ATLAS Measurements Using Jet Grooming and Substructure

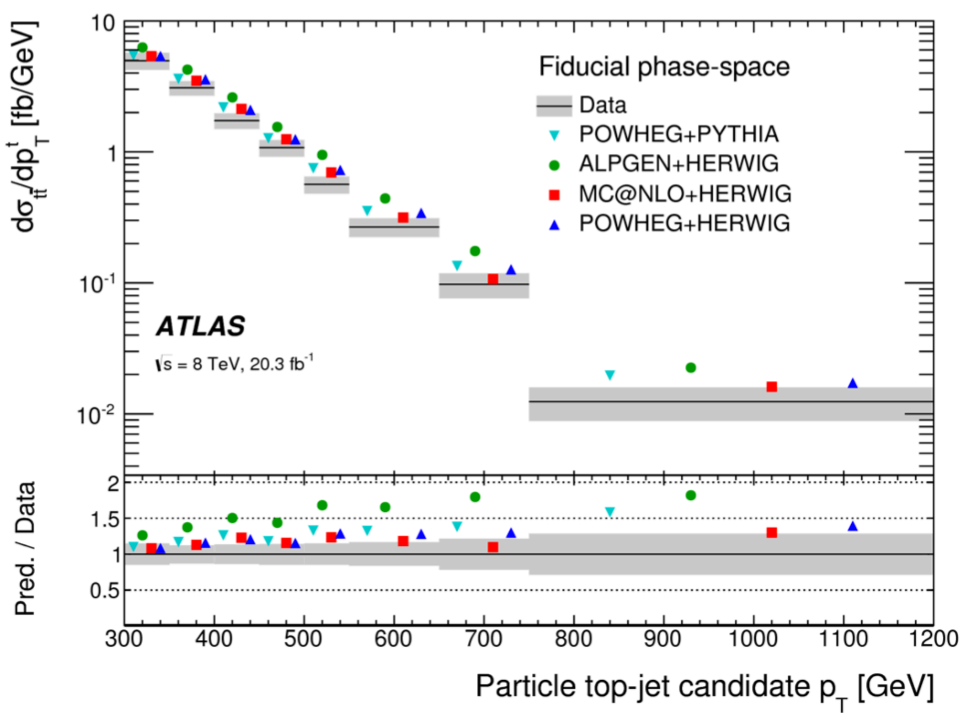
Giulia Ucchielli
(INFN and University of Bologna)

On Behalf of the ATLAS Collaboration

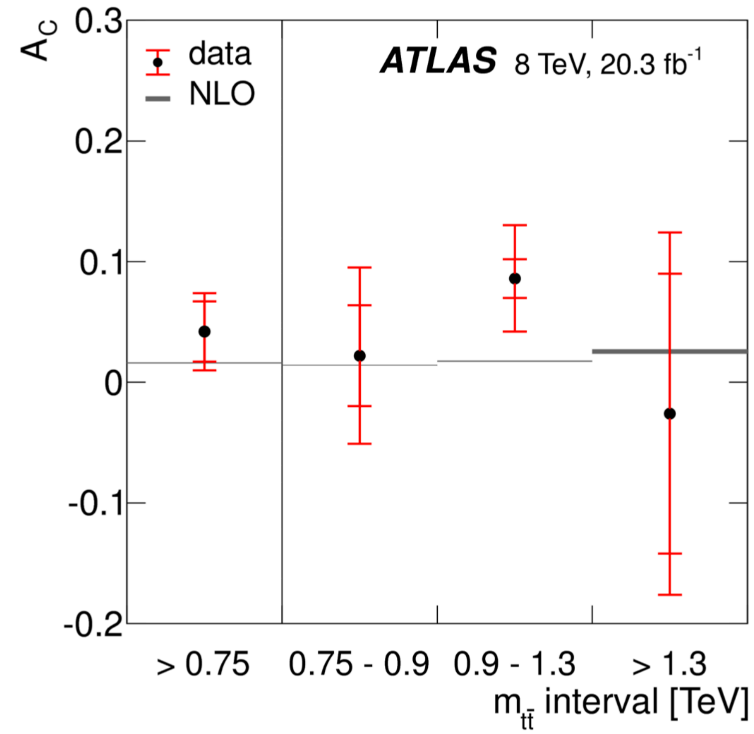
Introduction:

Many interesting ATLAS measurements using jet substructure **@ 8 TeV** already presented at BOOST (*)!

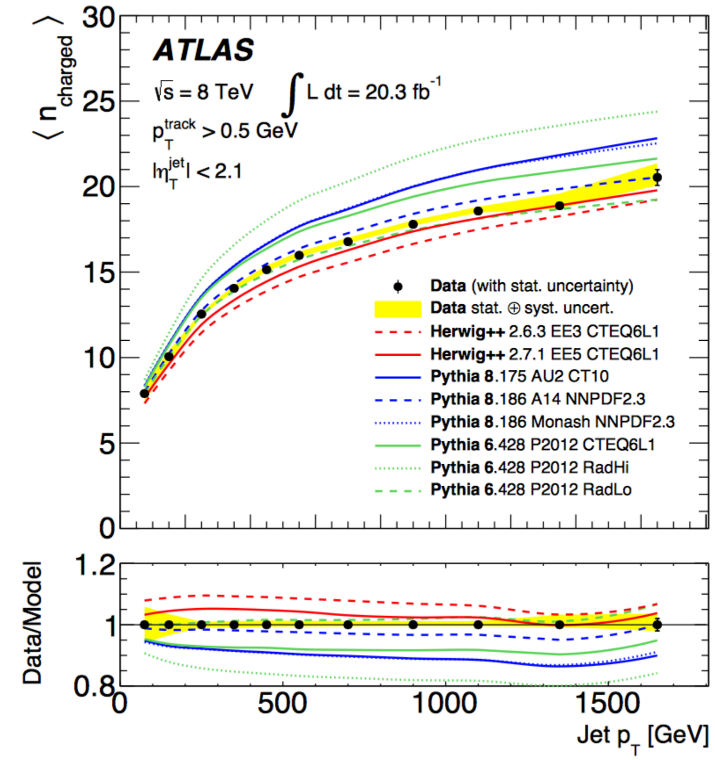
Phys. Rev. D 93, 032009 (2016)



Phys. Lett. B 756 (2016) 52-71



Eur. Phys. J. C76, 1-23 (2016)



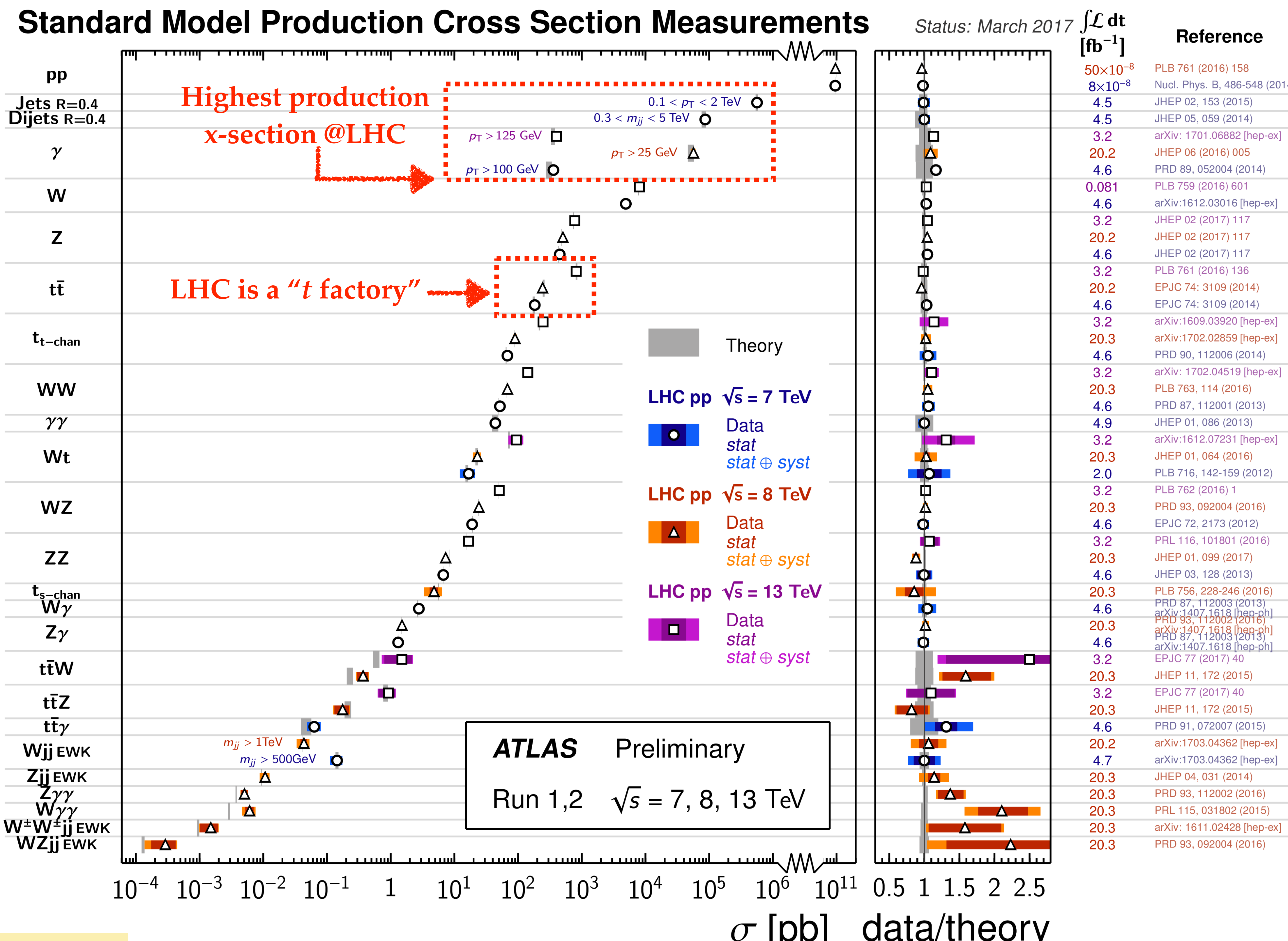
Outline of the talk:

- tt differential σ production → *now at 13 TeV first time at BOOST!*
- W collinear σ production → *first measurement of the weak structure!*
- To be published soon/ongoing efforts:
  - ttH → *first boosted channel in ttH*
  - trigger → *new triggers using jet substructure*

(*) some of them here:
https://indico.cern.ch/event/439039/contributions/2223299/attachments/1311086/1961864/boost16_negrini.pdf
https://indico.cern.ch/event/439039/contributions/2223300/attachments/1310580/1962242/BOOST16.pdf

ATLAS Standard Model:

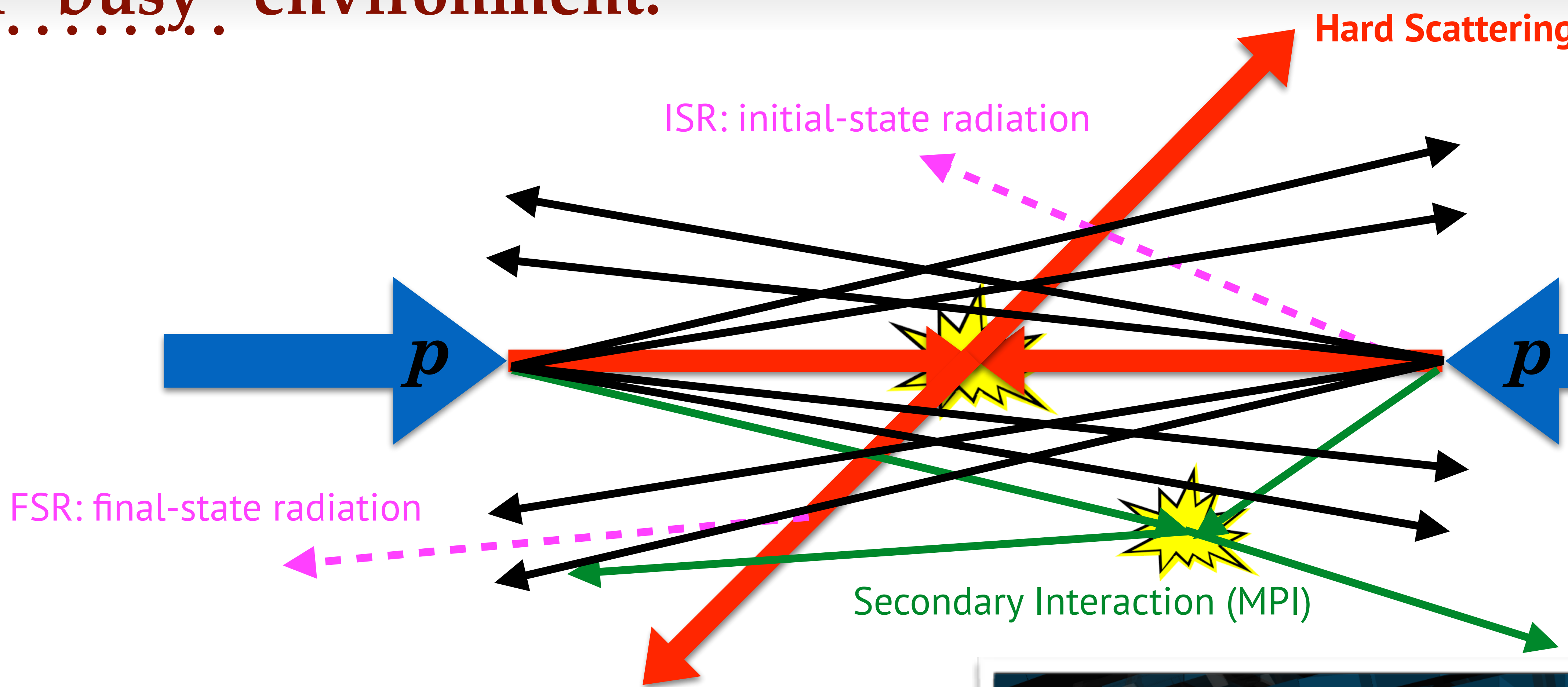
atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SM/



Jets importance:

- ✤ **SM:** test of QCD and QED theoretical calculations
- ✤ **Beyond SM:** many topologies involving hadronic final states (as seen in Jonathan's & Junpei's talks).

A “busy” environment:



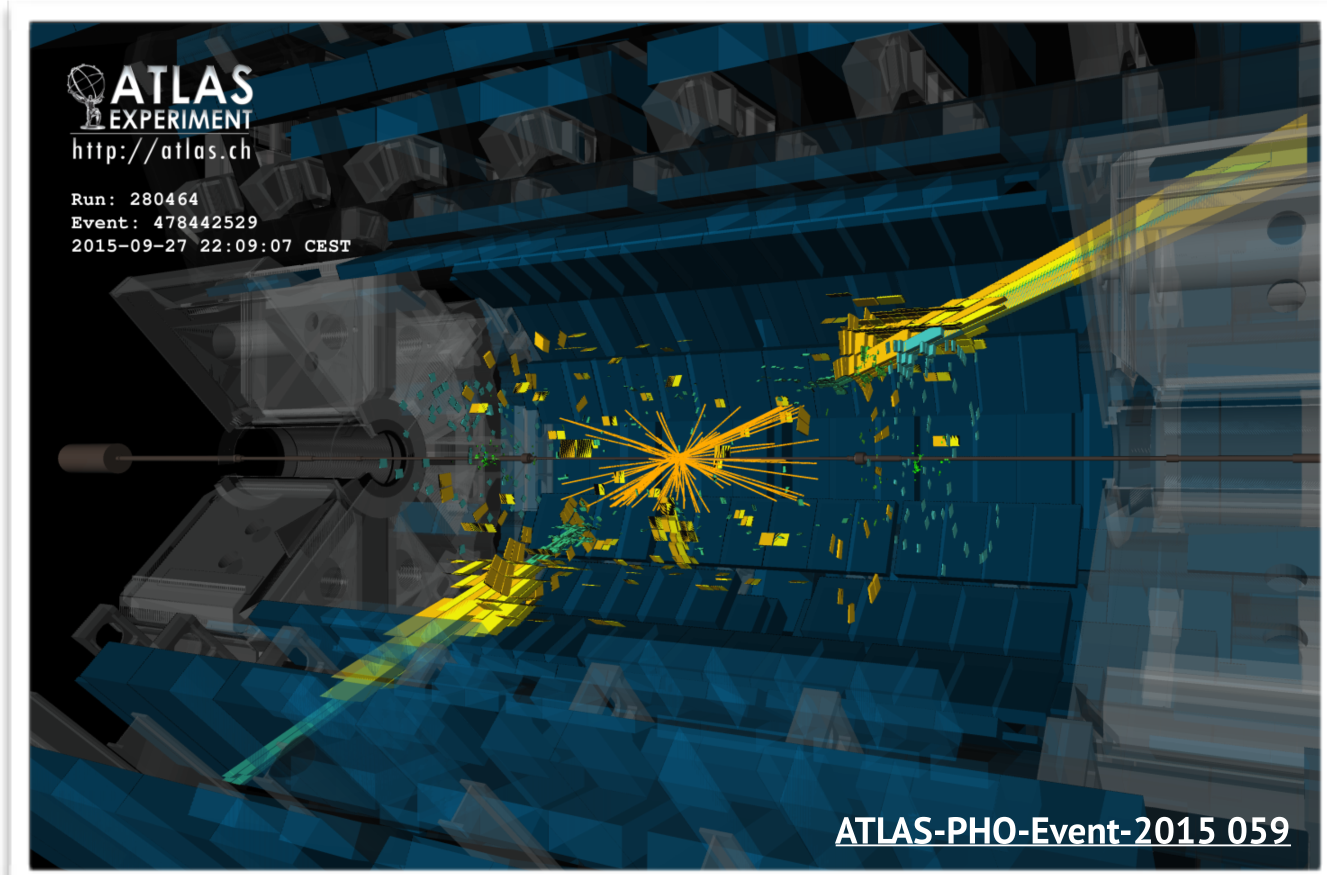
LHC @ √s=13 TeV:

