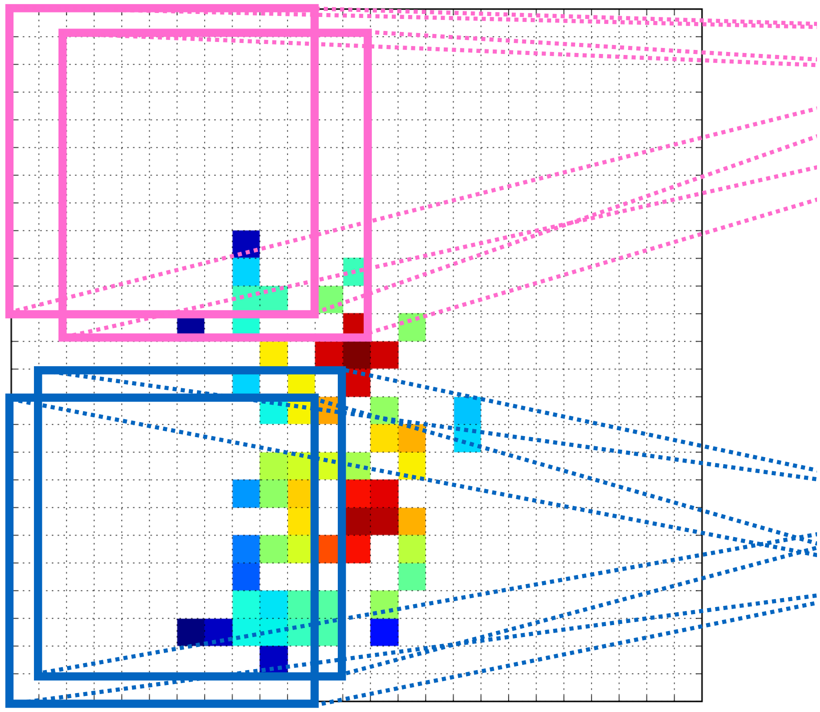
Top tagging and machine learning
Overview and future prospects

Benjamin Nachman

Lawrence Berkeley National Laboratory

bpnachman.com
bpnachman@lbl.gov
@bpnachman
bnachman

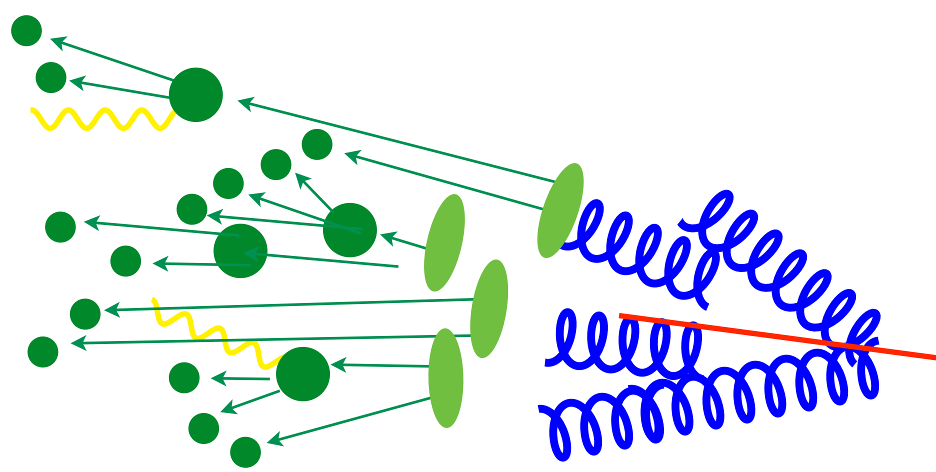


BERKELEY LAB

Top2020
September 17, 2020

Outline

- **Boosted Top Jet Classification**
  - Jet Representations

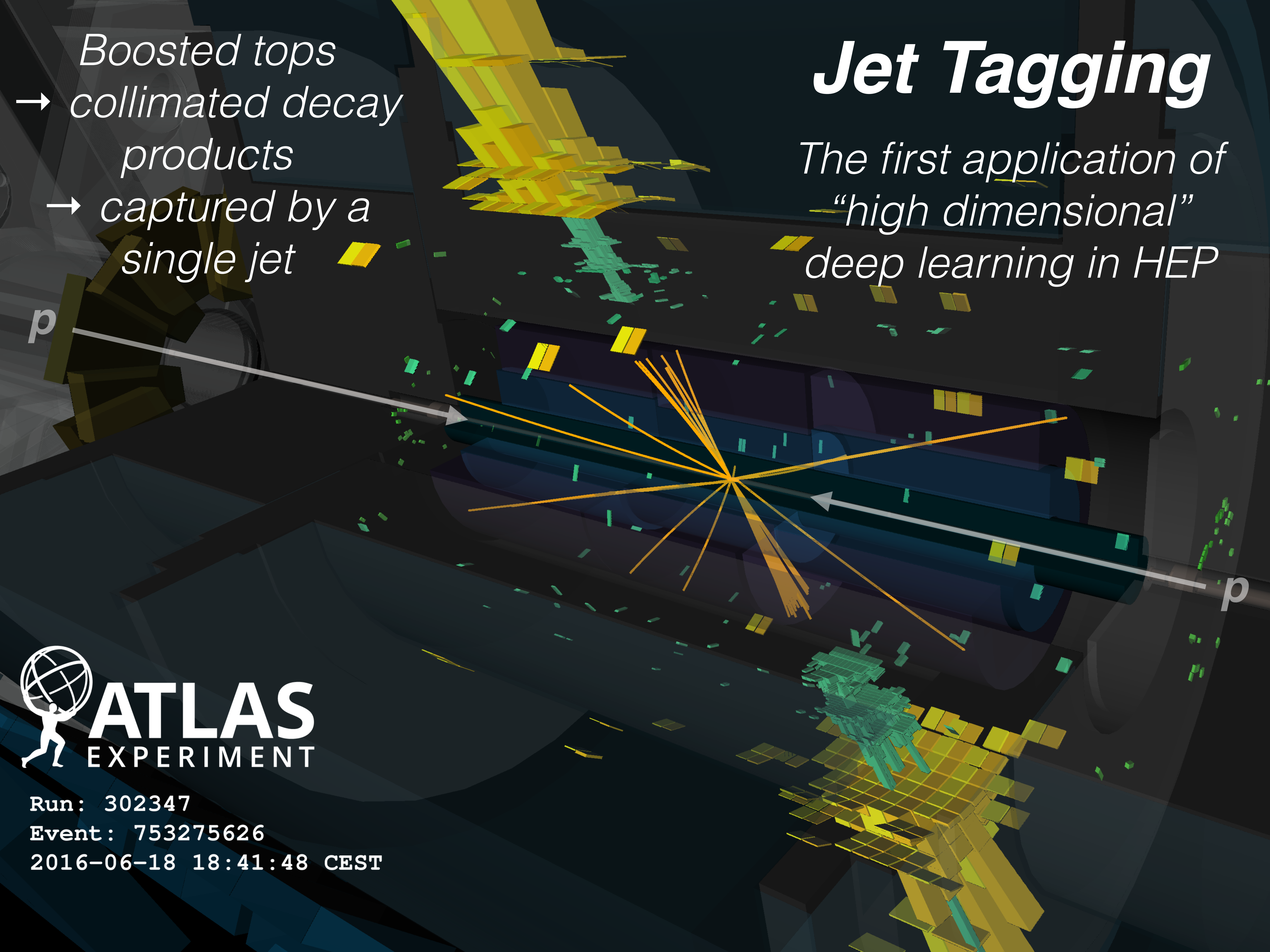


- **Beyond Standard Classification**
  - Less-than-supervised
    - Generation (more in Anja's talk)
    - Weak supervision
    - Anomaly detection (more in Jernej and Bryan's talks)
  - Parameter estimation / unfolding
  - Domain adaptation

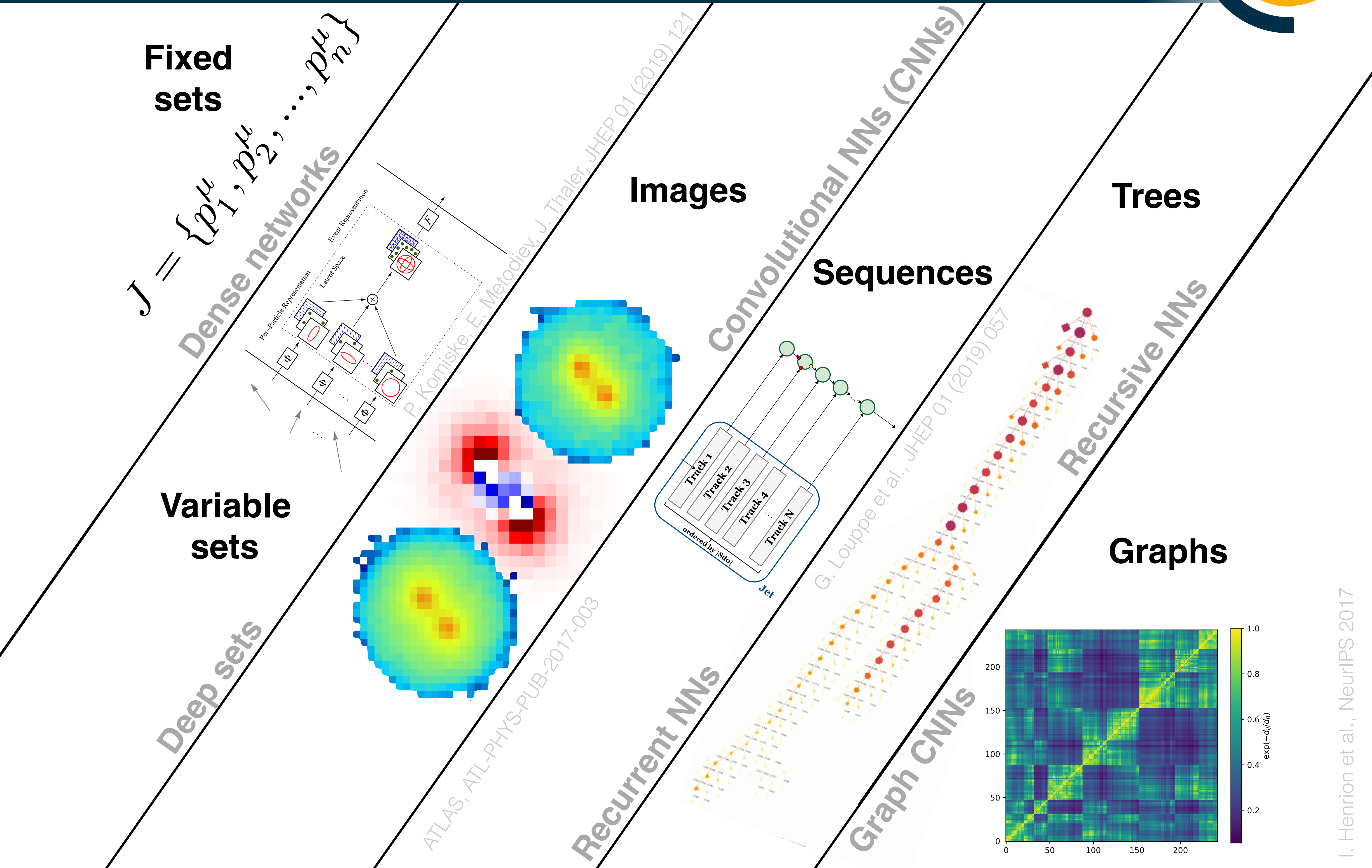
Jet Tagging

The first application of "high dimensional" deep learning in HEP

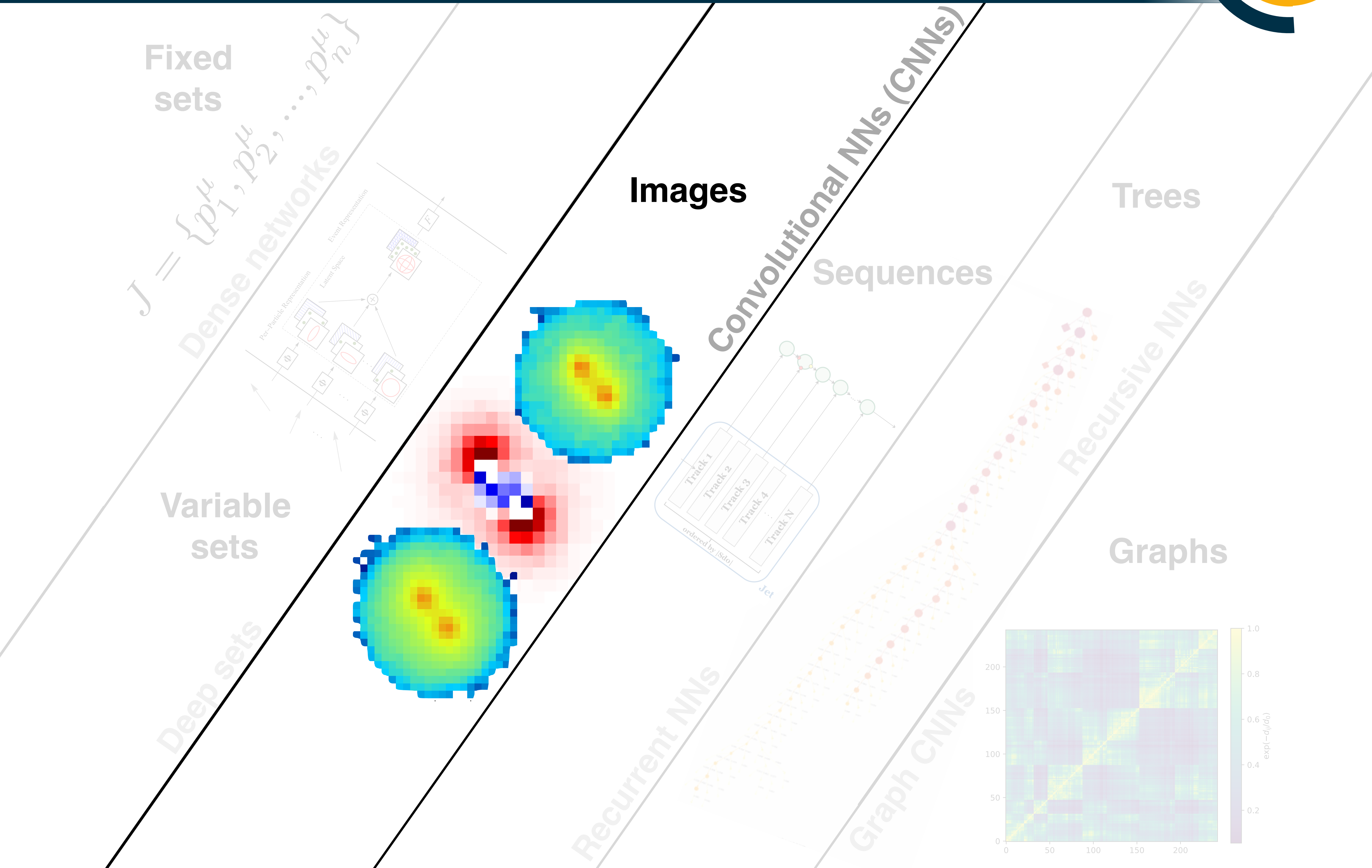
Boosted tops
→ collimated decay products
→ captured by a single jet



Question 1: how to represent our data?



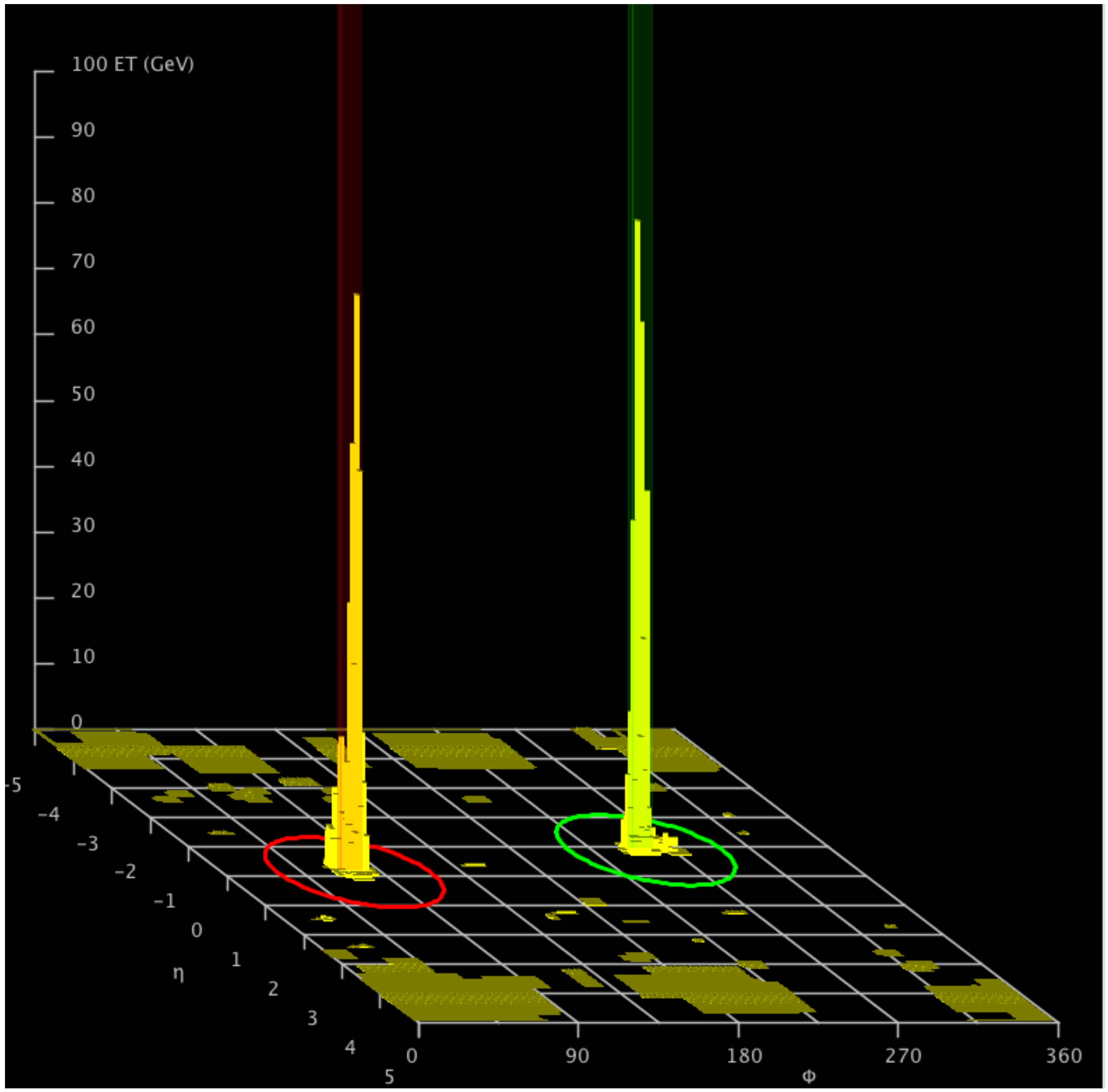
One answer: as an image!

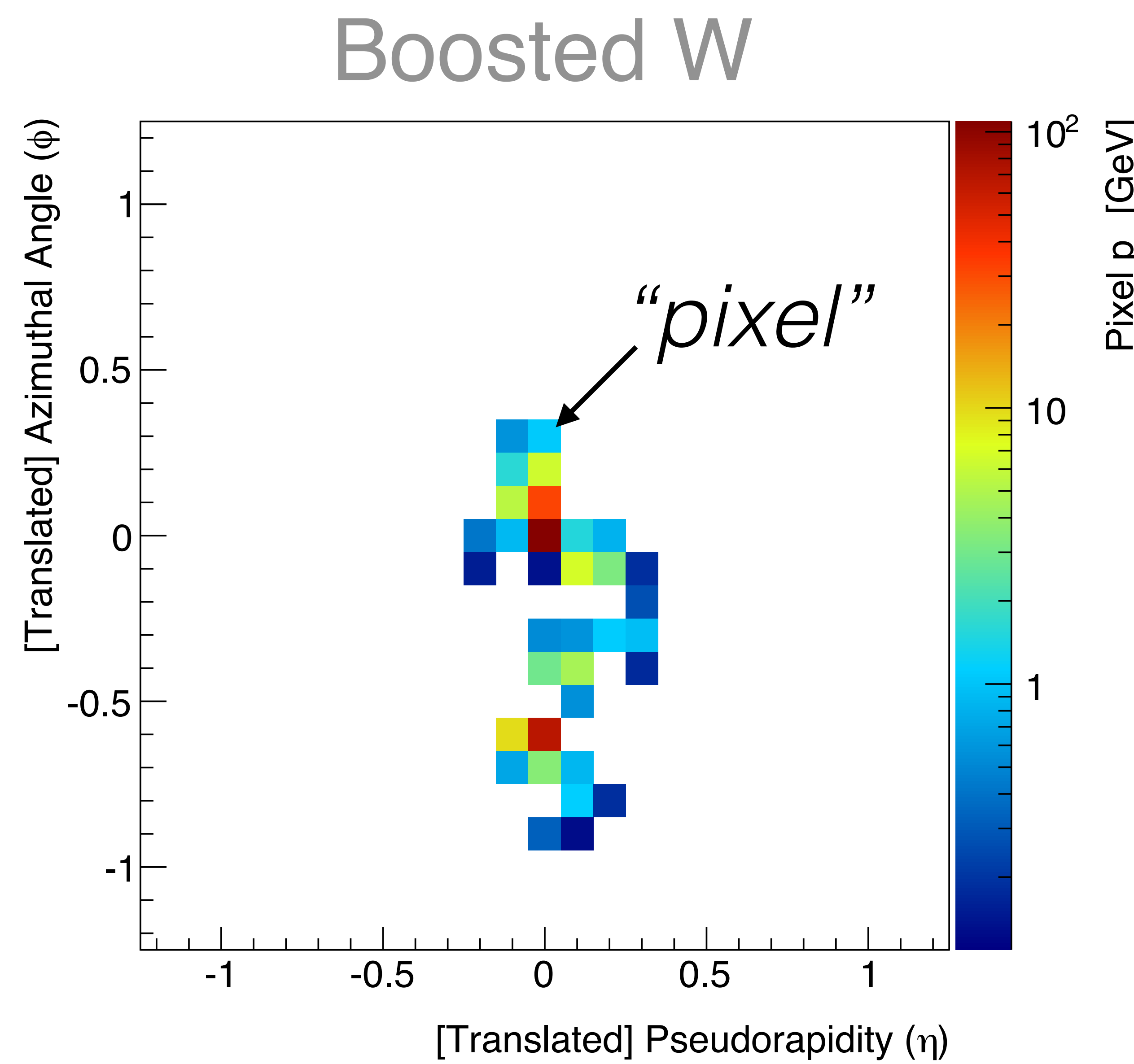


the Jet Image

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

Looks like a digital image!





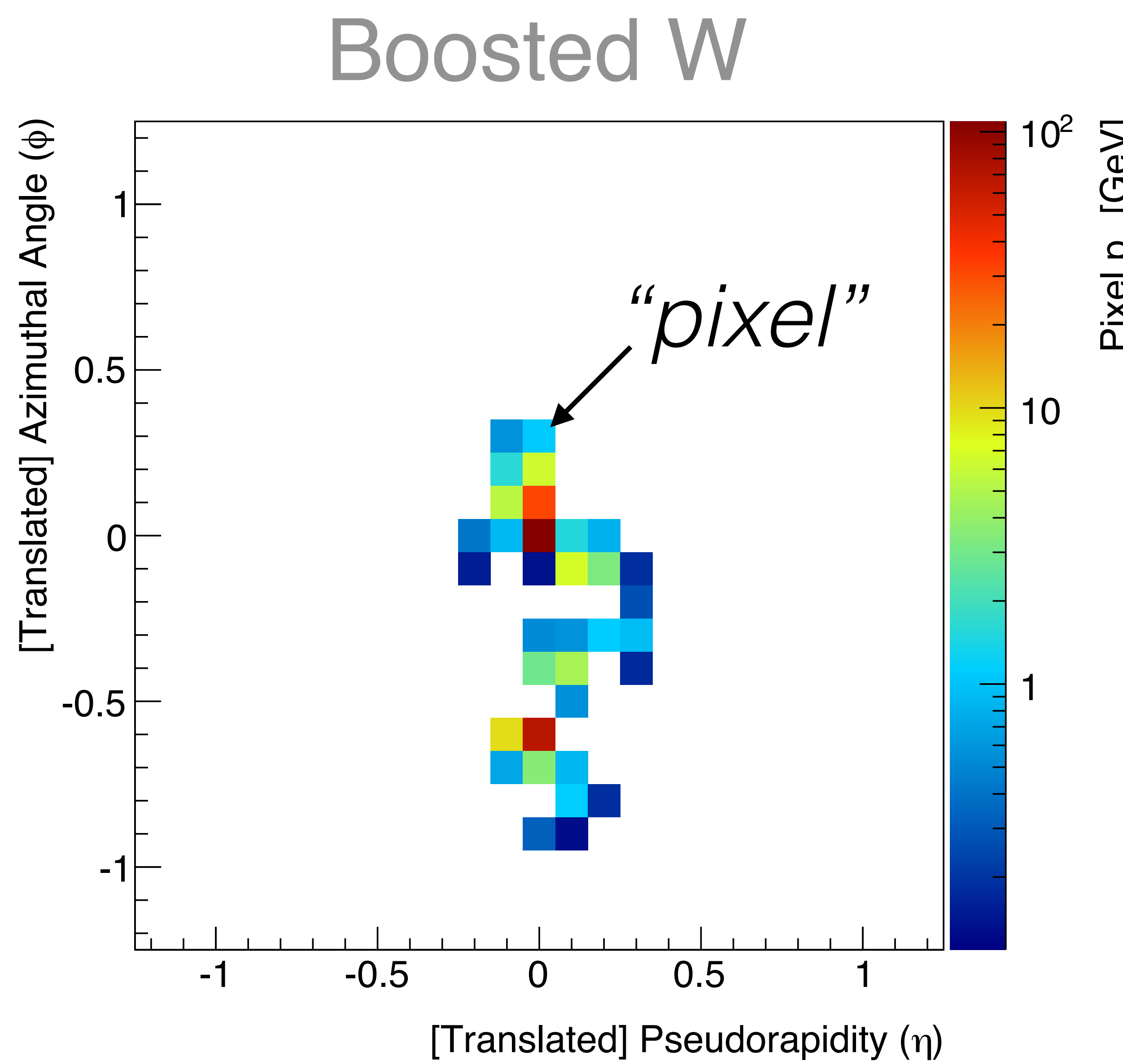
L. de Oliveira, M. Paganini, BPN, Comp. and Software for Big Science (2017) 1

nothing like a 'natural' image!

the Jet Image

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

Credit: Peter G Trimming (Wikipedia)



