ATLAS Jet properties:
jet substructure, jet mass

Jim Lacey
on behalf of the ATLAS Collaboration
SM@LHC, Florence, Italy
April 21, 2015
Introduction

Jet substructure techniques and variables

Performance highlights

Physics analysis highlights

Summary and conclusions
At the LHC $\sqrt{s} \gg M_{EW}$ ⇒ boosted objects with high $p_T$

decay products merge into a single large-R jet

jet substructure tools developed to help understand complex hadronic final states at the LHC

substructure tools needed to identify boosted objects

techniques more important as we begin Run II and move towards higher luminosity
The wealth of variables studied in ATLAS is interesting in their own right and helps further QCD understanding. A large motivation is their use in grooming and tagging.
Jet substructure variables

- Jet mass
- Jet charge
- Jet width
- mass drop ($\mu_{12}$)
- momentum balance ($\sqrt{y_f}$)
- $k_t$-splitting scales ($\sqrt{d_{12}}$)
- N-subjettiness ($T_N$)
- Q-jets volatility
- track based variables ($R_{pT}, r_{trk}$)
- energy correlation functions
- angularities
- planar flow
- Jet pull

\[ \sqrt{d_{1,2}} = \min(p_{T,1}, p_{T,2}) \times \Delta R_{12} \]

(re)cluster jets using $k_t$ algorithm

$k_t$-distance between the two proto-jets of the last clustering step

used to identify jets containing W’s and reject QCD jets

ATLAS measurement with 2010 data:

jet grooming

- grooming removes soft contributions from the jet, providing pileup mitigation and improving resolution:
  - trimming,
  - filtering,
  - pruning

- Example:
  trim anti-kt R=1.0 jets using:
  \[
  R_{\text{sub}} = 0.3 \\
  \frac{p_T^i}{p_T^{\text{jet}}} < f_{\text{cut}} = 0.05
  \]
jet grooming: trimming

- improved signal-background separation
- improved mass resolution
- correctly repositions the Z boson mass peak

- after trimming Data/MC comparisons with 20 fb⁻¹ look very good!
boosted W boson tagging

- Data/MC comparison using W bosons from semi-leptonic tt-bar sample
  - High-purity (~98%) selection using *HepTopTagger*
  - Good agreement for relevant variables
- measure the optimal combination performance
  - similar performance across taggers, especially when “grooming+tagger” are combined (i.e. within mass window)

\[ \sqrt{d_{12}} \text{ anti-kt R=1.0 trimmed} \]

Note: systematics not included on curves
shower deconstruction

- anti-$k_t$ R=1.0 jet with C/A R=0.2 subjets of four momenta $\{p\}_N = \{p_1, \ldots, p_N\}$

- splitting probabilities assigned from the likelihood ratio:

$$\chi_{SD}(\{p\}_N) = \frac{P(\{p\}_N|S)}{P(\{p\}_N|B)} = \frac{\sum_{\text{histories}} P(\{p, c^j\}_N|S)}{\sum_{\text{histories}} P(\{p, c^j\}_N|B)}$$

one possible shower history (out of >1500)

ATLAS Preliminary Simulation
Z' → tt event, $m_{Z'} = 1.75$ TeV
$m_W = 77.3$ GeV, $m_{Wb} = 186.5$ GeV

- Anti-$k$, R = 1.0
- Calorimeter clusters

Subjets, C/A R = 0.2
- W boson
- $b$ jet
- Top radiation
- ISR

ATLAS-CONF-2014-003

J.Lacey, Carleton University

SM@LHC, Apr 21-24, Florence
shower deconstruction

- splitting probabilities assigned from the likelihood ratio:

\[
\chi_{SD}(\{p\}_N) = \frac{P(\{p\}_N \mid S)}{P(\{p\}_N \mid B)} = \frac{\sum_{\text{histories}} P(\{p, c^j\}_N \mid S)}{\sum_{\text{histories}} P(\{p, c^j\}_N \mid B)}
\]

- good data / MC agreement
- stable against pileup

\[<\mu>: \text{average number of interactions per bunch crossing}\]
Note: systematics not included

• previous ATLAS studies investigated a large range of taggers:
  ATLAS-CONF-2013-084

• shower deconstruction exhibits very good performance over efficiency range
- many ATLAS analyses utilize jet shapes in boosted object tagging
- measurement of substructure variables useful for tuning parton showers
data: $\sqrt{s} = 7$ TeV, $L = 4.6$ fb$^{-1}$

simulation samples:

- signal: Herwig+Jimmy (LO scaled to NLO prediction)
- background: Pythia6; additional variations for systematics

anti-$k_t$ R=0.6 jet used for W/Z reconstruction:

- jet $p_T > 320$ GeV
- $|\eta| < 1.9$
- $50$ GeV < $m_{\text{jet}}$ < 140 GeV

**study substructure grooming techniques using enriched W/Z sample**

**measure W/Z combined cross section $\times$ BR:**

$$
\sigma_{W+Z} = \sigma_W(p_T > 320 \text{ GeV}, |\eta| < 1.9) \times \mathcal{B}(W \rightarrow q\bar{q}') \\
+ \sigma_Z(p_T > 320 \text{ GeV}, |\eta| < 1.9) \times \mathcal{B}(Z \rightarrow q\bar{q})
$$
**W/Z vs. QCD jet discrimination**

- substructure method used to aid in W/Z-jets / QCD-jet discrimination
- jets boosted to their rest frame: \( p_{\text{rest}} = (m_{\text{jet}}, 0, 0, 0) \)
  - W/Z jet constituents look *dijet like*
  - QCD jet constituents *isotropic*
- **Jet Shapes used as input to likelihood discriminant**

**Jet Shape:** *sphericity* \( S \) is a measure of the distribution of \( p_T \) of the clusters within a jet w.r.t the jet axis

- transform to jet rest frame
  - 2 back-to-back subjets \( S=0 \)
  - isotropic \( S=1 \)

\[
L = -\ln \left[ \frac{\mathcal{L}_s(i)}{\mathcal{L}_s(i) + \mathcal{L}_b(i)} \right]
\]
**W/Z vs. QCD jet discrimination**

- Substructure method used to aid in W/Z-jets / QCD-jet discrimination.
- Jets boosted to their rest frame: \( p_{\text{rest}} = (m_{\text{jet}},0,0,0) \)
  - W/Z jet constituents look dijet like
  - QCD jet constituents isotropic

Jet Shapes used as input to likelihood discriminant

- Optimal cut on likelihood \((L>0.15)\) obtained by maximizing significance:
  - \( \varepsilon = 56\% \)
  - Rejection = 89\%
- After selection multiple candidate jets in 2.5\% of data.
**W/Z cross sections measurement**

- binned maximum likelihood fit to the jet mass distribution used to extract the W/Z signal
- cross section calculated from:
  \[ \sigma_{W+Z} = \frac{N_{W+Z}}{L \cdot \varepsilon} \]
- NLO QCD calculation:
  \[ \sigma_{W+Z} = 5.1 \pm 0.5 \text{ pb} \]
- measurement agrees within 2\(\sigma\)

\[ \sigma_{W+Z} = 8.5 \pm 0.8 \text{ (stat.)} \pm 1.5 \text{ (syst.) pb} \]

**ATLAS**

\[ \sqrt{s} = 7 \text{ TeV, 4.6 fb}^{-1} \]
\[ p_T > 320 \text{ GeV} \quad |\eta| < 1.9 \]
\[ L > 0.15 \]
**boosted top differential cross sections**

- ttbar production provides an experimental signature similar to many interesting BSM processes
- possible distortion the $p_T^{\text{top}}$ spectrum form BSM production mechanisms, particularly in boosted regimes
- also a dominant background to many new physics searches

- large production cross section at the LHC
- first measurement using large-R jets and substructure techniques
- extends kinematic reach into the TeV range
boosted top differential cross sections

leading anti-\textit{kt} $R = 0.4$ jet satisfying $\Delta R(l, \text{jet}\ R=0.4) < 1.5$ is $b$-tagged

- $e$ and $\mu$ overlap removal
- jet $p_T>25$ GeV & $|\eta|<2.5$
- JVF$<0.5$ if jet $p_T<50$ GeV

$E_T^{\text{miss}}>20$ GeV

$m_{W_T+E_T^{\text{miss}}}>60$ GeV

trigger on lepton

$$m_W = \sqrt{2p_T^{\text{lepton}}E_T^{\text{miss}}(1 - \cos\Delta\phi(l, E_T^{\text{miss}}))} \quad m_{T}$$

data: $\sqrt{8}$ TeV, $\mathcal{L} = 20.3$ fb$^{-1}$

trimmed anti-\textit{kt} $R = 1.0$ jet

- $p_T>300$ GeV
- $|\eta|<2.0$, $m>100$ GeV
- $\Delta R(\text{jet}\ R=1.0, \text{jet}\ R=0.4) > 1.5$
- $\Delta\phi(l, \text{jet}\ R=1.0) > 2.3$
- $\sqrt{d_{12}} > 40$ GeV

at least one anti-\textit{kt} $R=0.4$ jet with $\Delta R(\text{jet}\ R=1.0, \text{jet}\ R=0.4)<1.0$ is $b$-tagged
that are introduced by instabilities in the inversion procedure. The unfolding regularization parameter, the 

\[ t \]

expected fraction of these events in 

\[ t \]

subtracted from the observed number of events in each 

\[ t \]

the following steps.