- ✔ *Hard Scattering Event*
- ✖ Underlying Event (UE):
  - ✖ ISR/FSR
  - ✖ Multiple interactions per bunch crossing
  - ✖ Pile-up (up to ~50 in 2016)

} **Jet “cleaning” or *grooming* =“ *trimming*” ATLAS standard procedure**

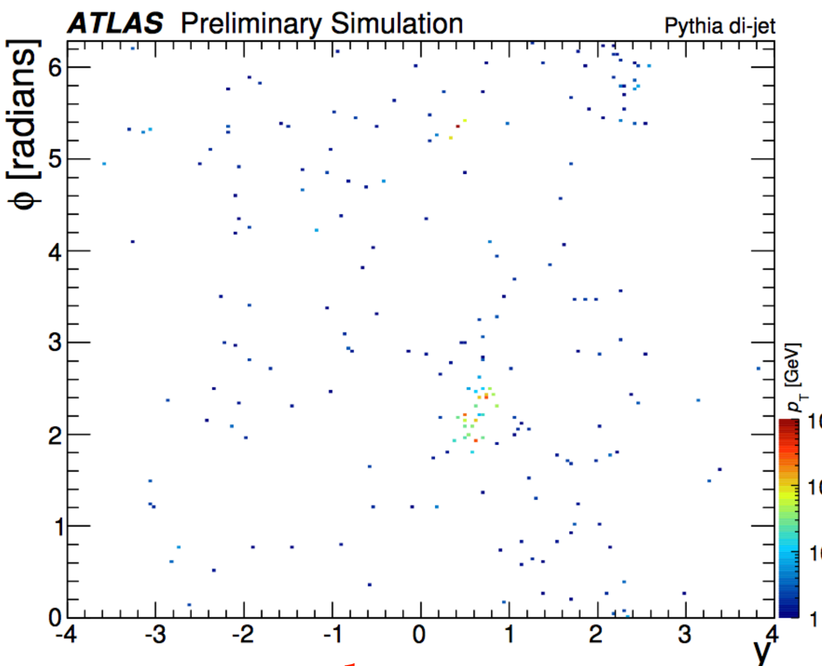
✔ *Boosted objects:*

Exploit jet substructure properties



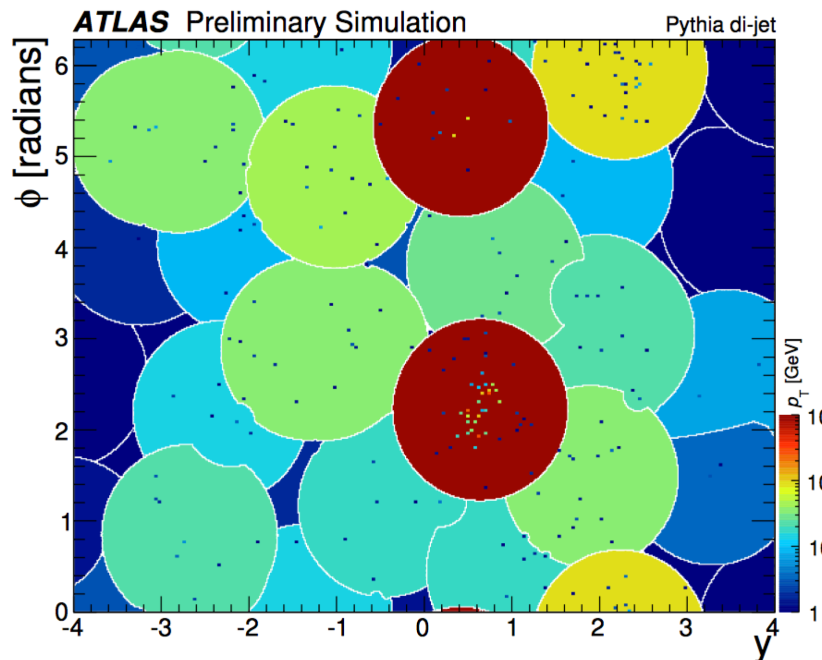
Jet Trimming - ATLAS Large-R jets:

1



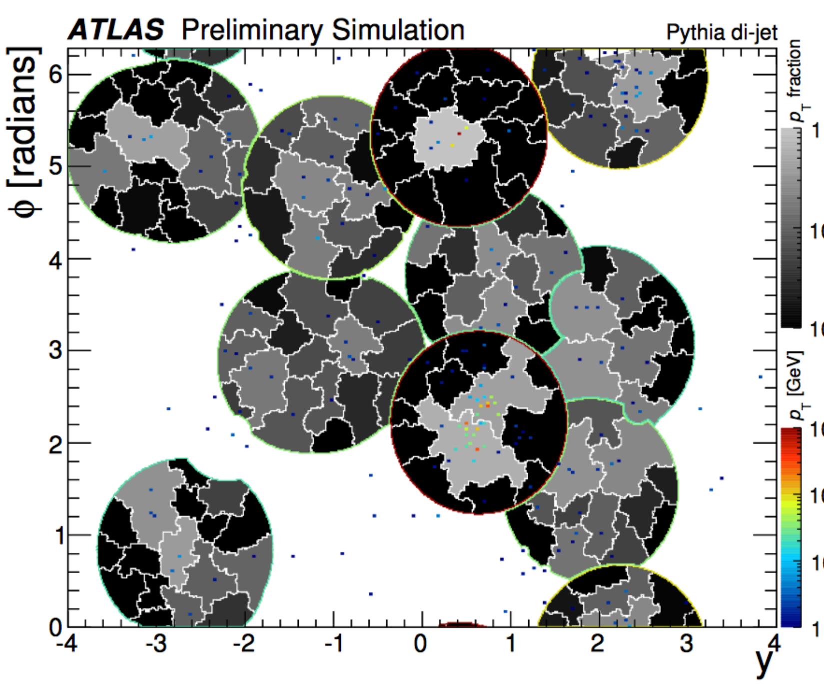
start with all clusters

2



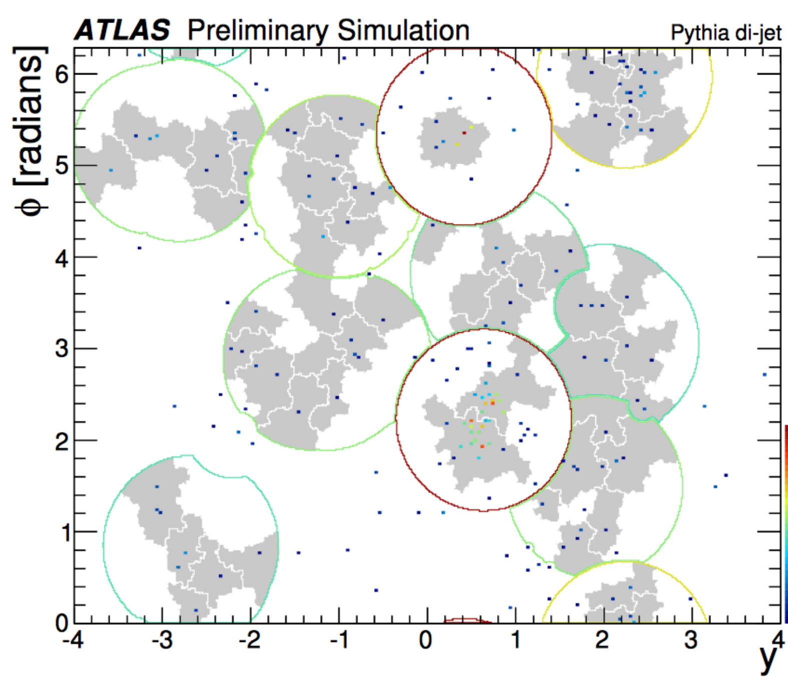
make R=1.0 jets

3



make k_T subjets

4



remove subjets if $p_T^i / p_T^{jet} < f_{cut}$

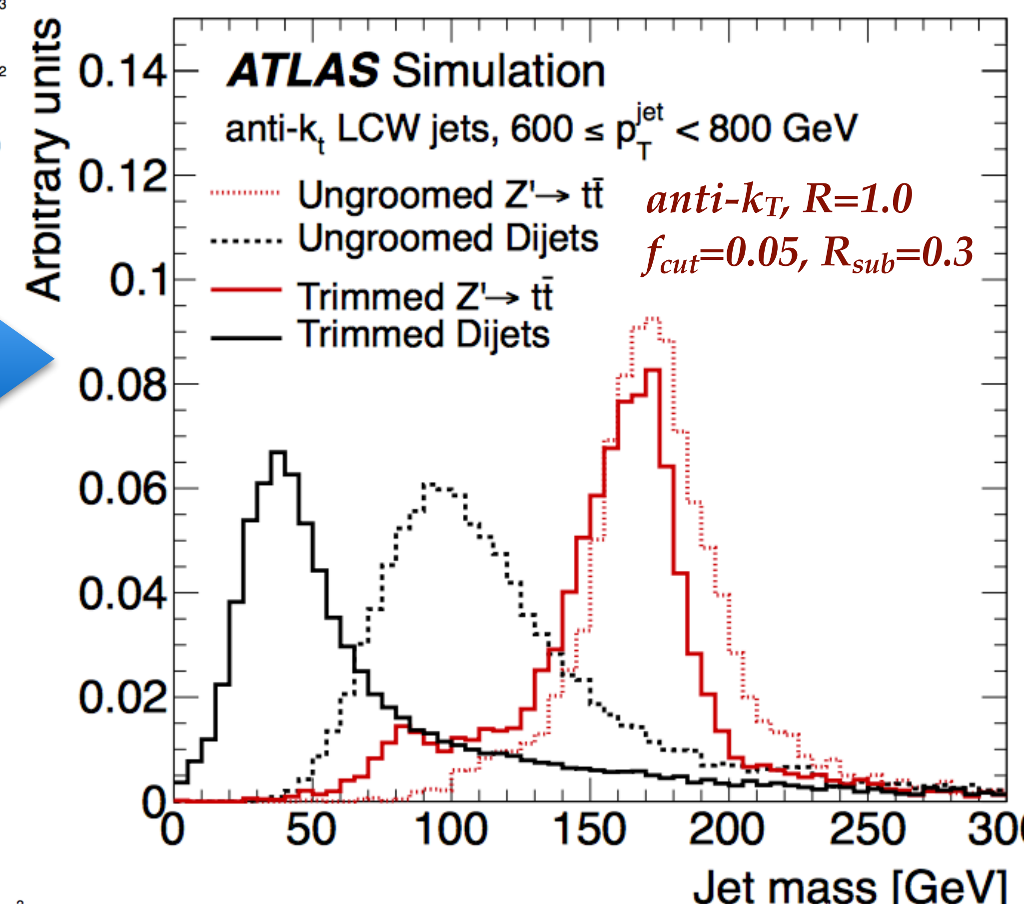
Grooming:

- *Trimming* →ATLAS standard procedure for many SM analysis

- mass-drop filtering
- pruning

} *alternatives*

CERN-PH-EP-2013-069



more in Joe's talk!

Jet re-clustering:

- jet radius might be not optimized for specific physics scenarios
- *in situ* calibration
- enhance the availability of large-R jet configurations
  → intermediate ***scale $r < R$*** input to reconstruct large-R jets
- calibrated small-R jets can make *calibration of re-clustered large-R jets automatic*

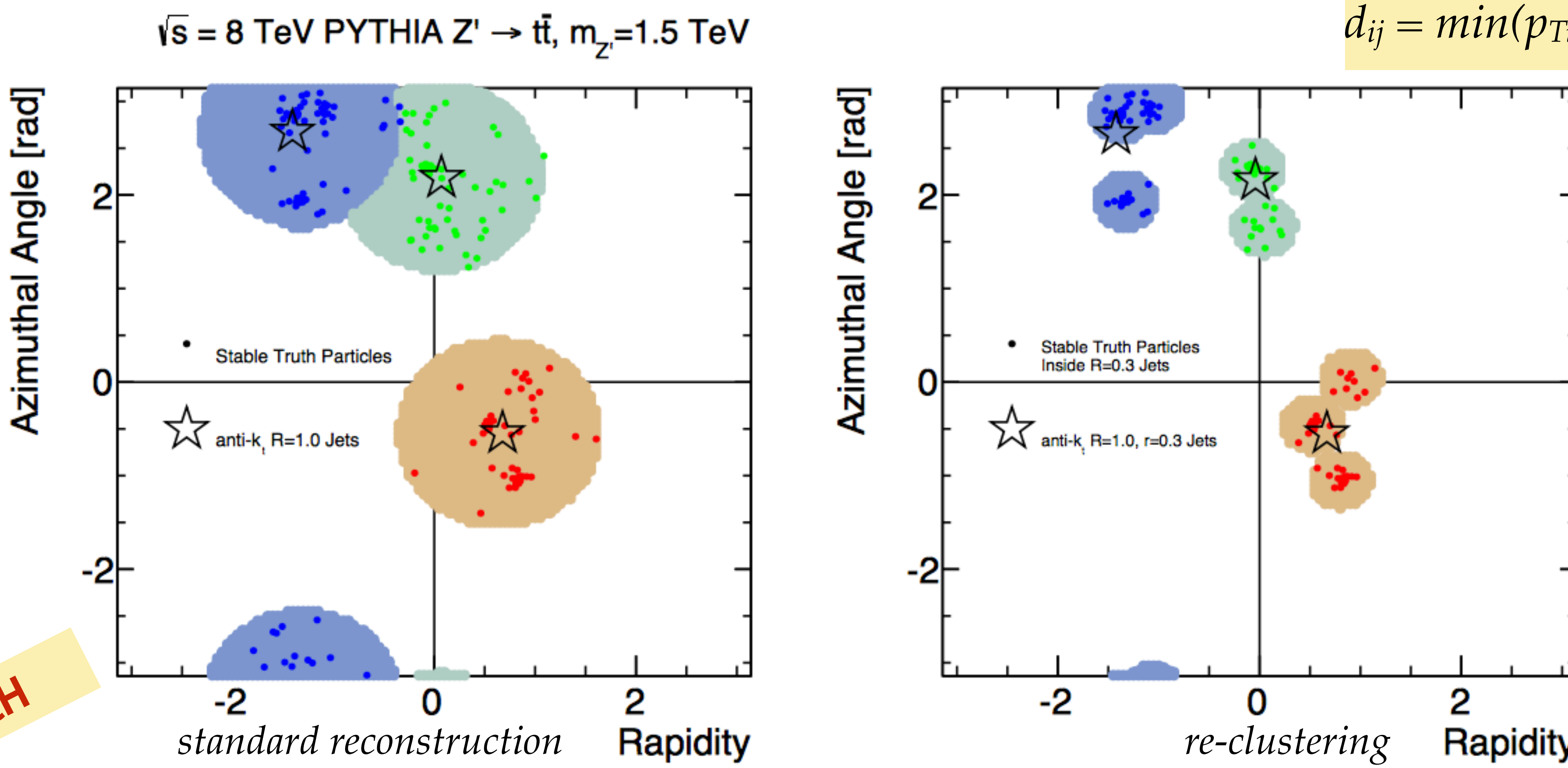
↪ ***any large-R, any clustering algorithm, and many grooming strategies can be used!***

N.B. does not mean uncertainty on re-clustered jet necessarily smaller than correspondent large-R jet.

improving potential discovery for NP

How re-clustering is applied:

$$d_{ij} = min(p_{Ti}^{2n}, p_{Tj}^{2n}) R_{ij}^2 / R^2$$



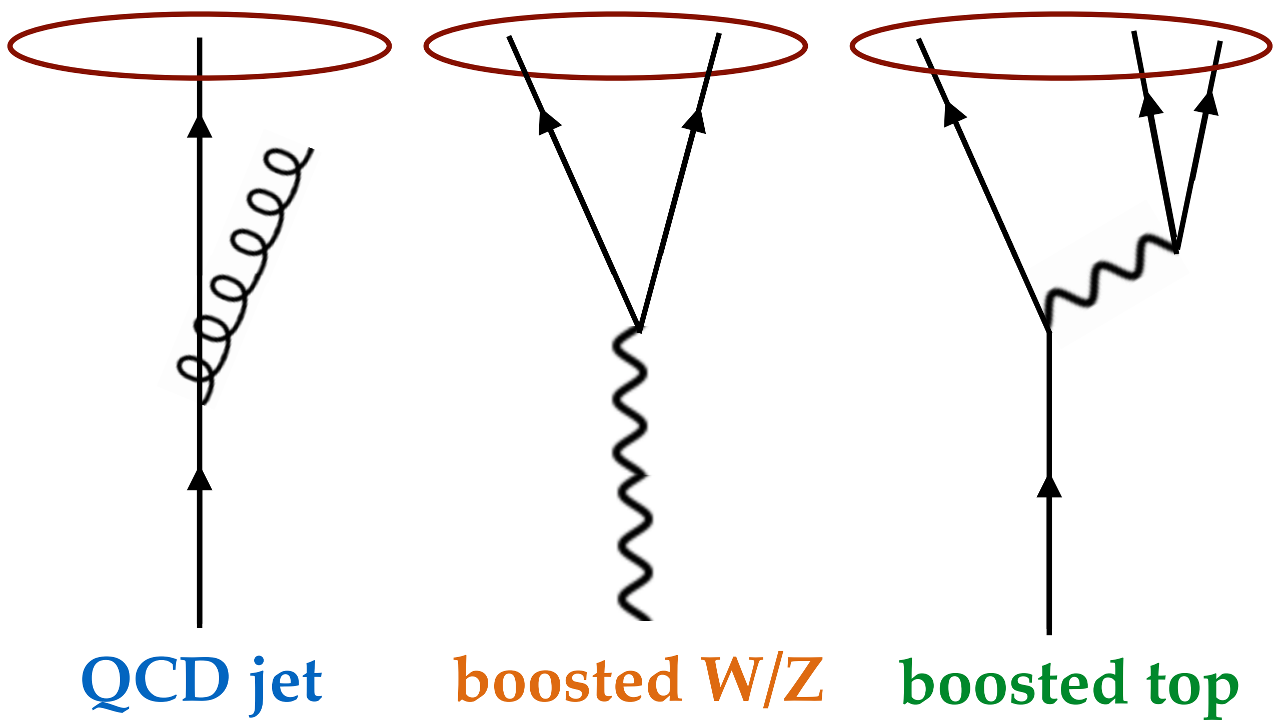
Jets from Jets: re-clustering

used in ttH

Jet tagging using substructure variables:

more in Nurfikri's talk!