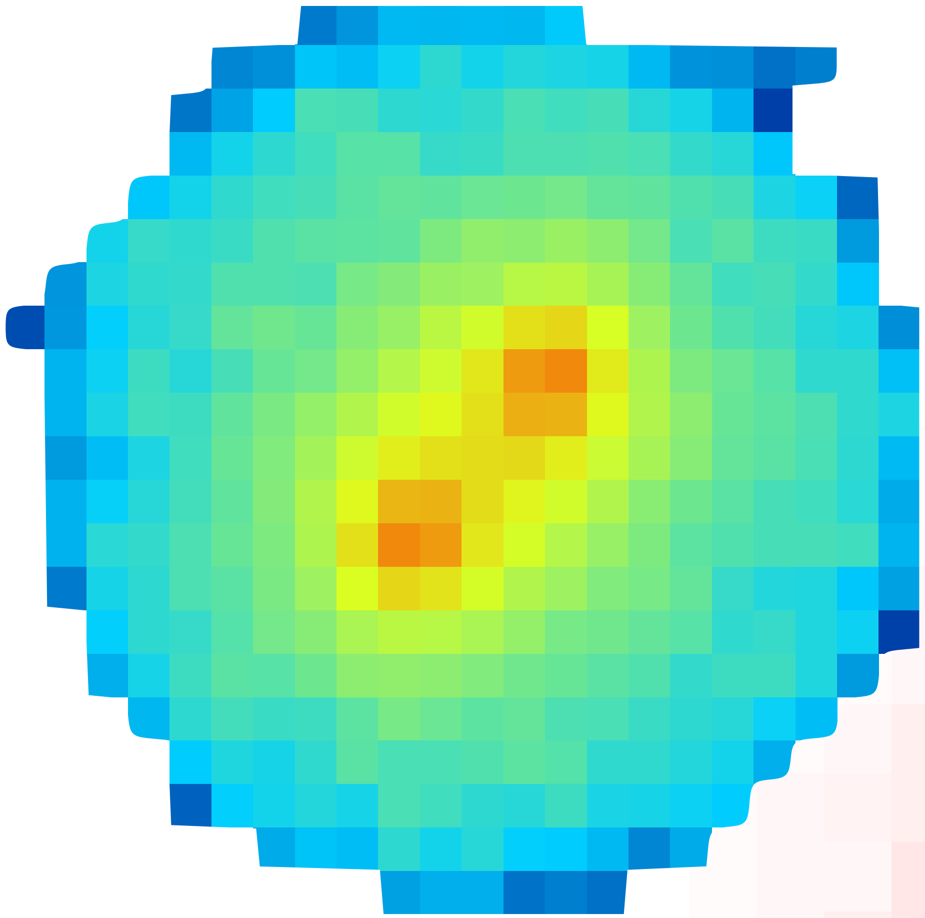
no smooth edges, clear features, low occupancy (number of hit pixels)

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

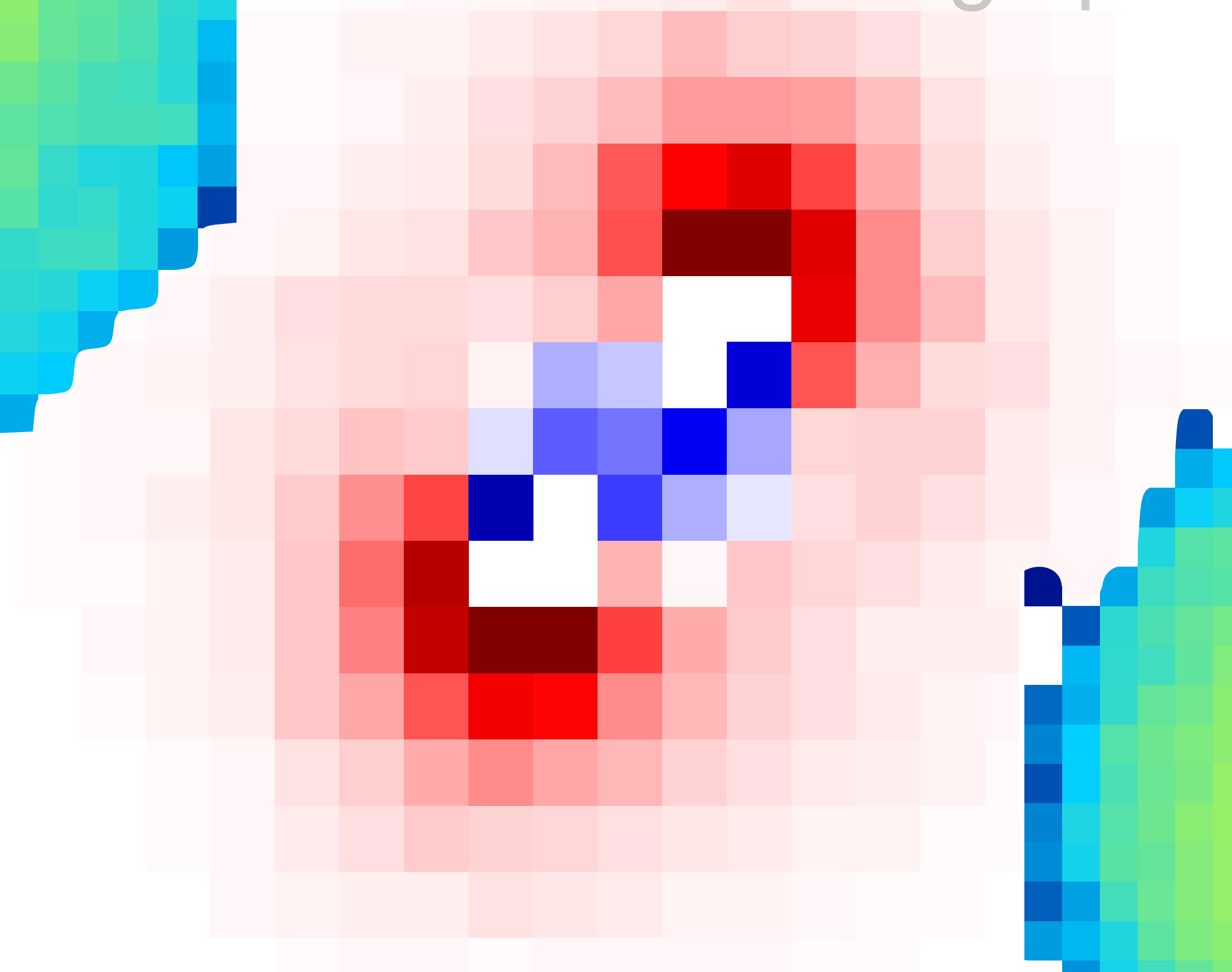
Why Images?

Can directly visualize physics

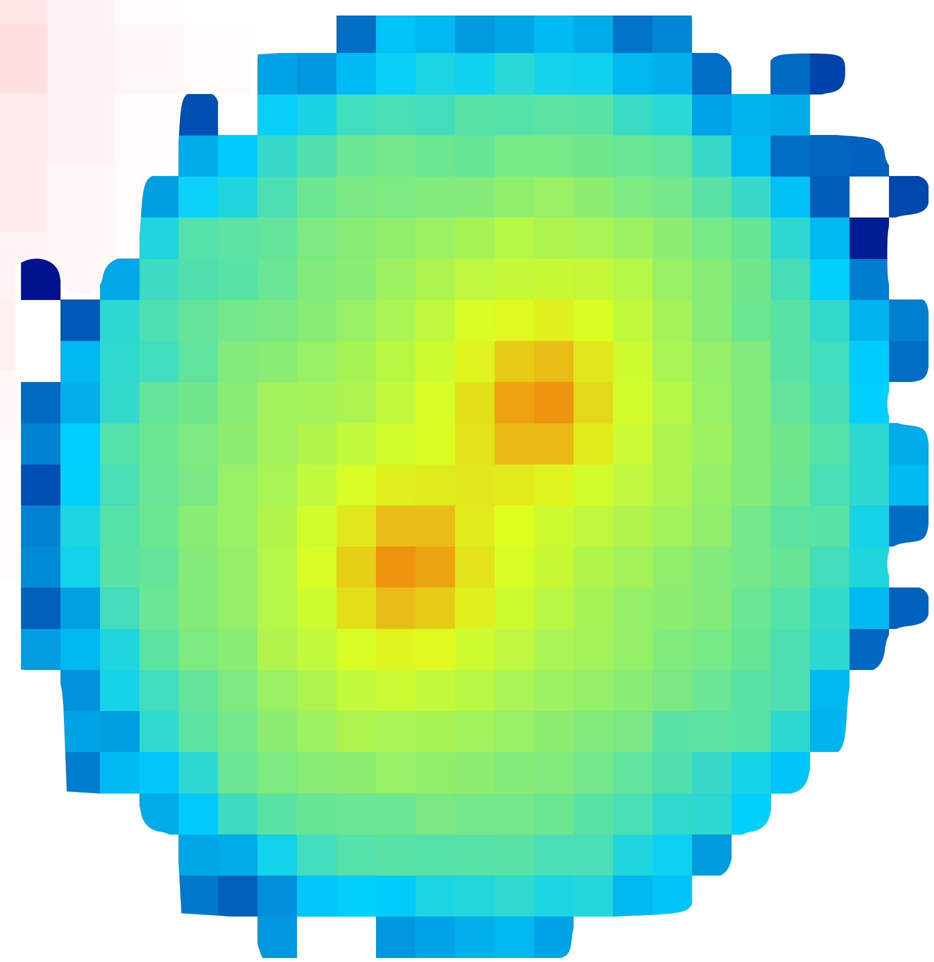
and we can benefit from the extensive image processing literature



$W \dashrightarrow q\bar{q}$

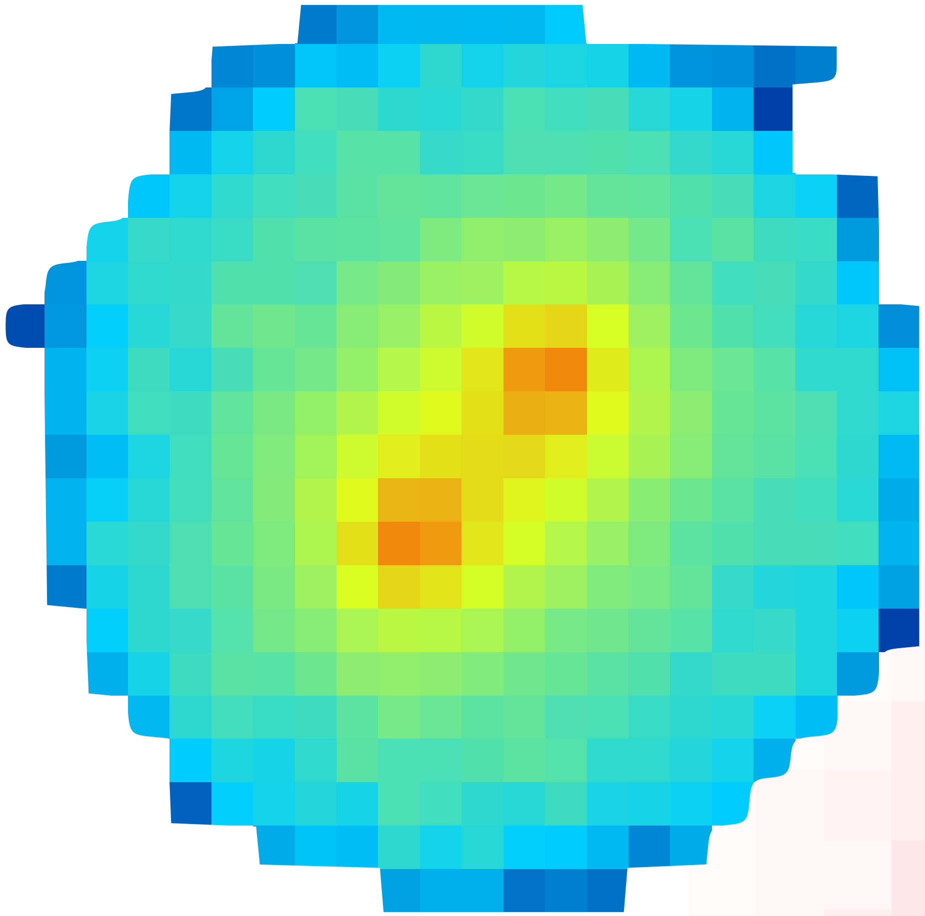


$g \dashrightarrow q\bar{q}$



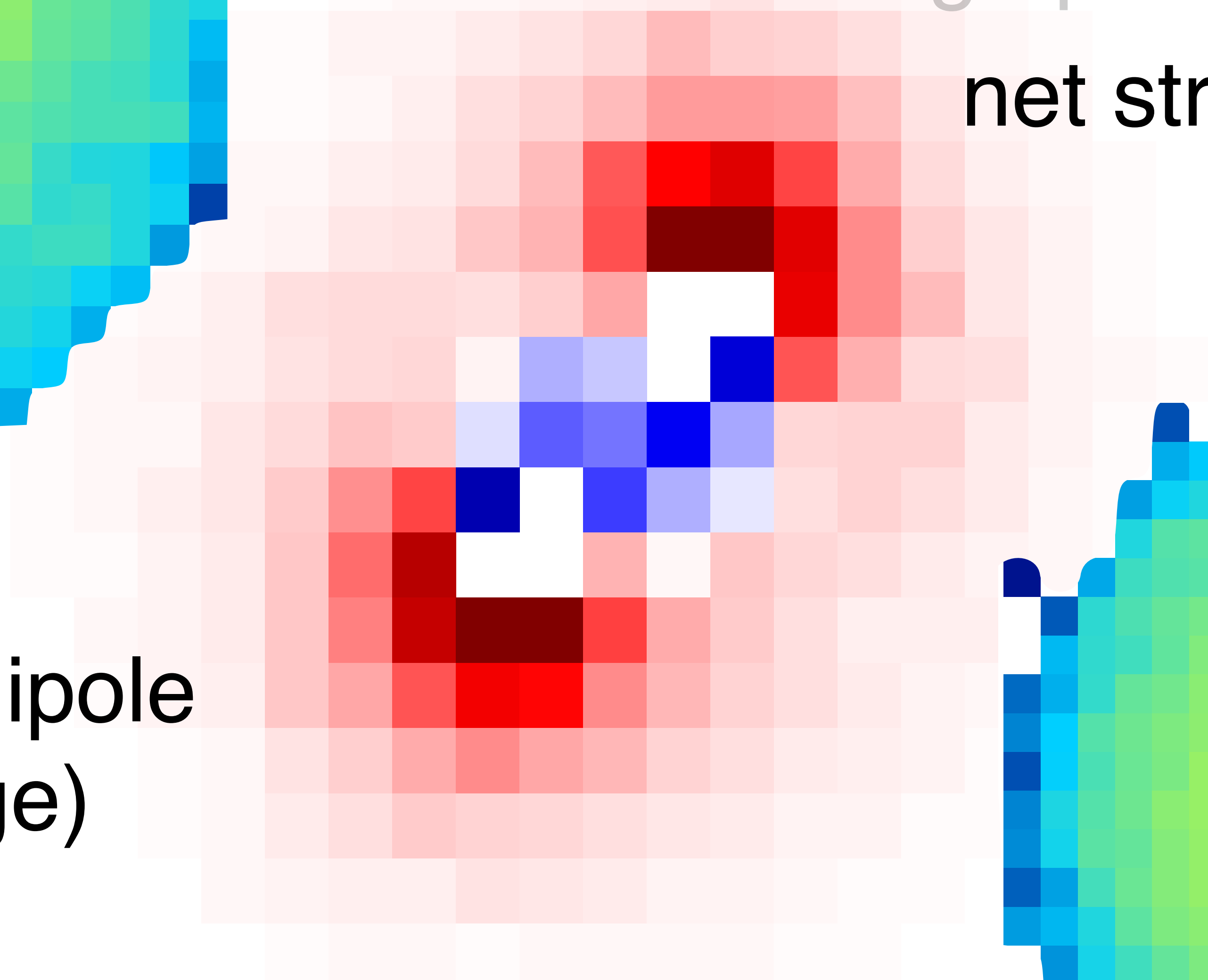
there is information encoded in the physical distance between pixels

Why Images?



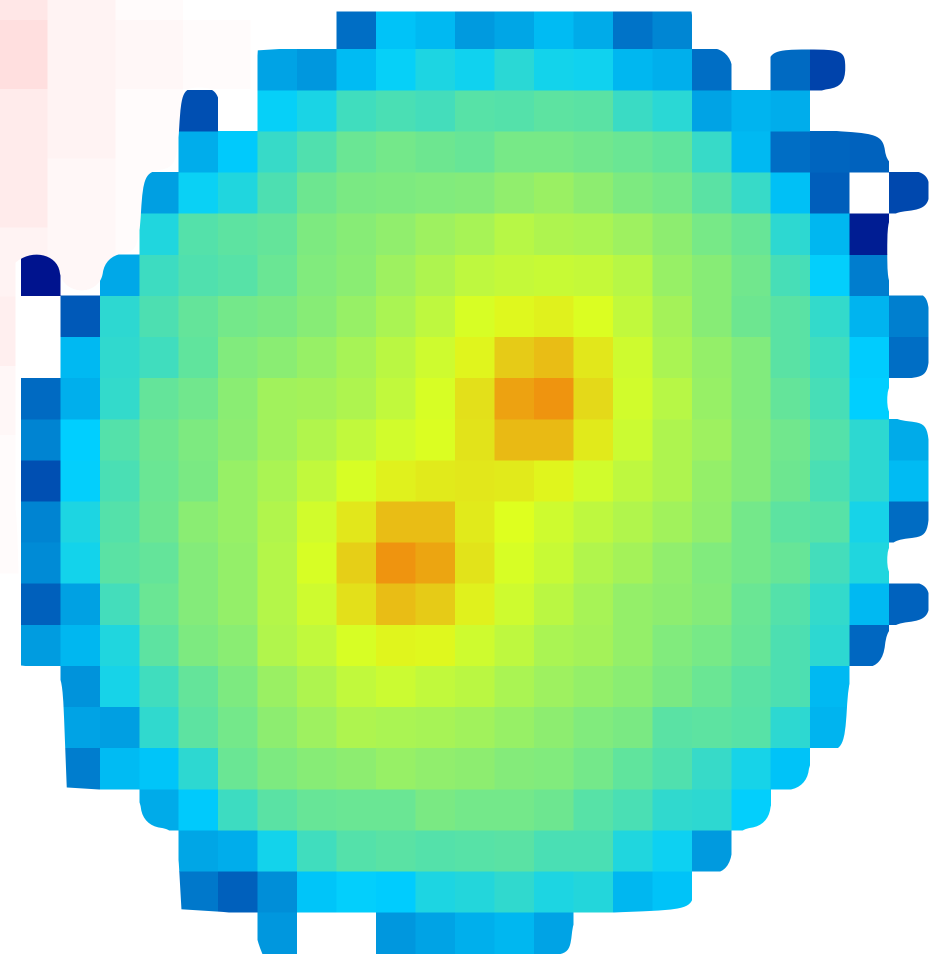
Can directly visualize physics

and we can benefit from the extensive image processing literature



net strong-force charge

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

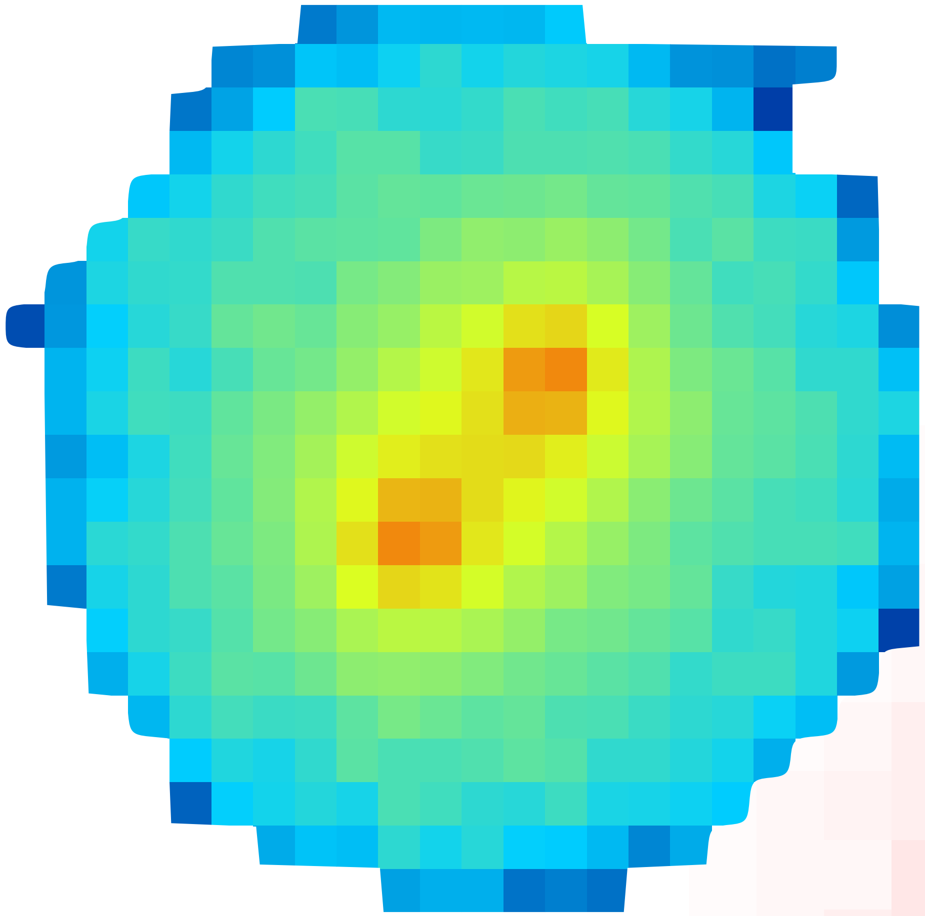


$\mathbf{W} \dashrightarrow q\bar{q}$

radiates like a dipole (no net charge)

there is information encoded in the physical distance between pixels

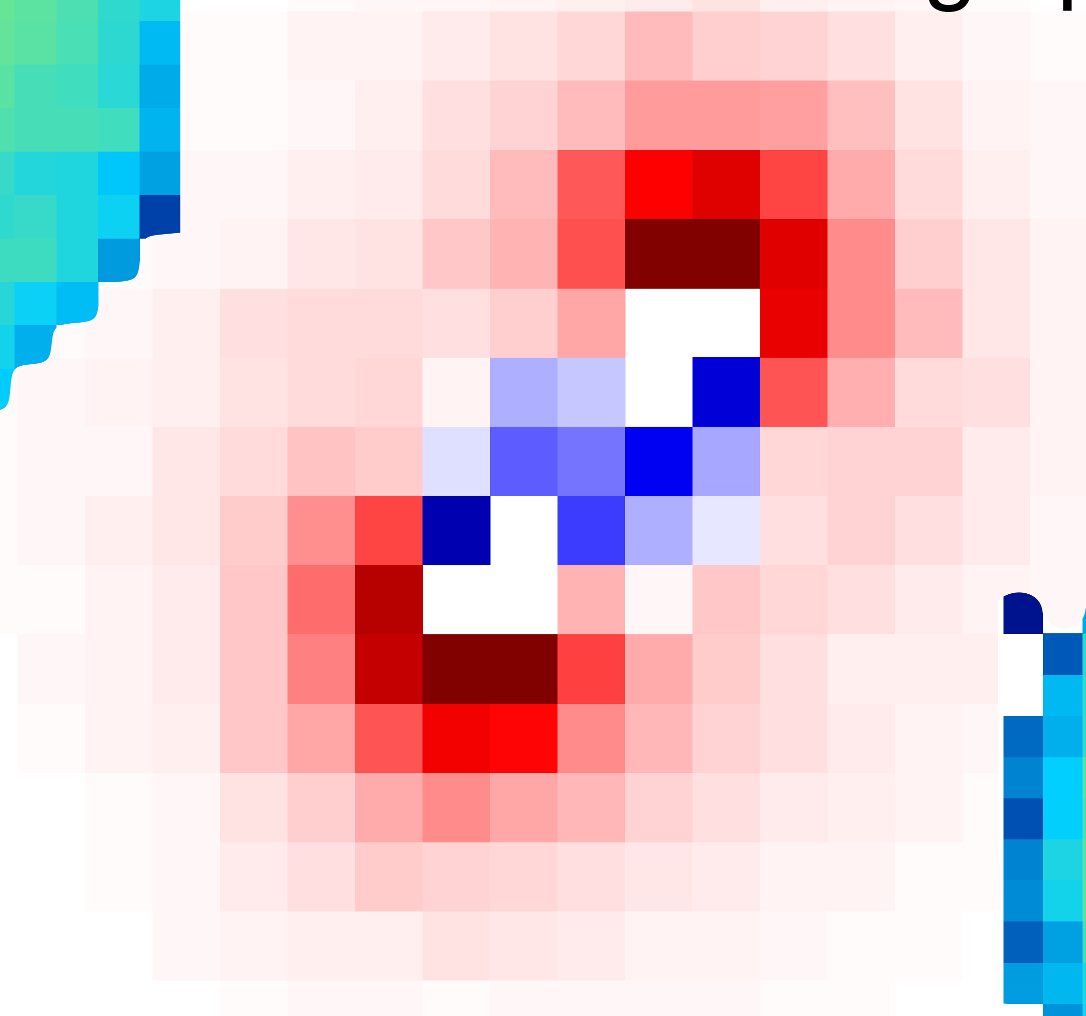
Why Images?



