Machine Learning in Particle Physics

BERKELEY LAB

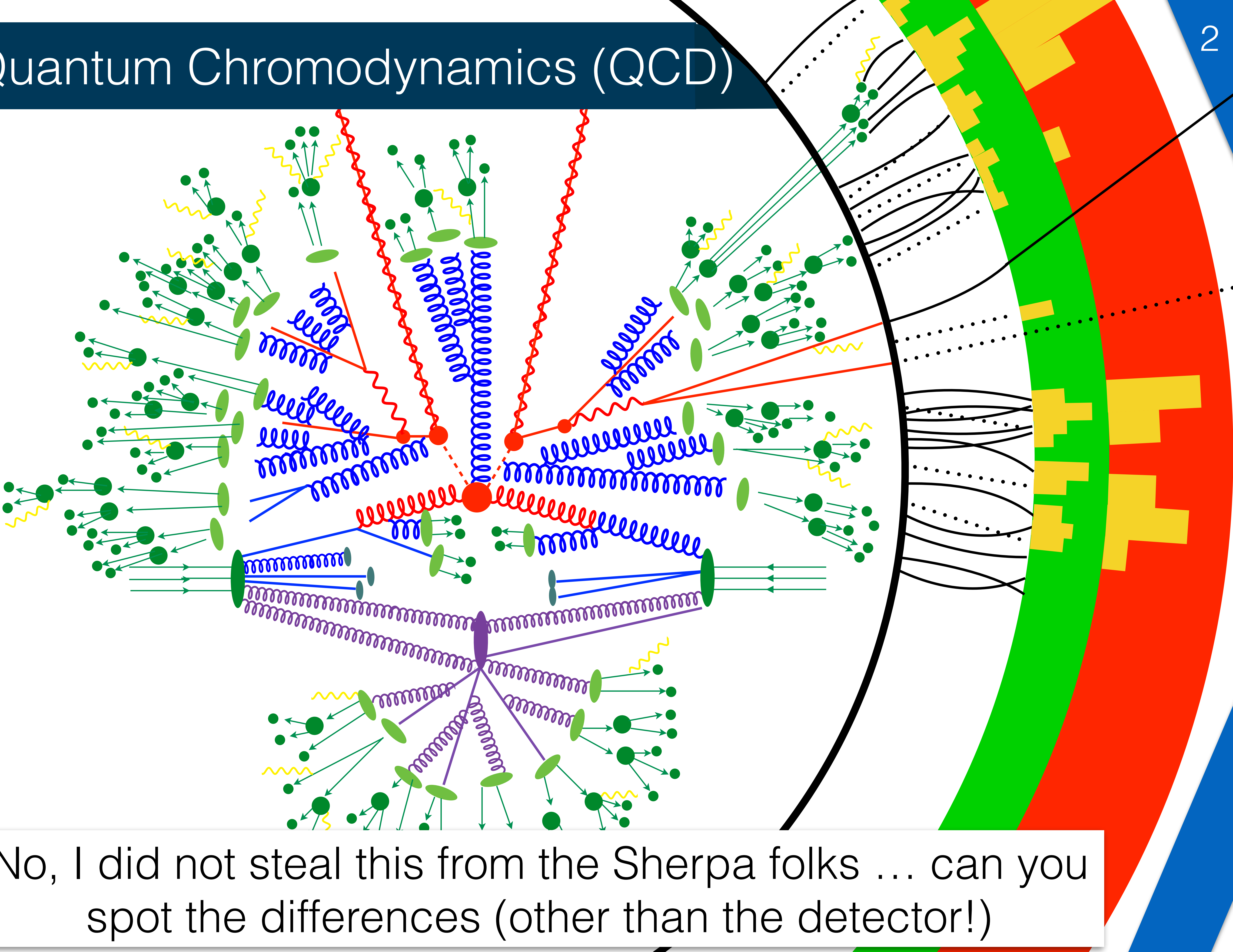
Benjamin Nachman

Lawrence Berkeley National Laboratory

MC4BSM 2017

Disclaimer: "Machine Learning" is an extremely broad (and **hot**) topic: nearly all analysis work is ML. I'll focus on state-of-the-art **image processing** with **jets** as this is really pushing the cutting edge.

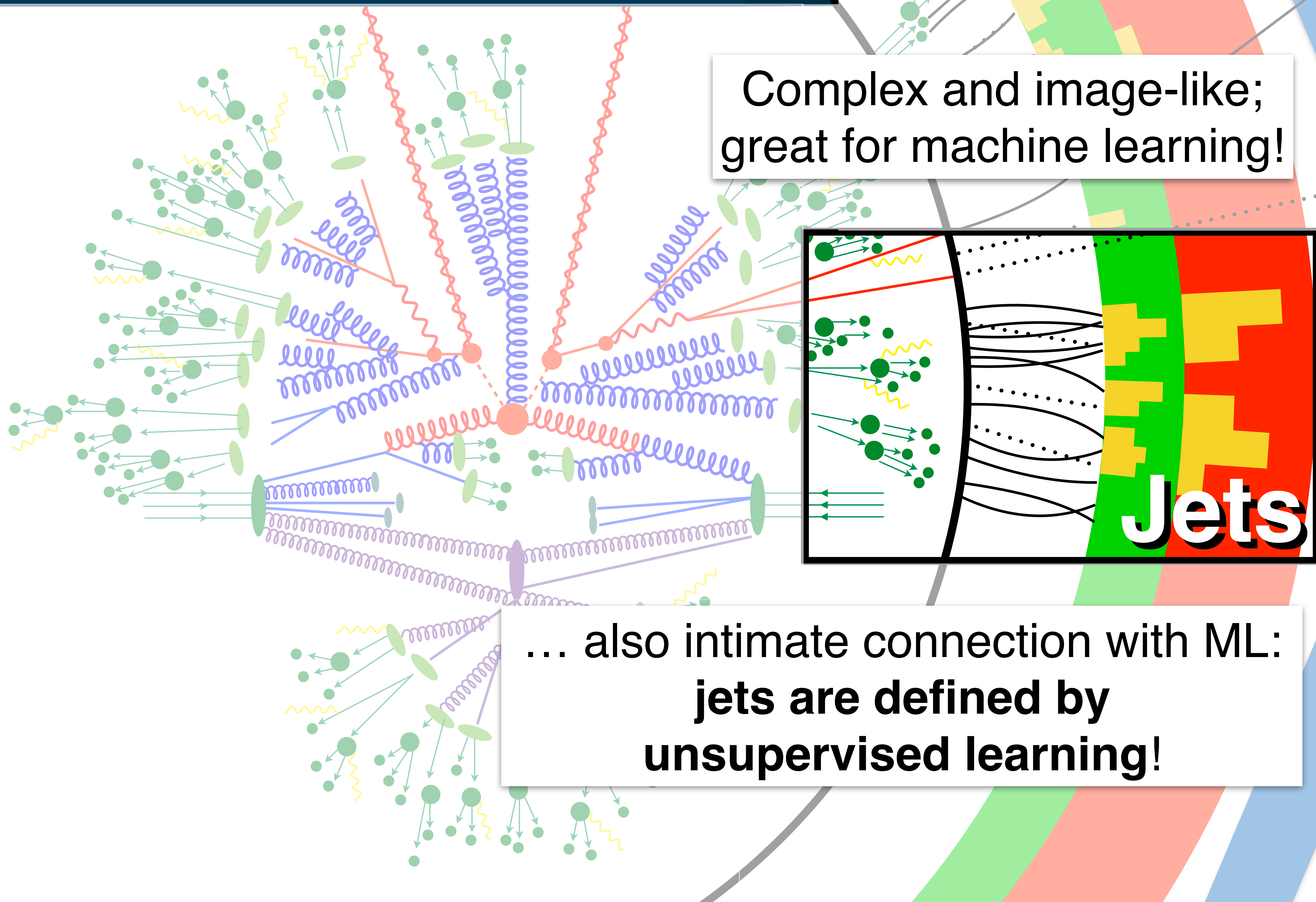
Quantum Chromodynamics (QCD)



No, I did not steal this from the Sherpa folks … can you spot the differences (other than the detector!)

Quantum Chromodynamics (QCD)

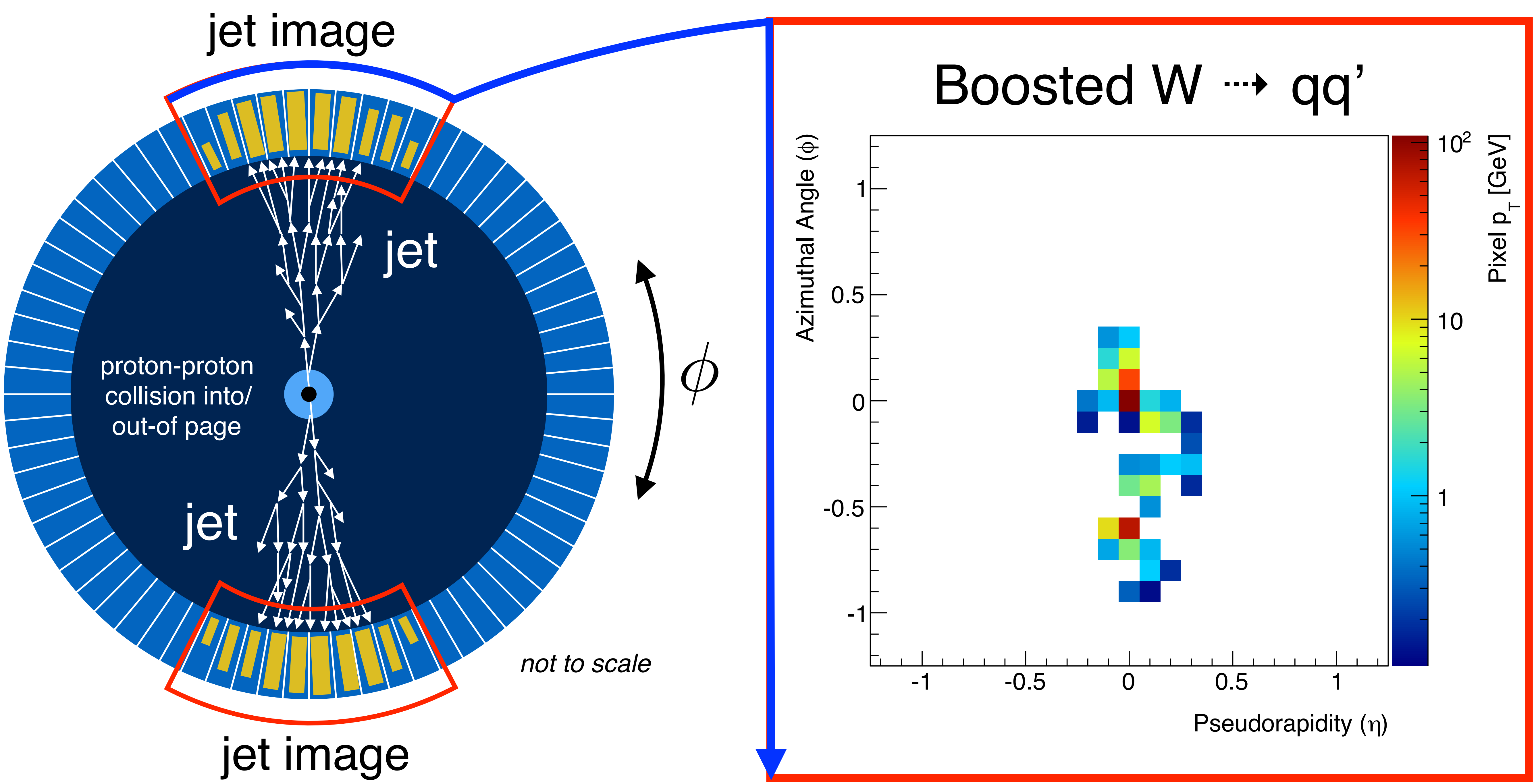
Complex and image-like;
great for machine learning!



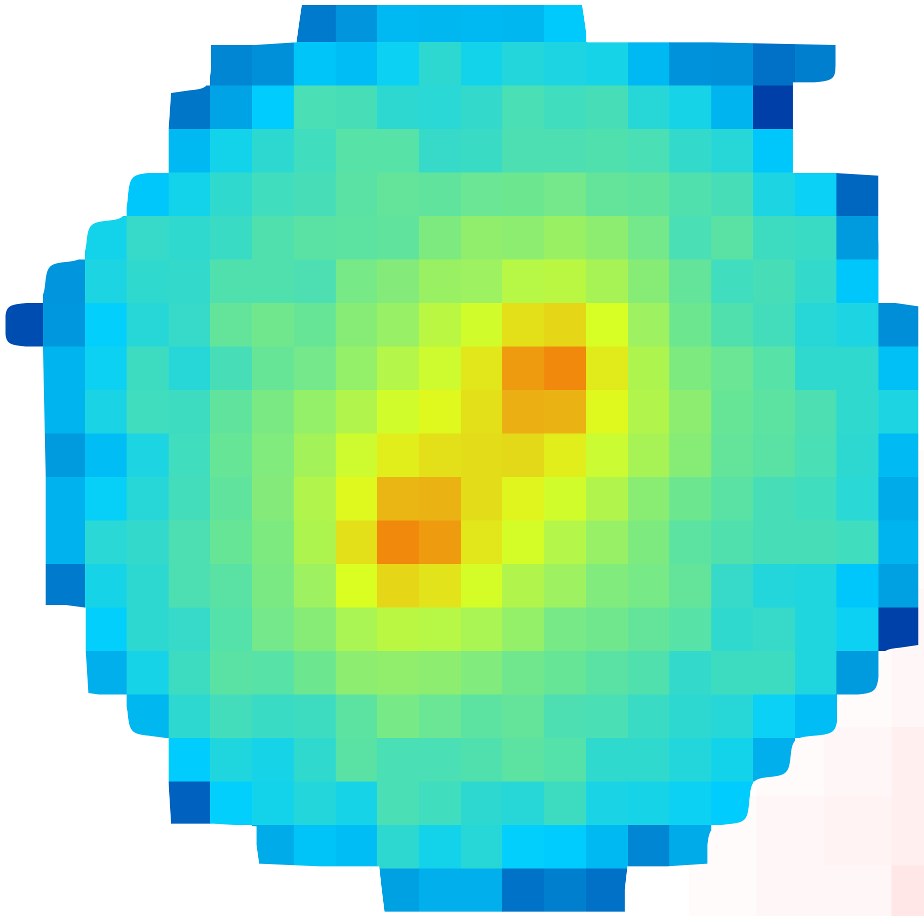
… also intimate connection with ML:
jets are defined by
unsupervised learning!

The Jet Image

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

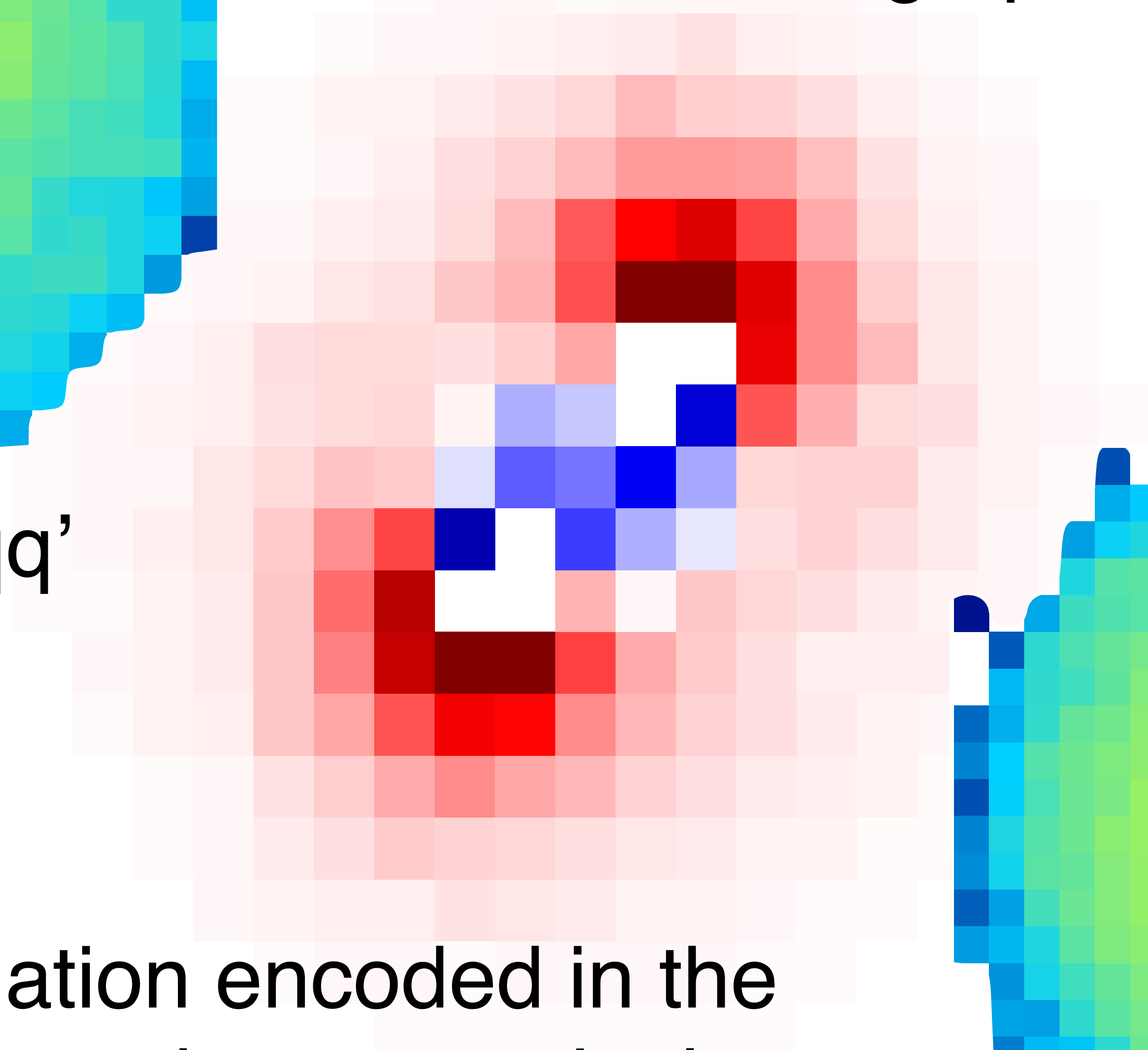


Why images?

