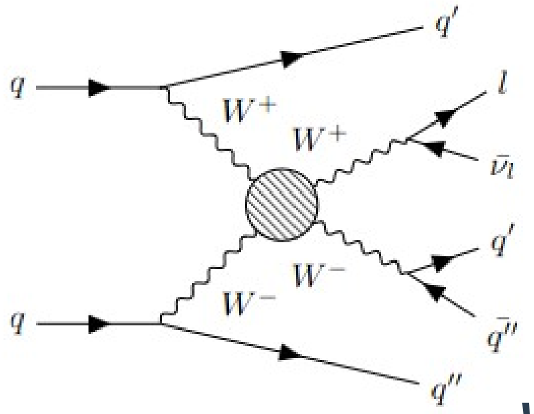
Light quark and gluon jet tagging at CMS

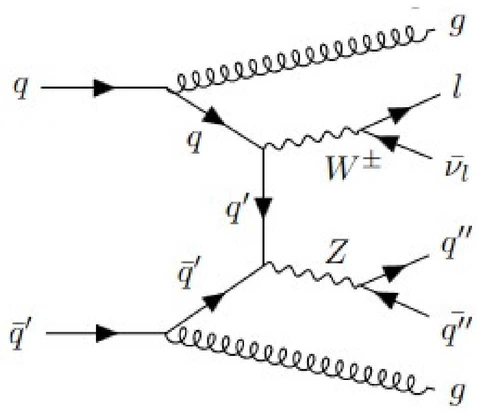
Nico Toikka, Helsinki Institute of Physics
QCD Midsummer School
June 26th 2024

