Jet fragmentation and substructure at the LHC

Andy Buckley, University of Glasgow
for the ATLAS & CMS collaborations

PANIC 2021, 8 September 2021

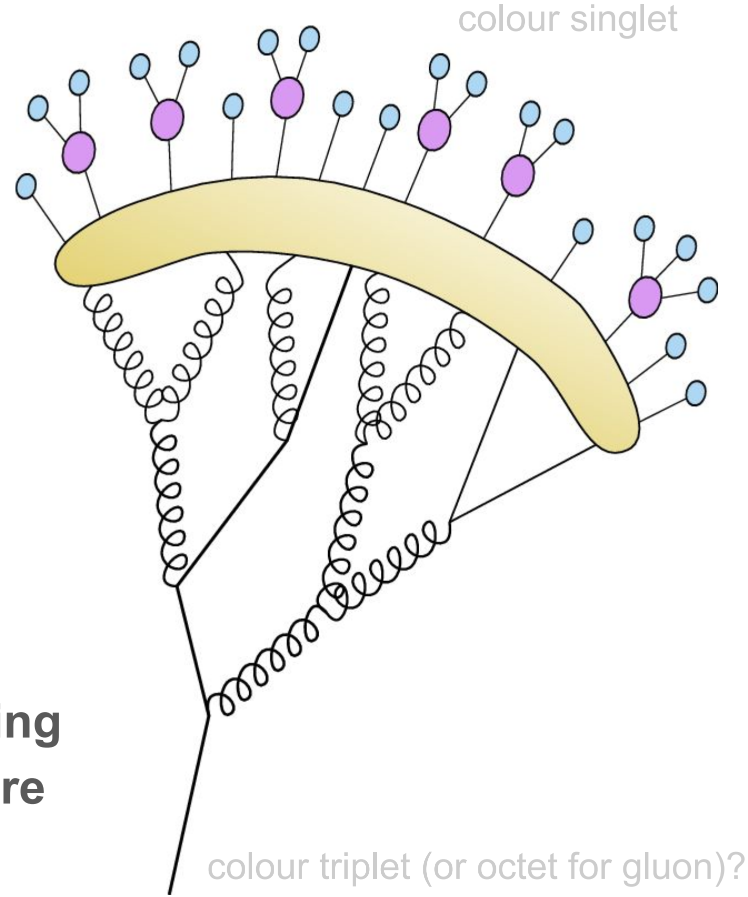
Jet fragmentation

In leading-order QCD, well-separated jets and partons are exactly equivalent

Broken by evolution from fixed-order to “real” jets, including both perturbative QCD radiation and non-perturbative hadronisation

Collectively this process can be considered as the *fragmentation* of a parton into the multi-hadron spray of a particle-level jet

Measuring jet fragmentation ⇒ understanding the emergence of jet structure from a mixture of parton flavours and configurations



Light-jet fragmentation

Earlier jet fragmentation measurements

Previous LHC measurements of jet fragmentation:

Eur. Phys. J. C 76 (2016) 322 — Measurement of the charged-particle multiplicity inside jets from √s = 8 TeV pp collisions with the ATLAS detector arXiv:1602.00988

Phys. Rev. D 93 (2016) 052003 — Measurement of jet charge in dijet events from √s=8 TeV pp collisions with the ATLAS detector, arXiv:1509.05190

Eur. Phys. J. C 71 (2011) 1795 — Measurement of the jet fragmentation function and transverse profile in proton-proton collisions at a center-of-mass energy of 7 TeV with the ATLAS detector, arXiv:1109.5816

Phys.Rev.D 83 (2011) 052003 — Study of jet shapes in inclusive jet production at 7 TeV with the ATLAS detector, arXiv:1101.0070

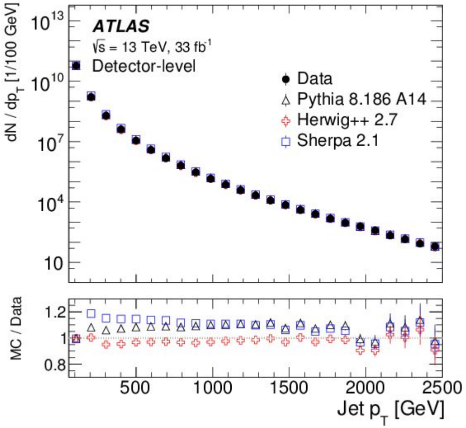
ATLAS jet fragmentation at 13 TeV — arXiv:1906.09254

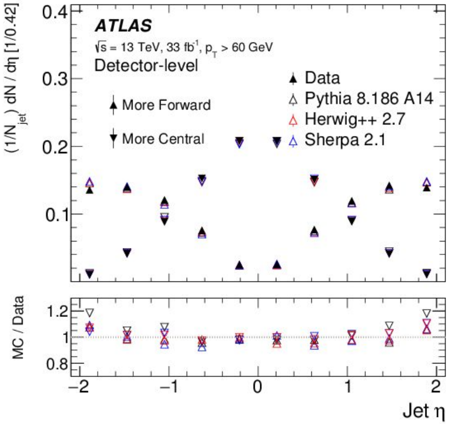
Uses 33 fb^{-1} dataset of 13 TeV pp collisions from 2016

- Increased phase space & jet p_T reach wrt 7, 8 TeV
- Makes use of Run 2 tracker upgrades, e.g. IBL
- Dense-environment tracking, for $\langle\mu\rangle \approx 25$

At least two reco jets with $|\eta| < 2.1$, and $p_T > 60$ GeV

- $|\eta|$ requirement for full containment in tracker
- $p_{T1}/p_{T2} < 1.5$ balance to simplify interpretation
- $p_T > 100$ GeV at fiducial level
- Charged tracks ghost-associated to calo jets



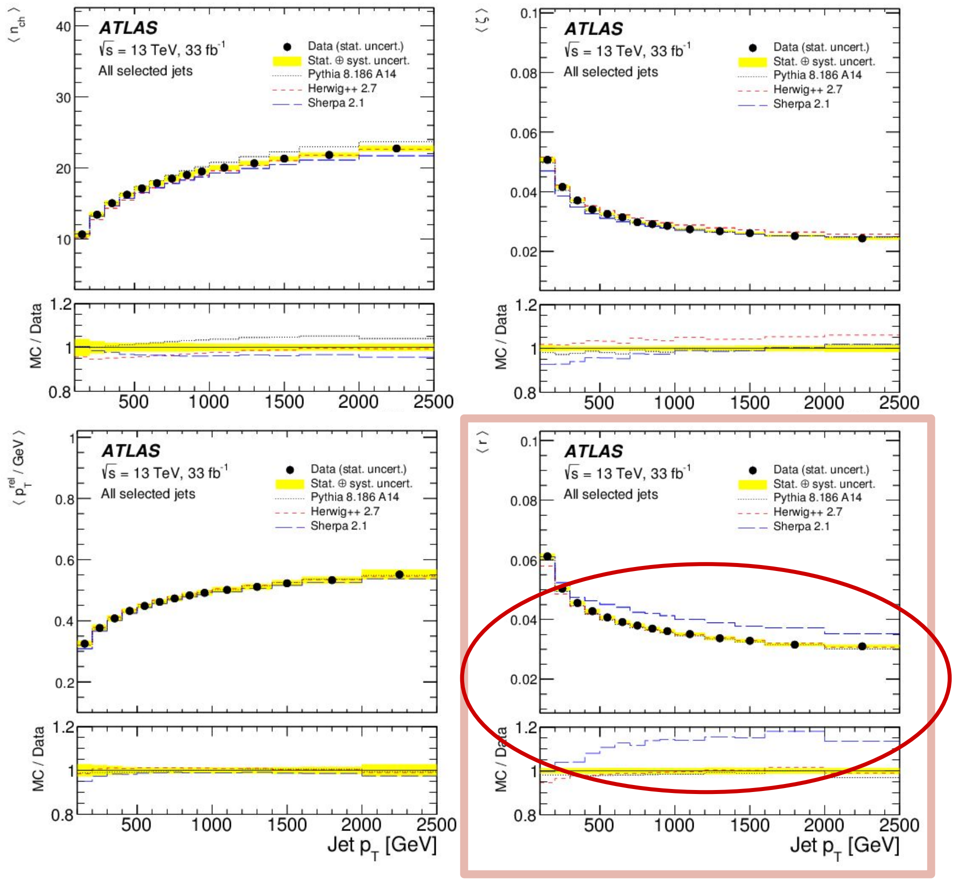


Unfolded average observables

Average observables vs p_T generally well-described by main shower MC codes (Pythia8.1xx, Herwig++ and Sherpa 2.1)

Hints of deviation from Sherpa, particularly in radial profiles — standard component of MC via 7 TeV jet shapes … for jet $p_T < 500$ GeV!

Need to check vs Sherpa 2.2.x



Unfolded observable moments & weighted sums

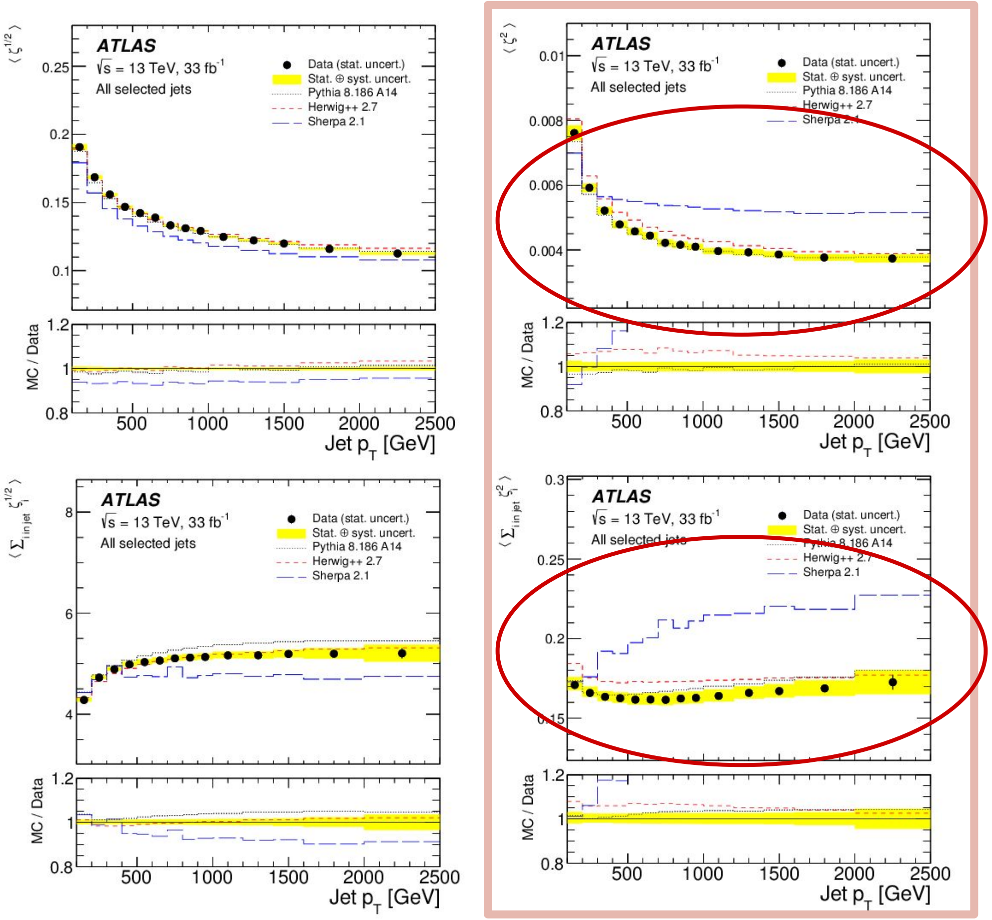
Also observables computed as moments and weighted sums with the p_T fraction ζ raised to powers $\kappa = 0.5$ and $\kappa = 2$:

$$\langle \zeta^\kappa \rangle = \int \zeta^\kappa F(\zeta) d\zeta \Big/ \int F(\zeta) d\zeta$$

$$\langle \textstyle\sum_{i \in \text{jet}} \zeta_i^\kappa \rangle = \int \zeta^\kappa F(\zeta) d\zeta$$

Pythia 8 and Herwig++ mostly well-behaved; major discrepancies seen for Sherpa 2.1, esp. for $\kappa = 2$

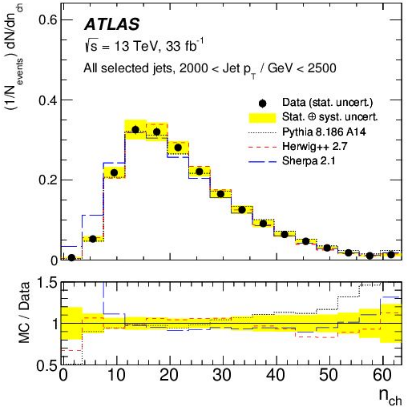
[effectively a var(ζ) measurement]

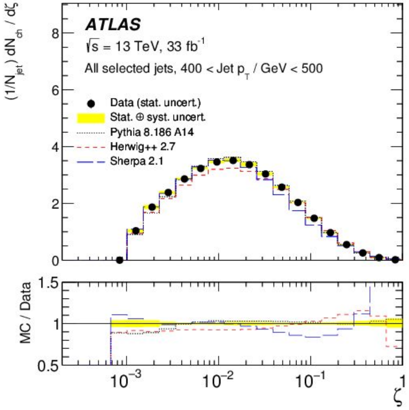


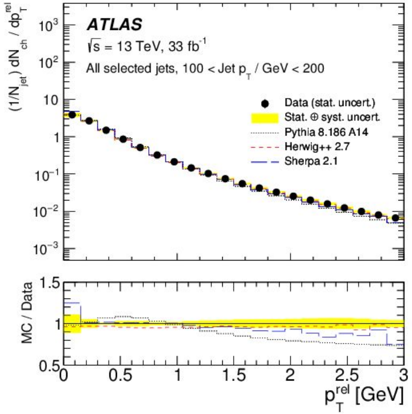
And more!

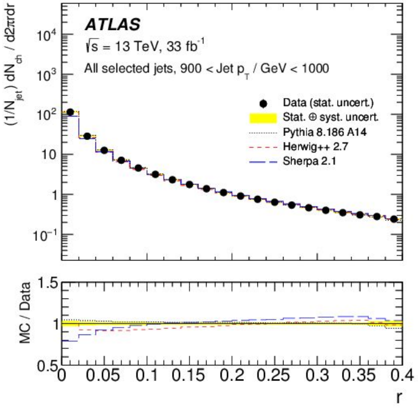
Differential distributions of every core variable in bins of jet p_T

A treasure-trove of data for jet modelling & resummation studies! **Rivet soon**









...

