Connections Experiment-Theory

Pia Zurita (University of Regensburg) and Markus Diefenthaler (Jefferson Lab)

How to fully take advantage of AI/ML in physics studies for the EIC?

For the EIC community: A rigorous benchmarking process

EIC analyses should be based on rigorous benchmarking that allows for quantitative comparison of different approaches.

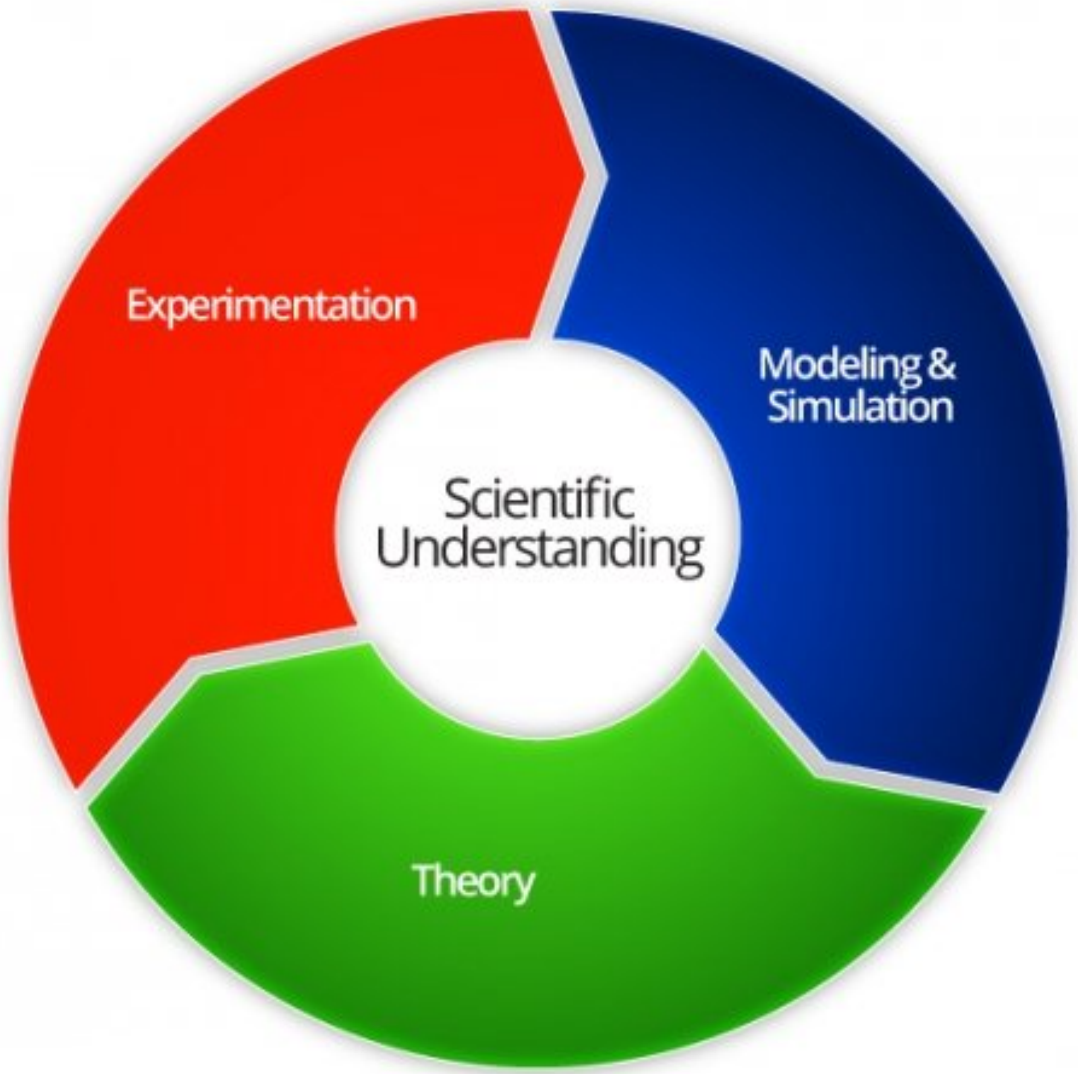
In our DVES example:

