PROCEEDINGS OF SCIENCE

Jet substructure and tagging at LHCb

Lorenzo Sestini* **on behalf of the LHCb collaboration**
INFN and University of Padova
E-mail: lorenzo.sestini@cern.ch

The LHCb experiment offers a complementary phase space to ATLAS and CMS to study jet processes, thanks to the forward acceptance and the large bandwidth of the trigger allowing low energy thresholds. Moreover, thanks to the precise tracks reconstruction near the interaction point performed by the Vertex Locator detector, it is possible to study the jet substructure. In these proceedings the latest measurements on the jet substructure performed during the LHC Run I and Run II data taking are presented.

Sixth Annual Conference on Large Hadron Collider Physics (LHCP2018)
4-9 June 2018
Bologna, Italy

PoS(LHCP2018)204

---

*Speaker.

1. Introduction

LHCb is a forward spectrometer, initially designed for b and c quarks physics [1]. Within the LHC experiments, LHCb alone provides precision coverage in the forward region of pp collisions corresponding to the $2 \leq \eta \leq 5$ pseudo-rapidity range. In both LHC Run I (2010-2012) and Run II (2015-today) LHCb demonstrated its capability in electroweak and jet physics as a general purpose forward detector. Jet reconstruction at LHCb takes advantage of several detector features: the excellent momentum and impact parameter resolution, the efficient vertex reconstruction, the electron reconstruction with bremsstrahlung recovery and the particle identification system.

At LHCb jets are reconstructed by selecting tracks, calorimeter clusters and metastable particles as inputs with a particle flow algorithm. Jets are clustered using the anti-k_t algorithm with a radius parameter of 0.5. Jet energy corrections calibrated on simulation are also applied. The jet reconstruction efficiency is above 90% for jets with transverse momentum (p_{T}) greater than 20 GeV$/c$. The energy resolution of final jets varies from 10% to 15% in the 20–100 GeV$/c$ transverse momentum range [2]. At LHCb the jet substructure has been studied for the identification of b- and c- jets (secondary vertices) and for QCD studies (J/ψ in jets).

2. Identification of b and c jets

In the study of processes with heavy quarks in the final state (b or c) it is fundamental to identify the flavour of the quark that "generates" a jet. This is achieved by using a heavy flavour tagging algorithm [3] that analyses the jet substructure in order to take a decision on the jet flavour. The algorithm checks if a secondary vertex (SV) reconstructed by the Vertex Locator (VELO), detached from the Primary Vertex and compatible with a B-hadron (b-jets) or a D-hadron not from a B decay (c-jet), is found inside the jet. In case of positive answer the jet is tagged as potentially generated from a heavy flavour quark. To further remove light jet contamination and to distinguish b-jets from c-jets multivariate algorithms (MVA) are used. Two boosted decision trees (BDTs) are employed: one for the heavy/light jets separation, BDT$(bc|udsg)$, and the other for the b/c jets separation, BDT$(b|c)$. Simulated samples of heavy quark/light quark jets and b/c jets generated with Pythia 8 are used as signal/background samples for the BDTs training. The observables in input to the BDTs are those related to the SV that provide the highest discrimination power between the different flavours. These observables are: the SV mass; the SV-corrected mass M_{cor}; the transverse flight distance of the 2-body particle closest to the primary vertex (PV) within those that form SV; the fraction of the jet p_{T} carried by SV; the number of tracks that form SV; the number of tracks that form SV with $\Delta R < 0.5$ from the jet axis; the total charge of tracks in SV; the SV flight distance χ^2. In particular the SV-corrected mass is one of the most sensitive observables to the jet flavour and is defined as

$$M_{cor} = \sqrt{M^2 + p^2 \sin^2\theta} + p\sin\theta,$$

where M and p are respectively the invariant mass and the momentum of the SV and θ is the angle between the SV momentum and flight direction. The 2-dimensional distributions of BDT$(bc|udsg)$ and BDT$(b|c)$ for b-jets, c-jets and light jets simulated samples are shown in figure 1.

