Performance of *b*-jet identification in ATLAS

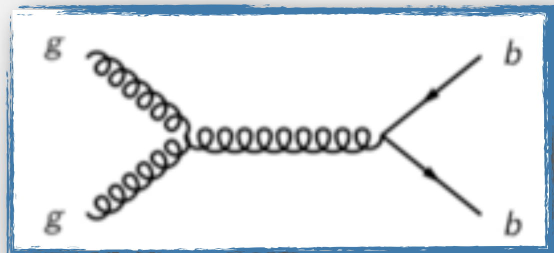
Jelena Jovićević
on behalf of the ATLAS collaboration

La Thuile
25 February, 2014

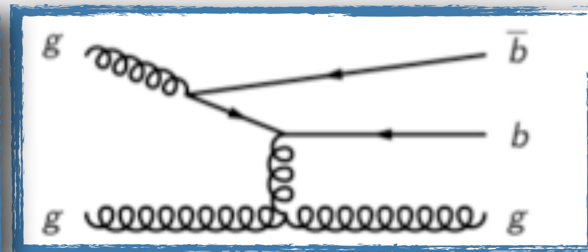
Production of b-quarks at LHC

Directly:

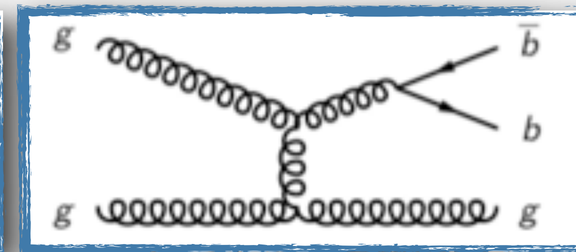
flavour creation



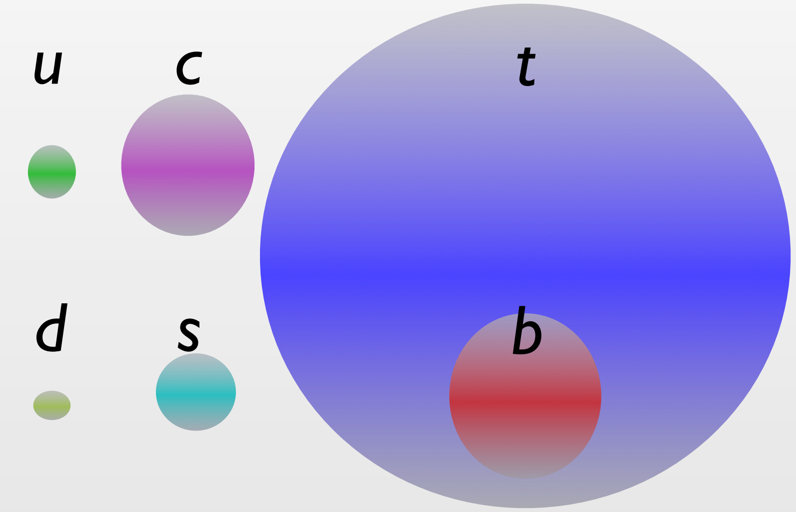
flavour excitation



gluon splitting



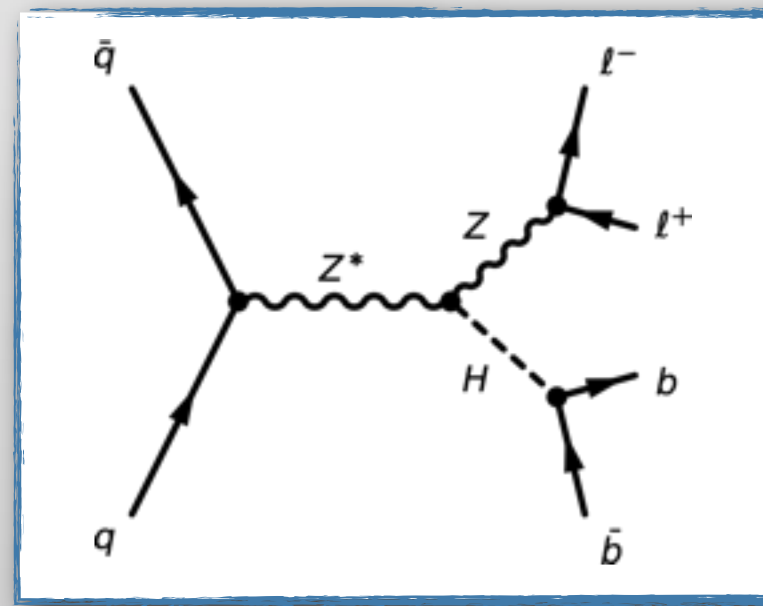
3 generations



In decays:

- top quark decays to *b* quark
  (e.g. important for precise measurements in top sector)
- Standard Model Higgs boson with m_H = 126 GeV, decays to $b\bar{b}$ pair 56% of the time.
  - ✤ Important channel to test the Higgs-to-fermion couplings
- SUSY particles can decay to third quark generation

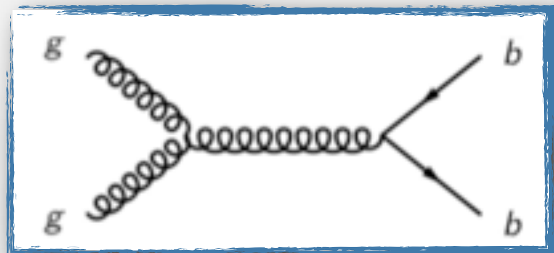
$$H(Z) \to b\bar{b}(\ell^+\ell^-)$$



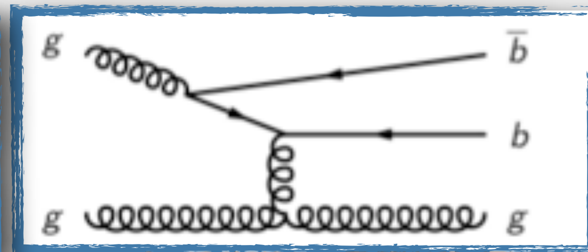
Production of b-quarks at LHC

Directly:

