Measurement of jet substructure observables using the ATLAS detector

Amal Vaidya,
the ATLAS collaboration

April 10, 2019

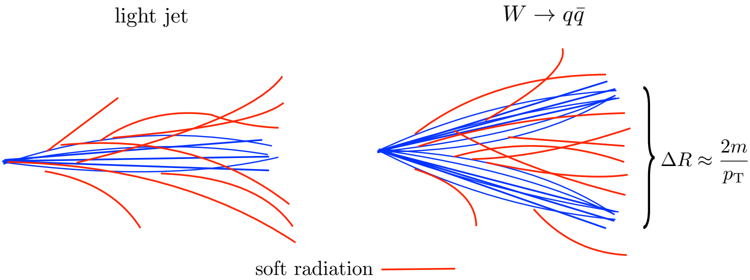
27th Workshop on Deep-Inelastic Scattering and Related Subjects

Jet substructure

Studying the internal structure of jets

- Uses for separating hadronic decays of light partons from W, top, Z
- Sensitive to pile-up and MPI
- sensitive to non-perturbative QCD

At the LHC electroweak particles can also be studied with JSS
New techniques facilitated recent calculation of numerous variables



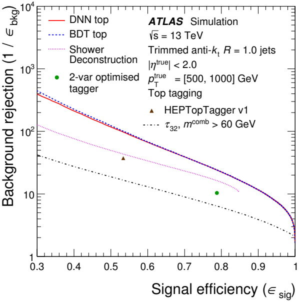
Recent pedagogical review: S. Marzani *et al.*; Summary of theory developments A. Larkoski *et al.*

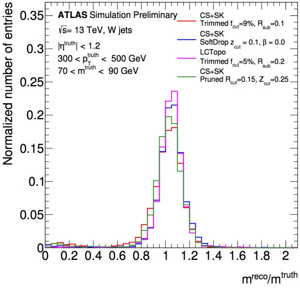
Jet substructure in ATLAS

Extensive performance work

- Large number of analyses studying W, Z, top jets in final states
- Tagging performance studies, including ML effort [1808.07858]
- Studying combinations of grooming and pile-up mitigation algorithms
- The development of better jet constituent reconstruction techniques

Experimentally challenging regime





The measurements

A measurement of the soft-drop jet mass in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector

[1711.08341]

- State of the art grooming tools have facilitated theoretical calculations
- First JSS quantity at the LHC to be calculated at NNLL

Measurement of jet-substructure observables in top quark, W boson and light jet production in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector

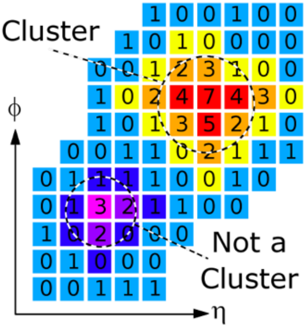
[1903.02942]

- Variables used for top, W and Z jet tagging
- Dedicated selections enriched with W, top and light jets

Topoclustering: making jet constituents

Selecting the significant signal from calorimeter noise

- Want to suppress pile-up and other noise
- Assign a significance based on the expected noise level
- Form 3D clusters corresponding showers
- Additional calibration steps at the cluster level

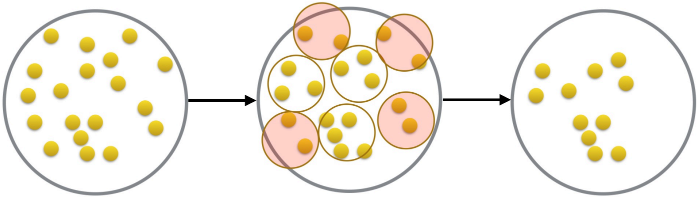


In these two analyses we consider topocluster based jets

Jet grooming

Trimming Used in ATLAS based on performance studies

(D. Krohn *et al*)

