Searches for new physics using jet substructure in ATLAS and CMS

Steven Schramm[*]**, on behalf of the ATLAS and CMS Collaborations**

Université de Genève (CH)

E-mail: steven.schramm@cern.ch

The ATLAS and CMS experiments are increasingly making use of jet substructure techniques in the search for new physics. By exploiting the energy and angular correlation structure within a jet, it is possible to differentiate between jets originating from light-quarks and gluons and jets from hadronically-decaying massive particles. Such techniques allow for the creation of new hadronic analyses, including di-tagged-jet resonances, single-jet mass scan searches, and searches for pair-produced vector-like quarks with complex hadronic final states. The sensitivities of these new searches are typically limited by the precision of the jet four-vector (especially the jet mass) and the ability to identify jets containing hadronically-decaying massive particles. Studies into new methods and tools to further extend the power of jet substructure techniques are very promising, and are set to further expand the relevance of jet substructure in the search for new physics.

ATL-PHYS-PROC-2018-163
25 November 2018

XIII Quark Confinement and the Hadron Spectrum - Confinement2018
31 July - 6 August 2018
Maynooth University, Ireland

[*]Speaker.

1. Introduction

The ATLAS [1] and CMS [2] experiments have now collected a substantial dataset of proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ as delivered by the LHC. Both collaborations are now actively working on analyzing this large dataset, and are conducting a wide variety of searches for new phenomena beyond the Standard Model (SM) of particle physics. In parallel, the two collaborations are also working on improving their understanding of the ATLAS and CMS detectors, and further refining their ability to reconstruct the nature of particles interacting with the detectors. This improved understanding of particle interactions has led to the creation of a new class of searches for physics beyond the SM, referred to as searches using jet substructure, where the energy and angular correlation structure within an individual hadronic jet is used to select interesting events.

Traditionally, jets have been used to represent the hadronization of individual light-quarks or gluons and their subsequent interactions with the detector, where light-quarks denotes all non-top-quarks. However, individual light-quarks and gluons are not the only processes which give rise to hadronic showers. Massive hadronically-decaying particles, such as electroweak W/Z bosons, Higgs bosons, and top-quarks can all decay to two or more light-quarks, which then interact with the detector. At low transverse momentum p_T these decays give rise to light-quarks which are individually distinguishable. However, as the p_T of the hadronically-decaying massive particle increases, the boost from the decay frame (where daughter particles are back to back) to the experimental frame results in highly collimated light-quarks. At sufficiently high parent-particle p_T, the light-quark hadronic showers begin to overlap, and thus they cannot be reconstructed as two separate jets. Instead, the entire decay system is reconstructed as a single large jet, and thus the large jet now represents not an individual light-quark or gluon but rather a hadronically-decaying massive particle. These two hadronic decay regimes are typically referred to as the *resolved* regime, where the parent particle is at low p_T and thus the light-quarks can be reconstructed separately, and the *merged* or *boosted* regime, where the parent particle p_T is sufficiently high that the entire decay system is reconstructed as a single large jet. A schematic showing the two regimes is provided in Figure 1.

In the boosted regime, it is very important to be able to distinguish large jets containing hadronically-decaying massive particles from those only containing individual light-quarks and gluons. This is made possible by studying the energy and angular correlations between the constituents which make up the jet, referred to as the jet's substructure. To first order, this can be thought of as the mass of the jet and the number of decay axes present within the jet. In the case of a hadronically-decaying massive particle, the jet mass should correspond to the particle's mass, and the number of decay axes should correspond to the light-quark multiplicity of the hadronic decay. On the other hand, individual light-quarks or gluons should have only a single axis and should have a small mass, corresponding to the stochastic nature of hadronic showers and the objects used to reconstruct the jet. This allows for the ability to identify real hadronic decays of massive particles, as represented in the schematic in Figure 2. More concrete examples of jet identification are presented in Section 2.

The ability to reconstruct entire hadronic decays as a single large jet, and the subsequent capability to identify the origin of the jet, is very useful in the search for physics beyond the SM. In the first seven months of 2018, ATLAS and CMS each individually published 15 results (papers
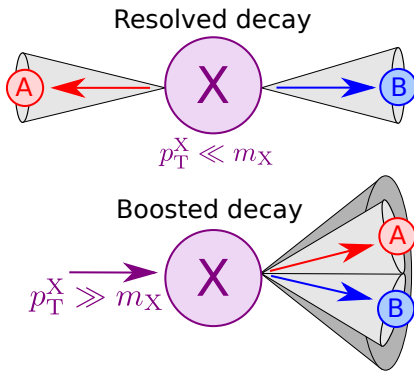
Figure 1: A schematic for a hadronically-decaying massive particle $X \to A + B$ showing the separation between the low p_T or *resolved* regime and the high p_T or *boosted* or *merged* regime. Light grey represents typical jets, while dark grey denotes large jets.
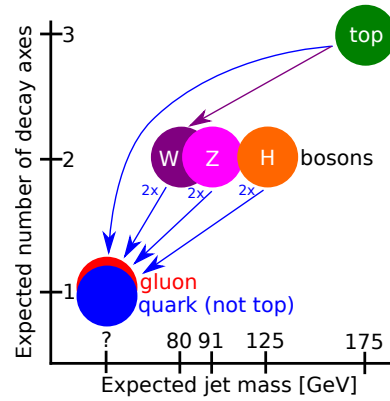
Figure 2: A schematic for hadronically-decaying massive particles in the Standard Model, showing the general idea of how the jet mass and number of axes in the jet can be used to identify the parent particle.

