Improving jet substructure performance in ATLAS using Track-CaloClusters

The ATLAS Collaboration

Jet substructure techniques play a critical role in ATLAS in searches for new physics, are increasingly important in measurements of the Standard Model, and are being utilized in the trigger. To date, ATLAS has mostly focused on the use of calorimeter-based jet substructure, which works well for jets initiated by particles with low to moderate boost, but which lacks the angular resolution needed to resolve the desired substructure in the highly-boosted regime.

We present a novel approach designed to mitigate the calorimeter angular resolution limitations, thus providing superior performance to prior methods. Similarly to the previously developed combined mass technique, the superior angular resolution of the tracker is combined with information from the calorimeters. However, the new method is fundamentally different, as it correlates low-level objects such as tracks and individual energy deposits in the calorimeter, before running any jet finding algorithms. The resulting objects are used as inputs to jet reconstruction, and in turn result in improved resolution for both jet mass and substructure variables. Preliminary studies indicate that these jets may be sufficiently robust against pile-up without any grooming applied, due to the pile-up rejection capabilities of the tracker.

1 Introduction

Jet substructure techniques have become an integral piece of the ATLAS [1] physics agenda, and are only becoming more important in Run 2. A common use of jet substructure is in the tagging of the hadronic decays of W/Z bosons [2], H→bb [3], and top quarks [4], as well as for discrimination between light quark and gluon initiated jets [5]. These tagging methods are then used in a plethora of searches for new physics, such as diboson resonances [6, 7], as well as measurements of Standard Model processes [8]. They have even recently been integrated into the ATLAS trigger [9], allowing for the experiment to record a sample of jets enriched in hadronic massive particle decays. ATLAS has also used jet substructure variables as the inputs to both Boosted Decision Trees (BDTs) and Deep Neural Networks (DNNs) to further improve the identification of hadronically decaying W bosons and top quarks [10].

As the name implies, jet substructure is the study of resolvable energy structures within a jet, and is a powerful way to discriminate between hadronic decays of massive particles and QCD multi-jet events. As an example, the decay of a Z boson to a pair of quarks generally produces two significant energy deposit regions in the detector, while a single quark or gluon from a QCD multi-jet event results in primarily one significant energy deposit region. As the momentum of the Z boson is increased, the decay system is increasingly boosted with respect to the experimental reference frame, and thus the two quarks are increasingly collimated until they are reconstructed as a single large-R jet. At this point, jet substructure becomes a powerful means of identifying the source of the large-R jet.

While the energy deposits of the two quarks in this example are now contained within a single jet, it is important that they can still be differentiated with respect to the single quark or gluon of a QCD multi-jet event. This relies on both the energy and angular resolution of the detector used to reconstruct the jet. To date, ATLAS has mostly focused on the use of calorimeter-based jet substructure, which exploits the exceptional energy resolution of the ATLAS calorimetry [11]. However, as the event becomes even more energetic, jets become so collimated that the calorimeter lacks the angular resolution to resolve the desired structure within the jet. On the other hand, tracking detectors have excellent angular resolution and good reconstruction efficiency at very high energy [12], while their energy resolution deteriorates. To significantly improve the jet mass resolution for highly energetic jets, ATLAS previously performed a simple linear combination of calorimeter and track quantities [13].

This first simple approach is useful in that it improves the mass resolution, but there are several aspects which can be improved upon. By more rigorously unifying information from the ATLAS calorimeter and tracking detectors, the precision of jet substructure techniques can be further improved for a wide range of energies, thus ensuring the continued viability of jet substructure within the ATLAS physics program. Algorithms using tracking detectors have been previously suggested in the context of top quark jet tagging, mainly to achieve an improved mass resolutions [14]. Since that time, there have been phenomenological studies using track-assisted jet substructure for highly boosted boson and top quark jets [15–17]. These methods all rely on matching the tracks of charged particles to (sub-)jets build from energy deposits in the calorimeter.

This document proposes a new unified object built from both tracking and calorimeter information, referred to as Track-CaloClusters (TCCs). This procedure is a type of particle flow, but it is distinct in several key areas from the energy subtraction algorithm described in the recent ATLAS particle flow publication [18]. In particular, the 4-vector construction and energy sharing procedures are very different. In contrast to the energy subtraction algorithm, energy sharing in the TCC approach is addressed solely based on the weighting scheme described in Section 4.3. In practice, this means that the TCC algorithm uses the spatial

coordinates of the tracker and the energy scale of the calorimeter, while the energy subtraction algorithm allows the tracking detector measurements to correct the full 4-vector. There are additional differences, such as in the track-cluster matching criteria and kinematic regime for which they are optimized, but in general both procedures aim to improve the understanding of hadronic showers by combining tracker and calorimeter information.

Section 2 describes the ATLAS detector, while Section 3 details the Monte Carlo simulation samples used for this study. The focus then shifts to the reconstruction and performance of TCCs in Section 4. The resulting jet substructure performance using jets built from TCCs is presented in Section 5, before finishing with a conclusion in Section 6.

2 ATLAS detector

The ATLAS detector [19] is a multi-purpose particle detector with nearly 4π coverage in solid angle.[1] Starting from the center, ATLAS consists of an inner tracking detector, sampling electromagnetic and hadronic calorimeters, and a muon spectrometer for precision muon reconstruction out to $|\eta| < 2.7$.

The inner tracking detector is immersed in a 2 T field, provided by a solenoid, and is used to measure charged-particle tracks. The inner detector is composed of pixel and silicon microstrip detectors covering the pseudorapidity range $|\eta| < 2.5$ and a transition radiation tracker covering the pseudorapidity range $|\eta| < 2.0$.

A series of calorimeters enclose the solenoid and inner detector providing high precision electromagnetic and hadronic energy measurements. Lead and liquid argon electromagnetic sampling calorimeters are used in both the barrel ($|\eta| < 1.5$) and endcap ($1.4 < |\eta| < 3.2$) regions, while copper and liquid argon is used for the forward region ($3.1 < |\eta| < 4.9$). Hadronic calorimeters use steel and scintillating tiles in the barrel ($|\eta| < 1.0$) and extended barrel regions ($0.8 < |\eta| < 1.7$), copper and liquid argon for the endcaps ($1.5 < |\eta| < 3.2$), and tungsten and liquid argon for the forward region ($3.1 < |\eta| < 4.9$). For the high p_{T} regime of primary interest in this document, the hadronic calorimeters are critical. The typical hadronic calorimeter cell size within the barrel and extended barrel regions is roughly 0.1×0.1 in $\Delta\eta \times \Delta\phi$.

3 Monte Carlo simulation samples

Samples of hadronically decaying W and Z bosons are created by simulating the fully hadronic decay $W' \to WZ \to qqqq$ and changing the mass of the W' in order to probe different regions of p_{T}. The W' mass is varied in steps of 0.2 TeV from 0.8 TeV to 1.2 TeV, steps of 0.1 TeV from 1.2 TeV to 3 TeV, and steps of 0.2 TeV from 3 TeV to 5 TeV. The W and Z bosons produced in this decay chain are both longitudinally polarized. All such simulated samples are produced using the Pythia8 Monte Carlo generator [20] with the A14 event tune [21] and the NNPDF2.3 Leading Order (LO) Parton Distribution Function (PDF) [22].

---

[1] ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the z-axis along the beam pipe. The x-axis points from the IP to the centre of the LHC ring, and the y-axis points upward. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z-axis. The pseudorapidity is defined in terms of the polar angle θ as $\eta = -\ln\tan(\theta/2)$.

The resulting simulated samples are processed using the full GEANT4 [23] simulation of the ATLAS detector [24]. The impact of additional proton-proton interactions (pile-up), both in the same bunch crossing and adjacent bunch crossings, is simulated by overlaying an average of roughly 24 minimum bias events generated using PYTHIA8 with the A2M event tune [25] and the MSTW2008 LO PDF [26].

4 Track-CaloClusters

4.1 Motivation

Tracking detectors and calorimeters have complimentary behaviour, and the idea of TCCs is to maximally exploit the strengths of each detector for improved measurements of hadronic interactions in highly energetic hard-scatter processes. Boosted hadronic decays provide a particularly challenging environment. In the simple example of a hadronically decaying vector boson V (representing both the W and Z bosons), $V \rightarrow qq$, the two quarks become increasingly collimated as the parent particle becomes more energetic, following the approximation $\Delta R \approx 2m_V/p_T^V$ where $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$.

Discrimination between such hadronically decaying particles and the overwhelming QCD multi-jet background depends crucially on both the energy and angular resolution of the inputs to jet reconstruction. In ATLAS, these inputs are typically locally calibrated (LC) topo-clusters built from noise-suppressed topological groupings of calorimeter cells [11]. The ATLAS calorimeter has an excellent energy resolution, which improves as the energy is increased until reaching a constant value, so long as the full hadronic shower is contained within the calorimeter. However, the granularity of the calorimeter is insufficient to resolve the angular separation between highly boosted hadronic decays of massive particles, and thus the angular resolution can be a limiting factor in the construction of variables sensitive to the differences between hadronically decaying massive particles and QCD multi-jet events.

In contrast, the ATLAS tracking detector has an excellent angular resolution when evaluated at the perigee. As the tracks become more energetic, they are less curved by the magnetic field, and thus their extrapolation to the calorimeter is also increasingly precise. However, the momentum of the tracks is measured from their curvature. As such, improving the precision of the track extrapolation to the calorimeter is tied to a degraded p_T resolution. Inefficiencies in the reconstruction of tracks in the p_T range considered are very small, and are not expected to impact the results [12].

Tracks and topo-clusters are at a different energy scale, as tracks represent only electromagnetic energy from charged particles while LC topo-clusters have been corrected on average to the hadronic scale, with a per-topo-cluster scale factor of order one. As a result, if the track p_T and the cluster p_T are the same value, then the track typically represents a higher energy particle. Rather than using truth information to compare the same kinematic values, it is instead important to retain these reconstruction level differences to understand the impact on jet building, as discussed in Section 5.1.

