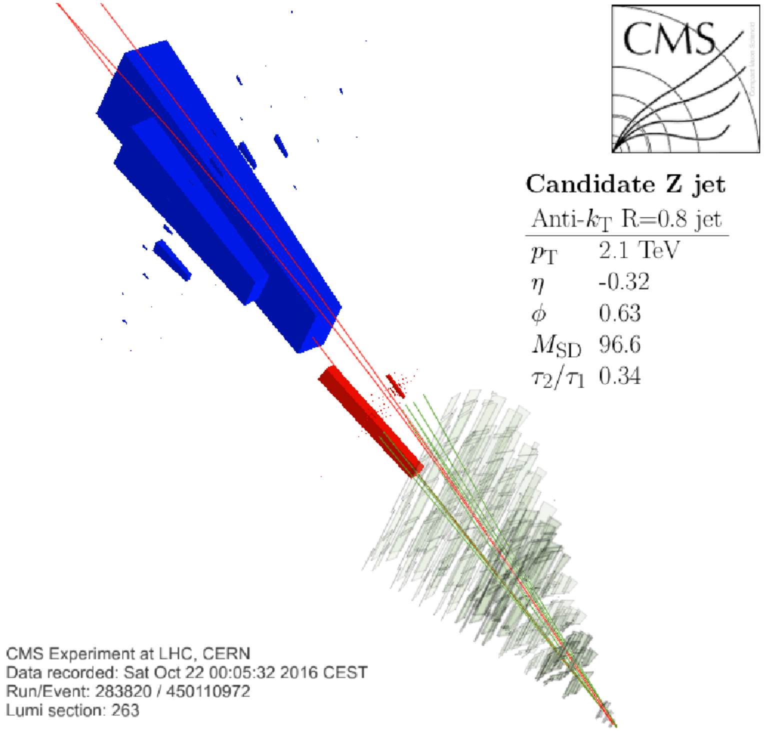
Jet substructure with CMS



Andreas Hinzmann (U. Hamburg)
26. Oct 2020

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Emmy Noether-Programm
DFG Deutsche Forschungsgemeinschaft

Outline

- Jet substructure as a tool
  - How we make use of jet substructure for jet identification?
  - What physics we can do with such jet identification?

- From jet shape to jet substructure measurements
  - What has been measured?
  - What we learn from it?

- Jet substructure reconstruction and the future
  - How we measure jet substructure at LHC and HL-LHC?

- Disclaimer: Not a full summary, but a few recent examples as input for discussion in LHC EW WG. Observables and physics well introduced in Jennifer's+Nima's talks! Mostly focus on pp here.

Quark and gluon taggers

- Gluon jets differ from quark jets
    - More particles
    - Wider

- CMS reference tagger combines

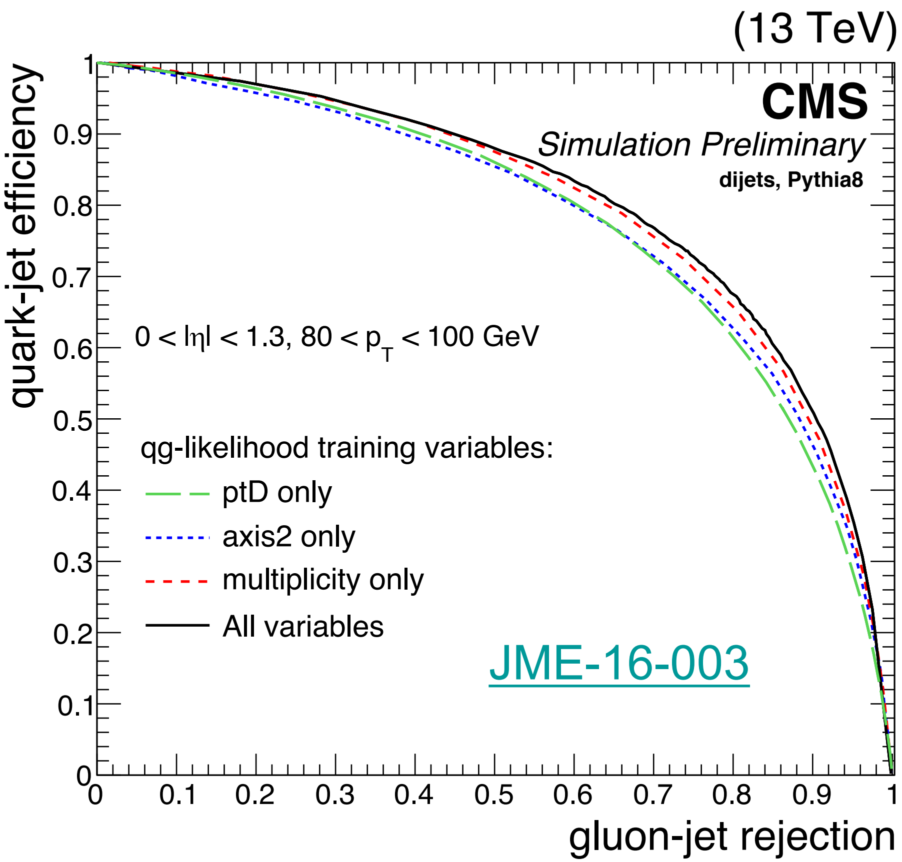
jet constituents multiplicity

$$p_T D = \frac{\sqrt{\sum_i p_{T,i}^2}}{\sum_i p_{T,i}}$$

jet minor angular opening (σ_2)

- Mild separation, not a yes/no
    - Used as ingredient to multivariate event selection

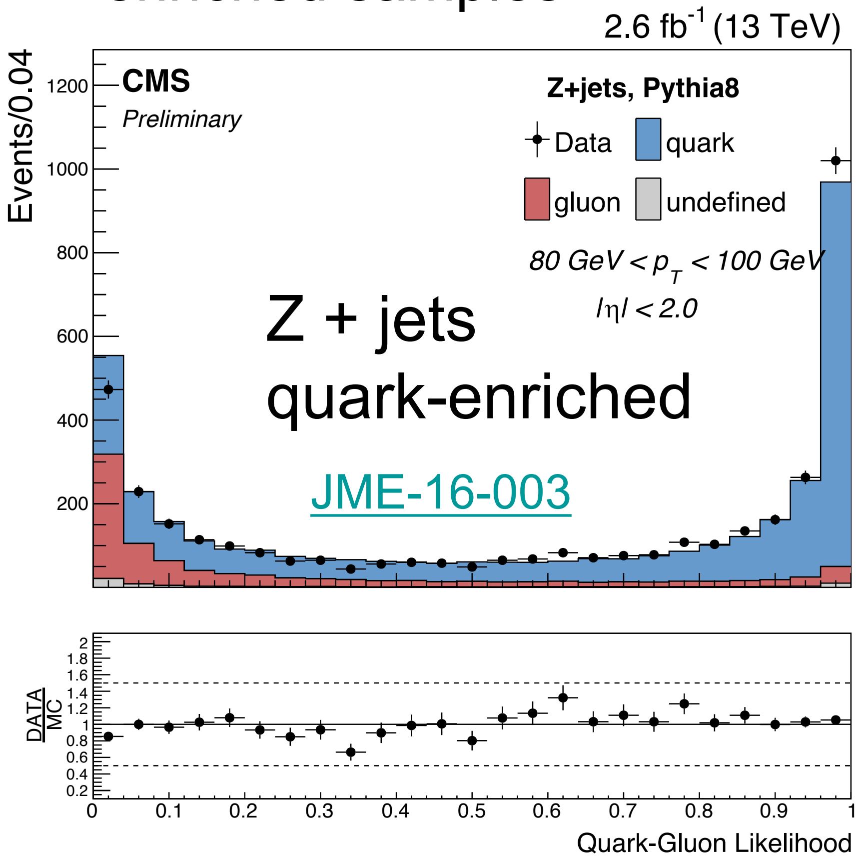
- Example physics
    - VBF Higgs→bb
    - Z+2 jets



JME-16-003

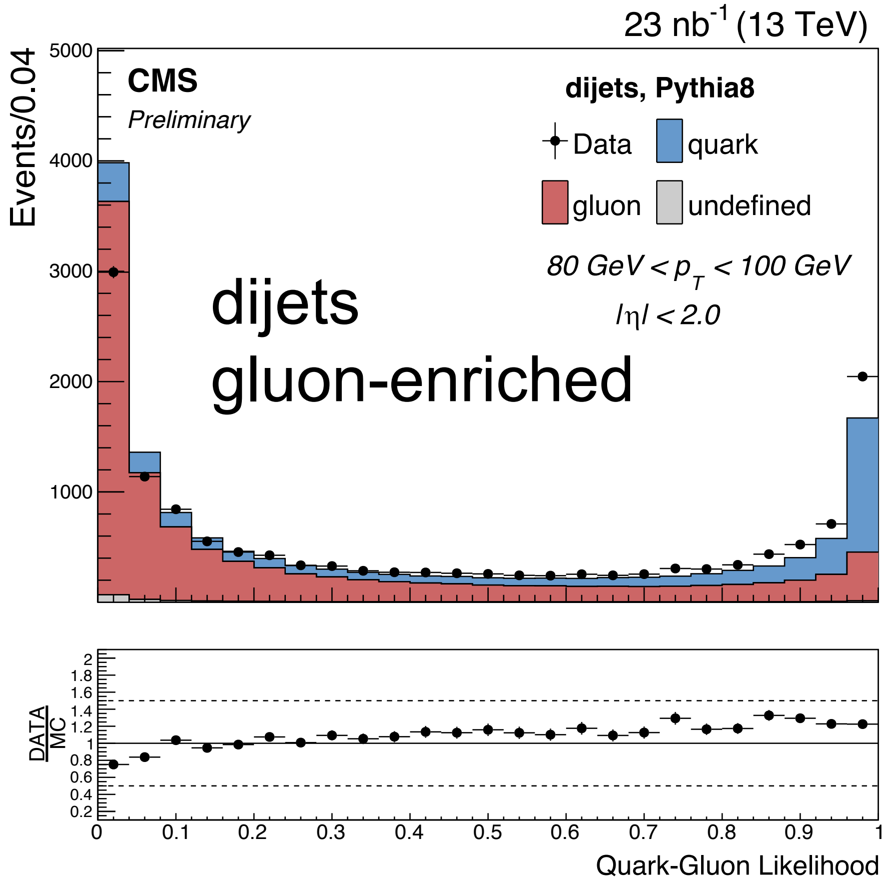
Quark and gluon tagger modeling

- Gluon jet shape/substructure not well modeled by parton shower generators (detector response modeling is subdominant effect)
- Morph jet substructure distributions of quark and gluon jets in generator predictions to match data measurement in quark and gluon enriched samples



