IL NUOVO CIMENTO **40 C** (2017) 12
DOI 10.1393/ncc/i2017-17012-y

Colloquia: IFAE 2016

b-tagging performance at 13 TeV for the ATLAS experiment

A. Lapertosa(1)(2)

(1) *INFN, Sezione di Genova - Genova, Italy*
(2) *Dipartimento di Fisica, Università degli Studi di Genova - Genova, Italy*

received 17 October 2016

Summary. — The correct identification of jets containing b hadrons (hence b-tagging) is of capital importance for the hadron collider experiments, such as ATLAS and CMS at the Large Hadron Collider. In particular, b-tagging is an important tool for many physics analyses: for top quark studies and Higgs boson searches, as well as the search for new physics fenomena beyond Standard Model. ATLAS developed its own algorithms for b-jet identification, exploiting the typical properties of the b quarks and the B hadrons emerging from the jet: the long lifetime, the high decay multiplicity and the high invariant mass above all. The status of the most recent b-tagging algorithms developed by the ATLAS experiment is presented, along with the methods used to measure the b-tagging efficiency on a charm jets sample at 13 TeV with 2015 data.

1. – Introduction

In the Standard Model frame, stand-alone quarks and gluons usually combine to form colourless states (mesons and barions): during the "hadronization" process a tight cone of hadrons (jet) emerges. It is impossible to detect stand-alone quarks before hadronization (exept for the top quark, which decays before hadronizing), but it is possible to identify the flavour of the quark from which the jet originated. Exploiting some typical properties of the jets, it is possible to separate jets coming from a beauty quark (b-jets), from a charm quark (c-jets) or from a gluon or an up, down or strange quark (light jets).

In particular, b-jets have a high invariant mass (as a consequence of the beauty quark high mass) and contain one or more B hadrons in their final state after hadronization. The long lifetime of some B hadrons gives to b-jets their special topology: the presence of a secondary vertex, quite separated from the primary vertex and reconstructable in most cases, and tracks with large impact parameters.

The light quarks (up, down and strange) with masses in the MeV scale or the massless gluons are much lighter than the b quark: the light jets are therefore quite different with respect to the b-jets in terms of invariant mass and decay length.

The c quarks are slightly lighter than the b quarks, so jets coming from c quarks (c-jets) have intermediate properties: since the hadrons containing c quarks have a lifetime similar to that of the B hadrons, they also produce a separated secondary vertex, but they are characterized by lower track multiplicity and invariant mass. For this reason c-jets are very hard to identify and they easily fake b-jets.

To be able to correctly identify b-jets is fundamental for Standard Model analysis, like Higgs (the most probably decay channel is $H \rightarrow bb$) and top quark physics (almost always decaying as $t \rightarrow Wb$), but also for beyond Standard Model physics as Super symmetry particles searches. With this purpose, the ATLAS experiment developed b-tagging algorithms in order to separate b-jets from c-jets and light jets: the most recent algorithms are introduced in sect. **2**. The efficiency to identify a b-jet is measured on data, with different methods based on selected samples of b-jets, c-jets and light jets: in sect. **3**, the methods used to measure the b-tagging efficiency on a sample of c-jets are described.

2. – ATLAS *b*-tagging algorithms

The identification of b-jets is performed with several algorithms, exploiting the long lifetime, high mass and decay multiplicity of b hadrons. The tracks originated from b-hadron decay have large impact parameters with respect to the tracks stemming from the primary vertex. The three basic ATLAS b-tagging algorithms are based on the secondary vertex reconstruction (SV1), the large impact parameters of their tracks (IP3D) and the topological structure of the decays of the b-hadron and the c-hadron inside the jet (JetFitter). To improve the performance of these algorithms, their output variables are combined through multivariate tecniques to obtain more discriminant variables.

During Run I, the variables were combined using a Neural Network with two hidden layers consisting of three and two nodes respectively, and an output layer with a single node which holds the final discriminant variable (MV1) [1]. To enhance the c-jet rejection, a new version of the MV1 Neural Network was developed, adding also a c-jet component as background hypotesis in the training sample. The resulting variable (MV1c) has a better c-jet rejection with a small decrease in light jet rejection.

For Run II a new multivariate b-tagging algorithm was developed: MV2c20 [2]. The output variables obtained from the basic algorithms are combined using a Boosted Decision Tree algorithm instead of a Neural Network. The performance for several background mixtures of c-jets and light jets in the training sample has been compared: the mixture adopted in MV2c20 (b-jets as signal and a mixture of 80% light jets and 20% c-jets as background) gave the best c-jet rejection performance.