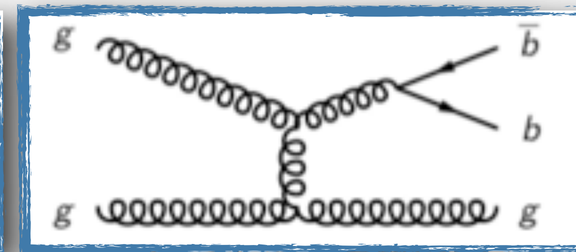
flavour creation



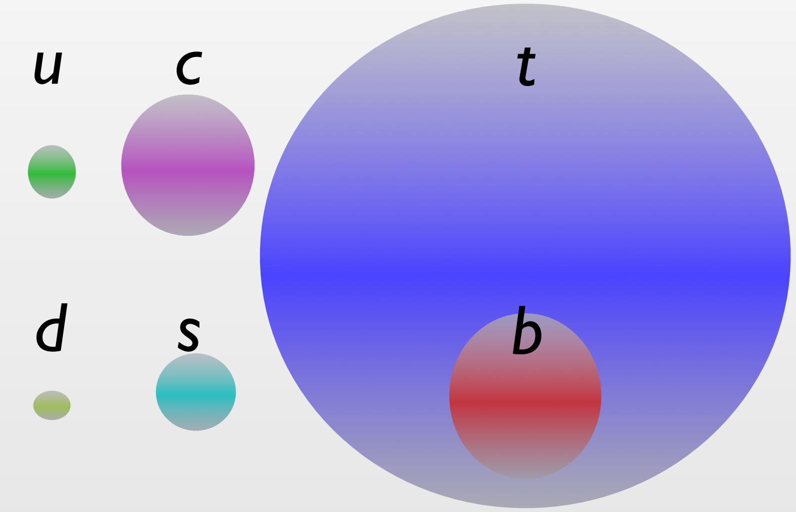
flavour excitation



gluon splitting



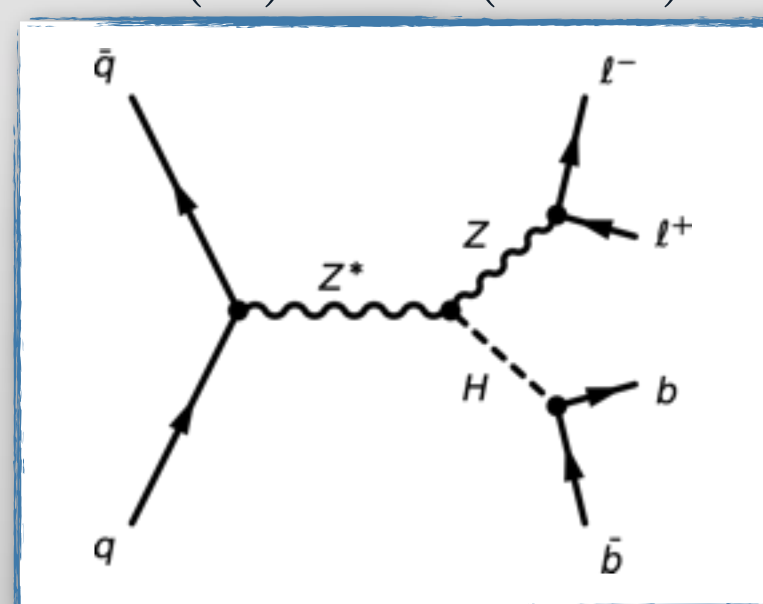
3 generations



In decays:

- top quark decays to *b* quark (e.g. important for precise measurements in top sector)
- Standard Model Higgs boson with m_H = 126 GeV, decays to $b\bar{b}$ pair 56% of the time.
  - ✤ Important channel to test the Higgs-to-fermion couplings
- SUSY particles can decay to third quark generation

$$H(Z) \rightarrow b\bar{b}(\ell^+\ell^-)$$



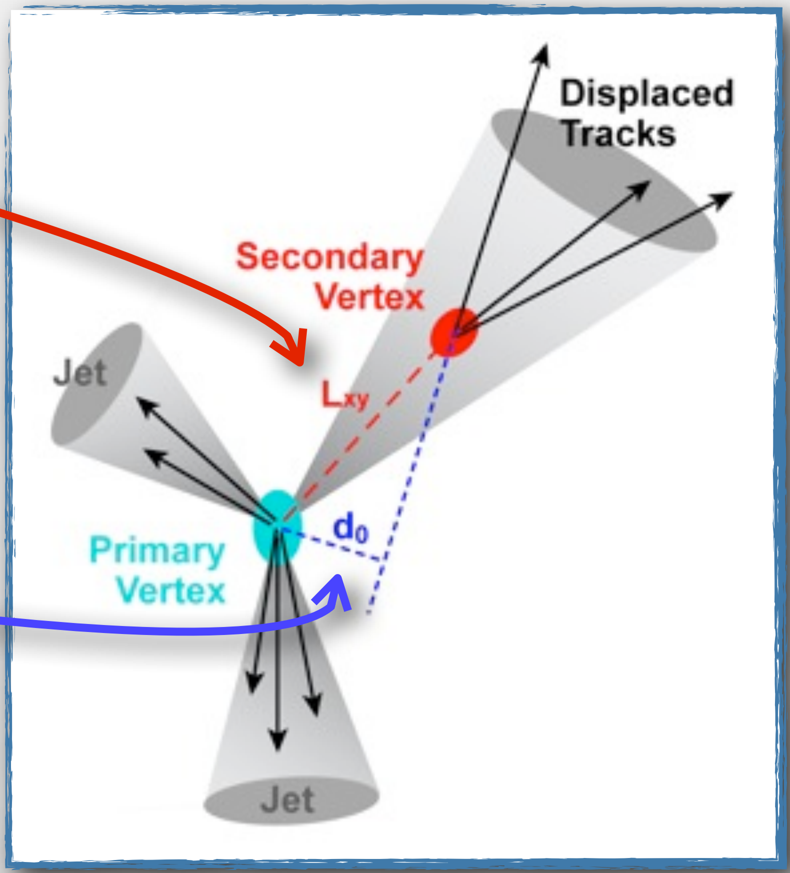
The b-quark opens a window for important physics measurements!

Signatures of b-quarks

Once produced, *b*-quarks hadronise forming B-hadrons inside jets

B-hadron characteristics:

- Large mass (typically 5 - 6 GeV)
- Long life time ~1.5 ps and large decay length (L_{xy})
  ➡ presence of a secondary vertex
- Secondary vertex generates displaced tracks
  ➡ large impact parameter (d_0)
- A chance of semi-leptonic decay
  ➡ nearby soft lepton



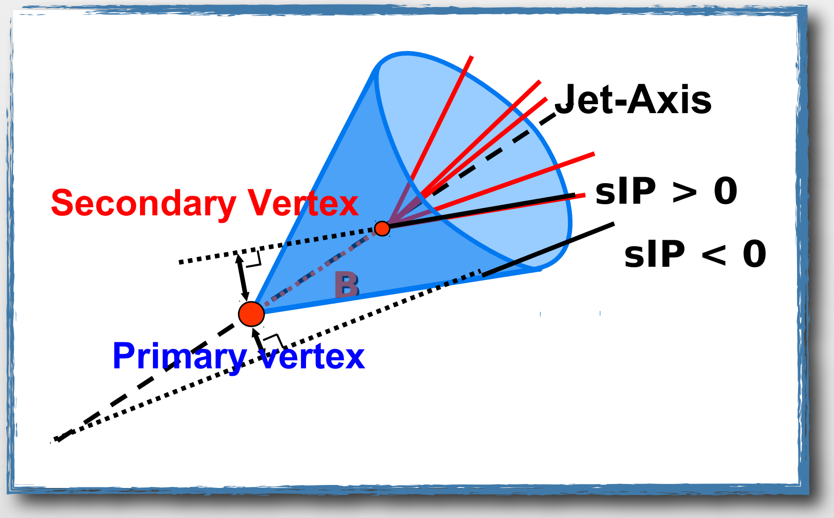
➡ **ATLAS is able to exploit all these characteristics!**
(thanks to its excellent tracking and vertexing performance)

The b - jet identification algorithms

Primary vertex selection: vertex with highest $\sum_{tracks} p_T^2$ - good pile-up rejection

Impact parameter (IP) based: **IP3D**

- uses **transverse and longitudinal IP significances** **$d_0/\sigma(d_0)$ & $z_0/\sigma(z_0)$** to discriminate between **b jets** and **light-jets**
- considers the "relative sign" of the IP ($d_0 > 0$ for b-jets)



Secondary vertex based: **SV0, SV1**

- aims to reconstruct **displaced vertices**
- exploits track-based **invariant mass** and **flight length significance**
- small mis-tag rate, limited efficiency

