Top tagging at CMS

Torben Dreyer

on behalf of the CMS Collaboration

BOOST 2017, Buffalo

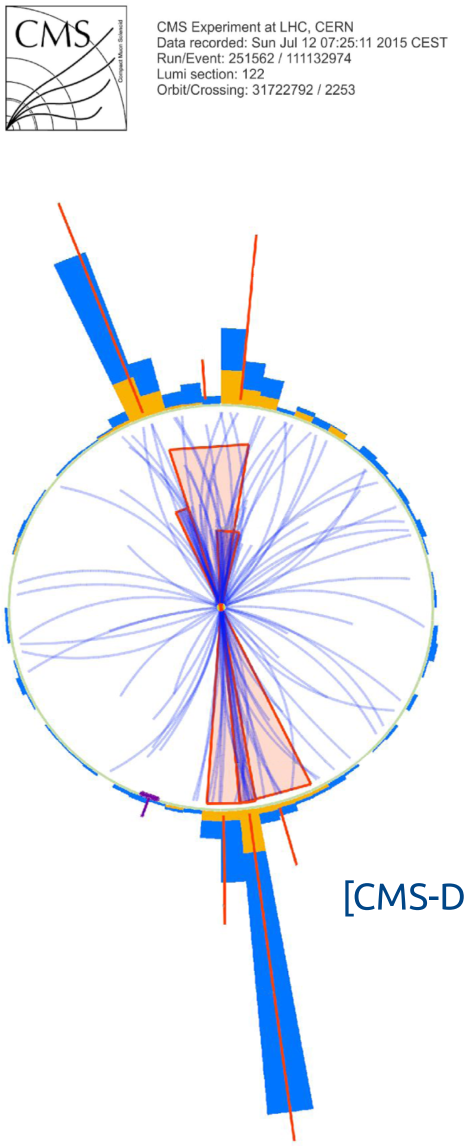
Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

Motivation

- LHC at $\sqrt{s}$ = 13 TeV
  - More top quarks with high momentum (Standard Model and new physics)
  - Top tagging more important than ever

- Understanding in data crucial
  - Measurement of top tagging efficiencies
  - Study performance of new algorithms



[CMS-DP-2016-035]

Outline

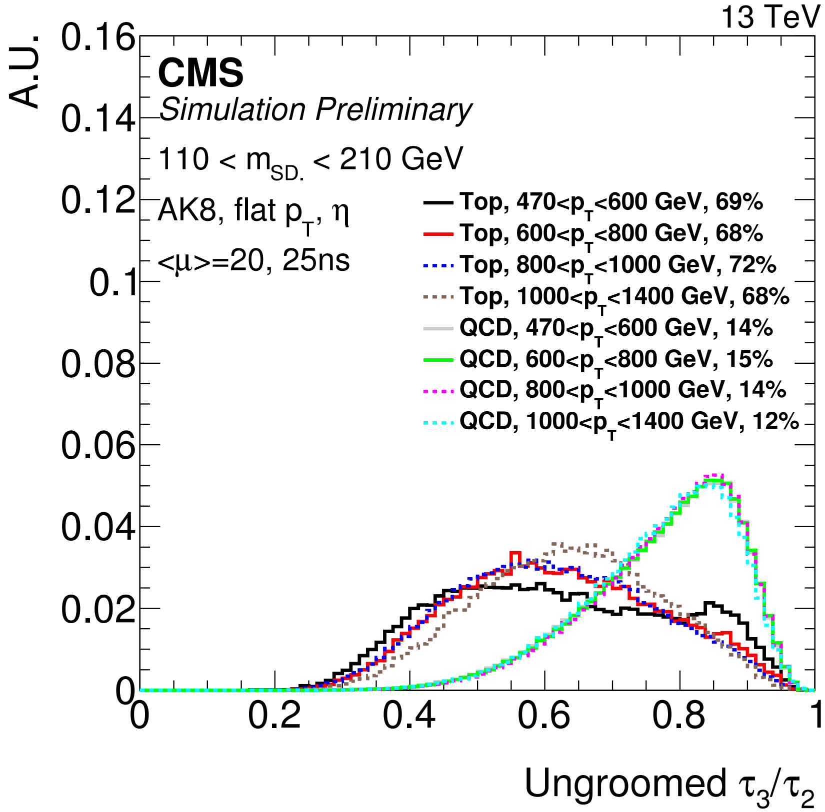
- Top tagging validation in 2016 data [CMS-DP-2017-026]
- Top tagging with Energy Correlation Functions [CMS-PAS-EXO-16-051]
- Boosted Event Shape Tagger (BEST) [CMS-DP-2017-027]
- XCone Studies [CMS-DP-2017-027]

Top tagging validation in 2016 data

CMS-DP-2017-026

CMS default top tagger

- Based on:
  - Groomed jet mass (mMDT / Soft drop)
  - Ungroomed N-subjettiness ratio τ_3/τ_2
  - Subjet b-tagging
- anti-k_T jets with R = 0.8 (AK8)
- Pileup correction:
  - Charged Hadron Subtraction or PUPPI
- Different N-subjettiness working points
  - with and without subjet b-tag

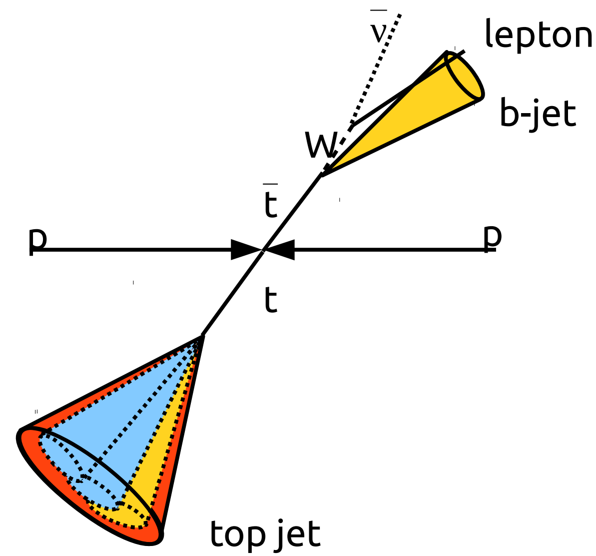


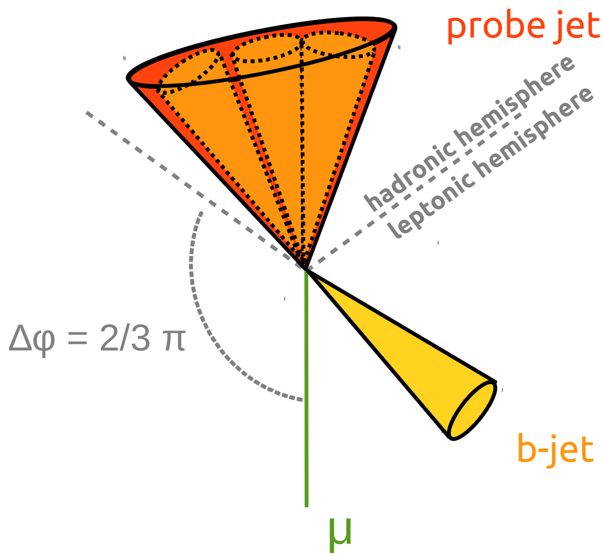
[CMS-PAS-JME-15-002]

Validation in data

- Muon+jets $t\bar{t}$ selection
- High p_T muon
- Large boost expected

  → use non isolated muons





- Tag and probe method
- Leptonic decay leg as tag
- B-tag in leptonic hemisphere
- Probe jet: AK8 (CA15) jet in hadronic hemisphere

Validation in 2016 data

- Soft drop groomed AK8 jets
- Different categories for $t\bar{t}$ simulation

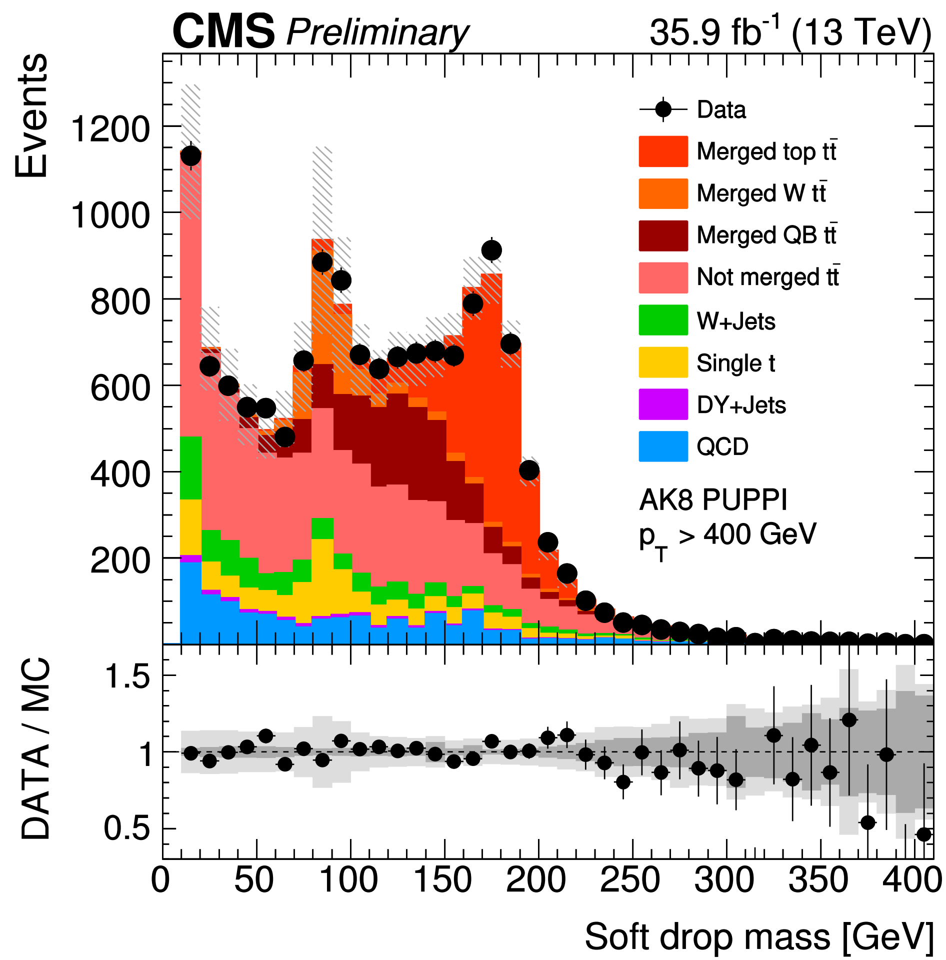
  (matching to generator particles)

  - Merged top
  - Merged W
  - Merged b+light quark
  - Not merged

- Template fit soft drop mass distribution

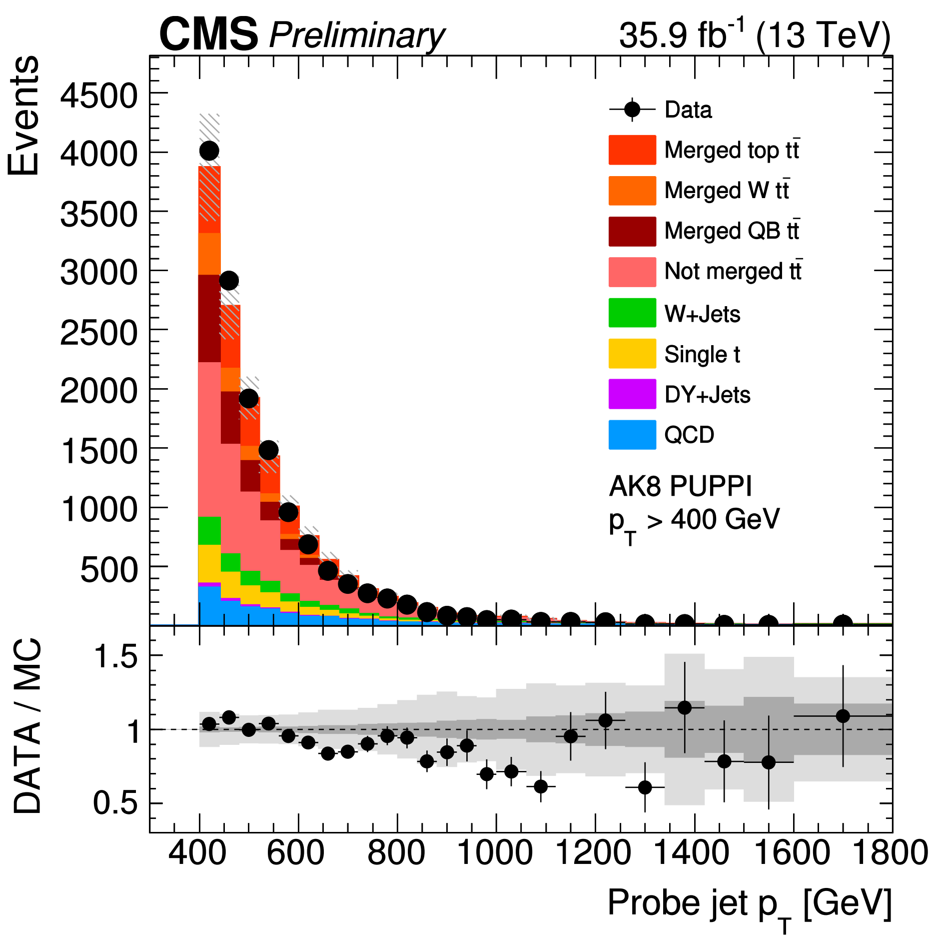
  → new method

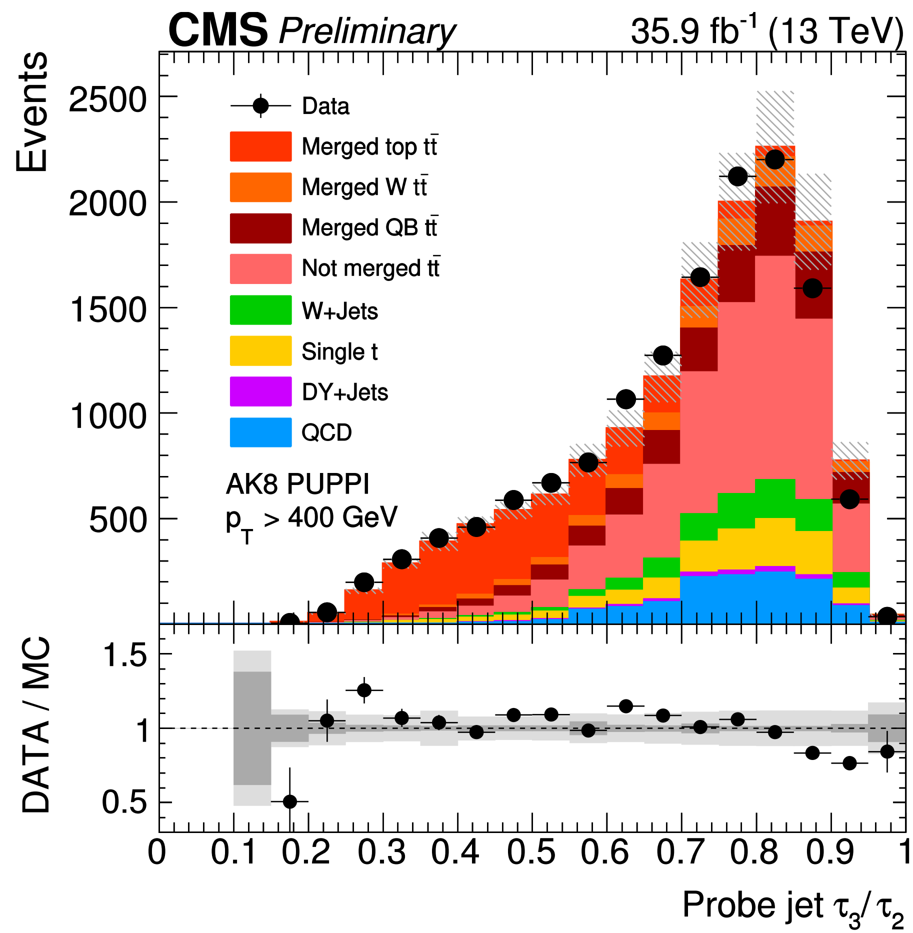
  → scale factors for different $t\bar{t}$ categories