The precision of good quality tracks [12] from the selected hard scatter primary vertex[2] when extrapolated to the calorimeter is compared to topo-cluster angular width in Figure 1. This comparison is done using a $W' \rightarrow WZ \rightarrow qqqq$ sample with a W' mass of 1 TeV. The track extrapolation uncertainty is defined using the covariance matrix at the calorimeter entry, which is derived by extrapolating the track through

---

[2] Of the multiple pp collision vertices reconstructed from the available tracks in a given event, the hard scatter primary vertex is selected as the one with the largest $\sum p_T^2$, where the sum is over all tracks with $p_T > 0.4$ GeV that are associated with the vertex.
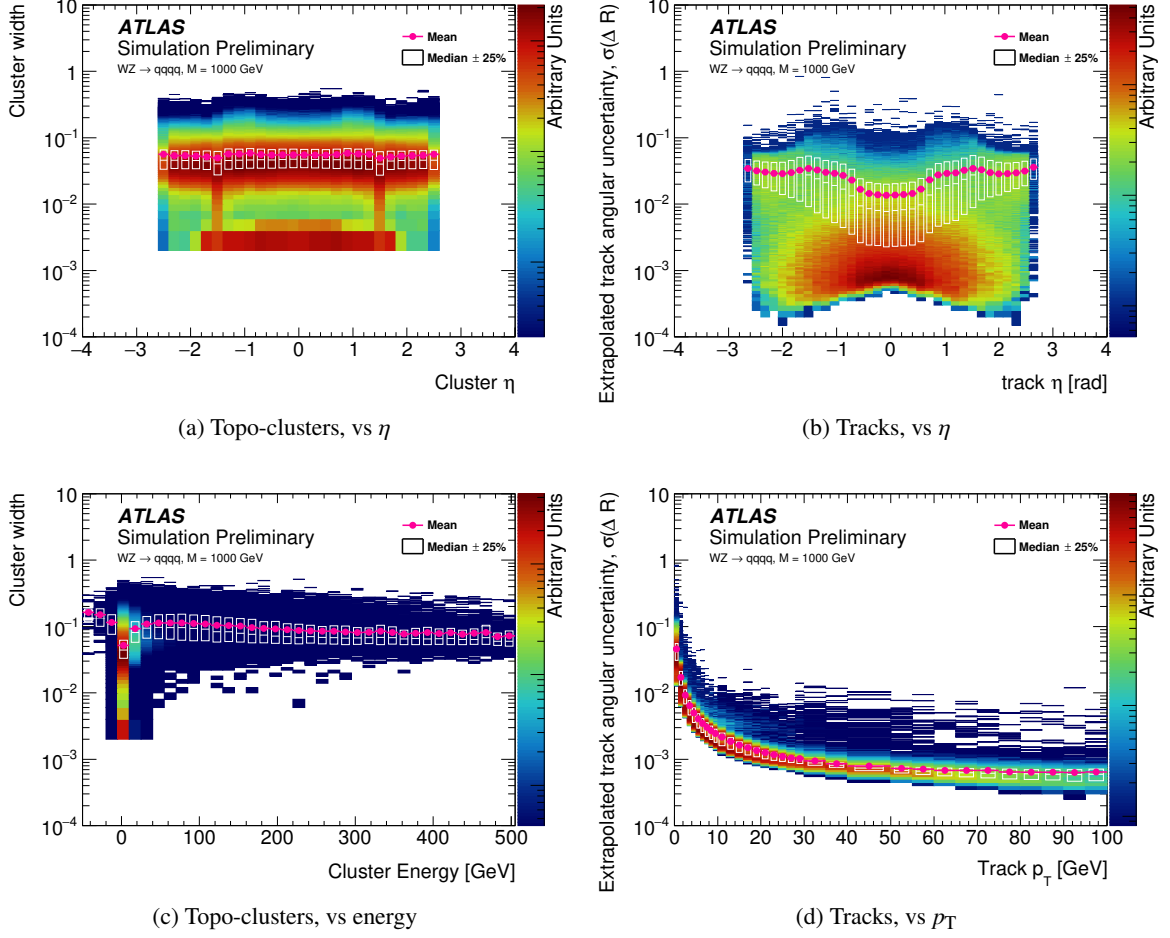
(a) Topo-clusters, vs η

(b) Tracks, vs η

(c) Topo-clusters, vs energy

(d) Tracks, vs p_T

Figure 1: A comparison of the topo-cluster width in W′ → WZ → qqqq events with a W′ mass of 1 TeV as a function of (a) topo-cluster η and (c) topo-cluster energy, and the uncertainty on the extrapolation of tracks to the calorimeter as a function of (b) track η and (d) track p_T. The comparison demonstrates that the extrapolated track position is typically known to much higher precision than the size of a cluster. As such, energetic tracks provide a means of more accurately specifying the location of an energy deposit than the corresponding cluster. Note that the distributions are sometimes highly non-Gaussian, so the mean value often does not align with the median of the distribution. This can be seen by comparing the mean (pink points) to the 50% median window (white boxes).

the full ATLAS detector material model using the complete magnetic field representation. Topo-clusters are larger objects, and their width is defined as follows:

$$\sigma_\eta \approx \sigma_\phi \approx \mathrm{atan}\left(\frac{\sqrt{\langle r^2 \rangle}}{|\vec{c}|}\right) \times \cosh(\eta) \ . \tag{1}$$

For the motivation of this width formulation and details on the topo-cluster variables from which it is defined, please see Reference [11]. The typical extrapolated track angular uncertainty is seen to be significantly smaller than the size of the average topo-cluster across the detector and for moderate to high p_T tracks. This increasing precision in the extrapolation is important and will be revisited later.

This forms the basis of the TCC. In order to maximally benefit at high p_T from the superb calorimeter energy resolution and excellent tracker spatial resolution, a new 4-vector is formed which (to first order) uses topo-clusters for the scale components (p_T, m - where the mass of a single topo-cluster in ATLAS is defined to be zero) while using the track parameters at the perigee to determine the angular (η, ϕ) coordinates. As the perigee angular uncertainty is smaller than the extrapolated angular uncertainty by construction, Figure 1 serves as an upper bound on what can be expected.

4.2 Matching

In order to build 4-vectors from a combination of track and topo-cluster information, a well-motivated track-cluster matching criterion must be defined. The TCC matching procedure attempts to match every good quality track to every topo-cluster following two steps. In the first stage, the uncertainty on the track extrapolation to the calorimeter is compared to the width of the topo-cluster. If the extrapolation uncertainty is larger than the topo-cluster width, then the track is discarded from the matching procedure. Otherwise, the matching continues to the second step, where a track-cluster pair is defined as matched whenever their angular separation $\Delta R < \sqrt{\sigma_{\text{cluster}}^2 + \sigma_{\text{track}}^2}$ with σ_{cluster} the topo-cluster width and σ_{track} the track extrapolation uncertainty. As previously shown in Figure 1, the track extrapolation uncertainty reduces significantly at high p_T, such that the second matching criterion becomes $\Delta R \lesssim |\sigma_{\text{cluster}}|$ in the high p_T limit.

Topo-cluster width is the relevant quantity rather than topo-cluster angular resolution as the intent is to resolve hadronically decaying particles at extremely high p_T, such events where a W boson is so boosted that, in the most extreme case, it falls within a single topo-cluster. In this case, the topo-cluster angular position is precisely known, as it is just a measure of how well the centroid of the cluster can be determined, and there is a lot of energy to fix the centroid. By matching tracks to this highly energetic topo-cluster, the aim is to resolve the hidden topo-cluster substructure and identify two regions of energy corresponding to the two quarks from the W decay.

The matching procedure does not attempt to remove tracks from leptons, and it makes use of tracks originating from any primary vertex, not only of tracks from the selected hard scatter primary vertex. While the tracking detector is able to differentiate between multiple in-time collisions, the calorimeter cannot, and thus tracks from multiple pp collision vertices may be matched to the same topo-cluster. In this way, tracks from pile-up are used to address energy sharing. These track-cluster matches are used to construct TCC objects, which fall into three categories. First, tracks compatible with the primary vertex and matched to topo-clusters are referred to as combined TCCs. Topo-clusters not matched to tracks from any vertex, and tracks from the primary vertex not matched to clusters, are also considered and are referred to as neutral and charged TCCs respectively. Tracks from pile-up vertices matched to topo-clusters are not used to build independent TCC objects.

The resulting fractions of neutral, charged, and combined TCCs when applied to $\text{W}' \rightarrow \text{WZ} \rightarrow \text{qqqq}$ events with a W$'$ mass of 1 TeV are shown in Figure 2. There are a large number of low-p_T neutral TCCs (from unmatched topo-clusters), while the number of charged TCCs (unmatched tracks) is small for all p_T. By $\sim 5\,\text{GeV}$, the combined TCC becomes the most common scenario, and this continues to rise as the p_T increases. This is expected as highly energetic topo-clusters are likely from the hadronically decaying W or Z boson and thus close to tracks. Multiple tracks can often be matched to such topo-clusters. At high enough boost, both neutral and charged decay products of the boson are highly collimated and often contribute to the same topo-clusters. At low energies, topo-clusters are primarily from pile-up.
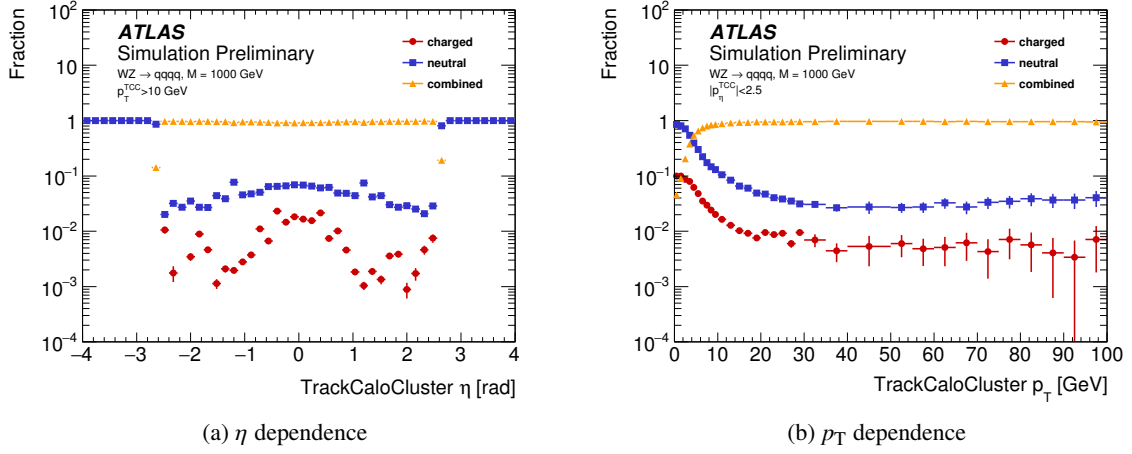
(a) η dependence　　　　　　　　(b) p_T dependence

Figure 2: The fractions of different Track-CaloCluster classes are shown as a function of (a) the TCC η and (b) the TCC p_T. Combined objects (triangles) represent tracks from the selected hard scatter primary vertex which are matched to topo-clusters, neutral objects (squares) are for topo-clusters not matched to tracks from any primary vertex, and charged objects (circles) are tracks from the selected hard scatter primary vertex not matched to any topo-cluster.

