Studying Hadronization by Machine Learning Techniques

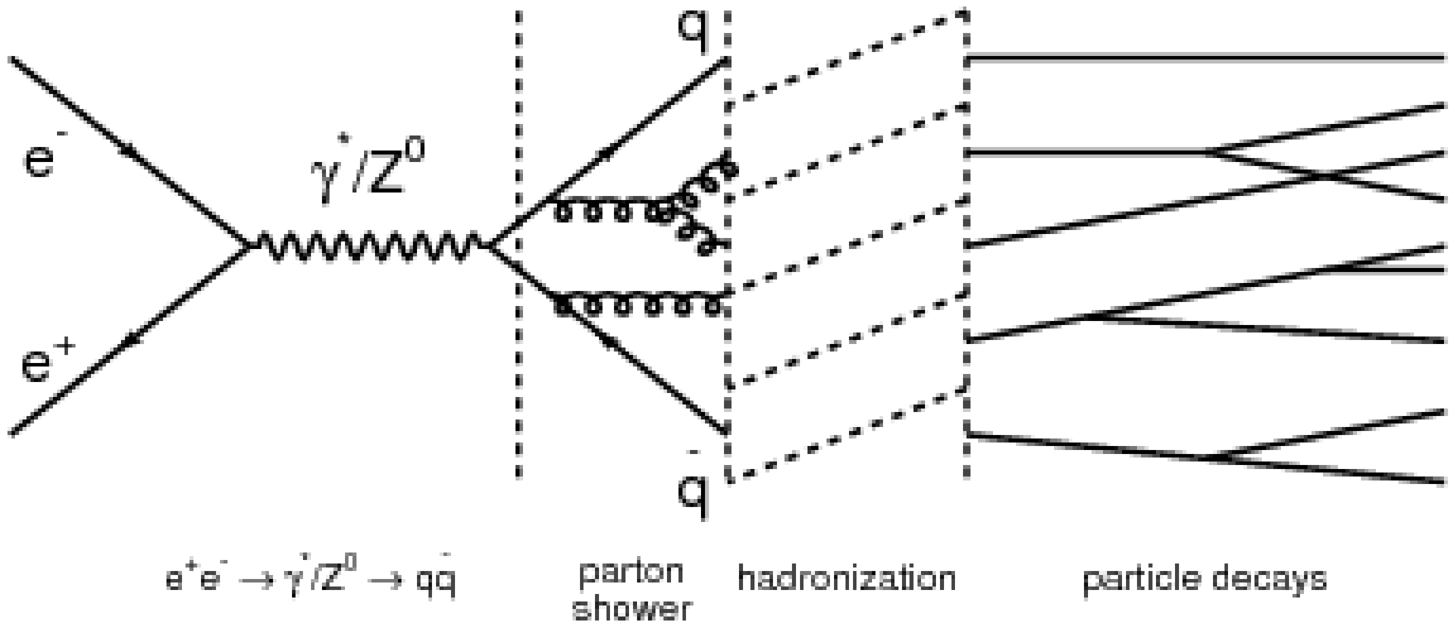
G.G. Barnaföldi, G. Bíró (Wigner FK), and B. Tankó-Bartalis (University of Oxford)

Support: *Hungarian OTKA grants, NK123815, K135515, Wigner Scientific Computational Laboratory*
Ref: arXiv:2111.15655

