Jet substructure and H/V/top-tagging

Danilo Ferreira de Lima[a] & **Simone Marzani**[b]
[a] *Ruprecht-Karls-Universität Heidelberg*
[b] *Università di Genova and INFN Sezione di Genova.*

[a] on behalf of the *ATLAS* and *CMS* collaborations

QCD@LHC 2018
Dresden, 30th August 2018

Outline

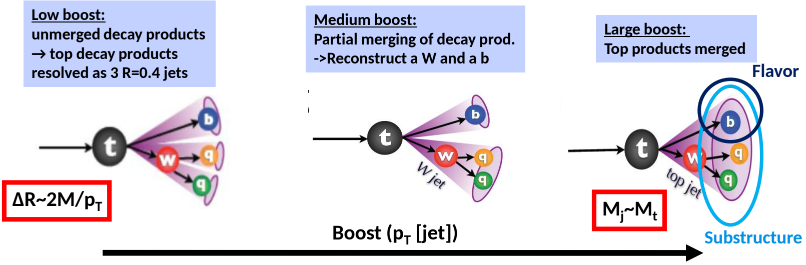
1. Jet substructure as a tool
   - Grooming and tagging algorithms
   - Efficiency and performance
2. Jet substructure for precision
   - Precision calculations meet the data
   - Towards the extraction of Standard Model parameters
3. New ideas and techniques
   - News from BOOST 2018
4. Summary

Jet substructure as a tool

Grooming and tagging algorithms

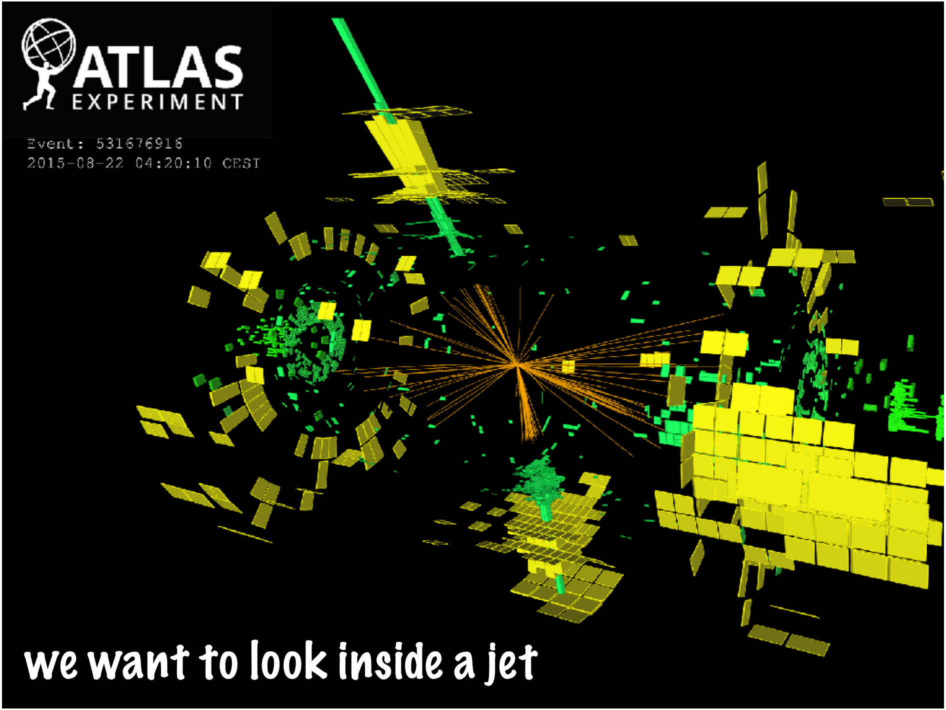
EW-scale particles get a boost

- Standard analysis: the heavy particle X decays into two partons, reconstructed as two jets.
- Search strategy: look for bumps in the dijet invariant mass distribution.
- What about EW-scale particles at the LHC?

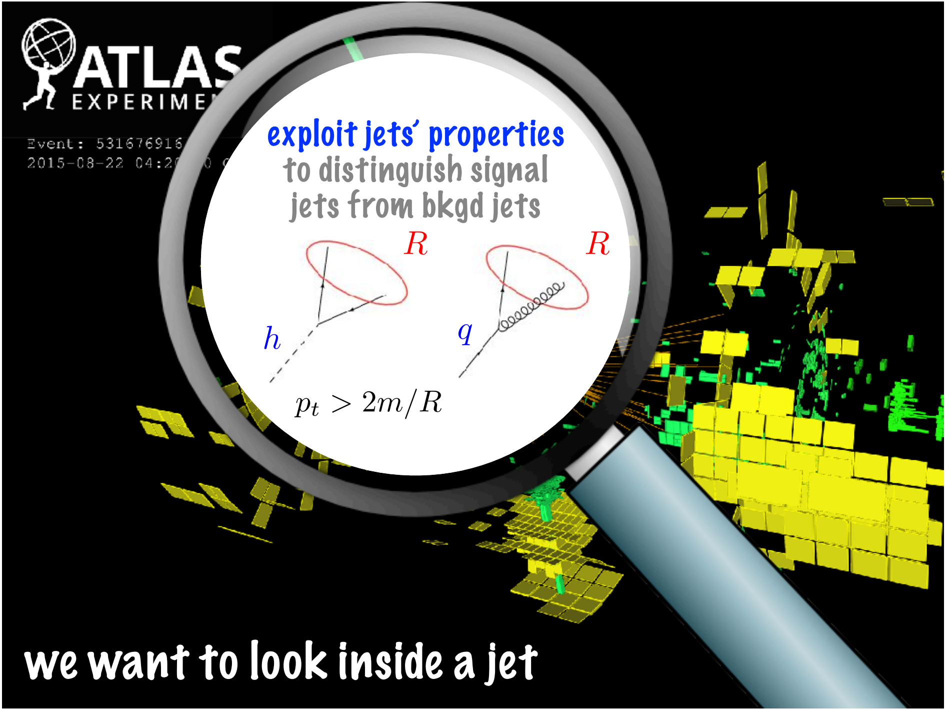


(Figure by L. Gouskos)

- $E_{\text{c.o.m.}} \gg m_{\text{EW}}$, hence they are abundantly produced with a large boost.
- Their decay-products are then collimated and, if they decay into hadrons, we end up with localised deposition of energy in the hadronic calorimeter: a jet.


ATLAS
EXPERIMENT
Event: 531676916
2015-08-22 04:20:10 CEST
we want to look inside a jet

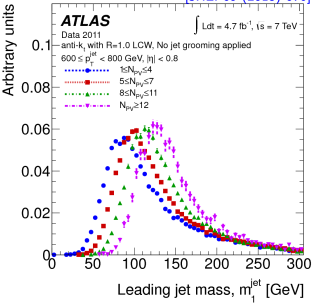

ATLAS
EXPERIME
Event: 531676916
2015-08-22 04:2 0 CEST
p_t
we want to look inside a jet


ATLAS
Event: 531676916
exploit jets' properties
to distinguish signal
jets from bkgd jets
R
R
h
q
$p_t > 2m/R$
we want to look inside a jet

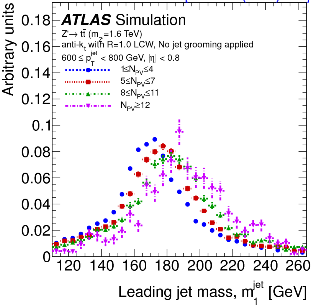
The jet invariant mass

- Signal jets have an intrinsic energy scale: the mass of the decaying object ($m_{\text{jet}} \sim m_{\text{EW}}$).
- Background (QCD) jets acquire their mass through parton branching ($m_{\text{jet}} \propto p_T$)
- Cut on the jet mass to separate signal and background!
- This simple observation still remains at the core of every substructure analysis, however it is not enough.

[JHEP09 (2013) 076]



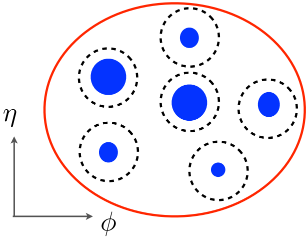
[JHEP09 (2013) 076]

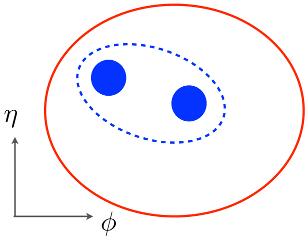


- Non-perturbative contributions (hadronisation, underlying event and pile-up) pollute and deform our jets.

