AI SSAI
Anomaly Detection Workshop
March 4-7, 2024
CLERMONT-FERRAND, FRANCE

ADVANCED ANOMALY DETECTION ALGORITHMS TO SEARCH FOR SEMIVISIBLE JETS IN THE CMS EXPERIMENT AT THE CERN LHC

Roberto Seidita - ETHZ

On behalf of the CMS Collaboration

ETH zürich

OUTLINE

- Anomaly detection as a tool for discovery: the case of semivisible hadronic jets
- Strengths and shortcomings of autoencoders (in short)
- How to normalize an autoencoder
- Beyond normalized autoencoders

- **Focus more on techniques than results given the occasion**

NEW PHYSICS AS AN ANOMALY

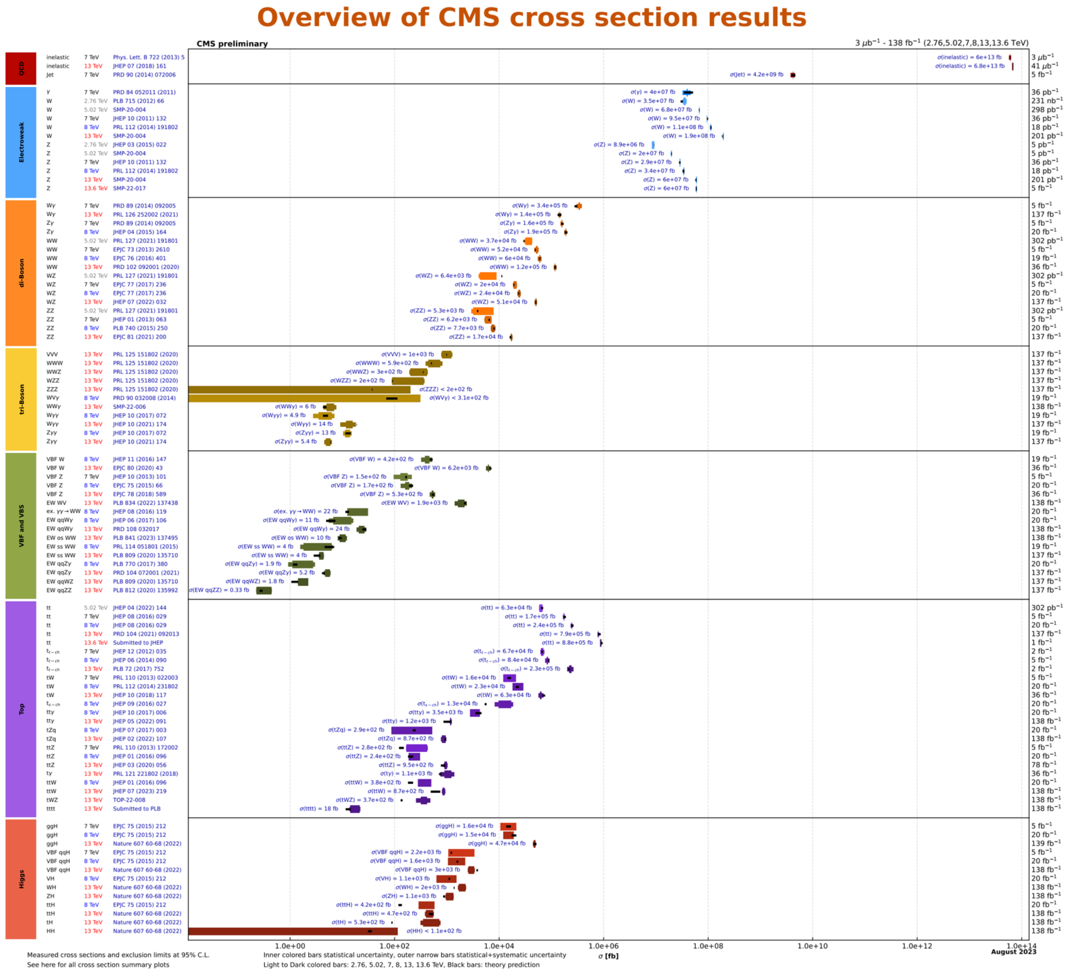
It's a unique time in HEP

Stunning agreement of measurements with the Standard Model (SM)

We know that there is more to this story

NEW PHYSICS AS AN ANOMALY

Stunning agreement of measurements with the Standard Model (SM)



NEW PHYSICS AS AN ANOMALY

Stunning agreement of measurements with the Standard Model (SM)

VBF and VBS

Higgs

Top

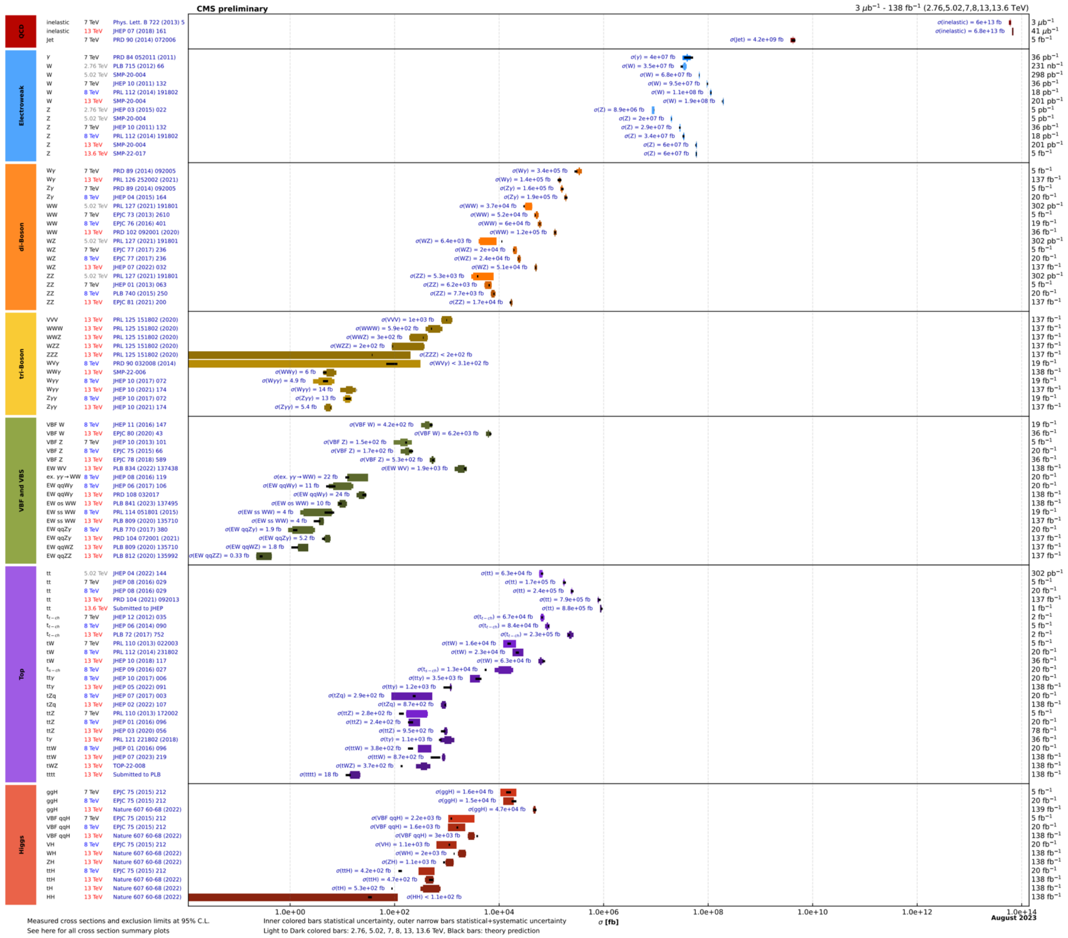
Electroweak

tri-Boson

QCD

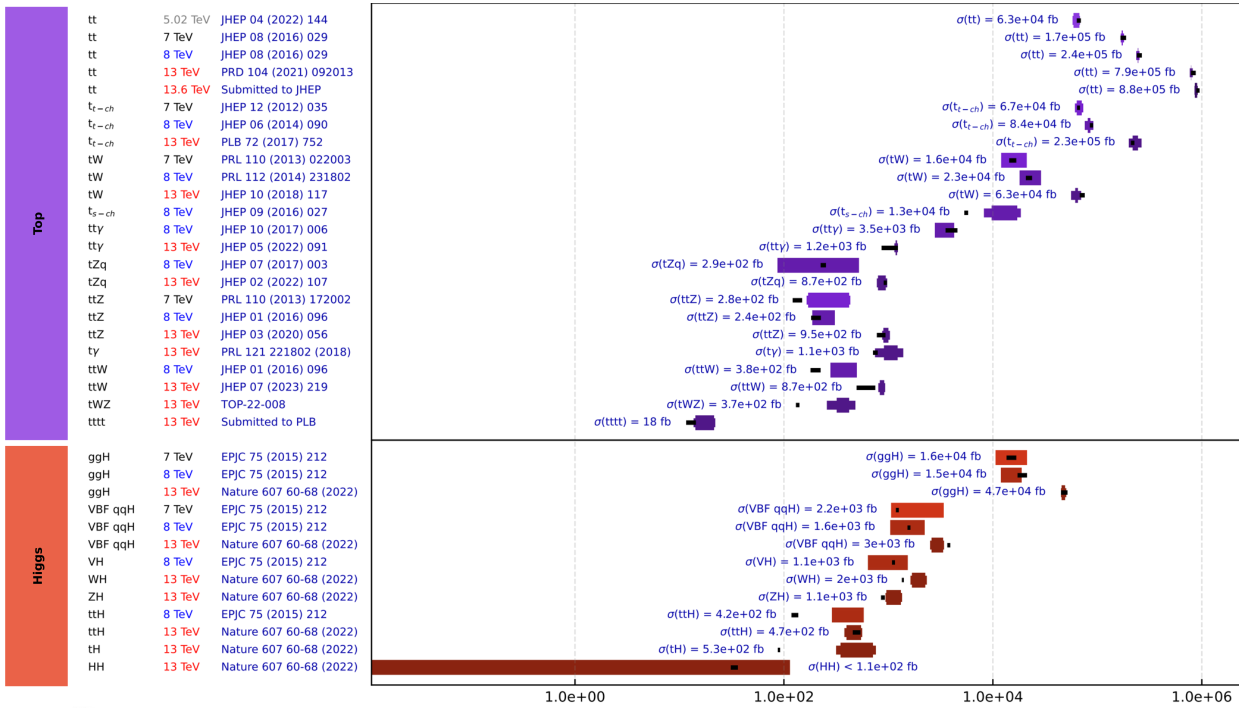
di-Boson

Overview of CMS cross section results

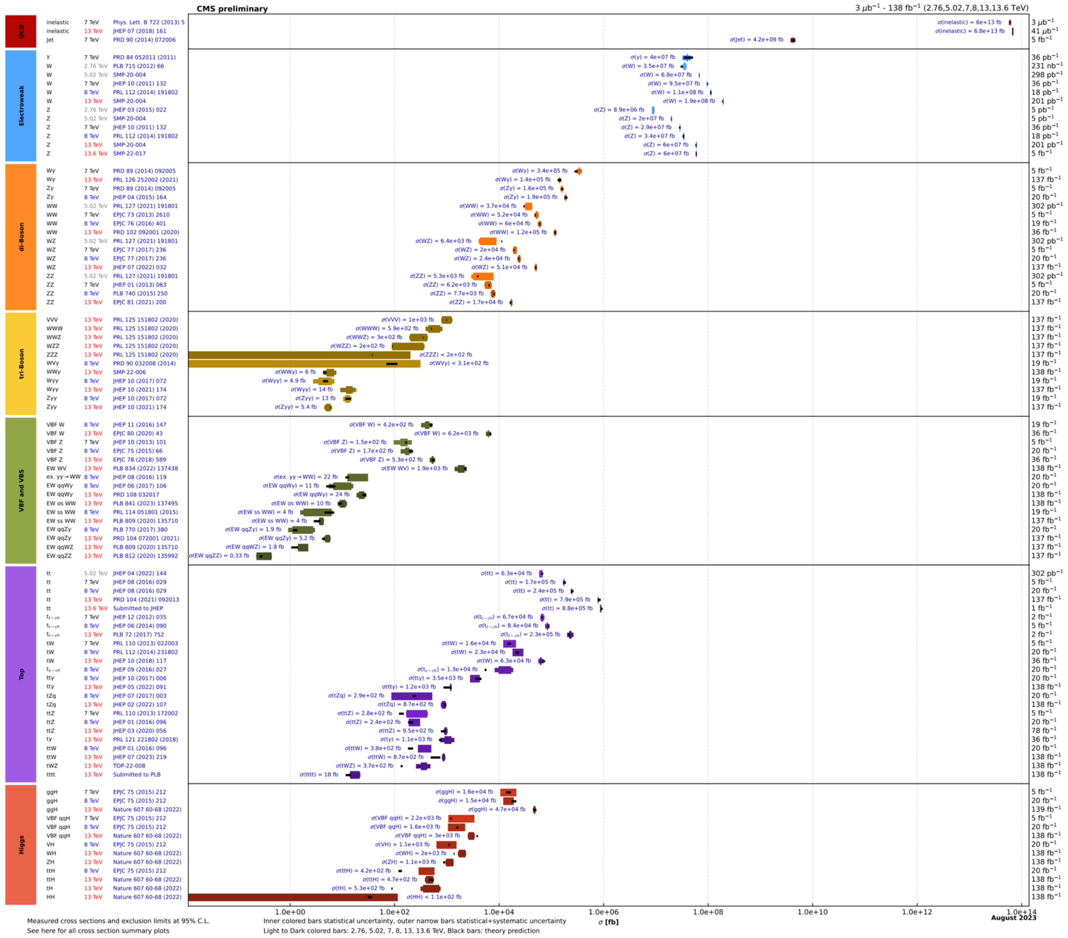


NEW PHYSICS AS AN ANOMALY

Stunning agreement of measurements with the Standard Model (SM)



Overview of CMS cross section results



NEW PHYSICS AS AN ANOMALY

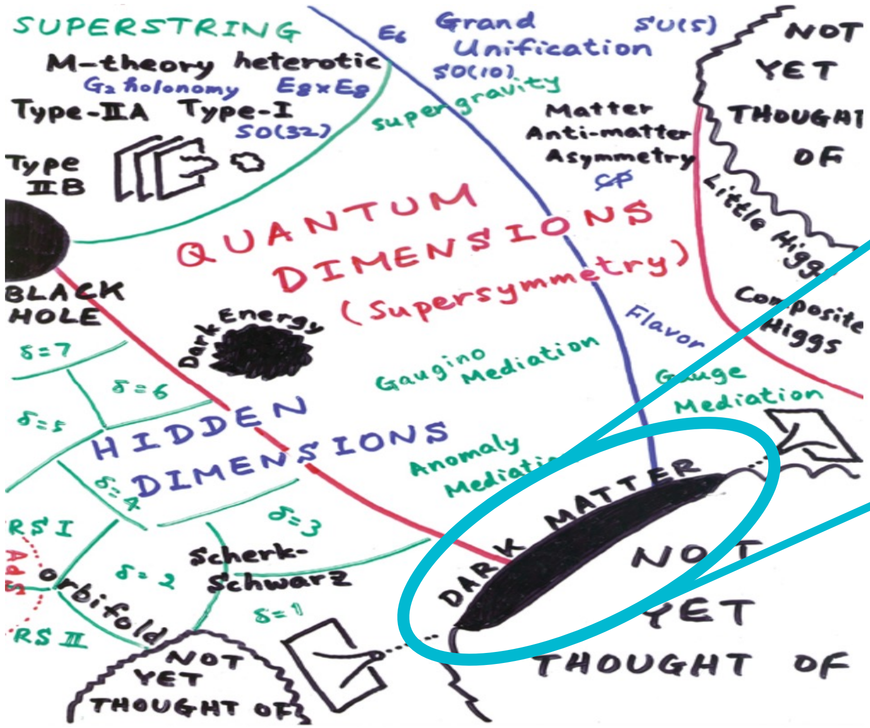
It's a unique time in HEP

Stunning agreement of measurements with the Standard Model (SM)

We know that there is more to this story

NEW PHYSICS AS AN ANOMALY

We know that there is more to this story



Picking one example: dark matter

- **We know:**
  - It's there (astrophysics/cosmology)
  - It's not in the SM
- **We don't know:**
  - Basically, anything else

What if new physics was there all along, only in an unexpected form?

