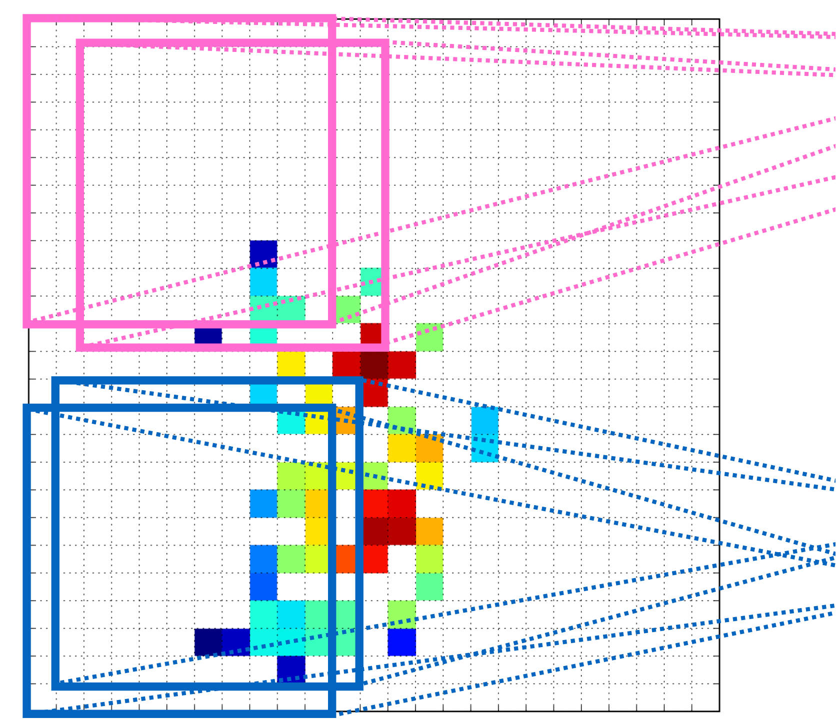
Advancing High Energy Physics with Machine Learning

Benjamin Nachman

Lawrence Berkeley National Laboratory



BERKELEY EXPERIMENTAL PARTICLE PHYSICS

UCR Physics Seminar
January 15, 2019

Outline

The deep learning revolution

Representations + preprocessing

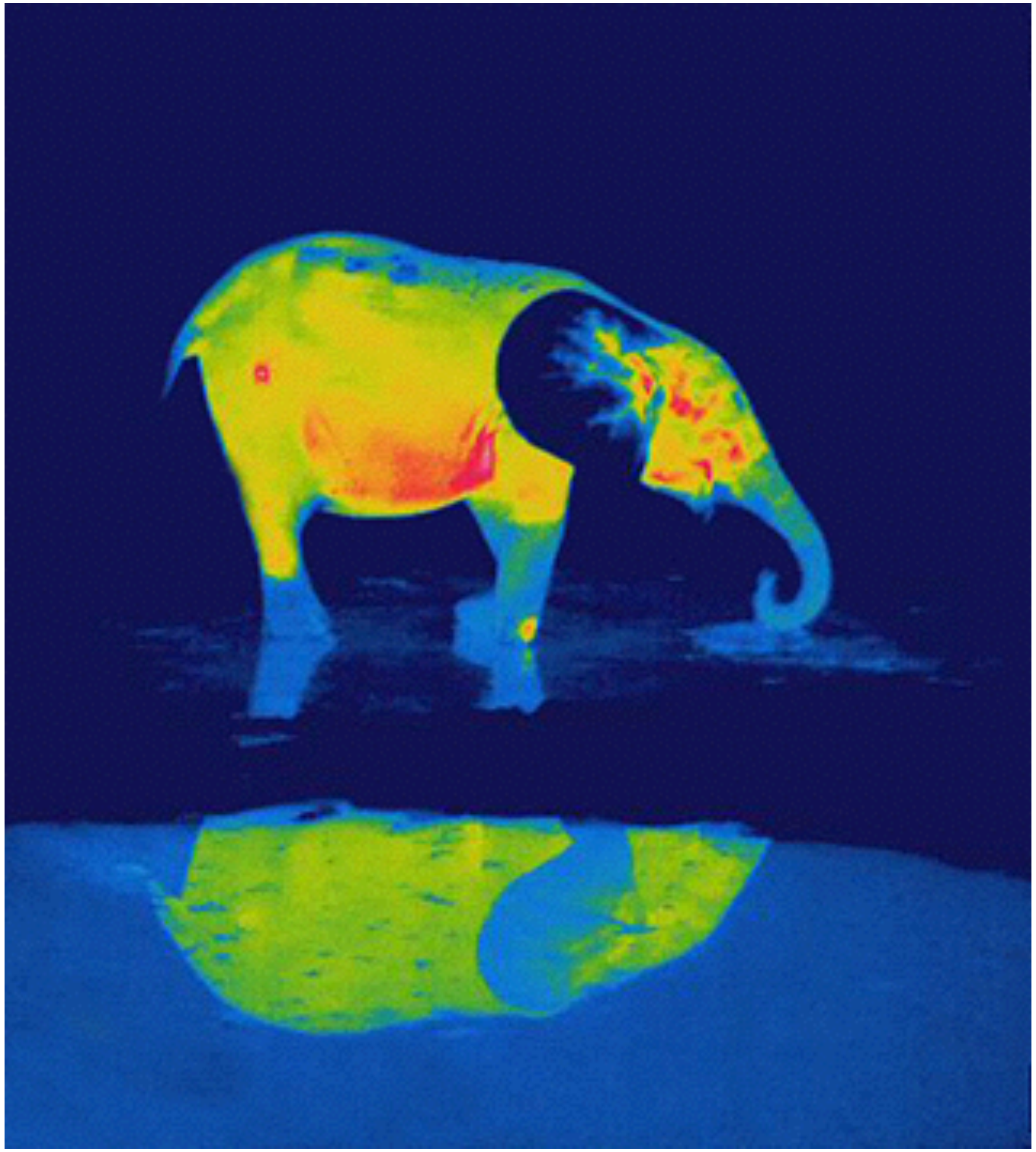
Survey of exciting developments

Classification

Regression

Generation

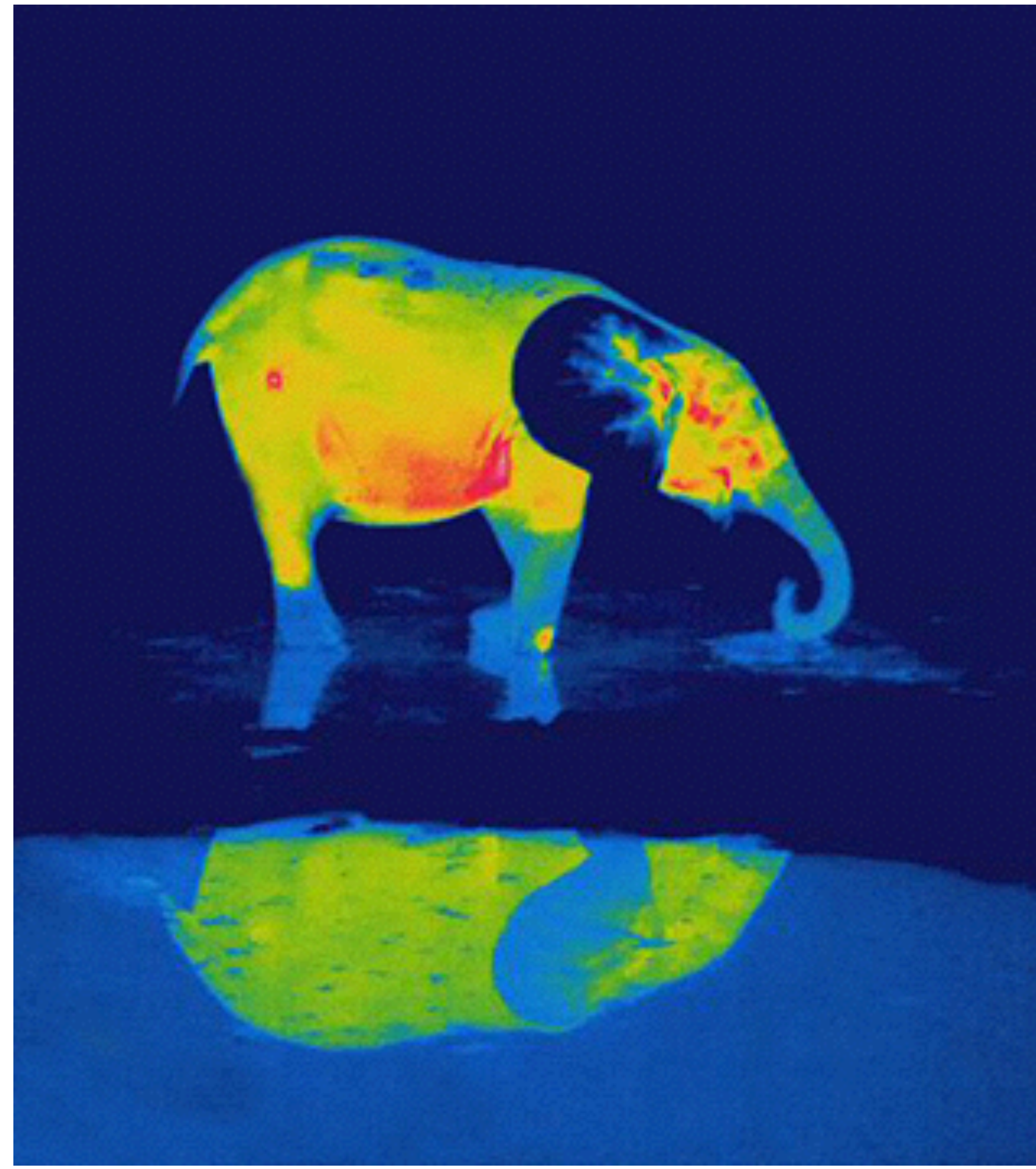
Outlook and conclusions

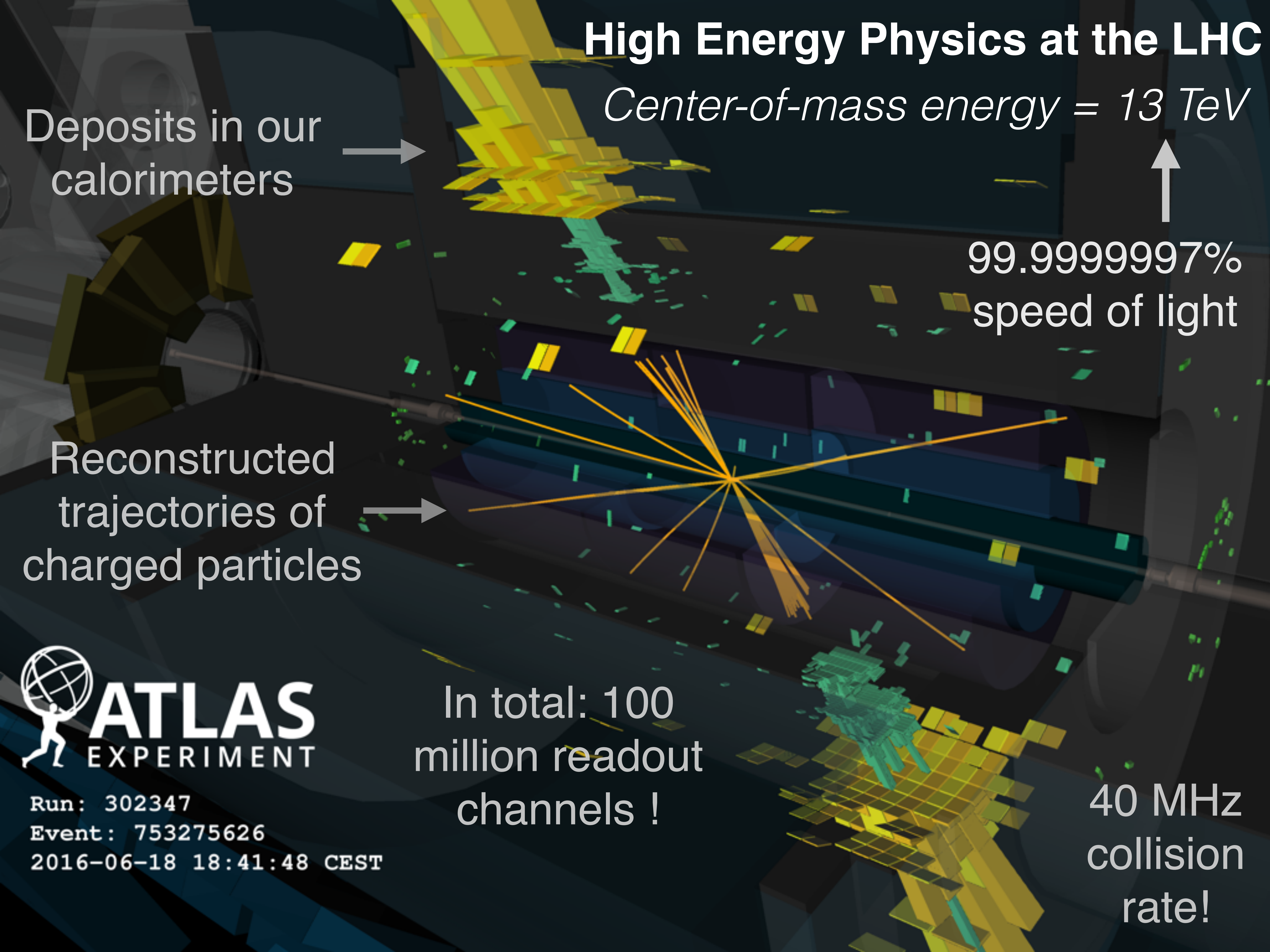


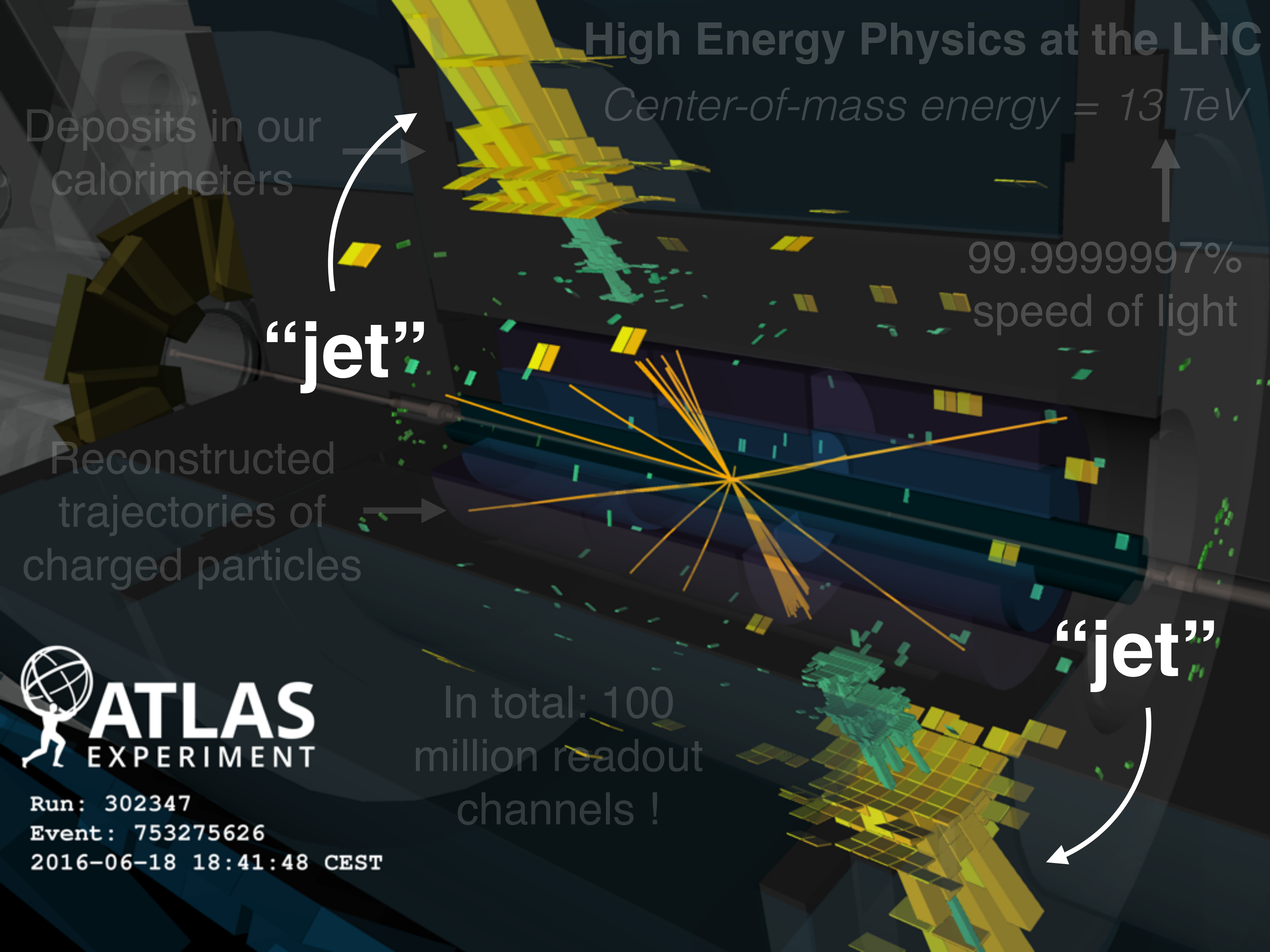
What is new about deep learning?

The HEP community has been conducting “multivariate” analysis of our data for many years.

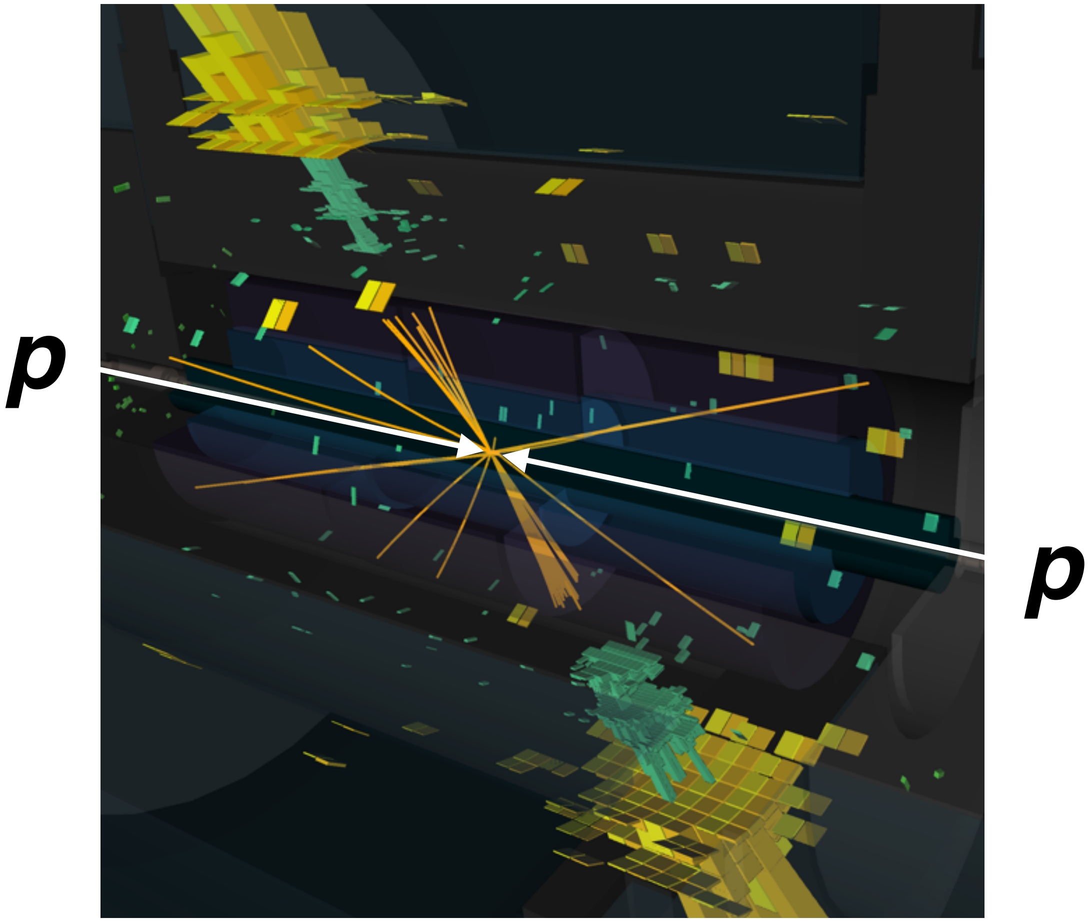
However, recent advances have opened up a **new way** of looking at our data. This **hypervariate vision** will lead to a deeper understanding of nature and perhaps surprises along the way…




High Energy Physics at the LHC
Center-of-mass energy = 13 TeV
Deposits in our calorimeters
99.9999997% speed of light
Reconstructed trajectories of charged particles
ATLAS EXPERIMENT
Run: 302347
Event: 753275626
2016-06-18 18:41:48 CEST
In total: 100 million readout channels !
40 MHz collision rate!

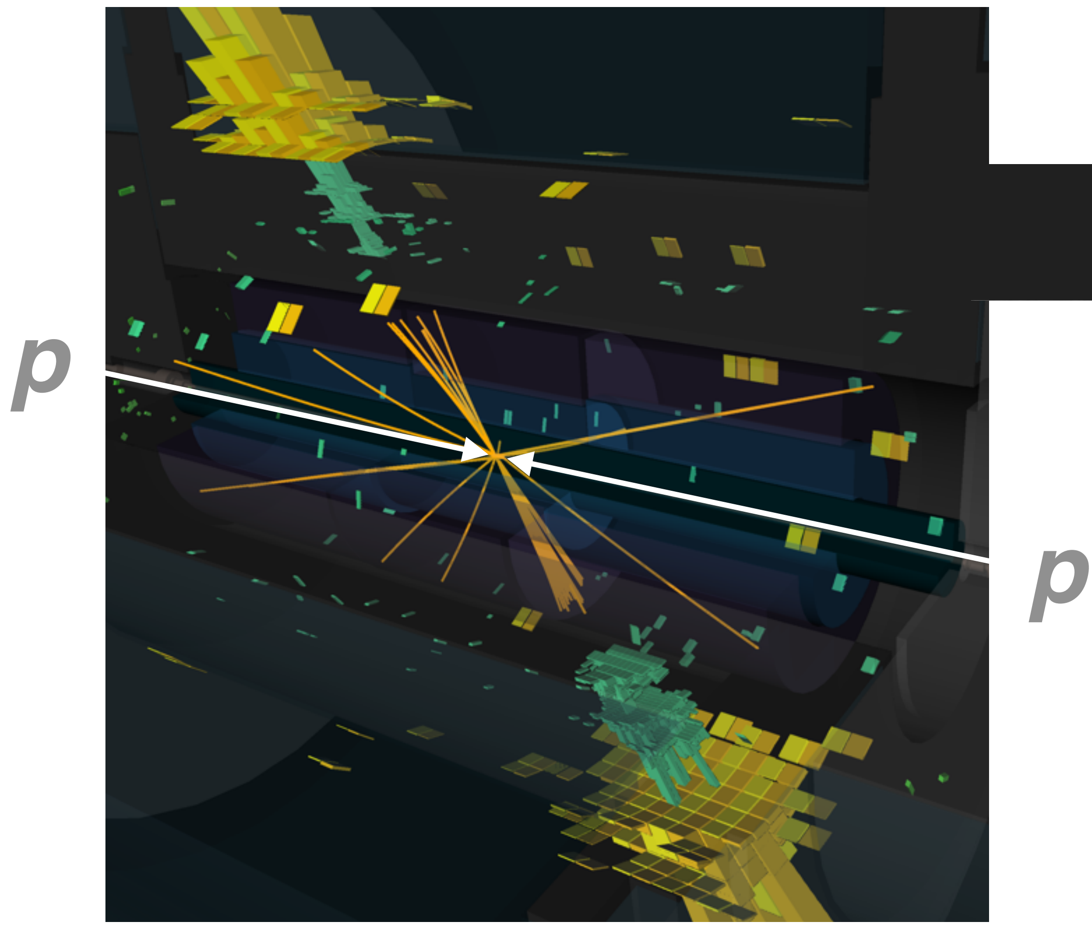

High Energy Physics at the LHC
Center-of-mass energy = 13 TeV
99.9999997%
speed of light
Deposits in our
calorimeters
"jet"
Reconstructed
trajectories of
charged particles
"jet"
In total: 100
million readout
channels !
ATLAS
EXPERIMENT
Run: 302347
Event: 753275626
2016-06-18 18:41:48 CEST

Picking a representation of the data



Yellow, **green** = active pixels

Picking a representation of the data



Yellow, **green** = active pixels

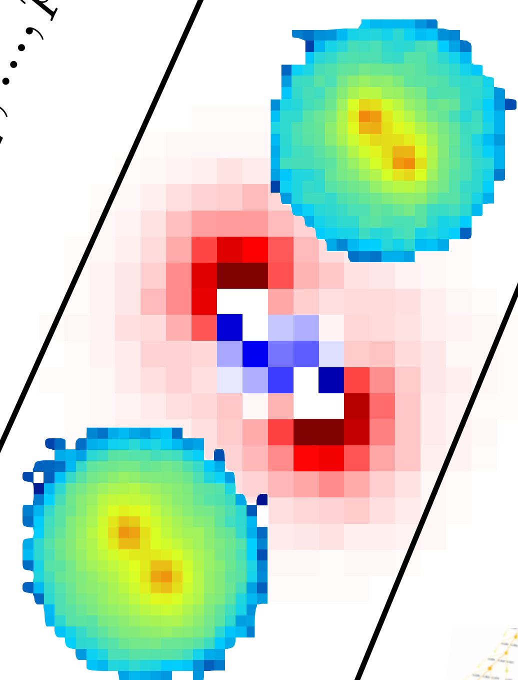
Think of events as e.g. an image

Sets

$$J = \{p_1^\mu, p_2^\mu, \ldots, p_n^\mu\}$$

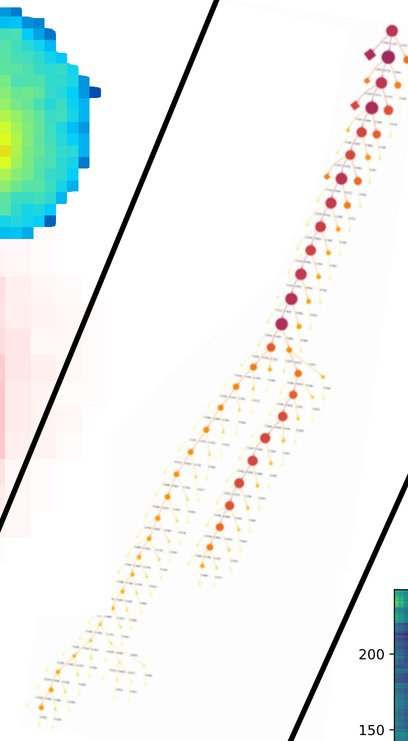
L. de Oliveira,
M. Kagan, L. Mackey,
BPN, A. Schwartzman,
JHEP 07 (2016) 069

Images



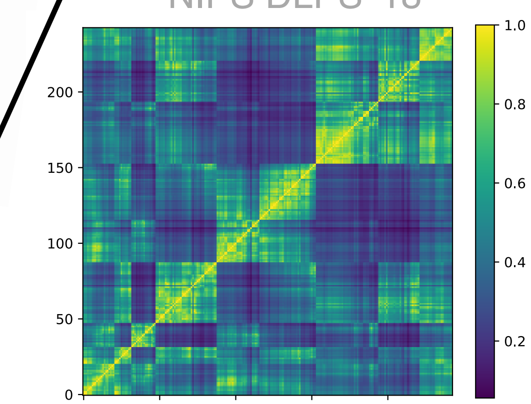
G. Louppe,
K. Cho, C. Becot,
K. Cranmer,
1702.00748

Sequences/ Trees

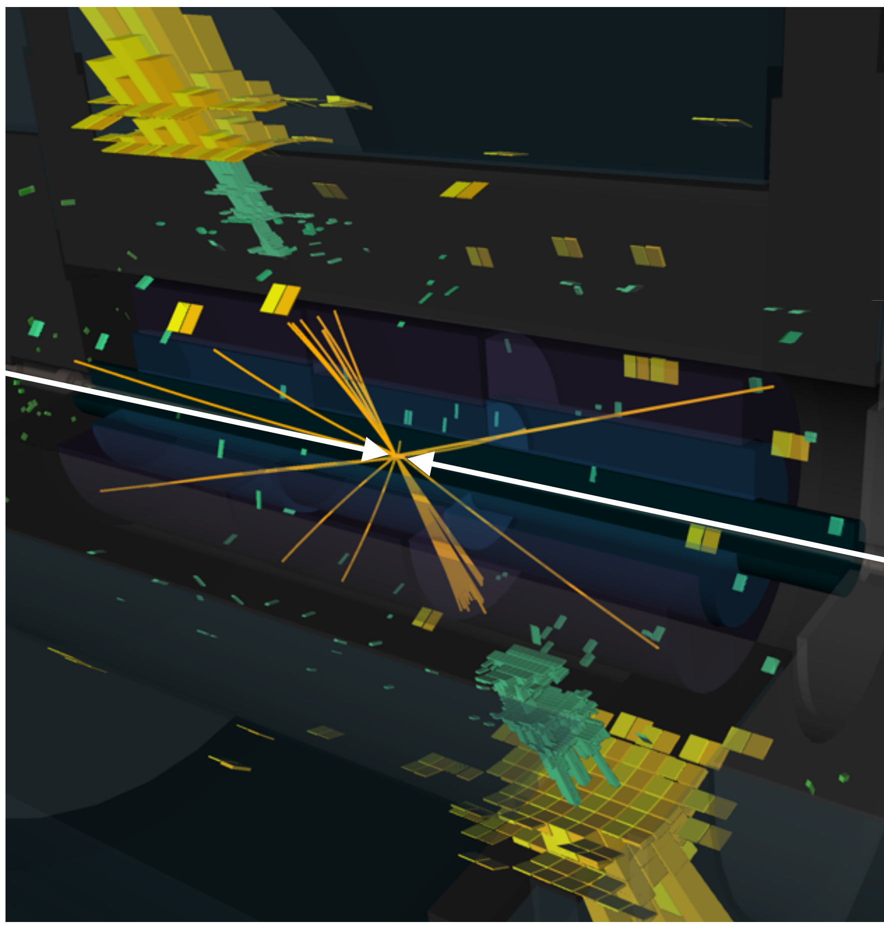


Graphs

I. Henrion, et al.,
NIPS DLPS '18



Picking a representation of the data



Yellow, **green** = active pixels

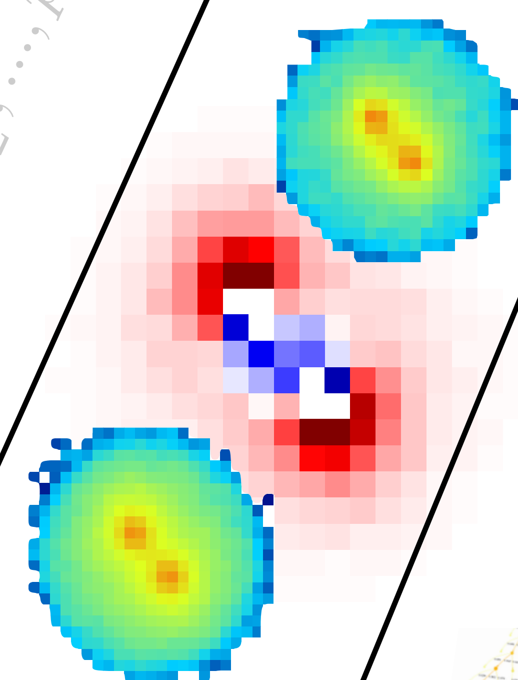
Think of events as e.g. an image

Sets

$J = \{p_1^\mu, p_2^\mu, \ldots, p_n^\mu\}$

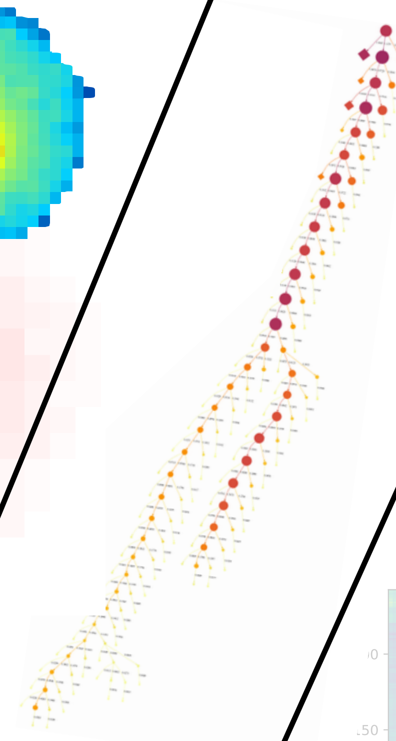
L. de Oliveira,
M. Kagan, L. Mackey,
BPN, A. Schwartzman,
JHEP 07 (2016) 069

Images



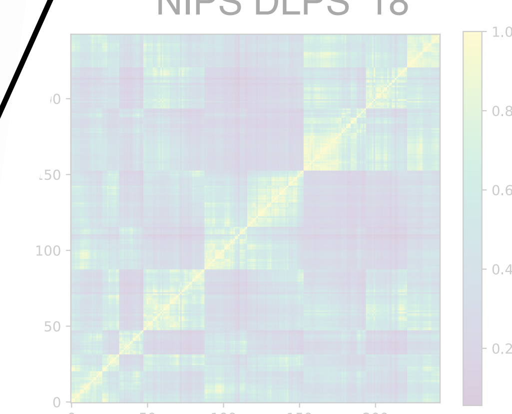
G. Louppe,
K. Cho, C. Becot,
K. Cranmer,
1702.00748

