Jet reconstruction and substructure measurements in ATLAS and CMS with first Run-2 data

Matthias U. Mozer[*][†]
KIT
E-mail: mmozer@cern.ch

Jets play an important role in LHC physics but jet reconstruction and calibration in the high pile-up environment of the 2015 and 2016 data-taking periods pose unique challenges. The ATLAS and CMS experiments have improved their jet reconstruction and correction methods compared to Run I in order to better clean jets of the additional particles generated in pile-up interactions. Beyond jets formed from the hadronization of quarks and gluons, hadronic decays of highly boosted heavy particles, such as top quarks or Z, W or H bosons, in a single fat jet is gaining importance in the search for new physics at the highest possible energies. The LHC experiments have used the time between the LHC Run I and Run II to refine the methods used to identify such boosted decays and make them more robust in the presence of the high pile-up encountered in Run II. The application of this work to early Run II data are presented.

PoS(LHCP2016)090

---

[*]Speaker.

[†]on behalf of the ATLAS and CMS collaborations.

1. Introduction

Jets are collimated particle bundles formed in the hadronization of energetic color-charged particles, i.e. quarks and gluons. At the LHC, the anti-k_t algorithm [1] is most commonly used to combine individual detector signals into jets. In the ATLAS experiment [2], calorimeter clusters form the building blocks of jets, while in the CMS experiment [3], signals from all subdetectors are first combined in a particle flow (PF) algorithm, reconstructing electrons, muons, and photons, as well as neutral and charged hadrons [4, 5]. In order to infer information on the outgoing partons from the kinematics of the measured jets, it is important to have a proper understanding of effects that could spoil a direct correspondence between measured hadrons and the kinematics of the initial quarks and gluons. In particular, the contributions from soft interactions of the proton remnants (the so called underlying event) and additional collisions within the same bunch-crossing (pile-up) need to be accounted for. Especially the pile-up contributions are higher in the LHC Run II compared to Run I due to the increased instantaneous luminosity of the accelerator.

Final states with jets provide some of the most interesting signatures of new physics at the LHC, such as composite or excited quarks (see for example Ref. [6]) or supersymmetric decay chains (see for example Ref. [7]). Jet measurements also advance our knowledge of standard model physics, where jet cross sections are used to constrain parton densities and are used to measure the mass of the top quark. Because of their central role in many LHC physics analyses, an accurate reconstruction of jet kinematics and structure is important to fully exploit the physics potential of the LHC experiments.

During the LHC shutdown in 2013/2014, the machine has been significantly upgraded compared to its earlier running period, which achieved beam energies up to 4 TeV and instantaneous luminosities up to 7 Hz/nb, now reaching a beam energy of 6.5 TeV and luminosities up to 12 Hz/nb. Especially the higher instantaneous luminosity pose a challenge to jet physics as the increased pile-up (on average ~ 50 pile-up at 12 Hz/nb) needs to be adequately removed in order to retain the resolution obtained in Run I.

2. Jet Reconstruction Developments

The ATLAS and CMS experiments have used the phase between the LHC Run I and Run II to improve their jet reconstruction software to be more robust against pile-up and have adequate response and resolution even for the highest jet energies. Two results of the ATLAS collaboration can be highlighted, both related to the suppression of pile-up. Pile-up can be problematic in two different ways for jet-physics: the kinematics of a jet originating from the primary interaction can be adulterated by additional particles from pile-up interactions and randomly overlapping particles from pile-up can form so-called pile-up jets.

The first of these issues is tackled by applying appropriate correction factors to the kinematic quantities of the reconstructed jet. In Run I, the ATLAS collaboration used a correction scheme that provided correction factors as function of the the number of observed vertices in the event and the number of expected collisions in the corresponding bunch-crossing. For Run II this procedure is being replaced with a correction method that estimates the pile-up energy flow event-by-event from the charged particles not associated by their track to the primary vertex and uses this energy flow

