Towards a deep learning model for hadronization

[an example of the application of the ML generative models in high-energy physics]

Andrzej Siódmok

Towards a Deep Learning Model for Hadronization

Aishik Ghosh,[a,b] Xiangyang Ju,[b] Benjamin Nachman,[b,c] and Andrzej Siodmok[d]

[a] *Department of Physics and Astronomy, University of California, Irvine, CA 92697, USA*
[b] *Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*
[c] *Berkeley Institute for Data Science, University of California, Berkeley, CA 94720, USA*
[d] *Jagellonian University, Krakow, Poland*

2203.12660

Fitting a Deep Generative Hadronization Model

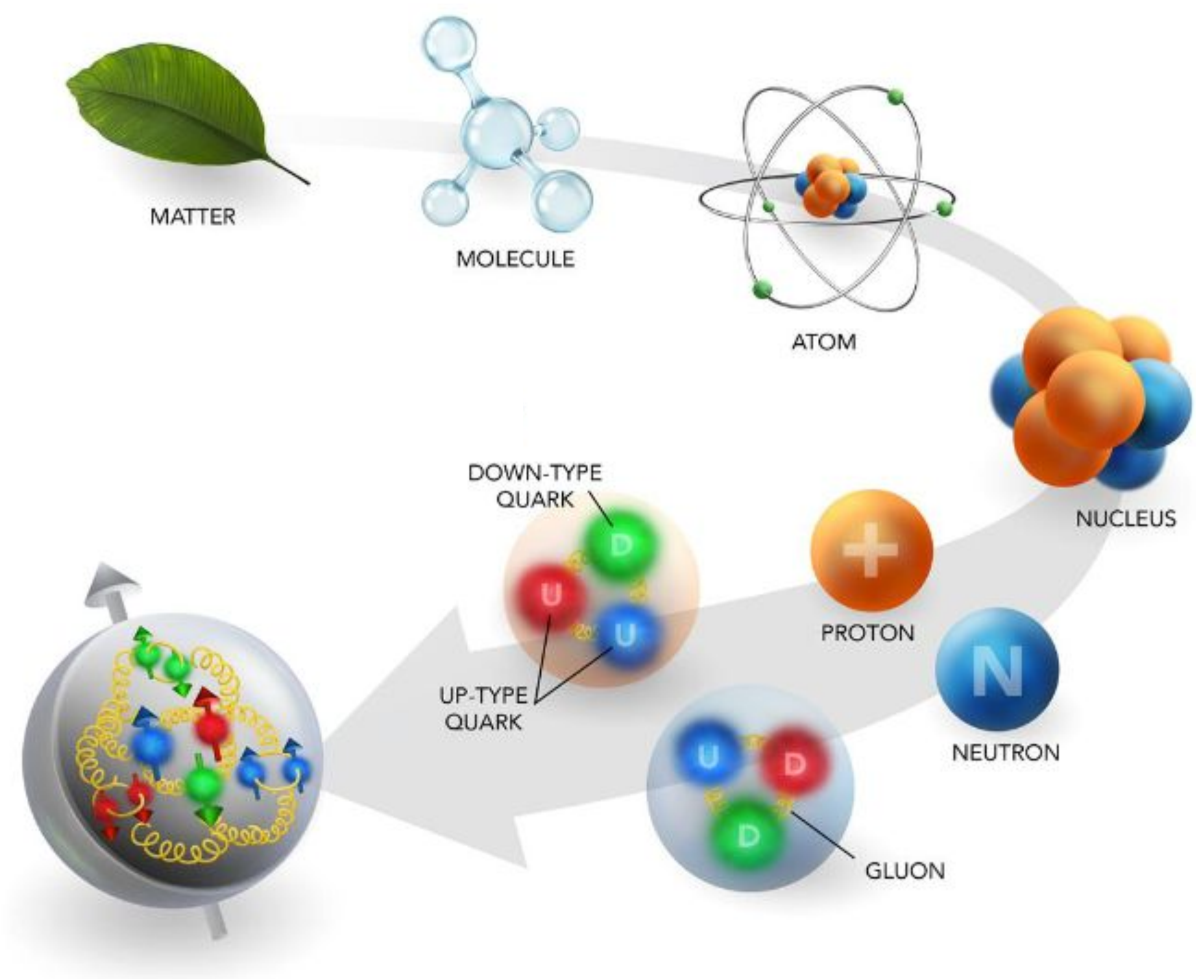
Jay Chan,[a,b] Xiangyang Ju,[b] Adam Kania,[e] Benjamin Nachman,[b,c] Vishnu Sangli,[d,b] and Andrzej Siodmok[d]

[a] *Department of Physics, University of Wisconsin-Madison, Madison, WI 53706, USA*
[b] *Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*
[c] *Berkeley Institute for Data Science, University of California, Berkeley, CA 94720, USA*
[d] *Department of Physics, University of California, Berkeley, CA 94720, USA*
[e] *Jagiellonian University, Krakow, Poland*

2305.17169

NCN: 2019/34/E/ST2/00457

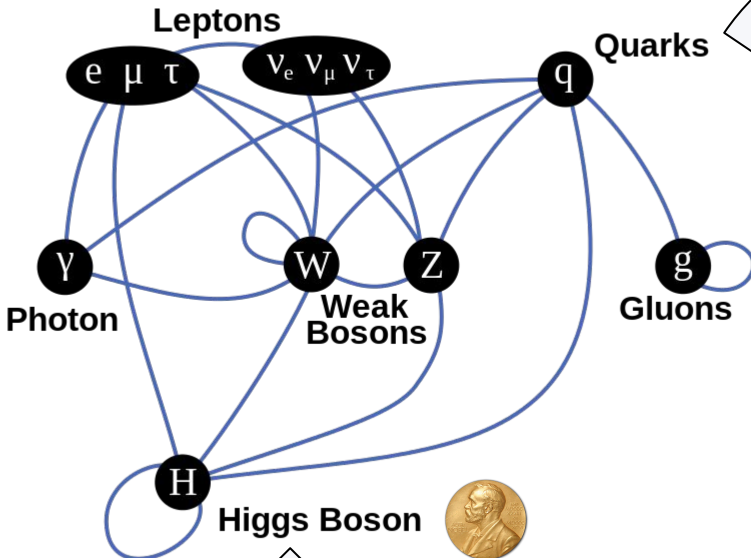
Motivation


MATTER
MOLECULE
ATOM
NUCLEUS
PROTON
NEUTRON
DOWN-TYPE
QUARK
UP-TYPE
QUARK
GLUON

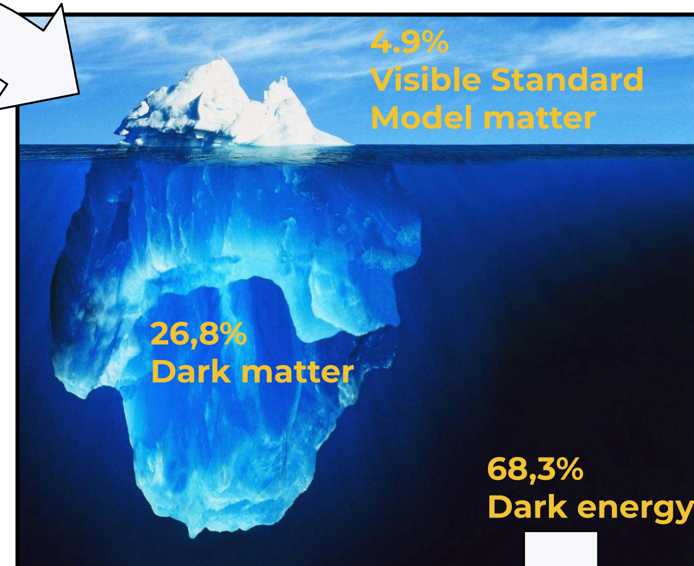
Motivation

Standard Model

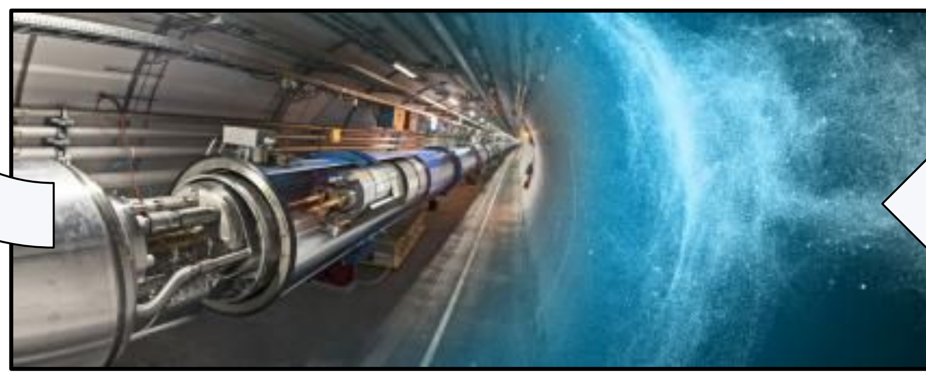
very successful theory



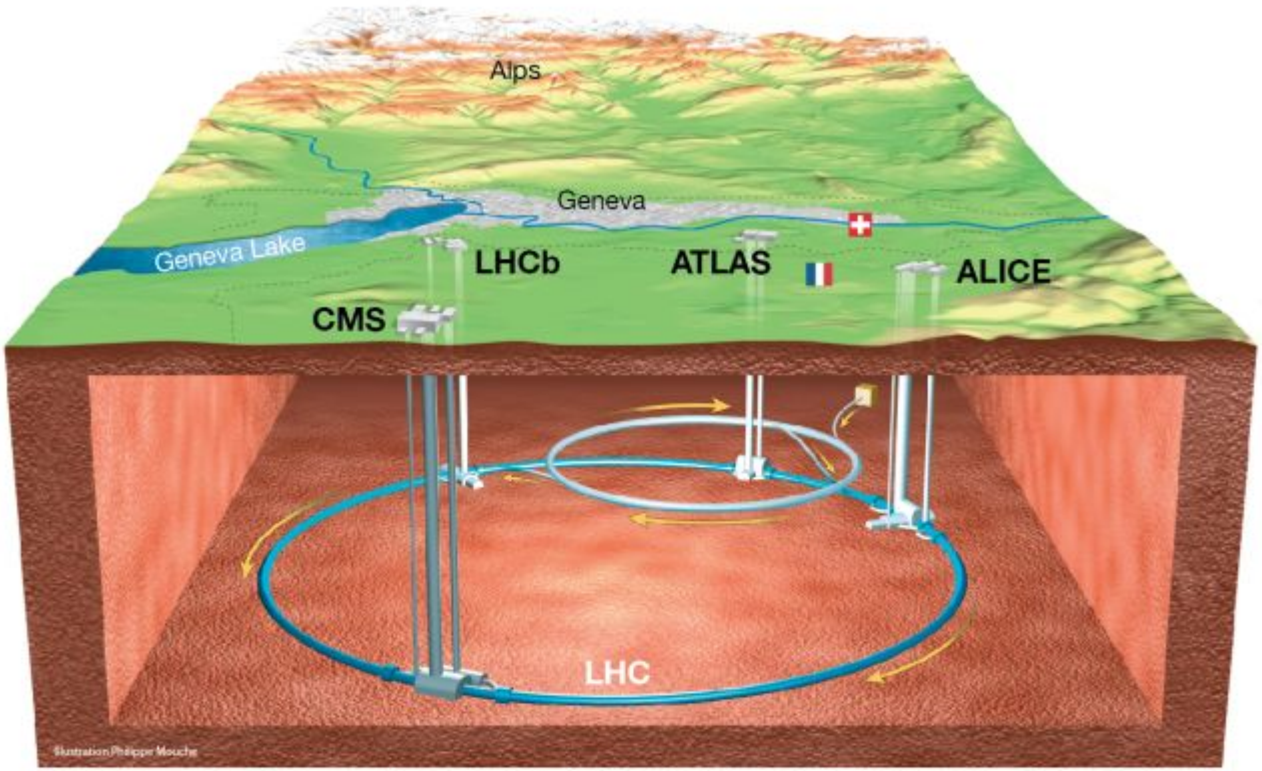
Physics beyond SM must exist



LHC



Large Hadron Collider - 27 km = 27 000 m!



European Strategy for Particle Physics

"Europe's top priority should be the exploitation of the full potential of the LHC"

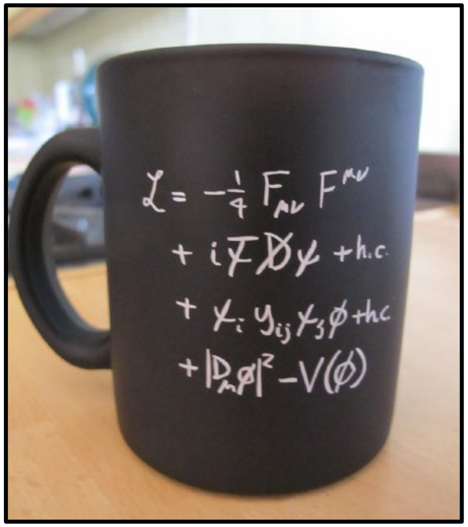
Motivation - Monte Carlo Event Generators (MCEG)

Standard Model

There is a **huge gap** between a one-line formula of a fundamental theory, like the Lagrangian of the SM, and the experimental reality that it implies

Theory

Standard Model Lagrangian



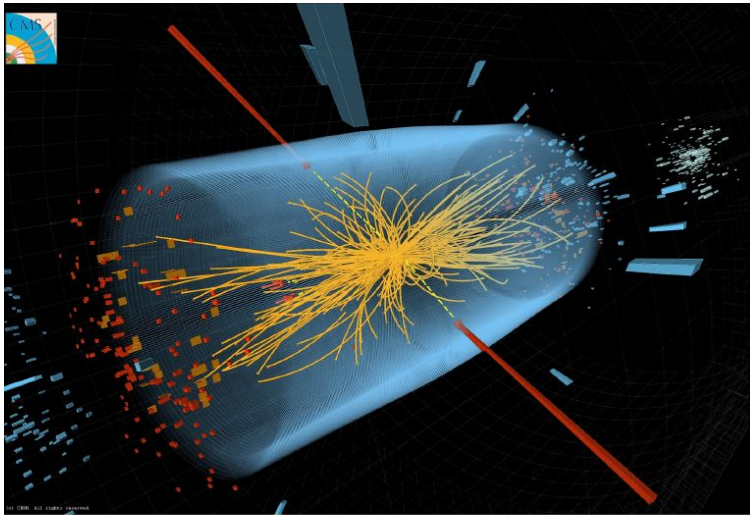
Data makes you smarter

> It doesn't matter how beautiful your theory is, it doesn't matter how smart you are. If it doesn't agree with experiment, it's wrong.

Richard P. Feynman

Experiment

LHC event



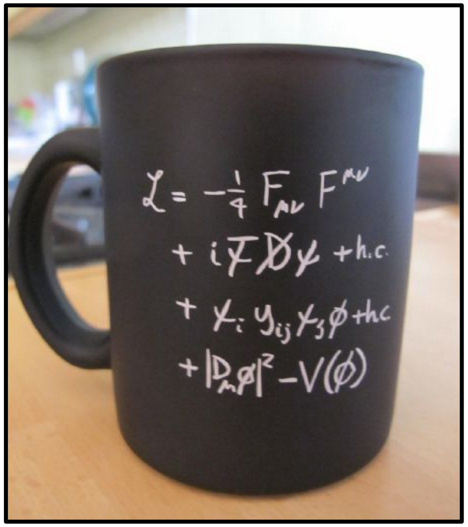
Motivation - Monte Carlo Event Generators (MCEG)

Standard Model

There is a **huge gap** between a one-line formula of a fundamental theory, like the Lagrangian of the SM, and the experimental reality that it implies

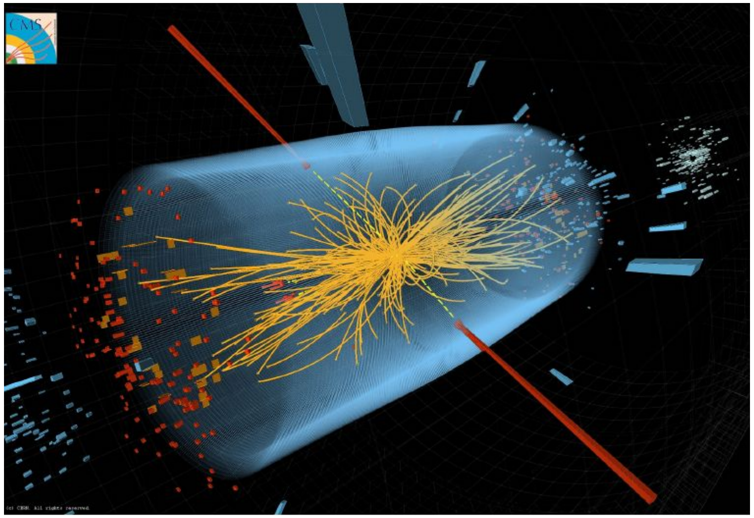
Theory

Standard Model Lagrangian



Experiment

LHC event



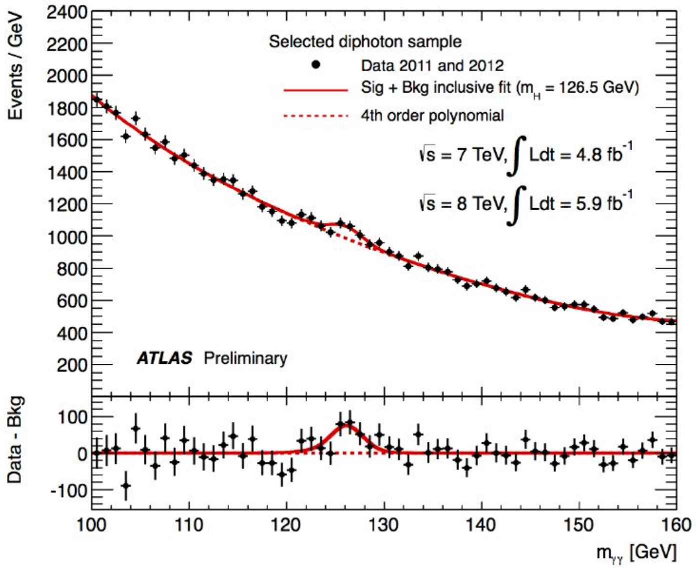
- MC event generators are designed to bridge that **gap**
- “Virtual collider” ⇒ Direct comparison with data

