Performance studies of *b*-tagging algorithms using $t\bar{t}$ processes in *pp* collisions

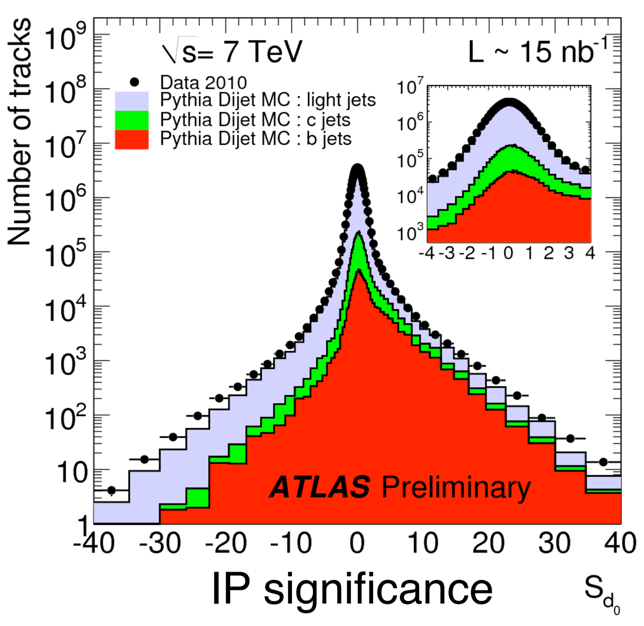
Kirika Uchida for ATLAS collaboration

universität bonn

ATLAS EXPERIMENT http://atlas.ch

***b*-tagging** is an identification of jets stemming from the hadronization of *b*-quarks (*b*-jets) making use of the distinctive properties; the long lifetime of *b*-hadron, the larger transverse momentum of their decay products due to the large mass of *b*-hadron, etc. Good tracking with the inner detector is crucial for the algorithms. Simple taggers are calibrated with early data taken in 2010 with 35 pb^{-1}.

- **JetProb** algorithm makes use of the impact parameters of the tracks associated with a jet. **Impact parameters (IP)** in *b*-jets distribute wider than the resolution function.
- **SV0** algorithm reconstructs **secondary vertices**. The longer lifetime of *b*-hadrons can produce secondary vertices.



Direct calibration in data is necessary because these variables highly depend on data taking conditions.

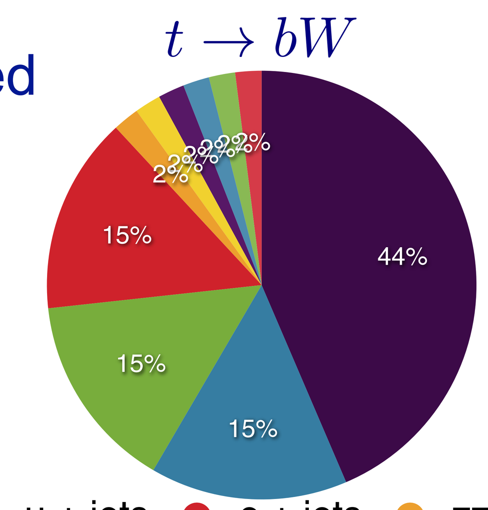
***b*-tagging calibration** has been done with QCD di-jet sample using the algorithms that break down for *b*-jets above 100 GeV. **$t\bar{t}$ process** provides calibrations in **higher p_T *b*-jet**.

A lot of $t\bar{t}$ events produced by LHC make it possible to do direct *b*-tagging calibrations for the high p_T *b*-jets in $t\bar{t}$ sample for the first time!

Top decay: $t \to bW,\ W \to q\bar{q}(2\text{ jets})\text{ or } l\nu(1\text{ lepton})$

Single lepton and dilepton channels are analyzed for the *b*-tagging performance study (lepton : *e* & *μ* including leptonic *τ* decay.)