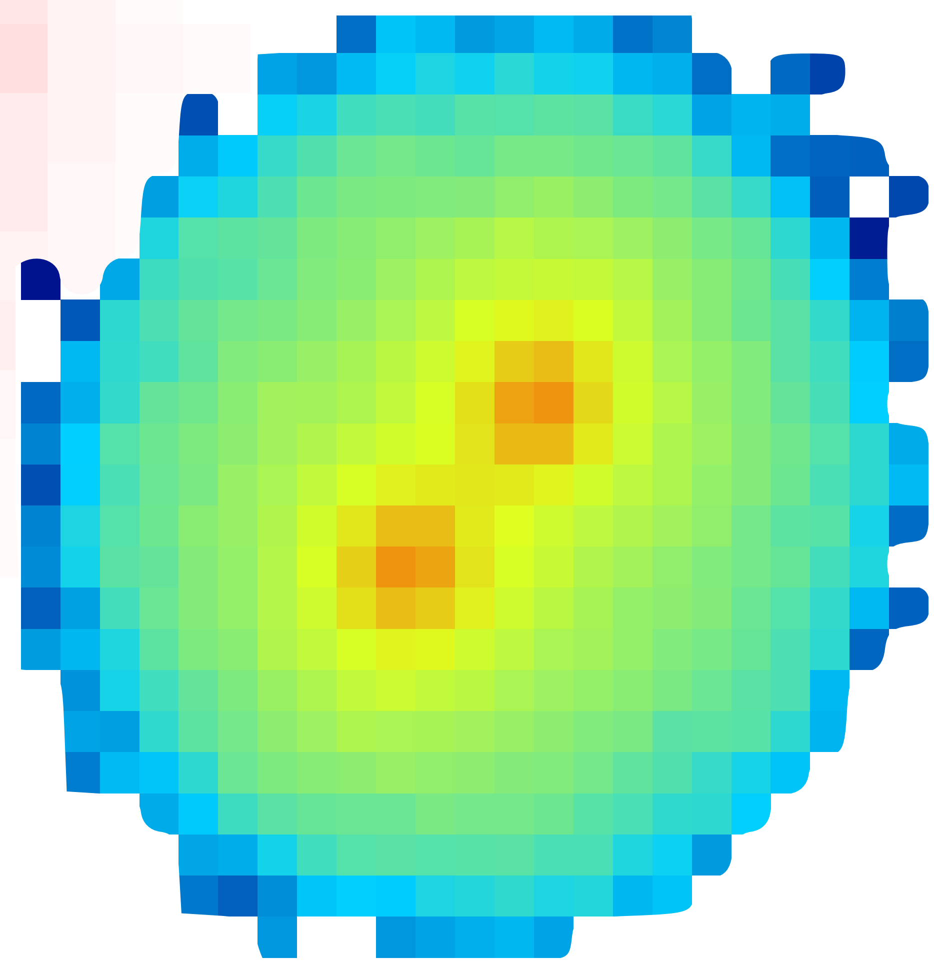
$W \dashrightarrow q\bar{q}$

Can directly visualize physics

and we can benefit from the extensive image processing literature



$g \dashrightarrow q\bar{q}$

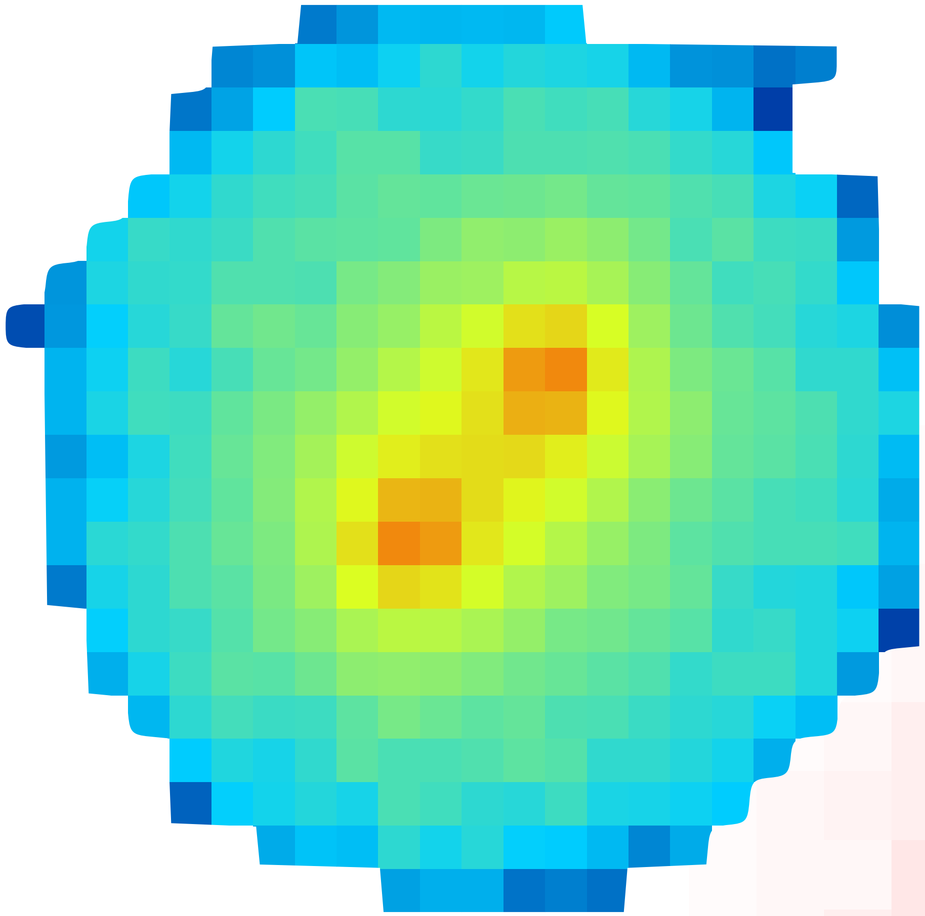


there is information encoded in the physical distance between pixels

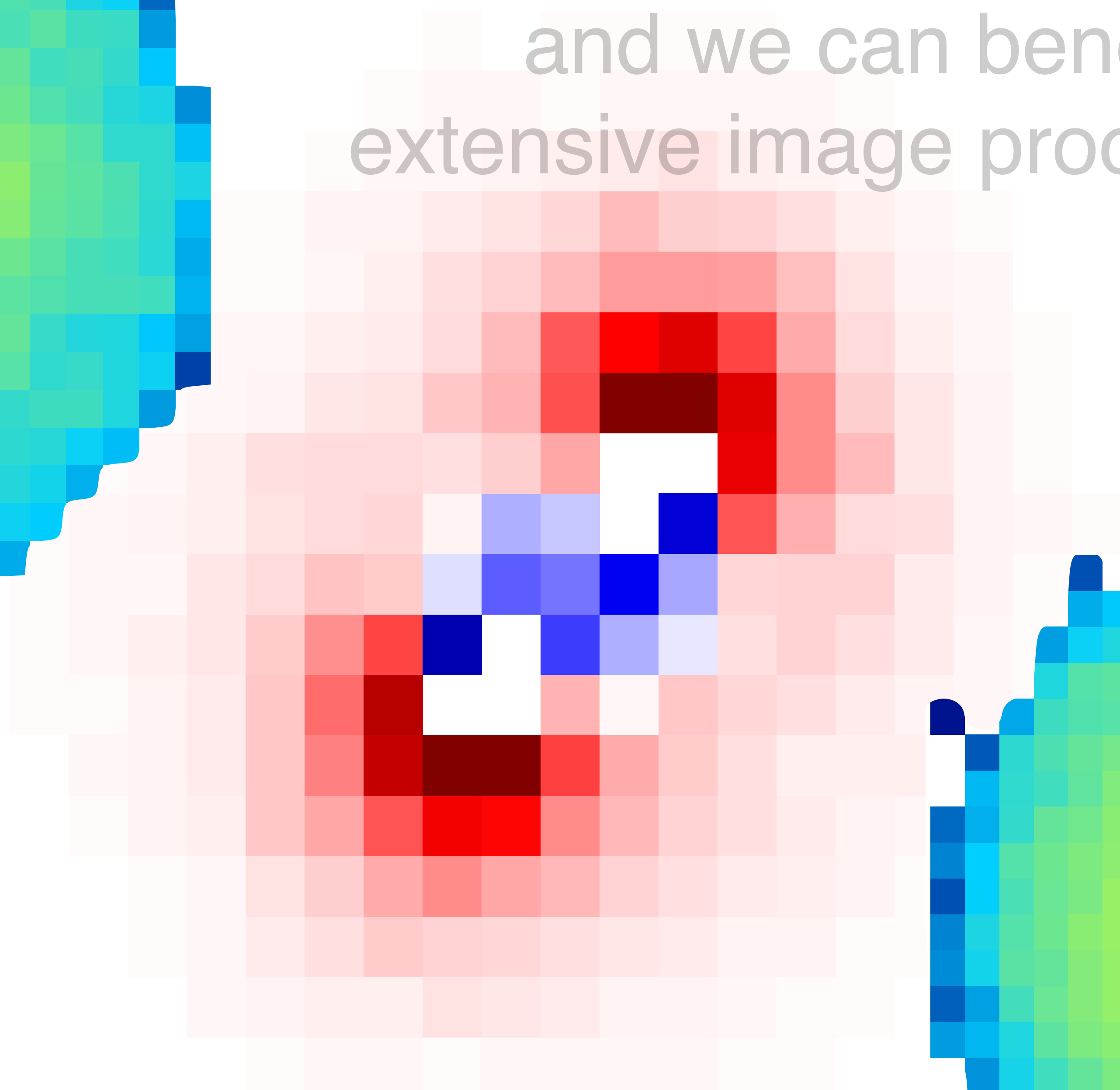
Why Images?

Can directly visualize physics

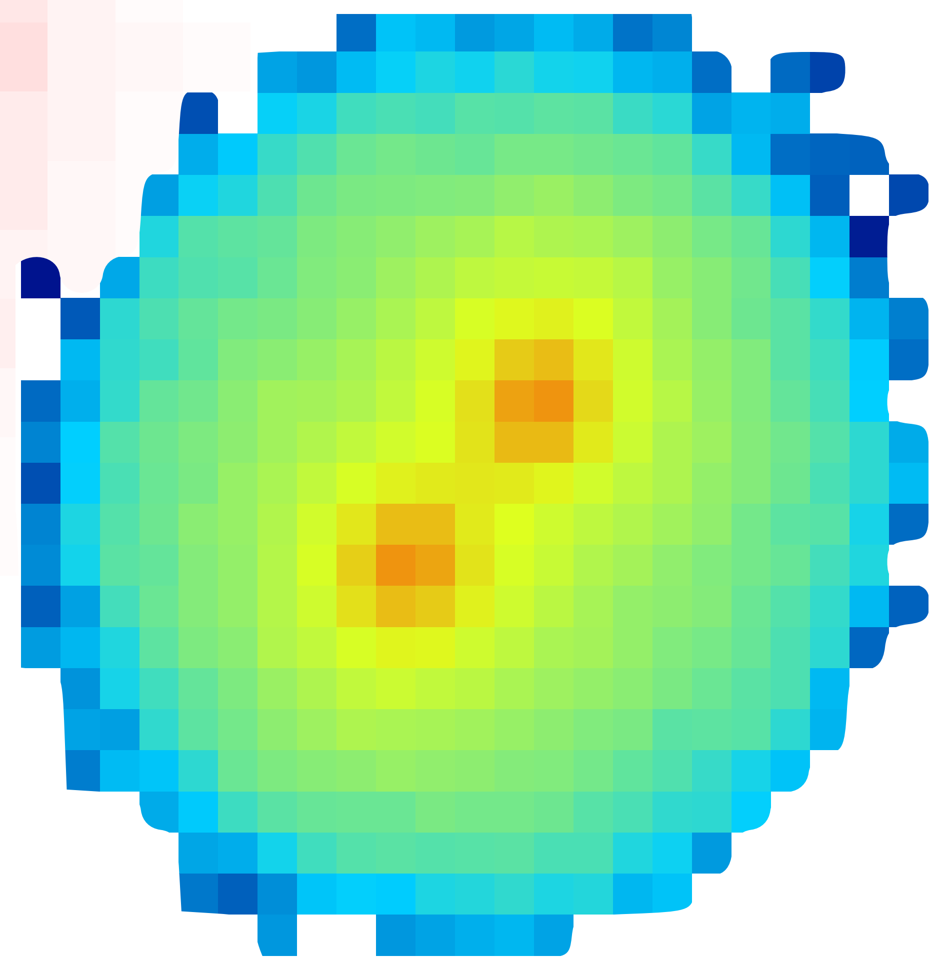
and we can benefit from the extensive image processing literature



W ⇢ $q\bar{q}$



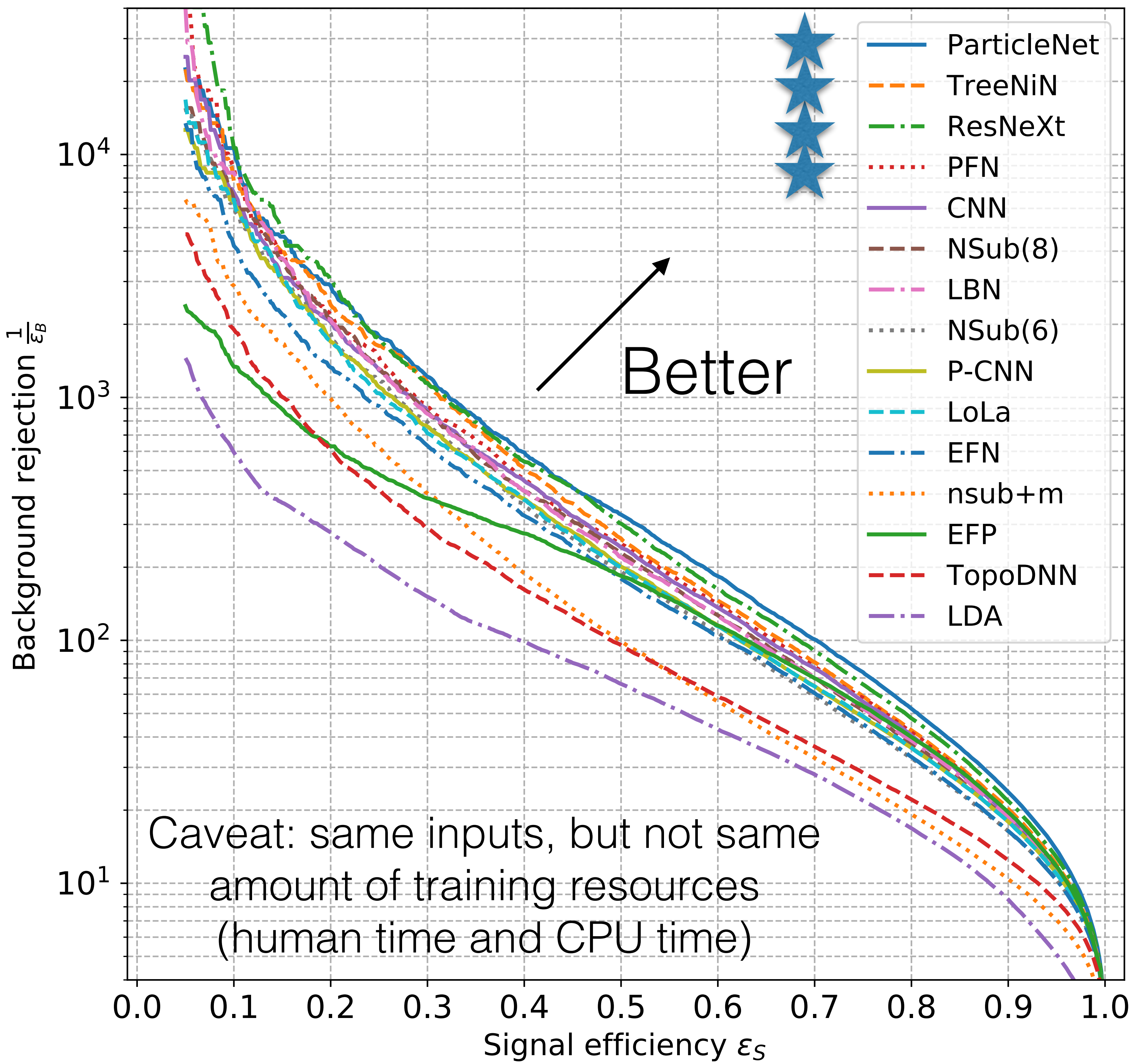
g ⇢ $q\bar{q}$



there is information encoded in the physical distance between pixels

Images and Beyond

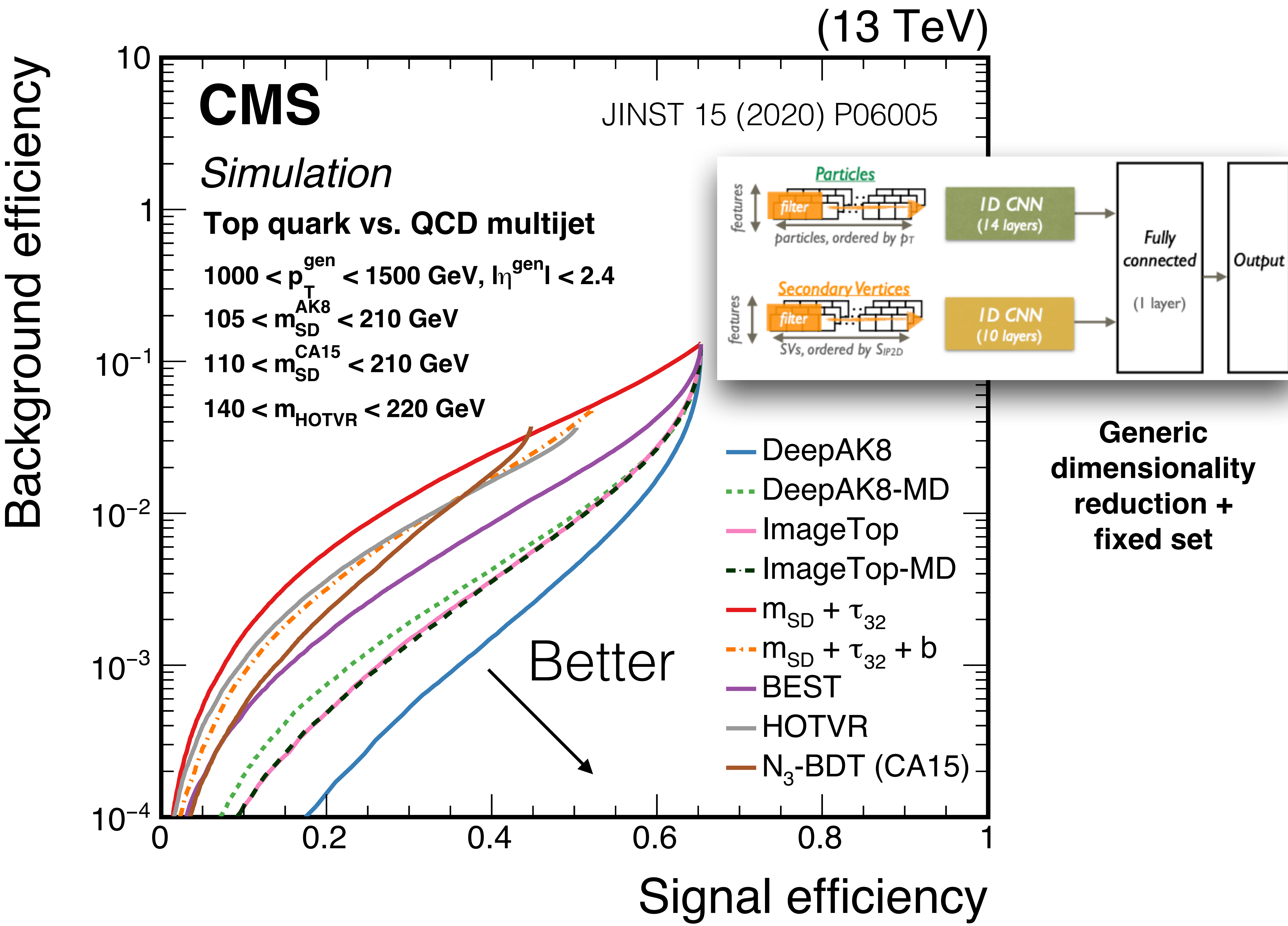
G. Kasieczka and T. Plehn et al., SciPost Phys. 7, 014 (2019)



Graphs
Tree
Images
Variable Sets
Images
Fixed set
*
Fixed set
Sequence
*
Variable Sets
Fixed set
Variable Sets
Fixed set
Variable Sets
(unsupervised)

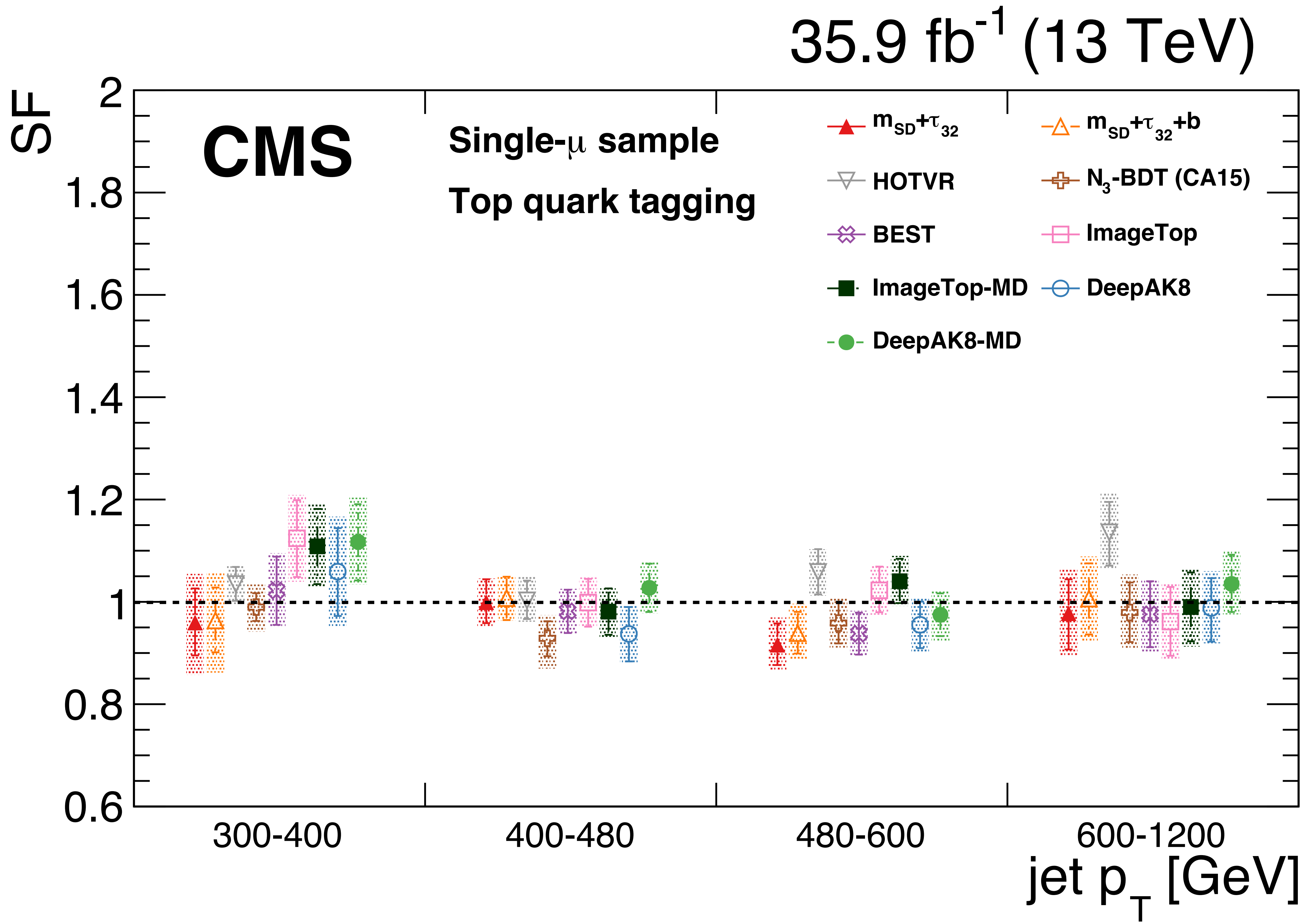
*** = structured dimensionality reduction + fixed set**

Classification in Practice



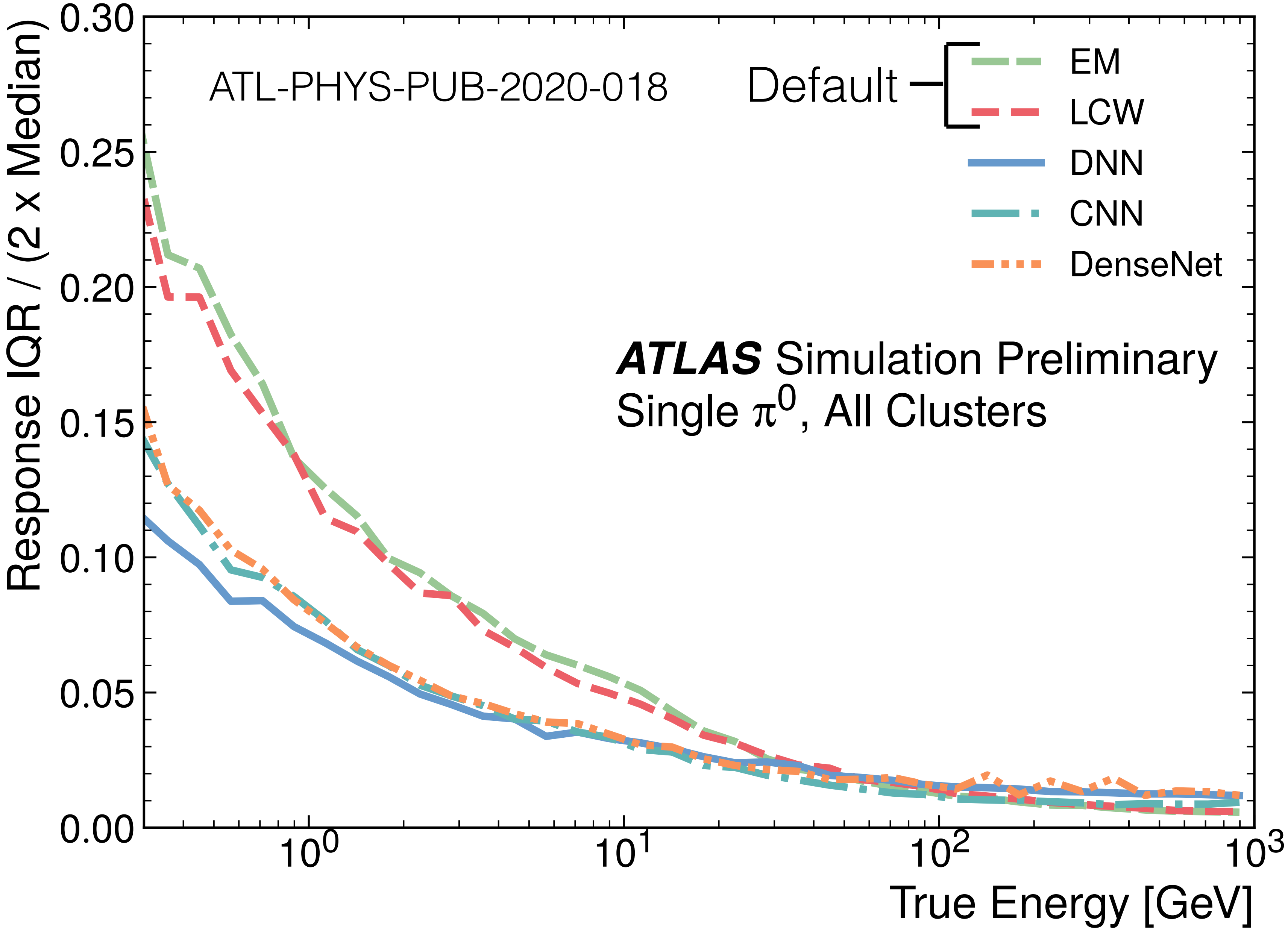
Generic dimensionality reduction + fixed set

