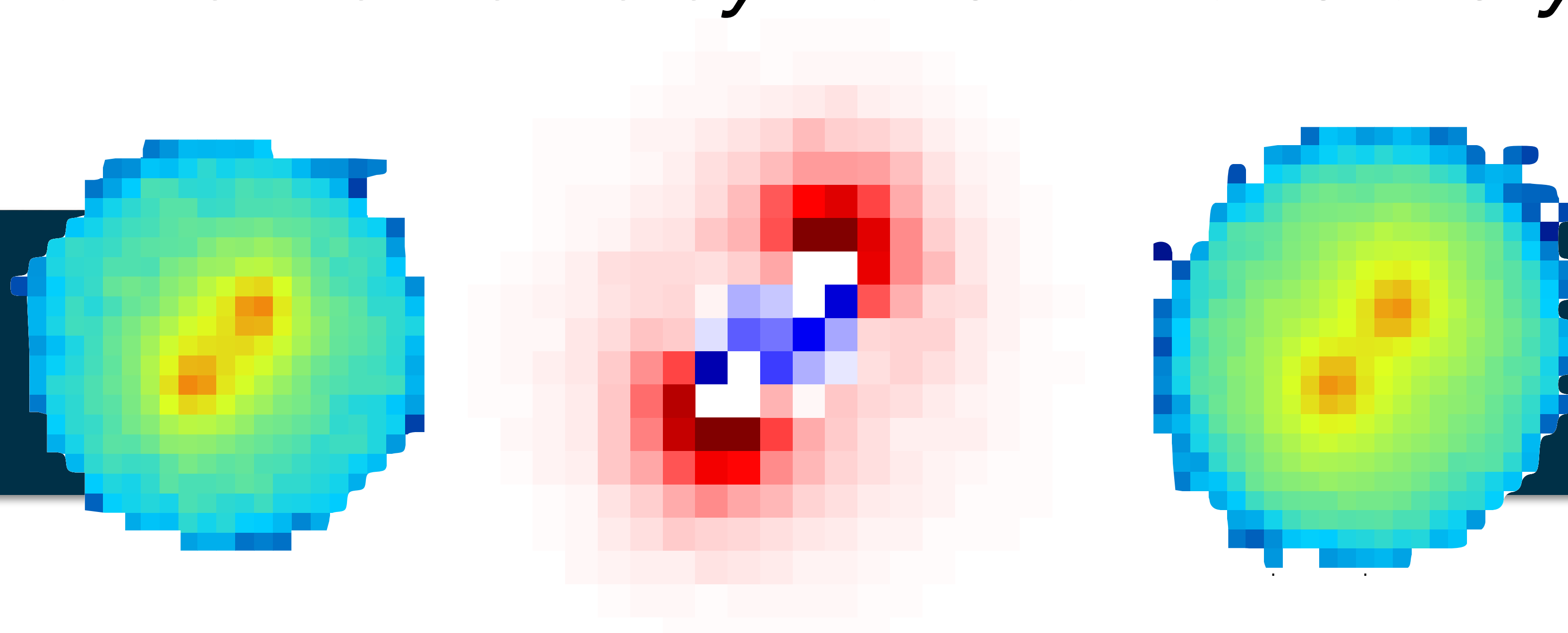
Modern Machine Learning

with Jet Images for High Energy Physics

Benjamin Nachman

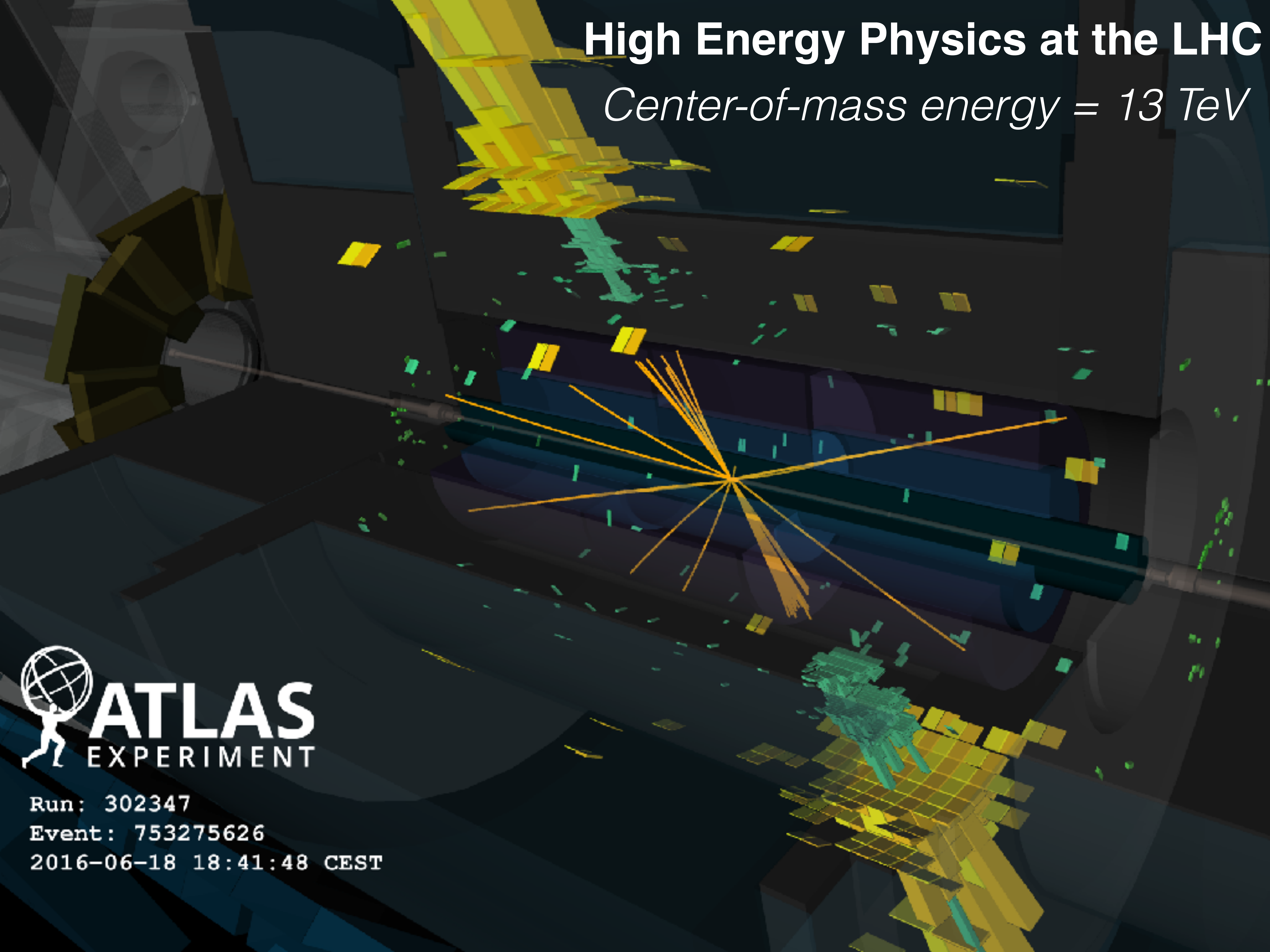
Lawrence Berkeley National Laboratory



University of Genova, November 10, 2017

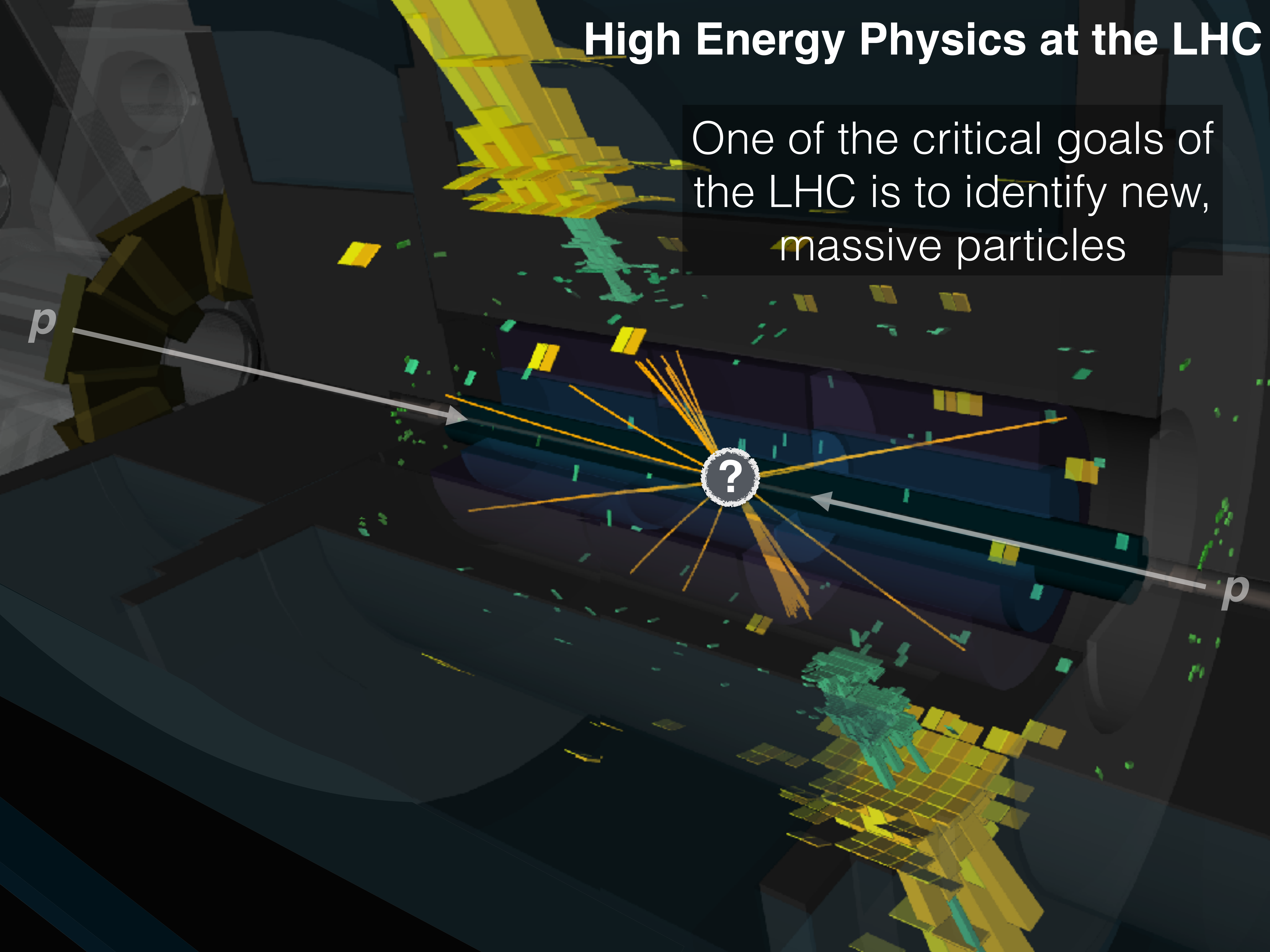
High Energy Physics at the LHC

Center-of-mass energy = 13 TeV

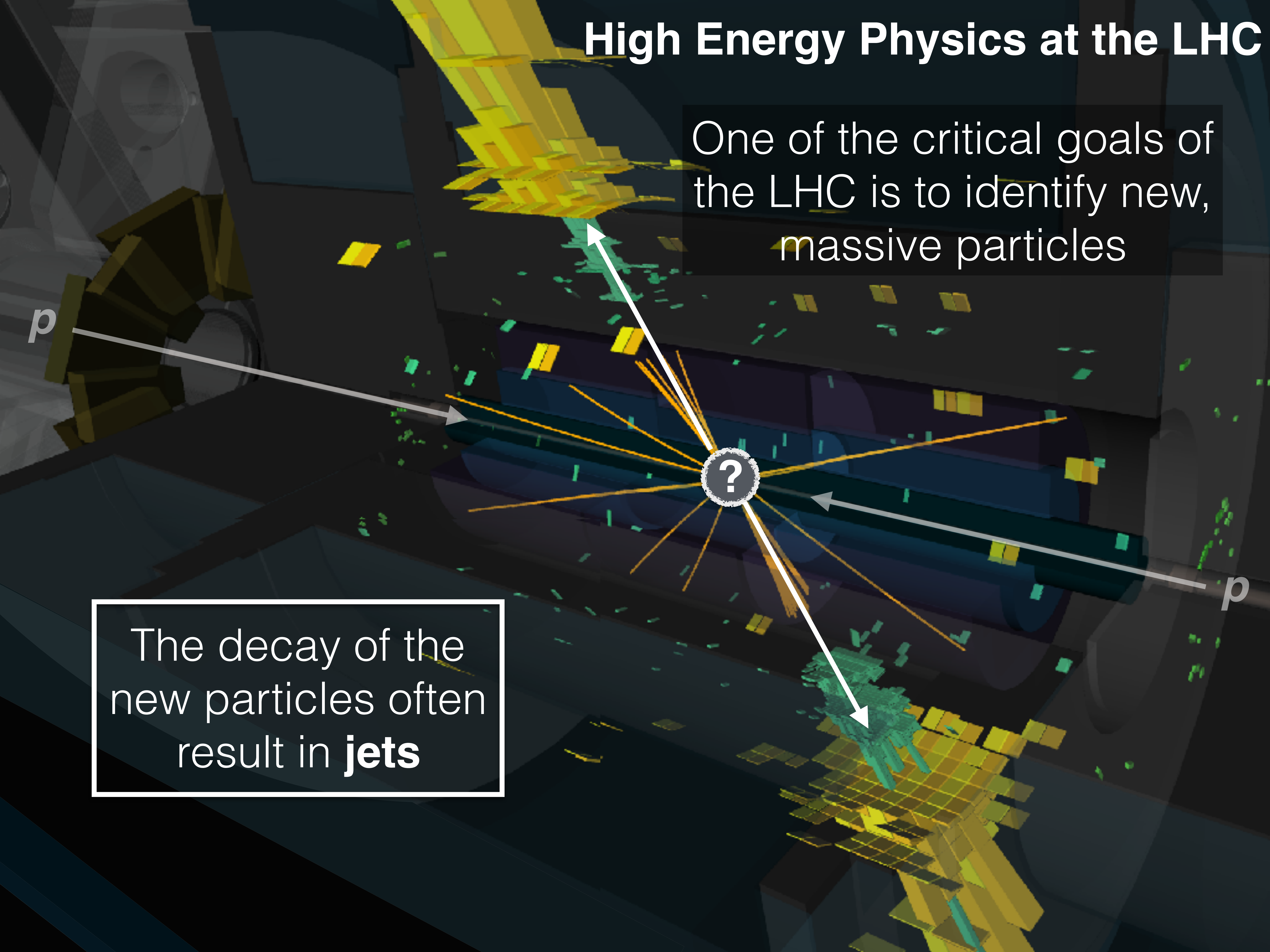


High Energy Physics at the LHC

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

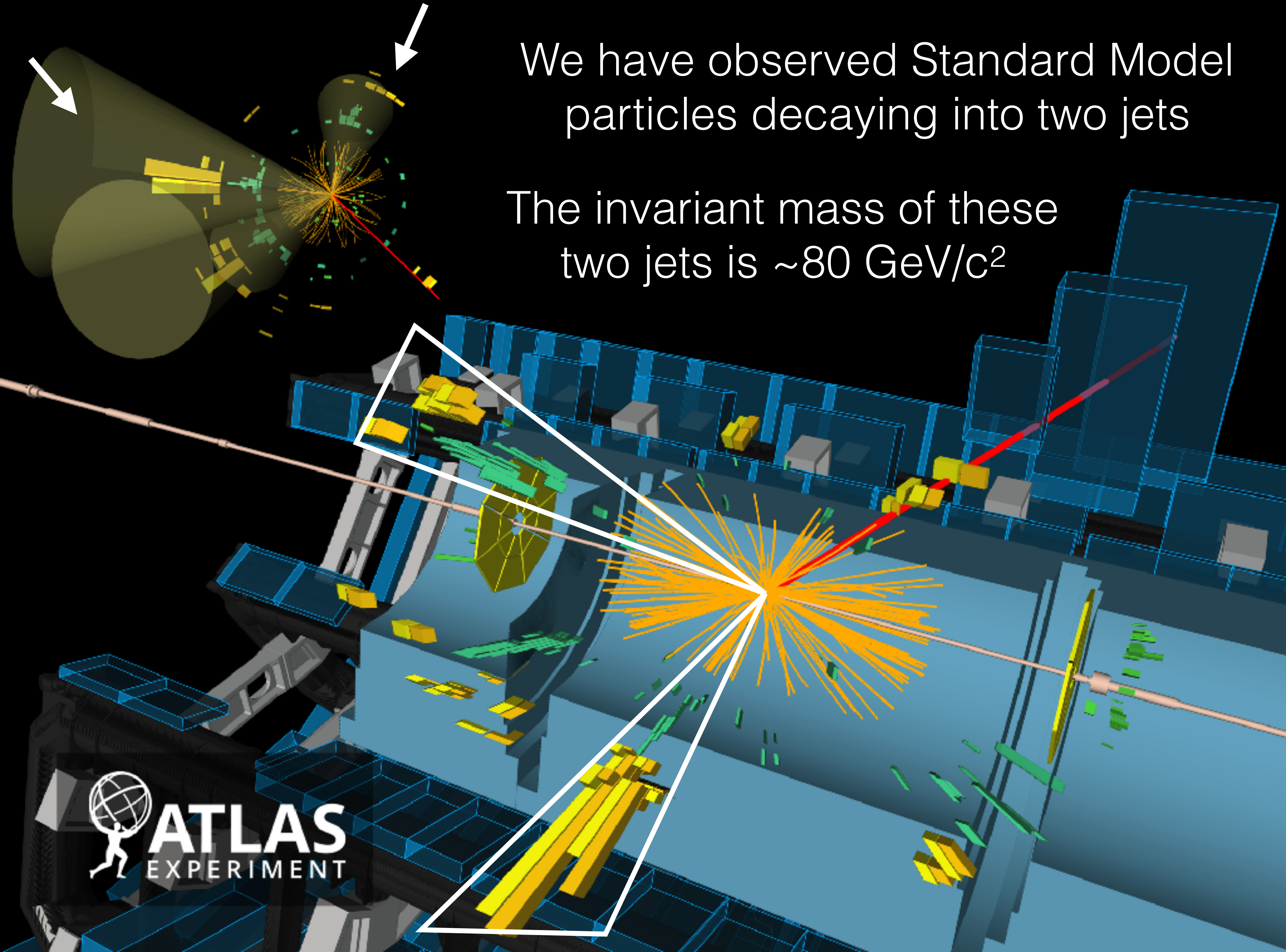


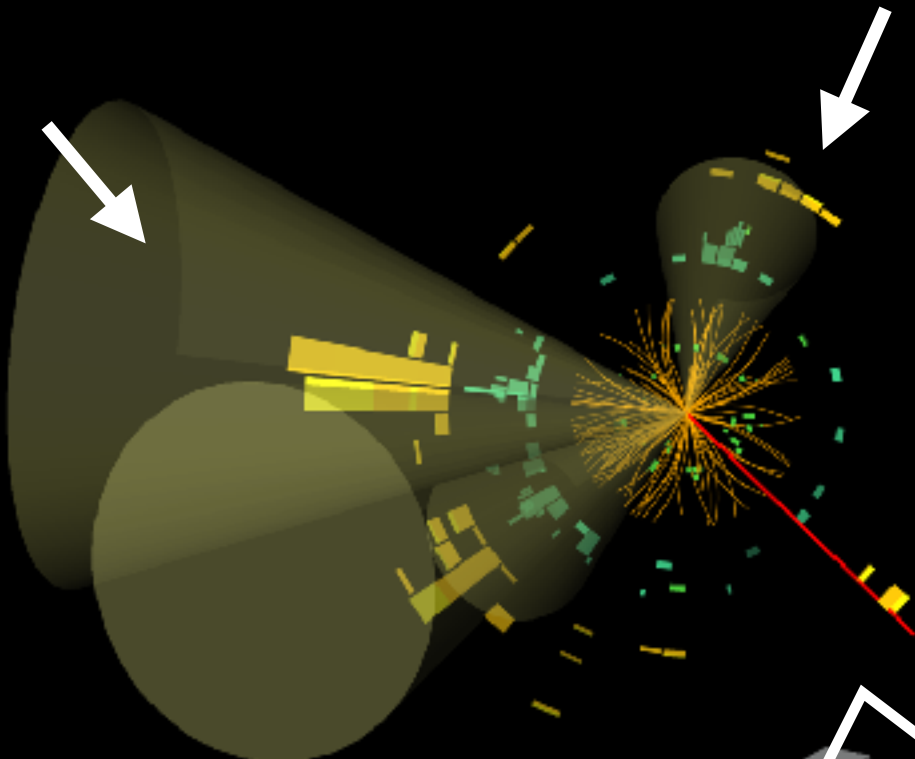
High Energy Physics at the LHC



We have observed Standard Model particles decaying into two jets

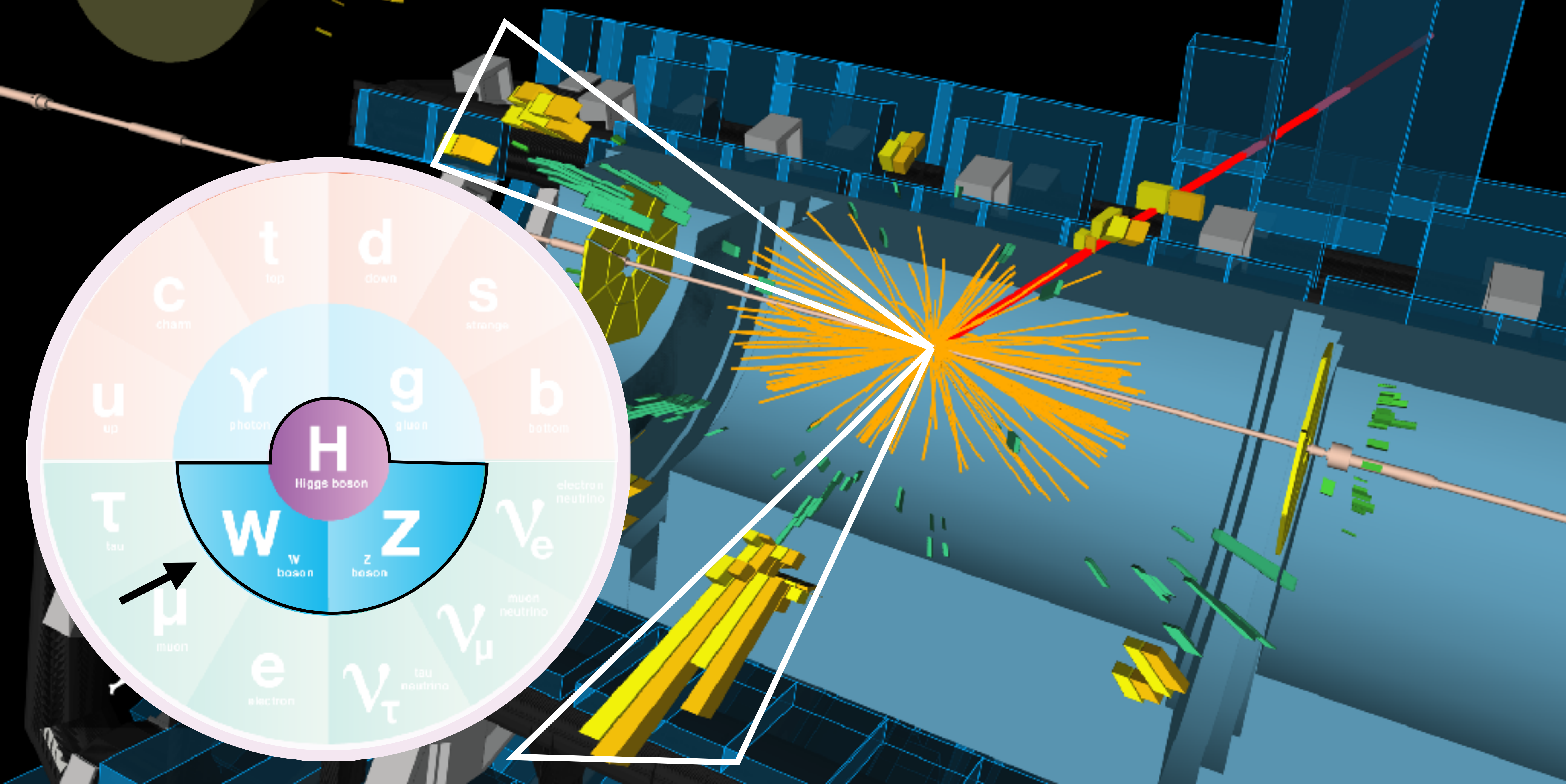
The invariant mass of these two jets is ~80 GeV/c^2





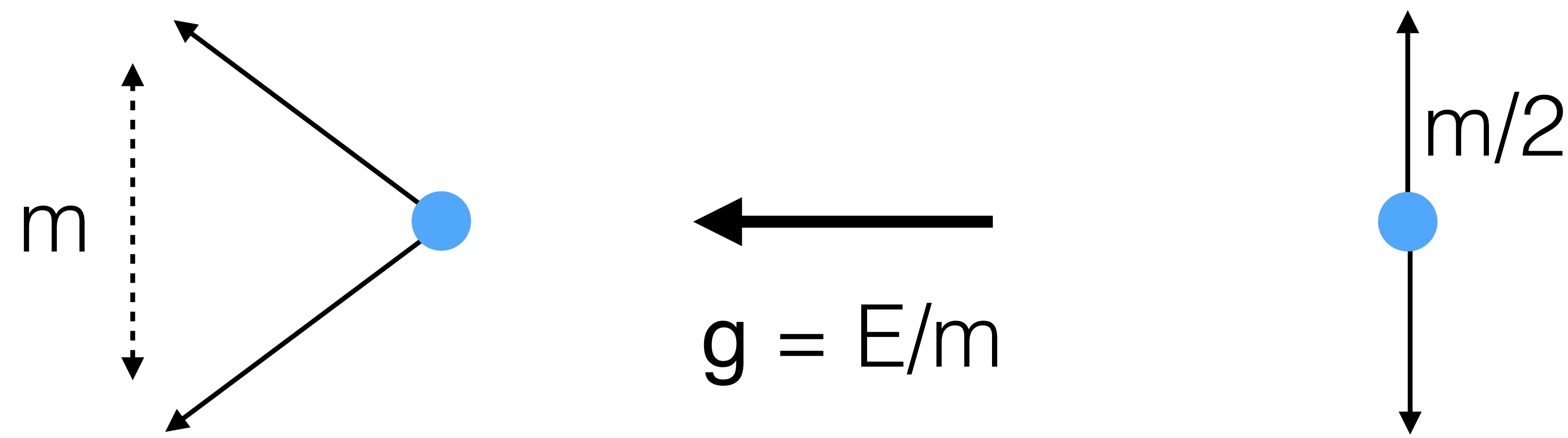
We have observed Standard Model particles decaying into two jets

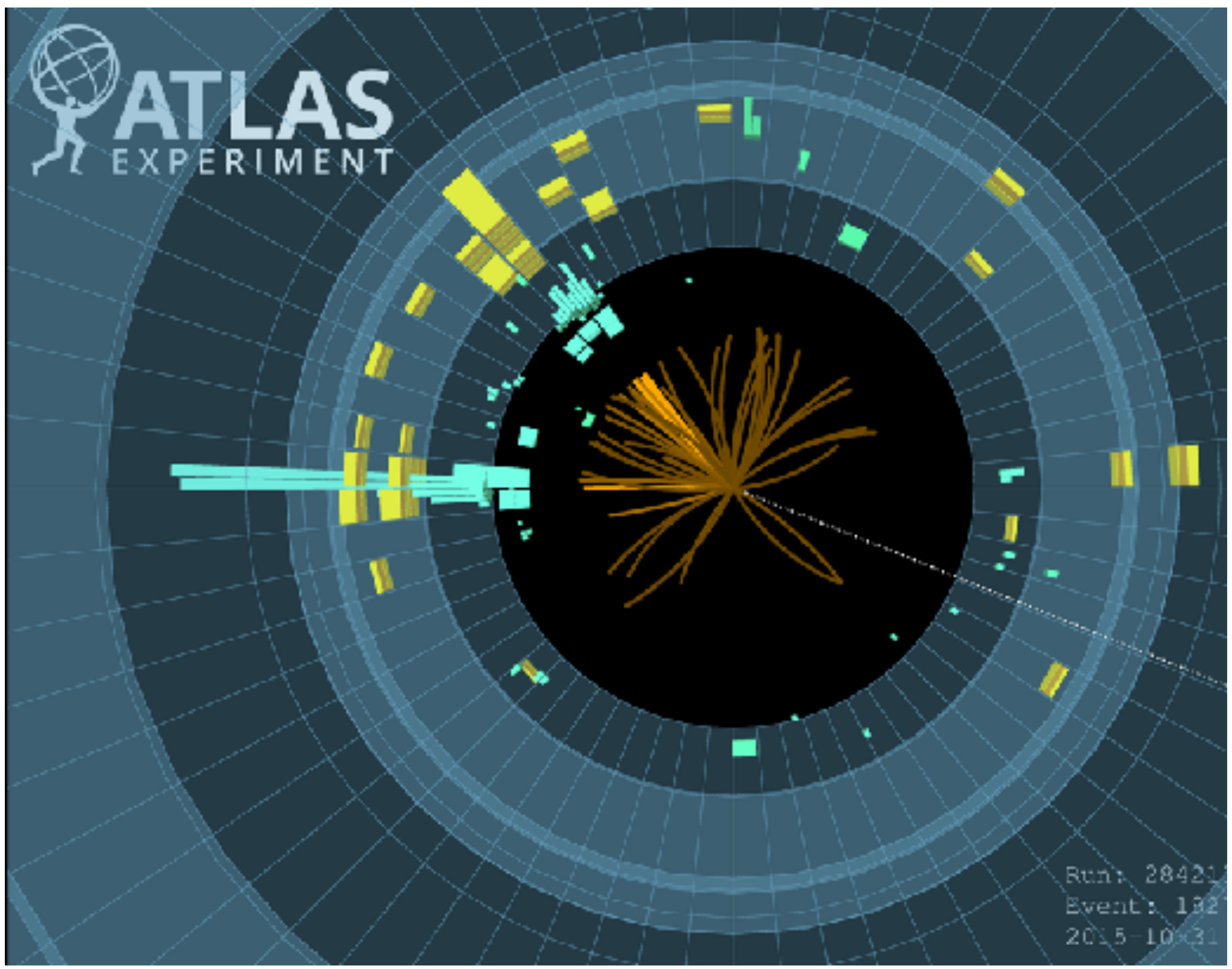
The invariant mass of these two jets is ~80 GeV/c²

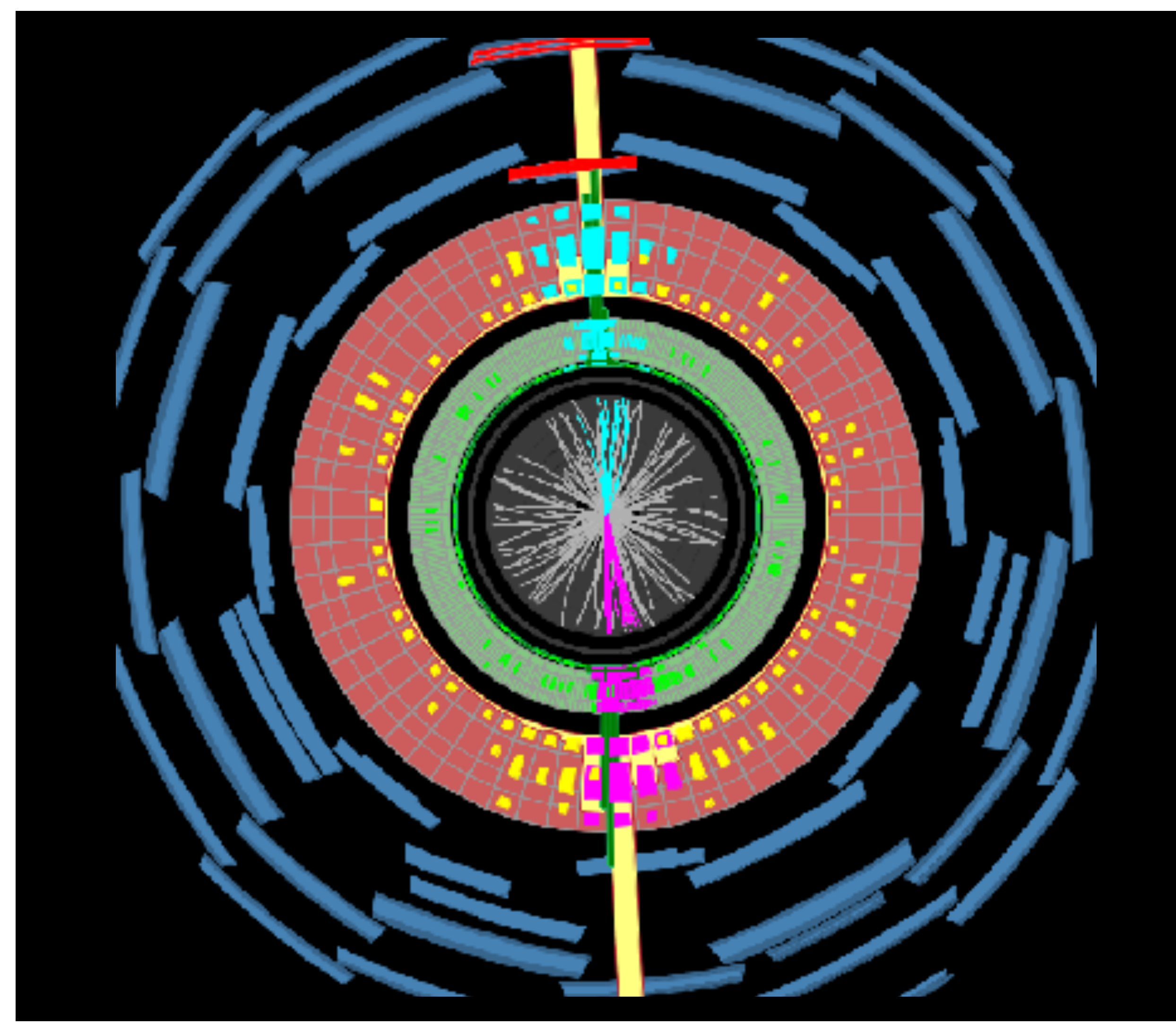


What if you take one of those SM dijet resonances and Lorentz boost it?

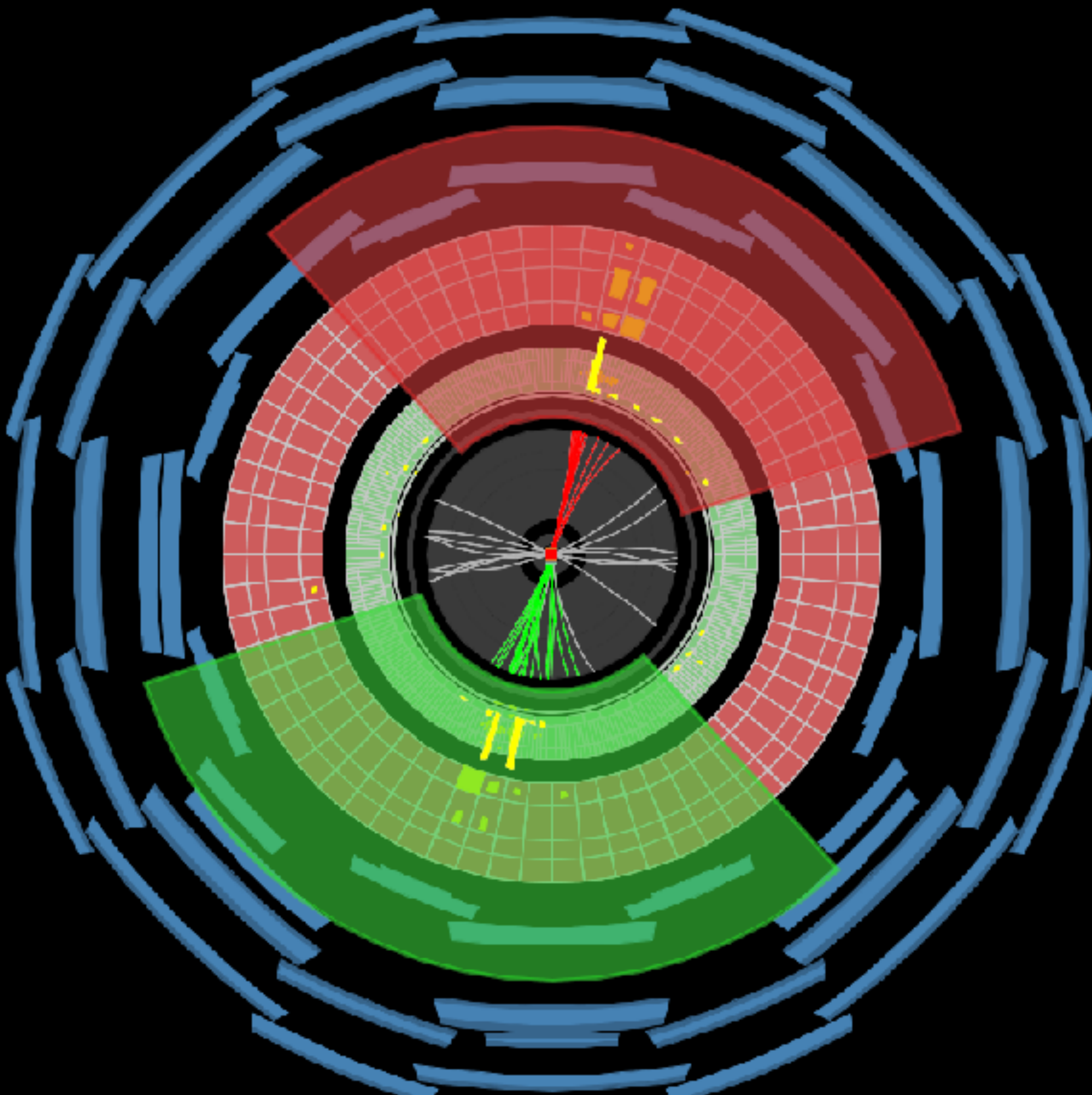
$f \sim 1/g = m/E$

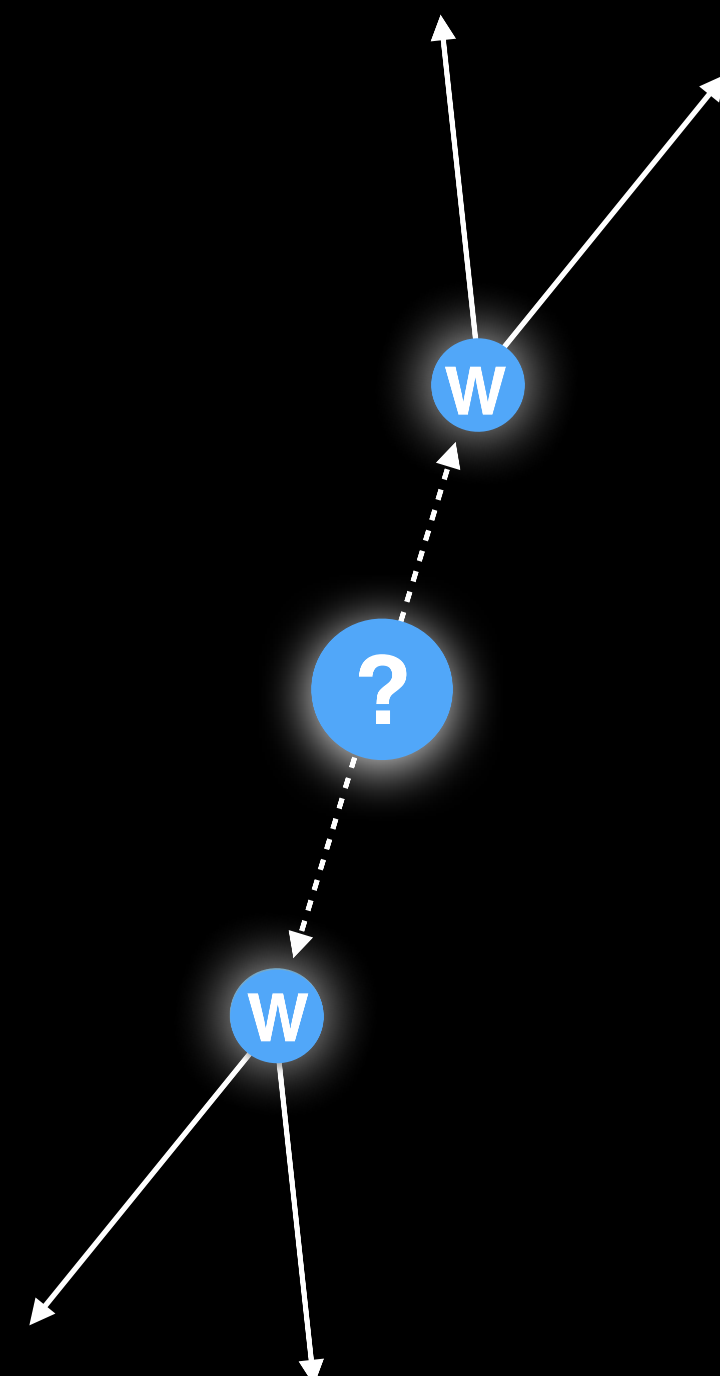




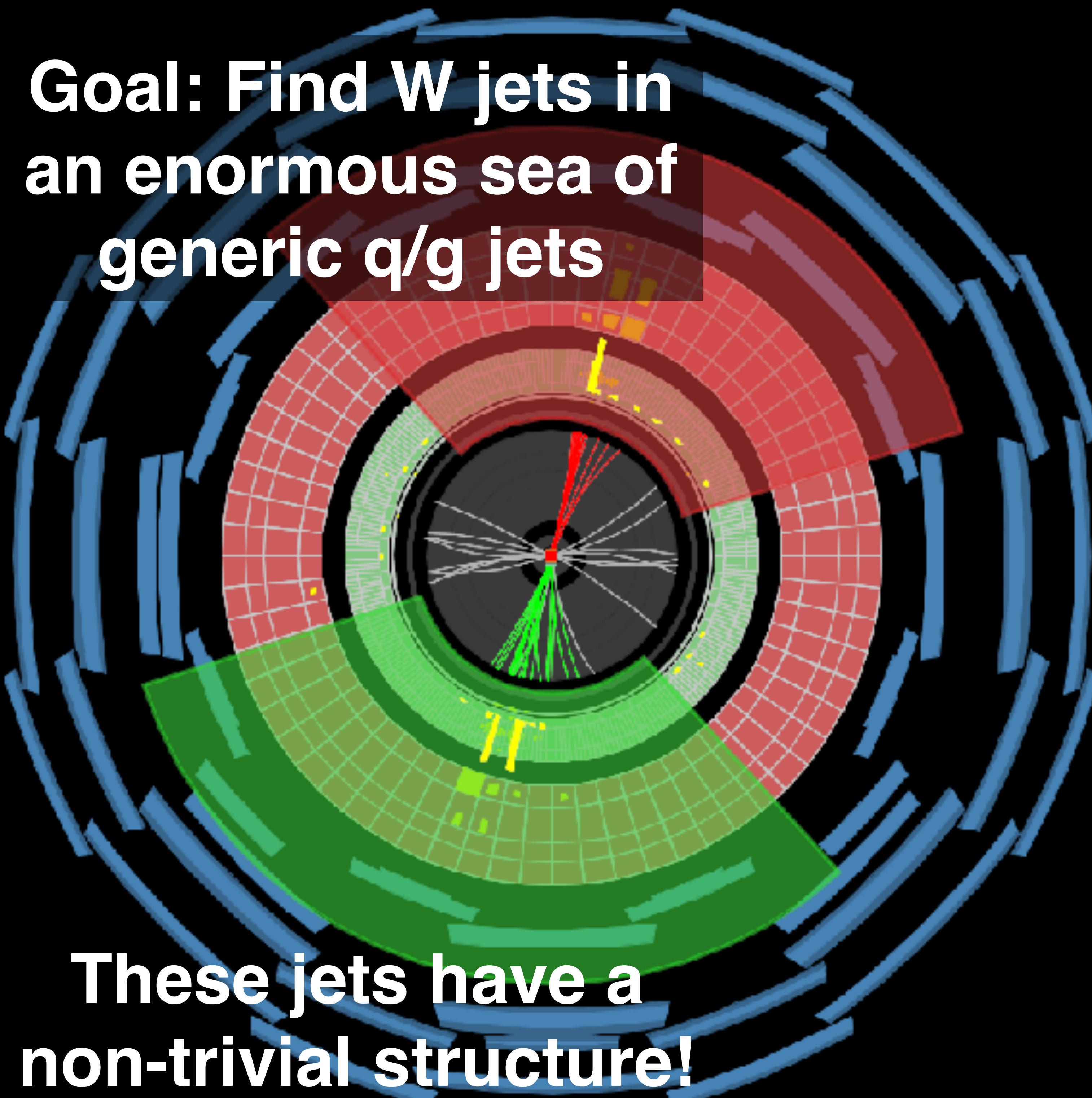


W bosons are naturally boosted if they result from the decay of something even heavier



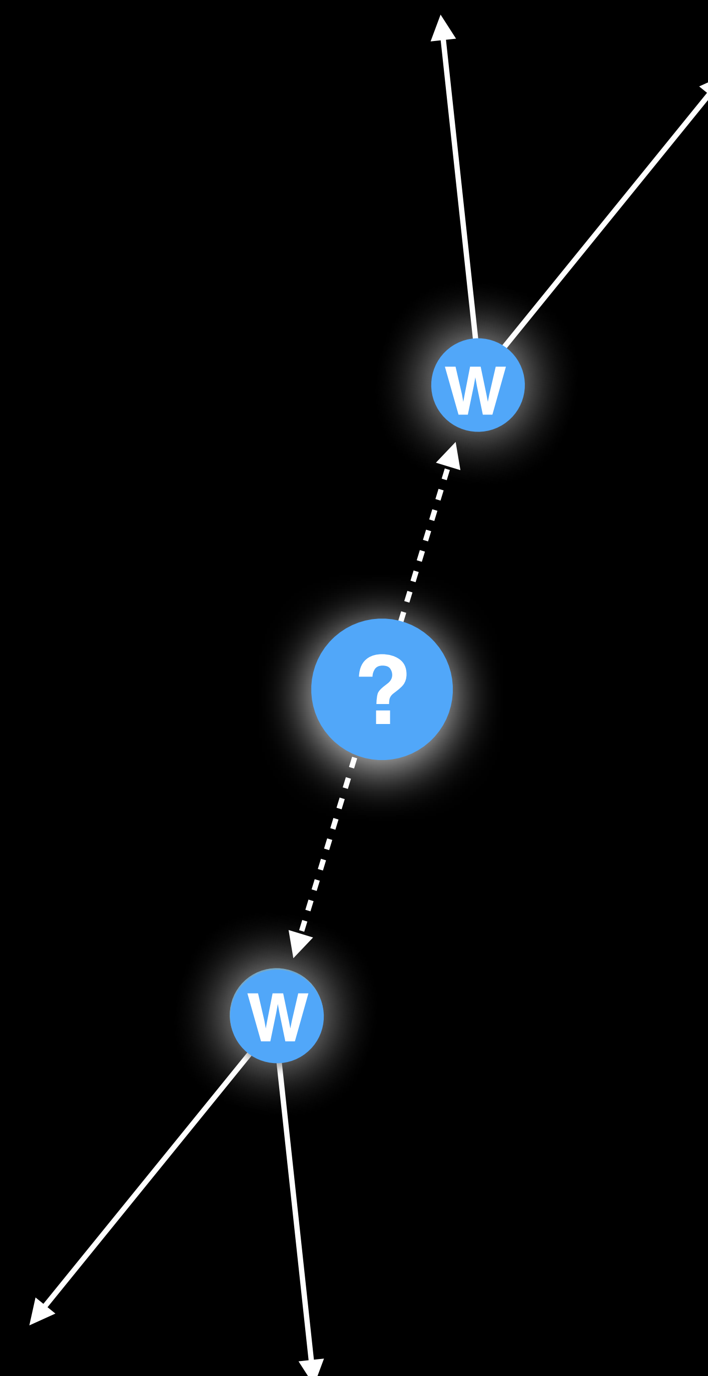


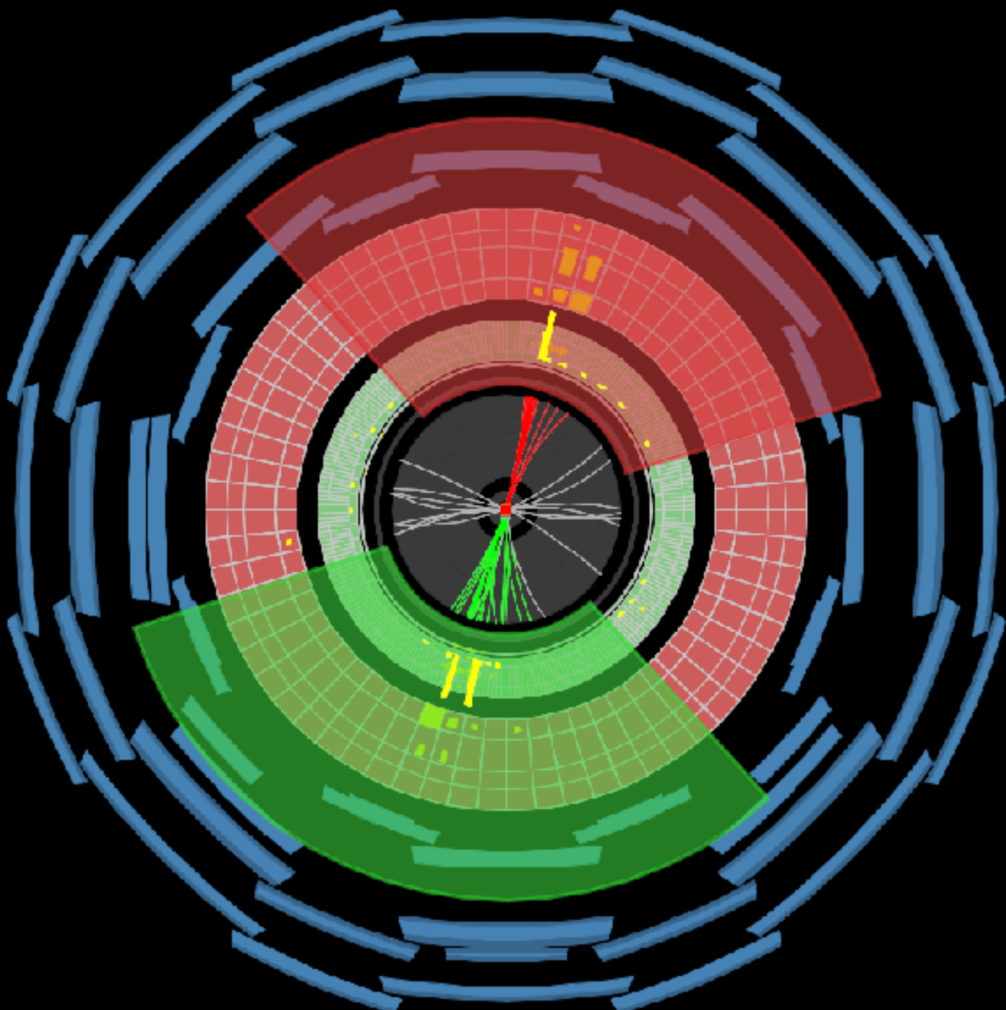
W bosons are naturally boosted if they result from the decay of something even heavier



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

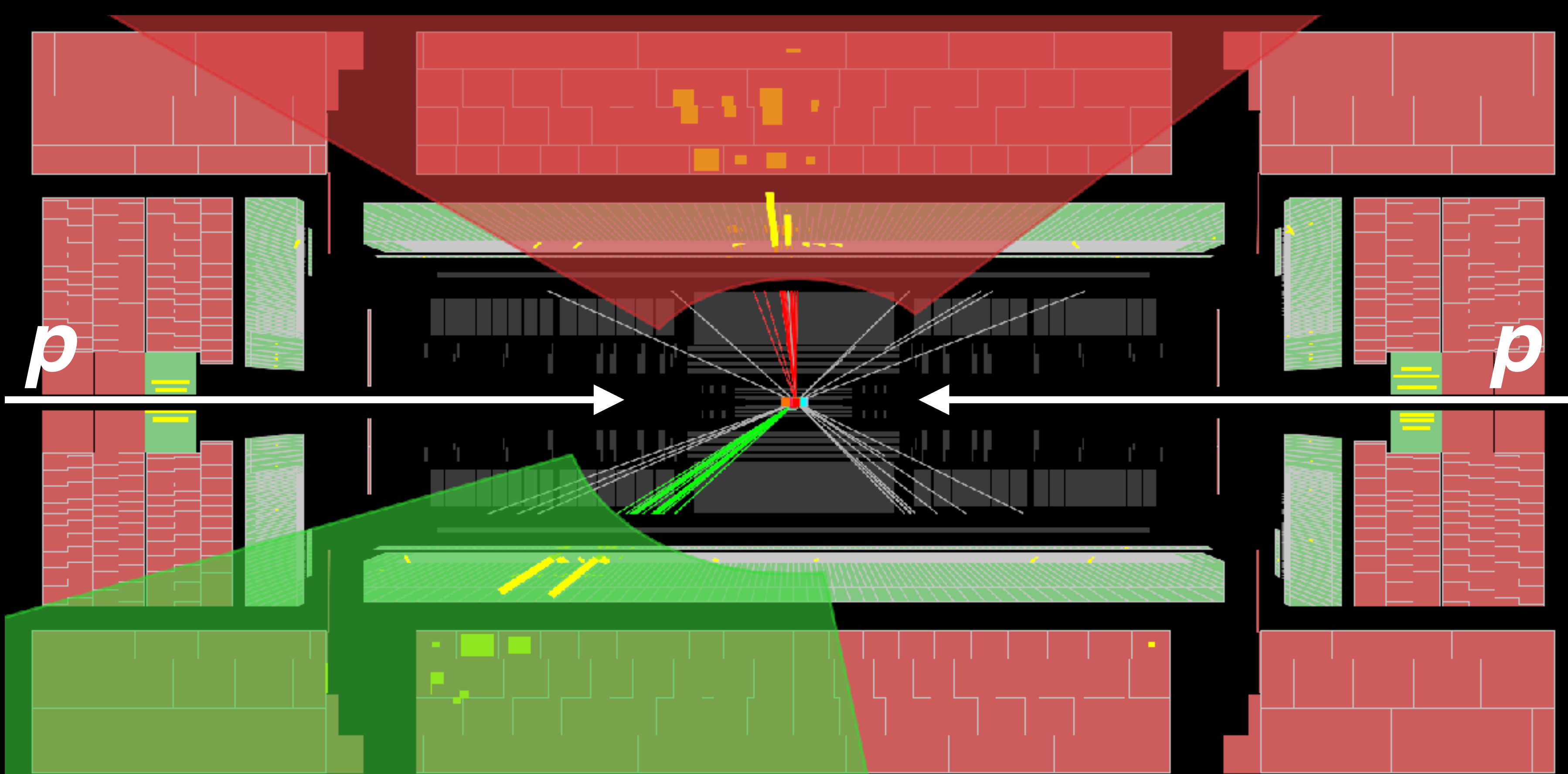
These jets have a non-trivial structure!

