Modeling Hadronization using Machine Learning

QCD@LHC 2022

Tony Menzo

PhD candidate, University of Cincinnati

In collaboration with:

Phil Ilten, Stephen Mrenna, Manuel Szewc, Michael Wilkinson, Ahmed Youssef, and Jure Zupan

Based upon work done in 2203.04983

Goals and outlook

The main motivation is to create a better simulation of collider events.

But also to promote a paradigm shift in the modeling of non-perturbative physics.

Present: In **2203.04983** we showed that machine learning techniques can be used to implement a model of hadronization using (artificial) training data

Near Future: Implement a machine learning-improved (i.e. data-improved) model of hadronization

Far future: Take what we've learned and develop BETTER theoretical models

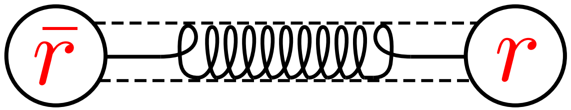
Event Generators

1. Hard proccess
2. Parton Showers
3. Hadronization → **Lund String Model**
4. Unstable particle decay

Early 80s brought many non-perturbative models: Cluster, percolation, ...

Lund String Model

(currently implemented in Pythia)



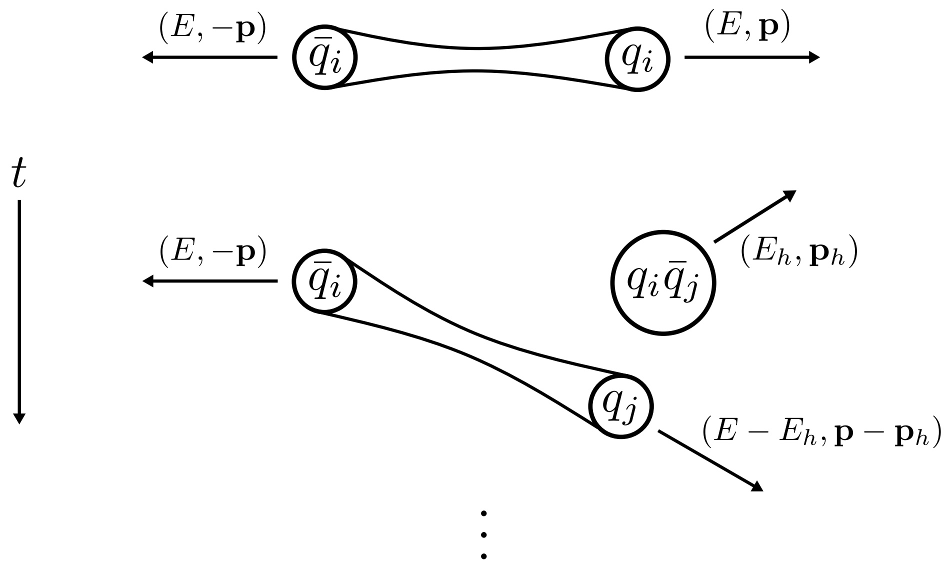
Stringy Hadronization

The momentum fraction z of each fragmenting hadron is sampled according to the

Lund fragmentation function

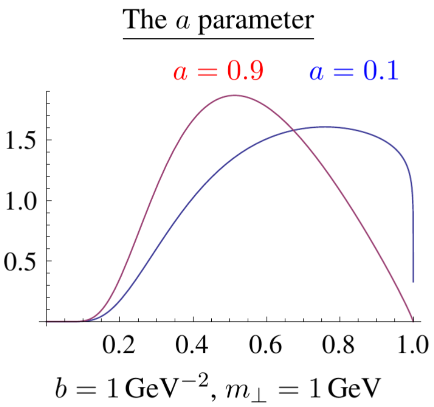
$$f(z) \propto \frac{(1-z)^a}{z} \exp\left(\frac{-bm_\perp^2}{z}\right)$$