Classification in Practice

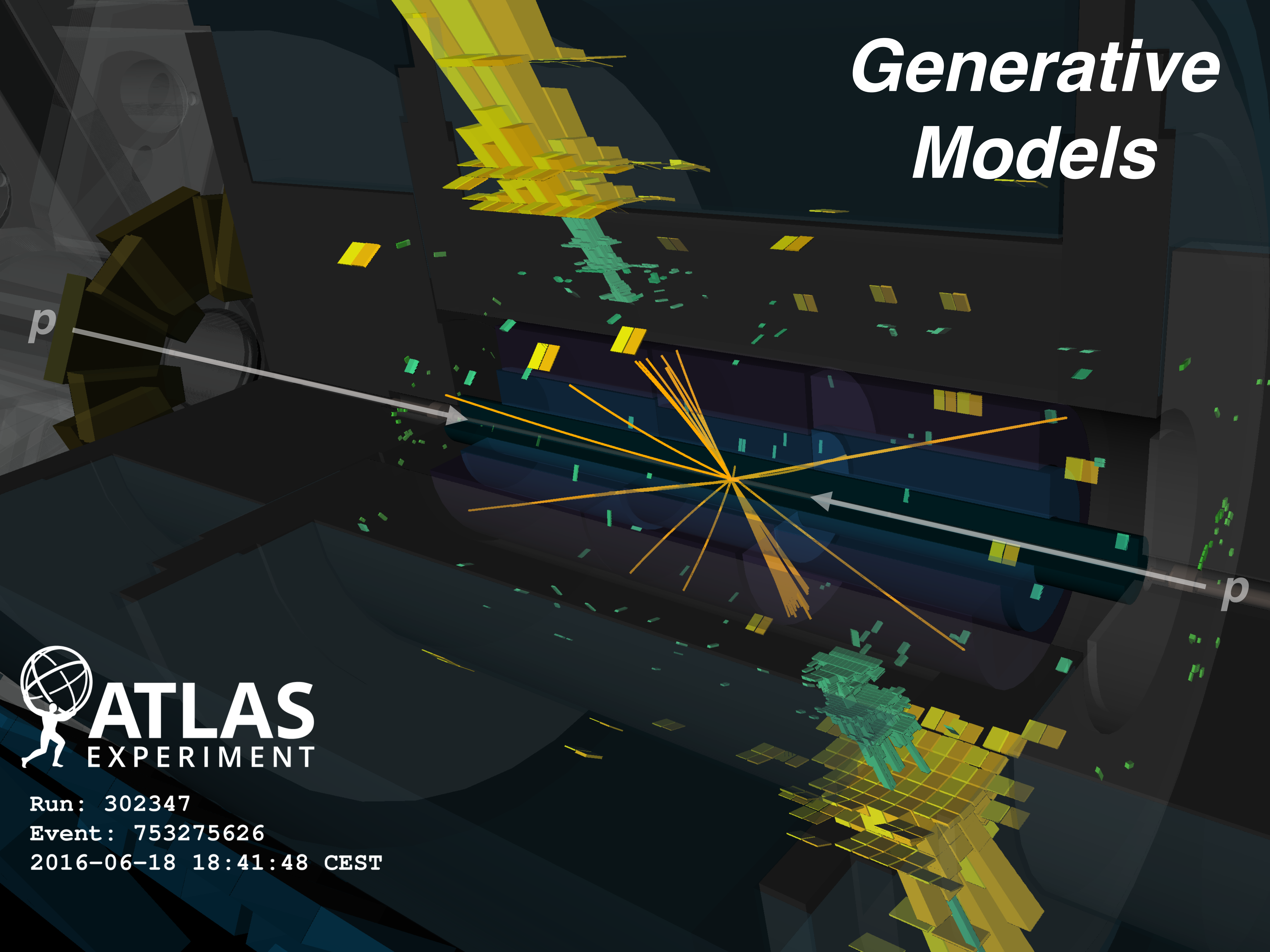


One last comment: low-level learning

Top tagging improves when you use all of the available information

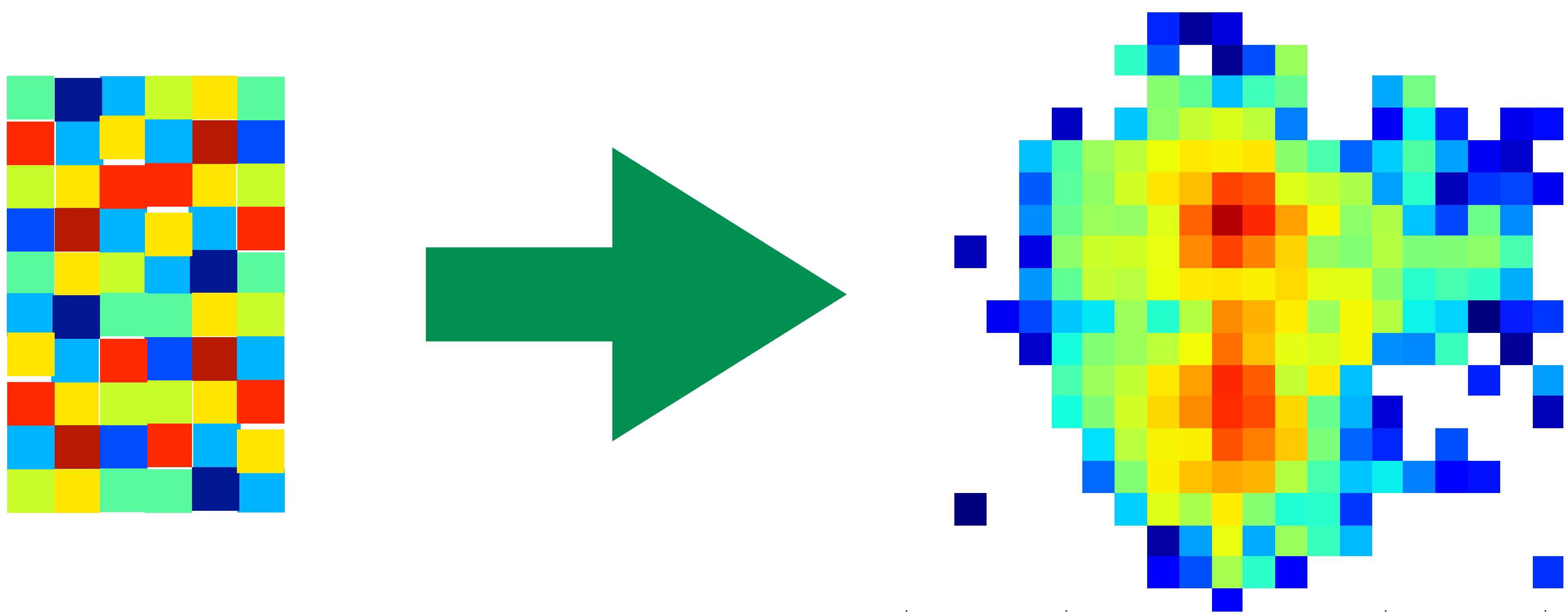


It is worth also taking a step back to ensure that all of the inputs are themselves optimized


Generative Models
p
p
ATLAS
EXPERIMENT
Run: 302347
Event: 753275626
2016-06-18 18:41:48 CEST

Generative models

You will hear all about this in Anja's talk, but I wanted to quickly highlight this exciting possibility.

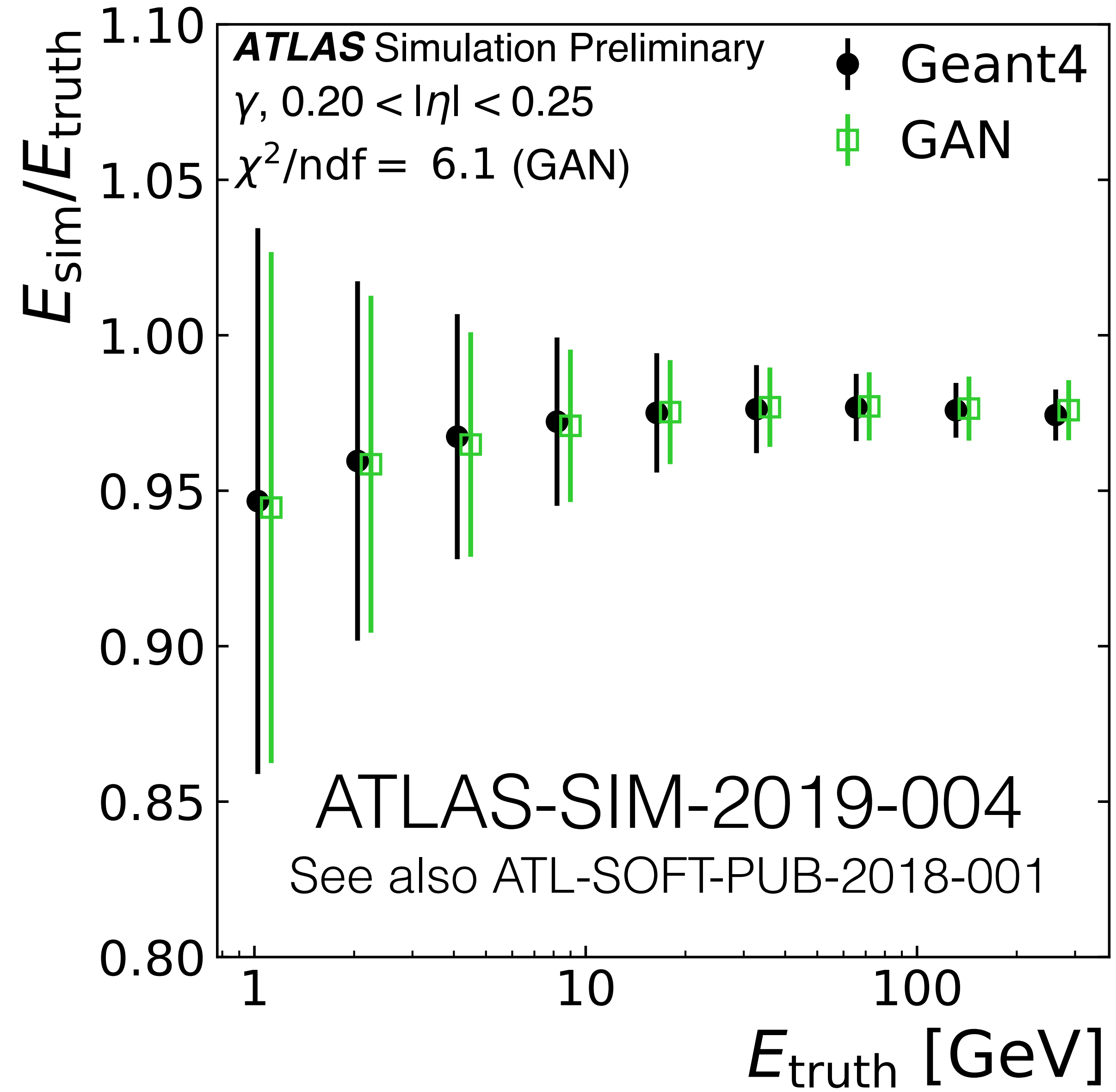


What is a generative model?
Answer: A function from noise to structure.

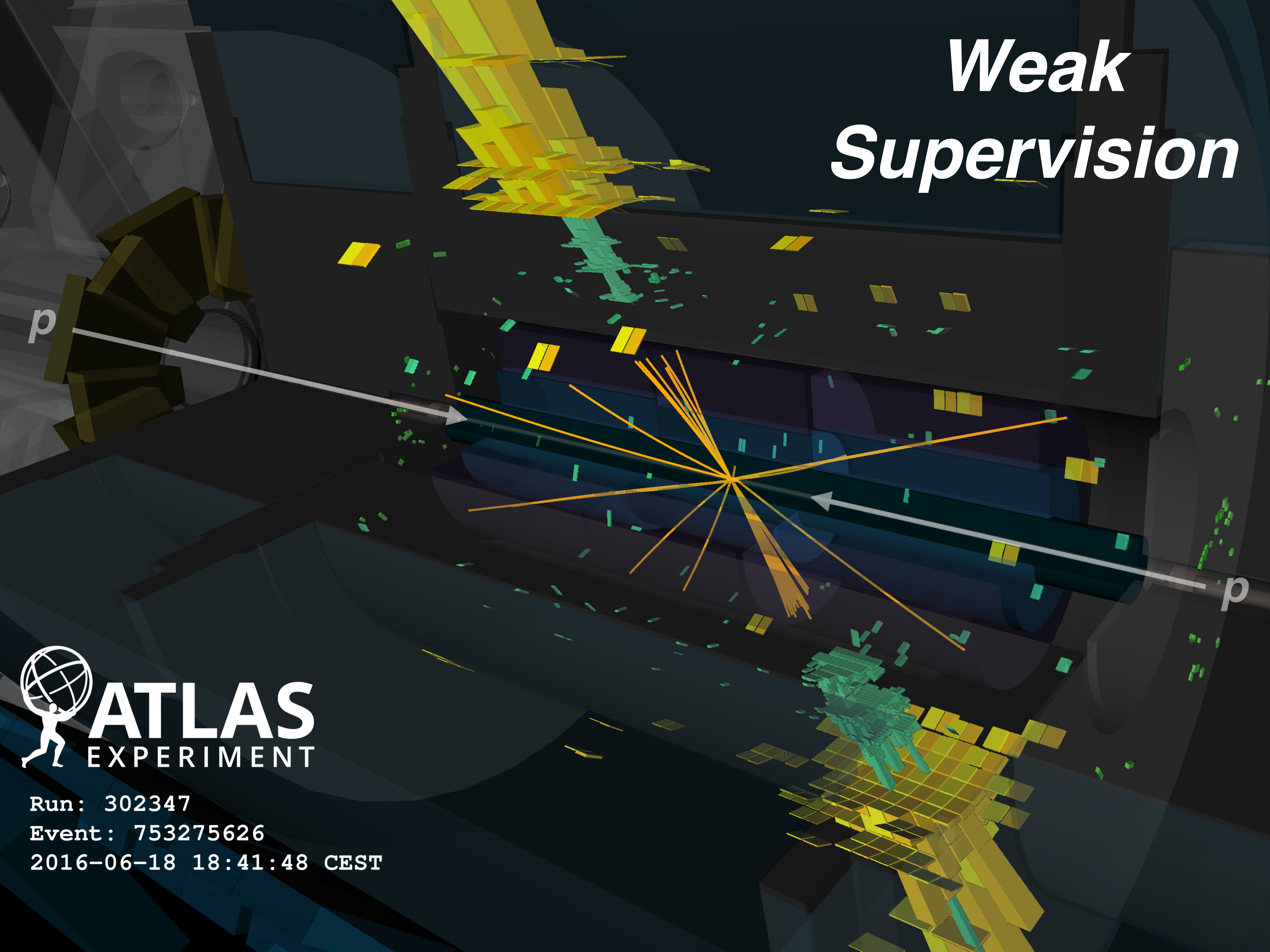
Generative models

These are being used for parton showers, background estimation, distribution subtraction, unfolding, …

These are already being integrated into the experimental workflow and may be able to improve top physics analyses in the future!



See also M. Paganini, L. de Oliveira, and B. Nachman, PRD 97 (2018) 014021
M. Paganini, L. de Oliveira, and B. Nachman, PRL 120 (2018) 042003


Weak
Supervision
p
p
ATLAS
EXPERIMENT
Run: 302347
Event: 753275626
2016-06-18 18:41:48 CEST

Why can't we learn directly from data?

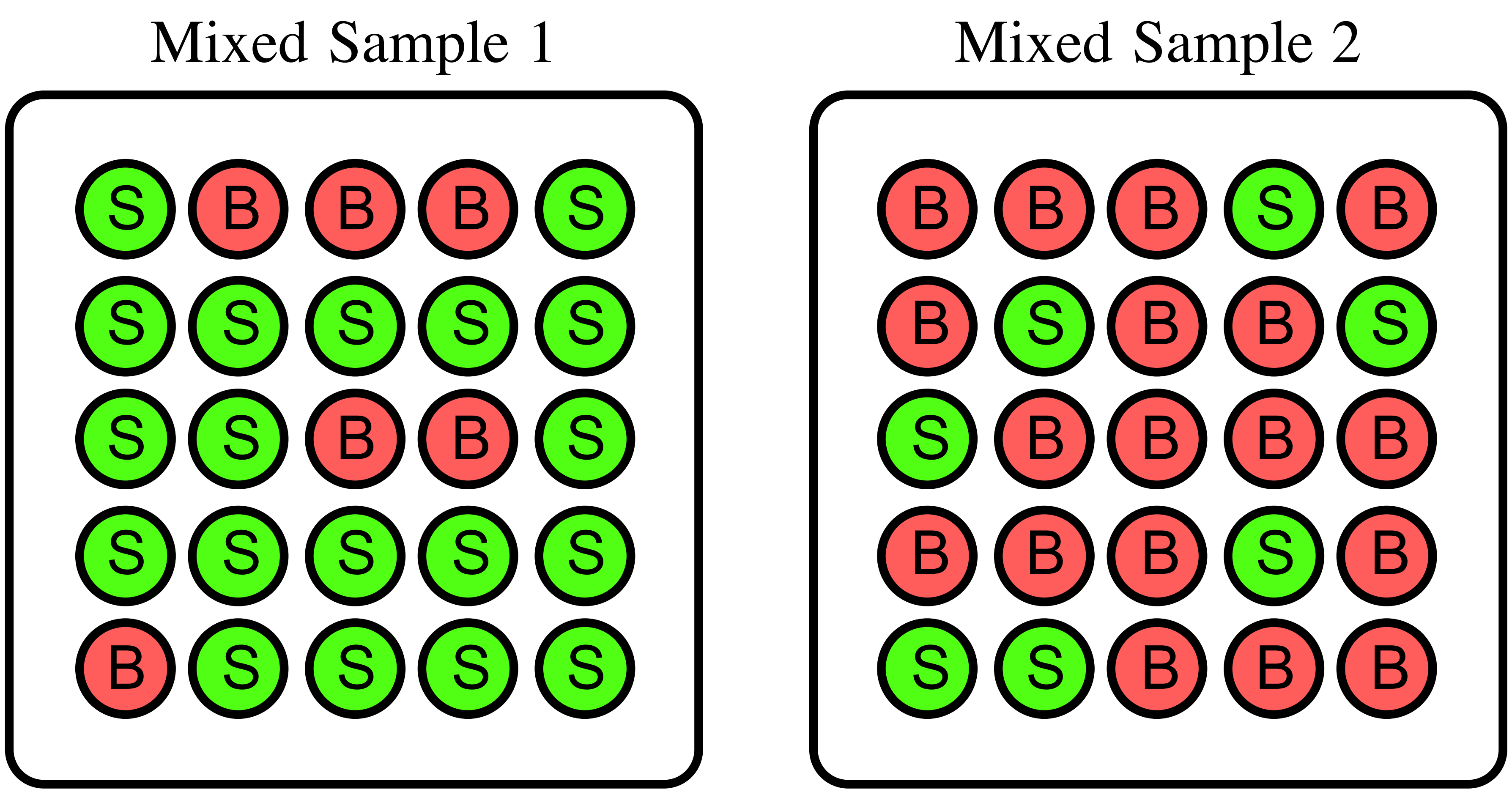
Why can't I just pay some physicists to label events and then train a neural network using those labels?

Image credit: pixabay.com

Answer: this is not cats-versus-dogs … thanks to quantum mechanics it is **not possible to know** what happened.

What is the problem?

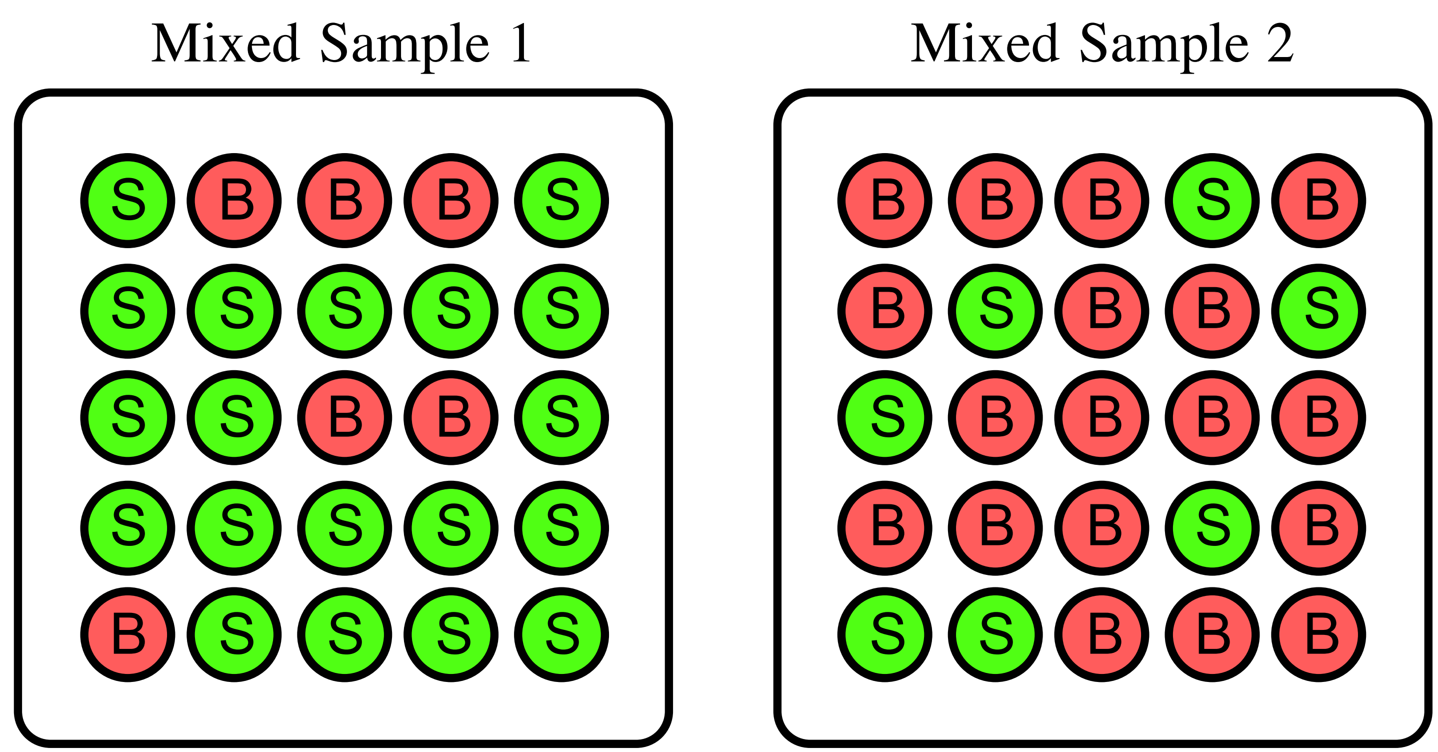
The data are unlabeled and in the best case, come to us as mixtures of two classes (“signal” and “background”).



(we don’t get to observe the color of the circles)

Weak supervision: *Classification Without Labels*

Can we learn without any label information?



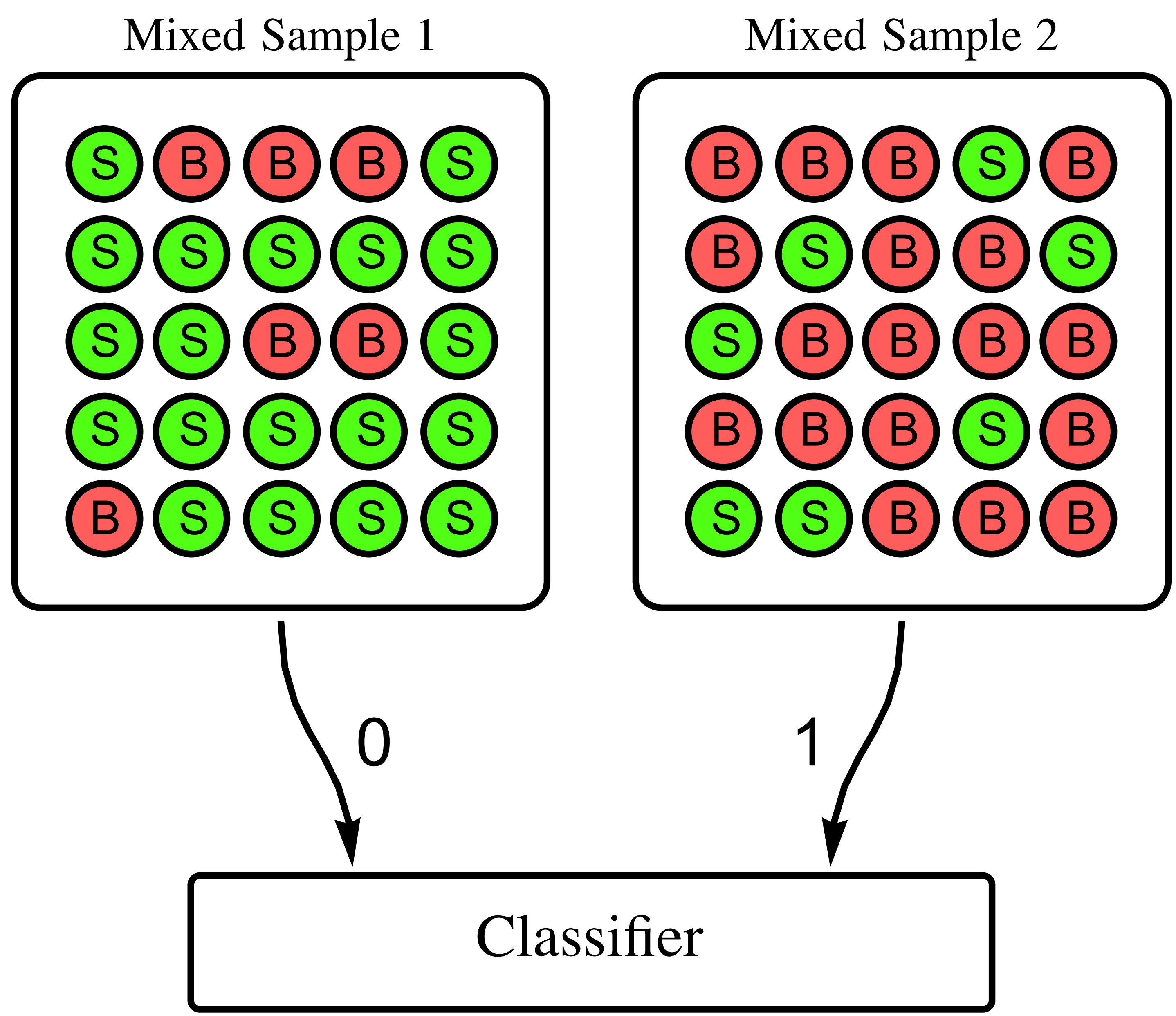
E. Metodiev, BPN, J. Thaler, JHEP 10 (2017) 51

Weak supervision: *Classification Without Labels*

Can we learn without any label information?

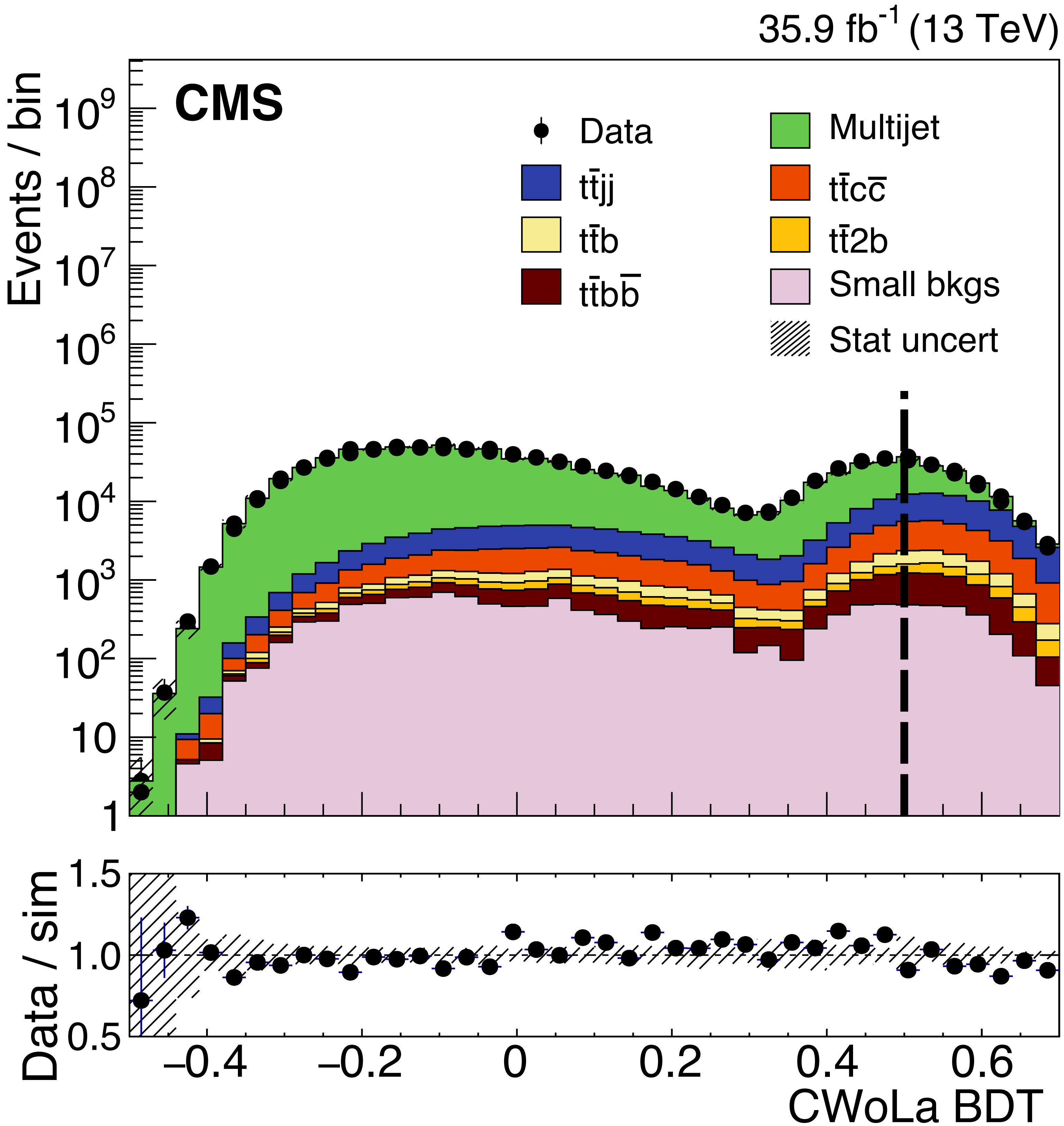
Yes !