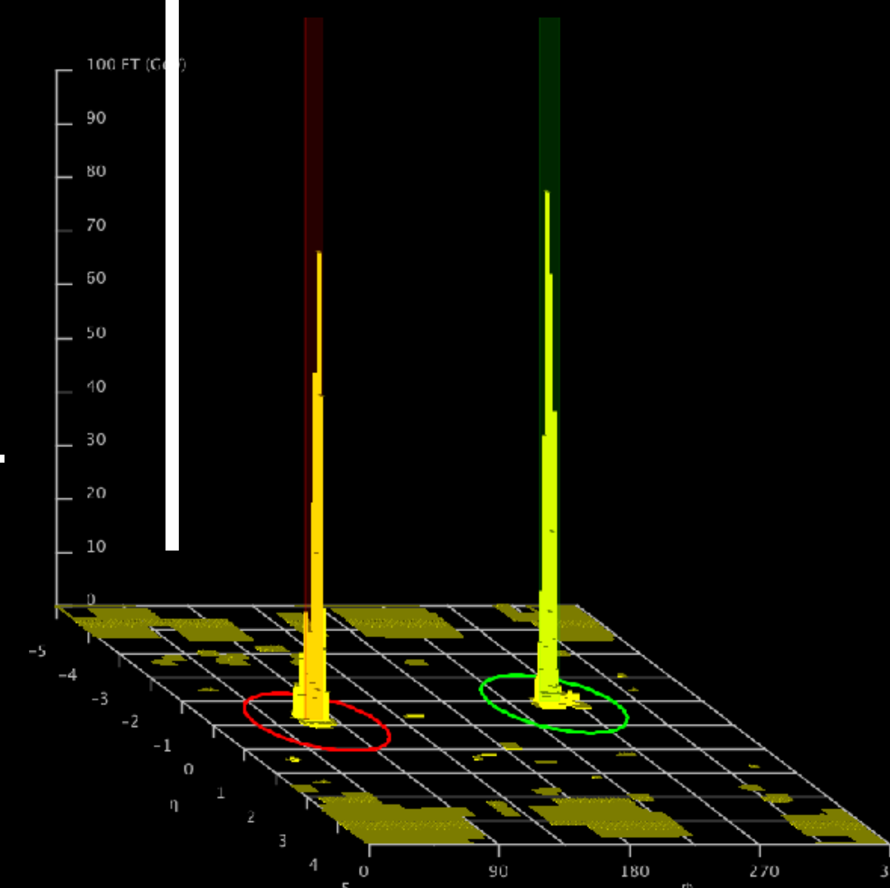




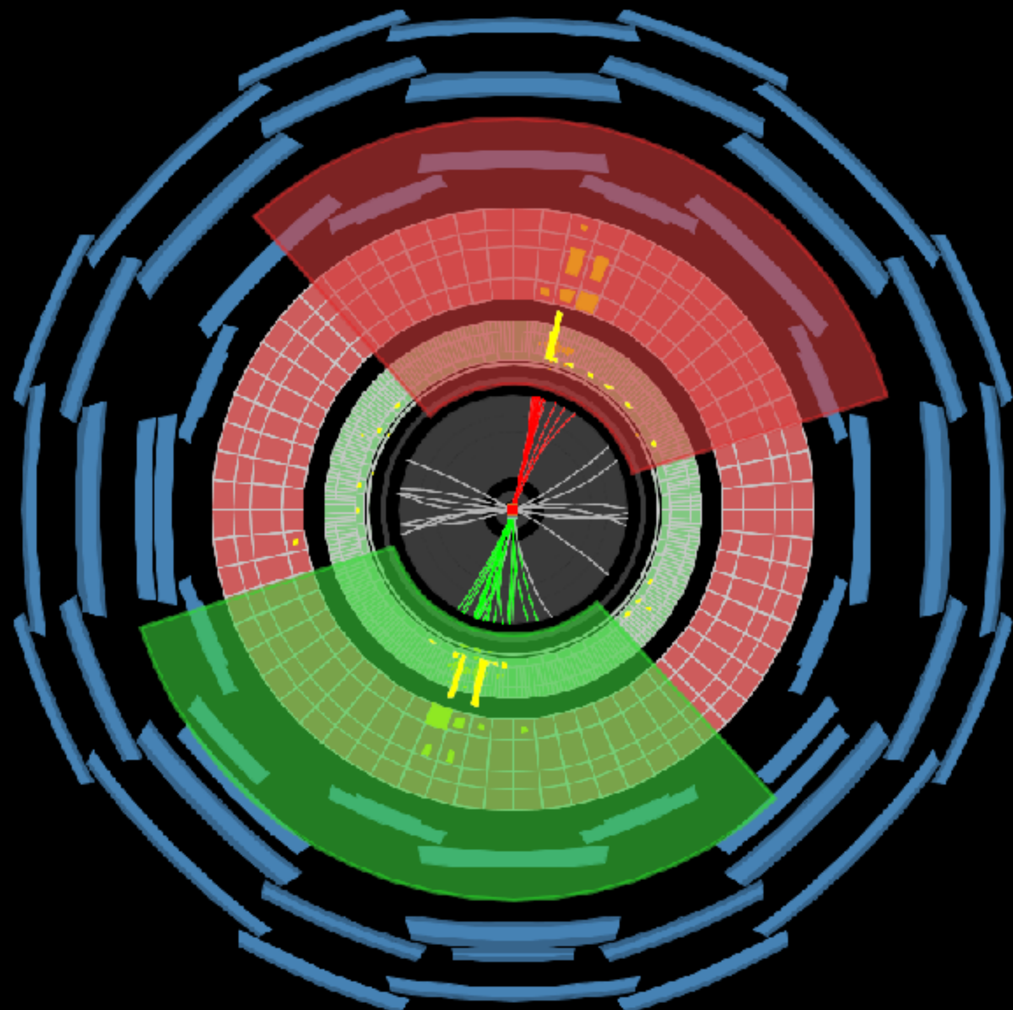
Searching for new particles decaying into boosted W bosons requires **looking at the radiation pattern inside jets**

momentum transverse to the beam (p_T)

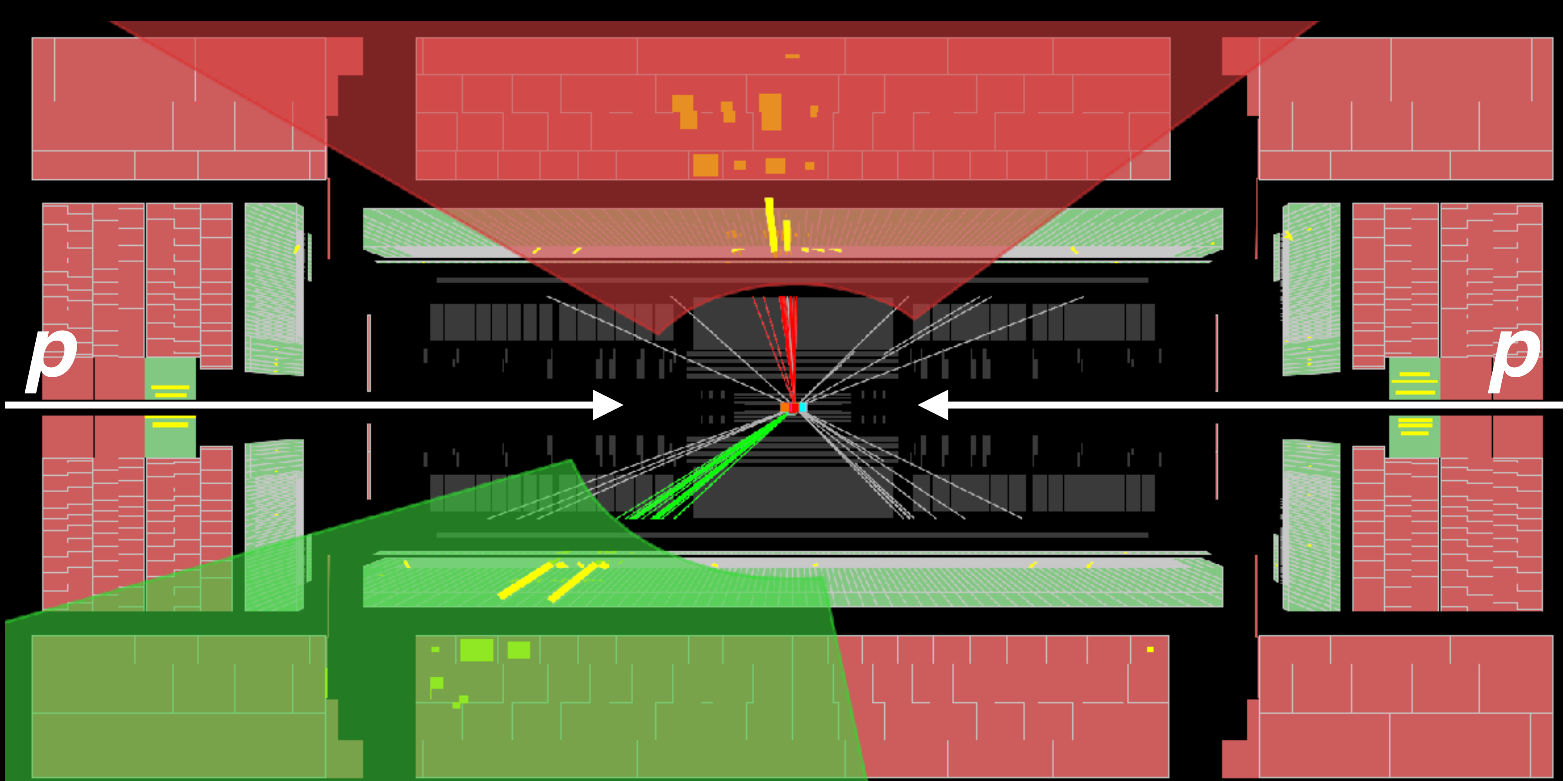


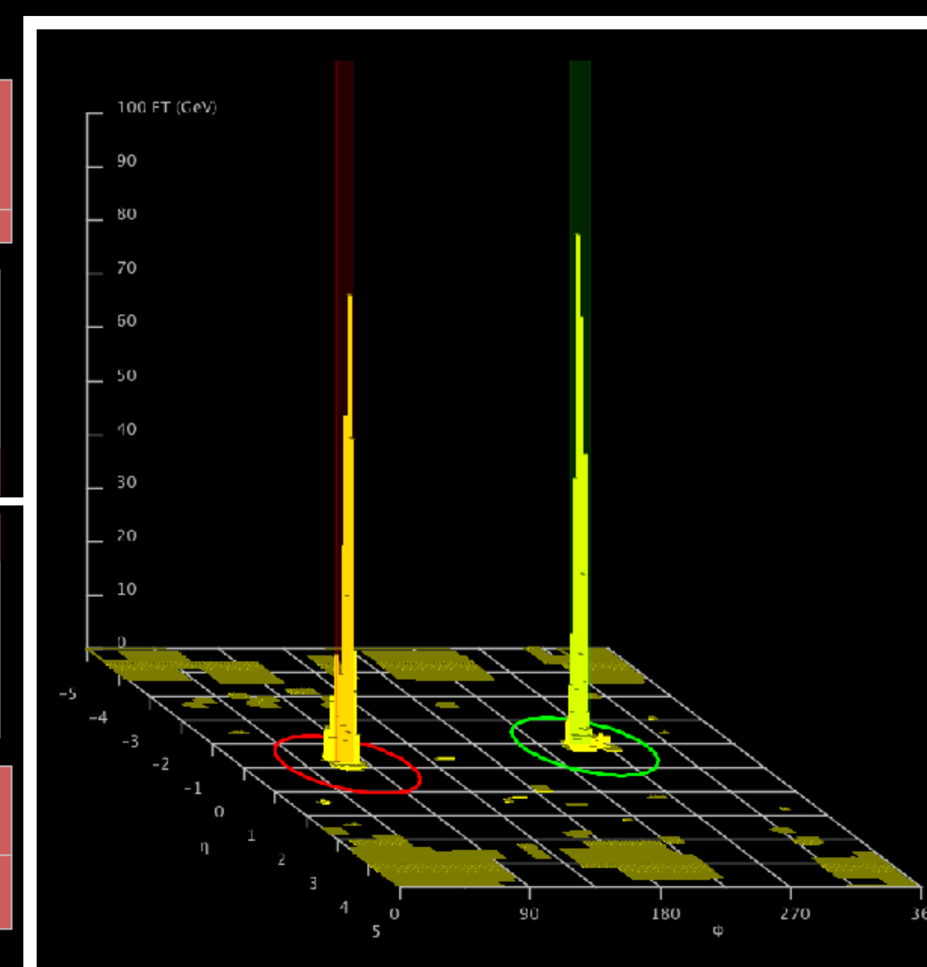


Up next: jet images



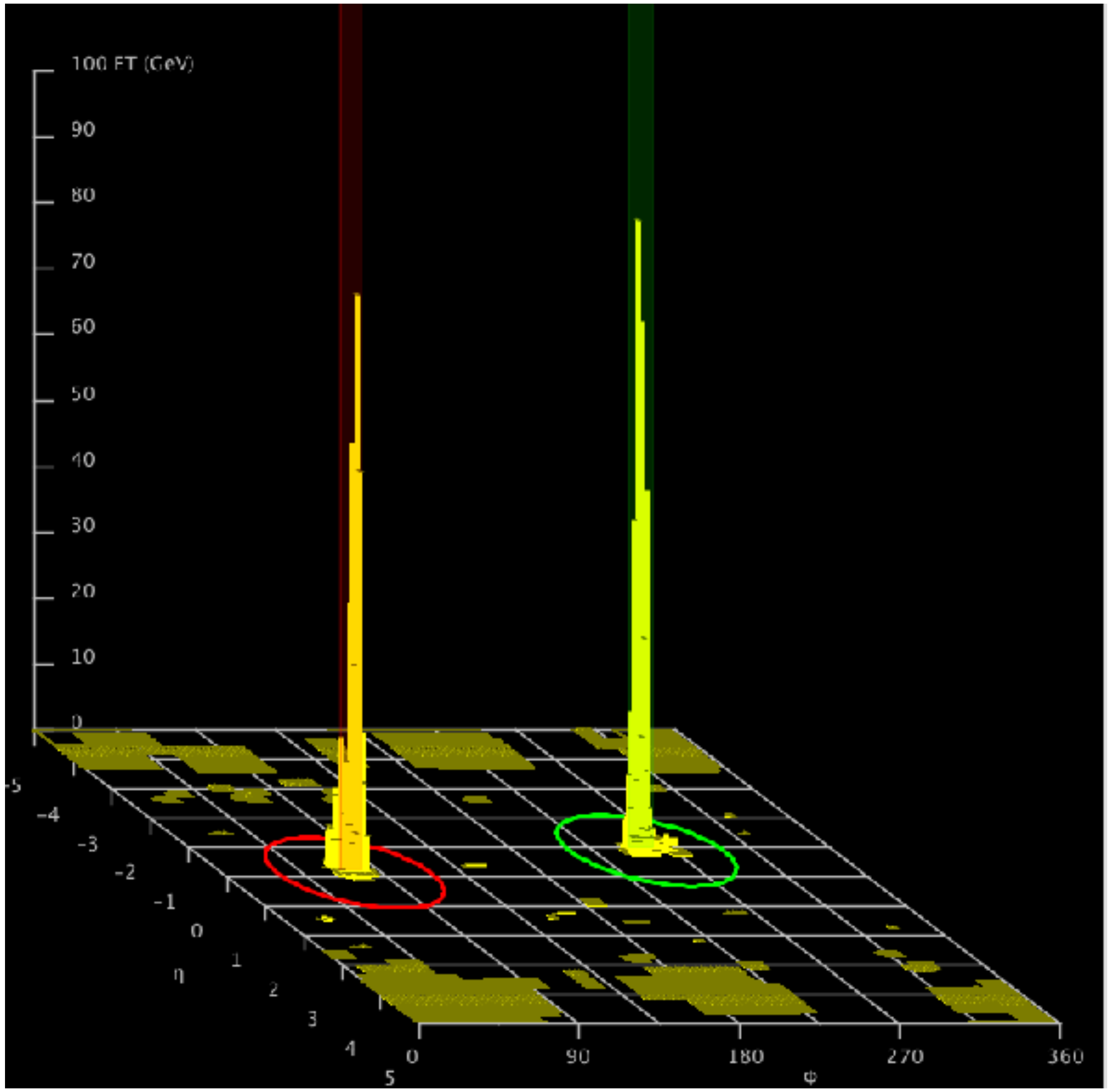
like a digital image!

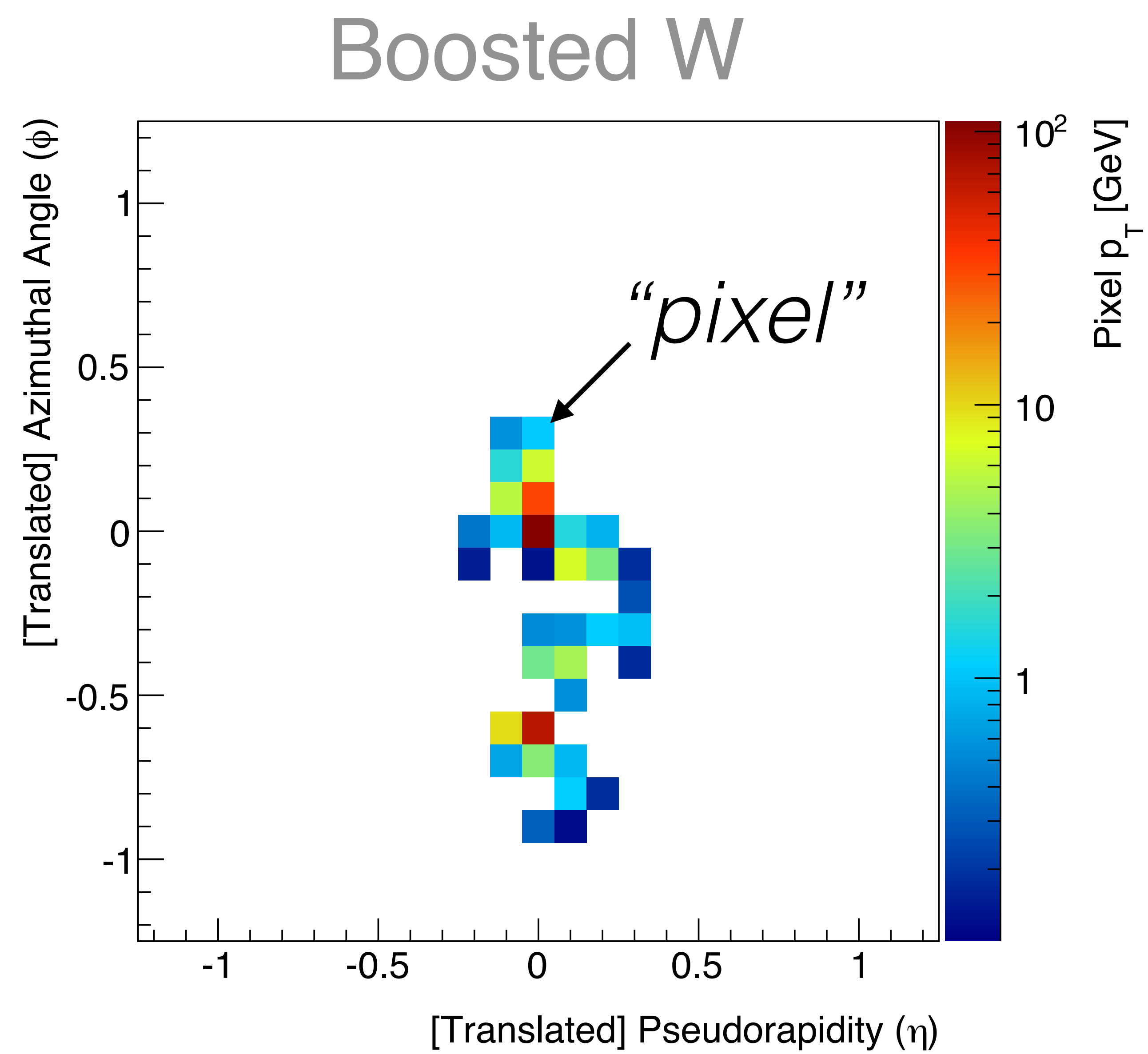




the Jet Image

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



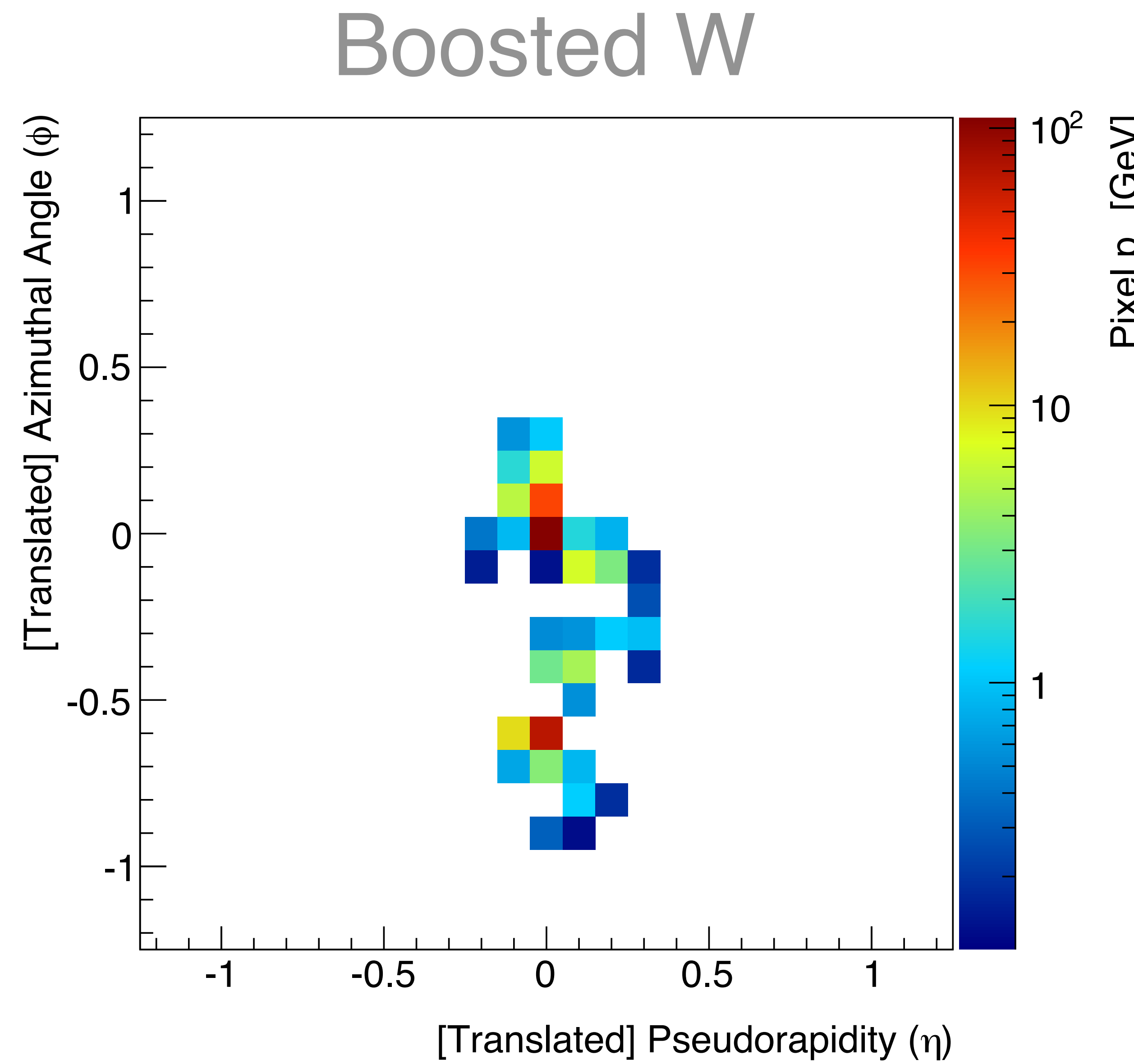


L. de Oliveira, M. Paganini, **BPN**, Comp. and Software for Big Science (2017) 1 *nothing like a 'natural' image!*

the Jet Image

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

Credit: Peter G Trimming (Wikipedia)



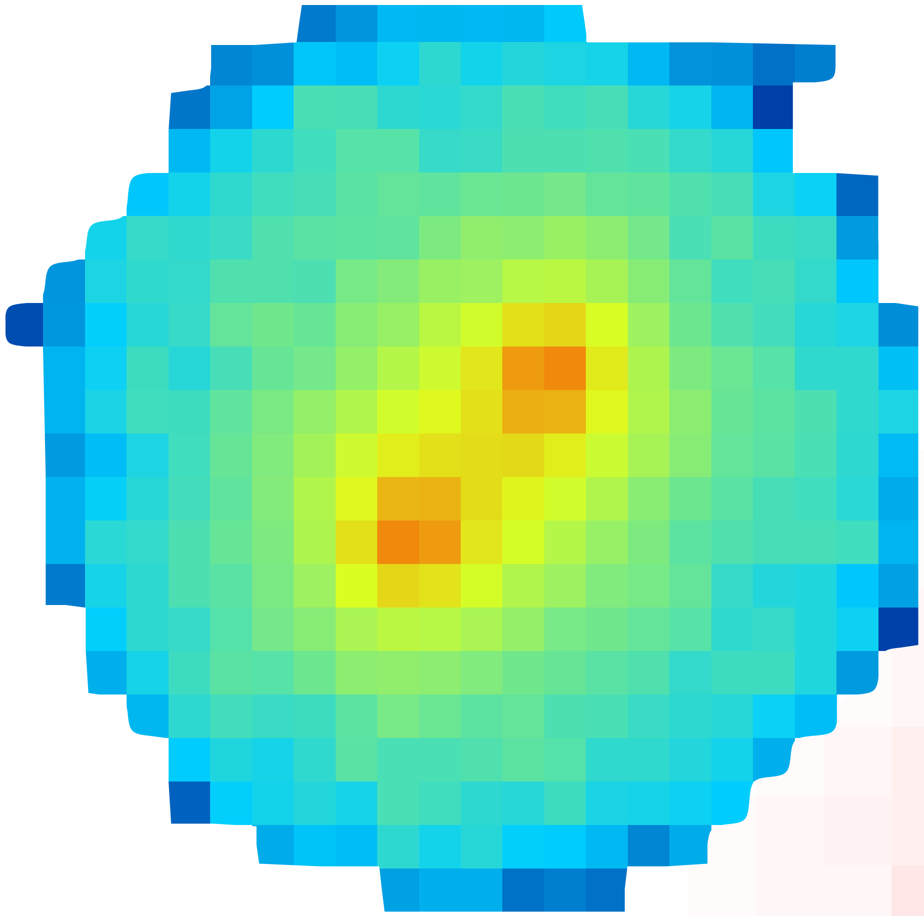
no smooth edges, clear features, low occupancy (number of hit pixels)

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

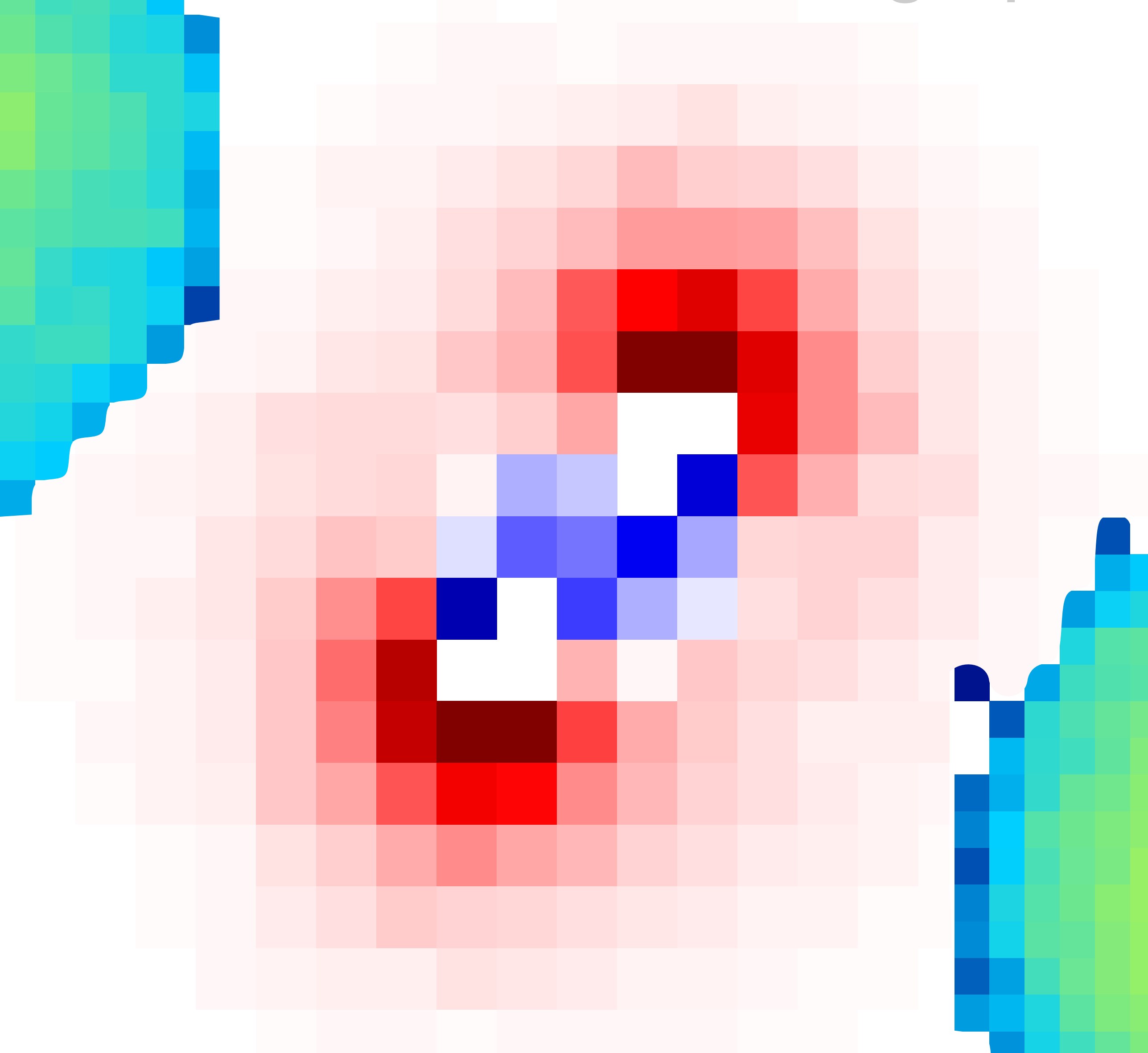
Why images?

Can directly visualize physics

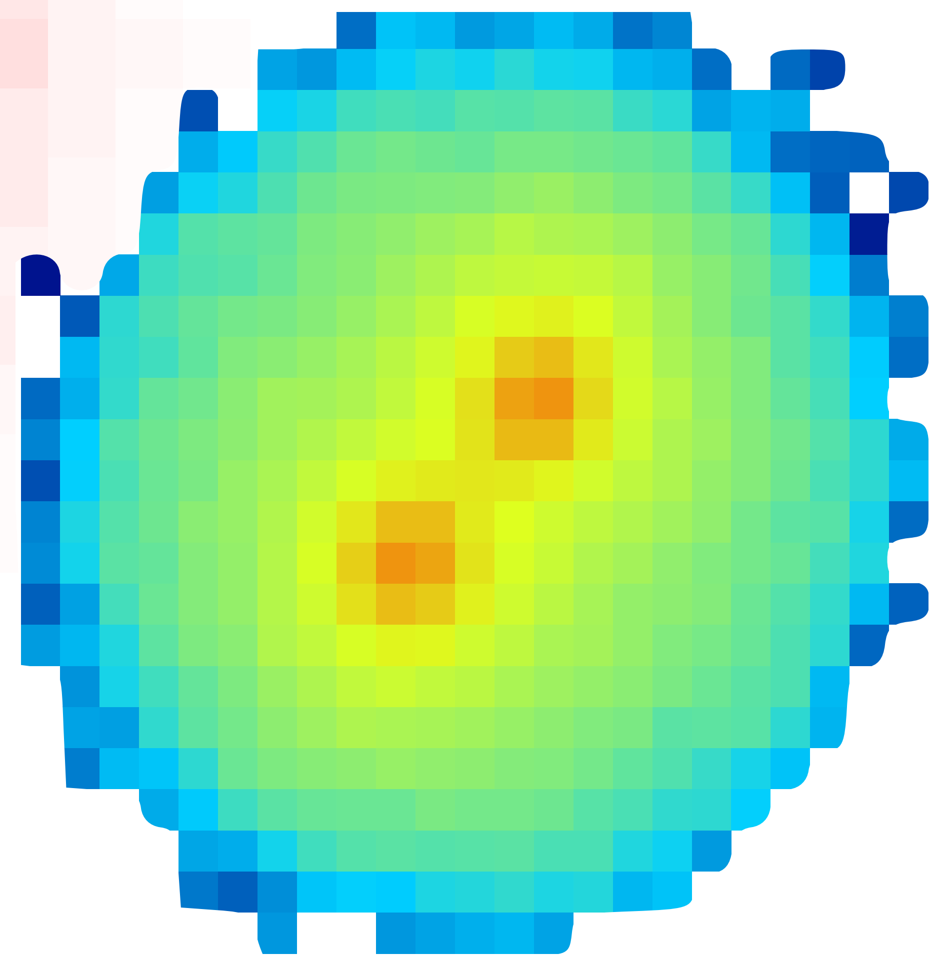
and we can benefit from the extensive image processing literature



W ⇢ $q\bar{q}$



g ⇢ $q\bar{q}$

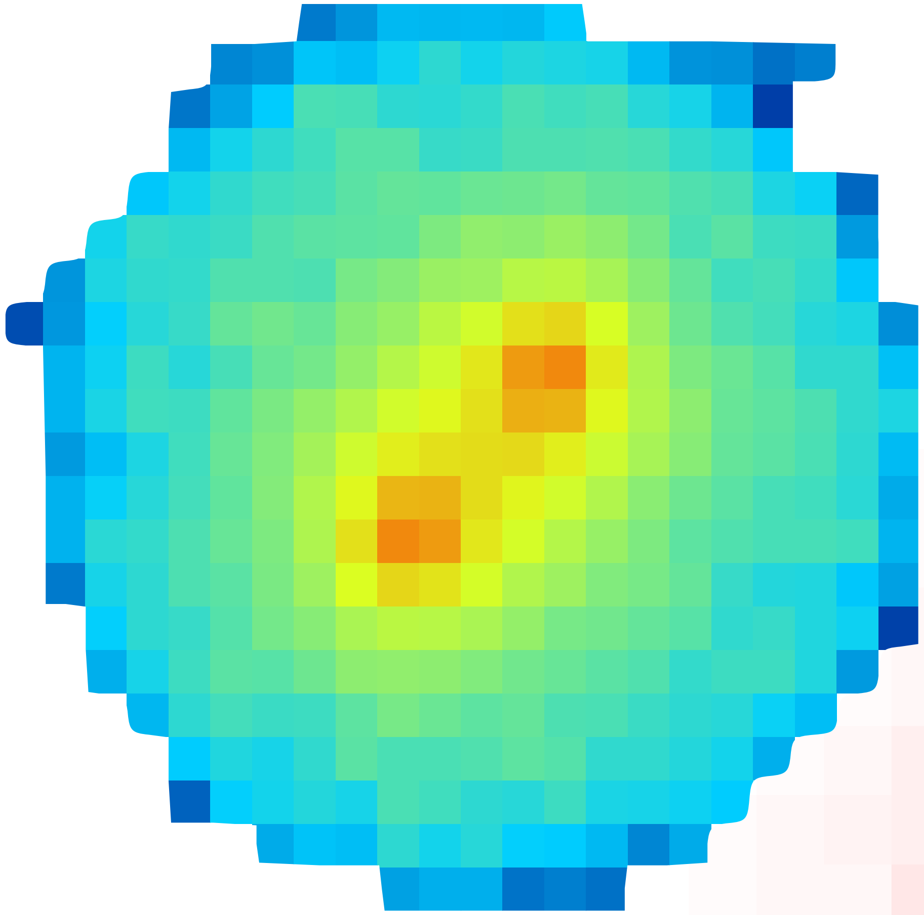


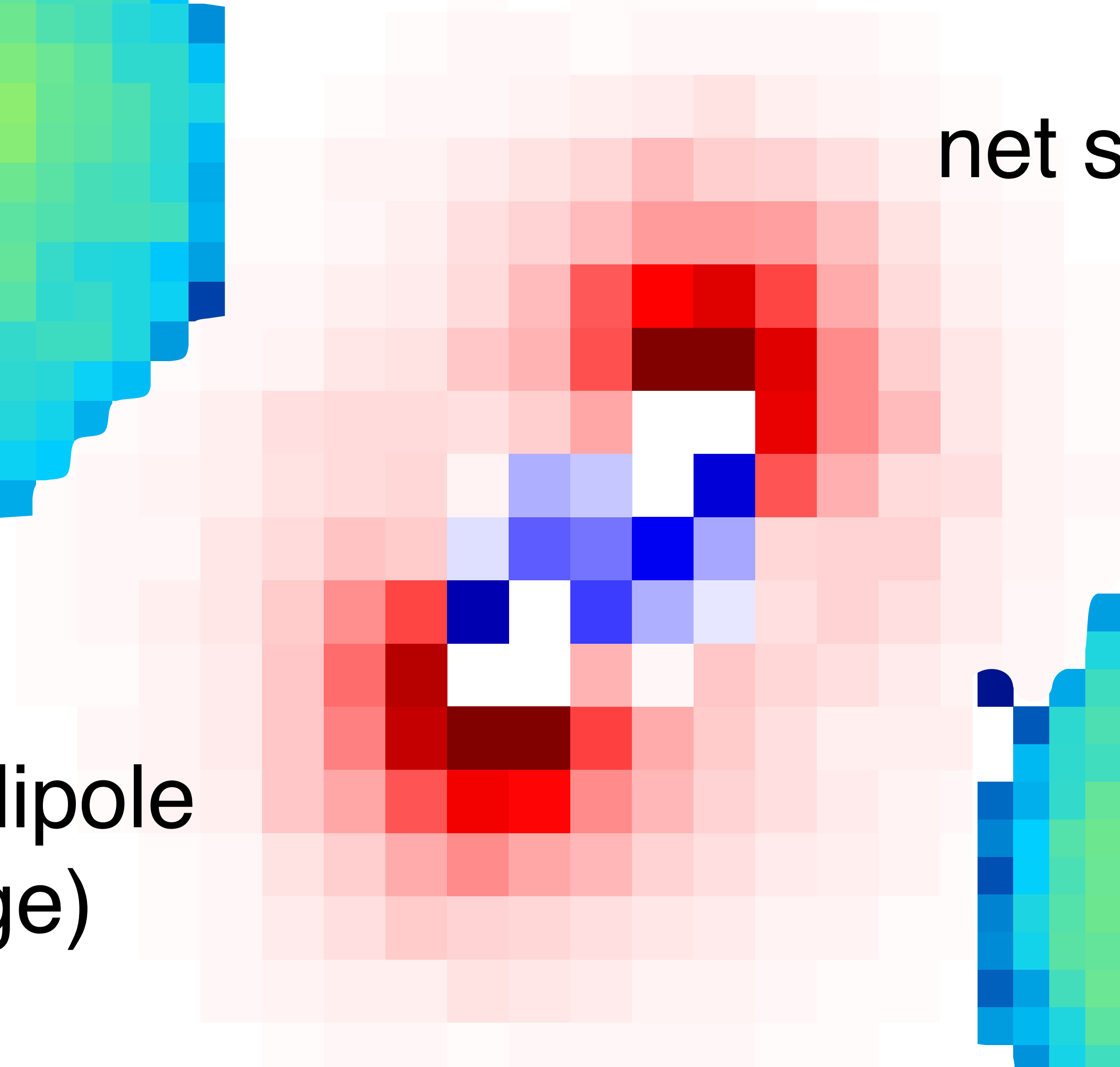
there is information encoded in the physical distance between pixels

Why images?

Can directly visualize physics

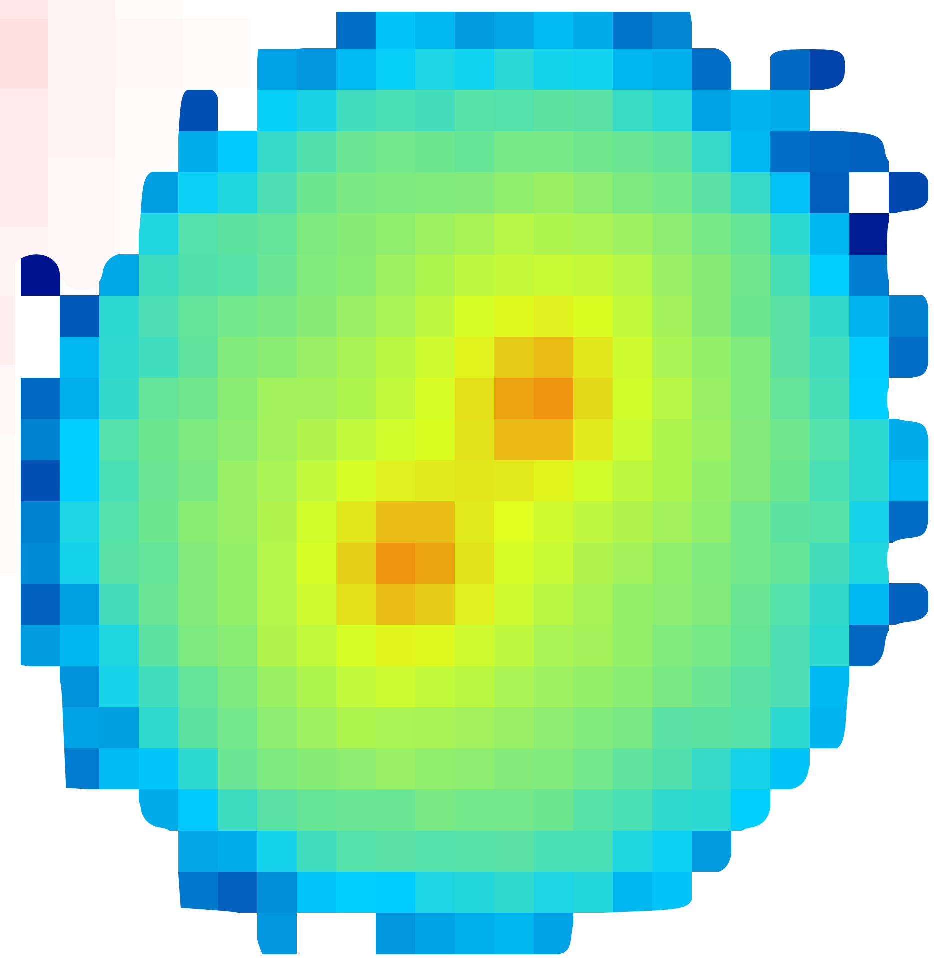
and we can benefit from the extensive image processing literature





net strong-force charge

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



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

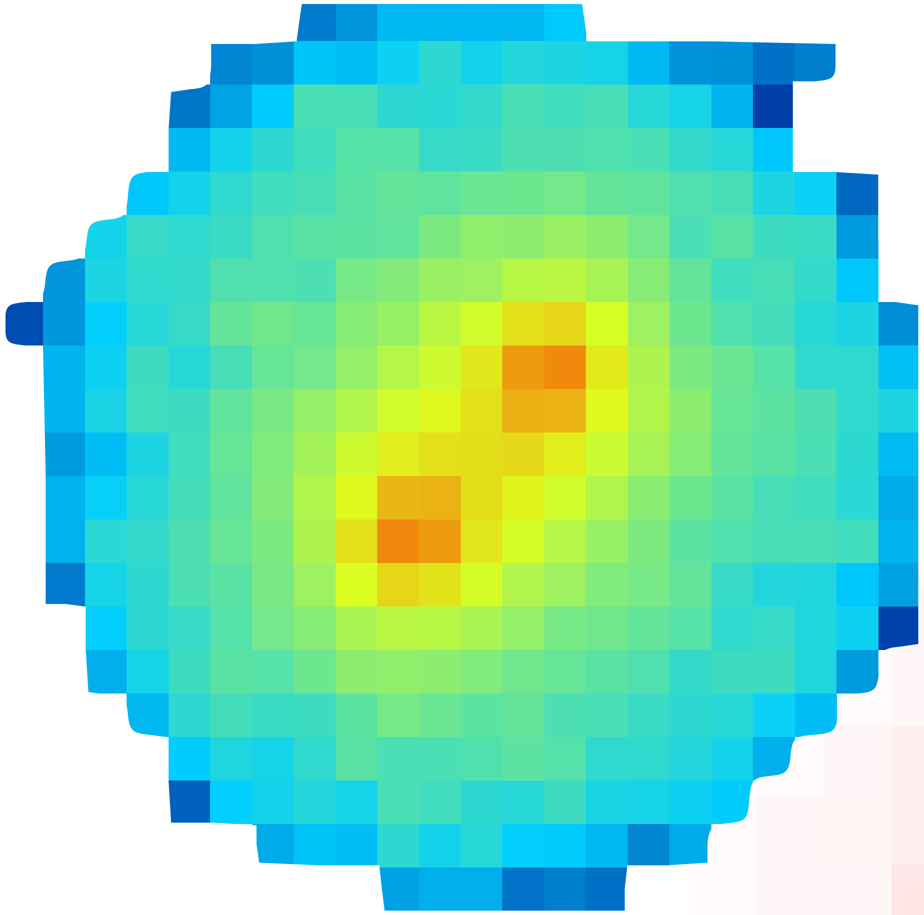
radiates like a dipole (no net charge)

there is information encoded in the physical distance between pixels

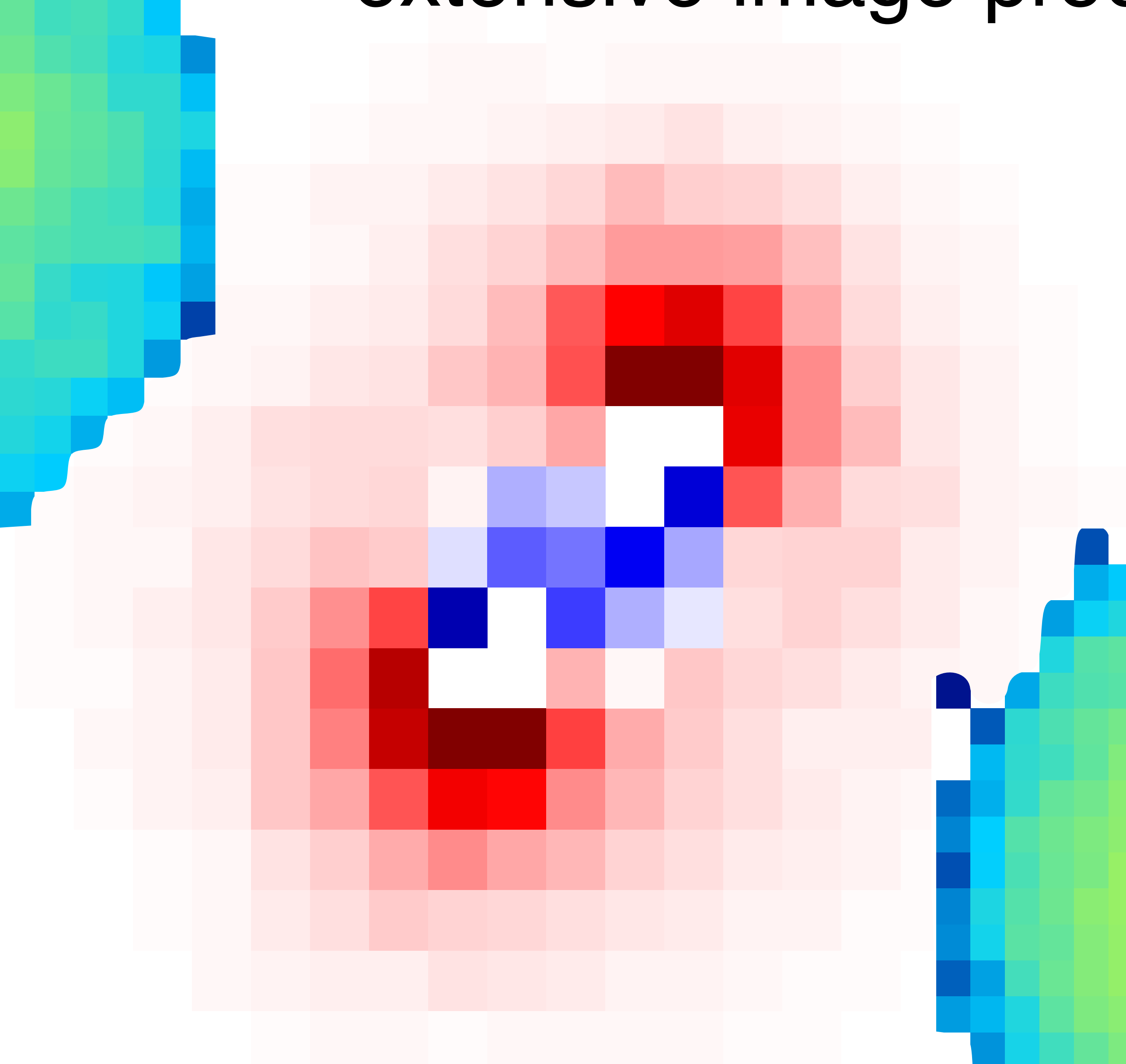
Why images?