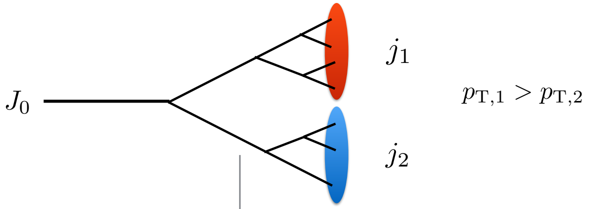


Soft-drop grooming

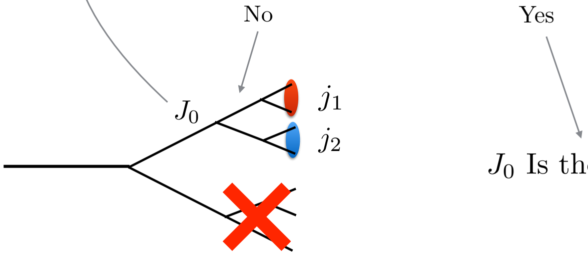
(A. J. Larkoski *et al*)

- Cluster together constituents using the C/A algorithm
- undo last clustering step to get resulting subjets j_1 and j_2
- If the subjets pass the soft-drop condition: $\frac{\min(p_{T1}, p_{T2})}{p_{T1}+p_{T2}} > z_{\mathrm{cut}} \left(\frac{\Delta R_{12}}{R_0}\right)^{\beta}$ then j is the final jet
- Otherwise, discard lowest p_T subjet and iterate.

soft-drop grooming



$$\frac{\min(p_{T,1}, p_{T,2})}{p_{T,1} + p_{T,2}} > z_{\mathrm{cut}} \left(\frac{\Delta R_{12}}{R_0} \right)^{\beta}$$



J_0 Is the final jet

Soft-drop mass

non-global logarithms (NGLs): resummation terms associated with particles that radiate out of and then radiate back into a jet.

- have prevented next-to-leading logarithm accuracy calculations
- soft-drop groomed variables are formally insensitive to NGLs
- soft-drop mass calculations now available at
  NLO+NLL S. Marzani *et al.* [1, 2]
  LO+NNLL C. Frye *et al.* [1, 2]

Variable of Interest: $\log_{10}(\rho^2)$ where

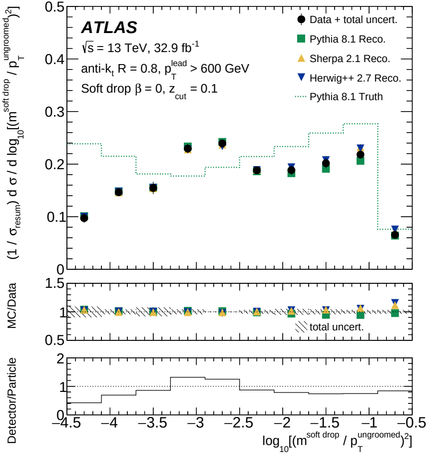
$$\rho = m^{\text{soft-drop}} / p_{\mathrm{T}}^{\text{ungroomed}}$$

ρ is dimensionless, only weakly correlated to p_{T} unlike the mass
p_{T} is collinear unsafe for $\beta = 0$

Measure soft-drop parameters $\beta = \{0, 1, 2\}$, $z_{\text{cut}} = 0.1$

Select dijet events

- Use single jet trigger to select events with $\geq$ 2 anti-k_t R=0.8 jets
- Leading jet in the event $p_{\text{T}} > 600$ GeV based on trigger
- Select leading two central $|\eta| < 1.5$ jets
- Require $p_{\text{T},j_1}/p_{\text{T},j_2} < 1.5$, removes events with additional jets



soft-drop mass

Experimental uncertainties

Require novel approach, ATLAS Jet systematics typically “top down”

Systematic variations on clusters

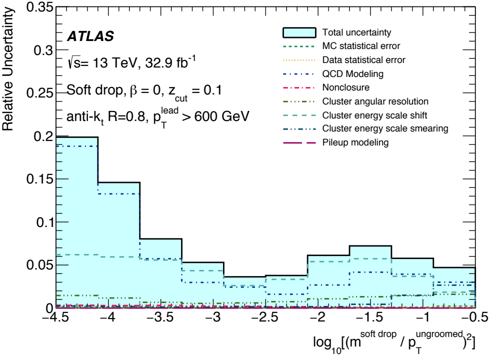
- Account for energy and angular resolution, reconstruction efficiency
- Based on isolated clusters matched to tracks the E/p ratio is used
- Extensive validation to show validity for non-isolated clusters
- Data from test beam studies is also used