$t \to bW$



all jets · τ + jets · μ + jets · e + jets · ττ · τμ · τe · μμ · μe · ee

Event topology

- Isolated leptons,
- at least two jets (*b*-jets),
- large Missing E$_T$.

b-tagging performance measurements

ATLAS-CONF-2011-089

Average *b*-jet tagging efficiency at 50 and 70 % for $t\bar{t}$ sample (p_T > 15 GeV) working points are calibrated for JetProb (**JetProb50** & **JetProb70**) and 50 % for SV0 (**SV050**.)

Tag counting method

Tag counting method counts the number of $t\bar{t}$ events with n *b*-tagged jets, <N_n> to simultaneously determine the top normalization and the tagging efficiency. Both in single lepton and dilepton channels are performed.

<N_n> for single lepton channel,

$$< N_n >= \sum_{i,j,k} \left\{ (\sigma_{t\bar{t}} \cdot BF \cdot A_{t\bar{t}} \cdot L \cdot F^{t\bar{t}}_{ijk} + N_{\rm bkg} \cdot F^{\rm bkg}_{ijk}) \times \sum_{i'+j'+k'=n} C_i^{i'} \cdot \epsilon_b^{i'} \cdot (1-\epsilon_b)^{i-i'} \cdot C_j^{j'} \cdot \epsilon_c^{j'} \cdot (1-\epsilon_c)^{j-j'} \cdot C_k^{k'} \cdot \epsilon_l^{k'} \cdot (1-\epsilon_l)^{k-k'} \right\}$$

- ***i, j,*** and ***k*** (***i', j', & k'***) are the number of pretagged (tagged) *b*-, *c*-, and light-flavour jets,
- $C^{\alpha'}_{\alpha}$ is the number of permutations with ***α*** for the three jet flavours.
- F_{ijk} is the fraction of pretagged events containing ***i*** *b*-jets, ***j*** *c*-jets, and ***k*** light-flavour jets, from simulation.
- ***BF*** is the branching fraction to each final state (*e*+jets and *μ*+jets),
- A_{tt} is the selection efficiency in this final state,
- ***L*** is the integrated luminosity.
- The efficiency to tag a *c*-jet and light-flavour jet, ε_c, and ε_l respectively, are fixed to the values found in simulation, corrected with data-to-simulation scale factors measured by other calibration.

Simultaneous fit of $t\bar{t}$ **cross section (σ_{tt})** and ***b*-tagging efficiency (ε_b)** with the likelihood function,

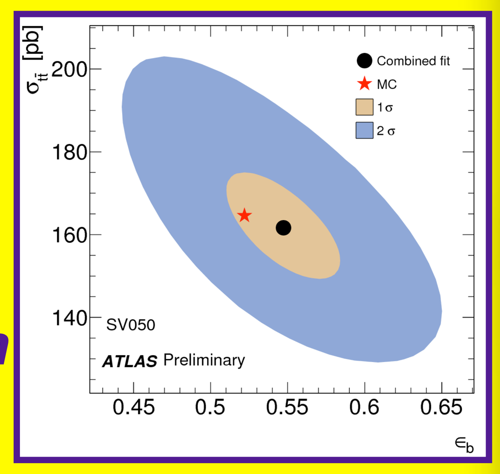
$$L = \prod_{n=1}^{3} \text{Pois}(N_n, < N_n >)$$

RESULT

single lepton

σ_{tt} uncertainty

Source	e+jets	μ+jets	ℓ+jets
Acceptance	+16.5/−13.3	+13.7/−10.1	+14.8/−10.8
ε_c	±0.6	±0.9	±0.8
ε_l	±0.8	±1.0	±0.9
W+jets normalisation	±4.8	±5.9	+5.5/−5.4
Other backgrounds normalisation	±2.2	±2.3	±2.3
W+jets flavour fractions	+6.5/−4.8	+7.9/−5.7	+7.3/−5.4
QCD flavour fractions	±5.0	±4.1	±4.4
Luminosity	±3.6	±3.6	±3.6
Total systematic	+22.0/−17.8	+19.2/−15.5	+20.0/−16.0
Statistical	±12.7	±10.1	±7.9



