ATLAS b-tagging performance during LHC Run-2 with the new Insertable B-layer

Yulia Rodina¹,², on behalf of the ATLAS Collaboration

¹Centre de Physique des Particules de Marseille, Aix-Marseille Université & CNRS/IN2P3
²Institut de Física d’Altes Energies, The Barcelona Institute of Science and Technology, Universitat Autònoma de Barcelona

rodina@cppm.in2p3.fr

Abstract. Expected b-tagging performance in ATLAS during LHC Run-2 is presented together with results of commissioning studies with early Run-2 data. The performance of b-tagging is expected to be significantly improved in Run-2 thanks to the addition of a new pixel detector layer and to updates in the tracking and b-tagging algorithms. First commissioning studies show a promising agreement of Monte Carlo simulation with the new 2015 data recorded at 13 TeV.

INTRODUCTION

Identifying jets containing b-hadrons, a capability known as b-tagging, is useful in various analysis domains with b-jets in final state performed by the ATLAS experiment, like Standard Model measurements (top quark physics and Higgs physics) and beyond the Standard Model searches.

For b-tagging various algorithms are used, which rely on special properties of b-hadrons, such as their high mass (≈ 5 GeV) and relatively long lifetime (≈ 1.5 ps). For the Run 2 of LHC with higher center of mass energy (13 TeV) of the proton-proton collisions, the b-tagging algorithms were revisited. The b-tagging performance during Run-2 is expected to be improved thanks to the insertion of a new innermost layer of pixel detector and algorithmic enhancements in both tracking and b-tagging algorithms.

THE INSERTABLE B-LAYER (IBL)

The major ATLAS inner detector upgrade for Run-2 is the addition of the Insertable B-Layer (IBL), a fourth pixel layer in order to have better track and vertex reconstruction performance at the higher luminosities [1].

The pixel detector used during Run-1 was designed for a peak luminosity of \( \mathcal{L} \approx 10^{34} \text{cm}^{-2}\text{s}^{-1} \), while during Run-2 peak luminosity is expected to be \( \mathcal{L} \approx 1.7 \times 10^{34} \text{cm}^{-2}\text{s}^{-1} \). Because of that the addition of the new pixel layer was necessary to maintain tracking and b-tagging performance despite the increased pile-up.

The IBL was inserted inside the existing pixel detector at a radius of \( \approx 3.3 \text{ cm} \) from the beamline, while the next to innermost layer (which was innermost pixel layer in Run-1) is located at a radius of \( \approx 5 \text{ cm} \). Another advantage of the IBL is its higher granularity, with pixels of size 50 \( \mu\text{m} \times 250 \mu\text{m} \) instead of 50 \( \mu\text{m} \times 400 \mu\text{m} \) for the former innermost pixel layer. Finally, the average number of pixel measurements on a single track became 4 instead of 3. This improves the tracking robustness with respect to pile-up and possible pixel module failures.

The Run-2 to Run-1 data comparison showed already that IBL significantly improves track impact parameter resolution: by up to a factor of 2 for both transverse and longitudinal components for low-pr tracks. For a typical 2 GeV track, the transverse impact parameter resolution is now \( \approx 30 \mu\text{m} \) and the longitudinal one \( \approx 80 \mu\text{m} \) [2].
RUN-2 B-TAGGING ALGORITHMS

In the ATLAS Run-2 b-tagging scheme there are three basic algorithms:

- Impact parameter-based (IP2D, IP3D), making use of the fact that tracks from the b-hadron decay are not pointing to the primary vertex.
- Secondary vertex finding (SV), reconstructing an inclusive displaced secondary vertex within the jet.
- Decay chain multi-vertex fit (JetFitter), attempting to reconstruct the full b-hadron decay chain.

Several observables from these algorithms are combined with a multivariate algorithm (MV2), that provides the final discriminant between the different jet flavours.

**Impact parameter-based (IP2D, IP3D)**

IP2D and IP3D algorithms use the signed impact parameter significance of tracks associated to a jet. The impact parameter is the distance of closest approach of a track to the primary vertex (PV). Its sign is defined positive (negative) if the point of closest approach of the track to the primary vertex is in front (behind) the primary vertex with respect to the jet direction.

IP2D algorithm is using as input only the transverse impact parameter, while IP3D uses both transverse and longitudinal components and their correlation. Probability density functions (PDFs) of the track’s impact parameter are built from simulation for the b- and light-flavour jet hypotheses and are then combined using a log-likelihood ratio method to define a tagging weight for a jet:

\[
  w_{\text{track}} = \frac{p_b}{p_{\text{light}}}, \quad w_{\text{jet}} = \sum_{\text{tracks}} \log w_{\text{track}}.
\]  

In the IP2D and IP3D algorithms different PDF sets are used for different track categories, depending on the quality of the tracks, which is defined using information on hits in the different silicon layers of the inner detector.