Unfolding

- Use iterative bayesian unfolding
- LO Pythia 8 (p_T-ordered) used as nominal MC
- Additional Sherpa 2 (p_T-ordered) and Herwig++ (angular ordered) samples used for comparison

soft-drop mass - unfolding

Modelling uncertainty



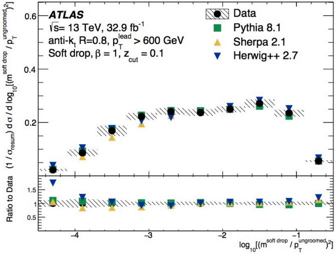
- Largest source of uncertainty is QCD fragmentation modelling
- Estimate by unfolding with alternate (Sherpa) MC response matrices

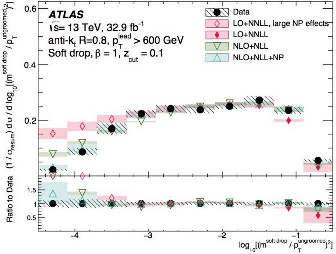
Modelling and cluster energy scale are largest sources of uncertainty
NP correction for NLL based on NP effects predicted by MC models

soft-drop mass - results

Different regions of the phase space are sensitive to different effects

- NP effects are large for $\log_{10}(\rho^2) < -3.7$
- fixed order ME effects expected to be important for $\log_{10}(\rho^2) > -1.7$
- Resummation region $-3.7 < \log_{10}(\rho^2) < -1.7$





- agreement worse for $\log_{10}(\rho^2) < -3.7$
- NNLL is not better than NLL everywhere in resummation region
- In the future improvements from better understanding of NP effects would be beneficial

Jet substructure measurement

What are we measuring?

Trimmed Softdrop ($\beta = 0$)	$\times$	W top dijet	$\times$	ECF$(2,3)$ $D_2^{(\beta=1)}$ $C_2^{(\beta=1)}$ $(\tau_{21}^{WTA}, \tau_{32}^{WTA})^*$ LH angularity N subjets

- JSS variables are used extensively for top/W/Z and higgs tagging
- No ATLAS unfolded measurement of variables at 13 TeV
- Useful for MC model and tagger development
- A set of softdrop groomed variables can be calculated analytically - strong theoretical motivation

*Only for W and top jet selection

Substructure variables

(generalised) *N*-subjetiness

(J. Thaler and K. V. Tilburg)

$$\tau_N^{(\beta)} = \frac{1}{d_0} \sum_i p_{\mathrm{T}i} \min\left\{(\Delta R_{1,i})^\beta, (\Delta R_{2,i})^\beta ... (\Delta R_{N,i})^\beta\right\}$$

A set of N subjet axis are defined using the exclusive k_t algorithm.

Normalised energy correlation functions (and ratios)

(A. J. Larkoski *et al*)

$$e_2^\beta = \frac{1}{p_{\mathrm{T}\ J}^2} \sum_{i<j\in J} p_{\mathrm{T}i} p_{\mathrm{T}j} R_{ij}^\beta \qquad e_3^\beta = \frac{1}{p_{\mathrm{T}\ J}^3} \sum_{i<j<k\in J} p_{\mathrm{T}i} p_{\mathrm{T}j} p_{\mathrm{T}k} R_{ij}^\beta R_{jk}^\beta R_{ik}^\beta$$

subjet independent way to discriminate between a ”two-pronged” jet and a single prong parton initiated jet.