Grooming & Tagging

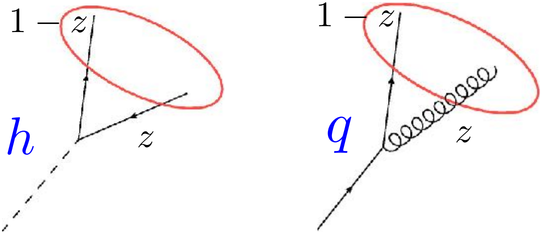
- We need to go beyond the mass and exploit jet substructure:
- Two key principle (often combined in actual algorithms):
  - grooming: clean the jets up by removing soft radiation;
  - tagging: identify the features of hard decays and cut on them.
- Core idea for grooming
  1. identify the "right" angular scale for a jet
  2. throw away what is soft and at large angle





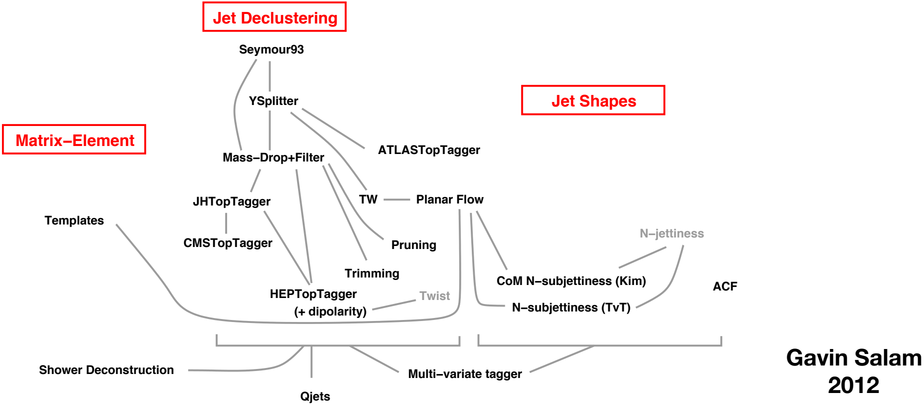
Grooming & Tagging

- We need to go beyond the mass and exploit jet substructure:
- Two key principle (often combined in actual algorithms):
  - grooming: clean the jets up by removing soft radiation;
  - tagging: identify the features of hard decays and cut on them.
- Core idea for (2-prong) tagging
  1. H/Z/W characterised by symmetric energy sharing: $P_{h\to q\bar{q}} = 1$
  2. QCD splitting is enhanced in the soft limit $P_{gq} = C_F \frac{1+(1-z)^2}{z}$



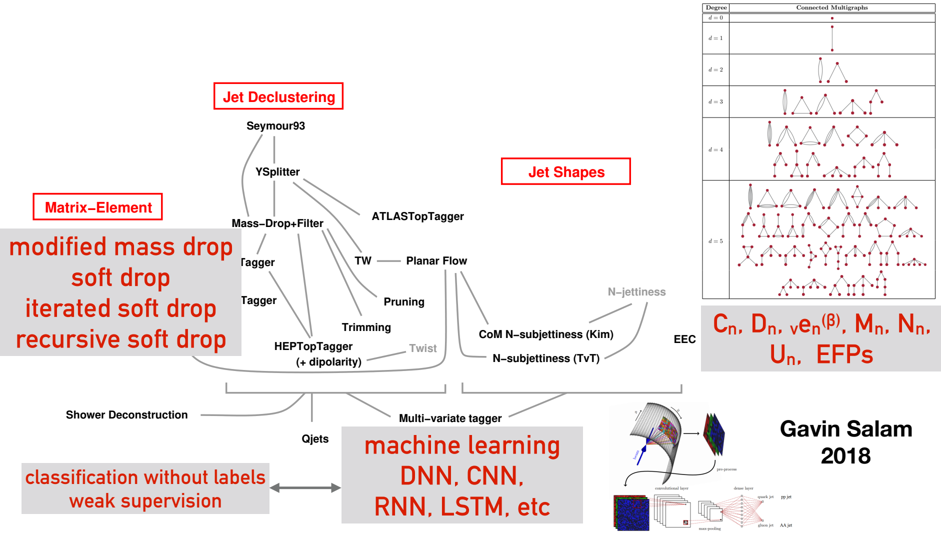
$\min(z, 1-z) > z_c$

An expanding universe of jet substructure techniques



- How well do we understand them?

An expanding universe of jet substructure techniques



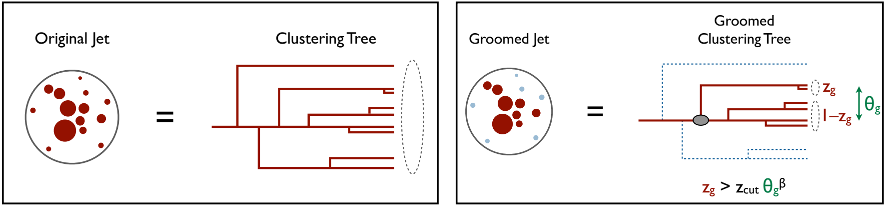
- How well do we understand them?

Example of a groomer: soft drop

- The algorithm starts from a C/A clustered jet and proceeds as follows:

[Larkoski, SM, Soyez, Thaler (2013)]

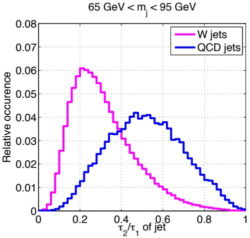
1. Undo the last stage of the clustering. Label the two subjets j_1 and j_2.
2. If $\frac{\min(p_{T1},p_{T2})}{(p_{T1}+p_{T2})} > z_c \left(\frac{\Delta_{12}}{R}\right)^{\beta}$ then deem j to be the soft-drop jet.
3. Otherwise redefine j to be the harder subjet and iterate



- Groomer built to be both efficient and robust.
- Amenable for theoretical calculations (see later).

Examples of two-prong taggers (H/Z/W)

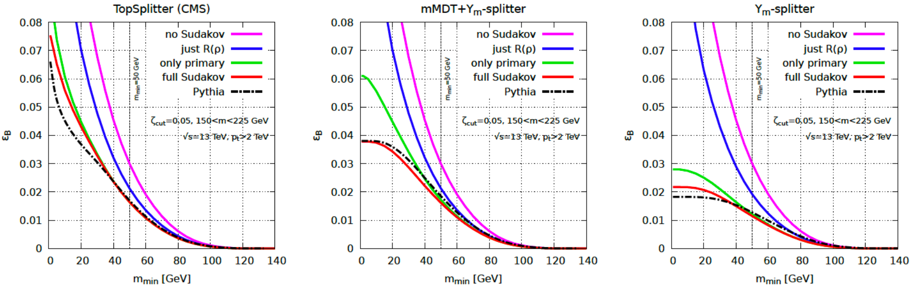
- General strategy:
  1. groom to remove contamination;
  2. select the mass window of interest;
  3. use a jet shape to determine the prong structure.
- *N*-subjettiness is prototype [Thaler and Van Tilburg (2011)].
  - Often used in the experiments: $\tau_{32} = \frac{\tau_3}{\tau_2}$ and $\tau_{21} = \frac{\tau_2}{\tau_1}$.
- Ratios of energy correlation functions also offer a powerful way of discriminating the prong structure [Larkoski, Salam, Thaler (2013)].
- Phase-space analysis can guide us in building the optimal ratio: D_2 [Larkoski, Moult, Neill (2014)].



- Grooming modifies the emission phase-space: shapes that work well on ungroomed jet may loose their discrimination power.
  - design correlation functions that work specifically on groomed jets M_2. [Moult, Necib, Thaler (2016)]
  - dichroic ratios, e.g. $\tau_{21} = \frac{\tau_2^{\text{SD}\,\beta>0}}{\tau_1^{\text{SD}\,\beta=0}}$. [Salam, Schunk, Soyez (2016)]

Top tagging

- Top-taggers try to discriminate the three-pronged nature of a top-initiated jet from the background.
- Use shapes and correlation functions that are sensitive to three prongs, e.g. τ_{32}.
- Actual algorithms used by the collaboration can involve many steps (CMS Top Tagger, HEP Top Tagger, HOVRT...).
- Gaining analytic insight is more difficult because of the more complicated phase-space, but recently a breakthrough [Dasgupta, Guzzi, Rawling, Soyez (2018)].