to apply corrections proportional to the area covered by a given jet [8]. This method substantially improves the jet resolution at high pile-up (see Fig. 1), as it is better able to account for the large variance in pile-up contributions from event to event.
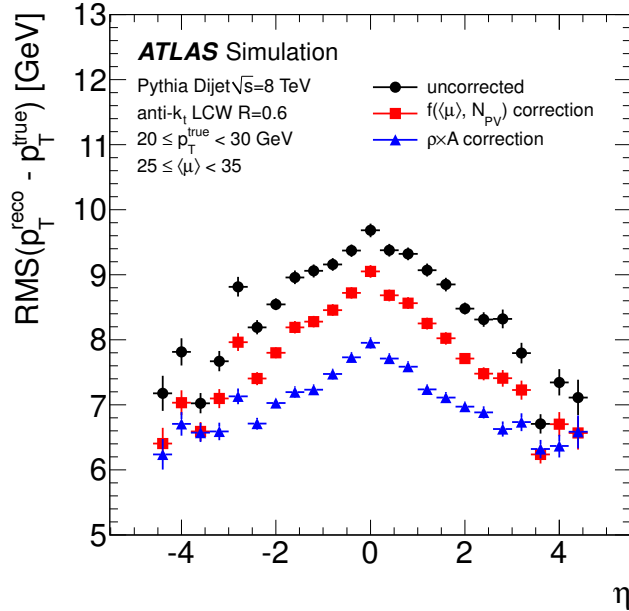

Figure 1: Jet Resolution for simulated events in the ATLAS detector as function of the jet pseudo-rapidity η using different pile-up correction techniques. [8]

The second issue can be ameliorated by using the so called Jet Veto Fraction (JVF), which computes the contribution a given jet receives from pile-up particles by comparing tracks from the primary and secondary vertices among all tracks geometrically linked to the jet. However, in its initial implementation as used during Run I, a simple cut on the JVF discriminant had an efficiency with significant dependence on the pile-up itself, complicating efficiency estimates. For Run II, the algorithm has been improved to take this variation into account, producing a discriminant that can easily be used over a wide range of pile-up scenarios [8].

To reduce the effect of pile-up particles on reconstructed jet kinematics, the CMS collaboration has implemented the Pileup Per Particle Identification (PUPPI) algorithm [9], which uses inter-particle correlations to assign a pile-up probability to each reconstructed particle. The probabilities are one (zero) for charged particles from the primary (pile-up) vertices, respectively, similar to the charged-hadron-subtraction method in common use by CMS. However, for neutral particles the probability is evaluated from the correlations with surrounding particles, leading to weighting factors between zero and one.

In addition to improvements in pile-up mitigation, the CMS collaboration has tuned the PF reconstruction to optimize performance for very high energy jets, which are more common at the higher beam energies of Run II. The improvements include a dedicated track reconstruction step focused on jets to reconstruct the many overlapping charged particles present in highly energetic jets. Additionally the reconstruction of calorimeter deposits is performed at a finer granularity

compared to Run I, significantly increasing the resolution of jet-structure variables, as shown in Fig. 2.
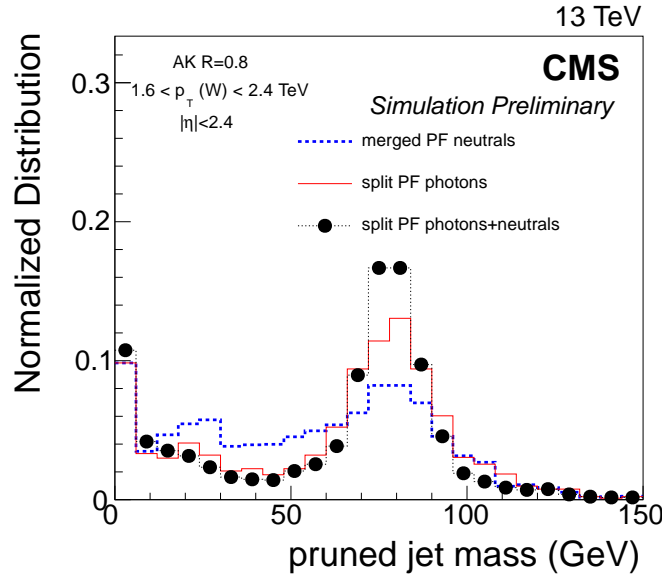

Figure 2: The reconstructed groomed jet mass for energetic hadronic W decays using different reconstruction parameters as simulated in the CMS experiment. [10]

3. Jet Substructure Developments

Energetic heavy particles that decay hadronically, such as W, Z or H bosons as well as top quarks may be reconstructed as single jets as the hadronization products of the decay quarks are collimated into a narrow region due to the high boost of the initial particle. As these particles have high hadronic branching fractions, it is advantageous to analyze these decays, but large backgrounds from QCD multijet production hinder such studies. In order to reduce the backgrounds, it is possible to use the known differences in the parton shower development of the multi-quark particle decays and single quark or gluon QCD multijet backgrounds to differentiate these processes.