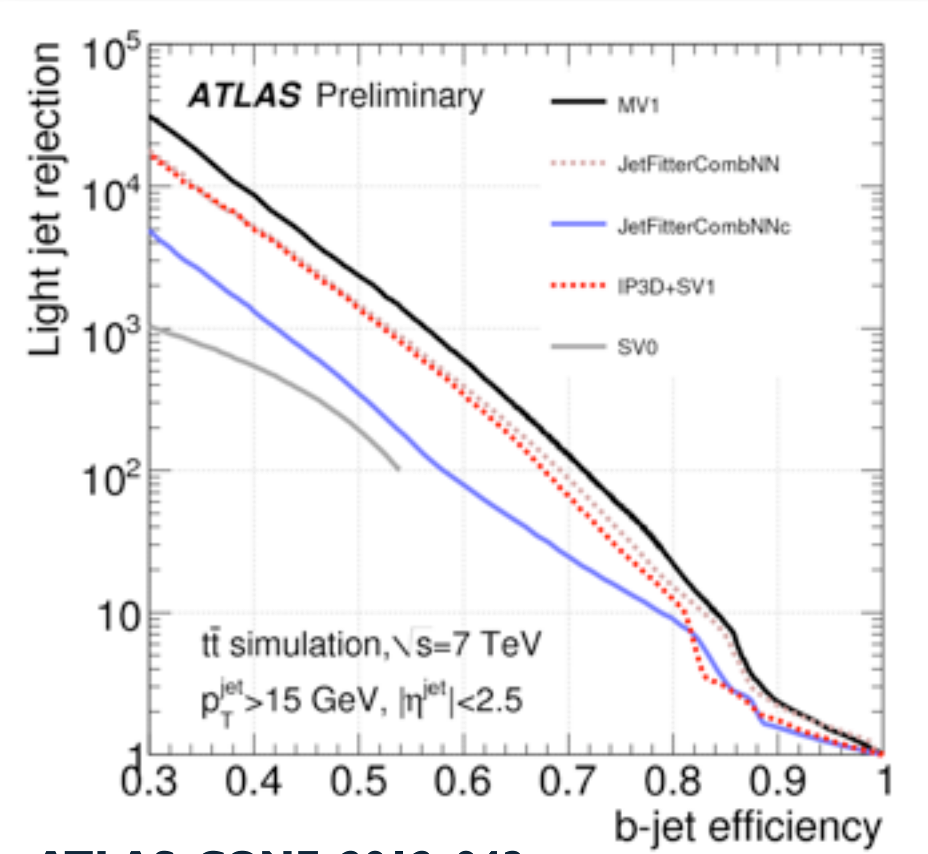
Decay chain reconstruction based: **JetFitter**

- aims at reconstructing **full hadron decay chain** (from b/c quarks)
- takes into account **track & vertices** info, **fraction of the energy** carried by charged particles within the jet, **flight length significance** in a neural net
- separate outputs for b, c and light jets

Combined algorithm - MV1

Combine individual algorithms to achieve higher rejection of light quark jets and to cover a wider range of b-tagging efficiencies !

Light jet rejection vs. b-jet efficiency



ATLAS-CONF-2012-043

MV1 algorithm:

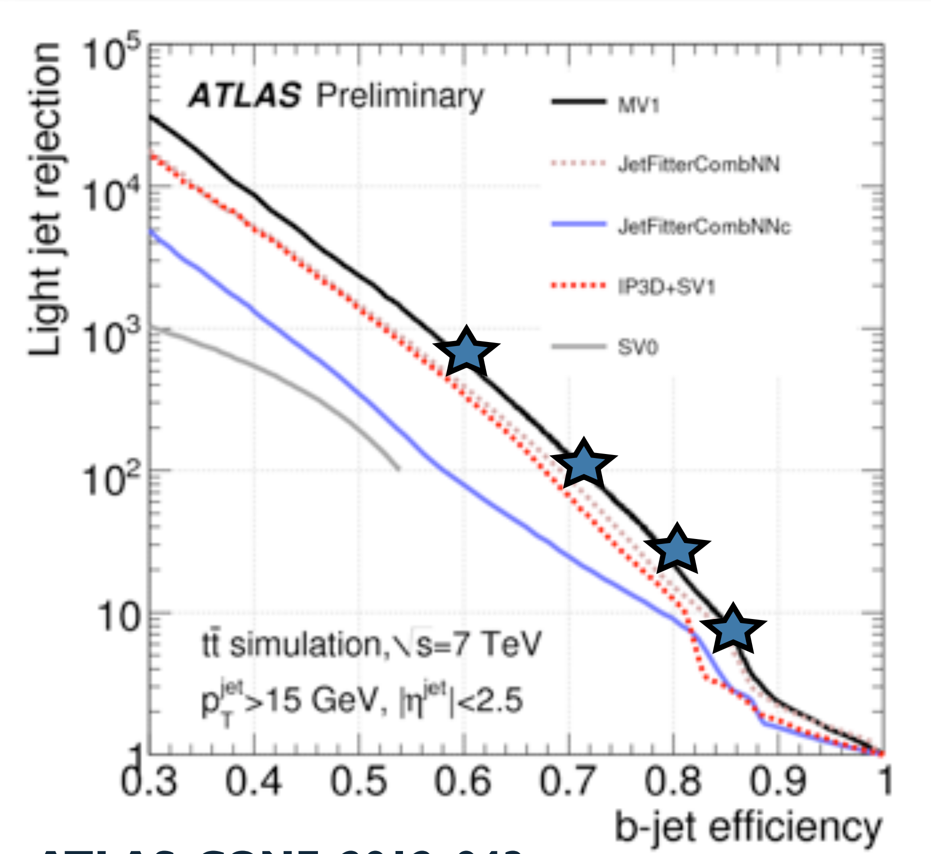
- multivariate technique, based on inputs from other 3 algorithms: IP3D, SV1 and IP3D+JetFitter
- takes into account input correlations
- output for b, c and light jets (p_T & η dependent)
- most commonly used in ATLAS

Provides the best rejection of light flavour jets for a given b-jet efficiency

Combined algorithm - MV1

Combine individual algorithms to achieve higher rejection of light quark jets and to cover a wider range of b-tagging efficiencies !

Light jet rejection vs. b-jet efficiency



ATLAS-CONF-2012-043

MV1 algorithm:

- multivariate technique,
  based on inputs from other 3 algorithms:
  IP3D, SV1 and IP3D+JetFitter
- takes into account input correlations
- output for b, c and light jets (p_T & η dependent)
- most commonly used in ATLAS

Provides the best rejection of light flavour jets for a given b-jet efficiency

Efficiency needs to be evaluated in data for reliable usage in physics analyses!