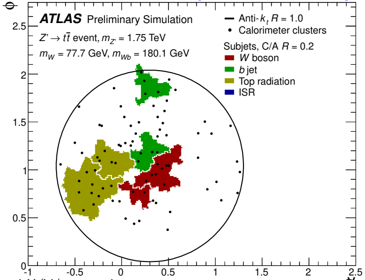


A Matrix-Element-Method-inspired solution: Shower Deconstruction

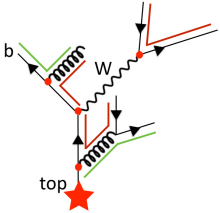
- Takes the probability for signal and background jets $\rightarrow$ defines a likelihood-ratio discriminant χ_{SD} [1102.3480] [1211.3140]
- Recluster jet in small sub-jets and use them as top decay partons $\rightarrow$ test compatibility with a top/H shower.
- Best tagging performance within non-MVA-based solutions $\rightarrow$ using a theoretically motivated approach.

$$\chi_{SD} = \frac{\sum_{\text{histories}} P(p_i|\text{signal})}{\sum_{\text{perm.}} P(p_i|\text{background})}$$

[ATLAS-CONF-2014-003]



[ATLAS-CONF-2014-003]

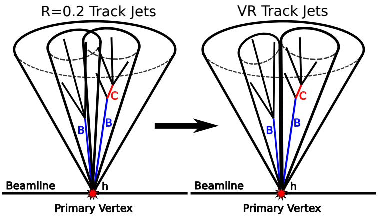


Jet substructure as a tool

Efficiency and performance

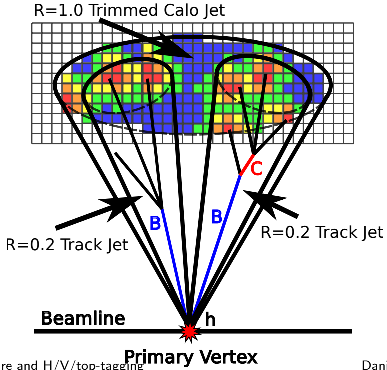
$X \to b\bar{b}$ methods in ATLAS

[ATL-PHYS-PUB-2017-010]

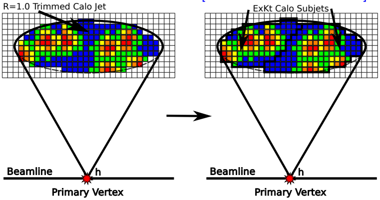


- ATLAS studied using $R = 0.2$ track-jets, variable-R track jets and exclusive k_T subjets for the $X \to b\bar{b}$.

[ATL-PHYS-PUB-2017-010]



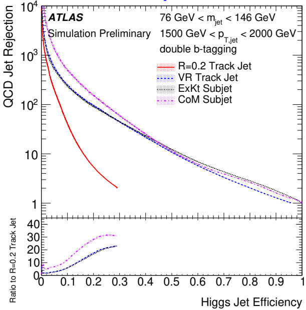
[ATL-PHYS-PUB-2017-010]



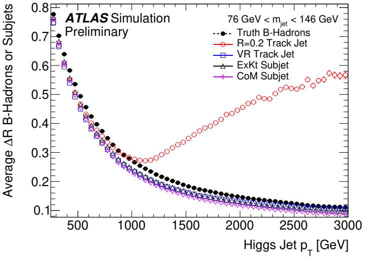
$X \to b\bar{b}$ methods in ATLAS – comparison

- Variable-R jets and exclusive k_T sub-jets show an improvement in performance.
- $R = 0.2$ track jets fail to resolve the Higgs decay products as the p_T become higher than ~ 1 TeV.
- Variable-R sub-jets lead to high rejection and reconstruct the direction of the sub-jets very well, relative to the b-hadron.

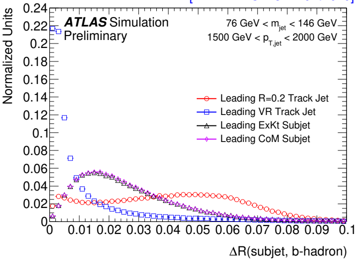
[ATL-PHYS-PUB-2017-010]



[ATL-PHYS-PUB-2017-010]

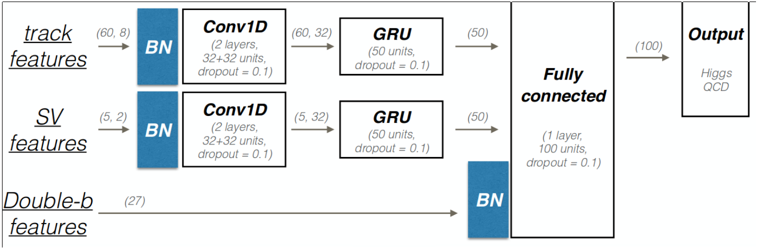


[ATL-PHYS-PUB-2017-010]



$h \to b\bar{b}$ and $h \to c\bar{c}$ identification in CMS

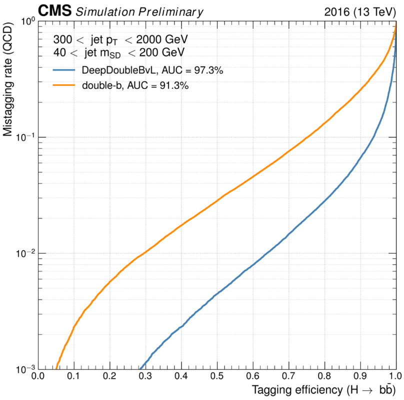
- Optimised neural network training for double b- and c-tagging for $h \to b\bar{b}$ and $h \to c\bar{c}$.
- Using convolutional neural networks with constituents and secondary vertex information as inputs.
- Uses the CMS double b-tagger (BTV-16-002) features to detect a boosted object to b-jet pair decay $\to$ exploits correlations between the flight directions of the b-jets, using N-subjettiness axes.



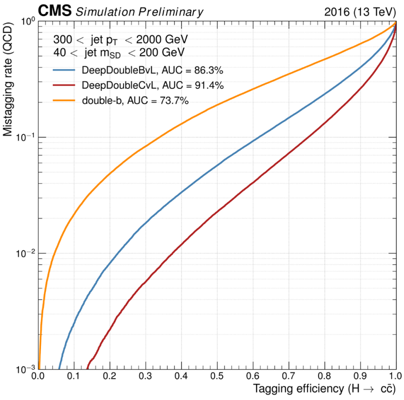
$h \to b\bar{b}$ and $h \to c\bar{c}$ identification in CMS

- Much better performance than the double *b*-tagger.

[DP 2018-046]



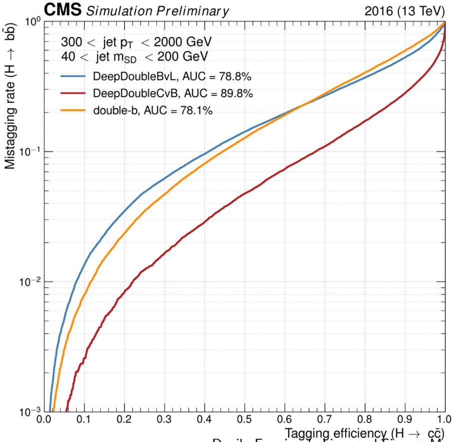
[DP 2018-046]



$b\bar{b}$ vs. $c\bar{c}$ separation

- After tagging $b\bar{b}$ decays, a veto on $c\bar{c}$ decays can be useful to orthogonalise analyses.
- The c vs. b tagger can be used to select c-jets $\rightarrow$ useful to reject $c\bar{c}$ background in a $b\bar{b}$ analysis or vice-versa.

[DP 2018-046]

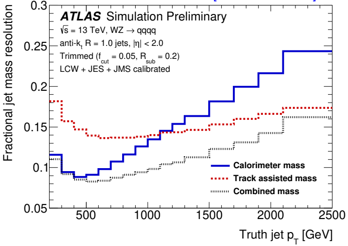


Top and W tagging

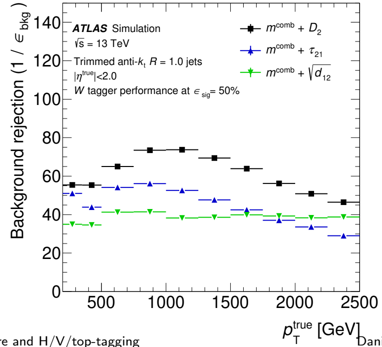
Cut-based taggers in ATLAS

- Improved mass resolution using combined track and calo. information.
- Using trimming to reduce the pile up effects.
- Cutting on the combined mass and D_2 or τ_{32} provides significant rejection.