\[ \text{total integrated luminosity of the data sample. The corrections that are applied to} \]

\[ \text{the} \]

\[ \text{third step corrects for detector resolution } \]

\[ \text{effects. A migration matrix is constructed to correlate} \]

\[ \text{particle- and parton-level} \]

\[ \text{accounting for the branching ratio of} \]

\[ \text{quark,} \]

\[ \text{bar} \]

sample. In the second step, the acceptance factors 

\[ \text{of} \]

\[ \text{events defined as background in Sec. 6 is subtracted by a multiplicative correction, quantifying the} \]

\[ \text{corrections described in the remaining steps are all extracted from the nominal P} \]

\[ \text{first, the post-selection non-} \]

\[ \text{the di} \]

\[ \text{is comprised of several steps, outlined} \]

\[ \text{distribution, the particle-level representation. It is calculated} \]

\[ \text{represents the probability for an event with} \]

\[ \text{ptcl!reco} \]

\[ \text{ptcl} \]

\[ \text{ptcl} \]

\[ \text{of} \]

\[ \text{reco!ptcl} \]

\[ \text{reco} \]

\[ \text{reco!ptcl} \]

\[ \text{N}_{\text{reco}} - N_{\text{reco,bgnd}} \]
measured cross-sections generally lower than LO and NLO MC predictions (after normalization to NNLO+NNLL QCD calculations)

discrepancy generally increases with the $p_T$

MC predictions overestimate the measured cross-sections in the highest-$p_T$ bin by up to 70%, depending on the MC generator

uncertainties ranging from 10-40%, increasing with $p_T$
development of jet substructure techniques has led to a substantial increase in our ability to probe hadronic final states and perform precision physics

techniques will play a essential role in analyses immediately in Run II and more-so as we move forward towards higher luminosities

- boosted object composition revealed through tagging
- grooming for resolution and pileup resistance

Considerable efforts in optimizing/validating these techniques in ATLAS Run I data

ATLAS has explored a wide variety of substructure variables, both in performance studies and physics measurements
ATLAS measurements involving jet properties and substructure have been steadily growing since the beginning of Run I highlights:

Performance:
- Shower deconstruction top-tagging: ATLAS-CONF-2014-003
- Jet Pull: ATLAS-CONF-2014-048
- Boosted W boson tagging: ATL-PHYS-PUB-2014-004

Standard Model:

Top:
- $t\bar{t}$bar cross section measurements (7 TeV): Phys. Rev. D 90 (2014), 072004
- Boosted top differential cross-sections: ATLAS-CONF-2014-057 (covered in this talk)

Top+Exotics:
- All hadronic $t\bar{t}$bar resonance (7 TeV): JHEP 01(2013) 116
- 1-lepton $t\bar{t}$bar resonance (8 TeV 20/fb): ATLAS-CONF-2015-009

Exotics:

SUSY:
- RPV 3 jets: JHEP 12 (2012) 086
- Zero lepton + multijet: JHEP 10 (2013) 130
backup
Jet mass
Jet charge
Jet width
mass drop ($\mu_{12}$)
momentum balance ($\sqrt{y_f}$)
$k_t$-splitting scales ($\sqrt{d_{12}}$)
N-subjettiness ($T_N$)
Q-jets volatility
track based variables ($R_{pT}, r_{trk}$)
energy correlation functions
angularities
planar flow
Jet pull

$\sqrt{d_{1,2}} = \min(p_{T,1}, p_{T,2}) \times \Delta R_{12}$
**boosted top: decay channel combination**

- build combined $l$+jets from e+jet and $\mu$+jet samples in data and MC:
  - data: logical “OR”
  - MC: proportions dictated by their efficiencies
  - data-driven bkg predictions derived in each channel, then combined and added to simulation predictions
  - combined sample used in deriving fiducial regions and unfolding factors
- method cross checked by unfolding each channel separately and comparing to combined result