A large number of possible discriminants have been devised, many of which fall into two broad categories:

- Variables that measure the invariant mass of the jet, correcting in different ways for the contributions of pile-up and the underlying event in a process generically labeled *grooming*.

- Variables that measure the compatibility of a given jet with a multi-prong structure, effectively reconstructing the decay quarks of the heavy particle as subjets.

The LHC experiments have used the shutdown before Run II for comprehensive reviews of these methods, evaluating their background suppression power, but more importantly also their resistance to high pile-up [10, 11, 12], which was not a major concern during Run I. These studies show that a reasonable performance can be obtained with relatively simple methods, combining

a groomed mass variable with a multi-prong discriminant. Nevertheless, substantial gains can be achieved if a large array of variables is combined in a multivariate discriminant.

Fig. 3 (left) compares the jet mass resolutions as simulated in the CMS experiment using different grooming algorithms. The best performance is achieved with the PUPPI algorithm. The charged hadron subtraction method and statistical corrections as used in Run I perform much worse. Fig. 3 (right), shows a similar study performed by the ATLAS experiment, determining the average groomed mass of simulated QCD jets and boosted boson decays as function of the number of pile-up events. While the ungroomed mass rises steadily with the pile-up, the mass can be kept very stable with suitably aggressive grooming as envisioned for the ATLAS Run II analysis.
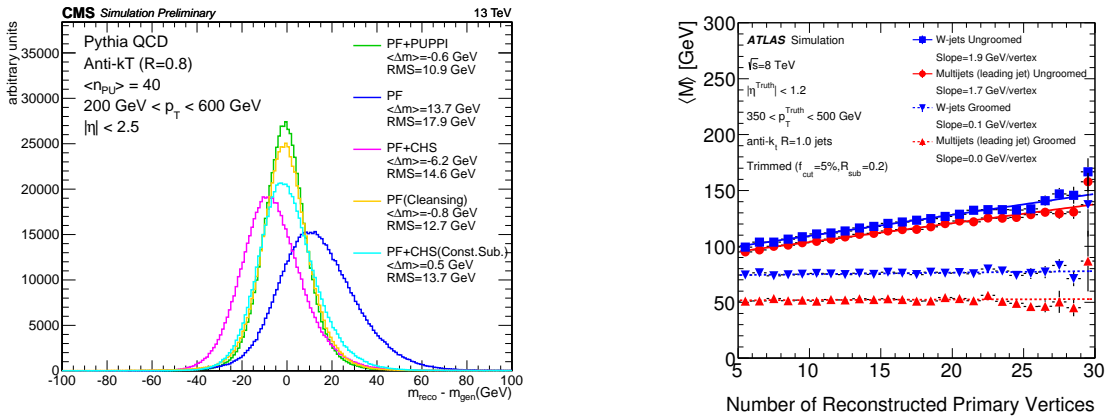

Figure 3: Left: Mass resolution for jets simulated in the CMS detector using different grooming methods for an average pile-up of 40 [11]. Mean jet mass for jets originating from boosted boson decays as well as from the hadronization of quarks and gluons as function of pile-up as simulated in the ATLAS experiment [12].

The performance of these methods are checked on data using samples of semi-leptonic $t\bar{t}$ events with high top-quark momenta. In such events, the leptonic top-quark decay as identified by a lepton and a b-jet recoils against the hadronic decay which provides a relatively pure sample of hadronic W boson decays. Thus, the events can be used for tag & probe studies of the performance of various boson tagging algorithm (see Fig. 4). While W boson decays are directly accessible with this method, there is no comparable sample for hadronic Z bosons and simulation is used to extrapolate results from hadronic W decays to Z decays.

