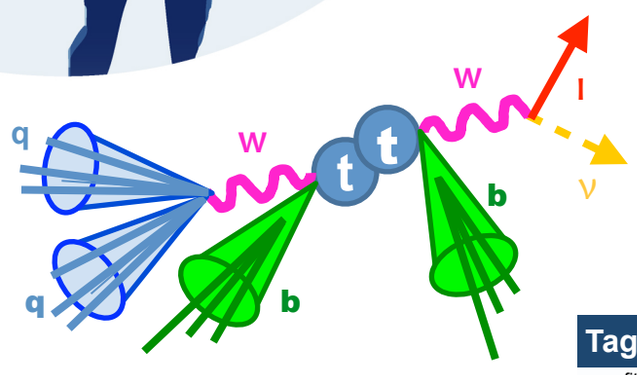
ICHEP 4-11 JULY 2012 MELBOURNE

Measuring *b*-tag efficiency in top quark pairs sample with 4.7 fb^{-1} of data from ATLAS detector



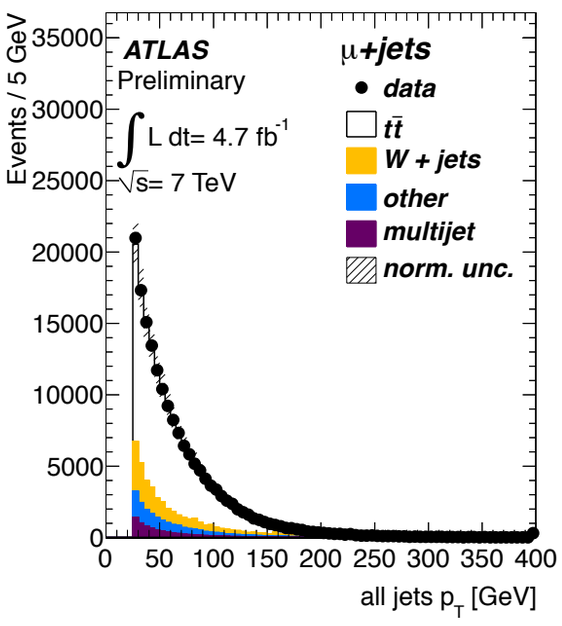
Introduction

- LHC is a top factory
- top decays ~100% to *Wb* ⇒ at least 2 *b*-jets in event
- clear event signature if at least 1 *W* boson decays leptonically
- in 4.7 fb^{-1} of data collected in 2011 ~0.5 million dilepton and single lepton events
- $t\bar{t}$ calibration methods nicely complement the currently used calibration with jets containing muons in dijet samples
- many physics analyses in ATLAS depend on the reconstruction of *b*-jets, often with high p_T
- $t\bar{t}$ methods are able to extend the *b*-tagging calibration to high p_T jets
- $t\bar{t}$ events are rich in *b*-jets and the final states (high p_T leptons, E_T^{miss} and multiple jets) are similar to other physics analyses of interest
- calibration of 4 *b*-tagging algorithms in all together 12 working points, including new multivariate algorithms MV1 making use of both impact parameters and secondary vertex information

Event Selection

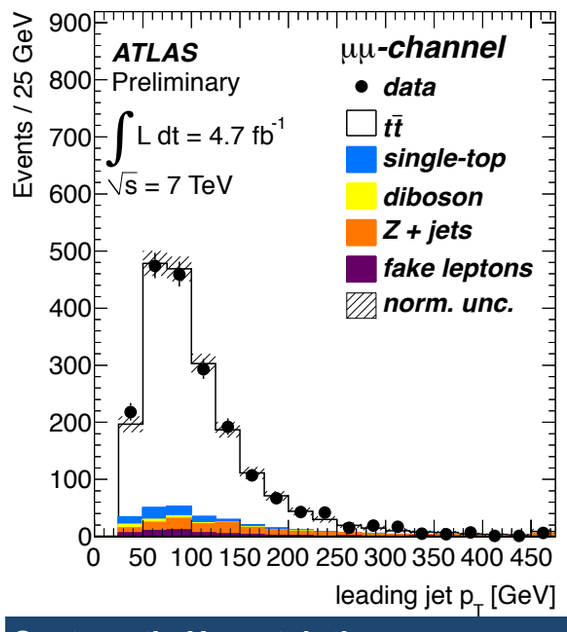
single lepton
- single *e* or single *μ* trigger
- exactly 1 isolated lepton with p_T > 25 GeV (*e*) or p_T > 20 GeV (*μ*)
- at least 4 jets with p_T > 25 GeV, |*η*| < 2.5
- *e*+jets: E_T^{miss} > 30 GeV, $m_T(l\nu)$ > 30 GeV
- *μ*+jets: E_T^{miss} > 20 GeV, E_T^{miss} + $m_T(l\nu)$ > 60 GeV

