IL NUOVO CIMENTO **39 C** (2016) 348
DOI 10.1393/ncc/i2016-16348-0

Colloquia: La Thuile 2016

Searches for new physics with boosted objects and substructure with the ATLAS detector

D. López Mateos on behalf of the ATLAS Collaboration
Harvard University - Cambridge, MA 02138, USA

received 26 July 2016

Summary. — The use of substructure in the ATLAS experiment has matured during the Run 1 analysis period into the most powerful new tool for understanding high-p_{T} physics at the LHC. In this document we present the studies that have been instrumental in reaching that maturity for boosted hadronic W/Z, Higgs and top tagging. We also summarize the results from Run 1 and Run 2 searches for new physics using substructure, thus demonstrating the power of these new techniques.

1. – Introduction

Hadronic decays of top quarks, Higgs and vector bosons were long thought of as being too hard to reconstruct precisely enough to play a significant role in the search for new physics at the LHC. However, following pioneering work on the use of substructure in hadronic reconstruction of boosted objects [1, 2], it became clear that the search for very massive objects produced at the LHC could greatly benefit from the use of these new techniques, in particular in light of the increased center-of-mass energy available in Run 2, giving renewed importance to hadronic channels. Since then, theoretical and experimental efforts have proceeded in parallel: the former searching for better and more robust ways of identifying high-p_{T} objects, the latter making sure that these new techniques are tested in an experimental environment and that systematic effects are properly understood. In what follows, first, the experimental studies performed with the ATLAS detector to commission boosted-object tagging techniques are discussed and the chosen algorithms and settings are motivated. In the second half of this document, some representative Run 1 searches for new physics using these techniques are presented, together with the newest results using Run 2 data.

2. – Jet reconstruction with ATLAS: Clusters and grooming

Jet reconstruction in ATLAS uses as input four-vectors clusters of cells that are topologically linked and with energy above the noise threshold. Cluster building uses

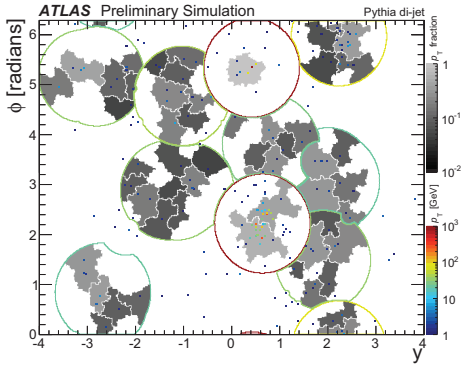

Fig. 1. – Dijet event in Monte Carlo simulation illustrating the effect of k_t trimming with $R = 0.3$ performed on anti-k_t jets with $R = 1.0$.

seed cells that are four times above the noise threshold and lets the cluster grow three-dimensionally as long as adjacent cell energy measurements are two times above the noise threshold. Once growth stops, an additional layer of adjacent cells is added to the cluster, independent of their energy. The existence of topological information about the cluster shape allows for a local calibration to be applied that brings the cluster energy closer to the true energy deposited in the calorimeter [3]. This is important for substructure applications, since it brings the reconstructed value of substructure variables closer to their truth values.

Three jet algorithms with different radius parameters (R) are used in optimization studies and for tagging different objects: the k_t [4], Cambridge/Aachen [5] and anti-k_t [6] algorithms. For the study of substructure, jet reconstruction is enhanced by *grooming* techniques that remove soft-energy contributions that are not relevant to understand the hard substructure of the jet. Three grooming techniques are used: pruning, split-filtering and trimming. Figure 1 shows the effect of trimming anti-k_t jets built with $R = 1.0$. The trimming algorithm uses k_t subjets of $R = 0.3$ and a cut-off of 5% of the jet p_{T} to remove soft contributions. The figure shows the calorimeter η and ϕ and the contours of the jets (lines) as well as the subjets (shaded areas). The points represent the position of clusters and only subjets that survive the trimming cut are shown.

3. – Tagging boosted hadronic objects: W/Z, Higgs and top

One of the main uses of substructure variables in searches is to discriminate between objects of interest (W/Z and Higgs bosons and top quarks) and backgrounds (most often QCD, but when looking for Higgs bosons, also $t\bar{t}$). For that reason, one of the primary focus of optimization studies is to understand the background rejection of a given algorithm at fixed values of the selection efficiencies. For W and Z tagging the same strategy is followed, except that the mass window chosen is adjusted according to which of the two objects one is trying to identify.