From by Simonetta Liuti (UVa)

- Physics Benchmarks
  - number and type of CFFs
  - Q^2 dependence of the cross section and observables: kinematic terms, PQCD evolution (LO, NLO, NNLO), and dynamical beyond LO terms (higher twists)

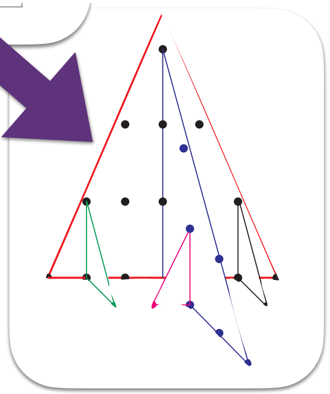
- Machine Learning Benchmarks
- ML architectures hyperparameters (number of layers, size of hidden nodes, activation functions, drop-out rates, loss functions, gradient descent methods)
- Features specific to data-centric analysis (feature selection and transformation, data augmentation, data synthetics, and data cleansing)
- Uncertainty Quantification
  - Inherent statistical fluctuations in physics (statistic)
  - Errors inherent from measurement system (systematic)
  - Errors in ML models
  - Errors in the training procedure

Monte Carlo Event Generators

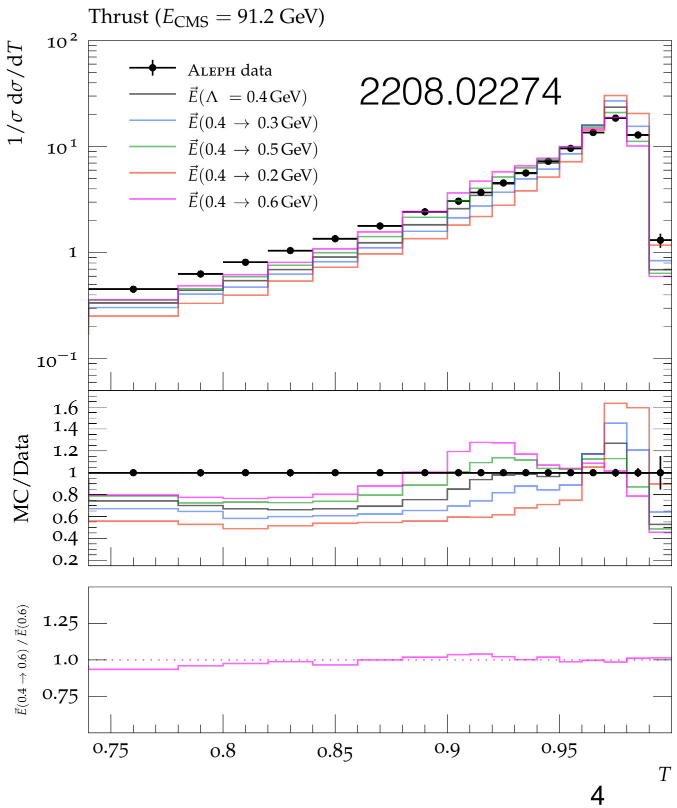


Ben Nachman (LBNL) – Towards a differential parton shower

Full parton shower is a bit tricky since variable (unbounded) number of random numbers. Let's start with "Discrete QCD" where the number is fixed.



Images from 2207.10694; algorithm from Nuclear Physics B 463 (1996) 217



As a first test, we show how this can be used to extract the strong coupling constant.

All of these samples have the same random numbers!

