# V-tagging in ATLAS and input from ML

Steven Schramm

Particle reconstruction for VBS

October 23, 2018

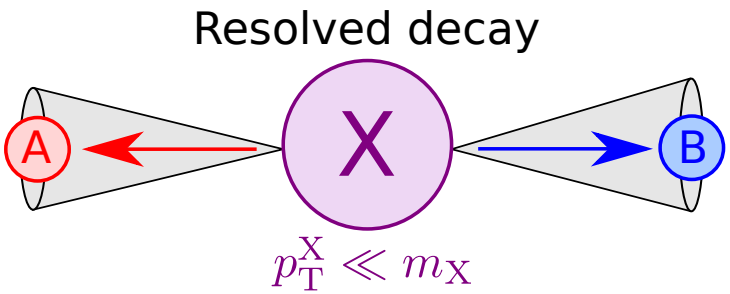
# Introduction

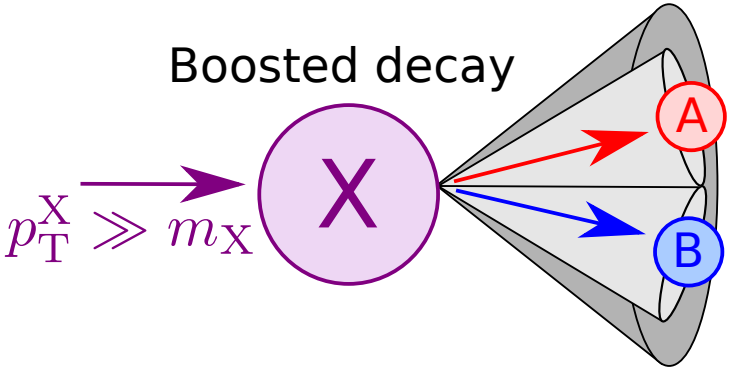
- This talk will focus on identifying hadronically decaying W/Z bosons
  - Other objects (H/top/X) are also frequently studied
  - Techniques are often transferable, just at different kinematic regimes
- Machine Learning (ML) is increasingly used for W/Z identification
  - ML is a powerful tool, and it comes with large benefits
- Important to remember that ML is not the only tool we have
  - ML alone will often not give the best possible results
  - Need to stop thinking of "ML vs non-ML techniques"
  - Instead, consider "how can we maximally benefit from all of our tools?"

- Jet inputs and tagging strategies
- Using machine learning for W/Z-taggers
- Ensuring that we can use complex taggers
- Future machine learning strategies

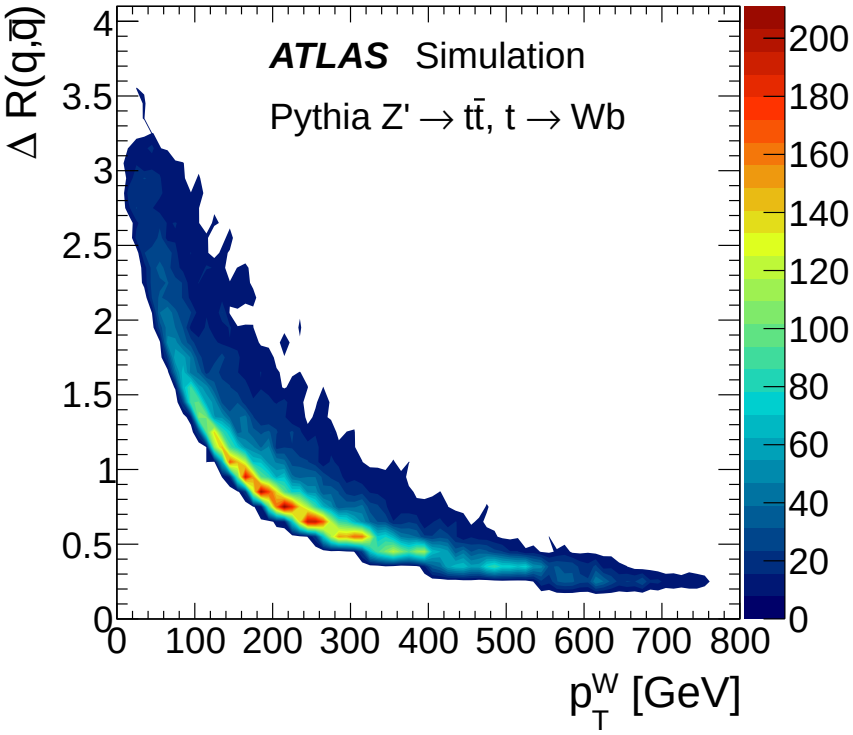
# Reminder: Boosted decays

Lower right: PERF-2012-02



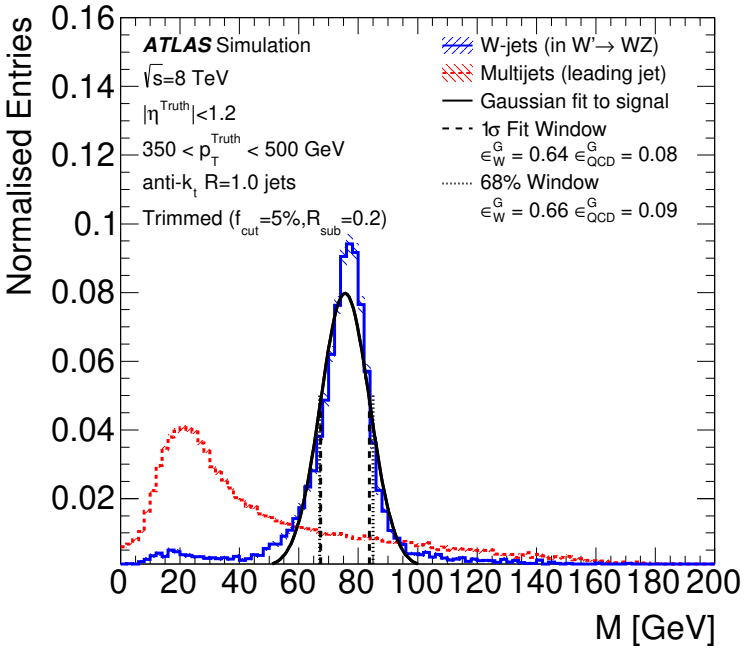

- Massive particle decay occurs back-to-back in its rest frame
- At high $p_T$, boost back to lab frame results in collimation
- Decay products can overlap
  - Large-R jet contains them all
  - ATLAS mostly uses anti-$k_t$ $R = 1.0$ topocluster jets
- Rule of thumb: $\Delta R \approx 2m/p_T$

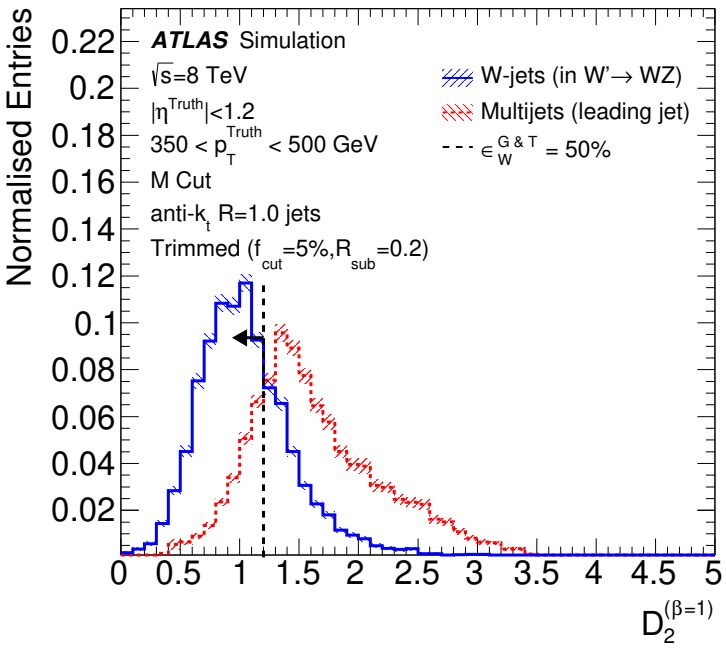


# Simple W/Z taggers

Both: PERF-2015-03

- When tagging W bosons, dominant background usually QCD multijet
  - Need to do a very good job at separating them: $\frac{\sigma(\mathrm{QCD\ multijet})}{\sigma(W/Z)} \approx 10^5$
- Numerous jet substructure variables have been studied
  - Calculated from the jet constituents (in this case, topoclusters)
- Combination of jet mass + $\mathrm{D}_2^{\beta=1}$ is best two-variable W tagger
  - Shown for W boson: mass before cuts and $\mathrm{D}_2^{\beta=1}$ after mass cut

