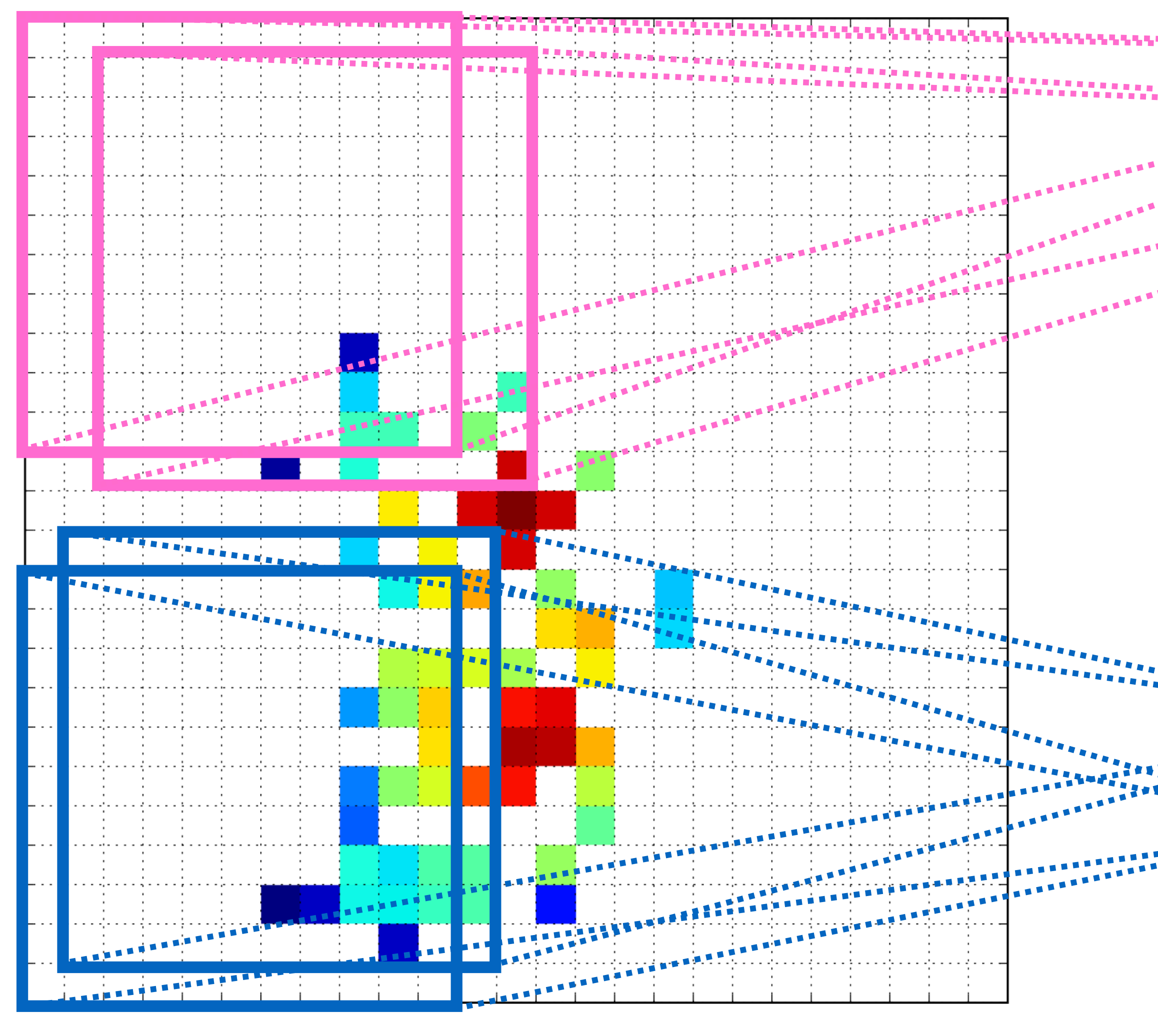
Advanced Maching Learning in High Energy Physics

Benjamin Nachman

Lawrence Berkeley National Laboratory

EFI Data Analytics Workshop
October 26, 2017



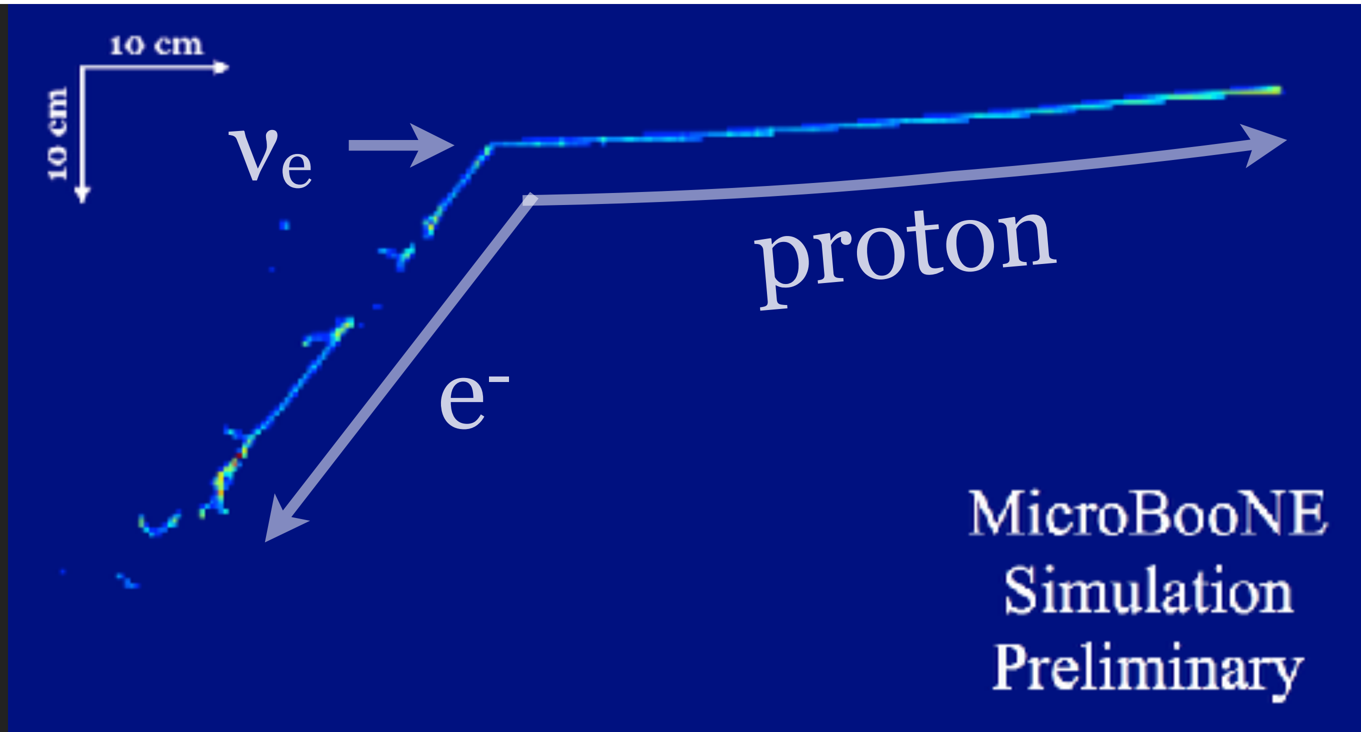
Disclaimer: I'm not going to talk about:

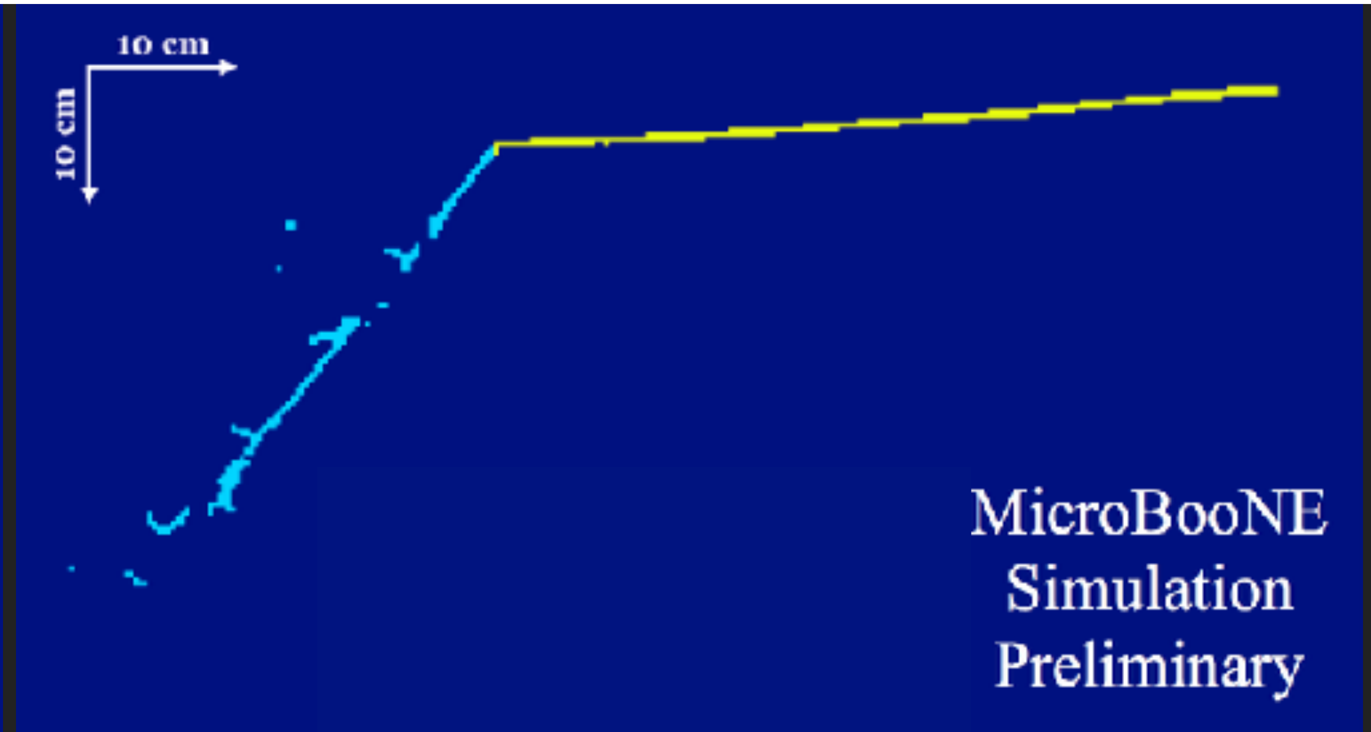
- deep learning "simply" replacing shallow learning
- interesting work from the large neutrino experiments
- non-image based classification at the LHC

a lot of the content still applies

Instead, I'll use hadronic final states at the LHC to illustrate DNN classification, regression, & generation

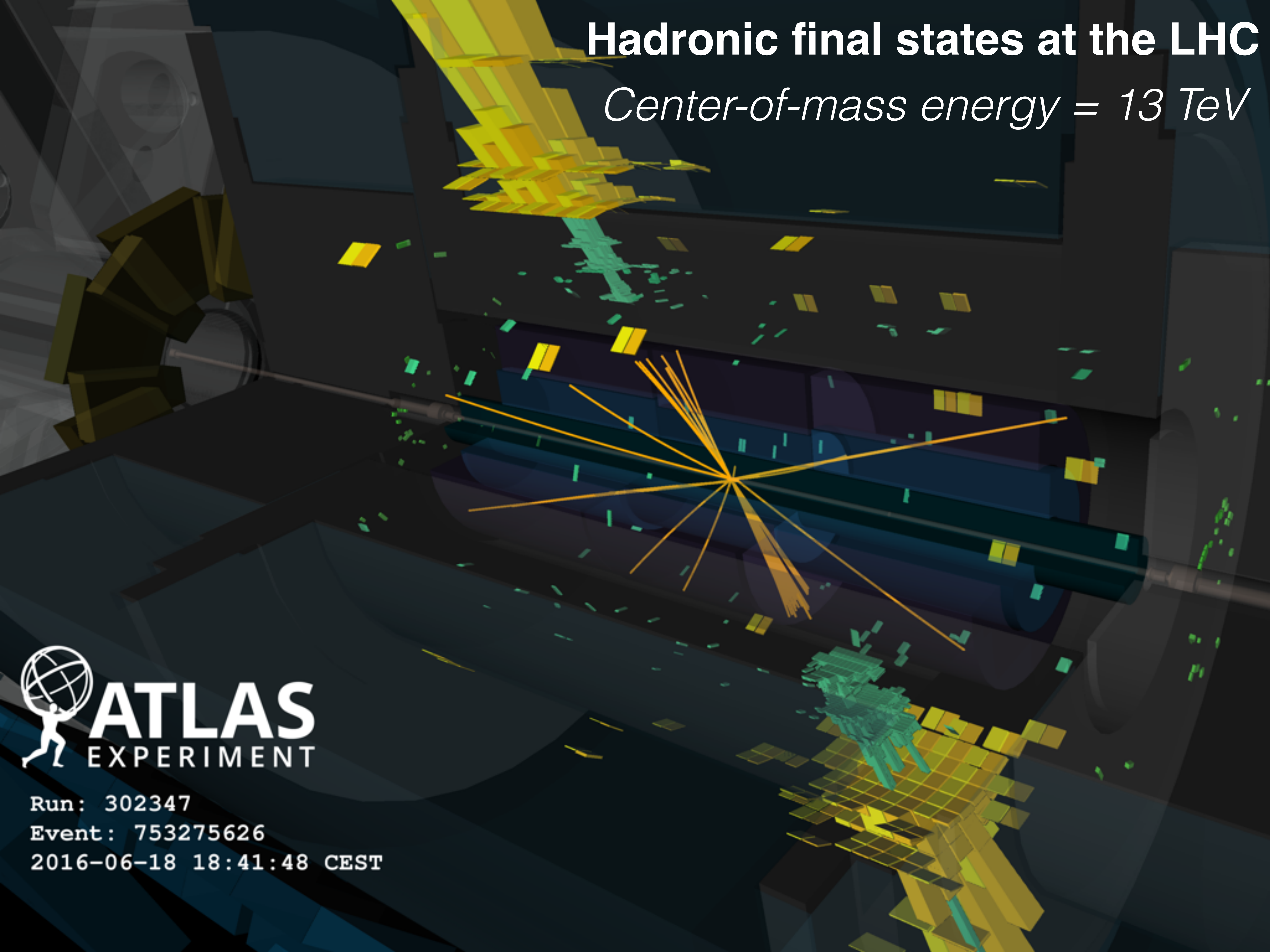
from T. Wongjirad's *DPF talk*

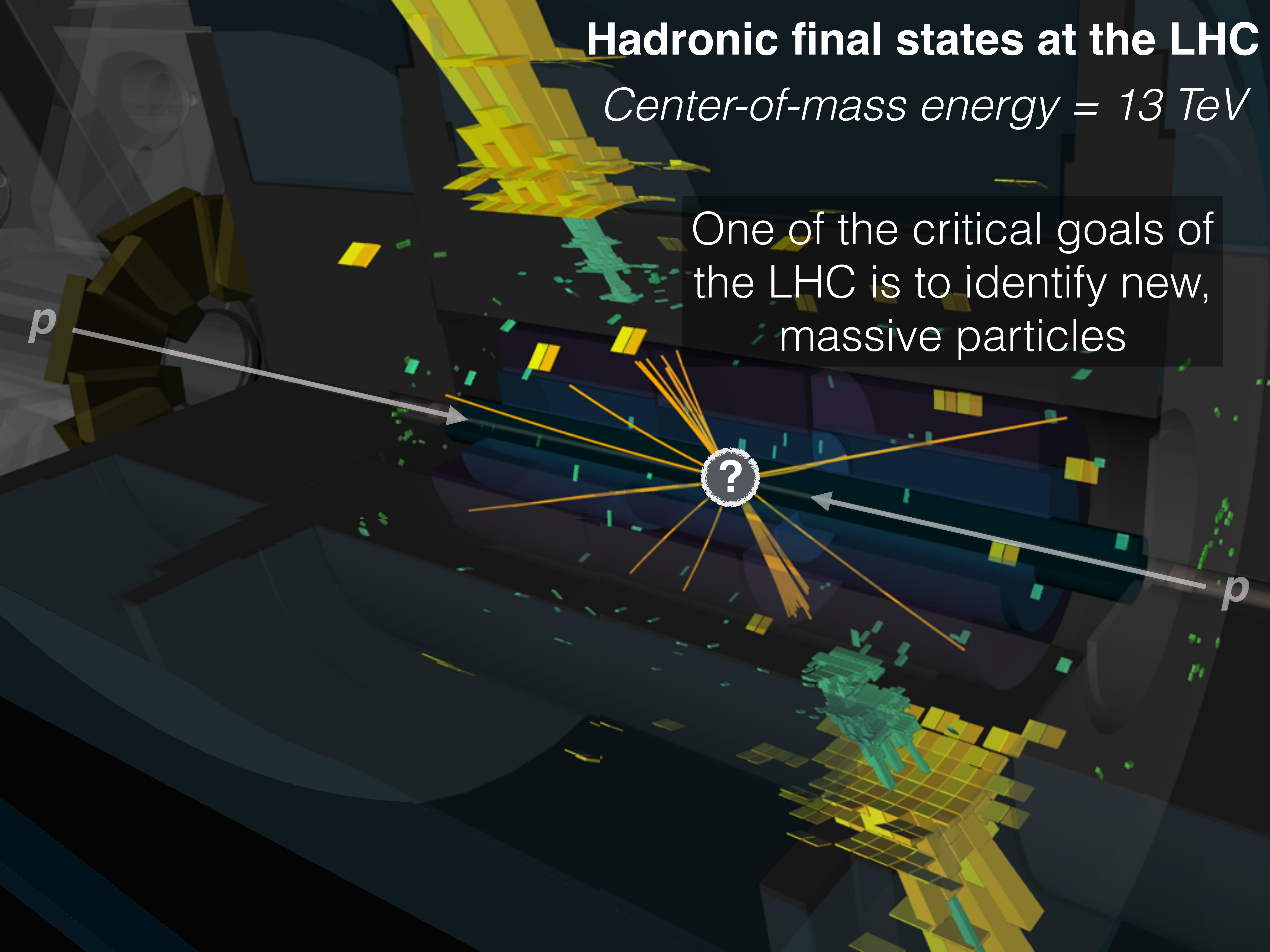




Hadronic final states at the LHC

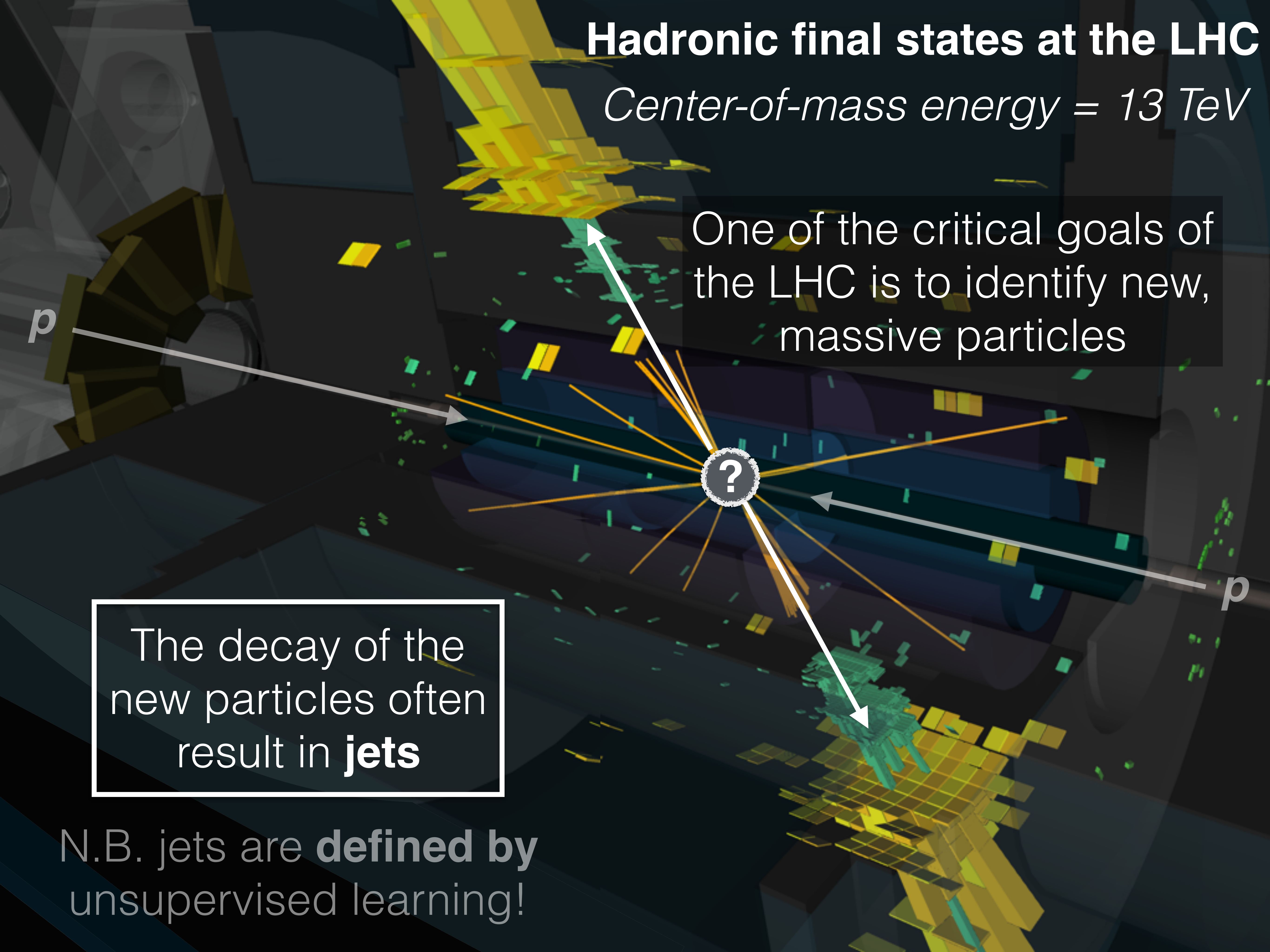
Center-of-mass energy = 13 TeV




Hadronic final states at the LHC
Center-of-mass energy = 13 TeV
One of the critical goals of the LHC is to identify new, massive particles
p
?
p

Hadronic final states at the LHC

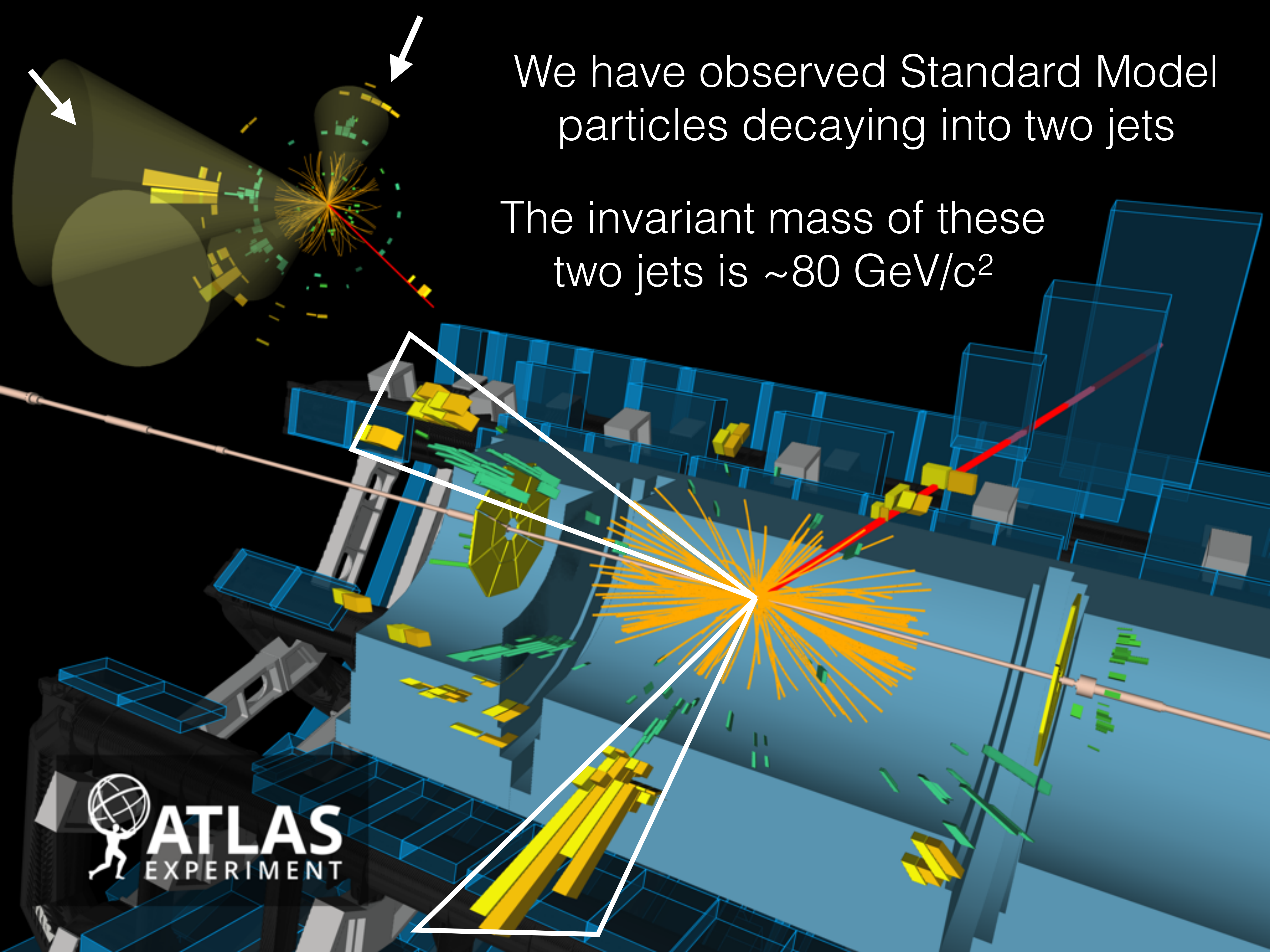
Center-of-mass energy = 13 TeV



One of the critical goals of the LHC is to identify new, massive particles

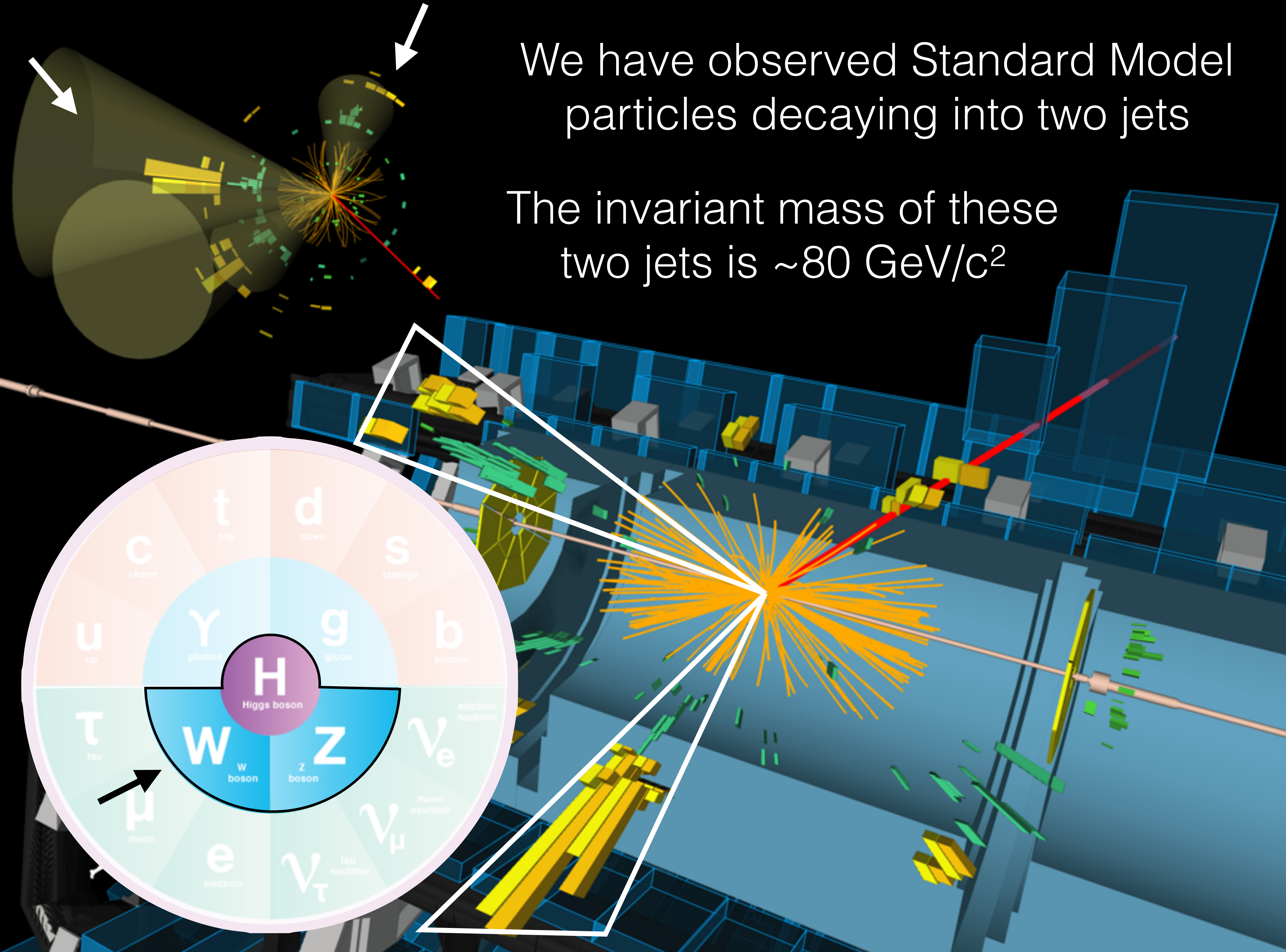
The decay of the new particles often result in **jets**

N.B. jets are **defined by** unsupervised learning!