Motivation

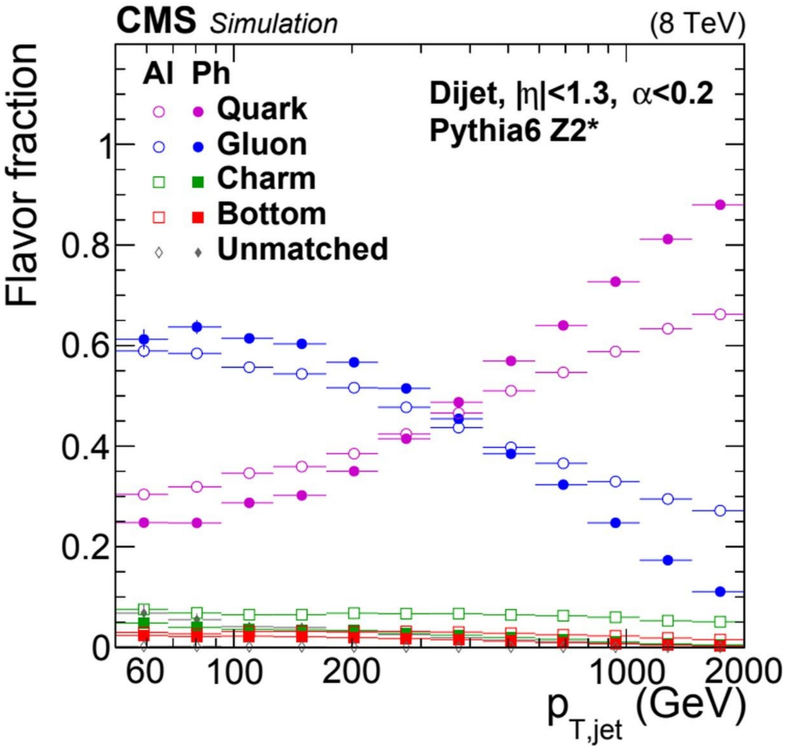
Final states



vs



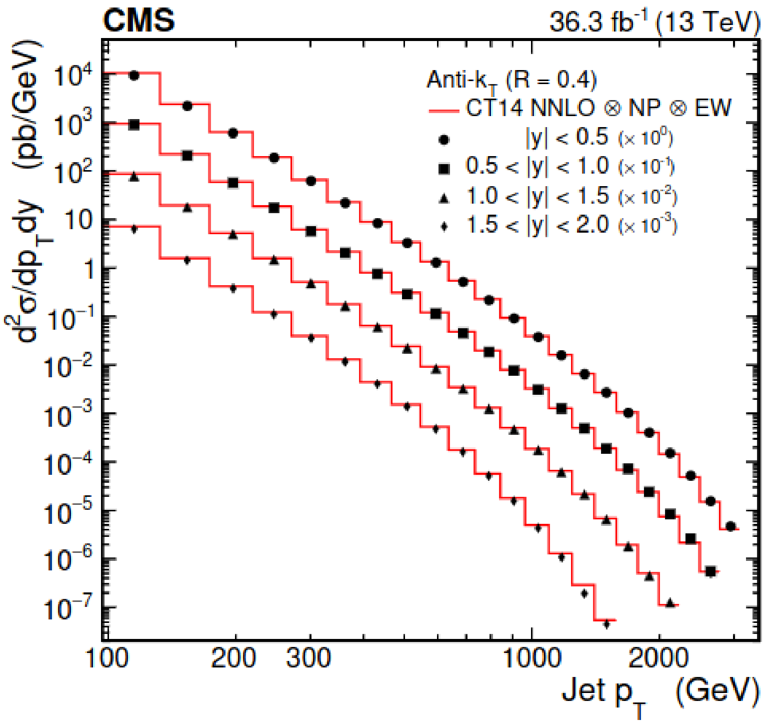
Environment



arXiv:1607.03663

Motivation

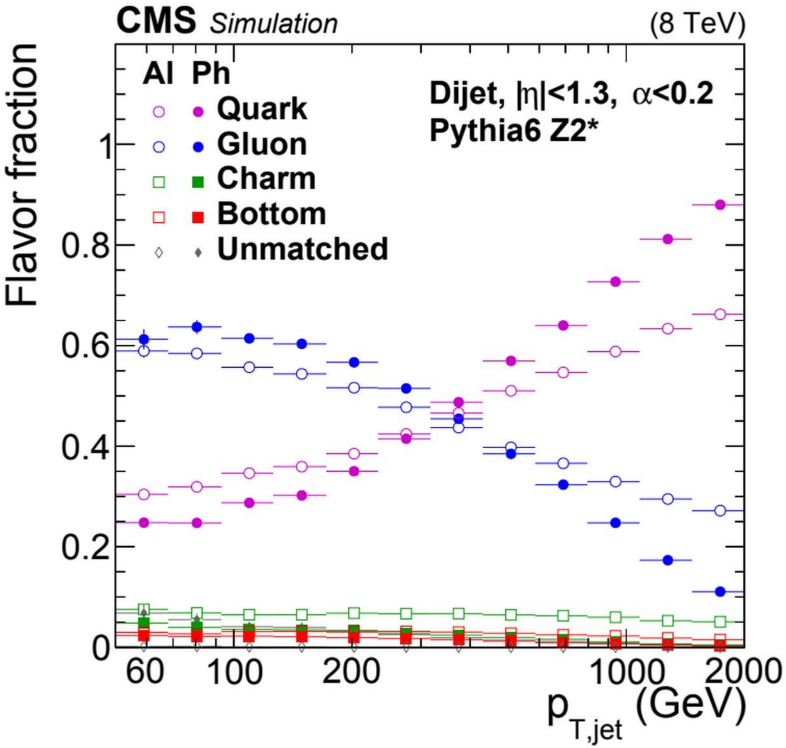
Final states (jets)



arXiv:2111.10431

+

Environment



arXiv:1607.03663

Motivation

Final states (jets)

Environment

Systematic uncertainties

- How sure we are about the measured final state
- How well is the environment modeled ie. FSR, MPI, pileup, detector response

Table 6: The impact of the dominant systematic uncertainties on the observed signal strength for inclusive Higgs boson production followed by decay to bottom quarks.

Source of systematic uncertainty	Impact on signal strength [%]
VBF parton shower	13.0
Jet energy scale	7.7
Trigger efficiency	6.7
Parton shower (final-state radiation)	5.6
b jet regression smearing	3.3
b tagging efficiency	3.0
Pileup modeling	2.3
b jet regression scale	2.0
Jet energy resolution	1.5

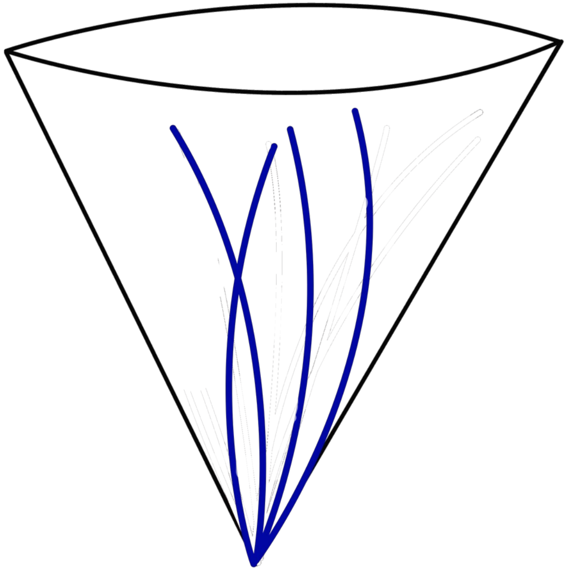
arXiv:2308.01253

Table 2: Breakdown of the uncertainties in the EW WV VBS signal

Uncertainty source	$\Delta\mu_{EW}$
Statistical	0.12
Limited sample size	0.10
Normalization of backgrounds	0.08
Experimental	
b-tagging	0.05
Jet energy scale and resolution	0.04
Integrated luminosity	0.01
Lepton identification	0.01
Boosted V boson identification	0.01
Total	0.06
Theory	
Signal modeling	0.09
Background modeling	0.08
Total	0.12
Total	0.22

arXiv:2112.05259

Definitions – objects

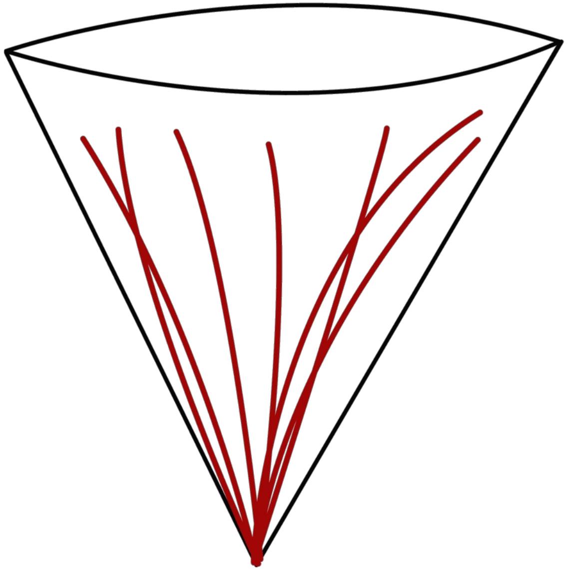


Quark jet

Quark and **gluon** jets are commonly defined by the **initiating parton**.