CMS quark & gluon jet substructure

CMS-PAS-SMP-20-010

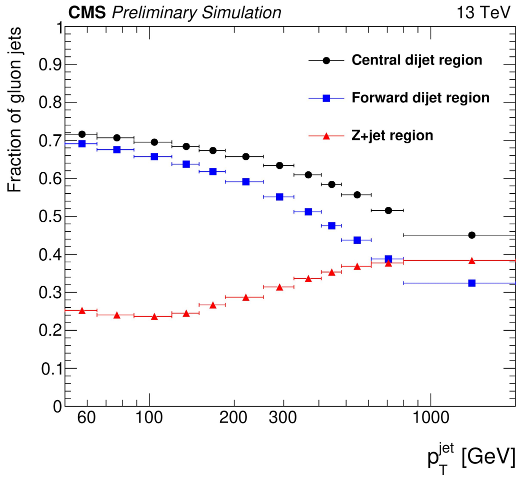
36 fb^{-1}, 13 TeV *pp* dataset, using akT4 & 8 jets with $p_T > 30$ GeV, $|y| < 1.7$ (cf. tracker acceptance)

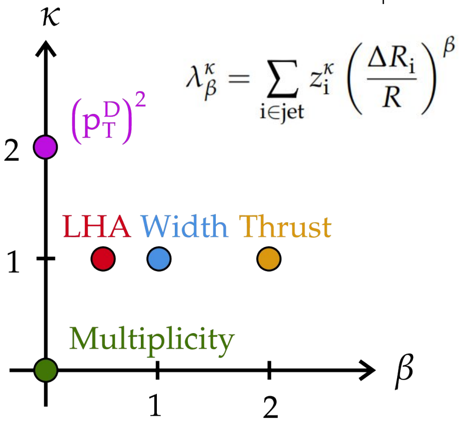
Dual event selection: dijets and *Z*+jets (two OS muons, $p_T > 26$ GeV, $|m_{\mu\mu} - m_Z| < 20$ GeV)

***JJ* and *ZJ* asymmetry and separation required**

Compare to LO event simulations with MG5+Py8 (+ ≤ 4 partons) and H++. Higher multiplicity MG5+Py8 unsurprisingly describes p_T spectra better ⇒ used as nominal MC

5 x generalised angularity observables cf. diagram:





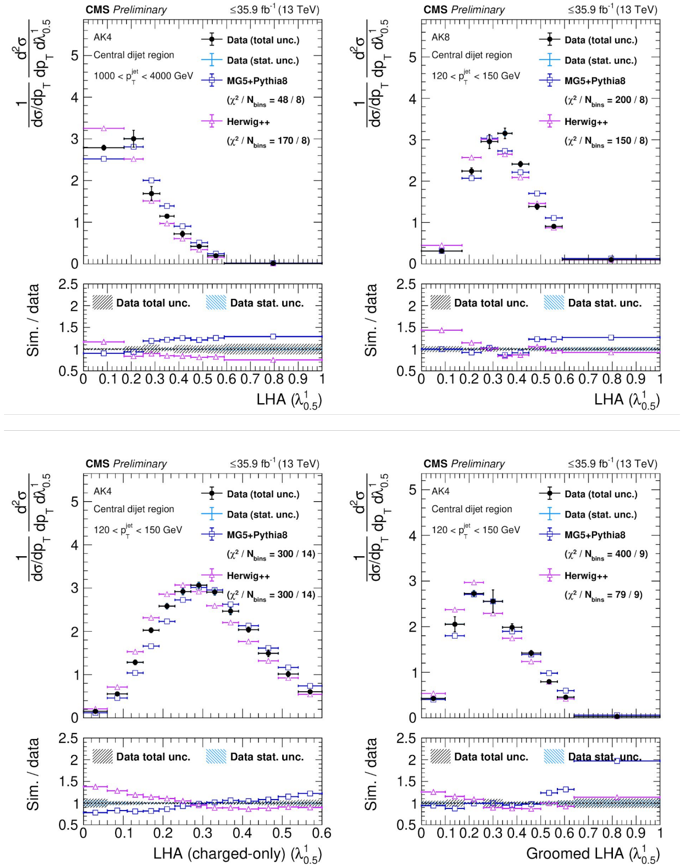
$$\lambda_\beta^\kappa = \sum_{i \in \mathrm{jet}} z_i^\kappa \left(\frac{\Delta R_i}{R} \right)^\beta$$

CMS quark & gluon jet substructure

CMS-PAS-SMP-20-010

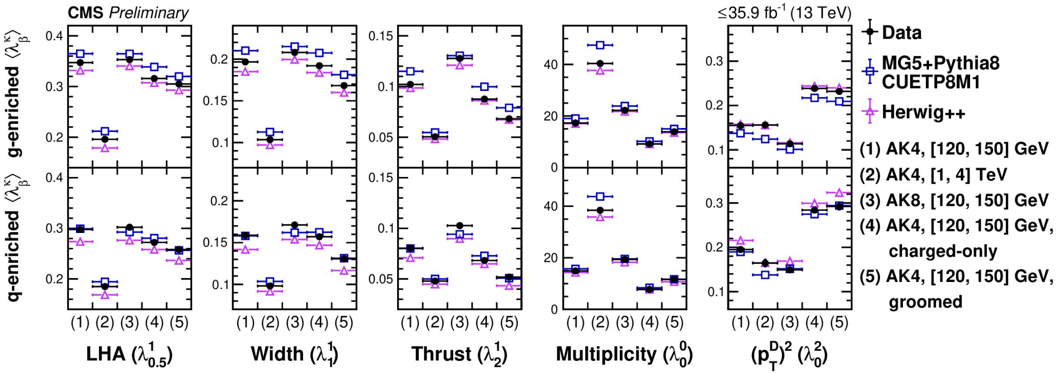
Observables distinguished by:

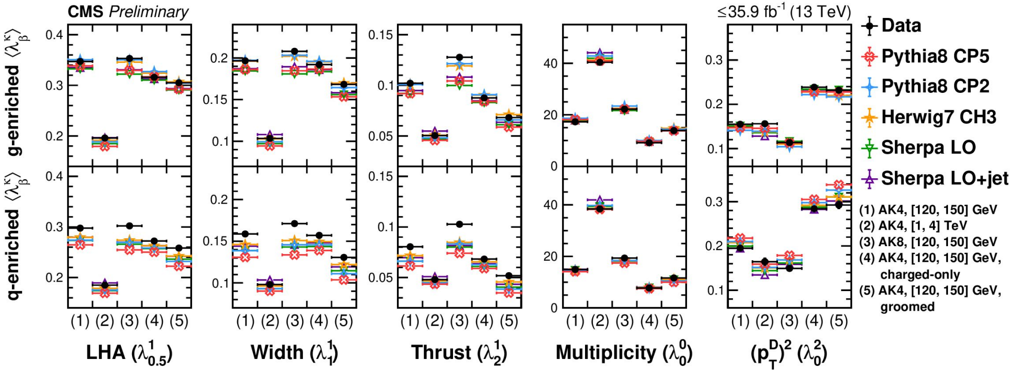
- event-selection
- central/forward dijet-event jet
- p_T-bin, and
- reco method: particle-flow, charged-only, and groomed



CMS quark & gluon jet substructure

CMS-PAS-SMP-20-010





Selections identified with *q*/*g* enhancement:

- “gluon” = central jet in dijet events
- “quark” = *Z*+jet at low-p_T (regions 1,3-5), and forward jet of dijet system at high p_T

Mean rates: MCs follow data trends, some large spreads. *q* worse? ⇒ tuning

b-jet fragmentation

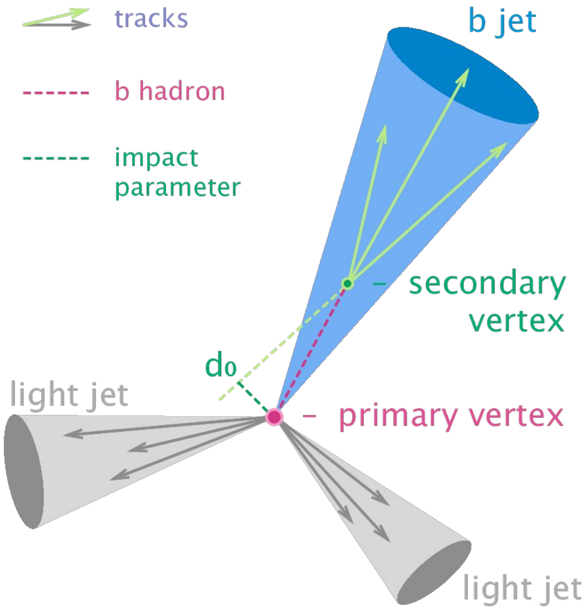
b-jet fragmentation

Heavy quark/jet production crucial both for QCD and EW/BSM physics

EW/BSM: *b*-jet signatures ubiquitous in many BSM models, top-quarks, and SM & BSM *H*→*bb* channels

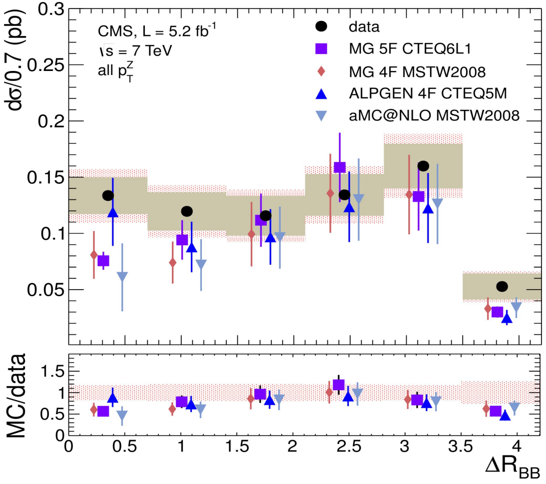
QCD: sensitivities to, for example…

- hard-scatter formalism between PDF / ME
- PS scale choices and mass effect on radiation pattern
- *b*-hadron production fractions



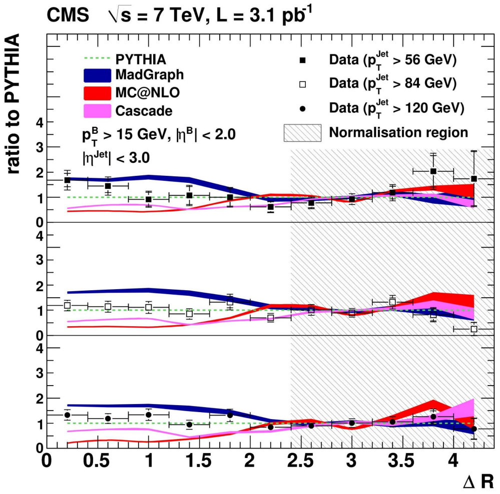
Earlier *b*-tagged jet fragmentation measurements

Previous measurements of *b*-jet fragmentation:



JHEP 12 (2013) 039 — CMS angular correlations in *Z*-boson with *b*-hadrons, arXiv:1310.1349

JHEP 03 (2011) 136 — CMS *BB* angular correlations from SV reconstruction, arXiv:1102.3194



ATLAS-CONF-2020-050

ATLAS *b*-fragmentation in *tt*

b-jet fragmentation moments, via secondary vertices and track-jets

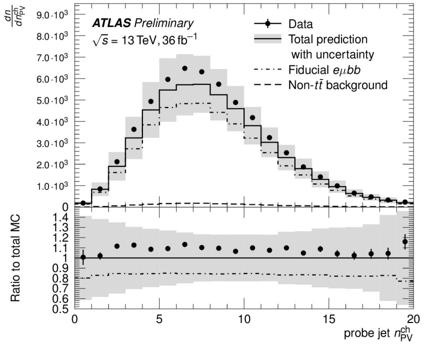
13 TeV analysis, partial Run 2 dataset of 36/fb

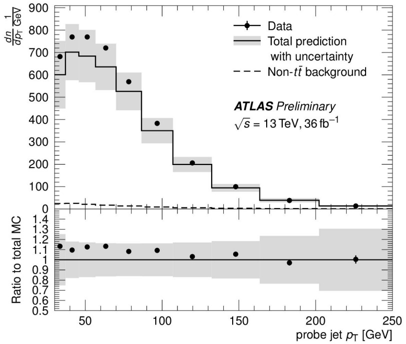
- R = 0.4 calo jets with standard ATLAS *b*-tagging (70% eff)
- Event selection: = 2 OS *e*+*μ*, = 2 jets, ΔR_{jj} > 0.5; **tag & probe, *both ways***
- **Variable-radius track jets ghost-associated to calo jets**
- **Track-jet PV and SV tracks: SV/(PV+SV) ~ *b*-hadron/*b*-quark**

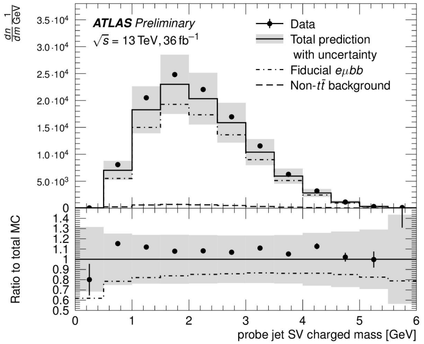
$$z_{\mathrm{T},b}^{\mathrm{ch}} = \frac{p_{\mathrm{T},b}^{\mathrm{ch}}}{p_{\mathrm{T,jet}}^{\mathrm{ch}}}$$

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

$$\rho = \frac{2p_{\mathrm{T},b}^{\mathrm{ch}}}{p_{\mathrm{T}}^{e} + p_{\mathrm{T}}^{\mu}}$$





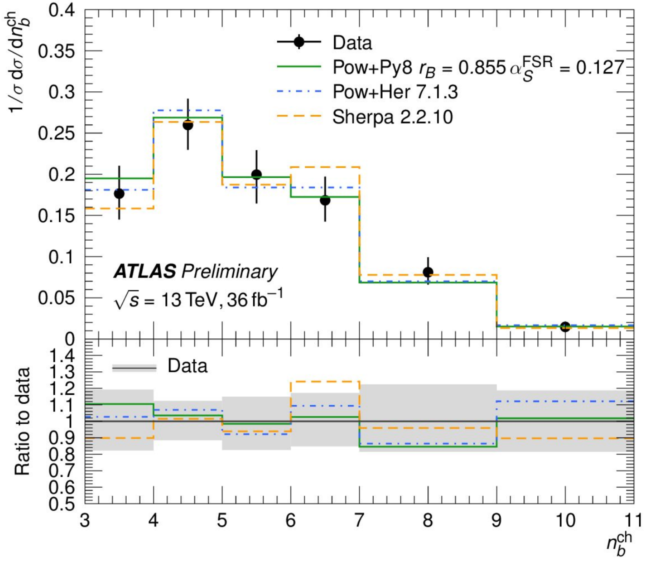


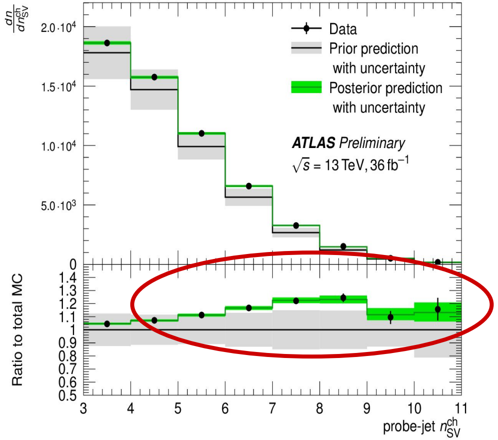
ATLAS-CONF-2020-050

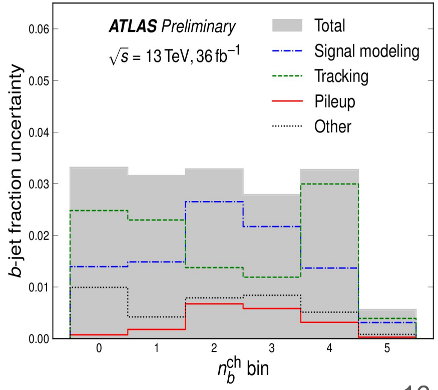
ATLAS *b*-fragmentation in *tt*

n_{ch} moment

- #SV tracks ≈ *b*-hadron
- Fit shifts away from the input-MC prior
- Reweightings of Pow+Py8 to probe-jet n_{SV} as systematic
- Check of hadron fractions & decays: Sherpa > 2.2.8 improvements