SQM2022, Busan, South Korea, 14th June 2022

Modeling hadronization in e^+e^- collisions

Final state processes & hadronization



Hadronization models – history

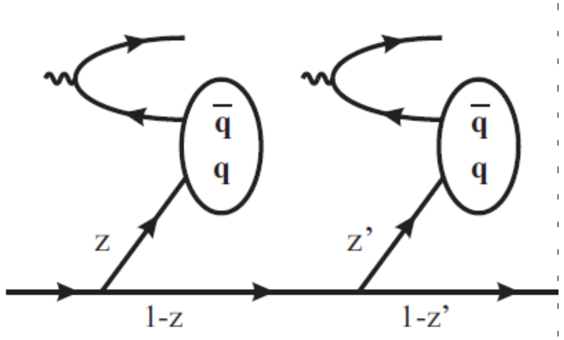
The evolution of hadronization models

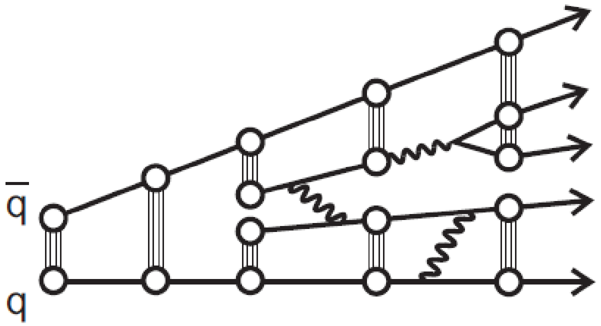
Feynman-Field

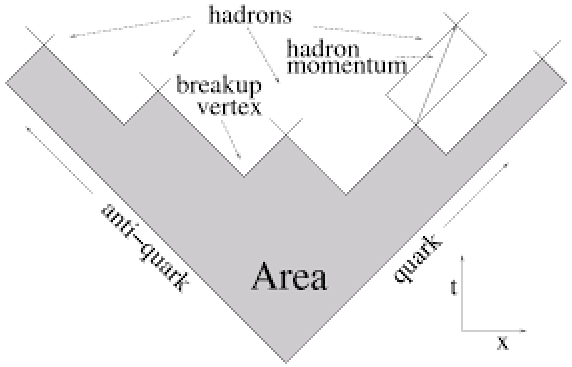
pair production

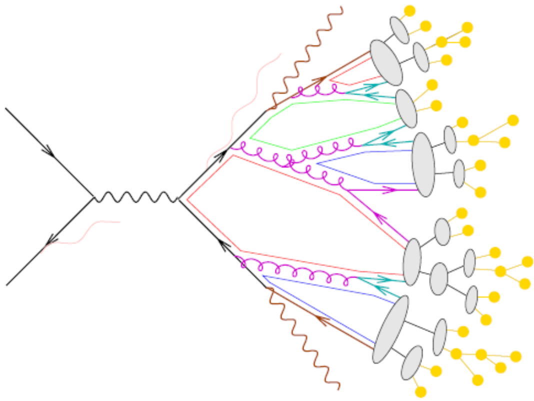
Lund model

cluster model









$$f(z) \propto \left[z \left(1 - \frac{1}{z} - \frac{\epsilon}{1-z} \right)^2 \right]^{-1}$$

