

Future perspectives for jet substructure techniques in LHC Run2

Matthias Mozer^{1,a}

¹ *Institut für Experimentelle Kernphysik, KIT
Wolfgang-Gaede-Str. 1, 76131 Karlsruhe, Germany*

Abstract. The increased pile-up expected in the LHC Run 2 and High Luminosity LHC creates a challenging environment for utilizing the jet-substructure techniques which were successfully demonstrated in the LHC Run 1. The ATLAS and CMS experiments are studying a range of methods to improve jet reconstruction to increase the resilience against high pile-up. Promising results are obtained in simulation but await validation on the first Run 2 data.

1 Introduction

In the search for new physics at the LHC, jets play a dominant role. They are of particular importance in studies involving electroweak bosons, top-quarks and the Higgs boson, as these particles decay into hadron with large branching ratios. At the high energies probed at the LHC the decay hadrons will often form a single jet, so called boosted topologies. The distribution of particles within such jets differs substantially from jets originating from the hadronization of single quarks or gluons and techniques that exploit these differences have been used successfully to suppress QCD induced backgrounds. The jet characteristics used in the substructure methods originate from the initial parton composition of the jets and the QCD evolution of the parton shower. However, in the presence of many simultaneous interactions (“pile-up”), the signatures can be diluted by the presence of additional jet constituents (calorimeter clusters, tracks...) not originating from the primary interaction. In the upcoming LHC Run 2, the number of simultaneous interactions is expected to rise to ~ 40 , compared to ~ 25 in Run 1, and even more for the future High Luminosity LHC (“HL-LHC”) upgrade. Under these harsh conditions, the performance of jet substructure techniques, but also basic reconstruction, such as the measurement of the jet transverse momentum (p_T) are negatively impacted. The LHC experiments have started to investigate various methods to reduce the effects of pile-up on jet measurements in order to retain or even improve the good performance observed in Run 1.

The LHC experiments are approaching this issue from three directions:

- improving the basic event reconstruction of the experiments.
- removing pile-up particles before the event is interpreted as a specific final state.

^ae-mail: Matthias.Mozer@kit.edu

- refining physics interpretation tools such as b-tagging and jet-substructure techniques to be more resilient against pile-up.

2 Improved Reconstruction

Before data are reconstructed for analysis, events have to pass the trigger selection in order to be permanently stored. The ATLAS [1] and CMS [2] experiments both use a multi-stage trigger design, where the first step of the selection is performed by hardware designed for this purpose (referred to as L1 trigger) and the last step is implemented as a somewhat reduced version of the full reconstruction run on a dedicated but otherwise generic cluster of computers (called HLT). The ATLAS experiment implements an intermediate step, while the CMS experiment uses just the two steps discussed above. The whole range of jet-substructure techniques can be relatively easily implemented in the HLT at the only cost of some additional processing time. However, for triggers that solely rely on boosted jet identification, the L1 trigger can impose a serious bottle neck. To improve on this situation, the ATLAS collaboration is currently implementing an upgrade to their L1 trigger system that allows for wider jets [3]. Figure 1 (left), shows how this upgraded trigger can increase the efficiency in identifying hadronic top quark decays.

The CMS collaboration has invested a significant effort into improving the existing Particle Flow (PF) [6, 7] reconstruction to more accurately measure the properties of very energetic jets. In previous versions of this reconstruction techniques, the large multiplicity of tracks in a very energetic jet could overwhelm the tracking algorithm, leading to a significant deficit in the fraction of charged particles within a very energetic jet. This issue has been cured by the addition of additional tracking steps, tuned specifically to recover tracks aligned with the jet axis. Additionally,