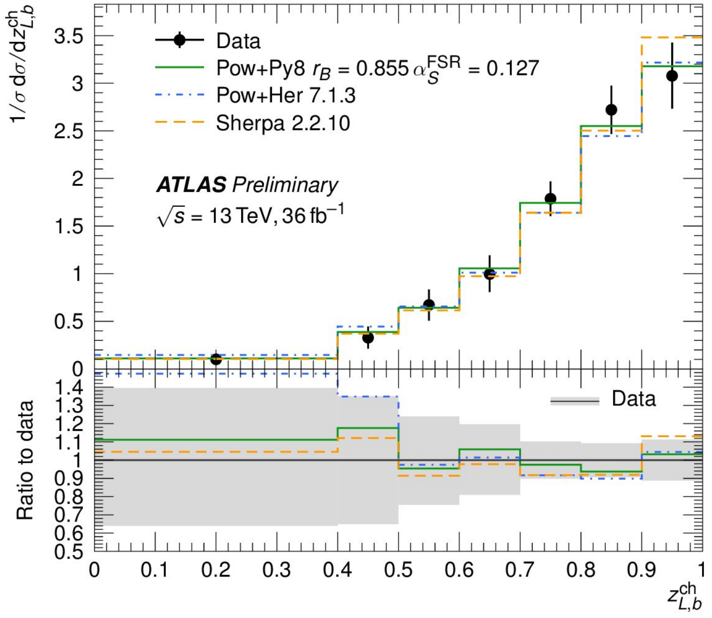


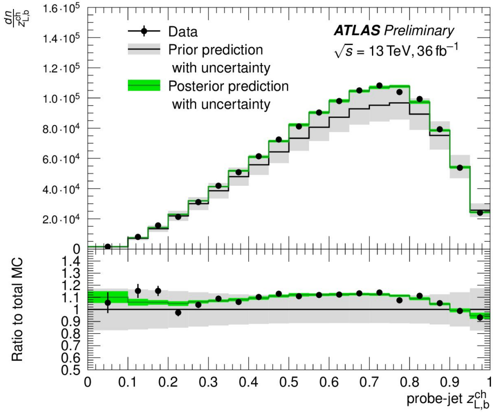
ATLAS-CONF-2020-050

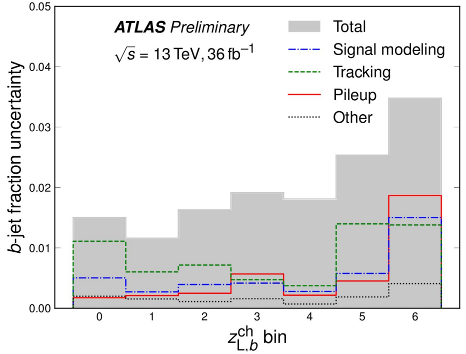
ATLAS *b*-fragmentation in *tt*

$z_{ch,L}$ (longitudinal) moment

- Experimental proxy for the theoretical fragmentation function: tunings from LEP and SLD valid?
- Descriptions generally good, with Herwig a slight outsider (much improved from v7.0.4)



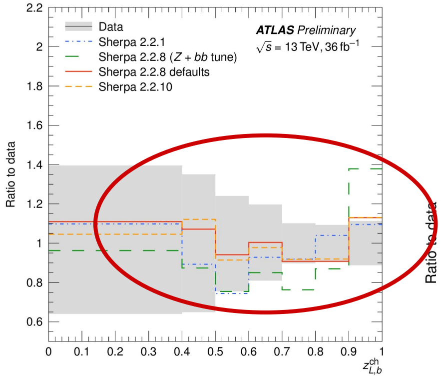


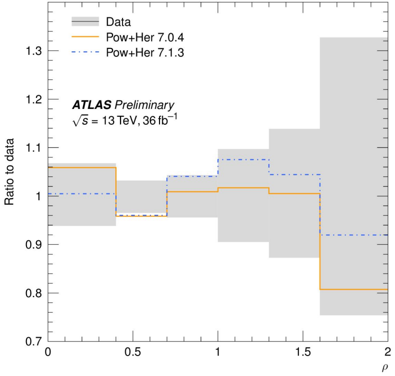


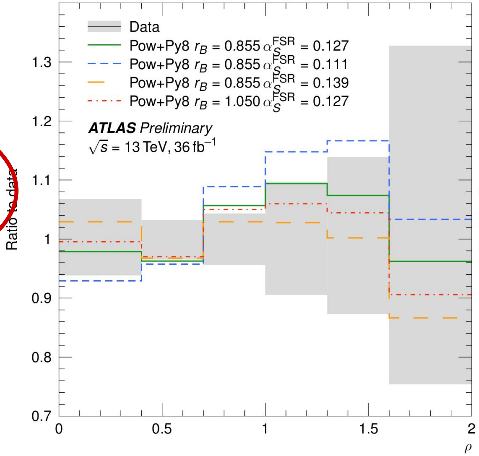
ATLAS-CONF-2020-050

ATLAS *b*-fragmentation in *tt*

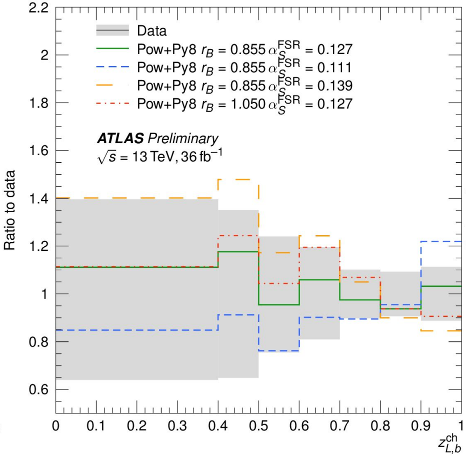
Sensitivities to Sherpa tunes, Pw+Hw versions, Pw+Py8 r_B tunes, α_S variations:

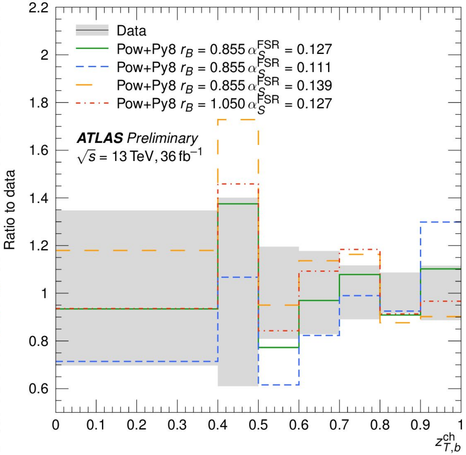






z variables: little sensitivity to r_B param





- Latest Sherpa models b-frag better
- Smooth effects of FSR α_S and frag function parameters on Py8 data/MC agreement: perfect tuning inputs

arXiv:2108.11650

ATLAS *b*-fragmentation with mesons

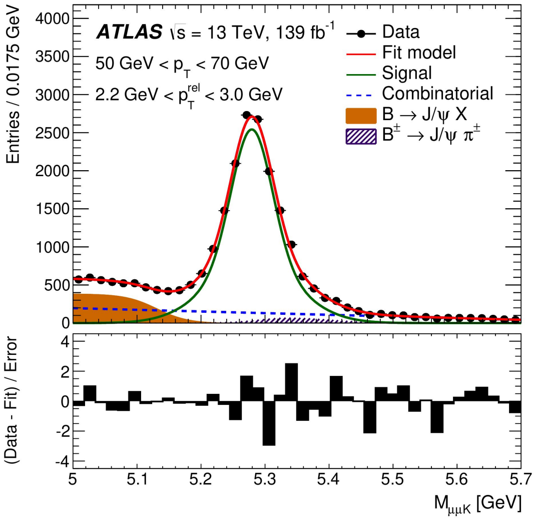
Explicit reconstruction of *$B^\pm \rightarrow J/\psi\ K$* decay channel

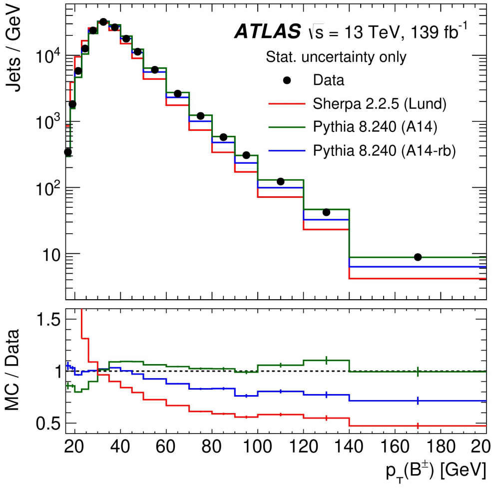
- Two muons: $p_T > 6$ GeV, $2 < m_{\mu\mu} < 9$ GeV
- Third track, cuts on vtx χ^2 & pair/triplet masses

Unfold to particle-level with kin cuts on μ and K p_T

Measure fragmentation functions:

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

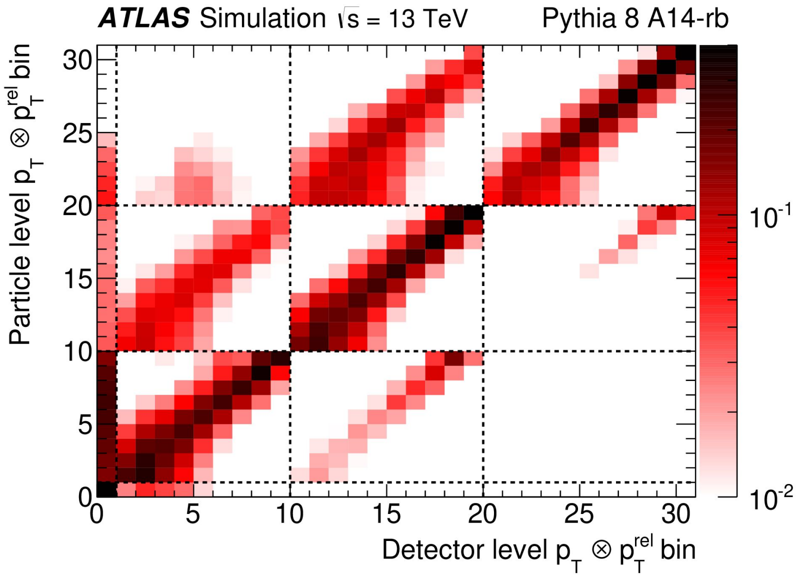


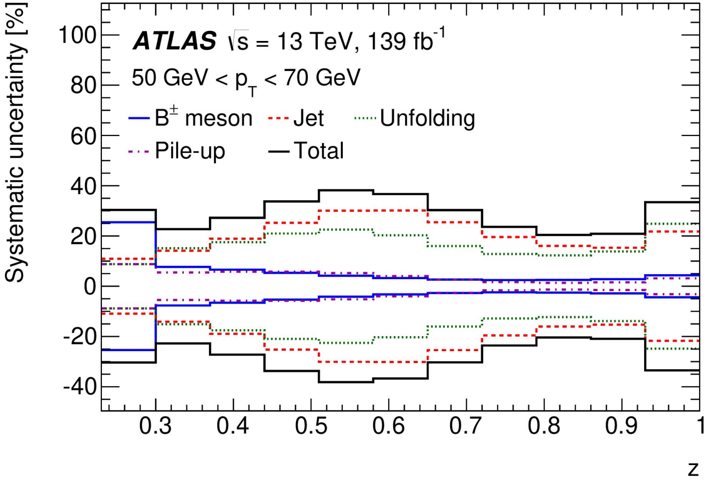


ATLAS *b*-fragmentation with mesons

arXiv:2108.11650

Unfolding more limited by transfer resolution for $p_{T,\mathrm{rel}}$ than z





Limiting systematics a balance btw *B* reco, jet reco, and unfolding uncertainties

