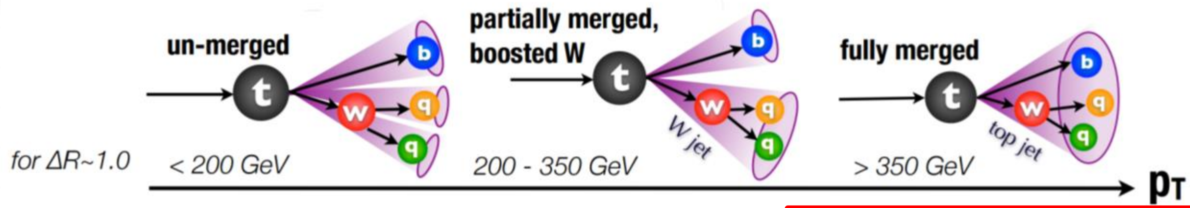
Boosted production at ATLAS and CMS

M. Romano
(INFN – Bologna)
On behalf of the ATLAS and CMS collaborations

LHCP2017
Shanghai, China
15-20 May 2017

Introduction

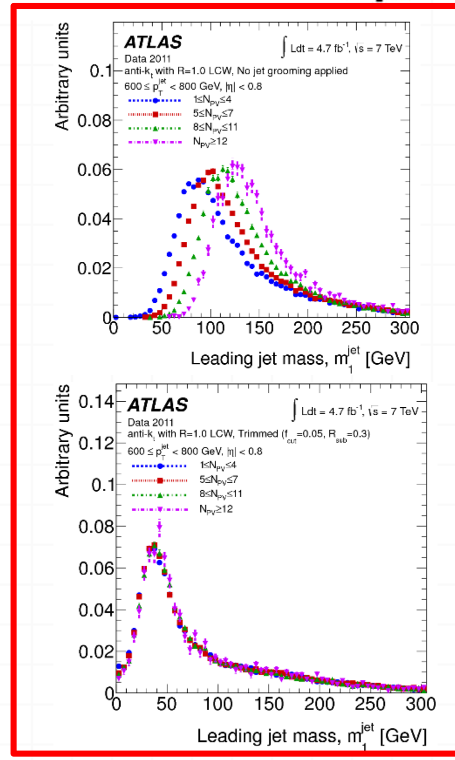


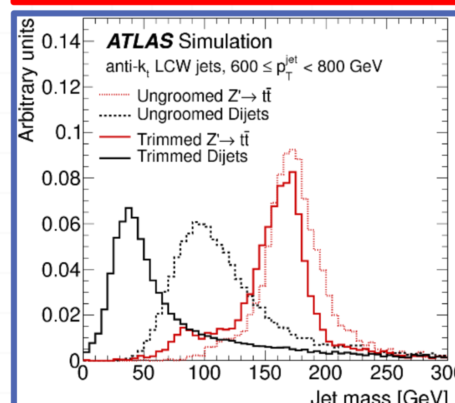
Why boosted tops?

- LHC is a 'top factory'
  - Large center-of-mass energy combined with high luminosity
    - $O(10)$ increase in cross section passing from 8 to 13 TeV
  - Access to phase space regions never explored before
- Feasibility of differential measurements
  - Stronger constraints to SM parameters
  - Sensitivity to BSM processes

A challenging topology...

- Individual top decay products cannot be resolved
  - Boosted tops appear as large-R jets
  - Rule of thumb: $R \sim 2M/p_T$
- Mitigate pileup contamination
  - Jet grooming: remove soft particles coming from pileup
- Top-tagging: exploit substructure properties (like m_{jet}) to reject jets originated by light partons

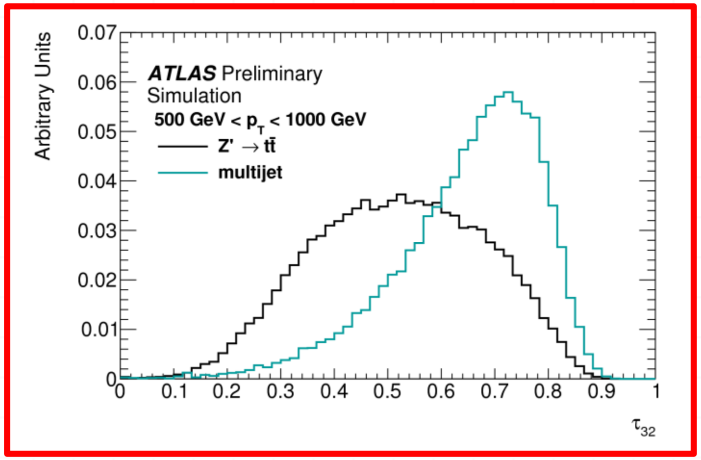




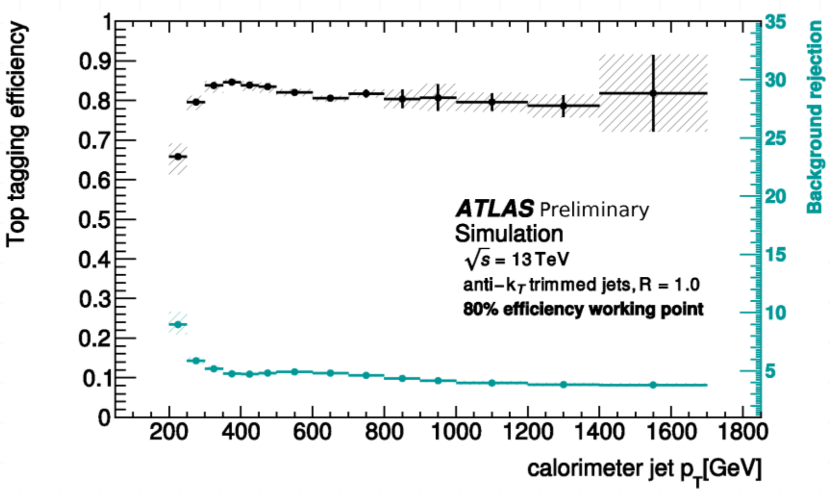
Top tagging in ATLAS

JHEP 1606 (2016) 093
ATL-PHYS-PUB-2015-053

- Large-R jet reconstruction:
  - Anti-kt with $R = 1.0,\ |\eta| < 2,\ p_T > 200$ GeV
  - **Trimming**: subjets with R=0.3-0.2 (at 8 and 13 TeV) and $p_T > 0.05 \cdot p_T^J$ are **removed**
  - Trimmed jet **mass** corrected to particle top jet using MC



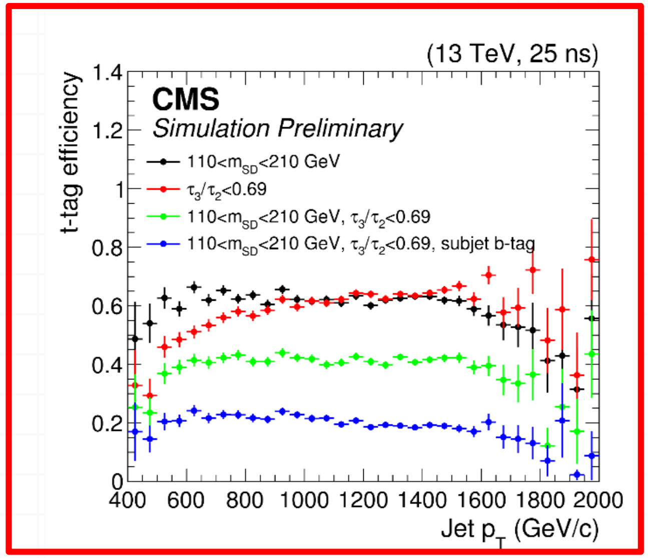
- Top tagging based on substructure variables:
  - **Large-R jet mass**
  - **N-subjettiness**: shape variable related to the hypothesis of having N subjets:
    - $\boldsymbol{\tau_{32}}$ provides discrimination between jets originated by 3 body decays and 2 body decays
  - **Kt splitting scale:** $\sqrt{d_{ij}} = \min(p_{T,i}, p_{T,j}) \cdot \Delta R_{ij}$
    - $\sqrt{d_{12}} \sim m_{top}/2$
    - Used only at 8 TeV



