Measurement of jet substructure in boosted $t\bar{t}$ events with the ATLAS detector using 140 fb^{-1} of 13 TeV pp collisions (ATLAS-CONF-2023-027)

Christopher Andrew Garner *on behalf of the ATLAS Collaboration*

UNIVERSITY OF TORONTO

Introduction and Motivation

The measurement of large-R jet substructure in boosted $t\bar{t}$ events is motivated by:

- **Experiment:** Large uncertainties in analyses selecting signal from background using top-taggers due to poor modelling of jet substructure by current MC generators
- **Theory:** Fixed-order analytic description of jet substructure is challenging
- **Improved Measurement using Tracks:** Uncertainties in calorimeter-cluster reconstruction and calibration have limited the precision of previous substructure measurements in boosted $t\bar{t}$ events
- **New Decay Channel:** Opportunity for the first boosted $t\bar{t}$ substructure measurement in the all-hadronic decay channel

Measurement Strategy

Provide unfolded data to tune MC generators and test fixed-order analytic predictions by measuring large-R jet substructure in boosted $t\bar{t}$ events [1].

Analysis is performed in two channels:

- **ℓ + Jets:** One top-quark decays semileptonically while the other top-quark decays hadronically
- **All-Hadronic:** Both top-quarks decay hadronically

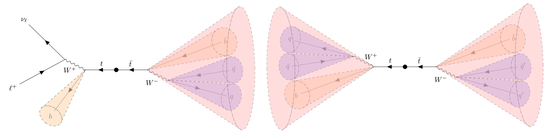

Figure 1. Event topology of $\ell + jets$ (**left**) and all-hadronic (**right**) channels.

Event Selection

ℓ+Jets Selection:

- 1 e/μ with $p_{\mathrm{T}} > 27$ GeV
- $E_{\mathrm{T}}^{\mathrm{miss}} > 20$ GeV
- $E_{\mathrm{T}}^{\mathrm{miss}} + m_{\mathrm{T}}^{W} > 60$ GeV
- ≥ 1 reclustered (RC) $R = 1.0$ jets with $p_{\mathrm{T}} > 350$ GeV far from e/μ
  - 122.5 GeV $< m_{\mathrm{jet}} <$ 222.5 GeV
  - Reclustered into at least 2 anti-k_t $R = 0.4$ jets
- ≥ 1 b-tagged R=0.4 jet close to e/μ
- $m_{\ell,b} < 120$ GeV

All-Hadronic Selection:

- 0 e/μ with $p_{\mathrm{T}} > 25$ GeV
- ≥ 1 anti-k_t $R = 1.0$ jets with $p_{\mathrm{T}} > 500$ GeV
  - 122.5 GeV $< m_{\mathrm{jet}} <$ 222.5 GeV
  - Ghost-matched to b-tagged variable-radius track jet
- ≥ 2 anti-k_t $R = 1.0$ jets with $p_{\mathrm{T}} > 350$ GeV
  - 122.5 GeV $< m_{\mathrm{jet}} <$ 222.5 GeV
  - Ghost-matched to b-tagged variable-radius track jet
- DNN top-tag on non-probe large-R jet

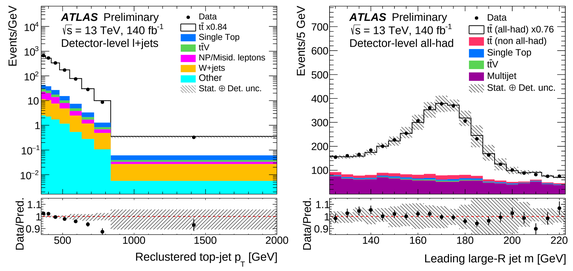

Figure 2. Detector-level RC large-R jet p_{T} and large-R jet mass distributions in the (**left**) l+jets and (**right**) all-hadronic channels, respectively. The predicted $t\bar{t}$ signal has been normalized so that the total predicted yield agrees with the observed number of events.

The transverse mass is $m_{\mathrm{T}}^{W} = \sqrt{2p_{\mathrm{T}}^{\ell}\left(1 - \cos\Delta\phi\left(p_{\mathrm{T}}^{\ell}, \right)\right)}$.