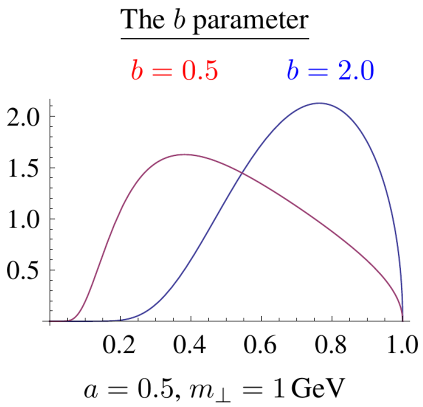
$$z = \frac{p_z + E_h}{2E}$$



Troubles in fit paradise

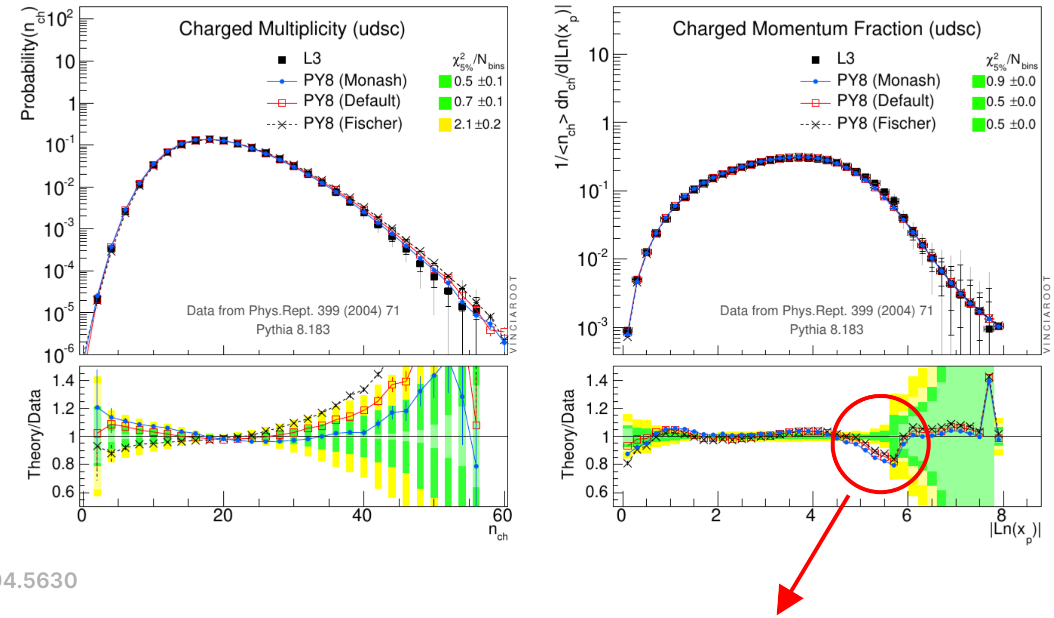
$$f(z) \propto \frac{(1-z)^a}{z} \exp\left(\frac{-bm_\perp^2}{z}\right)$$





1404.5630

Monash Tune (a,b,σ_{perp})

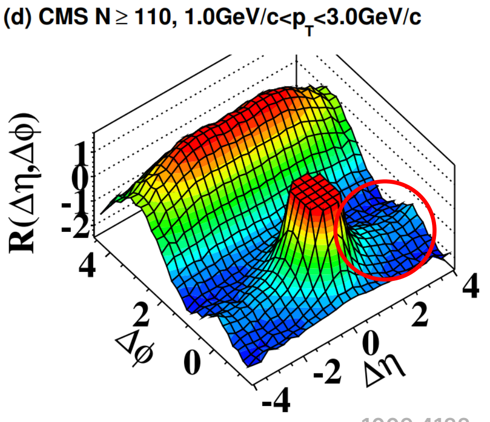


Cannot be improved by retuning

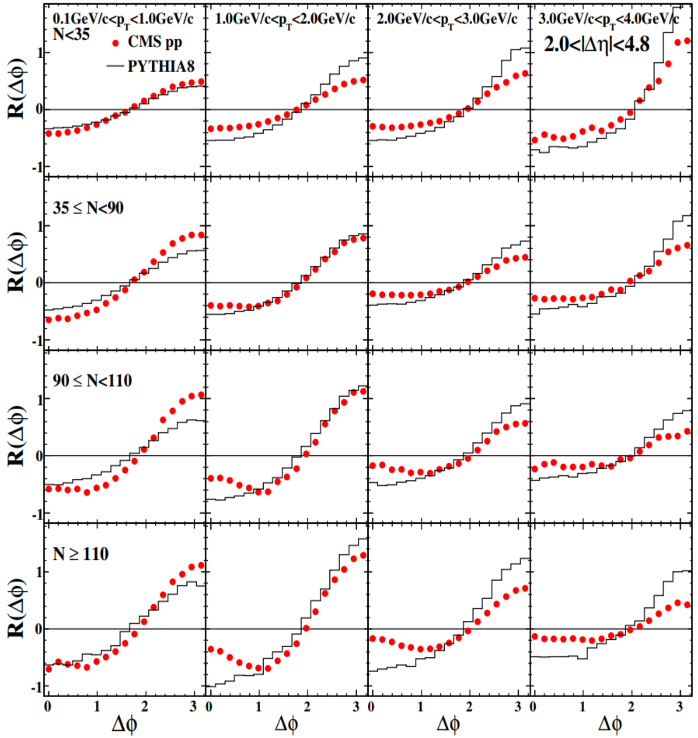
Some ~new disagreements for high multiplicty events…

Similar properties to heavy ion collisions:

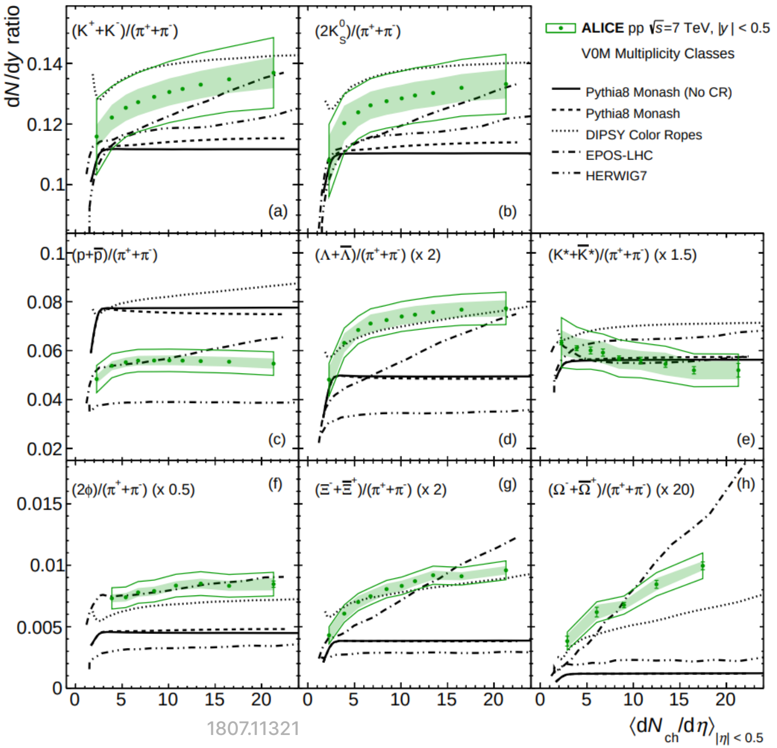
- “The ridge” i.e. enhanced particle production around the azimuthal angle of a trigger jet (CMS)