Sequences/ Trees



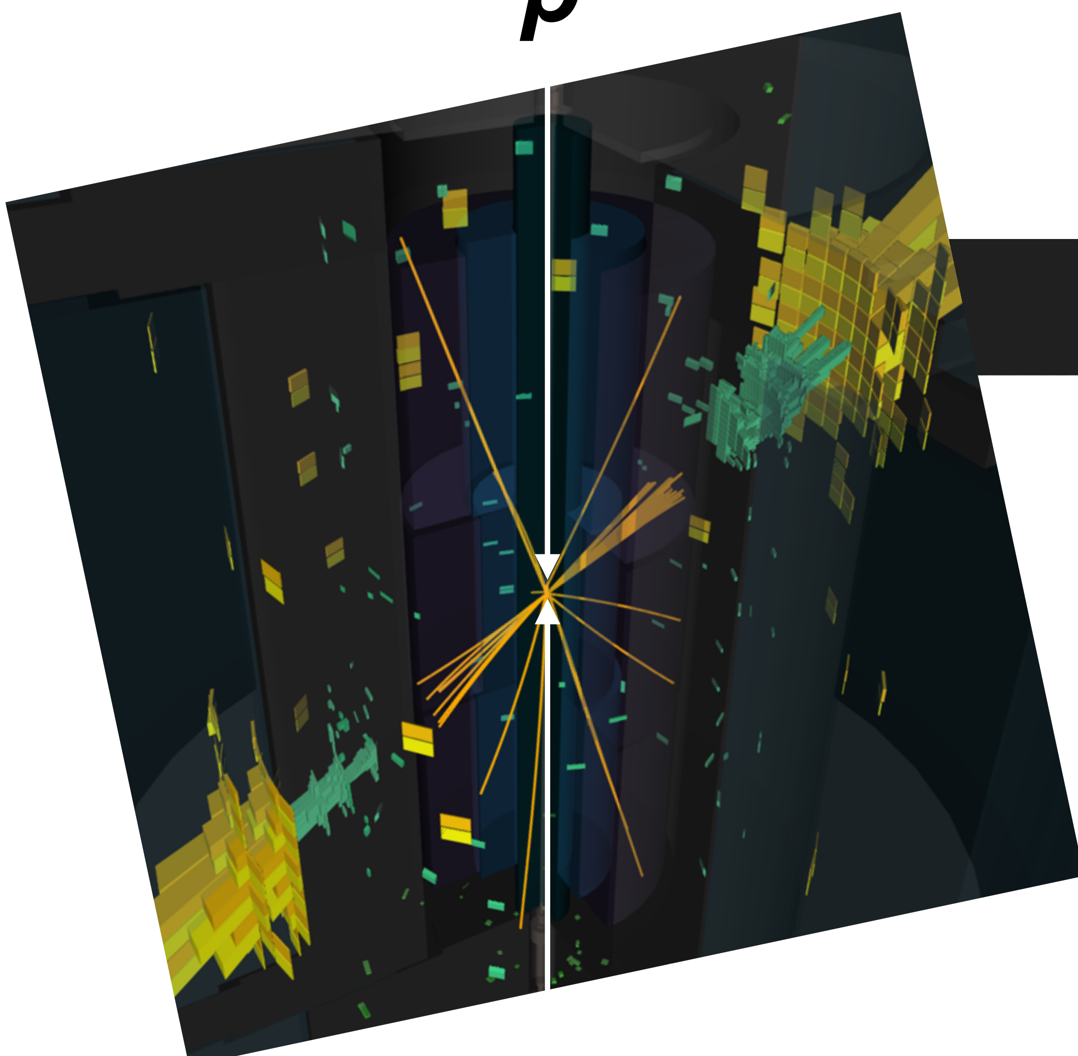
Graphs

I. Henrion, et al.,
NIPS DLPS '18



Picking a representation of the data

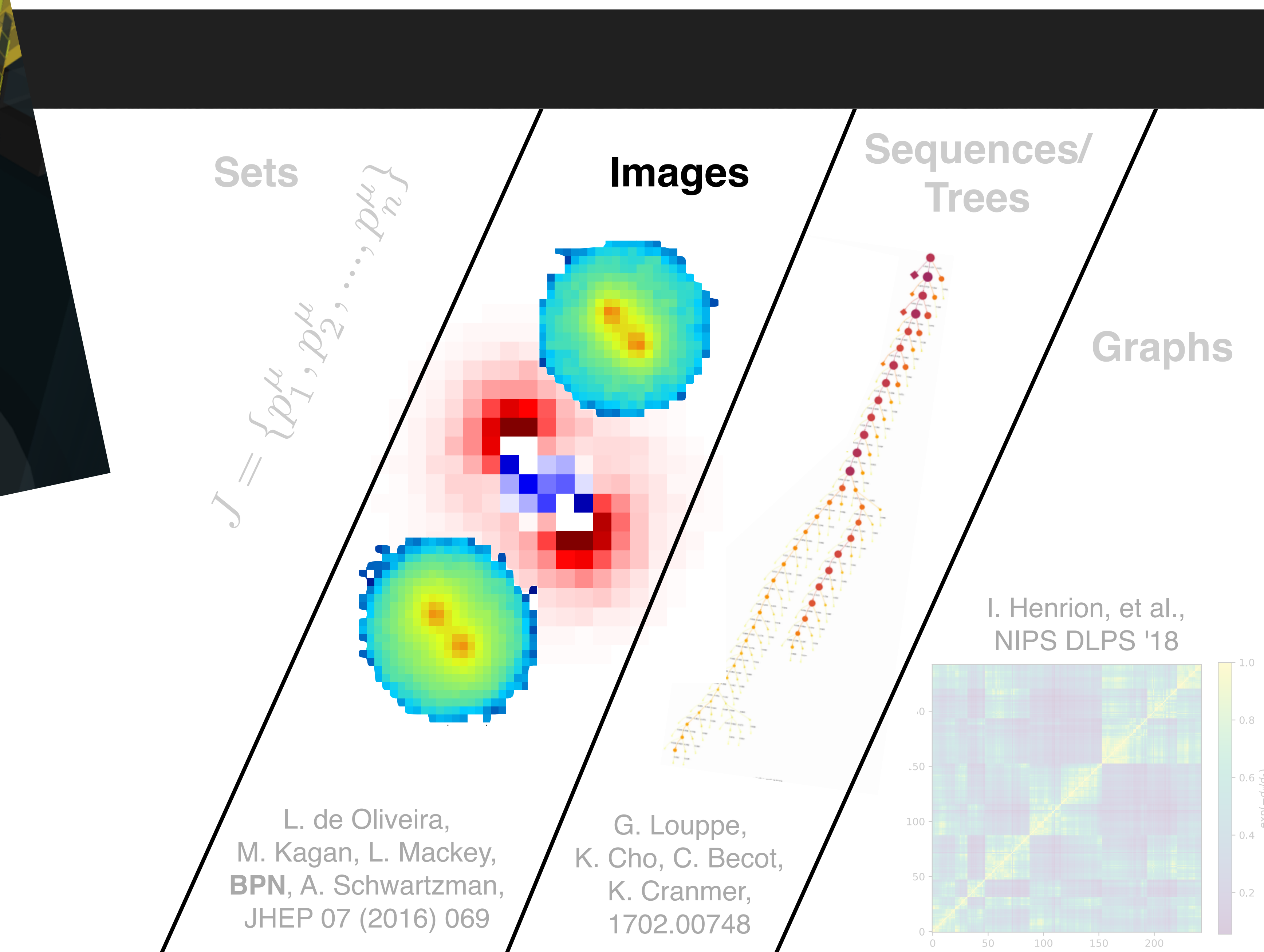
p



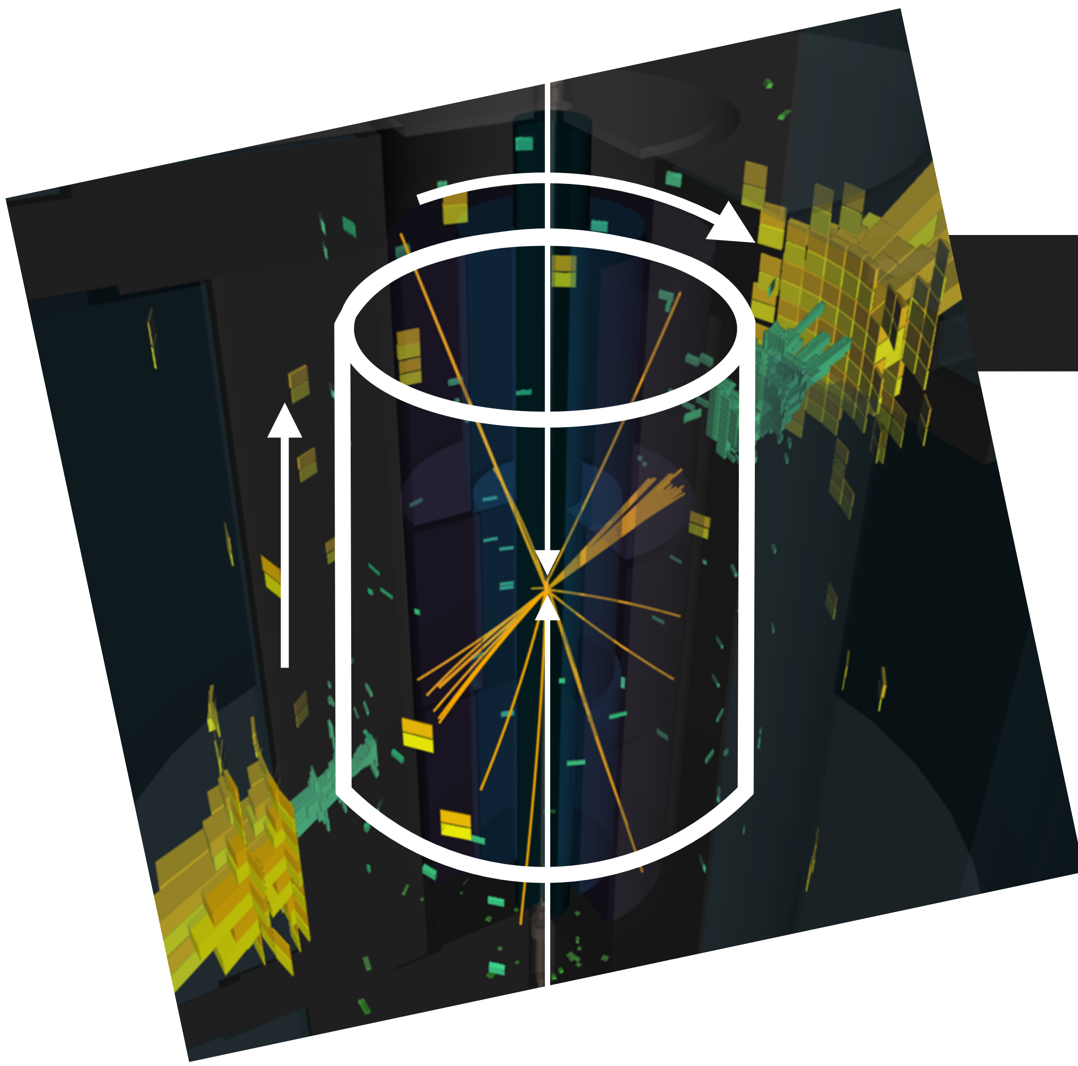
p

Yellow, **green**
= active pixels

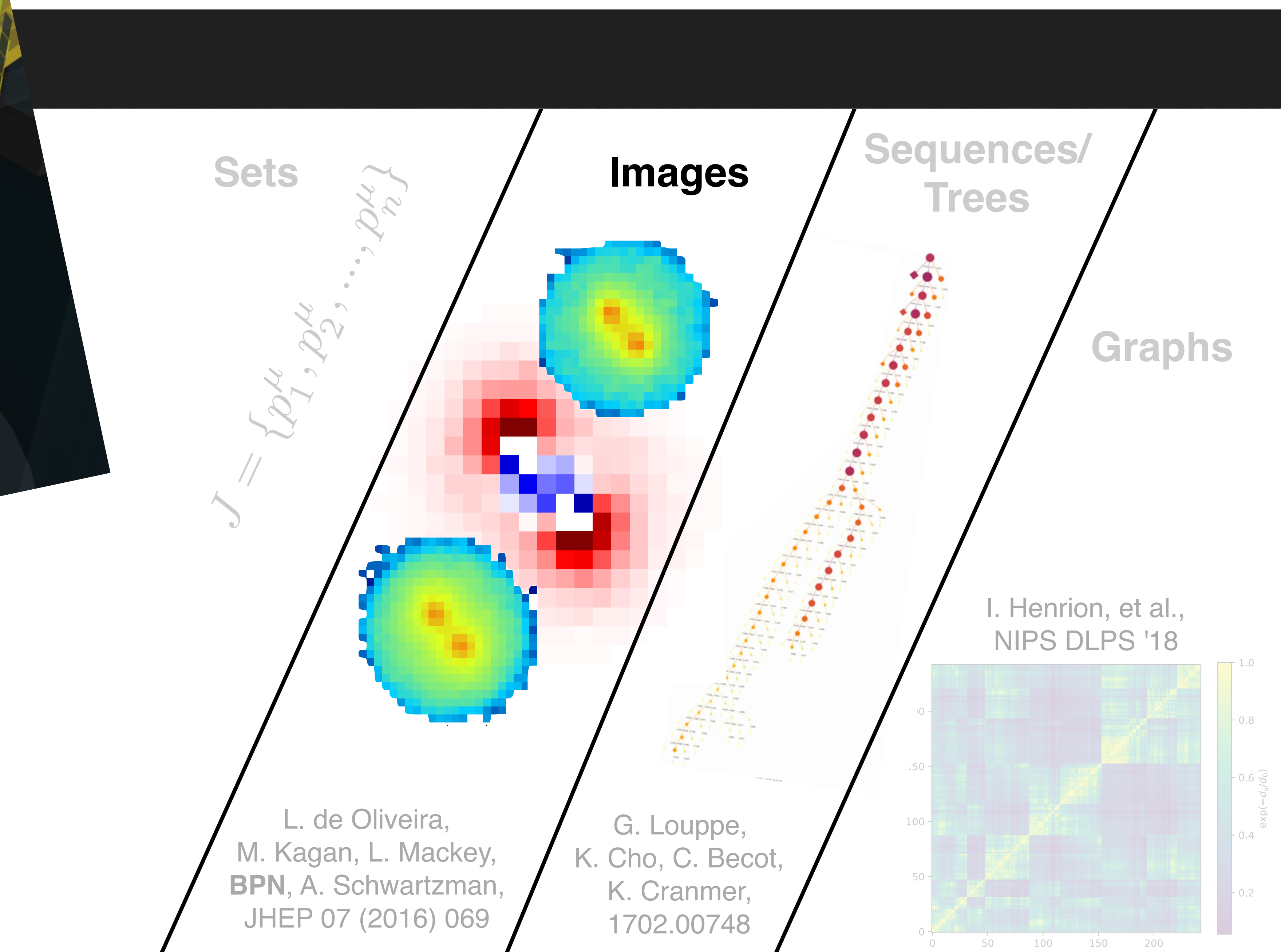
Think of events as e.g. an image



Picking a representation of the data

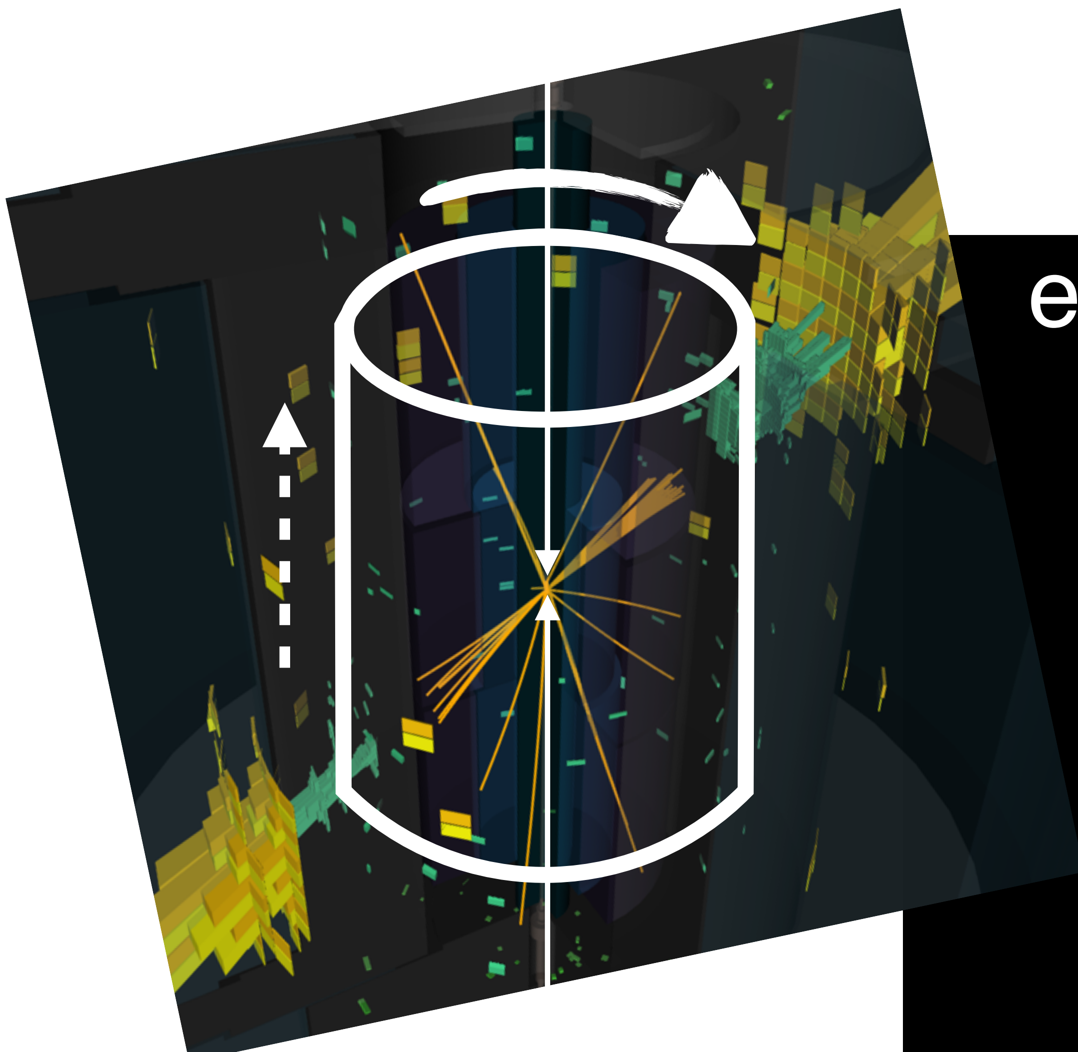


Think of events as e.g. an image

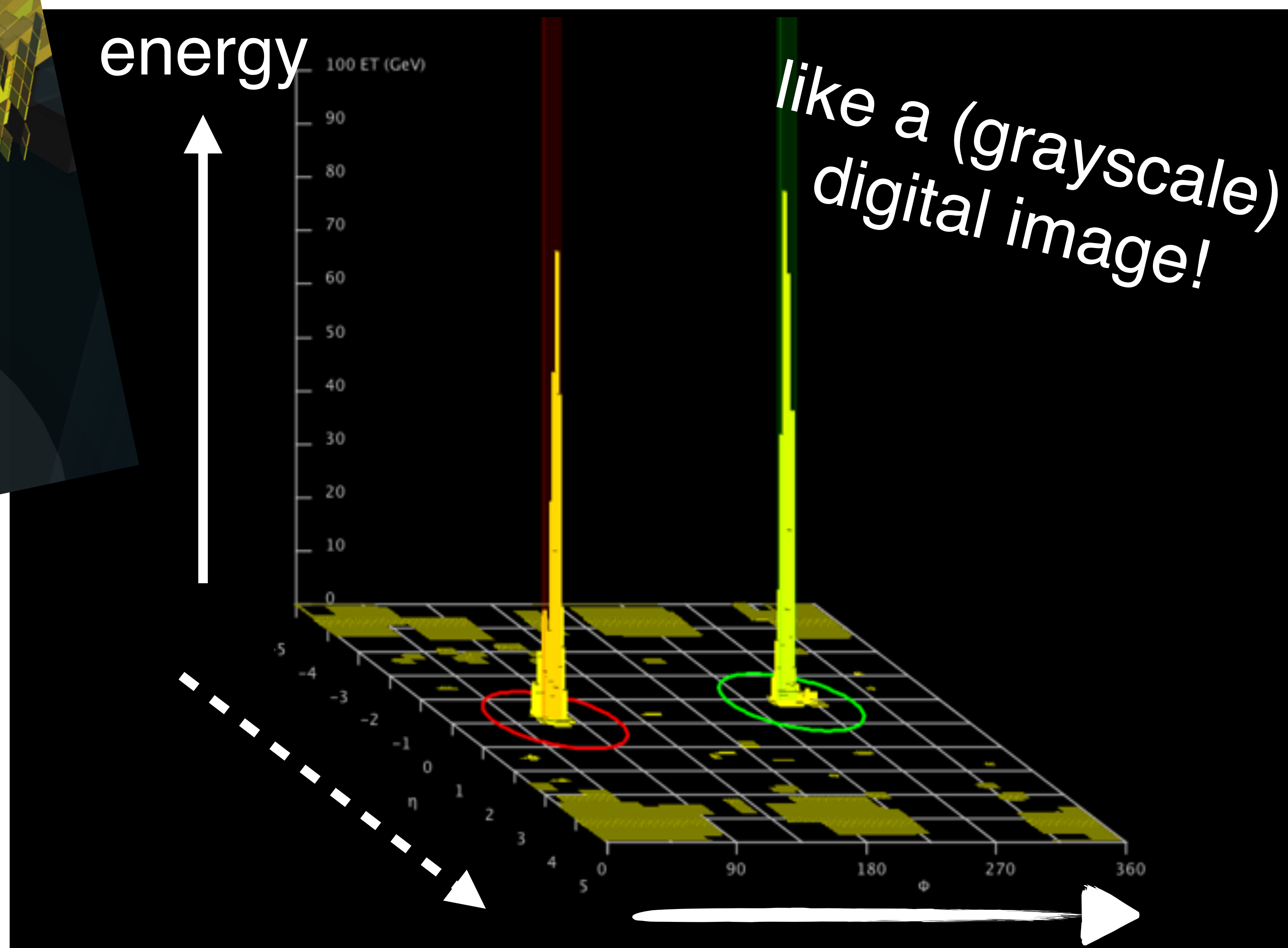


Yellow, **green**
= active pixels

Picking a representation of the data



Think of events as e.g. an image

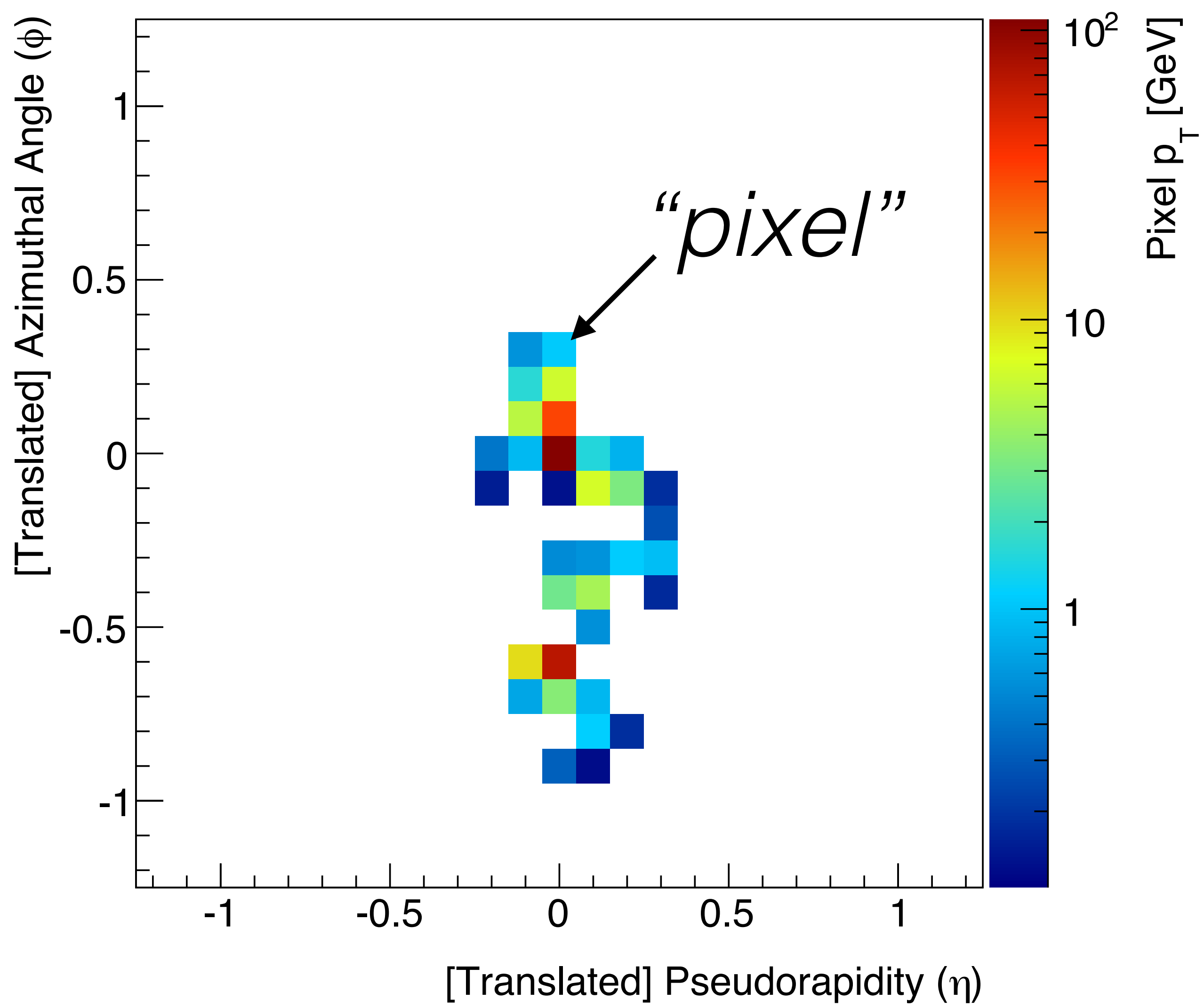


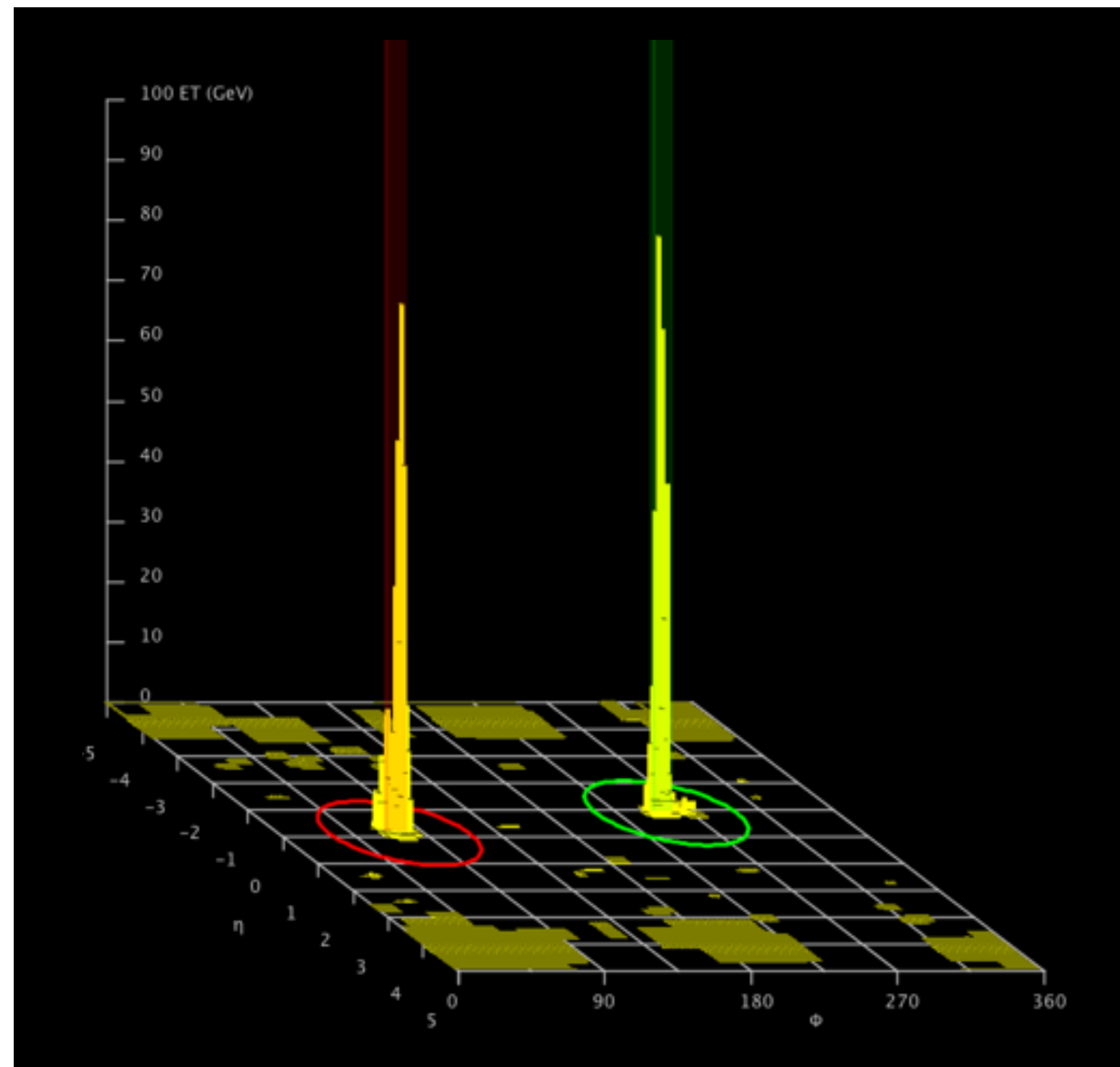
Yellow, **green**
= active pixels

Jet Images

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

Jet from a boosted W boson



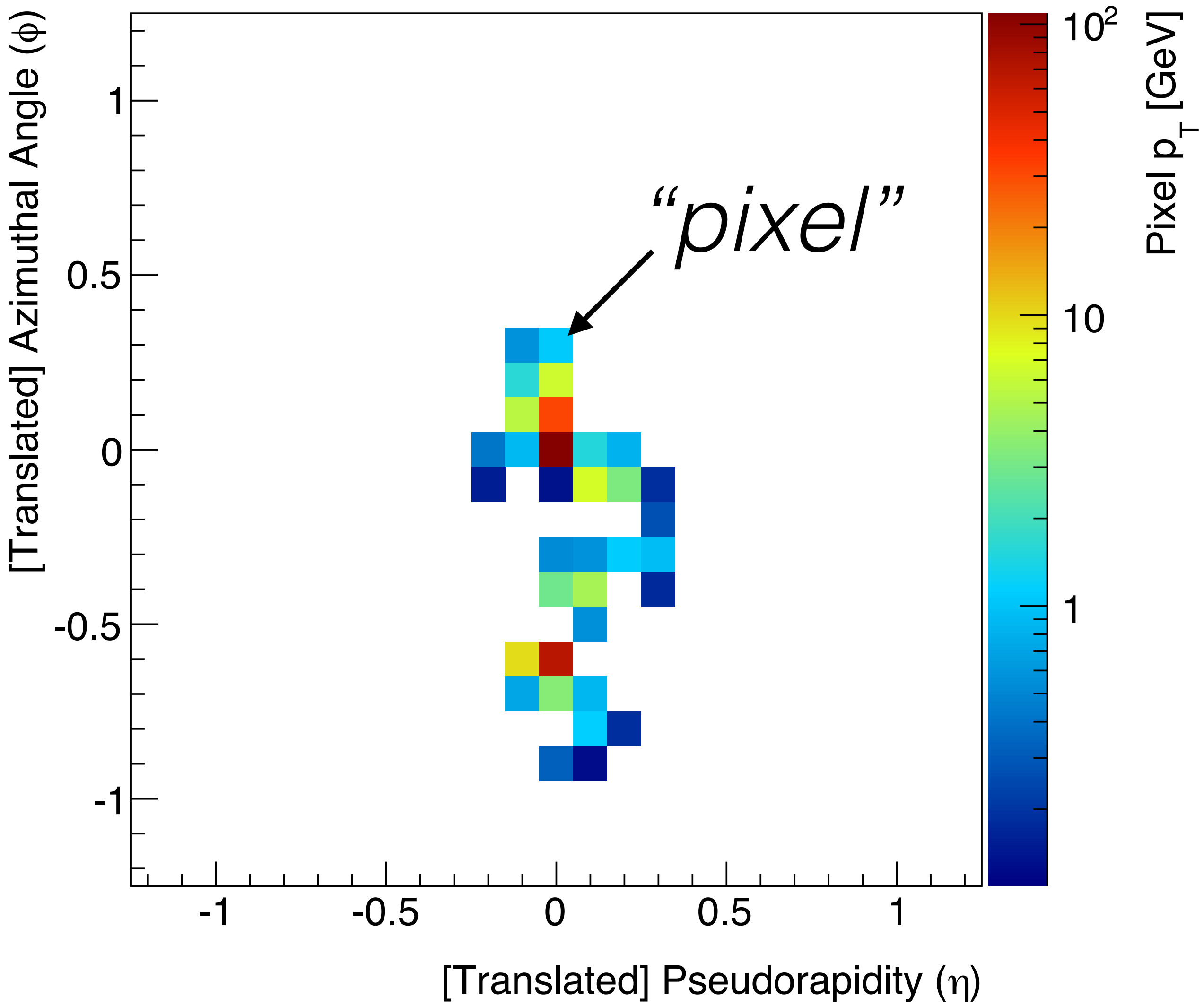


L. de Oliveira, M. Paganini, **BPN**,
Comp. and Software for Big Science (2017) 1

Jet Images

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

Jet from a boosted W boson



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

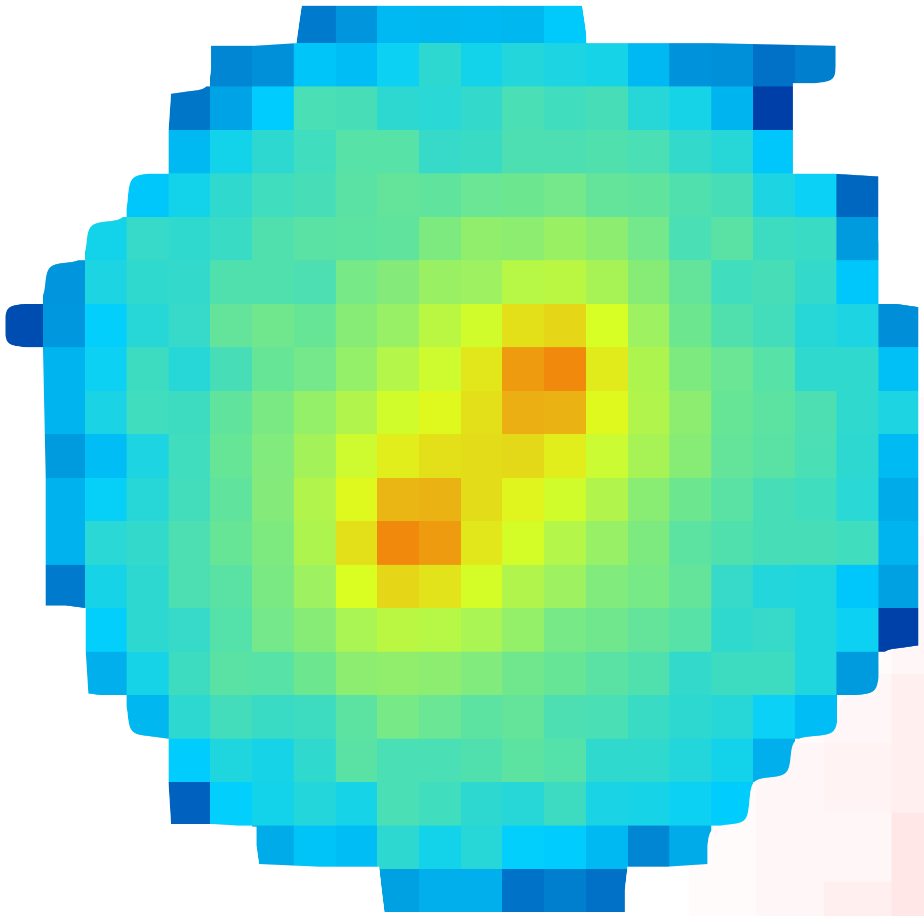
Credit: Peter G Trimming (Wikipedia)

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

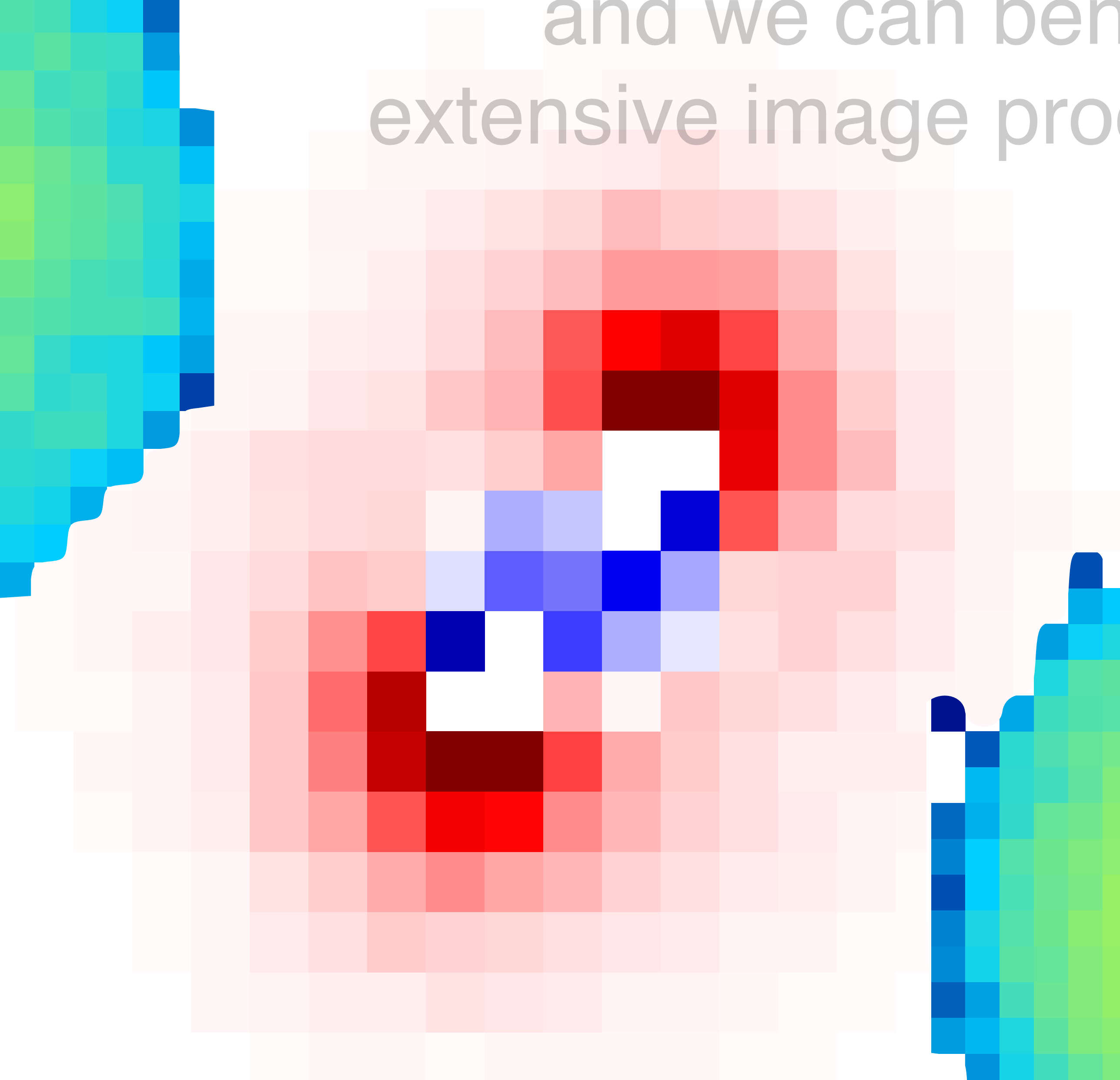
Why Images?

Can directly visualize physics

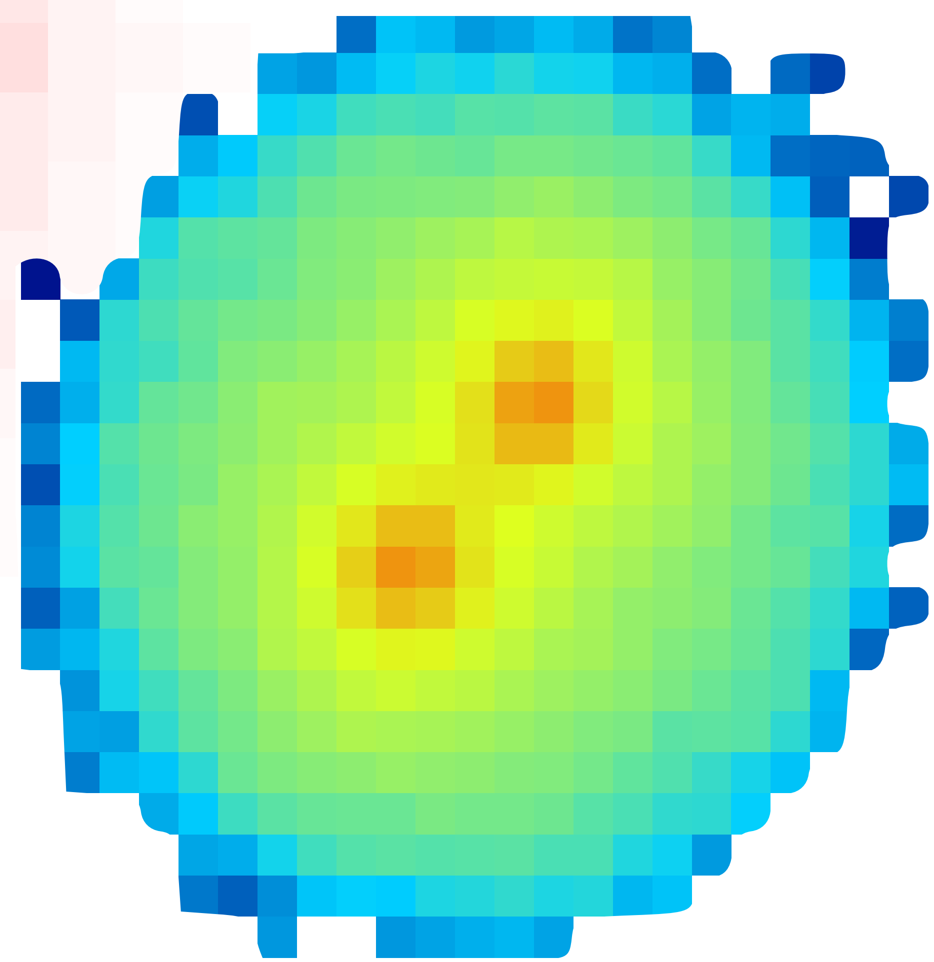
and we can benefit from the extensive image processing literature



$W \dashrightarrow q\bar{q}$

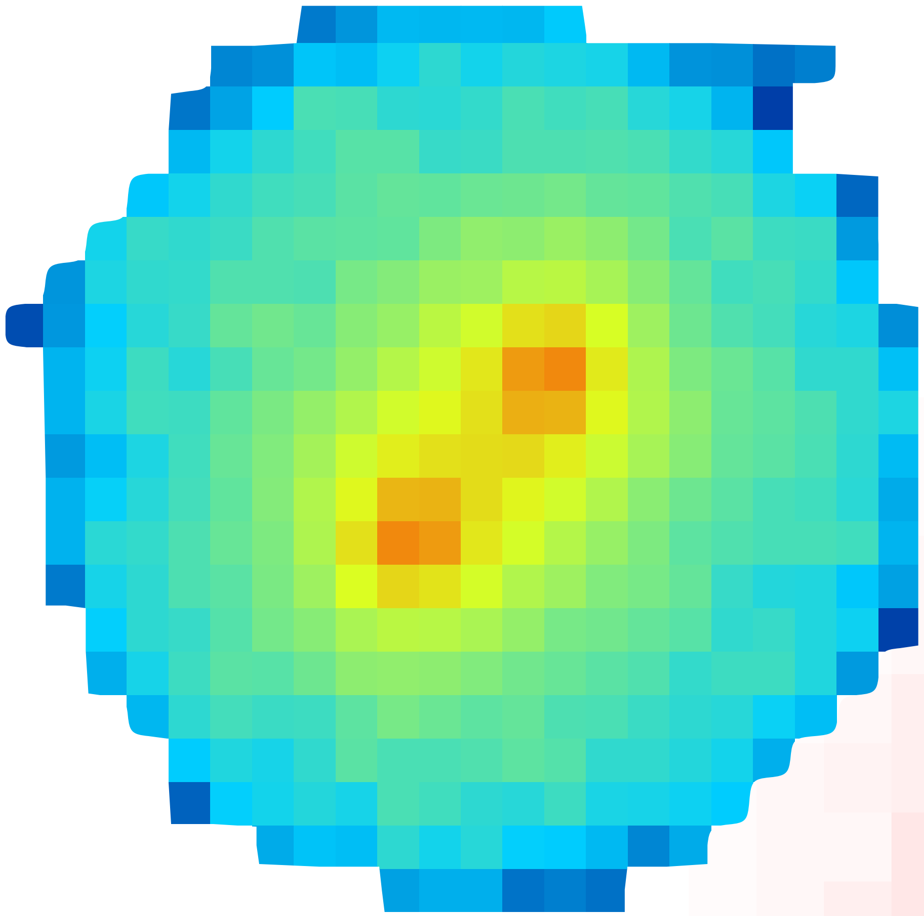


$g \dashrightarrow q\bar{q}$



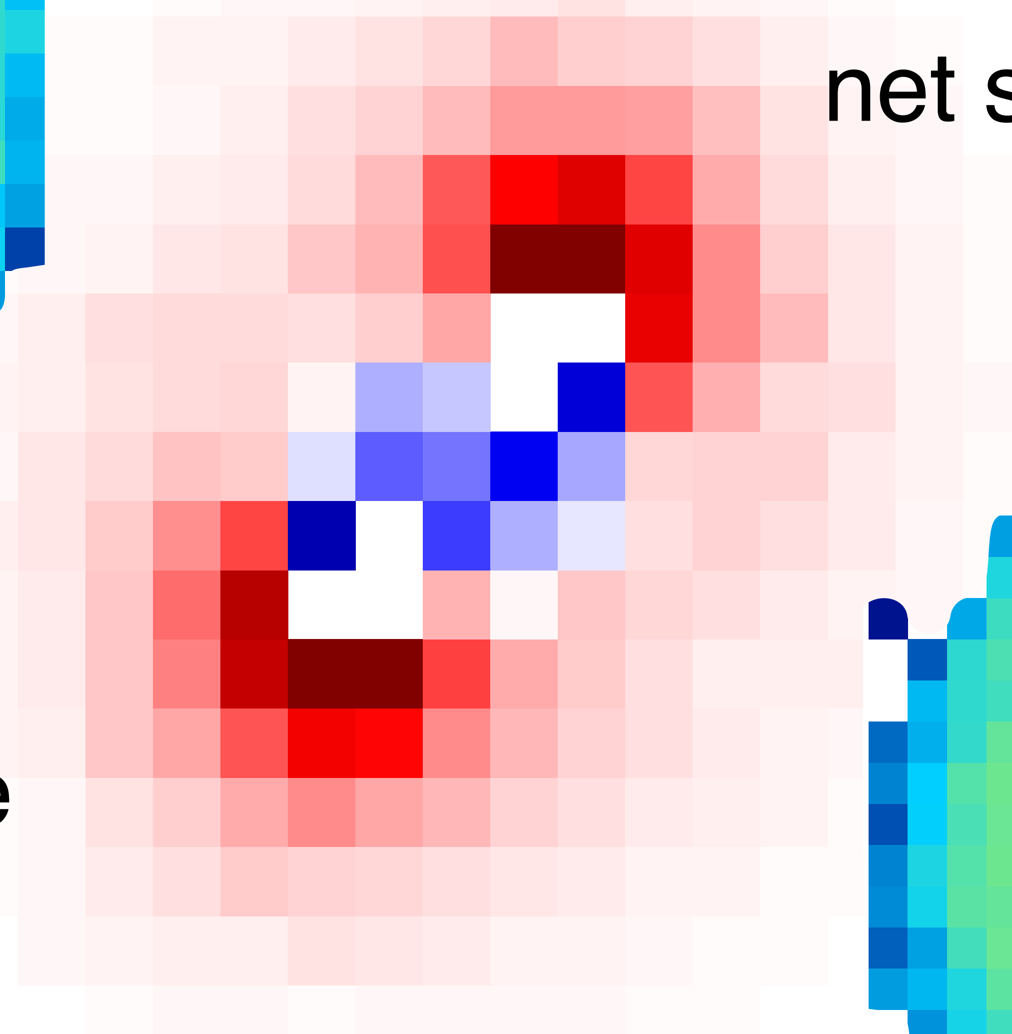
there is information encoded in the physical distance between pixels

Why Images?



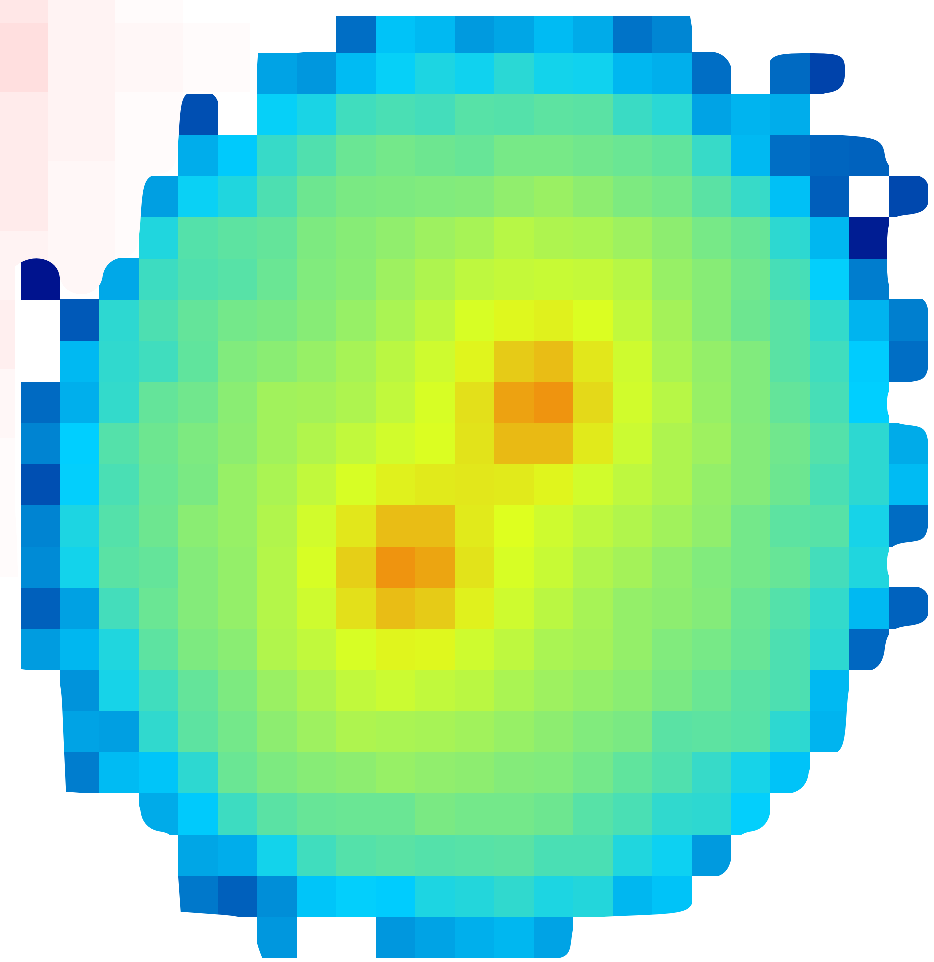
Can directly visualize physics

and we can benefit from the extensive image processing literature



net strong-force charge

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

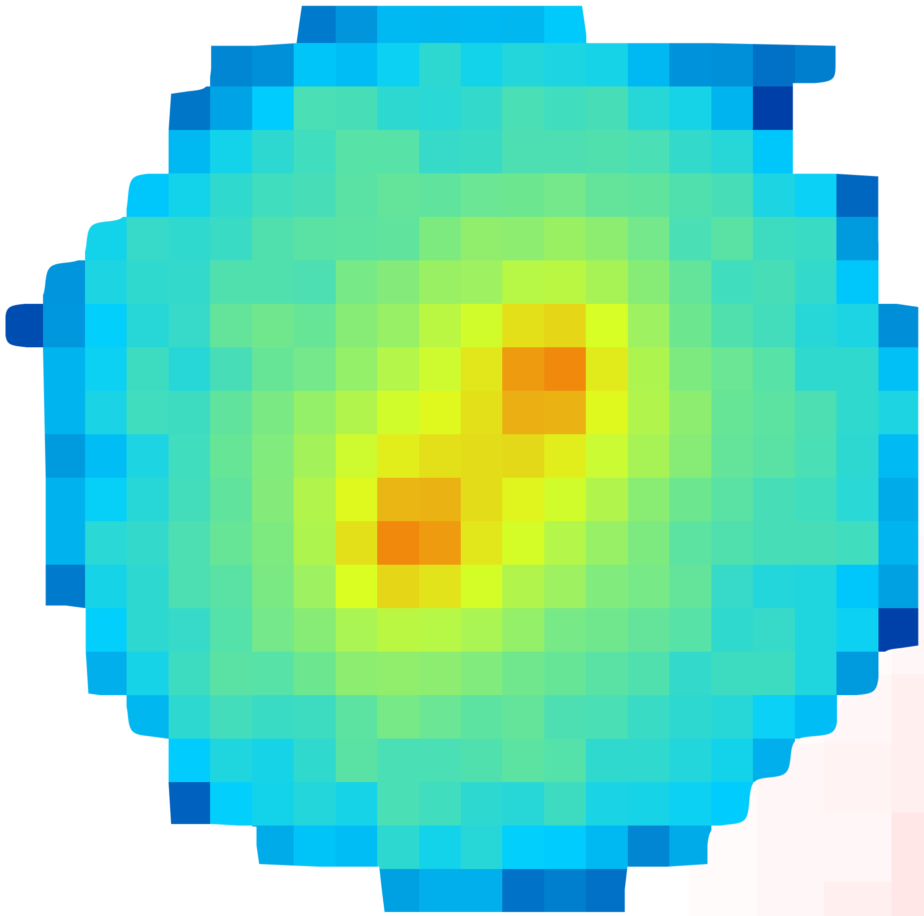


$\mathbf{W} \dashrightarrow q\bar{q}$

radiates like a dipole (no net charge)

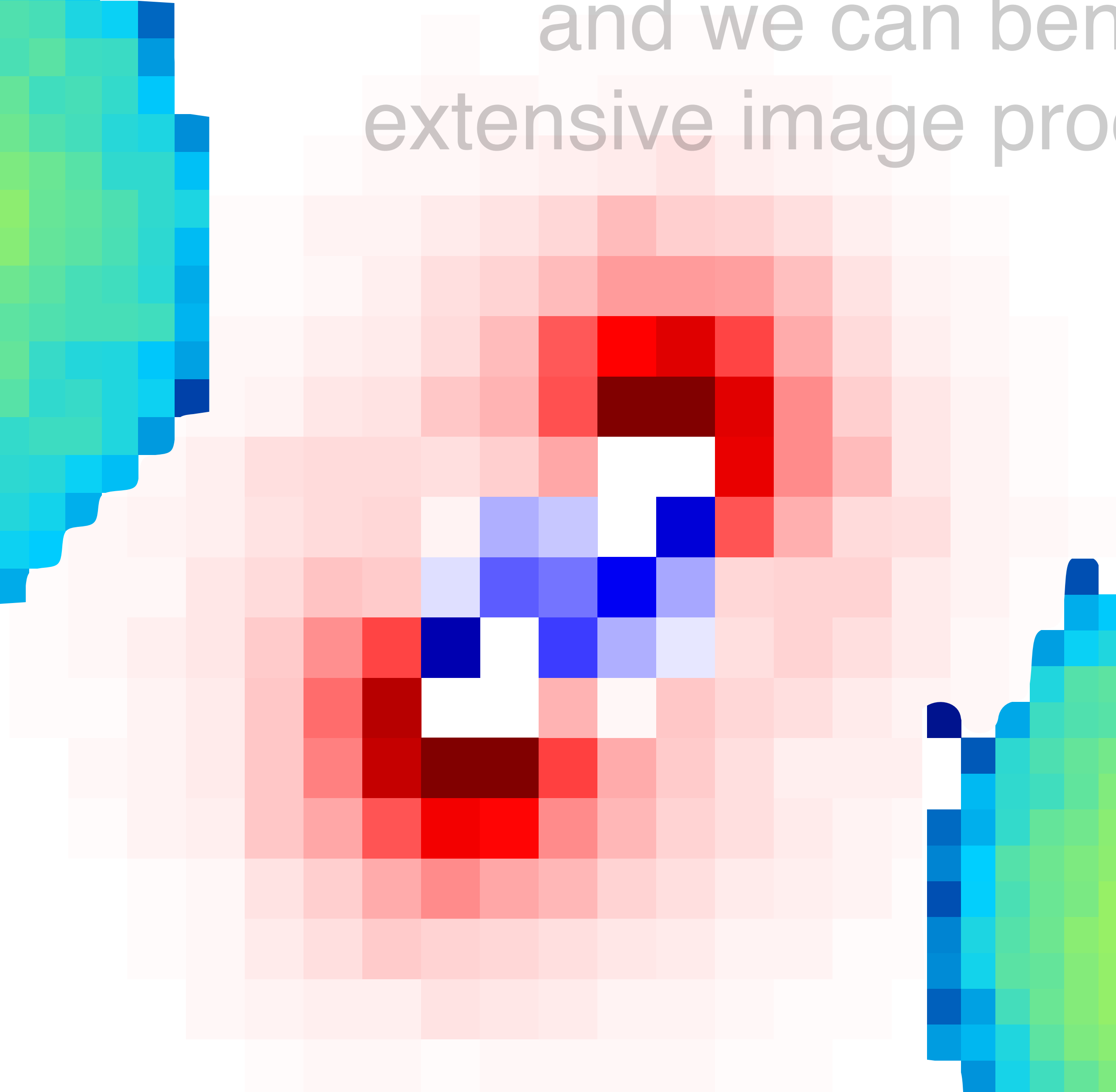
there is information encoded in the physical distance between pixels

Why Images?