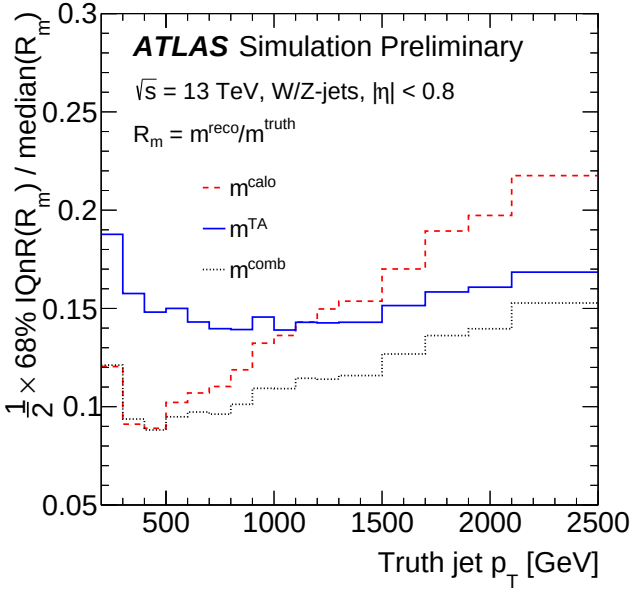


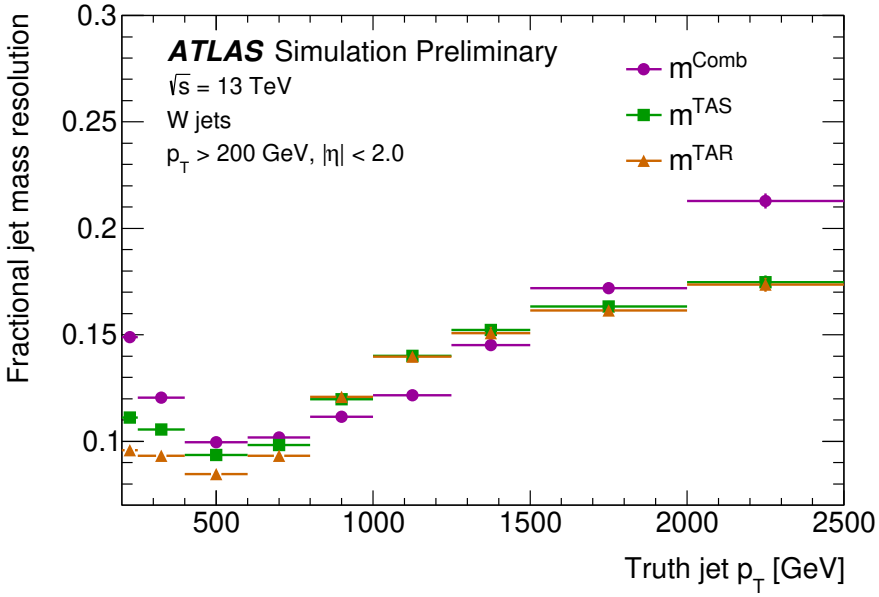



# The importance of jet inputs

Left: CONF-2016-035
Right: PUB-2018-012

- Reminder: jet substructure variables are calculated from constituents
  - By default, this means topoclusters in ATLAS
- However, substructure variables depend on both scale and direction
  - Calorimeter has best scale, tracks have best spatial resolution
- Increasingly complex track+calo combinations: better mass resolution
  - Improved resolution $\implies$ sharper W/Z peak $\implies$ better QCD veto

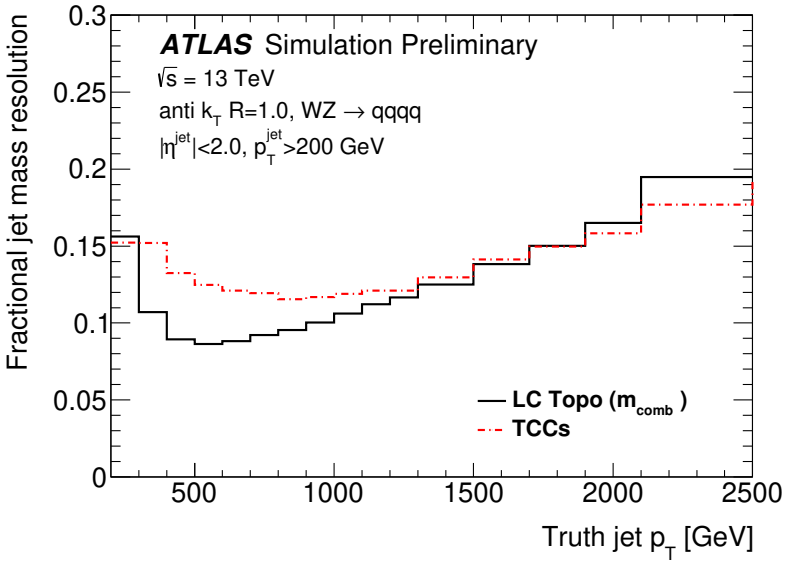


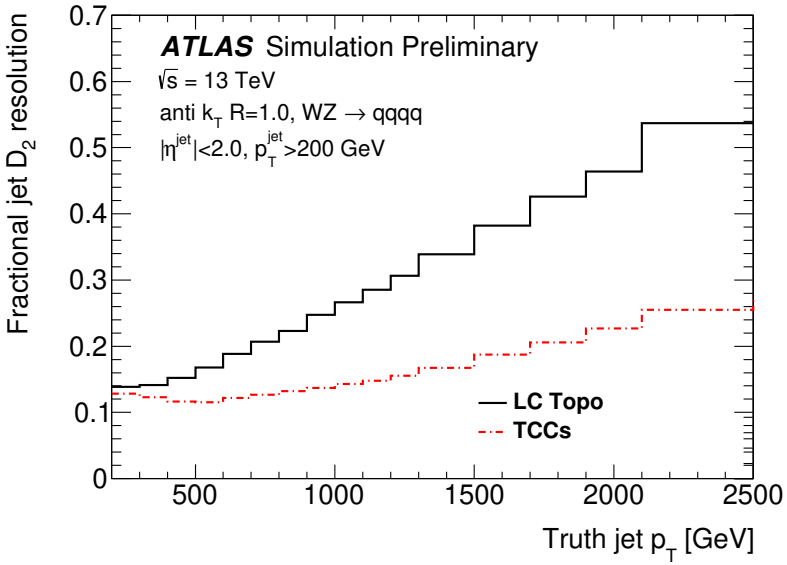



# Lower-level changes to inputs

Both: CONF-2018-016

- Combining track and calorimeter information is useful for substructure
  - Last slide was ways of exploiting tracks matched to jets
  - What about particle flow (direct track-cluster matches)?
- ATLAS is studying two types of particle flow for jet substructure
  1. "PFlow", tracks correct cluster 4-vectors (beneficial at "low" $p_{\mathrm{T}}$)
  2. "TCCs", using track angles and cluster scales (beneficial at "high" $p_{\mathrm{T}}$)
- Techniques aimed at different regimes, there will be a cross-over point
  - PFlow not yet public, TCC improves high-$p_{\mathrm{T}}$ mass and $\mathrm{D}_2^{\beta=1}$ resolution

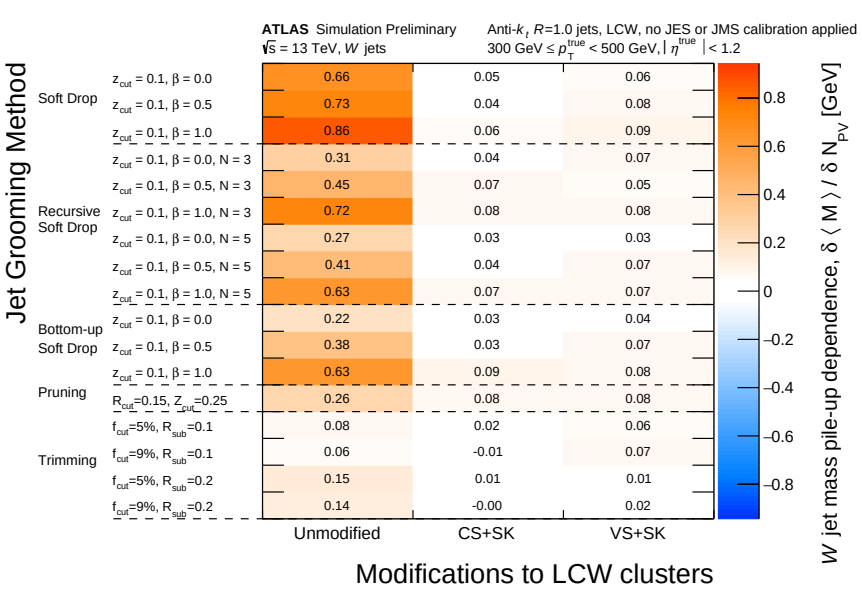


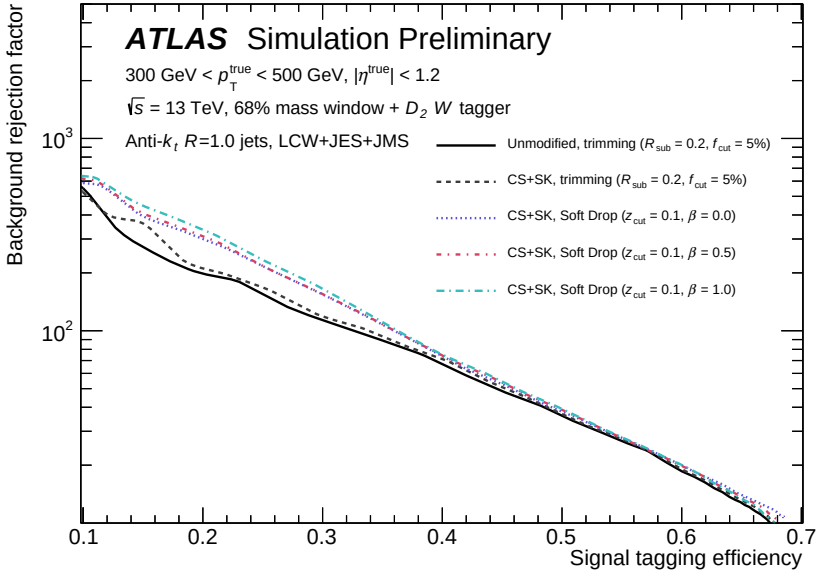



# Revisiting the jet definition

Both: JETM-2018-003

- Combining calo and track information is not the only way to improve
  - Constituent suppression methods reduce the impact of pileup
  - Large-R jet grooming parameters and types have an impact on tagging
  - Need to remove pileup and UE without over-suppressing hard scatter
- New jet definitions set to improve W/Z tagging performance
  - Low $p_{\mathrm{T}}$ and low signal efficiency sees most benefits, less at high $p_{\mathrm{T}}$
  - Stay tuned for comparisons including particle flow techniques