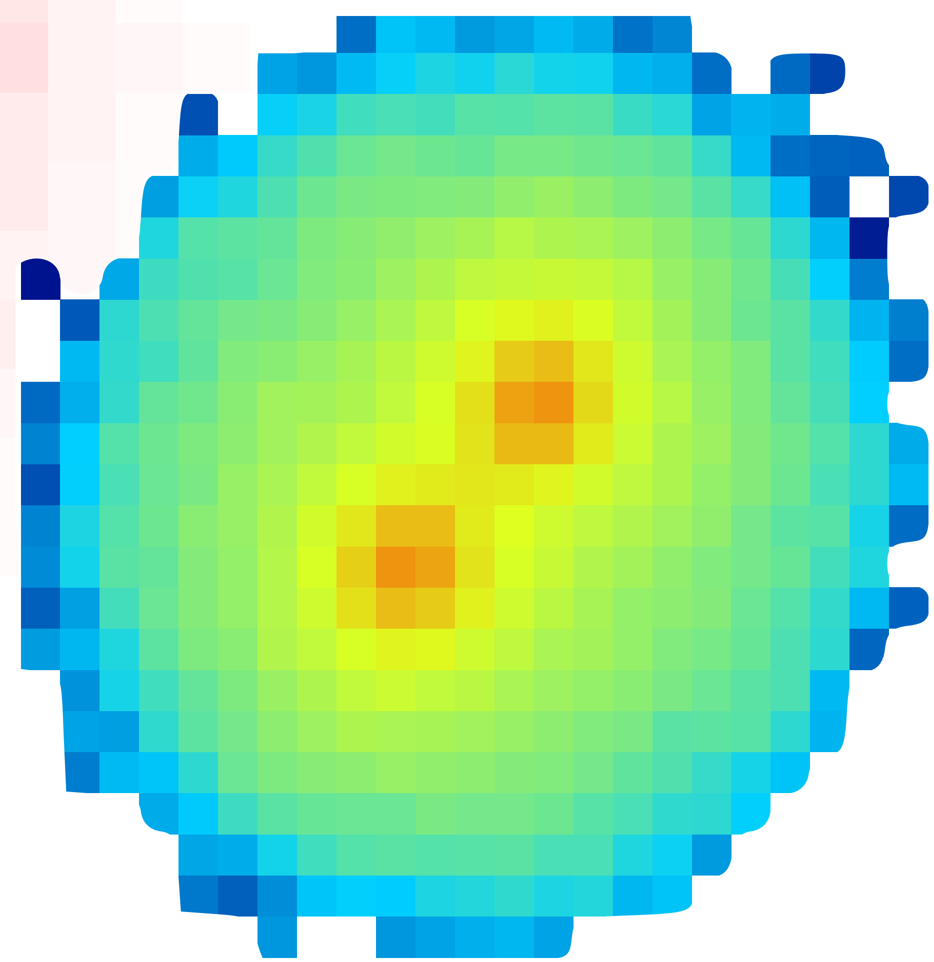


Can directly visualize physics

and we can benefit from the extensive image processing literature



octet ⇢ qq’



singlet ⇢ qq’

there is information encoded in the physical distance between pixels

(will mention other fixed representations later)

Pre-processing and Special Relativity

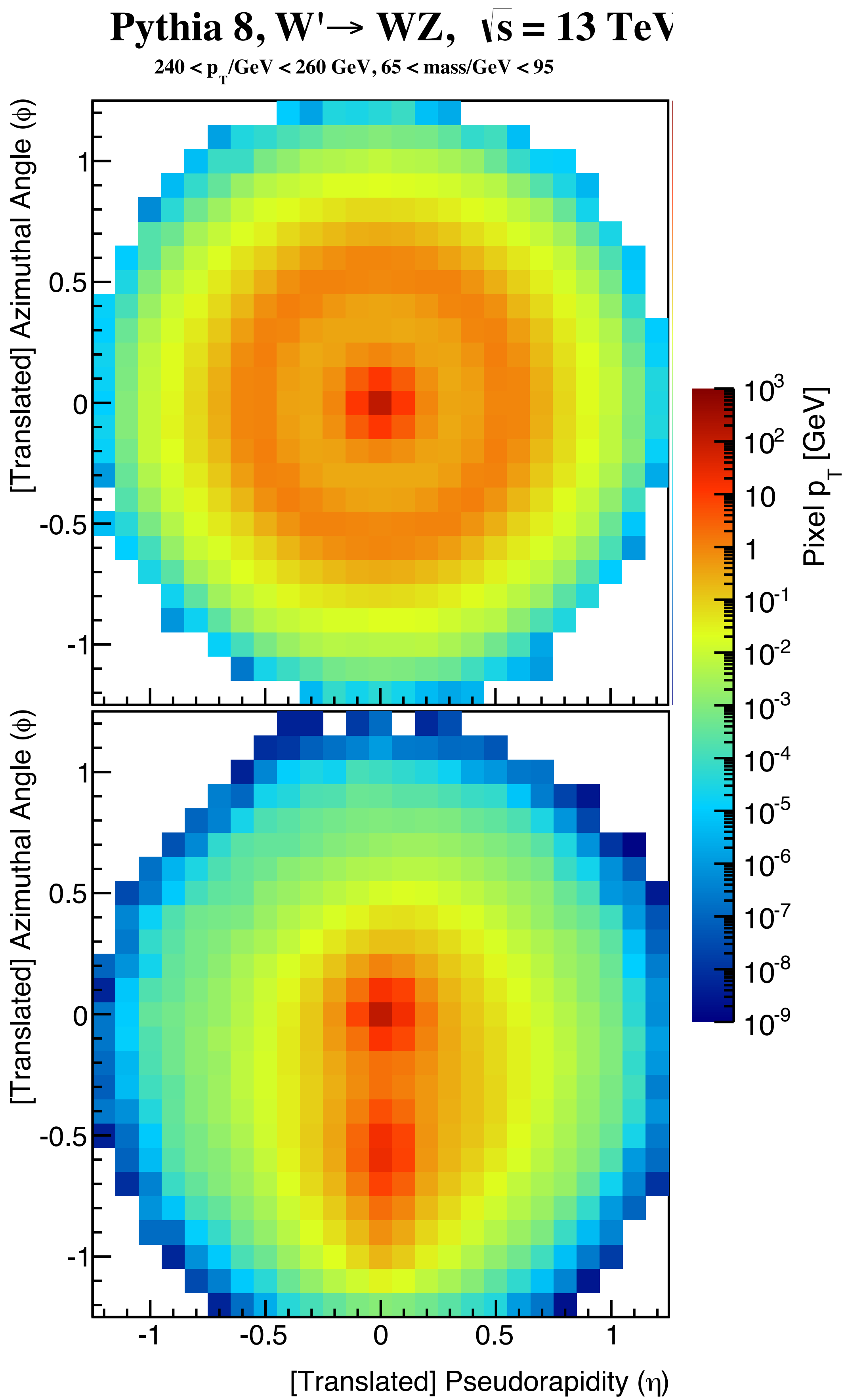
Pre-processing is an important aspect of image recognition

*we can inject **domain knowledge***

However, some steps can ***damage*** *the physics information content* of a jet image

I'll briefly illustrate some of these challenges in the next slides

there are also non-standard ideas such as "Zooming" - see *J. Barnard et al.* 1609.00607

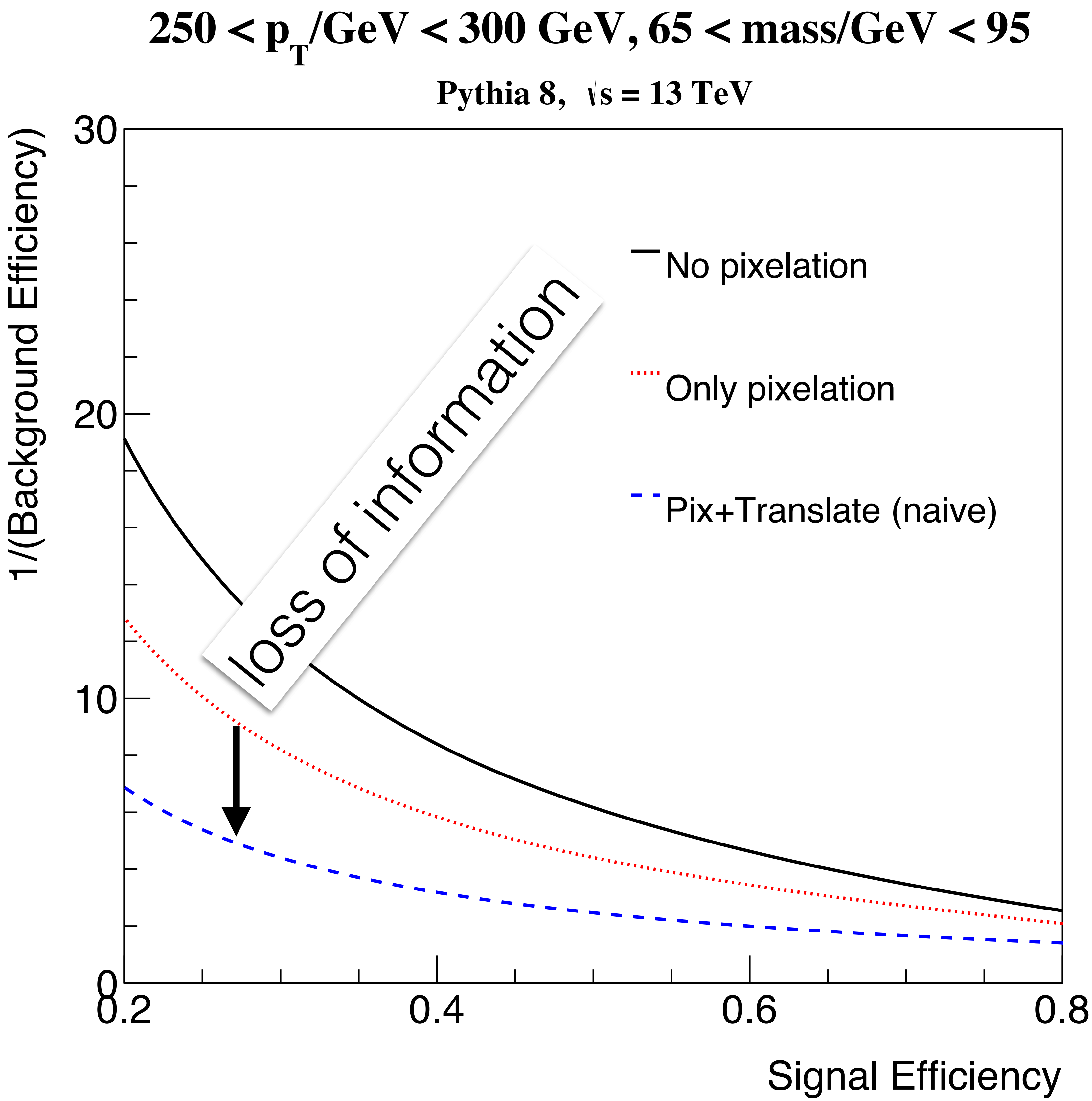


Pre-processing: E versus p_T

For calorimeter images, it is natural to think about energy as intensity.

However, **centering the image** in η corresponds to **boosting along z**!

Therefore, it is very important to use **p_T and not E**.

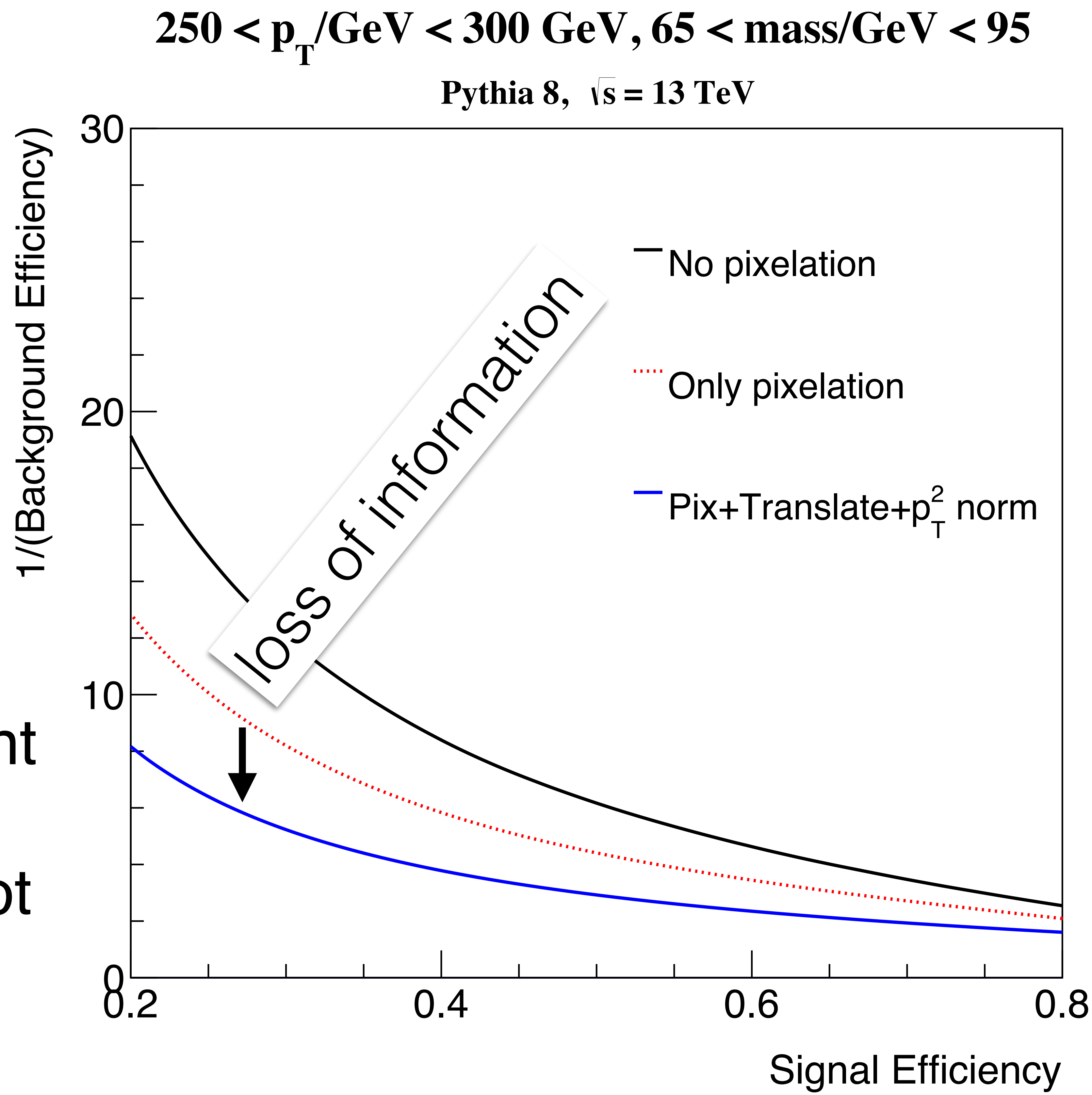


Pre-processing: Normalization

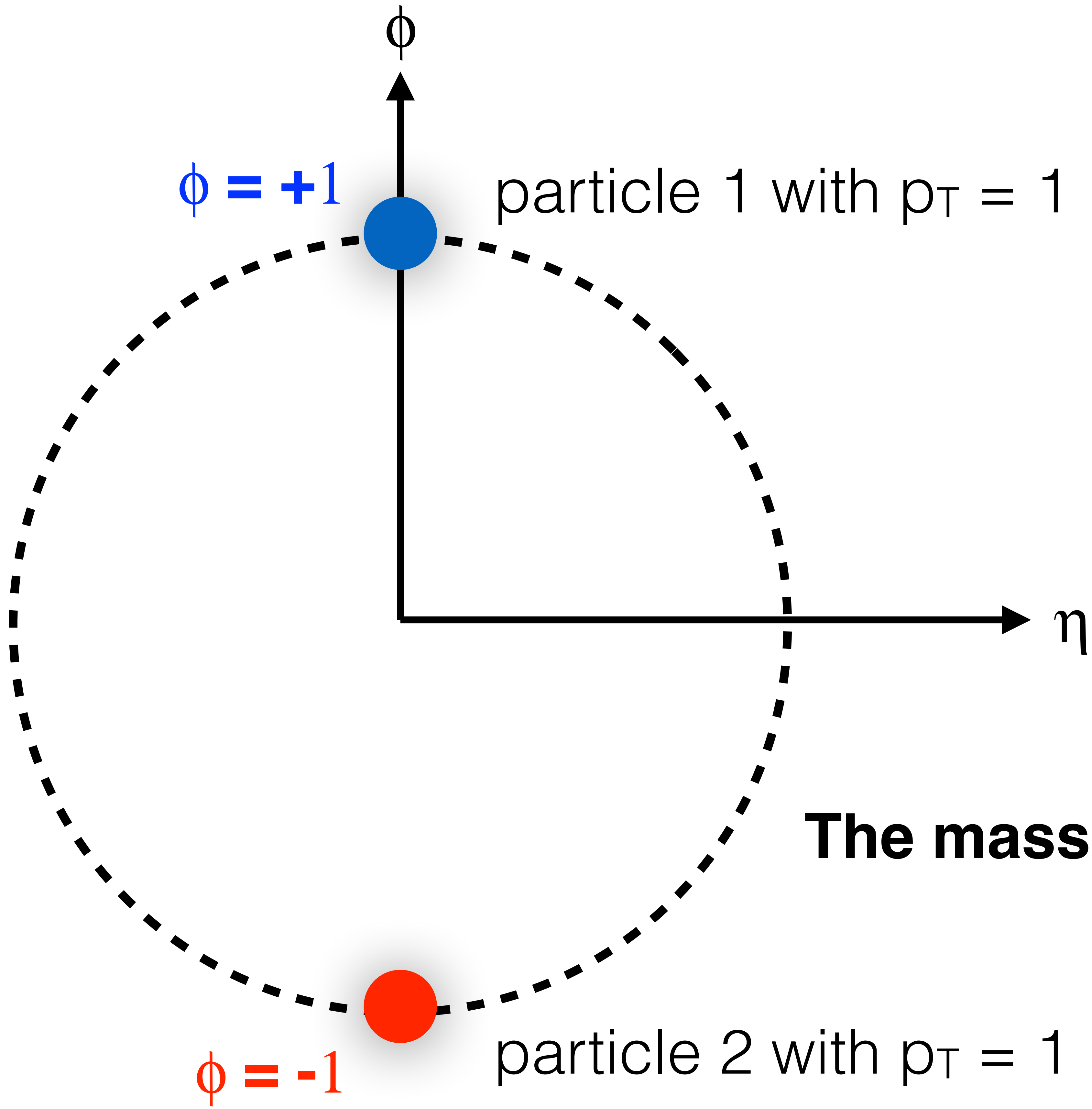
Similar story for image normalization.

As with E instead of p_T, this adds in ~random noise to e.g. the mass

Therefore, it is important to do **ensemble-normalizations** and not **jet-by-jet norms**.