Motivation - Monte Carlo Event Generators (MCEG)

```
--- final:
 17        mu+     -13 [16]
                                 -1.023     7.527     5.229     9.222     0.106
 18      gamma      22 [16]
                                  0.000     0.001     0.001     0.001     0.000
 19        mu-      13 [7]
                                 -1.824   -13.541    22.872    26.643     0.106
 84        pi+     211 [37]
                                 -0.034    -0.014   -69.955    69.955     0.140
 85        pi-    -211 [37]
                                  0.605    -0.133 -1776.296  1776.296     0.140
 86        pi+     211 [42]
                                  0.065     0.602    -1.119     1.280     0.140
 87        pi-    -211 [42]
                                 -0.009    -0.027    -1.185     1.193     0.140
 89        pi-    -211 [44]
                                  0.026     0.394     1.549     1.604     0.140
 90        pi+     211 [45]
                                 -0.018     0.461     1.578     1.650     0.140
102        pi-    -211 [59]
                                  0.367     0.068     3.037     3.063     0.140
103        pi+     211 [59]
                                 -0.363    -0.023     1.478     1.529     0.140
105         K+     321 [60]
                                  0.524    -0.256    56.930    56.935     0.494
108        pi+     211 [62]
                                  0.073    -0.838  1920.309  1920.310     0.140
109        pi-    -211 [62]
                                 -0.183    -0.108   245.969   245.969     0.140
110        pi+     211 [63]
                                 -0.068     0.178     2.109     2.122     0.140
112        pi-    -211 [67]
                                  0.202    -0.606     0.277     0.710     0.140
116        pi-    -211 [69]
                                 -0.239    -0.246    -0.063     0.376     0.140
117        pi+     211 [69]
                                  0.260     0.455     0.092     0.550     0.140
122        pi-    -211 [74]
                                  0.157    -0.080    -0.367     0.430     0.140
125        pi-    -211 [75]
                                  0.160    -0.216     1.711     1.738     0.140
128         p+    2212 [77]
                                 -0.133     1.254   254.182   254.187     0.938
131        pi+     211 [78]
                                 -0.071    -0.068    -1.126     1.139     0.140
133        pi-    -211 [130]
                                  0.333     0.303    -4.928     4.951     0.140
134      gamma      22 [132]
                                 -0.087     0.121    -2.305     2.310     0.000
135      gamma      22 [132]
                                  0.037     0.019    -1.422     1.423     0.000
138        pi+     211 [136]
                                  0.388     0.380    -3.539     3.584     0.140
139        pi-    -211 [136]
                                  0.051     0.279    -1.517     1.550     0.140
```

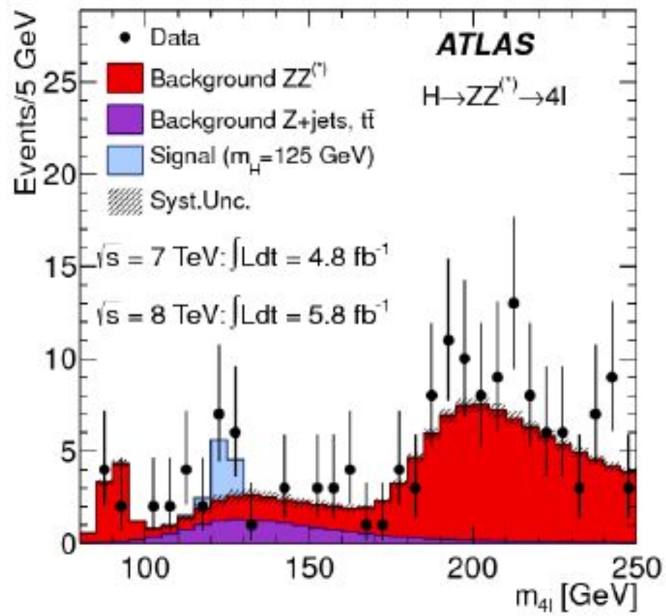
Motivation - Monte Carlo Event Generators (MCEG)

Almost all **HEP measurements and discoveries** in the modern era have **relied on MCEG**, most notably the discovery of the Higgs boson.

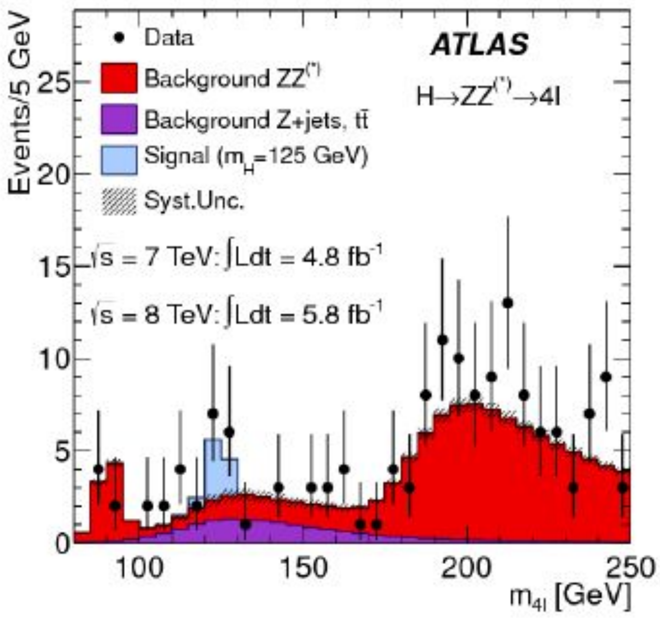


Motivation - Monte Carlo Event Generators (MCEG)

Almost all **HEP measurements and discoveries** in the modern era have **relied on MCEG**, most notably the discovery of the Higgs boson.



Almost all **HEP measurements and discoveries** in the modern era have **relied on MCEG**, most notably the discovery of the Higgs boson.



Last year Pythia 6 manual reached ~**13000** citations!
[JHEP 0605 (2006) 026]

- Main generators: Herwig, Pythia, Sherpa are cited by most papers from LHC experiments.

Published papers by ATLAS, CMS, LHCb: **2252**
Citing MCnet projects: **1888** (**84%**)

Motivation - Monte Carlo Event Generators (MCEG)

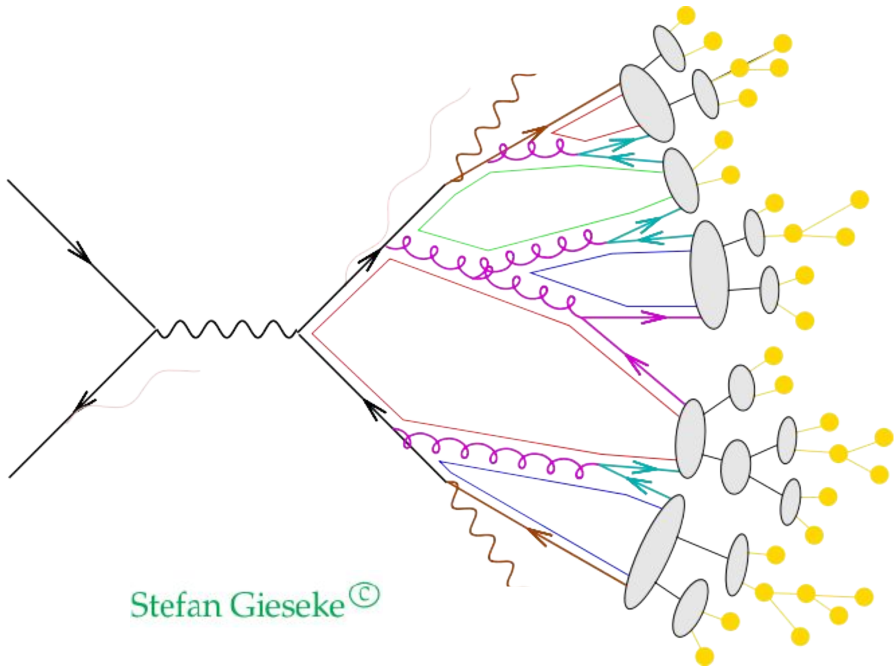
QCD correctly describes strong interactions in each energy range but its complex mathematical structure makes it very difficult to obtain precise predictions (Millennium Prize Problem $1,000,000)

High energy

- **perturbative QCD**
- **in theory we know what to do**
- **in practice very difficult**

Low energy

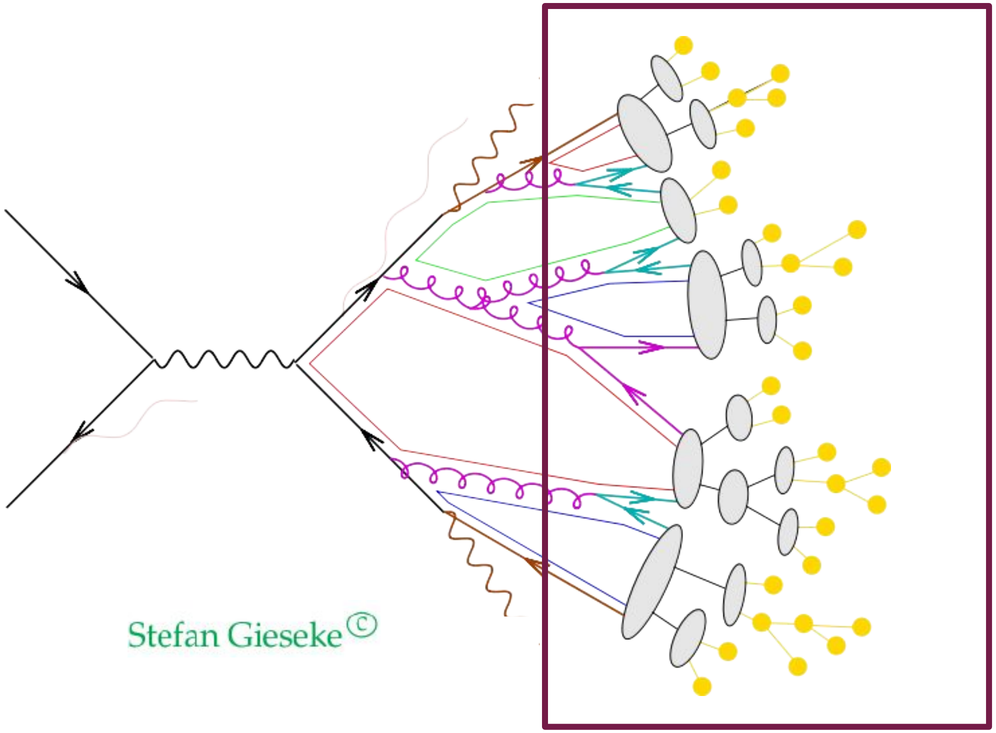
- **non-perturbative QCD**
- **we don't know what to do**
- **phenomenological models (with many free parameters)**



QCD correctly describes strong interactions in each energy range but its complex mathematical structure makes it very difficult to obtain precise predictions (Millennium Prize Problem $1,000,000)

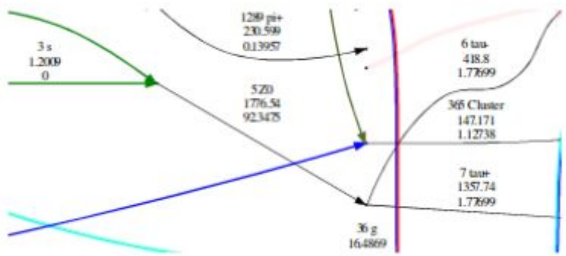
High energy
- **perturbative QCD**
- **in theory we know what to do**
- **in practice very difficult**

Low energy
- **non-perturbative QCD**
- **we don't know what to do**
- **phenomenological models (with many free parameters)**

