Soft *b*-hadron tagging with the ATLAS detector

Tomohiro Yamazaki∗ **on behalf of the ATLAS Collaboration**
The University of Tokyo
E-mail: tomohiro.yamazaki@cern.ch

The identification of *b*-hadrons at low p_T can play a crucial role in a variety of analyses where soft *b*-hadrons are produced by physics beyond the Standard Model such as stop/sbottom production in compressed SUSY senarios, and also in analyses that require heavy flavour jet vetos. Such *b*-hadron identification is a particularly challenging task, owing to the relatively short decay length of such *b*-hadrons and difficulties identifying hadronic jets, and current flavour tagging algorithms are not tuned to identify such *b*-hadron decays. This article presents new developments for soft *b*-hadron tagging techniques based either on the presence of jets composed only of charged tracks, or on the identification of secondary vertices from the *b*-hadron decay without requiring the presence of a jet. The techniques are described and their performance is reviewed.

European Physical Society Conference on High Energy Physics - EPS-HEP2019 -
10-17 July, 2019
Ghent, Belgium

∗Speaker.

1. Introduction

The identification of b-hadrons (b-tagging) [1] is a crucial technique in many physics programmes at the LHC. The b-tagging exploits the long lifetime of the b-hadrons, utilising the presence of the secondary vertices and significant displacement of the track impact parameters. The standard b-tagging technique in ATLAS [2] identifies jets that contain a b-hadron, therefore the b-tagging is only available for jets with $p_T > 20$ GeV. In the compressed stop and sbottom production where low-p_T b-hadrons are contained in the final states, the search sensitivity is limited because the standard b-tagging is not optimised for b-hadrons with $p_T < 20$ GeV [3–5]. To increase the sensitivity for these compressed stop and sbottom searches, three low-p_T b-tagging algorithms have been developed [6]. This article describes the algorithms of the three low-p_T b-tagging methods in Section 2, and the performance of them is reviewed in Section 3. Finally, the conclusions of this work are drawn in Section 4.

2. Algorithms

2.1 Track jet b-tagging

A track jet is built from tracks reconstructed from hits of charged particles in the inner detector using the anti-k_t algorithm [7]. In the standard b-tagging, track jets with $p_T > 10$ GeV can be the inputs of the b-tagging algorithm. In order to identify lower p_T b-hadrons than the p_T range of the standard b-tagger, the p_T threshold of the track jet is lowered to 5 GeV. The variable size of the track jet cone [8] is used. The size of the cone is $R = \rho/p_T$, and $\rho = 30$ GeV, $R_{min} = 0.02$, and $R_{max} = 0.4$ for all jets. In the low-p_T regime, however, this means effectively $R = 0.4$ for all jets. The MV2c10 algorithm [1], which is a multivariate-based tagging algorithm used in the standard b-tagging, is also used for the track jet b-tagging at low p_T.

2.2 Low-p_T vertex taggers

The second approach to access low-p_T b-hadrons relies on the secondary vertex reconstruction from tracks which are not associated with a jet. Two vertex-based algorithms have been developed as follows.

2.2.1 Track-based Low-p_T Vertex Tagger (T-LVT)

The first method of the vertex-based approaches is the *Track-based Low-p_T Vertex Tagger* (T-LVT). There are three steps in the T-LVT to identify low-p_T b-hadrons. The first step is the track selection, which is an essential part of this algorithm because reducing the number of input tracks before the vertex fitting can avoid an explosion of the fake rate and the processing time. As shown in Figure 1 (a) tracks that have large displacement from the primary vertex are selected by requiring large d_0 and z_0 significance. If the ΔR between the track and any calorimeter jets with $p_T > 30$ GeV is smaller than 0.4, the track is removed from the input list. The second step is vertex fitting. In order to reduce the combination of the tracks which are used in the vertex fitting, all pairings of two tracks are tested first. After that, tracks that have no reconstructed vertex at the two-track vertex fitting stage are discarded and the remaining tracks are used in the n-track vertex fitting ($n \geq 2$). The final step of the algorithm is the selection of the reconstructed vertices. The vertex

