Measurements of jet substructure observables in ATLAS and CMS experiments

Cristian Baldenegro (LLR-École Polytechnique)
on behalf of the ATLAS and CMS Collaborations

QCD@LHC 2022
IJCLab Orsay, Paris

Laboratoire Leprince-Ringuet

What is jet substructure?

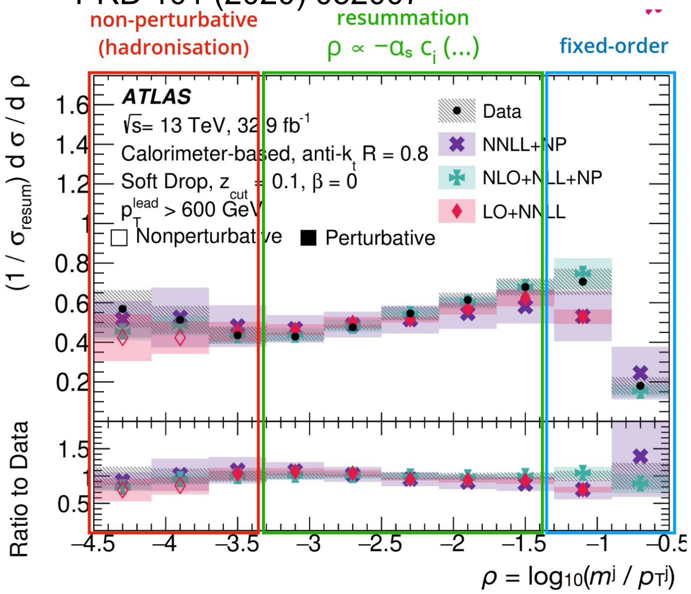
Jet constituents four-momenta are mapped onto physically meaningful observables:

$$\{p_i\} \rightarrow \lambda$$

sketch by Jesse Thaler

Why do we do jet substructure?

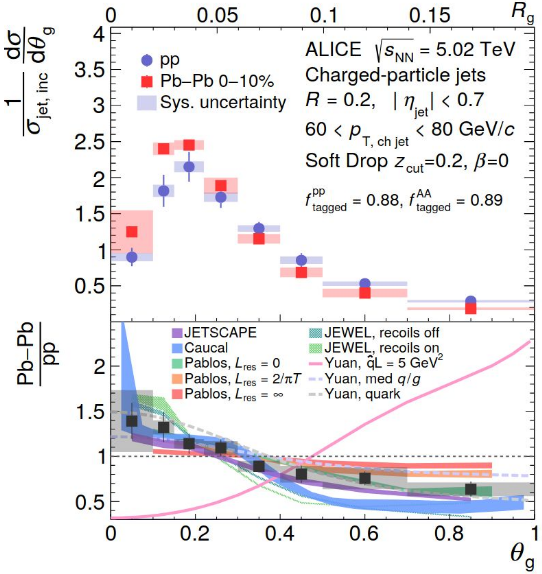
ATLAS, arXiv:1912.09837,
PRD 101 (2020) 052007



Test state-of-the-art pQCD calculations (resummation + fixed-order).

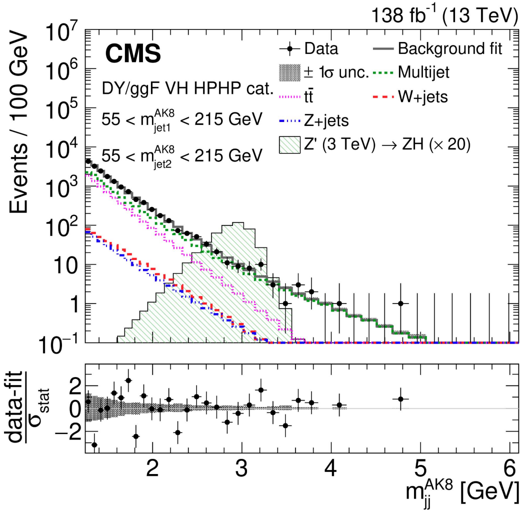
Constrain parton shower & hadronization modeling in MC generators.

ALICE, arXiv:2107.12984,
PRL 128 (2022) 102001



Space-time picture of interactions in the quark-gluon plasma, color coherence, …
(see talk by Raymond Ehlers).

CMS, arXiv:2210.00043,
submitted to PRL



BSM searches in boosted decay topologies
(see talks by *Antimo Cagnotta*, *Francesco Carnevali*).

A field in rapid development!

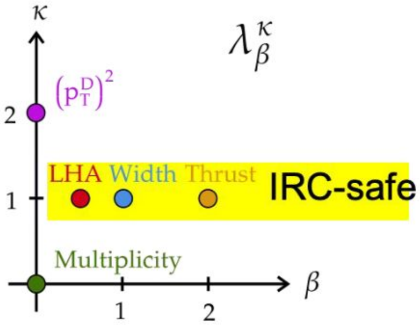
What will be covered in this talk:

- Generalized angularities in dijet and Z-jet by CMS,
  ***arXiv:2109.03340*, JHEP 01 (2022) 188**
- Inclusive *b*-hadron decays in top quark pair events by ATLAS
  arXiv:2202.13901, **PRD 106 (2022) 032008**
- Exclusive b-hadron decays in inclusive jet events by ATLAS
  arXiv:2108.11650, ***JHEP 12 (2021) 131***

Snapshot of recent jet substructure measurements at the LHC (LHC EW WG):
https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCJetSubstructureMeasurements

Generalized angularities in Z+jet and dijets

CMS, *arXiv:2109.03340*, *JHEP 01 (2022) 188*



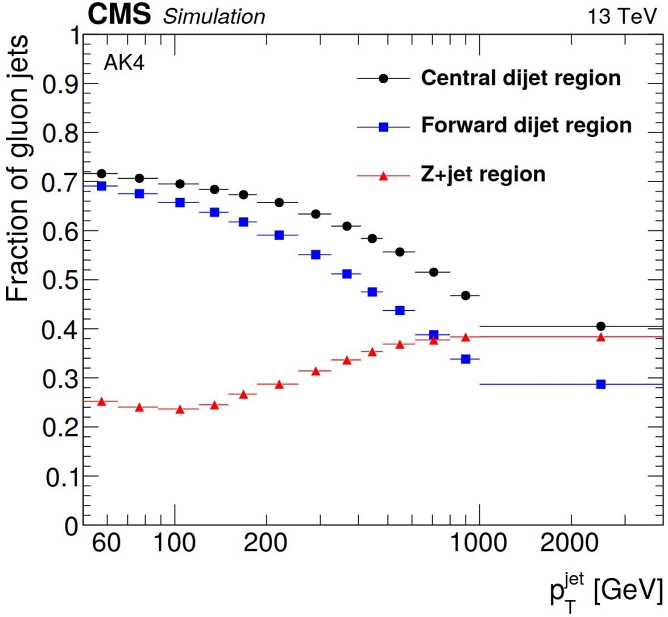
Observables suggested by theorists to separate quark and gluon jets (sum over jet constituents)

$$\lambda_\beta^\kappa = \sum_{i\in \text{jet}} z_i^\kappa \left(\frac{\Delta R_i}{R}\right)^\beta \qquad z_i \equiv \frac{p_{Ti}}{\sum_{j\in \text{jet}} p_{Tj}}$$

κ and β parameters control weight on momentum and angular scales

Z+jet, quark enriched ***dijet, gluon enriched***

Purities of up to ~75% for quark (gluon) jets



Example: Les Houches Angularity distribution

κ=1, β = 0.5

CMS, ***arXiv:2109.03340***, ***JHEP 01 (2022) 188***

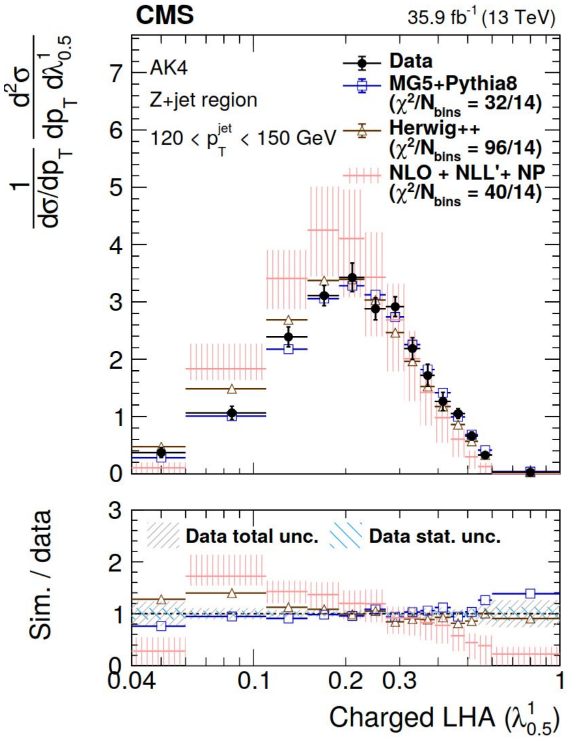
Data unfolded to particle level (regularized unfolding with Tikhonov algorithm).

MG5+Pythia8 and **Herwig++** describe well the quark-enriched data; they envelop the gluon-enriched data.