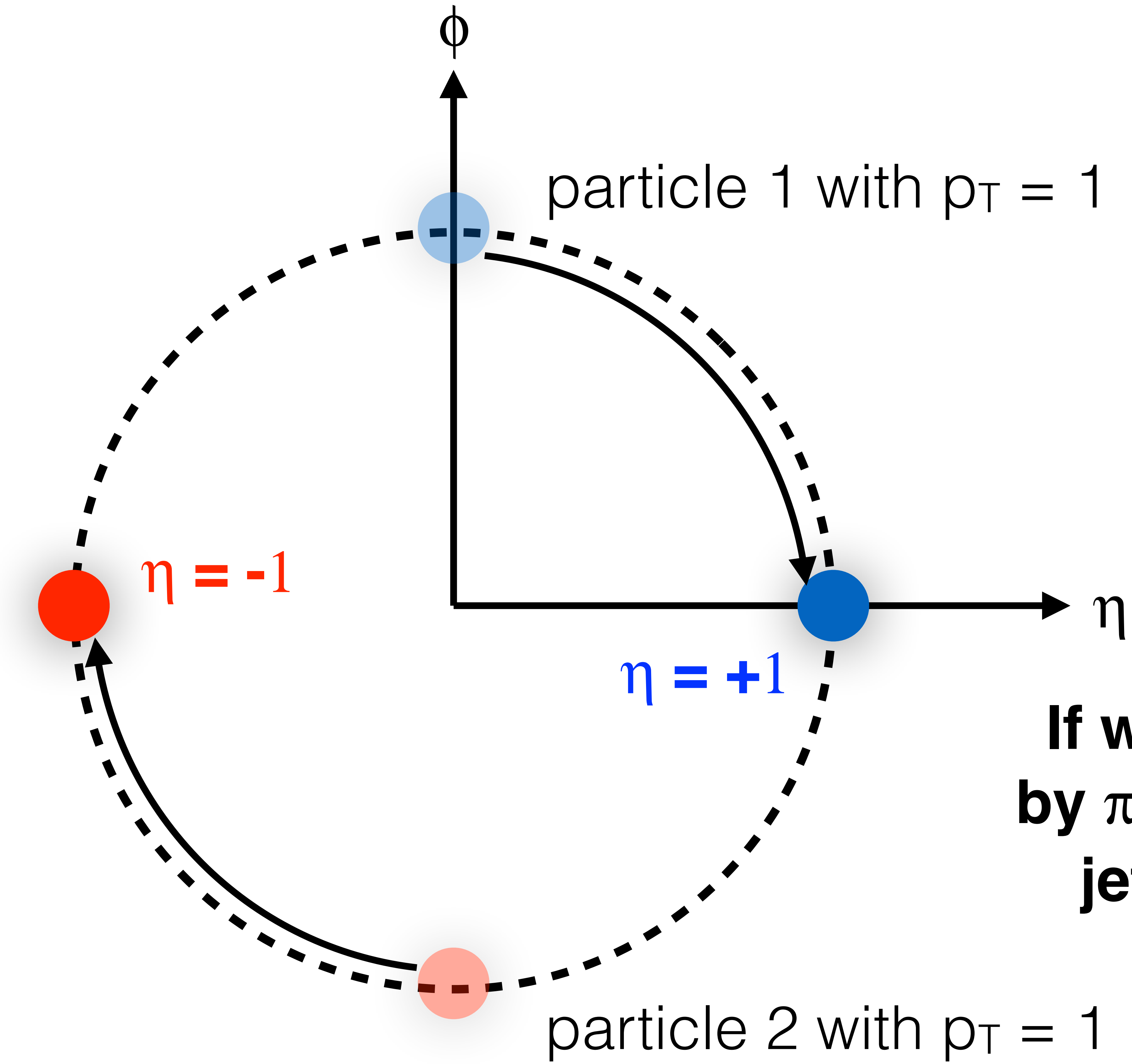


Pre-processing: Rotations



The mass of this 'jet' is ~1.7

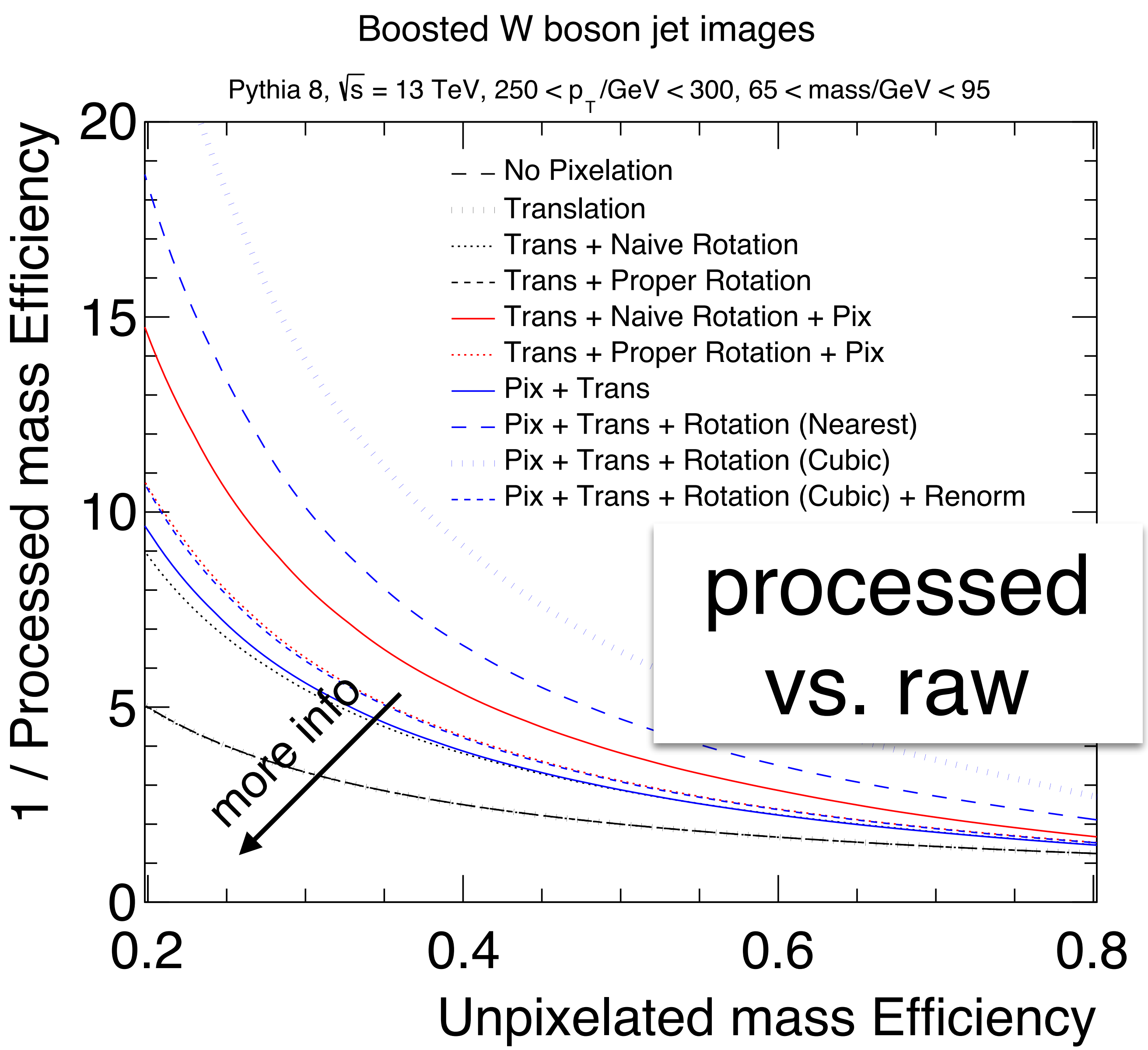
Pre-processing: Rotations



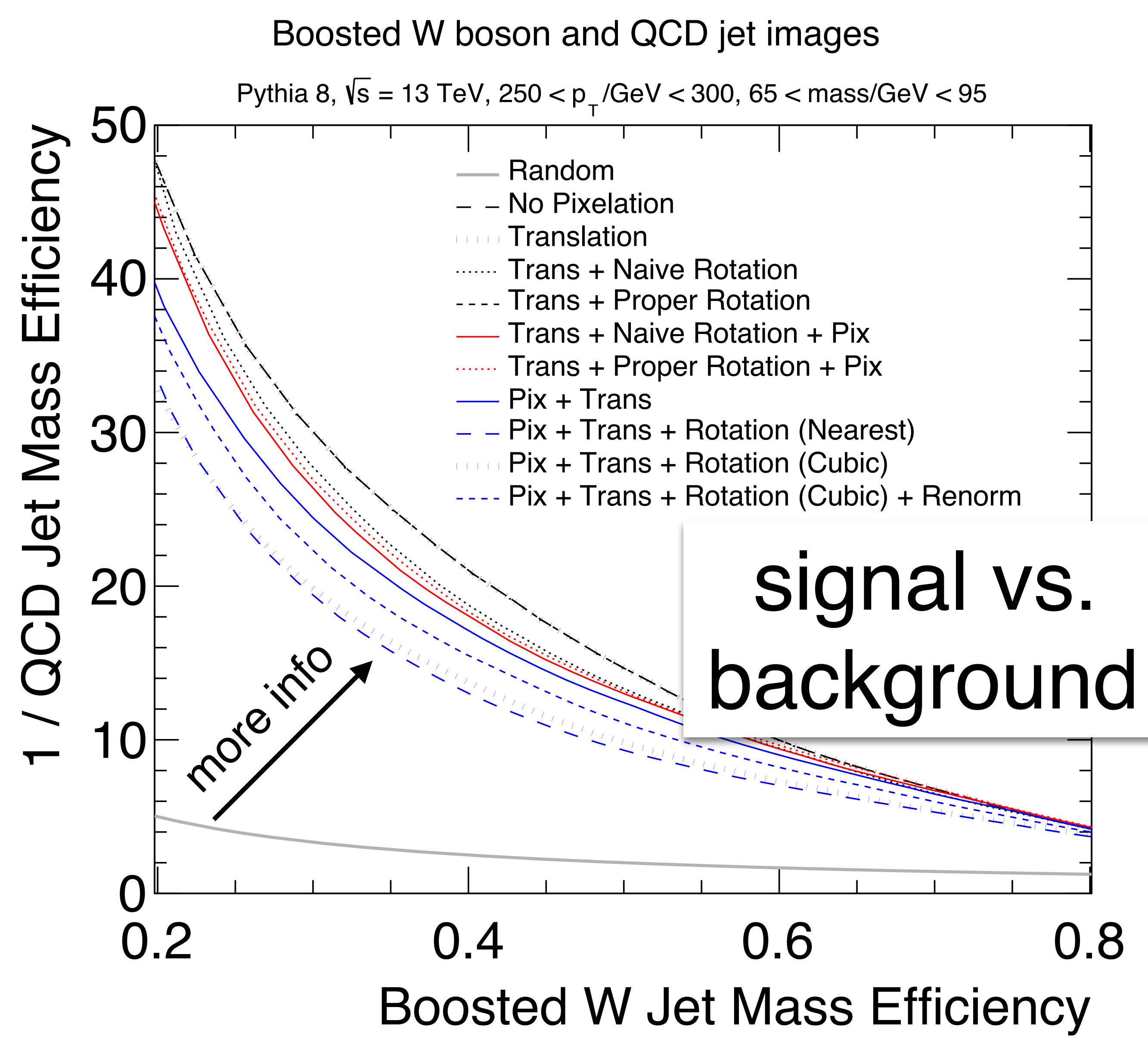
If we rotate the jet by $\pi/2$, then the new jet mass is ~2.4

Pre-processing: Rotations (cont.)

de Oliviera et al. (**BN**) arXiv:1701.05927



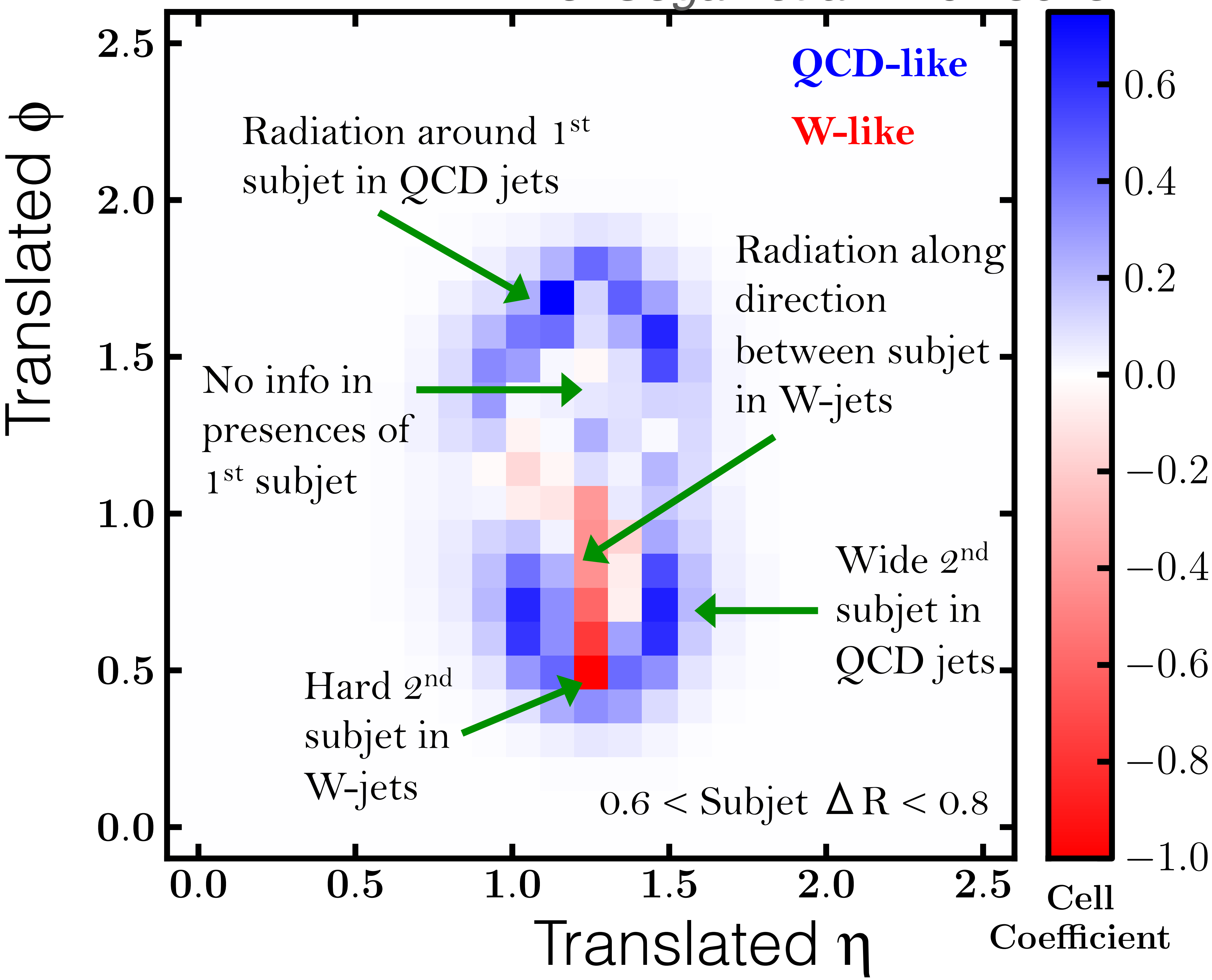
Can do a 'proper' rotation that preserves mass, but changes e.g. τ_{21}.



Can add in information to "undo" rotation or augment the dataset with "ghost images" (1612.01551)

Linear ML Methods for Classification

J. Cogan et al. 1407.5675



"Fisher Jet"

Build discriminant from projecting onto this image

Directly interpretable!

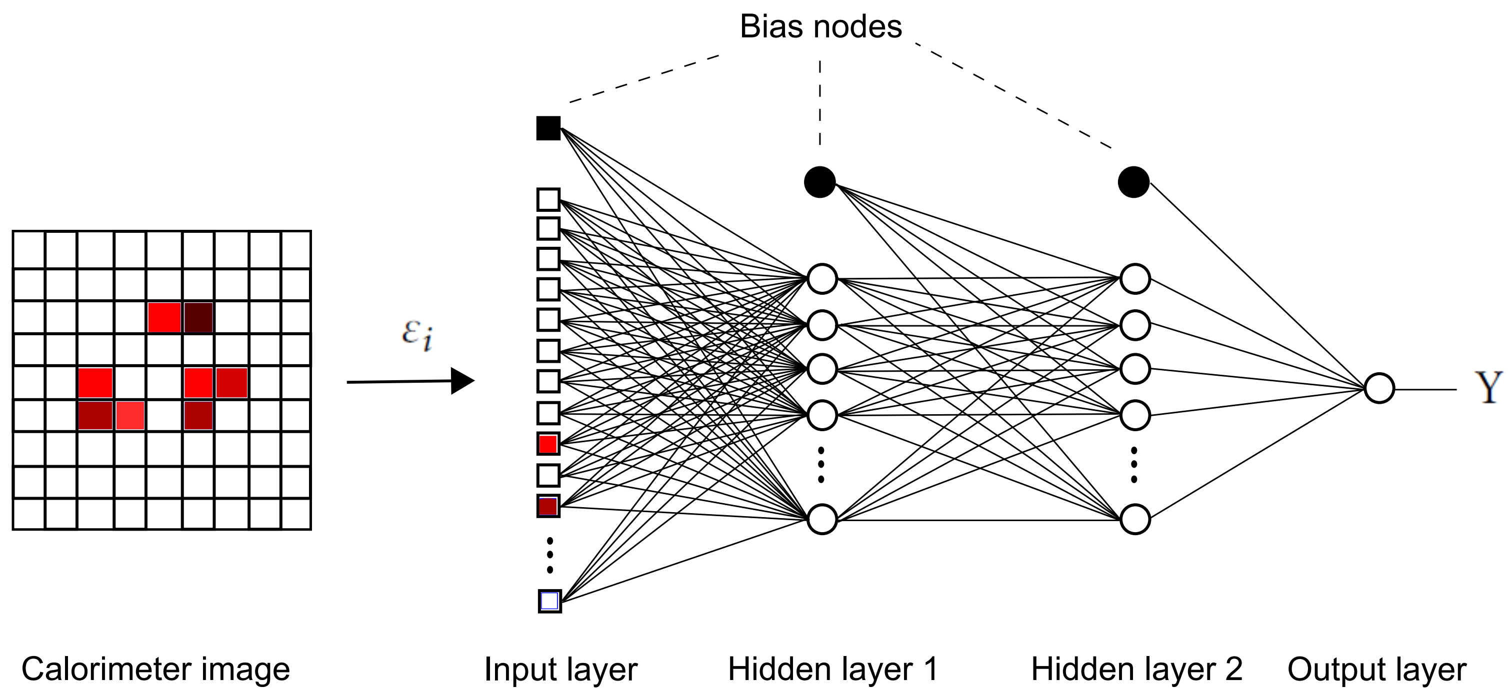
Add in a small non-linearity by binning in ΔR

(eyes on a face closer when further away!)

