

Generate Faces!

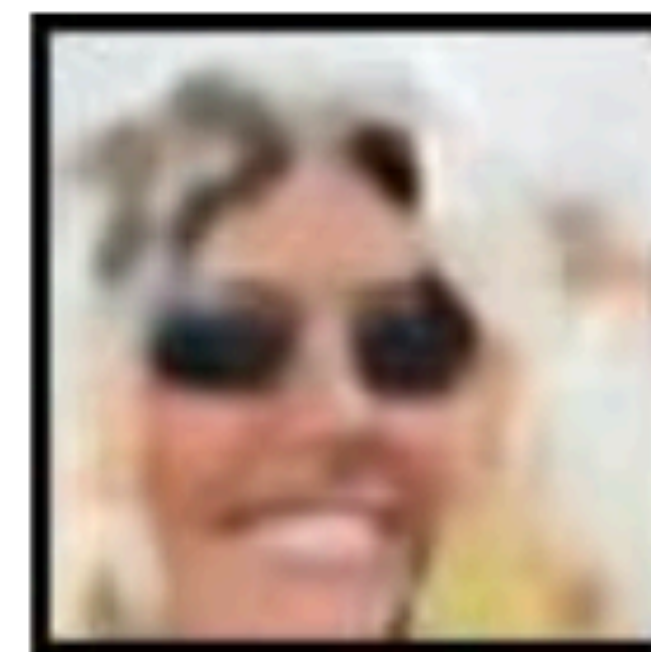
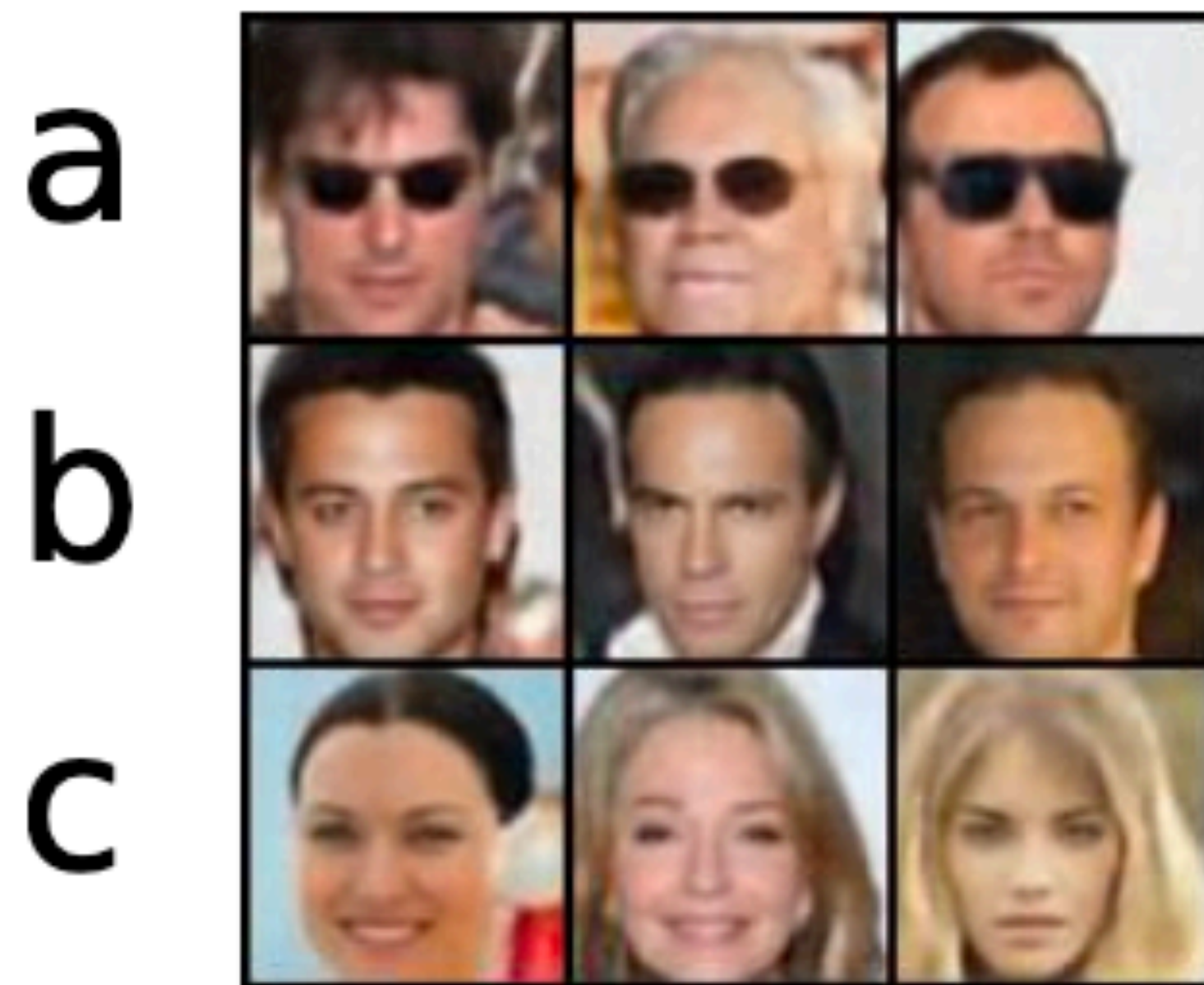


Nearest
training
set

Ian Goodfellow et. al, 1406.2661

A few things GANs can do!

latent space arithmetic : Reduce images to its inherent hidden representation (same-dimensions) so we can perform mathematical operations!

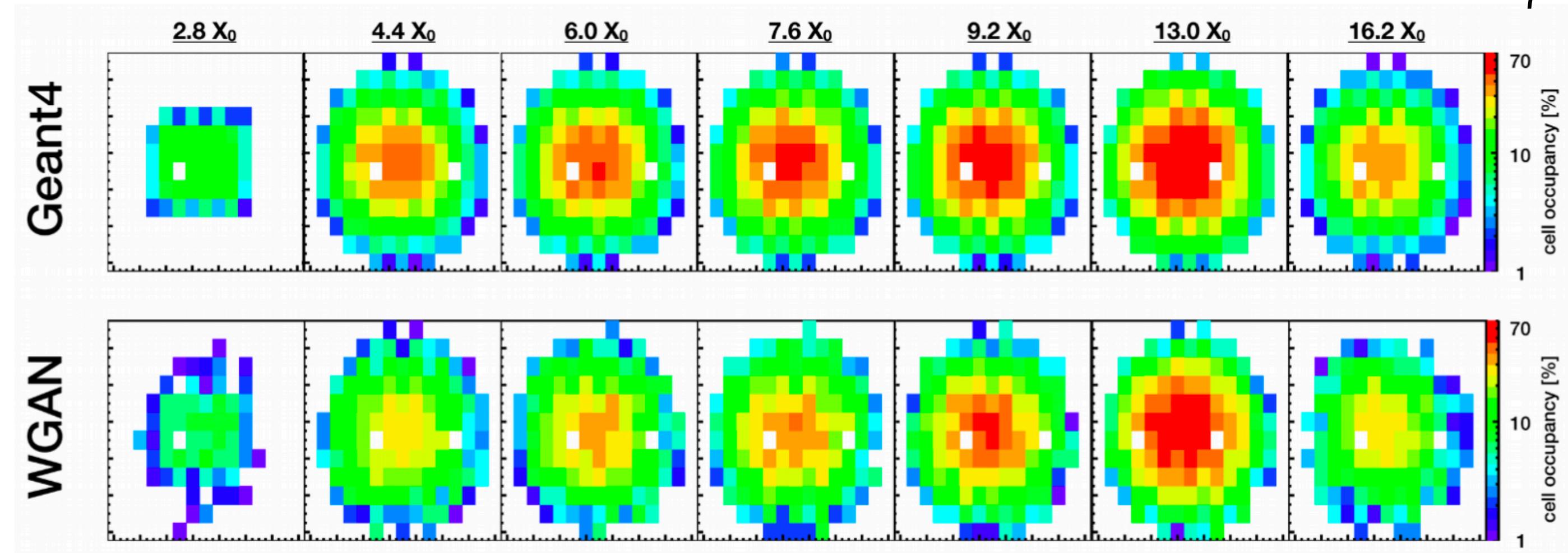


$a - b + c$

Piotr Bojanowski et. al, 1707.05776

Facebook AI

CMS
High-Granularity HF



Yannik Rath (Aachen), ML4Jets18

Average positron shower in each calorimeter layer

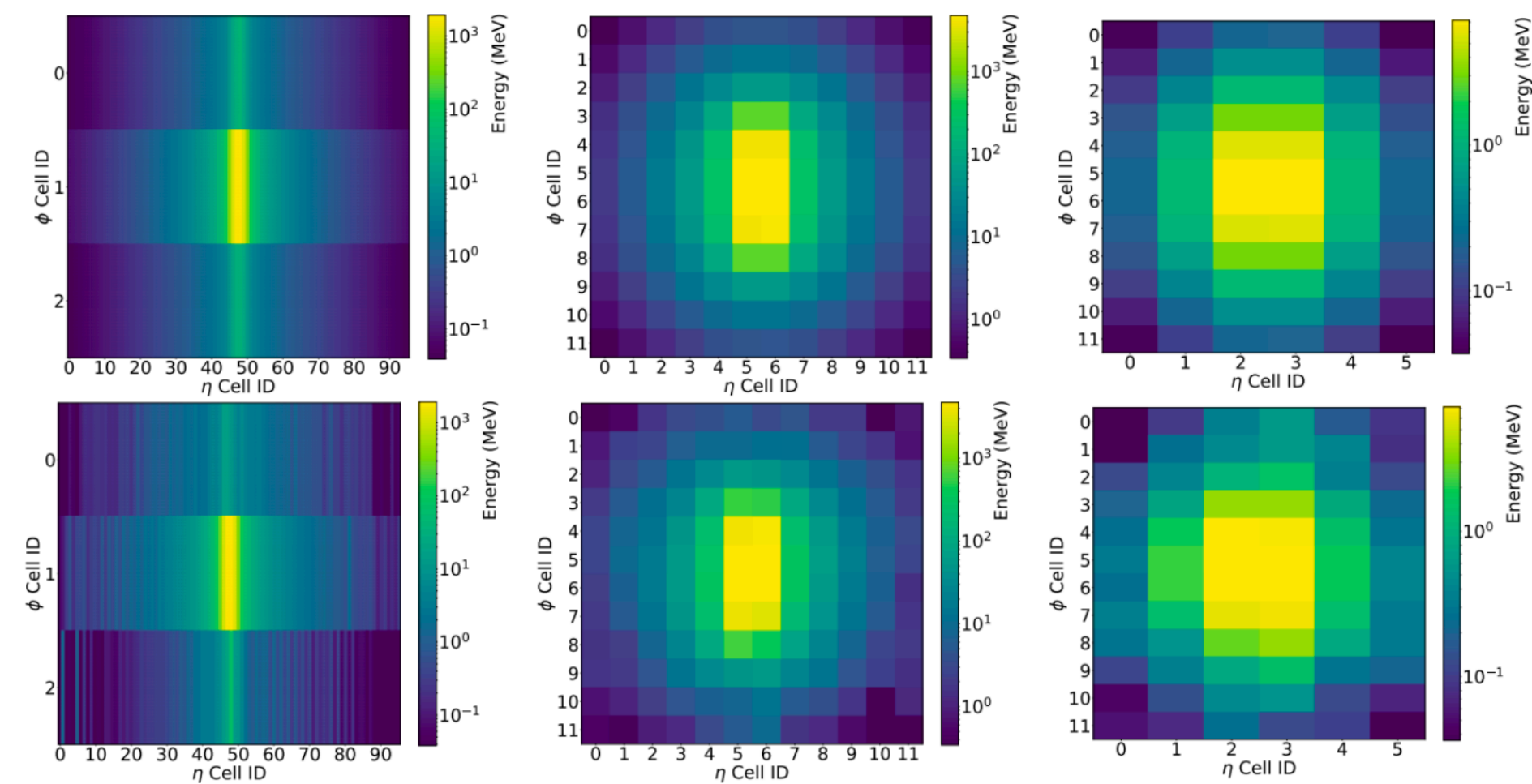
ATLAS
Longitudinally
segmented Calorimeter

Pagnini M, Nachman B, Olivera L
Phys. Rev. D 97, 014021 (2018)

Michela Pagnini (Yale, LBNL), ML4Jets17

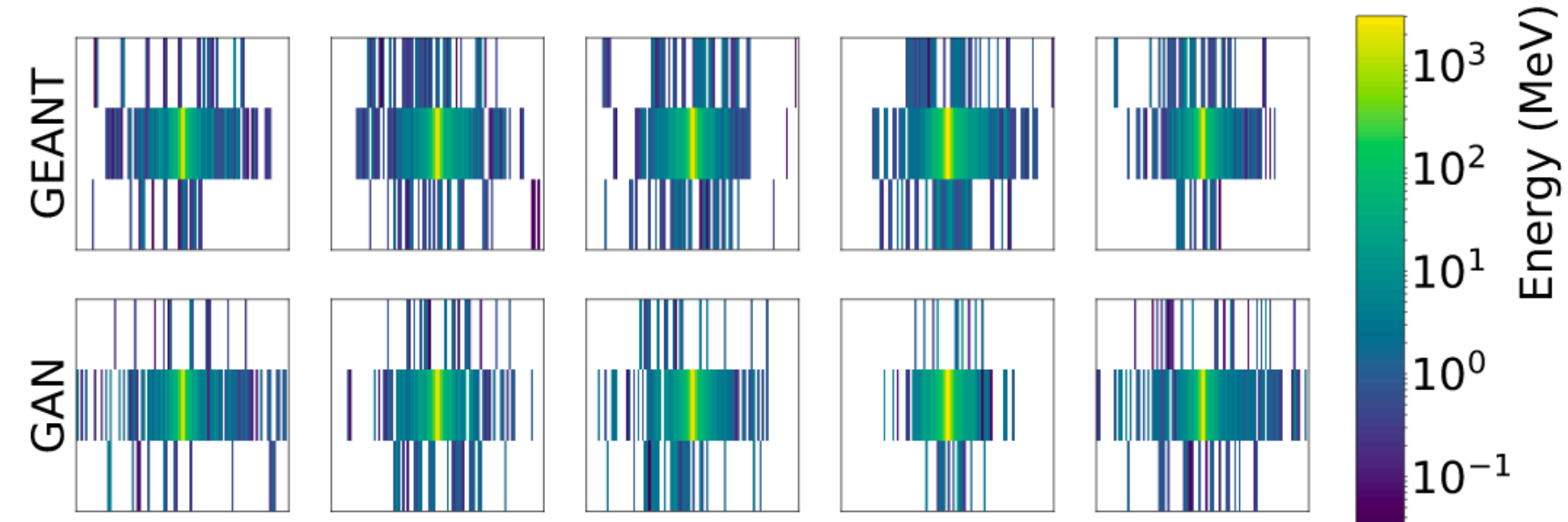
GEANT

CaloGAN



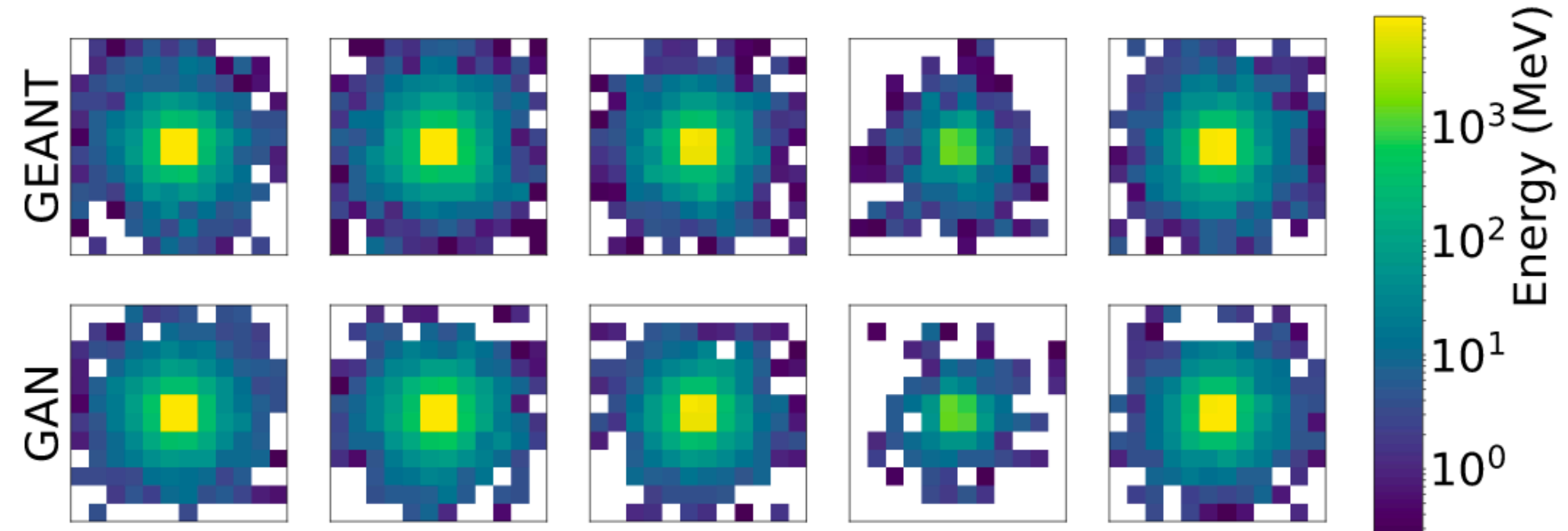
Individual positron showers and generated nearest neighbors

GEANT 1st layer deposition



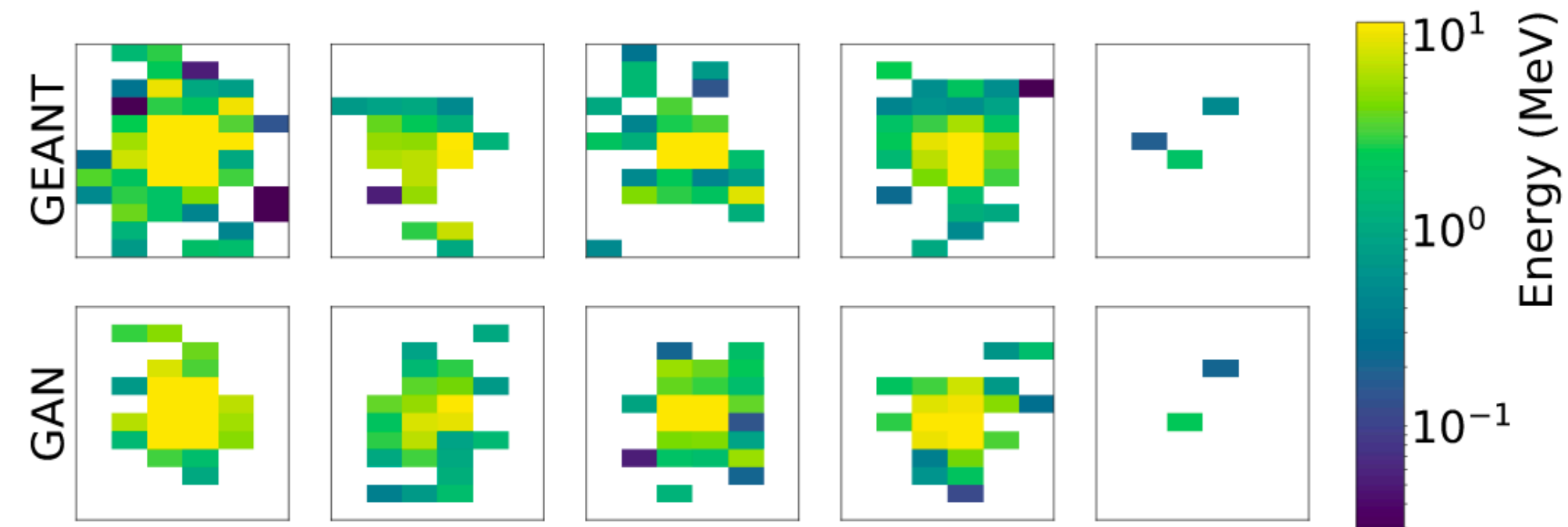
CaloGAN 1st layer deposition

GEANT 2nd layer deposition



CaloGAN 2nd layer deposition

GEANT 3rd layer deposition



CaloGAN 3rd layer deposition

Michela Pagnini (Yale, LBNL), ML4Jets17

The basics

What is machine learning?