Since the recent discovery of the Higgs boson [14, 15], the study of exotic processes involving the Higgs boson has received increased attention. The Higgs boson is very suited to the reconstruction with jet-substructure techniques in boosted scenarios due to its large branching fraction to b quarks. The identification of boosted Higgs bosons follows a similar strategy as for vector-bosons: a two-prong structure and a groomed jet mass close to the Higgs mass of ~ 125 GeV. As an additional identification criterion, the presence of b-quark signatures (displaced vertices, tracks with large impact parameters, etc.) can be used. However, with increasing boosts it becomes increasingly difficult to distinguish the decay particles of the two b-quarks and ambiguities arise when trying to identify both b-quarks separately (See Fig. 5). The issue may be approached by either decomposing the fat jet into subjets and using b-tag algorithms on each of these, or by using a suitable tagging algorithm on the whole fat jet [16, 17].
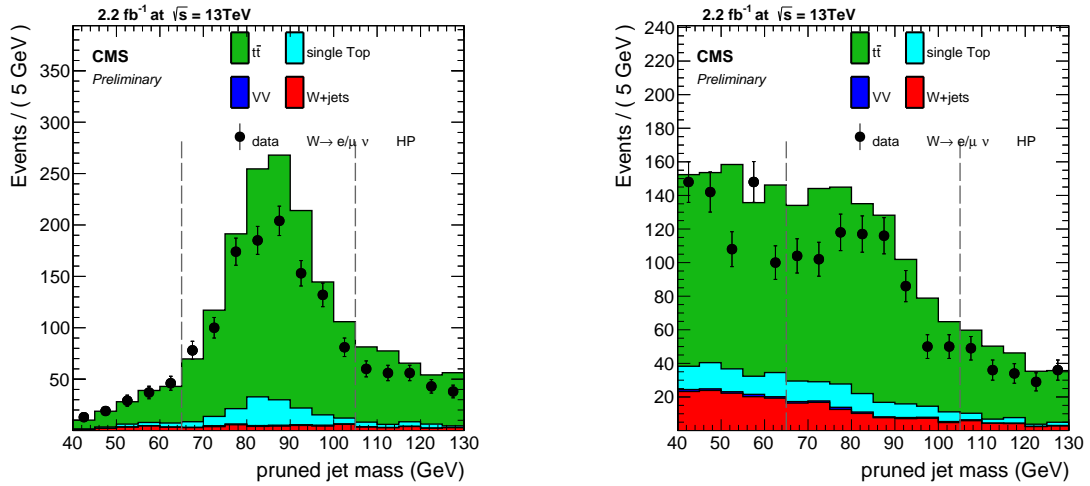
Figure 4: Measurement of the boosted boson tagging performance with the CMS experiment at 13 TeV. Left: groomed jet mass spectrum for a top-enriched sample passing tagging requirement. Right: groomed jet mass spectrum for a top-enriched sample failing tagging requirement. The boson tagging efficiency can be derived by comparing the peaking component of the two distributions. [13]
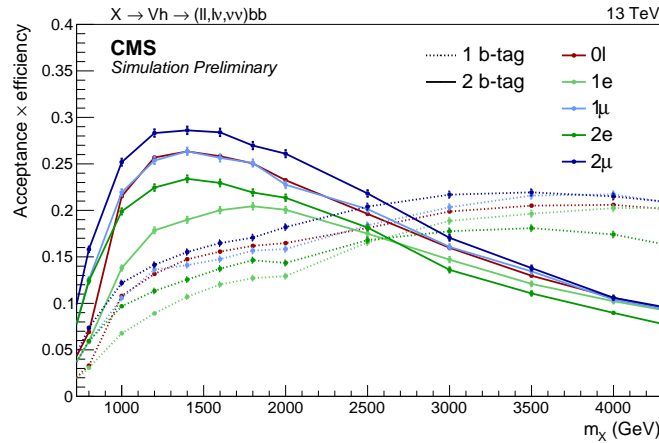