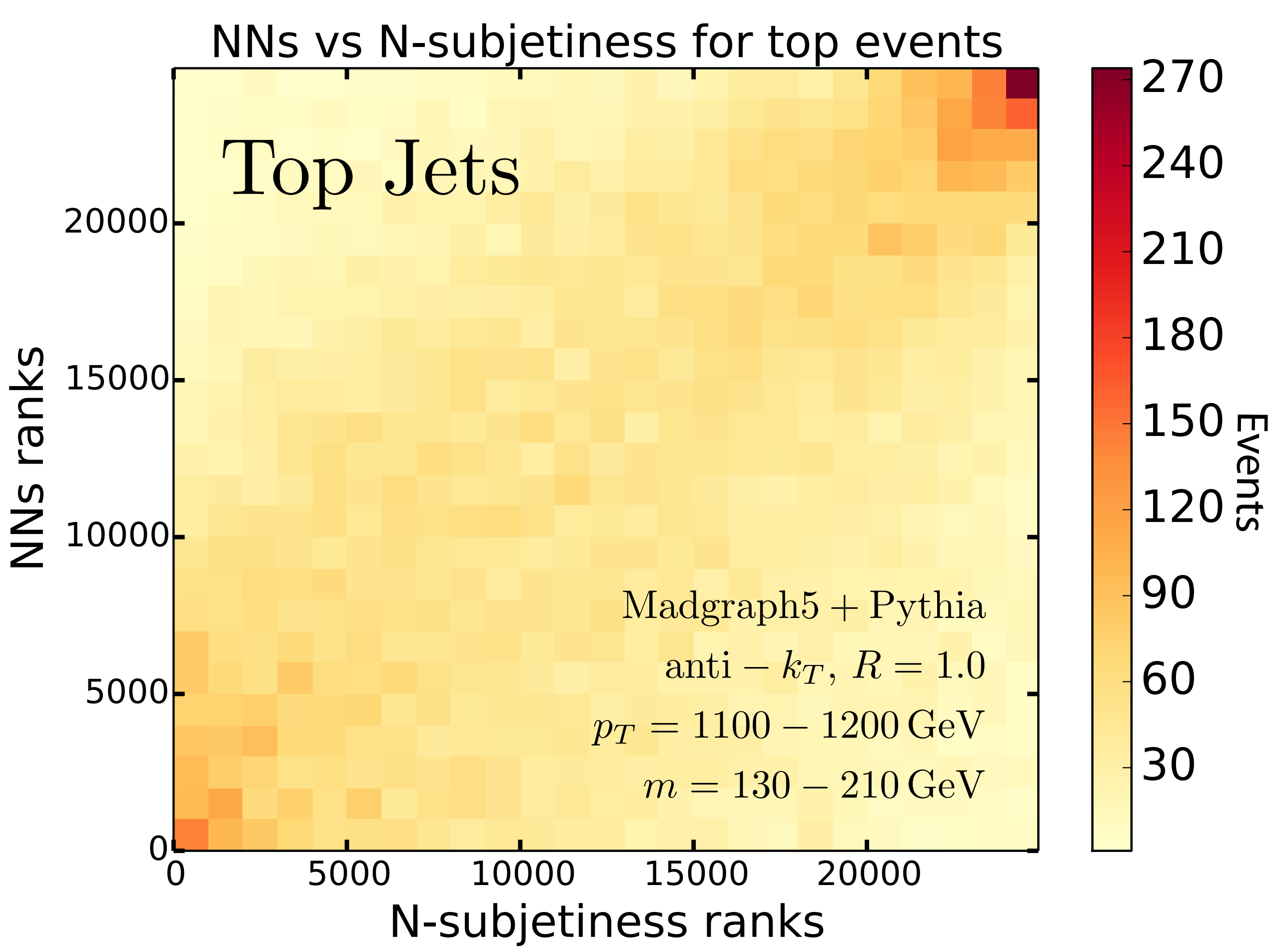
Maximize between class versus within class variance

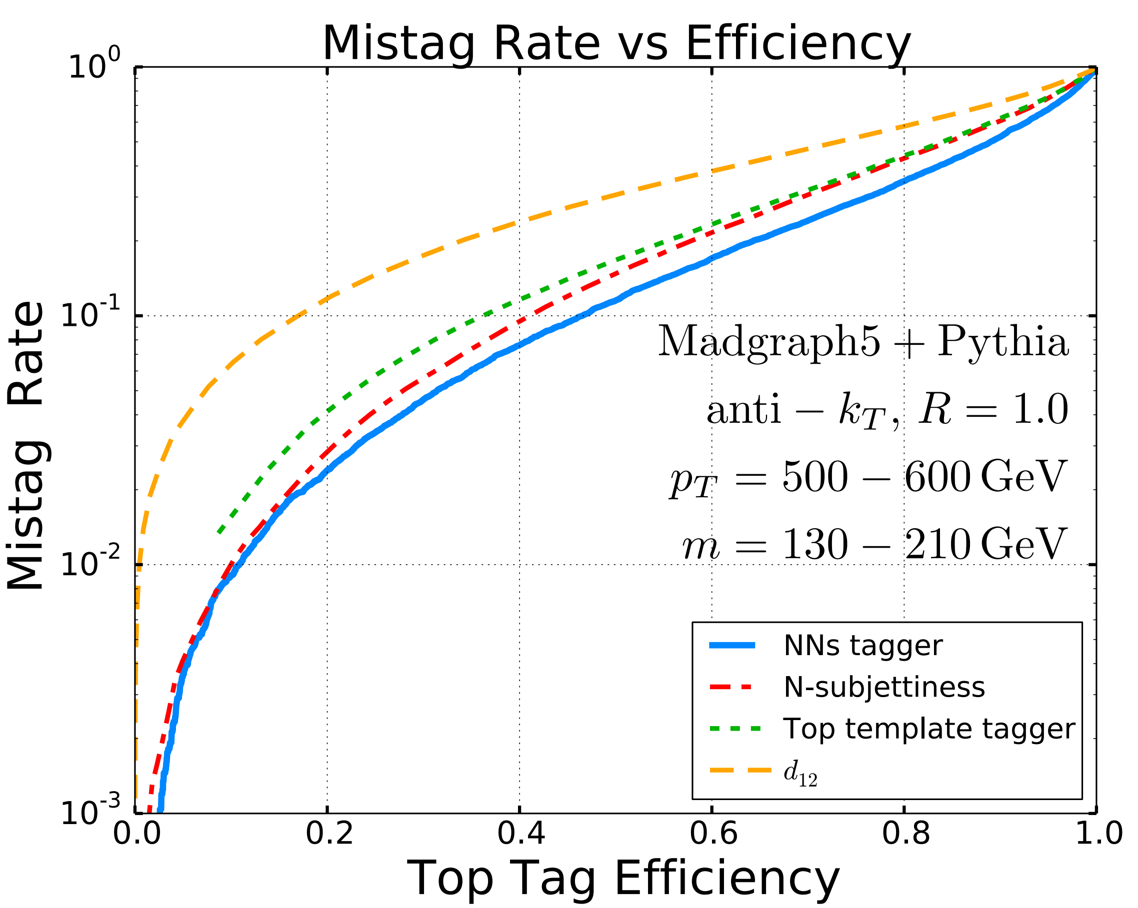
Shallow NN's for Classification

L. Almeida et al. 1501.05968



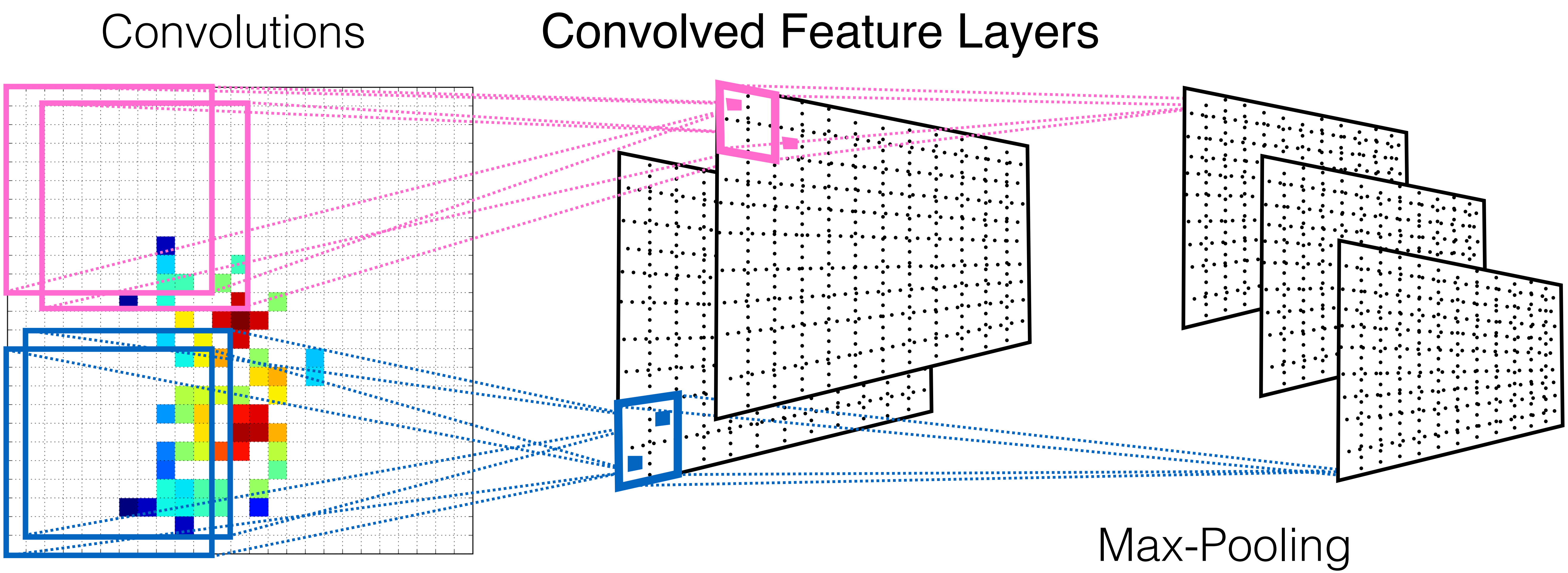
First application of the jet images idea using (shallow) NN's with top-quark tagging





Modern Deep NN's for Classification

de Oliviera et al. (***BN***) 1511.05190



W'→ *WZ* event

Subsequent developments:

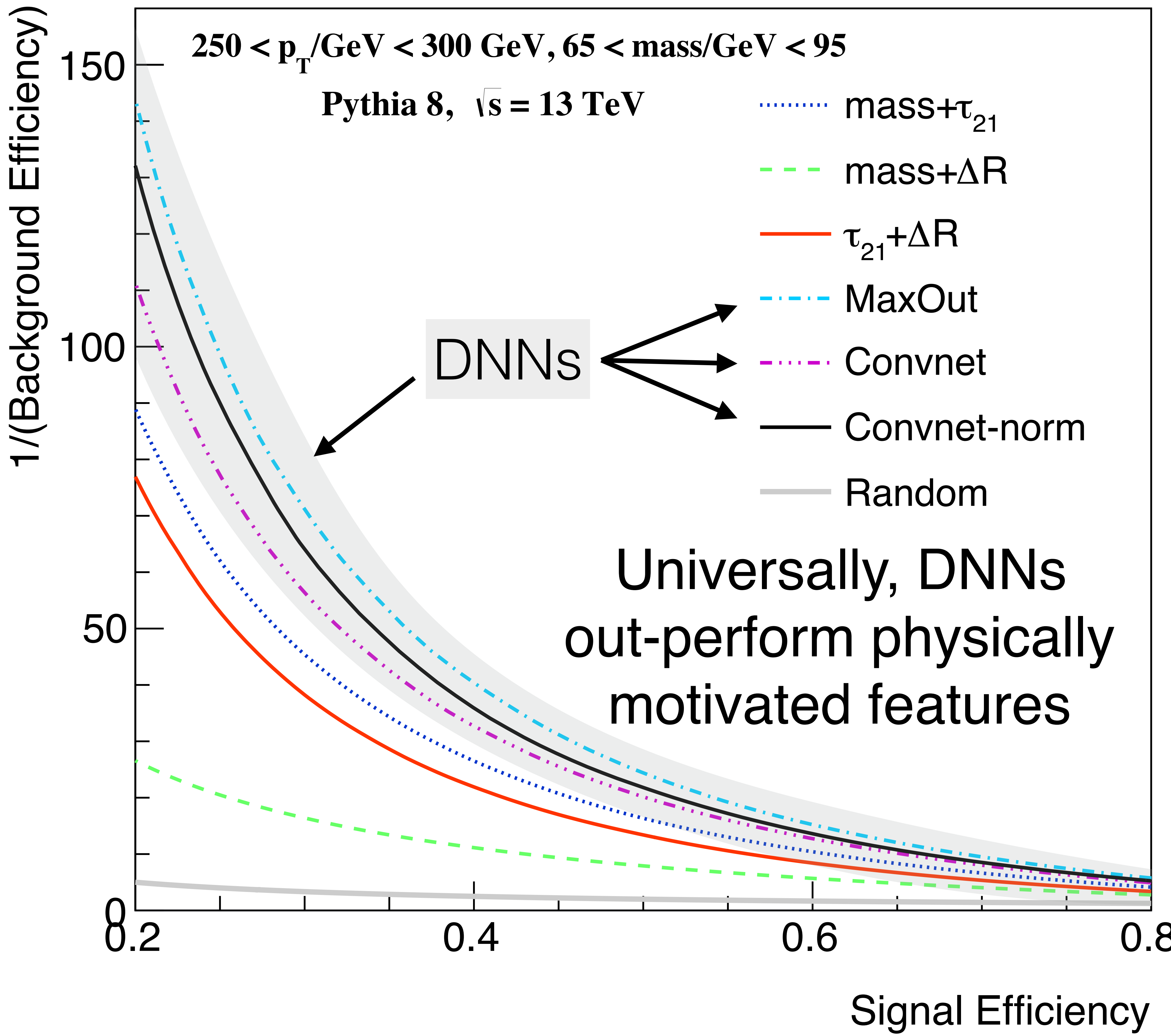
P. Baldi et al. 1603.09349 (W-tagging)

J. Barnard et al. 1609.00607 (W-tagging)

P. Komiske et al. 1612.01551 (q/g-tagging)

G. Kasieczka et al. 1701.08784 (top-tagging)

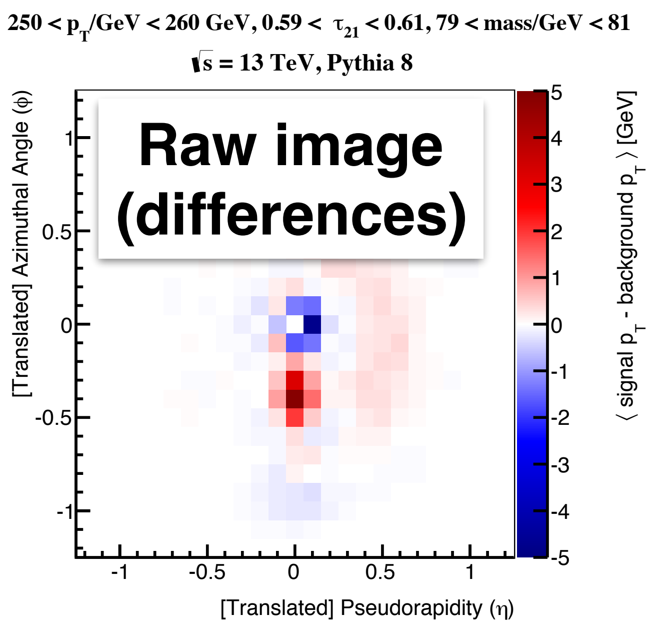
Modern Deep NN's for Classification

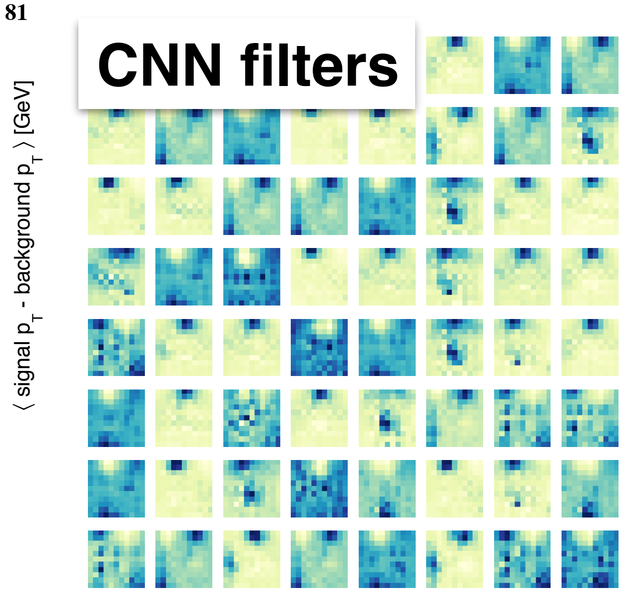


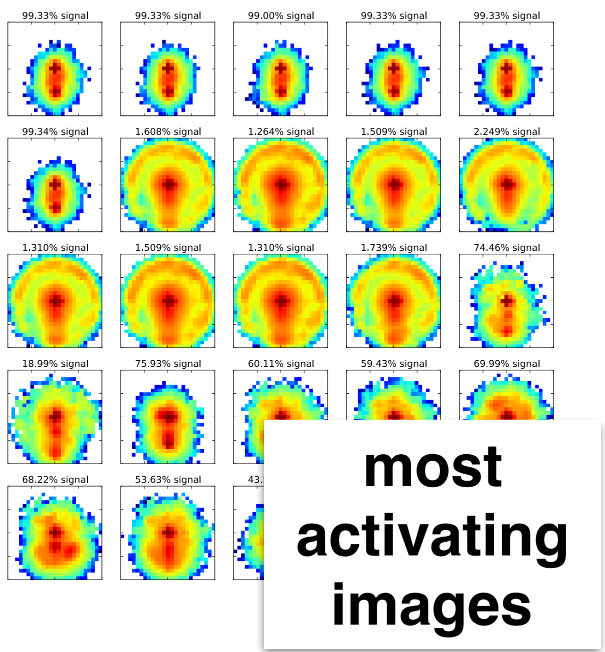
by how much varies by study; also no two groups have the same setup or metrics…