Why are these tools useful in high energy colliders?

How to quantify performance?

Physics with ML

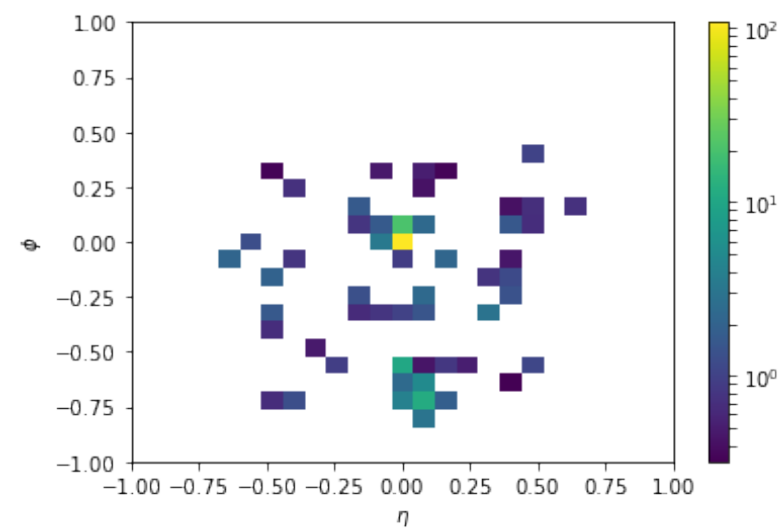
Classifier - Can select Heavy-Flavor or Quark vs Gluons

Regressor - multi-dimensional correction and unfolding

Generator - learn underlying physics of MC generators

Can we tag the flavor of the jets?

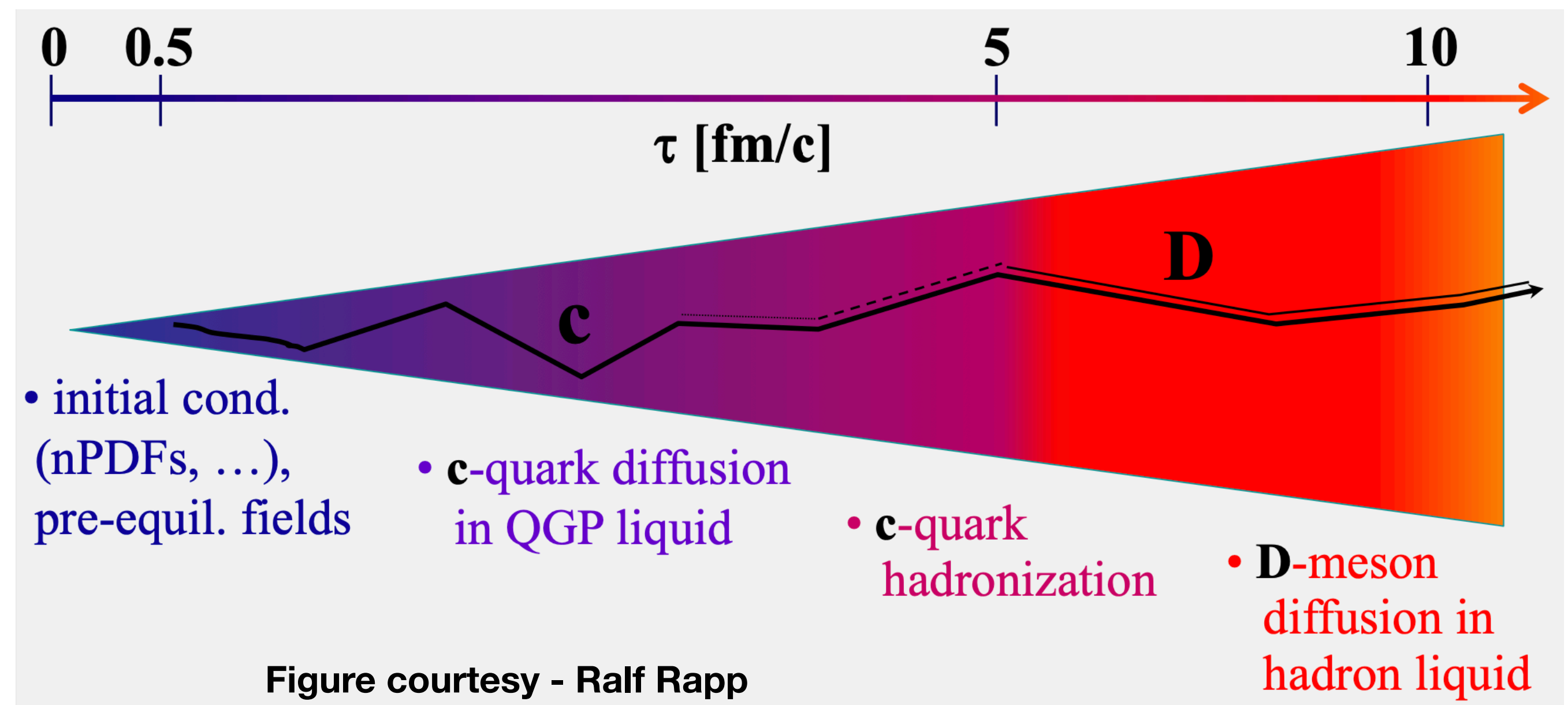
- What do we learn from this - flavor dependent fragmentation
- Proton's PDF and possibly extending all the way to GPDs
- Can we find the mother q/g?