Run-2 track categorisation is different from the one used in Run-1, making use of new tracking variables related to the presence of the IBL. Table 1 shows 14 exclusive track categories used at Run-2 and indicates percentage of tracks from b-, c- and light flavour jets in each category for simulated \(t\bar{t}\) events.

The number of hits in the innermost (L0) and the next to innermost (L1) layers of pixel detector, as well as information on whether the hit in a layer is expected or not (based on the detector coverage and dead module maps) are important variables for defining the track quality. Also the information on the presence of shared hits (clusters which are shared among more than one track) and split hits (clusters which have been identified as coming from different particles and have therefore been split into sub-clusters) is used when dividing tracks into categories.

**TABLE 1.** Run-2 IP2D and IP3D track categories and fraction of tracks from b-, c- and light flavour jets in each category for the \(t\bar{t}\) sample [3].

<table>
<thead>
<tr>
<th>#</th>
<th>Category</th>
<th>light jets</th>
<th>b-jets</th>
<th>c-jets</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No hits in first two layers; expected hit in both L0 and L1</td>
<td>1.6%</td>
<td>1.5%</td>
<td>1.6%</td>
</tr>
<tr>
<td>1</td>
<td>No hits in first two layers; exp. hit in L0 and no exp. hit in L1</td>
<td>0.1%</td>
<td>0.1%</td>
<td>0.1%</td>
</tr>
<tr>
<td>2</td>
<td>No hits in first two layers; no exp. hit in L0 and exp. hit in L1</td>
<td>0.03%</td>
<td>0.03%</td>
<td>0.03%</td>
</tr>
<tr>
<td>3</td>
<td>No hits in first two layers; no exp. hit in L0 and L1</td>
<td>0.02%</td>
<td>0.03%</td>
<td>0.03%</td>
</tr>
<tr>
<td>4</td>
<td>No hit in L0; exp. hit in L0</td>
<td>2.1%</td>
<td>2.4%</td>
<td>2.3%</td>
</tr>
<tr>
<td>5</td>
<td>No hit in L0; no exp. hit in L0</td>
<td>0.9%</td>
<td>0.9%</td>
<td>0.9%</td>
</tr>
<tr>
<td>6</td>
<td>No hit in L1; exp. hit in L1</td>
<td>0.5%</td>
<td>0.5%</td>
<td>0.5%</td>
</tr>
<tr>
<td>7</td>
<td>No hit in L1; no exp. hit in L1</td>
<td>2.3%</td>
<td>2.4%</td>
<td>2.4%</td>
</tr>
<tr>
<td>8</td>
<td>Shared hit in both L0 and L1</td>
<td>0.04%</td>
<td>0.01%</td>
<td>0.01%</td>
</tr>
<tr>
<td>9</td>
<td>Shared hits in other pixel layers</td>
<td>1.8%</td>
<td>2.1%</td>
<td>1.6%</td>
</tr>
<tr>
<td>10</td>
<td>Two or more shared SCT hits</td>
<td>2.2%</td>
<td>2.4%</td>
<td>2.2%</td>
</tr>
<tr>
<td>11</td>
<td>Split hits in both L0 and L1</td>
<td>0.8%</td>
<td>1.2%</td>
<td>1.1%</td>
</tr>
<tr>
<td>12</td>
<td>Split hits in other pixel layers</td>
<td>1.1%</td>
<td>2.1%</td>
<td>1.6%</td>
</tr>
<tr>
<td>13</td>
<td>Good: a track not in any of the above categories</td>
<td>86.6%</td>
<td>84.3%</td>
<td>85.5%</td>
</tr>
</tbody>
</table>
Secondary Vertex Finding Algorithm (SV)

The secondary vertex algorithm attempts to reconstruct the inclusive vertex formed by the decay products of the b-hadron, including those from the subsequent c-hadron decay. Firstly it searches for all two-track pairs that form a good vertex, using tracks displaced from the primary vertex. Then the algorithm removes those tracks that are compatible with decays of long lived particles (K_s, Λ etc) or interaction with the detector material. After this selection the algorithm fits an inclusive secondary vertex. Several properties of this vertex are useful to tag b-jets, such as its mass, number of tracks, distance to primary vertex, energy fraction of tracks at vertex with respect to all tracks in the jet.

Multi-vertex fit (JetFitter)

Another algorithm, called JetFitter, attempts to reconstruct the full PV to b- to c-hadron decay chain. A Kalman filter is used to find a common line on which the primary vertex and all secondary vertices are placed, approximating the b-hadron path. This approach allows to separate b- and c-hadron vertices even if only one track is attached to each of them.

