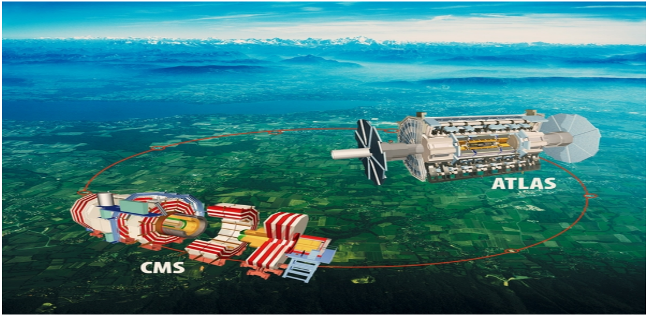
Recent jet and jet substructure measurements at the LHC, and ML based tagging

Meena (On behalf of ATLAS and CMS collaboration)

ISMD2021: 50th International Symposium on Multiparticle Dynamics

16^{th} July, 2021



Outline

- Study of quark and gluon jet substructure in dijet and Z+jet events from pp collisions (CMS)→ **New results**
- Measurement of the Lund jet plane using charged particles in 13 TeV proton-proton collisions with the ATLAS detector
- Mass regression of highly-boosted jets using graph neural networks (CMS) → **Very new results**
- Boosted hadronic vector boson and top quark tagging with ATLAS using Run2 data (ATLAS)
- Summary

Goal: To study jet substructure observables sensitive to quark and gluon as well as the ability of various models to describe these

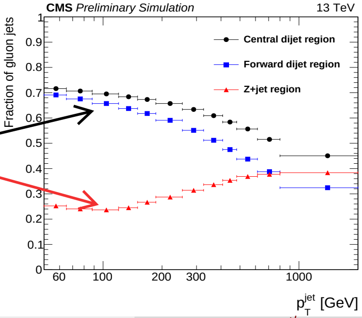
Z($\rightarrow \mu\mu$)+jet (quark enriched) selection:

- Z($\mu^+\mu^-$): $\geq$ 2 muons, $|\eta_\mu| < 2.4$, $p_T^\mu > 26$ GeV, $|m_{\mu^+\mu^-} - m_Z| < 20$ GeV,
- Reconstructed (generator)-level jets (j): $\geq$ 1 jet with $p_T > 30$ (15) GeV, $|y| < 1.7$
- Not overlapping with muons of Z
- $\Delta\phi(j_1, Z) > 2$, $|p_T^{j1} - p_T^Z|/|p_T^{j1} + p_T^Z| < 0.3$

Dijet (gluon enriched in central & forward) selection:

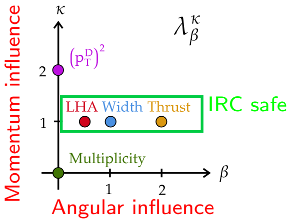
- Reconstructed (genrator)-level jets: $\geq$ 2 jets with $p_T > 30$ (15) GeV, $|y| < 1.7$
- $\Delta\phi(j1, j2) > 2$, $|p_T^{j1} - p_T^{j2}|/|p_T^{j1} + p_T^{j2}| < 0.3$

- Z+jet sample $\rightarrow$ 64-76% quark jets
- Central and forward dijet jets $\rightarrow$ Dominated by 69-72% gluon jets at low p_T & 55-68% quark jets at high p_T



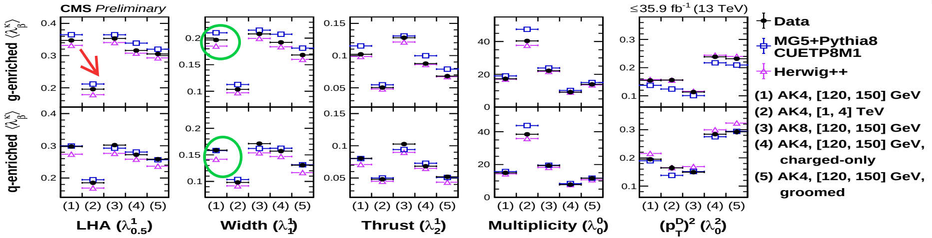
Results: Study of quark and gluon jet substructure in dijet and Z+jet events from pp collisions CMS-PAS-SMP-20-010

