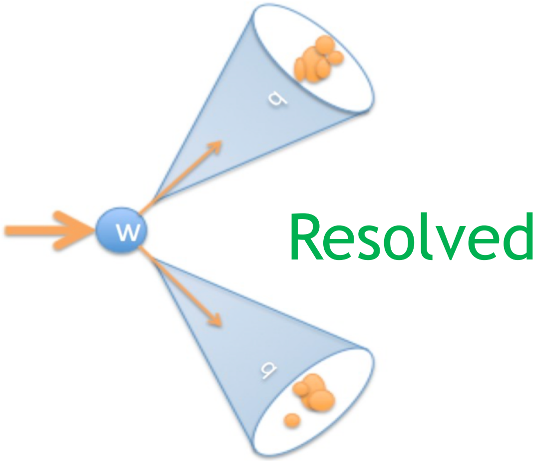
Jet substructure measurements at ATLAS and CMS

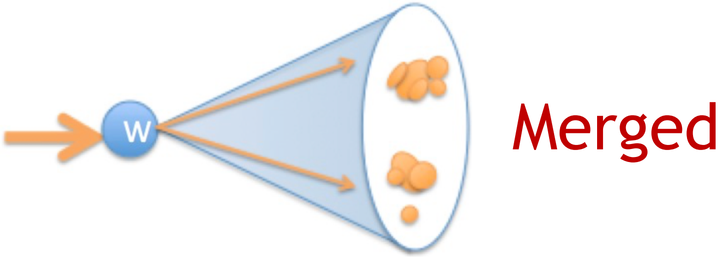
Aparajita Dattagupta (University of Oregon)
On behalf of ATLAS and CMS

SM@LHC
2-5 May 2017

Introduction

- With increased energies of the LHC collisions, advancement on the theory side and in experimental methods, we are able to treat a jet as more than a four-vector substitute for a parton and understand the internal structure of the jet.
- Jet substructure techniques can be used to improve sensitivity of measuring hadronic decays of boosted heavy particles like Higgs, W/Z and top. Requires understanding of QCD radiation inside the jet and structure of constituent particles.
- Large-R jets of usually R=1 or 0.8 are used to reconstruct these boosted objects,
  - More prone to pileup effects due to a larger area.
  - Jet mass of these large-R jets important to measure well.
- Quark-gluon tagging: example of substructure techniques useful in signal-background discrimination.
- W/Z/H- and top tagging can be used to separate their hadronic final states from multi-jet background.



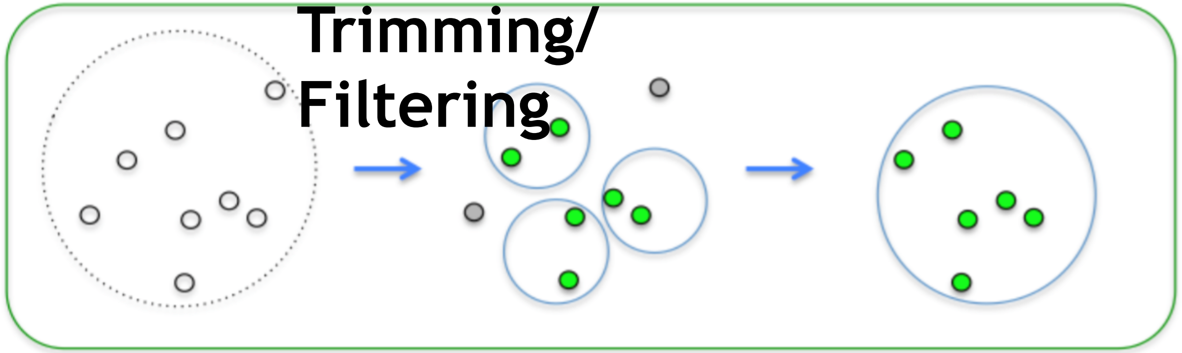


Jet grooming techniques

ATL-PHYS-PUB-2014-004

- Remove soft component targeting Pileup (PU) and Underlying event (UE)
- Better signal-background separation
- Improves resolution of signal jet mass peak

Trimming/ Filtering



Type 1 (Trimming) : If p_T (subjet i) / p_T (jet) < f_{cut} : discard subjet.

Type 2 : If $N_{subjets} \leq N_{min}$: discard jet.

Resulting jet is sum of subjets.

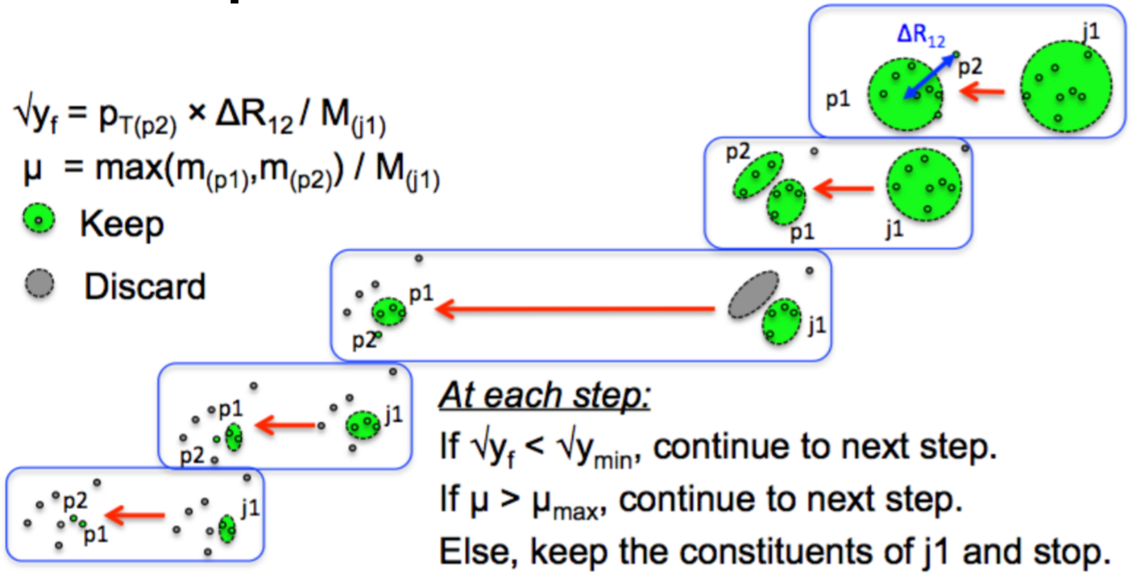
Mass drop/momentum

$\sqrt{y_f} = p_{T(p2)} \times \Delta R_{12} / M_{(j1)}$

$\mu = \max(m_{(p1)}, m_{(p2)}) / M_{(j1)}$

🟢 Keep

● Discard



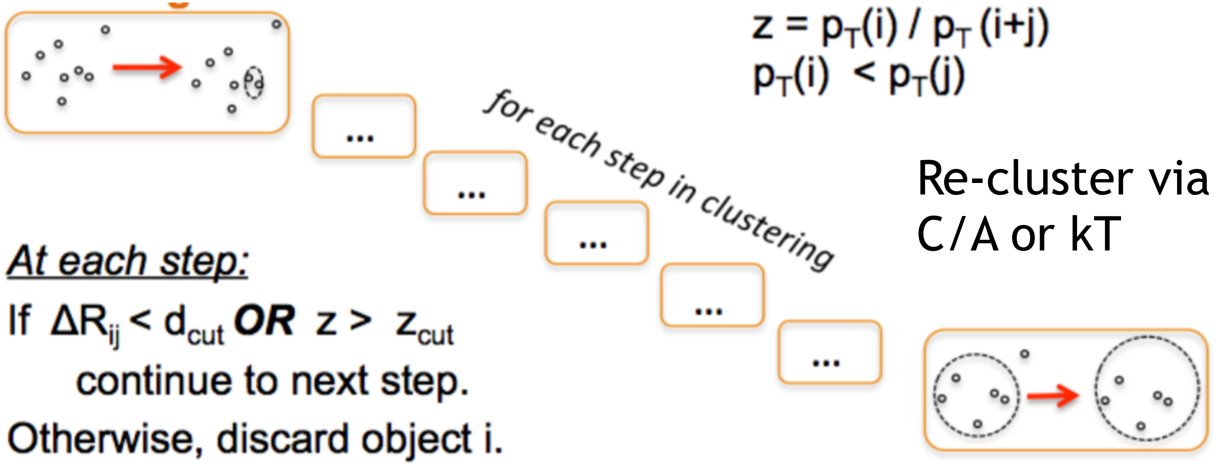
__At each step:__

If $\sqrt{y_f} < \sqrt{y_{min}}$, continue to next step.

If $\mu > \mu_{max}$, continue to next step.

Else, keep the constituents of j1 and stop.

Pruning



$z = p_T(i) / p_T(i+j)$

$p_T(i) < p_T(j)$

Re-cluster via C/A or kT

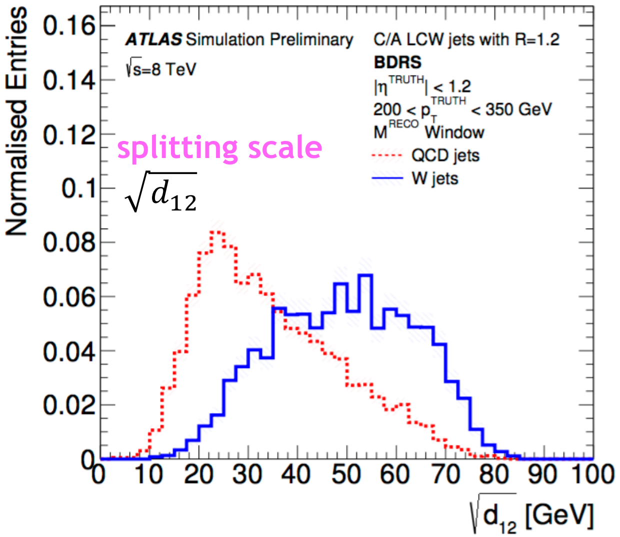
__At each step:__

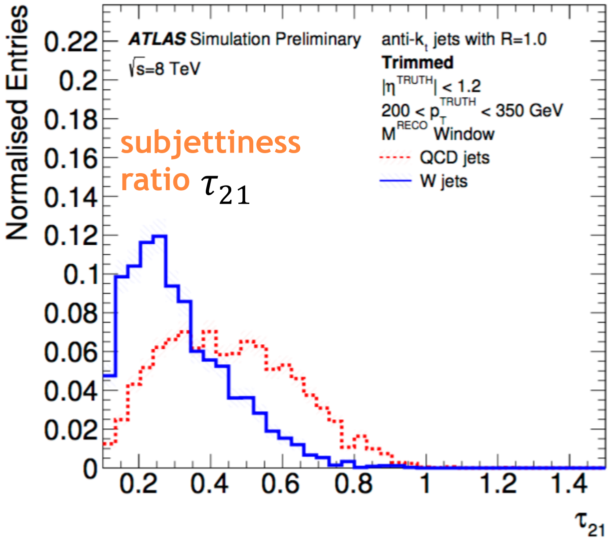
If $\Delta R_{ij} < d_{cut}$ ***OR*** $z > z_{cut}$

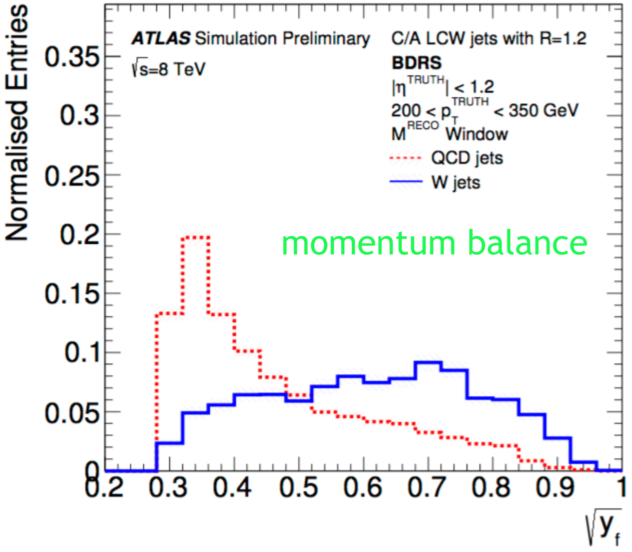
continue to next step.