Table 1. Detector-level event yields for data, simulated $t\bar{t}$ signal, and background processes. Uncertainties affecting the detector and background modeling are included. Signal purity is defined by the number of data events minus background events divided by the total sample size.

Category	Event yields ℓ+jets selection	Number of large-R jets all-hadronic selection
Data	83069	30525
Predictions	97200±3700	36700±1400
$t\bar{t}$ (ℓ+jets)	90600±3400	1630±140
$t\bar{t}$ (all-hadronic)	–	26000±1400
Multijet	–	8100±300
Single-top quark	2200±300	720±70
NP/Misid. leptons	1500±600	–
W+jets	1500±700	–
$t\bar{t}V$ ($t\bar{t}Z$ + $t\bar{t}W$ + $t\bar{t}H$)	920±120	310±40
Other	400±200	–
Data/Predictions	0.85±0.03	0.83±0.03
(Data-Background)/Signal	0.84±0.03	0.76±0.05
Signal purity	0.92±0.01	0.65±0.01

Substructure Variables

Criteria for selected substructure variables:

- Variables not well described by the nominal Powheg+Pythia 8 prediction
- Variables sensitive to the differences among alternative NLO+PS predictions
- Variables used in recent tagging algorithms within ATLAS
- Low (anti-)correlation between selected variables

Substructure reconstruction is performed using tracks, allowing for a resolution improvement of 50%, while also reducing the uncertainties with respect to previous results from calorimeter clusters.

Generalized Angularities [2]:

$$\lambda_{\beta}^{\kappa} = \sum_{i \in J} z_i^{\kappa} \left(\frac{\Delta R(i,\hat{n})}{R}\right)^{\beta} \qquad LHA = \lambda_{0.5}^{1} \qquad p_{\mathrm{T}}^{\mathrm{d}} = \lambda_{0}^{2}$$

$$p_{\mathrm{T}}^{\mathrm{d},*} = \sqrt{\left(p_{\mathrm{T}}^{\mathrm{d}} - \frac{1}{N}\right)\frac{N}{N-1}}$$

Variables measured: LHA, and $p_{\mathrm{T}}^{\mathrm{d},*}$

Energy Correlation Functions [3, 4]:

$$\mathrm{ECF}(N) = \sum_{i_1<i_2<\ldots<i_N \in J}\left(\prod_{a=1}^{N} p_{\mathrm{T},i_a}\right)\left(\prod_{b=1}^{N-1}\prod_{c=b+1}^{N}(i_b, i_c)\right)$$

$$ECF2 = \mathrm{ECF}(2)/\mathrm{ECF}(1)^2 \quad C_3 = \mathrm{ECF}(4)\,\mathrm{ECF}(2)/\mathrm{ECF}(3)^2$$

$$D_2 = \mathrm{ECF}(3)\,\mathrm{ECF}(1)^3/\mathrm{ECF}(2)^3$$

Variables measured: $ECF2$, C_3, and D_2

N-Subjettiness [5]:

$$\tau_N = \frac{1}{d_0}\sum_k p_{\mathrm{T},k} \min\left\{\Delta R_{1,k}, \Delta R_{2,k}, \cdots, \Delta R_{N,k}\right\}$$

$$d_0 = \sum_k p_{\mathrm{T},k} R_0$$

$$\tau_{32} = \tau_3/\tau_2 \qquad \tau_{21} = \tau_2/\tau_1$$

Variables measured: τ_{32}, τ_{21}, and τ_3

Reconstructed Jet Substructure

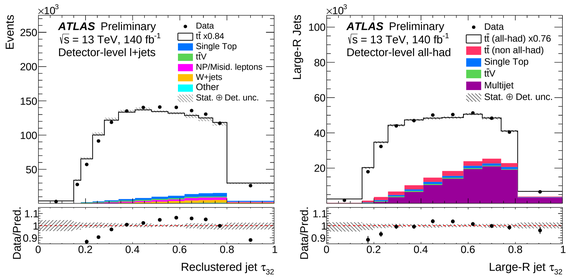

Figure 3. Detector-level τ_{32} distributions in the (**left**) l+jets and (**right**) all-hadronic channel. The predicted $t\bar{t}$ signal has been normalized so that the total predicted yield agrees with the observed number of events.

