

Performance of boosted object and jet substructure techniques in Run 1 and 2 ATLAS data

Steven Schramm*, on behalf of the ATLAS Collaboration

Universite de Geneve (CH)

E-mail: steven.schramm@cern.ch

Hadronic decays of heavy particles with momenta much larger than their mass result in their decay products being reconstructable as a single large-radius jet. The study of the substructure of these jets allows the separation of these boosted decays with respect to more common jets from light-quarks and gluons. Several techniques have been developed by the phenomenology and experimental community to identify jets coming from hadronic decays of boosted top quarks, W, Z and Higgs bosons. The performance of several such techniques has been studied in ATLAS using fully-simulated Monte Carlo events, and validated on data using pure samples of top quarks, W bosons from top decays and dijet events. Results of these studies will be presented for Run 1 as well as Run 2 of the LHC.

*38th International Conference on High Energy Physics
3-10 August 2016
Chicago, USA*

*Speaker.



1. Introduction

During 2015 and 2016, the LHC delivered proton-proton collisions at a centre of mass energy of $\sqrt{s} = 13 \text{ TeV}$ and an instantaneous luminosity surpassing the design value of $1 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. The resulting dataset contains high-energy hard scatter events, including the production of massive Standard Model particles, such as W/Z/H bosons and top quarks. Hadronic decays of these particles produce 2 or 3 quarks respectively, which each hadronize to form a collimated stream of particles. These particle groups deposit energy in the ATLAS [1] calorimeters, and are reconstructed as *jets*.

While massive particle decays occur back-to-back in the parent particle rest frame, the Lorentz boost to the laboratory frame can present a different picture. If the parent particle has a low p_T , then the boost factor is small, and the decay products remain mostly back to back. As the p_T of the parent particle is increased, the boost factor grows correspondingly and thus the decay products become more collimated. Roughly speaking, the angular separation $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ between the decay products of a particle ρ with a mass m_ρ and transverse momentum p_T^ρ can be related using $\Delta R \approx 2m_\rho/p_T^\rho$. For hadronic decays, a small value of ΔR results in overlapping jets, where the individual decay products cannot be resolved. In this scenario, large-radius (large- R) jets become a useful way to represent the entire hadronic decay, and thus become a proxy for the parent particle.

The ATLAS Collaboration currently uses large- R jets built using the anti- k_r [2] jet clustering algorithm with a distance parameter R of 1.0. The jets are trimmed [3] to remove pileup contributions and then calibrated to account for the non-uniform and non-compensating response of the ATLAS calorimeters to hadronic particles.

2. Jet mass optimization

The primary use of large- R jets in ATLAS is for their ability to represent the decay of a single massive particle. One of the most important properties of large- R jets is thus their mass, which should correspond to the mass of the parent particle, thus allowing for the identification of final states resulting from the top quark or W/Z/H boson decays. The ATLAS calorimetry is sufficient to accurately reconstruct the mass of hadronic decays for most jets. However, at very high p_T (very high collimation), the spatial resolution of the calorimeter becomes a limitation. In searches for new physics, these very high p_T jets typically form the most interesting regime of parameter space, thus it is important to find alternative ways of improving the large- R jet mass calculation.

The use of tracks is a logical next step to consider, as the ATLAS tracking detector has much finer spatial resolution than the calorimeters. However, tracks require the presence of a charged particle, and jets contain a sizable fraction of neutral constituents. Calculating a jet mass purely from tracks (m^{track}), is thus less effective (worse mass resolution) than calculating a jet mass purely from calorimeter information (m^{calo}), even at very high p_T , as seen in Figure 1.

Ultimately, the ideal situation is to use the calorimeter energy resolution and the tracker spatial resolution. This can be done using the *track-assisted mass* m^{TA} , with $m^{\text{TA}} = m^{\text{track}} \cdot p_T^{\text{calo}}/p_T^{\text{track}}$. This scales the track mass, with its superior spatial resolution, by the ratio of calorimeter to track p_T , thus fixing the scale of the variable by adding neutral contributions. In Figure 1, this is shown to have a better mass resolution than either the calorimeter or track mass definitions. The track-assisted mass also has a smaller uncertainty [4], making this a powerful new discriminant.