A large scan of substructure variables and jet/grooming algorithms is performed in order to build a robust, yet highly performant tagging algorithm for W/Z bosons [7]. The study, performed with MC simulations, uses several criteria of robustness, such as the shape of the mass distribution, or the dependence on pile-up, to pre-select a set of configurations. Amongst those preferred configurations, a tagger is designed that applies

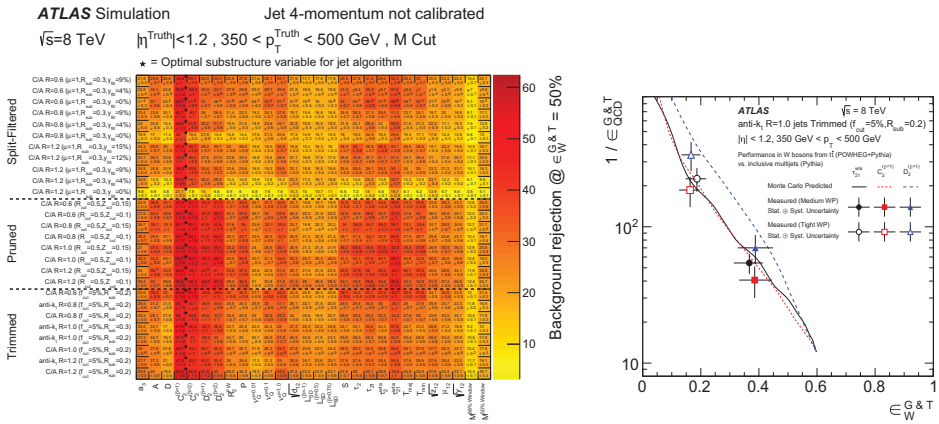

Fig. 2. – Rejection of QCD backgrounds for W-taggers using different reconstruction and substructure techniques for 50% tagging efficiency (left) and rejection as a function of tagger efficiency in data and MC simulation for three selected taggers (right) [7].

a cut on the mass to achieve 68% efficiency and a cut on a substructure variable sensitive to the two-prong structure of the W/Z boson to reach 50% or 25% selection efficiency. The rejection power for QCD backgrounds (defined as the inverse of the efficiency to select QCD jets) is then estimated for all the jet reconstruction/substructure variable configurations. These results are shown in fig. 2 (left). For a subset of some of the taggers providing the highest rejection, the efficiency and rejection of the taggers is measured using $t\bar{t}$ data and dijet data and systematic uncertainties on the measurements estimated. Those measurements are shown together with the MC predictions for one p_{T} bin in fig. 2 (right), demonstrating a good agreement between data and MC simulation and very powerful rejection of QCD backgrounds.

The tagging of $H \to b\bar{b}$ also exploits the two-prong structure of the decay to reject QCD backgrounds [8,9]. However, in order to also exploit the information that b-tagging algorithms can provide, the two-prong structure is first established by reconstructing two jets with the anti-k_t algorithm with $R = 0.2$ using tracks only. The two jets are geometrically matched to an anti-k_t $R = 1.0$ calorimeter jet (the $H \to b\bar{b}$ candidate). As shown in fig. 3 (left), the simple requirement of having two track jets is able to reject both QCD and $t\bar{t}$ backgrounds. Given the high rejection power of b-tagging algorithms in the ATLAS detector, it is reasonable to use the tracks in each of the small track jets to build a b-tagging discriminant, and add the requirement that the two track jets are b-tagged according to that discriminant. Figure 3 (right) shows that this is extremely powerful at rejecting light jets when combined with a cut on the calorimeter jet mass, and even retains power when trying to reject $g \to b\bar{b}$ due to the larger asymmetry between the decay products observed for $g \to b\bar{b}$. Substructure information can also be used to complement the b-tagging, mass and track-jet counting information. However, it has been observed that this is mostly useful in the low-efficiency regime [8]. Finally, despite the lack of a calibration sample for $H \to b\bar{b}$ decays, studies using a sample enriched in $g \to b\bar{b}$ have demonstrated that there are no additional systematic uncertainties that need consideration in these topologies [9].