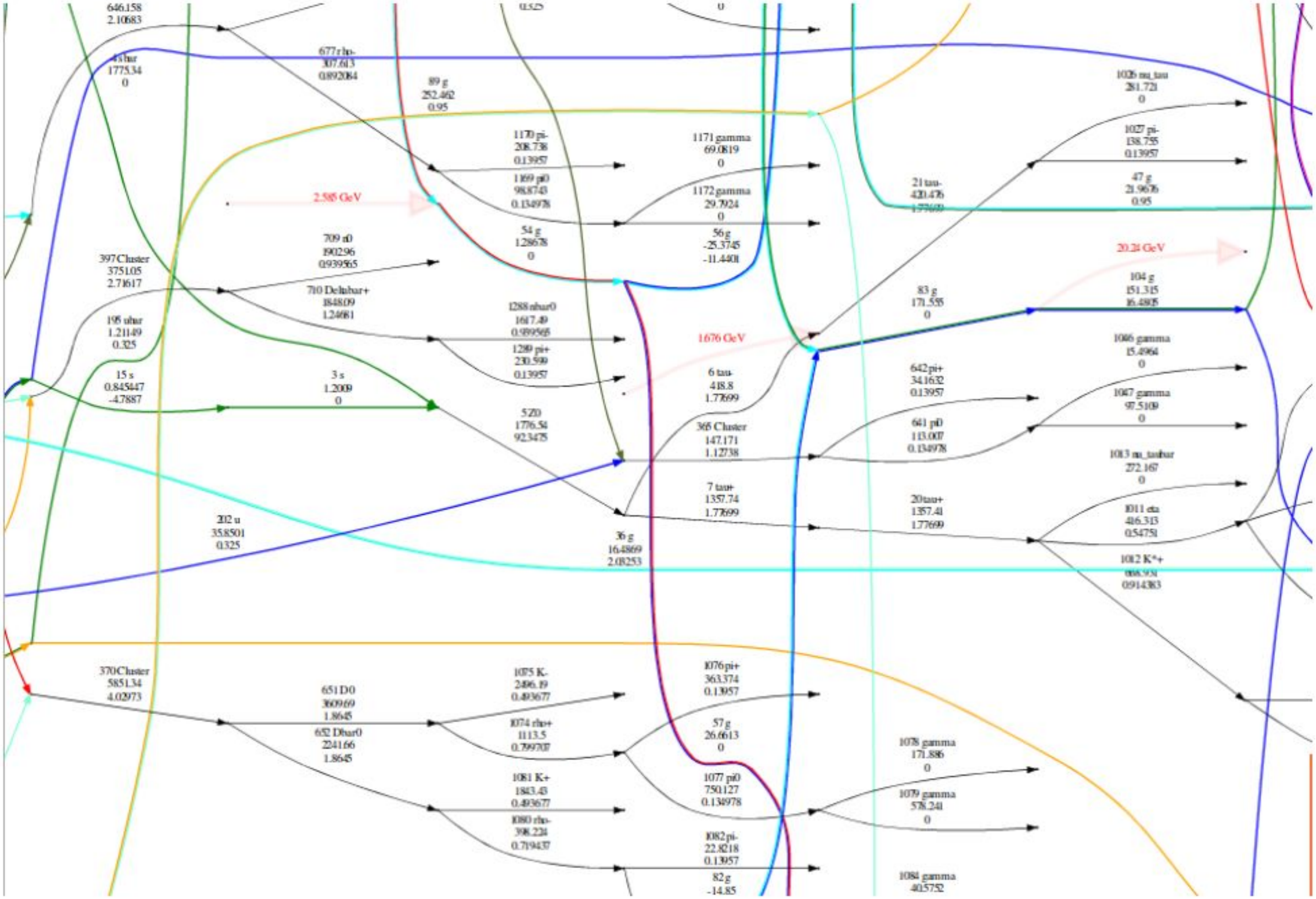


Hadronization:
one of the least understood elements of MCEG

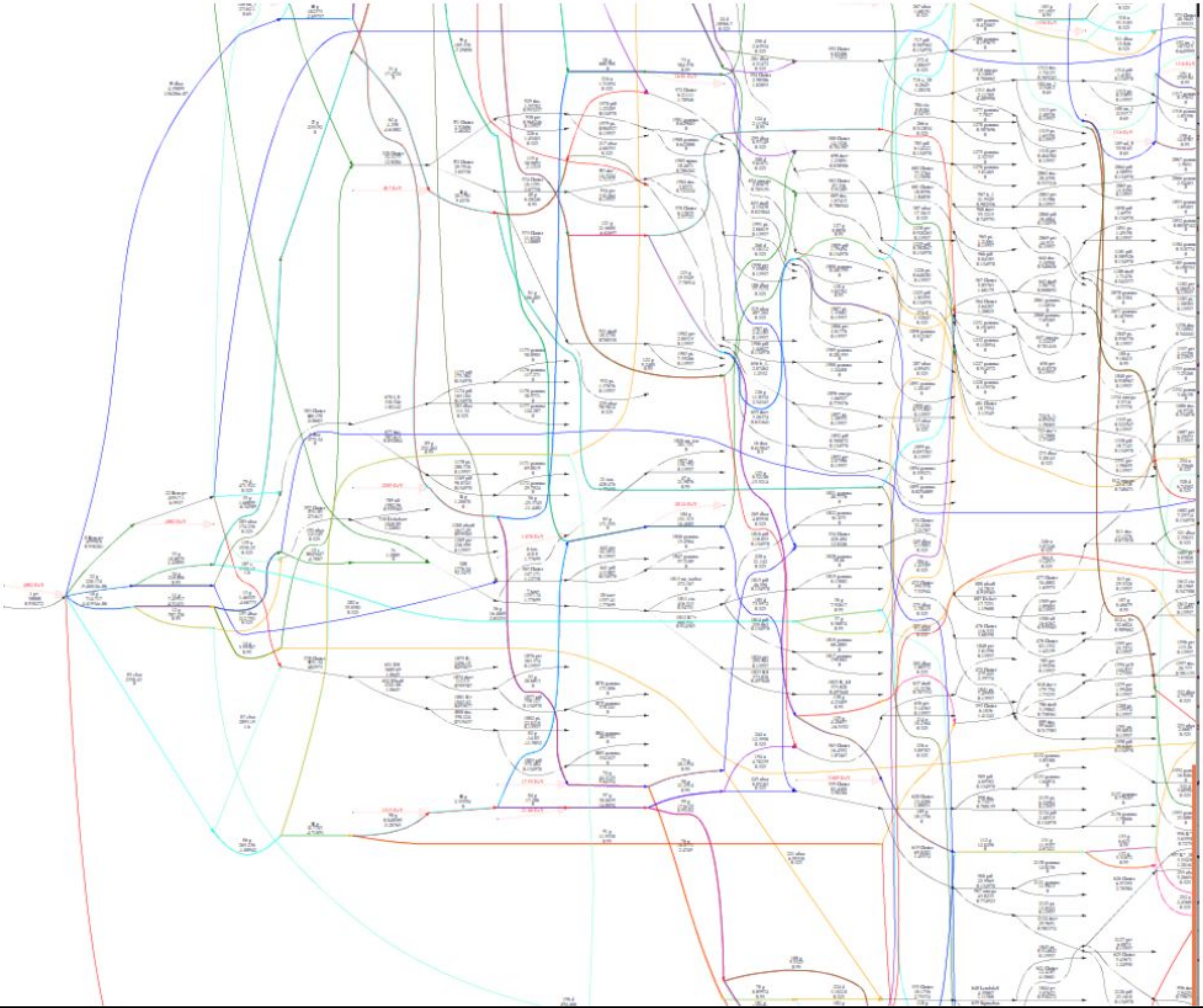
Motivation - Monte Carlo Event Generators (MCEG)



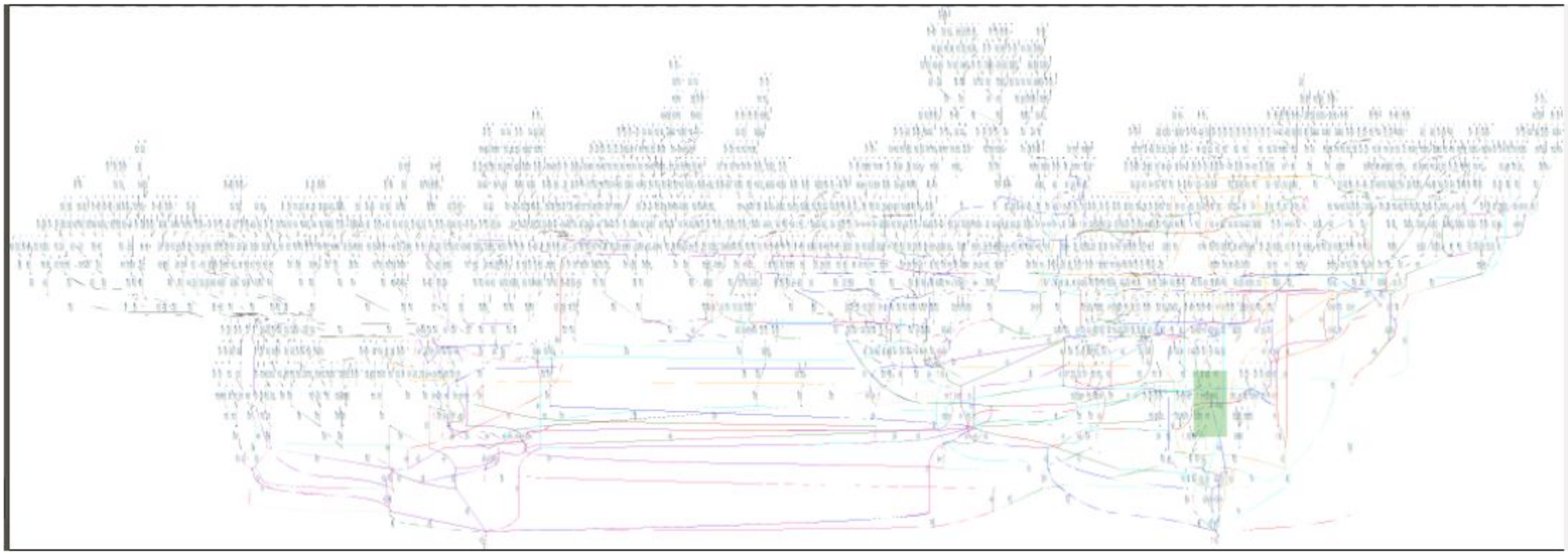
Motivation - Monte Carlo Event Generators (MCEG)



Motivation - Monte Carlo Event Generators (MCEG)



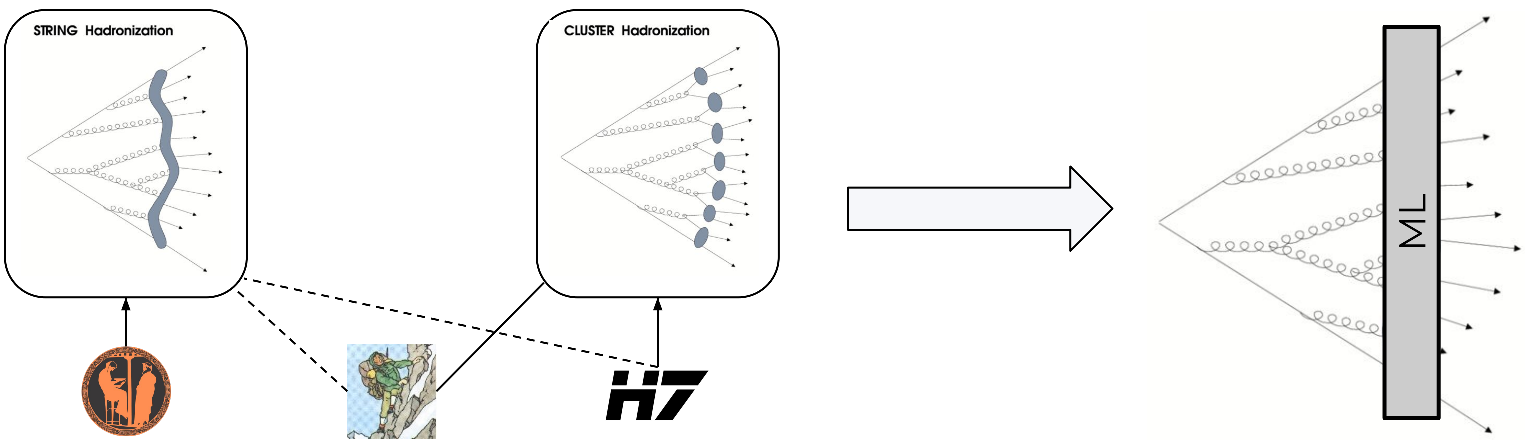
Motivation - Monte Carlo Event Generators (MCEG)



Hadronization:

Early 1980's

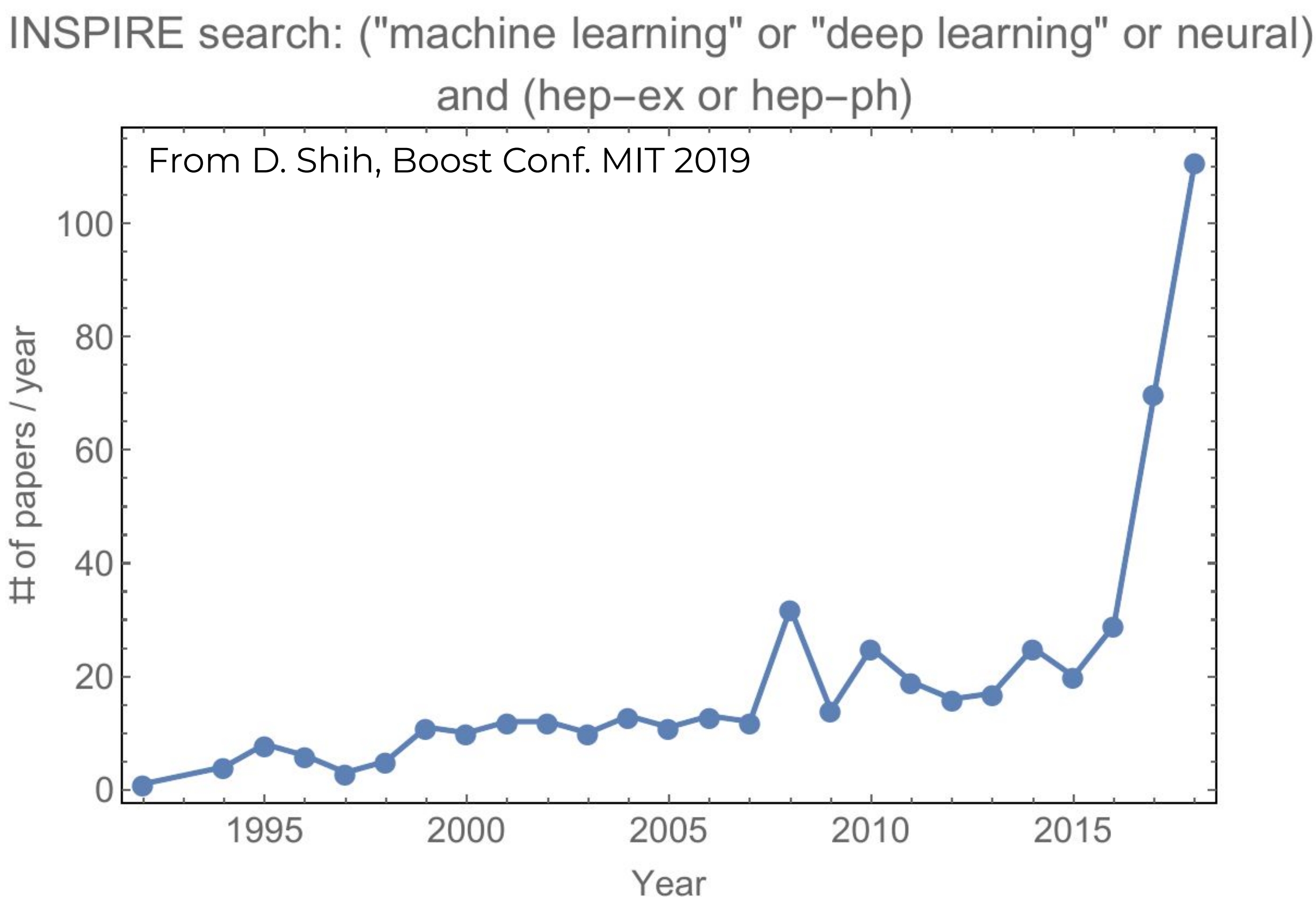
Early 2020's
(lot of progress in ML)



Pythia **Sherpa** **Herwig**

Idea of using Machine Learning (ML) for hadronization.

Pioneering ideas of using Machine Learning (ML) to improve hadronization.
➔ Why ML?



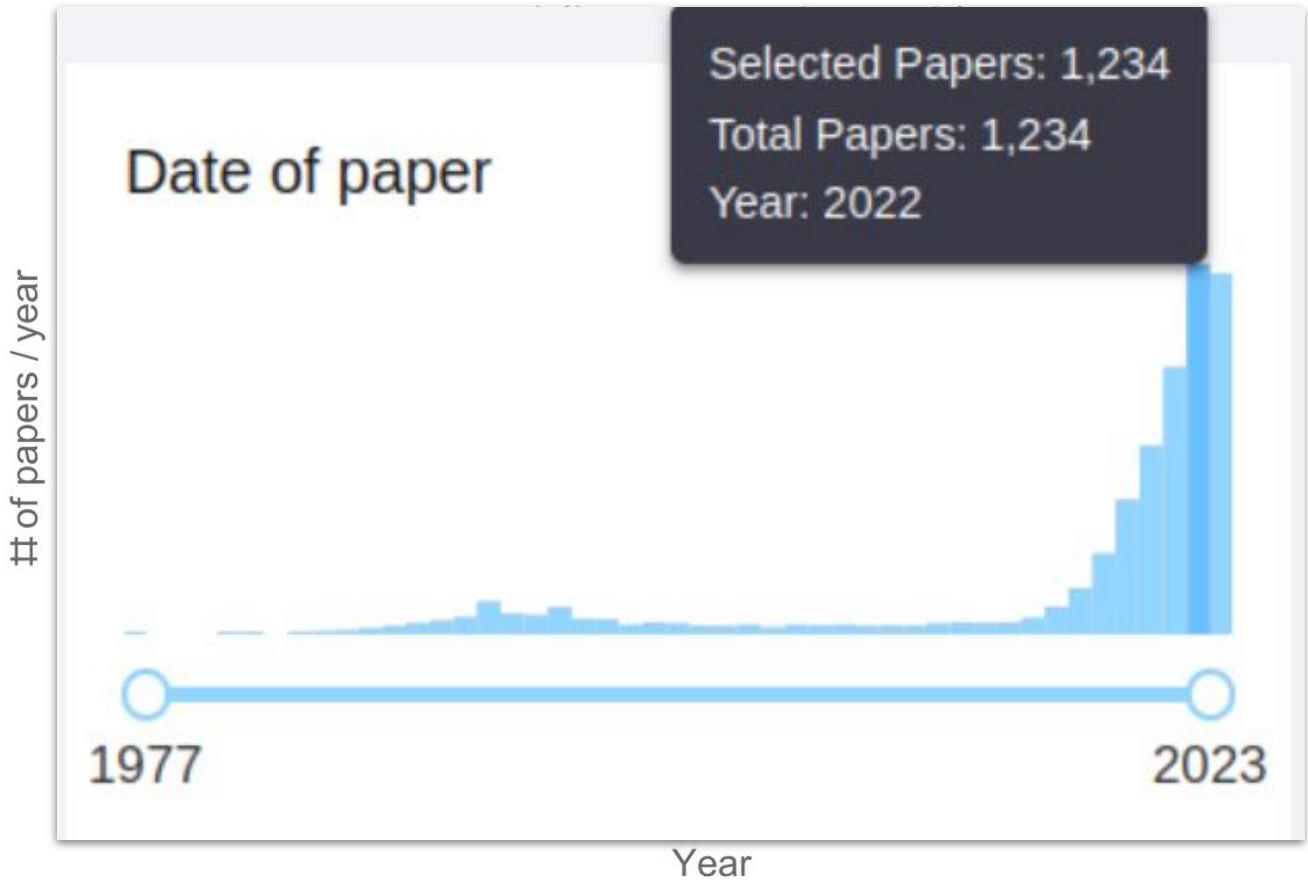
In HEP: Higgs boson [Nature 560], Quark/Gluon jet discrimination, PDF (inverse to hadronization),...

QCD ex Machina - building blocks

Pioneering ideas of using Machine Learning (ML) to improve hadronization.

➔ Why ML?

INSPIRE search: ("machine learning" or "deep learning" or neural)



In HEP: Higgs boson [Nature 560], Quark/Gluon jet discrimination, PDF (inverse to hadronization),...