Vertexing
DCA_{XY}, DCA_Z

Tracking
 p_T , η , ϕ

Fragmentation
 z , ΔR , $z\Delta R^2$



NetVLAD

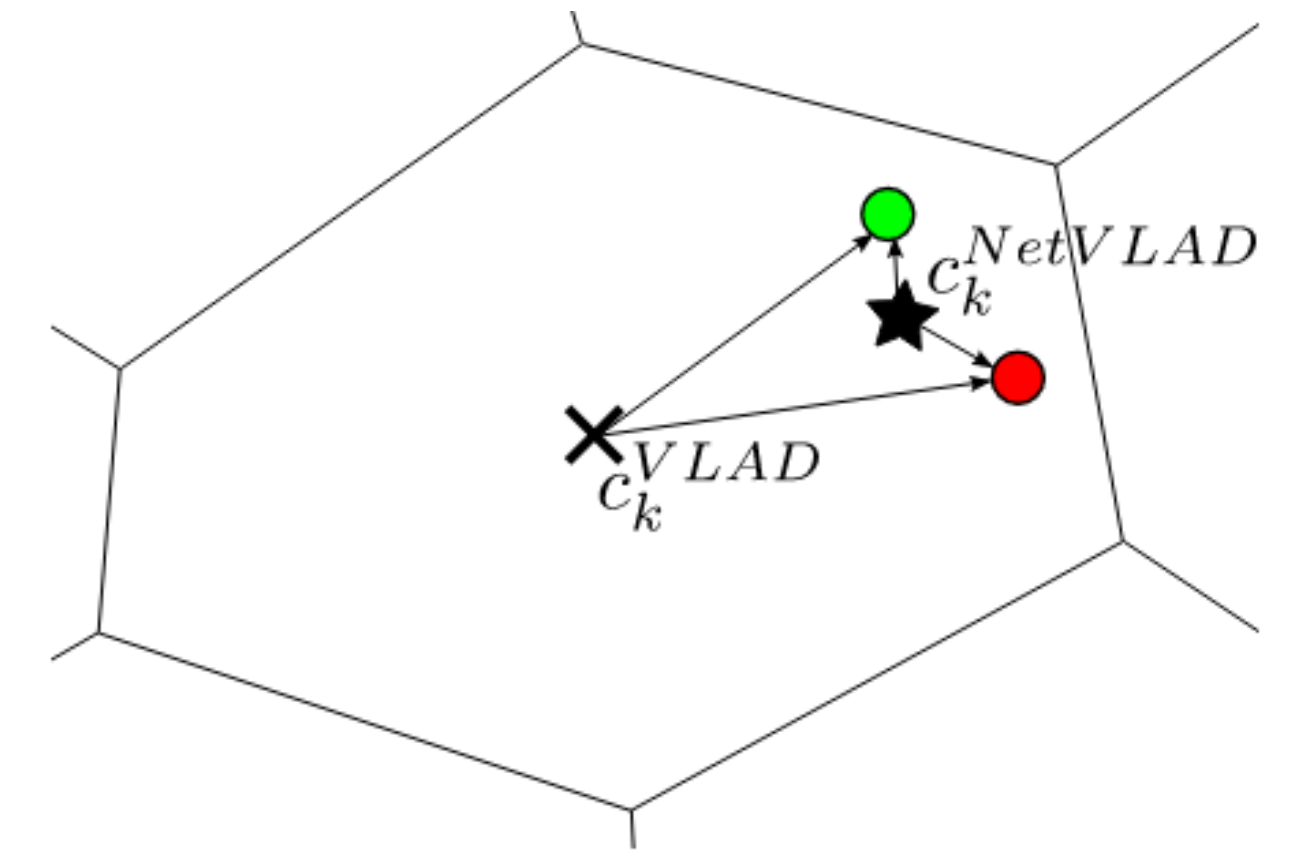
Arandjelović et. al 1511.07247



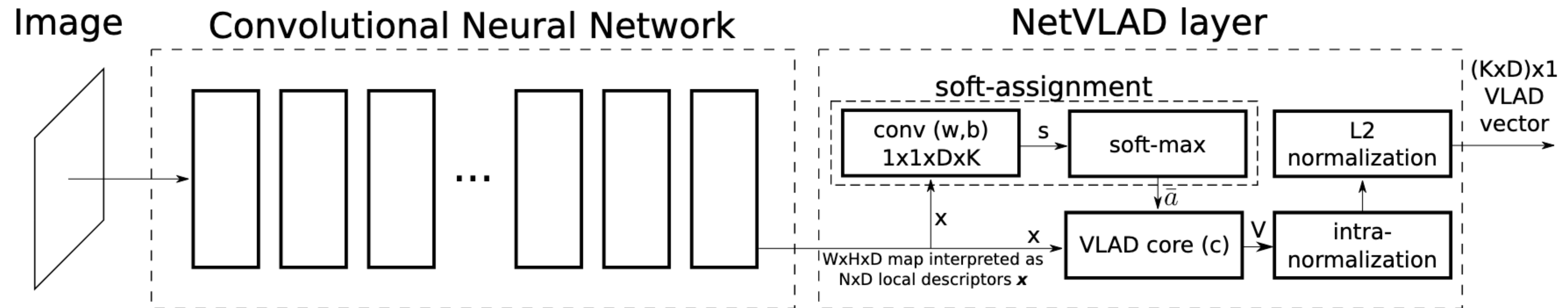
(a) Mobile phone query



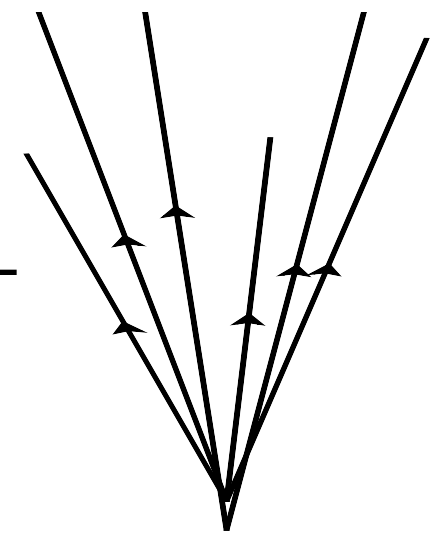
(b) Retrieved image of same place



- CNN architecture for weakly supervised place recognition



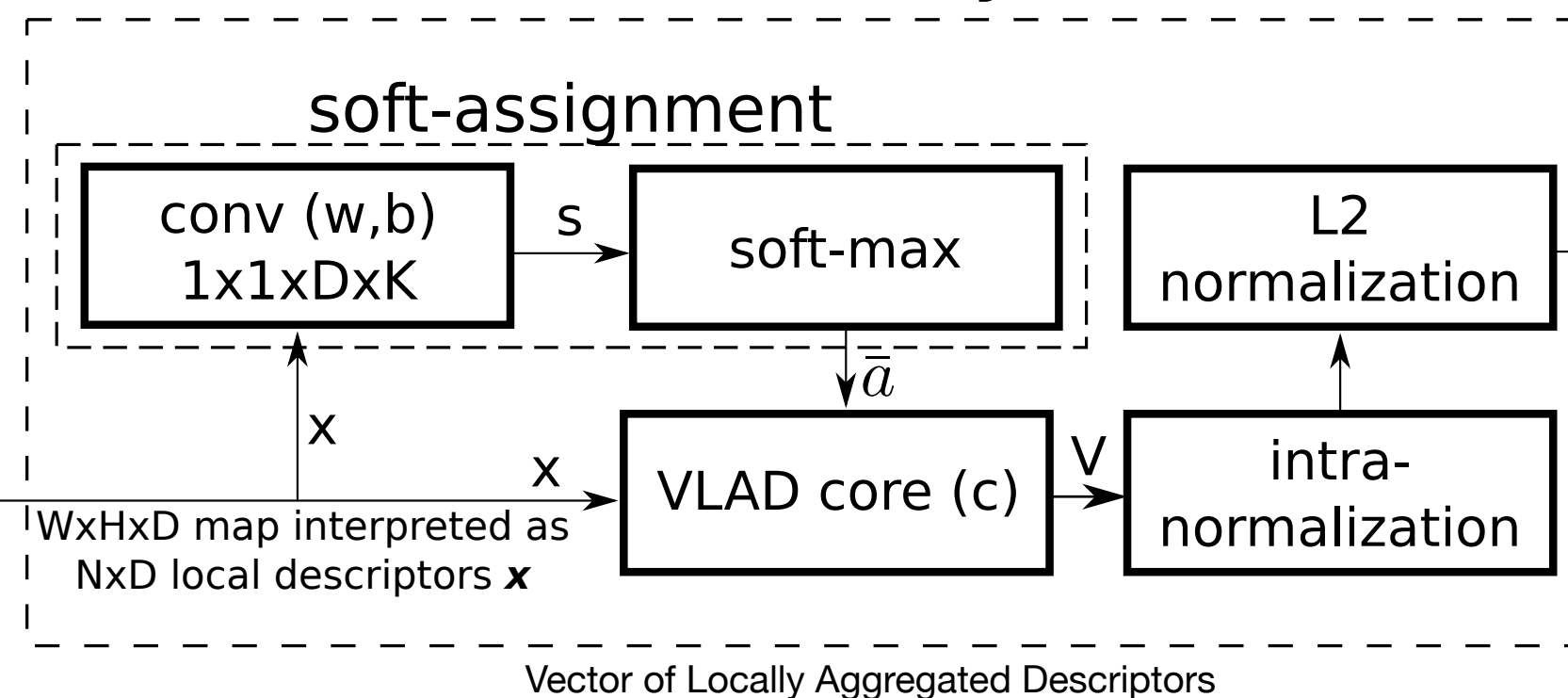
Tagging Heavy-Flavor JetVLAD



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

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

NetVLAD layer



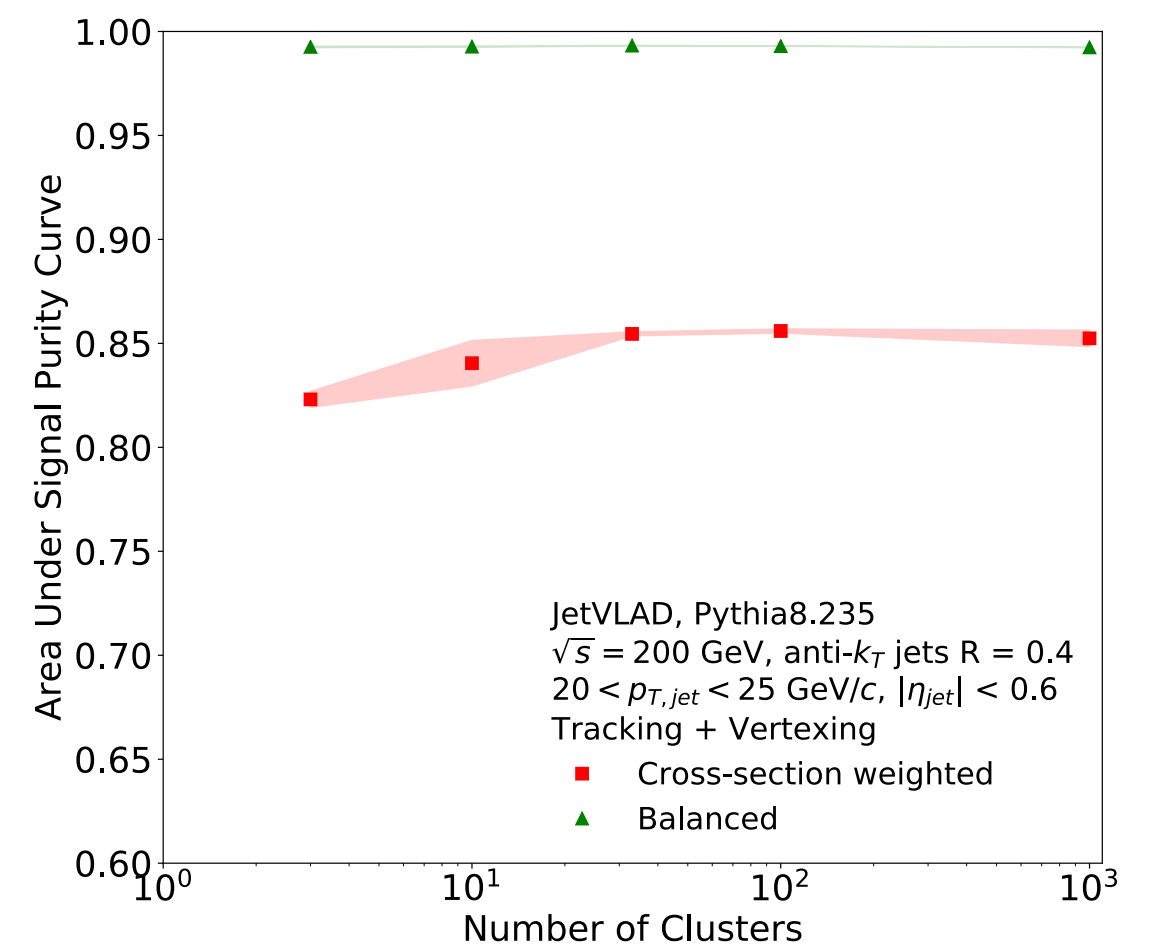
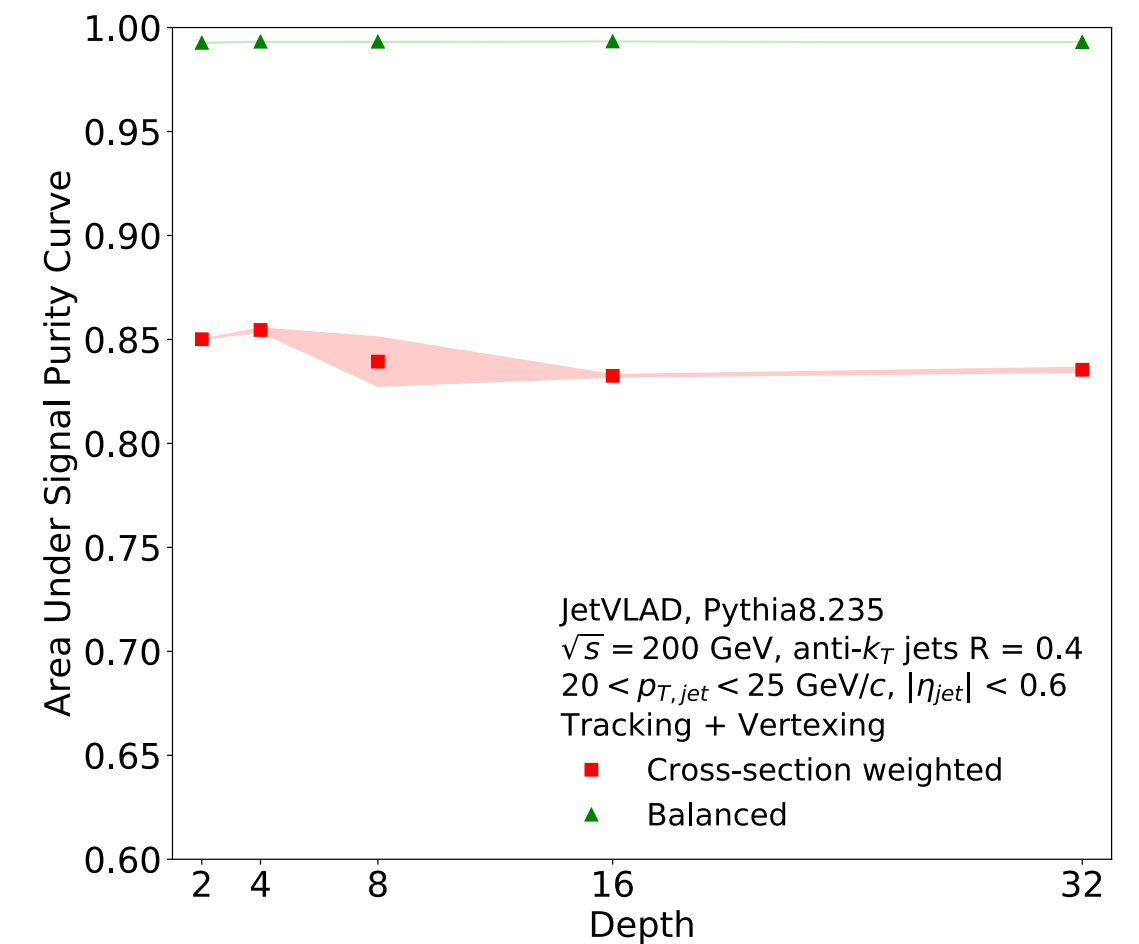
Vector of Locally Aggregated Descriptors



Jitka Mrazkova
PhD student @ NPI



Georgy Ponimatkin
PhD. In ML @
Ecole des Ponts
ParisTech



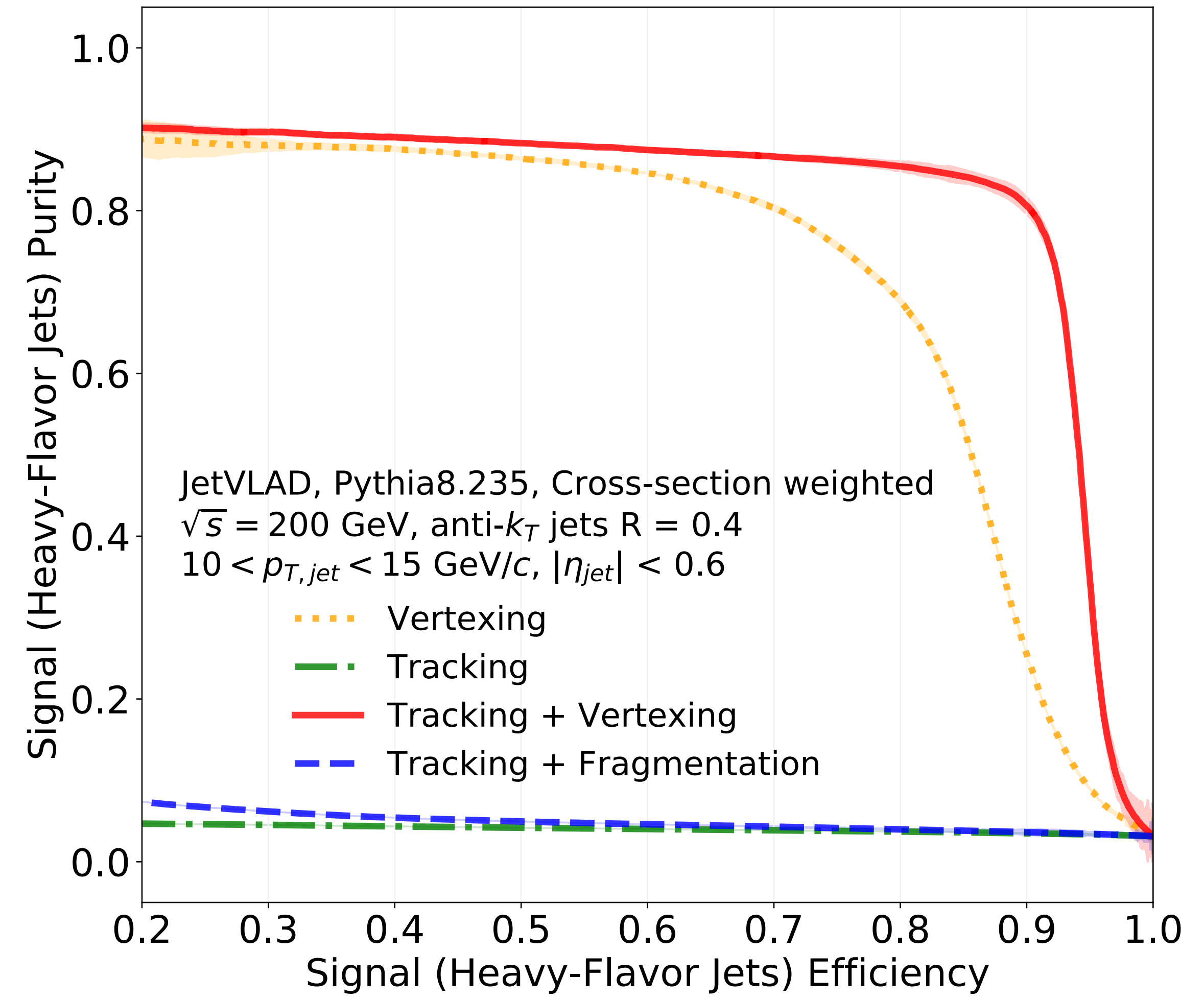
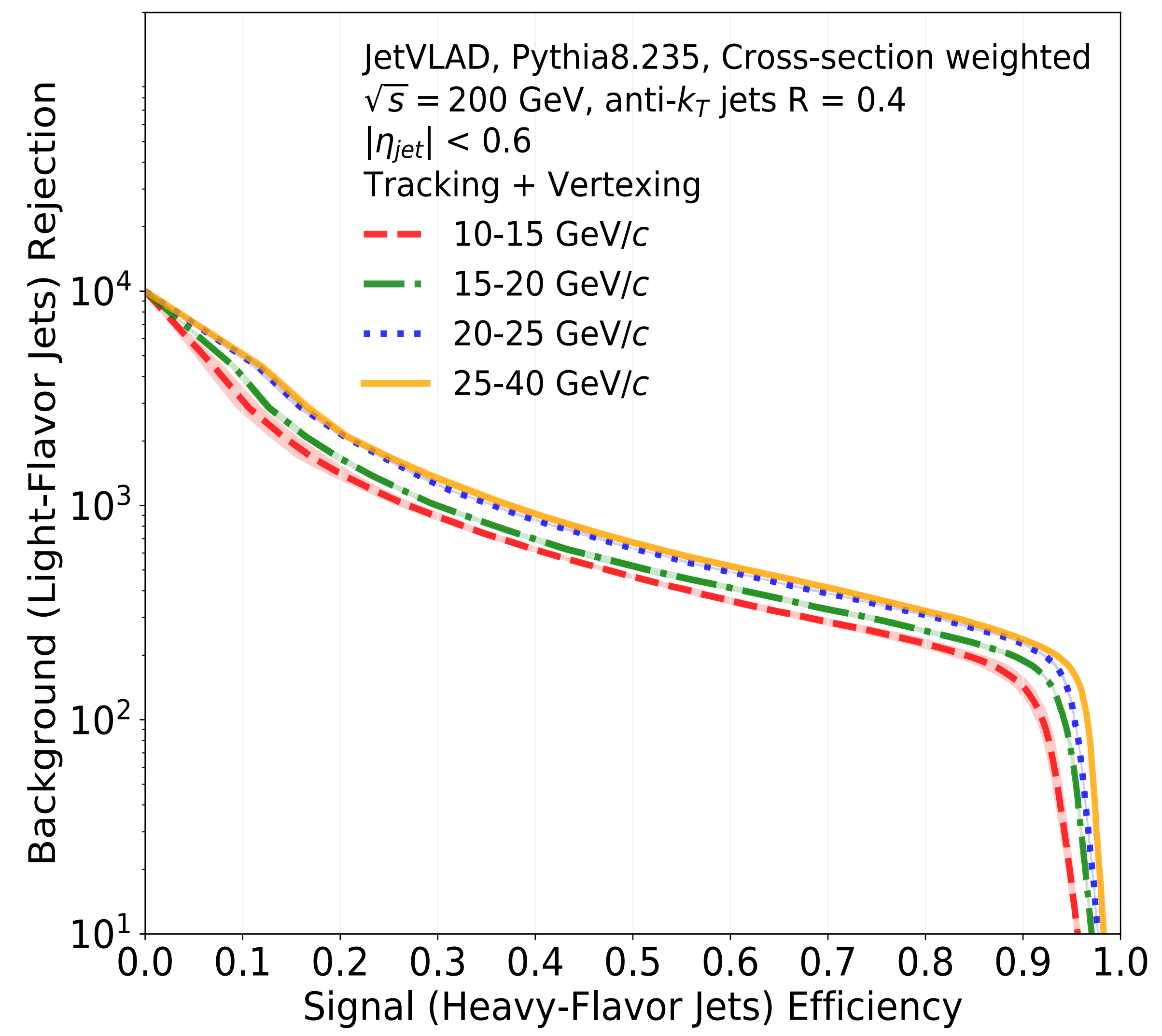
D - Depth
K - # Clusters

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

Ponimatkin, et. al JINST 2005.01842

Total of 111608 trainable parameters

Performance benchmarks

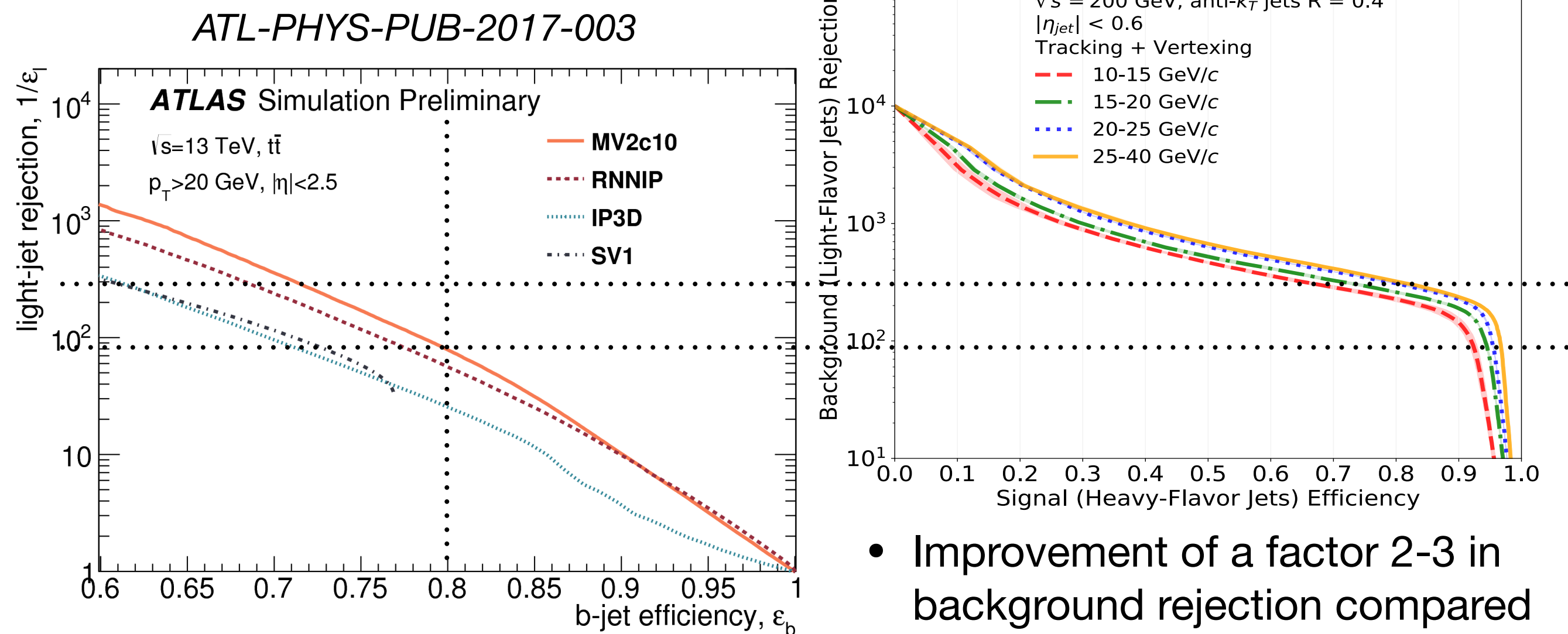


- Improvement (factor of 2) attributed to algorithmic differences primarily in comparison to RNN (which are quite hard to train)