Figure 5: Efficiency for the identification of boosted Higgs bosons originating from the decay of a heavy resonance as function of the resonance mass as simulated in the CMS experiment. Solid lines show the efficiency for a requirement of two subjet b-tags, while dotted lines represent events that have only one subjet b-tag. [18]

At the energies available at the LHC even the top quark can be energetic enough for its decay products to be reconstructed within a single jet. The structure of such a top-jet is especially rich, as we expect three subjets, one of which is a b-jet and two of which combine into a W boson. Additionally, the top quark carries a color-charge so that the b-quark in the top decay retains a color connection to the rest of the event, leading to a more complicated parton shower compared to the decays of the W, Z and H bosons, which form color-neutral systems where the decay quarks are only color-connected to each other.

This rich phenomenology has produced an equally rich set of different top-tagging algorithms,

ranging from simple taggers based on just groomed mass and three-prong criteria to sophisticated likelihood discriminants, that take into account the complete parton shower dynamics of the top decay. A large number of these different approaches have been studied by the experiments [19, 20, 21]. One of the more commonly used taggers is the so called HEP-top-tagger [22], which is based around the identification of the W boson in the pairwise masses of three subjets, as shown in Fig. 6. However, in its initial implementation the algorithm used very wide jets optimized for relatively low energy top quarks, which limited performance in highly boosted final states. More recently, the algorithm has been improved to be highly efficient over a wide energy range by introducing a variable jet radius criterion [24]
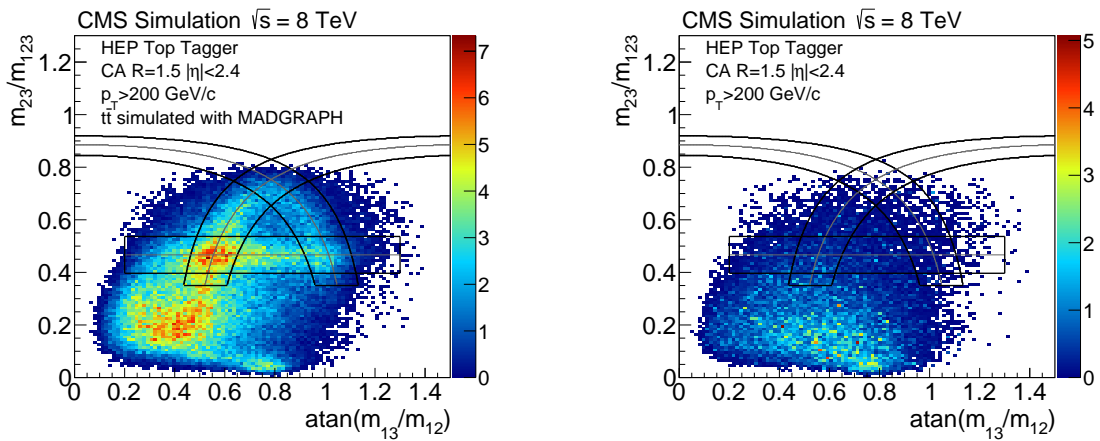

Figure 6: HEP-top-tagger variables as simulated in the CMS experiment. Signal (left) and background (right) events as function the invariant mass of different subject combinations. The black bounded areas denote the kinematic regions corresponding to the W-boson being captured by each of the different subject combinations and represent the selection region of the tagger. [19]

Similar to the boson tagging algorithms, for studies during Run II and, in the farther future, the HL-LHC require that the algorithms behave well at high pile-up or that they are insensitive to the commonly used pile-up reduction algorithms discussed above. As shown in Fig 7, the ATLAS collaboration shows that with sufficiently aggressive grooming, the mass resolution for boosted top quark decays can be well preserved even for extreme pile-up scenarios.

Even though much work has been performed during the shutdown to study sophisticated multi-variate top-tagging methods, initial 13 TeV analysis by the ATLAS and CMS experiments [26, 27], rely on relatively simple algorithms while the more complicated methods are still being commissioned. Nevertheless, convincing results are already achieved with simple three-prong discriminants, as shown in Fig 8, which shows a groomed top mass spectrum from the CMS experiment using early Run II data, demonstrating that the top mass can be reliably reconstructed even under the pile-up conditions seen during the 2015 data-taking period.