Their different properties as jets are also related to their **Casimir factors $C_A=3$ and $C_F=4/3$.**

- Gluon jets contain more particles in a larger cone
- Quark jets have a larger fraction of charged hadrons and have “harder” particles



Gluon jet

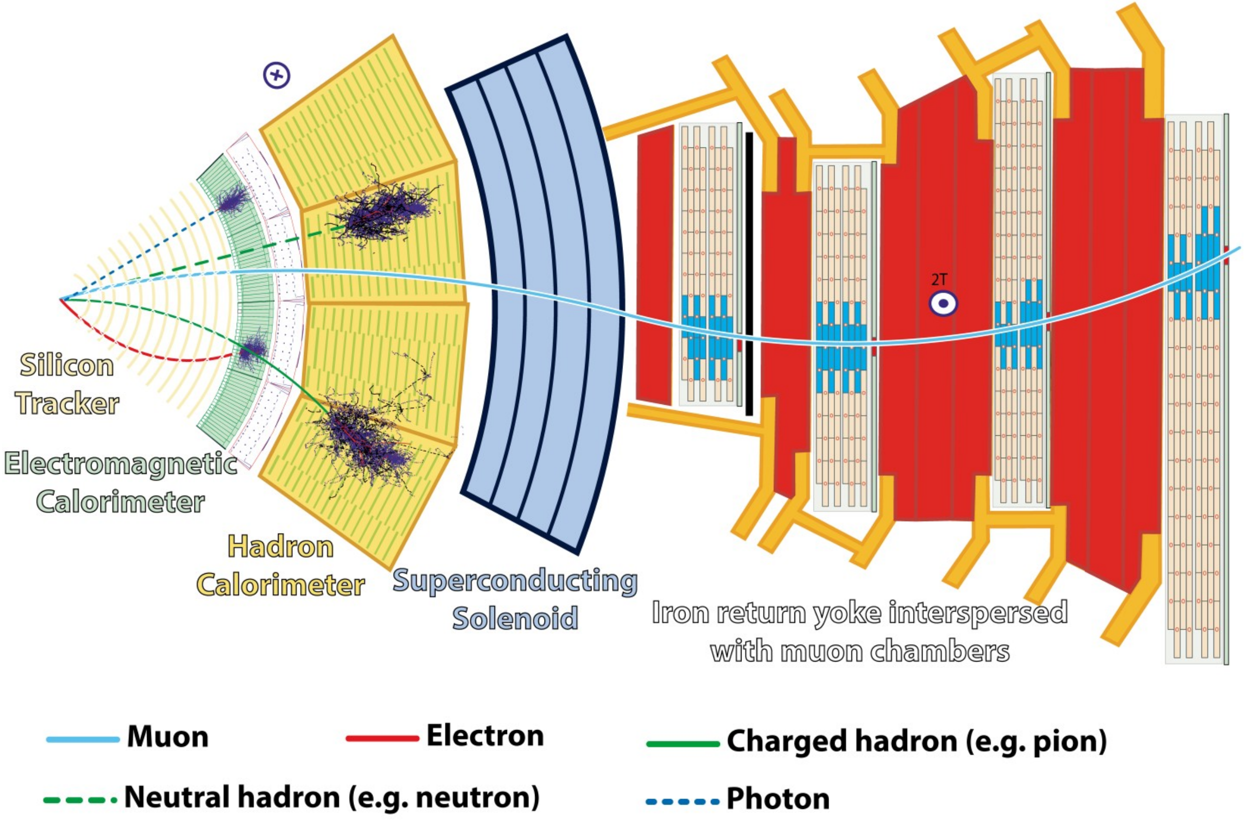
Definitions – detector

Key points from the detector:

I. Charged particles leave a **track** in the Si tracker, which does not cover the **forward region**

II. The energy of charged and neutral particles measured at **ECAL** and **HCAL** respectively

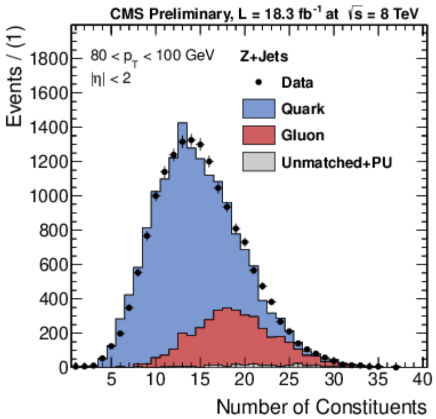
III. **Additional activity** due to pileup and MPI – low energy, mostly forward region