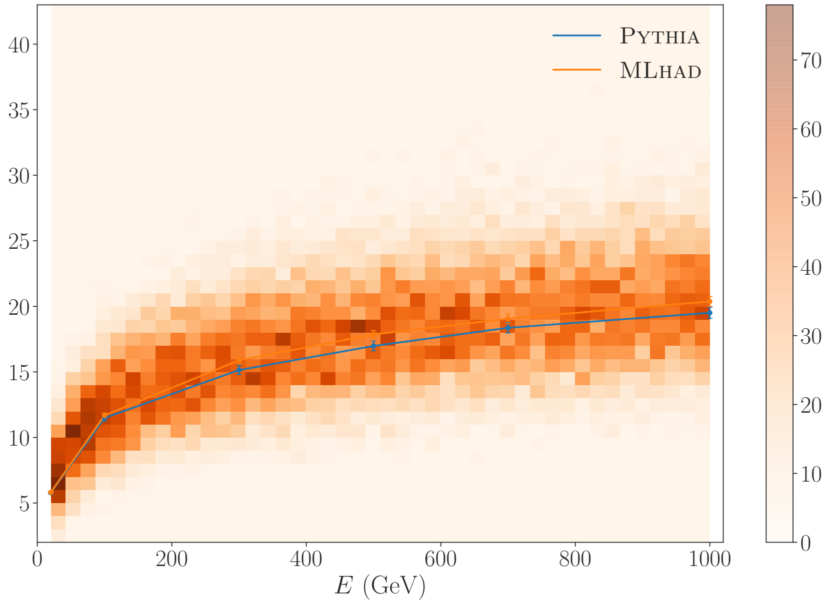
Tony Menzo (Cincinnati) – Modeling Hadronization using ML

Pythia / Lund String Model

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 ☺



Hadron multiplicty vs string energy cSWAE

Check out our code!

MLHAD

https://gitlab.com/uchep/mlhad

Check out our paper!

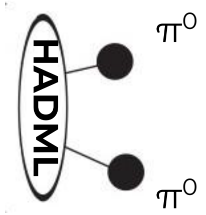
(Recently accepted for publication in SciPost Physics)

arXiV: 2203.04983

Andrzej Siódmok (Jagiellonian) – Modeling Hadronization Using ML and the Cluster Model

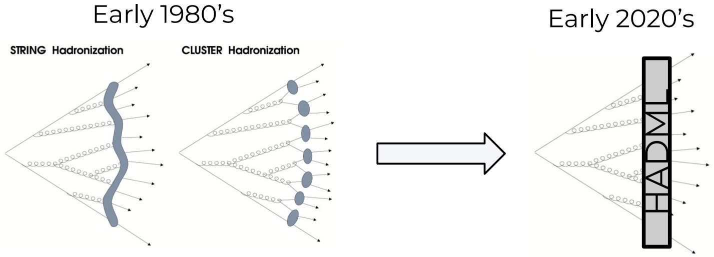
Summary

- We presented **first step** on the path towards a **neural network-based hadronization model**
- We emulated cluster hadronization model from Herwig with a GAN (**HADML**)
- HADML is designed to reproduce the two-body decay of clusters into pions
- The kinematic properties of other hadrons are emulated using the pion model and conservation of energy.
- HADML is able to reproduce Herwig's light cluster decays
- Integrated with the full Herwig simulation is able to reproduce results from LEP data