---

ATLAS Preliminary

$20.3 \text{ fb}^{-1}$, $\sqrt{s} = 8 \text{ TeV}$, e+jets

**e+jets**

- Data
- $t\bar{t}$ Single lepton
- $t\bar{t}$ Dilepton
- Single top
- W+jets
- Multijet
- Z+jets
- Diboson

ATLAS Preliminary

$20.3 \text{ fb}^{-1}$, $\sqrt{s} = 8 \text{ TeV}$, $\mu$+jets

**$\mu$+jets**

- Data
- $t\bar{t}$ Single lepton
- $t\bar{t}$ Dilepton
- Single top
- W+jets
- Multijet
- Z+jets
- Diboson

---

J.Lacey, Carleton University

SM@LHC, Apr 21-24, Florence
boosted top: particle-level object definitions

- truth objects built from stable particles ($\tau > 0.2 \times 10^{-10}$ s) directly resulting from pp interactions or from subsequent decay of particles with shorter mean lifetimes

- leptons from hadron decays are discarded
  - keep $e$ or $\mu$ from $\tau$ decays only if $\tau$ was not a hadron decay product
  - electrons “dressed” with photons within $\Delta R<0.1$

- jets reconstructed using all stable particles except for leptons as selected above:
  - includes $\mu$’s and $\nu$’s from hadron decay, as well as charged $\pi$'s
  - all particle-level constituents considered in large-R jet trimming
  - $R=0.4$ jets with $p_T>25$ GeV and $|\eta|<2.5$ are considered b-tagged if one or more B-hadron(s) ($p_T>5$GeV) are “ghost associated” with the jet

- $E_T^{miss}$ is built from 4-vector addition of all $\nu$’s not from hadron decay

- particle-level corrections are derived from a ttbar sample where exactly one of the W-bosons decays leptonically

- cross section measured as a function of particle-level top jet: $p_{T,ptcl}$
parton-level corrections are derived from a ttbar sample where exactly one of the W’s decays leptonically

boosted top quark definition:

- top decays to a W that subsequently decays hadronically
- instance just preceding the decay
- after QCD radiation
- top quark pT ($p_{T,\text{parton}}$) > 300 GeV

extrapolation to full partonic phase space obtained by accounting for branching ratios of ttbar events to the lepton+jets channel
### boosted top: fiducial region definitions

<table>
<thead>
<tr>
<th>Cut</th>
<th>Detector-level</th>
<th>Particle-level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$e + \text{jets}$</td>
<td>$\mu + \text{jets}$</td>
</tr>
</tbody>
</table>
| **Leptons**          | $|z_0| < 2 \text{ mm}$  
                      | $I_{\text{mini}} < 0.05$  
                      | $|\eta| < 1.37 \text{ OR } 1.52 < |\eta| < 2.47$  
                      | $p_T > 25 \text{ GeV}$  | $|z_0| < 2 \text{ mm} \text{ & } |d_0/\sigma(d_0)| < 3.$  
                      | $I_{\text{mini}} < 0.05$;  
                      | $|\eta| < 2.5$  
                      | $p_T > 25 \text{ GeV}$ | $|\eta| < 2.5$  
                      | $p_T > 25 \text{ GeV}$  |
| **Anti-$k_t$ R=0.4 jets** | $p_T > 25 \text{ GeV}$  
                      | $|\eta| < 2.5$  
                      | JVF > 0.5 (if $p_T < 50 \text{ GeV}$) | $|\eta| < 2.5$  
                      | $p_T > 25 \text{ GeV}$  |
| **Overlap removal**  | if $\Delta R(e, \text{jet}_{R=0.4}) < 0.4$:  
                      | jet$'_{R=0.4} = \text{jet}_{R=0.4} - e$  
                      | if $\Delta R(e, \text{jet}_{R=0.4}) < 0.2$:  
                      | $e$ removed | if $\Delta R(\mu, \text{jet}'_{R=0.4}) < 0.04 + 10. \text{ GeV}/p_T(\mu)$:  
                      | $\mu$ removed | None |
| $E_T^{\text{miss}}, m_T^W$ | $E_T^{\text{miss}} > 20 \text{ GeV}$, $E_T^{\text{miss}} + m_T^W > 60 \text{ GeV}$ |                      |
| **Leptonic top**     | At least one anti-$k_t$, R = 0.4 jet satisfying $\Delta R(l, \text{jet}_{R=0.4}) < 1.5$ |                      |
| **Hadronic top**     | The leading trimmed anti-$k_t$ R = 1.0 jet  
                      | $p_T > 300 \text{ GeV}$, $m > 100 \text{ GeV}$, $\sqrt{d_{12}} > 40 \text{ GeV}$  
                      | $\Delta R(\text{jet}_{R=1.0}, \text{jet}_{R=0.4}) > 1.5$, $\Delta \phi(l, \text{jet}_{R=1.0}) > 2.3$ | At least one of:  
                      | 1) the leading anti-$k_t$, R = 0.4 jet satisfying $\Delta R(l, \text{jet}_{R=0.4}) < 1.5$ is $b$-tagged;  
                      | 2) at least one anti-$k_t$ R = 0.4 jet satisfying $\Delta R(\text{jet}_{R=1.0}, \text{jet}_{R=0.4}) < 1.0$ is $b$-tagged |                      |
| **B-tagging**        | At least one of:  
                      | 1) the leading anti-$k_t$, R = 0.4 jet satisfying $\Delta R(l, \text{jet}_{R=0.4}) < 1.5$ is $b$-tagged;  
                      | 2) at least one anti-$k_t$ R = 0.4 jet satisfying $\Delta R(\text{jet}_{R=1.0}, \text{jet}_{R=0.4}) < 1.0$ is $b$-tagged |                      |
statistical fluctuations when estimating systematic uncertainties. The typical expected fractional resolution of each bin, and that the width of each bin corresponds to at least one and a half times the root mean square. The distribution has been optimized such that systematic uncertainties are larger than statistical uncertainties in the region described in Sec. 9.2, in which characterizes the size of the expansion of the solution to the inversion problem, has been optimized by using various MC generators.