Can directly visualize physics

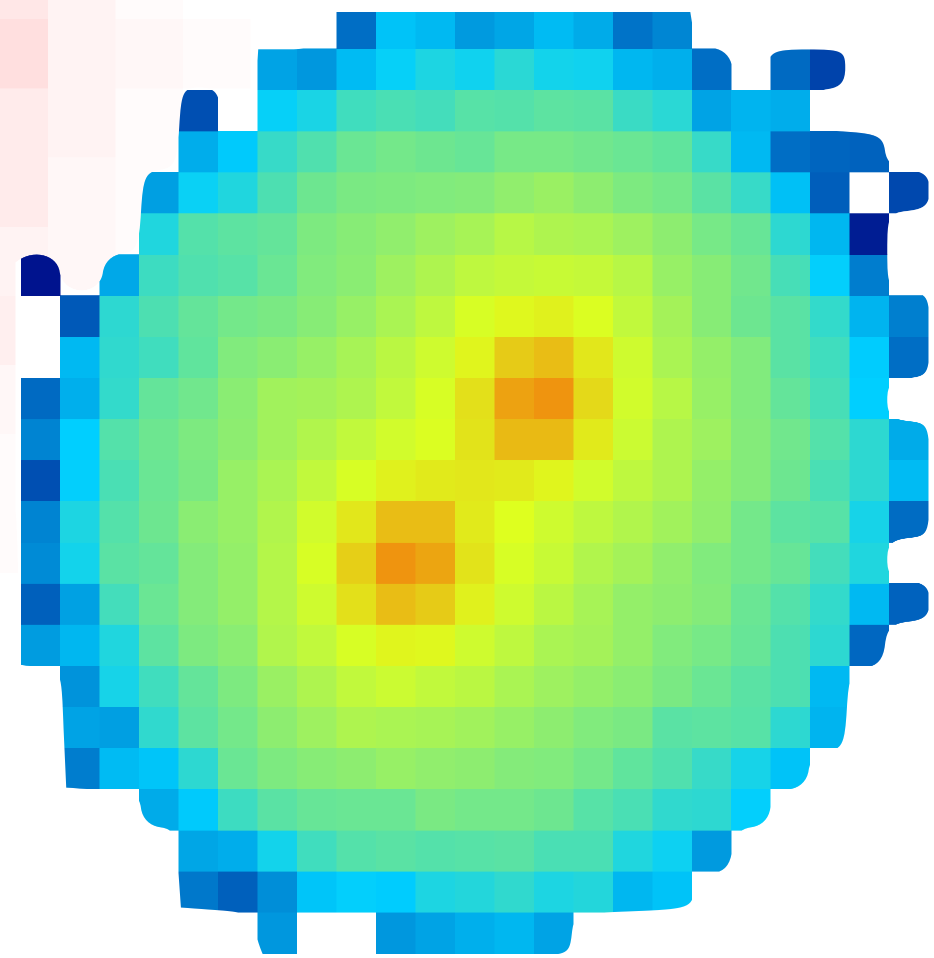
and we can benefit from the extensive image processing literature



W ⇢ $q\bar{q}$



g ⇢ $q\bar{q}$

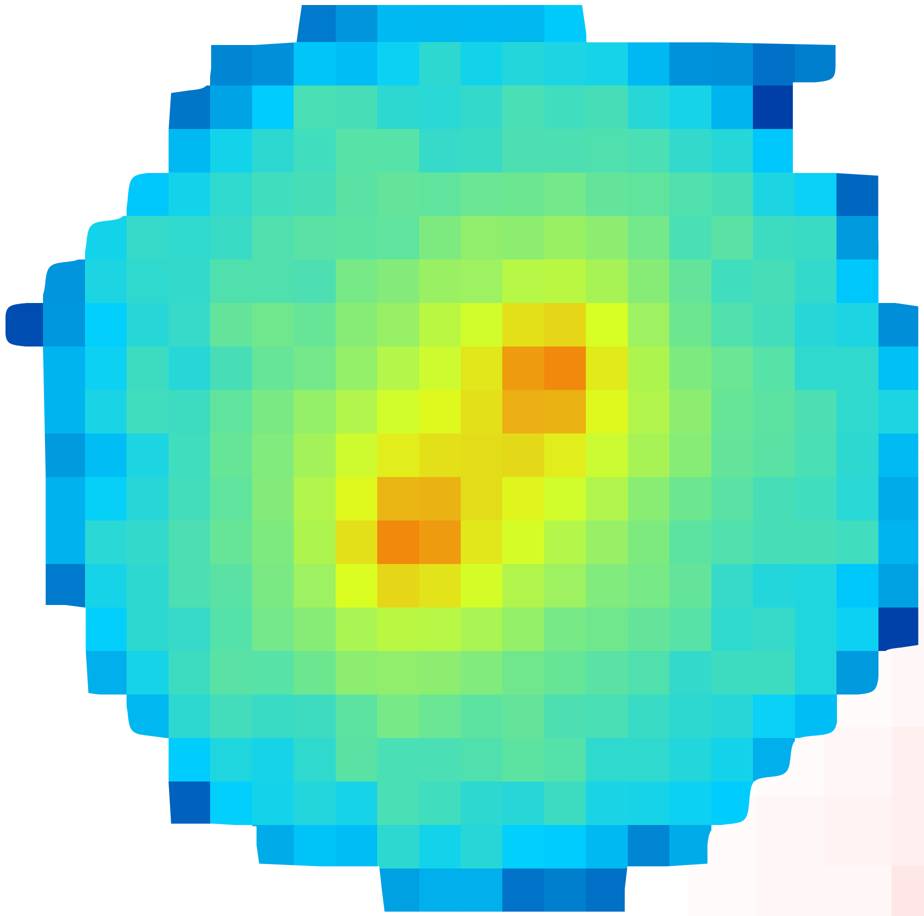


there is information encoded in the physical distance between pixels

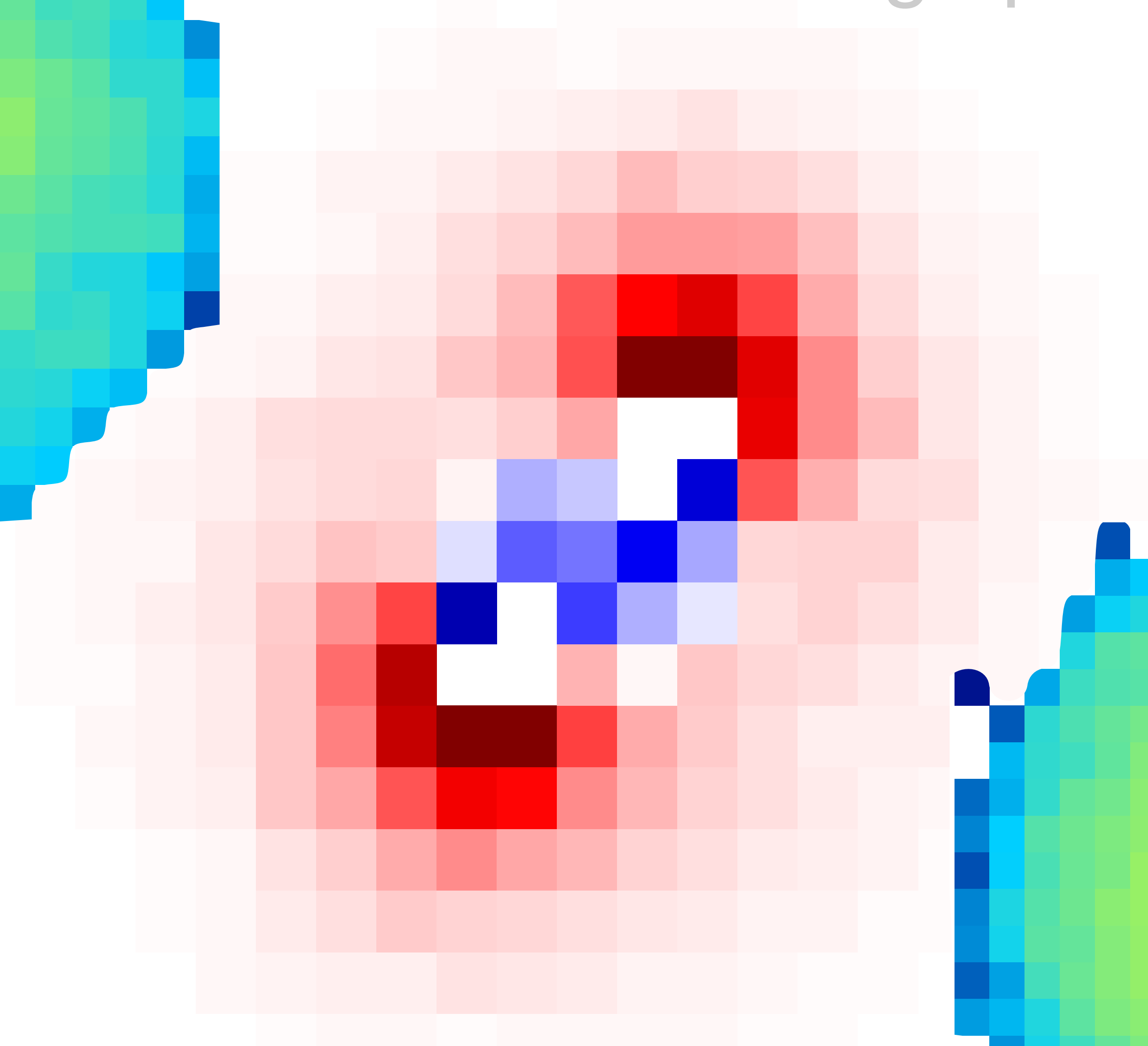
Why images?

Can directly visualize physics

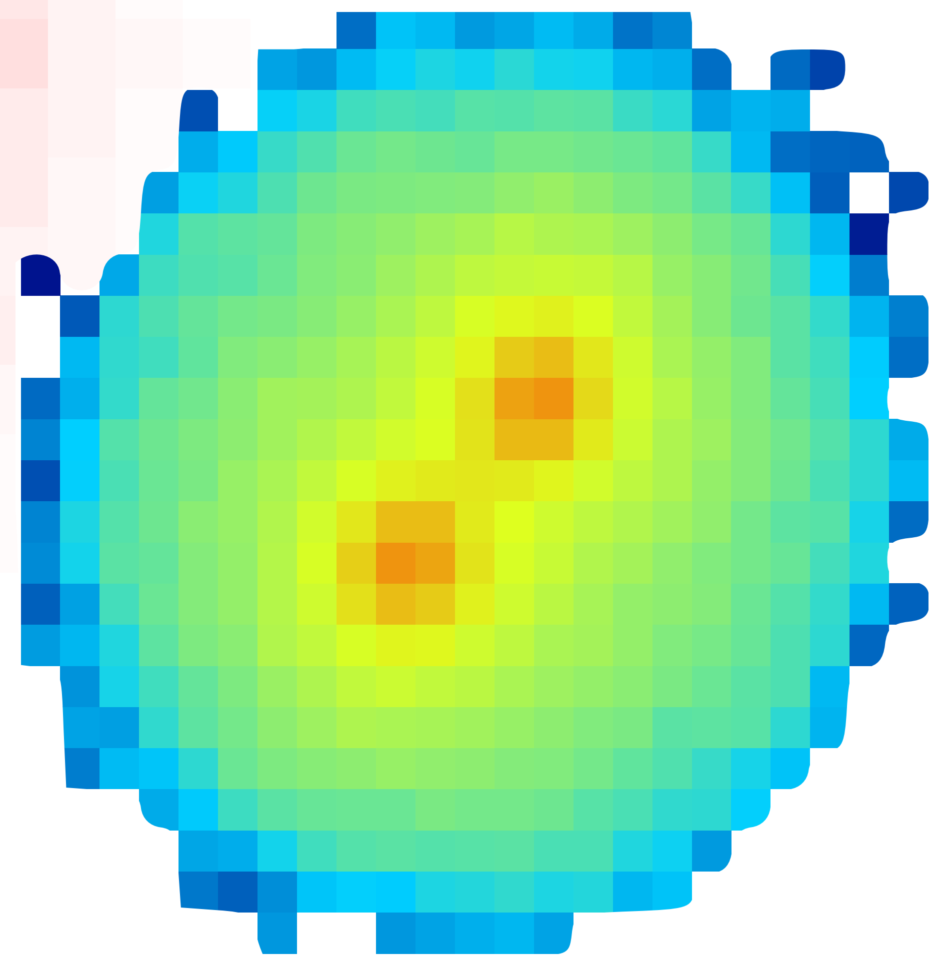
and we can benefit from the
extensive image processing literature



W ⋯→ $q\bar{q}$



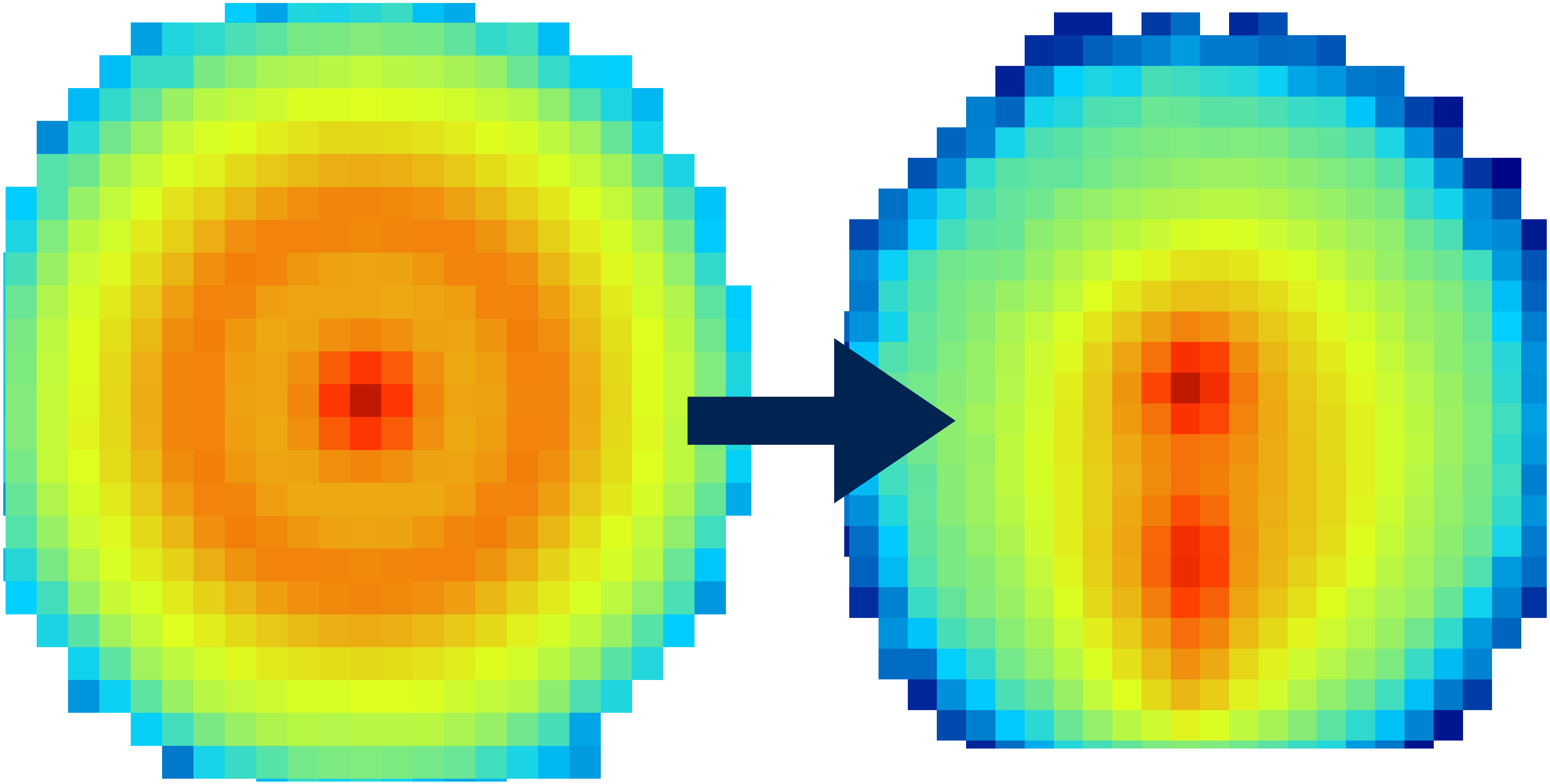
g ⋯→ $q\bar{q}$



there is information encoded in the
physical distance between pixels

Pre-processing & spacetime symmetries

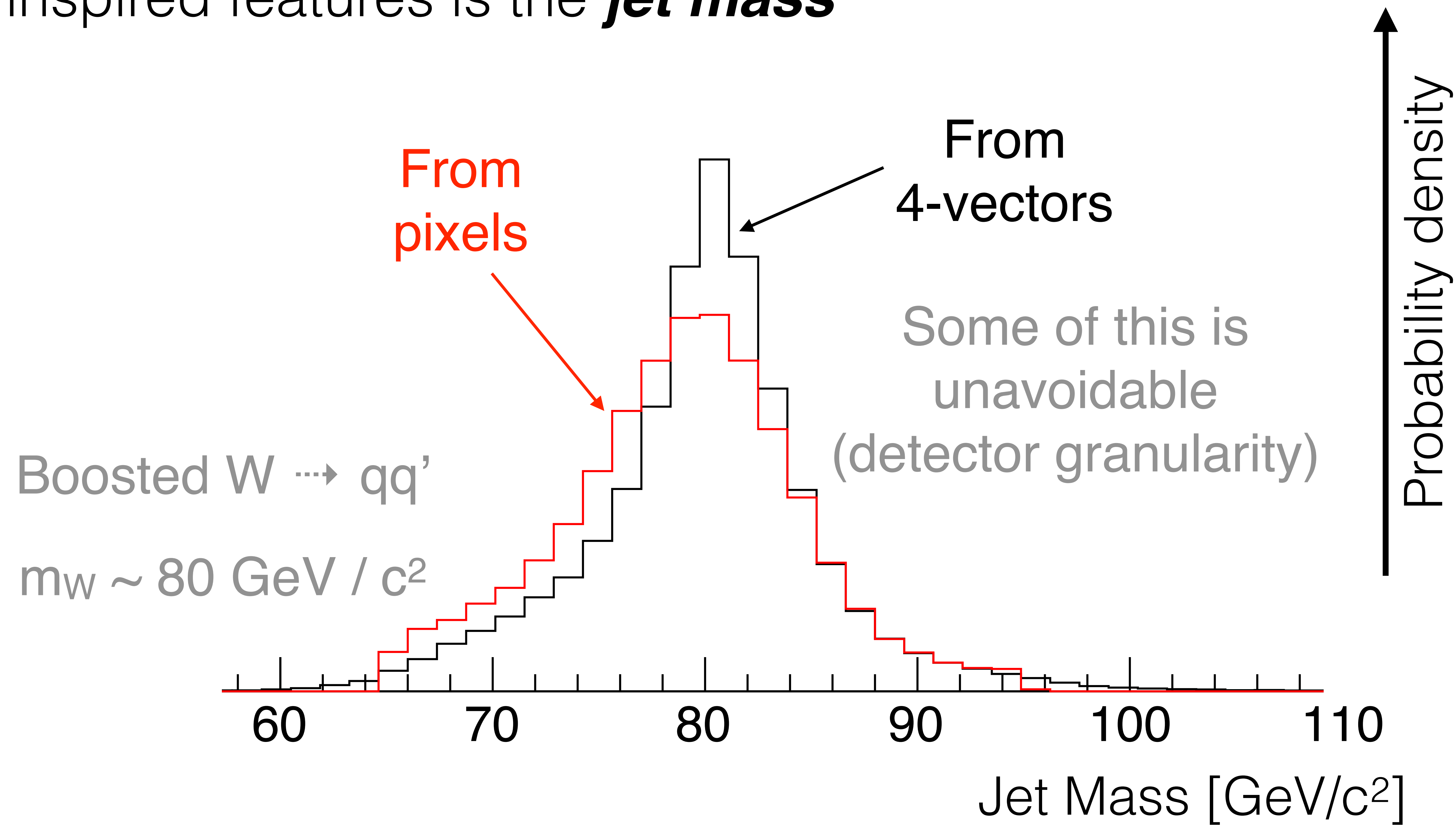
One of the first typical steps is pre-processing



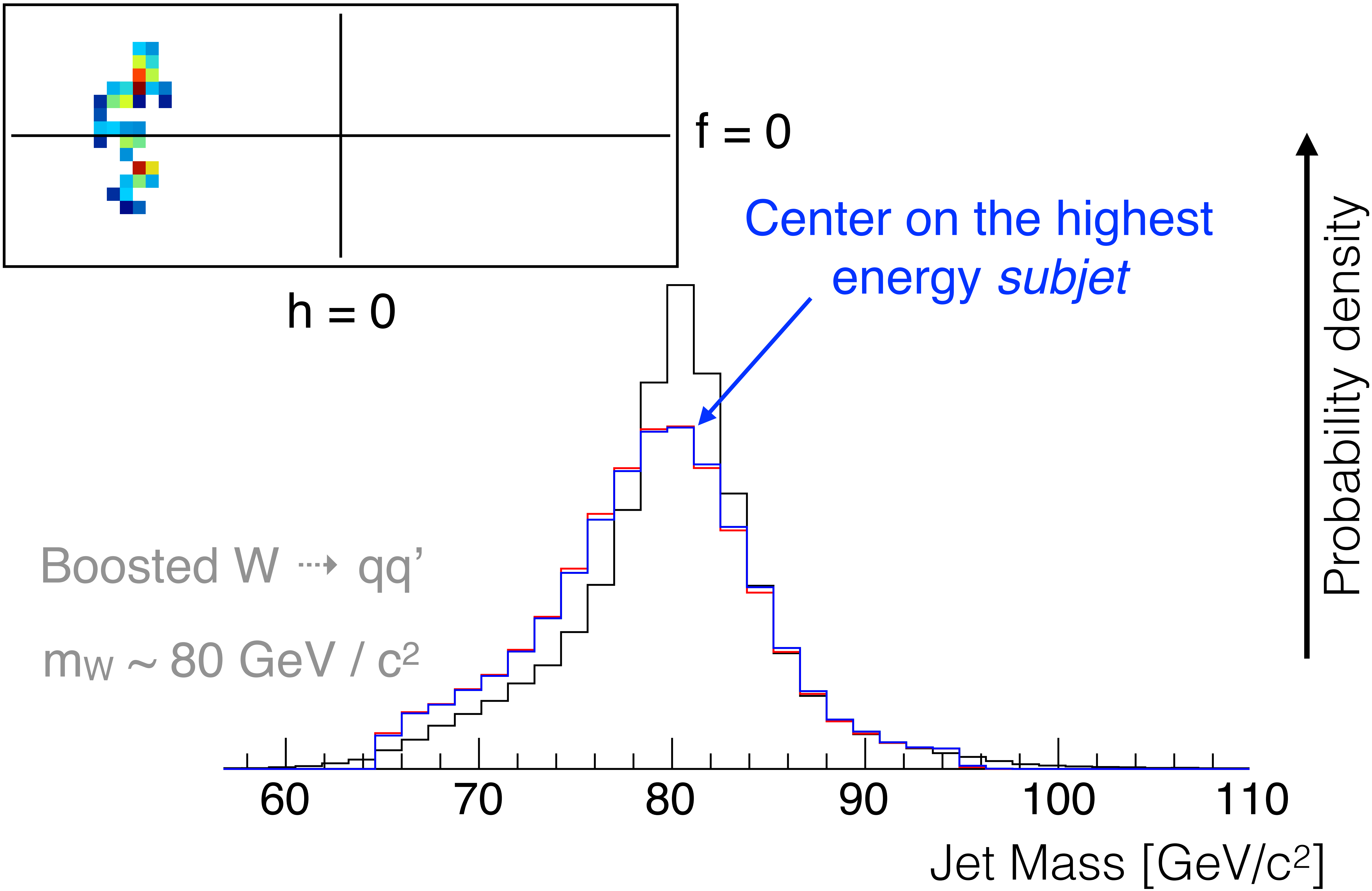
Can help to learn faster & smarter; but must be careful!

Pre-processing & spacetime symmetries

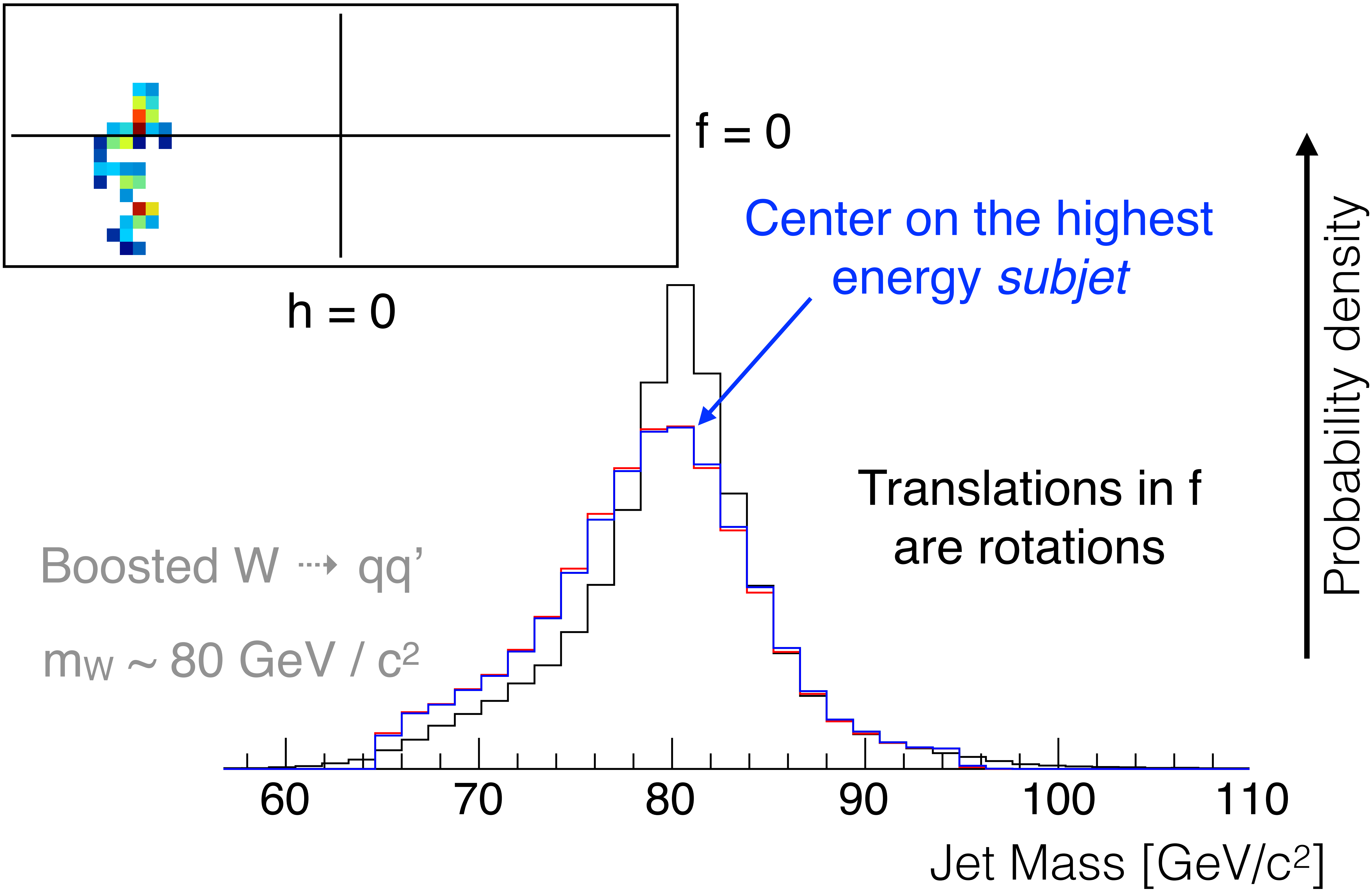
One of the most useful physics-inspired features is the ***jet mass***



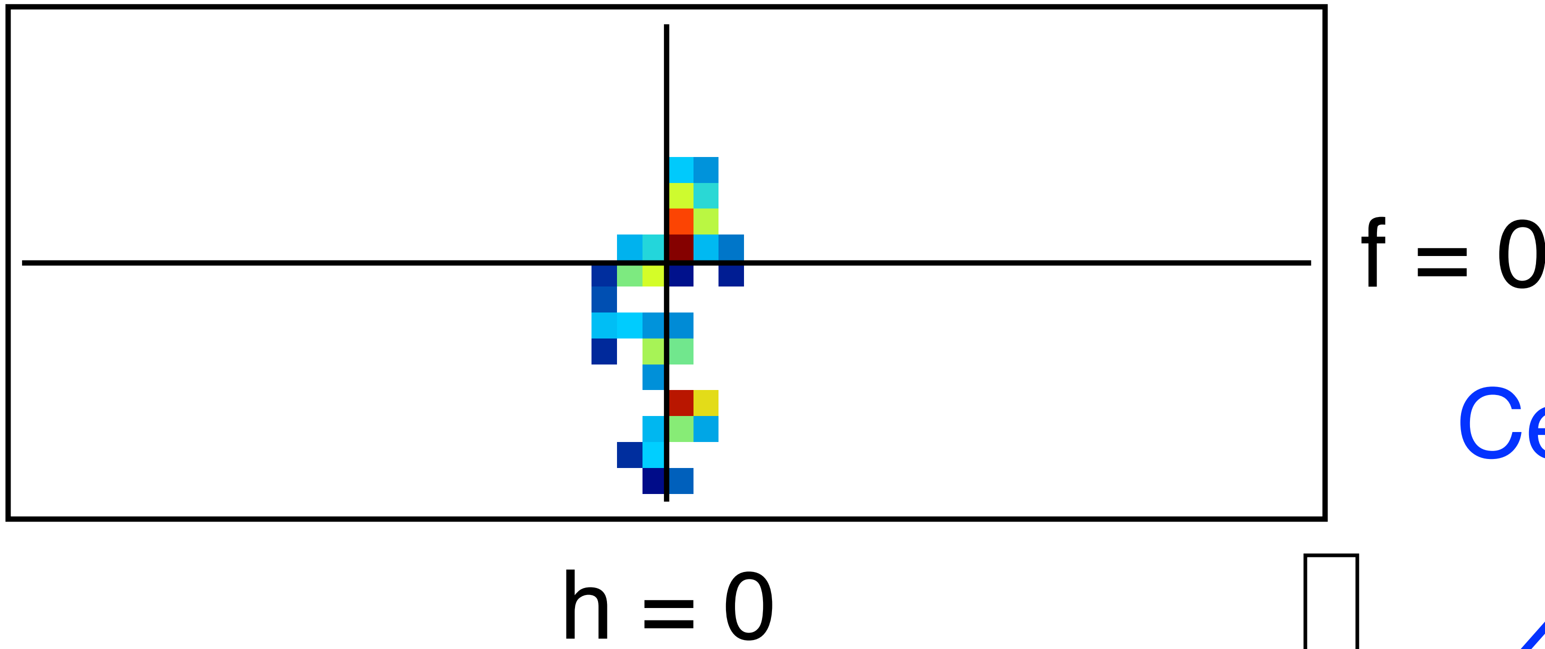
Pre-processing & spacetime symmetries

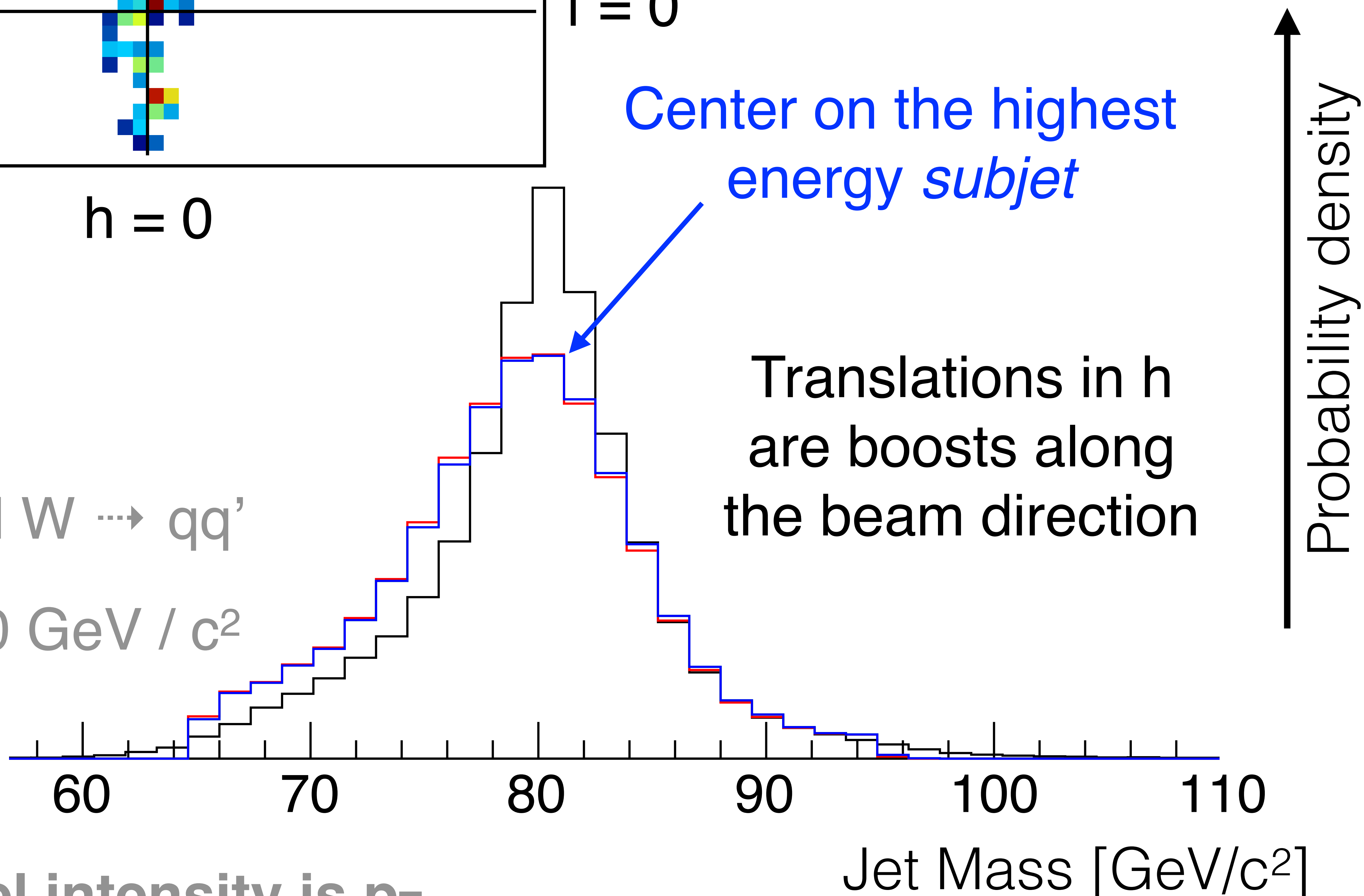


Pre-processing & spacetime symmetries



Pre-processing & spacetime symmetries





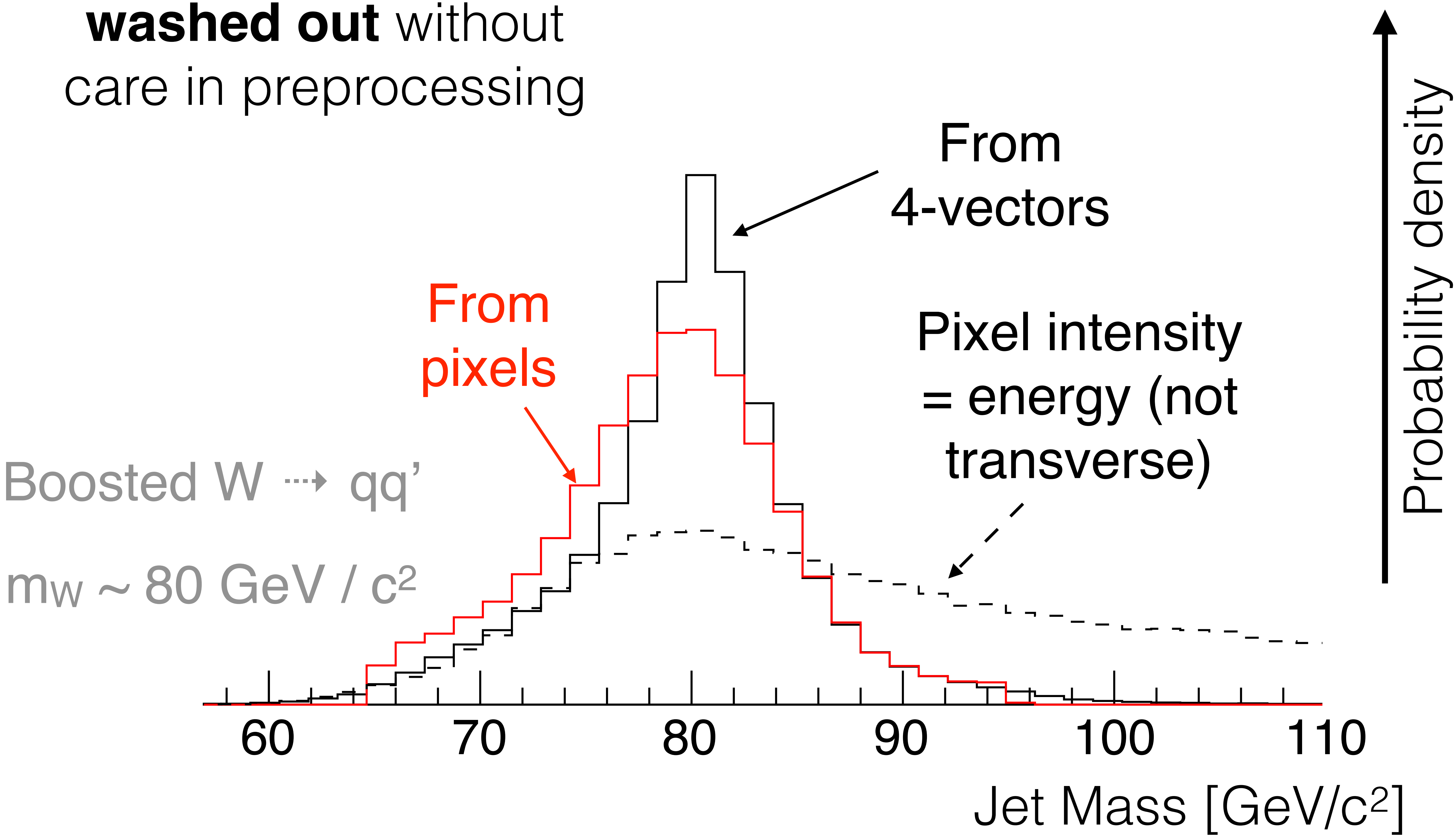
Boosted W ⇢ qq'

m_W ~ 80 GeV / c^2

N.B. pixel intensity is p_T

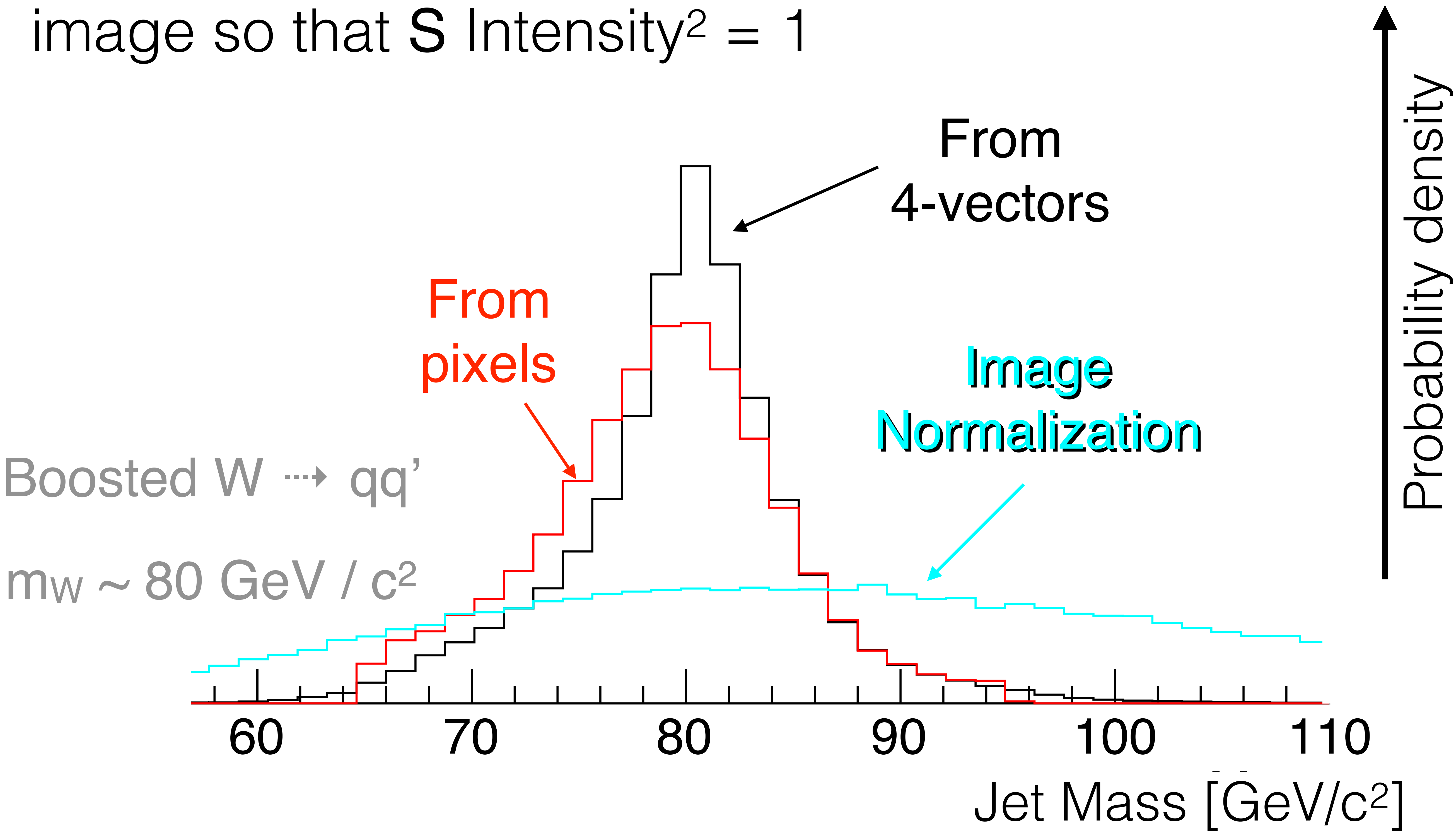
Pre-processing & spacetime symmetries

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

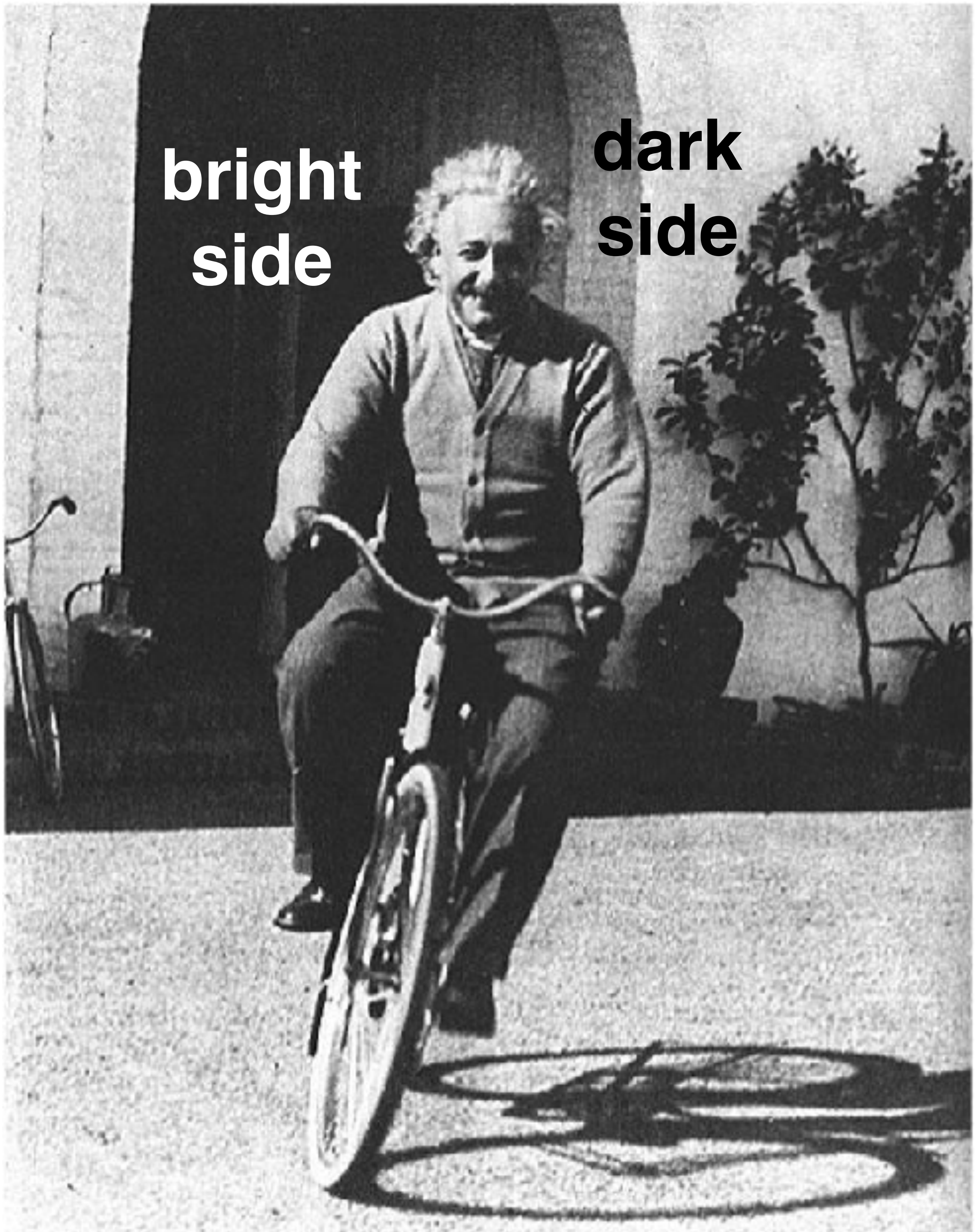


Pre-processing & spacetime symmetries

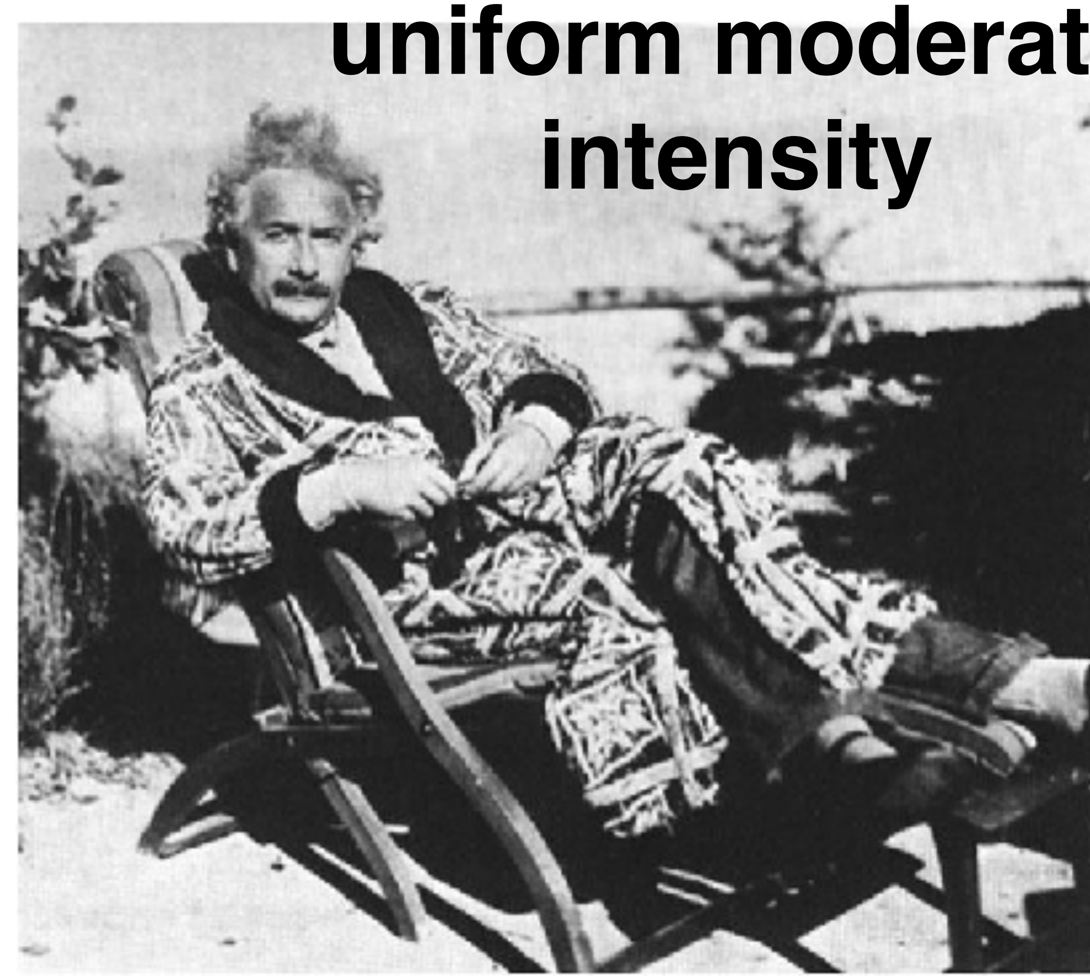
It is common to normalize each image so that **S** Intensity² = 1



Intuition via analogy *why normalization can hurt*



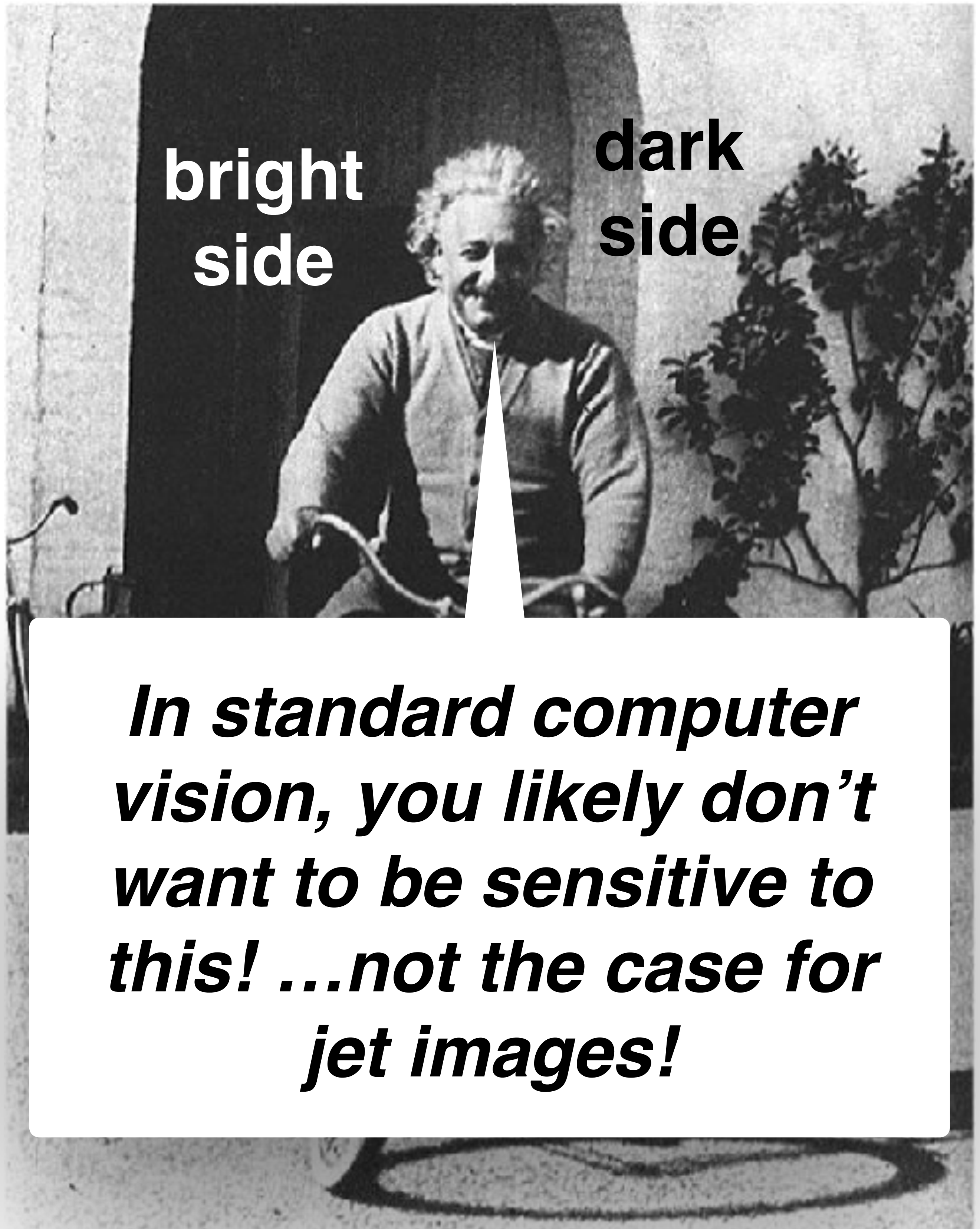
In both pictures, total intensity of Einstein's face is about the same.