We have observed Standard Model particles decaying into two jets
The invariant mass of these two jets is ~80 GeV/c²

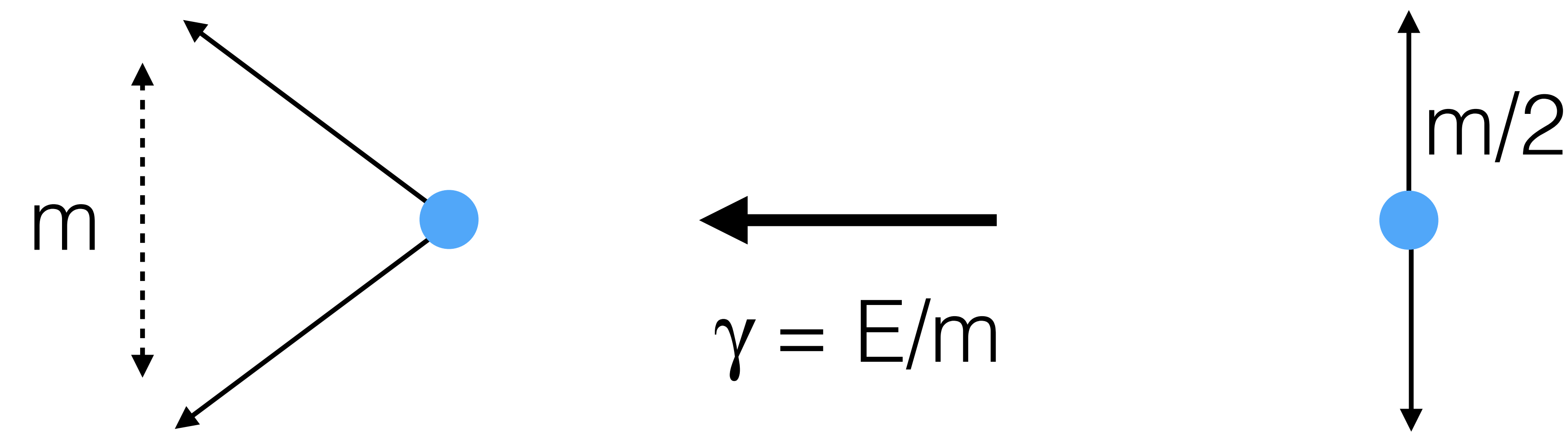
We have observed Standard Model particles decaying into two jets

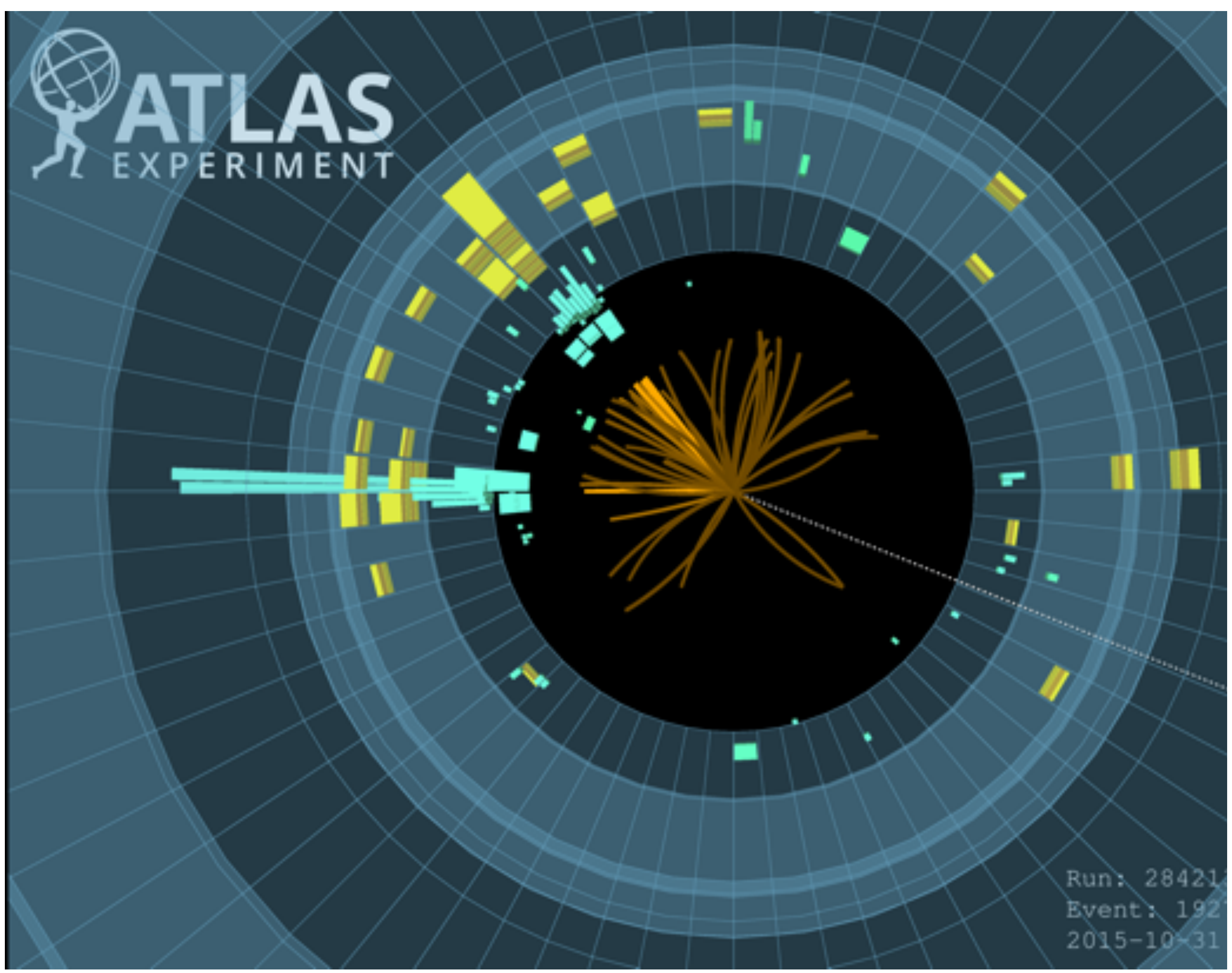
The invariant mass of these two jets is ~80 GeV/c²

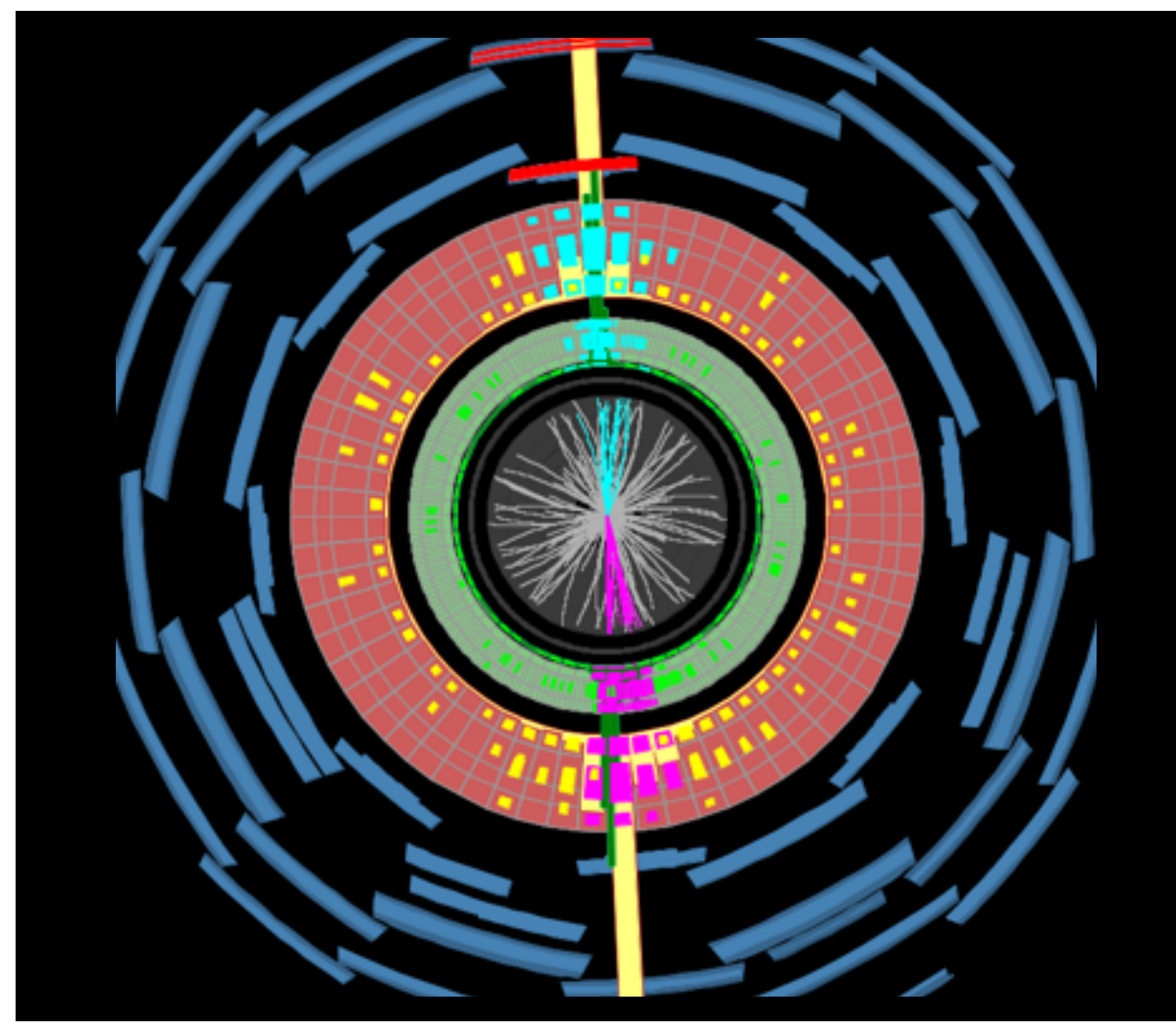


What if you take one of those SM dijet resonances and Lorentz boost it?

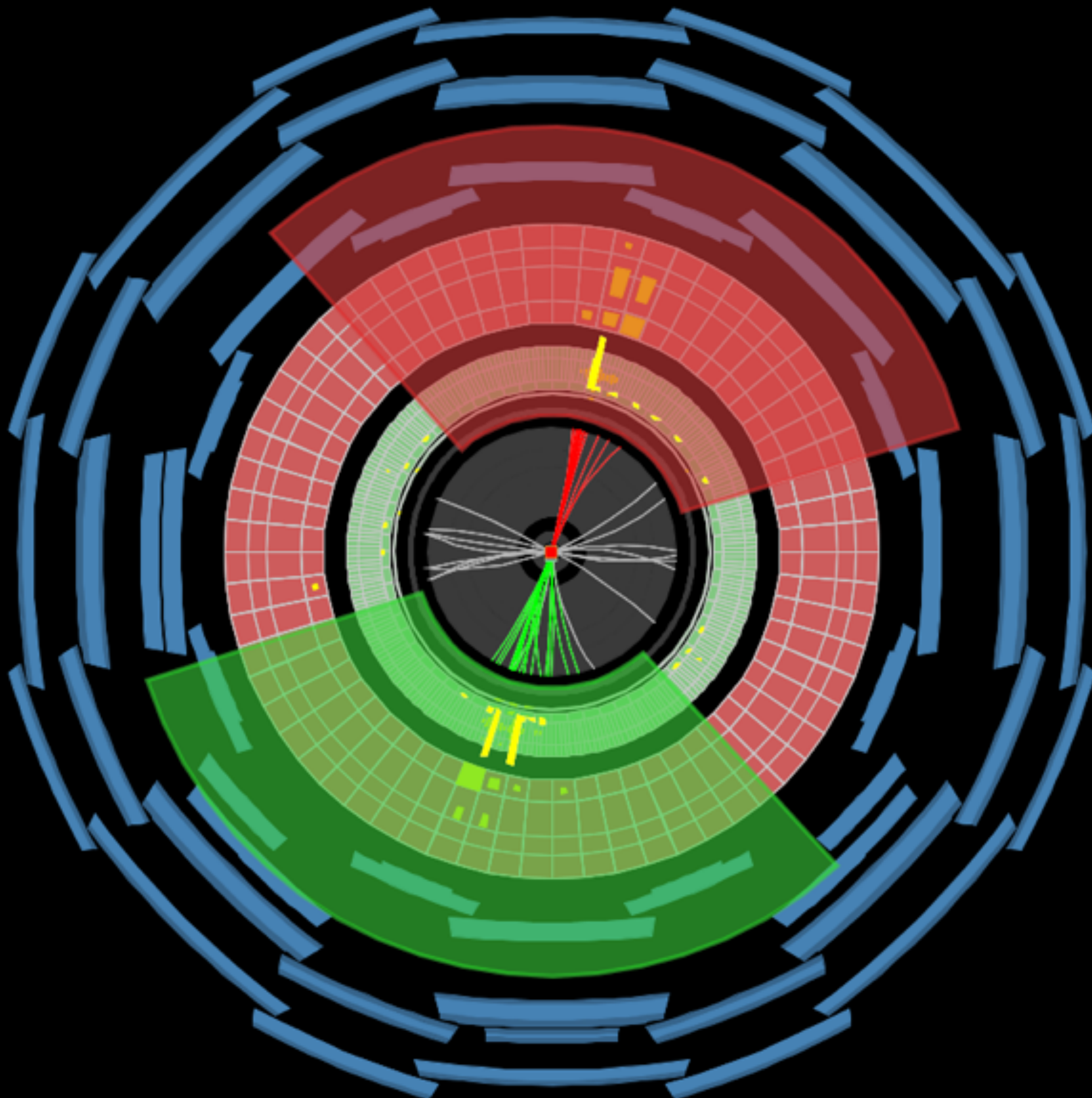
$\phi \sim 1/\gamma = m/E$

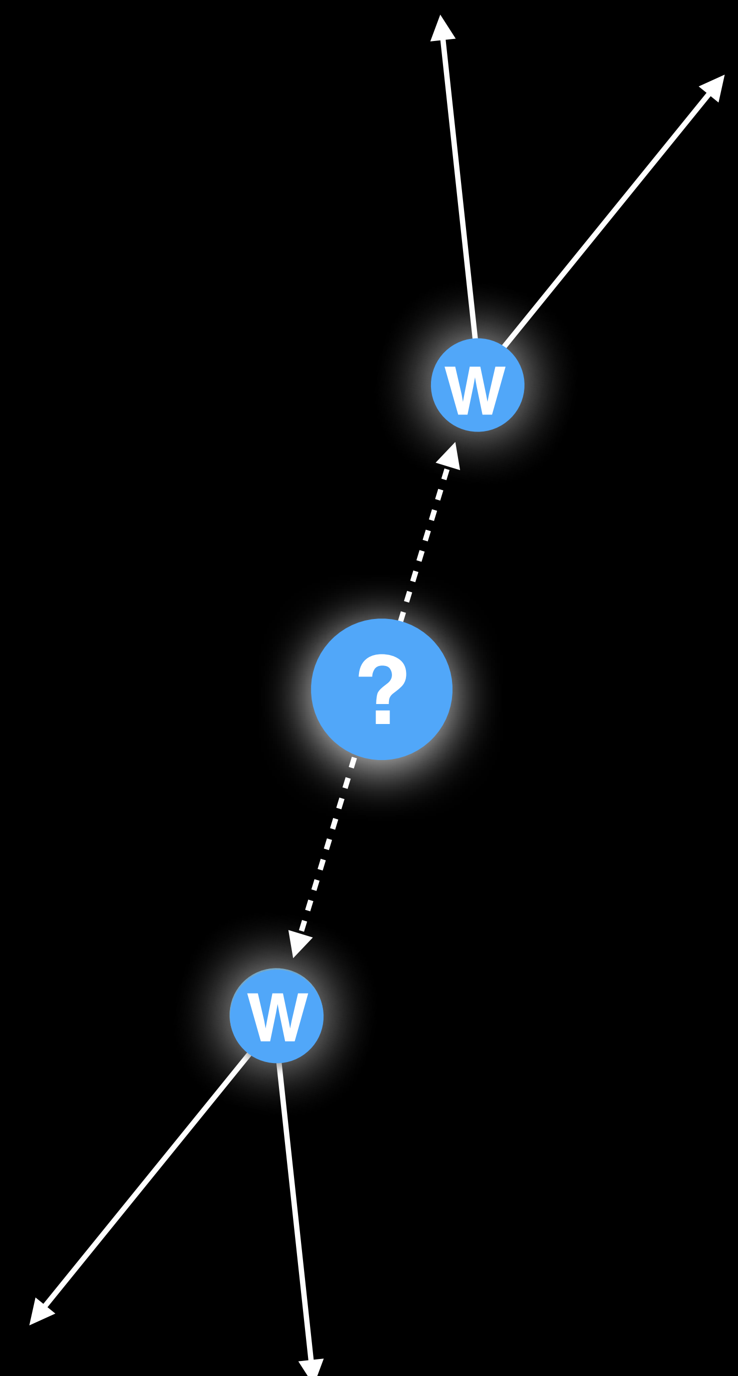






W bosons are naturally boosted if they result from the decay of something even heavier

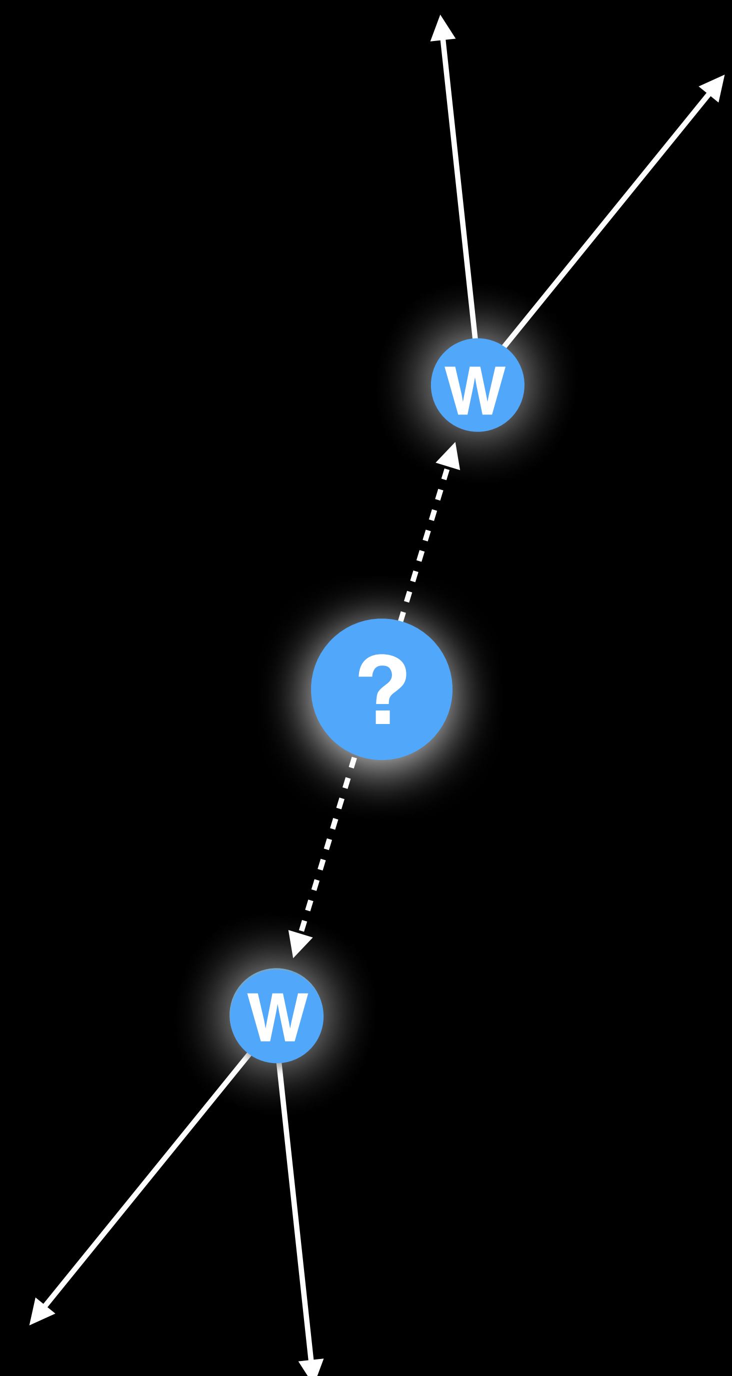


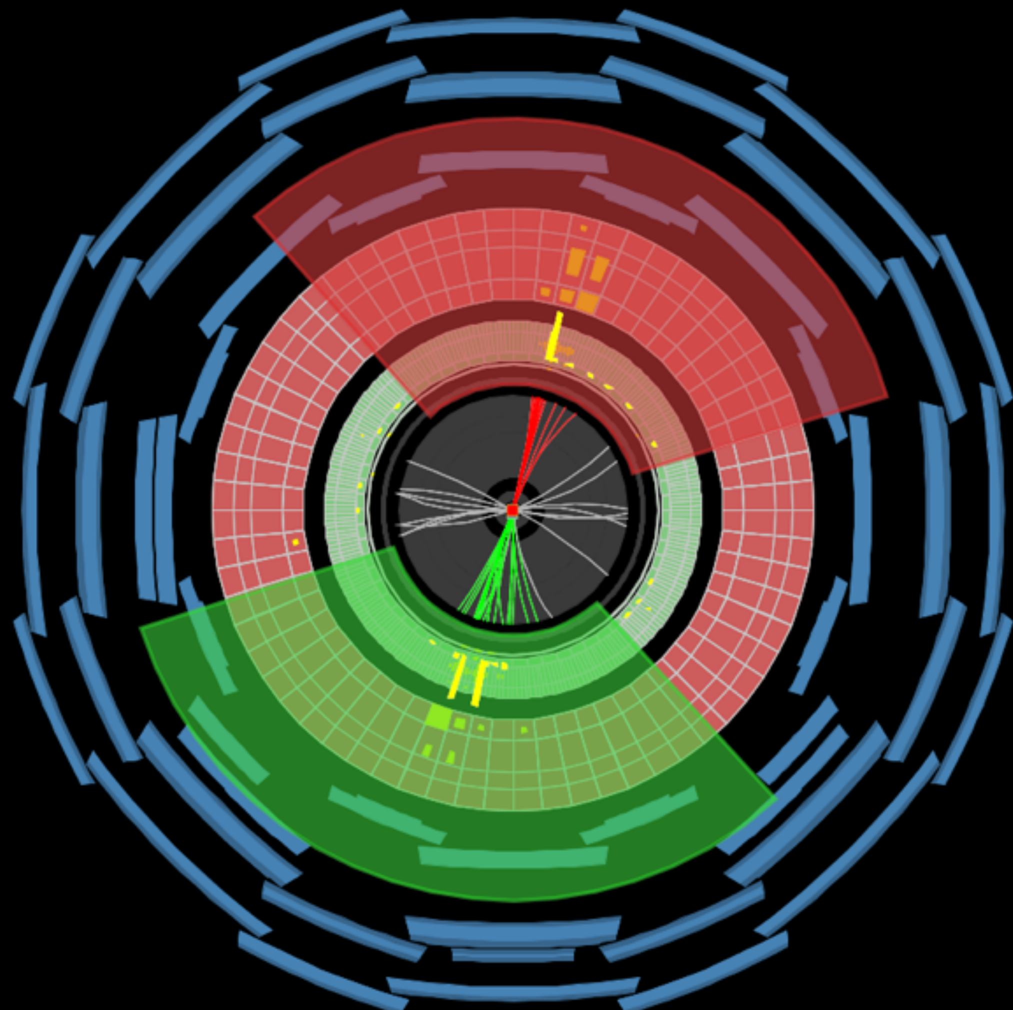


W bosons are naturally boosted if they result from the decay of something even heavier

Goal: Find W jets in an enormous sea of generic q/g jets

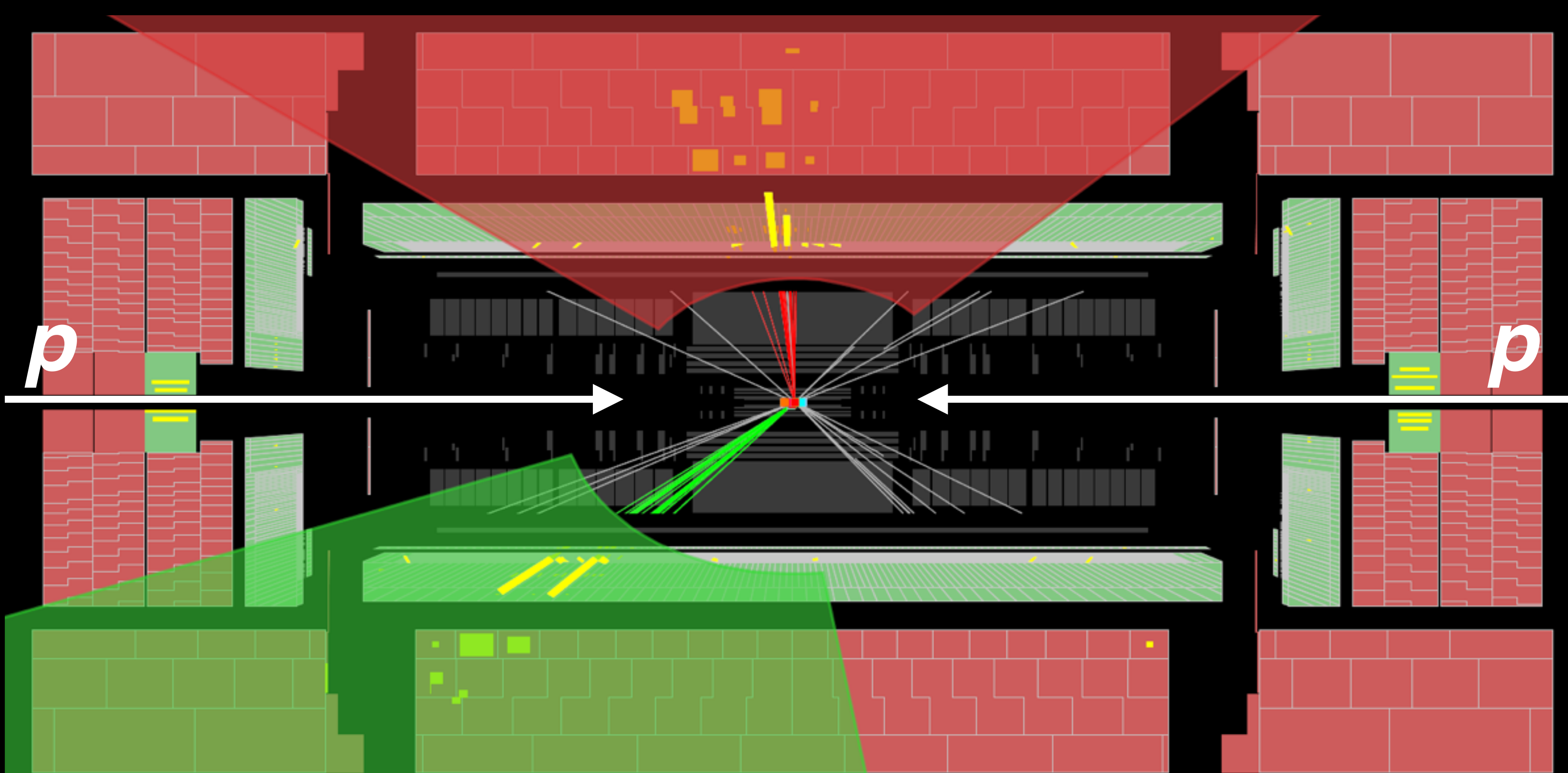
These jets have a non-trivial structure!

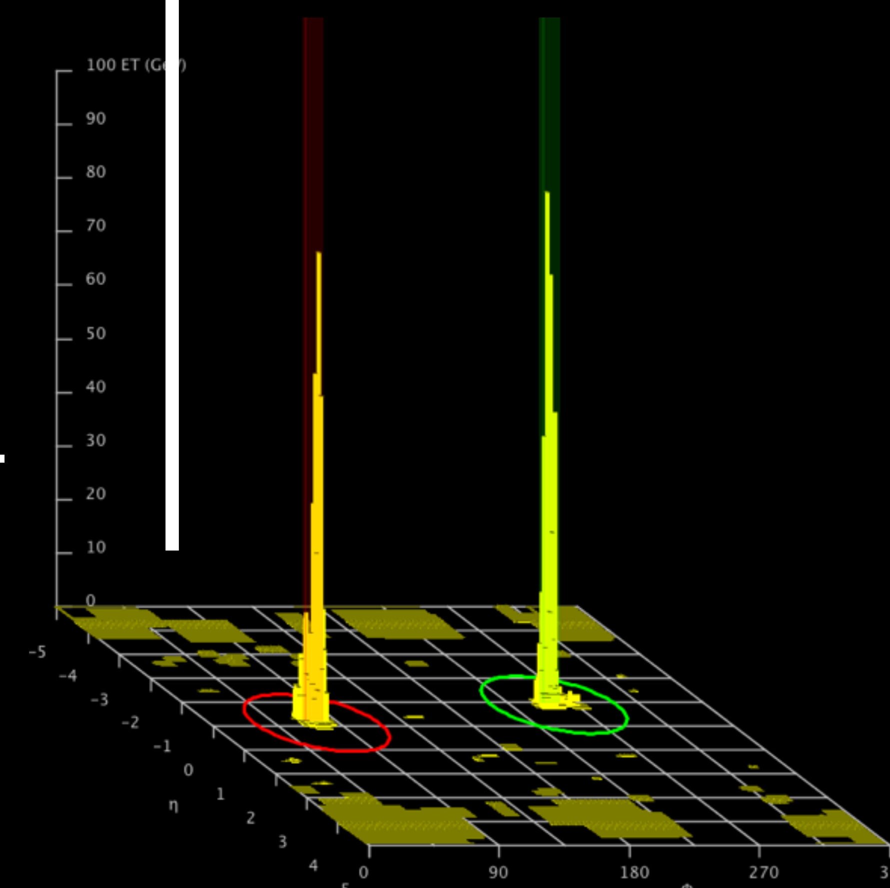




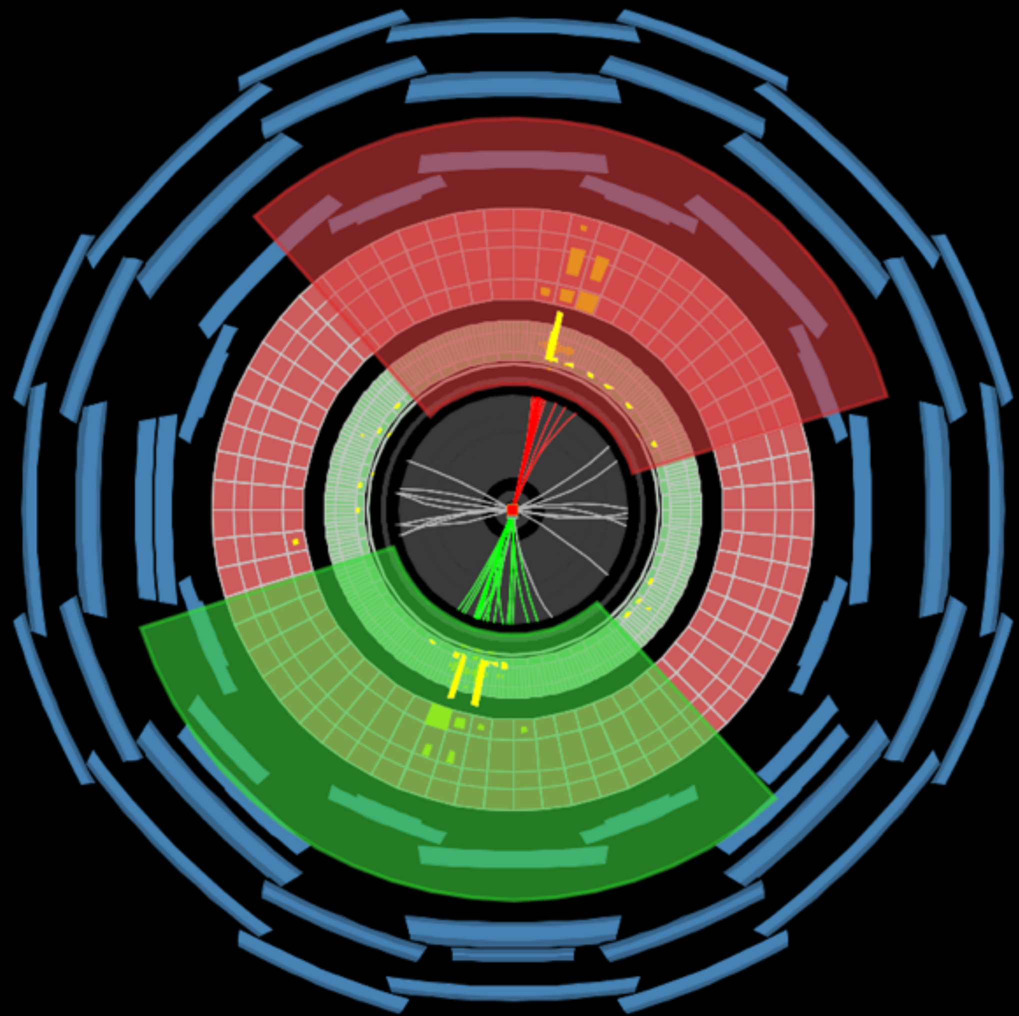
Searching for new particles decaying into boosted W bosons requires **looking at the radiation pattern inside jets**

momentum transverse to the beam (p_T)