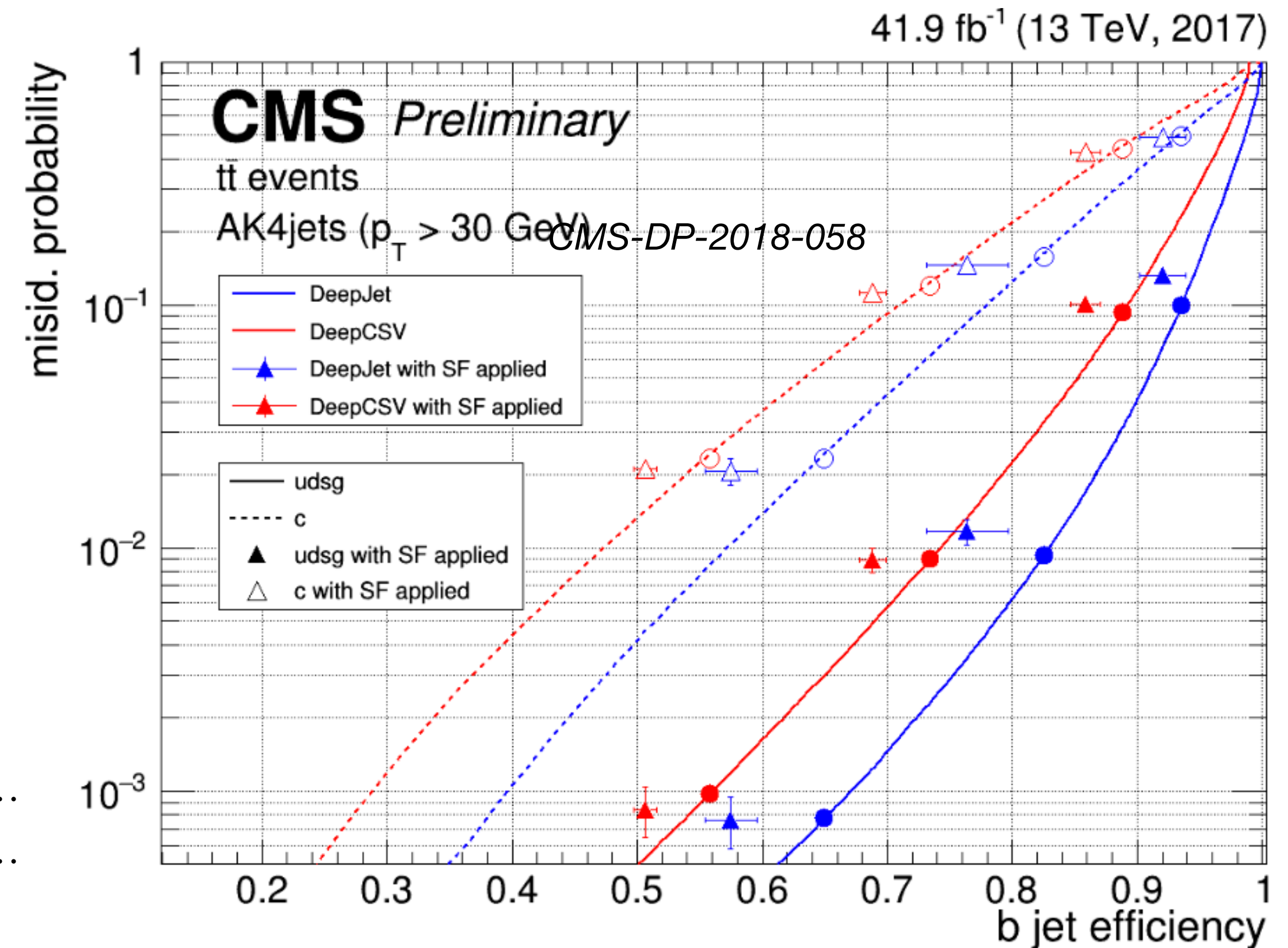
Performance benchmarks

Comparison w/ ATLAS and CMS

- Recurrent neural network along with IP3



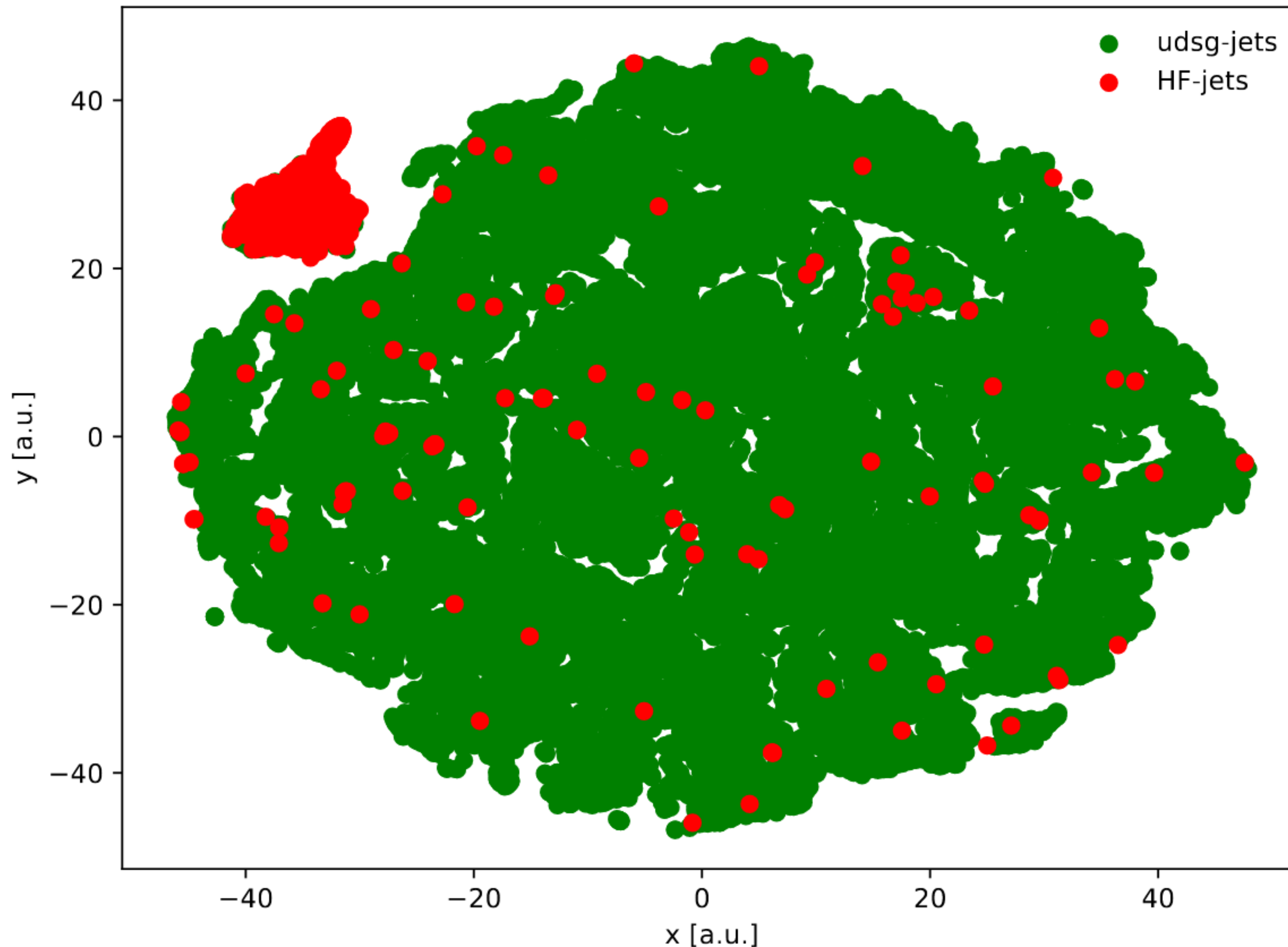
- Improvement of a factor 2-3 in background rejection compared to current leading models!



- 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 of JetVLAD Features



T-SNE projection

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

The basics

What is machine learning?

Why are these tools useful in high energy colliders?

How to quantify performance?

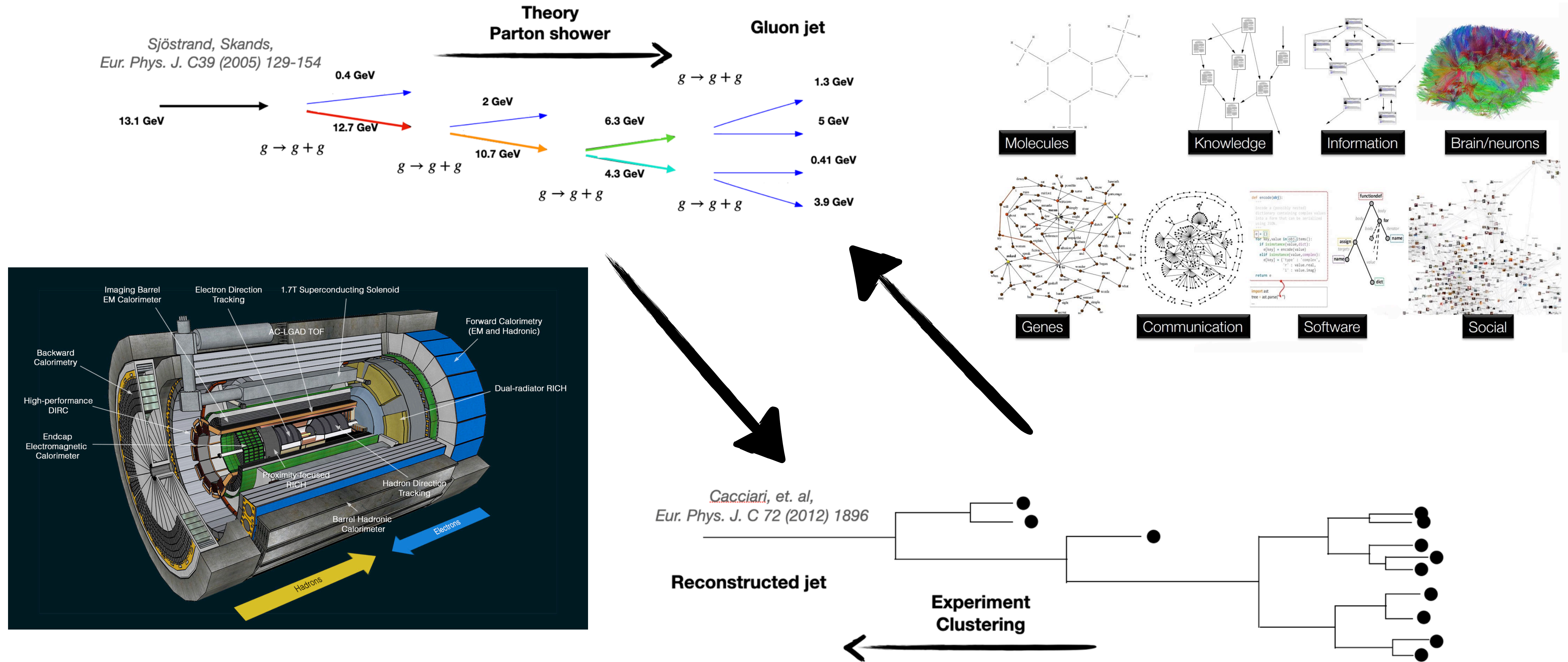
Physics with ML

Classifier - Can select Heavy-Flavor or Quark vs Gluons

Regressor - multi-dimensional correction and unfolding

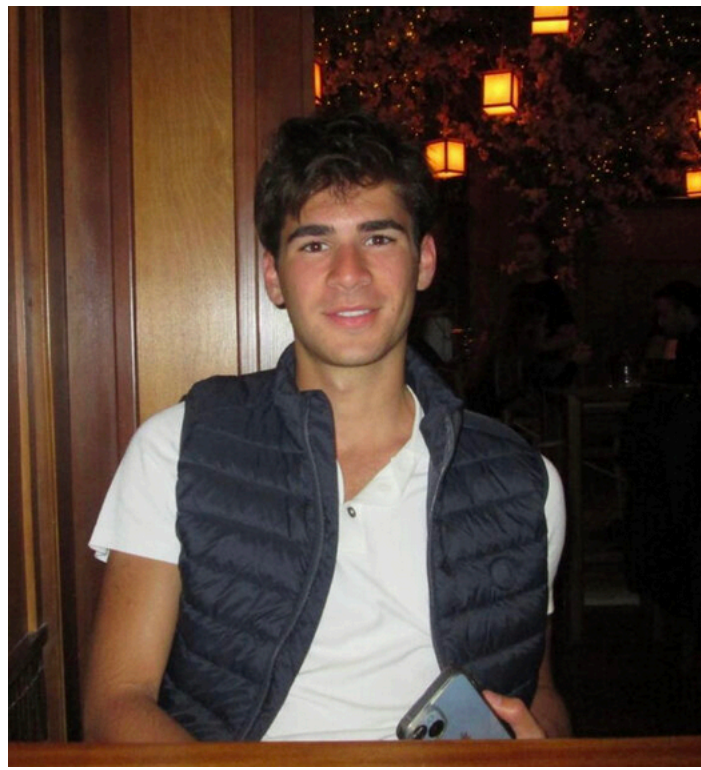
Generator - learn underlying physics of MC generators

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, Umar Soheil Qureshi,
Vandy Class of 2026 Vandy Class of 2025

Event Generation (PYTHIA 8.312)

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



pythia.org/latest-manual/welcome.html

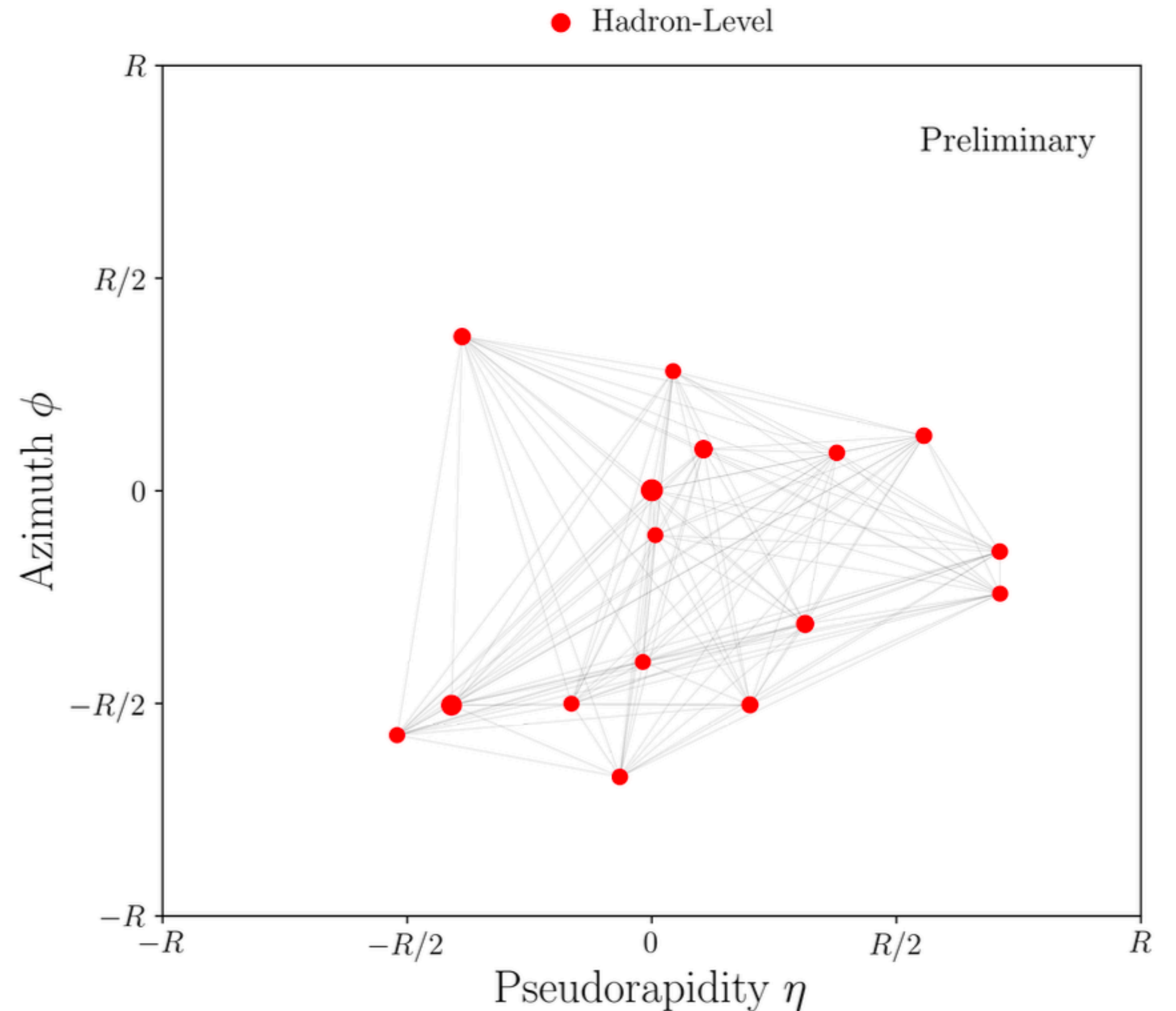
Graph Representation of Pythia Quark Jets

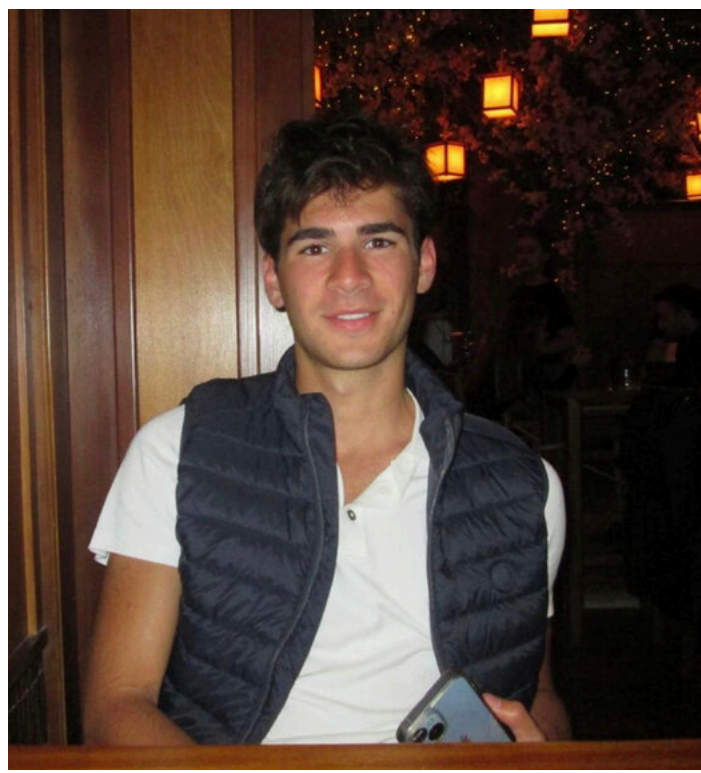
Jets represented as graphs, connected by ΔR :

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

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

Fully connected graphs, no self-loops.





