# Heavy-flavour jets at ATLAS & CMS

Andy Buckley, University of Glasgow for the ATLAS Collaboration Duong Nguyen, University at Buffalo for the CMS Collaboration

Jets and substructure at the LHC, 1 June 2021







# Introduction

# Heavy quark/jet production crucial both from QCD and EW/BSM physics perspectives

**EW/BSM:** b-jet signatures ubiquitous in many BSM models, top-quarks, and SM & BSM H→bb channels

**QCD:** understanding heavy-parton effects:

- matching production modes between high-scale/wide-sep matrix element, and softer/collinear parton shower
- hard-scatter formalism between PDF / ME
- PS scale choices and mass effect on radiation pattern

### Presenting ATLAS & CMS results focusing on final-state HF jets

split by experiment since different focuses



# Channels and MC models

#### Processes/channels

- Dominated by b-jet processes, since jet c-tagging less prevalent so far
- V+b(b)
- Z→bb
- Also V+c,b hadrons (probe production processes and PDFs)

#### MC models: NLO or matched/merged multileg NLO

- POWHEG: NLO 1-leg; 5F only?
- MG5\_aMC: MG LO multileg, NLO 1-leg, FxFx NLO multileg; 4/5F
- Sherpa: MEPS@NLO NLO multileg (MLM matching scheme); 4/5F and "fusing" hybrid
- Alpgen: LO multileg, 4F. For "academic interest" only in 2021!

13 TeV, 35/fb partial dataset, separate ≥1 and ≥2 resolved b-jet selections

Kinematic variable	Acceptance cut
Lepton $p_{\rm T}$	$p_{\rm T} > 27 { m ~GeV}$
Lepton $\eta$	$ \eta  < 2.5$
$m_{\ell\ell}$	$m_{\ell\ell} = 91 \pm 15 \text{ GeV}$
<i>b</i> -jet $p_{\rm T}$	$p_{\rm T} > 20 { m GeV}$
<i>b</i> -jet rapidity	y  < 2.5
<i>b</i> -jet–lepton angular distance	$\Delta R(b\text{-jet}, \ell) > 0.4$

# Flavour-fraction fits using b-tag discriminant distributions:



b-tagged ak<sub>T</sub> jets, cf. flavour-k<sub>T</sub> discussion point JHEP 07 (2020) 44, arXiv:2003.11960, <u>ATL-STDM-2017-38</u>

#### Measurement uncertainties dominated by:

- flavour-fraction fit in ≥1b selection (b-tag efficiency a significant second)
- b-tag efficiency in ≥2b selection



Full Run-2 and boosted-bb versions in prep 4

JHEP 07 (2020) 44, arXiv:2003.11960, <u>ATL-STDM-2017-38</u>

#### Integrated fiducial cross-sections for ≥1b and ≥2b selections:



Flavour-number scheme most important for ≥1b selection, unsurprisingly:

- 4F significantly undershoots data, esp. at LO. 5F & Sherpa "fusing" within error
- ≥2b selection more tolerant, mostly good match to data

JHEP 07 (2020) 44, arXiv:2003.11960, Plot page

#### Differential results, ≥1b selection:



5F predictions concur, arguably better differentially too. Fusing not helping shapes? MGaMC and Sherpa differences at low and high scales... origin?

JHEP 07 (2020) 44, arXiv:2003.11960, Plot page

#### Differential results, ≥2b selection



bb pair angular separations marginally best from Sherpa 5F. Kinks in MGaMC 5F 4F not helping (bottom panel), slope in Alpgen longitudinal separation

5

 $\Delta R_{hh}$ 

JHEP 07 (2020) 44, arXiv:2003.11960, Plot page

#### Differential results, ≥2b selection:



bb pair  $p_T$ , mass, and scale variable: Sherpa best  $p_T$  shape, all underestimate high-mass. 5F scale variable modelling better than 4F equivalents.