reconstruction algorithm targeting *b*-hadron decays in the vicinity of the primary vertex easily produces fake vertices formed by random crossings of tracks from the primary vertex. To reduce the fake vertices, the reconstructed secondary vertices are selected at the final step. A minimum distance from the primary vertex to the reconstructed vertex and the significance of the distance are required for the selected vertices to reduce the random crossings, and also there is a requirement on the reconstructed mass of the vertex. The vertex mass is calculated from the tracks associated with the vertex assuming the charged pion mass for the associated tracks. Vertices with a mass lower than 600 MeV are rejected as shown in Figure 1 (b). This requirement can strongly reduce vertices from $K_S \rightarrow \pi^+\pi^-$, which is a long-lived particle with $m_{K_S} = 498$ MeV [9].
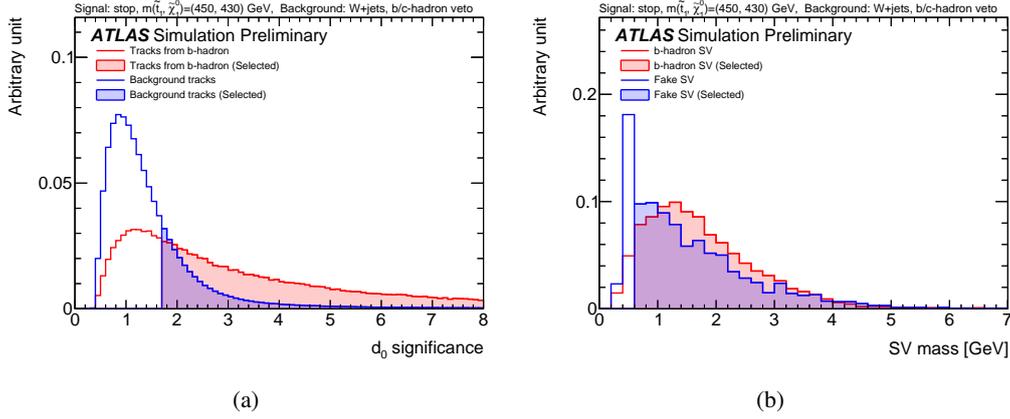

(a)   (b)

Figure 1: The discriminant variables used in the T-LVT algorithm [6]. (a) d_0 significance of the tracks originating from *b*-hadrons in red and tracks from the background process (*W*+jets without any heavy flavour hadrons) in blue are shown. (b) The mass of the reconstructed secondary vertices. The secondary vertices matched to a truth secondary vertex in the signal event is shown in red, and fake secondary vertices in the background process in blue. In both (a) and (b), the other cuts used in the track/vertex selection are applied and the selected tracks/vertices are shown as the colour-filled histograms.

2.2.2 Track-Cluster-based Low-p_T Vertex Tagger (TC-LVT)

The second vertex-based tagger is the *Track-Cluster-based Low-p_T Vertex Tagger* (TC-LVT). The basic idea of the vertex reconstruction is similar to that of the T-LVT, but a different strategy is used in the track selection. In the TC-LVT, a seed track is selected as a first step of the track selection. The seed track is required to have a large impact parameter significance. After selecting the seed tracks, tracks that satisfy a different selection (referred to as cluster tracks) are used to form a cluster of tracks around the seed track. The cluster tracks are selected by requiring further high-d_0 significance of the track. There are also requirements on the angular variables, such as the angle between the seed track and the cluster track. In the vertex fitting and the vertex selection, a similar strategy to the T-LVT is used in the TC-LVT.

3. Performance

The performance of the algorithms described above is evaluated with the Monte Carlo (MC) simulation and the recorded data with ATLAS in 2015-2018. Figure 2 shows the inclusive efficiency (= acceptance × tagging efficiency) as a function of the truth *b*-hadron p_T. Since the

track-jet b-tagging is optimised for high p_T b-hadrons, it has high efficiency above 20 GeV of the b-hadron p_T, but it covers low p_T down to approximately 15 GeV as well. Since the T-LVT is a dedicated tagger aiming to tag b-hadrons that are not tagged by the standard b-tagging, it has a different p_T dependence of the efficiency compared to the track jet b-tagging. The maximum efficiency of the T-LVT is about 25% at the p_T range of 15–20 GeV. The efficiency of the T-LVT drops at high p_T because tracks overlapping with a calorimeter jet with $p_T > 30$ GeV are removed.
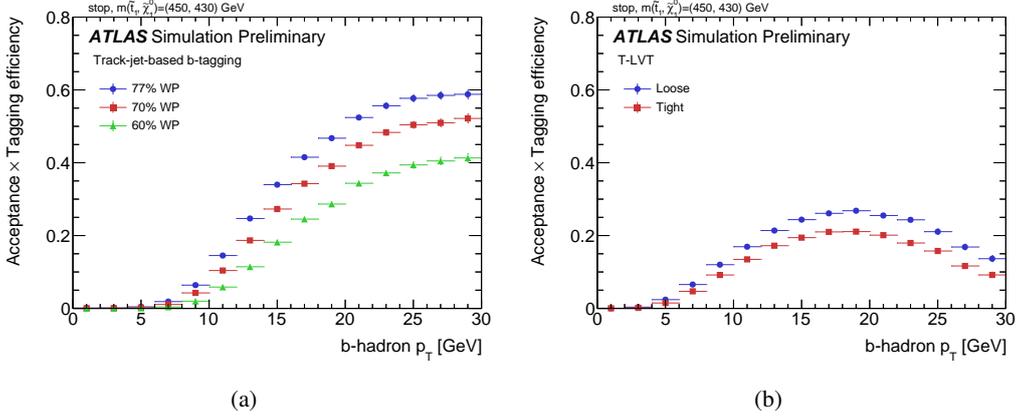

