PROCEEDINGS OF SCIENCE

Performance studies of b-tagging algorithms using top quark pair production in pp collisions at the center of mass energy 7 TeV with the ATLAS experiment

Kirika UCHIDA on behalf of the ATLAS Collaboration*
Universitaet Bonn
E-mail: kirika.uchida@cern.ch

Algorithms for b-tagging are calibrated with $t\bar{t}$ events for the first time with early data in the ATLAS experiment. Compared to di-jet events that have been used for b-tagging calibrations, $t\bar{t}$ events provide more similar topology to most of the important signals in the experiment. The higher multiplicity of jets, higher jet p_T, and high purity of the $t\bar{t}$ event sample make it an attractive calibration sample. The high $t\bar{t}$ production rate in the LHC enables an initial calibration with 35 pb^{-1}. Two methods and their results are presented.

XXIst International Europhysics Conference on High Energy Physics
21-27 July 2011
Grenoble, Rhône Alpes France

*Speaker.

PoS(EPS-HEP2011)465

1. Introduction

b-tagging is the identification of jets containing a *B* hadron (*b*-jets) using the distinctive properties of *B*'s decays, the long lifetime of *B* hadron, the larger transverse momenta of its decay products with respect to the jet axis due to the large mass of *B* hadron, etc. Two different simple tagging algorithms are studied in early data. The JetProb algorithm uses the impact parameters of track associated with a jet, taking advantage of the fact that the impact parameters of tracks in *b*-jets are distributed more widely than the resolution function. The SV0 algorithm reconstructs secondary vertices from those high impact parameter tracks. Good tracking performance in the inner detector is crucial for the algorithms, and direct calibrations in data is mandatory because the tracking and vertex finding depend on data taking conditions that are not simulated perfectly.

Calibrations of the *b*-tagging algorithms have been done with the QCD di-jet sample. The sample has the advantage of high statistics. However, the calibrations break down for *b*-jets above 100 GeV because it is difficult to estimate the jet flavor composition[1]. The $t\bar{t}$ methods do not have these flavor estimation difficulties and the further advantage of $t\bar{t}$ is the topology more similar to most physics analysis. Especially, the $t\bar{t}$ sample covers *b*-jet p_T's up to 200 GeV. The calibrations with the $t\bar{t}$ sample are already possible with early data taken in 2010 with the high $t\bar{t}$ production rate at LHC. Single lepton and dilepton channels are analyzed for this *b*-tagging performance study and $t\bar{t}$ event selection is done based on the topology; isolated leptons, at least two jets, and the large missing transverse mass. The selections are the standard in top analysis in the ATLAS experiment.

2. Tag-counting method

The tag-counting method counts the number of events with n *b*-tagged jets, N_n. It determines the $t\bar{t}$ cross-section and the *b*-jet tagging efficiency simultaneously with a likelihood function,

$$L = \prod_{n=1}^{3} \text{Poisson}(N_n, < N_n >),$$

where the expectation of N_n is

$$< N_n > = \sum_{i,j,k} \Big\{ (\sigma_{t\bar{t}} \cdot BF \cdot A_{t\bar{t}} \cdot \mathscr{L} \cdot F_{ijk}^{t\bar{t}} + N_{\text{bkg}} \cdot F_{ijk}^{\text{bkg}}) \times \sum_{i'+j'+k'=n} C_i^{i'} \cdot \varepsilon_b^{i'} \cdot (1-\varepsilon_b)^{i-i'} \cdot C_j^{j'} \cdot \varepsilon_c^{j'} \cdot (1-\varepsilon_c)^{j-j'} \cdot C_k^{k'} \cdot \varepsilon_l^{k'} \cdot (1-\varepsilon_l)^{k-k'} \Big\}.$$

$< N_n >$ is estimated by summing over the number of pre-tagged jets $[i, j, k]$ and tagged jets $[i', j', k']$ with i for *b*-jets, j for *c*-jets (originating from *c*-quarks), k for light-jets (all other jets), the number of permutations $[C_\alpha^{\alpha'}]$ with α for the number of the three jet flavours, the fraction of pre-tagged events $[F_{ijk}]$ with the jet flavor composition, the branching fraction of the $t\bar{t}$ decay $[BF]$, the $t\bar{t}$ signal acceptance $[A_{t\bar{t}}]$, the integrated luminosity $[\mathscr{L}]$, and the efficiency to tag a b, c and light-jet $[\varepsilon_b, \varepsilon_c, \varepsilon_l]$. F_{ijk} is estimated in simulation, ε_c and ε_l are fixed to the values found in the simulation corrected with data-to-simulation scale factors measured by other calibrations.

