Fitting a Deep Generative Hadronization Model

Andrzej Siódmok

Jagiellonian University

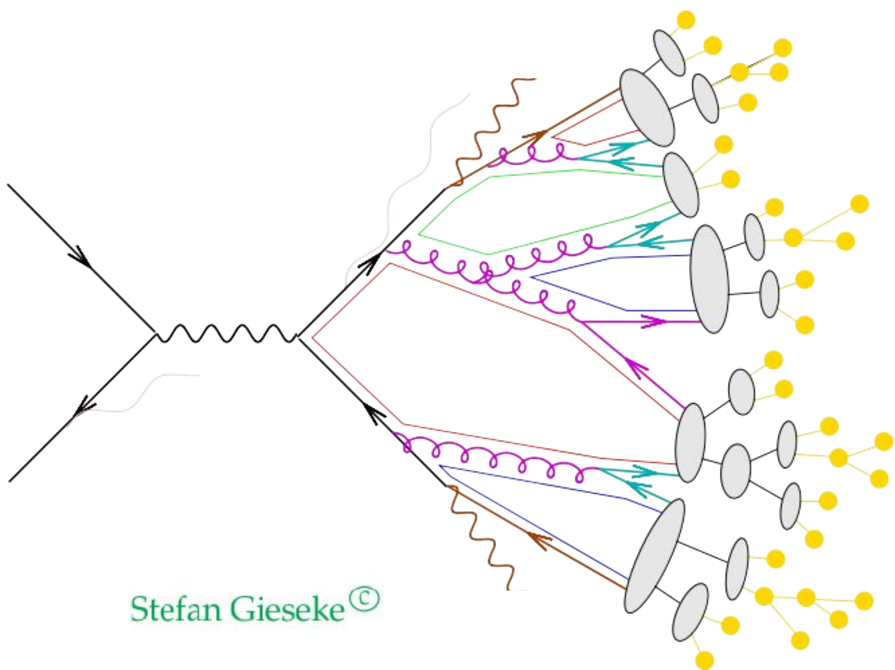
NCN: 2019/34/E/ST2/00457

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)**



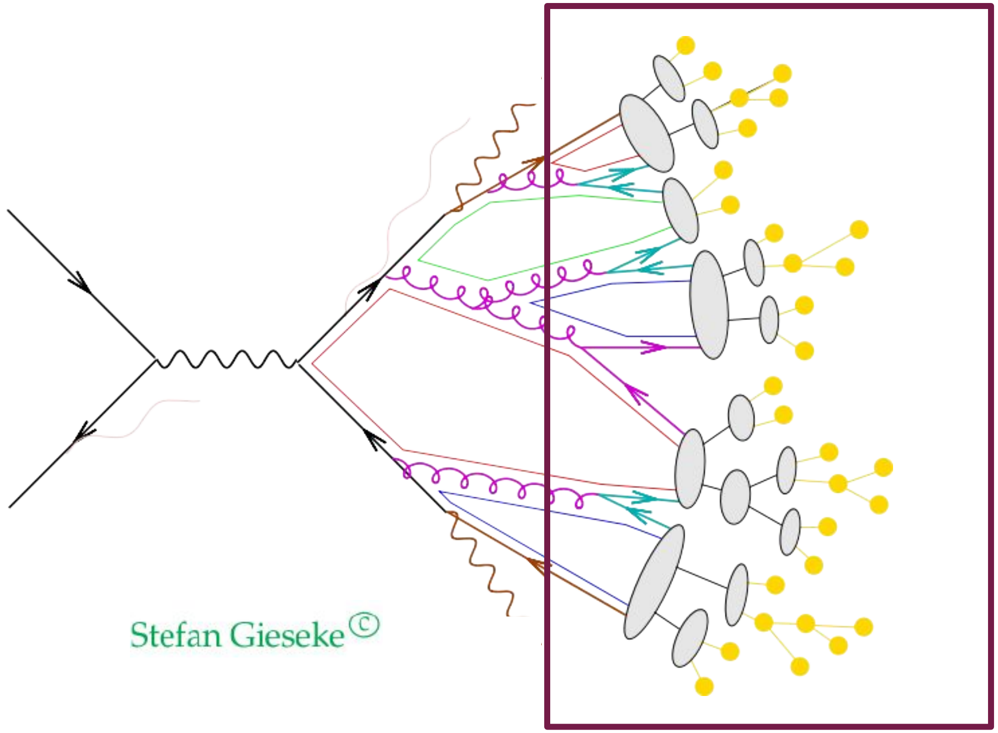
[See talks on MC generators by A. Masouminia, E. Bothmann and P. Skands]

Why hadronization?

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 - Hadronization

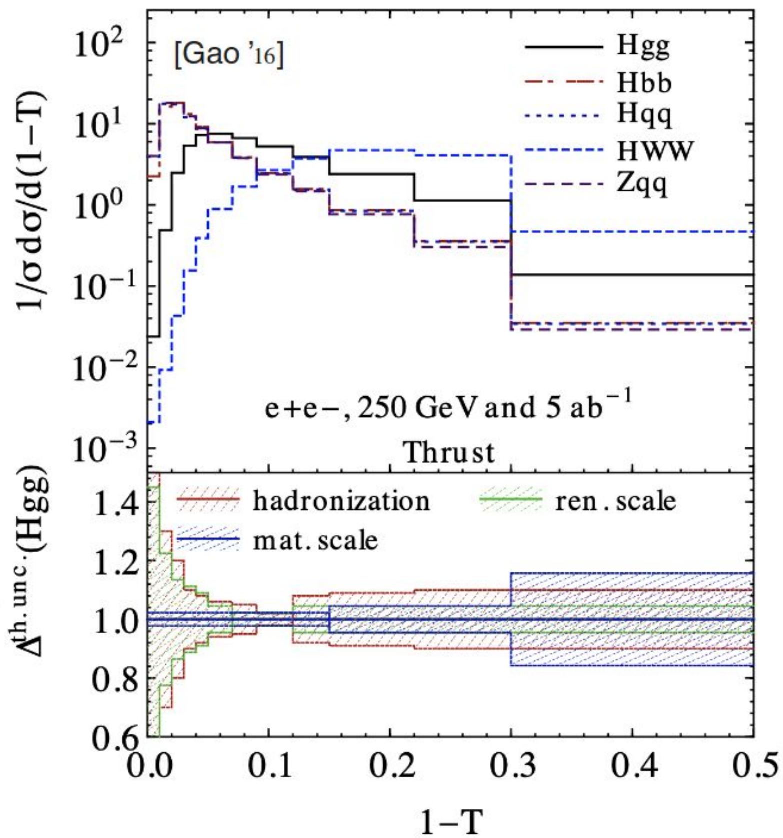
Hadronization:

➔ Increased control of perturbative corrections ⇒ more often LHC measurements are limited by non-perturbative components, such as hadronization.

- W mass measurement using a new method [Freytsis at al. JHEP 1902 (2019) 003]
- Extraction of the strong coupling in [M. Johnson, D. Maître, Phys.Rev. D97 (2018) no.5]
- Top mass [S. Argyropoulos, T. Sjöstrand, JHEP 1411 (2014) 043]
- ...

Pier Moni's talk
FCC Physics Workshop 2023

- However, hadronisation remains the main bottleneck
  - e.g. thrust in Higgs decays (MC variation in plot)
- Increase in energy insufficient for suppression ($Q \sim m_H$)
- Runs at lower energies are essential for a robust tuning of NP models in MCs
- Also crucial for training of ML algorithms for jet tagging, instrumental in extraction of Higgs couplings



Motivation - Hadronization

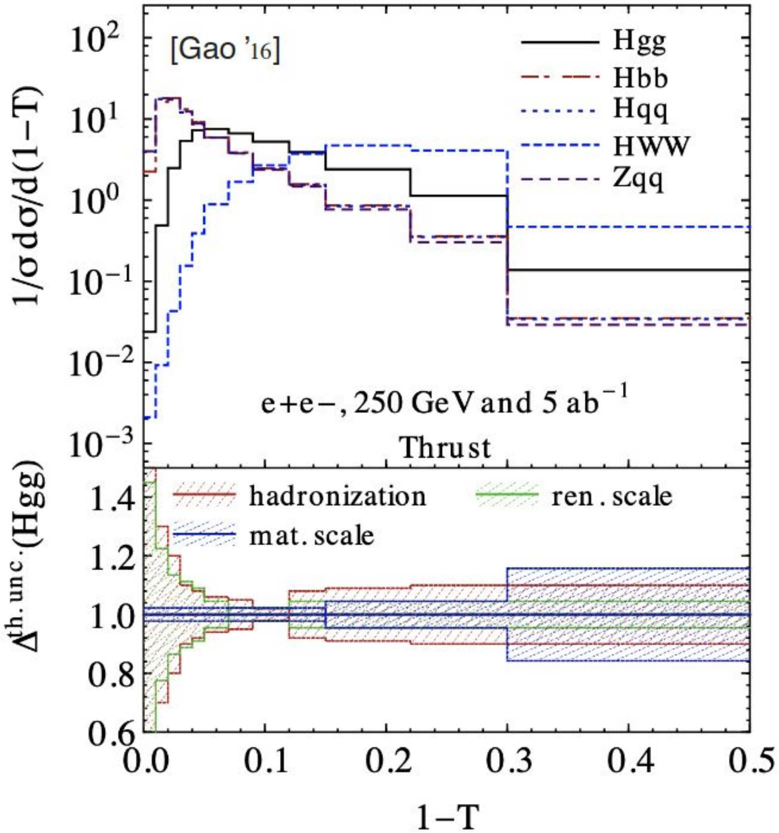
Hadronization:

➔ Increased control of perturbative corrections ⇒ more often LHC measurements are limited by non-perturbative components, such as hadronization.

- W mass measurement using a new method [Freytsis at al. JHEP 1902 (2019) 003]
- Extraction of the strong coupling in [M. Johnson, D. Maître, Phys.Rev. D97 (2018) no.5]
- Top mass [S. Argyropoulos, T. Sjöstrand, JHEP 1411 (2014) 043]
- ...

Pier Moni's talk
FCC Physics Workshop 2023

- However, hadronisation remains the main bottleneck
  - e.g. thrust in Higgs decays (MC variation in plot)
- Increase in energy insufficient for suppression ($Q \sim m_H$)



Taken from Andrea's Banfi talk

We don't talk about

hadronisation

Slide adopted to my talk :)

We have to talk about

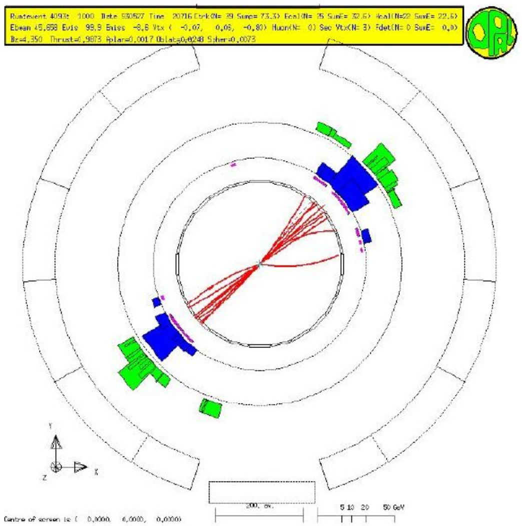
hadronisation

String Model in nutshell

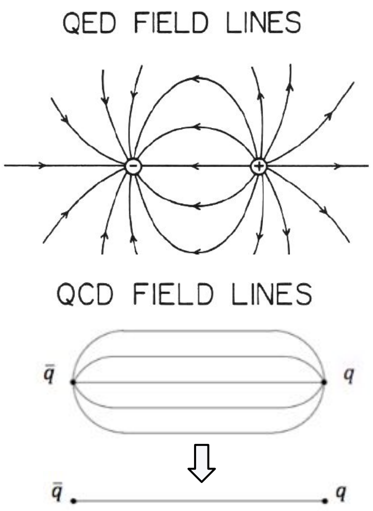
[Andersson, Gustafson, Ingelman, Sjostrand, Phys.Rept.97(1983)31]

Originally invented without perturbative physics of parton showers in mind.

We start with 2-jet events in
$e^+ e^- \to$ hadrons.