Z + jets
quark-enriched

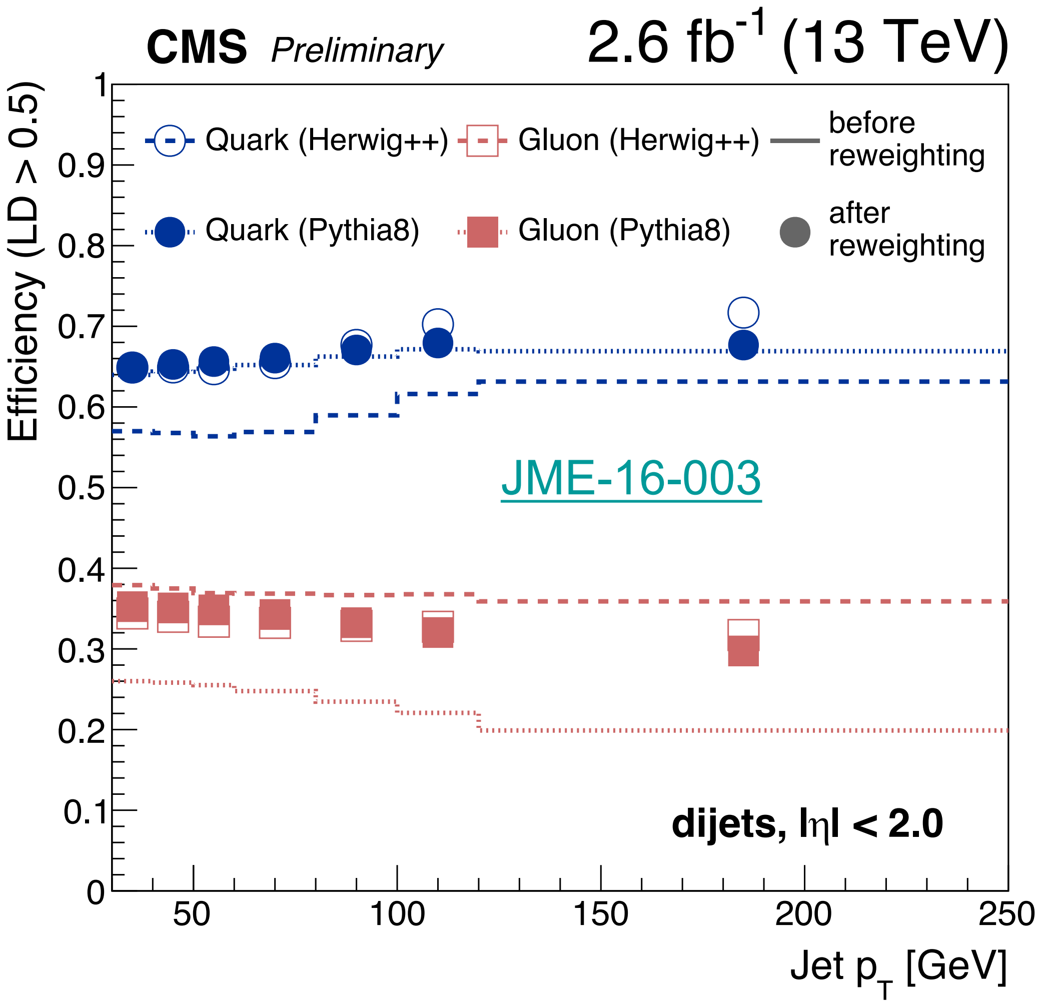
JME-16-003



dijets
gluon-enriched

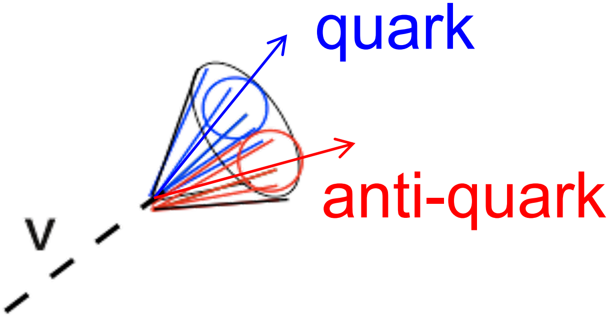
Quark and gluon tagger modeling

- Associated systematic uncertainties will profit from better predictions = smaller corrections
  → jet substructure measurements and theoretical understanding



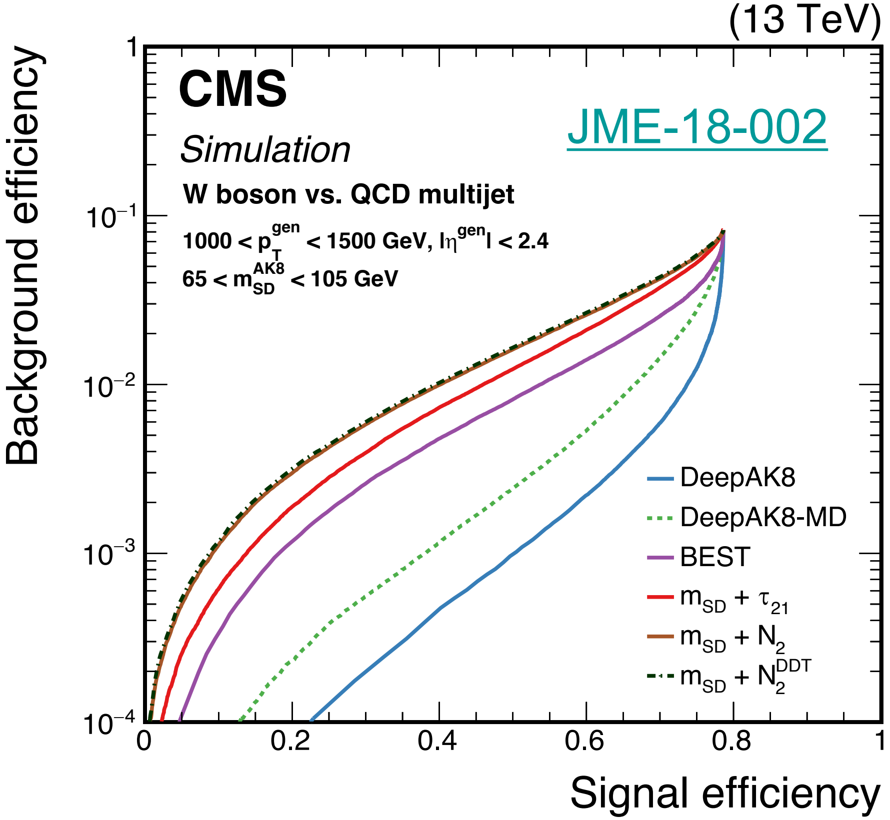
Boosted W/Z/H and top taggers

- W/Z/H and top jets differ from q/g jets
  - (groomed) jet mass
  - Number of prongs + more



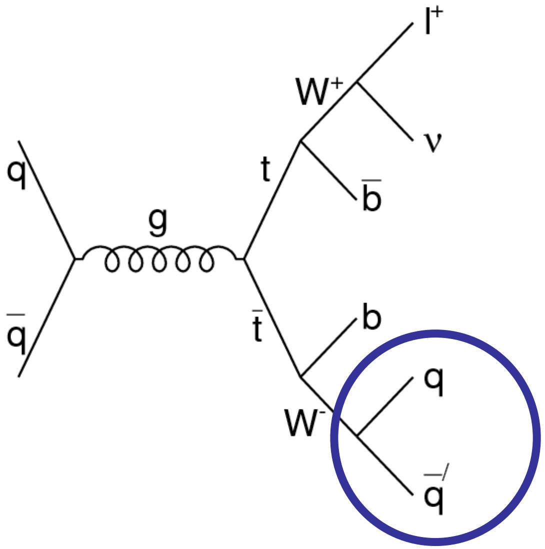
- Taggers range from “cut on 2 variables” to machine learning from parton showers

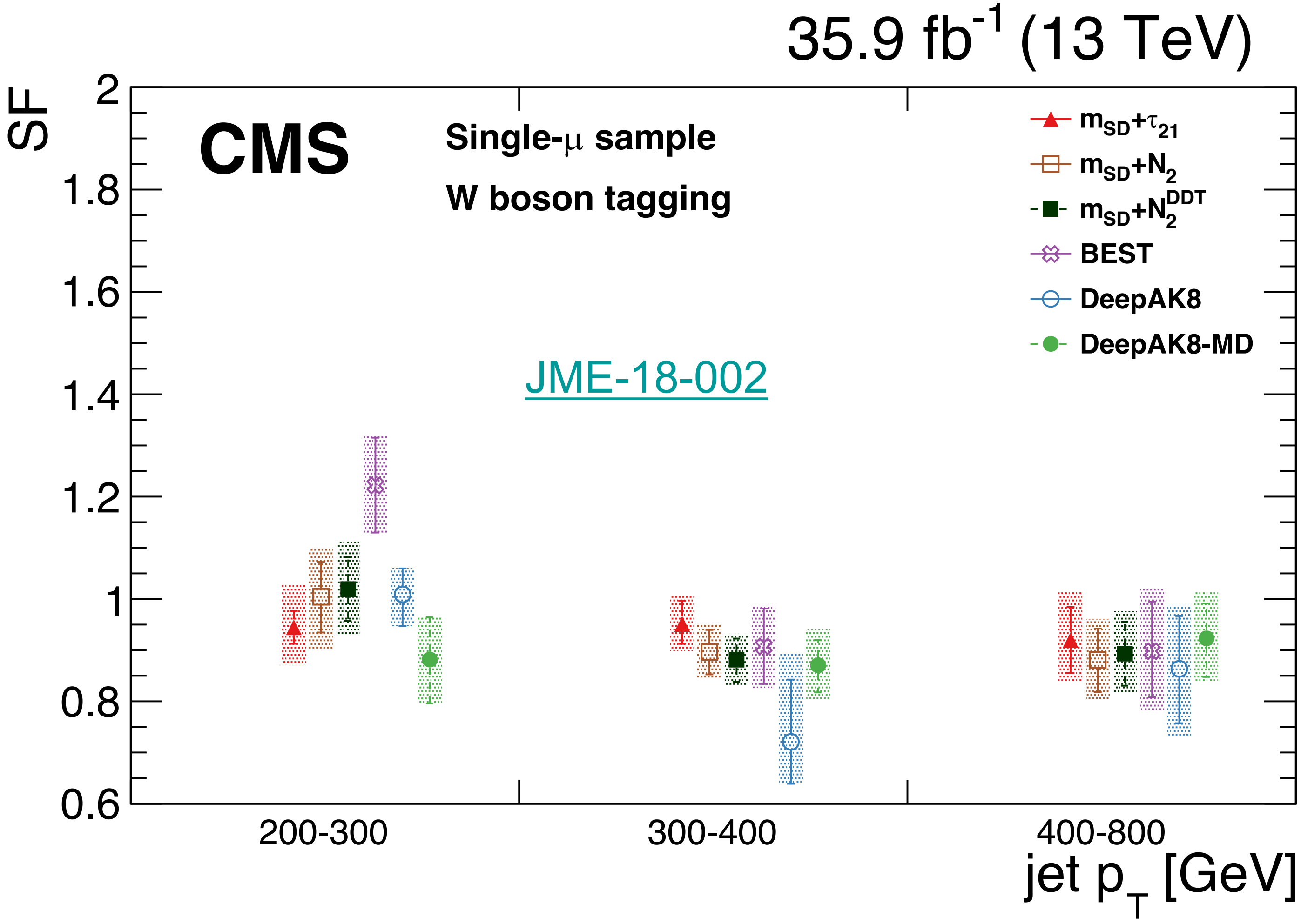
- Example physics
  - tt, VV, VLQ resonances
  - Top
  - aTGCs/aQGCs
  - DM
  - SUSY
  - Higgs→bb



Boosted W tagger modeling

- Tagging efficiency data/simulation SF measured in ttbar samples
  - Not perfect, but good enough for most purposes



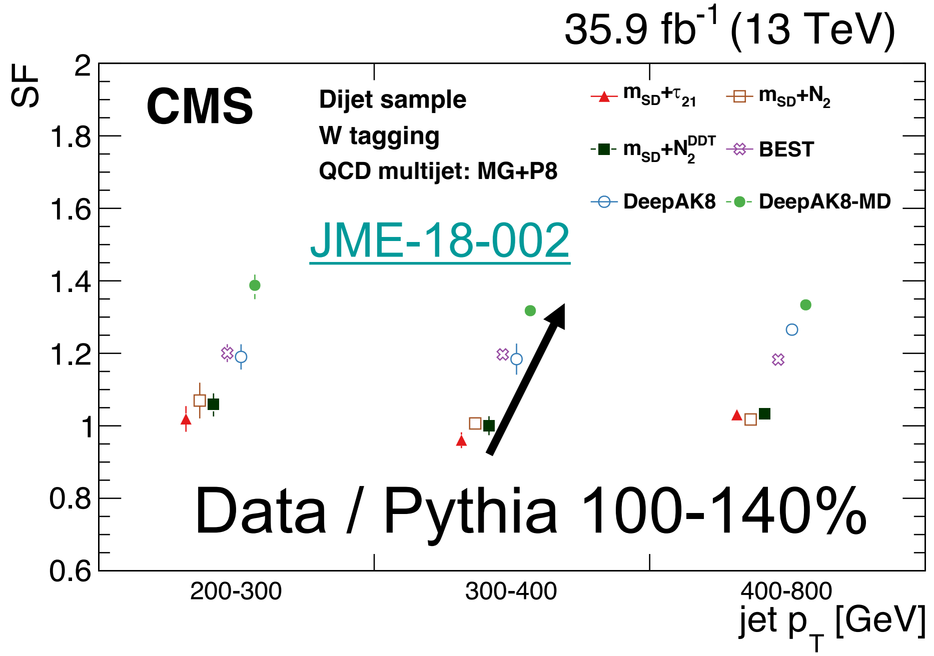


Boosted W tagger modeling

- Mistag rate, i.e. rate to identify a q/g jet as W boson measured in dijet or photon+jet samples

