Measuring the b-jet tagging efficiency using top anti-top events with ATLAS data

Dominik Duda*

on behalf of the ATLAS collaboration

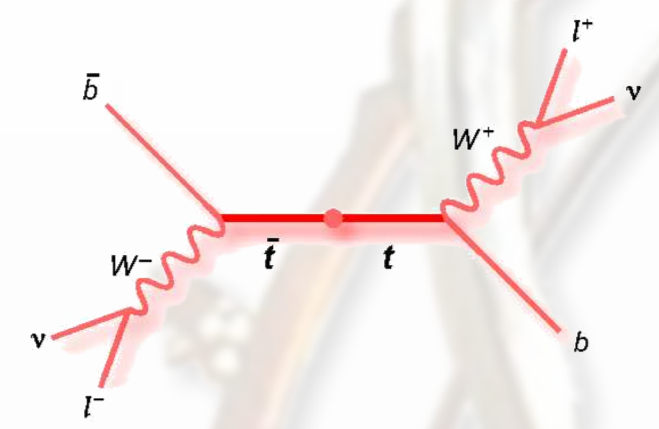
*Bergische Universität Wuppertal

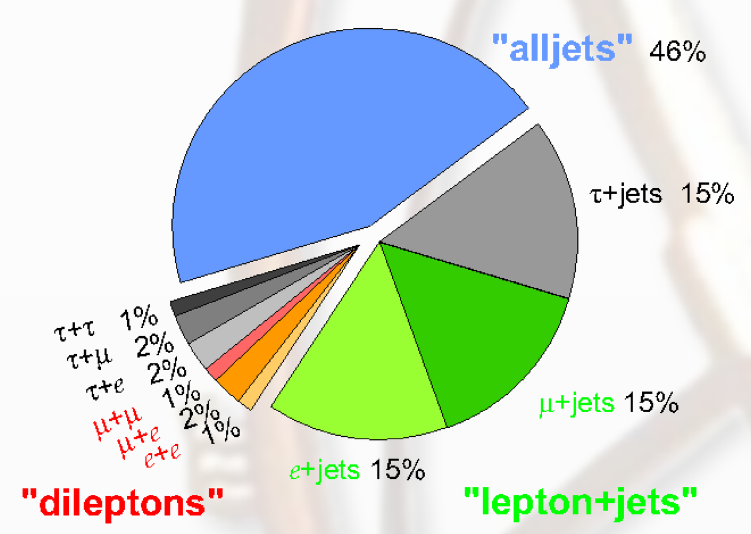
FSP 101 ATLAS

GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

Introduction

Since the decay of a **ttbar pair has a very clear signature** and because of the fact that the top quark **almost exclusively decays to a W boson and a b-quark**, a sample of top anti-top events is **ideal for calibrating the various b-tagging algorithms** used in ATLAS analyses.





Decay of an top anti-top pair in the lepton + jets channel (top) and the branching ratios into all possible final states (bottom).

The calibration results are presented in the form of data-to-simulation scale factors

$$\kappa_{\varepsilon_b}^{data/sim.} = \frac{\varepsilon_b^{data}}{\varepsilon_b^{sim.}}$$

in which the measured b-tagging efficiency is divided by the b-tagging efficiency in simulated events. Currently **three different ttbar calibration methods are in use**, where both the lepton+jets and the dileptonic decay channels are considered [1] [2].