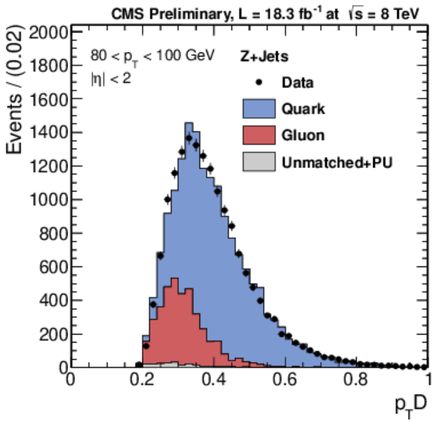


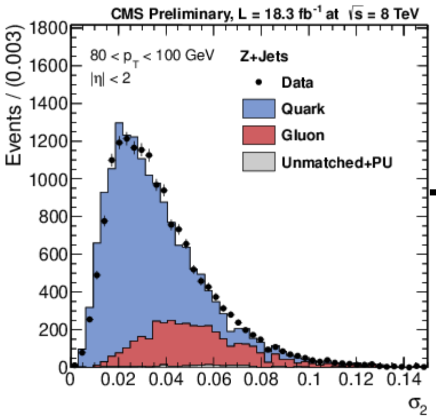
Evolution of QG tagging: Likelihoods

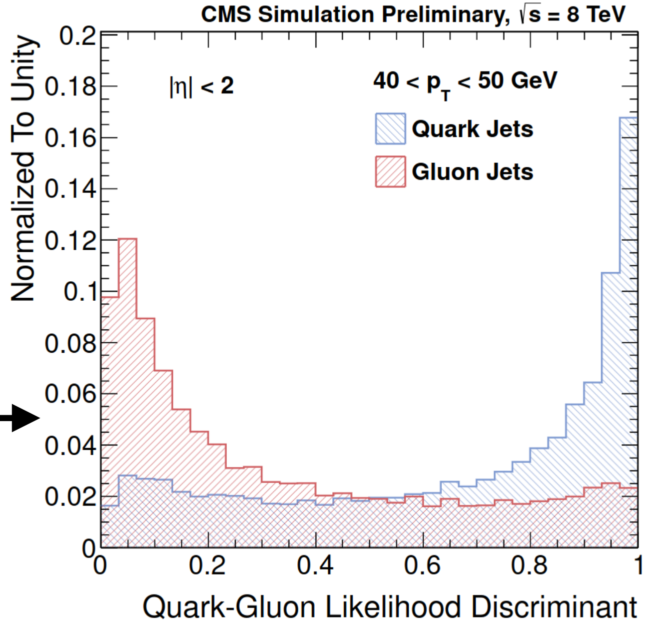
Likelihood based methods (QGL)

Utilize Q/G jet properties to create a likelihood function







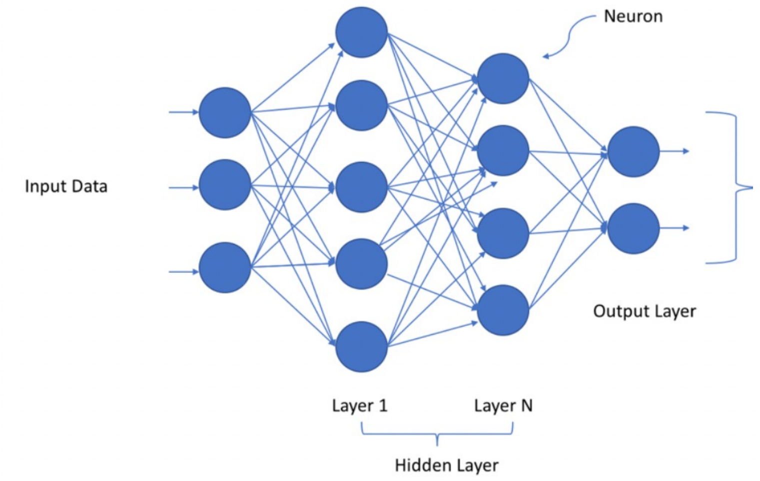


arXiv:1409.3072

Evolution of QG tagging: Neural networks

Deep neural networks (DeepJet)

