Real-time flavour tagging selection in ATLAS

Claudia Bertella on behalf of the ATLAS Collaboration
Johannes Gutenberg Universität, Staudingerweg 7 55128 Mainz, Germany
E-mail: bertella@cern.ch

Abstract. In high-energy physics experiments, online selection is crucial to the identification of the few interesting collisions from the large data volume processed. In the overall ATLAS trigger strategy, b-jet triggers are designed to identify heavy-flavor content in real-time and, in particular, provide the only option to efficiently record events with fully hadronic final states containing b-jets. In doing so, two different, but related, challenges are faced. The physics goal is to optimise as far as possible the rejection for light jets from QCD processes, while retaining a high efficiency on selecting jets from beauty, while maintaining affordable trigger rates without raising jet energy thresholds. This poses a challenging computing task, as charged tracks and their corresponding vertices must be reconstructed and analysed for each jet above the desired threshold, regardless of the increasingly harsh pile-up conditions. The performance of b-jet triggers during the LHC Run I data-taking campaigns is presented, together with an overview of the new online b-tagging strategy and algorithms, designed to face the above mentioned challenges, which will be adopted during Run II.

1. Introduction
The trigger system employed by the ATLAS experiment [1] at the LHC [2] has the ability to perform real-time track reconstruction and identify jets originating from the hadronization of b-quarks (b-jets). The selection of events using b-jet triggers makes it possible to collect signal events containing b-jets, particularly those characterised by fully hadronic final states. This class of trigger provides a way to reduce the overwhelming multi-jet backgrounds while preserving signal processes. Moreover, the trigger rate can be kept at a sustainable level without increasing the jet $p_T$ thresholds. Some interesting physics signatures which could benefit from the b-jet trigger are $t\bar{t}$ signatures with fully hadronic final states [3], VBF and associated production (with a vector boson W/Z or with top quarks) of a Standard Model (SM) Higgs boson with $H \rightarrow b\bar{b}$, supersymmetric processes (e.g. associated production of a Higgs boson with a b-quark, $bH \rightarrow bb\bar{b}$) and also Higgs boson pair production $X \rightarrow HH \rightarrow b\bar{b}b\bar{b}$.

2. How the b-tagging works
The identification of b-jets (b-tagging) is performed taking into account the distinguishing features of a b-hadron. The b-hadrons have a relative long lifetime of about 1.5 ps and as a consequence they travel several millimetres in the detector volume before decaying. This leads to the presence of a secondary vertex (SV) displaced from the primary interaction point, with a high transverse impact-parameter $d_0$, as illustrated in Figure 1. Typically, b-jets are characterised by high-multiplicity decay chains, with a large fraction of the jet energy being carried away by the b-hadron’s decay products. The b-tagging algorithms mainly use the impact parameter and
the secondary or tertiary vertex information to discriminate b-jets from gluon or light flavour jets.

3. ATLAS trigger system upgrade

The ATLAS trigger system during Run I consisted of three levels. The first level (L1) is hardware-based, using custom-built electronics. It uses only calorimeter and muon information with coarse granularity to select Regions of Interest (RoI). The second level (L2) can access the data from all sub-detectors at full granularity but only in RoIs identified by L1; it has access to the information provided by the Inner Detector (ID), allowing track reconstruction. The Event Filter (EF) is guided by the L2 results and has the ability to access all sub-detector data at full granularity. In order to deal with the increased centre-of-mass energy, increased luminosity and number of interactions per bunch crossing, major upgrades of the trigger system were carried out during the long shutdown (2013-2015). Several of them have a relevant impact on the flavour-tagging triggers:

- Upgraded hardware components have been included in the L1 Calorimeter trigger to provide a new topological trigger processor (L1Topo). They are able to perform event selection based on topological variables, such as the angular correlation and invariant mass between the L1 trigger objects [4].
- L2 and EF have been merged into a unique farm, the High Level Trigger (HLT), reducing the latency since no packing and transporting information between the two levels is needed anymore.
- Offline heavy flavour tagging algorithms have been implemented directly at the HLT, such as JetFitter [5] which reconstructs the b- to c-hadron decay chain under the hypothesis that their vertices are approximately aligned on a single flight path. Multivariate algorithms, e.g. MV1\(^1\) [6] and its successors that will be used for Run II, have been also implemented at the trigger level. The deployment of offline algorithms online will lead to an improved trigger performance and increase the correlation between online and offline and simplify the calibration procedure.
- The installation of the Fast Tracker (FTK) [7], hardware-based track finding after L1, will start in autumn 2015, with full pseudo-rapidity (|\(\eta\)|) coverage expected by the end of 2016. It receives all data from the pixel and silicon strip detectors, finds and reconstructs charged track candidates using pattern matching to pre-processed data. As a consequence, the tracking information will be available directly after L1, and the HLT can then improve the

\(^1\) The MV1 algorithm employs an artificial neural network based on IP3D, SV1 and JetFitter [5]. It is trained with b-jets as signal and light-flavour jets as background, and computes a tag weight for each jet.
Figure 2: The transverse impact parameter (a) and its significance (b) are shown for tracks associated to light-flavor (black) and heavy-flavor (red) jets in the barrel [8]. The clusters\(^3\) associated to the FTK tracks are fitted with the global $\chi^2$ algorithm. A $\chi^2/ndof < 2$ cut is applied to the FTK tracks after refit to increase track purity. The narrower $d_0$ significance core for FTK refitted tracks is a result of overestimated impact parameter uncertainties for the refitted FTK tracks.

track quality by refitting the pre-existing pattern. The $d_0$ and $d_0^2$ significance distributions for offline and FTK tracks are shown in Figure 2(a) and 2(b), respectively.