Furthermore, there are no shapes in the combined TCC η distribution within the acceptance of the tracking detector, confirming that the matching criteria is generally insensitive to the detector structure.

4.3 Reconstruction

Once tracks and topo-clusters have been matched, the next stage is to build the actual TCC 4-vectors. Examples are provided below, aligned with the schematic shown in Figure 3 in order to assist in understanding the procedure. When referring to the figure, ① refers to TCC object 1, c_1 refers to topo-cluster c_1, and t_1 refers to track t_1.

For a direct single match between a track from the selected hard scatter vertex and a topo-cluster, such as ①, the topo-cluster energy and the track direction are used to form a single TCC:

$$\text{TCC}_① = (p_T^{c_1}, \eta^{t_1}, \phi^{t_1}, m^{c_1} = 0) \,. \tag{2}$$

In the case of topo-clusters which do not match any tracks, such as ②, the topo-cluster 4-vector is directly used to create a TCC:

$$\text{TCC}_② = (p_T^{c_7}, \eta^{c_7}, \phi^{c_7}, m^{c_7} = 0) \,. \tag{3}$$

Independent tracks from the selected hard scatter vertex which do not match any topo-clusters, such as ③, are also used directly to create TCCs:
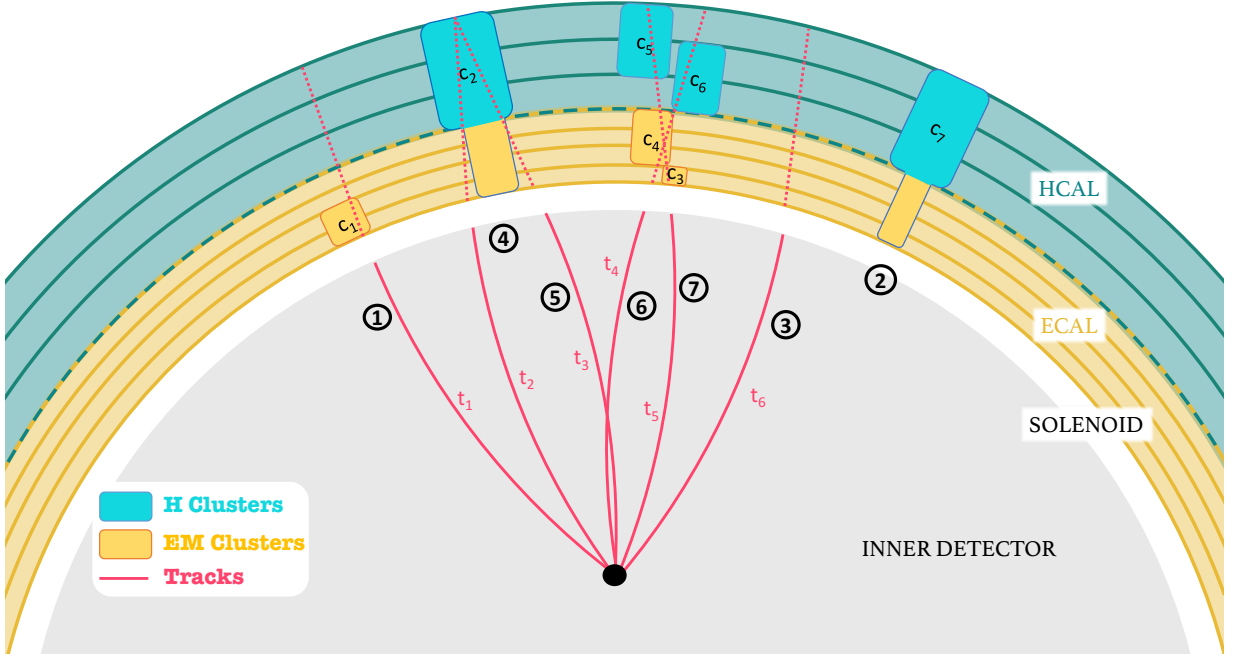
Figure 3: A schematic demonstrating the creation of seven TCC objects representing ① a simple track-cluster match, ② a topo-cluster without a matching track, ③ a track without a matching cluster, ④ and ⑤ are each tracks matching a single cluster but sharing that cluster's energy, and ⑥ and ⑦ showing a much more complex scenario with multiple track-cluster matches. Details on the exact reconstruction procedure and the seven TCC 4-vectors are provided in the text.

$$\text{TCC}_{③} = (p_T^{t_6}, \eta^{t_6}, \phi^{t_6}, m^{t_6} = 0) \ . \tag{4}$$

Once there is a match between multiple tracks and a single cluster, multiple topo-clusters and a single track, or especially multiple topo-clusters with multiple tracks, as in the case of ⑥ and ⑦, the situation becomes more complex. The TCC reconstruction procedure still creates exactly one TCC object per track originating from the primary vertex, where the track angular coordinates are used, but the scale coordinates must be adapted to account for energy sharing between the different matches.

For each track used to seed a TCC, all matching topo-clusters are found. For each of those matching clusters, the energy is then divided between all of the tracks which match that cluster, with the split defined by the fraction of p_T contributed by a given track compared to all of the other matching tracks. This provides a means of splitting the topo-cluster energy between tracks without requiring a precise measurement of the track p_T, as the p_T of any individual track only enters as a ratio with respect to the p_T of the sum of all matching tracks. By only using ratios, the algorithm is less sensitive to the scale of any particular track, and the more important metric is whether or not any individual track represents a significant fraction of the total p_T. Furthermore, this means that the tracker and calorimeter energy measurements are never directly compared, rather only scale factors derived from one detector are used to weight the energy split from the other detector.

As a simple starting example where there are two tracks matching a single topo-cluster, consider the case of ④ and ⑤. Regarding notation, $\mathbf{p}^a$ is the 4-vector corresponding to a given particle a, while $p_T[\mathbf{p}^a + \mathbf{p}^b]$

means the p_T of the 4-vector sum of a and b. The TCC 4-vectors, recalling that the mass of a single topo-cluster in ATLAS is defined to be zero, are thus:

$$\text{TCC}_{\textcircled{4}} = \left(p_T^{c2} \frac{p_T^{t_2}}{p_T\left[\mathbf{p}^{t_2} + \mathbf{p}^{t_3}\right]}, \eta^{t_2}, \phi^{t_2}, m^{c2} \frac{p_T^{t_2}}{p_T\left[\mathbf{p}^{t_2} + \mathbf{p}^{t_3}\right]} = 0 \right), \tag{5}$$

$$\text{TCC}_{\textcircled{5}} = \left(p_T^{c2} \frac{p_T^{t_3}}{p_T\left[\mathbf{p}^{t_2} + \mathbf{p}^{t_3}\right]}, \eta^{t_3}, \phi^{t_3}, m^{c2} \frac{p_T^{t_3}}{p_T\left[\mathbf{p}^{t_2} + \mathbf{p}^{t_3}\right]} = 0 \right). \tag{6}$$

The TCC energy sharing equation can be generalized to handle a seed track τ matched to a set of topo-clusters C_τ. Each topo-cluster c in C_τ can then be matched to a set of tracks T_c. Furthermore, each of those tracks t in T_c can be matched to a set of topo-clusters C_t. In general, the notation X_y represents the set of objects of type X matched to a given object y, where X is either C for topo-clusters or T for tracks.

The energy sharing procedure makes use of three general concepts. First, each cluster c matched to the seed track τ should contribute to the resulting TCC object proportionally to its p_T fraction f_τ^c out of all matched clusters. Second, each cluster c may contribute to multiple TCC objects, so its contribution to any given TCC should be weighted by the fraction of p_T demanded by the seed track τ compared to all other tracks t matched to the cluster, $\mathcal{F}_{c,t}^\tau$. Third, the contribution of each of those tracks t should itself be weighted by the fraction of energy that the cluster c represents compared to all other clusters matching the track, f_t^c. The equations for these three scalar values representing p_T fractions are:

$$f_\tau^c = \frac{p_T^c}{p_T\left[\displaystyle\sum_{k \in C_\tau} \mathbf{p}^k\right]}, \qquad \mathcal{F}_{c,t}^\tau = \frac{p_T^\tau}{p_T\left[\displaystyle\sum_{t \in T_c} \mathbf{p}^t f_t^c\right]}, \qquad f_t^c = \frac{p_T^c}{p_T\left[\displaystyle\sum_{k \in C_t} \mathbf{p}^k\right]}. \tag{7}$$