Self coupling of gluons
"attractive field line"

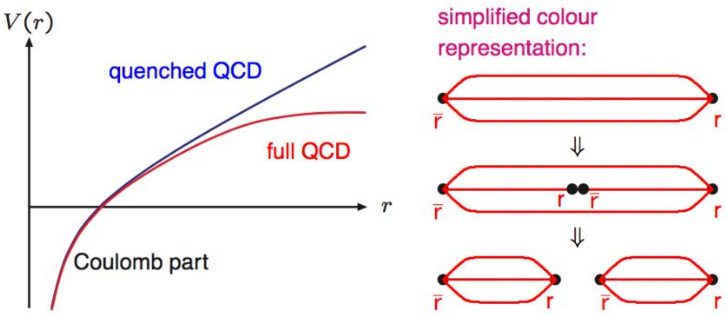


1+1 - dim. object: **string**

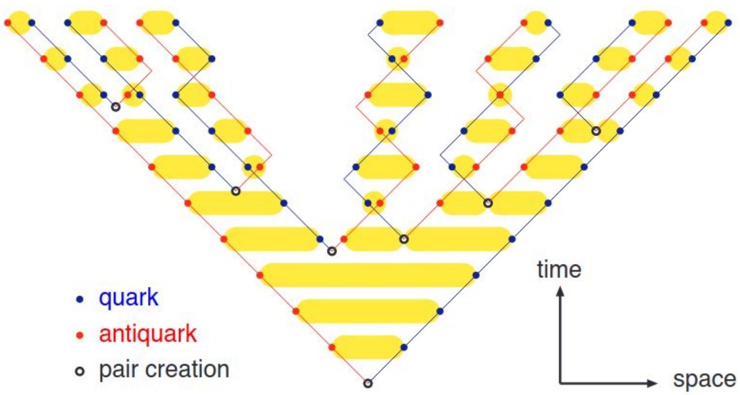
Linear static potential:

$F(r) \approx \text{const} = \kappa \approx 1\ \text{GeV/fm} \iff V(r) \approx \kappa r$

Picture supported by lattice QCD



Lund string model: like rubber band that is pulled apart and breaks into pieces



Plots from T. Sjostrand

String Model in nutshell

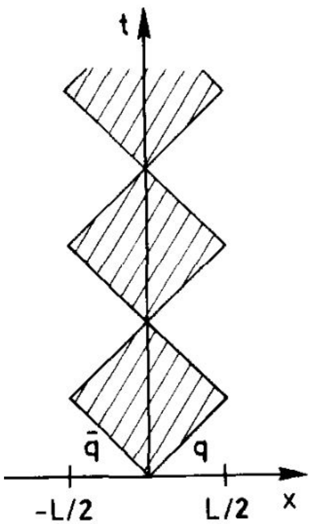
[Andersson, Gustafson, Ingelman, Sjostrand, Phys.Rept.97(1983)31]

String motion

From linear static potential $V(r) \approx \kappa r$ and linearity between space–time and energy–momentum:

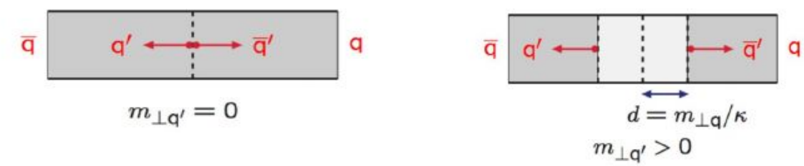
$$\left|\frac{\mathrm{d}E}{\mathrm{d}z}\right| = \left|\frac{\mathrm{d}p_z}{\mathrm{d}z}\right| = \left|\frac{\mathrm{d}E}{\mathrm{d}t}\right| = \left|\frac{\mathrm{d}p_z}{\mathrm{d}t}\right| = \kappa$$

We get a “YoYo“ state which we interpret as a meson.



String breakdowns

The quarks obtain a mass and a transverse momentum in the breakup through a tunneling mechanism



with a probability:

$$\mathcal{P} \propto \exp\left(-\frac{\pi m_{\perp q}^2}{\kappa}\right) = \exp\left(-\frac{\pi p_{\perp q}^2}{\kappa}\right) \exp\left(-\frac{\pi m_q^2}{\kappa}\right)$$

- Suppression of heavy quarks: uu : dd : ss : cc ≈ 1 : 1 : 0.3 : 10^{-11}
- Common Gaussian pT spectrum, <pT>~ 0.4 GeV
- Diquark (qq - $\bar{q}\bar{q}$ breakups) ~ antiquark ⇒ simple model for baryon production.

Iterative process (left-right symmetry) leads to distribution of momentum fraction taken by each hadron as:

$$f(z) \propto \frac{(1-z)^a}{z} \exp\left(-\frac{bm^2}{z}\right)$$

Summary [for a recent progress see P. Skands talk]:

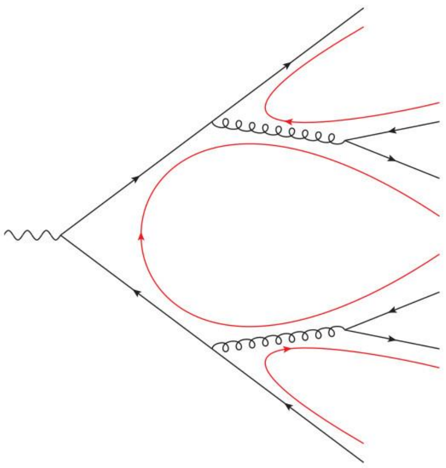
String model has very good energy-momentum picture however it is unpredictive in understanding of hadron mass effects ⇒ many parameters, 10-30 depending on how you count.

Cluster model in nutshell

[Webber NPB238(1984)492]

What if we have PS (more perturbative input before hadronization).

The philosophy of the model: use information from perturbative QCD as an input for hadronization.
QCD **pre-confinement** discovered by Amati & Veneziano [*Phys.Lett.B* 83 (1979) 87-92]:



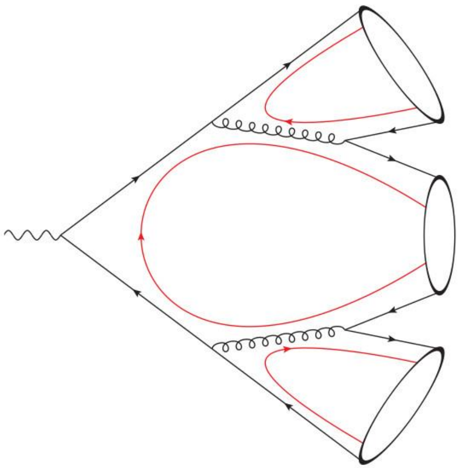
- QCD provide pre-confinement of colour

Cluster model in nutshell

[Webber NPB238(1984)492]

What if we have PS (more perturbative input before hadronization).

The philosophy of the model: use information from perturbative QCD as an input for hadronization.
QCD **pre-confinement** discovered by Amati & Veneziano [*Phys.Lett.B* 83 (1979) 87-92]:



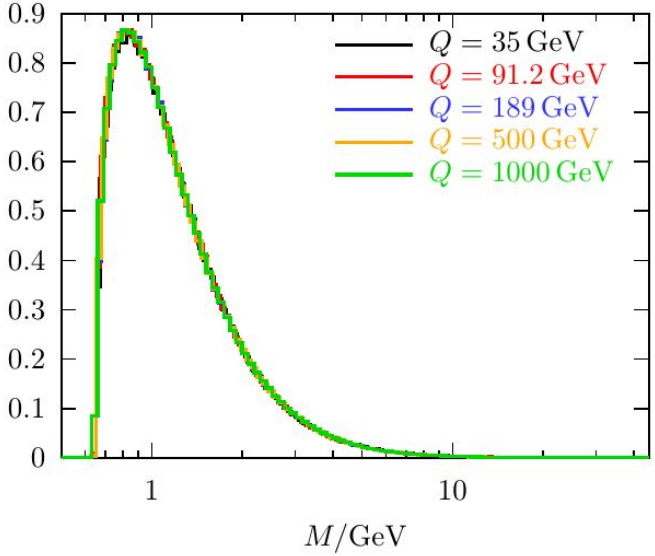
- QCD provide pre-confinement of colour
- Colour-singlet pair end up close in phase space and form highly excited hadronic states, the clusters

Cluster model in nutshell
[Webber NPB238(1984)492]

What if we have PS (more perturbative input before hadronization).

The philosophy of the model: use information from perturbative QCD as an input for hadronization. QCD **pre-confinement** discovered by Amati & Veneziano [*Phys.Lett.B* 83 (1979) 87-92]:

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



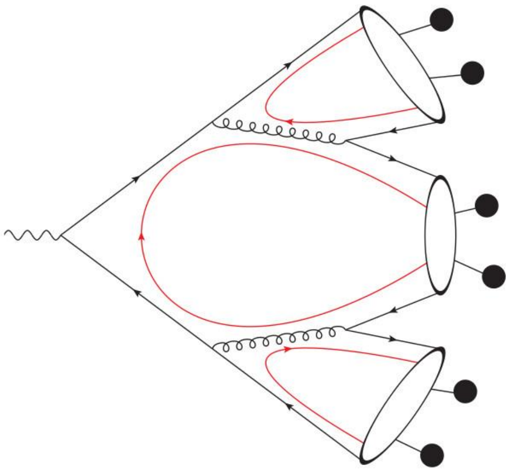
- QCD provide pre-confinement of colour
- Colour-singlet pair end up close in phase space and form highly excited hadronic states, the clusters
- Pre-confinement states that the spectra of clusters are independent of the hard process and energy of the collision

Cluster model in nutshell

[Webber NPB238(1984)492]

What if we have PS (more perturbative input before hadronization).

The philosophy of the model: use information from perturbative QCD as an input for hadronization.
QCD **pre-confinement** discovered by Amati & Veneziano [*Phys.Lett.B* 83 (1979) 87-92]:



- 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

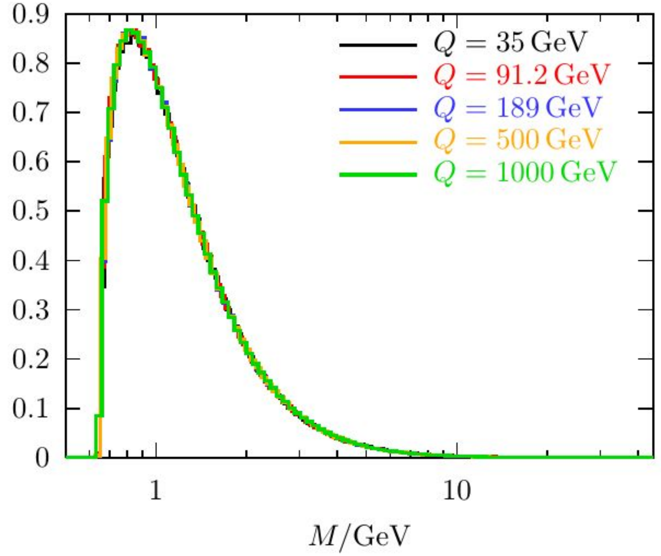
Cluster model in nutshell

[Webber NPB238(1984)492]

What if we have PS (more perturbative input before hadronization).

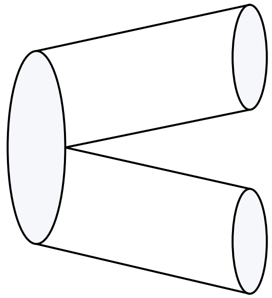
The philosophy of the model: use information from perturbative QCD as an input for hadronization. QCD **pre-confinement** discovered by Amati & Veneziano [*Phys.Lett.B* 83 (1979) 87-92]:

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



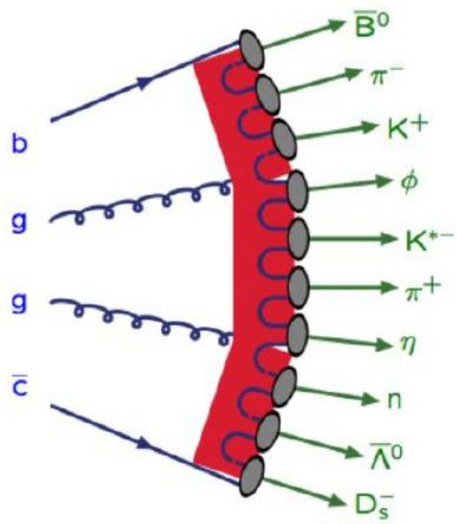
- QCD provide pre-confinement of colour
- Colour-singlet pair end up close in phase space and form highly excited hadronic states, the clusters
- Pre-confinement states that the spectra of clusters are independent of the hard process and energy of the collision
- Peaked at low mass (1-10 GeV) typically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

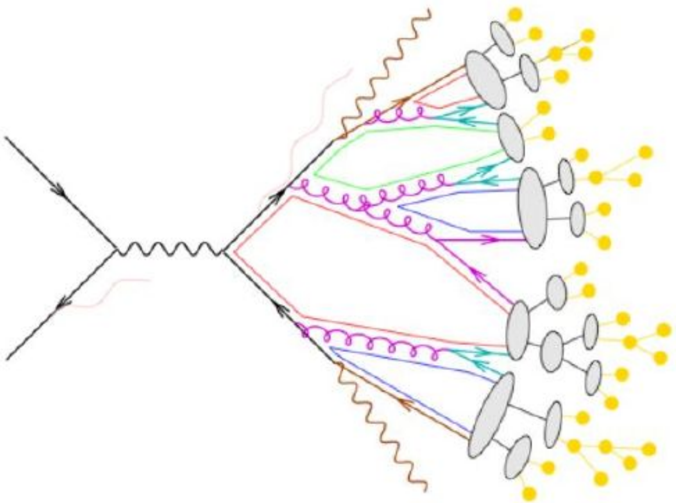
- Small fraction of clusters too heavy for isotropic two-body decay, heavy cluster decay first into lighter cluster C → CC, or radiate a hadron C → HC, it is rather string-like.
- ~ 15% of primary clusters get split but ~ 50% of hadrons come from them! *(see S. Kiebacher talk for some progress)*



[For a recent progress see A. Masouminia, S. Kiebacher talks]