Measurement of b-jet fragmentation moments, via secondary vertices and track-jets

13 TeV analysis, partial Run 2 dataset of 36/fb

#### Physics objects & selection:

- R = 0.4 calo jets with standard ATLAS b-tagging (70% eff)
- Event selection: = 2 OS e+ $\mu$ , = 2 jets,  $\Delta R_{jj}$  > 0.5; tag & probe, both ways Variable-radius track jets ghost-associated to calo jets
- Identification of VR track-jet PV and SV tracks: SV/(PV+SV) ~ b-hadron/b-guark



#### ATLAS-CONF-2020-050

#### n<sub>ch</sub> moment

- Number of charged tracks associated to the secondary vertex = b-hadron
- Data produces a significant shift from the input Bayesian prior
- Reweightings of Pow+Py8 to probe-jet n<sub>sv</sub> in data used as a systematic
- Useful cross-check of hadron fractions and decay modes: Sherpa improvements in > 2.2.8



0.00

0

2 3

*n<sub>b</sub><sup>ch</sup> bin* 

5

4

#### ρ moment

- Ratio of b-hadron p<sub>T</sub> to average of charged lepton p<sub>T</sub>s — comparison of the b momentum to the tt parent event scale
- Lepton p<sub>T</sub> more precisely measurable than the b-jets or the tt system
- Sensitive to QCD radiation not contained in the b-jets
- Mostly good MC descriptions (within errors and unfolding prior)





0.5

0.000

0

ATLAS Preliminary

 $\sqrt{s} = 13 \text{ TeV}, 36 \text{ fb}^{-1}$ 

Data

ATLAS Preliminary

 $\sqrt{s} = 13 \text{ TeV}, 36 \text{ fb}^{-1}$ 

1.5

Tracking

Pileup

Signal modeling

Total

······ Other

\_.\_.\_

probe-jet a

Prior prediction

with uncertainty

with uncertainty

Posterior prediction

#### ATLAS-CONF-2020-050

### z<sub>ch,L</sub> (longitudinal) moment

- Ratio of b-hadron p<sub>T</sub> projection along its parent track-jet axis to the track-b-jet p<sub>T</sub>
- Experimental proxy for the theoretical fragmentation function: tunings from LEP and SLD valid?
- Descriptions generally good, with Herwig a slight outsider (and this is much improved from version 7.0.4)



z<sup>ch</sup> bin

#### 13

ATLAS-CONF-2020-050

# **ATLAS b-fragmentation**

### z<sub>ch,T</sub> (transverse) moment

- Ratio of b-hadron p<sub>T</sub> wrt track-b-jet p<sub>T</sub> (the p<sub>T,rel</sub> projected p<sub>T</sub> wrt the parent jet axis proved difficult to measure)
- Correlated with longitudinal projection, plus some residual information about directional kicks from radiation
- Descriptions generally good, within significant uncertainties



0.5

1.5

#### Also systematic studies of sensitivity to Sherpa tunes, Powheg+Herwig versions, and Powheg+Pythia8 r<sub>B</sub> tunes and alphaS variations:



# ATLAS $g \rightarrow bb$ at small opening angles

Focus on final-state g $\rightarrow$ bb splitting kinematics in boosted region p<sub>T</sub> > 450 GeV R=1 akT jets, 33/fb of 13 TeV pp data

g manna trank b g q q

Again using ghost-associated track-jets for improved angular resolution

Require two b-tagged VR track-jets as b proxies, using 60% working point

Flavour fit again central. Based on signed impact-parameter distributions for subleading track in each track jet

Fit  $s_{d0}$  flavour fractions per observable bin, e.g.  $\Delta R_{bb}$  right with Py8 templates. Mostly slight BB-fraction overestimates **Observables:** 

- m<sub>bb</sub> / p<sub>T</sub>
- z(p<sub>T</sub>)
- $\Delta R_{bb}$
- Polarisation
   angle Δθ ⇒



Phys. Rev. D 99 (2019) 052004,



Plot page

arXiv:1812.09283,

# ATLAS $g \rightarrow bb$ at small opening angles

Phys. Rev. D 99 (2019) 052004, arXiv:1812.09283,

Flavour-fraction fits, per bin of each observable (BB ~good, large differences in B and L+C) Plot page