The SV tagging performance have been evaluated on data. Data events that contain a fully reconstructed b or c hadron or a high-p_T muon, which are referred as “event-tag”, have been used.

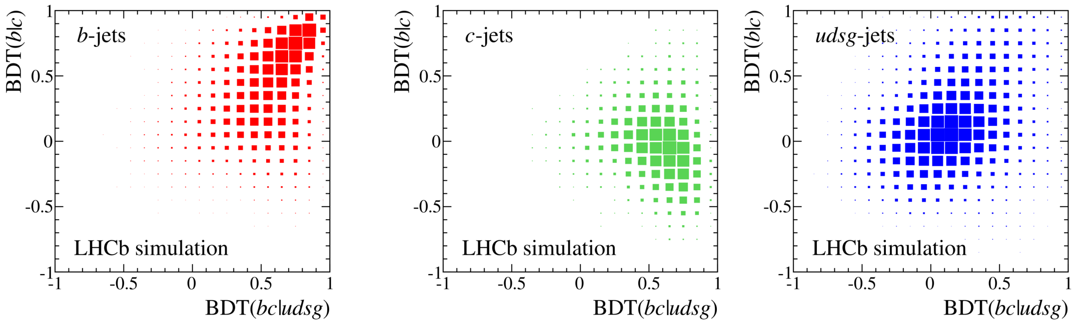

Figure 1: *2-dimensional distributions of BDT$(bc|udsg)$ and BDT$(b|c)$ for b-jets, c-jets and light jets simulated samples.*

A test jet associated to the same PV of the event-tag is required, with a $|\Delta\phi|$ with respect to the event-tag greater than 2.5 to reduce the possibility of background contamination from the event-tag. First of all a combined fit is performed to BDT$(b|c)$ and BDT$(bc|udsg)$ distributions of B+jet, D+jet and μ+jet events where the test jet is tagged. In the fit the yield of tagged b, c and light jets events is measured. Then a combined fit is performed to the $\chi^2_{IP(max-pt)}$ distributions of B+jet, D+jet and μ+jet events where the SV tagging is not applied to the test jet. $\chi^2_{IP(max-pt)}$ is defined as a χ^2 associated to the impact parameter of the highest-$p_{\rm T}$ track in the jet. In the fit the yield of total b, c and light jets events prior to the SV requirement is measured. The efficiencies are calculated by the ratio of the yields measured in the BDTs fit and in the $\chi^2_{IP(max-pt)}$ fit, and are shown in figure 2. For high-$p_{\rm T}$ jets the efficiency is around 60% for b-jets and around 25% for c-jets, with a light jets misidentification rate below 0.3%. The uncertainties on the efficiencies measurement are dominated by the size of the data samples and by the modeling of the templates used in the fits. The jet identification algorithm described here has been used in several LHCb measurements as the $Z \to b\bar{b}$ cross section measurement [4] and the $t\bar{t}$ cross section measurement [5].

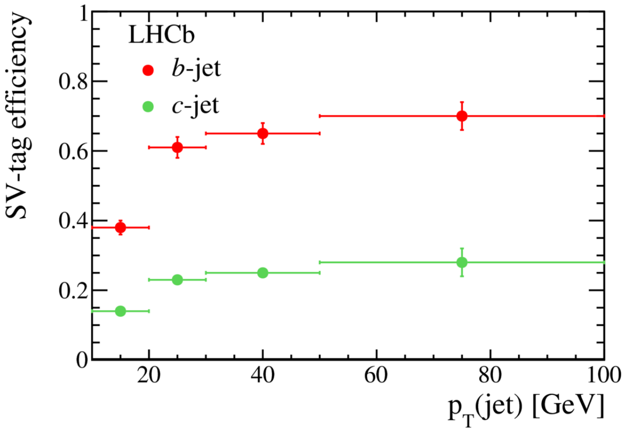

Figure 2: *b-jet and c-jet selection efficiencies as a function of the jet $p_{\rm T}$.*