4. Summary

Jet physics and physics with boosted hadronic decays remain a staple of LHC physics in Run II, if they are not becoming even more important due to the higher LHC beam energies. The LHC
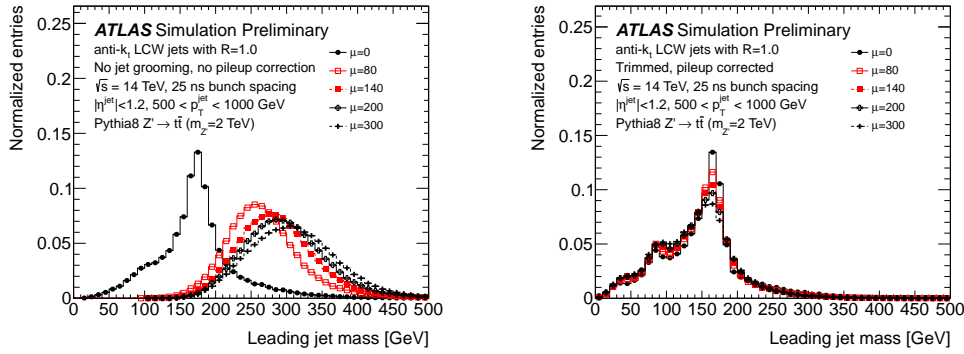
Figure 7: Raw (left) and groomed (right) reconstructed top quark mass for boosted top quarks in different pile-up scenarios as simulated in the ATLAS experiment [25].
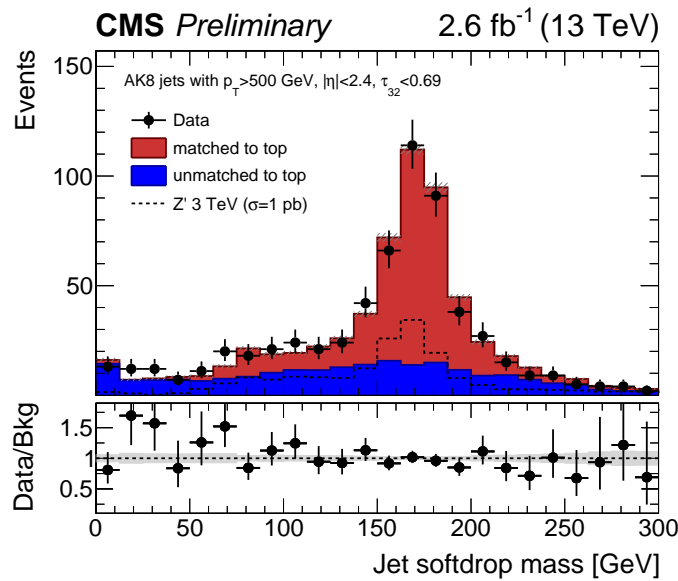

Figure 8: Mass spectrum for top-quark events in the CMS detector compared to simulation. Top quarks are enriched using an N-subjettiness requirement and the mass is groomed using the softdrop technique. [27]

experiments have improved their respective reconstruction methods, putting special emphasis in corrections for and robustness against the much increased pile-up. First results from the 2015 data-taking period are promising, but the highest numbers of pile-up are only expected to appear during 2016.

References

[1] M. Cacciari, G. P. Salam and G. Soyez, *The Anti-k(t) jet clustering algorithm,* JHEP **0804** (2008) 063 doi:10.1088/1126-6708/2008/04/063 [hep-ph:0802.1189].

[2] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider,* JINST **3** (2008) S08003. doi:10.1088/1748-0221/3/08/S08003

[3] CMS Collaboration, *The CMS experiment at the CERN LHC,* JINST **3** (2008) S08004. doi:10.1088/1748-0221/3/08/S08004

[4] CMS Collaboration, *Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV,* CMS-PAS-PFT-10-002.

[5] CMS Collaboration, *Particle-flow commissioning with muons and electrons from J/Psi and W events at 7 TeV,* CMS-PAS-PFT-10-003.

[6] CMS Collaboration, *Search for narrow resonances decaying to dijets in proton-proton collisions at $\sqrt{(s)} = 13$ TeV,* Phys. Rev. Lett. **116** (2016) no.7, 071801 doi:10.1103/PhysRevLett.116.071801 [hep-ex:1512.01224].