or notes) on searches for new phenomena using jet substructure, meaning that there were a total of 30 new results between the two collaborations. These results can primarily but not completely be divided into three categories: dijet resonance searches, single-jet mass scan searches, and searches for hadronically decaying vector like quarks. These three categories will be discussed in Sections 3, 4, and 5 respectively with all examples taken from results published in the first part of 2018.

Searches making use of jet substructure are typically limited by the jet identification power or the precision to which the large jet four-vector is known. These are both areas of active study, and a brief perspective on future developments in the field of jet substructure is presented in Section 6.

2. Jet substructure

As mentioned, the jet mass is a very powerful variable for the identification of the jet's initiating particle. The jet mass is part of the four-vector, and is thus a well-defined quantity with a physical meaning and truth expectation value. In contrast, jet substructure variables which quantify the consistency of the jet energy and angular distribution with the hypothesis of coming from a given number of particles (being aligned with a given number of axes) is less robust. They are typically intrinsically linked with the spatial resolution of the objects used to reconstruct the jets, and will differ drastically with respect to different jet reconstruction procedures. Substructure variables are thus powerful tools, but they have no absolute expectation value, rather there is typically an expectation that a light-quark or gluon jet will have a large or smaller value than the hadronically-decaying massive particle of interest. Two common examples of substructure variables will be considered.

The reconstructed jet mass and its corresponding separation between hadronically-decaying W bosons and the QCD jet background (individual light-quarks and gluons) is shown in Figure 3. While this does remove the majority of the QCD, the QCD cross-section is much larger than the cross-section for any hadronically-decaying massive particle, and thus it is necessary to define
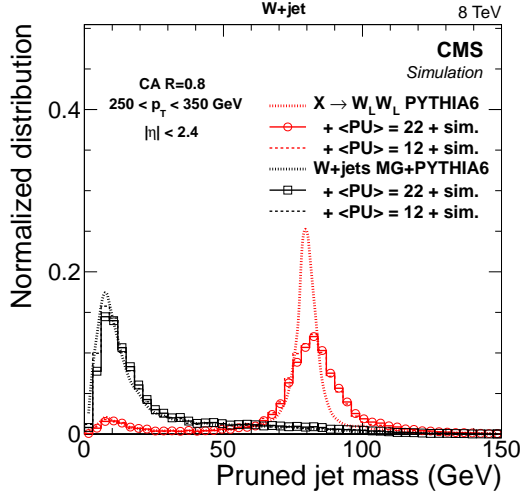
Figure 3: The jet mass for QCD jets (black) and hadronically-decaying W boson jets (red), demonstrating the power of jet mass for identifying hadronically-decaying massive particles [3].
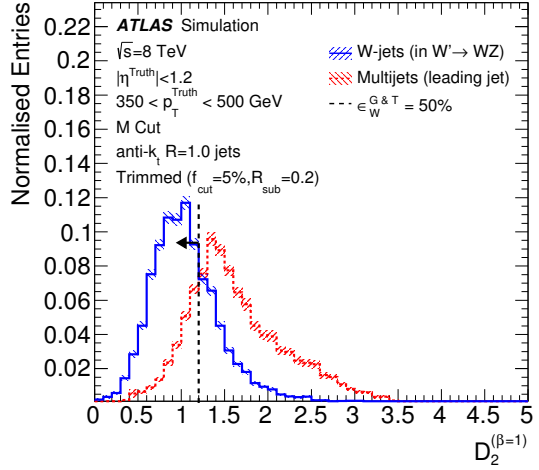
Figure 4: The $D_2^{(\beta=1)}$ energy correlation ratio after the application of a mass window selection, for QCD jets (red) and hadronically-decaying W boson jets (blue), further enhances jet identification [4].

further selections that reduce the QCD background. One powerful example is the $D_2^{(\beta=1)}$ variable, defined below, which is one way of representing the consistency of a jet under the one-axis vs multi-axis interpretation. The discrimination power of the $D_2^{(\beta=1)}$ variable after the application of a jet mass cut is shown in Figure 4, and the variable itself is defined as [5]:

$$D_2^{\beta} = \text{ECF}_3 \cdot \left[\frac{\text{ECF}_1}{\text{ECF}_2} \right]^3$$

$$\text{ECF}_1^{\beta} = \sum_{i \in \{\text{jet constituents}\}} p_T^i$$

$$\text{ECF}_2^{\beta} = \sum_{i,j \in \{\text{jet constituents}\}} p_T^i p_T^j \, (\Delta R_{ij})^{\beta}$$

$$\text{ECF}_3^{\beta} = \sum_{i,j,k \in \{\text{jet constituents}\}} p_T^i p_T^j p_T^k \, \left(\Delta R_{ij} \Delta R_{jk} \Delta R_{ik} \right)^{\beta} .$$