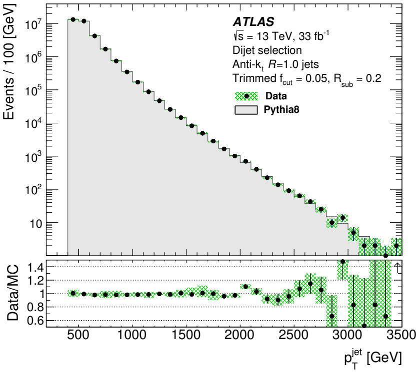
Ratio of energy correlation functions: $D_2^\beta = \frac{e_3^\beta}{\left(e_2^\beta\right)^3}$, $C_2^\beta = \frac{e_3^\beta}{\left(e_2^\beta\right)^2}$

$\beta = 1$ typically used for tagging

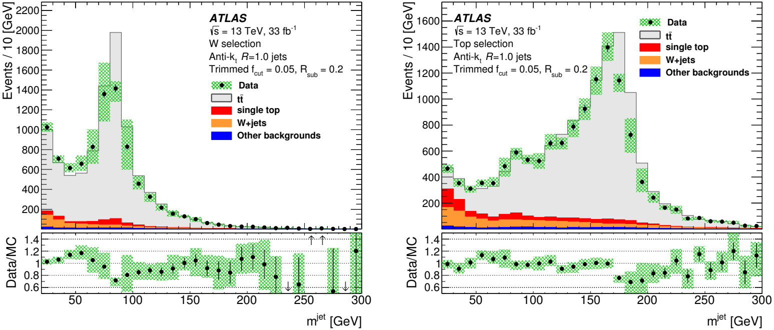
Dijet selection

Simple event selection

- Events with 2 large R jets with a $p_T > 200$ GeV
- Leading jet $p_T > 450$ GeV (trigger plateau)
- veto on leptons
- Consider central ($|\eta| < 1.5$) leading jet from each event



$t\bar{t}$ selection



Tag leptonic top and consider recoiling fatjet:

- Select events with 1 muon and tag leptonic top
- Separate *W* jets from top jets
- Differentiate based on ΔR of closest *b*-tagged small *R* jet
- Additional cuts applied on the jet mass

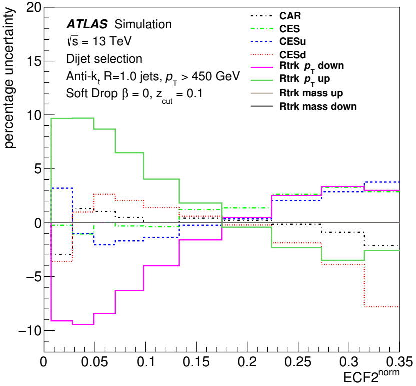
Jet substructure measurement

Additional background subtraction for $t\bar{t}$

- Data driven W+jets subtraction
- Fake factor, accounting for jets faking muons

Detector systematics

- Also use cluster based systematics
- Additional validation (cluster splitting/merging etc.) done for the extra variables
- Large R jet kinematics and cluster uncertainties are dominating detector systematics



Jet substructure measurement

Unfolding

- PowhegPythia8 used as nominal $t\bar{t}$ MC
- Pythia8 as nominal dijet MC
- Modelling uncertainty by unfolding with alternate MC Sherpa2.1 for dijet and Herwig + +
- Typically around 5% for dijet selection and 14% for $t\bar{t}$ selection.

Predictions

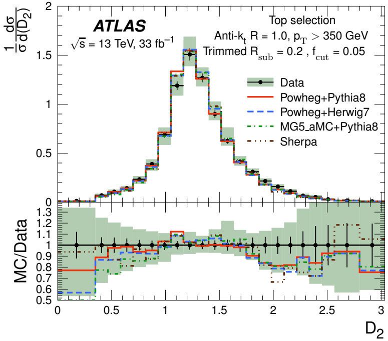
Dijet Additionally compare unfolded distributions with Herwig 7 and Sherpa with two different hadronisation models