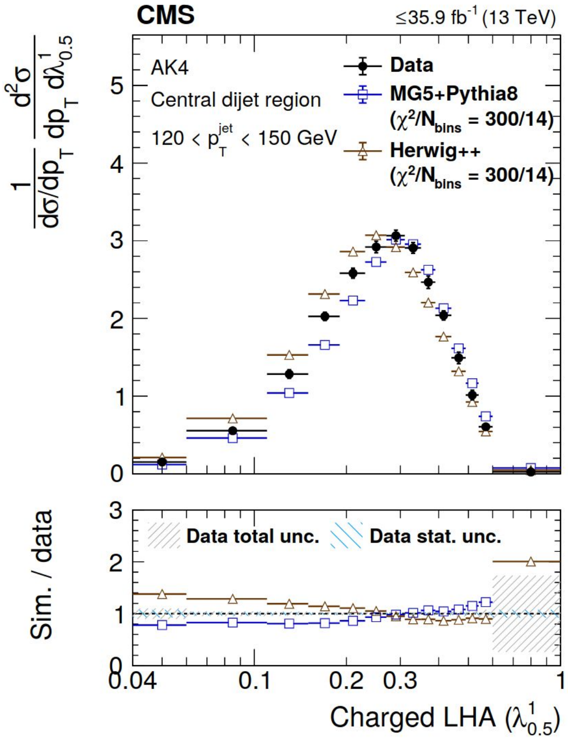
For Z+jet: resummation at NLL matched to fixed-order NLO matrix elements, with NP corrections from Sherpa (**NLO+NLL'+NP**)

D. Reichelt, S. Caletti, O. Fedkevych, S. Marzani, S. Schumann, G. Soyez ***arXiv:2112.09545***

Z+jet, quark enriched

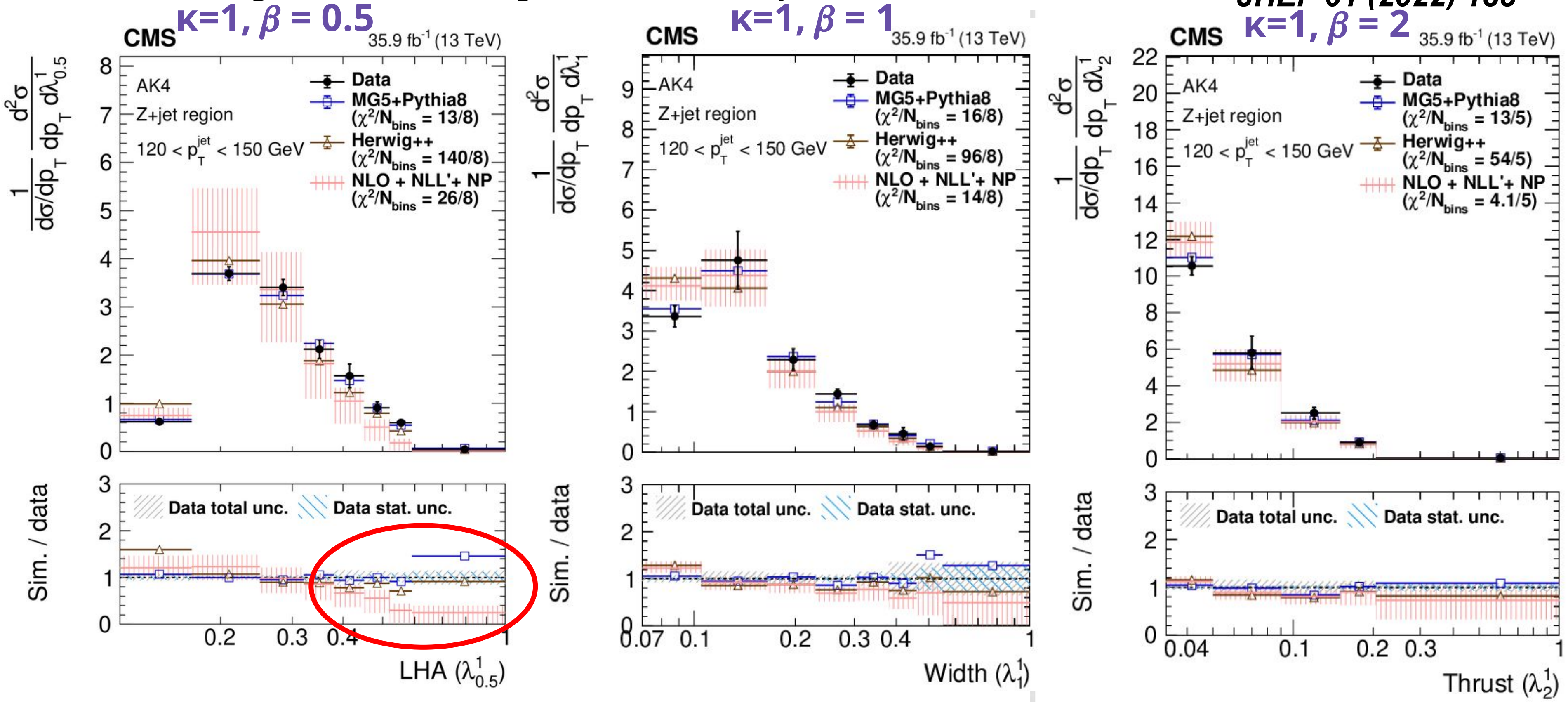


Dijet, gluon enriched



Ungroomed generalized angularities in Z-jet events

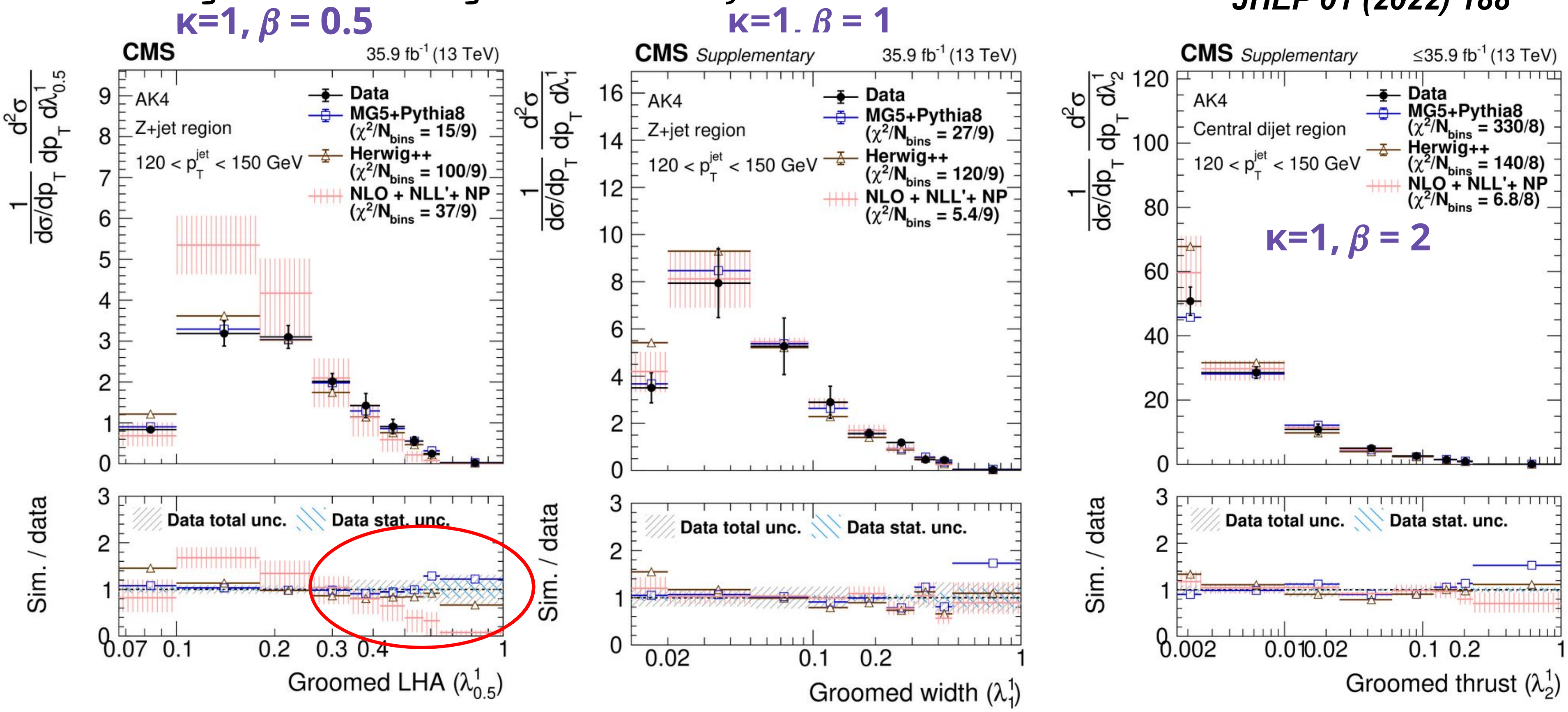
CMS, arXiv:2109.03340, *JHEP 01 (2022) 188*



The more weight is given to angular scale, the better agreement of theory with data.

Groomed generalized angularities in Z-jet events

CMS, arXiv:2109.03340, *JHEP 01 (2022) 188*



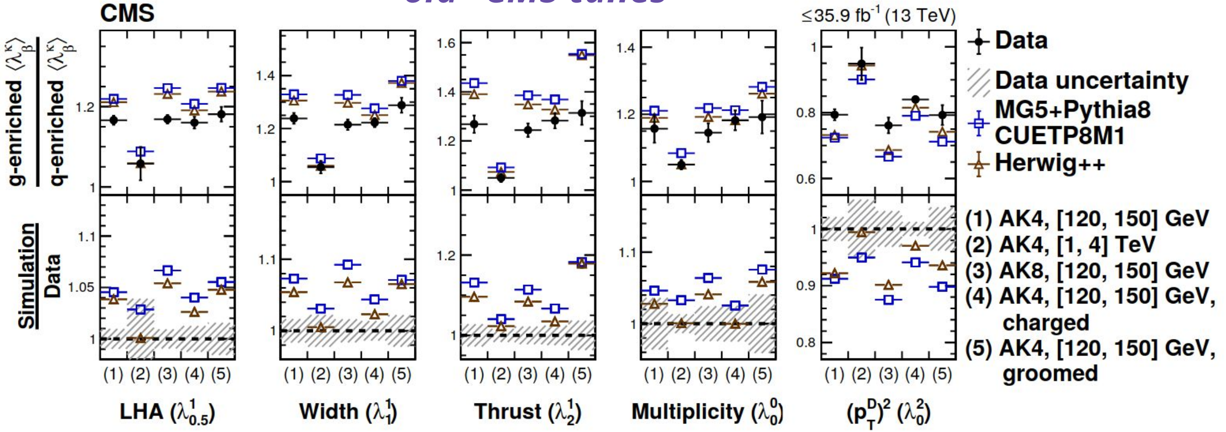
Soft-drop grooming to remove soft, wide-angle radiation (z_{cut} = 0.1, β = 0)
Tension at small β=0.5 persists, related to hard collinear splittings description?