In the decays of massive resonances, boosted prongs can be collimated into a single jet:



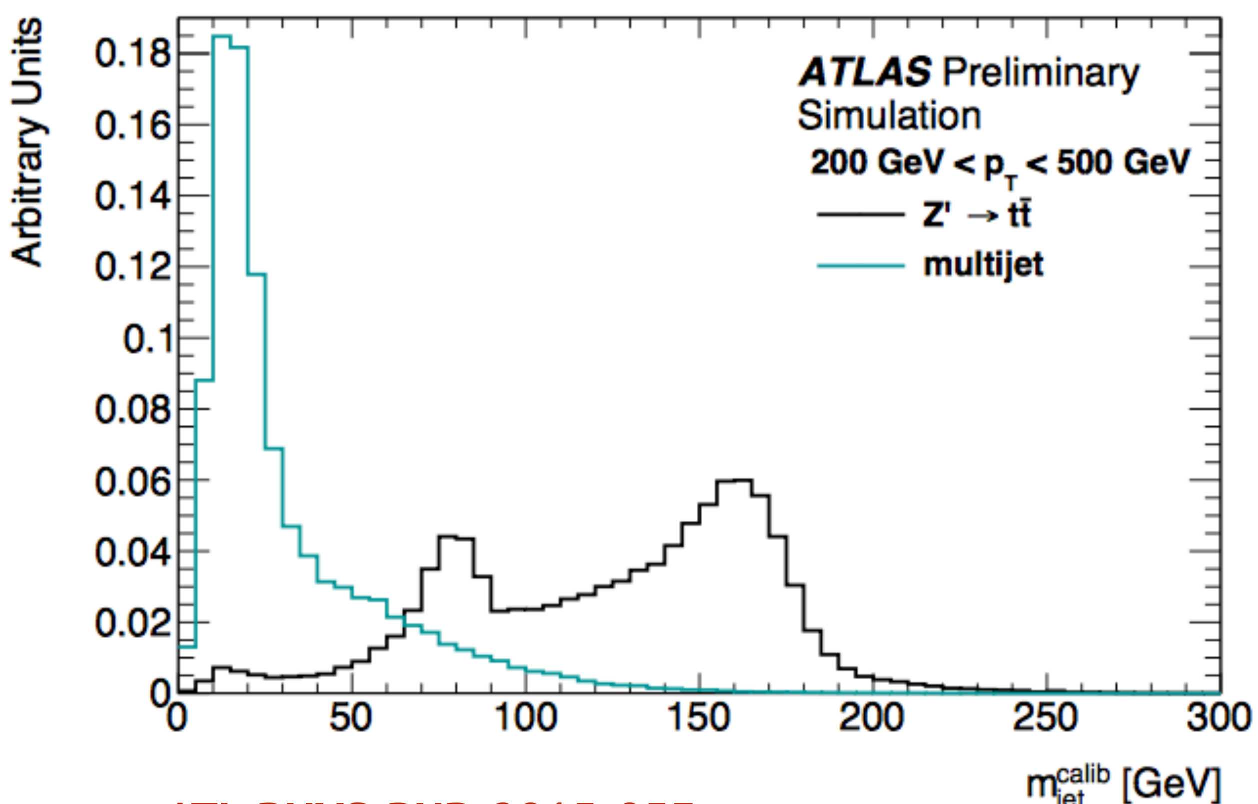
Tagging on:

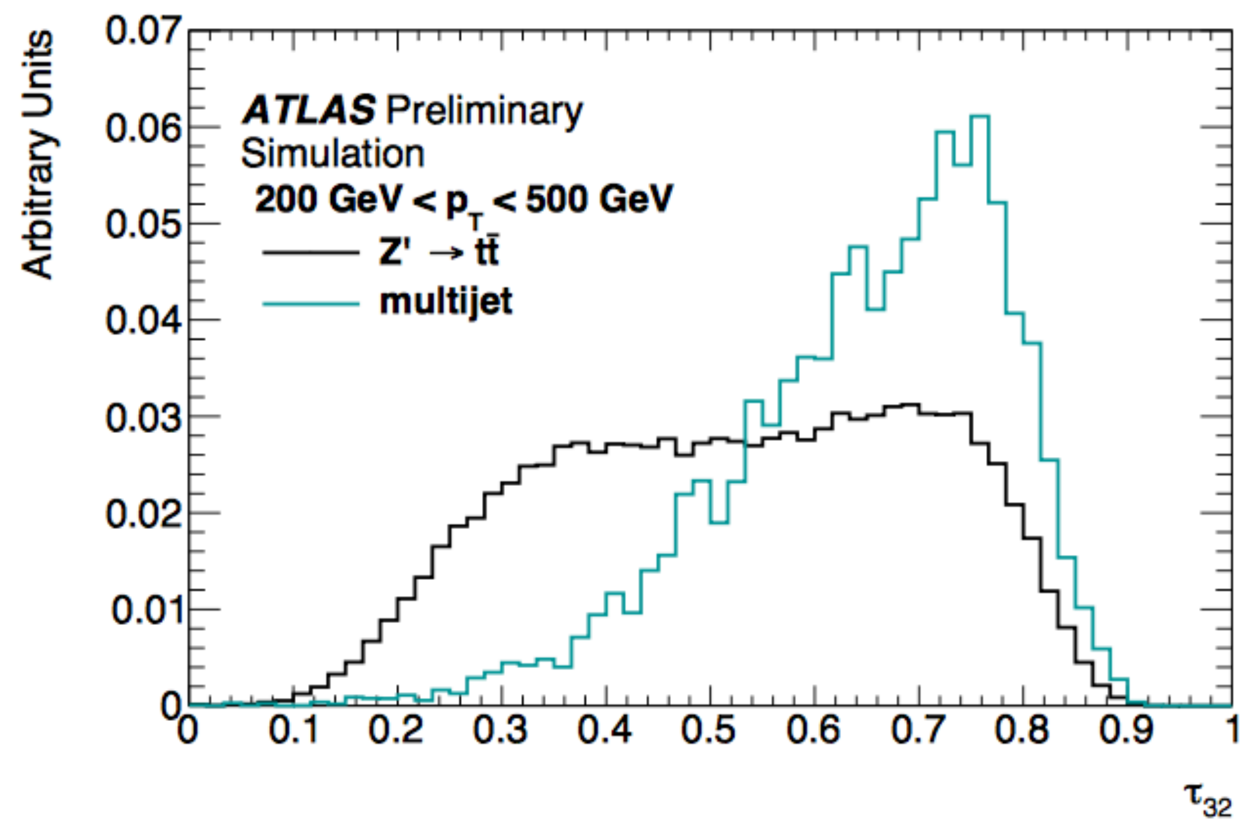
- Jet mass: calibrated mass (*)
- N-subjettiness ratio (**):

$$\tau_N = \frac{1}{d_0} \sum_k p_{\mathrm{T}k} \times \min(\delta R_{1k}, \delta R_{2k}, \ldots, \delta R_{Nk})$$

- τ_2: 2 prong decay
- τ_3: 3 prong decay
- splitting scale
- minimum dijet mass from three subjects

→ $\tau_{32} = \tau_3/\tau_2$





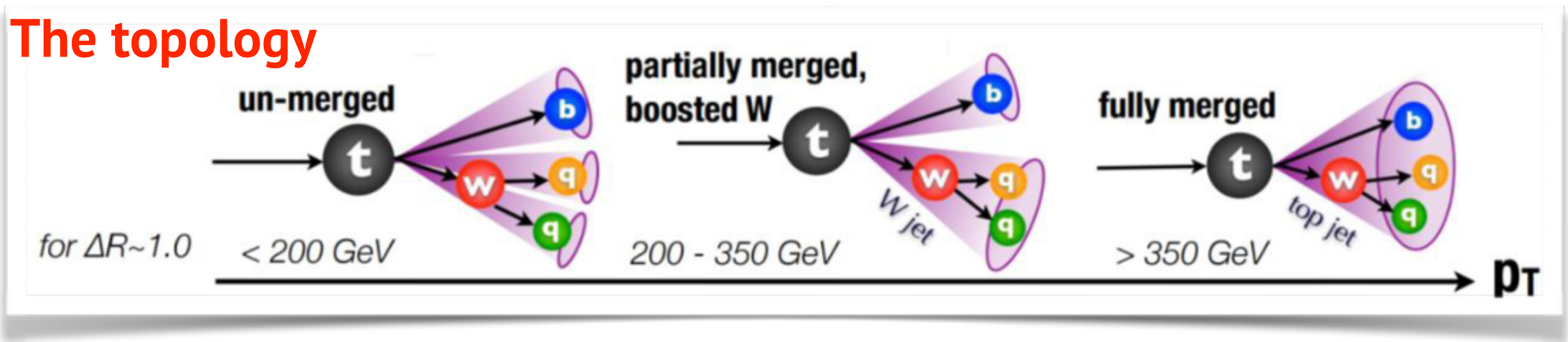
ATL-PHYS-PUB-2015-053

e.g. top-tagging (*), (**)

Measurements of tt differential cross-section:

The importance of tt:

- **high production cross section**: $\sigma(13\ TeV) = 3.3 \times \sigma(8\ TeV) \rightarrow \sim$ *800 pb*,
- test of SM up to the ***TeV scale***,
- differential measurements **sensitive to BSM** scenarios, not detectable in inclusive ones.



Final states:

- *l+jets:* resolved and boosted **(LJ)**
- *all-hadronic:* boosted **(AH)**

- High Lorentz boosted top-quark (p_T>300 GeV) decay products increasingly difficult to resolve → merged into a *large-R jet*
- Both analysis define two types of anti-k_T jets:
  - ***Small-R jets:*** R=0.4, $|\eta| < 2.5$, p_T> 25 GeV (**LJ** and **AH**)
  - ***Large-R jets:*** R=1.0, $|\eta| < 2.0$, p_T> 300 GeV (**LJ** and **AH**)

Large-R jet reconstruction:

- **Trimming: R_{sub}= 0.2 , f_{cut} = 0.05** (**LJ** and **AH**)
- Trimmed jet mass corrected to particle top jet using MC

Top-tagging based on substructure variables:

- *Large-R jet mass*
- *N-subjettiness*

} chosen because of low correlation, strong performance and robustness across pT range

Analysis Flow:

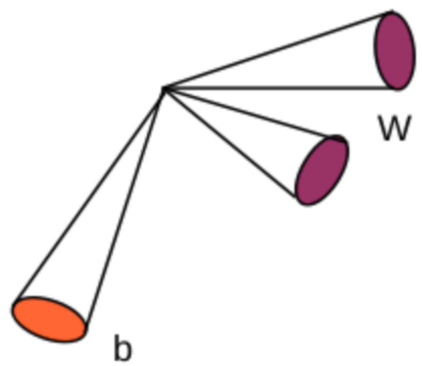
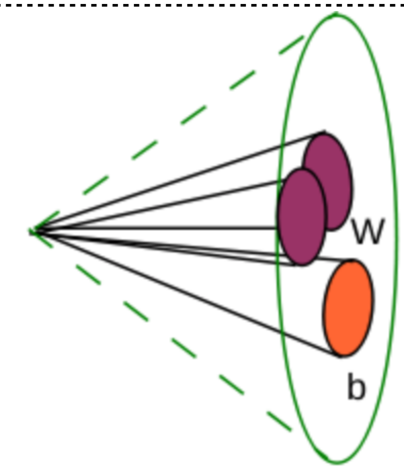
1. *Grooming + substructure:* Analysis jets definition
2. *Backgrounds estimation:* multijet, single top, W+jets to be subtracted from data
3. *Unfolding:* differential σ measurement

common to LJ and AH analysis

$t\bar{t}$ differential cross-section - LJ - Event Selection :