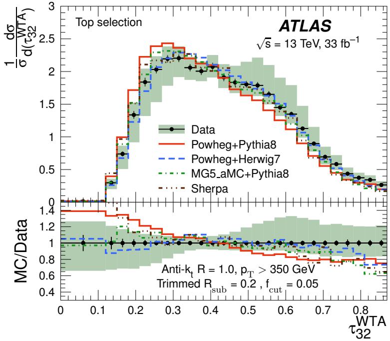
$t\bar{t}$ Additional comparisons with MadGraph5_aMC@NLO+Pythia8, Powheg + Herwig 7

Jet substructure measurement - results

$t\bar{t}$ results

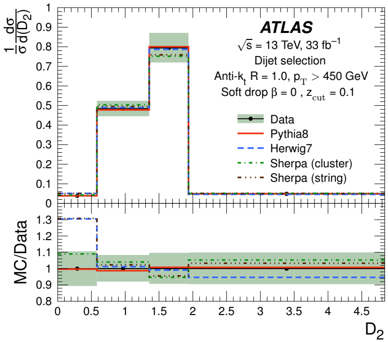
- Pythia8 tends to perform the best
- Sherpa Has a comparable performance to Pythia8
- Herwig 7 Tends to disagree with the data the most

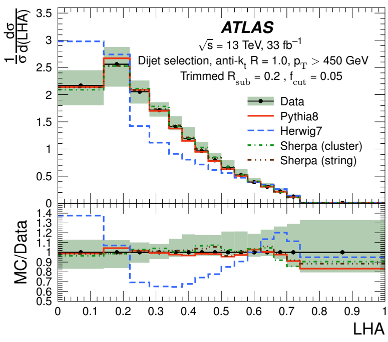




Dijet results

- Pythia8 tends to perform the best
- Negligible difference between the two Sherpa hadronisation models
- Herwig 7 Tends to disagree with the data the most
  Using older 7.0.4 version





Conclusions

Studying jet substructure variables allows us to explore a new region of phase space

- NP effects can be significant in certain regions of the soft-drop mass phase space
- MC predictions tend to agree with the data

- Considering jet shapes, generators tend to perform consistently for the multiple variables tested (with some exceptions)
- Dijet variables seem to be insensitive to NP effects
- Will be interesting to see performance of newer Herwig 7 version

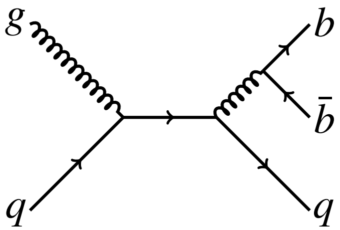
Further motivates development of constituent based uncertainties

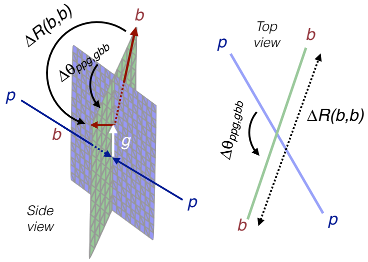
Final word: $g \to b\bar{b}$ fragmentation

Properties of $g \to b\bar{b}$ at small opening angles in pp collisions with the ATLAS detector at $\sqrt{s} = 13$ TeV

Phys. Rev. D 99 (2019) 052004

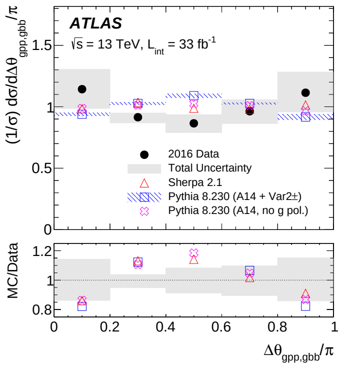
- Select events with a gluon , decaying to $b\bar{b}$, recoiling of a jet
- Reconstruct $g \to b\bar{b}$ within a large R jet which contains b-jets
- Unfold variables sensitive to b-jet fragmentation