Dijet/Z+jet ratio (gluon-like/quark-like jet ratio)

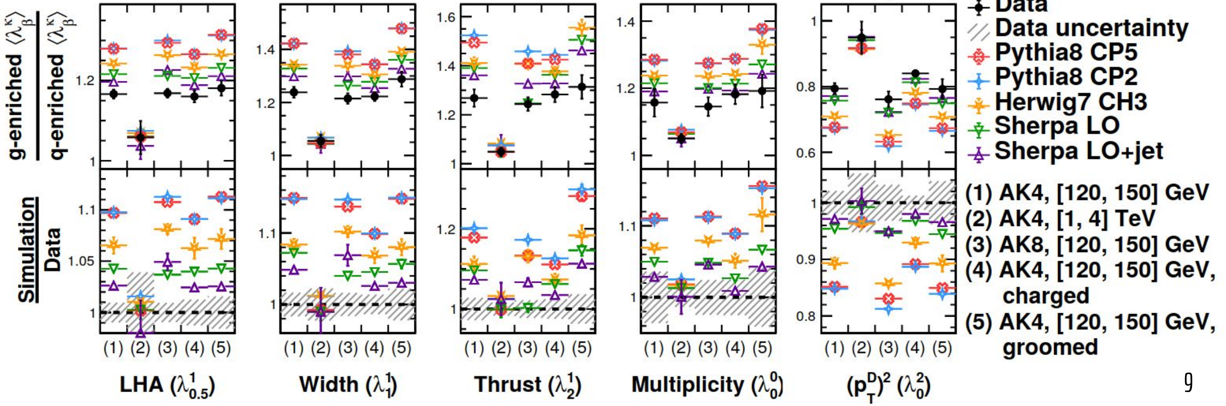
CMS, arXiv:2109.03340
JHEP 01 (2022) 188

- experimental uncertainties partially cancel in dijet/Z+jet ratio
- LO+PS preds. overestimate the g-enriched/q-enriched ratio
- g-enriched / q-enriched ratio is better modelled with “old” PYTHIA8 and HERWIG7 CMS tunes.

“old” CMS tunes



“new” CMS tunes



Fragmentation of b-quarks

- Fragmentation pattern of *b*-jets significantly different than their light-quark counterparts (harder fragmentation, dead cone effect as observed by ALICE for c-jets, arXiv:2106.05713, Nature 605 (2022) 440).

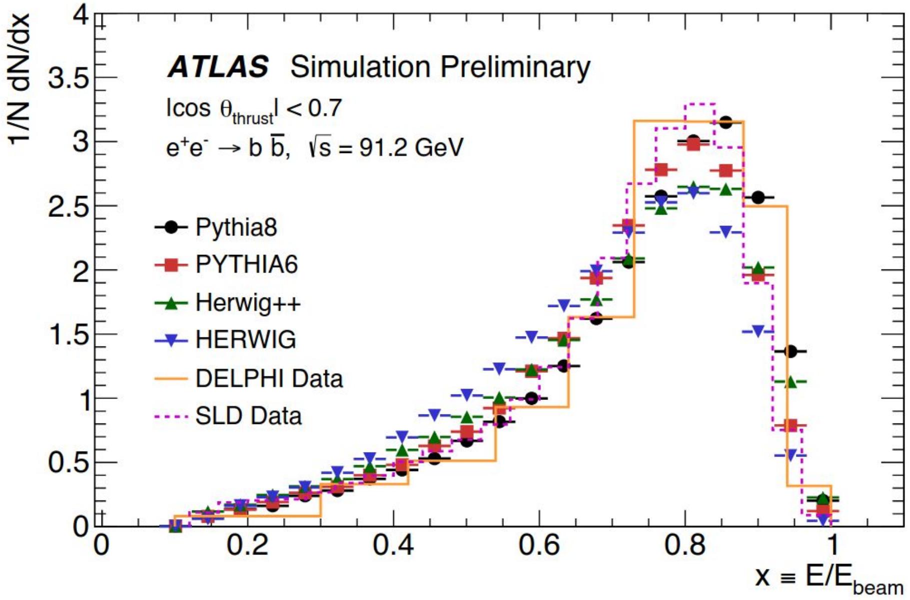
- Important for top quark and and H→ $b\bar{b}$ measurements.

- *b*-fragmentation tuned to e+e- data, then extrapolated to LHC.
  To what extent is this correct?

  (different color environment at LHC due to beam remnants + UE + ISR)

ATLAS-PHYS-PUB-2014-008

e+e- collisions



Predicted fragmentation function in e+e- collisions (after parton shower + hadronization)

Differences of O(10%) between different PS and hadronization models.

b-quark fragmentation in top quark pair events

ATLAS, arXiv:2202.13901, PRD 106 (2022) 032008

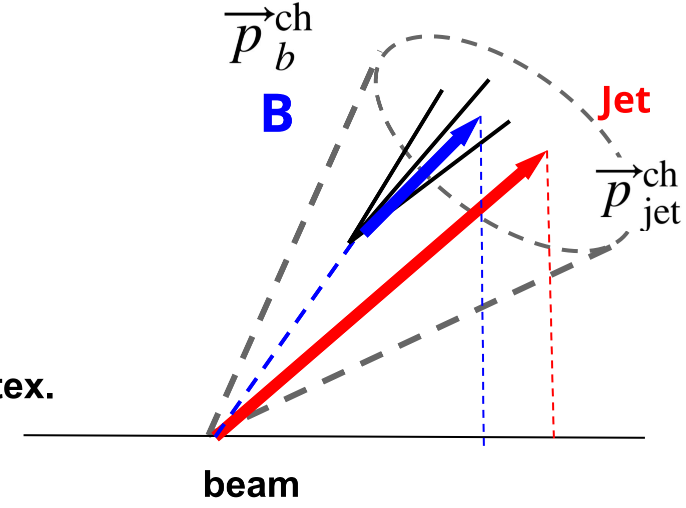
- Top quark sample rich in “primary” b-quark jets.
  1μ 1e golden channel is used for this measurement.

- exactly 2 *b*-jets with p_T>30 GeV, $|\eta| < 2.1$, and ΔR(jet, jet) > 0.5, clustered with anti-k_t algorithm with R = 0.4.

- **Proxy for *b*-hadron using charged particles in secondary vertex.**
  .

  Reconstruct 3-momentum of “charged” b-hadron $\vec{p}_b^{\,ch}$

  Compare with charged component of jet $\vec{p}_{jet}^{\,ch}$

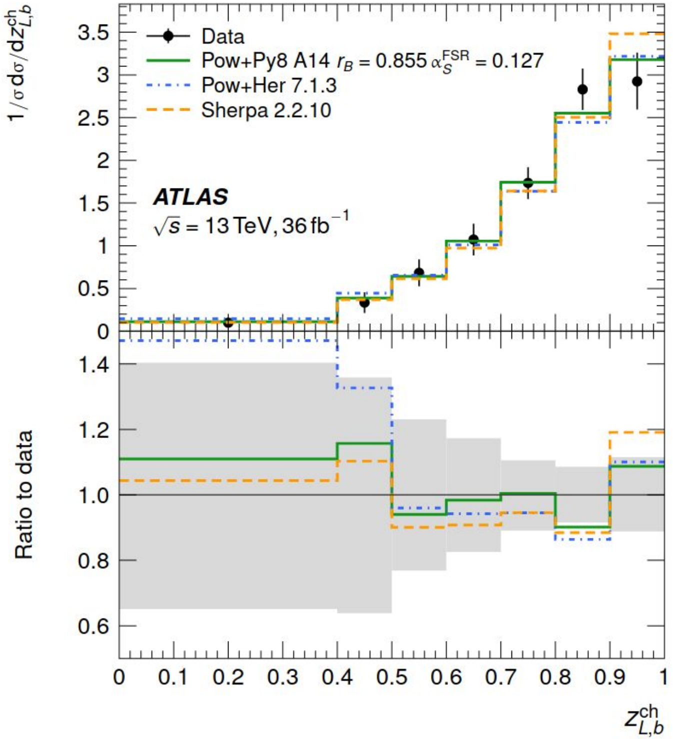


Longitudinal and transverse fragmentation variables

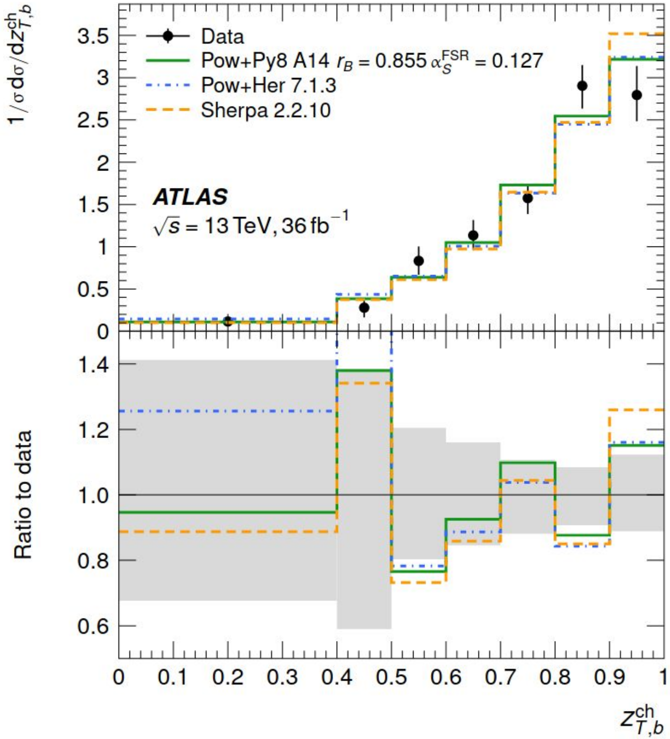
Fragmentation variables are fully corrected to particle-level with fully Bayesian unfolding method.