[CMS-DP-2017-026]

Kinematic distributions

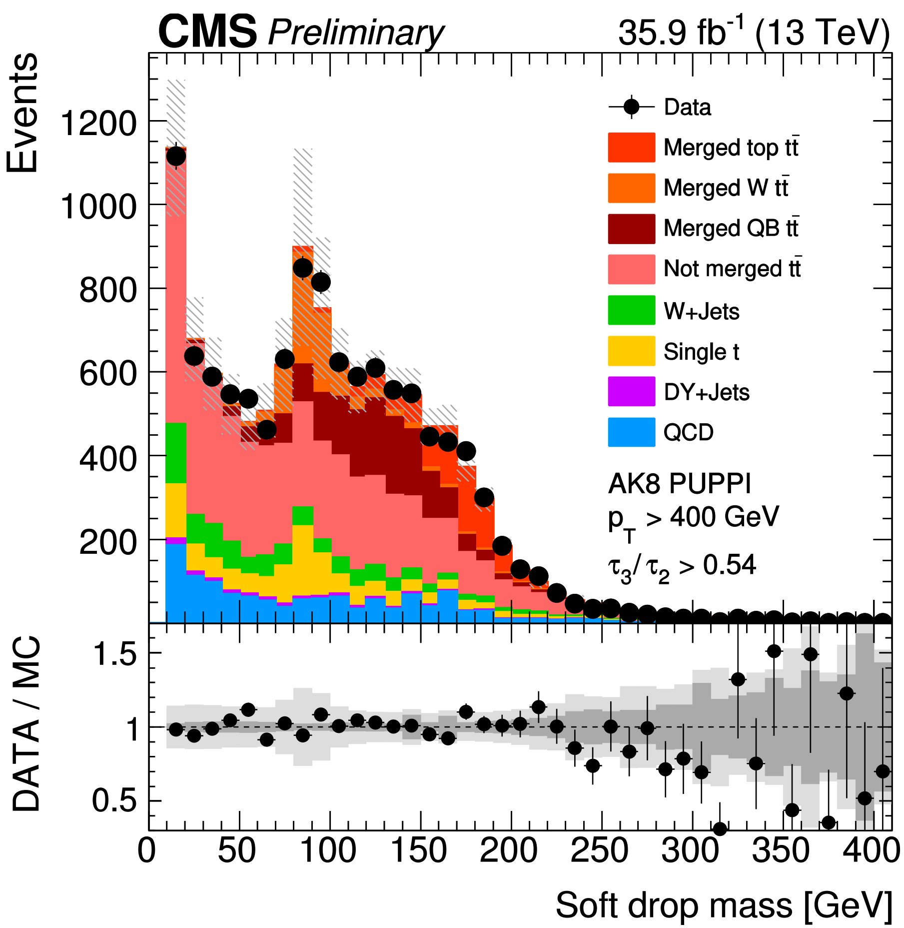


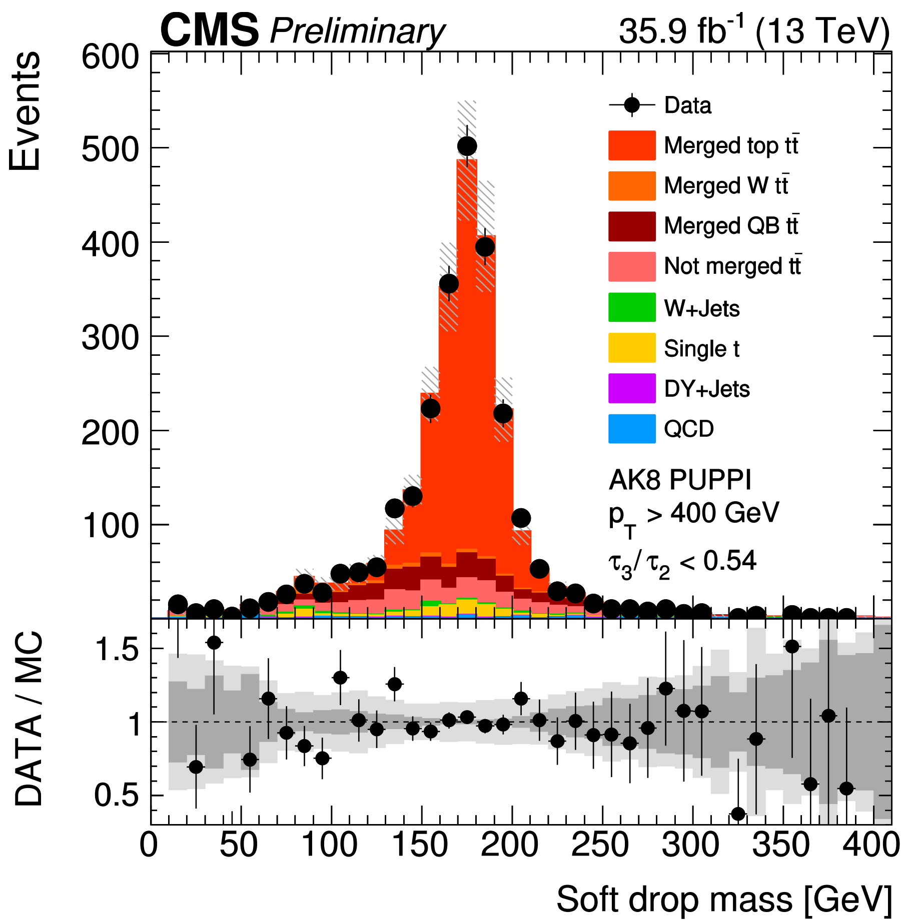


[CMS-DP-2017-026]

Softdrop mass – pass and fail region

- Fit in pass and fail regions
  - Fail: $\tau_3/\tau_2 > 0.54$
  - Pass: $\tau_3/\tau_2 < 0.54$





[CMS-DP-2017-026]

Efficiency measurement

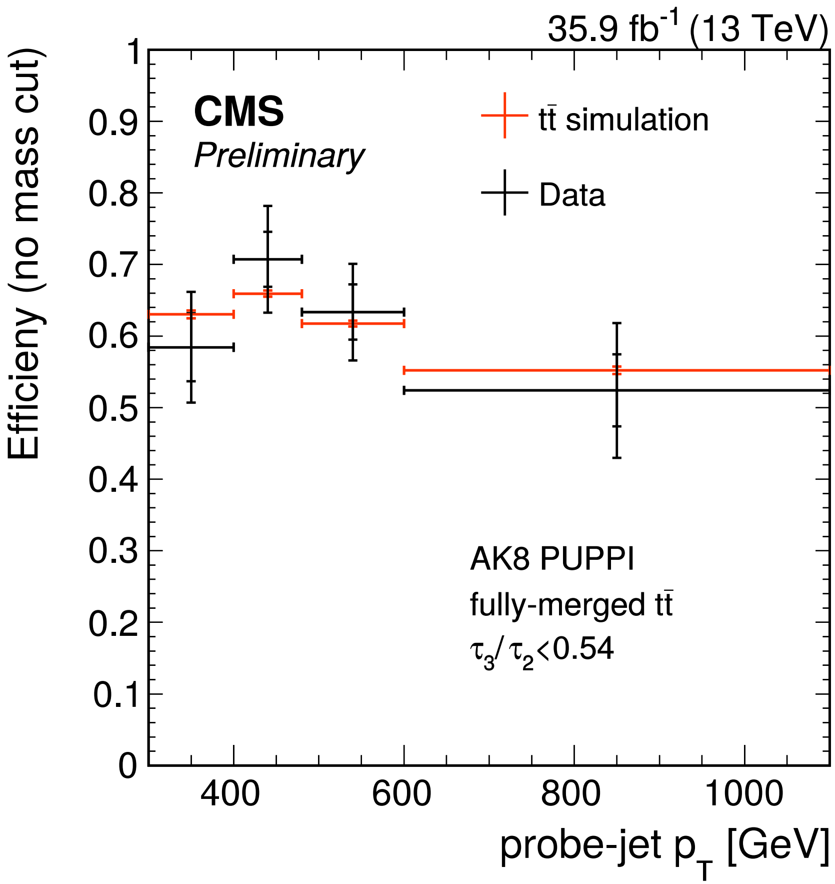
- Efficiency:

  $$\varepsilon = N_{pass}/(N_{pass}+N_{fail})$$

- For each $t\bar{t}$ contribution

- Scale factors: $sf = \varepsilon_{data}/\varepsilon_{simulation}$

- Fits performed in different p_T bins



[CMS-DP-2017-026]

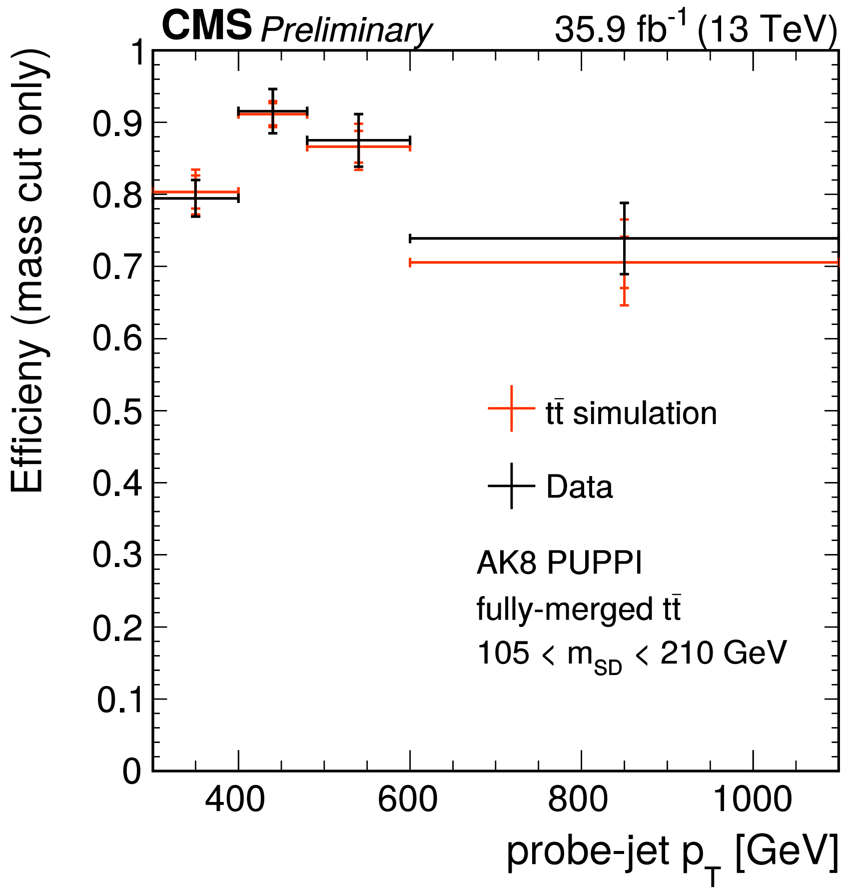
- Full set of systematic uncertainties studied

New for 2016 data

  – JEC, JER, pileup, muon/trigger scale factors, b-tagging, PDF, and scale uncertainties