1009.4122



- Strangeness production increases as a function of event multiplicity (ALICE)



1807.11321

How to improve the generator: two* approaches

- Improve model
  - MPIs, rope hadronization, transverse mass supression, flavor asymmetries, hadronic rescattering, multiscale models (string → hydrodynamical), flavor selector, etc.
  - Utilize techniques from gauge-gravity duality

  Hard to come up with mathematically precise model without established calculational techniques

- Data-driven generator
  - Sample directly from global distributions

  Non-universal and extremely difficult to convert into representative particle flow data

*** or a combination of both (our approach)**

Hybrid approach

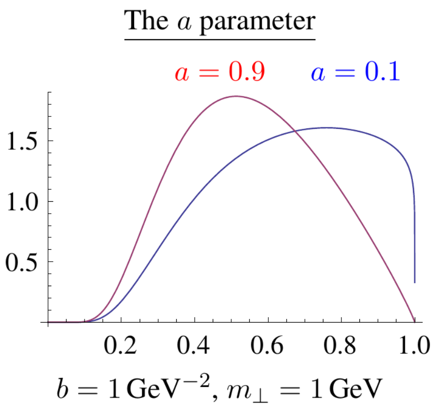
Hadronization models already do really well!

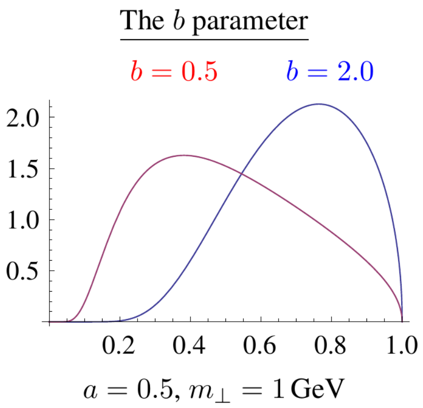
Model + **Experimental data**

↓

Complete (or at the very least better) phenomenological model of hadronization

For example, modify the fragmentation function f(z)...





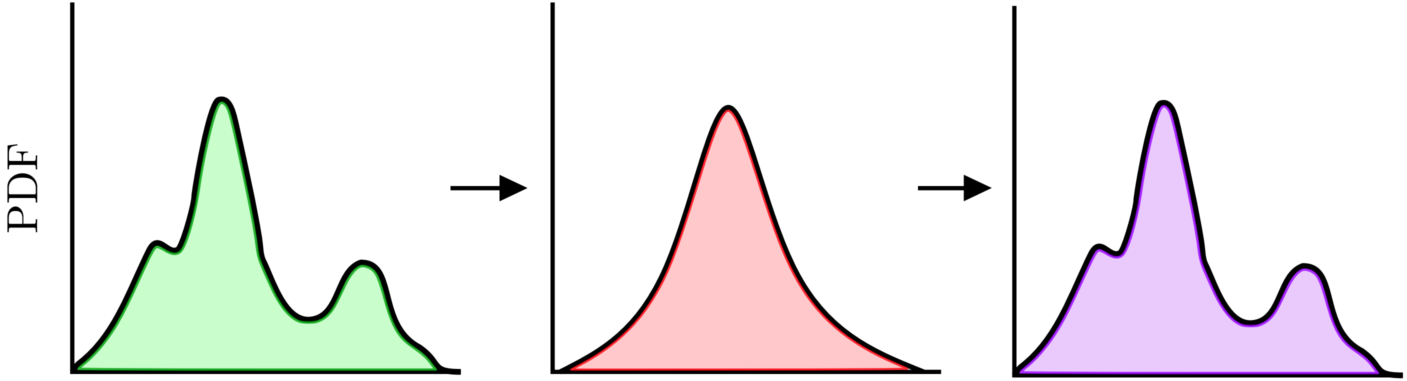
1404.5630

+ ϵ

Why machine learning?

To make any headway we need a tool which will allow us to efficiently sample probability distributions whose analytic form is unknown.

Generative machine learning algorithms are the perfect tool!