NEW PHYSICS AS AN ANOMALY

It's a unique time in HEP

Stunning agreement of measurements with the Standard Model (SM)

We know that there is more to this story

We should make sure to look everywhere new physics may be hiding

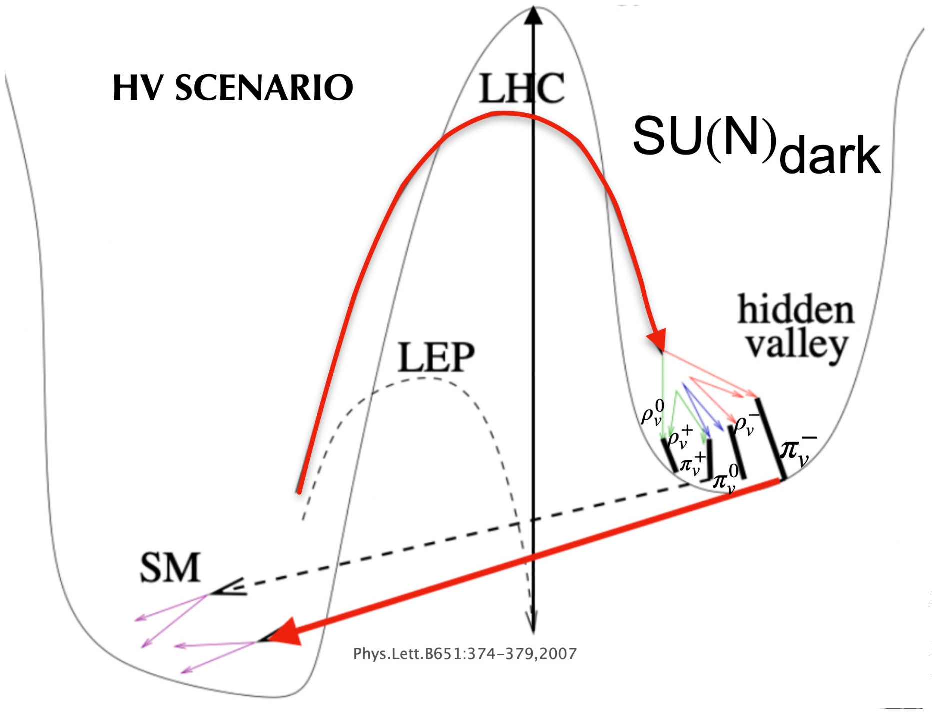
↓

Signature-based searches

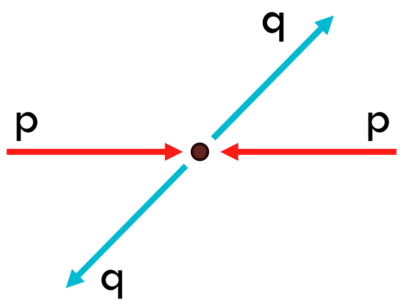
Quite general: hidden sector with at least one confining interaction

Would lead to experimental signatures unexplored as of recently

One of these is what we call a semivisible jet

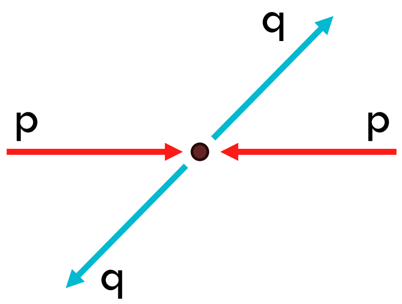


WHAT EVEN IS A JET?

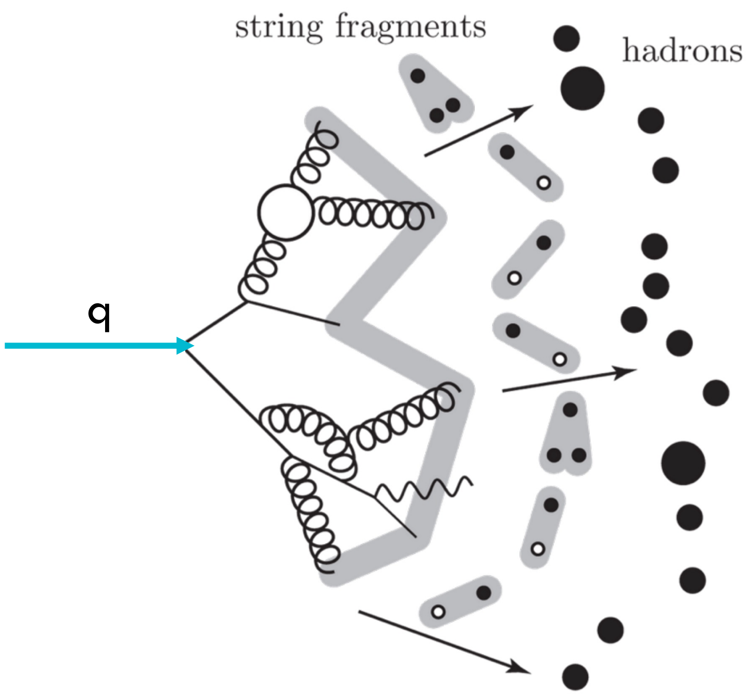


High-energy colored particles (quarks or gluons) are produced in a proton-proton collision

WHAT EVEN IS A JET?

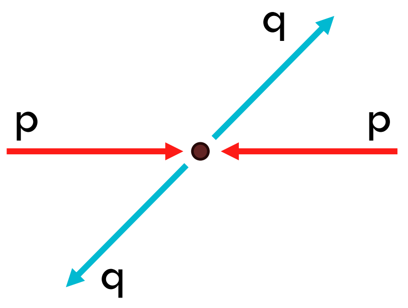


High-energy colored particles (quarks or gluons) are produced in a proton-proton collision

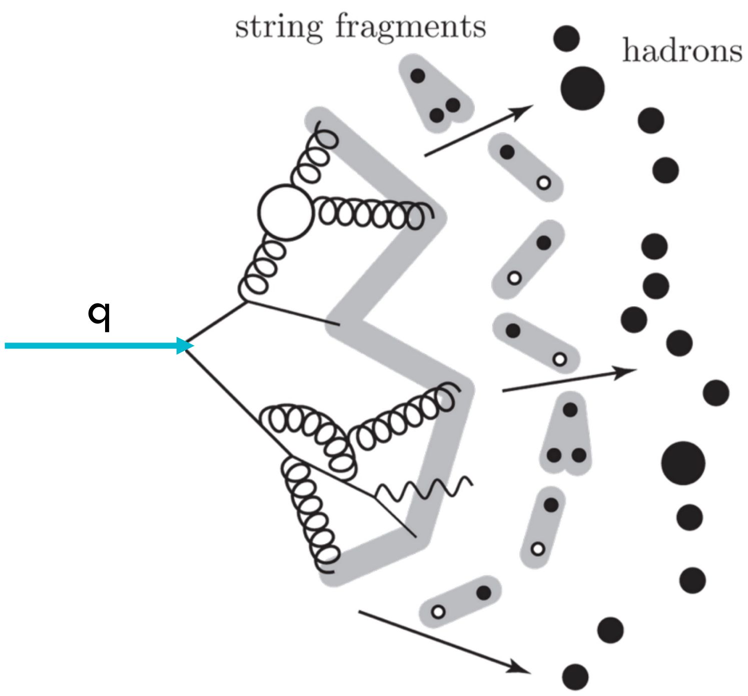


Low-energy QCD kicks in: the initial energy of the quark is split between colorless states (parton shower + hadronization)

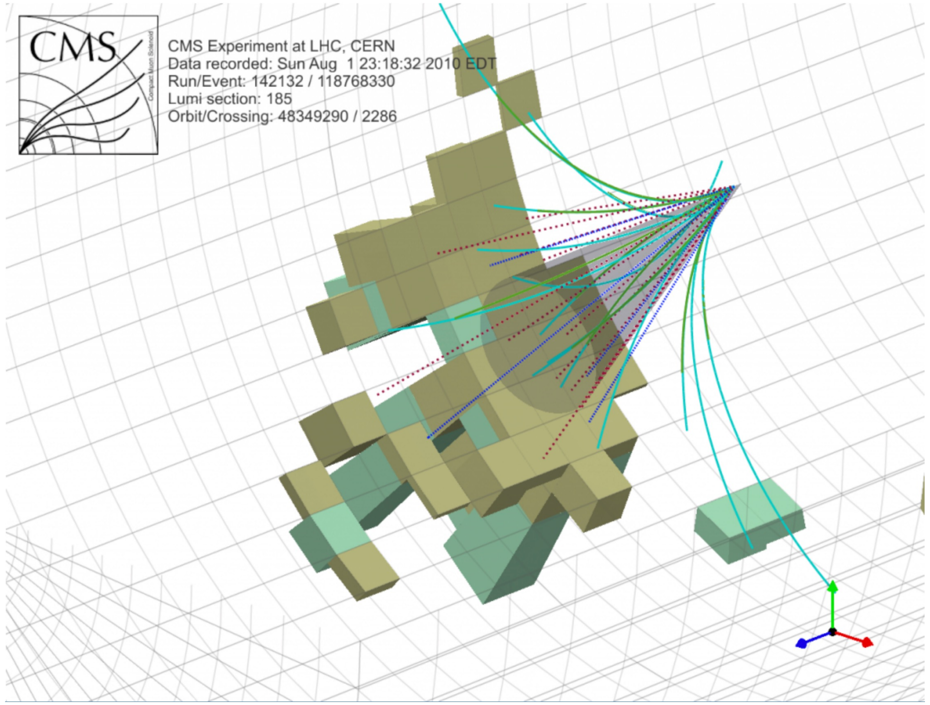
WHAT EVEN IS A JET?



High-energy colored particles (quarks or gluons) are produced in a proton-proton collision

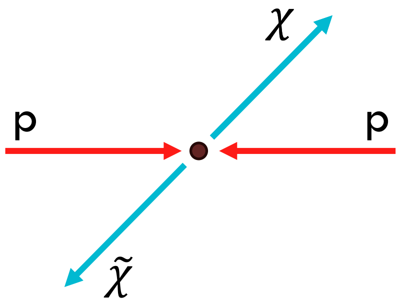


Low-energy QCD kicks in: the initial energy of the quark is split between colorless states (parton shower + hadronization)

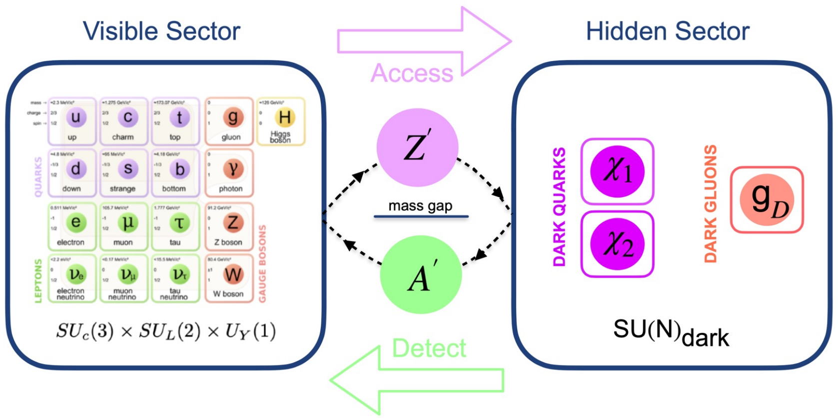


In the detector: collimated spray of hadrons, leptons, and hadrons

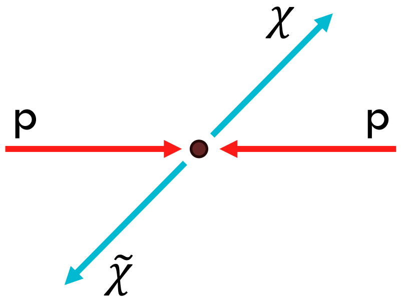
WHAT EVEN IS A SEMIVISIBLE JET??



