PROCEEDINGS OF SCIENCE

The boosted $X \rightarrow b\bar{b}$ tagger calibration using $Z \rightarrow b\bar{b}$ events collected with the ATLAS detector

PoS(ICHEP2022)1089

ATL-PHYS-PROC-2022-100
26 October 2022

Daariimaa Battulga[*] **on behalf of the ATLAS Collaboration**
Institut für Physik, Humboldt-Universität zu Berlin
Newtonstr.15, 12489 Berlin, Germany
E-mail: daariimaa.battulga@cern.ch

Many analyses in the ATLAS physics program at the LHC are dependent on the identification of jets containing b-hadrons (b-tagging). The corresponding algorithms are referred to as b-taggers. The baseline b-taggers are optimized for jets containing one b-hadron. A new double b-tagging algorithm, the $X \rightarrow b\bar{b}$ tagger, provides better identification efficiency to reconstruct boosted resonant particles decaying into a pair of b-quarks. In the boosted regime, it is a challenging task because of high collimation of the two b-hadrons. This neural network based $X \rightarrow b\bar{b}$ tagger uses the kinematic information of the large radius (R=1.0) jet and the flavour information of associated track-jets. The performance of this tagger was evaluated using Monte Carlo simulation, therefore it could vary in collision data. Thus this poster presents the in situ tagging efficiency calibration using $Z \rightarrow b\bar{b}$ events with a recoiling photon or jet for this boosted $X \rightarrow b\bar{b}$ tagger. The efficiency data to simulation scale factor is derived using the Run 2 pp collision data collected by the ATLAS experiment at $\sqrt{s} = 13$ TeV, with the integrated luminosity of 139 fb^{-1}.

41st International Conference on High Energy physics - ICHEP2022
6-13 July, 2022
Bologna, Italy

---

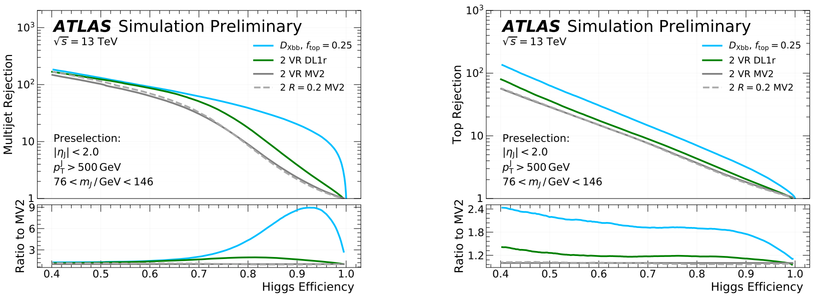
[*] Speaker

1. Introduction

The identification of jets containing b-hadrons (b-tagging) is an essential part of many physics analyses at the ATLAS experiment [1] at the LHC. One clear example would be the searches for processes involving the Higgs boson, whose predominant decay channel is into a pair of bottom quarks. When the Higgs boson is produced with a high Lorentz boost (high transverse momentum, p_{T}), its decay products can be highly collimated. In that case, fragmentations of these b-jets are clustered within a large-radius (large-R) jet with parameter $R = 1.0$. At this boosted topology, one needs an efficient tagger to identify and reconstruct b-jets inside the large-R jet. The so-called $X \to b\bar{b}$ tagger [2] aims to improve b-tagging efficiency at a high p_{T} regime. The high-level input, large-R jet kinematic variables, and the flavour information of up to three variable-radius track jets are used for the $X \to b\bar{b}$ tagger neural network algorithm. The outputs are the probabilities of the large-R jet being Higgs matched-jet (p_{Higgs}), top-matched jet (p_{top}), and multijet (p_{multijet}). These are combined in one single discriminant, D_{Xbb} score. The formula of this D_{Xbb} is defined as the log-likelihood ratio of these probabilities,

$$D_{Xbb} = \ln\left(\frac{p_{\mathrm{Higgs}}}{f_{\mathrm{top}} \cdot p_{\mathrm{top}} + (1 - f_{\mathrm{top}}) \cdot p_{\mathrm{multijet}}}\right). \tag{1}$$

Here, in Eq. 1, the $f_{\mathrm{top}} = 0.25$ is a fixed value for the fraction of top background. Double b-tagging efficiency of the tagger is defined using the D_{Xbb} score. Three working points (WPs) are defined, corresponding to 50%, 60% or 70% efficiency. This note will discuss the results for the 60% WP.



