

b Tagging in CMS

Ian Tomalin

Rutherford-Appleton Laboratory, Chilton, Oxfordshire, OX11 0QX, UK

E-mail: ian.tomalin@cern.ch

Abstract. Many of the exotic particles expected at the LHC, such as SUSY, Higgs bosons and top quarks, will decay to b quarks. This paper presents the methods used to identify b -jets at CMS. The algorithms exploit the long B hadron lifetime, semi-leptonic B decays and jet kinematics. The prospect for measuring the performance of these b -tags directly from CMS data is examined. Finally, the use of b -tagging in the High-Level Trigger is explained.

1. Introduction

The ability to identify jets containing B hadrons is important to many of the physics analyses planned for the LHC. Examples include: (i) $b\bar{b}$ and $t\bar{t}$ production, (ii) Higgs bosons produced in association with $b\bar{b}$ or decaying to $b\bar{b}$, and (iii) searches for SUSY such as $\tilde{t}_1 \rightarrow t\chi_2^0$.

Several B hadron properties [1] can be exploited to tag the b -jets: namely the long B lifetime (1.57 ± 0.01 ps), high mass (> 5.2 GeV/ c^2) and high charged decay multiplicity (4.97 ± 0.06). Furthermore, $35.6 \pm 0.5\%$ decay to e^\pm/μ^\pm , including subsequent c or τ^\pm decays. Nonetheless, c -jets are a significant background, notably because of the long D^+ lifetime (1.04 ± 0.01 ps). So too are gluon jets, since these can split to $b\bar{b}$ or $c\bar{c}$.

The lifetime-based b -tags are particularly reliant on the all-silicon CMS Tracker [2]. Tracks of $p_t = 10$ (1) GeV/ c have an impact parameter resolution in both $r - \phi$ and $r - z$ planes of ≈ 20 (100) μm . Unfortunately, the hadronic tracking efficiency in jets is limited by nuclear interactions to $\approx 84\%$. Jet and lepton reconstruction are also important to b -tagging. Jets of $E_t = 100$ GeV are reconstructed in the calorimeters with E_t and angular resolutions of ≈ 14 GeV and $\approx 3^\circ$, respectively. The published efficiency, to identify leptons in jets of $\approx 85\%$ for $p_t > 6$ GeV/ c , is now improving due to algorithmic changes.

With the exception of the High-Level Trigger, further details on the topics described here can be found in the CMS Physics TDR [3].

2. b -Tagging Algorithms

The simplest lifetime-based b -tags are based on the 3-D signed impact-parameter d_0 . This is the distance of closest approach of a track to the primary vertex, signed positive/negative if the track passes closest to its associated jet-axis down/up-stream of the primary vertex. Quality cuts, such as $|d_0| < 2$ mm and jet-track miss-distance less than 0.7 mm, are applied to reject tracks that are badly reconstructed or originate from γ -conversions etc.

The *track-counting* b -tag orders the tracks in a jet by decreasing impact-parameter significance d_0/σ . Its output is the d_0/σ of the N^{th} track, where N is typically 2 or 3. As shown in Fig. 1, this has a large, positive tail for b -jets.

The slightly more sophisticated *probability* b -tag calculates a confidence-level for each track in the jet, that it is compatible with the primary vertex. It then combines these confidence-levels to calculate a pseudo-confidence-level that all the tracks in the jet come from the primary vertex. This is its b -tag discriminator. A key ingredient of this tag is knowledge of the $|d_0|/\sigma$ resolution function. This can be derived from data, since it is to good approximation given by the d_0/σ distribution of tracks with $d_0 < 0$. The resolution function is found to differ for various categories of tracks (depending on their χ^2 , no. of hits etc.), so is measured separately for each. Fig. 2 shows that for a 50% b -tagging efficiency, the *probability* tag has only a 0.4% uds -jet mis-tagging rate. For tight b -tag cuts, the largest mis-tag rate is from gluon jets, since these can split into heavy quarks. This effect is particularly important, since in a typical QCD multi-jet event, most jets are produced by gluons.

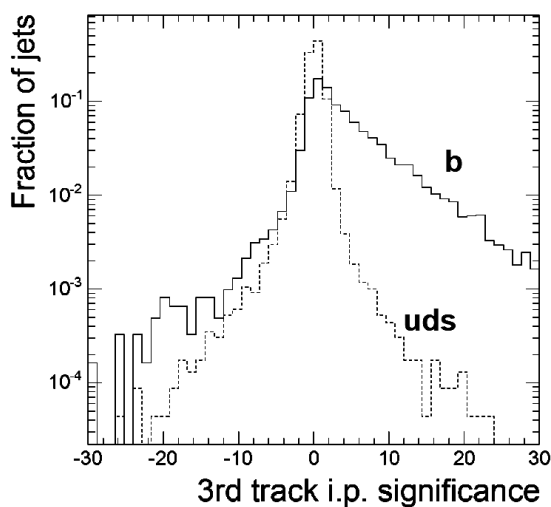


Figure 1. d_0/σ of the 3rd most significant track in the jet.

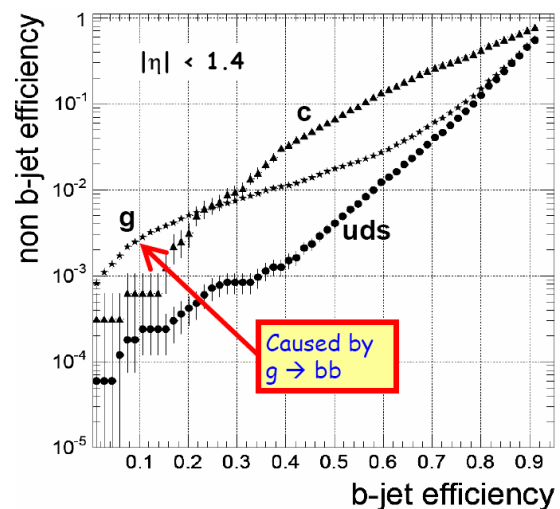


Figure 2. Performance of the *Probability* b -tag.