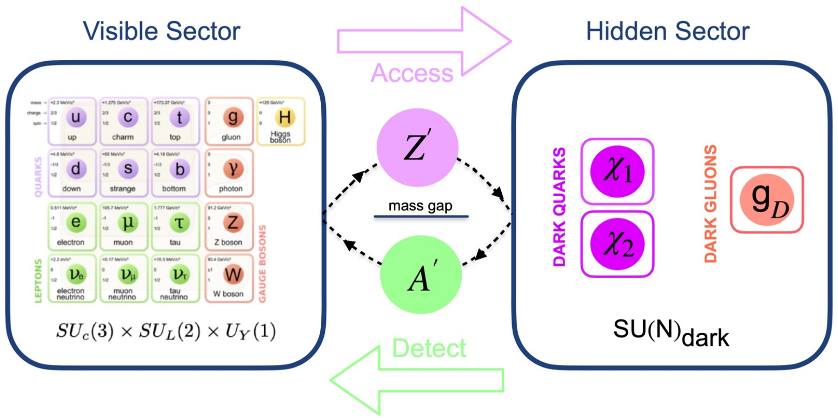
Dark quarks are produced by a proton-proton collision

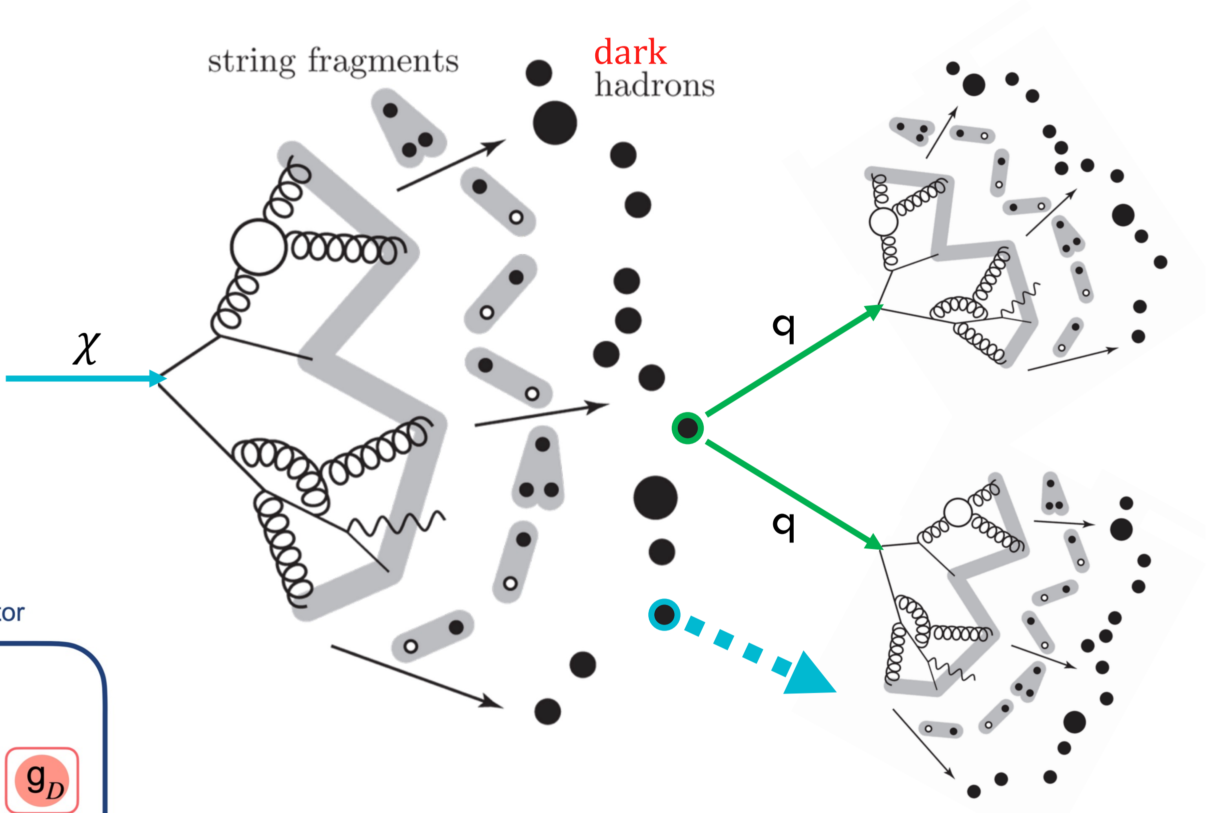


WHAT EVEN IS A SEMIVISIBLE JET??

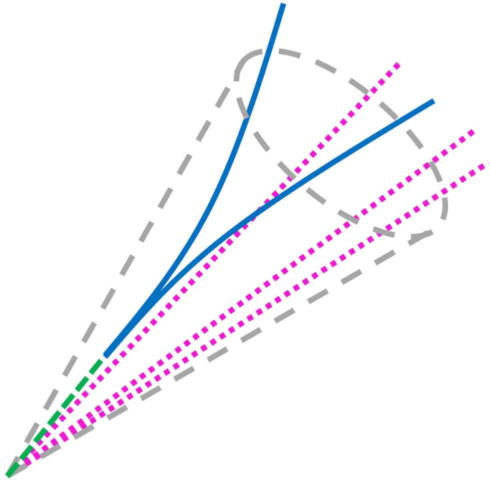


Dark quarks are produced by a proton-proton collision



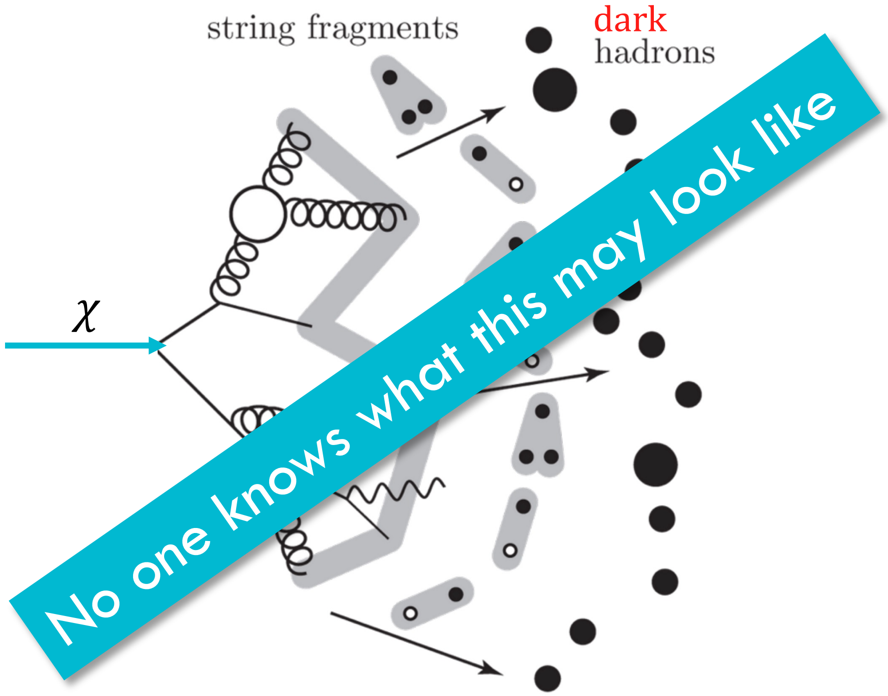


Stable dark hadrons are invisible
Unstable ones decay back to quarks and hadronized again



We have produced a semivisible jet (SVJ)

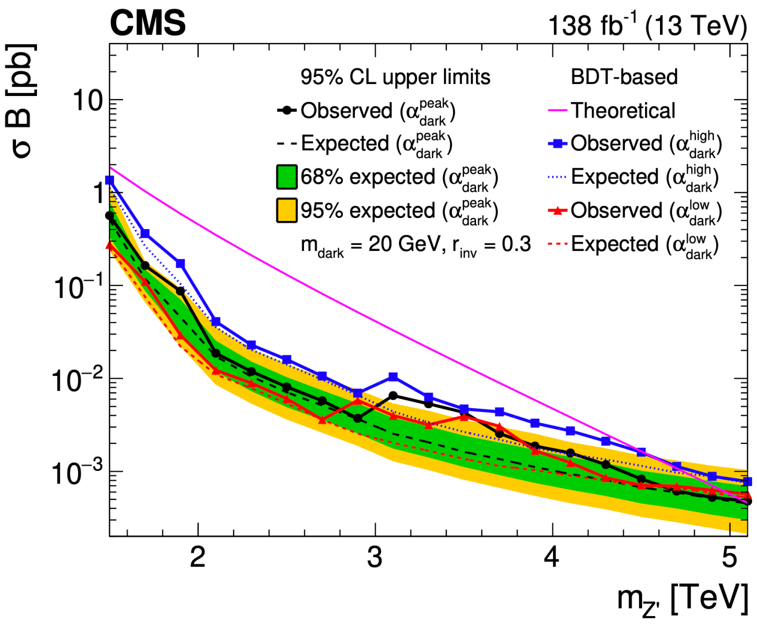
WHY GO UNSUPERVISED



- We need to tell these jets apart from “normal” jets
- Supervised approach needs truth label: simulate both (SM and SVJ) and train on that
    - Simulating QCD well is **hard**
    - Each simulation of SVJs assumes a given dark sector interaction $(N_{flav}^{dark}, N_{col}^{dark}, m_h^D, \Lambda_{QCD}^{dark})$
- Unsupervised approach: train on SM jets from data and tag anything anomalous
- Solves both issues – at the price of performance

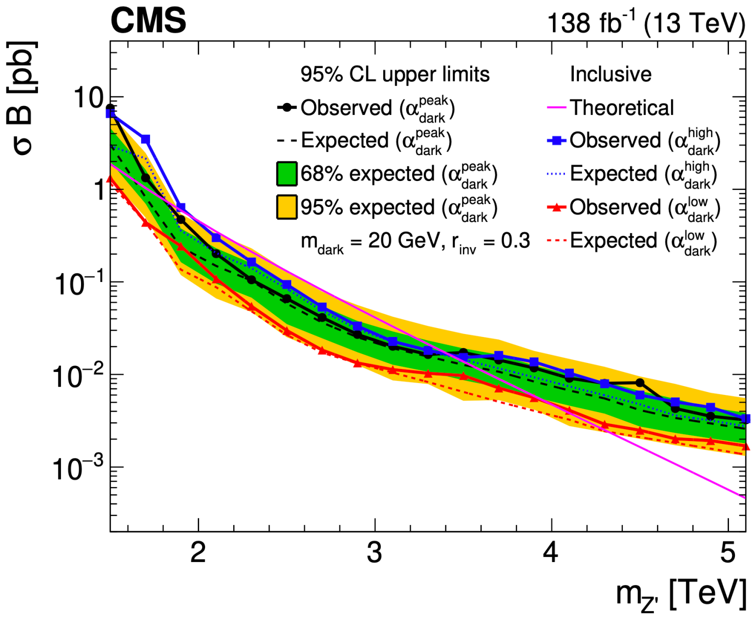
A BIT OF HISTORY

First ever search for SVJs published by CMS (10.1007/JHEP06(2022)156)



Supervised approach: BDT to tag SVJs

Something in between?

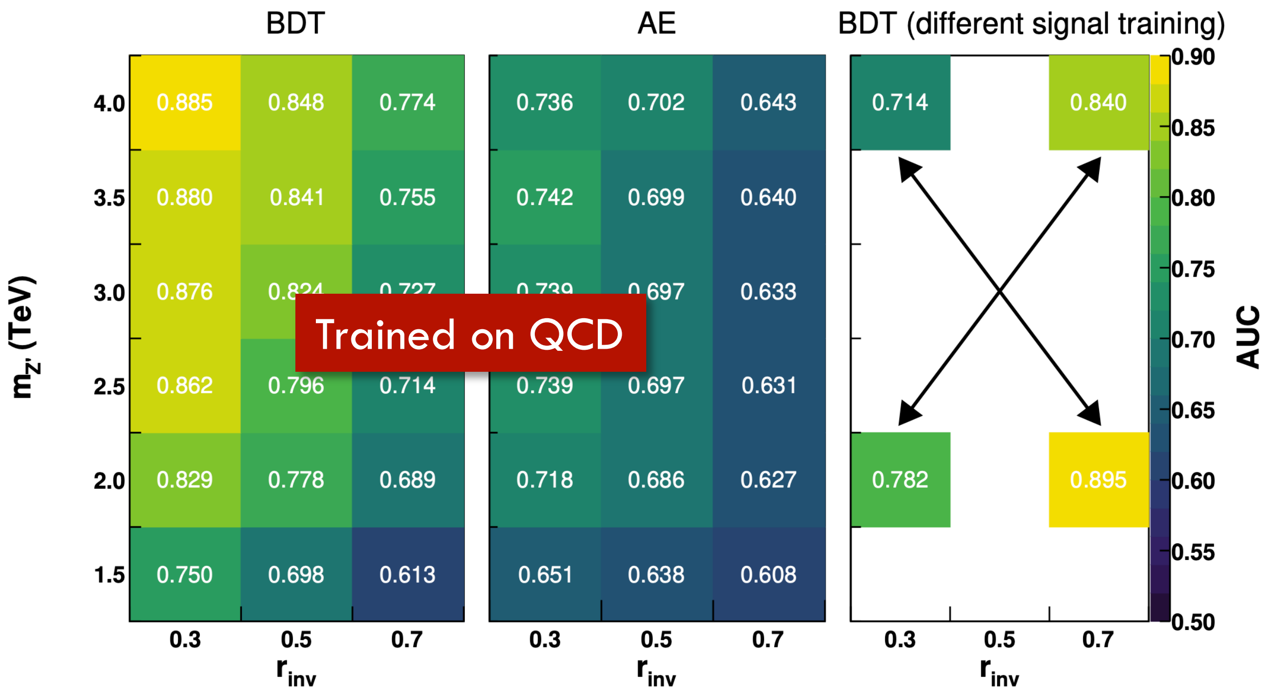


Model-agnostic approach: no jet tagging

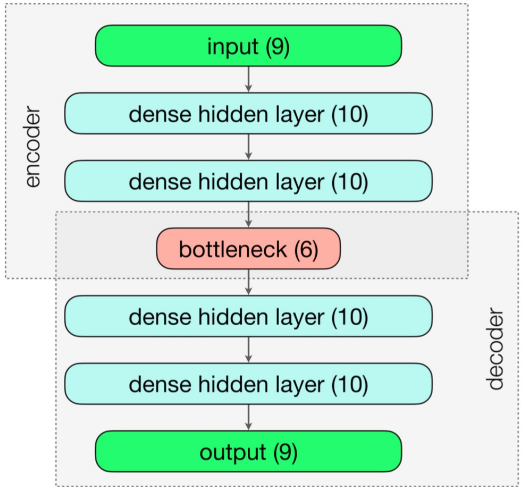
A BIT OF HISTORY

First ever search for SVJs published by CMS (JHEP06(2022)156)

First attempt to use autoencoders (AE) to tag SVJs as anomalous jets (JHEP02(2022)074)

