Tagging boosted jets from top quarks and *W* bosons using jet substructure and multivariate techniques

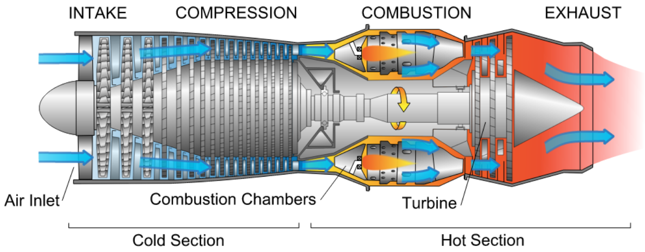
IOP joint HEPP and APP conference

March 2018

Amal Vaidya

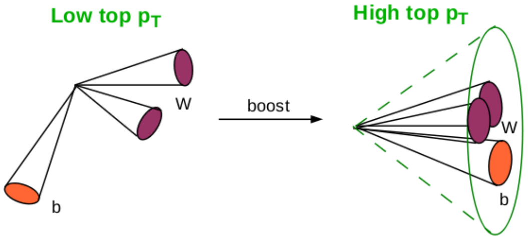
Overview

- Jet tagging and substructure
- Improving ATLAS jet tagging
- Run 2 performance
- Newer techniques
- Results



Hadronic jet tagging

Boosted, hadronically decaying massive particles can be reconstructed in a large radius jet



SM measurements and BSM searches study final states with *W* and top jets. Probe the **substructure** of each of each jet to identify it

The challenges

- Huge QCD multi-jet backround
- Pileup: stochastic noise smears signal
- Finite calorimeter angular resolution

Jet substructure variables

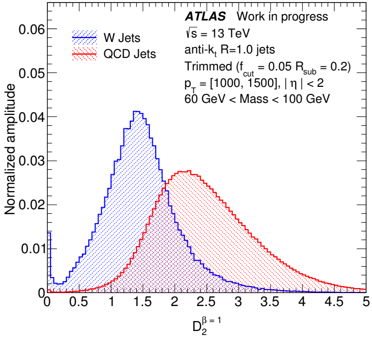
Variables designed to be sensitive to discriminating properties of jets

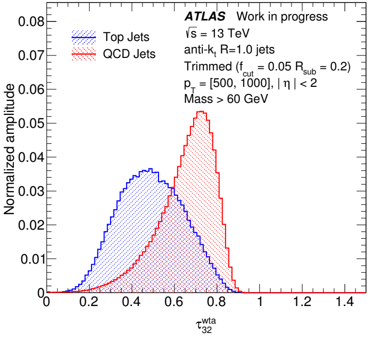
example: *N*-subjetiness

- A set of *N* subjet axis are defined using the exclusive k_t algorithm.
- Sum over constituents of each jet

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

D_2 variable is a subjet independent method of probing same structure





Jet substructure variables

Many many other variables exist (this is a small subset) and were studied in the following analyses

Technique	Variable	Used for
Jet Mass	Calorimeter Mass Track Assisted Mass	W, Top W, Top
Energy Correlation functions	ECF_{1-3} + newer D_2, C_2, M_2, N_2	W, Top W, Top
N-subjettiness	τ_1, τ_2, τ_3 τ_{21}, τ_{32}	W, Top W, Top
Splitting measures	Z_{cut}, Q_W μ_{12} $\sqrt{d_{12}}, \sqrt{d_{23}}$	W W W, Top
Shower histories	Shower Deconstruction	W, Top

Improving run 2 performance

- Better mass reconstruction
- 2 variable tagger optimisation

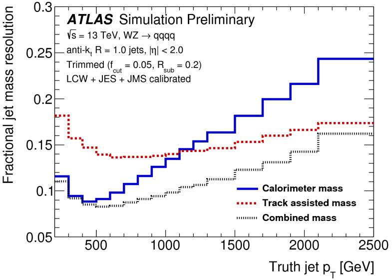
Better mass reconstruction

Track-assisted mass: $m_{TA} = m_{track} \frac{p_T^{calo}}{p_T^{track}}$

- consider tracks associated to jet, scale to calorimeter p_T
- lack of neutral reconstruction mitigated by better angular resolution