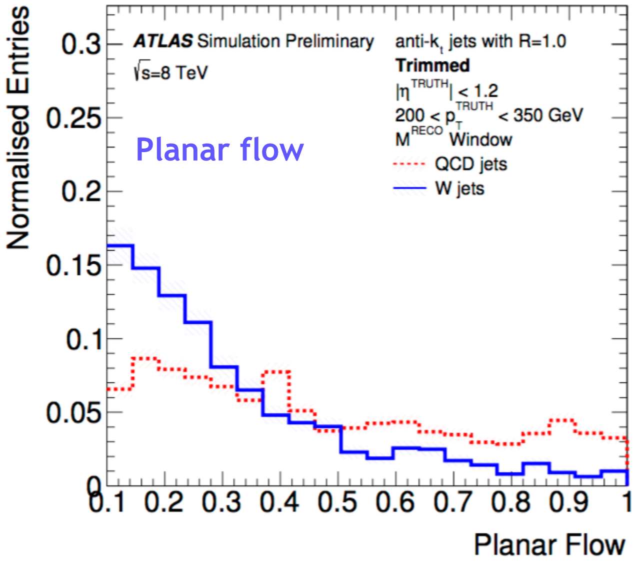
Otherwise, discard object i.

Substructure variables for boosted object tagging









- **Splitting scale**: for a jet re-clustered with the kT algorithm-kT distance between two proto jets in the final clustering step
- **N-subjettiness**: extent to which the substructure of a jet resembles N or fewer subjets.
- **Momentum balance**: ratio of splitting scale and jet mass of the final clustering step.
- **Mass-drop**: last step of recombination when two proto jets are combined into one. Fraction of mass carried by the most massive proto jet.
- **Jet width**: radial moment in η, ϕ space weighted by p_T.
- **Planar flow**: how uniformly spread out the jet energy is perpendicular to it's axis.

ATL-PHYS-PUB-2014-004

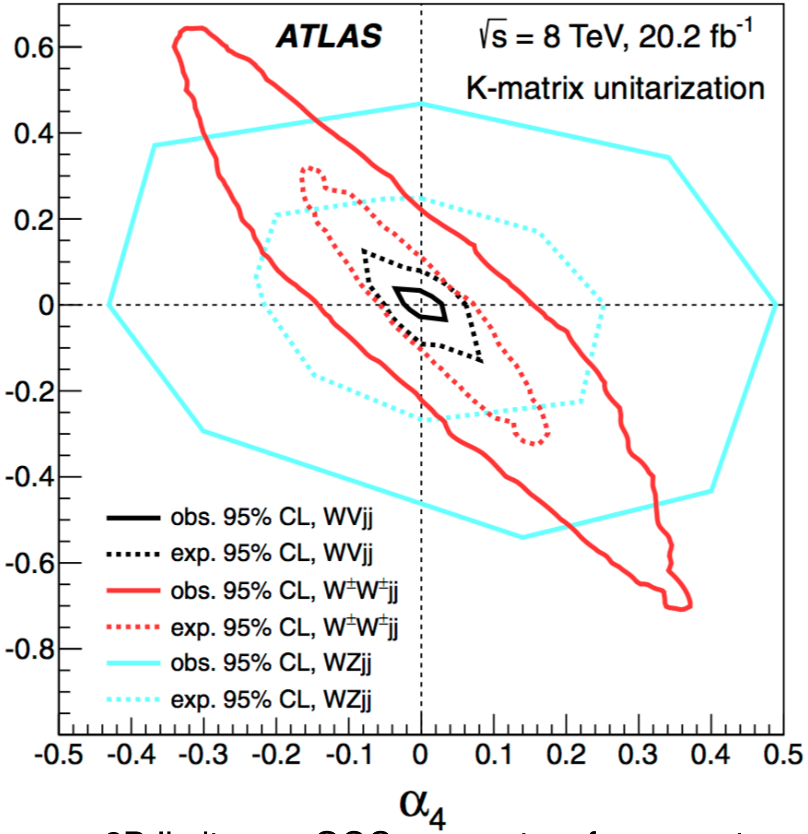
Jet substructure in boosted WV->lvqq search: ATLAS

VBS_ATLASLink

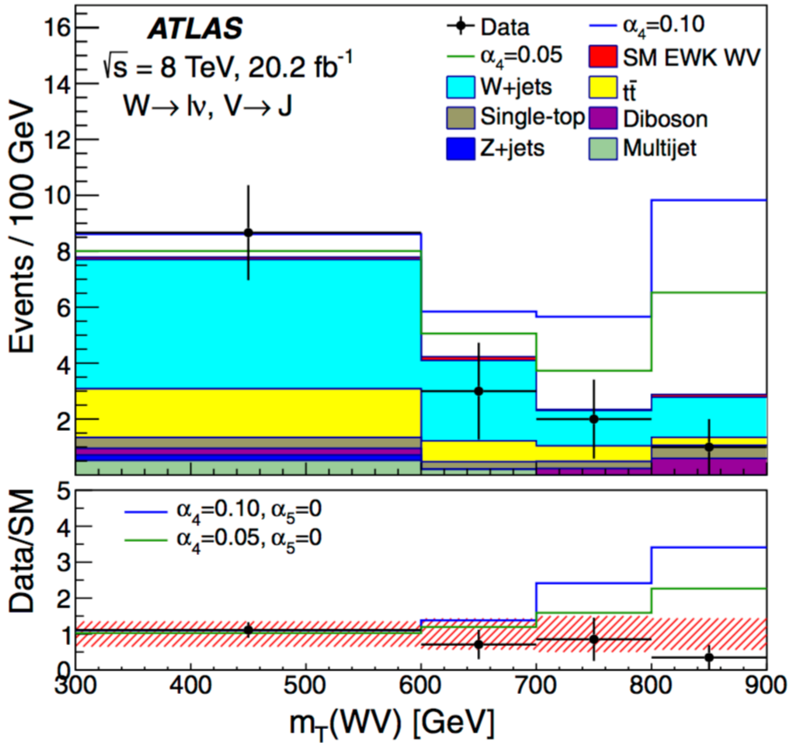
- Vector boson scattering (VBS)
  - key probe of EW symmetry breaking
  - Sensitive to anomalous quartic gauge couplings
- Previous searches for aQGCs in VBS have focussed on leptonic decays: smaller branching fraction
- Mass-drop filtered jets are used. For multiple large-R jets, the one with mass closest to the nominal W mass is used.

Merged regime results

- aQGC limits 40% better than using resolved events only.
- merged category is powerful at highest-Mt (WV) bins.
- More stringent limits than previous constraints on these parameters in searches for VBS in the $W^{+/-}W^{+/-}$ and WZ leptonic decay channels



2D limits on aQGC parameters for current result (black) and previous results (red).

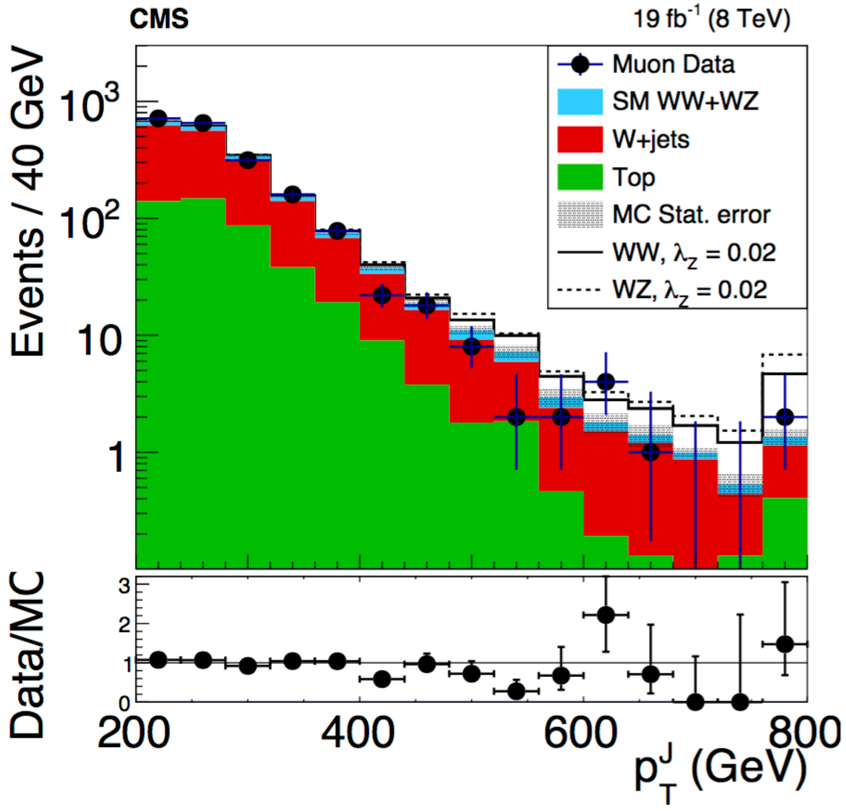


mT (W V) distribution used to extract aQGC limits: in merged regime

Jet substructure in boosted diboson production search: CMS

VBS_CMSLink

- Search for new physics as aTGC in WW and WZ diboson production at $\sqrt{s}$ =8 TeV.
- Events are selected with one W boson decaying leptonically and a W or Z boson whose decay products are merged into a single reconstructed jet.
- Higher cross section than a leptonic final state but larger backgrounds from W+Jets process
- Pruned CA R=0.8 jets are used to reconstruct the boosted W boson while the Z boson decays to invisible products
- N-subjettiness observables $\tau_{21} = \tau_2/\tau_1$ is used to tag the W/Z-jet.



V(had) p_T distribution shown: effects of aTGCs are shown by the solid and dotted lines

	c_{WWW}/Λ^2 (TeV^{-2})	c_B/Λ^2 (TeV^{-2})	c_W/Λ^2 (TeV^{-2})
*	$[-2.7, 2.7]$	$[-14, 17]$	$[-2.0, 5.7]$
CMS W+W-(lep)	$[-5.7, 5.9]$	$[-29.2, 23.9]$	$[-11.4, 5.4]$
ATLAS W+W-(lep)	$[-4.61, 4.60]$	$[-20.9, 26.3]$	$[-5.87, 10.54]$
[43]	$[-4.6, 4.2]$	$[-260, 210]$	$[-4.2, 8.0]$
[44]	$[-3.9, 4.0]$	$[-320, 210]$	$[-4.3, 6.8]$

Summary of one-dimensional limits in the EFT formulation for this analysis (*) compared to previous results.

Jet mass: ATLAS

ATLAS-CONF-2016-035

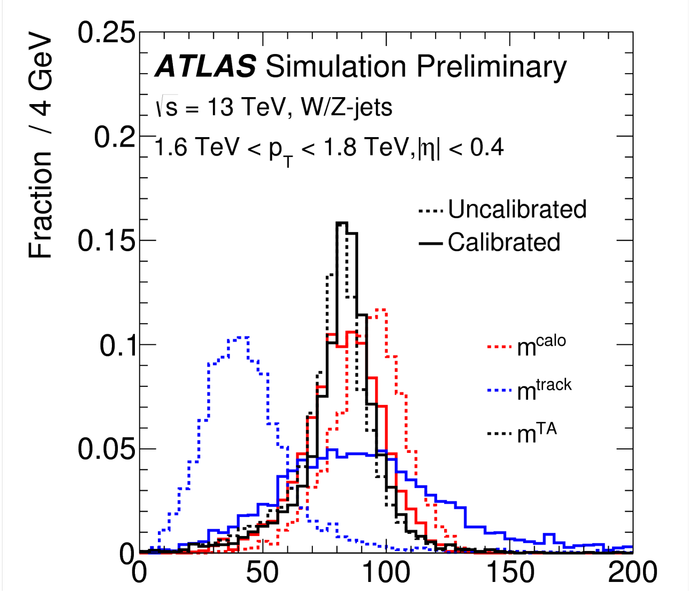
- Jet mass of large-R jets useful for tagging Lorentz boosted W/Z and top quarks and rejecting multi jet background.
- Reduce resolution and systematics to improve jet mass.
- Calo-based jet mass definition :