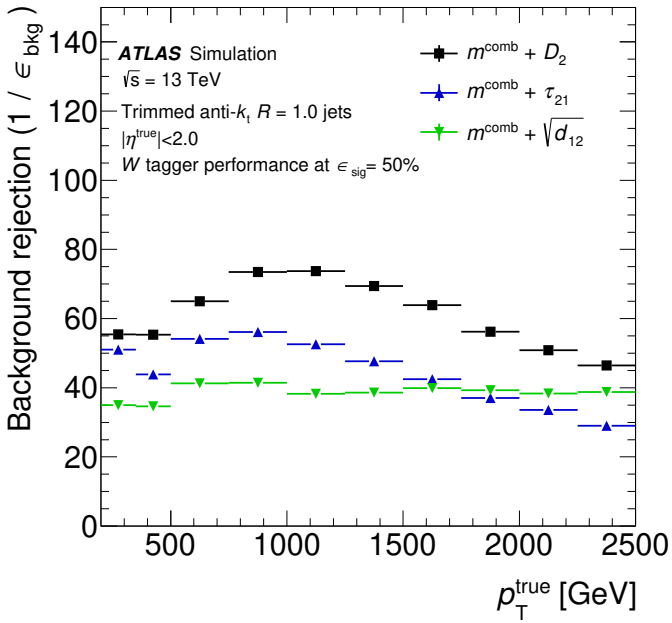



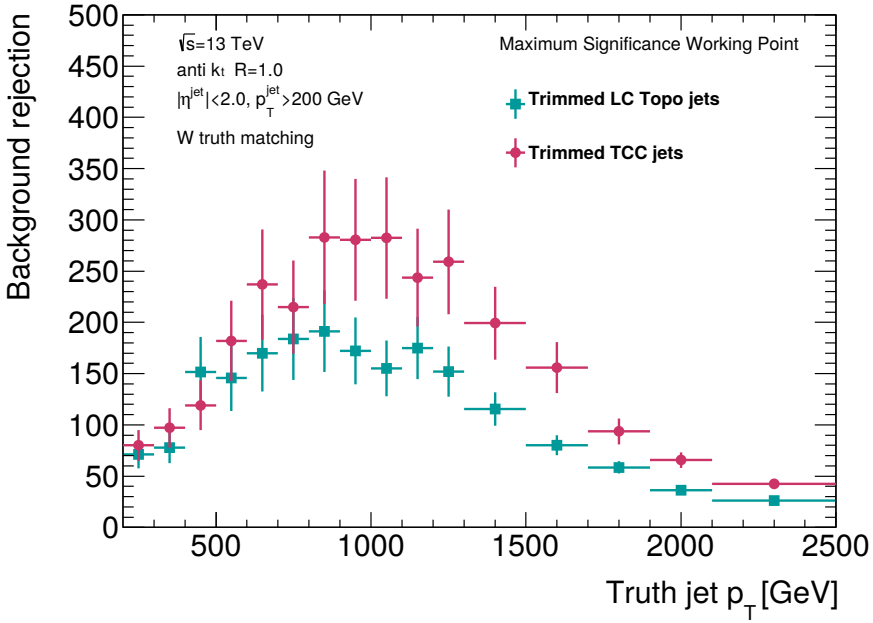
# Revisiting tagger definitions

Left: JETM-2018-03
Right: Thesis, N. Calace

Bourses postdoctorales Banting Postdoctoral Fellowships
UNIVERSITÉ DE GENÈVE

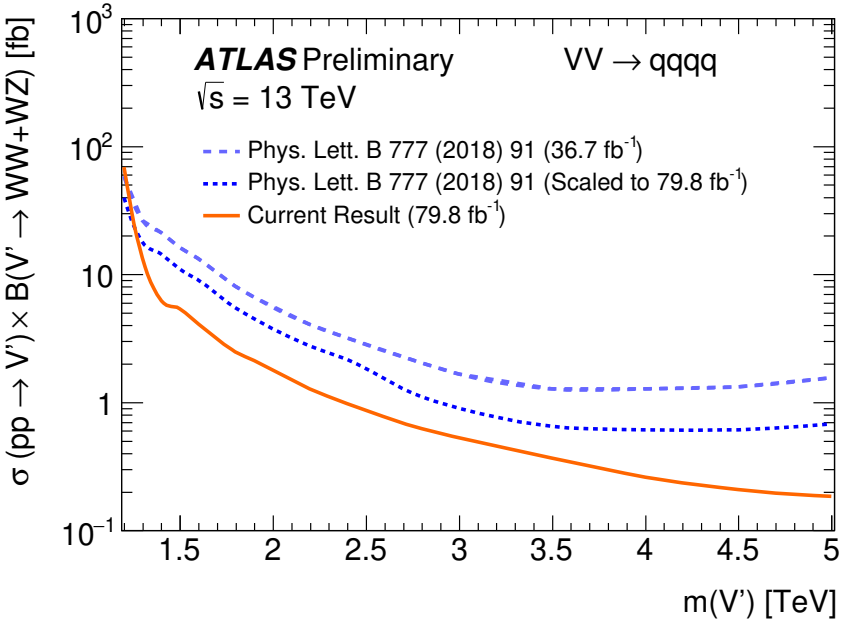
- ATLAS has primarily used flat-efficiency W/Z taggers
  - 50% W-efficiency using $m_{\text{jet}}^{\text{comb}}$ and $\mathrm{D}_2^{\beta=1}$ $\implies$ 50-80$\times$ QCD rejection
- Alternatively, consider maximum significance-based taggers
  - Significance $= \frac{\epsilon_{signal}}{a/2+\sqrt{B}}$, with $a = 3$ the number of $\sigma$ to optimize for
  - Tagger adapts to QCD expectation $B$ $\implies$ loosen cuts at high $p_{\mathrm{T}}$
  - Right plot: significance-based taggers using mass and $\mathrm{D}_2^{\beta=1}$



# New inputs and taggers in practice Plot: CONF-2018-016

- What is the real impact of these developments?
- Fully hadronic diboson resonance search used a new tagger
  - New inputs: TCCs (high-$p_{\mathrm{T}}$ particle flow) instead of topoclusters
  - New tagging strategy: significance-based instead of flat signal efficiency
- Result: hugely improved sensitivity to benchmark models
  - Gains after considering statistics: $2\times$ at low $p_{\mathrm{T}}$, $4\times$ at high $p_{\mathrm{T}}$
  - Inputs and tagging strategy both useful: gain is roughly evenly split