A different way to tag b -jets is to search for the e^\pm/μ^\pm from semi-leptonic B decay. This often has a large momentum component p_\perp perpendicular to its associated jet axis. Several variables relating to the lepton, such as its p_\perp and d_0/σ are combined using a neural network. As shown in Fig. 3, this b -tag is less efficient than the lifetime tags, because of the small B semi-leptonic branching ratio.

The most sophisticated algorithm is the *combined S.V.* b -tag. This attempts to reconstruct a secondary vertex in each jet [4]. Using a likelihood ratio technique, it then combines several variables related the vertex, such as its decay length significance, mass (shown in Fig. 4) and multiplicity, with additional variables such as track impact parameter significances. The resulting b -tag performance is marginally better than that of the *probability* tag.

3. Measuring the b Tag Performance with Data

Several methods are being developed to measure the b efficiency and uds and c mis-tag rates with real data. One of the prettiest consists of selecting a pure sample of b -jets from $t\bar{t}$ events and measuring what fraction can be b -tagged to determine the b efficiency.

The method uses $t\bar{t} \rightarrow b\bar{t}b\bar{t}$ and $t\bar{t} \rightarrow b\bar{t}b\bar{t}j$ events. For the latter, a kinematic fit is applied using the t and W mass constraints, and one jet is required to be b -tagged to help suppress the Wj background. One obtains an $\approx 85\%$ pure sample of b -jets, which allows the b efficiency to be measured with a relative precision of $\approx 7\%$ with only 1 fb^{-1} of data. The dominant systematic uncertainty is from QCD radiation.

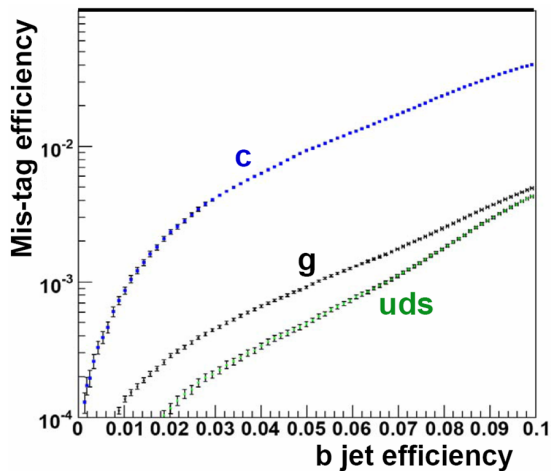


Figure 3. Performance of the $b \rightarrow \mu$ tag.

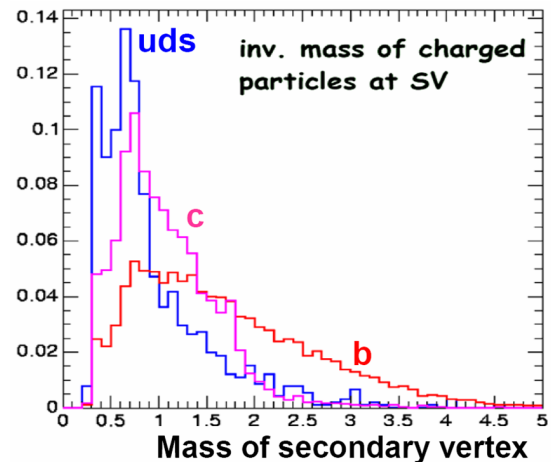


Figure 4. Invariant mass of charged tracks associated with secondary vertex.

4. Use of b Tagging in the CMS High-Level Trigger

The CMS High-Level Trigger (HLT) relies on C++ software running on a huge computer farm. A trigger path exploiting b tagging has been developed [5]. It consists of the following steps:

- (i) Events must pass a trigger based on jet E_t .
- (ii) Tracks are reconstructed using only the Pixel Tracker, yielding only 3 hits per track ! They are used to reconstruct the primary vertex and for b tagging. This takes only 18 ms/event on a typical PC. Events must have a b -tagged jet to pass the trigger.
- (iii) The b -tags are confirmed (with tighter cuts) using tracks reconstructed in a cone around each previously b -tagged jet. Both Pixel + Strip Trackers are now used, but tracking stops when 8 hits are assigned to each track. The event must have a confirmed b -tagged jet. This takes 300 ms/event, but since this final trigger level only runs on a fraction of the events, this is acceptable.

The b -tagging at HLT suppresses minimum bias background by a factor ≈ 20 , whilst suppressing a typical $t\bar{t}$ signal by only a factor ≈ 3 .

5. Conclusions

Several b -tag algorithms are ready for use in CMS (using impact parameters, secondary vertices and $b \rightarrow l$). For a b -jet tagging efficiency of 50%, they can achieve a uds -jet mistag rate of 0.3%.

Various techniques are being developed to measure with data the b , c and uds efficiencies of the b -tags. One based on $t\bar{t}$ events looks very promising.

It will be possible to select b -jets in the CMS high-level trigger.

References

- [1] Particle Data Group, "Review of Particle Physics", J. of Phys. **G33** (2006) 861-867.
- [2] A.Satpathy, "CMS Silicon Tracker: Overview and Status", proceedings of this EPS-2007 conference.
- [3] CMS Physics TDR, Vol. 1, CERN-LHCC-2006-001 (2006).
- [4] W.Erdmann, "Vertex Reconstruction at the CMS Experiment", proceedings of this EPS-2007 conference.
- [5] "CMS High Level Trigger", CERN/LHCC 2007-021 (2007).