EPJ manuscript No.
(will be inserted by the editor)

b-tagging at DØ

K. Hanagaki[1] for the DØ Collaboration

Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, Illinois 60510

Received: date / Revised version: date

Abstract. Many high p_T physics analyses at the Tevatron contain a b-quark and hence a b-jet in the final states. We report on the b-jet identification methods in DØ and their performance. For 0.5% of light jet tagging rate, 40 or 45% of b-jet tagging efficiency is achieved for jets with $35 < E_T < 55$ GeV and $|\eta| < 1.2$.

PACS. 29.40.Gx Tracking and position-sensitive detectors – 07.05.Kf Data analysis: algorithms and implementation

1 Introduction

In many high p_T physics analyses in the Tevatron, such as low mass Higgs searches, $t\bar{t}$ production, and so on, the final state involves b-quarks or actually b-jets. The cross sections of these interesting processes are much smaller than the dominant QCD production cross sections where many light quark jets (u, d, s, or gluon origin) are created. For example, $t\bar{t}$ cross section is ~ 7 pb, while the total cross section of $p\bar{p}$ collision at the Tevatron is ~ 80 mb. Therefore, the identification of b-jets (or b-tagging) is one of the most important factors in these high p_T physics analyses.

There are two ideas to discriminate b-jets from light quark jets. The first is to make use of the lifetime difference between b-hadrons and the other light hadrons. The b-hadrons have typically 400 or 500 μm of lifetime in $c\tau$. Because of the lorentz boost, they tend to travel a few mm before they decay. On the other hand, the hadrons originated from light quarks decay immediately by the strong force, cascading into hadrons with much longer lifetime, such as pions or kaons. As a result, b-hadrons have a decay vertex displaced from the original $p\bar{p}$ interaction point (primary vertex), while light hadrons do not. The displaced vertex or charged particles which do not originate from the primary vertex is the signature to identify b-jets.

The other widely used idea is to find a lepton (either electron or muon) near the jet. The branching ratio of semileptonic decay of b-hadrons is about 11% (there is also $b \to c \to \mu X$ cascade decays), while the chance to have leptons from light hadron decays is much smaller because of their long lifetime and the lorentz boost. Therefore, existence of associated lepton is a signature of b-jets.

In this report, we discuss the b-tagging methods using the first idea in DØ, as well as their performance. We also make some remarks for b-tagging in both general and specific in hadron collider.

2 The DØ Detector

The detailed description of DØ detector can be found in [1]. Here we describe only the charged particle tracking system which consists of a silicon microstrip tracker (SMT) and a central fiber tracker (CFT), both located within a 2 T superconducting solenoidal magnet. The design is optimized for tracking and vertexing capabilities at pseudorapidities $|\eta| < 3$, where $\eta = -ln(\tan(\theta/2))$ and θ is the polar angle with respect to the proton beam direction (z).

The SMT is composed of six barrels, 12 centrals disks, and four forward disks. The barrels and central disks cover the ~ 25 cm RMS long luminous region or $|\eta|$ up to 1.5. The forward disks provides coverage for $|\eta| < 3$. Each barrel is 12 cm long and consists of 72 ladders arranged in 8 layers with pairs of layers forming four super-layers, occupying the radial space from 2.7 cm to 10.5 cm. The strip pitch varies depending on the detector type, and is typically 50 μm.

The CFT consists of eight super-layers of scintillating fibers, occupying the radial space between 20 and 52 cm. Each super-layer is composed of one doublet fibers aligned along z and another doublet with a stereo angle. The two inner (six outer) layers are 1.66 m (2.52 m) long. The outer layers provides coverage for $|\eta| < 1.7$. The fiber's diameter is 835 μm, leading to doublet layer resolution of about 100 μm.

3 Methods

In DØ we have three methods for b-tagging, two based on the impact parameter of charged tracks (d_0), and the other

Send offprint requests to:

based on reconstruction of the secondary vertex [2]. The charged tracks are reconstructed using stereo information, and hence d_0 and secondary vertex are 3D quantities. The projection of d_0 or the decay length from the primary to secondary vertex in the plane perpendicular to z is used as discriminant.

The primary vertices are reconstructed in two passes using tracks having at least two SMT hits and $p_T > 0.5$ GeV/c. In the first pass, $S(d_0)$ of track is calculated with respect to the coordinate origin. The seed vertices are formed from tracks with $S(d_0) < 100$. Tracks that contribute to a χ^2/d.o.f. greater than a certain threshold are iteratively removed one by one, and new vertices are formed until a stable set of seeds is obtained. In the second pass, vertex fit is performed using tracks with $S(d_0)$ less than a certain threshold with respect to each seed vertex. This improves the position resolution on the vertex, because the fit is less affected by poorly reconstructed tracks. In order to select hard scatter vertex, the p_T distribution of the associated tracks is used. Comparing the p_T of the associated tracks with the distribution obtained from minimum bias events, the probability for the vertex to be consistent with that of soft interaction is computed. The vertex that has the smallest probability is selected as the primary interaction vertex.