★ We calibrate several operating points of the inclusive b-tagging efficiency of MV1

Calibration techniques

For the light jet efficiency: **2 methods**

- inclusive jet samples (method based on symmetries of track resolution function and vertex mass based method) https://cds.cern.ch/record/1435194

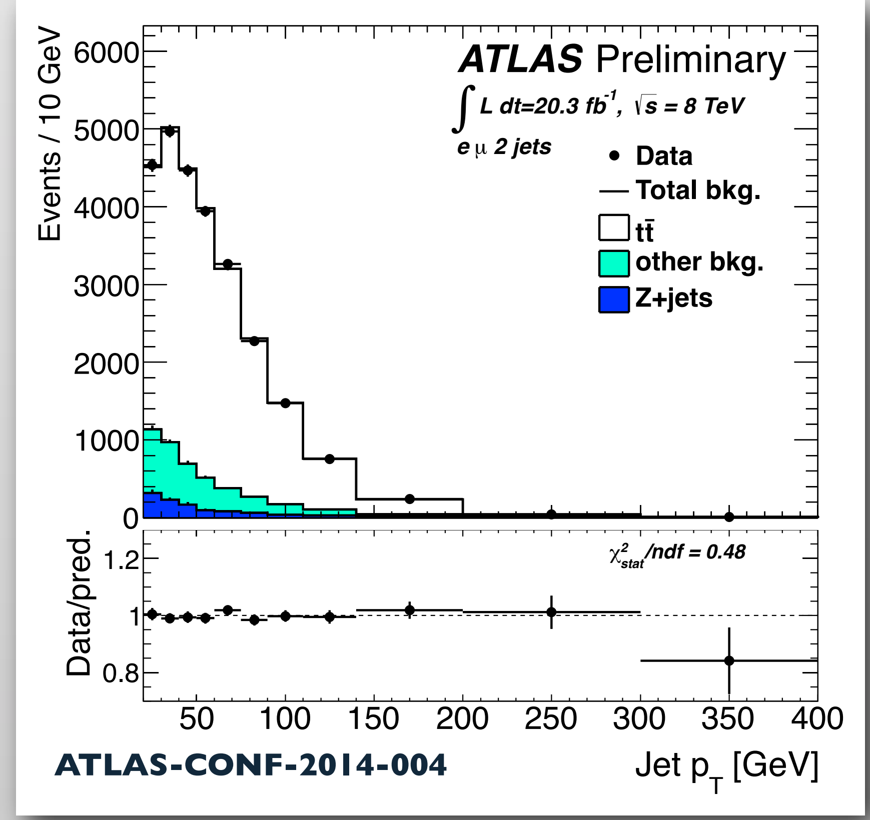
For the c-jet efficiency: **2 methods**

- in sample with D^* mesons https://cds.cern.ch/record/1435193
- W+c samples https://cds.cern.ch/record/1640162

For the b-jet tagging efficiency: **6 methods**

- in multijet samples with muons - 2 methods https://cds.cern.ch/record/1435197
- **in $t\bar{t}$ events** - 4 methods https://cds.cern.ch/record/1460443
  - relatively pure source of b assuming BR($t \rightarrow Wb$) = 1
  - purity in final states
    eμ + 2/3 jets - 73/54% pure in *b*
    ee + μμ +2/3 jets - 67/52% pure in *b*

Jet p_T distribution in the $t\bar{t}$ enriched sample



PDF based calibration method

- **New** $t\bar{t}$ based method for the b-jet efficiency calibration.
- Make better use of data by **exploiting per event jet flavour correlation.**
  - ✤ Allows b-tagging efficiency for a cut on the weight (w) to be measured to a high precision
- Model the system using likelihood - employs PDFs - $\mathcal{P}$ (2 jet example)

$$\begin{aligned}\mathcal{L}\left(p_{\mathrm{T},1}, p_{\mathrm{T},2}, w_1, w_2\right) = \quad & [\quad f_{bb}\mathcal{P}_{bb}\left(p_{\mathrm{T},1}, p_{\mathrm{T},2}\right)\mathcal{P}_b\left(w_1|p_{\mathrm{T},1}\right)\mathcal{P}_b\left(w_2|p_{\mathrm{T},2}\right) \\ & + \quad f_{bj}\mathcal{P}_{bj}\left(p_{\mathrm{T},1}, p_{\mathrm{T},2}\right)\mathcal{P}_b\left(w_1|p_{\mathrm{T},1}\right)\mathcal{P}_j\left(w_2|p_{\mathrm{T},2}\right) \\ & + \quad f_{jj}\mathcal{P}_{jj}\left(p_{\mathrm{T},1}, p_{\mathrm{T},2}\right)\mathcal{P}_j\left(w_1|p_{\mathrm{T},1}\right)\mathcal{P}_j\left(w_2|p_{\mathrm{T},2}\right) \\ & + \quad 1 \leftrightarrow 2 \;]/2,\end{aligned}$$

$\mathcal{P}_{ff}(p_{T,1}, p_{T,2})$ **MC** — 2D PDF for [$p_{T,1}$, $p_{T,2}$] for flavour combination *bj, bb, jj*

$\mathcal{P}_b(w, p_T)$ **Measured on data**

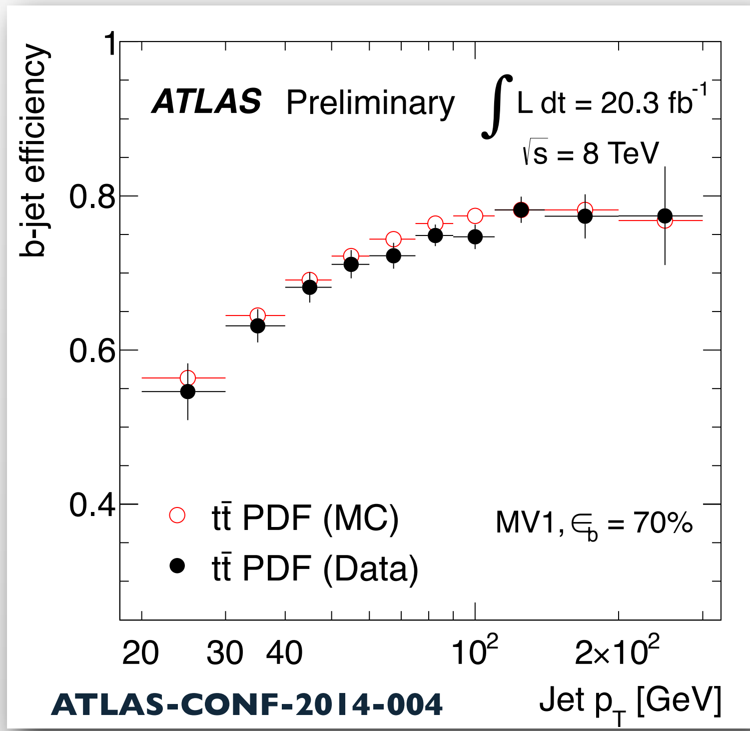
$\mathcal{P}_j(w, p_T)$ **MC** — PDF for the b-tagging discriminant for b(j) jet

$f_{bb},\ f_{bj},\ f_{jj} = 1 - f_{bb} - f_{bj}$ **MC** — flavour fractions in 2 jet case

- Efficiency determination $\epsilon_b(p_T) = \int_{w_{cut}}^{\infty} dw' \mathcal{P}_b(w', p_T)$
- Perform statistical combination of channels - reduced uncertainties

Results - PDF based calibration

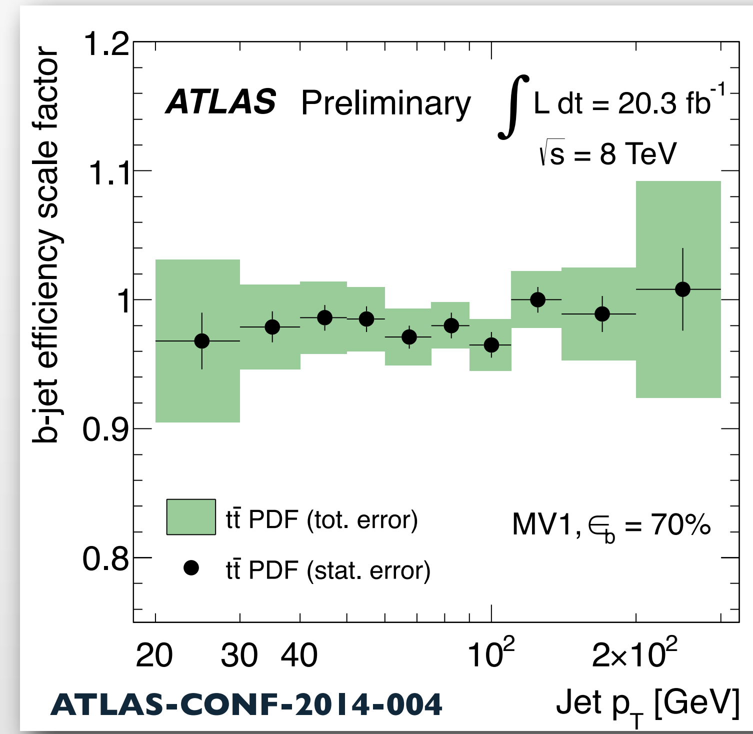
b-jet efficiency vs. jet p_T



Data to Monte Carlo correction factors

Calibration

Data/MC correction factors



Correction factors applied in physics analyses to account for mismodeling of the b-jet identification efficiency.