3. Study of J/ψ production in jets

The production of J/ψ mesons in jets has been studied in the forward region of proton-proton collisions using data collected with the LHCb detector at a center-of-mass energy of 13 TeV [6]. In this analysis the fraction of the jet transverse momentum carried by the J/ψ meson, $z \equiv p_{\mathrm{T}}(J/\psi)/p_{\mathrm{T}}(jet)$, is measured using jets with p_{T}(jet)>20 GeV$/c$ in the pseudorapidity range 2.5<η(jet)<4.0. J/ψ mesons are selected requiring two muons with p_{T}>0.5 GeV$/c$ and p>5 GeV$/c$. The J/ψ must be inside the selected jet cone. One of the most notable features of this analysis is that the jet reconstruction and the J/ψ reconstruction have been performed online at trigger level. The J/ψ yield is determined in several intervals of p_{T}(jet) and $z(J/\psi)$ by fitting the di-muon mass distribution $m(J/\psi)$. Prompt and J/ψ yields are obtained in the same intervals by fitting the pseudo decay time $\tilde{t} = \lambda m(J/\psi) p_{\mathrm{L}}(J/\psi)$, where λ denotes the difference in position along the beam axis between the J/ψ decay and primary vertices and $p_{\mathrm{L}}(J/\psi)$ is the J/ψ longitudinal momentum with respect to the beam axis. Examples of these fits are shown in figure 3. The $z(J/\psi)$ distribution has been obtained by applying an unfolding procedure to correct for the detector response. The measured distributions are then compared with theoretical predictions obtained with Pythia 8 (displaced case) and a leading-order non-relativistic QCD calculation (prompt case). The results are shown in figures 4 and 5.

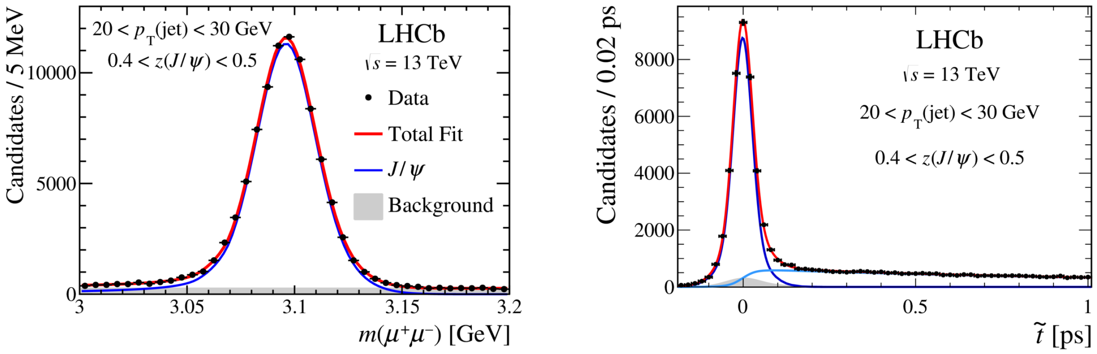

Figure 3: *Results of fits to the J/ψ invariant mass (left) and to the J/ψ pseudo decay time in one particular interval of p_{T}(jet) and $z(J/\psi)$.*

The observed $z(J/\psi)$ distribution for J/ψ mesons produced in b-hadron decays (assumed to be the displaced case) is consistent with expectations. However, the results for prompt J/ψ production do not agree with predictions based on fixed-order non-relativistic QCD, with higher order effects, such as the generation of J/ψ mesons in the parton shower, potentially important for understanding such processes. This is the first measurement of the p_{T} fraction carried by prompt J/ψ mesons in jets at any experiment.

