PROCEEDINGS OF SCIENCE

Expected b-tagging performance for ATLAS in LHC Run 2

Thomas CALVET*, **on behalf of the ATLAS Collaboration**
Centre de Physique des Particules de Marseille
CNRS/IN2P3 et Aix-Marseille Université
163, Avenue de Luminy - Case 902
13288 Marseille Cedex 09, France
E-mail: calvet@cppm.in2p3.fr

The ability to identify jets containing b-flavored hadrons is important for the physics program of the ATLAS experiment at the LHC. Both Standard Model and new physics measurement need efficient techniques to identify jets containing b-flavored hadrons. Thanks to the addition of a new pixel layer in the ATLAS detector, and the improvements of b-tagging algorithms, a significant improvement of b-jet identification is expected for Run 2. A review of b-tagging algorithms is presented here, together with a discussion of their expected performance, as well as a comparison between Run 1 performance and the expected performance of Run 2.

The European Physical Society Conference on High Energy Physics
22–29 July 2015
Vienna, Austria

*Speaker.

1. Introduction

The identification of jets containing b-hadrons (b-jets) is an important task for a general-purpose experiment at LHC like ATLAS. It is in fact a key ingredient for many analyses, ranging from precise top-quark measurements to new physics searches, along with Higgs studies, which can have b-jets in the final state. In the current Run 2 of the LHC and its increased center of mass energy $\sqrt{s} = 13$ TeV, the reach of these analysis is increased. Improved b-tagging techniques will thus be essential to fully exploit the Run 2 data.

The discrimination of b and c-jets or light-jets (usdg-jets) is mainly possible thanks to the long lifetime of b-flavored hadrons, of the order of 1.5 ps. A b-hadron of transverse momentum $p_{\mathrm{T}} = 50$ GeV, and a mass around 5 GeV, will have a flight path length $\langle l \rangle = \beta\gamma c\tau$ around 4 mm before decaying. This property is used either by reconstructing explicitly the displaced secondary vertex, or by measuring the impact parameters of charged particle tracks.

In Run 2, the ATLAS detector is equipped with a fourth pixel layer, the Insertable B-Layer (IBL), closer to the beam pipe (~ 3.3 cm instead of ~ 5 cm for the Run 1 innermost pixel layer), and with smaller pixels (50 µm $\times$ 250 µm instead of 50 µm $\times$ 400 µm for the Run 1). Thus, we expect a large improvement of the impact parameter resolution of tracks, and on vertexing. Thanks to the IBL and the improvements to the b-tagging algorithms [1], the b-tagging performance is expected to be significantly enhanced in LHC Run 2.

This paper presents the b-tagging algorithms, and the improvements expected in Run 2. The performance is estimated using $t\bar{t}$ Monte Carlo simulations corresponding to 13 TeV proton-proton collisions. To be considered, jets must fulfill $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.5$ [1].

2. Basic b-tagging Algorithms

ATLAS relies on three basic b-tagging algorithms, classified into two categories, which provide complementary information: impact parameter based algorithms (Section 2.1); and secondary vertex based algorithms (Section 2.2), using either an inclusive reconstruction of the vertex (Section 2.2.1), or a multi-vertex reconstruction (Section 2.2.2).

Each of the basic b-tagging algorithms uses tracks to identify one signature of b-hadrons inside a jet. Tracks are first associated to jets with a ΔR(track,jet)= $\sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ requirement that depends on the jet p_{T}. Then each algorithm has its own set of selection on track p_{T}, impact parameters and the number of hits in the inner detector. Typically, secondary vertex based algorithms use loose requirements and take advantage of the reconstruction of a secondary vertex to enhance the purity. On the other hand, impact parameter based algorithms rely on a tight track selection to remove undesired tracks.

2.1 Impact Parameter Based Algorithms: IP2D and IP3D

The impact parameter based algorithms identify the (in)compatibility of each track candidate

[1] ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the center of the detector, and the z-axis coinciding with the axis of the beam pipe. The x-axis points from the IP to the center of the LHC ring, and the y-axis points upward. Cylindrical coordinates (r,ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the beam pipe. The pseudo-rapidity is defined in terms of the polar angle θ as $\eta = -\log\left(\tan\frac{\theta}{2}\right)$.

with the primary vertex. The impact parameter is separated in two components : d_0 in the transverse plane and z_0 along the beam axis. They are defined positive if the point of closest approach to the primary vertex is in the same direction as the jet momentum with respect to the primary vertex. To give more importance to well-measured tracks, b-tagging uses the signed impact parameter significances, $s(d_0) = d_0/\sigma_{d_0}$ and $s(z_0) = z_0/\sigma_{z_0}$, where σ_{d_0} and σ_{z_0} are the errors on d_0 and z_0. The $s(d_0)$ distribution is shown in Fig 1.