The D_2^{β} variable is typically used with $\beta = 1$ and is referred to as an energy correlation ratio, as it is built from a ratio of the 1-, 2-, and 3-point Energy Correlation Functions (ECFs). This variable avoids defining any explicit axes, but an alternative approach is to directly impose a given number of axes on the jet constituents and evaluate the consistency of the jet with that interpretation. This is done in the n-subjettiness variable τ_{32}^{β}, which also is typically used with $\beta = 1$, and which measures the consistency of a given jet with the interpretation of being composed of 3 subjets as opposed to 2 or fewer. This variable is typically used for top-quark identification, and is defined as [6]:

$$\tau_{32}^{\beta} = \tau_3^{\beta} / \tau_2^{\beta}$$

$$\tau_0^{\beta} = \sum_{i \in \{\text{jet constituents}\}} p_T^i \, \Delta R_{a_{\text{jet}},i}^{\beta}$$

$$\tau_2^{\beta} = \frac{1}{\tau_0} \sum_{i \in \{\text{jet constituents}\}} p_T^i \, \min\left(\Delta R_{a_1,i}^{\beta}, \Delta R_{a_2,i}^{\beta}\right)$$

$$\tau_3^{\beta} = \frac{1}{\tau_0} \sum_{i \in \{\text{jet constituents}\}} p_T^i \, \min\left(\Delta R_{a_1,i}^{\beta}, \Delta R_{a_2,i}^{\beta}, \Delta R_{a_3,i}^{\beta}\right) \, .$$

$$a_i = \text{a given axis of, or within, the jet}$$

Taggers built from cuts on a pair of jet substructure variables, typically including the jet mass and one other variable (such as $D_2^{(\beta=1)}$ or $\tau_{32}^{(\beta=1)}$) are already very powerful and have a history of usage within both ATLAS and CMS. More complex taggers including machine learning classifiers are increasingly being used, but most of the searches to date for physics beyond the SM using jet substructure use these simple taggers.

3. Dijet resonance searches

Dijet resonance searches have a long history back to well before the LHC, and have been conducted several times by ATLAS and CMS. However, the advent of powerful tools for the identification of hadronically decaying massive particles has led to a new class of dijet resonance searches. There are many models of physics beyond the SM that predict couplings to pairs of massive SM particles, whether they are pairs of W bosons, Z bosons, Higgs bosons, top-quarks, or combinations thereof. All of these SM particles have large or dominant branching fractions to hadronic final states, and thus it is interesting to consider searches where one or both of the massive particles decays hadronically.

A prototypical example is the fully hadronic di-boson resonance search, which studies the WW, WZ, and ZZ final states where each boson decays to a pair of quarks. The ATLAS analysis [7] makes use of general topological selections, followed by very tight cuts on the jet mass and $D_2^{(\beta=1)}$, as shown in Figure 5. This selection retains less than 20% of possible signal events for one benchmark model, but it also suppresses the QCD background by several orders of magnitude. The sensitivity of the analysis is primarily driven by the sensitivity of the electroweak boson tagger, and thus the latest iteration of the analysis invested the time in developing new inputs to jet reconstruction and re-optimizing the boson tagger specifically for the analysis. The end result is significantly improved due to these changes, with $2 - 4\times$ improvements over a simple luminosity scaling for most of the parameter space, as shown in Figure 6.

Another example is to exploit the Higgs boson, and conduct searches for di-Higgs resonances where each Higgs boson decays to a pair of *b*-quarks. As with the fully hadronic di-boson resonance search, the sensitivity to di-Higgs decays is primarily defined by the ability to identify Higgs bosons. The CMS analysis [8] (superseded by [9]) does this by double-*b* tagging the large jet to significantly suppress the QCD background, and then applying an additional selection on the large
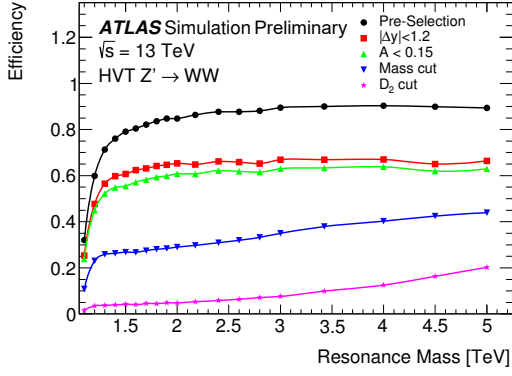
Figure 5: The sequential selection efficiency of the fully-hadronic di-boson resonance search, demonstrating the tight requirements imposed on jet mass and $D_2^{(\beta=1)}$ for hadronic W boson identification [7].
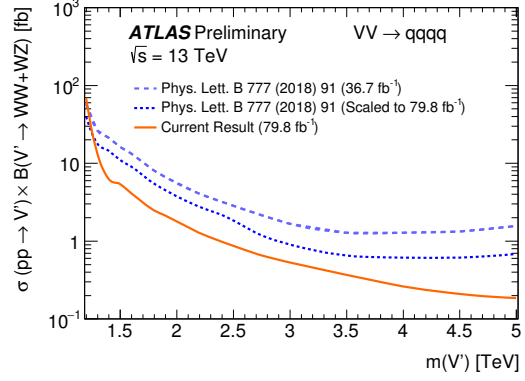