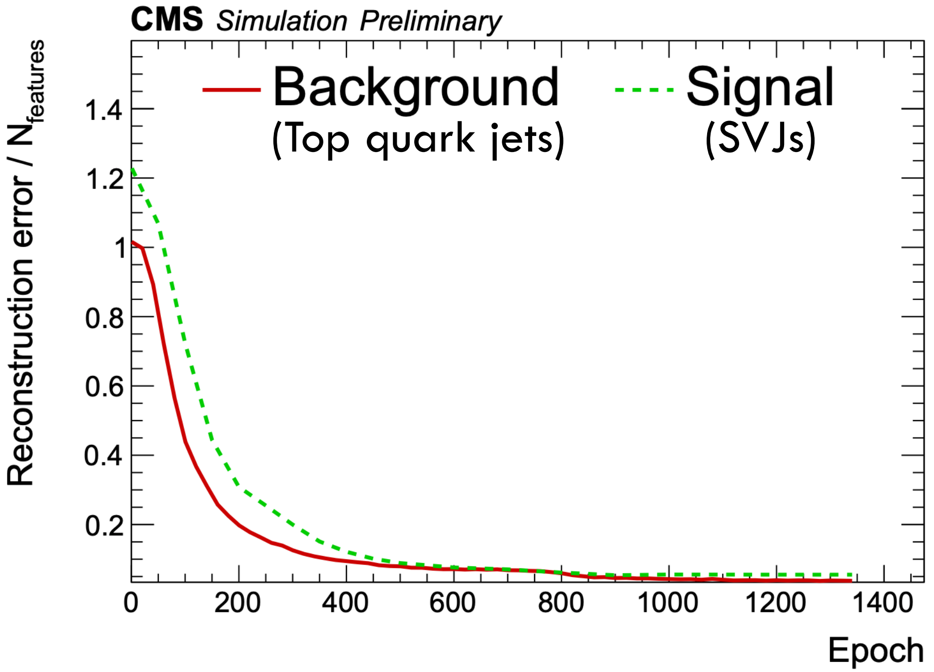


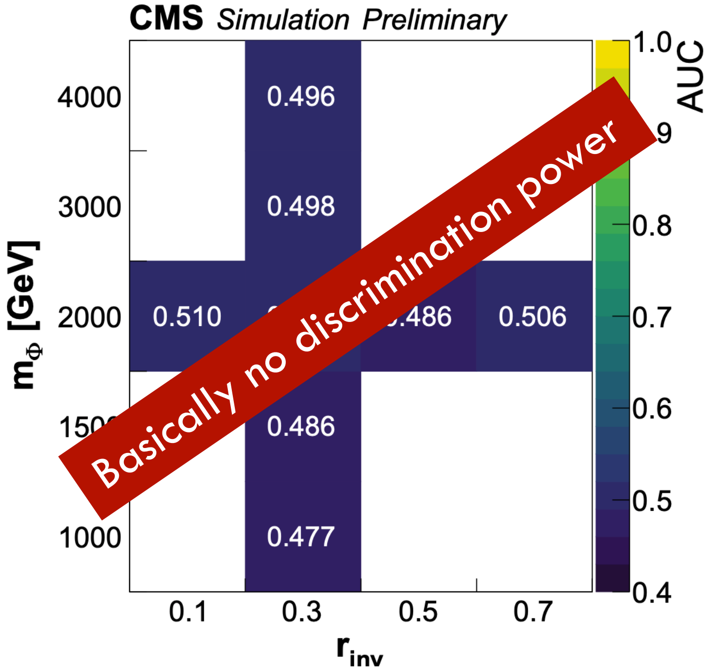
Even a simple AE can outperform a BDT trained on the "wrong" signal hypothesis



NEW TOOLS, NEW PROBLEMS

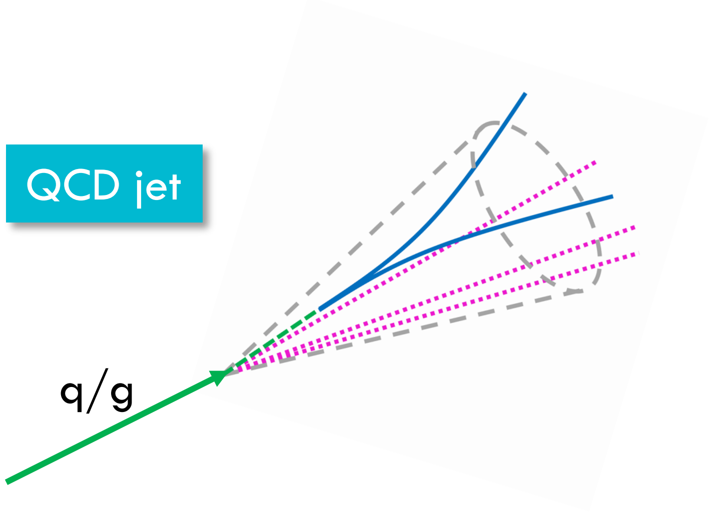
Training to have minimum reconstruction error on the background does not always work

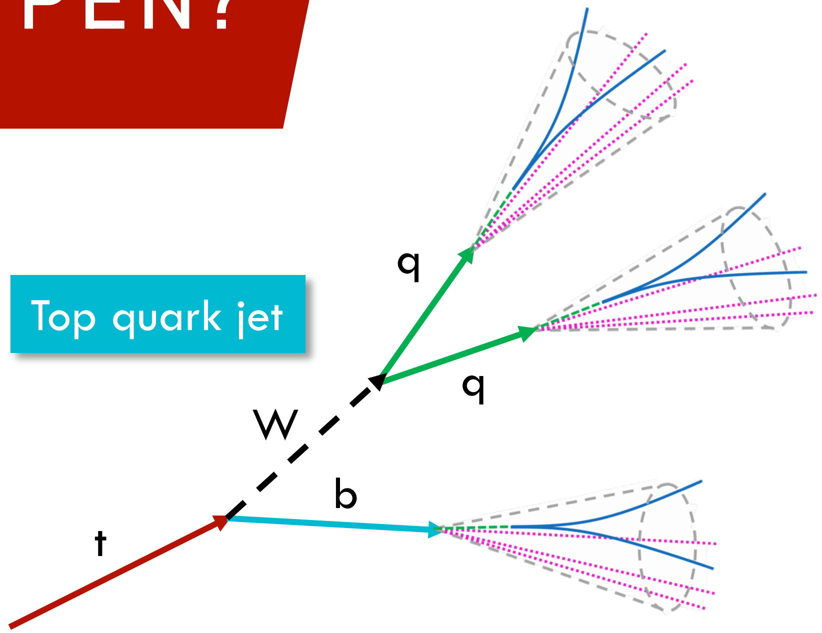




- The AE trained on top quark jets generalizes too well to SVJs, yielding no discrimination
- NB: still performing well against QCD jets

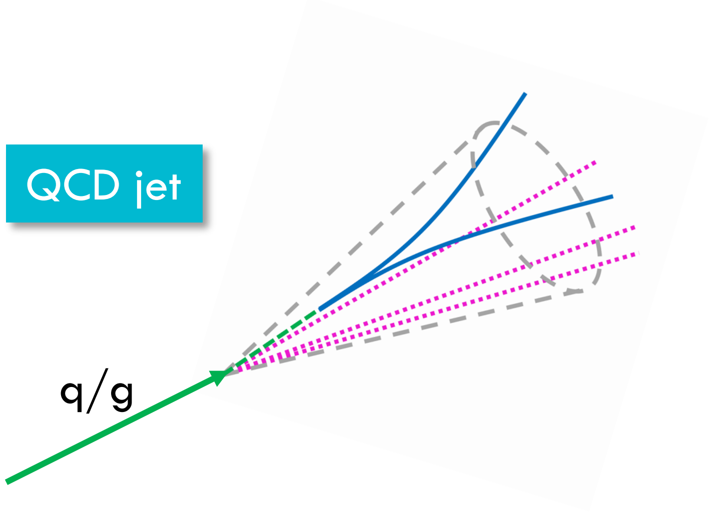
WHY DOES THIS HAPPEN?



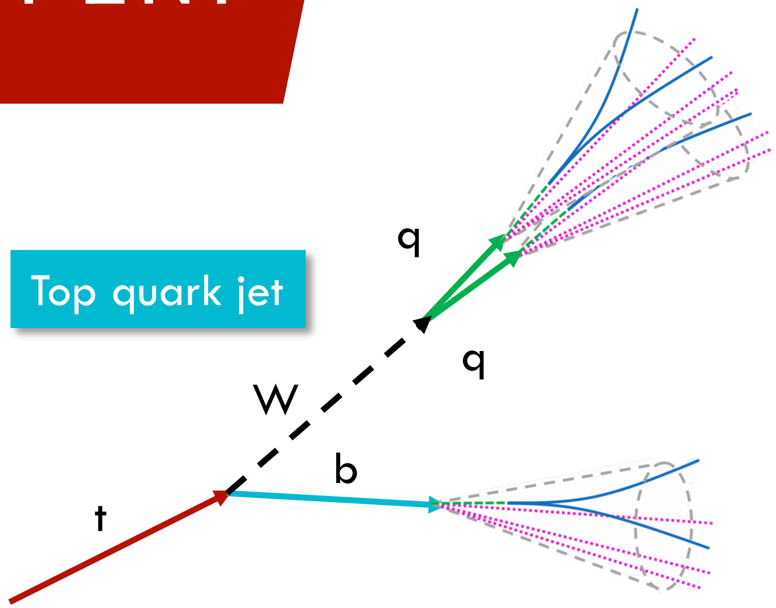


WHY DOES THIS HAPPEN?

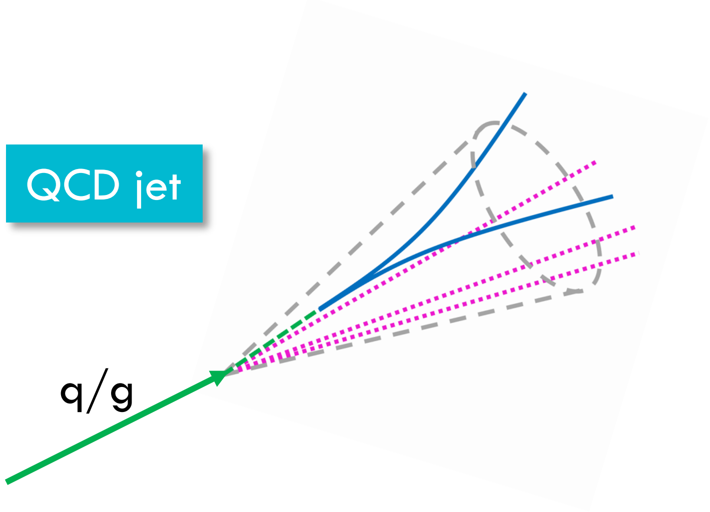
QCD jet

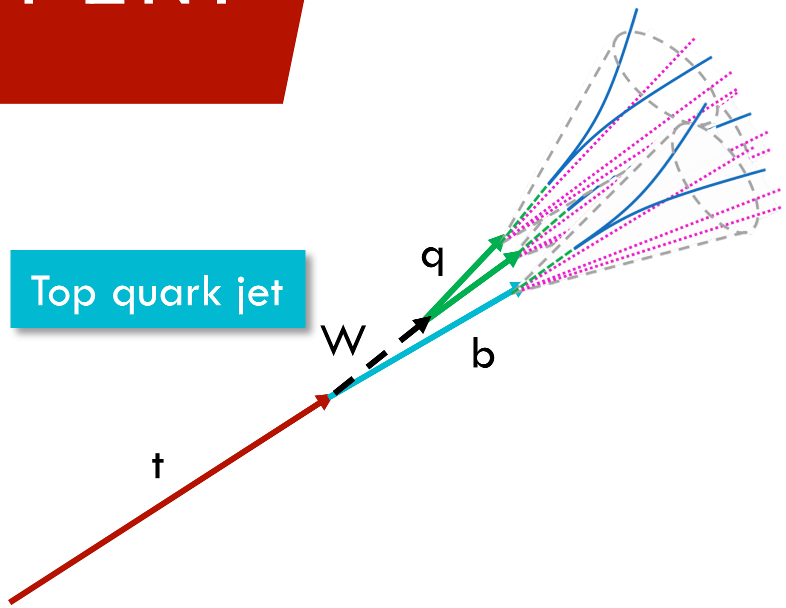


Top quark jet

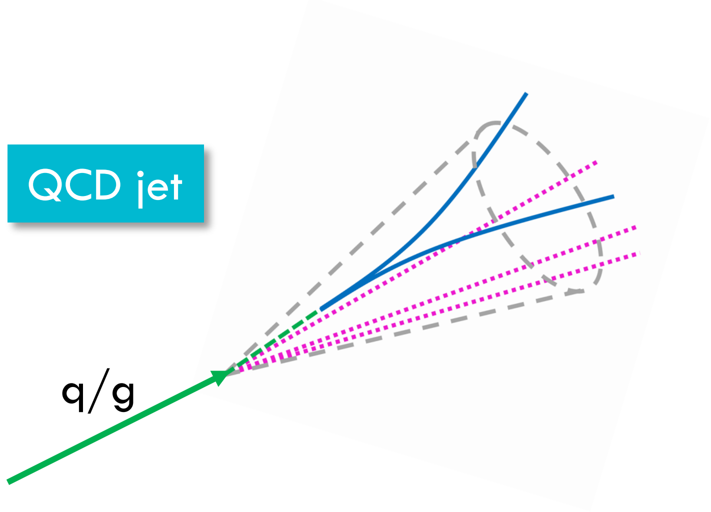


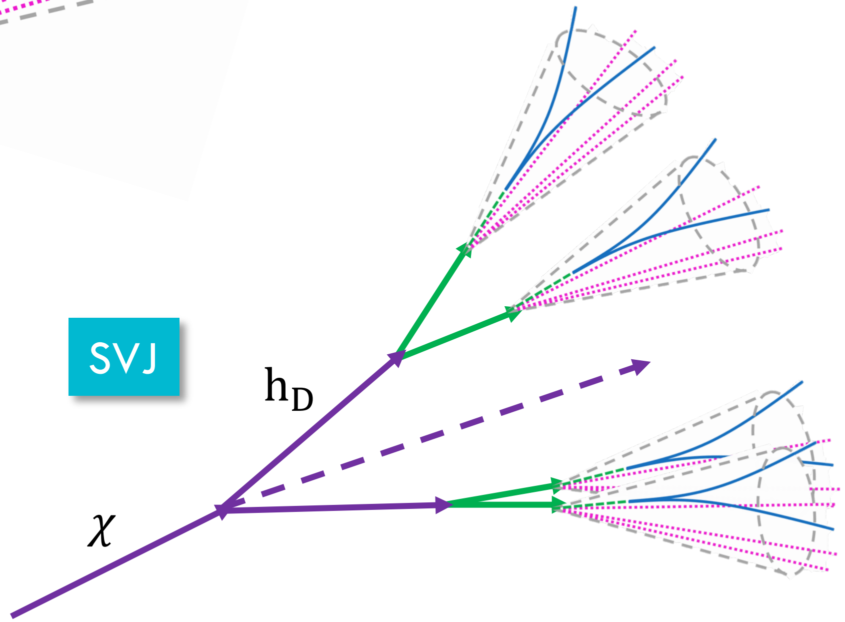
WHY DOES THIS HAPPEN?