Scale-fraction observables mostly well modelled — deviations at low ratio values.  $\Delta R$  good to 10% accuracy,  $\Delta R > 0.2$ . Poor  $\Delta \theta$  data/MC: *insensitive* to Py8 gpol modelling? <sup>16</sup>

# ATLAS jet mass in Z(bb)γ

10 GeV

Events /

Data / Fit

1000

600

400

200

1.4

1.2

0.8

0.6

Phys. Lett. B 812 (2021) 135991, arXiv:1907.07093, <u>Plot page</u>

# Measurement of boosted $Z \rightarrow bb$ jet mass and fiducial cross-sections

200 GeV large-R jet & 175 GeV photon tag to reduce QCD backgrounds; still dominated by  $\gamma$ +jet process:

b-tagged track-jets used as a double-b tagger

 $N_{bjet}$ =1 and non-tight photon CRs used to estimate the  $\gamma$ +jet bkg for subtraction

Two jet grooming methods: trimming and soft-drop (beta = 0,  $z_{cut}$  = 0.1)+constituent subtraction+SoftKiller

#### Sensitive to reco and jet-grooming schemes:



Reconstructed jet mass [GeV]

# ATLAS jet mass in $Z(bb)\gamma$

#### Phys. Lett. B 812 (2021) 135991, arXiv:1907.07093, Plot page

#### **Results:**

- Dedicated MC-based calibration. background subtraction (dominant  $\gamma$ +jets from data, others from floated MC templates), and iterative-Bayes unfolding
- Differences in line-shape and resulting cross-section precision depending on grooming
- Soft-drop jets (+CS+SK) broader and more symmetric than trimmed jets. Precision on extracted xsec slightly better via the trimmed jets
- Agreement of shape with LO Sherpa: need for other MC cmps



160

160

### CMS W+c cross section measurement

- Probe s-quark content of protons
- Most recent results at 13 TeV using ~35 fb<sup>-1</sup> of data. Other results at 7 and 8 TeV are available.
- W bosons are measured in the leptonic channels (electron, muon) while c-quarks are identified through hadronic decays of charm hadrons
- Dominant background from W+g(->cc), suppressed using charge correlation between charm and W (OS)
  - Use sign of total charge of tracks associated with D-meson decays for that of c quark









### CMS W+c inclusive cross section at 13 TeV



Eur. Phys. J. C 79 (2019) 269

- Presence of jet is not required: charm quark p<sub>T</sub> defined as p<sub>T</sub> sum of tracks in cone ∆R≤0.4 around the D<sup>\*</sup>(2010)<sup>±</sup>
- NLO predictions: MCFM 6.8, mc = 1.5 GeV,  $\mu_R = \mu_F = m_W$
- Uncertainty in predictions: PDF, , scales ( $\mu_R$ ,  $\mu_F$  varied simultaneously by a factor of 0.5 or 2)
- Good agreements between predictions and data except ATLASepWZ16 (ATLAS W+c results included in PDF fits)

### CMS W+c differential cross sections at 13 TeV

Predicted W+c differential cross sections agree with data, except the ones using ATLASepWZ16





21

### Strange quark PDF fits using CMS W+c data

The strange quark PDF uncertainty is reduced in the new PDF fit using CMS 7 and 8 TeV lepton charge asymmetry and W+c data at 7 and 13 TeV



### CMS Z+c cross section measurement at 13 TeV

- Direct access to charm PDF
  - $\circ$  Intrinsic charm contributions enhance Z+c production rate, especially at high  $p_{T}$  of c jets and Z boson
- Measure inclusive and differential cross sections as a function of c-jet and Z boson  $p_T$  using 35.9 fb<sup>-1</sup> data at 13 TeV
- c-quarks are tagged using a charm tagger on reconstructed jets
- Backgrounds: Z+b, Z+light (udsg) jets, tt, dibosons
  - Z+b and Z+light jets are estimated using a template fit on secondary vertex mass of tagged jets distribution.
  - Other backgrounds are estimated in simulation