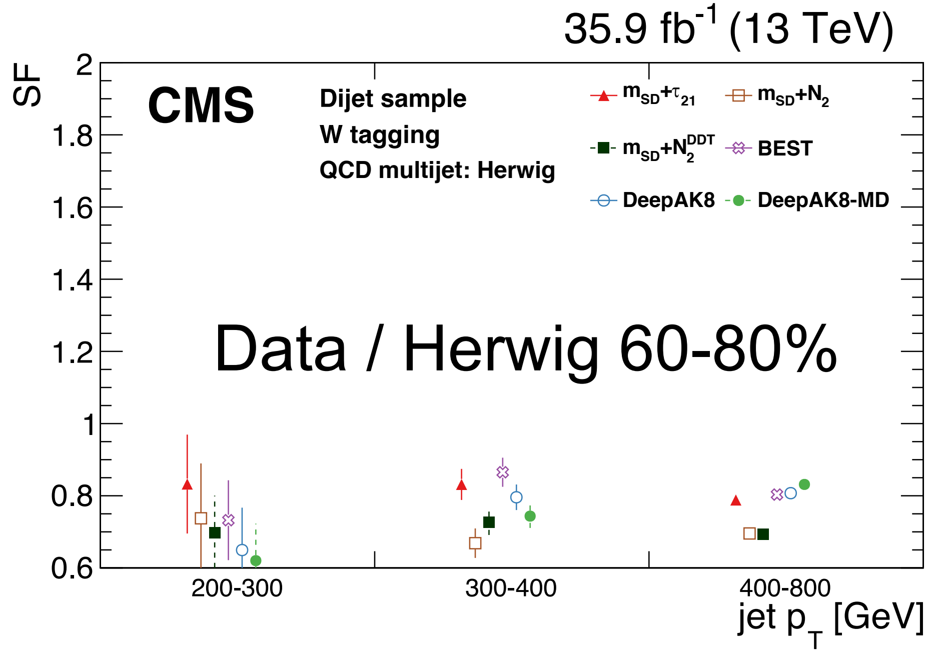
- Significant mismodeling
  - SF depends on q/g content of sample
  - Most pronounced for ML taggers trained with Pythia, but no such trend compared to Herwig

- In practice, physics analyses make use of data-driven methods for background estimation to avoid relying on predictions of mistag rate



JME-18-002

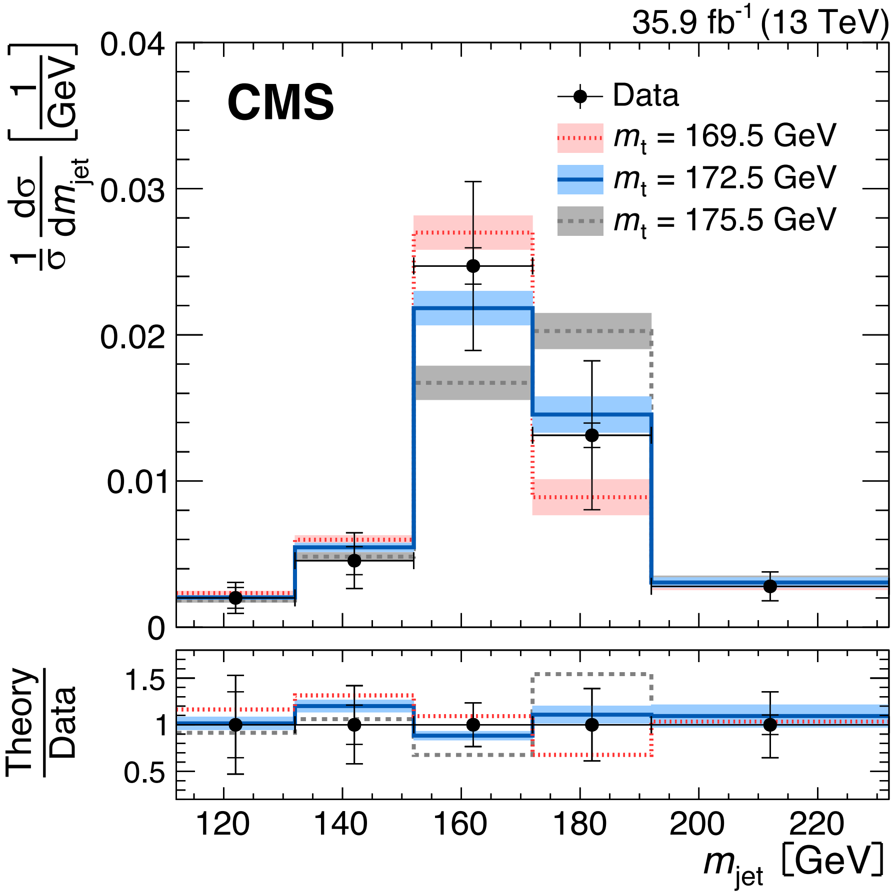
Data / Pythia 100-140%



Data / Herwig 60-80%

Boosted top mass

- Idea: boosted top jet mass could avoid ambiguity in top mass definition in direct measurements
- Variable radius X-cone algorithm to find top jets and subjets with “grooming” (subjet p_T>30 GeV)
- Dominant experimental uncertainty: subjet energy scale
- Dominant model uncertainty from parton shower FSR

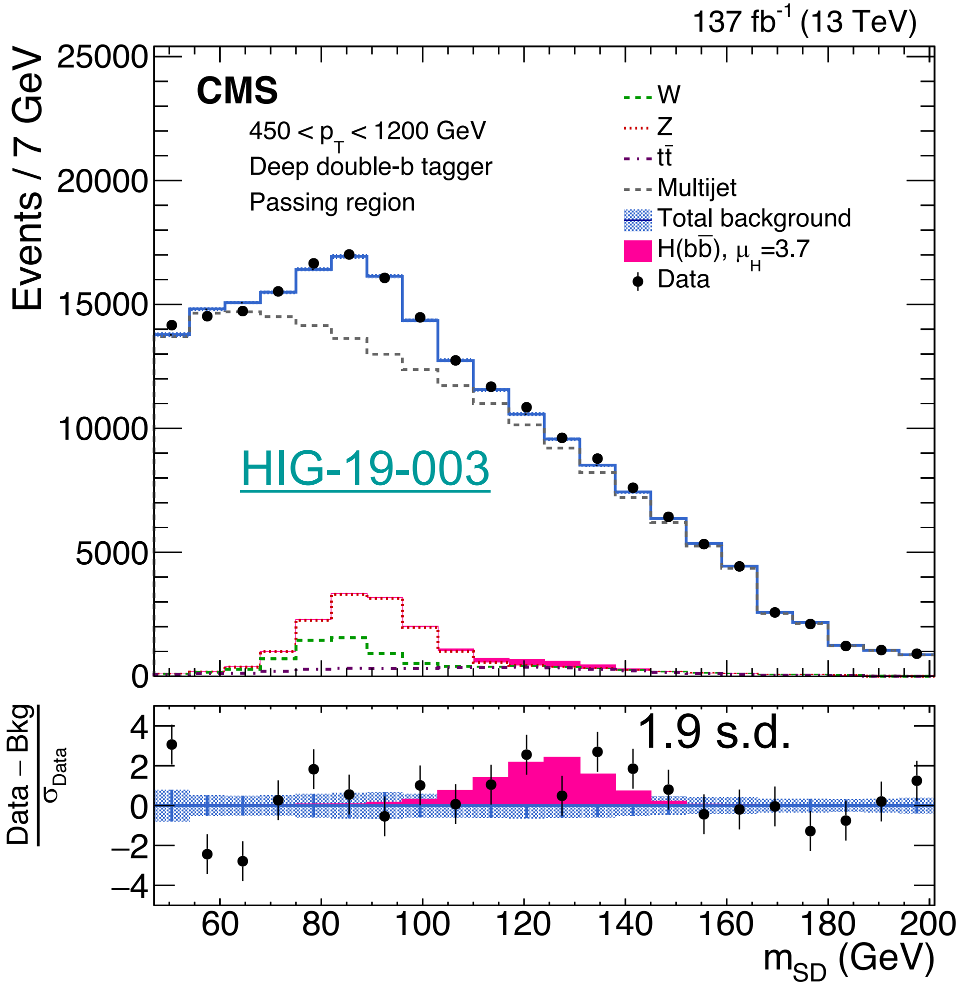


$$m_t = 172.6 \pm 0.4\,(\text{stat}) \pm 1.6\,(\text{exp}) \pm 1.5\,(\text{model}) \pm 1.0\,(\text{theo})\,\text{GeV}$$

Boosted H→bb

- High p_T Higgs→bb extracted out of QCD multijet background with machine-learned H→bb tagger
- Tagger uncorrelated with softdrop mass in QCD multijet background → smooth spectrum to look for peak
- Data-driven background estimate
  - Large source of uncertainty

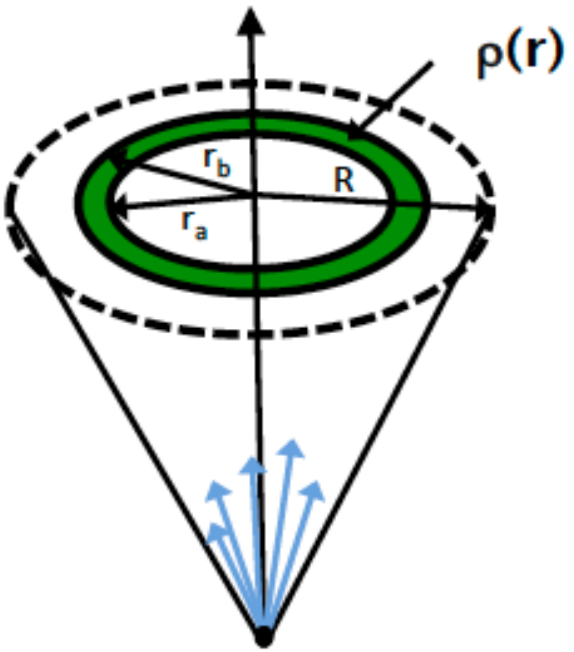
Uncertainty source	$\Delta\mu_H$	
Statistical	+1.2	−1.2
QCD pass-fail ratio (data correction)	+0.7	−0.7
$t\bar{t}$ normalization and misidentification	+0.4	−0.4
Systematic	+0.6	−0.7
QCD pass-fail ratio (simulation)	+0.5	−0.5
DDBT efficiency	+0.3	−0.4
Jet mass scale and resolution	+0.3	−0.3
Jet energy scale and resolution	+0.1	−0.1
Simulated sample size	+0.2	−0.1
Other experimental uncertainties	+0.1	−0.1
Theoretical	+0.8	−0.5
V+jets modeling	+0.6	−0.4
H modeling	+0.5	−0.3
Total	+1.6	−1.5