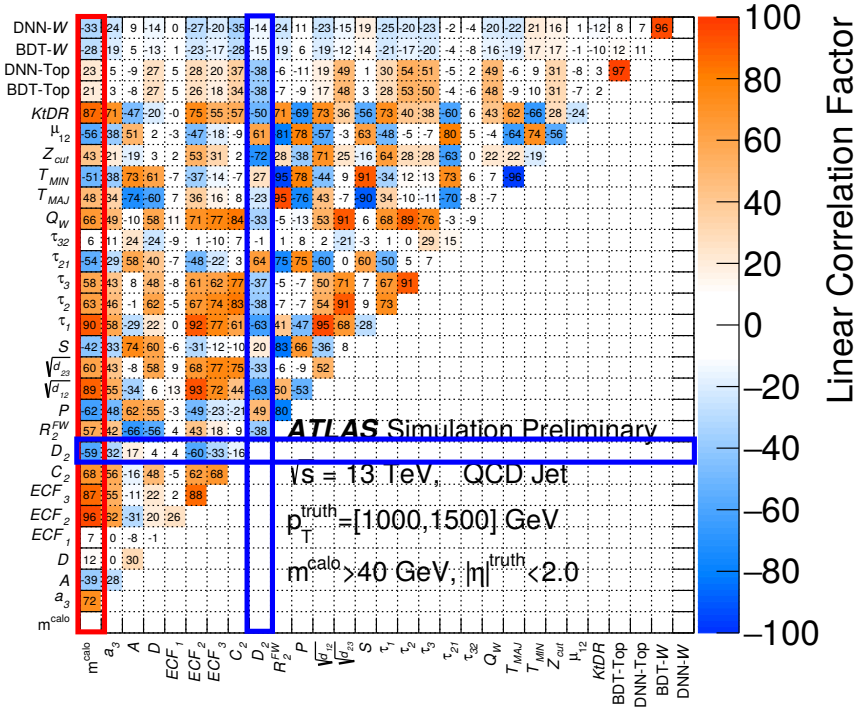


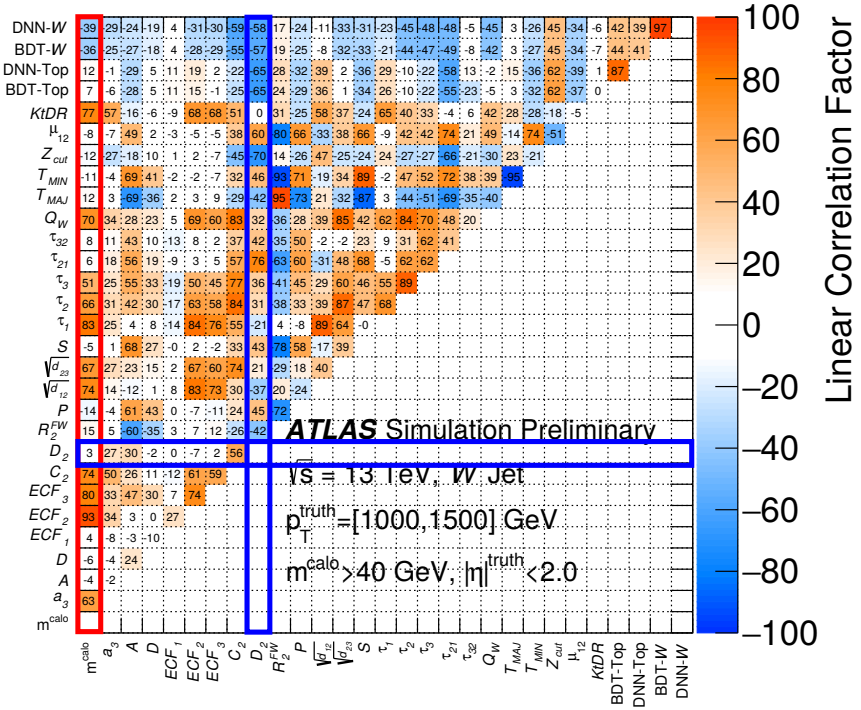
- Jet inputs and tagging strategies
- Using machine learning for W/Z-taggers
- Ensuring that we can use complex taggers
- Future machine learning strategies

# Exploiting correlations

Both: PUB-2017-004

- The mass + $D_2^{\beta=1}$ tagger does a good job at suppressing QCD
  - Natural question: can we do better with more varibles?
  - Equivalently, is there information not contained in mass+$D_2^{\beta=1}$?
    - Looking at the QCD (left) and W (right) correlations, it seems so!
- We can exploit these correlations with ML techniques
  - Start simple: train a classifier with substructure variables as inputs

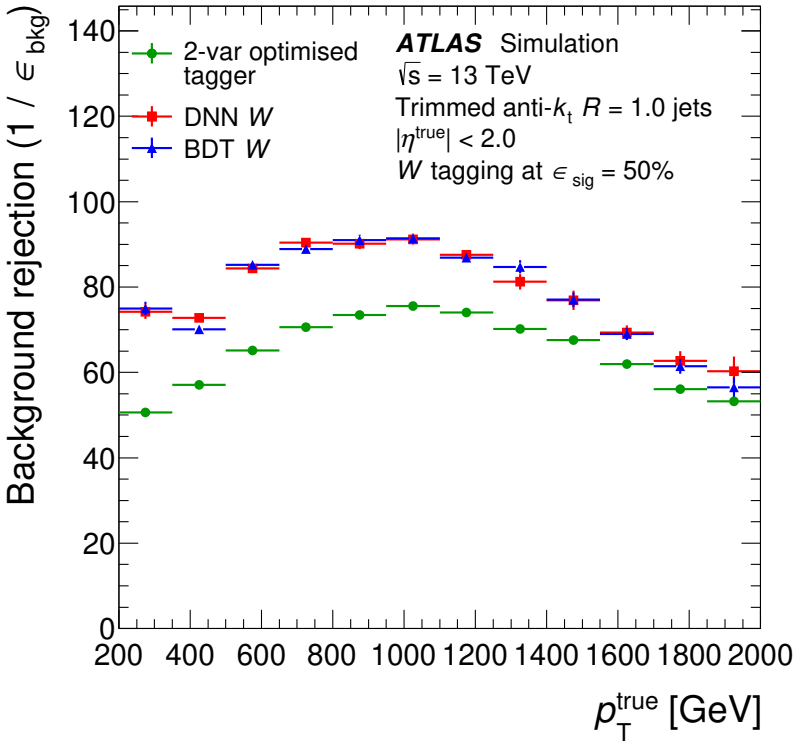


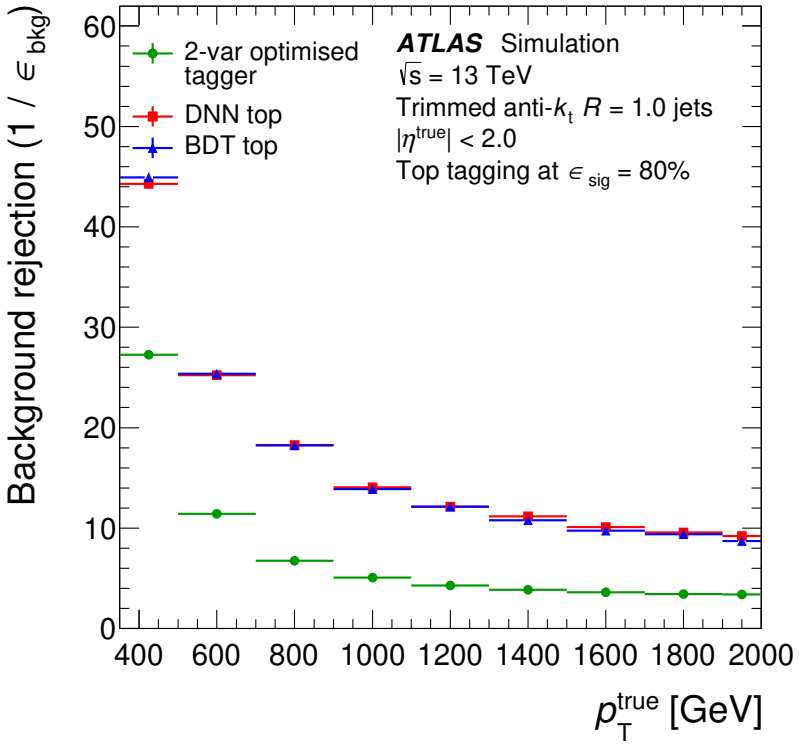



# ML using jet substructure variables

Both: JETM-2018-03

- Simple two-variable taggers can be improved using ML
- Build ML taggers from jet substructure variables
  - In this case, sufficiently complex BDTs and DNNs perform similarly
  - ∼30% improvements for W-tagging compared to two-variable tagger
  - Much larger improvements for top-taggers (right)



# Comparing W and top taggers

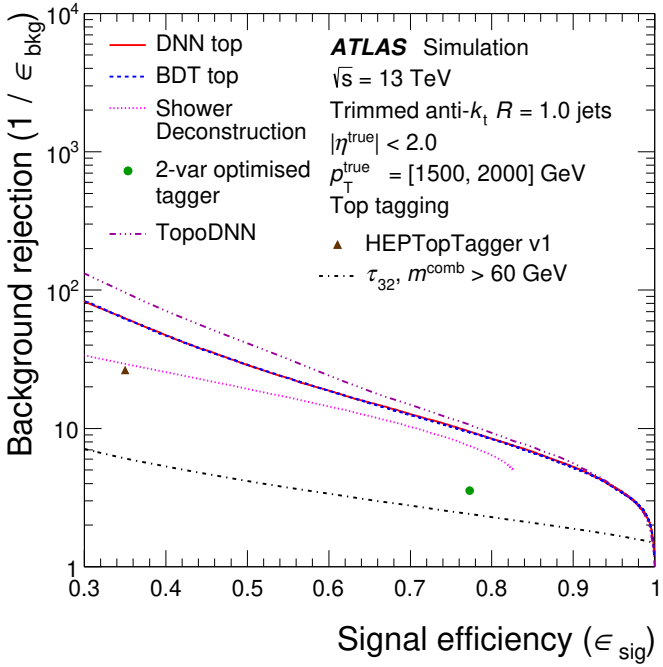
- Why does the W-tagger benefit much less from ML than top?
- Collimation: by $1\,\mathrm{TeV}$, W is already reaching calo granularity
  - Visible as the benefit of ML starts to reduce
  - New inputs exploiting tracking information would likely help
- W-tagging at 50% signal efficiency may not be optimal
  - Significance-based taggers would likely help
- Complexity: W is a less complex decay than top
  - Less correlations and structure to exploit

# Going further with low-level taggers

Both: JETM-2018-03

- Jet substructure variables are representations of the energy and spatial distributions of the jet constituents
  - Using them restricts ML to combinations of known representations
- Instead, let the ML classifier learn new representations
  - TopoDNN uses leading 10 topocluster 4-vectors for top-tagging
  - Also helps in cases with too few clusters to calculate substructure vars