$$f(z) \propto z^{-1}(1-z)^a \cdot \exp\left(\frac{-b\, m_T^2}{z} \right)$$

pQCD models

Non-pQCD models

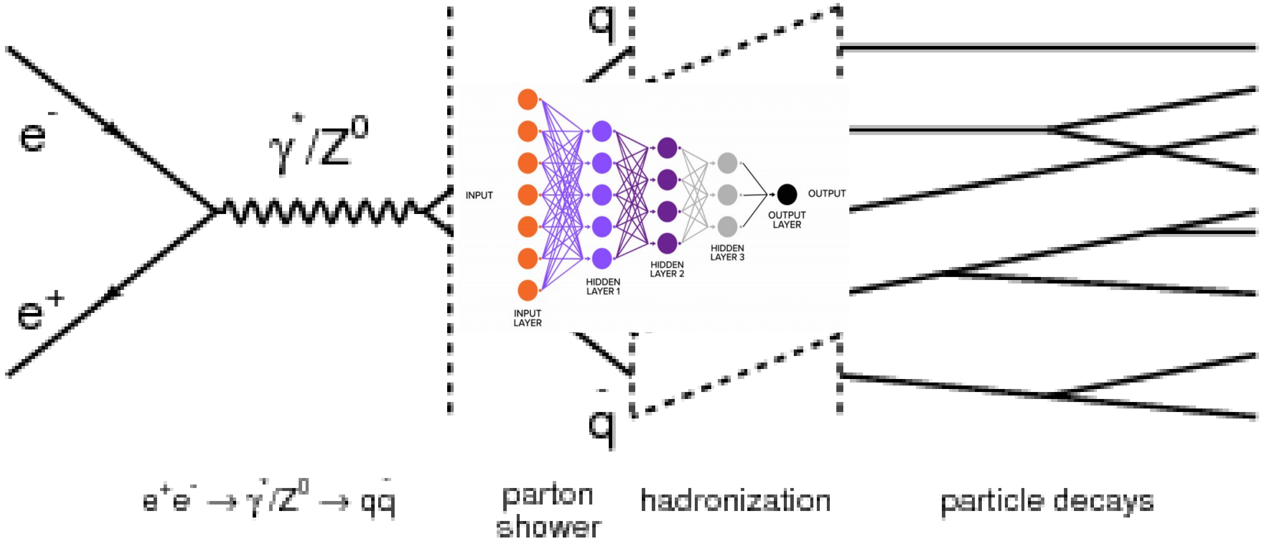
PYTHIA/HIJING

HERWIG

Idea & motivation

Three key layers

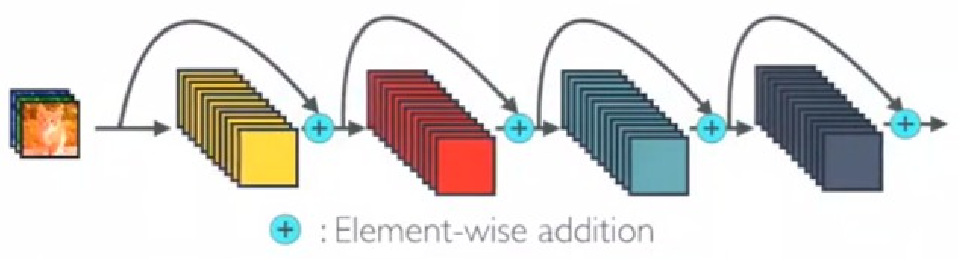
- **Input**: Takes the features
- **Hidden layers**: Connects to each neuron through different weights
- **Output**: Gives the result as a number or class

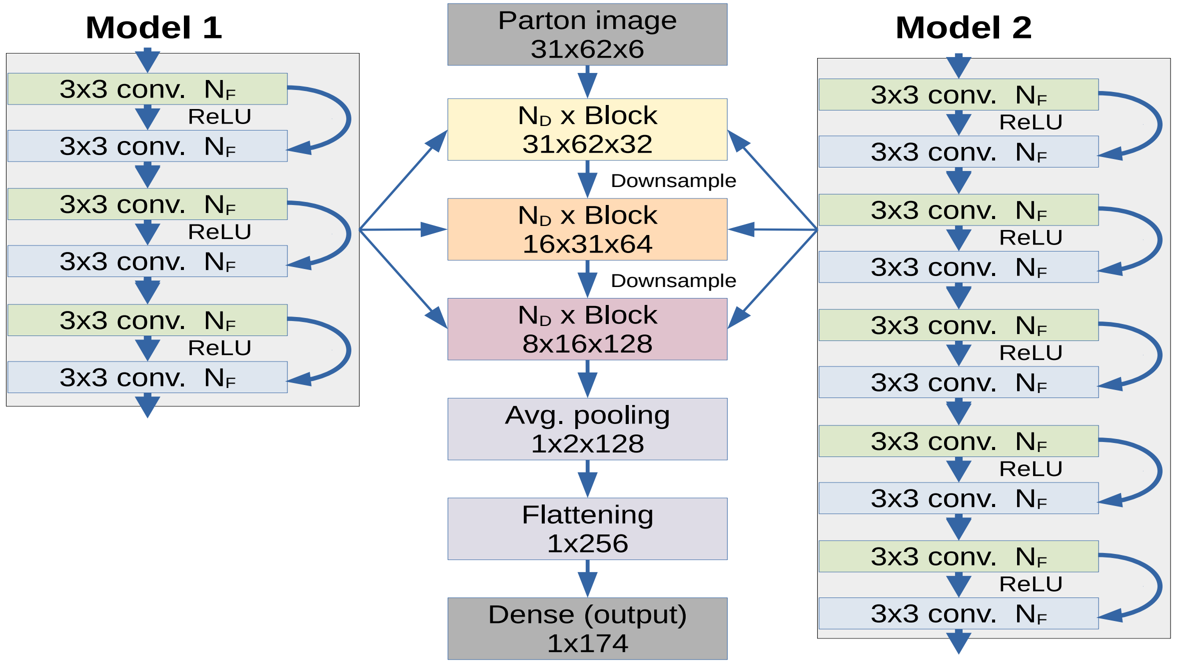


Building up the ML structure

Algorithms behind: ResNet

- **Weights** dictate the importance of an input → more important features get more weights
- **Activation function**: mathematical function that guides the outcome at each node → Standardize the values
- **Cost function**: Evaluates the accuracy between machine prediction and true value
- **Optimizer**: Method (or algorithm) that minimizes the cost function by automatically updating the weights