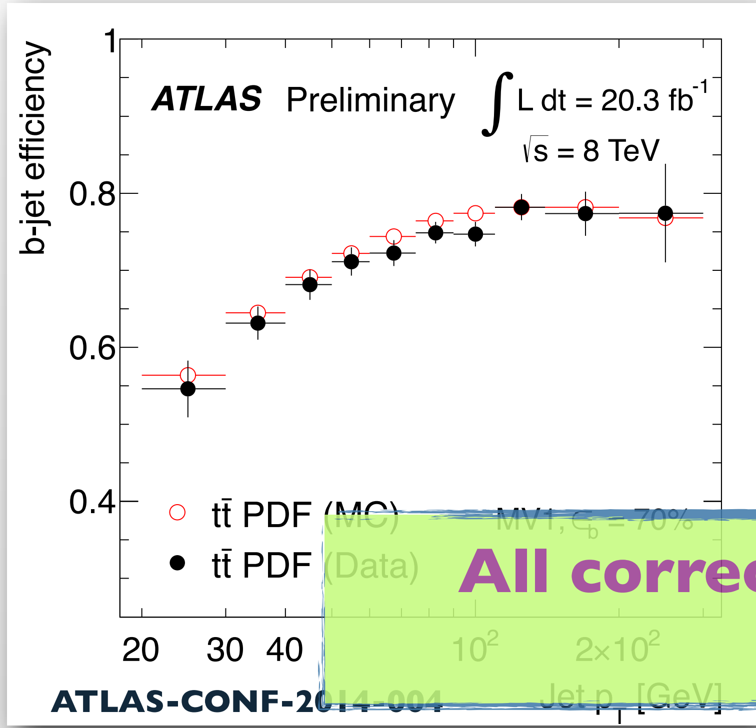
p_T interval	Corr. Factor	Stat error	Syst error	Tot error
[20; 30]	0.968	0.022	0.059	0.063
[30; 40]	0.979	0.012	0.030	0.033
[40; 50]	0.986	0.010	0.027	0.028
[50; 60]	0.985	0.010	0.023	0.025
[60; 75]	0.971	0.009	0.020	0.022
[75; 90]	0.980	0.010	0.015	0.018
[90; 110]	0.965	0.010	0.018	0.020
[110; 140]	1.000	0.010	0.020	0.022
[140; 200]	0.989	0.014	0.033	0.036
[200; 300]	1.008	0.032	0.077	0.084

Dominant systematics

Theory %()	
Hadronisation ($t\bar{t}$)	0.3 - 2.0
Modeling ($t\bar{t}$)	0.4 - 1.7
Modeling PS ($t\bar{t}$)	0.5 - 1.9
Top p_T reweighting ($t\bar{t}$)	0.2 - 4.6
Modeling Z+jets	0.2 - 2.4
Modeling diboson	0.7 - 3.1
Z+jets normalisation	0.4 - 1.7
Experimental (%)	
Jet energy scale	0.3 - 4.1
Jet energy resolution	0.1 - 2.6
Mistag rate	0.3 - 2.8

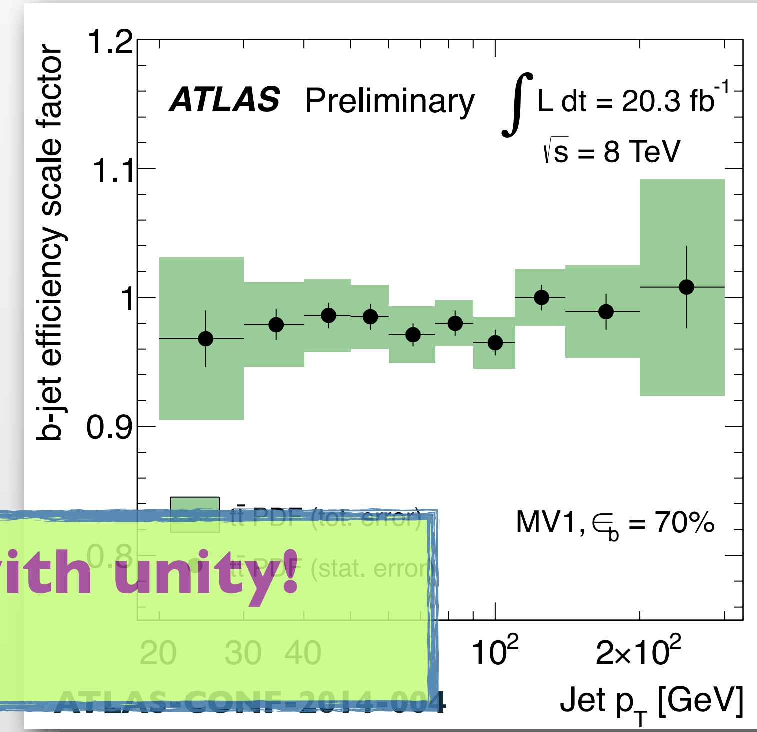
Results - PDF based calibration

b-jet efficiency vs. jet p_T



Data to Monte Carlo correction factors

Data/MC correction factors



All correction factors consistent with unity!
good modeling!

Correction factors applied in physics analyses to account for mismodeling of the b-jet identification efficiency.

p_T interval	Corr. Factor	Stat error	Syst error	Tot error
[20; 30]	0.968	0.022	0.059	0.063
[30; 40]	0.979	0.012	0.030	0.033
[40; 50]	0.986	0.010	0.027	0.028
[50; 60]	0.985	0.010	0.023	0.025
[60; 75]	0.971	0.009	0.020	0.022
[75; 90]	0.980	0.010	0.015	0.018
[90; 110]	0.965	0.010	0.018	0.020
[110; 140]	1.000	0.010	0.020	0.022
[140; 200]	0.989	0.014	0.033	0.036
[200; 300]	1.008	0.032	0.077	0.084

Statistical uncertainties very small - efficiently using data! Systematically limited.