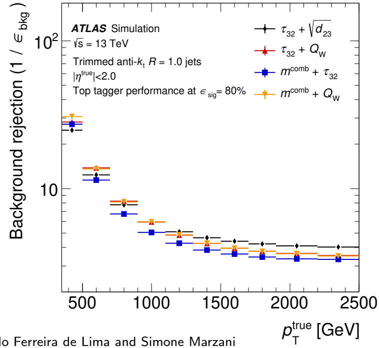
[JETM-2017-002]



[arXiv:1808.07858]



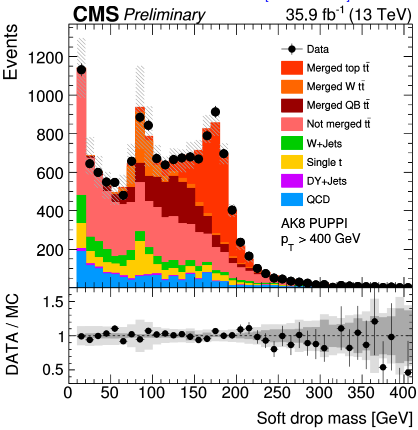
[arXiv:1808.07858]



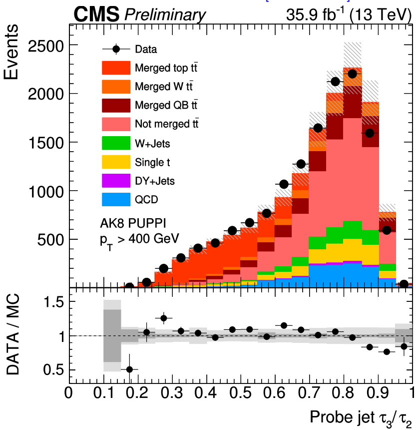
Cut-based top-taggers in CMS

- CMS uses a soft-drop mass estimate, PUPPI for pile up correction and N-subjettiness as a top-tagger.

[DP 2017-026]



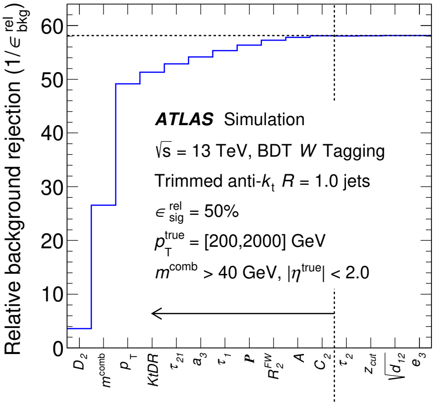
[DP 2017-026]



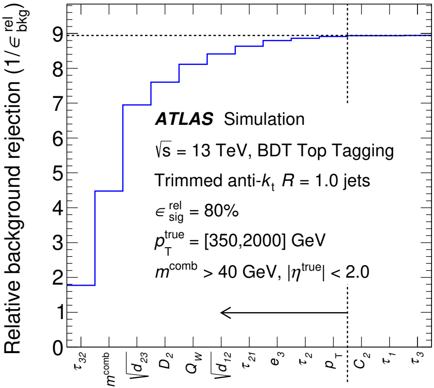
Top and W MVA-based taggers in ATLAS

- Top and W tagging using BDTs and neural networks combining several discriminant variables provide significant rejection.
- Variables can be added sequentially $\rightarrow$ no point adding more inputs without a significant benefit in classification.

[arXiv:1808.07858]



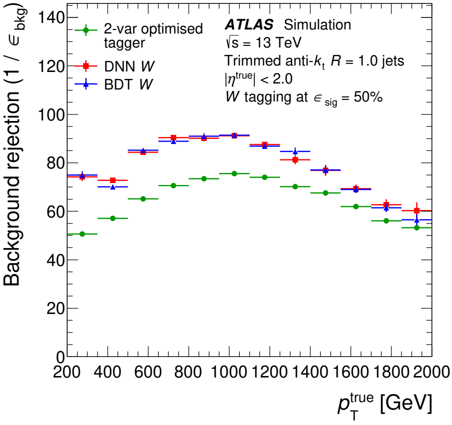
[arXiv:1808.07858]



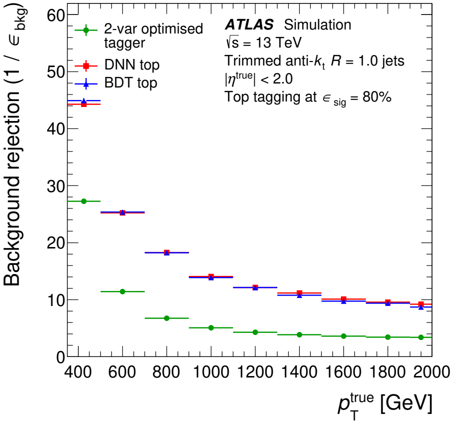
How much benefit do the MVA taggers provide?

- Up to 30% improvement in W tagging and 100% improvement in top-tagging rejection.
- MVA taggers provide a significant gain in top tagging.
- **Where is the extra gain coming from?**

[arXiv:1808.07858]



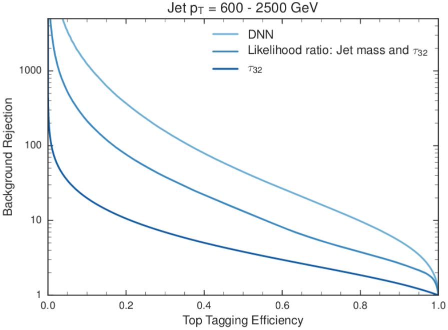
[arXiv:1808.07858]



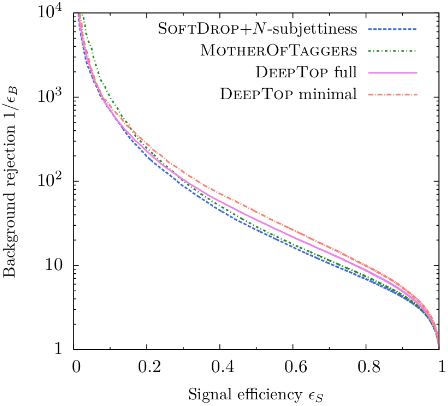
MVA methods with lower level inputs

- Using a lower level input for neural network, we may gain significantly in classification power. [1511.05190] [1603.09349] [1701.08784] [1704.02124]
- But there can be a significant impact from detector effects → need to be evaluated with detector simulation.

[arXiv:1704.02124]



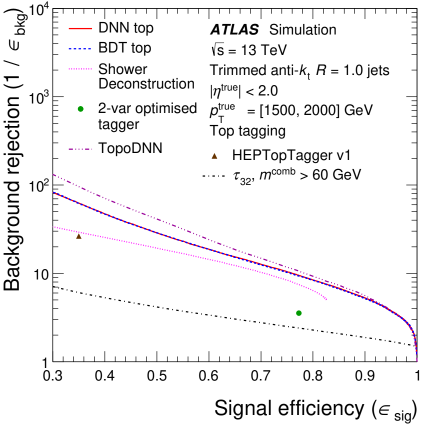
[arXiv:1701.08784]



A comparison of techniques

- Huge performance gain by MVA techniques at low and high p_T, but also at low and high rejection working points.
- Actually implemented TopoDNN in ATLAS.
  - Best performance even compared with BDT or NN methods which use a higher level input (other jet substructure variables).

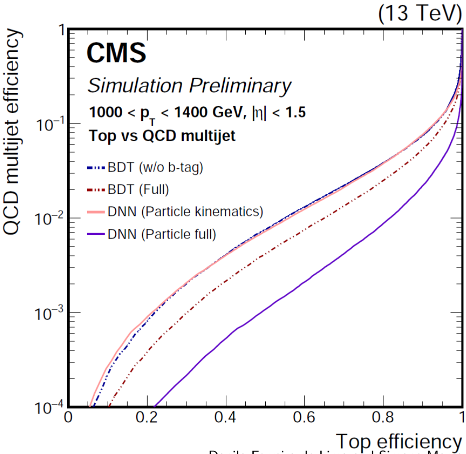
[arXiv:1808.07858]



Top tagging using MVA in CMS

- Convolutional NN used in CMS does better than BDT approach.
- Inputs can be constituent's kinematics (p_T, η, ϕ, ΔR between subjets) in the "kinematics" version.
- "Full" version adds information about the secondary vertex and track displacement and quality.