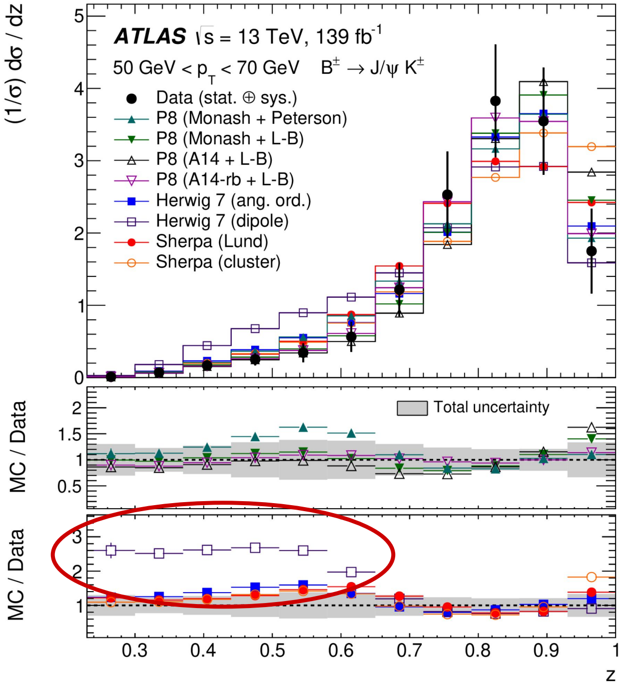
Obs and p_T dependence

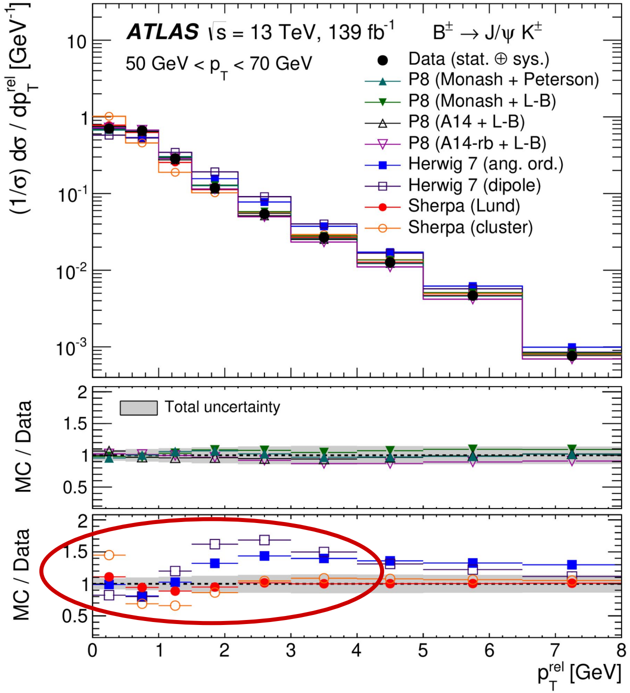
ATLAS *b*-fragmentation with mesons

arXiv:2108.11650

Longitudinal and transverse frag functions in jet-p_T bins

⇒ Herwig7 showers and Sherpa cluster hadronisation show deviations wrt data

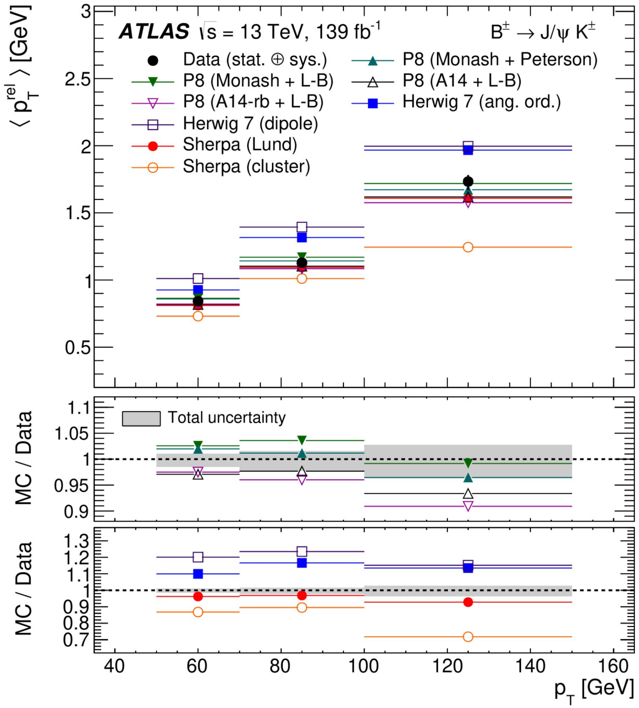


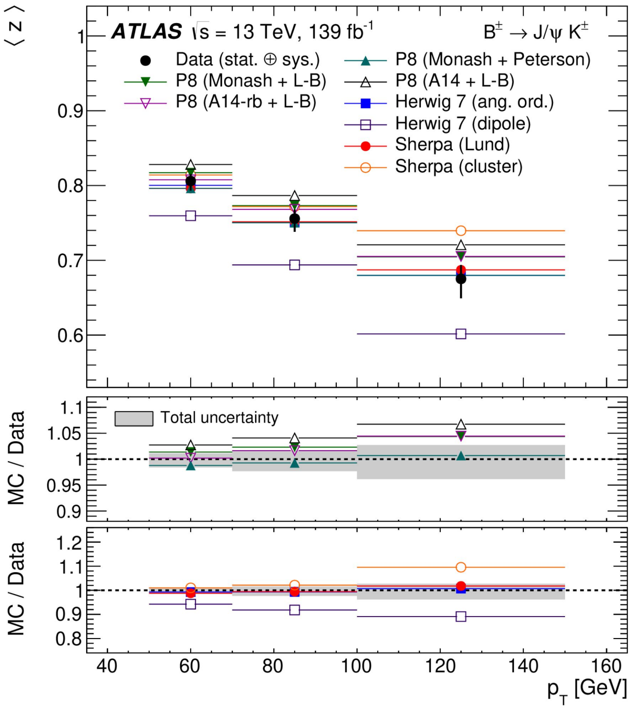


arXiv:2108.11650

ATLAS *b*-fragmentation with mesons

Average values of frag functions vs jet p_T ⇒ flags O(10%) mismodelling





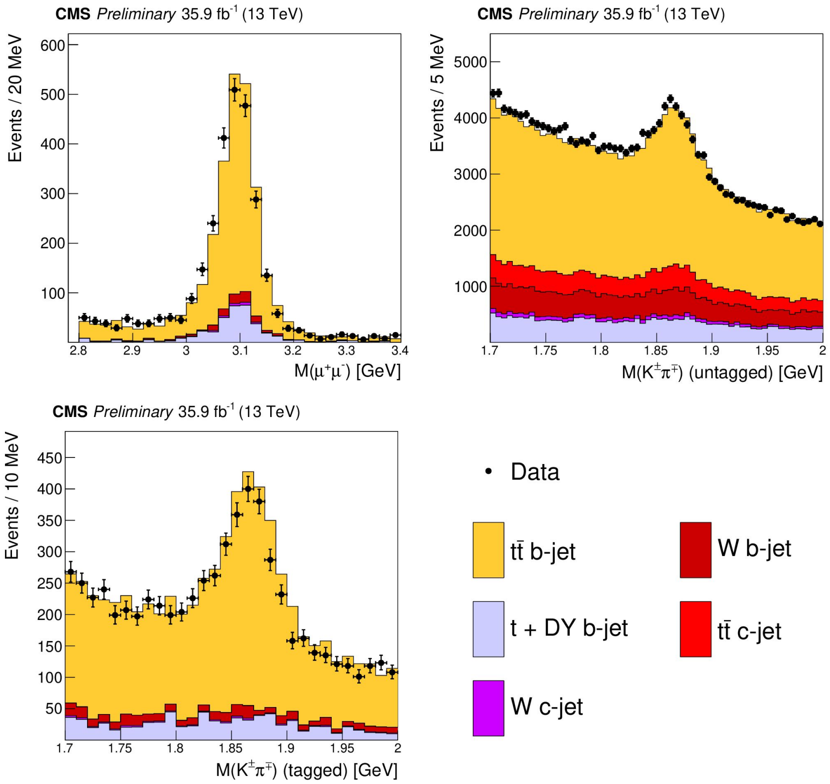
CMS *b*-jet fragmentation from *tt* decays

CMS-PAS-TOP-18-012

CMS' version is more directly tied to the Lund-Bowler fragmentation function ansatz and fitting of its r_b parameter

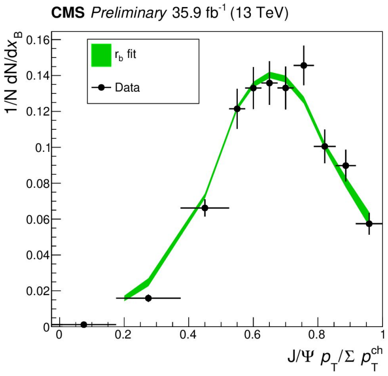
Explicit reconstruction of J/ψ and D^0 mesons (the latter with and without a muon-tagged decay)

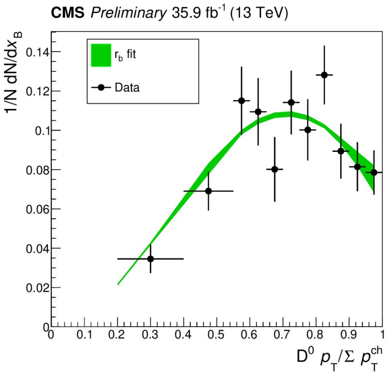
Use mesons as final-state proxies for the decayed *b*-hadron. Distributions of momentum fractions used to fit r_b

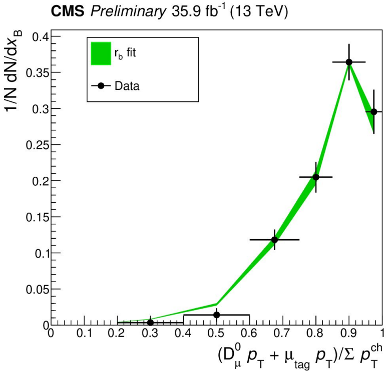


CMS *b*-jet fragmentation from *tt* decays

CMS-PAS-TOP-18-012

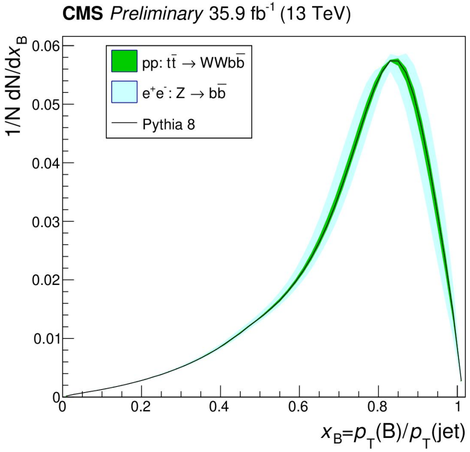


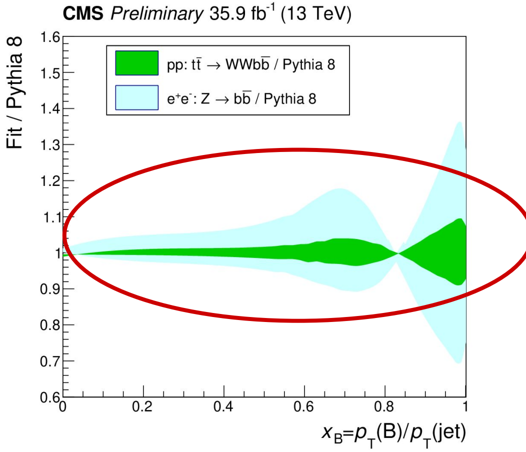




Set of three proxy fragmentation fractions extracted, used to constrain x_B

Significant precision improvements wrt LEP/SLD *Z*-pole tunes:



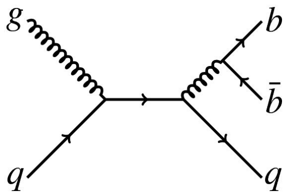


ATLAS *g*→*bb* at small opening angles

Phys. Rev. D 99 (2019) 052004,
arXiv:1812.09283,
Plot page

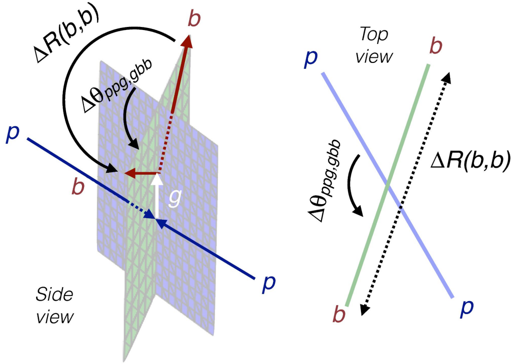
Focus on final-state *g*→*bb* splitting kinematics in boosted region

p_T > 450 GeV R=1 akT jets, 33/fb of 13 TeV *pp* data



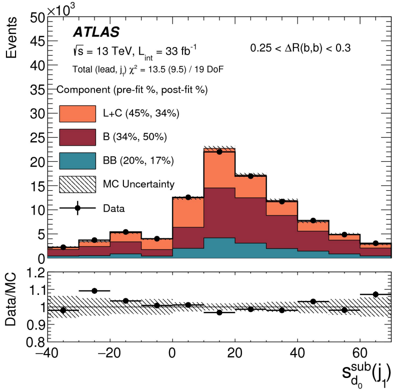
Observables:

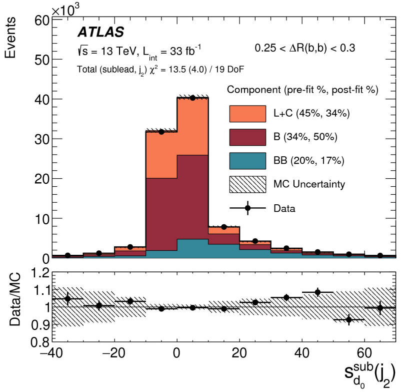
- m_{bb} / p_T
- $z(p_T)$
- ΔR_{bb}
- Polarisation angle Δθ ⇒



Require two *b*-tagged, ghost-assoc VR track-jets as *b* proxies, 60% working point

Flavour fit, via signed-impact-parameter distributions for subleading track, per observable bin, e.g. ΔR_{bb} right:
Mostly slight *BB*-fraction overestimates



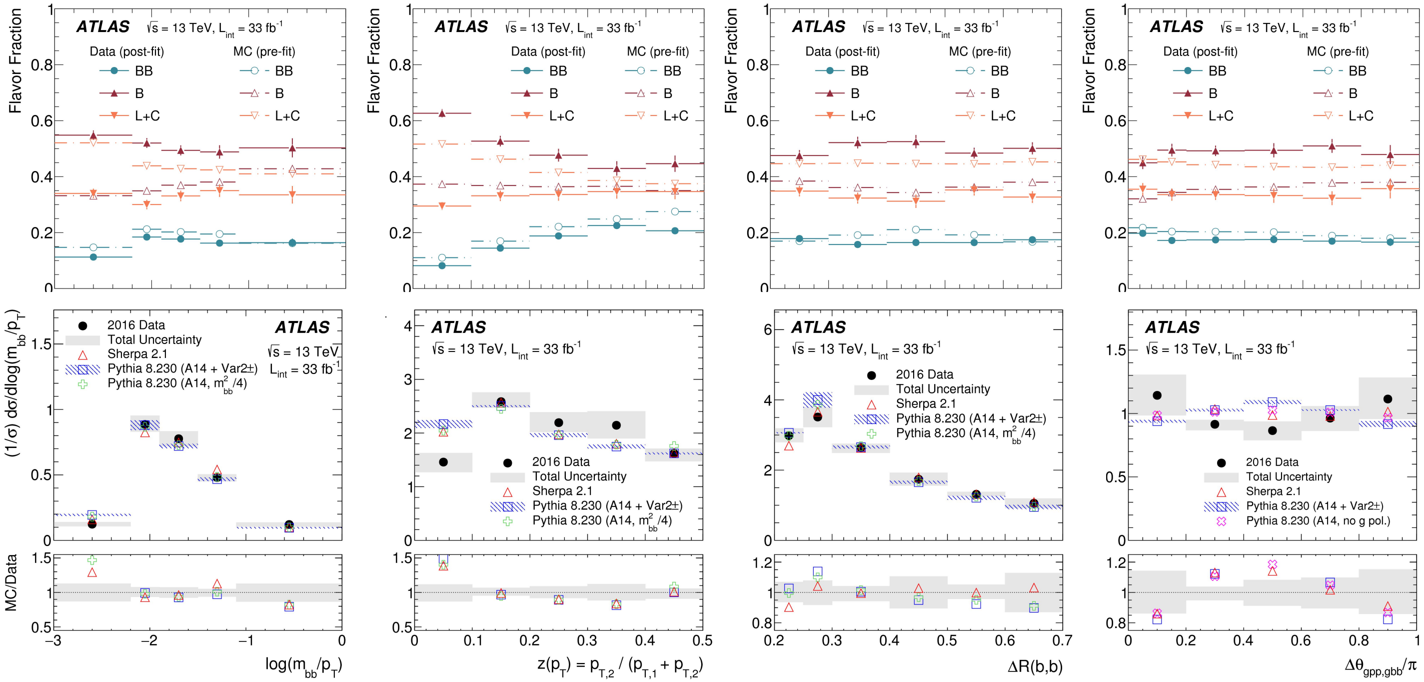


ATLAS *g→bb* at small opening angles

Phys. Rev. D 99 (2019) 052004,
arXiv:1812.09283,
Plot page

Flavour-fraction fits, per bin of each observable

(BB ~good, large differences in B and L+C)



Jet fragmentation in tt

CMS jet substructure in *tt* at 13 TeV

arXiv:1808.07340 / Phys. Rev. D 98 (2018) 092014

“one to rule them all...”