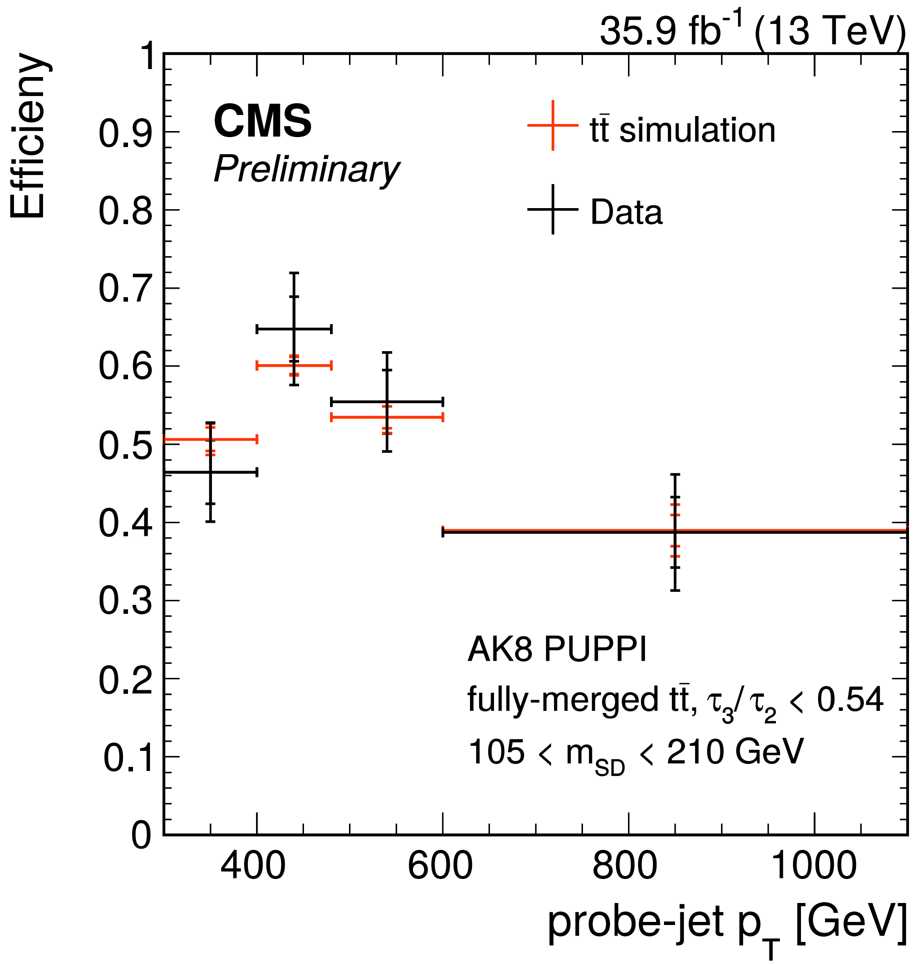
Mass cut efficiency

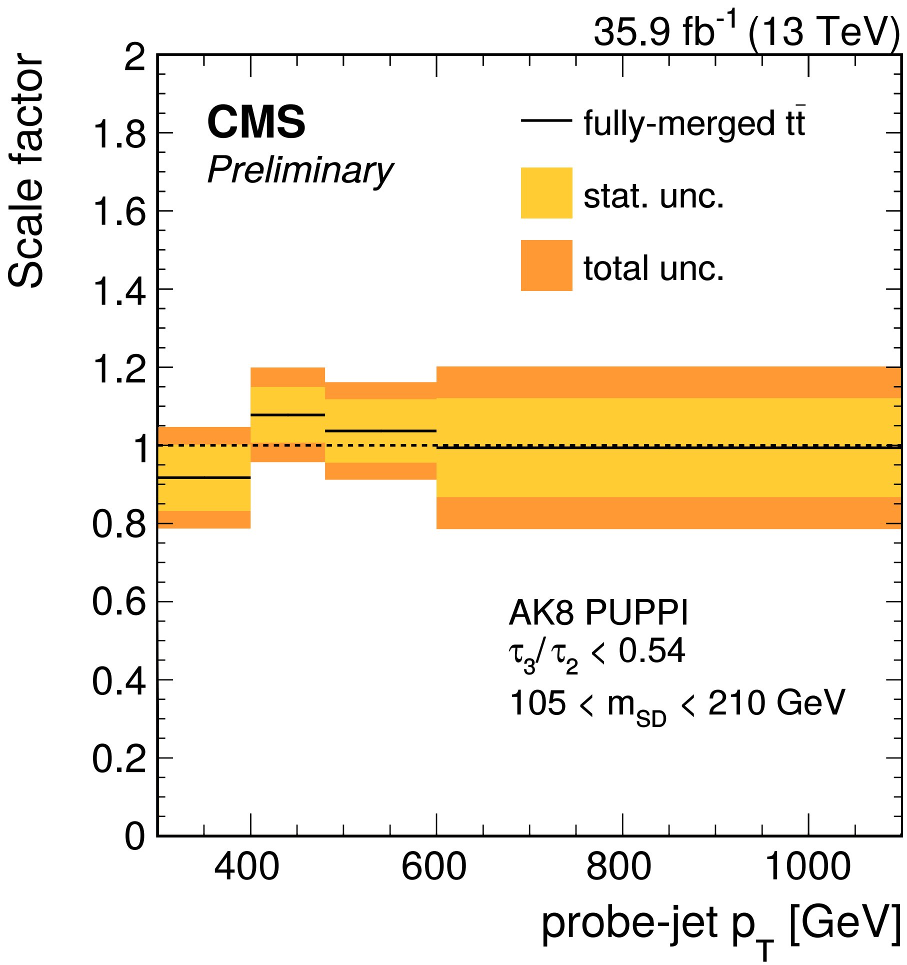
- Mass cut not included for the fit
  - Tails important for semi-merged and not merged contributions
- Extra mass cut efficiency
  - $\varepsilon_{mass} = N_{pass,\ mass\ cut}/N_{pass}$
  - For data use data distribution directly
  - Post fit MC (all categories) for simulation
  - Total efficiency: $\varepsilon = \varepsilon_{no\ mass}\ \varepsilon_{mass}$



[CMS-DP-2017-026]

Efficiency and scale factors





[CMS-DP-2017-026]

Top tagging with Energy Correlation Functions

CMS-PAS-EXO-16-051

ECF BDT tagger

- Developed in the context of a dark matter search [CMS-PAS-EXO-16-051]
- Use of energy correlation functions in CA15 jets

  e.g. 3 point ECF:

$$_2e_3^{(\beta)} = \sum_{1\le i<j<k\le n_j} z_i z_j z_k \min\{\Delta R_{ij}^{\beta}\Delta R_{ik}^{\beta}, \Delta R_{ij}^{\beta}\Delta R_{jk}^{\beta}, \Delta R_{ik}^{\beta}\Delta R_{jk}^{\beta}\}$$

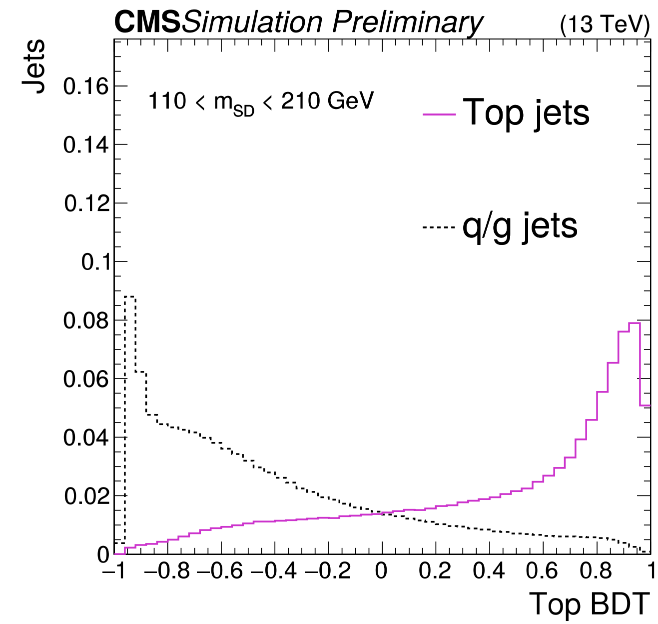
with: $z_i \equiv \frac{p_{Ti}}{\sum_{j\in jet} p_{Tj}}$

- ECF ratios proposed in [arXiv:1609.07483]

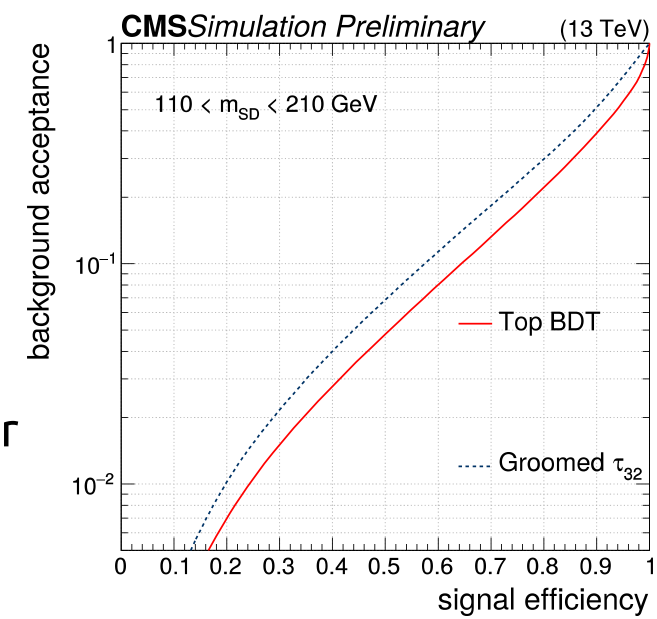
$$N_3 = \frac{_2e_4^{(\beta)}}{(_1e_3^{(\beta)})^2}$$ → similar to τ_3/τ_2

- several dimensionless ECFs ratios show separation power

→ BDT with 11 ECF ratios, f_{rec}, and τ_{32}^{SD}



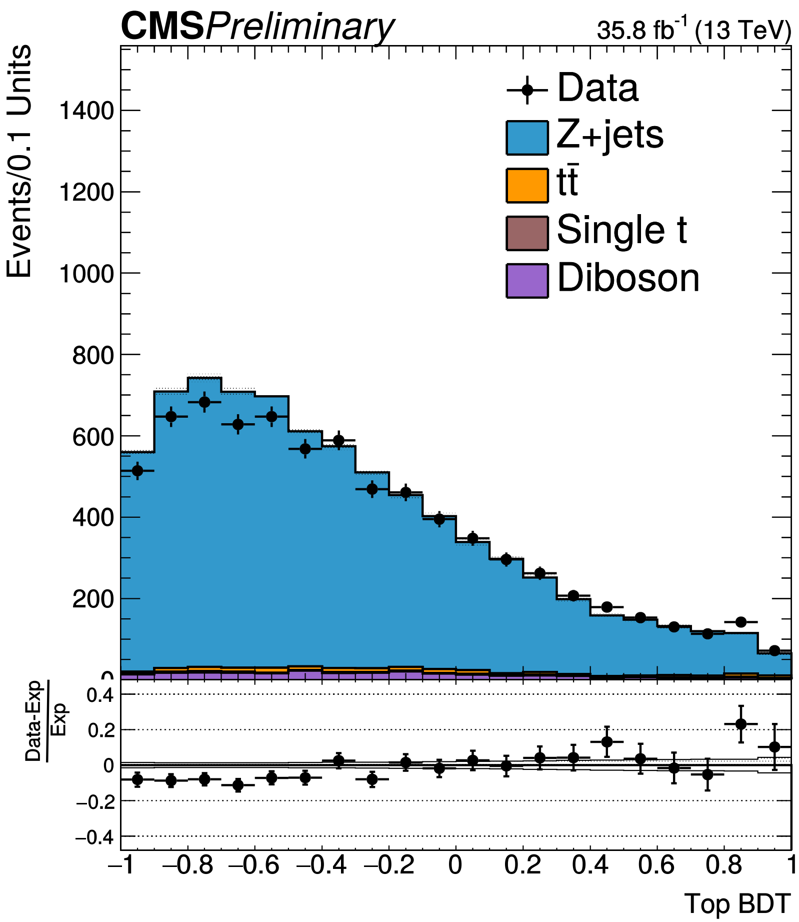
[CMS-PAS-EXO-16-051]



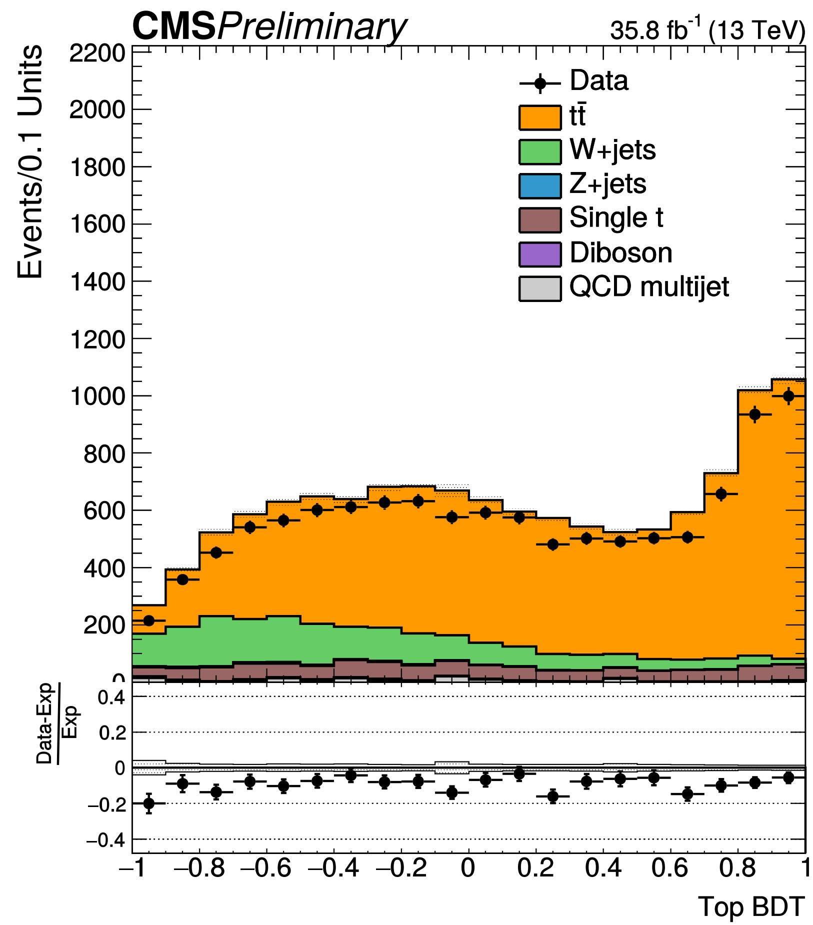
[CMS-PAS-EXO-16-051]

ECF BDT validation in 2016 data

- Validation on top jets in lepton+jets $t\bar{t}$ decays and quark/gluon jets in Z+jets
- Good performance in data



[CMS-PAS-EXO-16-051]



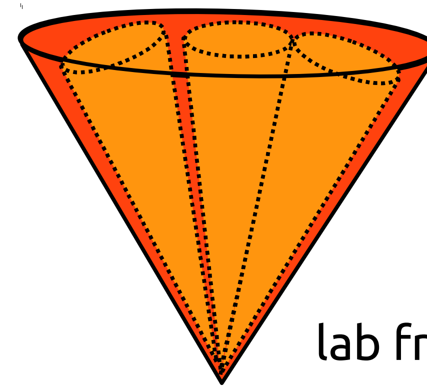
[CMS-PAS-EXO-16-051]

Boosted Event Shape Tagger (BEST)