Training on impure samples is (asymptotically) equivalent to training on pure samples



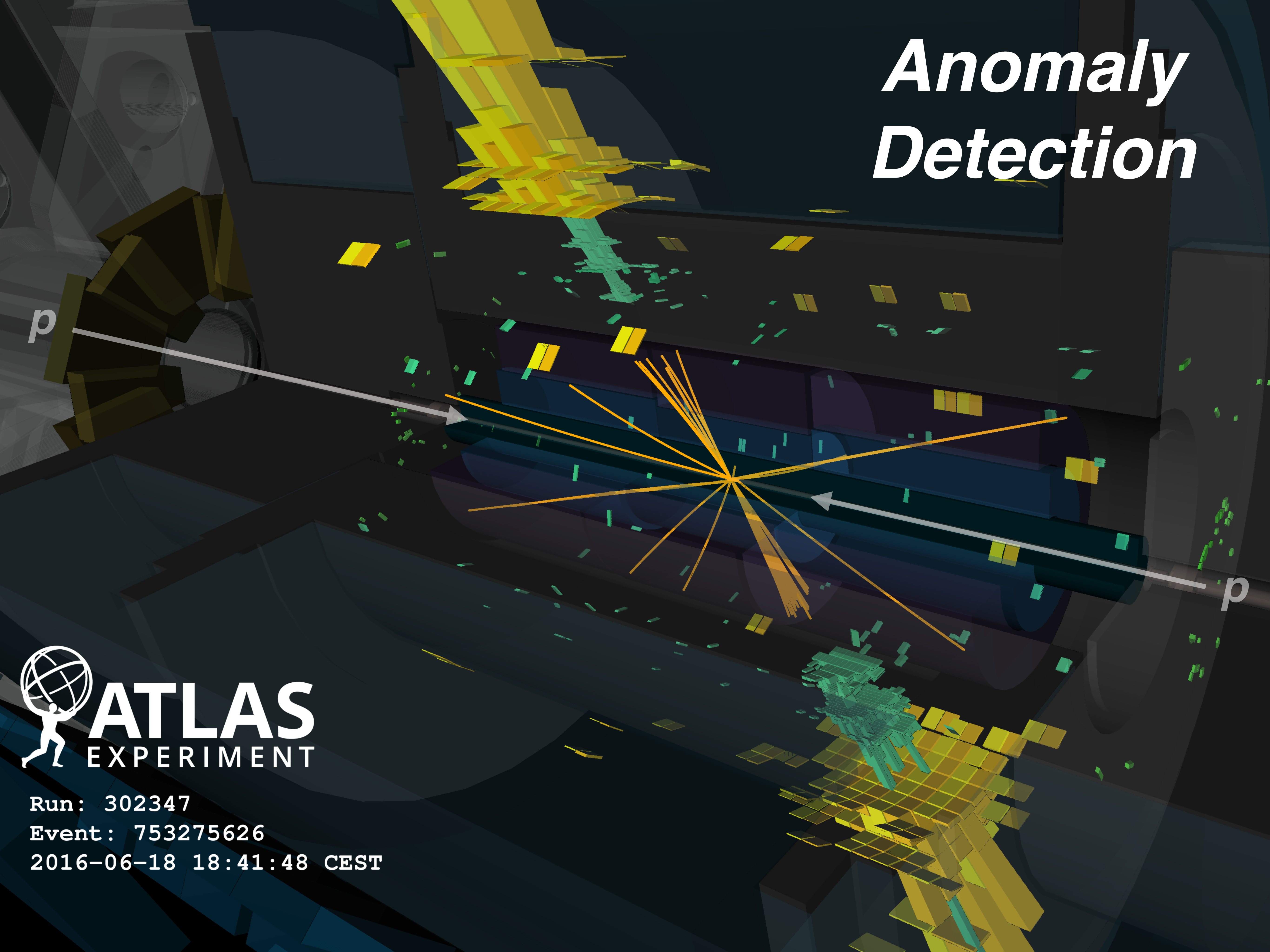
E. Metodiev, **BPN**, J. Thaler, JHEP 10 (2017) 51

CWoLa in action: tt + bb



Multijet background is hard to model - learn a classifier directly from data using jet substructure to make two samples and use jet kinematics to train the CWoLa classifier

Phys. Lett. B 803 (2020) 135285


Anomaly Detection
p
p
ATLAS
EXPERIMENT
Run: 302347
Event: 753275626
2016-06-18 18:41:48 CEST

Anomaly Detection

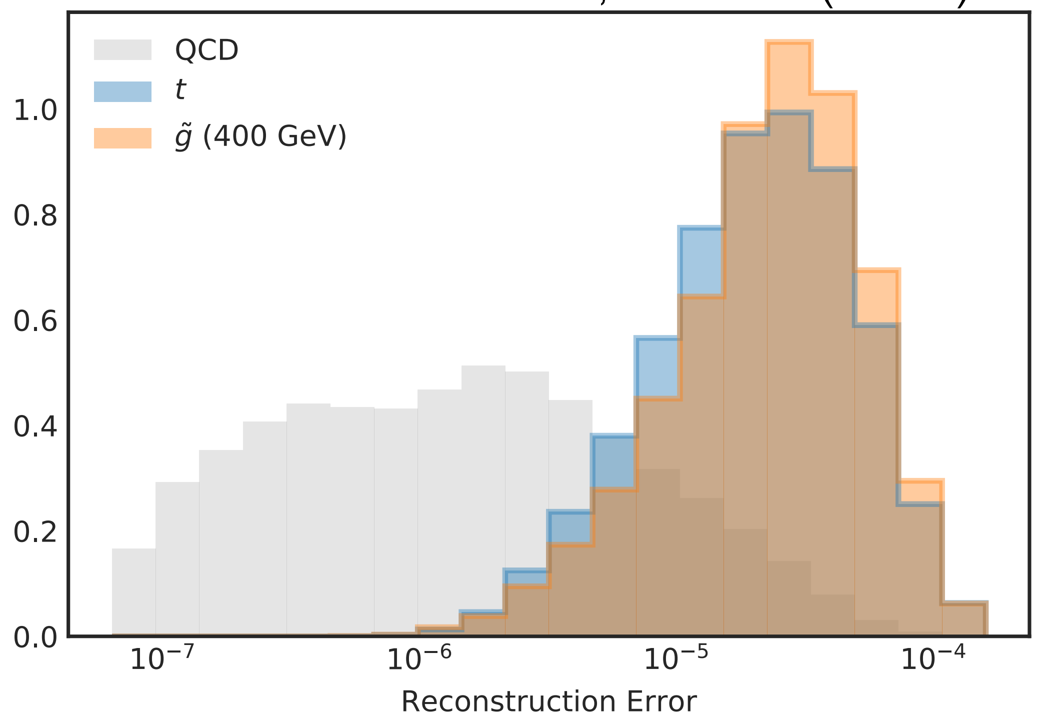
Weak supervision was motivated by inaccurate models; what if we do not even know what we are looking for?

We can use machine learning to ask if there are anomalous features in our data.

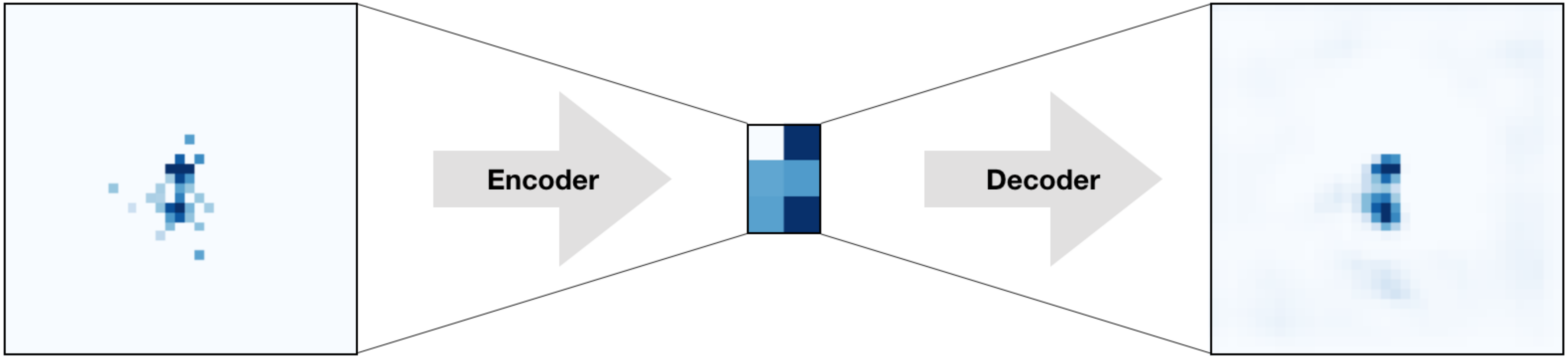
(Boosted) top quarks have been a key benchmark to study the performance of these new tools.

Unsupervised Anomaly Detection

M. Farinia, Y. Nakai, D. Shih,
PRD 101, 075021 (2020)

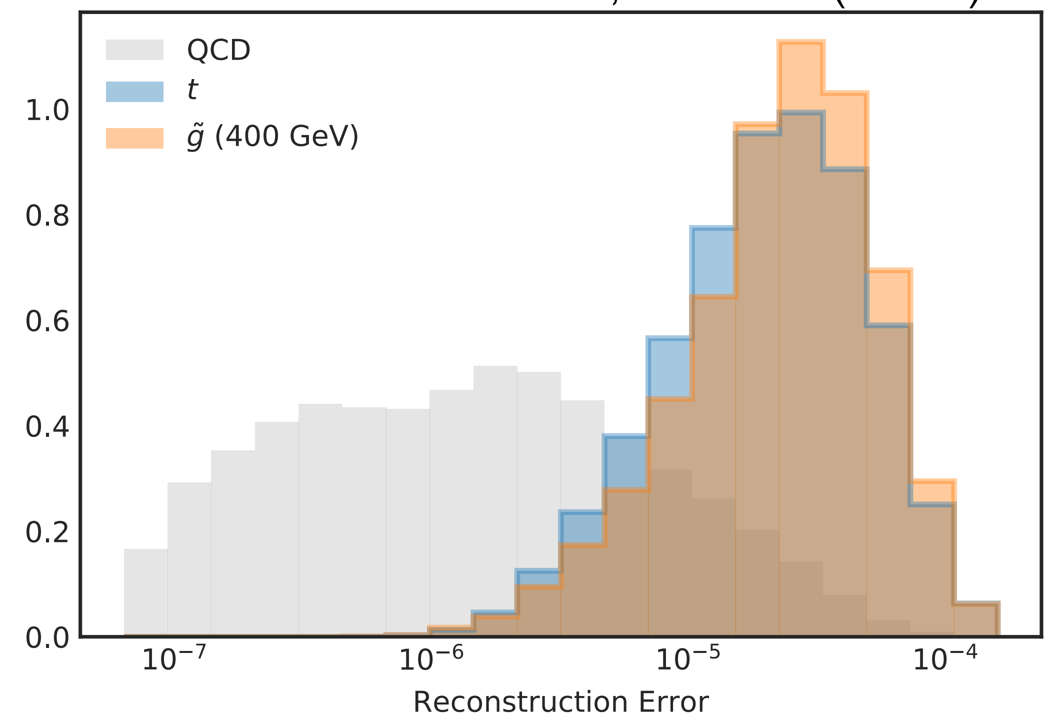


Multiple proposals for asking “which jets/events are strange”?

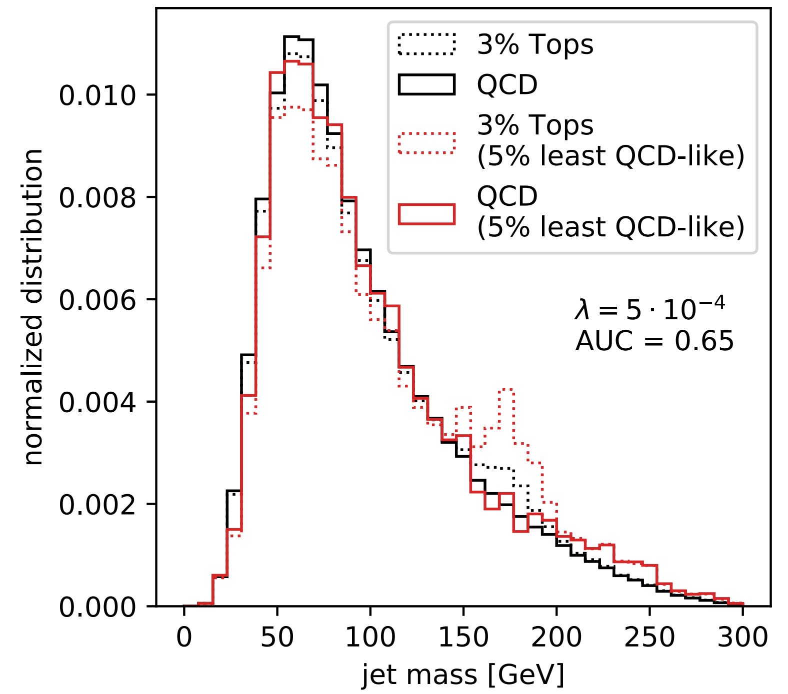


Unsupervised Anomaly Detection

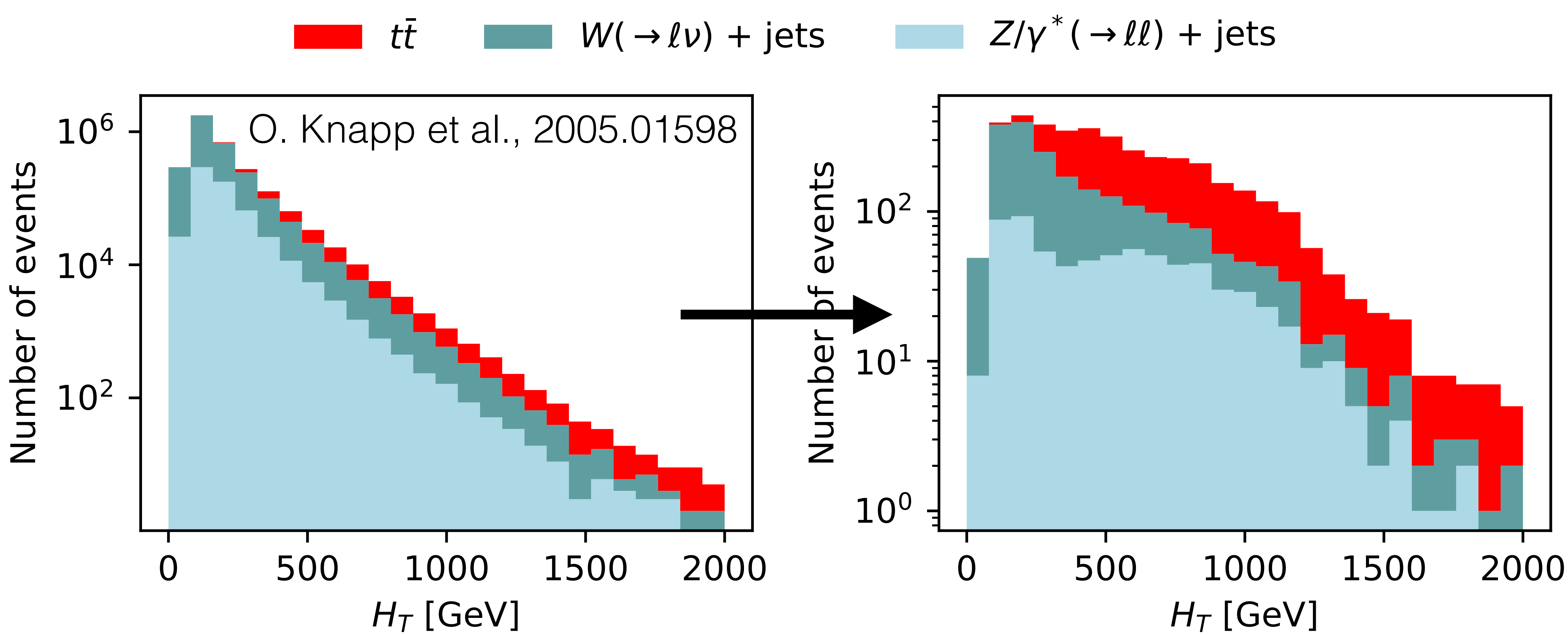
M. Farinia, Y. Nakai, D. Shih,
PRD 101, 075021 (2020)



T. Heimel, et. al,
SciPost Phys. 6, 030 (2019)



Multiple proposals for asking "which jets/events are strange"?



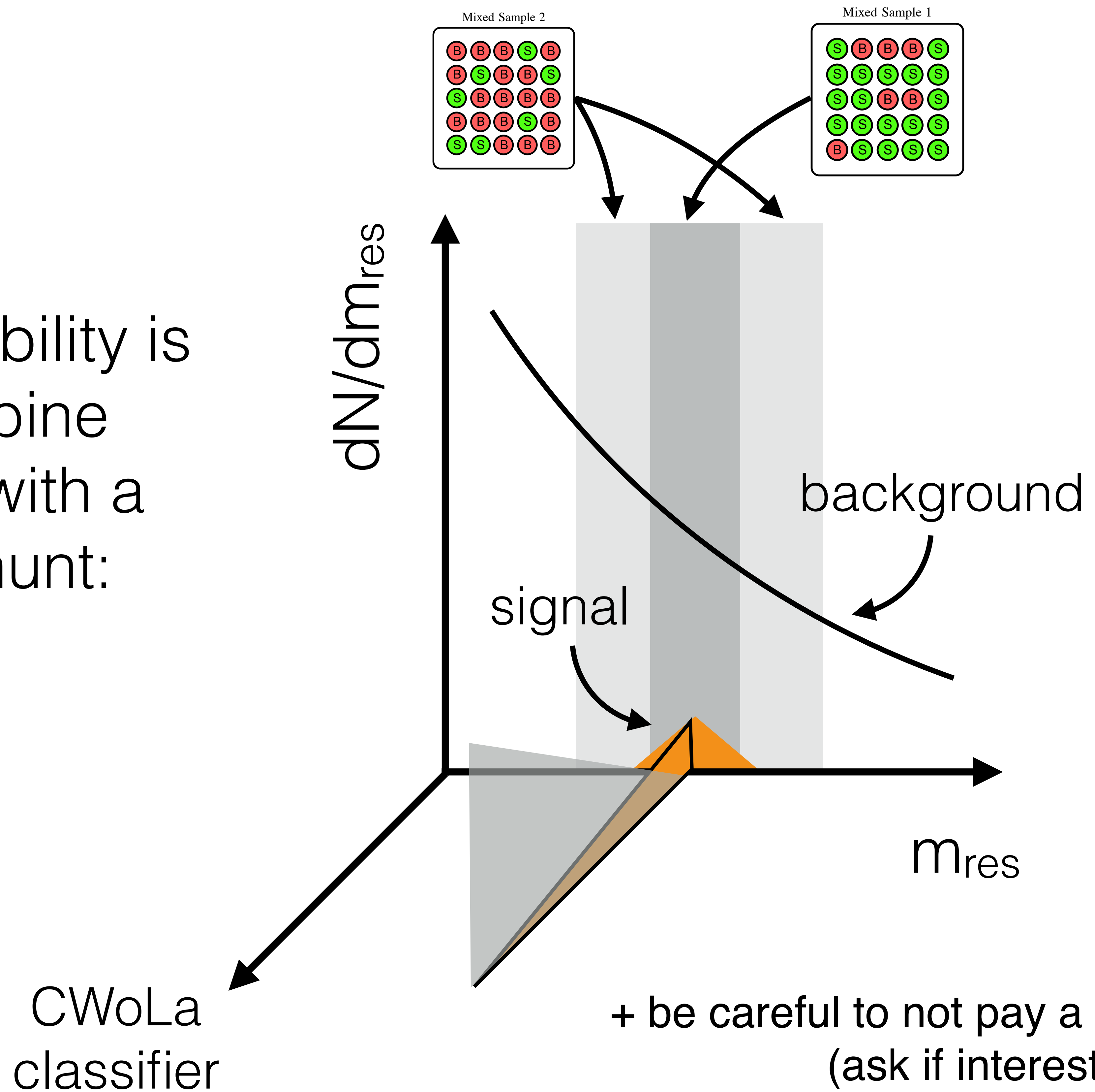
Weakly Supervised Anomaly Detection

One feature of unsupervised learning is that it gives you access to events with low p(x). However, the signal may have high p(x) but p(x|signal) is far from p(x|background).

Semi-supervised methods may be useful in this case.

Weakly Supervised Anomaly Detection

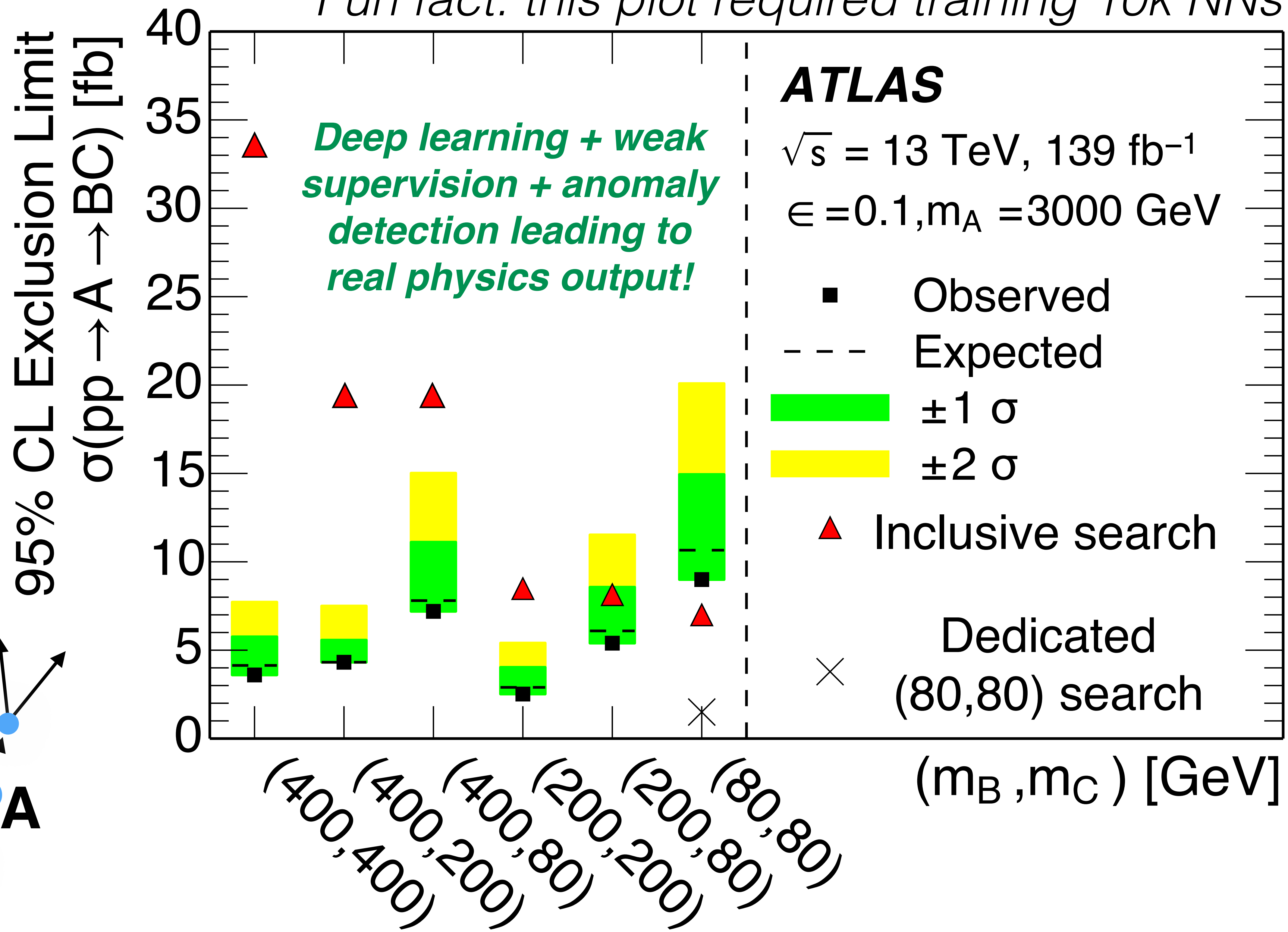
One possibility is to combine CWoLa with a bump hunt:

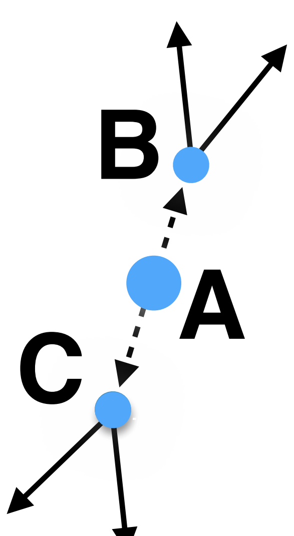


+ be careful to not pay a big trials factor (ask if interested)

Collision data results New

Fun fact: this plot required training 10k NNs



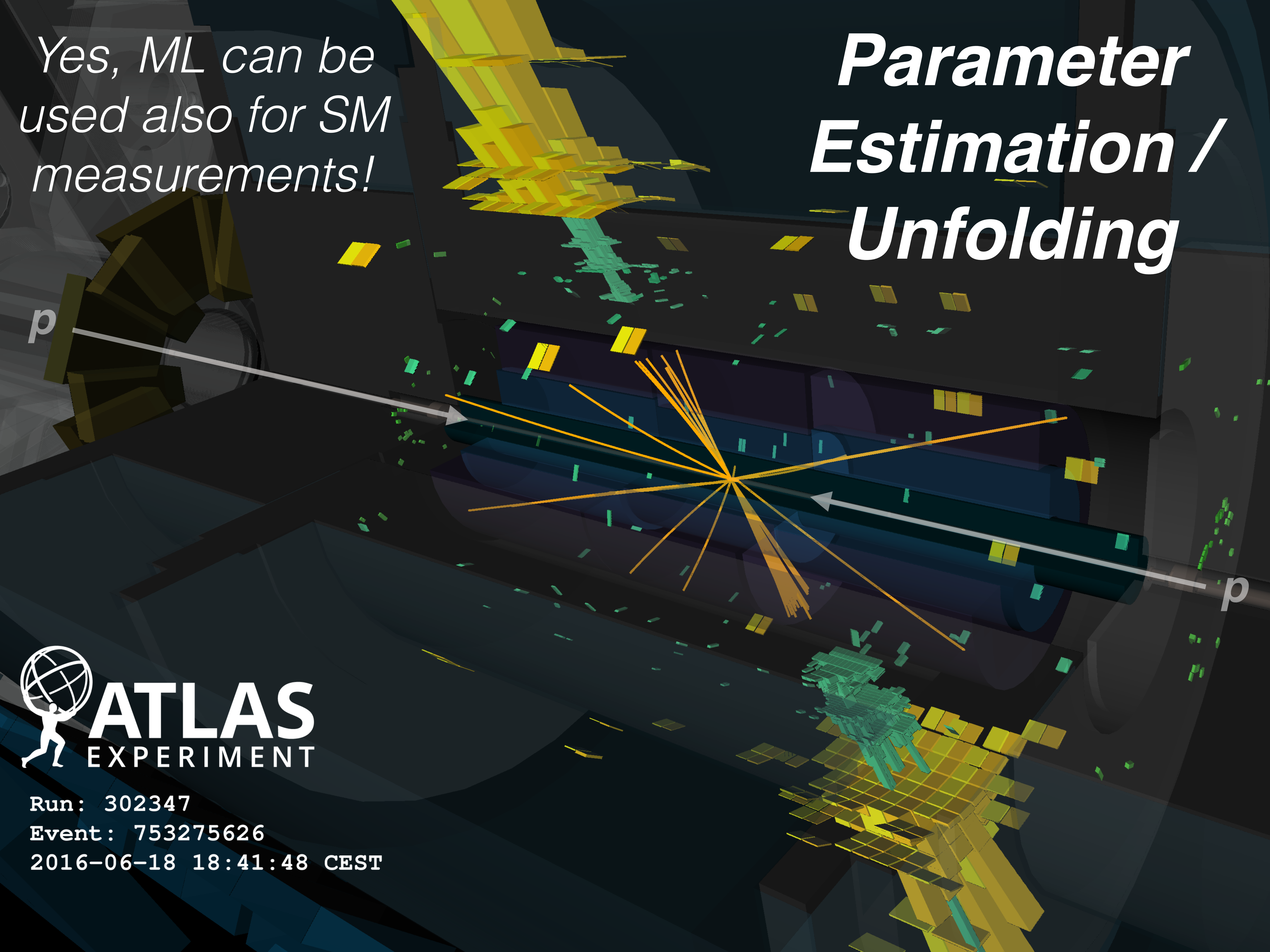


ATLAS Collaboration, 2005.02983
J. Collins, K. Howe, B. Nachman, PRL 121 (2018) 24803
J. Collins, K. Howe, B. Nachman, PRD 99 (2019) 014038
E. Metodiev, B. Nachman, J. Thaler, JHEP 10 (2017) 174

B = top, C = BSM? See J. Kim et al., JHEP 04 (2020) 30

Parameter Estimation / Unfolding

Yes, ML can be used also for SM measurements!