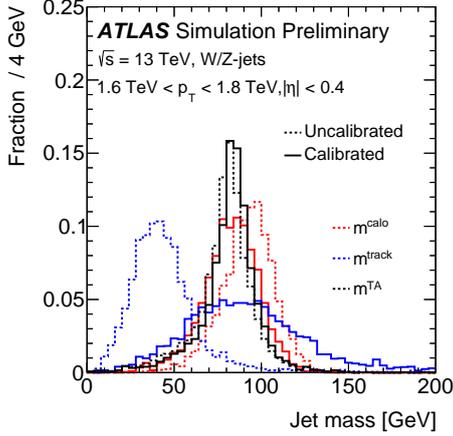


Figure 1: A comparison of hadronically decaying W and Z bosons for the normalized calorimeter, track, and track-assisted jet masses (m^{calo} , m^{track} , and m^{TA}) for a parent boson p_T between 1.6 and 1.8 TeV [4].

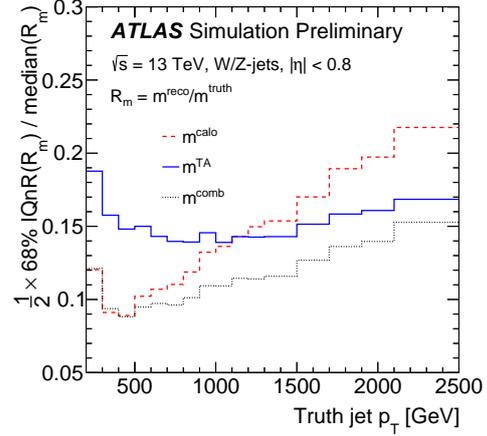


Figure 2: A comparison of hadronically decaying W and Z bosons for the calorimeter, track-assisted, and combined jet mass resolutions (m^{calo} , m^{TA} , and m^{comb}) as a function of the parent boson p_T [4].

The track-assisted mass may work very well at high p_T , but it is still a quantity which is only accounting for neutral contributions on average. The calorimeter mass can account for localized neutral contributions, and thus it performs better until the calorimeter's spatial resolution is insufficient. This is seen to occur at a p_T of approximately 1.2 TeV for hadronically decaying W and Z bosons in Figure 2. Note that this crossing point will depend on the parent particle, as the calorimeter performance depends on the collimation ΔR which varies with the parent particle mass.

The calorimeter and track-assisted mass each have a regime of parameter space in which they are the best option, and so it is sensible to form a *combined mass* m^{comb} from the two variables. This is shown to perform optimally over the full parameter space in Figure 2, and is calculated using weights based on their mass resolutions σ :

$$m^{\text{comb}} = m^{\text{calo}} \cdot \frac{\sigma_{\text{calo}}^{-2}}{\sigma_{\text{calo}}^{-2} + \sigma_{\text{TA}}^{-2}} + m^{\text{TA}} \cdot \frac{\sigma_{\text{TA}}^{-2}}{\sigma_{\text{calo}}^{-2} + \sigma_{\text{TA}}^{-2}}.$$

3. Hadronic tagging performance

In preparation for the $\sqrt{s} = 13$ TeV dataset, ATLAS has defined new taggers for the hadronic decays of W/Z bosons, $H \rightarrow b\bar{b}$, and top quarks. The W/Z boson tagger provides a QCD multijet rejection factor falling roughly linearly from 60 to 30, as the p_T is varied from 200 GeV to 2 TeV, for a flat signal efficiency of 50% [5]. The $H \rightarrow b\bar{b}$ tagger combines b-tagging with substructure information to provide rejection factors of 10^5 (6) for light-quark and gluon jets (bb-jets, such as $g \rightarrow b\bar{b}$) for a signal efficiency of roughly 50% [6]. Tagging of hadronic top decays at a flat 50% (80%) signal efficiency provides QCD multijet rejection factors falling roughly linearly from 16 to 8 (5 to 4) as the p_T is varied from 400 GeV to 1.7 TeV [7].