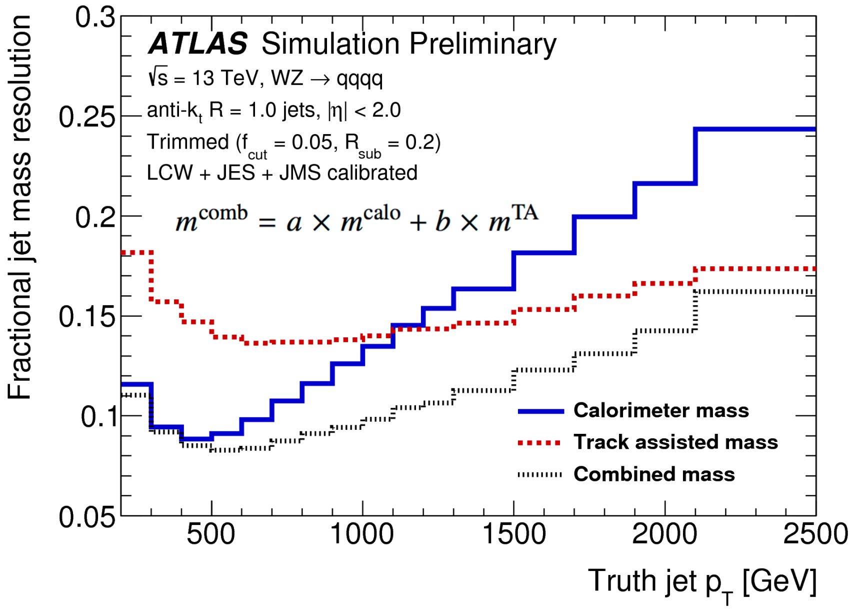
$$m^{\mathrm{calo}} = \sqrt{\left(\sum_{i\in J} E_i\right)^2 - \left(\sum_{i\in J} \vec{p}_i\right)^2}.$$

- Track assisted jet mass: complements m^{calo}. At high lorentz boost, spread of decay products of heavy particles are of the order of calorimeter granularity

$$m^{\mathrm{TA}} = \frac{p_{\mathrm{T}}^{\mathrm{calo}}}{p_{\mathrm{T}}^{\mathrm{track}}} \times m^{\mathrm{track}},$$

- Top jets: m^{calo} has better resolution across all p_T ranges
- W/Z jets: m^{TA} has better resolution at low p_T and worse resolution than m^{calo} at high p_T .
- Reconstructed jet mass is corrected to particle level using calibration factors derived from simulated QCD multi jet events.
- Combined mass $m^{\mathrm{comb}} = a \times m^{\mathrm{calo}} + b \times m^{\mathrm{TA}}$



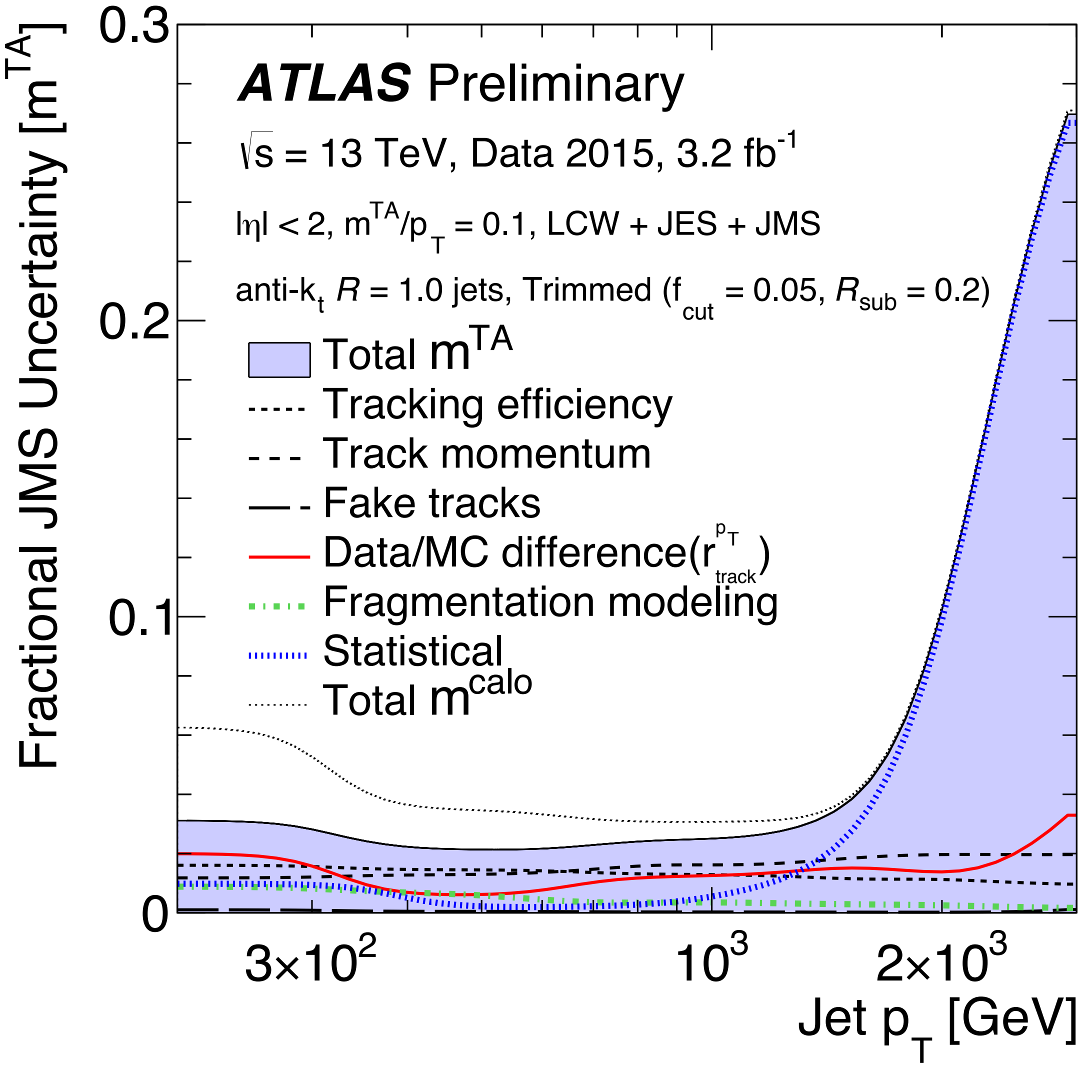


Jet mass systematic uncertainties: ATLAS

ATLAS-CONF-2016-035

- Uncertainty is ~4% for m^{calo} and 2% for m^{track} for 300 GeV < p_T < 1 TeV.
- tracking uncertainties for m^{TA} through the r_{track} method are smaller than for m^{calo} since large extent of the uncertainties cancel in the ratio m^{track}/p_T^{track}
    - At high p_T, uncertainties for m^{calo} and m^{track} are limited by the size of the dataset used to study the modeling of r_{track} .
- m^{comb} uncertainties propagated through

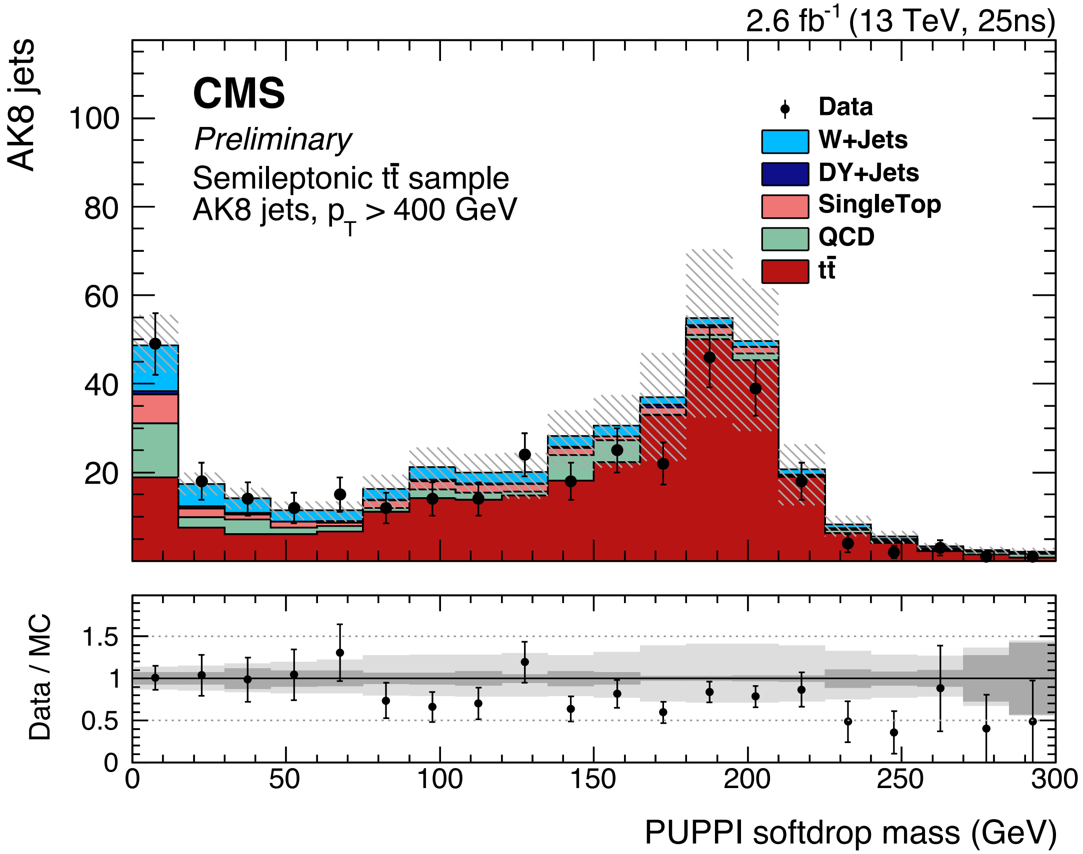
$$m^{\mathrm{comb}} = a \times m^{\mathrm{calo}} + b \times m^{\mathrm{TA}}$$



components of the track-assisted jet mass scale uncertainties.

Jet mass after pileup removal: CMS

PUPPI JME-16-003

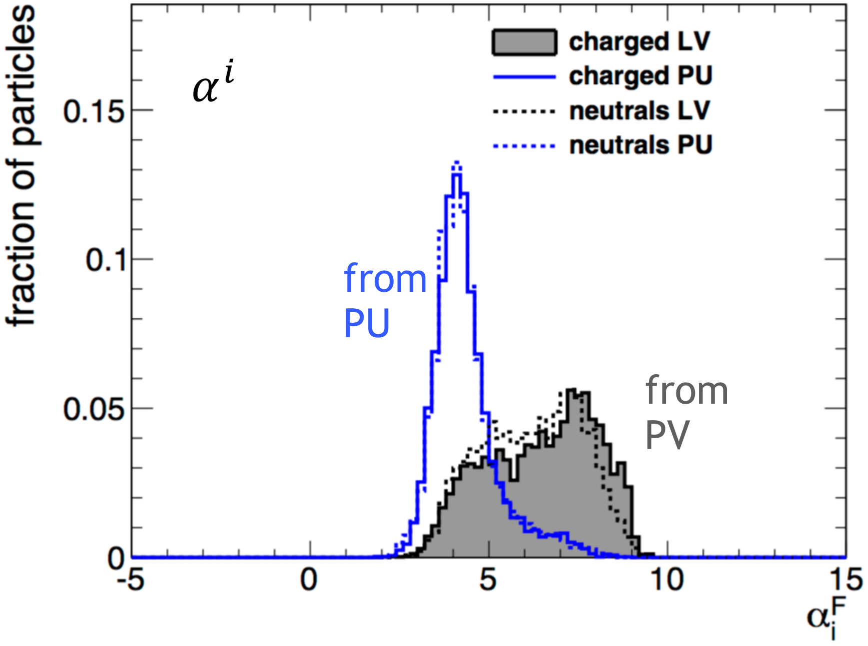


- PUPPI or pileup (PU) per particle identification, reduces PU effects on jet observables: uses local shape information, event pileup properties and tracking information.
    - before any jet clustering is performed.

- PUPPI defines α for each particle i using j particles around it

$$\alpha_i = \log \sum_{j\in\text{event}} \frac{P_T^j}{\Delta R_{ij}} \times \Theta(R_{\min} \le \Delta R_{ij} \le R_0),$$

- Transforms distribution of α as a weight
- These per-particle weights are used to rescale the particles four-momenta to correct for pileup, superseding the need for jet-based pileup corrections.