Significant improvements are introduced in the b-tagging performance for the Run II, due to the addition of an extra pixel layer in the detector (the Insertable B-Layer [3]) and from several enhancements to the tracking and b-tagging algorithms. With respect to MV1c, the new MV2c20 variable is expected to have better light jet and c-jet rejections. The new pixel layer, closer to the collisions point, allows an improvement in the impact parameter resolution of tracks up to 10 GeV, resulting in a better b-tagging performance of jets in the low p_T region. A consistent improvement was also achieved for jets in the medium and high p_T regions, mainly due to the enhancements in tracking and b-tagging algorithms.

3. – Measurement of the *c*-jet tagging efficiency

The performance of the b-tagging algorithms is calibrated on data, measuring the efficiency to tag a jet originated from a b quark as a b-jet (b-jet tagging efficiency) or to mistakenly tag a jet originated from a charm quark (c-jet tagging efficiency) or from a light quark or a gluon (mistag rate). The samples used to perform each efficiency measurement must be characterized by a strong predominance of jets of that specific flavour: the first step is to select an almost pure sample of single flavoured jets, estimating the flavour composition over data.

During Run I, the c-jet tagging efficiency was measured by the ATLAS experiment with two different techniques: the "D^* method" [4, 5] and the "$W + c$ method" [6]. The reference measurements of the c-jet tagging efficiency were performed with the "D^* method" on a high statistics sample of c-jets containing D^* charged mesons. The main disadvantage of this method is the high uncertainty due to the generalization of the result to an inclusive sample of c-jets, since D^* mesons occur in only a limited subset of c-jets.

The c-jet tagging efficiency measurement was also performed with the "$W+c$ method", using a sample of c-jets produced in association with a W boson. The W boson is reconstructed via its decay into an electron and a neutrino, and the c-jet is identified via a soft muon stemming from a semileptonic c-hadron decay. This analysis had the benefit of applying to all semileptonically-decaying c-jets, which requires significantly less extrapolation to inclusive c-jets than the D^* analysis.

The Run II program is to measure the c-jet tagging efficiency of MV2c20 with both methods, with a particular attention to the "W+c method", since lower extrapolation uncertainties are expected. First preliminary results of the measurement performed with the "$W + c$ method" on 13 TeV 2015 data confirm the expected improvements in the c-jet rejection of the MV2c20 b-tagging algorithm with respect to MV1.

4. – Conclusion

Beauty quarks are present in the final states of many interesting physics processes (Higgs boson decay, top quark decay and Super Symmetry physics). The identification of jets originated from b quarks is of fundamental importance for many measurements performed at high energy physics experiments. The ATLAS experiment developed specific algorithms able to discriminate between b-jets, c-jets and light jets, with the purpose of identifying interesting b-jets and reject c-jets and light jets as background. The most recent algorithm developed for Run II (MV2c20) is expected to have a better performance with respect to the one used during Run I (MV1). The improvements are mainly due to the addition of an extra pixel layer and to enanchements of the tracking and b-tagging algorithms. The c-jet tagging efficiency was measured with two different methods during Run I, using a sample of c-jets containing D^* charged meson and using a sample of c-jets produced in association with a W boson. The preliminary results of the first measurement of the c-jet tagging efficiency on 13 TeV 2015 data, performed with the "$W + c$ method", indicate that MV2c20 has a better c-jet rejection than MV1.

REFERENCES

[1] The ATLAS Collaboration, *JINST*, **11** (2016) P04008, CERN-PH-EP-2015-216, arXiv:1512.01094 [hep-ex].

[2] The ATLAS Collaboration, *Expected performance of the ATLAS b-tagging algorithms in Run II*, ATL-PHYS-PUB-2015-022 (2015).

[3] The ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, CERN-LHCC-2010-013, ATLAS-TDR-19.

[4] The ATLAS Collaboration, *b-jet tagging calibration on c-jets containing D^{*+} mesons*, ATLAS-CONF-2012-039 (2012).

[5] The ATLAS Collaboration, *Calibration of the performance of b-tagging for c and light-flavour jets in the 2012 ATLAS data*, ATLAS-CONF-2014-046 (2014).

[6] The ATLAS Collaboration, *Calibration of the b-tagging effficiency for c jets with the ATLAS detector using events with a W boson produced in association with a single c quark*, ATLAS-CONF-2013-109 (2013).