Roughly speaking, the first term f_τ^c represents the relevance of the topo-cluster c compared to all of the topo-clusters matching the seed track τ. If the p_T of c is small compared to the set of matched topo-clusters, then this cluster is less relevant, and the contribution of c to the final TCC is small. On the other hand, the term $\mathcal{F}_{c,t}^\tau$ represents the relevance of the seed track τ when compared to all other tracks matching the cluster c, which is essentially a normalization factor representing the amount of energy that the topo-cluster is able to contribute to any given TCC based on the number of other objects that it matches. Applying this to the actual topo-cluster 4-vectors matching the seed track τ results in the final TCC energy sharing equation:

$$\mathbf{M}_\tau = \sum_{c \in C_\tau} \mathbf{p}^c \, f_\tau^c \, \mathcal{F}_{c,t}^\tau = \sum_{c \in C_\tau} \mathbf{p}^c \frac{p_T^c}{p_T\left[\displaystyle\sum_{k \in C_\tau} \mathbf{p}^k\right]} \frac{p_T^\tau}{p_T\left[\displaystyle\sum_{t \in T_c} \mathbf{p}^t \frac{p_T^c}{p_T\left[\displaystyle\sum_{k \in C_t} \mathbf{p}^k\right]}\right]}. \tag{8}$$
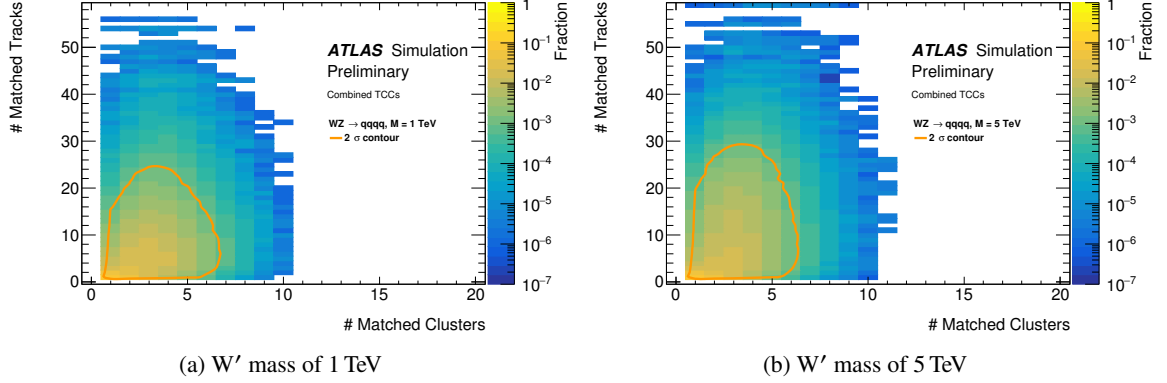
9

Figure 4: A comparison of the number of topo-clusters matching the seed track when building a TCC object, and the number of tracks matched to each one of those clusters, for $W' \to WZ \to qqqq$ events with two different W' masses. An orange contour encloses the region representing 2σ of the matches.

The 4-vector $\mathbf{M}_\tau$ resulting from the energy sharing equation is then used to define the scale components of the final TCC 4-vector, while the angular components come from the seed track τ:

$$\text{TCC}_\tau = (p_\text{T}[\mathbf{M}_\tau], \eta^\tau, \phi^\tau, m[\mathbf{M}_\tau]) . \tag{9}$$

This formula handles all TCC cases in which there is a match between track(s) and topo-cluster(s), and reduces to the simpler cases previously described. Written in this form, it is clear that the p_T and mass of the TCC 4-vector is defined by the 4-vector sum of matched topo-clusters times a pair of previously described analytically defined scalar weights. A consequence of this formula is that the combined TCC mass is zero if there is only one topo-cluster matching the seed track, and non-zero if there are at least two topo-clusters matching the seed track.

In Equation 8, there are two different restrictions on the tracks. The seed track, τ, must originate from the primary vertex. However, the tracks used in the energy sharing scale factor, t, can come from any vertex. This difference means that combined TCCs must spatially originate from the hard-scatter interaction, but the scale of such objects will take energy overlaps in the calorimeter into account. Examples of how to use the energy sharing algorithm are provided in Appendix A, where the utility of the p_T fractions defined in Equation 7 is made clear.

Following the construction of TCCs, it is natural to ask about the typical sizes of the sets C_τ and T_c. An example is shown from the $W' \to WZ \to qqqq$ with W' masses of 1 TeV and 5 TeV in Figure 4. The 2σ bulk of the distribution lies within the region bounded by C_τ containing 6 clusters and T_c containing 25 to 30 tracks depending on the W' mass. The plot contains one entry per seed-matched topo-cluster in each combined TCC object. That is, if a single combined TCC object seeded by the track τ contains N_c clusters in C_τ and each of those clusters is matched to N_t tracks in T_c, then the plot is filled N_c times with a fixed x-axis value of N_c and a variable y-axis value of N_t.

It is worth noting that there are some very large multiplicity matches in the tails of Figure 4. In particular, there are cases where single topo-clusters are matched to over 50 tracks. This is very rare, as evidenced from the logarithmic colour scale, and is roughly six orders of magnitude suppressed compared to cases

where topo-clusters are matched to only a few tracks. However, it does occur, and there are two reasons for this. First, while the seed track must originate from the primary vertex, there is no such requirement applied to the tracks used in the energy sharing procedure. As such, a moderate fraction of those matches are typically from additional simultaneous vertices. Second, the more significant effect is typically the size of the cluster. The matching criterion depends on the width of the cluster, and topo-clusters can grow to be quite wide in rare cases, thus the number of tracks matched to such large topo-clusters can be enormous. As these are rare cases, they do not significantly affect any of the results in this document.

5 Jet substructure performance

5.1 Jet reconstruction

ATLAS uses large-R jets built using the anti-k_t algorithm [27] with a distance parameter R of 1.0. Reconstructed large-R jets are typically built using LC topo-clusters, and are subsequently trimmed [28] to remove contributions from pile-up using the parameters $f_{cut} = 5\%$ and $R_{sub} = 0.2$. It is often useful to also create truth jets following the same procedure, but with inputs of stable truth particles ($c\tau > 10\,\mathrm{mm}$) excluding muons and neutrinos, and excluding particles from pile-up.

Different sets of TCC objects can be substituted for topo-clusters as the inputs to jet reconstruction. There are two particularly interesting selections of TCC inputs to jet reconstruction: all TCCs or only combined TCCs. The former should capture the large majority of the hard scatter energy and provide the best representation of the total energy flow in the event, as there are both charged and neutral contributions. However, by including the neutral TCCs (unmatched topo-clusters), these jets will have a pile-up dependence similar to that of standard topo-cluster jets. The latter case of building jets only with combined TCCs is particularly interesting, as while some of the real hard scatter energy will be missed, the objects are significantly pile-up suppressed due to the track-to-primary-vertex matching requirement.

Two ways of quantifying the performance of a given quantity χ for a given reconstructed jet definition are used. The first is a response, defined as the ratio $\mathcal{R}^r = \chi_{\mathrm{probe}}/\chi_{\mathrm{ref}}$, and the second a residual, defined as the difference $\mathcal{R}^d = \chi_{\mathrm{probe}} - \chi_{\mathrm{ref}}$. The former is useful when discussing quantities such as the mass which have a scale, while the latter is useful for quantities such as η which do not have a scale (and where the truth value can be zero). In the majority of cases, the probe is a reconstructed jet definition, and the reference is the truth jet definition. However, in some cases, we will also discuss the pseudo-response $\mathcal{R}^r_p$ where the reference jet is a different reconstructed jet definition. This is useful when comparing TCC jets against the standard topo-cluster jet definition.

A comparison of three types of TCC jets is presented in Figure 5, where each type of reconstructed jet is matched to the corresponding truth jet. The jet mass response, $\mathcal{R}^r_m$, is then shown. Comparing the untrimmed and trimmed distributions for jets built using all TCCs shows how trimming significantly cleans up the mass response. Alternatively, jets built using only combined TCCs are seen to have a well-behaved mass response without the need for trimming.

Whether the jet is built from topo-clusters, all TCCs, or only combined TCCs, a calibration using dedicated correction factors is needed to account for the non-uniform and non-compensating energy response of the ATLAS calorimeters. Topo-cluster jets are fully calibrated using both Monte Carlo derived jet energy and mass scale corrections [13, 29], while TCC jets are used at the uncalibrated scale for this preliminary
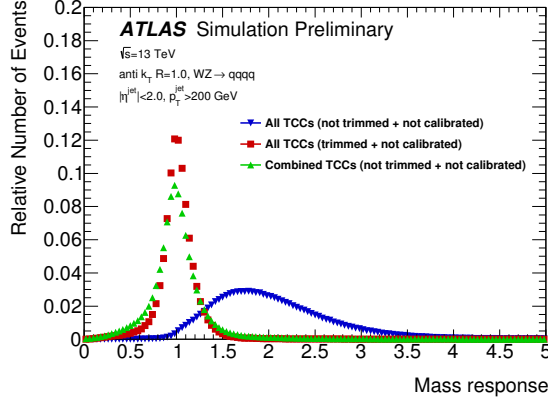
Figure 5: A comparison of the the mass response, $\mathcal{R}_m^r$, for jets built with all TCCs (blue upside-down triangles), all TCCs after trimming (red squares), and only combined TCCs (green triangles). Trimming is seen to be necessary for jets built from all TCCs, while the response of combined TCC jets is already good without trimming.

study. This difference has a minimal impact on the performance comparisons made in this note, and is discussed where relevant.

To demonstrate that the TCC jet definitions are under control, the p_T pseudo-response $\mathcal{R}_p^r$ is plotted in Figure 6(a), taking the trimmed uncalibrated topo-cluster jet definition as a reference. In this plot, the expectation for the all-TCC collection is to be peaked at $\mathcal{R}_p^r(p_T) \approx 1$ as the numerator and denominator are at the same scale, with a small shift to smaller p_T coming primarily from the removal of pile-up topo-clusters matching tracks from pile-up primary vertices. The combined-TCC jet definition is very similar to the trimmed all-TCC jet definition, despite the lack of grooming, suggesting that the pile-up contribution is well controlled and the majority of the hard scatter topo-clusters are retained. The third curve representing untrimmed all-TCC jets is much more susceptible to pile-up, which increases the jet p_T, and thus values larger than one are expected and observed.

The η residual distribution for the three jet definitions is shown in Figure 6(b), where the calibrated topo-cluster jets are shown to have the best performance. The centroid of high energy topo-clusters and thus high energy topo-cluster jets is very well known, surpassing that of jets built with track spatial coordinates, due to the inclusion of both charged and neutral contributions. Topo-cluster jets have a better η residual than TCC jets, but TCCs are not aimed at improving the jet angular uncertainties. They are instead aimed at resolving substructure below the topo-clusters size scale.