String vs Cluster model





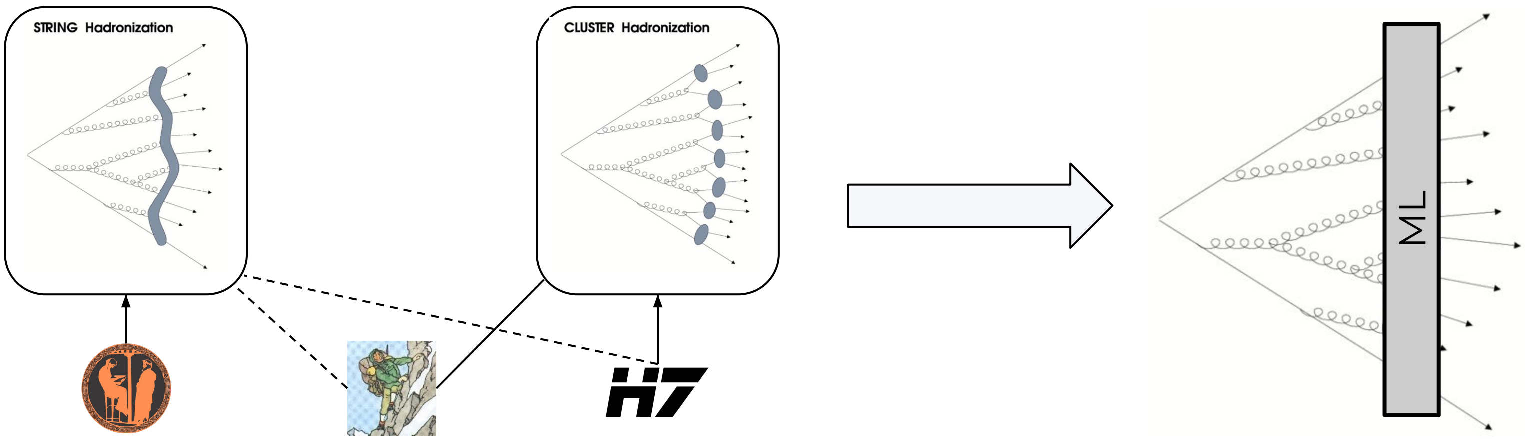
program	PYTHIA	Herwig
model	string	cluster
energy–momentum picture	powerful	simple
	predictive	unpredictive
parameters	few	many
flavour composition	messy	simple
	unpredictive	in-between
parameters	many	few

Taken from T. Sjostrand

Hadronization:

Early 1980's
(see talks by P. Skands, A. Masouminia, S. Kiebacher,...)

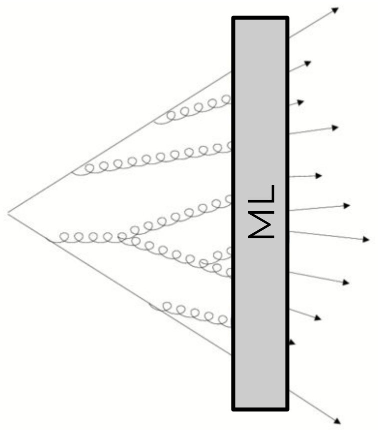
Early 2020's
(lot of progress in ML)



Pythia **Sherpa** **Herwig**

Idea of using Machine Learning (ML) for hadronization.

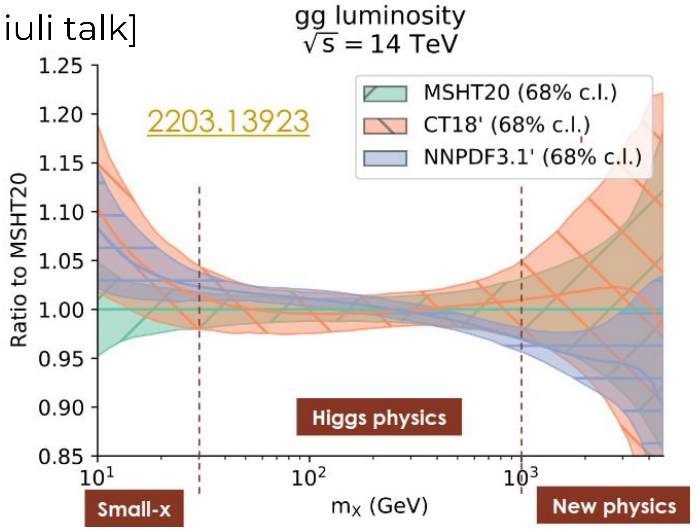
Idea of using Machine Learning (ML) for hadronization.



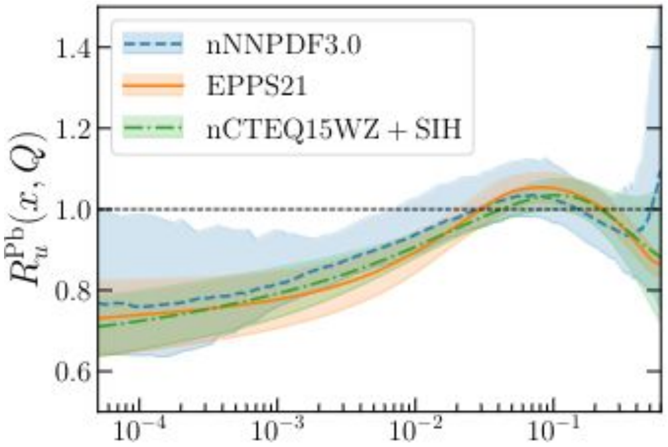
- Existing hadronization models are highly parameterized functions.
- Hadronization is a fitting problem [Ch. Oppedisano talk]
  - Can ML hadronization be more flexible?
  - Can ML hadronization extract more information from the data?
  [can accommodate unbinned and high-dimensional inputs]
- NNPDF

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.

[F. Giuli talk]



[E. Nocera, talk]



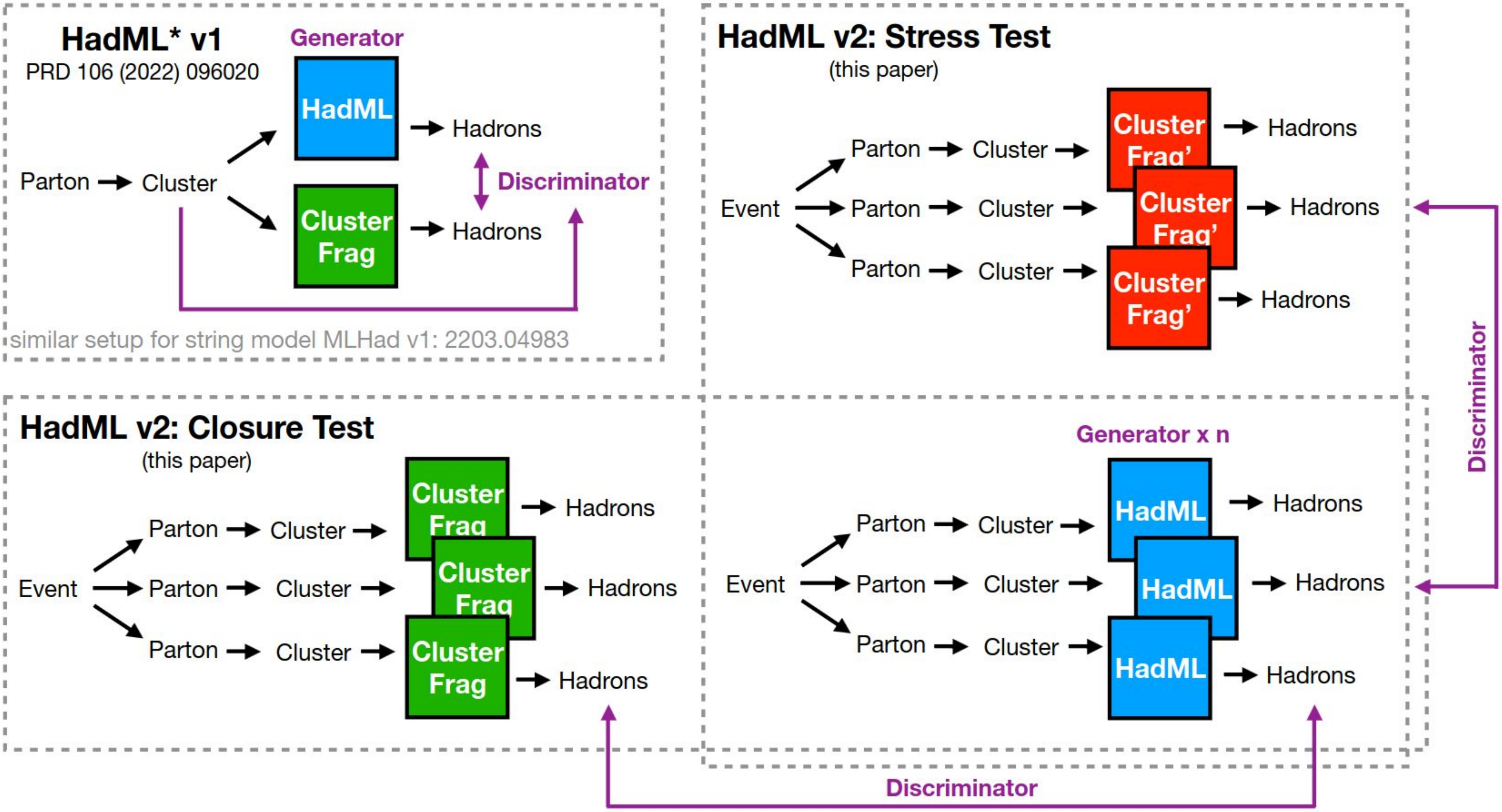
Recent progress: Machine learning hadronization

First steps for ML hadronization:

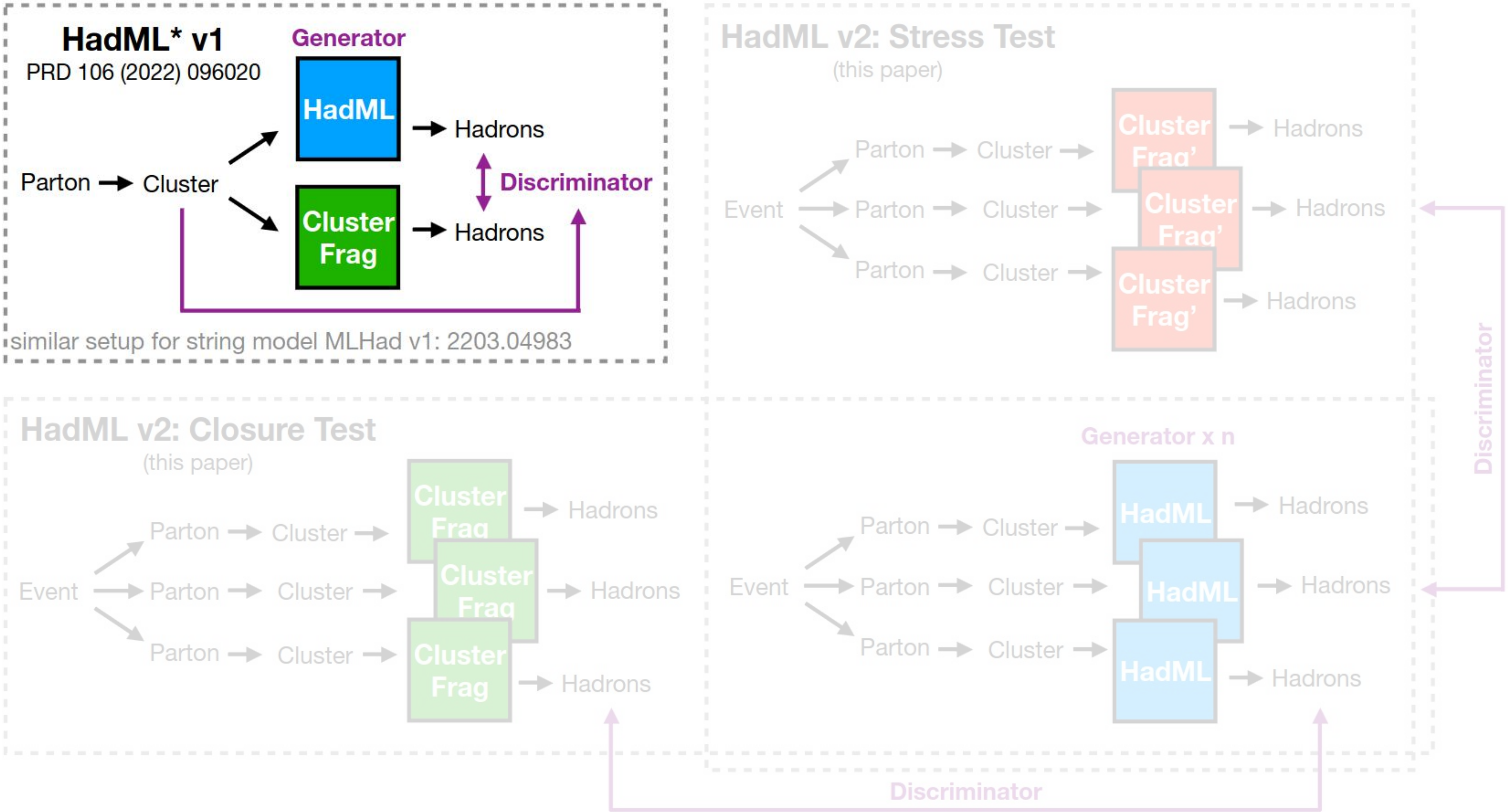
- HADML - [A. Ghosh, Xi. Ju, B. Nachman **AS**, *Phys.Rev.D* 106 (2022) 9]
- MLhad - [P. Ilten, T. Menzo, A. Youssef and J. Zupan, SciPost Phys. 14, 027 (2023)]

