Jet Substructure in Boosted $t\bar{t}$ with ATLAS

Adam Rennie on behalf of the analysis team

6th April, 2022

IOP HEPP & APP 2022, RAL

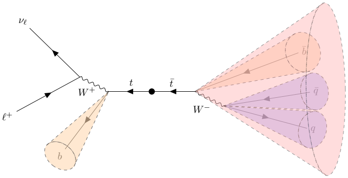
Measurement of large-R jet substructure in boosted $t\bar{t}$ events

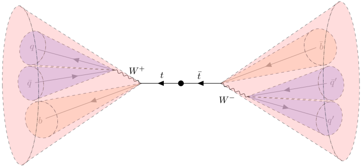
Craig Buttar, Federica Fabbri, **Adam Rennie**, Dwayne Spiteri
University of Glasgow

Christopher Garner, Pekka Sinervo
University of Toronto

What is jet substructure?

- Although the kinematics of a jet may be of interest, the distribution of a jet's constituents also contain much information
- In the case of a boosted top quark decaying hadronically, its decay products will be highly collimated and can be reconstructed as a single large-R jet
- This jet is expected to have a distinct *three-pronged* structure corresponding to the small-R jets from the decay products





Why measure jet substructure?

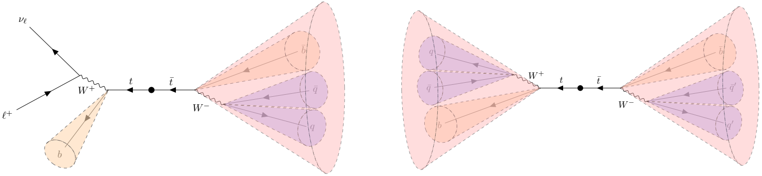
- The substructure of a jet provides valuable information on its origin
- Substructure is used extensively for *tagging*, which is ubiquitous in experimental HEP analysis
- There is significant mismodelling of substructure in many cases
- Many searches for new physics expect these sorts of signatures, with highly-boosted massive objects
- Can serve as a probe for QCD calculations

How does one measure jet substructure?

- There are many substructure observables which have been developed to probe various aspects of a jet's underlying substructure.
- These look at the constituents which make up a jet in order to characterise the radiation patterns within.
- Some basic examples would be the number of subjets or the number of constituents within a jet.

Analysis overview

- Measurement of jet substructure observables in boosted $t\bar{t}$ using ATLAS' full Run-2 dataset
- Both the lepton + jets and all-hadronic $t\bar{t}$ decay topologies are measured
- Results will be unfolded back to particle-level
- Results will be compared against a variety of MC predictions as well as variations on the nominal prediction



Event selection

Complete selection details in backup, with some important highlights:

Lepton + jets

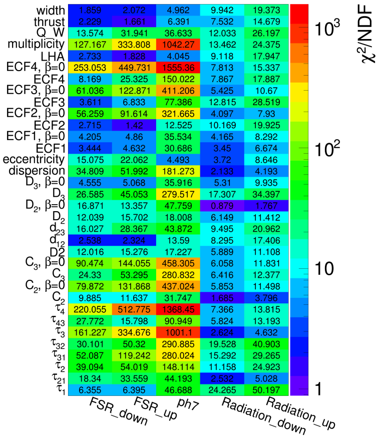
- $R = 1.0$ large-R jet reclustered from $R = 0.4$ small-R jets with a $p_\mathrm{T} > 350$ GeV
- Candidate top-jet to have 110 GeV $< m <$ 230 GeV
- One lepton with $p_\mathrm{T} > 27$ GeV
- $m_{\ell b} < 120$ GeV

All-hadronic

- Two $R = 1.0$ large-R topocluster jets with $p_\mathrm{T} > 350$ GeV, one with $p_\mathrm{T} > 500$ GeV
- b-tagged jet required inside both large-R jets
- DNN top-tag required on non-probe large-R jet

Selecting substructure observables

- A long list of possible substructure observables was drawn up, with criteria for trimming that into a final set to be measured:
    - Described poorly by the nominal prediction
    - Sensitive to variations in NLO+PS predictions
    - Useful in recent tagging algorithms
    - Good resolution
    - Minimally correlated with other observables