Motivation - Monte Carlo Event Generators (MCEG)

HEP ML Living Review

Search

GitHub
288 89

Home Recent About Contribute Code of conduct Resources Cite Us

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 GitHub

Expand all sections | Collapse all sections

Reviews

Modern reviews

Specialized reviews

Classical papers

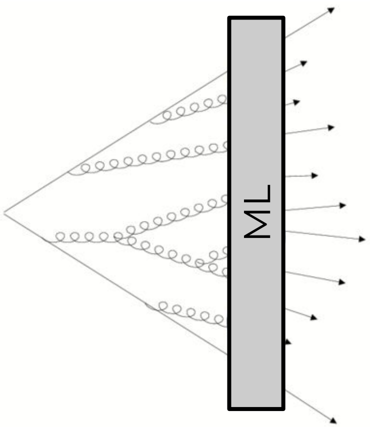
Datasets

Table of contents

- Reviews
  - Modern reviews
  - Specialized reviews
  - Classical papers
  - Datasets
- Classification
  - Parameterized classifiers
  - Representations
  - Targets
  - Learning strategies
  - Fast inference / deployment
- Regression
  - Pileup
  - Calibration
  - Recasting
  - Matrix elements
  - Parameter estimation
  - Parton Distribution Functions (and related)
  - Lattice Gauge Theory
  - Function Approximation
  - Symbolic Regression
  - Monitoring

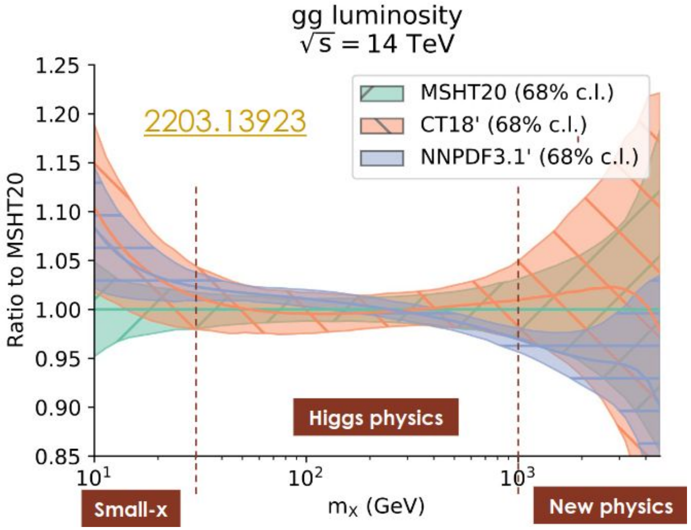
Idea of using Machine Learning (ML) for hadronization.

- Existing hadronization models are highly parameterized functions.
- Hadronization is a fitting problem
  - Can ML hadronization be more flexible to fit the data?
  - Can ML hadronization extract more information from the data? [can accommodate unbinned and high-dimensional inputs]



NNPDF

NNPDF used successfully ML to nonperturbative Parton Density Functions (PDF).
Hadronization is closely related to fragmentation functions (FF) which were considered the counterpart of PDFs.



Recent progress: Machine learning hadronization

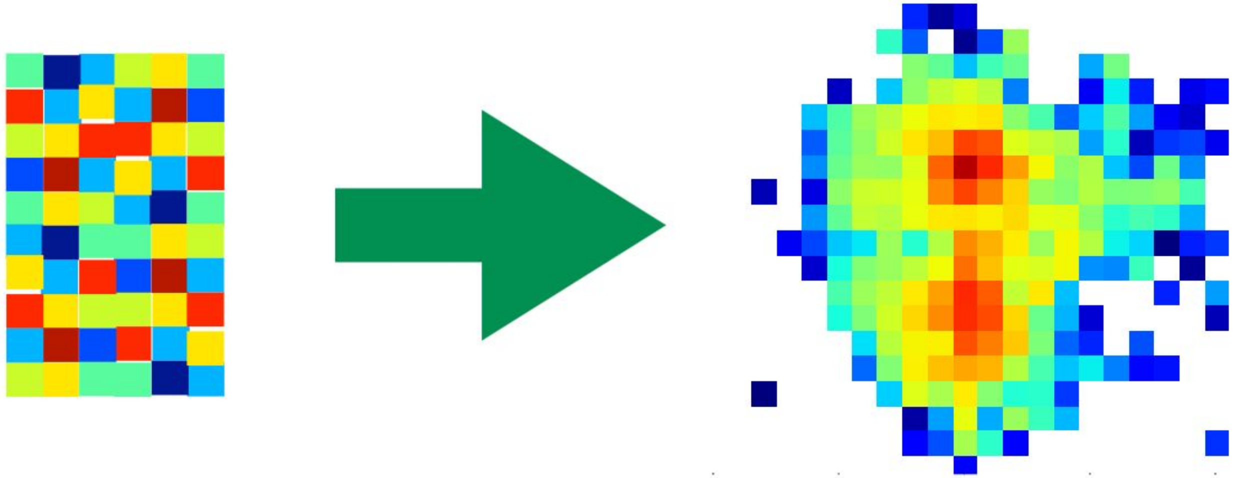
First steps for ML hadronization:

- HADML - [A. Ghosh, Xi. Ju, B. Nachman **AS**, *Phys.Rev.D* 106 (2022) 9]
- MLhad - [P. Ilten, T. Menzo, A. Youssef and J. Zupan, SciPost Phys. 14, 027 (2023)]

	MLhad	HADML
Deep generative model:	Variational Autoencoder	Generative Adversarial Networks
Trained on:	String model	Cluster model
Recent progress:	*“Reweighting Monte Carlo Predictions and Automated Fragmentation Variations in Pythia 8”* [Bierlich, Ilten, Menzo, Mrenna, Szewc, Wilkinson, Youssef, Zupan, 2308.13459] (see Christian’s talk)	*“Fitting a Deep Generative Hadronization Model”* [J. Chan, X. Ju, A. Kania, B. Nachman, V. Sangli and **AS**, JHEP 09 (2023) 084]

What is a deep generative model?

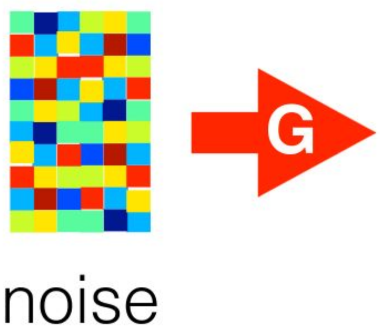
A **generator** is nothing other than a function that maps random numbers to structure.

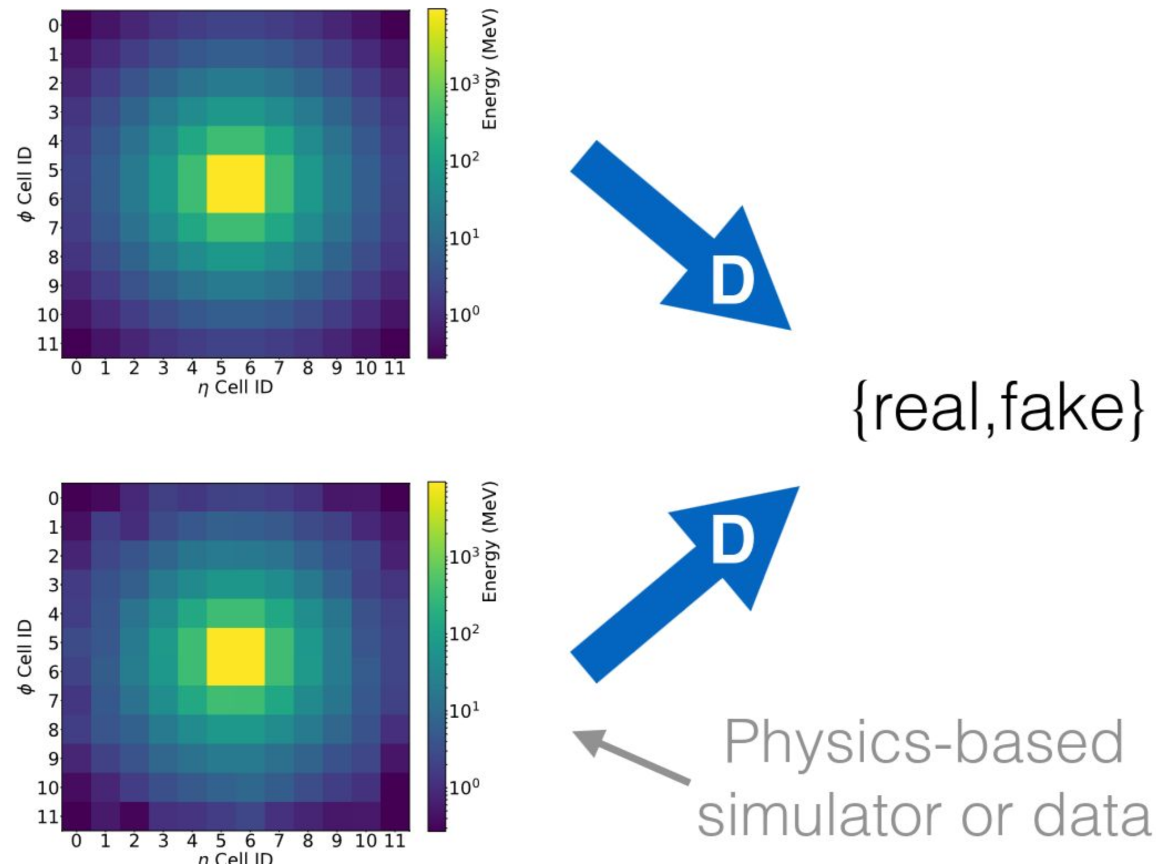


Deep generative models: the map is a deep neural network.

[Goodfellow et al. "Generative adversarial nets". arxiv:1406.2661]

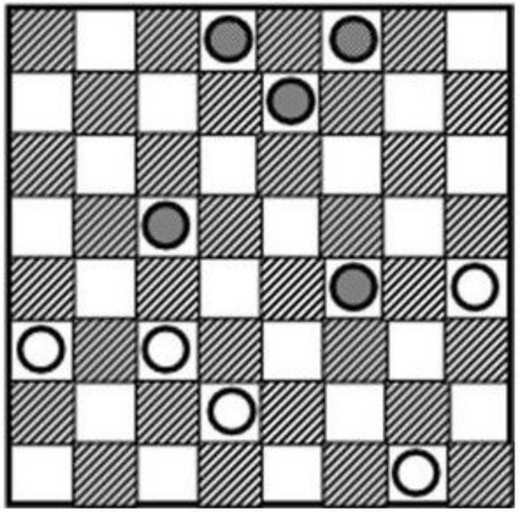
Generative Adversarial Networks (GANs):
A two-network game where one ***maps noise to structure*** *and one* ***classifies images as fake or real.***





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

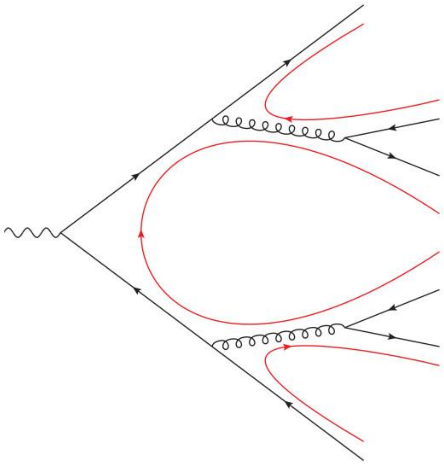
Arthur Lee Samuel (1959) wrote a program that learnt to play checkers well enough to beat him.



- He popularized the term **"machine learning"** in 1959.
- The program chose its move based on a **minimax** strategy, meaning it made the move assuming that the opponent was trying to optimize the value of the same function from its point of view.
- He also had it play thousands of **games against itself** as another way of learning.

The philosophy of the model: use information from perturbative QCD as an input for hadronization.

QCD **pre-confinement** discovered by Amati & Veneziano:

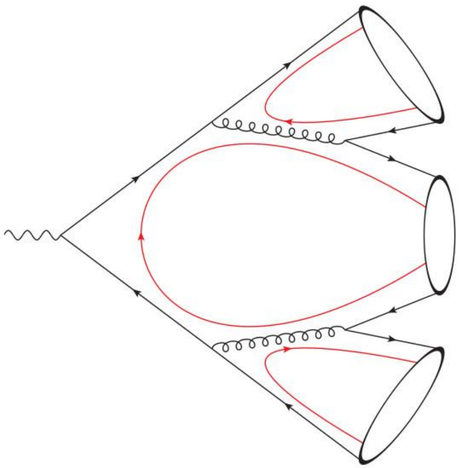


- QCD provide pre-confinement of colour

. . .

The philosophy of the model: use information from perturbative QCD as an input for hadronization.

QCD **pre-confinement** discovered by Amati & Veneziano:

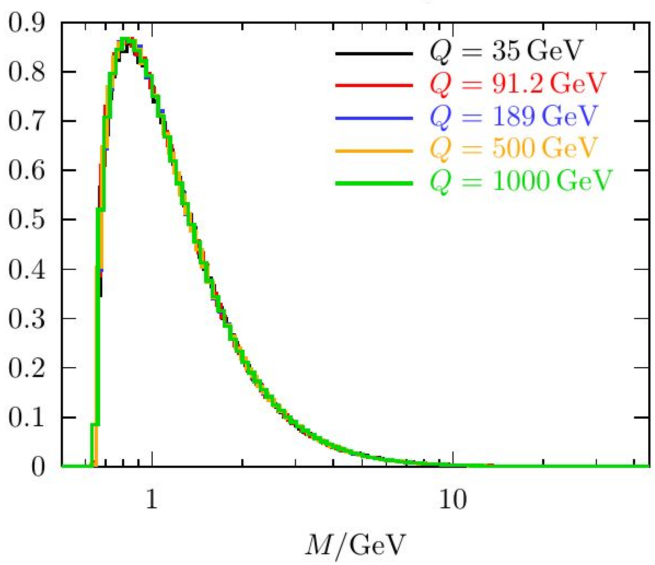


- QCD provide pre-confinement of colour
- Colour-singlet pair end up close in phase space and form highly excited hadronic states, the clusters

. . .

The philosophy of the model: use information from perturbative QCD as an input for hadronization.

QCD **pre-confinement** discovered by Amati & Veneziano:



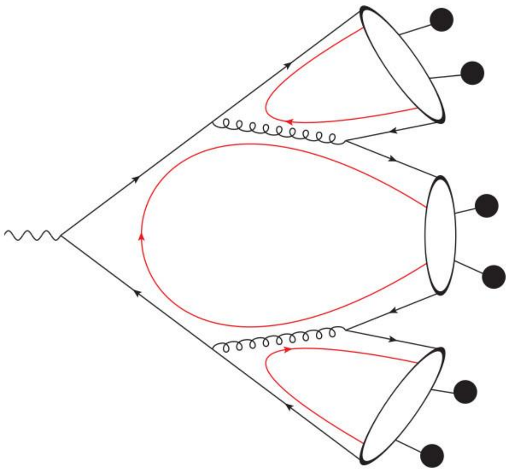
[S. Gieseke, A. Ribon, MH Seymour, P Stephens, B Webber JHEP 0402 (2004) 005]

- QCD provide pre-confinement of colour
- Colour-singlet pair end up close in phase space and form highly excited hadronic states, the clusters
- Pre-confinement states that the spectra of clusters are independent of the hard process and energy of the collision

Cluster hadronization model

The philosophy of the model: use information from perturbative QCD as an input for hadronization.

QCD **pre-confinement** discovered by Amati & Veneziano:

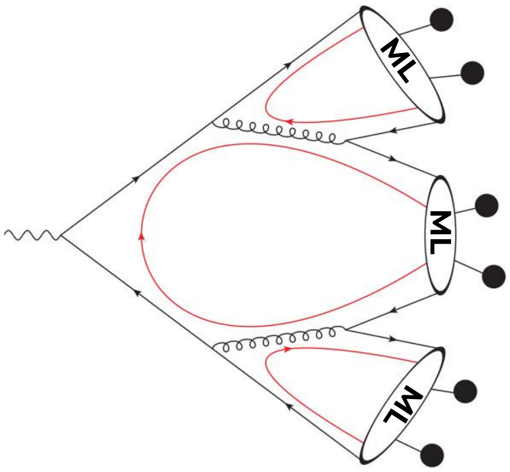


- QCD provide pre-confinement of colour
- Colour-singlet pair end up close in phase space and form highly excited hadronic states, the clusters
- Pre-confinement states that the spectra of clusters are independent of the hard process and energy of the collision
- Peaked at low mass (1-10 GeV) typically decay into 2 hadrons

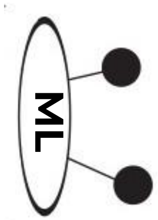
Cluster hadronization model

The philosophy of the model: use information from perturbative QCD as an input for hadronization.