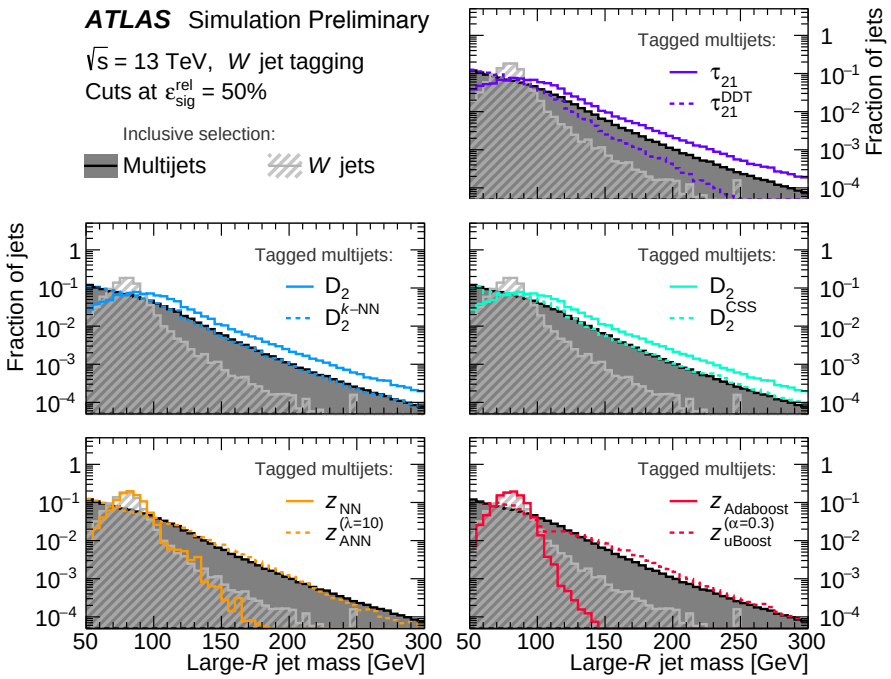
- ML language:
  - "Low-level" inputs = topocluster 4-vectors
  - "High-level" inputs = substructure variables
- Low-level expected to be similarly beneficial for W-tag

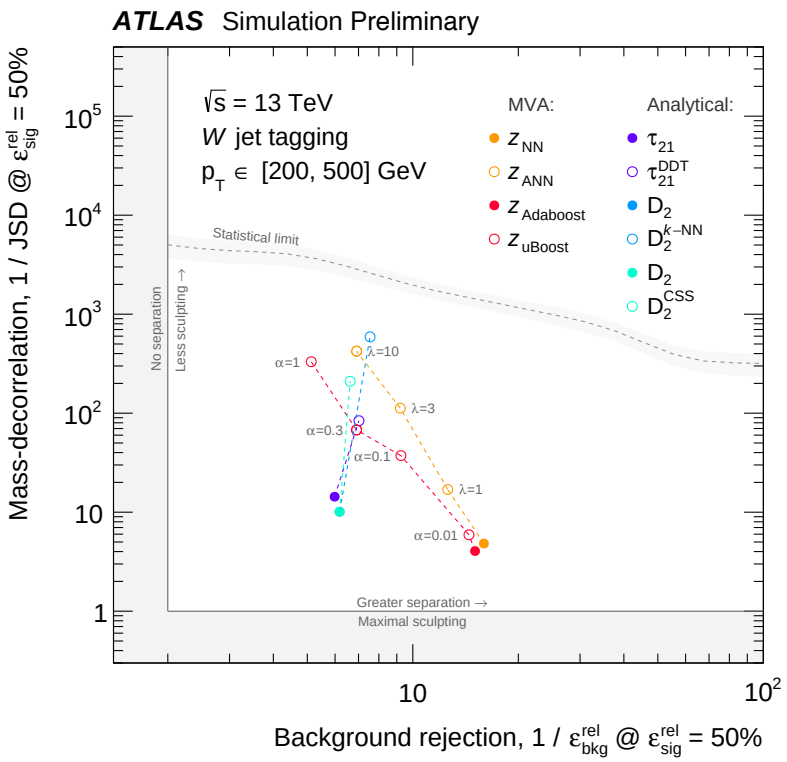


# Mass-decorrelated W-taggers

Both: PUB-2018-014

- Note that ∼all substructure information is correlated with jet mass
  - Cutting on any taggers, ML or not, therefore shapes the jet mass
- It is possible to develop mass-decorrelated taggers
  - Can be analytical or ML-based (such as adversarial neural networks)
  - Important for: mass-scan searches, mass measurements, etc
- Trade-off between ($x$ = QCD rejection) and ($y$ = mass decorrelation)

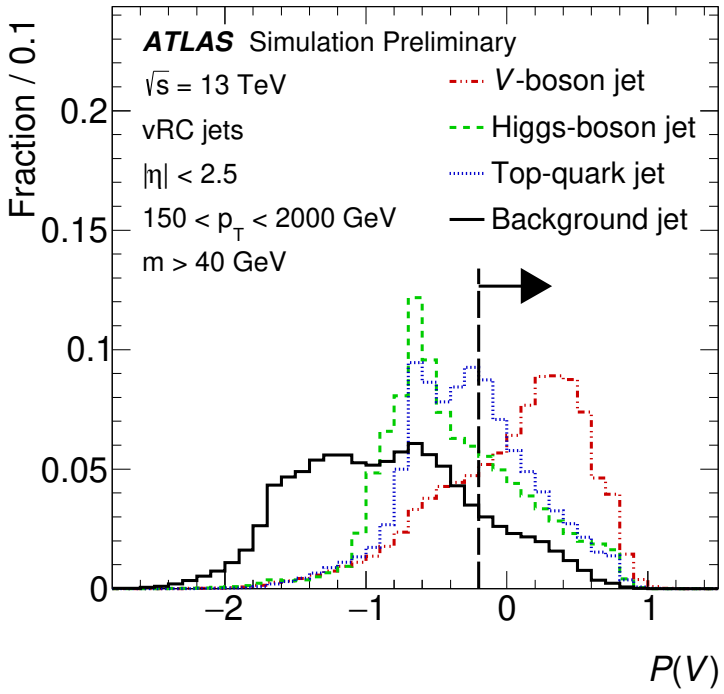


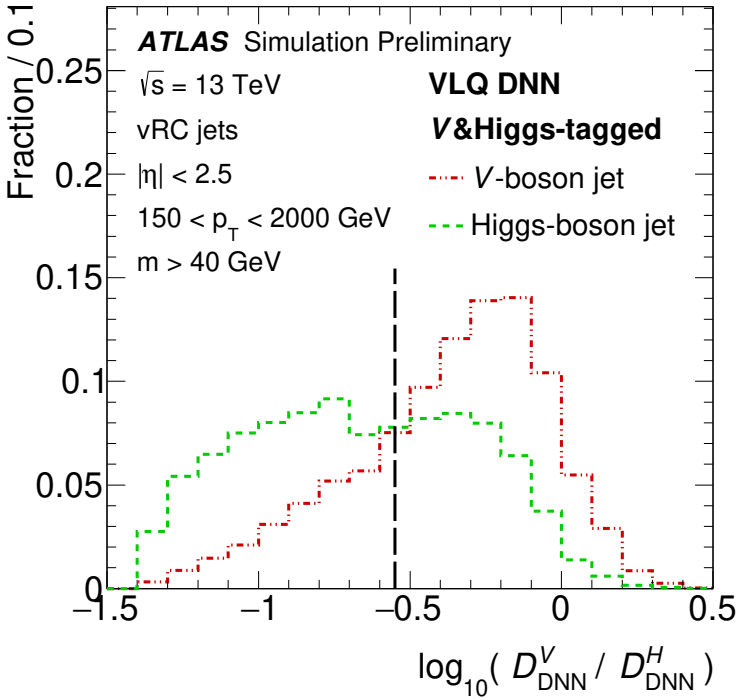



# Multi-class tagging

Both: EXOT-2017-14

- What if QCD is only one of your relevant backgrounds?
- Sometimes there are multiple similar objects to differentiate
  - Boosted W/Z, Higgs, and top jets have non-negligible overlap
- Multi-class DNN trained on a mix of low-level and high-level variables
  - First stage: discriminate vs QCD for each signal type independently
  - Second step: likelihoods of discriminants for signal ambiguity resolution



- Jet inputs and tagging strategies
- Using machine learning for W/Z-taggers
- Ensuring that we can use complex taggers
- Future machine learning strategies

# Tagger efficiencies

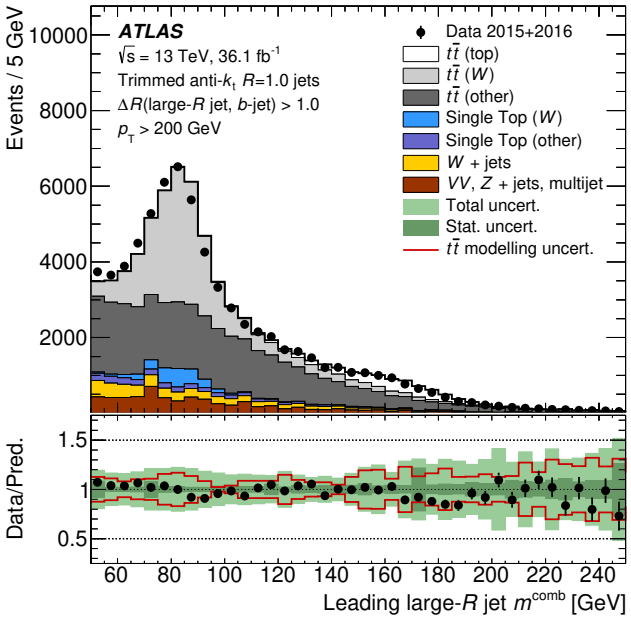
- In ATLAS, taggers have historically been simple
  - Independent, uncorrelated cuts on a small number of variables
  - Based purely on calorimeter quantities
  - Result: correspondingly simple uncertainty procedures were fine
- Now, we are entering a new age of complex jet taggers
  - ML taggers are exploiting correlations between variables
    - These correlations may be real, or may be MC "features"
  - Both calorimeter and tracking information is used
  - Result: need a more robust means of deriving uncertainties
- Derive tagging efficiency scale factors and uncertainties *in situ*
  - Use "standard candles" to directly evaluate data/MC differences