The σ_{d_0} and σ_{z_0} distributions are used to define Probability Density Functions (PDFs) of single tracks to fulfill the b- (P_b) or light-jet (P_u) hypotheses. The PDFs are then combined in a Log Likelihood Ratio (LLR) discriminant for each jet: $\log(P_b/P_u) = \sum_{tracks} \log(P_b/P_u)$. The IP2D LLR is shown in Fig 2. The IP3D algorithm combines the transverse and longitudinal impact parameter information into two dimensional PDFs, taking into account their correlations. However, this longitudinal component makes the IP3D algorithm less robust against pile-up compared to the IP2D algorithm.

To increase the performance, tracks are divided into categories depending on some of their properties :

- split and shared hits : hits identified as created by several particles.
- missing hit: a hit is said to be missing if the extrapolation of a track crosses active modules of the inner detector, but no hit is found.

Each category has its own set of reference histograms for the d_0 and z_0 distributions, which are then used to compute the PDFs for all tracks. In Run 2, these categories are refined to improve the separation between b, c and light-jets.

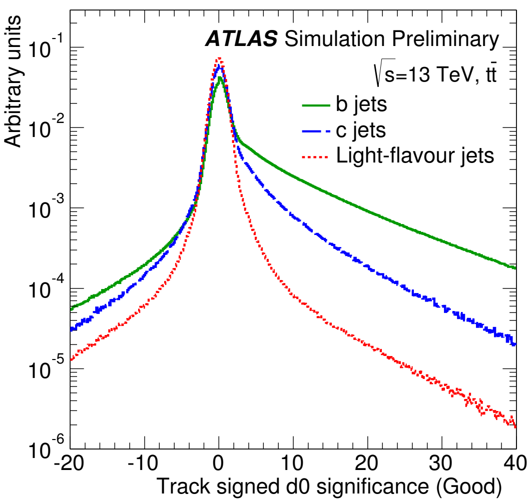

Figure 1: Signed transverse impact parameter significance of tracks in b, c, and light jets. Taken from [1]

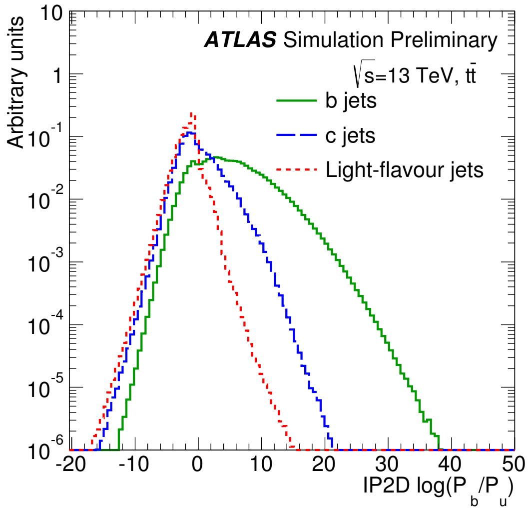

Figure 2: Log likelihood ratio for the IP2D b-tagging algorithm for b, c, and light jets. Taken from [1]

2.2 Secondary Vertex Based Algorithms

Two approaches are used in b-tagging to exploit the b-hadron secondary vertex: the explicit reconstruction of a single secondary vertex and the reconstruction of the full b-hadron decay chain.

2.2.1 Inclusive Secondary Vertex Reconstruction

This algorithm reconstructs explicitly one secondary vertex per jet, using a selected set of tracks. All track candidates are first used to reconstruct two-track vertices. If a secondary vertex is compatible with long lived particle decay vertex (e.g. K_S or Λ), photon conversions or hadronic interaction with the material, its associated tracks are removed. The invariant mass of the track four-momenta is used to identify vertices that are likely to come from long lived particle decay vertices and photon conversions, while a superposition of the vertex position with a simplified map of the innermost pixel detector layers and the beam pipe is used to reject vertices originating from material interaction.

The inclusive secondary vertex reconstructed from the remaining tracks is then used to build discriminating variables for the b and light jet hypotheses. For example, given the high mass of b-hadrons compared to light-hadrons, one expect the mass of the secondary vertex to be larger in b-jets than in light-jets.

The standalone version of this algorithm uses a LLR formalism, similar to the one explained for the IP3D algorithm, based on some properties of the secondary vertex.