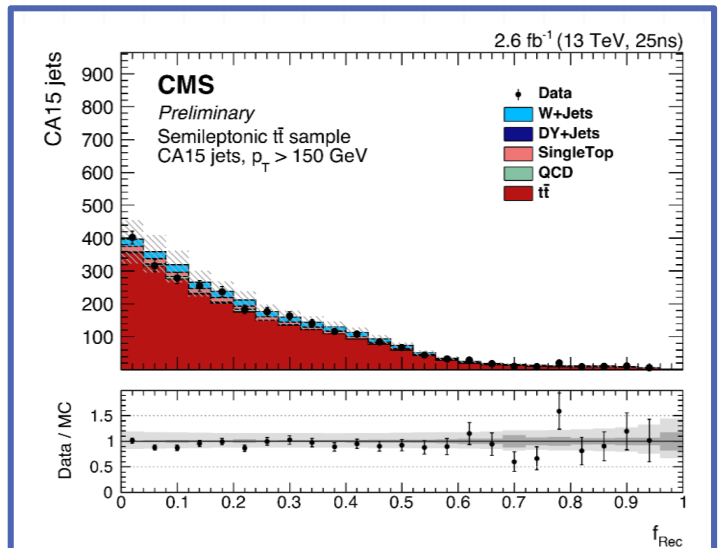
Top tagging in CMS at 13 TeV

CMS-PAS-JME-16-003

- Large-R jet reconstruction:
  - Low p_T (~500 GeV): Cambridge/Aachen particle flow jets with $R = 1.5$ (CA15)
  - High p_T: Anti-kt particle flow jets with $R = 0.8$ (AK8)



- High p_T top tagging:
  - N-subjettiness and softdrop mass ($\beta = 0, z = 0.1$)
- Low p_T top tagging:
  - HEP Top tagger version 2. Discriminating variables:
    - m_{123}: reconstructed top mass from three subjets obtained after a massdrop unclustering
    - Reconstructed W to top mass double-ratio f_{Rec}
  - N-subjettiness with AC15 jets after softdrop ($\beta = 1$, $z = 0.2$)



- b-tagging: Multivariate CSV algorithm
  - For CA15: use the three HTT subjets
  - For AK8: use the subjets after the softdrop mass

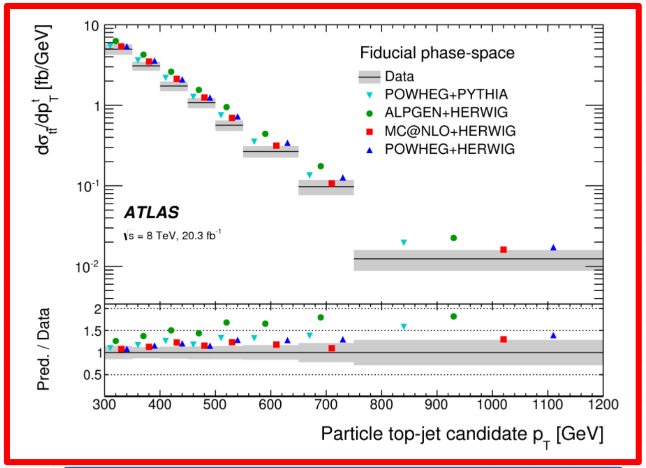
Top cross section measurements

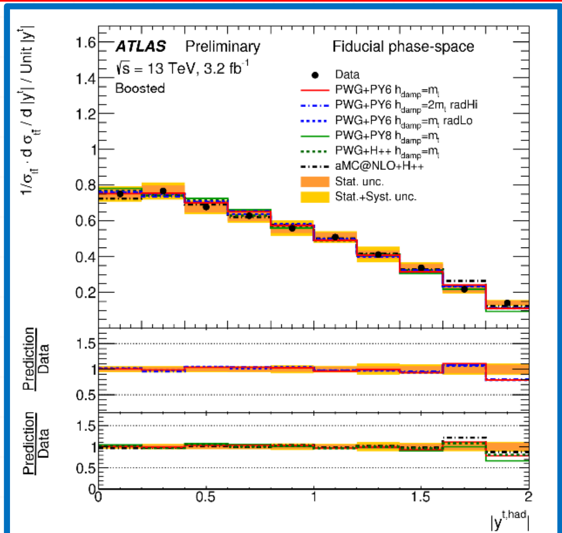
Differential cross section (l+jets) in ATLAS

Phys. Rev. D 93, 032009 (2016)
$\sqrt{s} = 8\ \text{TeV},\ L = 20.3\ \text{fb}^{-1}$

ATLAS-CONF-2016-040
$\sqrt{s} = 13\ \text{TeV},\ L = 3.2\ \text{fb}^{-1}$

- Precisely measure differential cross-section of top pair at high pT:
  - Critical test of Standard Model.
  - Monte Carlo generator tuning, constraints to the PDF of the proton.
  - Sensitive to new physics search / background to BSM.
- **8 TeV**: **Parton** and **particle** level *absolute* differential cross section $\left(\frac{d\sigma}{dp_T^t}\right)$
- **13 TeV**: **Particle** level *absolute* and *relative* differential cross section as a function of p_T^t and $|y_t|$
- Semi-leptonic (e/μ) channel
- Hadronic top tagging:
  - 8 TeV: $m_{jet} > 100$ GeV and $\sqrt{d_{12}} > 40$ GeV
  - 13 TeV: 80% WP based on m_{jet} and τ_{32}
- 13 TeV measurement already systematic-limited
  - main uncertainty: large-R JES
  - 8 TeV: extrapolation to parton level affected by an increased signal modelling uncertainty

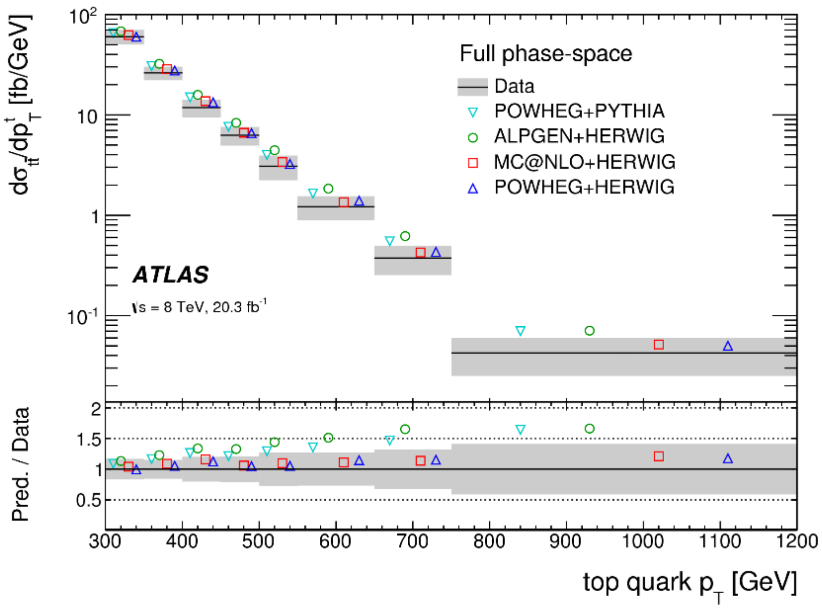




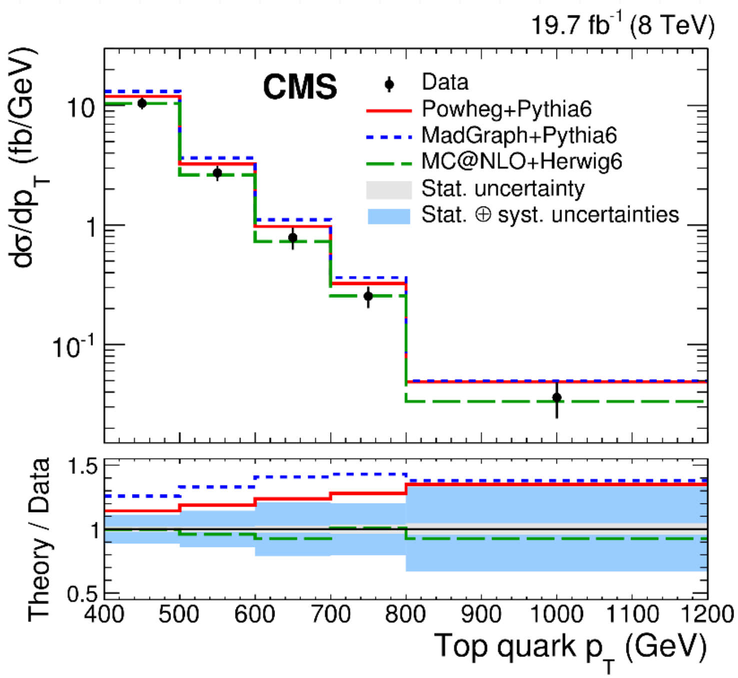
Parton level $d\sigma/dp_T^t$ in ATLAS and CMS