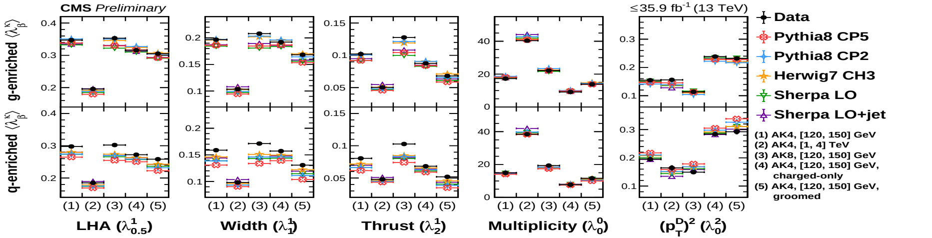
- Observables: 5 generalized angularities (LHA, Width, Thrust, Multiplicity & $(p_T^D)^2$), jets with charged+neutral & charged-only constituents, groomed & ungroomed jet, angularities with different jet radii
  - Five λ_β^κ observables on basis of β= 0, 0.5, 1, 2 & κ=0, 1, 2 define as:
    $\lambda_\beta^\kappa = \sum_{i \in jet} z_i^\kappa (\frac{\Delta R_i}{R})^\beta$, $z_i = \frac{p_{Ti}}{\sum_{j \in jet} p_{Tj}}$
    - Hypothesis: Larger values of LHA, width, thrust, multiplicity & low value of $(p_T^D)^2$ for gluon than quark jets



  - Groomed jet: Find the widest hard jet splitting by declustering and following the hardest branches (suppresses effects from pileup, underlying event, and initial-state radiation, which are more difficult to model)
    - Test the Soft Drop condition (z cut = 0.1, β = 0): Only splittings where the subleading prong carries at least 10% of the combined transverse momentum are accepted.
    - $z = \frac{min(p_T^{(1)}, p_T^{(2)})}{p_T^{(1)} + p_T^{(2)}} > z_{cut} (\frac{\Delta R_{1,2}}{R_{jet}})^{\beta_{sd}}$, if condition is not met, the declustering procedure is repeated with the subjet that has the larger p_T of the two, and the other subjet is rejected

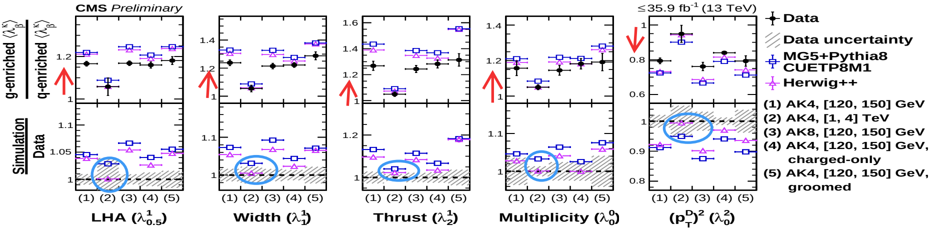
Results: Study of quark and gluon jet substructure in dijet and Z+jet events from pp collisions CMS-PAS-SMP-20-010

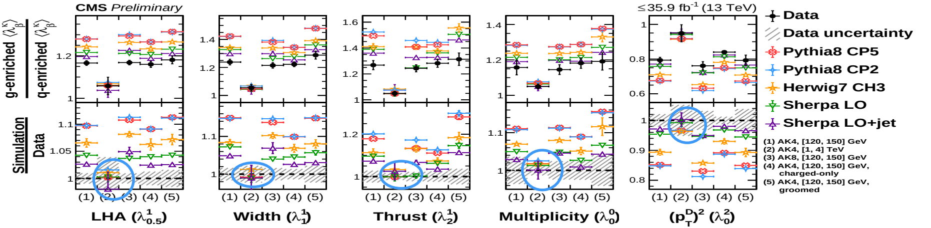




- Quark enriched sample: MG5+PYTHIA8 provides the best description, followed by HERWIG7, PYTHIA8 CP2 ($\alpha_s(m_Z)$=0.130), SHERPA , HERWIG++, and PYTHIA 8 CP5 ($\alpha_s(m_Z)$=0.118)
- Gluon enriched sample: HERWIG7 CH3, PYTHIA8 CP5, PYTHIA8 CP2, & SHERPA provide a better description than either HERWIG++ or MG5+PYTHIA8
- Improved modelling of gluon jets at the cost of poorer modelling of quark jets is observed