Source	*e*+jets			*μ*+jets		
$t\bar{t}$	17300	±	1700	28600	±	2800
W+jets	2800	±	400	5400	±	700
multijet	2300	±	1100	1800	±	400
single top	1430	±	110	2420	±	190
Z+jets	510	±	310	558	±	330
diboson	55.9	±	2.8	86	±	4
Σ	24500	±	2100	38900	±	2900
Observed 4.7 fb^{-1}	21978			38188		



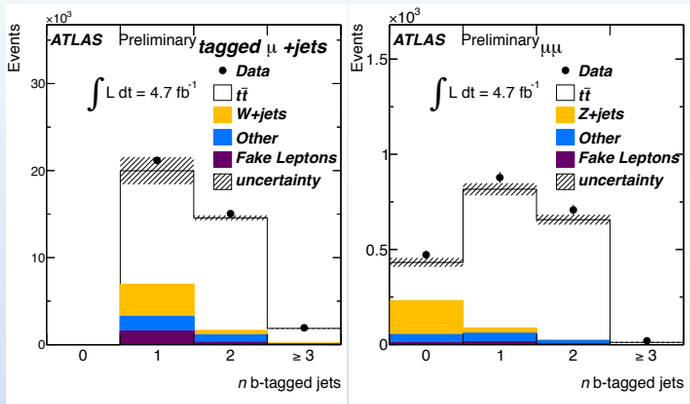
dilepton
- single *e* or single *μ* trigger
- exactly 2 oppositely charged isolated leptons with p_T > 25 GeV (*e*) or p_T > 20 GeV (*μ*)
- at least 2 jets with p_T > 25 GeV, |*η*| < 2.5
- *ee/μμ*: E_T^{miss} > 60 GeV, $|m_{ll} - m_Z|$ > 10 GeV and m_{ll} > 15 GeV
- *eμ*: H_T(leptons,jets) > 130 GeV

Source	N_{ee}	$N_{\mu\mu}$	$N_{e\mu}$
$t\bar{t}$	530 ± 50	1680 ± 170	4200 ± 400
$Z \to ee$ + jets	16 ± 6	–	–
$Z \to \mu\mu$ + jets	–	71 ± 28	–
$Z \to \tau\tau$ + jets	18 ± 7	70 ± 26	180 ± 70
diboson	8.4 ± 0.4	23.4 ± 1.2	67.2 ± 3.4
single top (*Wt*-channel)	26.8 ± 1.9	78 ± 6	204 ± 15
fake leptons	80 ± 40	43 ± 22	340 ± 170
∑ MC + fake leptons	680 ± 60	1970 ± 180	5000 ± 400
observed	716	1970	5341



