Efficiency measurement of b-tagging algorithms

Saptaparna Bhattacharya
for the CMS collaboration

Brown University, DPF 2011

August 10, 2011

Properties of b-hadrons

- The lifetime $\sim$1.5 ps ($c\tau = 450\ \mu$m) for a momentum (p) of 20 GeV/c, with the corresponding decay length being $\sim$1.8 mm.
- The high mass of $\sim$ 4.2 GeV and a decay multiplicity of $\sim$ 5 charged tracks.
- The decay kinematics, in particular the pseudo rapidity (η).
- Hard b-fragmentation function : High p_T of decay products, relative to the flight direction of b-hadrons.
- The semi-leptonic decays, branching fraction of $\sim$ 11 %, $\sim$ 20% including cascade decays.

Procedure and Ingredients

Procedure:

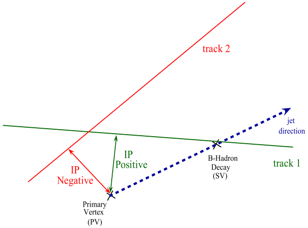
- A variable (discriminator) which is sensitive to the flavor content of the jet is computed from the tracks associated with the jet.
- A working point is chosen. A loose operating point implies that 10% light quarks are included, while medium and tight require the inclusion of 1% and 0.1% light quarks respectively.

Ingredients:

- Jets: Reconstructed by the anti-k_T clustering method, with a cone radius parameter of ΔR=0.5, where R is defined in terms of intervals in azimuthal angle ϕ and pseudorapidity η as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.
- Charged particle tracks : High quality tracks reconstructed with a Kalman Filter based method.
- Primary Vertex: Reconstructed from tracks compatible with beam spot using the Adaptive Vertex Fitter algorithm.

Lifetime based Taggers

- Lifetime based taggers rely on tracks with large impact parameters or on the presence of a reconstructed secondary vertex within a jet.
- Impact parameter: 2D or 3D distance between the track and the vertex at the point of closest approach.
- The impact parameter taggers are : Track Counting (TC) and Jet Probability (JP).
- The secondary vertex based taggers : Simple Secondary Vertex (SSV), based on the reconstruction of at least one secondary vertex.
- High Efficiency and high purity versions of the taggers exist.



Efficiency measurement from muon-jet events : p_{Trel} method

- Due to the high b quark mass, p_T^{rel} is larger for muons from b-hadron decays.
- From the p_{Trel} spectra of b and non-b (c + light flavor jets), one may extract their fractions (f_b^{tag}, f_b^{untag}) with a maximum likelihood fit.
- The fractions and the total yields (N_{data}^{tag}, N_{data}^{untag}) are used to calculate the efficiency: $\varepsilon_b^{tag} = \frac{f_b^{tag}.N_{data}^{tag}}{f_b^{tag}.N_{data}^{tag}+f_b^{untag}.N_{data}^{untag}}$

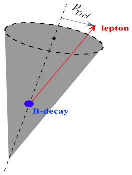
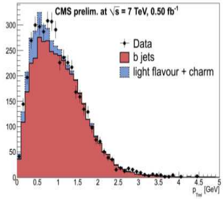


Figure: p_T^{rel} is defined as the transverse momentum of the muon with respect to the jet direction.



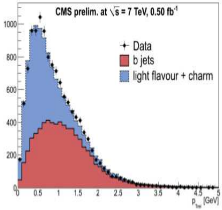


Figure: Fits of the muon p_{Trel} distributions to b and light flavor templates for jets containing muons that (top row) pass or (bottom row) fail the b-tagging algorithm: SSVHPT (Simple Secondary Vertex High Purity Tight OP).

"System 8"

- Applied to a sample of muon jet events.
- A system of 8 non-linear equations are set up and solved using numerical methods.
- Two data samples are used:
  - The muon jet+ away-jet sample : Contains two reconstructed jets and a muon within $\Delta R < 0.4$ of one of the jets. The highest p_T muon is taken when there exist more muons in the jet.
  - The muon jet+tagged-away-jet sample : This sample is created by tagging a b quark in the away jet. Since b quarks are produced in pairs a b quark can be tagged in the same event in another jet.