4. Real-time b-tagging algorithm at 8 TeV

A crucial component for the b-tagging algorithms is the correct reconstruction of the primary vertex. Primary-vertex reconstruction in ATLAS is divided into two steps. A “finding” step in which tracks are associated with a vertex candidate, and a “fitting” step in which a vertex position is assigned with an associated covariance matrix and estimated fit quality. A full discussion of the vertexing strategy can be found in [9]. The resolution for the primary vertex $z$ position, estimated as a function of the number of online tracks at L2 and EF, is shown in Figure 3.

During 2012, a combination of two likelihood-based algorithms was used for data taking. They exploit the impact parameter significance distribution (IP3D) and the secondary vertex proprieties (SV1).

4.1. IP3D tagger

The IP3D algorithm uses a likelihood ratio technique in which input variables are compared to pre-defined distributions for both the b- and light jet hypotheses, obtained from Monte Carlo simulation. The discriminating variables used are the 2D distributions of the transverse $z_0/\sigma(z_0)$ and the longitudinal $d_0/\sigma(d_0)$ impact parameter significance of tracks within jets.

4.2. SV1 tagger

The SV1 algorithm exploits secondary vertex properties: the invariant mass of all tracks associated to the vertex, the ratio of the sum of the energies of the tracks in the vertex to

\(^2\) The transverse impact parameter significance is defined as the transverse impact parameter divided by its associated uncertainty. The impact parameter is signed such that track displacements in the direction of the jet have positive values, while tracks with displacements opposite of the jet direction are negative.

\(^3\) Clusters consist of pixels connected side-by-side or diagonally. Clusters are truncated at 4 pixels in the $\phi$ direction and 5 pixels in the $z$ or $r$ direction for the barrel and endcap, respectively [7].
the sum of the energies of all tracks in the jet, and the number of two-track vertices.

4.3. Combined tagger
The jet combined tagger weight is calculated from prescaled EF (L2) tracks in EF (L2) jets with $p_T > 55$ GeV and $|\eta| < 2.5$ used for the combination of the impact parameter significance and the secondary vertex likelihood-based taggers. Figures 4(a) and 4(b) show the distribution of the jet weight distribution for the combined tagger at L2 and EF, respectively [10].

5. Calibration
In order to analyse data collected with b-jet triggers it is necessary to measure the trigger selection efficiencies, estimating the performance differences between data and simulation. The calibration results are then expressed in the form of scale factors which are parameterised in terms of $p_T$ and $|\eta|$. The scale factors are applied to simulation to match the tagging rate observed in data. The calibration addresses mainly:
• b-tagging efficiency: i.e. the efficiency with which a jet originating from a b-quark is tagged by a b-tagging algorithm. This is computed using the $p_T^\text{rel}$ method [11] which exploits the different properties of muons embedded in b-jets and light-jets (muon-jets). The muon-jet sample offers the advantage of being enriched in heavy-flavour jets since 20% of b-hadrons decay into muons. Figure 5 shows the b-tagging efficiency in data and simulation for the MV1 tagger at 70% efficiency.

![Figure 5: The b-tag efficiency in data and simulation (for the MV1 tagging algorithm at 70% efficiency) obtained with the $p_T^\text{rel}$ method.](image)

• c-tagging efficiency: this measurement makes use of the $D^\ast$ method [12] which exploits exclusive charm meson decays within a jet such as $D^{\ast+} \rightarrow D^0 (\rightarrow K\pi^+)\pi^+$ to provide a sample enriched in charm content. The beauty to charm jet fraction in the simulation is constrained to the value obtained by a pseudo-proper time fit on data for the $D^0$ mesons arising from the $D^{\ast+}$ decays. Figure 6 shows the EF trigger b-tagging weight distribution on a background-subtracted jets data sample containing $D^+$ mesons.

• mistag rate: i.e. the probabilities of mistakenly b-tagging a jet originating from a c-quark or a light-flavour parton. Light-flavour jets are mistakenly tagged as b-jets mainly because of the finite resolution of the ID and the presence of tracks stemming from displaced vertices from long lived particles or material interactions. It is estimated with the negative tag method [13] which measures the negative tag rate by inverting the selections of the discriminant parameters (the lifetime sign of the impact parameter of the tracks and the decay length significance of SV). Figure 7 shows the mistag rate for the MV1 tagger at 70% efficiency.