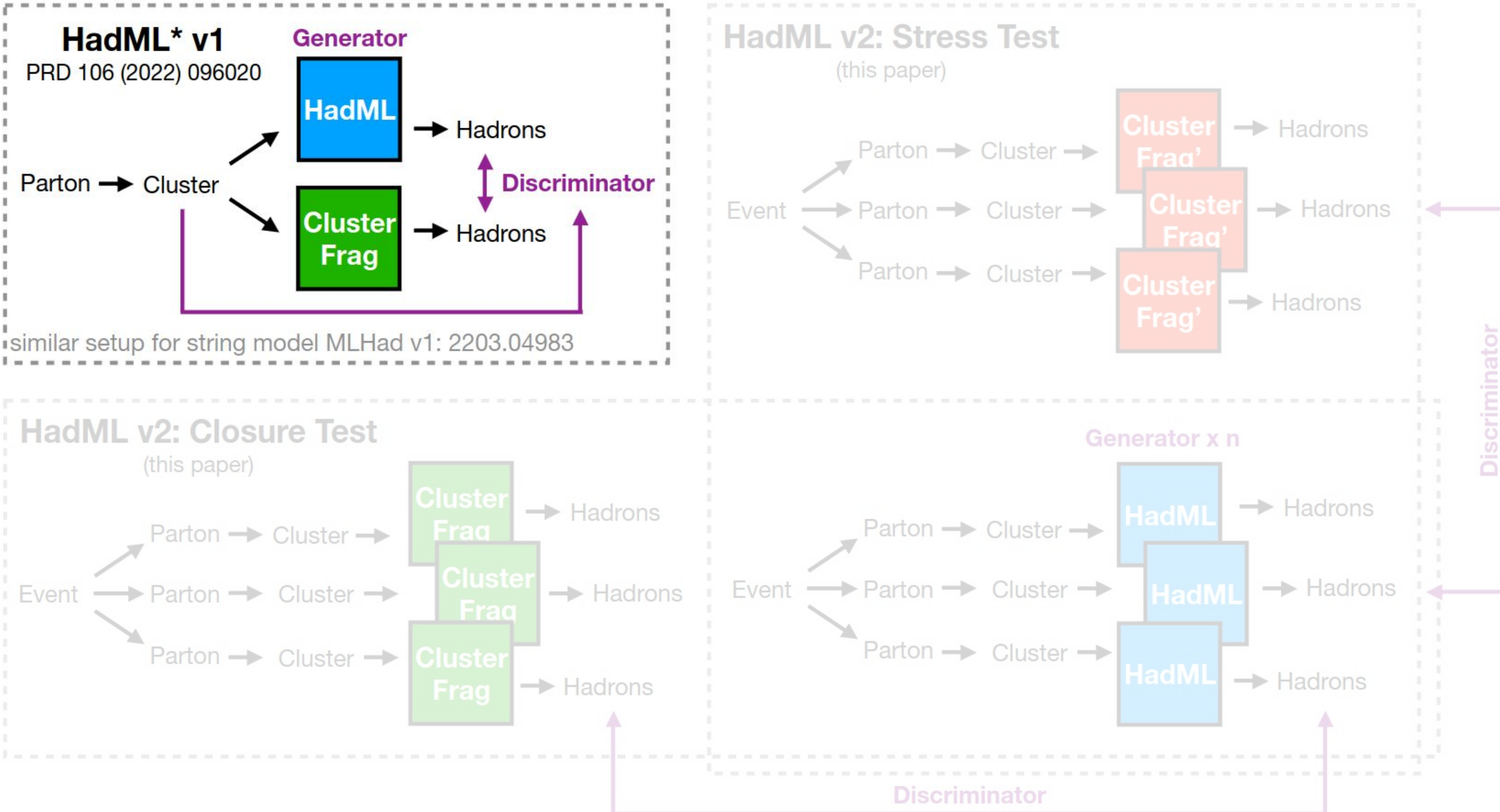
QCD **pre-confinement** discovered by Amati & Veneziano:



- QCD provide pre-confinement of colour
- Colour-singlet pair end up close in phase space and form highly excited hadronic states, the clusters
- Pre-confinement states that the spectra of clusters are independent of the hard process and energy of the collision
- Peaked at low mass (1-10 GeV) typically decay into 2 hadrons
- **ML hadronization**
  1st step: generate kinematics of a cluster decay:

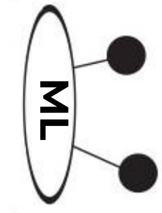


Road map for today



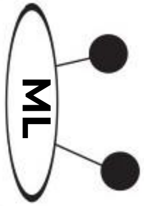
ML hadronization

1st step: generate kinematics of a cluster decay to 2 hadrons



ML hadronization

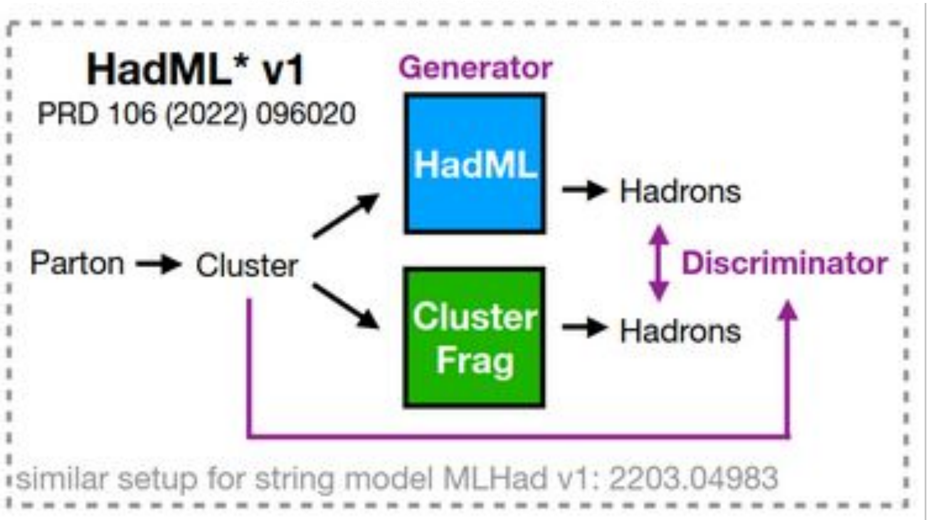
1st step: generate kinematics of a cluster decay to 2 hadrons



How?

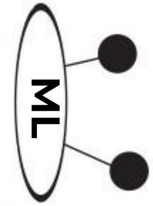
We have a conditional GAN, with cluster 4-vector input and two hadron 4-vector outputs.

Generative Adversarial Net



ML hadronization

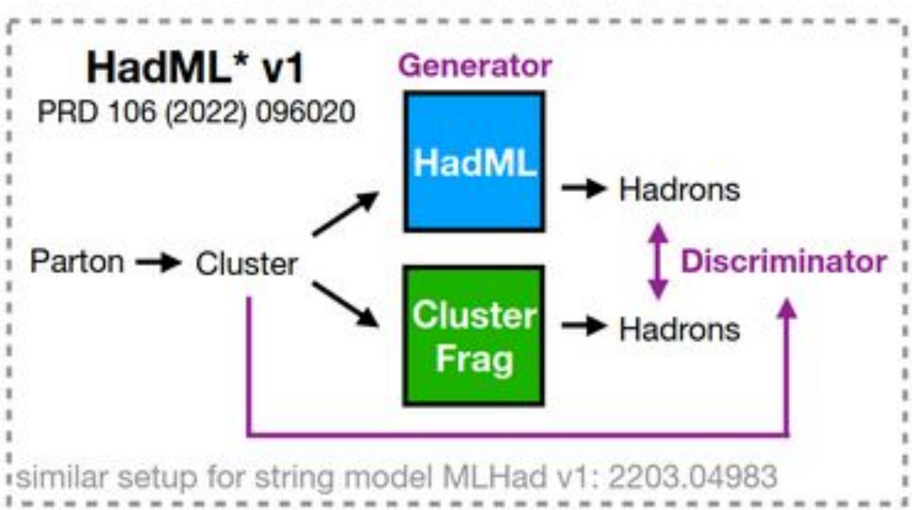
1st step: generate kinematics of a cluster decay to 2 hadrons



How?

Generative Adversarial Net

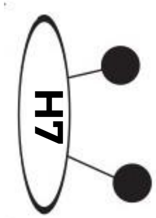
We have a conditional GAN, with cluster 4-vector input and two hadron 4-vector outputs.



Training data:

e^+e^- collisions at $\sqrt{s} = 91.2$ GeV

Cluster $(E,\ p_x,\ p_y,\ p_z)$

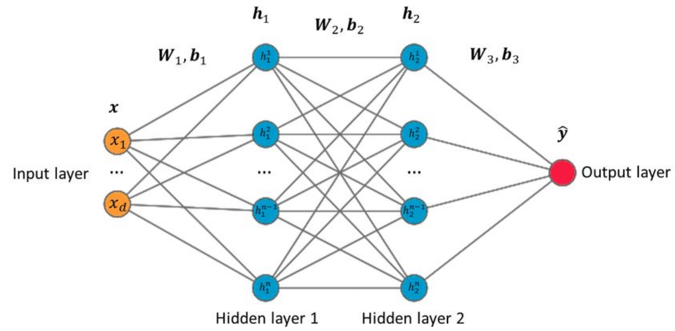


$\pi^0(E,\ p_x,\ p_y,\ p_z)$

$\pi^0(E,\ p_x,\ p_y,\ p_z)$

Simplification: considering only pions and generating two angles in the cluster rest frame.

Generator and the Discriminator are composed of two-layer perceptron
(each a fully connected, hidden size 256, a batch normalization layer, LeakyReLU activation function)



Generator

Input

Cluster $(E,\ p_x,\ p_y,\ p_z)$ and 10 noise features sampled from a Gaussian distribution

Output (in the cluster frame)

ϕ - polar angle
θ - azimuthal angle

we reconstruct the four vectors of the two outgoing hadrons

Discriminator

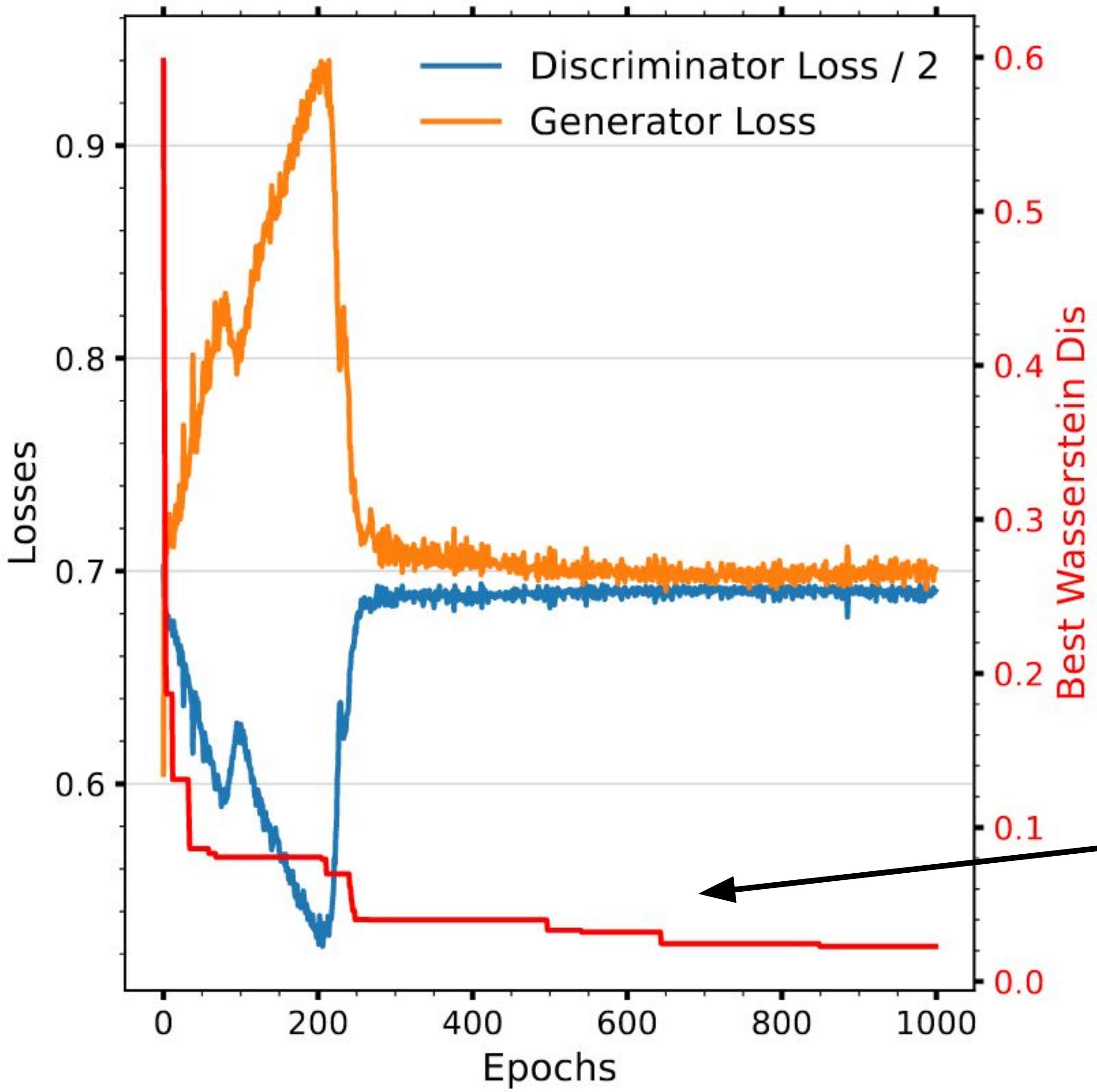
Input

ϕ and θ labeled as signal (generated by Herwig) or background (generated by Generator)

Output

Score that is higher for events from Herwig and lower for events from the Generator

Training HADML v1



We have a conditional GAN, with cluster 4-vector input and two hadron 4-vector outputs.

Simplification:
considering only pions and generating two angles in the cluster rest frame.

This is a typical learning curve for GAN training

Integration into Herwig

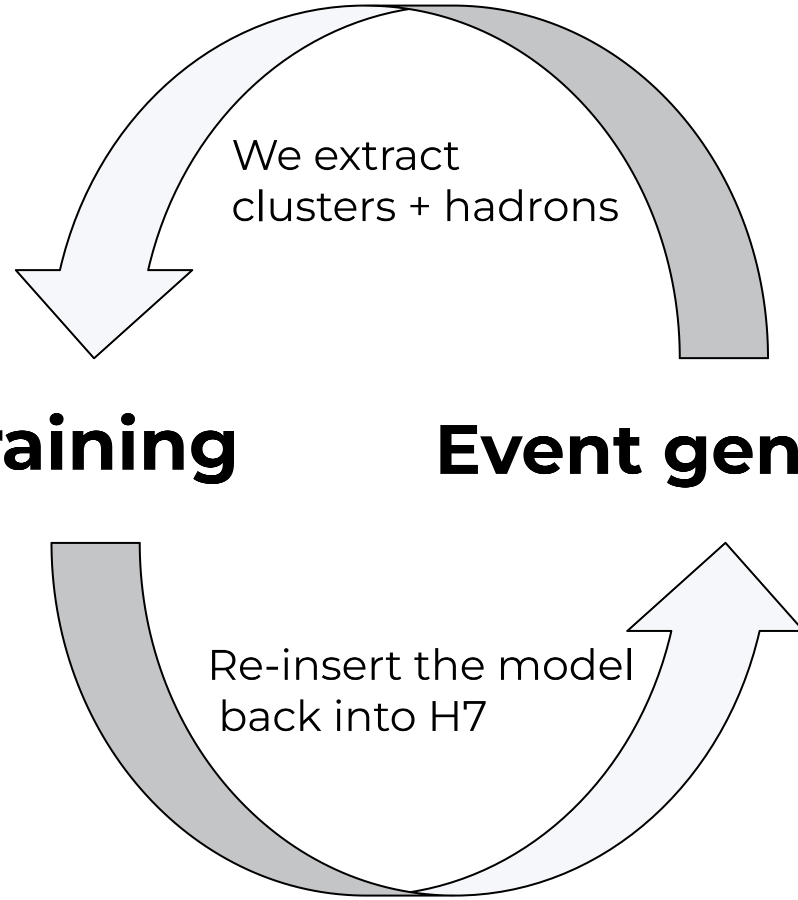
GPU

TensorFlow

python

CPU

H7



Training

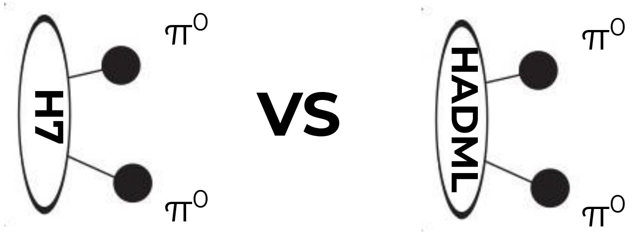
Event generation

ONNX RUNTIME

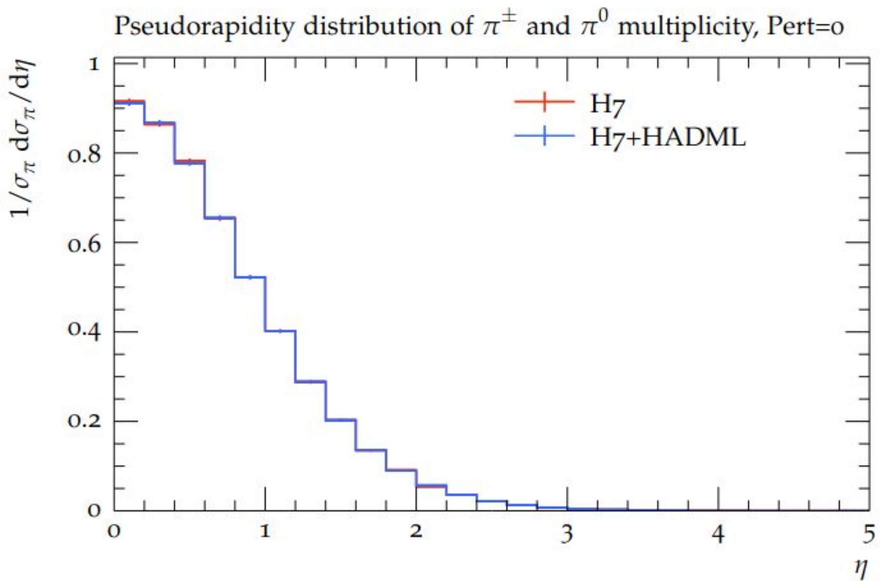
This then allows us to run a full event generator and produce plots