	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 P. Skands talk)	*"Fitting a Deep Generative Hadronization Model"* [J. Chan, X. Ju, A. Kania, B. Nachman, V. Sangli and **A.S,** 2305.17169]

Road map for today



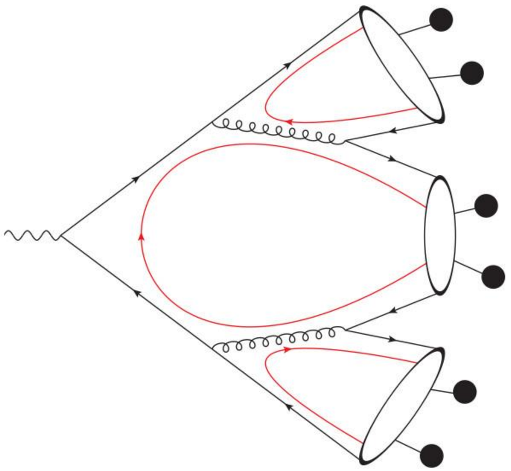
Road map for today



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:

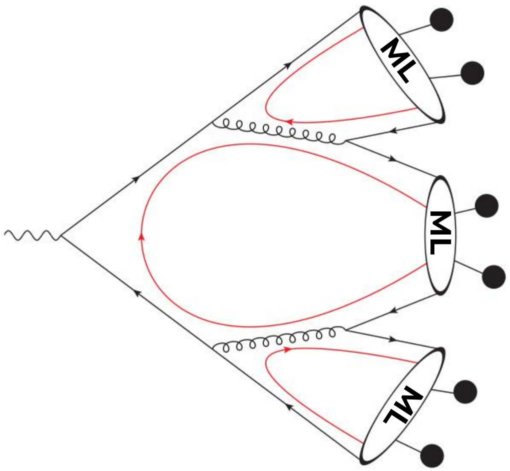


- QCD provide pre-confinement of colour
- Colour-singlet pair end up close in phase space and form highly excited hadronic states, the clusters
- Pre-confinement states that the spectra of clusters are independent of the hard process and energy of the collision
- Peaked at low mass (1-10 GeV) typically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

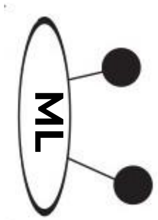
Cluster hadronization model

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

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

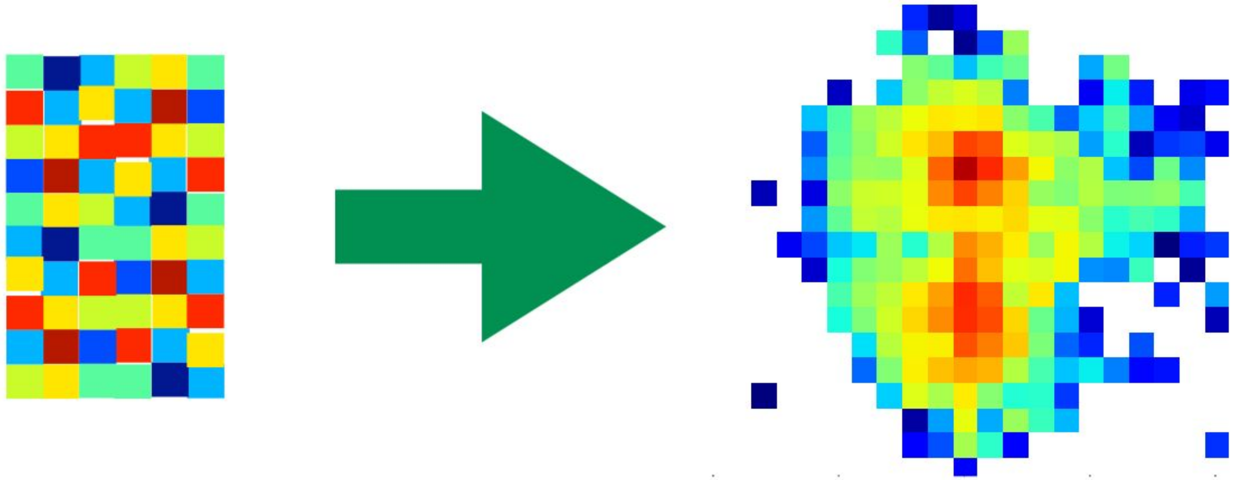


- QCD provide pre-confinement of colour
- Colour-singlet pair end up close in phase space and form highly excited hadronic states, the clusters
- Pre-confinement states that the spectra of clusters are independent of the hard process and energy of the collision
- Peaked at low mass (1-10 GeV) typically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



- **How?**
  Use Generative Adversarial Networks (**GAN**)

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

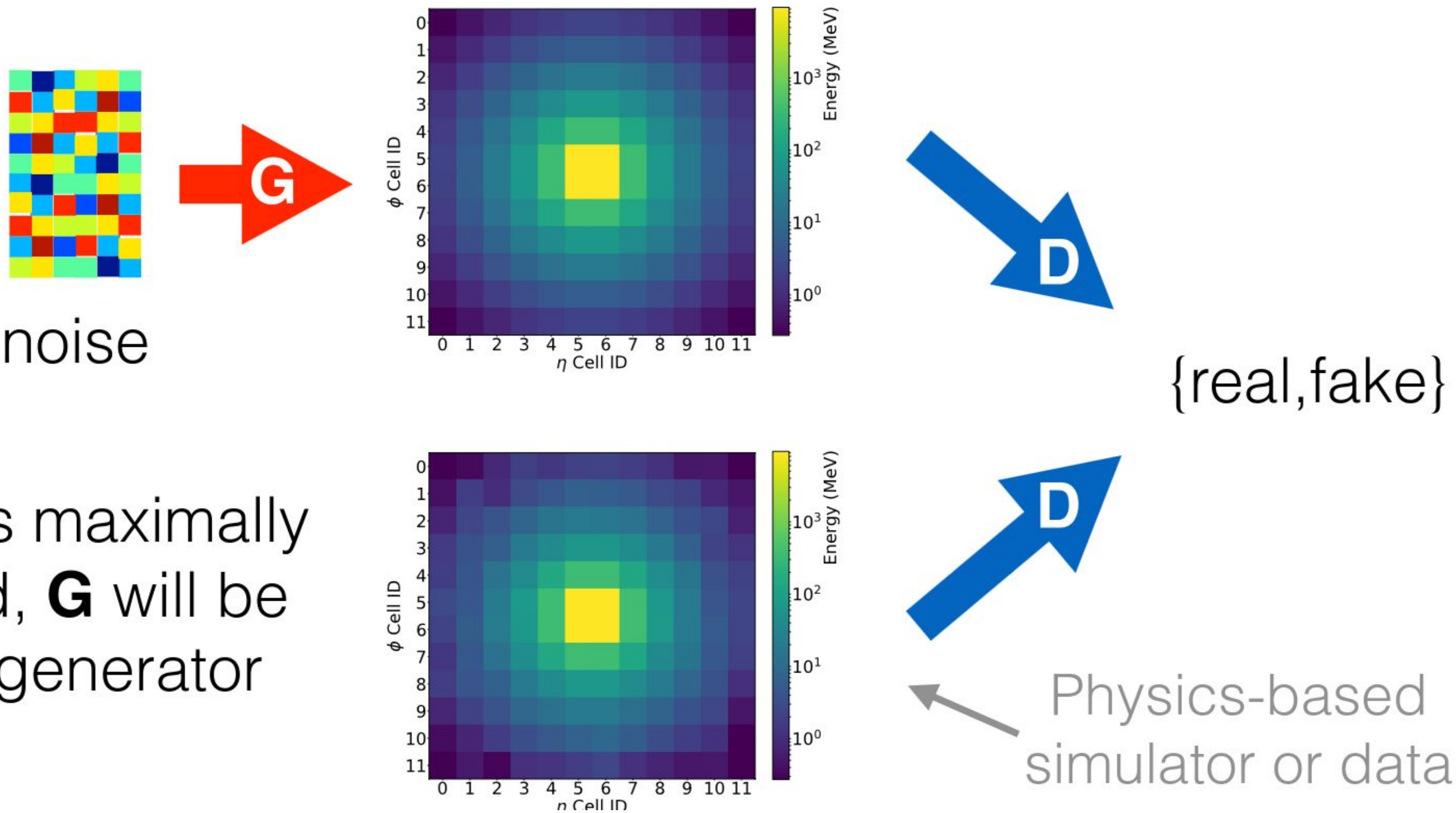


Deep generative models: the map is a deep neural network.

Our tool of choice: GANs

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

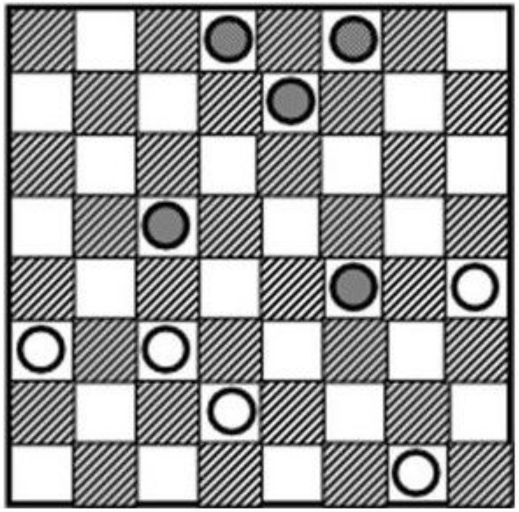
Generative Adversarial Networks (GANs):
*A two-network game where one **maps noise to structure** and one **classifies images as fake or real**.*



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

Physics-based simulator or data

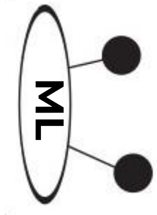
Arthur Lee Samuel (1959) wrote a program that learnt to play checkers well enough to beat him.



- 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.

ML hadronization

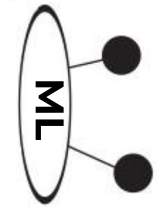
1st step: generate kinematics of a cluster decay to 2 hadrons



Towards a Deep Learning Model for Hadronization

ML hadronization

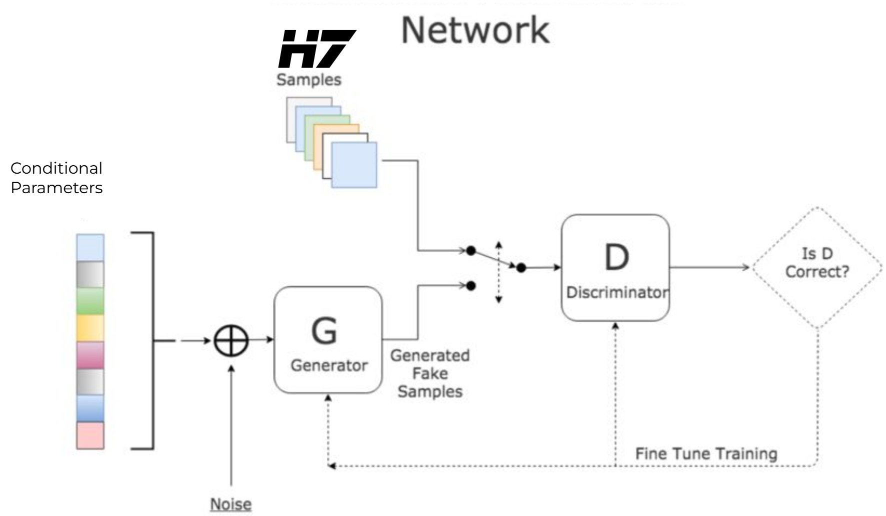
1st step: generate kinematics of a cluster decay to 2 hadrons



How?

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

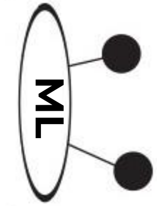
Generative Adversarial



Towards a Deep Learning Model for Hadronization

ML hadronization

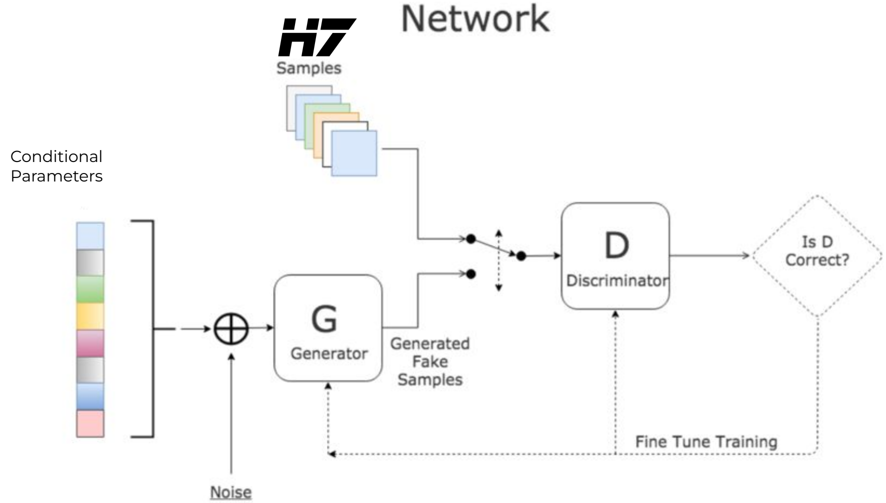
1st step: generate kinematics of a cluster decay to 2 hadrons



How?