ATLAS, arXiv:2202.13901, PRD 106 (2022) 032008

$$z^{\mathrm{ch}}_{\mathrm{L},b} = \frac{\vec{p}_b^{\,\mathrm{ch}} \cdot \vec{p}_{\mathrm{jet}}^{\,\mathrm{ch}}}{|p^{\mathrm{ch}}_{\mathrm{jet}}|^2}$$



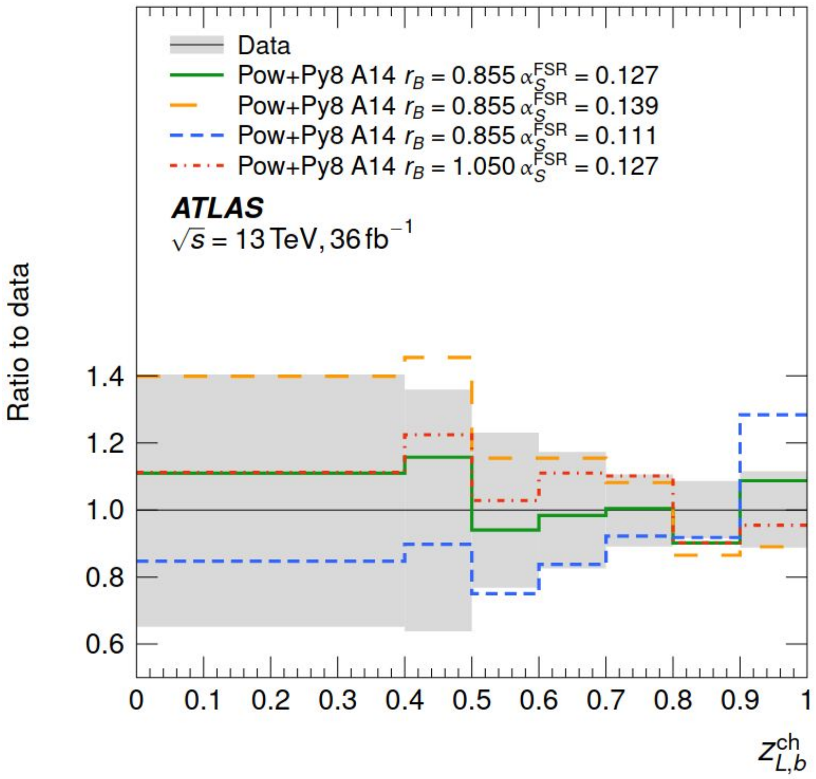
$$z^{\mathrm{ch}}_{\mathrm{T},b} = \frac{p^{\mathrm{ch}}_{\mathrm{T},b}}{p^{\mathrm{ch}}_{\mathrm{T,jet}}}$$



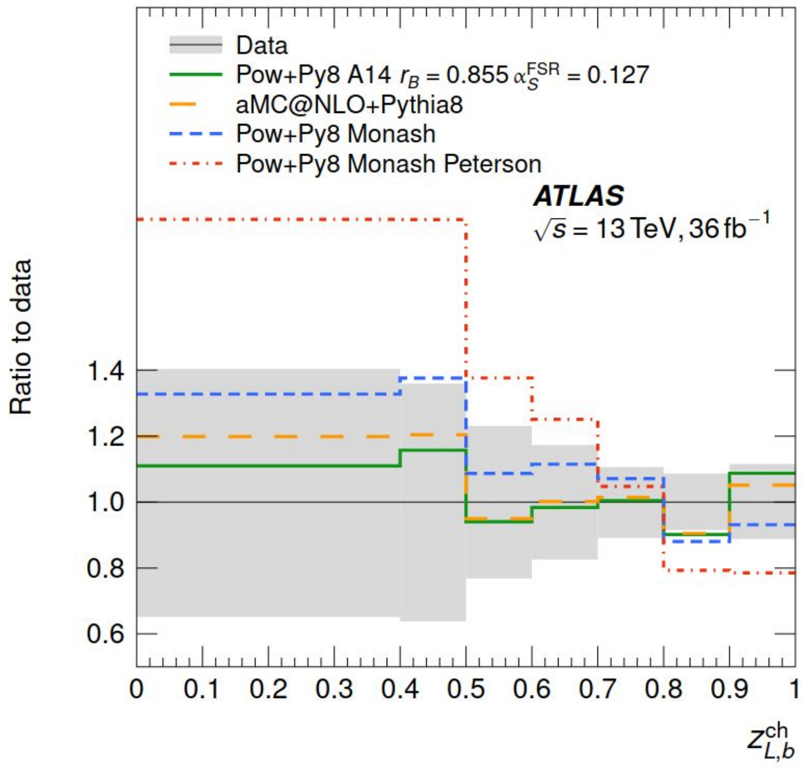
Simulations do a reasonable job in describing fragmentation variables

Sensitivity to fragm. function parametrization and α_s^{FSR}

ATLAS, arXiv:2202.13901
PRD 106 (2022) 032008



Lund-Bowler parametrization of frag. funct.

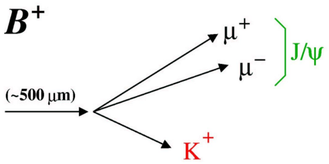


Tension with **Peterson** parametrization (when used with NLO+PS) at low $z_{L,b}^{ch}$

b-quark fragmentation via $B^\pm \to J/\psi\, K^\pm \to \mu^+\mu^- K^\pm$ decays

Full reconstruction of $B^\pm$ hadrons in b-jets using exclusive decay channel:

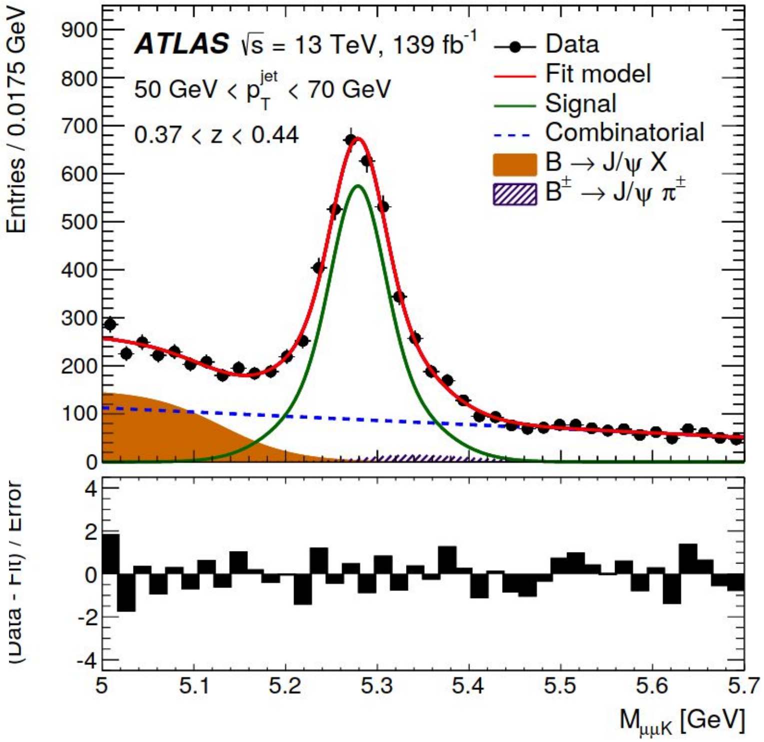
ATLAS, arXiv:2108.11650, *JHEP 12 (2021) 131*

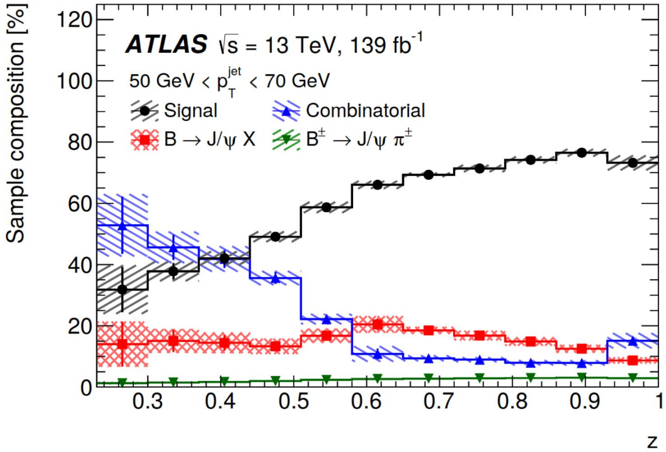


Analysis in inclusive jet sample (contributions from flavor excitation + flavor creation + gluon splitting) using R = 0.4 anti-k_T jets, p_T > 35 GeV and $|\eta|$< 2.1.