Feed jet properties to a DNN, use MC to train the model



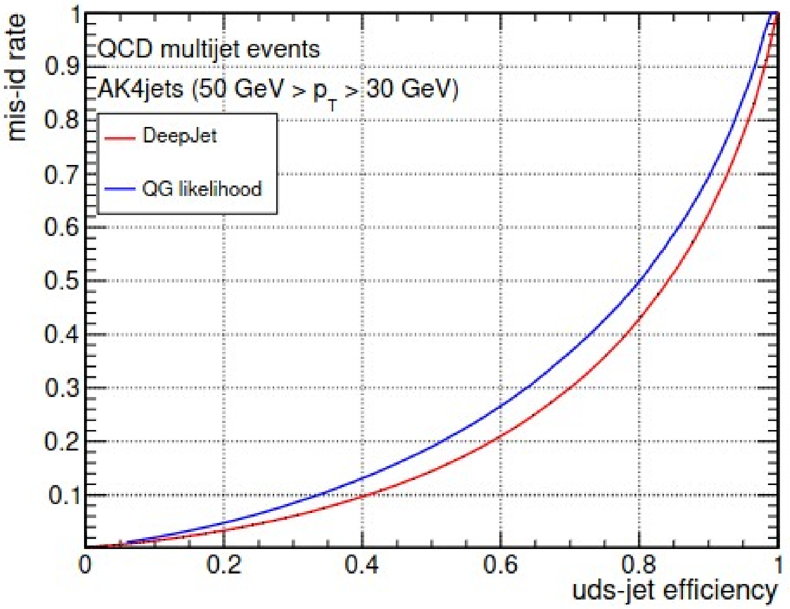


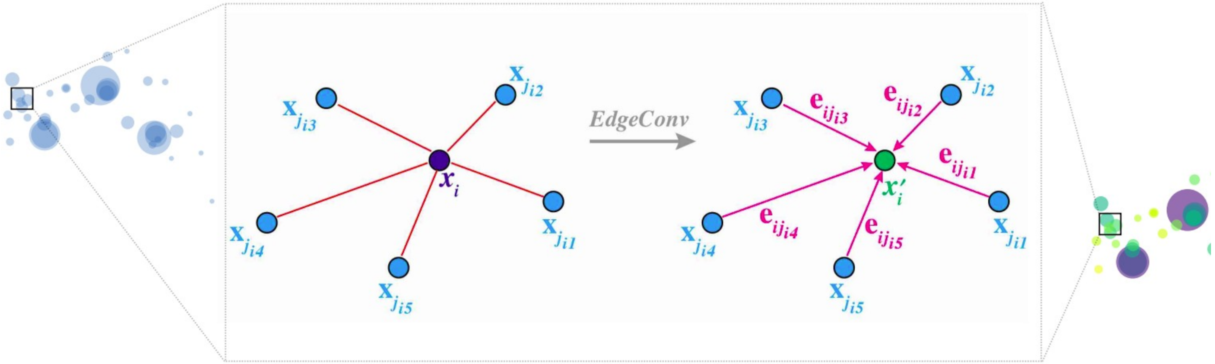
Figure 6. Quark gluon discrimination performance of DeepJet compared to the CMS "quark-gluon likelihood" method in a sample of pure light quark (uds) and gluon jets.

arXiv:2008.10519

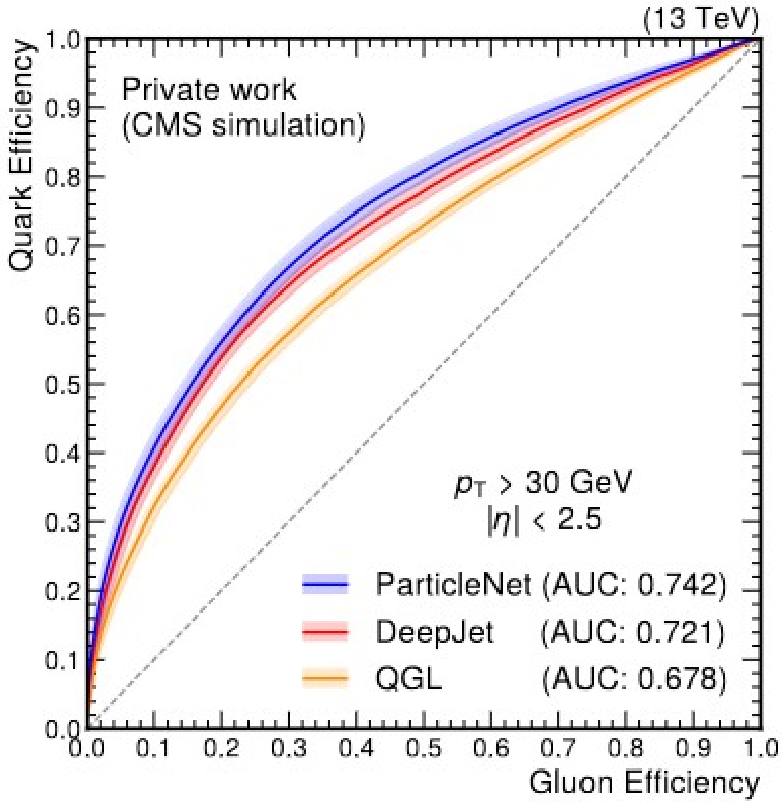
Evolution of QG tagging: Graph NN

Graph Neural Networks (ParticleNet)

Input particles as a graph → find k-nearest neighbors in detector coordinates and perform edge convolution



ParticleNet documentation



K. Kallonen PhD thesis

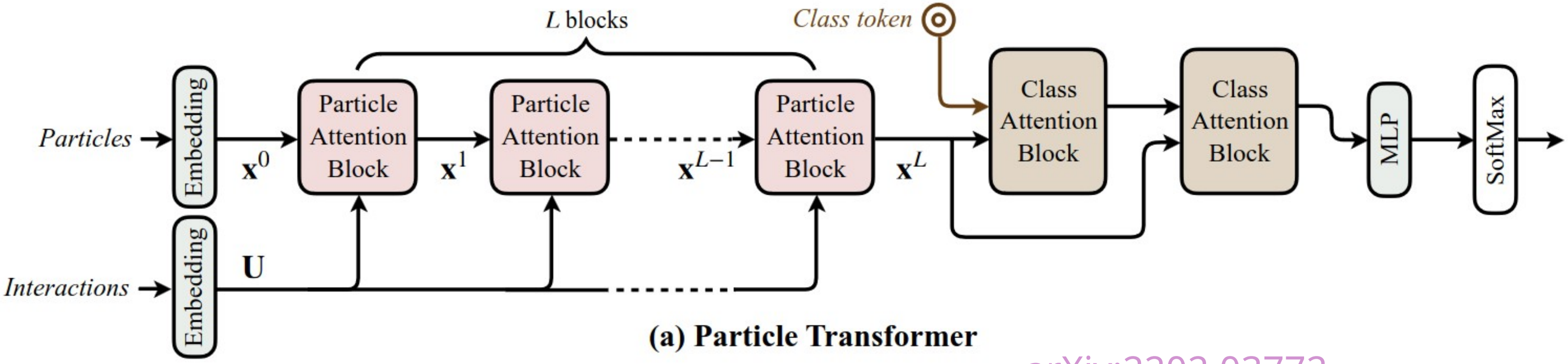
Evolution of QG tagging: Transformers

Future methods

Transformers (ParT)

Take in a point cloud representation of the particles that is used as a input to a transformer

Best performance out of the current CMS taggers, but no “field tests” so far (as far as I know).