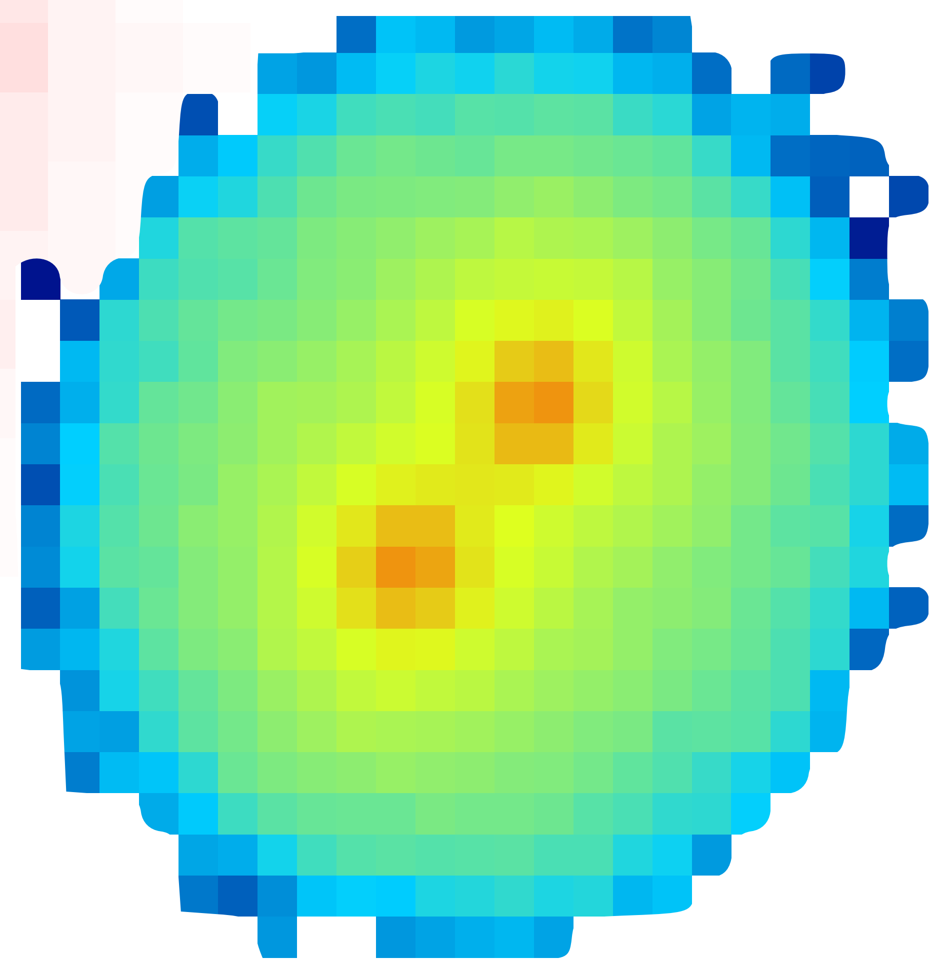


Can directly visualize physics

and we can benefit from the extensive image processing literature



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



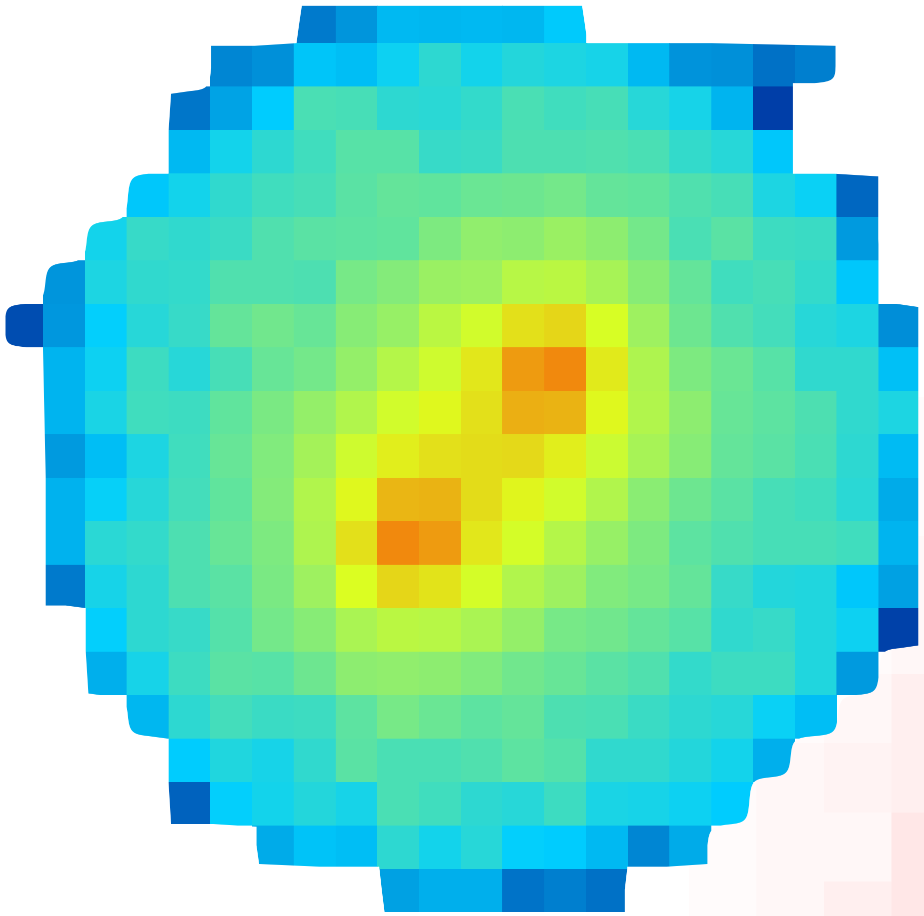
$\mathbf{W} \dashrightarrow q\bar{q}$

there is information encoded in the physical distance between pixels

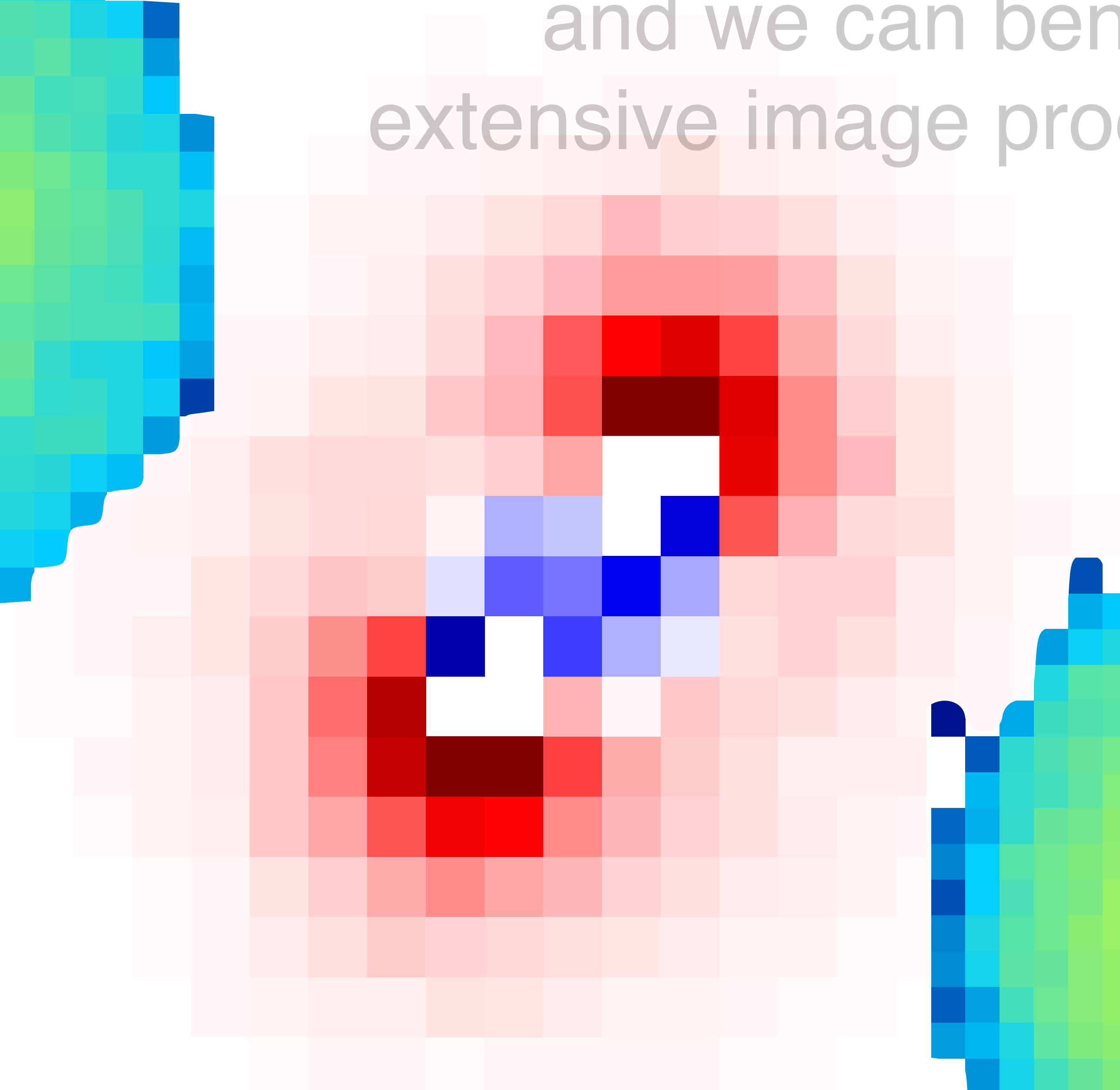
Why Images?

Can directly visualize physics

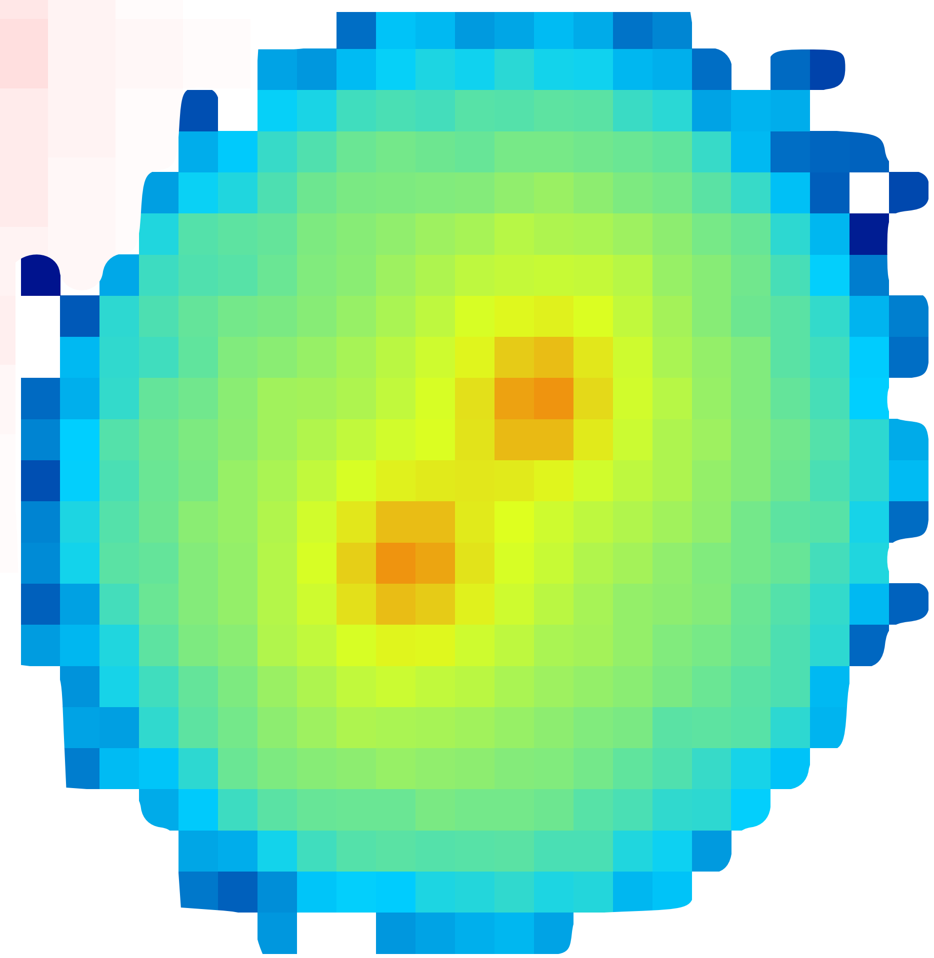
and we can benefit from the extensive image processing literature



$W \dashrightarrow q\bar{q}$



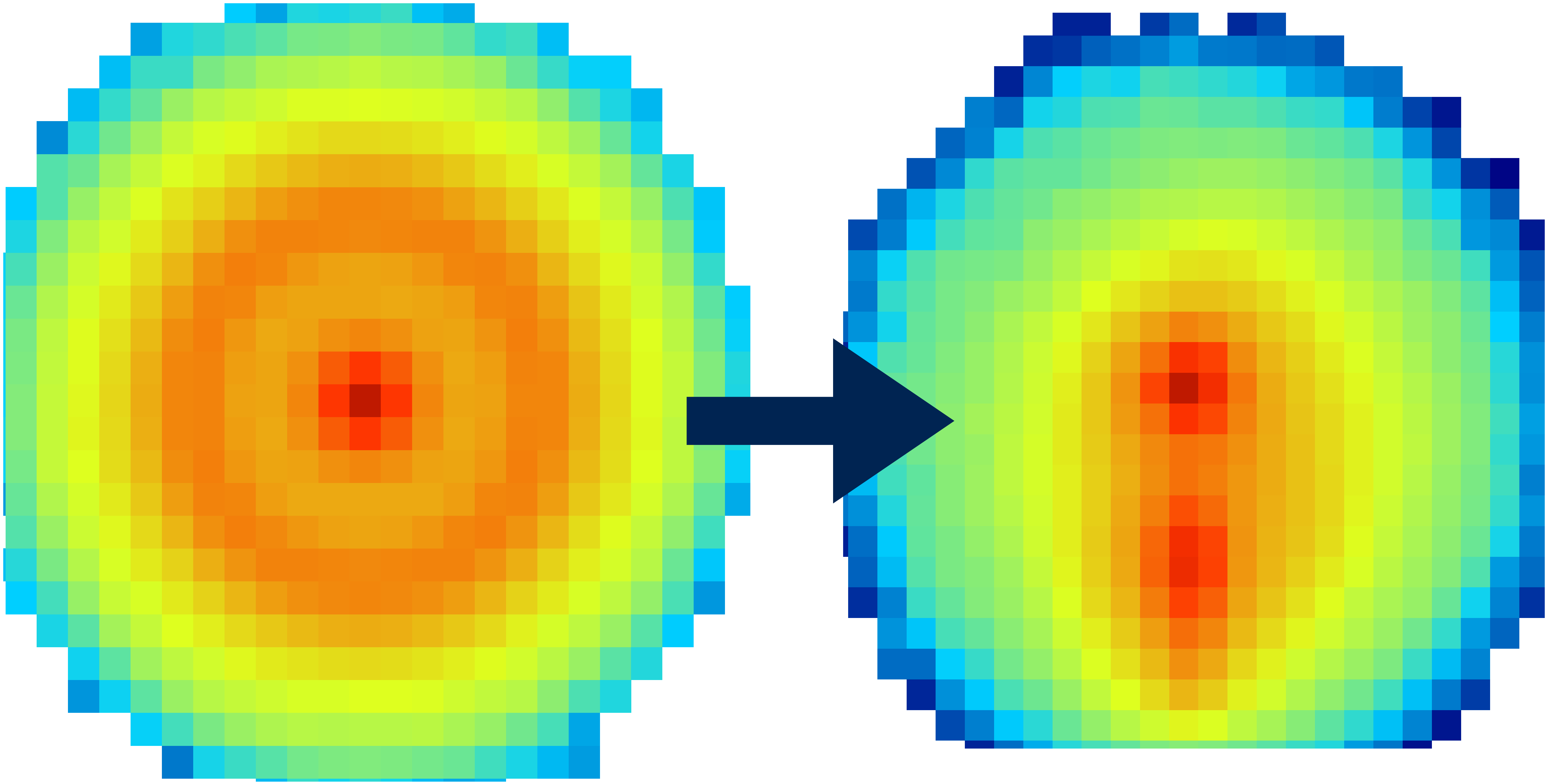
$g \dashrightarrow q\bar{q}$



there is information encoded in the physical distance between pixels

Pre-processing & spacetime symmetries

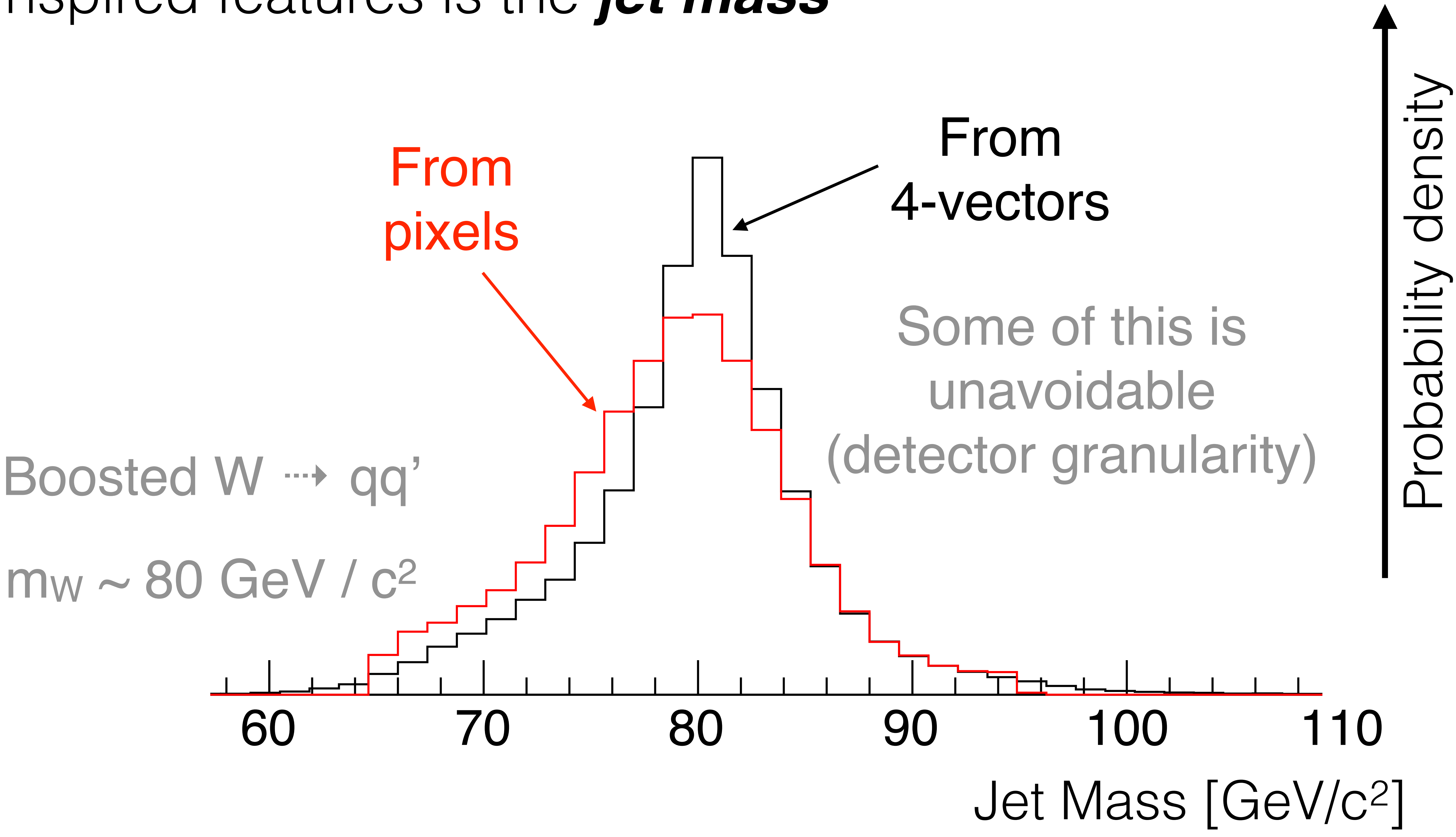
One of the first typical steps is pre-processing



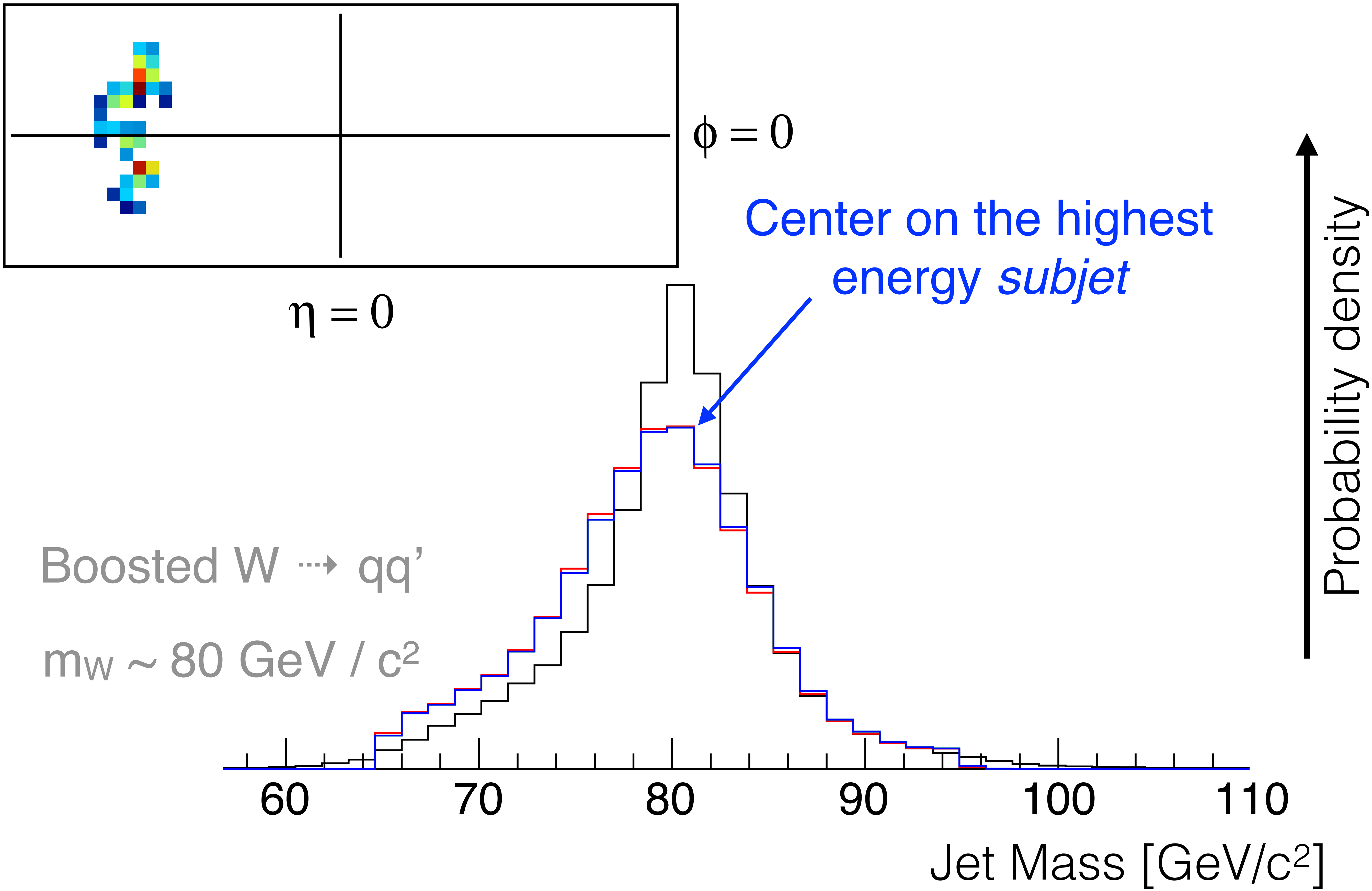
Can help to learn faster & smarter; but must be careful!

Pre-processing & spacetime symmetries

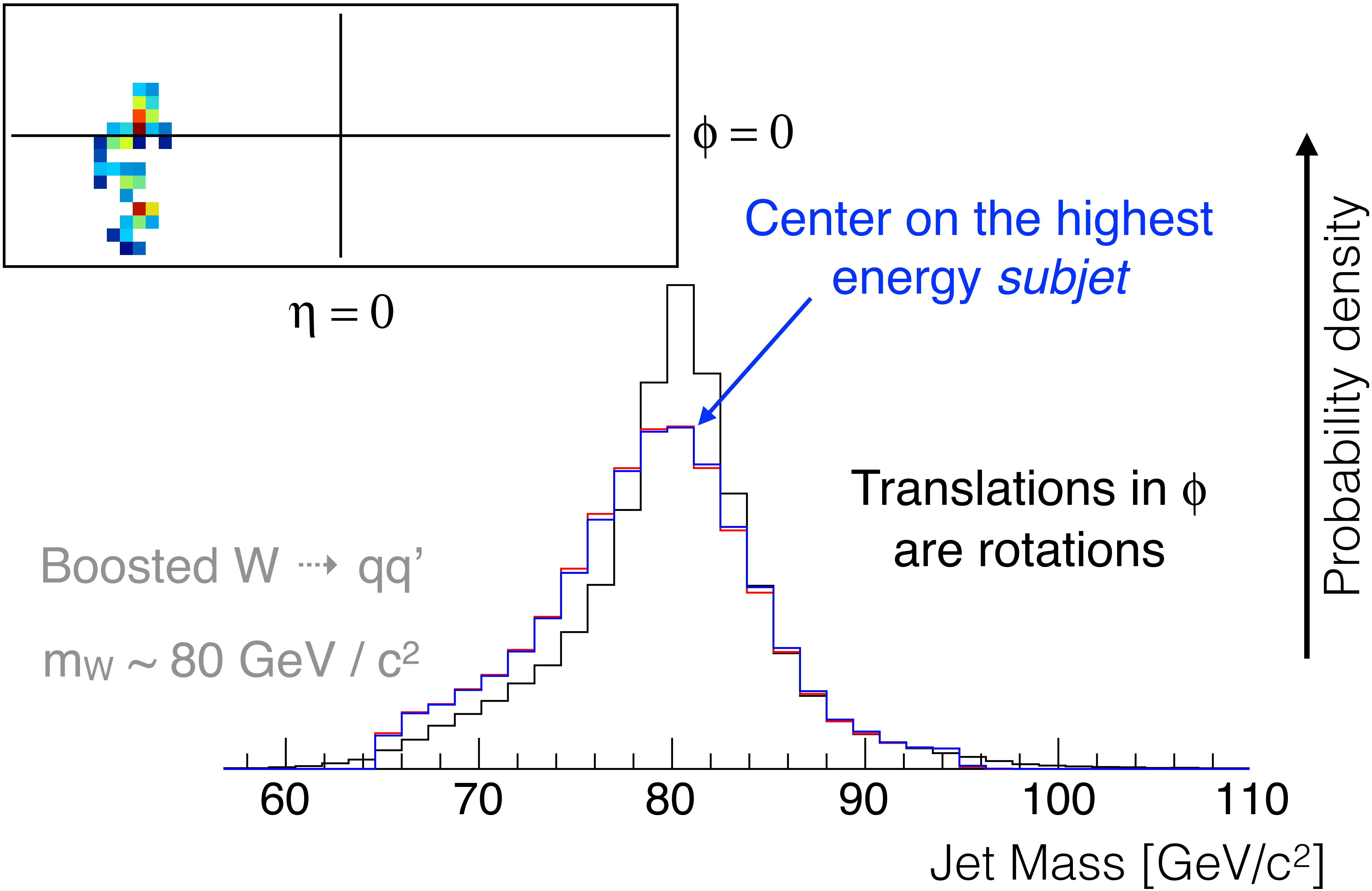
One of the most useful physics-inspired features is the ***jet mass***



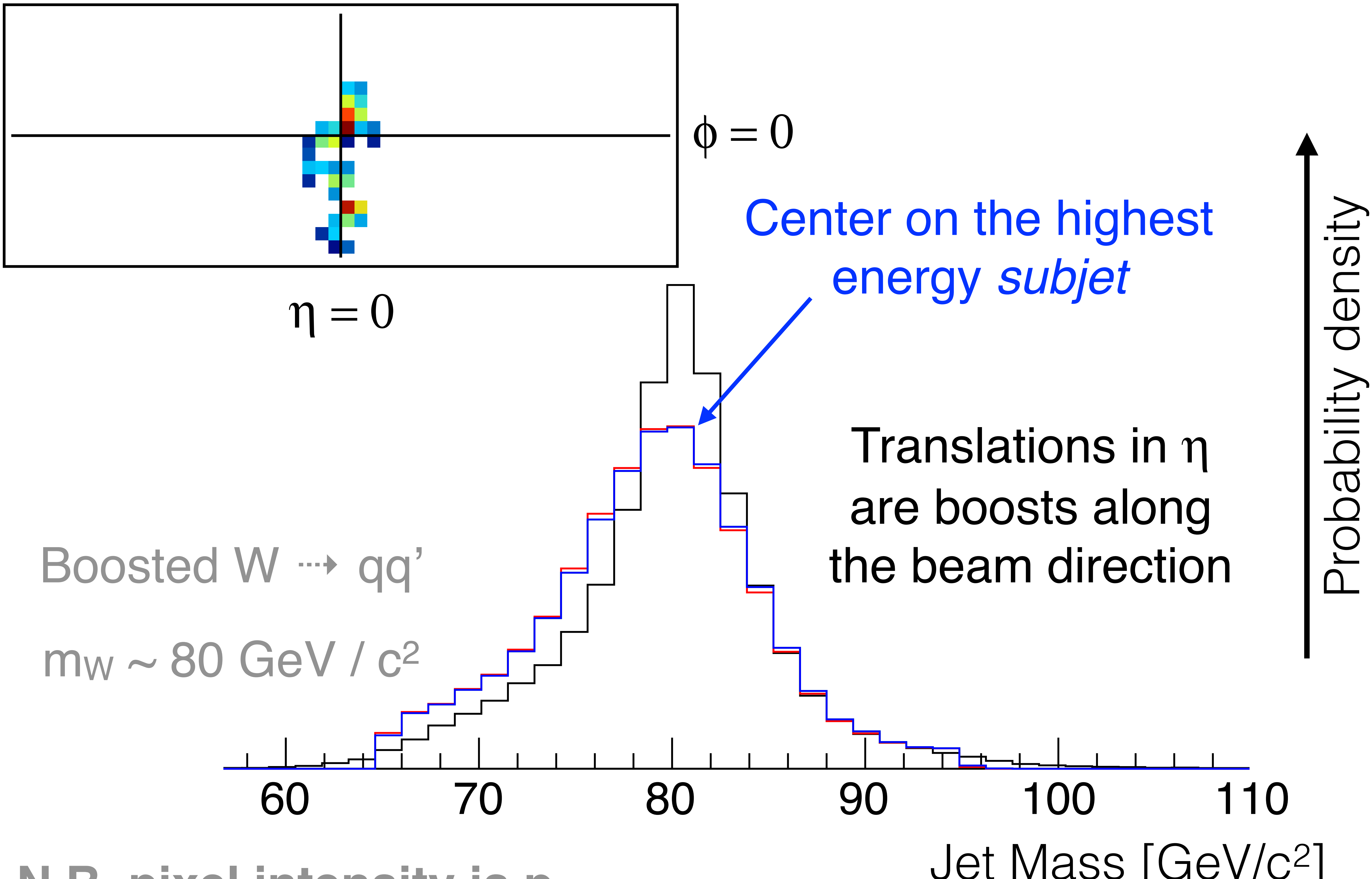
Pre-processing & spacetime symmetries



Pre-processing & spacetime symmetries



Pre-processing & spacetime symmetries



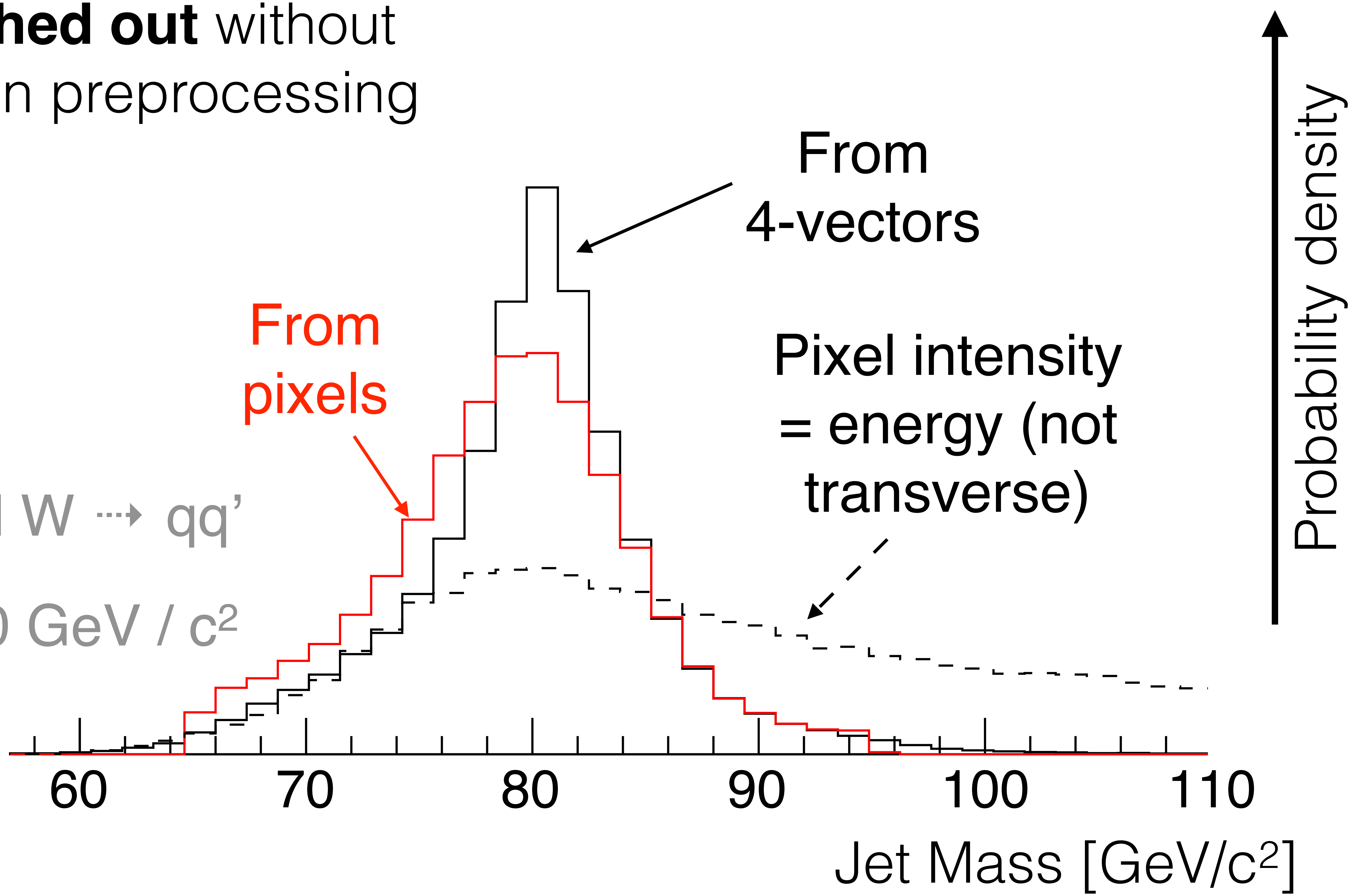
N.B. pixel intensity is p_T

Pre-processing & spacetime symmetries

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

Boosted W ⇢ qq'

$m_W \sim 80$ GeV / c^2

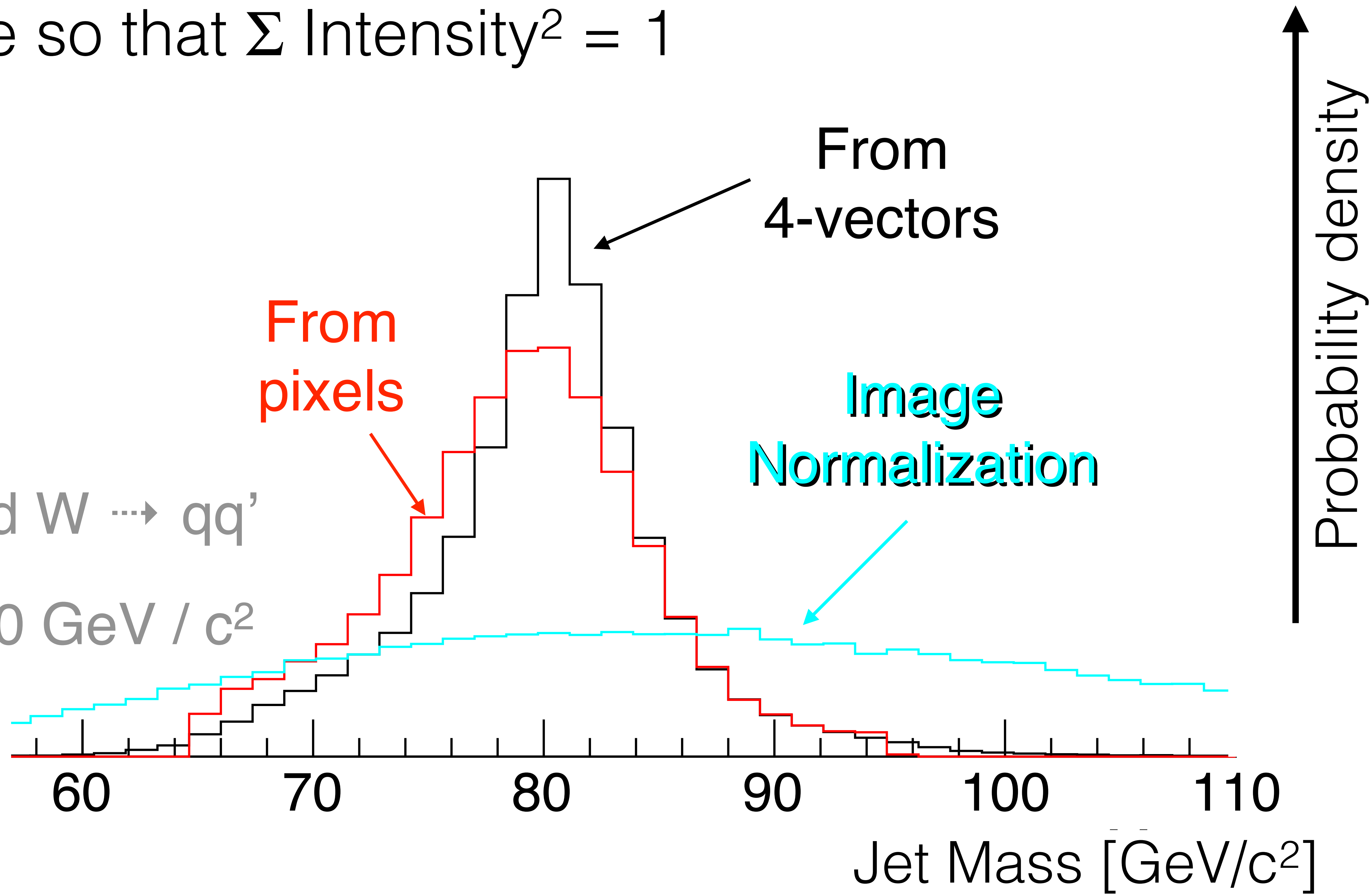


Pre-processing & spacetime symmetries

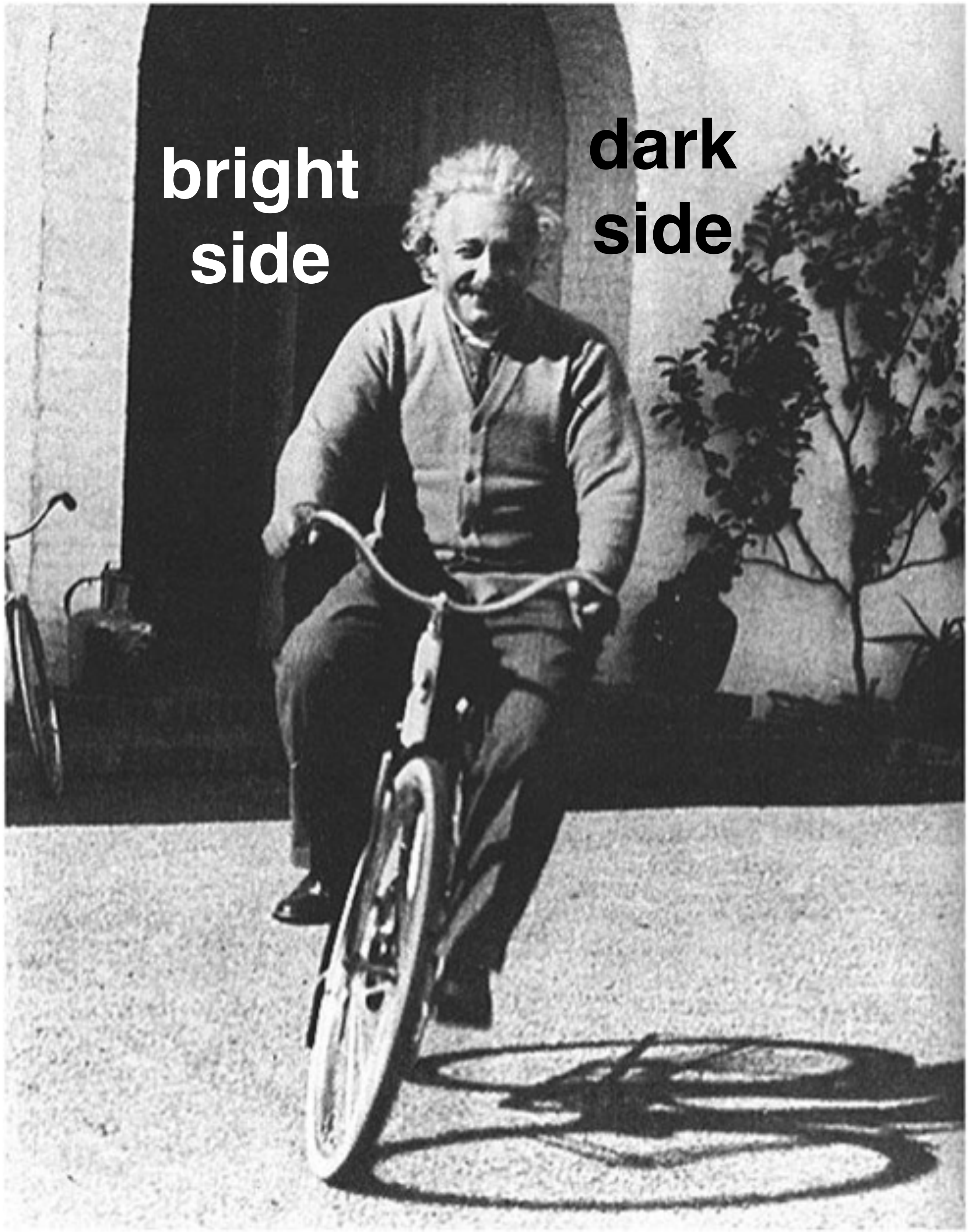
It is common to normalize each image so that Σ Intensity^2 = 1

Boosted W → qq'

$m_W \sim 80$ GeV / c^2



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*

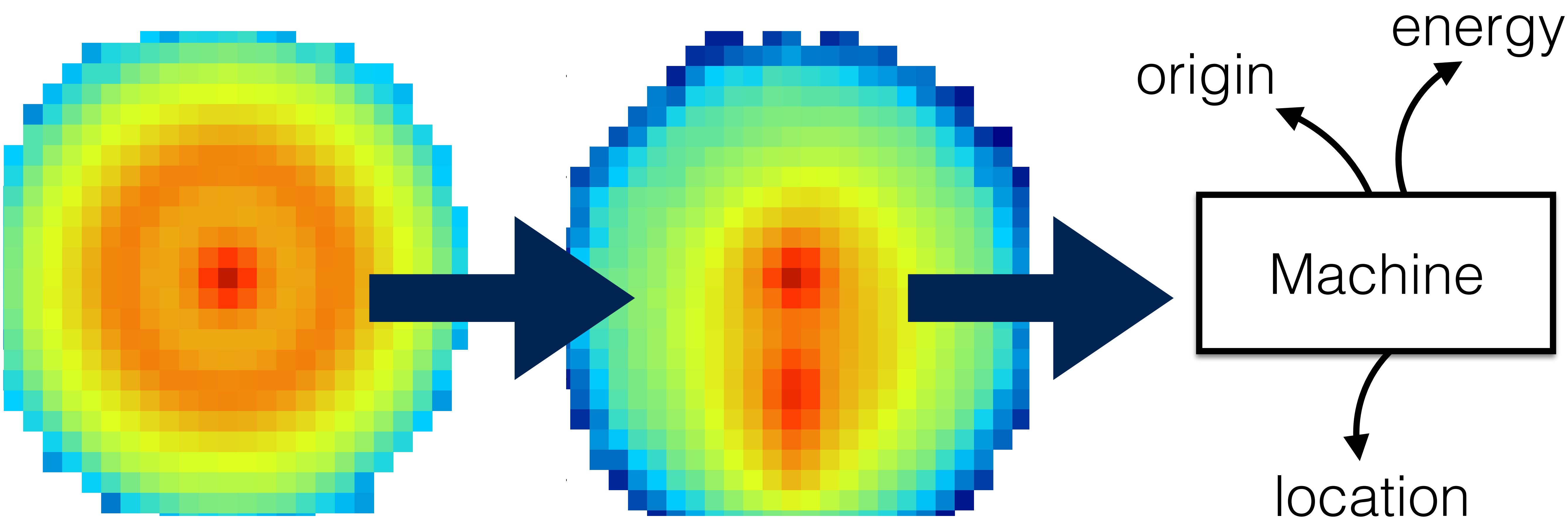
In both pictures, total intensity of Einstein's face is about the same.