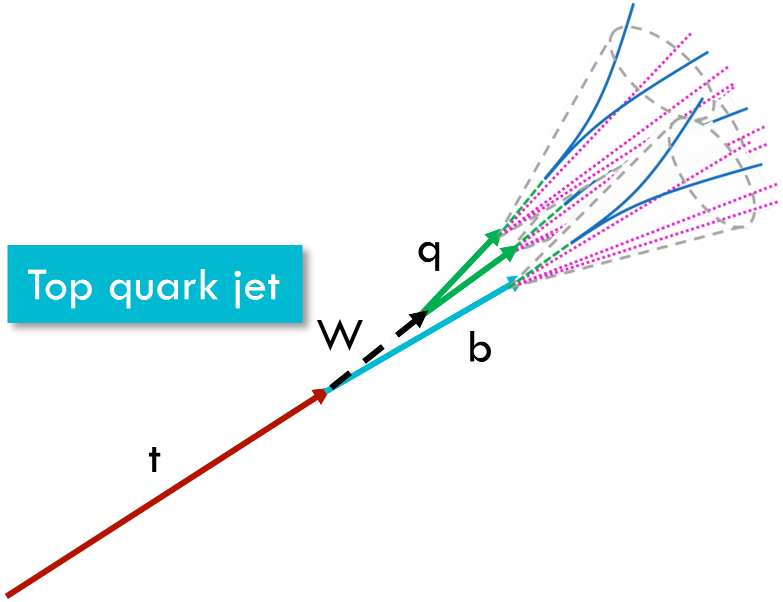




WHY DOES THIS HAPPEN?

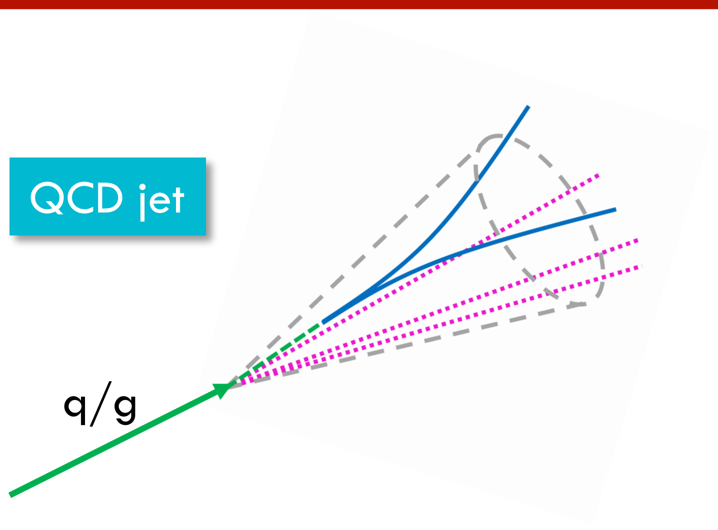




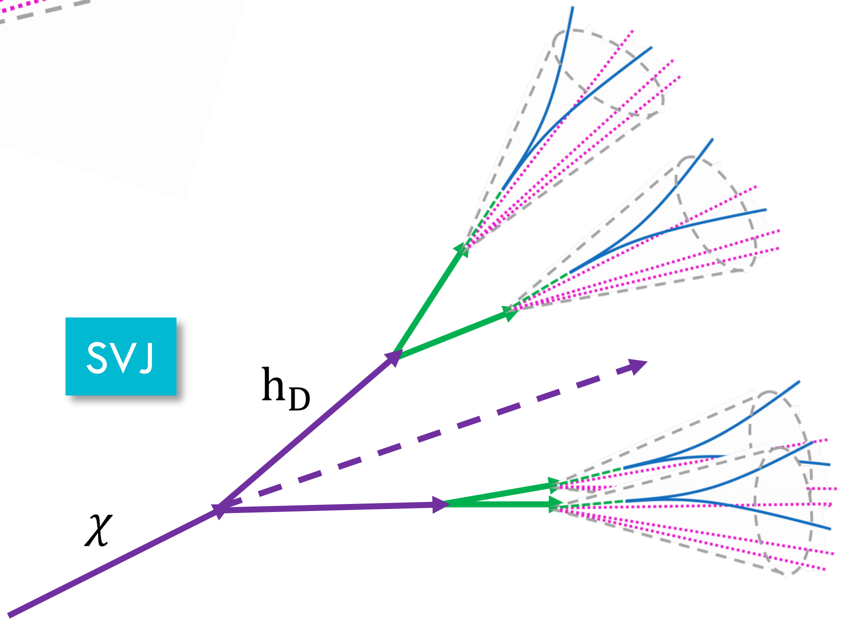


WHY DOES THIS HAPPEN?

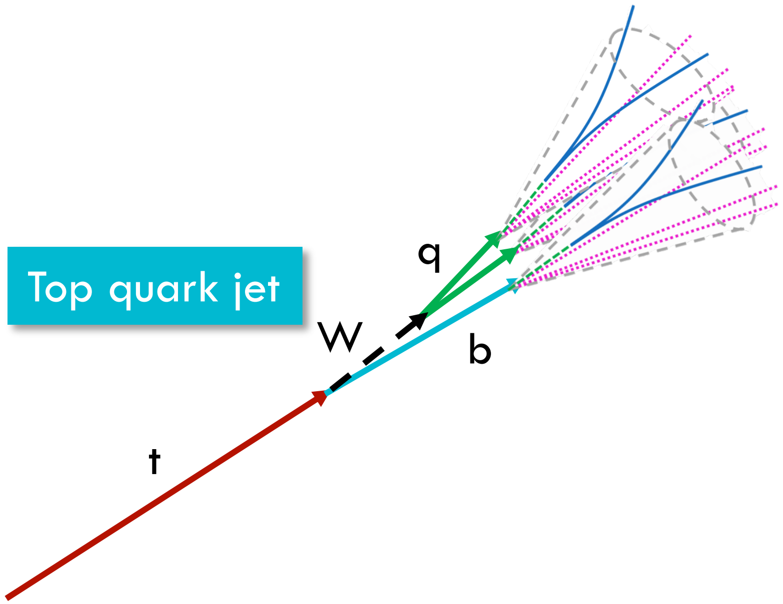
QCD jet



SVJ



Top quark jet

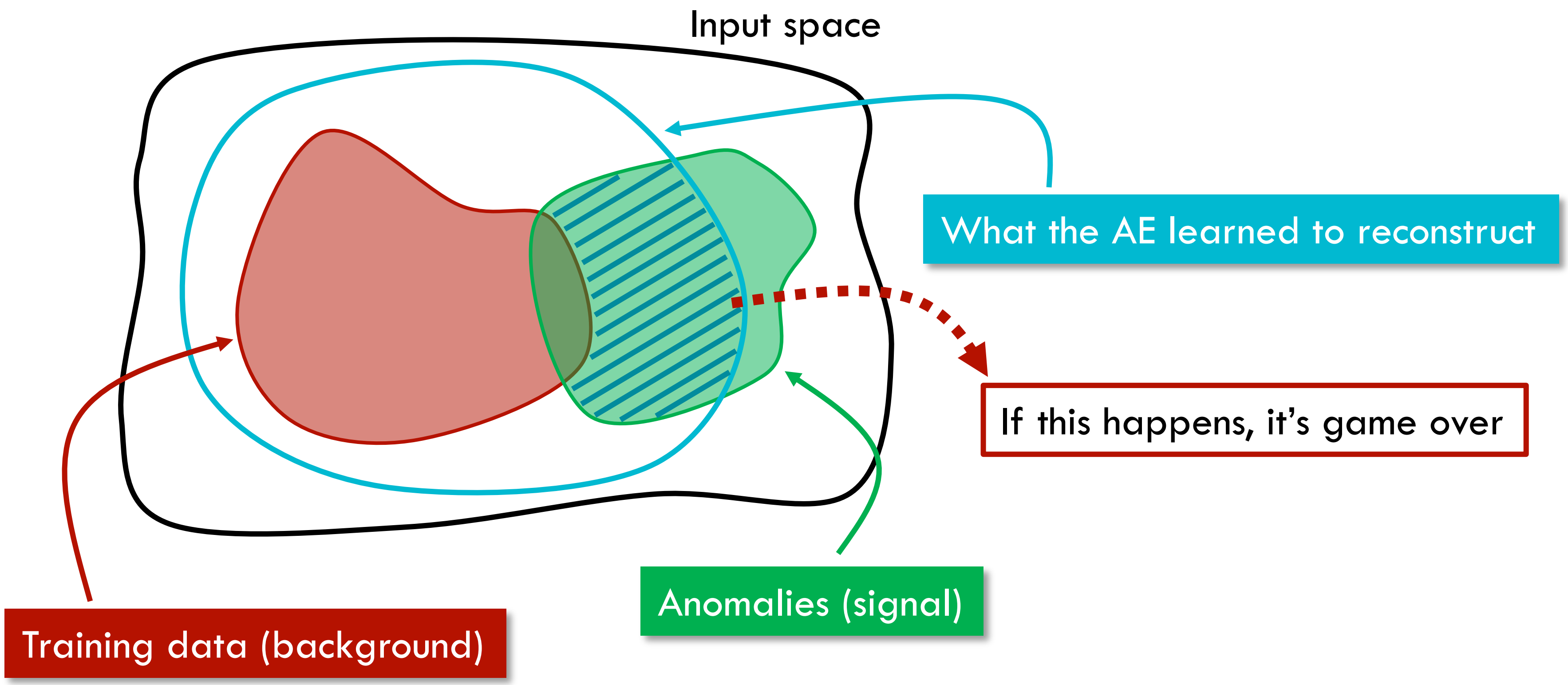


Top jets are a superset of QCD jets

↓

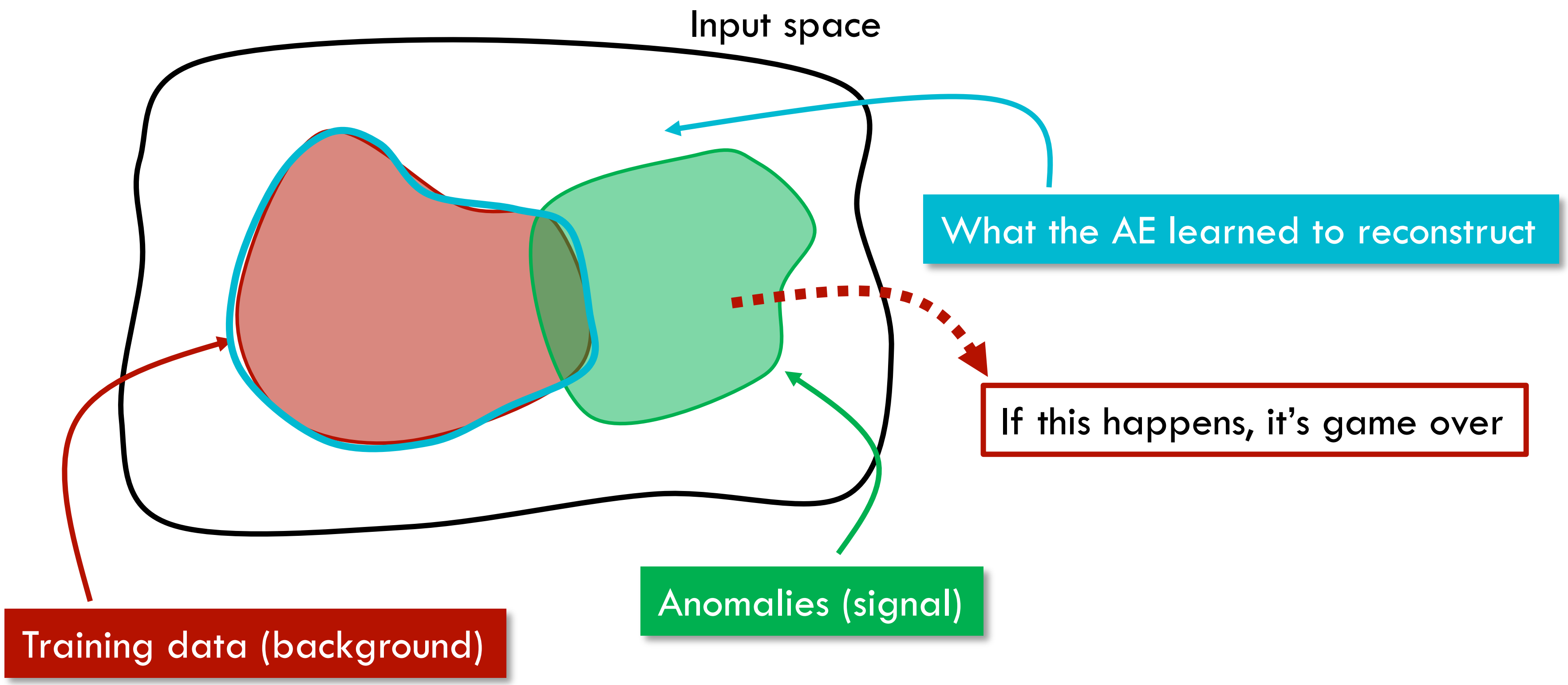
The AE learns to interpolate → reconstructs SVJs as well as top jets

OUTLIER RECONSTRUCTION



- We need a way to enforce that the AE only learns the background
- In other words: cyan should match red

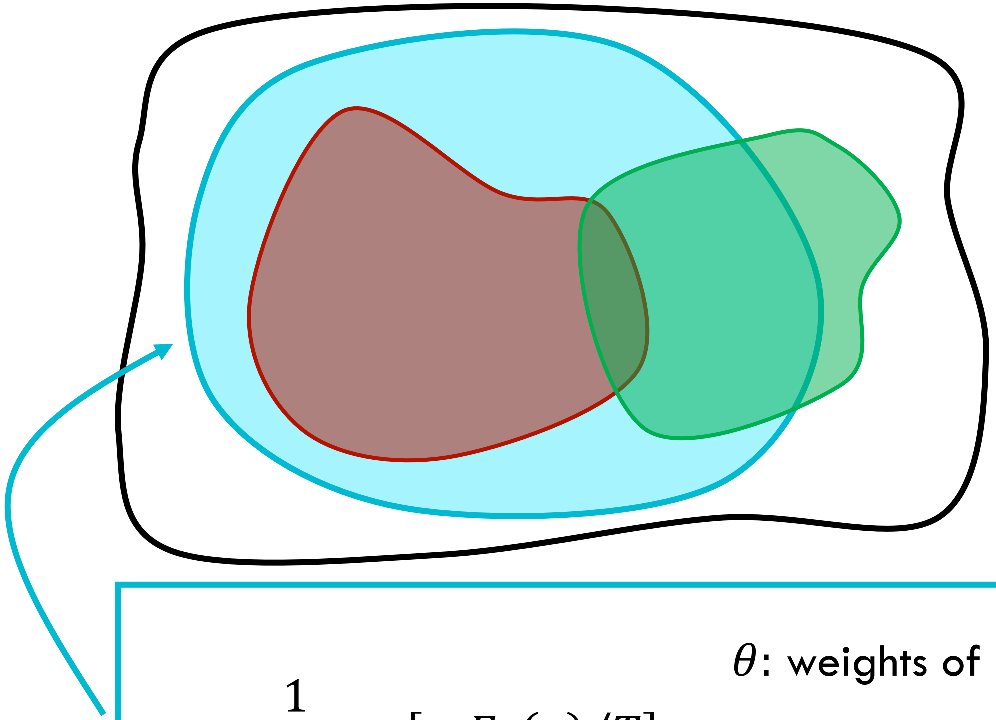
OUTLIER RECONSTRUCTION



- We need a way to enforce that the AE only learns the background
- In other words: cyan should match red

How do we achieve this?