(a) **(b)**

Figure 1: The distributions show the Higgs efficiency as a function of multijet and top rejection comparing different taggers in 1a and 1b, respectively. The performance of the $X \to b\bar{b}$ tagger (cyan) is compared to DL1r (green) and MV2 (grey) taggers [5].

The performance of boosted $X \to b\bar{b}$ tagger is shown in Figure 1, the Higgs efficiency as a function of multijet and top rejection comparing different taggers in 1a and 1b, respectively. The D_{Xbb} tagger (cyan) shows better multijet rejection at higher Higgs efficiency compared to DL1r (green) and MV2 (grey) taggers [5] for large-R jets with $p_{\mathrm{T}} > 500$ GeV in Randall-Sundrum (RS) graviton signal sample.

2. Signal efficiency calibration of the boosted jet tagger

The performance of the tagger is evaluated using Monte Carlo (MC) simulation. The efficiency of the tagger could differ when it is used in the observed pp collision data. Hence, data-simulation scale factors (SFs) are derived to match the b-tagging efficiency in data. The SF is defined as the efficiency ratio between data and MC,

$$\mathrm{SF} = \frac{\varepsilon_{\mathrm{data}}}{\varepsilon_{\mathrm{MC}}} = \frac{N^{\mathrm{data}}_{\mathrm{passed}}/N^{\mathrm{data}}_{\mathrm{total}}}{N^{\mathrm{MC}}_{\mathrm{passed}}/N^{\mathrm{MC}}_{\mathrm{total}}} = \frac{N^{\mathrm{data}}_{\mathrm{passed}}/N^{\mathrm{MC}}_{\mathrm{passed}}}{N^{\mathrm{data}}_{\mathrm{total}}/N^{\mathrm{MC}}_{\mathrm{total}}} = \frac{\mu_{\text{post-tag}}}{\mu_{\text{pre-tag}}}. \tag{2}$$

The efficiency is the number of events passing the b-tagging requirement over the total number of events. This is measured in data and MC. As shown in Eq. 2, the SF is derived using the signal strength ratio between post- ($\mu_{\text{post-tag}}$) and pre-tag ($\mu_{\text{pre-tag}}$). The $\mu_{\text{post-tag}}$ is measured in the $X \to b\bar{b}$ tagged region, where the full decay of $Z \to b\bar{b}$ is predominantly selected. However, before the tagging requirement the selected large-R jet may contain the full decay of the Z-boson. $N^{\mathrm{data}}_{\mathrm{total}}$ can be defined as,

$$N^{\mathrm{data}}_{\mathrm{total}} = \frac{N^{\mathrm{MC}}_{\mathrm{total}}}{N^{\mathrm{MC},\ell\ell}_{\mathrm{total}}} \cdot N^{\mathrm{data},\ell\ell}_{\mathrm{total}} = \mu_{\text{pre-tag}} \cdot N^{\mathrm{MC}}_{\mathrm{total}}, \tag{3}$$

using $Z \to ee(\mu\mu)$ events. Since the branching ratio of the $Z \to \ell\ell$ is well known, we can ignore other decay channels for this pre-tag signal strength. The post-tag signal strength was derived in the tagged region by performing a likelihood fit to the large-R jet mass distribution. While $\mu_{\text{pre-tag}}$ is derived using only the dileptonic ($ee, \mu\mu$) decay channel of the Z-boson as defined as Eq. 3. One advantage of this method is to cancel theoretical uncertainties for the SFs. The signal efficiency calibration is performed using $Z \to b\bar{b}$ events with a recoiling photon or a recoiling jet to cover soft and hard p_{T} regions, respectively. Dominant background contributions arise from multijet and γ+jet for $Z \to b\bar{b}$+jets and $Z \to b\bar{b} + \gamma$, respectively. The SFs in four p_{T} bins of the large-R jet are derived by the ratio of signal strengths in post- and pre-tag regions. Figure 2 shows the $X \to b\bar{b}$ tagging signal efficiency correction as a function of large-R jet p_{T}. Dominant uncertainties of SFs for $Z+\gamma$ are spurious signal, and statistical; for the Z+jets, the fit model, jet mass resolution as well as Z-modelling.

