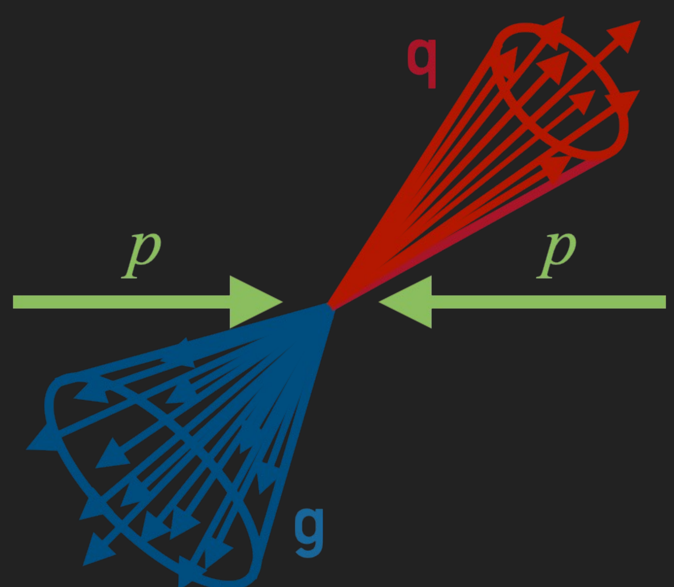


Constituent based Quark / Gluon Jet Tagging

SAMUEL JANKOVYCH for the ATLAS Collaboration

ATLAS EXPERIMENT

Charles University

Quark / Gluon Jet Tagging

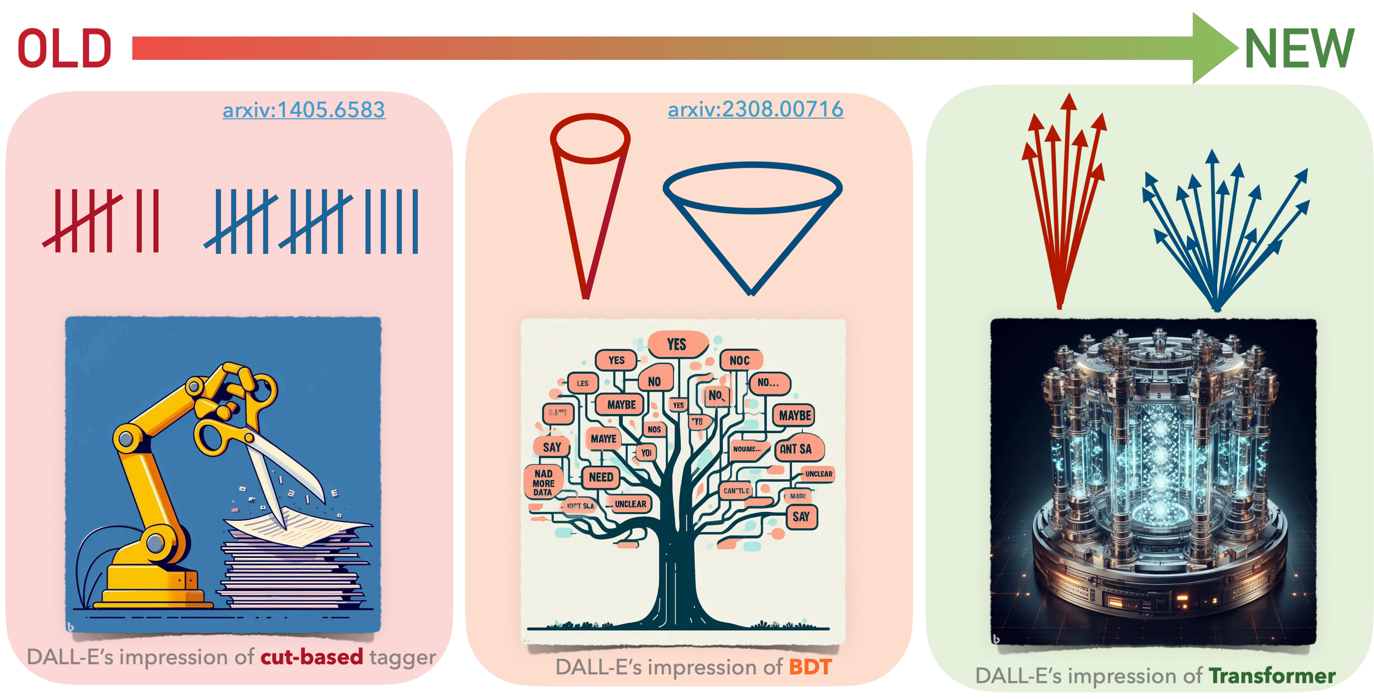
Goal = **distinguish quark/gluon jet origin based on jet properties**