Could we use all of the information?

For our measurements of top quark event properties, would it be possible to use all of the information in the event?

…all hadrons, their 4-vectors, charge, …

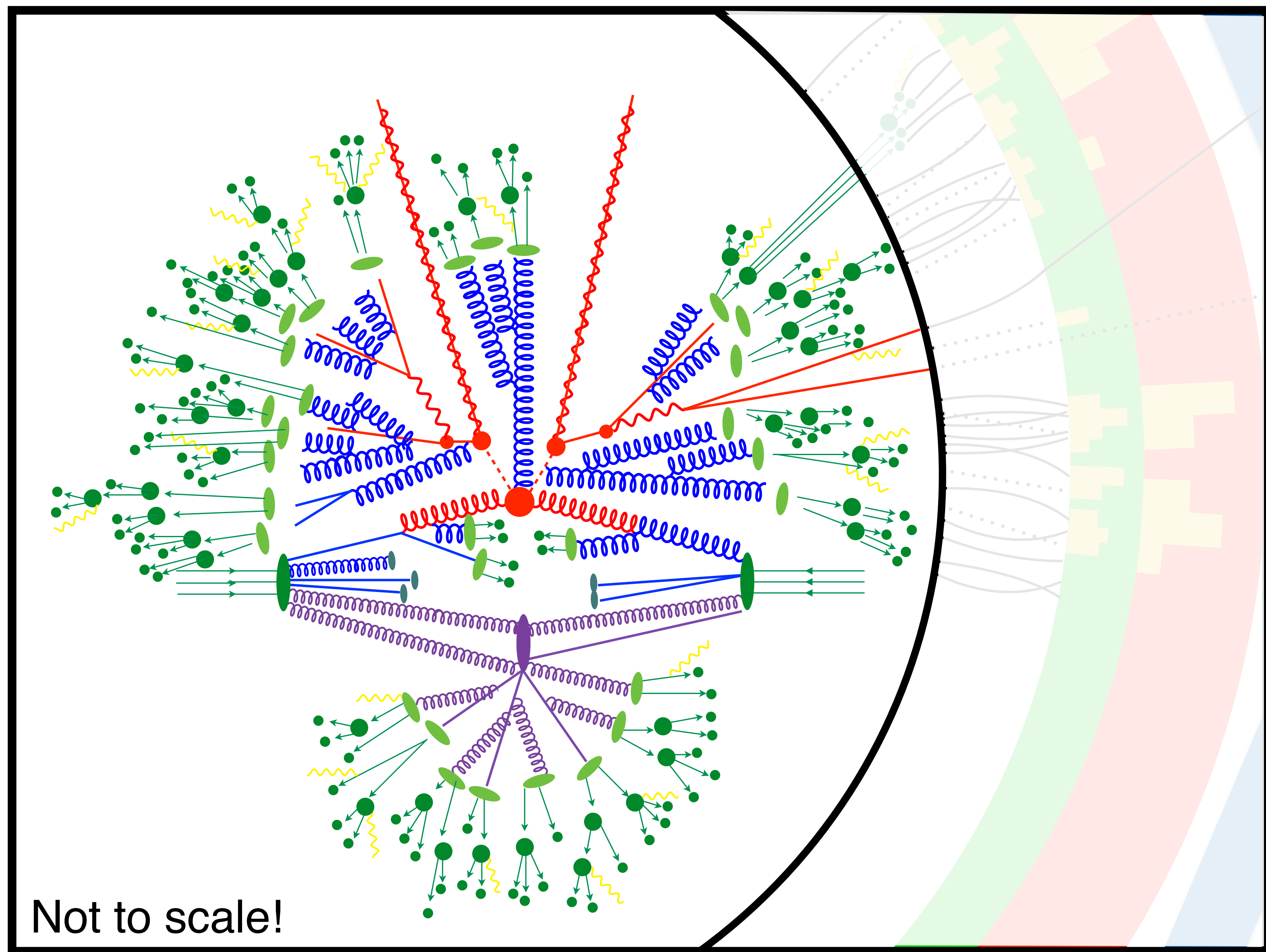
A *hyper* challenge

Key **challenge** and **opportunity**: *hypervariate phase space*
& *hyper spectral data*

Typical collision events at the LHC produce **O(1000+)** particles

We detect these particles with **O(100 M)** readout channels

Image inspired by JHEP 02 (2009) 007



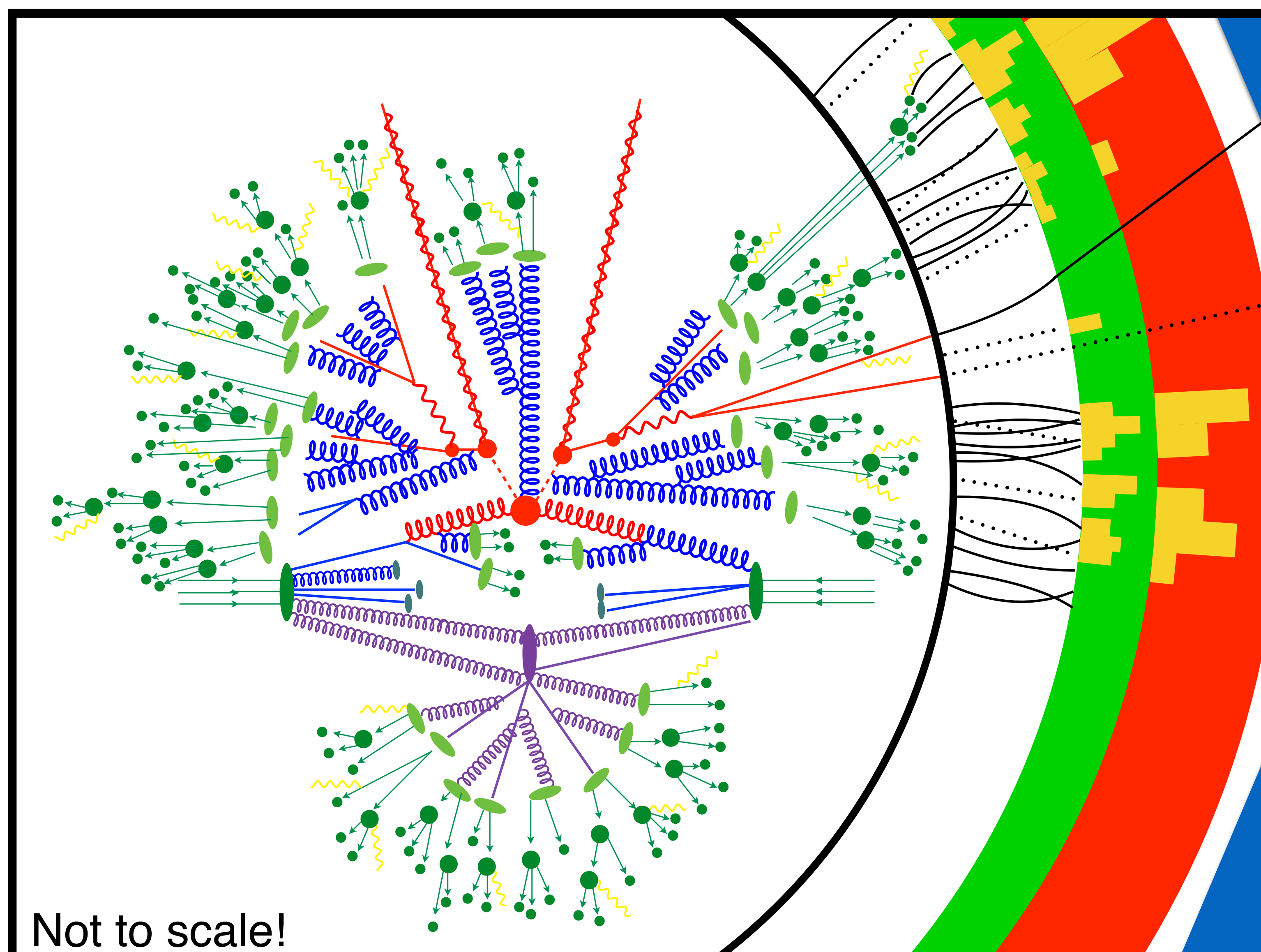
A *hyper* challenge

Key **challenge** and **opportunity**: *hypervariate phase space* & *hyper spectral data*

Typical collision events at the LHC produce **O(1000+)** particles

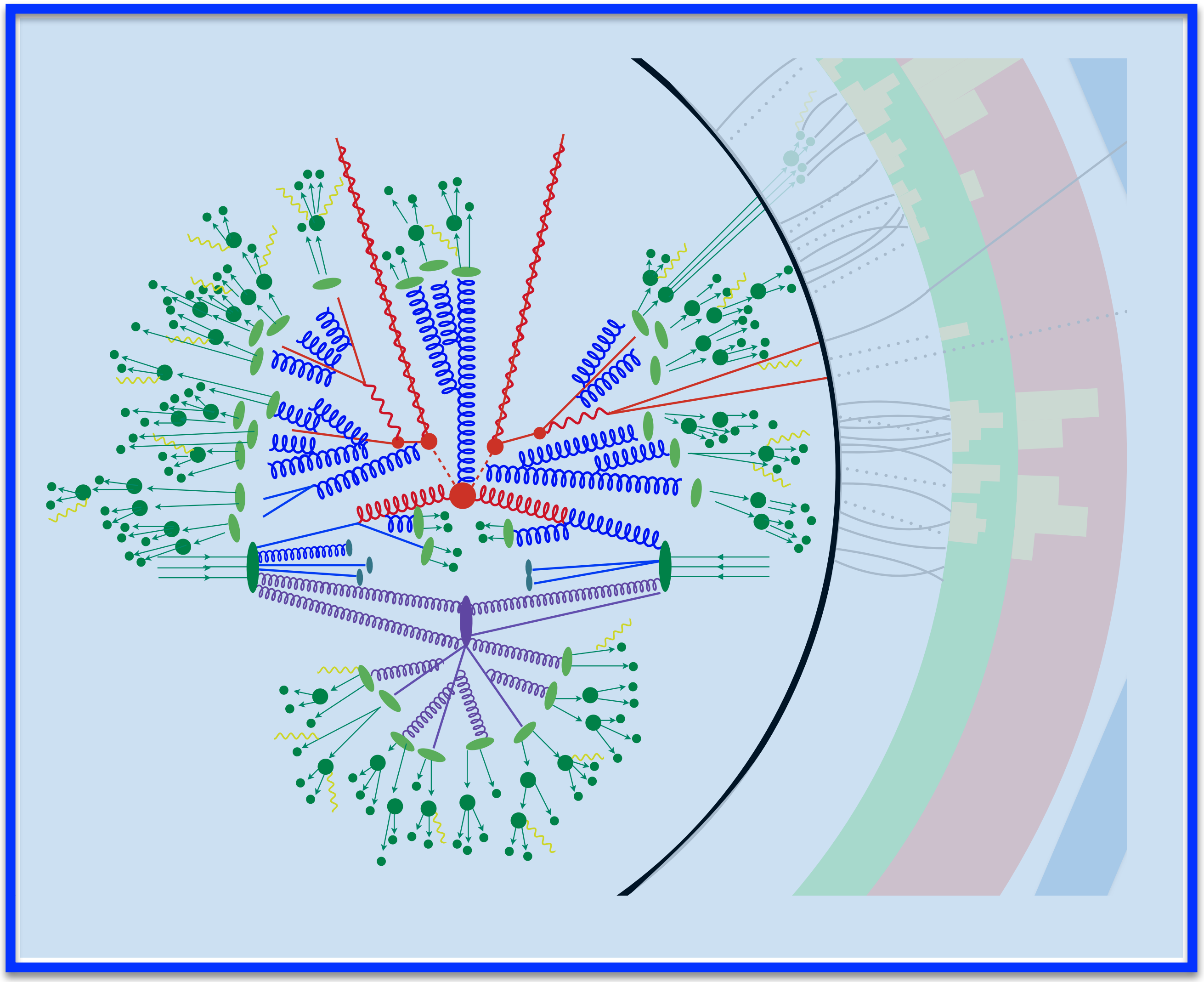
We detect these particles with **O(100 M)** readout channels

Image inspired by JHEP 02 (2009) 007

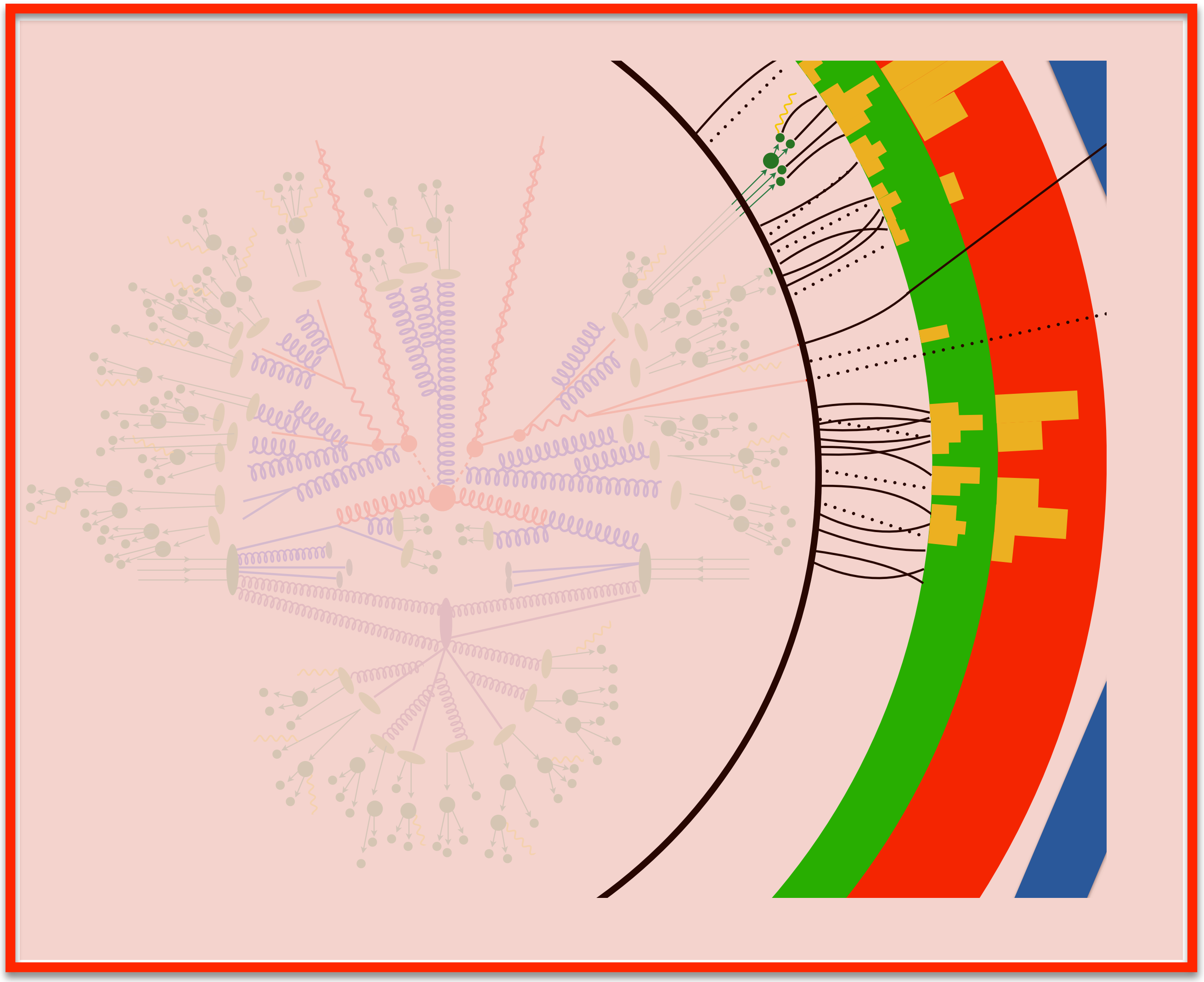


Example: Unfolding

Want this



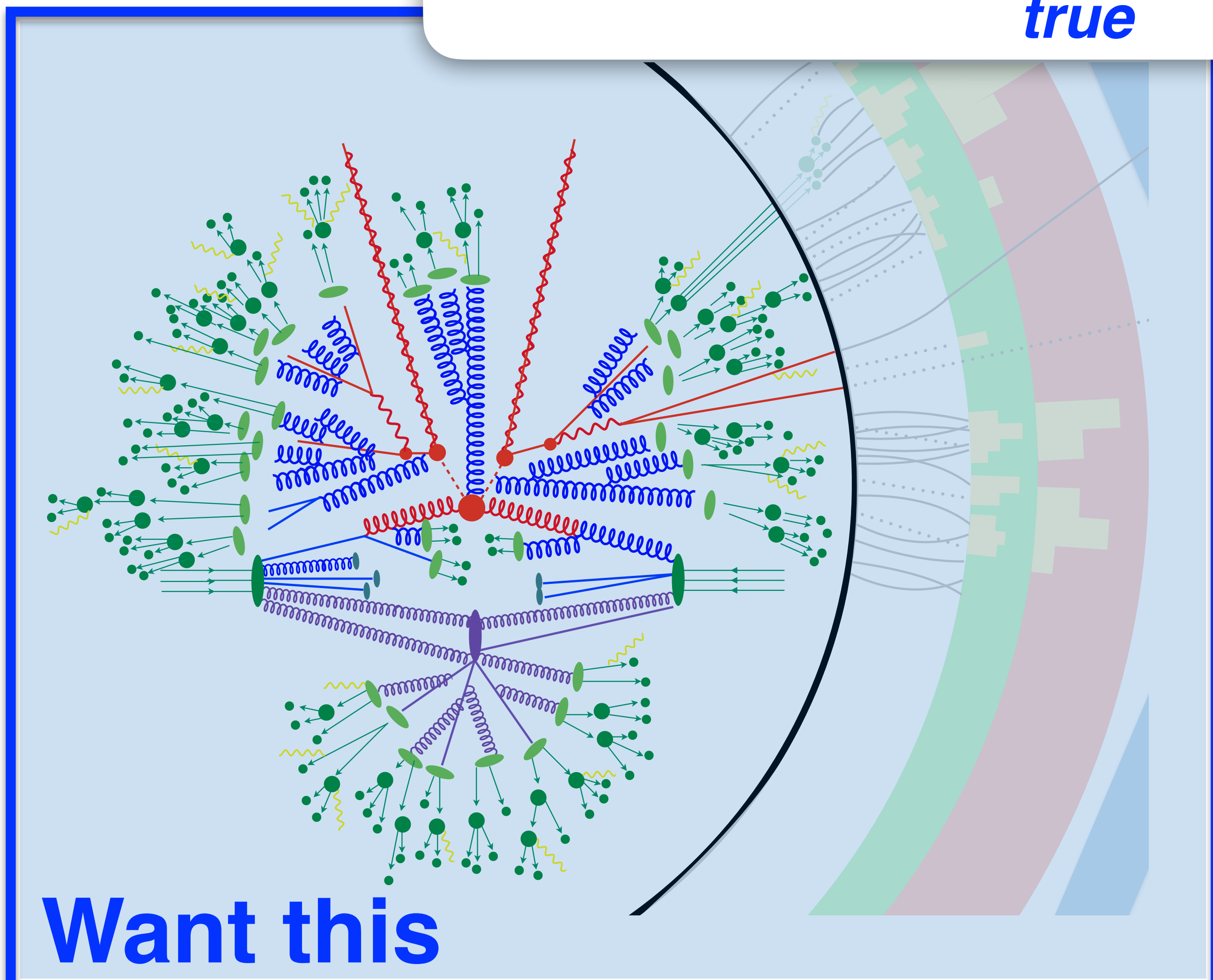
Measure this

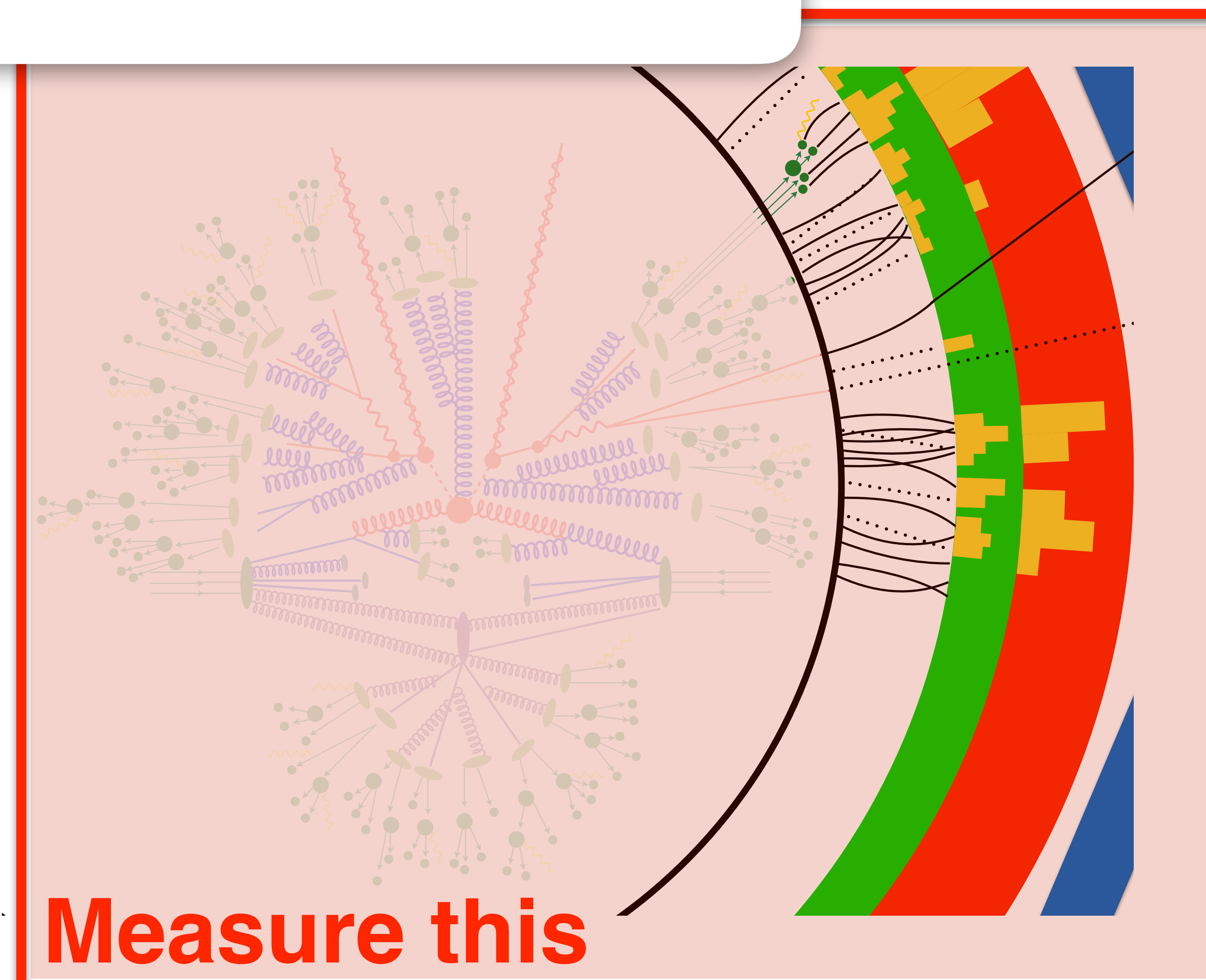


Example: Unfolding

If you know *p(**meas.** / **true**)*, could do maximum likelihood, i.e.

$$\mathit{unfolded} = \underset{\mathbf{true}}{\operatorname{argmax}}\; p(\mathbf{measured} \,/\, \mathbf{true})$$





*p(**meas.** / **true**)* = “response matrix”

Example: Unfolding

If you know *p(**meas.** / **true**)*, could do maximum likelihood, i.e.

$$unfolded = \underset{\mathbf{true}}{\operatorname{argmax}}\; p(\mathbf{measured} \mid \mathbf{true})$$

Challenge: **measured** is hyperspectral and **true** is hypervariate … *p(**meas.** | **true**) is **intractable** !*

*p(**meas.** / **true**)* = "response matrix"

Example: Unfolding

If you know *p(**meas.** | **true**)*, could do maximum likelihood, i.e.

$$\textit{unfolded} = \underset{\textit{true}}{\operatorname{argmax}}\ p(\textit{measured} \mid \textit{true})$$

Challenge: **measured** is hyperspectral and **true** is hypervariate … *p(**meas.** | **true**) is **intractable** !*

However: we have **simulators** that we can use to sample from *p(**meas.** | **true**)*

→ **Simulation-based (likelihood-free) inference**

*p(**meas.** | **true**)* = "response matrix"

Reweighting

One solution is based on ***reweighting***

dataset 1: sampled from ***p(x)***
dataset 2: sampled from ***q(x)***

Create weights ***w(x) = q(x)/p(x)*** so that when dataset 1 is weighted by ***w***, it is statistically identical to dataset 2.