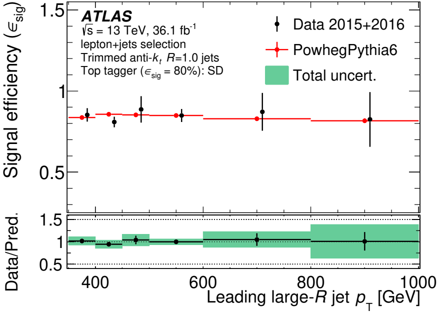
[DP2017-049]



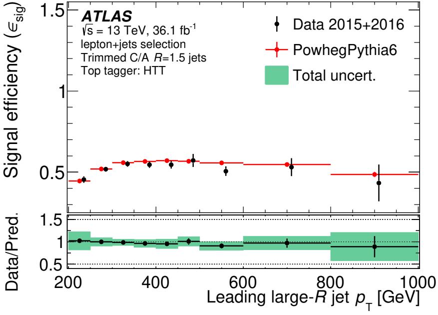
Efficiency – ATLAS SD and HTTv1

- And so are more complex taggers, such as Shower Deconstruction and HEPTopTagger v1.

[arXiv:1808.07858]



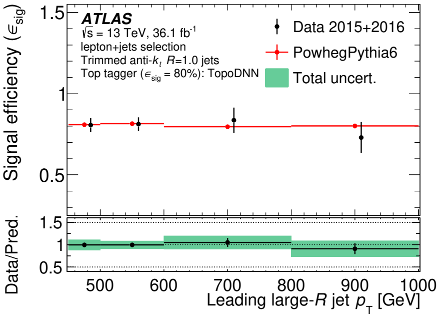
[arXiv:1808.07858]



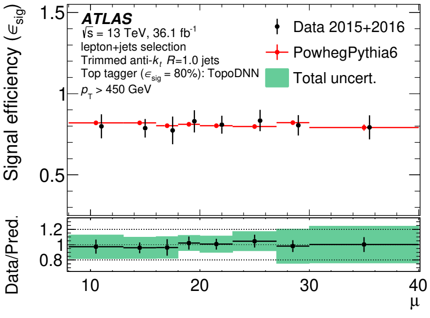
Efficiency – ATLAS TopoDNN

- Best performant MVA tagger has a very good agreement with data, both as a function of p_T and $< \mu >$.

[arXiv:1808.07858]



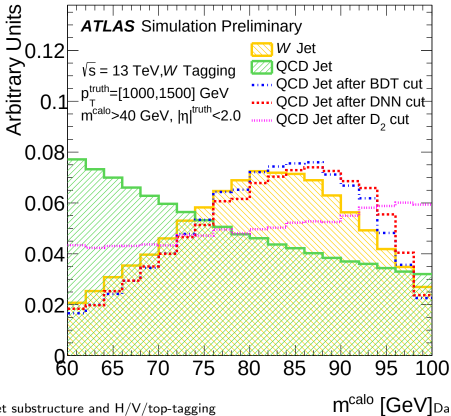
[arXiv:1808.07858]



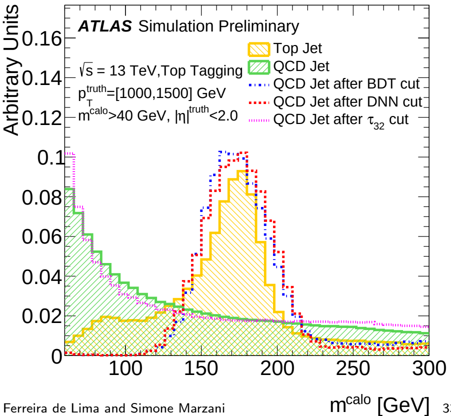
Tagging effects on mass

- Taggers often use the mass information to optimise the selection.
- This leads to sculpting of the mass spectrum in both signal and background.
- We might want to tag a generic $X \to qq'$ resonance without using its (unknown) mass.
- Or we might want to avoid using the mass information to define a background control region.

[ATL-PHYS-PUB-2017-004]



[ATL-PHYS-PUB-2017-004]



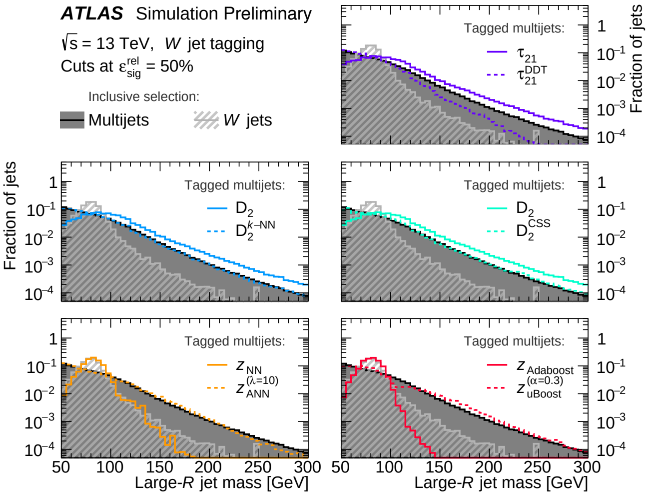
Mass decorrelation in ATLAS – methods

- The taggers can be made more general by not using the mass information.
- Several methods were studied in ATLAS to decorrelate the mass and other jet substructure information.
  - Fixed-efficiency regression [1710.00159] .
  - Designed decorrelated taggers [1603.00027] .
  - Convolved substructure [1710.06859] .
  - Adversarially trained NNs [1703.03507] .
  - Boosting for uniform efficiency [1305.7248] .
- ATLAS defined a metric for decorrelation using the symmetrised Kullback-Leibler divergence (JSD):
  - $\text{JSD}(\text{fraction}^{\text{pass}}_{\text{bkg}}||\text{fraction}^{\text{fail}}_{\text{bkg}}) \rightarrow$ measure distance between distributions of background events that pass and fail the selection.

Mass decorrelation in ATLAS

- The mass histograms below show the fraction of signal (W+jets) and background (QCD multi-jets) before and after the selection.

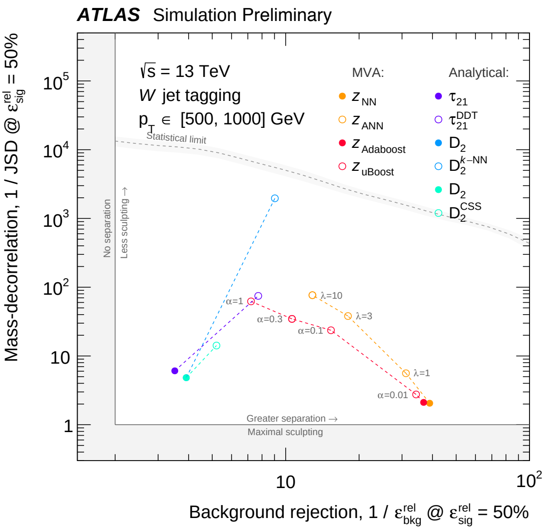
[ATL-PHYS-PUB-2018-014]



Mass decorrelation vs. background rejection

- It is possible to perform a fair comparison on several decorrelation methods using this metric.

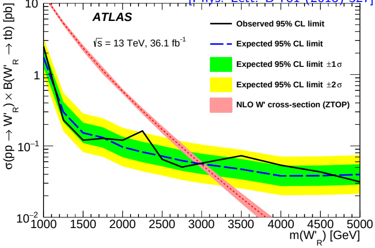
[ATL-PHYS-PUB-2018-014]



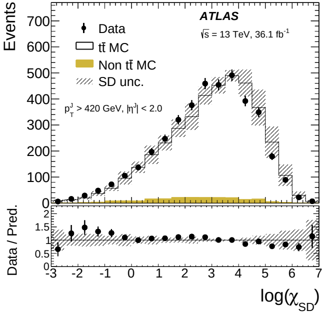
$W' \to tb$ search

- ATLAS search for $W' \to tb$ uses Shower Deconstruction to achieve a huge background reduction.
- A cut is made in the χ_{SD} likelihood ratio observable.