The migration matrix is the number of observed events in each bin, to reduce possible biases deriving from predictions of non $tt$ background subtraction, is shown in Fig. 3(a) for the $ptcl!reco$ spectrum, on a bin-by-bin basis, for the $ptcl$ basis, for the $j$th parton, for the $i$th binned distribution to the particle level is carried out by an unfolding to particle level.

The corrections described in the remaining steps are all extracted from the nominal P. The corrections that are applied, in the fiducial region described in Sec. 9.2, are applied, and the total integrated luminosity of the data sample. The corrections that are applied, where

$$L = 20.3 \, fb^{-1}, \Delta X^i: \text{bin width}$$

$$\frac{d\sigma_{tt\bar{t}}}{dp^i_{T,ptcl}} = \frac{N^i_{ptcl}}{\Delta X^i L} = \frac{1}{\Delta X^i L f^i_{ptcl!reco}} \sum_j (M^{-1})^{ptcl,i}_{reco,j} f^j_{reco!ptcl} (N^j_{reco} - N^j_{reco,bgnd})$$

dominant background: W+jets

- ALPGEN+PYTHIA sample normalized to inclusive W+jets NNLO cross section
- normalization and heavy flavour fractions obtained from data
- normalization obtained by exploiting expected $W^+$ $W^-$ charge asymmetry
- extract heavy flavour using b-jet multiplicity and b-tag/mistag efficiencies

QCD multijet backgrounds estimated from data

MC estimation of top, Z+jets, diboson, all sub-dominant backgrounds
$\mathcal{L} = 20.3$ fb$^{-1}$, $\Delta X^i$: bin width

\[
\frac{d\sigma_{t\bar{t}}}{dp_{T,ptcl}^i} = \frac{N^i_{ptcl}}{\Delta X^i \mathcal{L}} = \frac{1}{\Delta X^i \mathcal{L} f^i_{ptcl!reco}} \sum_j (M^{-1})^{ptcl,i}_{reco} f_{reco!ptcl}^j (N^i_{reco} - N^i_{reco,bgnd})
\]

- correct for tt events that pass reco level, but fail particle level selections

\[
f_{reco!ptcl} = \frac{N_{Pass}^{det}}{N_{Pass}^{ptcl}} + \frac{N_{Pass}^{det}}{N_{Pass}^{ptcl}}
\]
\[ \mathcal{L} = 20.3 \, \text{fb}^{-1}, \ \Delta X^i: \text{bin width} \]

\[ \frac{d\sigma_{t\bar{t}}}{dp^i_{T,\text{ptcl}}} = \frac{N^i_{\text{ptcl}}}{\Delta X^i \mathcal{L}} = \frac{1}{\Delta X^i \mathcal{L} f^i_{\text{ptcl!reco}}} \sum_j (M^{-1})^{{\text{ptcl,i}}}_{\text{reco,j}} f^j_{\text{reco!ptcl}} (N^j_{\text{reco}} - N^j_{\text{reco,bgnd}}) \]