Unfolding Procedure

The effects arising from detector and selection acceptance, resolution, and efficiency on the observed distributions are corrected using an unfolding technique.

$$\frac{\mathrm{d}\sigma_j}{\mathrm{d}X_j} = \frac{1}{\int \mathcal{L}\mathrm{d}t \cdot \Delta X_j}\frac{1}{f_j^{\mathrm{eff}}}\sum_i M_{ij}^{-1} f_{\mathrm{acc}}^{i}(D^i - B^i),$$

Unfolding is performed using an **iterative Bayesian method** using **six iterations** as a compromise between minimizing the measurement uncertainty and the measurement bias.

Single-Differential Results

Single-differential cross-sections are compared with several NLOME+PS predictions of the Standard Model (SM), including several tuning variations of the nominal Powheg+Pythia 8 prediction.

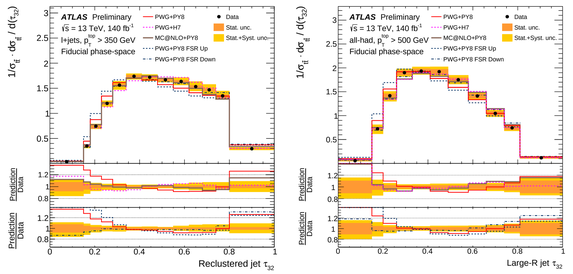

Figure 4. Particle-level single-differential cross-sections for τ_{32} in the ℓ+jets (**left**) and all-hadronic (**right**) channels. The distributions have each been normalized to unity.

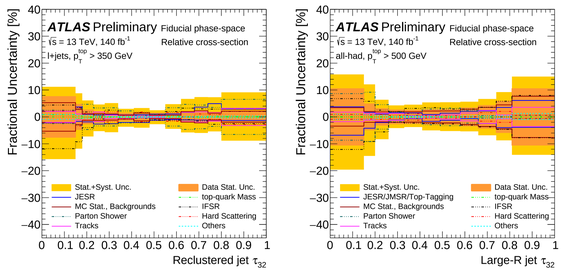

Figure 5. Fractional uncertainties on the particle-level differential $t\bar{t}$ production cross-section as a function of τ_{32} in the ℓ+jets (**left**) and all-hadronic (**right**) channels.

Double-Differential Results

Double-differential cross-sections are provided for τ_{32} and D_2 as a function of m^{top} and $p_{\mathrm{T}}^{\mathrm{top}}$. Correlations of τ_{32} and D_2 with m^{top} and $p_{\mathrm{T}}^{\mathrm{top}}$ are of particular interest due to their roll as leading discriminants in multi-variable taggers used to identify W and top-quark initiated jets.

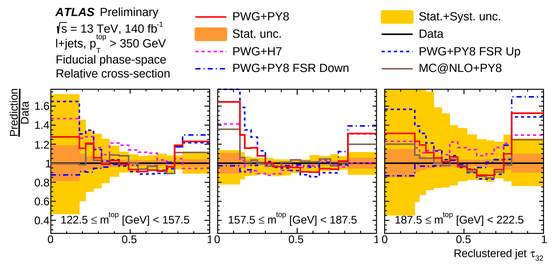

Figure 6. The ratios between the predictions and the data of the particle-level double-differential cross-section as a function of τ_{32} and of jet mass in the the l+jets channel.

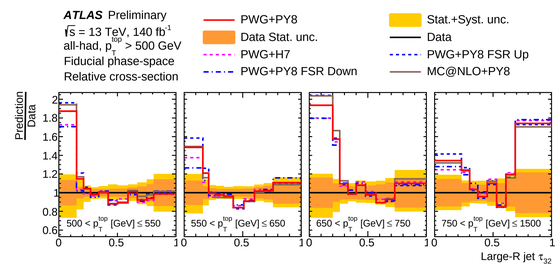

Figure 7. The ratios between the predictions and the data of the particle-level double-differential cross-section as a function of τ_{32} and of jet p_{T} in the the all-hadronic channel.