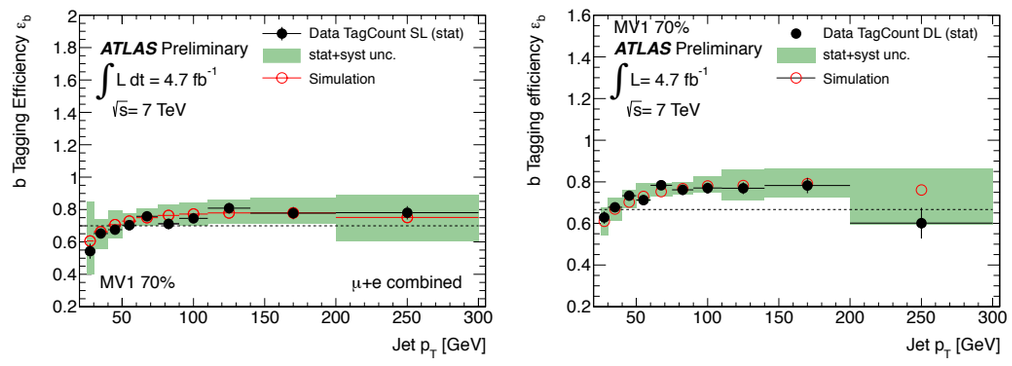
Tag Counting (single lepton & dilepton)

- fit observed *b*-tagged jet multiplicity distribution to expected using a binned maximum likelihood method
- floating parameters: top cross section, *b*-tagging efficiency
- BF - branching fraction, A_{tt} - event selection efficiency, $\mathcal{L}$ - int. luminosity
- ε_c, ε_l - mis-tag efficiencies for *c*- and light jets (from sim. corrected by the corresponding measured scale factors)
- F_{ijk} - fractions of events containing *i* *b*-jets, *j* *c*-jets and *k* light flavour jets (from sim.)



- single lepton: multijet background subtracted before fitting. Dominant background is *W*+jets. 0-tag bin is not considered in the measurement due to high background contribution
- dilepton: dominant background is *Z*+jets

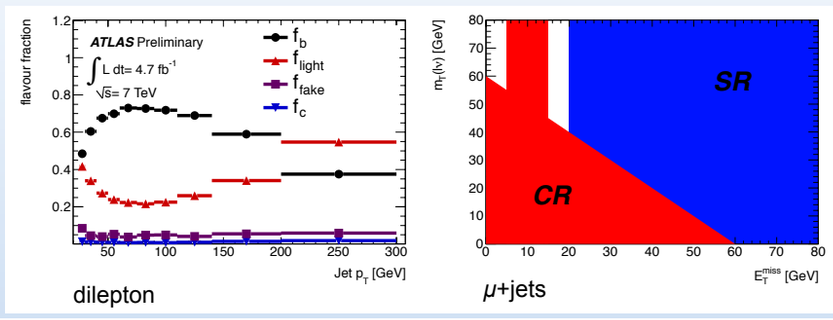
$$< N_n > \;=\; \sum_{i,j,k} \Big\{ (\sigma_{t\bar{t}} \cdot \mathrm{BF} \cdot A_{t\bar{t}} \cdot \mathcal{L} \cdot F_{ijk}^{t\bar{t}} + N_{\mathrm{bkg}} \cdot F_{ijk}^{\mathrm{bkg}}) \times \sum_{i'+j'+k'=n} \binom{i}{i'} \cdot \varepsilon_b^{\,i'} \cdot (1-\varepsilon_b)^{i-i'} \cdot \binom{j}{j'} \cdot \varepsilon_c^{\,j'} \cdot (1-\varepsilon_c)^{j-j'} \cdot \binom{k}{k'} \cdot \varepsilon_l^{\,k'} \cdot (1-\varepsilon_l)^{k-k'} \Big\}$$

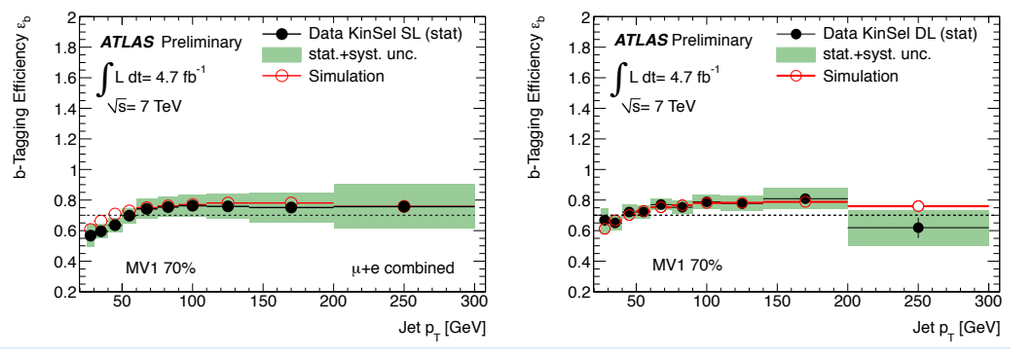


Kinematic Selection (single lepton & dilepton)