Two fragmentation variables are analyzed (showing *z* in this presentation)

$$z = \frac{\vec{p}_B \cdot \vec{p}_j}{|\vec{p}_j|^2}; \quad p_T^{\mathrm{rel}} = \frac{|\vec{p}_B \times \vec{p}_j|}{|\vec{p}_j|},$$





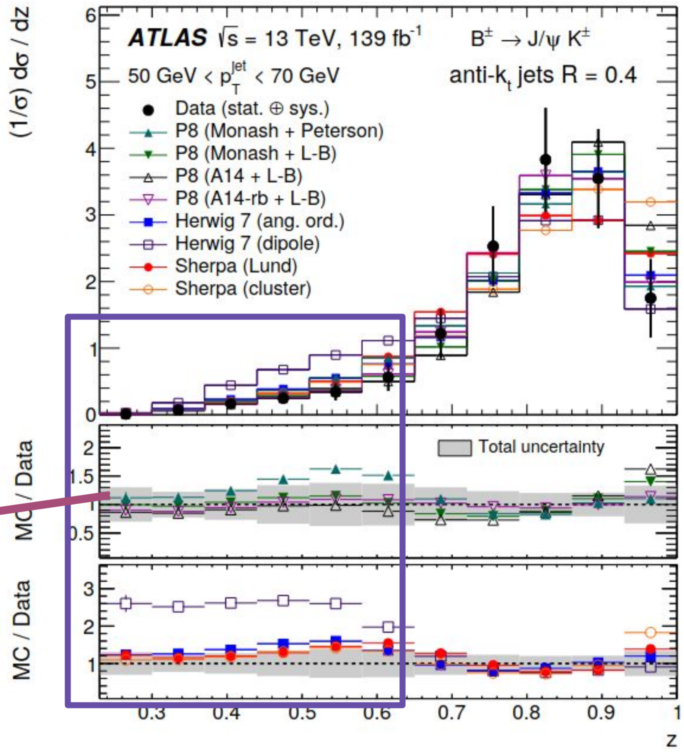
Fragmentation variable $z = \frac{\vec{p}_B \cdot \vec{p}_j}{|\vec{p}_j|^2}$

ATLAS, arXiv:2108.11650, *JHEP 12 (2021) 131*

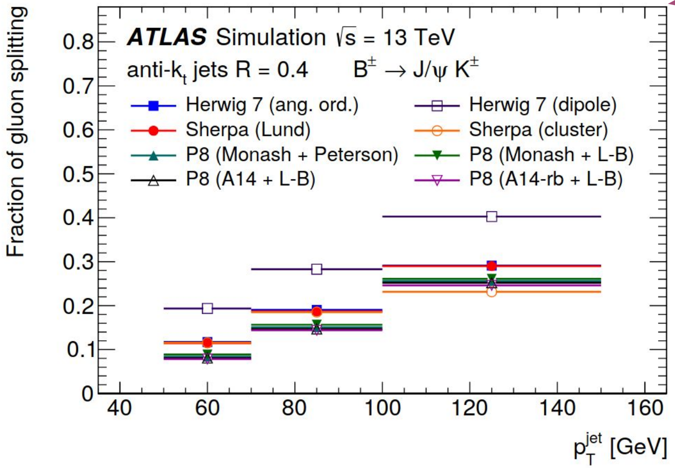
Observables fully corrected to particle level with iterative D'Agostini unfolding.

Tested PYTHIA8 tunes are consistent with data within the uncertainties.

Large differences between **angle-ordered** & **dipole** PS at low-z in Herwig7 associated with g→ bb modeling.



Sensitivity to hadronization



Main uncertainties related to mass fits, unfolding modeling, jet energy calibration.

Summary and outlook

- Valuable input for a better understanding of quark-jet and gluon-jet substructure from Z-jet and dijet events.

- *b*-quark fragmentation properties in b-jets from top quark decays and by analyzing *b*-hadron exclusive decays.

- Measurements allow us to constrain pQCD calculations and PS and hadronization modeling in MC generators.

Snapshot of recent jet substructure measurements at the LHC (LHC EW WG):
https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCJetSubstructureMeasurements

Back up

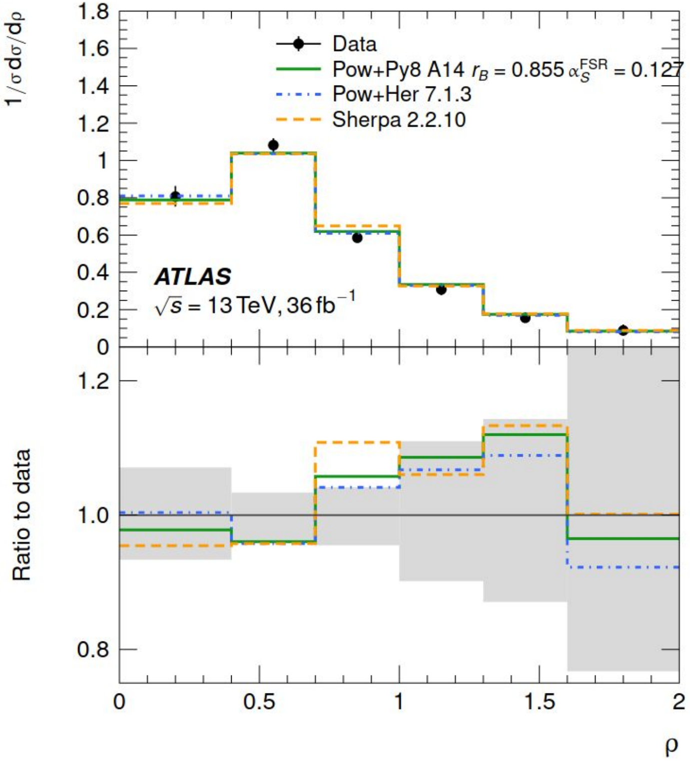
b-hadron p_T relative to leptons p_T and number of b-hadron decay products

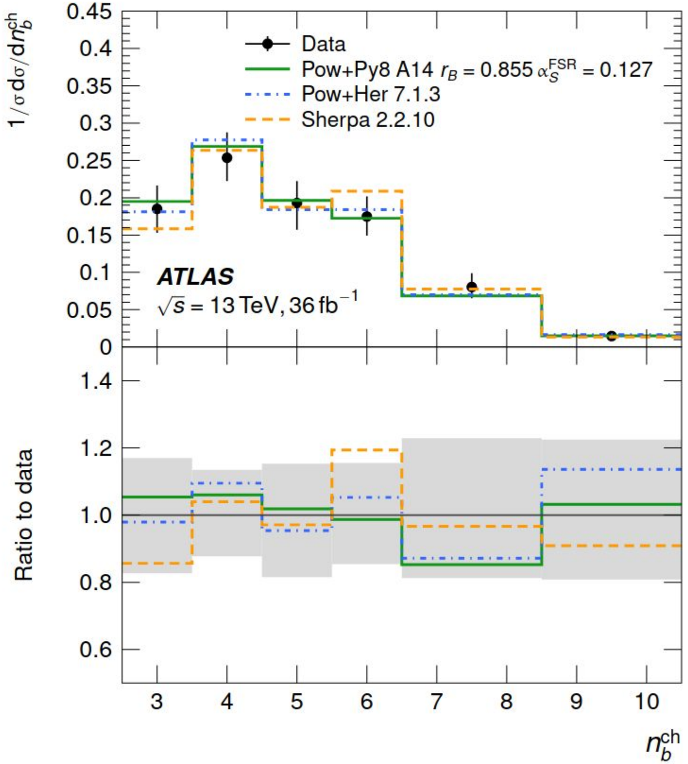
ATLAS, arXiv:2202.13901, PRD 106 (2022) 032008

$$\rho = \frac{2p^{\rm ch}_{{\rm T},b}}{p^e_{\rm T} + p^\mu_{\rm T}}$$

sensitive to α_s(FSR) and wide-angle radiation from top quark decay

number of b-hadron decay products, sensitive to b-hadron species production rates

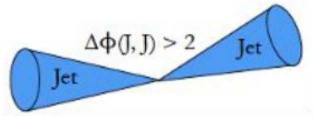




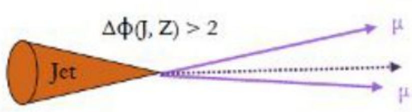
Simulations do a reasonable job in describing fragmentation variables

CMS, arXiv:2109.03340,
JHEP 01 (2022) 188

Event samples

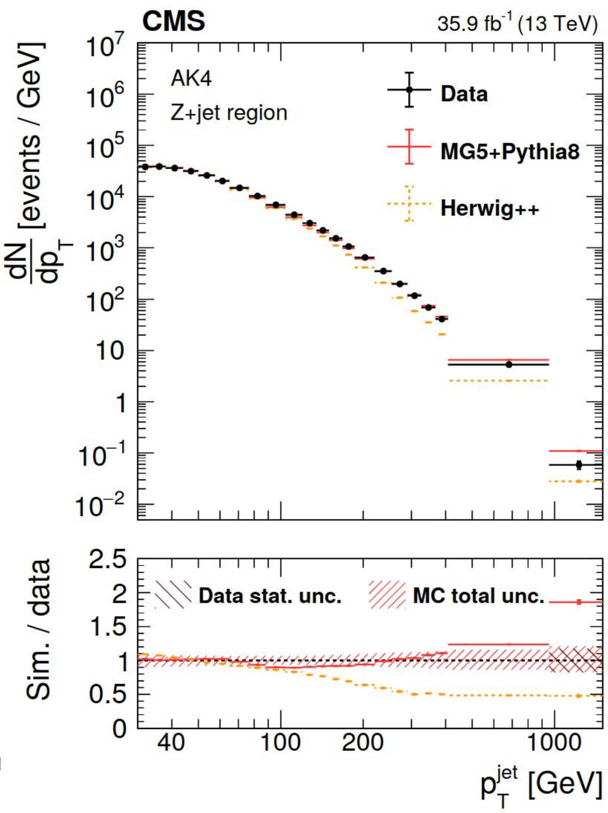


$\geq$2 jets with $|y| < 1.7$ and $p_T^j > 30\,\text{GeV}$
$\Delta\phi(j_1, j_2) > 2$
$|p_T^{j_1} - p_T^{j_2}|/(p_T^{j_1} + p_T^{j_2}) < 0.3$