Parton level measurements allow direct comparisons among experiments

Phys. Rev. D 93, 032009 (2016)



Phys. Rev. D 94, 072002 (2016)



Compatible trends observed wrt different generators

Differential cross section (full hadronic) in ATLAS

ATLAS-CONF-2016-100

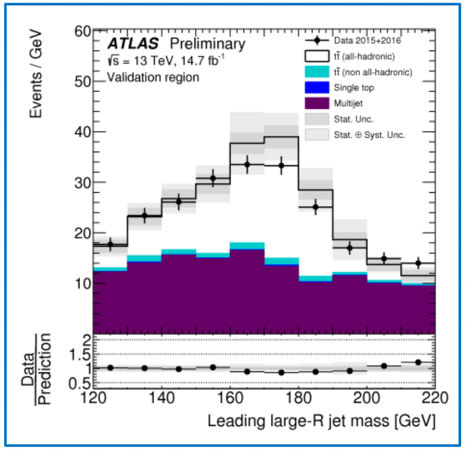
$\sqrt{s} = 13\ \text{TeV},\ L = 14.7\ \text{fb}^{-1}$

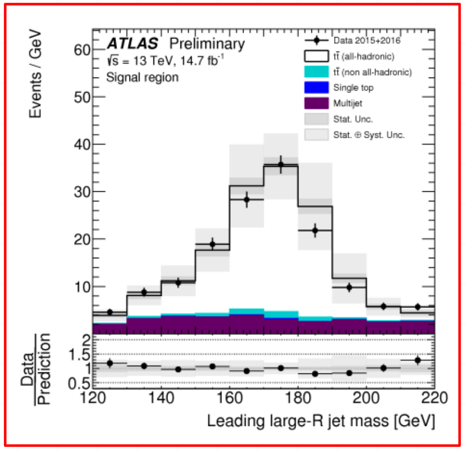
- *Relative* differential cross section in the full had channel
  - Two large-R jet ($p_T^1 > 500$ GeV and $p_T^2 > 350$ GeV, mass in [122.5, 222.5] GeV)
  - Top tagging WP @50%
  - At least two small-R jet (used for b-tagging)
- Main challenge: QCD background estimation
  - Data driven ABCD method

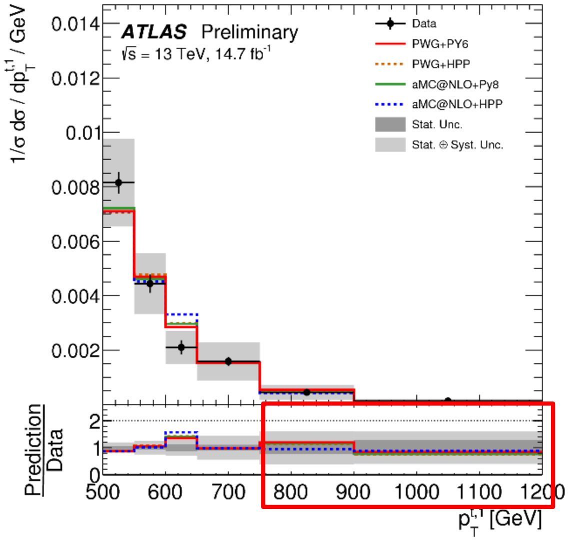
$$S_{QCD} = \frac{1}{2}\left(\frac{G}{A} + \frac{H}{B}\right) \times C$$

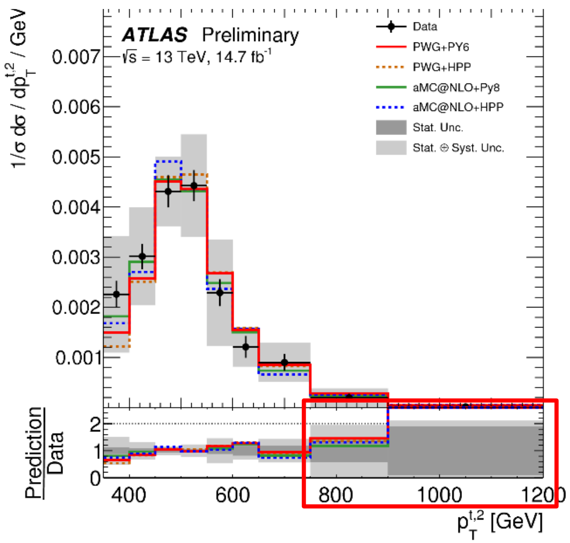
	0 t	1 t	2 t
0 b	A	D	G
1 b	B	E	H
2 b	C	F	S

  - Validated in region F
- Main uncertainties: large-R jes, signal modelling, btag









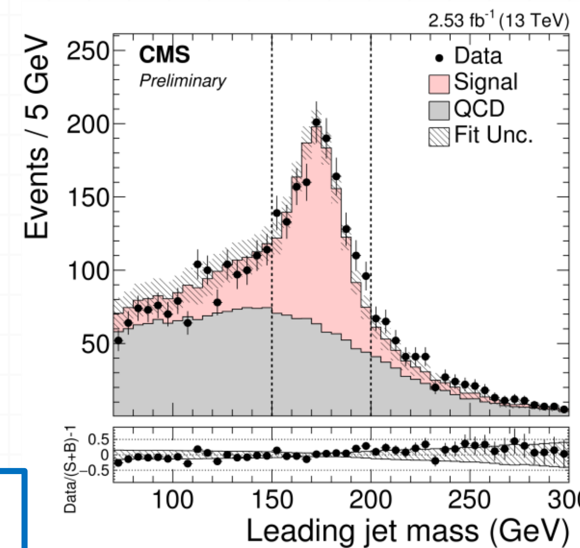
Differential and inclusive cross section (full hadronic) in CMS

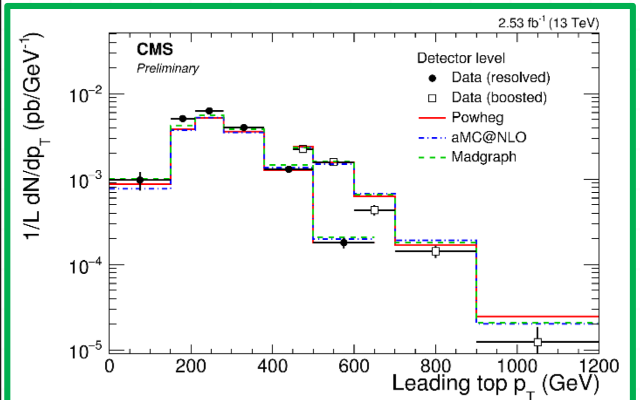
CMS-Top-16-013

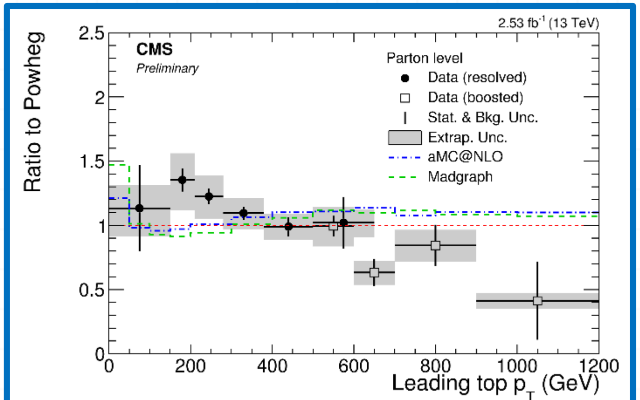
$\sqrt{s} = 13$ TeV, $L = 2.53$ fb^{-1}

Inclusive cross section, *detector* and *parton* level differential cross section

- Two anti-*kt* (R=0.8) jets ($p_T^1 > 450$ GeV and $p_T^2 > 200$ GeV) containing a *b*-subjets
  - Top tagging:
    - Leading jet soft-drop mass $150 < m_{SD} < 200$ GeV
    - Event Fisher discriminant built from τ_{21} and τ_{32}
- Signal and QCD bkg extracted via a fit of m_{SD}
  - QCD templates built from data in control region and corrected with MC to the signal region
- Limited by the statistical, QCD modelling (low pT), JES and bTag uncertainties







$\boldsymbol{\sigma_{tt} = 727 \pm 125}$ **(stat+syst+lumi) pb**

$\sigma_{tt}^{th} = 832^{+20}_{-29}(\text{scale}) \pm 35\ (\text{PDF}+\alpha_s)$ pb