Combined mass: Use tracking information to improve jet mass resolution

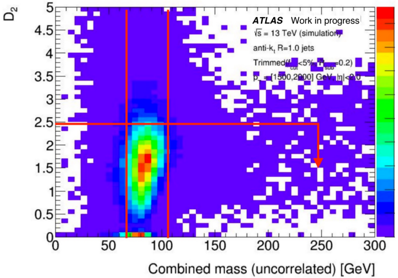
- Use linear combination of the two: $m_{comb} = w_{calo} m_{calo} + w_{TA} m_{TA}$
- weights based on resolution in jet phase space $w_{calo} = \frac{\sigma_{calo}^{-2}}{\sigma_{calo}^{-2} + \sigma_{TA}^{-2}}$



JETM-2017-002

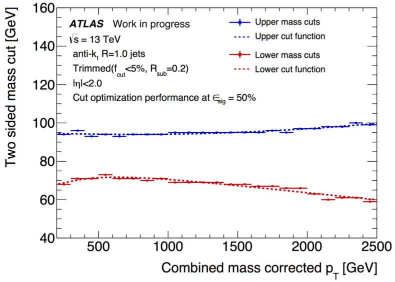
- better mass resolution across the jet p_T spectrum than either definition
- Allows for tighter cuts on mass, leads to better tagging performance

2 variable optimisation



Optimise two variable cuts in parallel

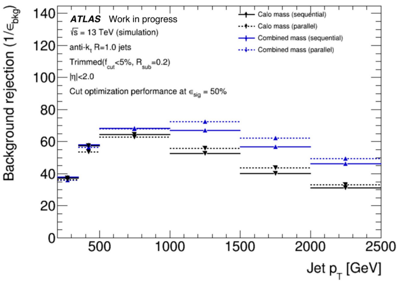
- Find each set of cuts that satisfies the required working point
- Select the set of cuts that maximises background rejection
- Do for narrow p_T bins

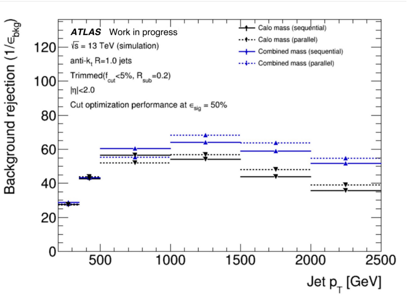


- Results in a set of p_T dependant cuts
- Fit to form a smooth set of cuts which define the tagger

Performance

Compare the performance relative to run 1 techniques





Look at background rejection at a fixed signal efficiency

- Better mass resolution allows for a tighter cut on the jet mass, increases background rejection
- New optimisation procedure also independently improves performance by optimising cuts in the variables in parallel

Studying newer techniques

- Shower deconstruction
- MVA techniques: Boosted decision tree (BDT) and deep neural network (DNN)

Newer techniques: Shower deconstruction

Calculate an observable for each jet in order to discriminate between signal and background jets

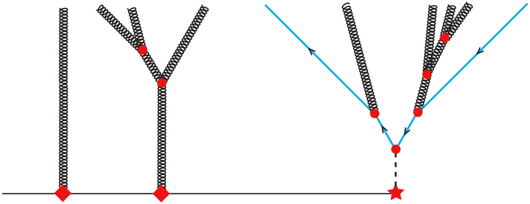
$$\chi(\{p\}_N) = \frac{P(\{p\}_N|S)}{P(\{p\}_N|B)}$$

for a set of N subjet momenta for an input jet, $\{p\}_N$

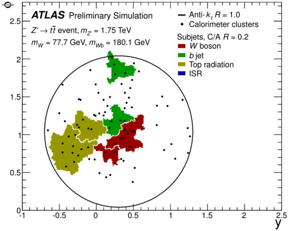
Use a simplified showering algorithm to calculate $P(\{p\}_N|S)$, $P(\{p\}_N|B)$

- Simulate hard scatter and ISR only
- Only consider partons which fall within the large R jet
- Approximate decay and splitting probabilities
- Repeat using signal and background model for each jet

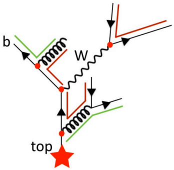
Finding physics signals with shower deconstruction



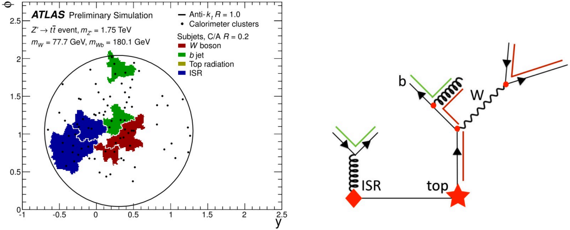
Sum over all possible shower histories in order to determine $P(\{p\}_N|S), P(\{p\}_N|B)$