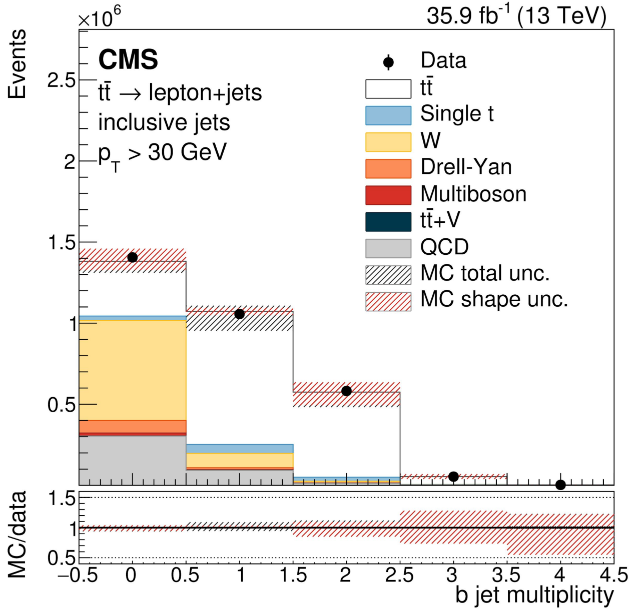
36 fb^{-1} of 13 TeV *pp* dataset, with semileptonic top-pair selection

Exactly 1 tight, isolated lepton; 63% *b*-tag working point ⇒ 94%-pure *tt*

akT4 jets, p_T > 30 GeV, flavour groups:

- inclusive
- *b*-jet: via tag, incl *g*→*bb*
- *q*-enh: via *W*-mass window, 50% pure
- *g*-enh: neither of the above, 58% pure

Particle-level flavour ⇒ ghost matching & m_{jj}



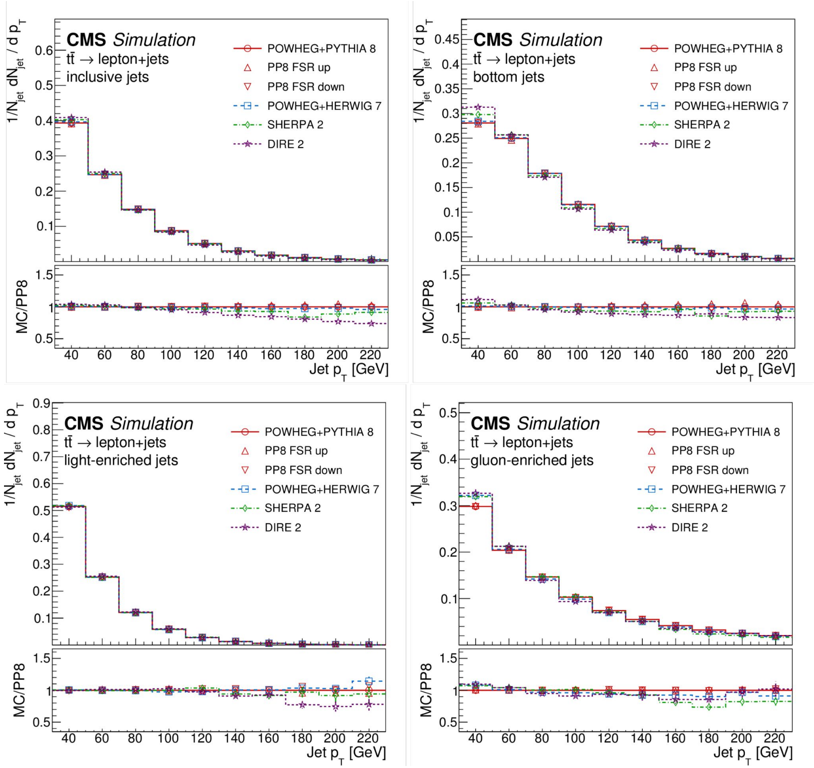
CMS jet substructure in *tt* at 13 TeV

arXiv:1808.07340 / Phys. Rev. D 98 (2018) 092014

***Lots* of substructure observables:**

- constituent-multiplicity & p_T-dispersion
- angularities: width, LHA, jet thrust
- eccentricity
- soft-drop fraction and groomed N_{ch}
- groomed-subjet angles
- N-subjettiness ratios
- ECFs $C_{1,2,3}$, M_2, $N_{2,3}$

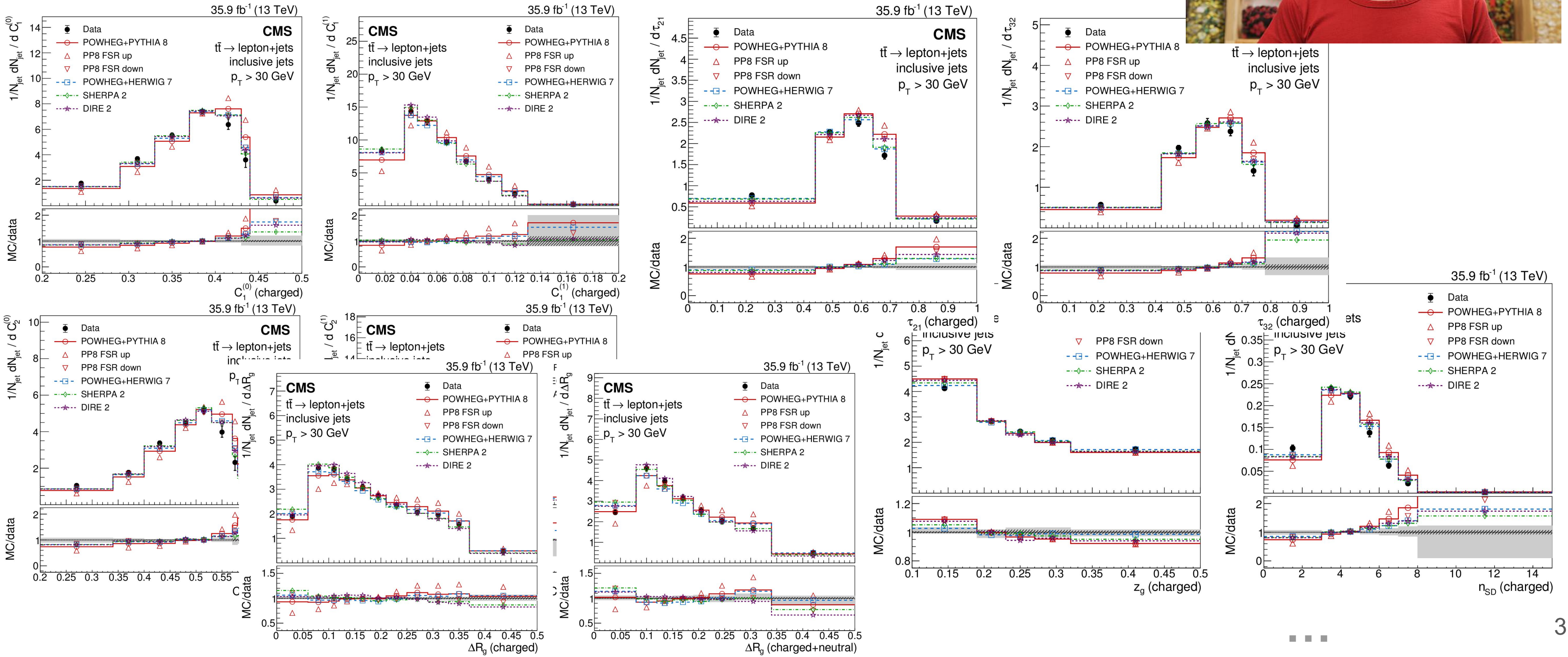
Several observables measured for both charged and charged+neutral constituents!



CMS jet substructure in *tt* at 13 TeV

arXiv:1808.07340 / Phys. Rev. D 98 (2018) 092014

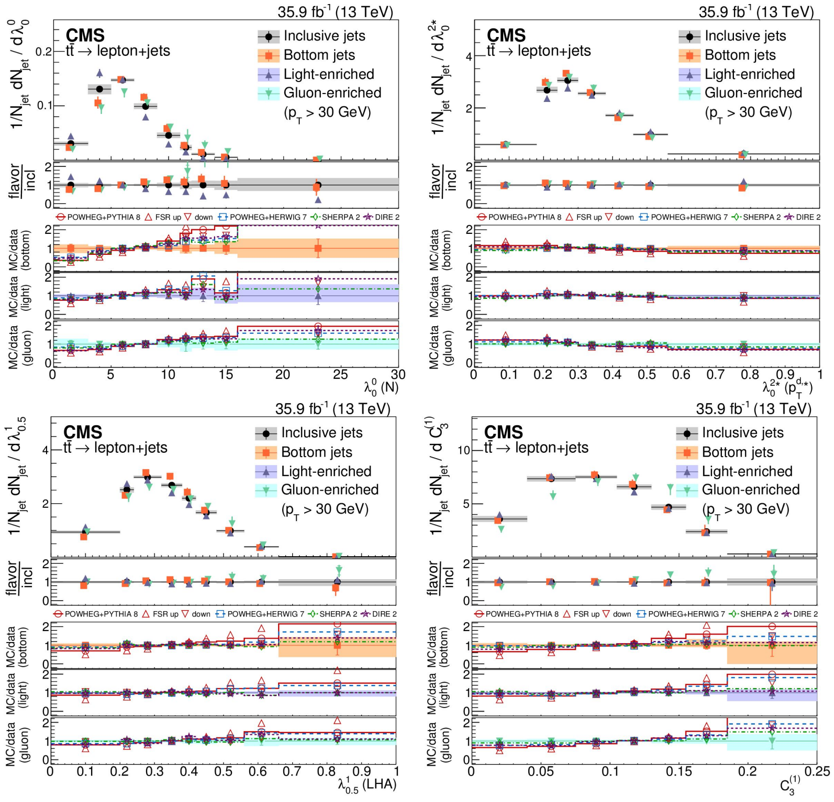
Exceptional resource for model development & validation!

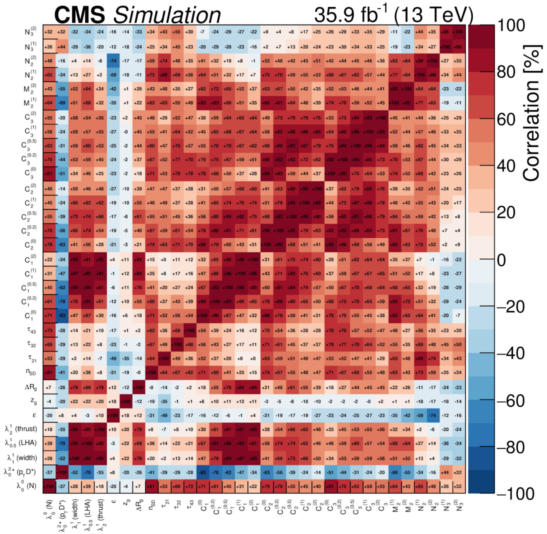


...

CMS jet substructure in *tt* at 13 TeV

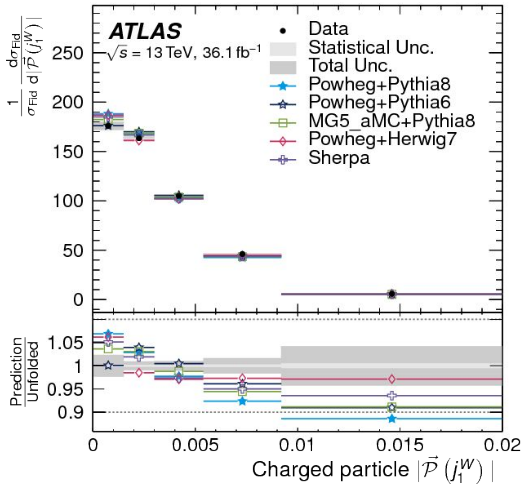
arXiv:1808.07340 / Phys. Rev. D 98 (2018) 092014





Breakdowns of observables by flavour category, plus correlations: will be ideal for MC dev & tuning via Rivet

Other jet structure measurements



Top-decay

Eur. Phys. J. C 78 (2018) 847 — ATLAS 13 TeV measurement of colour flow between *W* → *qq’* jets in *tt* with jet-pull observables arXiv:1805.02935

Boosted large-*R* jets

JHEP 08 (2019) 033 — ATLAS 13 TeV measurements of jet substructure in top-quark, *W*-boson, and light jets arXiv:1903.02942

JHEP 11 (2018) 113 — CMS 13 TeV jet mass distributions in dijet events, arXiv:1807.05974

Phys. Rev. D 86 (2012) 072006 — ATLAS 7 TeV properties of boosted jets arXiv:1206.5369

JHEP 1205 (2012) 128 — ATLAS 7 TeV jet mass and substructure arXiv:1203.4606

Conclusions

- **Measurements sensitive to jet fragmentation have grown far beyond direct measurements cf. LEP**
  - A multitude of angularities, correlation functions, substructure observables from both ATLAS and CMS
  - Use of track-based & all-particle reconstructions, and dig into jet-flavour dependence
  - Ghost-association of track-jets or reconstructed hadrons, esp. for *b*-quark studies
  - Grooming algorithms reduce pile-up dependence, provide another facet
  - Also large-*R* jets and boosted phase-space
- **Ways to improve syst precision? Most analyses only 36/fb so far!!**
- **Dawn of Run 3 a good time to revisit 7 TeV MC tunes & studies**

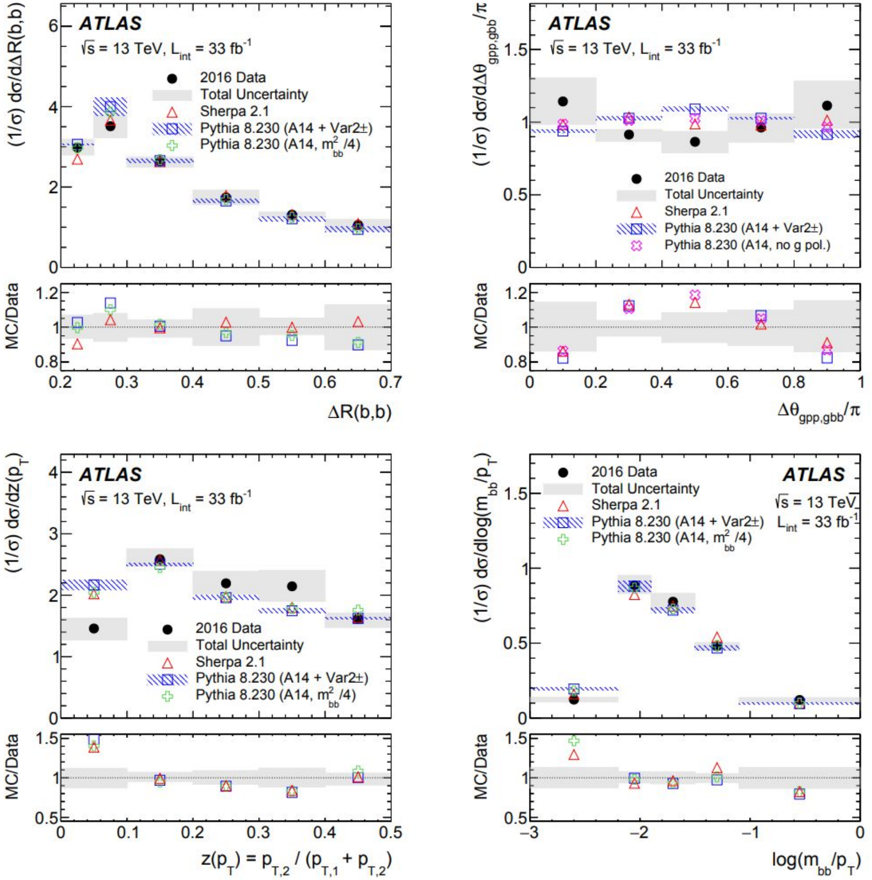
Backup

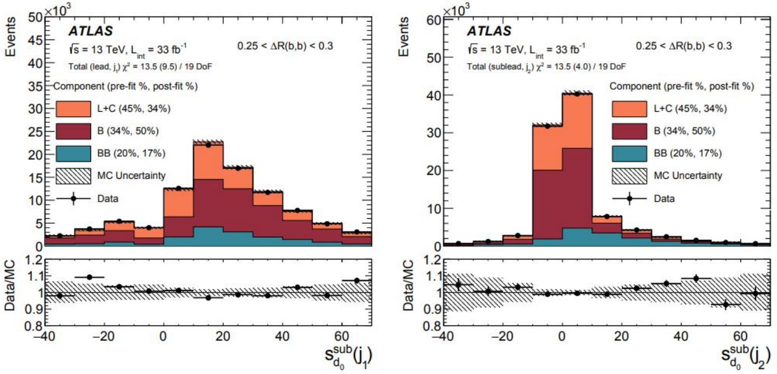
ATLAS g→bb fragmentation — arXiv:1812.09283