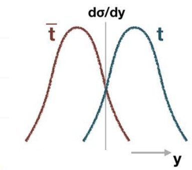
Top properties measurements

Charge asymmetry in ATLAS

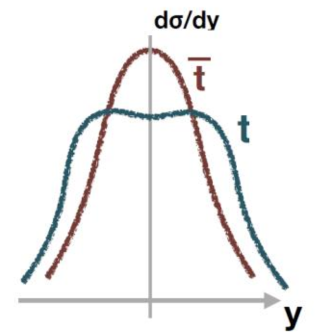
$$A(\Delta) = \frac{N(\Delta > 0) - N(\Delta < 0)}{N(\Delta > 0) + N(\Delta < 0)}$$

Phys. Lett. B756 (2016) 52
$\sqrt{s} = 8$ TeV, $L = 20.3$ fb^{-1}

- Top pair production via $\boldsymbol{q\bar{q}}$ gives non-zero *forward-backward* asymmetry
  - Measured at Tevatron ($\Delta = y_t - y_{\bar{t}}$):
    $A_{FB} = 0.164 \pm 0.047$, $A_{FB}^{MCFM} = 0.073 \pm 0.022$ (Phys. Rev. D 87, 092002)

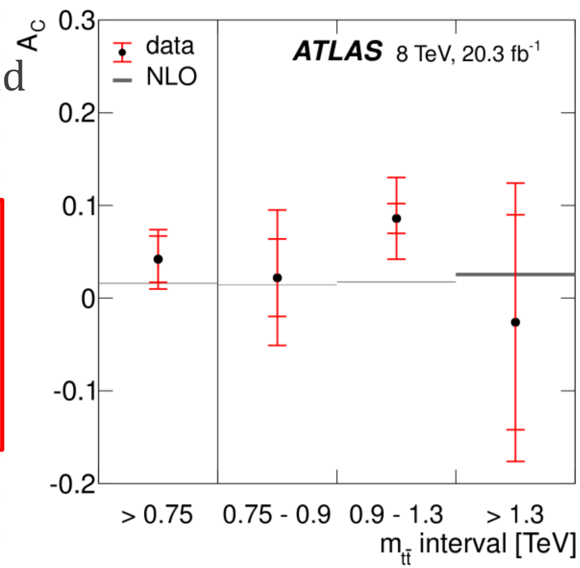


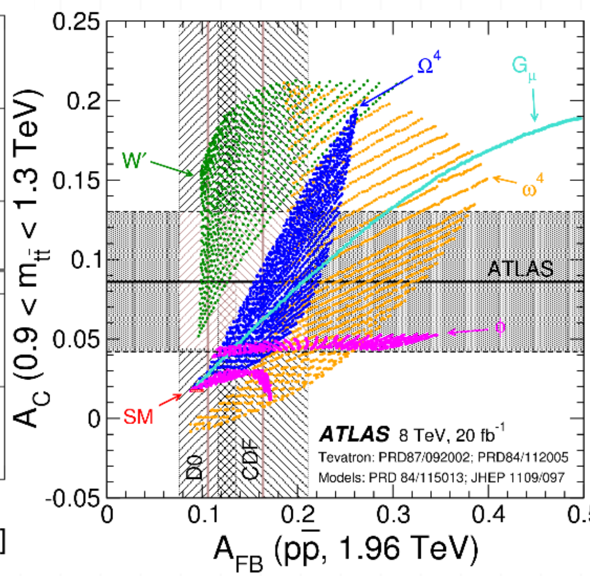
- FB asymmetry not defined at LHC (symmetric collider)
  - A_C (charge asymmetry) can be measured instead ($\Delta = |y_t| - |y_{\bar{t}}|$)
  - Diluted by the dominant $\boldsymbol{gg}$ production



- ATLAS has measured A_c in the boosted l+jets topology
  - Boosted tops can probe A_C at high invariant mass
  - Sensitive to BSM effects
  - Dominated by theoretical and large-R JES uncertainties

‘Inclusive’ measurement
($m_{tt} > 750$ GeV)
$A_C = (4.2 \pm 3.2)\%$
$A_C^{NLO} = (1.60 \pm 0.04)\%$

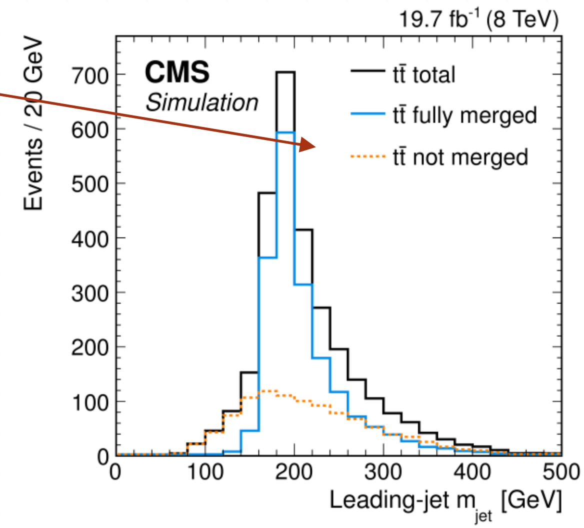




Jet mass distribution in CMS

arXiv:1703.06330
$\sqrt{s} = 8$ TeV, $L = 19.7$ fb^{-1}

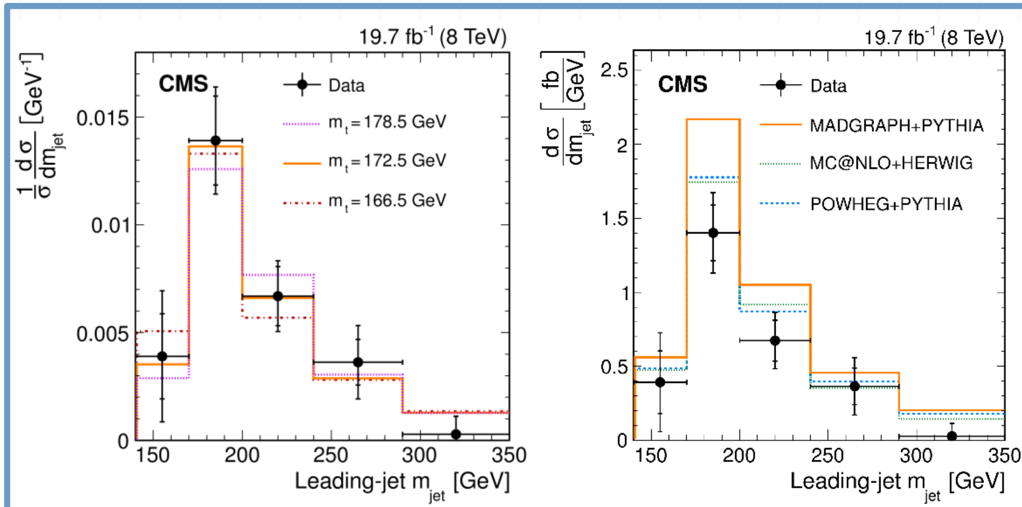
- A better modeling of jet mass m_{jet} distribution is crucial for measurements in the boosted topologies
  - m_{jet} is very sensitive to the angular spread of the jet wrt the top (35% of Pwg+Py events have $\Delta R(jet, top) > 1.2$)
- m_{jet} is sensitive to the top mass → additional way to measure $\boldsymbol{m_t}$



Inclusive and **differential** $\frac{d\sigma}{dm_{jet}}$ cross section measurement

- Exactly two C/A ($R = 1.2$) jets with $p_T > 150$ GeV ($p_T^1 > 400$ GeV)
  - **No top tag**
- Limited by the statistical uncertainty
- m_t extracted from a χ^2 fit of the *relative* differential cross section
  - $\boldsymbol{m_t = 171.8 \pm 9.5}$ **GeV**

Total cross section
($140<m_{jet}<350$ GeV)
Data: 101 ± 19 fb
Powheg+Pythia: 133^{+18}_{-28} fb
Madgraph+Pythia: 159^{+17}_{-18} fb

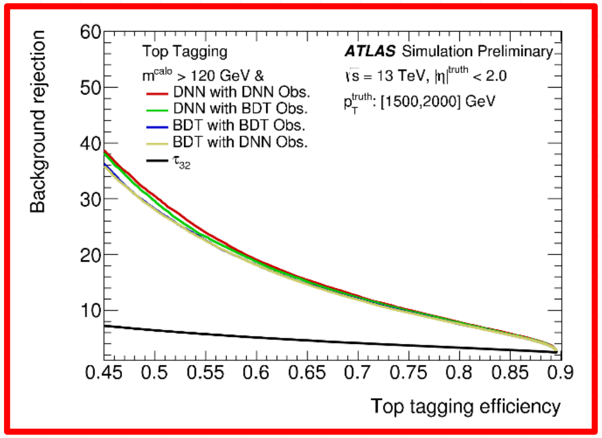


Summary

LHC offers a unique opportunity to explore extreme topologies through boosted tops

- Several boosted top reconstruction algorithms have been (and are being) developed by ATLAS[1][2] and CMS[3]
  - Jet grooming procedures allow for stability in high pileup conditions
  - Use of substructure variables improves the background discrimination
- Boosted tops have been used in SM measurements and BSM searches (not presented in this talk)
  - Lepton+jets differential cross section in ATLAS[4][5] and CMS[6]
  - Full hadronic differential and inclusive cross section in ATLAS[7] and CMS[8]
  - Lepton+jets charge asymmetry in ATLAS[9]
  - Jet mass distribution in CMS[10]

ATL-PHYS-PUB-2017-004



We are entering an era where data statistics is not the limiting factor

- Main uncertainties: large-*R* jets and signal modelling
- The measurements are "**self improving**" → can be used to improve future analyses
  - By adding better constraints to the generator parameters
  - By improving the understanding of the jet mass distribution
  - Stay tuned for new and improved measurements using the full 13 TeV data!