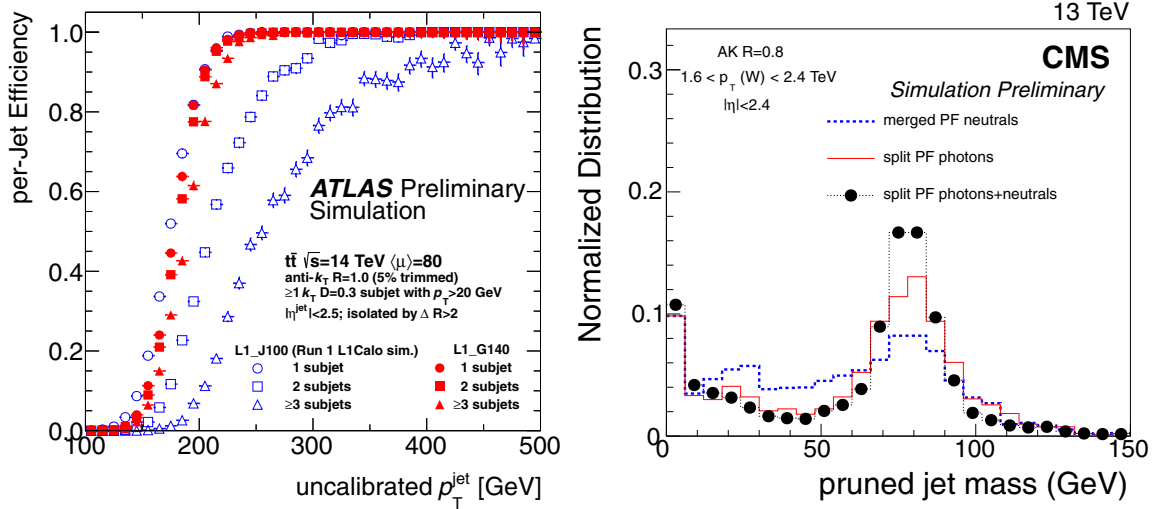


Figure 1. Left: Simulated efficiency of new (L1_G140) and previous (L1_J100) L1 triggers of the ATLAS detector as function of jet p_T using boosted top quarks as case study. Large efficiency gains are observed in the case of boosted topologies with two or three subjects. Right: Reconstructed jet mass for a hypothetical resonance with an invariant mass of 4 TeV decaying to two W bosons. The dotted line shows the mass reconstructed with the pruning [4, 5] algorithm using the PF configuration as used in Run 1; the solid line adds higher granularity treatment of deposits in the electromagnetic calorimeters and the points show the results when the hadronic calorimeters are treated with finer granularity as well.

overlapping calorimeter deposits are treated with a finer granularity, greatly improving the reconstruction of the jet mass for very energetic jets. As an example, Figure 1 (right) from [8] shows the jet mass distribution for a hypothetical resonance with an invariant mass of 4 TeV decaying to two W bosons. Using the methods described above, the reconstruction of the W mass improved substantially.

Additionally, algorithms that identify the hadronization of b-quarks (“b-tagging”) are being transferred from their initial use in jets originating from a single b-quark to the case of boosted hadronic decays with at least one b-quark. Initial studies were driven by boosted top decays, but with the discovery of the Higgs boson and its high branching ratio to b-quark pairs, additional attention has been focused on boosted hadronic Higgs decays.

The CMS collaboration has largely relied on reusing the same algorithms tuned to the identification of jets originating from single b-quarks also to identify b-quarks within larger decay chains [9, 10]. While this approach foregoes the opportunity for further optimization, it allows for a very rapid use of b-tagging in boosted topologies and focuses limited manpower resources. Even without specific optimizations, this approach leads to large gains in the identification of boosted top quark (see Figure 2, left) and Higgs boson decays. Conversely, the ATLAS collaboration is optimizing additional b-tagging algorithms specifically for their performance within hadronic decays of boosted heavy particles [11, 12]. The results are very promising, as shown in Figure 2 (right), and excellent efficiencies and background rejection rates can be achieved.

3 Pile-Up Removal

Once reconstructed, particles from pile-up may be removed from further consideration by directly identifying them, or alternatively, by statistically subtracting their contributions from other variables, such as jet momenta or lepton isolation variables. Such corrections were already necessary in the LHC Run 1 in order to meet expected resolutions and efficiencies. For Run 1, the predominant corrections were applied based on the jet- or isolation-cone area and proportional to the pile-up activity measured in a given event [13, 14].

In the CMS experiment, the PF reconstruction additionally allowed individually reconstructed charged particles to be classified as pile-up according to the association of their measured tracks to the vertices in the event. While this method (called charged hadron subtraction, CHS) has the advantage that the contributions from charged pile-up particles is subtracted exactly, neutral pile-up contributions can only be subtracted statistically.