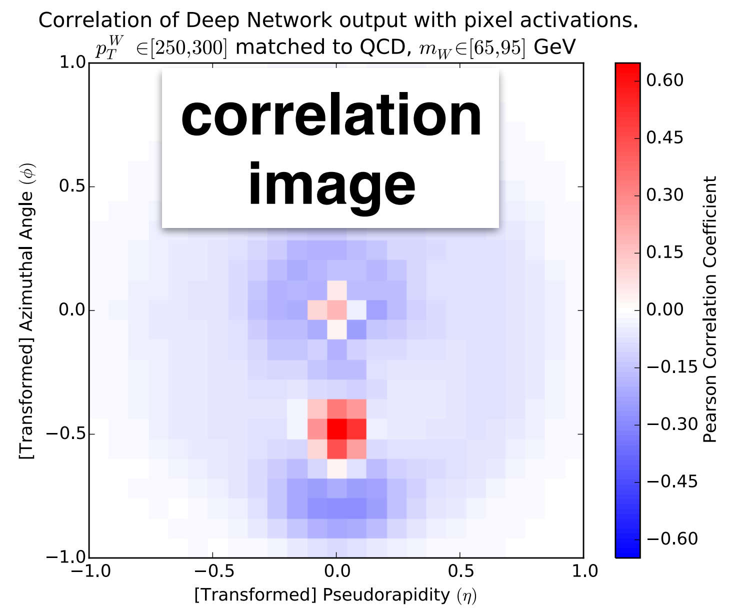
Learning about Learning

Jet images afford a lot of natural visualization

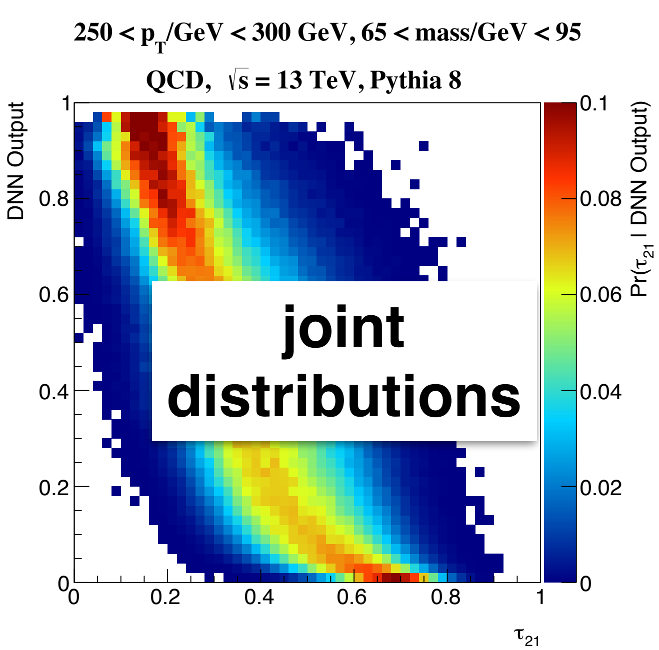


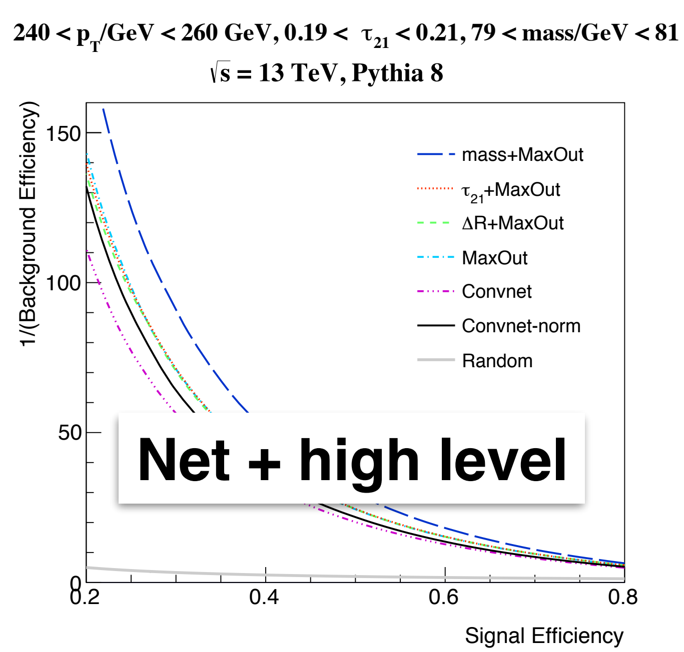


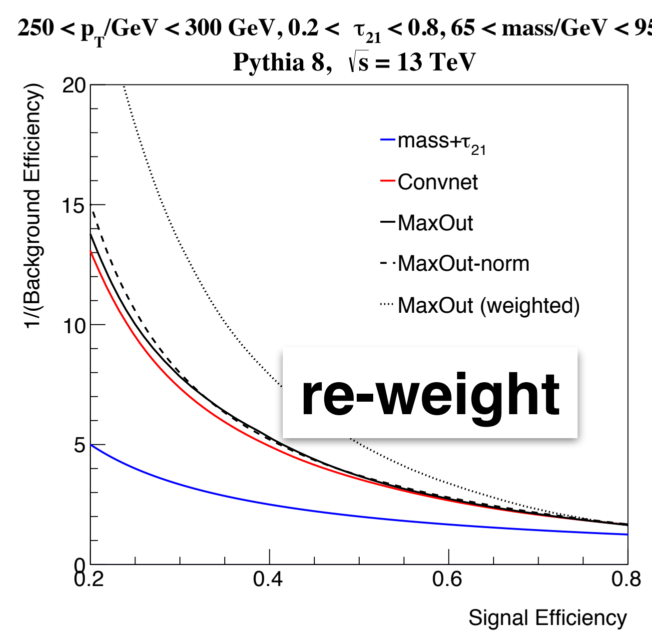


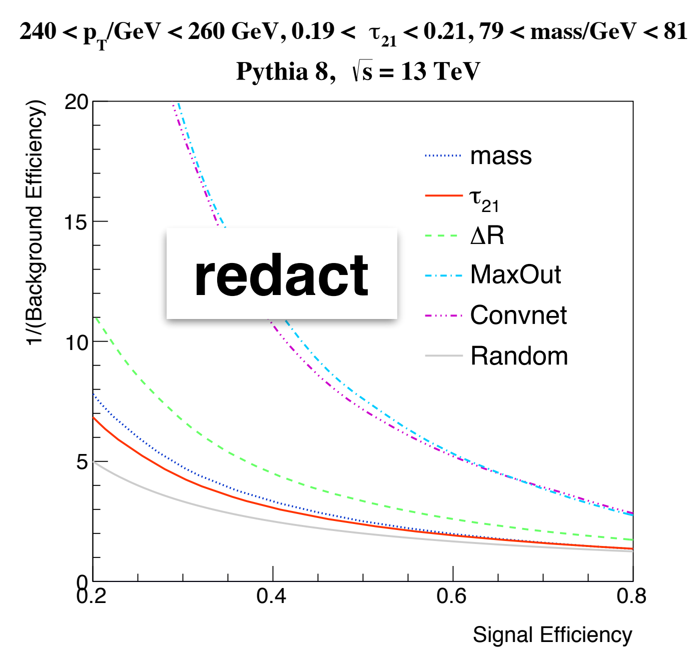


as a community, we have also developed many techniques





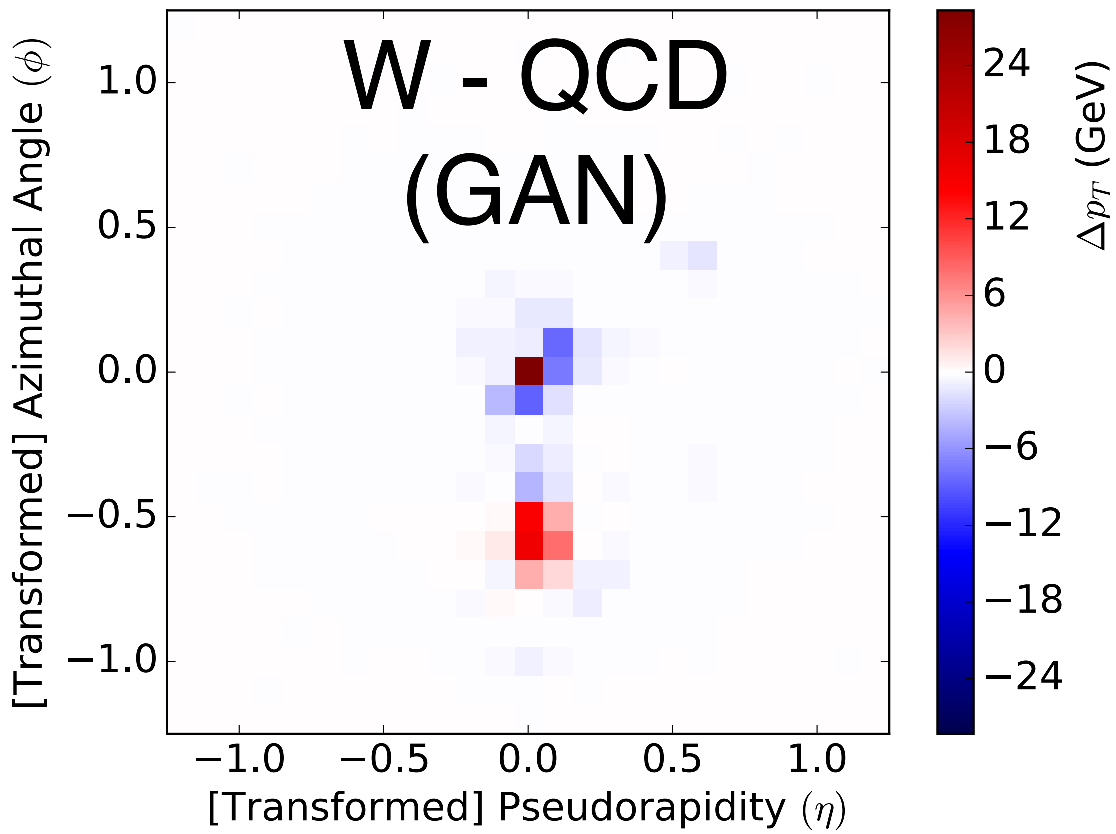


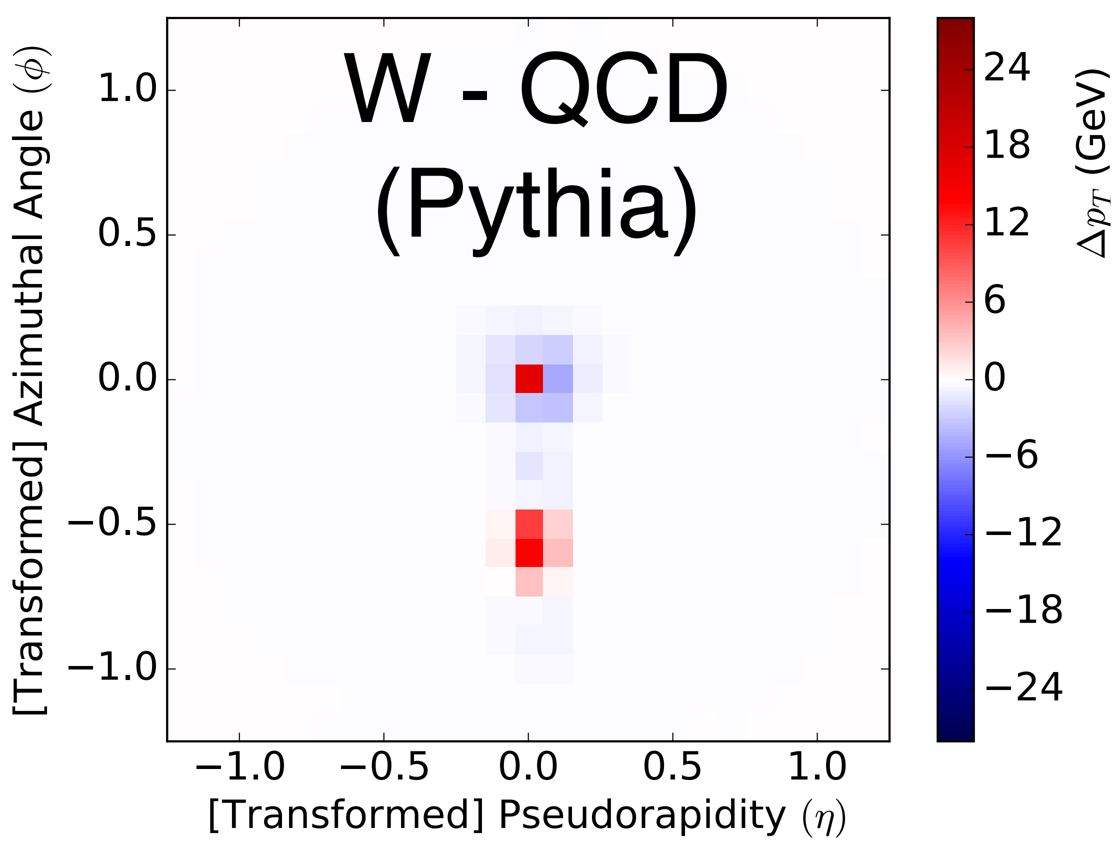


More detail in my DS@HEP15 talk

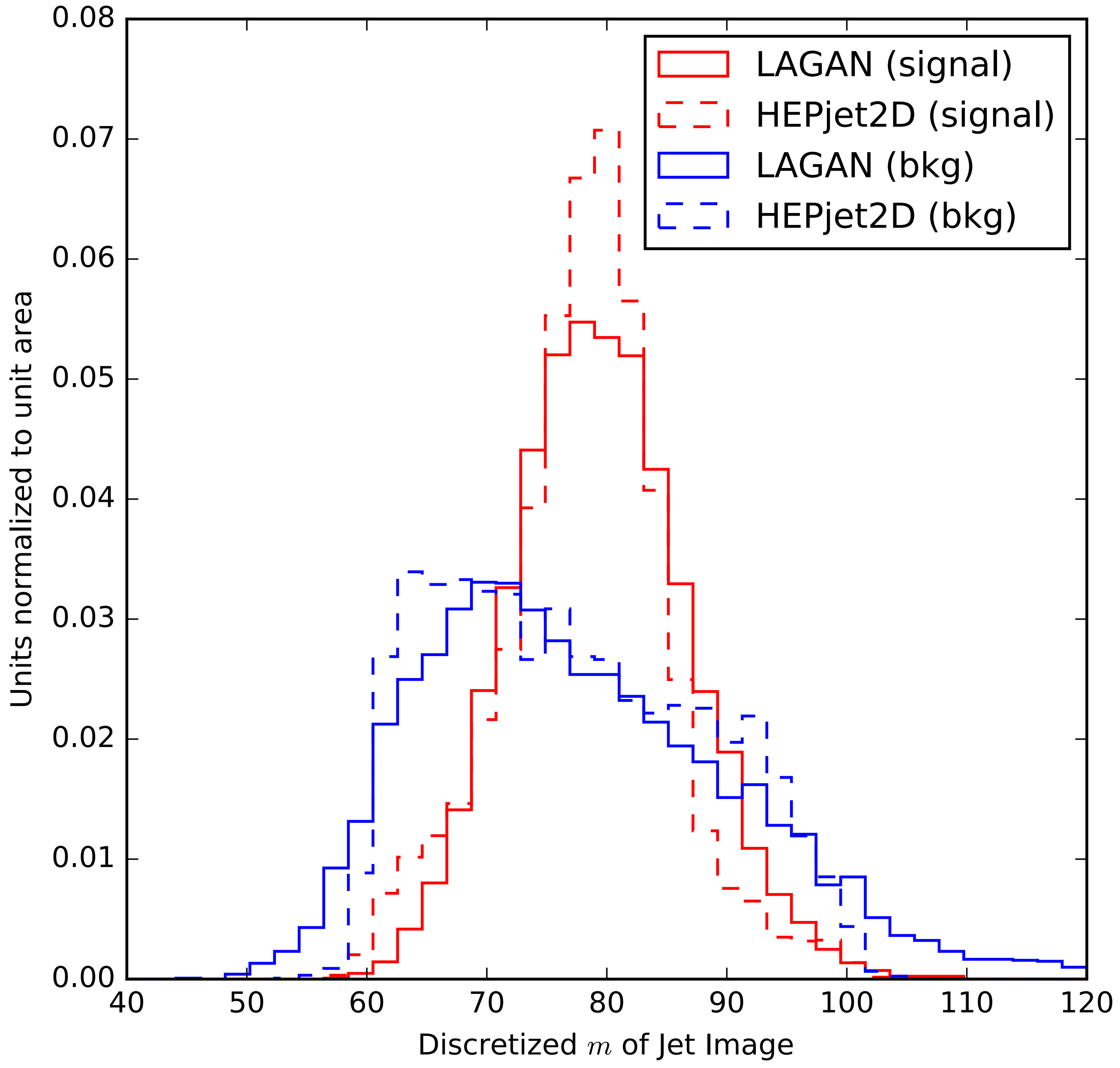
Modern Deep NN's for Generation

de Oliviera et al. (**BN**) arXiv:1701.05927



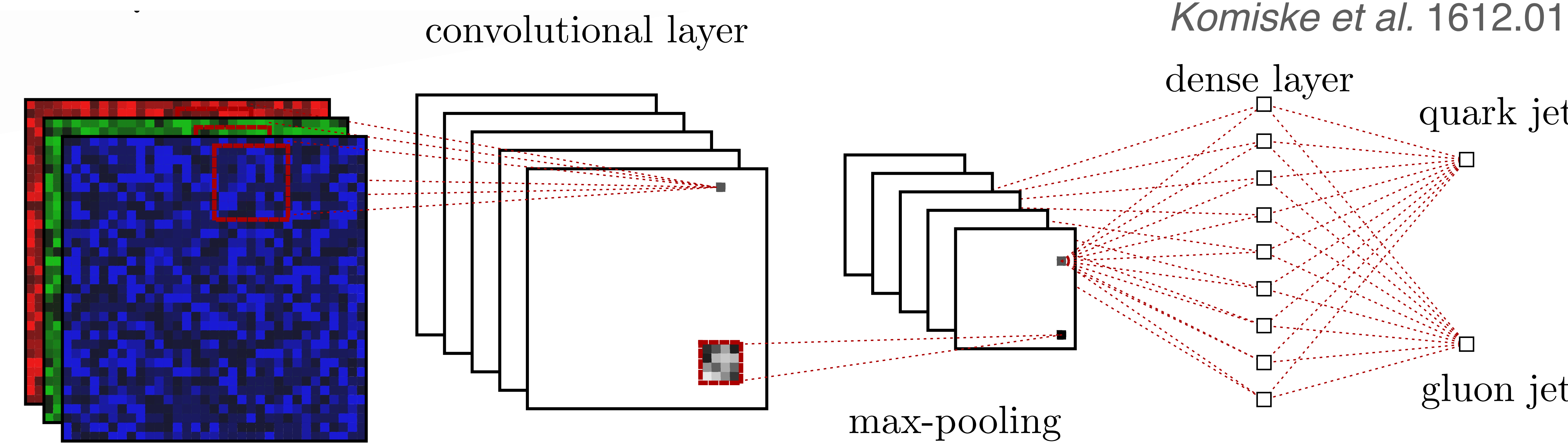


Generative Adversarial Networks (GANs)