- correct for detector resolution effects
- migration matrix inverted using an unfolding scheme based on Tikhonov regularization, implemented through the singular value decomposition (SVD) of the matrix
\[ \mathcal{L} = 20.3 \text{ fb}^{-1}, \Delta X^i: \text{bin width} \]

\[ \frac{d\sigma_{t\bar{t}}}{dp_{T,ptcl}^i} = \frac{N_{ptcl}^i}{\Delta X^i \mathcal{L}} = \frac{1}{\Delta X^i} \sum_j (M^{-1})_{ptcl,i}^{reco,j} f_{ptcl!reco}^i f_{reco!ptcl}^j (N_{reco}^j - N_{reco,bgnd}^j) \]

- **efficiency correction to restore events that pass particle level selections but fail at reco level**

\[ f_{ptcl!reco} = \frac{N_{Pass}^{det} + N_{Pass}^{ptcl}}{N_{Pass}^{ptcl}} \]
**boosted top: unfolding to parton level**

**BR=0.438:** branching ratio for exactly one W to decay to e, μ, τ

**ΔX^k:** bin width, $\mathcal{L} = 20.3$ fb$^{-1}$

\[
\frac{d\sigma_{\bar{t}t}}{dp_{T,\text{parton}}^k} = \frac{N^k_{\text{parton}}}{\text{BR}\Delta X^k \mathcal{L}} = \frac{1}{\text{BR}\Delta X^k \mathcal{L} f_{\text{parton!ptcl}}^k} \sum_i (M^{-1})_{\text{parton},i} f_{\text{ptcl!parton}}^i N_{\text{ptcl}}^i
\]

- similarly, unfold particle-level distributions back to parton-level
  - hadronization effects relating $p_{T,\text{ptcl}}^k$ to $p_{T,\text{parton}}^k$ are corrected for using the same matrix technique as done in correcting for detector effects
- cross-check two step unfolding procedure by comparing to results obtained by directly unfolding detector-level distributions to parton level

\[
\frac{d\sigma_{\bar{t}t}}{dp_{T,\text{parton}}^k} = \frac{N^k_{\text{parton}}}{\text{BR}\Delta X^k \mathcal{L}} = \frac{1}{\text{BR}\Delta X^k \mathcal{L} f_{\text{parton!reco}}^k} \sum_j (M^{-1})_{\text{reco},j} f_{\text{reco!parton}}^j (N_{\text{reco}}^j - N_{\text{reco,bgnd}}^j),
\]

- consistent results obtained
- primary results from two step since it provides a more detailed description of uncertainty evolution
differential boosted top: uncertainties

- particle-level cross-section dominant uncertainty: large-R jet energy scale
  - particularly due to two components:
    1. $\gamma$+jet balance uncertainty at high-$p_T$
    2. topology uncertainty at low-$p_T$
- parton-level experimental uncertainties comparable in size
- parton-level total uncertainty significantly larger due to signal modelling
measured differential cross-sections are compared to the predictions of Powheg+Pythia with and without the electroweak corrections applied

- electroweak corrections lead to a softer $p_T$ spectrum
- effect is too small to bring the prediction in agreement with data
boosted top quark: PDF sets

compared to predictions using either the HERAPDF or CT10 PDF sets, and two different values of the Powheg+Pythia $h_{damp}$ parameter ($h_{damp}$ effectively regulates the high-$p_T$ radiation)

- EW corrections not applied

larger improvement obtained using HERAPDF set instead of CT10, which improves the data/MC agreement by up to ~20%
**W/Z signal pdfs:**
- Breit-Wigner ⊗ Gaussian
- peak position, W/Z relative fractions and widths obtained from fit to MC
- free parameters: W/Z combined rate

**QCD background pdfs:**
- Exp. + Exp. + Sigmoid
- shoulder due to likelihood + kinematic selection
- well modelled in MC for various selections

---

**Figure 2:** Jet mass distribution in simulation for (a) hadronically decaying W and Z bosons, (b) QCD jet background and (c) hadronically decaying W bosons from top-quark pair events. The QCD jet background and W/Z signal distributions are fitted with the pdfs described in the text. The fit results are shown as solid lines. For the signal, the contributions from W (dotted line) and Z (dashed line) jets are shown. The distributions are normalized to the number of events expected in the data and the uncertainties are statistical only.