	FSR_down	FSR_up	ph7	Radiation_down	Radiation_up
width	1.859	2.072	4.962	9.942	19.373
thrust	2.229	1.661	6.391	7.532	14.679
Q_W	13.574	31.941	36.633	12.033	26.197
multiplicity	127.167	333.808	1042.27	13.462	24.375
LHA	2.733	1.828	4.045	9.118	17.947
ECF4, β=0	253.053	449.731	1555.36	7.813	15.337
ECF4	8.169	25.325	150.022	7.867	17.887
ECF3, β=0	61.036	122.871	411.206	5.425	10.67
ECF3	3.611	6.833	77.386	12.815	28.519
ECF2, β=0	56.259	91.614	321.665	4.097	7.93
ECF2	2.715	1.42	12.525	10.169	19.925
ECF1, β=0	4.205	4.86	35.534	4.165	8.292
ECF1	3.444	4.632	30.686	3.45	6.674
eccentricity	15.075	22.062	4.493	3.72	8.646
dispersion	34.809	51.992	181.273	2.133	4.193
D_3, β=0	4.555	5.068	35.916	5.31	9.935
D_3	26.585	45.053	279.517	17.307	34.397
D_2, β=0	16.871	13.357	47.759	0.879	1.767
D_2	12.039	15.702	18.008	6.149	11.412
d_{23}	16.027	28.367	43.872	9.495	20.962
d_{12}	2.538	2.324	13.59	8.295	17.406
D2	12.016	15.276	17.227	5.889	11.108
C_3, β=0	90.474	144.055	458.305	6.058	11.831
C_3	24.33	53.295	280.832	6.416	12.377
C_2, β=0	79.872	131.868	437.024	5.853	11.498
C_2	9.885	11.637	31.747	1.685	3.796
τ_4	220.055	512.775	1368.45	7.366	13.815
τ_{43}	27.772	15.798	90.949	5.824	13.193
τ_3	161.227	334.676	1001.1	2.624	4.632
τ_{32}	30.101	50.32	290.885	19.528	40.903
τ_{31}	52.087	119.242	280.024	15.292	29.265
τ_2	39.094	54.019	148.114	11.158	24.923
τ_{21}	18.34	33.559	44.193	2.532	5.028
τ_1	6.355	6.395	46.688	24.265	50.197

ATLAS Work in progress

Observables to be measured

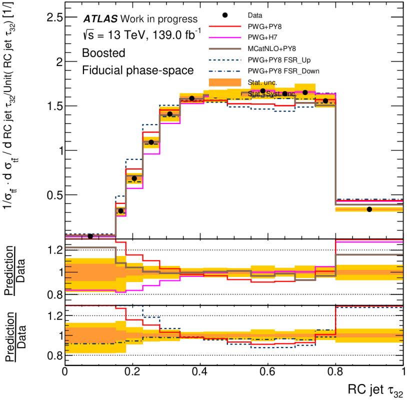
Observable	Description
N-subjettiness	Probes N-pronged structure. Useful in tagging, sensitive to variations in MC prediction. **Measured**: τ_3, τ_{32}, τ_{21}
Energy Correlation Functions (ECF)	Probes N-pronged structure. Useful in tagging, sensitive to variations in MC prediction. **Measured**: ECF2, D_2, C_3.
Generalised Angularities	Can be configured to probe different aspects of a jet's structure. Sensitive to variations in MC prediction. **Measured**: Les Houches Angularity, p_{T} dispersion.

Unfolding strategy

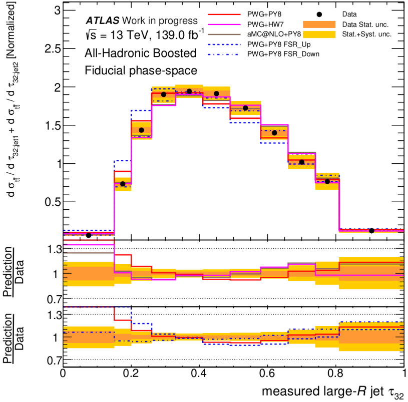
- Distributions unfolded to particle-level using Iterative Bayesian Unfolding (IBU)
- The number of iterations of the unfolding mechanism is optimised for each observable - between 4 and 10 iterations
- The procedure is the validated by:
    - Closure test: verifying that the unfolding can recover the underlying physics against statistical fluctuations
    - Pull test: checking for bias in measured bin values at particle-level and that errors are estimated correctly
    - Stress test: evaluate bias from regularization technique