In a recent study [15], the CMS has investigated more sophisticated methods to reduce the pile-up distribution. The so called constituent subtraction method [16] works similarly to the statistical jet correction described above, but instead of defining a jet area, an effective area and corresponding correction is assigned to each reconstructed particle. Additionally the Pileup Per Particle Identification (PUPPI) algorithm [17] was studied, 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 CHS method. However, for neutral particles the probability is evaluated from the correlations with surrounding particles, leading to weighting fac-

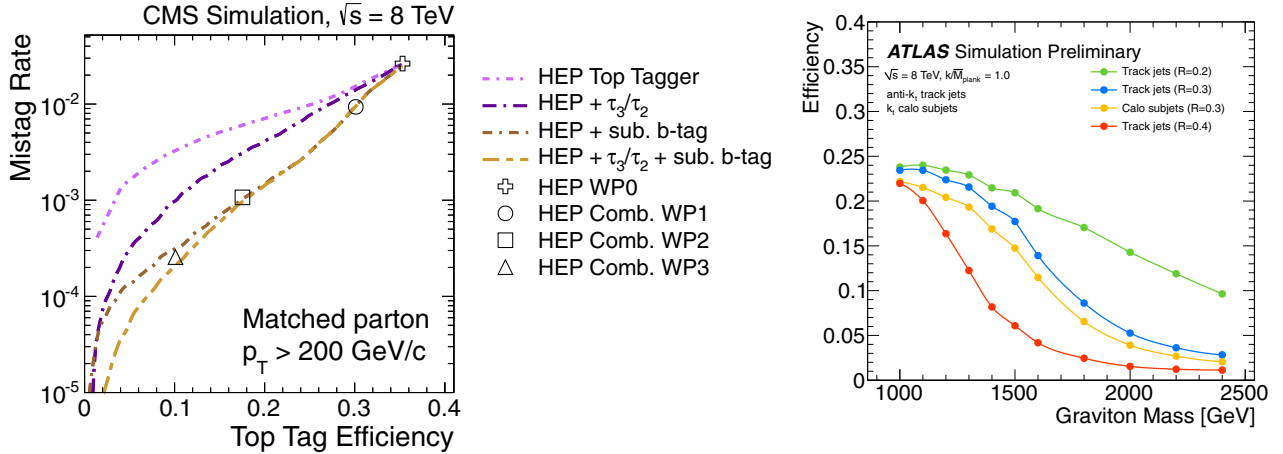


Figure 2. B-tagging performance in boosted objects.

Left: CMS top tagging performance with (brown lines) and without (purple lines) using subjet b-tags to select boosted top-quark decays.

Right: Efficiency to select a resonance decaying to two hadronically decaying Higgs bosons using several different working points for subjet b-tagging with the ATLAS detector.

tors between zero and one. Figure 3 compares the jet mass spectra and resolutions for QCD jets and boosted hadronic W decays using the different algorithm. The best performance is achieved with the PUPPI algorithm, closely followed by the constituent subtraction. CHS and statistical corrections as used in Run 1 perform much worse.

In Run 1, the so called Jet Veto Fraction (JVF) was widely used in the ATLAS collaboration to suppress jets largely consisting of pile-up particles. The JVF is defined as the sum of the p_T of the tracks within a jets area coming from the primary vertex, divided by the sum of the p_T of all tracks within the jets area. A requirement for the JVF to exceed a given threshold was then employed to select jets originating from the primary vertex. However, the fixed cut requirement leads to a noticeable dependence of the selection efficiency on the number of simultaneous collisions in the event, as shown in Figure 4, left. This issue is solved by two improvements: For future use, the JVF computation is rescaled to explicitly take into account the dependence on the number of vertices in the event. Additionally the JVF variable is combined with a variable related to the fraction of charged particles within a jet to further reduce the pile-up dependence of the variable and increase overall tagging efficiency [18].

4 Jet-Substructure Techniques

In the LHC Run 1, a number of different jet-substructure techniques have been employed. In the CMS collaboration, the so called pruning algorithm [19] has seen widespread use in order to reconstruct the invariant mass of hadronic W - and Z -boson. The algorithm is designed to clean the jet from soft and wide-angle particles. Jets are reclustered with a modified Cambridge-Aachen (CA) algorithm [20], where combinations only proceed if the constituents pass criteria on their relative angle and transverse momentum.

In addition the jet pruning, results of the ATLAS collaboration have been prepared with the trimming [21] and mass-drop [23] algorithms (see Ref. [24] for corresponding performance studies). In the trimming algorithm a jet with wide radius parameter is reclustered with a narrower radius parameter. Of the resulting subjets, only the ones carrying a certain fraction of the wide jets transverse momentum are retained. The mass drop algorithm is applied to jets reconstructed with the CA algorithm. In each step, the most recent combination of the CA algorithm is undone and it is checked whether the more massive of the two resulting subjets carries more than a certain fraction of the mass of the parent jet, in which case the lighter subjet is discarded and the procedure continues on the remaining subjet.