THE NORMALIZED AUTOENCODER PARADIGM



- We need a way to explore the space of examples that the AE is able to reconstruct well
- Note that **this is purely a feature of the AE**
  - No need to specify a family of anomalies
- Solved by Yoon et al. in [this paper]

$$p_\theta = \frac{1}{\Omega_\theta} \exp[-E_\theta(x)/T]$$

θ: weights of the AE

x: point in input space

E: reconstruction error

We can sample this distribution via Monte Carlo and enforce $p_\theta = p_{data}$

HOW IT WORKS

Define the positive (negative) energy as the average reconstruction error on examples drawn from p_{data} (p_θ)

$$E_+ = \mathbb{E}_{x\sim p_{data}}[E_\theta(x)] \qquad E_- = \mathbb{E}_{x'\sim p_\theta}[E_\theta(x')]$$

Train on minimizing the difference of the energies

$$\mathcal{L} = E_+ - E_-$$

→ Equivalent to minimizing the likelihood of the training data (details in backup)

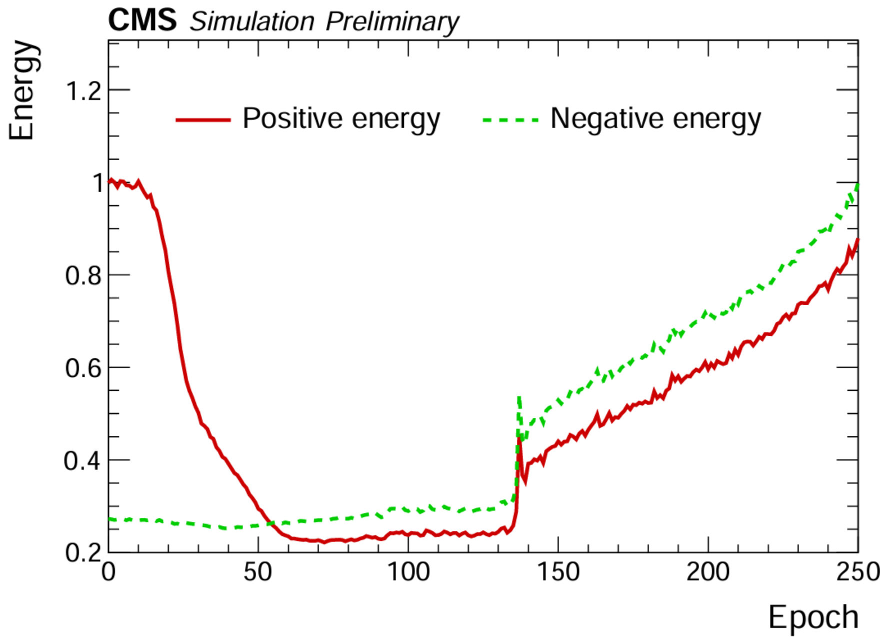
↓

Profound paradigm shift: the normalized autoencoder (NAE) is a fully-fledged statistical model of the training data

The actual AE can be kept very simple: (10 10 6 10 10) fully connected in this example

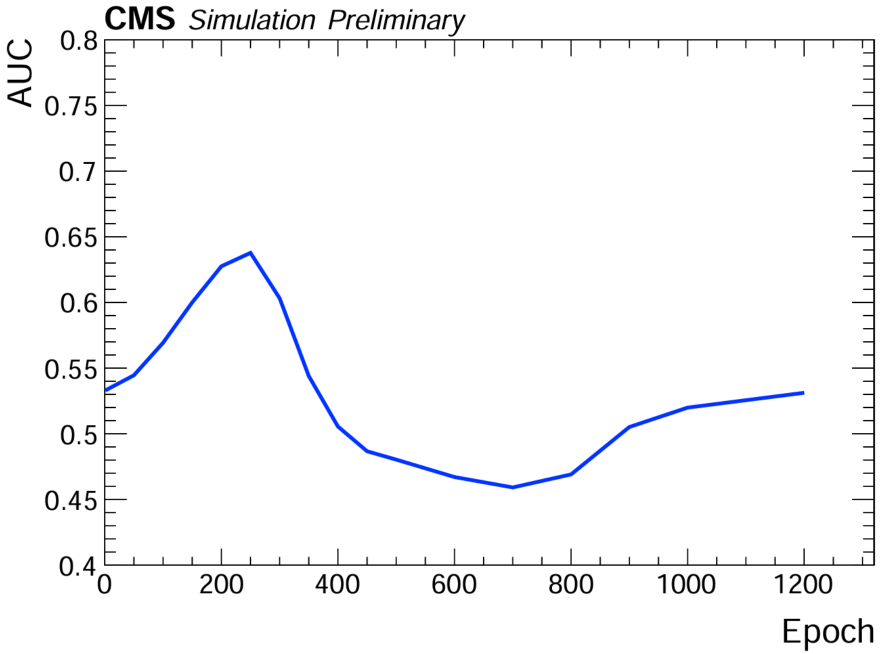
TWO IMPORTANT CAVEATS

Defining the loss as $E_+ - E_-$ can lead to a runaway effect when $E_- > E_+$



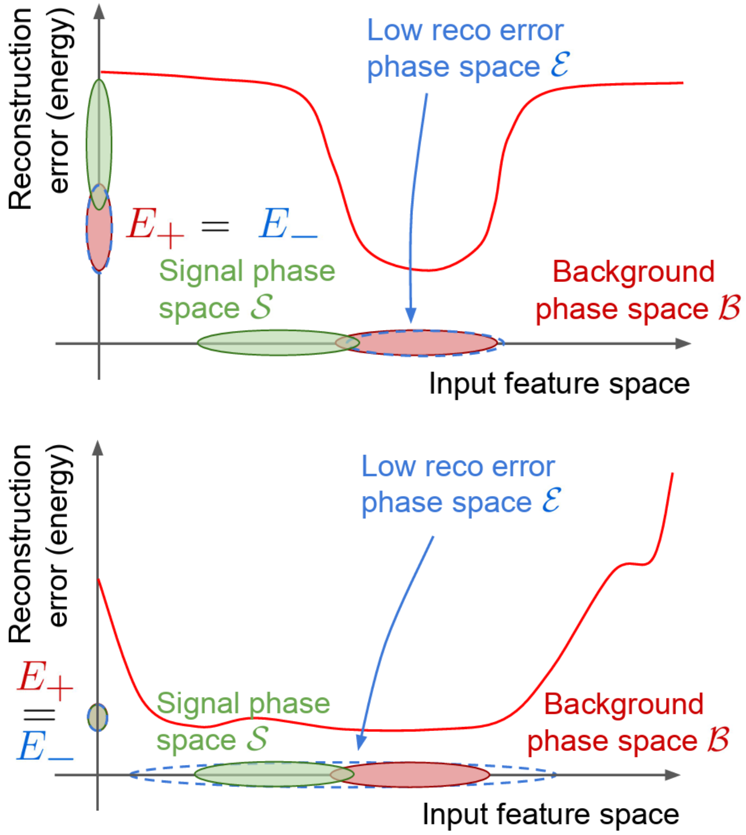
There is a mode collapse in the training

- Discrimination performance increases up to a certain point
- Sharp drop afterwards



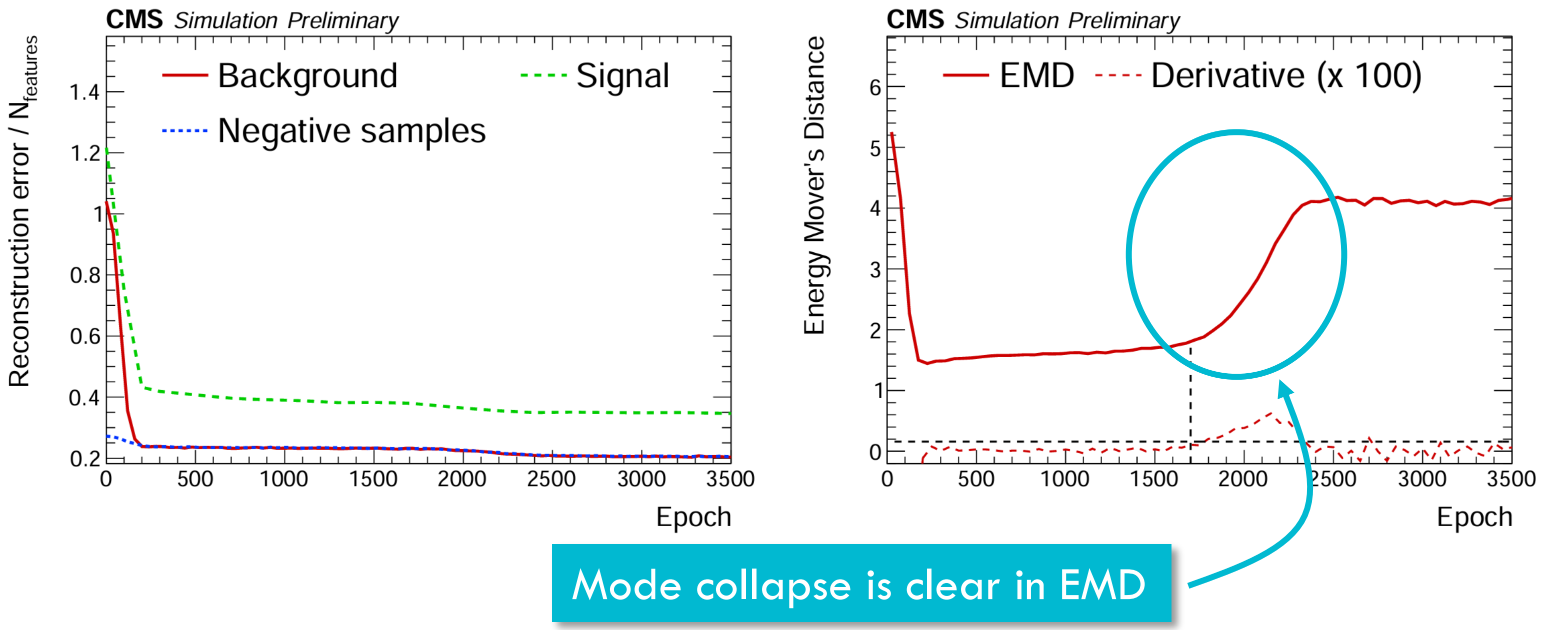
A BETTER METRIC

This drop is invisible in the energy difference



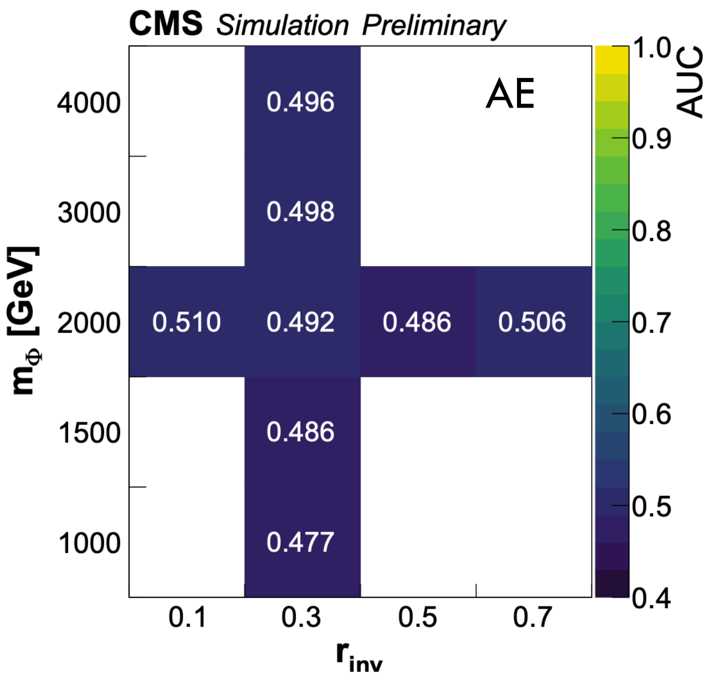
Need a more robust way to measure the distance between p_{data} and p_θ → the Earth Mover's (or Wasserstein 1) distance (EMD)

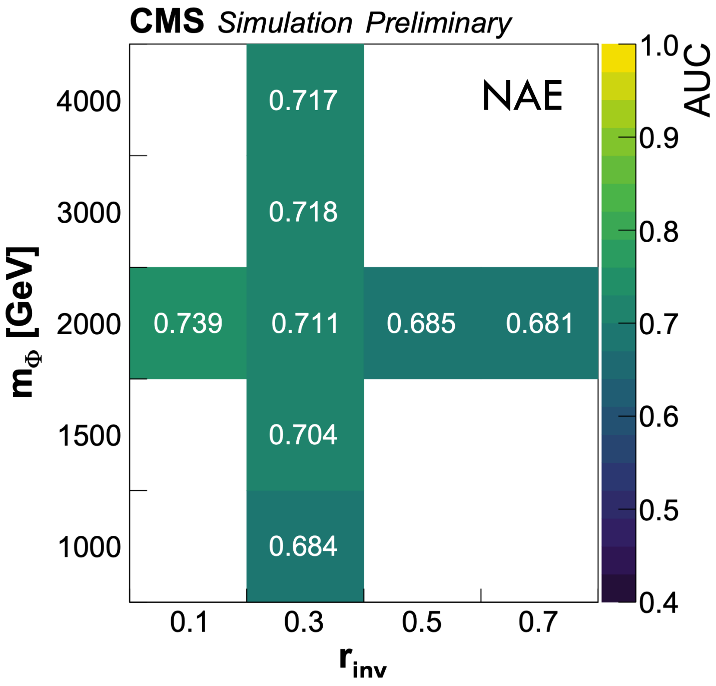
$$EMD(p_{data}, p_\theta) = \inf_{\gamma \in \prod(p_{data}, p_\theta)} \mathbb{E}_{(x,x')\sim\gamma}[||x - x'||]$$