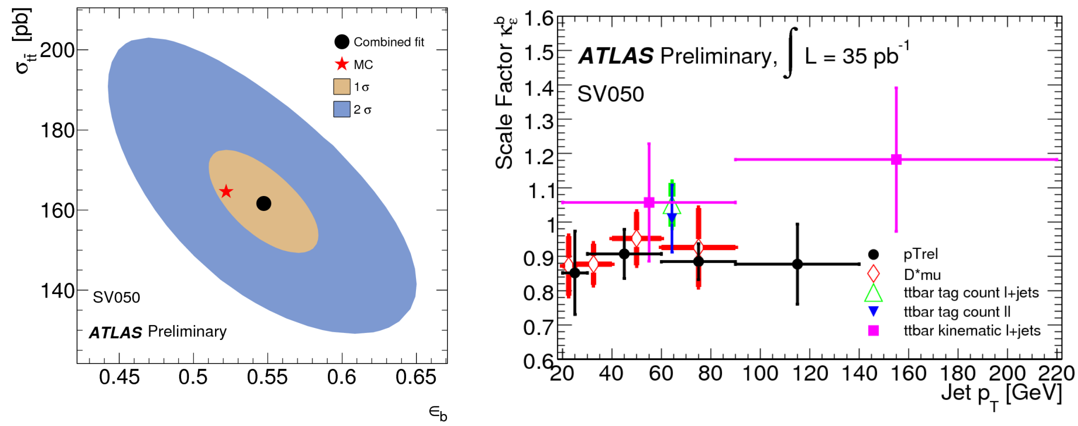

Figure 1: The measured $t\bar{t}$ cross-section vs. b-tagging efficiency in single-lepton (Left) and measured ε_b as a function of p_T with other measurements (Right)[2].

3. Kinematic method

The kinematic method requires the 1st or 2nd highest-p_T jet to be tagged to obtain high purity $t\bar{t}$ sample and measures the b-jet tagging efficiency, ε_b with remaining jets in the events. The measurement is done in single lepton channel in 2 p_T bins. The formula to calculate ε_b is

$$\varepsilon_b = \frac{1}{f_b} \times \left(\frac{N_{\text{tagged}}}{N} - \varepsilon_c f_c - \varepsilon_l f_l - \varepsilon_{\text{QCD}} f_{\text{QCD}} \right),$$

where [ε, f] are the fake rates and fraction of each subscript, where [b, c, l, QCD] are for b, c, light-jets estimated from the simulation (not including QCD), and QCD-jets are from data driven measurement. [N] and [N_{tagged}] are the number of pre-tagged and tagged jets.

4. Result

Figure 1 shows the results of the calibrations of the SV0 algorithm tuned to tag $\sim 50\%$ of b-jets in $t\bar{t}$ sample. The $t\bar{t}$ cross-section vs. b-jet efficiency from the tag-counting method on the single-lepton channel is shown on the left. The right figure shows measured b-jet efficiencies together with other calibrations. The $t\bar{t}$ kinematic method covers higher jet p_T than the di-jet methods. The detail of the tag-counting and kinematic methods can be found in [2].

References

[1] The ATLAS collaboration *Calibrating the b-Tag Efficiency and Mistag Rate of the SV0 b-Tagging Algorithm in 3/pb of Data with the ATLAS Detector* ATLAS-CONF-2010-099 (2010), http://cdsweb.cern.ch/record/1312145

[2] The ATLAS collaboration *Calibrating the b-Tag Efficiency and Mistag Rate in 35/pb of Data with the ATLAS Detector* ATLAS-CONF-2011-089 (2011), http://cdsweb.cern.ch/record/1356198