Results: Study of quark and gluon jet substructure in dijet and Z+jet events from pp collisions CMS-PAS-SMP-20-010





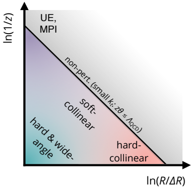
- Similar data-to-simulation agreement: AK8 vs AK4 jets, charged-only vs charged+neutral, groomed vs ungroomed observables
- Ratio (for all generator) → Significantly larger than unity except $(p_T^D)^2$: Showing the clear need for improvements in the MC at low p_T
- The best overall data-to-simulation agreement for the ratio is achived by SHERPA, followed by HERWIG^{++}, MG5^{+} PYTHIA 8, HERWIG 7 CH3, PYTHIA 8 CP2, and PYTHIA 8 CP5.

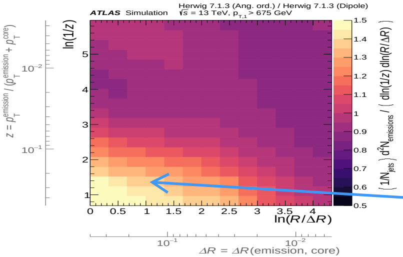
Goal: To use the Lund Plane of primary jet emissions to isolate various QCD effects in jets

- Calculation of number of emissions within regions of the Lund jet plan (LJP) provides optimal discrimination between quark and gluon jets
- In QCD parton emissions are characterized by a) emission opening angle (θ) b) momentum fraction (z) of emitted gluon w.r.t to primary quark and gluon
- Plane (ln(1/z), ln(1/θ)) is the 'Lund jet plane' $\rightarrow$ Useful to study jet substructure
  - Jets: Jets constituents are reclustered using the C/A algorithm using angle-ordered hierarchy
  - Followed in reverse ('declustered'), starting from the hardest proto-jet.
  - For each proto-jet pair, at each step $\rightarrow$ Individual jets are represented as a set of points within LJP

Why Lund jet plane ?

- Contributions from various QCD effects: ISR, UE, MPI, hadronization, & perturbative emissions are well-localized in the LJP

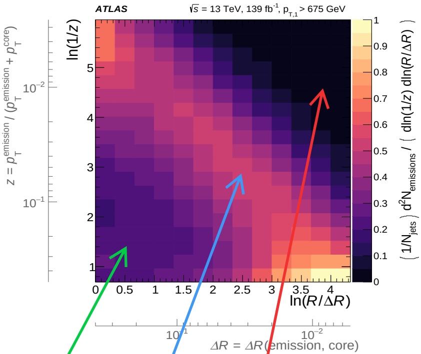


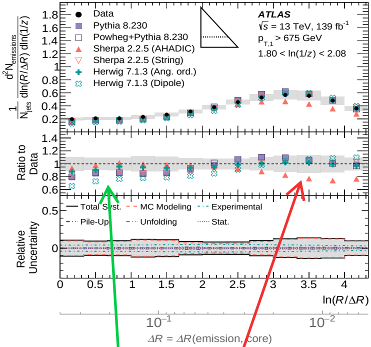


- Ratio of simulation at charged-particle level varied only one component (PS model) of simulation$\rightarrow$ Differences of up to 50% in the perturbative hard and wide-angle emissions

The Lund Jet plane measured in dijet events:

- Jets: Anti-k_T=0.4, $p_T^{leading} > 675$ GeV, $|\eta| < 2.1$
- LJP observables: Rebuilding C/A jets using tracks in dijets events & then declustering