Our approach:

- Modern Machine Learning approach = **Transformers**
- **Jet constituent** level information
- Full **ATLAS** detector simulation

So far, q/g taggers in ATLAS use jet level (or high-level) variables:

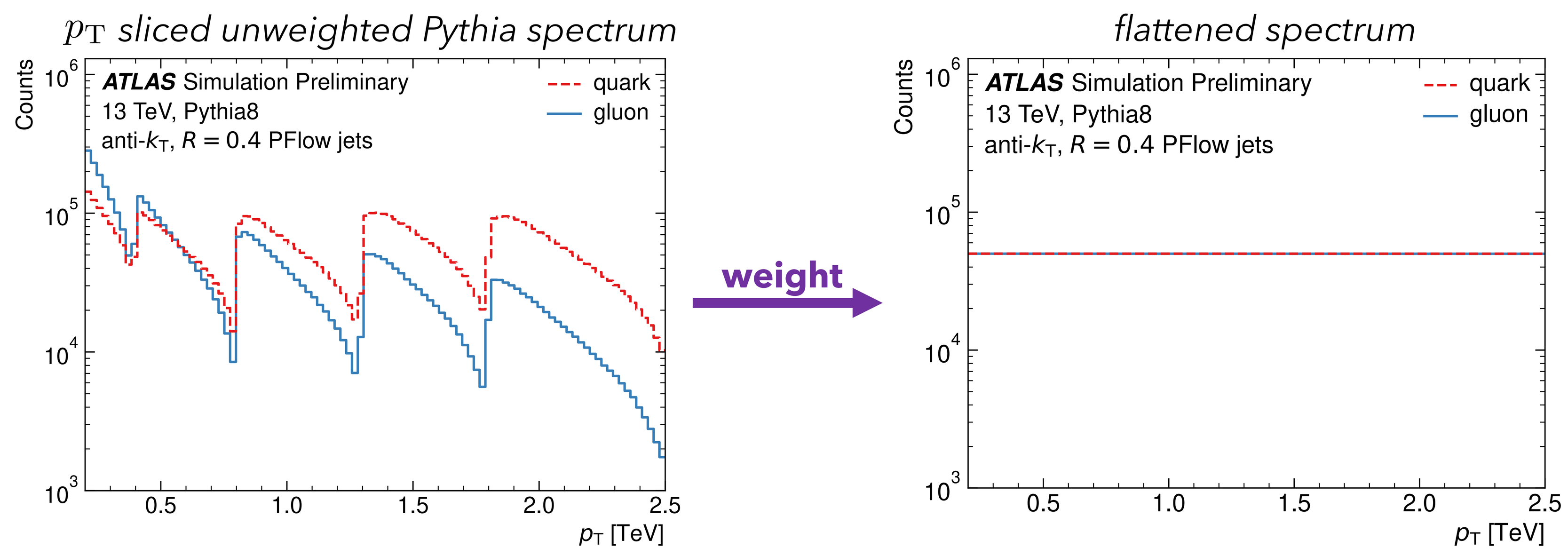
- Cut-based, using just the number of tracks asociated to a jet
- BDT, using 5 high-level variables (p_{T}, n_{trk}, W_{trk}, $C_1^{\beta=0.2}$)



Data

- MC simulated pp collisions = **Pythia dijet events**
- Full **ATLAS** detector simulation = Geant4
- **Jet constituents** = Particle Flow Objects (PFO)

- Force the ML tagger to not be reliant on jet p_{T} = **train** with **flat** p_{T} **spectrum**
- Flat p_{T} spectrum - separate for quark and gluons
- Allows physical analyses to use in both, high p_{T} and low p_{T} region
- **10M** jets for training



- Evaluate the models performance = **test** dataset with **physical** p_{T} **spectrum**
- 1M jets for evaluation

Deep Learning Model Bestiary

Constituent level

- Variables created from constituents 4-momenta for each constituent

+ Interaction variables

- Variables created for each pair of constituents

Constituent Interaction Variables
$\log\Delta^{ab} = \log\sqrt{(\eta^a-\eta^b)^2+(\phi^a-\phi^b)^2}$
$\log k_{\mathrm{T}}^{ab} = \log(\min(p_{\mathrm{T}}^a, p_{\mathrm{T}}^b)\Delta^{ab})$
$z^{ab} = \min(p_{\mathrm{T}}^a, p_{\mathrm{T}}^b)/(p_{\mathrm{T}}^a+p_{\mathrm{T}}^b)$
$\log m^{2,ab} = \log(p^{\mu,a}+p^{\mu,b})^2$