W/Z CX: modelling the jet mass distribution
use enriched W/Z sample to study performance of grooming techniques:

- pruning \((R_{\text{cut}} = 0.3, z_{\text{cut}} = 0.02)\)
- trimming \((R_{\text{sub}} = 0.2, f_{\text{cut}} = 0.03)\)
- jet “area based” subtraction

resulting cross sections compatible with default within statistical uncertainties
**W/Z cross sections: pileup dependence**

![Figure 5: The jet mass distributions for low-pileup (N_{vtx} < 5) and high-pileup (N_{vtx} > 10) conditions in MC (PYTHIA 8.153 for background plus HERWIG 6.520 for signal) and data, where N_{vtx} is defined as the number of reconstructed collision vertices in an event. The cases without grooming, after applying pruning, trimming and area subtraction are shown. Here, N_{vtx} is defined as the number of reconstructed primary vertices in an event. The uncertainties are statistical only.](image)

**8. Conclusion**

This paper presents a measurement of the production cross-section of a hadronically decaying boosted W or Z boson with transverse momentum p_T > 320 GeV and |\eta| < 1.9 in pp collisions at a centre-of-mass energy of 7 TeV with the ATLAS detector at the LHC. The measurement is performed by reconstructing boosted W and Z bosons in single jets. The reconstructed jet mass is used to identify the W and Z bosons and a jet substructure method based on energy cluster formation in the jet centre-of-mass frame is used to suppress the large multi-jet background. The measured cross-section is: \( \sigma_{W+Z} = 8.5 \pm 0.8 \text{ (stat.)} \pm 1.5 \text{ (syst.) pb} \).

The measured value is found to be in agreement with the theoretical prediction for the same kinematic range of \( \sigma_{W+Z} = 5.1 \pm 0.5 \text{ pb} \), obtained from the NLO QCD calculation, within 2 \( \sigma \). The total uncertainty in the measured cross-section is of the same order of magnitude as the uncertainties in measurements performed with leptonic decay channels for a similar kinematic region [2, 3, 4].
jet substructure techniques

- split/filter jets to identify hard constituents
- Ex. mass drop filtering (BDRS) used in boosted W boson identification
  - Cambridge-Aachen R=1.2 jet
  - $\mu_{1,2} > 2/3$
  - $\sqrt{y_f} > 0.3$
**Jet substructure variables**

1. **Jet mass**
2. **Jet charge**
3. **Jet width**
4. **Jet mass drop**
5. **Momentum balance**
6. **N-subjettiness**
7. **Q-jets volatility**

**Track based variables**

1. **Energy correlation functions**
2. **Angularities**
3. **Planar flow**
4. **Jet pull**

**Pull vector:**

\[ v_p(J) = \sum_{i \in J} \frac{p_T^i |\vec{r}_i|}{p_T^j} \vec{r}_i \]

- \( \theta_P \): angle the pull vector for \( J_1 \) makes with respect to the vector connecting jets \( J_1 \) and \( J_2 \)
- Sensitive to the underlying colour flow
- Small \( \theta_P \leftrightarrow \) colour connected \( J_1 \) and \( J_2 \)

**Jet constituents are calorimeter clusters, ghost associated tracks, or stable mc particles**

Legend:
- Pull Vector \( v_p(J_1) \)
- \( \theta_P \) jet pull angle
- Constituent of \( J_1 \) (size weighted by \( p_T \))

\( \Delta \phi = \phi - \phi_{J_1} \)

\( \Delta y = y - y_{J_1} \)

\( r_i = (\Delta y_i, \Delta \phi_i) \)
Jet pull

*Jet substructure variables*

- Jet pull
- Planar flow
- Angularities
- Track based variables
- Q-jets volatility
- \( \tau \)
- N-subjettiness (\( k \))
- Momentum balance
- Mass drop
- Jet width
- Jet charge
- Jet mass

**Introduction**

The bulk of this note uses a MC truth-object selection with signal separation.

- The two leading (in \( p_T \)) and at least two non-\( b \)-tagged jets, with \( |\Delta \eta| < 2.5 \) and \( \Delta R_j < 0.4 \), are required to have an invariant dijet mass close to 173 GeV. The muon is used to trigger and a cut on the missing energy from the \( b \)-tagged jets.

- The jets constituting the dijet pair closest to the lepton are required to have an invariant dijet mass close to 173 GeV. The muon is used to trigger and a cut on the missing energy from the \( b \)-tagged jets.

**MC@NLO**

Simulated events produced by Monte Carlo generators are processed with a full ATLAS detector simulation based on the ATLAS hardware description model. The minimum-bias events simulated with Geant4 and of the remaining non-\( b \)-hadrons, with at least two non-\( b \)-tagged jets, two are required to have an invariant dijet mass close to 173 GeV. The two leading (in \( p_T \)) and at least two non-\( b \)-tagged jets, with \( |\Delta \eta| < 2.5 \) and \( \Delta R_j < 0.4 \), are required to have an invariant dijet mass close to 173 GeV. The muon is used to trigger and a cut on the missing energy from the \( b \)-tagged jets.