Figure 2: The acceptance × tagging efficiency of the (a) track jet b-tagging and (b) T-LVT as a function of the truth b-hadron p_T [6]. The performance is evaluated with the stop signal. Three (two) working points corresponding to different efficiencies are available in the track jet b-tagging (T-LVT).

The performance of the fake rejection is shown in Figure 3. Figure 3 (a) shows the fake rate of the taggers, which is defined as the average number of the fake vertices per signal event. Since the efficiencies of the track jet b-tagging and the low-p_T vertex taggers have different dependence on the p_T of the b-hadrons, the track jet b-tagging efficiency is tuned to the T-LVT efficiency to compare the fake rate with the same efficiency in the given p_T range. The fake rate is comparable between these three algorithms for p_T of $15 - 20$ GeV, however, the vertex taggers have a stronger fake rejection in the p_T range of $5 - 10$ GeV. Figure 3 (b) shows the inclusive efficiency as a function of the fake rate. The new three low-p_T taggers improve the b-tagging performance at low p_T significantly, and T-LVT shows the best performance among these taggers.

In order to use the new algorithms in the physics analyses, the performance of the algorithms in data needs to be well modelled using MC simulation. The region which is dominated by the dileptonic $t\bar{t}$ process is used to evaluate the data and MC agreement. In this region, events are required to contain a different-flavour opposite-sign lepton pair ($e\mu$), and at least two calorimeter jets with $p_T > 30$ GeV, one of those to be b-tagged. Figure 4 (a) shows the data and MC comparison of the number of the secondary vertices reconstructed by the T-LVT algorithm, and Figure 4 (b) shows the distribution of the vertex mass. In Figure 4 (b), the normalisation of the MC events is scaled by 1.16 to compare the shapes of the distributions. The agreement between the data and MC is overall reasonable.

4. Conclusion

Three low-p_T b-taggers have been developed to improve the b-tagging performance for low-p_T
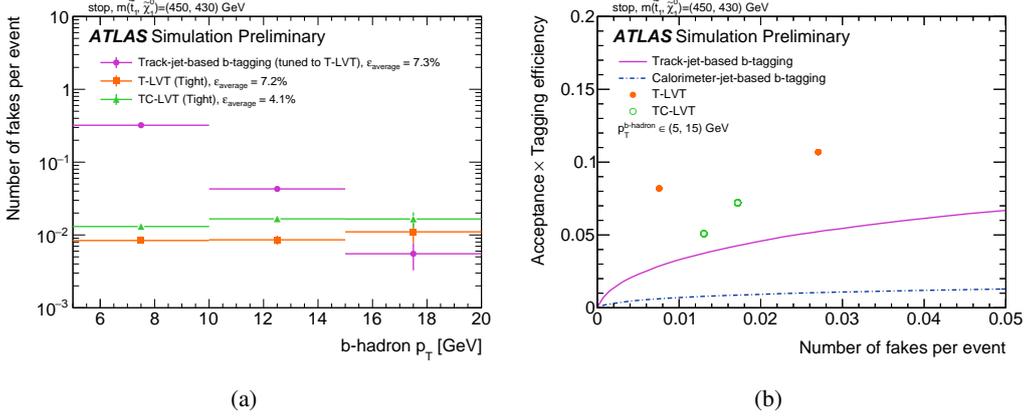
(a)

(b)