In preparation for the HL-LHC and the large amount of pile-up expected, the ATLAS collaboration has revisited the topic of jet substructure algorithms. The results show (see Figure 5, left) that even in the extreme case of 200 simultaneous interactions, the mass of hadronically decaying top-quarks can be well reconstructed.

In preparation for Run 2, the CMS collaboration has performed a comprehensive study of several jet cleaning algorithms in combination with methods that reduce pile-up at the reconstruction level [15]. Together with the pruning and trimming algorithms, discussed above, the so called soft drop algorithm [22] (an evolution of the mass drop algorithm described above) was investigated in combination with the CHS and PUPPI techniques described above. Figure 6 shows the mass resolution for the different combinations of algorithms. Similar to the ATLAS results, the trimming algorithm emerges as particularly advantageous. In addition to good stability against pile-up, the trimming algorithm has less pronounced tails in its resolution as estimated by comparing a Gaussian fit of the resolution to its RMS.

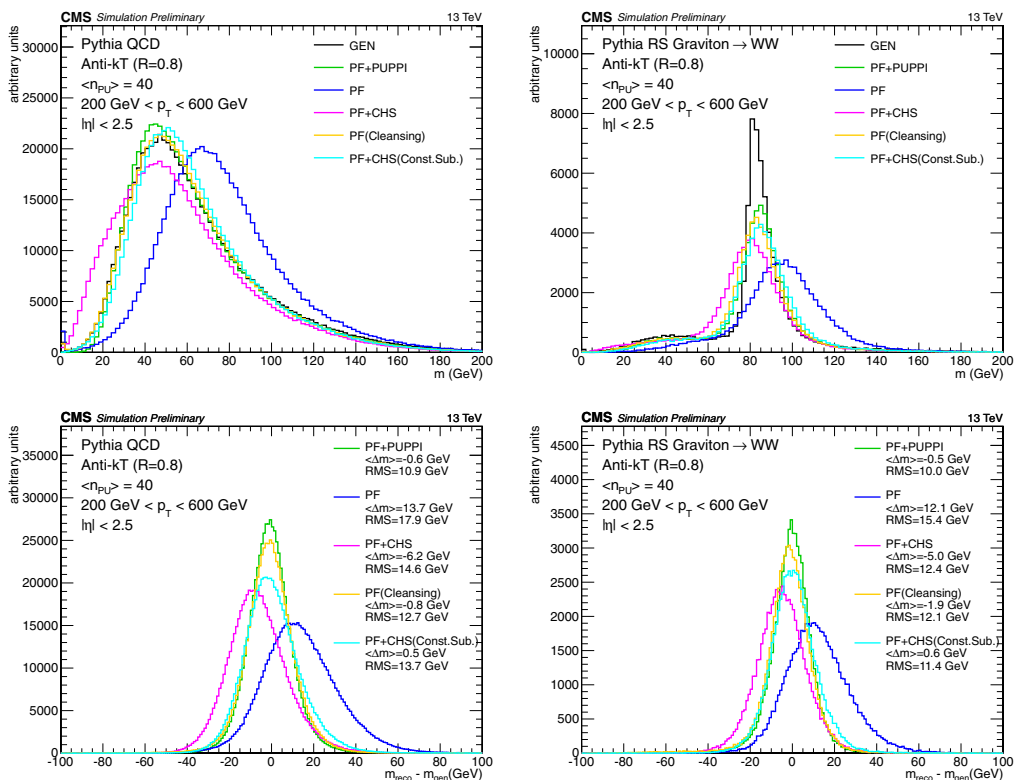


Figure 3. Mass distributions (top) and resolutions (bottom) for QCD jets (left) and boosted hadronic W decays (right) using several different pile-up mitigation techniques.