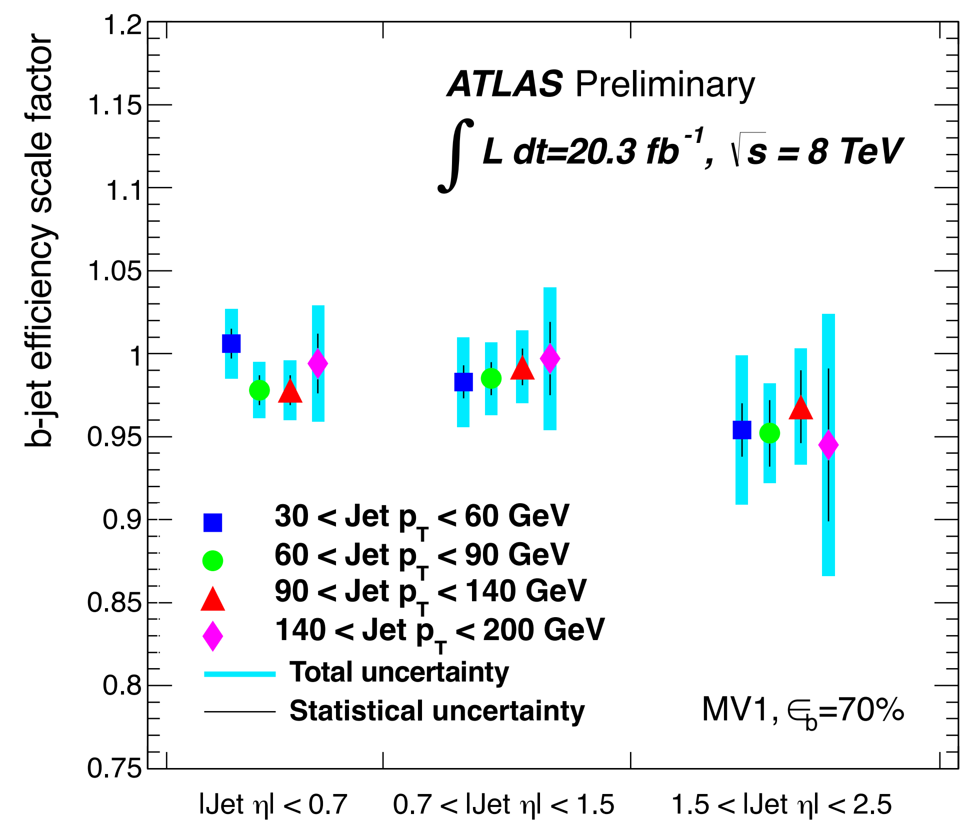
Dominant systematics

Theory %()	
Hadronisation ($t\bar{t}$)	0.3 - 2.0
Modeling ($t\bar{t}$)	0.4 - 1.7
Modeling PS ($t\bar{t}$)	0.5 - 1.9
Top p_T reweighting ($t\bar{t}$)	0.2 - 4.6
Modeling Z+jets	0.2 - 2.4
Modeling diboson	0.7 - 3.1
Z+jets normalisation	0.4 - 1.7
Experimental (%)	
Jet energy scale	0.3 - 4.1
Jet energy resolution	0.1 - 2.6
Mistag rate	0.3 - 2.8

Results - PDF based calibration

- Method allows for an arbitrary binning in any jet kinematic quantity
- Performed binning in both p_T and η - to verify no η dependence from the less sensitive previously used calibrations

Data/MC correction factors as a function of the jet pseudorapidity

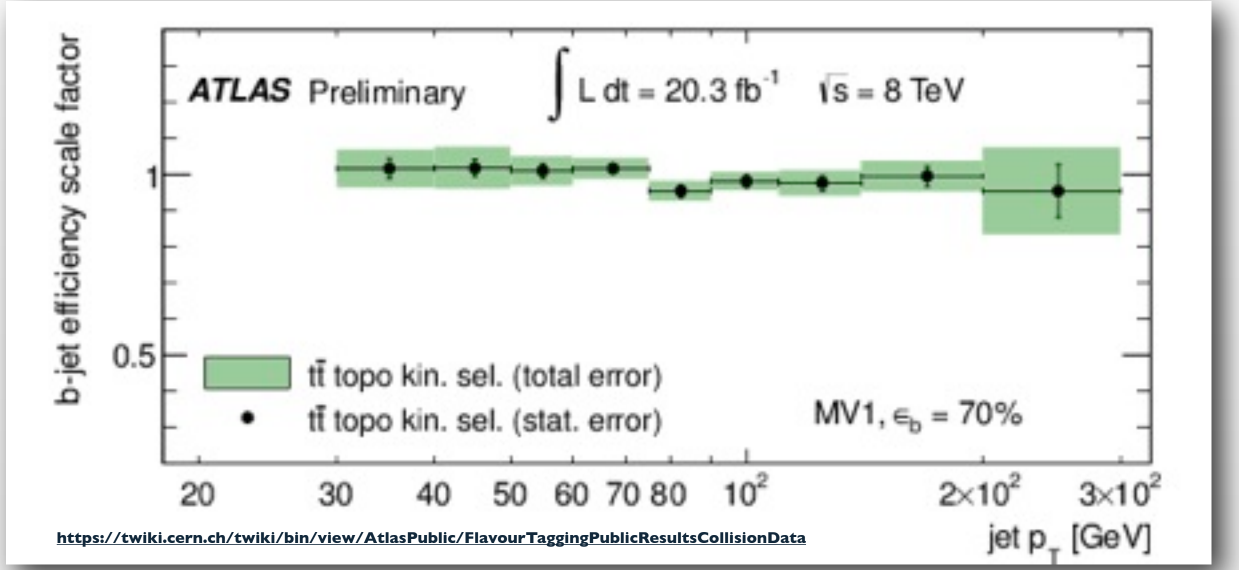


ATLAS-CONF-2014-004

- The scale factors were tested as a function of η inclusively in p_T using the $\chi 2$ test and no significant dependence is observed.