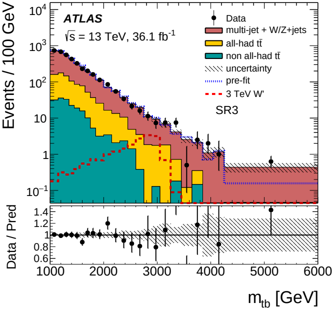
[Phys. Lett. B 781 (2018) 327]



[Phys. Lett. B 781 (2018) 327]



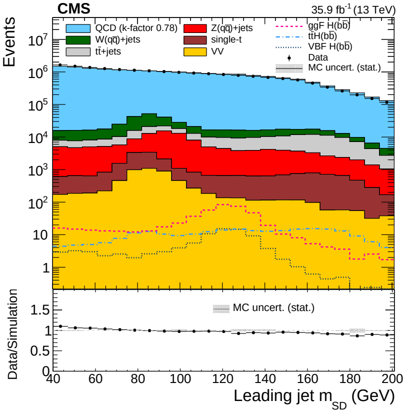
[Phys. Lett. B 781 (2018) 327]



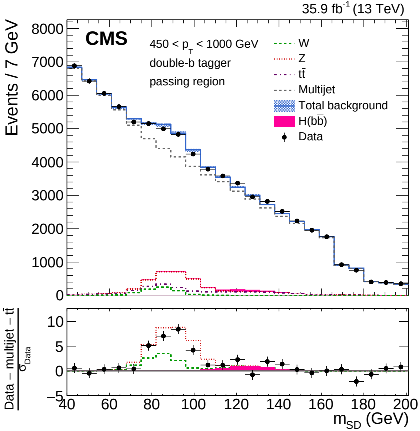
$h \to b\bar{b}$ search

- CMS uses the soft drop mass for a search for a highly boosted SM Higgs decaying into $b\bar{b}$.
- The soft drop groomed jets allow us to see quite clearly the $Z \to b\bar{b}$ events.

[Phys. Rev. Lett. 120 (2018) 071802]



[Phys. Rev. Lett. 120 (2018) 071802]

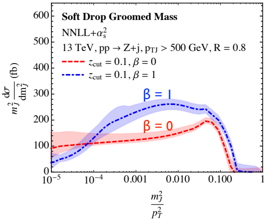


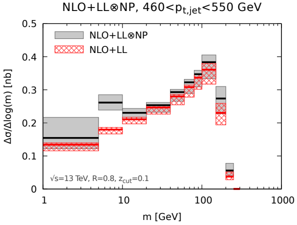
Jet substructure for precision

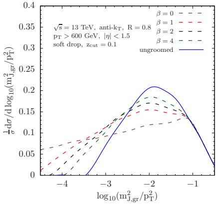
Precision calculations meet the data

Precision calculations for the soft-drop jet mass

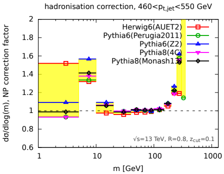
- All-order calculations, with meaningful uncertainty bands, are now available for the soft-drop jet mass. [Frye, Larkoski, Schwartz, Yan (2016)], [SM, Schunk, Soyez (2017)], [Kang, Lee, Liu, Ringer (2018)].

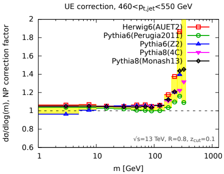






- Non-pert. corrections greatly reduced, perturbative convergence is improved.



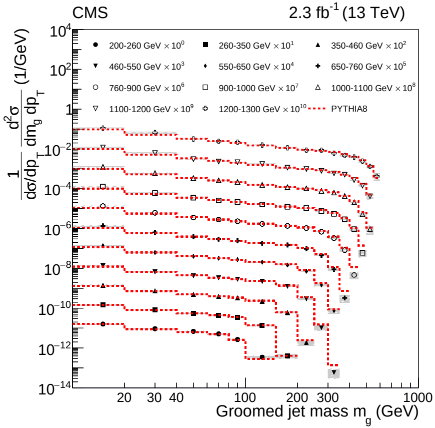


- Great opportunity to test and constrain simulation tools. e.g. parton showers and their hadronisation models.

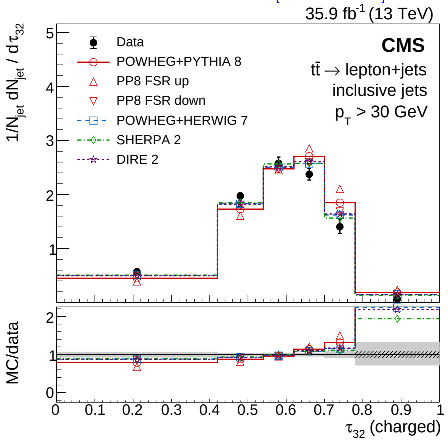
CMS jet substructure measurements

- CMS has provided measurements of several jet substructure variables. Showing the soft drop mass and *N*-subjettiness.
- Good agreement with simulation over a large phase space range.
- Some disagreement at higher p_T.
- Also unfolded the *N*-subjetiness observables.

[arXiv:1807.05974]



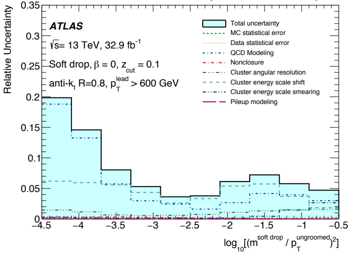
[arXiv:1808.07340]



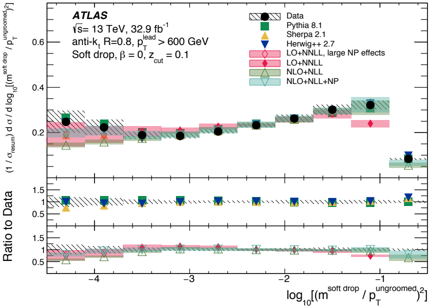
ATLAS soft drop mass measurement

- ATLAS has published measurements of $\log_{10}[m_{\text{soft drop}}/p_{T,\text{ungroomed}}]^2$ for $\beta \in \{0, 1, 2\}$.
- Disagreements where non-perturbative effects are expected to appear.
- Large uncertainty from the QCD modeling.

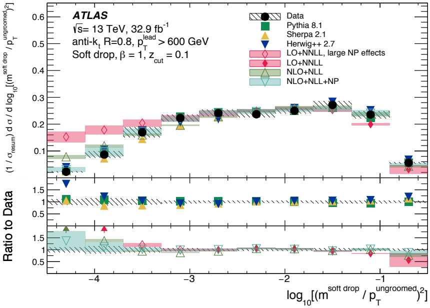
[arXiv:1711.08341]



[arXiv:1711.08341]



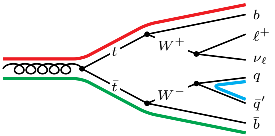
[arXiv:1711.08341]



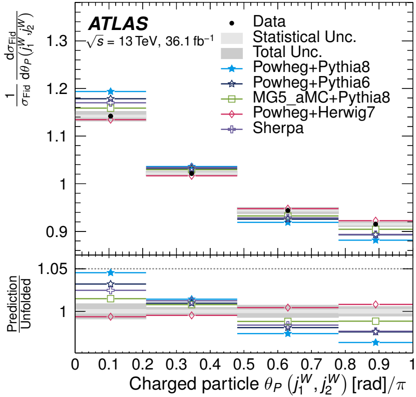
ATLAS colour flow measurement

- ATLAS used $t\bar{t}$ events to measure magnitude angle of the pull vector: allows a test of the colour flow.
- Colour flow angle between the W jets shown below after unfolding.
- Very subtle effect not well predicted by most simulators and measured with great precision.