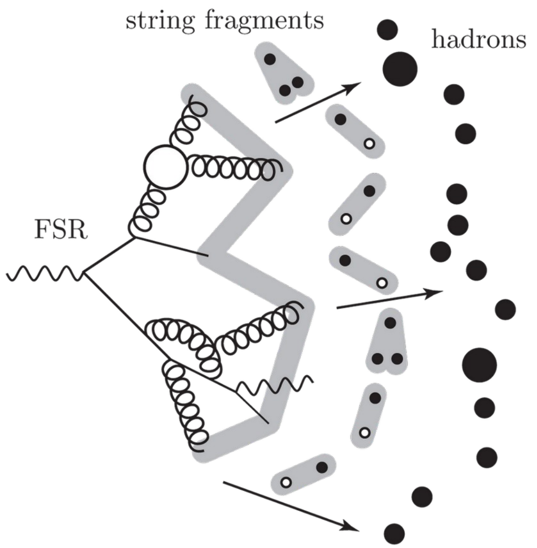


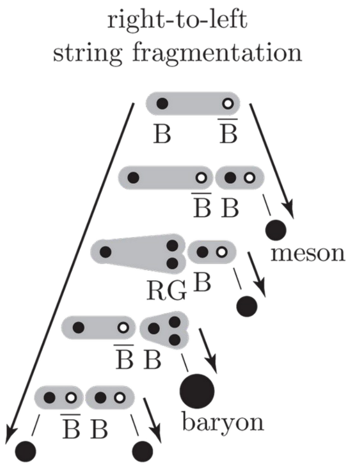


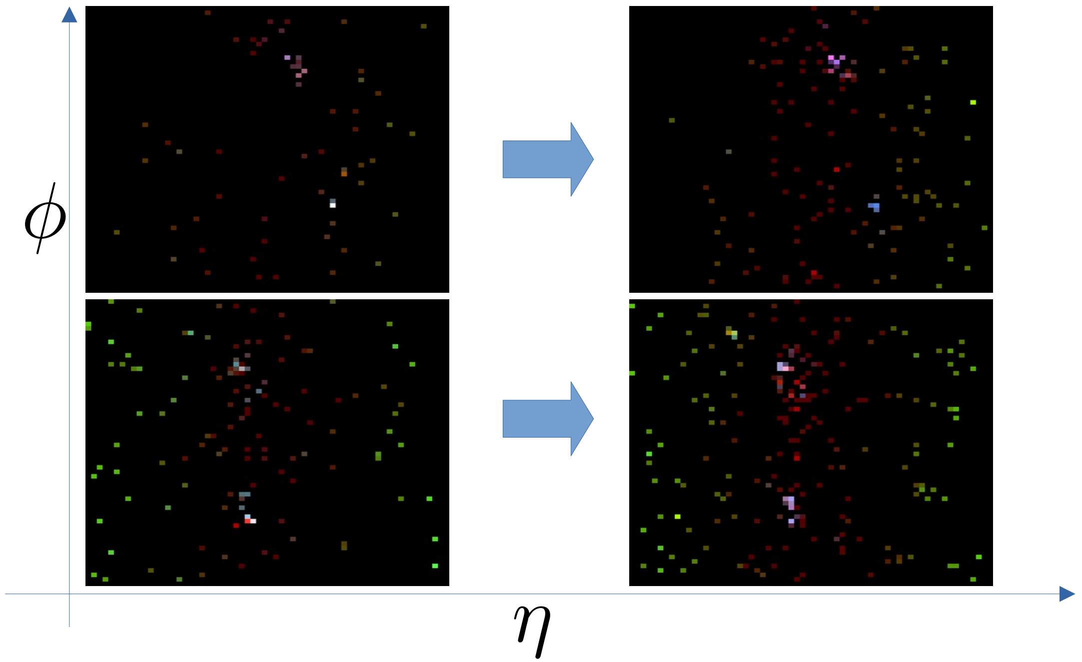
Input/output of the ML structure

Simulated data at parton/hadron level

- Event properties (now from PYTHIA)
- Inputs → Parton/hadron level input
- (η-ϕ) space is the primary input space