+ More Layers for Classification

Komiske et al. 1612.01551

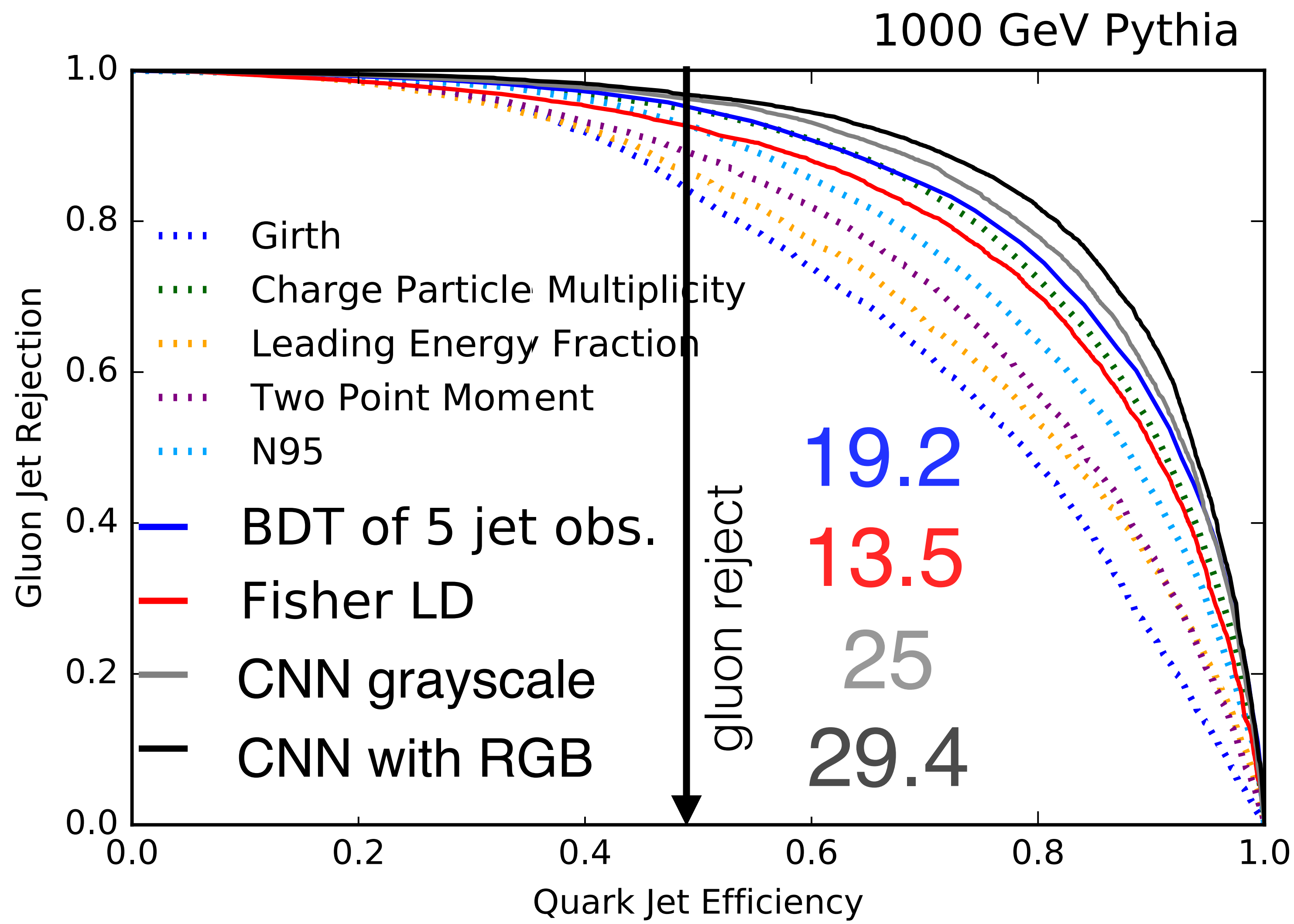


red = charged p_T

green = neutral p_T

blue = $n_{charged}$

Fixed granularity for the three layers: **linear** < **shallow** < **deep** < **deep**

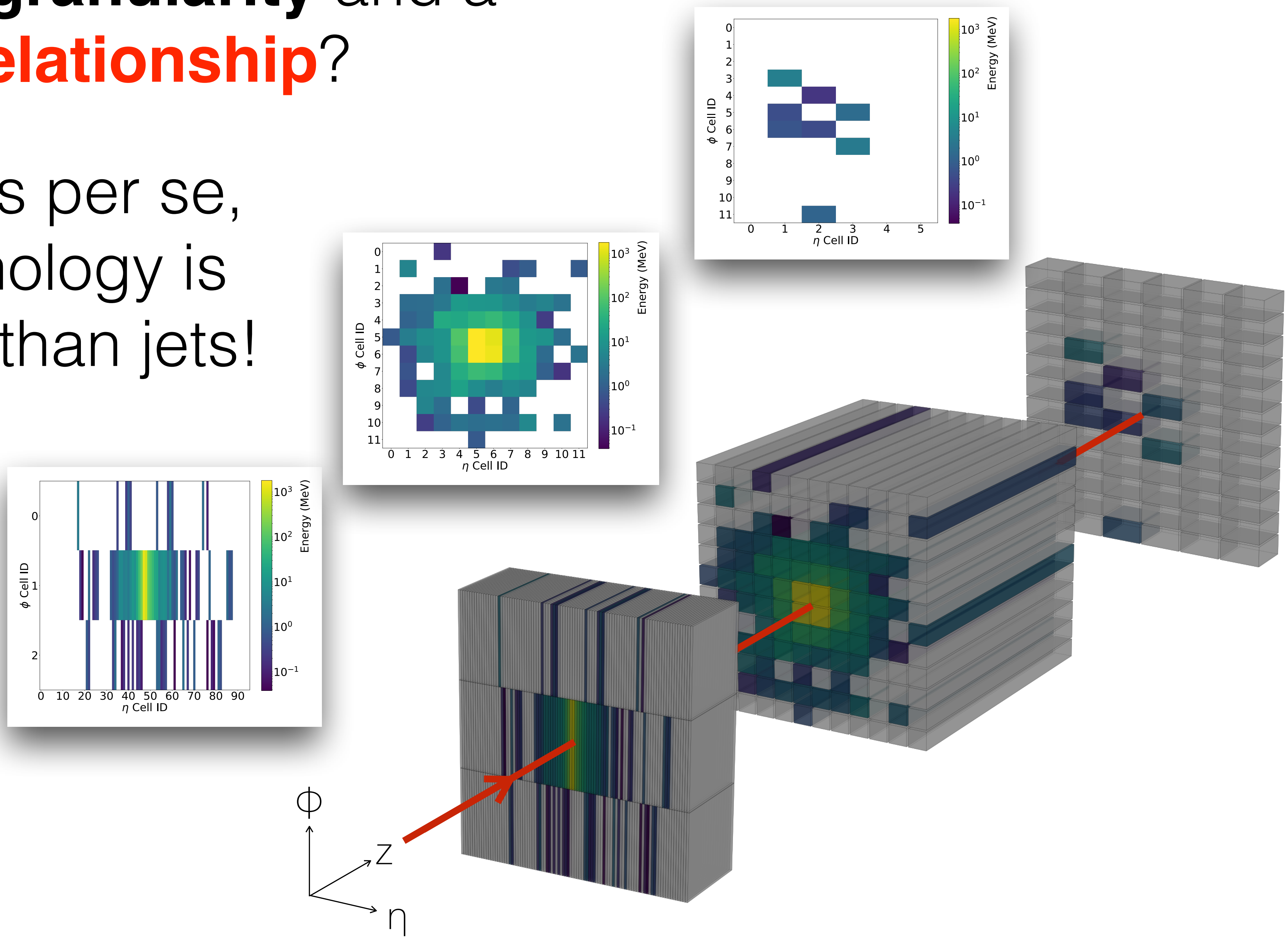


+ More Layers for Generation

Paganini et al. (**BN**) arXiv:1705.02355

What about **multiple layers** with **non-uniform granularity** and a **causal relationship**?

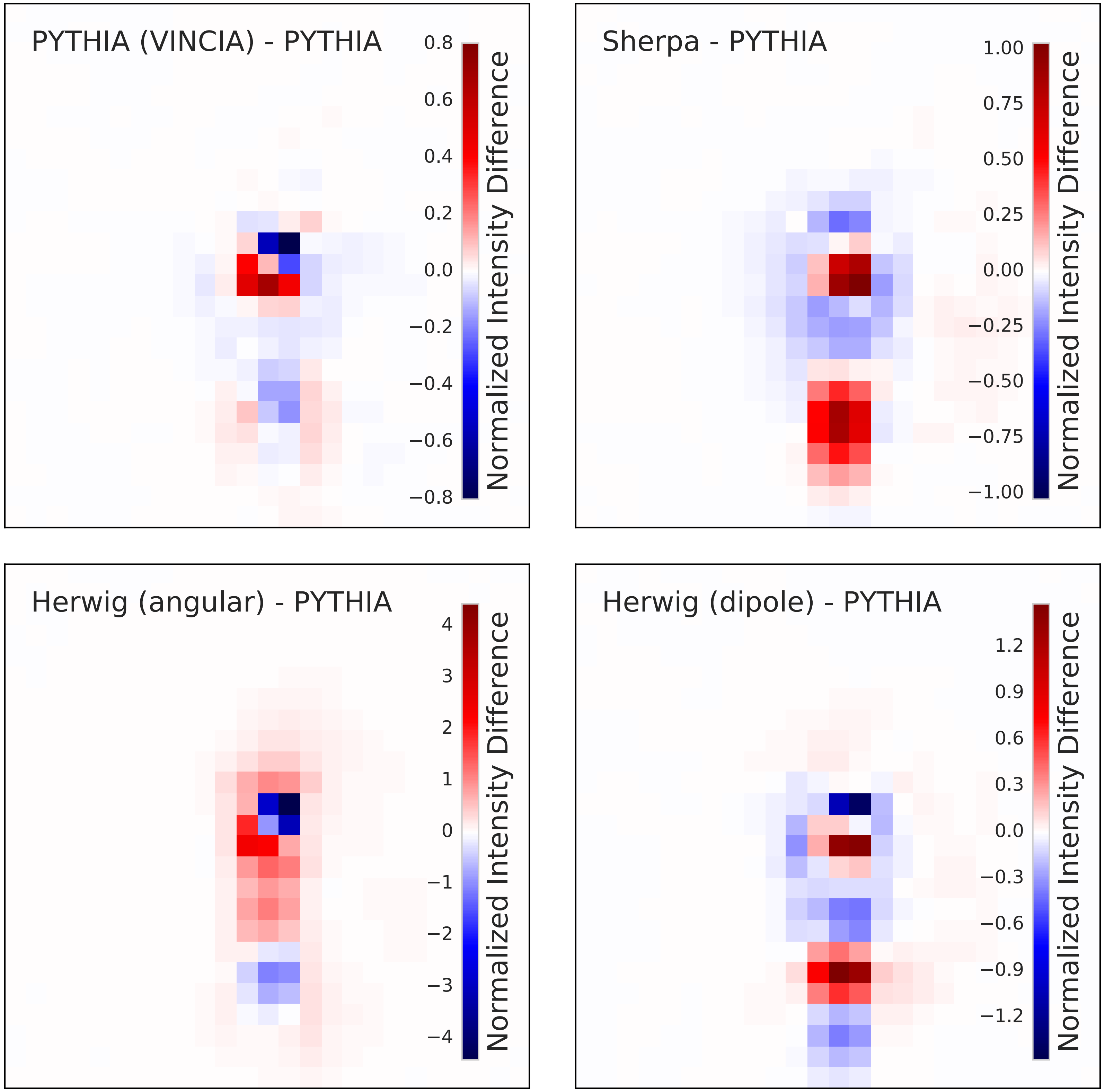
Not jet images per se, but the technology is more general than jets!



Sensitivity to Modeling

Boosted *W* boson jets

J. Barnard et al. 1609.00607



N.B. not all of these have been tuned to the same data

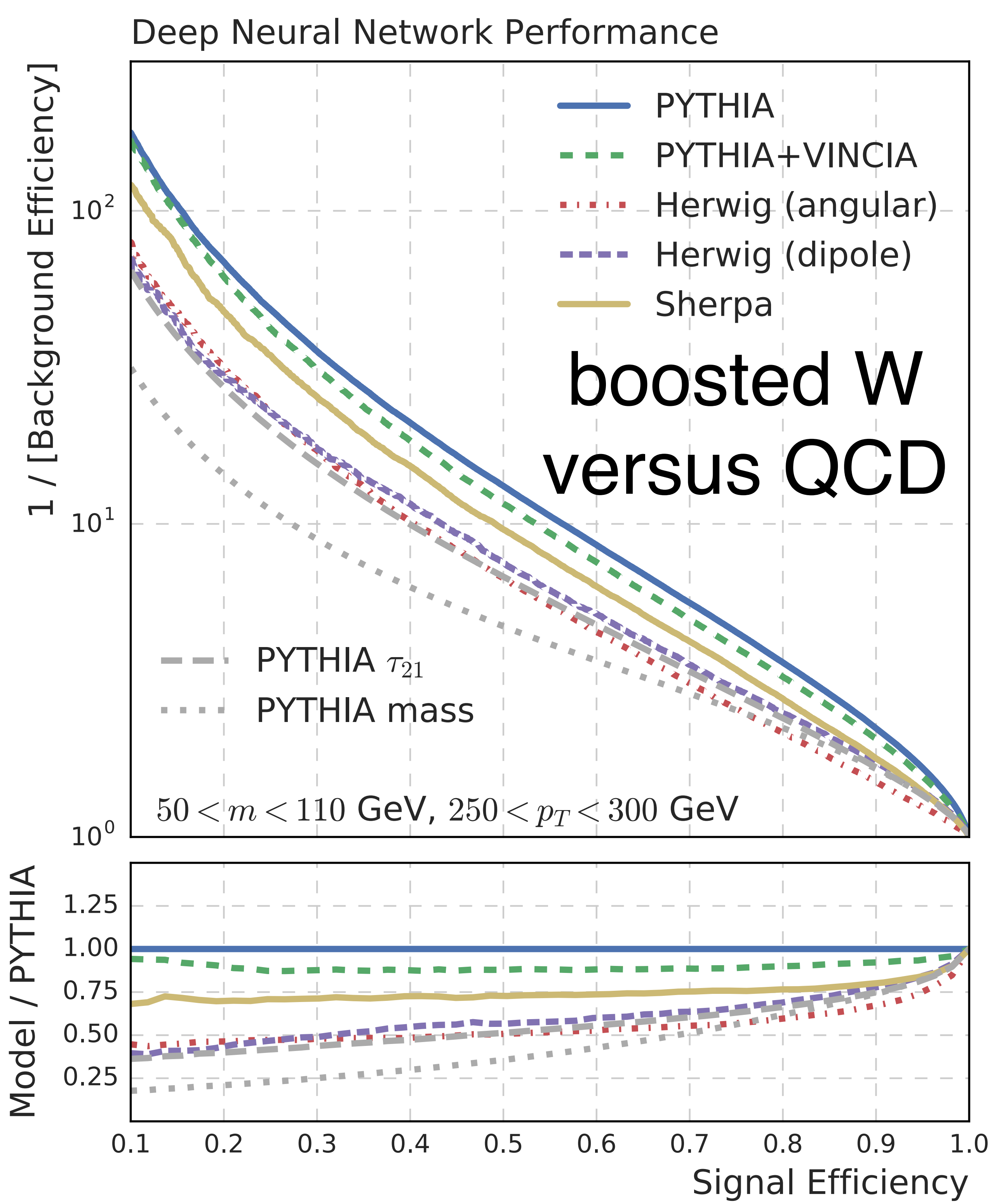
DNN classifiers can **exploit** subtle features

subtle features are **hard to model** !

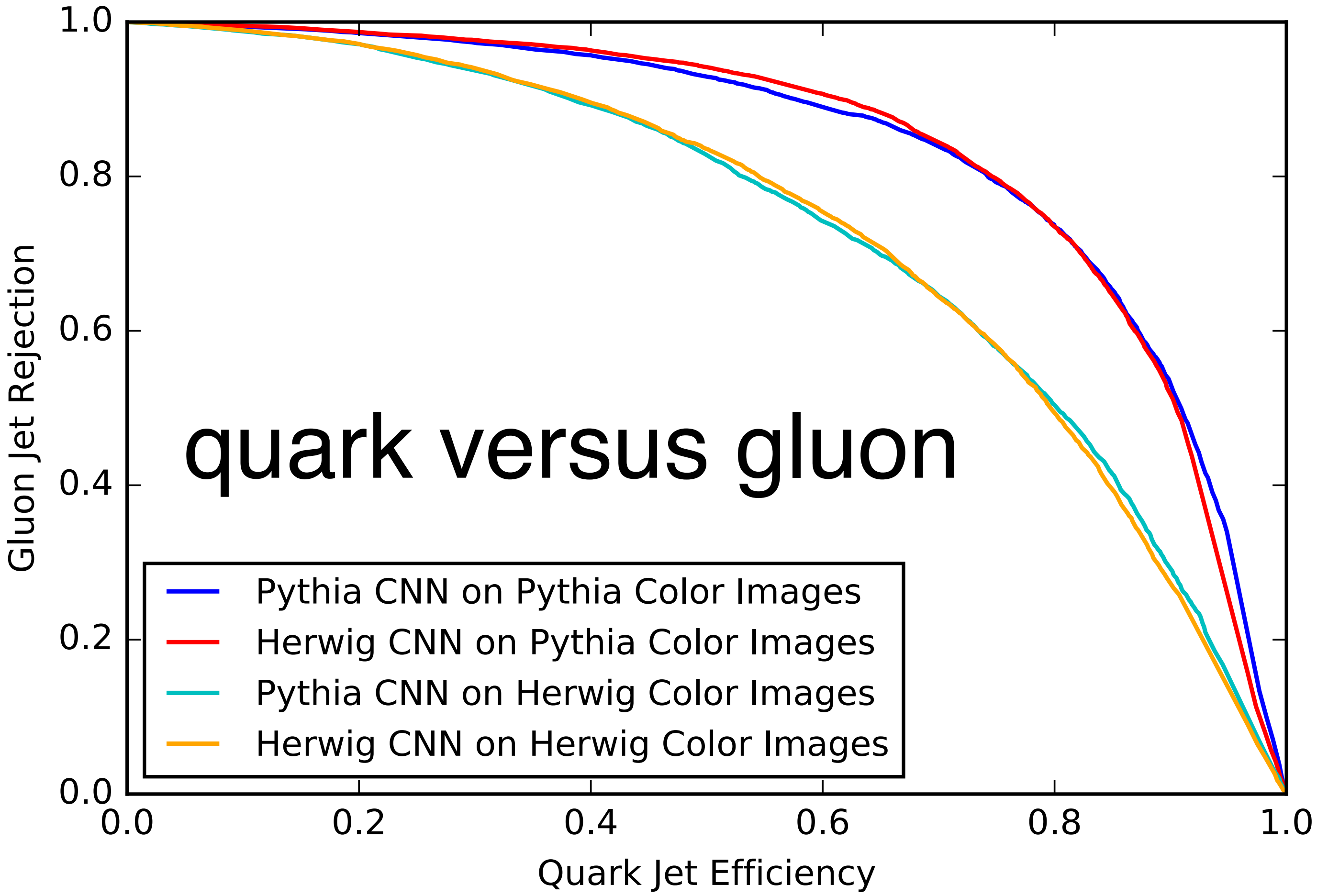
we need to be careful about which models we use - **only data is correct**

Sensitivity to Modeling (cont.)