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*



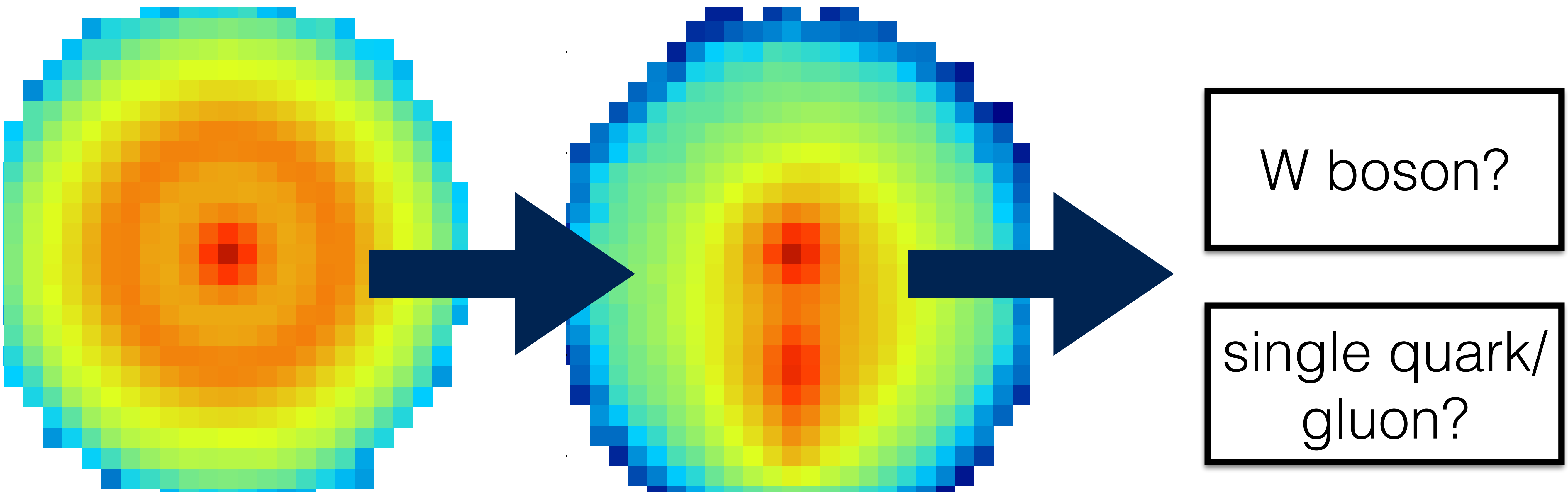
In standard computer vision, you likely don't want to be sensitive to this! …not the case for jet images!

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?

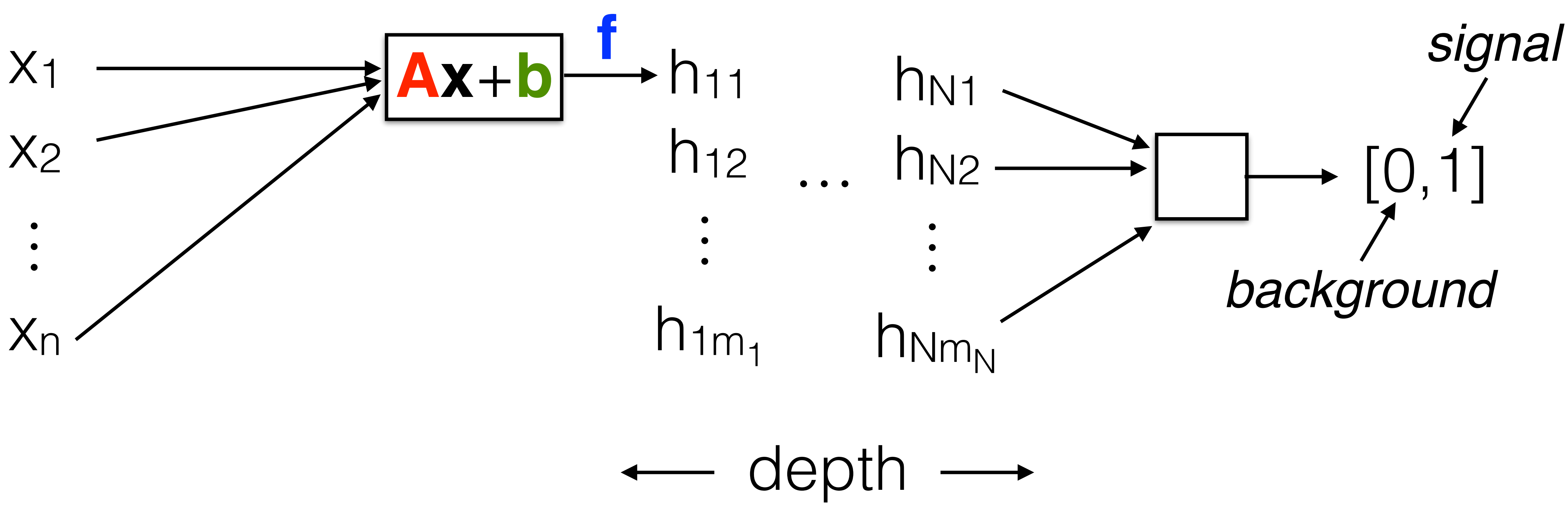


ultimate classification is achieved with modern machine learning using **all pixels as input**!

Modern Deep NN's for Classification

Neural Network: composition of functions **f**(**Ax**+**b**) for inputs **x** (features) matrix **A** (weights), bias **b**, non-linearity **f**.

N.B. I'm not mentioning biology - there may be a vague resemblance to parts of the brain, but that is not what modern NN's are about.



Modern Deep NN's for Classification

Neural Network: composition of functions **f**(**Ax**+**b**) for inputs **x** (features) matrix **A** (weights), bias **b**, non-linearity **f**.

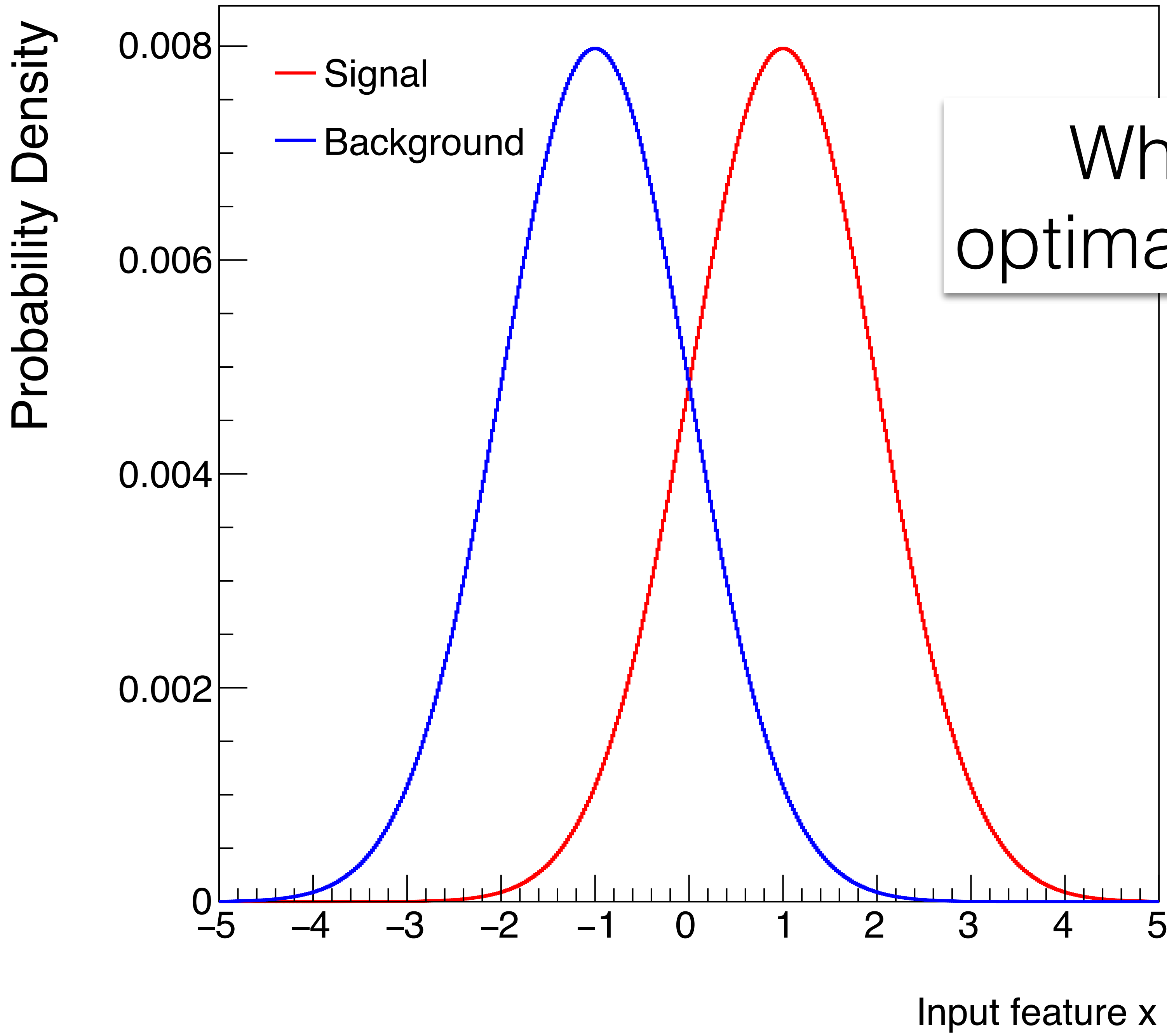
Fact: NN's can approximate "any" function.

Why useful?

For classification, there is an optimal function to learn: the likelihood ratio, LL(x) = $p_S(x) / p_B(x)$.

Getting into the machine's mind

Let's consider an important special case:
binary classification in 1D

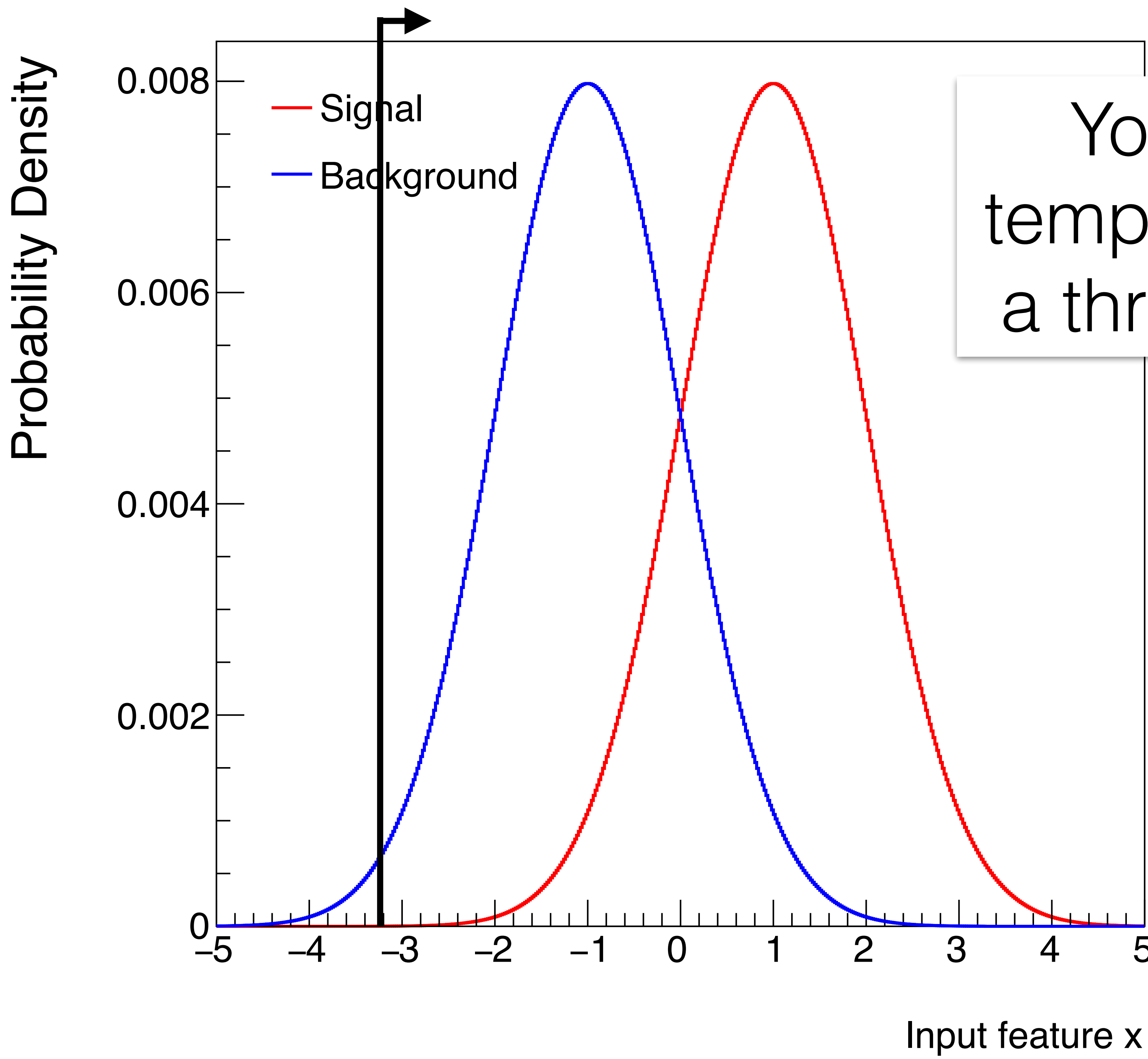


What is the optimal classifier?

could be e.g. the jet mass

Getting into the machine's mind

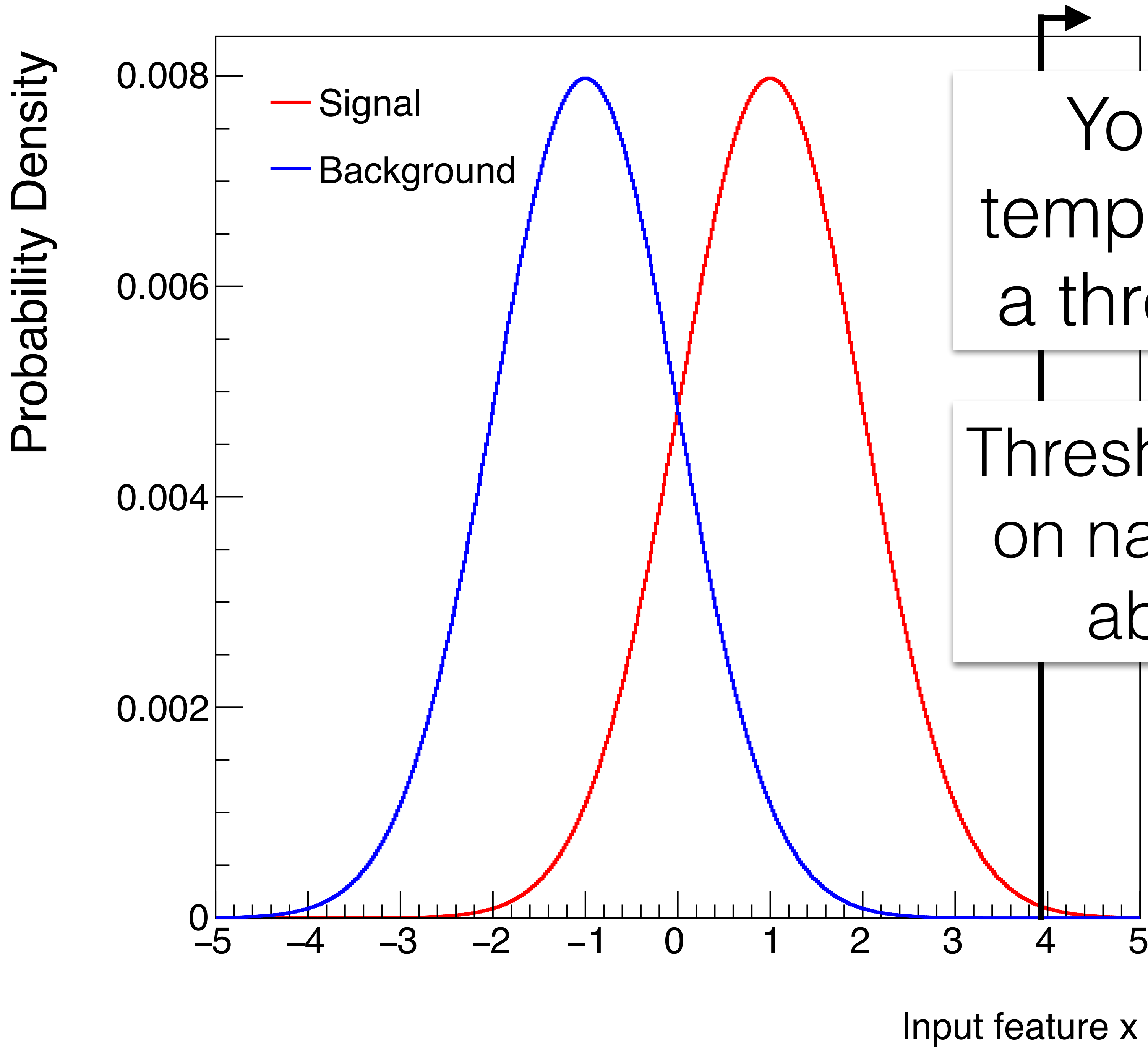
Let's consider an important special case:
binary classification in 1D



You may be tempted to place a threshold on x

Getting into the machine's mind

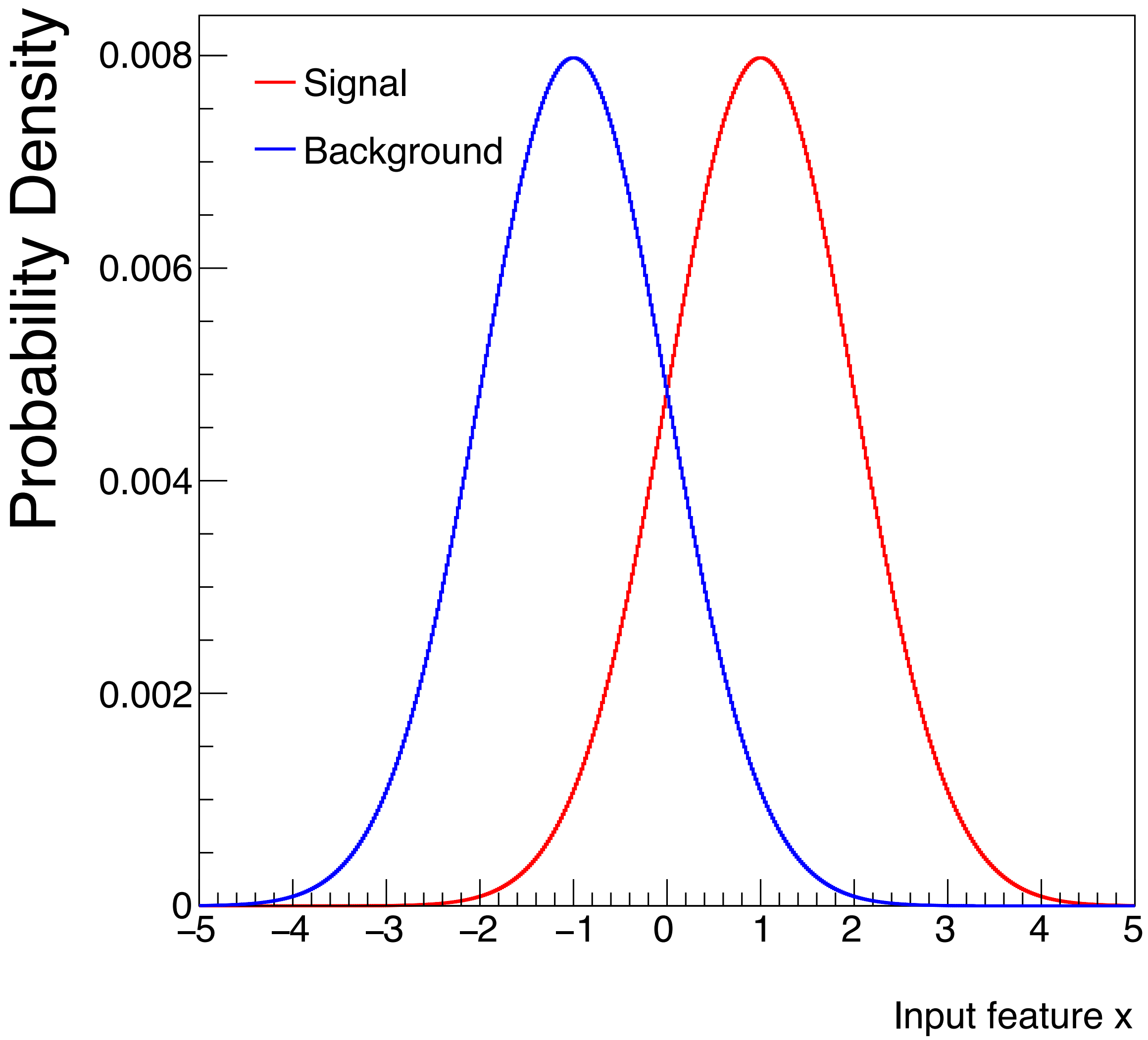
Let's consider an important special case:
binary classification in 1D



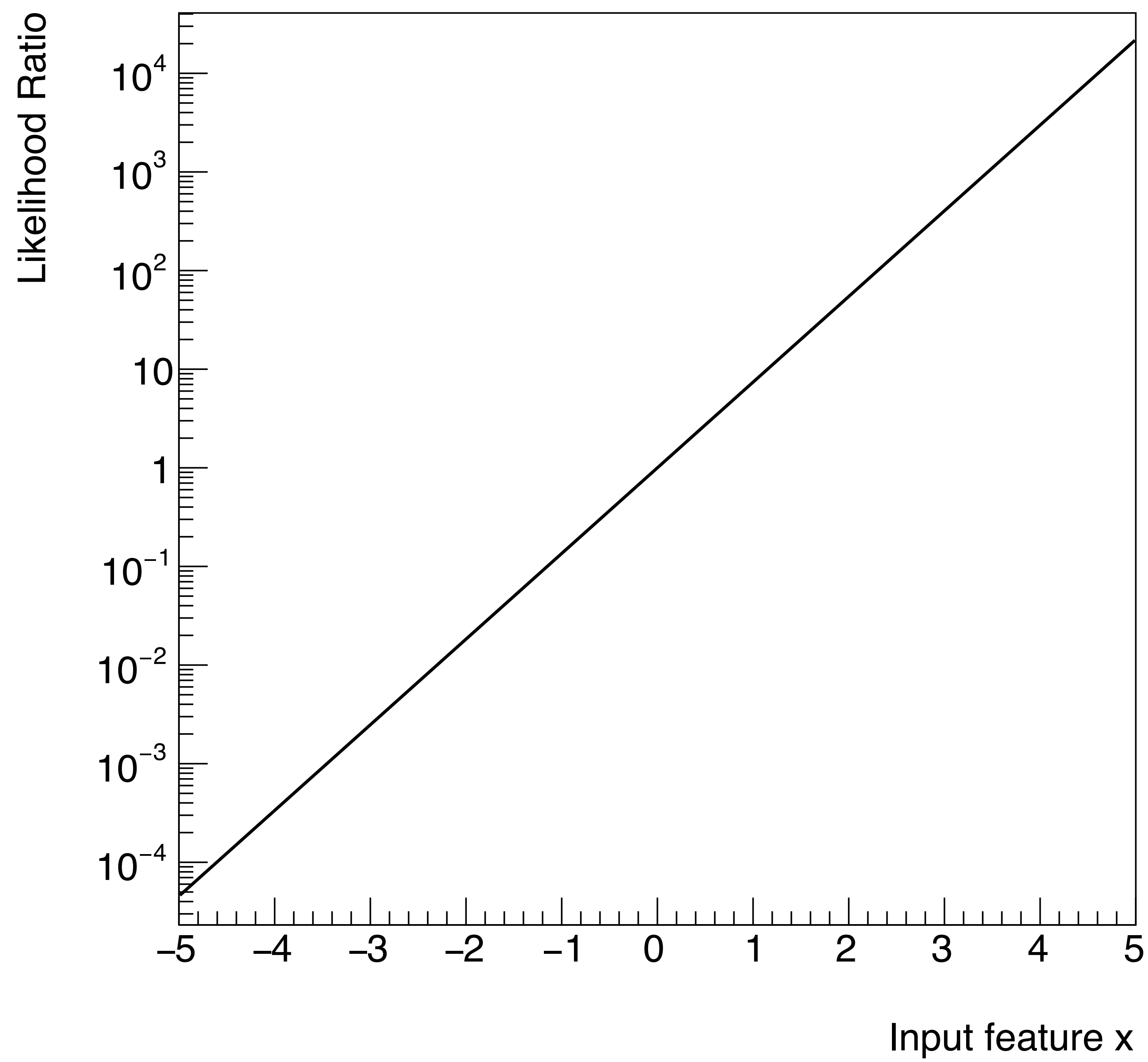
You may be tempted to place a threshold on x

Threshold depends on natural relative abundance

Getting into the machine's mind



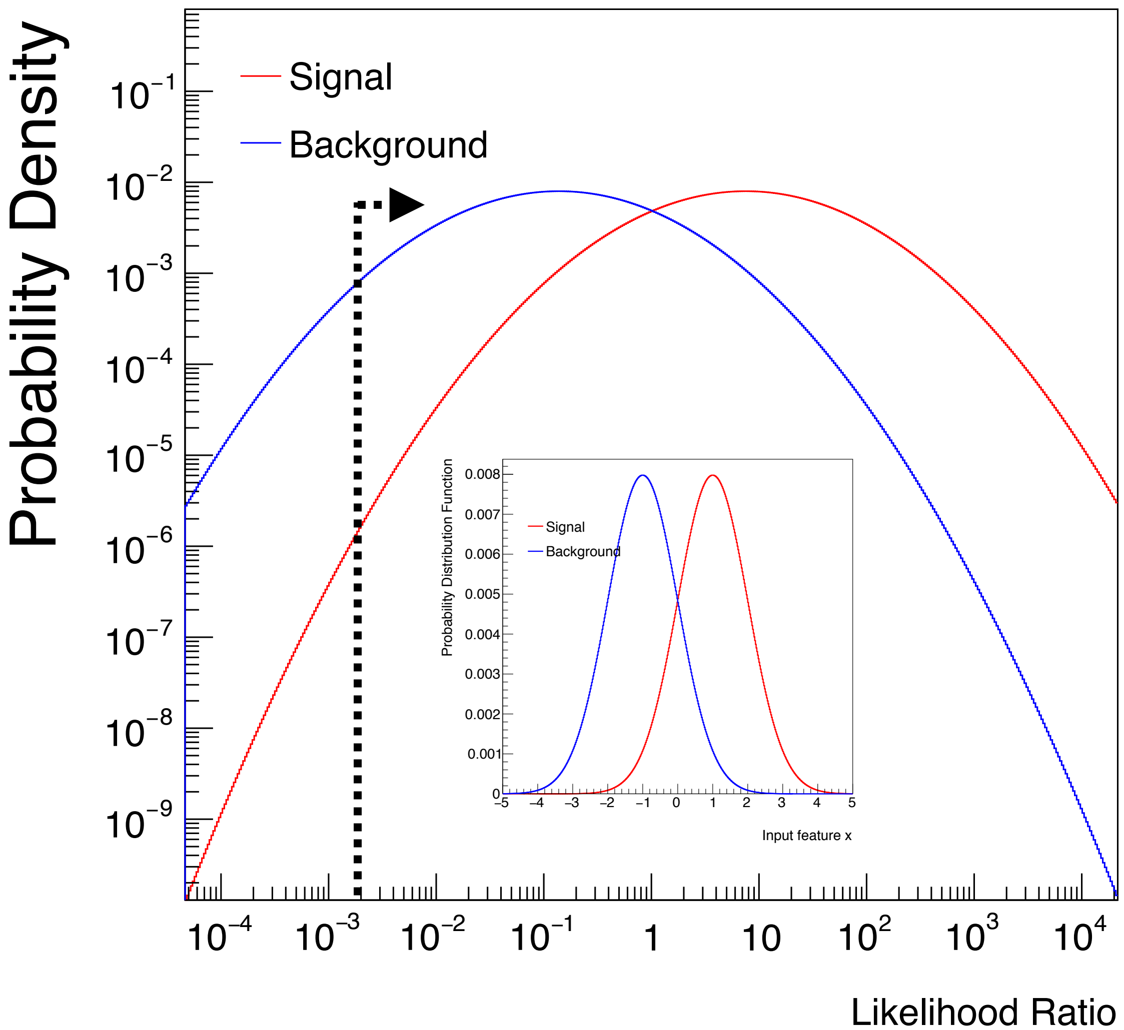
Is the simple threshold cut optimal?



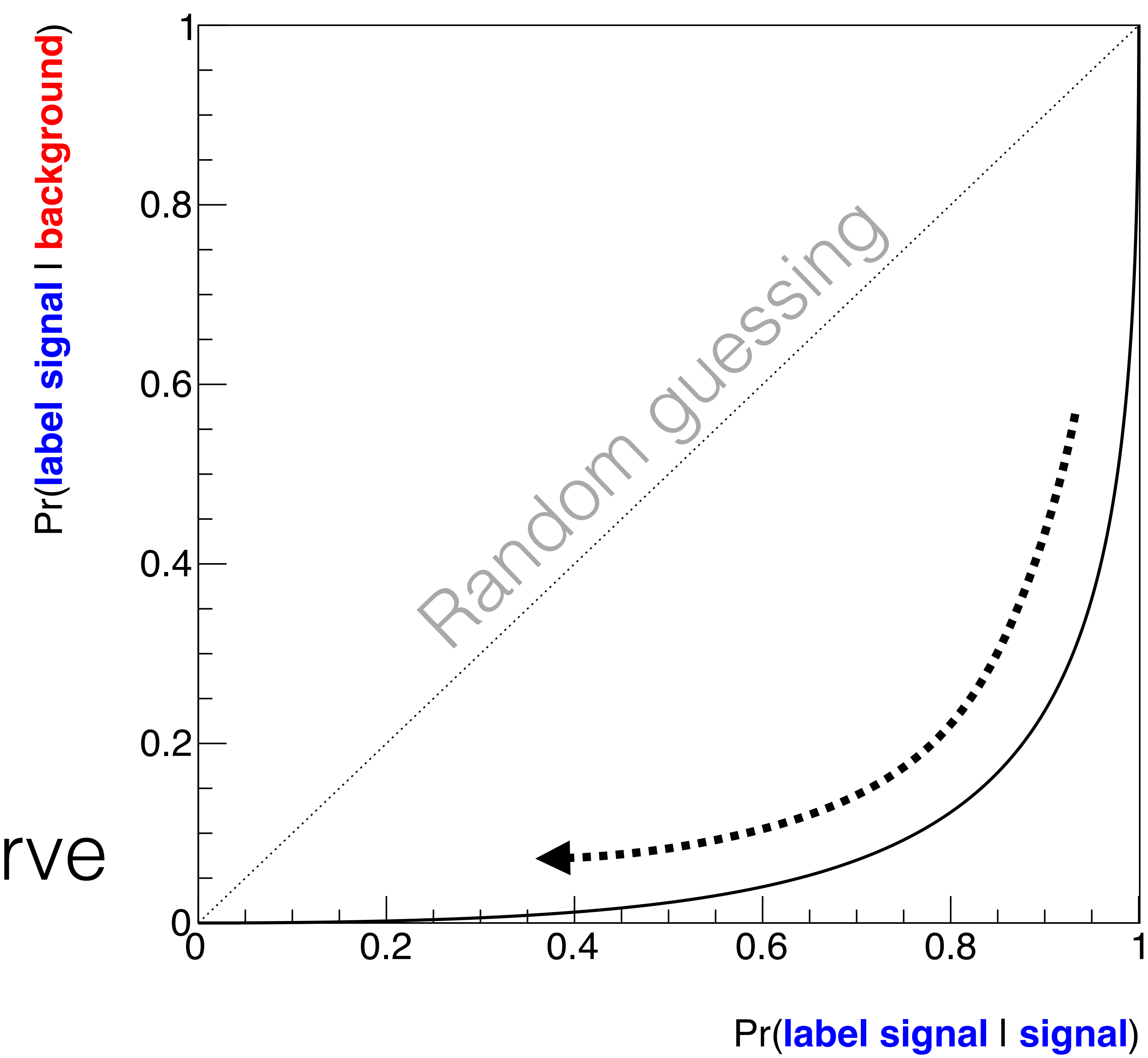
In this simple case, the log LL is proportional to x: **no need for non-linearities**!

Threshold cut is optimal

Getting into the machine's mind



The optimal procedure is a threshold on the LL

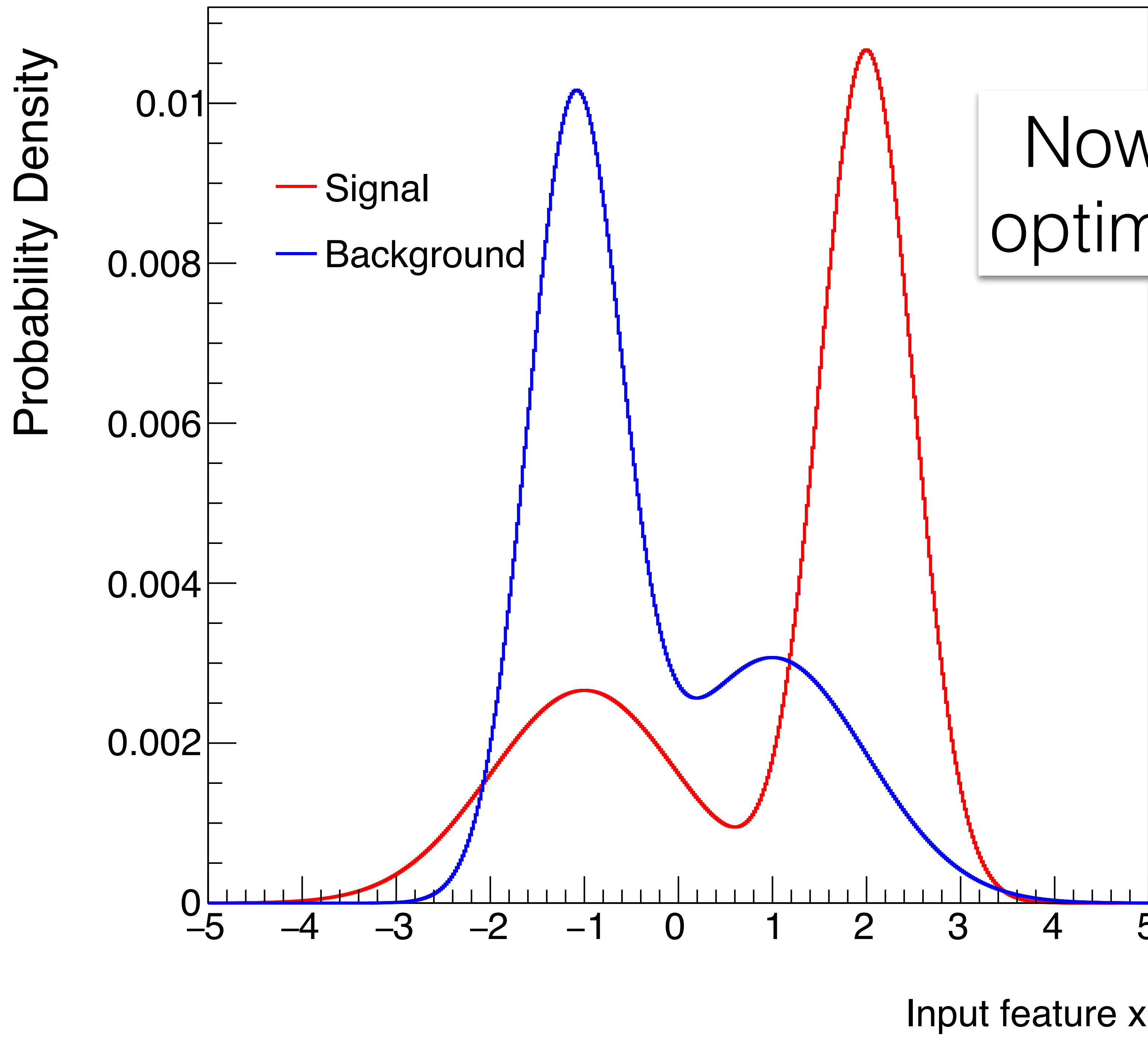


"Receiver Operating Characteristic" (**ROC**) Curve

Getting into the machine's mind

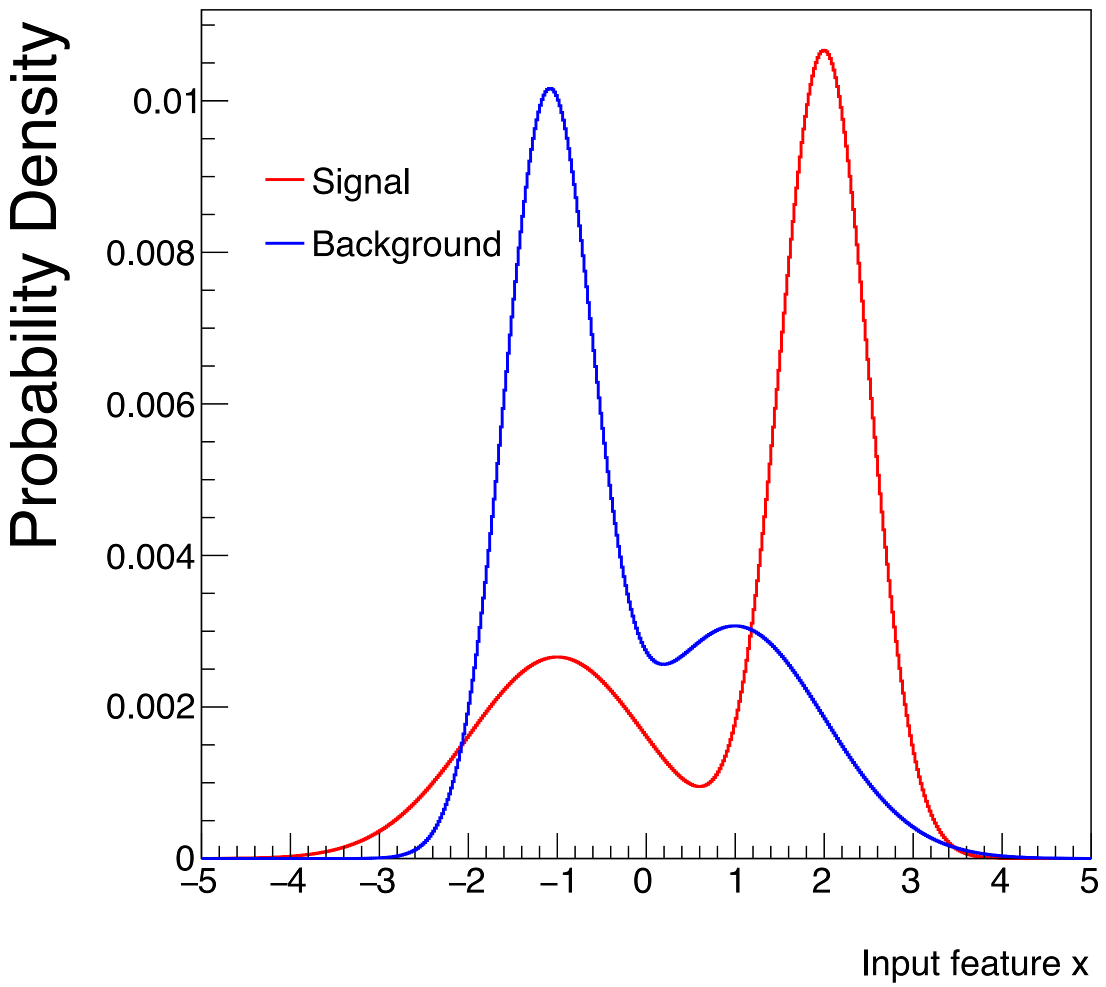
What if the distribution of x is complicated?

Real life is complicated!

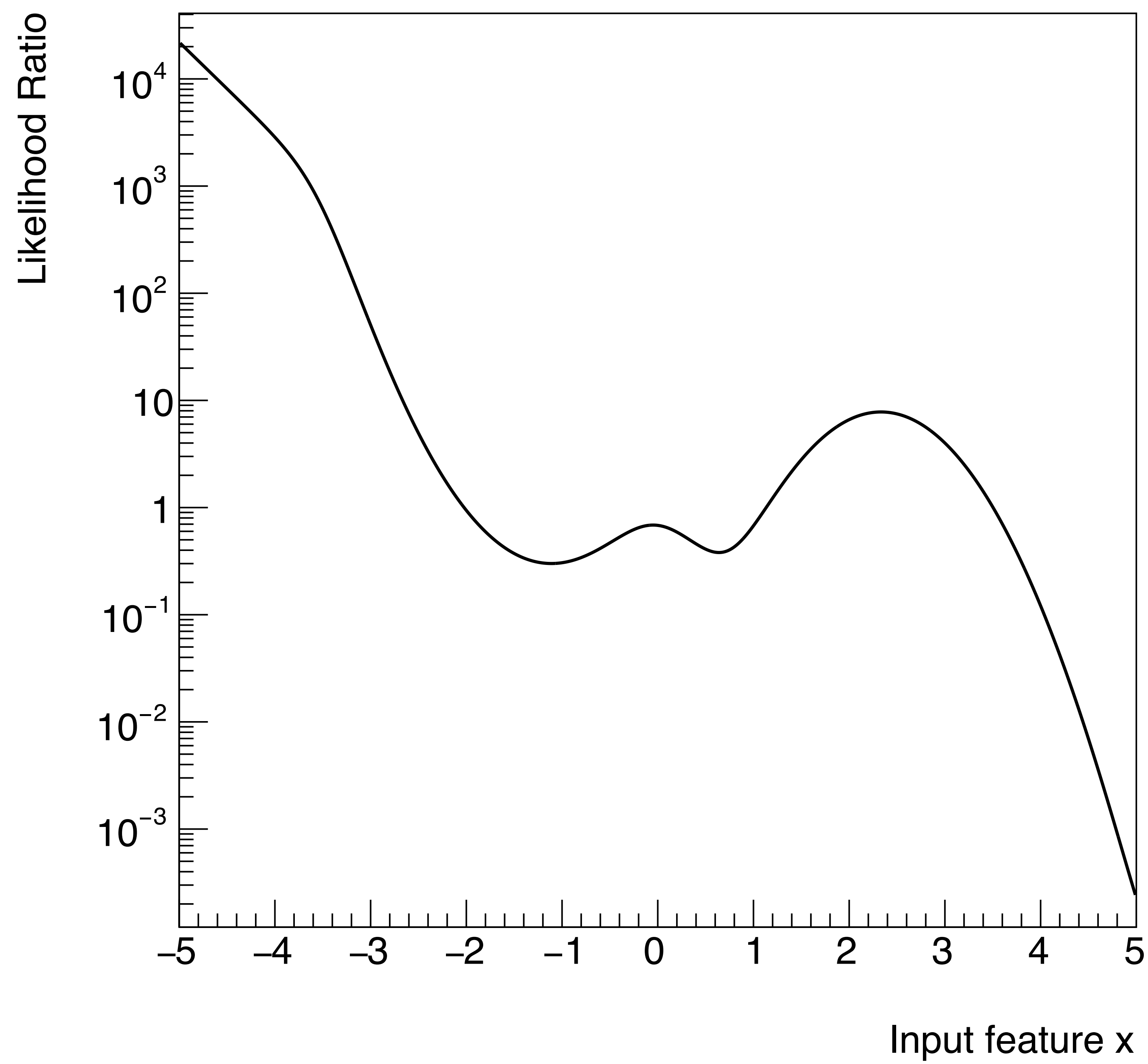


Now what is the optimal classifier?

Getting into the machine's mind

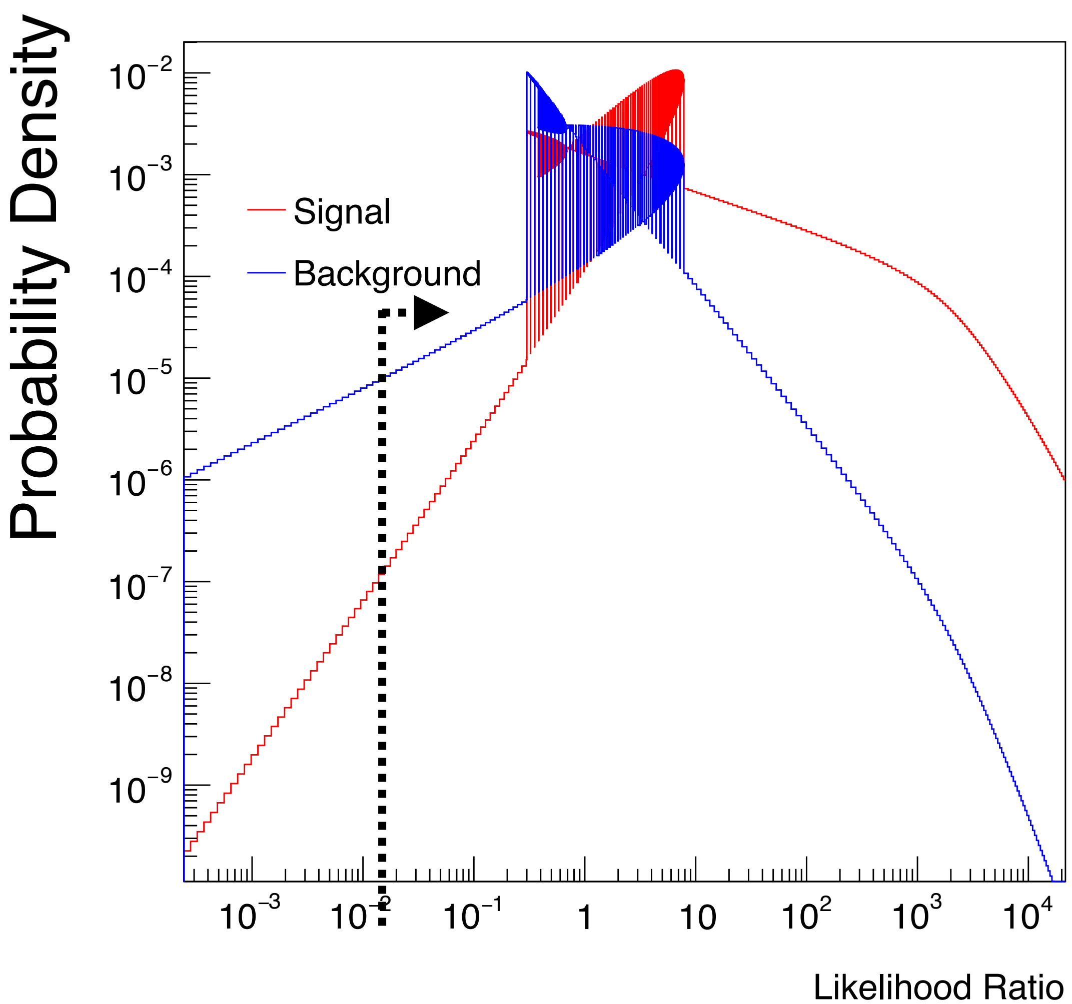


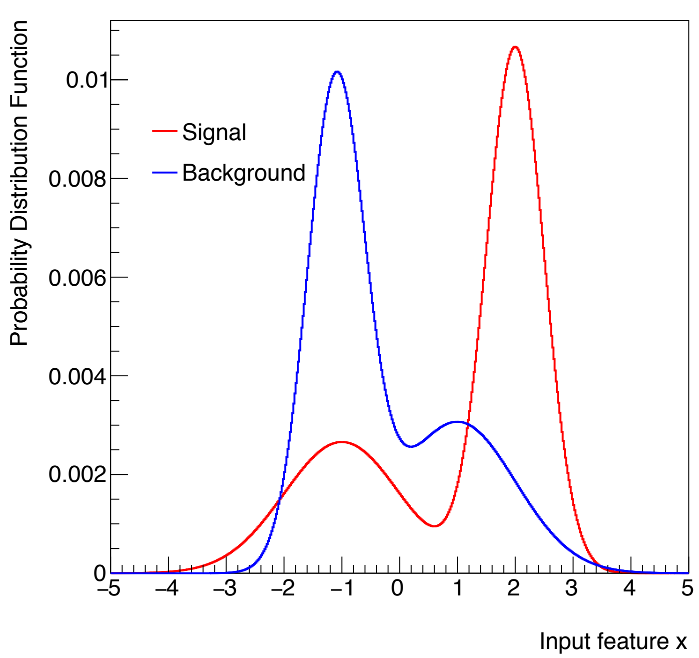
In this case, LL is highly non-linear function of x

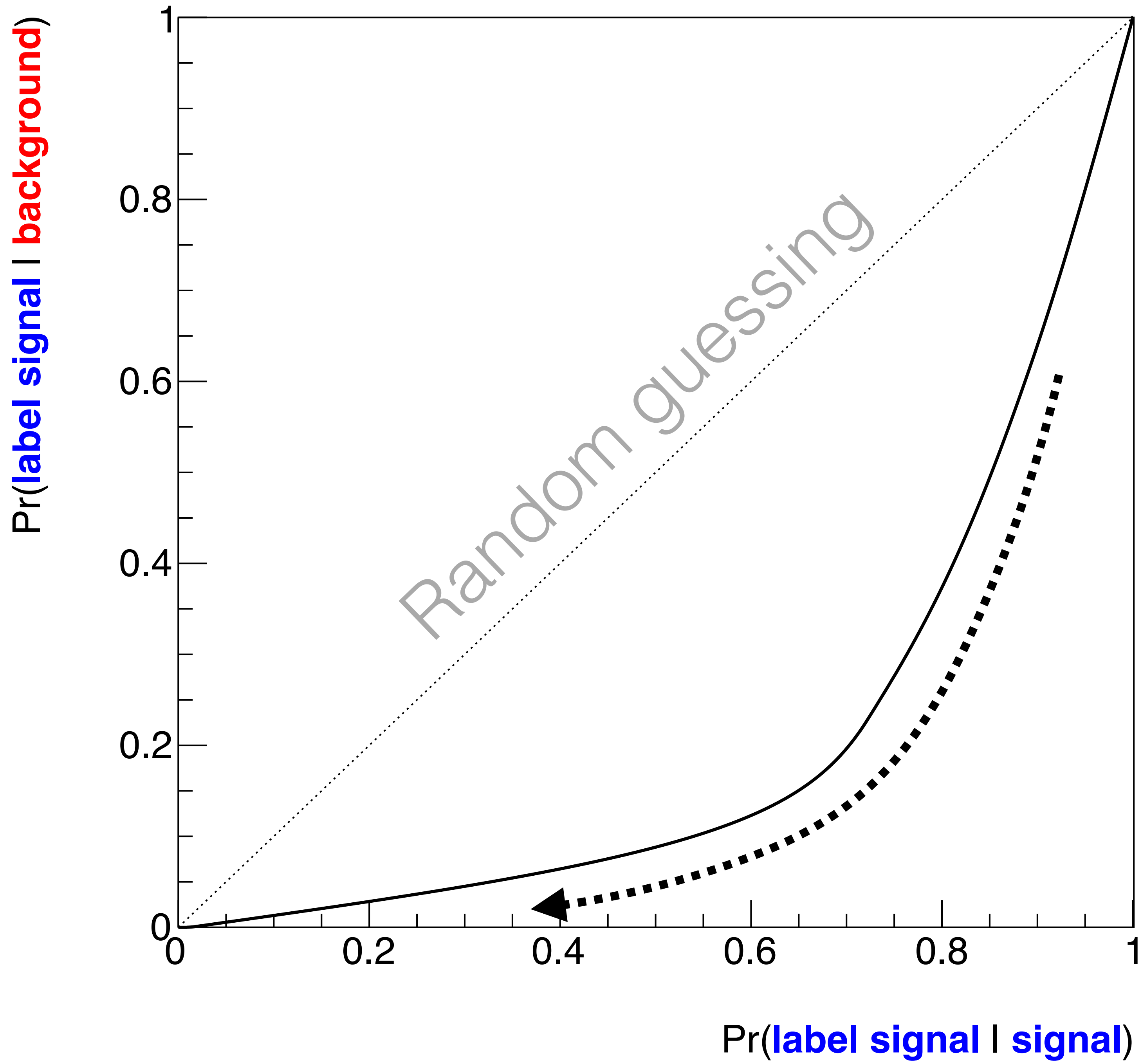


A threshold on x would be sub-optimal

Getting into the machine's mind





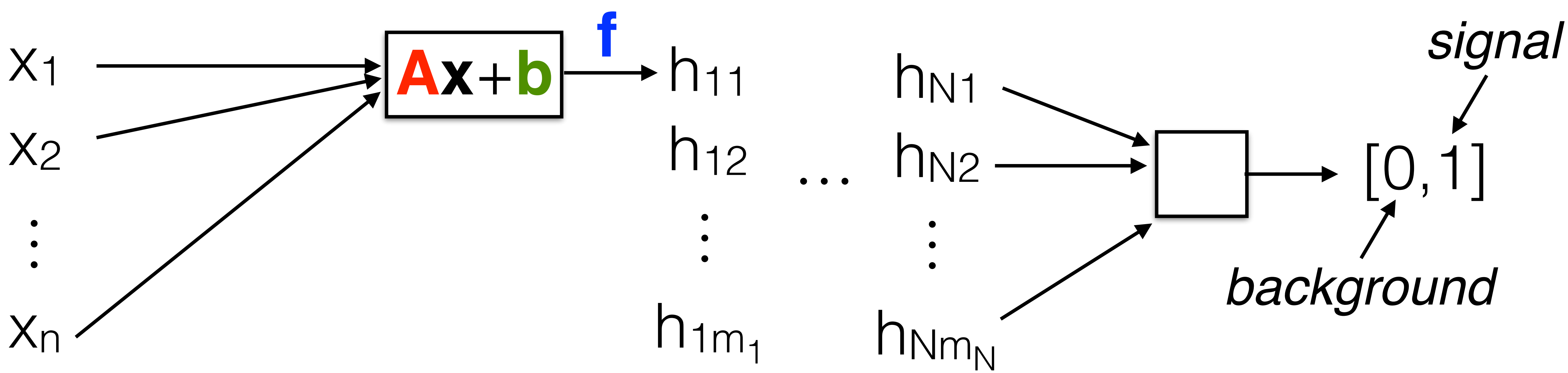


The curse of dimensionality

In principle, you can do the same thing in N > 1 dimensions. However, it very quickly gets out of hand!