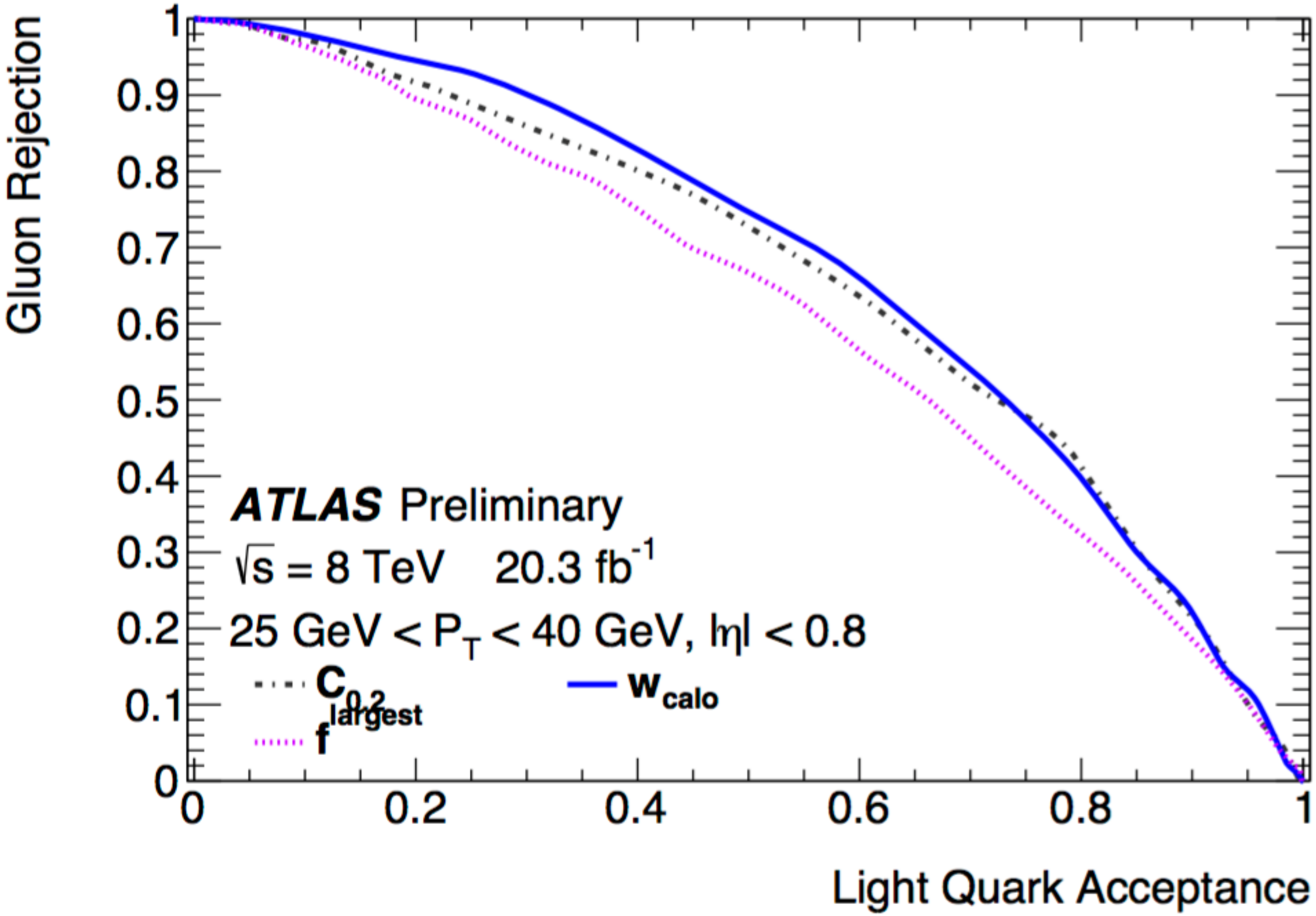


The distribution of α^i, over many events, for particles i from the LV (gray filled) and particles from pileup (blue) in a dijet sample.

Quark-gluon tagging ATLAS

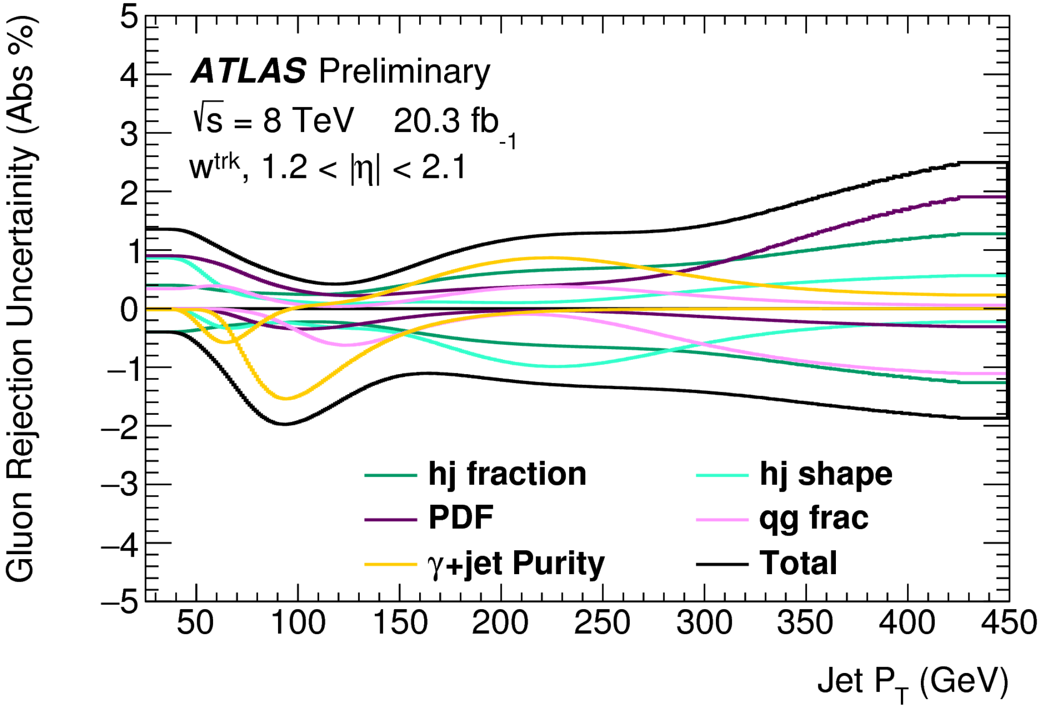
ATLAS-CONF-2016-034

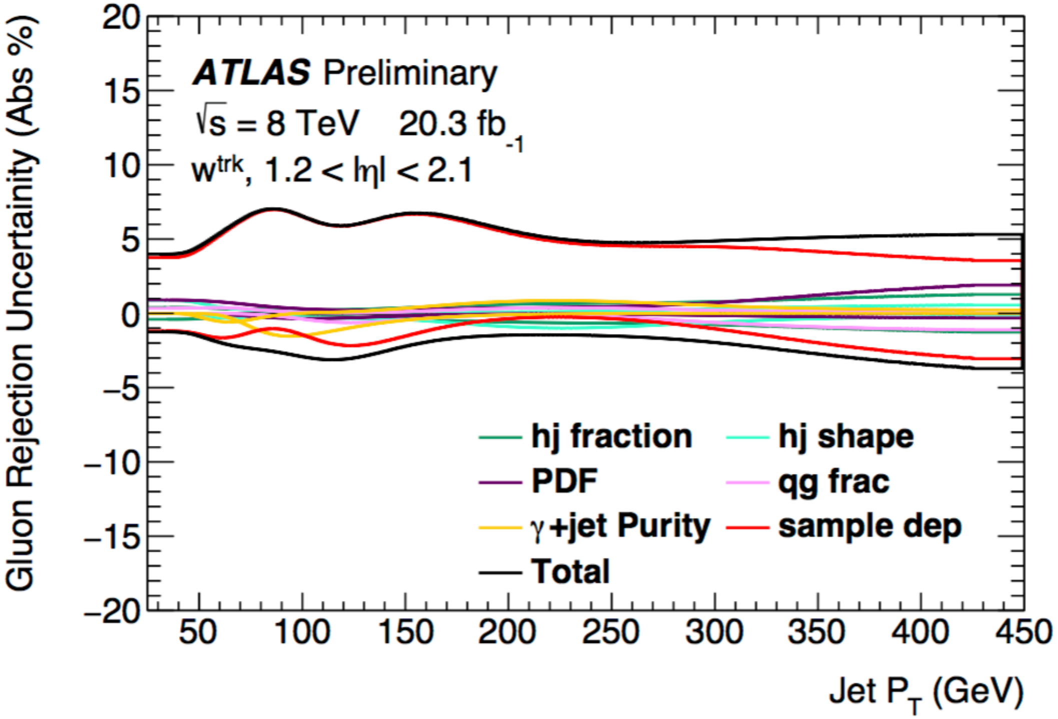
- Important handle in several physics analyses
  - Higgs VBF produced in association with mainly light-quark initiated jets. QCD multi jet background is mainly composed of gluon-initiated jets
- Discriminating variables: largest energy constituent energy fraction (f), width of the jet (w) and energy correlation function (C).
- Three-sample extraction done using Dijet and gamma+jet or Dijet and Z+jet events depending on statistics in the p_T and η bin.



uncertainties on gluon rejection and quark acceptance

ATLAS-CONF-2016-034



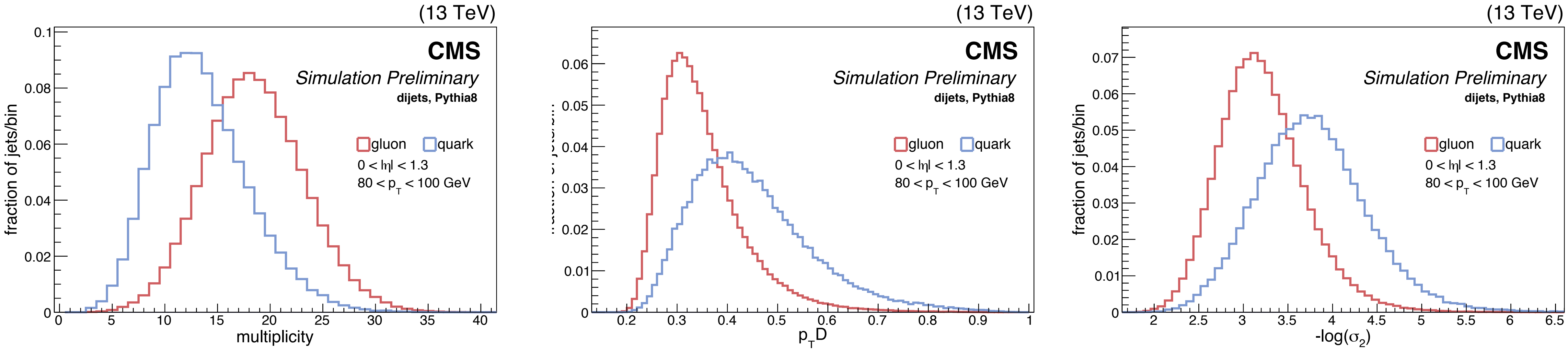


Uncertainty sources

- Statistical uncertainty: ~1% for jets with |η| < 2.1, and 5% at larger η .
- Uncertainty on light-quark like, gluon-like jet fraction: Herwig++ and Pythia differences , fractions derived from PDFs, PDF reweighing is used for uncertainty estimation. Uncertainty on gluon rejection is ~2%, negligible for Gamma+jet sample.
- Uncertainty on c/b jet fractions: 20% on Dijet sample and 50% in Gamma+jet sample, uncertainty is ~2%
- c-quark and b-quark contributions: 1% uncertainty on rejection and acceptance
- Gamma+jet, Z+jet purity : vary photon identification criteria, 1% uncertainty at low p_T. Z+jet uncertainty is 0.2%
- Largest uncertainty comes from samples dependence (right plot).