(1) JHEP 1606 (2016) 093
(2) ATL-PHYS-PUB-2015-053
(3) CMS-PAS-JME-19-003
(4) Phys. Rev. D 93, 032009 (2016)
(5) ATLAS-CONF-2016-040
(6) Phys. Rev. D 94, 072002 (2016)
(7) ATLAS-CONF-2016-100
(8) CMS-Top-16-013
(9) Phys. Lett. B756 (2016) 52
(10) arXiv:1703.06330

Backup

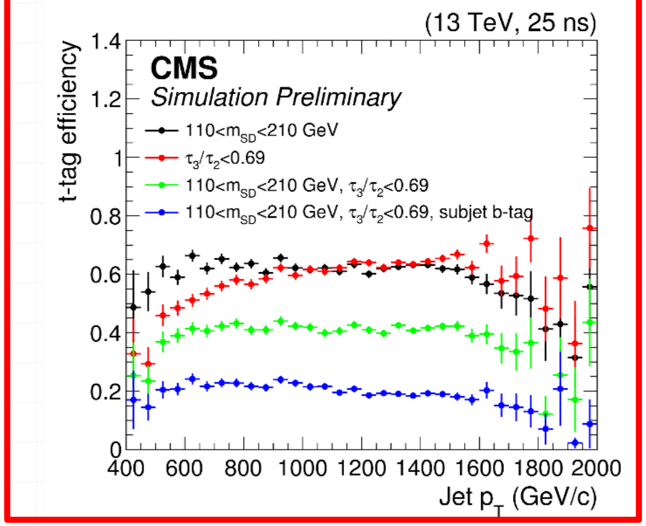
Top tagging in CMS at 13 TeV

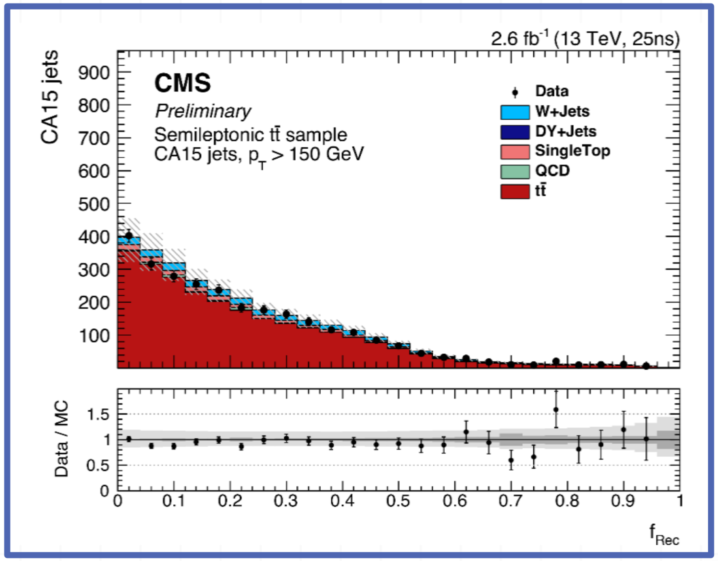
CMS-PAS-JME-16-003

- Large-R jet reconstruction:
  - Low p_T (~500 GeV): Cambridge/Aachen particle flow jets with $R = 1.5$ (CA15)
  - High p_T: Anti-kt particle flow jets with $R = 0.8$ (AK8)

- High p_T top tagging:
  - N-subjettiness and softdrop mass ($\beta = 0, z = 0.1$)
    - Recursive pair-wise declustering, rejecting the softer jets not satisfying $\frac{\min(p_T^1, p_T^2)}{p_T^1 + p_T^2} > z\left(\frac{\Delta R_{12}}{R_0}\right)^{\beta}$
- Low p_T top tagging:
  - HEP Top tagger version 2. Discriminating variables:
    - m_{123}: reconstructed top mass from three subjets obtained after a massdrop unclustering
    - Reconstructed W to top mass ratio $f_{Rec} = \min_{ij}\left|\frac{m_{ij}/m_{123}}{m_W/m_t} - 1\right|$
  - N-subjettiness with AC15 jets after softdrop ($\beta = 1, z = 0.2$)
- b-tagging: Multivariate CSV algorithm
  - For CA15: use the three HTT subjets
  - For AK8: use the subjets after the softdrop mass

- Top tagging at 8 TeV based on subjets from primary and secondary decomposition



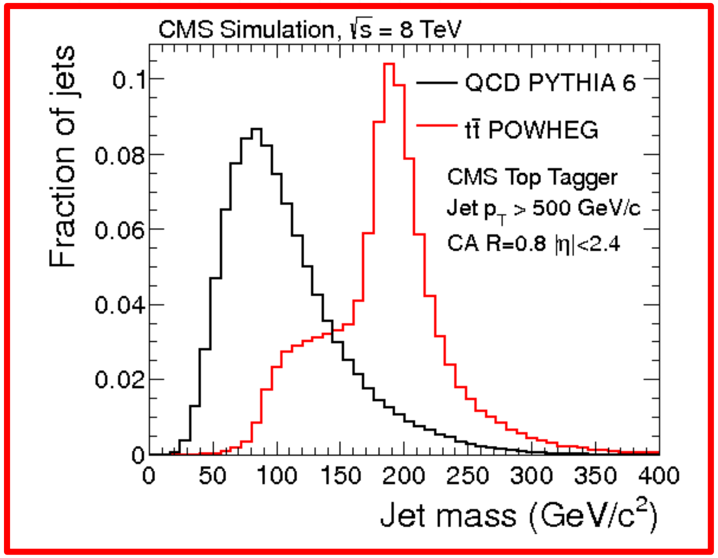


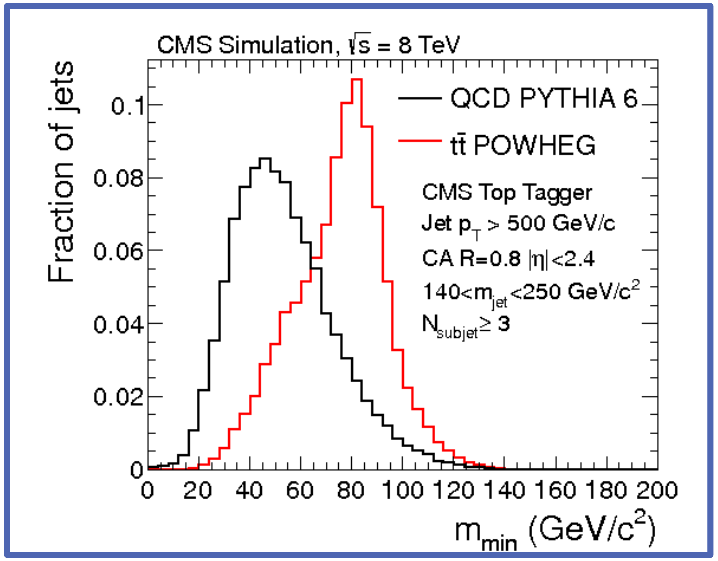
Top tagging in CMS at 8 TeV

- Large-R jet reconstruction:
  - Cambridge/Aachen (C/A) jets with $R = 0.8$, $|\eta| < 2.4$, $p_T > 350$ GeV
- Top tagging algorithm:
  - Primary decomposition: recursively declusters the jet to find two well separated hard subclusters
  - Secondary decomposition: declusters the previously found subclusters
- Top tagging based on subcluster variables:
  - **Jet mass**
  - **Number of subclusters**
  - **Minimum Pairwise Mass** among the three hardest subjets:

$$m_{min} = \min(m_{12}, m_{13}, m_{23})$$

- Other top tagging algorithms: N-subjetness and HEP Top tagger

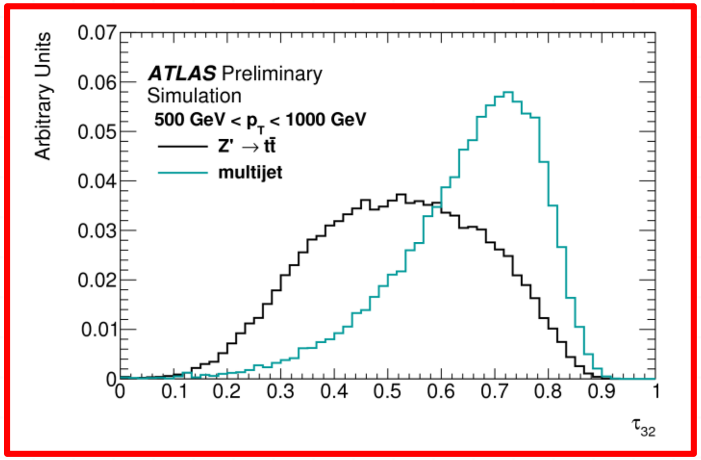




Top tagging in ATLAS

JHEP 1606 (2016) 093
ATL-PHYS-PUB-2015-053

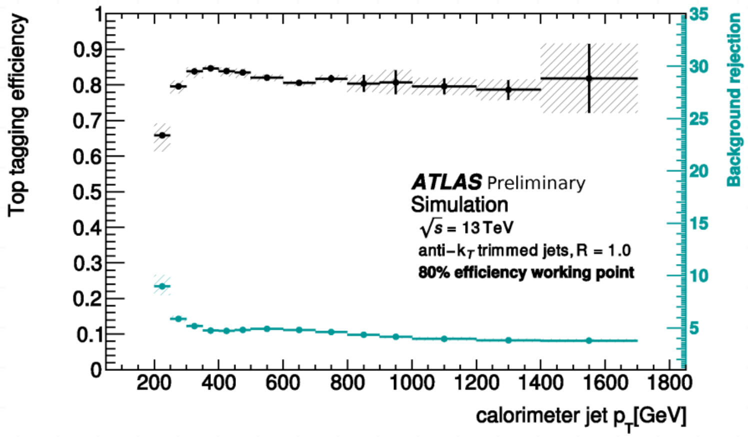
- Large-R jet reconstruction:
  - Anti-kt with $R = 1.0,\ |\eta| < 2,\ p_T > 200$ GeV
  - **Trimming**: subjets with R=0.3-0.2 (at 8 and 13 TeV) and $p_T > 0.05 \cdot p_T^J$ are **removed**
  - Trimmed jet **mass** corrected to particle top jet using MC



- Top tagging based on substructure variables:
  - **Large-R jet mass**
  - **N-subjettiness**: shape variable related to the hypothesis of having N subjets:

$$\tau_N = \frac{\sum_{i=1}^{n_{costintuents}} p_{T,i}\ \Delta R_i^{min}}{\sum_{i=1}^{n_{costintuents}} p_{T,i} R}$$

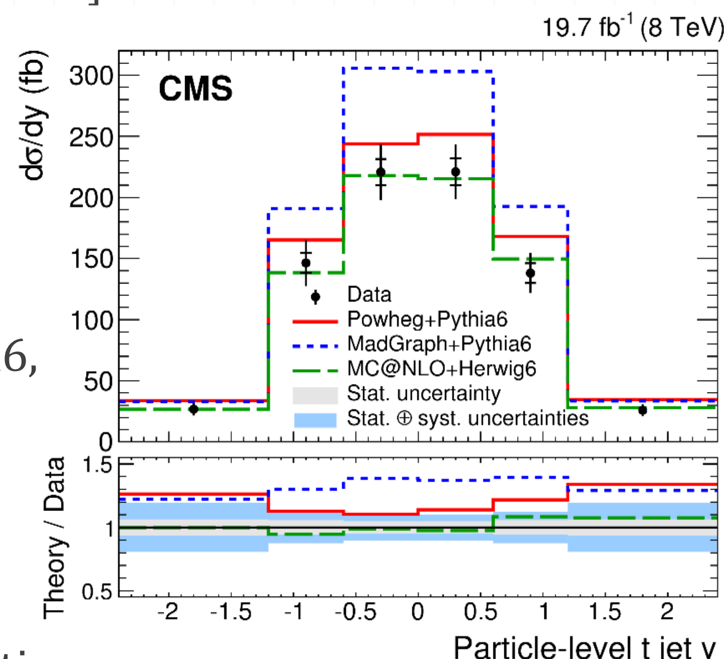
    - $\boldsymbol{\tau_{32} = \tau_3/\tau_2}$ provides discrimination between jets originated by 3 body decays and 2 body decays
  - **Kt splitting scale:** $\sqrt{d_{ij}} = \min(p_{T,i}, p_{T,j}) \cdot \Delta R_{ij}$
    - $\sqrt{d_{12}} \sim m_{top}/2$
    - Used only at 8 TeV



Differential and inclusive cross section (l+jets) in CMS

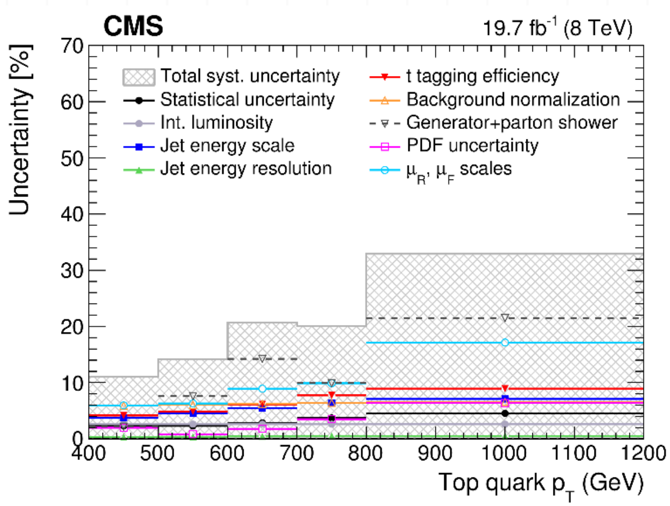
Phys. Rev. D 94, 072002 (2016)
$\sqrt{s} = 8$ TeV, $L = 19.7$ fb^{-1}

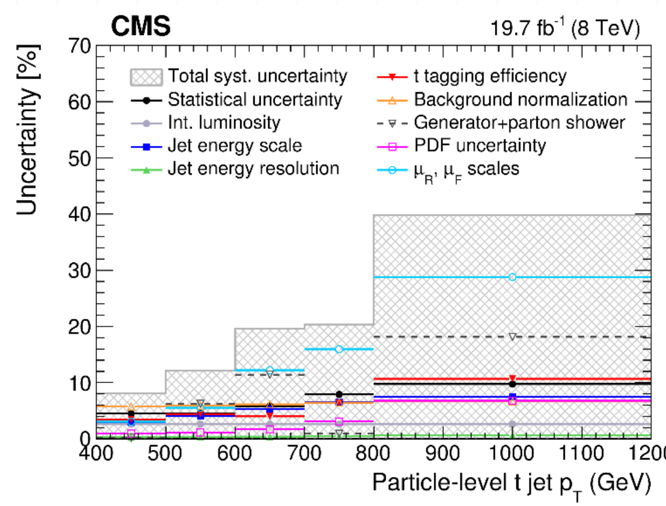
- **Integrated** ($p_T > 400$ GeV) and **differential** $\left(\frac{d\sigma}{dp_T^t}\right.$ and $\left.\frac{d\sigma}{dy^t}\right)$ cross section at **particle** and **parton** level
- Event selection 1 lepton (e/μ) + jets
  - C/A large-R jet ($R = 0.8$), $p_T > 400$ GeV; mass ~ [140, 250] GeV
  - CMS Top Tagger
- Signal yield extracted via maximum likelihood fit in (0t, 1t+0b, 1t+1b) exclusive categories
  - Background normalizations and uncertainties treated as nuisance parameters
- Inclusive cross section compared to Powheg+Pythia6, normalized to NNLO total cross section
  - Parton level: $\frac{\sigma_{meas}}{\sigma_{th}} = 0.86 \pm 0.19$
  - Particle level: $\frac{\sigma_{meas}}{\sigma_{th}} = 0.86 \pm 0.16$
- Dominated by theoretical and top tagging uncertainties