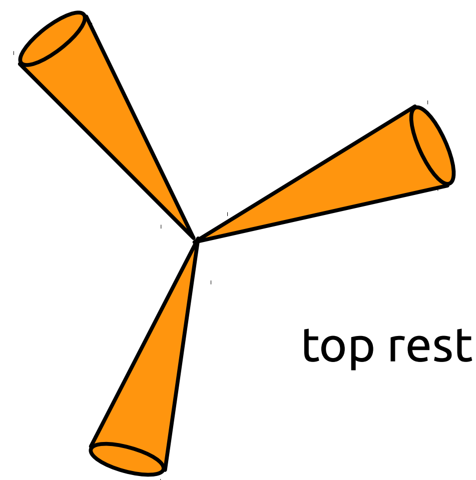
CMS-DP-2017-027

Boosted Event Shape Tagger

- Heavy particle jet classification and separation to light jets
- Targets top, H, W, Z
- Idea:
  - go to different hypothesized reference frames
  - Isotropic in correct rest frame
  - N-prong structure
- Neural network for classification



lab frame



top rest frame

- Based on [Phys. Rev. D 94, 094027 (2016)]
- First discussed at BOOST 2016

Input observables

- Fox-Wolfram Moments:

$$H_\ell = \sum_{i,j} \frac{|\vec{p}_i||\vec{p}_j|}{s} \cdot P_\ell(\cos(\phi_{i,j}))$$

- Eigenvalues of Sphericity tensor:

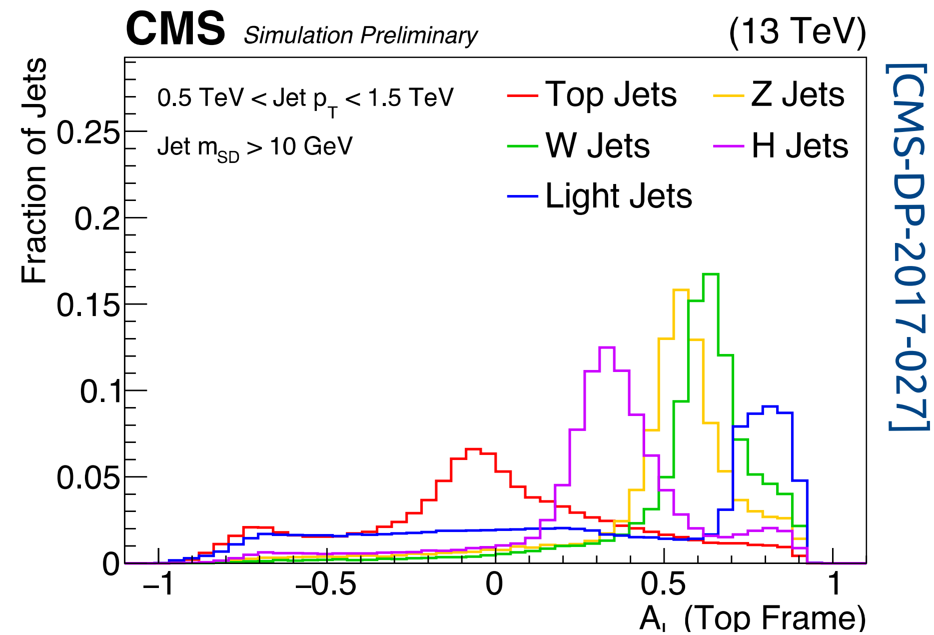
$$S^{\alpha,\beta} = \frac{\sum_i p_i^\alpha p_i^\beta}{\sum_i |\vec{p}_i|^2}$$

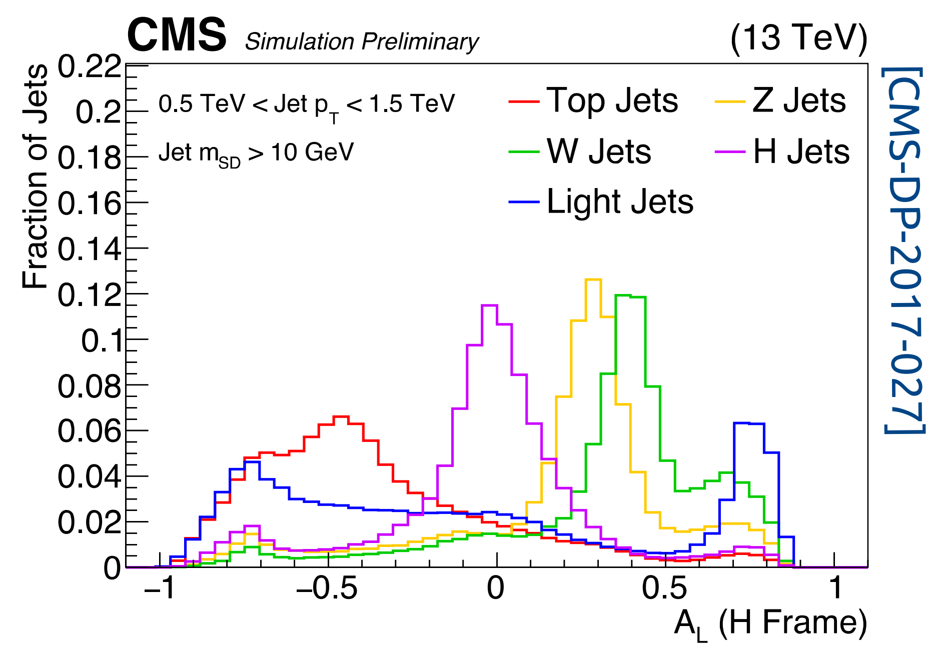
- Longitudinal Jet Asymmetry:
  - Recluster constituents
  - AK4 jets

$$A_L = \frac{\sum_{jet} p_z^{jet}}{\sum_{jet} p^{jet}}$$

Top frame

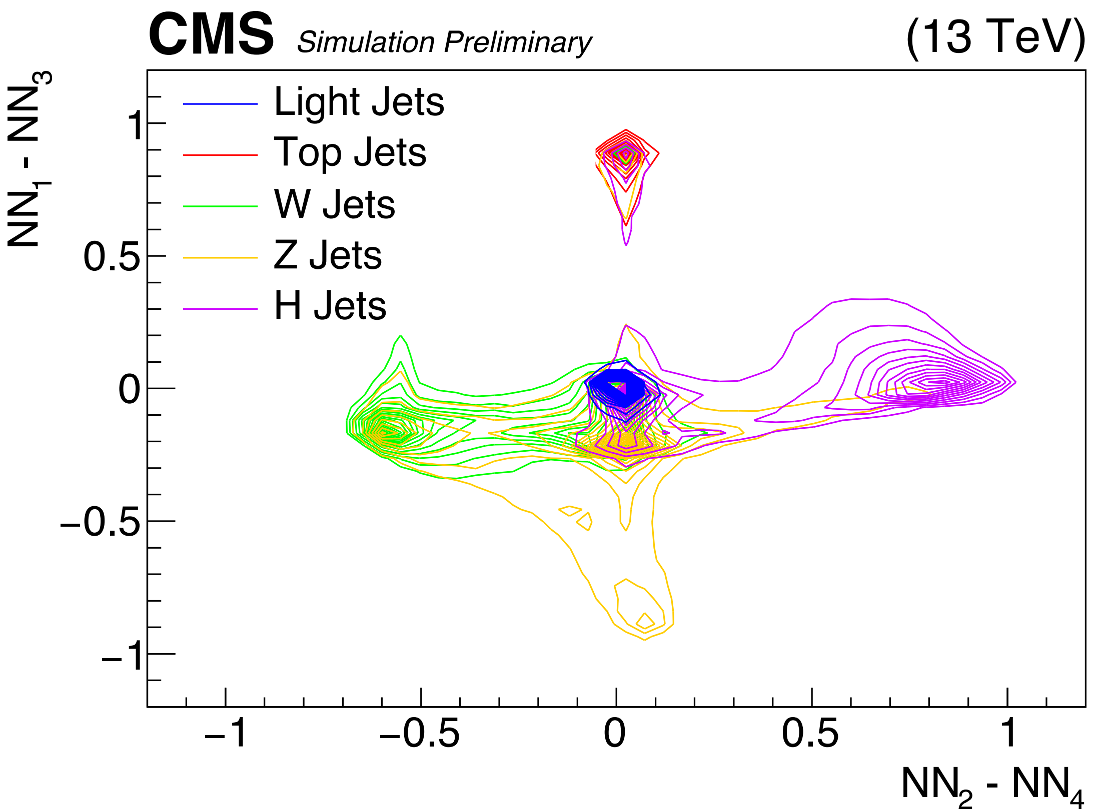
Higgs frame





Neural network

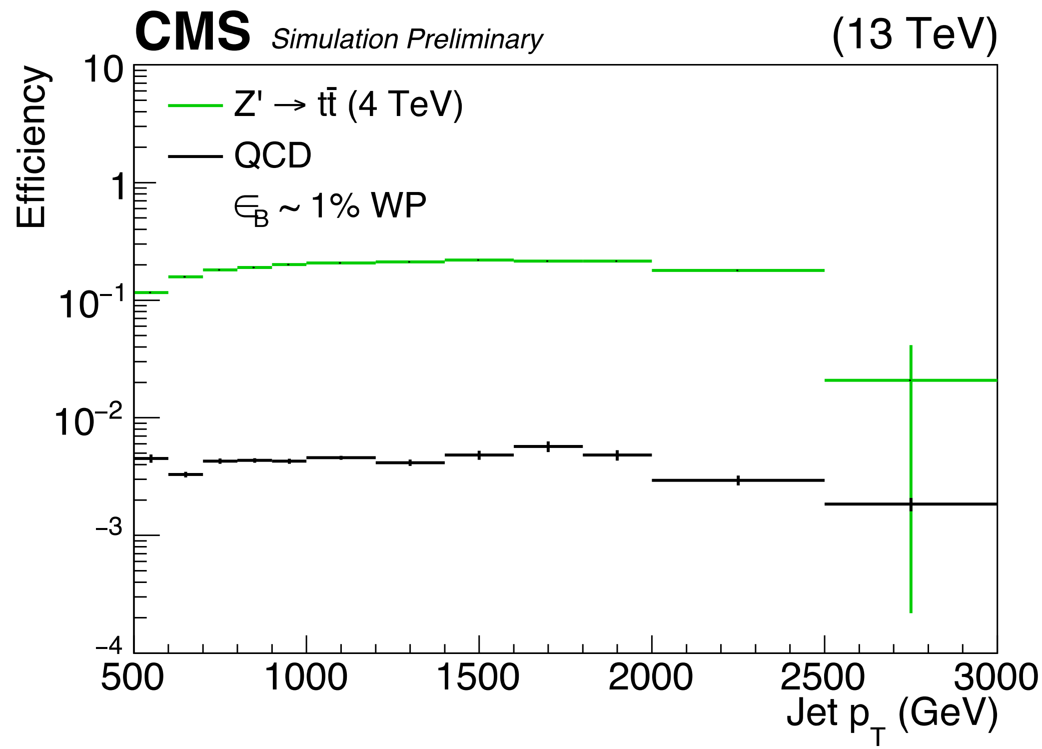
- Neural network for classification
- 41 input observables
- Simultaneous separation between top, W, H, Z, and light flavor
- 5 output nodes
- Each node optimized for another particle classification
- Visualization in 2D plane
- Best separation for top and Higgs



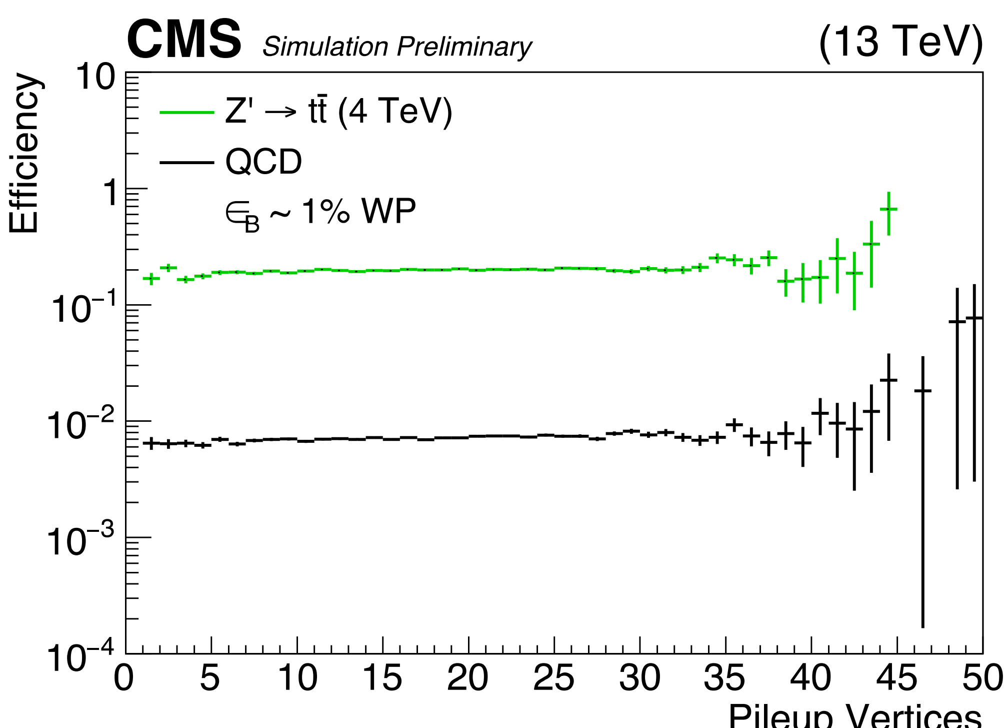
[CMS-DP-2017-027]

Tagging efficiency for top quarks

- Top tagging efficiency
- Based on m_{SD}, τ_{32}, b-tag, BEST NN output
- Working point with background efficiency ~1%



[CMS-DP-2017-027]



[CMS-DP-2017-027]