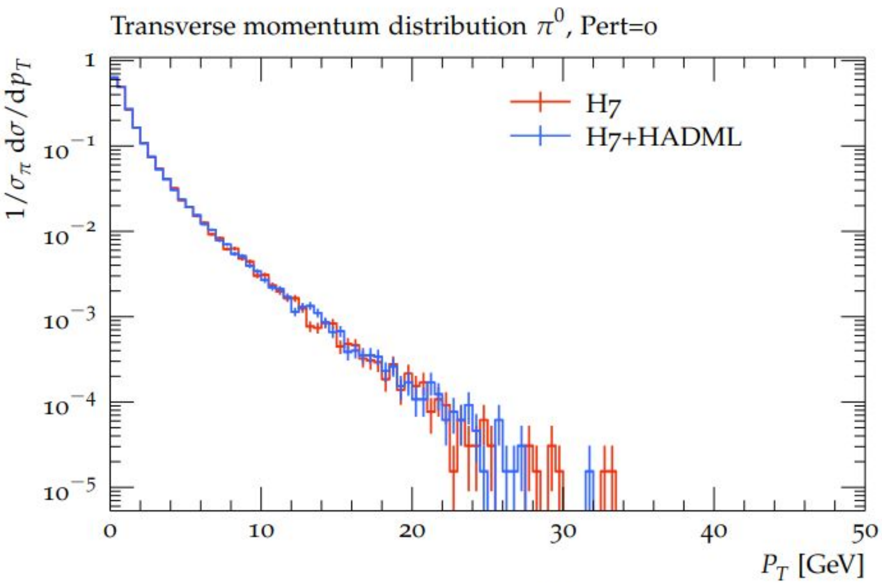
Low-level Validation
(similar to training data)

e^+e^- collisions at
$\sqrt{s} = 91.2$ GeV

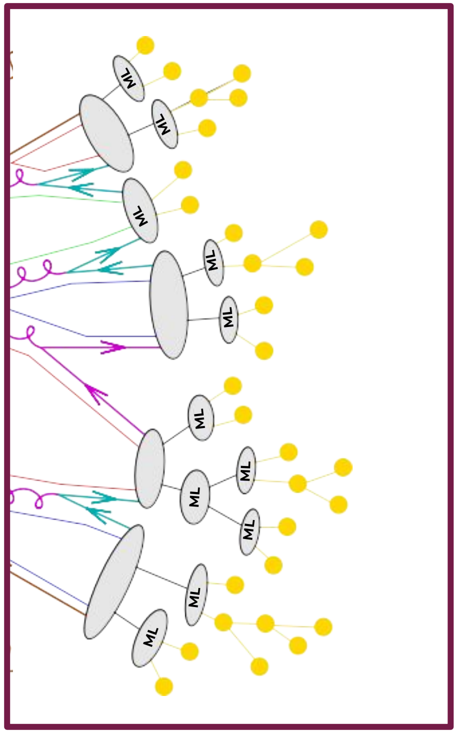


π^0 kinematic variables

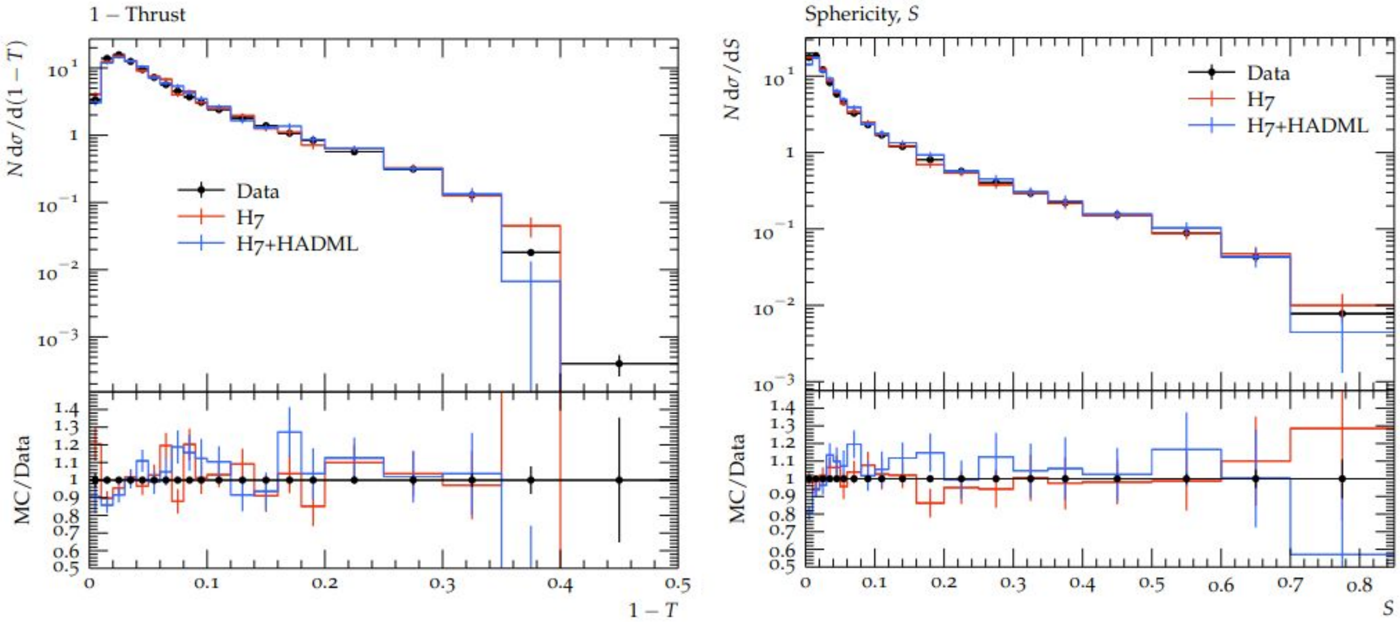




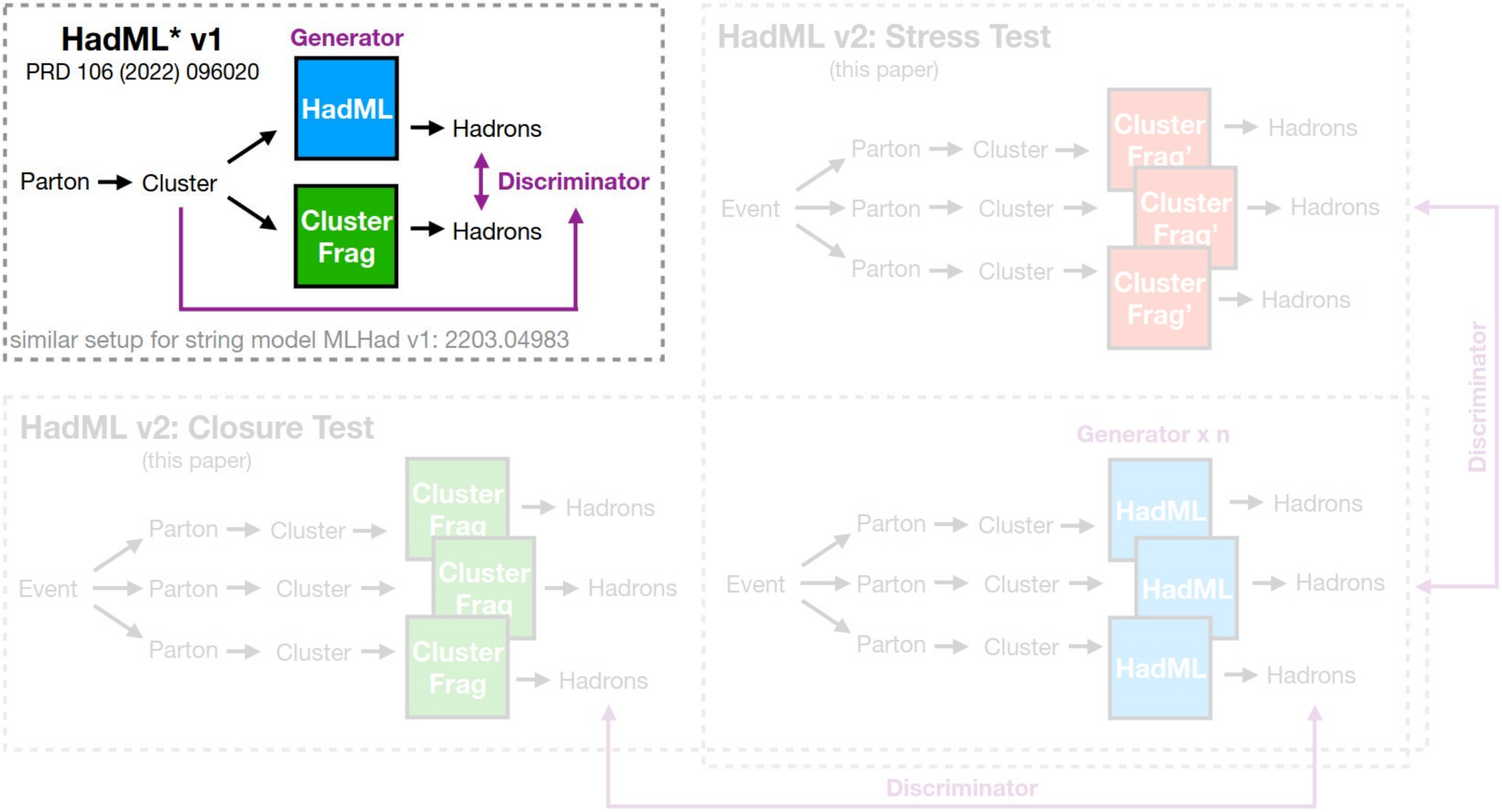
With a “full” model, we can compare directly to data!



LEP DELPHI Data

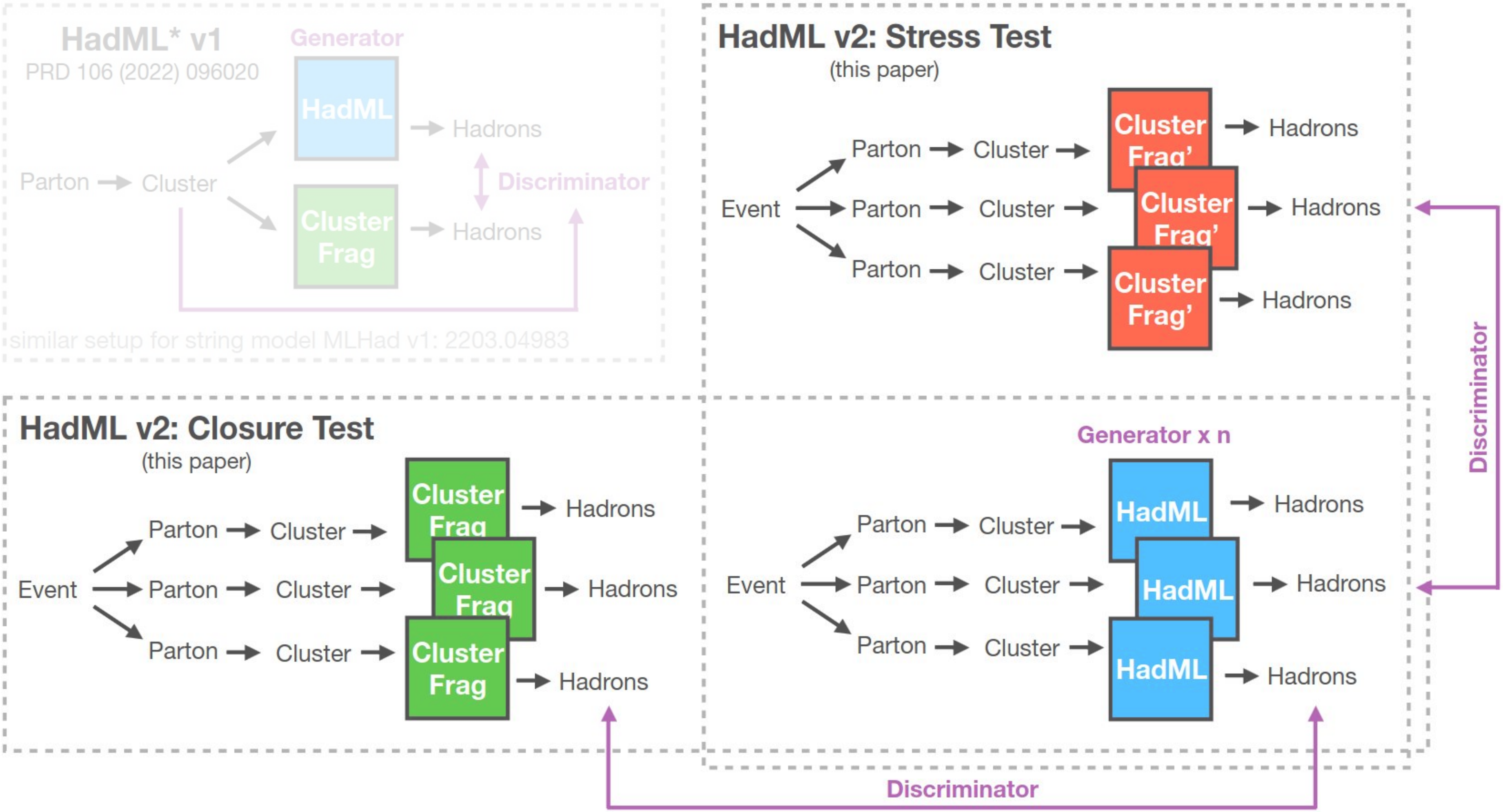


N.B. we have trained on H7, so we don’t expect to be any better than it at modeling the data.



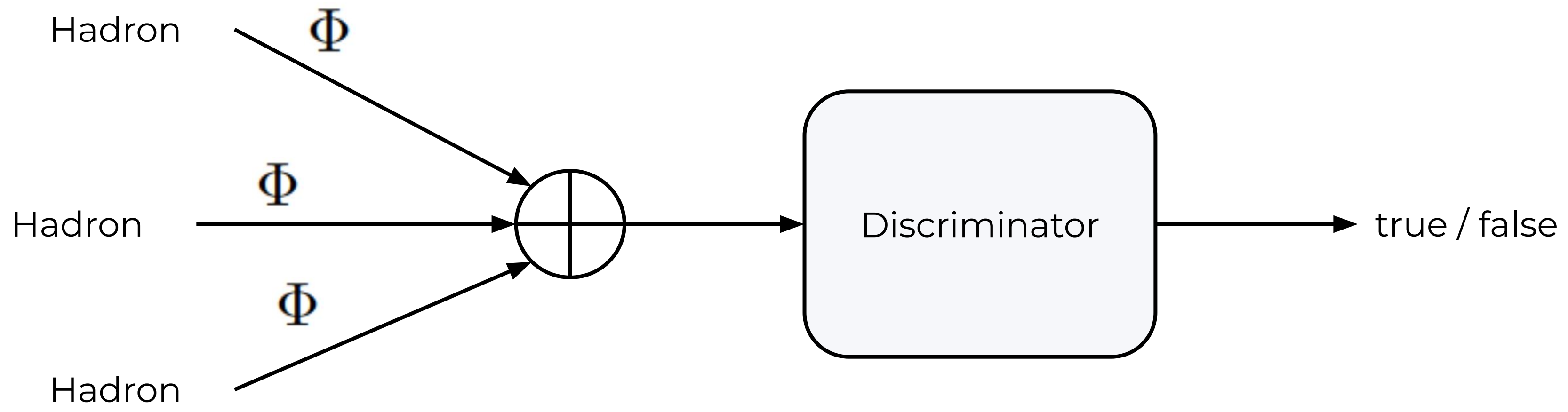
HadML* v1
PRD 106 (2022) 096020
Generator
HadML
Hadrons
Parton
Cluster
Discriminator
Cluster Frag
Hadrons
similar setup for string model MLHad v1: 2203.04983
HadML v2: Stress Test
(this paper)
Event
Parton
Cluster
Cluster Frag'
Hadrons
Discriminator
HadML v2: Closure Test
(this paper)
Cluster Frag
Generator x n
HadML
Discriminator

Road map for today



Protocol for fitting a deep generative hadronization model in a realistic data setting, where we only have access to a set of hadrons in data.