The reconstruction of a single simulated event using topo-cluster jets, all TCC jets, and only combined TCC jets is shown in Figure 7. Both TCC techniques are seen to result in an additional subjet within one of the two large-R jets with respect to the topo-cluster jets, suggesting a slightly different reconstructed energy flow. This is visible as an additional grey region (subjet), with a white division between the two grey regions stating that they are two separate subjets. The combined TCC reconstruction is also seen to significantly suppress the total number of input objects to jet reconstruction, which is expected as it suppresses essentially all independent pile-up contributions in the calorimeter.

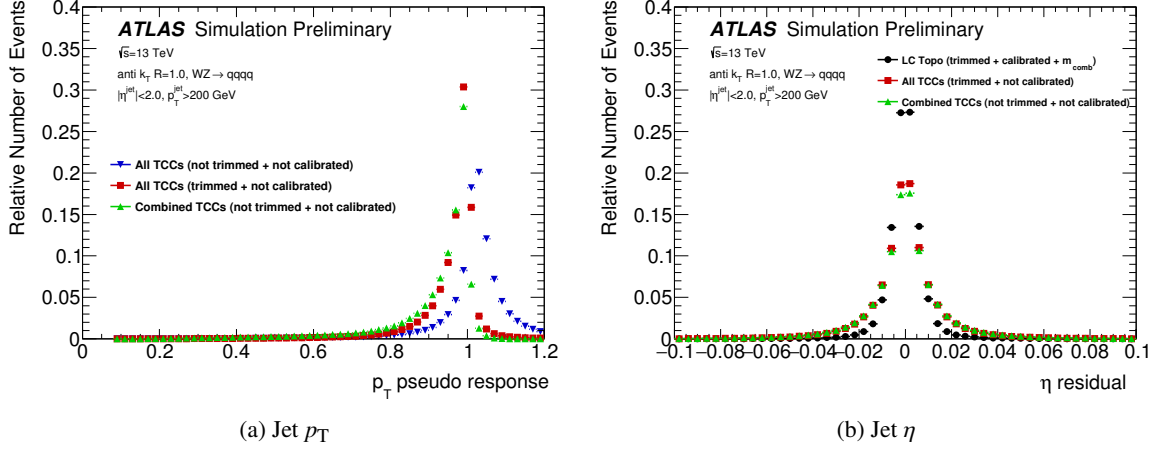|	
(a) Jet p_T	(b) Jet η

Figure 6: The (a) TCC jet uncalibrated p_T pseudo-response $\mathcal{R}_p^r(p_T)$ with respect to uncalibrated trimmed topo-cluster jets and (b) jet η residual distribution $\mathcal{R}^d(\eta)$ for both topo-cluster and TCC jets. The p_T pseudo-response demonstrates that TCC jets are constrained by the calorimeter energy scale as designed, while the ungroomed TCC collection has a higher p_T due to the presence of pile-up. The η residual distribution shows that the position of uncalibrated TCC jets is slightly worse than the position of calibrated topo-cluster jets, as explained in the text.

5.2 Jet substructure performance

The most intuitive jet substructure variable, and also one of the best at identifying hadronic decays of massive particles, is the jet mass. Jets built from topo-clusters currently make use of both tracking and calorimeter information through the combined mass [13]. The combined mass m_{comb} of topo-cluster jets is built from the linear combination of the calorimeter jet mass m_{calo} and the track-assisted jet mass m_{TA}, with weights defined by the respective mass resolutions σ, as follows:

$$m_{TA} = m_{track} \frac{p_T^{calo}}{p_T^{track}} \, ,$$

$$m_{comb} = m_{calo} \frac{\sigma_{m_{calo}}^{-2}}{\sigma_{m_{calo}}^{-2} + \sigma_{m_{TA}}^{-2}} + m_{TA} \frac{\sigma_{m_{TA}}^{-2}}{\sigma_{m_{calo}}^{-2} + \sigma_{m_{TA}}^{-2}} \, .$$

This linear combination does a good job of maximally exploiting the calorimeter at low p_T and the tracker at high p_T. However, the m_{TA} definition does not account for local fluctuations, while the TCC procedure does. As such, the mass resolution of TCC jets should be quantitatively superior to the mass resolution of topo-cluster jets using m_{comb} at high p_T. A comparison of the mass response for jets built from different inputs is shown for $700\,\text{GeV} < p_T < 800\,\text{GeV}$ in Figure 8(a) and for $2.1\,\text{TeV} < p_T < 2.5\,\text{TeV}$ in Figure 8(b). The performance of topo-cluster, all-TCC, and combined-TCC jets is seen to be qualitatively similar. At high p_T, the responses of the all-TCC and combined-TCC jets peak at values larger than one. Since these jets are at the uncalibrated scale, this effect is expected.

Another particularly powerful variable for the identification of vector bosons is $D_2^{\beta=1}$ [30, 31], hereafter referred to as D_2. This variable is a complex ratio of energy correlation functions that does a good job of differentiating two-body hadronic decays from one-body QCD jets. While the variable is powerful,
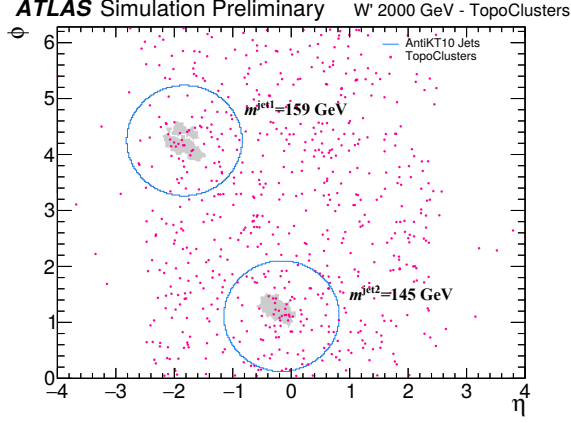
(a) Jets built from topo-clusters
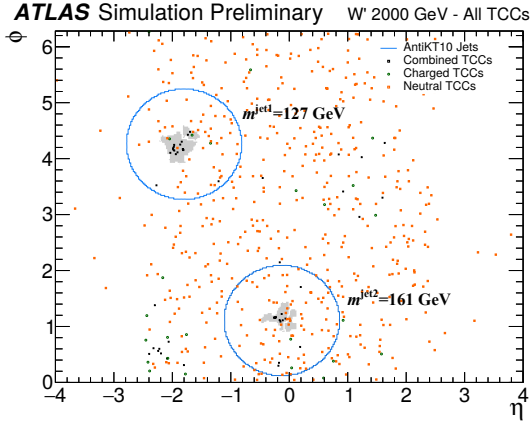
(b) Jets built from all TCCs
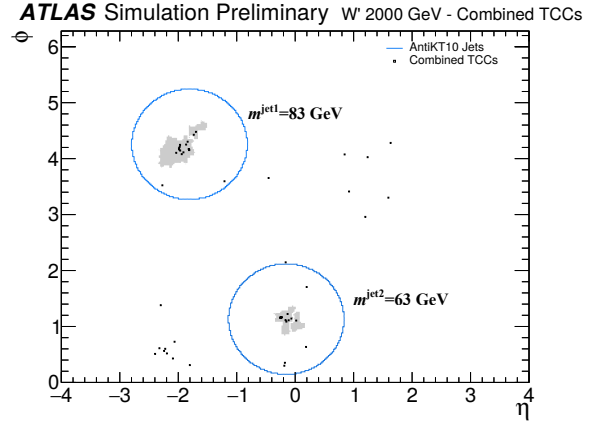
(c) Jets built from only combined TCCs

Figure 7: A single simulated $W' \to WZ \to qqqq$ event with a W' mass of 2TeV, reconstructed using (a) topo-clusters, (b) all TCCs, and (c) only combined TCCs. Jets are built using the anti-k_t algorithm with a distance parameter $R = 1.0$ and are considered before trimming, at the uncalibrated scale, to show the raw detector view of the event. Only the two leading jets are shown, with their area demarcated by a blue circle, and with grey regions denoting the corresponding k_t sub-jets of size $R_{sub} = 0.2$ and with $f_{cut} > 5\%$. Topo-clusters are shown as pink circles, combined TCC objects are shown as black squares, charged TCC objects (unmatched tracks) are shown as green circles, and neutral TCC objects (unmatched topo-clusters) are shown as orange squares.
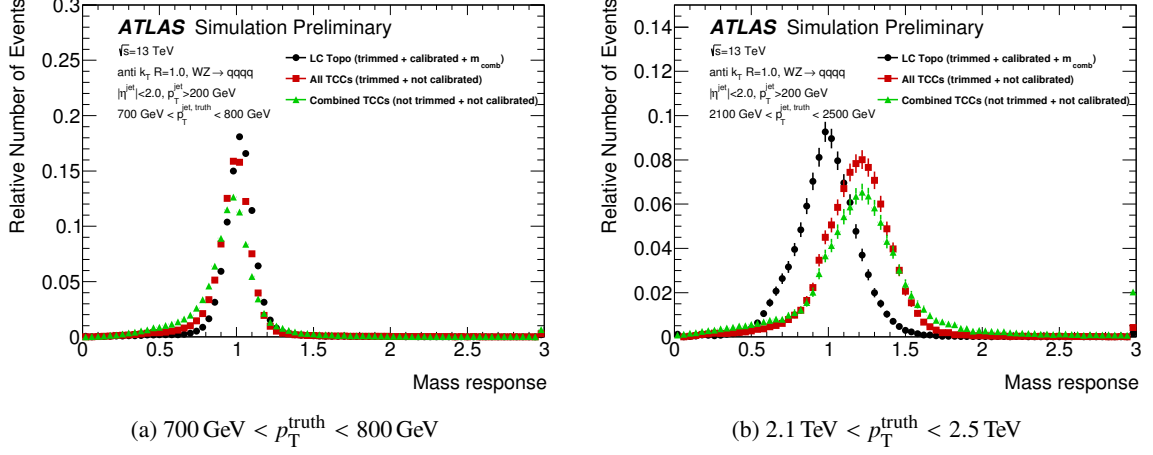
14

(a) $700\,\text{GeV} < p_\text{T}^\text{truth} < 800\,\text{GeV}$

(b) $2.1\,\text{TeV} < p_\text{T}^\text{truth} < 2.5\,\text{TeV}$

Figure 8: A comparison of the mass response, $\mathcal{R}_m^r$, for calibrated topo-cluster jets using the combined mass (black circles), uncalibrated jets built using all-TCC objects (red squares), and uncalibrated jets built using only combined-TCC objects (green triangles), for two different jet p_T^truth bins. The last bin in each plot represents the overflow for all entries beyond the range of the axis, which is only observed to be relevant for combined-TCC jets. Due to uncalibrated scale shifts the widths cannot be directly visually compared, but it is clear that the three jet definitions perform at a similar level for both jet p_T^truth bins. A quantitative comparison of their performance accounting for the scale shift is provided in Figure 10(a).

particularly at truth level, it is very sensitive to both the energy scale and angular resolution. This is very apparent when looking at jets built from topo-clusters and where the value of D_2 is thus calculated from those topo-clusters. On the other hand, the TCC definition significantly improves the angular resolution of the inputs to jet reconstruction, and thus to the D_2 calculation.