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

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

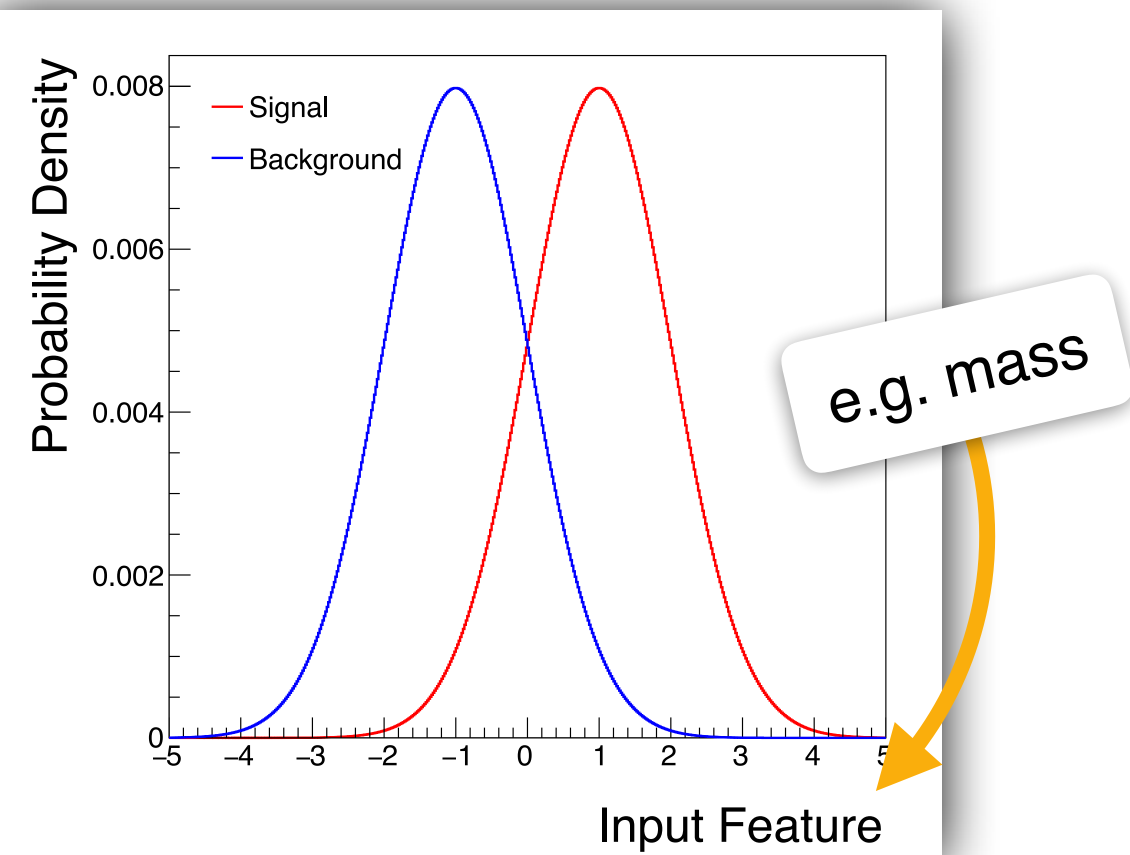
Using jet images

Now, with a carefully processed image, we can start using them for machine learning.



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

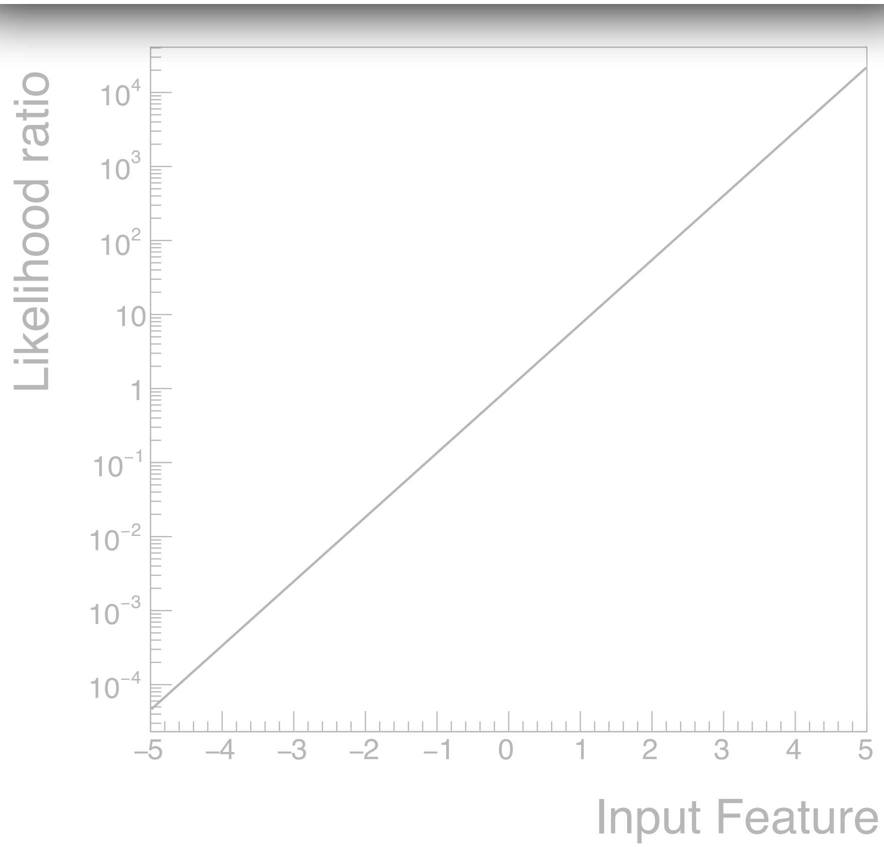
Optimal Classification with ML

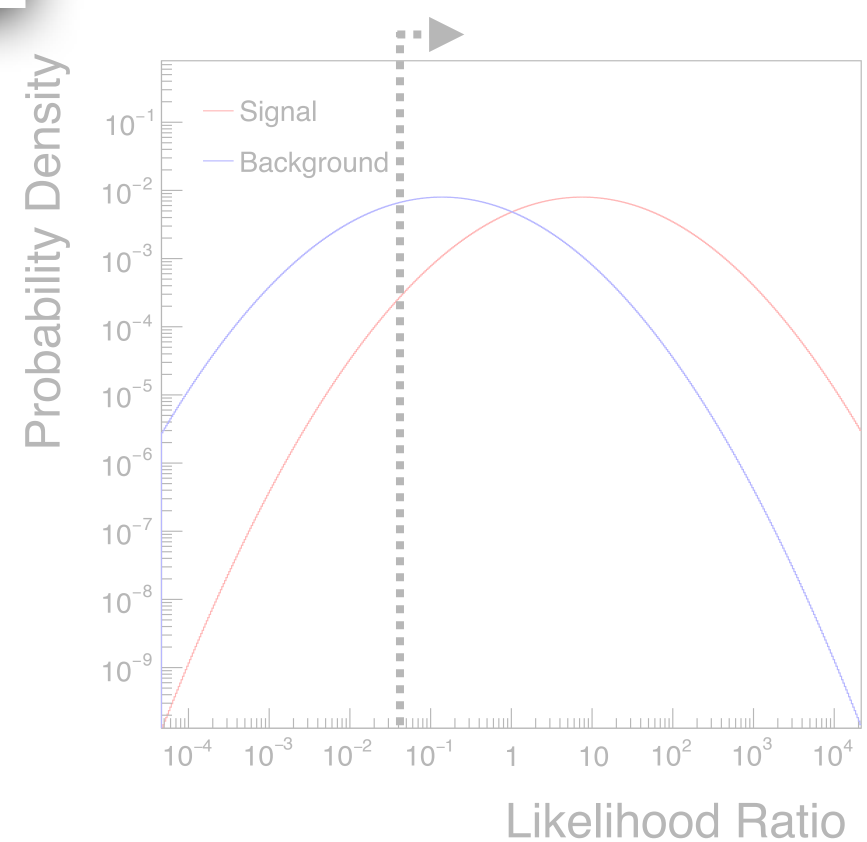


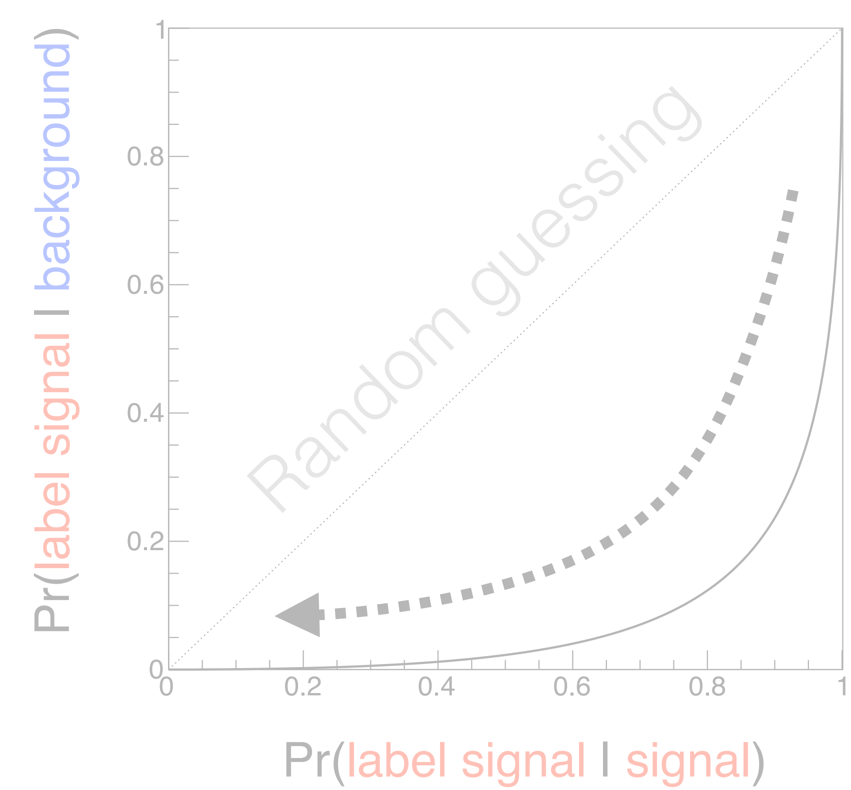
Think of events as e.g. an image

↓

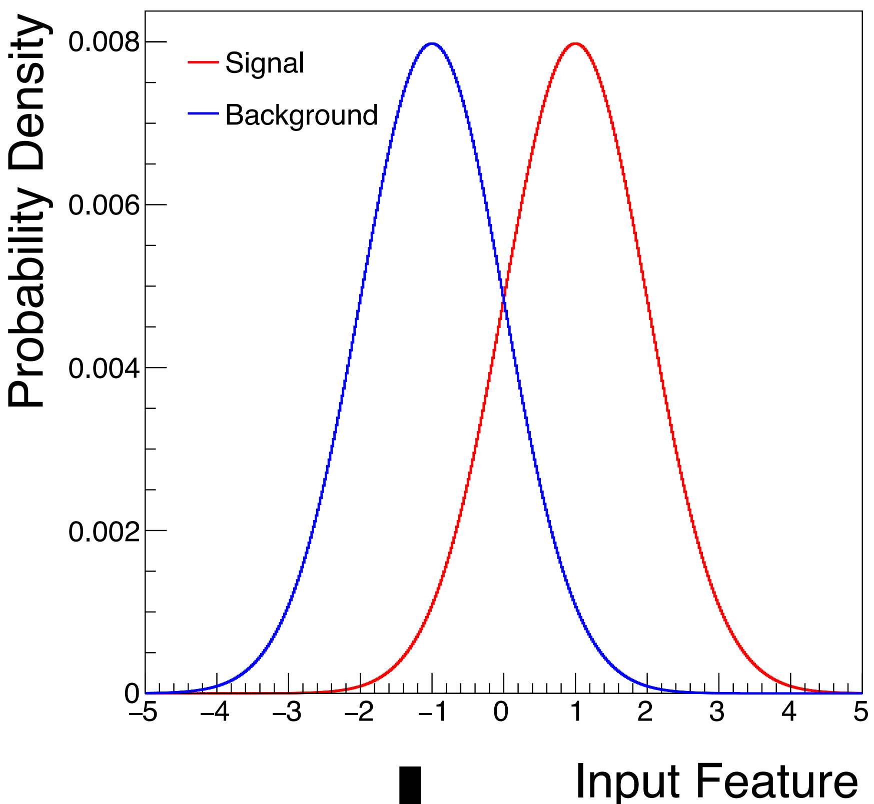
Train a classifier to distinguish known physics (“**background**”) from new particles (“**signal**”)

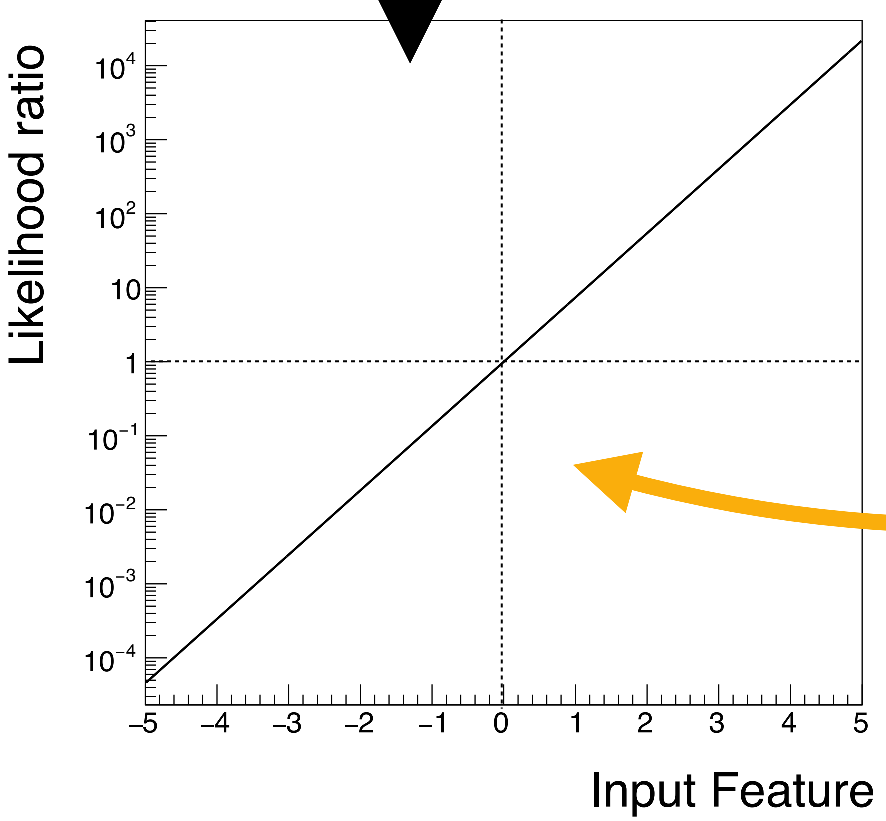






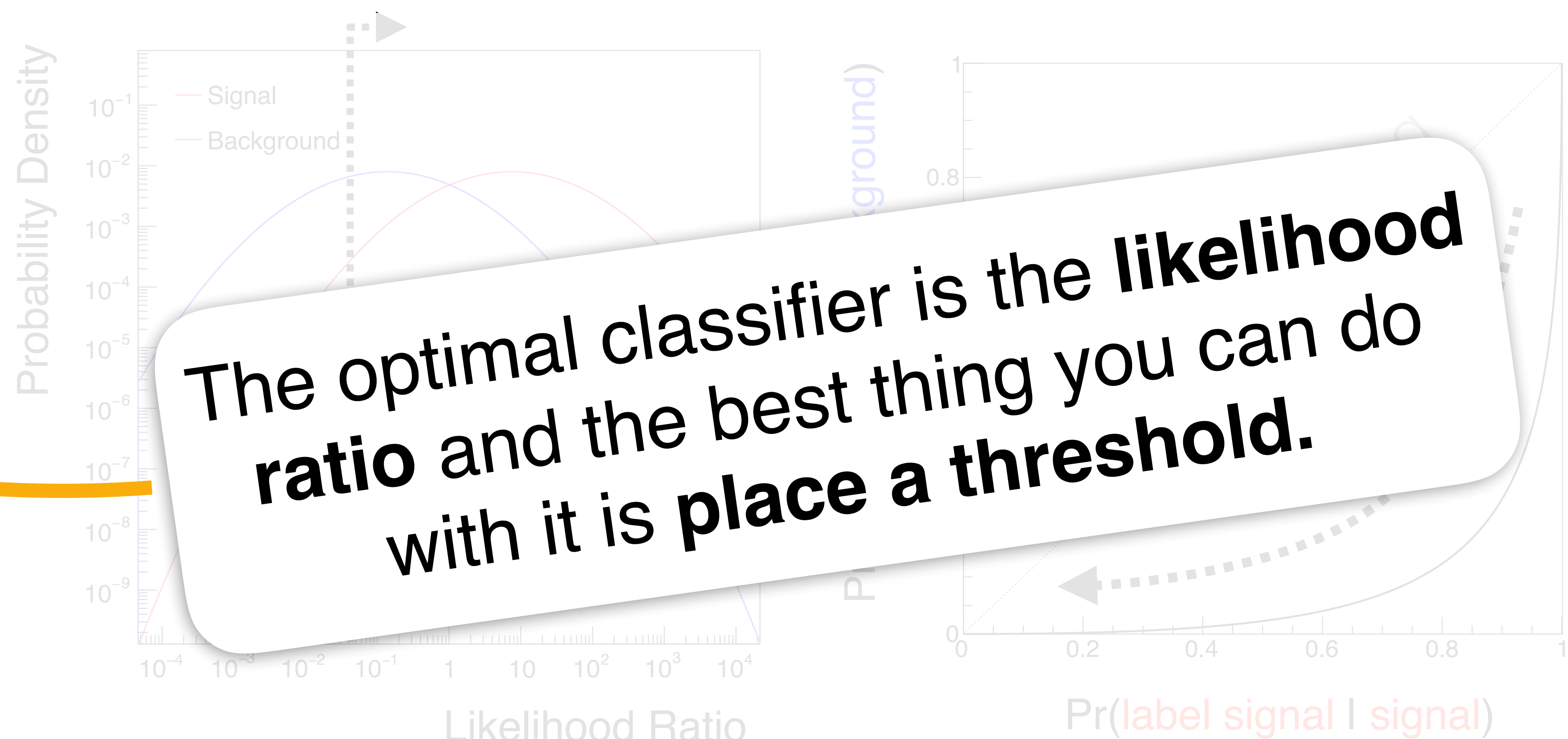
Optimal Classification with ML





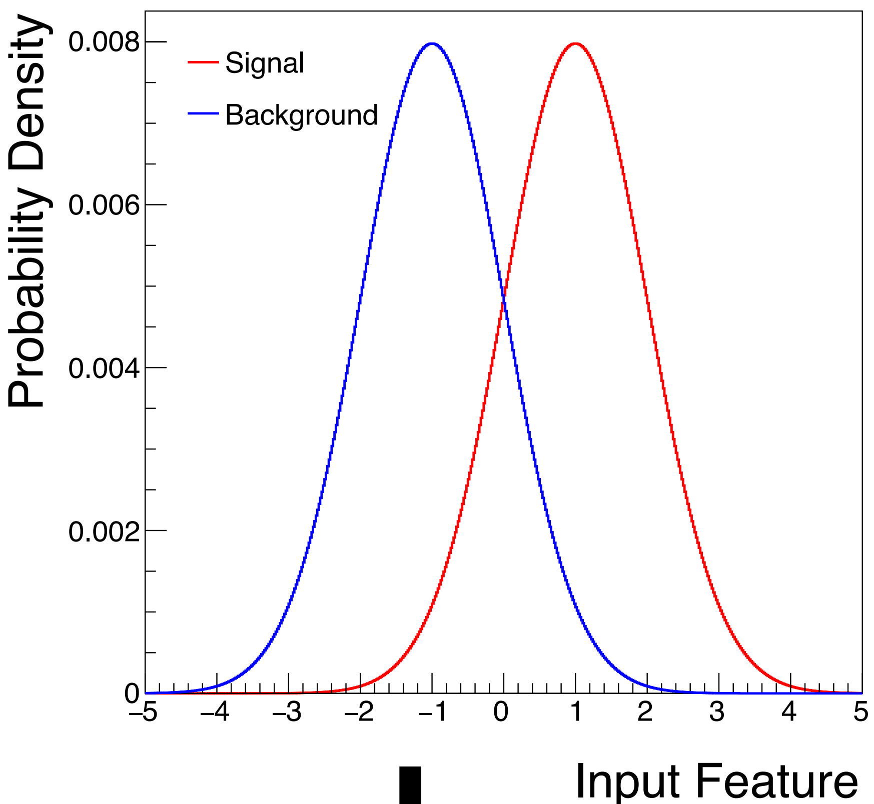
Think of events as e.g. an image

Train a classifier to distinguish known physics ("**background**") from new particles ("**signal**")



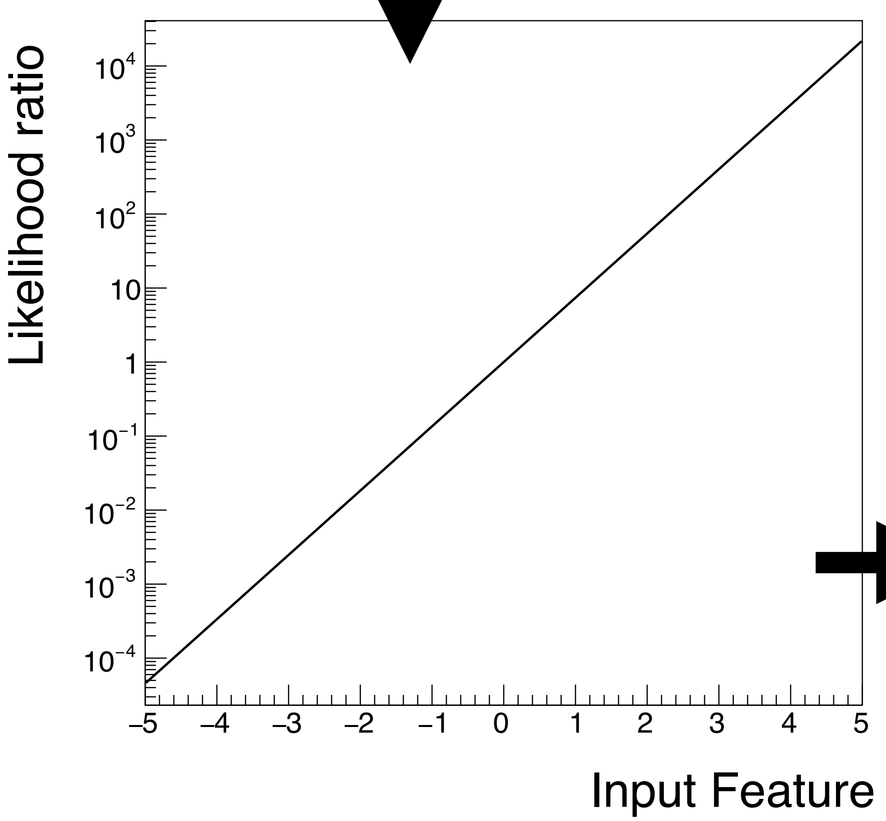
The optimal classifier is the **likelihood ratio** and the best thing you can do with it is **place a threshold.**

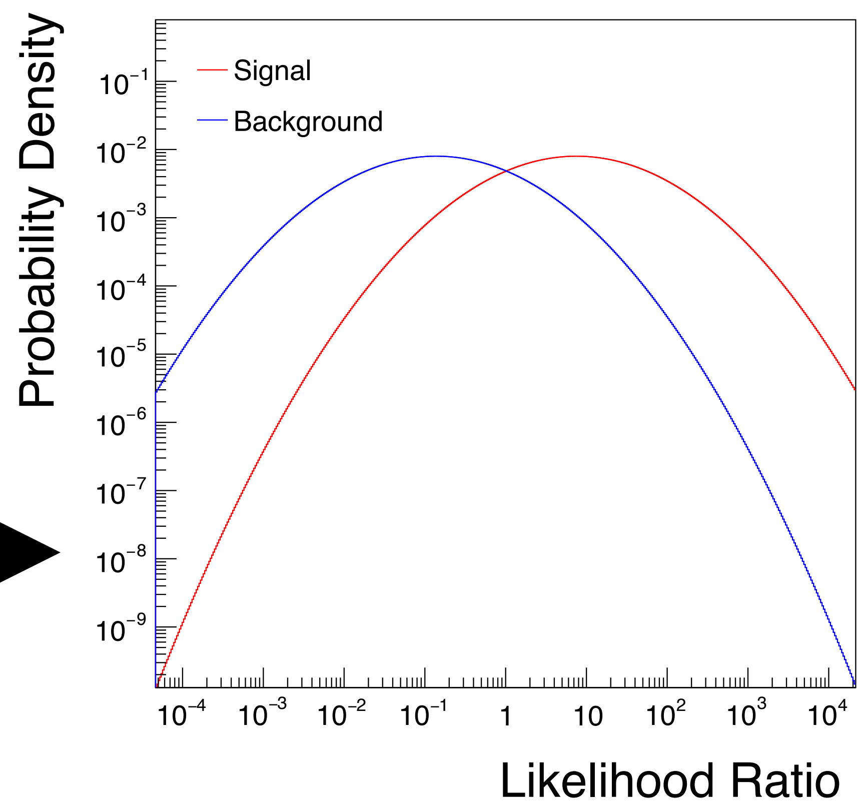
Optimal Classification with ML



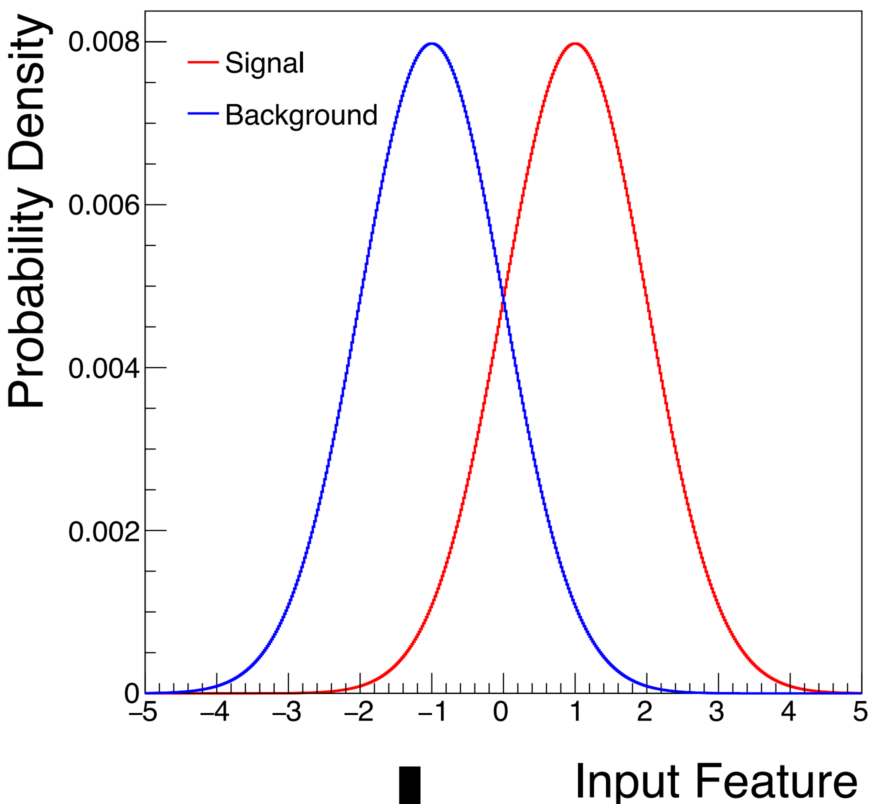
Think of events as e.g. an image

Train a classifier to distinguish known physics (“**background**”) from new particles (“**signal**”)



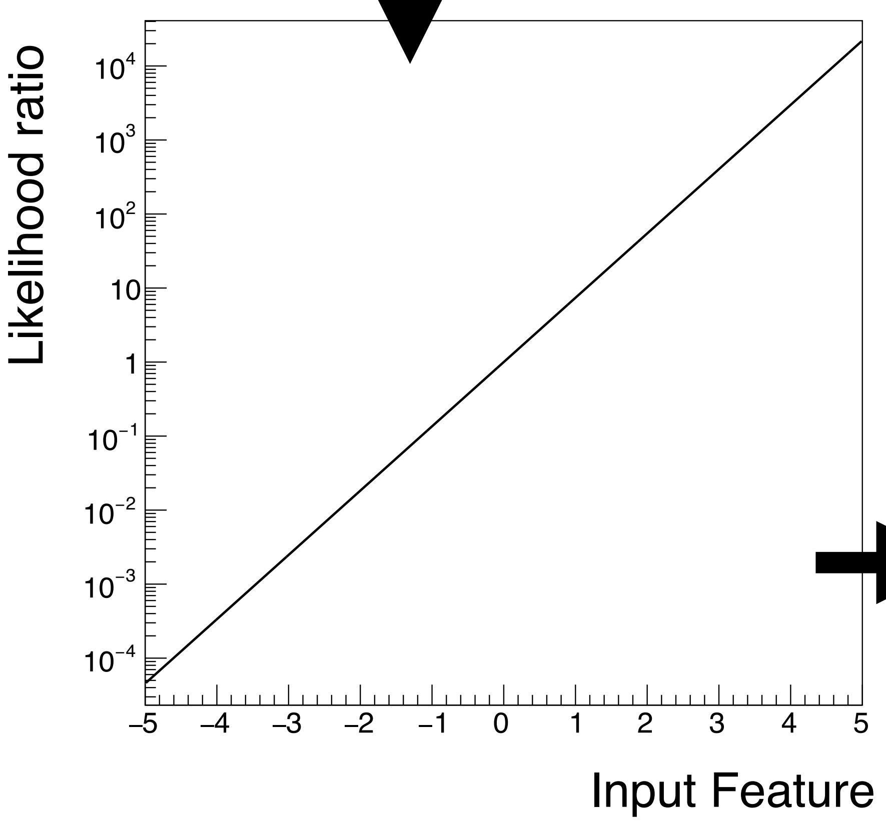


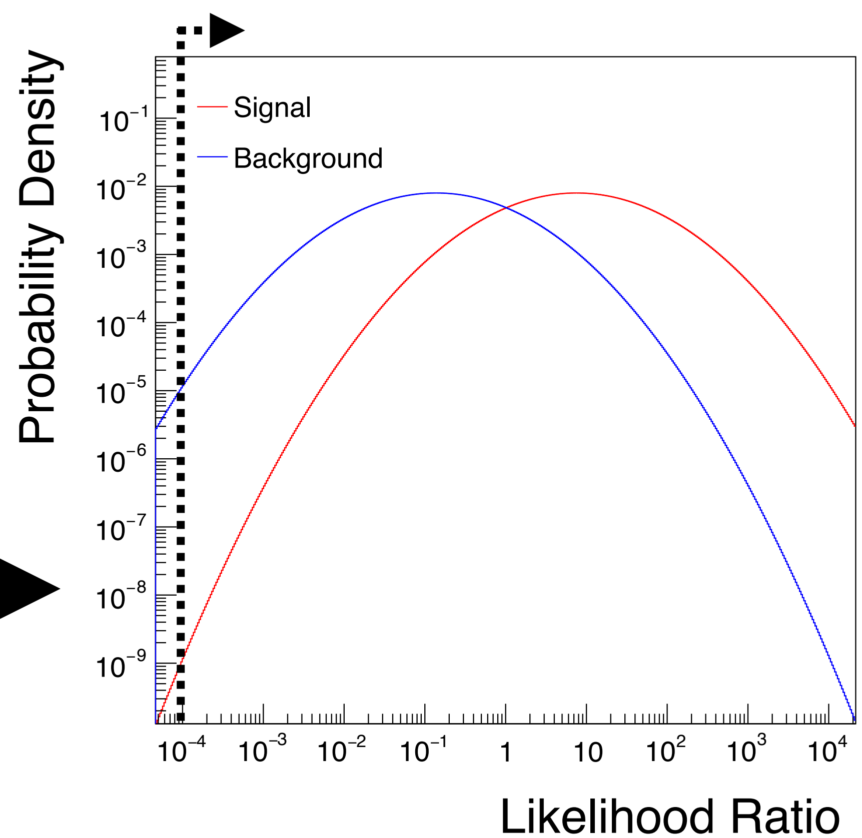
Optimal Classification with ML

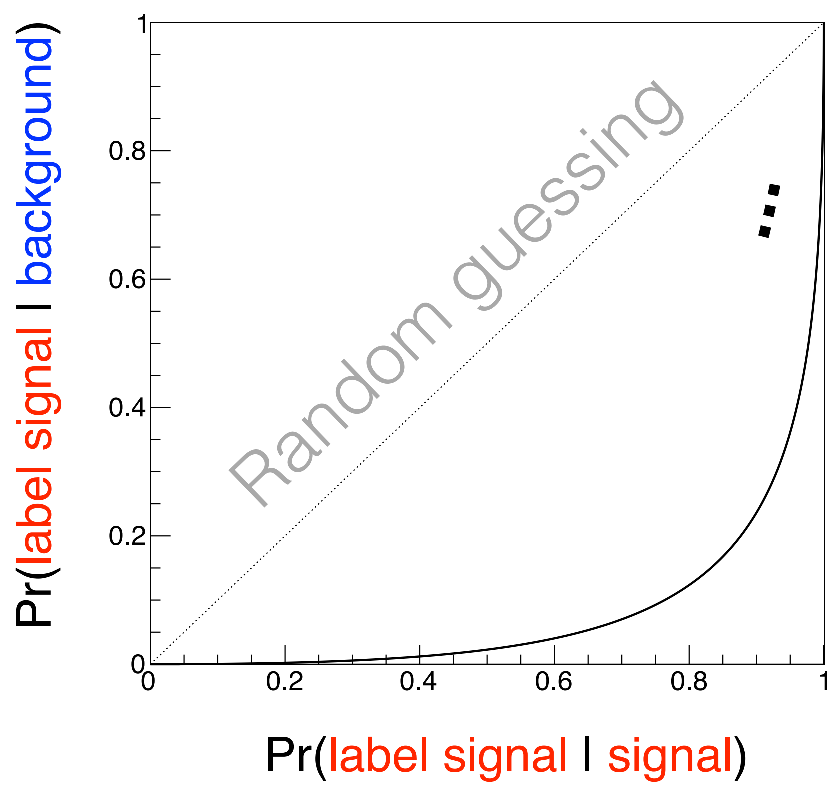


Think of events as e.g. an image

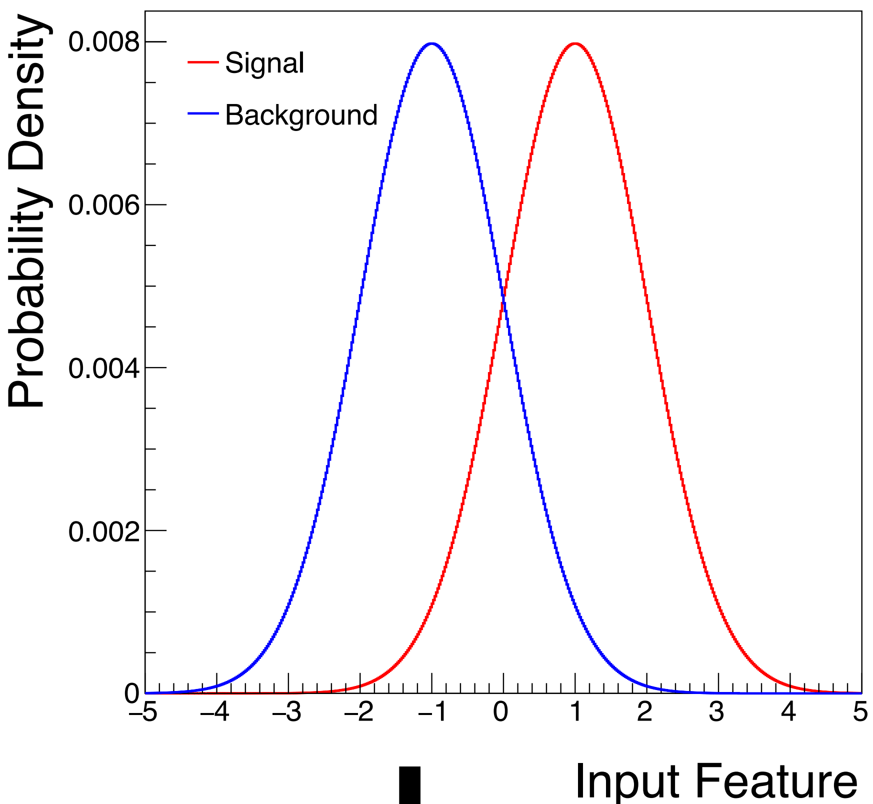
Train a classifier to distinguish known physics (“**background**”) from new particles (“**signal**”)





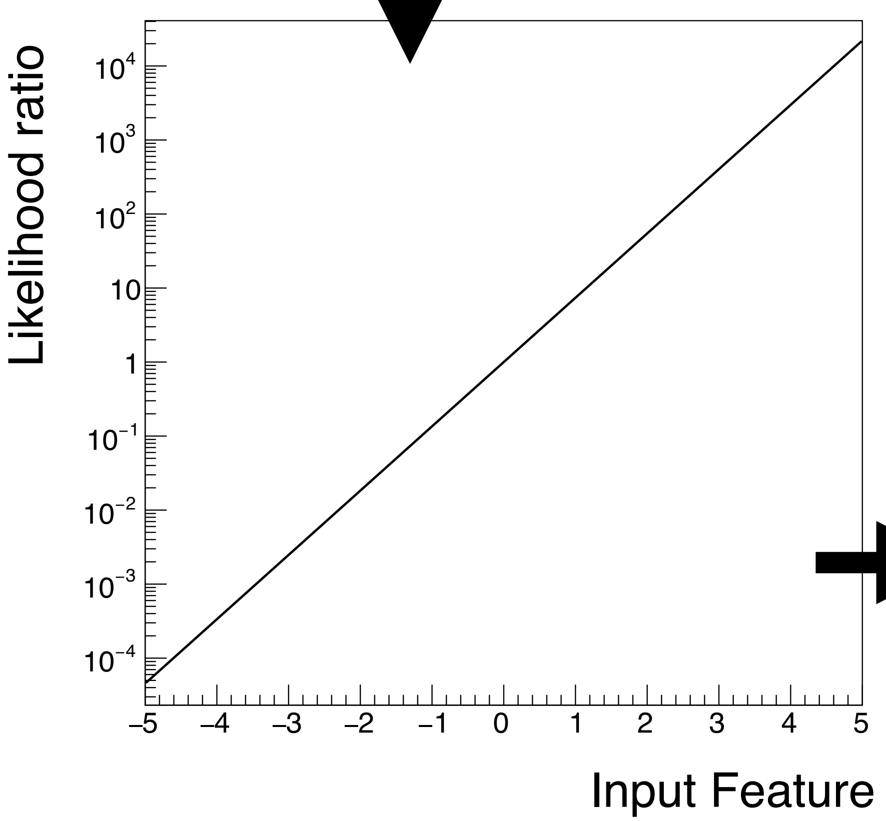


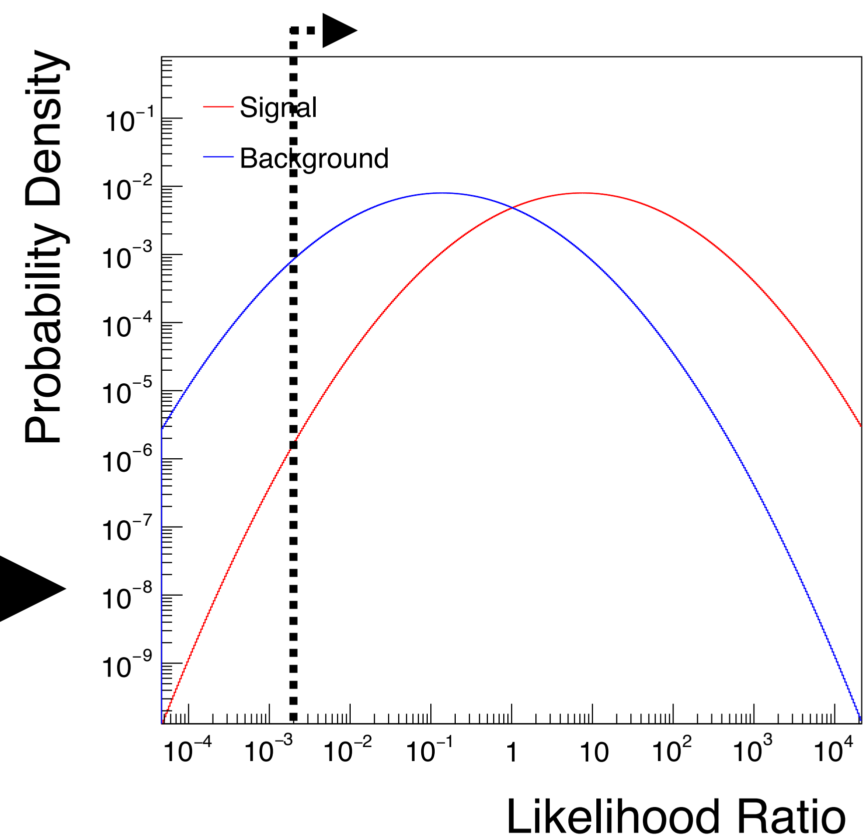
Optimal Classification with ML

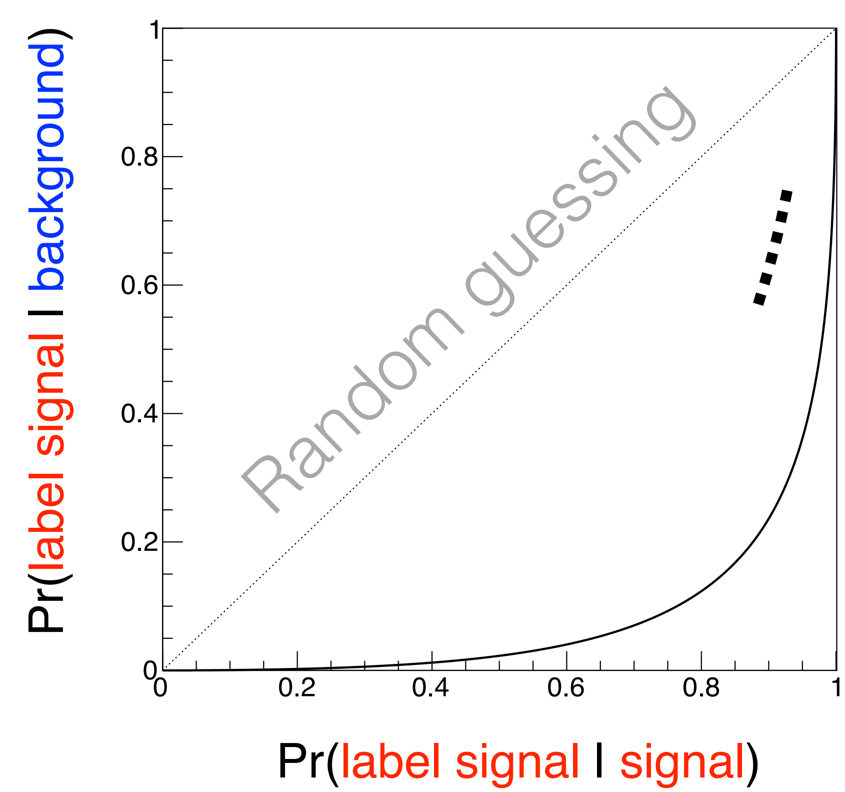


Think of events as e.g. an image

Train a classifier to distinguish known physics ("**background**") from new particles ("**signal**")