Ljets differential cross section in CMS

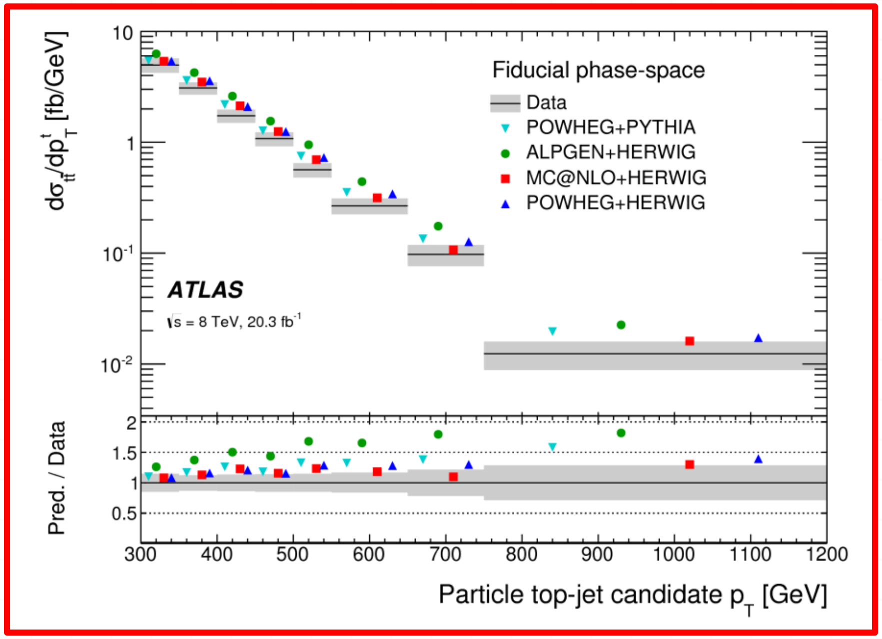
- signal extraction - maximum likelihood fit in 3 categories based on top and b tag
  - Signal and all background yields determined by fit
  - Discriminant variables: lepton |eta| used in (0t, 1t+0b), mvtx used in 1t+1b
  - Background normalizations and experimental uncertainties treated as nuisance parameters.
- Unfolding in 2 steps: reco → particle, particle → parton
  - Regularized unfolding (SVD)
  - Electron and muon channel unfolded separately and combined with weighted mean

