# B-TAGGING EFFICIENCY IN CDF RUN II


HEATHER RAY

hray@umich.edu

Department of Physics, University of Michigan, Ann Arbor, MI 48197 USA



ABSTRACT

B-tagging plays a vital role in the reduction of background in the top quark cross section measurement. The b-tagging efficiency is a measure of how well we are able to properly identify b-jets coming from top decays. The tagging efficiency is measured using Monte Carlo events. The Monte Carlo has been sent through a realistic detector simulation.


## 1 Why Measure the B-Tagging Efficiency

The $t\bar{t}$ cross section is measured using the following formula :

$$\sigma_{t\bar{t}} = \frac{N_{obs} - N_{bgd}}{\epsilon_{t\bar{t}} \cdot \int L} \quad (1)$$

- $N_{obs}$ : number of observed candidate events in data
- $N_{bgd}$ : number of expected background events
- $\int L$ : integrated luminosity of data sample
- $\epsilon_{t\bar{t}}$ : total event detection efficiency
  - product of individual $\epsilon$ of cuts applied to candidate events

$$\epsilon_{t\bar{t}} = \epsilon_{quality\ cuts} \cdot \epsilon_{tag} \quad (2)$$

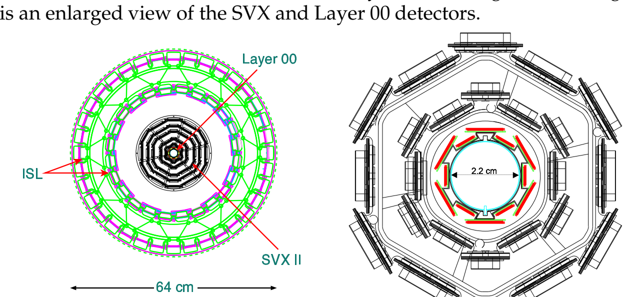
- $\epsilon_{tag}$ : probability to tag *any* jet in a top event
- $\epsilon_{tag}$ : $\epsilon_{b-jet} + \epsilon_{non-bjet}$ (charm, non-heavy flavor)

## 2 B - Tagging

### 2.1 How It Works

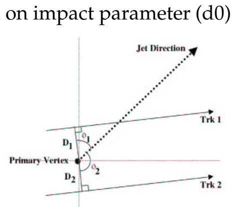
- Bottom quarks produced from top decay hadronize → B mesons
- B's travel ~0.2 cm in CDF detector before decaying.
- Identified by showers of particles, called *jets*
- Jets contain tracks → secondary vertex close to interaction point
- **B-tagging : identify b-jets by reconstructing a secondary vertex**

Shown below are views of our silicon detector, which is primary in the identification of secondary vertexes. The figure on the left shows the silicon detector with each of its three subsystems. The figure on the right is an enlarged view of the SVX and Layer 00 detectors.




### 2.2 SecVtx Algorithm

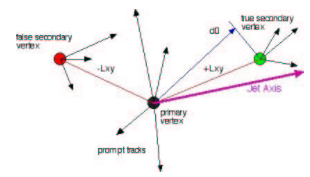
- **B-tagging algorithm (SecVtx) : 2-d, works in $r - \phi$ coordinates**
- Looks for tracks in the cone of jet
- Tracks must pass selection based on impact parameter (d0)
- d0 : distance of closest approach that a particle (track) has with the origin (point of collision)
  (See Figure on right)
- **Need $\geq$ 2 displaced tracks to reconstruct a vertex.**




### 2.2 (cont.) SecVtx Algorithm

Once a secondary vertex is reconstructed an additional cut is made on the $L_{xy}$, or distance from the beamline to the secondary vertex in $r - \phi$ space. **Jets having a secondary vertex which passes the $L_{xy}$ cut are then identified as b-tagged.**

- Tags are positive (+$L_{xy}$) or negative (-$L_{xy}$) (See Figure below)
- $\epsilon_{tag}$ only uses positive tags
- Negative tags : secondary vertexes on opposite side of interaction point w.r.t jet cone axis
- Negative tags : used to estimate mistag background




## 3 Measuring the B-Tagging Efficiency

### 3.1 Method

**Measure the b-tagging efficiency by applying a binomial probability**

$$\epsilon_{b-tag}^{event} = F_{1b} \cdot \epsilon_{btag} \cdot SF + F_{2b} \cdot \epsilon_{btag}^2 \cdot SF^2 + 2 \cdot F_{2b} \cdot \epsilon_{btag} \cdot (1 - \epsilon_{btag}) \quad (3)$$

- $F_{1b}$ : fraction of top events with 1 taggable b-jet
- $F_{2b}$ : fraction of top events with 2 taggable b-jets
- SF : scale factor for SecVtx between data and Monte Carlo,
- $\epsilon_{btag}$ : per jet b-tagging efficiency, measured in Monte Carlo.