Quark-gluon tagging CMS

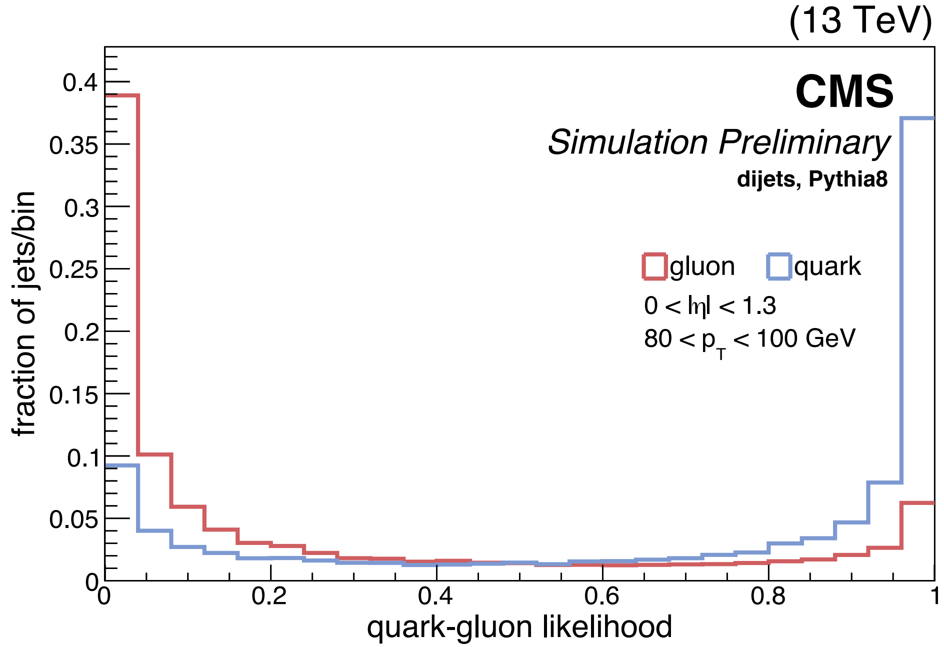
JME-16-003



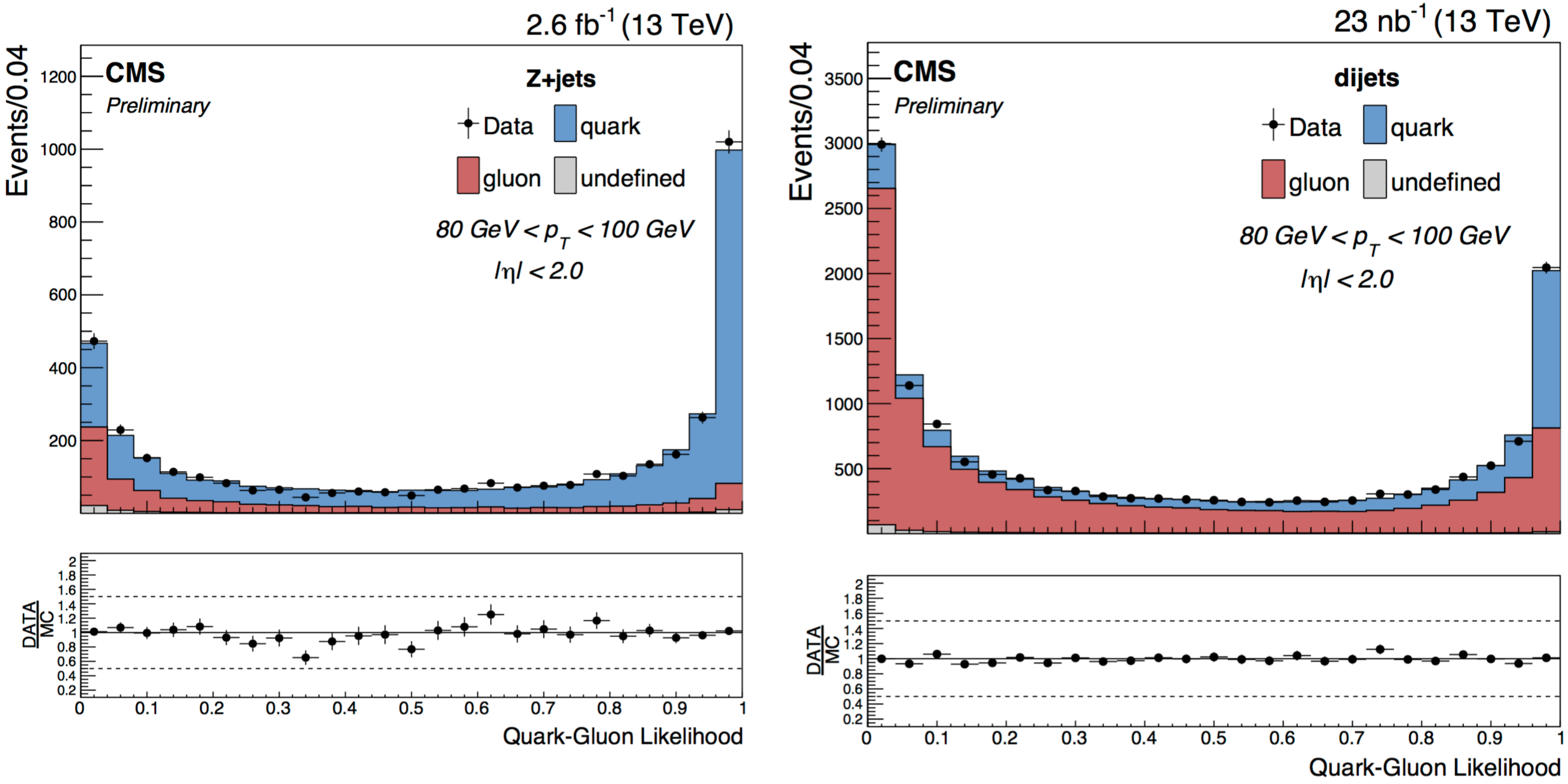
- Quark-gluon discriminator is a likelihood built from the product of the probability density functions of three variables:
  - jet constituent multiplicity
  - jet fragmentation distribution, range between 0-1, higher value for quark jets.

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

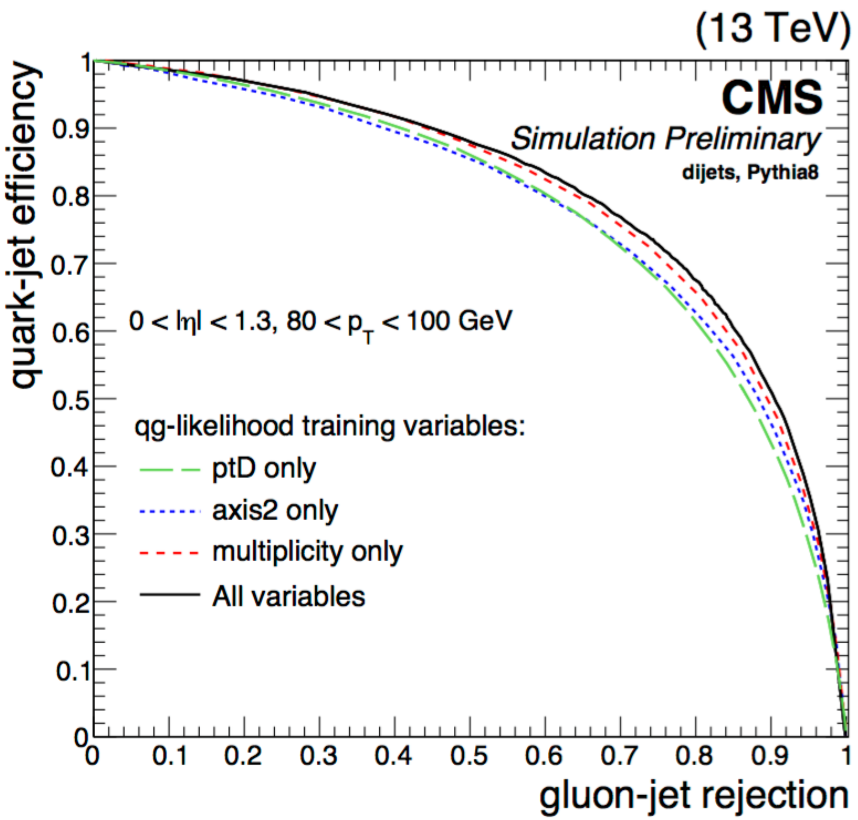
  - jet minor angular opening (sigma2), useful for jets with lower p_T where gluon jets are wider than quark jets.



Quark-gluon tagging CMS: validation in data JME-16-003



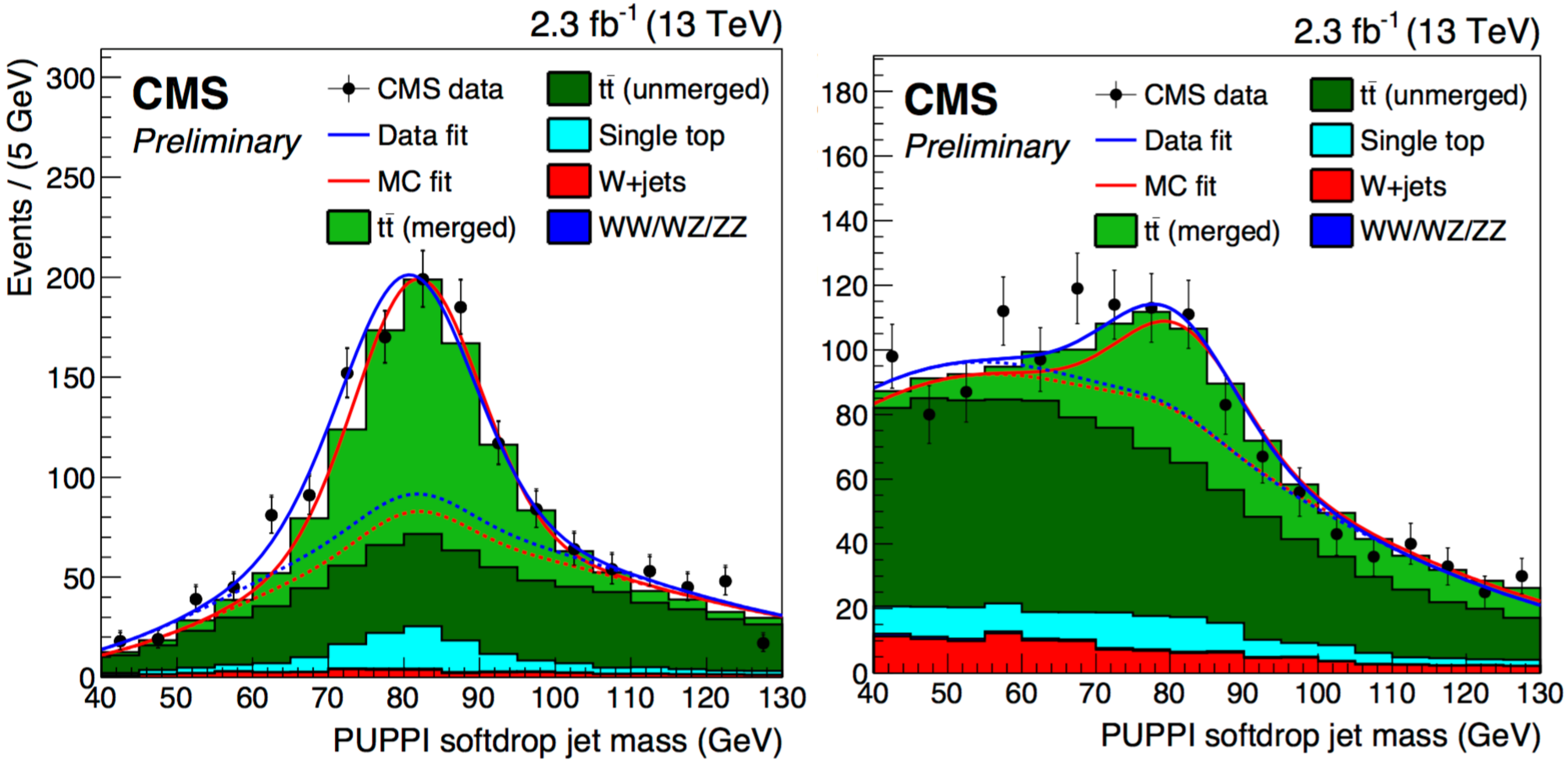
Data/MC comparison for Quark-gluon likelihood in Z+jet events (left) and Dijet events (right) after the data-driven systematics reshaping procedure.