Proof of concept (2203.04983)

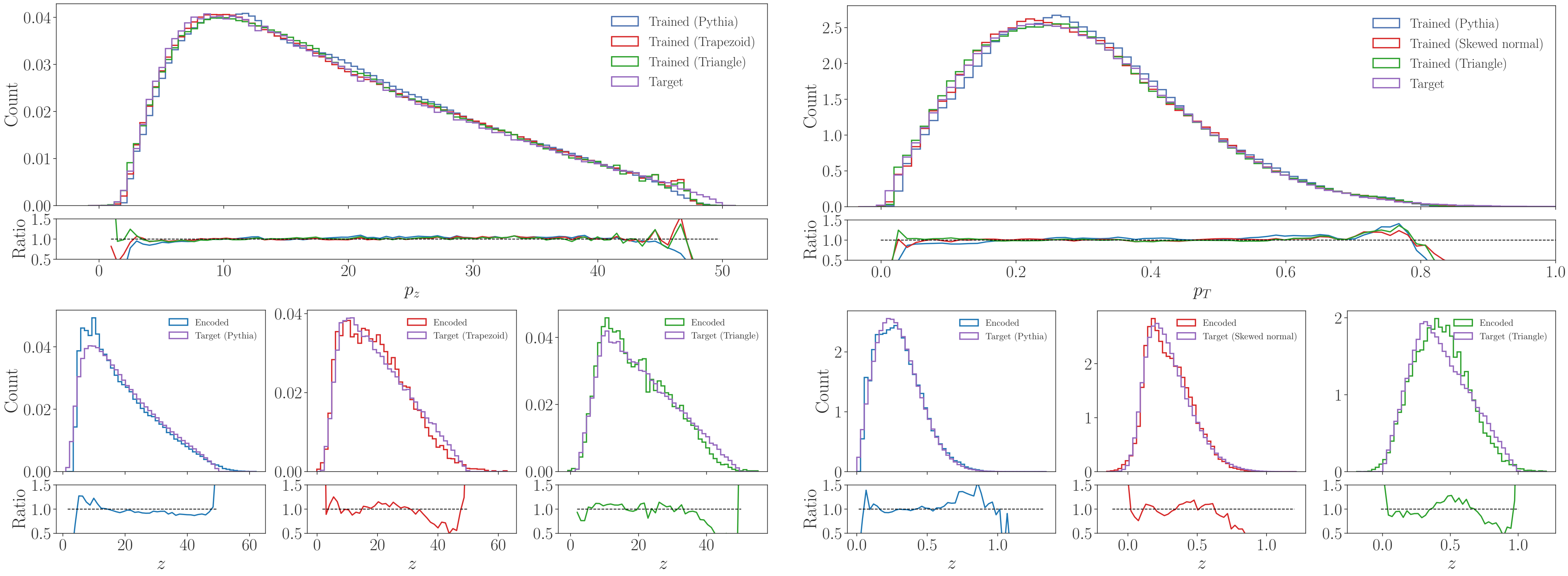
Consider Pythia output as 'experimental data' and try to reproduce hadronization observables by training on single emission kinematics (~learn the fragmentation function f(z)).

Start from simplest hadronizing system:

1. $q\bar{q} \rightarrow \pi$'s
2. Assume no correlations between emissions
3. E_{cut}~5 GeV (To avoid termination effects)

Train on p_z and p_T distributions of 1st emitted π

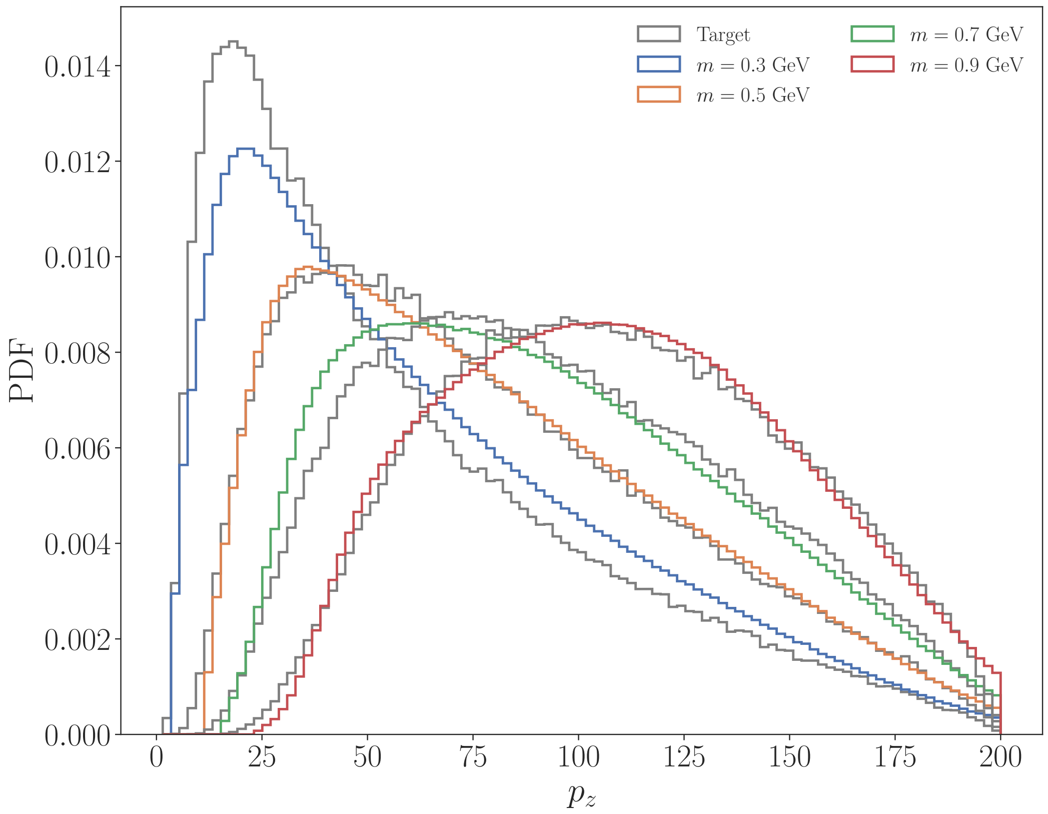
Training Results (cSWAE)