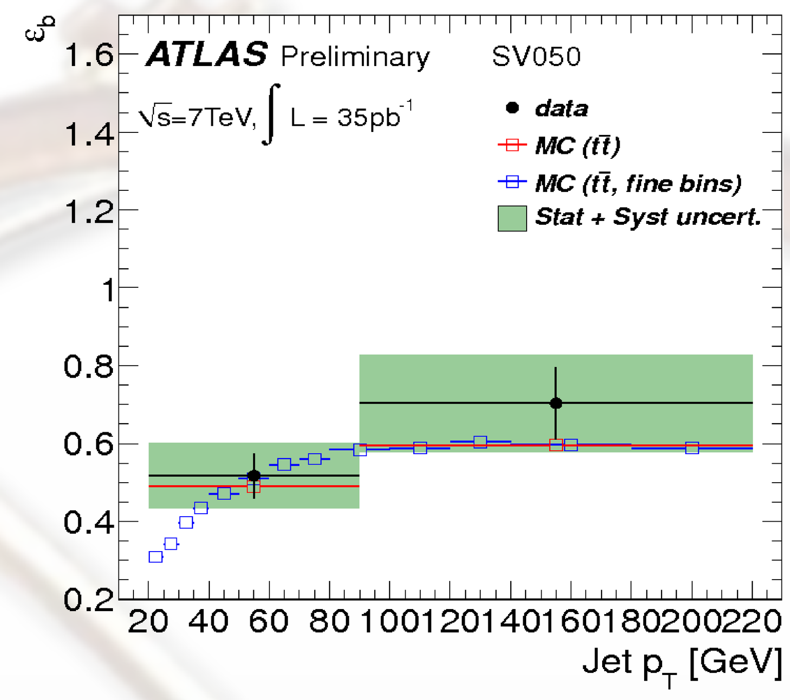
Kinematic selection method

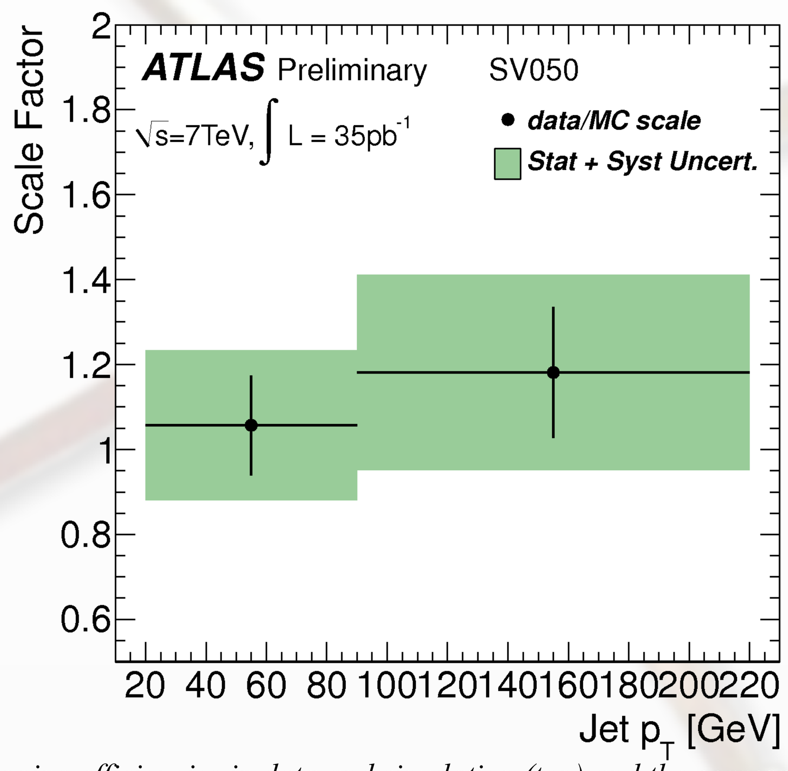
The b-tagging efficiency can be estimated using the equation

$$\varepsilon_b = \frac{1}{x_b}\left(x_{tag} - \varepsilon_c x_c - \varepsilon_{light} x_{light} - \varepsilon_{fake} x_{fake}\right)$$

where x_b, x_c and x_{light} are the fractions of b-, c- and light jets within the selected sample, while x_{fake} is the fraction of jets coming from the fake lepton background. The mistag efficiencies ε_c for charm and ε_{light} for light jets as well as the fractions of the various jet flavors are calculated using all the selected events in simulation, while the tagging efficiency of the jets coming from the fake lepton background ε_{fake} and the fraction of tagged jets x_{tag} are obtained from data.

While **in a dileptonic selection the b-jet purity is due to lower background contamination already up to 80 %**, one has to apply a **pre-tag in the lepton+jets selection to enrich the signal purity.** The tagging efficiency is determined in case of the lepton+jets **selection** by using the **leading jet** (if the second leading jet is tagged) or the **three subleading jets** (if the leading jet is b-tagged), while in **the dileptonic case only the two leading jets** are taken into account.



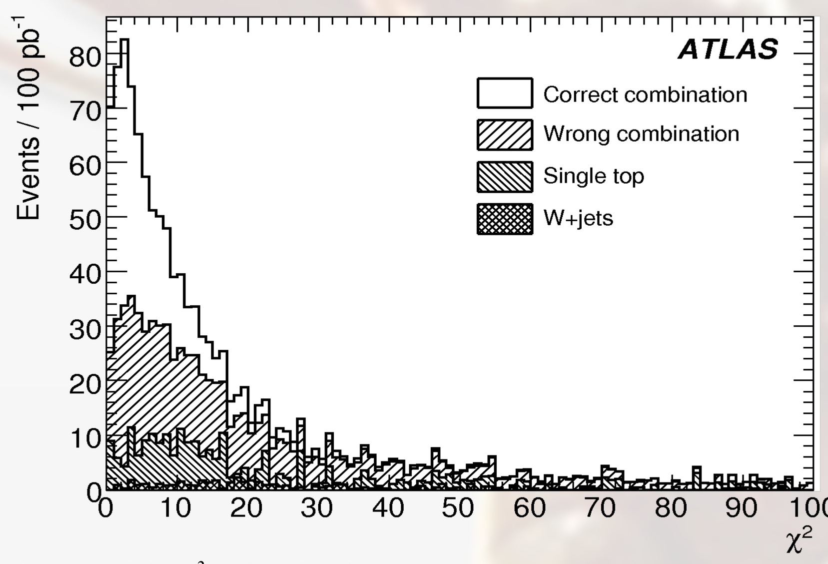


b-tagging efficiencies in data and simulation (top) and the corresponding scaling factors (bottom) for the SV0 tagging algorithm at 50 % signal efficiency [1].