Figure 6: Combined limits on the production of WW+WZ resonances comparing the past result before (long-dashes) and after (short-dashes) scaling to the new luminosity, and the new much-improved result (solid) using new jet inputs and taggers [7].

jet mass to further suppress QCD jets and also reduce the $t\bar{t}$ background, as shown in Figure 7. The end result is an impressive set of limits on di-Higgs resonances, where the analysis is nearing the level of sensitivity required to probe standard bulk KK graviton models, as shown in Figure 8. This is an impressive feat, given that the Higgs boson was only recently discovered, and its decay mode to pairs of b-quarks had not yet been confirmed by both ATLAS and CMS at the time of this conference.
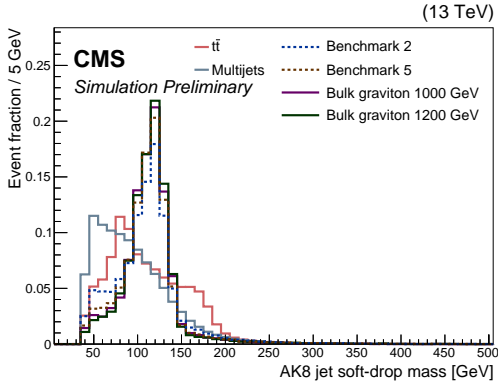

Figure 7: The jet mass is used on double-b-tagged jets to provide further discrimination power between the signal Higgs boson jets (black) and the background jets from QCD (blue) or $t\bar{t}$ (red) sources [8] (superseded by [9]).
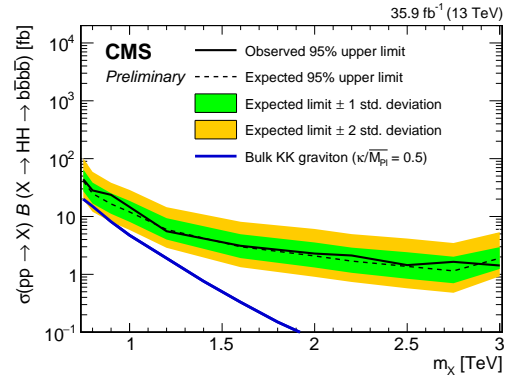

Figure 8: Limits on the production of di-Higgs boson resonances where each Higgs boson decays to a pair of b-quarks, which are approaching the needed sensitivity to probe standard graviton models [8] (superseded by [9]).

4. Jet mass scan searches

The typical resonance search for a pair of objects works well at high energies and in the strong

coupling regime. However, it is also important to search for weakly-coupled low-mass resonances. This cannot be done with standard dijet resonance searches due to the high trigger thresholds, which currently limit most dijet searches to searches for new resonances in the TeV regime. There are several methods to work around the trigger limitations, including one method which relies heavily on jet substructure, known as mass-scan searches. In such searches, the trigger limitation is bypassed using a hard Initial State Radiation (ISR) object which is easy to trigger on. As the ISR object has a large p_T, the recoiling system is also boosted, and thus is collimated into a single large jet. The large jet mass can therefore be probed for bumps, which would correspond to low-mass resonances contained within a single jet.

One example of such a search is the inclusive *X*+ISR search, where *X* denotes the potential new particle with an unknown mass. The ATLAS *X*+ISR search [10] is conducted in two channels, where the triggering ISR object is either a photon or another jet. In order to search for a bump in the single large jet mass spectrum, it is important that any selections applied in the analysis do not bias the jet mass. This is a very challenging constraint, as it is still necessary to suppress the QCD background, and essentially all substructure variables are correlated with the jet mass. Instead, the Designed Decorrelated Tagger (DDT) procedure [11] is used to decorrelate τ_{21} from the jet mass, therefore defining a QCD discriminant τ_{21}^{DDT} which does not bias the mass spectrum. This decorrelation is demonstrated in Figure 9. The limits from conducting a resonance search on the resulting single large jet mass spectrum is shown in Figure 10, where this result covers the range of potential new particle mass between 100 and 220 GeV.
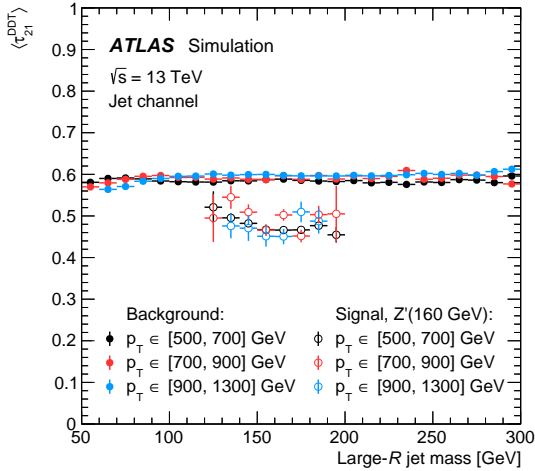