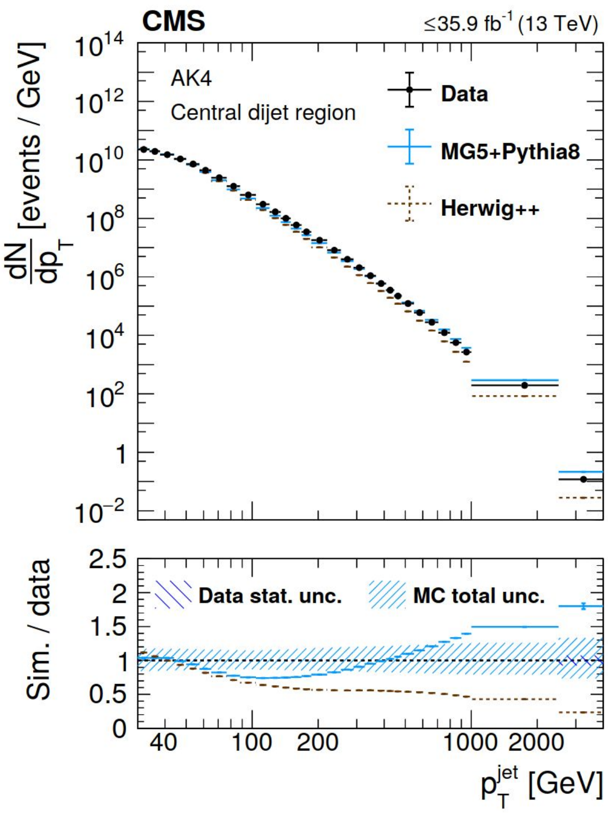


$\geq$2 muons with $|\eta| < 2.4$ and $p_T^\mu > 26\,\text{GeV}$
Opposite charge muons
$|m_{\mu\mu} - m_Z| < 20\,\text{GeV}$
$\geq$1 jet with $|y| < 1.7$ and $p_T^j > 30\,\text{GeV}$,
not overlapping with muons of the Z boson cai
$\Delta\phi(j_1, Z) > 2$
$|p_T^{j_1} - p_T^Z|/(p_T^{j_1} + p_T^Z) < 0.3$

Z+jet, quark enriched

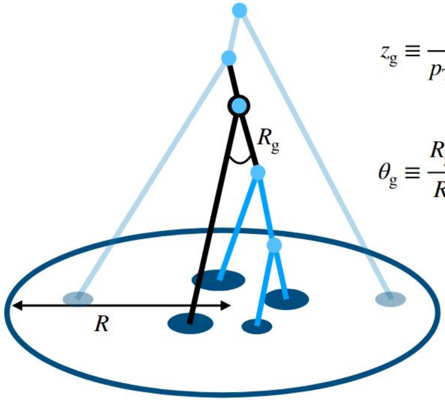


dijet, gluon enriched



Several points to consider in practice

- Ungroomed vs groomed jet substructure observables.
- Charged substructure (precision measurements, but IRC-unsafe) versus charged+neutrals (IRC-safe, more limitations on angular and momentum resolution).
- Quark vs gluon fraction (color factors), controllable via flavor tagging (Z-jet vs dijet, top quark decays, …).
- Jet distance parameter (trade-off between pileup sensitivity vs out-of-cone radiation)



$$z_g \equiv \frac{p_{T,\text{subleading}}}{p_{T,\text{leading}} + p_{T,\text{subleading}}}$$

$$\theta_g \equiv \frac{R_g}{R} \equiv \frac{\sqrt{\Delta y^2 + \Delta\varphi^2}}{R}$$

Soft drop grooming (algorithm to find the 1st hard subjet).

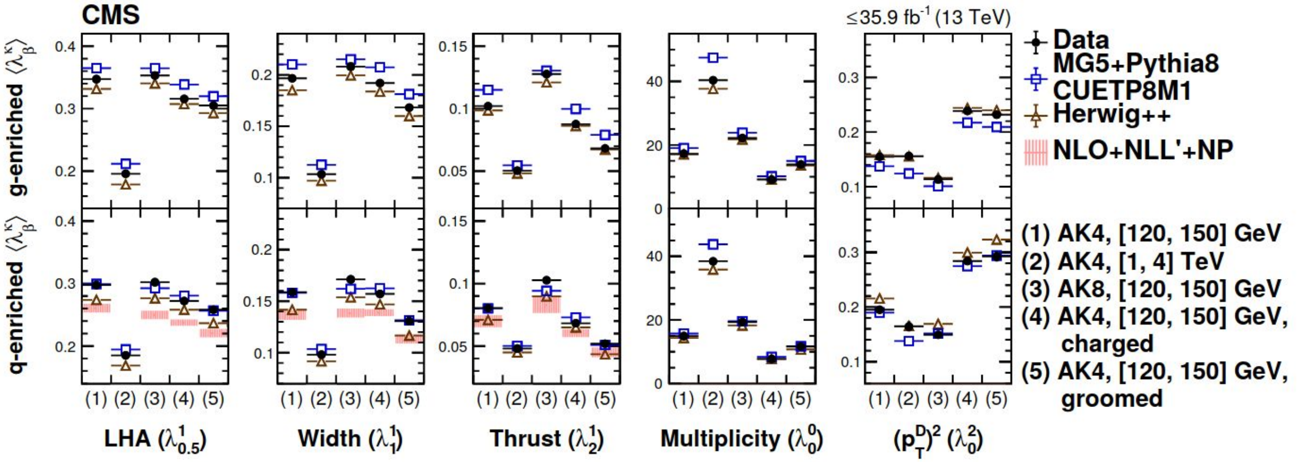
summary plot

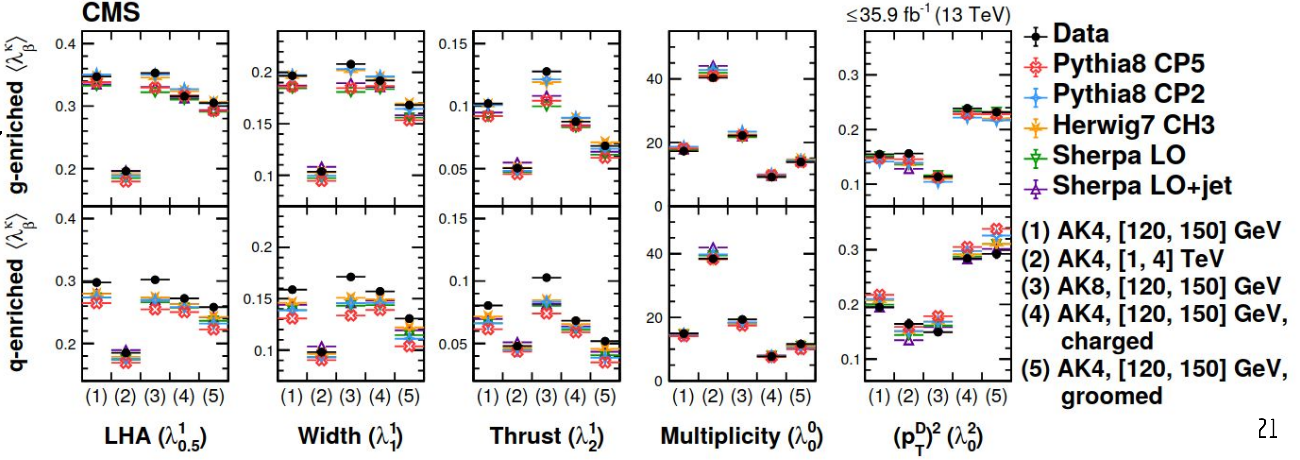
[arXiv:2109.03340]

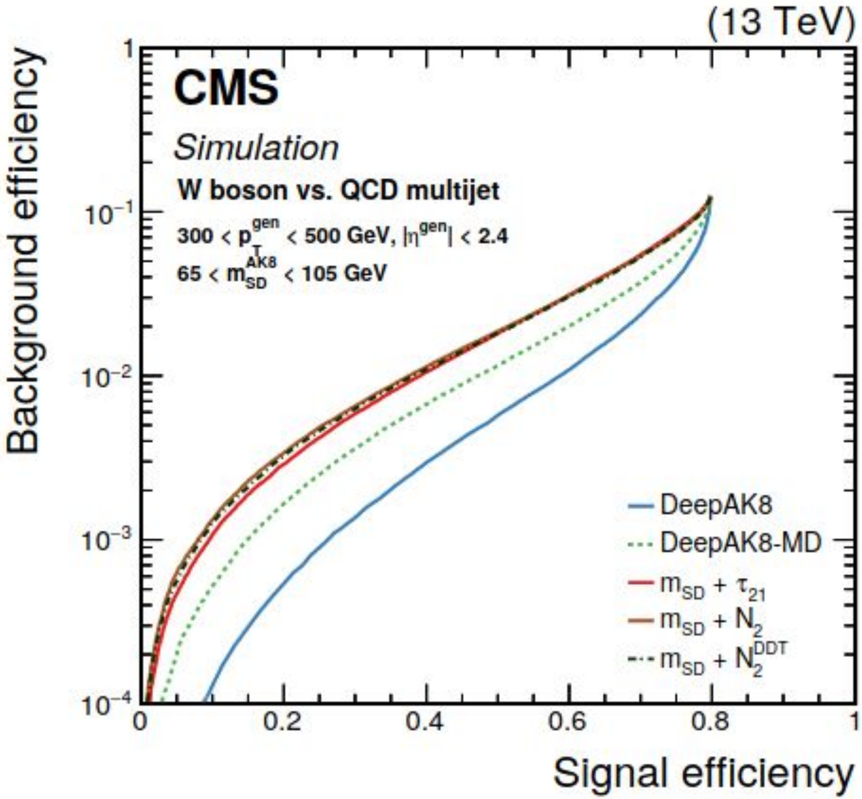
Angularities are generally larger in gluon-enriched jet samples, consistent with LO picture.