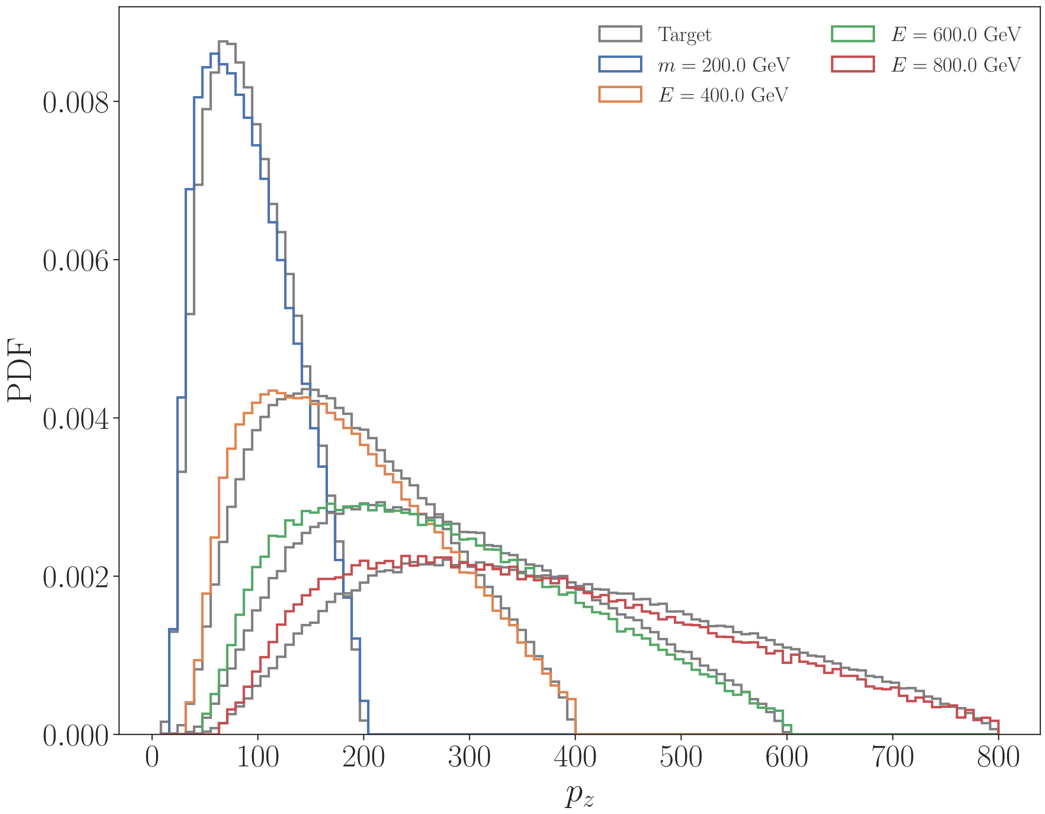


Training Results

(cSWAE with labels and boundaries)