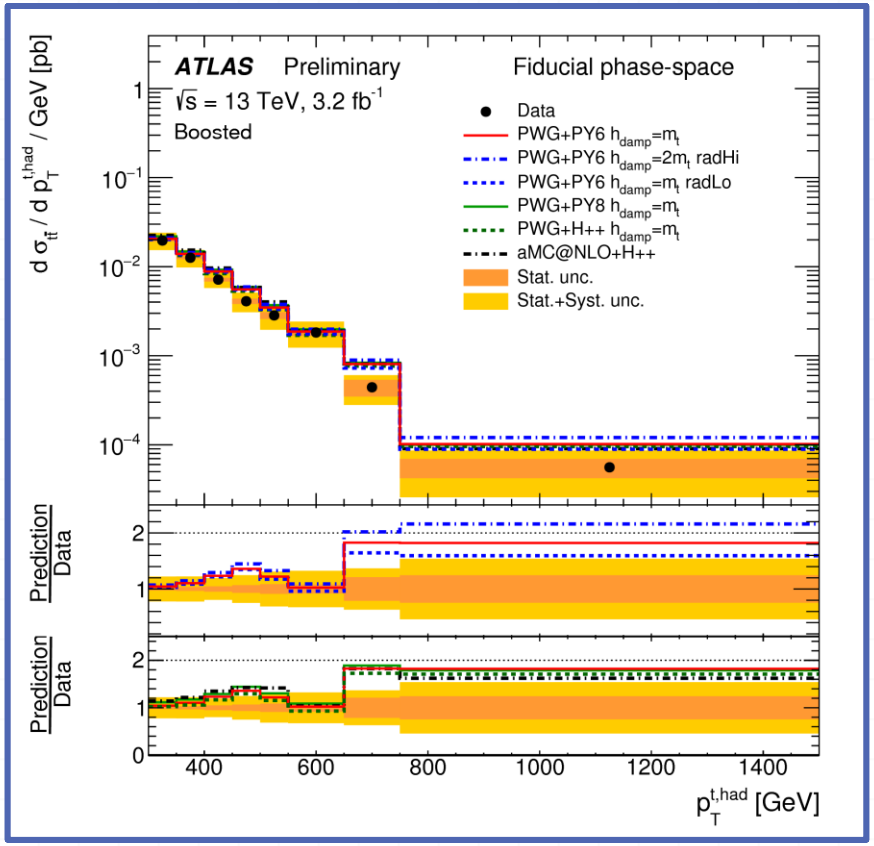




Ljets differential cross section in ATLAS

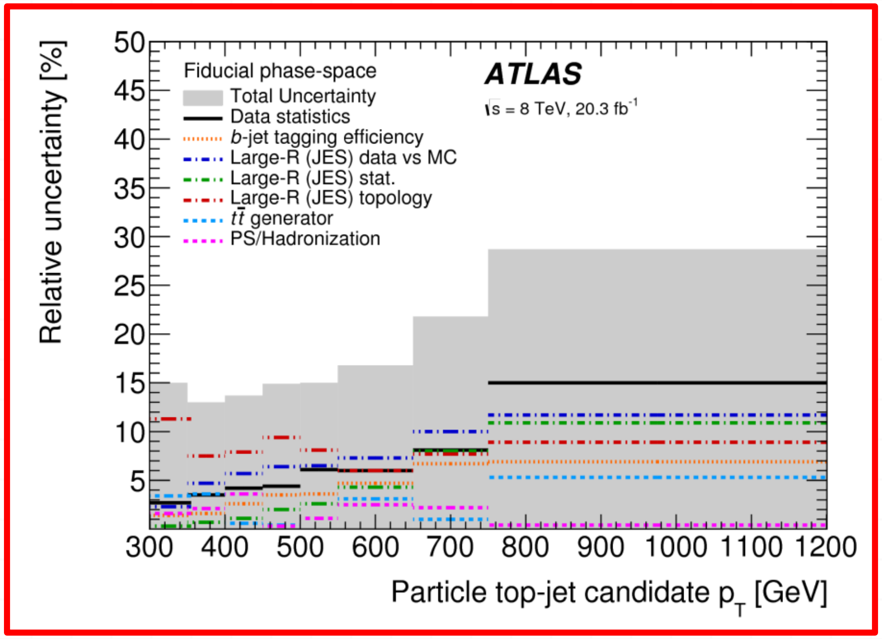
- 8 TeV vs 13 TeV comparison

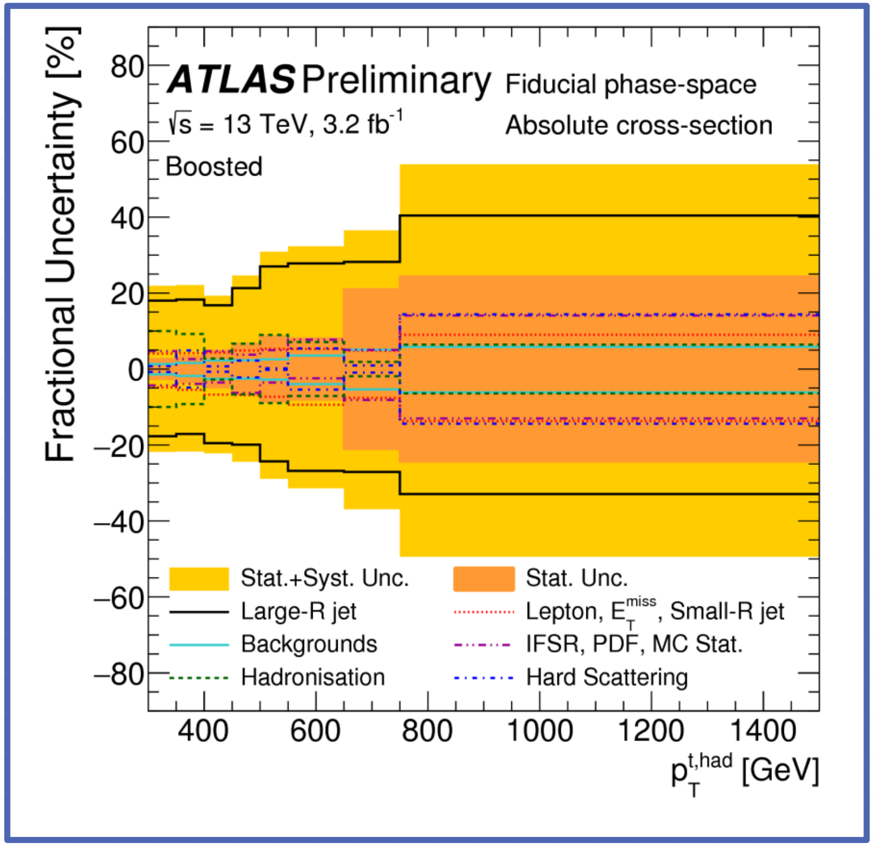




Ljets differential cross section in ATLAS

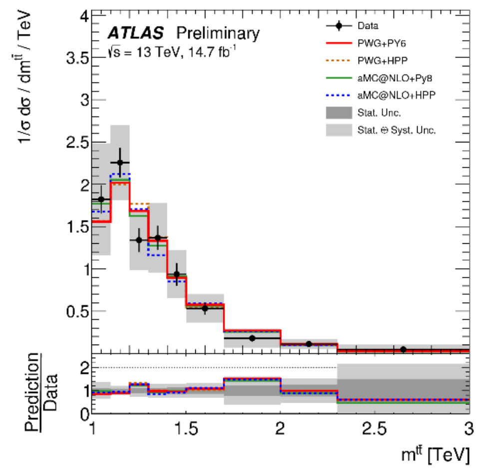
- 8 TeV vs 13 TeV uncertainty comparison

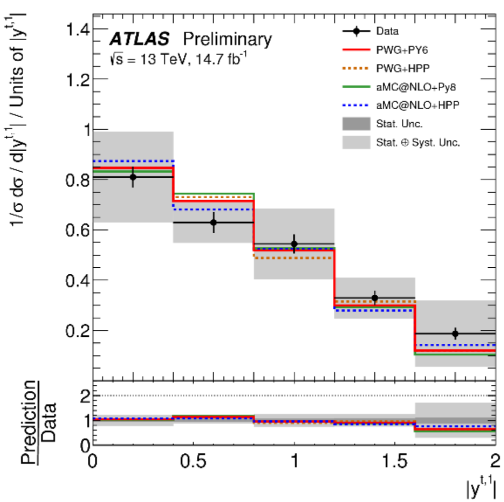


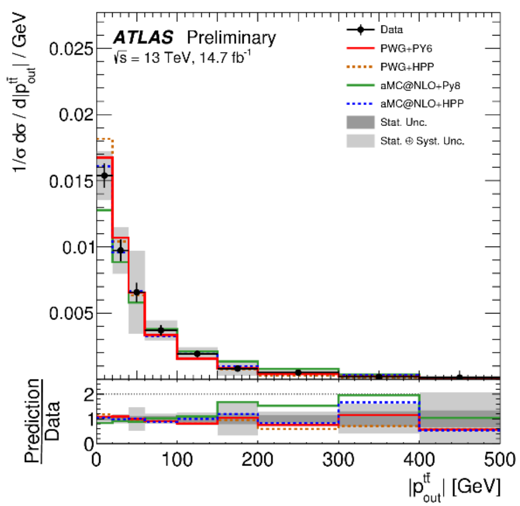


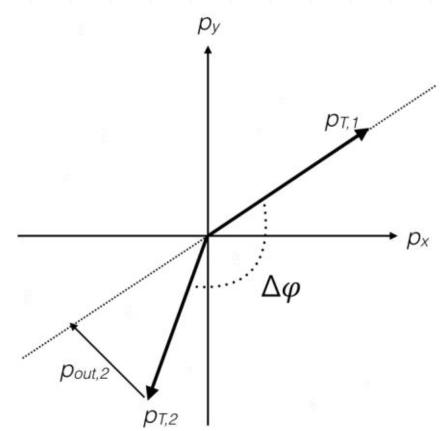
Full had cross section in ATLAS

- Differential cross section measurements performed as a function of several kinematic variables of the tt system:
- $p_T, |y|$ $(t_1, t_2, t\bar{t})$ and $m(t\bar{t})$
- $\chi^{tt} = \exp 2|y^*|$ (y^*: rapidity of the top in the $t\bar{t}$ rest frame)
- $Y_B^{tt} = \frac{1}{2}(y^{t_1} + y^{t_2})$: longitudinal boost in the lab frame
- $\Delta\phi^{tt}$ azimuthal angular separation between the tops
- $|p_{out}^{tt}|$: projection of top-quark momentum onto the direction perpendicular to a plane defined by the other top quark and the beam



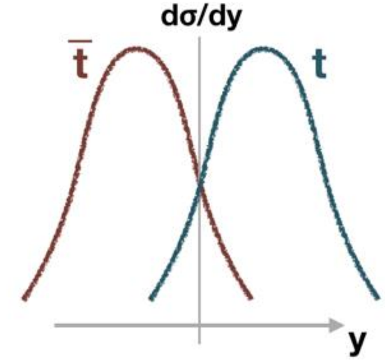






Forward-backward asymmetry at Tevatron

Definition: $A_\Delta = \frac{N(\Delta>0)-N(\Delta<0)}{N(\Delta>0)+N(\Delta<0)}$



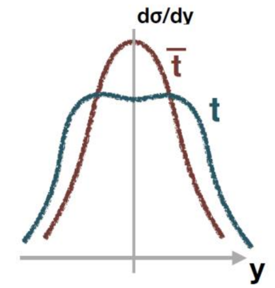
@ Tevatron leading production is $q\bar{q}$ (asimmetric initial state)
At Tevatron $\Delta_{FB} \equiv y_t - y_{\bar{t}}$

CDF measured $A_{FB} = 0.158 \pm 0.074$
(both *l*+jets and di-leptonic channels)
$A_{FB}^{\mathrm{MCFM}} = 0.058 \pm 0.009$ (agreement ~2σ)

Important features: dependence on $m_{t\bar{t}}$, Δy
$m_{t\bar{t}}$< 450 GeV compatible with SM ~ 1.8 sigma
$m_{t\bar{t}}$ > 450 GeV difference >3 sigma

Superseded by $A_{FB} = 0.160 \pm 0.045$
$A_{FB}^{NNLO} = 0.095 \pm 0.007$
(CDF/ANAL/TOP/PUB/11161)

Charge asymmetry at LHC



- @ LHC leading production channel is *gg* in *symmetric pp* collisions
- Forward-backward is not visible at the LHC, but we can measure A_Q (charge asymmetry) in the central and forward region
- $A_Q = \frac{N(\Delta|y|>0) - N(\Delta|y|<0)}{N(\Delta|y|>0) + N(\Delta|y|<0)}$
- $A_{QF} = \frac{N_t(|y|>y_0) - N_{\bar{t}}(|y|>y_0)}{N_t(|y|>y_0) + N_{\bar{t}}(|y|>y_0)}$
- $A_{QC} = \frac{N_t(|y|<y_0) - N_{\bar{t}}(|y|<y_0)}{N_t(|y|<y_0) + N_{\bar{t}}(|y|<y_0)}$

A_Q was found to be the most sensitive variable to new physics effects
Tests performed using a parametrized BSM asymmetry:

$$1 - f(m_{tt}) \tanh \Delta y \, , f(m_{tt}) = \frac{m_{tt}}{200} - 2$$

http://www.hep.phy.cam.ac.uk/theory/webber/MCEGforLHC.pdf

Introduction

Why boosted tops?

- LHC is a 'top factory'
  - Large center-of-mass energy combined with high luminosity
    - The cross section in boosted phase space will benefit most from the energy increase
  - Access to phase space regions never explored before
- Feasibility of differential measurements in boosted topologies
  - Stronger constraints to SM parameters

Cross section	8 TeV	13 TeV
Total	240.6 pb[1]	818[3] pb
Boosted (pt>300GeV)	5.5pb[2]	O(10)x

(1) ATLAS-CONF-2014-053, CMS-PAS-TOP-14-016
(2) PRD 93, 032009 (2016)
(3) PLB 761 (2016) 136

This talk:

- Boosted top tagger algorithms in ATLAS and CMS
- ATLAS and CMS precision measurements with boosted quarks
  - Inclusive and differential cross sections at 8 and 13 TeV
  - Top properties: charge asymmetry and jet mass distribution
  - Exotic searches will not be covered