#### $_{\underline{\times 10}^3}$ CMS 35.9 fb<sup>-1</sup> (13 TeV) Events / 0.2 GeV Data (Electron channel) C jets B jets Light jets Top and dibosons Data / MC m<sub>sv</sub> [GeV] CMS 35.9 fb<sup>-1</sup> (13 TeV) ×1∩<sup>3</sup> Events / 10 GeV 40 Data (Electron channel) C jets B jets iaht iets Fop and dibosons 20 Data / MC 150 50 100 200 250 p\_\_\_\_\_\_\_\_\_[GeV]

JHEP 04 (2021) 109

23

### CMS Z+c cross sections at 13 TeV

Measured inclusive cross sections:  $405.4 \pm 5.6$  (stat)  $\pm 24.3$  (syst)  $\pm 3.7$ (theo) pb

MADGRAPH5 amc@NLO (NLO): 524.9 ± 11.7 pb

- Good agreement between data and MG5 aMC (LO)
- Both MG5\_aMC (NLO) and SHERPA overestimate data



300

### CMS Z+BB measurement at 7 TeV

- Measure the production of a Z boson in association with B hadrons as a function of angular correlation especially the B-hadron pair production at small angular separation where there are significant theoretical uncertainties in the description of collinear production
  - $\label{eq:relation} \Delta R_{_{BB}} \text{ direct test of the modeling of different } pp \rightarrow ZbbX \\ \text{production modes}$
  - $\min \Delta R_{ZBB}$  is sensitive to radiation of Z from b quark
- B hadrons are identified by displaced secondary vertices without reconstructed jets → probe separation regions restricted by jet reconstruction









### CMS Z+BB cross sections at 7 TeV



- ALPGEN well describes data in collinear regions ( $\Delta R_{BB} < 0.7$ ) while MADGRAPH and amc@NLO are lower than data
- At large  $\Delta R_{BB}$ , all predictions agree with data

### CMS Z+BB cross sections at 7 TeV



• Data are consistent with predictions in  $min \Delta R_{7BB}$  except for region above 2

### CMS measurement of angular correlation between BB at 7 TeV

JHEP03 (2011) 136

- Measure BB production at 7 TeV
- B hadrons are identified using reconstructed secondary vertex
   → probe small separation angle
- Simulation is normalized to regions > 2.4



- None of the predictions describe the data very well
- amc@NLO does not describe the data at small angle where gluon splitting process is significant

### CMS Z+b jets cross section measurement at 8 TeV

- Sensitive to b quark PDF and details of pQCD calculations (5FS vs 4FS)
- Cross sections are measured in the dilepton channels and b-quarks are tagged using a b jet tagger on reconstructed jets
  - B tagging efficiency ~50% with 0.1% (1%) for light
     (c) jets
- Differential cross sections as a function of Z and b jets kinematic and angular correlation variables
- Z+b signals are extracted from a binned maximum likelihood fit based on secondary vertex mass of b-jets
- Dominated uncertainties: unfolding and secondary vertex mass fit

 $\sigma_{\rm fid}(\rm pp \rightarrow Z + (\geq 1b)) = 3.55 \pm 0.12 \,(stat) \pm 0.21 \,(syst) \, \rm pb$   $\sigma_{\rm fid}(\rm pp \rightarrow Z + (\geq 2b)) = 0.331 \pm 0.011 \,(stat) \pm 0.035 \,(syst) \, \rm pb$ MADGRAPH5\_aMC@NLO:  $4.23^{+0.27}_{-0.37} \, \rm pb \,(Z+1b)$   $0.356^{+0.030}_{-0.031} \,(Z+2b)$  EPJC 77 (2017) 751





### CMS Z+b jets cross sections at 8 TeV



- Predictions are obtained using NNPDF 3.0 PDF with CUETP8M1 tune
- LO predictions are provided by Madgraph with 4FS and 5FS. Overall cross section is rescaled to NNLO by FEWZ 3.1
- NLO predictions from Madgraph\_amc@NLO and POWHEG in 5FS
- Both NLO calculations tend to overestimate data