Particle-level results

τ_{32} exhibits sensitivity to variations in the MC predictions as well as variations of the FSR



Lepton + jets



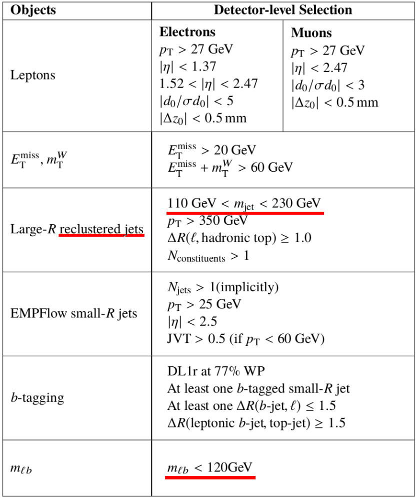
All-hadronic

Conclusions

- Analysis currently undergoing ATLAS review, with paper draft prepared
- Measurements presented in both lepton+jets and all-hadronic channels
- Results are compared to various MC generators and variations on the nominal MC
- With respect to previous ATLAS measurement:
  - Full Run-2 dataset analysed
  - Additional observables measured
  - Reduced uncertainties
  - All-hadronic channel also measured

Backup

Lepton + Jets Selection



Objects	**Detector-level Selection**	
	Electrons	**Muons**
Leptons	$p_T > 27$ GeV $\lvert\eta\rvert < 1.37$ $1.52 < \lvert\eta\rvert < 2.47$ $\lvert d_0/\sigma d_0\rvert < 5$ $\lvert\Delta z_0\rvert < 0.5\,\text{mm}$	$p_T > 27$ GeV $\lvert\eta\rvert < 2.47$ $\lvert d_0/\sigma d_0\rvert < 3$ $\lvert\Delta z_0\rvert < 0.5\,\text{mm}$
E_T^{miss}, m_T^W	$E_T^{\text{miss}} > 20$ GeV $E_T^{\text{miss}} + m_T^W > 60$ GeV	
Large-R reclustered jets	$110 \text{ GeV} < m_{\text{jet}} < 230$ GeV $p_T > 350$ GeV $\Delta R(\ell, \text{hadronic top}) \geq 1.0$ $N_{\text{constituents}} > 1$	
EMPFlow small-R jets	$N_{\text{jets}} > 1$(implicitly) $p_T > 25$ GeV $\lvert\eta\rvert < 2.5$ JVT > 0.5 (if $p_T < 60$ GeV)	
b-tagging	DL1r at 77% WP At least one b-tagged small-R jet At least one $\Delta R(b\text{-jet}, \ell) \leq 1.5$ $\Delta R(\text{leptonic } b\text{-jet, top-jet}) \geq 1.5$	
$m_{\ell b}$	$m_{\ell b} < 120$GeV	

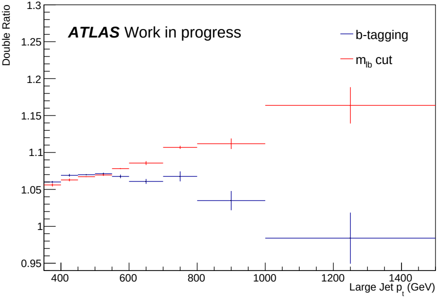
ATLAS Work in progress

m_{lb} Cut

- No *b*-matching means not biasing the substructure

- Higher efficiency than *b*-matching at high-p_{T}, double-ratio:

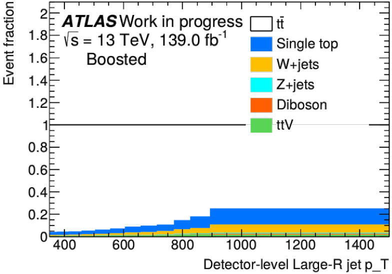
$$R = \frac{\left(\frac{\text{Signal}}{\text{Total}}\right)_{\text{After Cut}}}{\left(\frac{\text{Signal}}{\text{Total}}\right)_{\text{Before Cut}}}$$



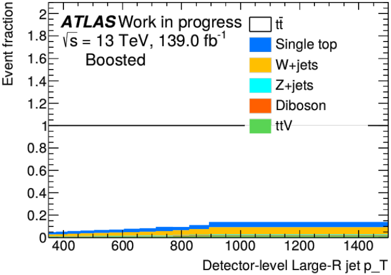
- Combined with mass cut on jet

m_{lb} Cut

- This cut not only reduces the fraction of events which are background, but the proportion of those events which are single top, a process carrying a large uncertainty.