**Conclusion**

- The bulk of this note uses a MC truth-object selection with signal separation.
- Dibosons are modelled with fortran \([17,18]\) with the NLO parton density function (PDF) set interfaced with \( \text{MC@NLO} \) and the version is 6.520.2 and the tune is the AUET2 tune \([28]\).
- The jets constituting the dijet pair closest to the lepton are required to have an invariant dijet mass close to 173 GeV. The muon is used to trigger and a cut on the missing energy from the \( b \)-tagged jets.

- The jets constituting the dijet pair closest to the lepton are required to have an invariant dijet mass close to 173 GeV. The muon is used to trigger and a cut on the missing energy from the \( b \)-tagged jets.

**ATLAS-CONF-2014-048**
ATLAS implements jet “area-based” pileup subtraction technique

- significantly reduces pileup dependence
- reduces impact on the jet resolution

\[
p_T^{\text{corr}} = p_T - \rho A - \alpha (N_{PV}-1) - \beta \langle \mu \rangle
\]
systematic uncertainties

large-R jet pT uncertainties
- estimated using in-situ techniques balancing a well measured reference object and a jet
- additional systematics to account for sample topology

uncertainties from track jets
- use jets built from tracks as a precise reference measurement
- used in ATLAS for mass, N-subjettiness and splitting scale uncertainties
**boosted W boson tagging**

- investigate various jet algorithms in selections of $p_T$ ranges
- **Signal**: KK Graviton $\rightarrow$ WW $\rightarrow$ lvqq
- **Background**: W+jets (Sherpa)

Jet mass used as main discriminant:
- 68% signal window (look for W peak)
- optimum variable+algorithm combinations investigated
tracks are fundamental constituents of jets and provide complementary information to topoclusters

currently exploited in ATLAS:

- pileup suppression:
  - JVF, Track Jets, ect.,
- substructure and tagging:
  - jet charge, quark/gluon discrimination

combine tracks and topoclusters to suppress / reject pileup:

- Jet Vertex Tagger (JVT)
jet-vertex tagger variables: $R_{pT}$ & corrJVF

- include tracks to build pileup jet tagging observables: $R_{pT}$

\[
R_{pT} = \frac{\sum_k p_T^{trk_k}(PV_0)}{p_T^{jet}}
\]

- corrJVF = \frac{\sum_k p_T^{trk_k}(PV_0)}{\sum_l p_T^{trk_l}(PV_0) + \frac{\sum_{n>1} \sum_l p_T^{trk_l}(PV_n)}{(k-n_{trk})^{PU}}}

- include track information and correct JVF for pileup
tagging jets with the jet-vertex tagger (JVT)

- jet-vertex tagger combines $R_{pT}$ and corrJVF to form a 2D likelihood

$$\Delta R(\text{HS jet, truth jet with } p_T > 10 \text{ GeV}) < 0.3$$
$$\Delta R(\text{PU jet, truth jet with } p_T > 4 \text{ GeV}) > 0.6$$

ATLAS Simulation Preliminary

ATLAS Preliminary

$\tilde{s} = 8 \text{ TeV}, L = 20.3 \text{ fb}^{-1}$
Sherpa $Z \rightarrow \mu\mu$
Anti-$k_t$ LCW+JES $R=0.4$
$20 < p_T < 50 \text{ GeV}$, $|\eta| < 2.4$
JVT > 0.4

JVT > 0.4
physics in Run II will require substructure:
- any objects with $p_T > 500$ GeV going to require substructure
- pileup is a major challenge current approaches shown to be performant at high luminosity

**ATLAS high pileup studies for top tagging**

**Jet grooming and pileup subtraction**

### ATLAS Simulation Preliminary

- **anti-$k_t$ LCW jets with $R=1.0$**
  - No jet grooming, no pileup correction
  - Trimmed, pileup corrected

- **$\sqrt{s} = 14$ TeV, 25 ns bunch spacing**
  - $|\eta_{\text{jet}}| < 1.2$, $500 < p_T^\text{jet} < 1000$ GeV
  - $m_Z' = 2$ TeV

- **Pythia8 $Z' \rightarrow t\bar{t}$**
  - $\mu = 0$
  - $\mu = 80$
  - $\mu = 140$
  - $\mu = 200$
  - $\mu = 300$