DeParT

- ParT enhancement

ParT arXiv:2202.03772

- Transformer

Constituent Variables
$\Delta\eta = \eta - \eta^{\mathrm{jet}}$
$\Delta\phi = \phi - \phi^{\mathrm{jet}}$
$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$
$\log p_{\mathrm{T}}$
$\log E$
$\log \frac{p_{\mathrm{T}}}{p_{\mathrm{T}}^{\mathrm{jet}}}$
$\log \frac{E}{E^{\mathrm{jet}}}$
m

ParticleNet (P.Net) arxiv:1902.08570

- Graph Network

PFN arxiv:1810.05165

- Deep Sets

Linear variables

- Allow construction of IRC safe model

EFN arxiv:1810.05165

- IRC safe

Linear Constituent Variables
$\Delta\eta = \eta - \eta^{\mathrm{jet}}$
$\Delta\phi = \phi - \phi^{\mathrm{jet}}$
$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$
$\frac{p_{\mathrm{T}}}{p_{\mathrm{T}}^{\mathrm{jet}}}$

Highlevel

- Variables describing the jet as whole

High-level Jet Variable
jet 4-momentum
total number of PFOs
number of charged PFOs with $p_{\mathrm{T}} > 1000\mathrm{MeV}$
number of charged PFOs with $p_{\mathrm{T}} > 500\mathrm{MeV}$
jet width computed from charged PFOs with $p_{\mathrm{T}} > 1\mathrm{GeV}$
a fraction of energy of a jet deposited in EM calo
charged fraction of a jet

Fully Connected (FC)

- MLP

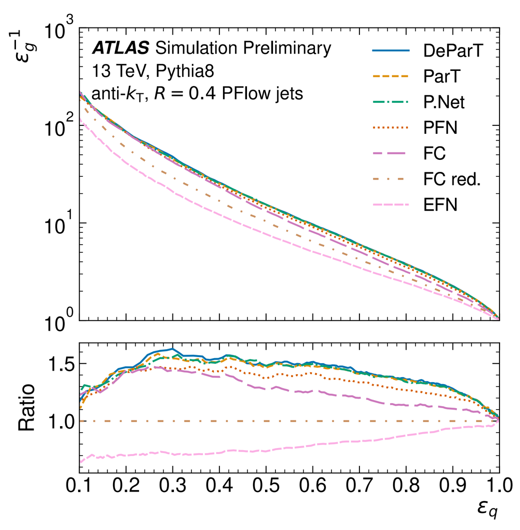
Reduced Fully Connected (FC red.)

- Closest to the previous BDT tagger

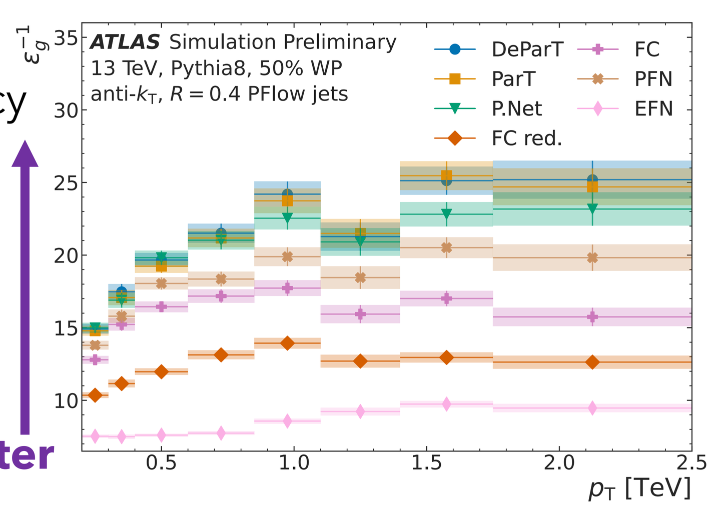
$p_{\mathrm{T}}, \eta, N_{\mathrm{PFO}}, W_{\mathrm{PFO}}, C_1^{\beta=0.2}$

Nominal Results

Model	AUC	$\varepsilon_g^{-1}@\varepsilon_q = 0.5$	# Params [10^6]	Inference Time [ms]	GPU Memory [MB]
DeParT	0.8489	15.4242	2.62	266.51	1684
ParT	0.8479	15.2457	2.62	233.84	1730
ParticleNet	0.8476	15.4402	2.59	768.74	5410
PFN	0.8406	14.2387	2.64	136.93	393
FC	0.8280	13.5199	2.63	65.53	76
FC reduced	0.8038	10.3639	2.63	84.84	47
EFN	0.7761	7.7222	2.60	101.53	337