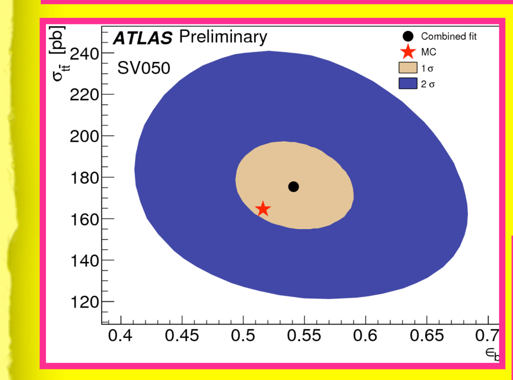
ε_b uncertainty

Source	SV050	JetProb50	JetProb70
ε_c	±1.0	±1.2	±1.3
ε_l	±0.1	<0.1	±0.1
W+jets background normalisation	±2.4	±3.1	+7.8/−6.7
Other backgrounds normalisation	±0.6	±0.6	±1.1
QCD flavour fractions	±2.8	±2.6	±4.3
W+jets flavour fractions	+2.2 /−2.7	+1.8/−2.3	+1.6/−1.9
Jet reconstruction	+0.8	+0.8	+0.8
Jet energy scale and resolution	+0.4/−0.6	+0.3/−0.6	+0.3/−0.6
Lepton trigger	<0.1	<0.1	<0.1
Lepton reconstruction and identification	<0.1	<0.1	<0.1
Lepton momentum scale and resolution	±0.1	±0.1	±0.1
Generator	−0.2	−0.4	−0.2
Parton shower	+0.5	+0.6	+0.6
PDF uncertainty	<0.1	<0.1	<0.1
ISR/FSR	+0.9/−1.0	+0.8/−0.7	+0.9/−0.8
Pileup	<0.1	<0.1	<0.1
Luminosity	<0.1	±0.1	±0.1
Total systematic	±4.8	±5.0	+9.3/−8.5
Statistical	±6.8	±6.8	±4.6

dilepton

ε_b uncertainty

Source	SV050	JetProb50	JetProb70
ε_c	±0.2	±0.2	±0.6
ε_l	±0.3	±0.5	±0.9
Z+jets normalisation	±3.5	±3.6	±2.7
Diboson normalisation	±0.7	±0.8	±0.7
Single top normalisation	±0.5	±0.5	±0.4
Fake lepton normalisation	±1.5	±1.3	±1.0
Z+jets flavour fractions	±0.4	±0.4	±0.3
Jet energy scale	±0.6	±0.8	±0.9
Generator	±0.4	±0.4	±0.4
Parton shower	±0.4	±0.5	±0.5
ISR/FSR	±1.4	±1.5	±1.5
Total systematic	±4.3	±4.4	±3.7
Statistical	±9.1	±8.8	±6.1



σ_{tt} uncertainty

Source	ee	$\mu\mu$	$e\mu$	$\ell\ell$
Acceptance	+13.1/−10.9	+8.8/−6.2	+8.0/−7.0	+10.5/−9.1
ε_c	±0.3	±0.2	±0.2	±0.2
ε_l	±0.3	±0.3	±0.3	±0.3
Z+jets normalisation	±4.6	±8.9	±2.9	±4.6
Diboson normalisation	±1.1	±1.3	±1.6	±1.5
Single top normalisation	±0.9	±0.8	±0.9	±0.9
Fake lepton normalisation	±3.5	±2.3	±4.3	±4.0
Z+heavy flavour fraction	±0.5	±1.1	±0.8	±0.8
Flavour composition other	±1.6	±1.3	±1.3	±1.4
Luminosity	±3.4	±3.4	±3.4	±3.4
Total systematic	+14.9/−13.0	+13.3/−11.8	+10.4/−9.6	+12.9/−11.7
Statistical	+31.7/−26.7	+25.0/−22.0	+16.8/−15.3	±12.5

Kinematic method

Kinematic method requires 1st or 2nd highest-p_T jet to be tagged to obtain high purity of $t\bar{t}$ sample and measure the tagging efficiency with other jets in the events. *b*-jet efficiency (ε_b) are measured in single lepton channel in 2 p_T bins.