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

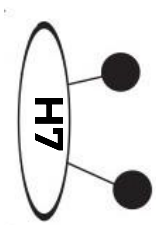
Generative Adversarial Network



Training data:

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

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



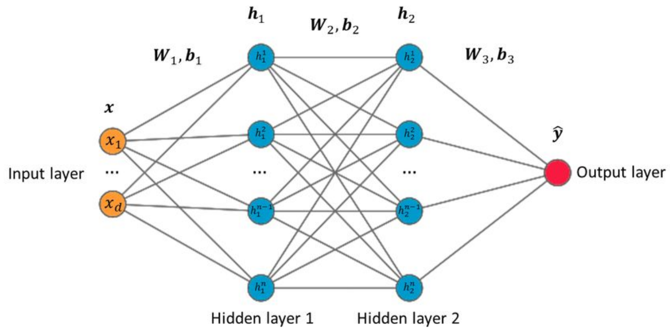
$\pi^0(E,\ p_x,\ p_y,\ p_z)$

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

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

Architecture: conditional GAN

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

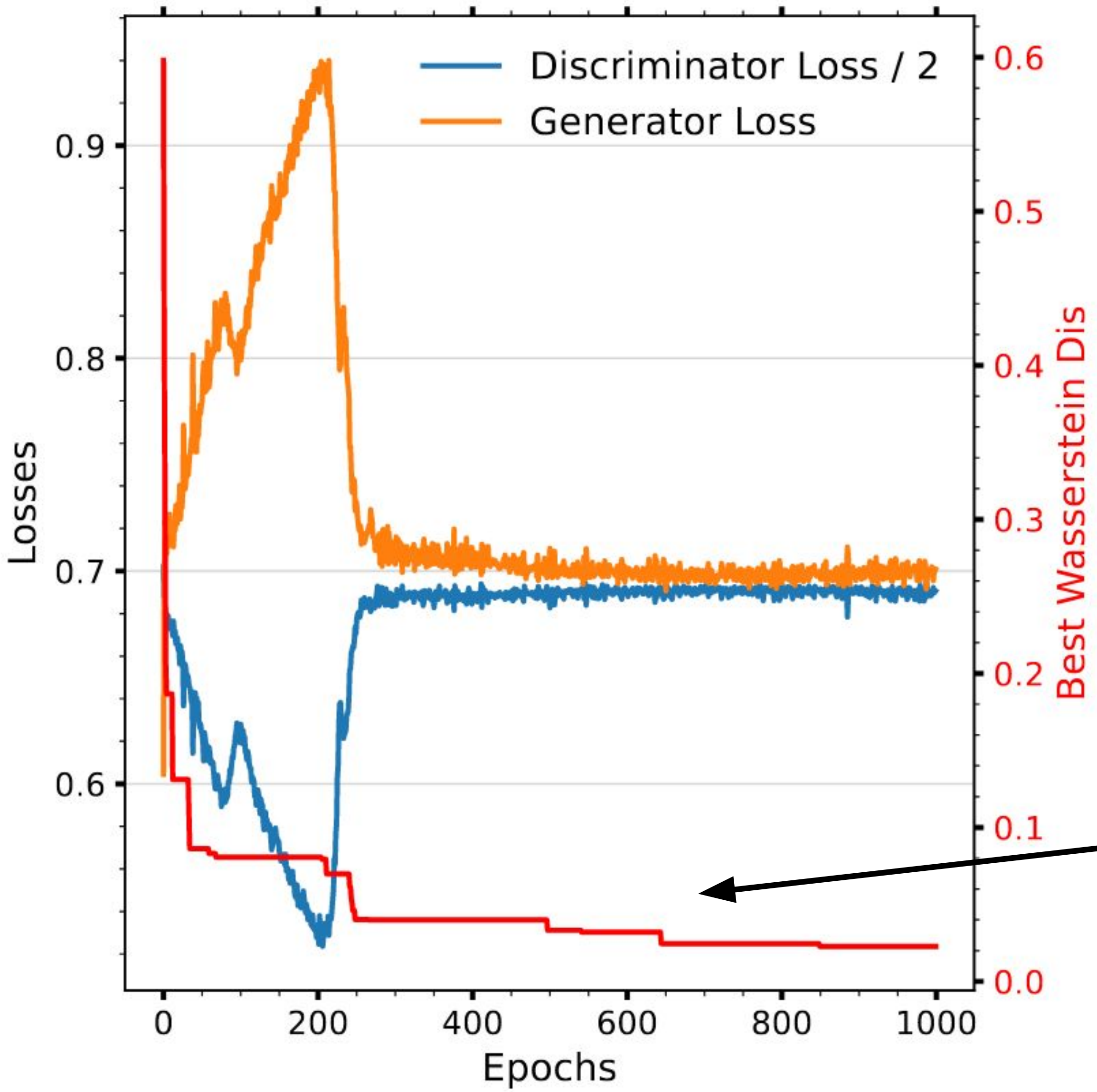
Discriminator

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

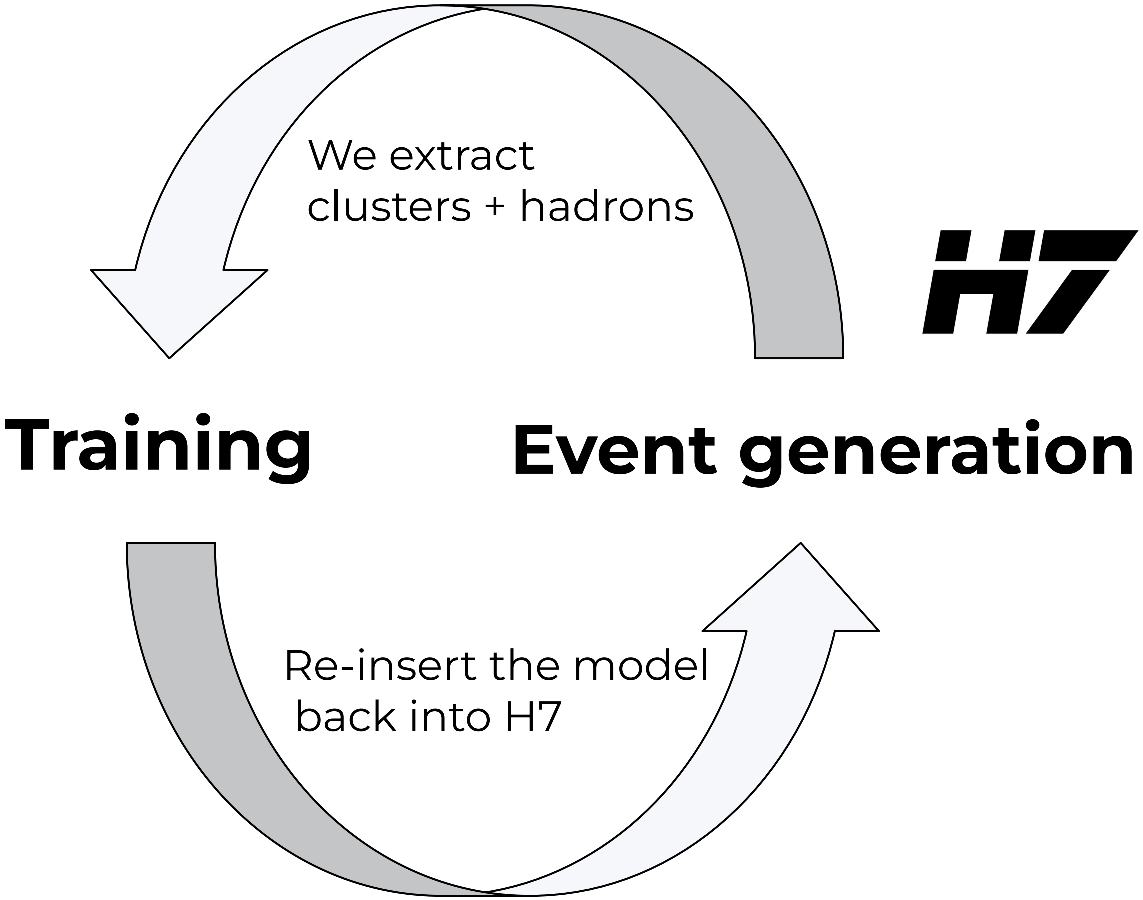


We have a conditional GAN, with cluster 4-vector input and two hadron 4-vector outputs.

<u>Simplification:</u>
considering only pions and generating two angles in the cluster rest frame.

This is a typical learning curve for GAN training

Integration into Herwig

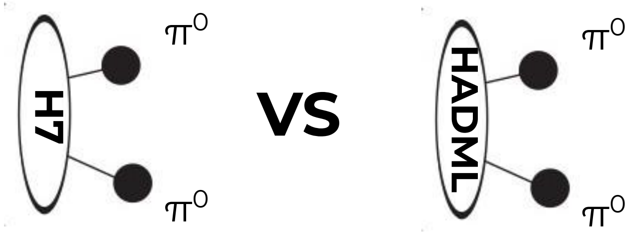


This then allows us to run a full event generator and produce plots

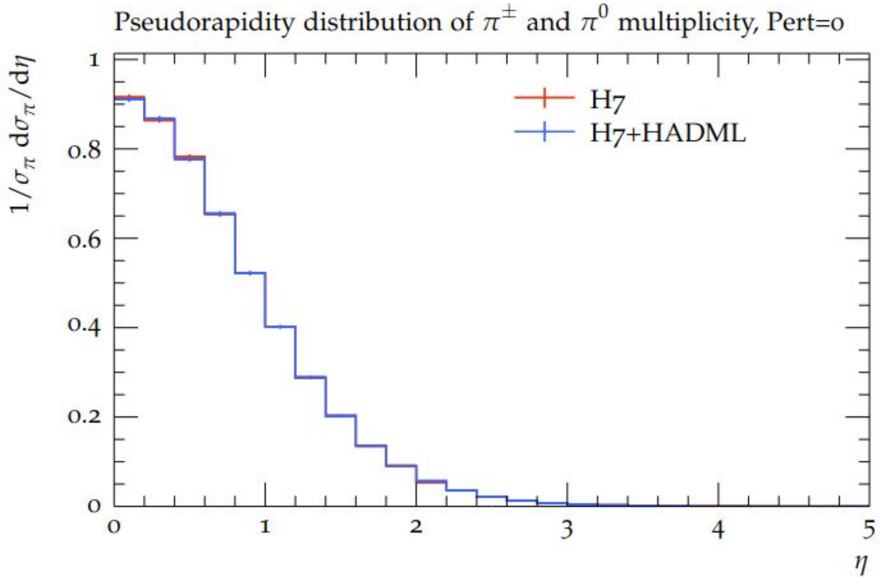
Performance: Pions

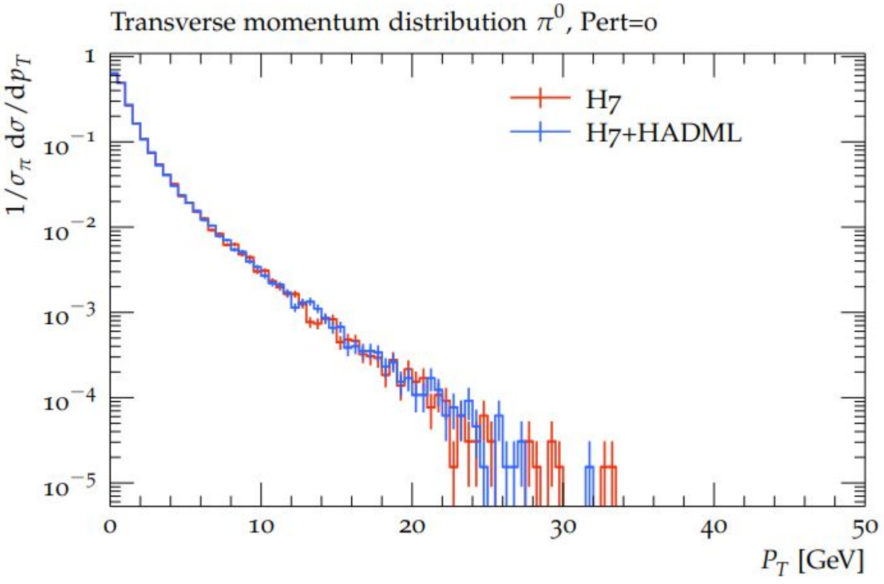
Low-level Validation
(similar to training data)