- Total number of parameters in all models is fixed to 2.6M
- Constituent based models outperform high-level ones
- **DeParT** obtains **best** AUC
- ParticleNet
  - Best gluon rejection at 0.5 quark efficiency
  - Time and memory consuming
- **Transformer** models perform **the best**
- EFN underperforms due to IRC limitations
- Performance grows with p_{T}
- At low p_{T} the performance is more similar

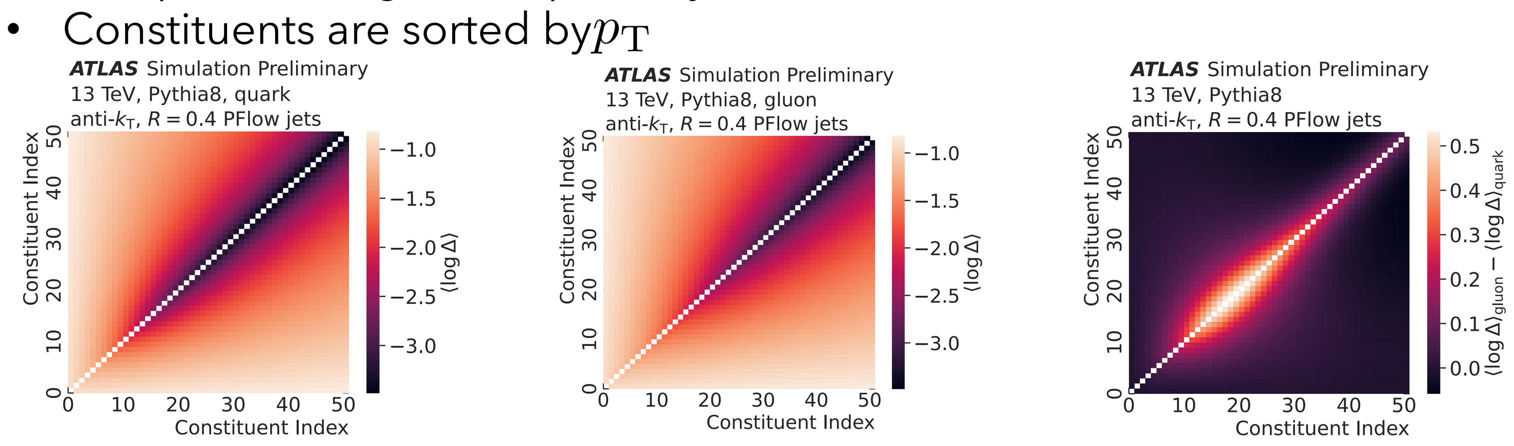


SCAN for more!

samuel.jankovych@cern.ch

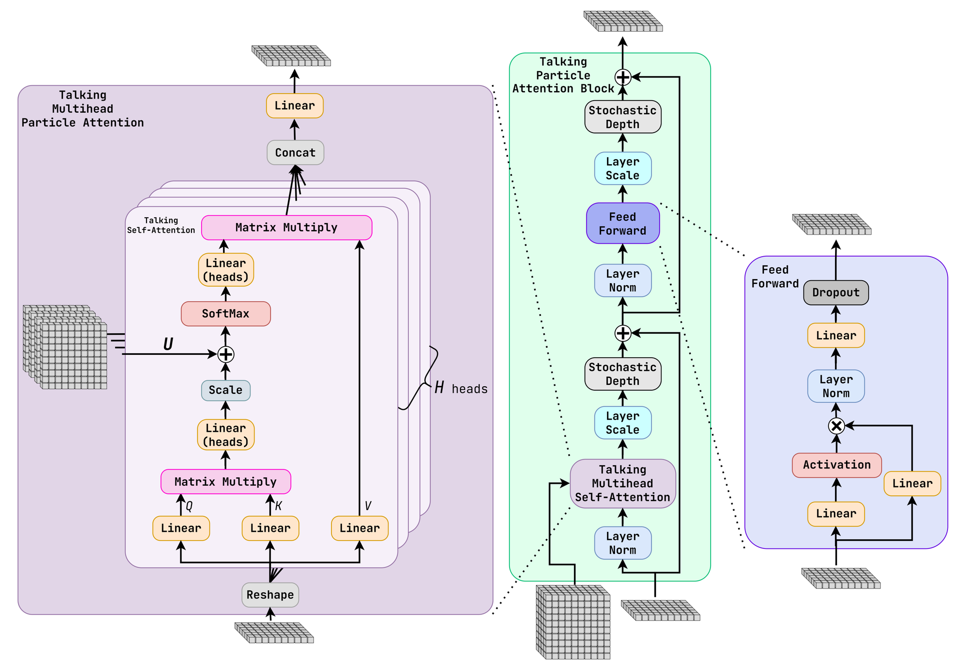
Constituent Interaction Variables

- Calculated for each **pair of constituents**
- Motivated by parton shower
- One variable is a N-by-N matrix, where N is the number of constituents
- The figures below show average value of $\log\Delta$ in the training dataset
  - For quarks and gluon separately and difference between them
  - Constituents are sorted by p_{T}