$$\epsilon_b = \frac{1}{f_b} \times \left(\frac{N_{\rm tagged}}{N} - \epsilon_c f_c - \epsilon_l f_l - \epsilon_{\rm QCD} f_{\rm QCD} \right)$$

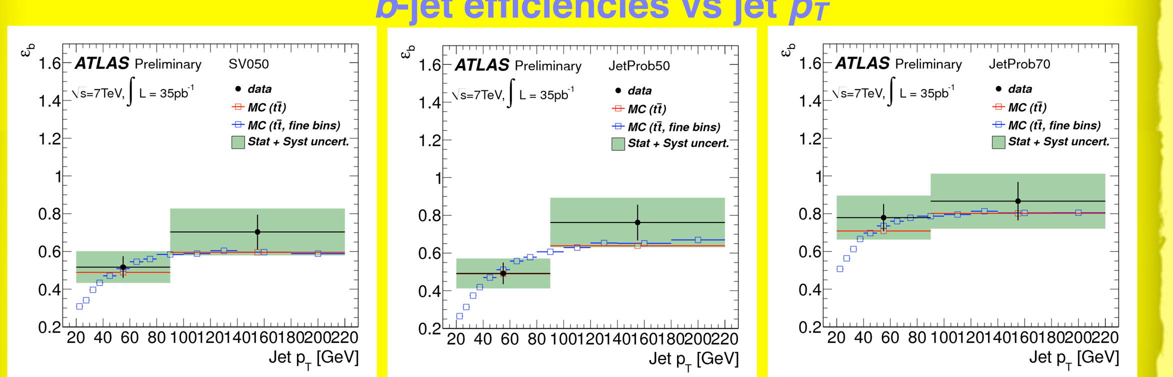
- ε_b / f_b, ε_c / f_c, and ε_l / f_l are the efficiencies/fractions of *b*-, *c*-, light-, which are estimated from MC except for ε_b.
- ε_{QCD} / f_{QCD} are the efficiency/fractions of QCD-jets, which are estimated from data.
- N/N$_{\text{tagged}}$ are the number of pretagged (tagged) jets

RESULT

ε_b uncertainty

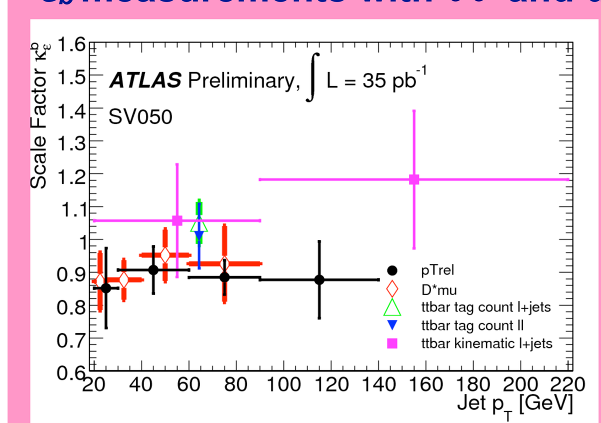
	SV050		JetProb50		JetProb70	
	Jet p_T [GeV]					
Source	20-90	90-220	20-90	90-220	20-90	90-220
ε_b SV050	±1.3	±1.7	±1.2	±1.8	±1.2	±1.7
ε_c SV050	±1.5	±2.3	±0.3	±0.2	±0.3	±0.2
ε_l SV050	±<0.1	±4.0	±0.5	±0.6	±0.5	±0.5
ε_c JetProb50/70	-	-	±1.8	±2.6	±2.8	±4.9
ε_l JetProb50/70	-	-	±1.5	±0.5	±2.8	±1.5
$t\bar{t}$ normalisation	±5.4	±5.2	±5.0	±5.4	±5.0	±5.1
W+jets normalisation	±2.8	±2.6	±2.8	±2.5	±2.7	±2.3
QCD normalisation	±5.8	±5.8	±4.7	±6.5	±5.0	±6.4
Z+jets normalisation	±0.1	±0.2	±0.1	±0.2	±0.1	±0.2
Single top normalisation	±0.1	±0.1	±0.1	±0.1	±0.1	±0.1
W+jets flavour fractions	±4.1	±3.4	±4.1	±3.4	±3.1	±1.9
QCD flavour fractions	±0.9	±0.7	±1.1	±0.4	±0.9	±0.8
Jet energy scale and resolution	±4.3	±0.3	±4.6	±0.5	±4.8	±0.1
Jet reconstruction	±1.2	±0.6	±1.2	±0.7	±1.3	±0.7
Lepton trigger	±1.3	±0.9	±1.3	±0.9	±1.3	±0.8
Lepton momentum scale and resolution	±1.9	±1.3	±1.9	±1.3	±3.0	±1.1
Lepton reconstruction and identification	±1.4	±0.9	±1.4	±0.9	±1.3	±0.8
Generator	±0.6	±2.3	±0.6	±2.2	±0.6	±2.0
Parton shower	±0.1	±0.5	±0.1	±0.5	±0.2	±0.6
ISR/FSR	±2.0	±4.6	±1.9	±4.5	±2.0	±4.1
Pileup	±1.3	±<0.1	±1.2	±0.4	±1.2	±0.5
Luminosity	±0.2	±0.2	±0.2	±0.2	±0.2	±0.2
Total systematic	±11.8	±11.9	±11.2	±11.9	±11.6	±12.1
Statistical	±11.1	±13.1	±11.6	±12.4	±9.3	±11.7