- Stable as a function of jet p_T and number of pileup vertices

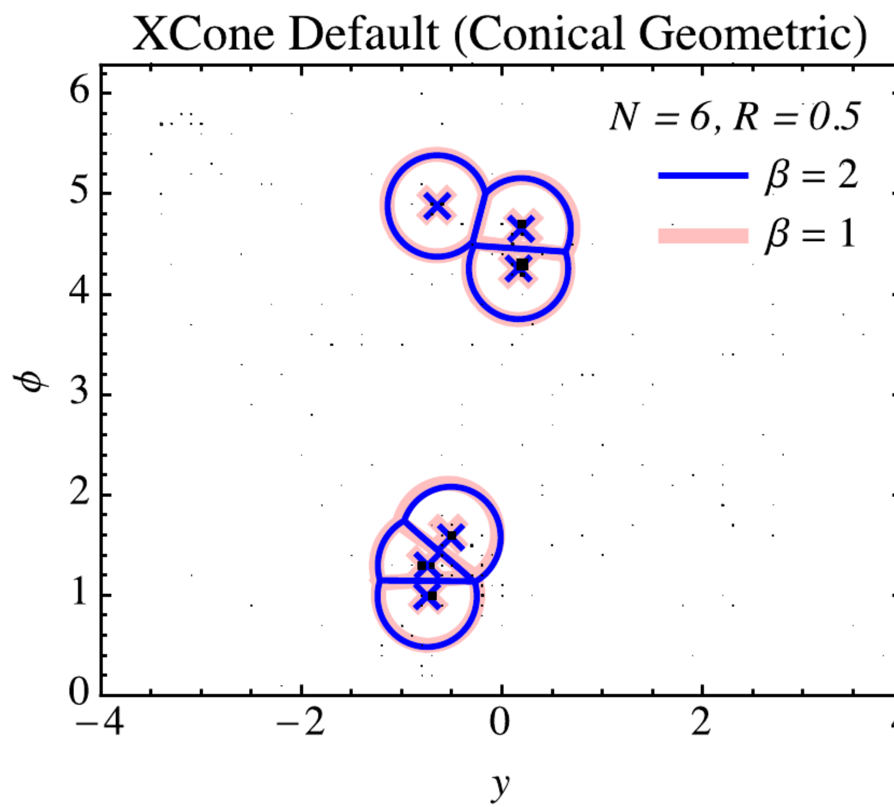
XCone Studies

CMS-DP-2017-027

XCone

- Exclusive jet algorithm based on N-jettiness [I. W. Stewart, F. J. Tackmann et al., JHEP 1511 (2015) 072]
  - N-jettiness: $\widetilde{\mathcal{T}}_N = \sum \min\{\rho_{\text{jet}}(p_i, n_1), \ldots, \rho_{\text{jet}}(p_i, n_N), \rho_{\text{beam}}(p_i)\}$
  - Measure how particles are distributed along N axis in the event

- Returns fixed number of jets
- Conical jets
- Jets stay separated in the boosted regime
- Smooth transition between resolved and boosted regime



[JHEP 1511 (2015) 072]

Optimal number of jets

[CMS-DP-2017-027]

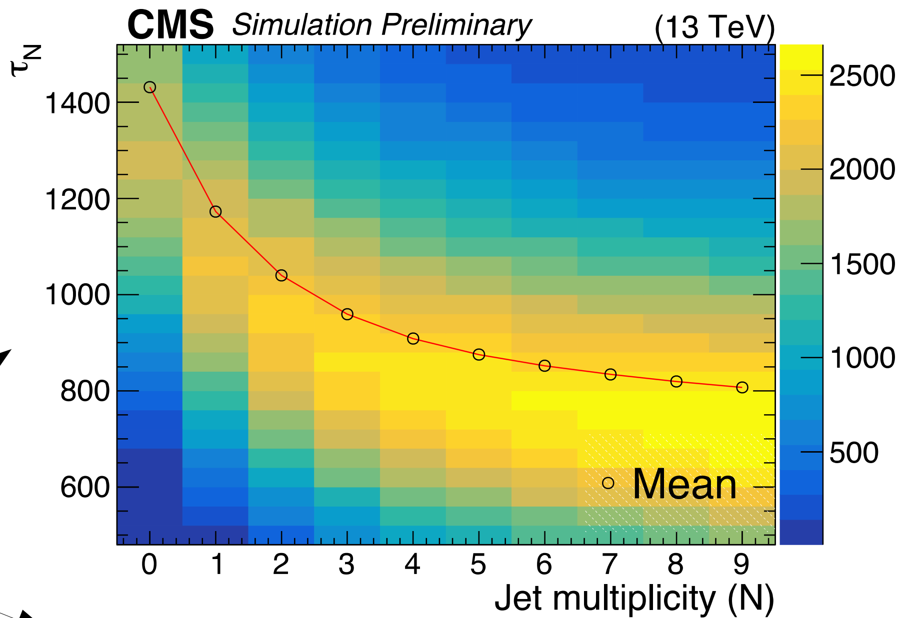
- Find the optimal N-jet to reconstruct the event
- Use N-jettiness
- Different methods studied
- $N_{optimal}$:

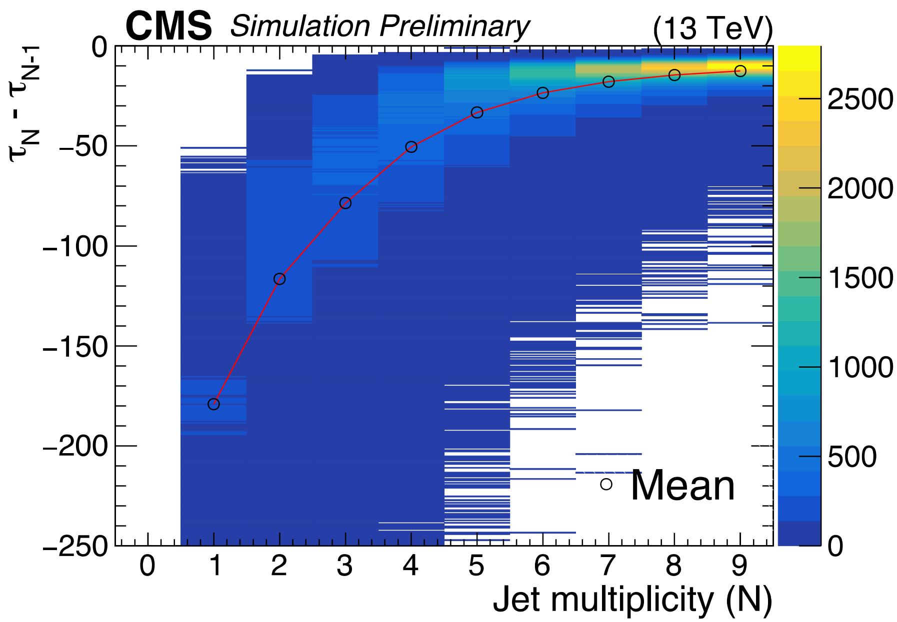
  smallest N with $\tau_N - \tau_{N-1} > -30$

→ Reconstruct event with $N_{optimal}$ XCone jets

- Can be used to count jets
  - resolved and boosted regime

$t\bar{t}$ simulation
μ+jets

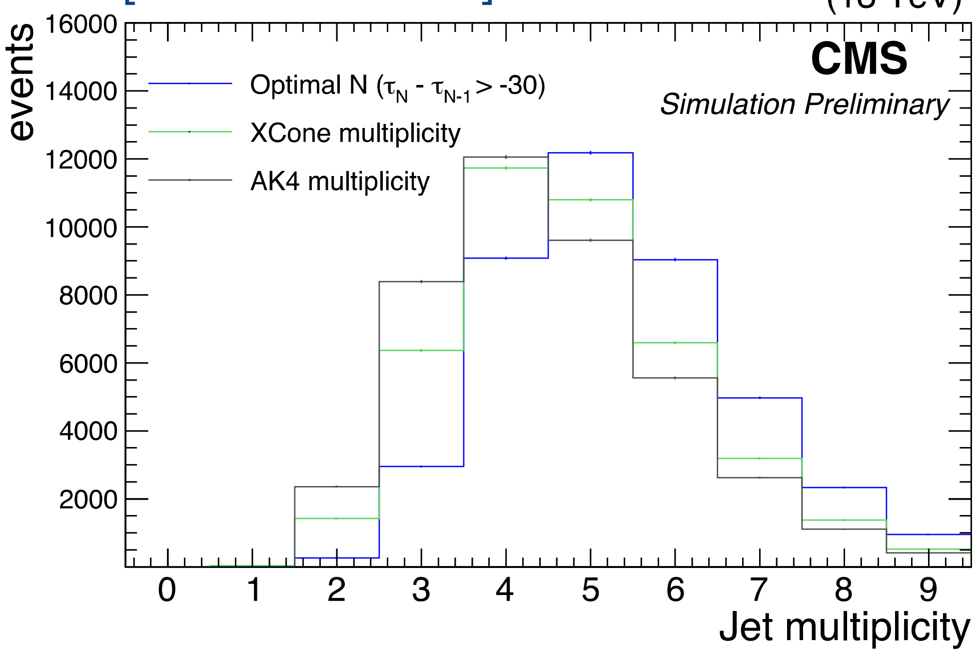


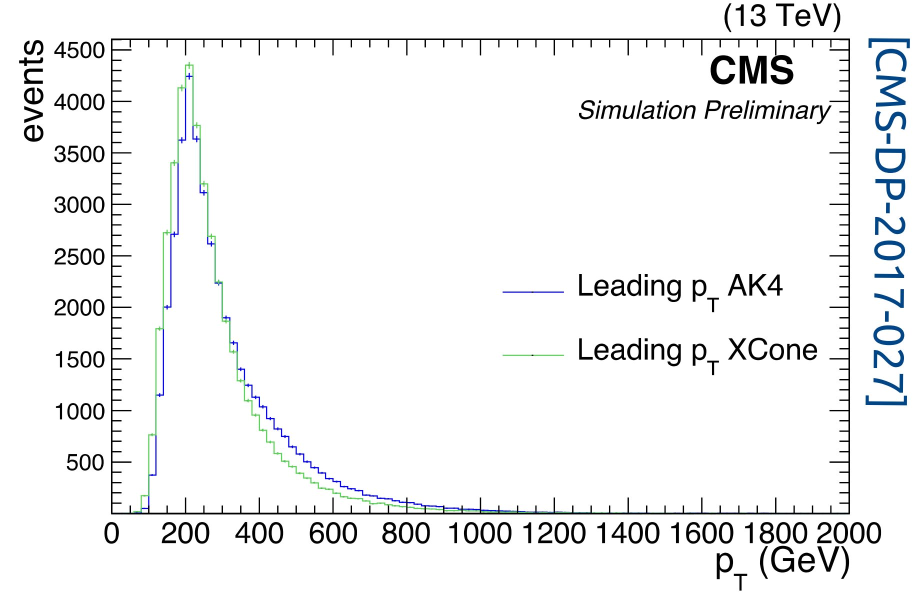


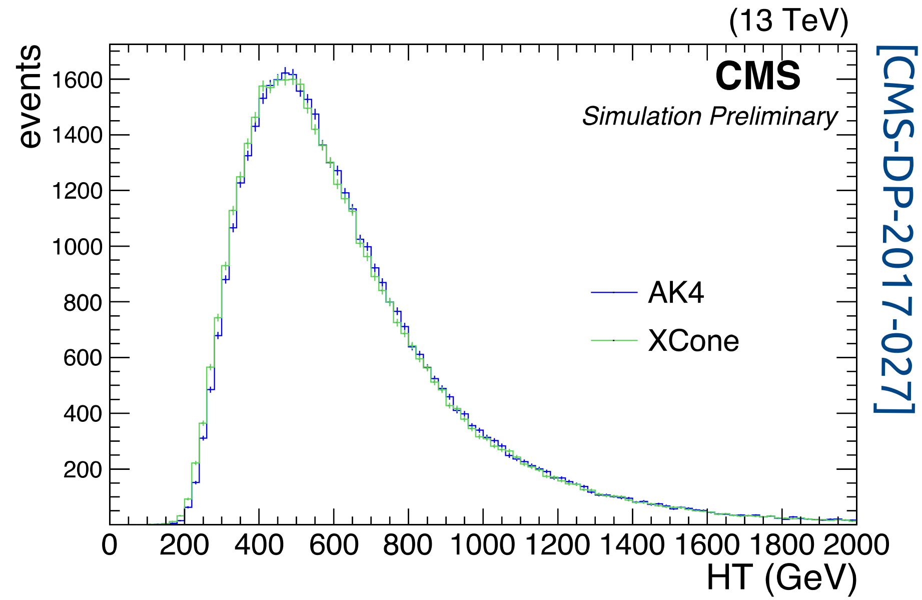
XCone jets

$t\bar{t}$ simulation, μ+jets

[CMS-DP-2017-027]