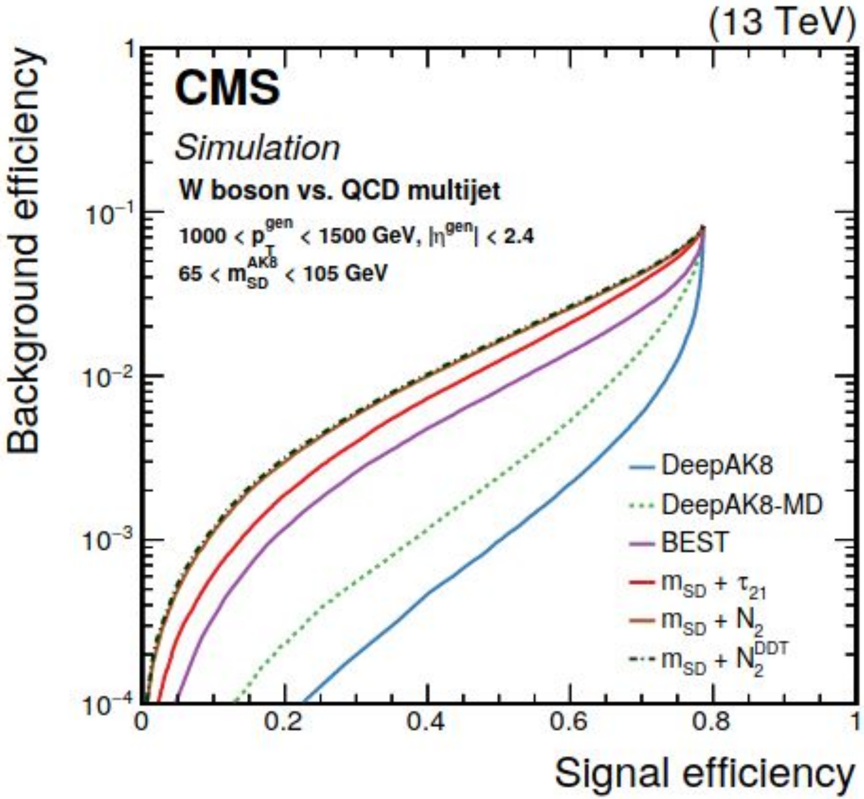
Quark- and gluon-initiated parton showers are not well described by generators, ***important for flavor tagging developments.***

Newest Pythia8 (CP2, CP5) and Herwig7 (CH3) tunes: description of gluon-like jets improves, not much improvement for quark-like jets.






(13 TeV)
CMS
Simulation
W boson vs. QCD multijet
300 < p_T^{gen} < 500 GeV, $|\eta^{gen}|$ < 2.4
65 < m_{SD}^{AK8} < 105 GeV
DeepAK8
DeepAK8-MD
$m_{SD} + \tau_{21}$
$m_{SD} + N_2$
$m_{SD} + N_2^{DDT}$
Background efficiency
Signal efficiency


(13 TeV)
CMS
Simulation
W boson vs. QCD multijet
1000 < p_T^{gen} < 1500 GeV, $|\eta^{gen}|$ < 2.4
65 < m_{SD}^{AK8} < 105 GeV
DeepAK8
DeepAK8-MD
BEST
$m_{SD} + \tau_{21}$
$m_{SD} + N_2$
$m_{SD} + N_2^{DDT}$
Background efficiency
Signal efficiency

Algorithm	Subsection	jet p_T [GeV]	t quark	W boson	Z boson	H boson
$m_{SD} + \tau_{32}$	6.1	400	✓			
$m_{SD} + \tau_{32} + b$	6.1	400	✓			
$m_{SD} + \tau_{21}$	6.1	200	✓	✓		
HOTVR	6.2	200	✓			
N_3-BDT (CA15)	6.3	200	✓			
$m_{SD} + N_2$	6.3	200		✓	✓	✓
BEST	6.5	500	✓	✓	✓	✓
ImageTop	6.6	600	✓			
DeepAK8$^{(*)}$	6.7	200	✓	✓	✓	✓
Jet mass decorrelated algorithms						
$m_{SD} + N_2^{DDT}$	6.3	200		✓	✓	✓
double-b	6.4	300			✓	✓
ImageTop-MD	6.6	600	✓			
DeepAK8-MD$^{(*)}$	6.7	200	✓	✓	✓	✓

Jet substructure in top quark pair + jet events

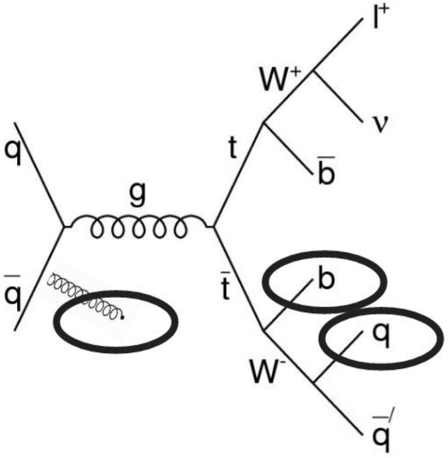
arXiv:1808.07340

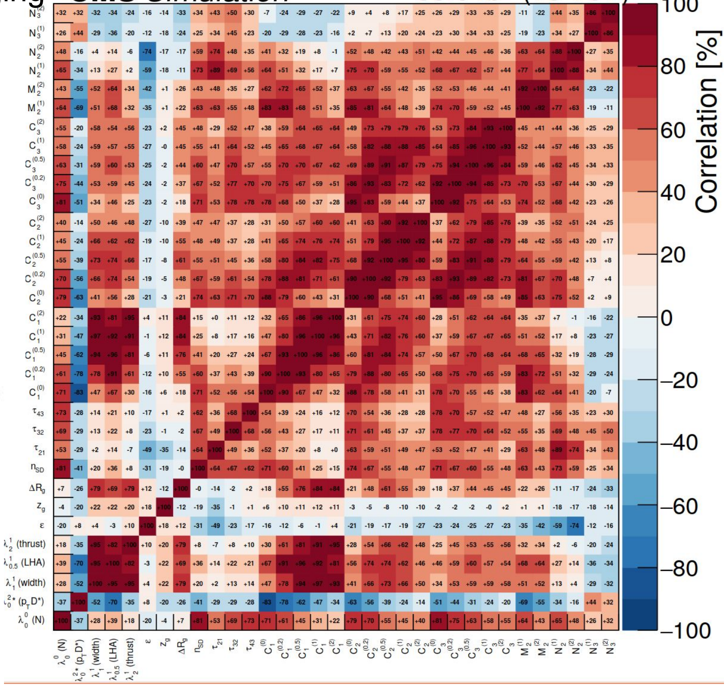
Top quark pair production provides: bottom, light-quark enriched, and gluon-enriched jet samples:

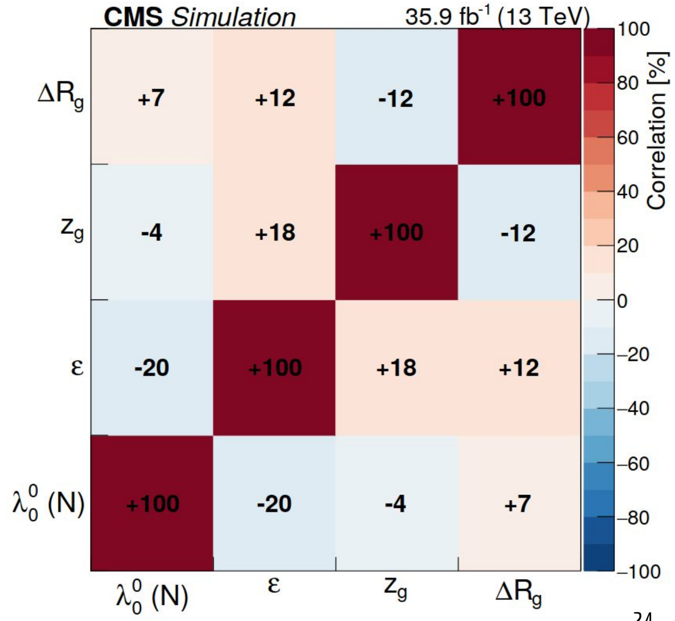
- ***Light-enriched:*** non b-tagged jets with |mjj − mW | < 15 GeV
- ***Gluon-enriched:*** non b-tagged jets with |mjj − mW | > 15 GeV
- ***B-quark* enriched** from b-tagging

33 observables were tested (Nsubjettiness, energy correlators, gen. angularities, …). A set of minimally-correlated variables is analyzed in detail:

ΔR_g, z_g, multiplicity (λ_0^0), eccentricity (ϵ)