[7] ATLAS Collaboration, *Search for squarks and gluinos in final states with jets and missing transverse momentum at $\sqrt{s} = 13$ TeV with the ATLAS detector,* Eur. Phys. J. C **76** (2016) no.7, 392 doi:10.1140/epjc/s10052-016-4184-8 [hep-ex:1605.03814].

[8] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector,* hep-ex/1510.03823.

[9] D. Bertolini, P. Harris, M. Low and N. Tran, *Pileup Per Particle Identification,* JHEP **1410** (2014) 059 doi:10.1007/JHEP10(2014)059 [hep-ph:1407.6013].

[10] CMS Collaboration, *V Tagging Observables and Correlations,* CMS-PAS-JME-14-002.

[11] CMS Collaboration, *Pileup Removal Algorithms,* CMS-PAS-JME-14-001.

[12] ATLAS Collaboration, *Identification of boosted, hadronically decaying W bosons and comparisons with ATLAS data taken at $\sqrt{s} = 8$ TeV,* Eur. Phys. J. C **76** (2016) no.3, 154 doi:10.1140/epjc/s10052-016-3978-z [hep-ex/1510.05821].

[13] CMS Collaboration, *Search for massive resonances decaying into pairs of boosted W and Z bosons at $\sqrt{s} = 13$ TeV,* CMS-PAS-EXO-15-002.

[14] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,* Phys. Lett. B **716** (2012) 30 doi:10.1016/j.physletb.2012.08.021 [hep-ex:1207.7235].

[15] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,* Phys. Lett. B **716** (2012) 1 doi:10.1016/j.physletb.2012.08.020 [hep-ex:1207.7214].

[16] CMS Collaboration, *b-tagging in boosted topologies,* CMS-DP-2015-038.

[17] ATLAS Collaboration, *Expected Performance of Boosted Higgs ($\to b\bar{b}$) Boson Identification with the ATLAS Detector at $\sqrt{s} = 13$ TeV,* ATL-PHYS-PUB-2015-035.

[18] CMS Collaboration, *Search for heavy resonances decaying into a vector boson and a Higgs boson in the (ll, lnu, nunu) bb final state,* CMS-PAS-B2G-16-003.

[19] CMS Collaboration, *Boosted Top Jet Tagging at CMS,* CMS-PAS-JME-13-007.

[20] CMS Collaboration, *Top Tagging with New Approaches,* CMS-PAS-JME-15-002.

[21] ATLAS Collaboration, *Identification of high transverse momentum top quarks in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector,* JHEP **1606** (2016) 093 doi:10.1007/JHEP06(2016)093 [hep-ex:1603.03127].

[22] T. Plehn, M. Spannowsky, M. Takeuchi and D. Zerwas, *Stop Reconstruction with Tagged Tops,* JHEP **1010** (2010) 078 doi:10.1007/JHEP10(2010)078 [hep-ph:1006.2833].

[23] D. E. Soper and M. Spannowsky, *Finding top quarks with shower deconstruction,* Phys. Rev. D **87** (2013) 054012 doi:10.1103/PhysRevD.87.054012 [`hep-ph:1211.3140`].

[24] G. Kasieczka, T. Plehn, T. Schell, T. Strebler and G. P. Salam, *Resonance Searches with an Updated Top Tagger,* JHEP **1506** (2015) 203 doi:10.1007/JHEP06(2015)203 [`hep-ph:1503.05921`].

[25] ATLAS Collaboration, *Jet substructure at very high luminosity,* `twiki.cern.ch/twiki/bin/view/AtlasPublic/JetSubstructureECFA2014`

[26] ATLAS Collaboration, *Search for heavy particles decaying to pairs of highly-boosted top quarks using lepton-plus-jets events in proton–proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector,* ATLAS-CONF-2016-014

[27] CMS Collaboration, *Search for t$\bar{\text{t}}$ resonances in boosted semileptonic final states in pp collisions at $\sqrt{s}$ = 13 TeV,* CMS-PAS-B2G-15-002.

PoS(LHCP2016)090