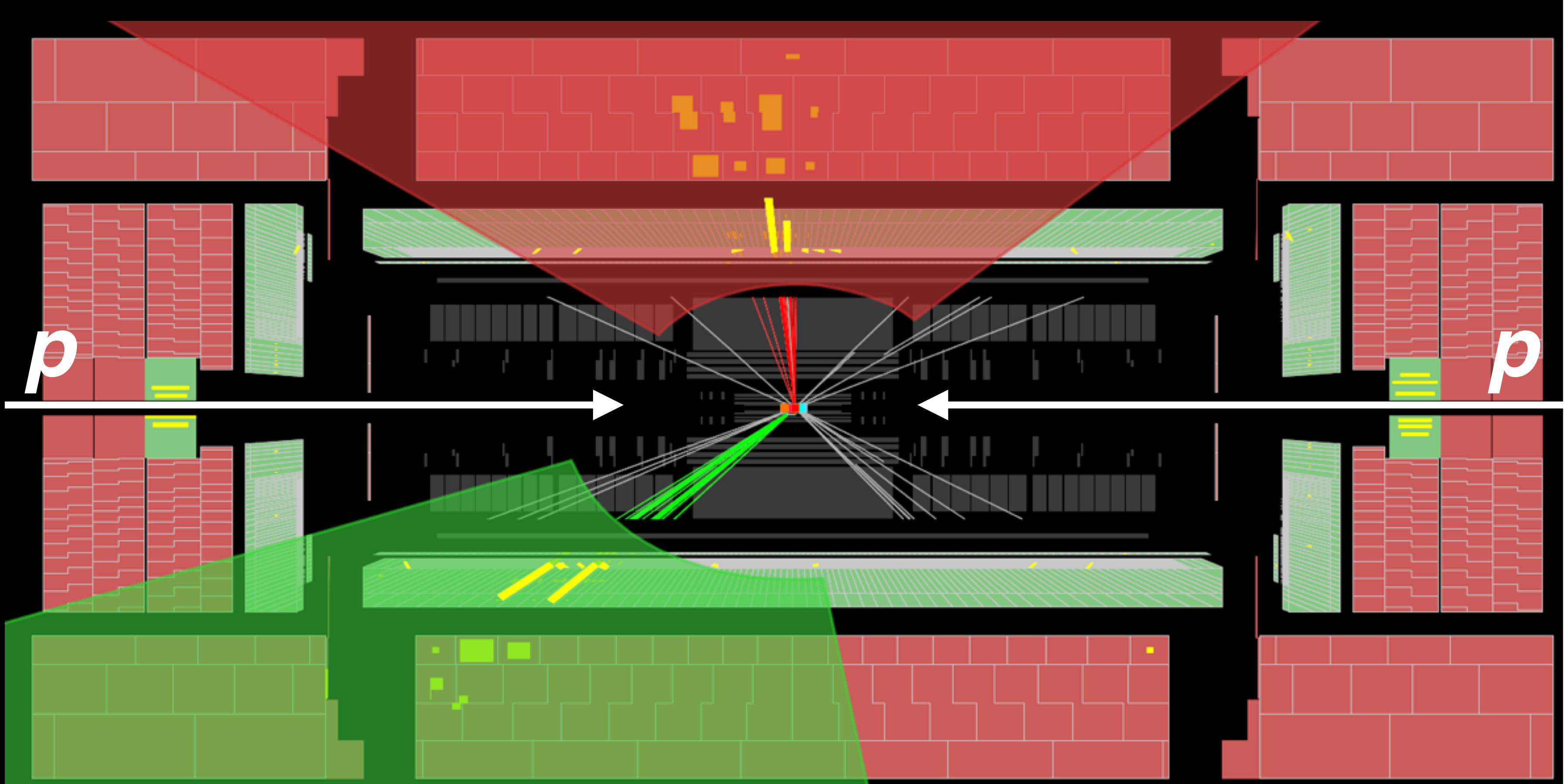


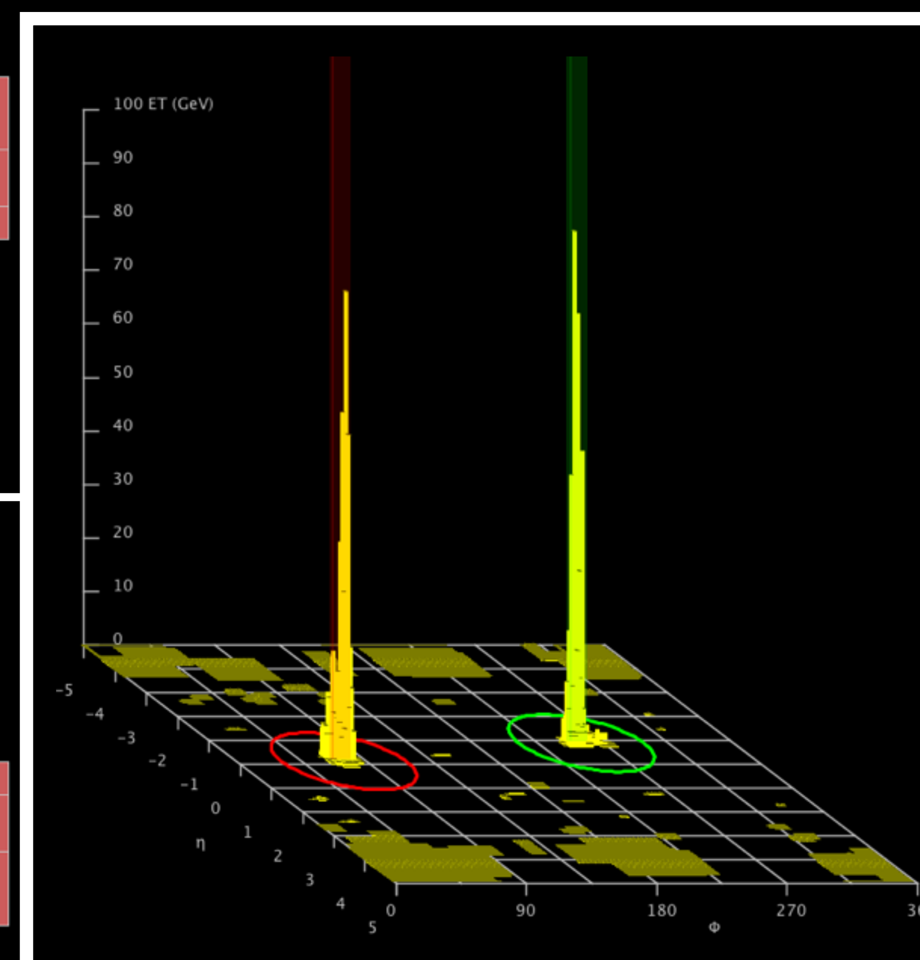


Up next: jet images



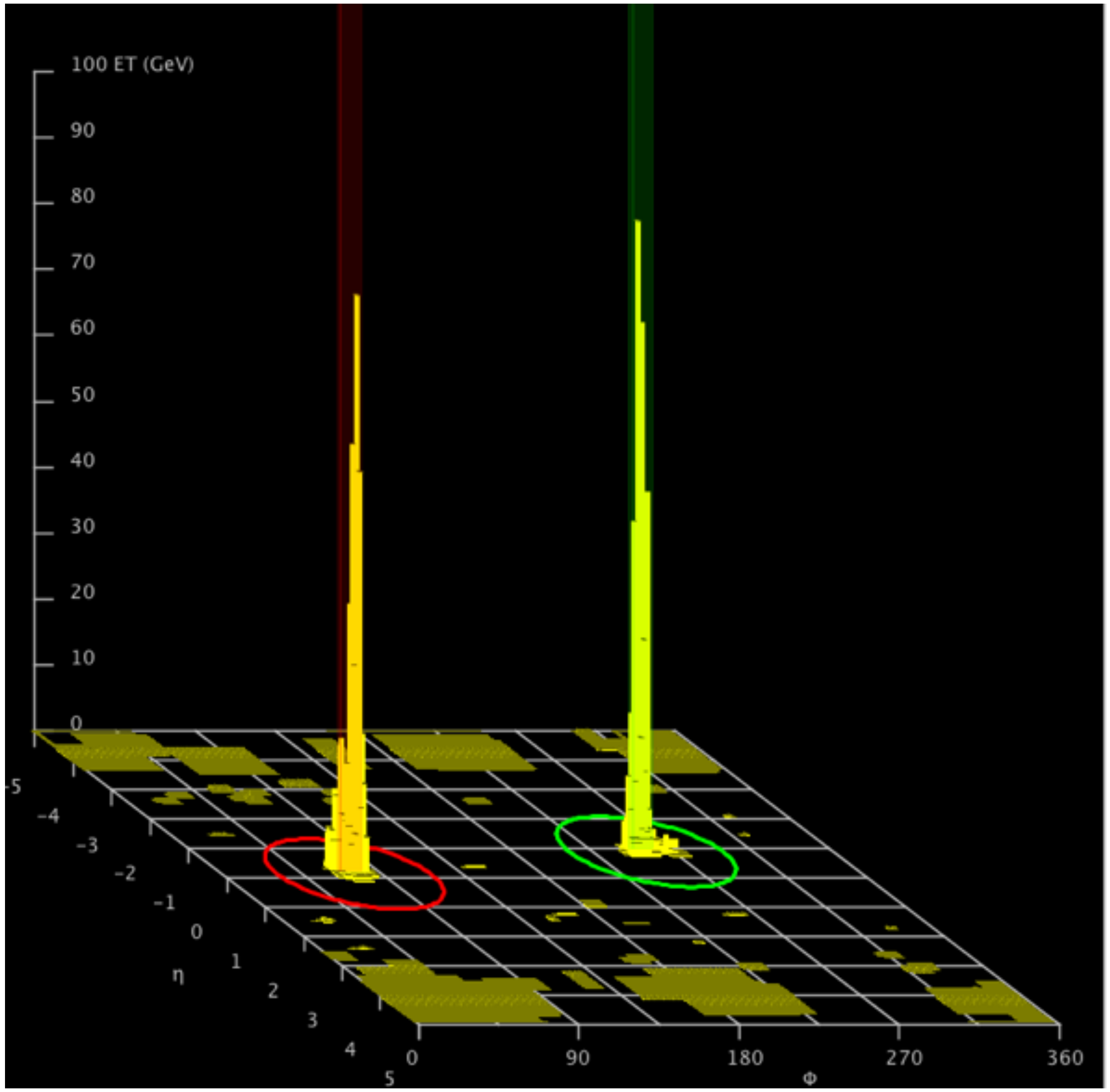
like a digital image!

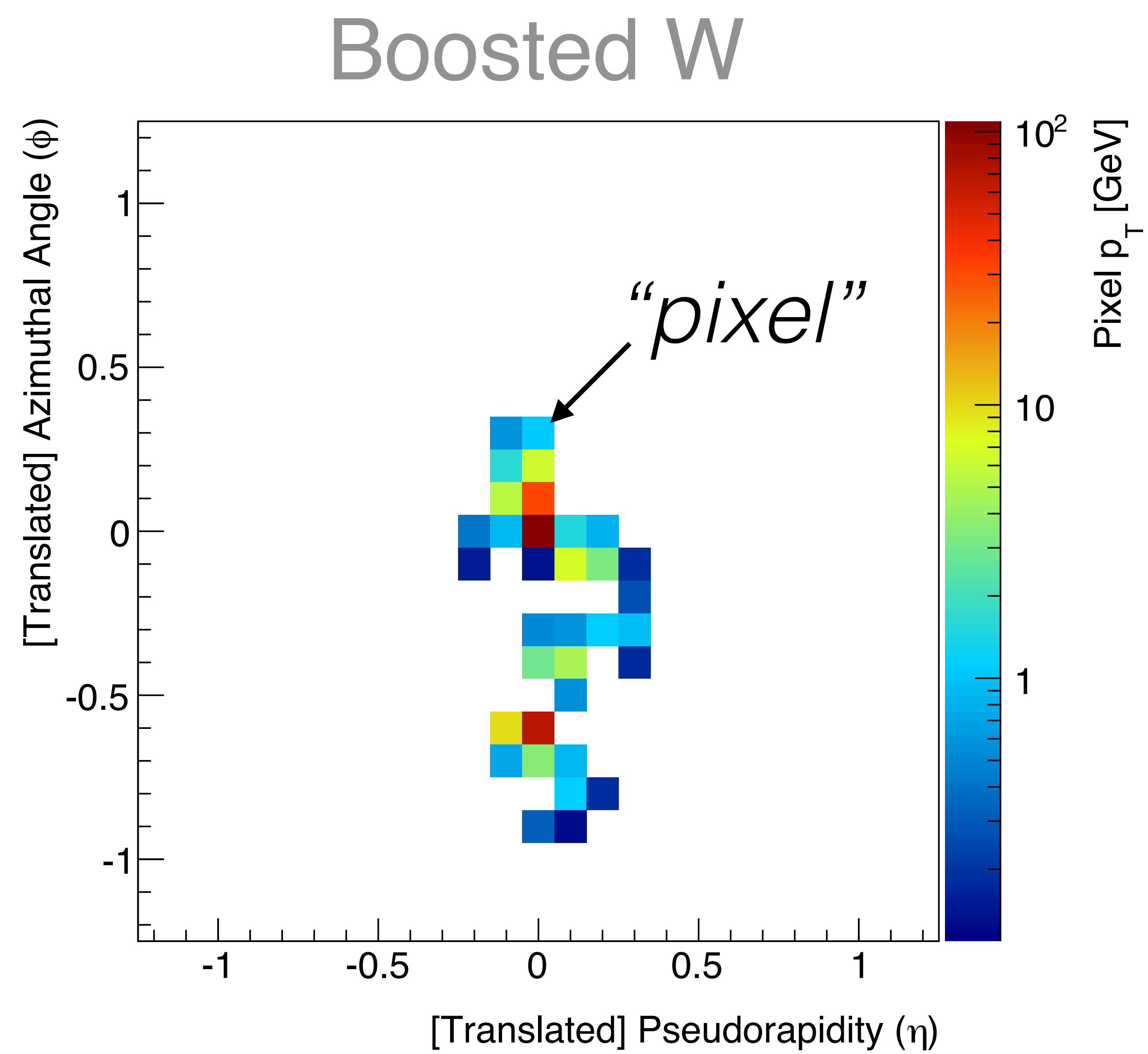




the Jet Image

J. Cogan et al. JHEP 02 (2015) 118



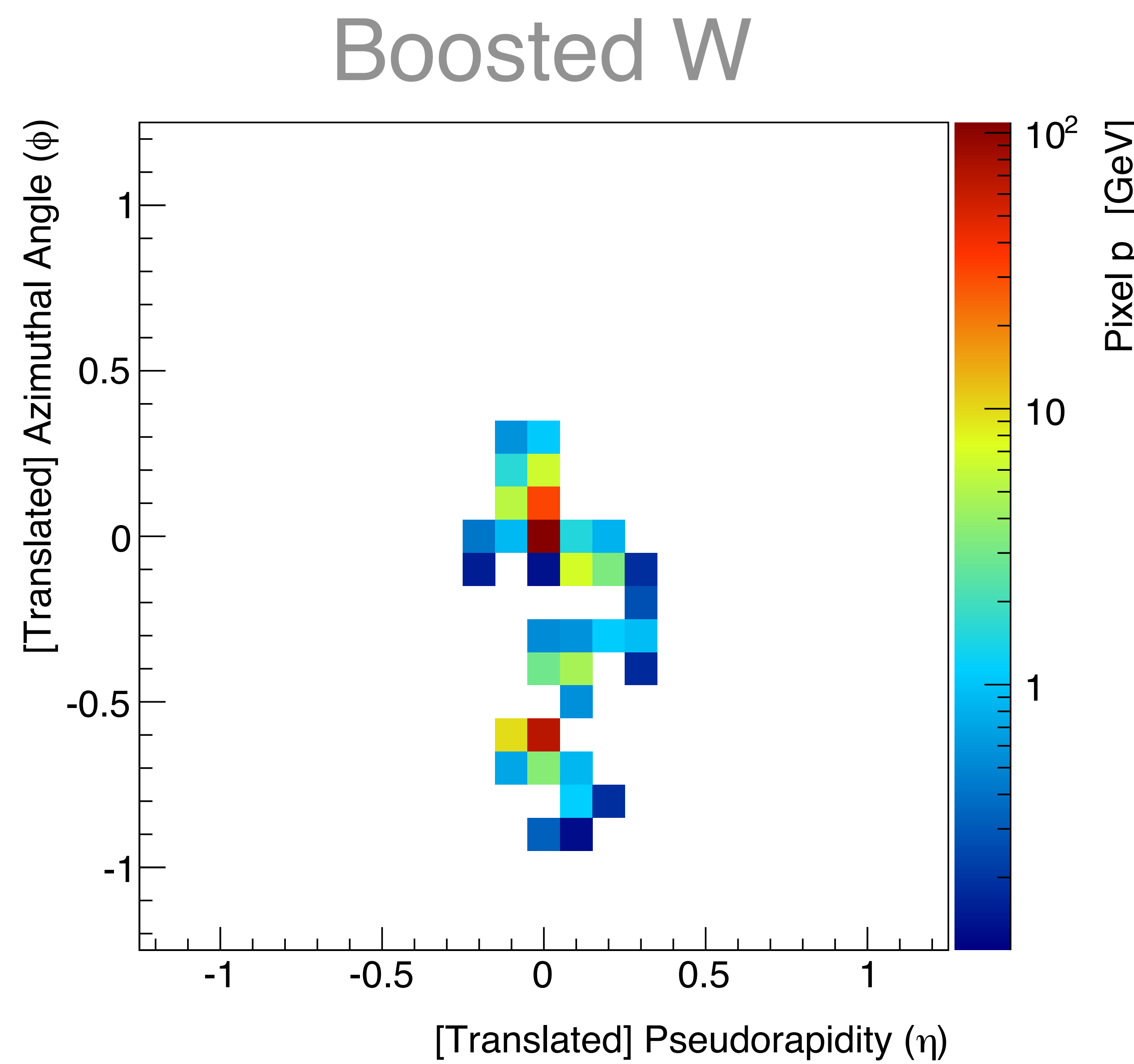


L. de Oliveira, M. Paganini, **BPN**, Comp. and Software for Big Science (2017) 1 *nothing like a 'natural' image!*

the Jet Image

J. Cogan et al. JHEP 02 (2015) 118

Credit: Peter G Trimming (Wikipedia)



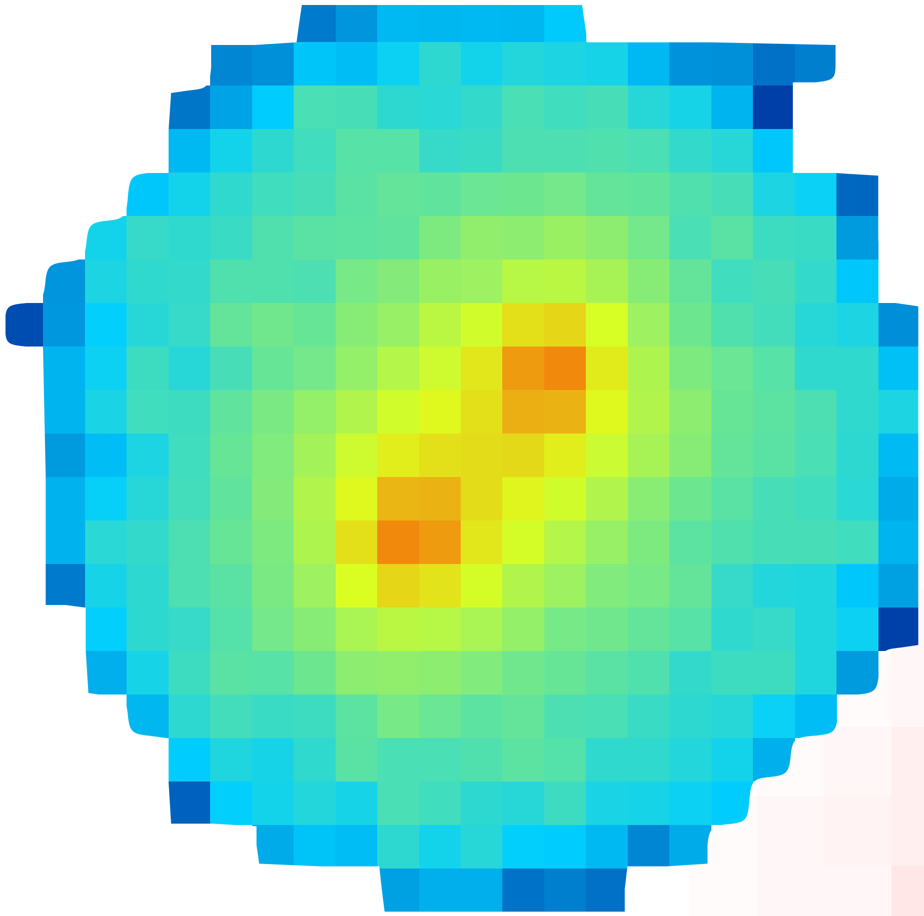
no smooth edges, clear features, low occupancy (number of hit pixels)

L. de Oliveira, M. Paganini, **BPN**, Comp. and Software for Big Science (2017) 1

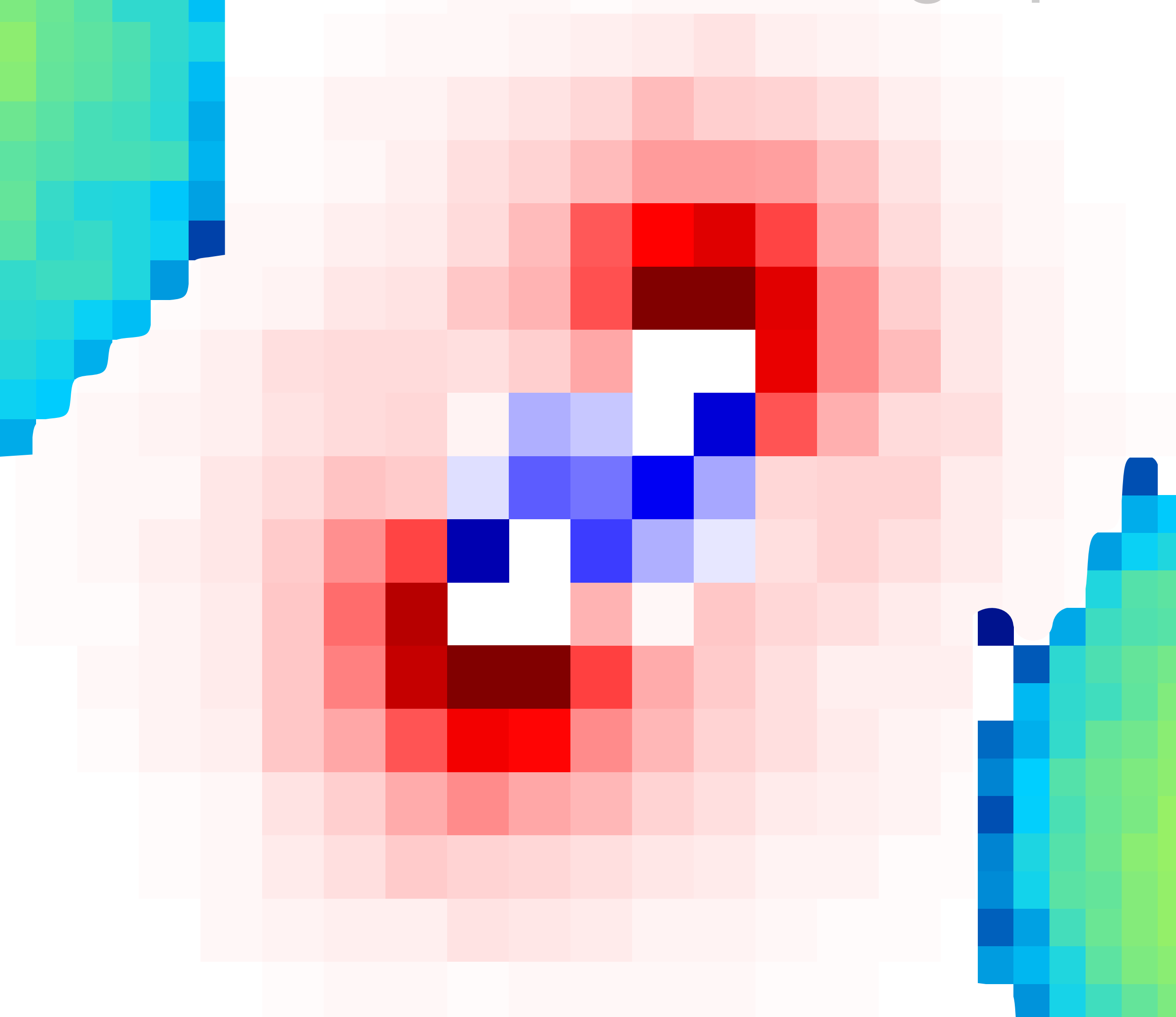
Why images?

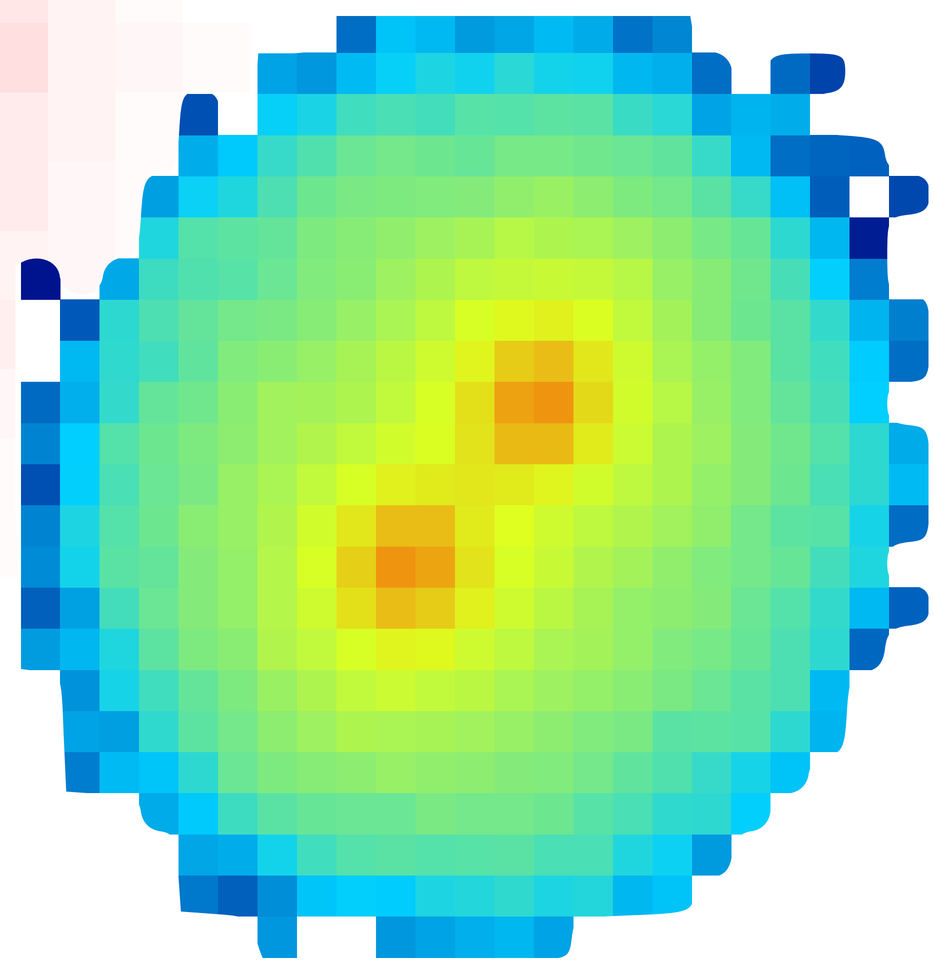
Can directly visualize physics

and we can benefit from the extensive image processing literature



$W \dashrightarrow q\bar{q}$





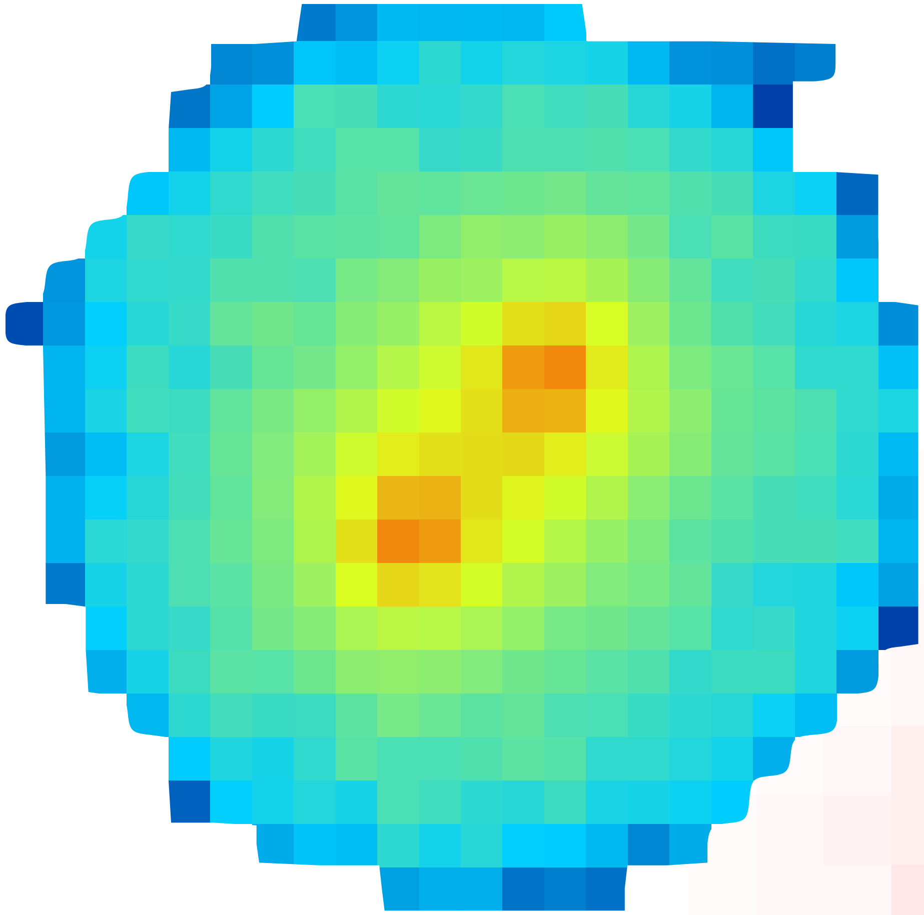
$g \dashrightarrow q\bar{q}$

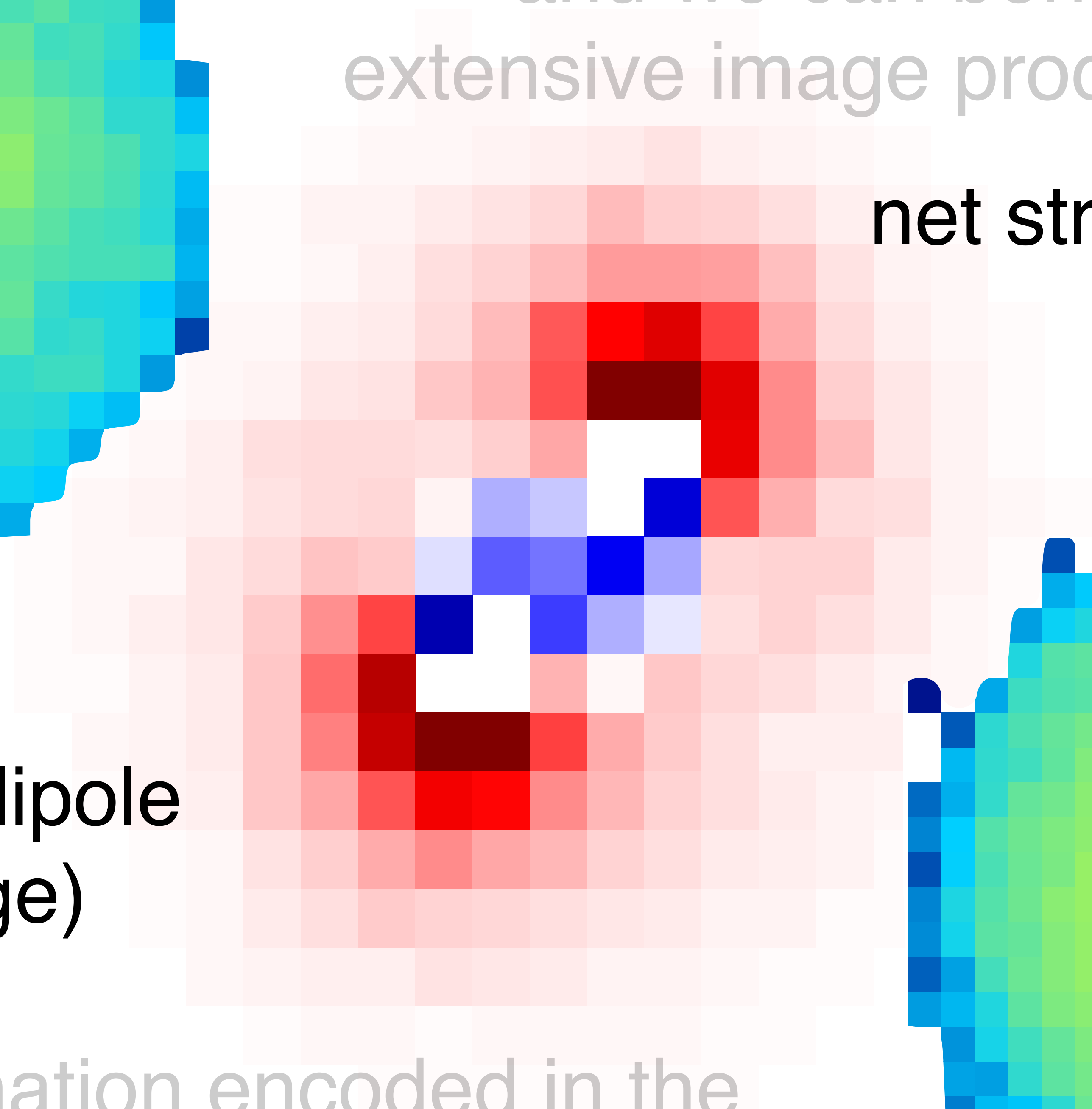
there is information encoded in the physical distance between pixels

Why images?