Figure 9: The jet mass as a function of the mass-decorrelated variable τ_{21}^{DDT}, showing minimal dependence on the jet p_T (different colours) and showing a clear separation between signal (open circles) and background (filled circles) [10].
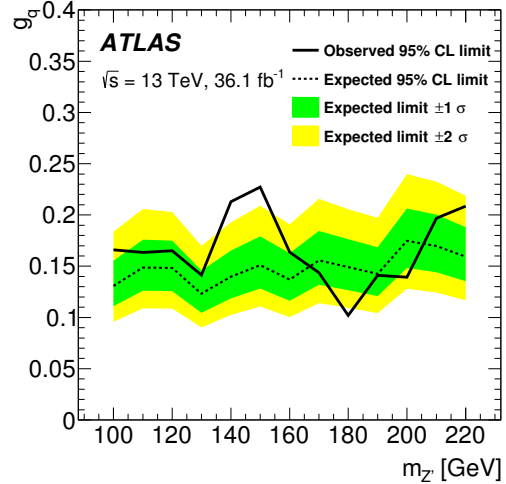
Figure 10: Limits on the production of new low-mass dijet resonances within a single boosted large jet, combining both the recoiling ISR photon and jet channels [10].

The same analysis can be further extended by applying a double-*b* tag to the large jet, therefore strongly suppressing the QCD background. Additionally, double-*b* tagging increases the sensitivity to new particles which couple strongly to more massive particles (including many scalar models).

The CMS analysis making use of this approach [12] very clearly observes the presence of the $Z \to bb$ peak as seen in Figure 11, which is a nice validation of the analysis methodology. The analysis is further conducted for two different jet definitions, including both large and very large jets. This combination of two jet definitions allows for the search to cover potential new particle masses between 50 and 350 GeV, as shown in Figure 12.
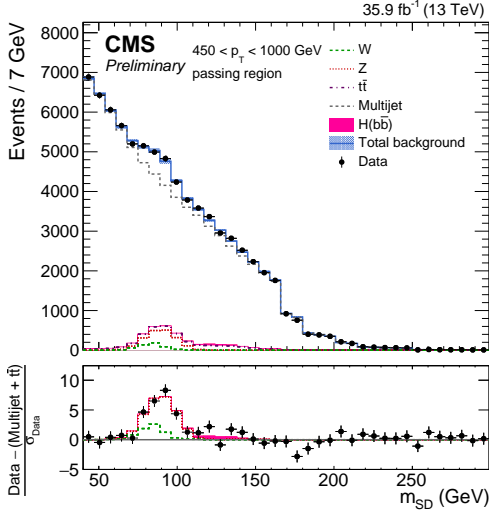

Figure 11: The mass of double-*b* tagged large jets, clearly showing the presence of the Z boson and potentially hinting at the presence of the Higgs boson decaying to a pair of *b*-quarks [12].
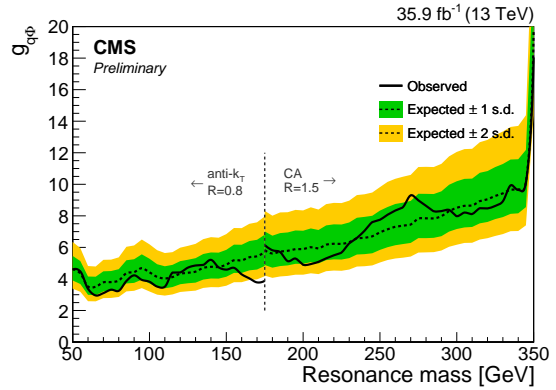
Figure 12: Limits on the production of new low-mass double-*b* tagged resonances within a single boosted large jet, combining two different jet sizes and algorithms to probe a larger mass range [12].

5. Vector-like quark searches

Vector-Like Quarks (VLQs) couple to one fermion, usually from the third generation, and one massive boson. This gives rise to final states of top-quarks (a third generation fermion), electroweak bosons, and Higgs bosons. As an example, the *B* VLQ can decay three ways: Wt, Zb, and Hb. As such, di-*B* production results in a final state with between two and four massive particles, all of which can decay hadronically. Jet substructure is an excellent tool to exploit the full potential of such complex final states.