J. Barnard et al. 1609.00607



Komiske et al. 1612.01551



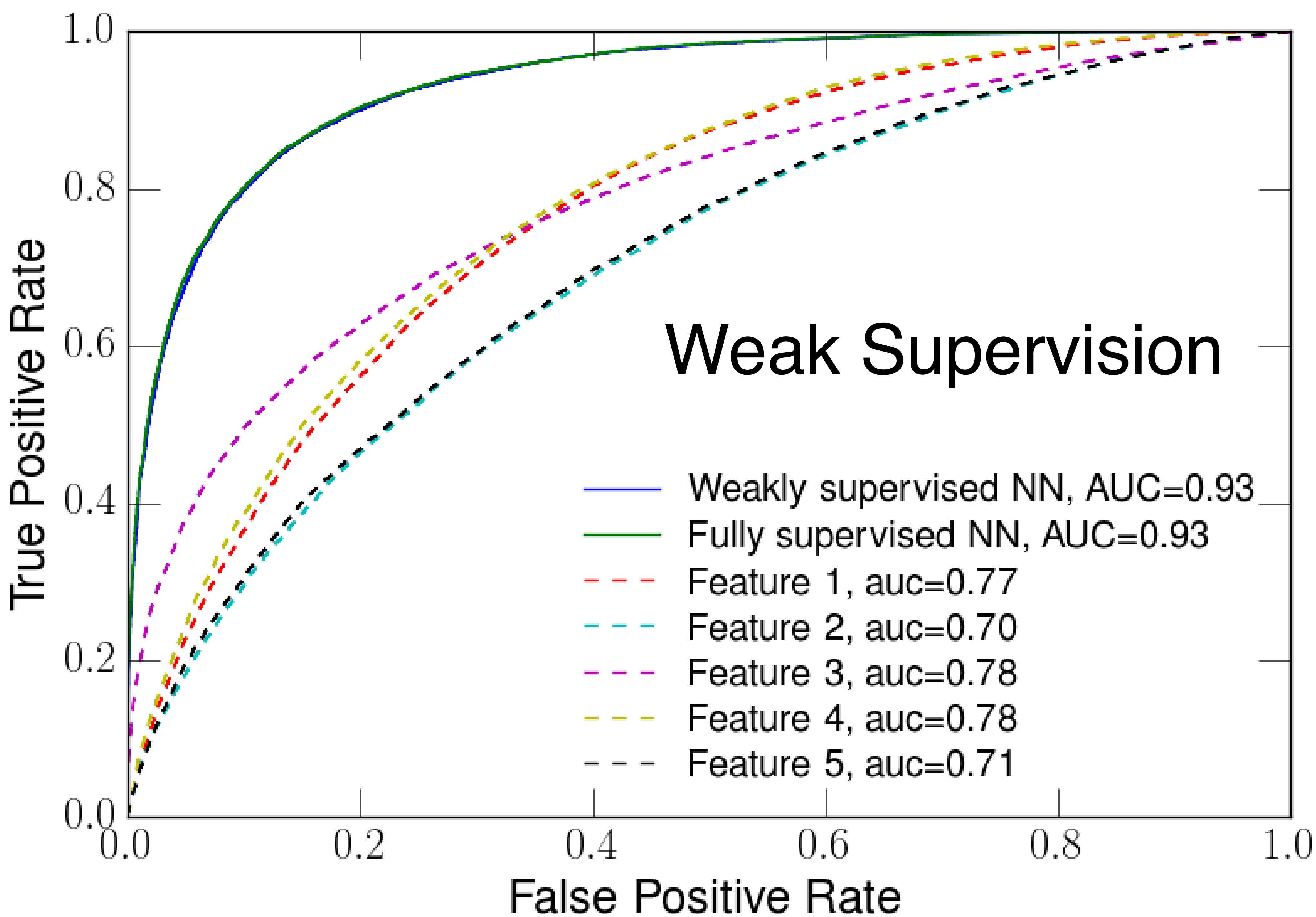
The story may be more complicated: even if performance differs, network itself my be similar

(could also be coincidence - fixed cuts may yield different performance)

Solution: Training Directly on Data

L. Dery et al. (**BN**) arXiv:1702.00414

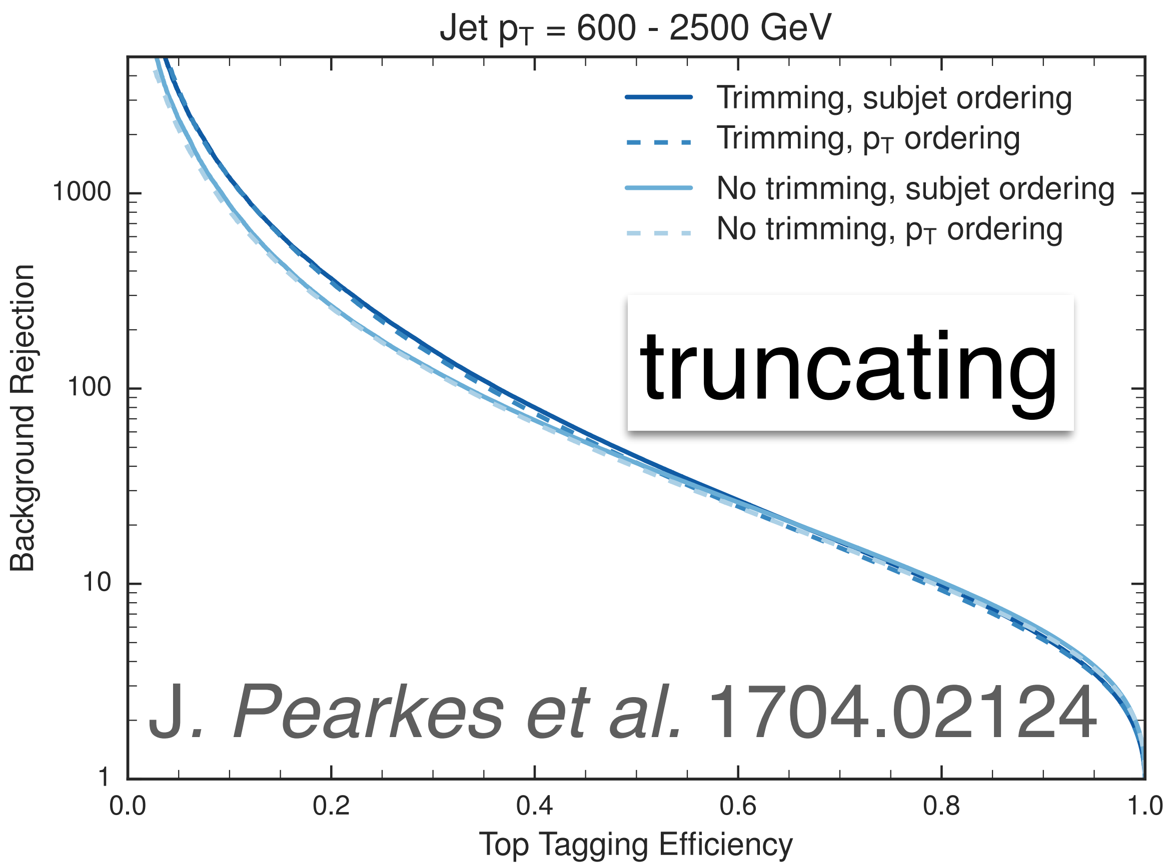
The only way for your classifier to be **optimal when applied to data** is to **train on data**.



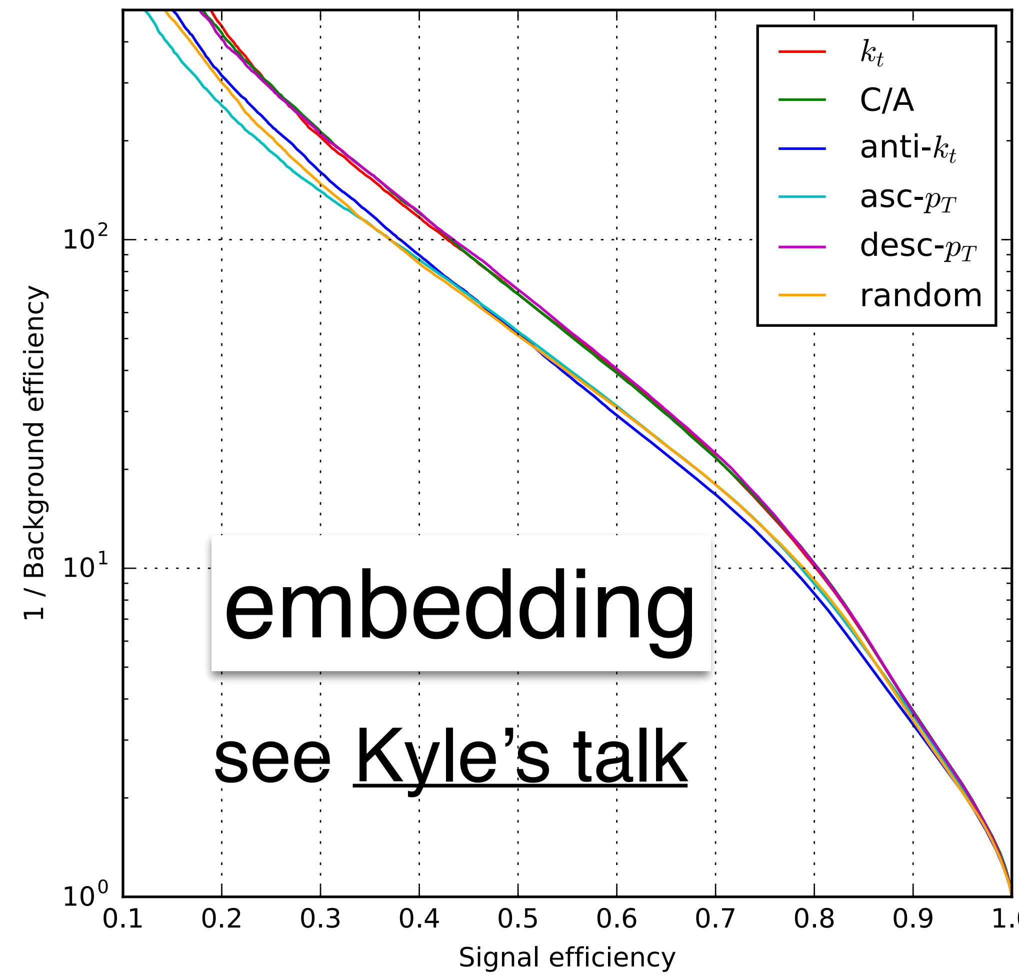
One regime where this is possible is when you have multiple samples with known class proportions.

When the proportions are non-unity it is still possible to modify the loss and learn!

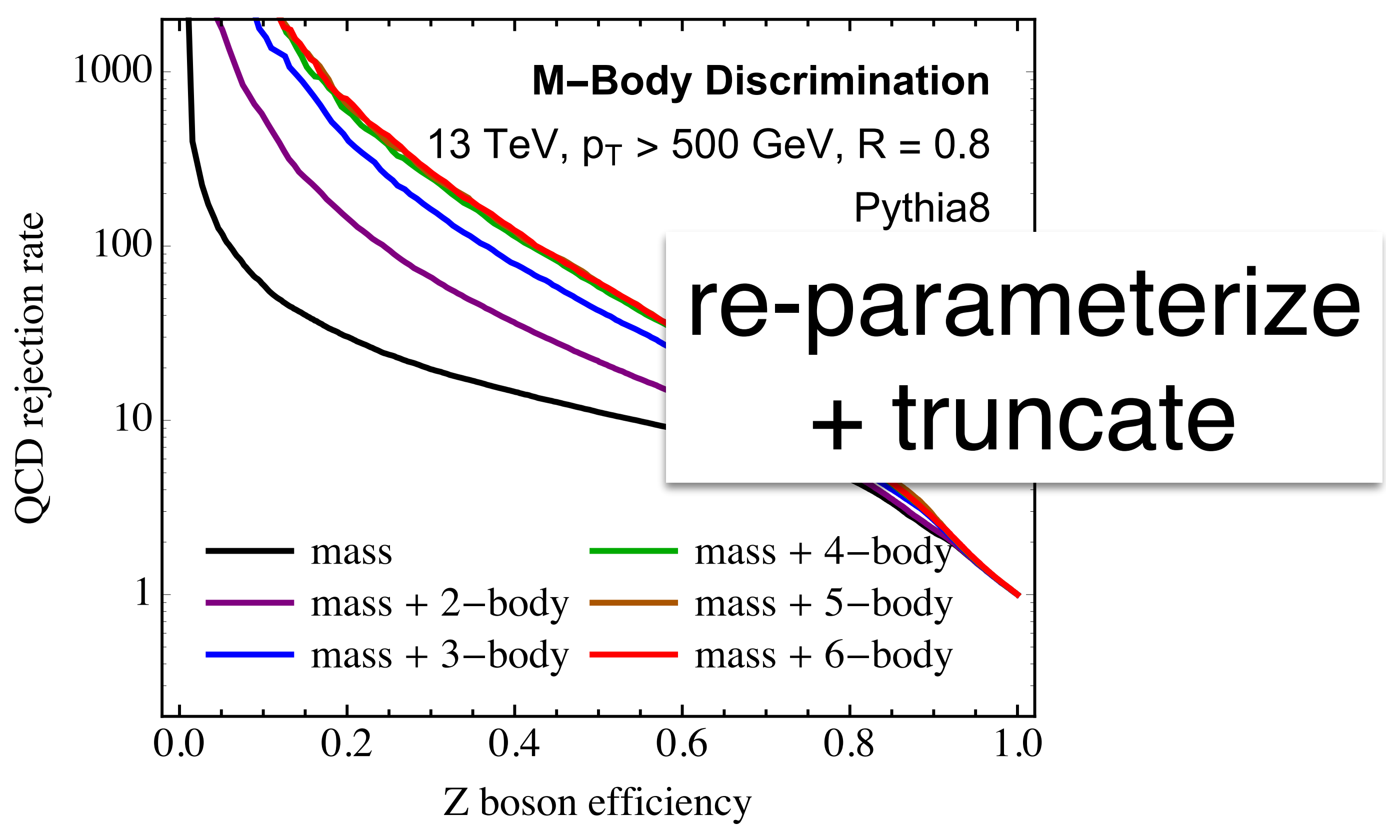
Related Work - other fixed representations



G. Louppe et al. 1702.00748

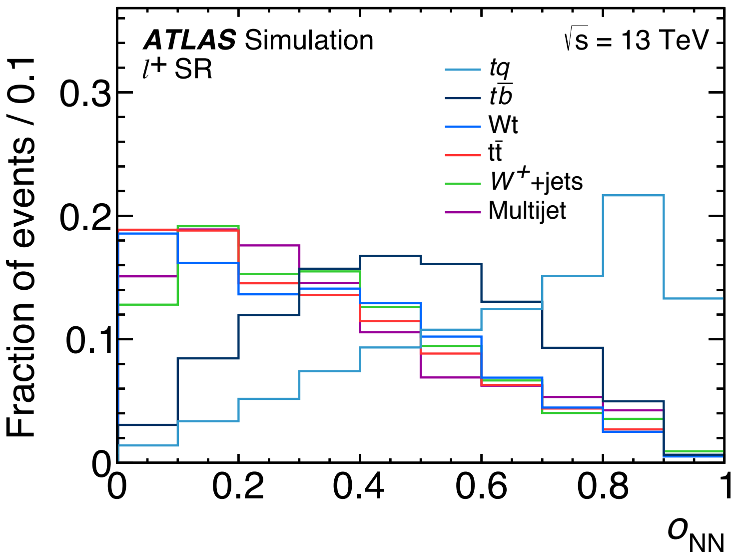


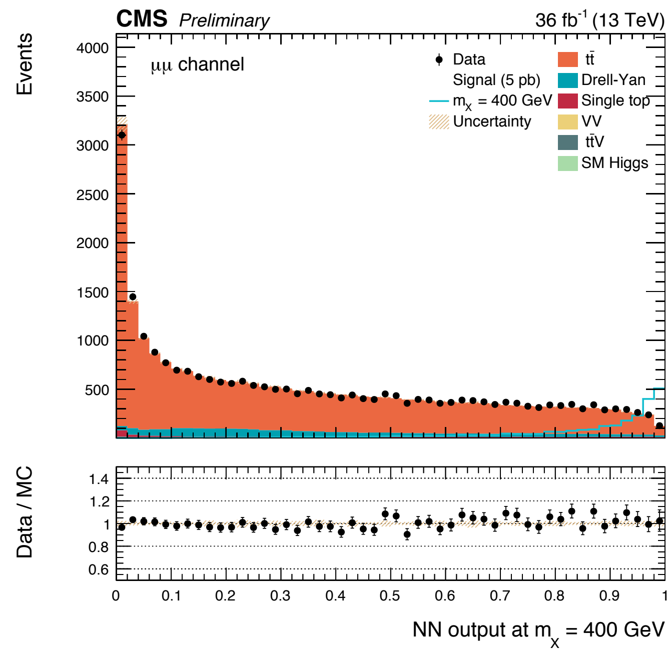
K. Datta et al. 1704.08249

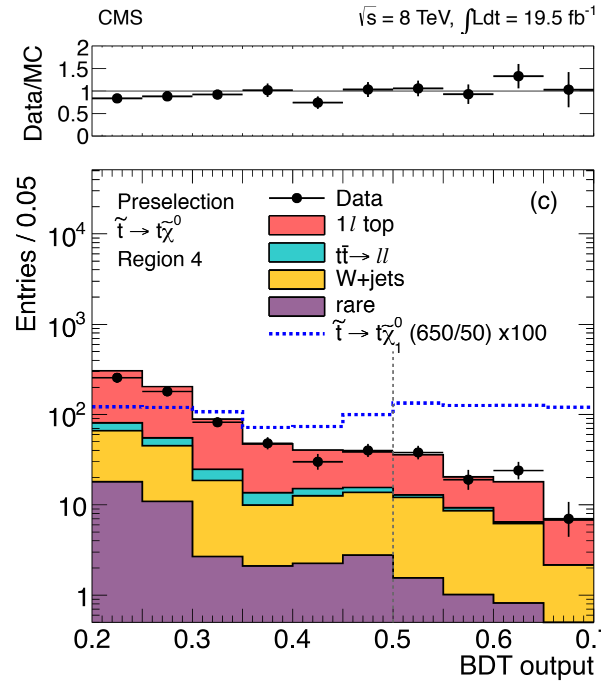


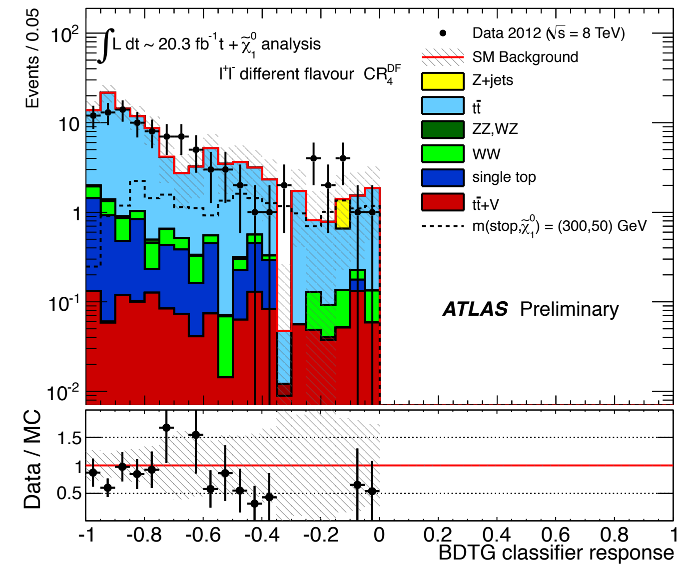
Unfortunately, I don't have time to talk about these interesting approaches!