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



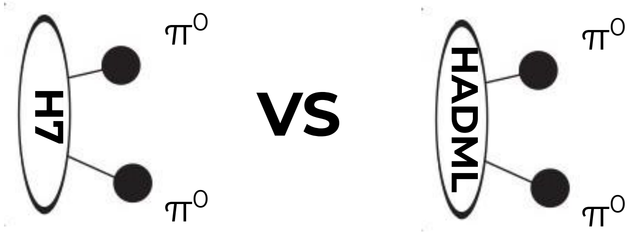
π^0 kinematic variables



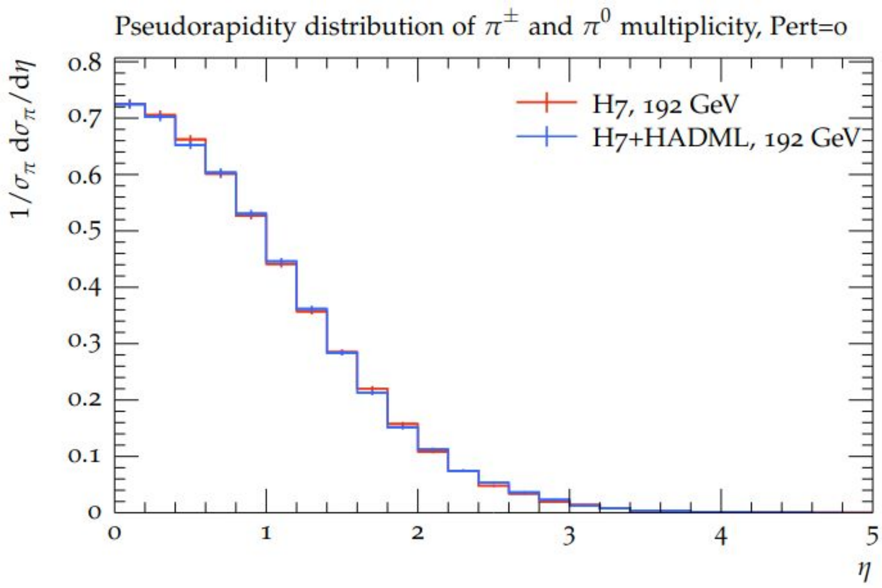


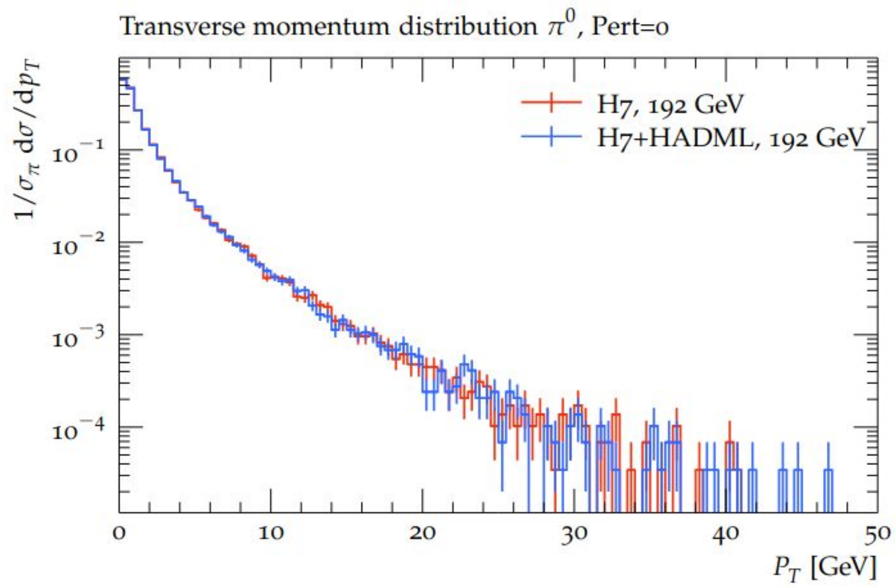
Low-level Validation
(**beyond** training data **different energy**)

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



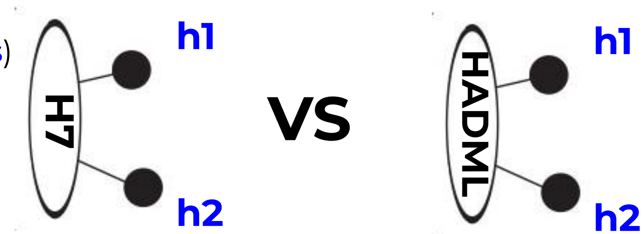
π^0 kinematic variables





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

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

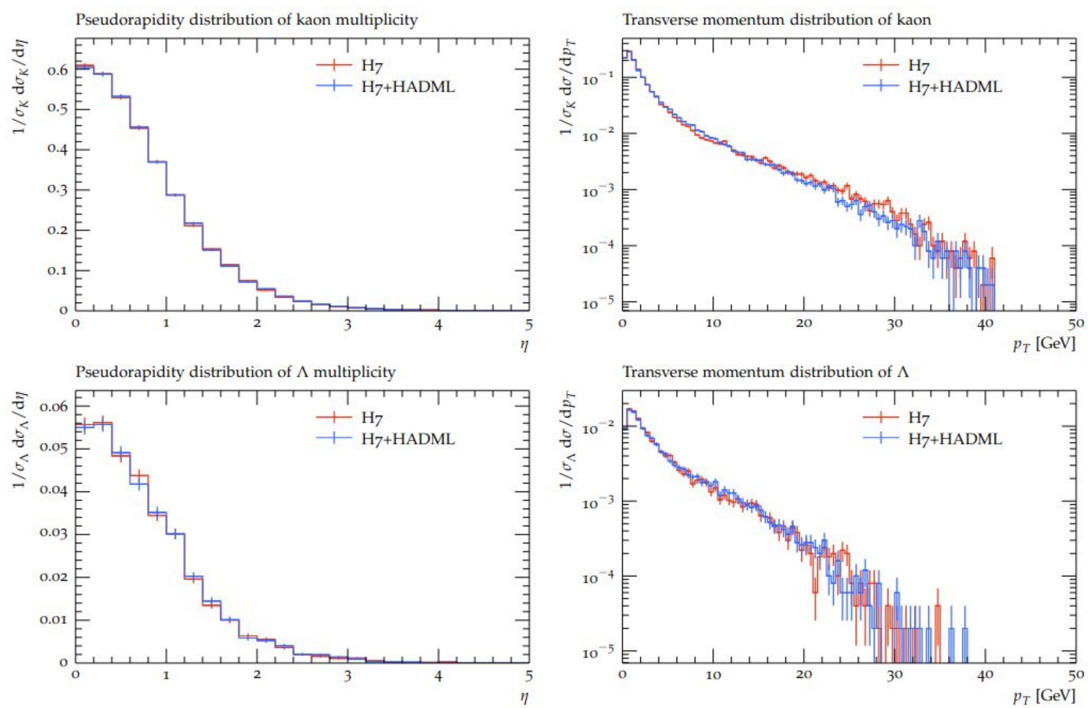


h kinematic variables

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

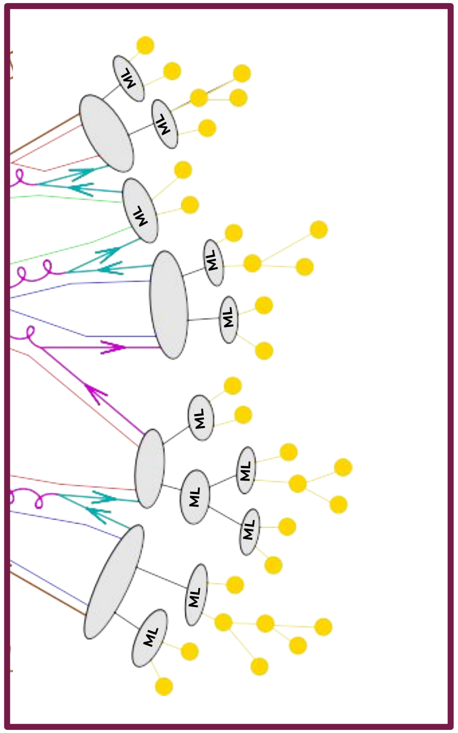
Kaons

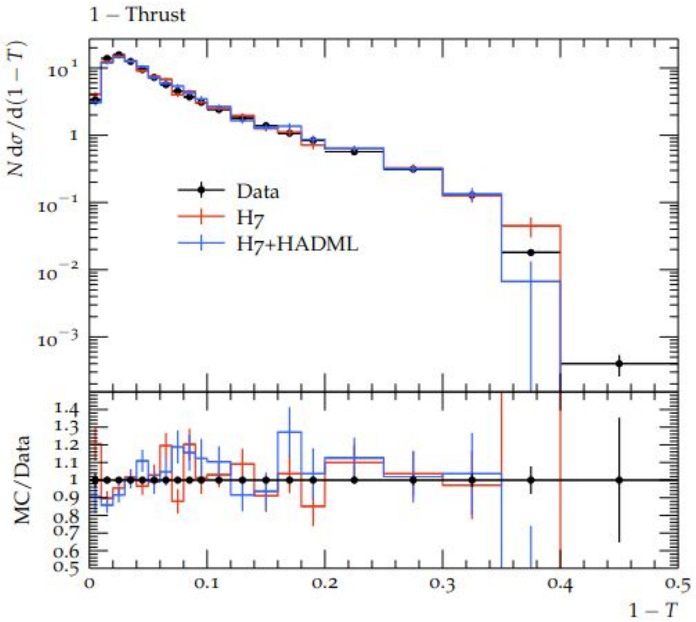
Lambda

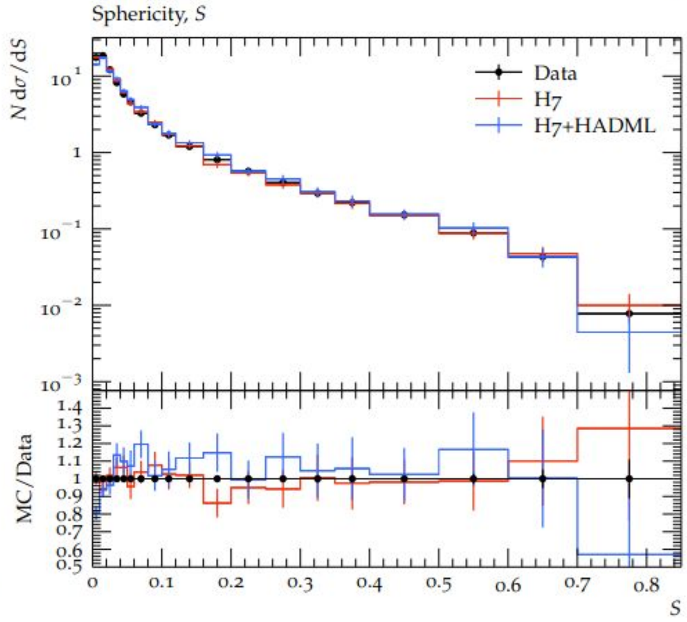


With a "full" model, we can compare directly to data!