The CMS search for di-*B* and di-*T* production in semi-leptonic final states [13] considers 1, 2, and 3-lepton channels. In the 1-lepton channel, the large jet mass and τ_{21} are both used to suppress the large $t\bar{t}$ background. The τ_{21} distribution after the application of a jet mass cut is shown in Figure 13. Jet substructure information is further part of the 16-way categorization of W boson vs Higgs boson jets used by the analysis. The resulting limit on a benchmark di-*B* production model is shown in Figure 14.

Semi-leptonic decay modes of pair-produced VLQs are useful as the presence of leptons significantly reduces the QCD background. However, the large hadronic branching fractions of the many massive particles also motivates the 0-lepton analysis. The ATLAS 0-lepton di-*B* and di-*T*
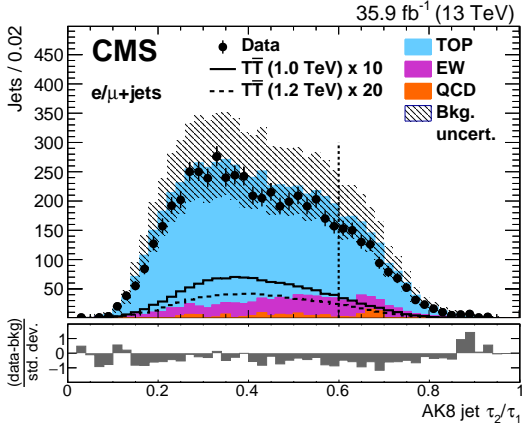
Figure 13: The $\tau_{21} = \tau_2/\tau_1$ jet substructure variable provides further discrimination between VLQ signals (black lines) and background (colour-filled regions) after the use of a jet mass selection [13].
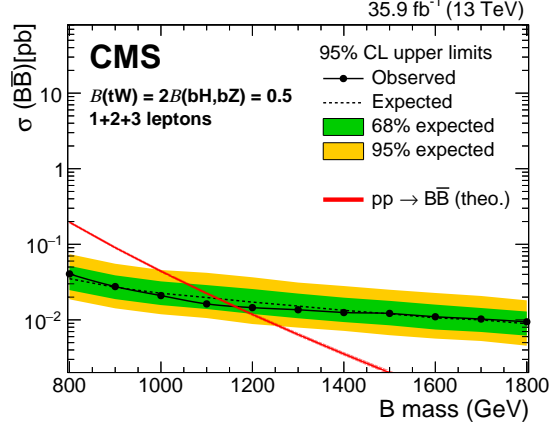
Figure 14: Limits on the production of VLQ $B\bar{B}$ pairs, combining the results from single-, di-, and tri-lepton final states, where jet substructure is used extensively in the single-lepton channel [13].

production search [14] makes use of a more complex jet tagger to suppress the large QCD background and resolve the ambiguity between different hadronically-decaying particles. This machine learning classifier is composed of a Deep Neural Network (DNN), with multiple output weights, thus making it a multi-class tagger. This is useful as the final discriminant can therefore distinguish between different categories of jets. In the first stage, QCD jets are rejected in favour of signal jets comprised of W/Z bosons (Vector bosons), Higgs bosons, or top-quarks, as shown in Figure 15. In the second stage, any ambiguity in the jet type is resolved by comparing the classifiers output for the different signal interpretations and picking the most probable result. The final limits for the same benchmark are shown in Figure 16. These limits are not currently as powerful as the semi-leptonic analysis for the benchmark model in question, but the fully hadronic search does provide better sensitivity to the $B \to Hb$ decay mode with respect to the semi-leptonic search.

6. Improving jet substructure for future searches

The sensitivity of dijet resonance searches and VLQ searches is typically limited by the ability to identify the hadronically-decaying massive particle of interest. It is therefore worthwhile investing significant effort into improving hadronically-decaying massive particle taggers, as many searches for physics beyond the SM will benefit. CMS has done exactly this by devising new taggers based on exploiting recent machine learning developments, building DNNs for both Higgs boson double-*b*-taggers [15] as shown in Figure 17 and DNNs for top-quark identification [16] as shown in Figure 18. In both cases, the DNN tagger drastically outperforms the reference tagger. These are only two of many examples; it is clear that machine learning is playing an increasingly prominent role in the identification of jets from hadronically-decaying massive particles.