"System 8" : Implementation

- The first 2 equations, hence are:

$$n = n_b + n_{cl} \quad (1)$$

$$p = p_b + p_{cl} \quad (2)$$

- (n, p) are the muon-in-jets in each sample.
- Taggers used: Two different taggers have been used : A test tagger ("tag") and a p_{Trel} selection.
- Hence the next set of equations are:

$$n^{tag} = \varepsilon_b^{tag} n_b + \varepsilon_{cl}^{tag} n_{cl} \quad (3)$$

$$p^{tag} = \beta_{12}\varepsilon_b^{tag} p_b + \alpha_{12}\varepsilon_{cl}^{tag} p_{cl} \quad (4)$$

- (n^{tag}, p^{tag}) are lifetime tagged.

"System 8 : Equations"

$$n^{p_{Trel}} = \varepsilon_b^{p_{Trel}} n_b + \varepsilon_{cl}^{p_{Trel}} n_{cl} \tag{5}$$
$$p^{p_{Trel}} = \beta_{23}\varepsilon_b^{p_{Trel}} p_b + \alpha_{23}\varepsilon_{cl}^{p_{Trel}} p_{cl} \tag{6}$$

- $(n^{p_{Trel}}, p^{p_{Trel}})$ are obtained by applying a p_{Trel} selection.

$$n^{tag,p_{Trel}} = \beta_{13}\varepsilon_b^{tag}\varepsilon_b^{p_{Trel}} n_b + \alpha_{13}\varepsilon_{cl}^{tag}\varepsilon_{cl}^{p_{Trel}} n_{cl} \tag{7}$$
$$p^{tag,p_{Trel}} = \beta_{123}\varepsilon_b^{tag}\varepsilon_b^{p_{Trel}} p_b + \alpha_{123}\varepsilon_{cl}^{tag}\varepsilon_{cl}^{p_{Trel}} p_{cl} \tag{8}$$

- The last set of equations are a result of the application of both tags.
- The correlation factors are $(\alpha_{12}, \beta_{12}, \alpha_{23}, \beta_{23}, \alpha_{13}, \beta_{13}, \alpha_{123}, \beta_{123})$ obtained from simulations.
- α_{13}, β_{13} are defined as:

$$\beta_{13} = \frac{\varepsilon_b^{tag,p_{Trel}}}{\varepsilon_b^{tag}\varepsilon_b^{p_{Trel}}} \quad \alpha_{13} = \frac{\varepsilon_{cl}^{tag,p_{Trel}}}{\varepsilon_{cl}^{tag}\varepsilon_{cl}^{p_{Trel}}} \tag{9}$$

- The other correlation factors are defined similarly.

Measured b-tagging efficiencies

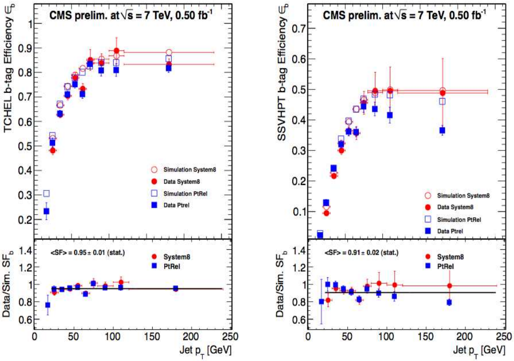


Figure: b-tagging for the TCHEL (left panel) and SSVHPT (right panel) taggers as a function of muon-jet p_T. Both lower panels show data/MC scale factors.

b-tagging efficiencies

- Measured b-tagging efficiencies and data/MC scale factors for several b-tagging algorithms. Uncertainties are statistical for ϵ_b^{tag} and statistical+systematic for SF_b.

b-tagger 50-80 GeV/c	ϵ_b^{tag} PtRel	SF_b^{tag} Ptrel	ϵ_b^{tag} System8	SF_b^{tag} System8
JPL	0.82 ± 0.01	$0.97 \pm 0.01 \pm 0.05$	0.85 ± 0.02	$1.00 \pm 0.02 \pm 0.07$
TCHEL	0.76 ± 0.01	$0.95 \pm 0.01 \pm 0.05$	0.77 ± 0.01	$0.96 \pm 0.02 \pm 0.05$
TCHEM	0.63 ± 0.01	$0.93 \pm 0.02 \pm 0.06$	0.63 ± 0.02	$0.93 \pm 0.02 \pm 0.07$
TCHPM	0.48 ± 0.01	$0.92 \pm 0.02 \pm 0.05$	0.49 ± 0.01	$0.93 \pm 0.03 \pm 0.09$
SSVHEM	0.62 ± 0.01	$0.95 \pm 0.02 \pm 0.07$	0.60 ± 0.01	$0.94 \pm 0.02 \pm 0.06$
SSVHPT	0.38 ± 0.01	$0.89 \pm 0.02 \pm 0.06$	0.37 ± 0.01	$0.90 \pm 0.03 \pm 0.05$
TCHPT	0.36 ± 0.01	$0.88 \pm 0.02 \pm 0.05$	0.37 ± 0.01	$0.88 \pm 0.03 \pm 0.07$