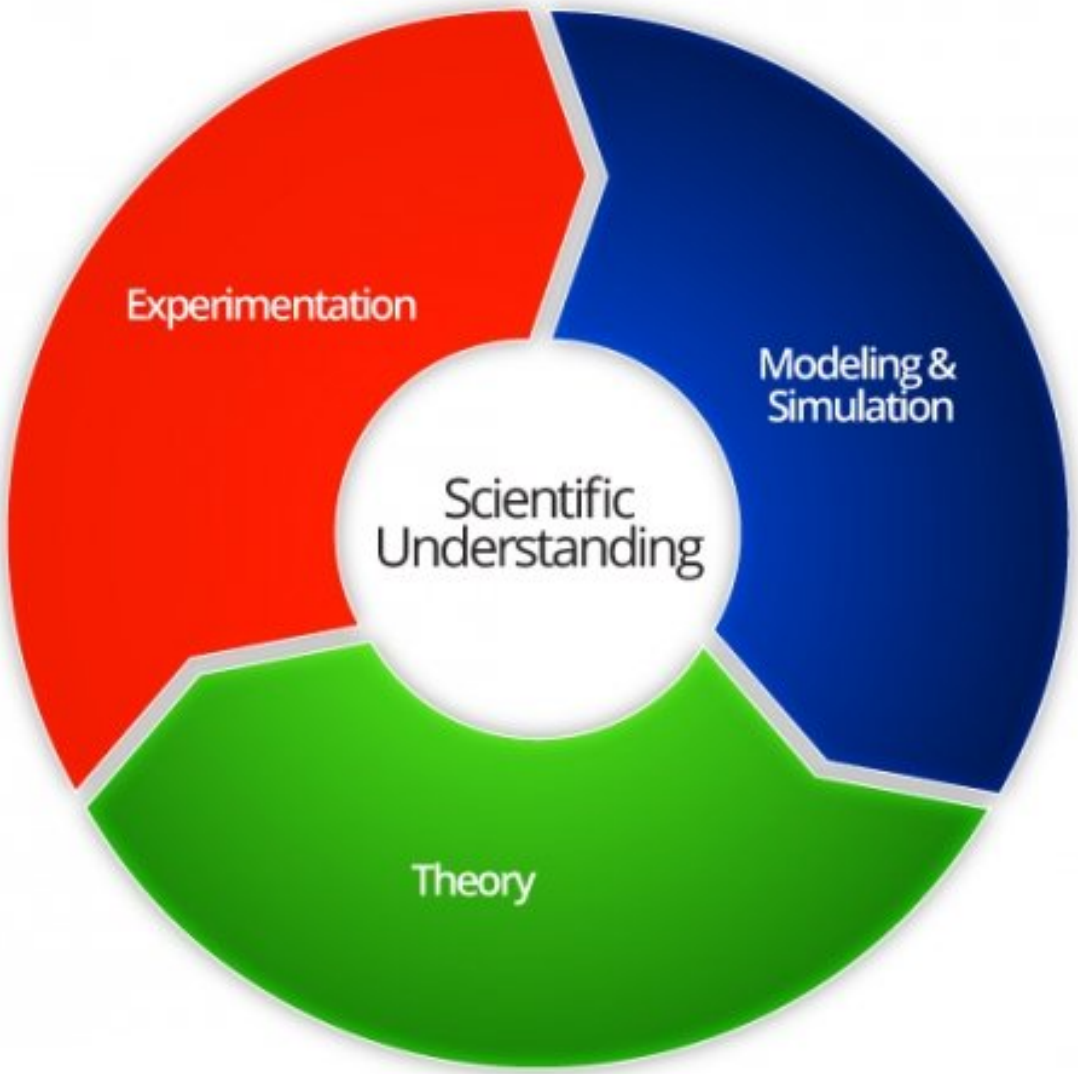


Outlook

- The ultimate goal of is to train the ML model directly on data to improve hadronization models
- Number of technical and methodological step needed:
  - ➔ Directly accommodate multiple hadron species with their relative probabilities
  - ➔ Heavy cluster decays
  - ➔ Hyperparameter optimization, including the investigation of alternative generative models
  - ➔ Methodological innovation is required to explore how to tune the model to data



Experimental-Theoretical Workflows



A(I)DAPT := AI for Data Analysis and PreservaTion

We created GAN event generators for:

DIS — closure test of an unfolding architecture for the first time;
CLAS data — captured intricate detector acceptance behavior and multi-particle correlations.