Jets as connected graphs

Michael Taleb, Umar Soheil Qureshi,
Vandy Class of 2026 Vandy Class of 2025

Event Generation (PYTHIA 8.312)

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



pythia.org/latest-manual/welcome.html

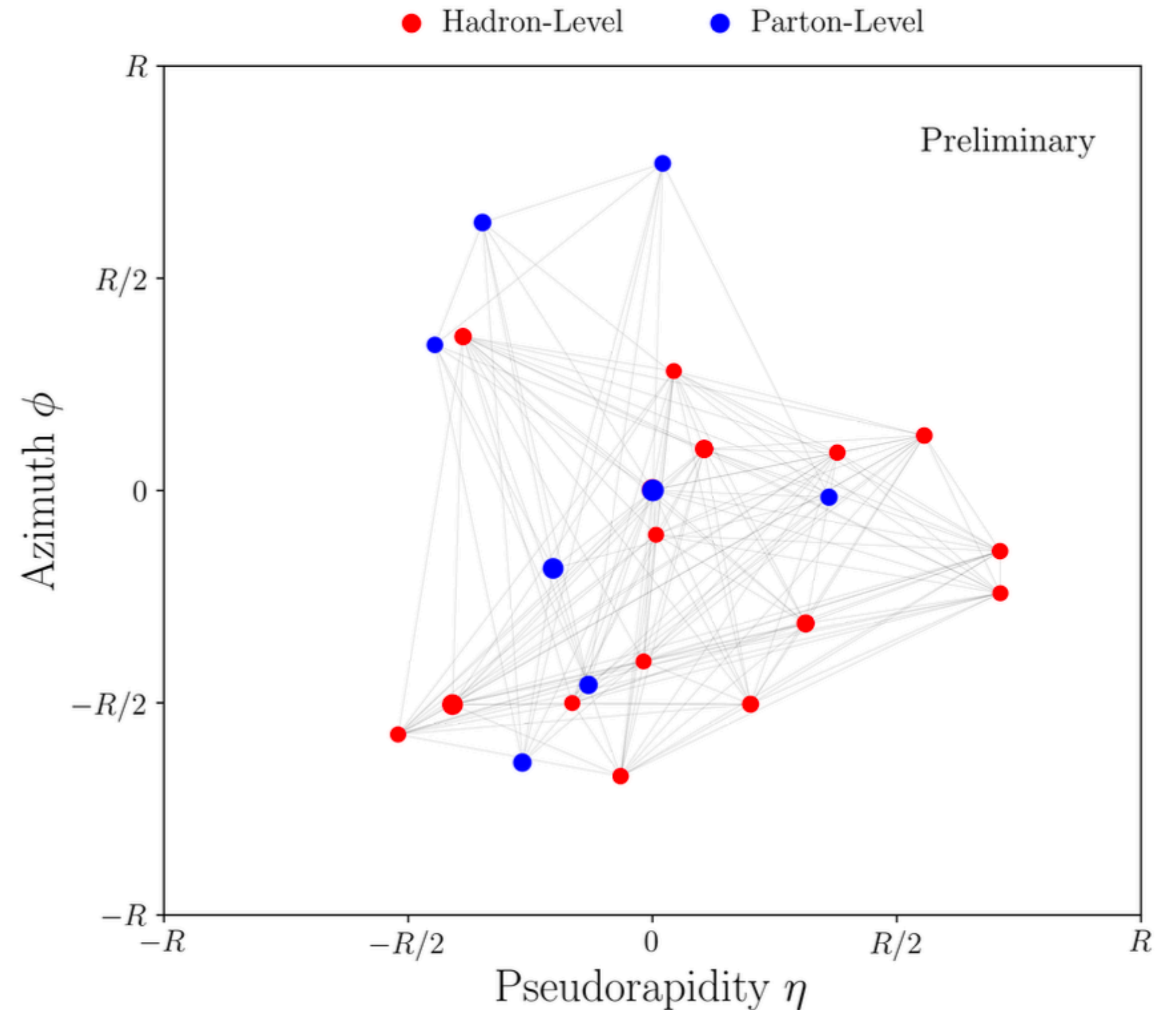
Graph Representation of Pythia Quark Jets

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

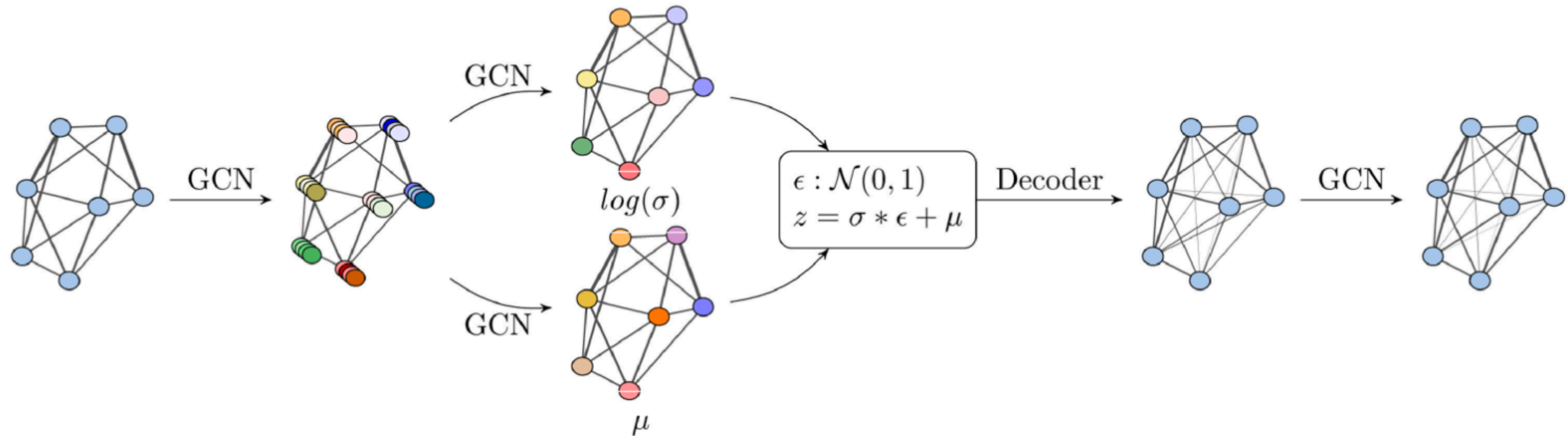
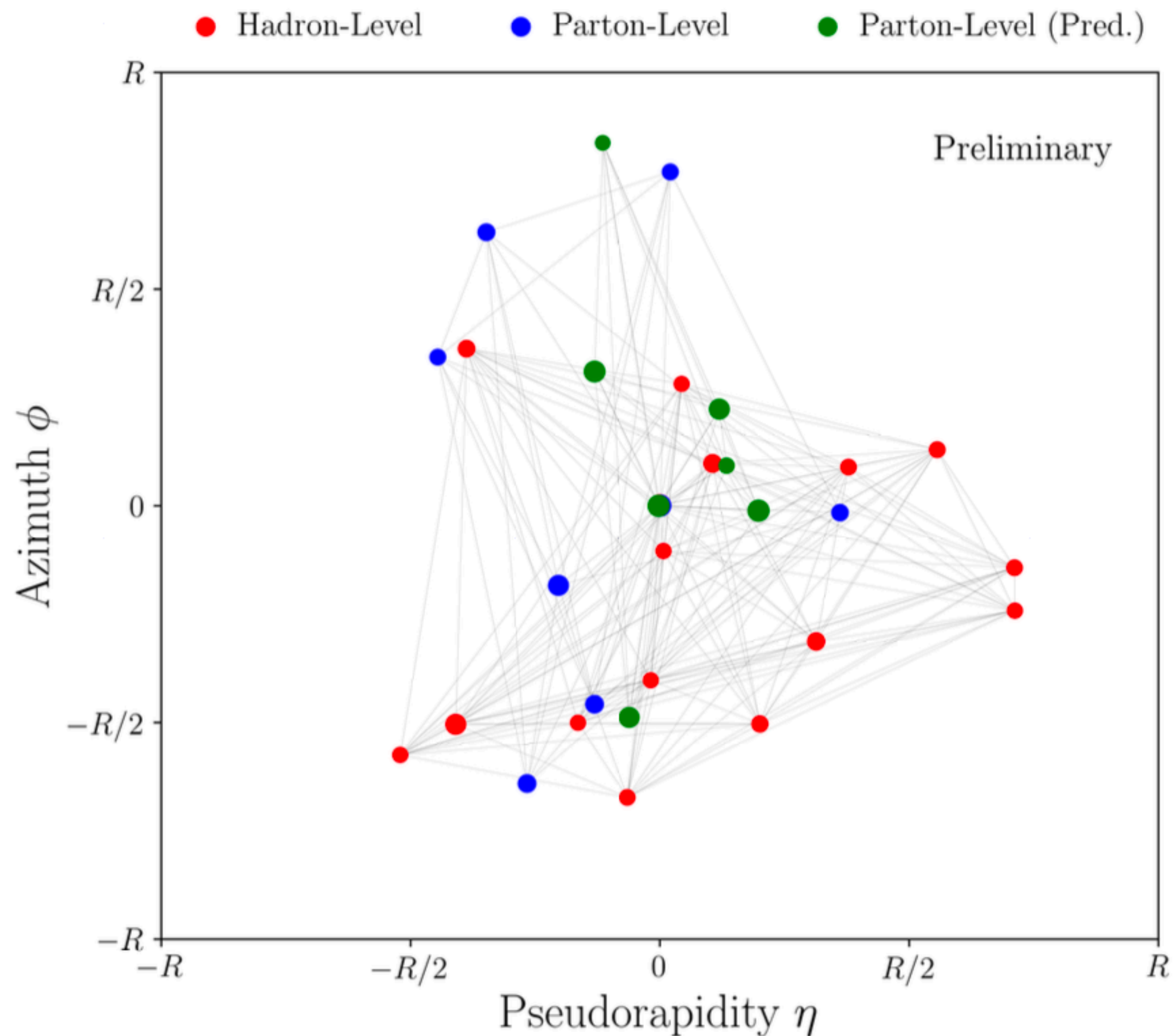


Image Credit: Tina Behrouzi et. al.

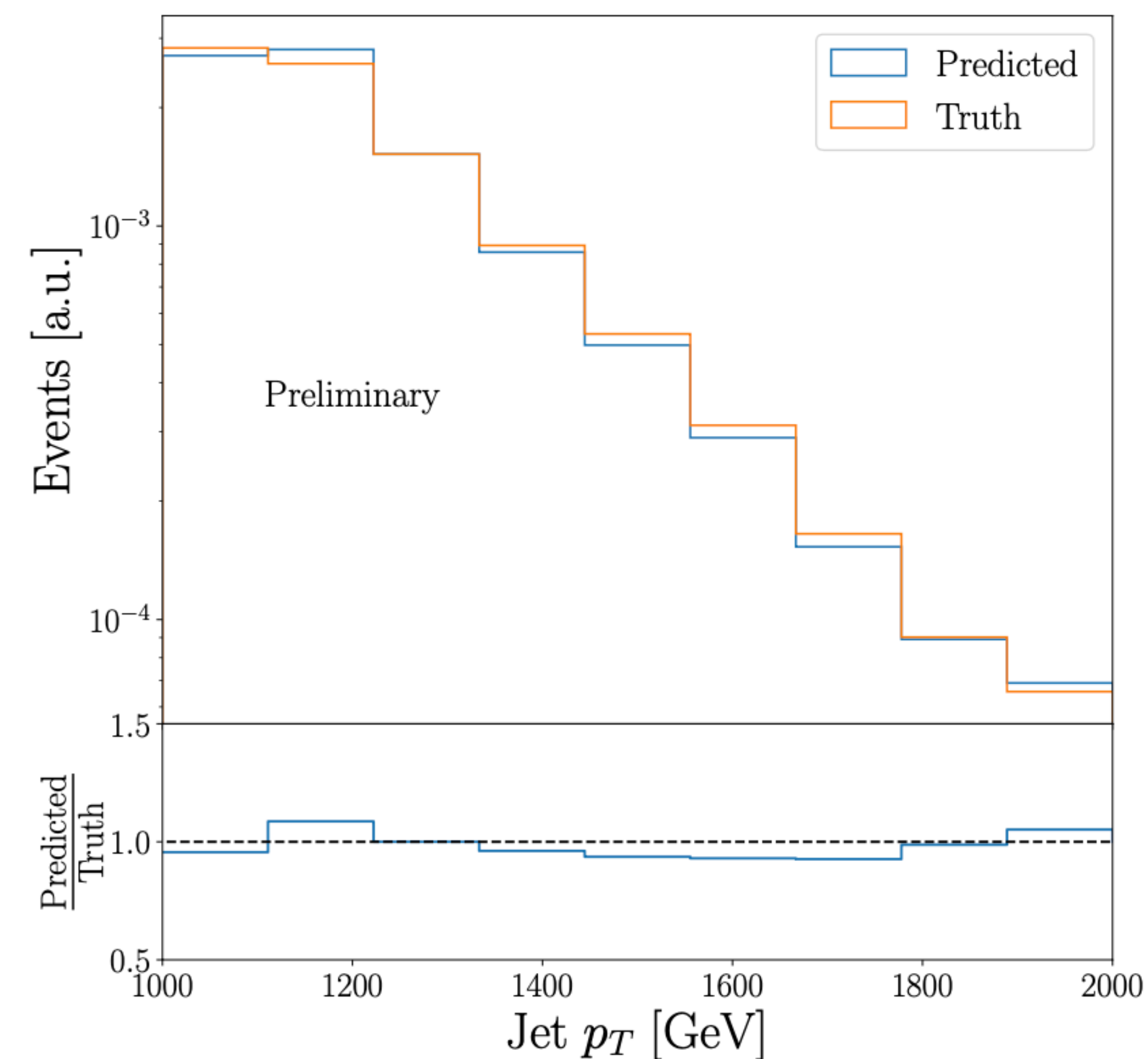
Variational Graph Autoencoder (VGAE)

- 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.

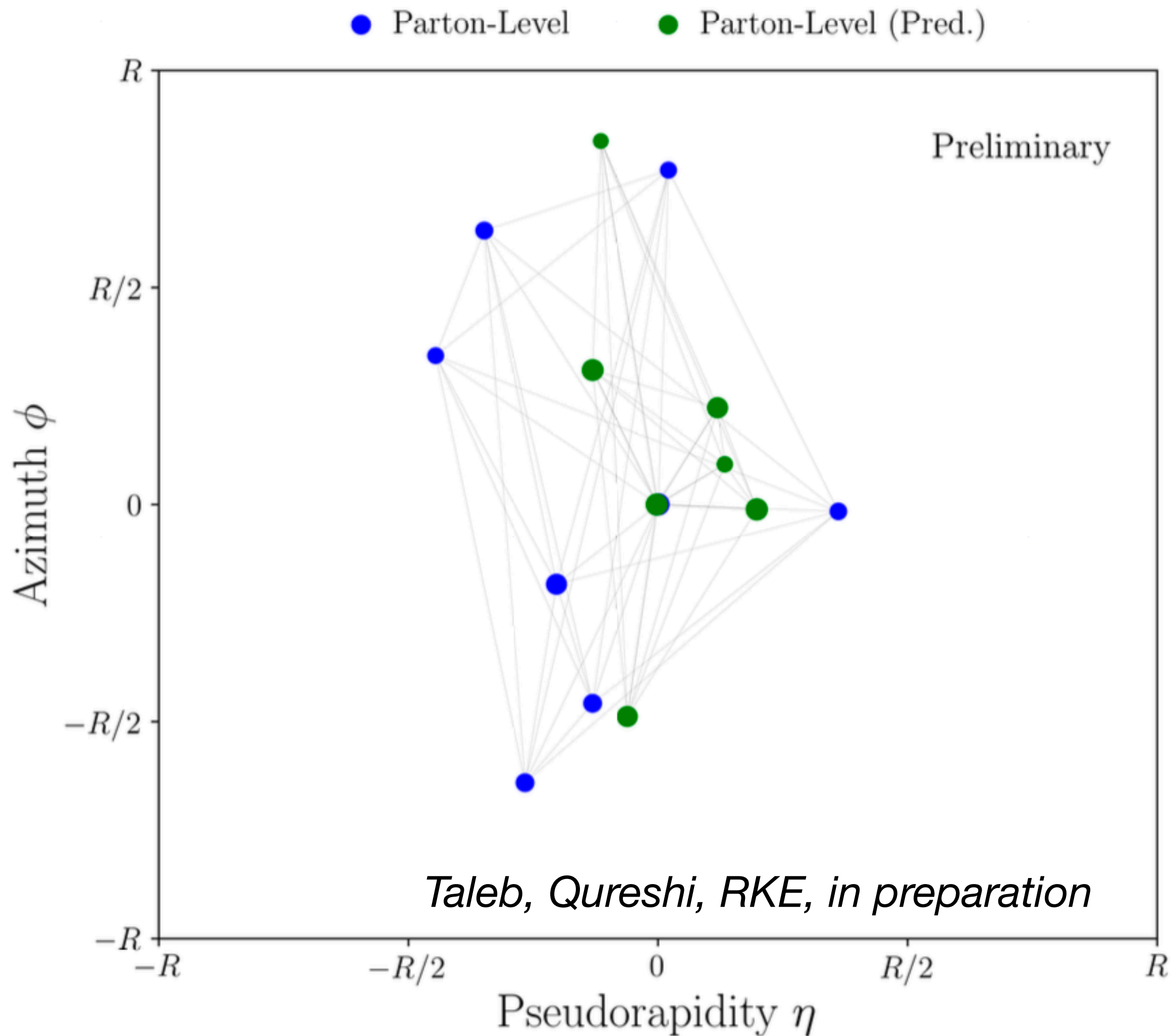


Predicted Parton level Jet



- 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 $\mathcal{G}'$.