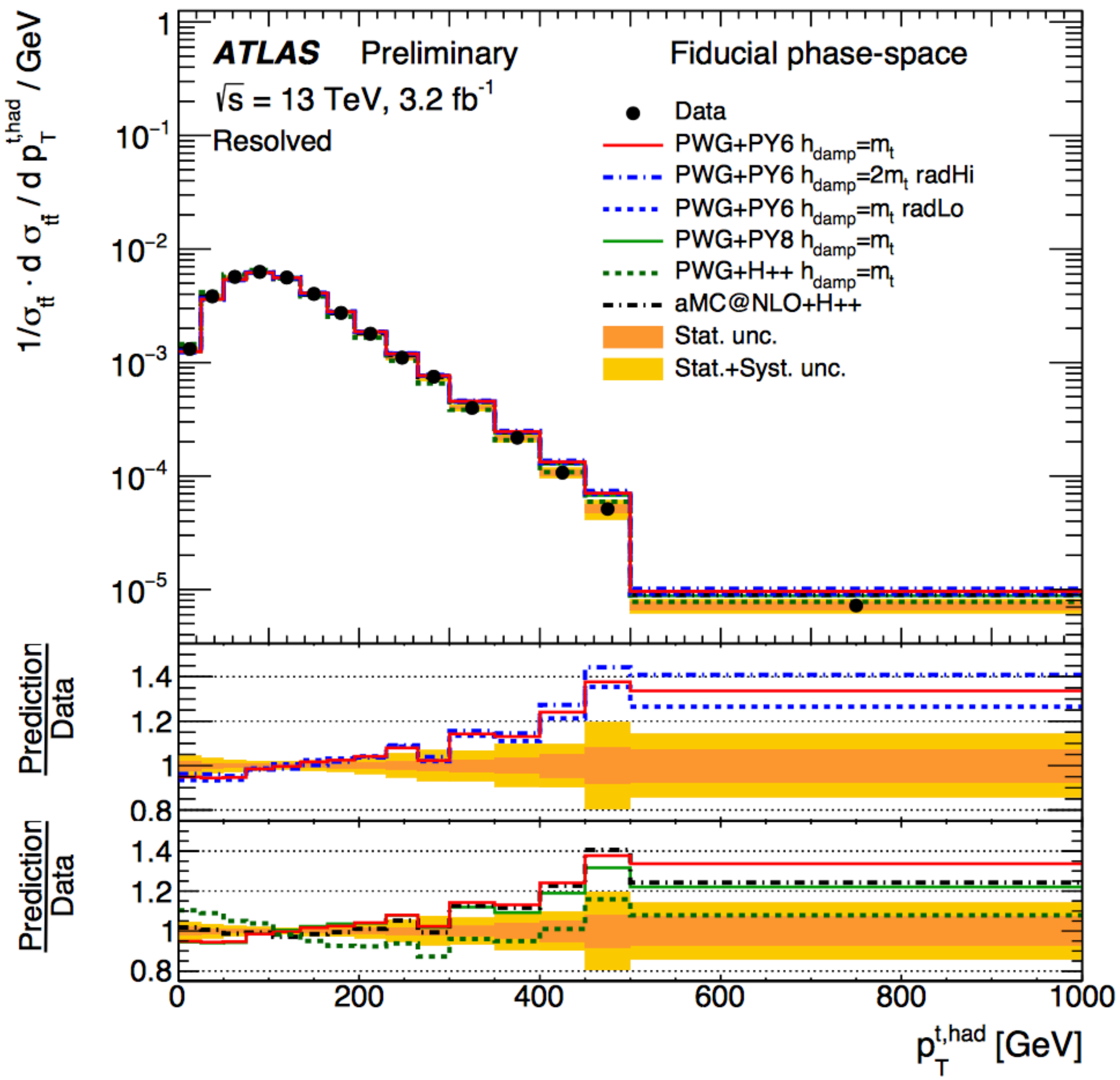
ATLAS-CONF-2016-040

Event Selection: $\sqrt{s}$=13 TeV, $\mathcal{L}$=3.2 fb^{-1}

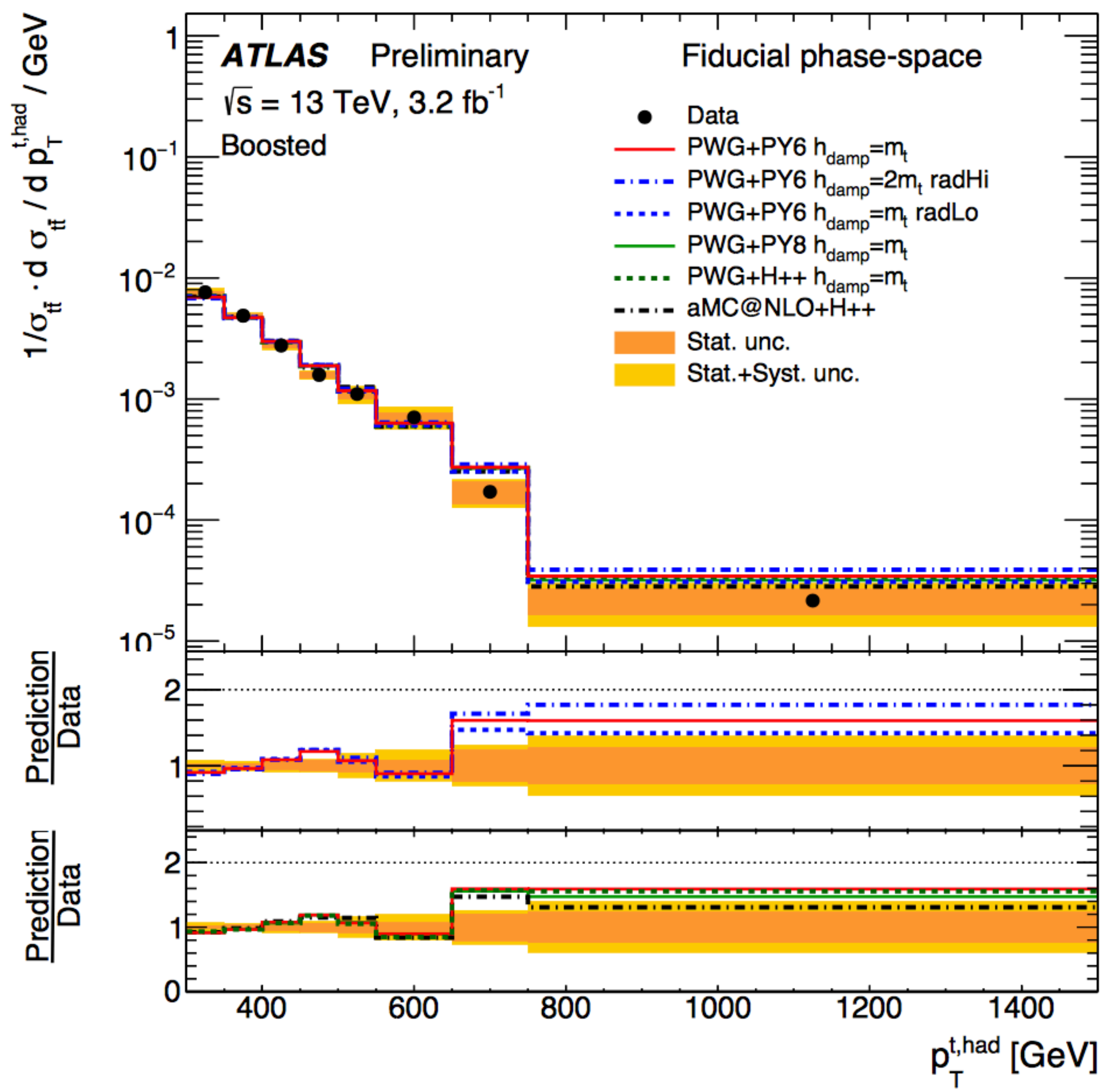
Resolved	**Boosted**
	
≥ 4 small-R jets (≥ 2 b-tagged)	*≥ 1 small-R jet & ≥ 1 large-R jets* *(at least one small-R b-tagged)* *MET > 20 GeV, MET + m_T^W > 60 GeV*
✤ **Leptonic t:** ✤imposes W-mass constraint to solve for ν $\lvert p_Z \rvert$ ✤pairs W and b-jet closest in ΔR to lepton ✤ **Hadronic t:** ✤pairs non b-tagged jets closest to m_W with remaining second hardest b-tagged jet	✤ **Leptonic t:** ✤ at least one small-R jet with ΔR(l, small-R jet) < 2.0 ✤ **Hadronic t:** ✤ top-tagged large-R jet (m>100 GeV, τ_{32}>0.75).
Variables: $p_T^{t,had}$, $\lvert y^{t,had} \rvert$, p_T^{tt}, m^{tt}, $\lvert y^{tt} \rvert$	*Variables:* $p_T^{t,had}$, $\lvert y^{t,had} \rvert$

tt differential cross-section - LJ:

Resolved



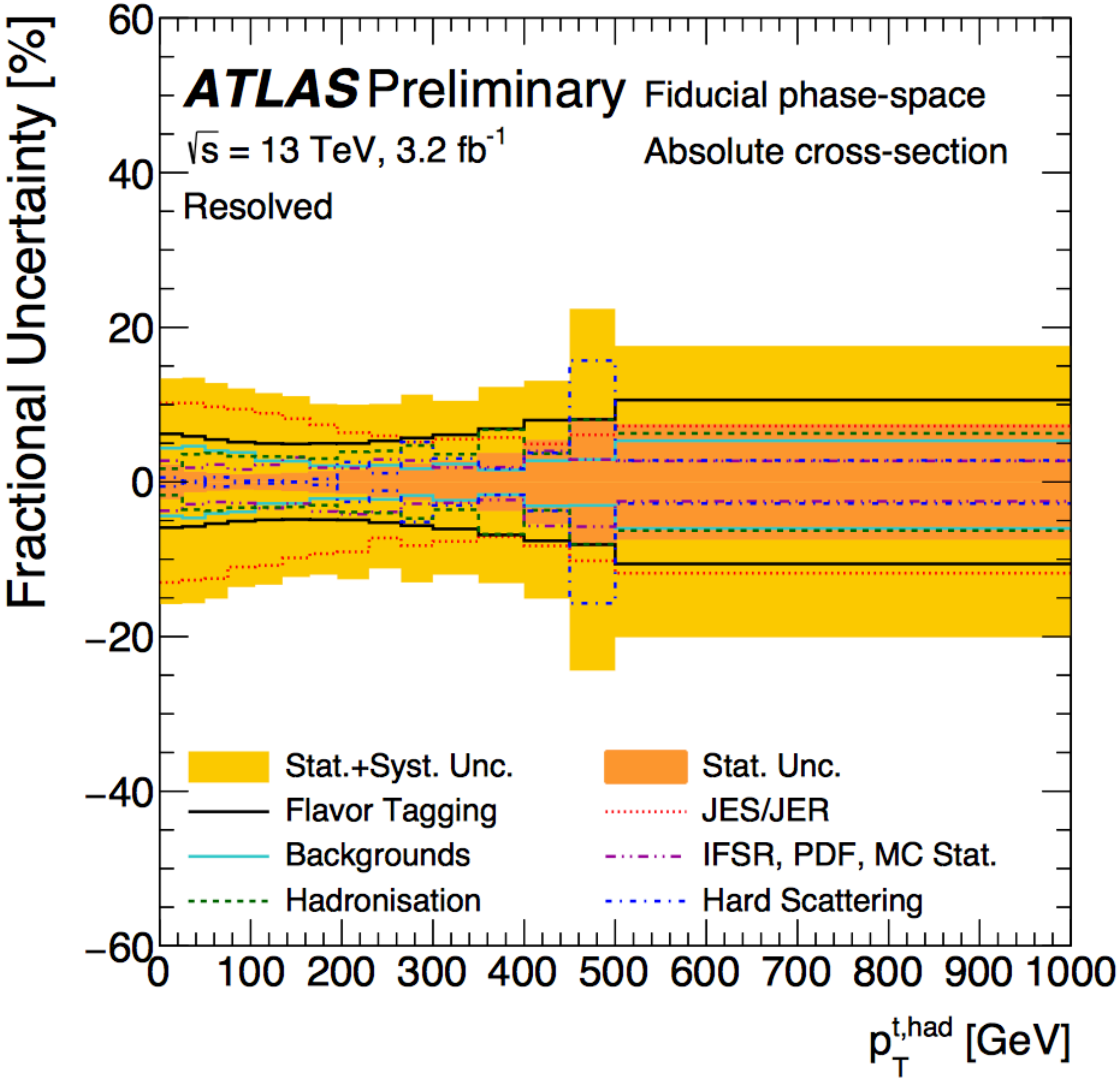
Boosted



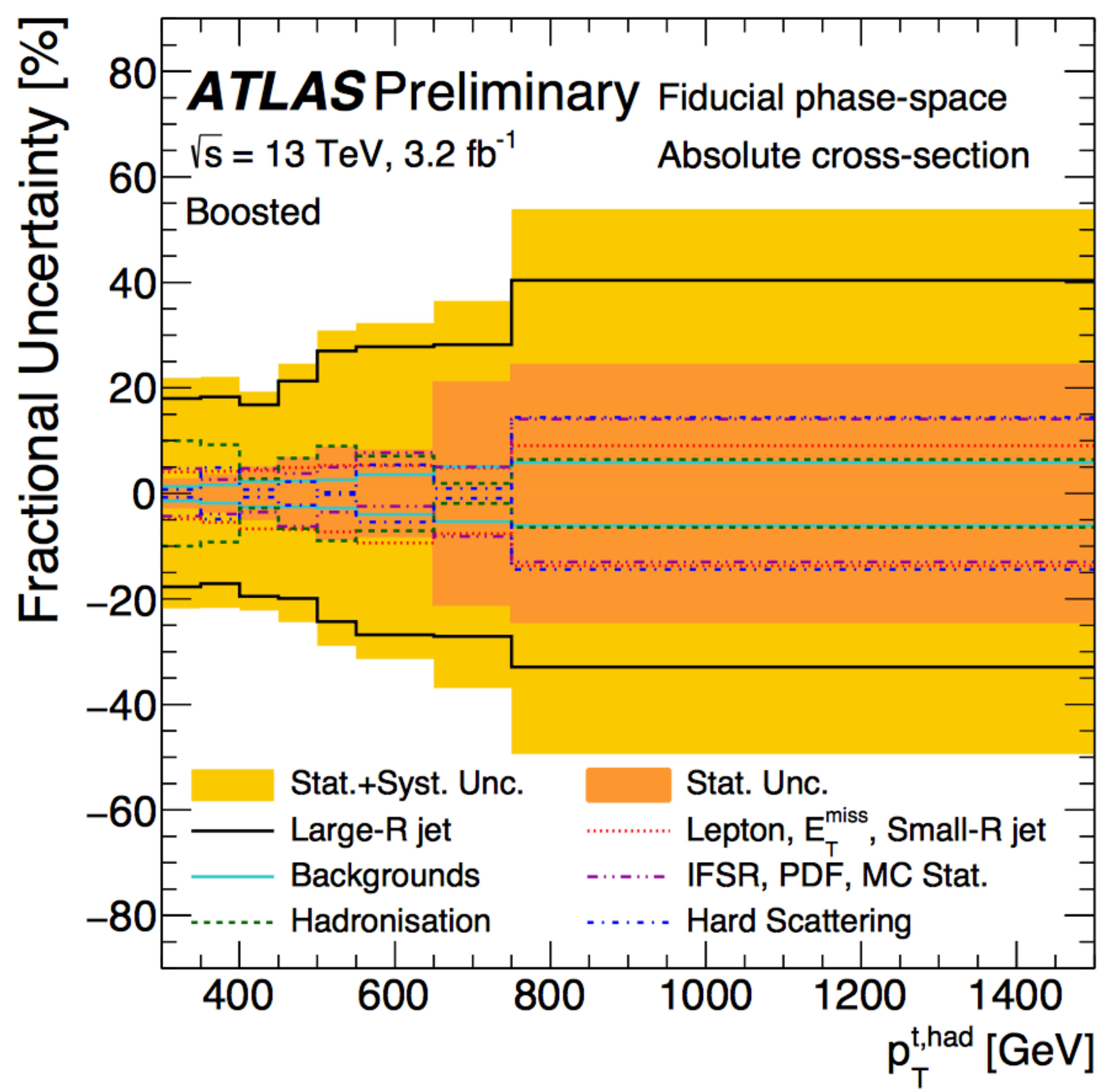
- Data seems softer at high p_T in both resolved and boosted channels
- $p_T^{t,had}$: trends of NLO MC generators similar among generators

tt differential cross-section - LJ - Systematics:

Resolved



Boosted



Dominant uncertainties:

- **Resolved:** Jet Energy Scale (JES) and flavour tagging
- **Boosted:** Large R-jet (→JES dominant)

Measurements of tt differential cross-section - AH : ATLAS-CONF-2016-100

Event Selection: $\sqrt{s}$=13 TeV, $\mathcal{L}$=14.7 fb^{-1}

✤ ≥ 2 large-R jet (top-tagged), $p_T^{lead} > 500$ GeV, $p_T^{sublead} > 350$ GeV, 122.5 GeV $< m_{large\text{-}R} <$ 225.5 GeV

✤ ≥ 2 small-R jet (b-tagged)

✤ ΔR(large-R, small-R) < 1.0

✤ τ_{32} p_T dependent cut (50% efficienciency top-tagging WP)

Variables: p_T^1, p_T^2, $|y^{t1}|$, $|y^{t2}|$, $|y^{tt}|$, m^{tt}, p_T^{tt}, H_T^{tt}, $\Delta\phi^{tt}$, y_B^{tt}, χ^{tt}, $|\cos\theta|^*$, p_{Tout}^{tt}

✤ Challenging QCD background

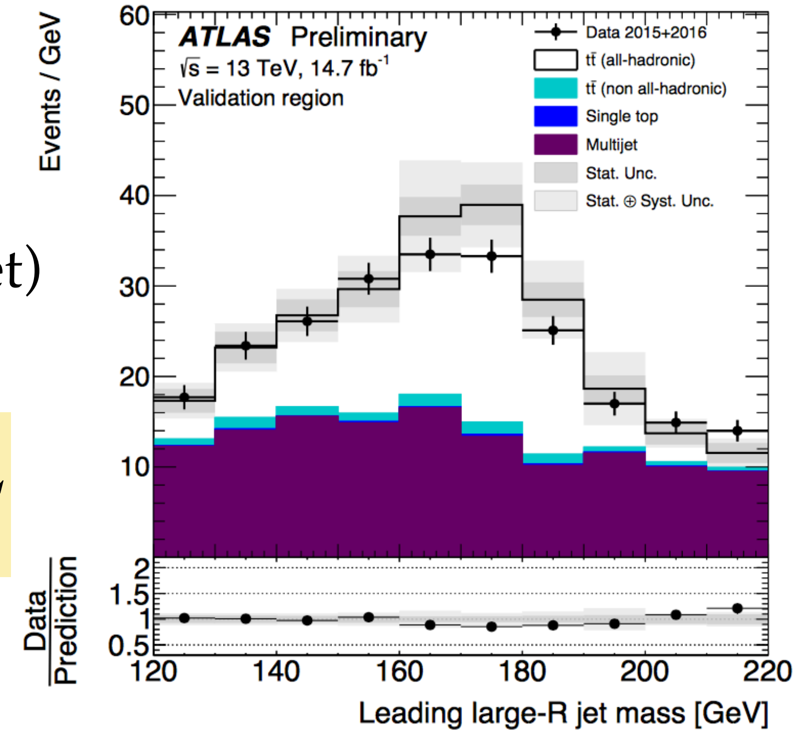
↪ data-driven: ABCD method

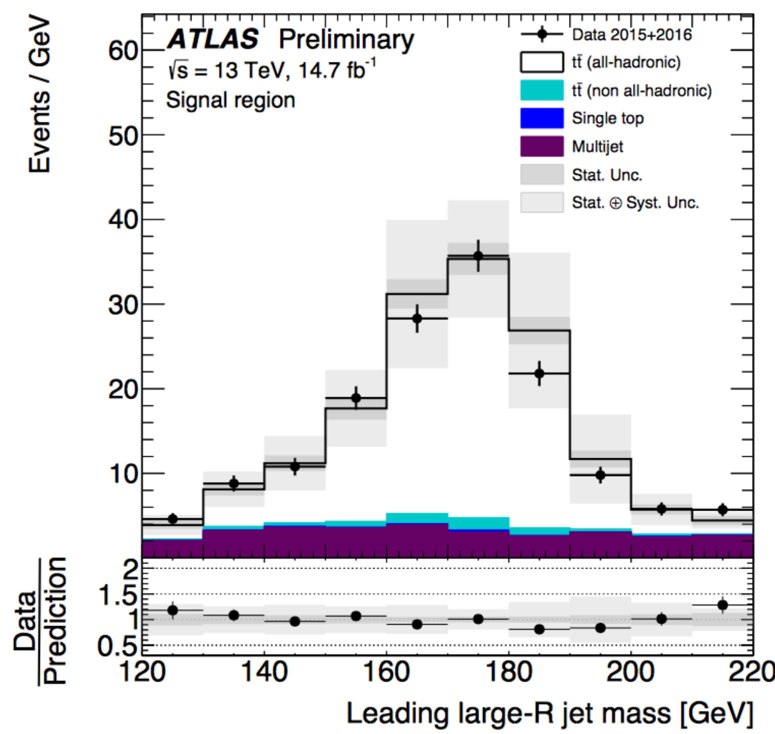
✤A,D,G,B: multijet dominated regions

✤F: validation region (~50% tt/~50% multijet)

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

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





✤ Event yields in signal region:

$t\bar{t}$ (all-hadronic)	1 190	±	240
$t\bar{t}$ (non all-hadronic)	60	±	15
Single top-quark	9	±	5
Multijet events	300	±	20
Prediction	1 570	±	260
Data (14.7 fb^{-1})	1 512		

✤ *tt non-all had* and *single-top* from MC:

✤ including contribution from τ

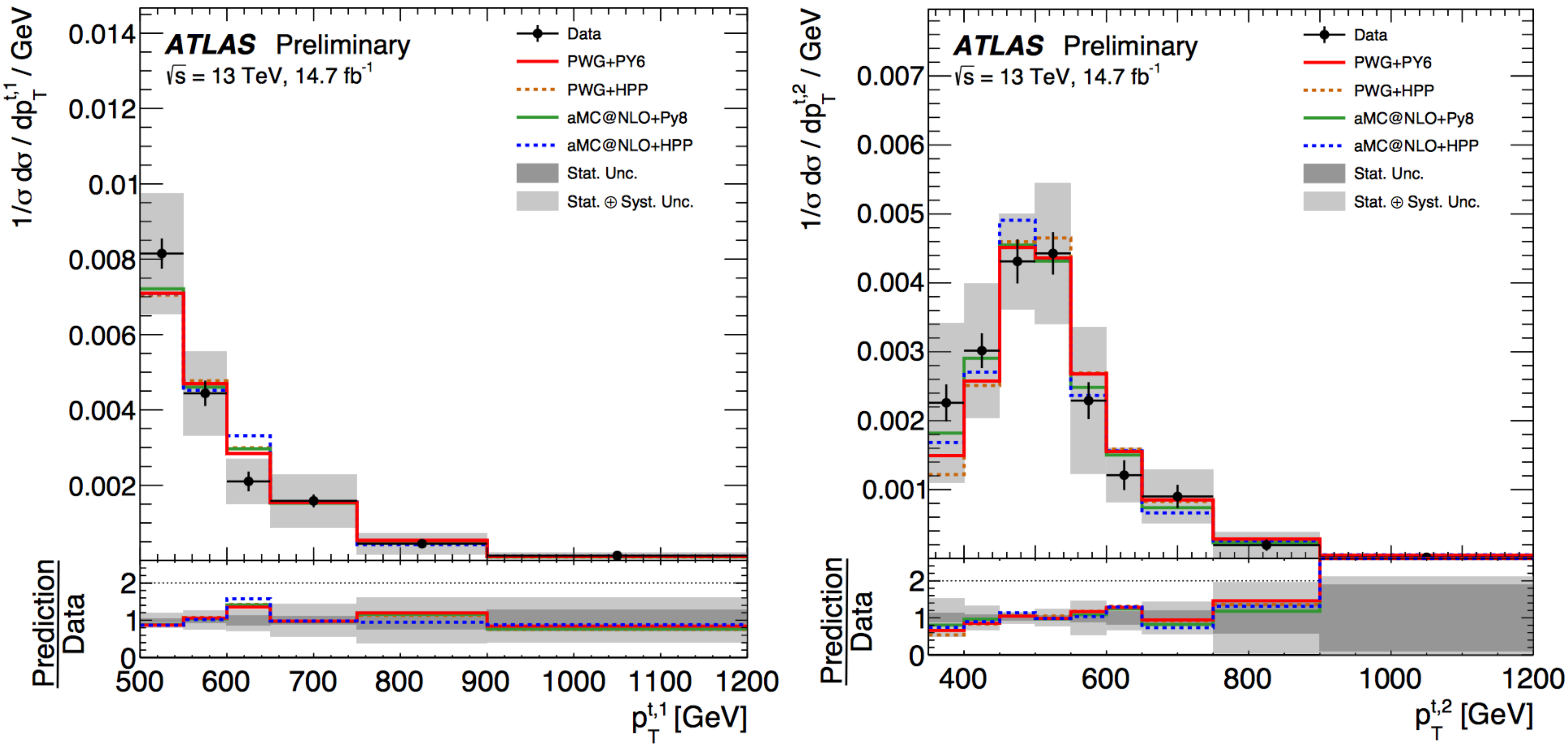
✤ ~4% of total yields in SR

✤ failing top-tagging requirements

tt differential cross-section - AH:

ATLAS-CONF-2016-100

Hadronic top variables:



✤ Good agreement for leading and sub-leading top p_T (sensitive to ~1 TeV)

Dominant uncertainties:

✤ Large-R jet,

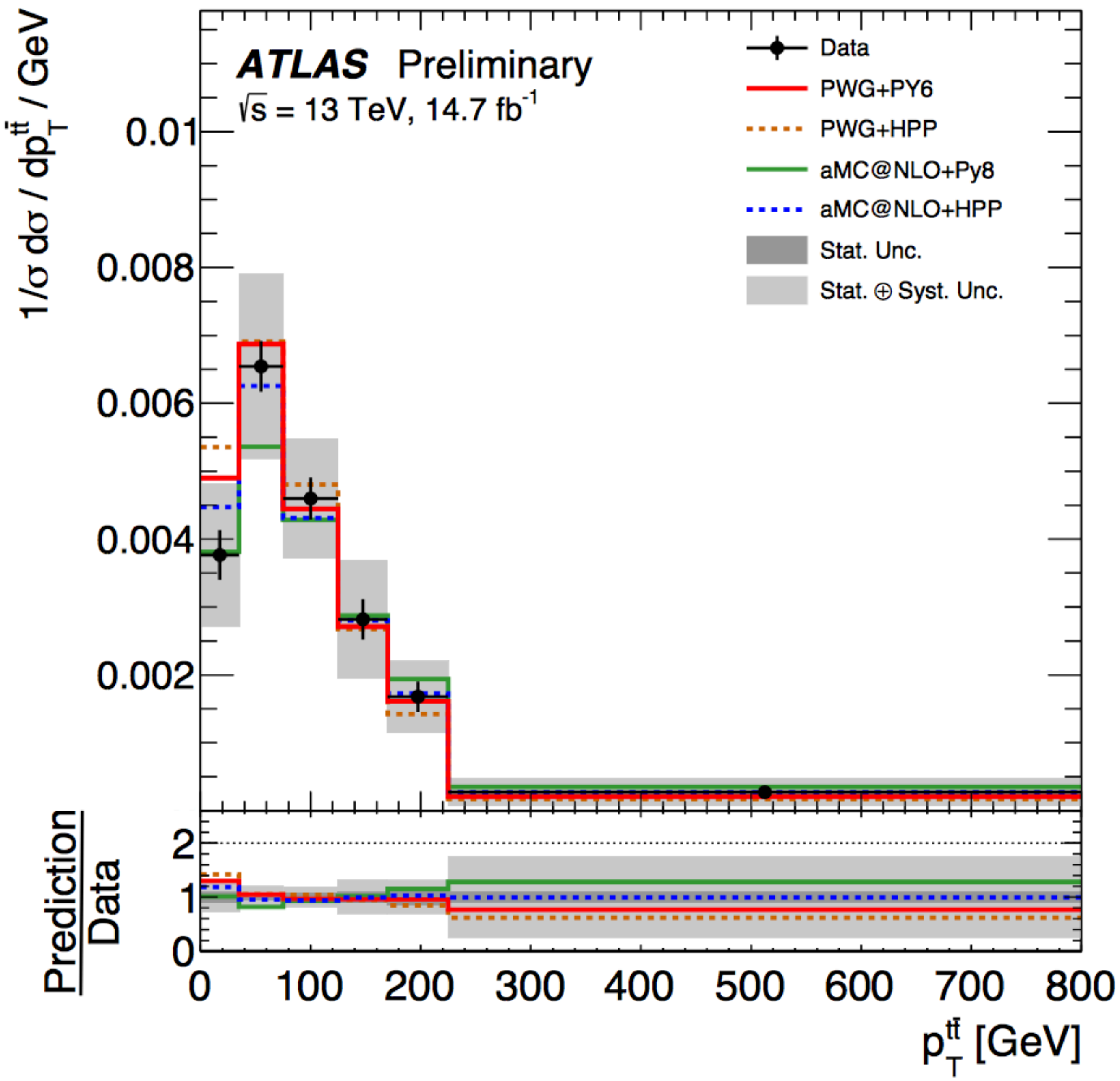
✤ signal modelling,

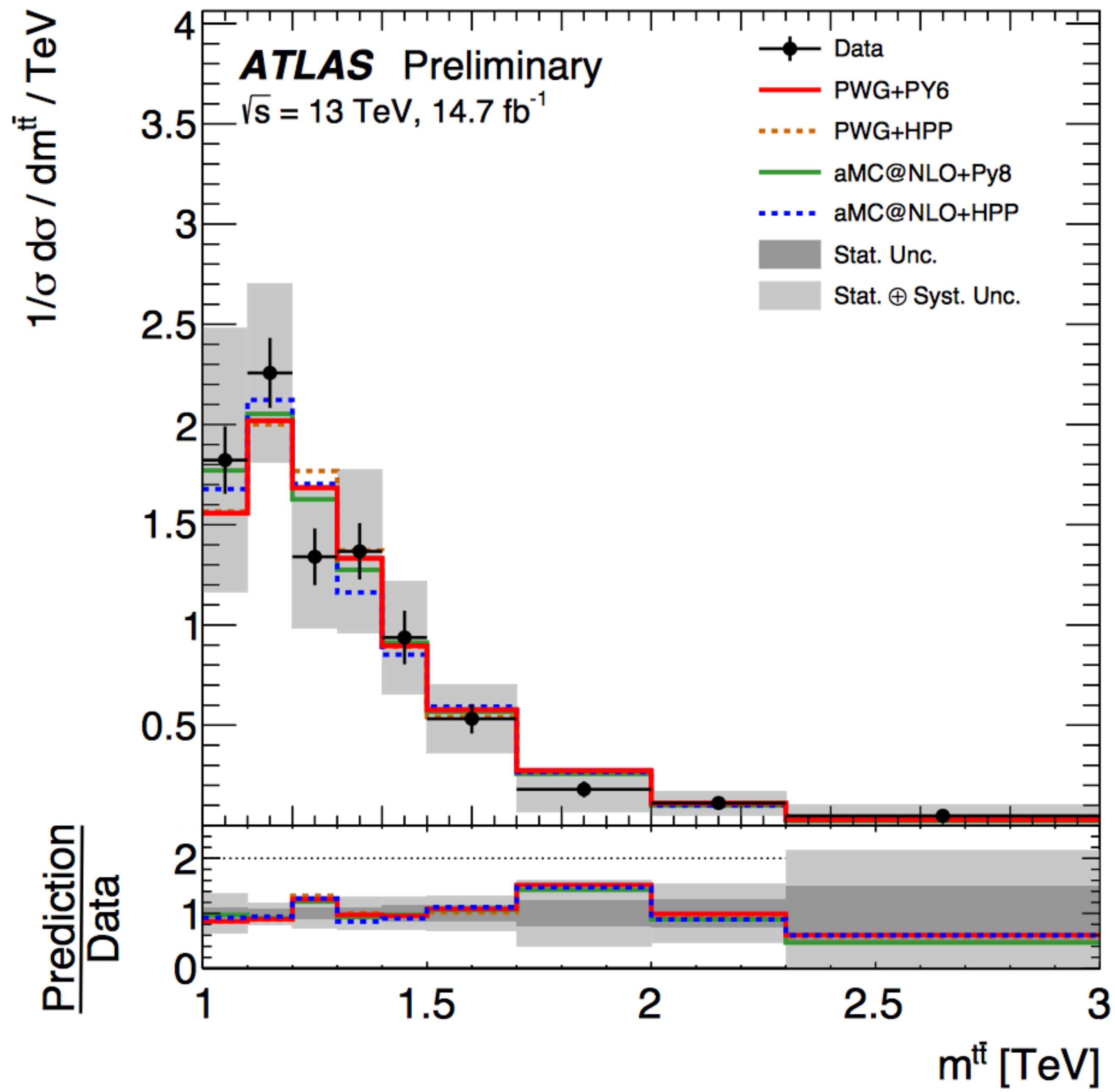
✤ b-tagging

$t\bar{t}$ differential cross-section - AH:

ATLAS-CONF-2016-100

Top system variables:





✤ $t\bar{t}$ system produced with modest p_T slowly falling $m_{t\bar{t}}$ → good agreement with SM

Dominant uncertainties:

✤ Large-R jet,

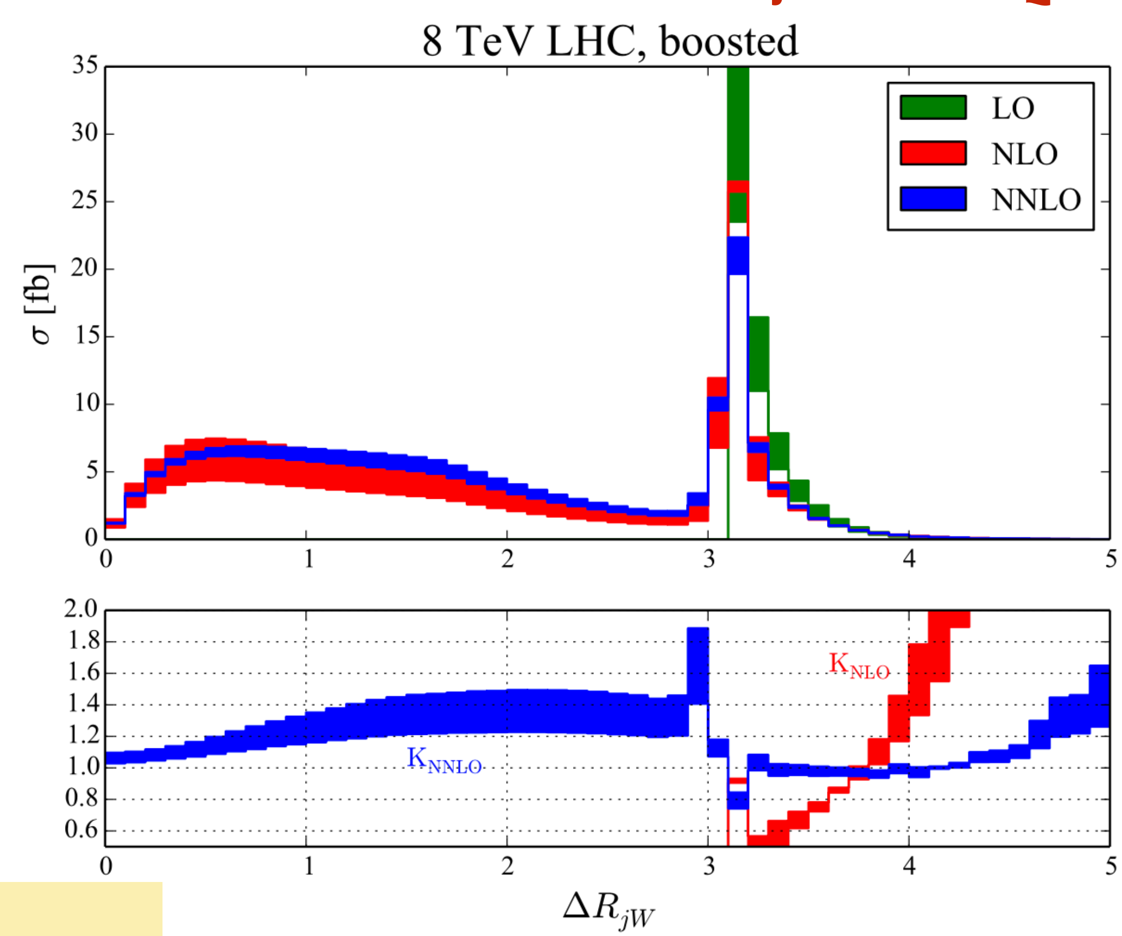
✤ signal modelling,

✤ b-tagging

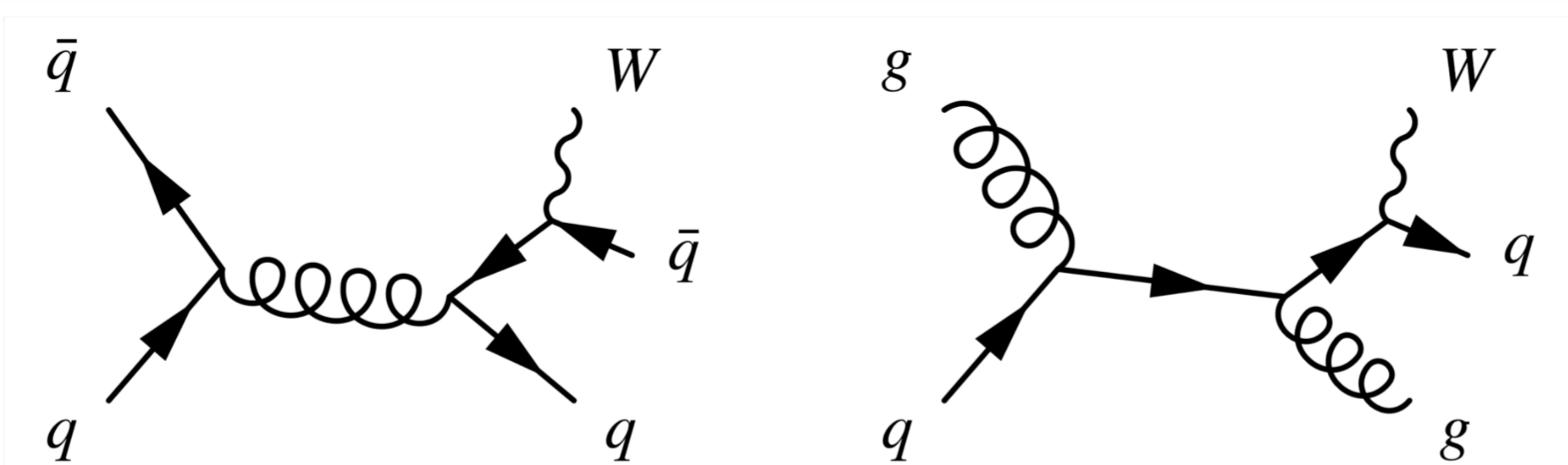
Collinear W @ 8 TeV [see also Junmou's talk!]