APPLYING TO SVJS

Applied to the SVJ case, strong improvement in discrimination power against top quark jets [CMS-DP-2023-071]

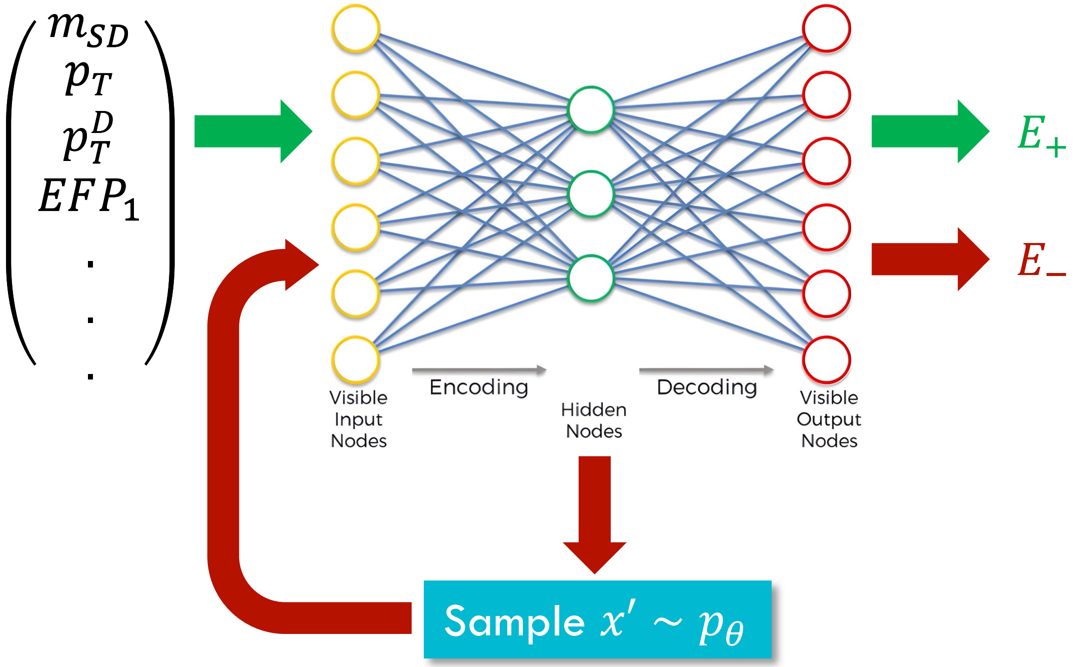




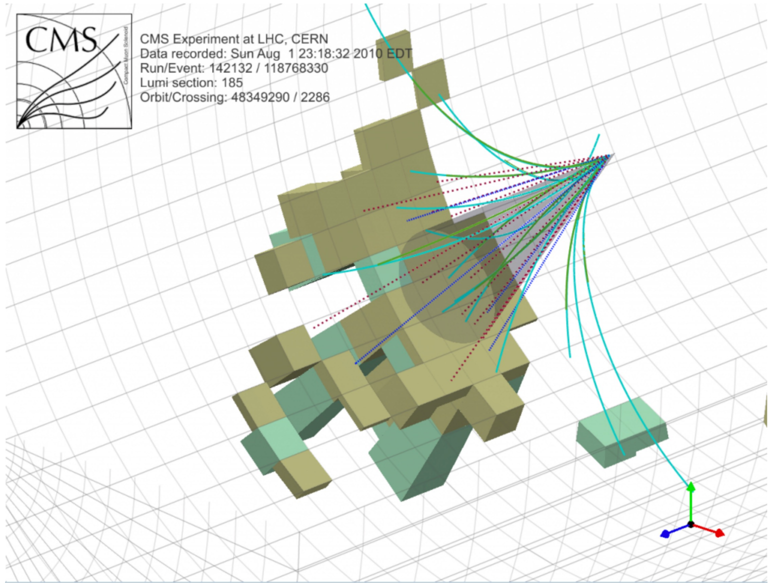
- Using the EMD to avoid mode collapse gives reliable results
- **Huge gain in performance**

Can we do better?

A MORE NATURAL REPRESENTATION

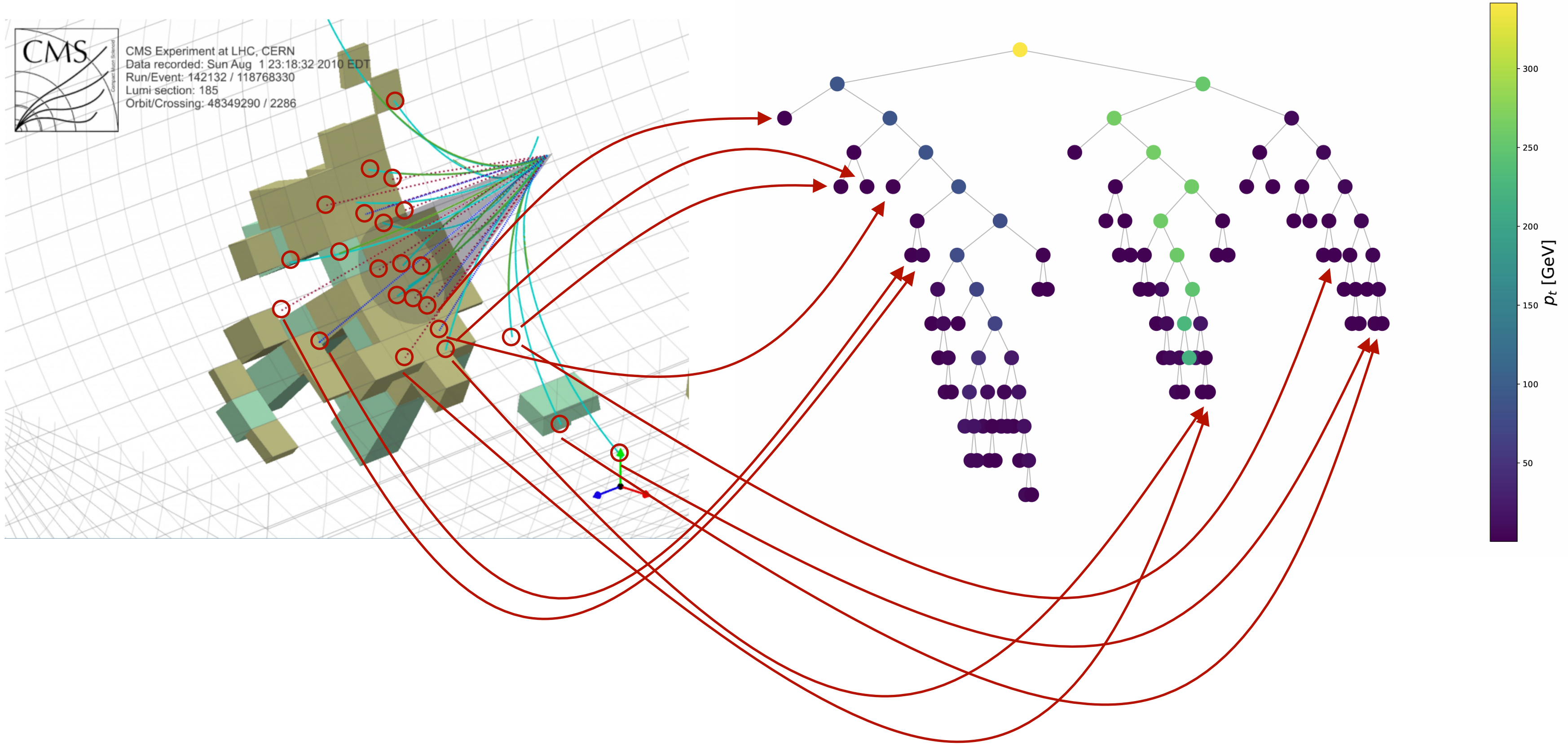


$\mathcal{L} = E_+ - E_-$

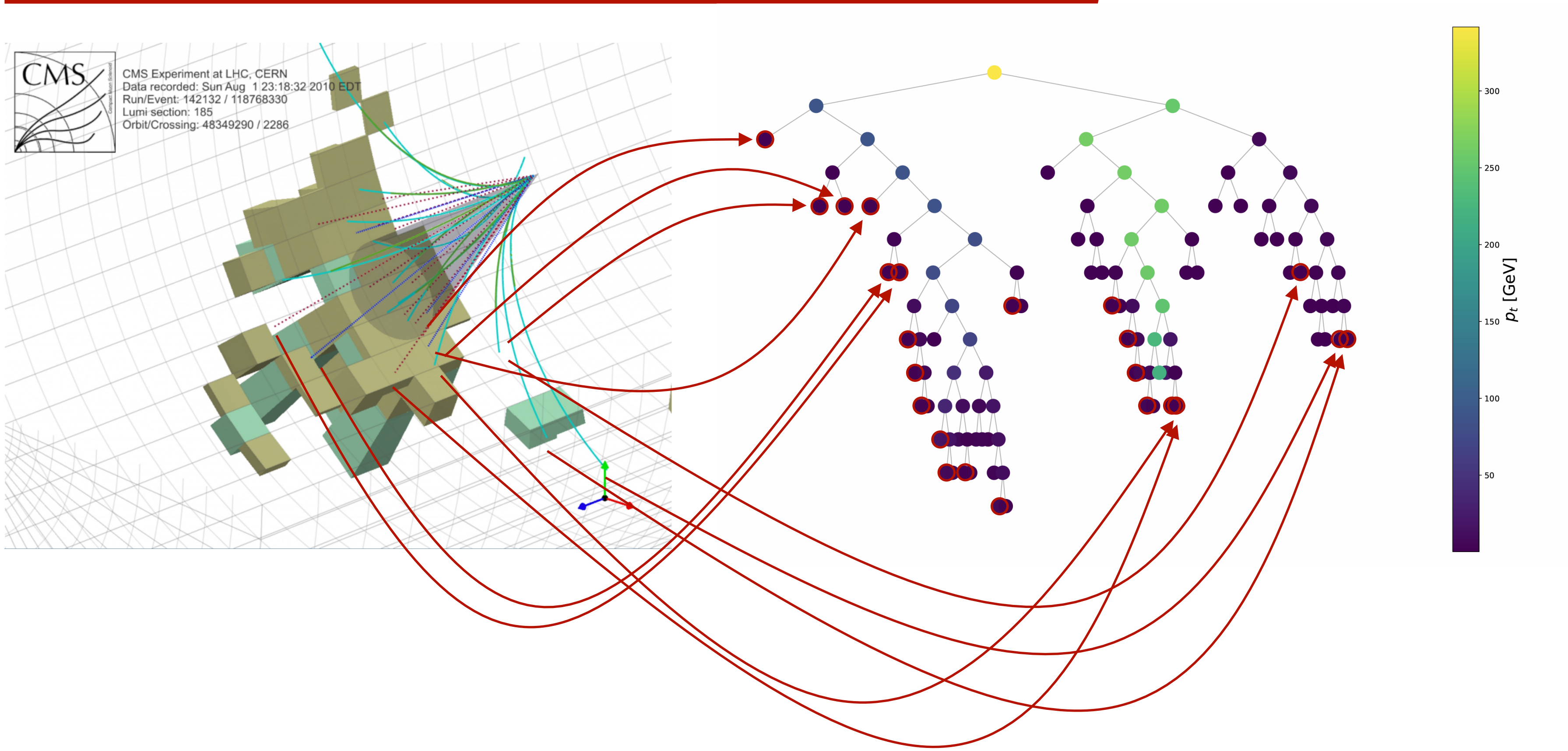


Can we do better then engineered features?

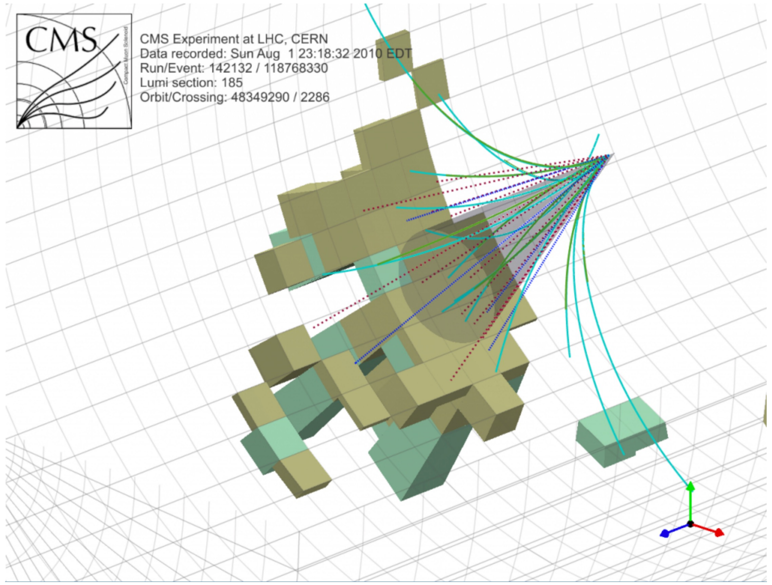
A MORE NATURAL REPRESENTATION

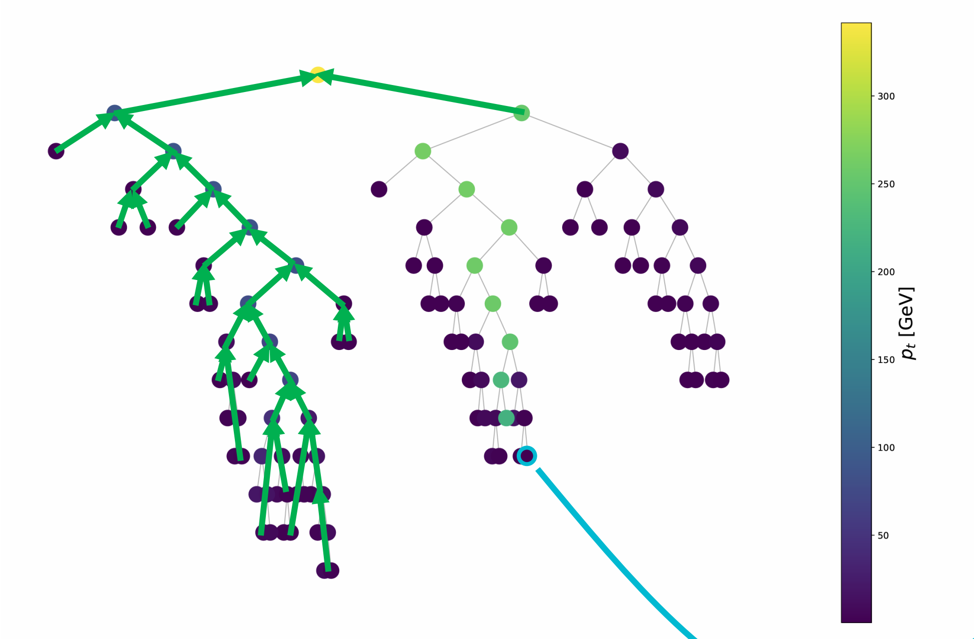


A MORE NATURAL REPRESENTATION



A MORE NATURAL REPRESENTATION



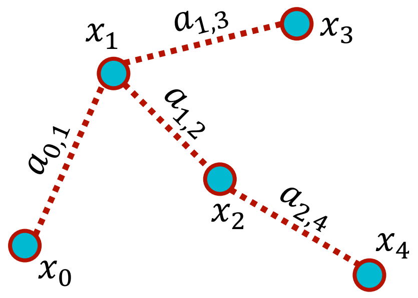


This «Lund graph» encodes the complete information of the jet

$\{p_T, \eta, \phi, m, \ldots\}$

Key problem: p_θ is now a distribution over graph space → Need a way to sample over graphs