Reweighting

One solution is based on ***reweighting***

dataset 1: sampled from ***p(x)***
dataset 2: sampled from ***q(x)***

Create weights ***w(x) = q(x)/p(x)*** so that when dataset 1 is weighted by ***w***, it is statistically identical to dataset 2.

What if we don't (and can't easily) know ***q*** and ***p***?

Classification for reweighting

Fact: Neutral networks learn to approximate the likelihood ratio $= q(x)/p(x)$

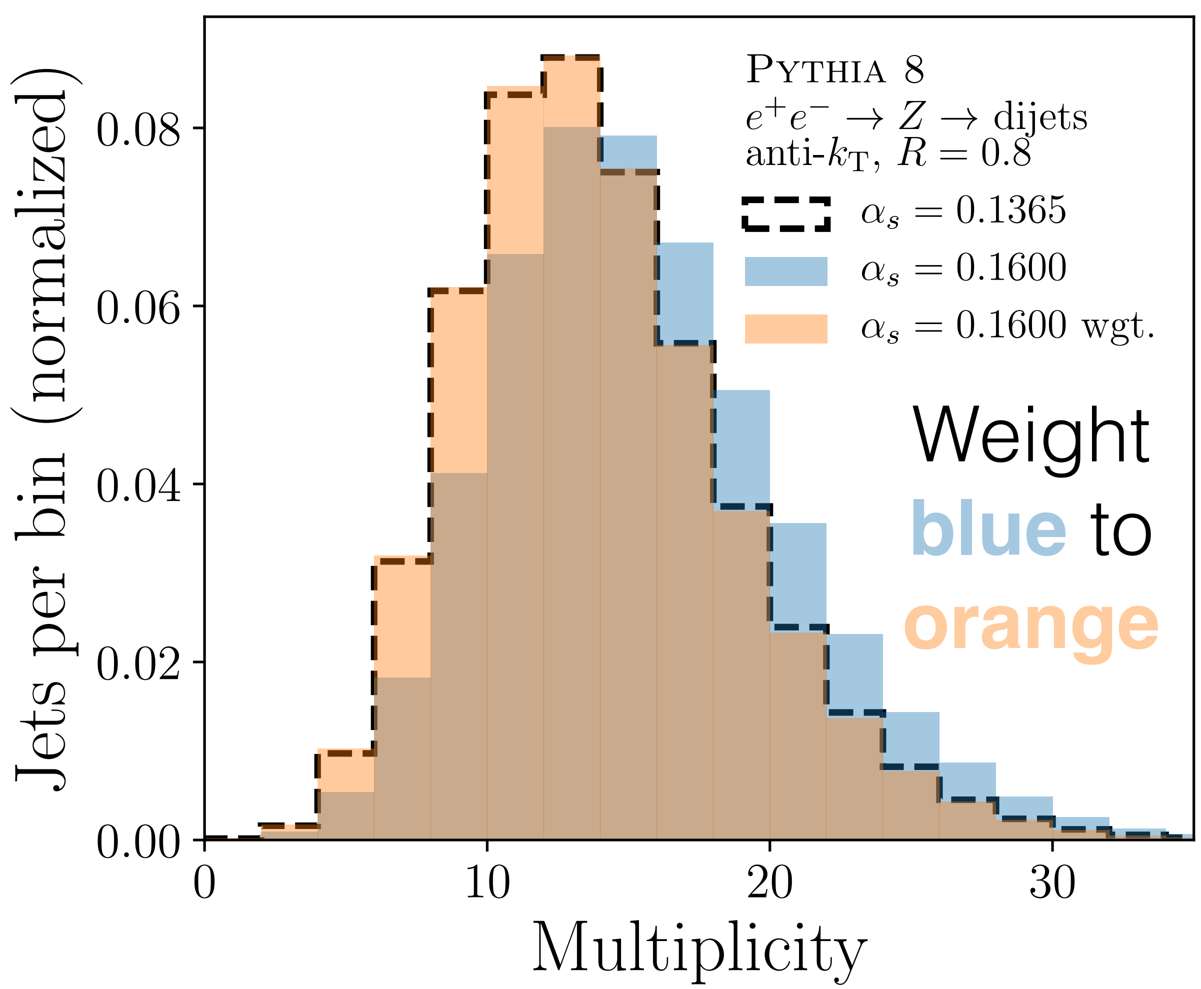
(or something monotonically related to it in a known way)

Solution: train a neural network to distinguish the two datasets!

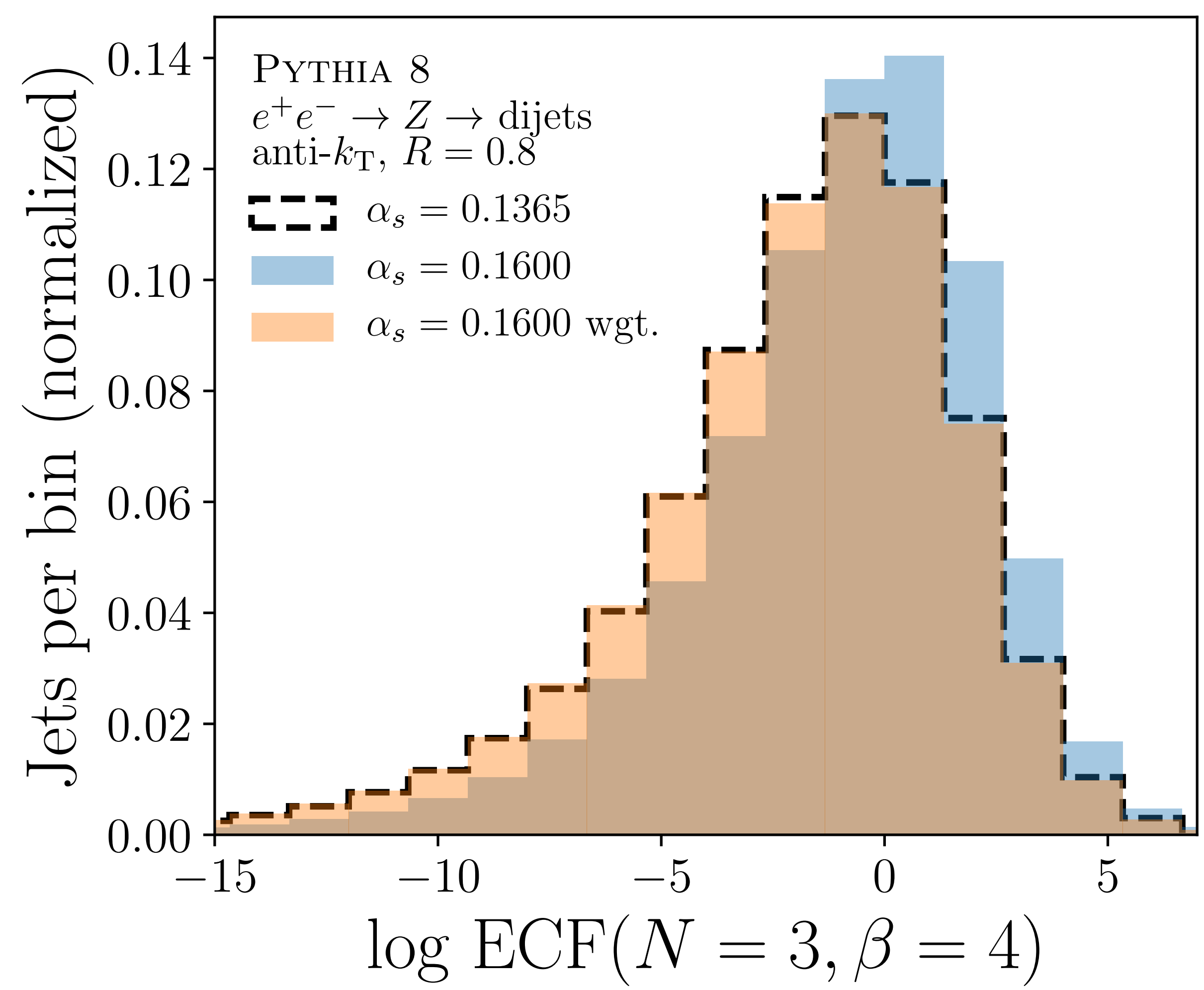
This turns the problem of **density estimation** (**hard**) into a problem of **classification** (**easy**)

Example: electron-positron collisions

Learn a classifier on the full observable phase space (momenta + particle flavor) and then check with some standard 1D observables.



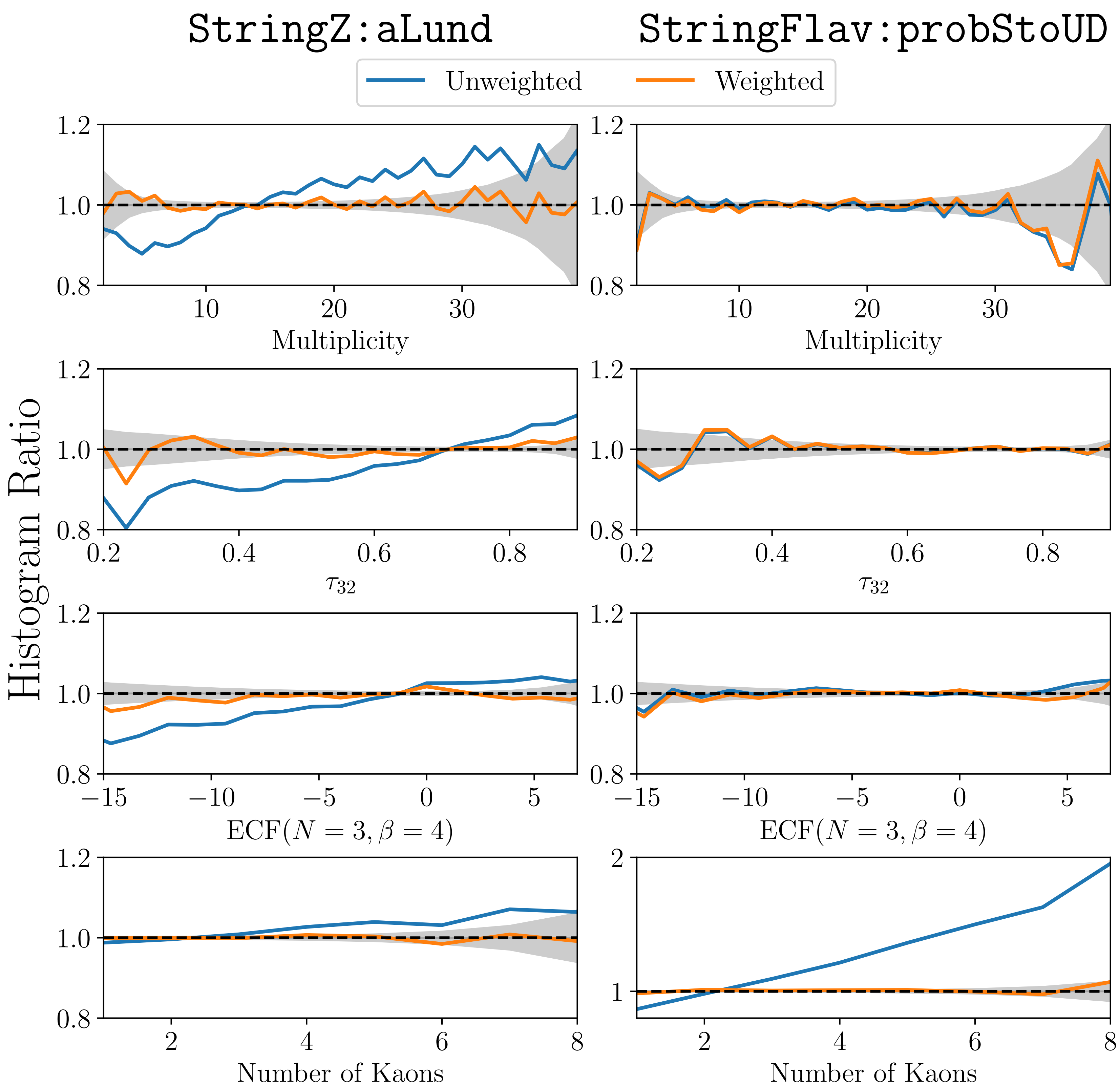
(# of particles)



(3-particle correlation function)

A. Andreassen, BPN
PRD RC 101 (2020) 091901

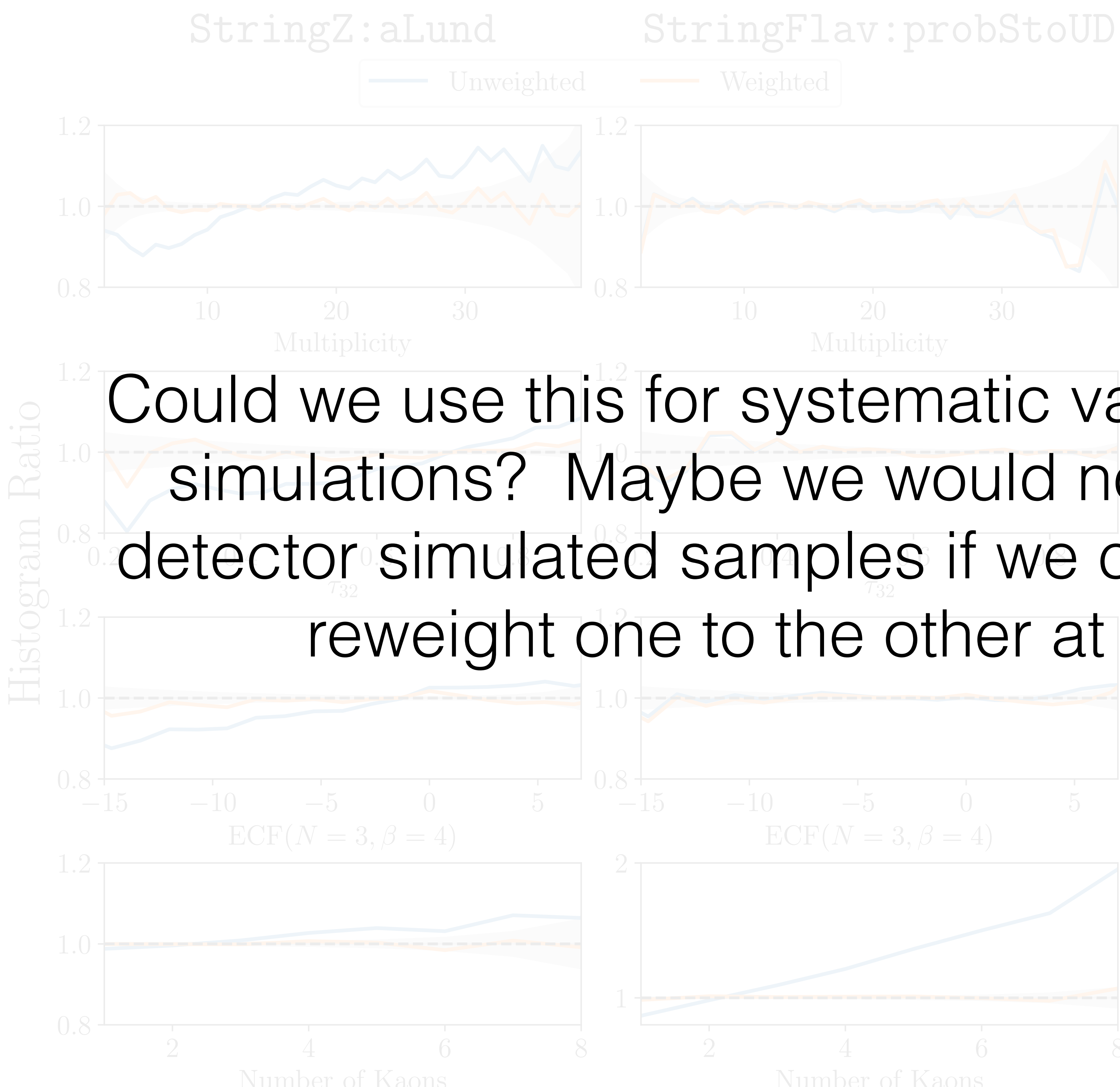
Achieving precision



Works also when the differences between the two simulations are small (left) or localized (right).

These are histogram ratios for a series of one-dimensional observables

Achieving precision

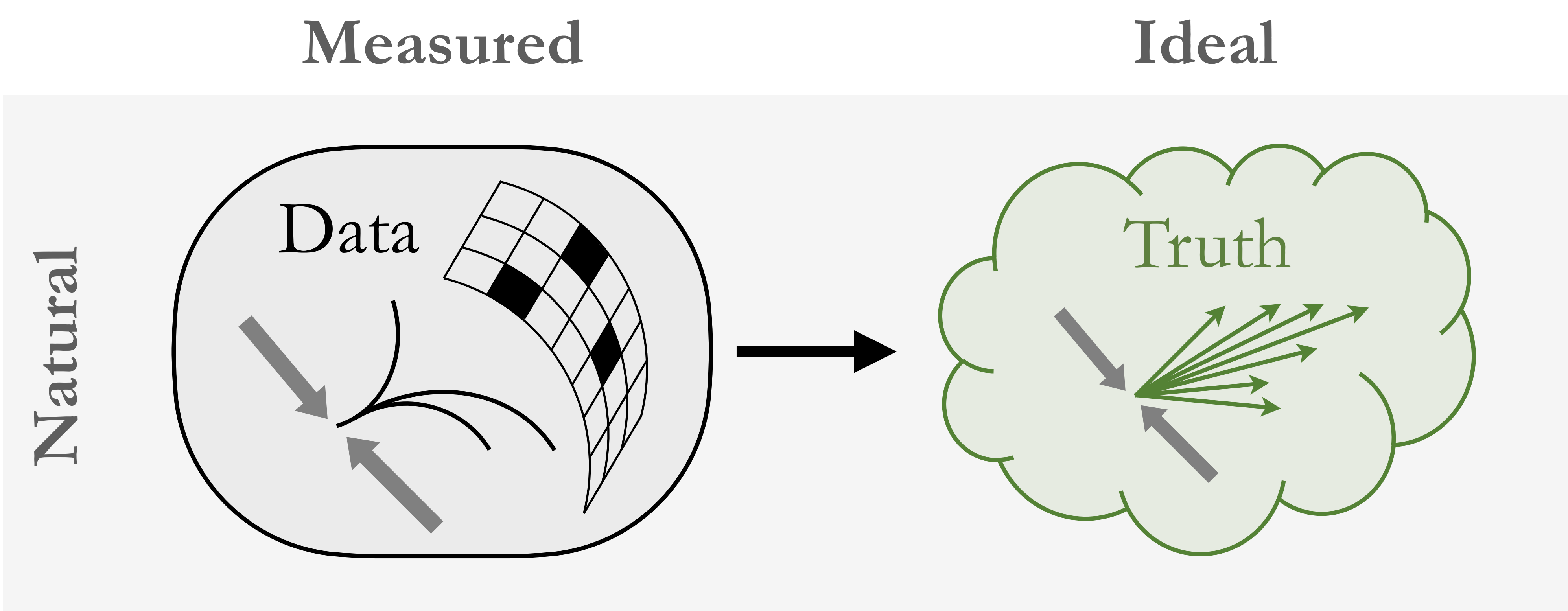


Works also when the differences between the two simulations are small (left) or localized (right).

Could we use this for systematic variations in top quark simulations? Maybe we would not need N different detector simulated samples if we can full phase space reweight one to the other at particle-level.

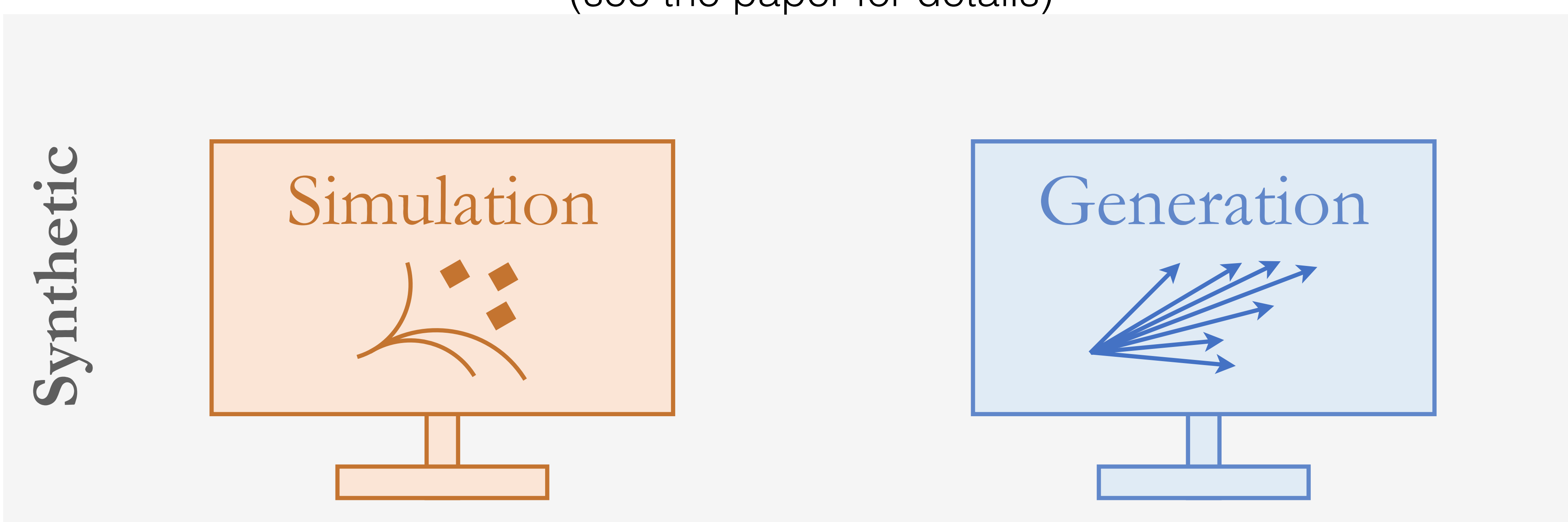
These are histogram ratios for a series of one-dimensional observables

Unfold by iterating: OmniFold



One OmniFold, compute any observable.