- check the tagging rate of jets in the selected events
  - single lepton: 4 jets with highest p_T
  - dilepton: 2 jets with highest p_T

$$\varepsilon_b = \frac{1}{f_{b-\mathrm{jets}}} \cdot \left(f_{b-\mathrm{tag}} - \varepsilon_c f_{c-\mathrm{jets}} - \varepsilon_l f_{l-\mathrm{jets}} - \varepsilon_{\mathrm{fake}} f_{\mathrm{fake}} \right).$$





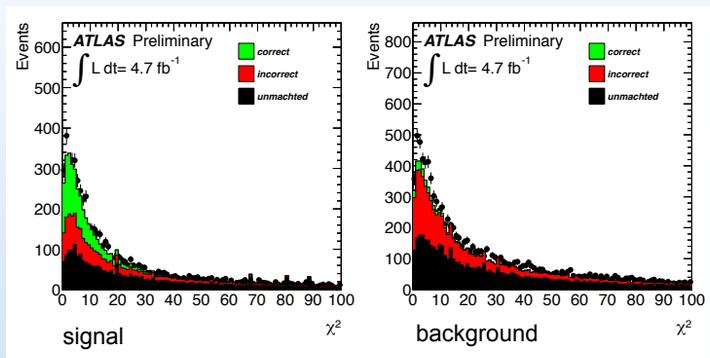
- $f_{b\text{-tag}}$ - fraction of *b*-tagged jets in data, fractions f_b, f_c, f_l and f_{fake} from a composition of simulated events and estimation from data for multijet (single lepton) or fake leptons (dilepton) and:

$$f_{b-\mathrm{jets}} + f_{c-\mathrm{jets}} + f_{l-\mathrm{jets}} + f_{\mathrm{fake}} = 1.$$

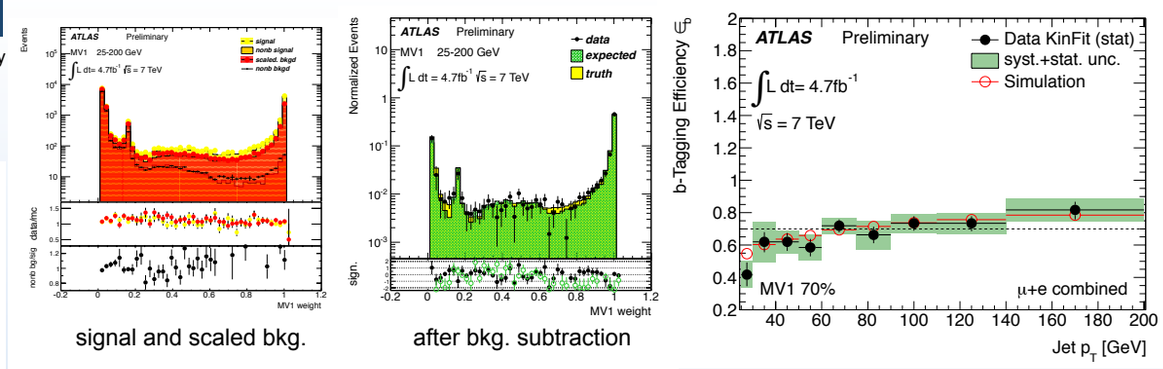
- ε_c and ε_l mis-tag efficiencies for *c*- and light jets (from sim. corrected by the corresponding measured scale factors)
- $\varepsilon_{\mathrm{fake}}$ - ratio of *b*-tagged jets to the total number of jets in the multijet (fake leptons) sample, estimated in data
  - single lepton: estimated in a control region with events after loose lepton (less strict isolation criteria) selection
  - dilepton: measured with events containing same sign leptons