Due to the more complex substructure present inside hadronic tops, simple mass and substructure-based taggers can be complemented by taggers that try to exploit that more complex substructure. A full description of the HEPTopTagger and Shower

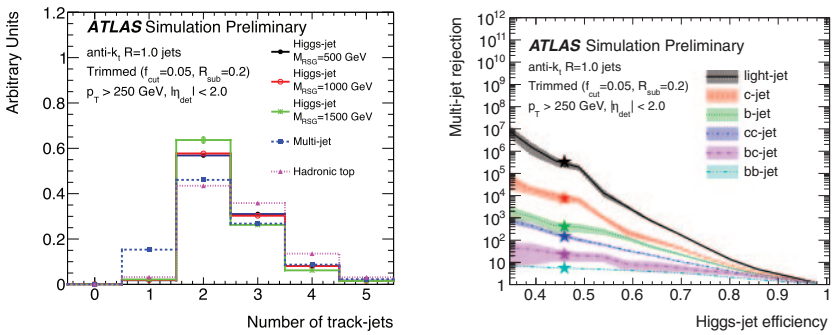

Fig. 3. – Number of track jets geometrically matched to jets originating from Higgs bosons, QCD and hadronic top decays (left) and QCD rejection as a function of Higgs-jet efficiency for a mass and b-tagging based Higgs tagger for different types of QCD jets (right) [8].

Deconstruction taggers used in ATLAS can be found in ref. [10]. The QCD rejection obtained with simple cut-based taggers as well as some of these more complex taggers as a function of the top-quark tagging efficiency can be seen for a representative p_{T} bin in fig. 4. As with W-boson tagging, the efficiency and rejection curves have also been probed in detail using data, demonstrating a good agreement between data and MC simulations.

4. – Searches with boosted hadronic objects in Run 1 and Run 2

Many searches and measurements have been performed with the ATLAS detector using techniques similar to those described in the previous section. This section thus briefly presents some results from Run 1 searches that capture well the improvements obtained through the exploitation of boosted topologies and exemplify the corresponding methodologies, and then goes on to focus on the latest Run 2 searches.

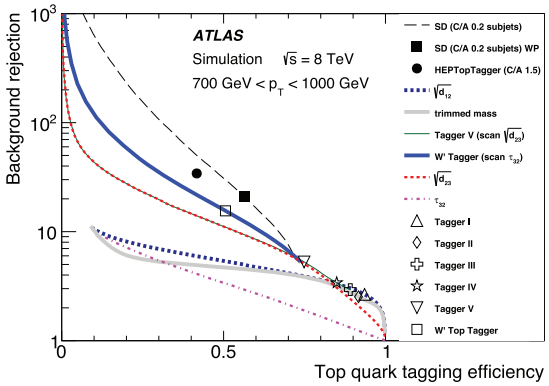

Fig. 4. – QCD rejection as a function of top tagging efficiency for different cut-based taggers (points), as well as substructure variable scans (lines). HEPTopTagger and Shower Deconstruction operating points are also shown as points and the efficiency-rejection curve arising from scanning the Shower Deconstruction output discriminant is also shown. More details on these taggers can be found in ref. [10].

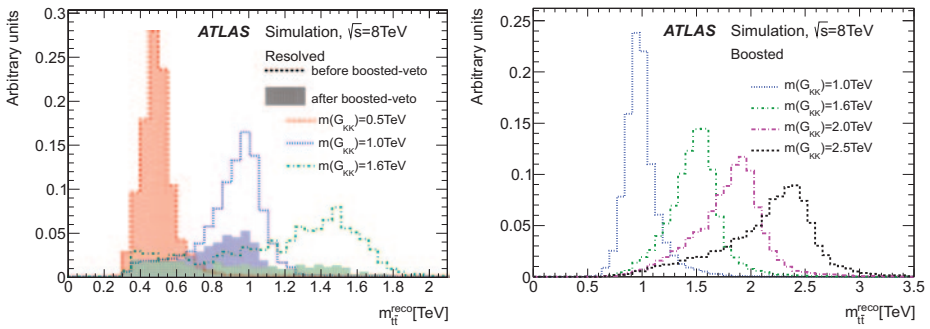

Fig. 5. – Reconstructed invariant mass of a top-quark pair originating from a Kaluza-Klein graviton of different masses reconstructed with standard reconstruction techniques (left) and with boosted reconstruction techniques (right) [11].