2.2.2 Multi-vertex Reconstruction

Knowing that b-hadrons decay to c-hadrons, typically leading to one or more tertiary vertices, the JetFitter algorithm aims at reconstructing the full b-hadron decay chain : PV$\rightarrow$B$\rightarrow$D. However, these vertices are hard to separate efficiently. Thus, this algorithm assumes that the c-hadron flies is the same direction as the b-hadron. A Kalman filter is used to find a line on which the primary vertex and the two secondary vertices lie. Track candidates are used to build single-track vertices along the first approximation (the jet direction) of the flight axis. Then vertices are merged two by two in decreasing order of probability to originate from the same vertex using a clustering algorithm. The PV$\rightarrow$B$\rightarrow$D system is finally described when a stable configuration is reached.

This strategy allows the JetFitter algorithm to resolve the b and c-hadron vertices, even when only one track is associated to each of them. The efficiency to reconstruct a decay chain is shown in Fig 3. One can see that the efficiency to reconstruct decay chain with at least one two-track-vertex is lower. However, requiring one two-track-vertex improves the purity.

The reconstructed decay chain is used to build discriminating variables for b and light jets, like shown in Figure 4. Some of the variables are used as inputs to an artificial neural network based discriminator. The three output nodes of this neural network are used as probabilities for the b, c and light jet hypothesis.

3. MVx b-tagging Algorithm

In order to improve the discrimination between b-jets and c or light jets, the properties of the single secondary vertex and of the multi-vertex decay chain, as well as the LLR from IP2D and

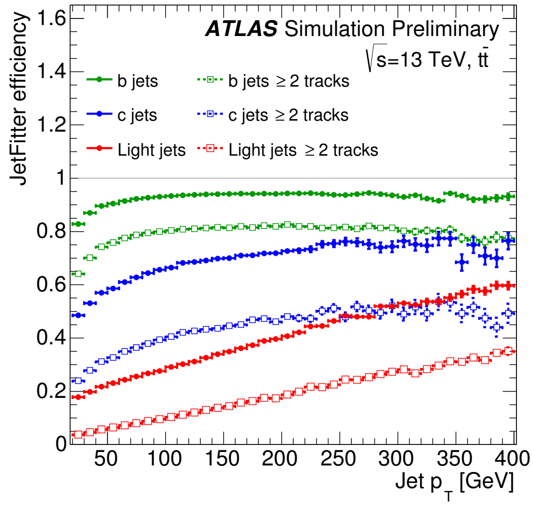

Figure 3: JetFitter vertex reconstruction efficiency as a function of jet p_{T} for the different jet flavors. Taken from [1]

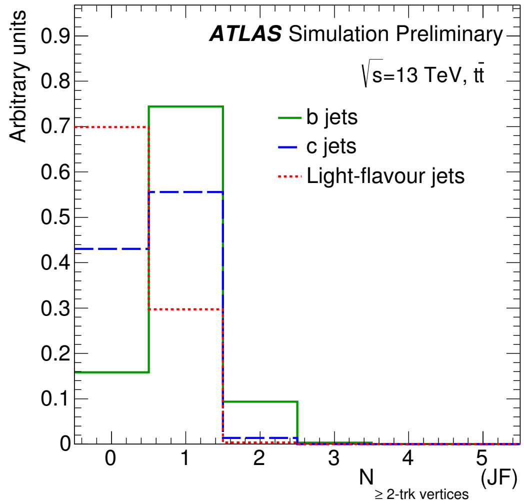

Figure 4: Number of vertices with at least two tracks reconstructed by JetFitter for the different jet flavors. Taken from [1]

IP3D, are combined in a boosted decision tree (BDT) algorithm. Its output is defined as the MV2 algorithm. Using $t\bar{t}$ events, several BDTs were trained with different mixture of c and light jets in the background. A good compromise between light-jet rejection (inverse of the efficiency), and c-jet rejection is obtained when the amount of c-jets in the background is equal to 20% of the amount of light-jets. The default b-tagging algorithm for Run 2, MV2c20, uses this last configuration.

In Run 1, b-tagging was providing two algorithms based on a neural network approach. The default Run 1 algorithm, MV1, was trained only against light-jet, while a second algorithm, MV1c, was trained against a mixture of c and light-jet to improve the c-jet rejection. Both of them were using with the single secondary vertex and impact parameters LLRs, as well as the neural network output of JetFitter. Beside the gain in both simplicity and performance, the new architecture allows to better exploit correlations between the input variables.