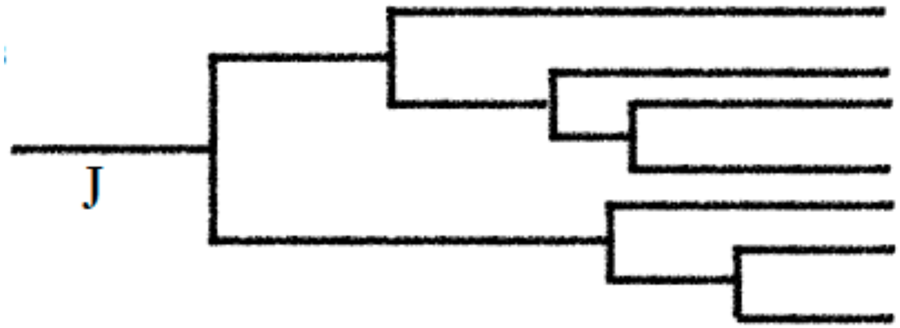
From jet shapes to jet substructure

"Jet shapes"
Width
Multiplicities
Fragmentation functions

"Jet substructure"
Jet tagging variables
→ better performance
→ better calculable
Pick out specific splittings



$$z_g = \frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{\mathrm{cut}} \left(\frac{\Delta R_{12}}{R_0} \right)^{\beta}$$

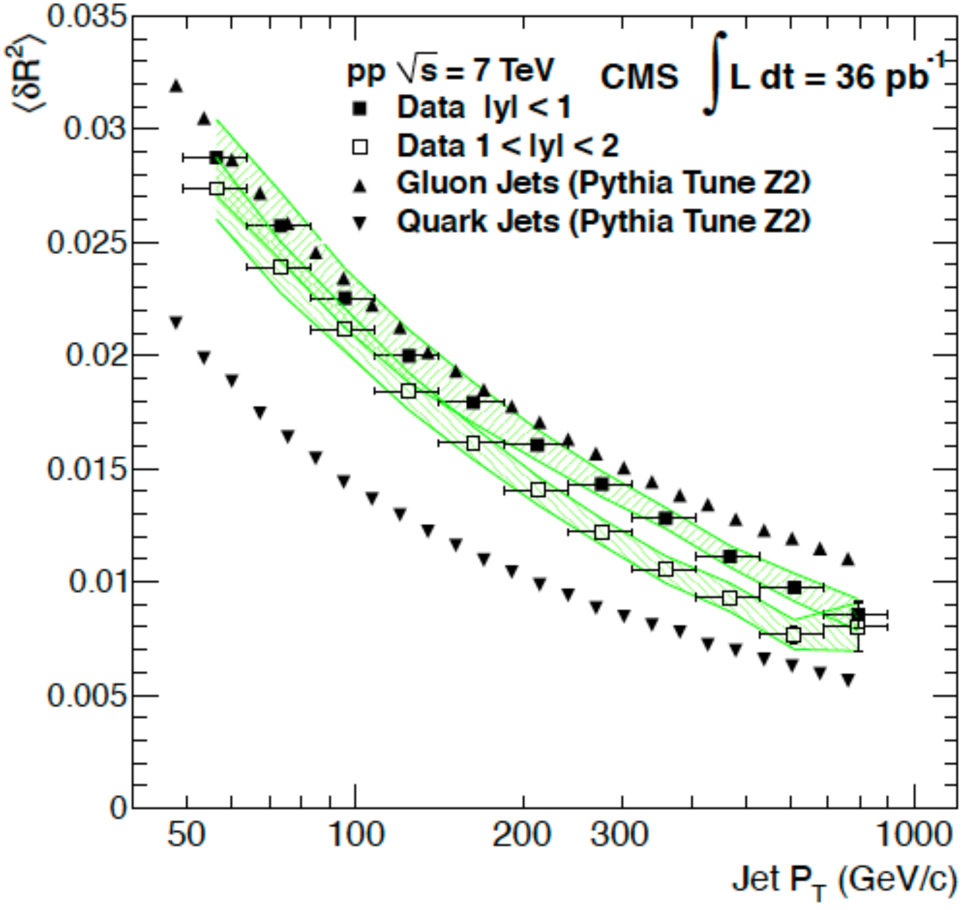


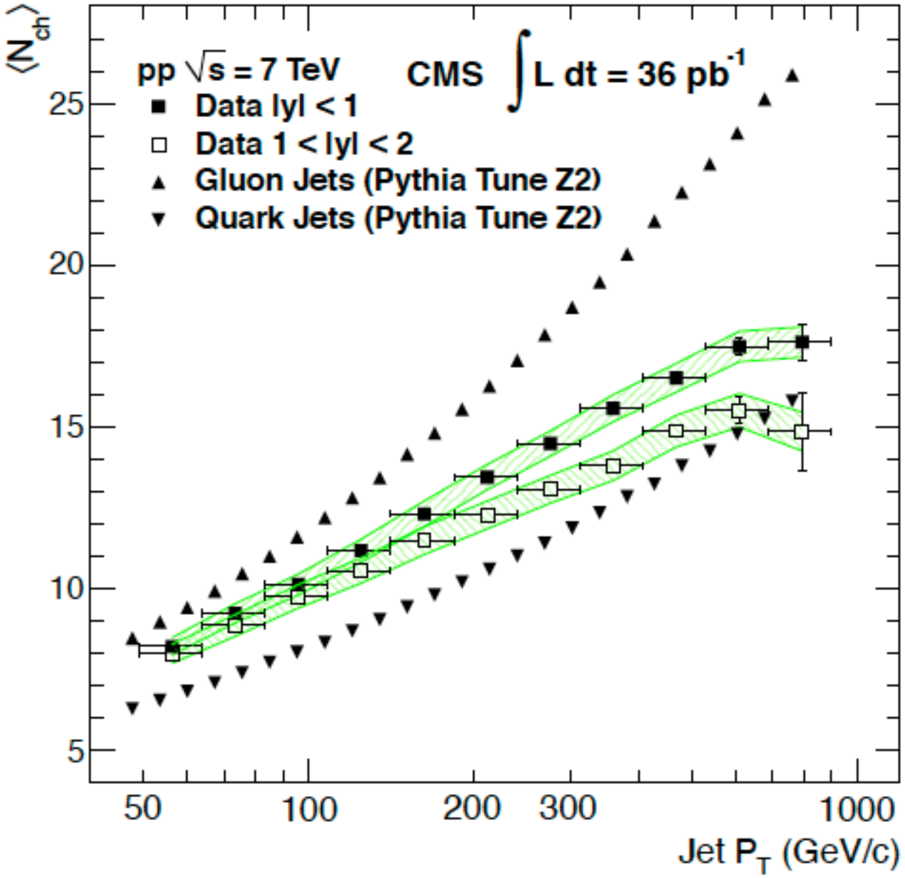
Jet substructure measurements

Reference	Final state	Jets, p_T (GeV)	Jet substructure observables
1204.3170 7 TeV pp	jets	q/g-jets (AK7), $20<p_T<1000$ q/g-jets (AK5), $50<p_T<1000$	jet shapes, charged hadron multiplicity and width
1205.5872 2.76 TeV pp/PbPb	dijets	q/g-jets (AK3), $40<p_T<320$	fragmentation functions
1310.0878 2.76 TeV pp/PbPb	jets	q/g-jets (AK3), $100<p_T<300$	fragmentation functions
1310.0878 2.76 TeV pp/PbPb	jets	q/g-jets (AK3), $p_T>100$	jet shapes
1809.08602 5.02 TeV pp/PbPb	γ+jet	q-jets (AK3), $p_T>30$	jet shapes
HIN-19-003 5.02 TeV pp/PbPb	dijets	q/g-jets (AK4), $p_T>50$	jet shapes
QCD-10-041 7 TeV pp	dijets	q/g-jets (KT6), $97<p_T<1032$	subjet multiplicities and p_T^{rel}
1706.05868 8 TeV pp	jet	q/g-jets (AK5), $400<p_T<1500$	jet charge
2004.00602 5.02 TeV pp/PbPb	jets	q/g-jets (AK4), $p_T>120$	jet charge
1703.06330 8 TeV pp	ttbar	top-jets (CA12), $p_T>400$	jet mass
1303.4811 8 TeV pp	dijets W/Z+jets	q/g-jets (AK7), $220<p_T<1500$ q-jets (AK7, CA8, CA12), $125<p_T<450$	jet mass, pruned/trimmed/filtered jet mass
1805.05145 5.02 TeV pp/PbPb	jets	q/g-jets (AK4), $140<p_T<300$	softdrop jet mass
1807.05974 13 TeV pp	dijets	q/g-jets (AK8), $200<p_T<1300$	jet mass, softdrop jet mass
1911.03800 13 TeV pp	ttbar	top-jets (XC12), $p_T>400$	XCone-grommed jet mass
1708.09429 5.02 TeV pp/PbPb	jets	q/g-jets (AK4), $140<p_T<500$	softdrop splitting function
1808.07340 13 TeV pp	ttbar	q-jets (AK4), $p_T>30$ g-jets (AK4), $p_T>30$ b-jets (AK4), $p_T>30$	jet substructure and softdrop observables

Jet shapes at 7 TeV

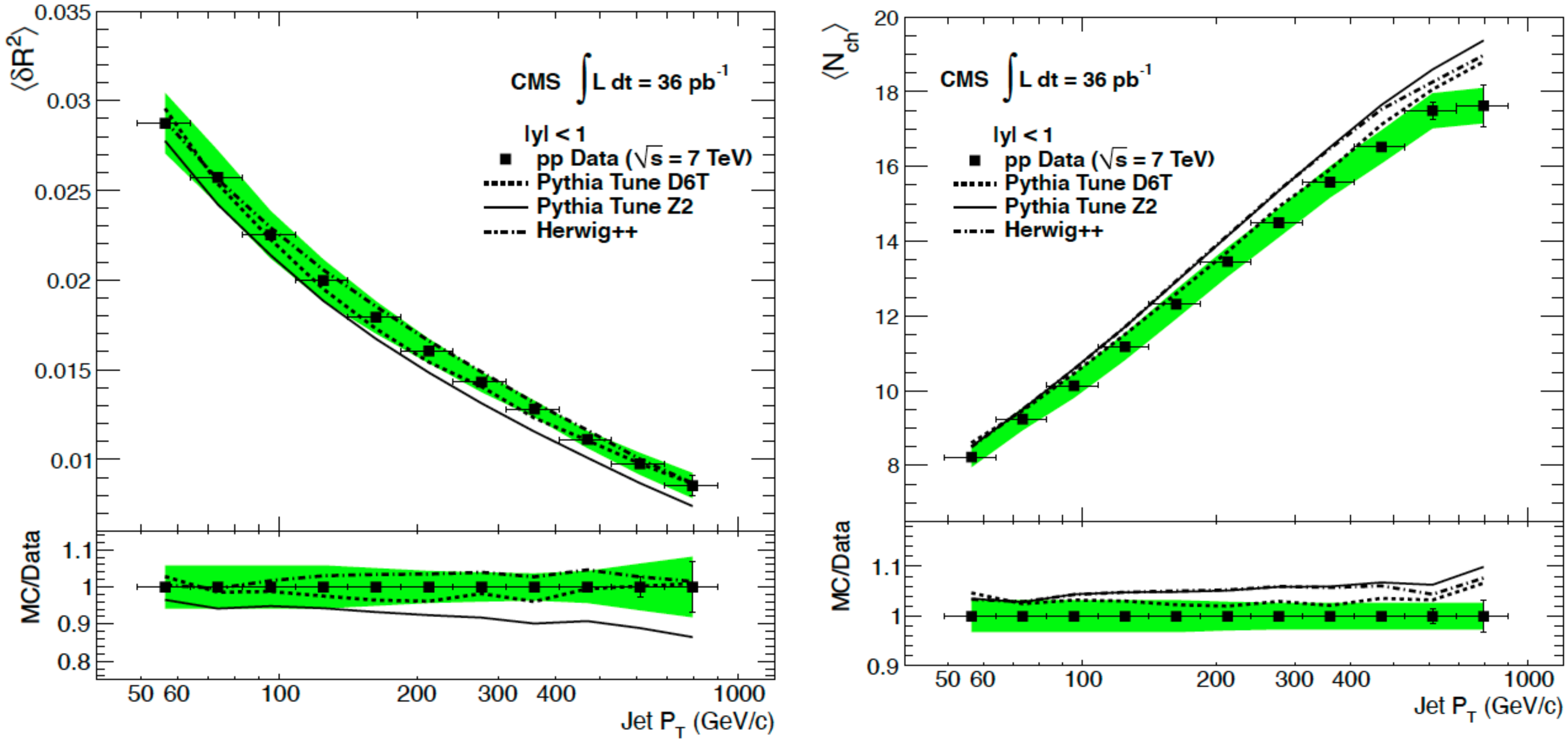
- Width and constituent multiplicity larger for gluon jets
- Central rapidity slightly more gluon enriched than forward rapidity