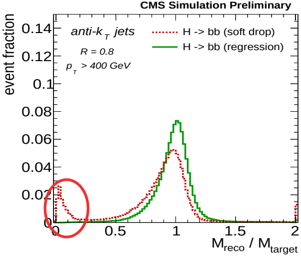
- $k_T \gtrsim \Lambda_{QCD}$: uniformly populated by perturbative emissions
- Transition to the nonperturbative regime: Large number of emissions
- Nonperturbative region: Emission suppressed

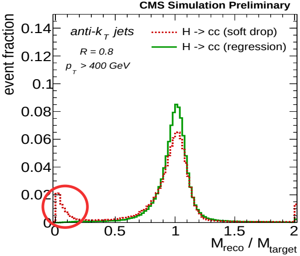
- Hard and widest emission: Angular-ordererd PS performs well
- Hadronisation region better described by string-based sherpa models

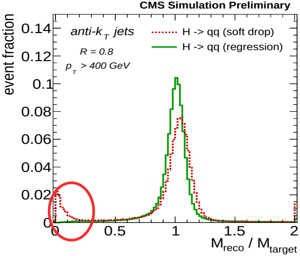
Conclusion: No generator match data in full plane

Mass regression: Machine learning (ML) technique to reconstruct the mass of hadronically decaying highly Lorentz-boosted heavy particles

- ParticleNet Tagger: Network is trained using PF candidates & secondary vertices associated with AK8/AK15 jet.
- ParticleNet Mass Regression: In addition to ParticleNet Tagger training sample consists of an equal mix of QCD and Higgs bosons event generated with MG+PYTHIA8. The Higgs boson sample generated with an equal mix of H→ bb/cc/qq (q=u,d,s) decays.
- Target Mass (M_{target}): Heterogeneous nature→ "soft drop" mass for the QCD sample otherwise Higgs boson generator mass in [15, 250] GeV range.

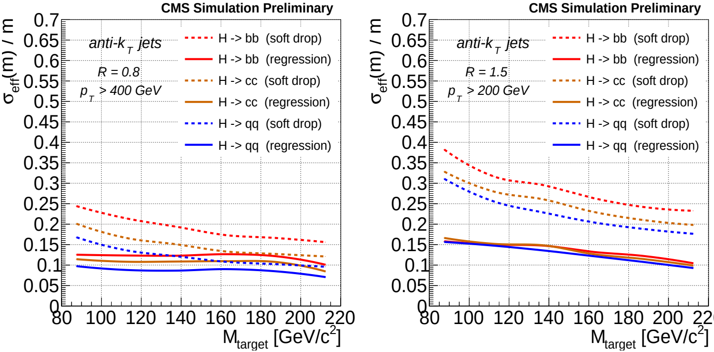






- Mass regression shows a substantial improvement in the mass resolution and in the absolute scale.
- Tails are also strongly mitigated with the mass regression, in particular at M∼0, where the soft drop algorithm incorrectly identifies the large R jet as single-prong.

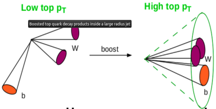
Results: Mass regression of highly-boosted jets using graph neural networks DP-2021-017



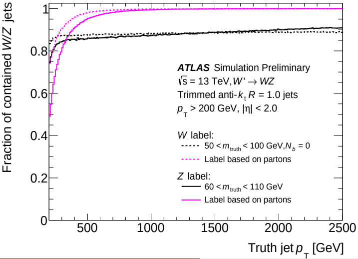
- The mass regression shows a substantial improvement in the resolution for all the considered mass range.

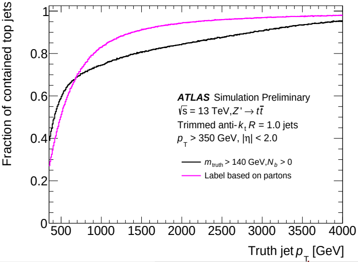
Results: Boosted hadronic vector boson and top quark tagging with ATLAS using Run 2 data ATL-PHYS-PUB-2020-017

Goal: In order to enhance the sensitivity of W & Z bosons, Higgs bosons, and top quarks final states, taggers are designed to identify large-R jets that originate from boosted hadronic decays.