Figure 5 shows a comparison of the Run 1 algorithm MV1c and the default Run 2 algorithm MV2c20. The MV1c performance curves are made with a simulated 8 TeV sample using the Run 1 detector simulation and reconstruction software, while the MV2c20 performance curves are made with a 13 TeV sample using the Run 2 configuration. To allow for a fair comparison, the 13 TeV sample is re-weighted to match the jet p_{T}, jet η, and average number of pp interaction in the 8 TeV sample. For the same 70% b-tagging efficiency, the light-jet rejection is increased by a factor of about 4 from MV1c to MV2c20. Fixing the light-jet rejection at the value achieved for the 70% b-jet efficiency in Run 1, a relative gain of 10% in b-tagging efficiency is achieved. This increase represents a gain of $40-50\%$ in signal acceptance for an analysis with four b-quarks in the final state such as the production of a Higgs boson decaying in two b-quarks in association with two top quarks. In Figure 6, the b-tagging efficiency is fixed at 70% and the light-jet rejection is shown as a function of jet p_{T}. This allows to decouple the contribution of the IBL, that is expected to dominate the improvements at low and medium p_{T}, from updated b-tagging and tracking algorithms [2] that are the largest contributions to the improvement at high p_{T}. The IBL together with the improved algorithms increase the rejection by a factor of 4 at low p_{T}, while the new algorithms double the light-jet rejection at high p_{T}, in comparison to Run 1.

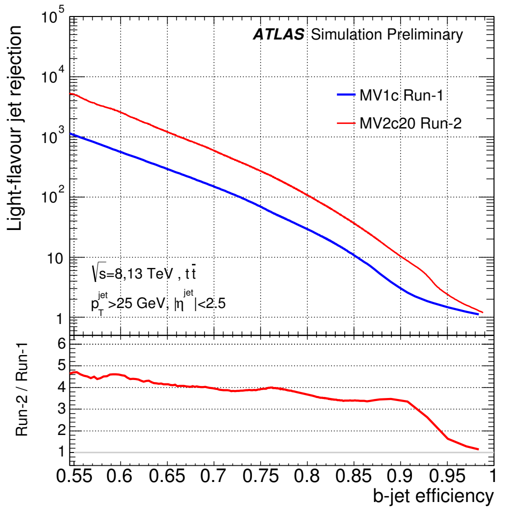

Figure 5: Light-jet rejection as a function of the b-jet efficiency obtained with the MV1c and MV2c20 algorithms. Taken from [1]

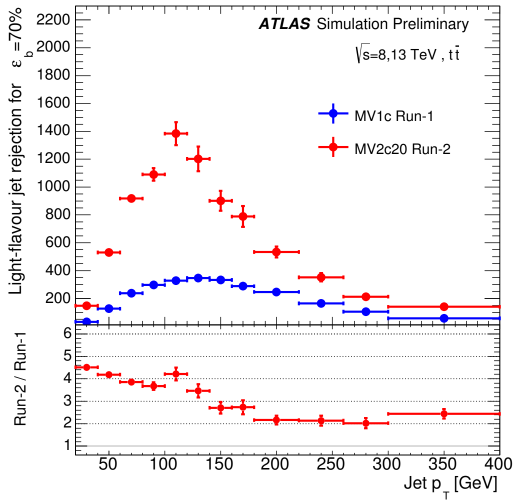

Figure 6: Light-jet rejection as a function of the jet p_T obtained with the MV1c and MV2c20 algorithms, requiring a 70% b-jet efficiency in each p_T bin. Taken from [1]

4. Conclusion

The b-jet identification procedure in ATLAS is described in the report, together with the expected improvements for LHC Run 2. The new MV2c20 algorithm achieves a light-jet rejection of about 140 and a c-jet rejection of 4.5, for a b-jet efficiency of 77% on a $t\bar{t}$ sample. This represents a relative increase of 10% in b-jet efficiency compared to Run 1 performance with the same c and light jet rejection. Any analysis with b-jets in the final state will greatly benefit from this improvement. This enhancement relies on several improvements to both basic and multivariate b-tagging algorithms, as well as the insertion of the new Insertable B-Layer.

ATLAS is now collecting LHC data at $\sqrt{s} = 13$ TeV. To ensure the achievement of these expected improvements, it is vital that the detector, tracking, and b-tagging algorithms are fully commissioned with the new data. This effort is ongoing and the first results can be found in [3].

References

[1] The ATLAS collaboration, *Expected performance of the ATLAS b-tagging algorithms in Run-2*, ATL-PHYS-PUB-2015-022

[2] The ATLAS collaboration, *The Optimization of ATLAS Track Reconstruction in Dense Environments*, ATL-PHYS-PUB-2015-006

[3] The ATLAS collaboration, *Track Reconstruction Performance of the ATLAS Inner Detector at* $\sqrt{s} = 13 TeV$, ATL-PHYS-PUB-2015-018