- After the reshaping procedure, difference in signal efficiencies predicted by Herwig and Pythia are at the percent-level, before the procedure the differences are sizable and more pronounced for gluon jets.
- Z+jet: higher fraction of quark jets
- Dijet: Higher fraction of gluon jets

W tagging inputs: CMS

JME-16-003



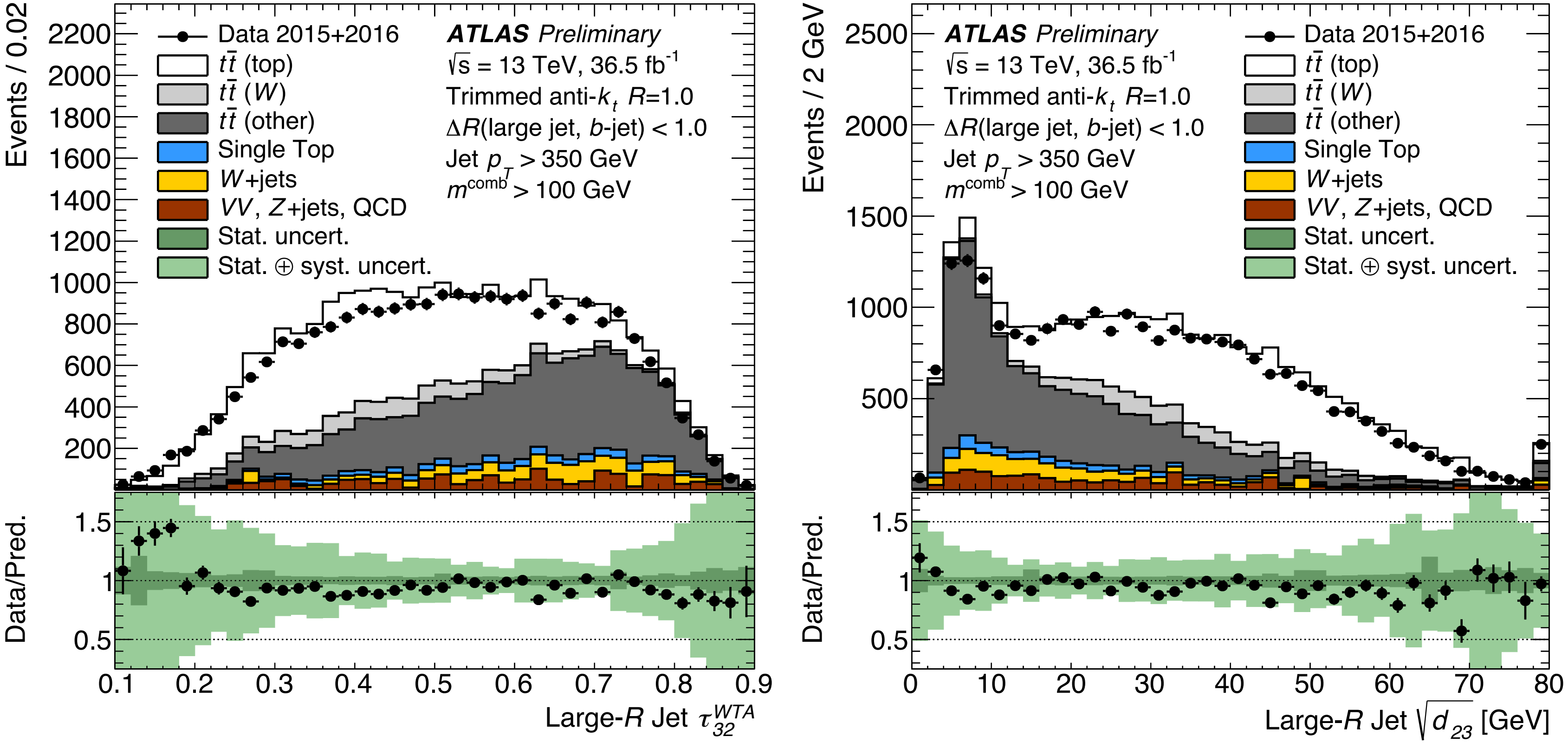
- Pruned jet mass events passing(left) and failing(right) PF+PUPPI τ_{21}< 0.40 selection.
- Data-to-simulation scale factors for the W-tagging procedure are extracted from a top enriched data sample and from simulation.
- Systematic uncertainties on the scale factor: $t\bar{t}$ simulation, choice of the signal and background fit model.

Category	Definition	W scale factor
High-purity Pruning	$(\tau_2/\tau_1 < 0.45)$	0.94 ± 0.05 (stat) ± 0.03 (sys) ± 0.003 (sys)
Low-purity Pruning	$(0.45 < \tau_2/\tau_1 < 0.75)$	1.27 ± 0.25 (stat) ± 0.13 (sys) ± 0.008 (sys)
High-purity Pruning	$(\tau_2/\tau_1 < 0.6)$	0.98 ± 0.03 (stat) ± 0.003 (sys) ± 0.02 (sys)
High-purity PUPPI softdrop	$(\tau_2/\tau_1 < 0.4)$	0.97 ± 0.06 (stat) ± 0.04 (sys) ± 0.06 (sys)
Low-purity PUPPI softdrop	$(0.4 < \tau_2/\tau_1 < 0.75)$	1.12 ± 0.24 (stat) ± 0.17 (sys) ± 0.12 (sys)

good agreement between data and MC tagging efficiency

top tagging substructure inputs: ATLAS

JETM-2017-004
JETM-2017-005



Cut on subjettiness ratio τ_{32} and second splitting scale $\sqrt{d_{23}}$

Systematic uncertainties

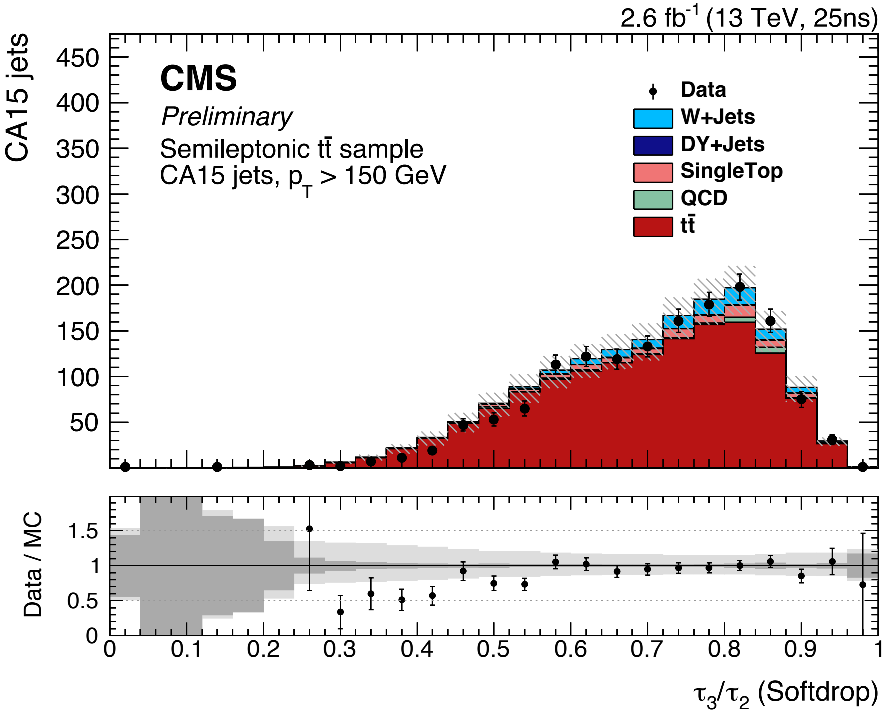
- scale and resolution uncertainties on large-R jet energy, mass and substructure observables
- $t\bar{t}$ modeling uncertainties on parton shower, hadronization model
- Initial-state and final-state radiation.
- Subdominant scale and resolution uncertainties on small-R jet energy and lepton kinematics

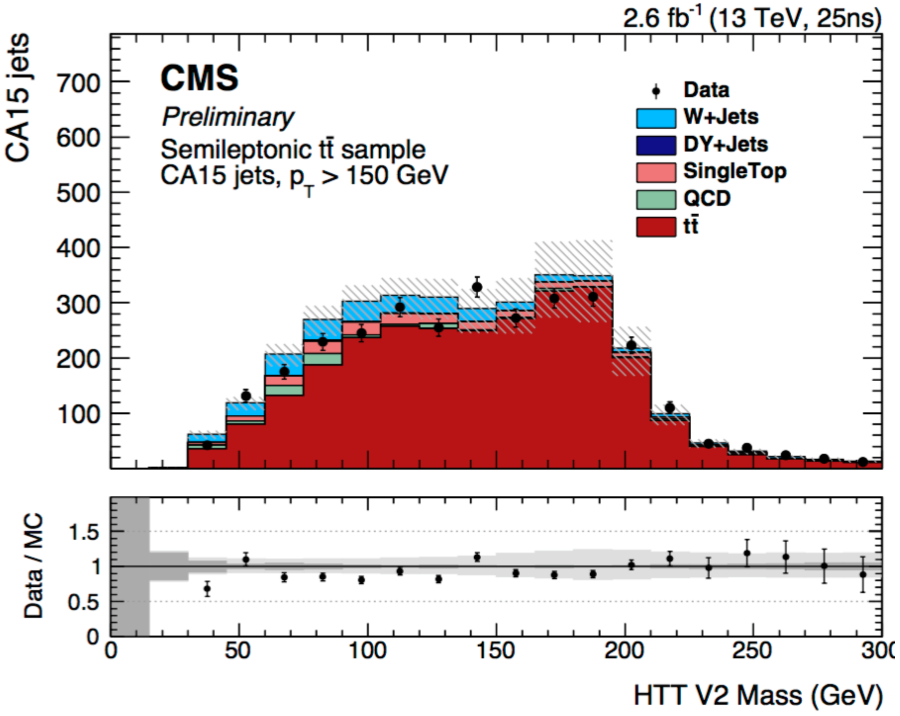
Lepton+jets selection in $t\bar{t}$ events

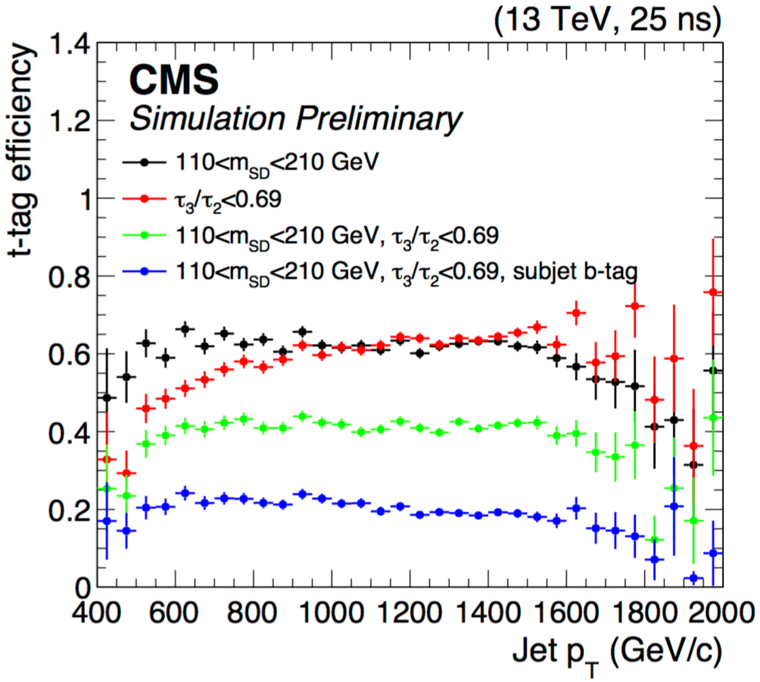
- Subjettiness ratio τ_{32} and second splitting scale $\sqrt{d_{23}}$ in the full 13 TeV dataset and MC simulation for $t\bar{t}$ (left) and Dijet (right) events
- Trimming parameters used: f_{cut} = 5%, R_{subjet} = 0.2
- $t\bar{t}$ sample: “W-matched” part, require the decay products of W boson to be within the large-R jet.
- “top-matched” part, require the qqb particles from top-decay to be within the large-R jet.
- Require R=0.4 b-tagged jet inside the large-R jet cone.

top tagging inputs and efficiency: CMS

JME-16-003





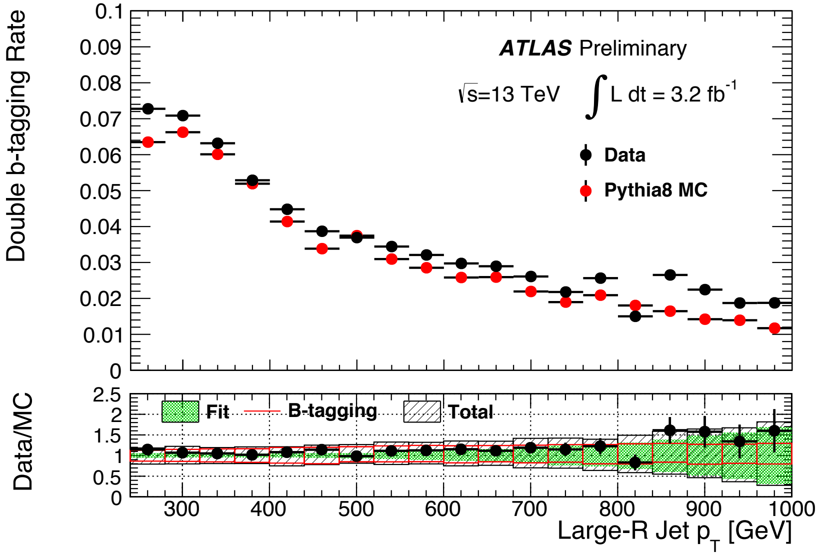


- top tagger: soft drop mass window cut, subjettiness ratio τ_{32} cut and subjet b tagging.
- Distributions are made for AK8 jets after applying the PUPPI procedure.
- tagging techniques relying on PUPPI pileup suppression were found to maintain performance up to at least 40 simultaneous interactions.
- Plot on right shows the signal efficiency measured in simulation for the various tagging variable cuts.