b-tagging efficiencies

- Measured data/MC scale factors for several b-tagging algorithms in the overall jet p_T range from 20 to 240 GeV/c for pseudorapidity $|\eta| < 2.4$, $|\eta| < 1.2$, $1.2 < |\eta| < 2.4$. Uncertainties are statistical for ϵ_b^{tag} and statistical+systematic for SF_b. Both p_{Tel} and System8 provide values compatible with each other.

b-tagger 20-240 GeV/c	SF_b $\lvert\eta\rvert < 2.4$	SF_b $\lvert\eta\rvert < 1.2$	SF_b $1.2 < \lvert\eta\rvert < 2.4$
JPL	0.99 ± 0.01± 0.10	0.99 ± 0.01 ± 0.10	0.98 ± 0.01± 0.10
TCHEL	0.95 ± 0.01± 0.10	0.95 ± 0.01 ± 0.10	0.95 ± 0.02± 0.10
TCHEM	0.94 ± 0.01± 0.09	0.94 ± 0.01 ± 0.09	0.93 ± 0.02± 0.09
TCHPM	0.91 ± 0.01± 0.09	0.91 ± 0.02 ± 0.09	0.90 ± 0.03± 0.09
SSVHEM	0.95 ± 0.01± 0.10	0.95 ± 0.01 ± 0.10	0.93 ± 0.02± 0.09
SSVHPT	0.90 ± 0.02± 0.09	0.89 ± 0.02 ± 0.09	0.90 ± 0.03± 0.09
TCHPT	0.88 ± 0.02± 0.09	0.88 ± 0.02 ± 0.09	0.87 ± 0.03± 0.09

Systematic Uncertainties

- Pile-up: Systematic uncertainties are estimated by constructing 2 samples with high and low pileup regions.
- Away jet tagger: Dependency of the away-jet tagger.
- Muon p_T: Muon p_T cut is varied from its central value at 5 GeV to 7 and 10 GeV.
- Gluon splitting: Account for the error in mismodeling gluon to $b\bar{b}$.
- Closure test: The method was checked for self-consistency.
- p_{Trel} specific: Mismodeling of light jet p_{Trel} spectra.
- System8 specific: Selection of p_{Trel} cut and sample dependence.
- Average systematic uncertainty is between 6%-7%.

Systematic Uncertainties

- Systematic uncertainties for the Ptrel (top) and System8 (bottom) methods.

b-tagger	pile-up	away jet	muon p_T	light	$g \to b\bar{b}$
JPL	0.2%	3.0%	2.3%	2.8%	0.3%
TCHEM	2.4%	3.6%	1.5%	3.3%	0.2%
TCHEM	0.9%	5.1%	1.5%	3.7%	0.1%
TCHPM	1.8%	3.3%	2.6%	3.4%	0.4%
SSVHEM	1.4%	5.8%	1.9%	3.4%	0.6%
SSVHPT	1.1%	4.8%	2.8%	3.4%	0.6%
TCHPT	0.6%	4.3%	2.3%	3.7%	0.3%

b-tagger	pile-up	away jet	muon p_T	p_{Trel}	$g \to b\bar{b}$	sample
JPL	5.1%	1.3%	0.8%	2.2%	0.1%	3.8%
TCHEM	3.3%	2.4%	2.8%	0.9%	0.6%	1.9%
TCHEM	5.8%	2.6%	0.9%	2.0%	0.7%	2.4%
TCHPM	4.8%	3.9%	4.9%	1.7%	2.1%	4.0%
SSVHEM	3.5%	4.6%	0.4%	1.8%	0.2%	3.0%
SSVHPT	1.2%	2.9%	2.8%	2.4%	0.2%	3.0%
TCHPT	3.5%	3.1%	4.0%	2.8%	2.5%	2.5%

Cross-checks with $t\bar{t}$ events

- In the standard model, t decays to Wb at least 99.8% of the time.
- Defining $R_b = \left(\frac{B(t \to Wb)}{B(t \to Wq)}\right)$, where q is any down type quark.
- R_b, if assumed to be 1, can be used to extract the b tagging efficiency.
  - The Profile Likelihood Ratio method : With dilepton $t\bar{t}$ events.
  - The R_b method : With dilepton $t\bar{t}$ events.
  - The Flavor Tag Consistency Method : With lepton+jets $t\bar{t}$ events.
  - The Simultaneous Heavy Flavor and Top method : With lepton+jets $t\bar{t}$ events.
- All of these methods give efficiency values compatible with Ptrel and System8 methods.