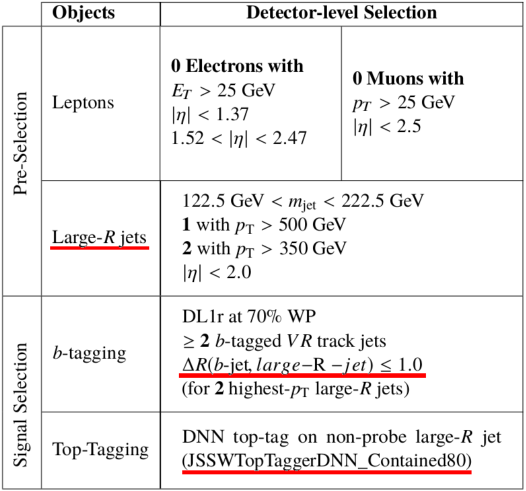


b-tag cut.



m_{lb} + mass window cut.

All-Hadronic Selection



	Objects	Detector-level Selection	
Pre-Selection	Leptons	**0 Electrons with** $E_T > 25$ GeV $\lvert\eta\rvert < 1.37$ $1.52 < \lvert\eta\rvert < 2.47$	**0 Muons with** $p_T > 25$ GeV $\lvert\eta\rvert < 2.5$
	Large-*R* jets	122.5 GeV $< m_{\text{jet}} <$ 222.5 GeV **1** with $p_{\mathrm{T}} > 500$ GeV **2** with $p_{\mathrm{T}} > 350$ GeV $\lvert\eta\rvert < 2.0$	
Signal Selection	*b*-tagging	DL1r at 70% WP $\geq$ **2** *b*-tagged *VR* track jets $\Delta R(b\text{-jet}, large{-}\mathrm{R}-jet) \leq 1.0$ (for **2** highest-p_{T} large-*R* jets)	
	Top-Tagging	DNN top-tag on non-probe large-*R* jet (JSSWTopTaggerDNN_Contained80)	

ATLAS Work in progress

Substructure Observables: Generalised Angularities

- For a hadron collider, the generalised angularities can be defined:

$$\lambda_\beta^\kappa = \sum_{i \in J} z_i^\kappa \left(\frac{\Delta R(i, \hat{n})}{R} \right)^\beta$$

  - z_i^κ the fraction of the jet's p_T carried by the i^{th} constituent
  - $\hat{n}$ is the jet's axis
  - ΔR the distance in $\eta - \phi$ space
  - R the jet radius parameter
  - κ and β are p_T and angular weighting terms, respectively

Substructure Observables: Generalised Angularities

- Measured in the analysis are:
  - The Les Houches Angularity, $\lambda^1_{0.5}$, which provides a measure of the width of a jet.
  - The scaled p_T dispersion,

$$\lambda_0^{2*} = p_\mathrm{T}^{\mathrm{d},*} = \sqrt{(p_\mathrm{T}^\mathrm{d} - \frac{1}{N})\frac{N}{N-1}}$$

  p_T^d is the standard p_T dispersion (λ_0^2) and N is the particle multiplicity. In this way, $\lambda_0^{2*} \to 0$ for equally distributed constituents and $\lambda_0^{2*} \to 1$ when most of the momentum is carried by a single particle.