The hadronic W and top taggers have now been evaluated *in situ* in 2015 data using a semi-leptonic $t\bar{t}$ topology, without the application of large- R jet mass selections. The simulated dataset is observed to match the data within the associated uncertainties, as shown in Figure 3, demonstrating

that the taggers have been properly optimized to identify hadronic decays. Figure 3 also clearly demonstrates that the subsequent application of mass cuts around the W or top mass peaks will provide further background rejection, leaving a sample relatively pure in the object of interest.

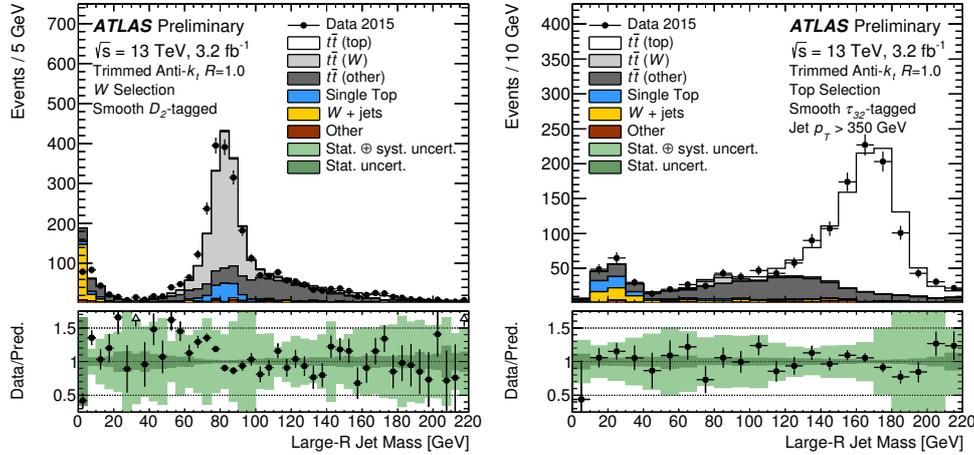


Figure 3: The application of the hadronic W tagger (*left*) and top tagger (*right*) in 2015 data using a semi-leptonic $t\bar{t}$ topology, before applying a large- R jet mass selection. The simulation and data are observed to agree within the associated uncertainties [8].

4. Summary

ATLAS has developed taggers for the hadronic decays of W/Z/H bosons and top quarks, which were optimized for and now evaluated in $\sqrt{s} = 13$ TeV data. The taggers are performing well, but are less optimal at high p_T due to the limitations of the calorimeter spatial resolution. A new approach to define a mass discriminant using both track and calorimeter information appears very promising, and will be of great use in probing high-energy hadronic decays in the years to come.

References

- [1] ATLAS collaboration, *ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) 003.
- [2] M. Cacciari, G. P. Salam and G. Soyez, *The Anti- $k(t)$ jet clustering algorithm*, *JHEP* **04** (2008) 063, [0802.1189].
- [3] D. Krohn, J. Thaler and L.-T. Wang, *Jet Trimming*, *JHEP* **02** (2010) 084, [0912.1342].
- [4] ATLAS collaboration, *Jet mass reconstruction with the ATLAS Detector in early Run 2 data*, *ATLAS-CONF-2016-035* (2016).
- [5] *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).
- [6] *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).
- [7] *Boosted hadronic top identification at ATLAS for early 13 TeV data*, *ATL-PHYS-PUB-2015-053* (2015).
- [8] ATLAS Collaboration, “Public Plots for Top and Boson Tagger Performance.” <http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/JETM-2016-005/>.