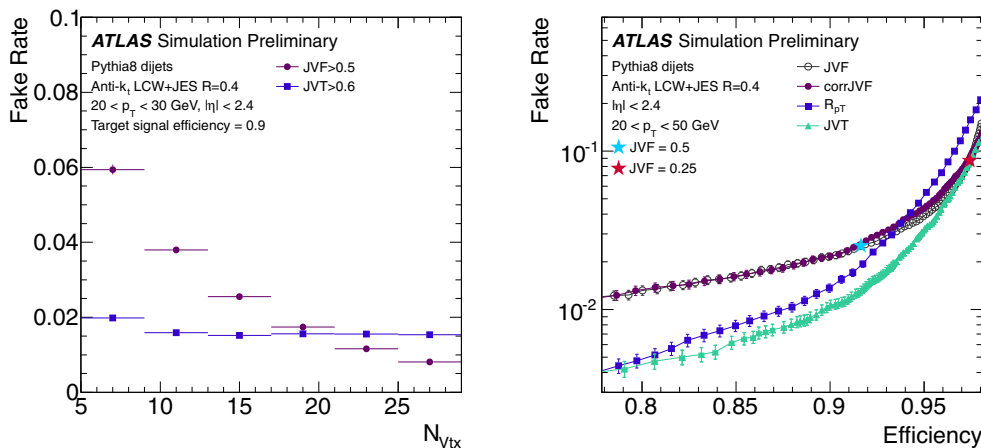


Figure 4. Left: Comparison of the pile-up-dependence of the JVF used in Run 1 to the improved combined variable (JVT). Right: Fake rate as function of tagging efficiency for the JVF used in Run 1 compared to the scaled JVF (corrJVF), the additional charge-fraction variable (R_{PT}) and the combination (JVT).

5 Summary and Outlook

The LHC experiments are demonstrating with simulation studies, that the previously used jet reconstruction and substructure techniques are affected by the increased pile-up expected in Run 2 and the HL-LHC. However, improved reconstruction and substructure methods have been devised to minimize the impact of the additional pile-up.

Nevertheless, these methods will have to be validated and possibly revised when high pile-up data-taking starts with the LHC Run 2 in 2015.

References

[1] ATLAS Collaboration, JINST 3 S08003 (2008)
 [2] CMS Collaboration, JINST 3 S08004 (2008)

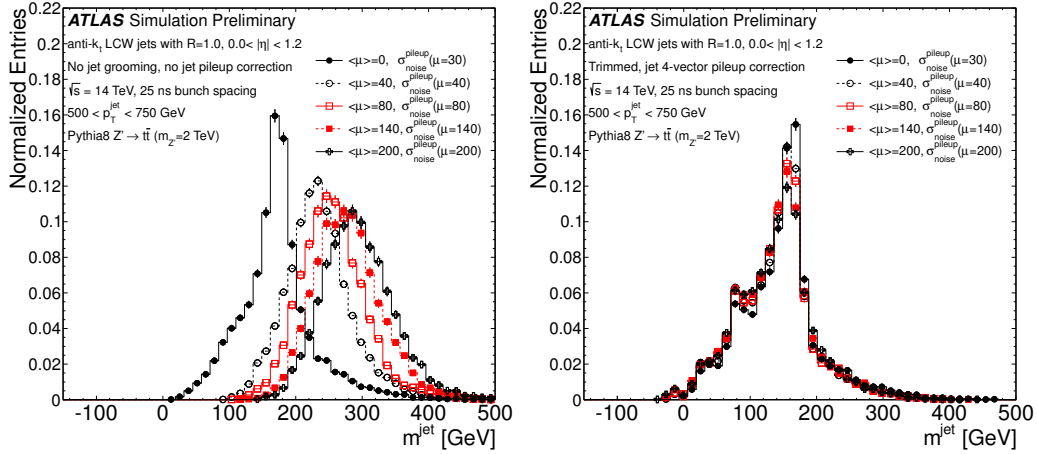


Figure 5. Reconstructed mass for boosted top quarks in the ATLAS detector with different amounts of pile-up without (left) and with (right) jet-grooming techniques and pile-up corrections applied.