### CMS Z+2 b jets cross section at 8 TeV



- Z+2 b jets differential cross sections as functions of dijet observables
  - Di-b jets invariant mass is important for  $ZH(\rightarrow bb)$  measurements and new physics searches
  - $\circ$  Small  $\varDelta R_{_{hh}}$  regions are important to test gluon splitting contribution
  - A<sub>7bb</sub> (anguĨar asymmetry) is sensitive to additional gluon emissions
- NLO generators provide good description of data

### CMS cross section ratios of Z+c and Z+b at 13 TeV

- Measure the cross section ratio between Z+c and Z + b productions to obtain high precision data (systematic uncertainties are reduced in ratios)
- Also based on secondary vertex mass template fits to separate Z+c and Z+b jets events
- Experimental results are unfolded to the particle levels
- Prediction is consistent with data
  - Parton level MCFM cross sections are obtained using different PDF sets



### CMS W+2 b jets cross section measurement at 8 TeV

- Fiducial definition: lepton  $p_T > 30$  GeV,  $|\eta| < 2.1$ , exactly 2 b-jets with  $p_T > 25$  GeV and  $|\eta| < 2.4$  and no other jets with  $p_T > 25$  GeV and and  $|\eta| < 2.4$
- W+bb yield is estimated from a binned maximum likelihood fit to data in the transverse mass distribution
- Theoretical predictions:
  - MCFM v7.0, MSTW2008 PDF
  - MADGRAPH interface with Pythia6 or Pythia8 in five- or four-flavour scheme. PDFs are CTEQ6L and NNLO for 4- and 5-flavour schemes, respectively
- The observed fiducial cross section is consistent with predictions within one standard deviation.



# Conclusions

- ATLAS summary: resolved Z+b-jet studies at 13 TeV, plus detailed digs into g→bb splitting and b-fragmentation (both using displaced vertices)
- **CMS summary:** W+c cross section measurements and demonstrated effects of the results on strange quark PDF, performed Z+c (13 TeV) and Z+b (8 TeV) together with Z+c/Z+b cross section ratio (13 TeV) measurements to test theoretical predictions

#### • Status of data vs MC/theory:

- Higher-tech calculations mostly improve data-description. Corresponding scale uncertainties: nominal LO scale-choices can sometimes "locally" fit better, without being better physics
  - amc@NLO NLO predictions normalized to NNLO cross sections overestimate CMS Z+c and Z+b jets data and have some tensions with data in collinear production of di-b jets
- Systematics a mix of exp (tracking, tagging) and theory (flav fractions, HF frag & decays, PDF and scales)
- Little case still seen for 4-flavour simulation. With 5F multileg and loops, the LO argument for 4F kinematics is lesser. Sherpa fusing scheme not obviously necessary/helping
- Follow-up on Gehrmann talk re. (N)NLO flavour-k<sub>T</sub> and IR-safety: how to fold or unfold safely? Requires a per-event mapping to particle/reco level from a sufficiently equivalent ME...

#### • Upcoming results:

- ATLAS full Run-2 and boosted Zbb in development; many CMS analyses using full Run-2 data are on-going: Z+b,  $\gamma$ +b, W+c, W+bb
- More event features and more statistics essential, especially at high scales and small  $\Delta R$

#### • Pheno measurement wishlist?



#### Selection based on dileptonic tt, using tag-probe strategy



Selection very pure in signal process <sup>36</sup>

# ATLAS $g \rightarrow bb$ at small opening angles

#### Phys. Rev. D 99 (2019) 052004, arXiv:1812.09283,

ATL-STDM-2017-17

Pr(Detector-level | Particle-level)

Pr(Detector-level | Particle-level)

0.6

0.4

0.2

0.8

0.6

0.4

0.2

0.2

0.4

0.6

0.8 Detector-level  $\Delta \theta_{gpp,gbb}/\pi$ 

0.3

0.4

Detector-level z(p\_)

0.5



Particle-level log( $m_{bb}^{}/p_{T}^{})$ 

0.