Mass scan searches depend on both the ability to reject the QCD background and the precision of the jet four-vector. In particular, it is important to have a very precise knowledge of the jet mass, as this is the primary variable use to study the presence of potential new signatures of physics
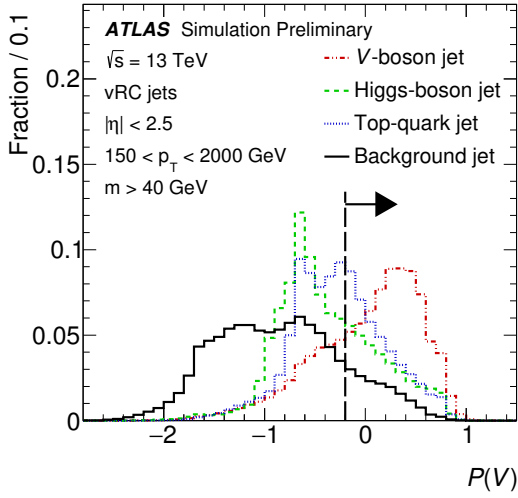
Figure 15: The multi-class classifier used for discriminating between jets from W/Z (Vector) bosons, Higgs bosons, top-quarks, and the QCD background, showing moderate separation power between hadronically-decaying massive particles [14].
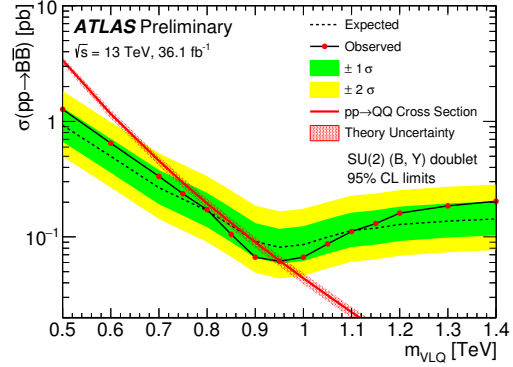

Figure 16: Limits on the production of VLQ $B\bar{B}$ pairs in the zero-lepton final state, where the sensitivity decays away from the 1 TeV mass point because the analysis was optimized directly for the one point rather than over a range of points [14].
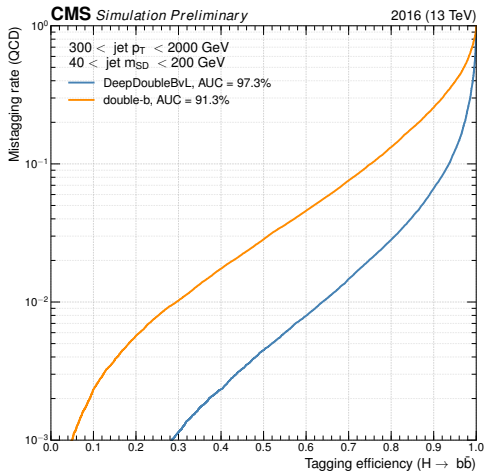

Figure 17: The classification power of the Deep-DoubleB tagger clearly demonstrates the enormous power of machine learning as applied to hadronic Higgs boson identification [15].
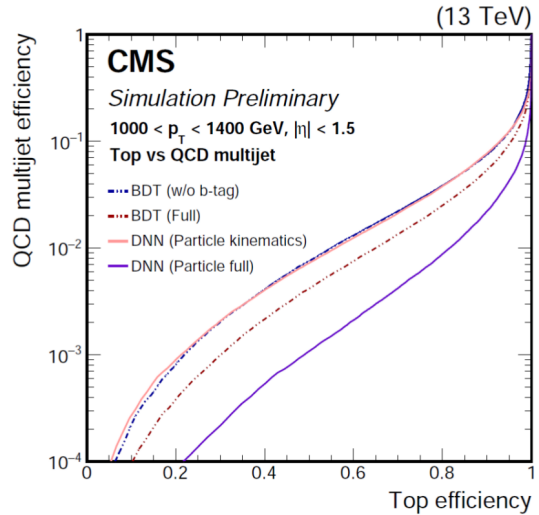

Figure 18: The classification power of different top taggers shows the importance of using Deep Neural Networks (DNNs) using full particle information to obtain the best top-quark identification power [16].

beyond the SM. ATLAS is thus working on bringing large jets into the precision regime [17], with percent-level p_T scale uncertainties over most of the parameter space as seen in Figure 19 and few-percent-level mass scale uncertainties as shown in Figure 20. The p_T scale is evaluated in multiple topologies by balancing the large jet against a reference object, while the mass scale is evaluated by fitting the W boson and top-quark mass peaks in semi-leptonic $t\bar{t}$ events and extrapolated to other kinematic regions using the R_{trk} procedure. The R_{trk} procedure uses double-ratios of the large jet mass as calculated using purely tracking detector or calorimeter information and then uses this ratio to constrain the mass scale differences between data and simulation.
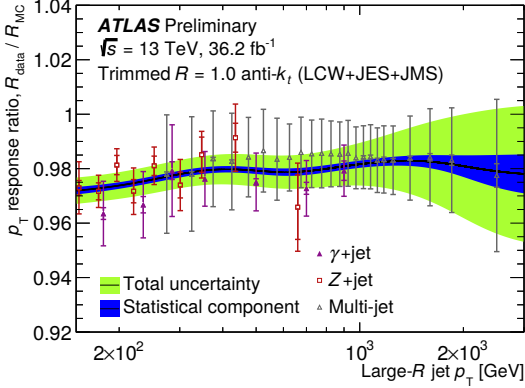