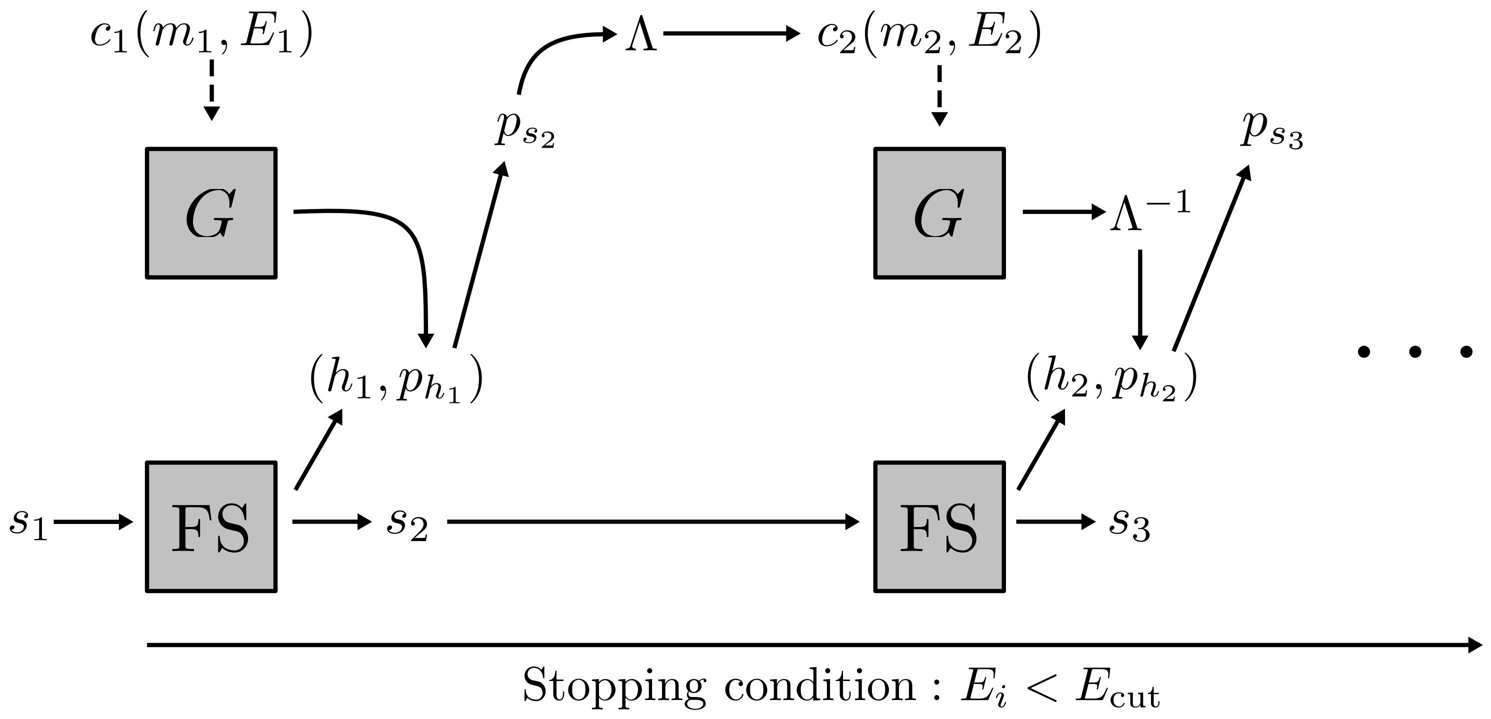
***Preliminary**



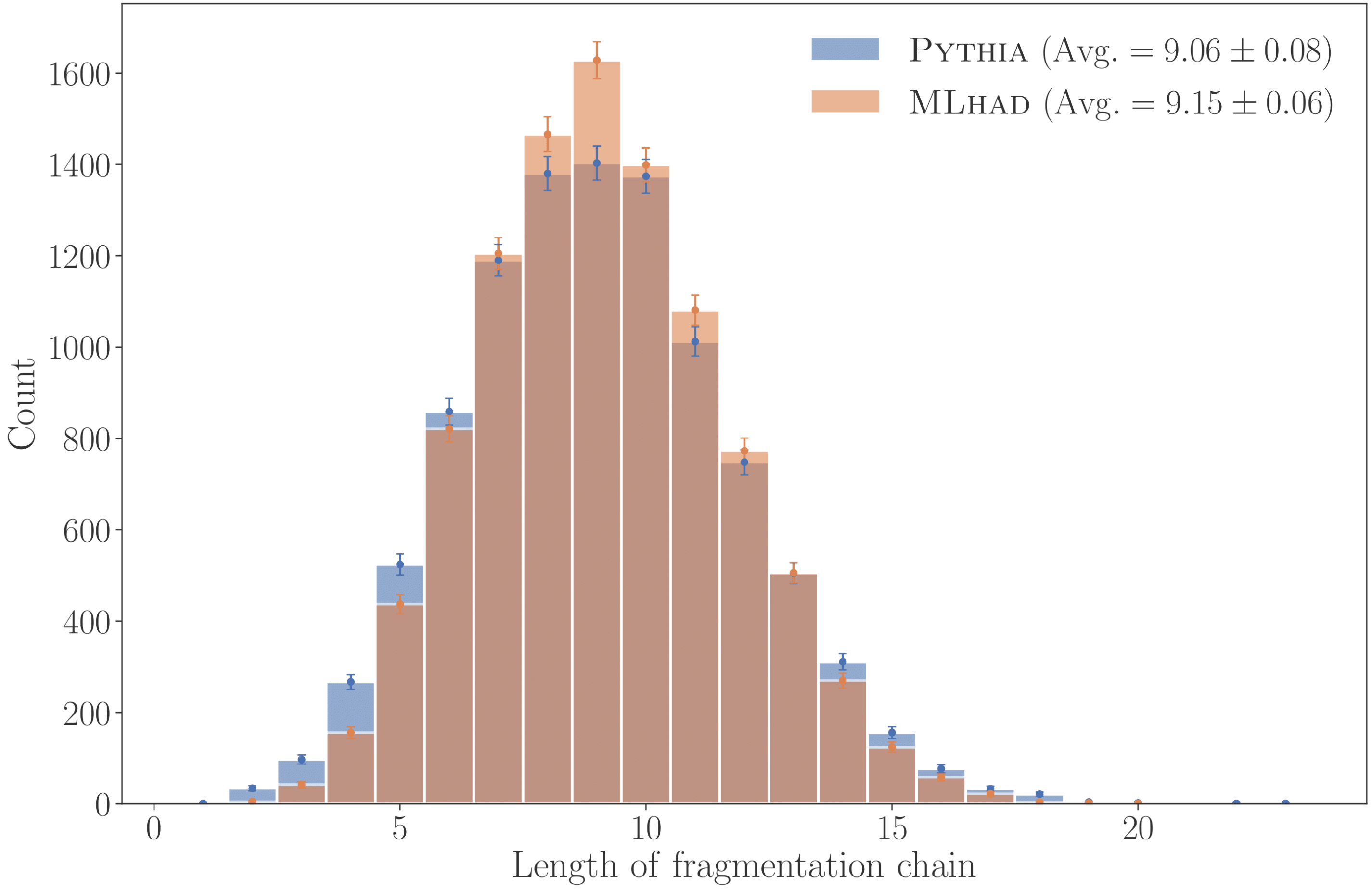


Hadronization (kinematics + flavor selector)