Super-quick summary: b-tagged track subjets in boosted jets

Fiducial differential cross-sections in b-subjet separation, mass, p_T balance, and polarisation angle

Key: flavour fit via signed impact param





Observables

Fragmentation function *D* defined vs $\zeta = p_T$ fraction of hadron *h* wrt its containing jet p_T, from parton *p*
⇒ DGLAP pQCD evolution; mirror image of PDFs

$$\mu\frac{\partial}{\partial\mu}D_p^h(\zeta,\mu) = \sum_{p'}\int_\zeta^1 \frac{d\zeta'}{\zeta'}\frac{\alpha_S(\mu)P_{p'\leftarrow p}(\zeta',\mu)}{\pi}D_{p'}^h\left(\frac{\zeta}{\zeta'},\mu\right)$$

This paper uses charged hadrons, but full (calo) jet
⇒ $\langle n_{ch}\rangle$ and differential $1/N_{jet}\, dN_{jet}/d\langle n_{ch}\rangle$

$$\langle n_{ch}\rangle(p_T^{jet}) = \sum_p f_p(p_T^{jet}) \sum_{h\text{ charged}} \int_{\text{threshold}/p_T^{jet}}^1 d\zeta\, D_p^h(\zeta, p_T^{jet})$$

- **summed fragmentation function:**
  differential in p_T fraction ζ *and* jet p_T ⇒ extract partial fractions, moments & weighted sums

$$F(\zeta, p_T^{jet}) = \sum_p f_p(p_T^{jet}) \sum_{h\text{ charged}} D_p^h(\zeta, p_T^{jet})$$

- Relative transverse momentum
  Radial profile (non-p_T-weighted)

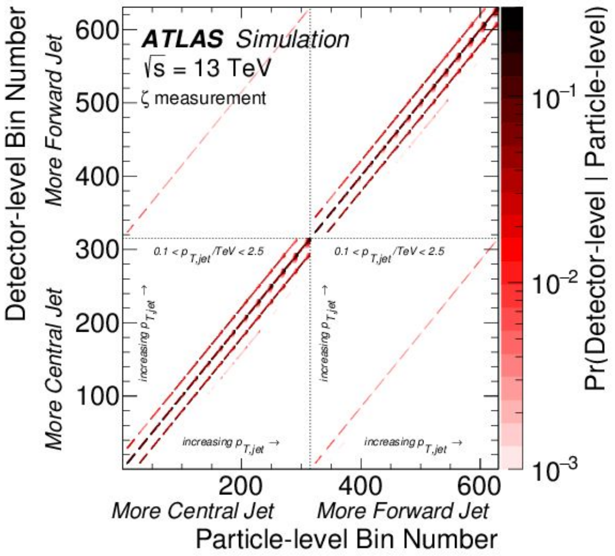
$$p_T^{rel} \equiv p_T^{\text{charged particle}} \sin\Delta\phi$$

$$\rho_{ch}(r, p_T^{jet}) = (1/N_{jet})dn_{ch}/2\pi r\,dr$$

Detector correction & uncertainties

Unfolding from detector obs to fiducial phase space:

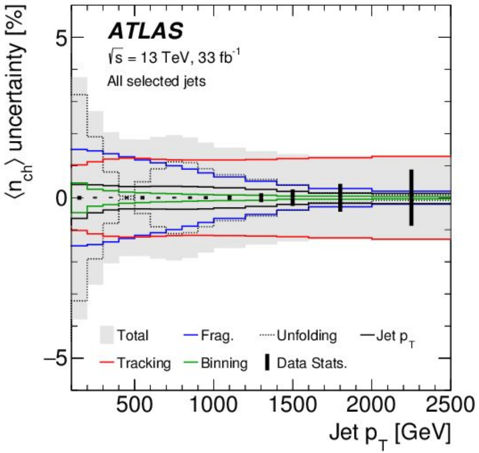
particle-level tracks & jets from particles with $c\tau_0 > 10$ mm; muons and neutrinos excluded from jets

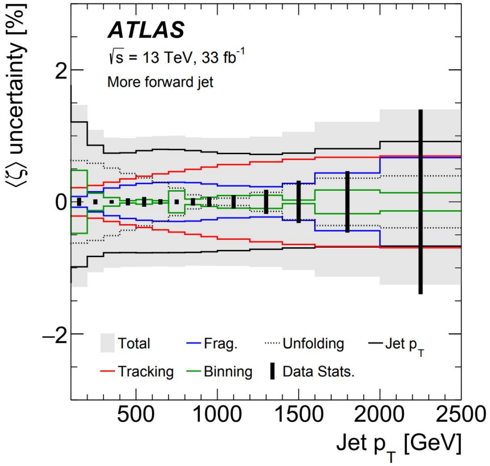


Unfolding by 2D iterative Bayes method (1 iter) sandwiched by explicit in/out migration corrs.

Main uncertainties:

tracking, jet scale, binning & unfolding, depending on observable



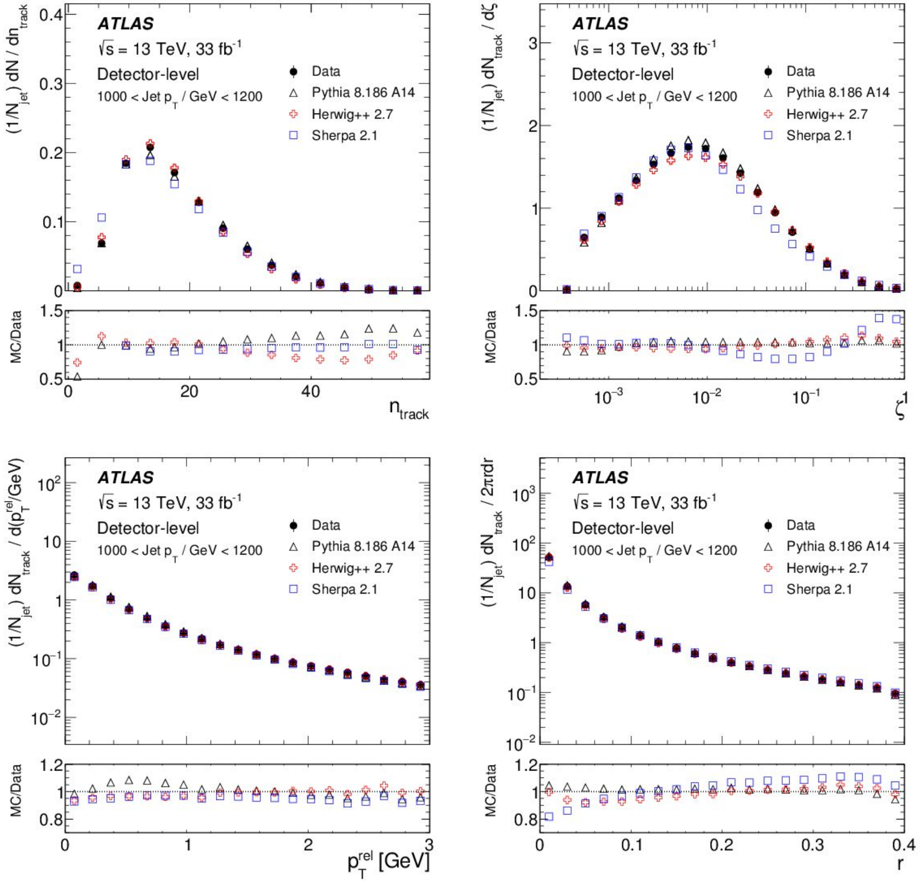


Detector-level variables

Raw distributions of n_{trk}, track momentum fraction, track $p_{T,rel}$, and track radial profile

For a 1 TeV jet, most probable n_{trk} is ~15, and most probable momentum fraction ~1%

Track $p_{T,rel}$ and r (radial profile) distributions peak at zero since radiation dominantly collinear



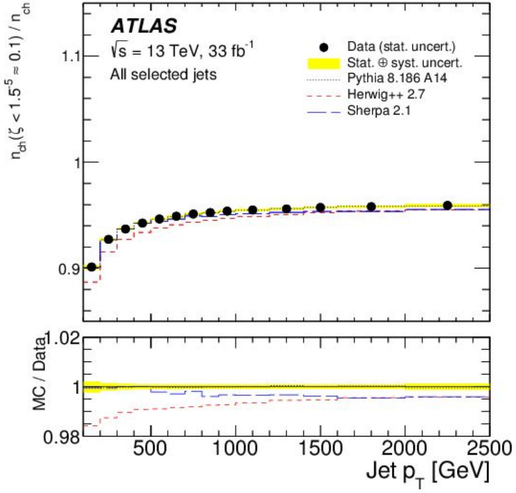
Unfolded partial sums: n_{ch} fraction in bins of ζ

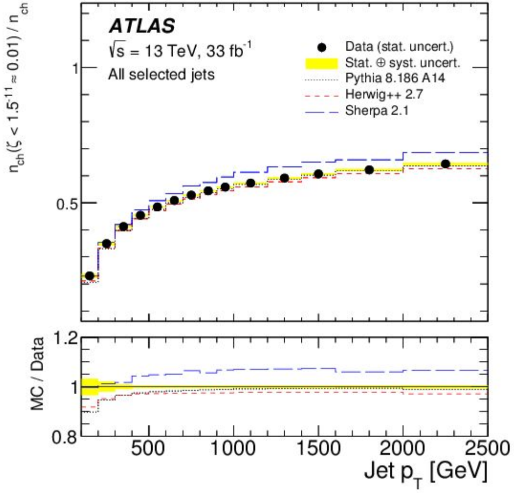
$$\int_0^X F(\zeta)d\zeta \,/ \int F(\zeta)d\zeta = n_{\mathrm{ch}}(\zeta < X)/n_{\mathrm{ch}}$$

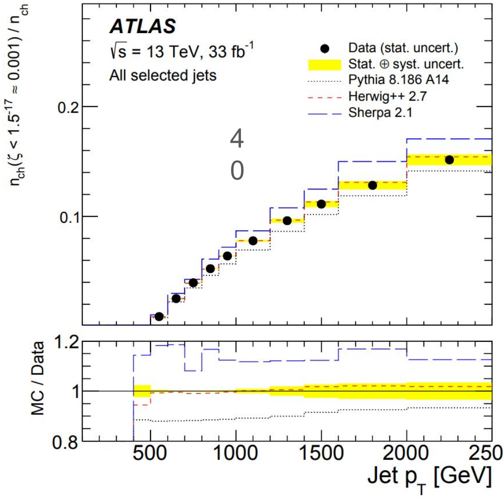
Fractions of charged particles with $\zeta \lesssim$ 10%, 1%, and 0.1% vs jet p_T

Fraction of small-fraction particles increases with jet p_T, cf. hadronisation scale

Small mismodelling of 10% by Herwig; with Sherpa & Py8 in less inclusive bins





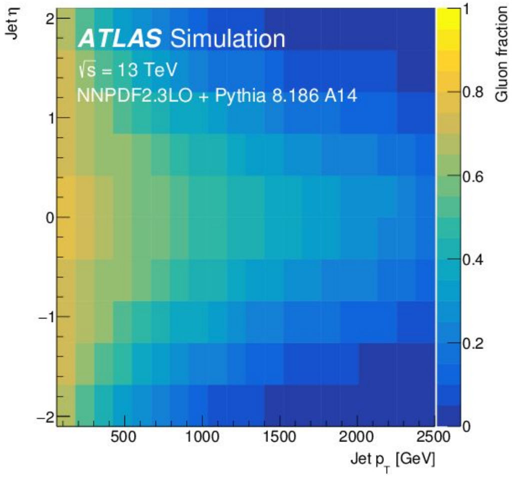


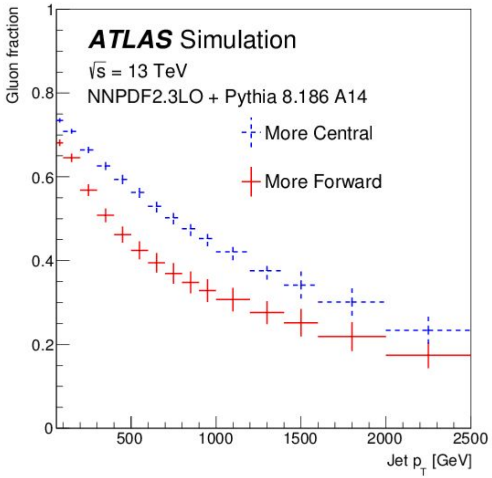
Quark/gluon jet discrimination

An important application of jet structure data is development of methods to extract information about quark/gluon jet origins

Ideally in a well-defined, QCD-aware way!