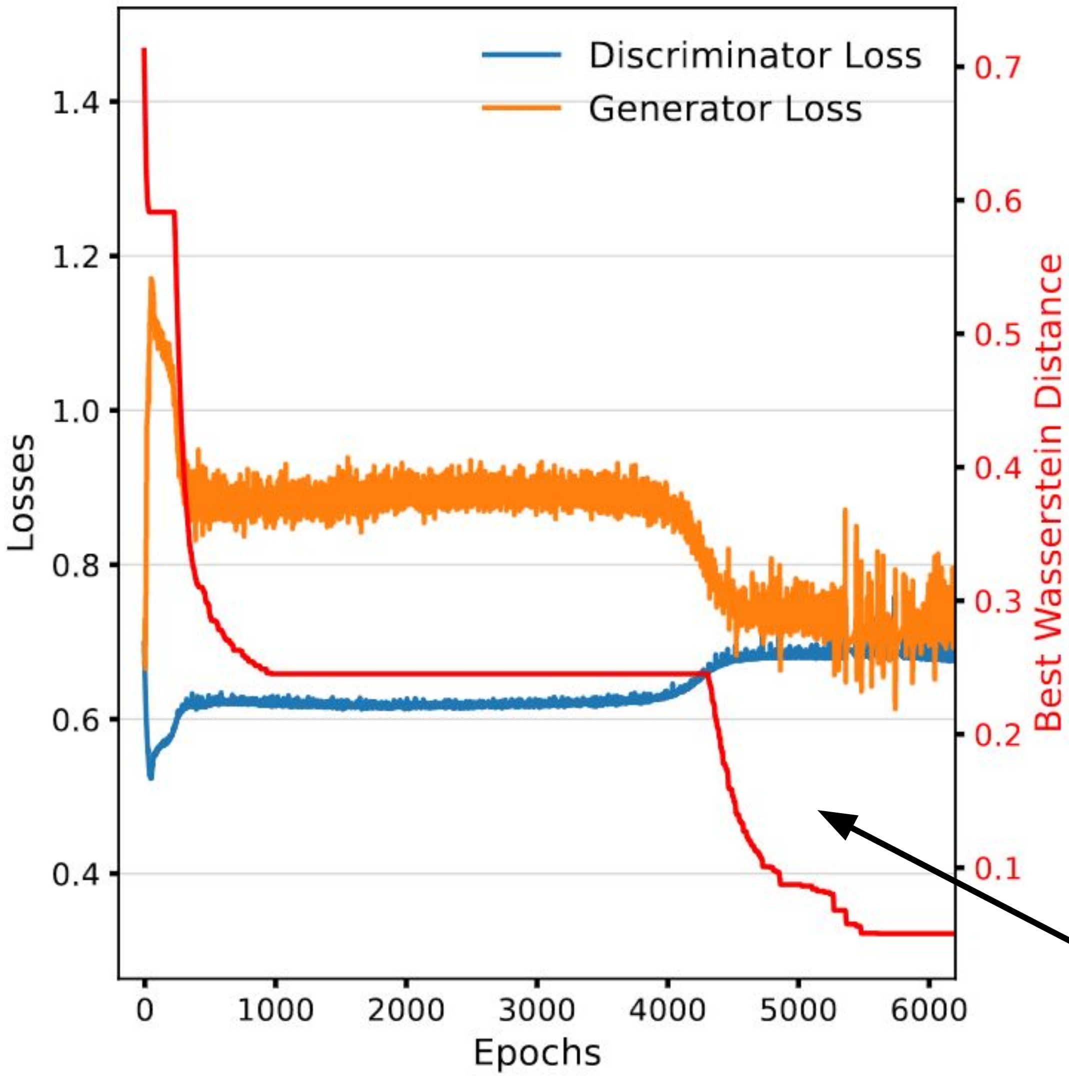
Discriminator HadML v2



The discriminator function is modified, we parameterize is as a Deep Sets model

$$D_E(x) = F\left(\frac{1}{n}\sum_{i=1}^{n}\Phi\left(h_i, \omega_{D_\Phi}\right), \omega_F\right)$$

invariant under permutations of hadrons

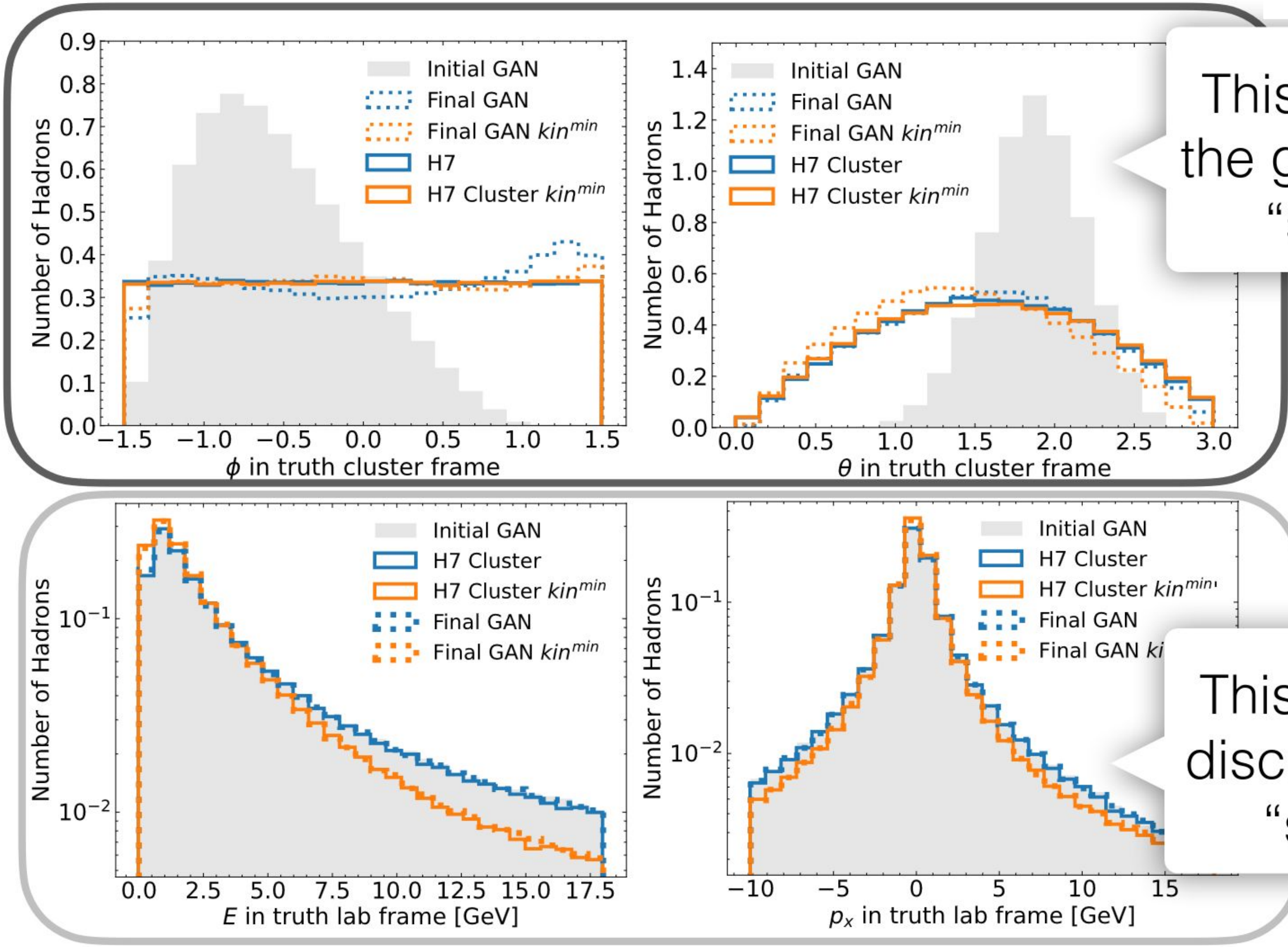


Now, the generator is local (per cluster), but the discriminator is global (whole event).

Discriminator is a permutation-invariant architecture called Deep Sets.

Simplification only Pions

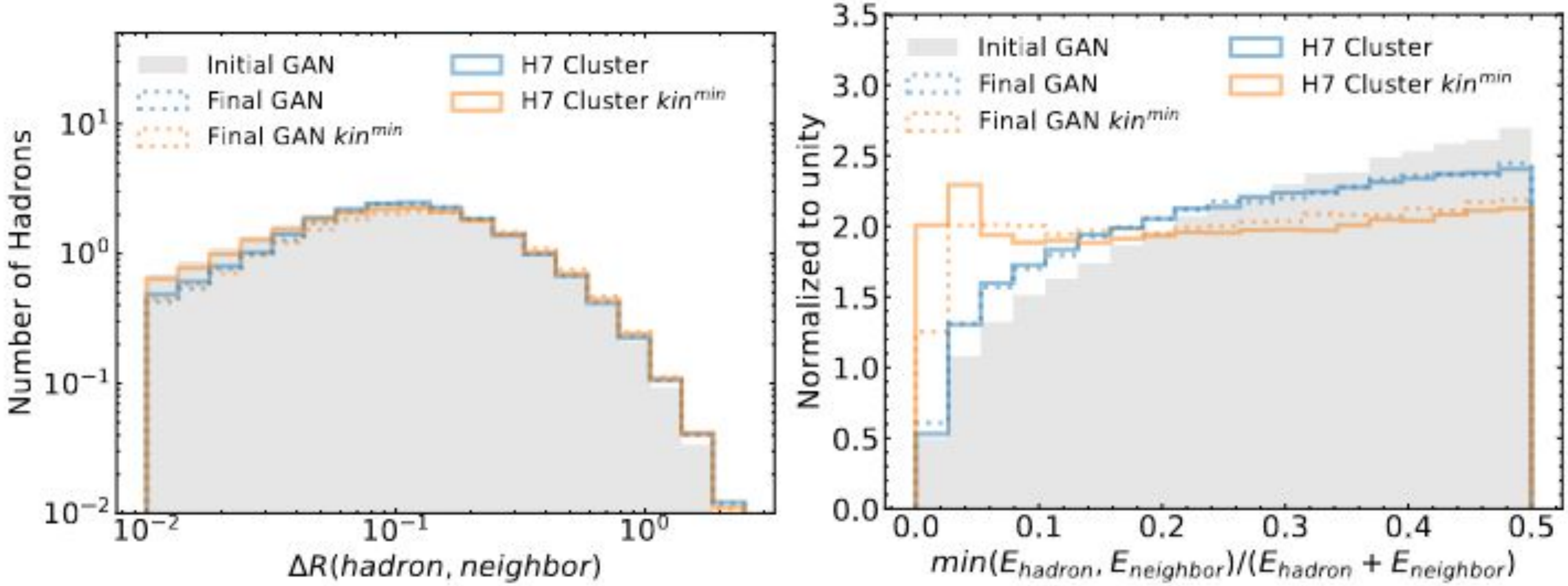
Still works !



This is what the generator “sees”

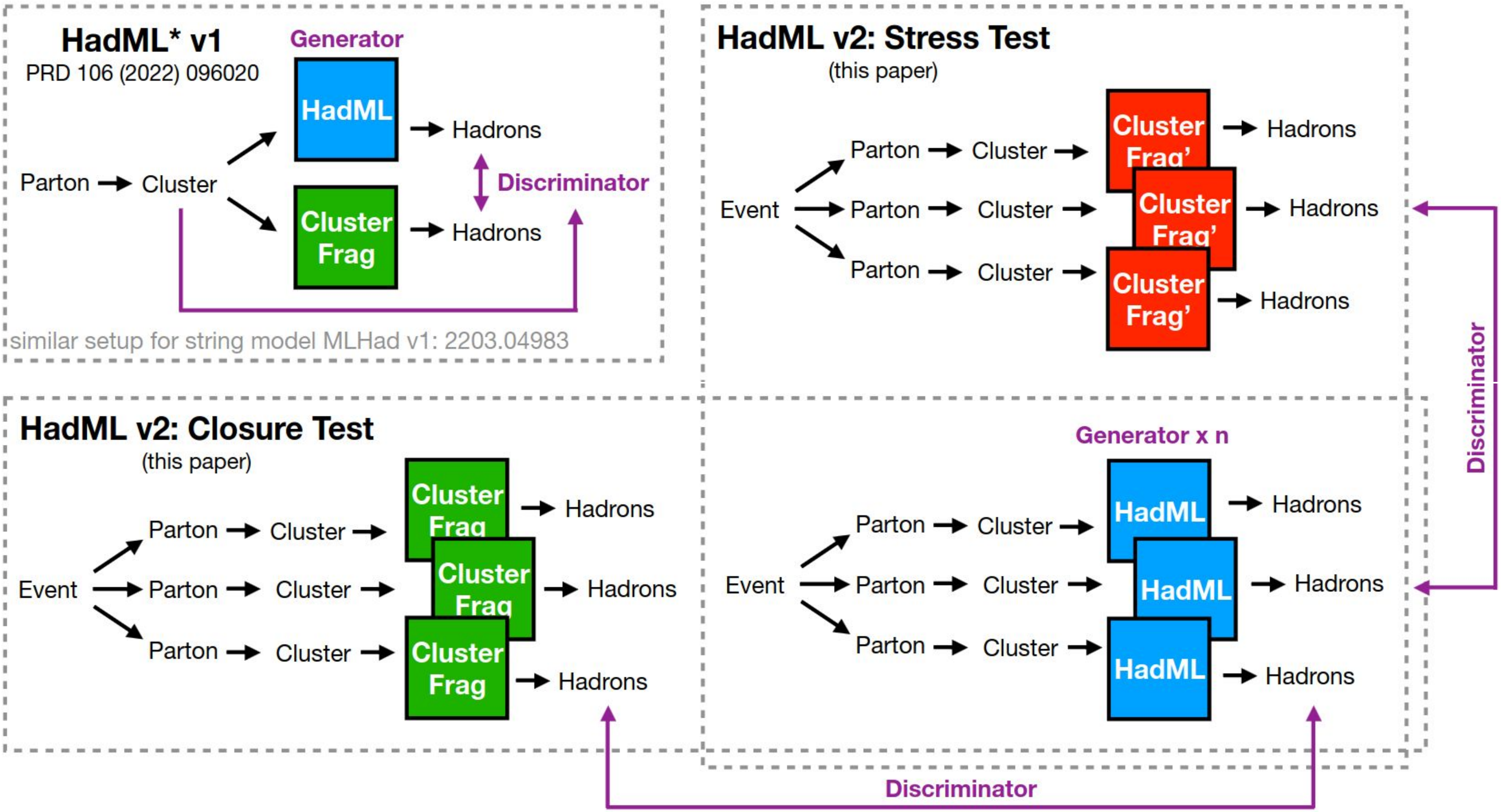
This is what discriminator “sees”

Performance: going beyond inputs and outputs



MINIMAL $\Delta R^2 = \Delta\phi^2 + \Delta\eta^2$

Summary

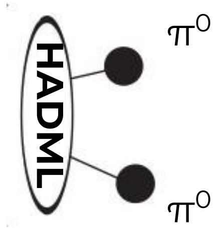


A key advantage of this fitting protocol over other methods is that it can accommodate unbinned and high-dimensional inputs.

The approach could also be used to tune (without binning) data to a parametric physics model (for example cluster) as well. However, this would require making the cluster model differentiable.

Outlook

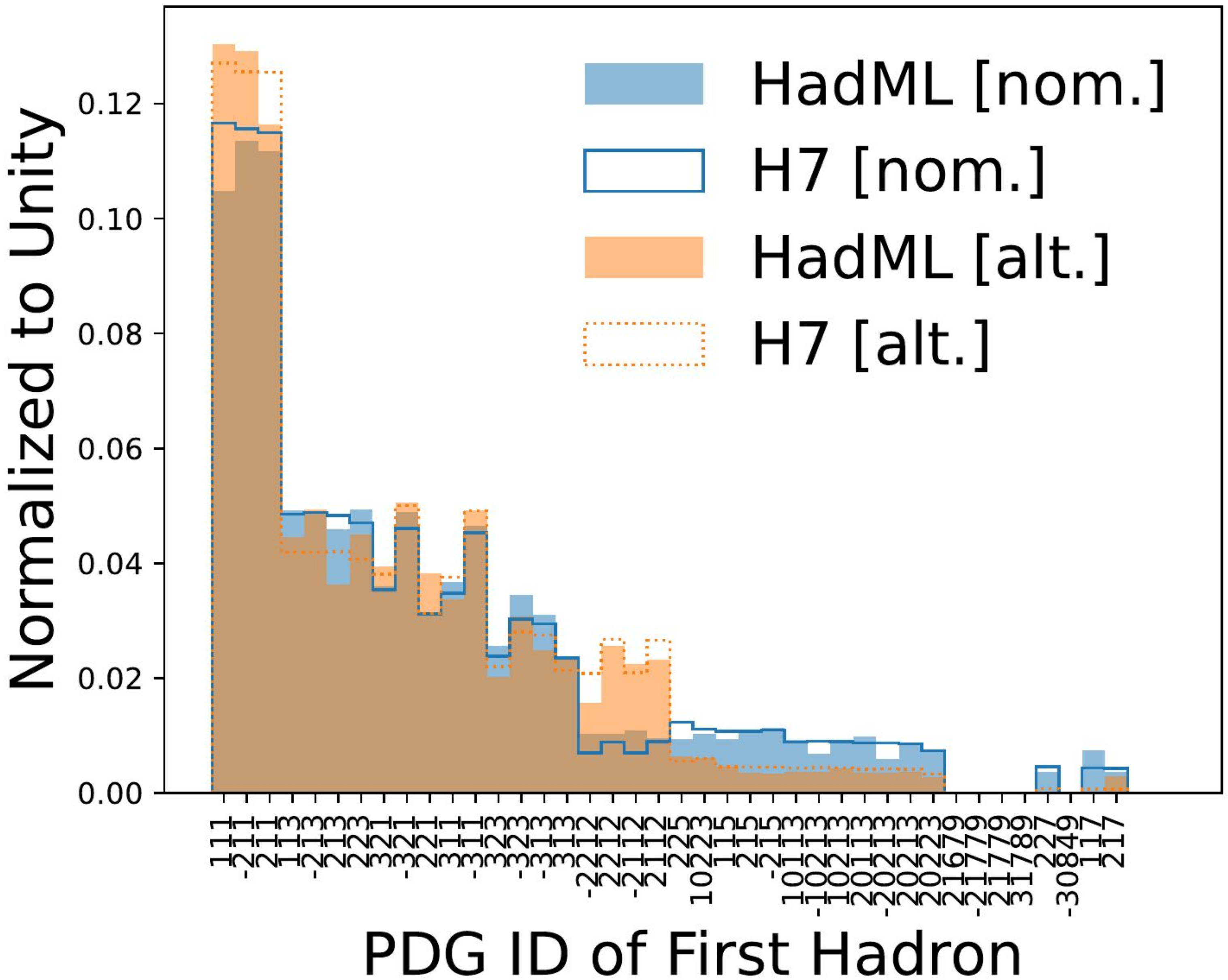
- For HADML, we have made significant progress, but there are still multiple steps to build and tune a full-fledged hadronization model.