A comparison of the D_2 residual distribution, $\mathcal{R}_{D_2}^d$ for jets built from topo-clusters, jets built from all-TCC objects, and jets built from only combined-TCC objects is presented in Figure 9. As the value of D_2 is calculated from the inputs, jet calibration does not change the results, and thus the topo-cluster and TCC jet definitions can be directly and quantitatively compared.

Jets built from all-TCCs are seen to have a much more stable D_2 residual distribution when changing the p_T by an order of magnitude, while topo-cluster jets change dramatically. In particular, the topo-cluster jet distribution is no longer remotely Gaussian at high p_T. Additionally, the all-TCC jet distribution remains centred around zero, while the topo-cluster jet distribution is increasingly biased towards values of D_2 which are too large. It is also interesting to see that the combined-TCC performance has a reasonable core of the distribution, but long tails degrade the performance, suggesting that relevant energy contributions are missed when only considering track-cluster matches.

In order to understand the evolution of the mass response and D_2 residuals for topo-cluster and TCC jets, and to compare mass responses centred at different values, a resolution metric must be defined. Two Inter-Quantile Range (IQR) definitions of the distribution width are considered, where Q_x is the x% quantile boundary, meaning that Q_{50} is the median. The IQR definition is slightly different depending on if we are using the response $\mathcal{R}^r$ or the residual distribution $\mathcal{R}^d$:

15

(a) $700\,\text{GeV} < p_\text{T}^\text{truth} < 800\,\text{GeV}$

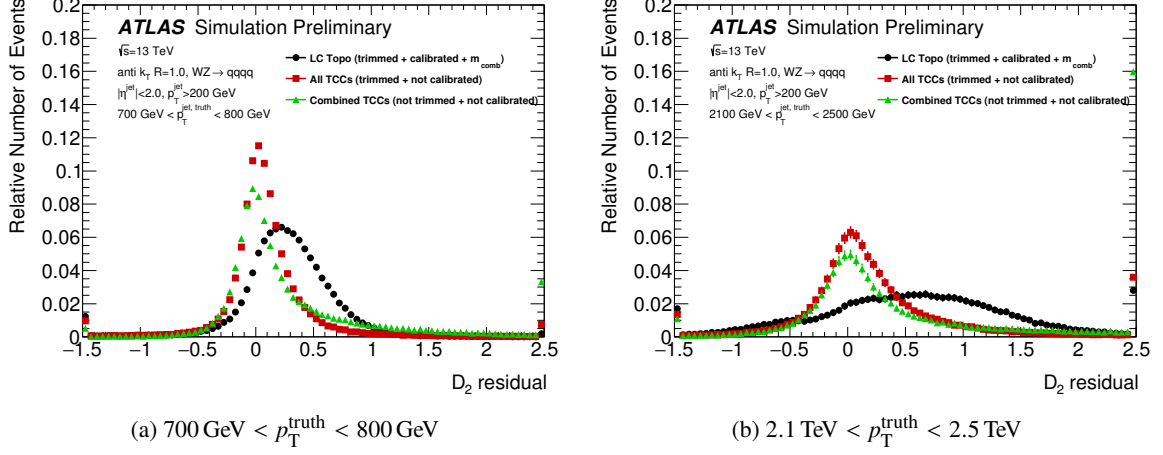(b) $2.1\,\text{TeV} < p_\text{T}^\text{truth} < 2.5\,\text{TeV}$

Figure 9: A comparison of the D_2 residual distribution, $\mathcal{R}_{D_2}^d$, for topo-cluster jets (black circles), jets built using all-TCC objects (red squares), and jets built using only combined-TCC objects (green triangles), for two different p_T^truth bins. The first (last) bin in each plot represents the underflow (overflow) for all entries before (beyond) the range of the axis. The underflow bin is seen to be minimal, while the overflow bin is seen to be very significant at high p_T for the combined-TCC jet definition, representing roughly 16% of the total number of events. While the D_2 resolution for all-TCC jets does grow when moving to high p_T, the difference is significantly less than for topo-cluster jets.

$$\text{IQR}^r = \frac{1}{2} \frac{Q_{84}(\mathcal{R}^r) - Q_{16}(\mathcal{R}^r)}{Q_{50}(\mathcal{R}^r)}, \tag{10}$$

$$\text{IQR}^d = \frac{1}{2} \left[Q_{75}(\mathcal{R}^d) - Q_{25}(\mathcal{R}^d) \right]. \tag{11}$$

These definitions correspond to half of a 68% window for the mass resolution, customary for jet mass resolution measurements in ATLAS, and half of a 50% window for the substructure variable resolution. These definitions properly include events previously grouped into underflow and overflow bins by expanding the axis ranges. The respective IQR resolutions are shown for jet mass and D_2 in Figure 10.

The mass resolution of all-TCC jets is seen to be superior to the topo-cluster jet m_comb for $p_\text{T}^\text{truth} > 2\,\text{TeV}$, and to perform slightly worse for most of the p_T^truth range below that. The D_2 residual distribution is very different, and all-TCC jets are seen to significantly out-perform the topo-cluster jets over the entire p_T^truth range. For the high end of the considered range, the resolution of all-TCC jets is a factor of two improved with respect to topo-cluster jets. Combined-TCC jets have a larger resolution than both other jet types due to a high tail as seen in Figure 9, which grows at high p_T. Reducing this tail is important to improving the performance of combined-TCC jets. The mass and D_2 resolutions of jets built from all three types of inputs degrade with increasing p_T^truth, due to the more collimated topology and the worsening respective detector resolutions. For the D_2 resolution of all-TCC jets, this degradation is much reduced.

It is also important to understand the pile-up robustness of the TCC procedure compared to jets built from topo-clusters, particularly for the combined-TCC jets where grooming was not applied. The median mass and the mass resolution for high p_T jets as a function of the Number of reconstructed Primary Vertices (NPV) in an event is shown in Figure 11.
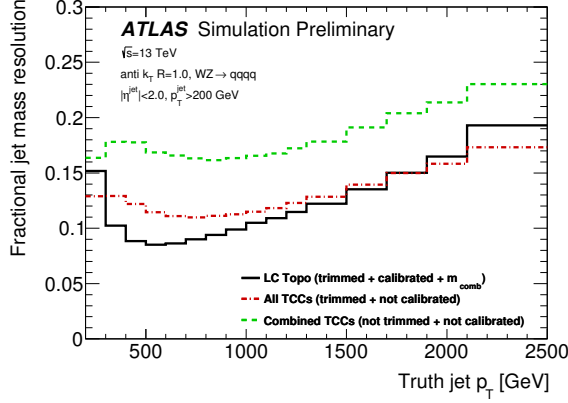
(a) Jet mass resolution, IQR^r
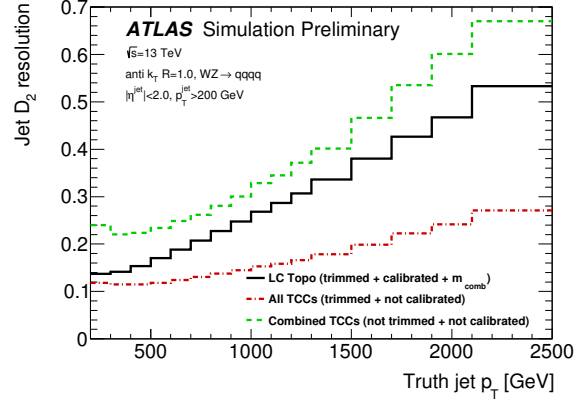
(b) Jet D_2 resolution, IQR^d

Figure 10: A comparison of the fractional jet (a) mass and (b) D_2 resolution for topo-cluster jets (solid black lines), jets built using all-TCC objects (dash-dotted red lines), and jets built using only combined-TCC objects (dashed green lines), as a function of p_T^{truth}. The TCC definitions are seen to excel at high p_T^{truth}, particularly for D_2.
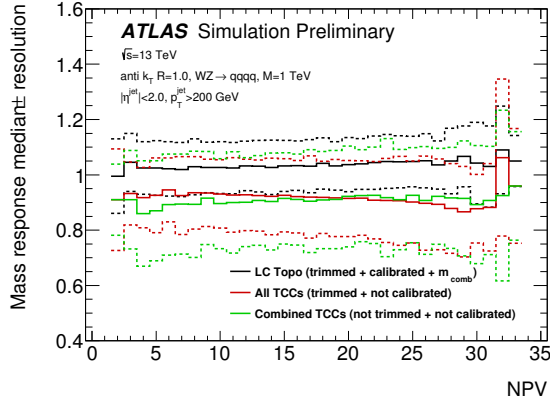
(a) W′ mass of 1 TeV
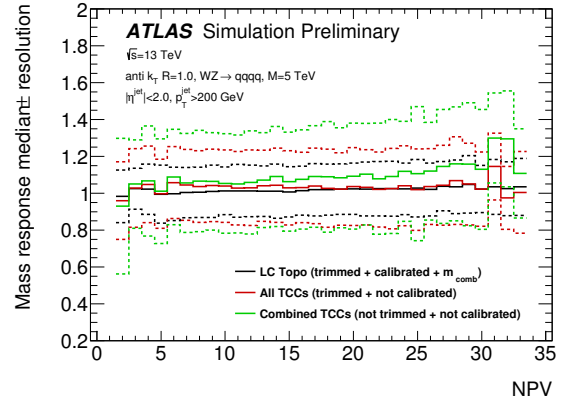
(b) W′ mass of 5 TeV

Figure 11: The median (solid line) and median $\pm 1\sigma$ IQR resolution (dashed lines) for (black) topo-cluster, (red) all-TCC, and (green) combined-TCC jets, as a function of the number of reconstructed primary vertices (NPV) for W′ → WZ → qqqq samples with two different W′ masses. A small positive (negative) slope is observed for the median of the topo-cluster (all-TCC) jet definition at low W′ mass, which disappears by higher mass values. The median of the ungroomed combined-TCC jet definition is seen to have only a small positive slope for both plots.