- Same kinematic cuts on XCone and AK4 jets

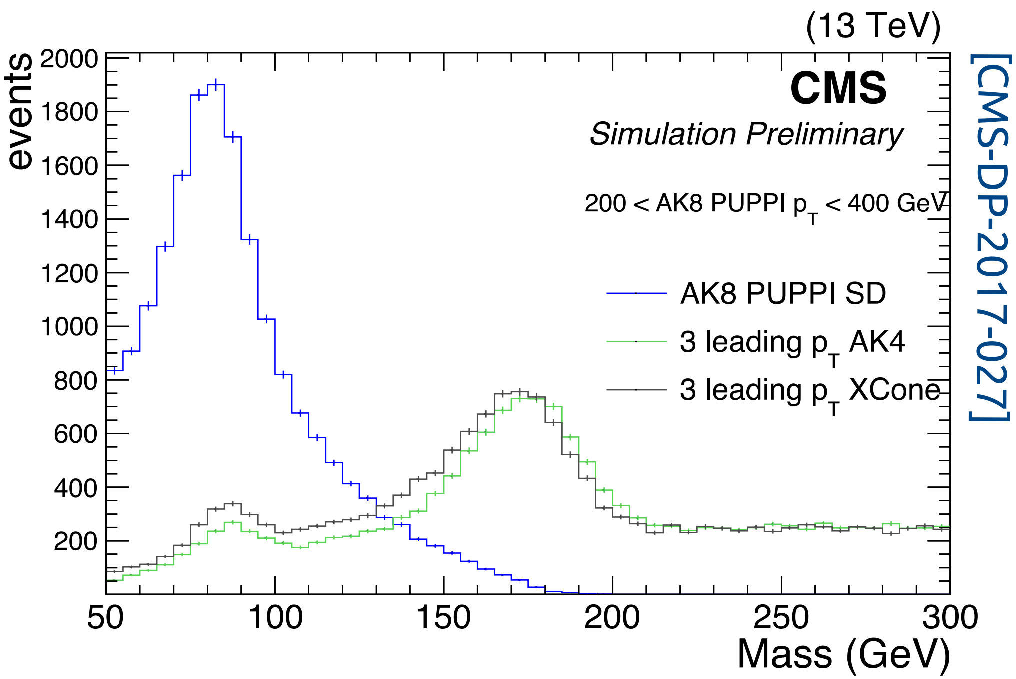
  → similar jet multiplicity

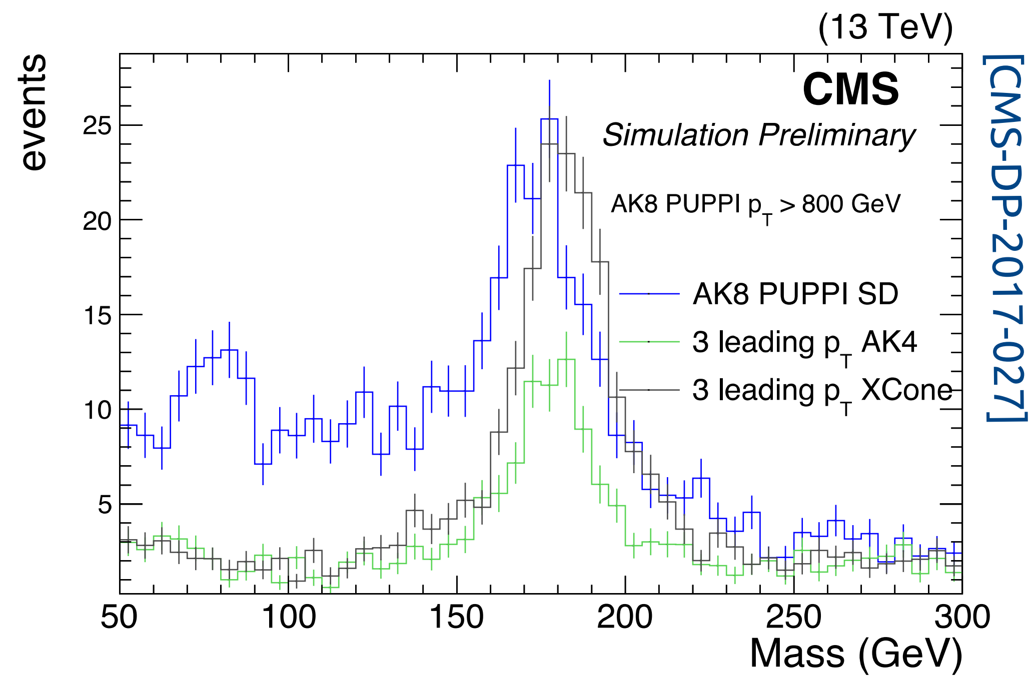
- Similar behavior of jet p_T and HT

Top mass from XCone jets

- Mass reconstructed from the 3 leading XCone, AK4 jets compared to AK8 Soft drop mass

$t\bar{t}$ simulation, μ+jets



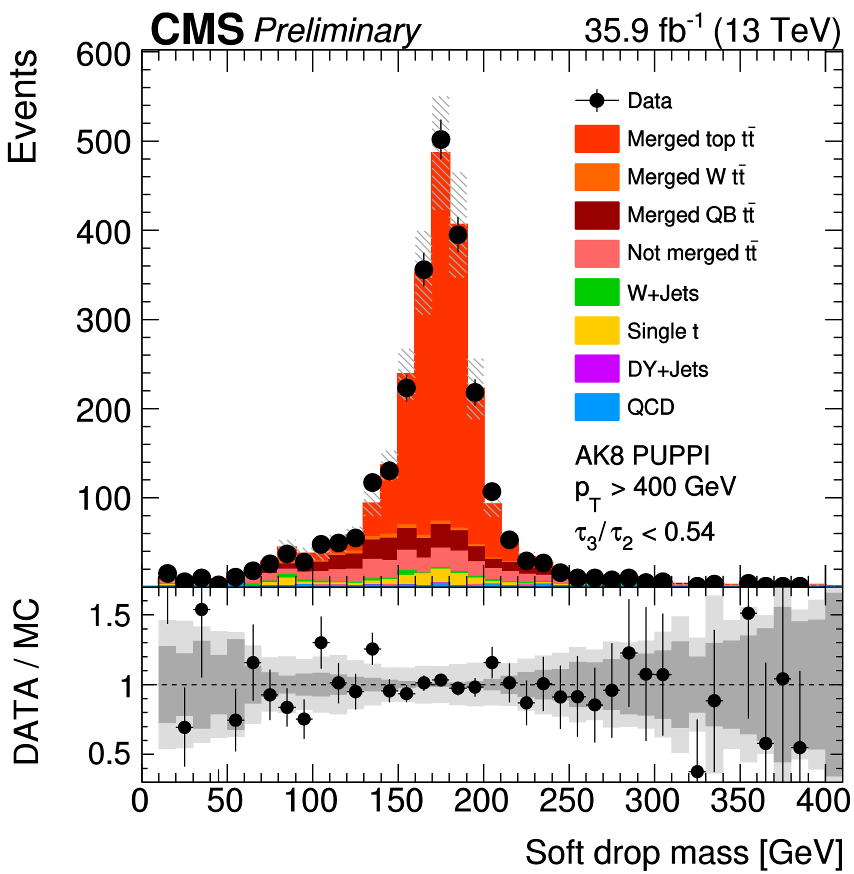


→ Similar performance between AK4 and XCone for low p_T

→ Better performance by XCone in the boosted region

Summary

- Top tagging very important at 13 TeV
- Top tagging used and studied in 2016 data
- Jet substructure well described by simulation
- Performance well understood in data and simulation
- New methods studied in simulation



[CMS-DP-2017-026]

Back up

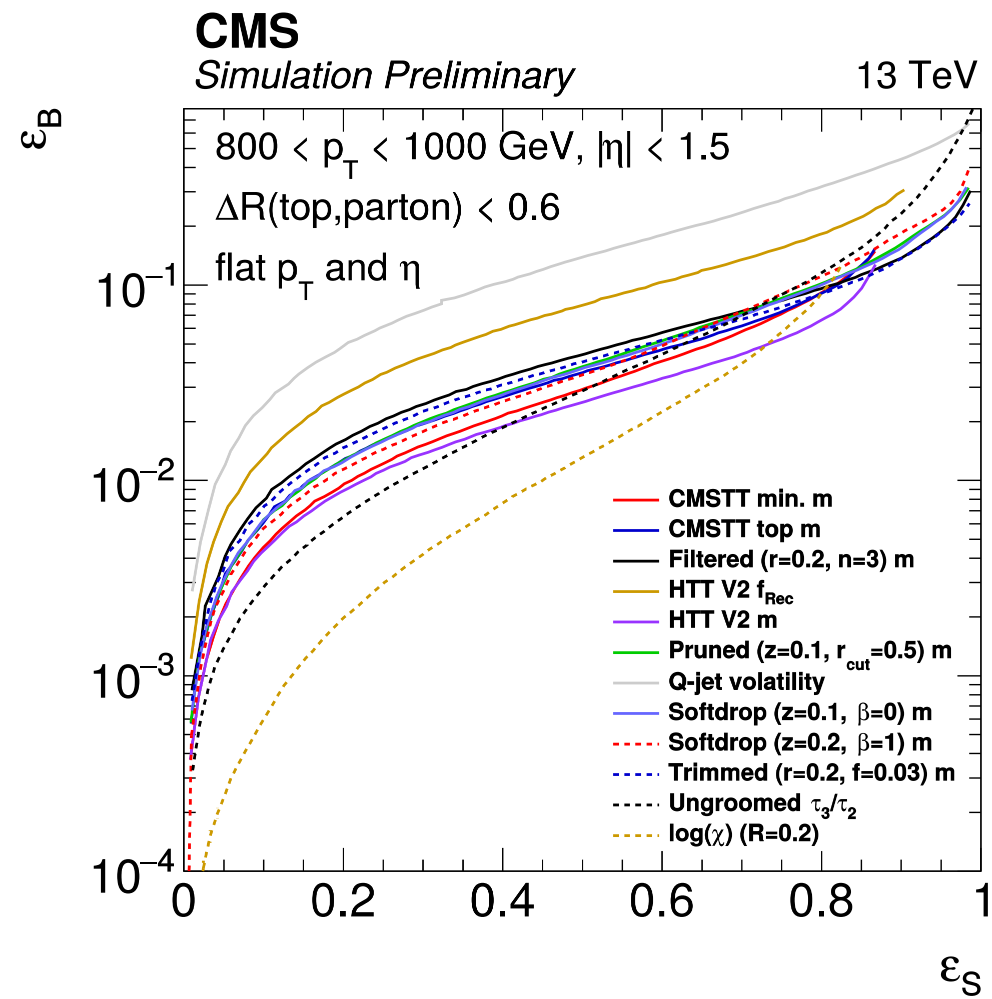
13 TeV simulation studies

CMS-PAS-JME-15-002

13 TeV simulation studies

Studied variables:

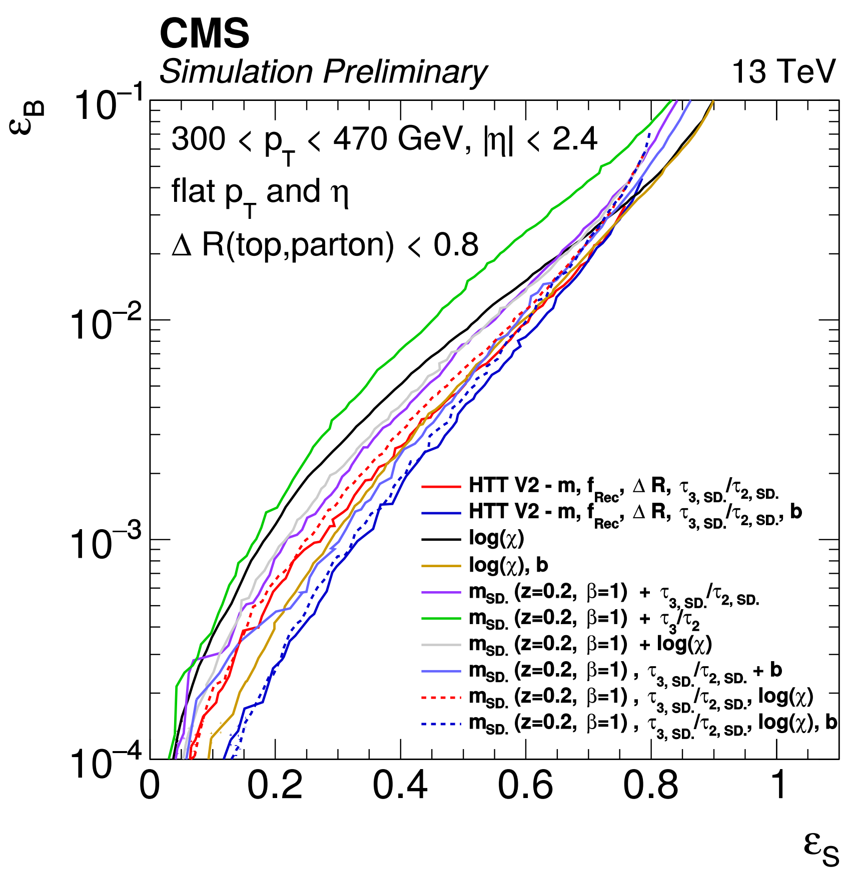
- CMSTT v1 variables
  - min. pairwise mass
  - top mass
- HepTopTagger V2 (HTT V2) variables
  - $f_{Rec} = \min_{ij} \left| \frac{\frac{m_{ij}}{m_{123}}}{\frac{m_W}{m_t}} - 1 \right|$
  - mass
- Groomed masses:
  - Pruned, Filtered, Trimmed, Soft drop
- N-subjettiness ratio τ_3/τ_2
- Qjets-volatility
- Shower deconstruction log(X)

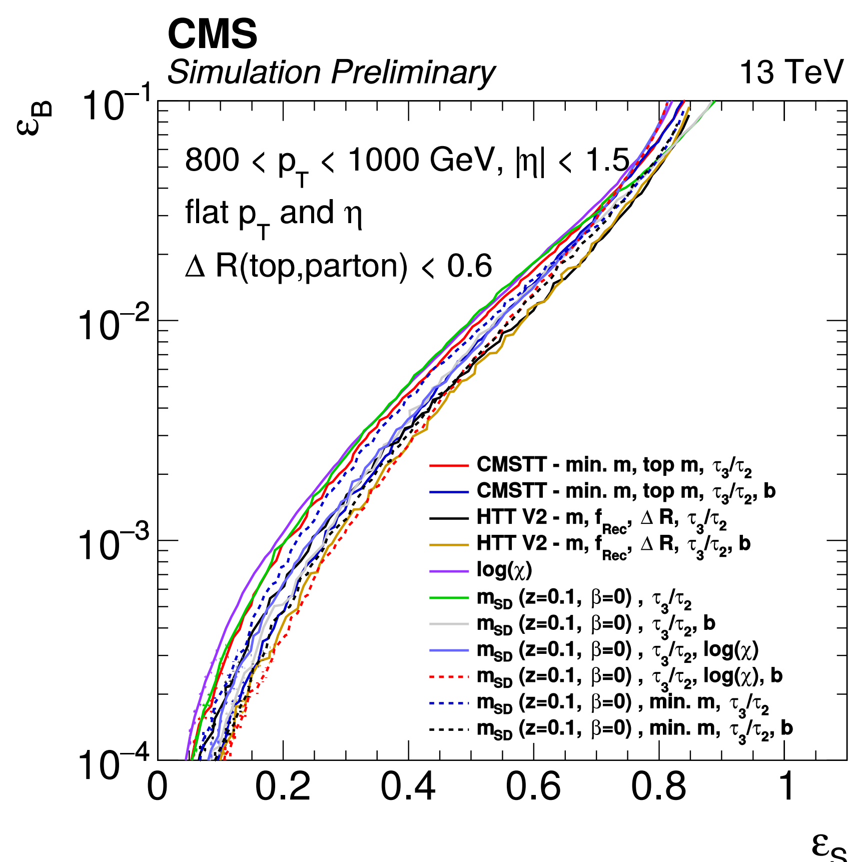


[CMS-PAS-JME-15-002]