6. Summary and Outlook
Real-time flavour tagging algorithms were used in the ATLAS trigger during 2011 and 2012 data taking and used for physics analyses. In Run II, taking data with the possibility to efficiently trigger on b-jets will be crucial for several analyses. The improvements performed during the long shutdown period allow b-jet triggers to select signal processes with higher efficiency without compromising on fake rates. For Run II, more sophisticated b-jet triggers
Figure 6: Comparison between the EF Trigger b-tagging weight distribution on a background-subtracted jets, containing $D^{*+}$ mesons, data sample with the corresponding simulated PYTHIA sample. The statistical uncertainty of the simulation is below a few percent and not shown in this figure. The data sample is collected in 2012 using single jet triggers, and the transverse momentum of the selected jets is required to be above 20 GeV [10].

Figure 7: The mistag rate in data and simulation (for the MV1 tagging algorithm at 70% efficiency) obtained with the negative tag method, for jets with $1.2 < |\eta| < 2.5$.

have been implemented due to the merged HLT and its ability to use complex offline tagging algorithm directly at HLT level, such as multivariate combinations of properties sensitive to the jet flavours. Using common algorithms online and offline will improve the performance and simplify the calibration procedure. The FTK will be incorporated into the ATLAS trigger system, providing the possibility to find and reconstruct tracks in the ID for events that fire the L1 trigger. The FTK, free from the CPU constraints of HLT tracking, will improve the event selection of the ATLAS HLT trigger, particularly b-jet triggers. In Run I, b-jet tagging in the HLT began with a calorimeter-based jet pre-selection that tightened the jet $p_T$ thresholds. With the FTK, track finding can be run with looser HLT jet thresholds without putting additional load on the HLT processors. This opens up the possibility for the ATLAS trigger to be sensitive to physics signatures with lower b-quark transverse momentum. Figure 8 shows the output rates of the various trigger items as a function of the event-level $t\bar{t}H (H \rightarrow b\bar{b})$ efficiency by comparing b-jet items with and without the FTK.
Figure 8: The output rates of the various trigger items as a function of the event-level \(t\bar{t}H(H \rightarrow b\bar{b})\) efficiency. All operating points assume the re-fitted FTK performance. The black points show b-jet items that will fit into the HLT constraints without the FTK, red points show options with significantly larger input rates possible with the FTK. The efficiencies are quoted with respect to the inclusive signal and include the L1 efficiency.

References

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

[2] L. Evans and P. Bryant, *LHC Machine*, JINST 3 (2008) S08001.

[3] ATLAS Collaboration, *Measurement of the \(tt\) production cross section in the tau+jets channel using the ATLAS detector*, Eur. Phys. J. C73 (2013), no. 3 2328, [arXiv:1211.7205].

[4] E. Simioni, *Upgrade of the ATLAS Level-1 Trigger with event topology information*, Proceedings of the CHEP2015 conference, J. Phys.: Conf. Ser. (2015).

[5] ATLAS Collaboration, *Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data*, Tech. Rep. ATLAS-CONF-2011-102, CERN, Geneva, Jul, 2011.

[6] ATLAS Collaboration, *Calibration of the performance of b-tagging for c and light-flavour jets in the 2012 ATLAS data*, Tech. Rep. ATLAS-CONF-2014-046, CERN, Geneva, Jul, 2014.

[7] M. Shochet, L. Tompkins, V. Cavaliere, P. Giannetti, A. Annovi, and G. Volpi, *Fast TracKer (FTK) Technical Design Report*, Tech. Rep. CERN-LHCC-2013-007. ATLAS-TDR-021, CERN, Geneva, Jun, 2013. ATLAS Fast Tracker Technical Design Report.

[8] ATLAS Collaboration. http://twiki.cern.ch/twiki/bin/view/AtlasPublic/FTKPublicResults.

[9] ATLAS Collaboration, *Characterization of Interaction-Point Beam Parameters Using the pp Event-Vertex Distribution Reconstructed in the ATLAS Detector at the LHC*, Tech. Rep. ATLAS-CONF-2010-027, CERN, Geneva, May, 2010.

[10] ATLAS Collaboration. http://twiki.cern.ch/twiki/bin/view/AtlasPublic/BJetTriggerPublicResults.

[11] ATLAS Collaboration, *Measurement of the b-tag Efficiency in a Sample of Jets Containing Muons with 5 fb\(^{-1}\) of Data from the ATLAS Detector*, Tech. Rep. ATLAS-CONF-2012-043, CERN, Geneva, Mar, 2012.

[12] ATLAS Collaboration, *b-jet tagging calibration on c-jets containing \(D^{**}\) mesons*, Tech. Rep. ATLAS-CONF-2012-039, CERN, Geneva, Mar, 2012.

[13] ATLAS Collaboration, *Measurement of the Mistag Rate with 5 fb\(^{-1}\) of Data Collected by the ATLAS Detector*, Tech. Rep. ATLAS-CONF-2012-040, CERN, Geneva, Mar, 2012.