With the primary vertex calculated, d_0 can be determined. The sign of d_0 is given by using jet momentum vector, $\overrightarrow{p_T^{jet}}$, and another vector, $\overrightarrow{v_{ip}}$, whose direction is defined as primary vertex to the closest approach of the track under consideration. The d_0's sign is the same as $\overrightarrow{p_T^{jet}} \times \overrightarrow{v_{ip}}$. Figure 1 shows d_0 divided by its measurement

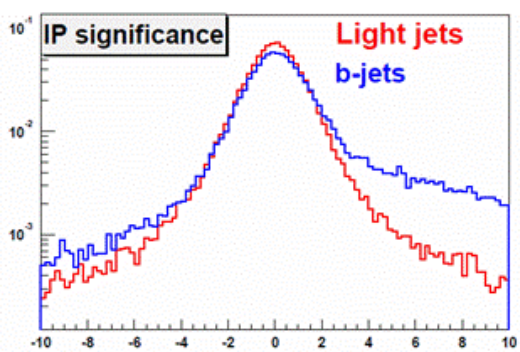

Fig. 1. Impact parameter (d_0) significance in Monte Carlo simulation for b-jets and light jets.

error, $S(d_0) \equiv d_0/\sigma(d_0)$, referred to as signed significance, in Monte Carlo simulation (MC). The symmetric distribution in the light jets is caused by the resolution of the tracking system, while the asymmetry in the b-jets is due to the b-hadron's lifetime as explained in Introduction.

In all the three b-tagging methods, K_S^0, Λ, and photon conversion are explicitly removed by checking the invariant mass of any two oppositely charged tracks with the $S(d_0) > 3$.

3.1 Counting Signed Impact Parameter (CSIP)

In the first method, named CSIP, tracks near a jet within a cone of $\Delta R = \sqrt{\phi^2 + \eta^2} < 0.5$, where ϕ is azimuthal angle, are required to have $p_T > 0.5$ GeV/c and $|d_0| < 0.2$ (0.4) cm in the plane transverse to (along) z. At least two hits in the SMT are also required. In addition, the tracks with $\Delta\Psi > 0.02$, where $\Delta\Psi$ is the opening angle between the track and jet axis, are accepted only if d_0 is positive [1]. A jet is assumed to be b-jet if it contains at least two tracks with $S(d_0) > 3$, or at least three tracks with $S(d_0) > 2$. The operating point, i.e. tightness of the selection criteria, is varied by changing the threshold of the p_T cut on the tracks.

3.2 Jet Lifetime Probability (JLIP)

In the second method, named JLIP, the same criteria as CSIP are used for the track selection, except for the $\Delta\Psi$ requirement. Each track was categorized by three quantities; $p(\sin\theta)^{3/2}$ where p is the particle momentum, hit configuration to SMT and CFT, and the number of tracks associated to the reconstructed primary vertex. In each category, a resolution function of $S(d_0)$ was formed using only the negative part of the d_0. Based on this resolution function, a probability for the track to originate from the primary vertex (P_{trk}) can be calculated. Then all N_{trk}^+ (N_{trk}^-) tracks associated with the jet with a positive (negative) $S(d_0)$ can be used to compute a jet lifetime probability P_{jet}^+ (P_{jet}^-);

$$P_{jet}^{\pm} = P^{\pm} \times \sum_{j=0}^{N_{trk}^{\pm}-1} \frac{(-\log P^{\pm})^j}{j!} ,$$

where $P^{\pm} = \prod_{i=1}^{N_{trk}^{\pm}} P_{trk}$. $P_{jet}^{\pm}$ for light jets are uniform by construction, while P_{jet}^+ for b-jets has a peak near zero. A jet is considered to be a b-jet if P_{jet}^+ is smaller than a threshold.

3.3 Secondary Vertex Tagger (SVT)

The third method named SVT reconstructs a secondary vertex. Tracks are selected in the same manner as JLIP with one tighter requirement, $|d_0| < 0.15$ cm in the plane transverse to z, and formed into track-jets using fixed-cone jet algorithm of $\Delta R = 0.5$. A seed for secondary vertex is built from pairs of tracks, which are used to form the track-jet, with $S(d_0) > 3$. Additional tracks pointing to the seed according to the χ^2 contribution by the vertex fit are attached iteratively. A jet is regarded as b-tagged when it has at least one secondary vertex, whose direction is within a cone of $\Delta R < 0.5$ relative to the jet axis, with a decay length (L_{xy}) divided by its uncertainty L_{xy}/σ_{xy} greater than a threshold.