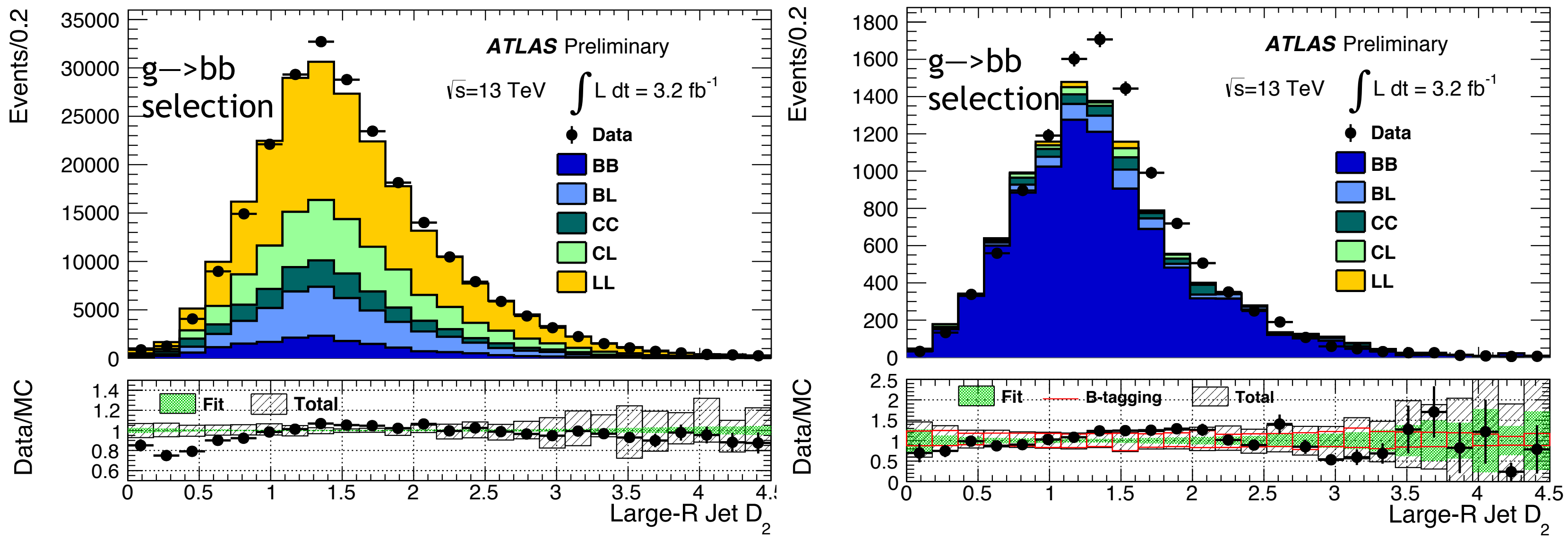
Boosted Higgs (to $b\bar{b}$) boson tagging

HbbLink

- Higgs bosons are reconstructed using anti$-k_t$ R=1 jets and tagged using requirements on
  - Track jet (R=0.2) b tagging: matched to anti$-k_t$ R=1 calorimeter jet.
  - Large-R jet (trimmed) mass window cut around the Higgs boson mass peak
  - Large-R substructure variable: trimmed jet energy correlation ratio $D_2^{(\beta=1)}$
    - Ratio of 2 and 3 point energy correlation functions, based on the energies and pair-wise angular distances of particles within a jet.
    - Optimized to separate between one-prong and two-prong decays.



double b-tagging rate as a function of the large-R jet pt for Data and MC.



Distributions of large-R jet $D_2^{(\beta=1)}$ before (left) and after (right) b-tagging.

uncertainties from template fitting procedure, MC sample stats and b-tagging systematics.

Summary

- Tagging of hadronic decays of particles is becoming more important as we probe particles with higher transverse momentum.
- Use of jet-substructure and tagging is steadily increasing in SUSY, exotics and SM analyses.
  - Using grooming and subjettiness improves aTGC limits in VBS analyses by taking advantage of high branching fractions of hadronic decays of bosons.
- Combined mass for large-R jets takes advantage of better resolution of calo mass at low Pt and track-assisted mass at high pt.
- Quark-gluon taggers have been validated in Run 1 data in ATLAS and in Run 2 data in CMS and systematic uncertainties have been derived.
- W tagging and top tagging techniques have been validated in Run 2 data.
- Substructure and tagging of jets are important areas of ongoing study and development
  - Many methods are being actively studied for W, Z, H and top tagging.

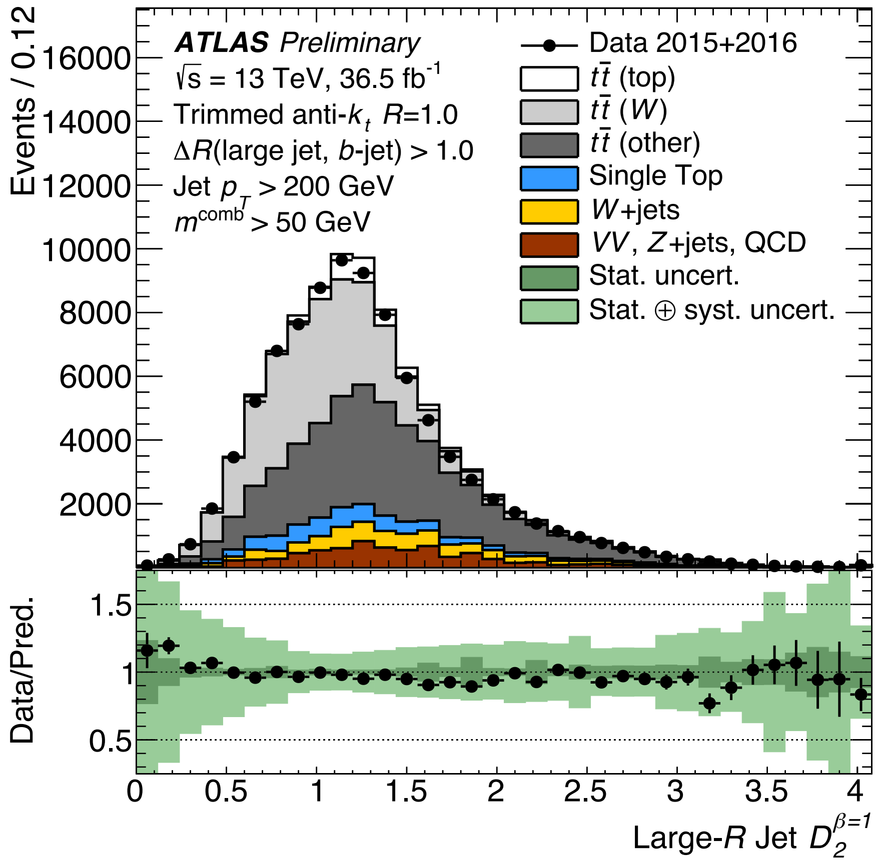
Backup

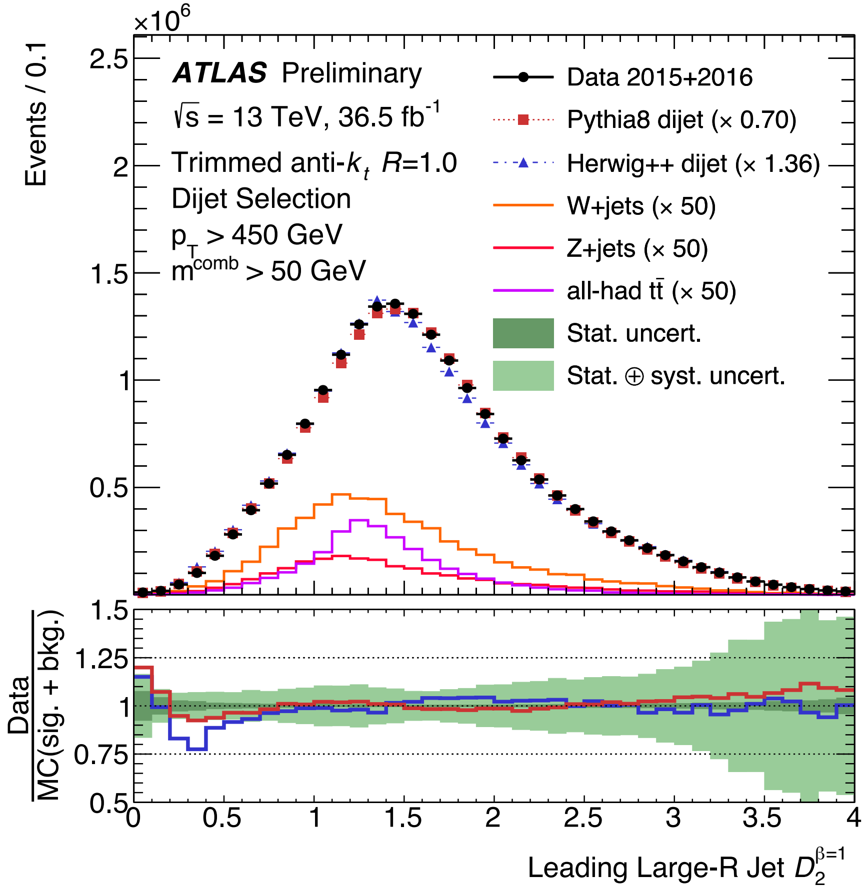
W/Z tagging substructure inputs: ATLAS

JETM-2017-004

JETM-2017-005

Cut on energy correlation D2





- Energy correlation D2 in the full 13 TeV dataset and MC simulation for $t\bar{t}$ (left) and Dijet (right) events.
- Anti$-k_t$ R=1 jets calibrated at LCW+JES scale, trimmed jet: fcut=5%, Rsubjet=0.2.
- $t\bar{t}$ sample is split into a "W –matched" part by requiring the decay products of W boson to be within the large-R jet.
- "top-matched" part requires the qqb particles from the top-decay to be within the large-R jet.
- In addition to enhance purity of jets matched to a W boson, the events is required to have a R=0.4 b-tagged jet but outside the large-R jet cone.