LEP DELPHI **Data**

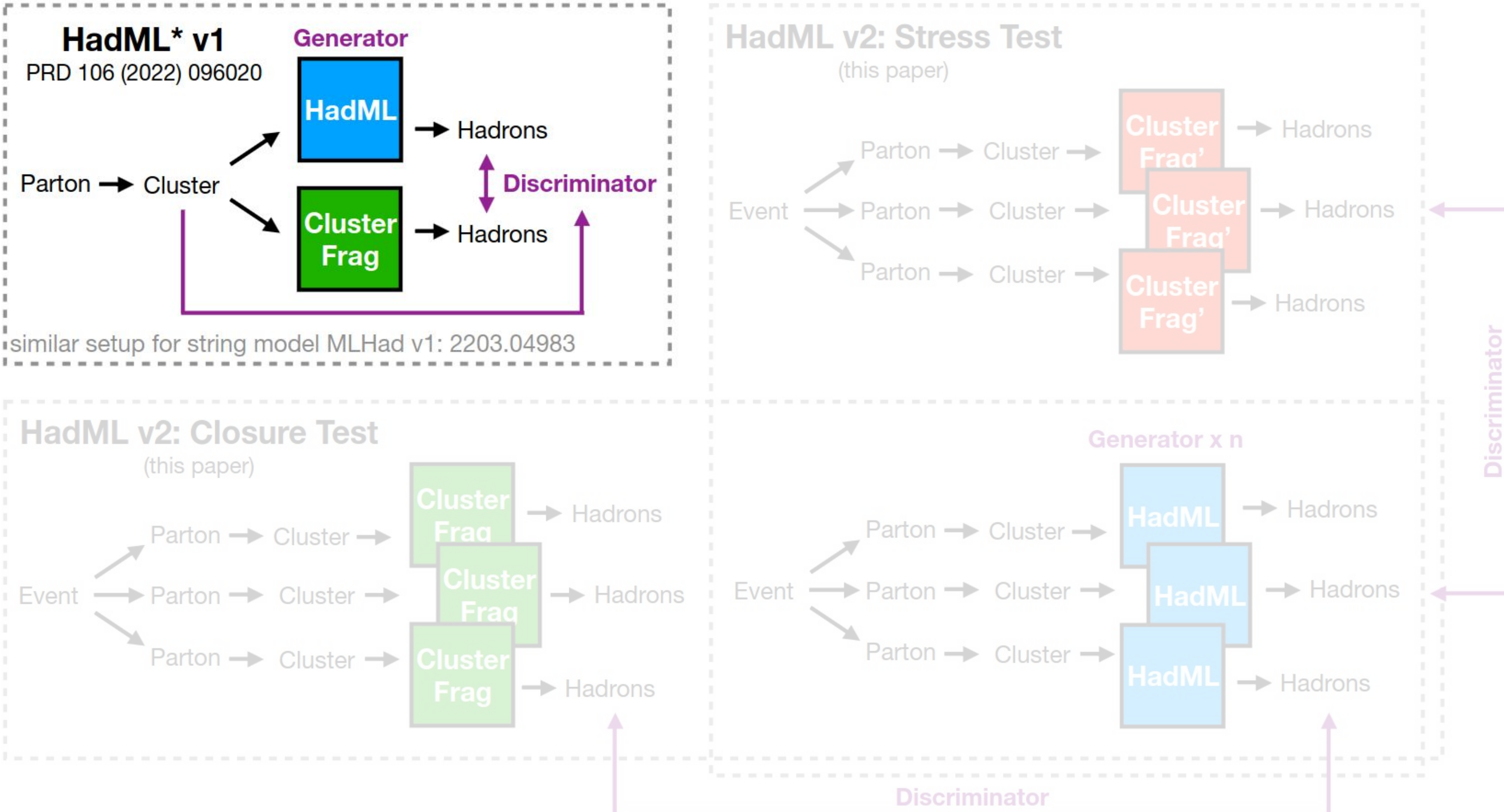


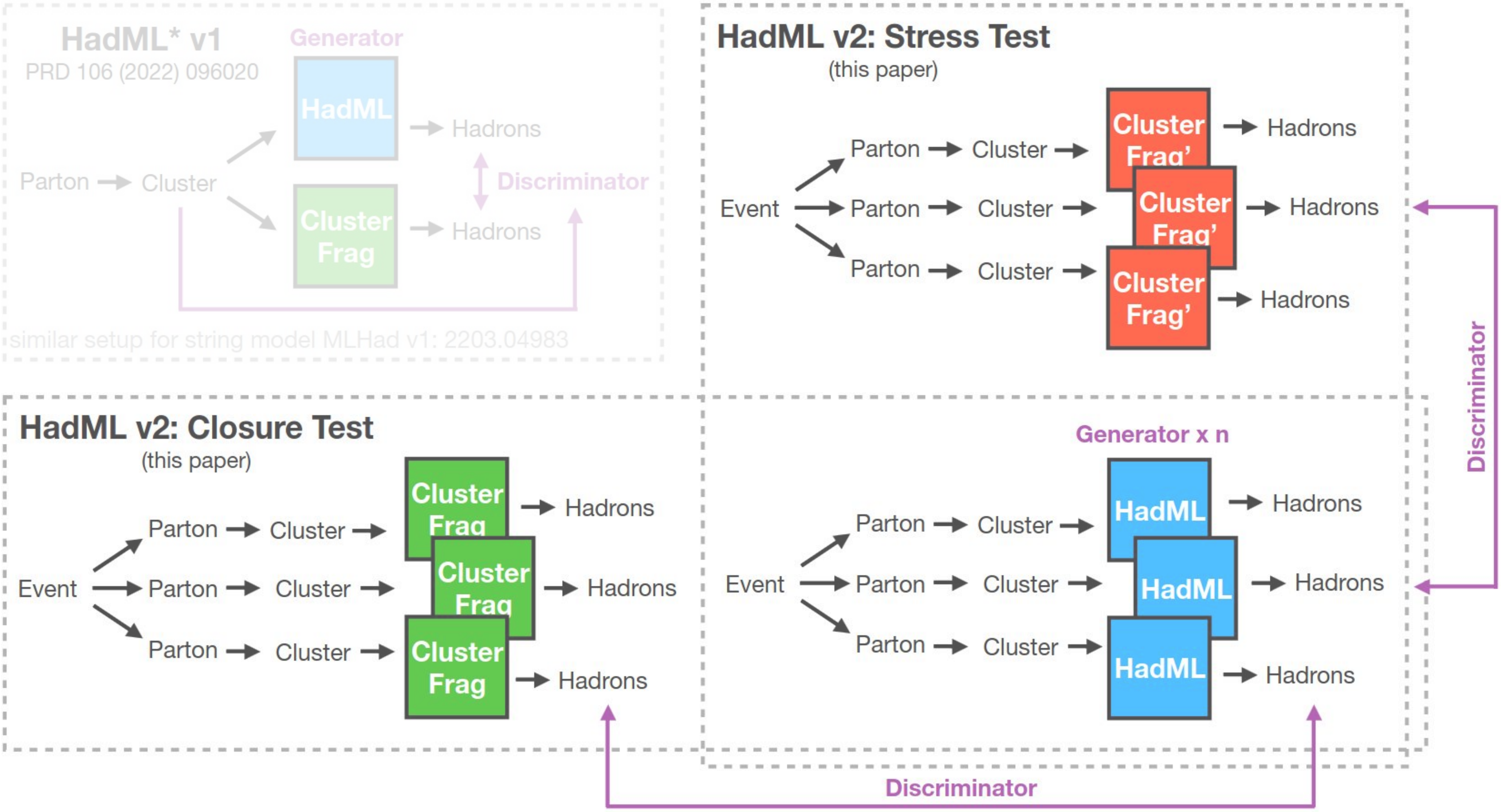




N.B. we have trained on H7, so we don't expect to be any better than it at modeling the data.

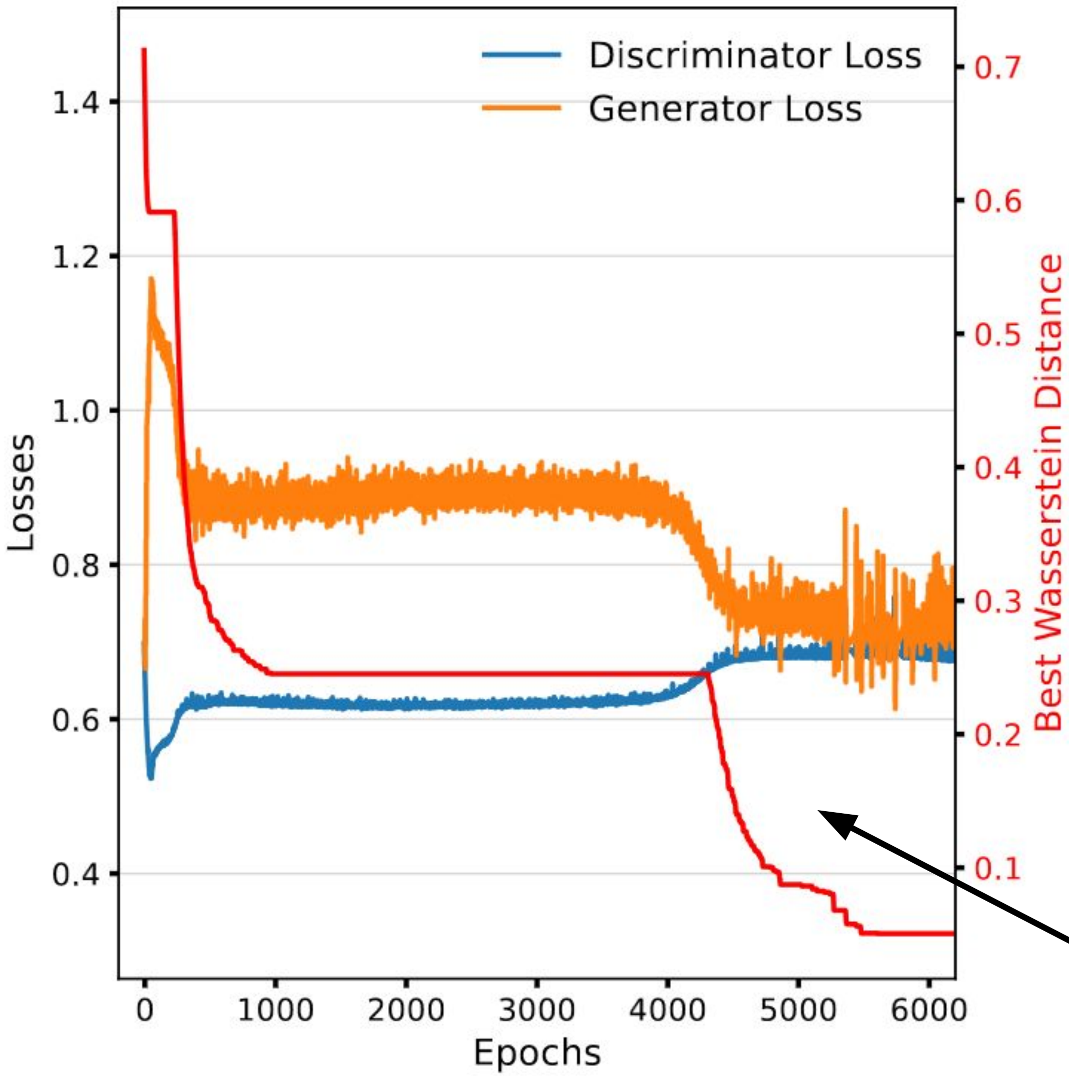
Road map for today





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.

Training HADML v2

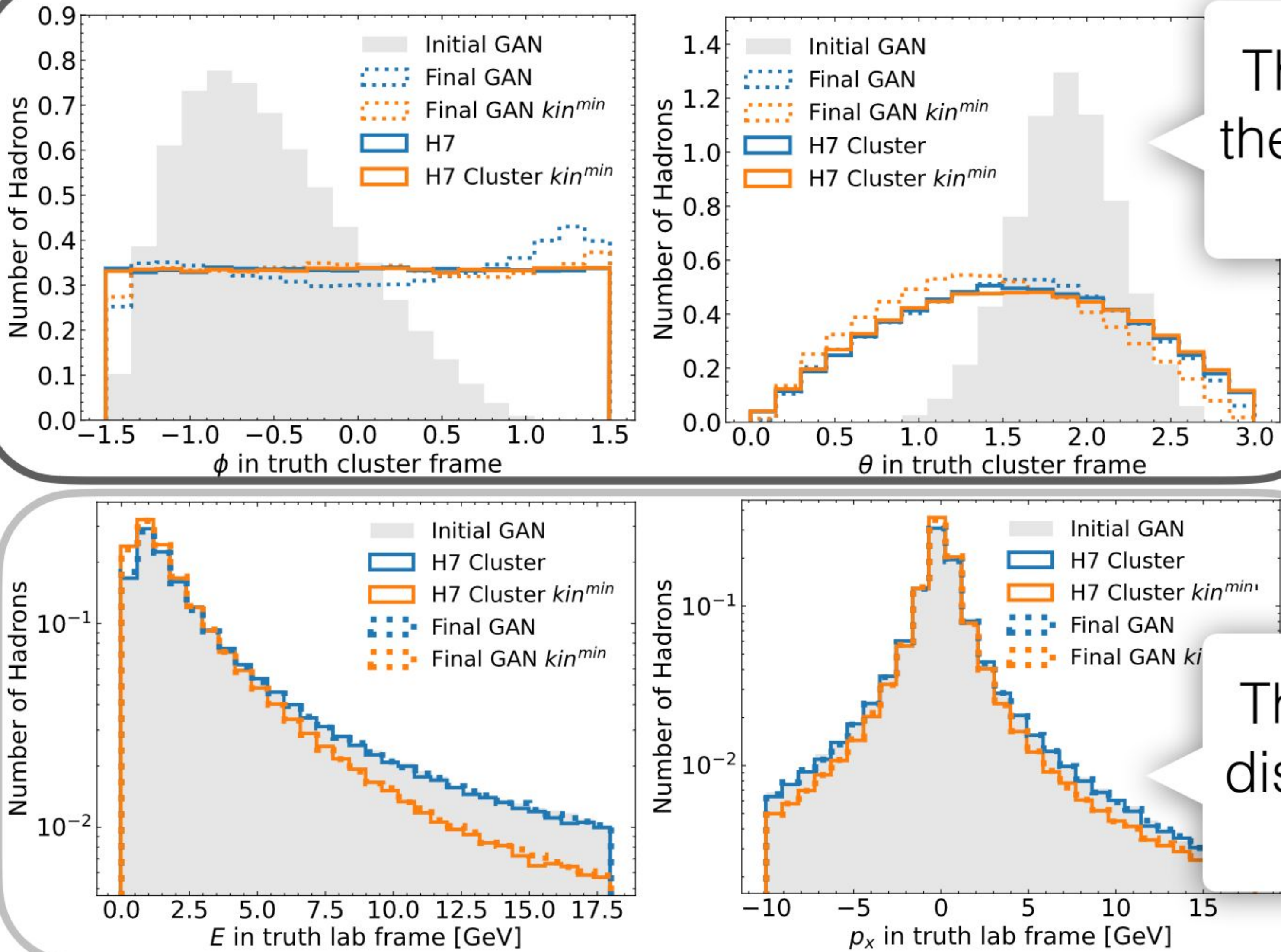


Now, the generator is local (per cluster), but the discriminator is global (whole event).

Discriminator is a permutation-invariant architecture called Deep Sets.