13 TeV simulation studies

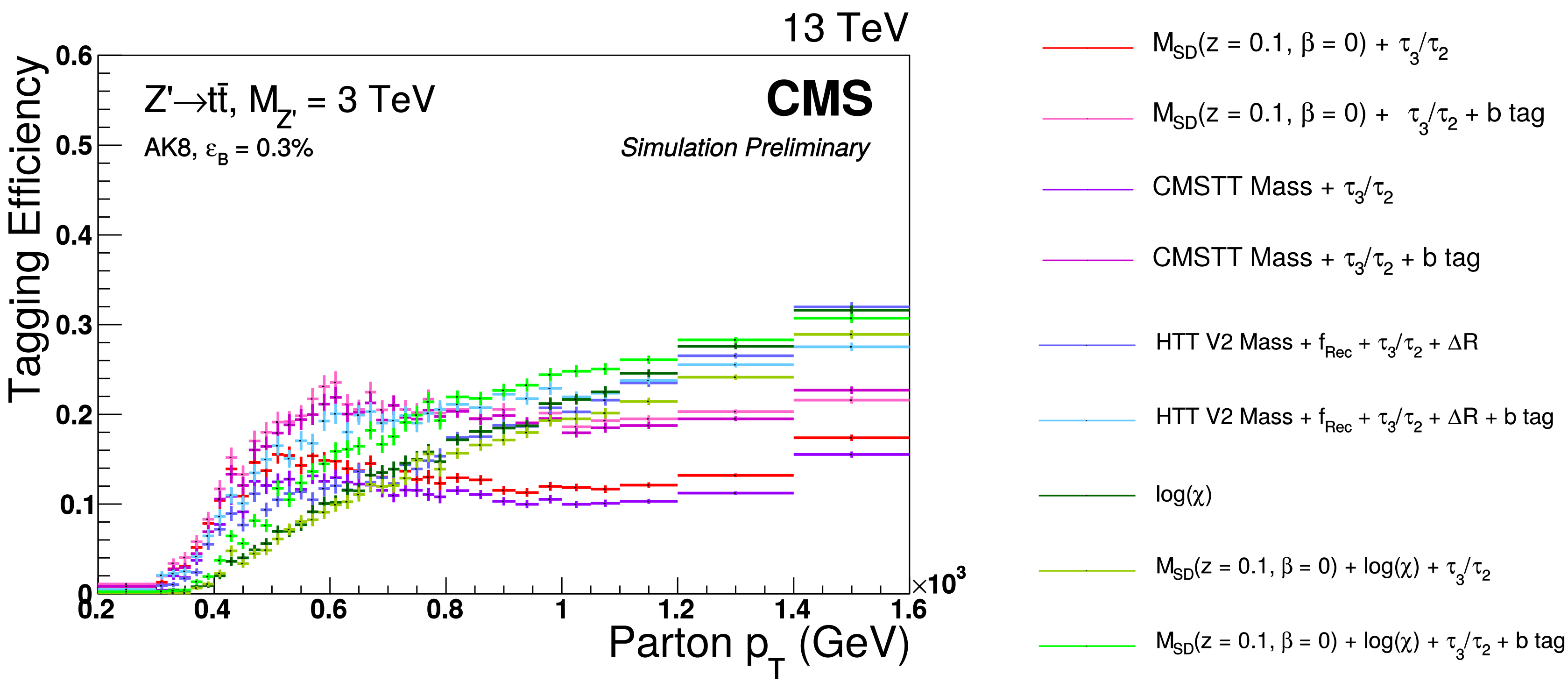
- Different combinations studied
- Similar performance
- HTT V2 good performance for low p_T
- Good performance for m_{SD} + N-subjettiness for high p_T





[CMS-PAS-JME-15-002]

13 TeV simulation studies



[CMS-PAS-JME-15-002]

- m_{SD} + N-subjettiness (+ subjet b-tag)
  - Steep turn on
  - Stable in p_T

→ default option for high p_T region

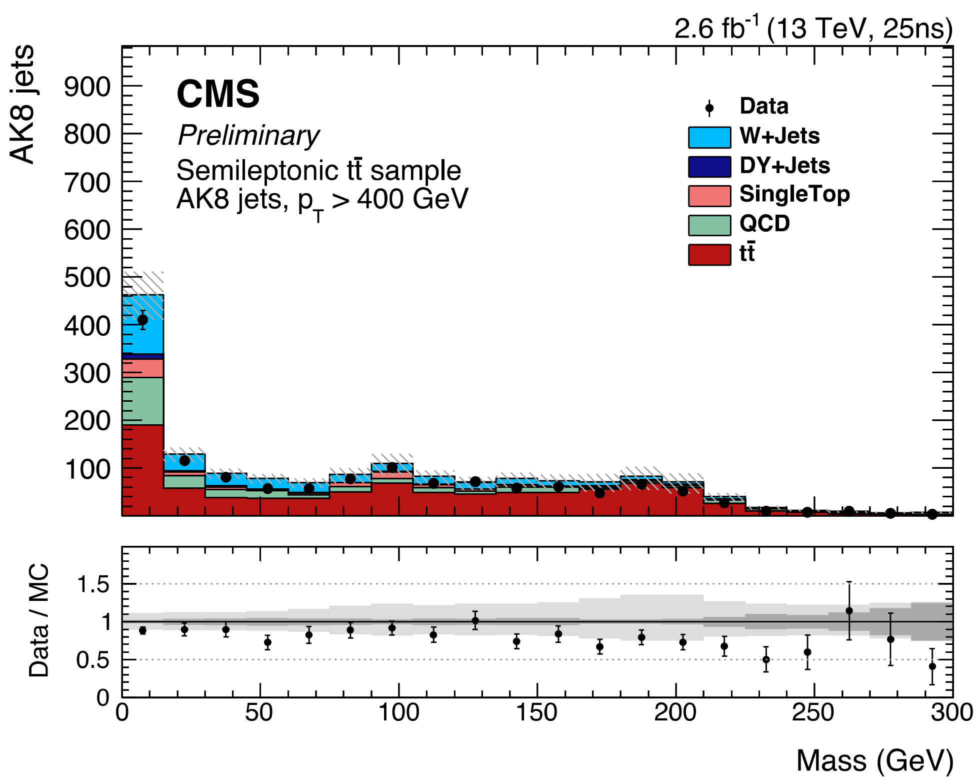
- Additional study of HepTopTagger V2 for low p_T region

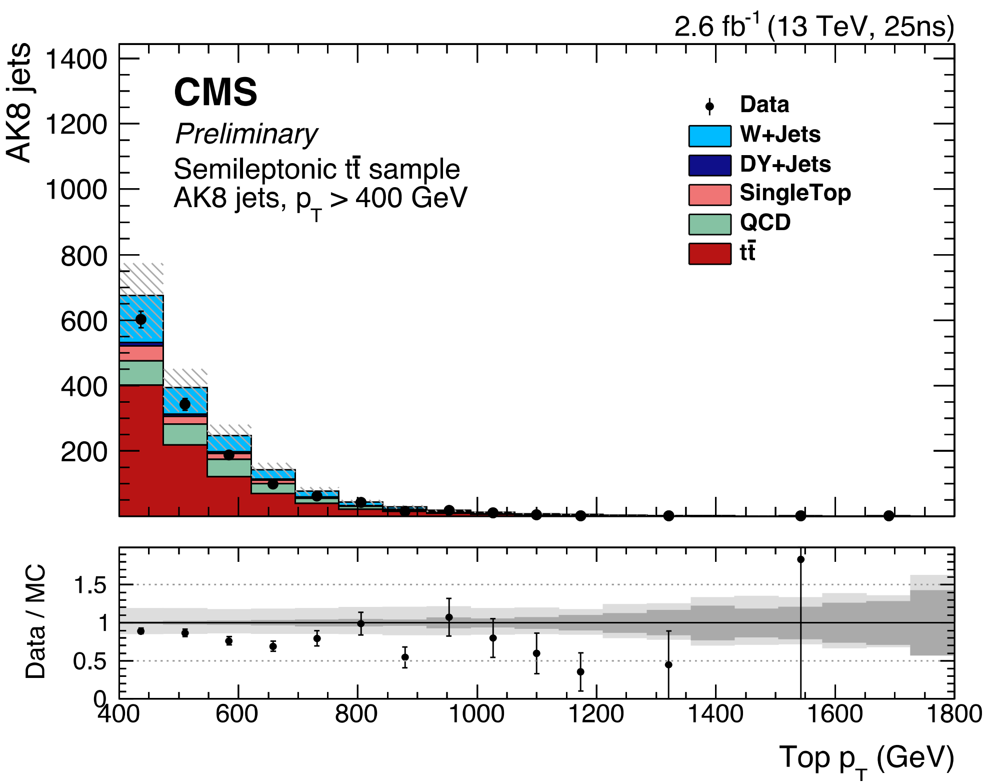
Validation in 2015 data

CMS-PAS-JME-16-003

Validation in 2015 data

- AK8 jets with PUPPI applied
- p_T > 400 GeV
- Soft drop mass and p_T after selection

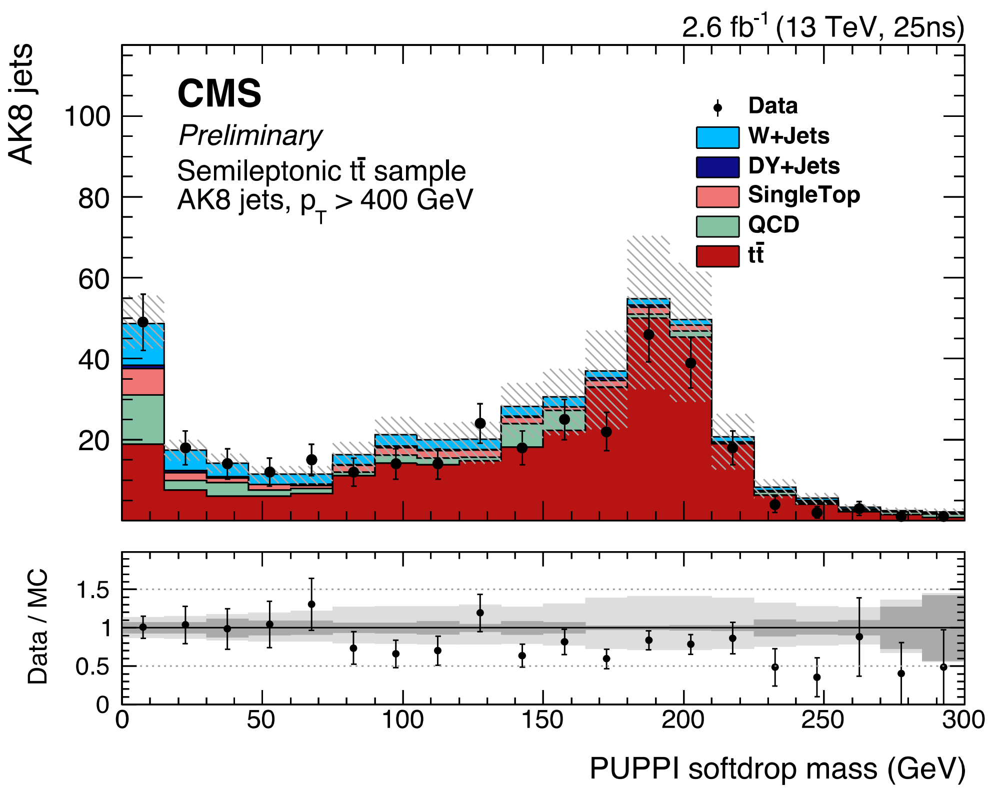


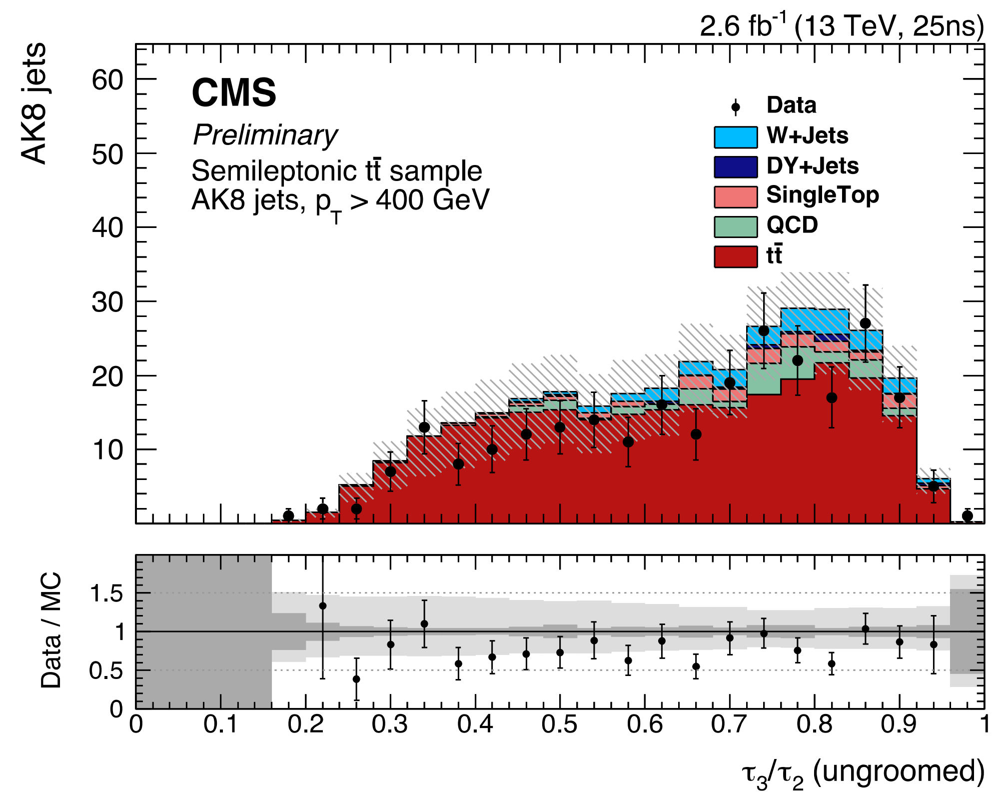


[CMS-PAS-JME-16-003]

Validation in 2015 data

- Loose top tag ($\tau_3/\tau_2 < 0.8$, $105 < m_{SD} < 210$ GeV)
- Cut on respective variable omitted