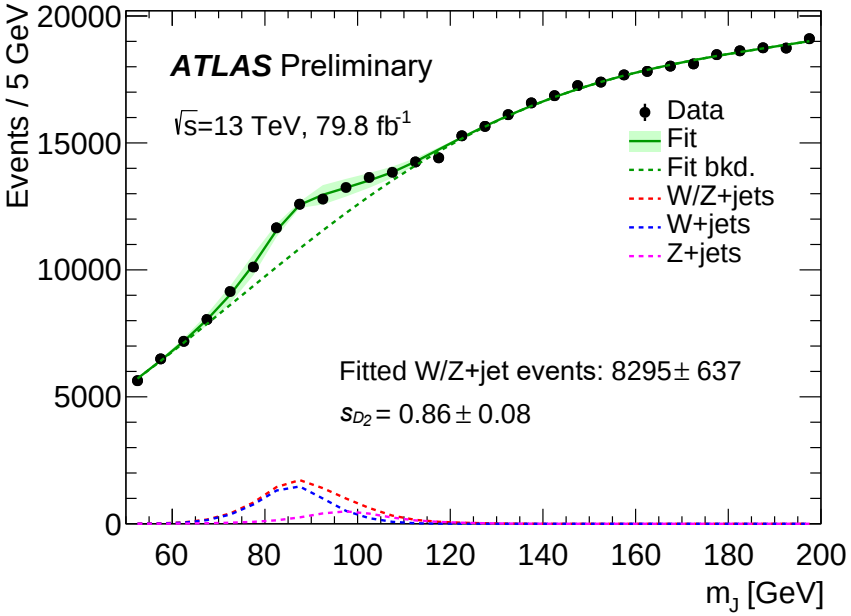
# Signal tagging efficiencies

Left: JETM-2018-03
Right: CONF-2018-016

Bourses postdoctorales Banting Postdoctoral Fellowships
UNIVERSITÉ DE GENÈVE

- Two examples of "standard candles" for W-tagging
  - Left: low-$p_{\mathrm{T}}$ semi-leptonic $t\bar{t}$ events
  - Right: higher-$p_{\mathrm{T}}$ V+jets events
- Fitting templates to the peaks allows for extracting W-jet efficiency
  - Then compare data and MC values, and determine uncertainties

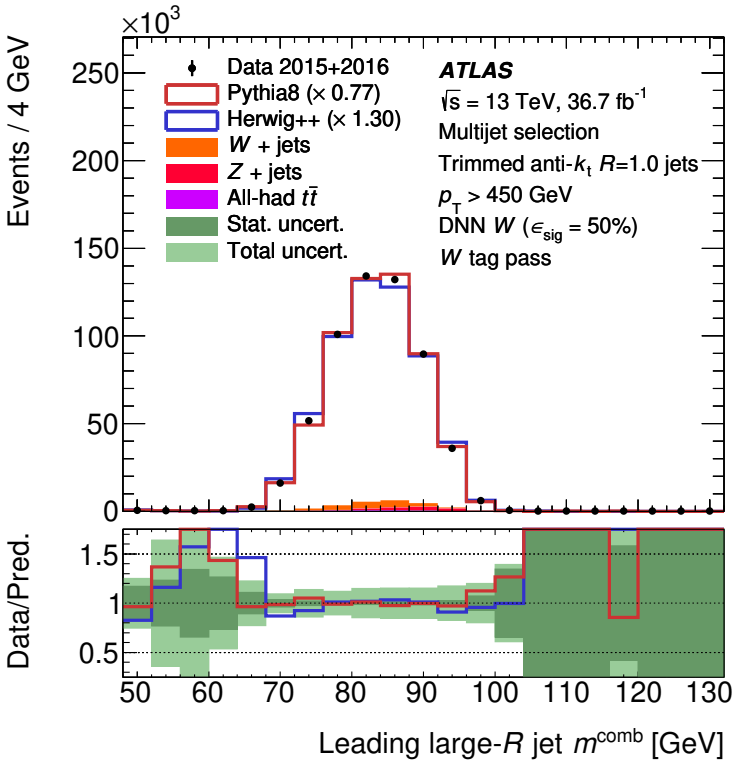


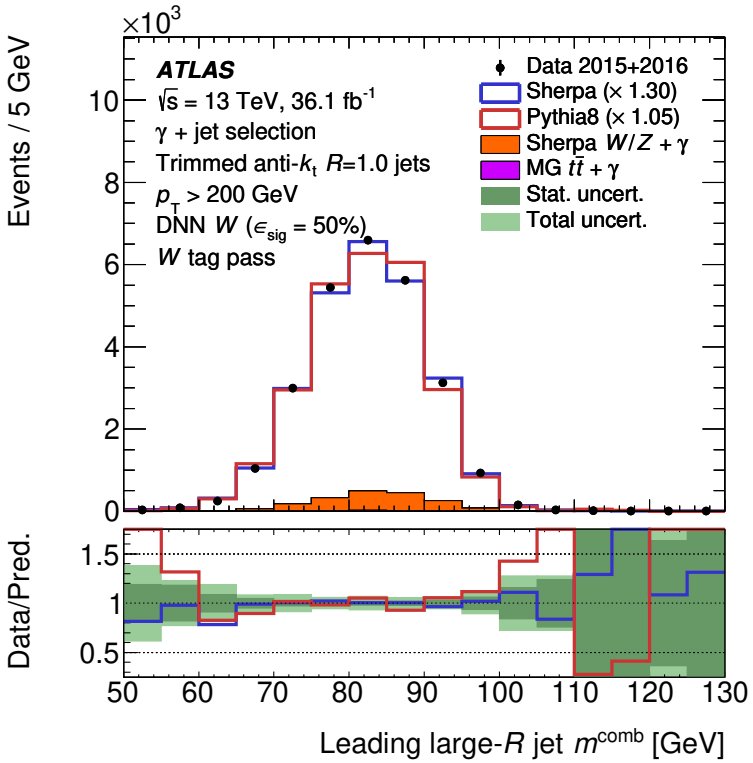



# Background tagging efficiencies

Both: JETM-2018-03

- Tagging efficiencies also have to be evaluated in background samples
  - Left: QCD multijet events, after requiring a W-tag
  - Right: $\gamma$+jets events, after requiring a W-tag
- The taggers clearly sculpt the background
  - Can be mass-decorrelated, at the cost of background rejection

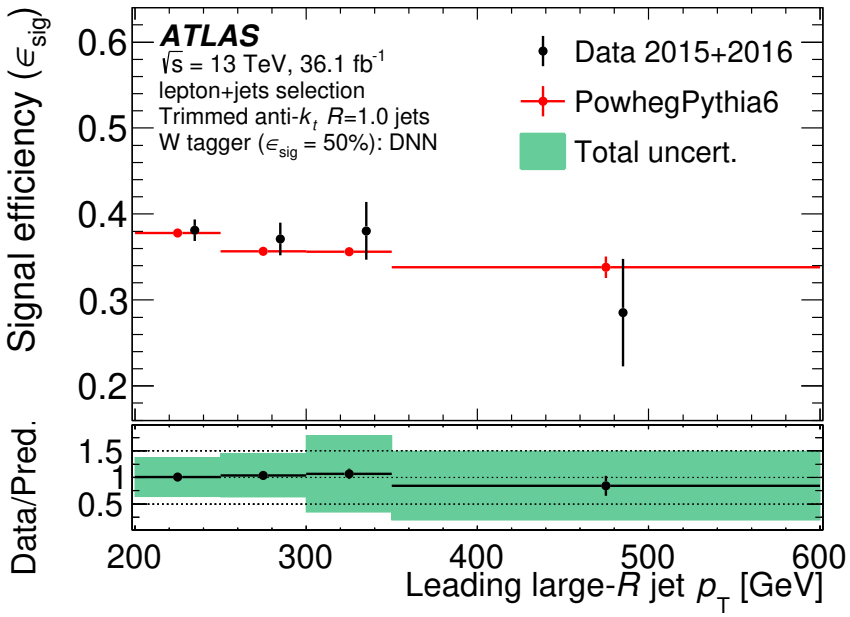


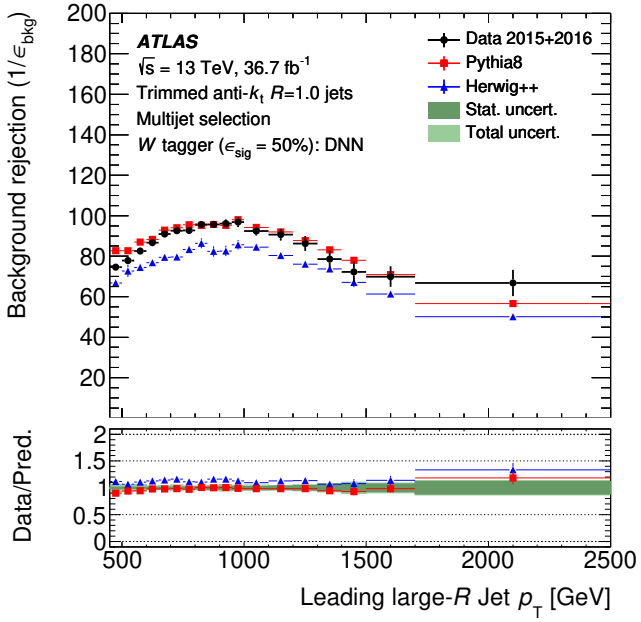



# Tagging efficiency scale factors

Both: JETM-2018-03

- Resulting signal (left) and background (right) efficiencies are shown
- With this done, we now have a handle on data/MC differences
  - Performed in specific kinematic regions
  - Some extrapolations are needed to cover full parameter space