W+jet NNLO in QCD

- W+jet: *test perturbative QCD* and *real/virtual EW* emissions
  - **LO** W+1 jet: *back-to-back* production
  - **NLO** W+1 jet: real W boson emission from ISR/FSR $O(\alpha \ln^2 p_{T,j}/m_W)$
    ↪ *collinear enhancement* in angular distance between W and closest jet.
  - tested in regions where cancellation between real/virtual correction incomplete →*small ΔR(W,jet)*
- ***probing a new phase space*** region:
  ↪ *ΔR(l,jet)>0.2:* usually ΔR(l,jet)>0.4
  ↪ *High p_T boosted jet*



- **strong background to WW at very high p_T**
- W/jet collimated → *resemble three prong structure* mimics *t* decay
- important for W + jets measurements at high p_T, vector boson scattering, QCD multijets at high m_{jj}



example of Feynman diagram for W collinear emission

Muon and initial W directions highly correlated
⇒ measure $\sigma_{W(\to\mu\nu)+jets}$ as a function of *ΔR(μ,closest jet)*

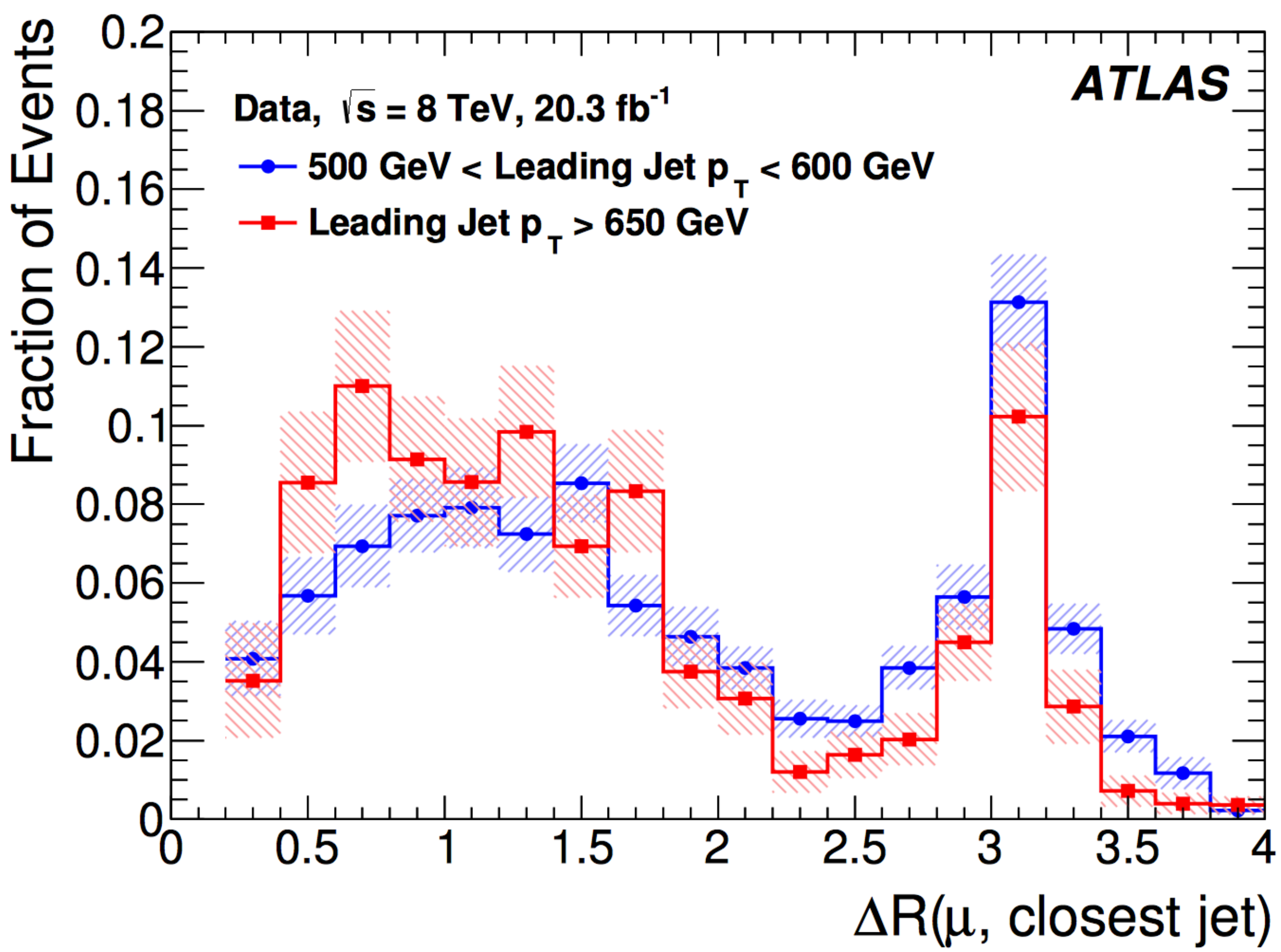
Collinear W @ 8 TeV

CERN-EP-2016-182

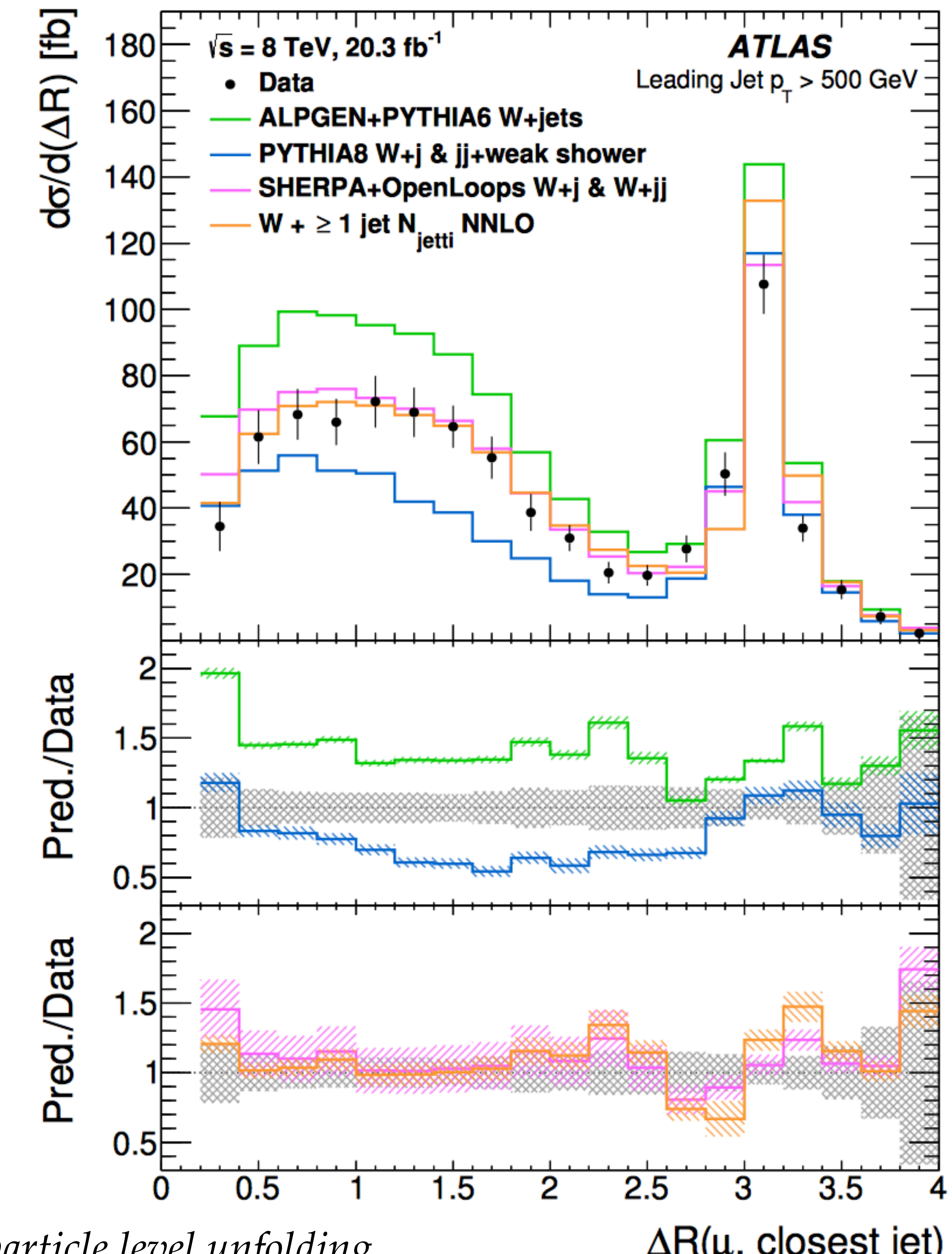
Event Selection: √s=8 TeV, $\mathcal{L}$=20.3 fb^{-1}

- *≥ 1 jet with p_T> 500 GeV and |η|<2.1.*
- *exactly one μ with dressed p_T> 25 GeV, |η|<2.4.*
- *jet with p_T> 100 GeV, |η|<2.1 closest to μ → closest jet*
- *ΔR(closest jet, μ) > 0.2*

- Normalization correction of W+jets, multijet, $t\bar{t}$ and Z+jets in data control regions
- Main systematic: JES and b-tagging
- Fraction of collinear events increases with leading jet p_T and √s



- Alpgen+Pythia6: *multi-leg LO*
- Pythia8: *includes dijet events with weak shower*
- Sherpa+OpenLoops: *NLO QCD + EW corrections*
- Njetti NNLO: *calculation up to $O(\alpha_s^3)$*



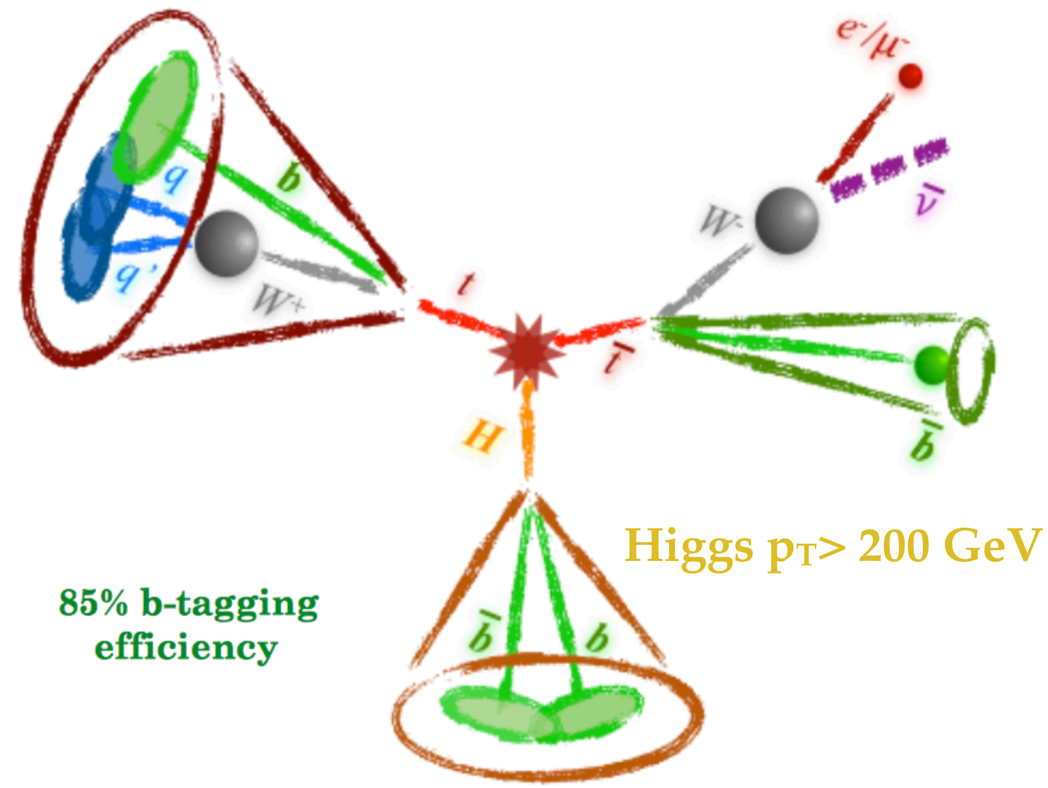
particle level unfolding

Search for ttH in high-p_T regime

Event Selection: $\sqrt{s}$=13 TeV, $\mathcal{L}$=36.1 fb^{-1}

- ttH → *direct access* to Higgs-top Yukawa coupling
- measuring Y_t provides indirect hints of *new physics*
- *first time* we study a boosted channel in ttH!

- *re-clustering preferred* then trimmed large-R jets
  - *better sensitivity* to signal strength μ
  - no systematic of large-R jets
  - anti-k_T jets (R=0.4) used to re-cluster the large-R jets (R =1.0, 200<p_T<1500GeV, |η| < 2, m < 50 GeV) in this analysis.



Analysis strategy:

- Signal identification: ***MVA*** using event kinematics and topology, b-tagging information:
  - identification of very low signal over a very large background
- ***Combination*** with the resolved channel:
  - single-lepton
  - di-lepton

- exactly one lepton;
- one Higgs candidate p_T > 200 GeV with two associated b-jets
- one Top candidate pT > 250 GeV with one associated b-jet and one non-b-jet
- one b-jet outside the two re-clustered jets.

IN PROGRESS!!
Latest result:
ATLAS-CONF-2016-080

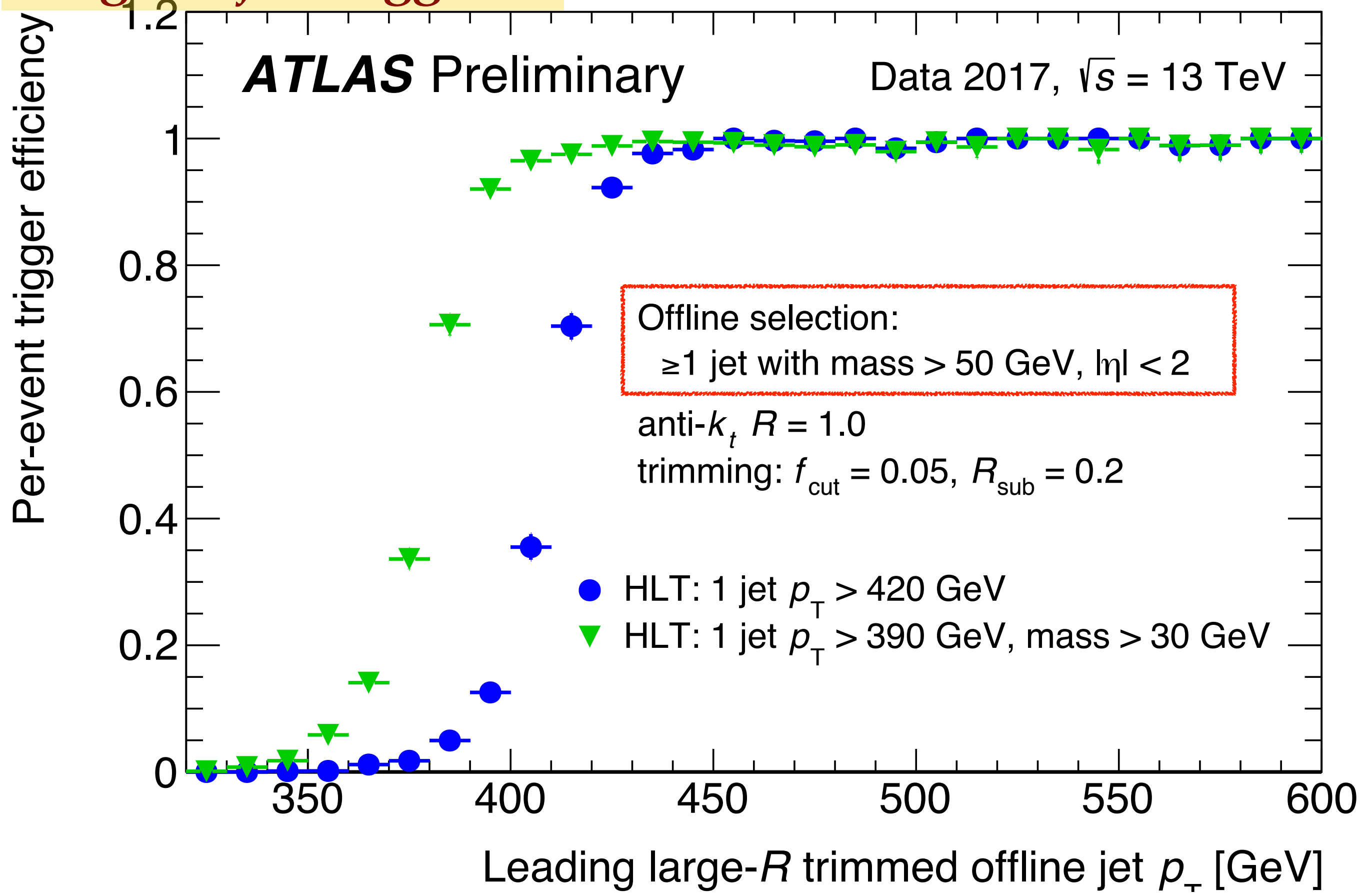
Motivation for adding the boosted category to the resolved channel:
- fewer **combinatorial background**;
- easier **system reconstruction** thanks to the re-clustered techniques;
- testing new methods, measuring the Higgs p_T in ttH events (useful for differential σ analysis).