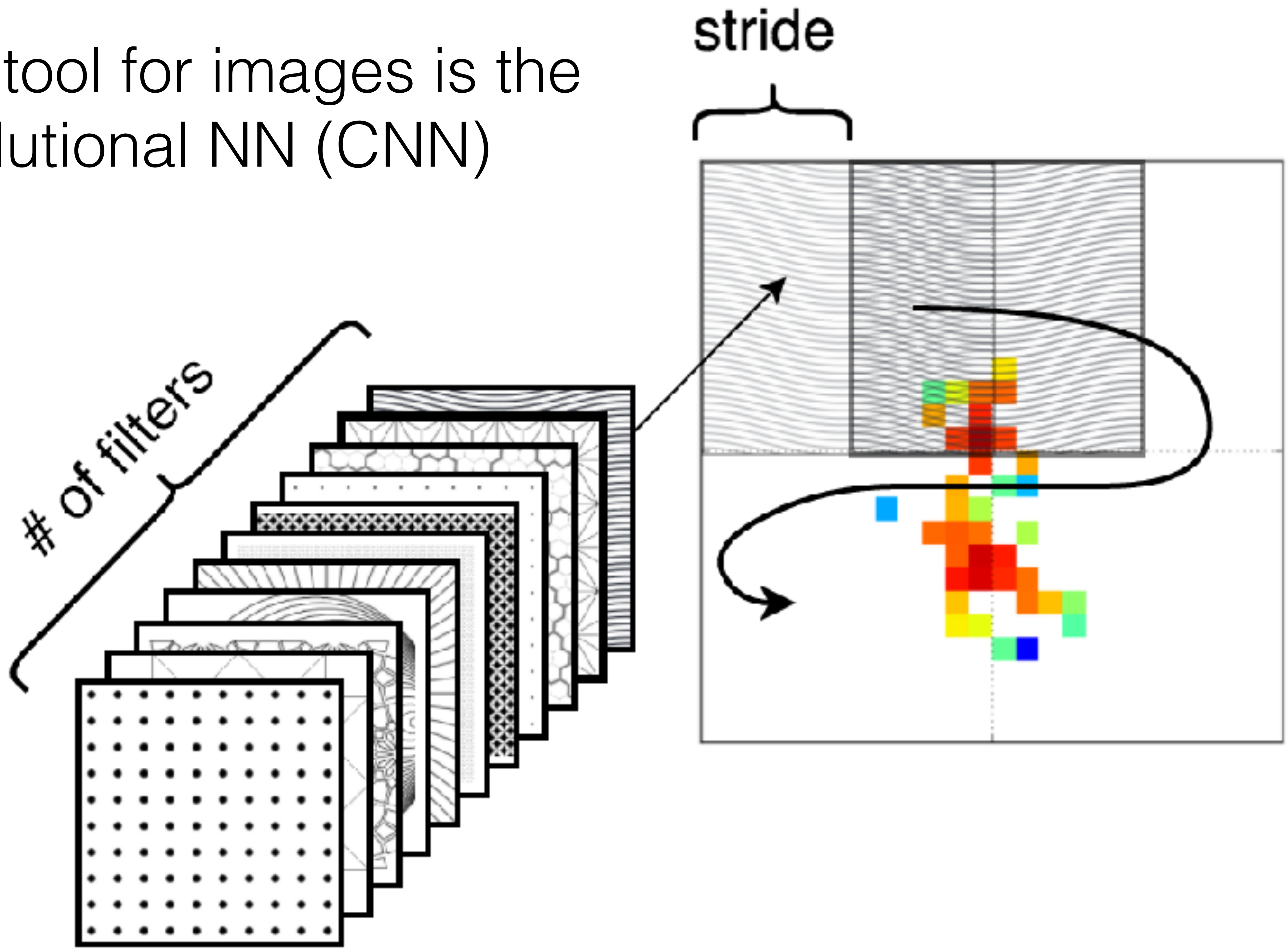
That is where NN's come in.

Image ~ 1000 dimensional



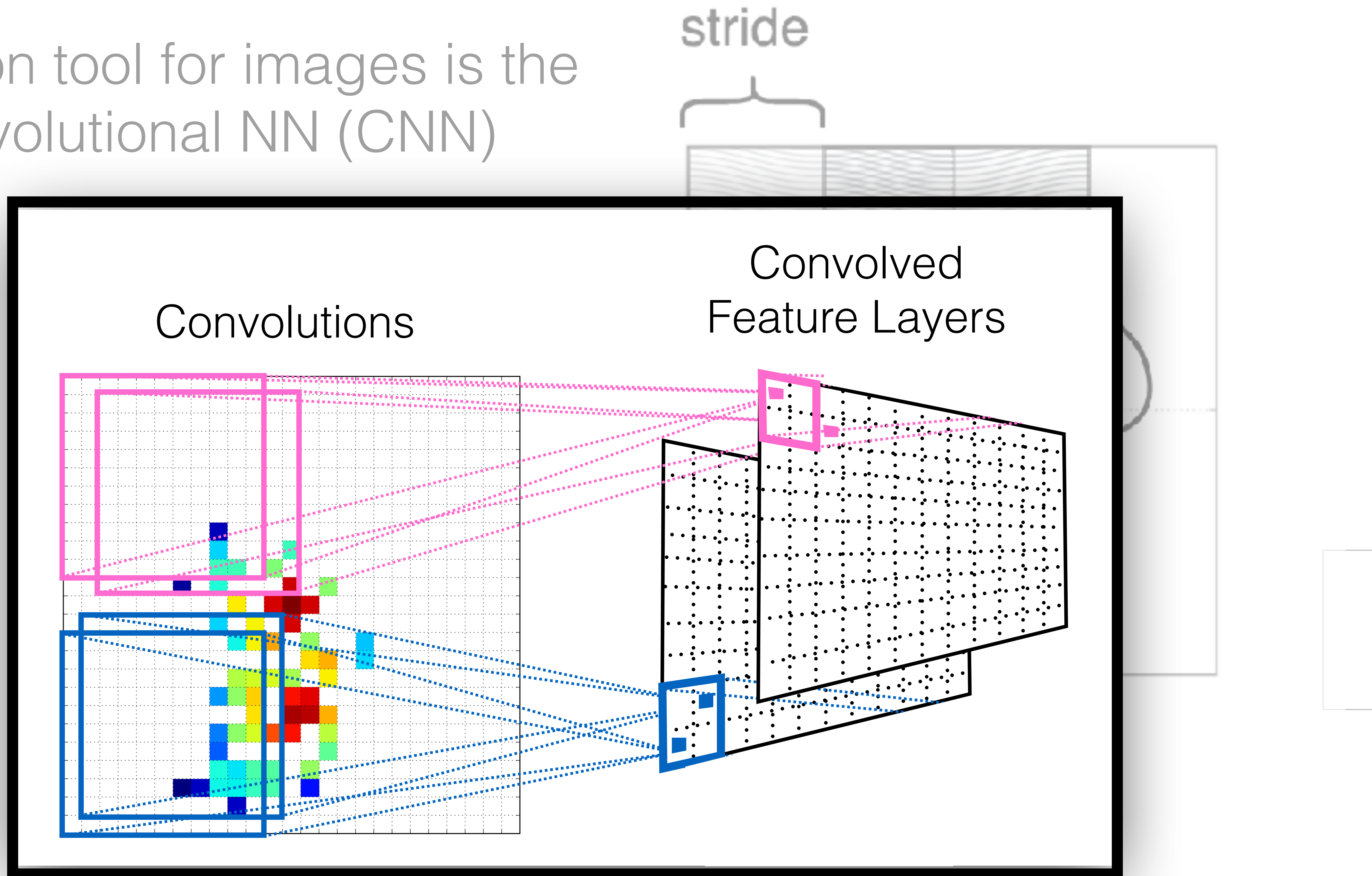
Let's see how we can use DNN's for jet image classification

Common tool for images is the convolutional NN (CNN)



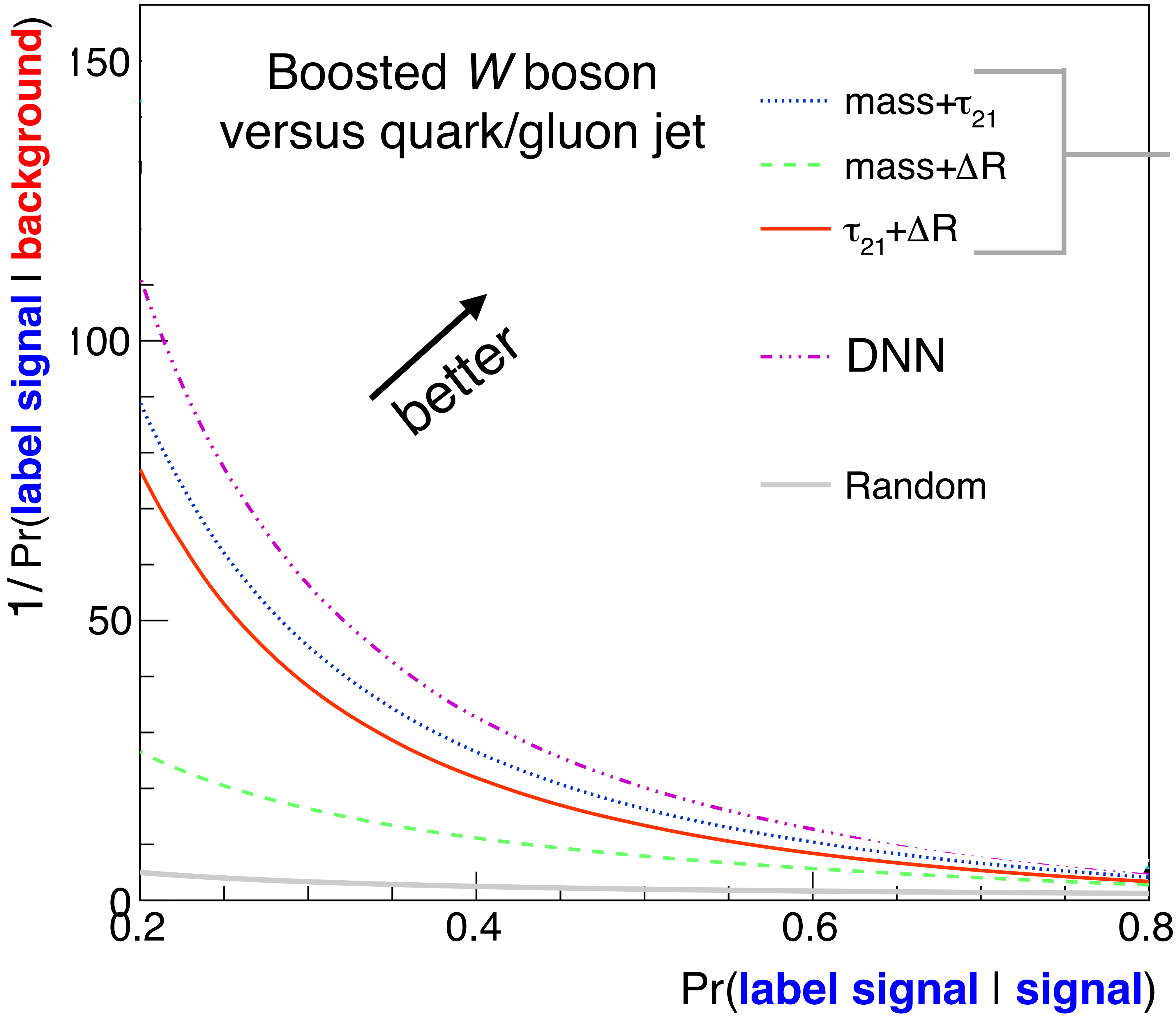
The filter is like the **A**, only the dimensionality is now the filter size (<< n) and not the image size (n).

Common tool for images is the convolutional NN (CNN)



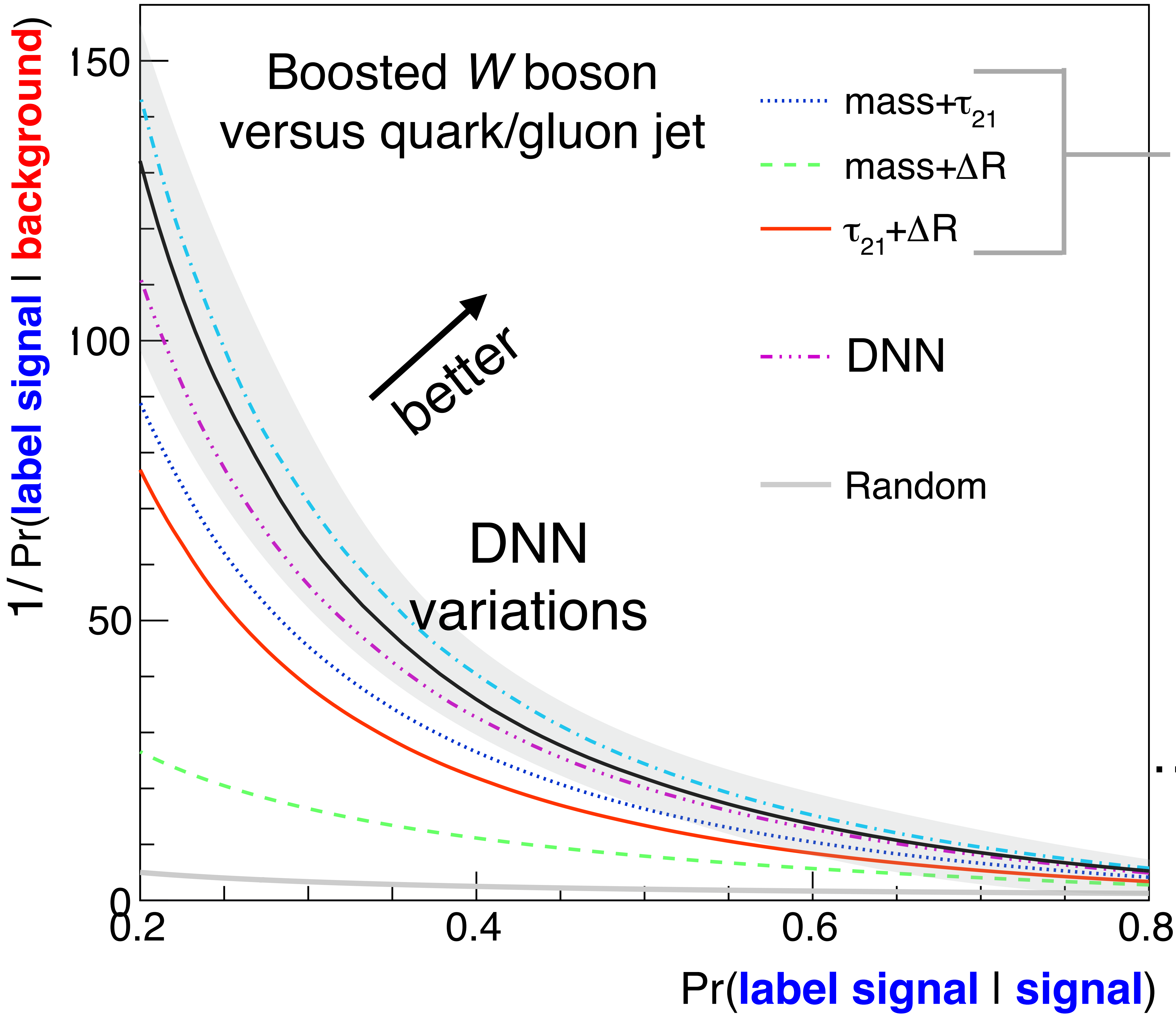
The filter is like the **A**, only the dimensionality is now the filter size (<< n) and not the image size (n).

Modern Deep NN's for Classification



mass, t21, DR are all simple functions of the image

Modern Deep NN's for Classification

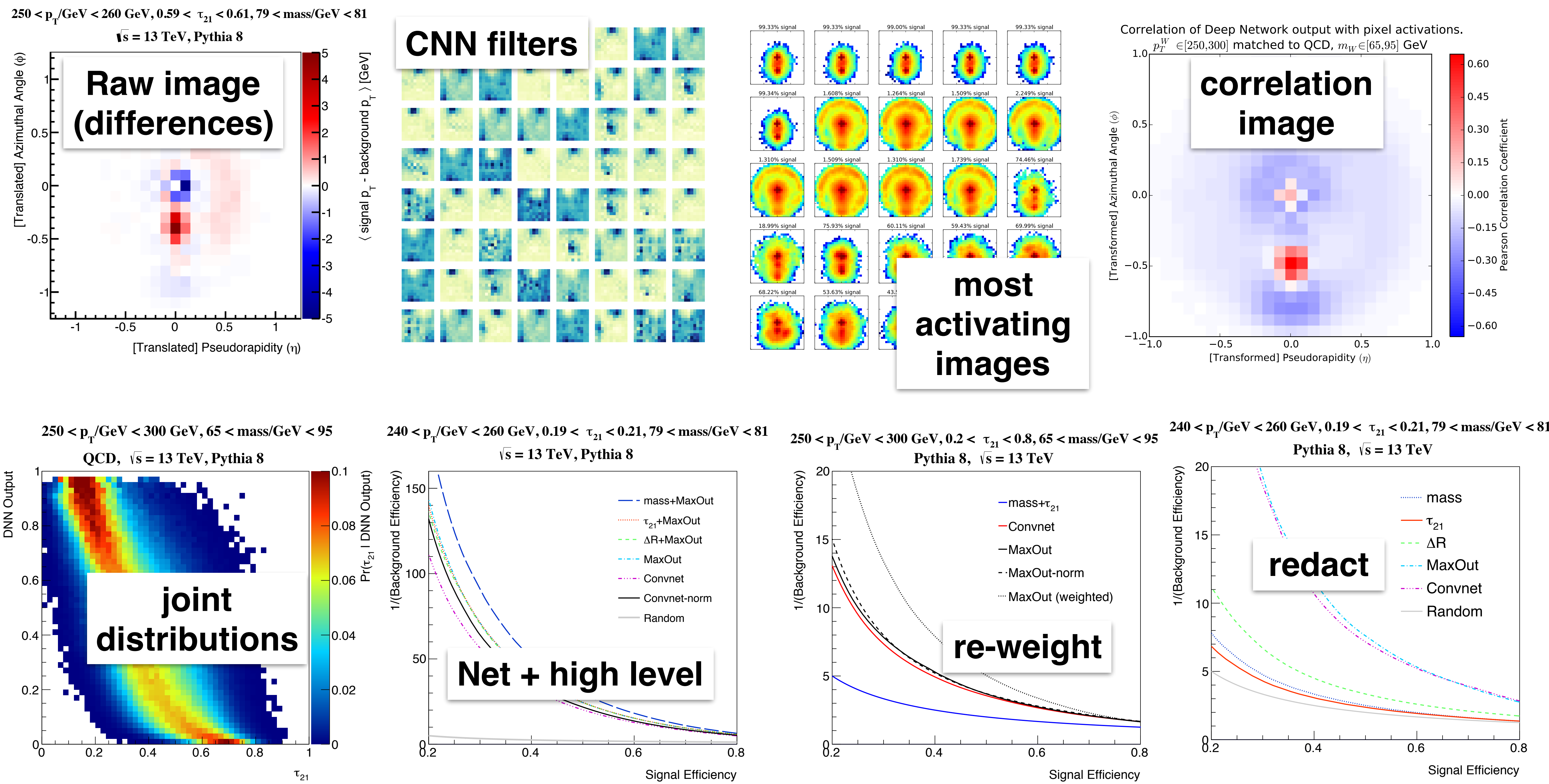


mass, t_{21}, DR are all simple functions of the image

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

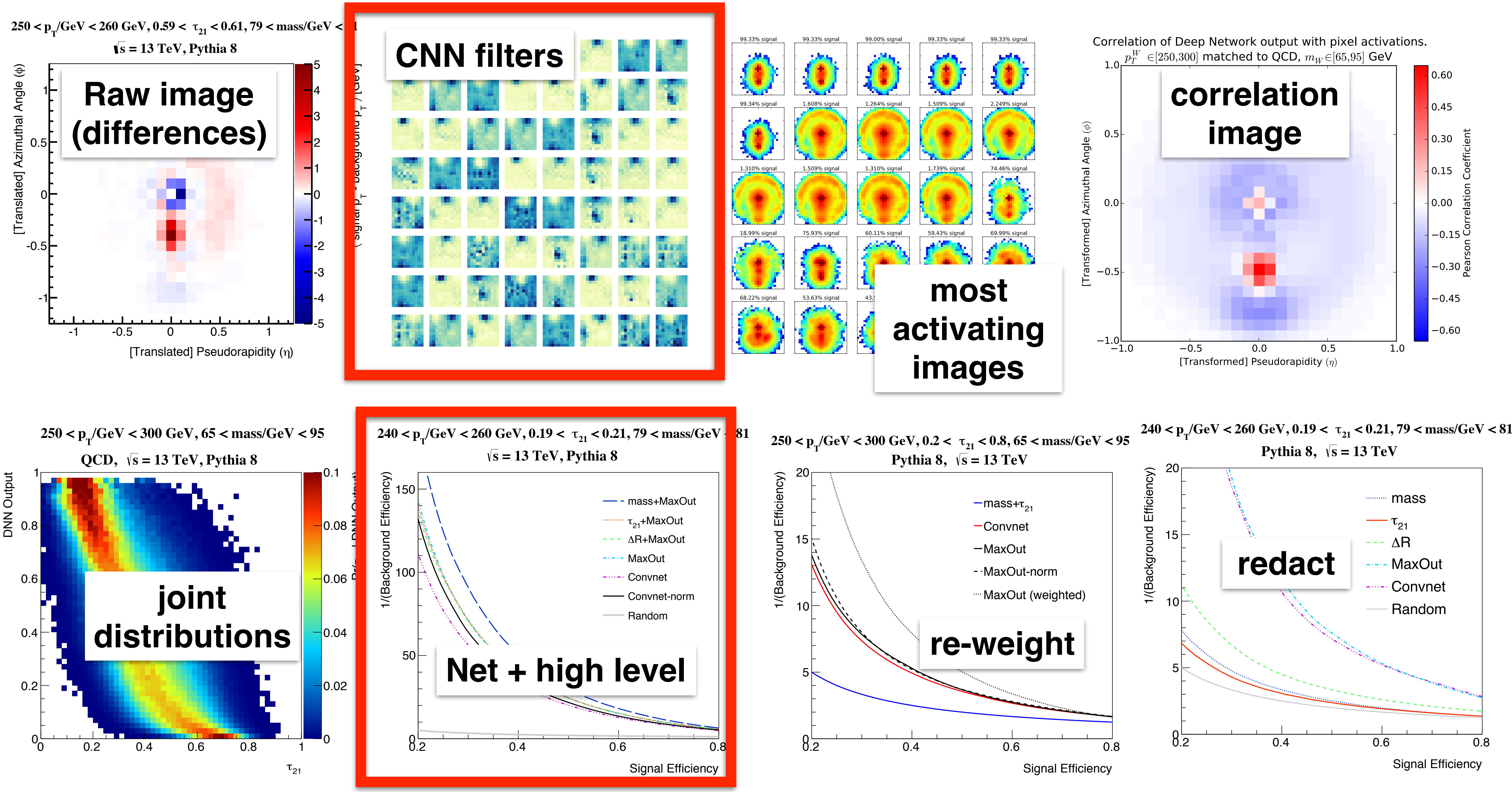
Learning about Learning

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



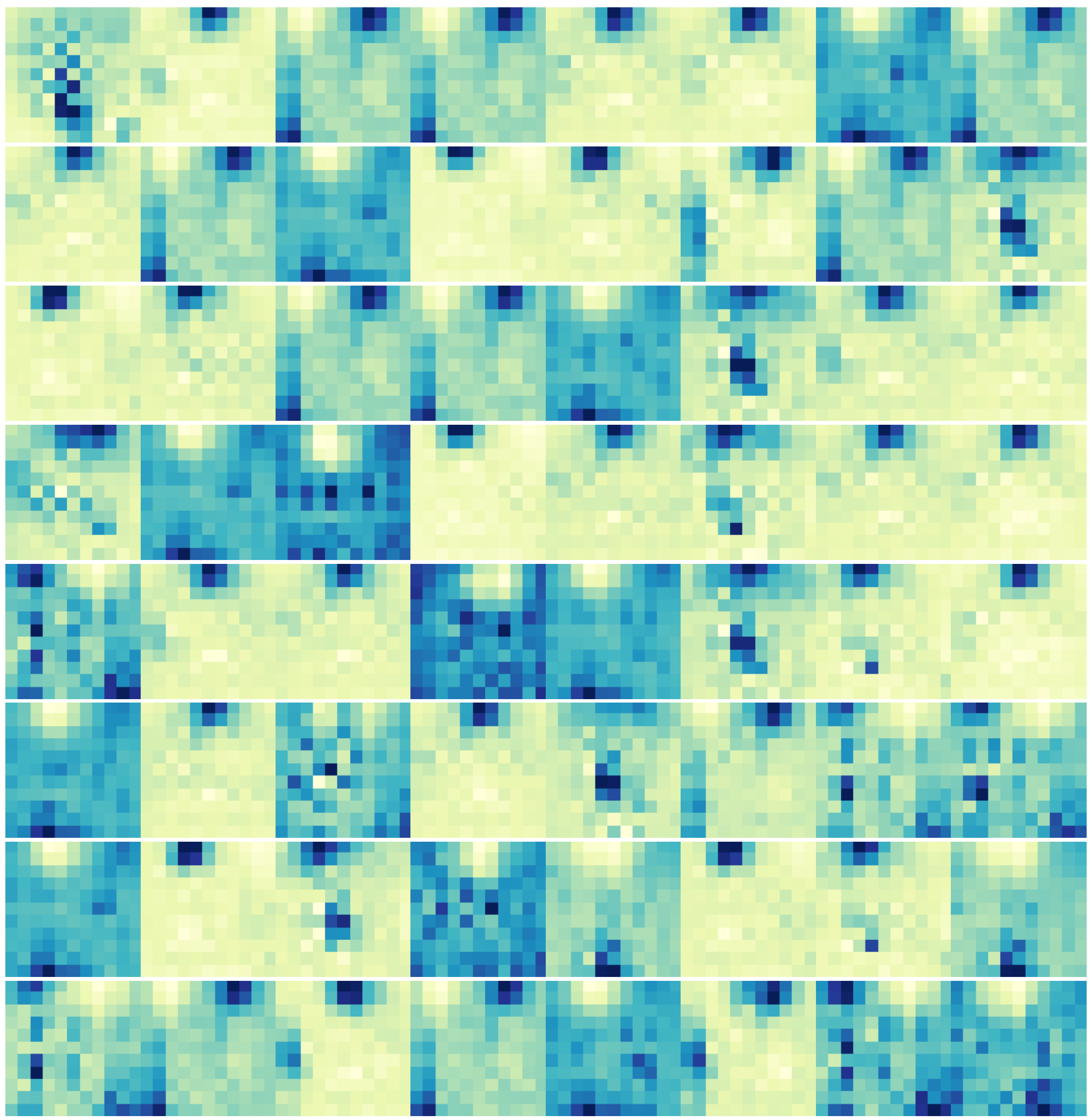
Learning about Learning

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



Convolutional Filters

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



Jet Images

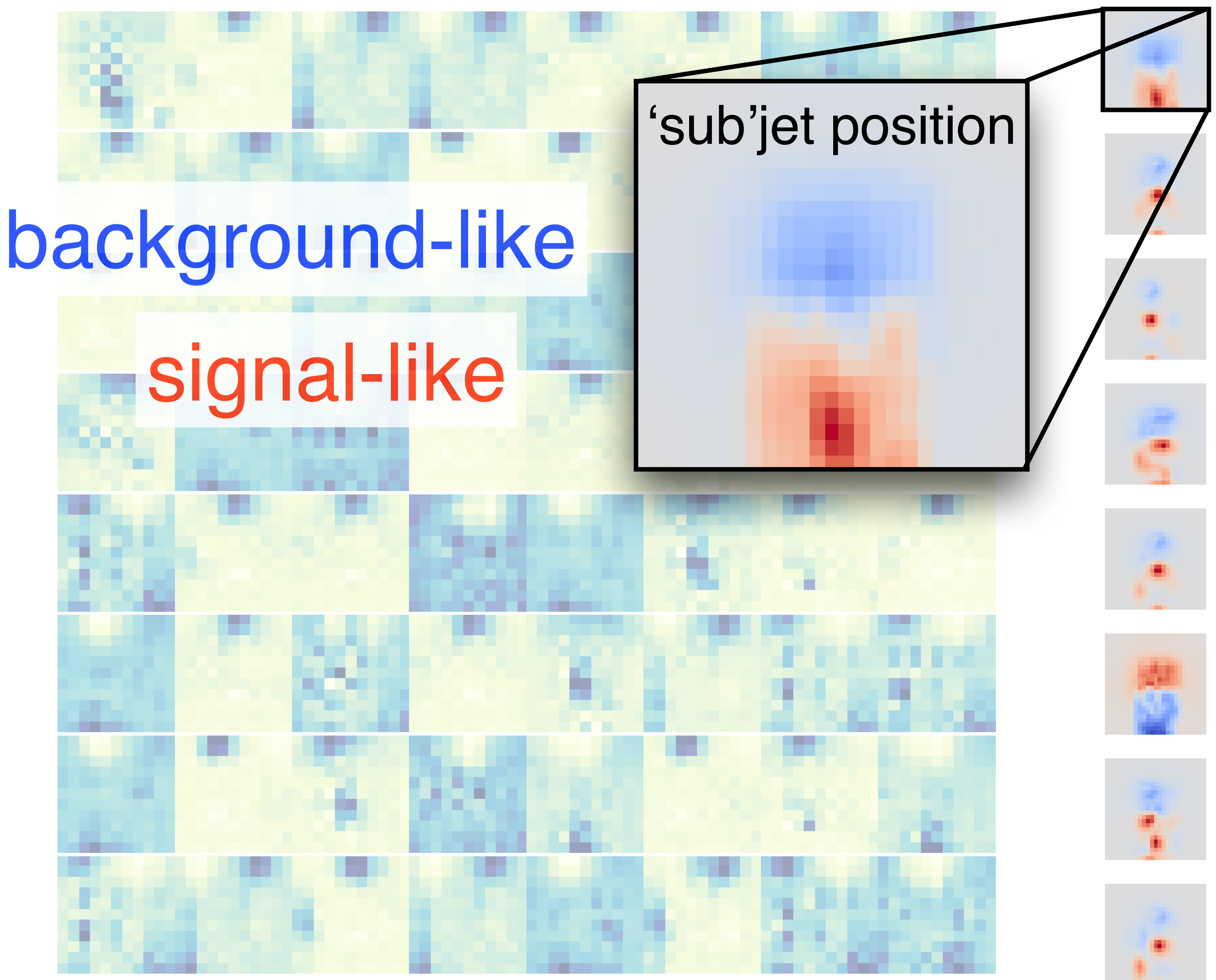
learned edge detection

"Natural" Images

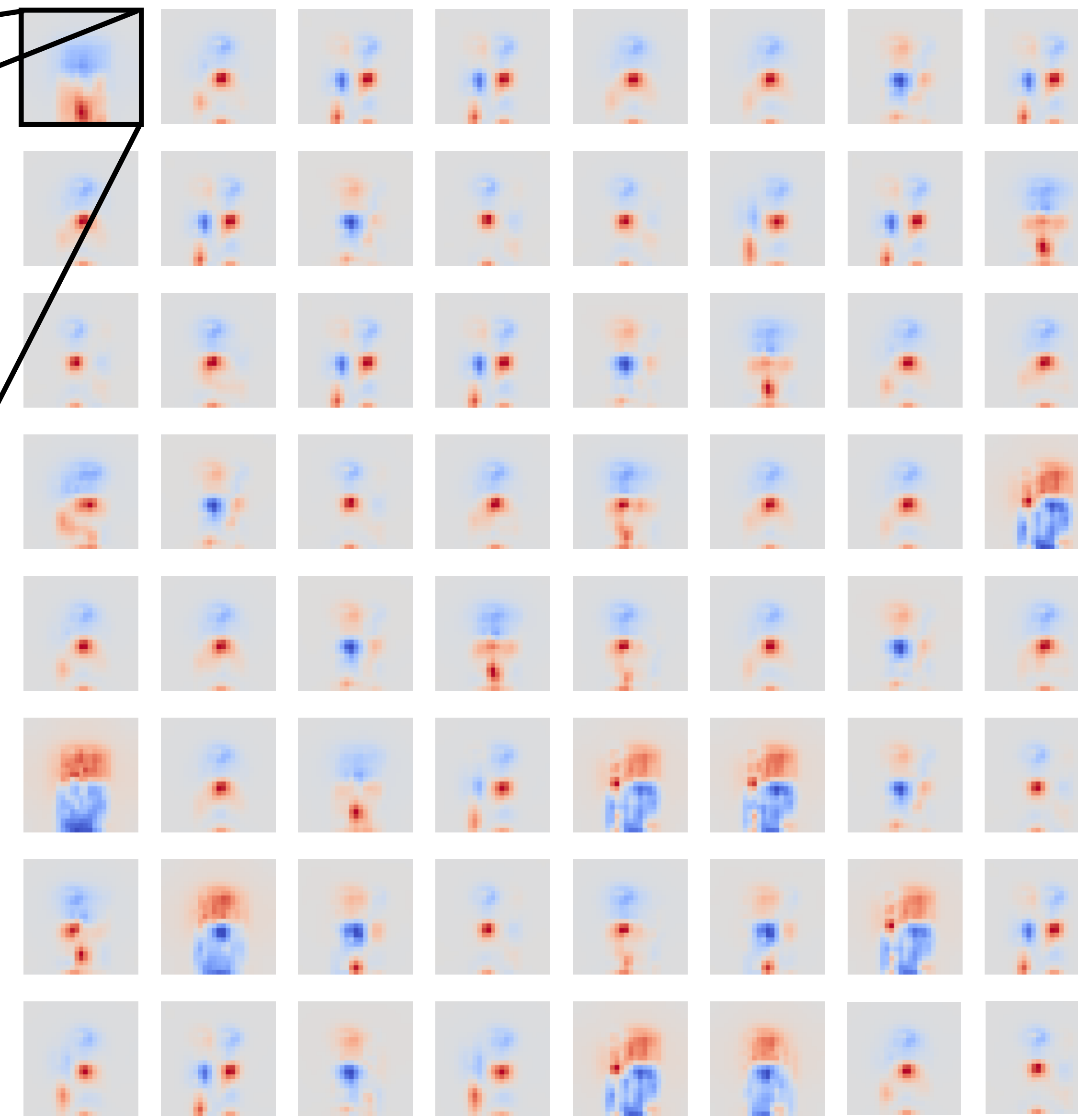
L. de Oliveira, M. Kagan, L. Mackey, **BN**, and A. Schwartzman, JHEP 07 (2016) 069 A. Krizhevsky et al. DNN for ImageNet

Convolutional Filters

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



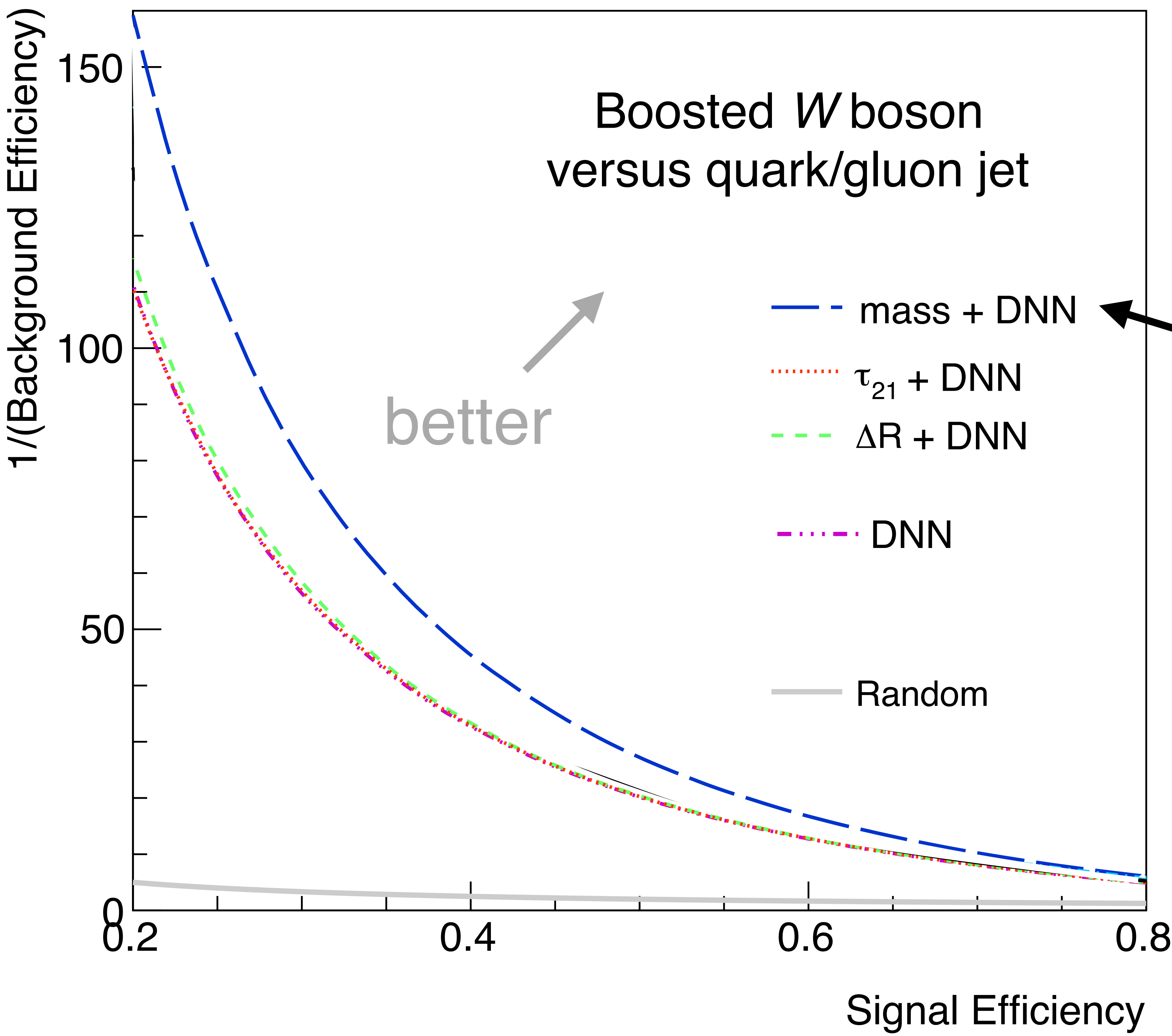
Jet Images Layer 1 Filters



Filters convolved with signal - background

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

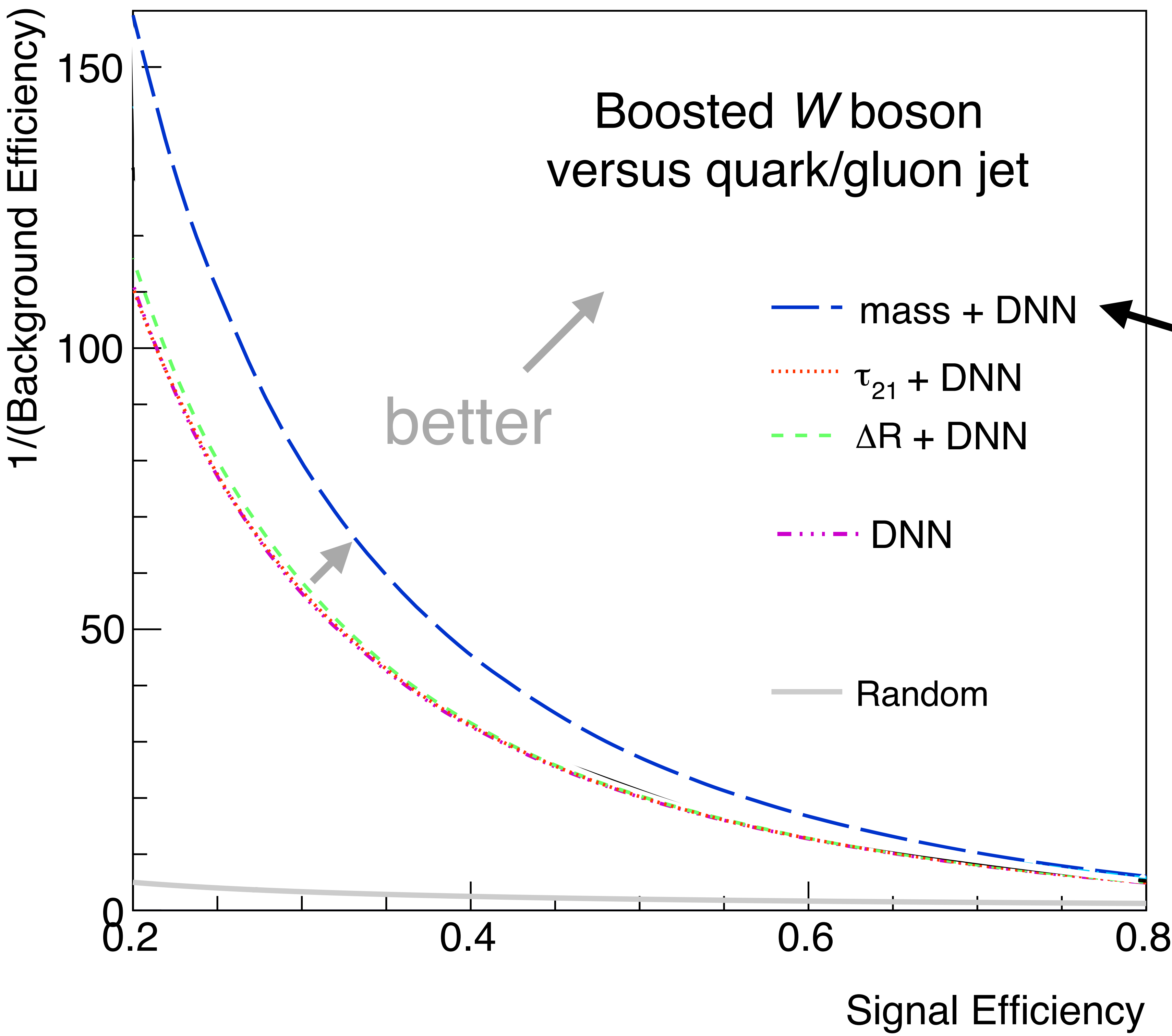
Is there more to learn that we know about?



Idea: explicitly combine NN with a known feature

Has learned image mass?

Is there more to learn that we know about?

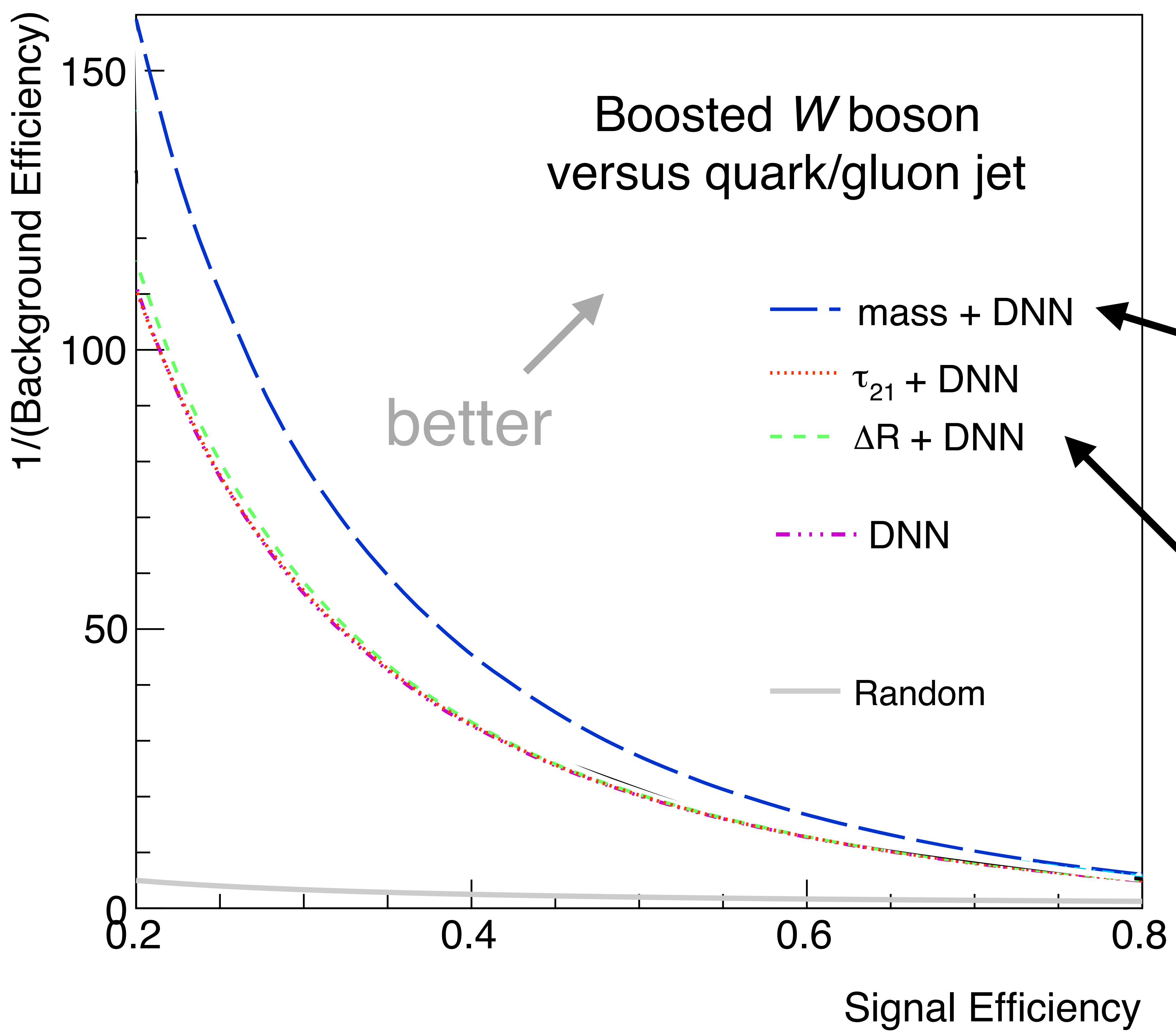


Idea: explicitly combine NN with a known feature

Has learned image mass?

not fully!

Is there more to learn that we know about?



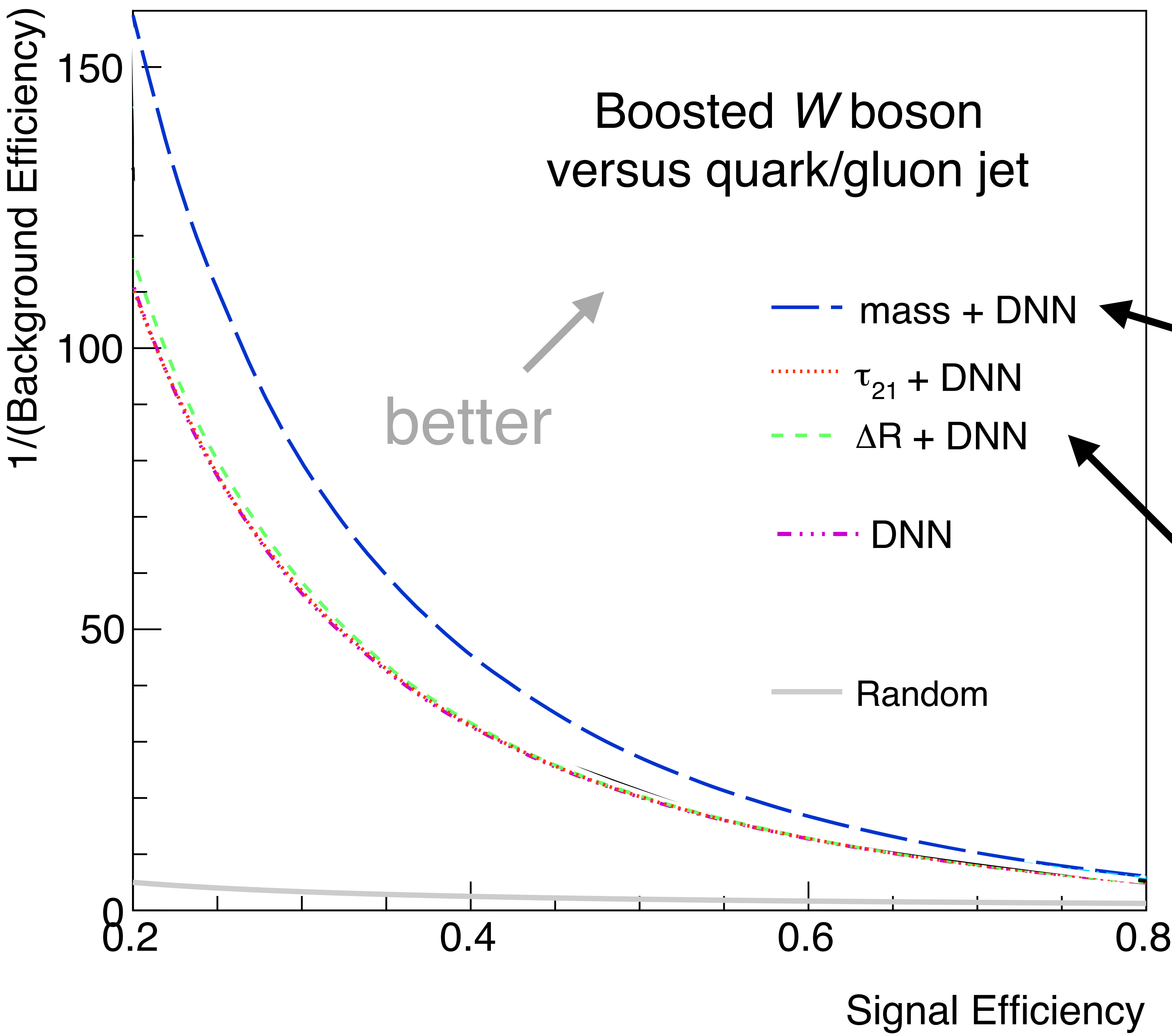
Idea: explicitly combine NN with a known feature

Has learned image mass?

not fully!

What about the distance between subjets?

Is there more to learn that we know about?



