PROCEEDINGS OF SCIENCE

Strategies for b-Tagging Calibration using data at CMS

Jason Keller[*†]
University of Nebraska-Lincoln
E-mail: Jason.Daniel.Keller@cern.ch

In experimental high energy physics, the capability to tag jets produced from bottom (b) quarks has become a necessity. Many Standard Model and new physics processes have a b-jet signature, including top quarks, the Higgs boson, and supersymmetry. Algorithms used to identify b-jets utilize either the lifetime of b-hadrons, or the large rate of decay to leptons, compared to hadrons from charm (c) or light ($udsg$) partons. It is therefore crucial to have the ability of measuring the performance of these algorithms. Methods have been developed in CMS to calculate the efficiency and mistag rate of these algorithms using data.

The 2009 Europhysics Conference on High Energy Physics,
July 16 - 22 2009
Krakow, Poland

[*]Speaker.
[†]On behalf of the CMS Collaboration

PoS(EPS-HEP 2009)441

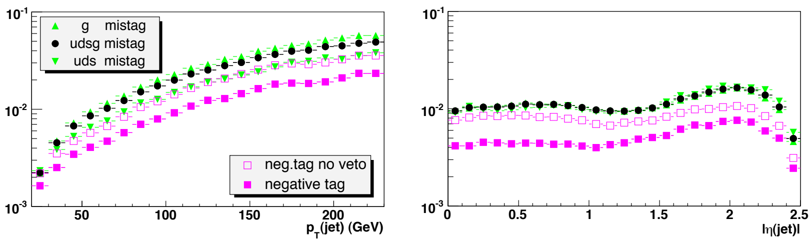
1. Evaluation of *udsg*-jet mistag rates using negative impact parameters

The first method presented describes measuring the rate b-tagging algorithms mistag *udsg*-jets as b-jets. This method uses a quantity called the *signed impact parameter significance*, which is the distance of closest approach of a track to the event primary vertex divided by the distance uncertainty. The impact parameter significance is then given the sign of positive (negative) if the angle between the distance vector and the jet axis with which the track is associated is less than (greater than) 90°. Decays which have a long lifetime, will have displaced vertices, and thus have large, positive impact parameters. Thus, decays from b and c quarks will tend to have a more positive signed impact parameter significance. Tracks associated with *udsg*-jets, however, will tend to have a symmetric signed impact parameter significance. Thus, a sample enriched in *udsg*-jets can be created by taking those jets with tracks having negative impact parameter significance.

The *udsg* mistag rate can then be found from data using the equation

$$\varepsilon_{data}^{mistag} = \varepsilon_{data}^{-} \cdot R_{light}, \tag{1.1}$$

with ε_{data}^{-} = the number of negative-tagged jets divided by the number of taggable jets, and R_{light} = the *udsg* monte carlo mistag rate divided by the monte carlo negative tag rate for all jets. R_{light} is the only parameter to come from monte carlo for this method. The results for this method, compared with monte carlo efficiencies, are shown in figure (1) as a function of p_T and $|\eta|$.



(a) Measured *udsg* mistag rate vs. p_T (b) Measured *udsg* mistag rate vs. $|\eta|$.

Figure 1: Performance of the mistag rate method

2. Performance measurement of *b*-tagging algorithms using data containing muons in jets

This second method uses data samples containing muons within jets, which should be enriched in heavy flavor. The first method using muon-in-jet samples is the $p_{T,Rel}$ template fitting method. $p_{T,Rel}$ is defined as the transverse momentum of the muon with respect to the muon plus jet axis. Because of the larger mass of mesons from b quarks, muons in b-jets will have a larger $p_{T,Rel}$ compared to jets from lighter flavors. It is possible to fit the $p_{T,Rel}$ distribution of the muon-in-jet sample with template functions to determine the b and non-b content of the sample. The sample is then tagged with a given b-tagging algorithm, and the distribution is refit. The ratio of flavor content before and after tagging gives the b-tagging efficiency and non-b mistag rate.

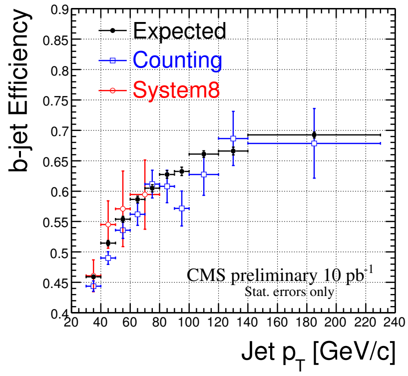


Figure 2: Performance measurements for both the template-fitting and System 8 methods.

Another way muon-in-jet samples can be used is with the System8 method, which utilizes a system of 8 equations to solve for the b-Jet tagging and c + light jet mistagging efficiencies. This method requires both a lifetime jet tagger and a soft muon tagger. Quantities found from data are the number of jets tagged from the lifetime tagger, the soft muon tagger, or both. Inputs from monte carlo include the correlation factors for the lifetime and soft muon tagging/mistagging efficiencies. Figure (2) below shows the results for both the template fitting method and the System8 method.

3. Estimating *b*-tagging performance with ttbar semileptonic and fully leptonic decays

This method utilizes the enriched b-Jet sample that comes from selecting ttbar events. Both semileptonic and fully leptonic ttbar events are used to reduce the QCD background. In the semileptonic case, additional kinematic fits are used to determine which jets in the sample are likely to the b-Jets. The χ^2 of the kinematic fit is then used as a parameter of the method. In fully leptonic events, both jets are assumed to be b-Jets, as the background from other event types is not large. Using the kinematic information from the event, a likelihood method is used to remove further background, and enrich the sample with ttbar events. The assumed b-Jets are then selected for lifetime tagging, and the b-Tagging efficiency is calculated using

$$\varepsilon_b = \frac{1}{x_b}[x_{tag} - \varepsilon_{cl}(1 - x_b)], \tag{3.1}$$

where $x_b(x_{tag})$ is the fraction of b-Jets (tagged jets) in the sample and ε_{cl} the non-b mistag rate. Only x_{tag} is determined from data. This method is currently begin updated for use with current CMS software.

References

[1] The CMS Collaboration, *Evaluation of udsg Mistags for b-tagging using Negative Tags*, *CMS PAS BVT_07_002*.

[2] The CMS Collaboration, *Performance Measurement of b tagging Algorithms Using data containing Muons within Jets*, *CMS PAS BVT_07_001*.

[3] S. Lowette, J. D'Hondt, J.Heyninck, P. Vanlaer, *Offline Calibration of b-Jet Identification Efficiencies*, *CMS Note 2006/013*.

[4] The CMS Collaboration, *Impact of Tracker Misalignment on CMS b-Tagging Performance*, *CMS PAS BVT_07_003*.