Quark-gluon tagging ATLAS

- Discriminating variables include:
  - Number of tracks in a jet
  - Pt weighted width of the jet from tracks.
  - Et weighted width of the jets
  - Fraction of energy carried by the largest energy constituent
  - Two point energy correlation function
  - Jet Charge: gluons have no charge while quarks have fractional charge

$$n_{\mathrm{trk}} = \sum_{\mathrm{trk} \in \mathrm{jet}}$$

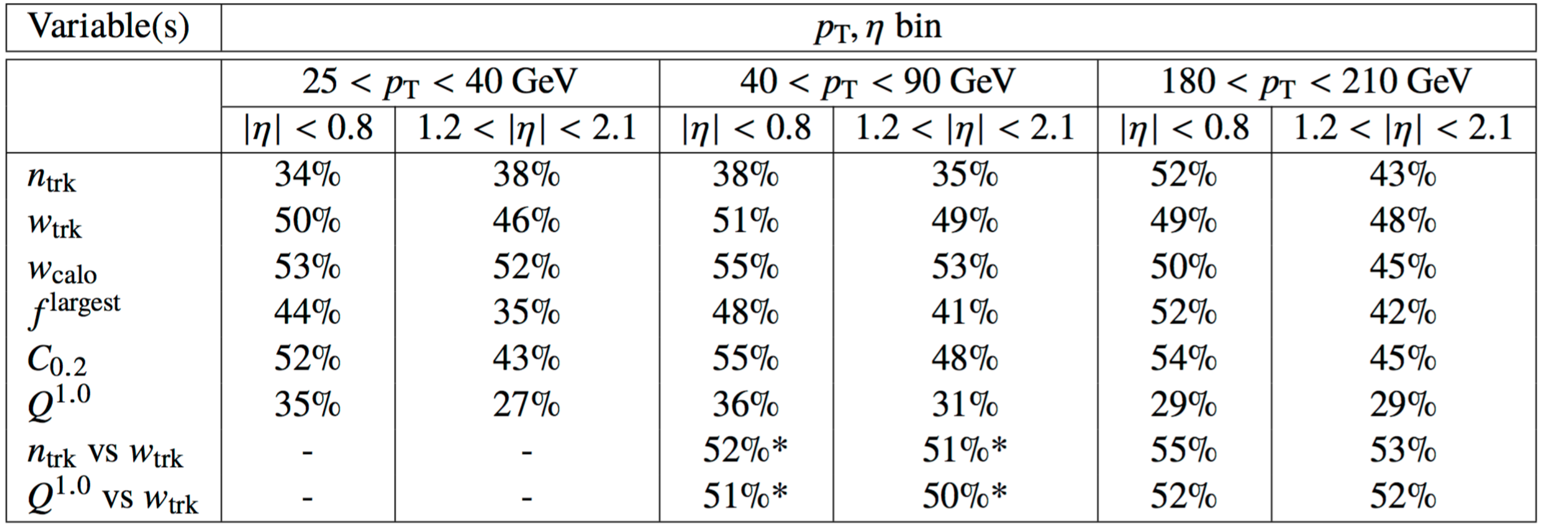
$$w_{\mathrm{trk}} = \frac{\sum_{\mathrm{trk} \in \mathrm{jet}} p_{\mathrm{T,trk}} \Delta R_{\mathrm{trk,jet}}}{\sum_{\mathrm{trk} \in \mathrm{jet}} p_{\mathrm{T,trk}}}$$

$$w_{\mathrm{calo}} = \frac{\sum_{\mathrm{const} \in \mathrm{jet}} p_{\mathrm{T,const}} \Delta R_{\mathrm{const,jet}}}{\sum_{\mathrm{const} \in \mathrm{jet}} p_{\mathrm{T,const}}}$$

$$f^{\mathrm{largest}} = \frac{E_{\mathrm{largest\ const}}}{E_{\mathrm{jet}}}$$

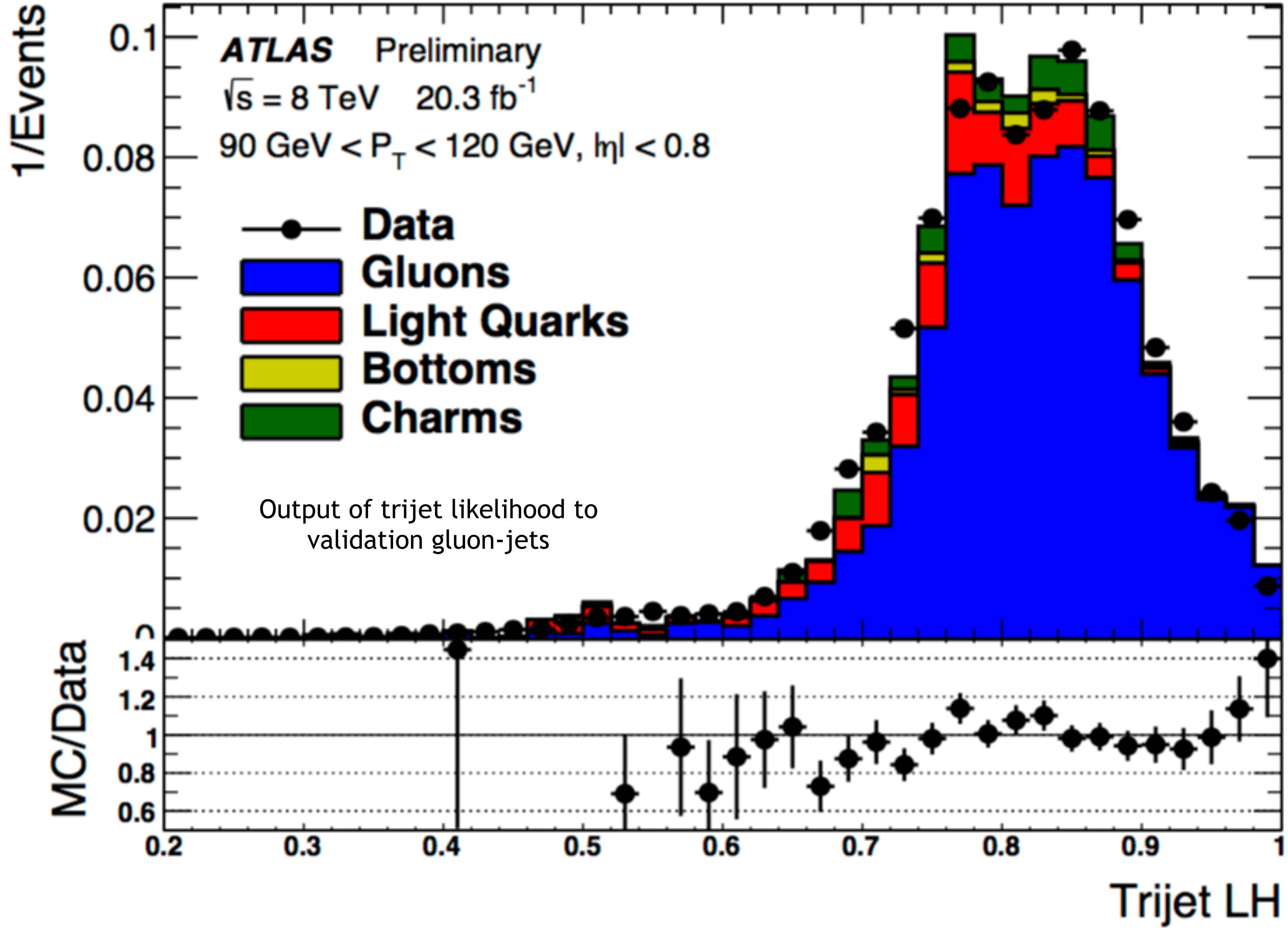
$$C_{\beta} = \frac{\sum_{i,j \in \mathrm{jet}} E_{\mathrm{T},i} E_{\mathrm{T},j} (\Delta R_{i,j})^{\beta}}{\left(\sum_{i \in \mathrm{jet}} E_{\mathrm{T},i}\right)^{2}}$$

$$Q^{\kappa} = \frac{1}{(p_{\mathrm{T}})^{\kappa}} \sum_{trk\ i\ \in \mathrm{jet}} q_i \times (p_{\mathrm{T}}^{i})^{\kappa}$$



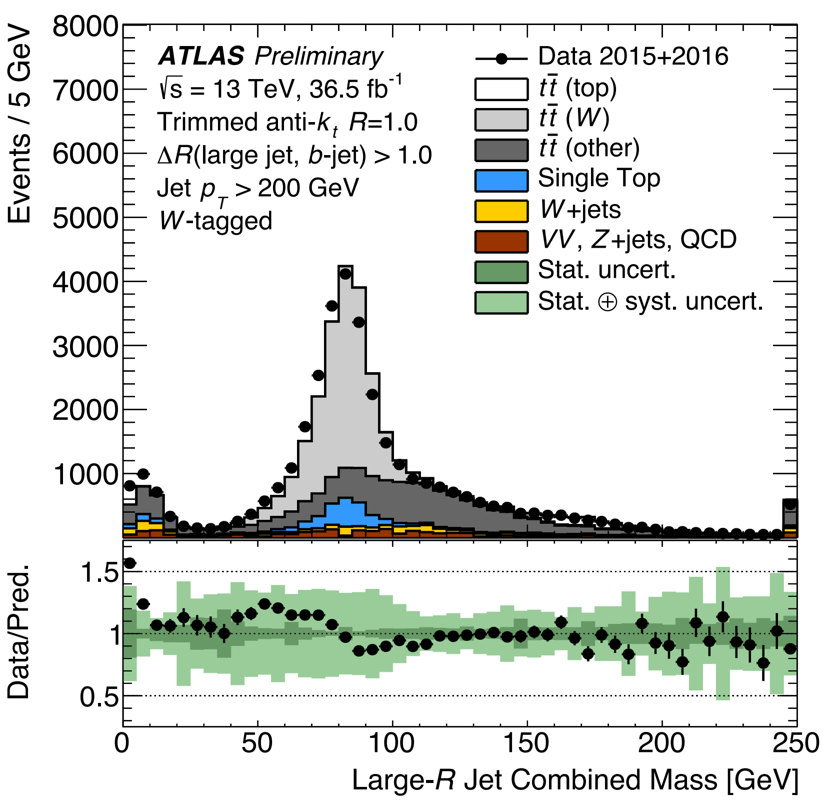
Variable(s)	p_{T}, η bin					
	$25 < p_{\mathrm{T}} < 40$ GeV		$40 < p_{\mathrm{T}} < 90$ GeV		$180 < p_{\mathrm{T}} < 210$ GeV	
	$\lvert\eta\rvert < 0.8$	$1.2 < \lvert\eta\rvert < 2.1$	$\lvert\eta\rvert < 0.8$	$1.2 < \lvert\eta\rvert < 2.1$	$\lvert\eta\rvert < 0.8$	$1.2 < \lvert\eta\rvert < 2.1$
n_{trk}	34%	38%	38%	35%	52%	43%
w_{trk}	50%	46%	51%	49%	49%	48%
w_{calo}	53%	52%	55%	53%	50%	45%
f^{largest}	44%	35%	48%	41%	52%	42%
$C_{0.2}$	52%	43%	55%	48%	54%	45%
$Q^{1.0}$	35%	27%	36%	31%	29%	29%
n_{trk} vs w_{trk}	-	-	52%*	51%*	55%	53%
$Q^{1.0}$ vs w_{trk}	-	-	51%*	50%*	52%	52%

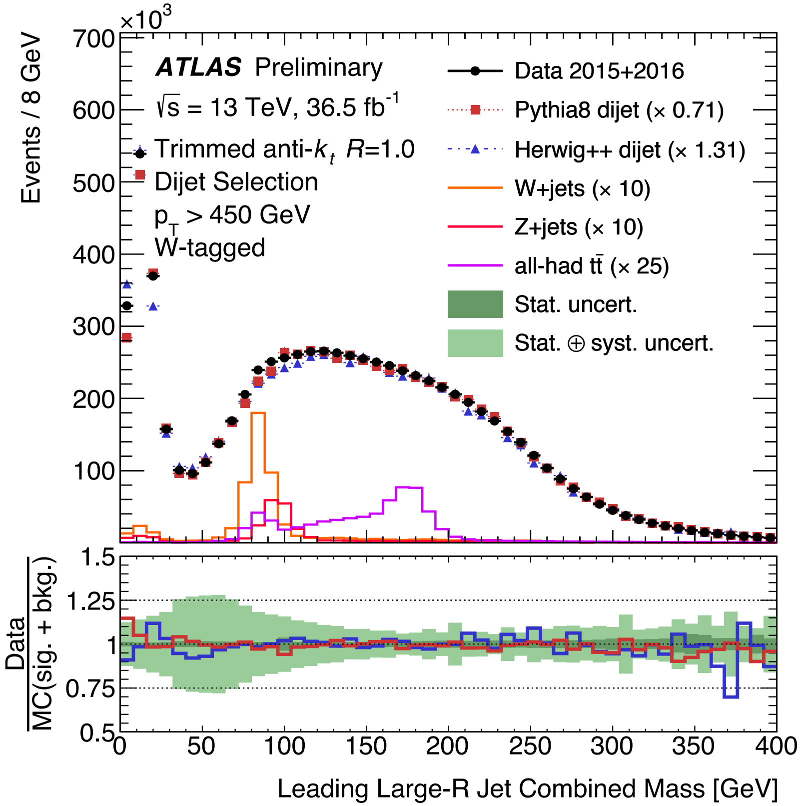
Quark-gluon tagging ATLAS



Output of trijet likelihood to validation gluon-jets

W/Z tagging substructure inputs: ATLAS





Systematic uncertainties

- scale and resolution uncertainties on large-R jet energy, mass and substructure observables
- $t\bar{t}$ modeling uncertainties on parton shower, hadronization model
- Initial-state and final-state radiation.
- Subdominant scale and resolution uncertainties on small-R jet energy and lepton kinematics

top tagging substructure inputs: ATLAS