All three jet definitions are seen to have relatively stable central values and widths across the full NPV range, with a very small negative slope for all-TCC jets and a very small positive slope for topo-cluster jets for lower p_T jets from a W′ → WZ → qqqq sample with a W′ mass of 1 TeV. These small slopes disappear by the time the W′ mass is 5 TeV. The combined-TCC jets have a small positive slope for both W′ mass points. From this, we conclude that all three jet definitions are reasonably stable with respect to pile-up, and even though the combined-TCC jets are not groomed they are only slightly more sensitive.

6 Conclusion

A new type of object unifying tracking and calorimeter information, the Track-CaloCluster (TCC), has been studied as an input to jet reconstruction. This object exploits the excellent tracker spatial resolution and calorimeter energy resolution to improve jet substructure techniques at high p_T. This is done by building 4-vectors using tracker spatial coordinates and calorimeter energy scales, and defining an energy sharing algorithm where the tracker p_T measurements only enter as ratios with other tracks.

Two types of TCC jets are studied, those built from all-TCC objects and those built from only combined-TCC objects, both of which are compared to standard topo-cluster jets. The all-TCC jet definition outperforms the topo-cluster jet combined mass definition for $p_T > 2$ TeV, with slightly worse performance for most of the lower p_T range. The all-TCC jet definition performs significantly better than topo-cluster jets for the D_2 resolution across the full p_T range considered, providing a factor of two improvement for the highest p_T jets. Combined-TCC jets remain interesting for their pile-up resilience without the use of grooming, but further studies are needed to make their performance more competitive.

These preliminary studies reinforce the potential power of unifying tracking and calorimeter information for the use of jet substructure at very high p_T. Such techniques promise to increase the sensitivity of the ATLAS physics program.

Appendix

A Energy sharing algorithm examples

In order to assist the reader in understanding the energy sharing algorithm presented in Equation 8, worked examples corresponding to Figure 3 are provided. To gain familiarity with the algorithm, consider the simple 4-vectors presented in Equations 5 and 6. In this case, there are a pair of tracks t_2 and t_3 from the selected hard scatter primary vertex matching a single topo-cluster c_2, meaning that the final result will be two TCC objects (one per track) sharing the energy from the topo-cluster. Using Equations 7 and 8, this simplifies as follows:

$$\mathbf{M}_{t_2} = \sum_{c \in \{c_2\}} \mathbf{p}^c f_{t_2}^c \mathcal{F}_{c,t}^{t_2} = \mathbf{p}^{c_2} f_{t_2}^{c_2} \mathcal{F}_{c_2,t}^{t_2} \ , \qquad\qquad \mathbf{M}_{t_3} = \sum_{c \in \{c_2\}} \mathbf{p}^c f_{t_3}^c \mathcal{F}_{c,t}^{t_3} = \mathbf{p}^{c_2} f_{t_3}^{c_2} \mathcal{F}_{c_2,t}^{t_3} \ . \qquad (12)$$

Next, the cluster p_{T} fractions f_τ^c can be determined:

$$f_{t_2}^{c_2} = \frac{p_{\mathrm{T}}^{c_2}}{p_{\mathrm{T}}\left[\sum_{k \in C_{t_2}} \mathbf{p}^k\right]} = \frac{p_{\mathrm{T}}^{c_2}}{p_{\mathrm{T}}\left[\mathbf{p}^{c_2}\right]} = 1 \ , \qquad\qquad f_{t_3}^{c_2} = \frac{p_{\mathrm{T}}^{c_2}}{p_{\mathrm{T}}\left[\sum_{k \in C_{t_3}} \mathbf{p}^k\right]} = \frac{p_{\mathrm{T}}^{c_2}}{p_{\mathrm{T}}\left[\mathbf{p}^{c_2}\right]} = 1 \ . \qquad (13)$$

In this case, the cluster p_{T} fractions are simple because there is only a single matching topo-cluster, thus the relevance of the given topo-cluster is maximal. The normalization term $\mathcal{F}_{c,t}^\tau$ should now split the energy from the topo-cluster c_2 between the two matching tracks t_2 and t_3, which in turn both match only the single topo-cluster c_2:

$$\mathcal{F}_{c_2,t}^\tau = \frac{p_{\mathrm{T}}^\tau}{p_{\mathrm{T}}\left[\sum_{t \in \{t_2,t_3\}} \mathbf{p}^t \frac{p_{\mathrm{T}}^{c_2}}{p_{\mathrm{T}}\left[\sum_{k \in C_t} \mathbf{p}^k\right]}\right]} \ ,$$

$$= \frac{p_{\mathrm{T}}^\tau}{p_{\mathrm{T}}\left[\mathbf{p}^{t_2} \frac{p_{\mathrm{T}}^{c_2}}{p_{\mathrm{T}}\left[\sum_{k \in \{c_2\}} \mathbf{p}^k\right]} + \mathbf{p}^{t_3} \frac{p_{\mathrm{T}}^{c_2}}{p_{\mathrm{T}}\left[\sum_{k \in \{c_2\}} \mathbf{p}^k\right]}\right]} \ ,$$

$$= \frac{p_{\mathrm{T}}^\tau}{p_{\mathrm{T}}\left[\mathbf{p}^{t_2} \frac{p_{\mathrm{T}}^{c_2}}{p_{\mathrm{T}}\left[\mathbf{p}^{c_2}\right]} + \mathbf{p}^{t_3} \frac{p_{\mathrm{T}}^{c_2}}{p_{\mathrm{T}}\left[\mathbf{p}^{c_2}\right]}\right]} \ ,$$

$$\mathcal{F}_{c_2,t}^\tau = \frac{p_{\mathrm{T}}^\tau}{p_{\mathrm{T}}\left[\mathbf{p}^{t_2} + \mathbf{p}^{t_3}\right]} \implies \mathcal{F}_{c_2,t}^{t_2} = \frac{p_{\mathrm{T}}^{t_2}}{p_{\mathrm{T}}\left[\mathbf{p}^{t_2} + \mathbf{p}^{t_3}\right]} \ , \quad \mathcal{F}_{c_2,t}^{t_3} = \frac{p_{\mathrm{T}}^{t_3}}{p_{\mathrm{T}}\left[\mathbf{p}^{t_2} + \mathbf{p}^{t_3}\right]} \ . \qquad (14)$$

Combining Equations 12, 13, and 14 provides the final result:

$$\mathbf{M}_{t_2} = \mathbf{p}^{c_2} \frac{p_T^{t_2}}{p_T\left[\mathbf{p}^{t_2} + \mathbf{p}^{t_3}\right]}, \qquad\qquad \mathbf{M}_{t_3} = \mathbf{p}^{c_2} \frac{p_T^{t_3}}{p_T\left[\mathbf{p}^{t_2} + \mathbf{p}^{t_3}\right]}, \qquad (15)$$

which matches the energy sharing 4-vectors needed to obtain Equations 5 and 6. The same can be done for any other arbitrary set of matched topo-clusters and tracks.

As a more complex example, this same methodology can be followed to derive the 4-vectors for TCC⑥ and TCC⑦ of Figure 3. In this case, our seed track is either $\tau = t_4$ for TCC⑥ or $\tau = t_5$ for TCC⑦. Then, $C_{t_4} = \{c_4, c_6\}$ and $C_{t_5} = \{c_3, c_4, c_5\}$. This means that the sets T_c become $T_{c_3} = \{t_5\}$, $T_{c_4} = \{t_4, t_5\}$, $T_{c_5} = \{t_5\}$, and $T_{c_6} = \{t_4\}$. Looping over each track t within each of these sets of tracks T_c then results in the following set of matching topo-clusters C_t: $C_{t_4} = \{c_4, c_6\}$ and $C_{t_5} = \{c_3, c_4, c_5\}$.

Using this information and following the same steps as above, but in a compressed form, results in the following series of equations ending in $\mathbf{M}_{t_4}$ for TCC⑥ and $\mathbf{M}_{t_5}$ for TCC⑦. Note that the formalism is such that the subscripts on f_τ^c and $\mathcal{F}_{c,t}^\tau$ parametrize the matches, which can be directly read off of Figure 3, while the superscripts can be left as generic until the final step where the final 4-vector sum is performed.

$$\mathbf{M}_{t_4} = \mathbf{p}^{c_4} f_{t_4}^{c_4} \mathcal{F}_{c_4,t}^{t_4} + \mathbf{p}^{c_6} f_{t_4}^{c_6} \mathcal{F}_{c_6,t}^{t_4} \qquad\qquad \mathbf{M}_{t_5} = \mathbf{p}^{c_3} f_{t_5}^{c_3} \mathcal{F}_{c_3,t}^{t_5} + \mathbf{p}^{c_4} f_{t_5}^{c_4} \mathcal{F}_{c_4,t}^{t_5} + \mathbf{p}^{c_5} f_{t_5}^{c_5} \mathcal{F}_{c_5,t}^{t_5}$$

$$f_{t_4}^c = \frac{p_T^c}{p_T\left[\mathbf{p}^{c_4} + \mathbf{p}^{c_6}\right]} \qquad\qquad\qquad f_{t_5}^c = \frac{p_T^c}{p_T\left[\mathbf{p}^{c_3} + \mathbf{p}^{c_4} + \mathbf{p}^{c_5}\right]}$$

$$\mathcal{F}_{c_3,t}^\tau = \frac{p_T^\tau}{p_T\left[\mathbf{p}^{t_5} f_{t_5}^{c_3}\right]} \quad \mathcal{F}_{c_4,t}^\tau = \frac{p_T^\tau}{p_T\left[\mathbf{p}^{t_4} f_{t_4}^{c_4} + \mathbf{p}^{t_5} f_{t_5}^{c_4}\right]} \quad \mathcal{F}_{c_5,t}^\tau = \frac{p_T^\tau}{p_T\left[\mathbf{p}^{t_5} f_{t_5}^{c_5}\right]} \quad \mathcal{F}_{c_6,t}^\tau = \frac{p_T^\tau}{p_T\left[\mathbf{p}^{t_4} f_{t_4}^{c_6}\right]}$$