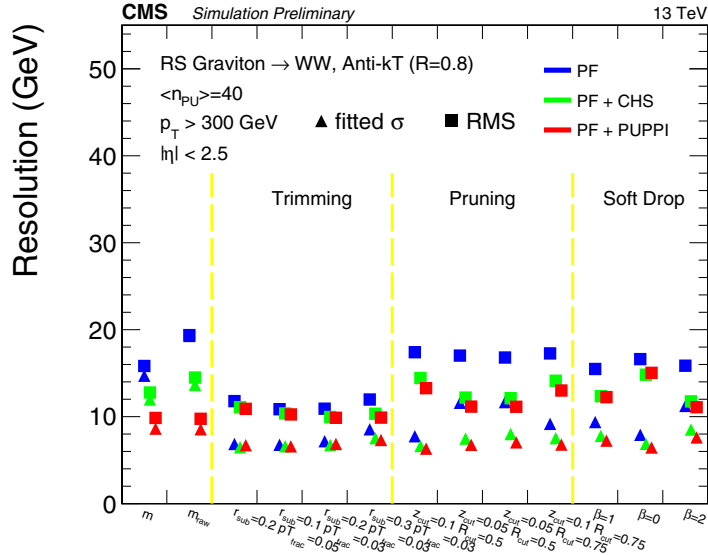


Figure 6. Comparison of the jet mass resolution of a variety of jet grooming algorithms in combination with pile-up reduction techniques. The trimming, pruning and soft drop algorithms are combined with CHS and PUPPI pile-up reduction techniques. Shown are mass resolutions from a Gaussian fit as well as the RMS values of the distributions, indicating the size of non-Gaussian tails in the resolution.

- [3] ATLAS Collaboration, *Global Feature Extraction (gFEX) Performance Plots*, ATLAS-COM-DAQ-2014-087 (2014), <https://cds.cern.ch/record/1749167>
- [4] Ellis, S. D. *et al.*, Phys. Rev. D **80**, 094023 (2010),
- [5] Ellis, S. D. *et al.*, Phys. Rev. D **81**, 094023 (2010)
- [6] CMS Collaboration, *Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taups, and E_T^{miss}* , CMS-PAS-PFT-09-001 (2009), <https://cds.cern.ch/record/1194487>
- [7] CMS Collaboration, *Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector*, CMS-PAS-PFT-10-001 (2010), <https://cds.cern.ch/record/1247373>

- [8] CMS Collaboration, *V Tagging Observables and Correlations*, CMS-PAS-JME-14-002 (2014), <https://cds.cern.ch/record/1754913>
- [9] CMS Collaboration, *Boosted Top Jet Tagging at CMS*, CMS-PAS-JME-13-007 (2014), <https://cds.cern.ch/record/1647419>
- [10] CMS Collaboration, *Performance of b tagging at $\sqrt{s}=8$ TeV in multijet, $\bar{t}t$ and boosted topology events*, CMS-PAS-BTV-13-001 (2013), <https://cds.cern.ch/record/1581306>
- [11] ATLAS Collaboration, *Flavor Tagging with Track Jets in Boosted Topologies with the ATLAS Detector*, ATL-PHYS-PUB-2014-013 (2014), <https://cds.cern.ch/record/1750681>

- [12] ATLAS Collaboration, *b-tagging in dense environments*, ATL-PHYS-PUB-2014-014 (2014), <https://cds.cern.ch/record/1750682>
- [13] M. Cacciari, G. P. Salam and G. Soyez, *JHEP* **0804**, 005 (2008)
- [14] M. Cacciari and G. P. Salam, *Phys. Lett. B* **659**, 119 (2008)
- [15] CMS Collaboration, *Study of Pileup Removal Algorithms for Jets*, CMS-PAS-JME-2014-001 (2014), <https://cds.cern.ch/record/1751454>
- [16] P. Berta, M. Spousta, D. W. Miller and R. Leitner, *JHEP* **1406** 092 (2014)
- [17] D. Bertolini, P. Harris, M. Low and N. Tran, *JHEP* **1410** 59 (2014)
- [18] ATLAS Collaboration, *Search for high-mass states with one lepton plus missing transverse momentum in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, ATLAS-CONF-2014-018 (2014), <https://cds.cern.ch/record/1700870>
- [19] CMS Collaboration, arXiv:1410.4227 [hep-ex].
- [20] M. Wobisch and T. Wengler, In **Hamburg 1998/1999, Monte Carlo generators for HERA physics** 270-279 [hep-ph/9907280].
- [21] D. Krohn, J. Thaler and L. T. Wang, *JHEP* **1002** (2010) 084
- [22] A. J. Larkoski, S. Marzani, G. Soyez and J. Thaler, *JHEP* **1405** (2014) 146
- [23] J. M. Butterworth, A. R. Davison, M. Rubin and G. P. Salam, *Phys. Rev. Lett.* **100** (2008) 242001
- [24] ATLAS Collaboration, *JHEP* **1309** (2013) 076