Training & validation of the model

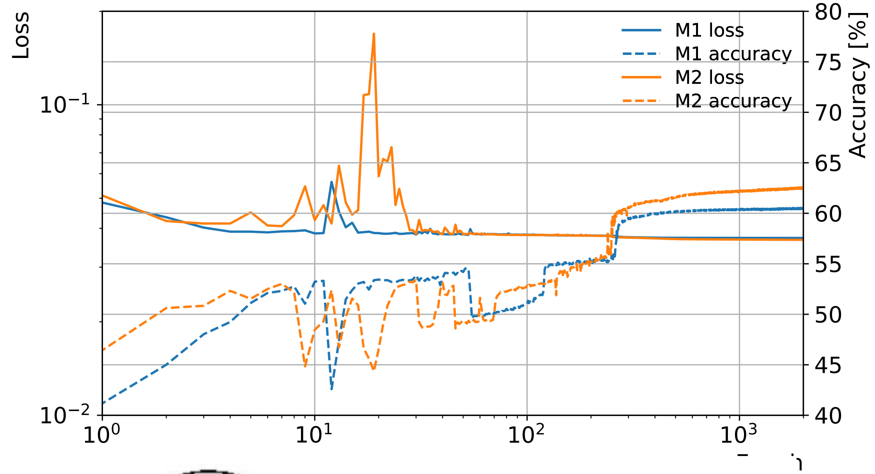
ML: training, optimalization, validation

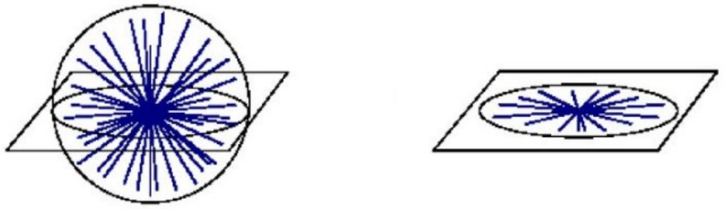
- **Training:** PYTHIA 8.303 Monash tune, All final particles, at least 2 jet, Anti-k_T R=0.6 p_T> 40 GeV/c, $|y| < \pi$
- **Input**: parton/hadron: *(p_x, p_y, p_z, E, m)*

  62 bin ($\phi \in [0, 2\pi]$)

  31 bin ($y \in [\pi, \pi]$)
- **Epoch:** 300
- **Training/Validation:** 150k events (20 GB)
- **Machine:** Used hardwares: Nvidia Tesla T4, GeForce GTX 1080, GeForce GTX 980 @ Wigner Scientific Computational Laboratory
- **Framework:** Tensorflow 2.4.1, Keras 2.4.0
- **Features:** Multiplicity/Jet distributions, Jet/p_T spectra, Event properties: Sphericity Transverse spherosity,