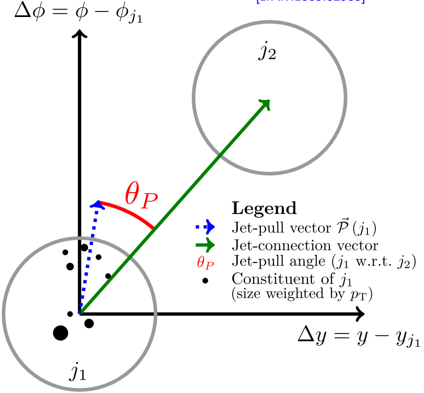
[arXiv:1805.02935]



[arXiv:1805.02935]



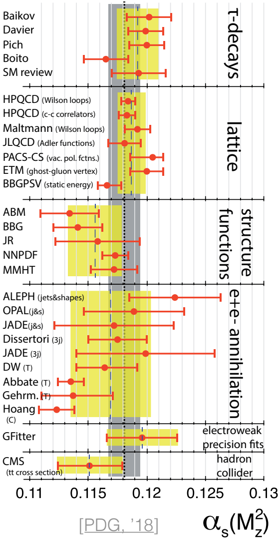
[arXiv:1805.02935]



Jet substructure for precision

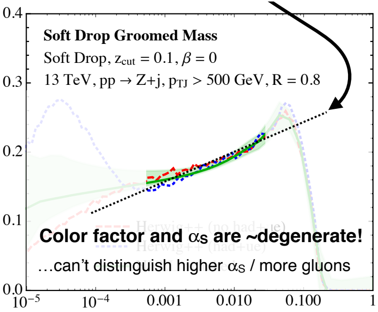
Towards the extraction of Standard Model parameters

Towards α_s extraction



(from B. Nachman, BOOST's talk)

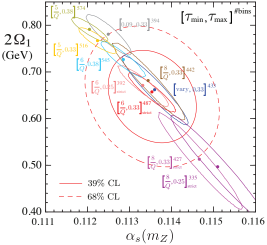
- Current precision is less than 1% and is dominated by the lattice.
- Next-most precise is LEP event shapes and differs by $\sim$5% (3σ).
- Preliminary studies show that using soft-drop mass (or other angularities) can lead to an extraction with 10% uncertainty. [Les Houches (2017)]

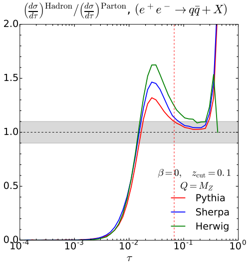


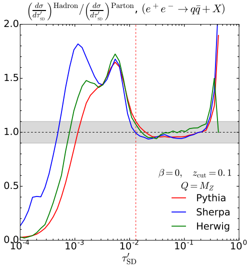
- Not yet competitive for the world average but worth pursuing.

Groomed e^+e^- event shapes

- This study can pave the way for a more competitive measurement in e^+e^-
- Groomed event shapes show reduced sensitivity to hadronisation and may help breaking degeneracy with non-perturbative effects and resolve long-standing puzzle of low α_s. [Baron, SM, Theeuwes (2018)]







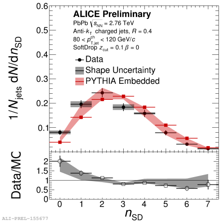
- New NNLO results for groomed event shapes are presented at this confernece

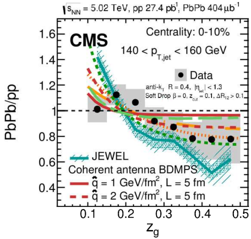
  [Kardos, Somogyi, Troócsányi (2018)]
- No time to cover it here but also: extraction of top mass with light grooming.

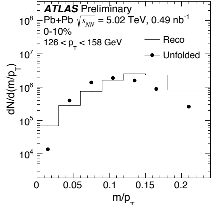
  [Hoang, Mantry, Pathak, Stewart (2017)]

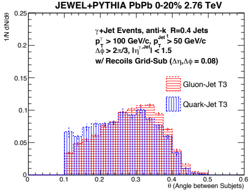
Jet substructure goes nuclear

- Ideas and observables to study jets in proton collisions have made their way into the heavy-ion community.
- Jets and their structure offer a unique probe for the quark-gluon plasma.
- Tough both theoretically and experimentally, but results are pouring in.





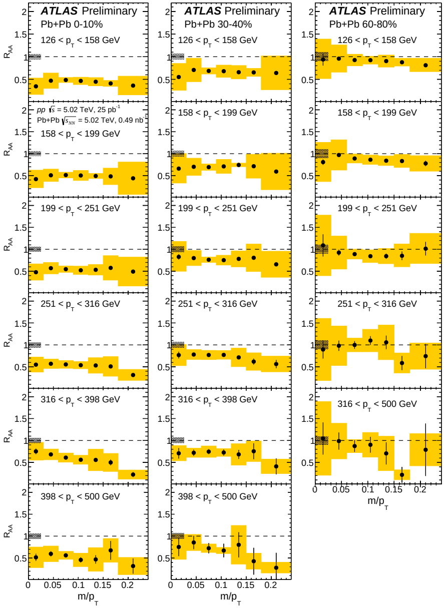




courtesy of Jesse Thaler

ATLAS measurement of m/p_T in PbPb and pp

[ATLAS-CONF-2018-014]



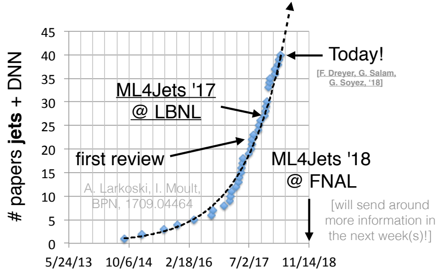
- The nuclear modification factor, R_{AA}, is measured.
- No visible change in the m/p_T distribution shapes between PbPb and pp.
- We are not yet sensitive to quark-gluon plasma effects in this variable.

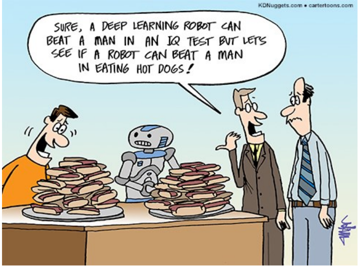
New ideas and techniques

News from BOOST 2018

Deep Thinking VS Deep Learning

- Particle physics (and jet physics) undergoing a machine-learning revolution.
- New ideas and techniques are pouring into the field.

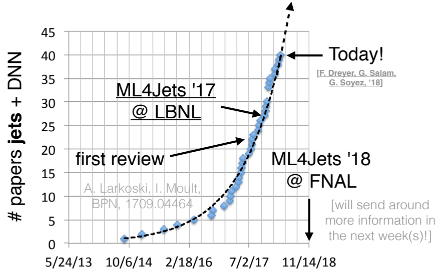


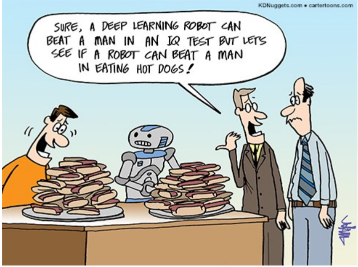


- Techniques met with a mixture of excitement and skepticism.
- Many studies to investigate the information content exploited by these methods (make the black box more transparent)
- Deep Thinking & Deep Learning can lead to Deep Understanding
- Join the Machine Learning for Jet Physics workshop in November if you are interested in contributing: https://indico.cern.ch/event/745718/

Deep Thinking VS Deep Learning

- Particle physics (and jet physics) undergoing a machine-learning revolution.
- New ideas and techniques are pouring into the field.

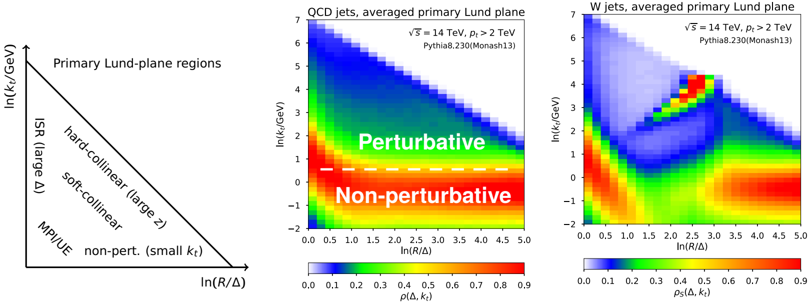




- Techniques met with a mixture of excitement and skepticism.
- Many studies to investigate the information content exploited by these methods (make the black box more transparent)
- Deep Thinking & Deep Learning can lead to Deep Understanding
- Join the Machine Learning for Jet Physics workshop in November if you are interested in contributing: https://indico.cern.ch/event/745718/

Physics intuition meets computer science I: Lund plane

- Standard approach is supervised learning: apply classification algorithms to large collections of (simulated) samples, e.g. the jet image. [de Oliveira, Kagan, Mackey, Nachman, Schwartzman (2015)]
- Physics intuition can lead us to construct better representations of a jet: the Lund plane. [Dreyer, Salam, Soyez (2018)] (see talk in parallel session)

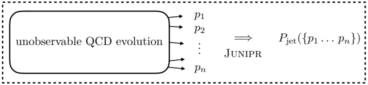


- Decluster the jet following the hard branch and record (k_t, Δ) at each step
- Use this representation as input of log-likelihood or ML algorithms.