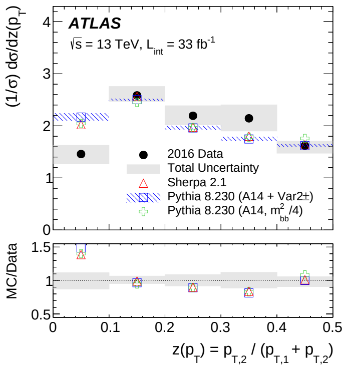




Final word: $g \rightarrow b\bar{b}$ fragmentation

- Compare unfolded results to Sherpa and Pythia with two different QCD renormalisation scales
- Significant disagreement between data and MC seen in some variables





BACKUP

MC generators for JetShapes

Process	Generator	Version	PDF	Tune	Use
Dijet	Pythia8	8.186	NNPDF23LO	A14	Nominal for unfolding
	Sherpa	2.2.1	CT10	Default	Validation of unfolding (with two different hadronisation models)
	Herwig7	7.0.4	MMHT2014	H7UE	Comparison
$t\bar{t}$	Powheg	v2	NNPDF30NLO		Nominal for unfolding
	+ Pythia8	8.186	NNPDF23LO	A14	
	Powheg	v2	CT10		Validation of unfolding
	+Herwig++	2.7	CTEQ6L1	UE-EE-5 tune	
	Powheg	v2	CT10		Comparison
	+Herwig7	7.0.4	MMHT2014	H7UE	
	MG5_aMC@NLO	2.6.0	NNPDF30NLO		Comparison
	+ Pythia8	8.186	NNPDF23LO	A14	
	Sherpa	2.2.1	CT10	Default	Comparison
Single top	Powheg	v1	CT10		Nominal for unfolding
	+ Pythia6	6.428	CTEQ6L1	Perugia2012	
Z+jets	Sherpa	2.2.1	CT10	Default	Background estimation
W+jets	Sherpa	2.2.1	CT10	Default	Background estimation (nominal)
W+jets	MG5_aMC@NLO	2.2.5	CT10		Background estimation (cross-check)
	+ Pythia8	8.186	NNPDF23LO	A14	
Diboson	Sherpa	2.2.1	CT10	Default	Background estimation

Systematics

Need to determine a systematic uncertainty on the shape of each variable
Rtrk only provides uncertainties on the scale, need something new

Cluster systematics

Apply the following variations to the ungroomed jet topoclusters separately:

- Cluster efficiency (CE) - drop some low p_T ($<$ 10 GeV) clusters
- Cluster energy scale up/down (CESu/CESd) - based on 13 TeV E/p
- Cluster energy smearing (CES) - Smear energy based on above
- Cluster Angular resolution (CAR) - track - isolated cluster angular difference

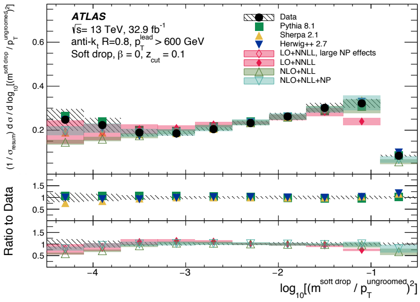
Reapply the jet grooming and re-calculate the observable

Cluster and large R systematics dominate in all topologies

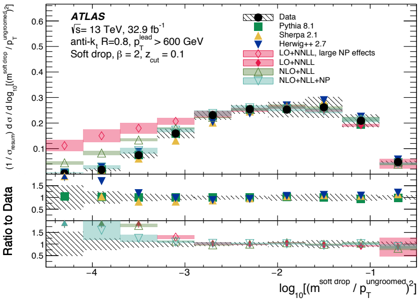
Soft-drop mass

Other β values

$\beta = 0$

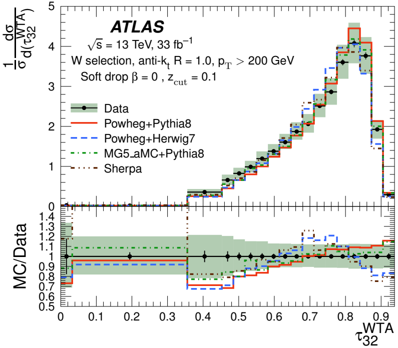


$\beta = 2$



Jet shapes *N*-subjettiness

W selection



Top selection