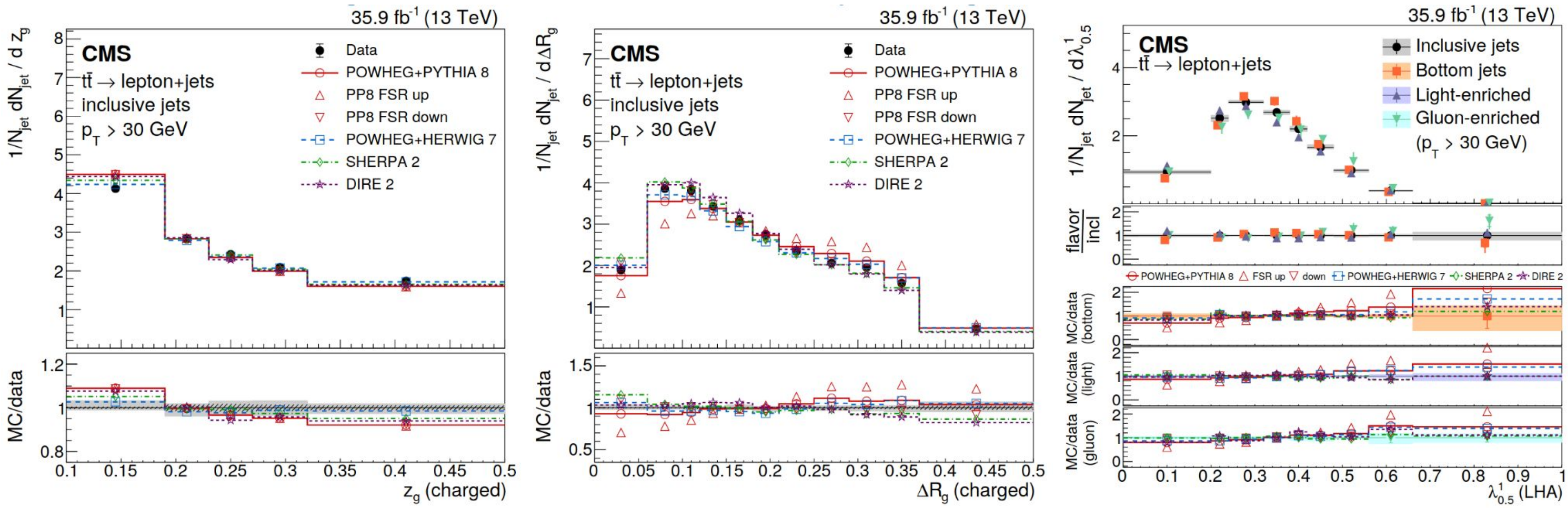






Measurement of jet substructure in top quark pair+jets

arXiv:1808.07340



- Groomed momentum fraction zg, related to the splitting function of QCD, at LO insensitive to αS
- Angle between groomed subjets ΔRg: sensitivity to αS, robust against non-perturbative corrections.
- LHA is expected to be larger for gluon-jets than quark jets. B-quark jets are expected to have smaller LHA.

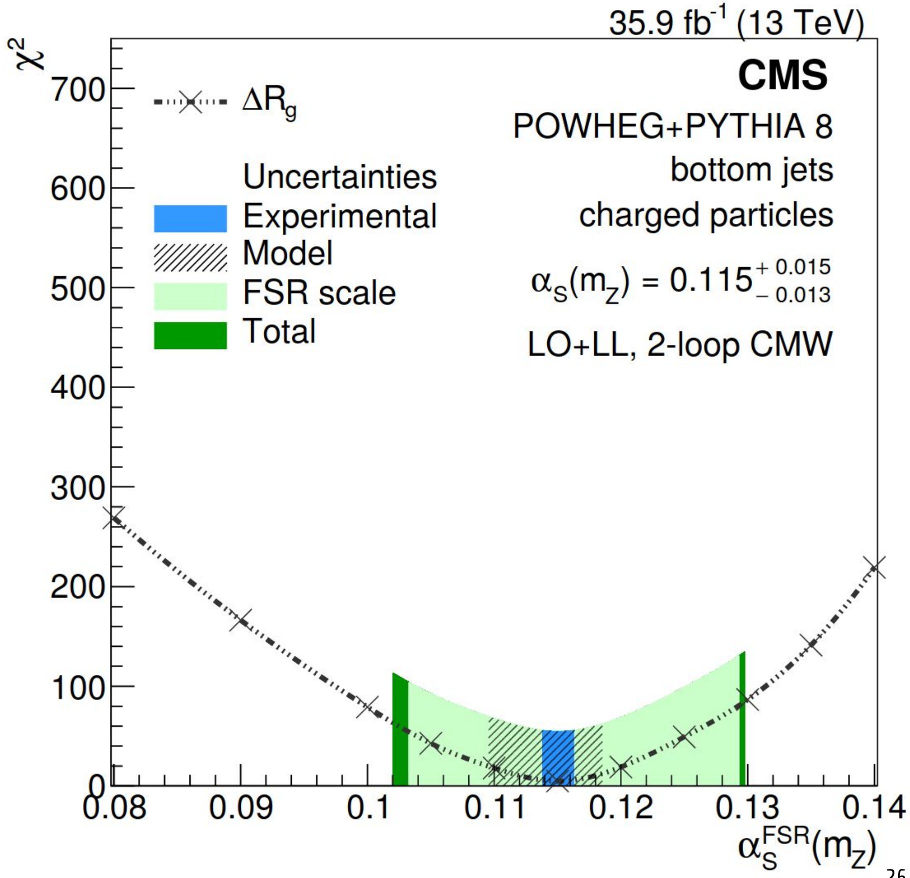
α_S extraction from ttbar lepton + jets

arXiv:1808.07340

- ΔRg is expected to be the least sensitive to NP effects.
- Substructure of b-jets is used for the extraction of αS (NLO matrix element of top quark decay in PYTHIA8, effectively LO+LL b-jet substructure)
- Higher-order corrections supplemented in an effective way using 2-loop running and CMW rescaling of ΛQCD.
- Result:

$$\alpha_S(m_Z) = 0.115^{+0.015}_{-0.013}$$

experimental uncertainties are less than 1% of αS, uncertainty in αS is dominated by FSR scale uncertainties.

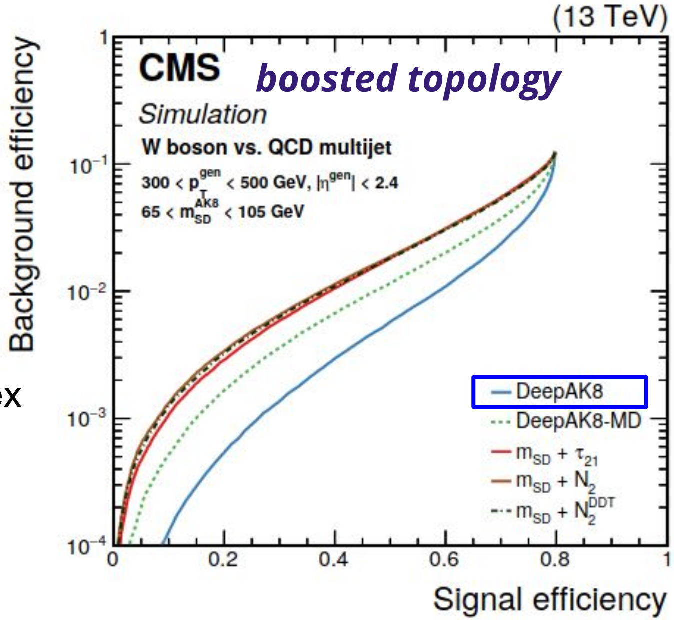


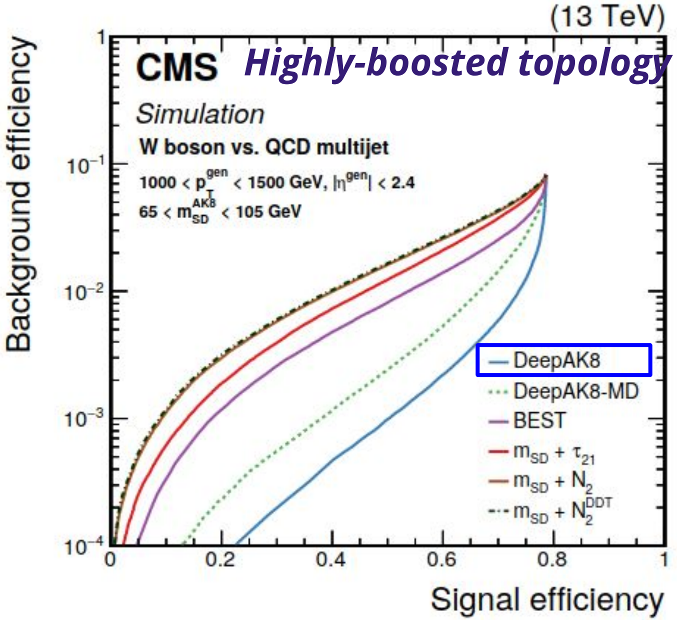
Boosted object tagging using machine learning algorithms

- Improved tagging performance with neural networks (NN) over cutoff methods
- DeepAK8 (convolutional NN) gives best performance for highly boosted topologies.

[arXiv:2004.08262](https://arxiv.org/abs/2004.08262)

DeepAK8:
100 particles
(4-momenta, charge,
track info, secondary vertex
info, angular separation
relative to jet axis, …),
42 variables total





Performance of taggers & scale factors is highly sensitive to modeling of quark- and gluon-initiated jet parton showers. Important to have as many experimental inputs to constrain these uncertainties!!