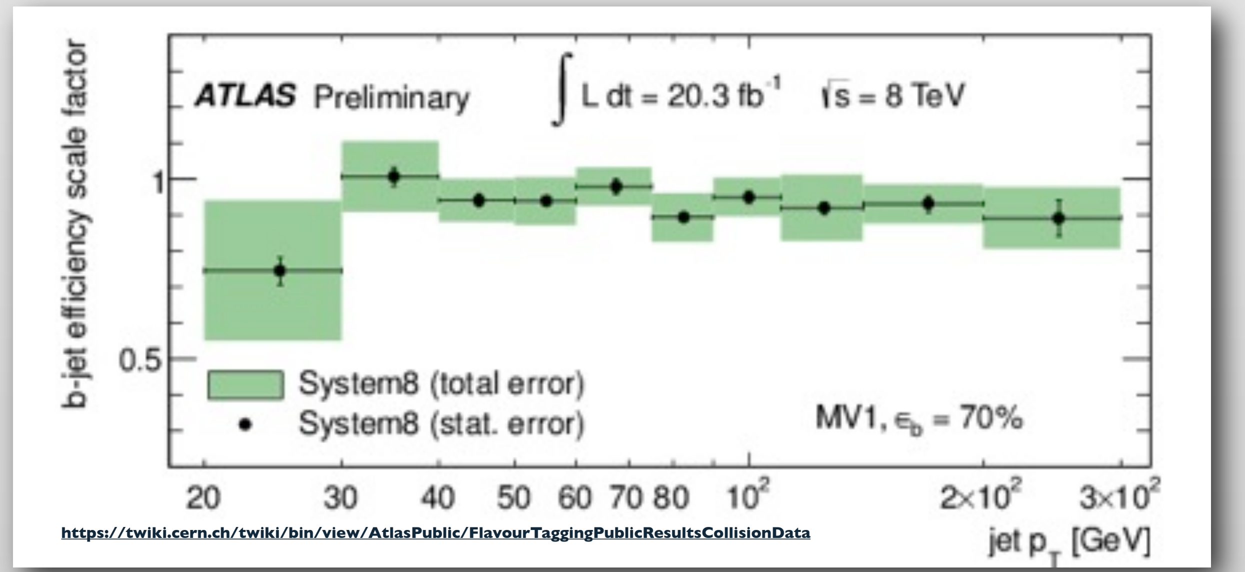
Improvement with PDF based method

Performance of the tag and probe method in $t\bar{t}$ event



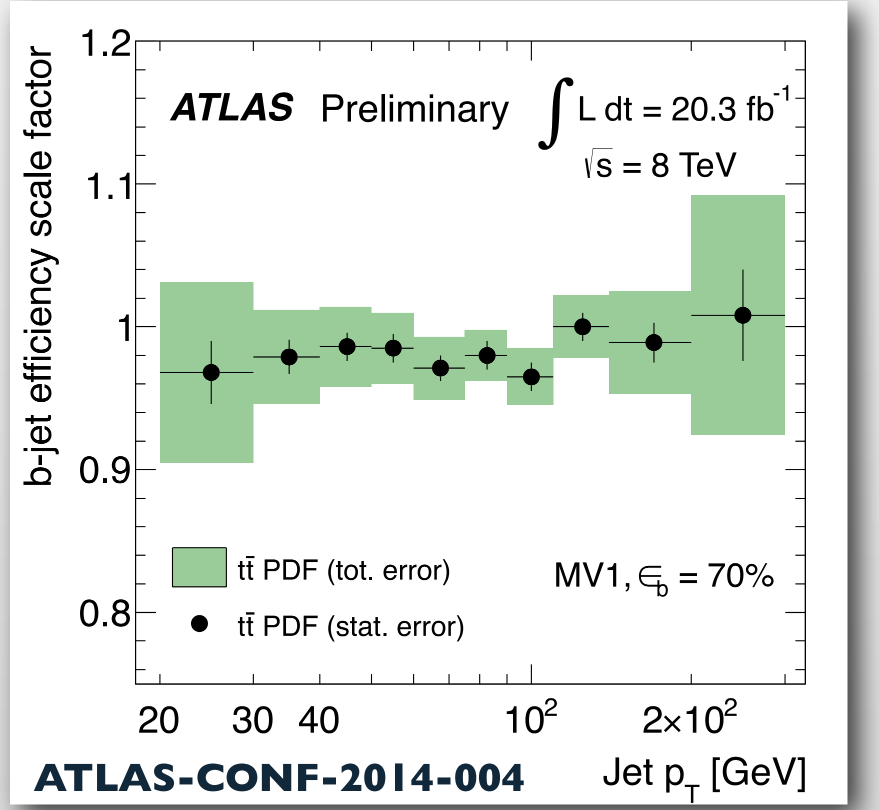
$t\bar{t}$ tag and probe - uncertainties 3% - 14%

Performance of the System8 method



System8 - uncertainties 5% - 20%

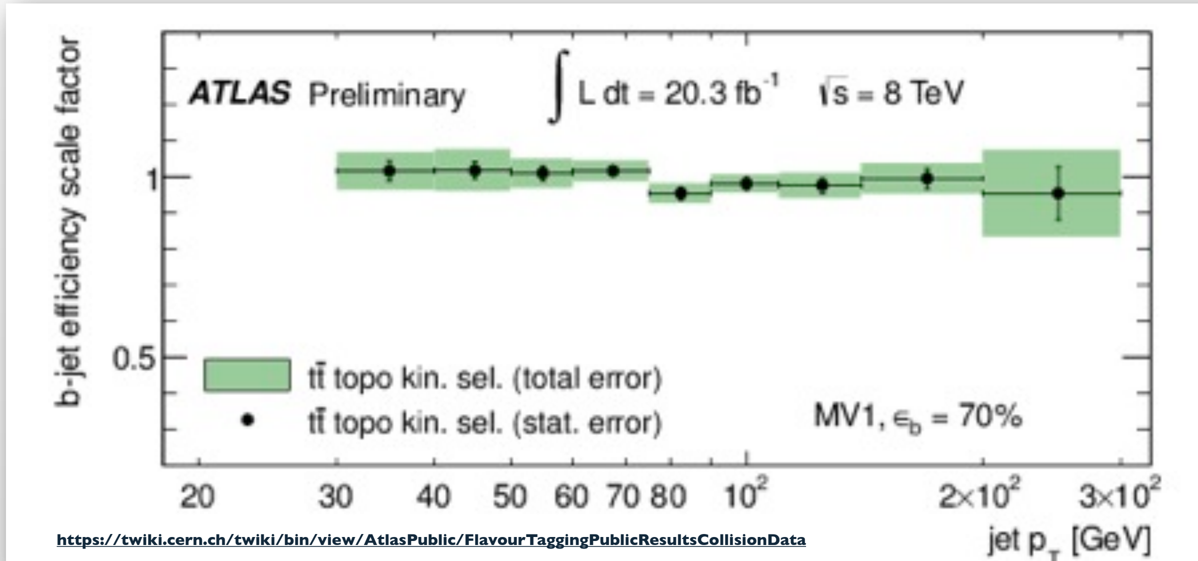
Performance of the $t\bar{t}$ PDF based method



$t\bar{t}$ PDF based - uncertainties <2% - 8%

Improvement with PDF based method

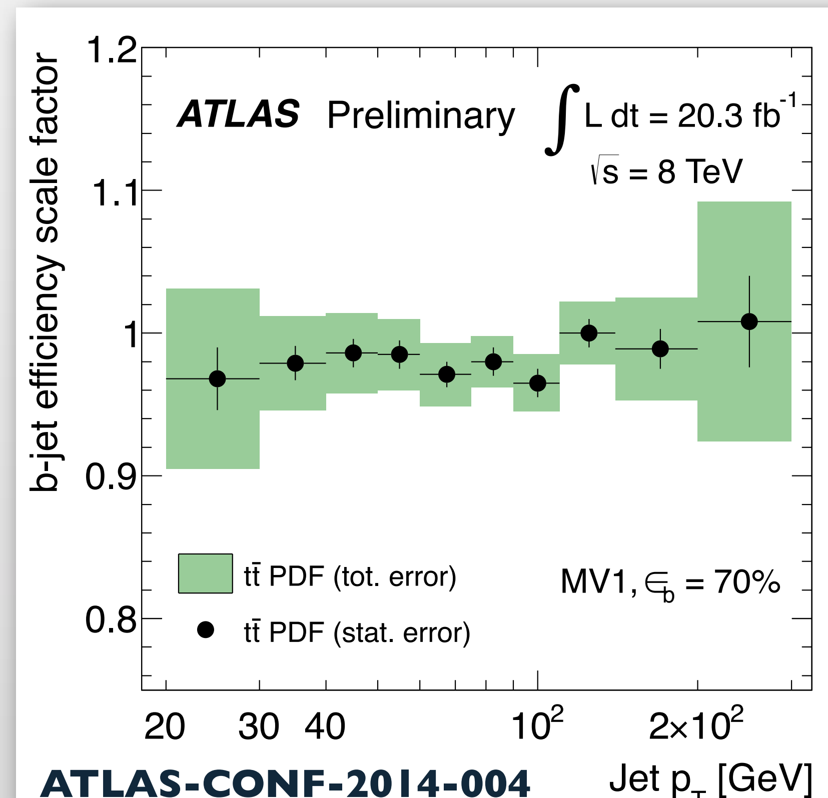
Performance of the tag and probe method in $t\bar{t}$ event