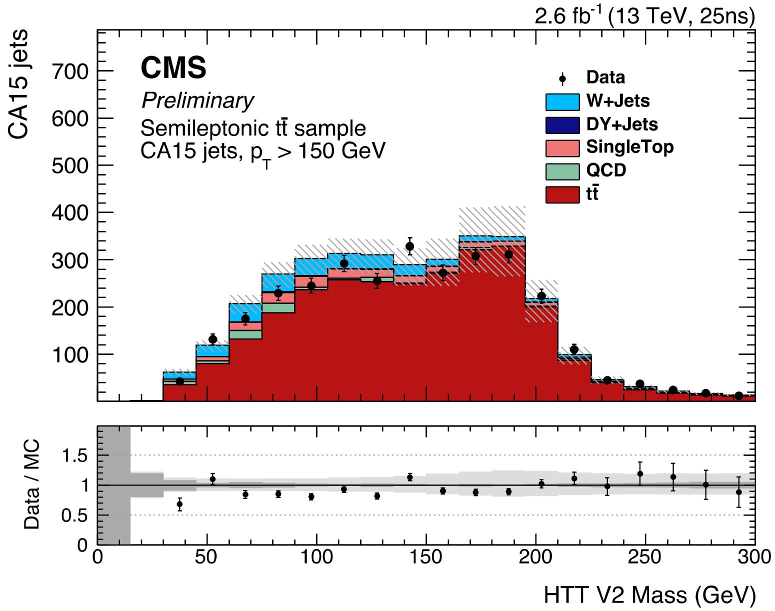




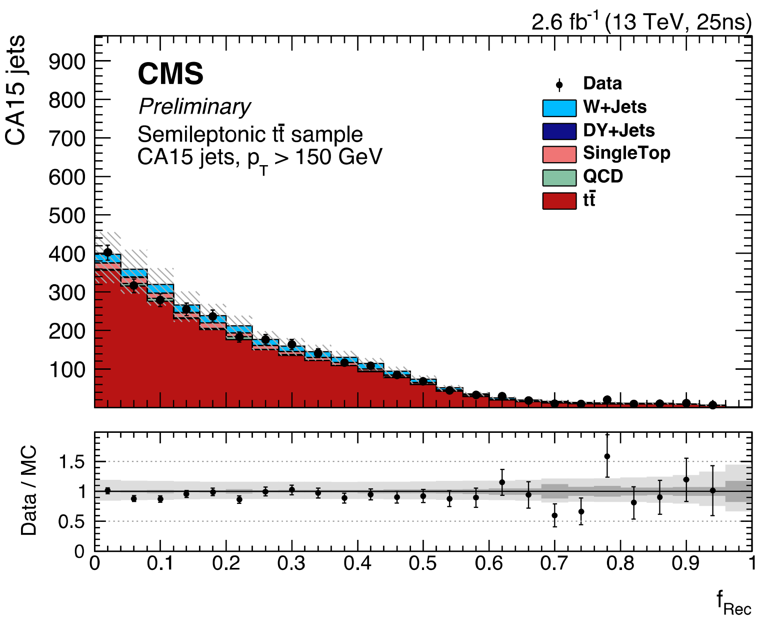
[CMS-PAS-JME-16-003]

Hep Top Tagger V2

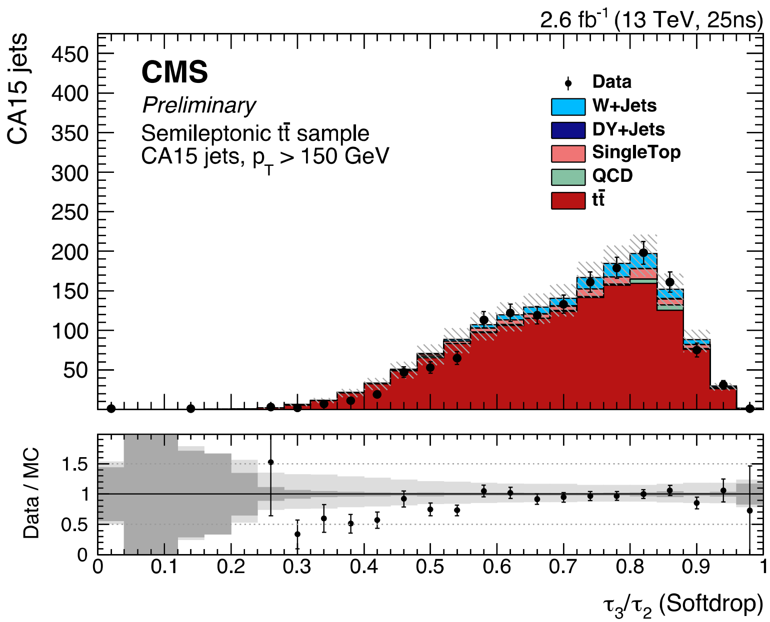
- CA15 jets
- HTT V2 variables after loose top tag
- Cuts on respective variable omitted
- Good description of mass HTT V2 mass and f_{Rec}



[CMS-PAS-JME-16-003]



[CMS-PAS-JME-16-003]



[CMS-PAS-JME-16-003]

2015 scale factors

- Efficiency: $$\epsilon = \frac{N_{\text{tagged jets}} - N_{\text{tagged jets (background)}}}{N_{\text{all jets}} - N_{\text{all jets (background)}}}$$
- Scale factors: sf = $\varepsilon_{data}/\varepsilon_{simulation}$

AK8 jets MSD + N-subjettiness

Input	$\epsilon(B)$	$\epsilon(S)$	p_T bin [GeV]	subjet b-tag > 0.46 400 - 550	550 - ∞	incl. (400 - ∞)	no subjet b-tag cut incl. (400 - ∞)
PUPPI	0.1 %	≈ 15 %	$M_{SD}[105,210]$, $\tau_{32} < 0.46$	0.96 ± 0.18	1.32 ± 0.40	1.04 ± 0.16	0.98 ± 0.15
	0.3 %	≈ 30 %	$M_{SD}[105,210]$, $\tau_{32} < 0.54$	1.02 ± 0.15	1.14 ± 0.32	1.05 ± 0.14	1.01 ± 0.13
	1.0 %	≈ 45 %	$M_{SD}[105,210]$, $\tau_{32} < 0.65$	1.06 ± 0.13	1.07 ± 0.28	1.06 ± 0.12	1.06 ± 0.11
	3.0 %	≈ 55 %	$M_{SD}[105,210]$, $\tau_{32} < 0.80$	1.04 ± 0.11	1.05 ± 0.26	1.05 ± 0.10	1.02 ± 0.09
CHS	0.1 %	≈ 15 %	$M_{SD}[105,220]$, $\tau_{32} < 0.50$	0.76 ± 0.14	1.10 ± 0.30	0.85 ± 0.13	0.86 ± 0.13
	0.3 %	≈ 25 %	$M_{SD}[105,220]$, $\tau_{32} < 0.57$	0.82 ± 0.13	1.00 ± 0.25	0.97 ± 0.11	0.88 ± 0.11
	1.0 %	≈ 45 %	$M_{SD}[105,220]$, $\tau_{32} < 0.67$	0.90 ± 0.11	1.03 ± 0.21	0.94 ± 0.10	0.93 ± 0.09
	3.0 %	≈ 60 %	$M_{SD}[105,220]$, $\tau_{32} < 0.81$	0.88 ± 0.09	1.09 ± 0.19	0.94 ± 0.08	0.96 ± 0.08

[CMS-PAS-JME-16-003]

CA15 jets HTT V2

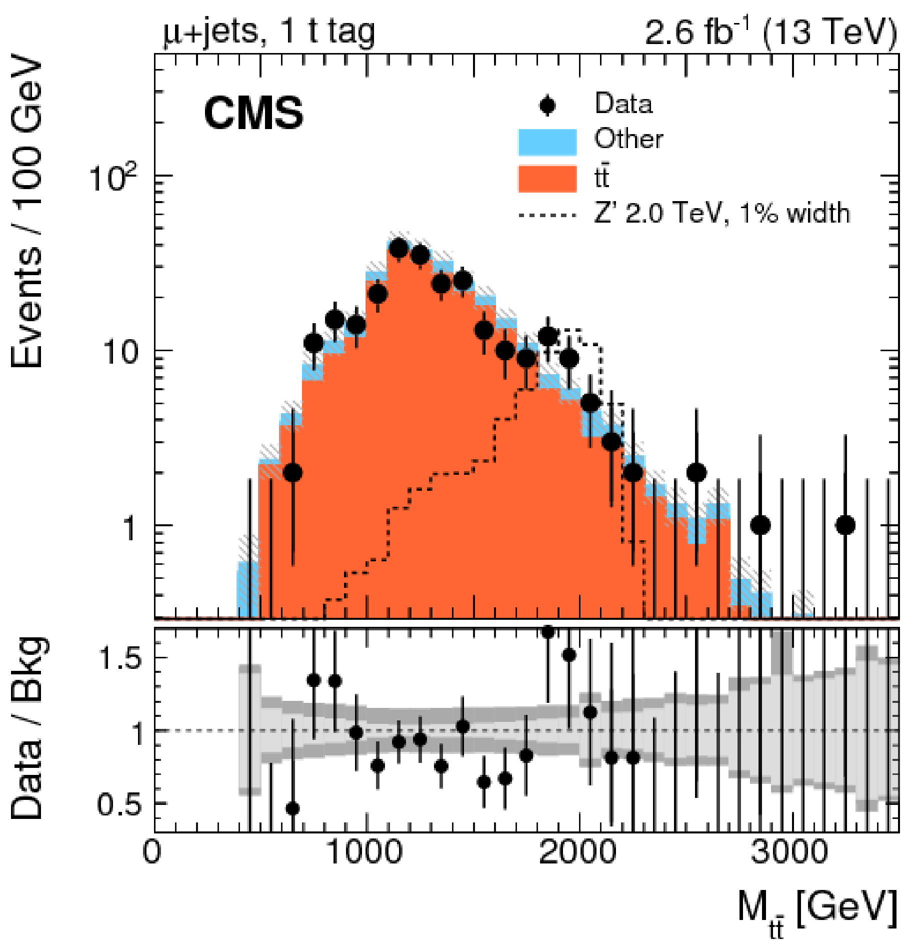
$\epsilon(B)$	$\epsilon(S)$	p_T bin [GeV]	subjet b-tag > 0.46 150 - 400	400 - 550	550 - ∞	incl. (150 - ∞)	no subjet b-tag cut incl. (150 - ∞)
0.1 %	≈ 15 %	M[130,185], $\tau_{32,SD}$ <0.55, f_{Rec} <0.17	0.62 ± 0.06	0.85 ± 0.20	1.47 ± 0.47	0.68 ± 0.06	0.73 ± 0.06
0.3 %	≈ 25 %	M[115,180], $\tau_{32,SD}$ <0.62, f_{Rec} <0.27	0.87 ± 0.05	0.83 ± 0.17	1.26 ± 0.39	0.87 ± 0.05	0.91 ± 0.04
1.0 %	≈ 35 %	M[110,185], $\tau_{32,SD}$ <0.93, f_{Rec} <0.20	0.87 ± 0.03	0.91 ± 0.11	1.50 ± 0.30	0.91 ± 0.03	0.95 ± 0.02
3.0 %	≈ 45 %	M[85,280], $\tau_{32,SD}$ <0.97, f_{Rec} <0.47	0.90 ± 0.02	1.02 ± 0.06	1.20 ± 0.16	0.92 ± 0.01	0.98 ± 0.01

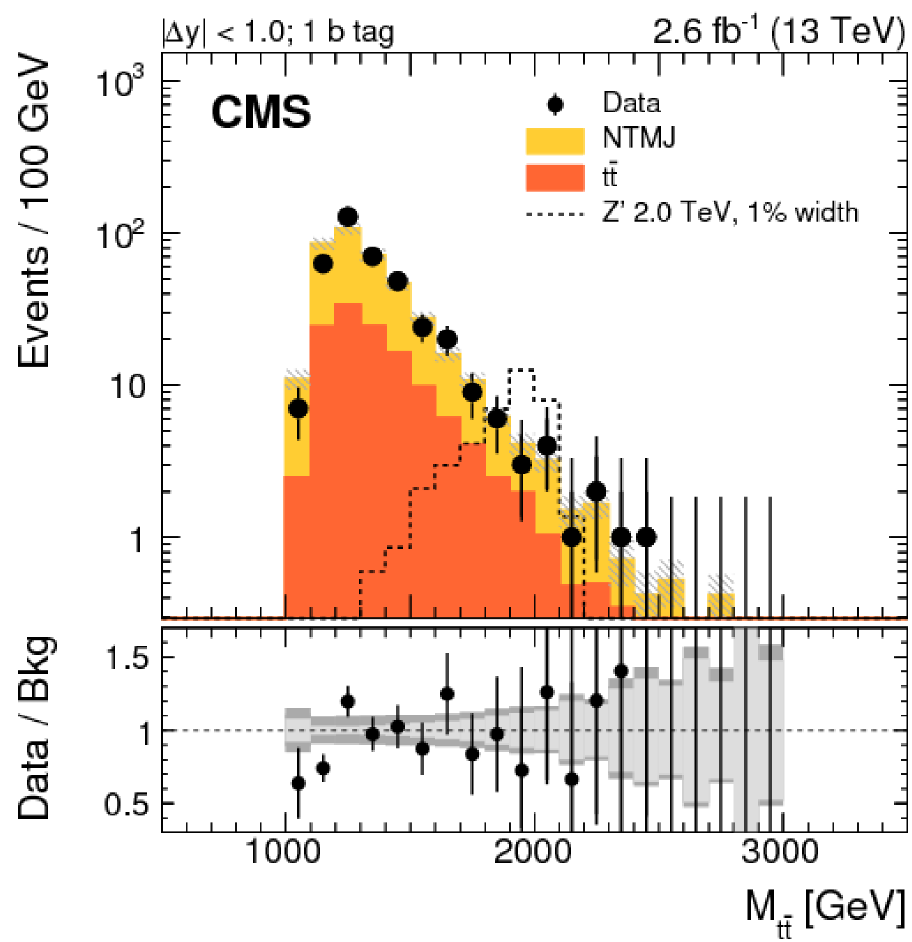
Use of top tagging in 2015 data

$t\bar{t}$ resonance search

- Top tagging used in search for $t\bar{t}$ resonances
- Lepton+jets and fully hadronic channel

[JHEP 1707 (2017) 001]





Validation for $t\bar{t}$ resonances

- Well description of soft drop mass
- Lepton+jets and fully hadronic channel

[JHEP 1707 (2017) 001]