Substructure Observables: Energy Correlation Functions

- For a hadron collider, the Energy Correlation Functions (ECFs) provide a way to probe N-pronged structure in a jet and may be defined:

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

  - The sum runs over all constituents in the jet
  - i always denoting a constituent of the jet
  - N the number of prongs
  - ΔR is the distance in $\eta - \phi$ space between pairs of constituents with combinations determined by the product terms
  - β is an angular weighting term

Substructure Observables: Energy Correlation Functions

- Measured in the analysis are:
  - $D_2^{(\beta)} = \frac{\mathrm{ECF}(3,\beta)\,\mathrm{ECF}(1,\beta)^3}{\mathrm{ECF}(2,\beta)^3}$: useful for identifying two-pronged structure, with lower values being more consistent with that structure.
  - $C_3^{(\beta)} = \frac{\mathrm{ECF}(4,\beta)\,\mathrm{ECF}(2,\beta)}{\mathrm{ECF}(3,\beta)^2}$: useful for identifying three-pronged structure, with lower values being more consistent with that structure.
  - $\mathrm{ECF2}^{\mathrm{NORM}}(\beta) = \frac{\mathrm{ECF2}(\beta)}{\mathrm{ECF1}^2}$: The normalised version of ECF2.
- For all of the above, β is set to one.

Substructure Observables: N-Subjettiness

- For a hadron collider, the N-subjettiness variables provide a way to probe N-pronged structure in a jet and may be defined:

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

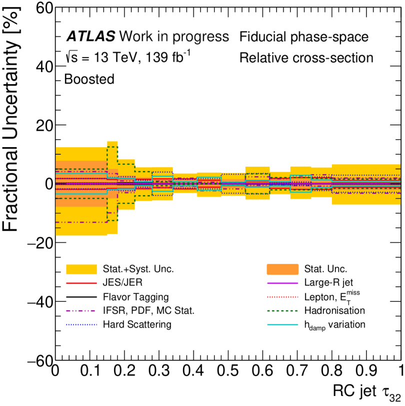
$$d_0 = \sum_k p_{t,k} R_0.$$

  - Constituents of the jet are reclustered using the exclusive k_t algorithm
  - Clustering continues until N subjets are returned
  - Calculation of τ_N then runs over the constituents of the jet
  - $p_{t,k}$ the p_{T} of the k^{th} constituent
  - $\Delta R_{N,k}$ is the distance in $\eta - \phi$ space between the axis of the N^{th} subjet and the k^{th} constituent
  - R_0 is the jet radius parameter

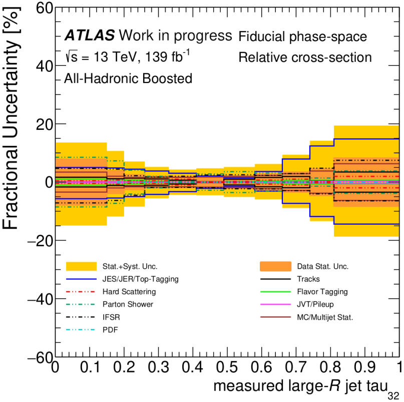
Substructure Observables: *N*-Subjettiness

- Measured in the analysis are:
    - τ_3: tends towards lower values when structure is consistent with 3 or fewer prongs
    - $\tau_{32} \equiv \frac{\tau_3}{\tau_2}$: tends to lower values when consistent with 3-pronged substructure
    - $\tau_{21} \equiv \frac{\tau_2}{\tau_1}$: tends to lower values when consistent with 2-pronged substructure

Systematics



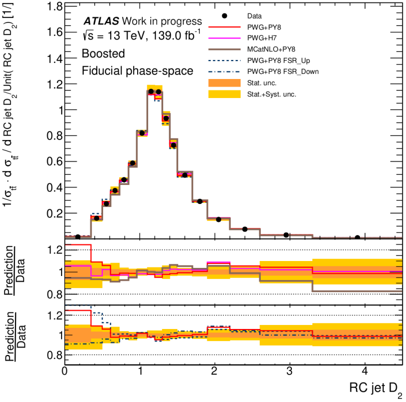
Lepton + jets



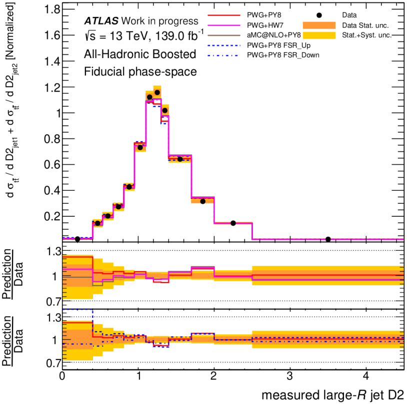
All-hadronic

Particle-level results

D_2 exhibits sensitivity to variations in the MC predictions



Lepton + jets



All-hadronic