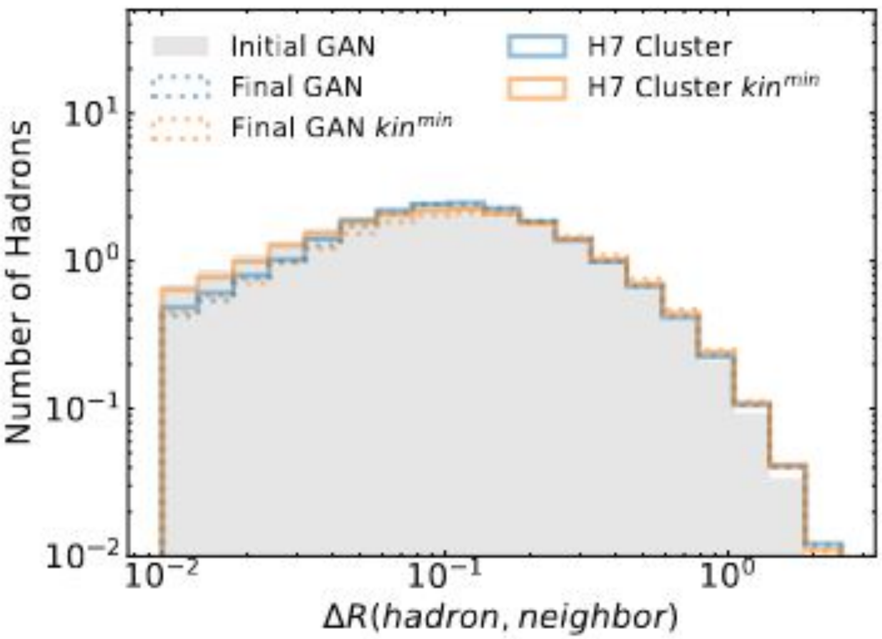
Simplification only Pions

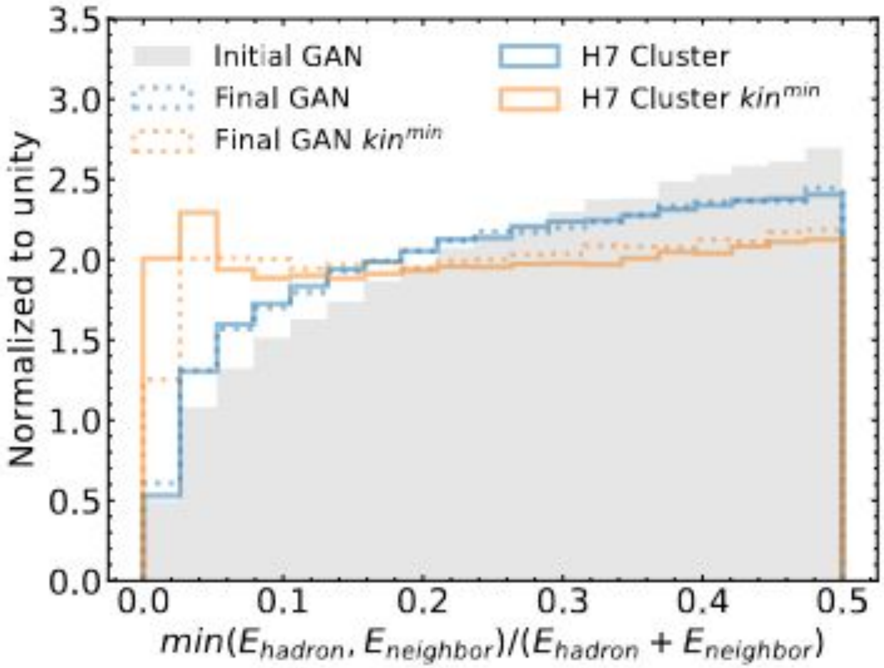
Still works !



This is what the generator "sees"

This is what discriminator "sees"



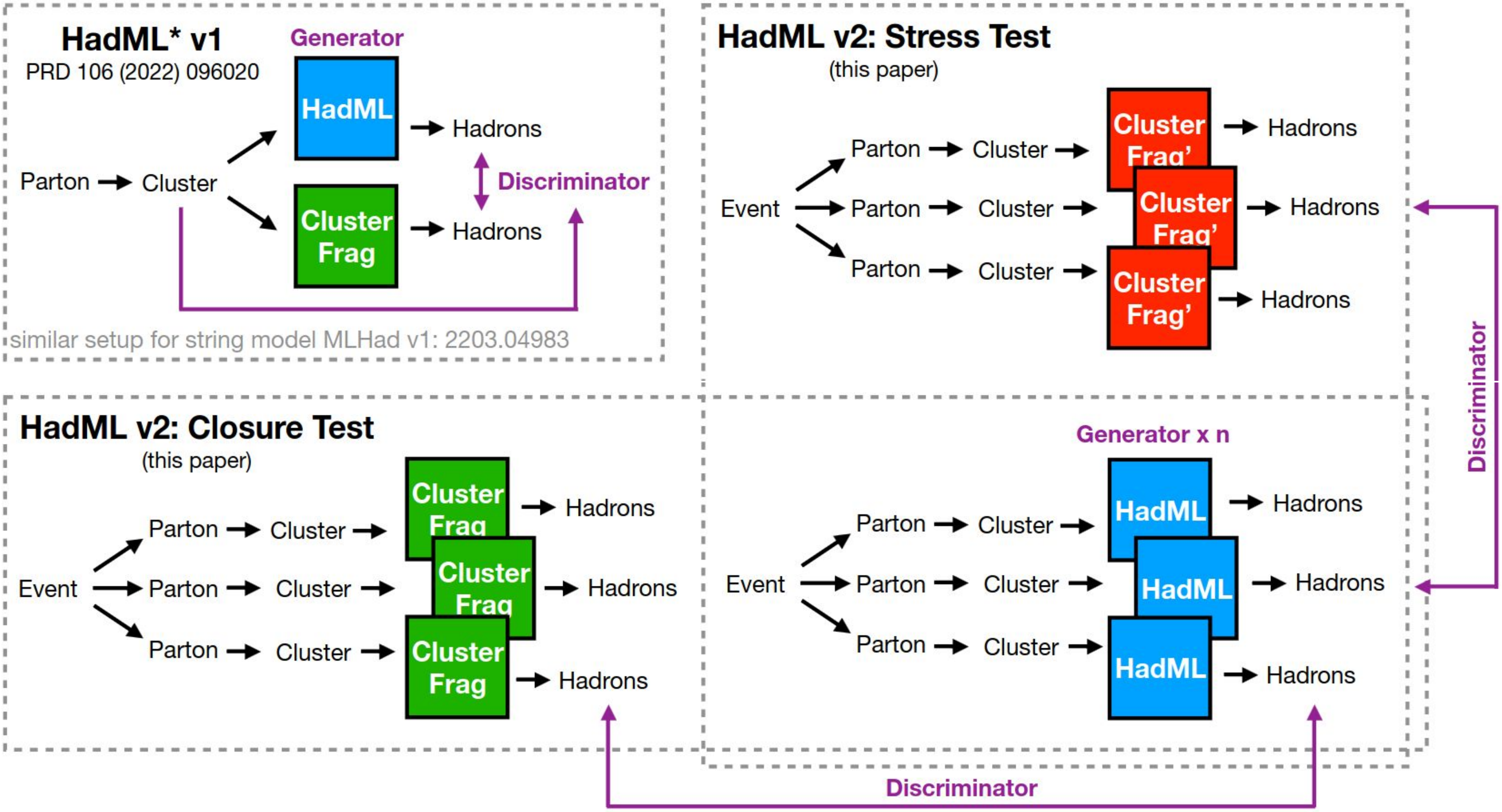


MINIMAL $\Delta R^2 = \Delta\phi^2 + \Delta\eta^2$

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

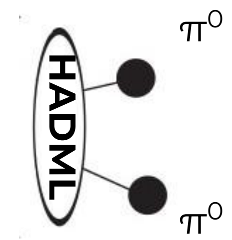
The approach could also be used to fit (without binning) data to a parametric physics model (for example cluster) as well.
However, this would require making the cluster model differentiable.

Summary


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
Parton
Cluster
Cluster Frag'
Hadrons
Parton
Cluster
Cluster Frag'
Hadrons
Discriminator
HadML v2: Closure Test
(this paper)
Event
Parton
Cluster
Cluster Frag
Hadrons
Parton
Cluster
Cluster Frag
Hadrons
Parton
Cluster
Cluster Frag
Hadrons
Generator x n
Event
Parton
Cluster
HadML
Hadrons
Parton
Cluster
HadML
Hadrons
Parton
Cluster
HadML
Hadrons
Discriminator

Outlook

- First ML hadronization models: HADML and MLHAD
- Recent progress:
  -HADML: “Data fitting protocol”
  - MLHAD: “Reweighting Monte Carlo Predictions and Automated Fragmentation Variations” - see P. Skands talk

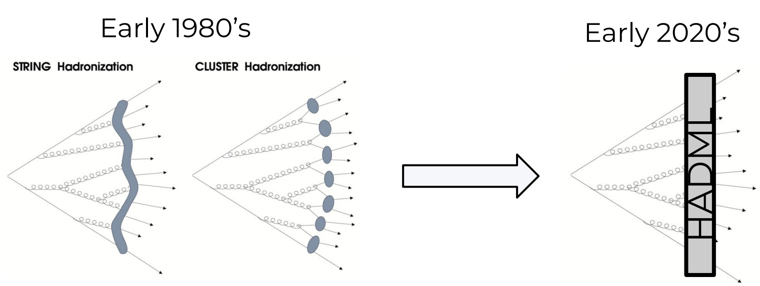


We have made significant progress,
but there are still multiple steps to build and tune a full-fledged hadronization model.

What is next for HADML?

- Number of technical and methodological step needed:
  - ➔ Directly accommodate multiple hadron species with their relative probabilities
  - ➔ Hyperparameter optimization, including the investigation of alternative generative models
  - ➔ More flexible model with a capacity to mimic the cluster or string models as well as go beyond either model.

There is still a multi-year program ahead of us!



So Stay tuned!

Advertisement

A postdoc in ML/HEP position

JAGIELLONIAN UNIVERSITY
IN KRAKÓW

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

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.

Wasserstein distance

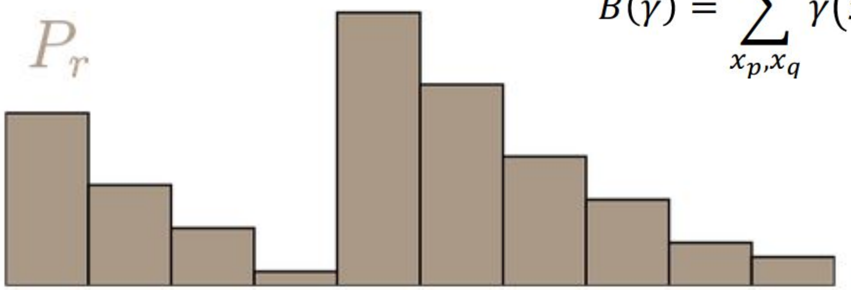
The Wasserstein distance

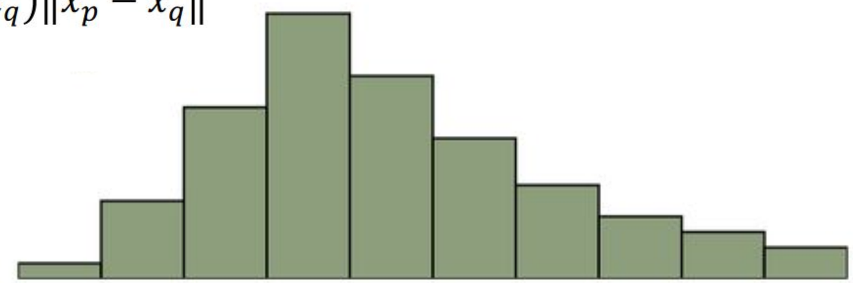
- For discrete probability distributions, the Wasserstein distance is called the earth mover's distance (EMD):
- EMD is the minimal total amount of work it takes to transform one heap into the other.

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

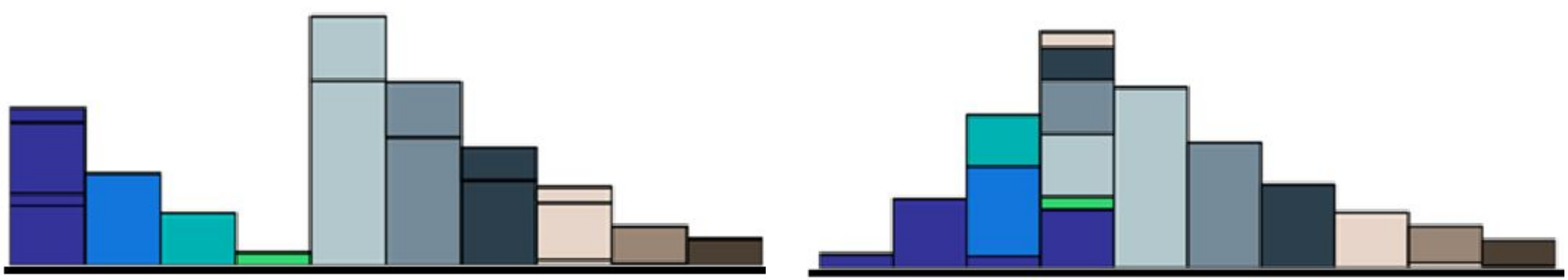
- Work is defined as the amount of earth in a chunk times the distance it was moved.

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





Best "moving plans" of this example



In the paper that introduced GANs, the generator tries to minimize the following function while the discriminator tries to maximize it:

$$E_x[log(D(x))] + E_z[log(1 - D(G(z)))]$$

In this function:

- `D(x)` is the discriminator's estimate of the probability that real data instance x is real.
- E_x is the expected value over all real data instances.
- `G(z)` is the generator's output when given noise z.
- `D(G(z))` is the discriminator's estimate of the probability that a fake instance is real.
- E_z is the expected value over all random inputs to the generator (in effect, the expected value over all generated fake instances G(z)).
- The formula derives from the cross-entropy between the real and generated distributions.

The generator can't directly affect the `log(D(x))` term in the function, so, for the generator, minimizing the loss is equivalent to minimizing `log(1 - D(G(z)))`.