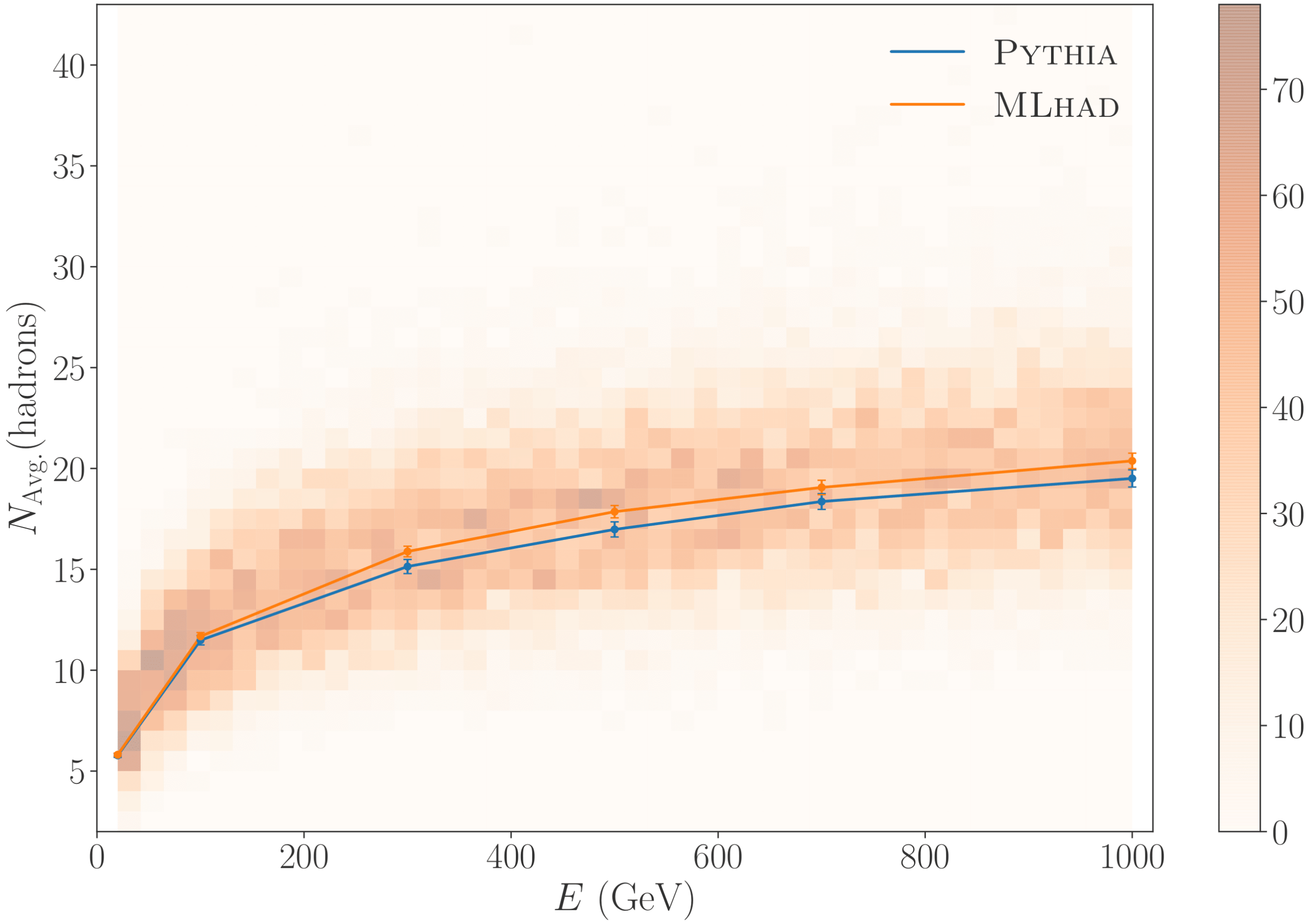
The trained model distributions now need to be integrated into a chain of fragmentations



Global observable (Hadron multiplicity cSWAE)



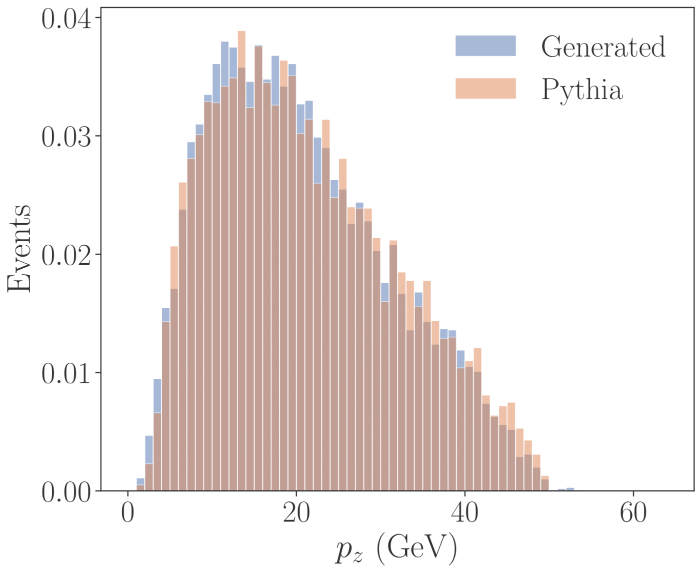
Global scaling (Hadron multiplicty vs string energy cSWAE)