ä $F_{1b} \cdot \epsilon_{btag} \cdot SF$ : probability to tag 1 b-jet when 1 is found

ä $F_{2b} \cdot \epsilon_{btag}^2 \cdot SF^2$ : probability to tag 2 b-jets when 2 are found

ä $F_{2b} \cdot \epsilon_{btag} \cdot (1 - \epsilon_{btag})$ : probability to tag 1 b-jet when 2 are found. This factor is multiplied by 2, as there are two ways in which only one of the two b-jets found can be tagged.

### 3.2 Sample

- 400 K of HERWIG generated $t\bar{t}$ Monte Carlo
- Sent through a realistic detector simulation
- Apply same quality cuts to events as used in data
- Purpose of cuts : increase $\frac{S}{B}$ by selecting more top-like events

**B-jets are ones with a B meson matched to a jet in a cone of dR< 0.4**

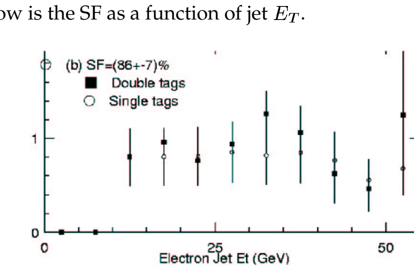
(dR = $\sqrt{(|\phi_B - \phi_{jet}|)^2 + (|\eta_B - \eta_{jet}|)^2}$ )

### 3.3 The Scale Factor

- Measure SecVtx tagging efficiency in high heavy flavor samples
- Use inclusive electron data sample
- Use enhanced heavy flavor Monte Carlo (2→2 generic HERWIG)
- Use single and double tagged events
- Divide $\epsilon_{data}$ by $\epsilon_{MC}$ to get the Scale Factor

ä SF : 86% ± 5% (sys)

Shown below is the SF as a function of jet $E_T$.




### 3.4 Systematics

**There are two sources of systematic errors quoted in this analysis.**

- Mis-measurements in calorimeter jet energy scale ($F_{1b}$, $F_{2b}$, SF)
- Matching B meson to the jet ($F_{1b}$ and $F_{2b}$)

The per-jet btagging efficiency, $\epsilon_{btag}$, is always multiplied by the scale factor, and the systematic on the scale factor dwarfs the small correction to $\epsilon_{btag}$ due to jet energy systematics. As a result, we quote no systematic errors for $\epsilon_{btag}$.

### 3.5 Results

ä $F_{1b}$ : 39.5% ± 0.5% (stat) ± 0.7% (sys)

ä $F_{2b}$ : 48.9% ± 0.6% (stat) ± 1.3% (sys)

ä $\epsilon_{btag}$ : 53.5% ± 0.6% (stat)

Our final event B-tagging efficiency, measured in $t\bar{t}$ events, is :

$$\epsilon_{btag}^{event} = 52.8\% \pm 1.1\% \text{ (stat)} \pm 5.6\% \text{ (sys)}$$

## 4 Checks

**Check method of measuring the b-tagging efficiency by conducting a simple counting experiment.**

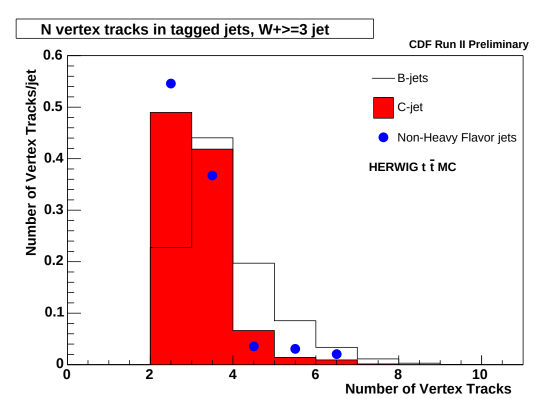
- Divide number of events with $\geq$ 1 tagged B-jet by number of candidate events, before tagging
- Multiply by SF (valid to first order)
- $\epsilon_{btag}^{event}$ : 51.7% ± 0.6% (stat) ± 4.2% (sys)