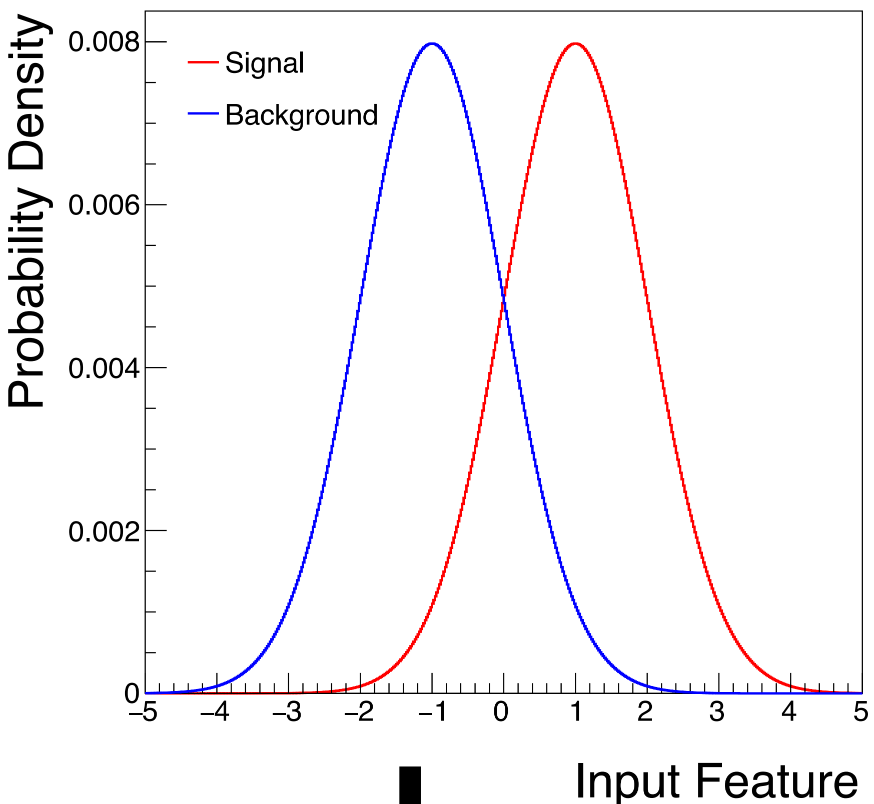


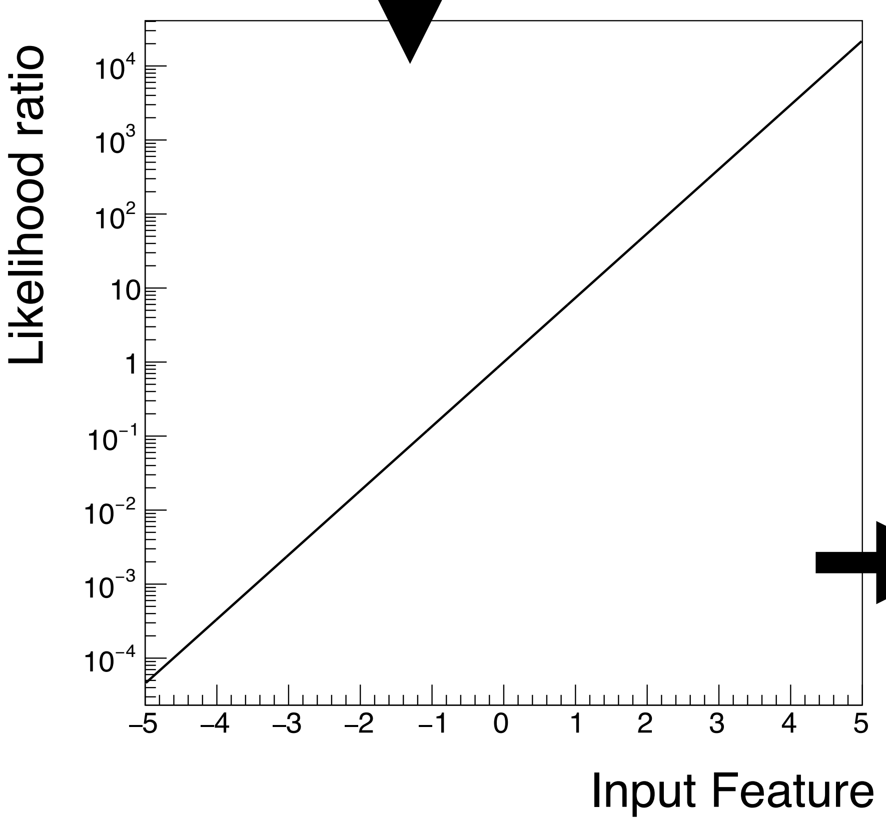


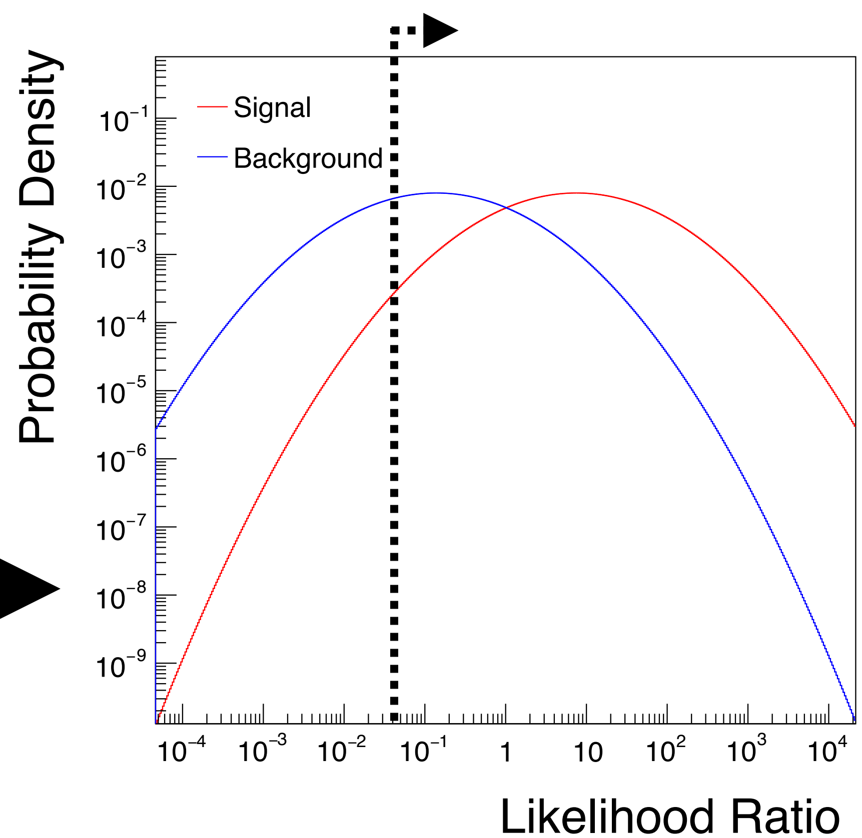
Optimal Classification with ML

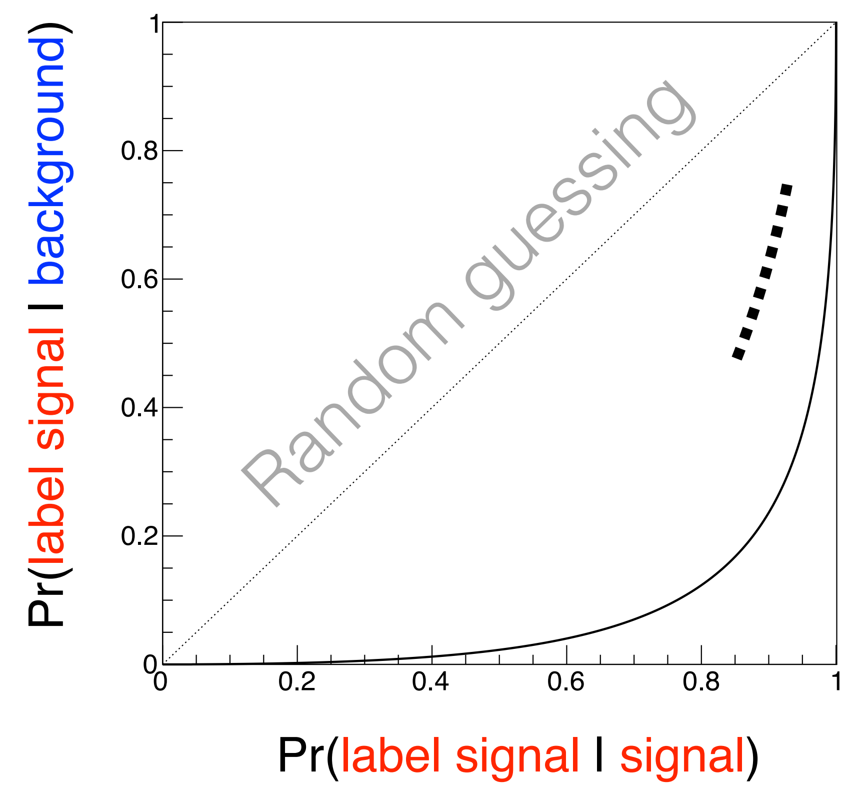
Think of events as e.g. an image

Train a classifier to distinguish known physics ("**background**") from new particles ("**signal**")

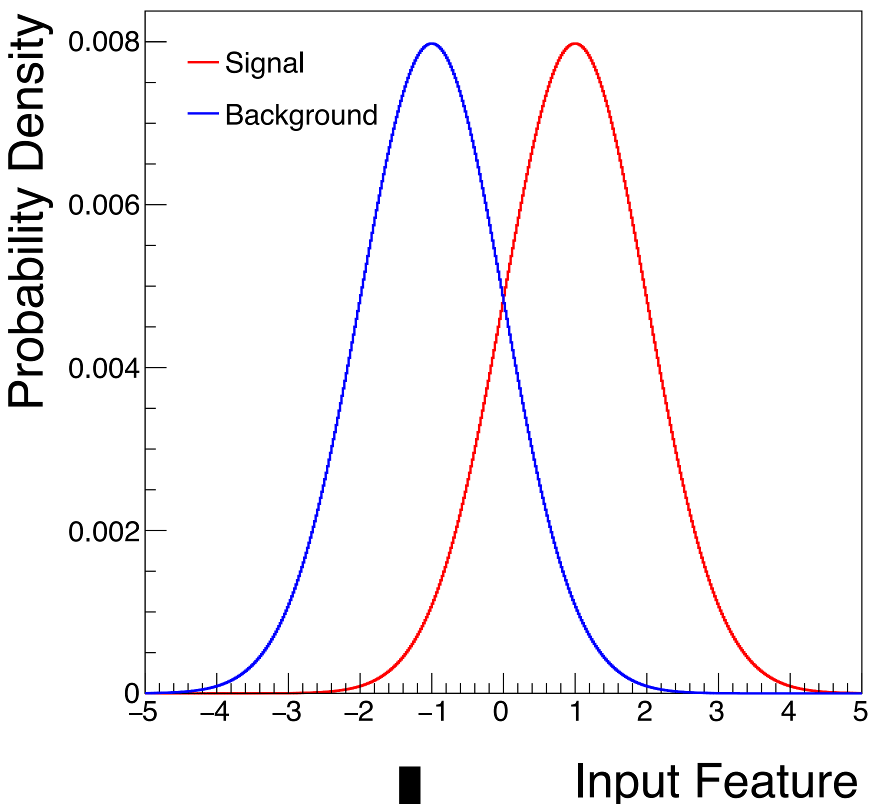






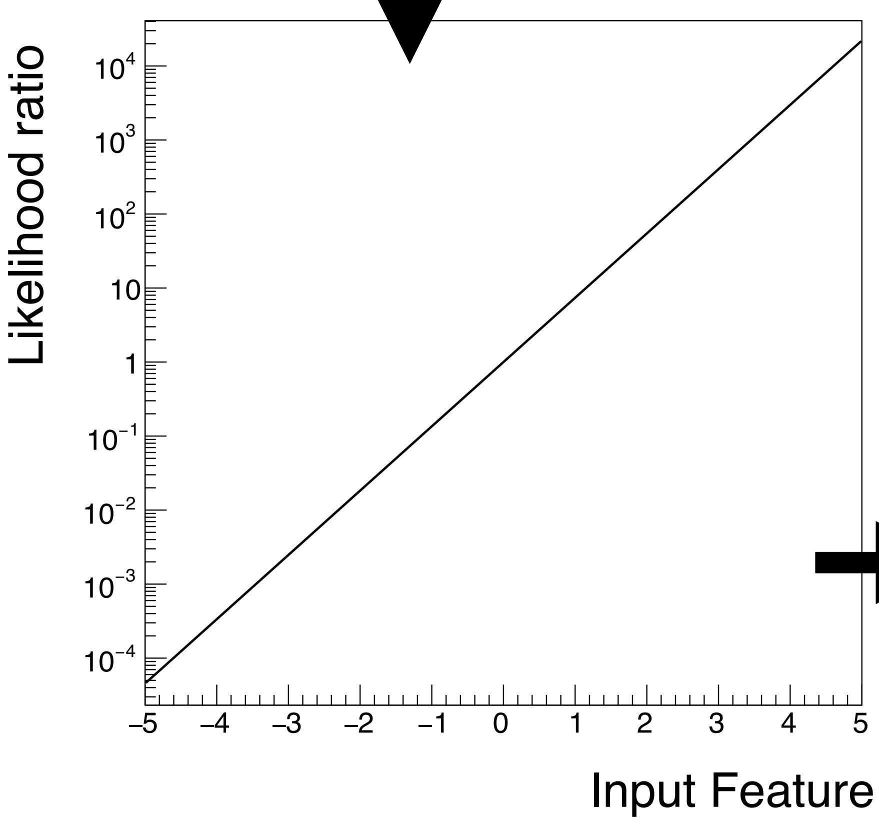


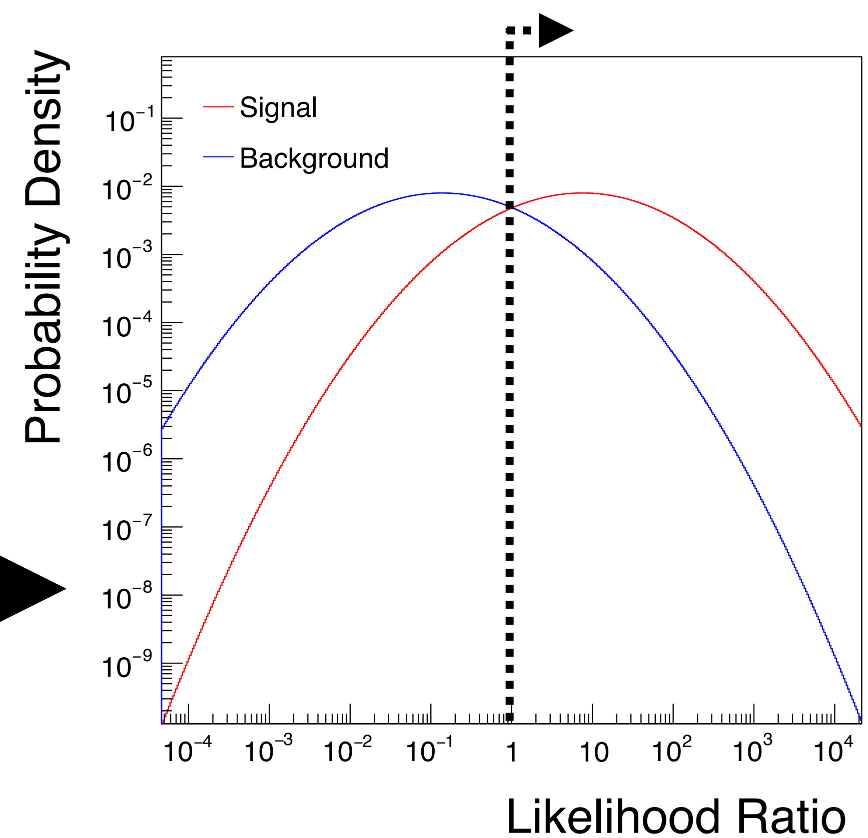
Optimal Classification with ML

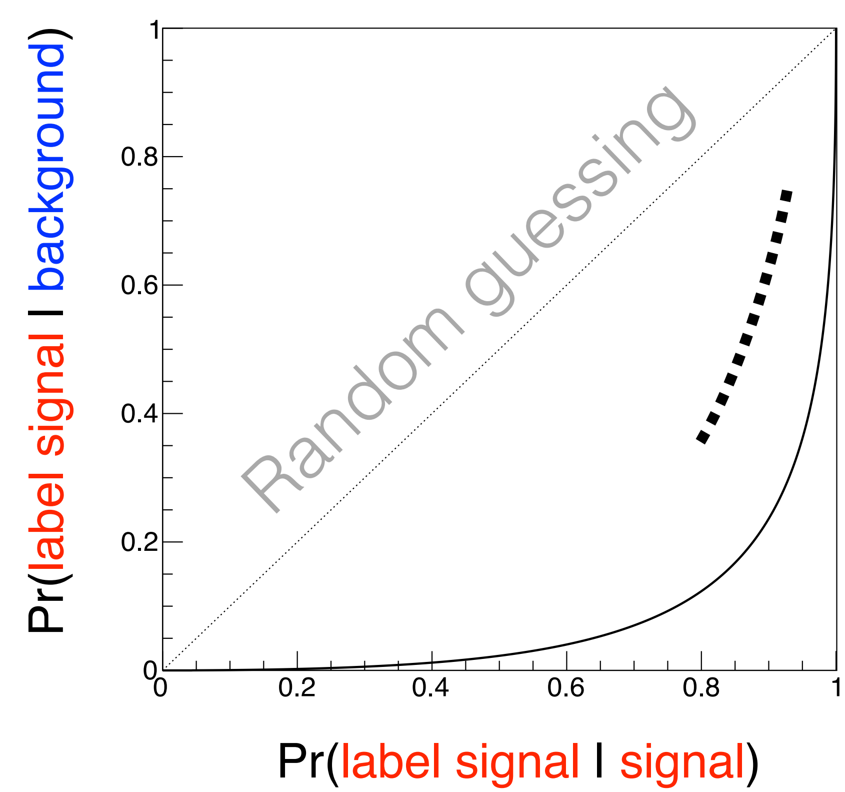


Think of events as e.g. an image

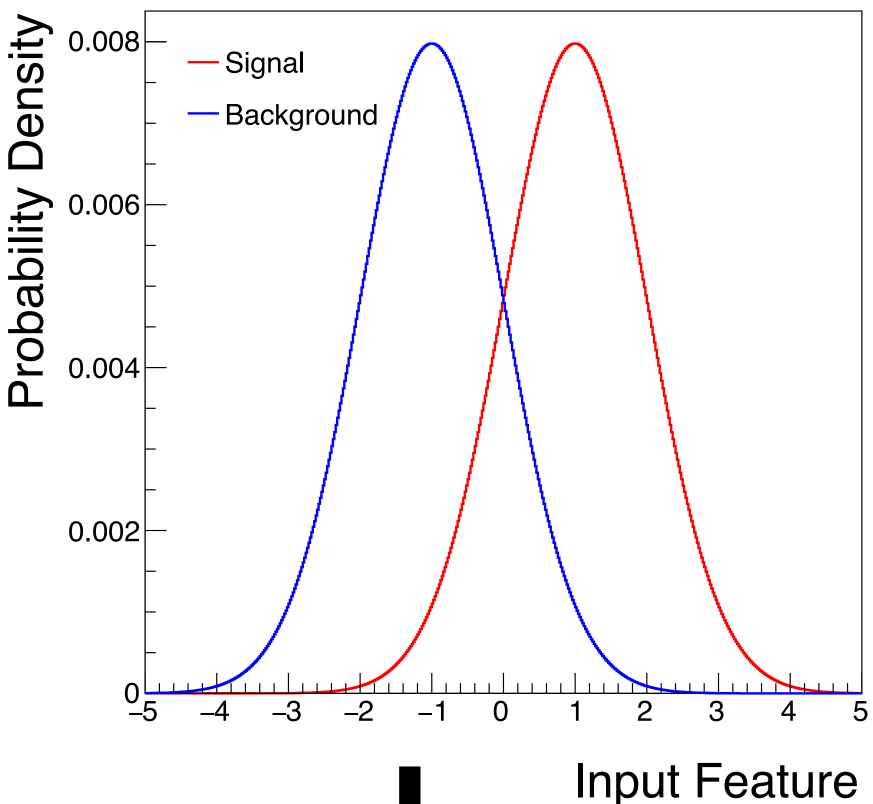
Train a classifier to distinguish known physics ("**background**") from new particles ("**signal**")





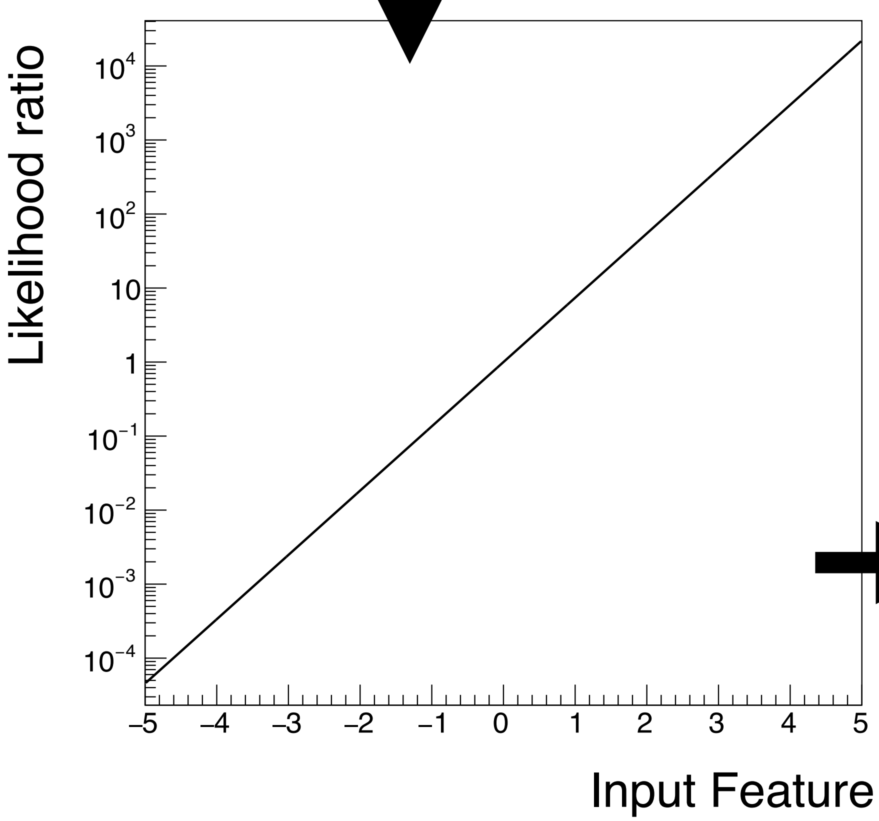


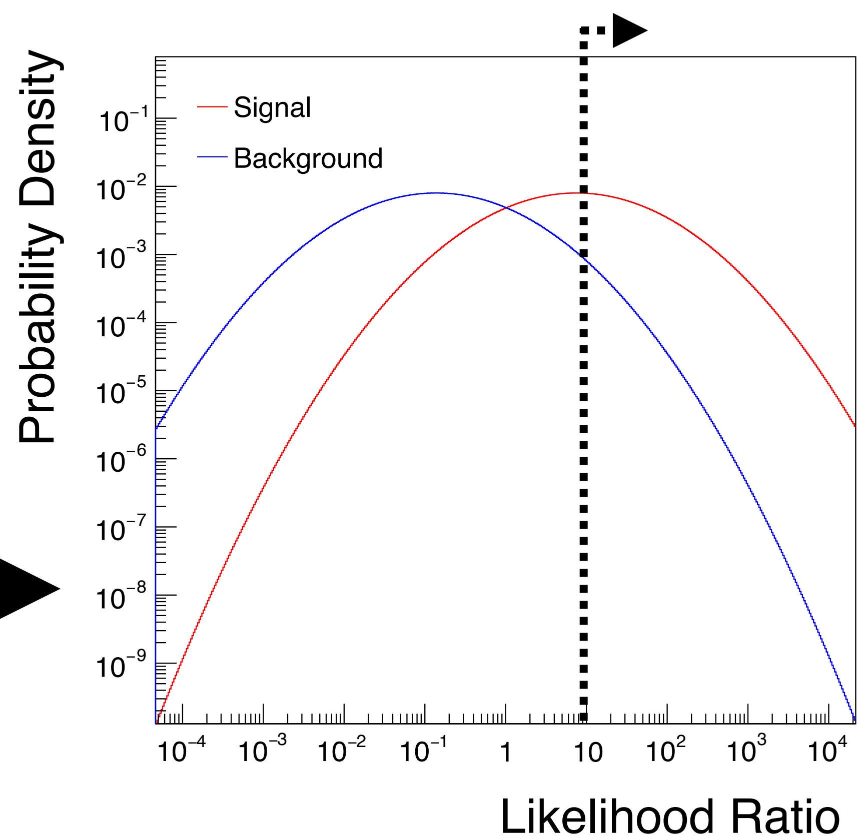
Optimal Classification with ML

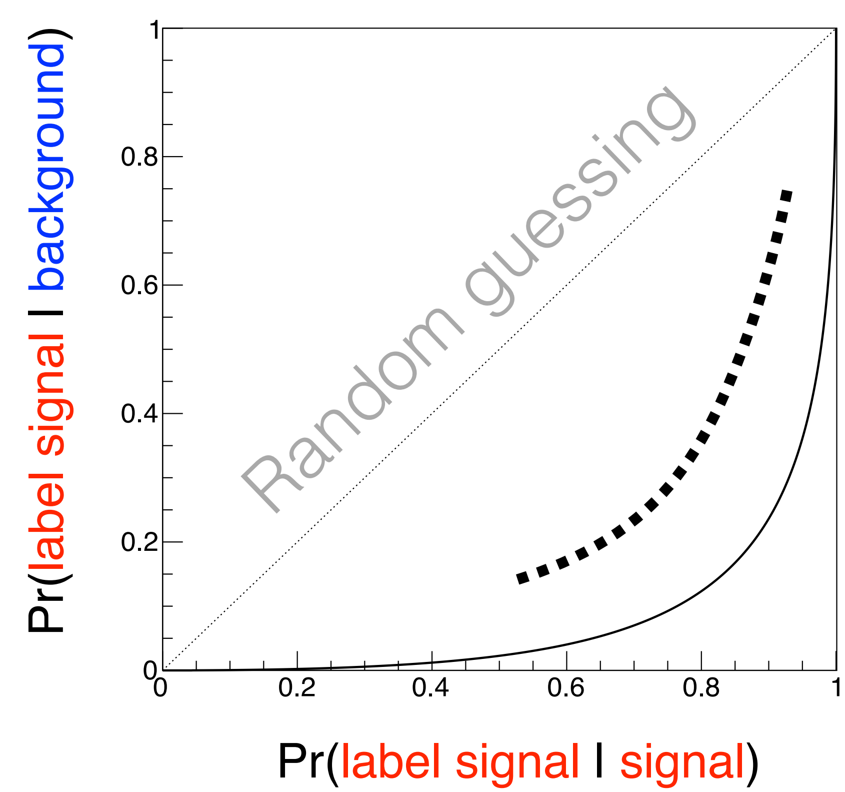


Think of events as e.g. an image

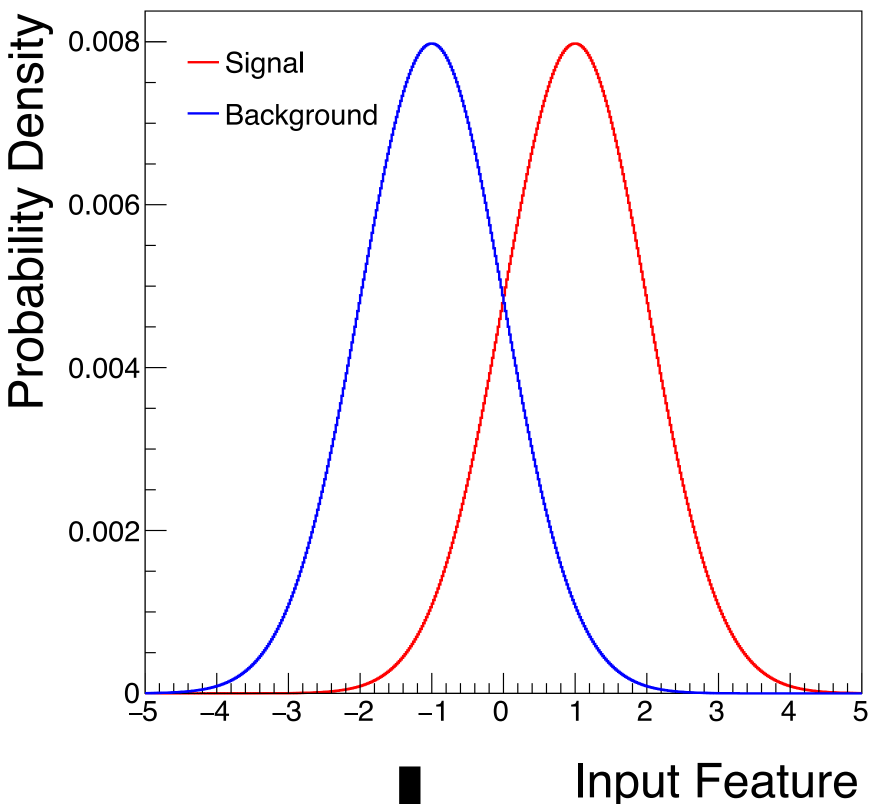
Train a classifier to distinguish known physics ("**background**") from new particles ("**signal**")





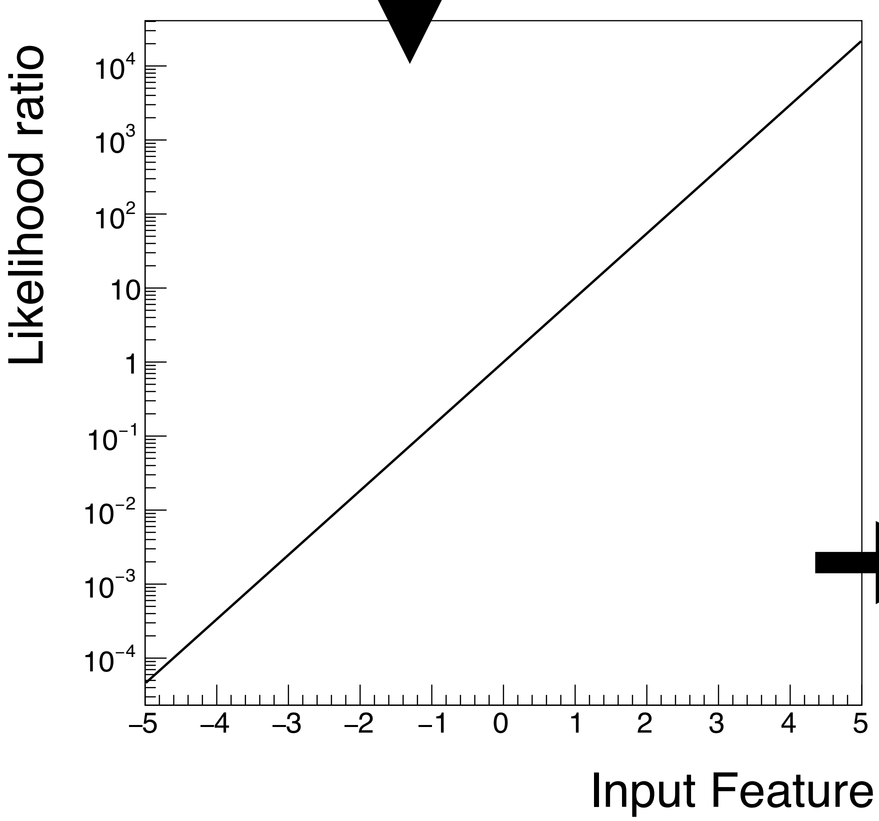


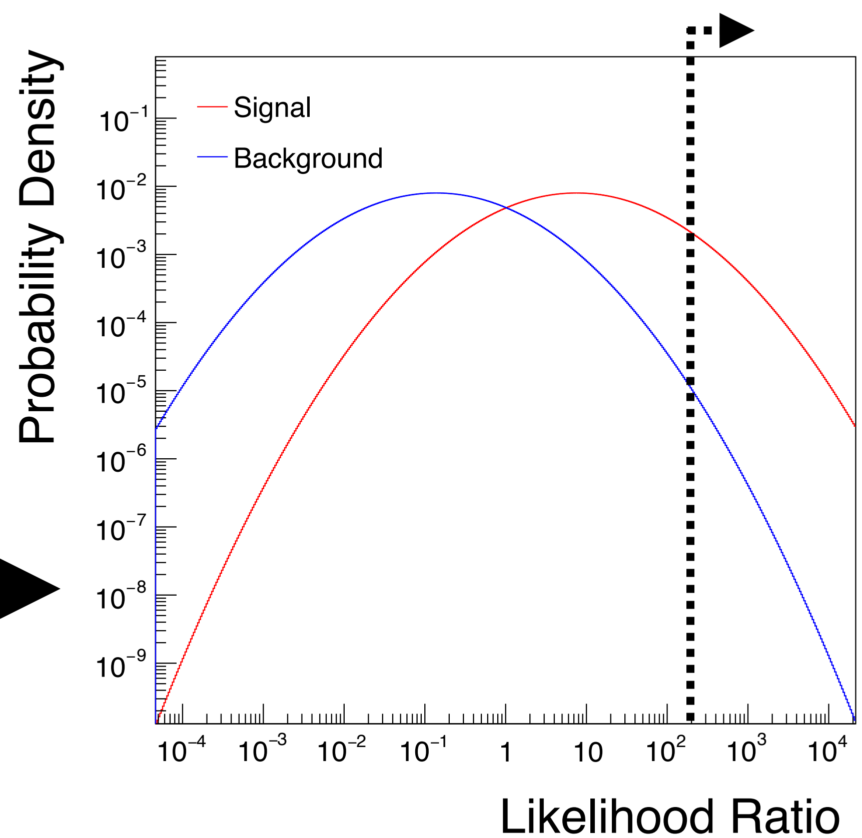
Optimal Classification with ML

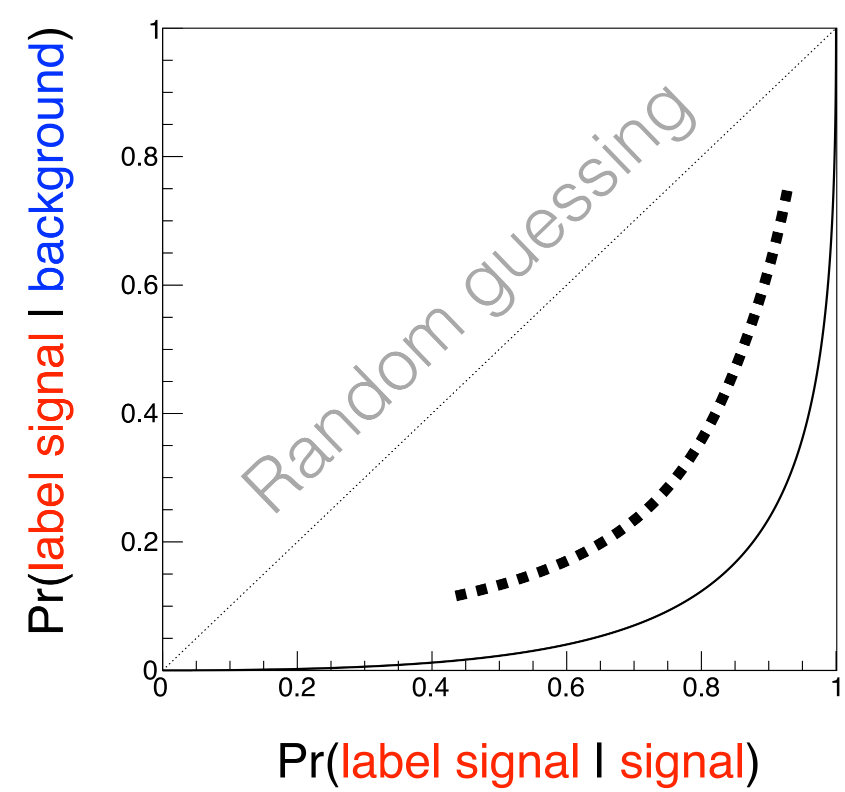


Think of events as e.g. an image

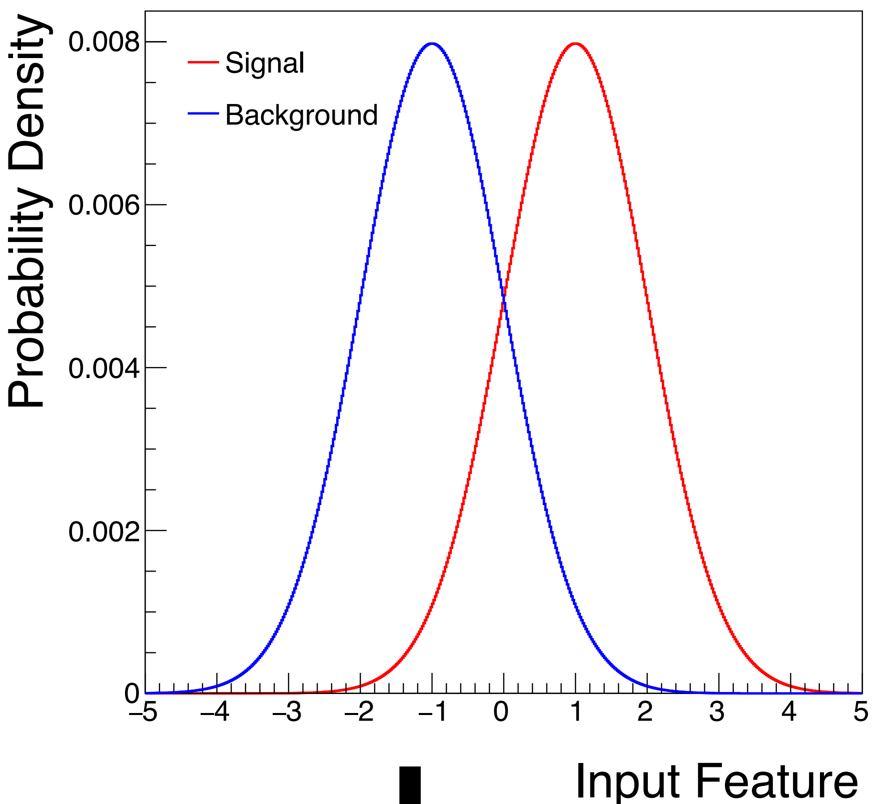
Train a classifier to distinguish known physics (“**background**”) from new particles (“**signal**”)





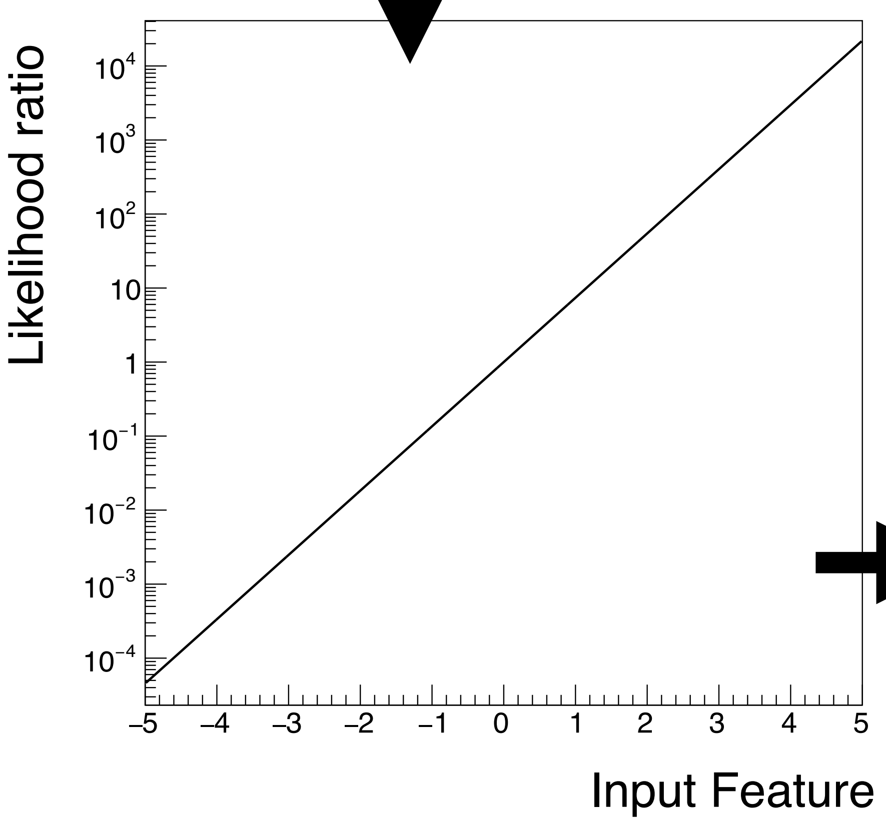


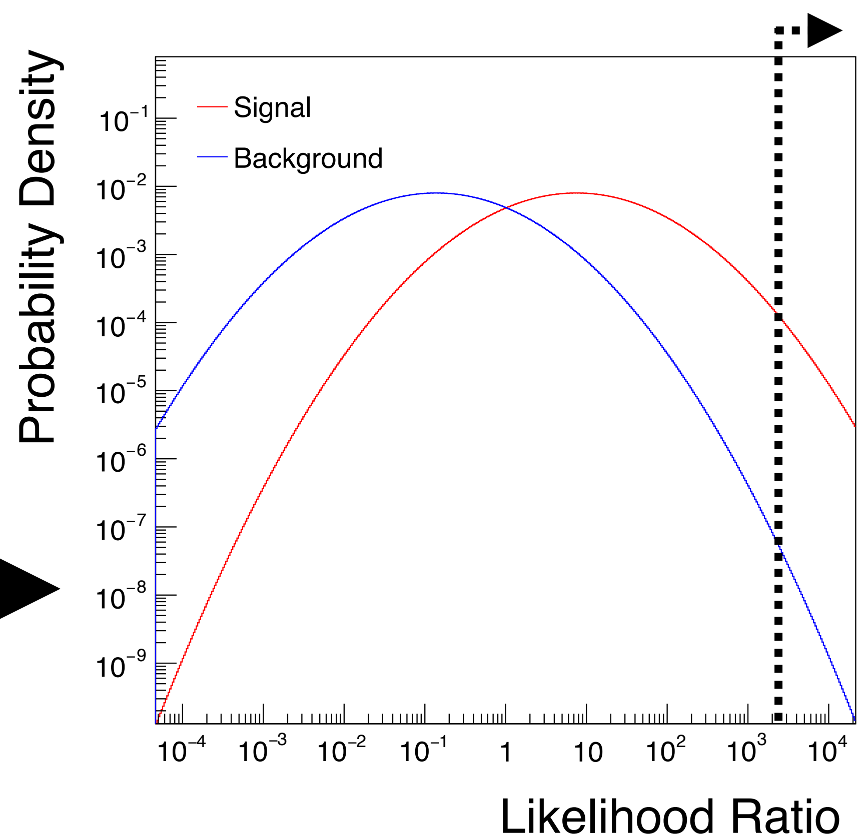
Optimal Classification with ML

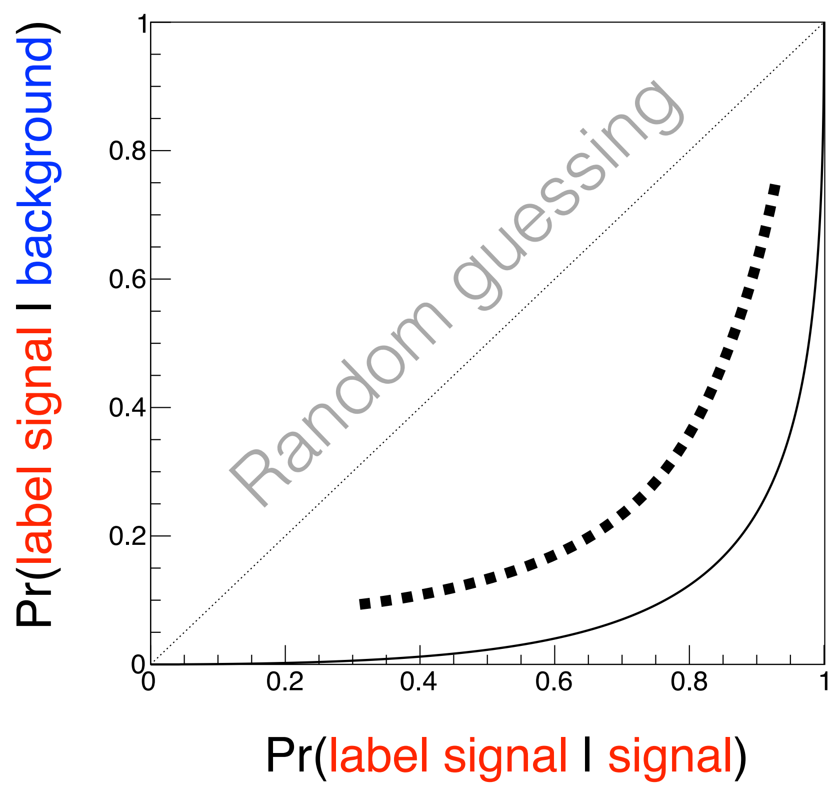


Think of events as e.g. an image

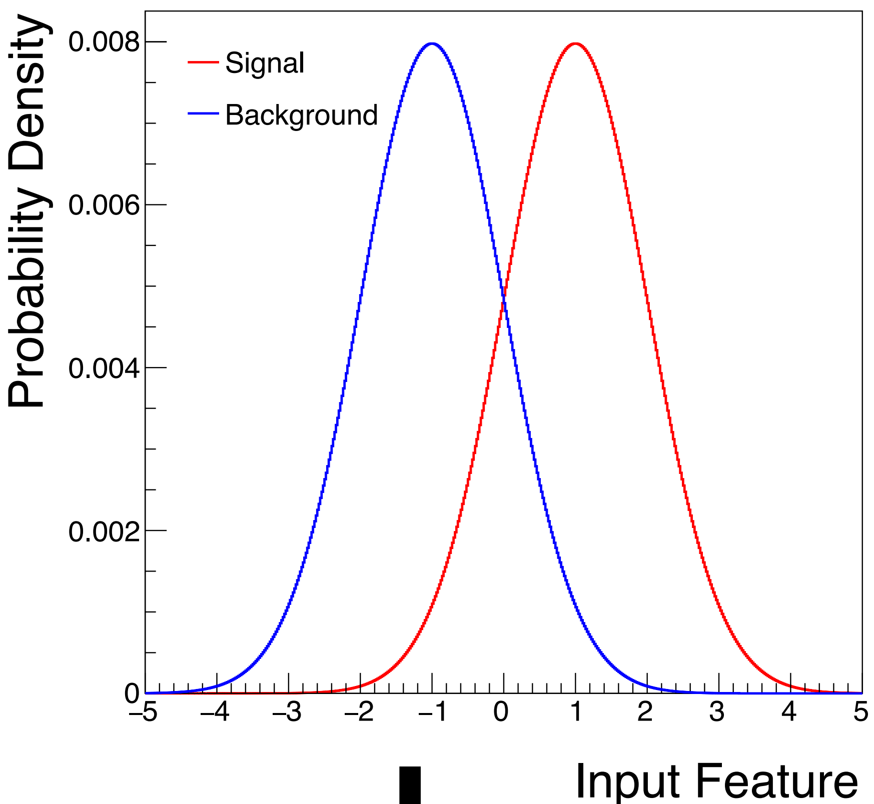
Train a classifier to distinguish known physics (“**background**”) from new particles (“**signal**”)