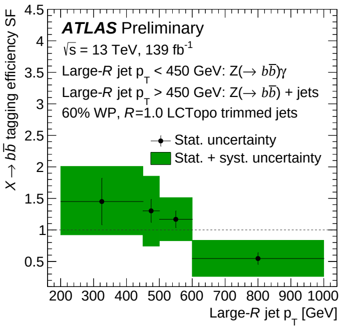
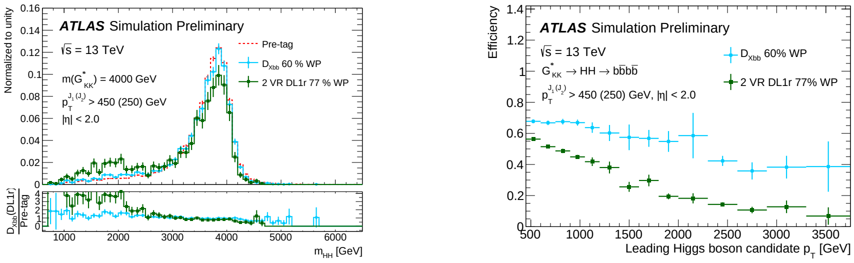


Figure 2: The $X \to b\bar{b}$ 60% working point tagging efficiency scale factors as a function of large-R jet transverse momentum (p_{T}) [3].

3. Performance study of the boosted tagger in boosted resonant diHiggs signal

This section presents a preliminary study of the $X \to b\bar{b}$ tagging efficiency compared to the current DL1r b-tagger used in the boosted resonant $HH \to b\bar{b}b\bar{b}$ search [6]. For this study, the MC simulation (produced at leading order with MadGraph5_aMC@NLO2.2.2 [7]), of spin-2 Kaluza-Klein gravitons decaying into a pair of Higgs bosons at the mass range between 1 TeV and 6 TeV are used [4]. The two highest p_{T} large-R jets are chosen as the Higgs boson candidates. The leading (subleading) large-R jet is required to have p_{T} greater than 450 (250) GeV and $|\eta| < 2.0$. The mass of the two large-R jets is greater than 50 GeV. Moreover, angular selection, $2m/p_{\mathrm{T}} < 1.0$, is applied to fully contain the decay products within the large-R jet radius. As shown in the Figure 3, the $X \to b\bar{b}$ tagger has a better mass resolution compared to the standard DL1r at higher graviton masses, e.g., $m(G^*_{KK}) = 4$ TeV. Figure 3b shows that the b-tagging efficiency increases by $20 - 110\%$ as a function of transverse momentum of the Higgs boson candidate in the graviton sample compared to DL1r.



(a) **(b)**

Figure 3: Invariant mass distribution of the diHiggs system shown in Figure 3a comparing the before (red dashed) and after b-tagging with Xbb 60% WP (cyan) and DL1r 77% WP (green) per variable-radius track jets [4]. Figure 3b show the b-tagging efficiency as a function of Higgs boson candidate. Only statistical uncertainties are considered here.

4. Conclusion

Identification of b-jets has always been a central component in the ATLAS physics program for the LHC. The boosted $X \to b\bar{b}$ tagger has been developed to improve the b-tagging efficiency in higher p_{T} regions. The signal efficiency calibration of boosted $X \to b\bar{b}$ tagger at 60% WP, using $Z \to b\bar{b}$ events with a recoiling photon or jet, is presented in this report. The SFs range between $1.45^{+0.51}_{-0.54}$ and $0.51^{+0.29}_{-0.28}$ in p_{T} range of 200 – 1000 GeV of the large-R jets. In addition, the preliminary $X \to b\bar{b}$ tagging efficiency has been studied in resonant $HH \to b\bar{b}b\bar{b}$ signal samples. The $X \to b\bar{b}$ tagging efficiency is significantly better than that of the current tagger DL1r using graviton decaying into a pair of Higgs bosons at various masses ranging between 1 TeV and 6 TeV.

References

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

[2] ATLAS Collaboration, *Identification of Boosted Higgs Bosons Decaying Into $b\bar{b}$ With Neural Networks and Variable Radius Subjets in ATLAS*, http://cds.cern.ch/record/2724739

[3] ATLAS Collaboration, *Efficiency corrections for a tagger for boosted $H \to b\bar{b}$ decays in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, https://cds.cern.ch/record/2777811

[4] ATLAS Collaboration, *Performance of $X \to b\bar{b}$ tagger in $G^*_{KK} \to HH \to b\bar{b}b\bar{b}$ signal*, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/FlavourTaggingPublicResultsCollisionData

[5] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s}$ = 13 TeV*, Eur. Phys. J. C **79** (2019) 970

[6] ATLAS Collaboration, *Search for resonant pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, Phys. Rev. D **105** (2022) 092002

[7] J. Alwall et. al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) 079