**When we run the tagger in data, how do we know that we really are tagging B-jets?**

**The number of tracks, the invariant mass of these tracks, and the $L_{xy}$ of tagged jet vertexes have distinct distributions for b, c, and non-heavy flavor jets.**

ä Plot below : compare number of vertex tracks in tagged b-jets, c-jets, non-heavy flavor jets (Monte Carlo)

- B-jets distribution peaks at 3 vertex tracks, substantial tail in higher track bins
- C-jets, non-heavy flavor jets peak at cut-off point for vertexing (2 tracks), falls rapidly

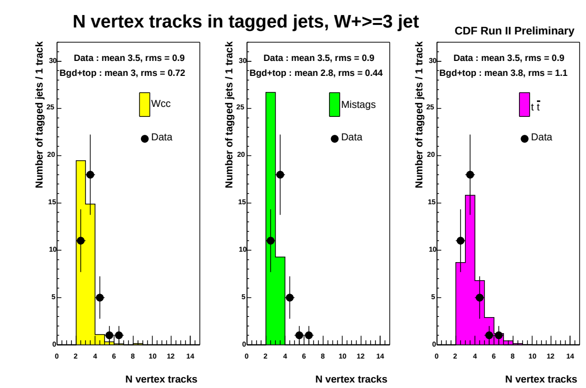



ä Compare to plot at top of next column

- → tagged jets in data vs $Wc\bar{c}$ Monte Carlo (left), non-heavy flavor events (mistags, middle), $t\bar{t}$ Monte Carlo (right)

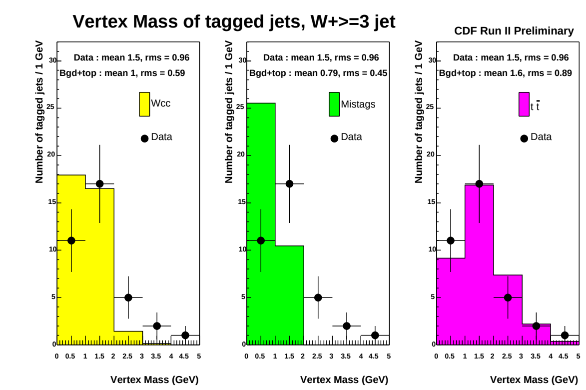
ä Distribution of number of vertex tracks in data agrees well with b-jets in top events.

## 4 (cont.) Checks



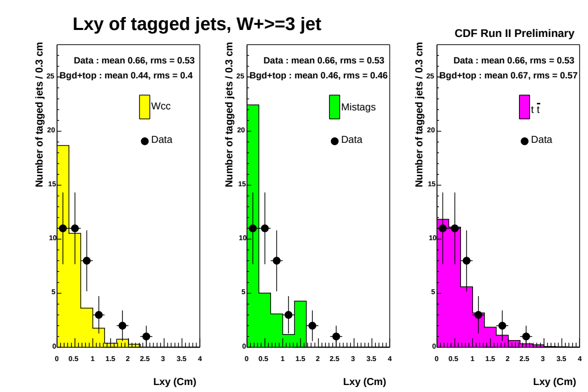

- Mass of the tracks in vertex is also a good discriminating variable
- Figure below : vertex mass of tagged jets in data vs charm jets (left), non-heavy flavor jets (middle), b-jets (right)
- Charm and non-heavy flavor jets distributions have sharp cut-off point at lower mass

ä Both the data and the b-jet distributions have tails extending to higher masses




- $L_{xy}$ distribution varies between these three types of jets

ä Charm, non-heavy flavor jets peak at very small $L_{xy}$, b-jets and data extend out to 3 cm




## 5 Conclusions

We have measured the per-event b-tagging efficiency in $t\bar{t}$ events to be 52.8% ± 1.1% (stat) ± 5.6% (sys). A thorough comparison of tagged data points to b-jet, c-jet, and non-heavy flavor jet distributions has verified that we are indeed tagging B-jets in the data.