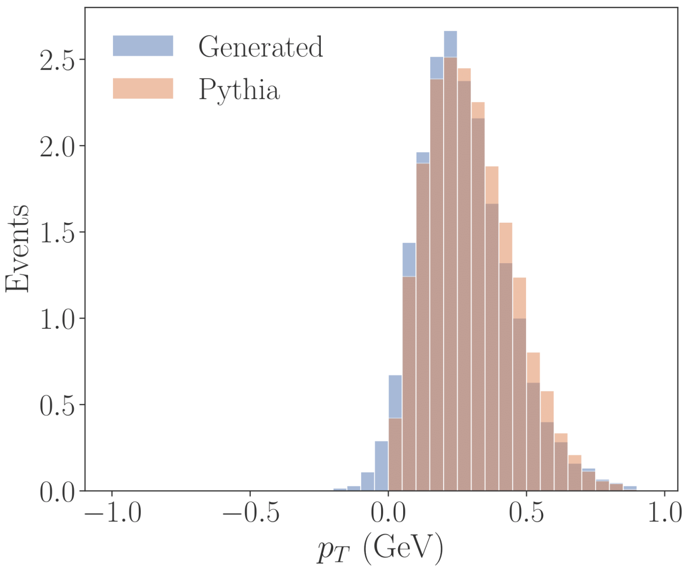


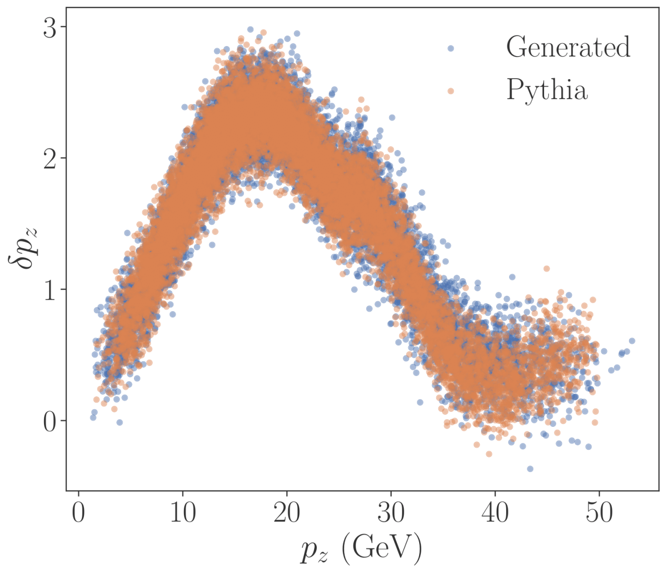
Error estimation (BINN)

Incorportating (theoretical/experimental) errors from training dataset errors into the hadronization simulation

***Preliminary**







Conclusion

Model + machine learning methods CAN be used to implement hadronization.

What's next:

- **ML-improved (data-improved) model of hadronization**
- **ML flavor selector**
- **Error estimation**
- **Much more**

Check out our repo!

MLHAD

https://gitlab.com/uchep/mlhad

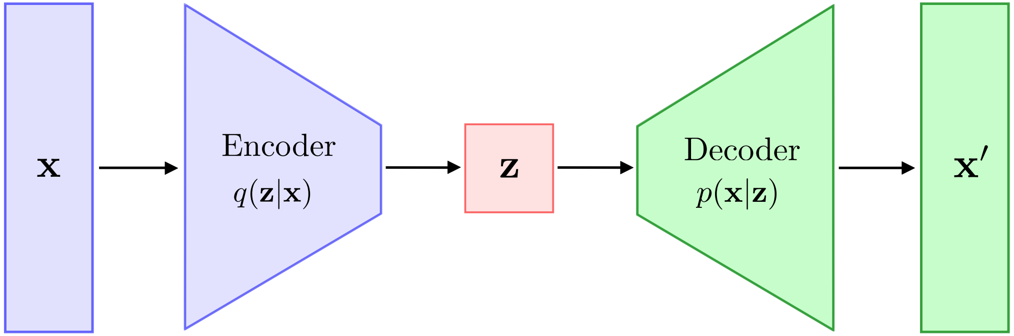
Check out our paper!

arXiV: 2203.04983

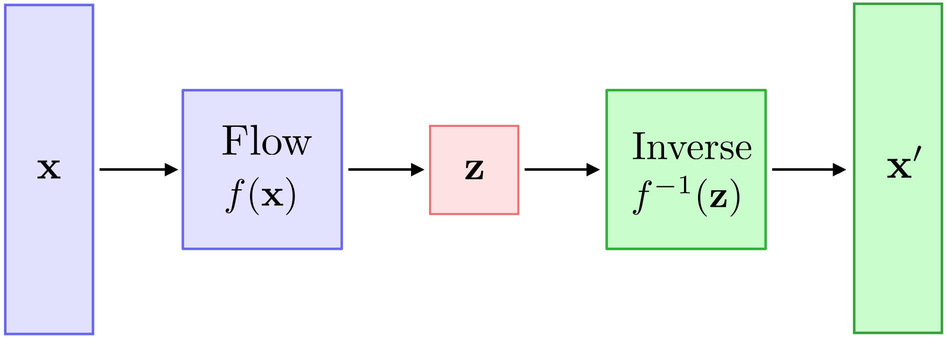
Back-up

Architectures

Conditional sliced-Wasserstein Autoencoder (cSWAE)



Conditional normalizing flow (cNF)



Training Results

(cNF with labels)

***Preliminary**