Data-MC Comparisons

Agreement between data and the SM predictions are quantified by a χ^2 test-statistic.

$$\chi^2 = V^T \times Cov^{-1} \times V$$

The covariance matrix Cov incorporates:

- Statistical uncertainties
- Bin-to-bin correlations induced by systematic uncertainties
- Regularization effects from unfolding the data

Table 2. χ^2 and p-values to quantify the agreement between the relative unfolded spectra in the ℓ+jets channel and several NLOME+PS predictions.

Observable	PWG+PY8		PWG+H7		MC@NLO+PY8		PWG+PY8(FSR Up)		PWG+PY8(FSR Down)	
	χ^2/NDF	p-value	χ^2/NDF	p-value	χ^2/NDF	p-value	χ^2/NDF	p-value	χ^2/NDF	p-value
τ_{32}	53/12	<0.01	20/12	0.08	15/12	0.24	164/12	<0.01	39/12	<0.01
τ_{21}	14/14	0.44	8/14	0.91	16/14	0.31	41/14	<0.01	7/14	0.92
τ_3	37/11	<0.01	41/11	<0.01	13/11	0.30	133/11	<0.01	23/11	0.02
$ECF2$	24/18	0.16	13/18	0.78	15/18	0.68	30/18	0.04	24/18	0.15
D_2	19/16	0.27	17/16	0.40	20/16	0.20	33/16	<0.01	18/16	0.35
C_3	14/14	0.47	6/14	0.97	4/14	1.00	42/14	<0.01	3/14	1.00
$p_{\mathrm{T}}^{\mathrm{d},*}$	27/12	<0.01	12/12	0.48	11/12	0.52	55/12	<0.01	25/12	0.01
LHA	14/17	0.69	10/17	0.91	21/17	0.22	13/17	0.75	19/17	0.34
D_2 vs m^{top}	58/42	0.05	62/42	0.03	59/42	0.04	114/42	<0.01	45/42	0.33
D_2 vs $p_{\mathrm{T}}^{\mathrm{top}}$	70/56	0.10	65/56	0.20	70/56	0.11	105/56	<0.01	92/56	<0.01
τ_{32} vs m^{top}	152/42	<0.01	75/42	<0.01	58/42	0.06	411/42	<0.01	76/42	<0.01
τ_{32} vs $p_{\mathrm{T}}^{\mathrm{top}}$	148/50	<0.01	101/50	<0.01	54/50	0.31	357/50	<0.01	112/50	<0.01

A RIVET routine has been written and validated for the ℓ+jets channel. A similar routine will be provided for the all-hadronic measurement.

Conclusions

- Most precise measurement of large-radius jet substructure in the boosted $t\bar{t}$ events by ATLAS
- First boosted $t\bar{t}$ substructure measurement in all-hadronic channel
- Herwig 7 parton-shower and hadronization model is generally a better match to the observed distributions
- Alternative matrix-element predictions employing aMC@NLO+Pythia 8 are in better agreement with the observed distributions in both channels
- The predictions produced with increased (decreased) FSR are universally poorer (better) descriptions of the data (favors increased value of α_{FSR})
- Agreement between the measurements and the predictions is better in the all-hadronic channel

References

[1] Measurement of jet substructure in boosted $t\bar{t}$ events with the ATLAS detector using 140 fb^{-1} of 13 TeV pp collisions. Technical report, CERN, Geneva, 2023. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2023-027.

[2] Andrew J. Larkoski, Jesse Thaler, and Wouter J. Waalewijn. Gaining (mutual) information about quark/gluon discrimination. *JHEP*, 2014(11), Nov 2014.

[3] Andrew J. Larkoski, Gavin P. Salam, and Jesse Thaler. Energy correlation functions for jet substructure. *JHEP*, 2013(6), Jun 2013.

[4] Andrew J. Larkoski, Ian Moult, and Duff Neill. Analytic boosted boson discrimination. *JHEP*, 2016(5), May 2016.

[5] Jesse Thaler and Ken Van Tilburg. Identifying boosted objects with n-subjettiness. *JHEP*, 2011(3), Mar 2011.