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

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

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

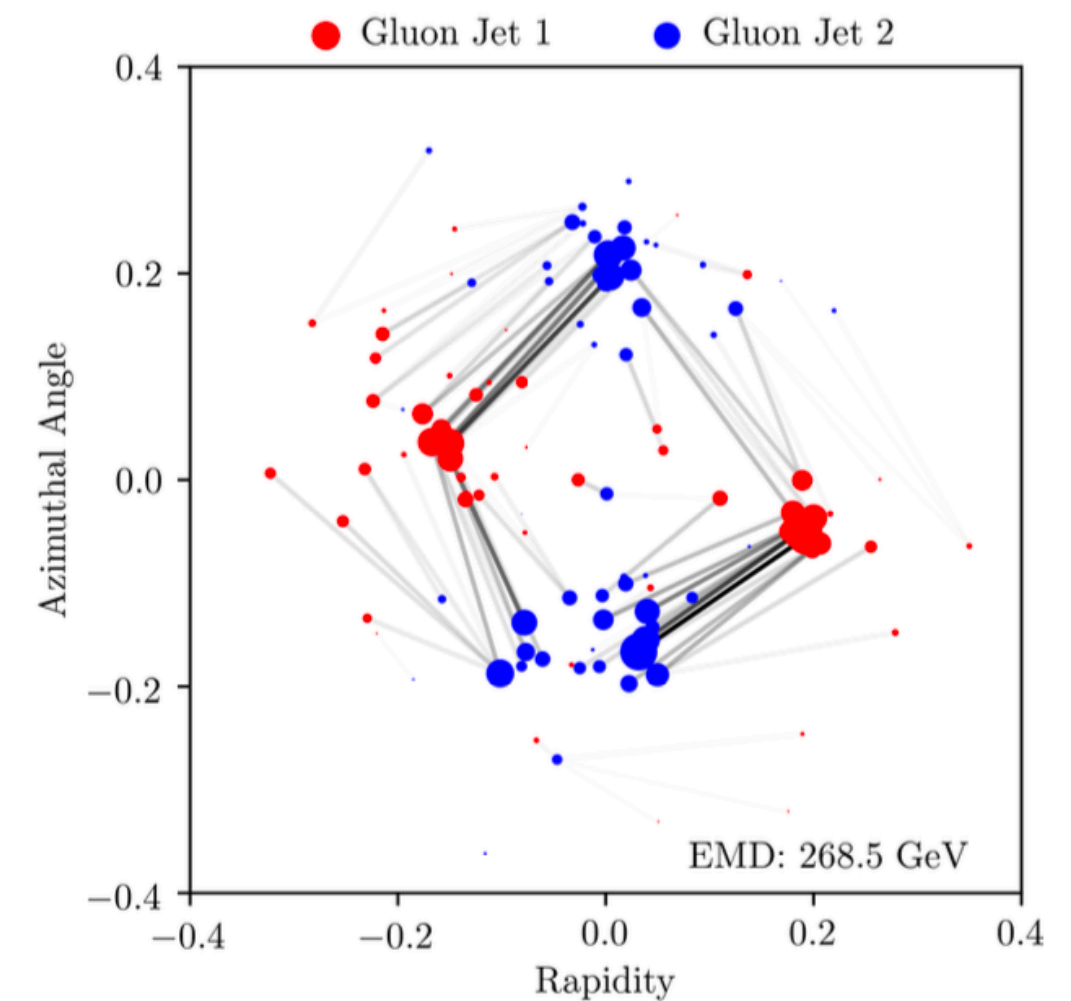
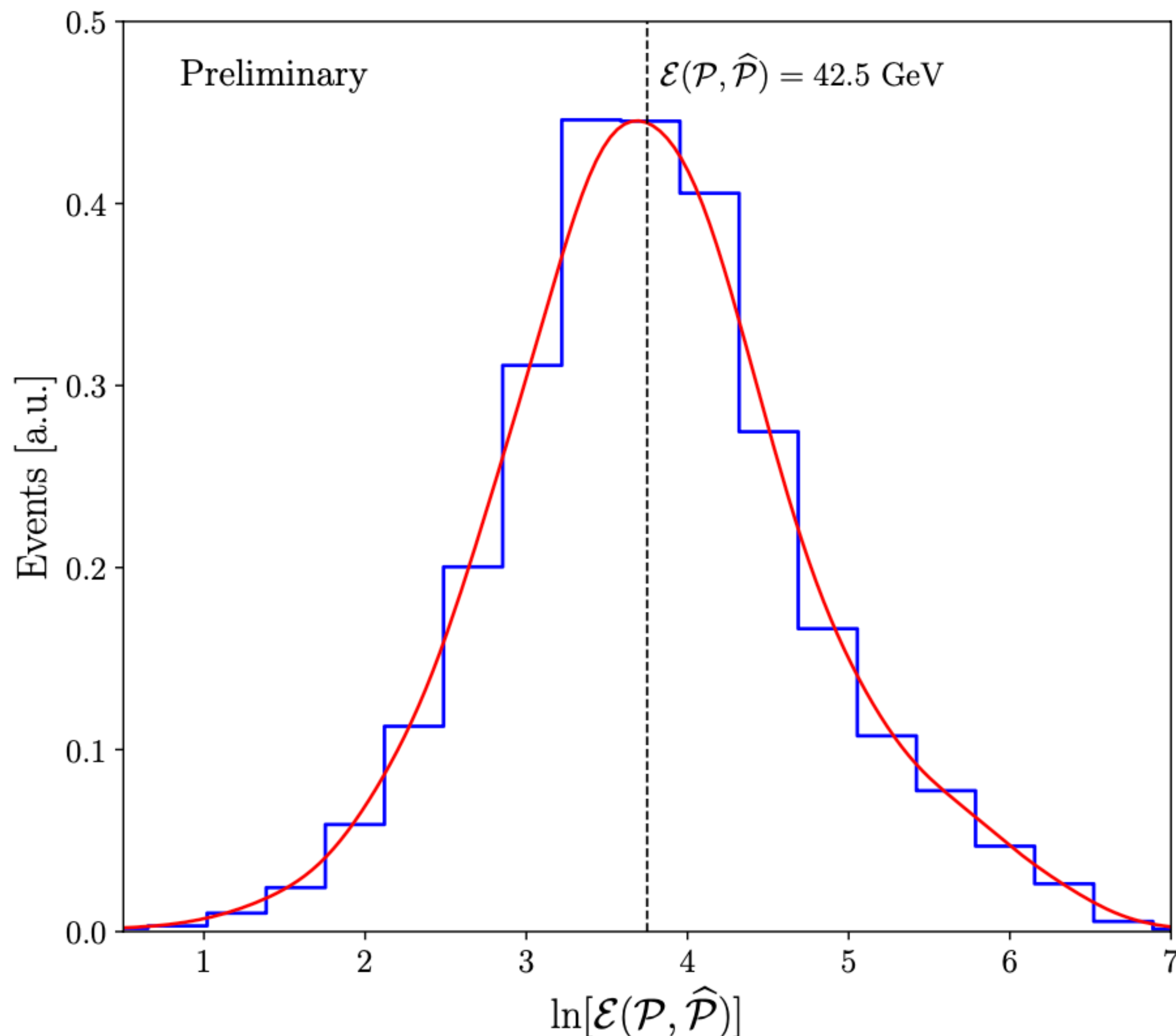


Figure 5: EMD between two gluon jets.

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



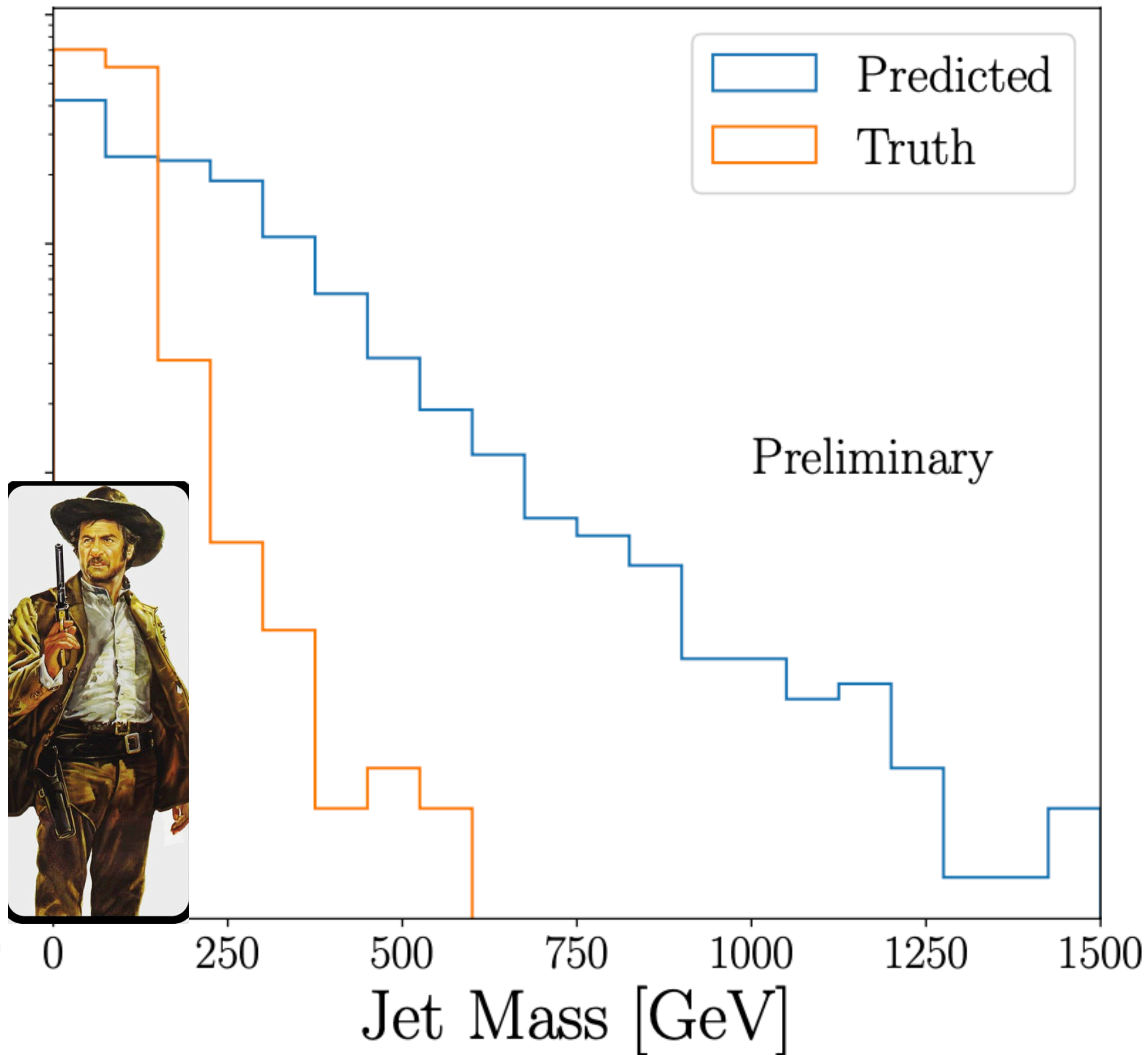
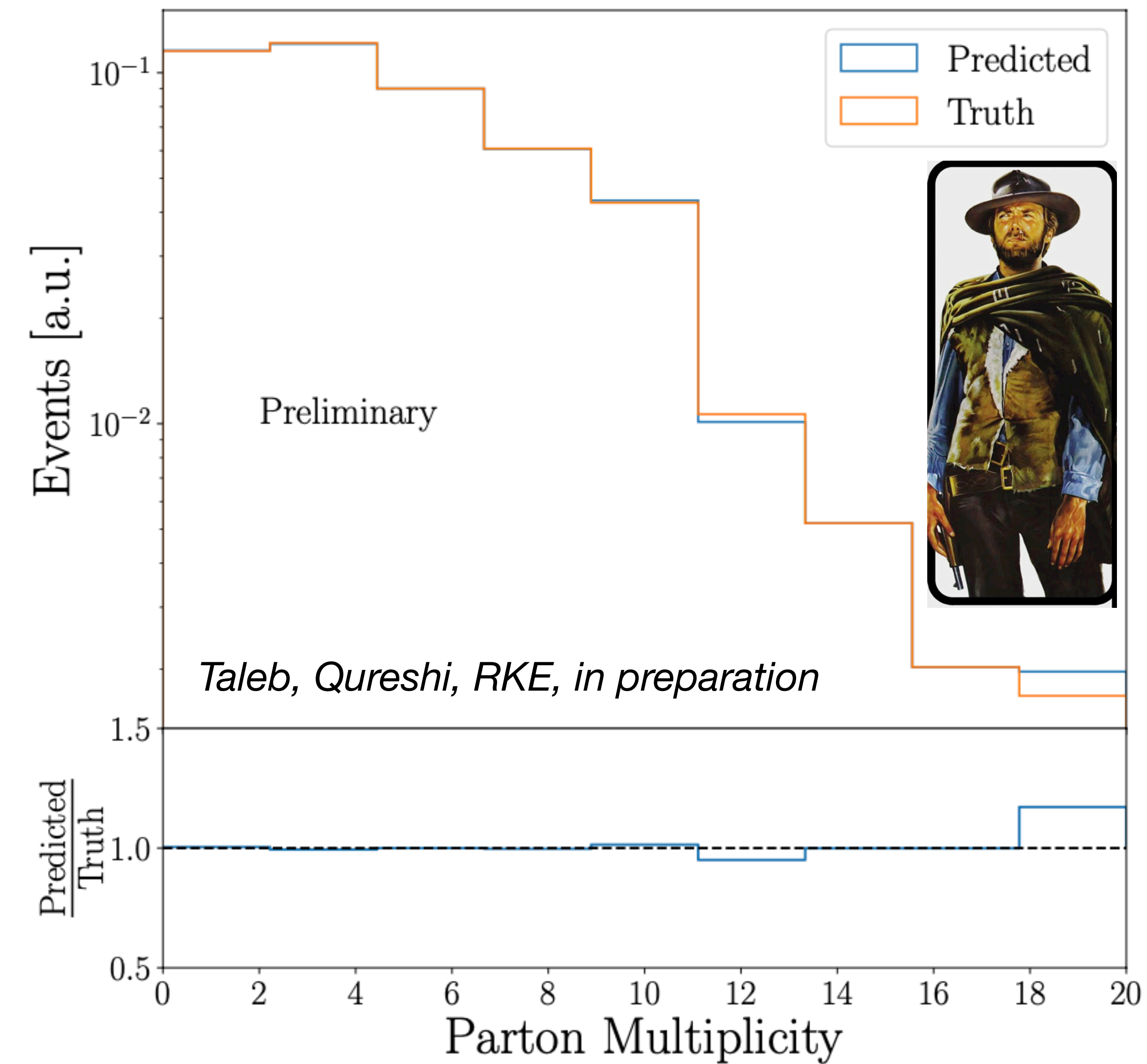
Predicted jets close to ground truth (Pythia)!

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

Ugly



The basics

What is machine learning?

Why are these tools useful in high energy colliders?

How to quantify performance?

Physics with ML

Classifier - Can select Heavy-Flavor or Quark vs Gluons

Regressor - multi-dimensional correction and unfolding

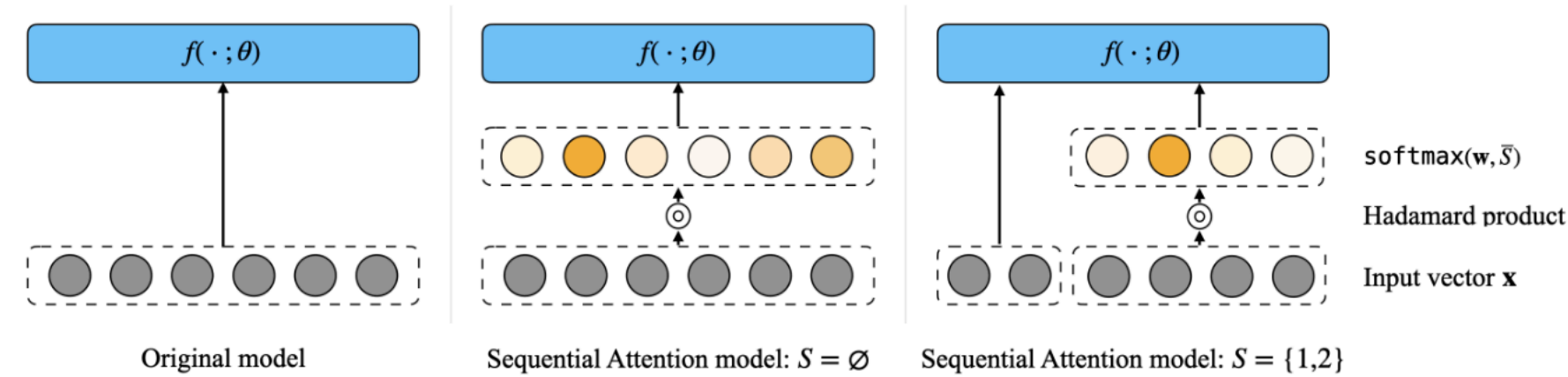
Generator - learn underlying physics of MC generators

Detour - Can we identify jets that have modified fragmentation



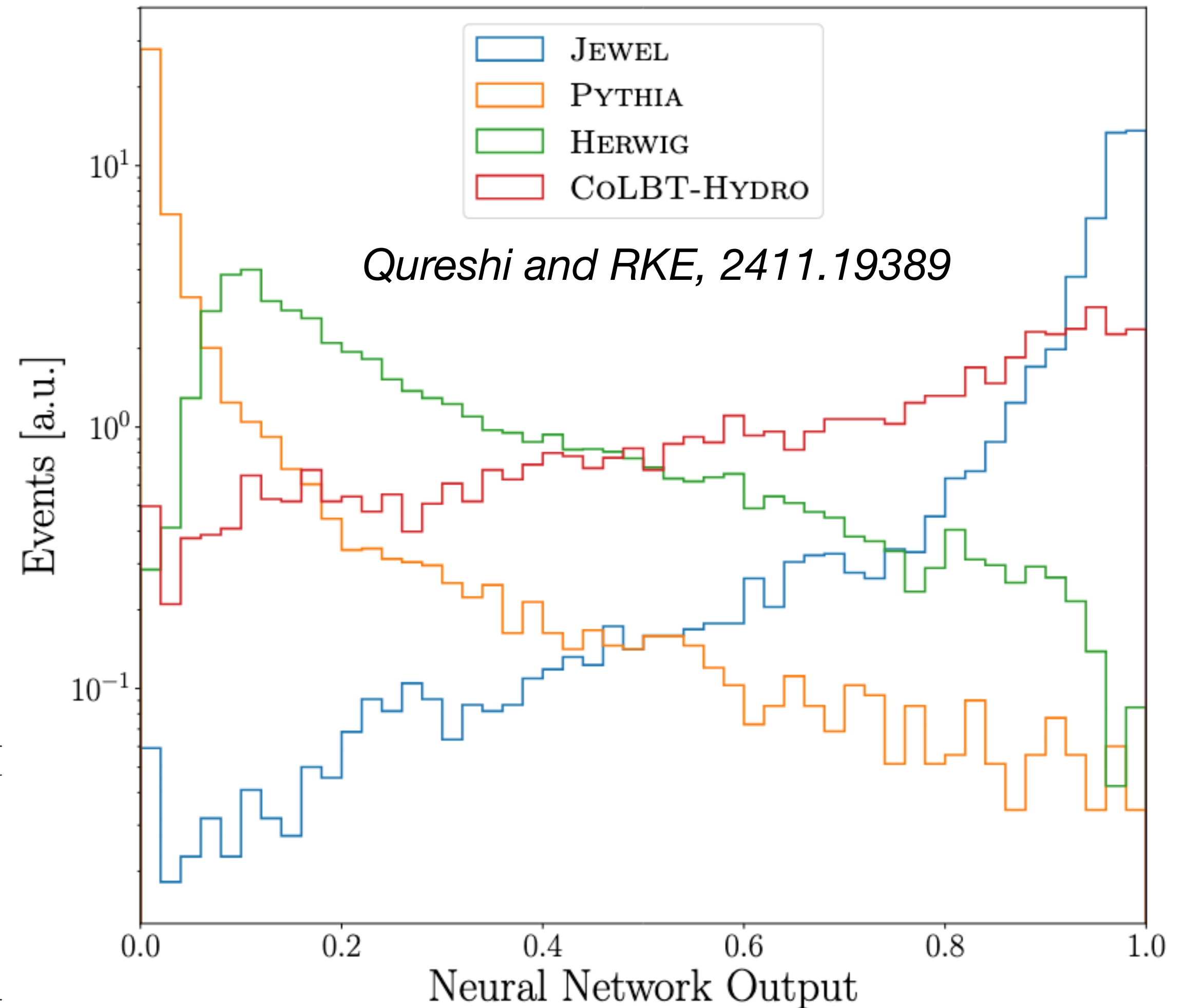
Umar Soheil Qureshi,
Vandy Class of 2025

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 $w \in \mathbb{R}^d$ 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!