Putting this all together and replacing the superscripts with the appropriate seed track τ and matching clusters c gives the final results for $\mathbf{M}_{t_4}$ and $\mathbf{M}_{t_5}$, which correspond to the 4-vector energy sharing results for TCC⑥ and TCC⑦ respectively:

$$\mathbf{M}_{t_4} = \mathbf{p}^{c_4} \; \frac{p_{\mathrm{T}}^{c_4}}{p_{\mathrm{T}}\left[\mathbf{p}^{c_4} + \mathbf{p}^{c_6}\right]} \; \frac{p_{\mathrm{T}}^{t_4}}{p_{\mathrm{T}}\left[\mathbf{p}^{t_4} \dfrac{p_{\mathrm{T}}^{c_4}}{p_{\mathrm{T}}\left[\mathbf{p}^{c_4} + \mathbf{p}^{c_6}\right]} + \mathbf{p}^{t_5} \dfrac{p_{\mathrm{T}}^{c_4}}{p_{\mathrm{T}}\left[\mathbf{p}^{c_3} + \mathbf{p}^{c_4} + \mathbf{p}^{c_5}\right]}\right]}$$

$$+ \; \mathbf{p}^{c_6} \; \frac{p_{\mathrm{T}}^{c_6}}{p_{\mathrm{T}}\left[\mathbf{p}^{c_4} + \mathbf{p}^{c_6}\right]} \; \frac{p_{\mathrm{T}}^{t_4}}{p_{\mathrm{T}}\left[\mathbf{p}^{t_4} \dfrac{p_{\mathrm{T}}^{c_6}}{p_{\mathrm{T}}\left[\mathbf{p}^{c_4} + \mathbf{p}^{c_6}\right]}\right]}$$

$$\mathbf{M}_{t_5} = \mathbf{p}^{c_3} \; \frac{p_{\mathrm{T}}^{c_3}}{p_{\mathrm{T}}\left[\mathbf{p}^{c_3} + \mathbf{p}^{c_4} + \mathbf{p}^{c_5}\right]} \; \frac{p_{\mathrm{T}}^{t_5}}{p_{\mathrm{T}}\left[\mathbf{p}^{t_5} \dfrac{p_{\mathrm{T}}^{c_3}}{p_{\mathrm{T}}\left[\mathbf{p}^{c_3} + \mathbf{p}^{c_4} + \mathbf{p}^{c_5}\right]}\right]}$$

$$+ \; \mathbf{p}^{c_4} \; \frac{p_{\mathrm{T}}^{c_4}}{p_{\mathrm{T}}\left[\mathbf{p}^{c_3} + \mathbf{p}^{c_4} + \mathbf{p}^{c_5}\right]} \; \frac{p_{\mathrm{T}}^{t_5}}{p_{\mathrm{T}}\left[\mathbf{p}^{t_4} \dfrac{p_{\mathrm{T}}^{c_4}}{p_{\mathrm{T}}\left[\mathbf{p}^{c_4} + \mathbf{p}^{c_6}\right]} + \mathbf{p}^{t_5} \dfrac{p_{\mathrm{T}}^{c_4}}{p_{\mathrm{T}}\left[\mathbf{p}^{c_3} + \mathbf{p}^{c_4} + \mathbf{p}^{c_5}\right]}\right]}$$

$$+ \; \mathbf{p}^{c_5} \; \frac{p_{\mathrm{T}}^{c_5}}{p_{\mathrm{T}}\left[\mathbf{p}^{c_3} + \mathbf{p}^{c_4} + \mathbf{p}^{c_5}\right]} \; \frac{p_{\mathrm{T}}^{t_5}}{p_{\mathrm{T}}\left[\mathbf{p}^{t_5} \dfrac{p_{\mathrm{T}}^{c_5}}{p_{\mathrm{T}}\left[\mathbf{p}^{c_3} + \mathbf{p}^{c_4} + \mathbf{p}^{c_5}\right]}\right]}$$

References

[1] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[2] ATLAS Collaboration, *Identification of boosted, hadronically-decaying W and Z bosons in $\sqrt{s}$ = 13 TeV Monte Carlo Simulations for ATLAS*, ATL-PHYS-PUB-2015-033 (2015), URL: http://cds.cern.ch/record/2041461.

[3] ATLAS Collaboration, *Expected Performance of Boosted Higgs ($\rightarrow b\bar{b}$) Boson Identification with the ATLAS Detector at $\sqrt{s}$ = 13 TeV*, ATL-PHYS-PUB-2015-035 (2015), URL: http://cds.cern.ch/record/2042155.

[4] ATLAS Collaboration, *Boosted hadronic top identification at ATLAS for early 13 TeV data*, ATL-PHYS-PUB-2015-053 (2015), URL: http://cds.cern.ch/record/2116351.

[5] ATLAS Collaboration, *Quark versus Gluon Jet Tagging Using Charged Particle Multiplicity with the ATLAS Detector*, ATL-PHYS-PUB-2017-009 (2017), URL: https://cds.cern.ch/record/2263679.

[6] ATLAS Collaboration, *Search for resonances with boson-tagged jets in 15.5 fb^{-1} of pp collisions at $\sqrt{s}$ = 13 TeV collected with the ATLAS detector*, ATLAS-CONF-2016-055 (2016), URL: http://cds.cern.ch/record/2206137.

[7] ATLAS Collaboration, *Searches for heavy diboson resonances in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, JHEP **09** (2016) 173, arXiv: 1606.04833 [hep-ex].

[8] ATLAS Collaboration, *Measurements of tt̄ differential cross-sections in the all-hadronic channel with the ATLAS detector using highly boosted top quarks in pp collisions at $\sqrt{s}$ = 13 TeV*, ATLAS-CONF-2016-100 (2016), URL: http://cds.cern.ch/record/2217231.

[9] ATLAS Collaboration, 'Jet Trigger Efficiency Plots, Adding Trimming and Mass Cuts to Large-R Jets', Public plots, 2017, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/JetTriggerPublicResults#Jet_Trigger_Efficiency_Plots_Add.

[10] ATLAS Collaboration, *Identification of Hadronically-Decaying W Bosons and Top Quarks Using High-Level Features as Input to Boosted Decision Trees and Deep Neural Networks in ATLAS at $\sqrt{s}$ = 13 TeV*, ATL-PHYS-PUB-2017-004 (2017), URL: http://cds.cern.ch/record/2259646.

[11] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, (2016), arXiv: 1603.02934 [hep-ex].

[12] ATLAS Collaboration, *Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC run 2*, Eur. Phys. J. **C (submitted)** (2017), arXiv: 1704.07983 [hep-ex].

[13] ATLAS Collaboration, *Jet mass reconstruction with the ATLAS Detector in early Run 2 data*, ATLAS-CONF-2016-035 (2016), URL: http://cds.cern.ch/record/2200211.

[14] S. Schaetzel and M. Spannowsky, *Tagging highly boosted top quarks*, Phys. Rev. **D89** (2014) 014007, arXiv: 1308.0540 [hep-ph].

[15] A. J. Larkoski, F. Maltoni and M. Selvaggi, *Tracking down hyper-boosted top quarks*, JHEP **06** (2015) 032, arXiv: 1503.03347 [hep-ph].

[16] S. Bressler, T. Flacke, Y. Kats, S. J. Lee and G. Perez, *Hadronic Calorimeter Shower Size: Challenges and Opportunities for Jet Substructure in the Superboosted Regime*, (2015), arXiv: 1506.02656 [hep-ph].

[17] M. Spannowsky and M. Stoll, *Tracking New Physics at the LHC and beyond*, Phys. Rev. **D92** (2015) 054033, arXiv: 1505.01921 [hep-ph].

[18] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, (2017), arXiv: 1703.10485 [hep-ex].

[19] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[20] T. Sjostrand, S. Mrenna and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv: 0710.3820 [hep-ph].

[21] ATLAS Collaboration, *ATLAS Run 1 Pythia8 tunes*, ATL-PHYS-PUB-2014-021 (2014), URL: http://cds.cern.ch/record/1966419.

[22] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. **B867** (2013) 244, arXiv: 1207.1303 [hep-ph].

[23] S. Agostinelli et al., *GEANT4: A Simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.

[24] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. **C70** (2010) 823, arXiv: 1005.4568 [physics.ins-det].

[25] ATLAS Collaboration, *Summary of ATLAS Pythia 8 tunes*, ATL-PHYS-PUB-2012-003 (2012), URL: http://cds.cern.ch/record/1474107.

[26] A. D. Martin, W. J. Stirling, R. S. Thorne and G. Watt, *Parton distributions for the LHC*, Eur. Phys. J. **C63** (2009) 189, arXiv: 0901.0002 [hep-ph].

[27] M. Cacciari, G. P. Salam and G. Soyez, *The Anti-k(t) jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: 0802.1189 [hep-ph].

[28] D. Krohn, J. Thaler and L.-T. Wang, *Jet Trimming*, JHEP **02** (2010) 084, arXiv: 0912.1342 [hep-ph].

[29] ATLAS Collaboration, *Monte Carlo Calibration and Combination of In-situ Measurements of Jet Energy Scale, Jet Energy Resolution and Jet Mass in ATLAS*, ATLAS-CONF-2015-037 (2015), URL: https://cds.cern.ch/record/2044941.

[30] A. J. Larkoski, G. P. Salam and J. Thaler, *Energy Correlation Functions for Jet Substructure*, JHEP **06** (2013) 108, arXiv: 1305.0007 [hep-ph].

[31] A. J. Larkoski, I. Moult and D. Neill, *Power Counting to Better Jet Observables*, JHEP **12** (2014) 009, arXiv: 1409.6298 [hep-ph].