Can directly visualize physics

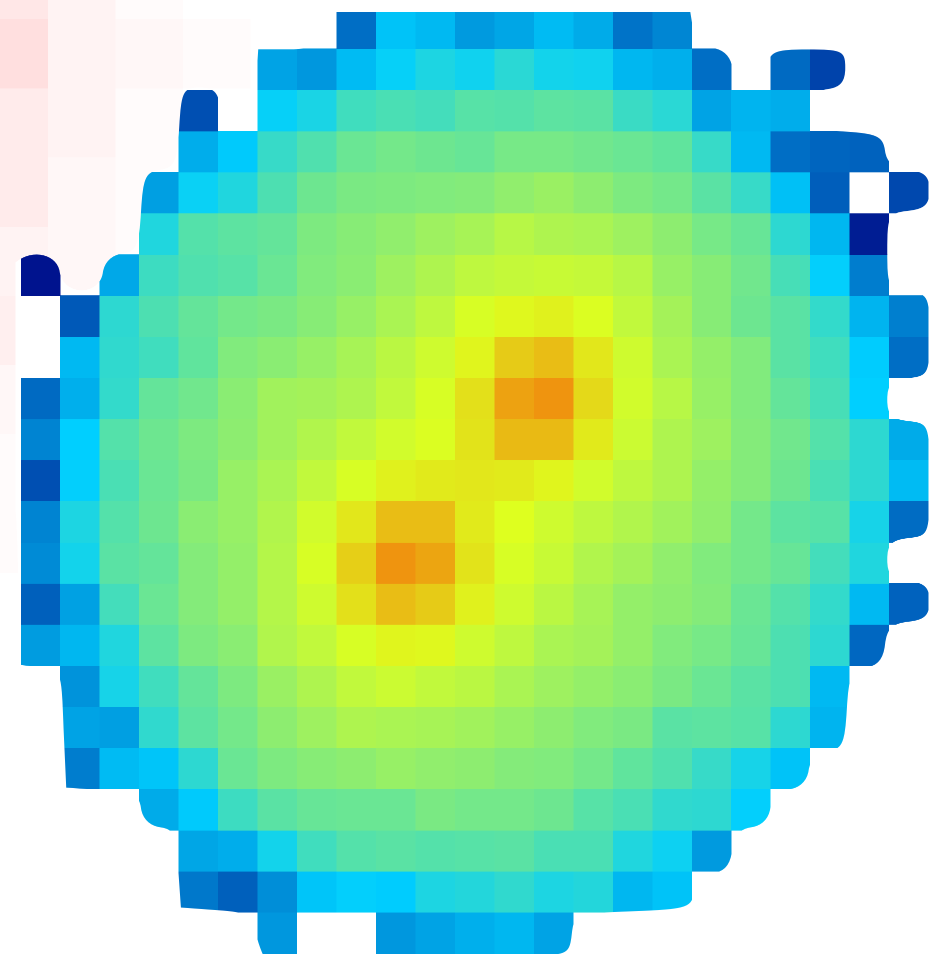
and we can benefit from the extensive image processing literature





net strong-force charge

g $\dashrightarrow$ $q\bar{q}$



W $\dashrightarrow$ $q\bar{q}$

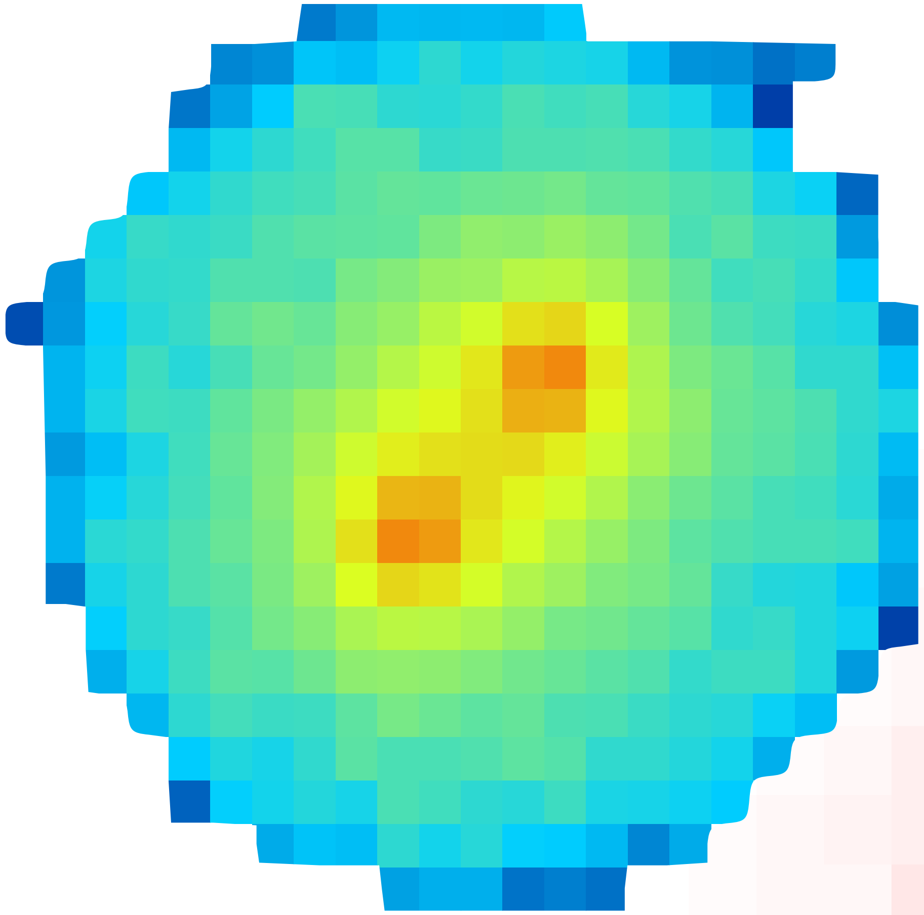
radiates like a dipole (no net charge)

there is information encoded in the physical distance between pixels

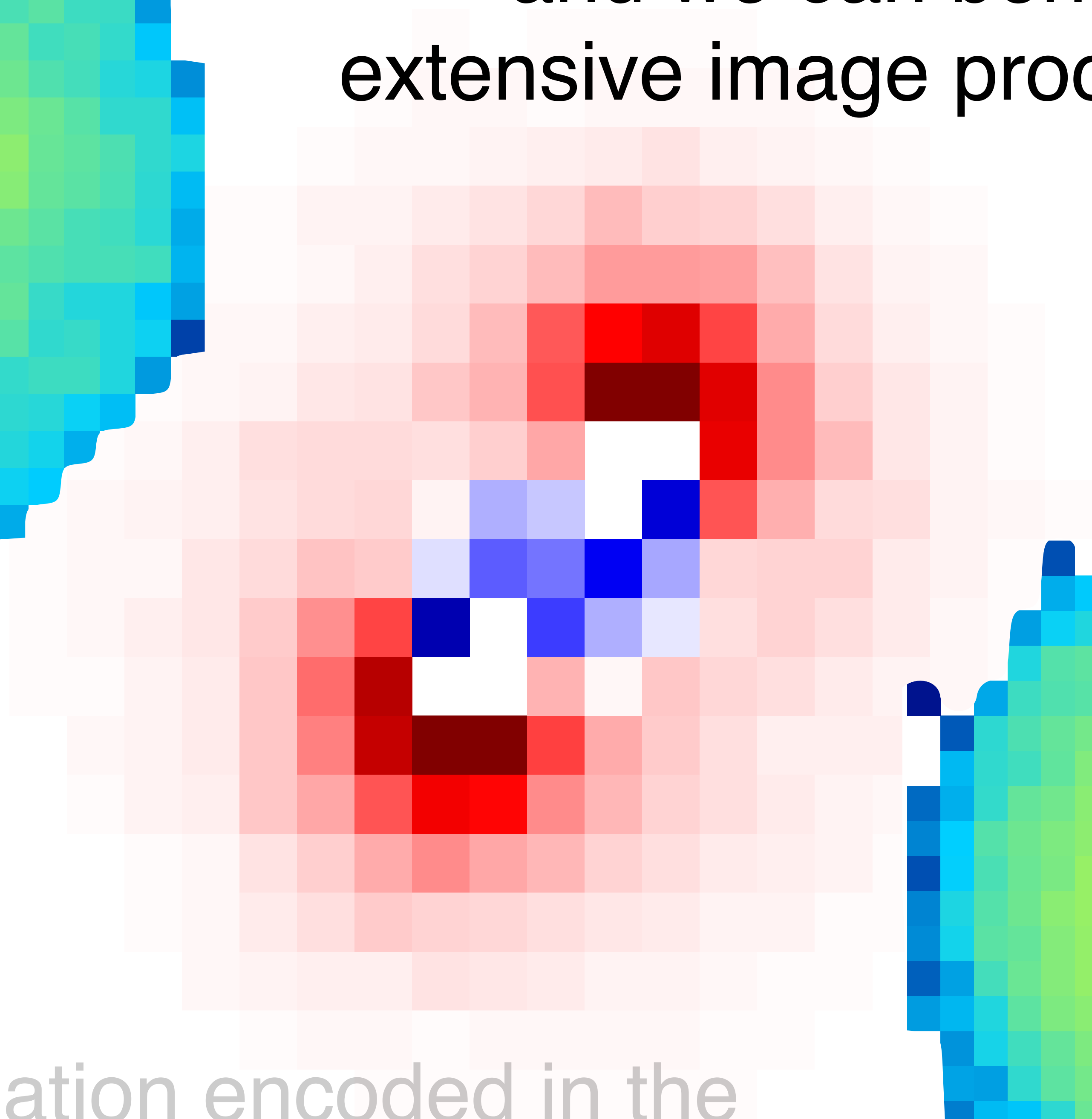
Why images?

Can directly visualize physics

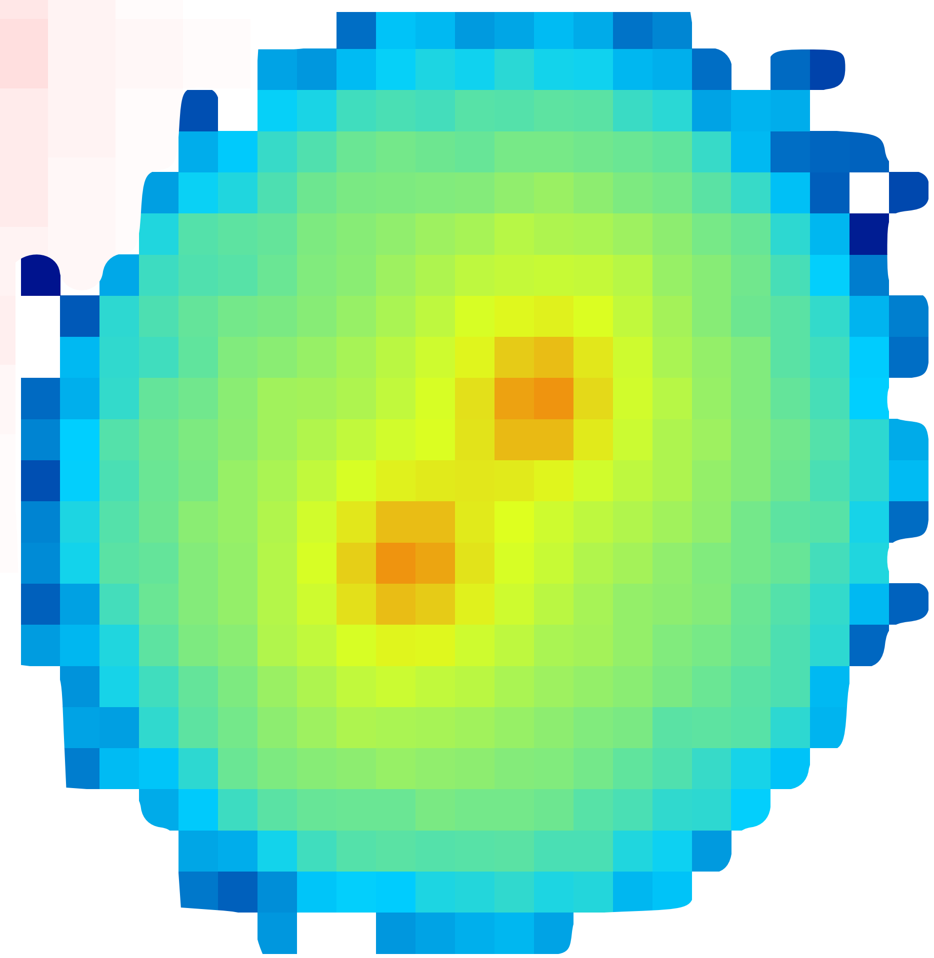
and we can benefit from the extensive image processing literature



$\mathbf{W} \dashrightarrow q\bar{q}$

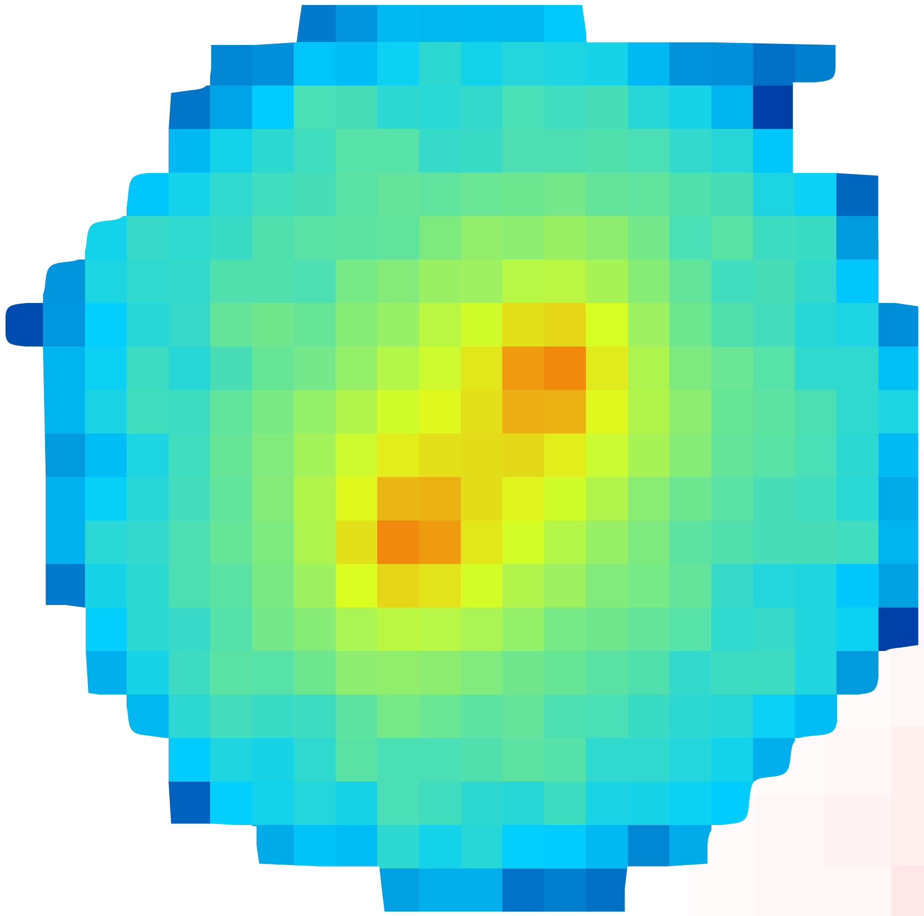


$\mathbf{g} \dashrightarrow q\bar{q}$



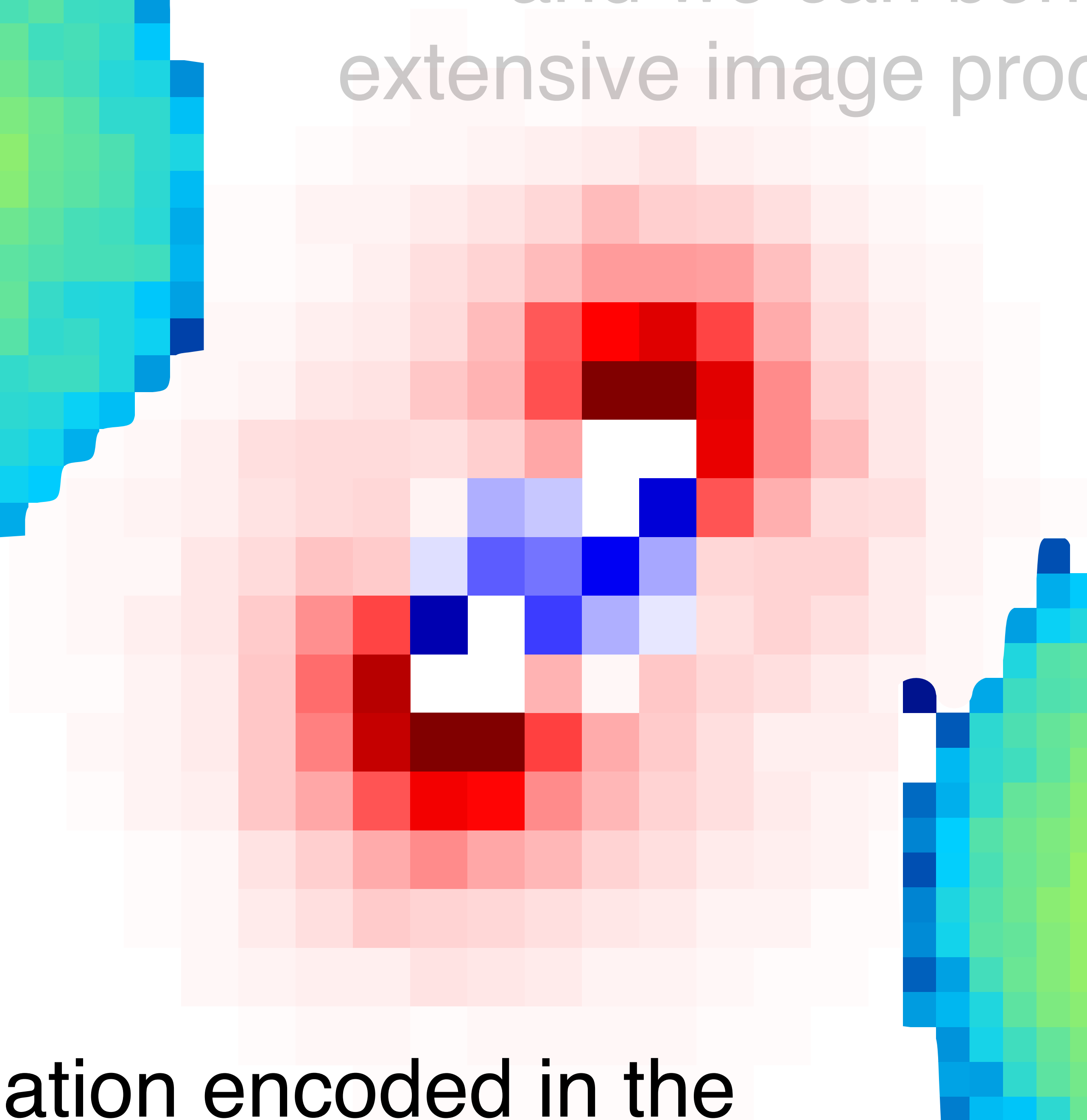
there is information encoded in the physical distance between pixels

Why images?

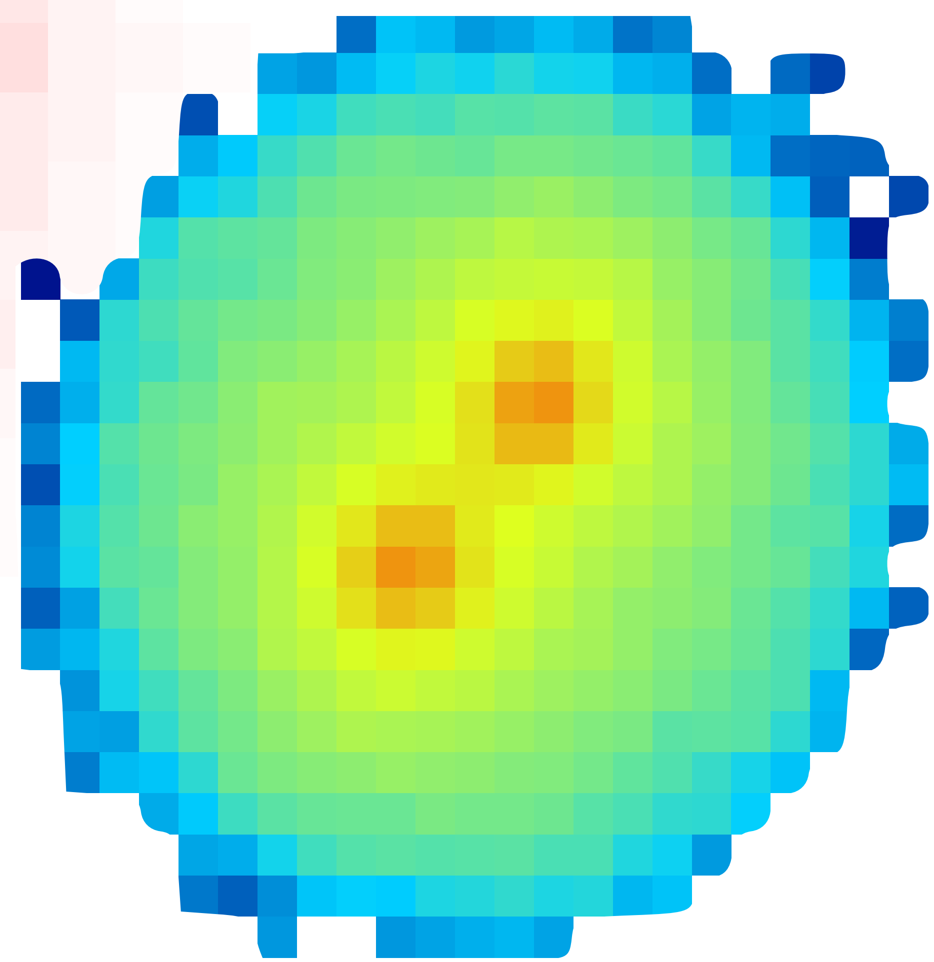


Can directly visualize physics

and we can benefit from the
extensive image processing literature



g $\dashrightarrow$ $q\bar{q}$

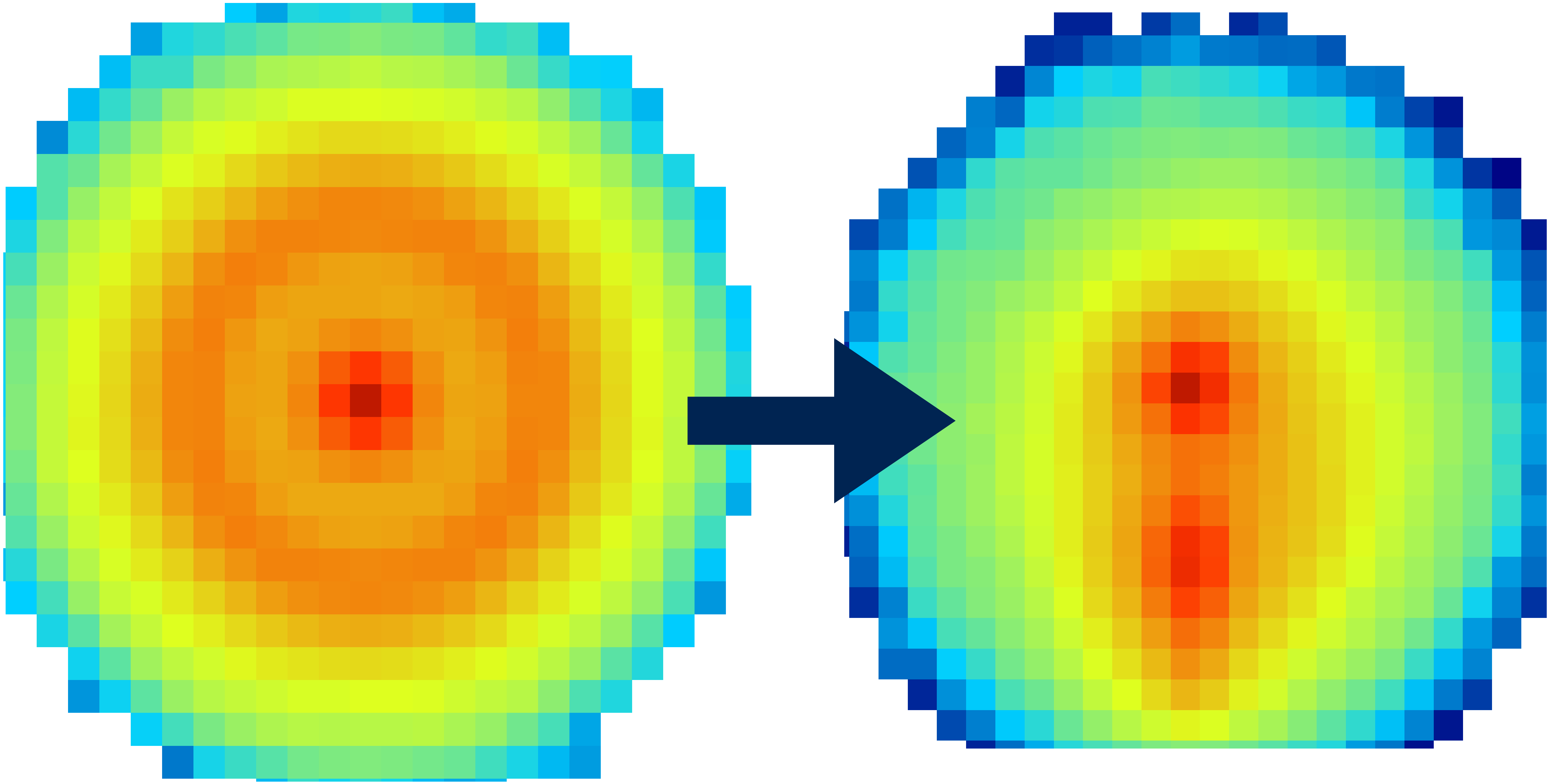


W $\dashrightarrow$ $q\bar{q}$

there is information encoded in the
physical distance between pixels

Pre-processing & spacetime symmetries

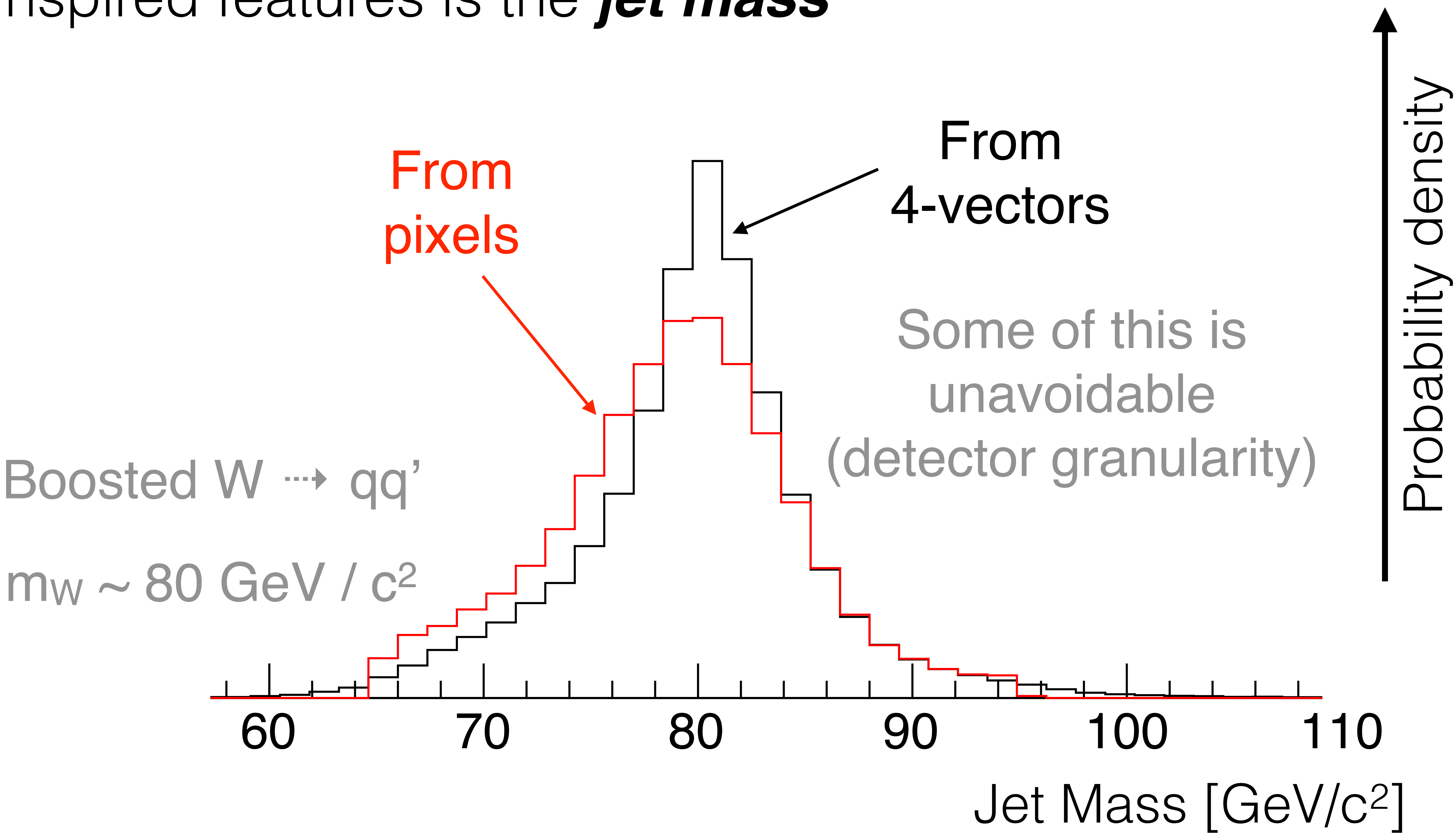
One of the first typical steps is pre-processing