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

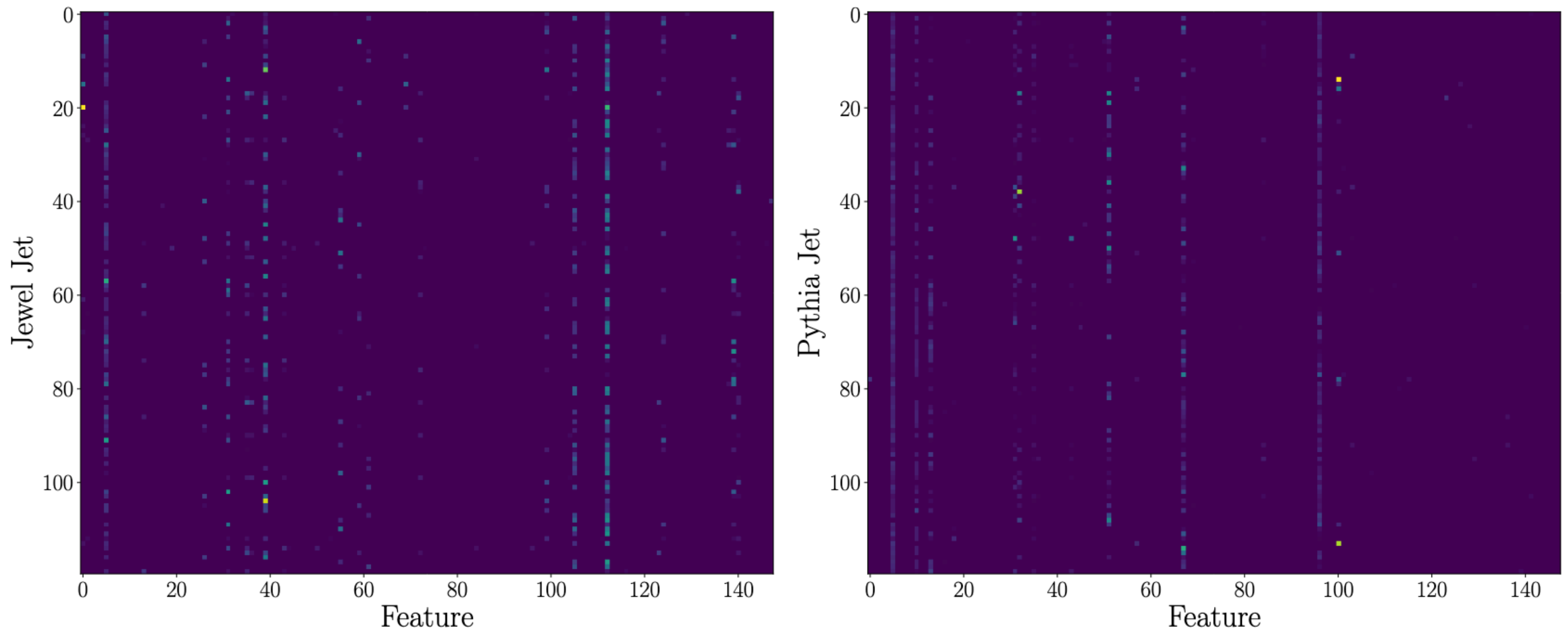


FIG. 4. Heatmap illustrations of the aggregate feature mask (Eq. [8](#)) for the first 125 JEWEL (left) and PYTHIA (right) truth jets. The sparsity in feature activation highlights the attention-based mechanism's focus on relevant features for classification.

The basics

What is machine learning?

Why are these tools useful in high energy colliders?

How to quantify performance?

Physics with ML

Classifier - Can select Heavy-Flavor or Quark vs Gluons

Regressor - multi-dimensional correction and unfolding

Generator - learn underlying physics of MC generators

Events as noisy images

Event Generation (JEWEL)

- $\sqrt{s_{\text{NN}}} = 5.02$ TeV PbPb beams.
- Dijets at 0-10% centrality.
- $\hat{p}_{\text{T}} > 100$ 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 $\langle p_{\text{T}} \rangle = 1.2$ GeV.

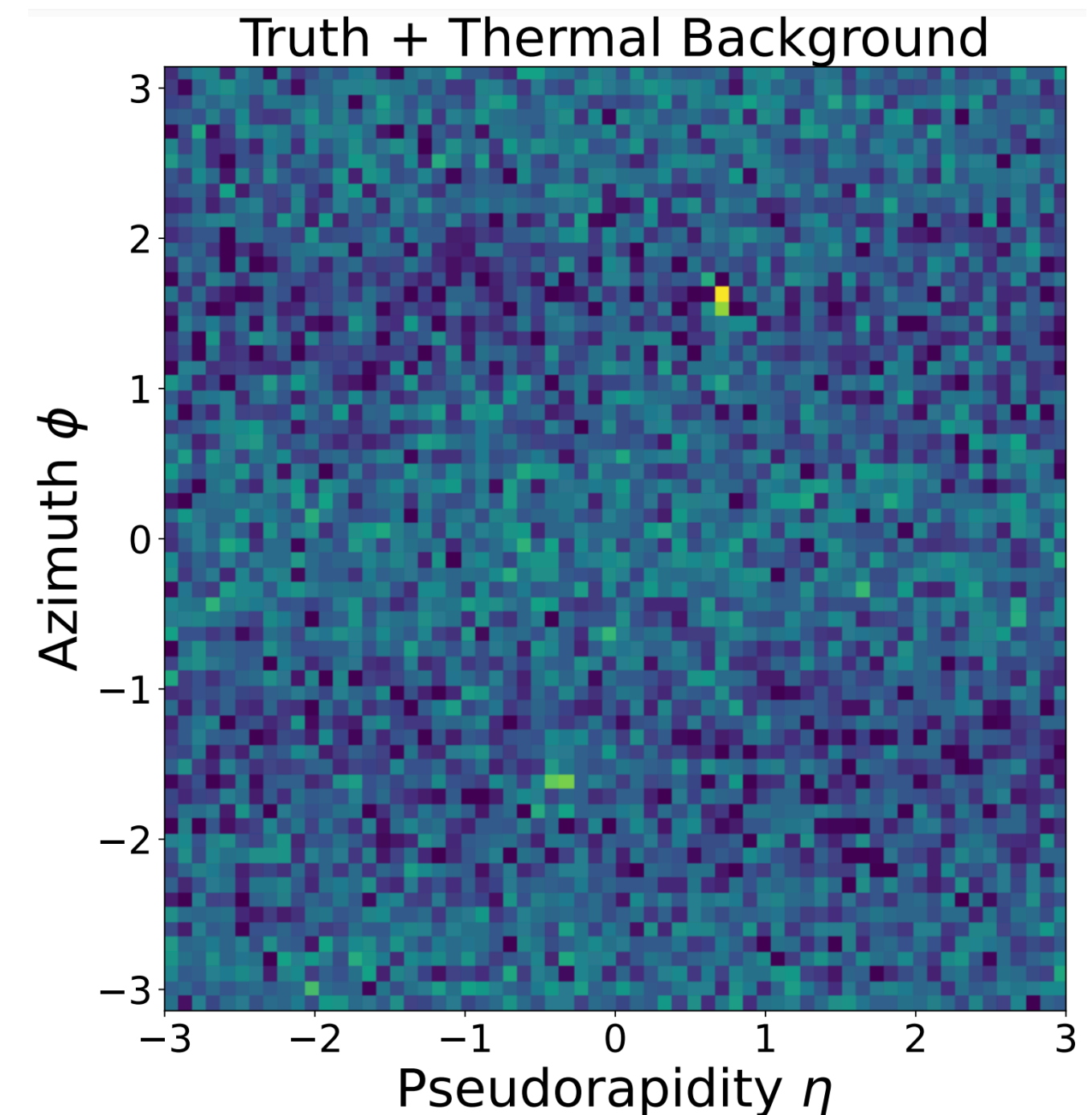
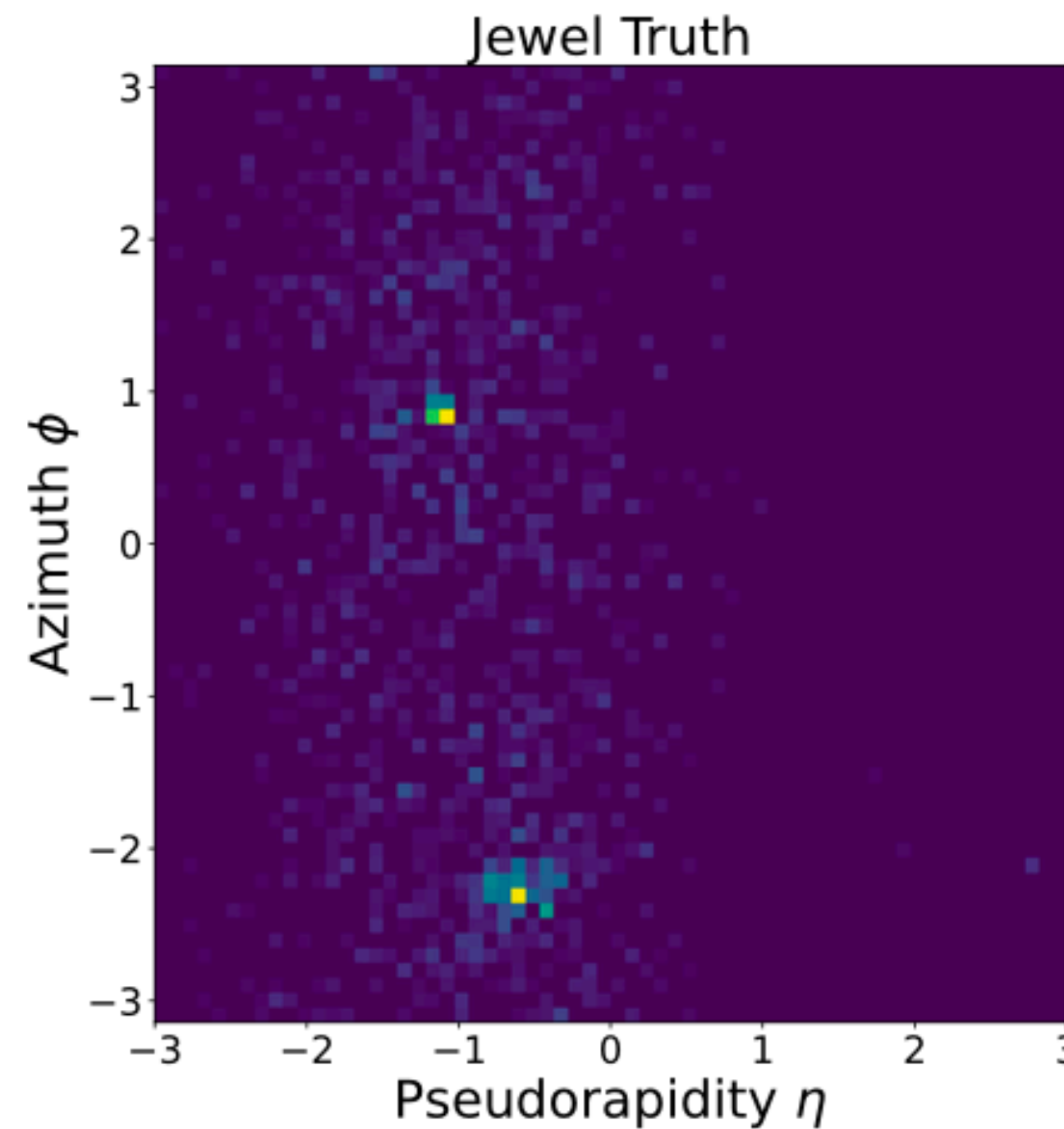


Image Representations

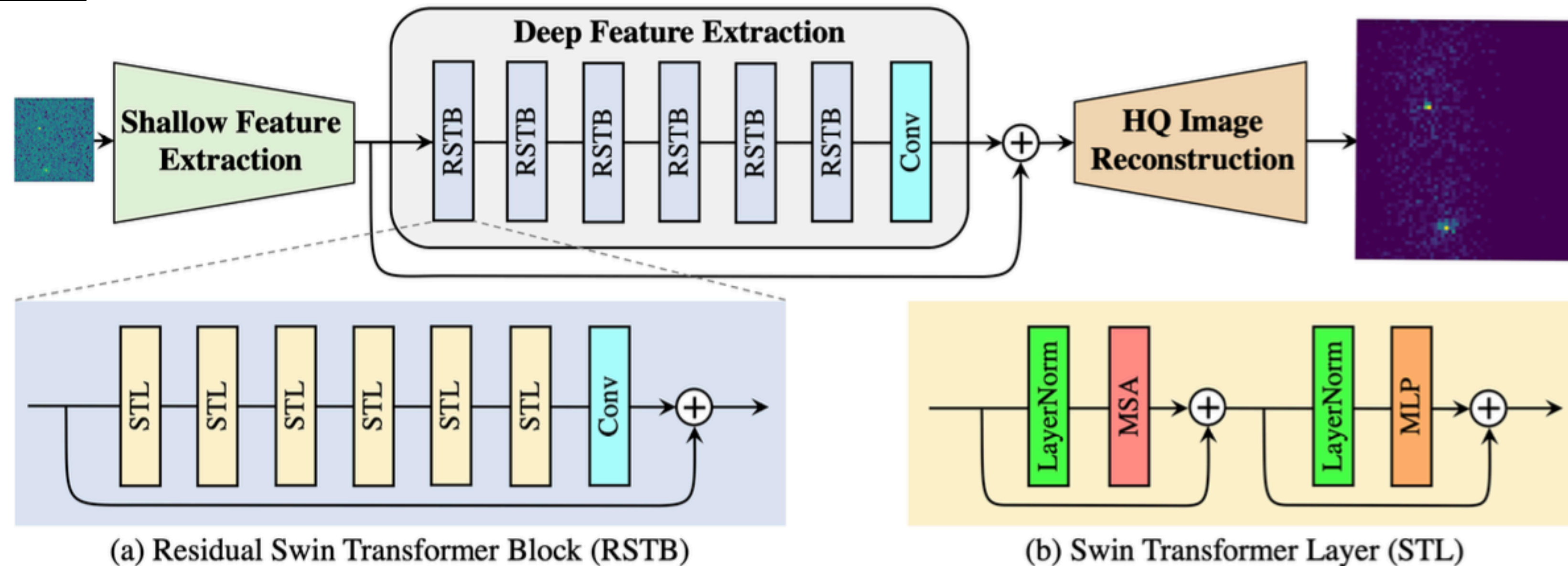
- Dijet events as images in the $(\eta, \phi) \in [-3, 3] \times [-\pi, \pi]$ plane.

