

B-tagging at CMS

Cristina Ferro^{1,a}

IPHC Strasbourg.

Abstract. The identification of *b* jets is a crucial issue to study and characterize various channels like top quark events and many new physics scenarios. Different b-tagging techniques are defined in CMS which benefit from the long life time, high mass and large momentum fraction of the b-hadron produced in b-quark jet. Efficient algorithms have been developed based on the measure of b-hadron secondary vertex or on tracks with a large impact parameter. Data collected in *pp* collisions at $\sqrt{s}=7\text{TeV}$ in 2011 are used to estimate both the b-tagging efficiency and the mistag rate from light flavor jets.

1 Introduction

The b-tagging algorithms in CMS mainly rely on the long life time, high mass and large momentum fraction of b hadrons produced in b-quark jets, as well as on the presence of soft leptons from semi-leptonic b decays[1]. Due to the high instantaneous luminosity during the 2011 data taking, the number of collision taking place in the same bunch crossing (pileup events) is of the order of 5 to 11 on average.

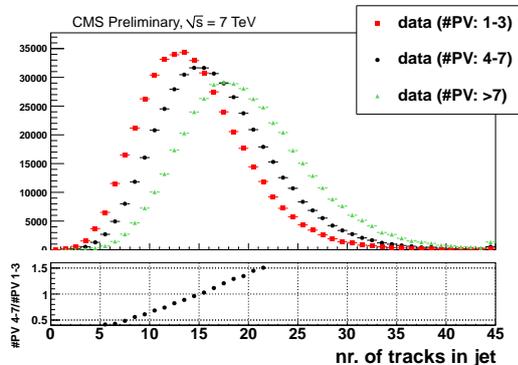


Fig. 1. Number of tracks associated to a jet without any selection cut.

The presence of pileup increases the track multiplicity in the events, as we can see in Fig.(1). This is why a special selection of the tracks was applied in order to remove the tracks originating from pileup[2]. In Fig.(2), the number of tracks passing the selection cuts shows a smaller pileup dependence.

2 The b-tagging observables

The b-tagging algorithms and their study are based on the measure of three main variables: the impact parameter significance of the tracks, the position of the secondary vertex,

^a e-mail: Cristina.Ferro@cern.ch

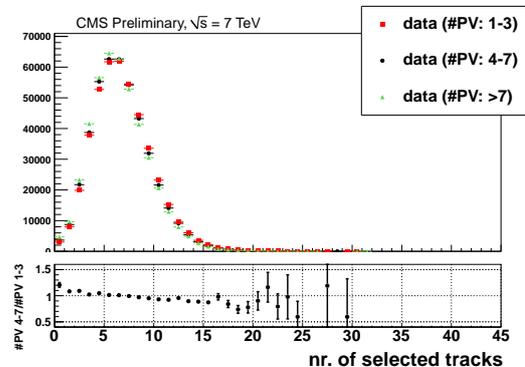


Fig. 2. Number of tracks associated to a jet after selection cuts.

and the transverse momentum of the muon relative to the jet direction. In the following a brief description of these variable is presented.

2.1 The impact parameter significance

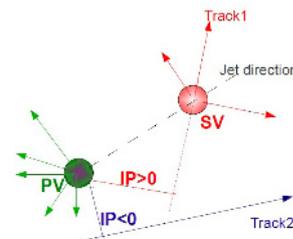


Fig. 3. Geometric meaning of the impact parameter significance.

The impact parameter (IP) is defined as the distance between the track and the primary interaction vertex (PV) at the point of closest approach. The IP is positive (negative) if the track is produced downstream (upstream) with respect to the PV along the jet direction (Fig.(3)). The IP is calculated in 3 dimensions thanks to the good x-y-z resolution provided by the pixel detector. An important features of the IP is that it is Lorentz invariant and due to the b-hadron lifetime the typical IP scale is set by $c\tau \sim 480 \mu\text{m}$. In practice, the impact parameter significance $IP/\sigma(IP)$ is used in order to take into account resolution effects. Thanks to the long lifetime of the b-hadrons the IP from b-jets is expected to be mainly positive, while for the light jets it is almost symmetric with respect to zero (Fig.4).

2.2 The secondary vertex

Thanks to the high resolution of the CMS tracking system, it is possible to directly reconstruct the secondary vertex, the point where the b hadron decays (Fig.3)). The vertex reconstruction is performed using the adaptive vertex fitter. The resulting list of vertices is then subject to a cleaning procedure, rejecting SV candidates that share 65% or more of their tracks with the PV.

2.3 The transverse momentum of the muon

Semileptonic decays of b hadrons give rise to b jets that contain a muon with a branching ratio of about 11%, or 20% when $b \rightarrow c \rightarrow l$ cascade decays are included. This is why the reconstructed muons inside a jet are used to study the performance of the lifetime-based tagging algorithms. The muons are seeded from the CMS muon chambers, and are then linked to tracks found in the tracking system to form global muons. The CMS muon system is able to measure muons with high acceptance resolution and efficiency.