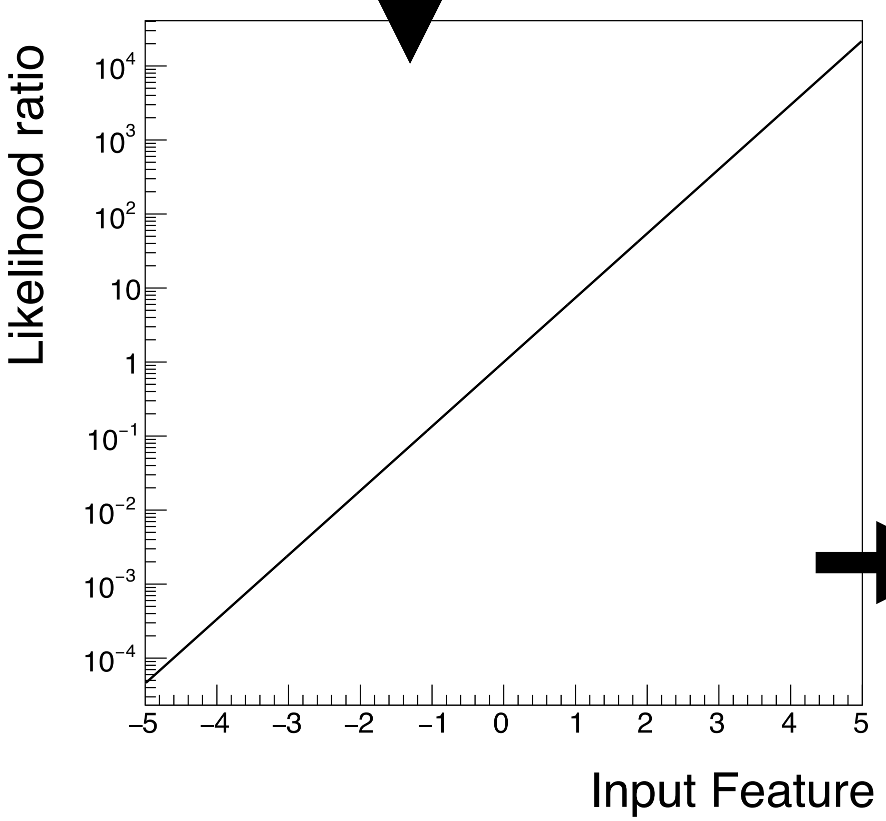


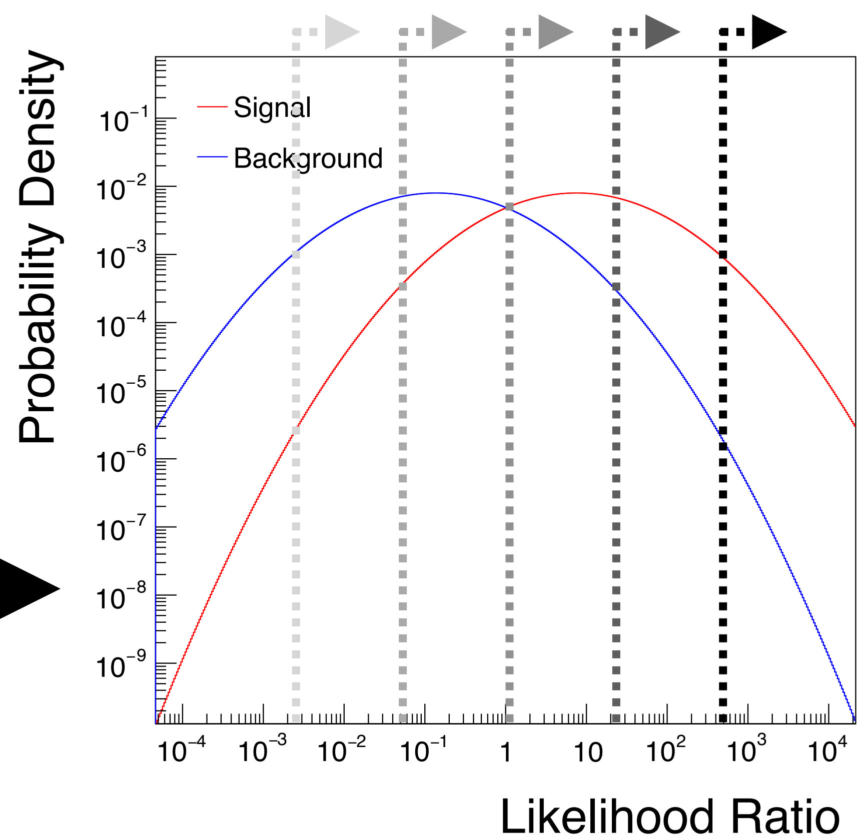
Optimal Classification with ML

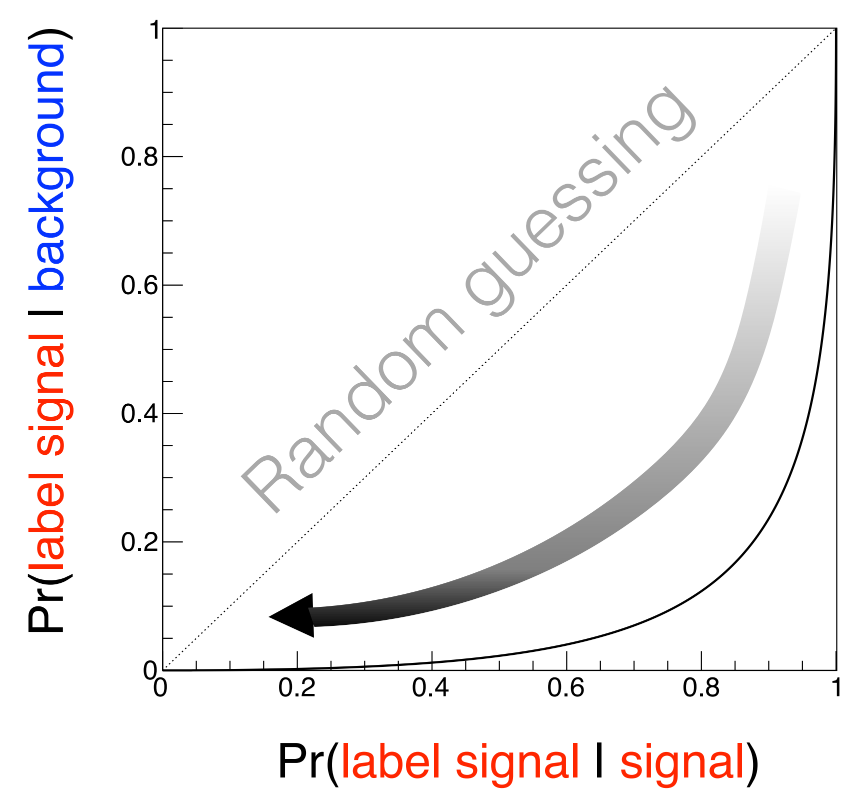


Think of events as e.g. an image

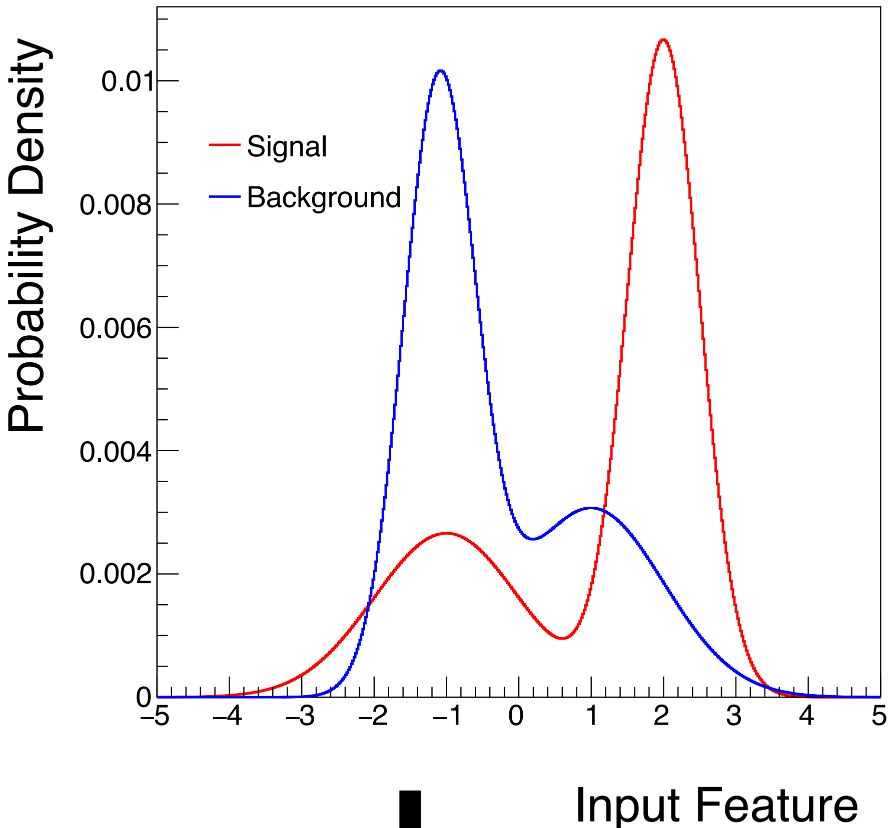
Train a classifier to distinguish known physics (“**background**”) from new particles (“**signal**”)





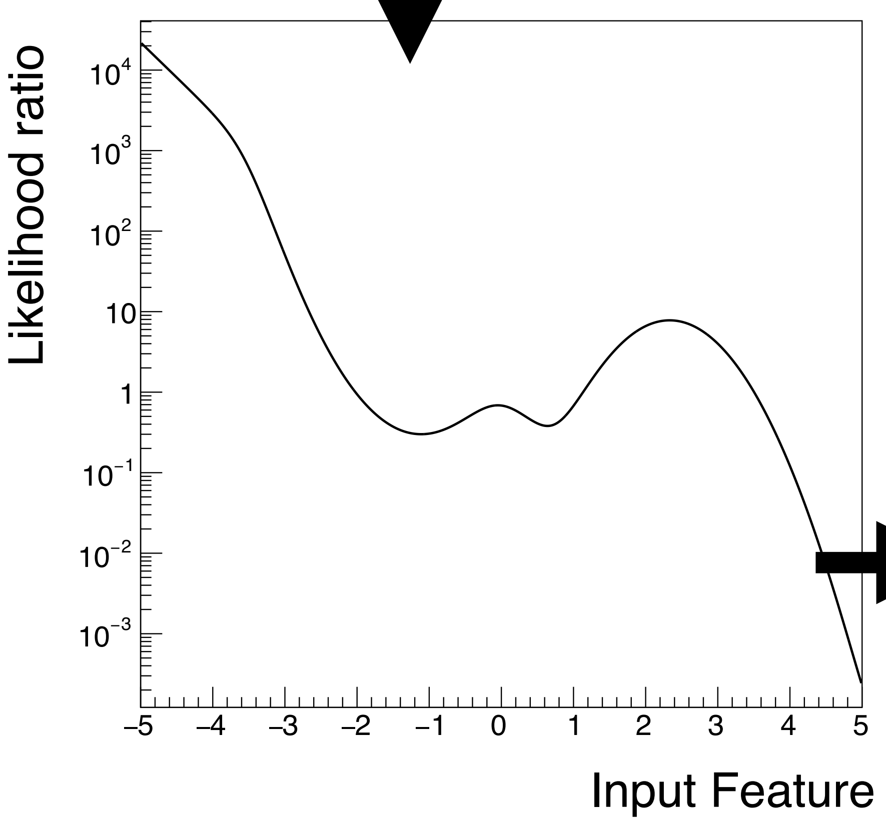


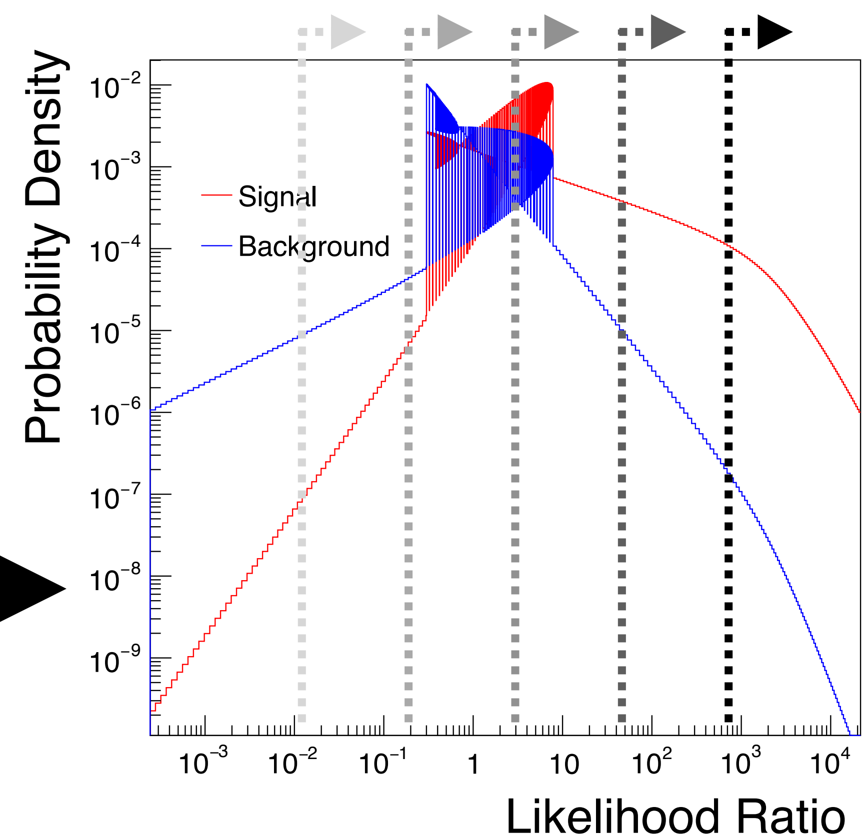
Optimal Classification with ML

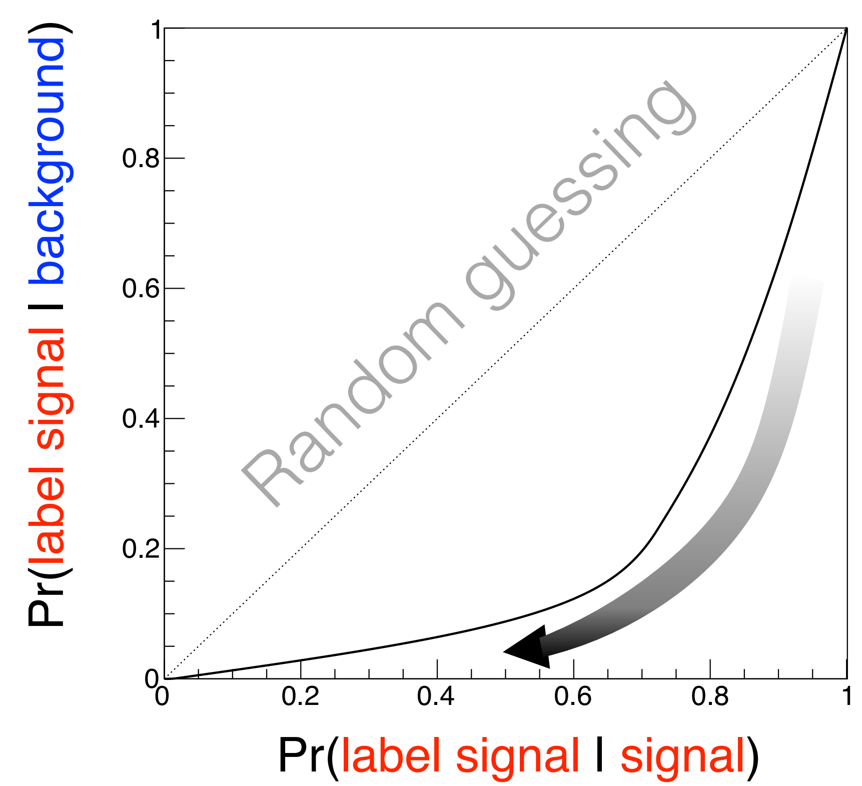


Think of events as e.g. an image

Train a classifier to distinguish known physics (“**background**”) from new particles (“**signal**”)





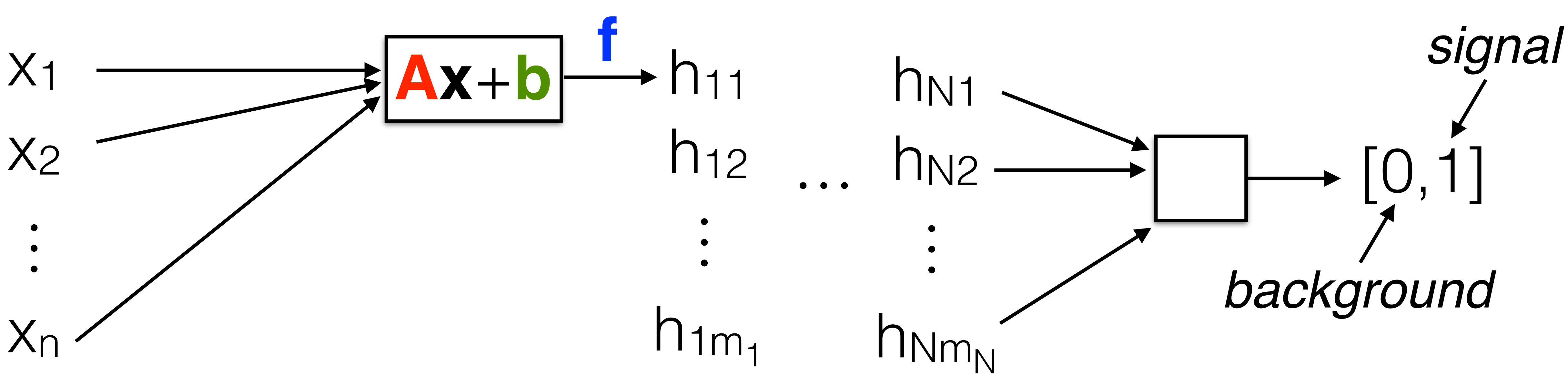


Deep Neural Networks

This gets out of hand quickly … use a neural network to do an unbanned version in high-dimensions!

Think of events as e.g. an image

Train a classifier to distinguish known physics ("**background**") from new particles ("**signal**")



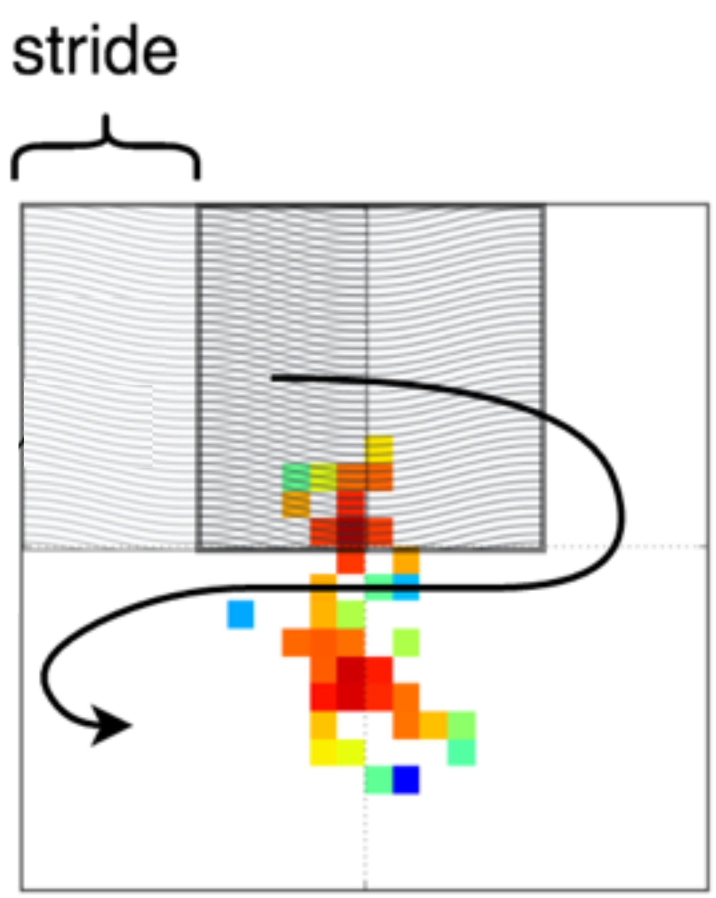
Deep Neural Networks

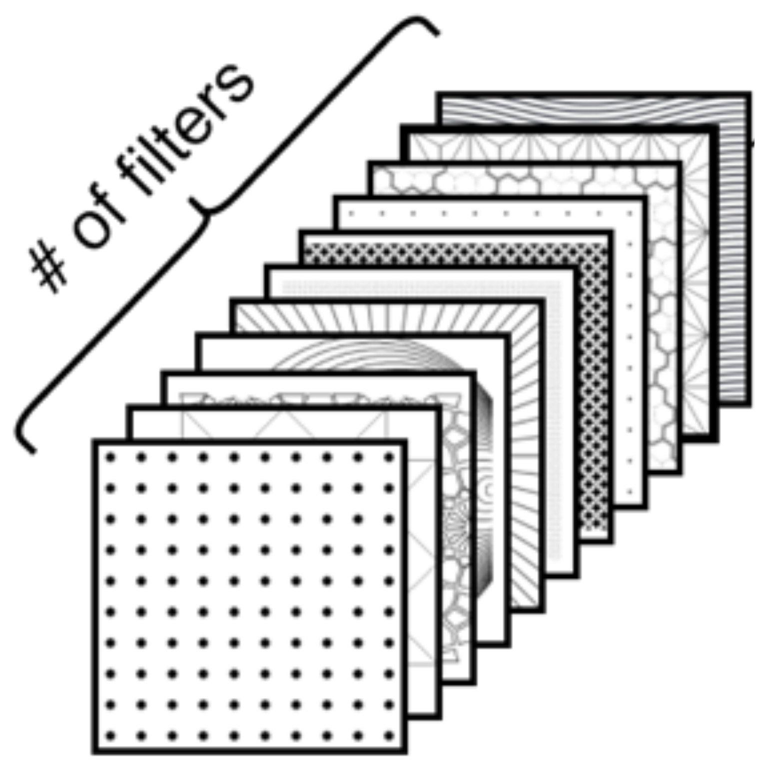
For images, often use a convolution neural network (CNN)

Think of events as e.g. an image

↓

Train a classifier to distinguish known physics ("**background**") from new particles ("**signal**")

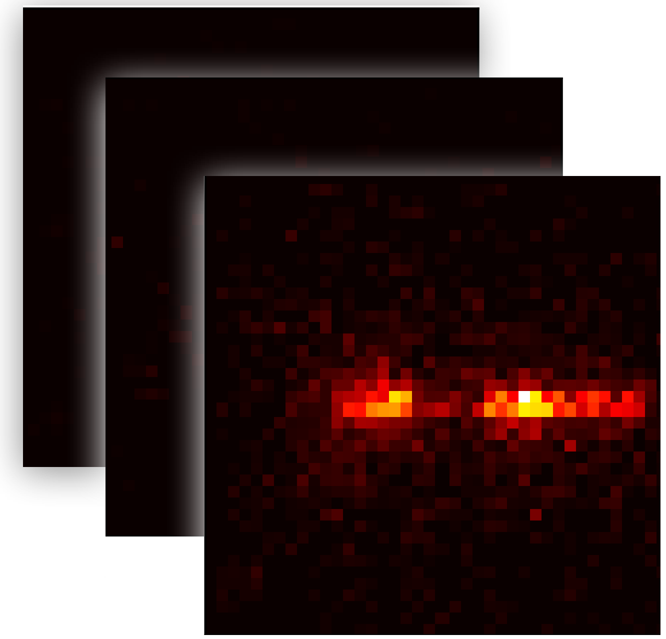




Even if there is no translational symmetry, overall fewer parameters

Applying ML to HEP Images

Deep learning is now widely used to do what shallow NN's and BDTs have been doing. DNN's allow us to use all of the available information.



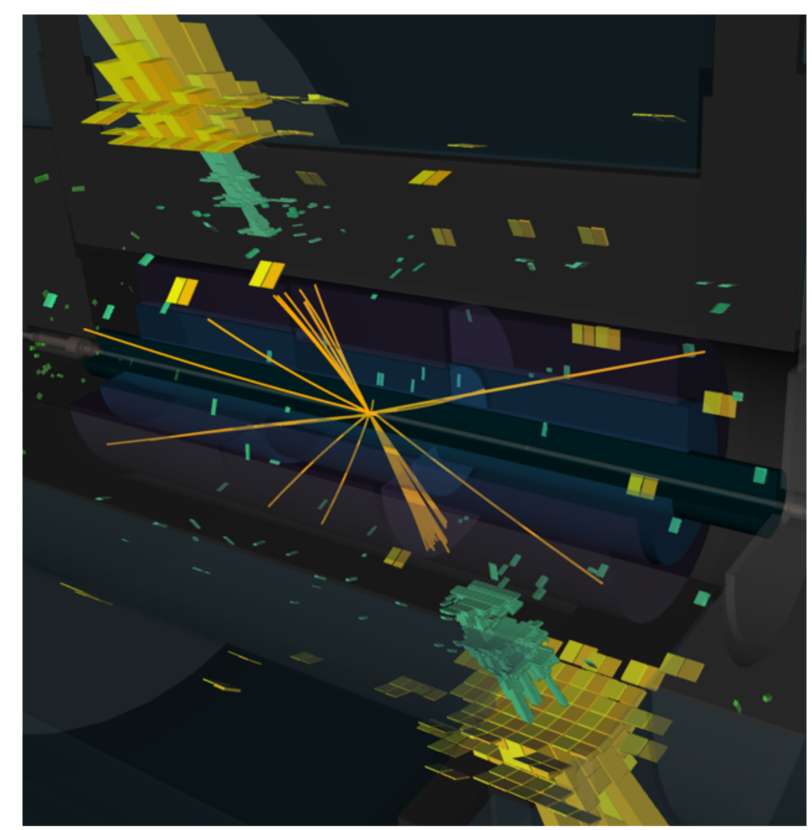
Classification

provide examples for training

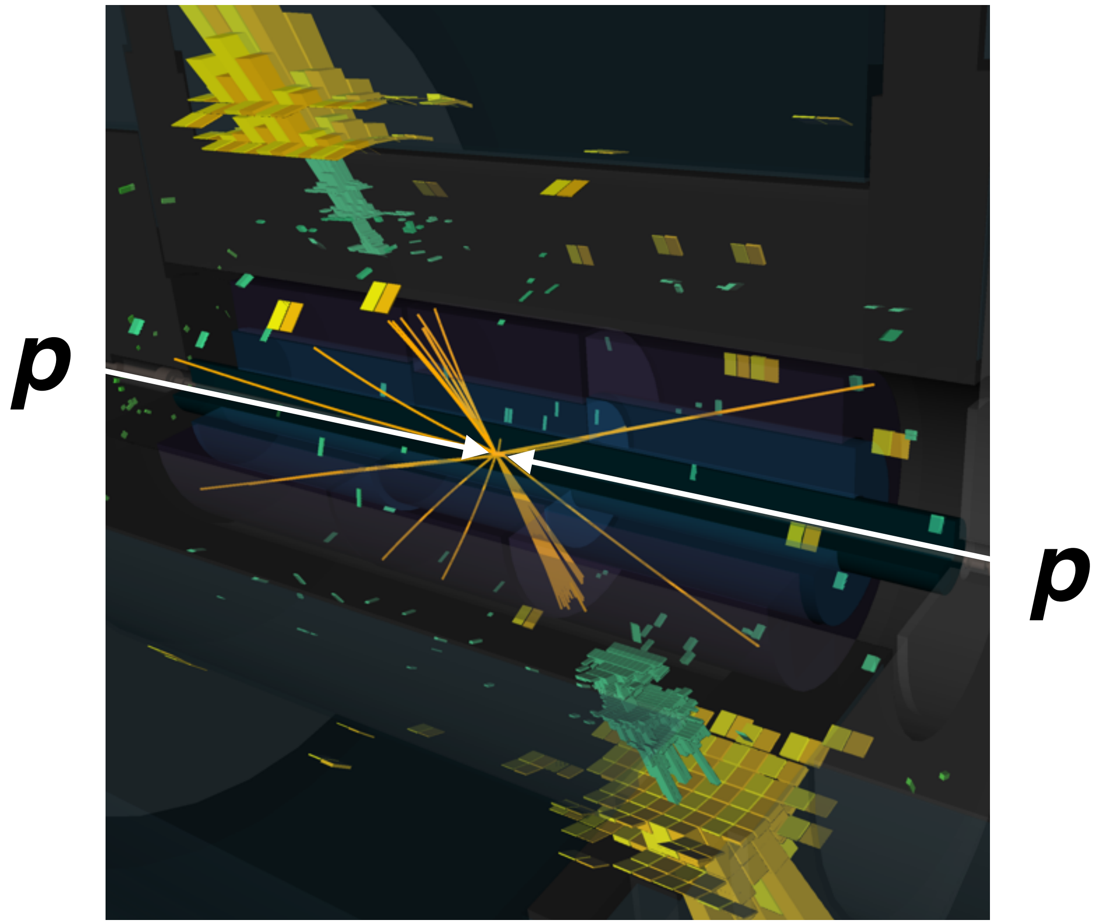
arbitrarily many categories

Generation ← **Regression**

map noise to structure



Classification of HEP images



Yellow, **green** = active pixels

Think of events as e.g. an image

↓

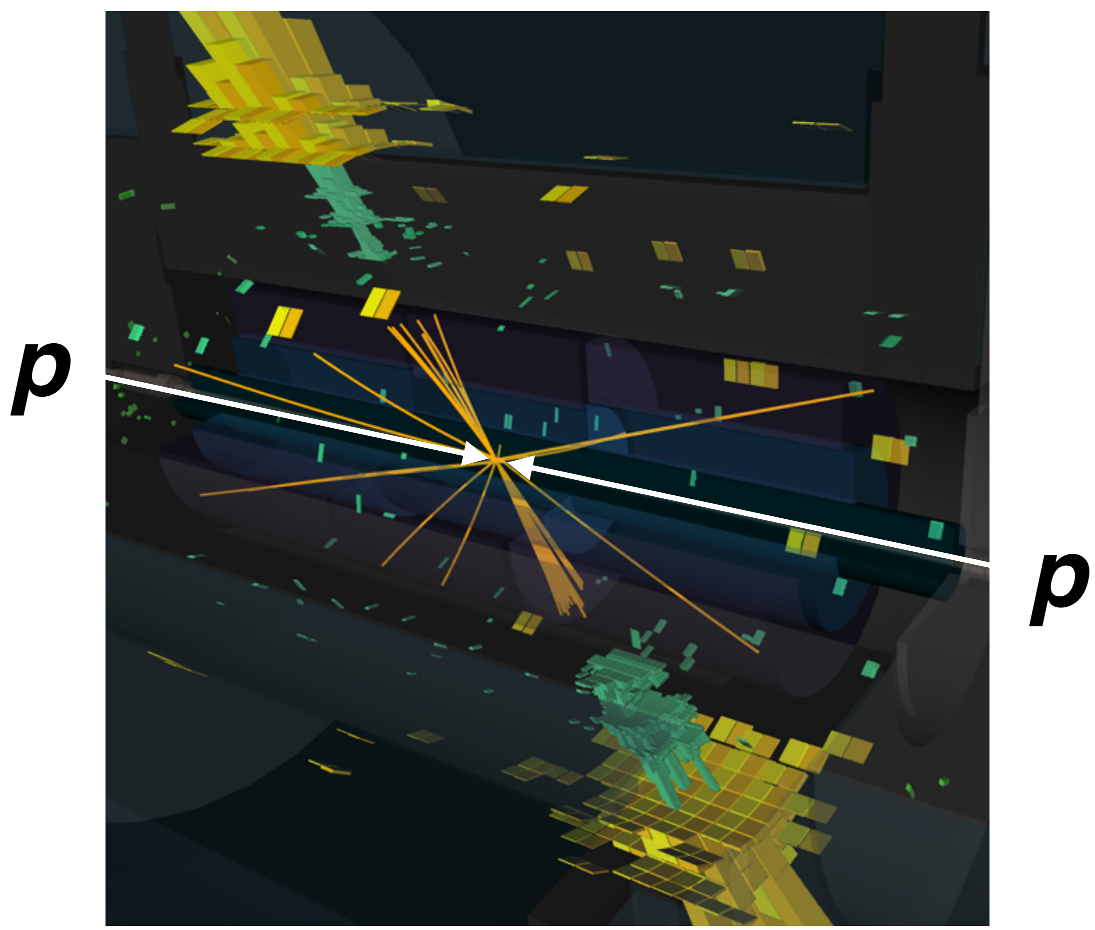
Train a classifier to distinguish known physics (“**background**”) from new particles (“**signal**”)

↓

Apply classifier to data

A NN is a(n analytic!) function from image to [0,1]

Classification of HEP images



Yellow, **green** = active pixels

Think of events as e.g. an image

↓

Train a classifier to distinguish known physics (“**background**”) from new particles (“**signal**”)

↓

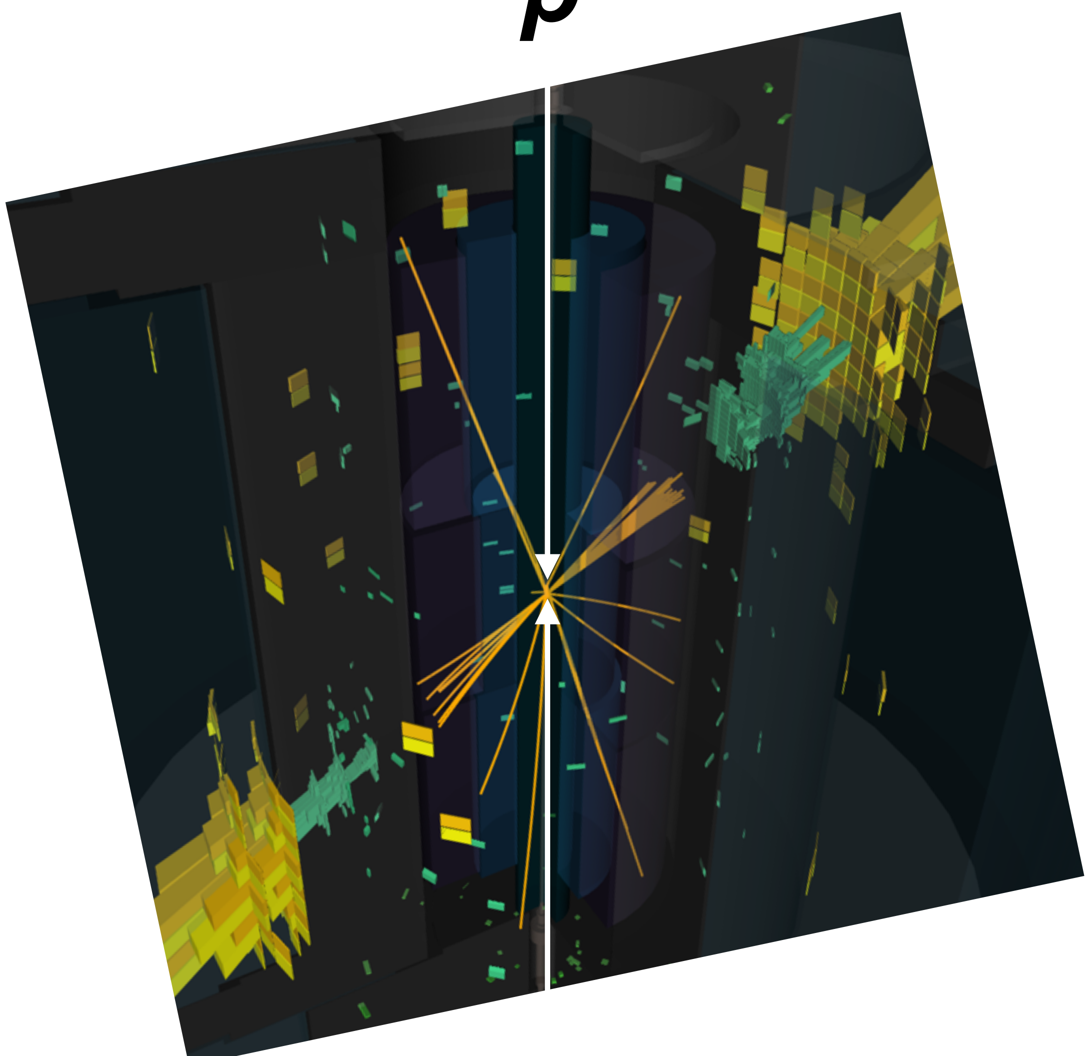
Apply classifier to data

↓

With improved signal-to-noise, increase sensitivity!

Classification of HEP images

p



p

Yellow, **green**
= active pixels

Think of events as e.g. an image

↓

Train a classifier to distinguish known physics ("**background**") from new particles ("**signal**")

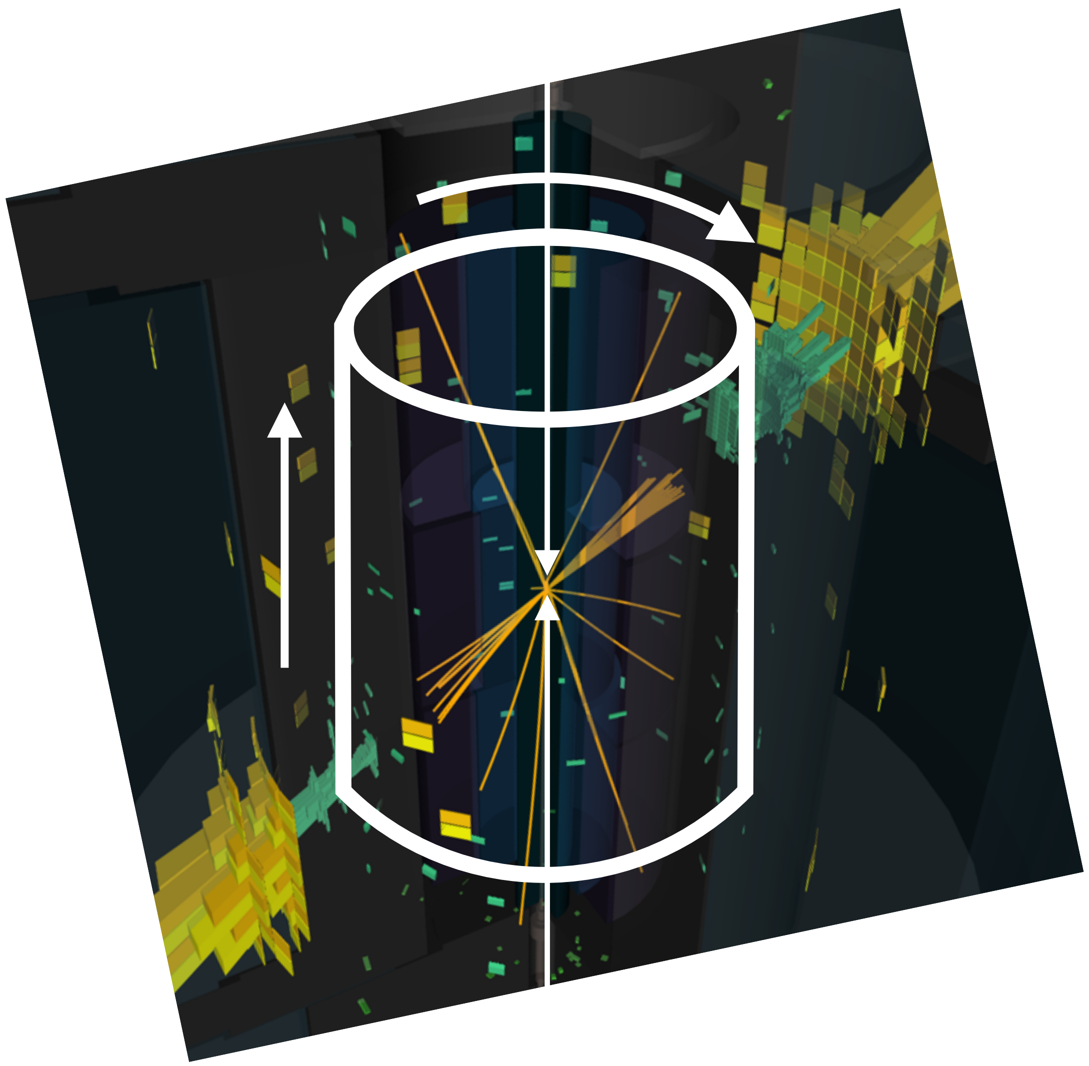
↓

Apply classifier to data

↓

With improved signal-to-noise, increase sensitivity!

Classification of HEP images



Yellow, **green**
= active pixels

Think of events as e.g. an image

↓

Train a classifier to distinguish known physics ("**background**") from new particles ("**signal**")

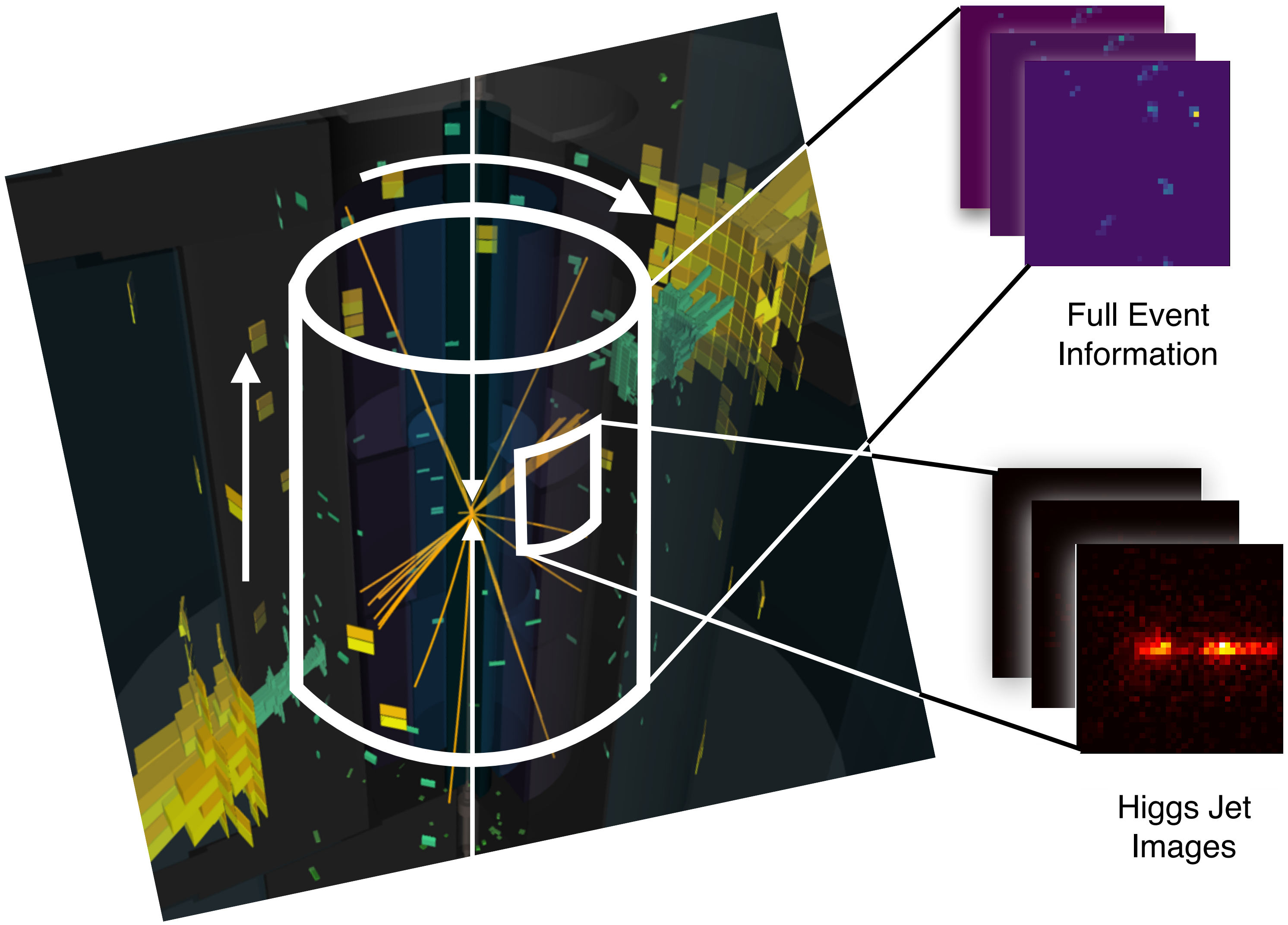
↓

Apply classifier to data

↓

With improved signal-to-noise, increase sensitivity!