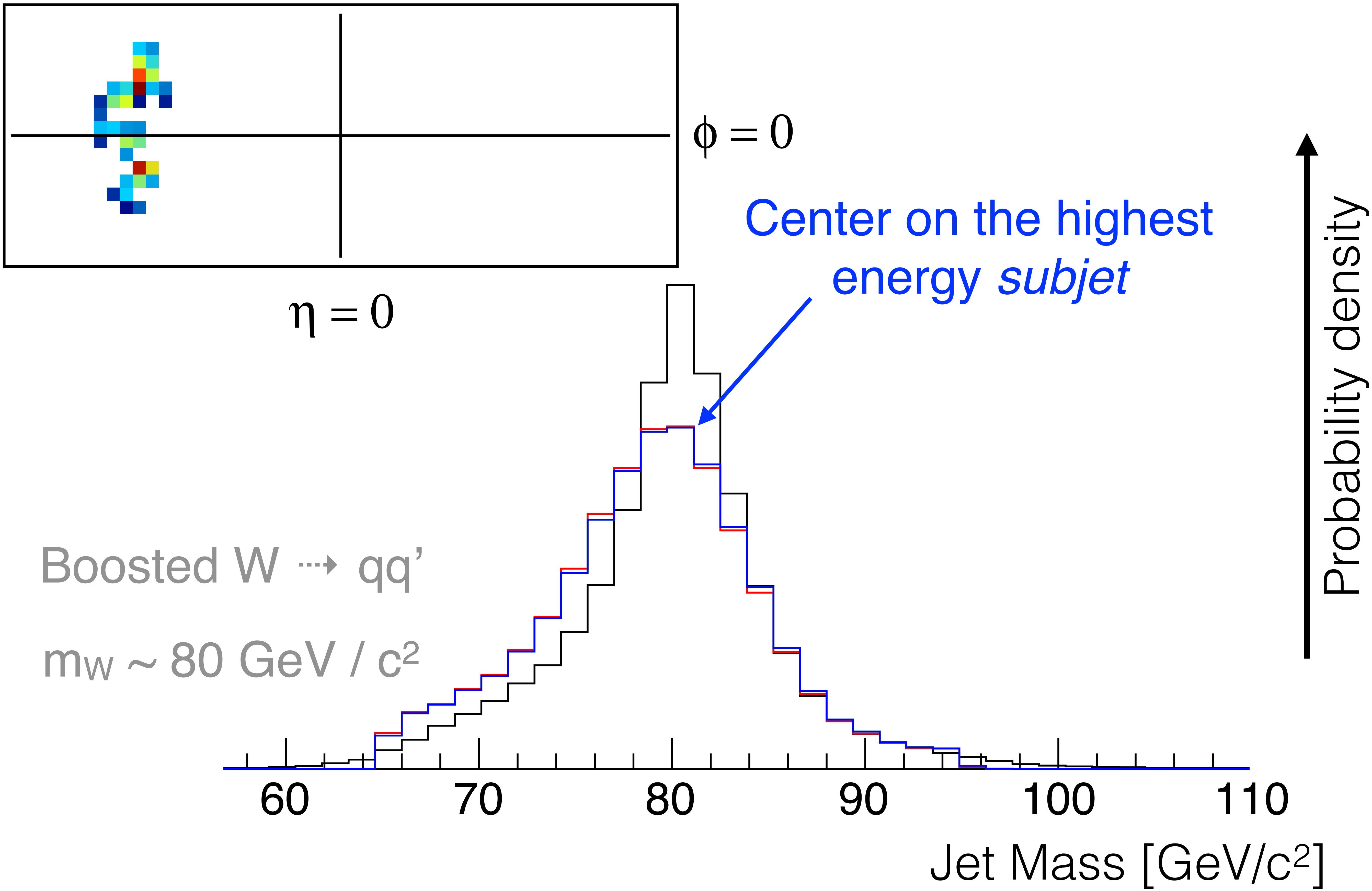
Can help to learn faster & smarter; but must be careful!

Pre-processing & spacetime symmetries

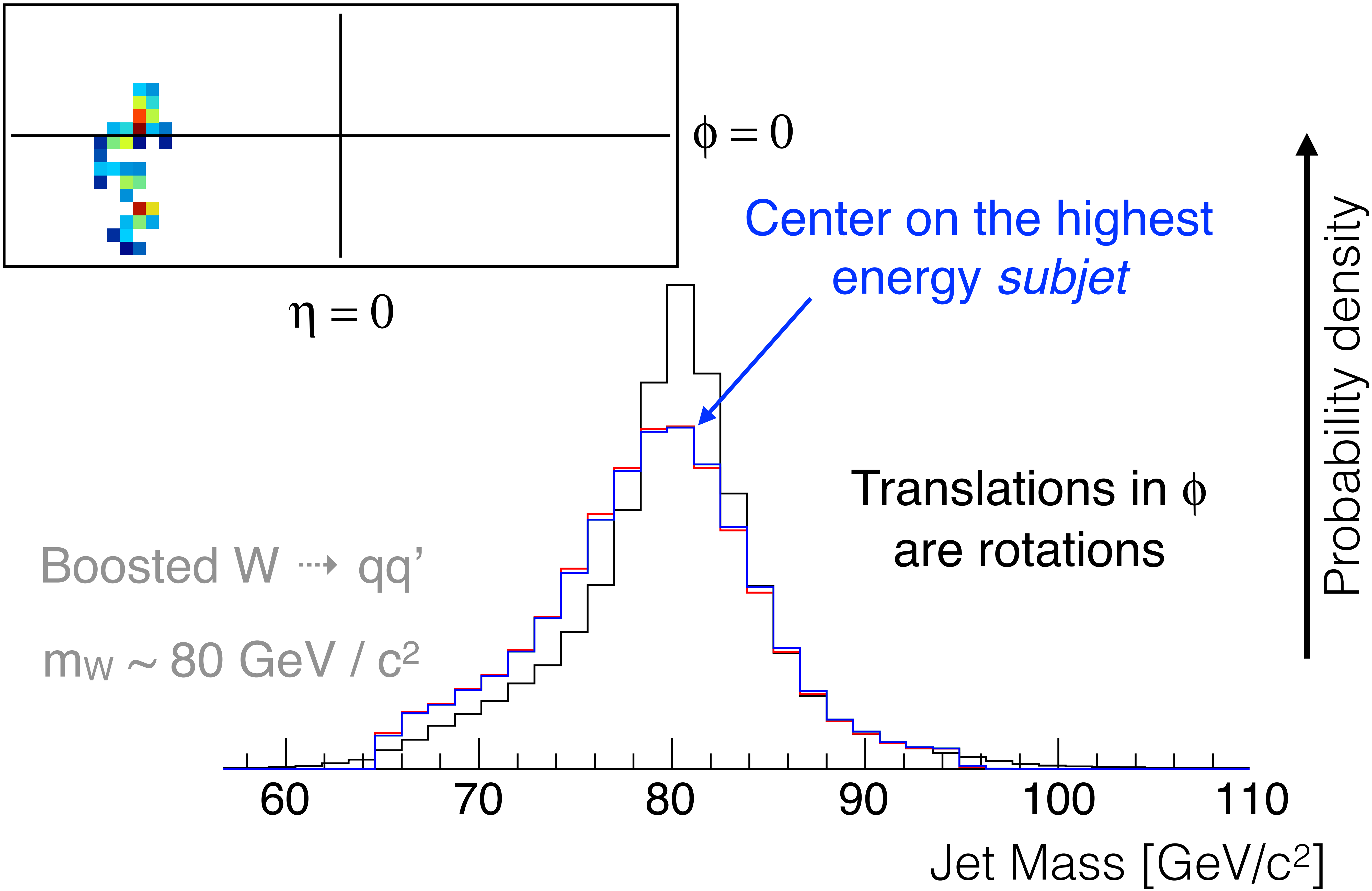
One of the most useful physics-inspired features is the ***jet mass***



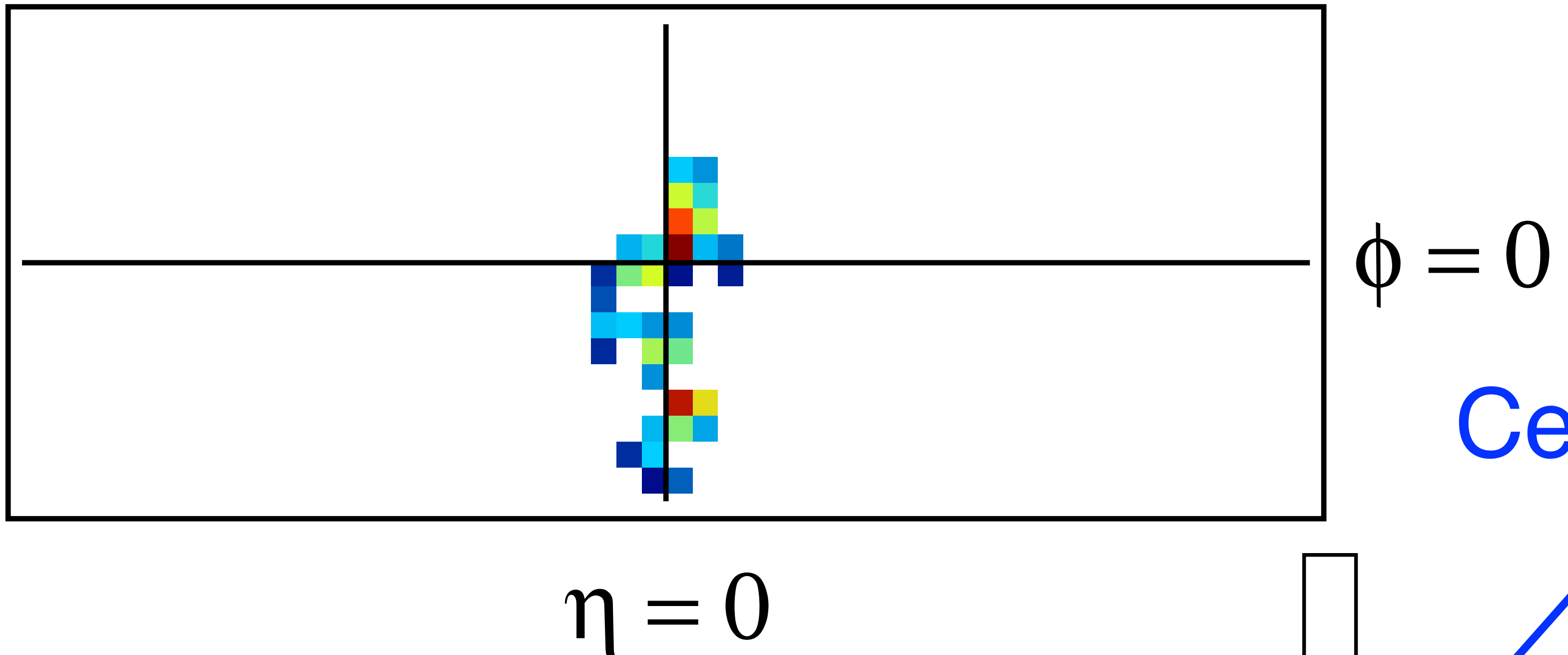
Pre-processing & spacetime symmetries



Pre-processing & spacetime symmetries



Pre-processing & spacetime symmetries

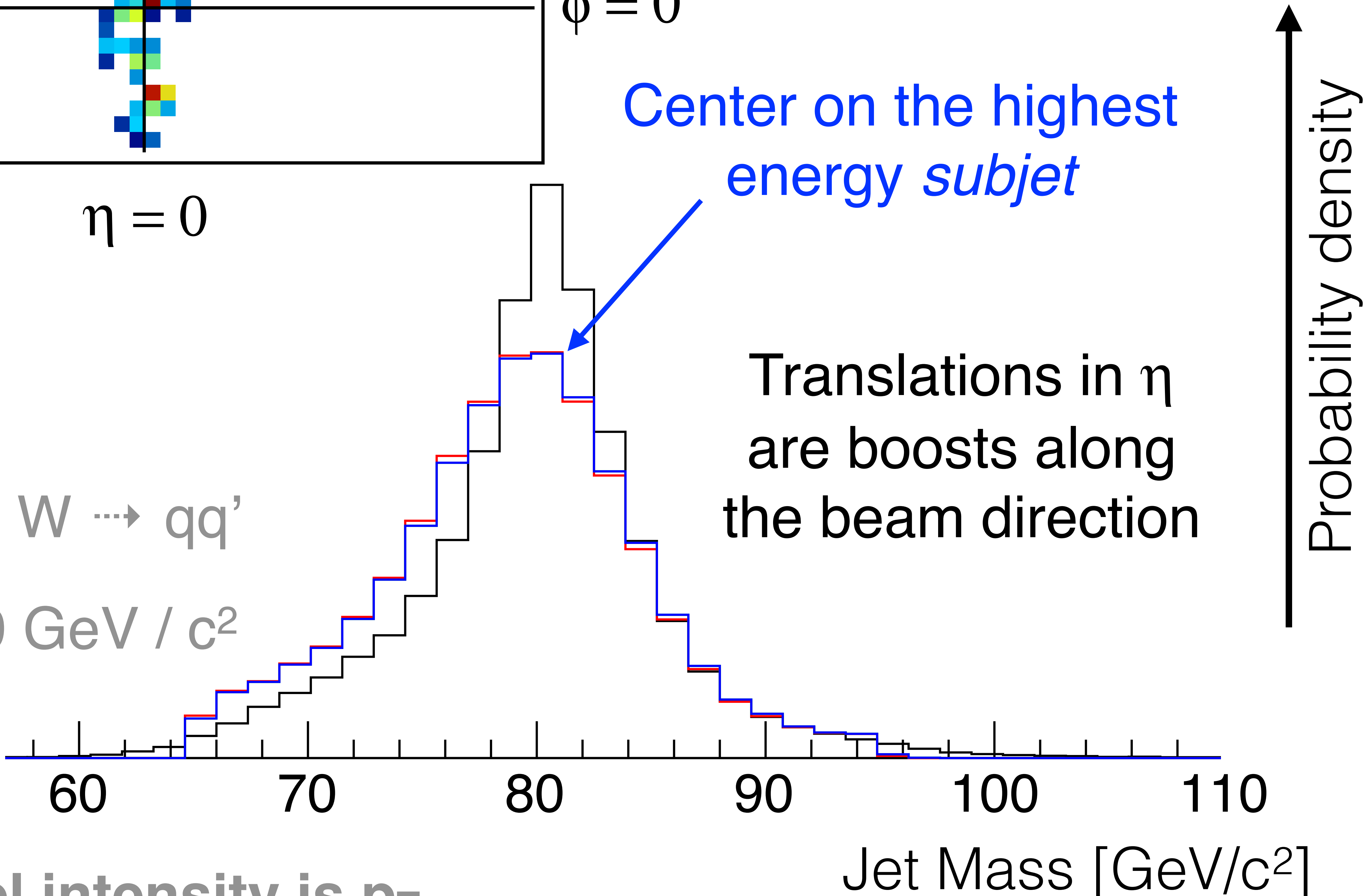


Center on the highest energy *subjet*

Translations in η are boosts along the beam direction

Boosted W ⇢ qq'

$m_W \sim 80$ GeV / c^2



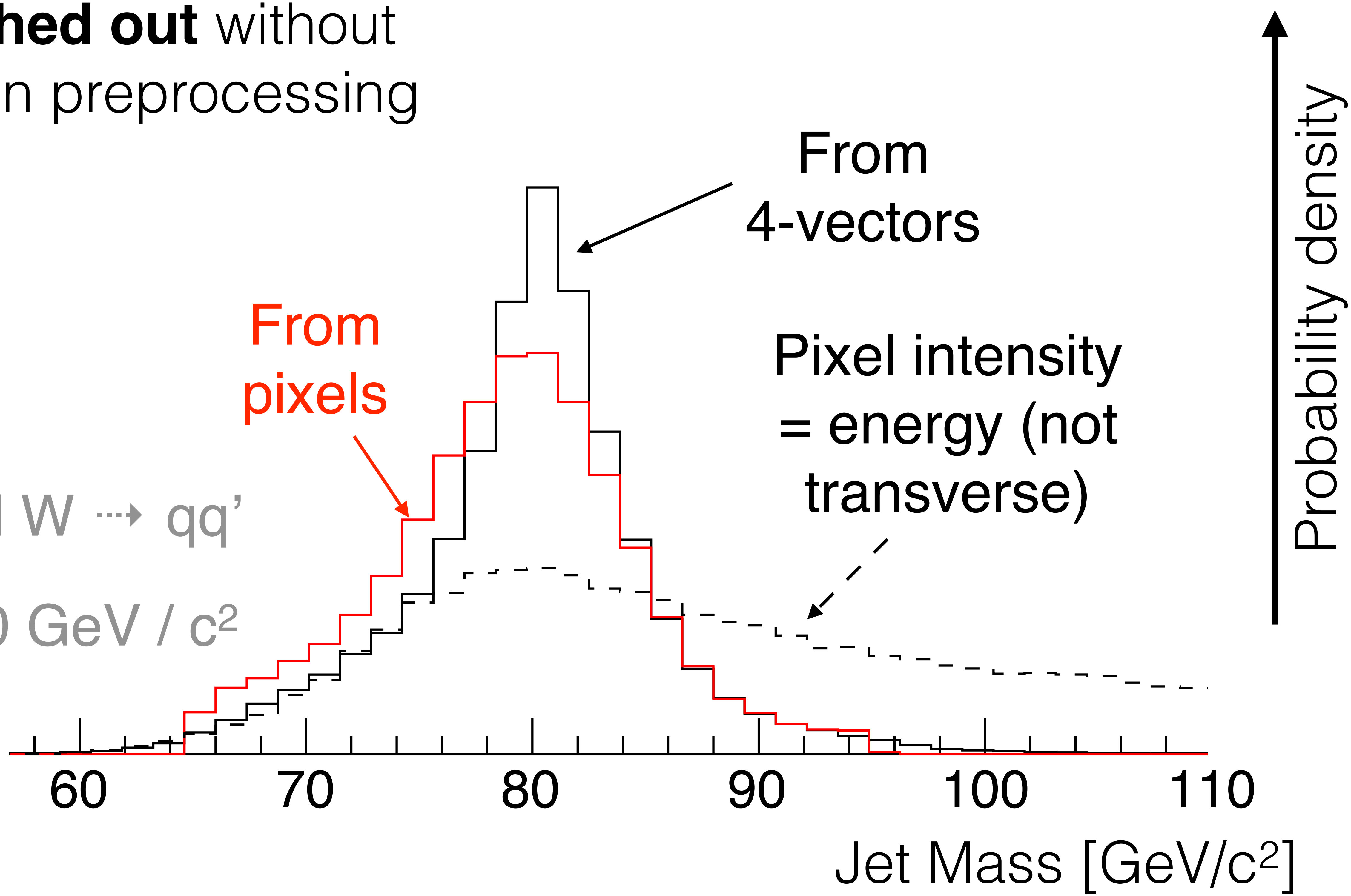
N.B. pixel intensity is p_T

Pre-processing & spacetime symmetries

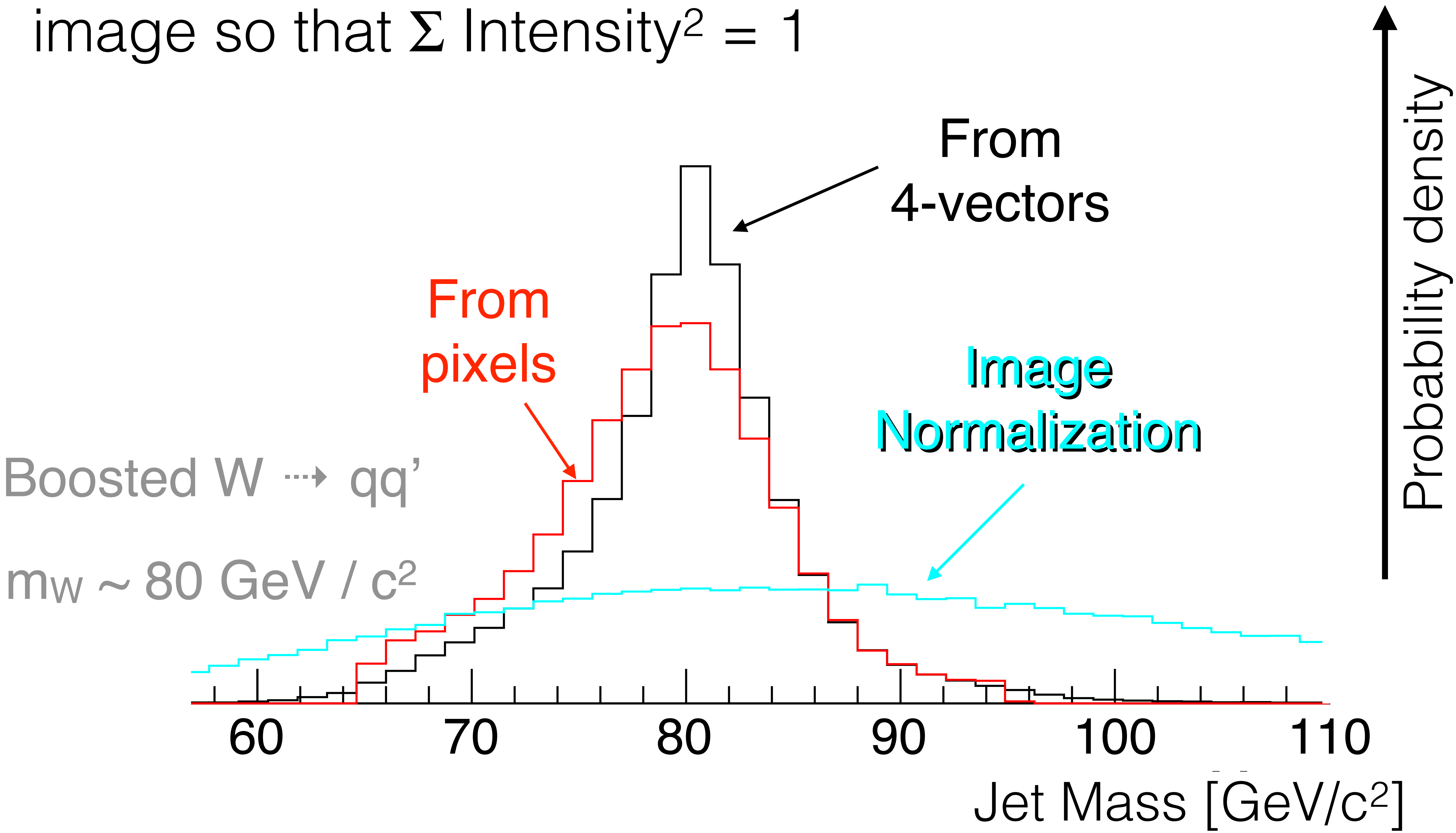
Information can be **washed out** without care in preprocessing

Boosted W ⇢ qq’

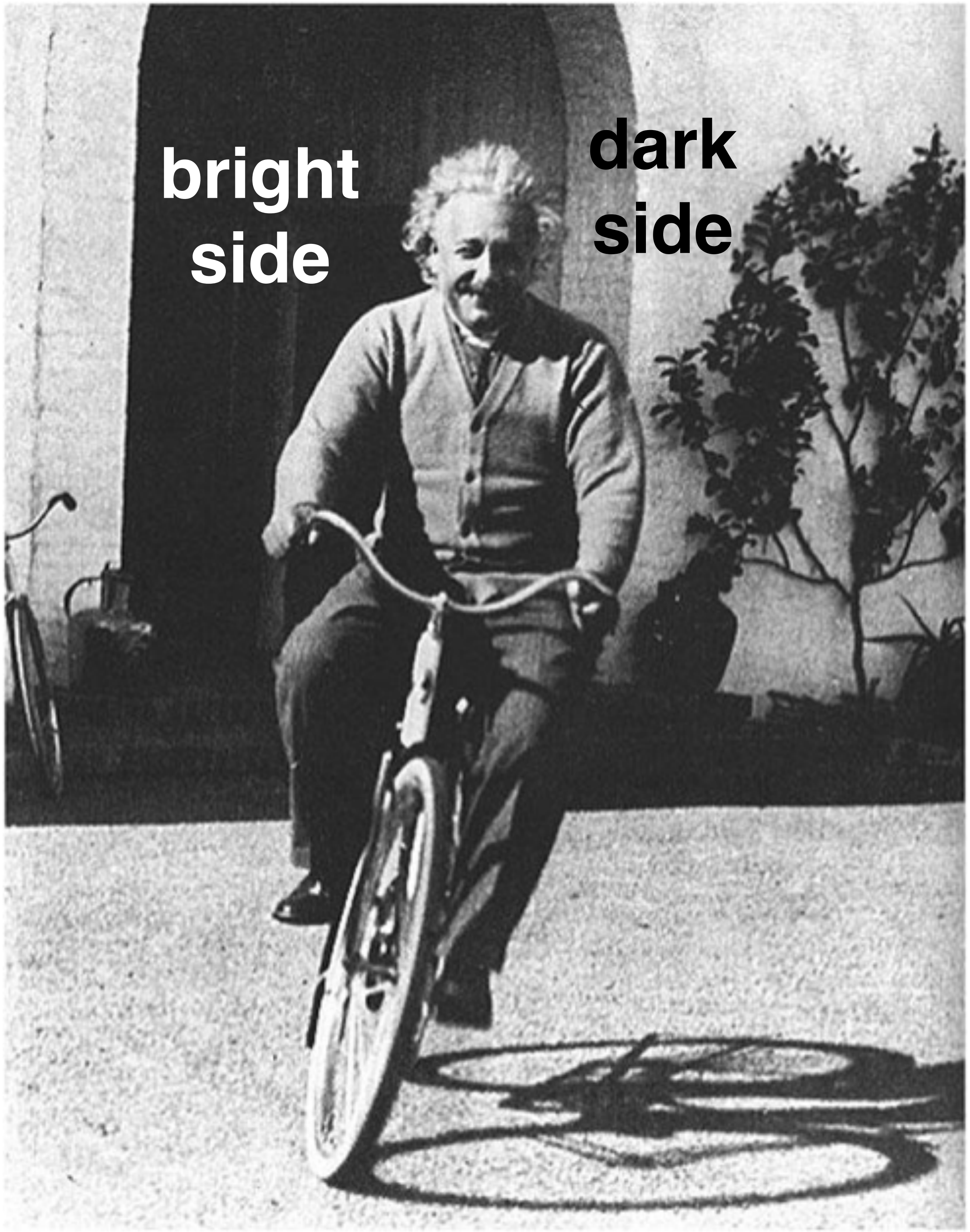
$m_W \sim 80$ GeV / c^2



It is common to normalize each image so that Σ Intensity^2 = 1



Intuition via analogy *why normalization can hurt*



In both pictures, total intensity of Einstein's face is about the same.

However, his face's **image mass** is quite different!

Photos from: http://mentalfloss.com/article/49222/11-unserious-photos-albert-einstein

Intuition via analogy *why normalization can hurt*

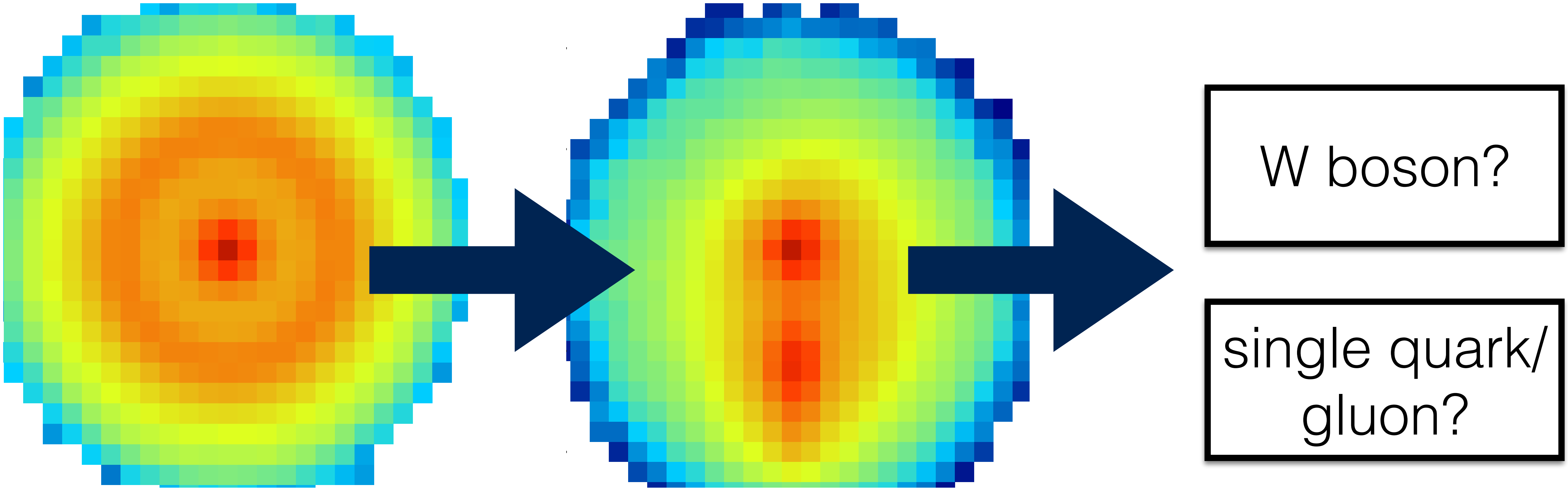
In standard computer vision, you likely don't want to be sensitive to this! …not the case for jet images!

In both pictures, total intensity of Einstein's face is about the same.

However, his face's **image mass** is quite different!