3 B-tagging algorithms

Several b-tagging algorithms are used in CMS [1], [2]. The output of each algorithm is a *discriminator* value on which the user can cut on to select different regions in the efficiency versus purity phase space. In Fig.(4) these discriminators are presented.

- The track counting algorithm identifies a b-jet if there are at least N tracks with a significance of the impact parameter above a given threshold. The tracks are ordered in decreasing $IP/\sigma(IP)$ and the discriminator is the impact parameter significance of the Nth track. To get an high b-jet efficiency we can use the $IP/\sigma(IP)$ of the second track (TCHE), to select b-jets with high purity the third track is the better choice (TCHP).
- The Jet Probability algorithm relies on the $IP/\sigma(IP)$ measurement of all tracks in a jet. One can use the negative tail of the $IP/\sigma(IP)$ distribution to extract the probability density function (PDF) for tracks not coming from b/c-jets. By integrating on the PDF, we can compute the *probability* for tracks to originate from the PV. Then combining the probability of the tracks we can assign to the jet a probability to come from

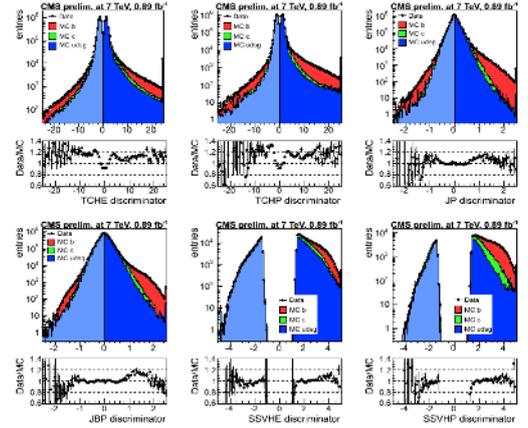


Fig. 4. Discriminators for: *Top left* Track Counting High Efficiency ($IP/\sigma(IP)$), *center* Track Counting High purity ($IP/\sigma(IP)$), *right* JetProbability, *Bottom left* JetBProbability, *center* Simple Secondary Vertex High efficiency, *right* Simple Secondary Vertex High purity.

the PV. The JetBprobability is then defined in a similar way but giving more weight to the four most displaced tracks.

- Soft-Lepton tagging algorithms rely on the properties of muons or electrons from semileptonic b-decay. Due to the large b-quark mass, the momentum of the muon transverse to the jet axis, p_T^{rel} , is larger for muons from b-hadron decays than for muons in light flavor jets.
- Secondary Vertex tagging algorithms rely on the reconstruction of at least one secondary vertex. The significance of the 3D flight distance is used as a discriminating variable. Two variants based on the number of tracks at SV are considered: $N \geq 2$ for *high efficiency* (SSVHE), and $N_{tr} \geq 3$ for *high purity* (SSVHP) [2]. The *combined secondary vertex* algorithm includes this information and provides discrimination even when no secondary vertices are found. The mass of reconstructed charged particles at the secondary vertex is used to measure the b-tagged sample purity.

4 Performance of the taggers

Varying the cuts on the discriminator, we obtain different efficiency of the taggers. We establish standard operating points as, *loose* (L), *medium* (M), and *tight* (T), being the value at which the tagging of udsg jets is estimated from MC to be 10%, 1%, or 0.1%, respectively, for jet transverse momentum of about 80 GeV. In Fig.(5) the performance for different taggers are shown. In Fig.(6) the effects of the pileup on the performance of the TCHE tagger is presented. Thanks to the good selection on tracks the performance of the taggers are not compromised by the pileup events.

5 Physics results

Many measurements have been obtained using the b-tagging algorithms at $\sqrt{s} = 7 \text{ TeV}$. Some of them used the b-tagging

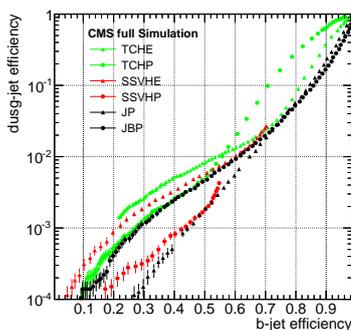


Fig. 5. Performance of all b-taggers obtained on the simulated QCD events. The performance are shown as udsg jets tagging efficiency versus b-jets tagging efficiency.