*Umar Soheil Qureshi,
Vandy Class of 2025*



Hierarchical Vision Transformer using Shifted Windows

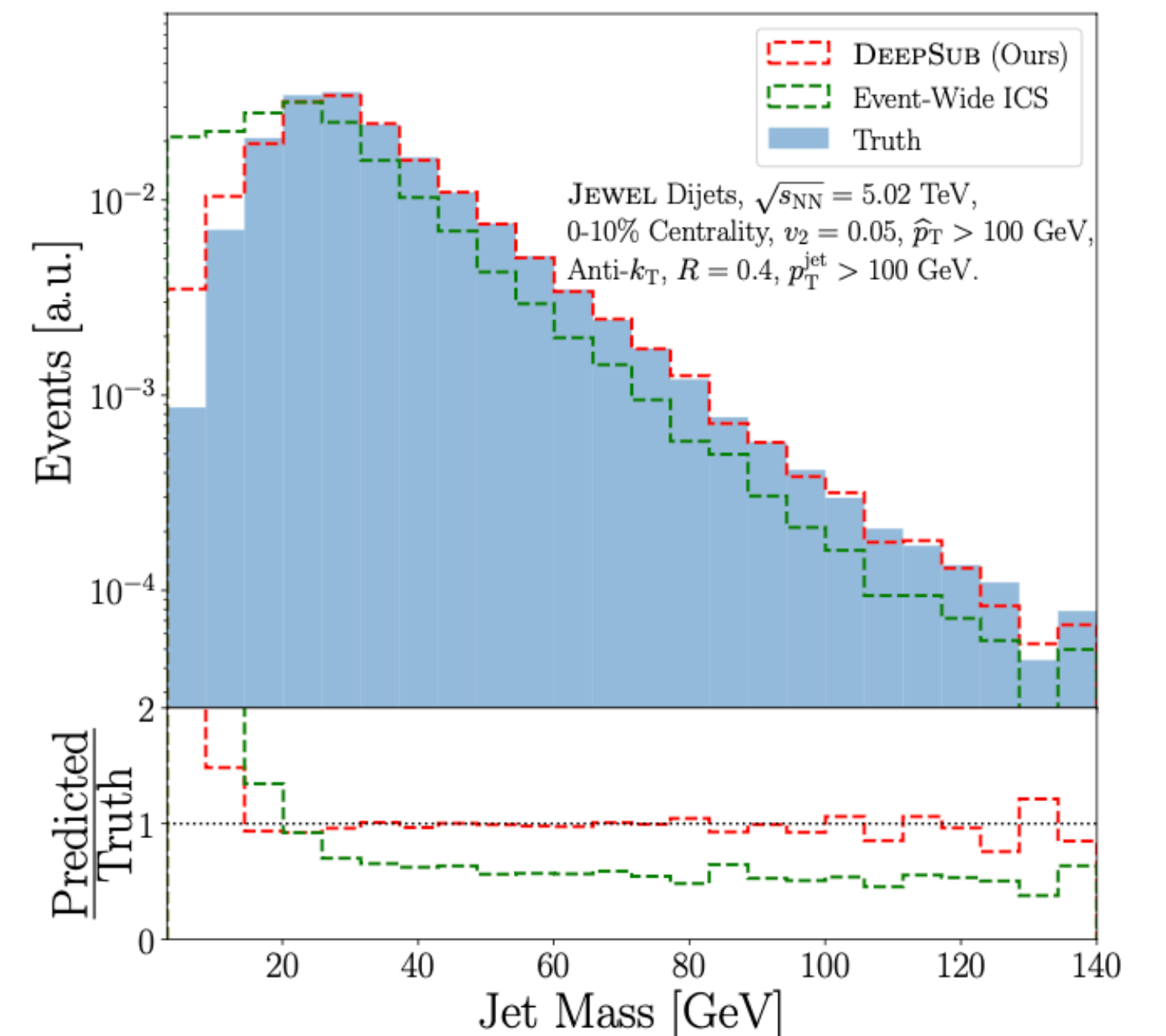
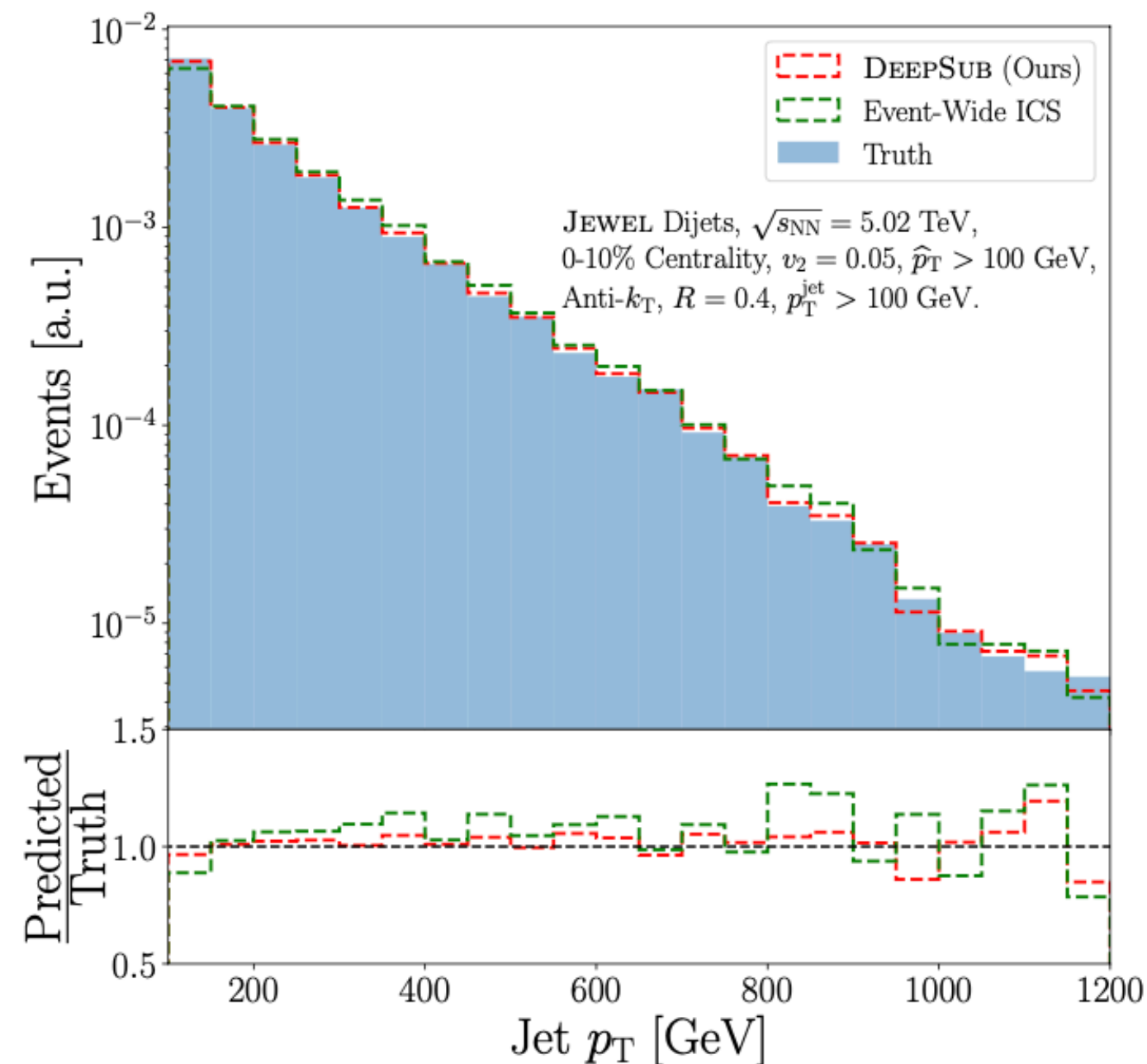
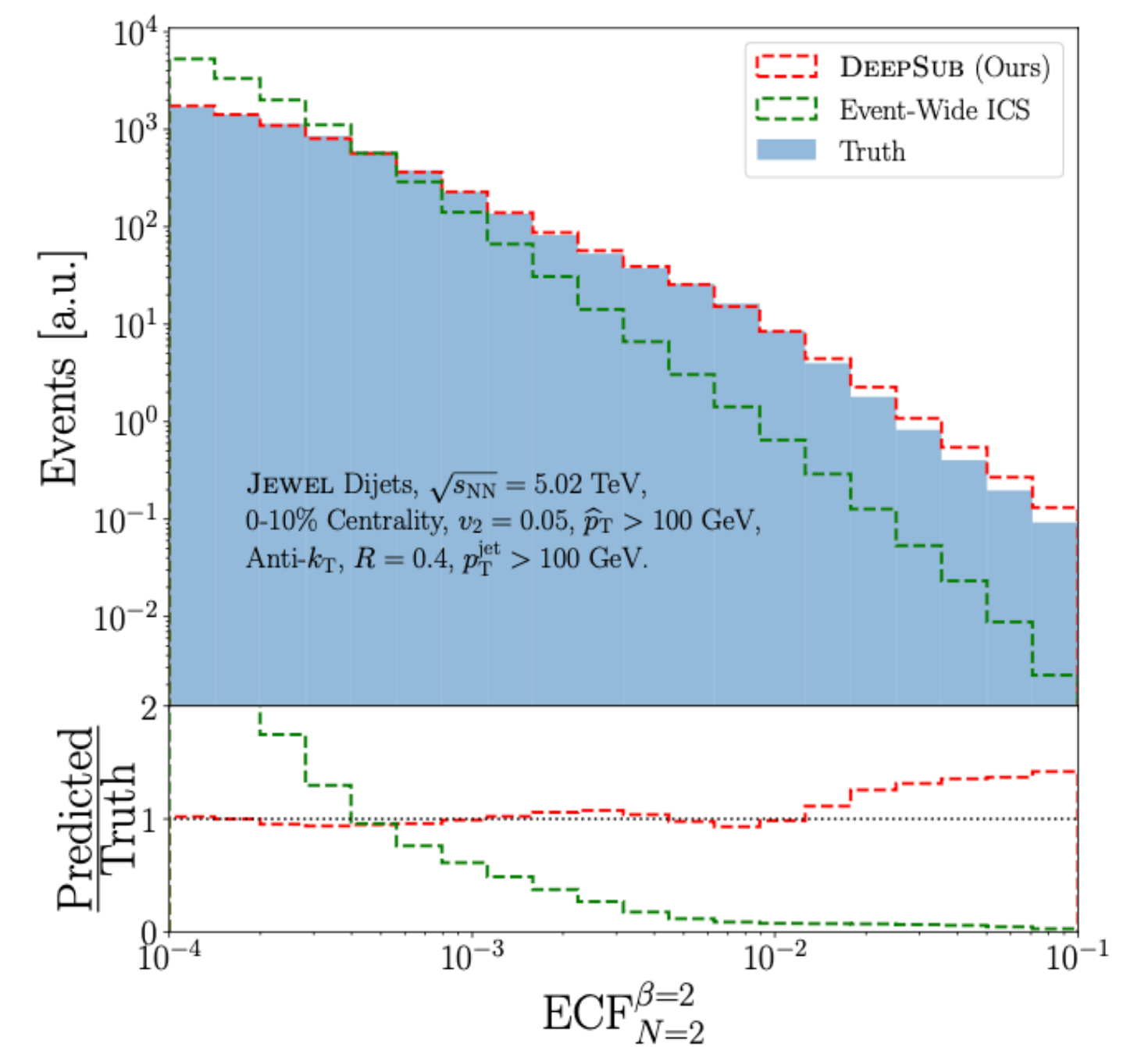
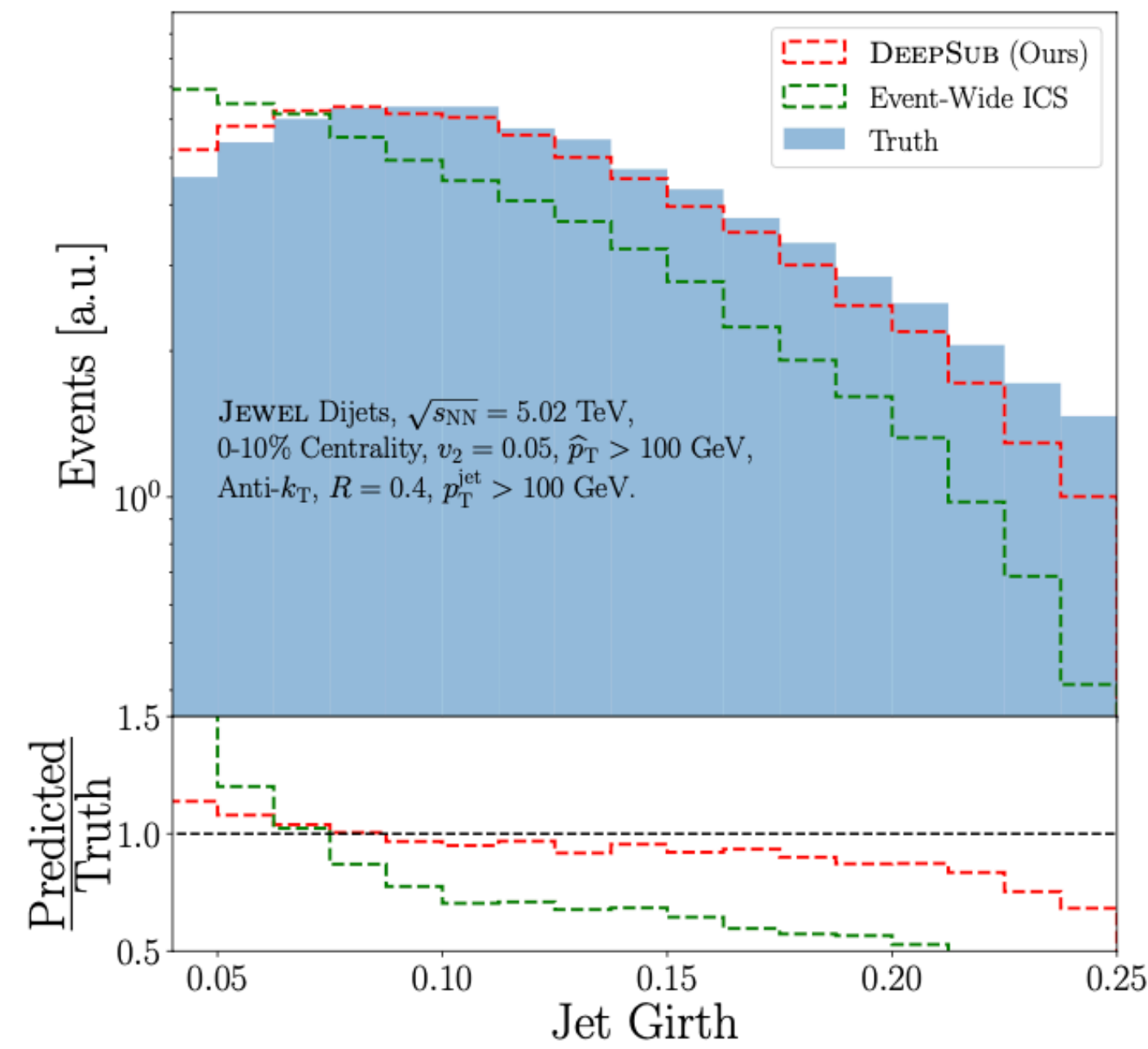
Liu, Ze et.al [2103.14030](#)



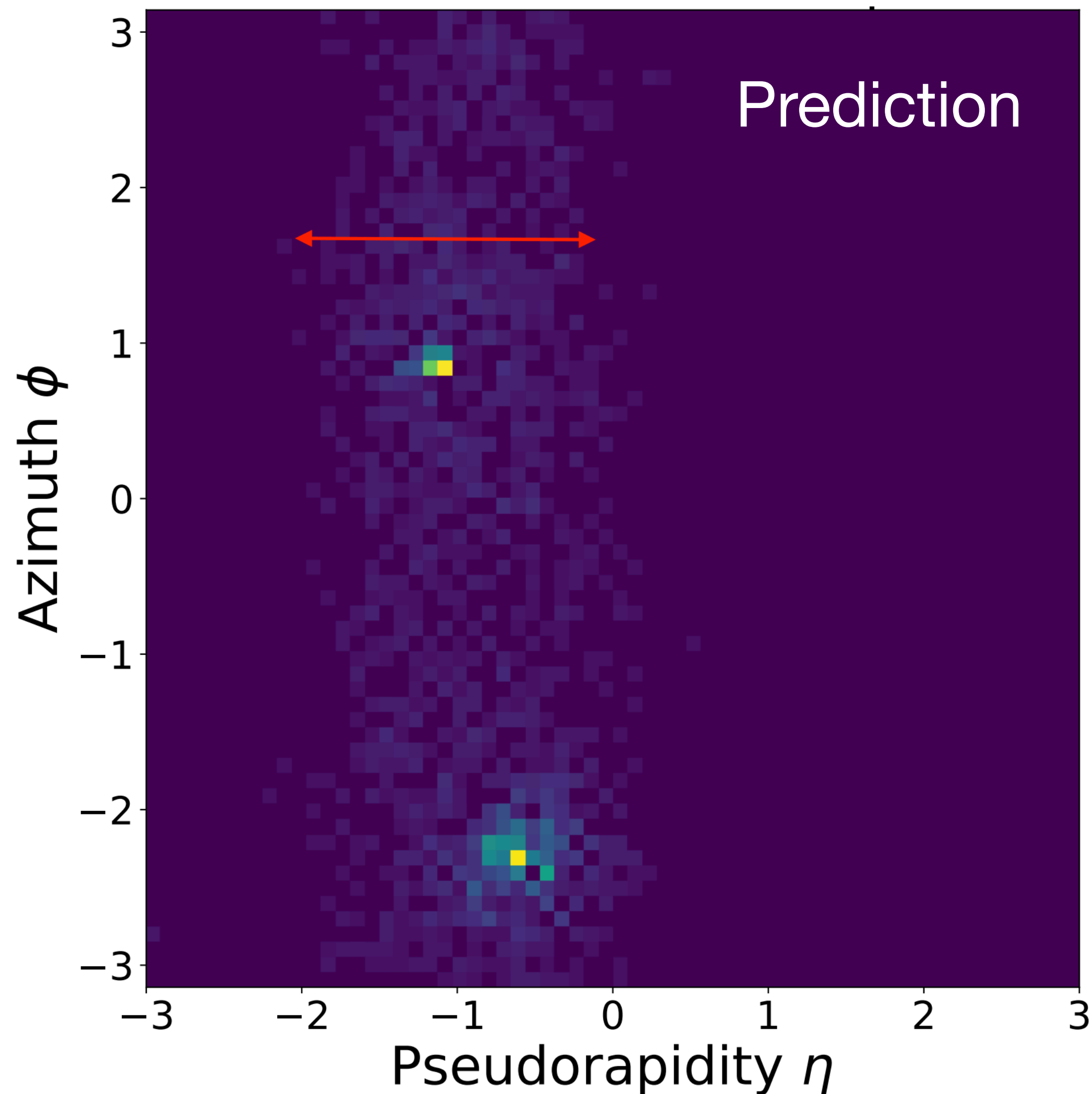
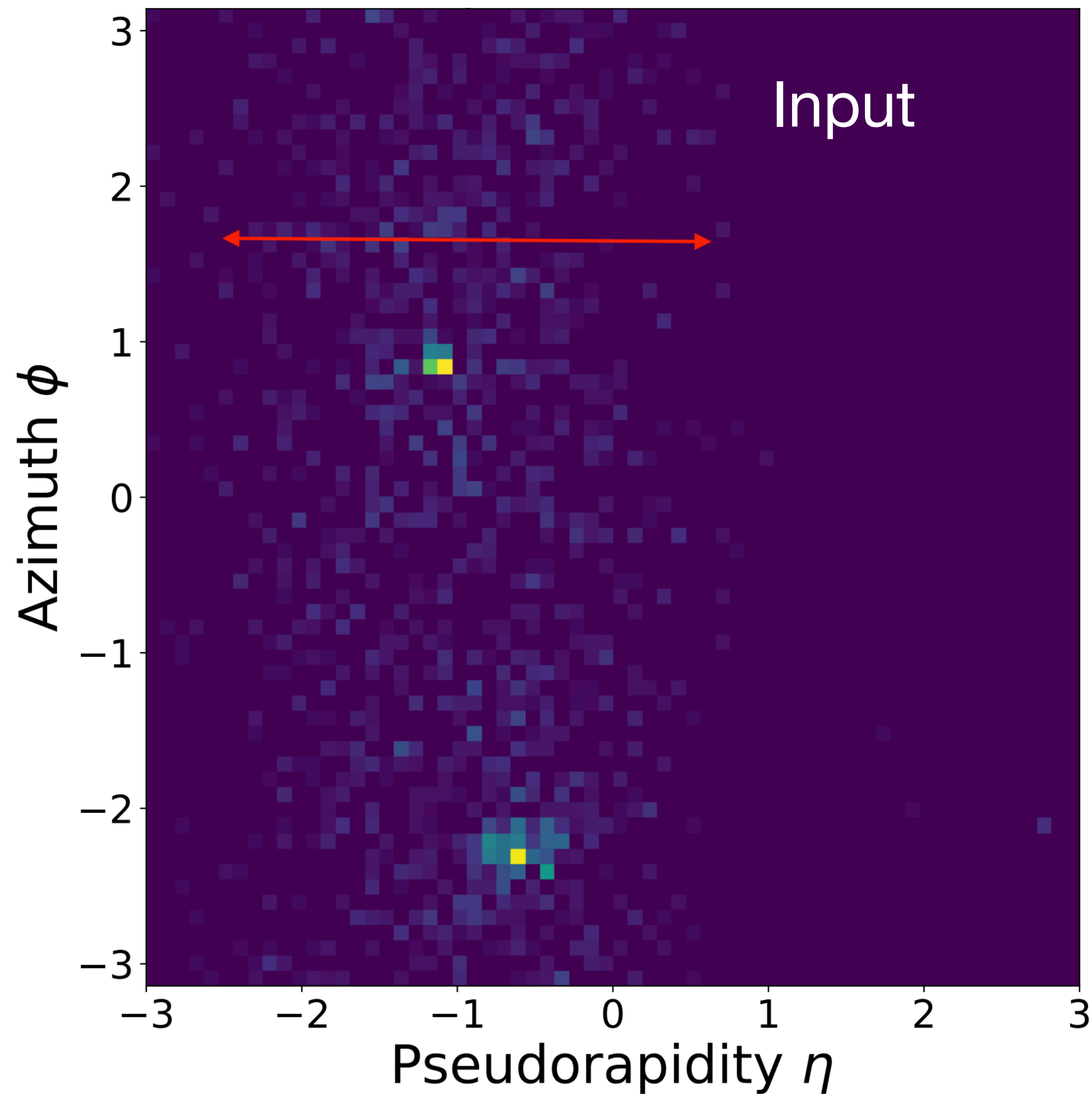
- The shifted windowing scheme brings greater efficiency by limiting self-attention computation to non-overlapping local windows while also allowing for cross-window connection.
- This hierarchical architecture has the flexibility to model at various scales and has linear computational complexity with respect to image size.

DEEPSUB vs Consti-Sub

- 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