Kinematic fit

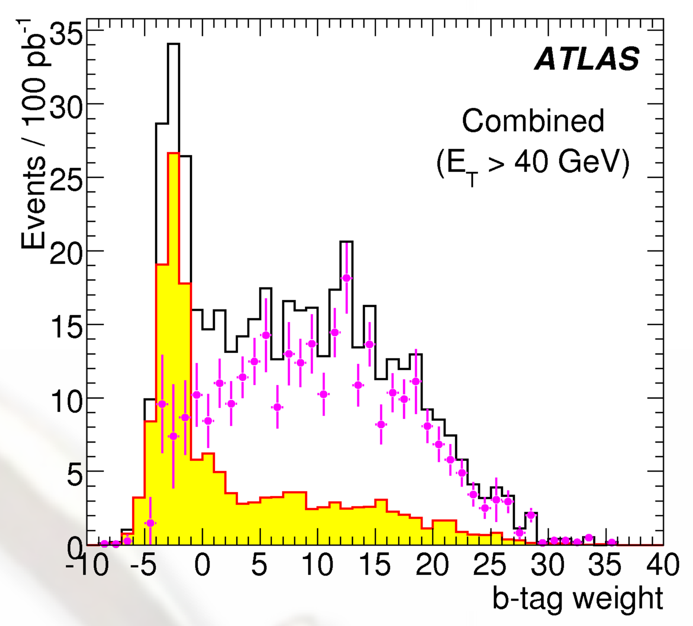
The kinematic fit method **takes advantage of the event structure of the semileptonic ttbar decay** by reconstructing both the hadronic and the leptonic top quark decay. **The fit assigns the leading jets in the event to originate either from the W-boson or the b-quarks**, and provides a probability measure that this is the correct assignment. When taking only the four leading jets into account 12 various permutations are possible. Only **the permutation with the lowest χ^2 is chosen to form the ttbar candidate.**

The **sample is futher purified by requiring that** the jet assumed to be **the b-jet on the hadronic side** of the event **is b-tagged, while the two jets assigned to the W-boson are required to not be b-tagged**. The **measurement of the b-tag efficiency is then performed on** the jet assigned to be **the b-jet on the leptonic side** of the event.



Distribution of the χ^2 value for the most important backgrounds, the correct and the wrong combinations [2]

This jet is however not always a b-jet due to the kinematic fit assigning the jets wrongly. The **fraction of wrong combinations in the signal sample is estimated using an orthogonal background sample**, where one of the jets associated to the W-boson is required to be b-tagged. The **number of events in the background sample with large fit χ^2 are normalized to the corresponding number of events in the signal sample**.The χ^2 **shape of the background sample is** then **used to estimate the** fraction of **background** in the low χ^2 region of the signal sample. The b-tag output weight distribution in **this background-subtracted signal sample is then used to derive the b-tag efficiency in data.**



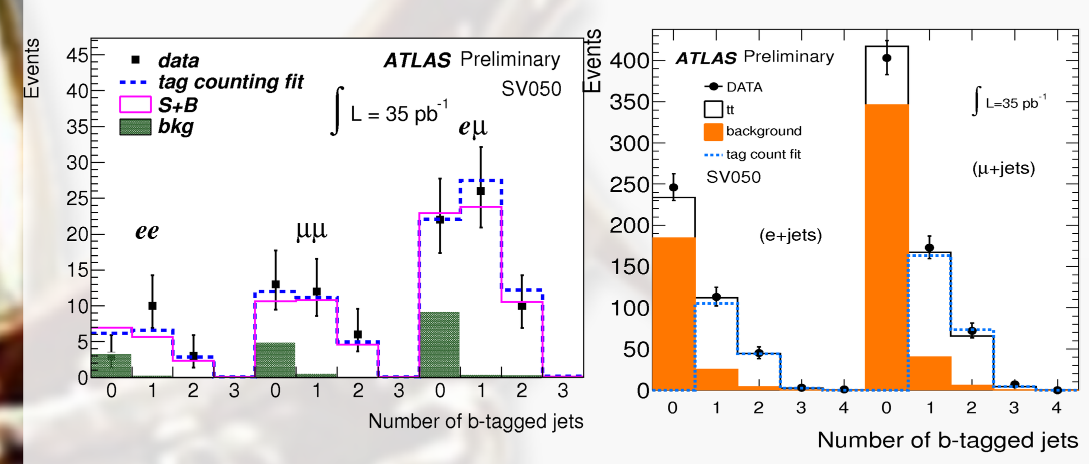
The b-tag weight distribution for the uncorrected sample (unfilled histogram), for the estimated background sample (filled histogram) and the corrected distribution calculated from the difference (data points) [3].