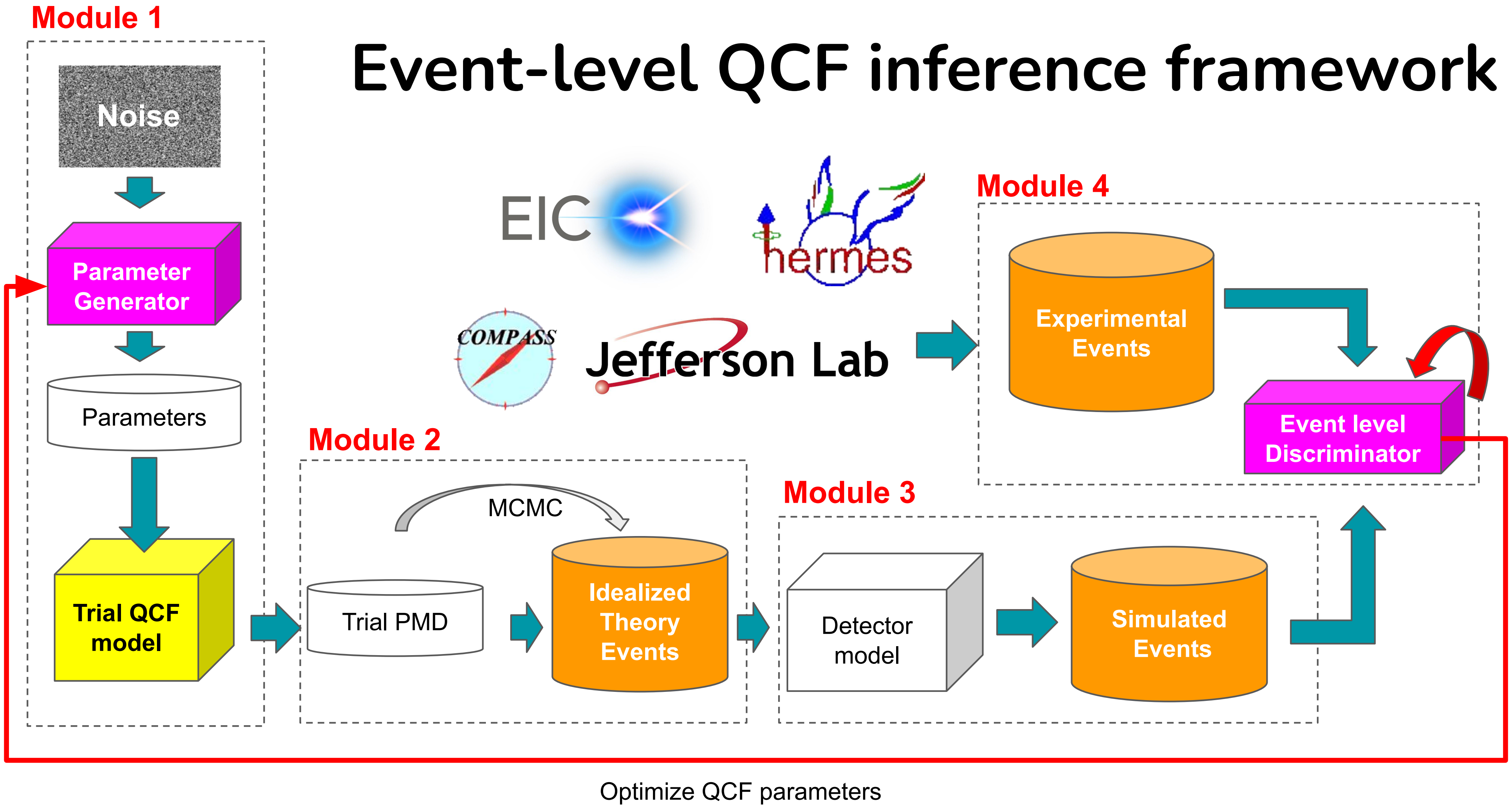
We are currently analysing:

– unfolding of CLAS data to extract physics at vertex level;
– information from hidden layers;
– uncertainty quantification metrics and physics validation.

In future, these benchmarks will allow for

– powerful minimum-bias interpolation tools;
– physics extraction and amplitude analysis from synthetic vertex-level data.

Event-level QCF inference framework



Summary

Pia Zurita (University of Regensburg
Markus Diefenthaler (Jefferson Lab)

D. Geesaman (NSAC) **“***It will be* ***joint progress of theory and experiment*** *that moves us forward, not in one side alone”*

- *We have ideas for new workflows on how to connect experiment and theory better.*
- *We have ML approaches for how to accelerate and improve simulations.*

Ben Nachman (LBNL) on Differentiable Simulations

- Thinking of our simulations as differentiable is a new and powerful paradigm.
- We can do optimal inference and run our codes on accelerators “for free”.

Bottleneck, as usual: Lack of open data from experiments