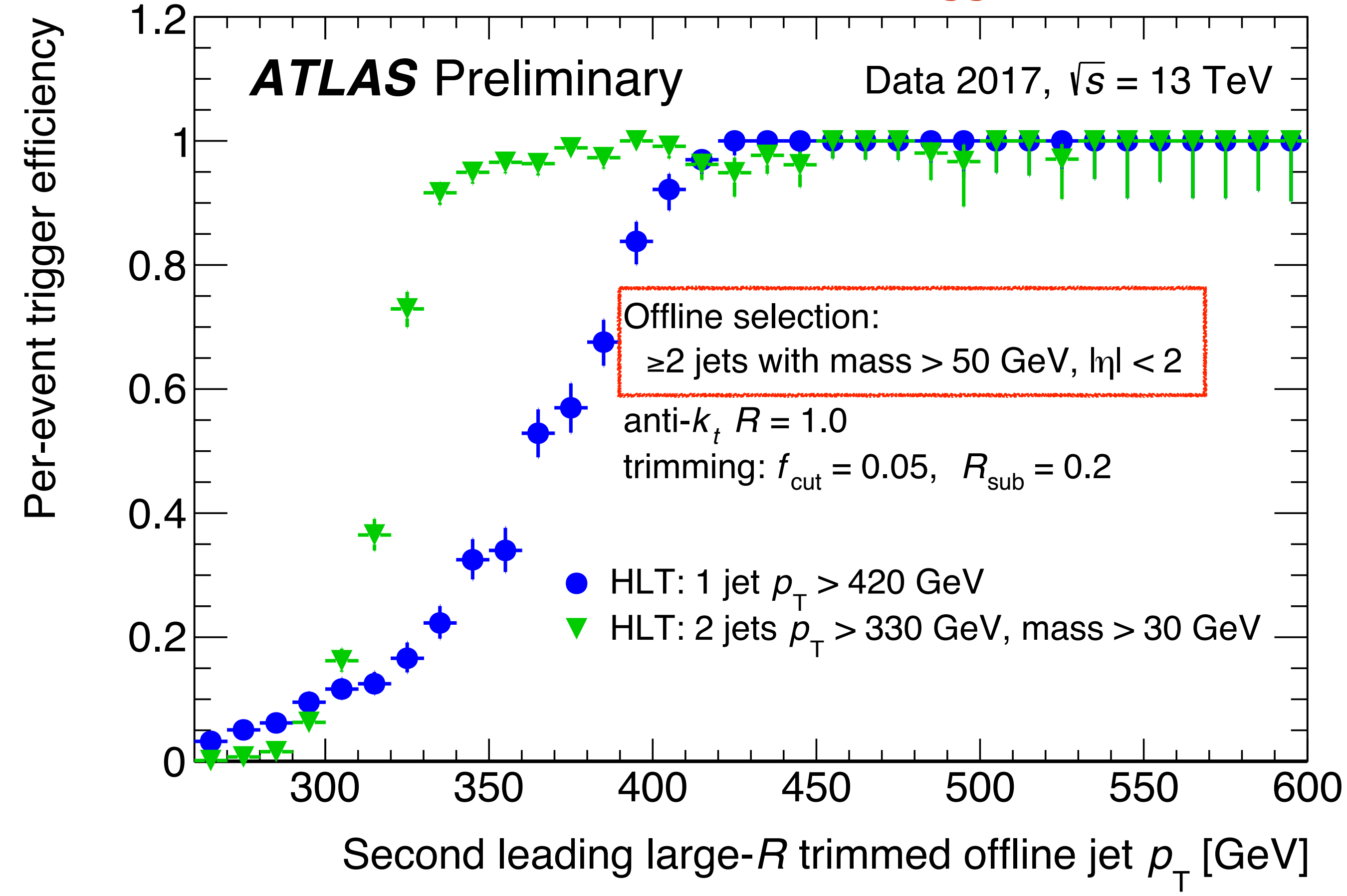
On the trigger side..

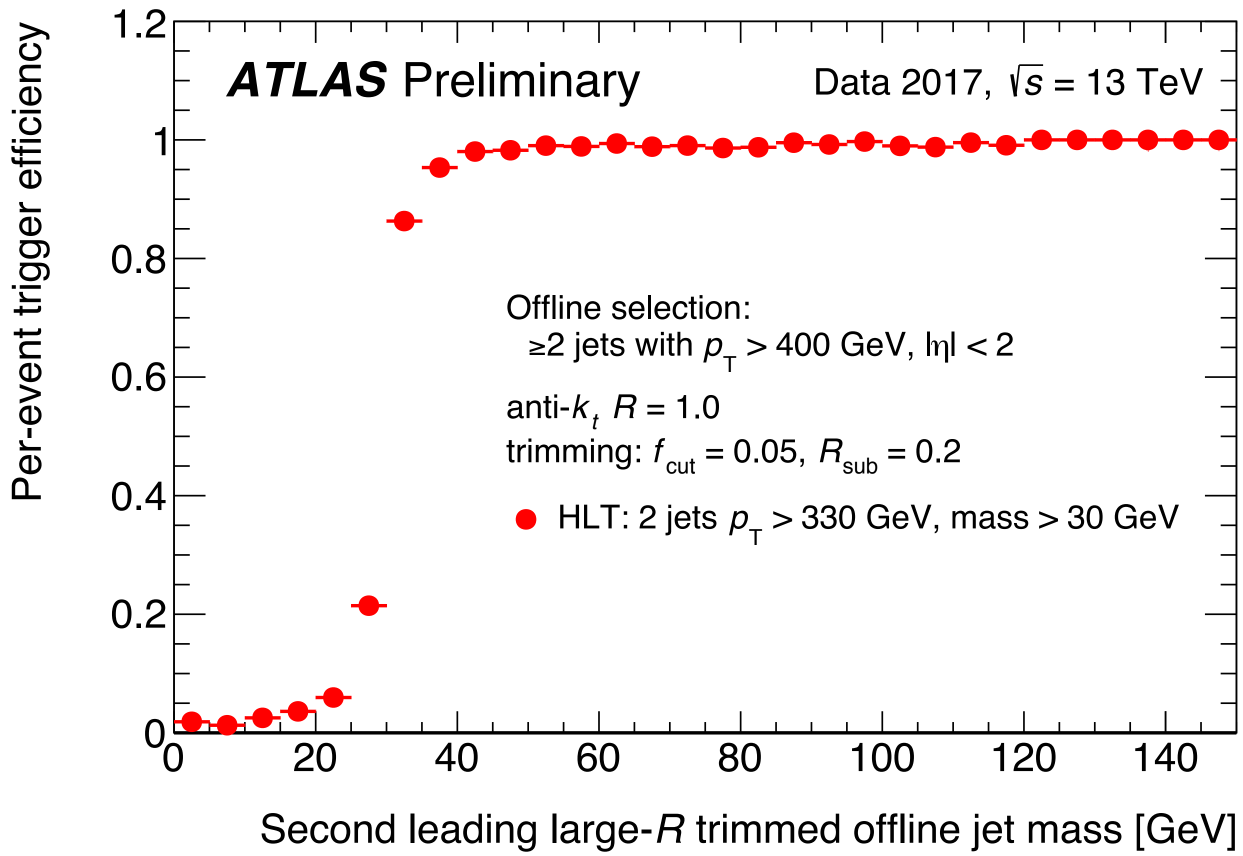
jet triggers: high p_T thresholds or prescale. Large-R jets at trigger level allows lower thresholds and good QCD bkg suppression.

Large-R jet triggers:

Trigger Public Plots







- ✤ large-R jet trigger applied to trimmed jets with |η|< 2.0 and mass > 50 GeV
  - ✤ efficiently **suppresses QCD** bkg
- ✤ offline/online trimming difference: trigger jets f_{cut}=0.04, offline f_{cut}=0.05 to avoid inefficiency on jet mass reconstruction.

Conclusions:

From the experiment…

✤ LHC is collecting ***more and more data***

✤ Detectors and analysis strategies must copy with the ***challenging environmental conditions*** @ 13 TeV

✤ jet physics allow us to *probe QCD and QED predictions* in new phase space regions @ 13 TeV (learning/improving from 8 TeV measurements) → **boosted**

✤ **New triggers** implementing online large-R jets trimming very efficient

…to the analyses:

✤ Now more than ever jet grooming/tagging are fundamental to select "interesting jets":

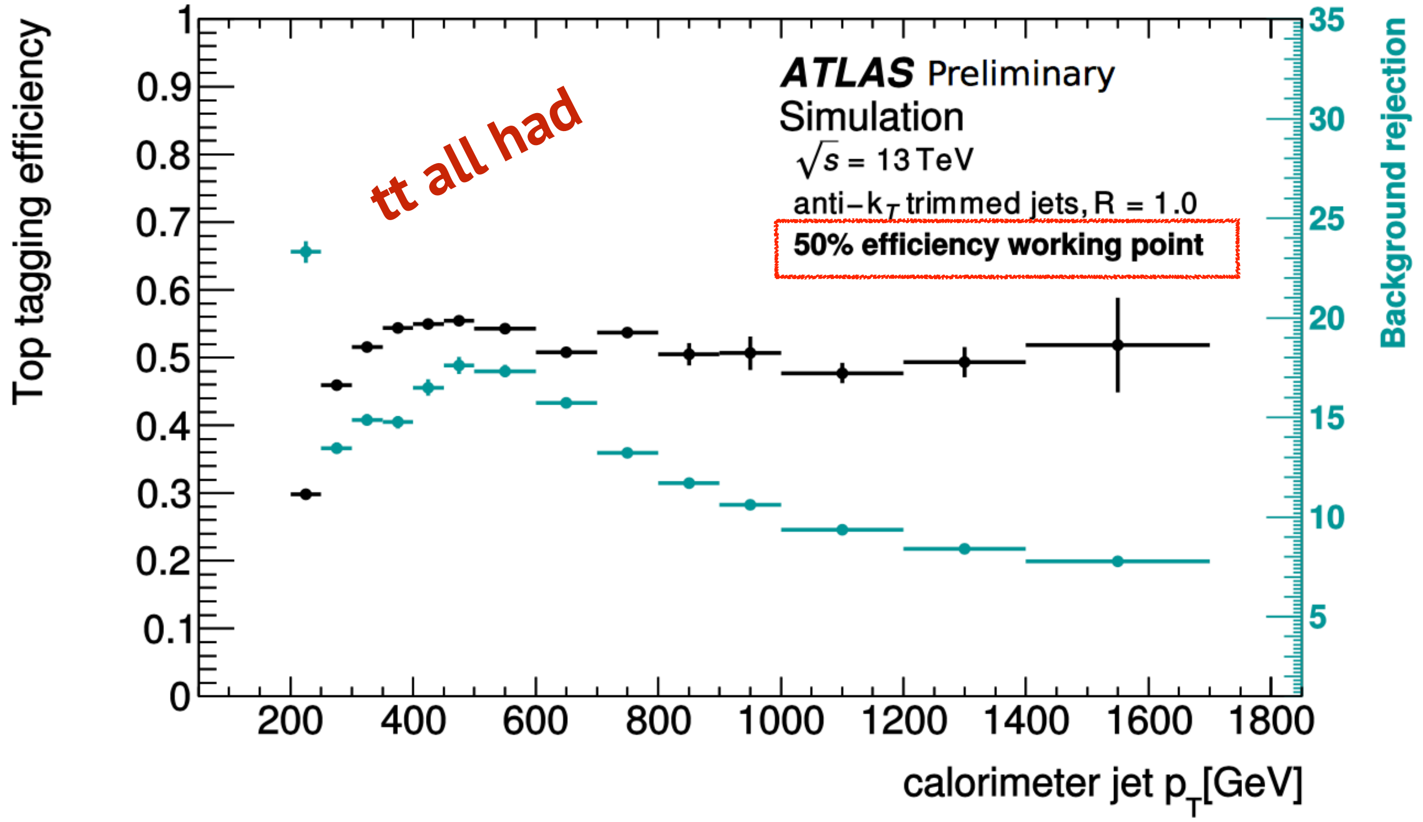
- ✤ *high performances* of ATLAS standard trimming/tagging techniques
- ✤ *jet-reclustering* interesting for analysis targeting specific processes/regions

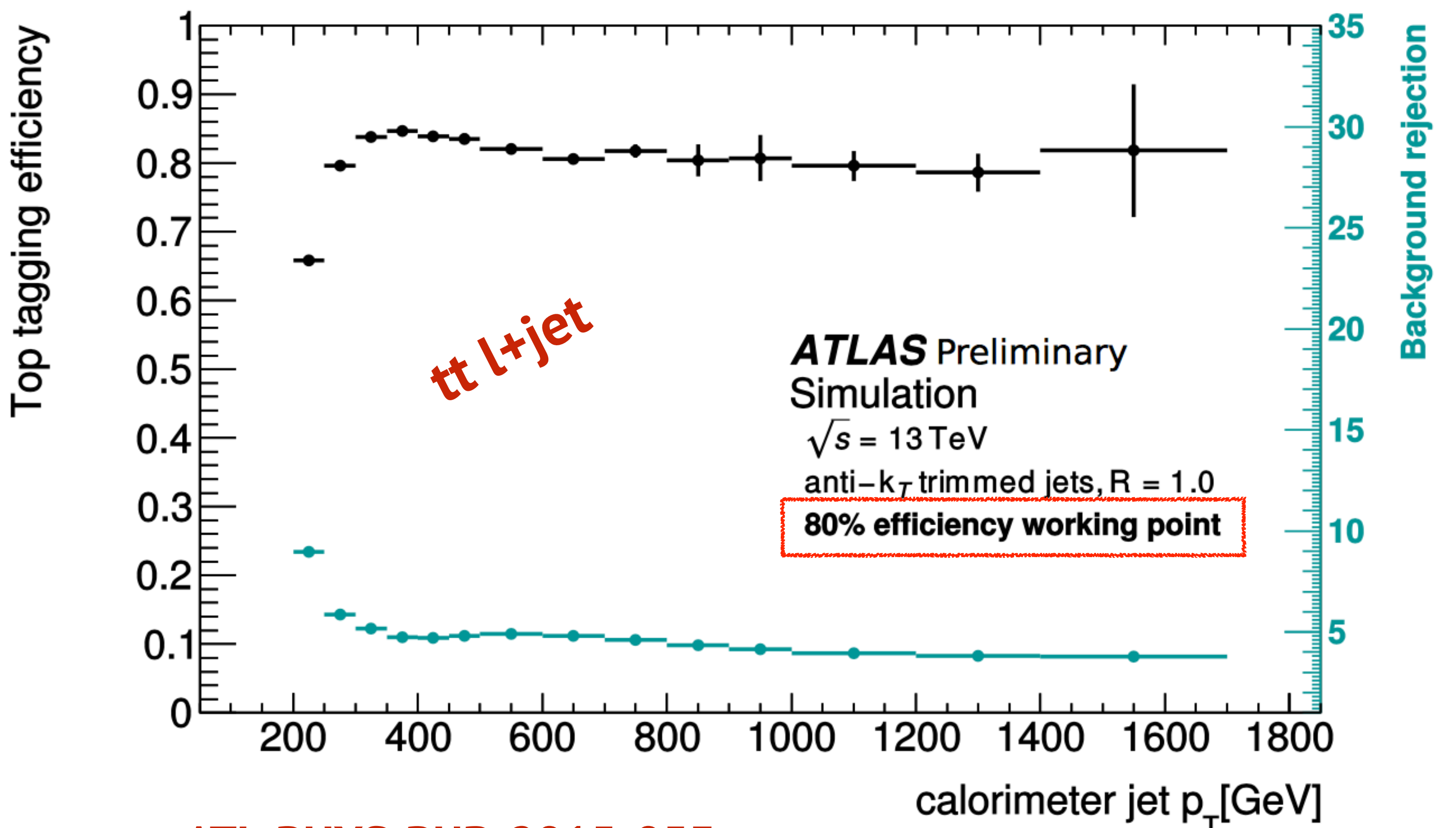
..many new results are coming!

Thanks for the attention!!

Backup/
Additional Material

tt differential cross-section - Tagger performance:





ATL-PHYS-PUB-2015-053

- Signal: $Z' \to tt$
- Background: dijet event

Jet matching:

- reconstructed jet within $\Delta R < 0.75$ of generator-level jet,
- signal generator-level jets $p_T > 200$ GeV and $\Delta R < 0.75$ of a hadronically decaying top quark.