Multivariate algorithm (MV2)

Discriminant observables from the above algorithms are combined together into a boosted decision tree (BDT) based algorithm. The default algorithm for Run-2, MV2c20, is a BDT which is trained using b-jets as signal and a mixture of light-flavour jets and c-jets as background (the amount of c-jets in the background is equal to 20% of the amount of light-jets). The kinematic properties (p_T and η) of the jets are included in the training to take advantage of correlations with the other input variables.

MV2c20 is an upgrade of Run-1 main b-tagging algorithm MV1, which was combining the outputs of the various b-tagging algorithms using neural network approach. The MV2c20 algorithm provides better performance and easier retraining and software maintenance.

EXPECTED B-TAGGING PERFORMANCE ENHANCEMENT

A b-tagging performance improvement is expected to be achieved in Run-2 due to addition of the IBL and many algorithmic updates in track reconstruction [4] and b-tagging, both in the basic taggers and final multivariate algorithm.

Figure 1 show a comparison of the default Run-2 b-tagging algorithm MV2c20 and the equivalent Run-1 b-tagging algorithm MV1c: light jet rejection vs b-jet efficiency (a) and light jet rejection as a function of jet p_T given a fixed b-jet efficiency of 70% in each bin (b). Light jet rejection is the number of light jets over the number of light jets tagged as b-jets. Improvement at low and medium p_T is mostly due to the addition of the IBL, while the improvement at high p_T is due to algorithm improvements. At 70% efficiency the light-flavour jet rejection in Run-2 is improved inclusively by a factor of 4 compared to Run-1. This corresponds to a 10% gain in efficiency at a constant light-jet rejection.

RUN-2 DATA TO MONTE CARLO COMPARISON

To confirm the Monte Carlo performance, simulation needs to be compared with data and the early Run-2 data commissioning studies are a first step in this direction. A first study of the b-tagging modeling was performed using pp collision data collected by ATLAS at the centre-of-mass energy of 13 TeV with 50 ns bunch-spacing on a high purity b-jet sample of e + µ di-leptonic t¯t candidate events. Only jets with p_T > 20 GeV and |η| < 2.5 are considered [5].

Input observables from the basic b-tagging algorithms and the output of the multivariate algorithm MV2c20 have been studied. Figure 2 shows the log-likelihood ratio of the IP3D algorithm. Figure 3 represents properties of the secondary vertices reconstructed by the SV algorithm: the invariant mass of the tracks from the vertex (a) and the energy fraction, defined as the energy from the tracks in the displaced vertex relative to all tracks reconstructed within the jet (b). Figure 4 illustrates vertex decay chain fit properties provided by the JetFitter algorithm: the number of vertices with no less than two tracks (a) and the number of tracks at vertices with no less than two tracks (b). Finally,
FIGURE 1. Performance of default Run-2 b-tagging algorithm MV2c20 and the equivalent Run-1 b-tagging algorithm MV1c in simulated $t\bar{t}$ events: light jet rejection vs b-jet efficiency (a) and light jet rejection as a function of jet $p_T$ for a fixed b-jet efficiency of 70% in each bin (b) [3].

Figure 5 shows the output distribution of the MV2c20 algorithm. On all plots the data are shown by the points and the simulation by the filled area, divided into b (red), c (light green) and light-flavour (blue) components. The dark green shaded area represents the total systematic and statistical uncertainty on the simulation and the error on the points corresponds to the statistical uncertainty on the data.

CONCLUSIONS

The b-tagging performance in ATLAS is expected to be significantly improved in Run-2 thanks to the addition of a new pixel detector layer and to various updates in the tracking and b-tagging algorithms.

The Monte Carlo studies of the b-tagging performance showed significant enhancement in light-flavour jet rejection in Run-2 compared to Run-1: by a factor of 4 at 70% efficiency, which corresponds to an improvement of $\approx 10\%$ in efficiency at a constant light-jet rejection. Thanks to this improvement, the final signal acceptance for different type of analysis with b-jets in final state is increased (for example, for the search for the Higgs boson decaying to b-quarks and produced in association with top quarks signal acceptance could be increased by up to $\approx 46\%$).

First commissioning studies of basic b-tagging algorithms as well as the multivariate algorithm on Run-2 data showed promising agreement with simulation within the present statistical and systematic uncertainties.

REFERENCES

FIGURE 2. IP3D impact parameter-based algorithm output for jets selected from the $t\bar{t}$ dominated $e + \mu$ sample [5].

FIGURE 3. Two properties of the secondary vertices reconstructed by the SV algorithm: the invariant mass (a) and the energy fraction of tracks at vertex with respect to all tracks in the jet (b) [5].
FIGURE 4. Vertex decay chain fit properties reconstructed by the JetFitter algorithm: the number of vertices with at least two tracks (a) and the number of tracks at vertices with at least two tracks (b) [5].

FIGURE 5. Output distribution of the MV2c20 algorithm [5].