Particle-level ΔR(b,b)

### W+c at 13 TeV: uncertainties

- Statistical uncertainties from 3% to 8%
- Systematic uncertainty:
  - Dominant uncertainty from fragmentation of c quark into D\*(2010)±
  - Phenomenological Bowler-Lund function
  - Vary the a and b parameters by 10%

$$f(z) = \frac{1}{z^{r_c \, b \, m_q^2}} (1 - z)^a \exp(-b \, m_\perp^2 / z) \, c$$

Pseudorapidity $[ \eta ]$	[0,2.4]	[0, 0.4]	[0.4, 0.8]	[0.8, 1.3]	[1.3, 1.8]	[1.8, 2.4]
Luminosity	$\pm 2.5$	$\pm 2.5$	$\pm 2.5$	$\pm 2.5$	$\pm 2.5$	$\pm 2.5$
Tracking	$\pm 2.3$	$\pm 2.3$	$\pm 2.3$	$\pm 2.3$	$\pm 2.3$	$\pm 2.3$
Branching	$\pm 2.4$	$\pm 2.4$	$\pm 2.4$	$\pm 2.4$	$\pm 2.4$	$\pm 2.4$
Muons	$\pm 1.2$	$\pm 1.2$	$\pm 1.2$	$\pm 1.2$	$\pm 1.2$	$\pm 1.2$
$N_{\rm sel}$ determination	$\pm 1.5$	$\pm 1.5$	$\pm 1.5$	$\pm 1.5$	$\pm 1.5$	$\pm 1.5$
${ m D}^{*}(2010)^{\pm}$	$\pm 0.5$	$\pm 0.5$	$\pm 0.5$	$\pm 0.5$	$\pm 0.5$	$\pm 0.5$
kinematics						
Background						
normalization	$\pm 0.5$	+0.9/-0.8	+1.9/-0.8	+1.4/-0.5	+0.8/-1.0	0.0/-0.6
$\vec{p}_{\rm T}^{\rm miss}$	+0.7/-0.9	+0.4/-1.2	+1.3/-0.3	+1.1/-1.0	0.0/-2.6	0.0/+1.5
Pileup	+2.0/-1.9	+0.4/-0.5	+2.9/-3.0	+2.0/-1.9	+4.6/-5.1	+2.7/-2.6
Secondary vertex	-1.1	+1.3	-1.2	-1.5	-2.7	-2.5
PDF	$\pm 1.2$	$\pm 1.3$	$\pm 0.9$	$\pm 1.4$	$\pm 1.5$	$\pm 1.7$
Fragmentation	+3.9/-3.2	+3.4/-1.8	+7.4/-5.2	+3.3/-3.0	+2.2/-1.2	+7.4/-5.7
MC statistics	+3.6/-3.3	+8.8/-7.5	+9.0/-11.9	+7.9/-6.8	+9.8/-14.1	+10.1/-8.5
Total	+7.5/-7.0	+10.7/-9.3	+13.2/-14.2	+10.1/-9.3	+12.7/-16.2	+13.8/-12.1

### Z+c at 13 TeV: uncertainties and inclusive cross section

Significant contributions from jet measurement, c tagging and theoretical uncertainties (QCD, PDF) in measured cross sections

The precision of the measurement is strongly affected by amount of MC samples (MC stat.) used in deriving the SV mass templates for fits to obtain scale factors

Channel	QCD	PDF	c tag/mistag	JER	JES	Pileup	Top Pair	$ID \setminus Iso$	$\mathcal{L}$	MC stat.
	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)
$\mu\mu, p_{\rm T}^{\rm c  jet}$	5.5	0.5	4.2	3.9	4.8	1.5	0.6	1.0	2.5	4.2
$\mu\mu, p_{\rm T}^Z$	1.9	0.5	4.2	1.1	3.9	1.6	0.8	1.0	2.5	3.1
ee, $p_{\rm T}^{\rm cjet}$	6.4	0.6	4.2	3.1	6.4	3.0	0.7	2.6	2.5	6.3
ee, $p_{\rm T}^Z$	2.6	0.5	4.1	1.1	4.8	1.8	0.6	2.6	2.5	3.8

Measured inclusive cross sections:

405.4 ± 5.6 (stat) ± 24.3 (syst) ± 3.7 (theo) pb