Figure 3: Comparison of the *b*-tagging algorithms at low p_T [6]. (a) Number of fake vertices per signal event as a function of the truth *b*-hadron p_T. The track jet *b*-tagging efficiency is tuned to the T-LVT efficiency. (b) Acceptance times efficiency as a function of the number of fake vertices per signal event. Truth *b*-hadron p_T in the range of $5 - 15$ GeV is shown. The standard *b*-tagging based on calorimeter jets is also shown for the comparison.
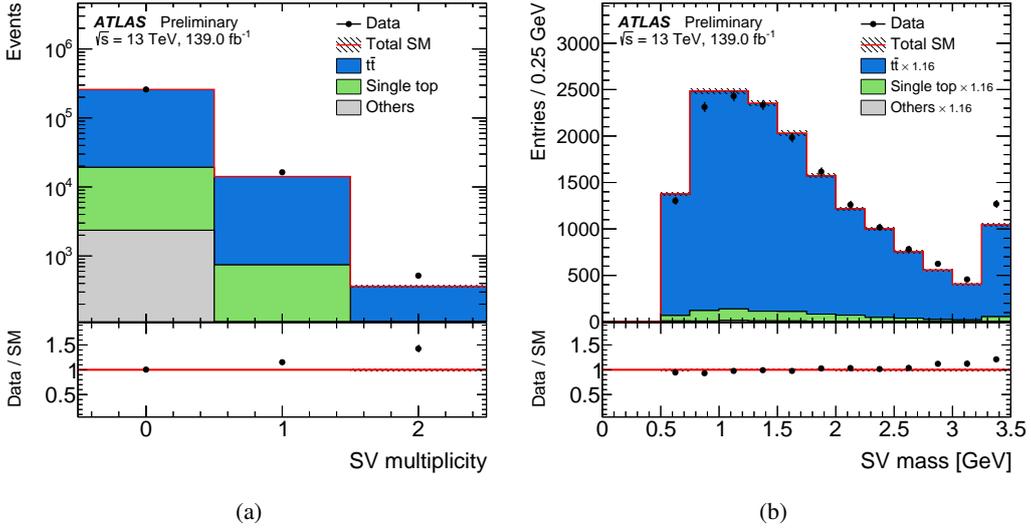

(a)

(b)

Figure 4: Comparison of the data and MC of the T-LVT algorithm with the $t\bar{t}$ selection [6]. (a) Number of the T-LVT vertices and (b) reconstructed mass of the vertices. In (b), the MC events are scaled by 1.16 to adjust the normalisation of the distribution. Only statistical uncertainties are displayed. The last bin includes the overflow.

b-hadrons. The p_T threshold of the track jet *b*-tagging is lowered from 10 to 5 GeV, while the other two methods exploit the secondary vertex reconstruction without relying on jets. The performance of these taggers is evaluated and a significant improvement compared to the standard *b*-tagging is achieved for $p_T < 15$ GeV. The new methods are expected to improve the search sensitivity of the compressed stop/sbottom searches.

References

[1] ATLAS Collaboration, *Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV*, JHEP **08** (2018) 089, `arXiv:1805.01845 [hep-ex]`.

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

[3] ATLAS Collaboration, *Search for top-squark pair production in final states with one lepton, jets, and missing transverse momentum using* 36 fb^{-1} *of* $\sqrt{s} = 13$ TeV *pp collision data with the ATLAS detector* , JHEP **06** (2018) 108, `arXiv:1711.11520 [hep-ex]`.

[4] ATLAS Collaboration, *Search for supersymmetry in events with b-tagged jets and missing transverse momentum in pp collisions at* $\sqrt{s} = 13$ TeV *with the ATLAS detector*, JHEP **11** (2017) 195, `arXiv:1708.09266 [hep-ex]`.

[5] ATLAS Collaboration, *Search for a scalar partner of the top quark in the jets plus missing transverse momentum final state at* $\sqrt{s} = 13$ TeV *with the ATLAS detector*, JHEP **12** (2017) 085, `arXiv:1709.04183 [hep-ex]`.

[6] ATLAS Collaboration, *Soft b-hadron tagging for compressed SUSY scenarios*, ATLAS-CONF-2019-027, 2019, `https://cds.cern.ch/record/2682131`.

[7] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, `arXiv:0802.1189 [hep-ph]`.

[8] D. Krohn, J. Thaler, and L.-T. Wang, *Jets with Variable R*, JHEP **06** (2009) 059, `arXiv:0903.0392 [hep-ph]`.

[9] Particle Data Group, *Review of Particle Physics*, Phys. Rev. D **98** (2018) 030001, `https://link.aps.org/doi/10.1103/PhysRevD.98.030001`.