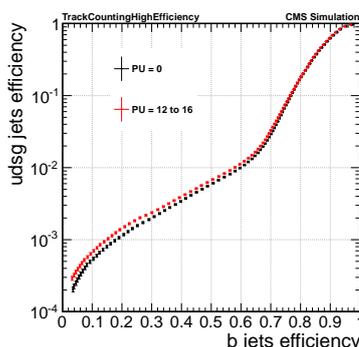


Fig. 6. Light flavor mistag efficiency versus b-tagging efficiency for different pileup scenario, for the TCHE tagger.

algorithms already at trigger level[3]. Indeed, at trigger level, the b-quark candidates can be selected if they have at least one or two tracks with a 3D impact parameter significance above a given threshold. The motivation for applying b-tagging in the trigger is a reduction of the trigger rates, while keeping the signal efficiency high at the same time. The typical rate reduction is a factor of 5-10. In the following a list of the main 2011 physics results obtained thanks to the b-tagging algorithms is presented:

- B-PHYSICS:
 - Inclusive production cross section of b-jets[4].
- EW PHYSICS:
 - Measurement of associated charm production in W final state[5].
- Top-PHYSICS:
 - Cross-section measurement of top pair production in various final states: dileptonic [6],[7], [8], [11], lepton+jets [9], all hadronic [3].
 - Single top in t channel [10].
 - Top mass measurement [11].
- New PHYSICS:
 - Search for supersymmetry in events with b-jets and missing transverse momentum [12].
 - Search for supersymmetry in all hadronic events [13].
 - Search for an Heavy Bottom-like quark [14].
 - Search for an Heavy Top-like quark [15].
 - Search for pair production of a fourth-generation t' quark in the lepton-plus-jets channel [16].
 - Inclusive search for a fourth generation of quarks [17].

References

1. **CMS PAS BTV-11-001**, CMS Collaboration, *Performance of b-jet identification in CMS*.
2. **CMS PAS BTV-11-002**, CMS Collaboration, *Status of b-tagging tools for 2011 data analysis*.
3. **CMS PAS TOP-11-007**, CMS Collaboration, *Measurement of the $t\bar{t}$ production cross section in the fully hadronic decay channel in pp collisions at $\sqrt{s} = 7\text{TeV}$* .
4. **J. High Energy Phys. 03 (2011) 090**, CMS Collaboration, *Inclusive b-hadron production cross section with muons in pp collision at $\sqrt{s} = 7\text{TeV}$* .
5. **CMS PAS EWK-11-013**, CMS Collaboration, *Measurement of associated charm production in W final state I pp collision at $\sqrt{s} = 7\text{TeV}$* .
6. **Phys. Lett. B 695 (2011) 424-443**, CMS Collaboration, *First measurement of the cross section for top-quark pair production in pp collisions at $\sqrt{s} = 7\text{TeV}$* .
7. **CMS PAS TOP-11-005**, CMS Collaboration, *Measurement of the $t\bar{t}$ production cross section in the dilepton channel in pp collisions at $\sqrt{s} = 7\text{TeV}$* .
8. **CMS PAS TOP-11-006**, CMS Collaboration, *First measurement of the $t\bar{t}$ production cross section in the dilepton channel with tau leptons in the final state in pp collisions at $\sqrt{s} = 7\text{TeV}$* .
9. **CERN-PH-EP-2011-085**, CMS Collaboration, *Measurement of the $t\bar{t}$ Pair Production Cross Section at $\sqrt{s} = 7\text{TeV}$ using b-quark Jet Identification Techniques in Lepton + Jet Events*.
10. **Phys.Rev.Lett.107(2011)091802**, CMS Collaboration, *Measurement of the t-channel single top quark production cross section in pp collision at $\sqrt{s} = 7\text{TeV}$* .
11. **J.High Energy Phys.07(2011)049**, CMS Collaboration, *Measurement of the $t\bar{t}$ production cross section and the top quark mass In the dilepton channel in pp collision at $\sqrt{s} = 7\text{TeV}$* .
12. **J.High Energy Phys.07(2011)091802**, CMS Collaboration, *Search for supersymmetry in events with b-jets and missing transverse momentum at LHC*.
13. **CMS PAS SUS-11-006**, CMS Collaboration, *Search for supersymmetry in all hadronic events with b-jet*.
14. **CMS PAS EXO-11-036**, CMS Collaboration, *Search for an Heavy Bottom-like quark in 1.14fb^{-1} of pp collision at $\sqrt{s} = 7\text{TeV}$* .
15. **CMS PAS EXO-11-050**, CMS Collaboration, *Search for an Heavy Top-like quark in the dilepton Final state in pp collision at $\sqrt{s} = 7\text{TeV}$* .
16. **CMS PAS EXO-11-051**, CMS Collaboration, *Search for pair production of a fourth-generation t' quark in the lepton-plus-jets channel with the CMS experiment*.
17. **CMS PAS EXO-11-054**, CMS Collaboration, *Inclusive search for a fourth generation of quarks with the CMS experiment*.