Estimation of mistag rate with Negative Taggers

- The mis-tag rate is obtained from tracks with negative impact parameters or secondary vertices with negative decay lengths.
- The mis-tag rate is evaluated as: $\varepsilon_{data}^{mistag} = \varepsilon_{data}^{-}.R^{light}$, where ε_{data}^{-} is the negative tag rate in data and $R_{light} = \varepsilon_{MC}^{mistag}/\varepsilon_{MC}^{-}$ is the ratio between the light flavor mis-tag rate and negative tag rate of all jets in the simulation.

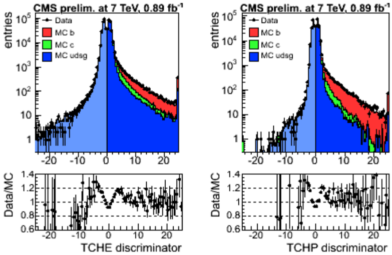


Figure: Signed b-tag discriminators in data (dots) and simulation of light flavor jets (blue), c-jets(green) and b-jets (red area) with a p_T threshold

Systematic Uncertainties

- b and c fractions: 1.9 %
- Gluon fraction: 0.2 %
- Long lived K_s^0 and Λ decays (displaced vertices) and photon conversion and nuclear interactions: 2.0%
- Mismeasured tracks: 0.3 %
- Sign flip: 4.3%
- Event sample (dominant systematic): 10%
- Pile up: 0.7%

Mistag rates

- Mistag rate and data/MC scale factor for different b-taggers with p_T between 50 and 80 GeV. The statistical+systematic uncertainties are quoted.

b-tagger	mistag rate	Scale Factor for light jets
JPL	0.077 $\pm$ 0.001$\pm$ 0.016	0.98 $\pm$ 0.01 $\pm$ 0.11
TCHEL	0.128 $\pm$ 0.001$\pm$ 0.026	1.11 $\pm$ 0.01 $\pm$ 0.12
TCHEM	0.0175 $\pm$ 0.0003$\pm$ 0.0038	1.21 $\pm$ 0.02 $\pm$ 0.17
SSVHEM	0.0144 $\pm$ 0.0003$\pm$ 0.0029	0.91 $\pm$ 0.02 $\pm$ 0.15
SSVHPT	0.0012 $\pm$ 0.0001$\pm$ 0.0002	0.93 $\pm$ 0.09 $\pm$ 0.12
TCHPT	0.0017 $\pm$ 0.0001$\pm$ 0.0004	1.21 $\pm$ 0.10 $\pm$ 0.18

Conclusion

- Several methods have been used to obtain the tagging efficiency of b jets using an integrated luminosity of 0.50 to 0.89 fb^{-1} collected by the CMS experiment in 2011.
- The data/MC scale factor is measured with an uncertainty of 10% for b jets with p_T upto 200 GeV/c.
- For light flavor jets with p_T upto 500 GeV the mistag rate is measured with an uncertainty of 10-20 %.
- Tagging efficiencies are cross checked with independent analyses using $t\bar{t}$ events.
- B tagging is of crucial importance in events with topologies involving b quarks. Single top being a glorious example!

References

For more information please look at CMS PAS BTV-11-001.

Estimation of mistag rate with Negative Taggers

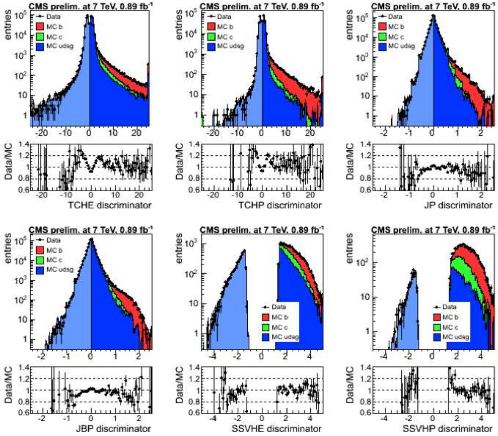


Figure: Signed *b*-tag discriminators in data (dots) and simulation of light flavor jets (blue), c-jets(green) and b-jets (red area) with a p_T threshold of 30 GeV/c