What is next?

- Number of technical and methodological step needed:

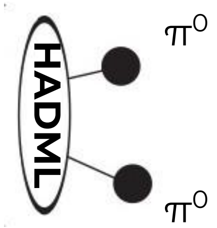
  ➔ Directly accommodate multiple hadron species with their relative probabilities


HadML [nom.]
H7 [nom.]
HadML [alt.]
H7 [alt.]
Normalized to Unity
0.00
0.02
0.04
0.06
0.08
0.10
0.12
PDG ID of First Hadron

Outlook

- For HADML, we have made significant progress, but there are still multiple steps to build and tune a full-fledged hadronization model.

- HADML is naturally suited for GPUs

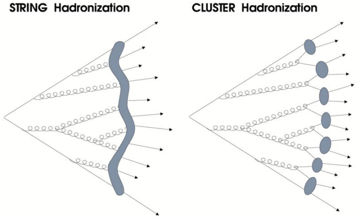


What is next?

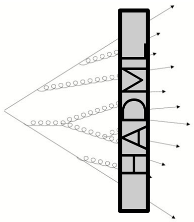
- Number of technical and methodological step needed:
  - ➔ Directly accommodate multiple hadron species with their relative probabilities
  - ➔ Include heavy clusters (so far done by Herwig)
  - ➔ Hyperparameter optimization, including the investigation of alternative generative models
  - ➔ More flexible model with a capacity to mimic the cluster or string models and beyond.
  - ➔ Tune to the LEP data

There is still a multi-year program ahead of us, but it will be worth it!

Early 1980's



Early 2020's



HADML on Quantum Computer?

So Stay tuned!

A postdoc in ML/HEP position

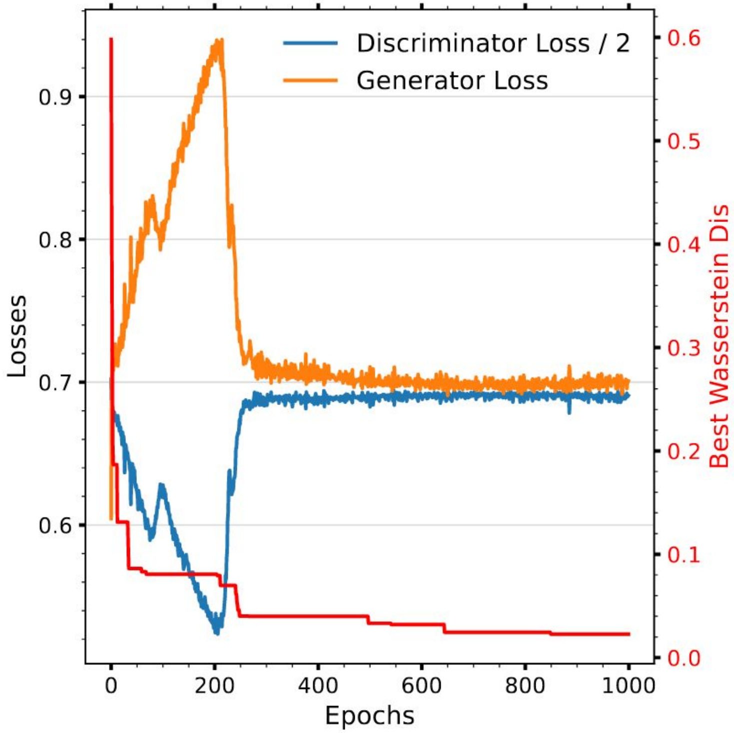
JAGIELLONIAN UNIVERSITY
IN KRAKÓW

If you are interested please contact me:
andrzej.siodmok@cern.ch

Training

- **Data normalization:**
  cluster's four vector and angular variables are scaled to be between -1 and 1 (tanh activation function as the last layer of the Generator)

- **Discriminator** and the **Generator** are trained separately and alternately by two independent Adam optimizers with a learning rate of 10^{-4}, for 1000 epochs



- **The best model** for events with partons of Pert = 0, is found at the epoch 849 with a total Wasserstein distance of 0.0228.

Wasserstein distance

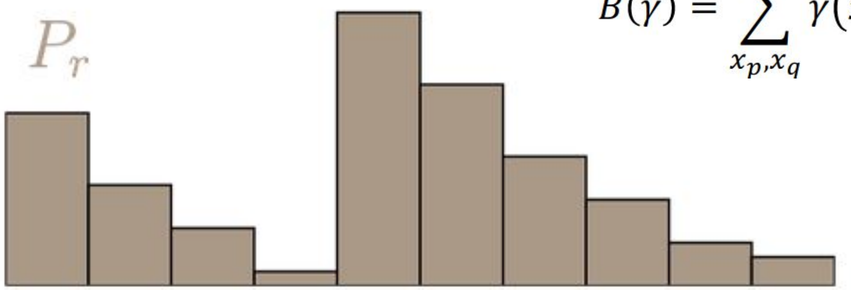
The Wasserstein distance

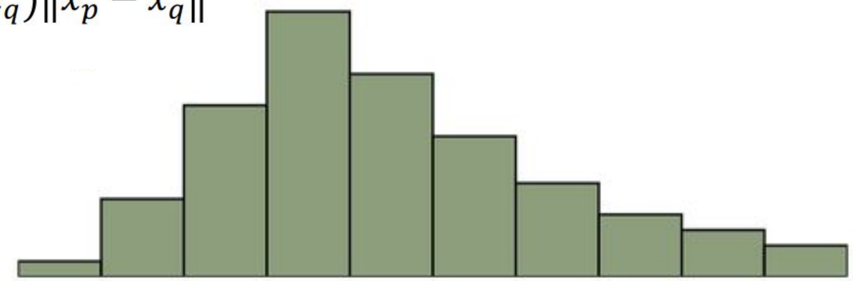
- For discrete probability distributions, the Wasserstein distance is called the earth mover's distance (EMD):
- EMD is the minimal total amount of work it takes to transform one heap into the other.

$$W(P,Q) = \min_{\gamma \in \Pi} B(\gamma)$$

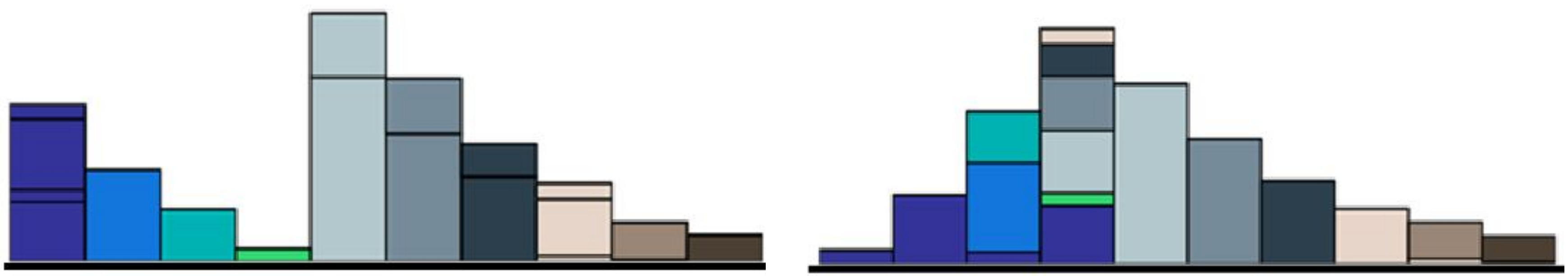
- Work is defined as the amount of earth in a chunk times the distance it was moved.

$$B(\gamma) = \sum_{x_p, x_q} \gamma(x_p, x_q) \|x_p - x_q\|$$

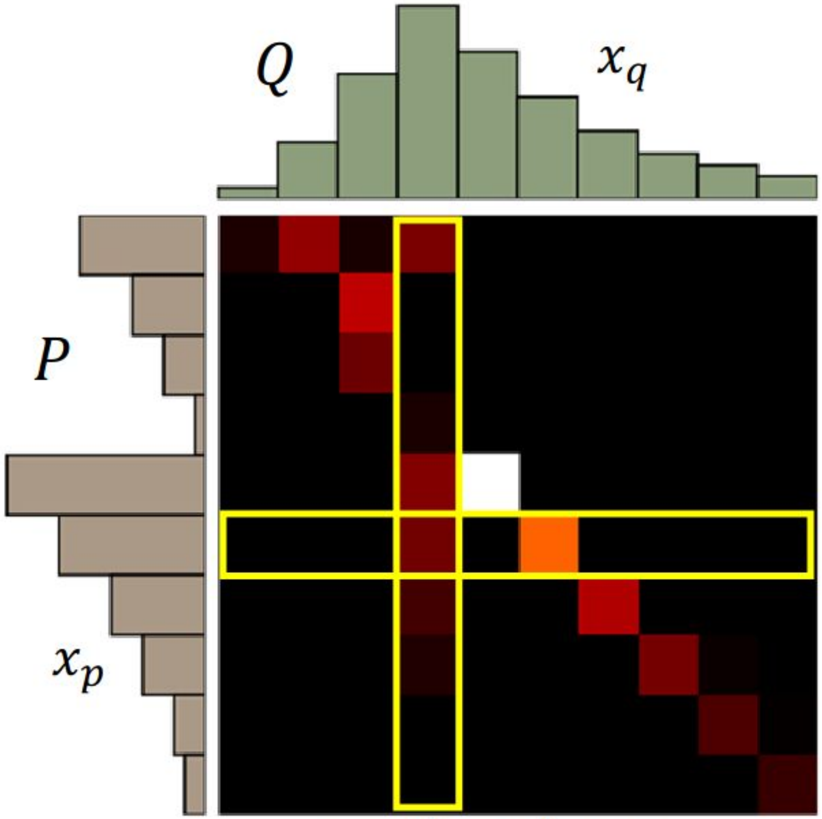




Best "moving plans" of this example



Wasserstein distance



moving plan γ

All possible plan Π

A "moving plan" is a matrix

The value of the element is the amount of earth from one position to another.

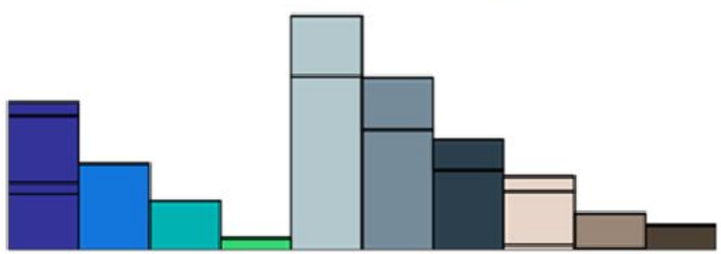
Average distance of a plan γ:

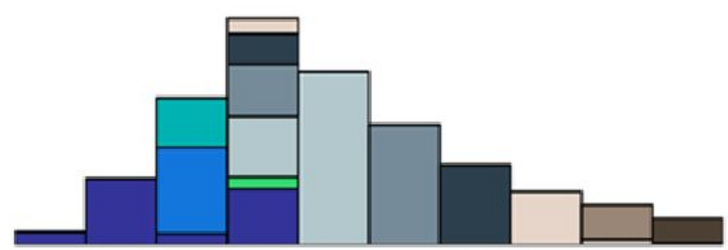
$$B(\gamma) = \sum_{x_p, x_q} \gamma(x_p, x_q) \|x_p - x_q\|$$

Earth Mover's Distance:

$$W(P,Q) = \min_{\gamma \in \Pi} B(\gamma)$$

The best plan





Discriminator HadML v1 vs v2

HadML v1

The loss function:

$$L = -\sum_{\lambda \sim \textsc{Herwig},\, z \sim p(z)} \left(\log\left(D\left(\tau\left(\lambda\right)\right)\right) + \log\left(1 - D\left(G\left(z, \lambda\right)\right)\right)\right)$$

HadML v2

The discriminator function is modified, we parameterize is as a Deep Sets model

$$D_E\left(x\right) = F\left(\frac{1}{n}\sum_{i=1}^{n} \Phi\left(h_i, \omega_{D_\Phi}\right), \omega_F\right)$$

invariant under permutations of hadrons

Φ embeds a set of hadrons into a fixed-length latent space and F acts on the average

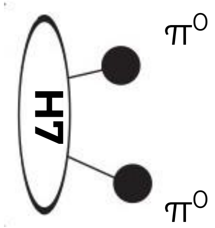
$$L = -\sum_{x \sim \text{data}} \log\left(D_E\left(x\right)\right) - \sum_{\{G\} \sim \textsc{Herwig},\, z \sim p(z)} \log\left(1 - D_E\left(\{G\left(z, \lambda\right)\}\right)\right)$$

The approach could also be used to fit (without binning) data to a parametric physics model (for example cluster) as well. However, this would require making the cluster model differentiable.

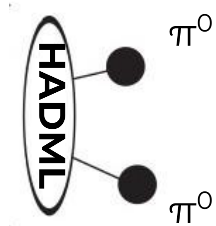
Performance: Energy of the collisions

Low-level Validation
(**beyond** training data **different energy**)

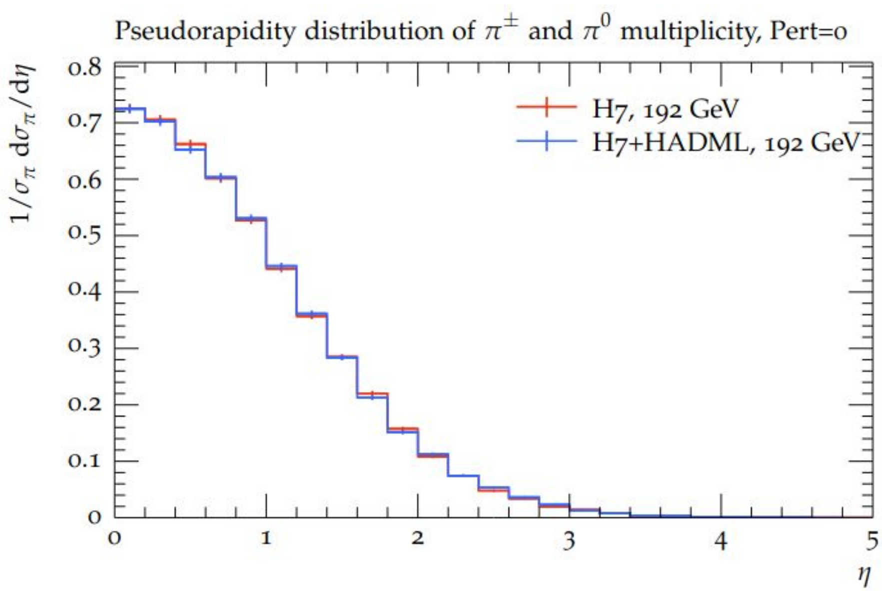
e^+e^- collisions at
$\boxed{\sqrt{s} = 192\ \text{GeV}}$

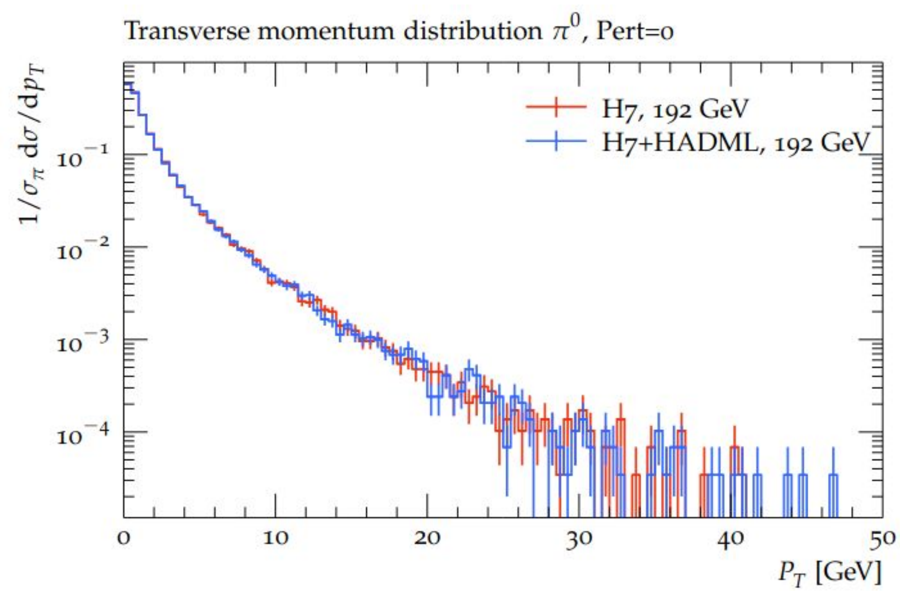


VS



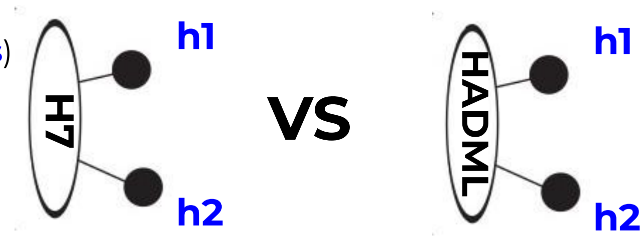
π^0 kinematic variables





Low-level Validation
(**beyond** training data **different hadrons**)

e^+e^- collisions at
$$\sqrt{s} = 91.2 \text{ GeV}$$



h kinematic variables

As a crude "full" model, we simply take the PIDs from Herwig and the kinematics from the GAN.