Figure 19: Multiple *in situ* balance techniques are combined to evaluate the p_T scale difference between data and MC, resulting in percent-level precision for most of the large jet parameter space [17].
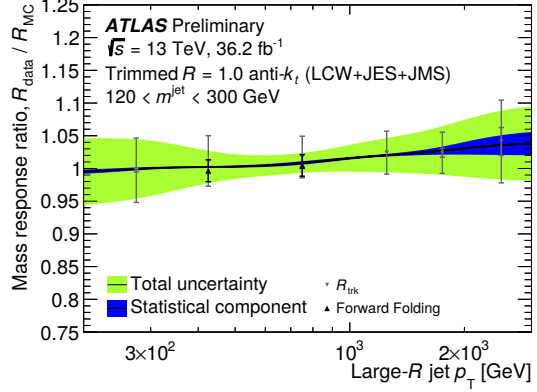
Figure 20: The forward folding technique is used to evaluate the large jet mass scale difference between data and MC, which is then extrapolated to other kinematic ranges using the R_{trk} procedure [17].

7. Summary

Jet substructure is playing an increasing prominent role in the ATLAS and CMS physics programs, including the search for new physics beyond the standard model. Of the many searches using jet substructure that have been published this year, they can be primarily divided into three categories: dijet resonance searches, mass scan searches, and VLQ pair-production searches. These three types of searches benefit from the increased knowledge of the ATLAS and CMS detectors, as well as an improved understanding of how to reconstruct jets and jet properties. Searches using jet substructure are typically limited by the ability to identify the hadronic decays of massive particles, notably W bosons, Z bosons, Higgs bosons, and top-quarks, and machine learning appears to be a promising means of significantly improving jet taggers. Some of the searches are also sensitive to the precision of the jet four-vector, and thus increasingly advanced methods are being devised to constrain the jet p_T scale to the percent-level and the jet mass scale to the few-percent-level. These advances in jet identification power and precision jet four-vector evaluation is expected to both further improve existing searches using jet substructure and create new jet substructure searches.

References

[1] ATLAS collaboration, *ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) 003.

[2] CMS collaboration, *The CMS Experiment at the CERN LHC*, *JINST* **3** (2008) 004.

[3] CMS collaboration, *Identification techniques for highly boosted W bosons that decay into hadrons*, *JHEP* **12** (2014) 017, [`1410.4227`].

[4] ATLAS collaboration, *Identification of boosted, hadronically decaying W bosons and comparisons with ATLAS data taken at $\sqrt{s} = 8$ TeV*, *Eur. Phys. J.* **C76** (2016) 154, [`1510.05821`].

[5] A. J. Larkoski, I. Moult and D. Neill, *Power Counting to Better Jet Observables*, *JHEP* **12** (2014) 009, [`1409.6298`].

[6] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with N-subjettiness*, *JHEP* **03** (2011) 015, [`1011.2268`].

[7] ATLAS collaboration, *Search for diboson resonances in hadronic final states in 79.8 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *ATLAS-CONF-2018-016* (2018) .

[8] CMS collaboration, *Search for resonant and non-resonant production of Higgs boson pairs in the four b quark final state using boosted jets in proton-proton collisions at sqrt(s)=13 TeV*, *CMS-PAS-B2G-17-019* (2018) .

[9] CMS collaboration, *Search for production of Higgs boson pairs in the four b quark final state using large-area jets in proton-proton collisions at $\sqrt{s} = 13$ TeV*, `1808.01473`.

[10] ATLAS collaboration, *Search for light resonances decaying to boosted quark pairs and produced in association with a photon or a jet in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, `1801.08769`.

[11] J. Dolen, P. Harris, S. Marzani, S. Rappoccio and N. Tran, *Thinking outside the ROCs: Designing Decorrelated Taggers (DDT) for jet substructure*, *JHEP* **05** (2016) 156, [`1603.00027`].

[12] CMS collaboration, *Search for low mass resonances decaying into bottom quark-antiquark pairs in pp collisions at $\sqrt{s} = 13$ TeV*, *CMS-PAS-EXO-17-024* (2018) .

[13] CMS collaboration, *Search for vector-like T and B quark pairs in final states with leptons at $\sqrt{s} = 13$ TeV*, *JHEP* **08** (2018) 177, [`1805.04758`].

[14] ATLAS collaboration, *Search for pair production of heavy vector-like quarks decaying into hadronic final states in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, `1808.01771`.

[15] CMS collaboration, *Performance of Deep Tagging Algorithms for Boosted Double Quark Jet Topology in Proton-Proton Collisions at 13 TeV with the Phase-0 CMS Detector*, *Public plots* (2018) .

[16] CMS collaboration, *Boosted jet identification using particle candidates and deep neural networks*, *CMS-DP-2017-049* (2017) .

[17] ATLAS collaboration, *In situ calibration of large-R jet energy and mass in 13 TeV proton-proton collisions with the ATLAS detector*, *Submitted to: Eur. Phys. J.* (2018) , [`1807.09477`].