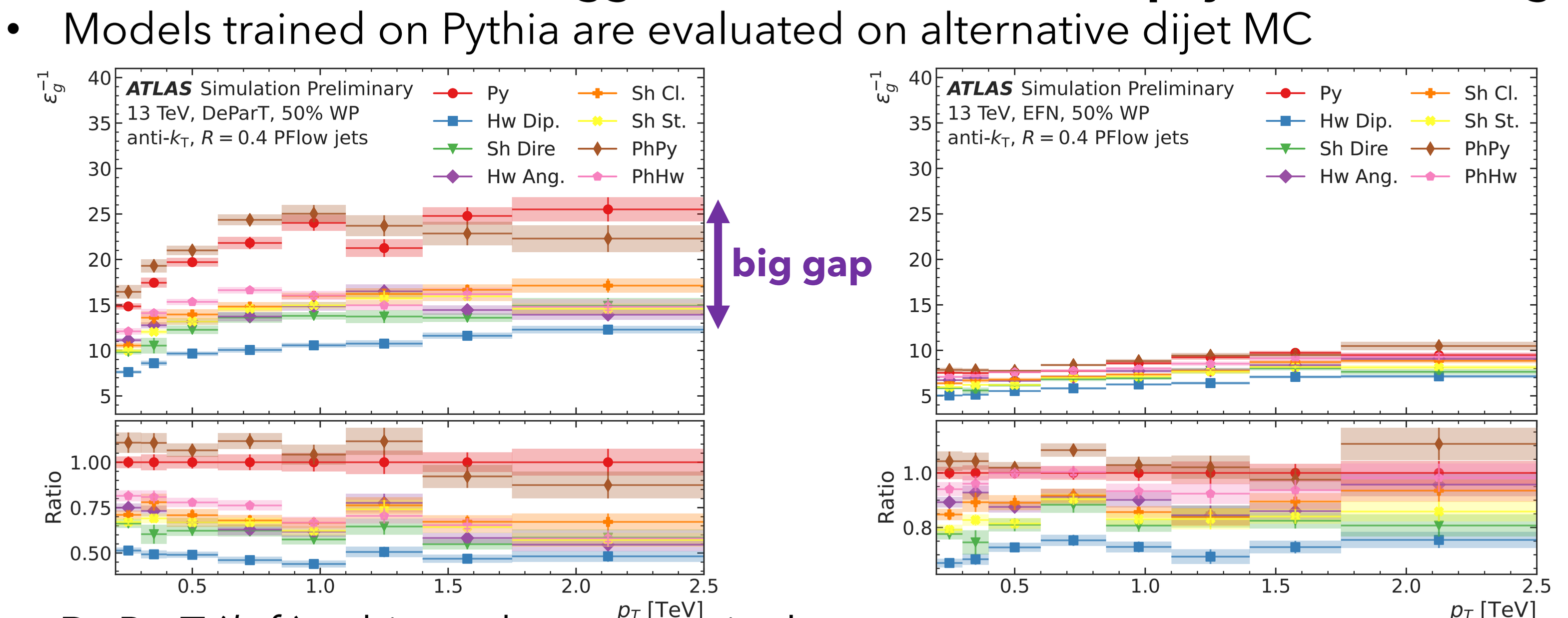
Dynamically enhanced Particle Transformer

- Enhancement of ParT
- Additions from DeiT III (arxiv:1810.05165) allow easier and more stable training of deeper models
  - **Stochastic depth**
  - **Layer scale**
- **Gated FFN** allows easier information flow through the network
- **Talking Heads**
  - More communication between heads
  - Share constituent variables



Monte Carlo Dependence

- **How sensitive are the taggers to the differences in physical modelling?**
- Models trained on Pythia are evaluated on alternative dijet MC



- DeParT (*left*) achieves better nominal performance but also has bigger difference between MCs than EFN (*right*)
- Only PhPy has similar performance to Py → ME has smaller effect
- Parton shower and hadronization modelling have big effect
- Tradeoff: **performance vs MC independence**