(a)



(b)



ATLAS-CONF-2014-003

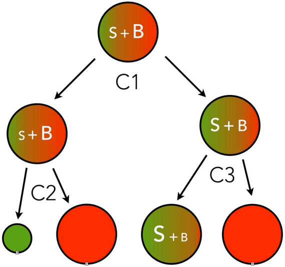
Some jets are instantly rejected if they are deemed incompatible with the signal hypothesis

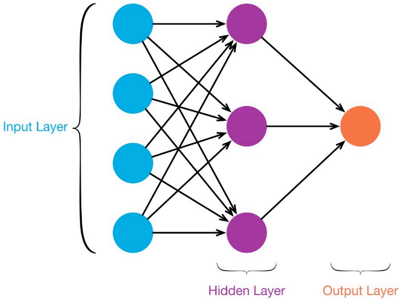
Newer techniques: multi-variate techniques

Assess the performance of boosted decision tree (BDT) and deep neural network (DNN) algorithms

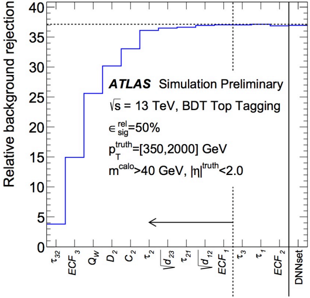
In each case the set of **input variables** and the **hyperparameters** of each technique were studied and optimised

- Large training and testing samples were obtained
- A large number of variables were tested, not including the jet mass
- The correlations between the variables were also studied



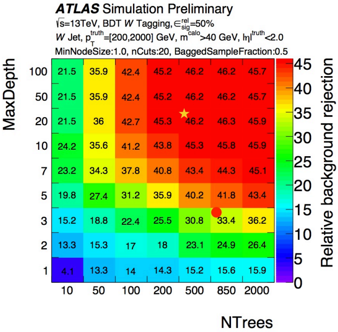


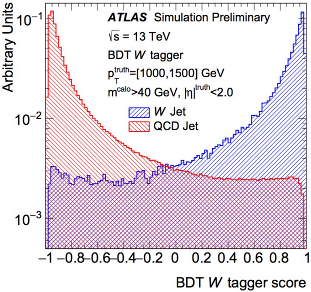
BDT tagger



- Sequentially add best performing variables
- Scan over the various hyperparameters for optimum
- Can cut on output score to obtain 50% working point

ATL-PHYS-PUB-2017-004



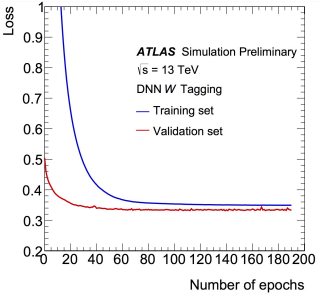


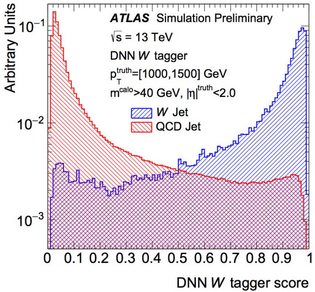
DNN tagger

	Top Tagging Observable Groups						
Observable	1	2	3	4	5	6	7 (BDT)
ECF_1				○		○	○
ECF_2				○		○	
ECF_3				○		○	○
C_2					○	○	○
D_2					○	○	○
τ_1		○	○	○	○	○	
τ_2		○	○	○	○	○	○
τ_3		○	○	○	○	○	
τ_{21}	○		○		○	○	○
τ_{32}	○		○		○	○	○
$\sqrt{d_{12}}$	○	○	○	○	○	○	○
$\sqrt{d_{23}}$	○	○	○	○	○	○	○
Q_w	○	○	○	○	○	○	○

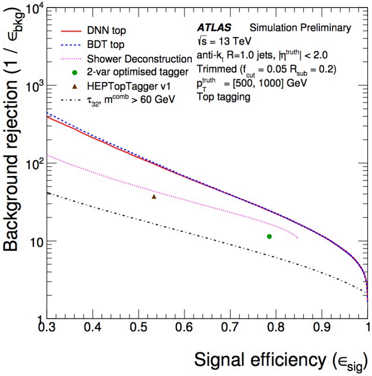
- Run training on groups of variables based on correlations (difficulty due to large training times)
- Overtraining is tested by studying the loss with a validation set
- Can cut on output score to obtain 50% working point