- Jet inputs and tagging strategies
- Using machine learning for W/Z-taggers
- Ensuring that we can use complex taggers
- Future machine learning strategies

# More advanced network architectures



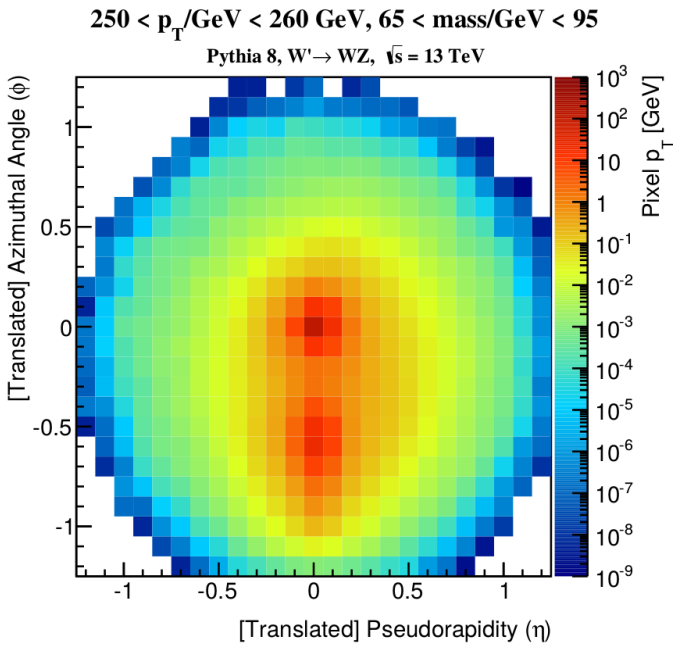
- As shown before, using low-level information is powerful
  - TopoDNN performed better than a DNN based on substructure vars
  - However, such DNNs are only one type of neural network
- Convolutional Neural Networks (CNNs):
  - Create pixelated image-based representations of jets
  - Commonly used in image recognition applications
- Recurrent Neural Networks (RNNs):
  - Networks that have a form of "memory" about what they have seen
  - Good for learning about direct correlations between subsequent inputs
  - Commonly used in language processing: a word depends on past words

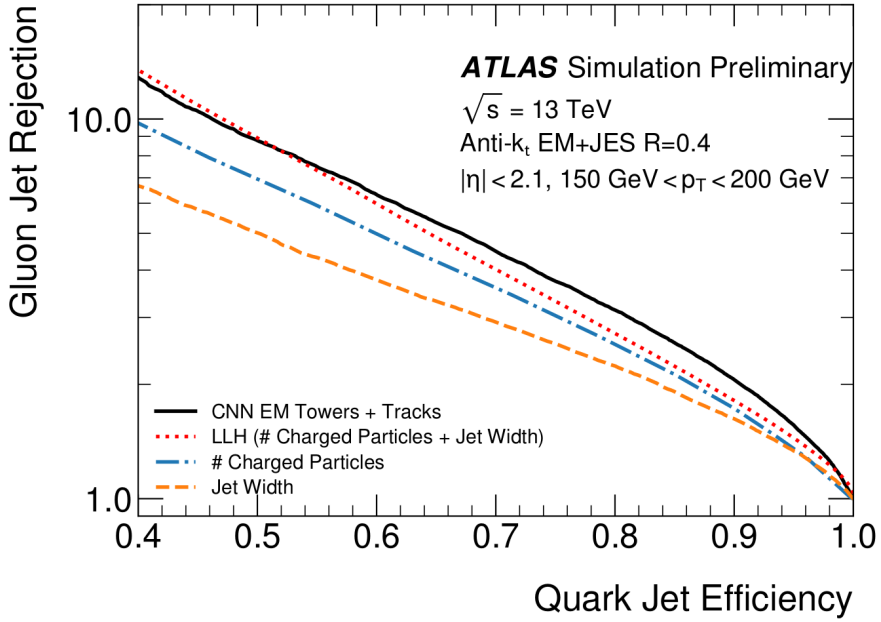
# CNNs

Left: arXiv:1511.05190

Right: PUB-2017-017

- Jet images can be used to identify a given particle type
  - Normally pre-processing is applied (translation, rotation, inversion, etc)
- Shown in 2015 phenomenological paper for W-tagging (left)
  - Nice and "intuitive", expect to see a second energy maximum for W
- So far ATLAS has only tried CNNs for q/g tagging (right)
  - q/g tagging is hard; W-tagging will likely benefit more

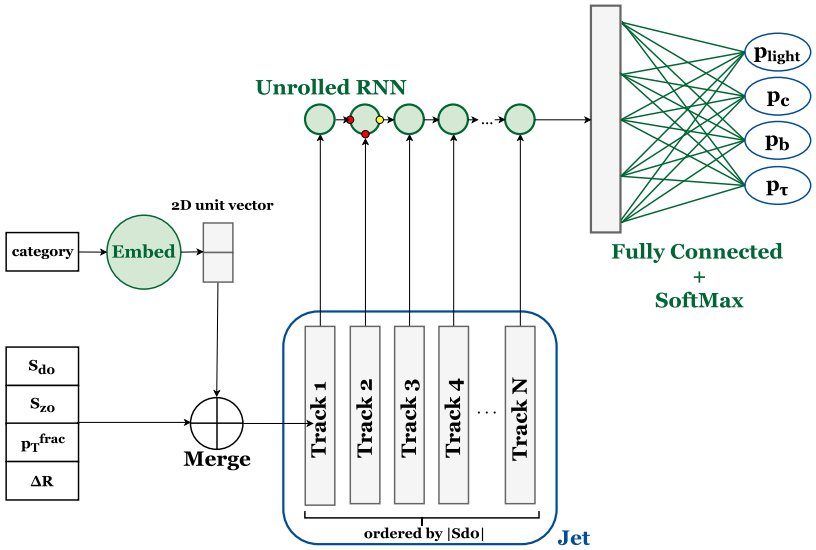


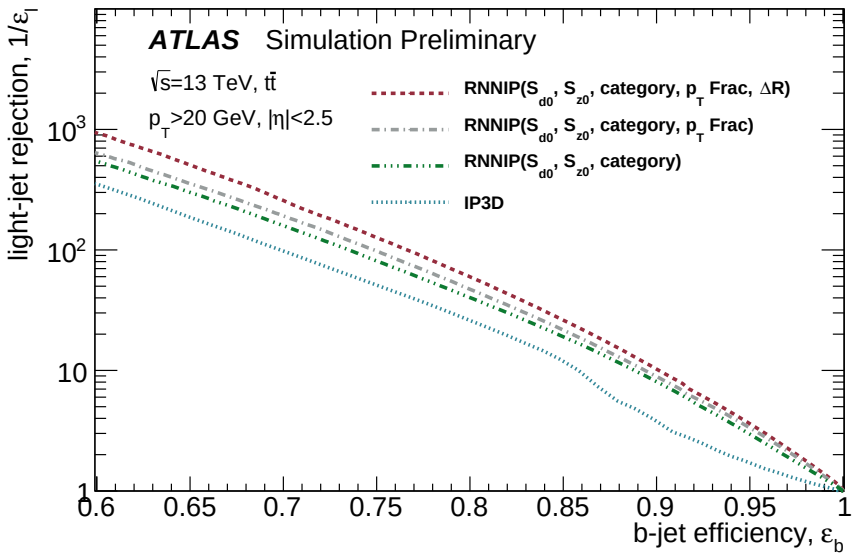



# RNNs

Both: PUB-2017-003

- ATLAS has developed a b-tagging RNN
  - RNN "memory" used to exploit correlations between ordered tracks
  - Good chance of secondary vertex if a few have large d0 values
- Similar could be done with W-tagging
  - Energy correlation functions (ECFs) are powerful W-tagging variables
  - RNN can further exploit correlations between jet constituents