Figure 5 shows the invariant mass reconstruction of $t\bar{t}$ resonances of different masses when using standard top-quark reconstruction algorithms (left) and boosted object reconstruction, described in the previous section (right). In this analysis, described in detail in ref. [11], a combination of resolved (for lower masses) and boosted (for higher masses) selections are defined to reconstruct semileptonic $t\bar{t}$ resonances. As the figure shows, already for masses of 1 TeV the boosted reconstruction (which is only applied to the top quark that decays hadronically) provides a much narrower resonant peak than the resolved reconstruction. This analysis is characterized by low QCD backgrounds. For this reason, the use of high-efficiency taggers, based on simple cuts on substructure variables, is preferred. However, for analyses with much more QCD background, such as the complementary analysis in which both top quarks decay hadronically, the use of more complex taggers, such as the HEPTopTagger, is necessary [12].

The improvements observed in the resolution in fig. 5 translate directly into improvements in the limits found for new resonances. This is illustrated for a search for $X \to hh \to 4b$ in fig. 6 [13]. In this analysis, QCD backgrounds are very large in the boosted regime, but they are effectively suppressed using the tagging techniques described

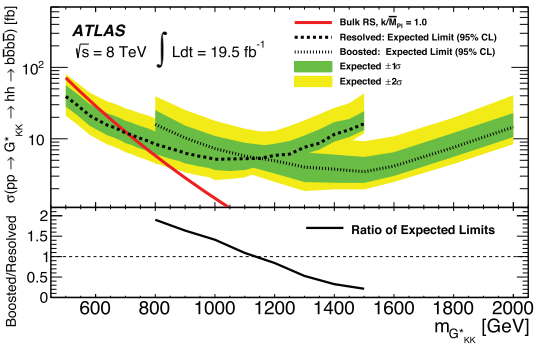

Fig. 6. – Limits obtained on a search of a Kaluza-Klein graviton decaying to a pair of Higgs bosons decaying each to a pair of b-quarks [13]. The cross section limit is shown as a function of the graviton mass for a resolved analysis and a boosted analysis using techniques from sect. **3**.

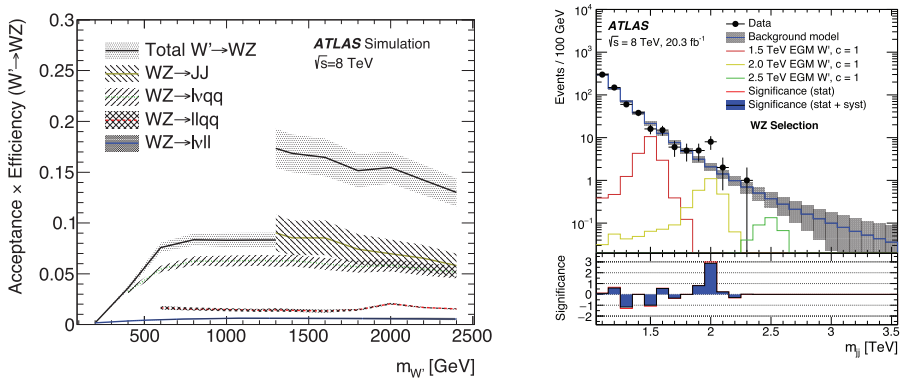

Fig. 7. – Acceptance times efficiency to select a $W' \to WZ$ for different decay channels as a function of resonant mass (left) [14]. The all-hadronic channel is used only with masses above 1.3 TeV. Invariant mass reconstructed in pp data and MC simulation at $\sqrt{s} = 8$ TeV in the all-hadronic search (right) [15].

in sect. **3**. The size and shape of the remaining QCD backgrounds are estimated using data-driven techniques that look at the mass sidebands of the signal region, as well as other control regions. These methodologies make the analysis competitive with traditional resolved analyses already for resonant masses of about 1.1 TeV, and bring an order of magnitude improvement in the cross-section limits at around 1.4 TeV.