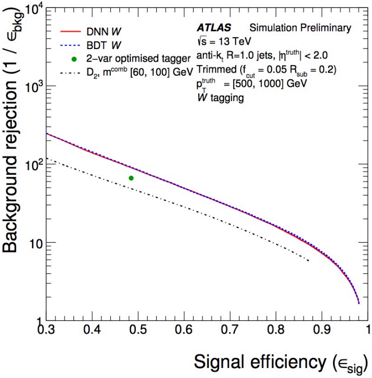
ATL-PHYS-PUB-2017-004





Results





ATLAS-CONF-2017-064

Direct comparison of the above techniques

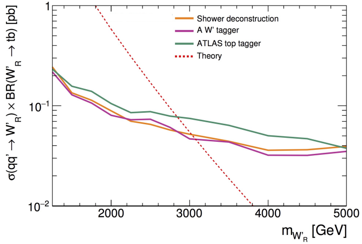
- Practically no difference in DNN and BDT performance
- Shower deconstruction is the best **single variable** top tagger studied (but not for *W*s)
- Newer techniques outperform traditional substructure cut based taggers

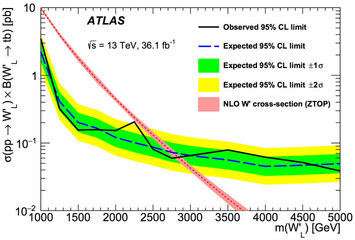
Summary

There has been a significant improvement in tagging performance using new reconstruction and MVA techniques

- Above MC results have been studied in data ATLAS-CONF-2017-064
- MVA techniques exploit correlations between variables for better performance, up to a point
- Important also to consider better reconstruction techniques which can have a significant effect on performance

Better performance directly impacts physics results! ($W' \rightarrow tb$ search)



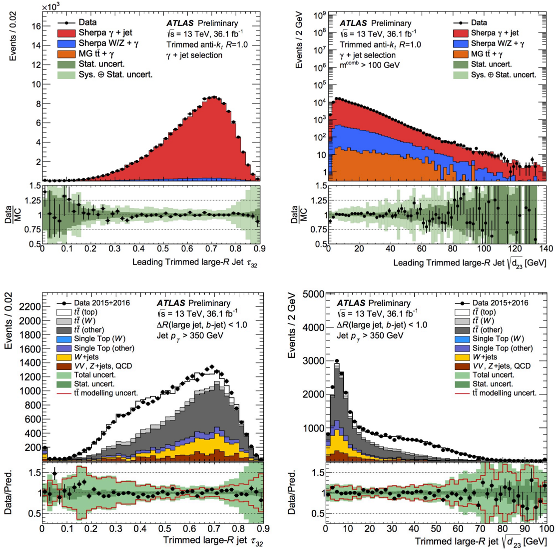


CERN-EP-2017-340

Backup

Study the JSS and MVA variables in data

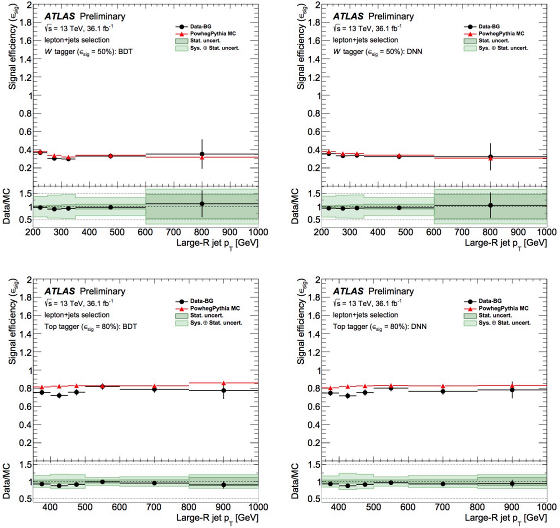
Use a selection of qcd jets from γ+jets events and W and top jets from $t\bar{t}$ events