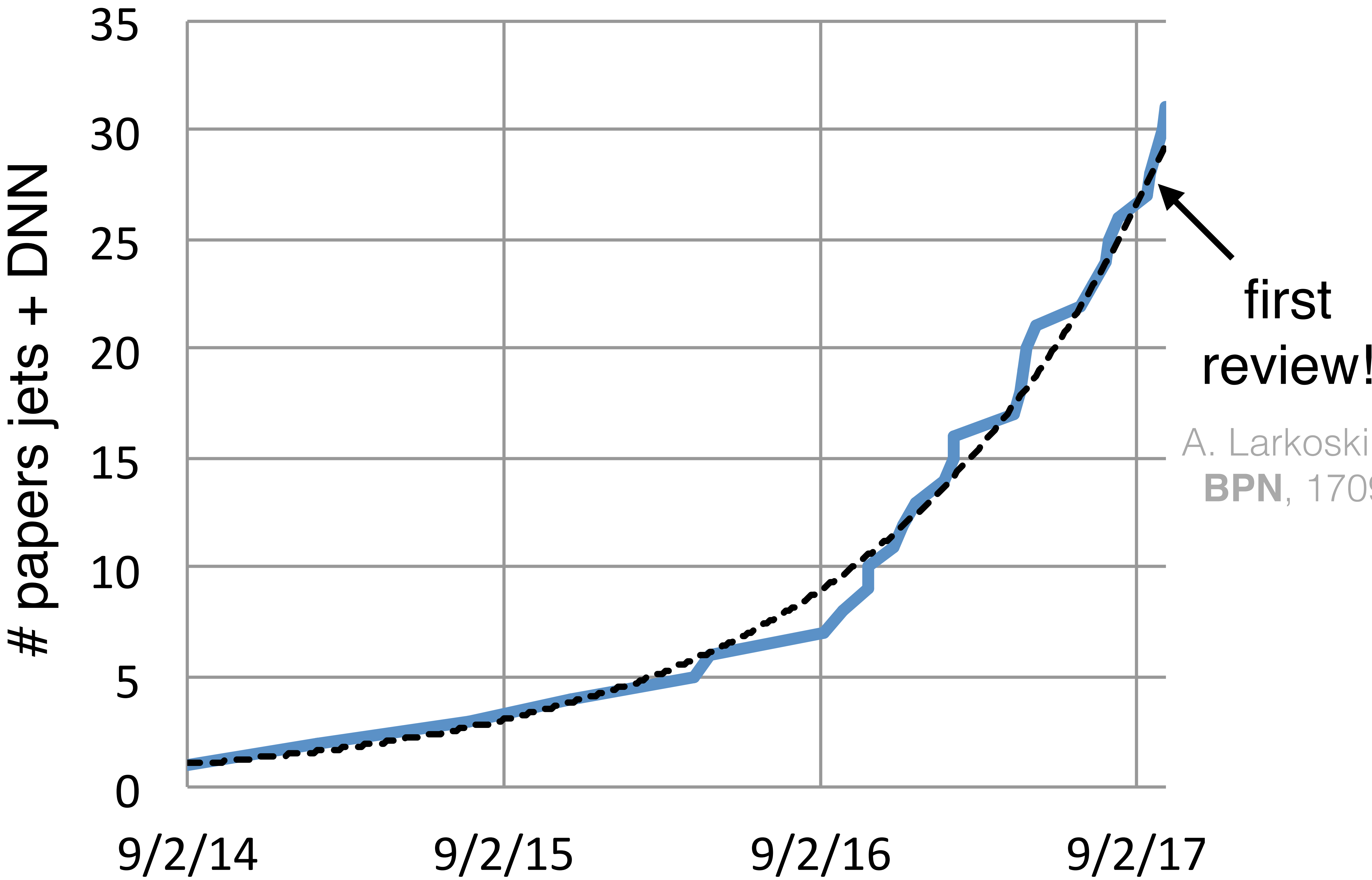
Photos from: http://mentalfloss.com/article/49222/11-unserious-photos-albert-einstein

Now, with a carefully processed image, we can ask: where did this jet come from?



ultimate classification is achieved with modern machine learning using **all pixels as input**!

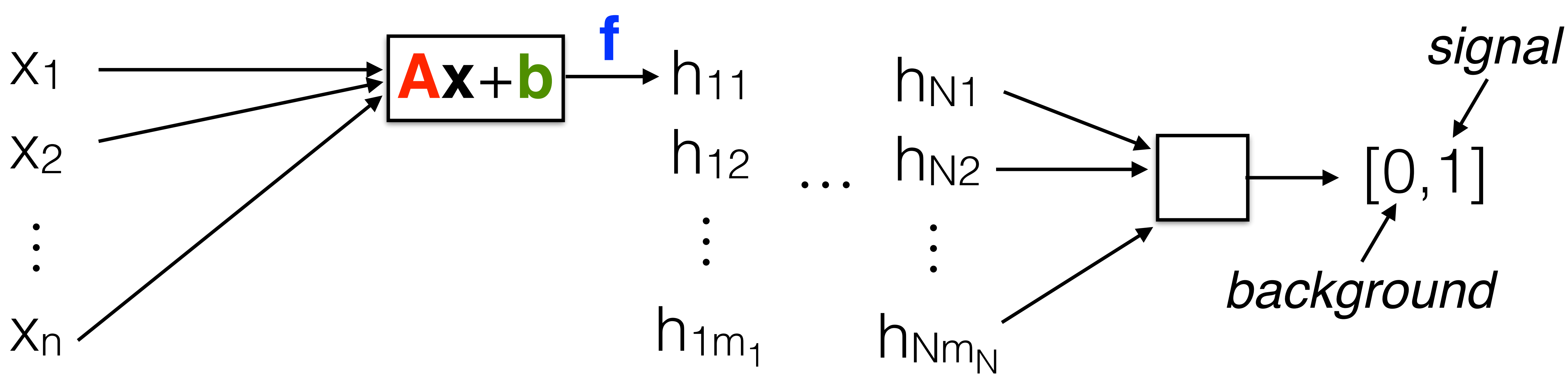
Deep learning with jets



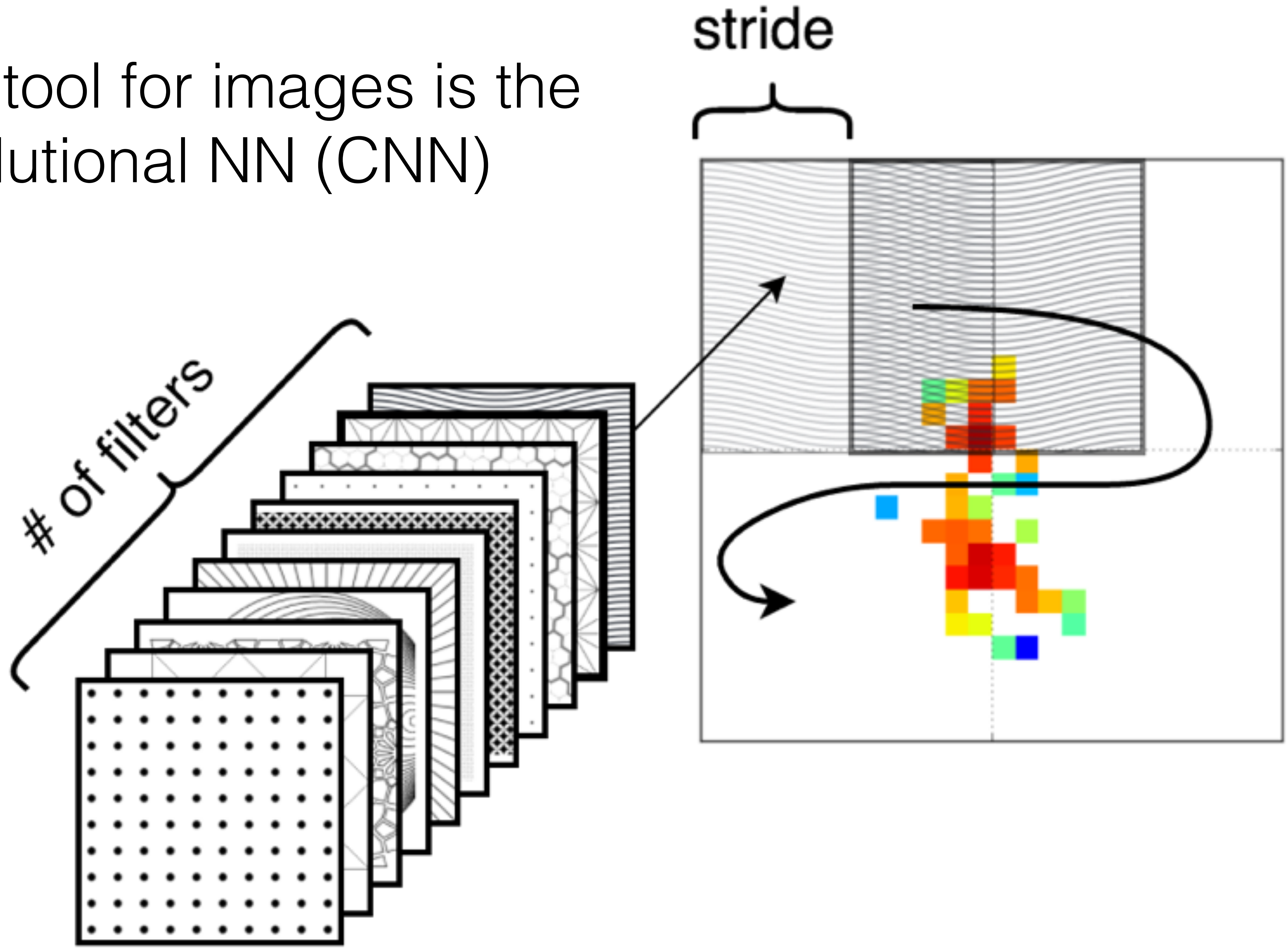
A. Larkoski, I. Moult, **BPN**, 1709.04464

(Deep) Neural Networks

Typical ‘fully connected’ network:

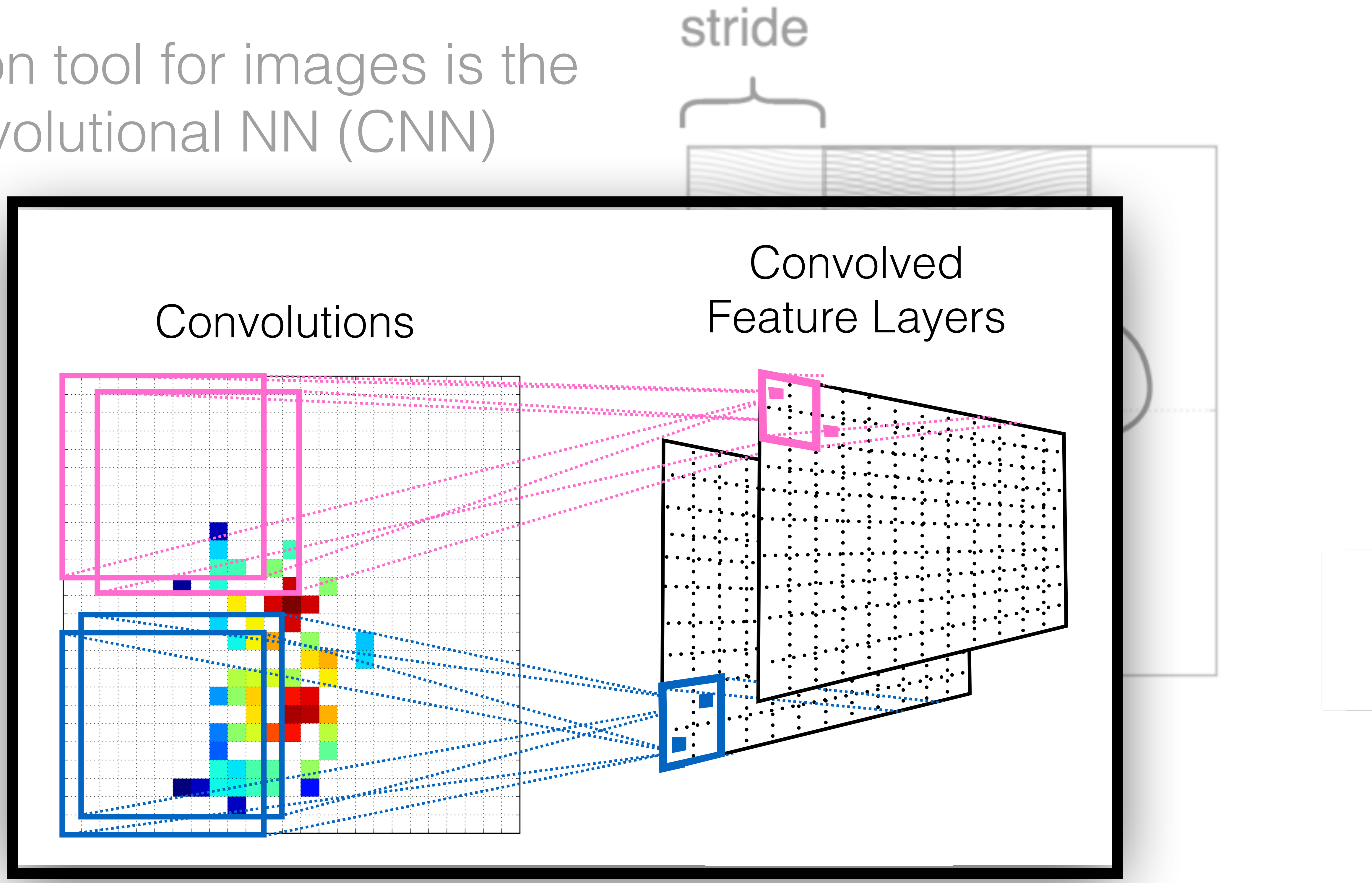


Common tool for images is the convolutional NN (CNN)



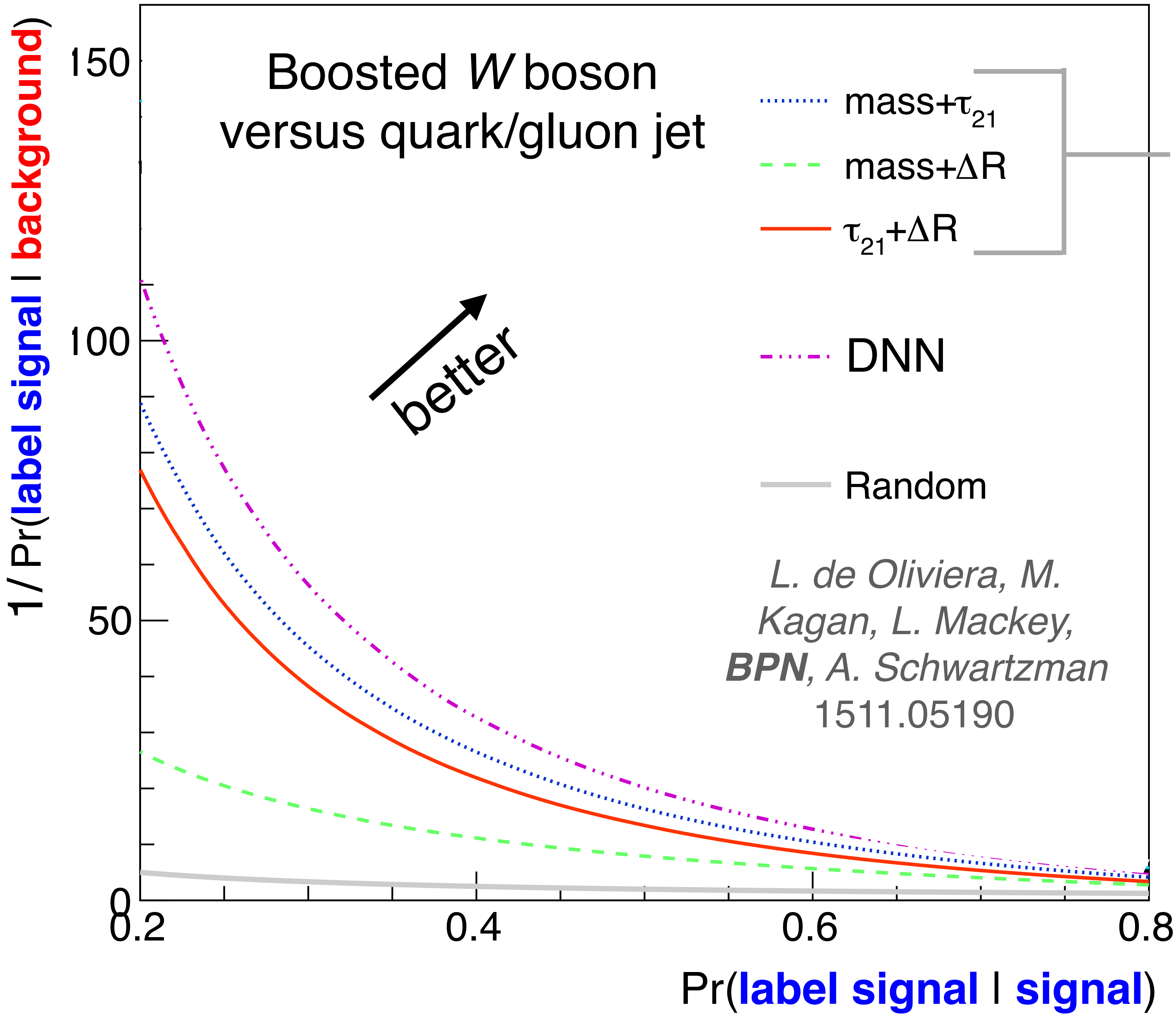
The filter is like the **A**, only the dimensionality is now the filter size (<< n) and not the image size (n).

Common tool for images is the convolutional NN (CNN)



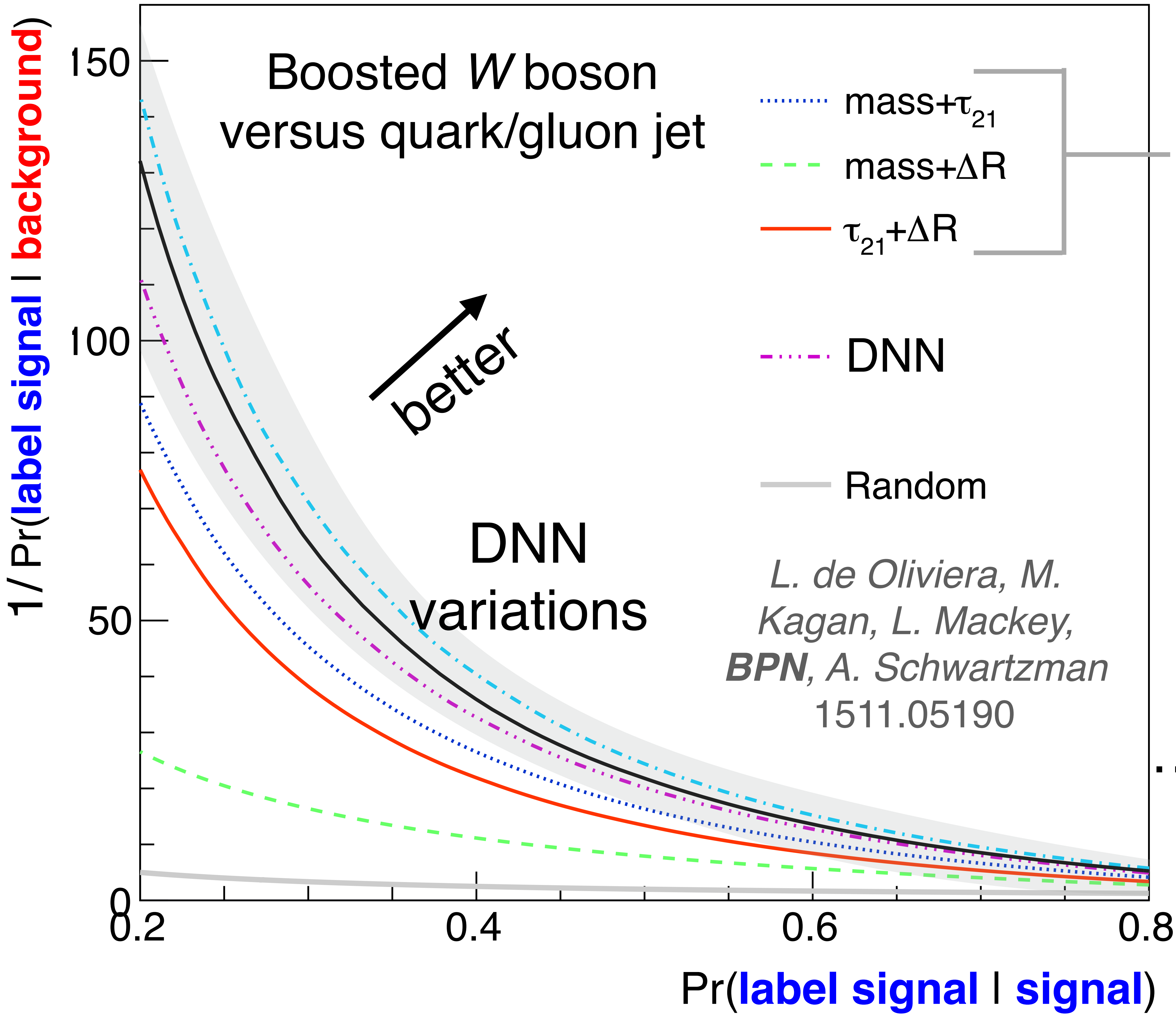
The filter is like the **A**, only the dimensionality is now the filter size (<< n) and not the image size (n).

Modern Deep NN's for Classification



mass, τ_{21}, ΔR are all simple functions of the image

Modern Deep NN's for Classification

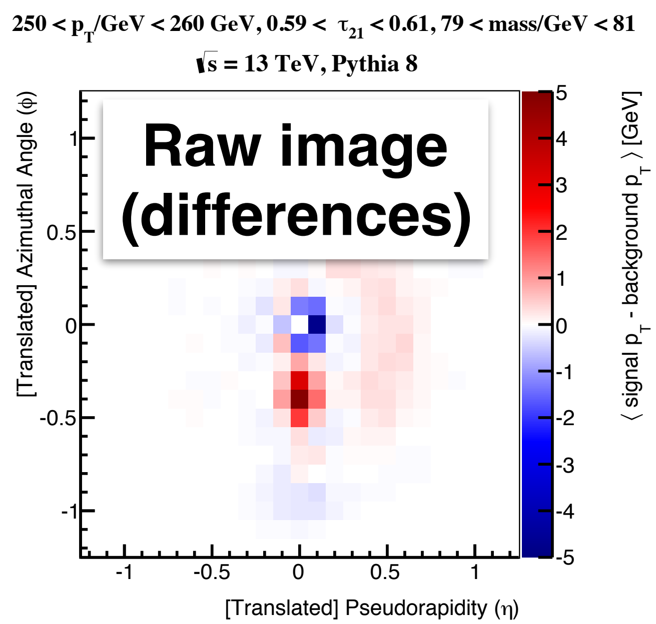


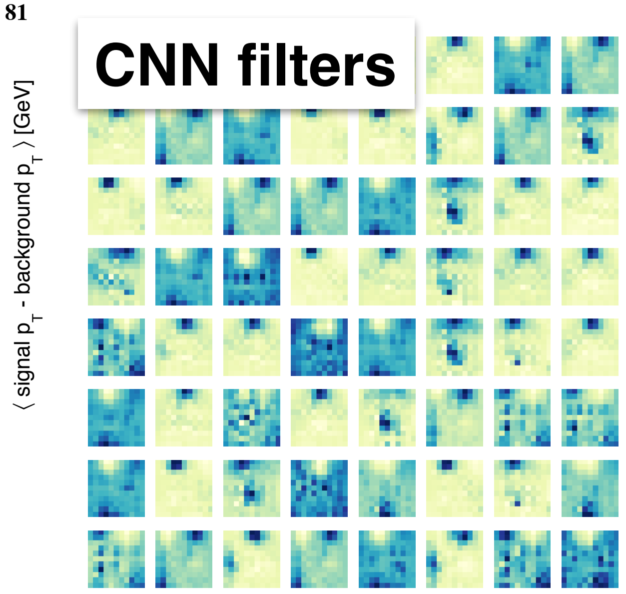
mass, τ_{21}, ΔR are all simple functions of the image

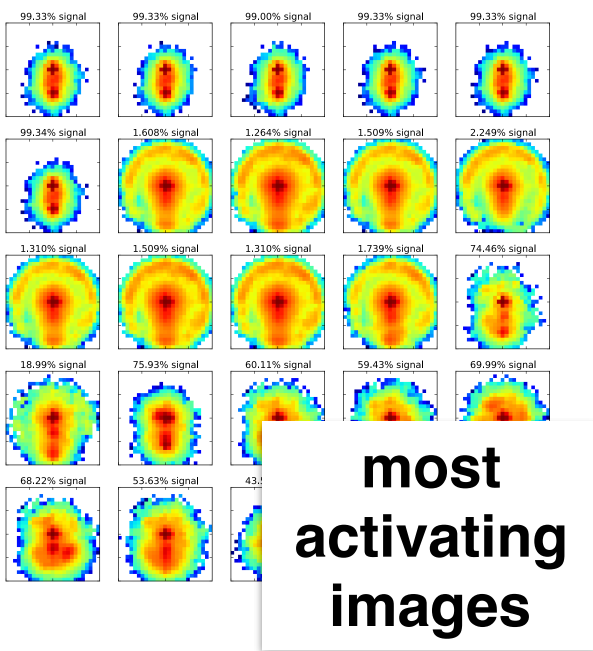
…what the DNN is learning is active R&D!

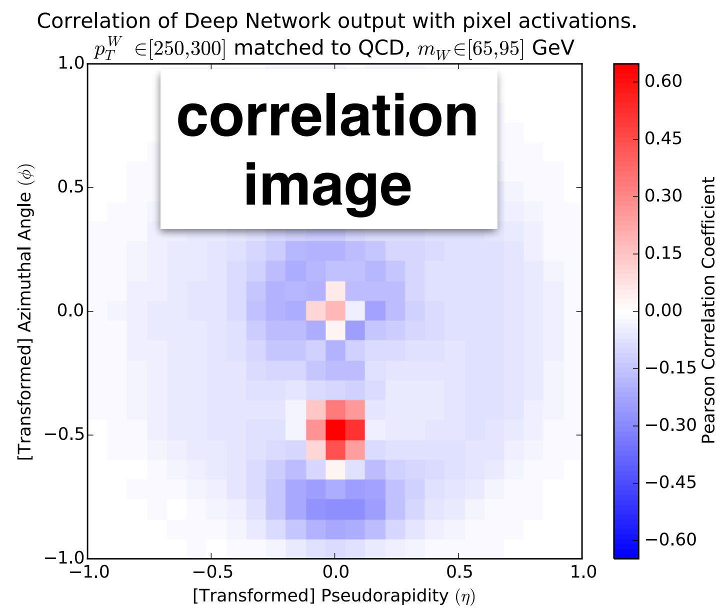
Learning about learning

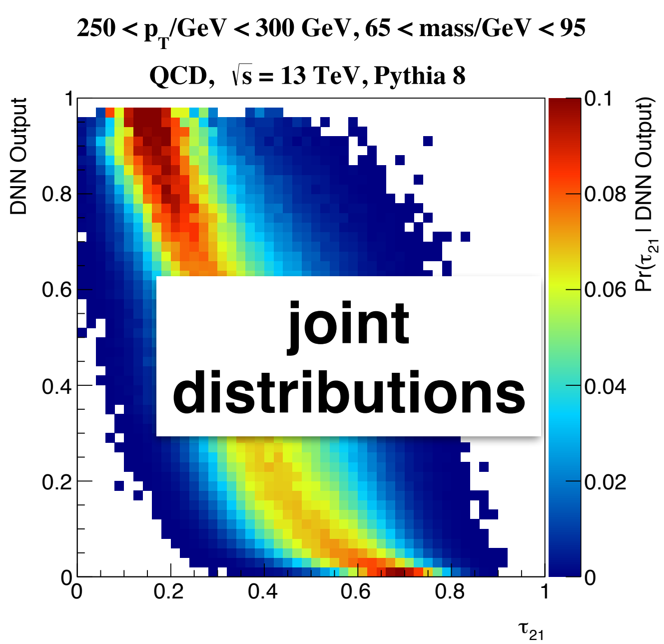
Opening the **box** is critical for improving robustness