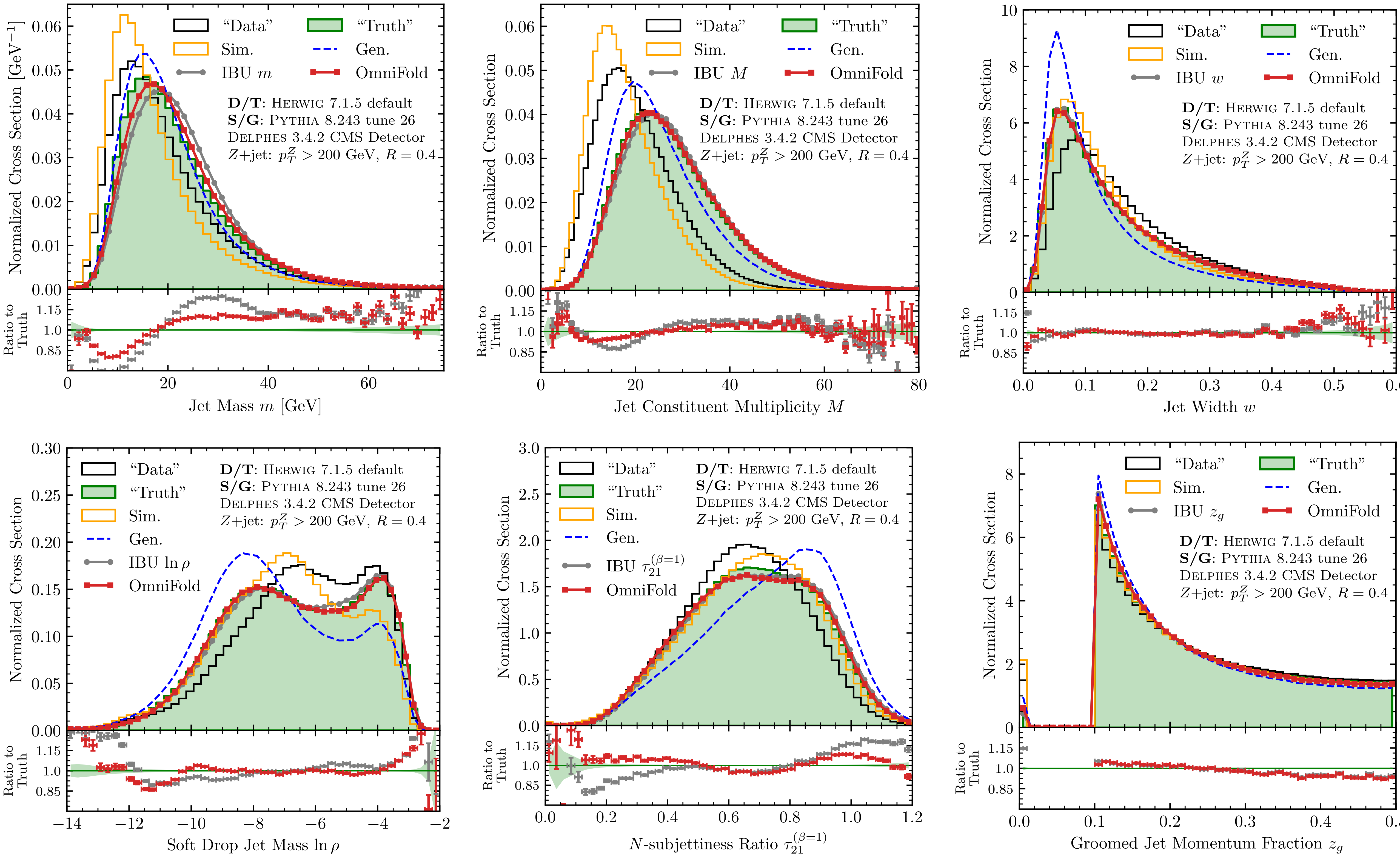
(see the paper for details)



A. Andreassen, P. Komiske, E. Metodiev, BPN, J. Thaler, PRL 124 (2020) 182001

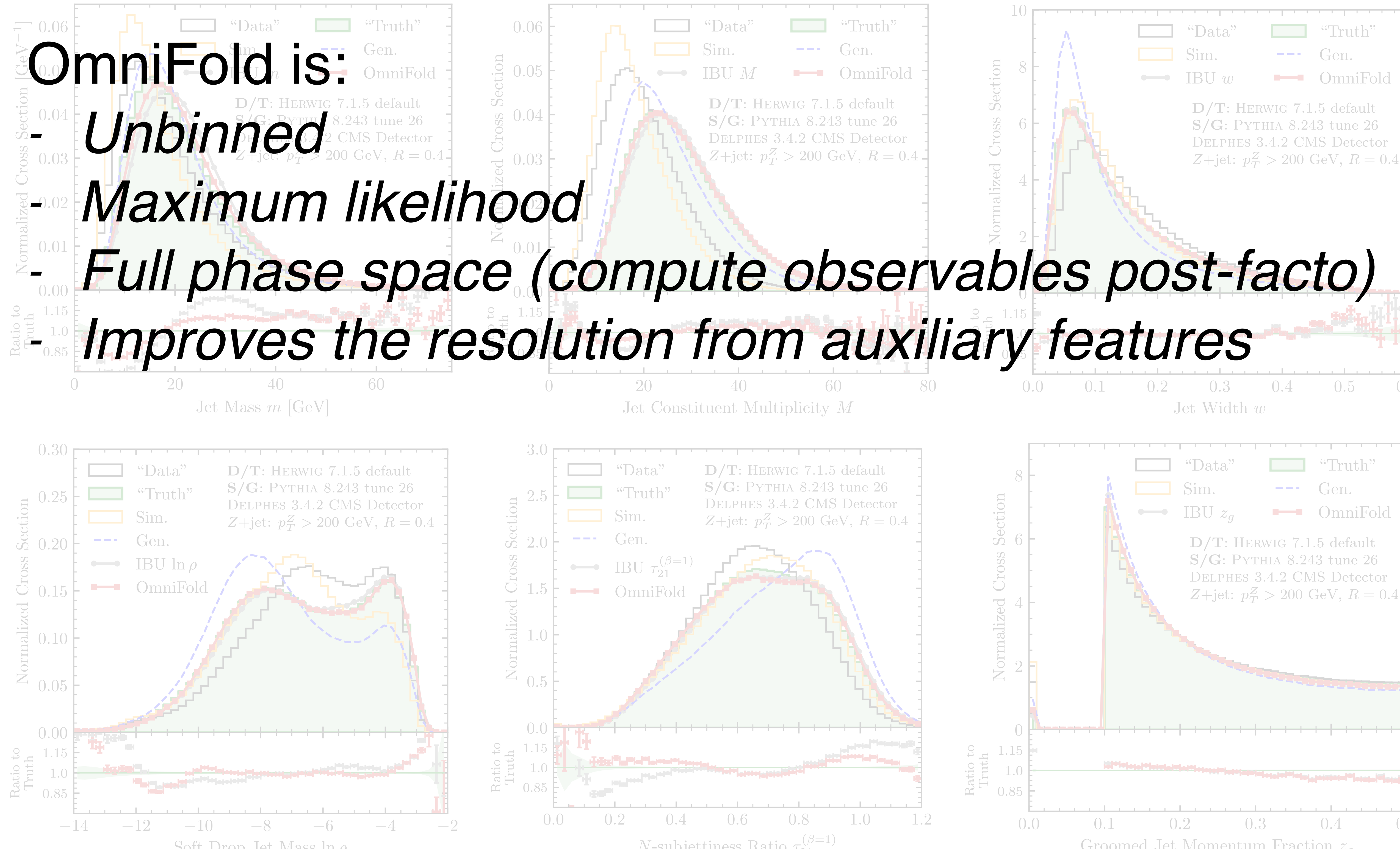
Iterative reweighting

A. Andreassen, P. Komiske, E. Metodiev, BPN, J. Thaler, PRL 124 (2020) 182001



Results

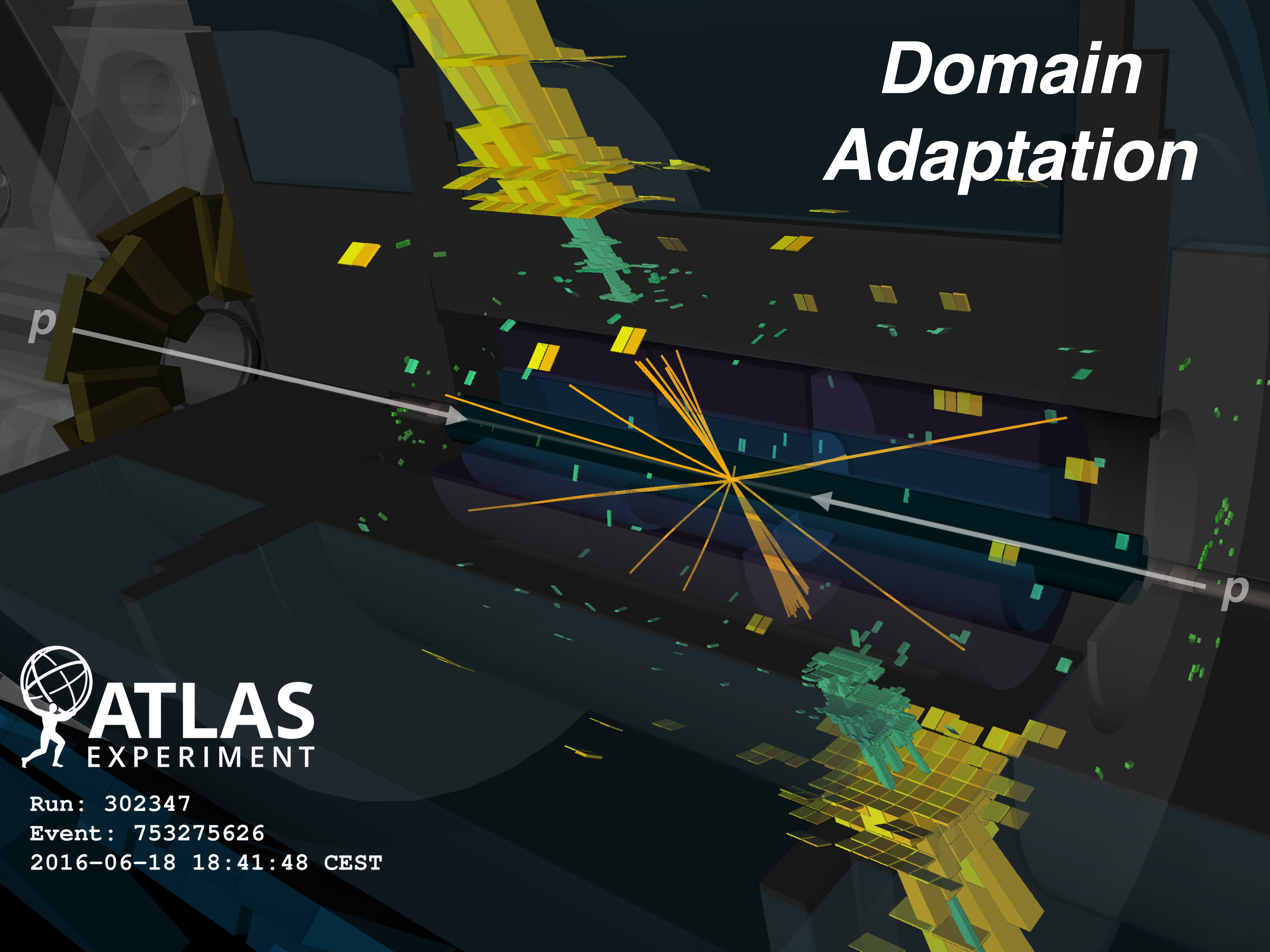
[A. Andreassen, P. Komiske, E. Metodiev, BPN, J. Thaler, 1907.08209]



OmniFold is:

- *Unbinned*
- *Maximum likelihood*
- *Full phase space (compute observables post-facto)*
- *Improves the resolution from auxiliary features*

Domain Adaptation



How to get around data/MC diffs?

A variety of recent proposals to cope with data/MC differences. This is related to anomaly detection and weak supervision, but the goal here is not to be completely model independent.

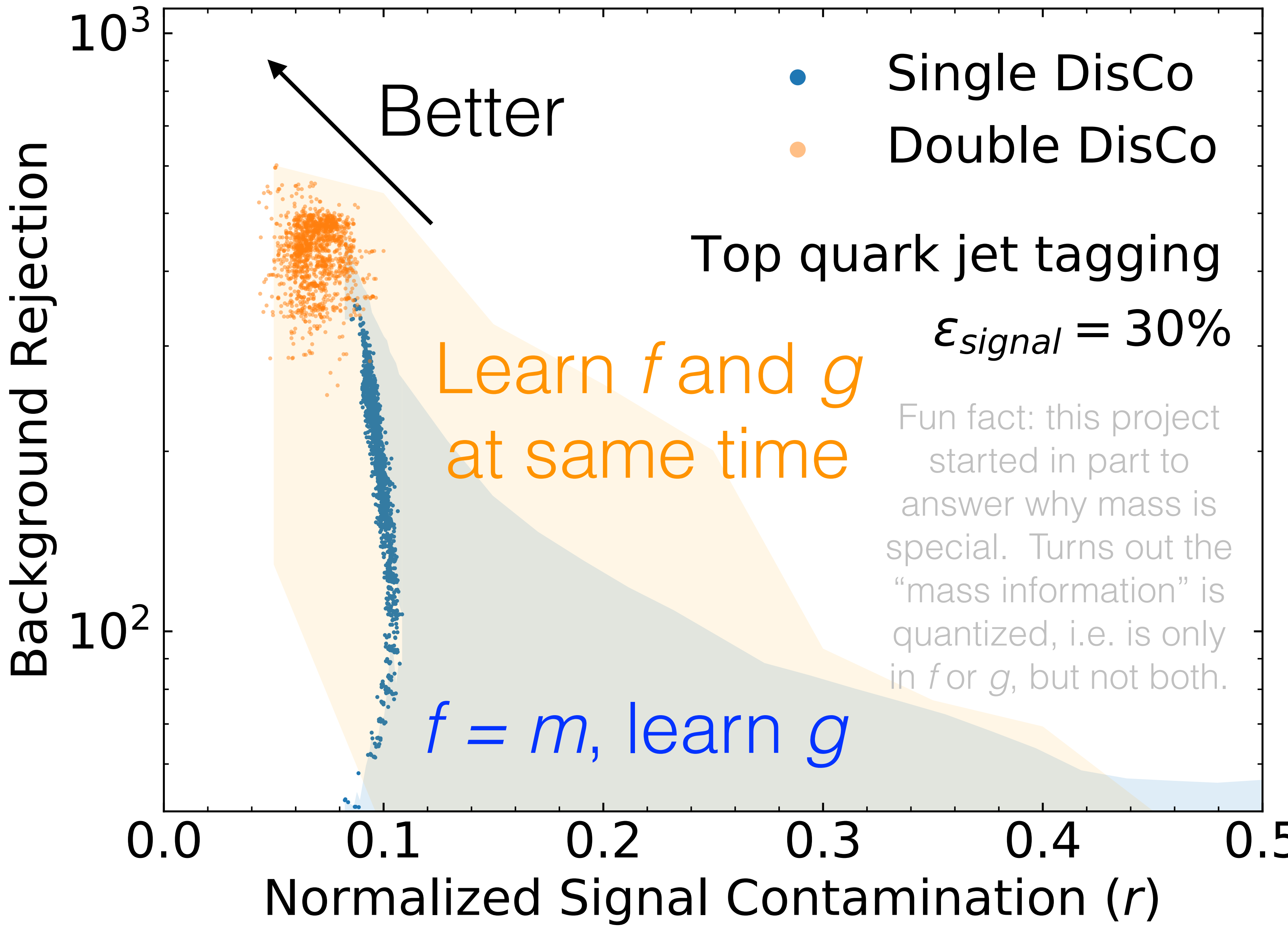
Example 1: Automated ABCD Method

One example: a common strategy for background estimation is the ABCD method. Is there a way to pick the two defining features automatically instead of by hand?

We need the two features to be **good at distinguishing** signal (e.g. top events) from background, but also they need to be **independent** of each other.

Example 1: Automated ABCD Method

$$\mathcal{L}[f,g] = \mathcal{L}_{\text{classifier}}[f(X),y] + \mathcal{L}_{\text{classifier}}[g(X),y] + \lambda\,\text{dCorr}^2_{y=0}[f(X),g(X)]$$

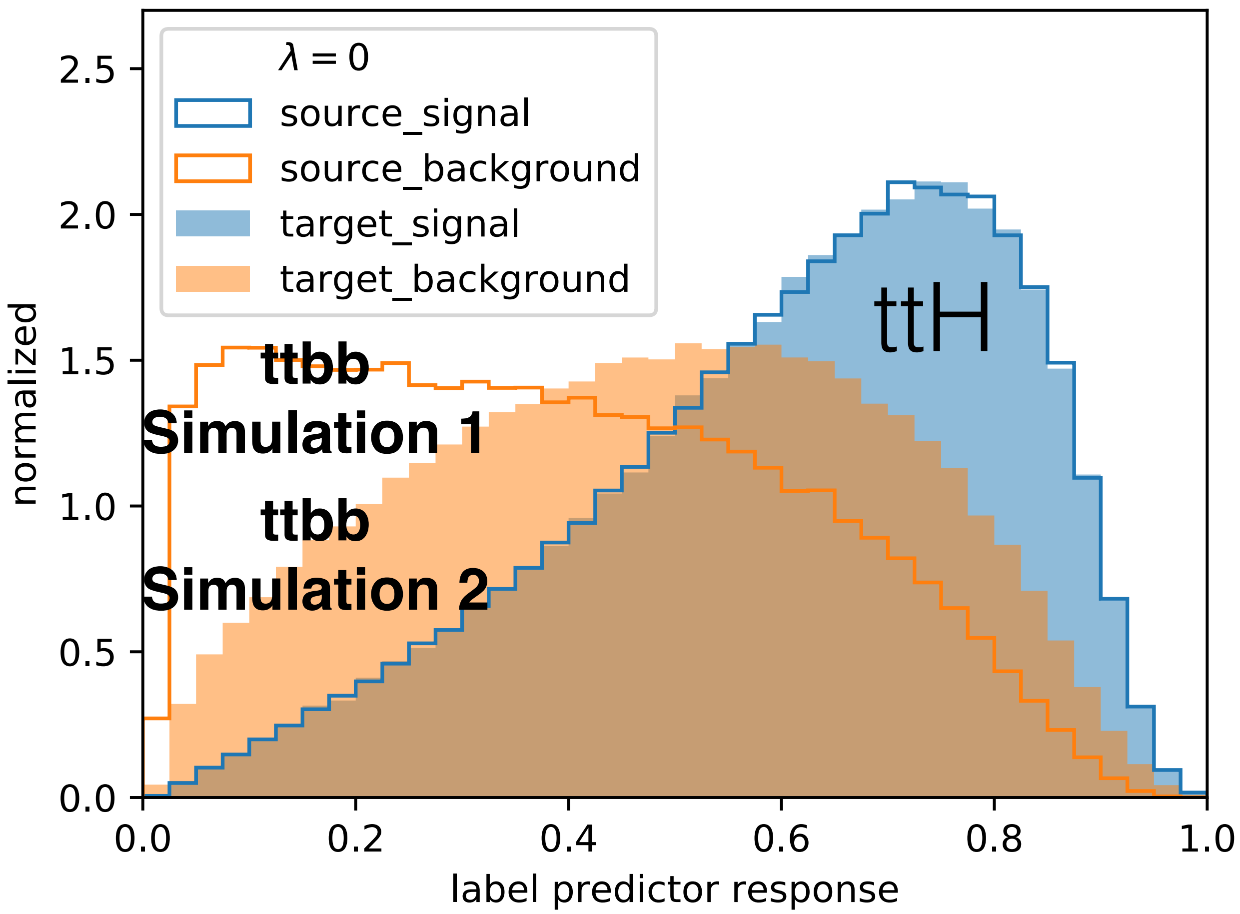


(contamination relative to signal region, not absolute!)

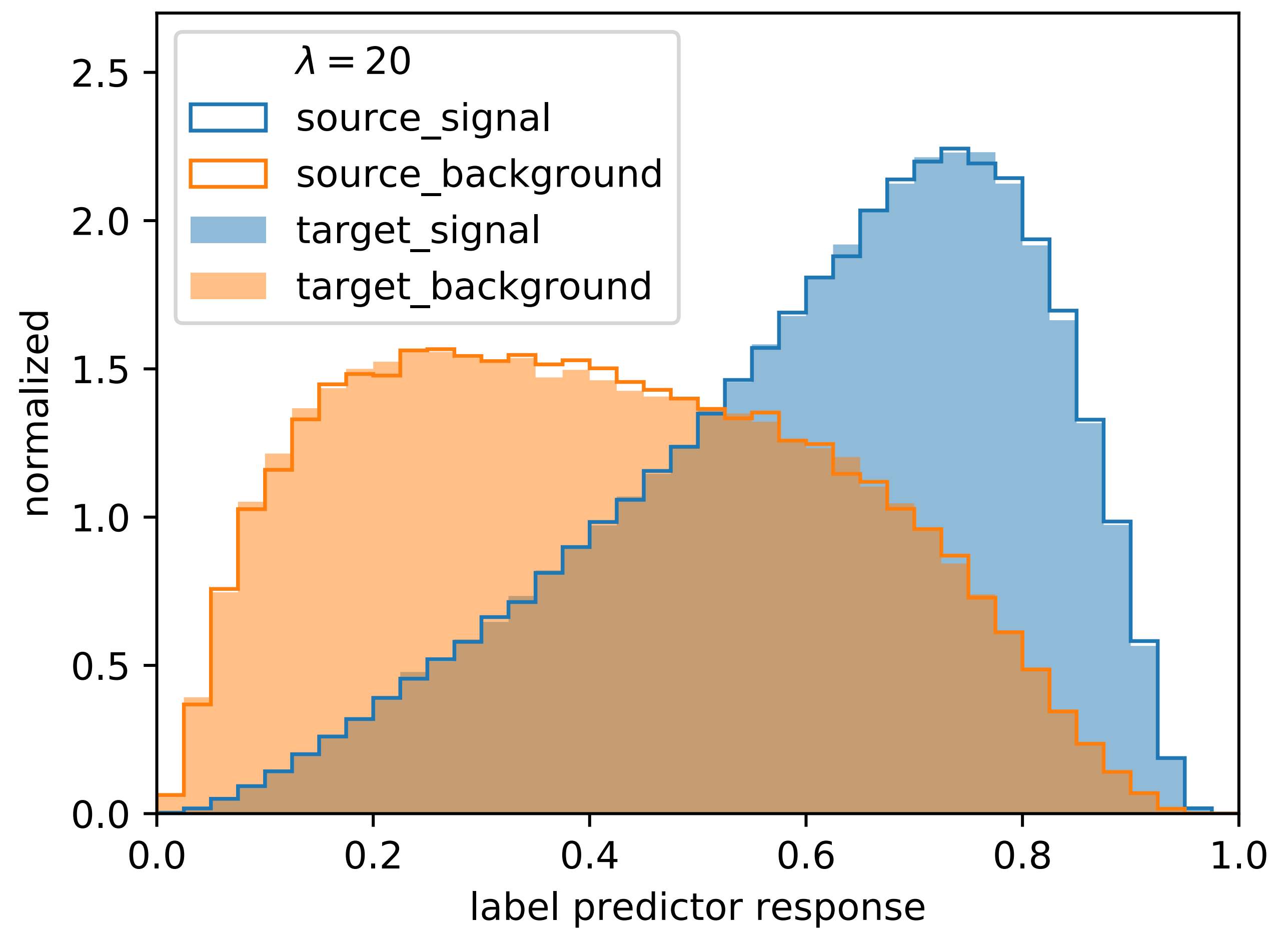
G. Kasieczka et al., 2007.14400

Example 2: penalize learning data/MC

J. Clavijo, P. Glaysher, J. Katzy, 2005.00568



Nominal

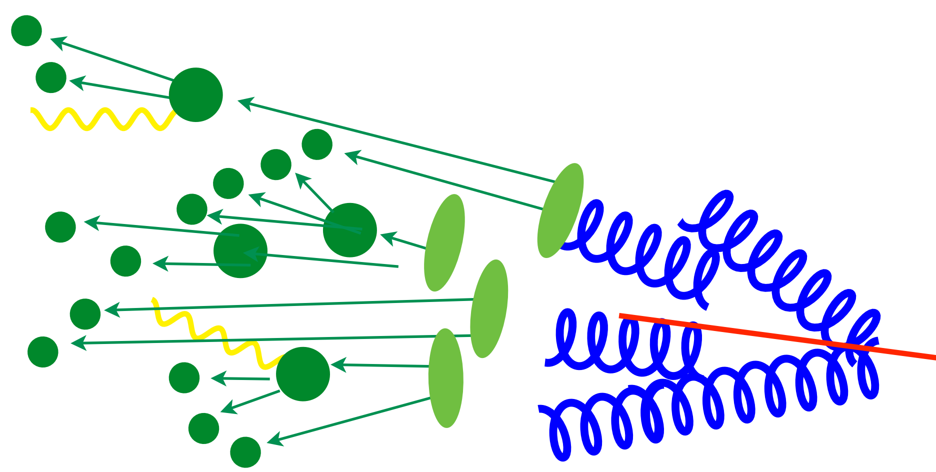


Penalty term in loss

See also CMS, 1912.12238

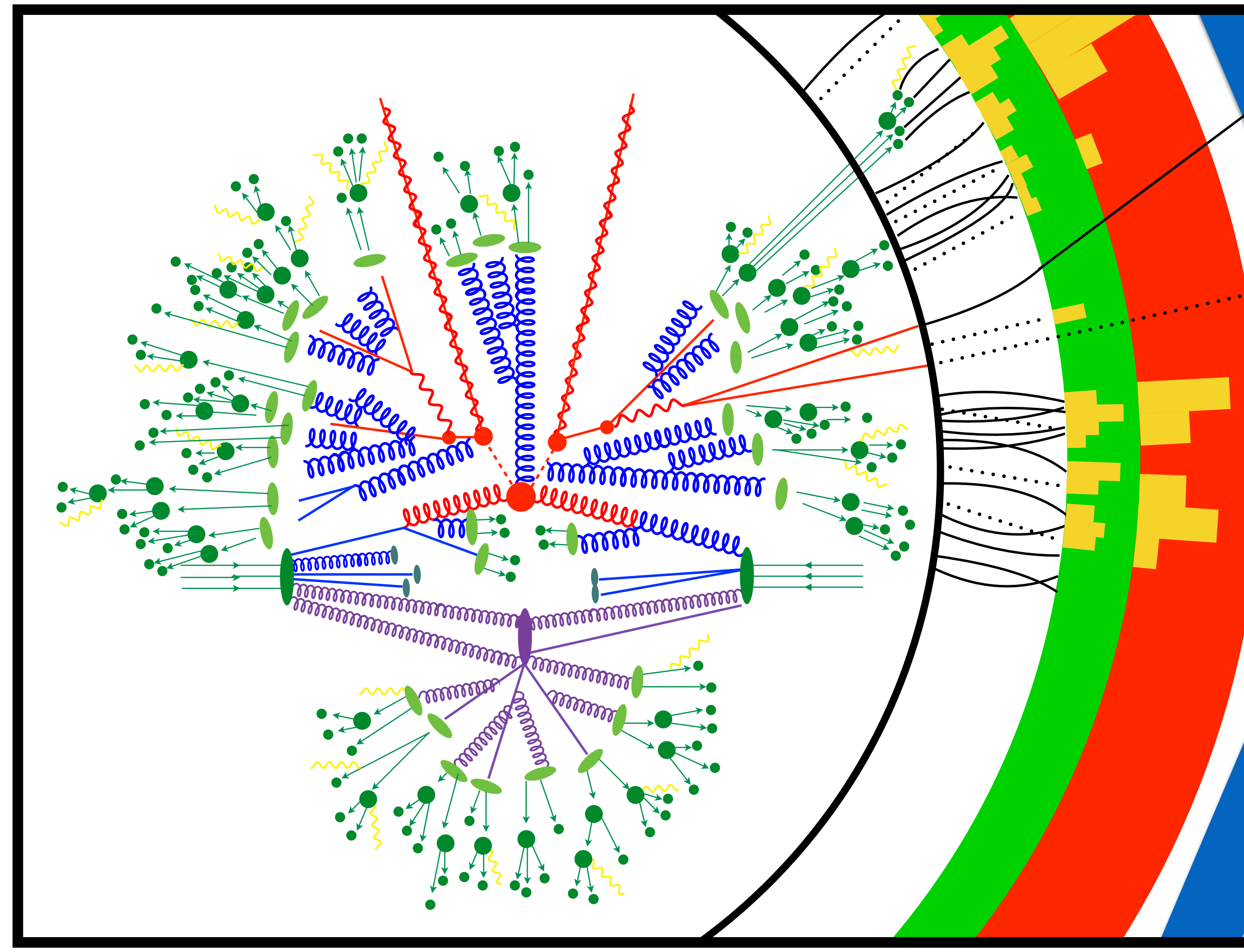
Outline

- **Boosted Top Jet Classification**
  - Jet Representations
- **Beyond Standard Classification**
  - Less-than-supervised
    - Generation (more in Anja's talk)
    - Weak supervision
    - Anomaly detection (more in Jernej and Bryan's talks)
  - Parameter estimation / unfolding
  - Domain adaptation



Conclusions and outlook

Deep learning has a great potential to **enhance**, **accelerate**, and **empower** top physics.



Disclaimer: I have given you a biased perspective of new developments!

There is still work to do on all fronts to consider top quark events holistically in their natural high dimensionality!

Interested in learning more?

HEPML-LivingReview

A Living Review of Machine Learning for Particle Physics

Modern machine learning techniques, including deep learning, is rapidly being applied, adapted, and developed for high energy physics. The goal of this document is to provide a nearly comprehensive list of citations for those developing and applying these approaches to experimental, phenomenological, or theoretical analyses. As a living document, it will be updated as often as possible to incorporate the latest developments. A list of proper (unchanging) reviews can be found within. Papers are grouped into a small set of topics to be as useful as possible. Suggestions are most welcome.

download review

The purpose of this note is to collect references for modern machine learning as applied to particle physics. A minimal number of categories is chosen in order to be as useful as possible. Note that papers may be referenced in more than one category. The fact that a paper is listed in this document does not endorse or validate its content - that is for the community (and for peer-review) to decide. Furthermore, the classification here is a best attempt and may have flaws - please let us know if (a) we have missed a paper you think should be included, (b) a paper has been misclassified, or (c) a citation for a paper is not correct or if the journal information is now available. In order to be as useful as possible, this document will continue to evolve so please check back before you write your next paper. If you find this review helpful, please consider citing it using \cite{hepmllivingreview} in HEPML.bib.

- Reviews.
  - Modern reviews
    - Jet Substructure at the Large Hadron Collider: A Review of Recent Advances in Theory and Machine Learning [DOI]
    - Deep Learning and its Application to LHC Physics
    - Machine Learning in High Energy Physics Community White Paper
    - Machine learning at the energy and intensity frontiers of particle physics
    - Machine learning and the physical sciences [DOI]
    - Machine and Deep Learning Applications in Particle Physics [DOI]

https://iml-wg.github.io/HEPML-LivingReview/

Questions?