NN, BDT, …



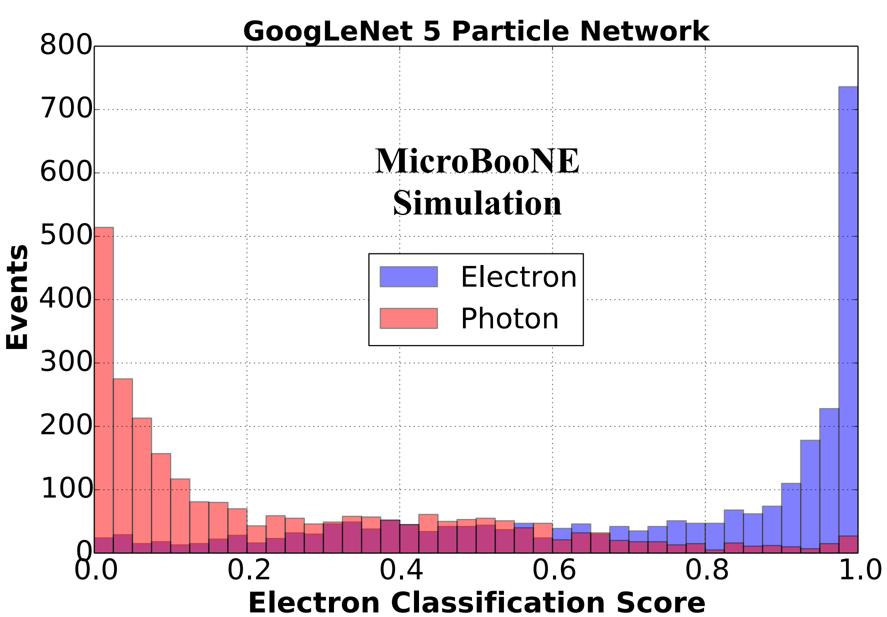


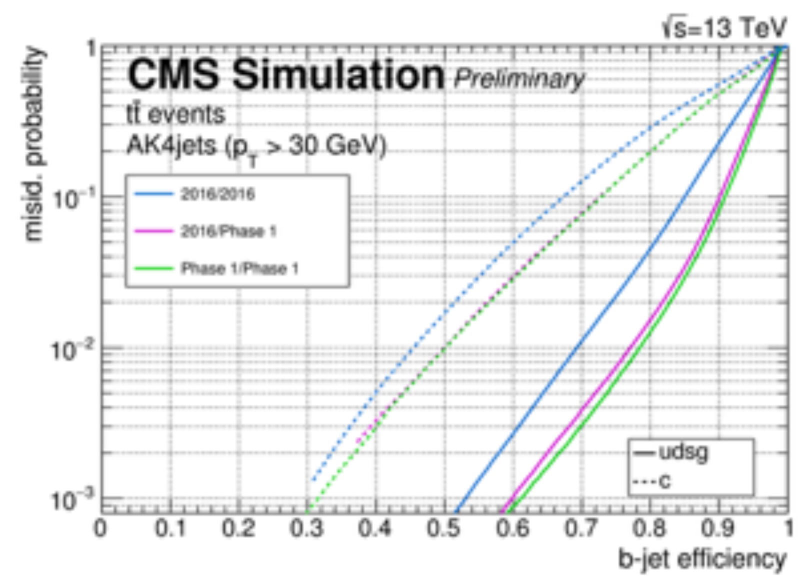


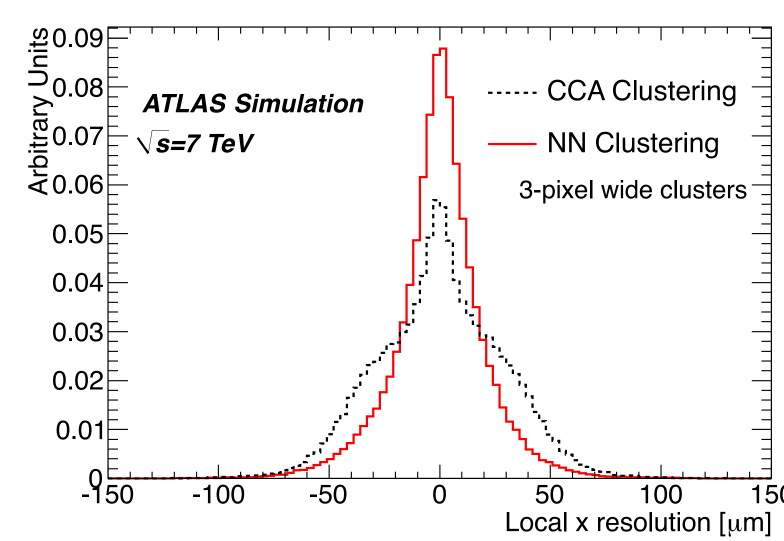


TMVA

NNPDF



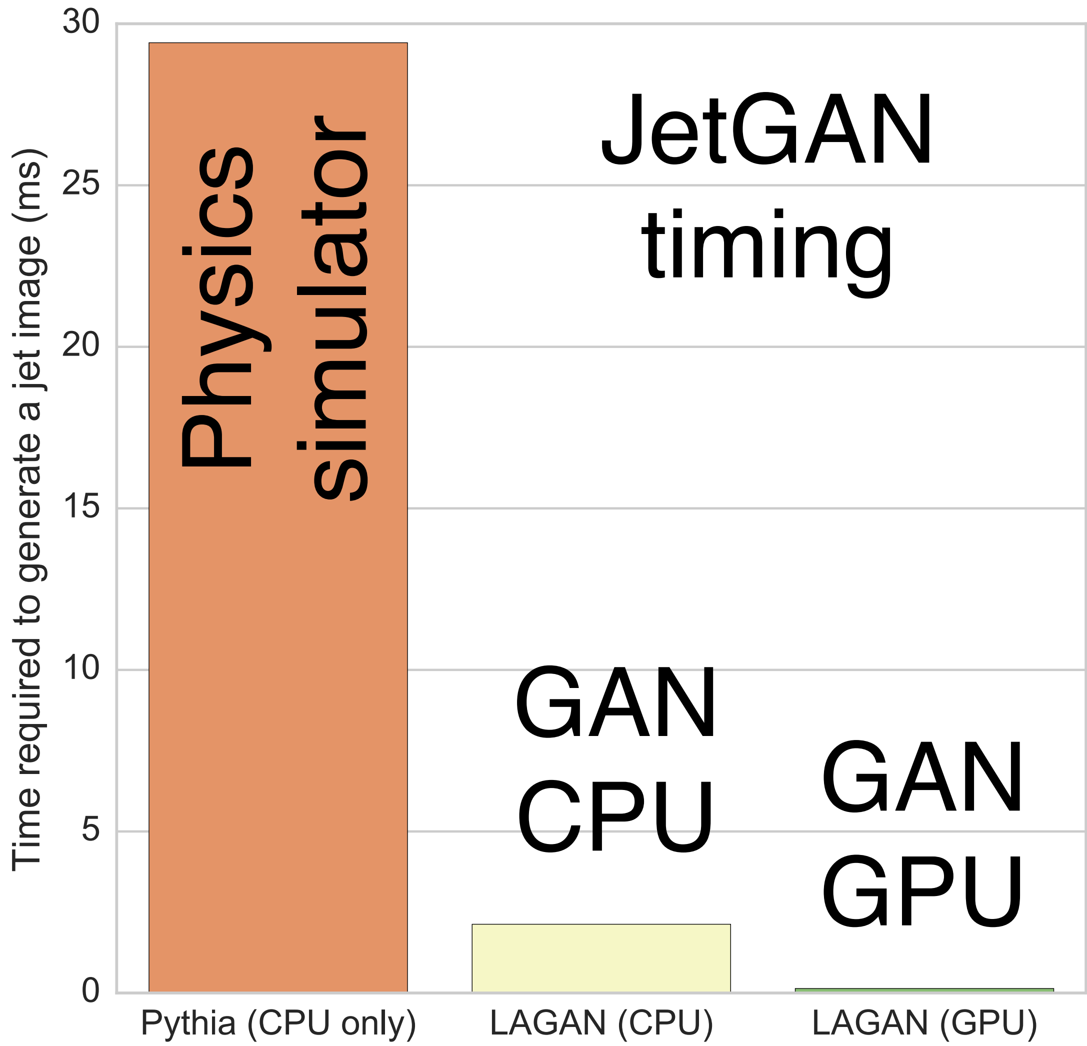


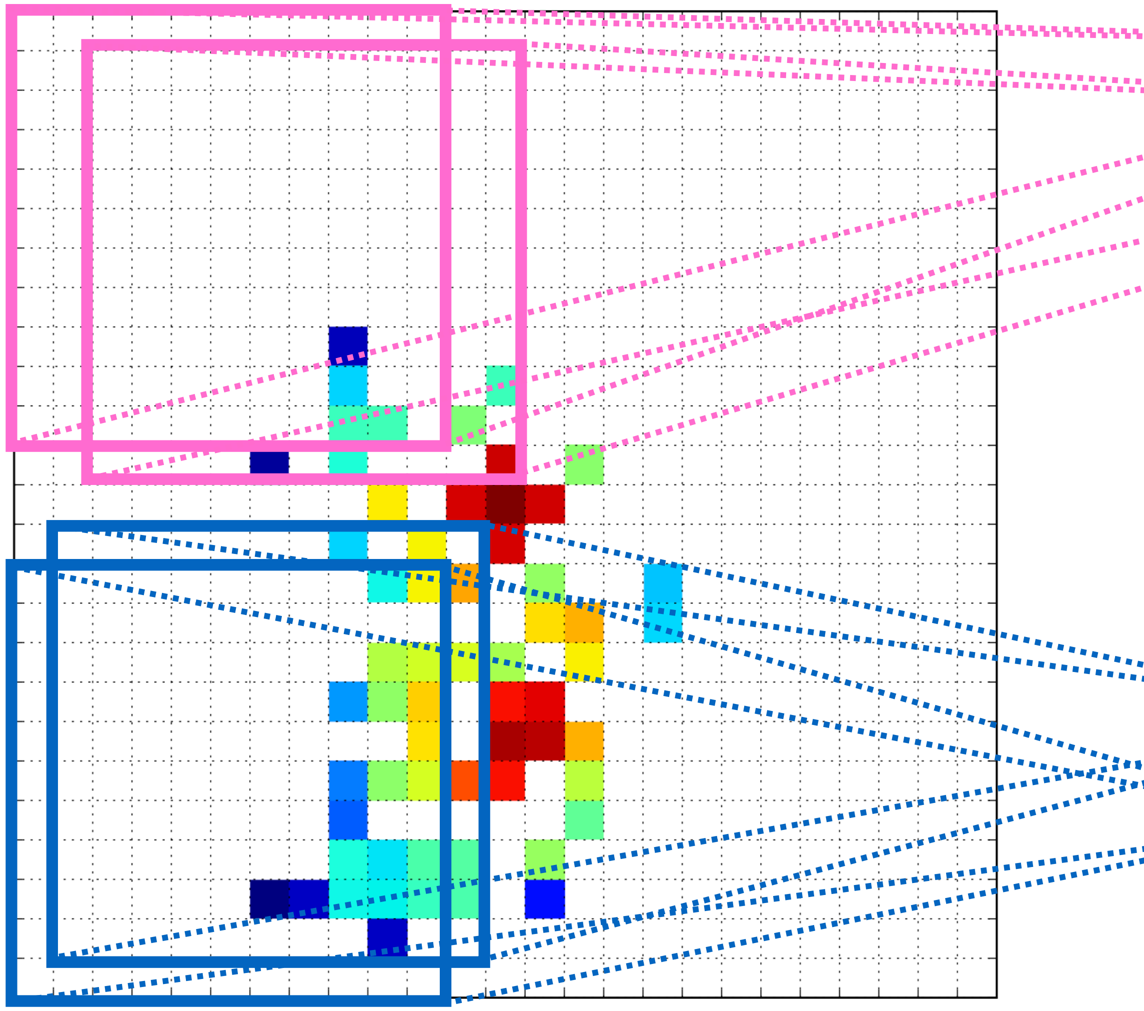


Conclusions

Machine Learning offers powerful tools for fully exploiting the physics program at the LHC

the three 'ions' of ML:
classification, regression, generation

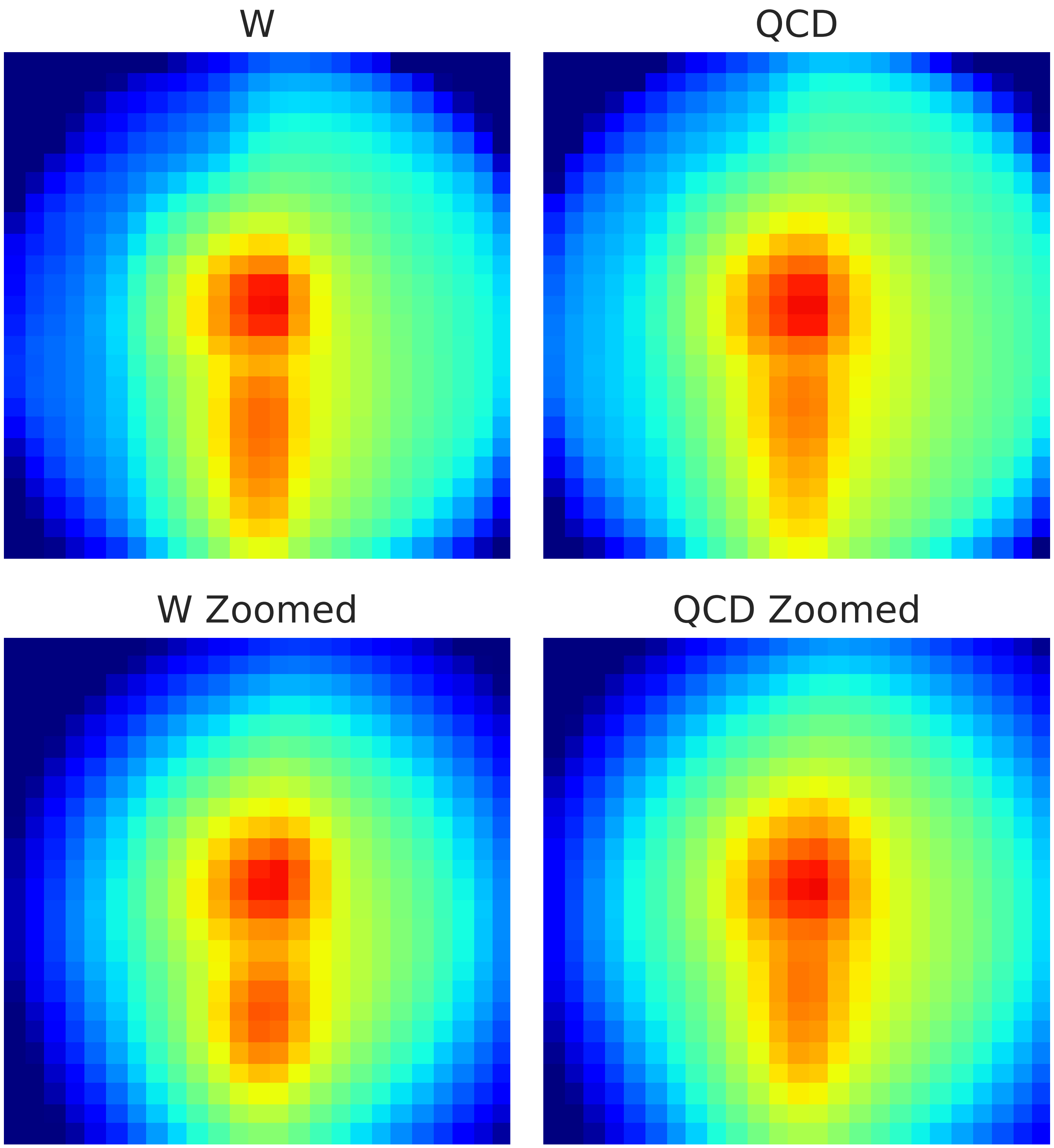




The key to robustness is to study what is being learned; this may even help us to learn something new about the SM!

Backup

Pre-processing: Zooming



$50 < m < 110$ GeV, $200 < p_T < 500$ GeV

"Zooming" (two-prong)

J. Barnard et al. 1609.00607