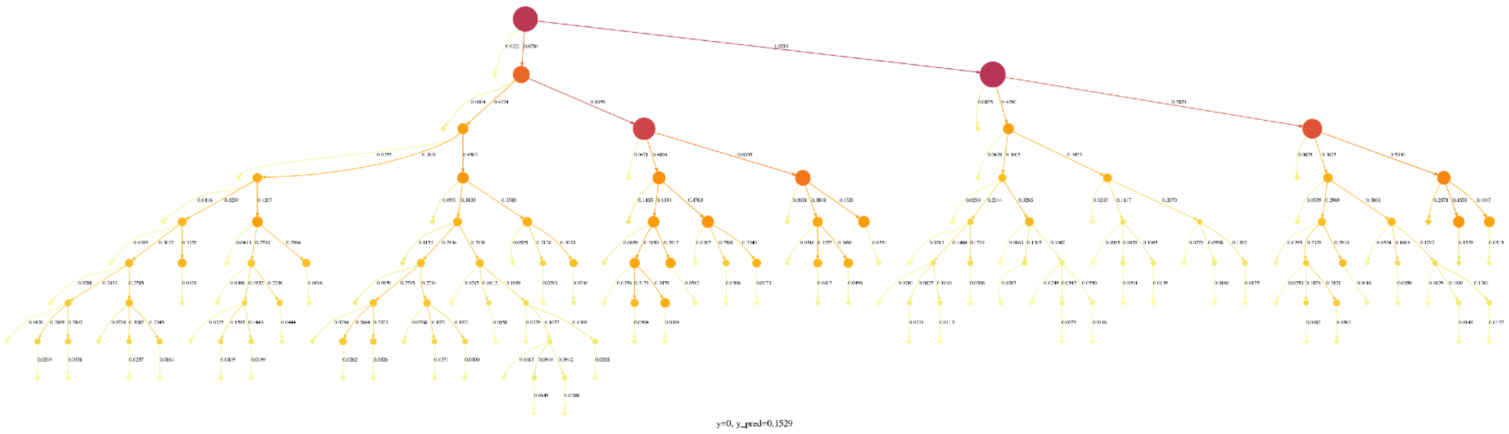



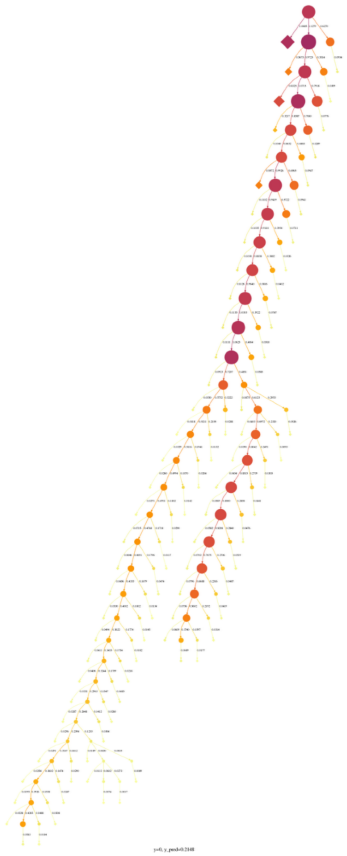
# RNN tree representation

Paper: arXiv:1702.00748
Figures: CERN seminar

Bourses postdoctorales Banting Postdoctoral Fellowships
UNIVERSITÉ DE GENÈVE

- Idea of considering jets as clustering trees using RNNs
  - Analogy: word = particle, parsing = jet algorithm
- Two representations of W-jet events
  - The choice of "algorithm" matters!
  - Left is using the $k_t$ metric, right is anti-$k_t$
  - $k_t$ metric gives better performance in this study
- Large gains over image-based approach at truth level
  - Projecting into pixels discards information





# Summary

- Huge progress has recently been made in hadronic W/Z identification
  - Improved inputs to jet reconstruction
  - Scans over large numbers of jet grooming types
  - Revised optimization metrics based on tagging significance
  - Increasingly complex ML taggers
  - Possibility of deriving mass-decorrelated taggers where needed
  - Increasing studies into the use of multi-class taggers
  - Usage of "standard candles" to evaluate taggers in data
- The ML community continues to evolve at a rapid rate
  - Several new ideas that appear promising for use in W/Z tagging
- Remember that ML can only exploit the information that it is given
  - Any conclusions drawn on the utility of a given approach may change after improving the inputs to jet reconstruction and/or the jet definition

$\Longrightarrow$ **Important to optimize all stages, not only the ML algorithm!**

# Backup Material

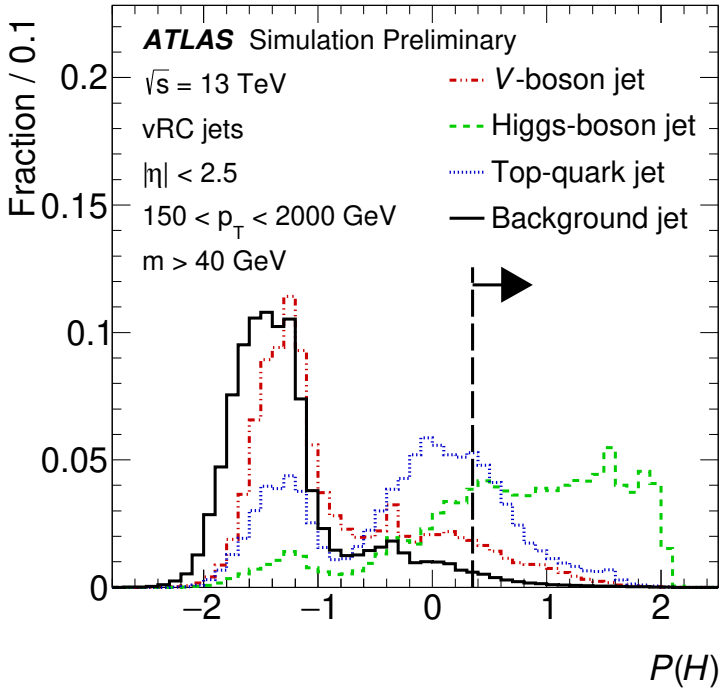
# Multi-class tagging, step 1

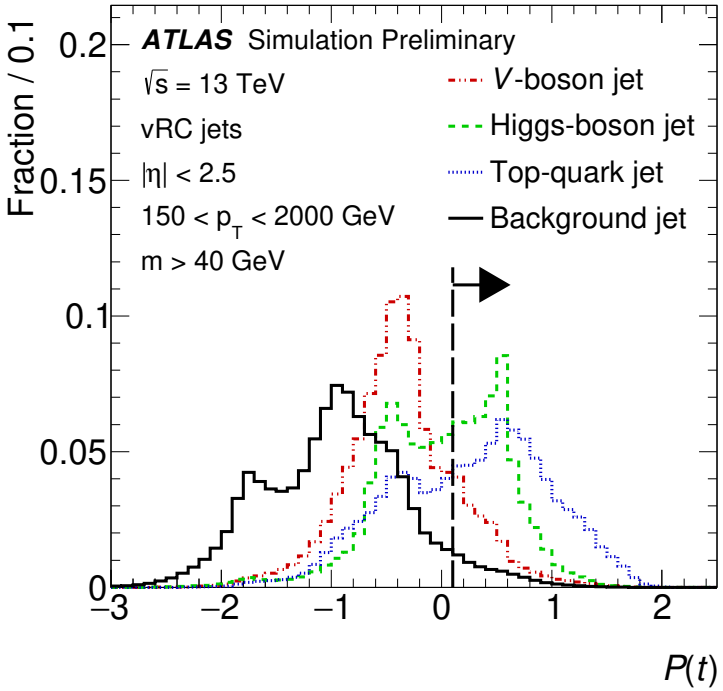
Both: EXOT-2017-14

$$P(V) = \log_{10}\left(\frac{D^{V}_{\mathrm{DNN}}}{0.9 \cdot D^{\mathrm{background}}_{\mathrm{DNN}} + 0.05 \cdot D^{t}_{\mathrm{DNN}} + 0.05 \cdot D^{H}_{\mathrm{DNN}}}\right)$$

$$P(H) = \log_{10}\left(\frac{D^{H}_{\mathrm{DNN}}}{0.9 \cdot D^{\mathrm{background}}_{\mathrm{DNN}} + 0.05 \cdot D^{V}_{\mathrm{DNN}} + 0.05 \cdot D^{t}_{\mathrm{DNN}}}\right)$$

$$P(V) = \log_{10}\left(\frac{D^{t}_{\mathrm{DNN}}}{0.9 \cdot D^{\mathrm{background}}_{\mathrm{DNN}} + 0.05 \cdot D^{V}_{\mathrm{DNN}} + 0.05 \cdot D^{H}_{\mathrm{DNN}}}\right)$$



# Multi-class tagging, step 2

Both: EXOT-2017-14

$$P(V) = \log_{10}\left(\frac{D^{V}_{\mathrm{DNN}}}{0.9 \cdot D^{\mathrm{background}}_{\mathrm{DNN}} + 0.05 \cdot D^{t}_{\mathrm{DNN}} + 0.05 \cdot D^{H}_{\mathrm{DNN}}}\right)$$

$$P(H) = \log_{10}\left(\frac{D^{H}_{\mathrm{DNN}}}{0.9 \cdot D^{\mathrm{background}}_{\mathrm{DNN}} + 0.05 \cdot D^{V}_{\mathrm{DNN}} + 0.05 \cdot D^{t}_{\mathrm{DNN}}}\right)$$

$$P(V) = \log_{10}\left(\frac{D^{t}_{\mathrm{DNN}}}{0.9 \cdot D^{\mathrm{background}}_{\mathrm{DNN}} + 0.05 \cdot D^{V}_{\mathrm{DNN}} + 0.05 \cdot D^{H}_{\mathrm{DNN}}}\right)$$