Jet shapes at 7 TeV

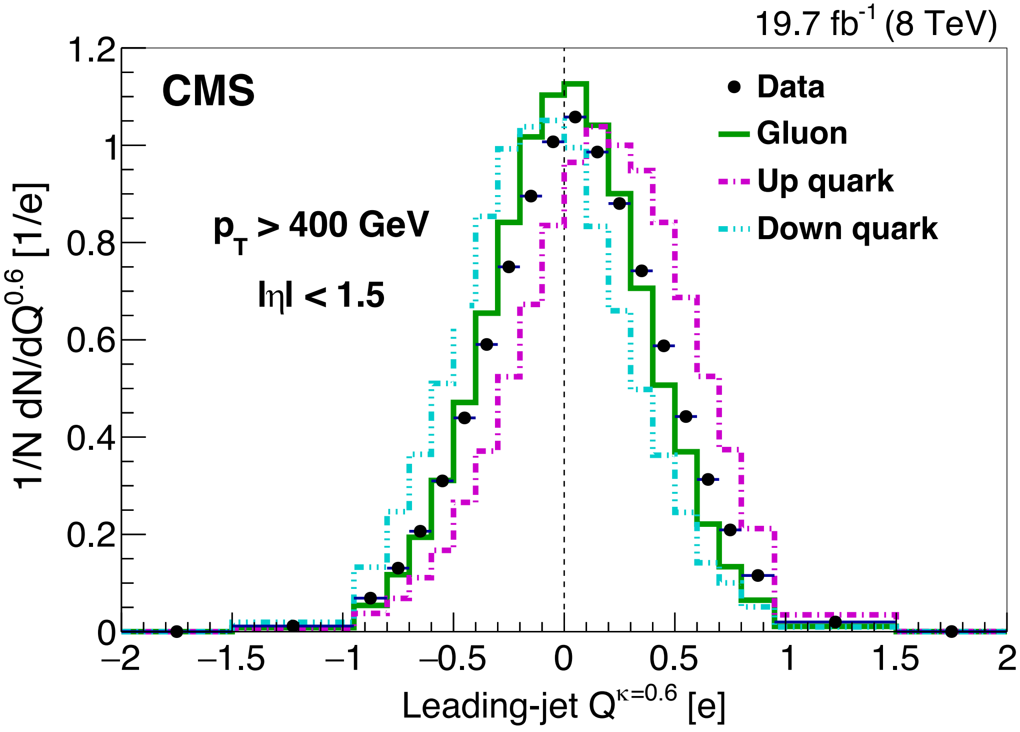
- Width and constituent multiplicity larger for gluon jets
- Central rapidity slightly more gluon enriched than forward rapidity
- Reasonable, but no perfect agreement for initial LHC generator tunes
- Observable checked, but not used for tuning simulation

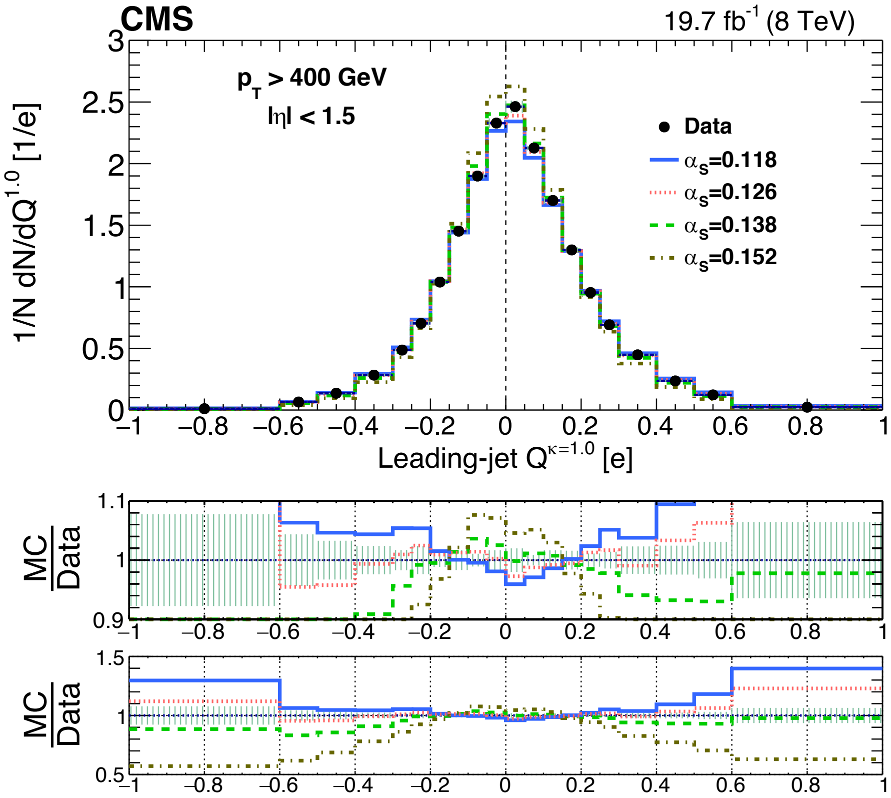


Jet charge in pp

- Jet charge estimator for parton charge
- Sensitive to α_S used in FSR shower

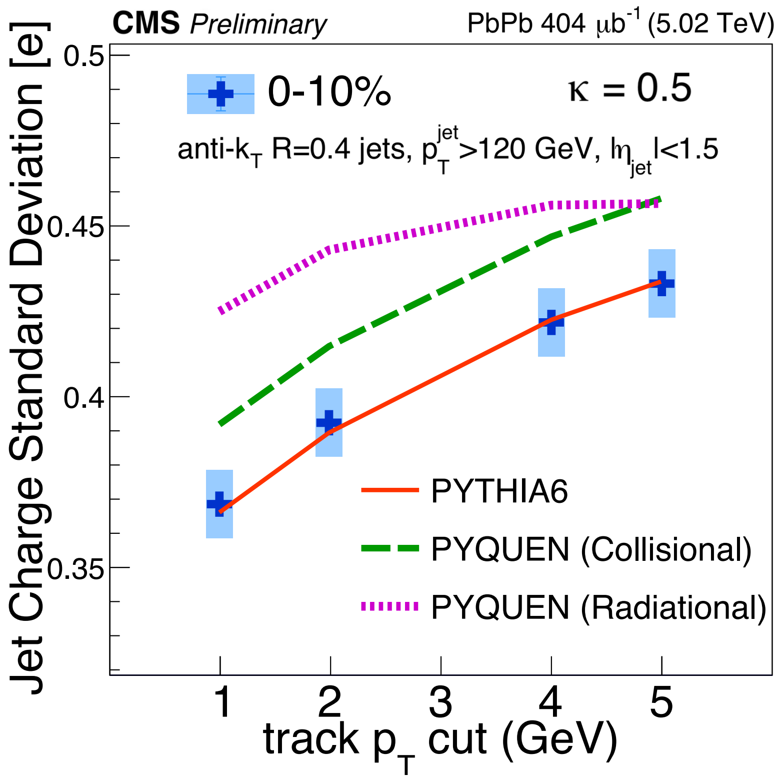
$$Q^\kappa = \frac{1}{(p_T^{\text{jet}})^\kappa} \sum_i Q_i (p_T^i)^\kappa$$

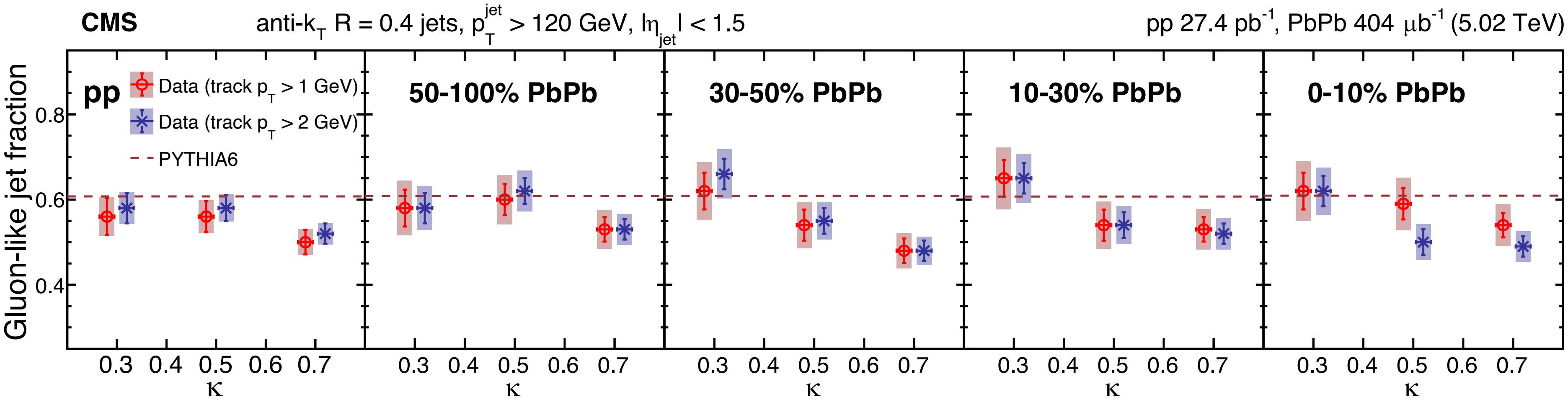




Jet charge in PbPb

- Question: Does QGP change fraction of quark and gluon jets?
- Jet charge compared in pp and PbPb
- Quark and gluon fractions extracted with templates from PS simulation
- No significant deviation between pp and PbPb observed





Softdrop jet mass

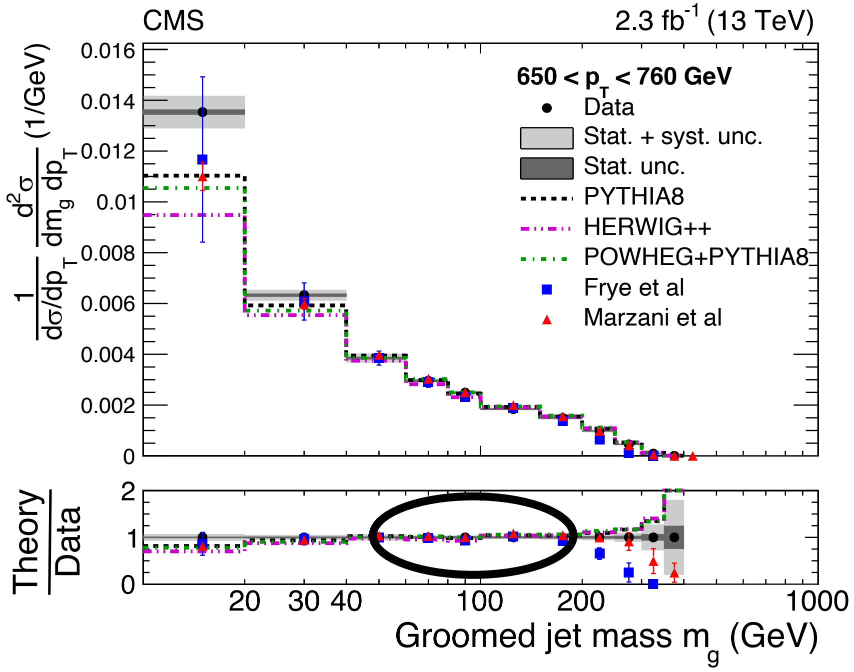
- Jet mass with softdrop (β=0)