	Model 1	Model 2
Trainable parameters	**1.13 M**	**1.90 M**



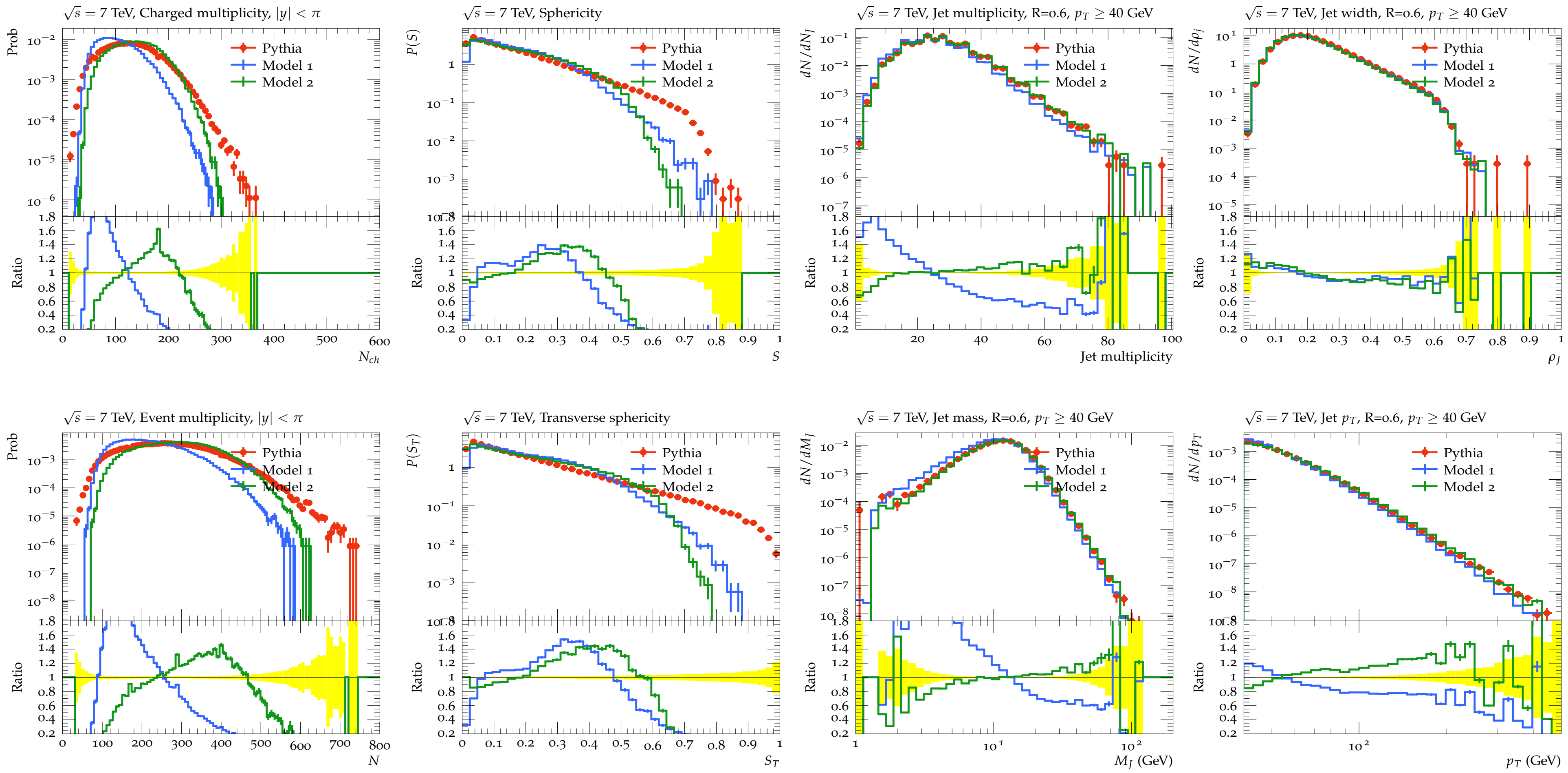


S=1 A=1/2

S=3/4 A=0

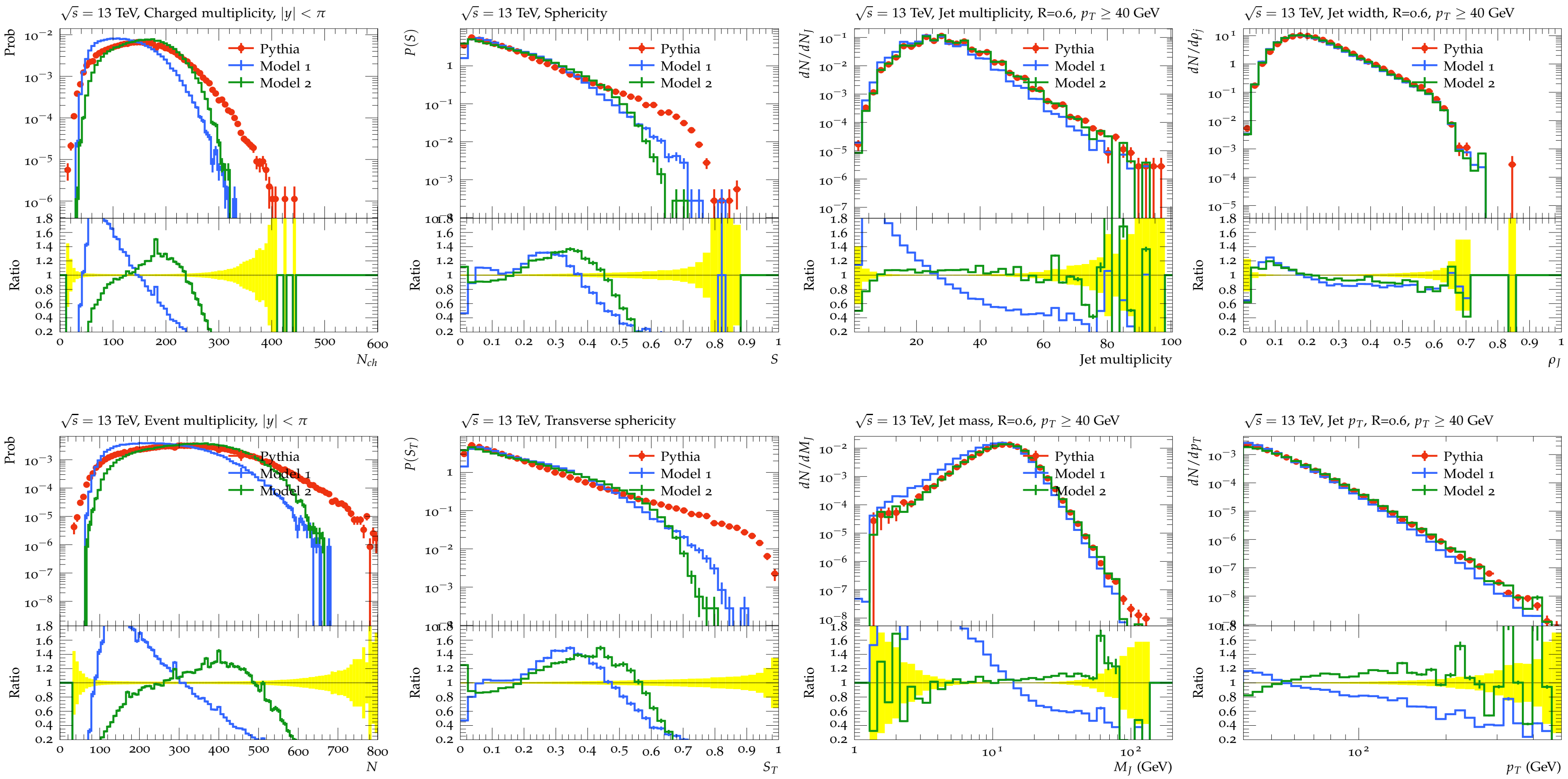
Results on ML-hadronization

Training and validation pp@7 TeV



Results on ML-hadronization

Predictions pp@13 TeV

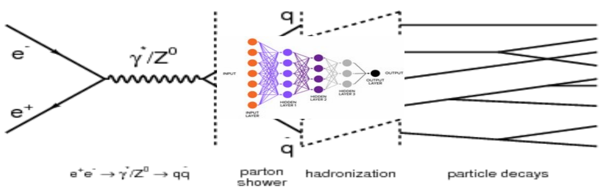


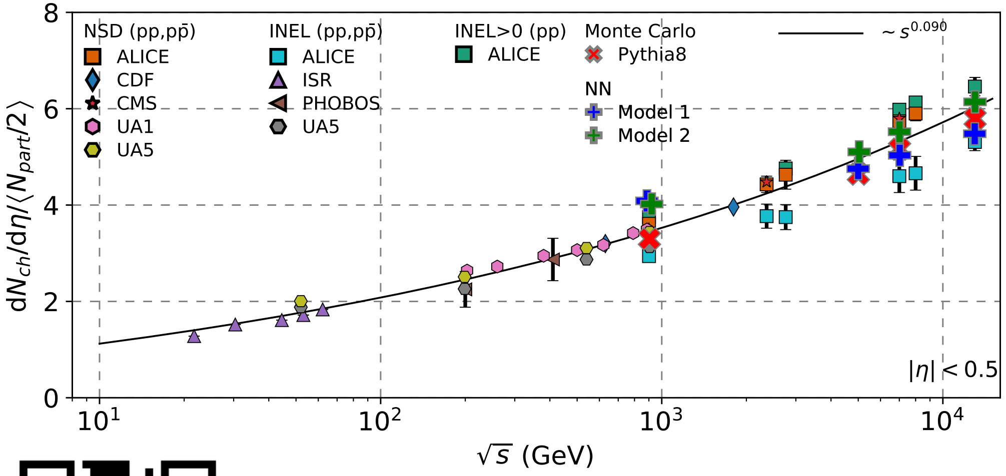
Conclusions

- **Aim: modelling hadronization by ML**
  - Highly non-linear/non perturbative problem

    → Many features are fitted well: multiplicity distributions, jet/pT distributions, spericity, transverse spherocity, etc

    → Preserved scaling in multiplicity on a wide energy range





- **Work in progress…**
  - Stability and test of noise on training
  - Better separation of shower/hadronization