(a) Particle Transformer

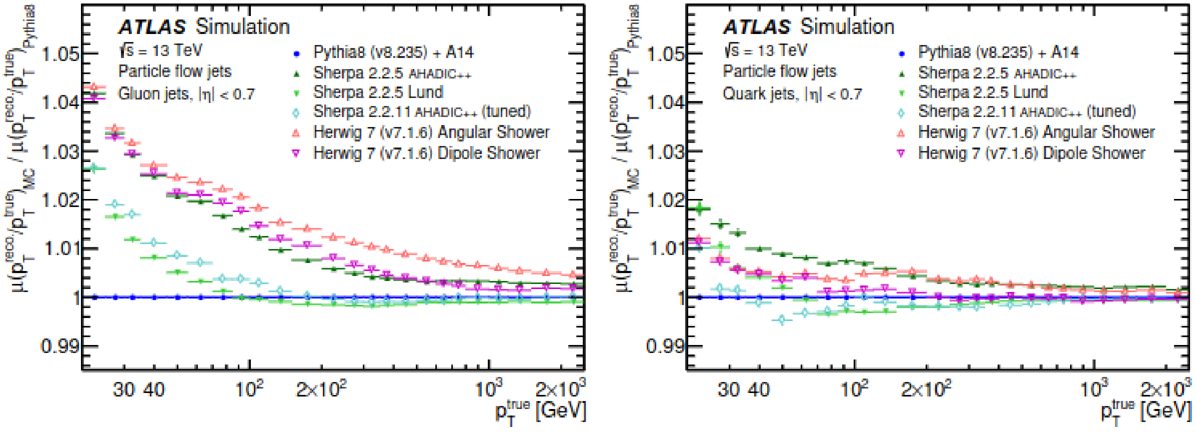
$$
\begin{aligned}
\Delta &= \sqrt{(y_a - y_b)^2 + (\phi_a - \phi_b)^2},\\
k_{\mathrm{T}} &= \min(p_{\mathrm{T},a}, p_{\mathrm{T},b})\Delta,\\
z &= \min(p_{\mathrm{T},a}, p_{\mathrm{T},b})/(p_{\mathrm{T},a} + p_{\mathrm{T},b}),\\
m^2 &= (E_a + E_b)^2 - \|\mathbf{p}_a + \mathbf{p}_b\|^2,
\end{aligned}
$$

arXiv:2202.03772

Utilizing the taggers

Jet energy scale & Generators

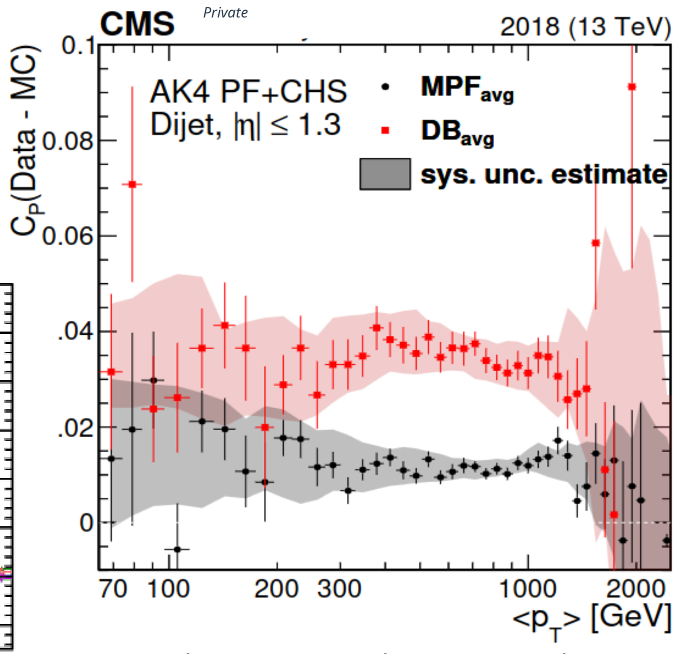
Quark and gluon jets respond differently to the detector. This information can be tied to monitoring jet and detector performance.



(a) (b)

Figure 1: The ratio of the average jet energy response for (a) gluon-initiated and (b) quark-initiated jets in various MC simulation samples to the jet energy response of the nominal PYTHIA sample, as a function of the true jet p_T for jets within the pseudorapidity range $|\eta| < 0.7$.

arXiv:2405.20206



There is a ~3 % disagreement between data and simulation when comparing the Q/G jets response to detector (CMS 2018 Data vs flat Pythia8)