4. Conclusions and future perspectives

In these proceedings the latest results on jet substructure at LHCb have been presented. In particular, at LHCb the jet substructure has been studied for the identification of b- and c- jets

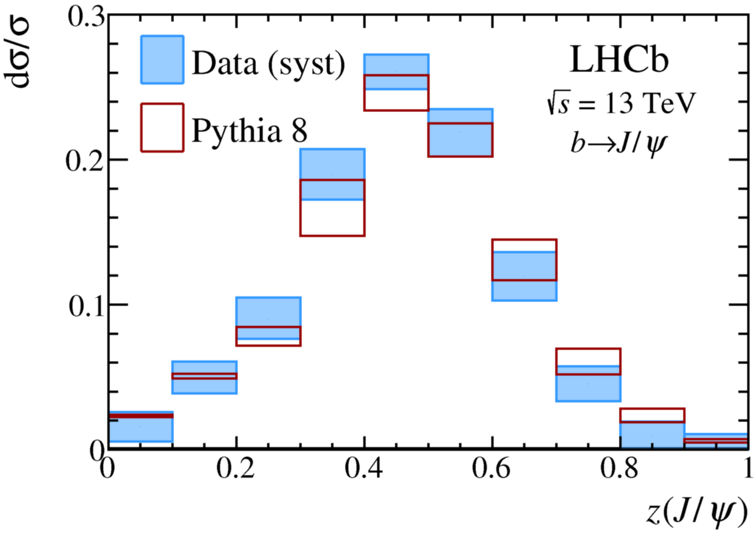

Figure 4: *Measured and theoretical z(J/ψ) distribution for the displaced case.*

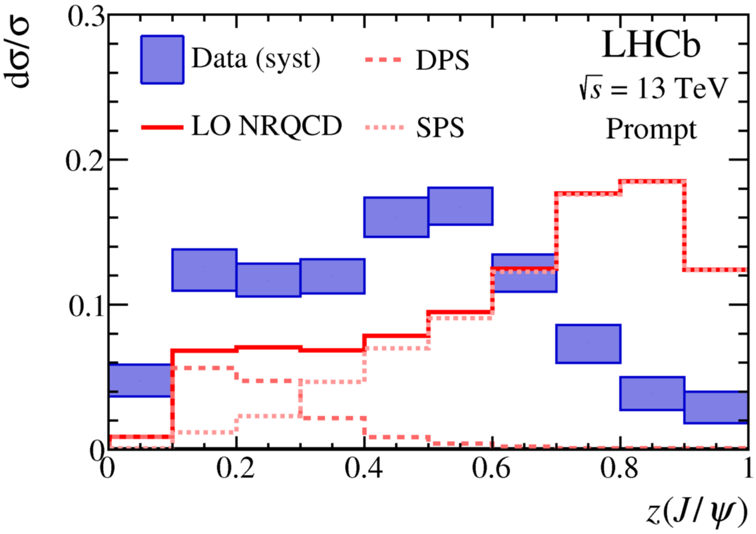

Figure 5: *Measured and theoretical z(J/ψ) distribution for the prompt case.*

(secondary vertices) and for QCD studies (J/ψ in jets). Thanks to the precise secondary vertices reconstruction LHCb developed an excellent jet heavy flavour tagging system. Work is in progress to further improve the tagging performance: for example there is the possibility to use not only the secondary vertices but also all the jet constituents to identify the flavour. Deep learning and image classifications algorithm with the jet substructure as input can be used to achieve better discrimination. Moreover jet substructure could be used to identify $b\bar{b}$ or $c\bar{c}$ fat jets for QCD studies or resonances searches.

References

[1] A. A. Alves Jr. *et al.*, *The LHCb detector at LHC*, JINST 3 (2008) S08005.

[2] R. Aaij *et al.*, *Study of forward Z+jet production in pp collisions at* $\sqrt{s}$ = *= 7 TeV*, JHEP01 (2014) 033.

[3] R. Aaij *et al.*, *Identification of beauty and charm quark jets at LHCb*, JINST 10 (2015) P06013.

[4] R. Aaij *et al.*, *Measurement of forward* $Z \to b\bar{b}$ *production in* pp *collisions at* $\sqrt{s} = 8$ TeV, Phys. Lett. B776 (2017) 430.

[5] R. Aaij *et al.*, *Measurement of forward top pair production in the dilepton channel in* pp *collisions at* $\sqrt{s} = 13$ TeV, arXiv:1803.05188.

[6] R. Aaij *et al.*, *Study of* J/ψ *production in jets*, Phys. Rev. Lett. 118, 192001 (2017).