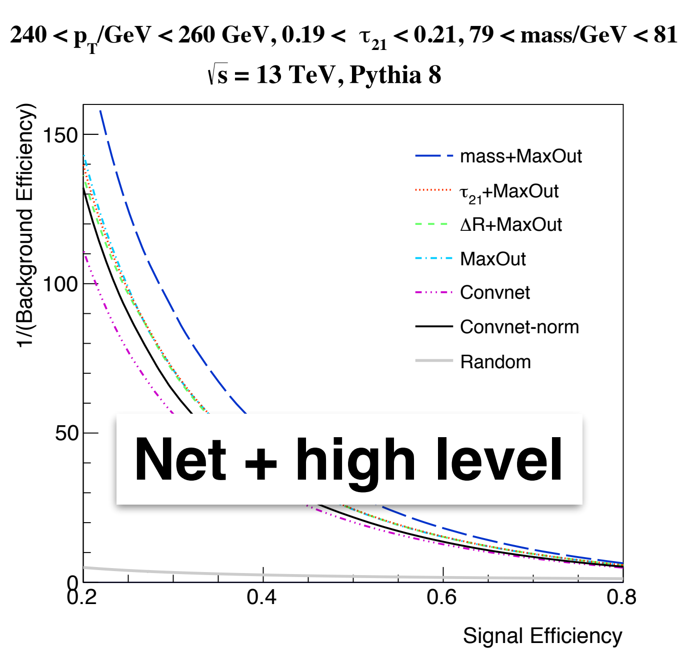


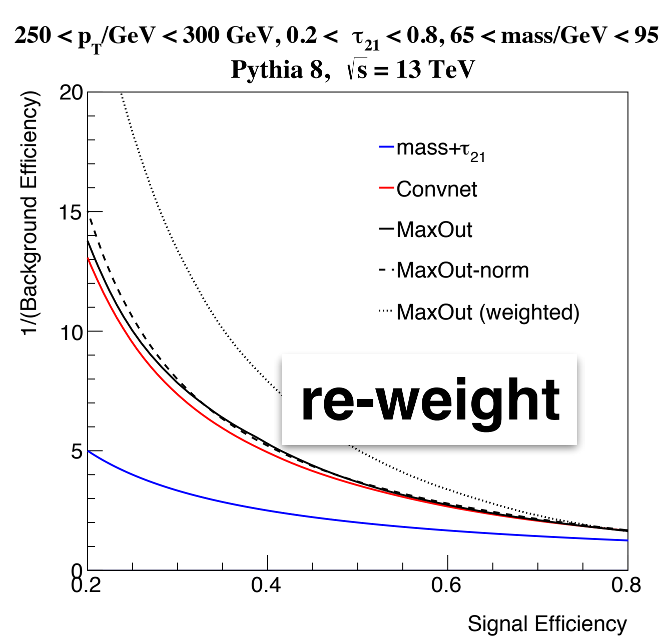


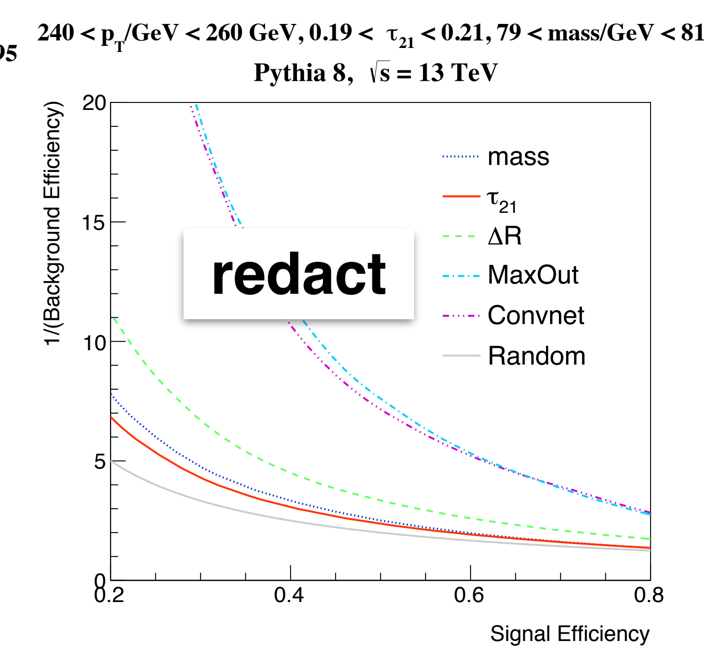






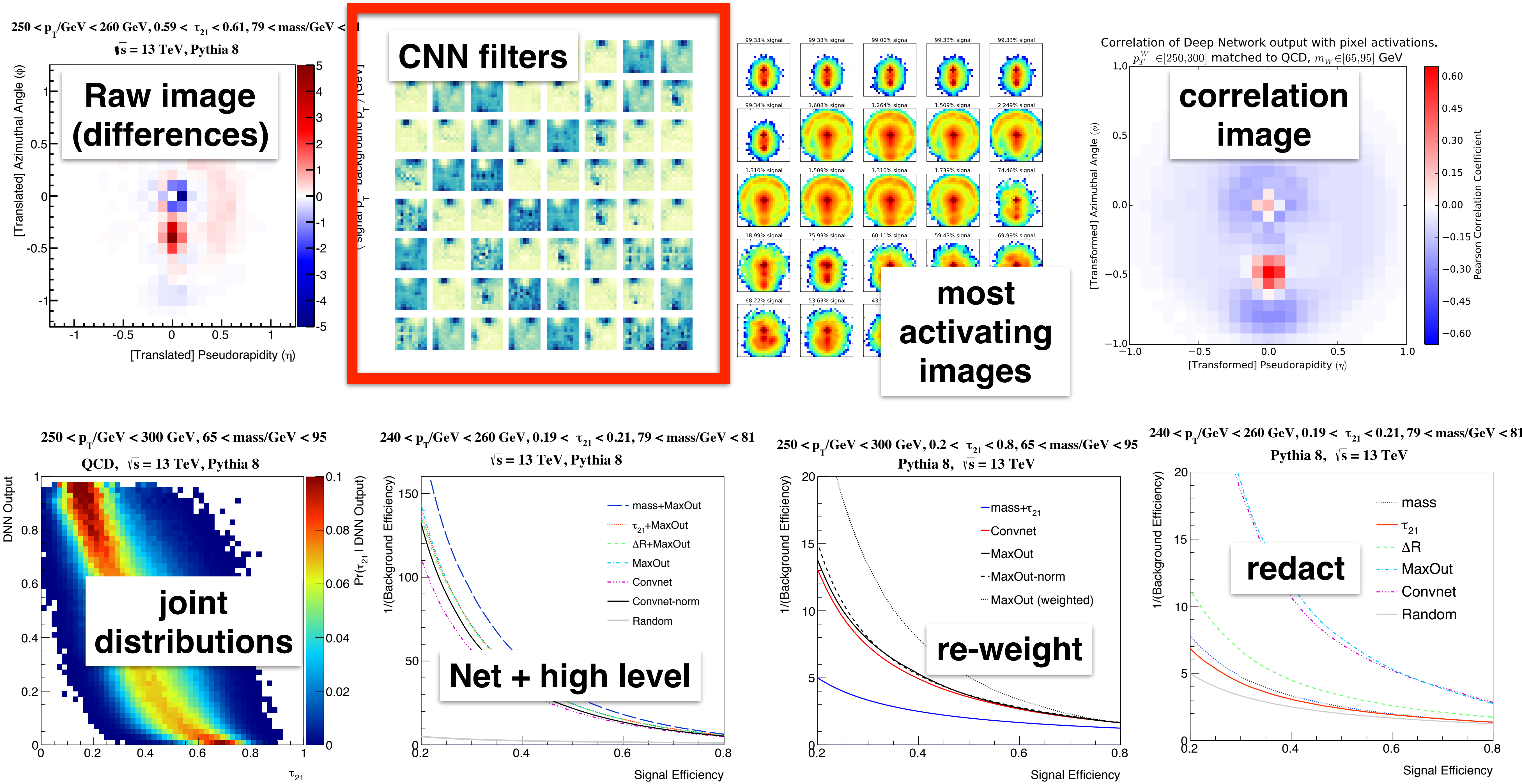






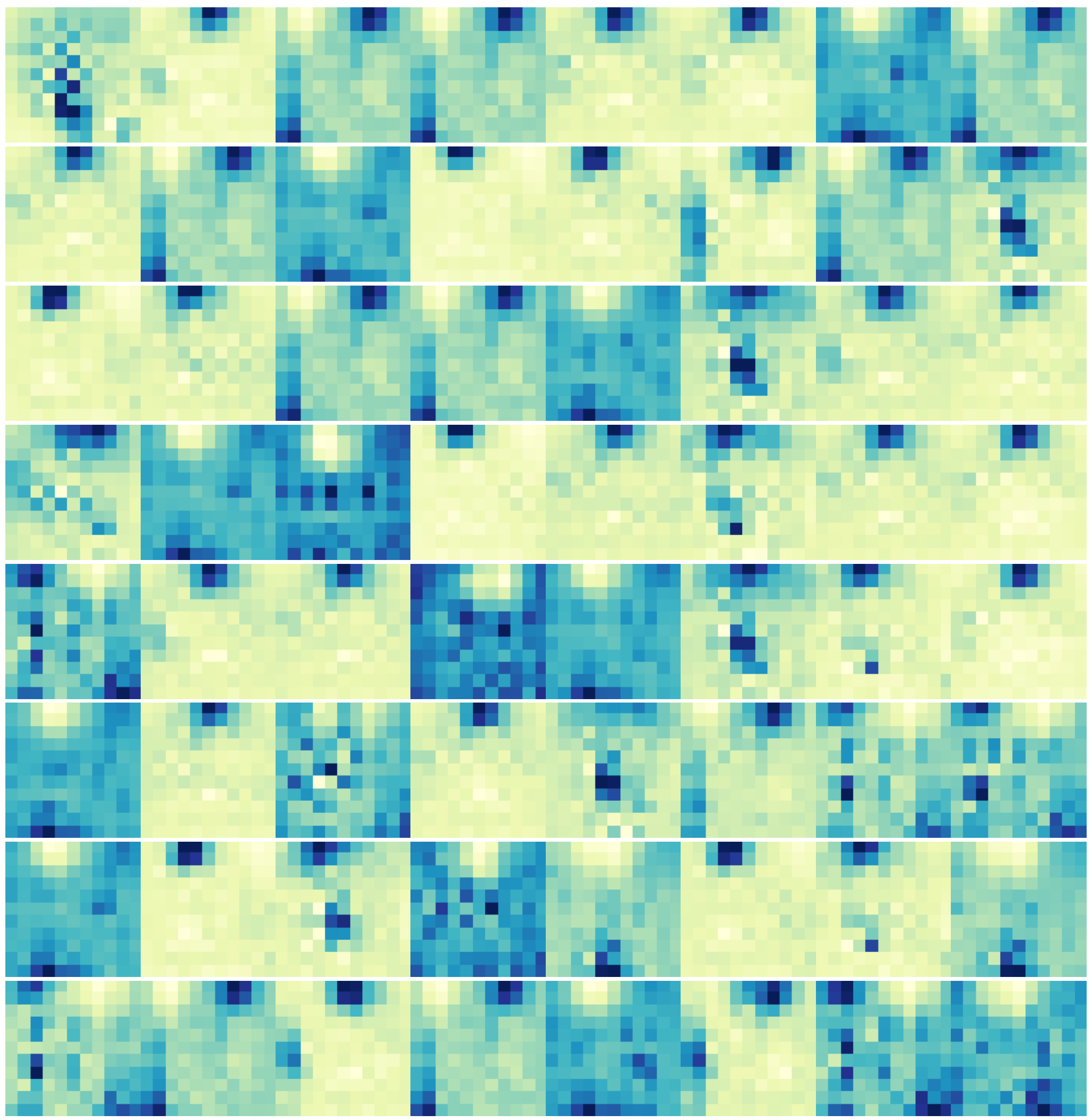
Learning about learning

Opening the **box** is critical for improving robustness



Convolution Filters

Filters are images! Can visualize 'higher-level features' learned by the network



Jet Images

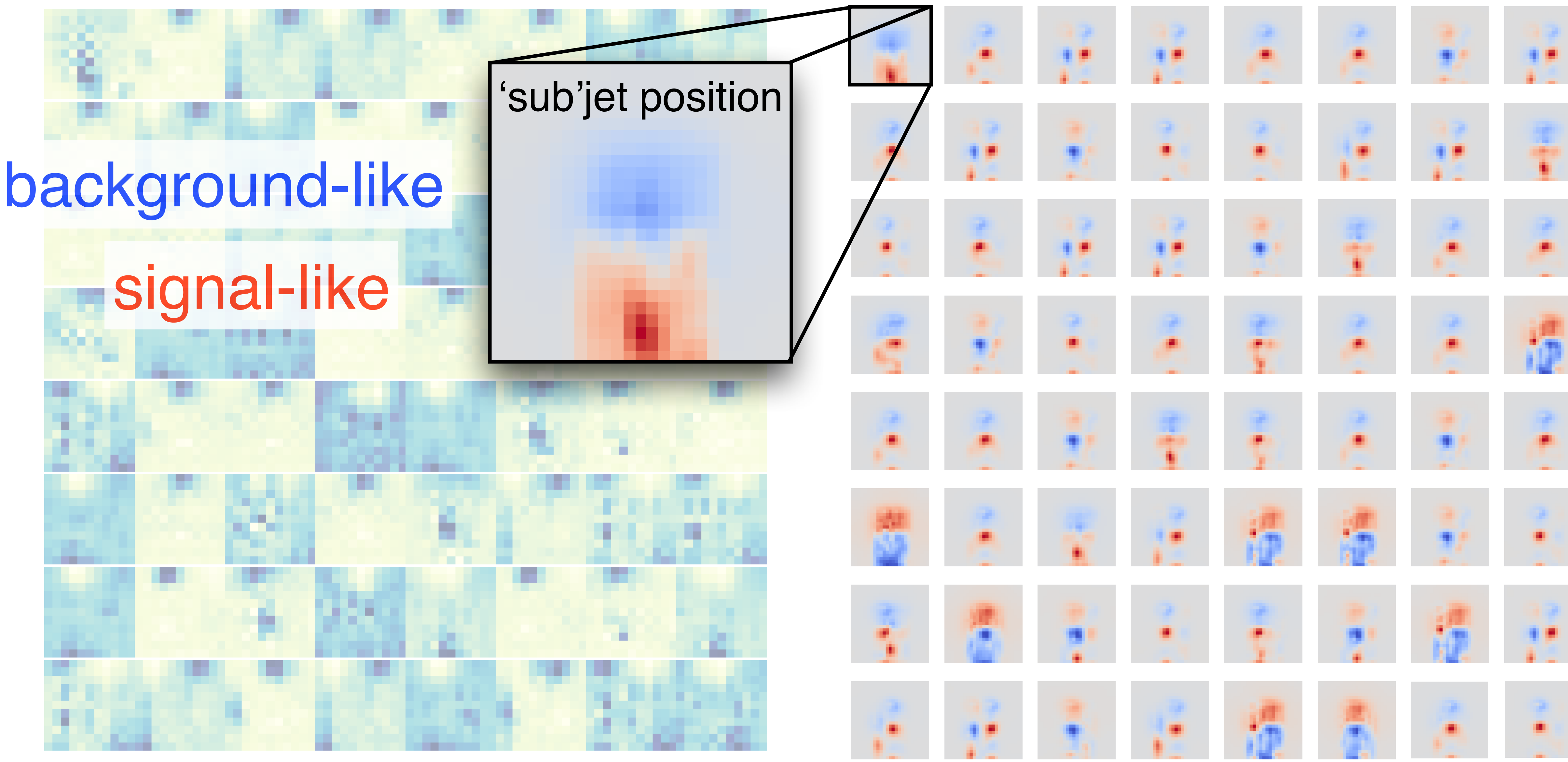
learned edge detection

"Natural" Images

L. de Oliveira, M. Kagan, L. Mackey, **BN**, and A. Schwartzman, JHEP 07 (2016) 069   A. Krizhevsky et al. DNN for ImageNet

Convolution Filters

Filters are images! Can visualize 'higher-level features' learned by the network



Jet Images Layer 1 Filters

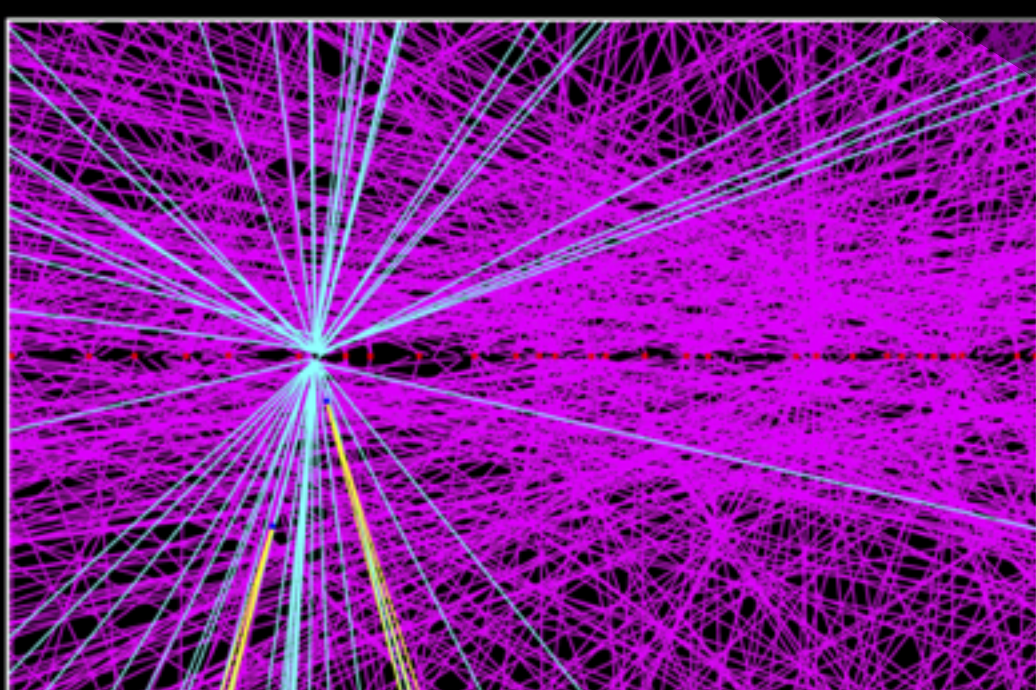
Filters convolved with signal - background

L. de Oliveira, M. Kagan, L. Mackey, **BN**, and A. Schwartzman, JHEP 07 (2016) 069

Beyond Classification I: Removing Noise

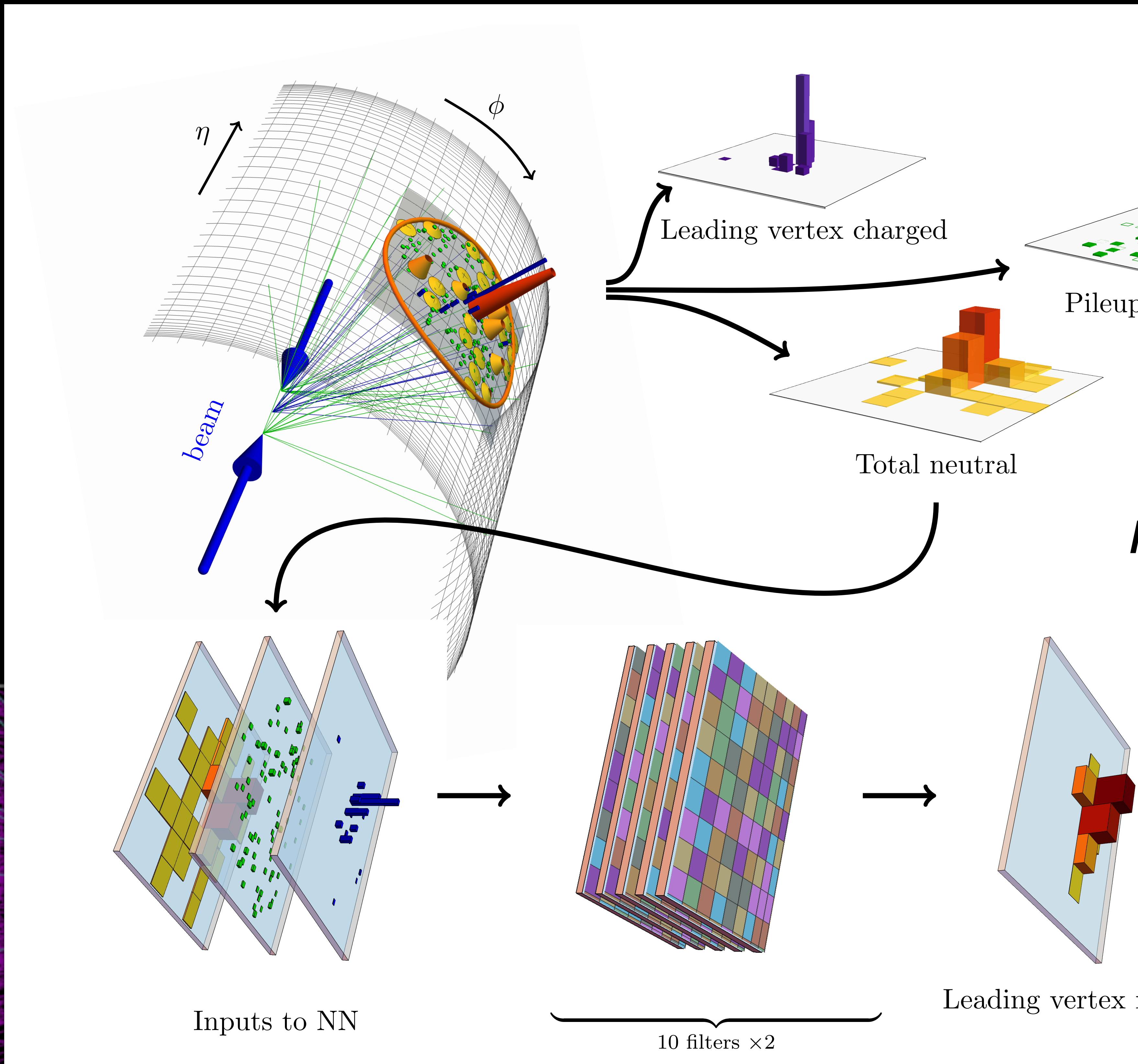
pp collisions at the LHC
don't happen one at a time!

the extra collisions are called **pileup**
and add soft radiation on top of our jets



this is akin to image
de-noising - we can
use ML for that!

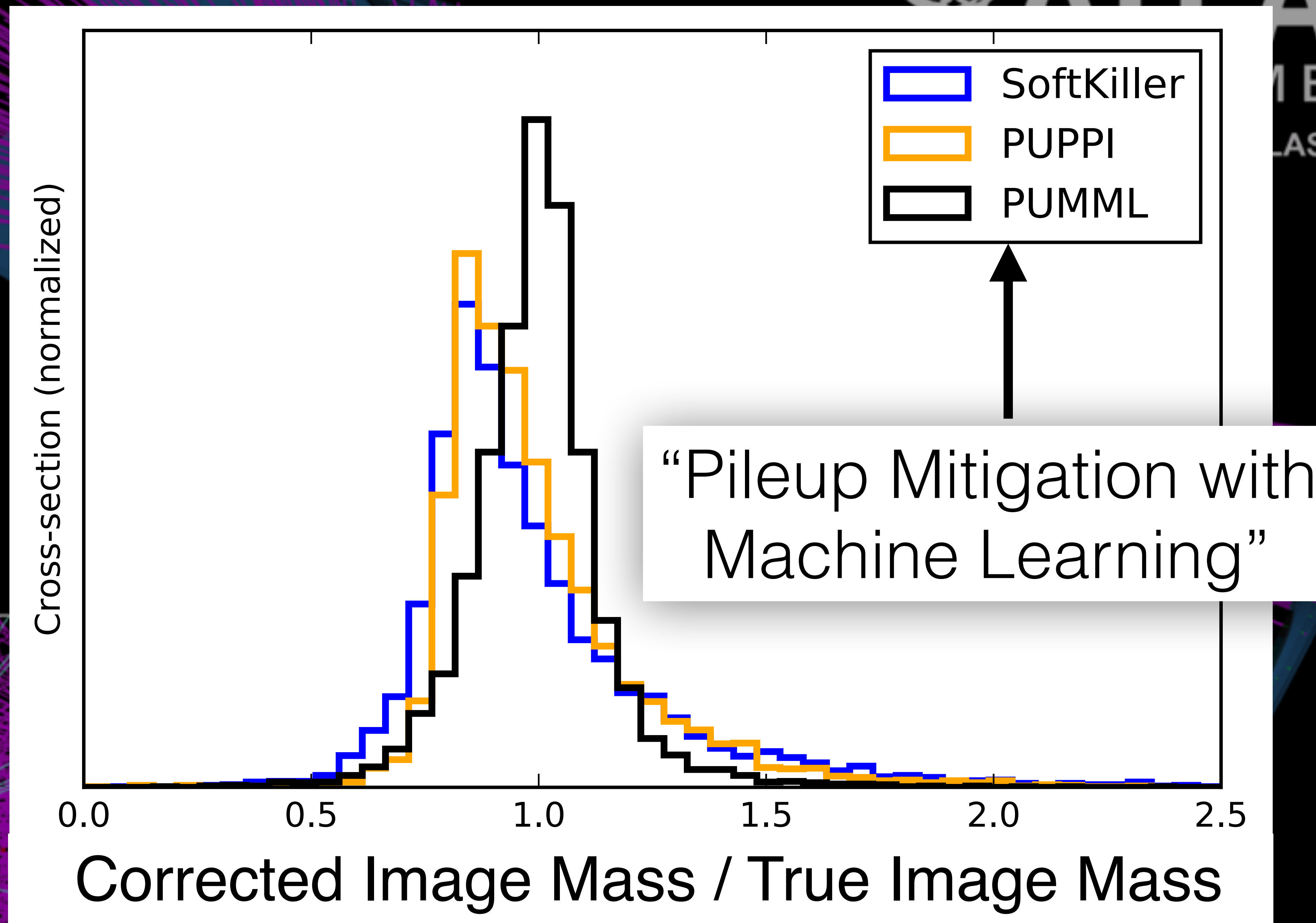
Beyond Classification I: Removing Noise

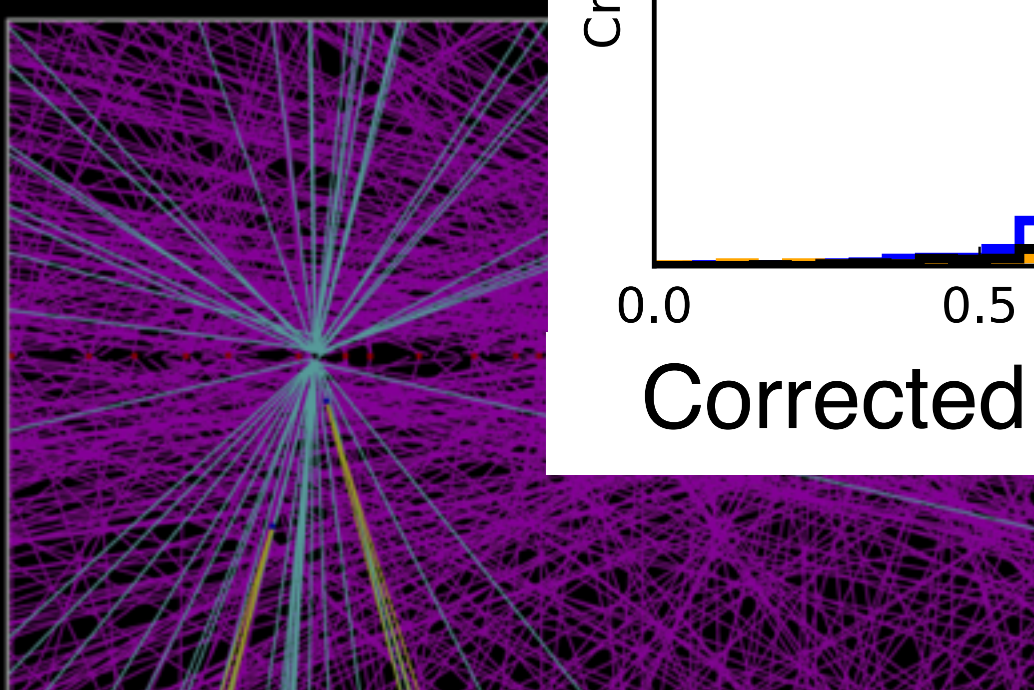


Strange noise because we can measure ~2/3 of it (charged pileup)

…also a natural application of convolutional NNs!

Beyond Classification I: Removing Noise



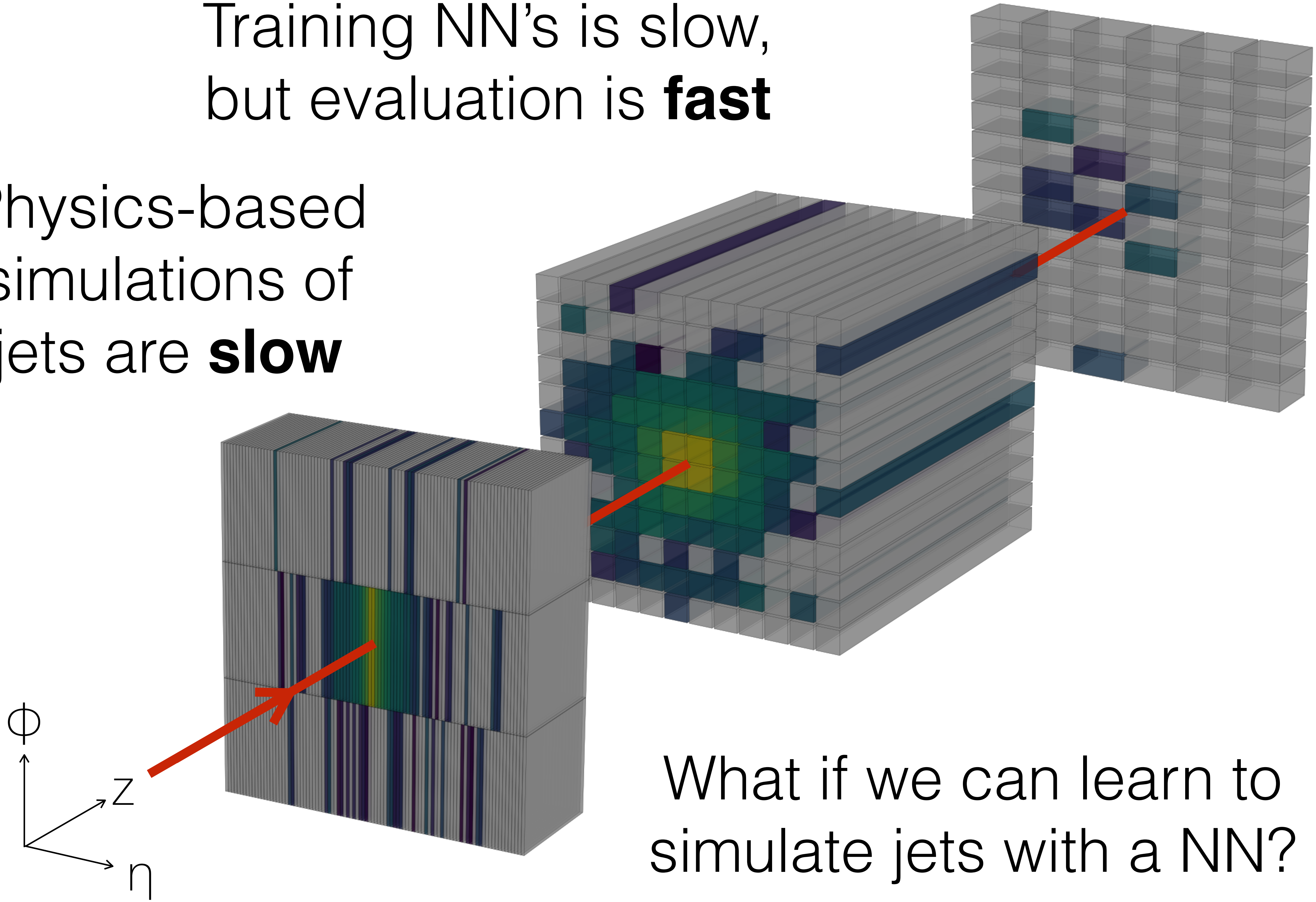


P. Komiske, E. Metodiev, **BPN**, and M. Schwartz,1707.08600

Beyond Classification II: Simulation NN

Training NN's is slow,
but evaluation is **fast**

Physics-based
simulations of
jets are **slow**



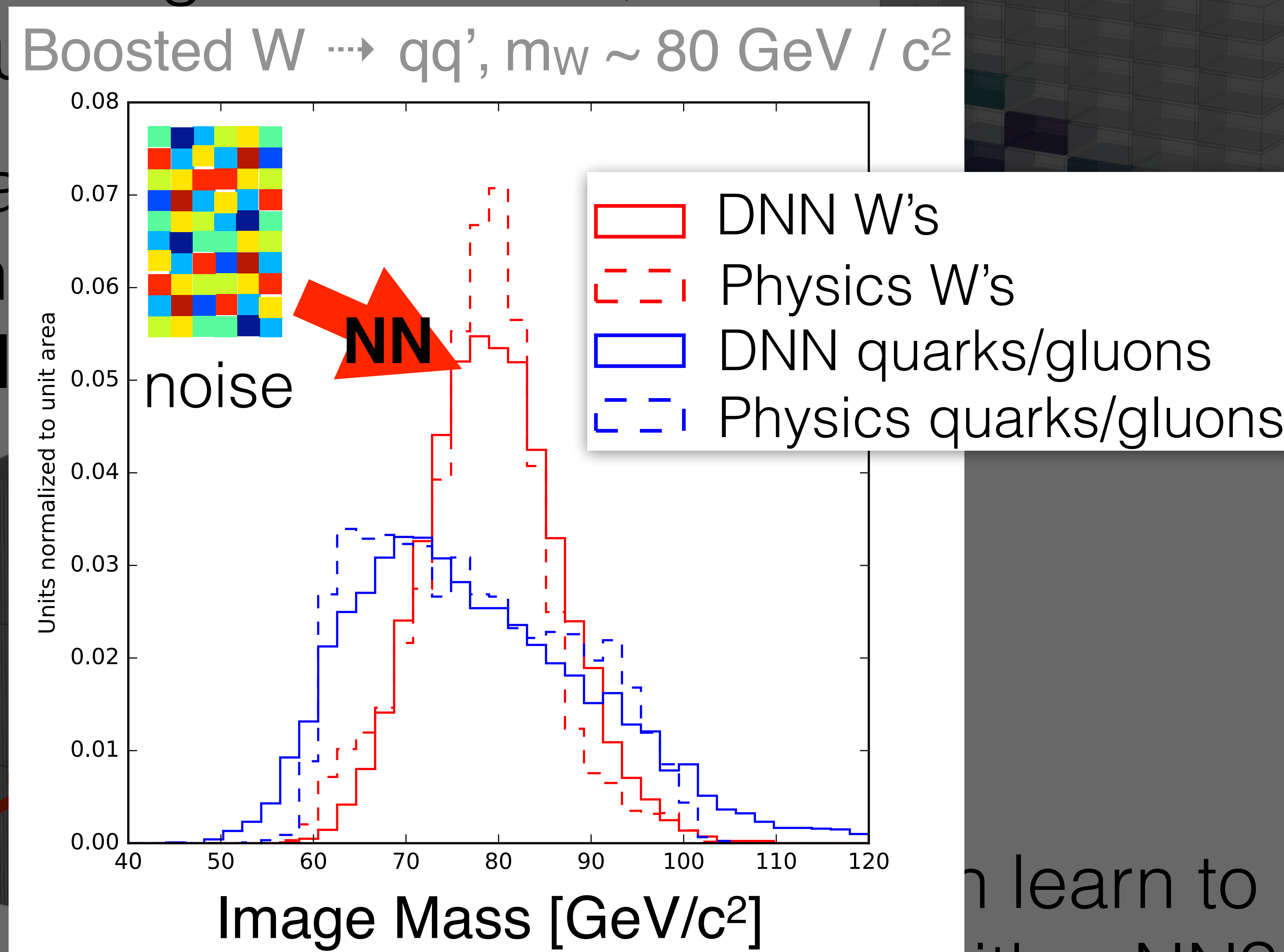
What if we can learn to
simulate jets with a NN?