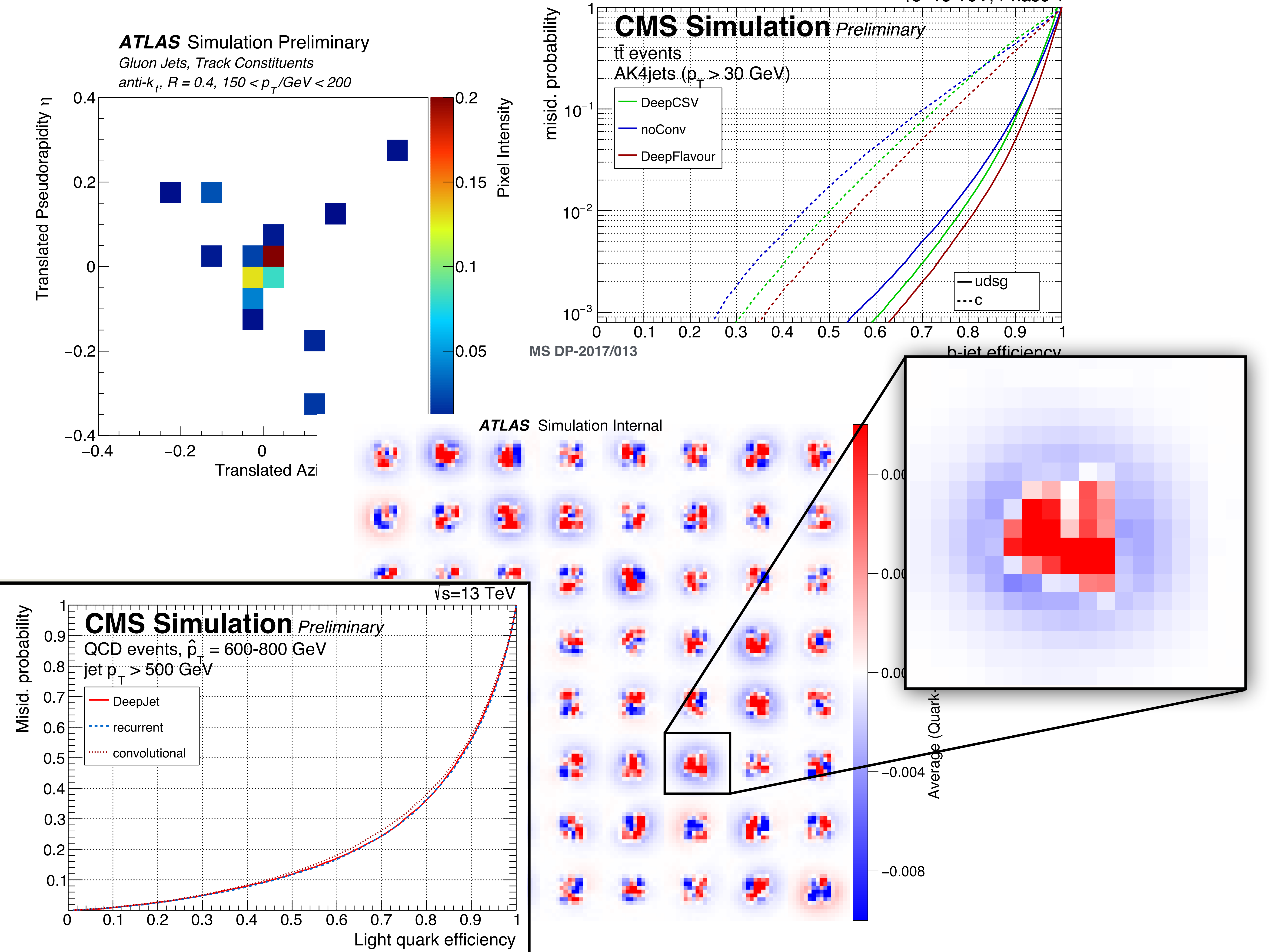
Idea: explicitly combine NN with a known feature

Has learned image mass?

not fully!

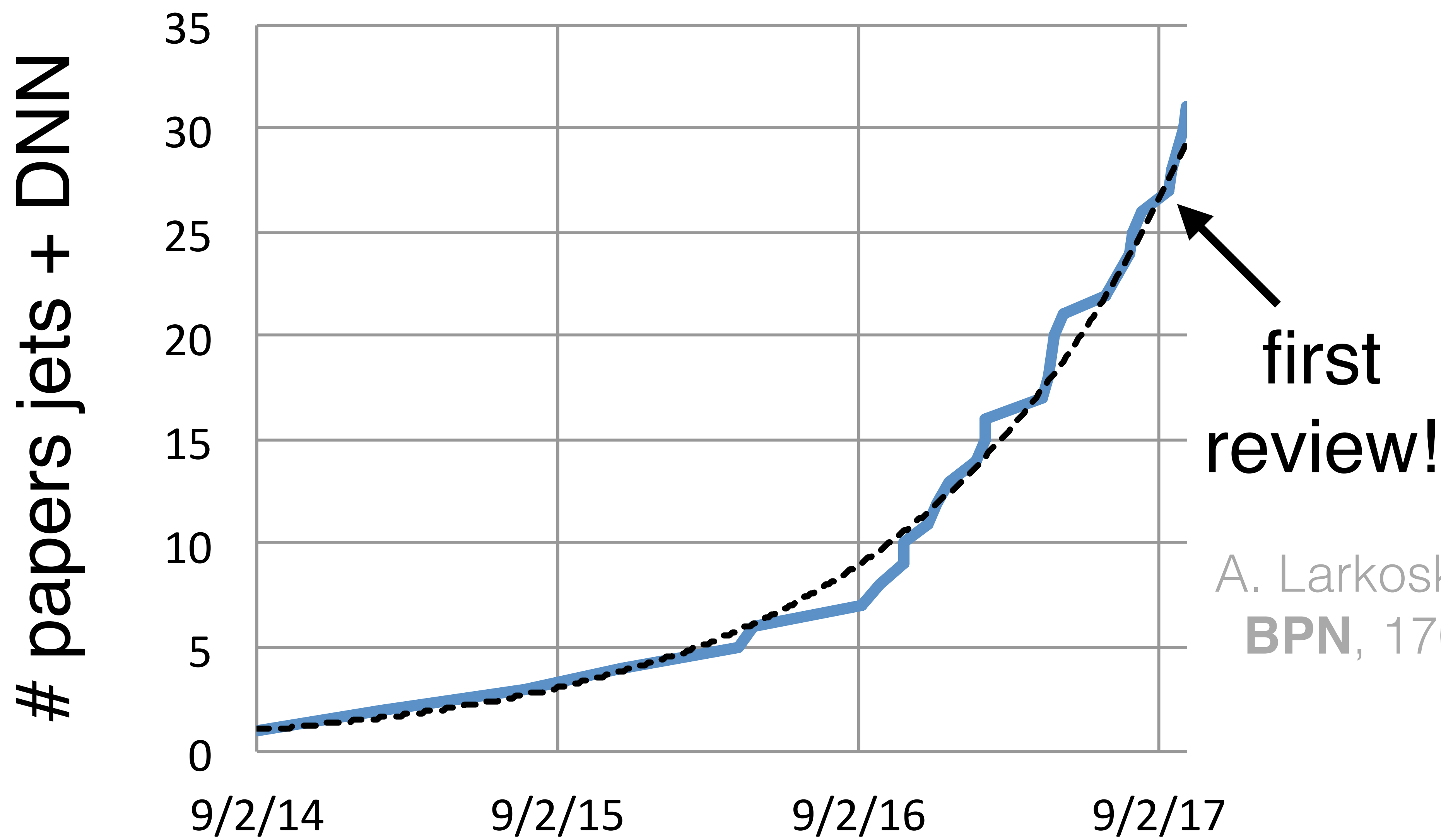
What about the distance between subjets?

yes!


ATLAS Simulation Preliminary
Gluon Jets, Track Constituents
anti-k_t, R = 0.4, 150 < p_T/GeV < 200
Translated Pseudorapidity η
Translated Azi
Pixel Intensity
CMS Simulation Preliminary
tt̄ events
AK4jets (p_T > 30 GeV)
DeepCSV
noConv
DeepFlavour
udsg
c
misid. probability
b-jet efficiency
MS DP-2017/013
ATLAS Simulation Internal
Average (Quark-
√s=13 TeV
CMS Simulation Preliminary
QCD events, p̂_T = 600-800 GeV
jet p_T > 500 GeV
DeepJet
recurrent
convolutional
Misid. probability
Light quark efficiency

Exciting New Directions

So far only scratches the surface
…this is a very active field of research!



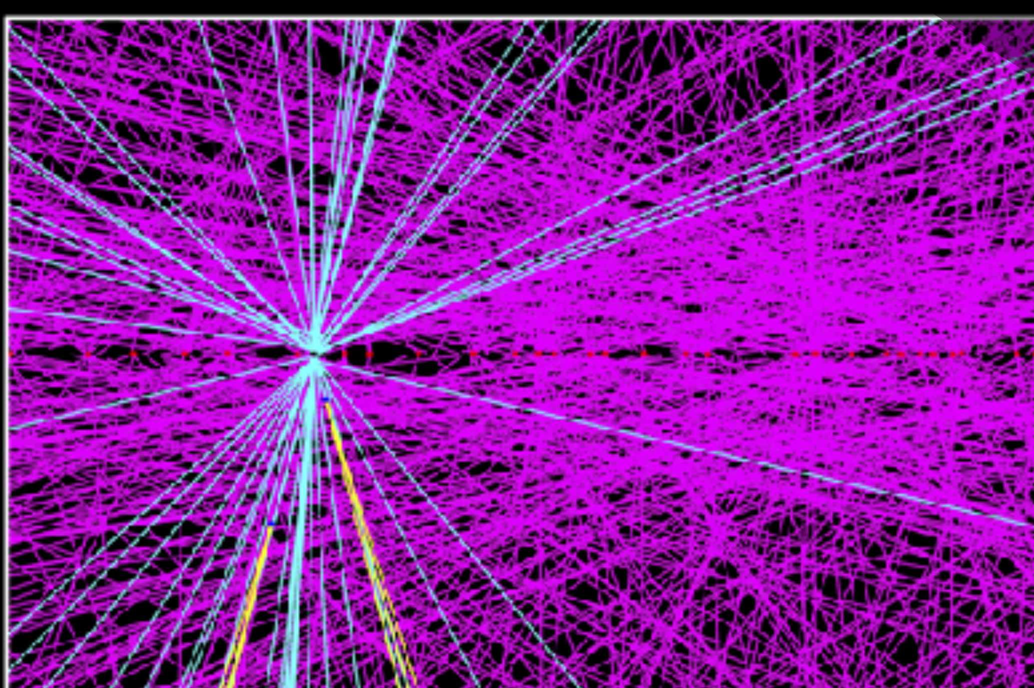
A. Larkoski, I. Moult, **BPN**, 1709.04464

Exciting New Directions I: Removing Noise

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

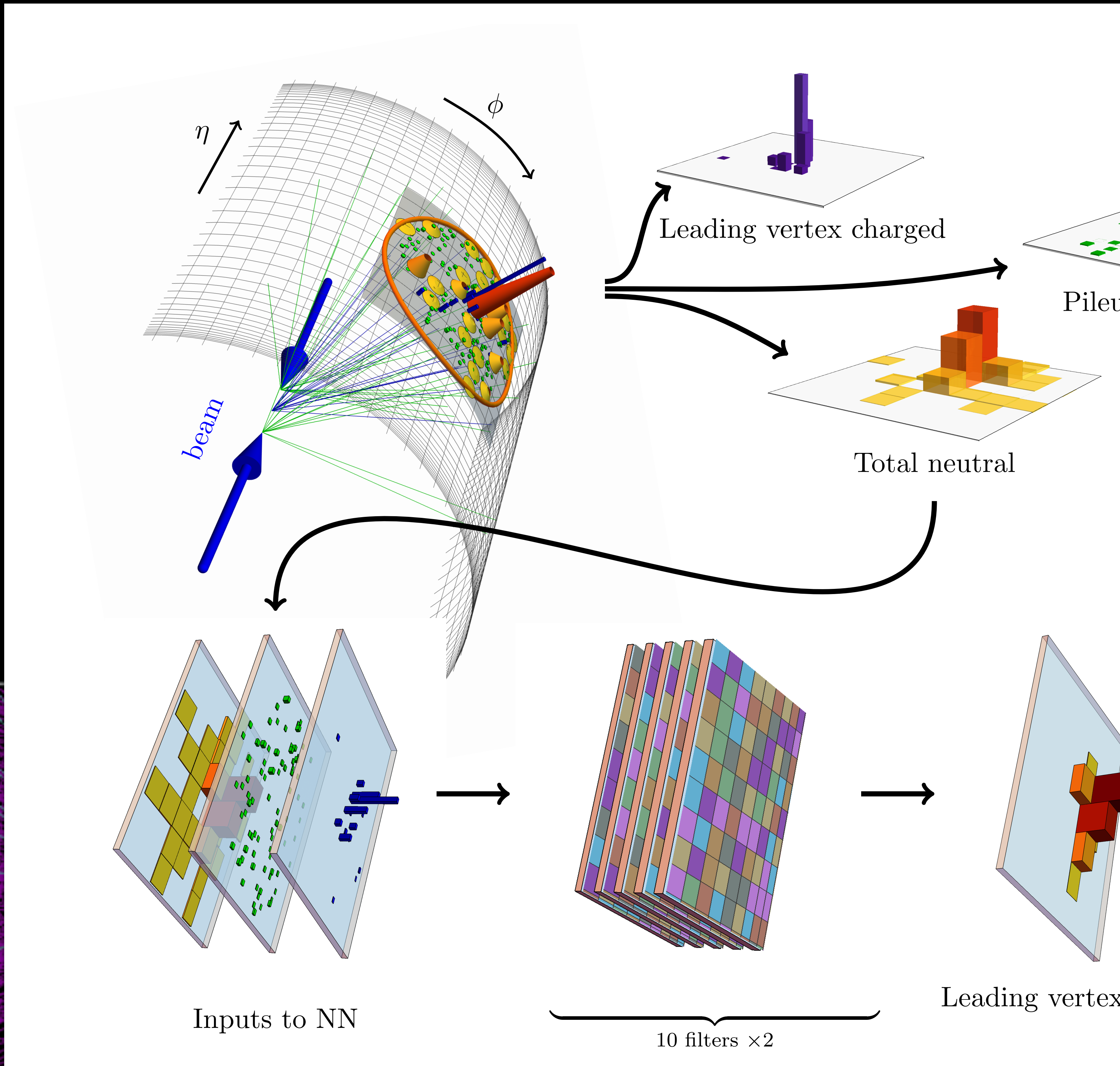
HL-LHC $t\bar{t}$ event in ATLAS ITK
at <μ>=200

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!

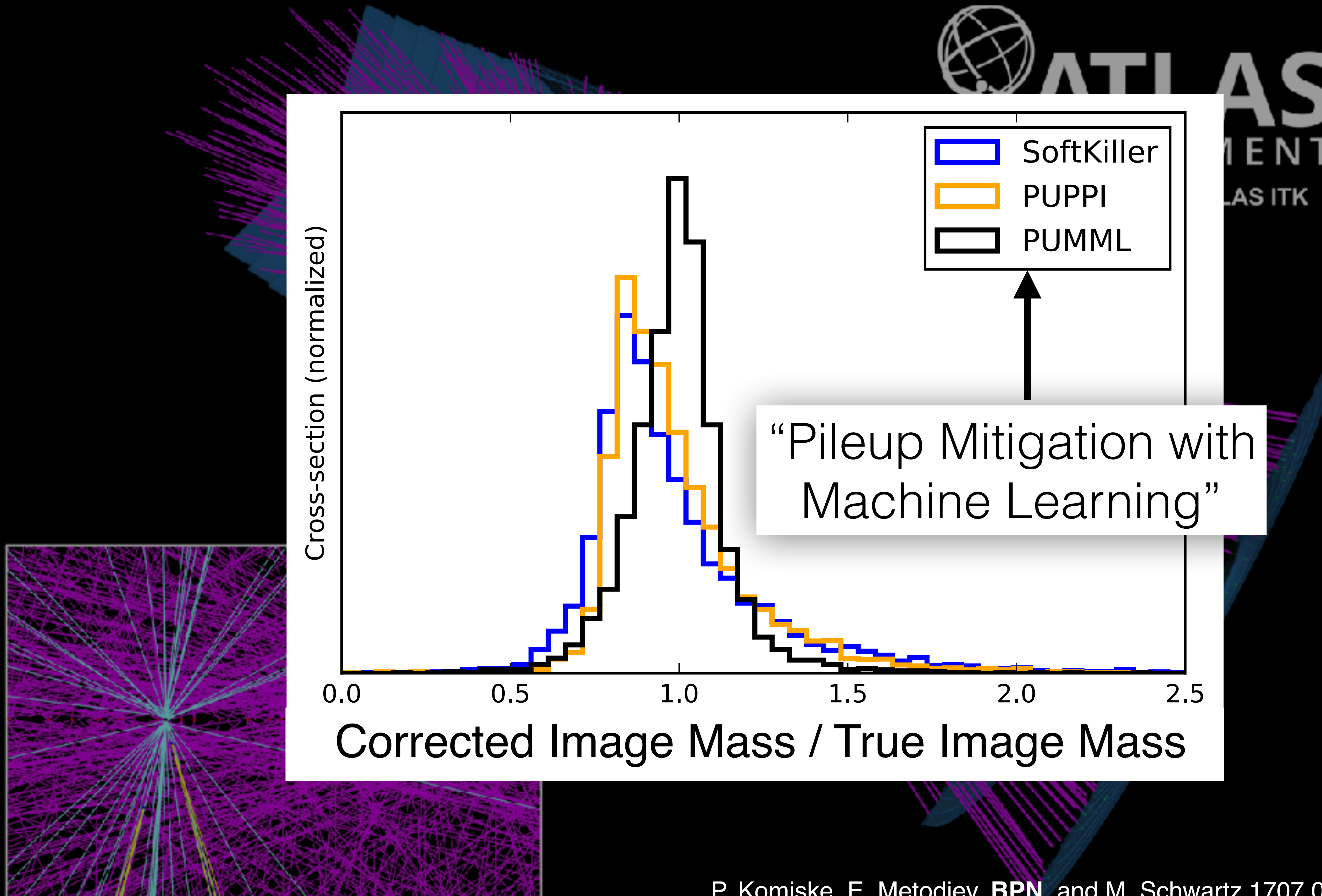
Exciting New Directions I: Removing Noise



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

…also a natural application of convolutional NNs!

Exciting New Directions I: Removing Noise

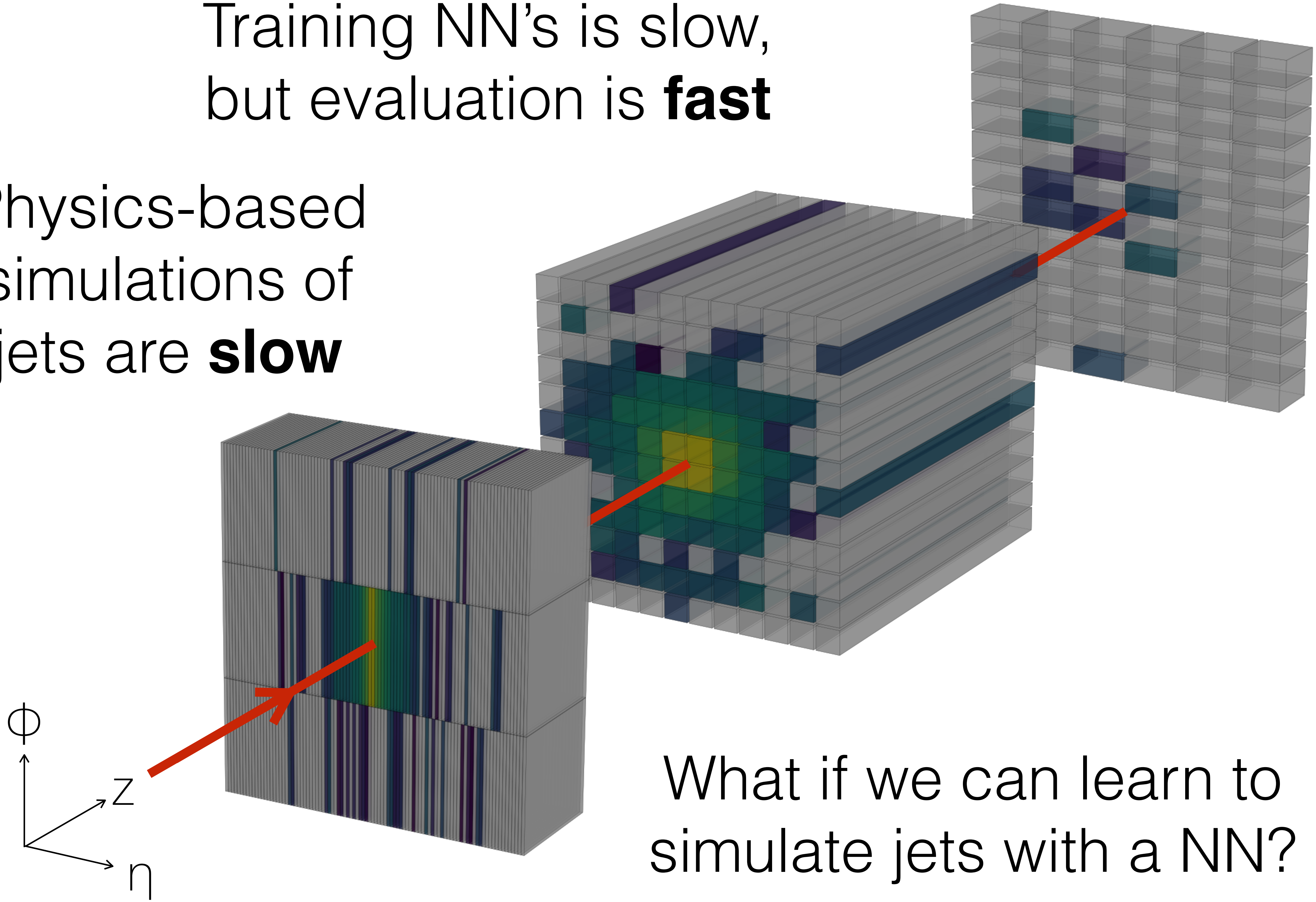


P. Komiske, E. Metodiev, **BPN**, and M. Schwartz,1707.08600

Exciting New Directions II: Simulation NN

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

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



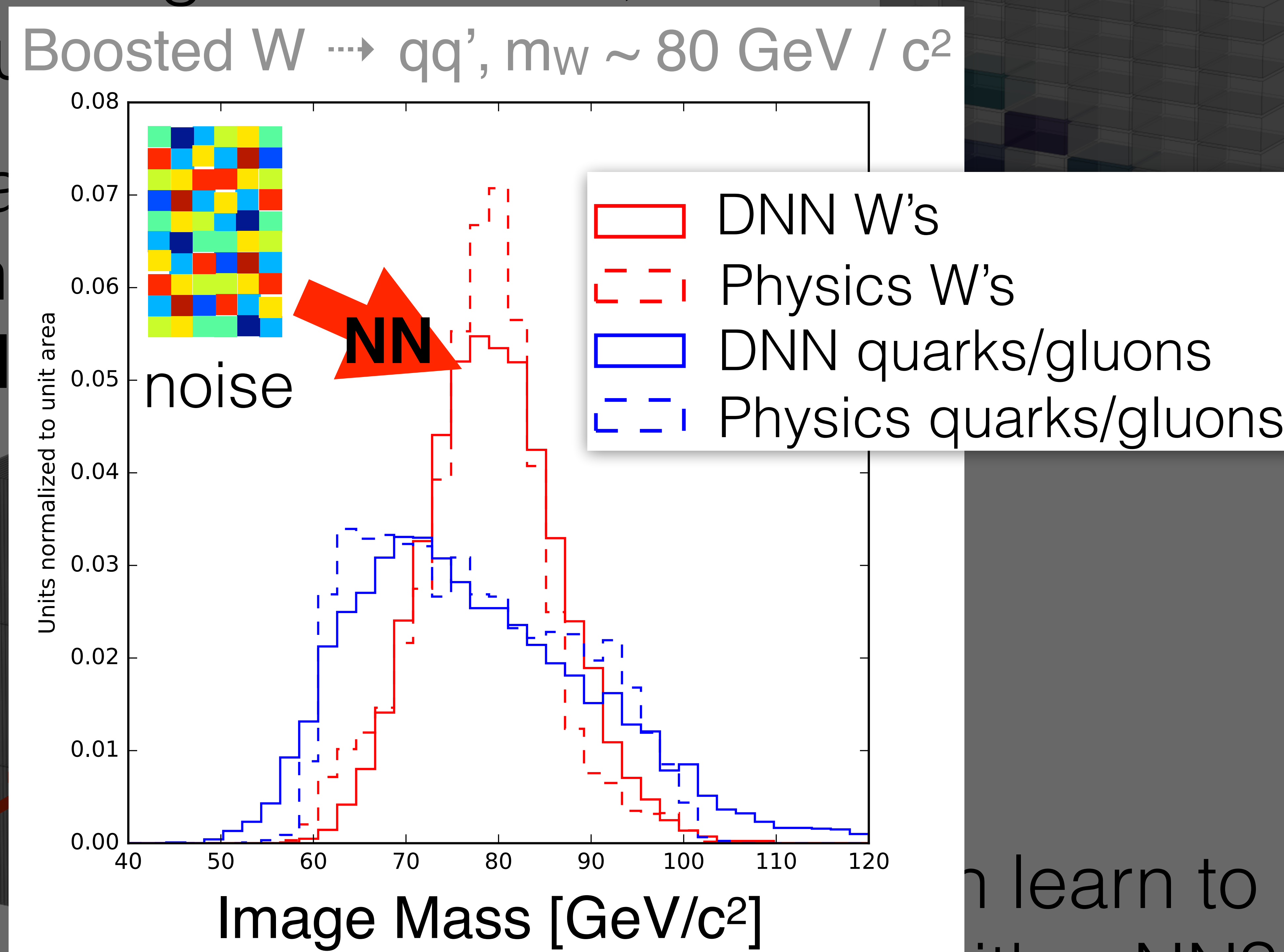
What if we can learn to simulate jets with a NN?

Exciting New Directions II: Simulation NN

Training NN's is slow,

bu

Physics-ba
simulation
jets are **sl**



n learn to
simulate jets with a NN?

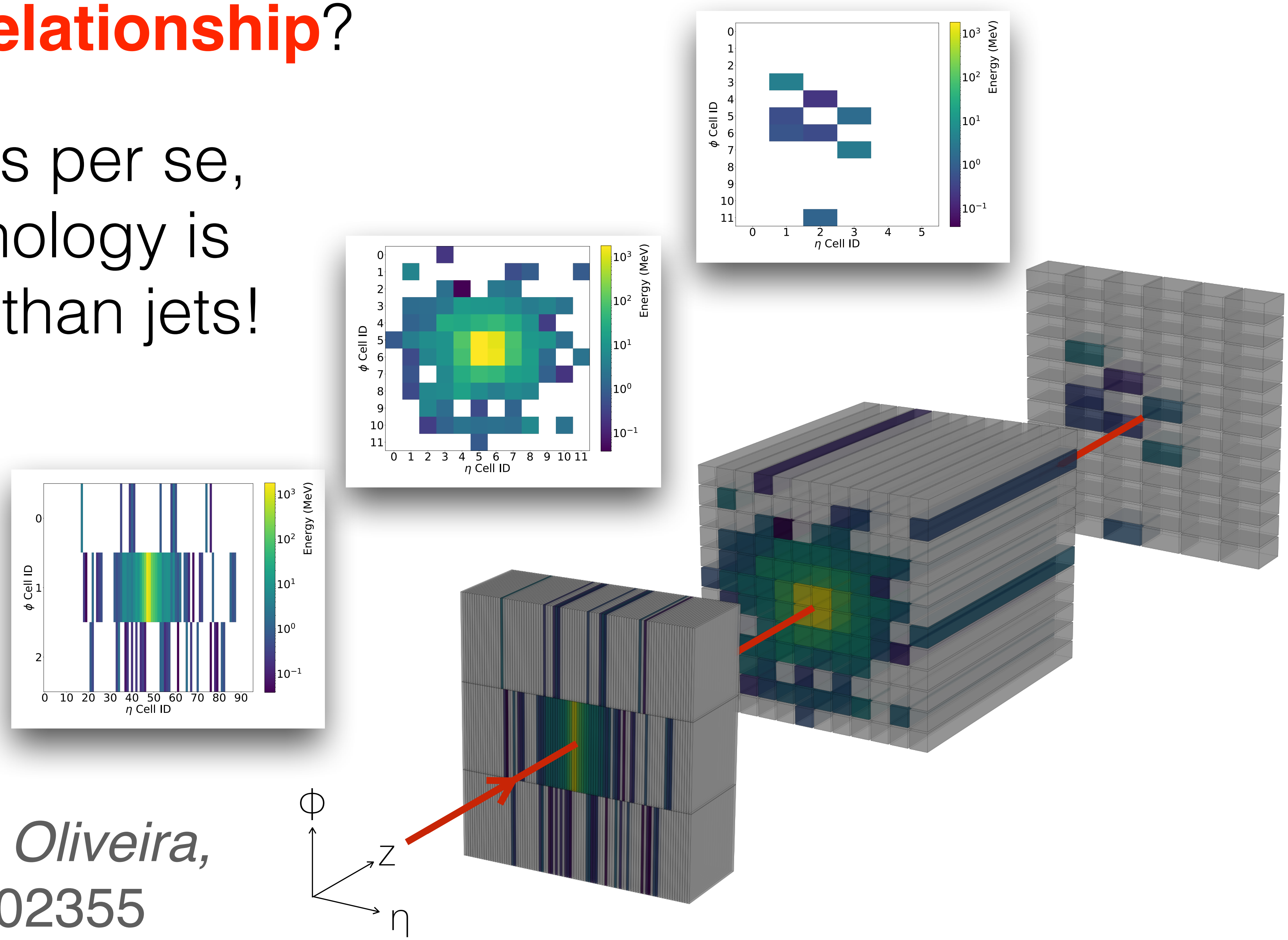
L. de Oliveira, M. Paganini, **BN**, Comp. and Software for Big Science (2017) 1

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

+ 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!

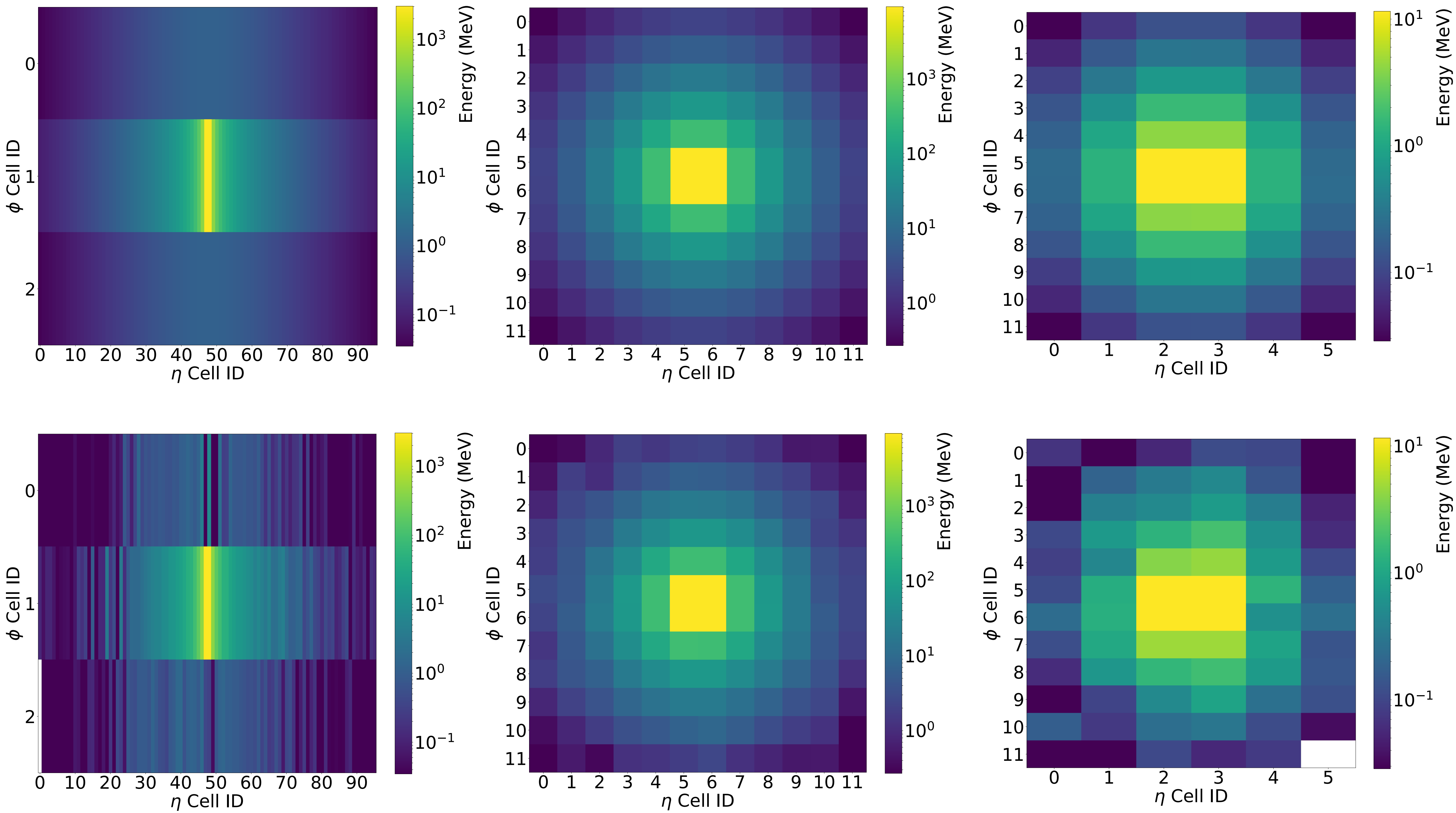


*M. Paganini, L. de Oliveira, and **BPN*** 1705.02355

Average Images

Geant4

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



CaloGAN

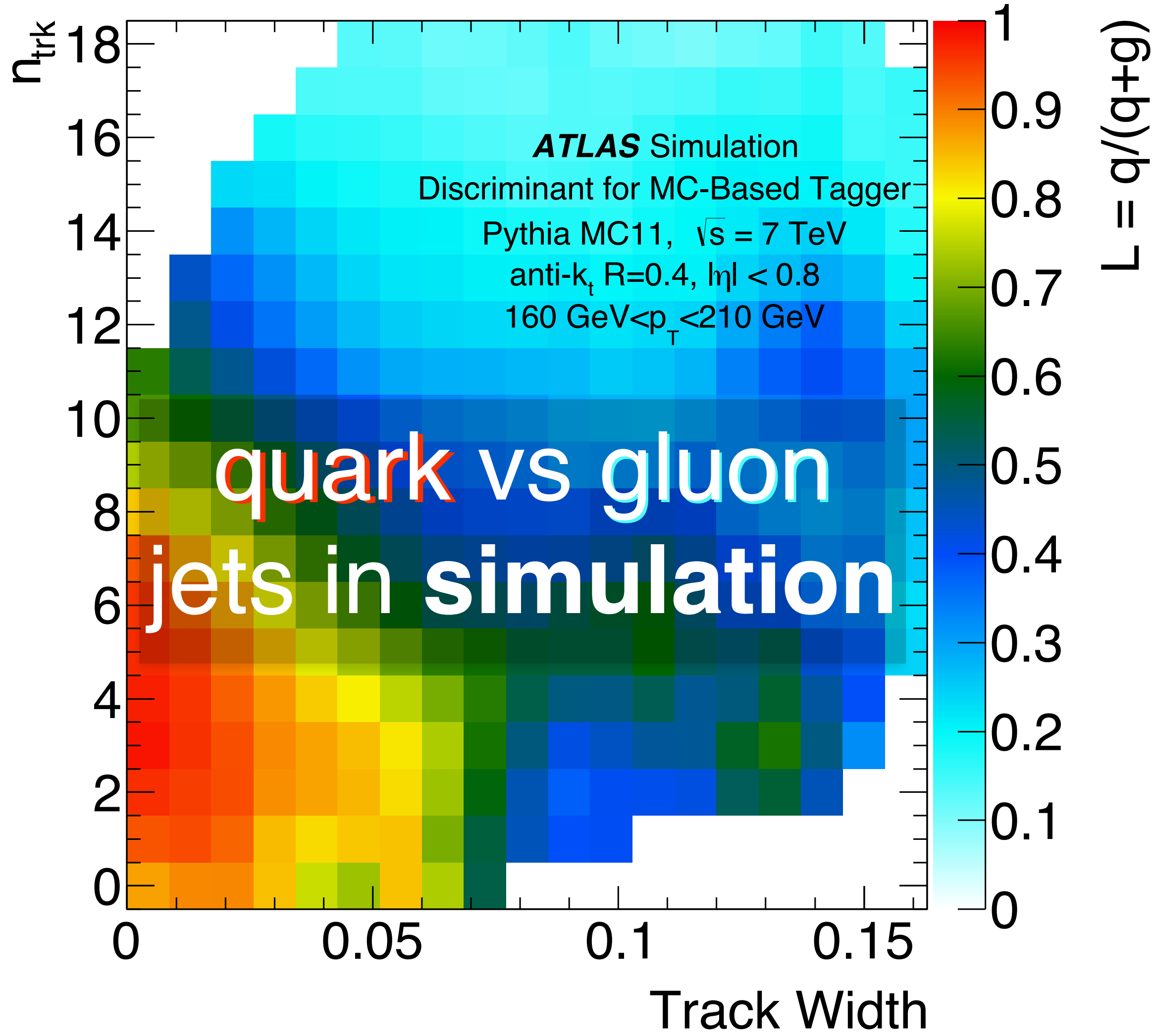
Timing

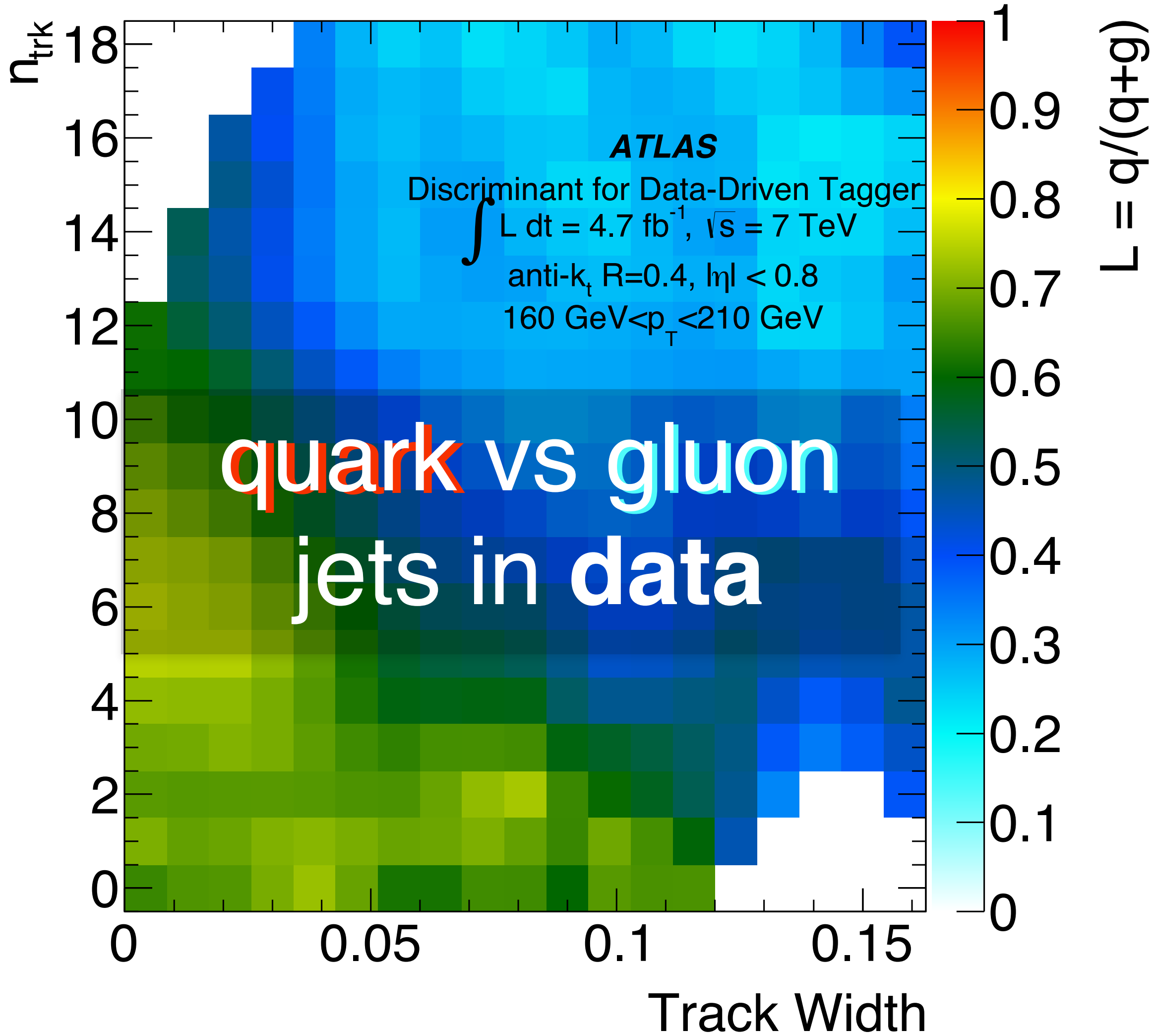
M. Paganini, L. de Oliveira, and ***BPN*** 1705.02355

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 ←

Where next III: Learning directly from data

For supervised learning, we depend on labels
labels usually come from simulation



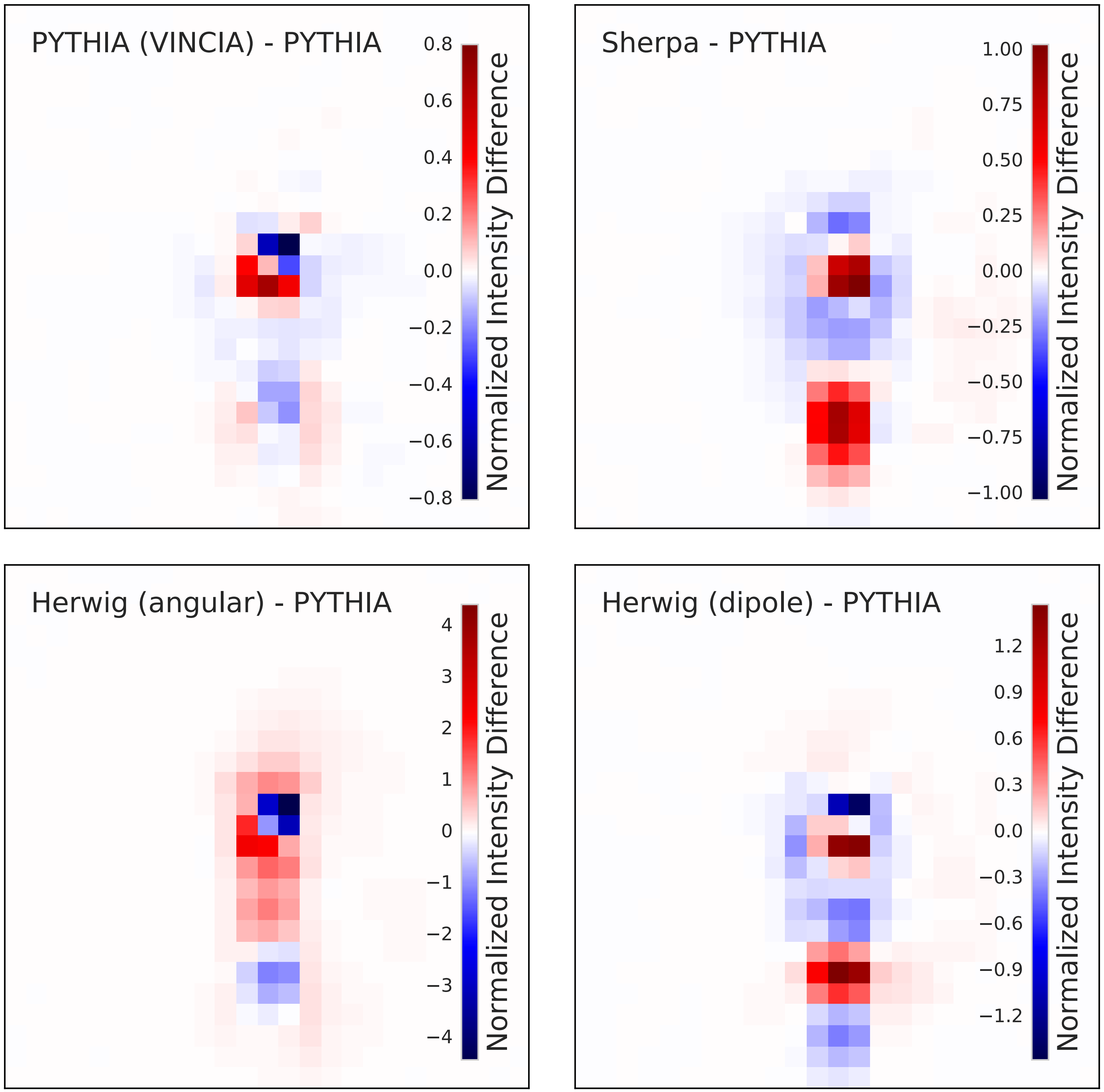


What if data and simulation are very different?
…your classifier will be sub-optimal

Where next III: Learning directly from data

Boosted *W* boson jets

J. Barnard et al.
Phys. Rev. D 95, 014018 (2017)



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

DNN classifiers can **exploit** subtle features

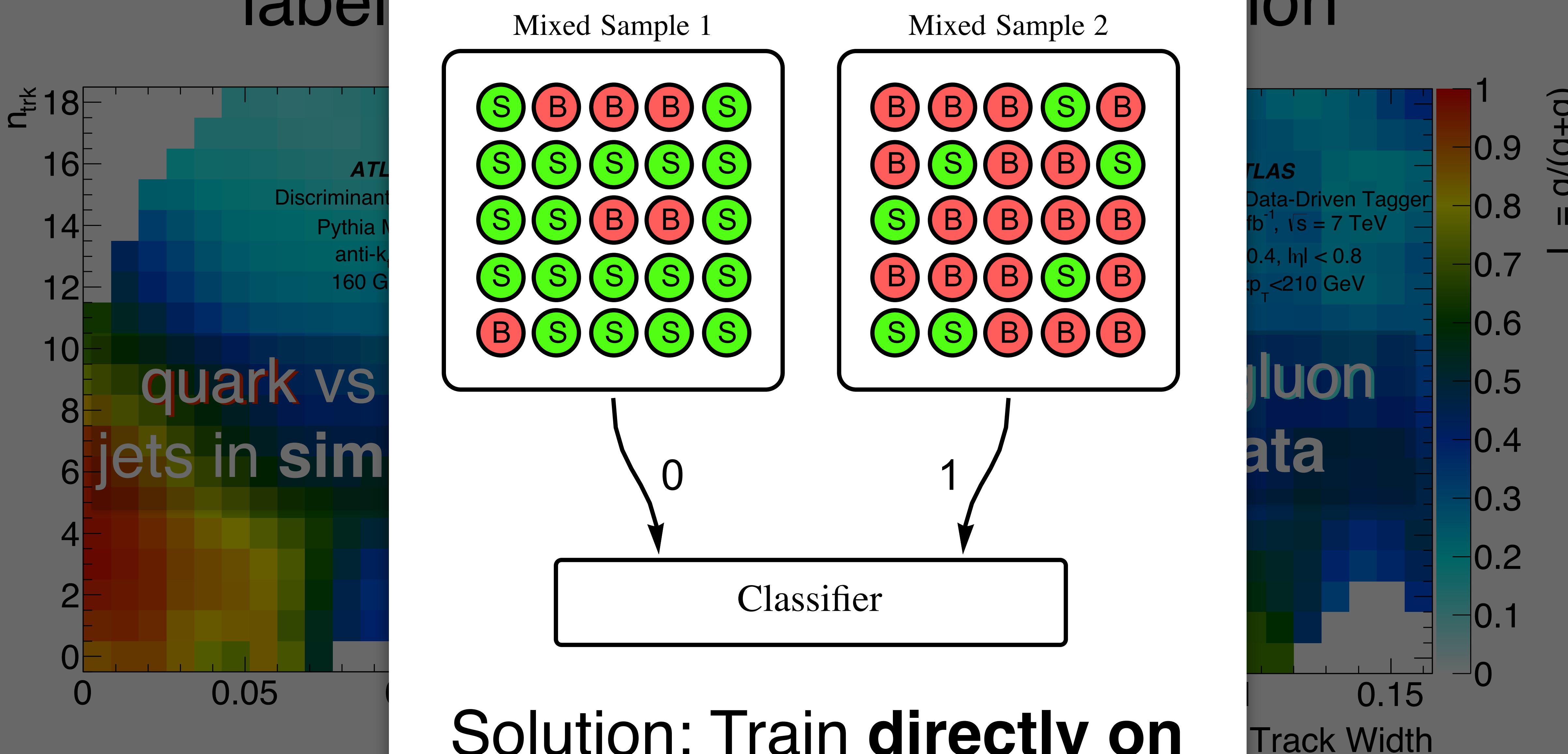
subtle features are **hard to model** !

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

Where next III: Learning directly from data

For supervised learning, we depend on labels

label… …ion



Solution: Train **directly on data** using mixed samples

What if da… …ifferent?