Classification of HEP images



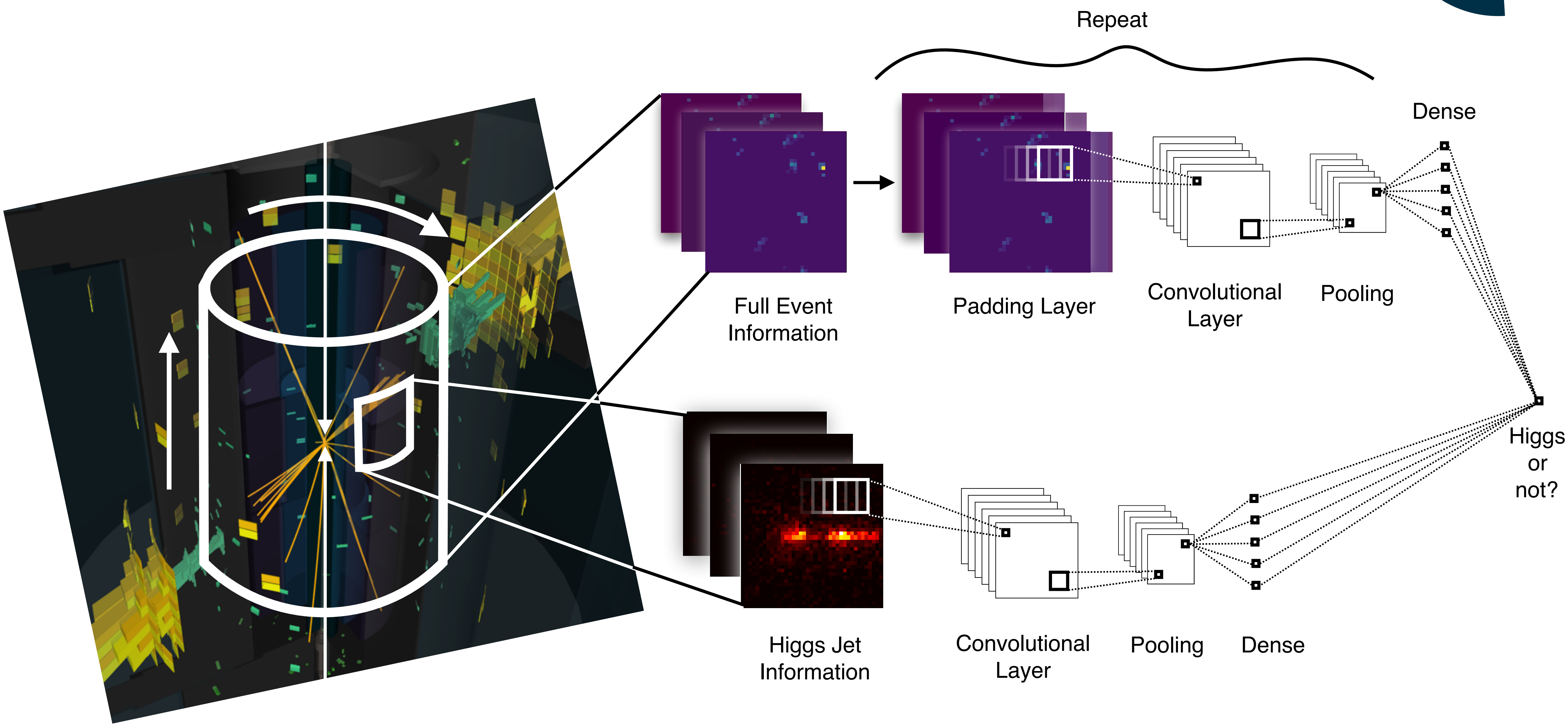
@LHC, surface of detector ~cylinder

Particles hitting detector can occur ~anywhere

Interplay between useful information at multiple length scales

Each image can encode information from different detectors, like RGB
(e.g. charged energy and neutral energy)

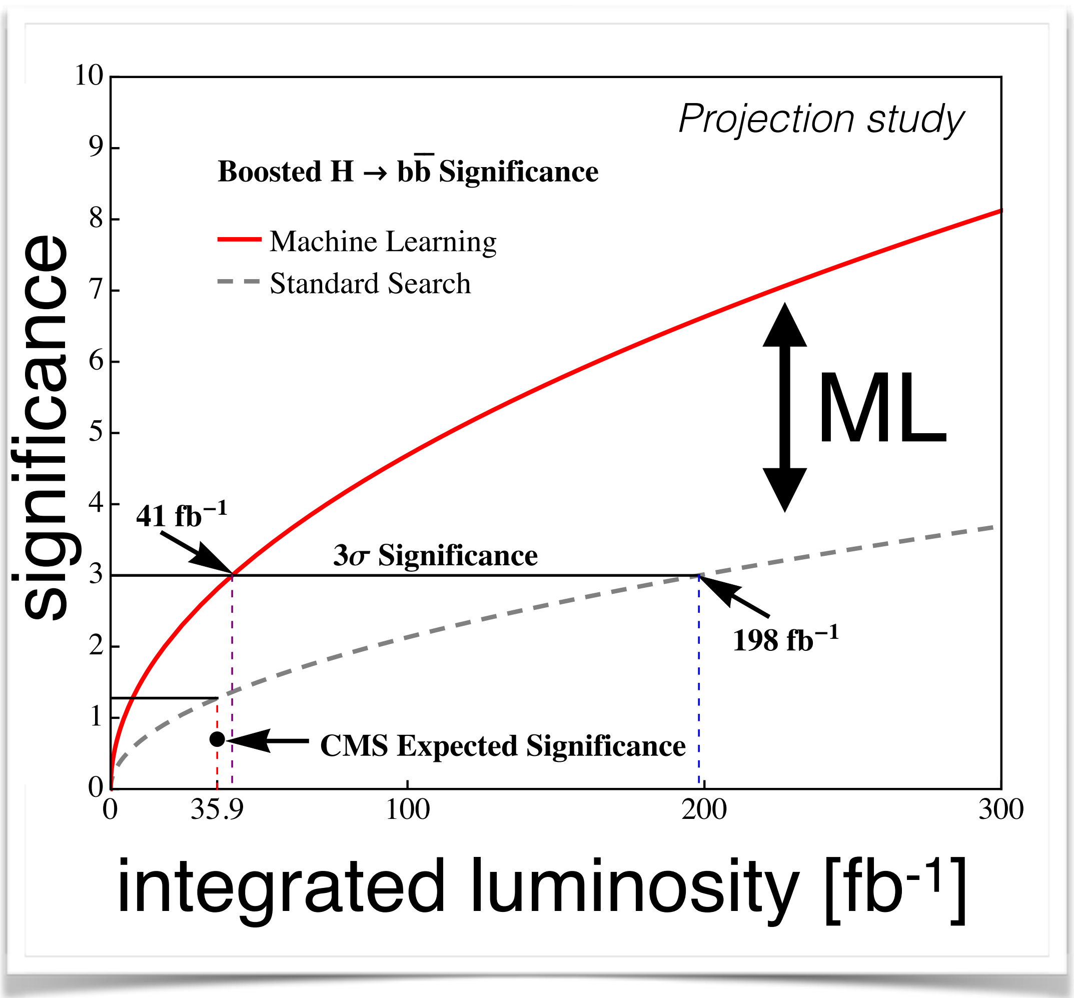
Classification of HEP images

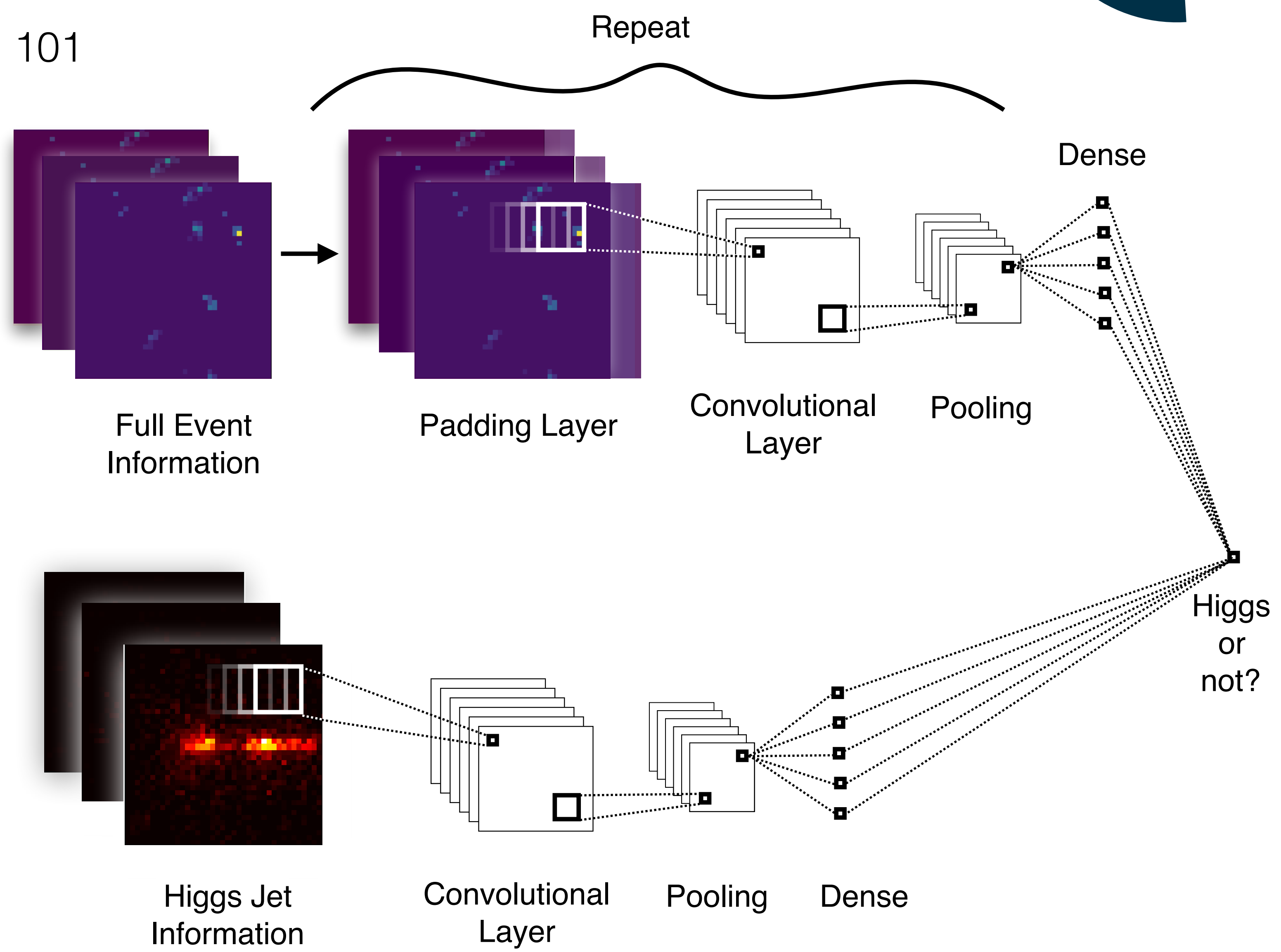


Deep CNNs can be used to pick out useful features in order to classify regions of interest / entire events as interesting (e.g. "Higgs Boson") or not.

Classification of HEP images

J. Lin, M. Freytsis, I. Moult, **BPN**, JHEP 10 (2018) 101

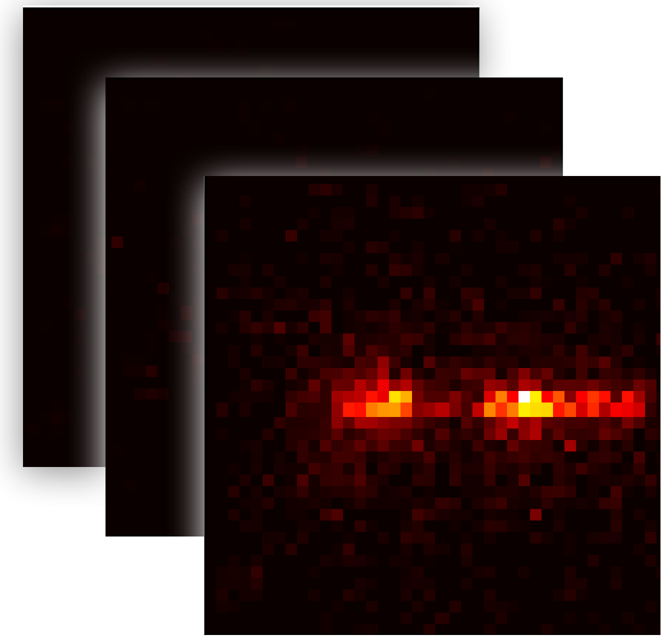




We have excellent baselines from physically-motivated algorithms. Still much to gain - critical to **understand what features NN is learning**!

ML beyond classification

Regression is nothing more than classification with many categories.



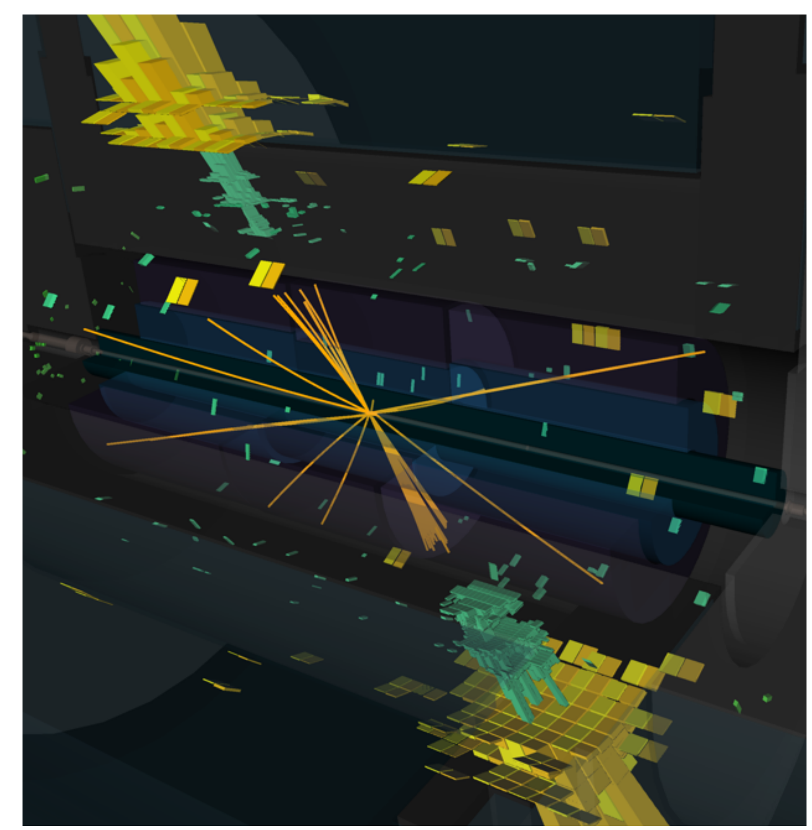
Classification

provide examples for training

arbitrarily many categories

Generation ← Regression

map noise to structure



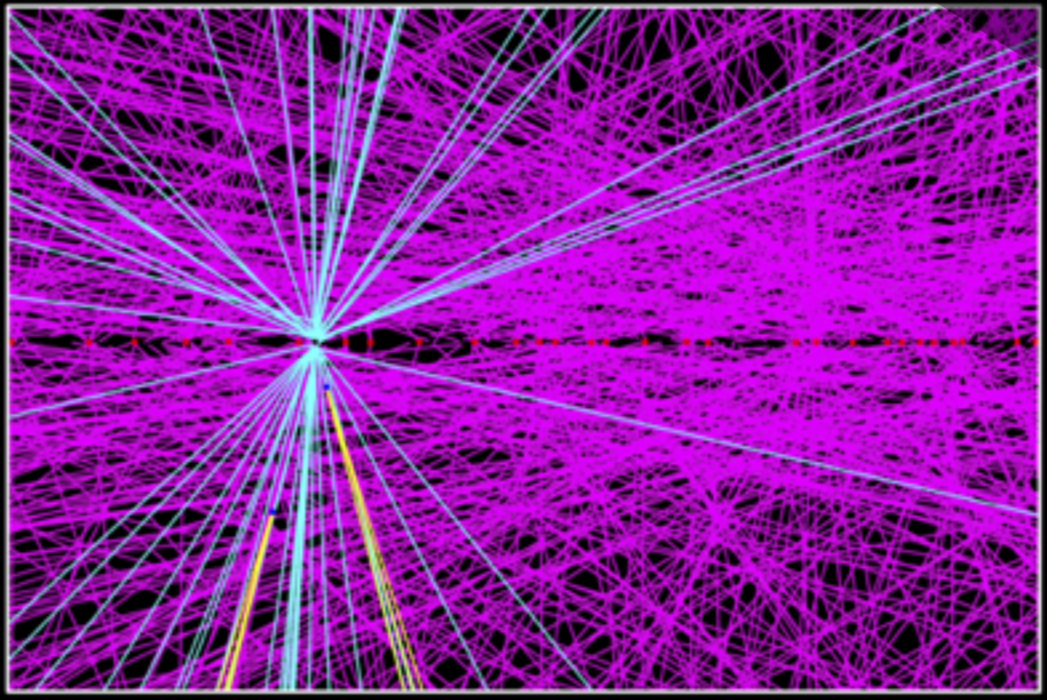
Regression with HEP images: denoising

ATLAS EXPERIMENT

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

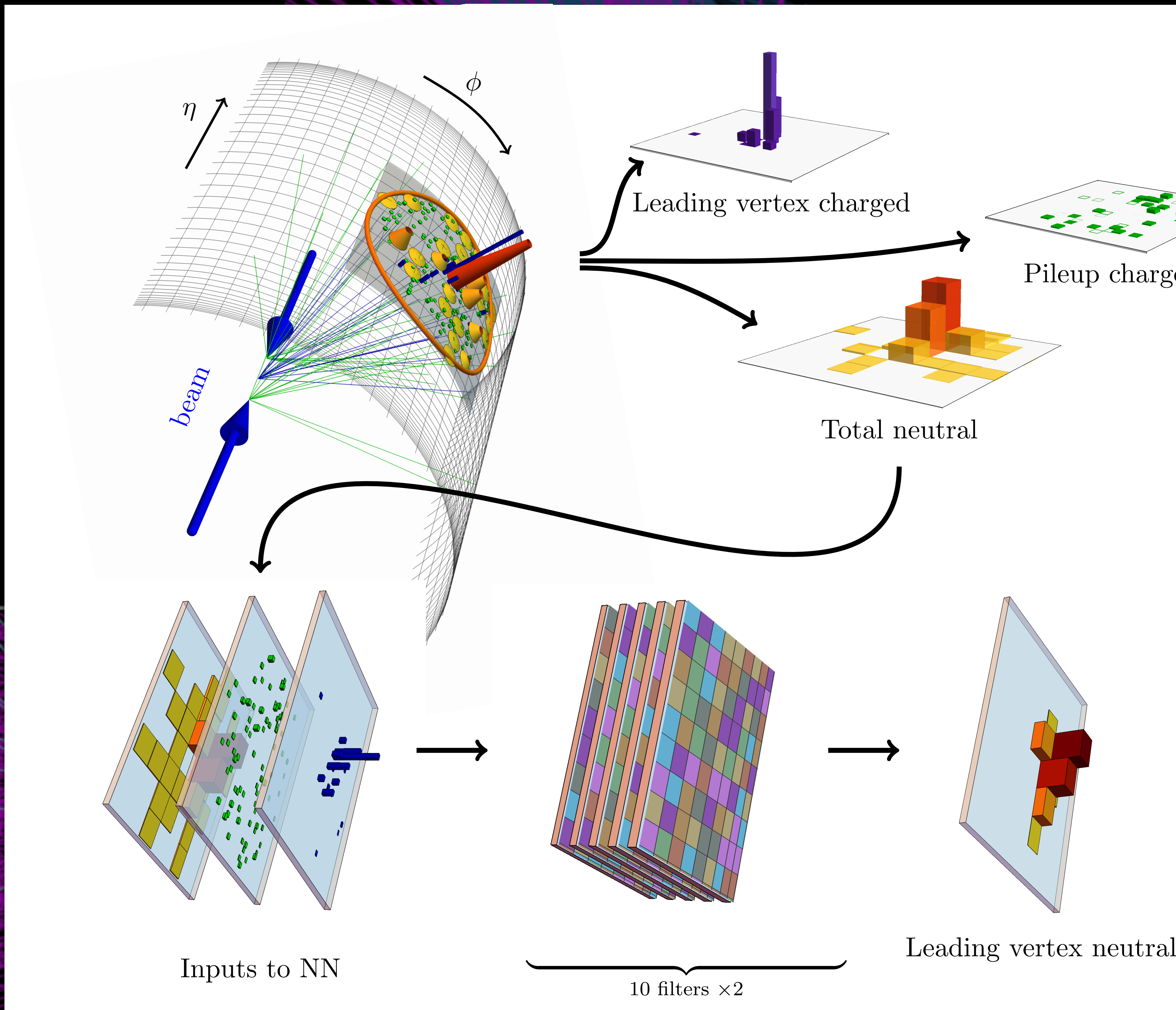
pp collisions at the LHC don't happen one at a time!

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



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

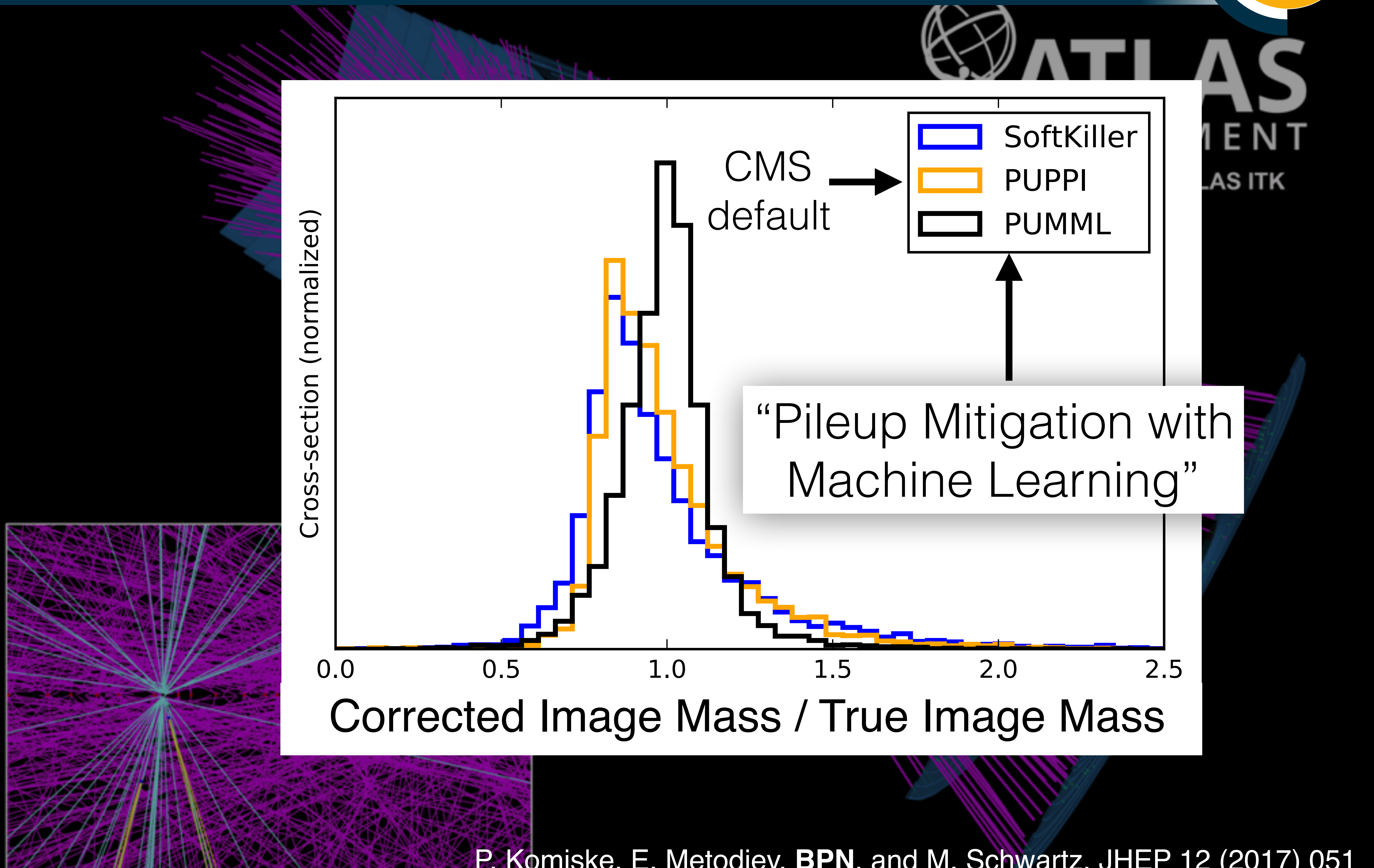
Regression with HEP images: denoising



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

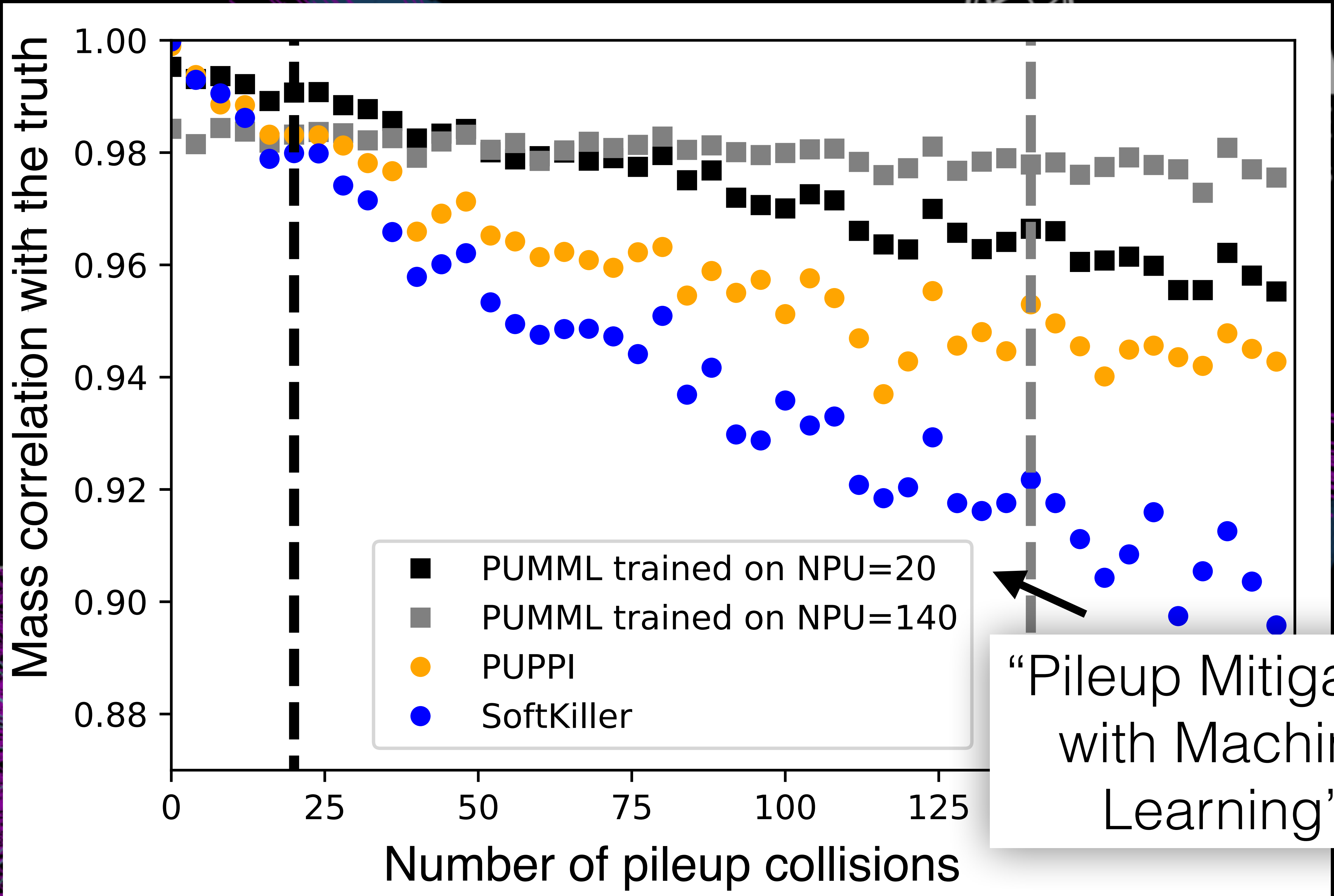
…also a natural application of convolutional NNs!

Regression with HEP images: denoising



P. Komiske, E. Metodiev, **BPN**, and M. Schwartz, JHEP 12 (2017) 051

Regression with HEP images: denoising

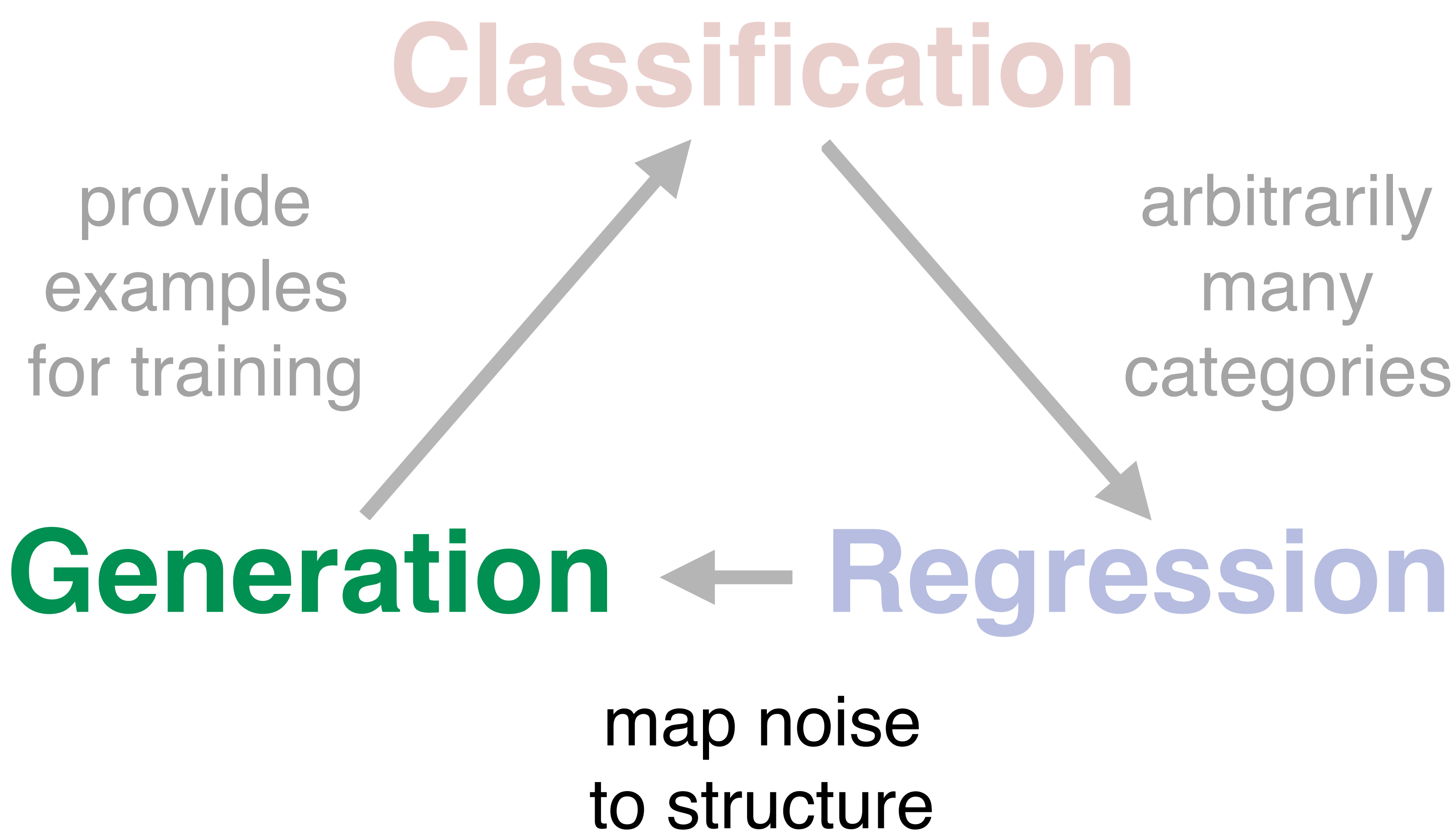


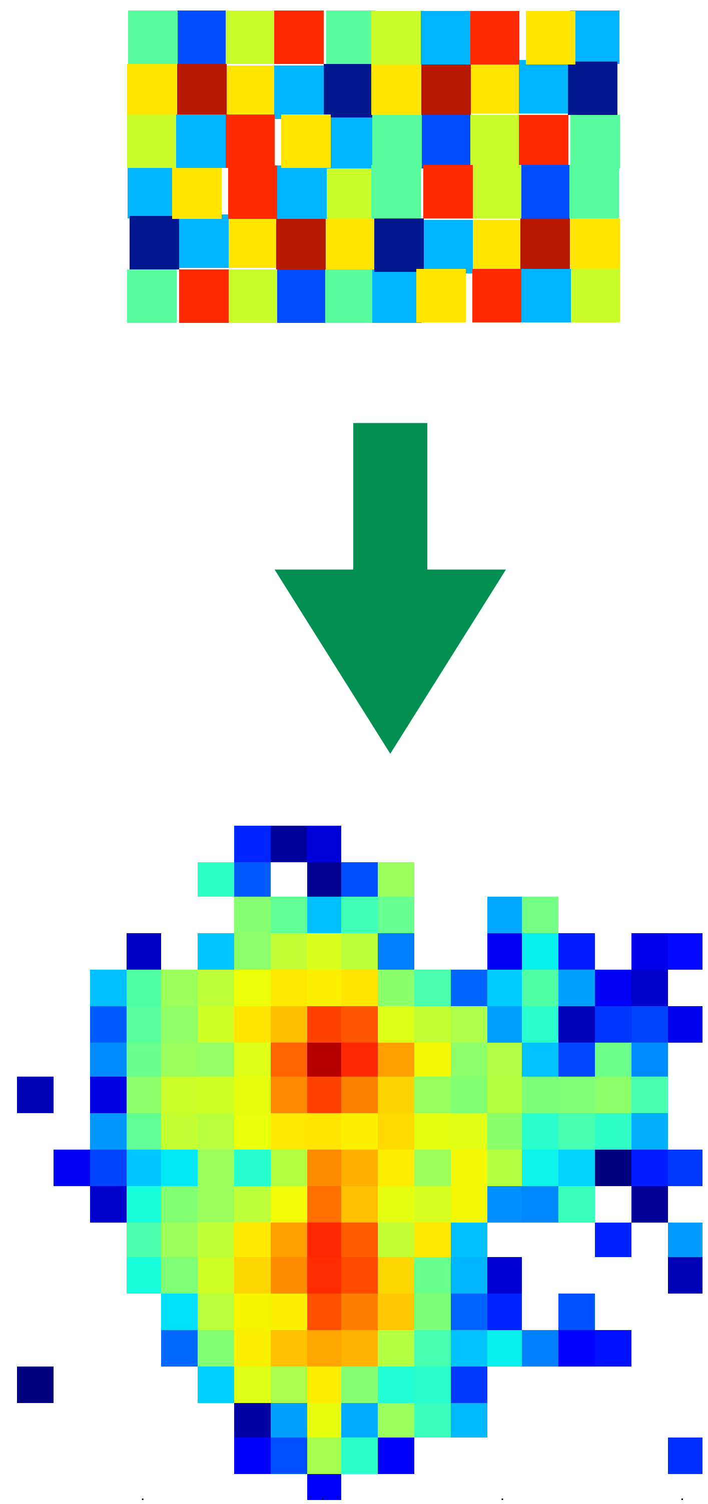
"Pileup Mitigation with Machine Learning"

P. Komiske, E. Metodiev, **BPN**, and M. Schwartz, JHEP 12 (2017) 051

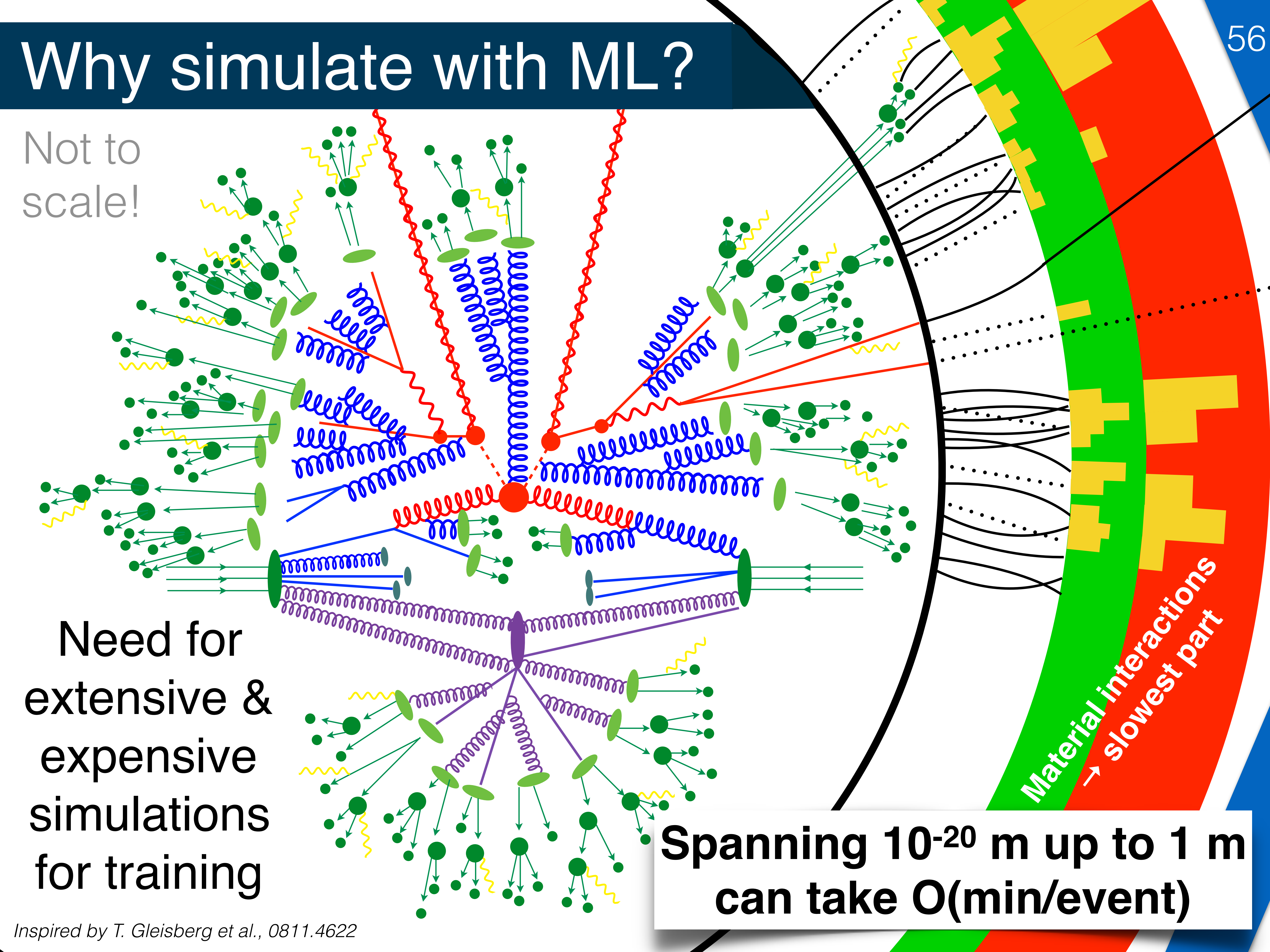
Generative Models

A generator is nothing other than a function that maps noise to structure.





Why simulate with ML?

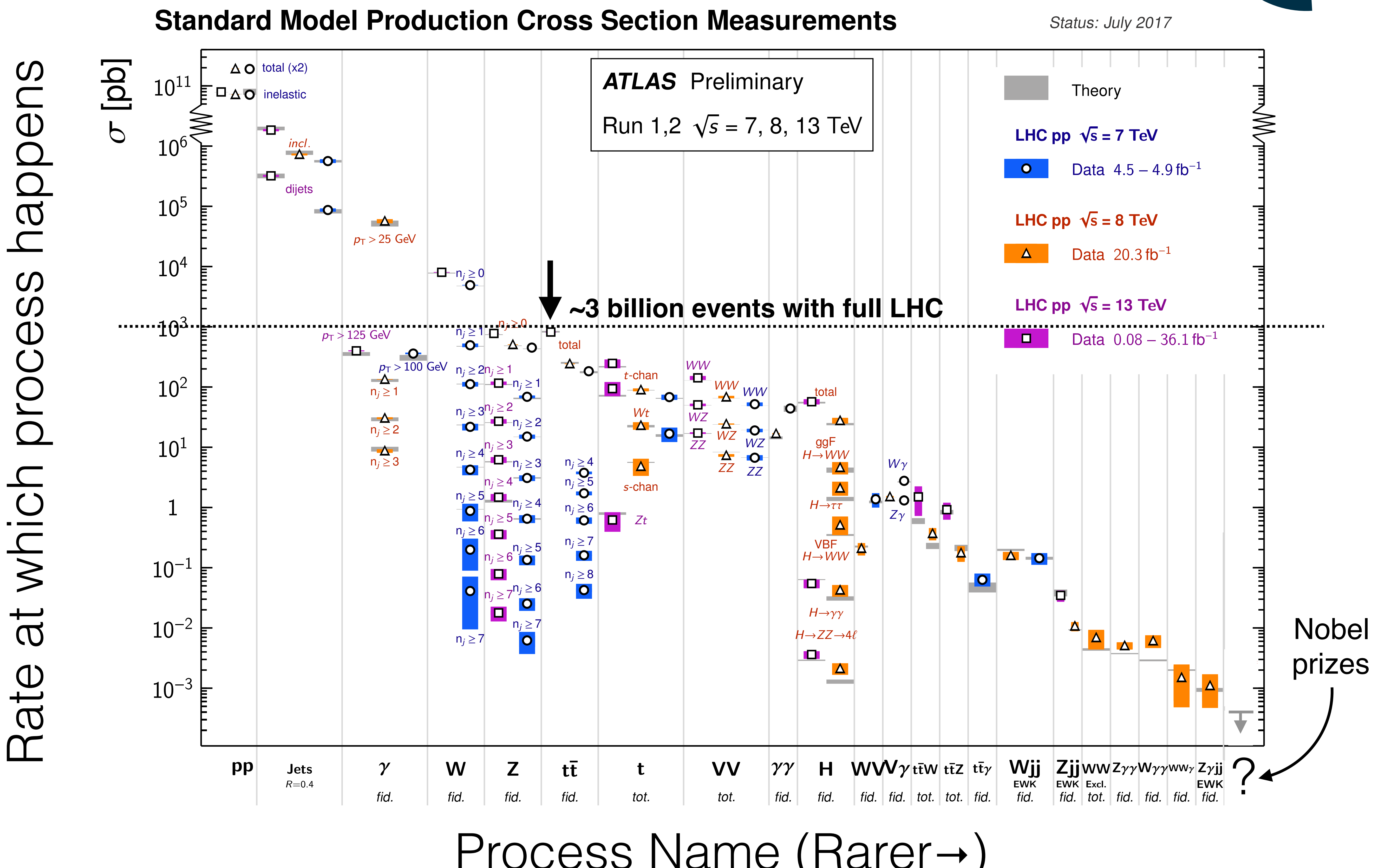


Need for extensive & expensive simulations for training

Spanning 10^{-20} m up to 1 m can take O(min/event)

Inspired by T. Gleisberg et al., 0811.4622

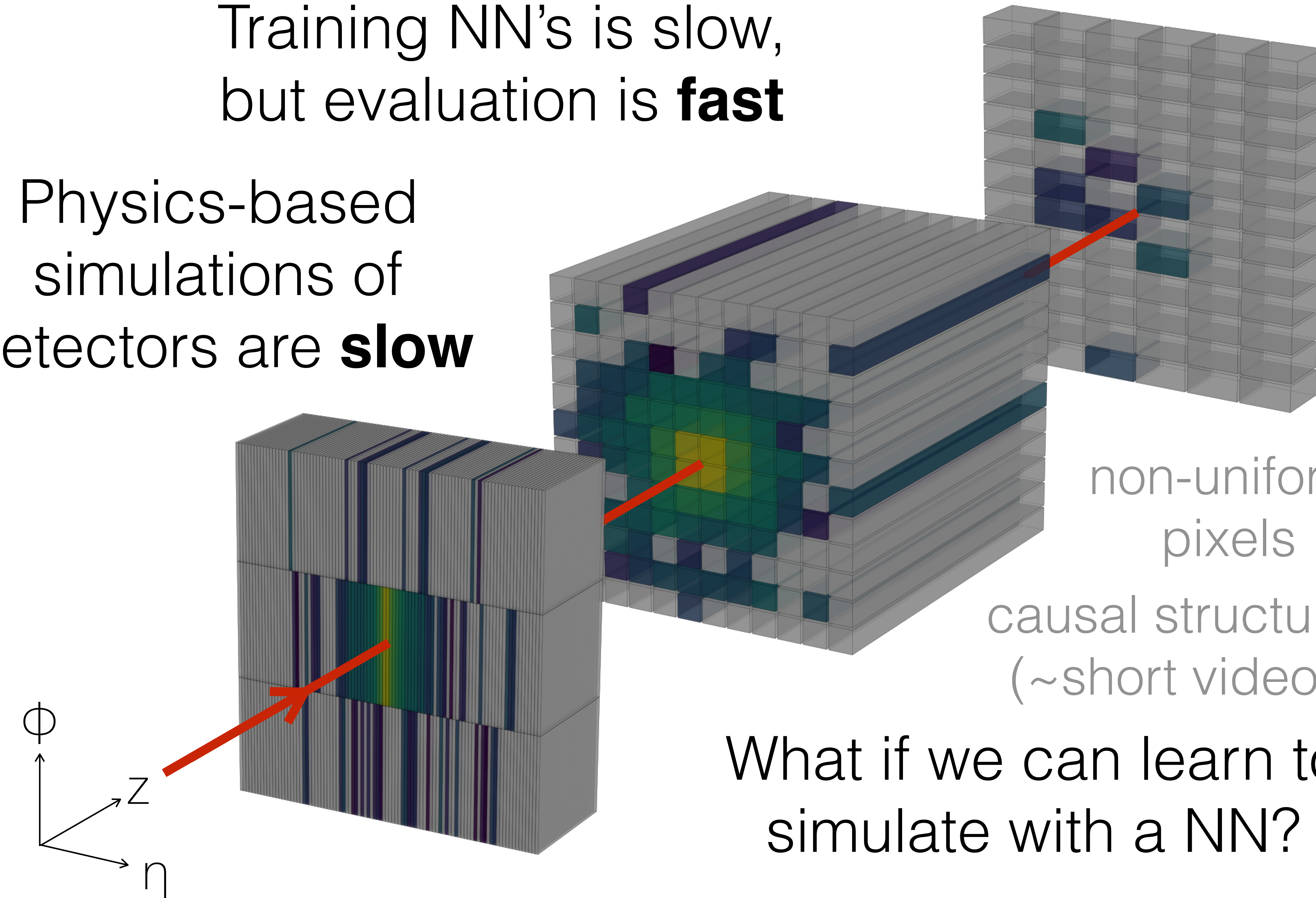
The simulation challenge



DNN's for the simulation challenge

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

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



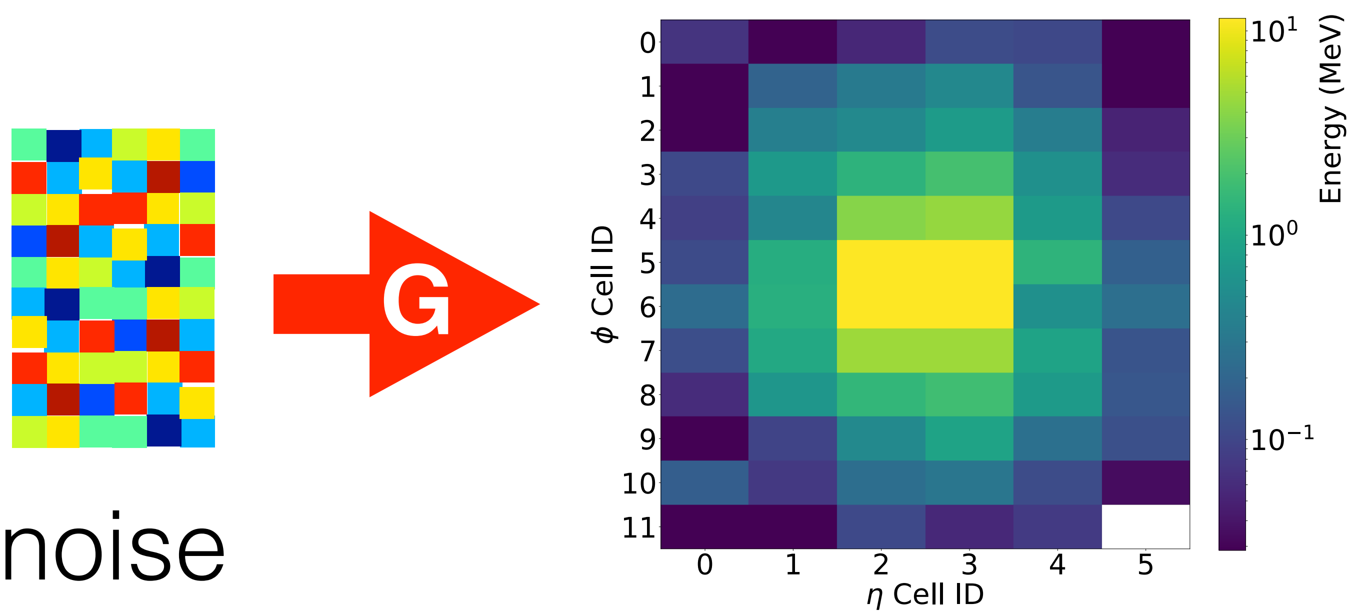
non-uniform pixels

causal structure (~short video)

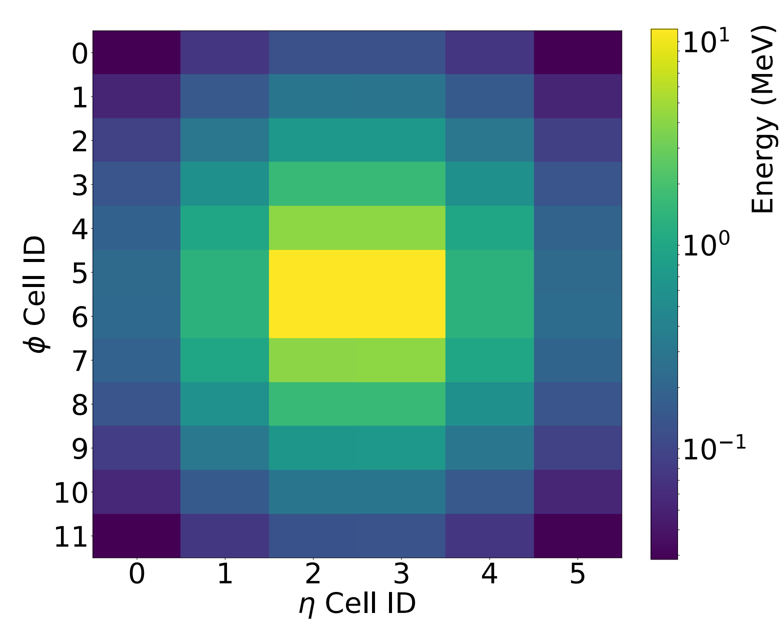
What if we can learn to simulate with a NN?