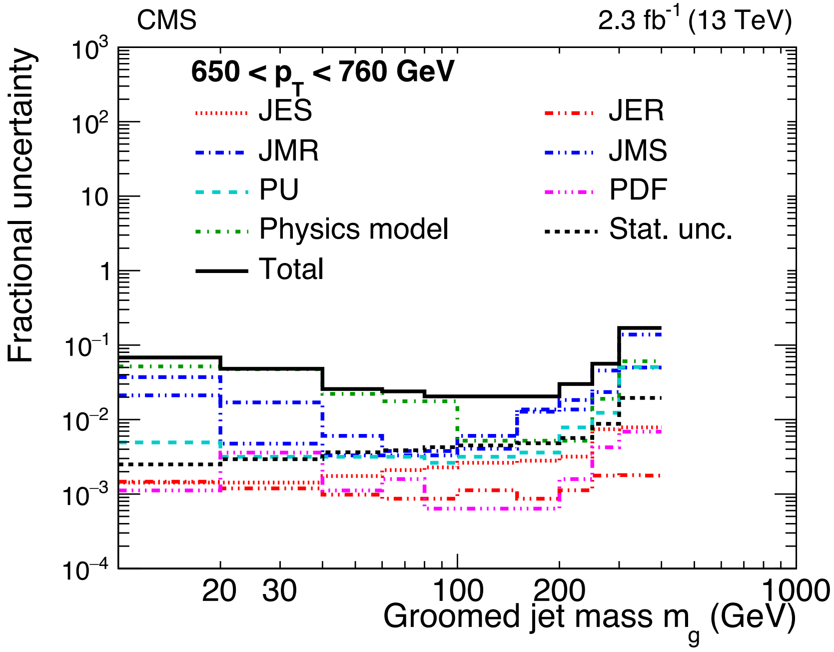
$$\frac{\min(p_{T1},p_{T2})}{p_{T1}+p_{T2}} > z_{\text{cut}}$$

- Low mass ~ non-perturbative
- Intermediate mass ~ resummation
- High mass ~ perturbative

- LO+NLL prediction describes intermediate mass very well

- Dominant uncertainty from physics model



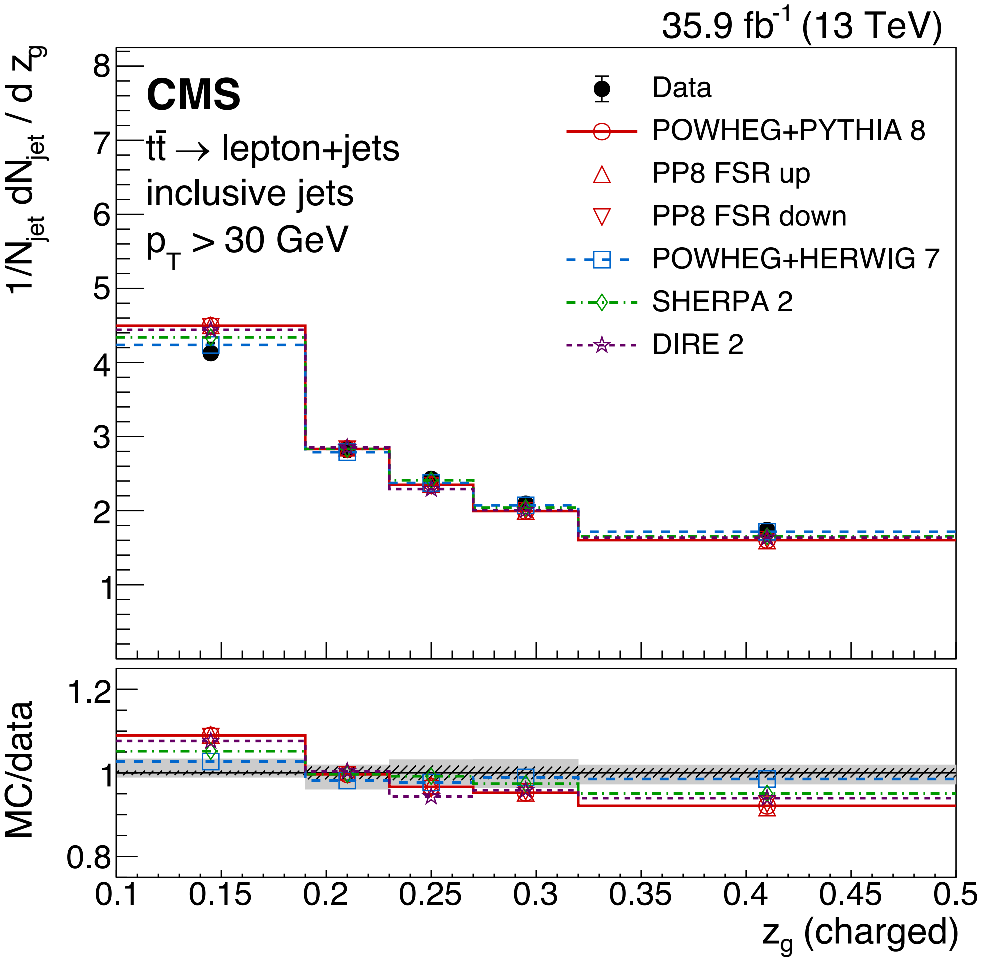


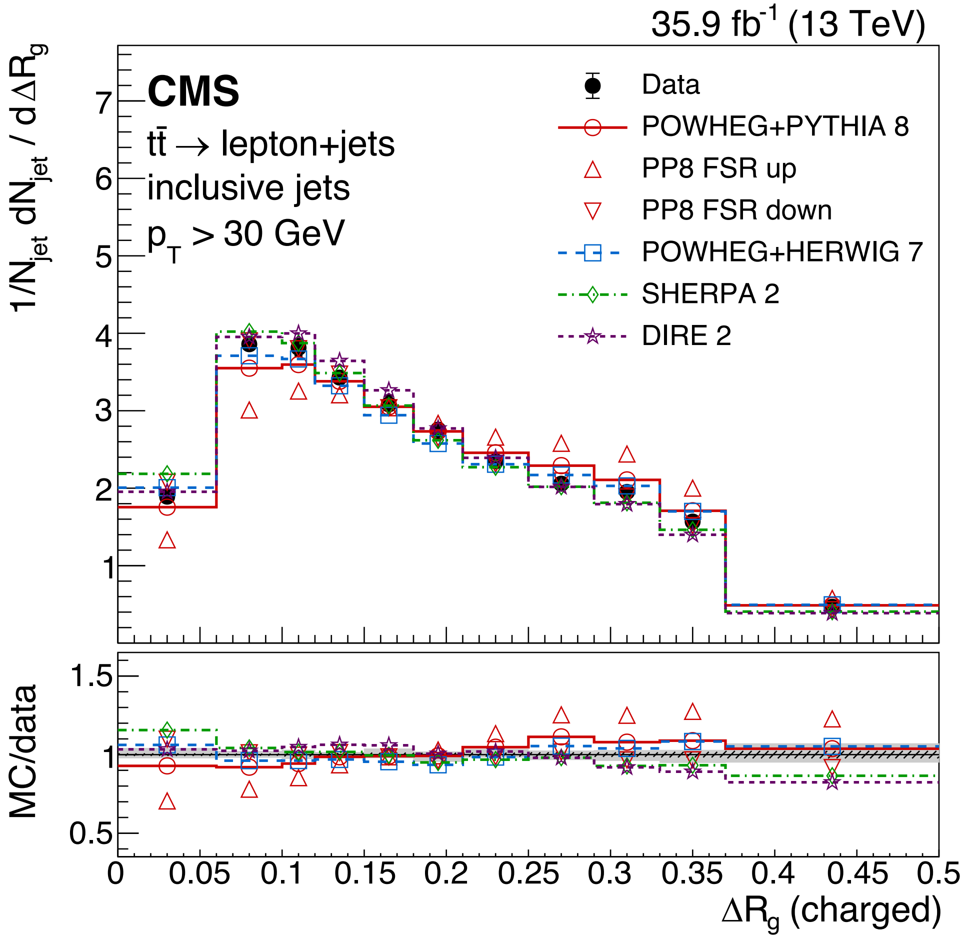
Jet substructure observables

- Measured many jet substructure observables with/without grooming
- Light quark, gluon and b-quark enriched samples from ttbar events

$$z_g = \frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{\mathrm{cut}} \left(\frac{\Delta R_{12}}{R_0} \right)^{\beta}$$