…your classifier will be sub-optimal

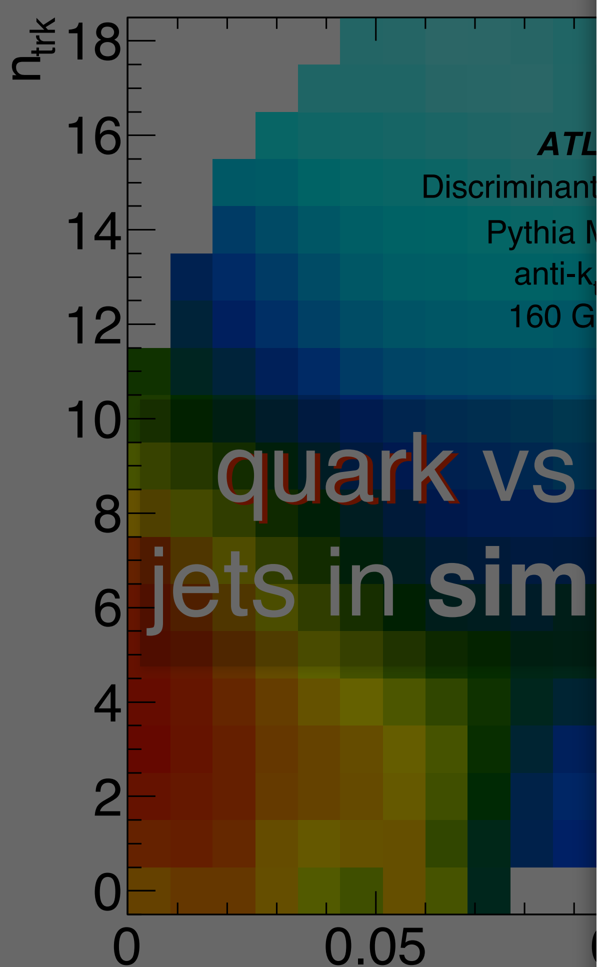
L. Dery, **BPN**, F. Rubbo, A. Schwartzman, JHEP 05 (2017) 145

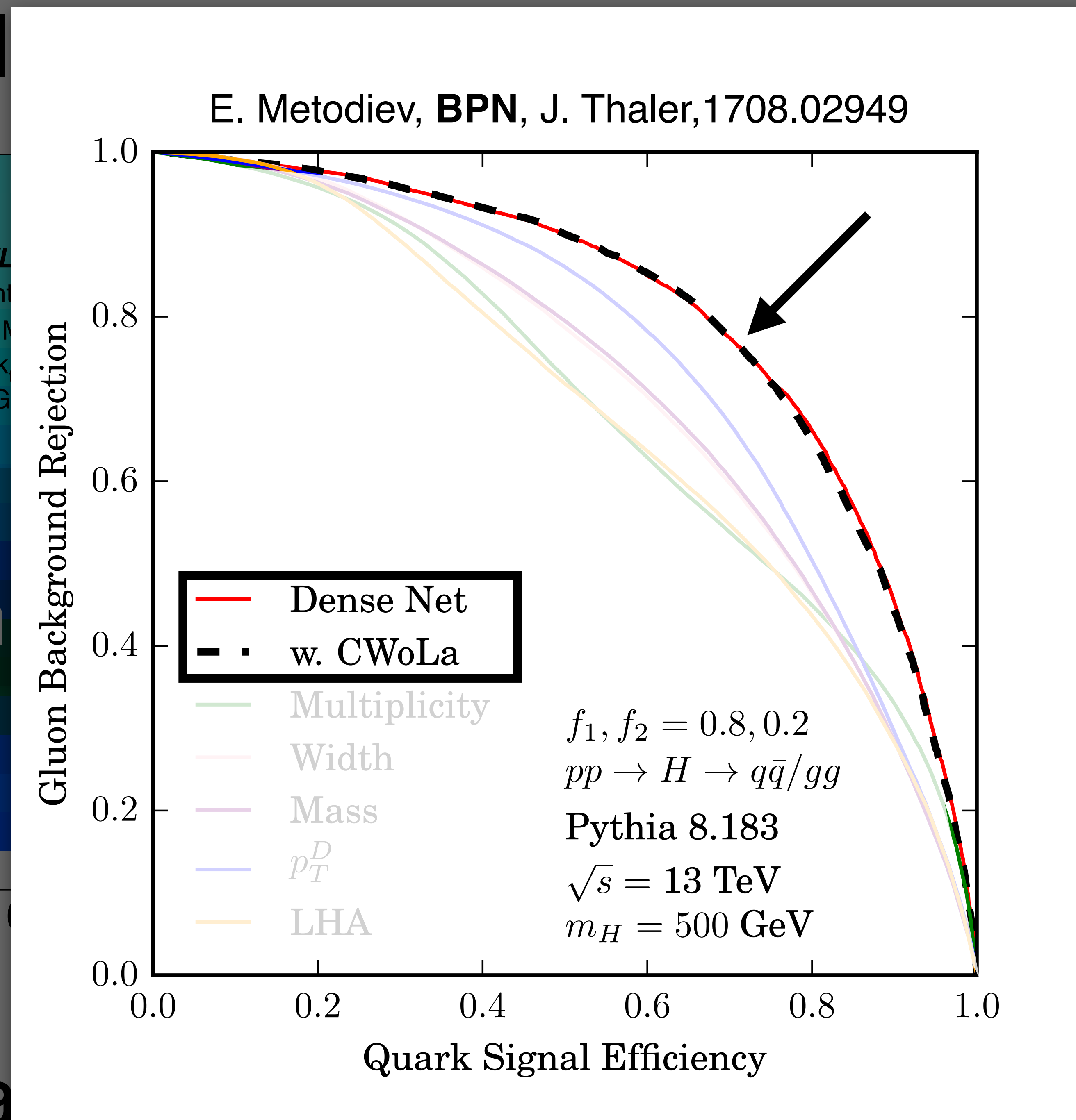
E. Metodiev, **BPN**, J. Thaler,1708.02949

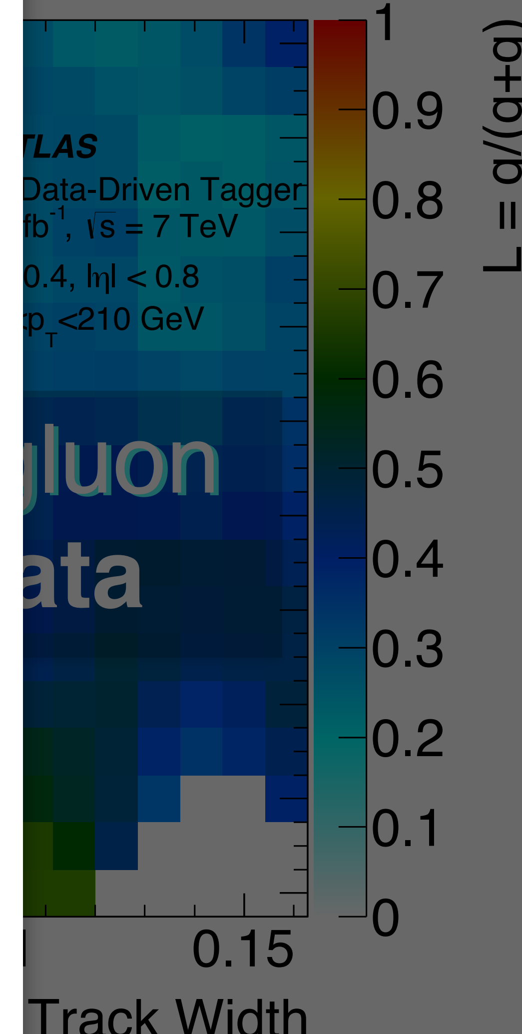
Where next III: Learning directly from data

For supervised learning, we depend on labels

label ... ion





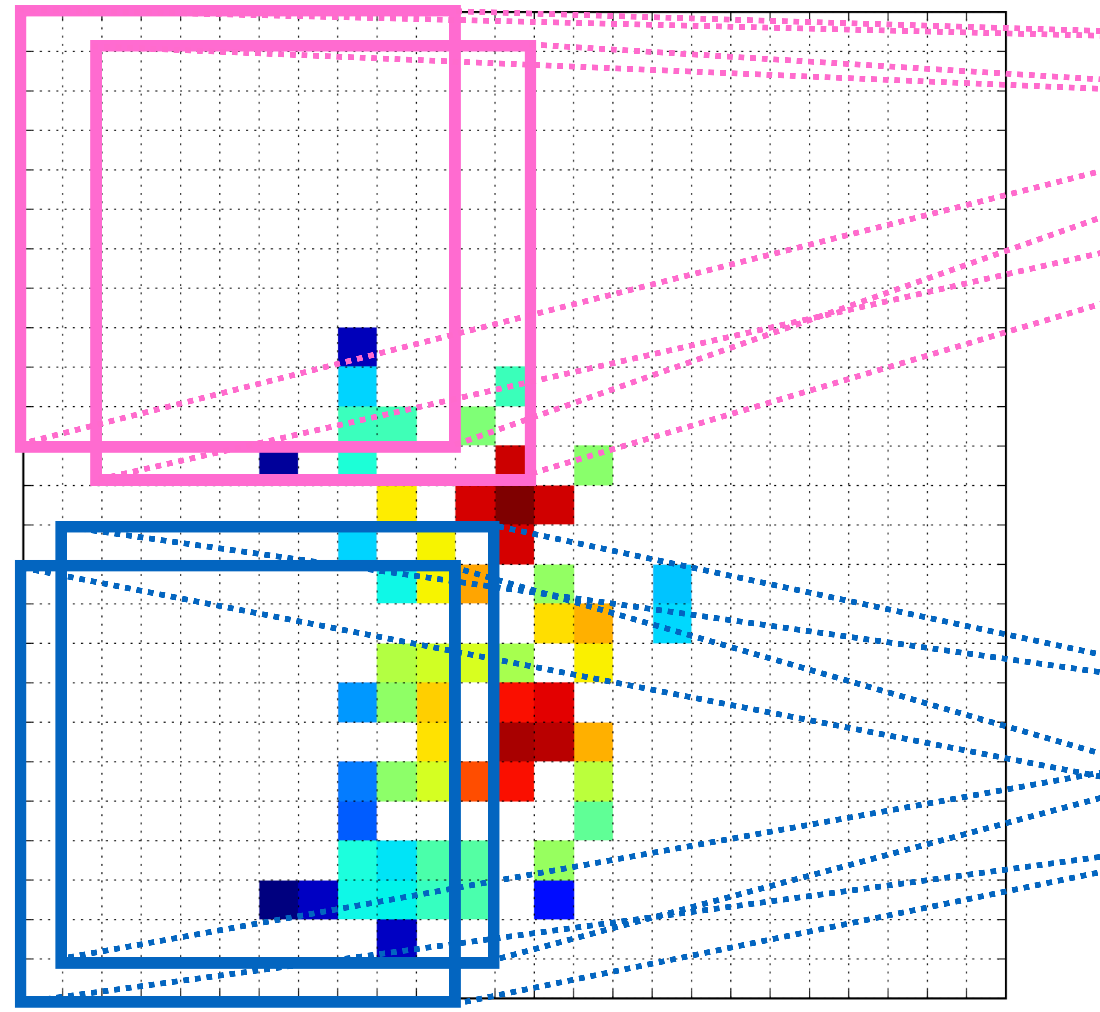


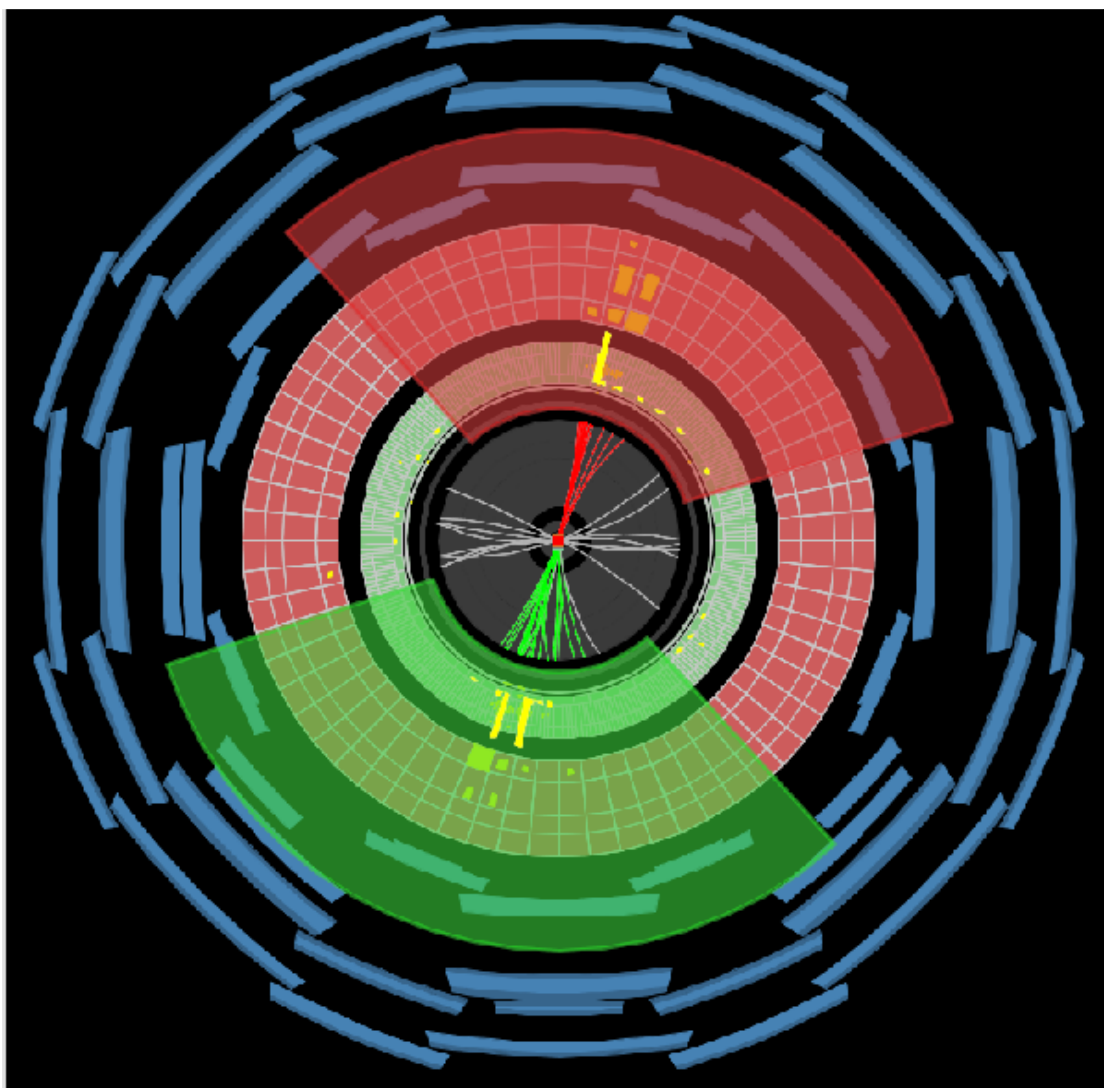
What if da ... ifferent?

…your classifier will be sub-optimal

Conclusions and outlook

(Jet) image-based NN classification, regression, and generation are powerful tools for fully exploiting the physics program at the LHC





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

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

Workshop Advertisement

Machine Learning for Jet Physics

11-13 December 2017
Lawrence Berkeley National Laboratory
US/Pacific timezone

link

- Overview
- Scientific Programme
- Call for Abstracts
  - View my Abstracts
  - Submit Abstract
- Timetable
- Contribution List
- Author List
- My Conference
- Book of Abstracts
- Registration
  - Modify my Registration
- Participant List

There has been a recent surge of interest in developing and applying advanced machine learning techniques in HEP, and jet physics is a domain at the forefront of the excitement. The goal of this workshop is to gather experts and new-commers to discuss progress, new ideas, and common challenges. The workshop is open to the community; we invite contributions and will try to accommodate everyone within reason.

Starts 11 Dec 2017 08:00
Ends 13 Dec 2017 18:00
US/Pacific

Lawrence Berkeley National Laboratory

Nachman, Benjamin
Dr. Cohen, Timothy
Dolan, Matt
Cranmer, Kyle

No material yet

There is no fee for attending the workshop. Coffee and light refreshments will be provided during breaks but meals and lodging are the responsibility of the attendant.

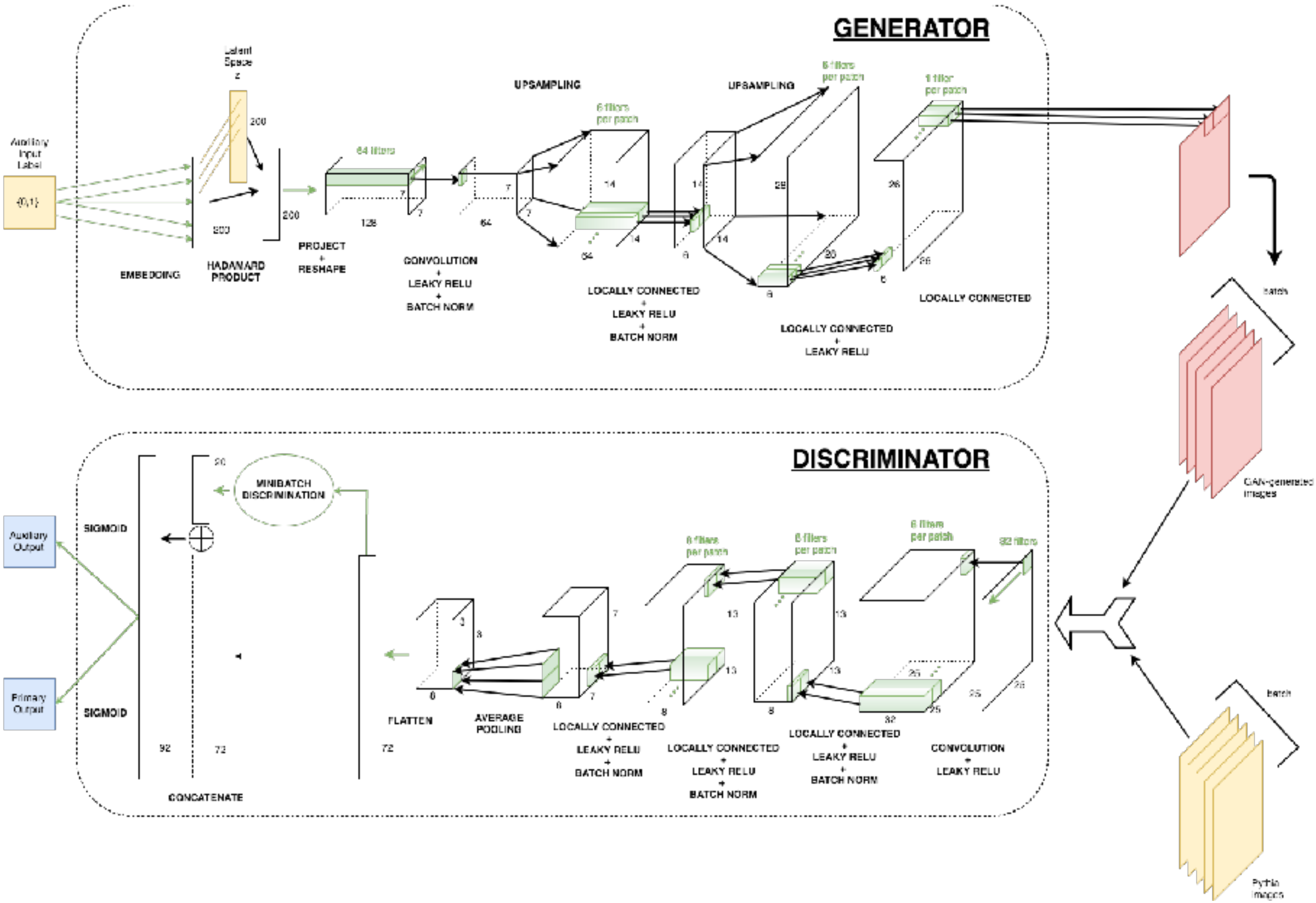
There are many hotels in the Berkeley area, including limited availability at the LBNL guesthouse (5 min walk from the workshop, $140/night). A complementary shuttle runs every 10 min from downtown Berkeley up to the lab. For lunches, the most convenient option will be to eat in the LBNL cafeteria (5 min walk from workshop, ~10$).

Related workshops:

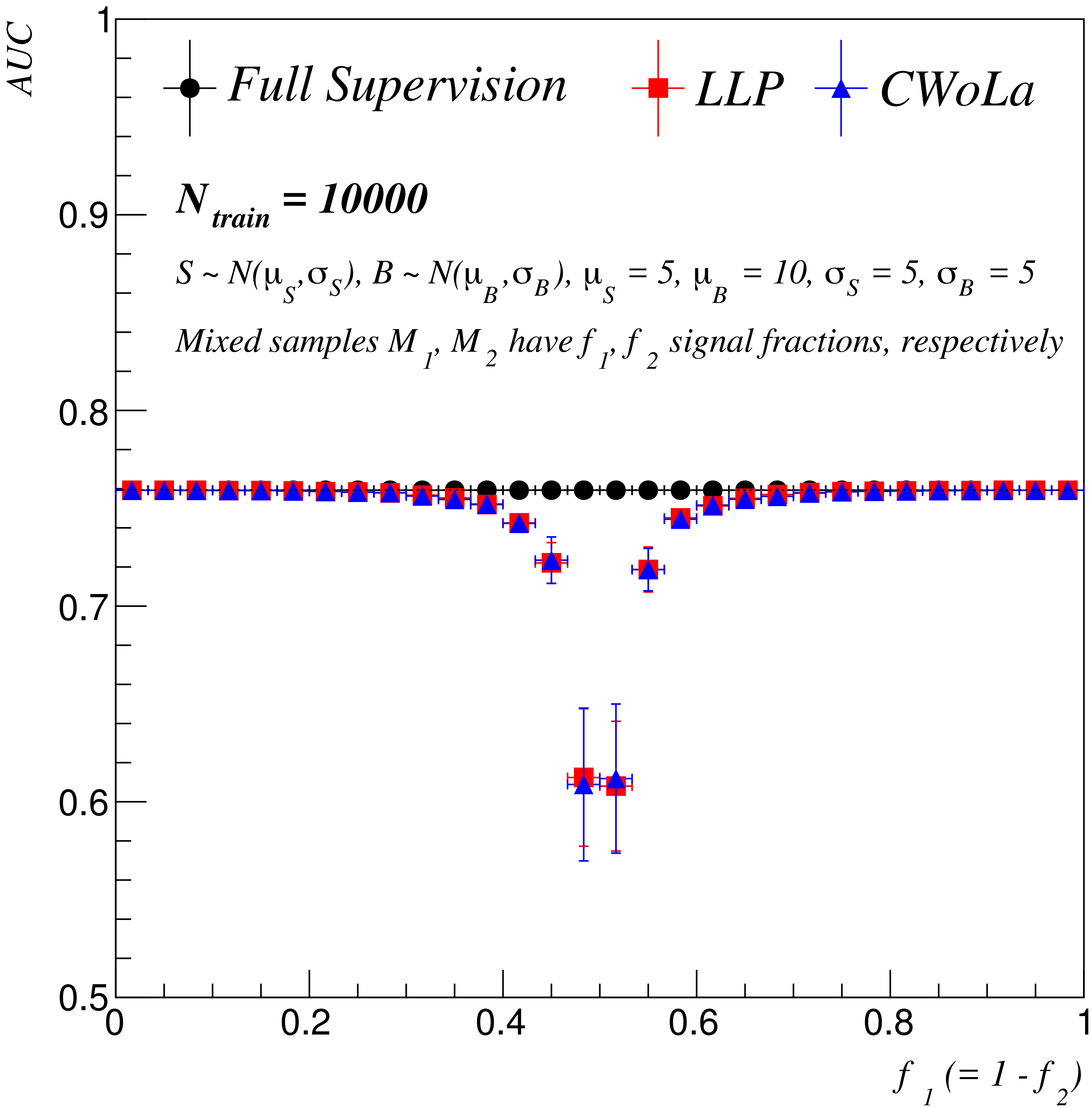
DS@HEP: https://indico.fnal.gov/conferenceDisplay.py?ovw=True&confId=13497
BOOST: https://indico.cern.ch/event/579660/

Backup

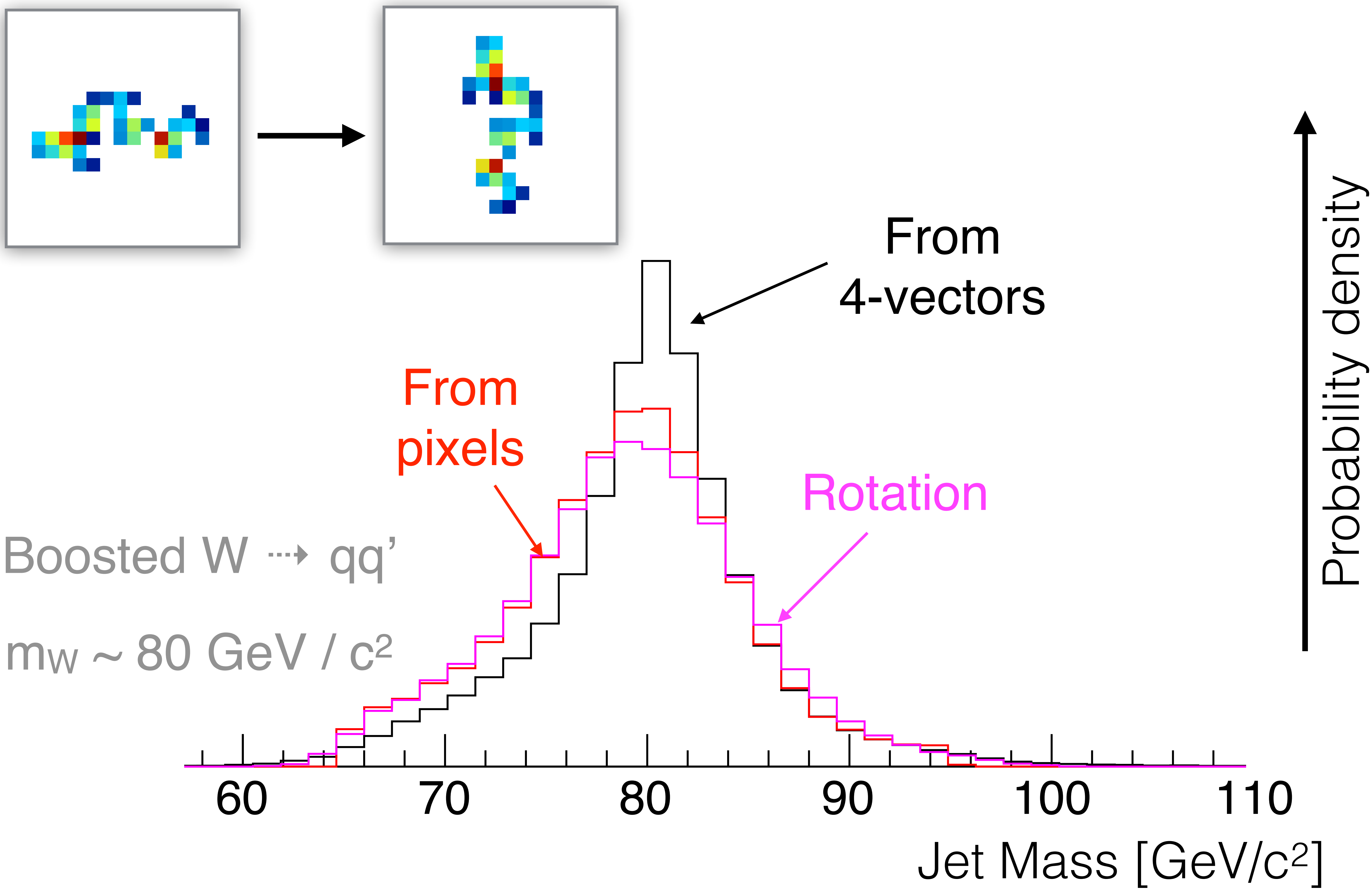
Locally Aware GAN (LAGAN)



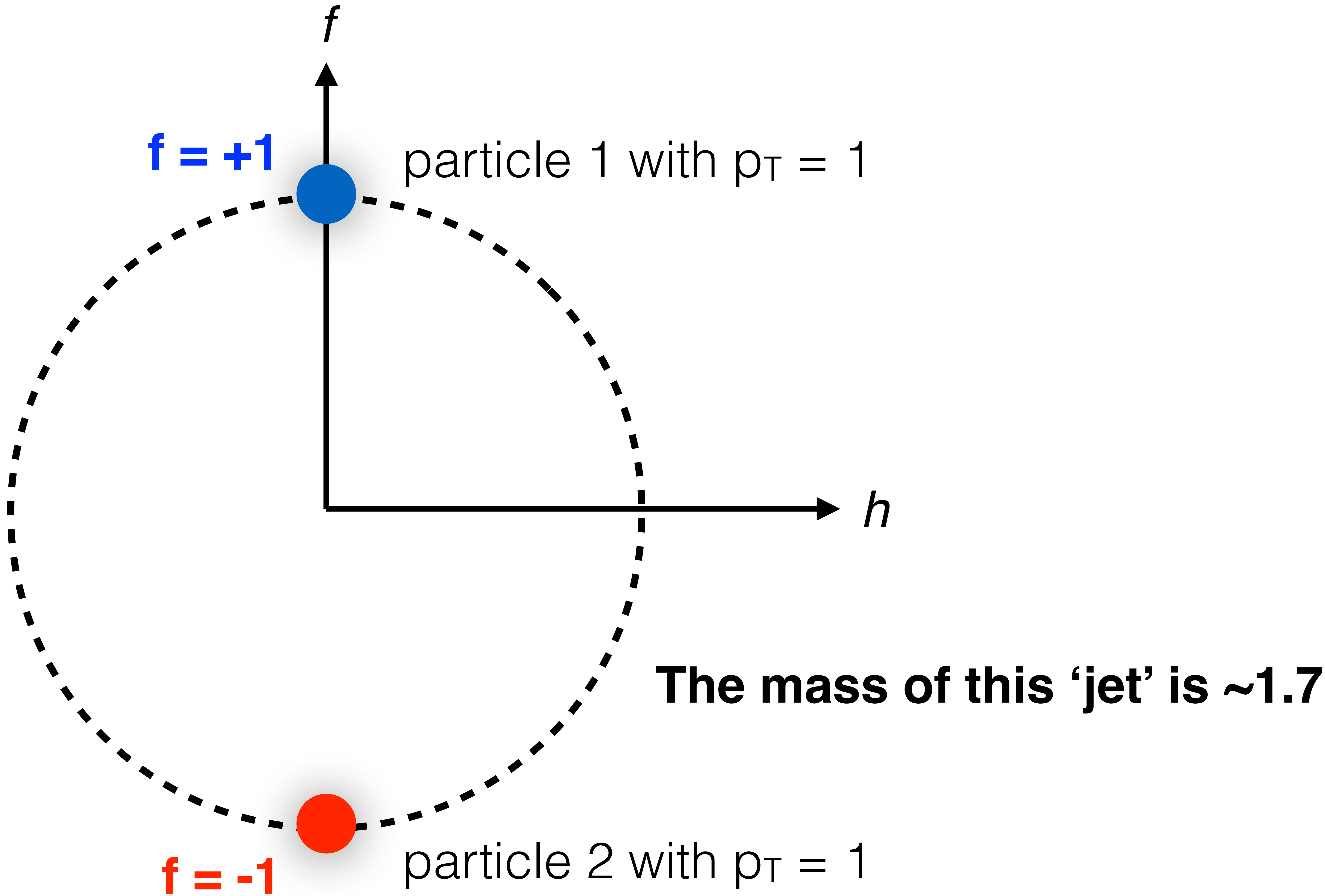
Learning when you know (almost) nothing



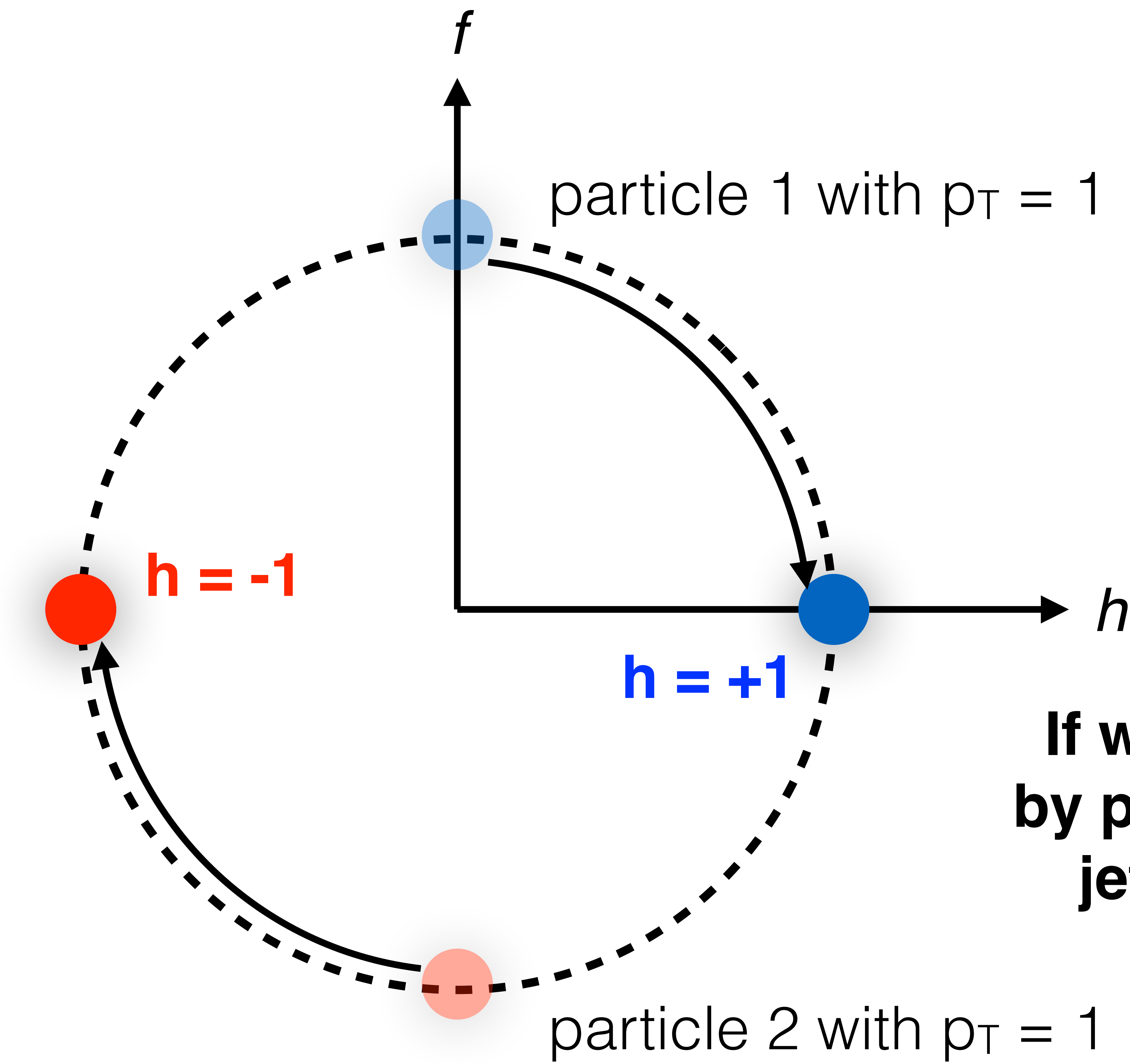
Pre-processing & spacetime symmetries



Pre-processing & spacetime symmetries



Pre-processing & spacetime symmetries

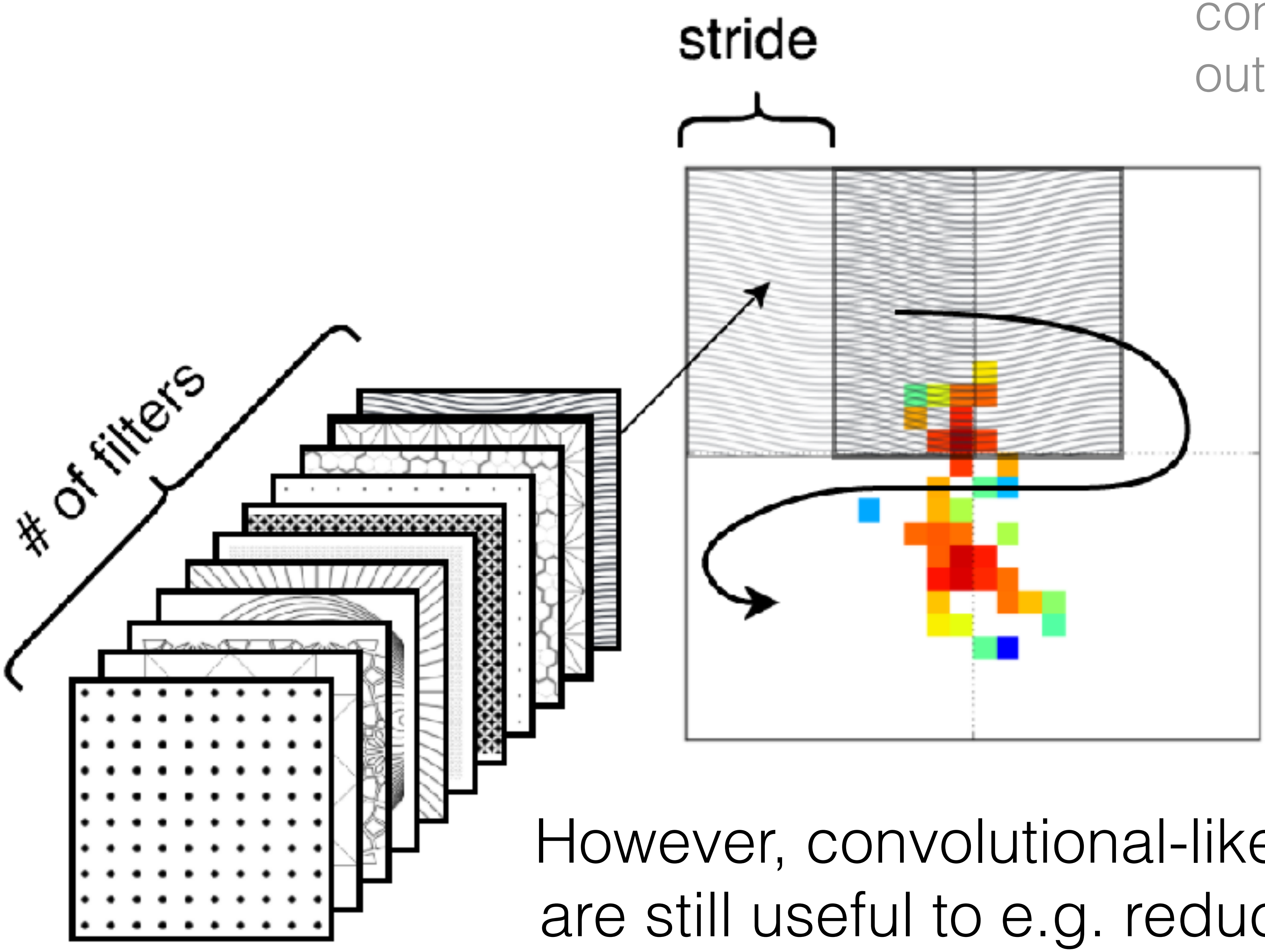


If we rotate the jet by p/2, then the new jet mass is ~2.4

Locally Connected Layers

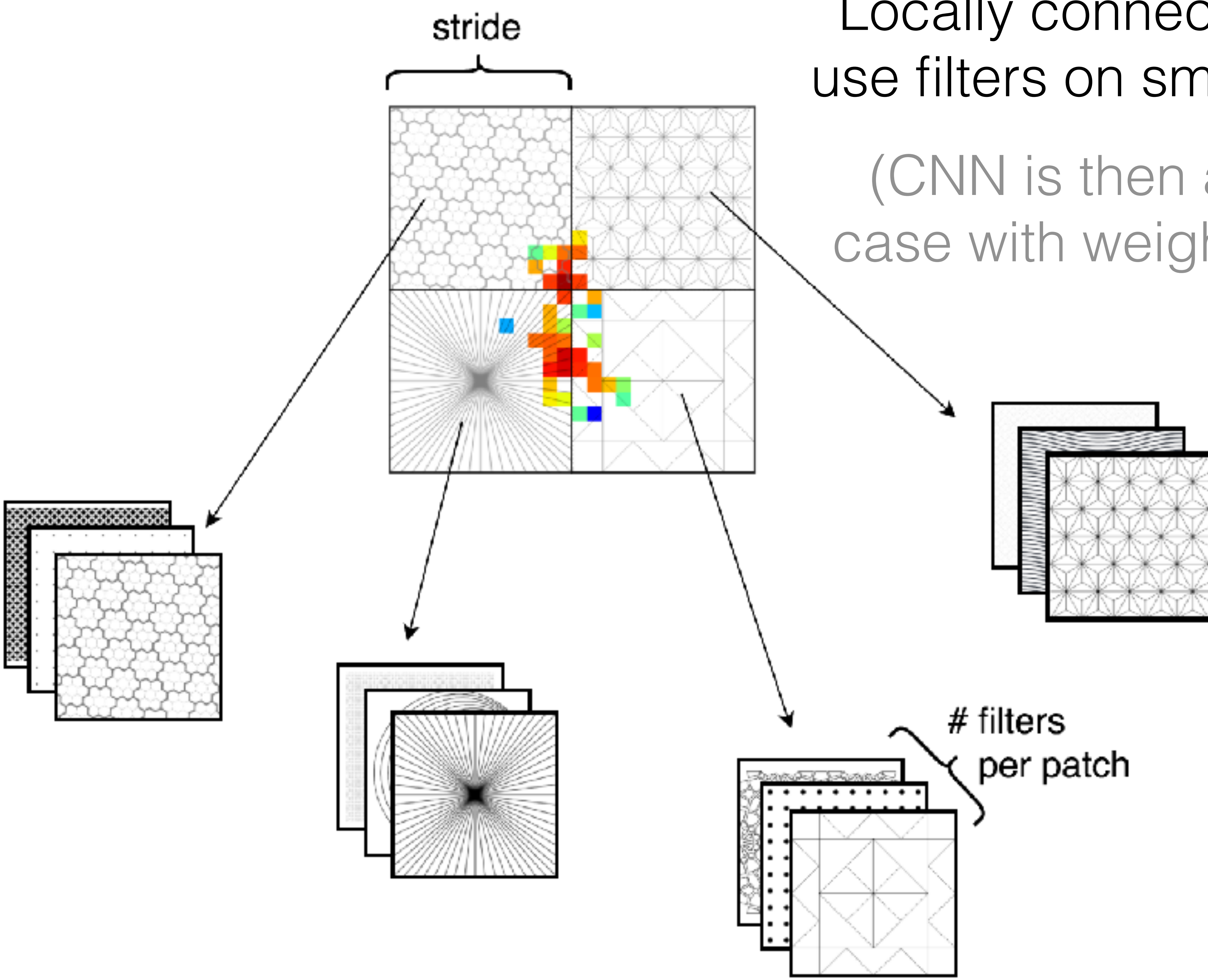
Due to the structure of the problem, we do not have translation invariance.

Classification studies found fully connected networks outperformed CNNs



However, convolutional-like architectures are still useful to e.g. reduce parameters

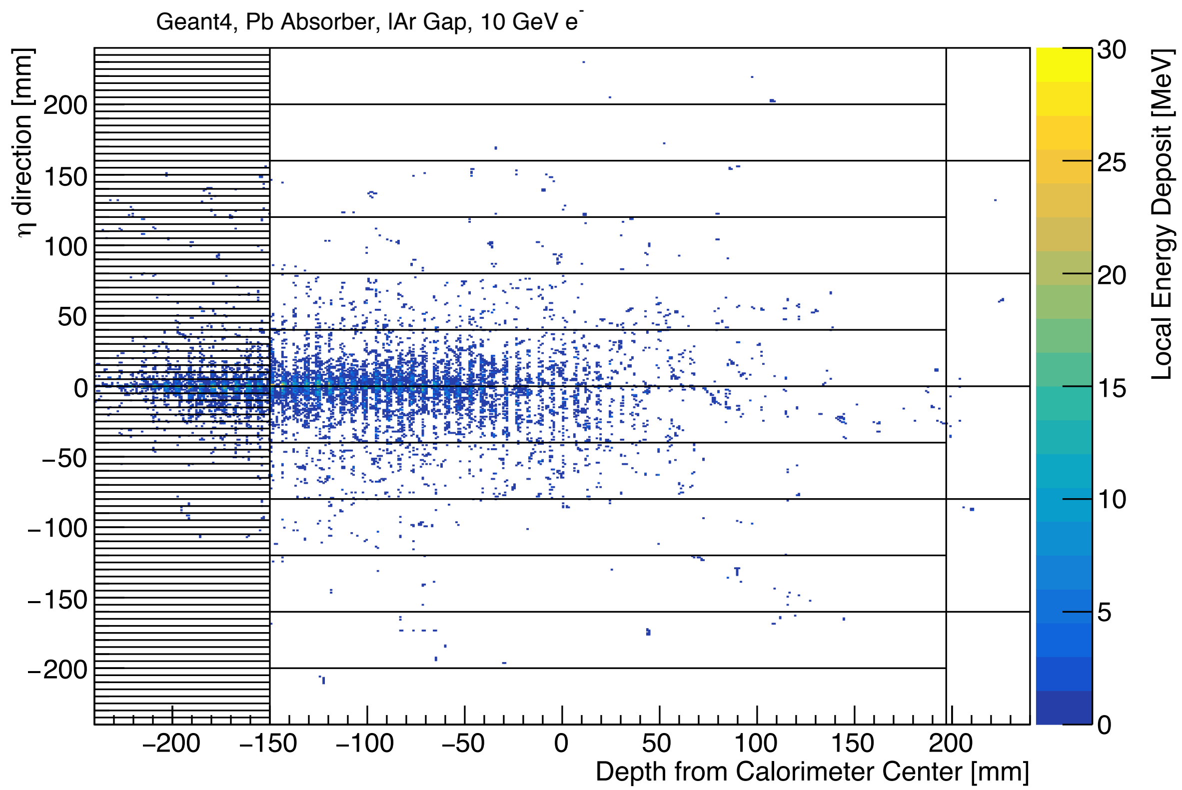
Locally Connected Layers



Locally connected layers use filters on small patches

(CNN is then a special case with weight sharing)

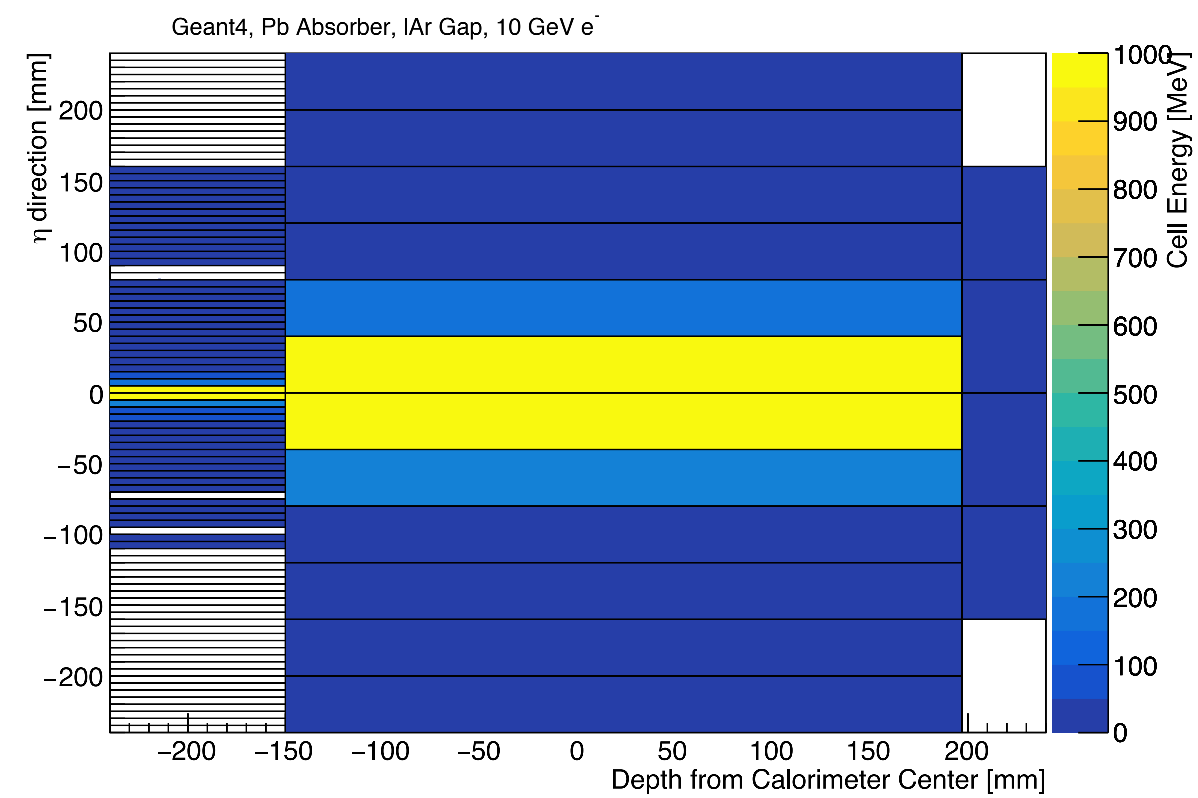
Calorimeter Simulation



We take as our model a 3-layer LAr calorimeter, inspired by the ATLAS barrel EM calorimeter

A single event may have O(10³) of particles showering in the calorimeter - too cumbersome to do all at once (now)

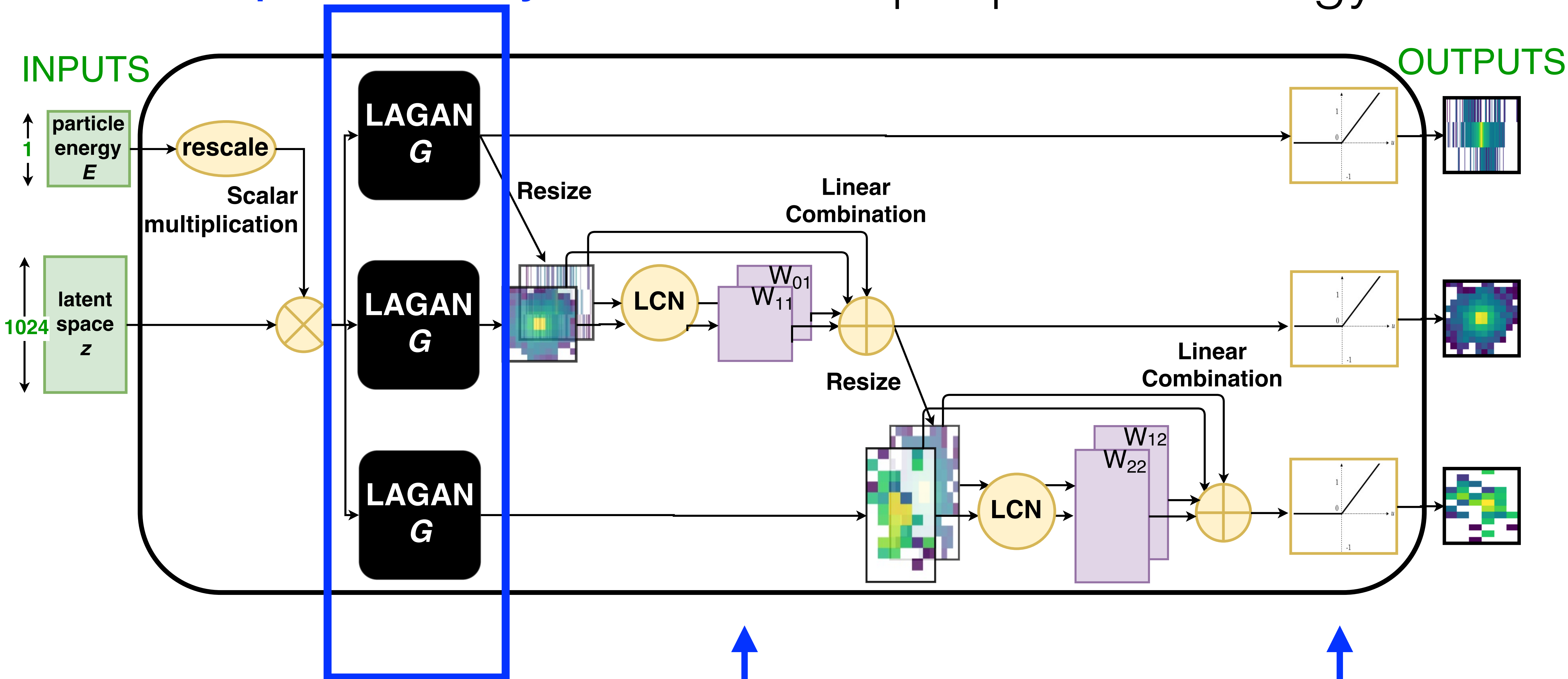
We exploit factorization of energy depositions



Generator Network for CaloGAN

One 'jet image' per calo layer

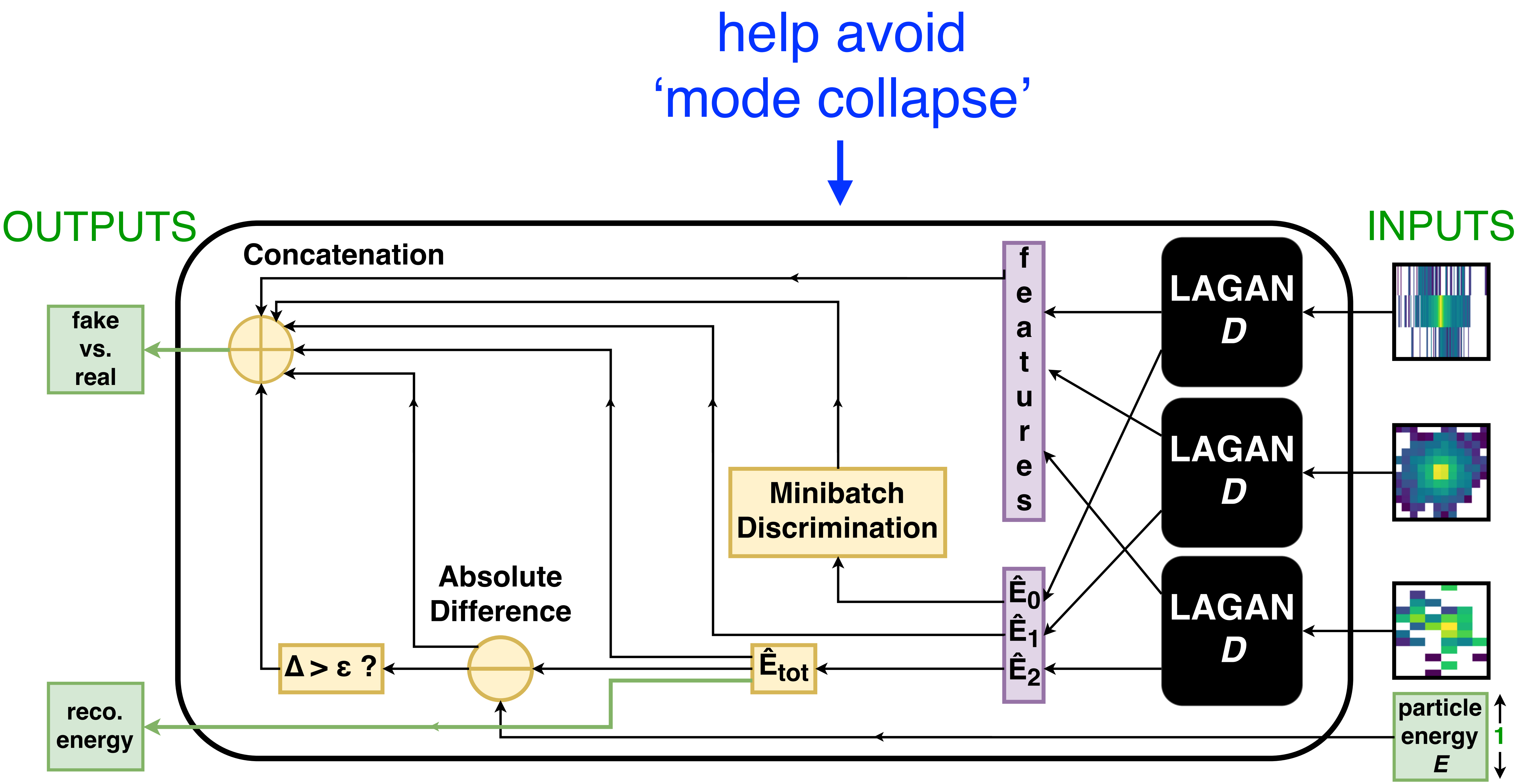
One network per particle type; input particle energy



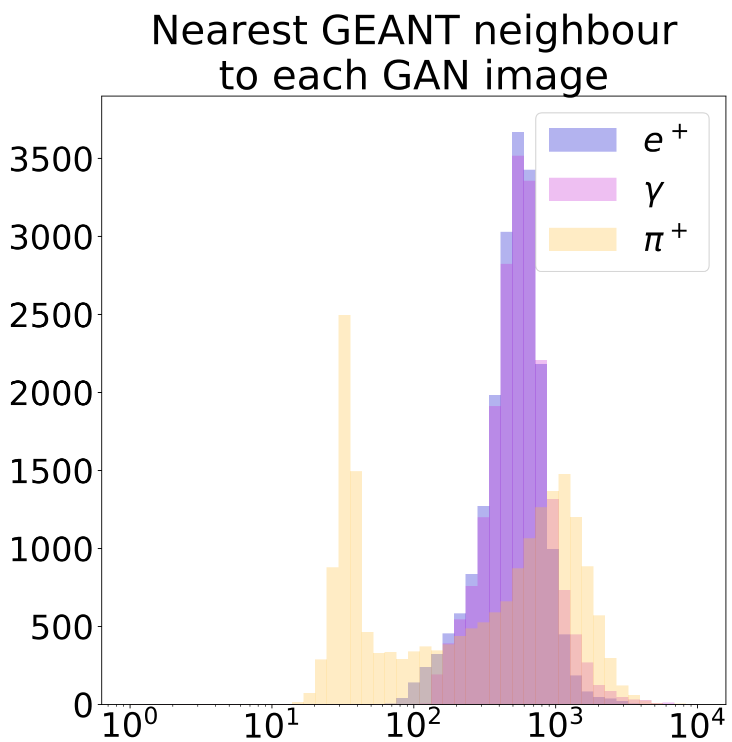
use layer i as input to layer i+1

ReLU to encourage sparsity

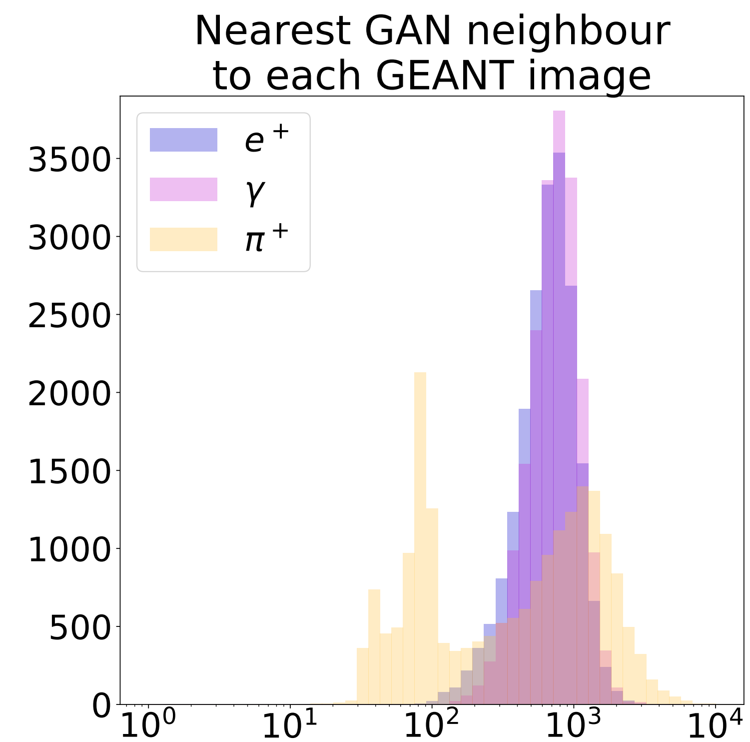
Discriminator Network for CaloGAN



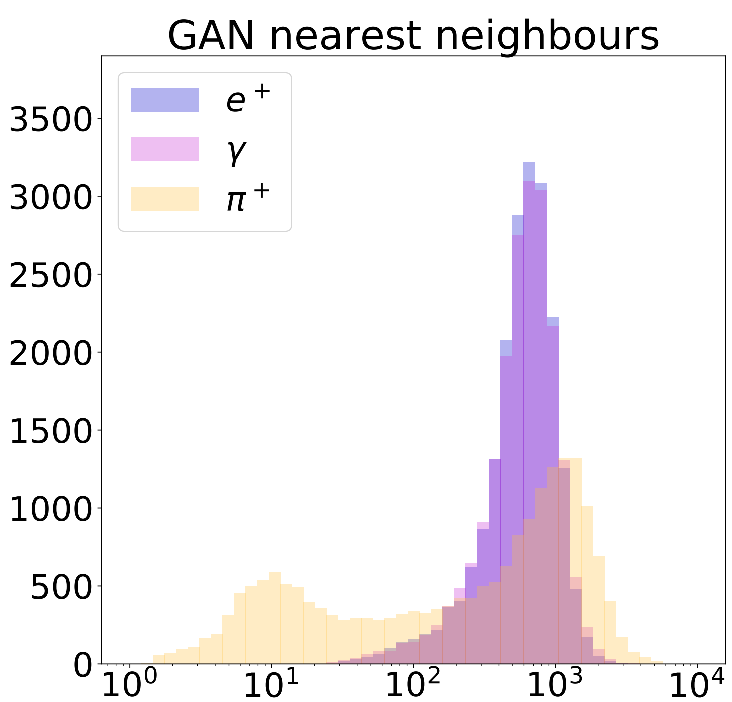
“Overtraining”



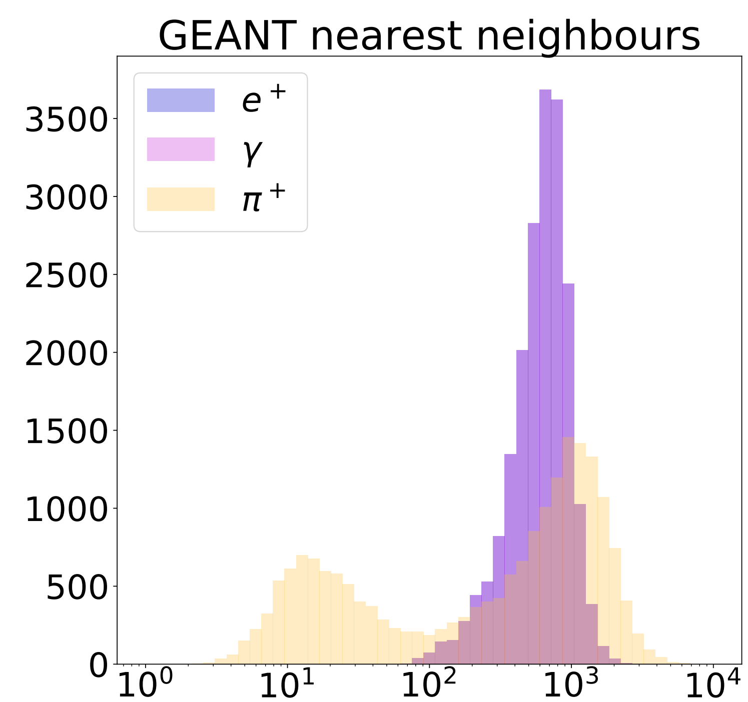
not memorizing



A key challenge in training GANs is the diversity of generated images. This does not seem to be a problem for CaloGAN.