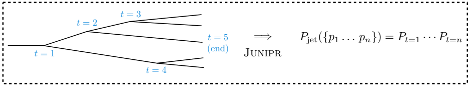
Physics intuition meets computer science II: Junipr

- Physics intuition can lead us to construct better representations of a jet: Jets using UNsupervised Interpretable PRobabilistic models. [Andreassen, Feige, Frye, Schwartz (2018)]

JUNIPR computes the probability of a jet...



...as a product over time steps in its clustering tree...



...where each time step is decomposed into 3 parts:

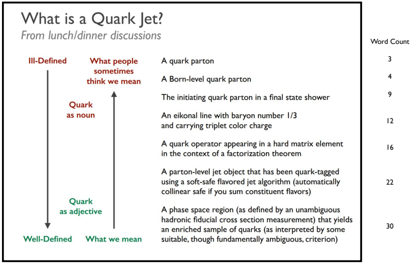
$$P_t = P_{\text{end}} \cdot P_{\text{mother}} \cdot P_{\text{branch}}$$

slide by Chris Frye

- Implemented using Recurrent Neural Networks, which are trained on (simulated) data.
- The trained model (10^6 parameters) can be used for discrimination as well as for generation!

Physics intuition meets computer science III: Jet Topics

- Quark/gluon tagging always a hot (and controversial) topic in jet substructure.
- Jet topics exploit the technology of topic modelling from texts. [Medotiev, Thaler (2018)]

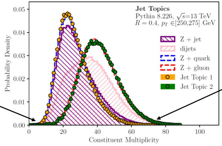


adapted from E. Metodiev and J. Thaler

$$p_{\text{quark}}(\boldsymbol{x}) \equiv \frac{p_A(\boldsymbol{x}) - \kappa_{AB}\, p_B(\boldsymbol{x})}{1-\kappa_{AB}}$$

$$p_{\text{gluon}}(\boldsymbol{x}) \equiv \frac{p_B(\boldsymbol{x}) - \kappa_{BA}\, p_A(\boldsymbol{x})}{1-\kappa_{BA}}$$

$$\kappa_{BA} \equiv \min_{\boldsymbol{x}} \frac{p_B(\boldsymbol{x})}{p_A(\boldsymbol{x})} = \frac{f_B^q}{f_A^q}$$



$$\kappa_{AB} \equiv \min_{\boldsymbol{x}} \frac{p_A(\boldsymbol{x})}{p_B(\boldsymbol{x})} = \frac{1-f_A^q}{1-f_B^q}$$

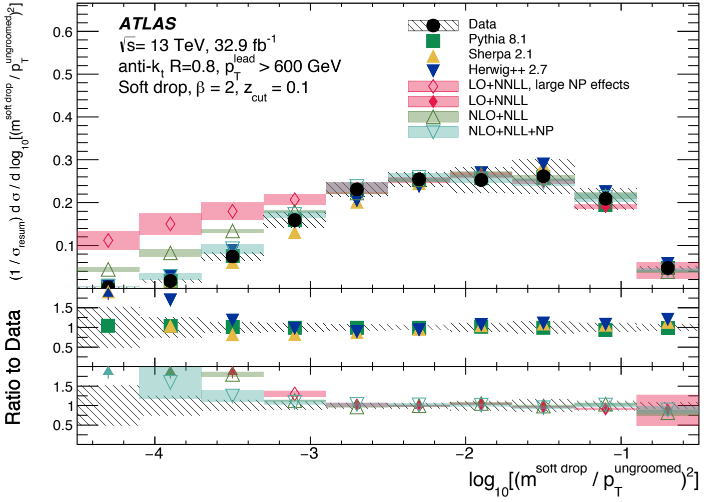
- Basic assumptions: categories exist and they are mutually irreducible ($\exists$ region of 100 % purity for each topic).

Summary

- Many ideas from both ATLAS and CMS on top/V/Higgs tagging!
- A lot of dialogue with the phenomenology community, recycling, improving and creating new methods.
- How can we *understand* the improvements observed from Machine Learning?
  - Analytical calculations are excellent tools:
    - to understand jet substructure and ...
    - to develop new taggers.
- The time has come to move from just *tagging* jets to also *measuring* jets.

Summary

[arXiv:1711.08341]

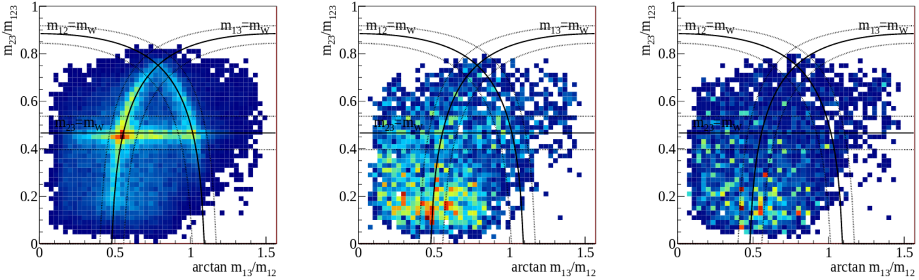


Extra slides

HEPTopTagger v. 1

- Tests the compatibility with a 3-prong decay based on the sub-jets mass ratios. [0910.5472] [1006.2833]
- Cuts on mass ratio of the 3 decay products.

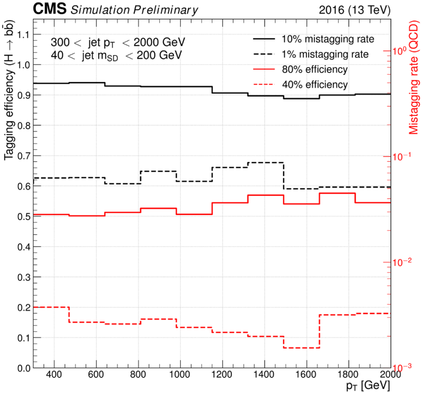
[1006.2833]



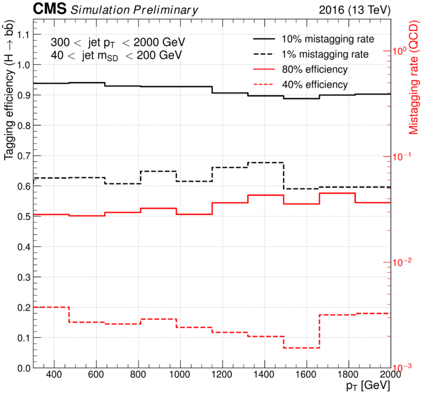
CMS: $h \to b\bar{b}$ and $h \to c\bar{c}$ tagger p_T dependency

- Not a huge dependency on p_T.

[DP 2018-046]

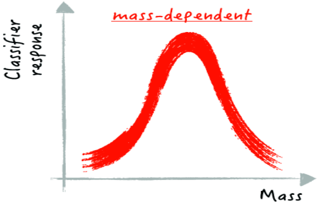


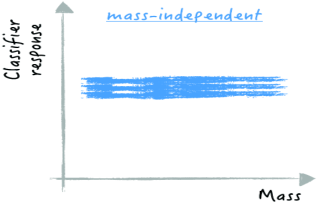
[DP 2018-046]



CMS: Mass sculpting

- But are the previous taggers relying on the jet mass?
- If the tagger uses the jet mass directly, the performance may be different for a Higgs or a $Z \to b\bar{b}$.

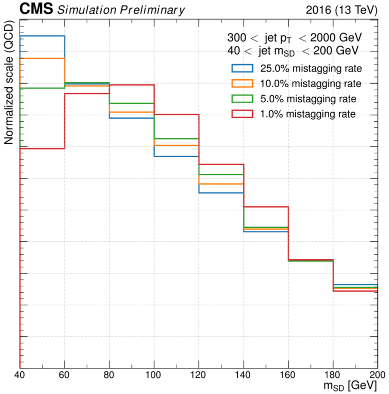




CMS: Mass sculpting – effect of decorrelation

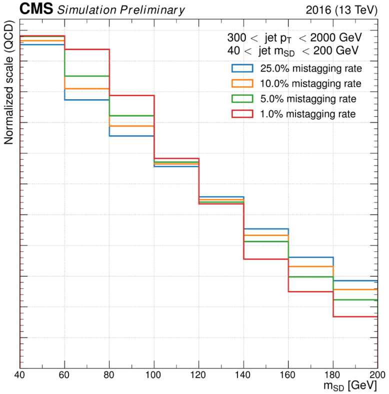
- Decorrelating the jet mass and the tagger may allow the tagger to be applicable in more analyses.

[DP 2018-046]



Before decorrelation

[DP 2018-046]



After decorrelation