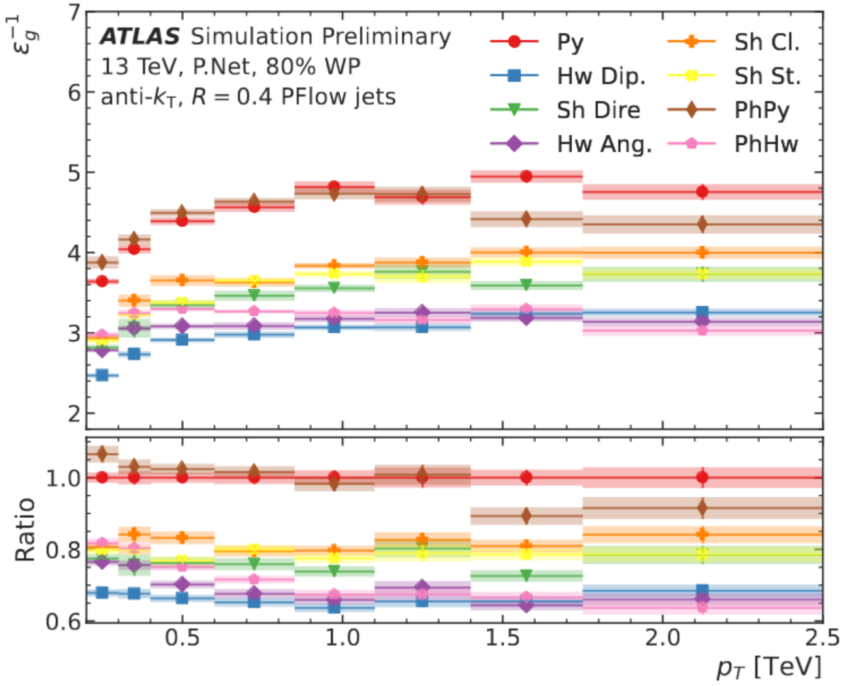
(Open) Question: Which generator agrees the best, and exactly where do the differences come from?

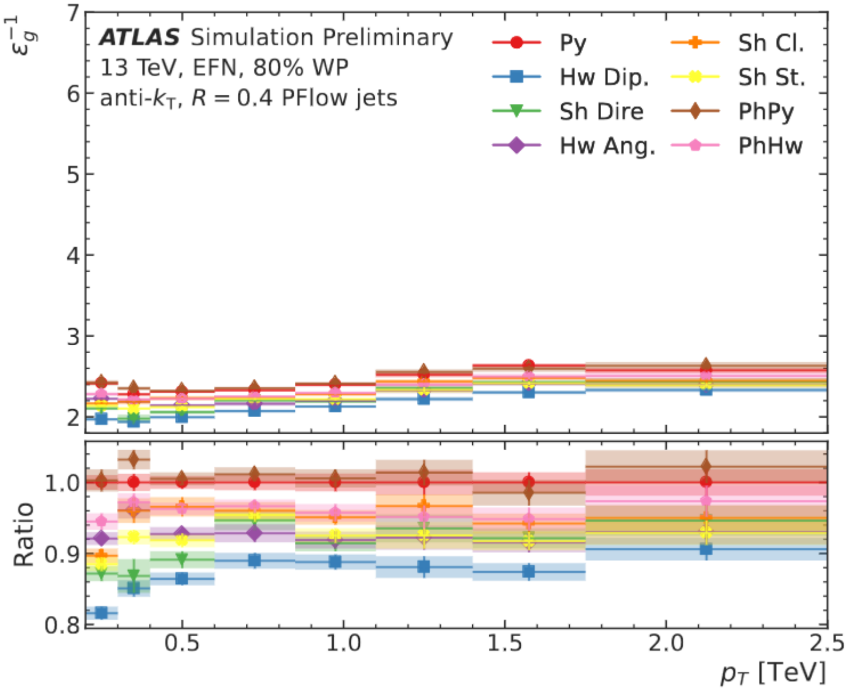
Utilizing the taggers

Jet energy scale & Generators

Quark and gluon jets respond differently to the detector. This information can be tied to monitoring jet and detector performance.

ATLAS study shows differences in Q/G tagging performance between generators

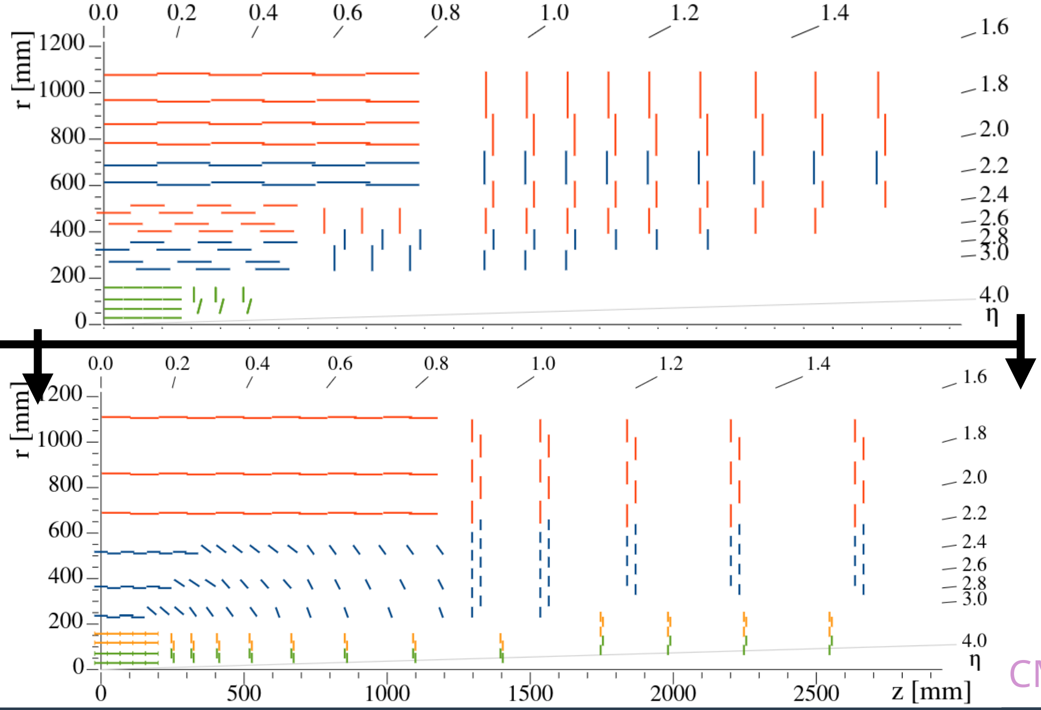




Future work: Technology

"Just measure better"