Finally, analyses in Run 1 and Run 2 have made use of boosted W/Z-boson taggers to search for resonances decaying to WZ, ZZ or WW. The use of boosted-boson tagging is critical for obtaining access to the all-hadronic decay channels. As shown in fig. 7 (left) the all-hadronic channel has the highest acceptance of all channels for a W' signal model after taking into account tagging efficiency cuts [14]. Furthermore, in the Run 1 analysis a local excess of around 3σ is found in the all-hadronic search that could arise from a new resonance of mass around 2 TeV. This excess is shown in fig. 7 (right). This result does not only illustrate the usefulness of boosted object reconstruction for expanding our new physics reach, but it puts it front and center for the Run 2 start-up.

The increase in center-of-mass energy in Run 2 provides large gains in searches for massive objects that were close to or just below the threshold of being produced. Gains can also be obtained for lower-mass objects, but not with datasets much smaller than the 2012 dataset of $20\,\text{fb}^{-1}$. For this reason, and given the size of the 2015 dataset (of just over $3\,\text{fb}^{-1}$), the 2015 analyses searching for very massive particles were of special relevance, and so was boosted object reconstruction. Many searches for new objects decaying to dibosons were finalized in December 2015, and due to the focus in high-mass objects, many of them use exclusively boosted object reconstruction, rather than more traditional reconstruction techniques [16-20]. Here we focus the discussion on the two most sensitive searches for $W' \to WZ$ due to their interest in light of the slight excess seen in Run 1, namely the all-hadronic channel, and a channel that was not explored in Run 1: the 0-lepton ($WZ \to J\nu\nu$) channel.

The all-hadronic channel search proceeds as in Run 1, but with a more powerful W/Z tagger, derived from the discussion in sect. **3**. A fit is thus performed to the background after requiring a W-boson tag on one of the reconstructed jets and a Z-boson tag on the other (the searches for WW and ZZ in this channel are also reported in ref. [18]). The

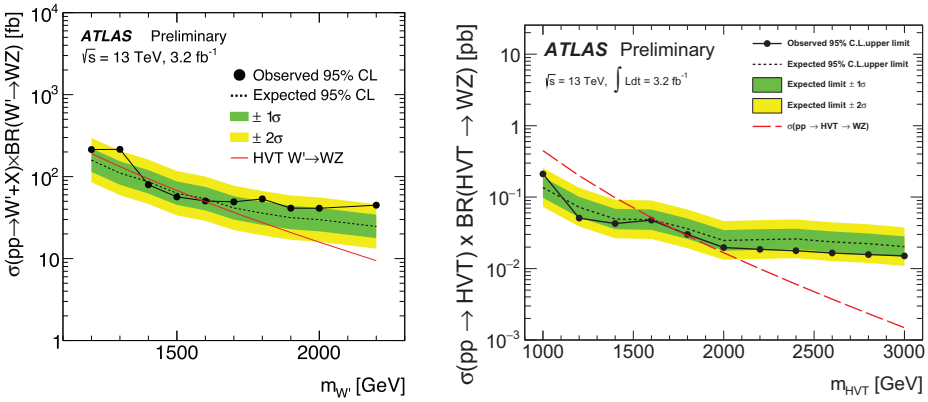

Fig. 8. – Cross-section limits on W' production decaying to WZ in the all-hadronic channel (left) [18] and in the 0-lepton ($WZ \to J\nu\nu$) channel (right) [16], obtained using the full 2015 dataset.

fit uses the same fit function as in the Run 1 analysis and is tested in a sample in which only one jet is required to be tagged. A new control region enhanced on $W+$jets events is also used for validating properties of the W-jet that are used in the tagger and to set associated systematic uncertainties. The cross-section limits obtained in this analysis are shown in fig. 8 (left). The statistics at masses around 2 TeV are clearly too small to rule out the presence of a new resonance. The small 1σ excess around 2 TeV is lower than naive expectations based on the Run 1 result. However, the signal hypothesis cannot be ruled out either.