Generative Adversarial Networks (GAN)

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



{real,fake}



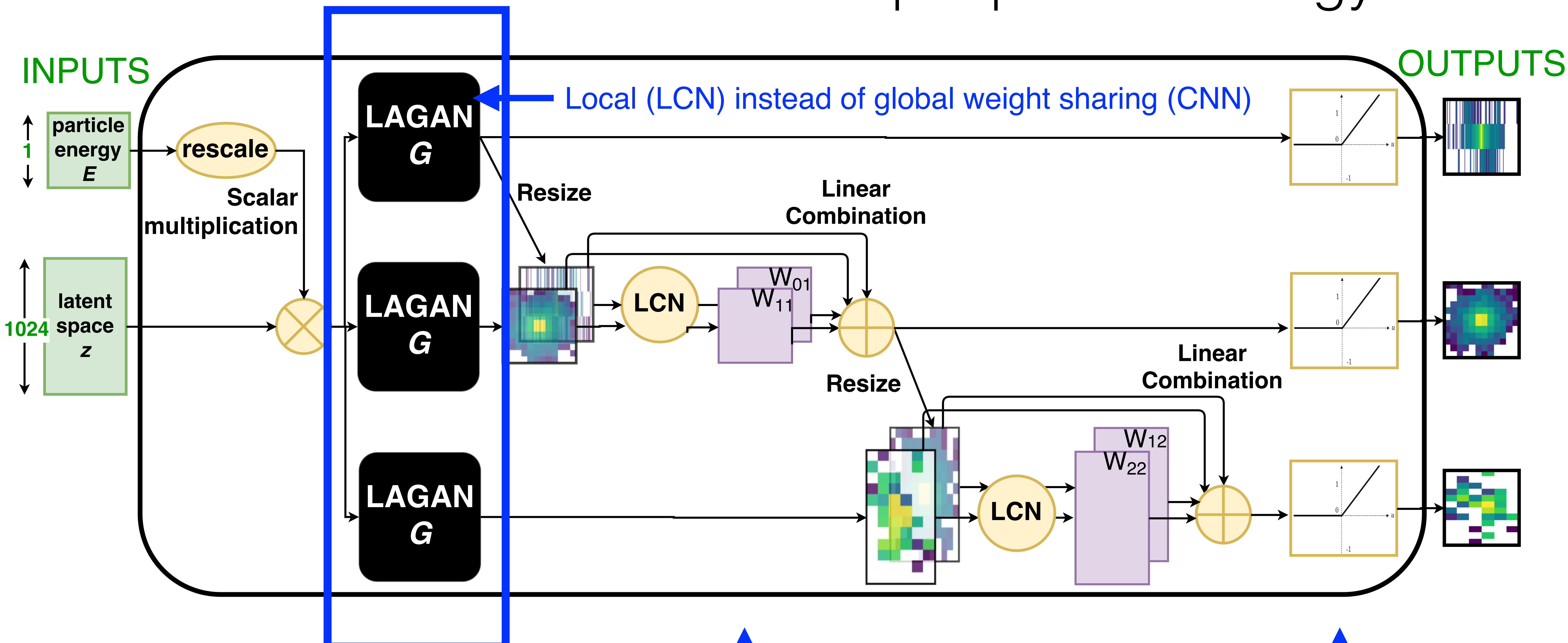
When **D** is maximally confused, **G** will be a good generator

M. Paganini, L. de Oliveira, and **BPN**, PRL 120 (2018) 042003

Accelerating Simulation with GANs

One image per detector layer

One network per particle type; input particle energy



Generator inside Generative Adversarial Network (GAN)

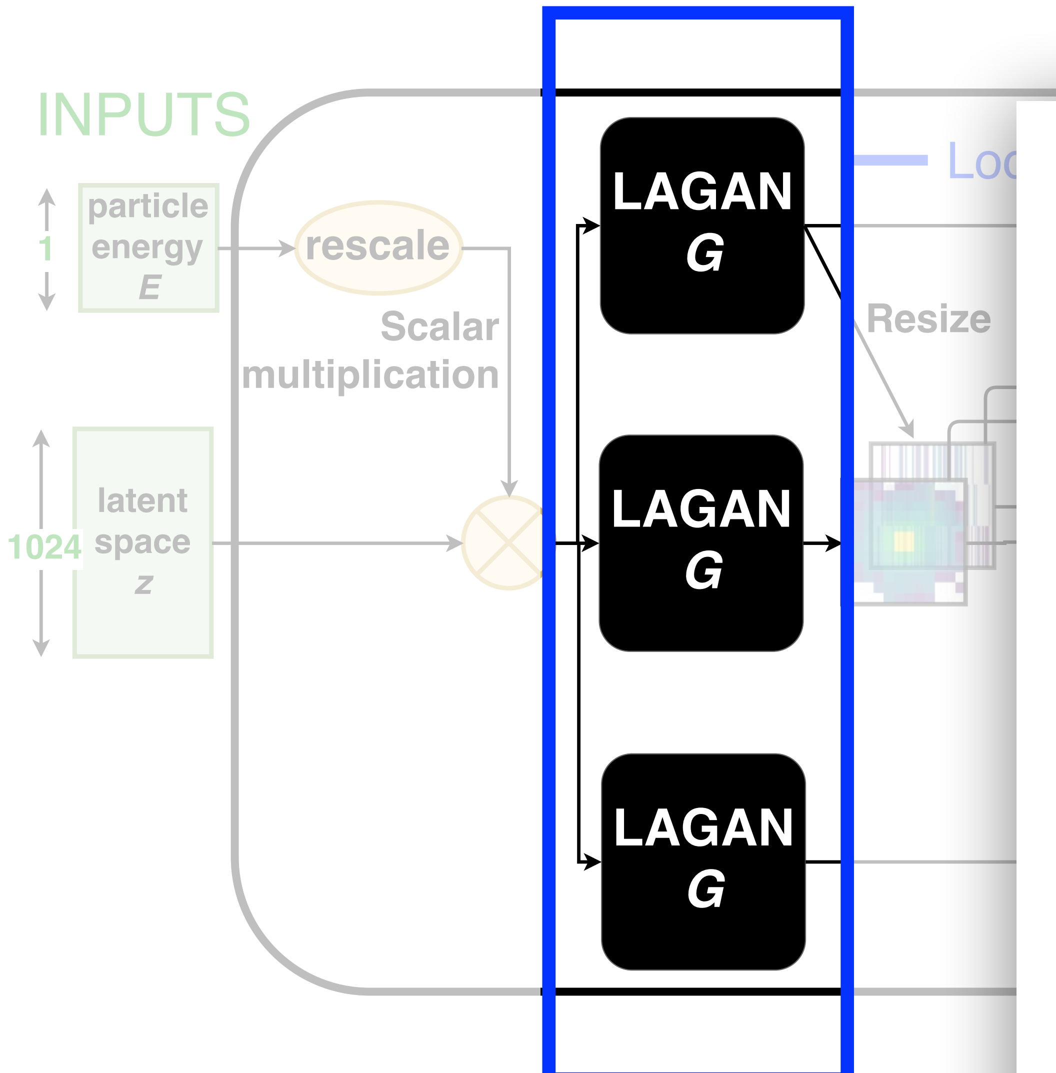
use layer i as input to layer i+1

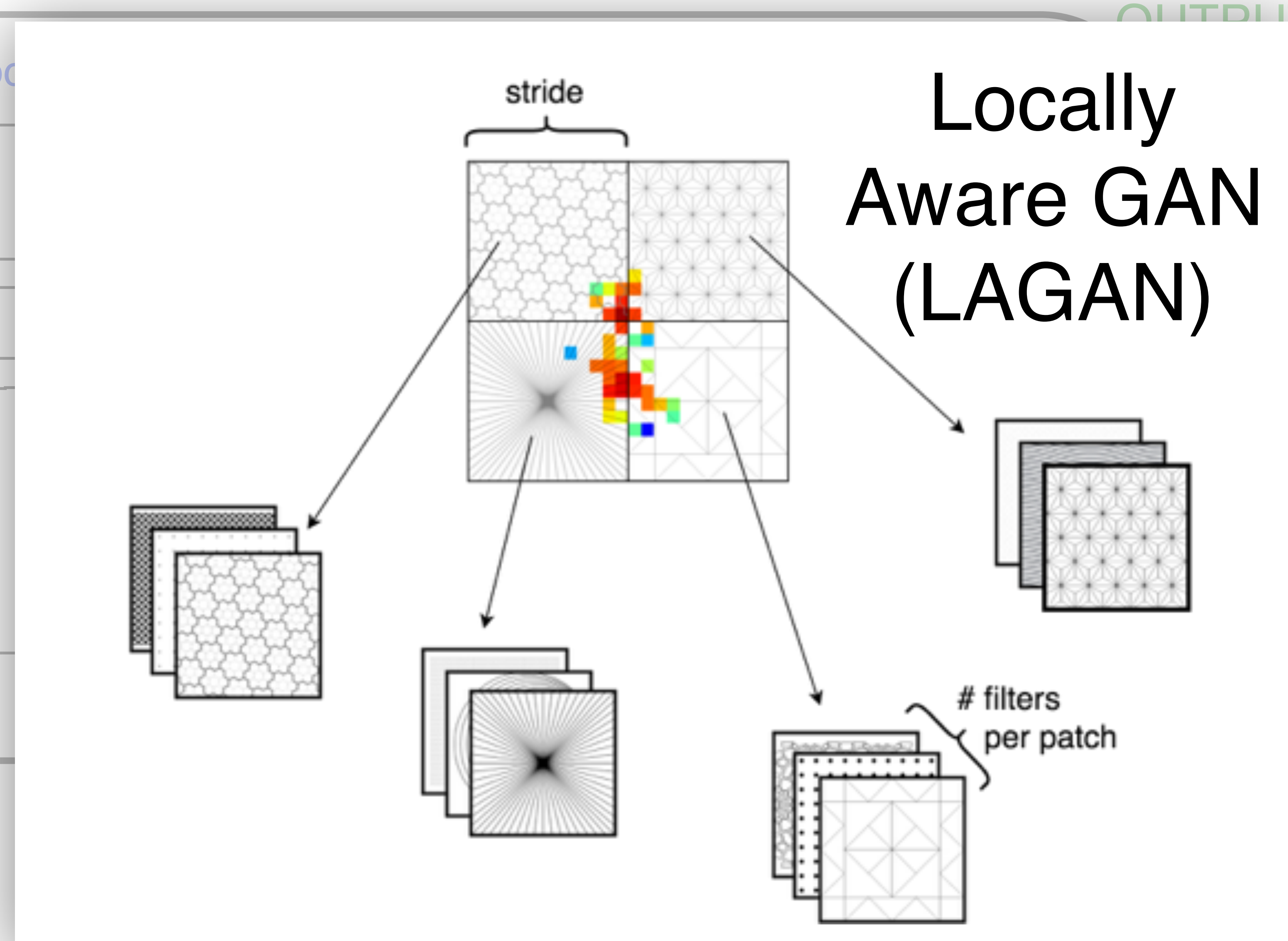
ReLU to encourage sparsity

Accelerating Simulation with GANs

One image per detector layer

One network per particle type; input particle energy



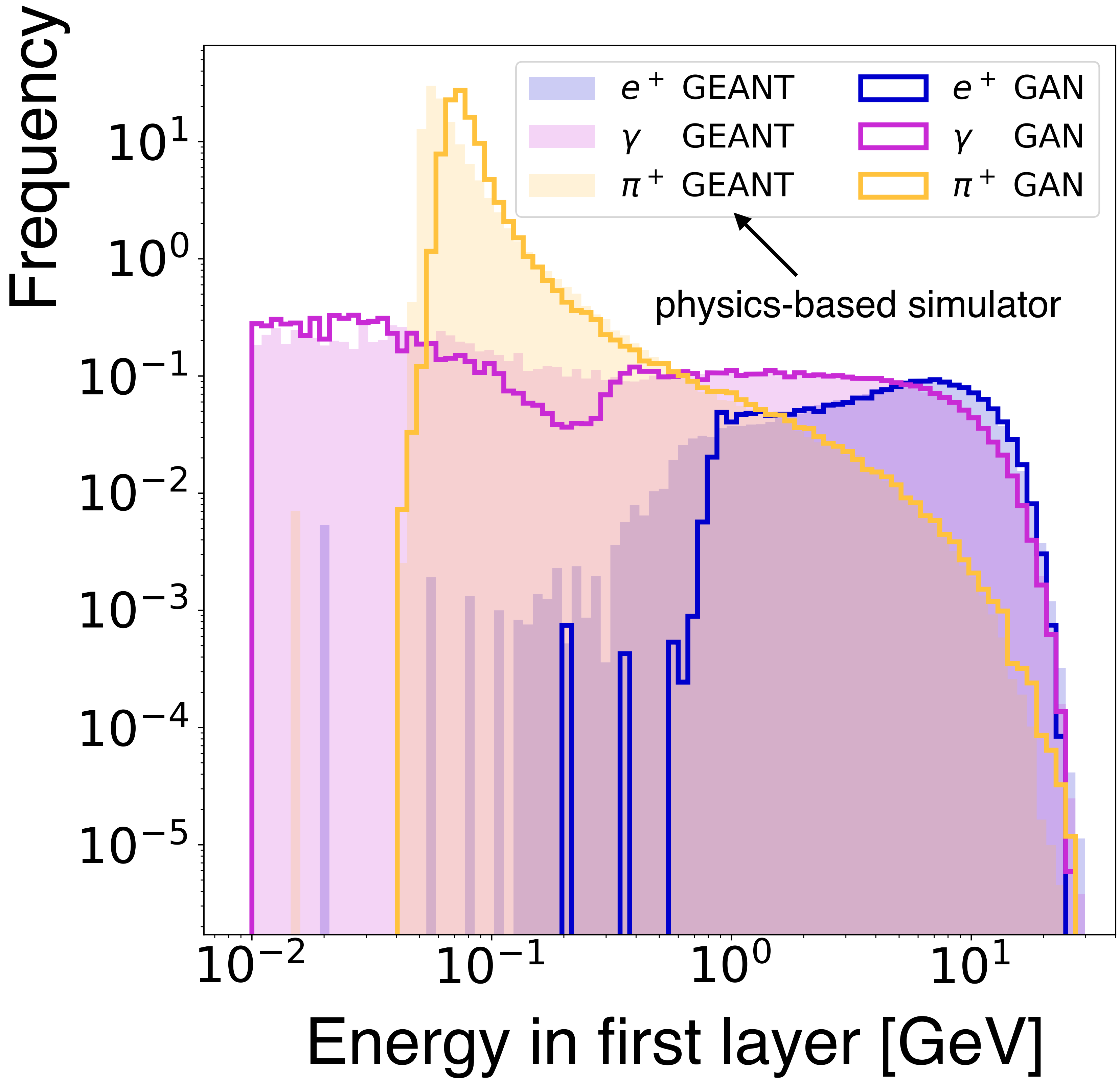


Generator inside Generative Adversarial Network (GAN)

use layer i as input to layer i+1

ReLU to encourage sparsity

CaloGAN Results

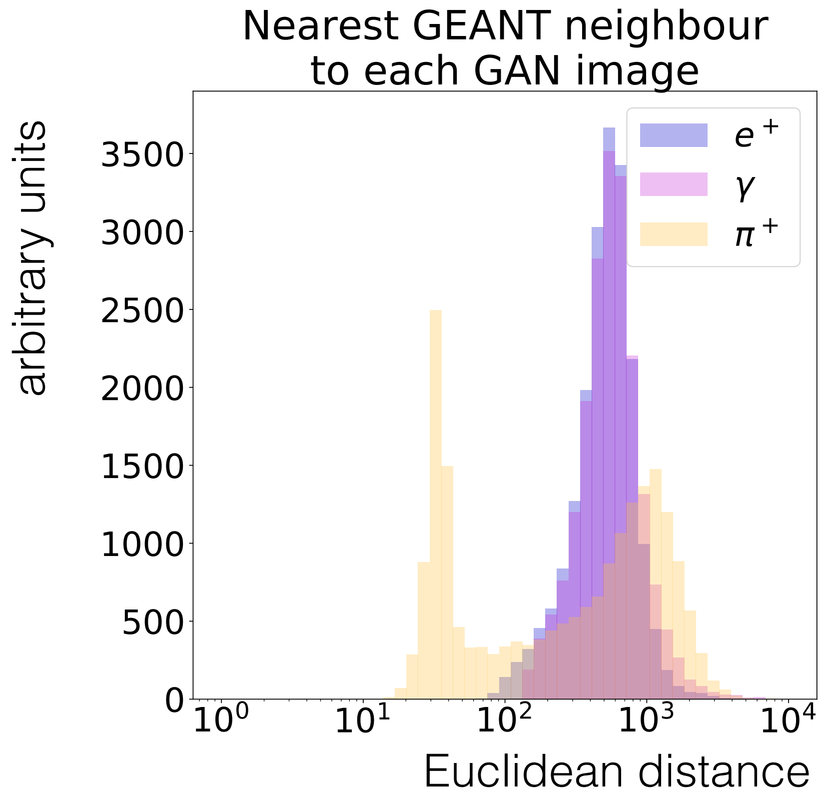


Qualitative agreement; clearly also room for improvement.

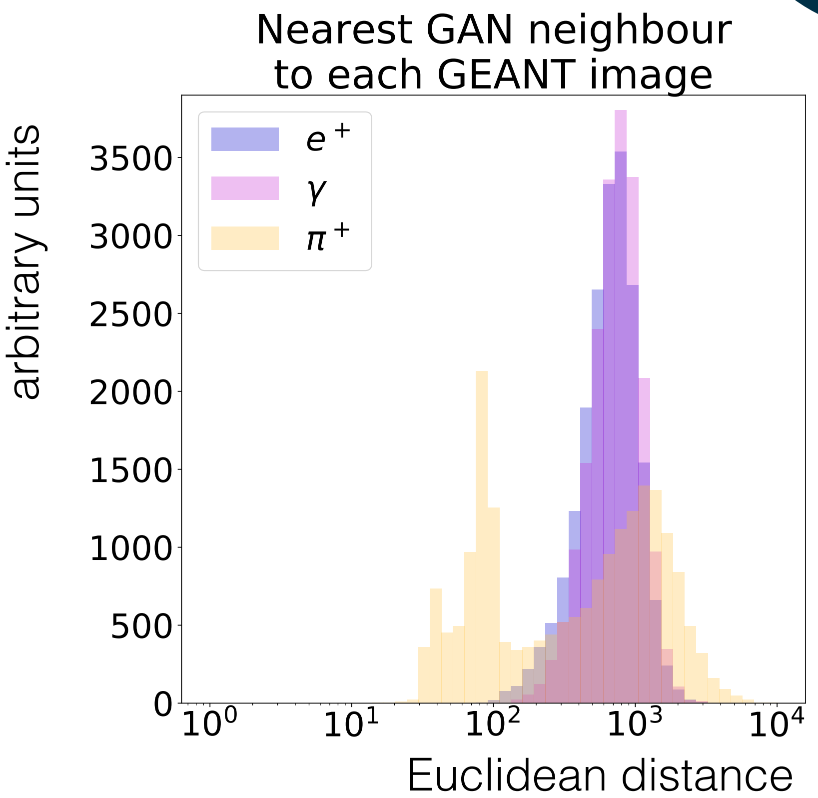
We have examined many 1D projections and most similar quality.

Pions and photons penetrate further than electrons.

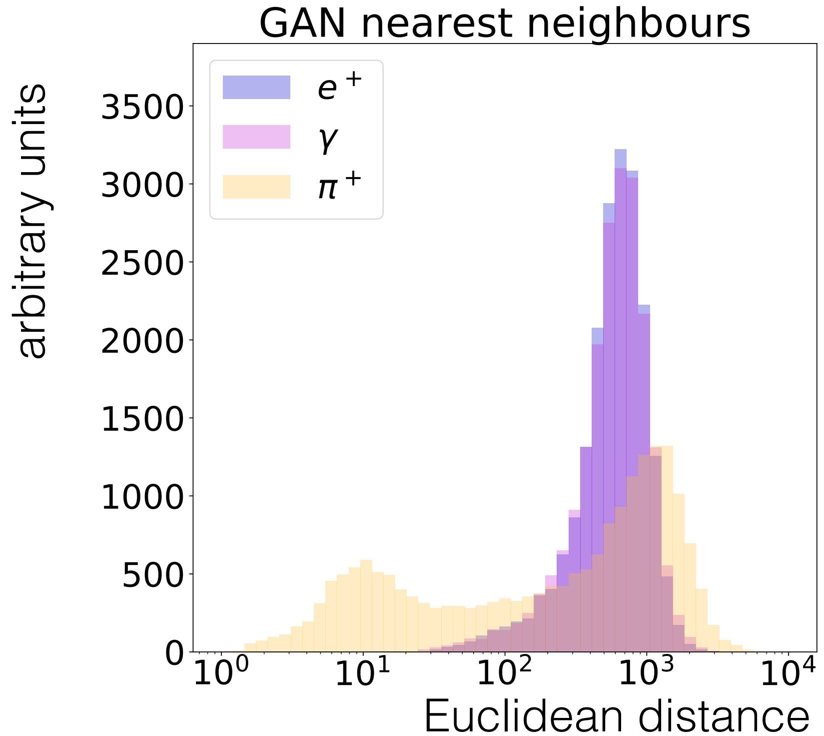
“Overtraining”



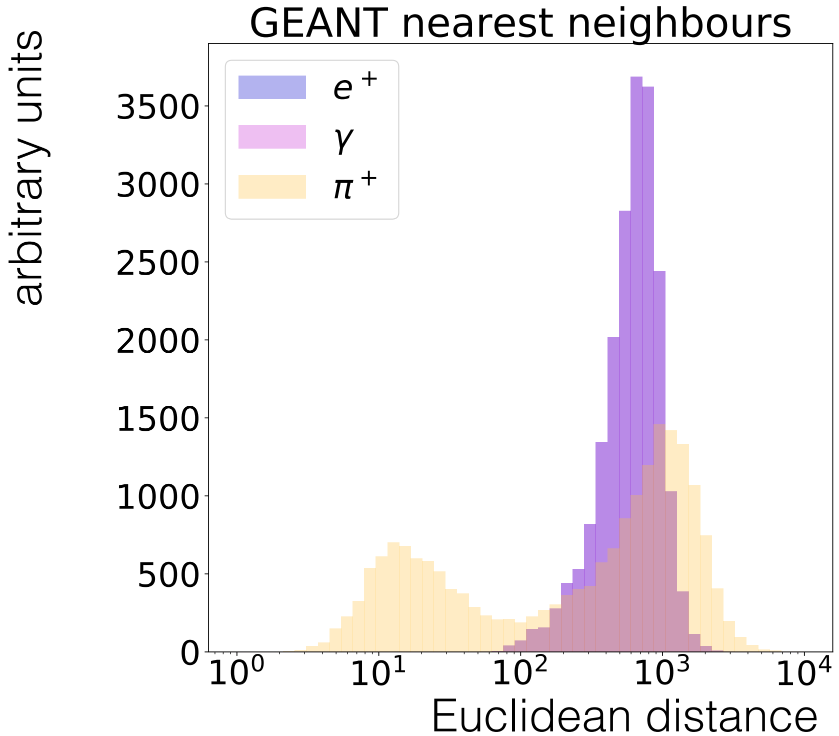
not memorizing



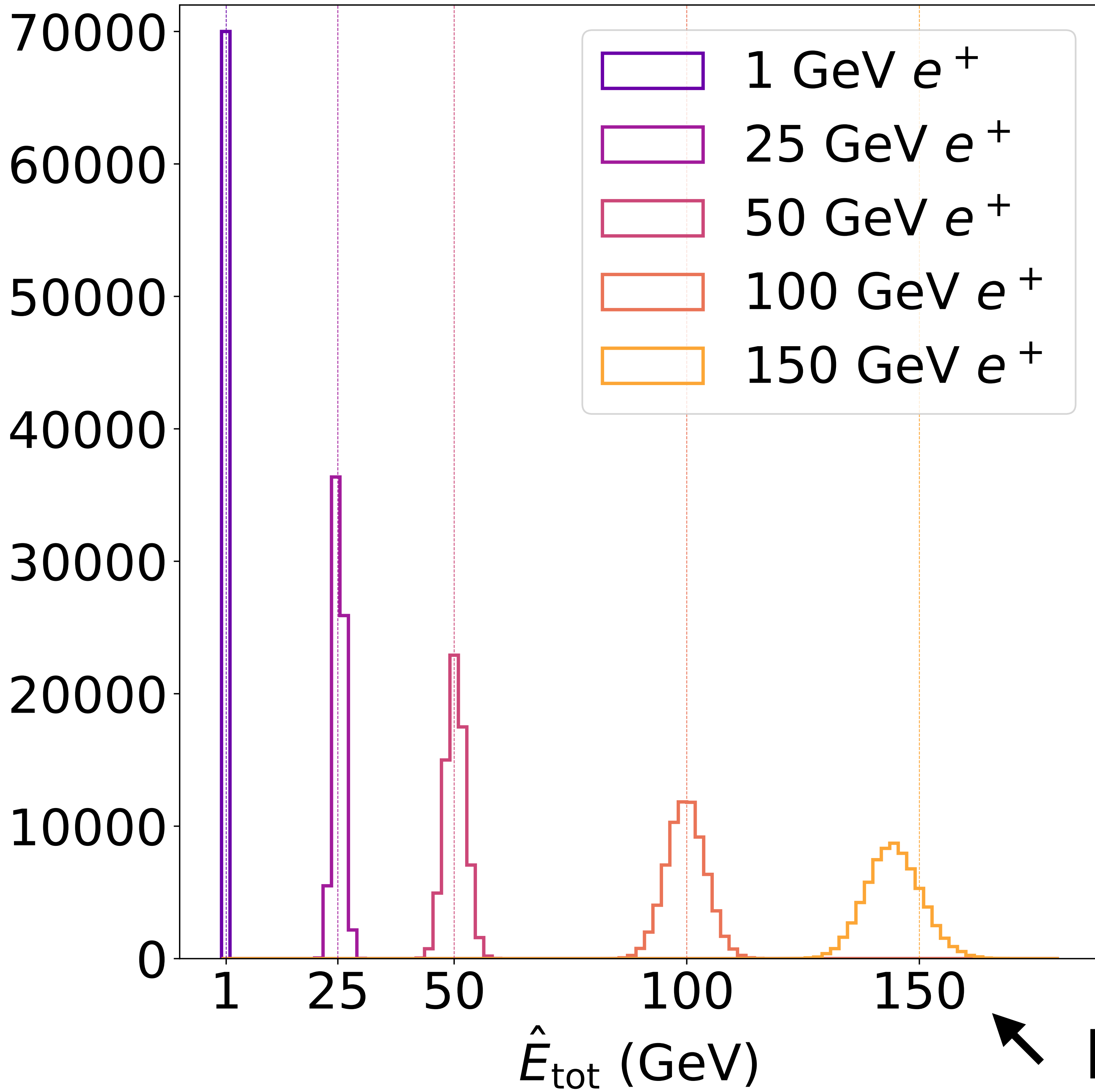
A key challenge in training GANs is the diversity of generated images.



no mode collapse



Extrapolation



Beyond our training sample!

Timing

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 ←

Active work in ATLAS and first public results for ICHEP (ATL-SOFT-PUB-2018-001)

Can you imagine a future where you can run full detector simulations on your laptop (in real time)?

Review

We have been machine learning for a long time, but deep learning allows us to use all of the available information.

Non-trivial interplay between physics & representations / preprocessing.

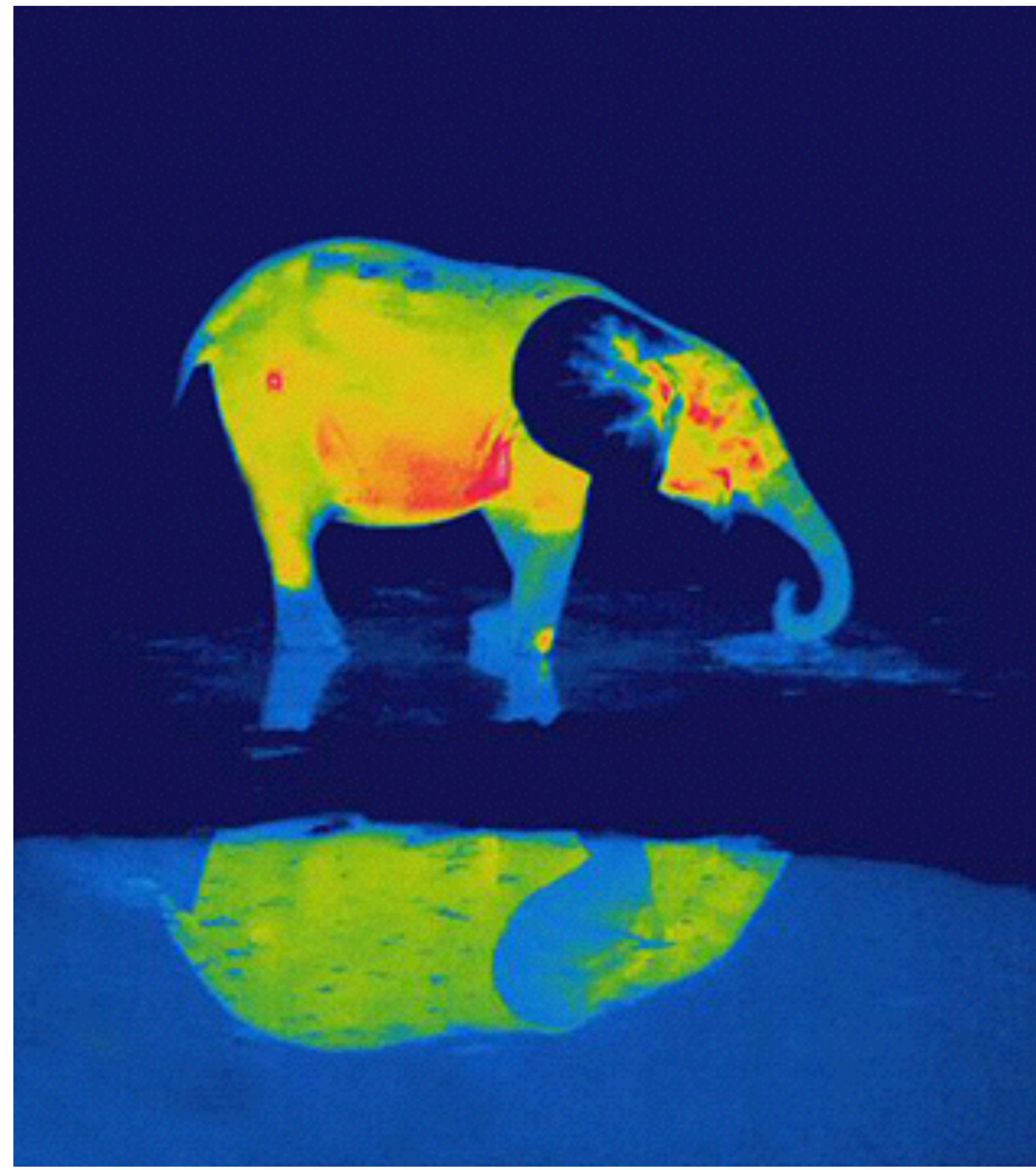
Many ways to use modern ML:

Classification

Regression

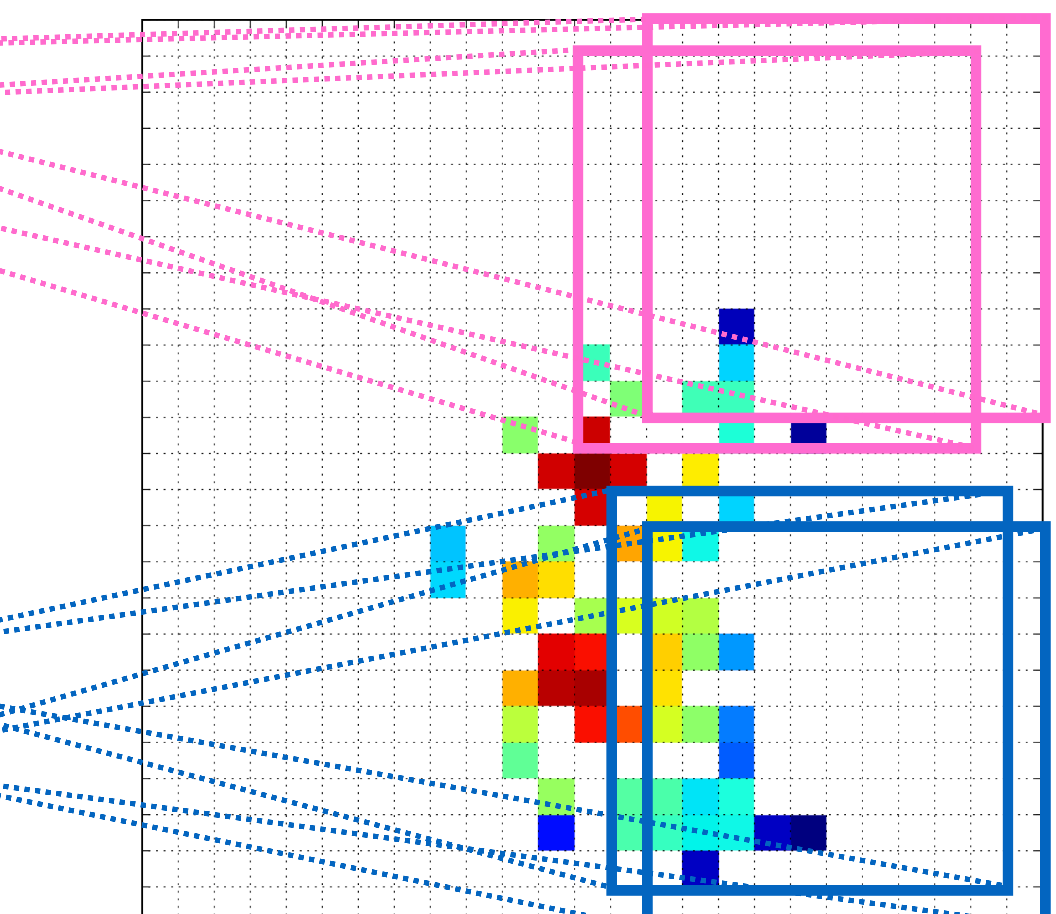
Generation

I've only picked a few examples using images - many exciting developments!

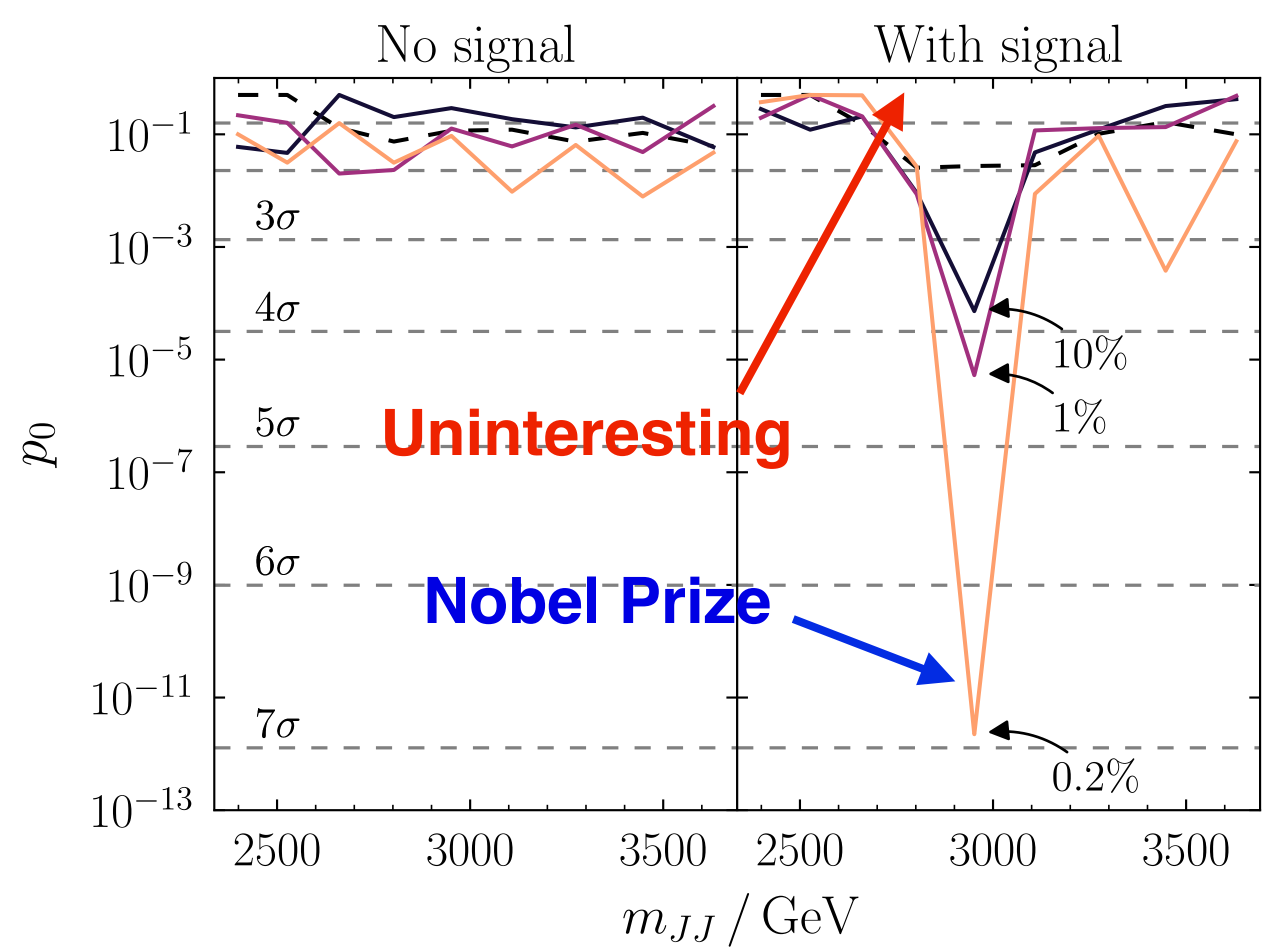


Outlook

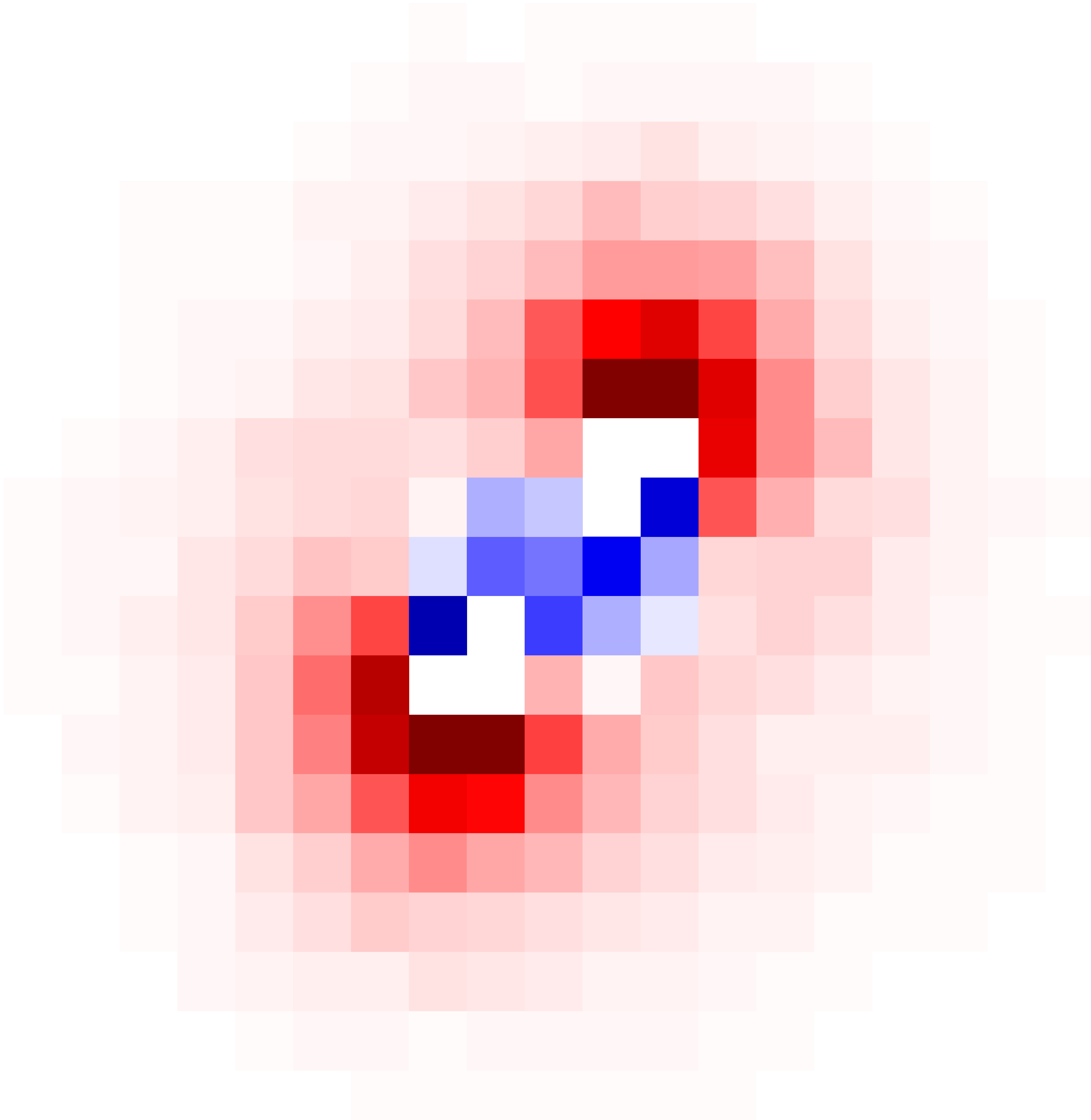
Deep learning will replace/ augment standard analysis pipelines by giving us hypervariat vision and new ways to represen & learn (directly) from data!



J. Collins, K. Howe, **BPN**, PRL 121, 241803 (2018)



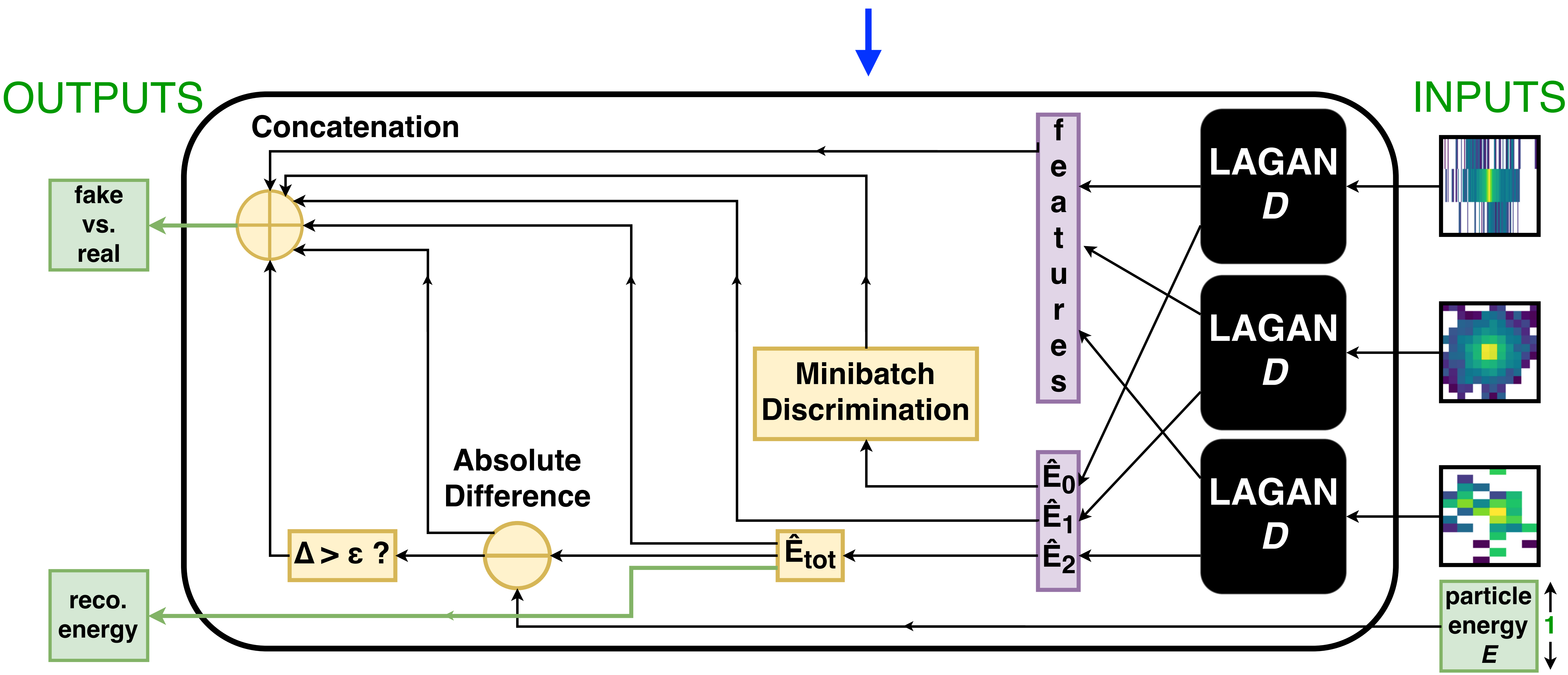
If you want to hear more about how we can use **weak supervision** for HEP machine learning, check out the seminar on Friday!



Fin.

Accelerating Simulation with GANs

help avoid
'mode collapse'



Generator network "backwards"