NORMALIZED GRAPH AE



$X = (x_0, \dots, x_N)$ → Node feature matrix

$$A = \begin{pmatrix} 1 & \cdots & a_{0,N} \\ \vdots & \ddots & \vdots \\ a_{N,0} & \cdots & 1 \end{pmatrix}$$ → Edge matrix

Sample p_θ via MCMC:

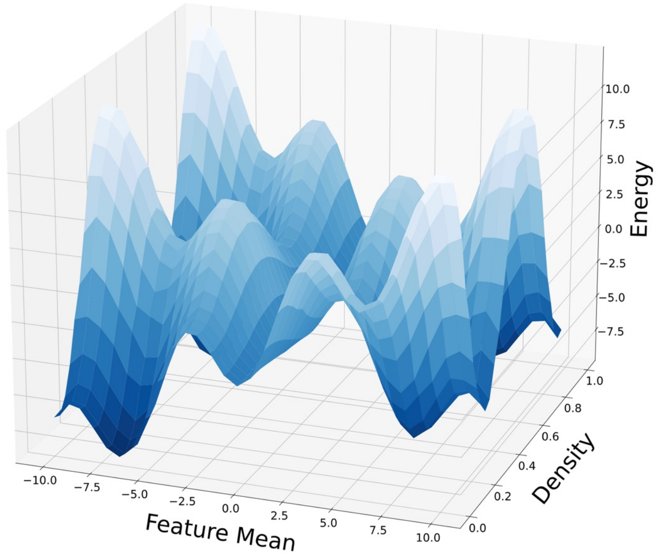
$$x_{n+1} = x_n + \lambda \nabla \log p_\theta + \varepsilon \sigma_x = x_n - \frac{\lambda}{T} \nabla E_\theta(x_n) + \varepsilon \sigma_x$$

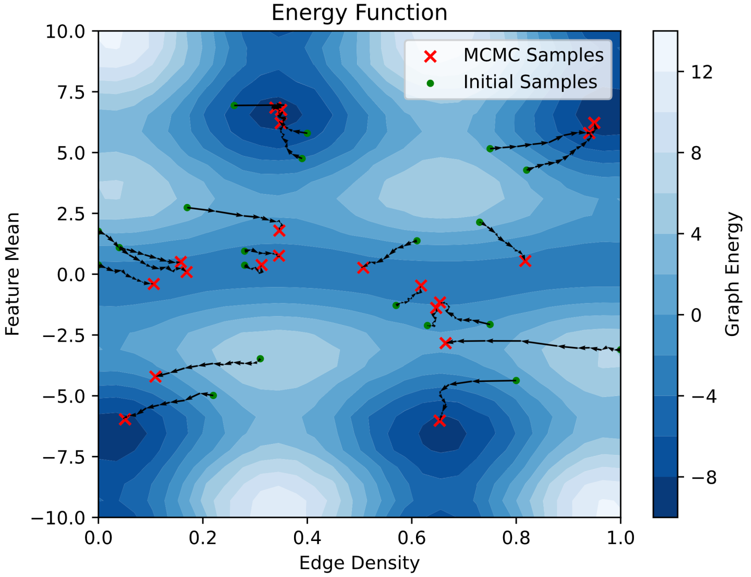
$$X_n = X_{n-1} - \frac{\alpha}{T} E_\theta(X_{n-1}, A_{n-1}) + \beta \sigma_X$$

$$A_n = A_{n-1} - \frac{\gamma}{T} E_\theta(X_{n-1}, A_{n-1}) + \delta \sigma_A$$

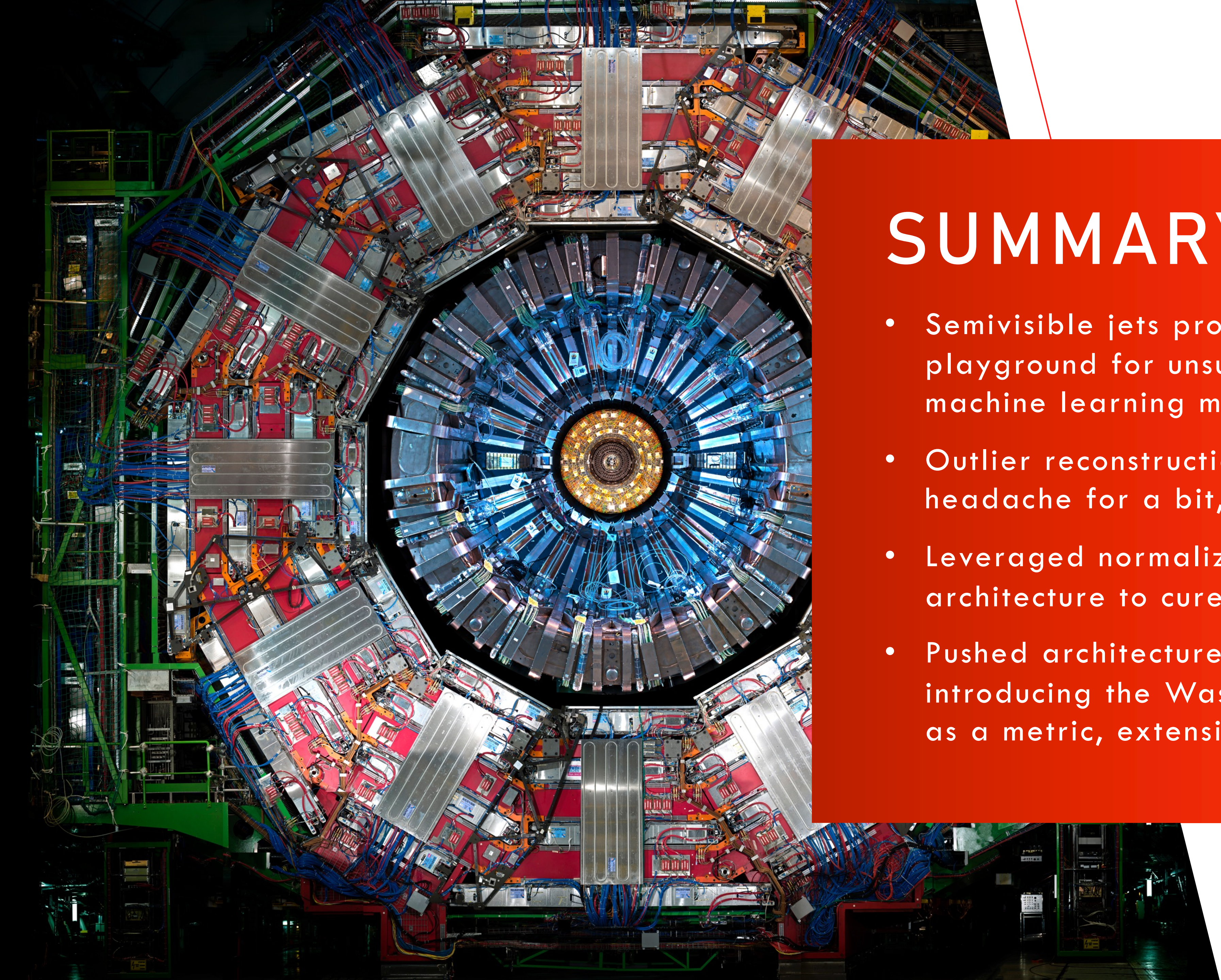
NORMALIZED GRAPH AE

$$X_n = X_{n-1} - \frac{\alpha}{T} E_\theta(X_{n-1}, A_{n-1}) + \beta\sigma_X \qquad A_n = A_{n-1} - \frac{\gamma}{T} E_\theta(X_{n-1}, A_{n-1}) + \delta\sigma_A$$





- The sampling of p_θ can be extended to graphs
- The rest of the NAE pipeline remains unchanged
- **Enables the extension of NAEs to graph networks**



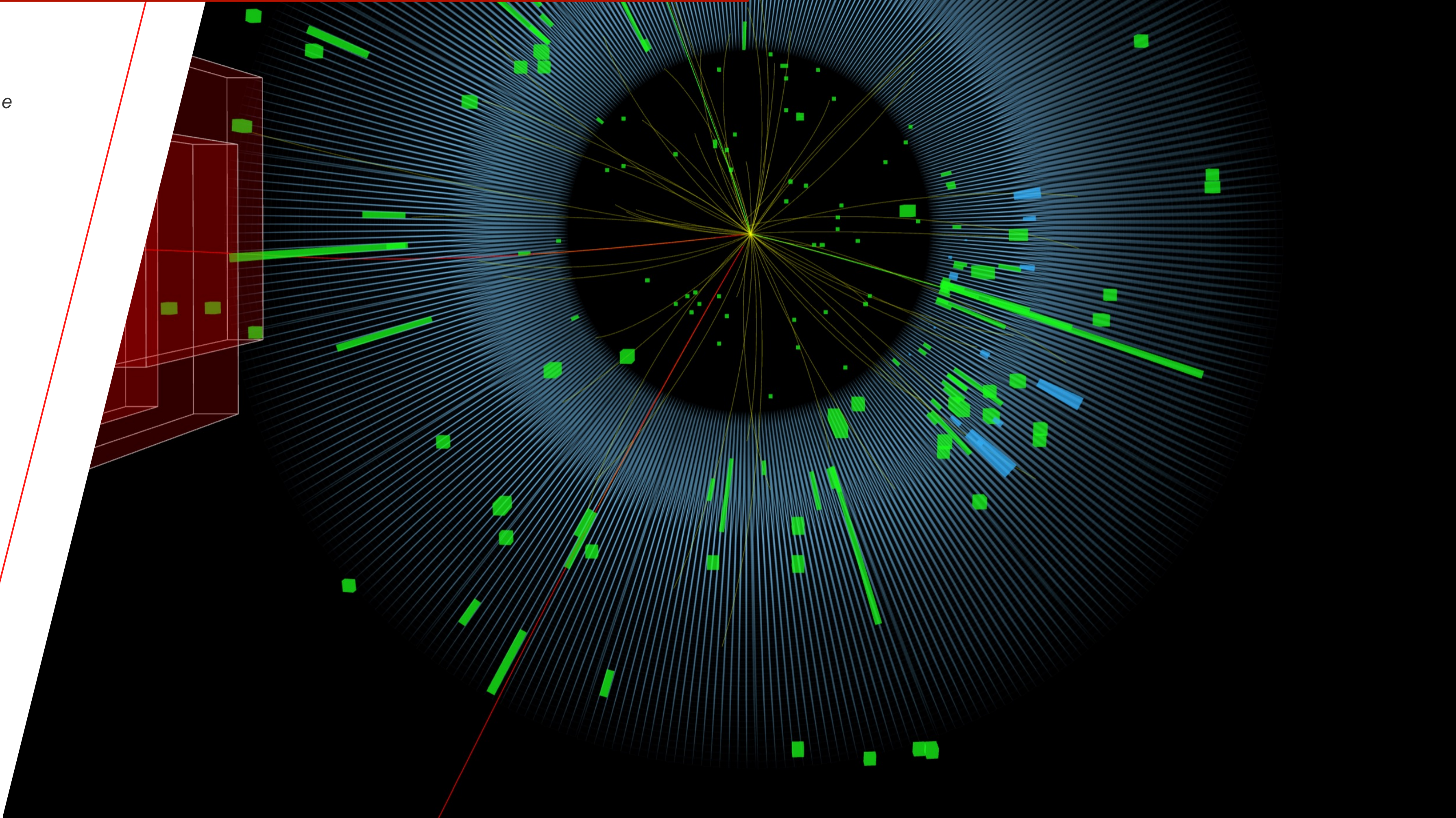
SUMMARY

- Semivisible jets provide a unique playground for unsupervised machine learning models
- Outlier reconstruction gave us a headache for a bit, but
- Leveraged normalized autoencoder architecture to cure the issue
- Pushed architecture forward by introducing the Wasserstein distance as a metric, extension to graphs

THANK YOU

*« Ce qui est admirable, ce n’est pas que
le champ des étoiles soit si vaste,
c’est que l’homme l’ait mesuré. »*

Jacques Anatole François Thibault



BACKUP

NAE DERIVATION

$$p_\theta = \frac{1}{\Omega_\theta}\exp(-E_\theta(x)/T)$$

$$\mathbb{E}_{x\sim p_{data}}[-\log p_\theta(x)] = \mathbb{E}_{x\sim p_{data}}[E_\theta(x)]/T + \log \Omega_\theta$$ → The painful part

$$\Omega_\theta = \int_{\mathcal{B}} dx \exp(-E_\theta(x)/T)$$ →

$$\begin{aligned}
\nabla_\theta \log \Omega_\theta &= \frac{1}{\Omega_\theta}\nabla_\theta \Omega_\theta \\
&= \frac{1}{\Omega_\theta}\int_{\mathcal{B}} dx \nabla_\theta \exp(-E_\theta(x)/T) \\
&= \frac{1}{\Omega_\theta}\int_{\mathcal{B}} dx \exp(-E_\theta(x)/T)\nabla_\theta(-E_\theta(x)/T) \\
&= -\frac{1}{T}\int_{\mathcal{B}} dx \exp(-E_\theta(x)/T)\nabla_\theta E_\theta(x) \\
&= -\frac{1}{T}\mathbb{E}_{x\sim p_\theta(x)}[\nabla_\theta E_\theta(x)],
\end{aligned}$$