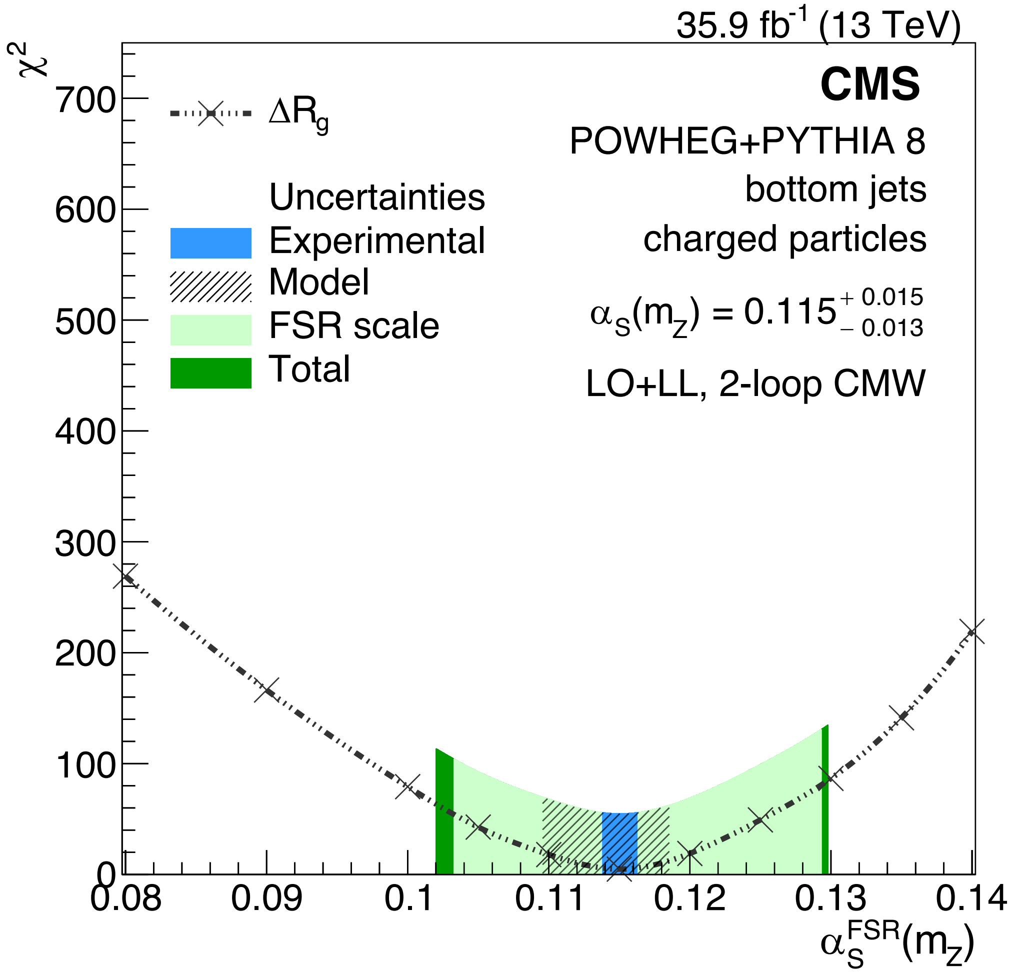
$\Delta\breve{R}_g$





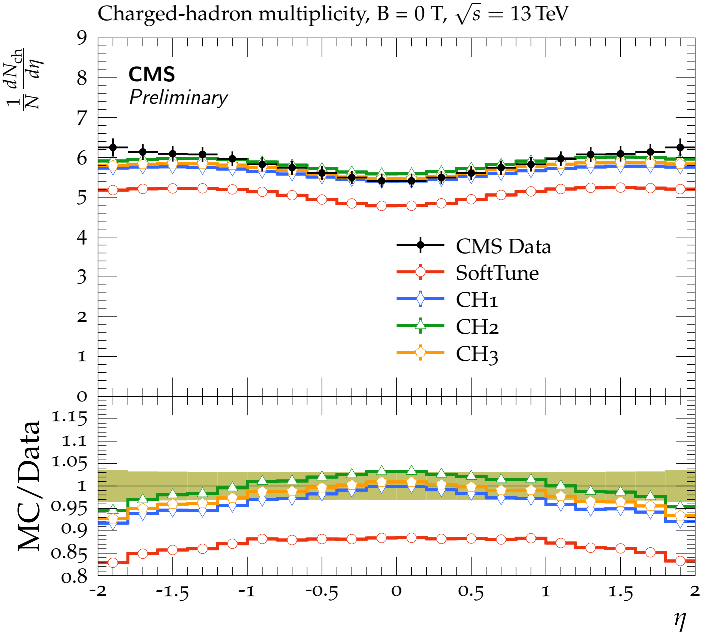
Jet substructure observables - α_S

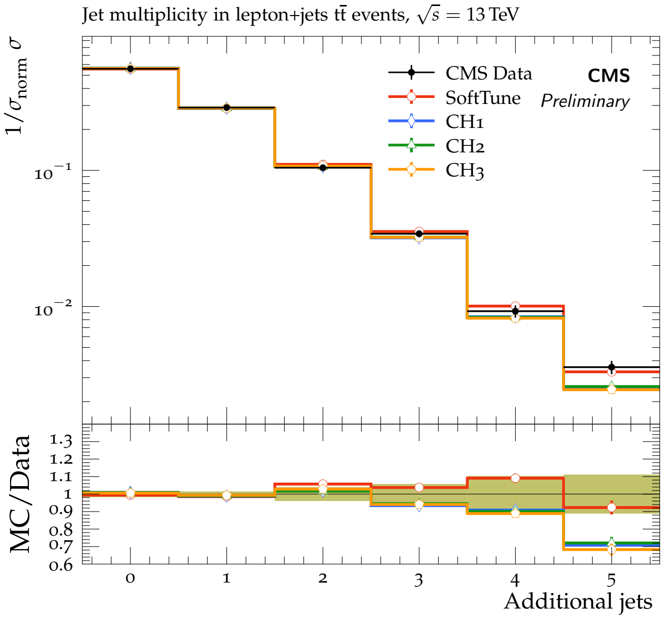
- Measure $\alpha_S(m_Z) = 0.115^{+0.015}_{-0.013}$ from most sensitive observable ΔR_g
- Dominant uncertainty from FSR scale uncertainty of PS prediction

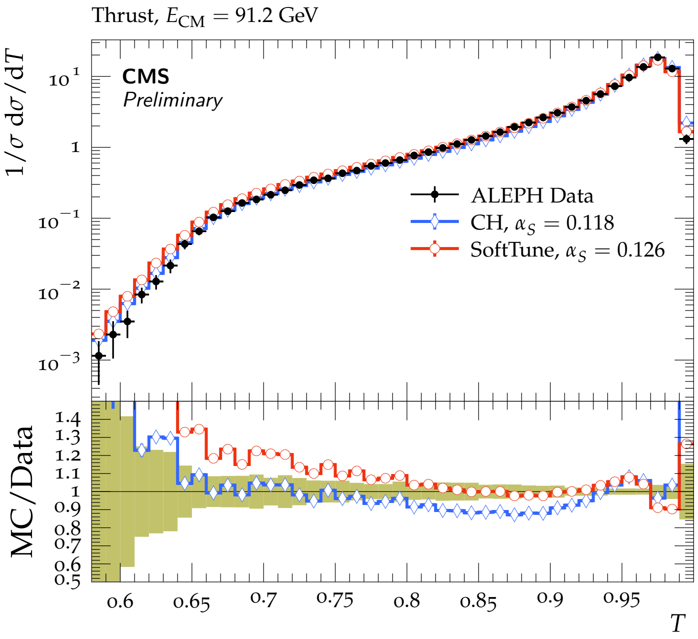


Jet substructure and tuning

- Relation between choice of α_S in PDF, ME, shower, MPI
- CMS generator tuning looks at more than just UE in CMS Pythia8 ”CP” and Herwig7 “CH” tunes





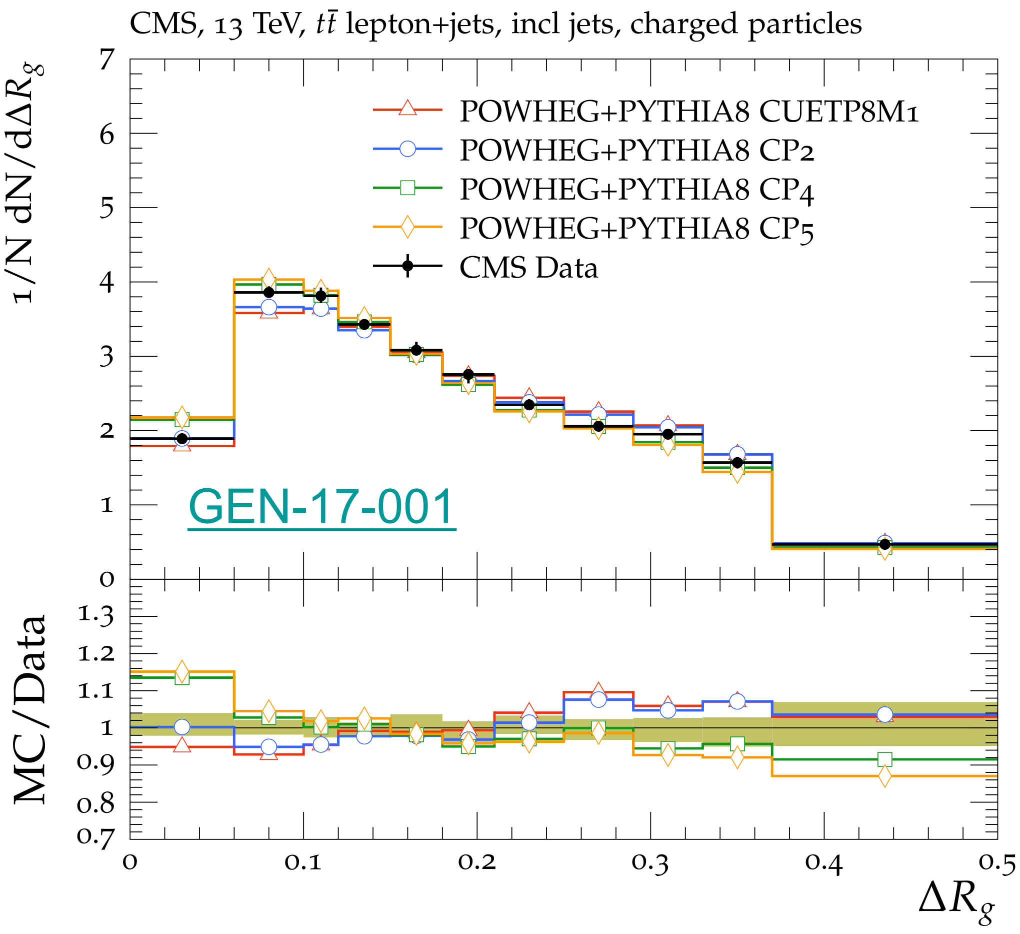


GEN-17-001

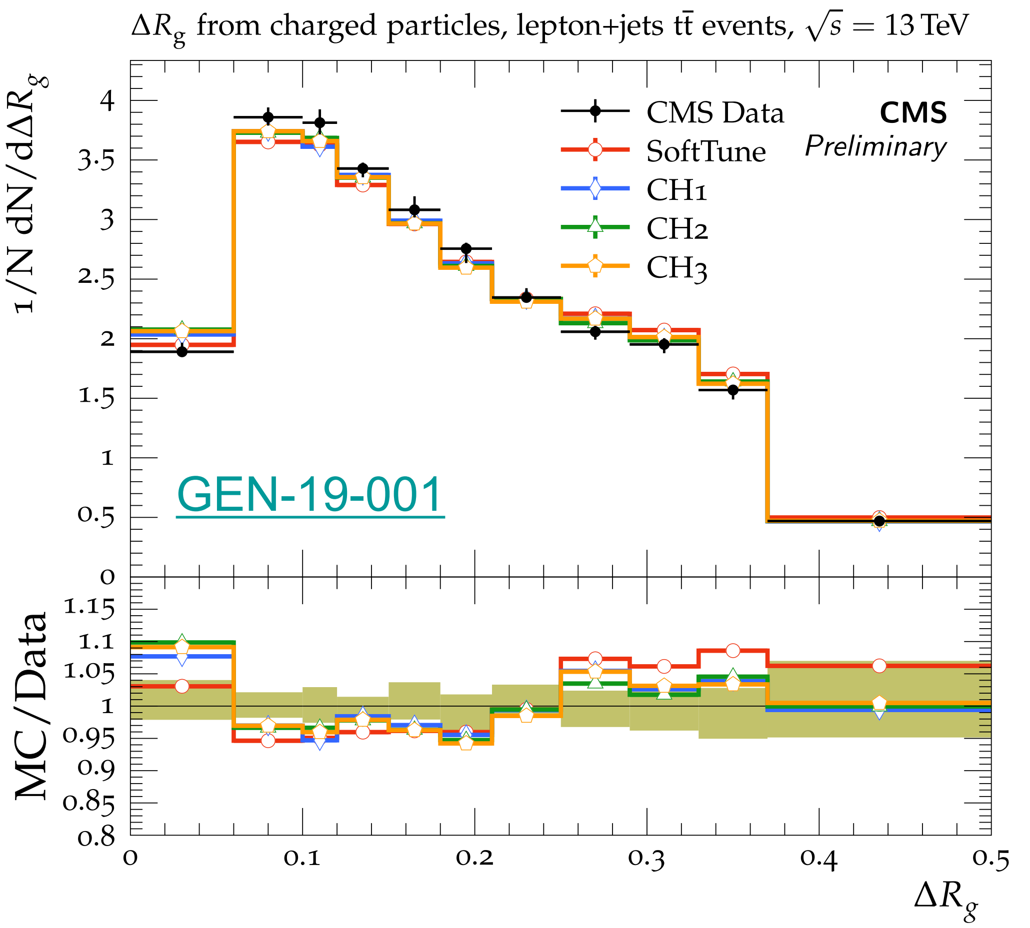
GEN-19-001

Jet substructure and tuning

- Jet substructure modeling affected by UE tuning
- CMS checks jet substructure observables for new tunes (ATLAS A14 tune includes them in the tuning itself)