[1] Tracks with $d_0 < 0$ are accepted to count so-called negative tagging rate which is explained in Section 4.

4 Performance & Issues

4.1 b-tagging Efficiency and Light Jet Tagging Rate

The b-tagging efficiency is measured using two sets of data. The first one contains muons associated with a jet (*muon+jet*). The second dataset is required to have another jet which is b-tagged as well as the requirement for the first sample (*awaytag*). Since the *awaytag* sample is inclusive of the *muon+jet* sample, we assume the relative fraction of c- to light jets are similar in the two samples. For the jets in the sample we apply two tagging algorithms. The first is the one we are trying to measure the efficiency, and second a simple muon-tagging [2], leading to 8 equations in total, i.e. no b-tagged sample, sample tagged by the method under testing, muon b-tagged sample, and sample tagged by both; and we have *muon+jet* and *awaytag* sample each. There are 8 unknowns; four are the number of b-jets and the number of backgrounds (sum of c-jets and light jets) in each *muon+jet* and *awaytag* samples, the other four are the b-tagging efficiency and the tagging rate to the backgrounds by each two methods. Solving the 8 equations with 8 unknowns gives the b-tagging efficiency without relying on much MC information.

One of the major source of systematic uncertainty in the efficiency measurement is the factorization of b-tagging efficiency, i.e. it is assumed that the efficiencies are the same before and after applying muon tagging. The other major source of systematic uncertainty is the assumption that the efficiencies are the same for *muon+jet* and *awaytag* samples. Combining these two and other minor contributions, the relative systematic error in the b-tagging efficiency measurement is about 3-4% in the relevant E_T range. In Fig. 2, for example, the absolute systematic uncertainty in CSIP is shown as a function of E_T. The statistical error is much smaller (less than 1% in relative).

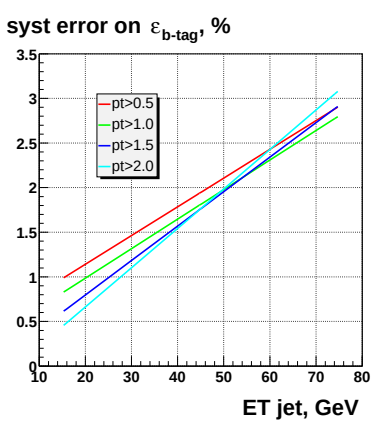

Fig. 2. The absolute systematic uncertainty of b-tagging efficiency measurement in CSIP as a function of jet E_T.

[2] In the tagging by muon, p_T of muon with respect to the jet axis is required to be above a certain threshold.

In principle, light jets are mis-tagged due to the effects of inefficiency and resolution in tracking and vertexing. Therefore, the light jet tagging rate can be estimated by using the tagging rate computed from negative $S(d_0)$ or L_{xy}/σ_{xy} (referred to as negative tagging rate). In reality, however, there are two contributions we have to take into account; the asymmetric distribution of $S(d_0)$ or L_{xy}/σ_{xy} in light jets which can be mainly caused by decays of $\pi^{\pm}$ and $K^{\pm}$, and the fraction of non-light jets. These two effects in the negative tagging rate are corrected by MC, resulting in the light jet tagging rate.

In order to eliminate the effect of fake jet, such as caused by calorimeter noise, or poor coverage of tracking system, the b-tagging efficiency below is defined relative to a so-called taggable jet in which existence of charged tracks in the jet cone is ensured. The taggability is typically 85 or 90% in data. The primary vertex position is also restricted to be $|z| < 60$ cm.

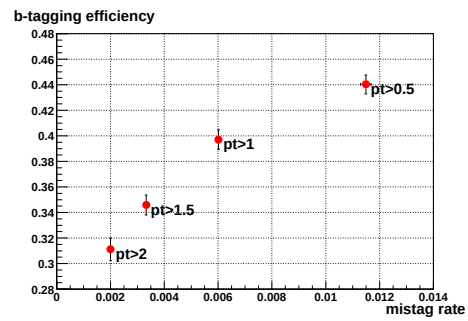

Fig. 3. b-jet tagging efficiency vs light jet tagging rate in CSIP with various p_T cuts on tracks which satisfy the $S(d_0)$ requirements.

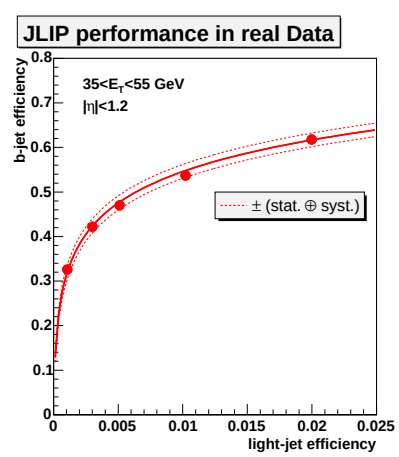

Fig. 4. b-jet tagging efficiency vs light jet tagging rate in JLIP with various cuts on P^{+}_{jet}.