At CMS the forward region and hadron calorimetry will be improved for HL-LHC with new tracker and endcap calorimeters

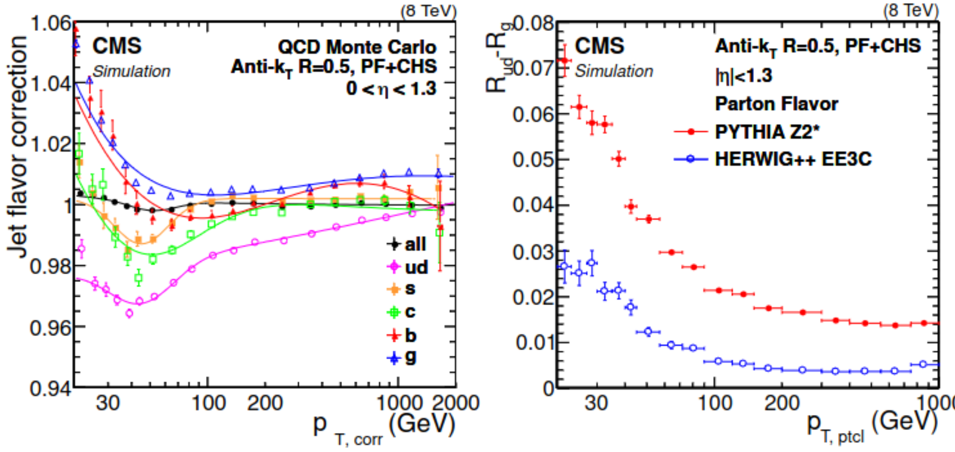


CMS Phase 2 tracker upgrade

Finer inputs

Hypothetically, flavor based calibrations can be introduced at MC level to improve Data/MC performance. Can be classic fit or ML model trained to minimize difference.

But fit/train based on which generator?

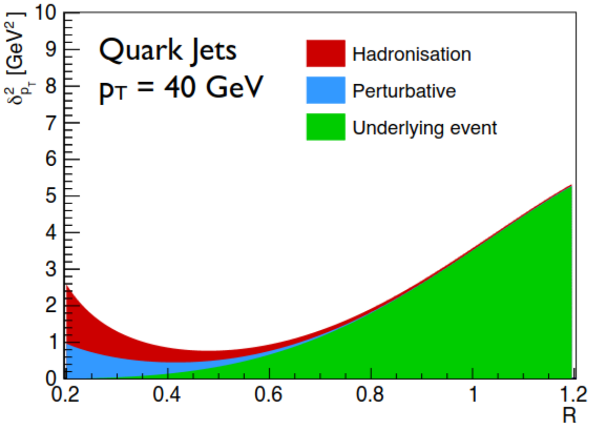


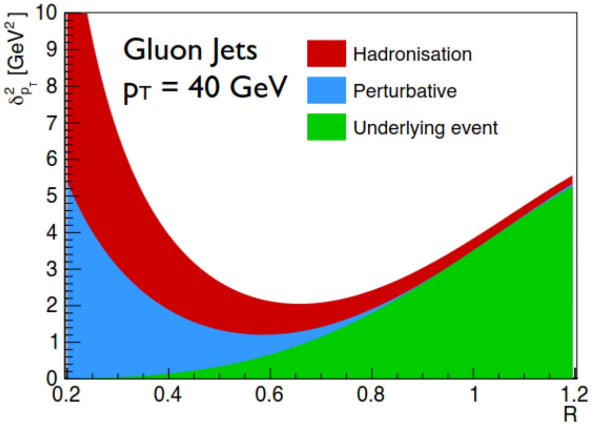
arXiv:1607.03663

Future work: Theory

Defining jets

Anti-k_t is *the* jet algorithm of the LHC era, but a fixed cone size can't accommodate the needs of both types of jets



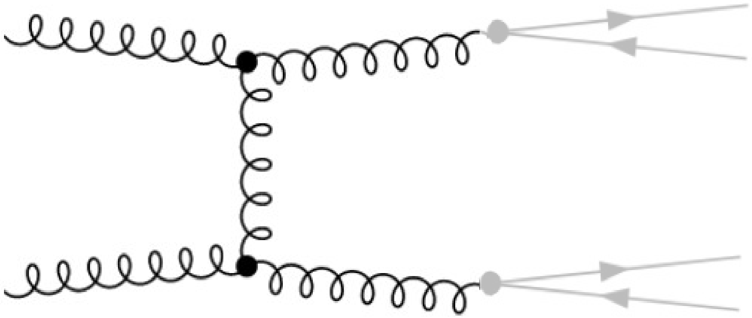


Private study by Roman Kogler (DESY/CMS) on the correction required for Q/G jets

Heavy object tagging with variable R

Defining flavor

Initiating parton -definition has been adopted by the experimental community, but it has the flaw of not being measurable from hadrons and a doesn't consider how the jet evolves



An operational definition of quark and gluon jets

Thank you for following!