Kinematic Fit (single lepton)

- select events using kinematic fit to obtain pure *b*-jet sample and measure efficiency from its *b*-tag weight distribution
- fit is based on a χ^2 minimisation (constraints from m_W and m_t)
- subtract non-b contamination by using an orthogonal background sample ⇒ no MC inputs



- test on simulated events
  - signal: high fraction of correct mappings
  - background: almost exclusively wrong mappings



signal:
- require *b*-jet on hadronic side to be *b*-tagged
- require the 2 jets from the *W* decay to not be *b*-tagged
- not more than 6 jets with p_T > 25 GeV

background:
- require 1 jet from *W* decay to be *b*-tagged
- no requirement on the maximum number of jets

- events with high values of χ^2 - wrong mappings in both samples
- background sample prediction is normalised at high χ^2 values (χ^2 > 25) by a scale factor

$$S_{\mathrm{BG}} = \frac{\int_{25}^{\infty} d\chi_S^2}{\int_{25}^{\infty} d\chi_B^2}$$

- b-tagging efficiency measured from background-subtracted distribution of the *b*-tag weight

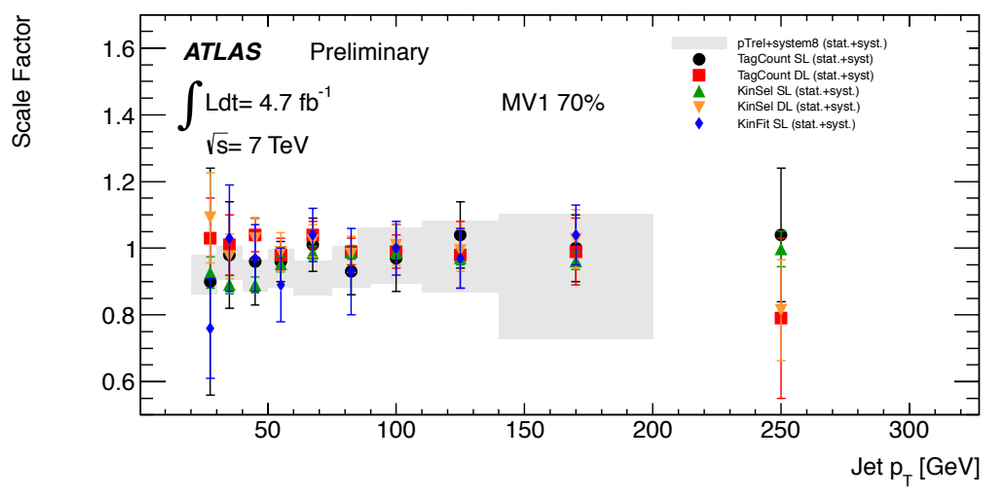
$$\varepsilon(w_c) = \int_{w_c}^{\infty} T(w)\, \mathrm{d}w.$$

Results

scale factors defined as

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

- $\varepsilon_b^{\mathrm{sim}}$ - fraction of *b*-jets which are tagged in simulated events
- pTrel+system8: combination of results from calibration methods using jets containing muons from dijet events (available only for jets with p_T < 200 GeV
- all methods show a good agreement with each other
- combination of results from $t\bar{t}$ methods with pTrel and system8 is in preparation
  - will reduce the *b*-tagging calibration uncertainty for many physics analyses



Systematic Uncertainties

varies from 5% to 20% depending on the p_T bin and method
systematic uncertainties from simulation modelling are the dominating ones for all methods
- modelling of initial/final state radiation
- choice of $t\bar{t}$ generator (MC@NLO vs. Powheg)
- Herwig vs. Pythia parton showering

significant systematic uncertainties also from modeling of jet properties
- jet energy scale, energy resolution, reconstruction efficiency, vertex fraction

significant contribution from the dominant background normalisation (estimated from data)
- *W*+jets (single lepton)
- multijet/fake leptons (including $\varepsilon_{\mathrm{fake}}$)

more details in ATLAS-COM-CONF-2012-089

Agnieszka Leyko, for the ATLAS Collaboration

ATLAS EXPERIMENT