Beyond Classification II: Simulation NN

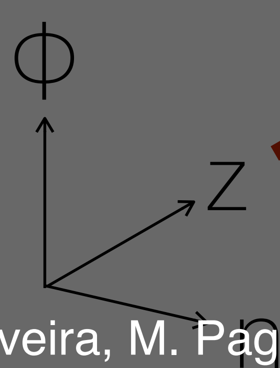
Training NN's is slow,

bu

Physics-ba
simulation
jets are **sl**



n learn to
simulate jets with a NN?

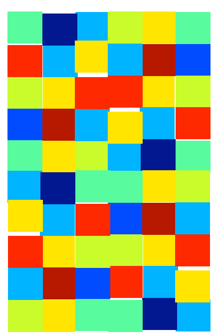


L. de Oliveira, M. Paganini, **BN**, Comp. and Software for Big Science (2017) 1

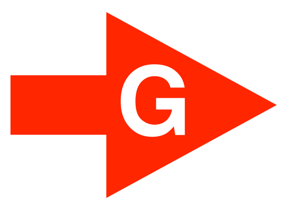
M. Paganini, L. de Oliveira, **BN**,1705.02355

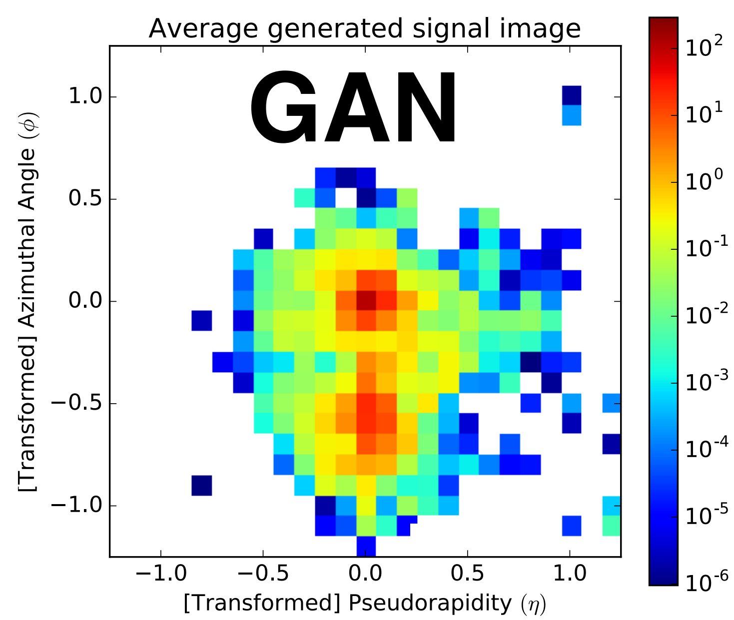
And now: Modern Deep NN's for Generation

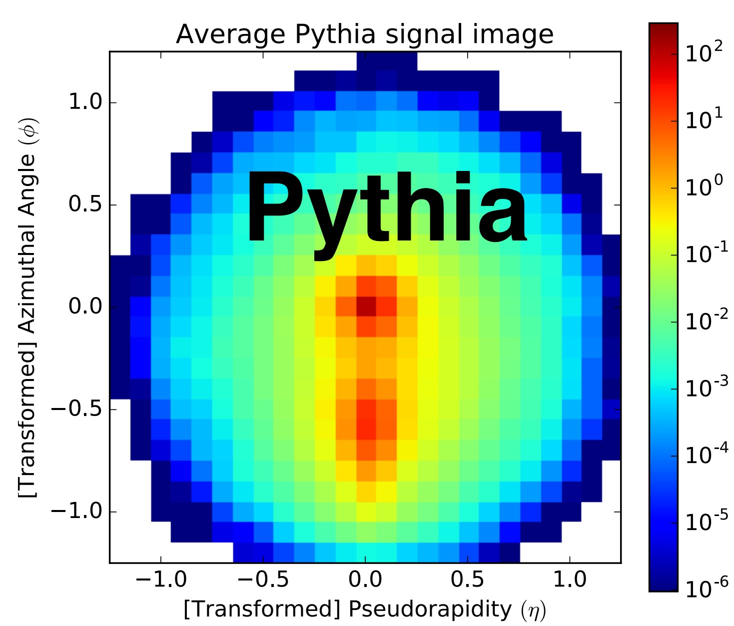
Generative Adversarial Networks (GAN):
*A two-network game where one **maps noise to images** and one **classifies images as fake or real**.*

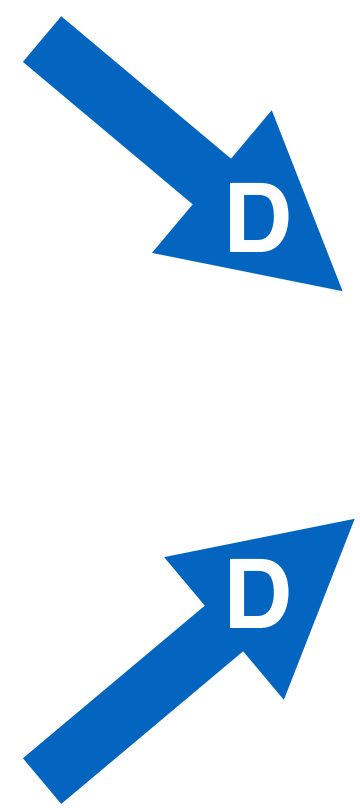


noise



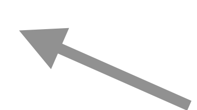






{real,fake}

When **D** is maximally confused, **G** will be a good generator

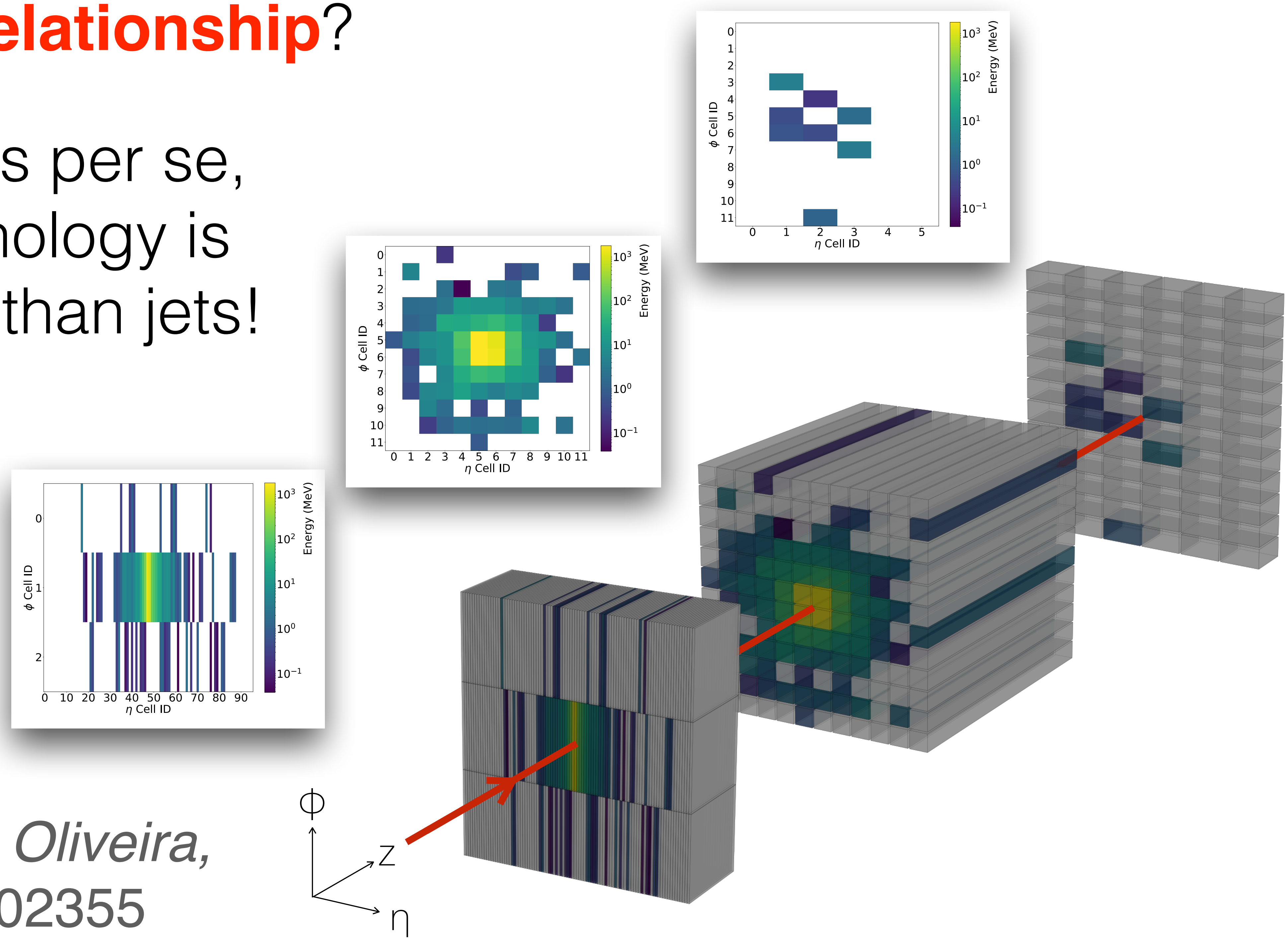


Physics-based simulator

+ More Layers for Generation

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!

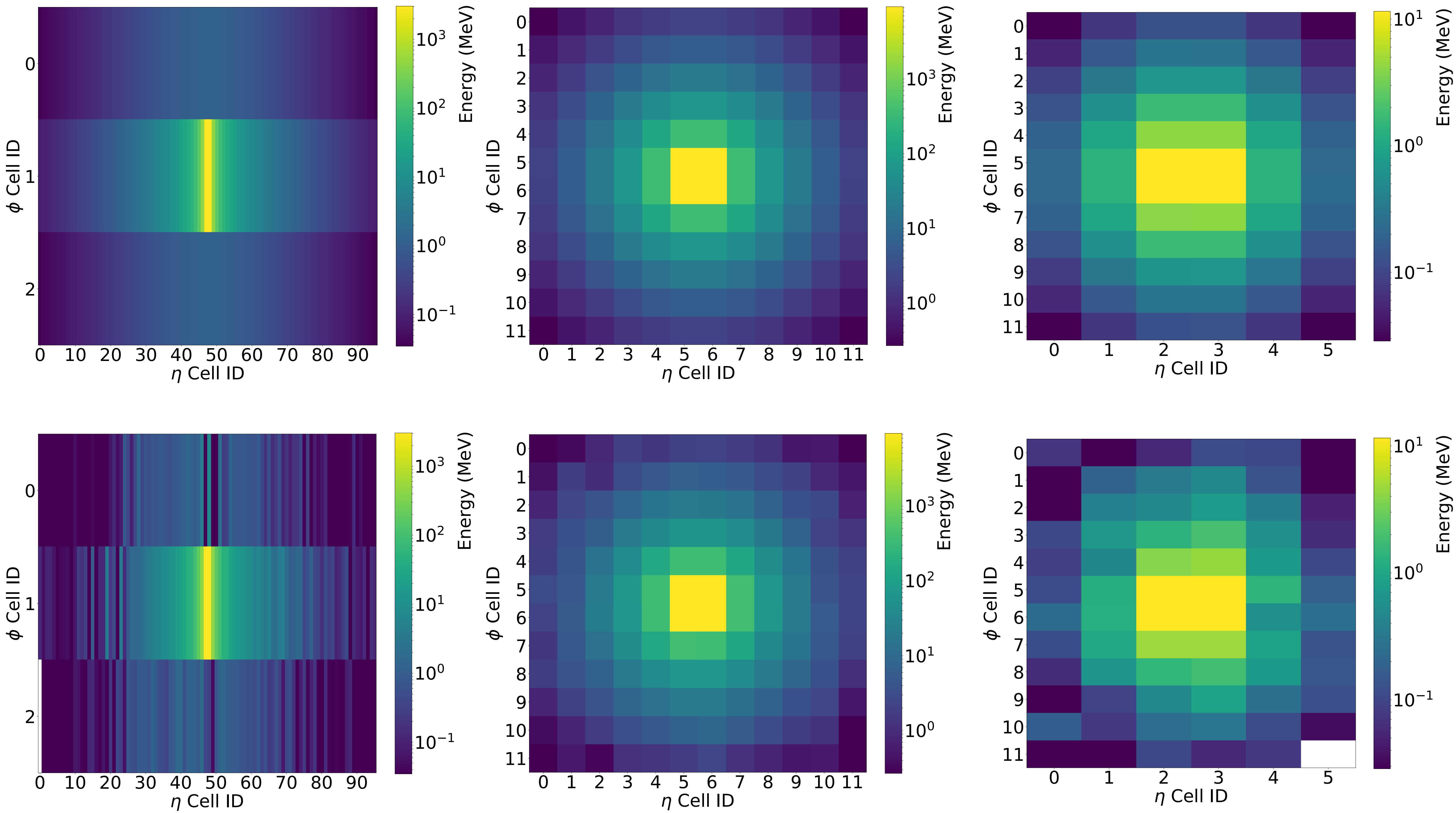


M. Paganini, L. de Oliveira, and ***BPN*** 1705.02355

Average Images

Geant4

M. Paganini, L. de Oliveira, and ***BPN*** 1705.02355



CaloGAN

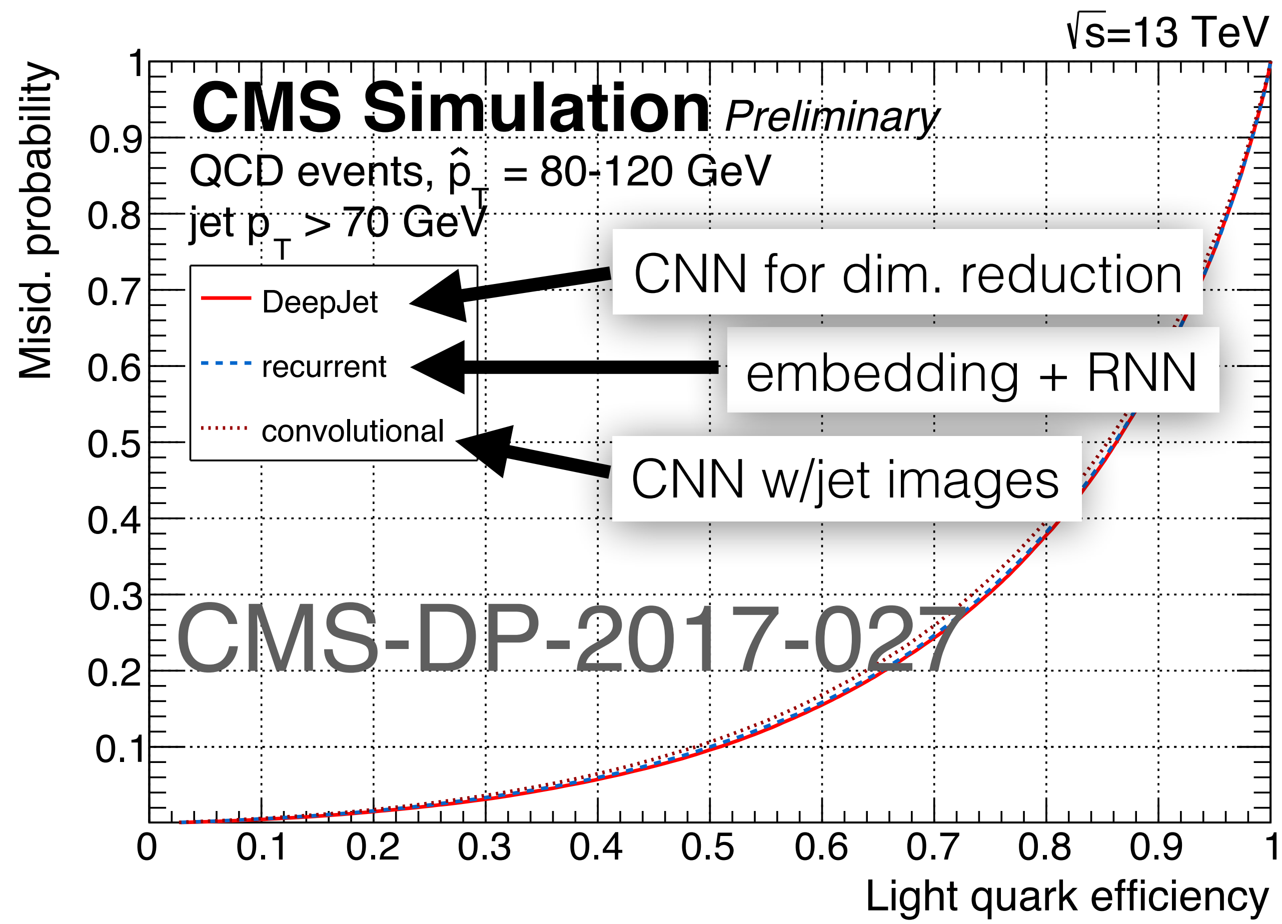
Timing

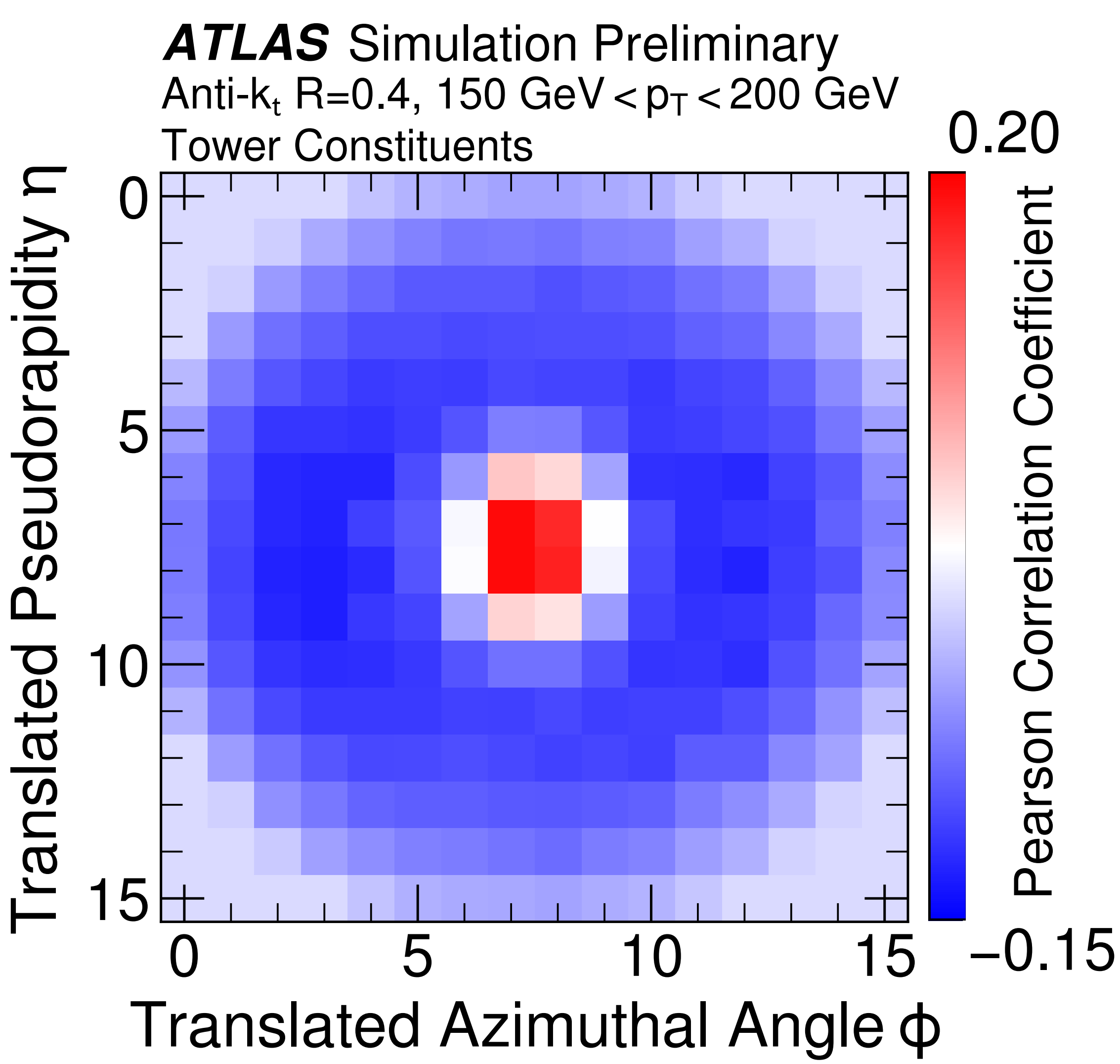
M. Paganini, L. de Oliveira, and ***BPN*** 1705.02355

Generation Method	Hardware	Batch Size	milliseconds/shower
GEANT4	CPU	N/A	1772 ←
CALOGAN	CPU	1	13.1
		10	5.11
		128	2.19
		1024	2.03
	GPU	1	14.5
		4	3.68
		128	0.021
		512	0.014
		1024	0.012 ←

R&D cycle is fast; sometimes integration can be slow

There may be multiple ways to get to the same solution



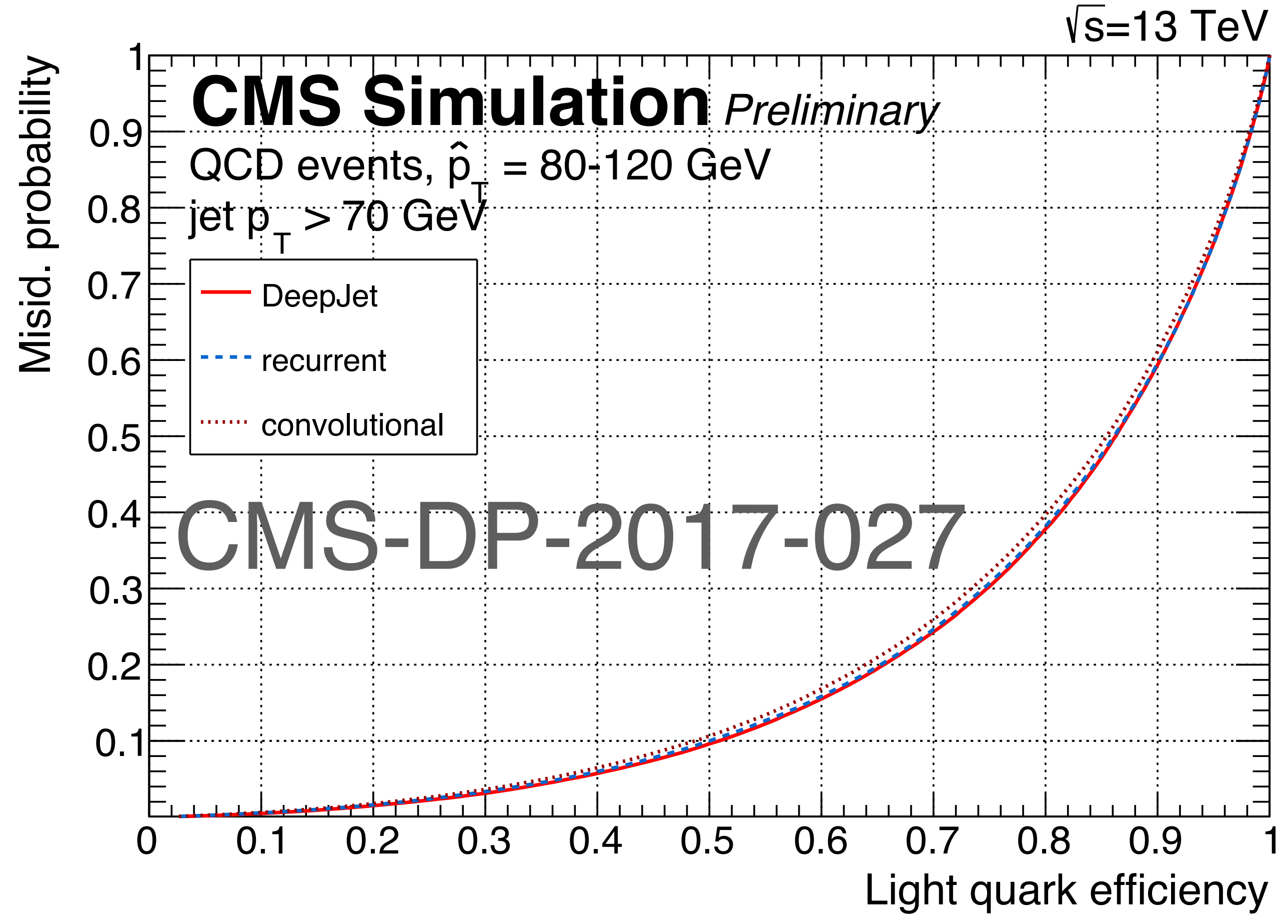


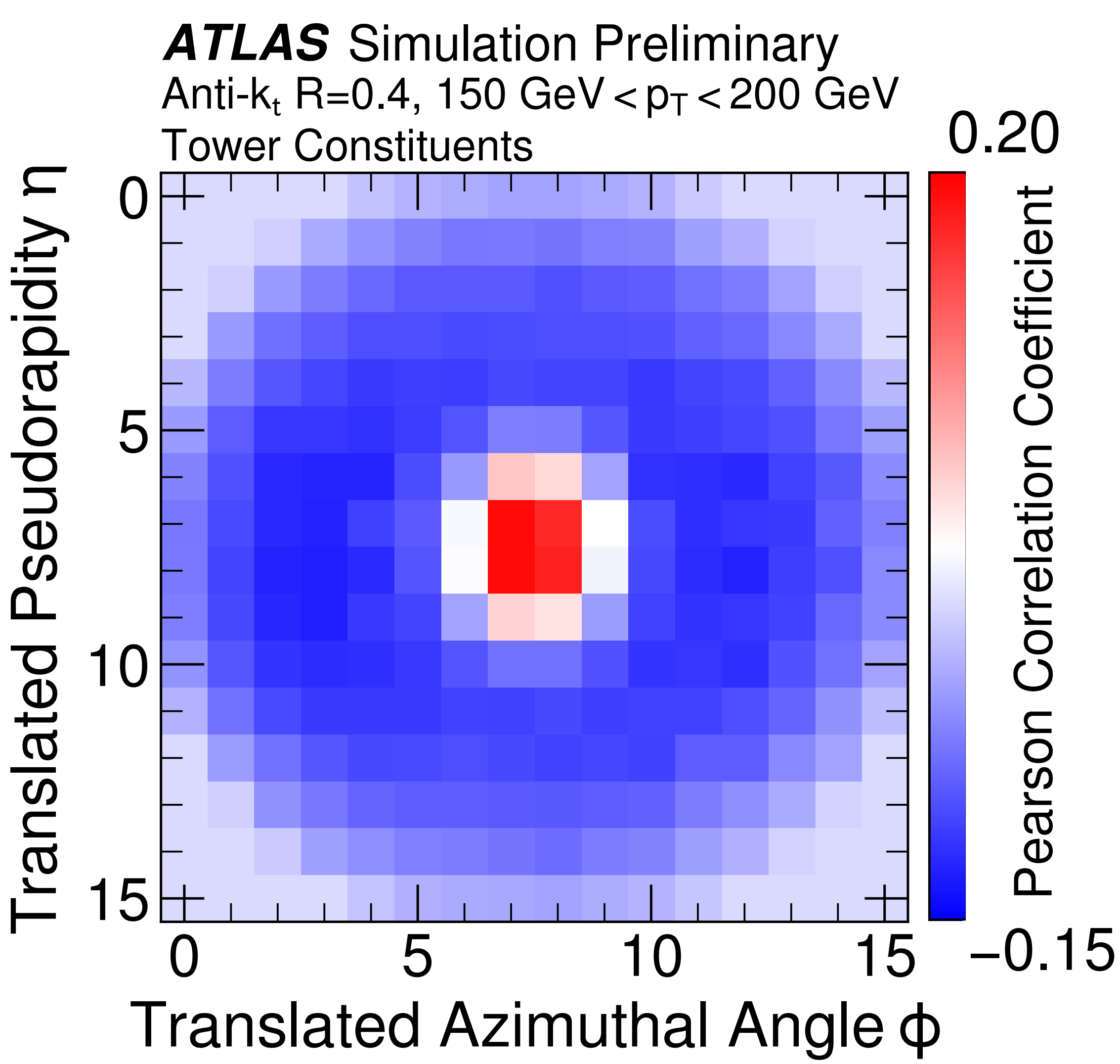
ATL-PHYS-PUB-2017-017

CMS, ATLAS, LHCb, MicroBooNE, NOvA, DUNE, etc. are increasing their use of DNNs for many applications

Conclusions and outlook

DNN classification, regression, and generation are powerful tools to fully exploit the physics program at the LHC & beyond

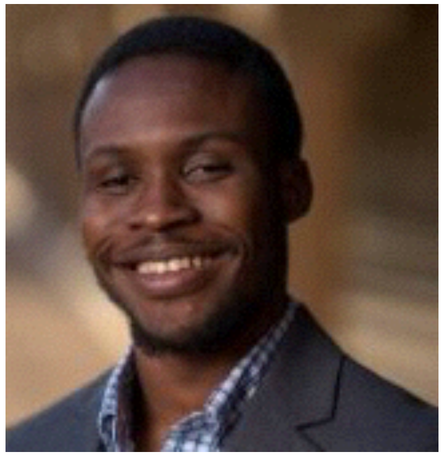




ATL-PHYS-PUB-2017-017

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

Collaborators



Lucio
Dery
Stanford

Michela
Paganini
Yale

Eric
Metodiev
MIT

Patrick
Komiske
MIT

Zihao
Jiang
Stanford

Francesco
Rubbo
SLAC

Luke
de Oliveira
VAI tech.

Michael
Kagan
SLAC

Jesse
Thaler
MIT

Matt
Schwartz
Harvard

Ariel
Schwartzman
SLAC