Figure 3, 4, and 5 show the b-tagging efficiency vs light jet tagging rate measured in data for the three different

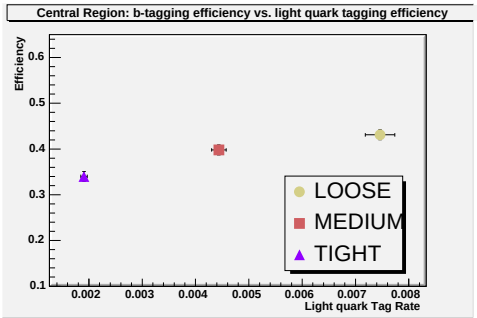

Fig. 5. b-jet tagging efficiency vs light jet tagging rate in SVT with various cuts on L_{xy}/σ_{xy}.

methods. The central region ($|\eta| < 1.2$) and moderate E_T range ($35 < E_T < 55$ GeV) of jets are considered here. With 0.5% of the light jet tagging rate, for example, the b-tagging efficiencies are about 40 or 45% depending on the method.

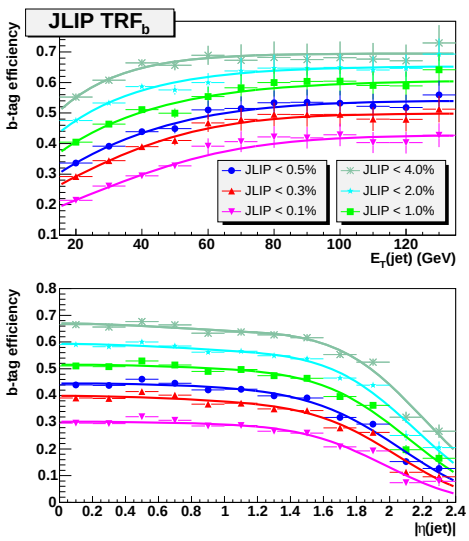

Fig. 6. b-jet tagging efficiency in JLIP as a function of E_T (top) and $|\eta|$ (bottom) of jets.

In Fig. 6, b-tagging efficiency in JLIP is shown as a function of E_T and $|\eta|$ of jets. The efficiency goes up until 60 GeV as the E_T increases, and then reaches the plateau. Thanks to the excellent acceptance coverage by the tracking system, the efficiency is almost flat up to $|\eta|$ of 1.5 or so. Even at $|\eta| = 2$, the efficiency relative to the central region is still 50% or higher.

Among the three methods, the correlation are found to be around 70% in b-tagging efficiency, and about 20% in the negative tagging rate. This implies the possibility of improvement by combining the three methods. The development of such combination is in progress where neural network is employed for the combination.

4.2 Issues

One of the difficulties in the analyses is the fact that the performance in MC does not reproduce reality. If the discrepancy were uniform in jet E_T and η, or jet E_T and η distributions are identical to the data sample used to measure the efficiency, the analyses would not be so complicated. Analyzer could use simply one scale factor to correct the efficiency in MC. However, the real world is not that simple. We therefore measure the b-tagging efficiency and the light jet tagging rate in bins of jet E_T and η, and parametrized them in the 2D space. This parametrization must be used in any analyses in DØ which needs b-jet tagging.

The second issue is the lack of calibration source. There is no process giving pure b-jets in the hadron collider. DØ has developed the novel technique described above to measure the efficiency relying minimally on MC. Still there exists some sources of systematic uncertainty. There is a future possibility to use b-quarks in top decays, but it is not realistic given the current statistics in the Tevatron. Another nasty feature in the hadron collider is that b-jet can be created by gluon splitting, where two b-quarks exist within a jet cone. The b-tagging efficiency should be different, but has not been measured independently in data yet.

The third issue is the performance in high luminosity environment, which will be very crucial in the LHC experiments. The number of interactions per bunch crossing is about three in the highest luminosity operation so far. Even with this luminosity, we already see some minor degradation of performance, which is under current study.

5 Conclusions

To conclude, we have described the methods of identification of b-jets in DØ, how to measure the b-tagging efficiency and tagging rate to light jets, and the results. Typically between 40 and 45% of efficiency was achieved with a fake rate of 0.5%. Some issues both in general and specific in the hadron collider environment were also discussed.

Acknowledgment

I thank the members of b-tagging working group. My special thanks go to those who work hard for designing, constructing, operating and calibrating the detectors.

References

1. V. Abazov *et al.* (DØ collaboration), "The Upgraded DØ Detector" submitted to Nucl. Instrum. Methods Phys. Res. A. physics/0507191.
2. V. Abazov *et al.* (DØ collaboration), Phys. Rev. Lett. **94**, 161801 (2005).