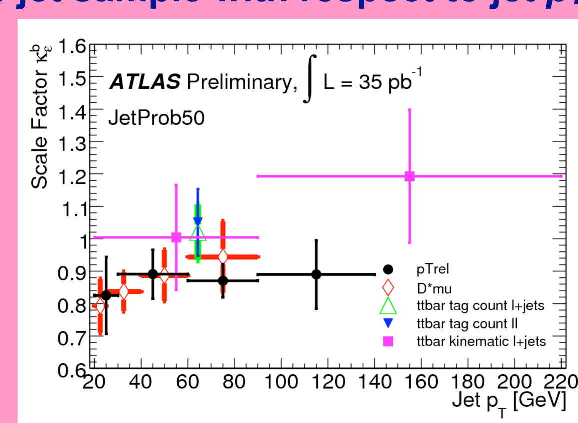
b-jet efficiencies vs jet p_T

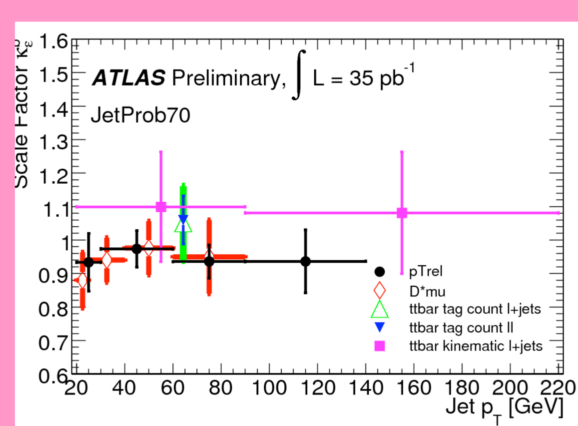


ε_b measurements in ATLAS

ε_b measurements with $t\bar{t}$ and di-jet sample with respect to jet p_T







The uncertainties in ε_b measurements with $t\bar{t}$ sample are larger than those with QCD sample.

Although, it has the advantages; direct measurements in higher *b*-jet p_T and higher jet multiplicity sample.

Statistical uncertainty will be smaller with much more sample collected by ATLAS.

Systematic uncertainty could also reduced by a better background estimation expected near future. Stay tuned!

International Europhysics Conference on High Energy Physics
Grenoble, Rhône-Alpes France July 21-27 2011
European Physical Society
HEP 2011

Deutsche Forschungsgemeinschaft DFG

Emmy Noether-Programm