- **Central/forward jet:** roughly, central and low-p_T jets are more likely to be gluon-initiated
- ⇒ Extract q/g components with an MC-template procedure
- **New:** model-independent q/g extraction by data-driven “topic” modelling





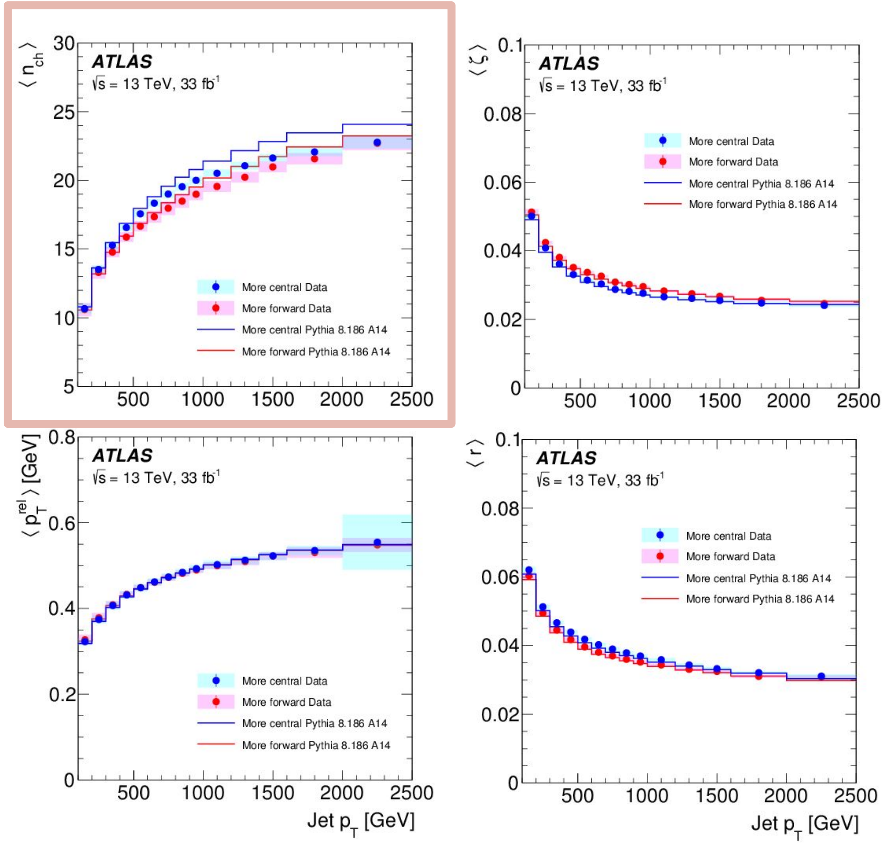
Mean observables with central/forward-jet split

Aim of central/forward jet distinction is to bias quark or gluon jet origin

Biases allow extraction of separate q/g-like fragmentation functions by comparison of forward and central jet ones

Note Pythia mismodelling of split n_{ch} distributions, unlike inclusive.

Most c/f-split mean observables are well-described



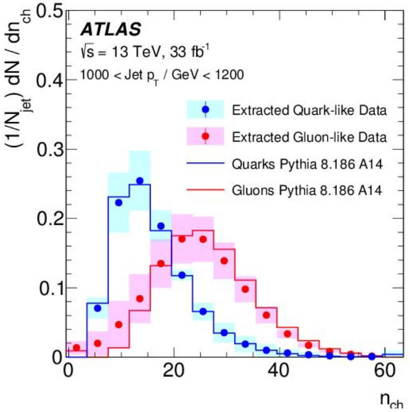
Model-dependent quark/gluon jet characterisation

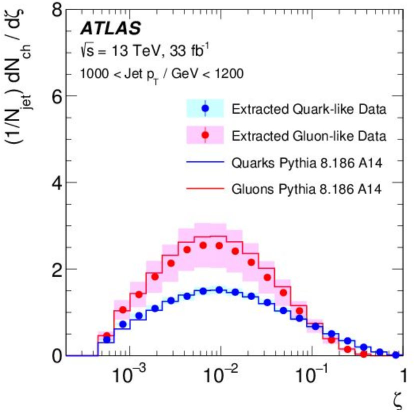
q/g extraction by use of MC flavour fractions *f*, nominally from Pythia:

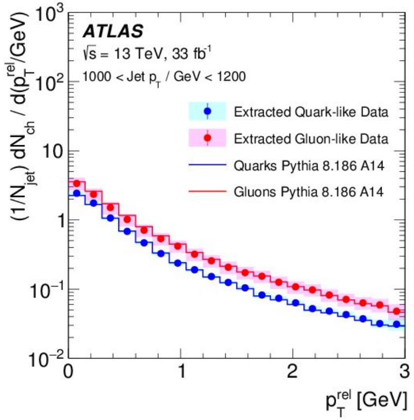
$$h_i^f = f_q^f h_i^q + (1 - f_q^f) h_i^g$$
$$h_i^c = f_q^c h_i^q + (1 - f_q^c) h_i^g$$

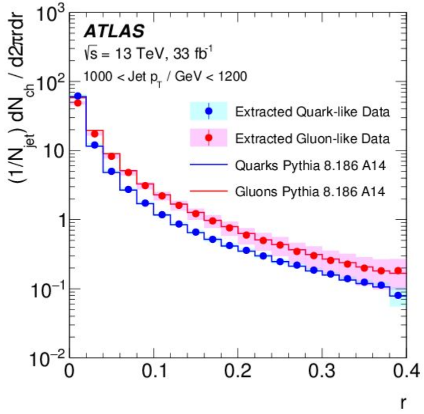
Jet flavour defined by hardest parton geometrically associated to the jet: many theory issues, and potential sources of uncertainty

Extracted q/g-like fragmentation observables fit expectations:









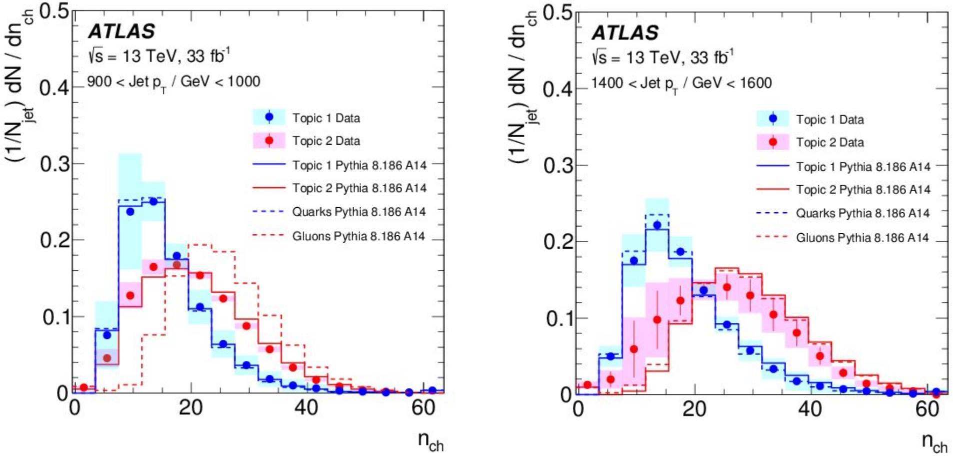
Model-independent quark/gluon jet characterisation

Novel approach is to use “topic modeling” extraction.

The categories are defined by data rather than MC internals:

$$h_i^{T_1} = \frac{h_i^f - \left(\min_j\{h_j^f/h_j^c\}\right) \times h_i^c}{1 - \min_j h_j^f/h_j^c}$$

$$h_i^{T_2} = \frac{h_i^c - \left(\min_j\{h_j^c/h_j^f\}\right) \times h_i^f}{1 - \min_j h_j^c/h_j^f}$$

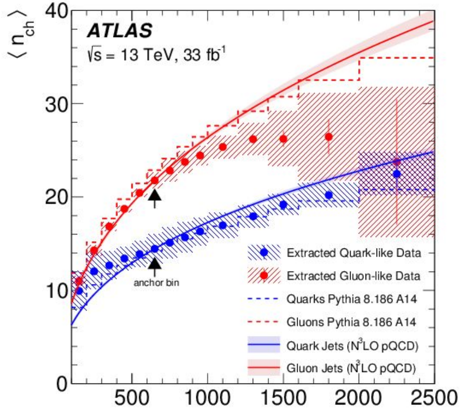


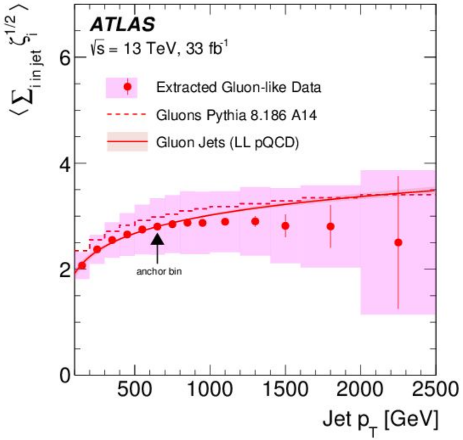
Interesting new approach. Limitation: alignment of topics to q and g template ideas relies on the existence of bins dominated by q or g: applies to n_{ch} distribution only

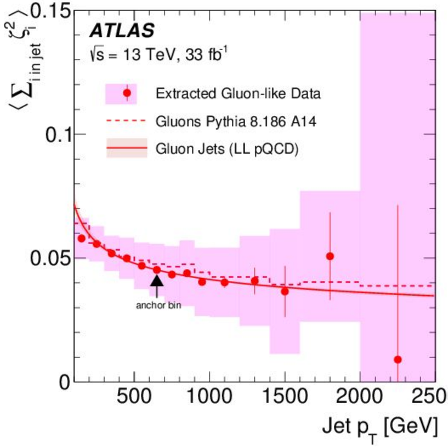
Comparing quark/gluon jet characterisations

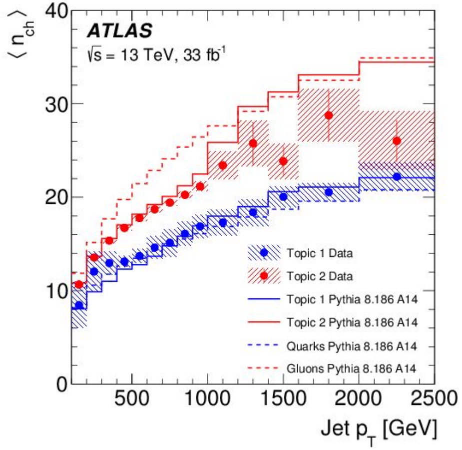
Pythia-based vs topic modeling: good description by Pythia for quarks in both; less good for gluons. “Quark” topic also aligns well with quarks, worse for gluons.

pQCD normalization-anchored, since can’t handle non-perturbative physics: compares well to q/g extractions







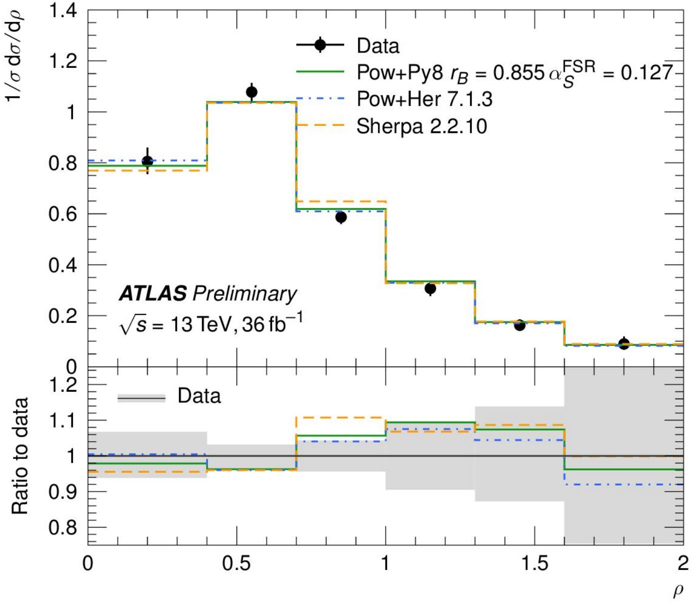


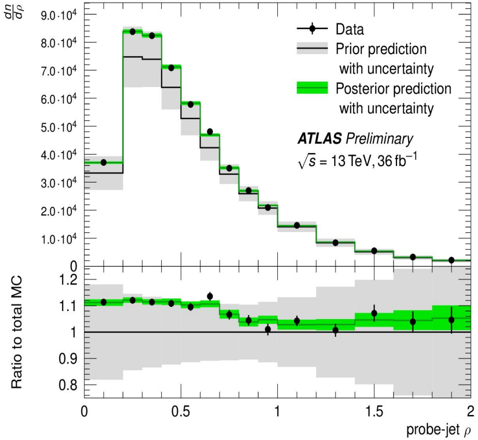
ATLAS *b*-fragmentation

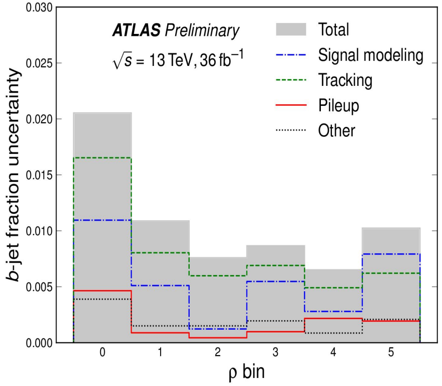
ATLAS-CONF-2020-050

ρ moment

- Ratio of *b*-hadron p_T to average of charged lepton p_T's — comparison of the *b* momentum to the *tt* parent event scale
- Lepton p_T more precisely measurable than *b*-jets or *tt*
- Sensitive to QCD radiation not contained in the *b*-jets





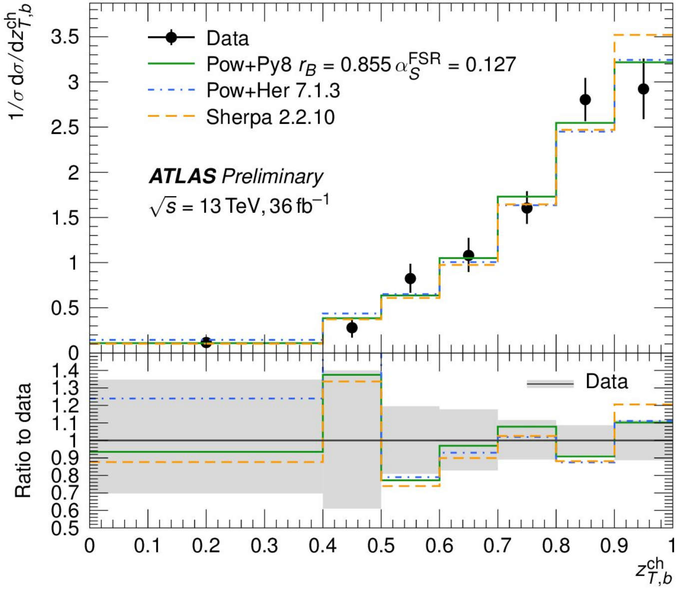


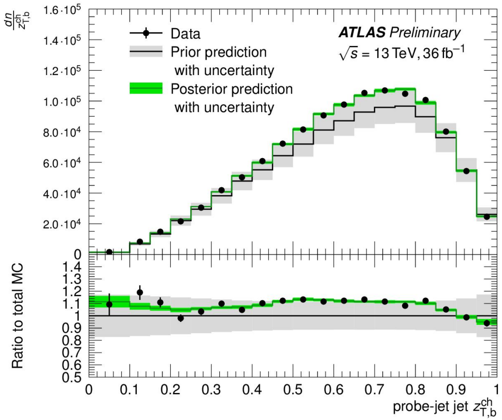
ATLAS-CONF-2020-050

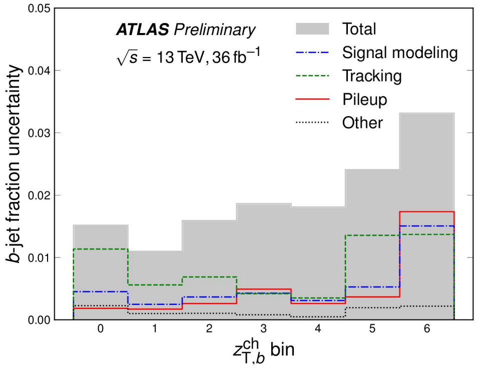
ATLAS *b*-fragmentation

$z_{\mathrm{ch,T}}$ (transverse) moment

- $p_{\mathrm{T,rel}}$ (projected p_{T} wrt the parent jet axis) proved difficult to measure
- Correlated with longitudinal projection, plus some residual information about directional kicks from radiation