Tag counting method

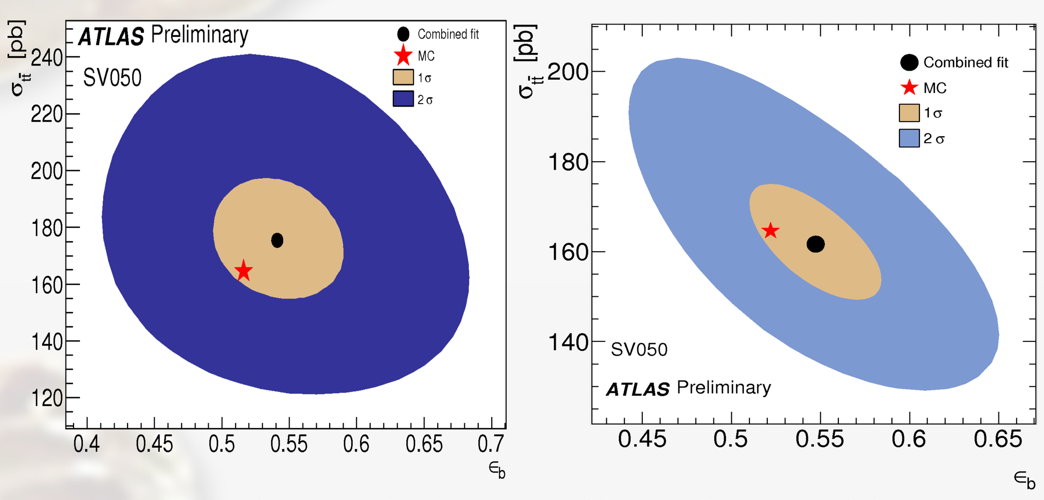
A Likelihood fit is performed to estimate the b-tagging efficiency by assuming that the expected number of events containing n b-tagged jets is given by

$$\langle N_{t\bar{t}}\rangle = \sum_{i,j,k}\left\{\left(\sigma_{t\bar{t}}\cdot BF\cdot A_{t\bar{t}}\cdot L\cdot F_{ijk}^{t\bar{t}} + N_{bkg}\cdot F_{ijk}^{bkg}\right) \times \sum_{i'+j'+k'=n} C_i^{i'}\cdot\varepsilon_b^{i'}\cdot\left(1-\varepsilon_b\right)^{i-i'}\cdot C_j^{j'}\cdot\varepsilon_c^{j'}\cdot\left(1-\varepsilon_c\right)^{j-j'}\cdot C_k^{k'}\cdot\varepsilon_b^{k'}\cdot\left(1-\varepsilon_{light}\right)^{k-k'}\right\}$$

where i, j and k are the number of b-, c- and light-flavor jets before applying b-tagging, while i', j' and k' represent the number of those jets after b-tagging. $C_\alpha^{\alpha'}$ is the number of permutations $\frac{\alpha!}{\alpha'!(\alpha-\alpha')!}$ with α=i,j,k for the three jet flavours. F_{ijk} is the fraction of events (before tagging) containing i b-jets, j c-jets and k light-flavour jets. BF is the branching ratio, $A_{t\bar{t}}$ is the selection acceptance and L is the integrated luminosity. The b-tag efficiency can then be determined by fitting this expected n-tag distribution to that observed in data.



Fitted b-tagged jet multiplicity distribution superimposed on the observed distribution in the dilepton (left) and lepton+jets (right) tag counting measurements [1].



2-dimensional contour for the measured b-tag efficiency and the top anti top cross section in the dilepton (left) and lepton+jets (right) tag counting measurements [1].

Resources

[1] The ATLAS Collaboration, "Calibrating the b-Tag Efficiency and Mistag Rate in 35 pb^{-1} of Data with the ATLAS Detector ", ATLAS-CONF-2011-089.

[2] The ATLAS Collaboration, "Expected performance of the ATLAS experiment: detector, trigger and physics", 2009.

[3] https://twiki.cern.ch/twiki/bin/view/AtlasPublic/FlavorTaggingPublicResults