In light of this result, it is interesting to turn to the 0-lepton channel, discussed in detail in ref. [16]. Due to the many backgrounds that are relevant for this search (W+jets, Z+jets and $t\bar{t}$), it relies heavily on control regions for the background estimation. Shapes are estimated from MC simulations and normalizations simultaneously fit in the control regions and the signal region in data. Shapes can also change in the final fit based on modeling systematic uncertainties. In the end a likelihood fit is performed for a signal and limits set based on that fit. The limit obtained with the 2015 data is shown in fig. 8 (right). The limit on the cross section is slightly better than that obtained on the all-hadronic search, but this time a deficit is observed in the data at around 2 TeV, suggesting that no excess is observed in the combined search of $W' \to WZ$ using the 2015 dataset. Since there are several options for the production channel for such a new particle, the results obtained with the 2015 dataset are not conclusive and the 2016 dataset should help clarify whether the 2012 results arose from an upwards statistical fluctuation or new physics.

5. – Conclusions and future prospects

The advent of the LHC has given rise to a new set of experimental and phenomenological techniques to get access to hadronic decays of massive objects with high p_{T}. During Run 1, studies have been performed to successfully commission these techniques for their use in analyses using data collected with the ATLAS detector. This has resulted in dedicated W/Z, top and Higgs taggers, and analyses to establish the systematic uncertainties associated to their usage. Those taggers have been effectively used in searches for new

physics in Run 1, improving the limits on the production cross section of new particles by over an order of magnitude. With the start of Run 2, and the emphasis placed on the search for very massive objects with the 2015 dataset, the use of boosted-object tagging techniques has become even more pervasive, allowing for improvements on searches for massive objects even with the small amount of data collected in 2015. In particular, the excess observed in the all-hadronic diboson search at around 2 TeV at the end of Run 1 seems to have disappeared, even though 2016 data will be necessary to unambiguously determine that it was caused by a statistical fluctuation.

Moving forward, boosted hadronic object reconstruction is here to stay, and will play an important role not only in searches for new physics. As data-driven approaches for understanding the hadronic object tagging become more sophisticated, these techniques will become more and more relevant also for precision measurements of SM processes. Some of these techniques have already been used in this context [21], and could play a decisive role in understanding the nature of new physics that may be found at the LHC. Furthermore, because of their importance, detectors for future accelerators will have to be designed to be able to probe the substructure of boosted objects, and calorimeter design choices will be influenced by the evolution and the need for these techniques.

REFERENCES

[1] Butterworth J. M., Davison A. R., Rubin M. and Salam G. P., *Phys. Rev. Lett.*, **100** (2008) 242001.
[2] Thaler J. and Wang L.-T., *JHEP*, **07** (2008) 92.
[3] ATLAS Collaboration, arXiv:1603.02934, submitted to *Eur. Phys. J. C*.
[4] Ellis S. D. and Soper D. E., *Phys. Rev. D*, **48** (1993) 3160.
[5] Dokshitzer Y. L., Leder G. D., Moretti S. and Webber B. R., *JHEP*, **08** (1997) 001.
[6] Cacciari M., Salam G. P. and Soyez G., *JHEP*, **04** (2008) 063.
[7] ATLAS Collaboration, *Eur. Phys. J. C*, **76** (2016) 1.
[8] ATLAS Collaboration, ATL-PHYS-PUB-2015-035.
[9] ATLAS Collaboration, ATLAS-CONF-2016-002.
[10] ATLAS Collaboration, *JHEP*, **06** (2016) 093, arXiv:1603.03127.
[11] ATLAS Collaboration, *JHEP*, **08** (2015) 148.
[12] ATLAS Collaboration, *JHEP*, **01** (2013) 116.
[13] ATLAS Collaboration, *Eur. Phys. J. C*, **75** (2015) 412.
[14] ATLAS Collaboration, *Phys. Lett. B*, **755** (2016) 285.
[15] ATLAS Collaboration, *JHEP*, **12** (2015) 055.
[16] ATLAS Collaboration, ATLAS-CONF-2015-068.
[17] ATLAS Collaboration, ATLAS-CONF-2015-071.
[18] ATLAS Collaboration, ATLAS-CONF-2015-073.
[19] ATLAS Collaboration, ATLAS-CONF-2015-075.
[20] ATLAS Collaboration, ATLAS-CONF-2015-074.
[21] ATLAS Collaboration, *Phys. Lett. B*, **750** (2015) 475.