ATLAS-CONF-2017-064

Study the JSS and MVA variables in data

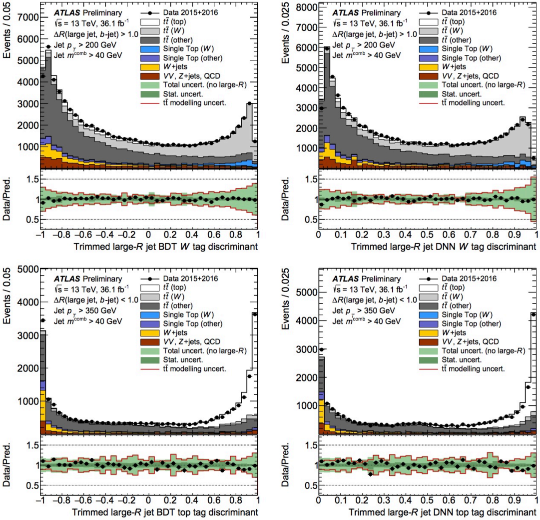
Can study the tagger performance in the data sample



ATLAS-CONF-2017-064

Study the JSS and MVA variables in data

And also look the MVA tagger output score



ATLAS-CONF-2017-064

Jet substructure variables

(generalised) N-subjetiness

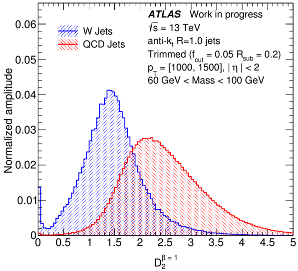
$$\tau_N^{(\beta)} = \frac{1}{d_0} \sum_i p_{\mathrm{T}i} \min\left\{ (\Delta R_{1,i})^\beta, (\Delta R_{2,i})^\beta ... (\Delta R_{N,i})^\beta \right\}$$

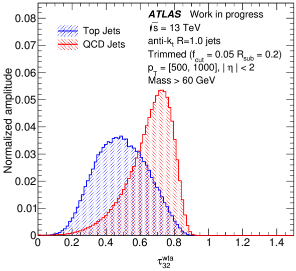
A set of N subjet axis are defined using the exclusive k_t algorithm.

Energy correlation functions (and ratios)

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

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





Hype rparameters

Optimised BDT hyperparameters

Setting Name	Value Tested
NTrees	[10, 50, 100, 200, 500, 850, 2000]
MaxDepth	[1, 2, 3, 5, 7, 10, 20, 50, 100]
MiniNodeSize	[0.5, 1.0, 2.5, 5.0, 10, 20]
nCuts	[5, 10, 20, 50, 100, 500]
Bagged Fraction	[0.1, 0.3, 0.5, 0.7, 0.9]
Shrinkage	[0.05, 0.1, 0.3, 0.5, 0.7, 0.9]

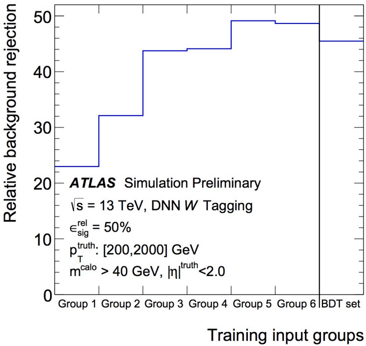
ATL-PHYS-PUB-2017-004

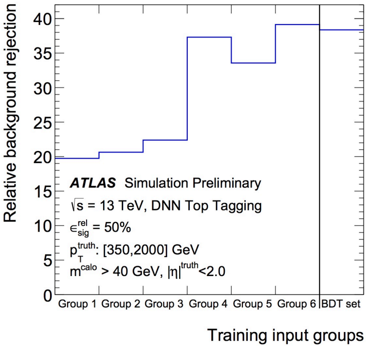
Optimised DNN hyperparameters

	W-Boson Tagging Chosen	Top-Quark Tagging Chosen	Reference
Layer type	Dense	Dense	[24]
Number of hidden layers	5	5	[24]
Activation function	rectified linear unit (relu)	rectified linear unit (relu)	[41]
Learning rate	10^{-5}	5×10^{-5}	[43]
L1 Regularizer	10^{-2}	10^{-3}	[41]
NN weight initialization	Glorot uniform	Glorot uniform	[44]
Batch size	200	200	[41]
Batch normalization	Yes	Yes	[45]
Training groups	Group 5	Group 6	-
Architecture	18, 25, 22, 19, 14, 7, 1	13, 18, 16, 14, 10, 5, 1	-

ATL-PHYS-PUB-2017-004

DNN input group performance





ATL-PHYS-PUB-2017-004