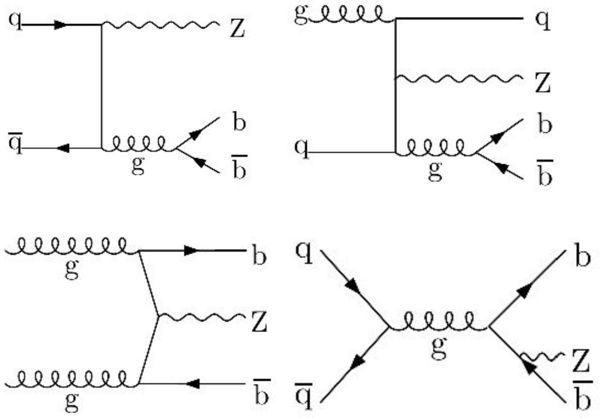






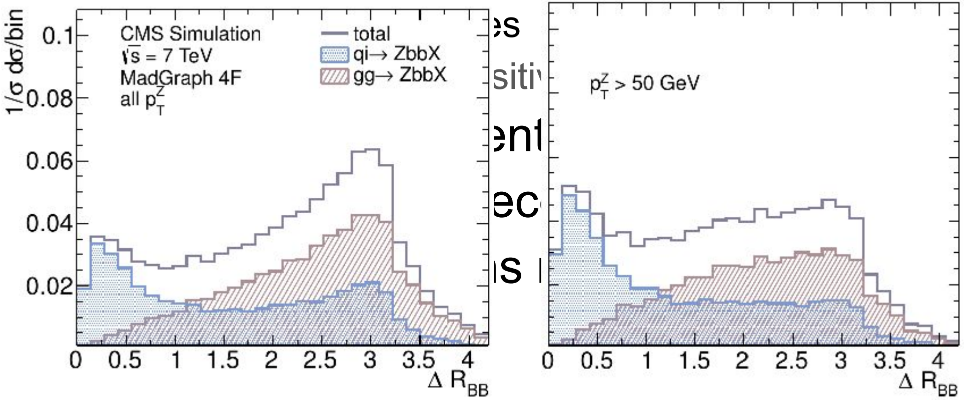
JHEP 12 (2013) 039

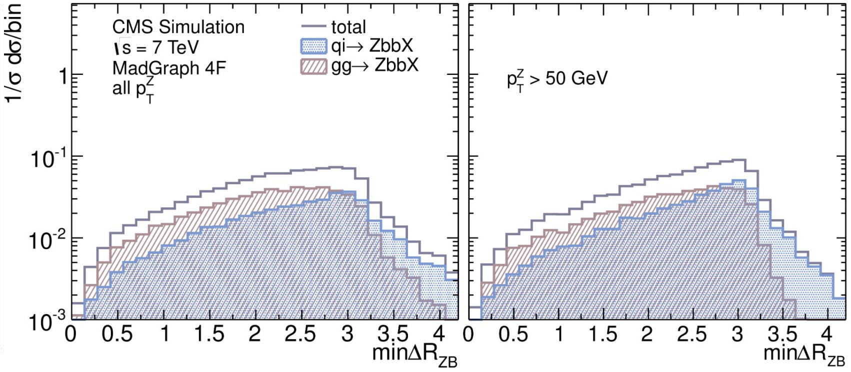
CMS *Z+BB* cross sections at 7 TeV



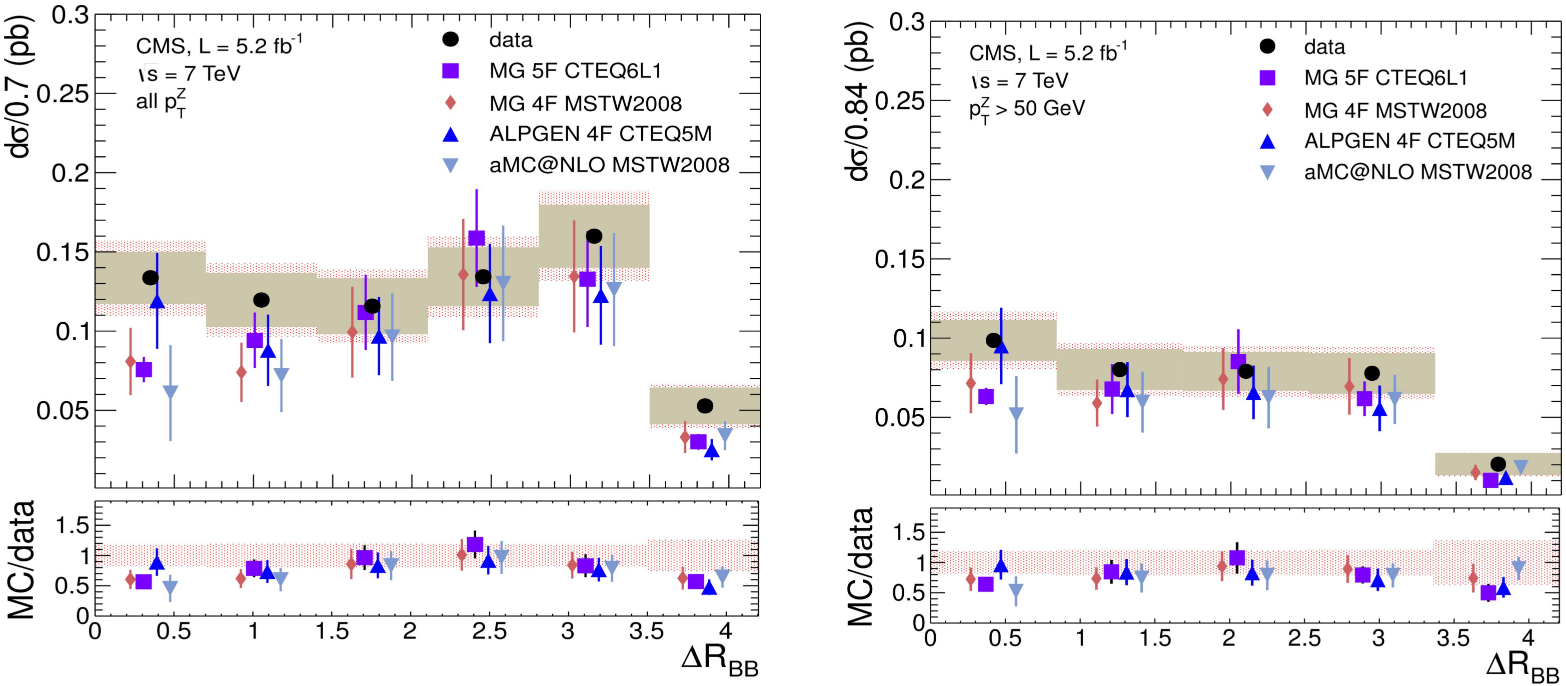
- Measure the production of a Z boson in association with B hadrons as a function of angular correlation especially the B-hadron pair production at small angular separation where there are significant theoretical uncertainties in the description of collinear production
    - ○ ΔR_{BB} direct test of the modeling of different pp→ ZbbX

es
sitiv
ent
ec
s l



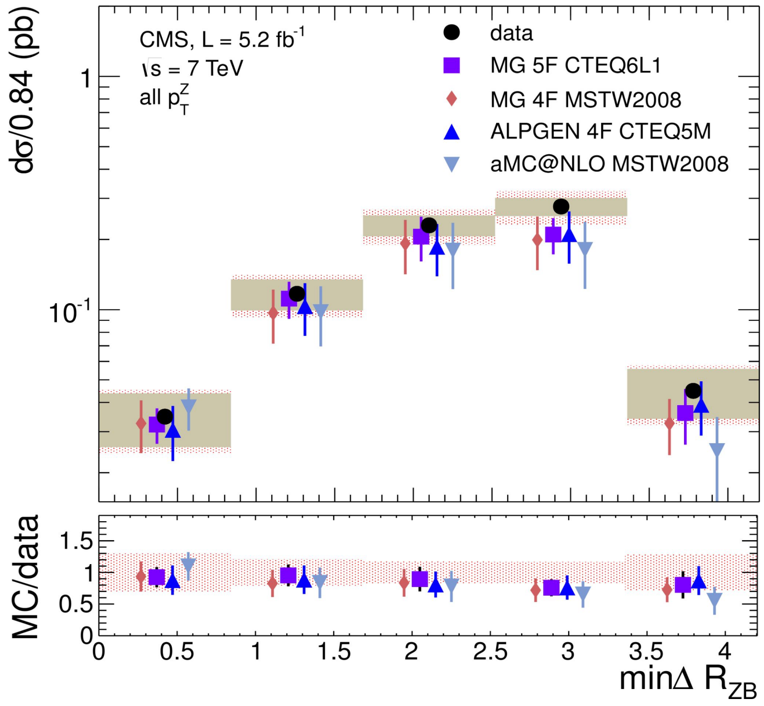


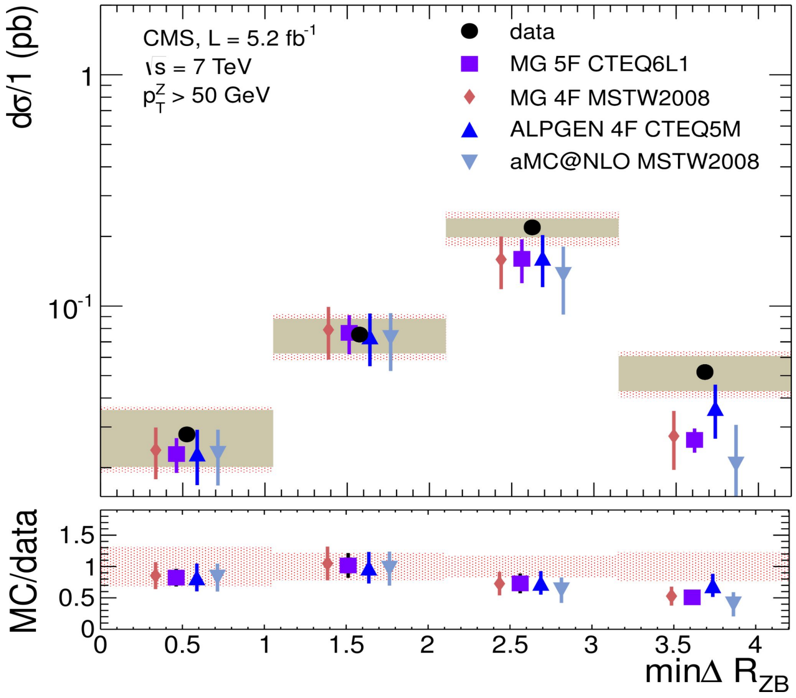
CMS *Z+BB* cross sections at 7 TeV



- ALPGEN well describes data in collinear regions ($\Delta R_{BB} < 0.7$) while MADGRAPH and amc@NLO are lower than data
- At large ΔR_{BB}, all predictions agree with data

CMS *Z*+*BB* cross sections at 7 TeV



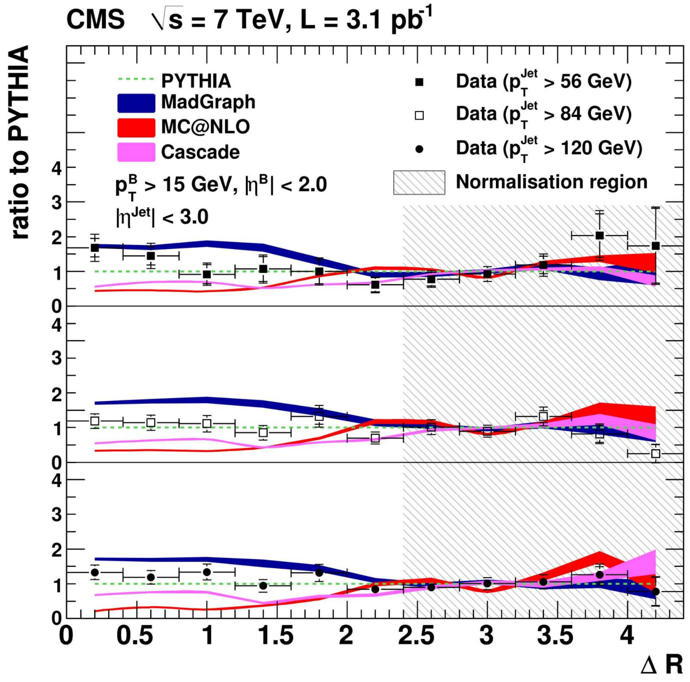


- Data are consistent with predictions in **min**ΔR_{ZBB} except for region above 2

CMS angular correlations between *BB* at 7 TeV

JHEP03 (2011) 136

- *BB* production at 7 TeV
- *B* hadrons are identified using reconstructed secondary vertex → probe small separation angle
- Simulation is normalized to regions > 2.4

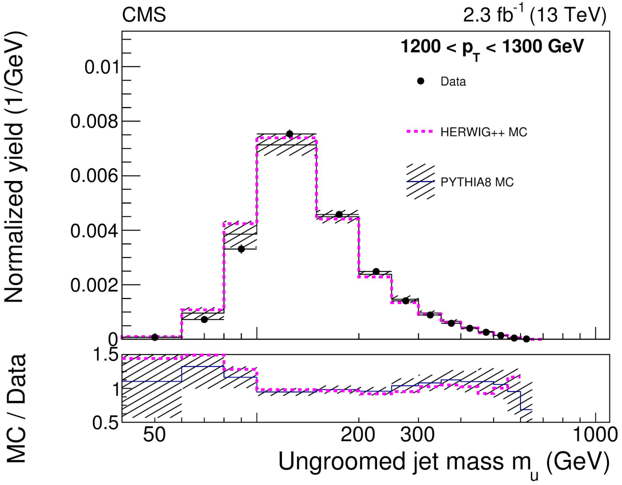


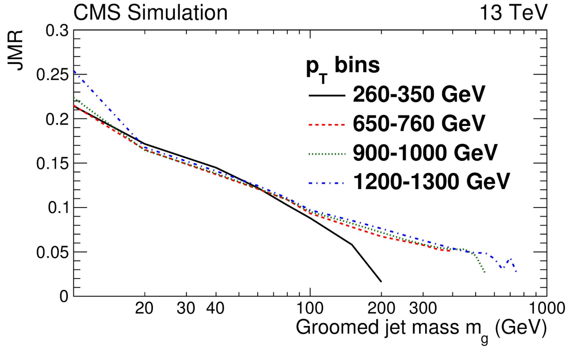
- None of the predictions describe the data very well
- amc@NLO mismodels at small angle where gluon splitting is significant

CMS jet mass in dijet events at 13 TeV

arXiv:1807.05974

Uses 33 fb^{-1} dataset of 13 TeV pp collisi





CMS jet mass in dijet events at 13 TeV

arXiv:1807.05974

Uses 33 fb^{-1} dataset of 13 TeV pp collisi