CMS, 13 TeV, $t\bar{t}$ lepton+jets, incl jets, charged particles

GEN-17-001

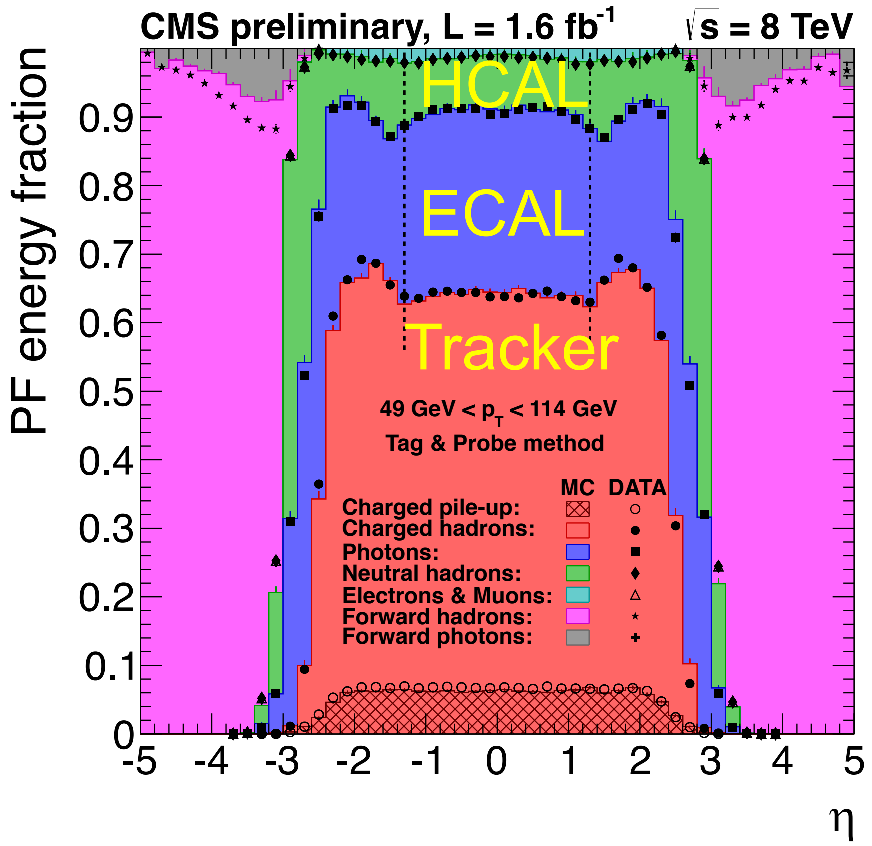

ΔR_g from charged particles, lepton+jets $t\bar{t}$ events, $\sqrt{s} = 13$ TeV

GEN-19-001

Jet substructure reconstruction

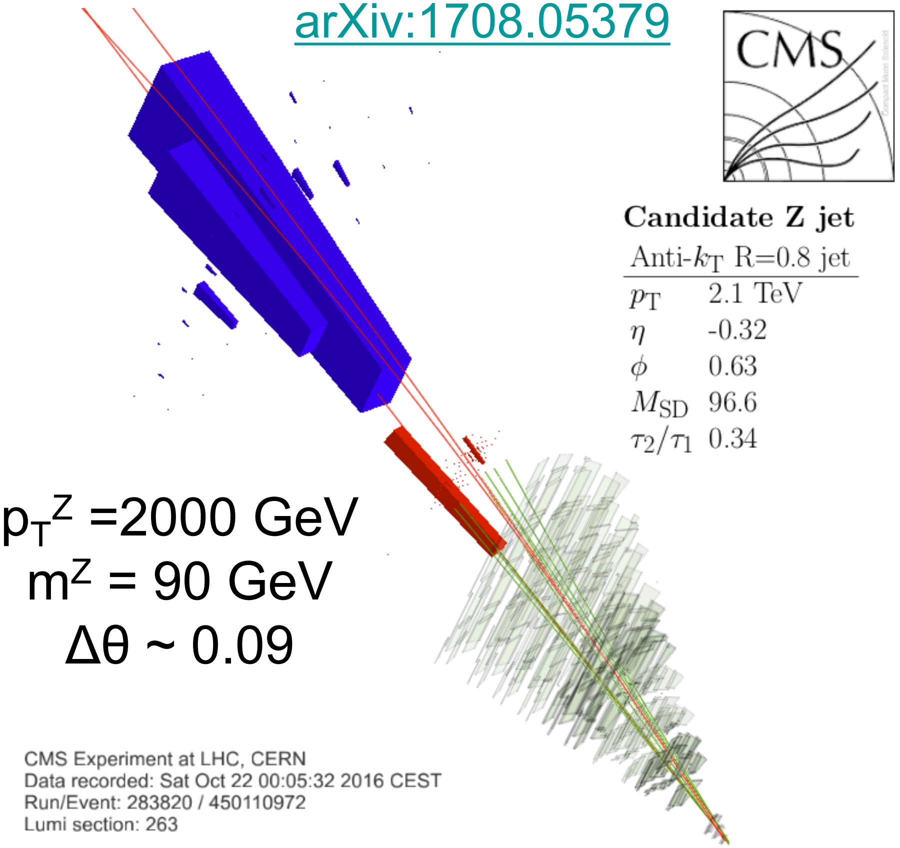
Detector	p_T-resolution	η/Φ-segmentation
Tracker	0.6% (0.2 GeV) – 5% (500 GeV)	0.002 x 0.003 (first pixel layer)
ECAL	1% (20 GeV) – 0.4% (500 GeV)	0.017 x 0.017 (|η|<1.48)
HCAL	30% (30 GeV) – 5% (500 GeV)	0.087 x 0.087 (|η|<1.74) 0.175 x 0.175 (|η|>3)

- Particle flow essential for substructure reconstruction



CMS-DP-2012/012

arXiv:1708.05379



Candidate Z jet

Anti-k_T R=0.8 jet	
p_T	2.1 TeV
η	-0.32
ϕ	0.63
M_{SD}	96.6
τ_2/τ_1	0.34

p_T^Z =2000 GeV

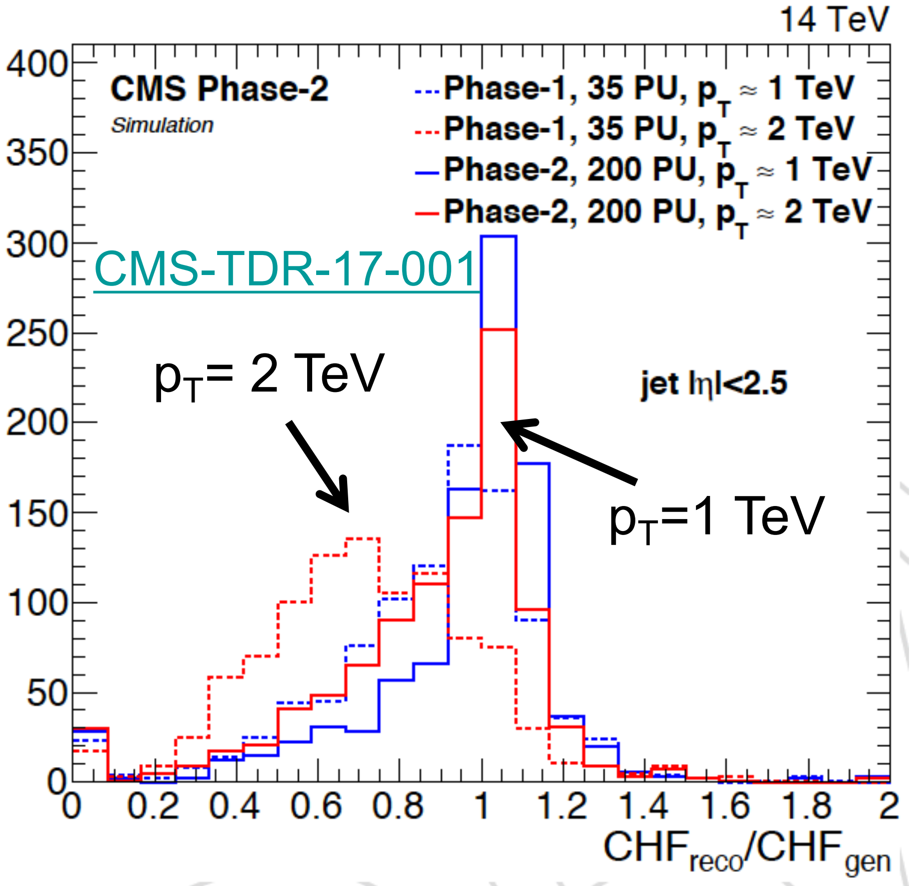
m^Z = 90 GeV

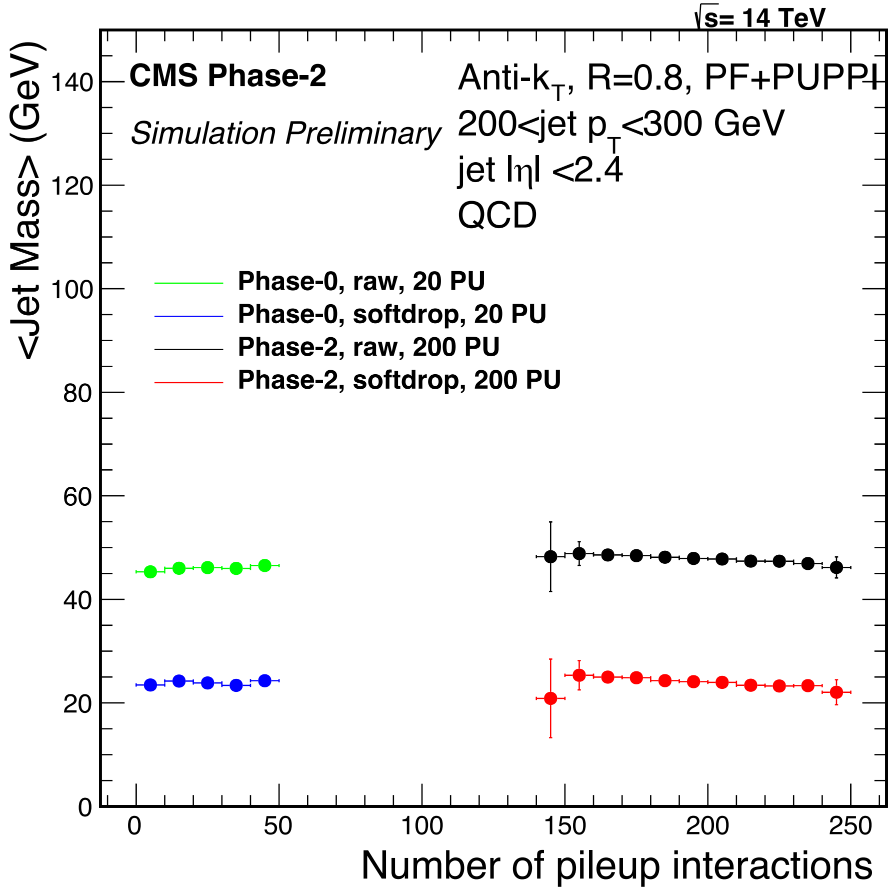
Δθ ~ 0.09

CMS Experiment at LHC, CERN
Data recorded: Sat Oct 22 00:05:32 2016 CEST
Run/Event: 283820 / 450110972
Lumi section: 263

Jet substructure at HL-LHC

- Jet substructure was not design goal in CMS TDR, but works well
- For CMS Phase-2 upgrade for HL-LHC considered in TDRs
- More granular tracker, highly granular endcap calorimeter
- Better substructure performance despite 200 PU





Fraction of energy reconstructed by tracker

Conclusions

- Jet substructure observables are widely used across the physics program of CMS
    - Calibration procedures in place, though systematic uncertainties dominated by theoretical understanding/modeling

- So far CMS has 16 measurements of jet substructure observables in pp/PbPb

- Proven useful for
    - Understanding of jet quenching in PbPb
    - Test of state-of-the art calculations
    - Estimation of α_S
    - Check of parton shower tunes

- A potential area of collaboration in LHC EWK WG
    - Choice of α_S and parton shower predictions/tunes good for substructure