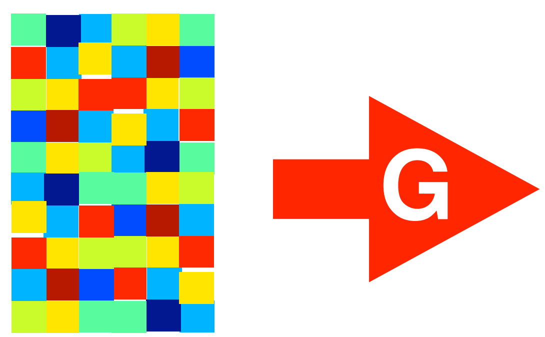
no mode collapse



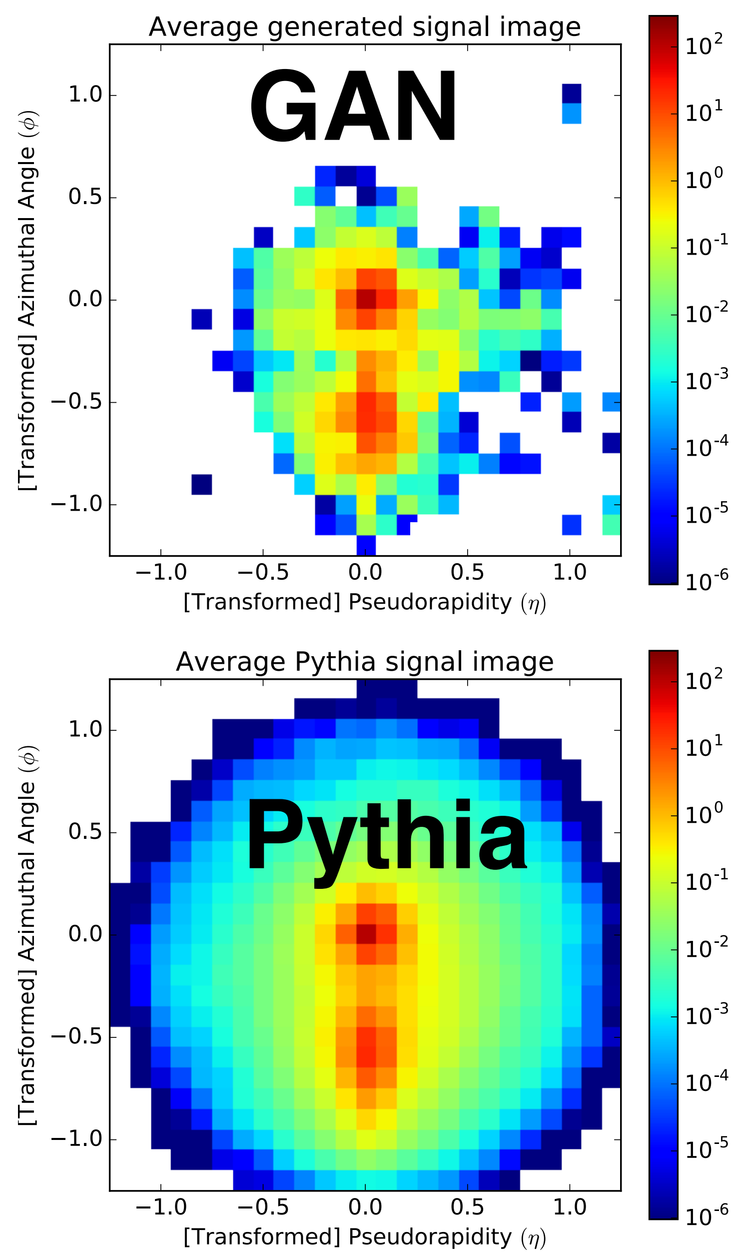
*M. Paganini, L. de Oliveira, and **BPN*** 1705.05927, 1705.02355

Generative Adversarial Networks (GAN):

A two-network game where one maps noise to images and one classifies images as fake or real.



noise

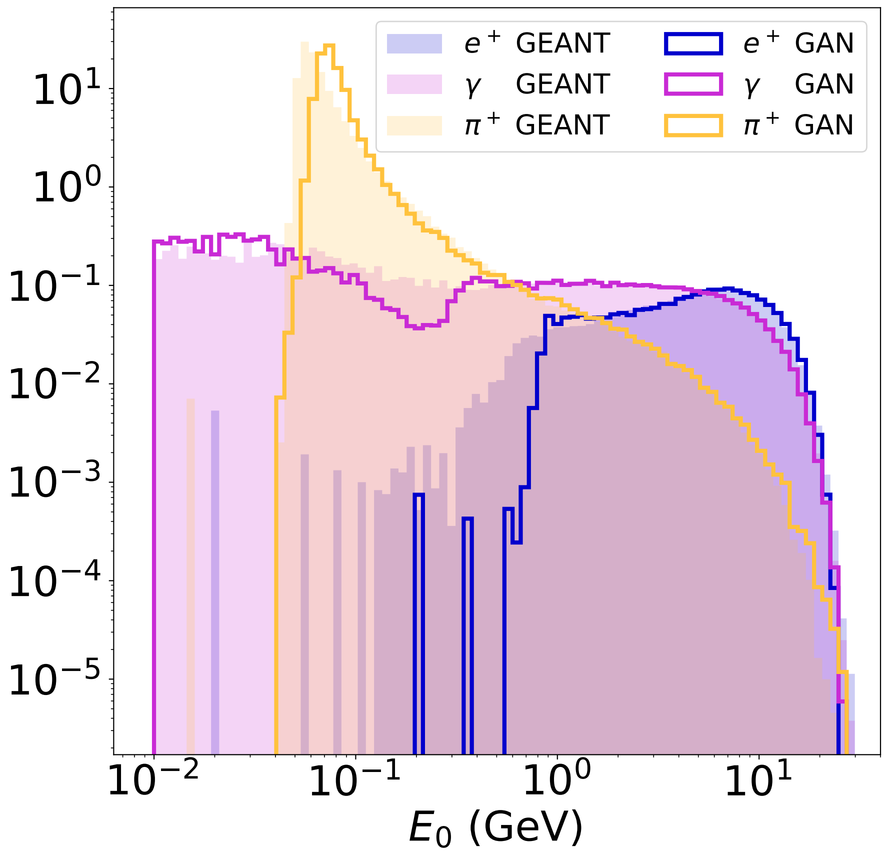


{real,fake}

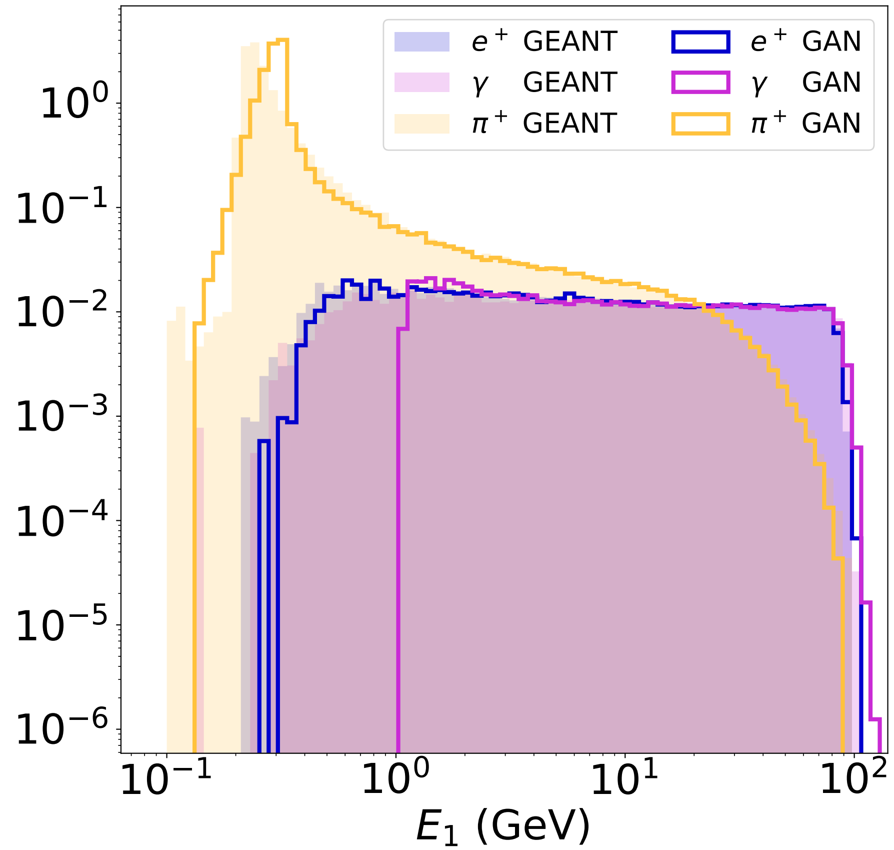
When **D** is maximally confused, **G** will be a good generator

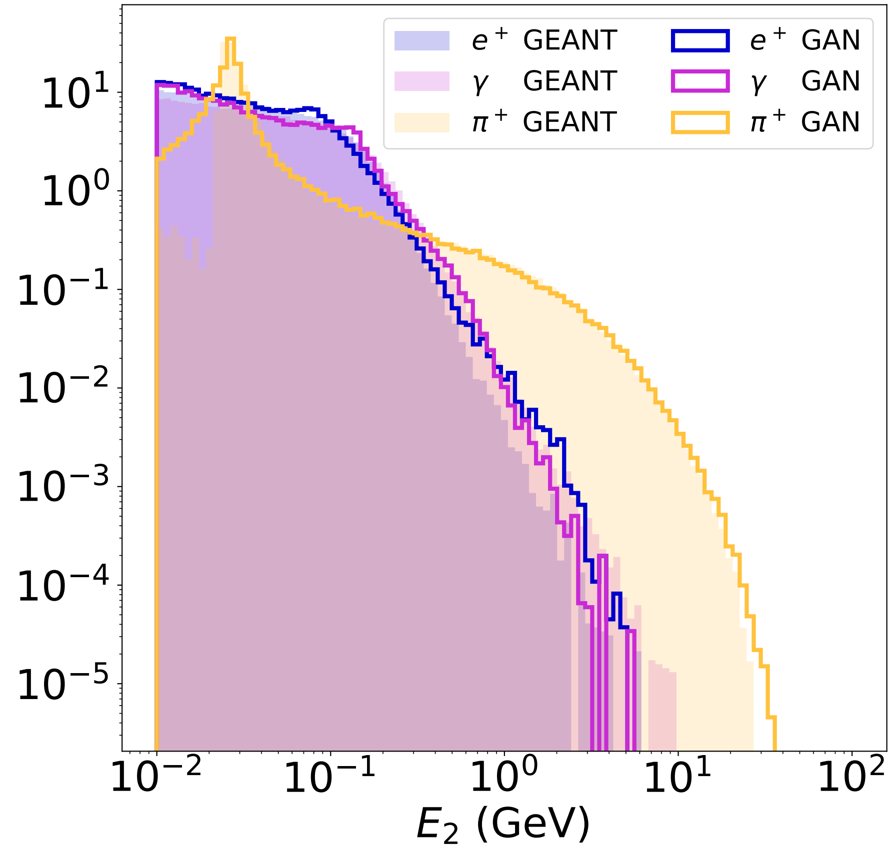
Physics-based simulator

Energy per layer



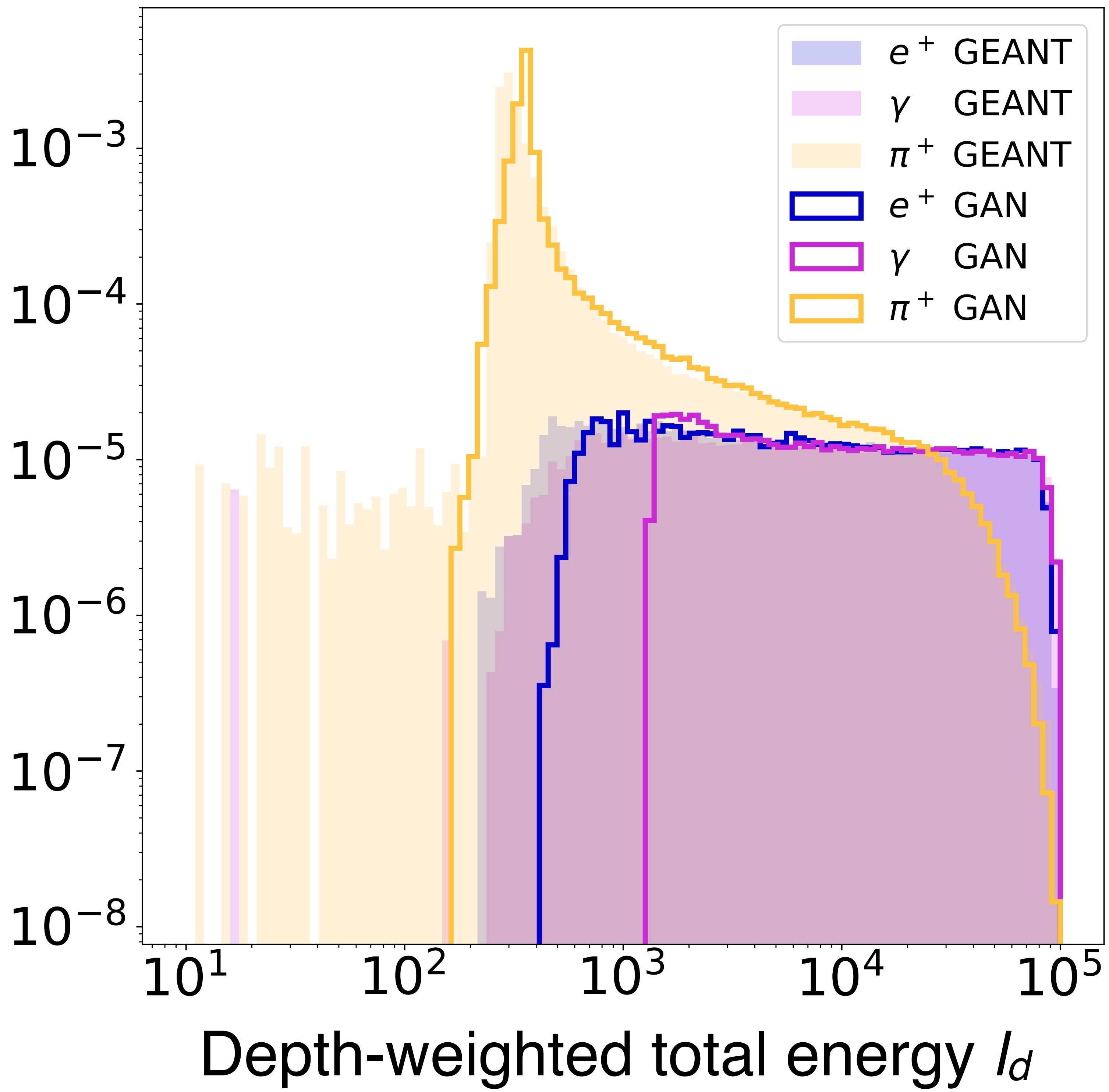
Pions deposit much less energy in the first layers; leave the calorimeter with significant energy



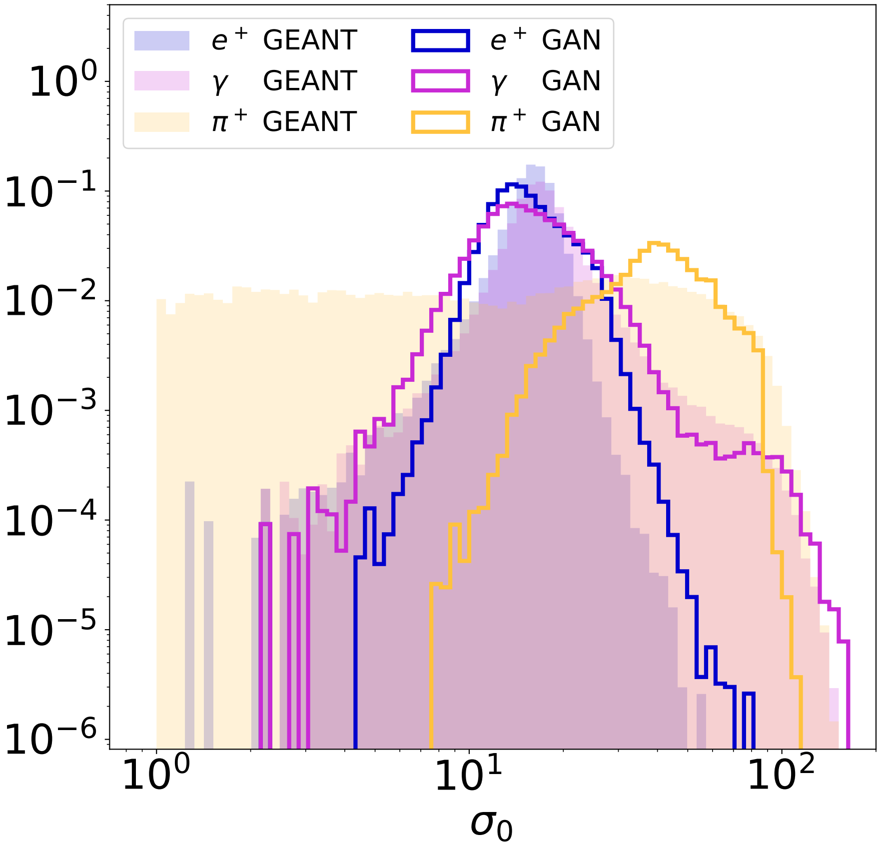


N.B. can always add these (and others) explicitly to the training

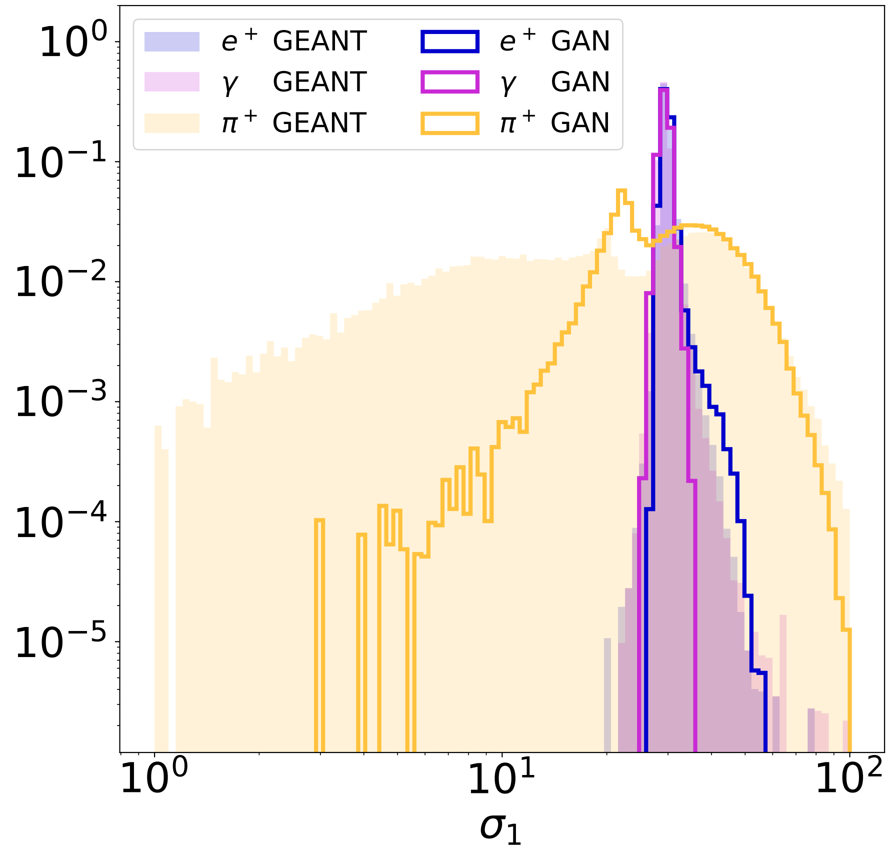
Depth of the shower

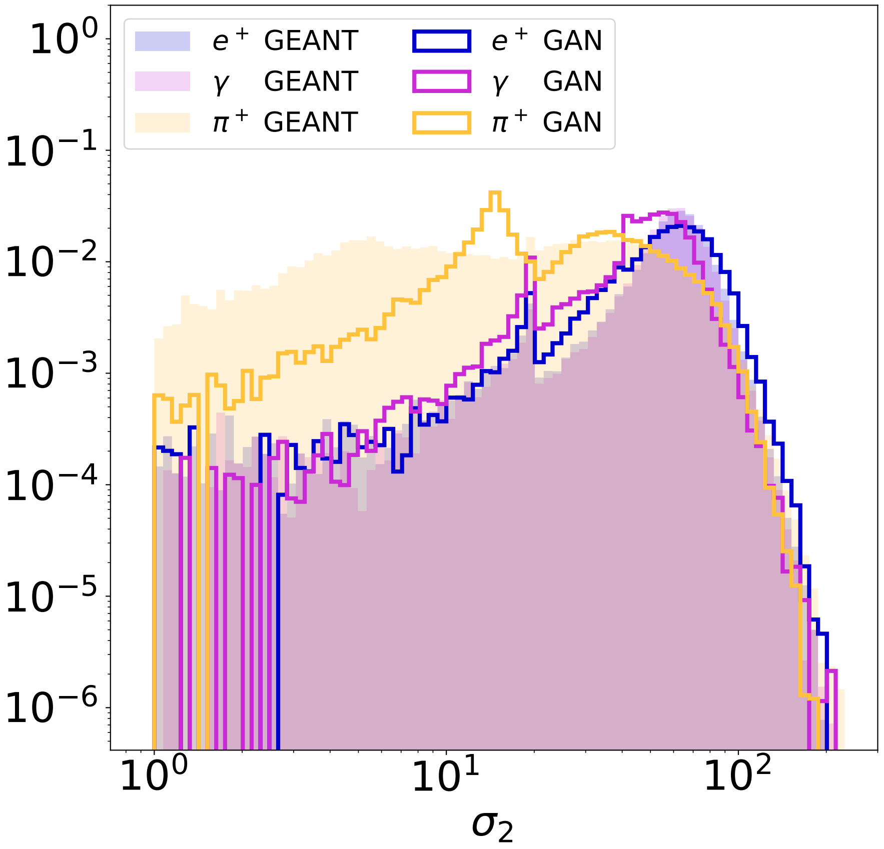


Lateral spread



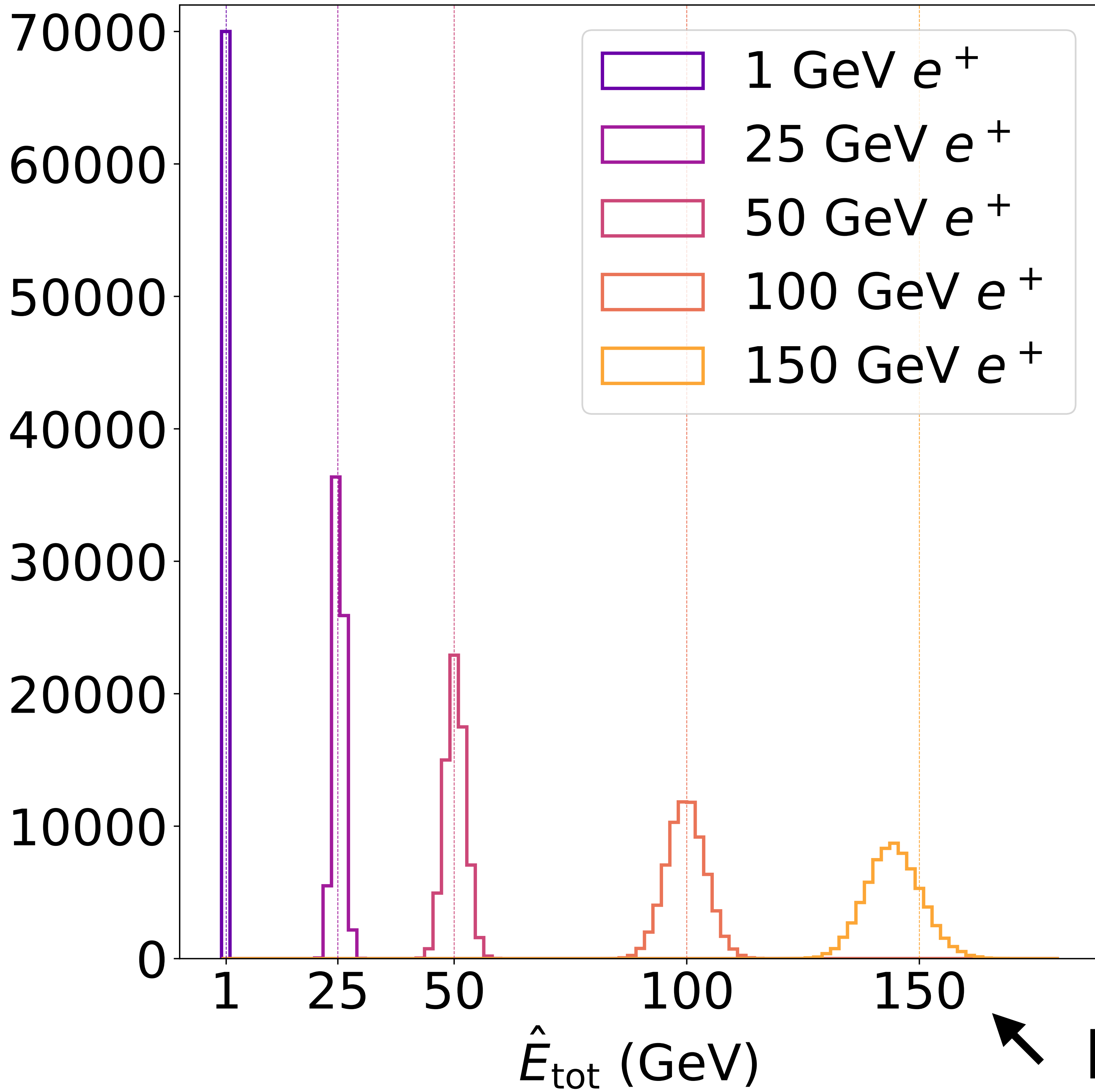
The much larger variation in the pion showers is a challenge for the network.





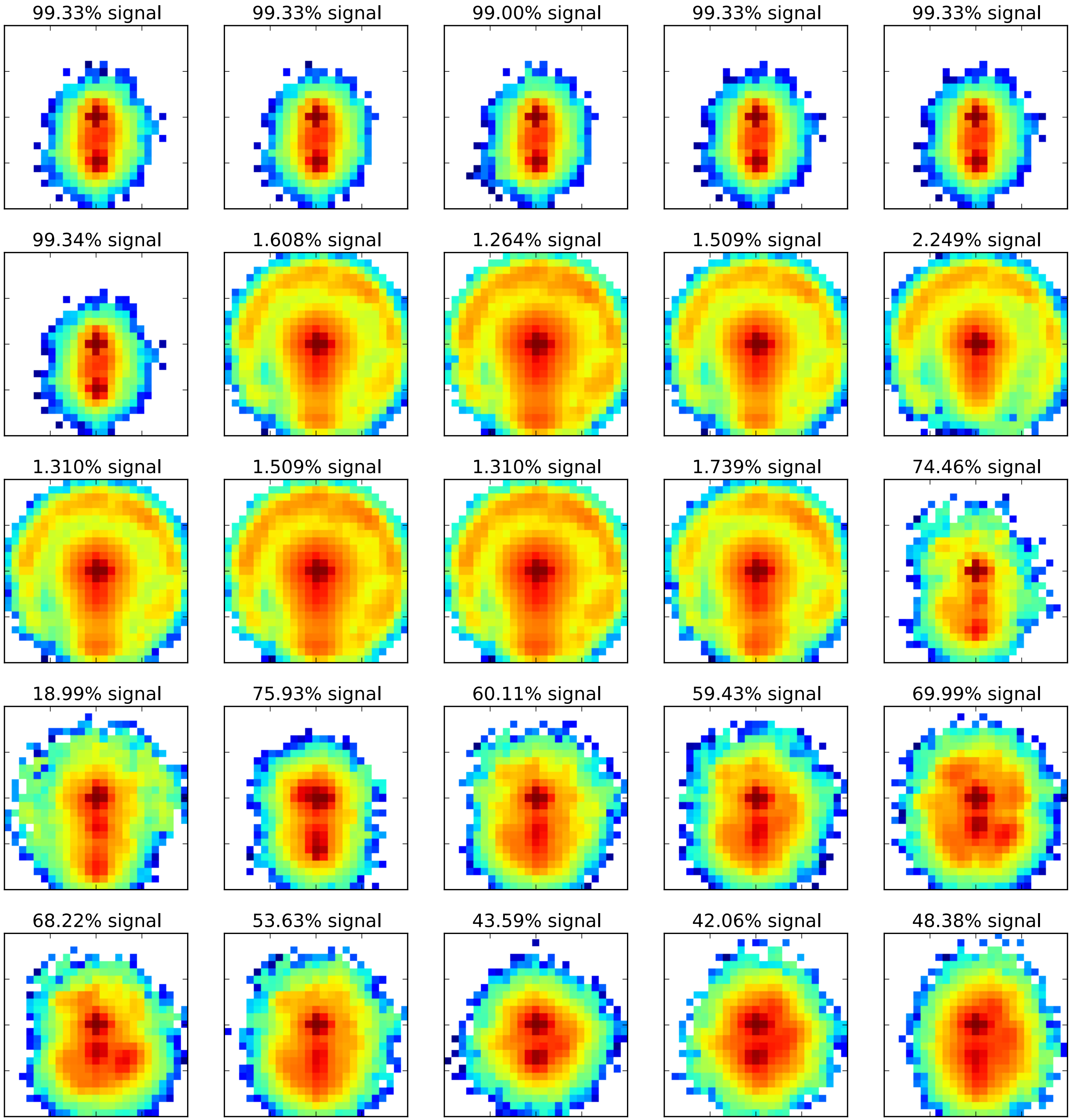
These moments and others are useful for classification; we have also tested this as a metric (NN on 3D images)

Shower Energy



Beyond our training sample!

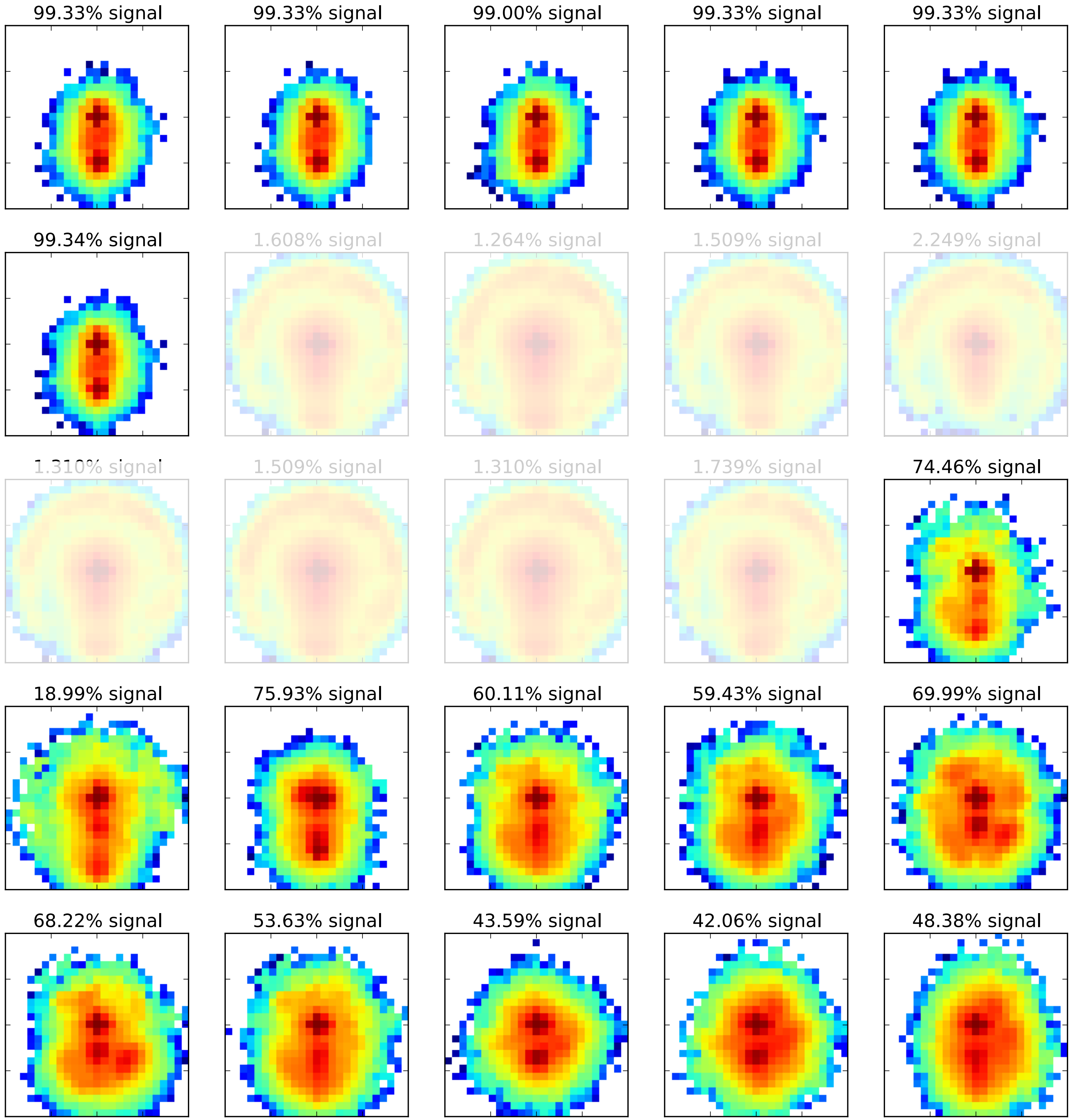
Most activating images



Take a node in the NN and ask which input images activate it the most

Some nodes learn about subjets and some learn about peripheral radiation

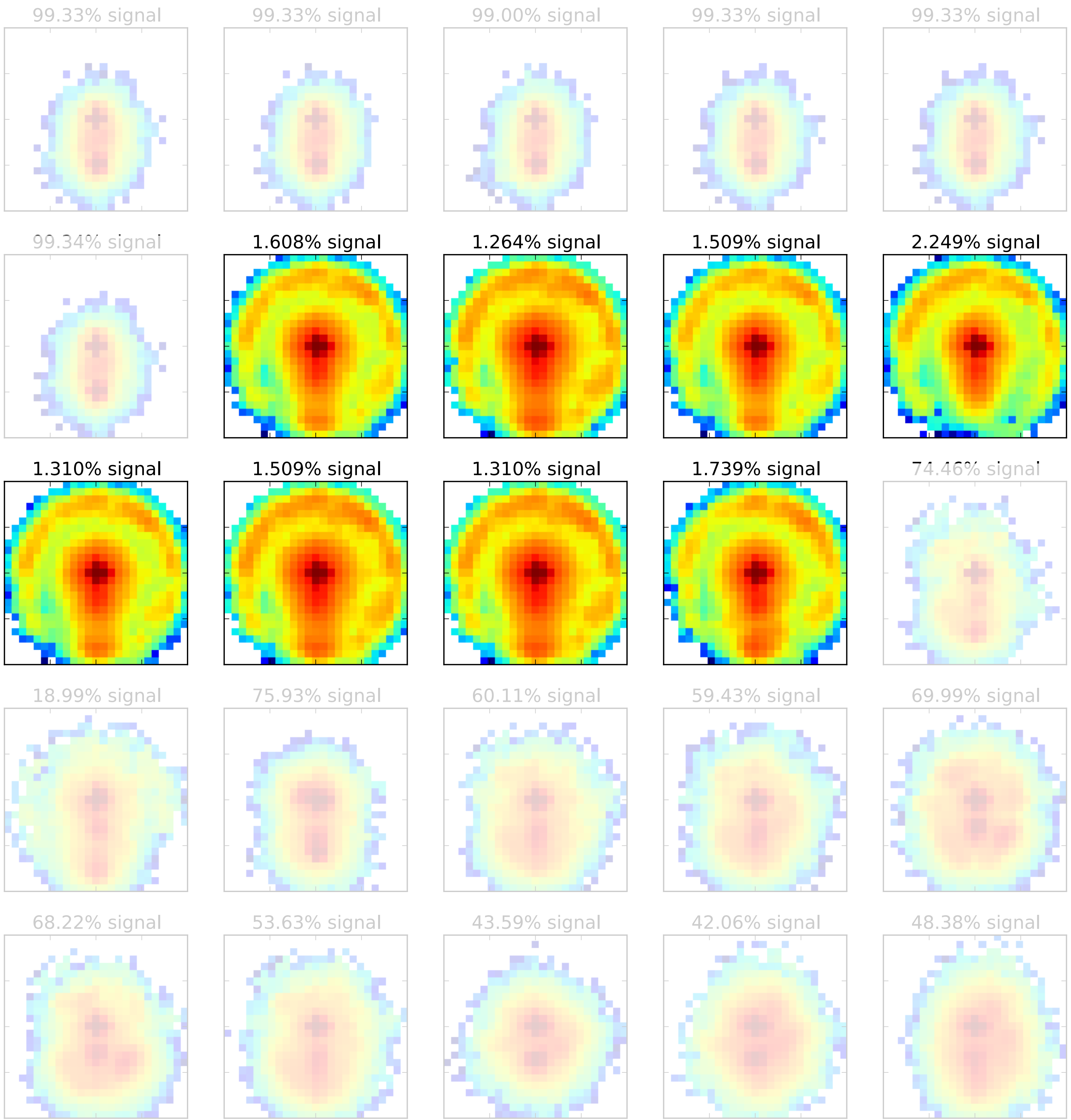
Most activating images



Take a node in the NN and ask which input images activate it the most

Some nodes learn about subjets and some learn about peripheral radiation

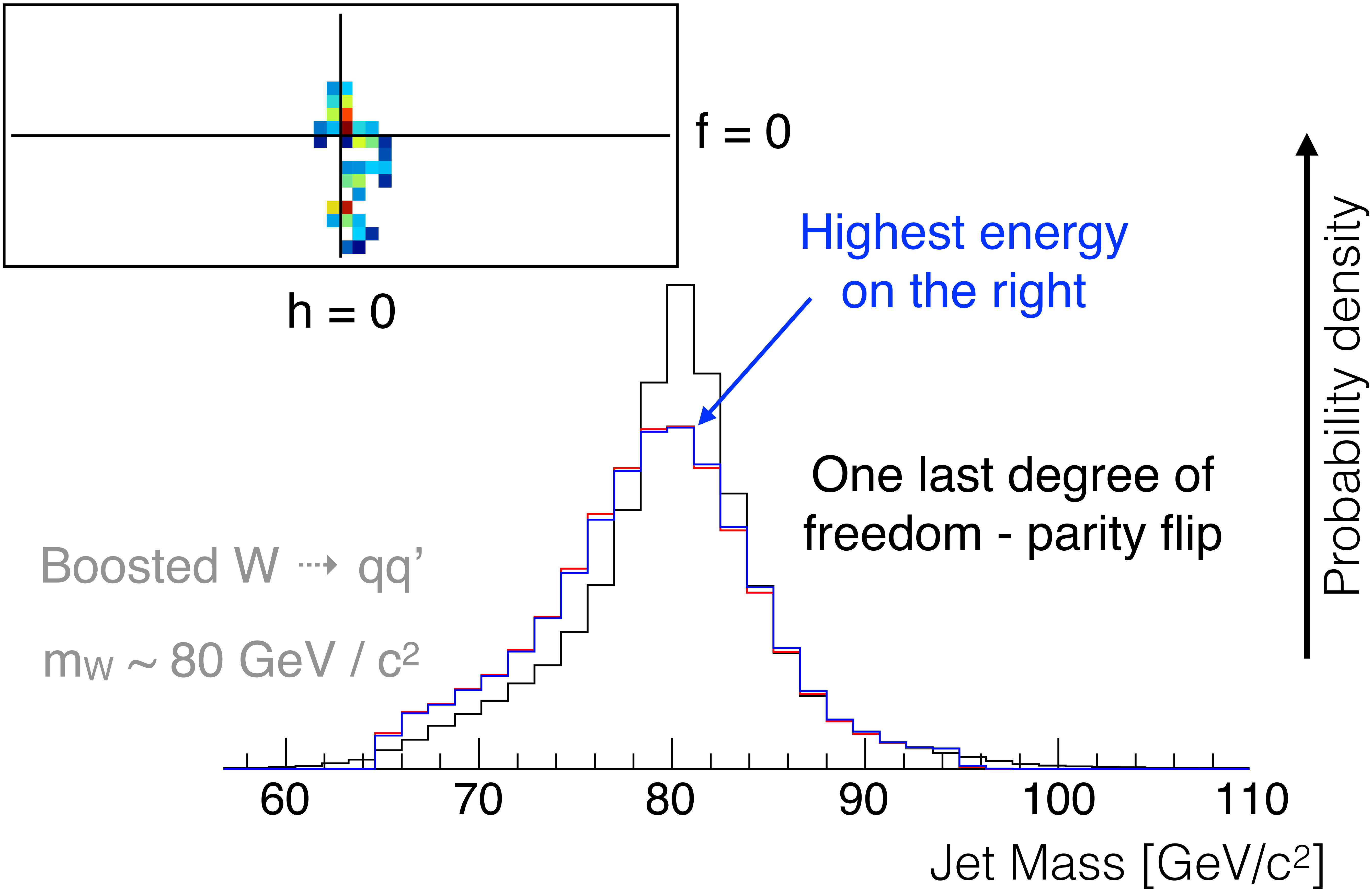
Most activating images



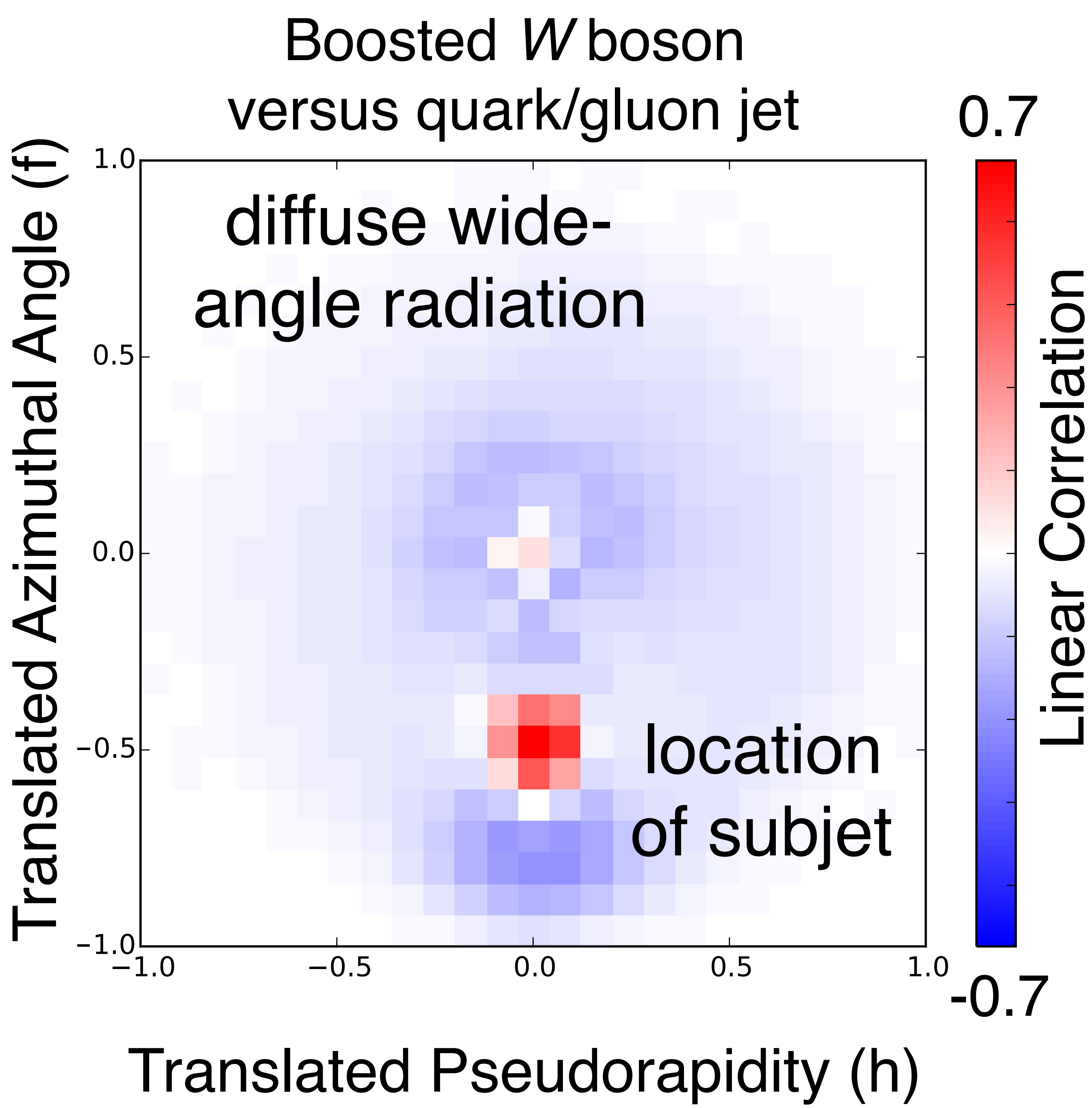
Take a node in the NN and ask which input images activate it the most

Some nodes learn about subjets and some learn about peripheral radiation

Pre-processing & spacetime symmetries



Correlation between input and output



Red = network is more activated (more signal-like)