Efficiency: $Jet_{sig}^{tagged} / Jet_{sig}$

Rejection: $Jet_{bkg}^{tagged} / Jet_{bkg}$

50% WP:

- $p_T > 350$ GeV top fully contained
- 200 GeV $< p_T < 350$ GeV fully contained fraction rising

80% WP: already flat at $p_T > 250$ GeV → suitable for searches not sculpting top p_T

Monte Carlo Samples

Physics process	Generator	Cross-section normalisation	PDF set for hard process	Parton shower	Tune
$t\bar{t}$ Signal	Powheg-Box v2	NNLO+NNLL	CT10	Pythia 6.428	Perugia2012
$t\bar{t}$ PS syst.	Powheg-Box v2	NNLO+NNLL	CTEQ6L1	Herwig++2.7.1	UE-EE-5
$t\bar{t}$ ME syst.	MadGraph5_ aMC@NLO	NLO	CT10	Herwig++2.7.1	UE-EE-5
$t\bar{t}$ rad. syst.	Powheg-Box v2	NNLO+NNLL	CT10	Pythia 6.428	'radHi/Lo'
s top t-channel	Powheg-Box v1	NLO	CT10f4	Pythia 6.428	Perugia2012
s top s-channel	Powheg-Box v2	NLO	CT10	Pythia 6.428	Perugia2012
s top Wt-channel	Powheg-Box v2	NLO+NNLL	CT10	Pythia 6.428	Perugia2012
$t\bar{t}$+$W/Z/WW$	MadGraph5_ aMC@NLO	NLO	NNPDF2.3LO	Pythia 8.186	A14
$W(\rightarrow \ell\nu)$+ jets	Sherpa 2.1.1	NNLO	CT10	Sherpa	Sherpa
$Z(\rightarrow \ell\ell)$+ jets	Sherpa 2.1.1	NNLO	CT10	Sherpa	Sherpa
WW, WZ, ZZ	Sherpa 2.1.1	NLO	CT10	Sherpa	Sherpa

Measurements of tt differential cross-section - LJ :

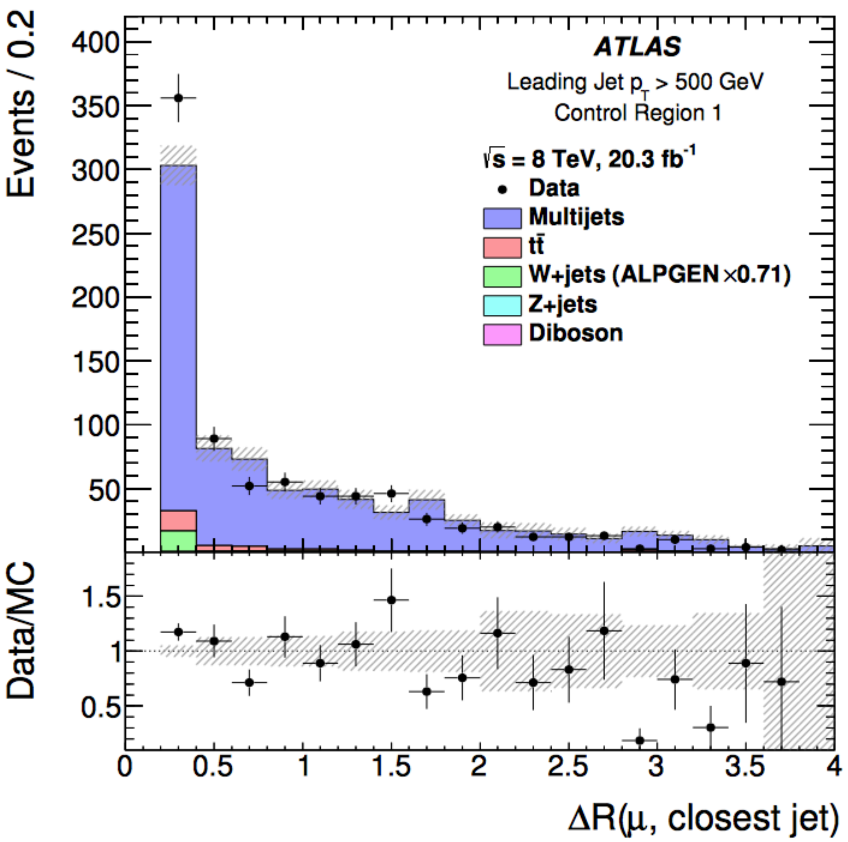
Level	Detector		Particle
Topology	Resolved	Boosted	
Leptons	$\|d_0/\sigma(d_0)\| < 5$ and $\|z_0 \sin\theta\| < 0.5$ mm Track-Calo-based Isolation $\|\eta\| < 1.37$ or $1.52 < \|\eta\| < 2.47$ (e) $\|\eta\| < 2.5$ (μ) E_T (e), p_T (μ)> 25 GeV		$\|\eta\| < 2.5$ $p_T > 25$ GeV
Small-R jets	$p_T > 25$ GeV $\|\eta\| < 2.5$ JVT cut (if $p_T < 60$ GeV and $\|\eta\| < 2.4$)		$\|\eta\| < 2.5$ $p_T > 25$ GeV
Num of small-R jets	≥ 4 jets	≥ 1 jets	
E_T^{miss}, m_T^W		$E_T^{miss} > 20$ GeV, $E_T^{miss} + m_T^W > 60$ GeV	same as detector level
Leptonic top		At least one small-R jet with $\Delta R(\ell$, small-R jet) < 2.0	
Hadronic top	kinematic top quark reconstruction for detector and particle level	the leading-p_T trimmed large-R jet has: 300 GeV $< p_T <$ 1500 GeV, $m > 50$ GeV, TopTagging at 80% efficiency ΔR(large-R jet, small-R jet)> 1.5, $\Delta\phi(\ell$, small-R jet)> 1.0	**Boosted:** $300 < p_T < 1500$ GeV Top-tagging: $m > 100$ GeV, $\tau_{32} < 0.75$
b-tagging	at least 2 b-tagged jets	at least one of: 1) the leading-p_T small-R jet with $\Delta R(\ell$, small-R jet)< 2.0 is b-tagged 2) at least one small-R jet with ΔR(large-R jet, small-R jet)< 1.0 is b-tagged	ghost-matched B-hadron

tt differential cross-section - AH:

- at least 2 anti-k_t $R = 1.0$ jets with $p_{\mathrm{T}} > 350$ GeV,
- at least 1 anti-k_t $R = 1.0$ jet with $p_{\mathrm{T}} > 500$ GeV,
- at least 2 anti-k_t $R = 0.4$ jets with $p_{\mathrm{T}} > 25$ GeV,
- the masses of both $R = 1.0$ jets be within 50 GeV of the top-quark mass,
- the two leading $R = 1.0$ jets be associated with a b-hadron in the final state using a ghost-matching technique as described in Ref. [43], and
- no electrons or muons with $p_{\mathrm{T}} > 25$ GeV be in the event.

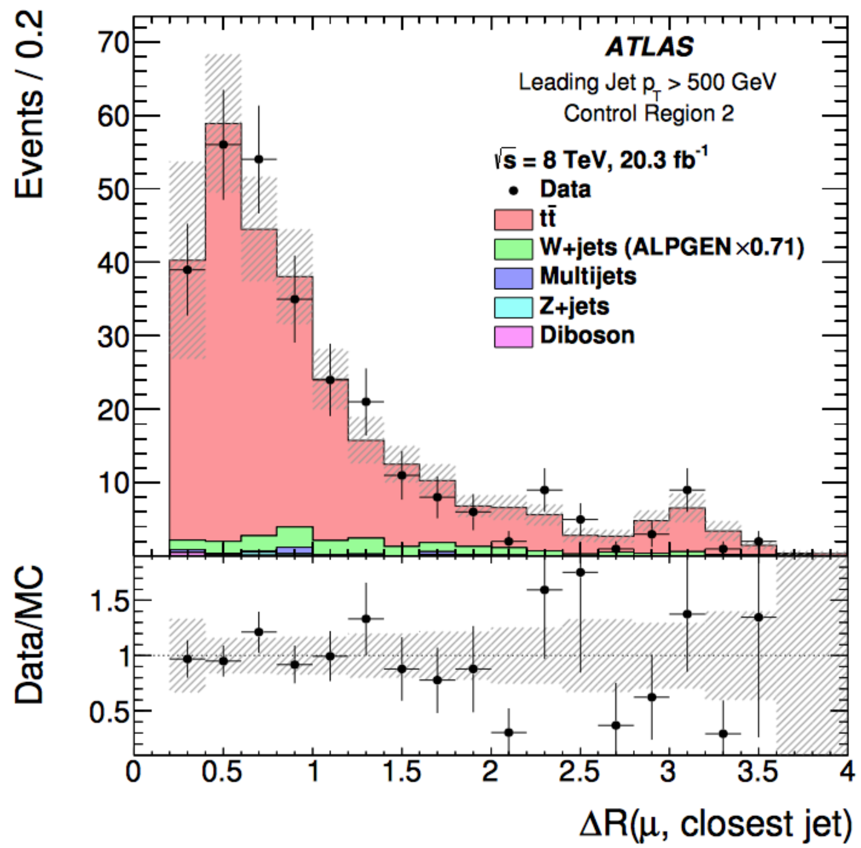
Collinear W @ 8 TeV

CERN-EP-2016-182



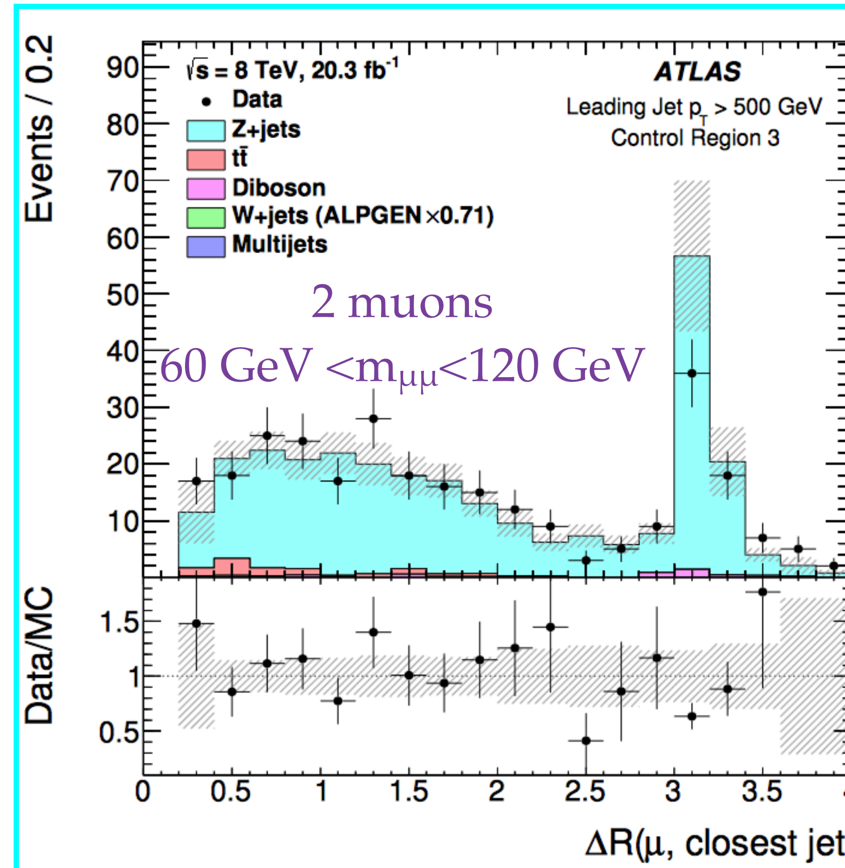
- ✤ 93% purity of dijet events
- ✤ revert signal region isolation
- ✤ μ p_T > 38 GeV
  (to pass the non iso trigger)
- ✤ ΔR(μ, closest jet) > 0.2

Normalization correction:
1.134 ± 0.054 (stat)



- ✤ 91% purity of tt events
- ✤ at least 2 b-tagged jets

Normalization correction:
0.861 ± 0.061 (stat)



- ✤ 94% purity of Z+jets events
- ✤ exactly two μ
- ✤ 60 GeV < $m_{\mu\mu}$ < 120 GeV
- ✤ higher p_T μ used to define ΔR

Normalization correction:
0.705 ± 0.052 (stat)

Collinear W @ 8 TeV

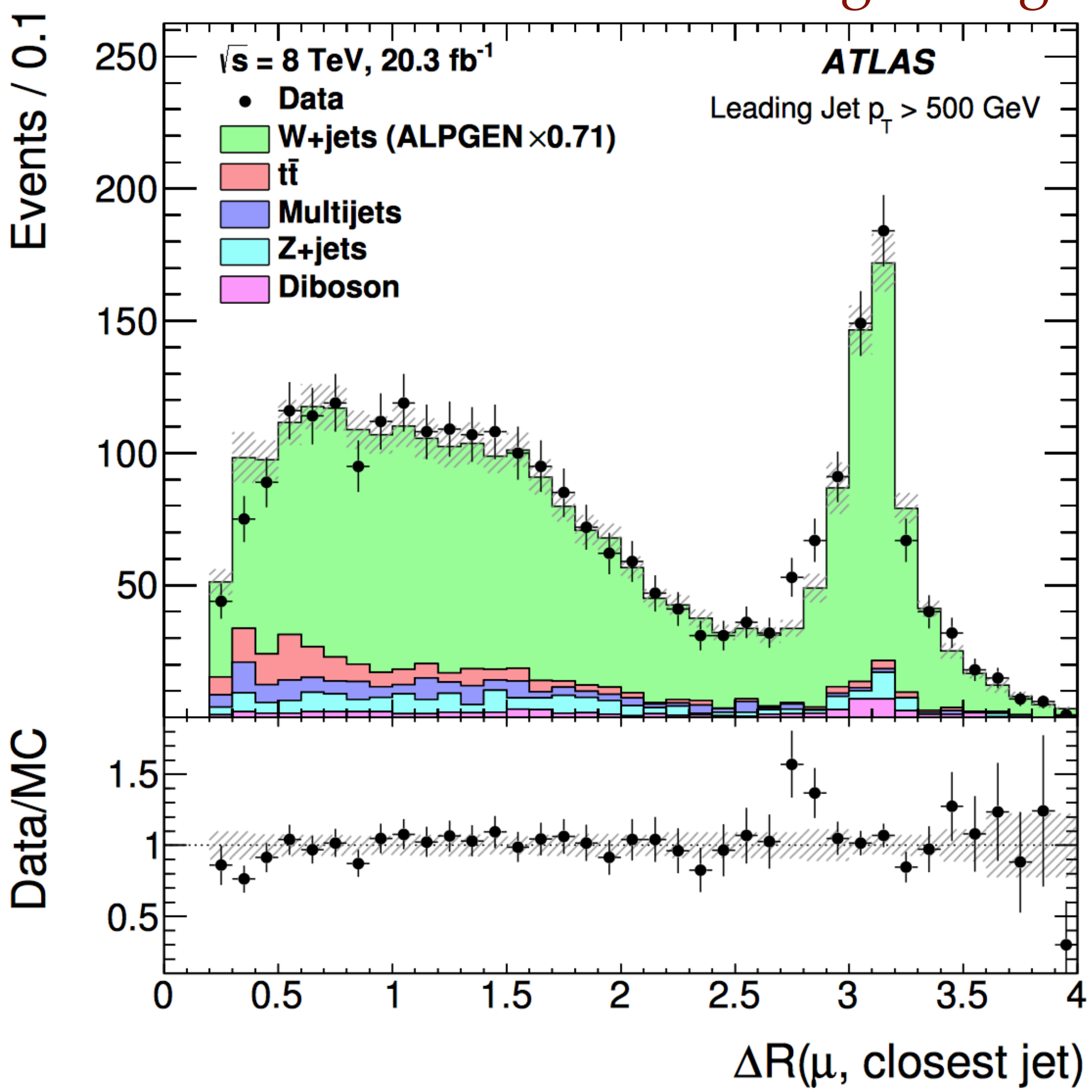
CERN-EP-2016-182

Systematic Source	$0.2 < \Delta R < 2.4$	$\Delta R > 2.4$	Inclusive
Scaling of dijets to data	0.4%	0.1%	0.3%
Scaling of $t\bar{t}$ to data	0.6%	0.2%	0.5%
Scaling of Z + jets to data	0.6%	0.3%	0.5%
Jet energy scale	4.6%	5.8%	5.0%
b-tagging efficiency	3.7%	1.2%	2.9%
Data/MC disagreement for dijets	0.9%	0.6%	0.8%
Data/MC disagreement for $t\bar{t}$	1.2%	0.4%	1.0%
Data/MC disagreement for Z + jets	0.6%	1.5%	0.9%
Diboson background estimate	2.2%	0.1%	1.5%
Unfolding dependence on prior	1.1%	1.8%	1.3%
Muon momentum scale and resolution	0.0%	0.1%	0.1%
Muon reconstruction efficiency	0.4%	0.4%	0.4%
Muon trigger efficiency	2.0%	1.9%	1.9%
Jet energy resolution	0.6%	0.8%	0.6%
MC background statistical	2.4%	1.8%	2.3%
MC response statistical	1.7%	2.2%	1.9%
Total systematic (excluding luminosity)	7.6%	7.4%	7.3%
Luminosity	1.9%	2.0%	2.0%
Data statistical	2.7%	3.6%	2.2%

Collinear W @ 8 TeV

CERN-EP-2016-182

Uncorrected distribution in signal region



Process	$0.2 < \Delta R < 2.4$	$\Delta R > 2.4$	Inclusive
Dijets	5%	2%	4%
$t\bar{t}$	7%	2%	5%
Z + jets	6%	4%	5%
Dibosons	2%	4%	3%
W + jets	80%	88%	82%
Data	1907	833	2740

Cross section measurement

Process	$\sigma(W(\to \mu\nu) + \geq 1$ jet) [fb]
Data ($\sqrt{s}$ = 8 TeV, 20.3 fb^{-1})	169.2 ± 3.7 (stat.) ± 12.3 (syst.) ± 3.3 (lumi.)
ALPGEN+PYTHIA6 W+jets	236.6 ± 1.1 (stat.)
PYTHIA8 $W+j$ & jj+weak shower	134.8 ± 0.9 (stat.) ± 7.3 (pdf)
SHERPA+OpenLoops $W+j$ & $W+jj$	183 ± 25 (scale)
$W + \geq 1$ jet N_{jetti} NNLO	181 ± 14 (scale)

Process	$\sigma(W(\to \mu\nu) + \geq 1$ jet, $0.2 < \Delta R < 2.4$) [fb]
Data ($\sqrt{s}$ = 8 TeV, 20.3 fb^{-1})	116.2 ± 3.2 (stat.) ± 8.8 (syst.) ± 2.3 (lumi.)
ALPGEN+PYTHIA6 W+jets	167.1 ± 0.9 (stat.)
PYTHIA8 $W+j$ & jj+weak shower	83.4 ± 0.7 (stat.) ± 4.4 (pdf)
SHERPA+OpenLoops $W+j$ & $W+jj$	128 ± 20 (scale)
$W + \geq 1$ jet N_{jetti} NNLO	123 ± 9 (scale)

Process	$\sigma(W(\to \mu\nu) + \geq 1$ jet, $\Delta R > 2.4$) [fb]
Data ($\sqrt{s}$ = 8 TeV, 20.3 fb^{-1})	53.0 ± 1.9 (stat.) ± 3.9 (syst.) ± 1.0 (lumi.)
ALPGEN+PYTHIA6 W+jets	69.5 ± 0.6 (stat.)
PYTHIA8 $W+j$ & jj+weak shower	51.4 ± 0.6 (stat.) ± 2.9 (pdf)
SHERPA+OpenLoops $W+j$ & $W+jj$	55 ± 5 (scale)
$W + \geq 1$ jet N_{jetti} NNLO	58 ± 5 (scale)

	σ_{LO} [fb]	σ_{NLO} [fb]	σ_{NNLO} [fb]
8 TeV	57^{+13}_{-10}	160^{+35}_{-27}	187^{+5}_{-12}

Collinear W @ 8 TeV

CERN-EP-2016-182