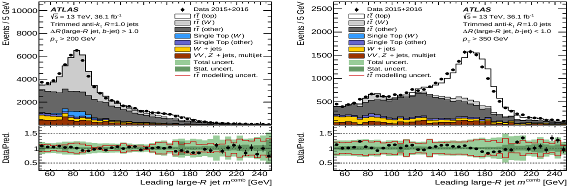


- Labeling procedure for a jet (have smaller generator dependence):
  - Inclusive tops aim at identifying tops where just parts of the decay are within the large-R jets
  - Contained top aim at reconstructing tops where all the decay products are within the jet
  - Inclusive top $\rightarrow \Delta R(\text{jet}_{truth}, \text{truth top quark}) < 0.75$ after FSR
  - Contained top jet $\rightarrow$ If an inclusive top jet has ≥ 1 b-hadron, & $m_{turth} > 140$ GeV
  - W $\rightarrow \Delta R(\text{jet}_{truth}, \text{truth W boson}) < 0.75$, no b-hadrons, $50 < m_{truth} < 100$ GeV
  - Z $\rightarrow$ If it fails the W label criteria & $\Delta R(\text{jet}_{truth}, \text{truth Z boson}) < 0.75$, $60 < m_{truth} < 110$ GeV



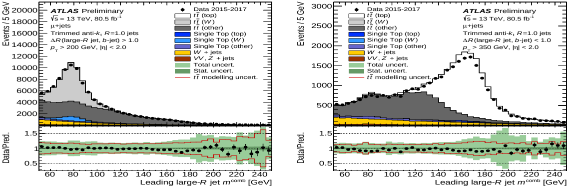


Eur.Phys.J.C79 (2019) 375



Without JES correction

ATL-PHYS-PUB-2020-017

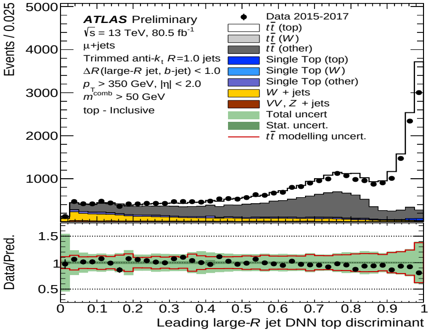


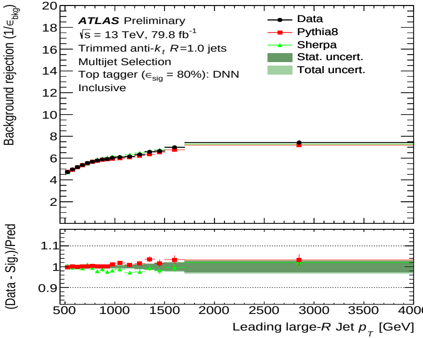
With JES correction → Improved data and MC agreement near the mass peaks

Left: W enriched events Right: Top enriched events

- W & Z tagger→ Use selections on the large-R jet mass, energy correlation function ratio ($D_2^{\beta=1.0}$) and the ghost-associated track multiplicity
- Top tagger → Deep neural networks trained on hadronic jet properties, including several jet substructure moments

ATL-PHYS-PUB-2020-017





- Inclusive top taggers→ Good agreement across the region within the uncertainties
- Background efficiency scale factors→ Close to unity, dominant uncertainties are statistical and modeling

Summary

- Study of quark and gluon jet substructure in dijet and Z+jet events from pp collisions (CMS):
  Means of the angularities in quark-and gluon-enriched data samples demonstrated their discrimination power, overestimated by all generators, showing the clear need for improvements in the MC.
- Measurement of the Lund jet plane using charged particles in 13 TeV proton-proton collisions with the ATLAS detector (ATLAS):
  No generator match data in full Lund jet plane so it can provide useful input to both perturbative and nonperturbative model development and tuning
- Mass regression of highly-boosted jets using graph neural networks (CMS):
  Shows a substantial improvement in the resolution for all the considered mass range as compare to more traditional grooming algorithms
- Boosted hadronic vector boson and top quark tagging with ATLAS using Run 2 data (ATLAS):
  Future improvements for top tagging could include the addition of track variables in the DNN after considering the background rejection gain along with the impact of the additional associated uncertainties.

Thank You