Reduction on average:
Statistical unc. 55%
Total unc. 36%

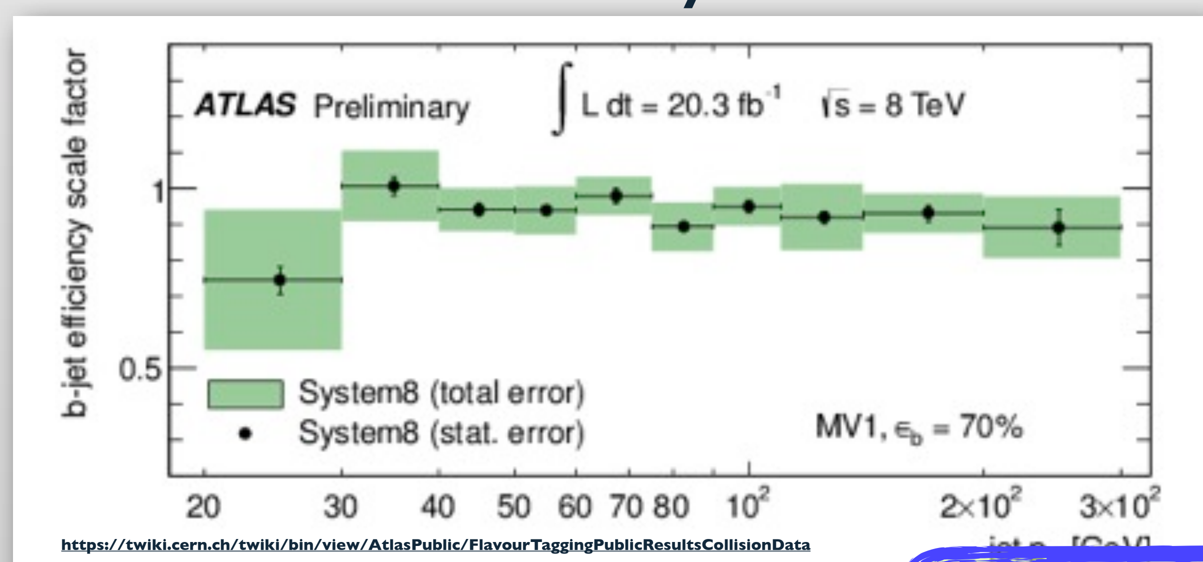
$t\bar{t}$ tag and probe - uncertainties 3% -14%

Performance of the $t\bar{t}$ PDF based method



$t\bar{t}$ PDF based - uncertainties <2% - 8%

Performance of the System8 method



System8 - uncertainties 5% - 20%

Large reduction of uncertainties with respect to the previously used methods!

Conclusions

- Processes with b-quark(s) in the final state important for the physics program.
- Identification of b-jets based on physics of the b-quark hadronisation and B-meson properties

  ➡ enabled by excellent tracking and vertexing performance in ATLAS.
- Several algorithms for the b-jet identification developed in ATLAS and combined into the sophisticated multivariate technique algorithm MV1.
- Efficiency of the b-jet identification measured in data using several methods
- Recently developed PDF based calibration method in $t\bar{t}$ enriched sample reduces significantly theoverall uncertainty with respect to previous methods

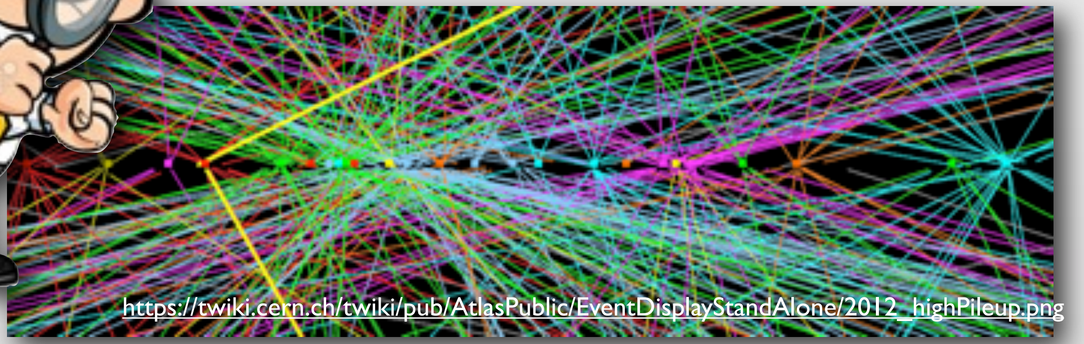
ATLAS is well equipped to successfully pursue the physics program which relies on performant b-jet identification!

ATLAS allows for it!

Excellent tracking

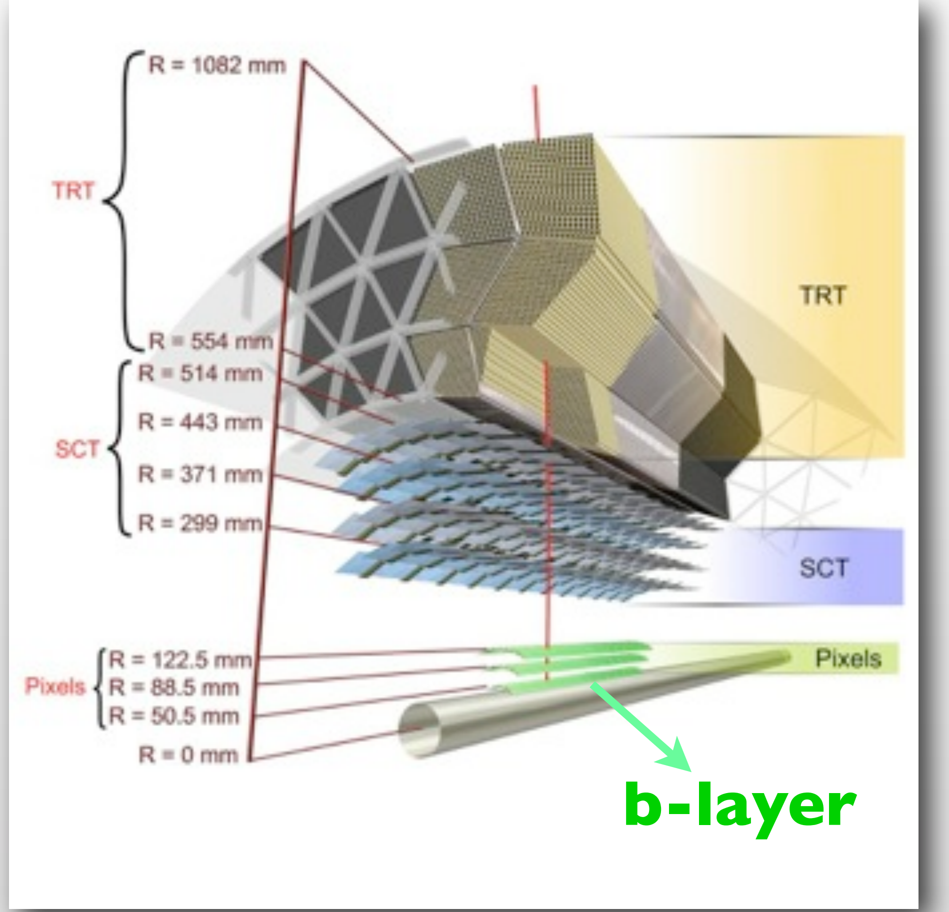
- Robust reconstruction algorithms performant under the high occupancy in the inner detector

ATLAS $Z \rightarrow \mu\mu$ event in 2012



https://twiki.cern.ch/twiki/pub/AtlasPublic/EventDisplayStandAlone/2012_highPileup.png

ATLAS Inner detector

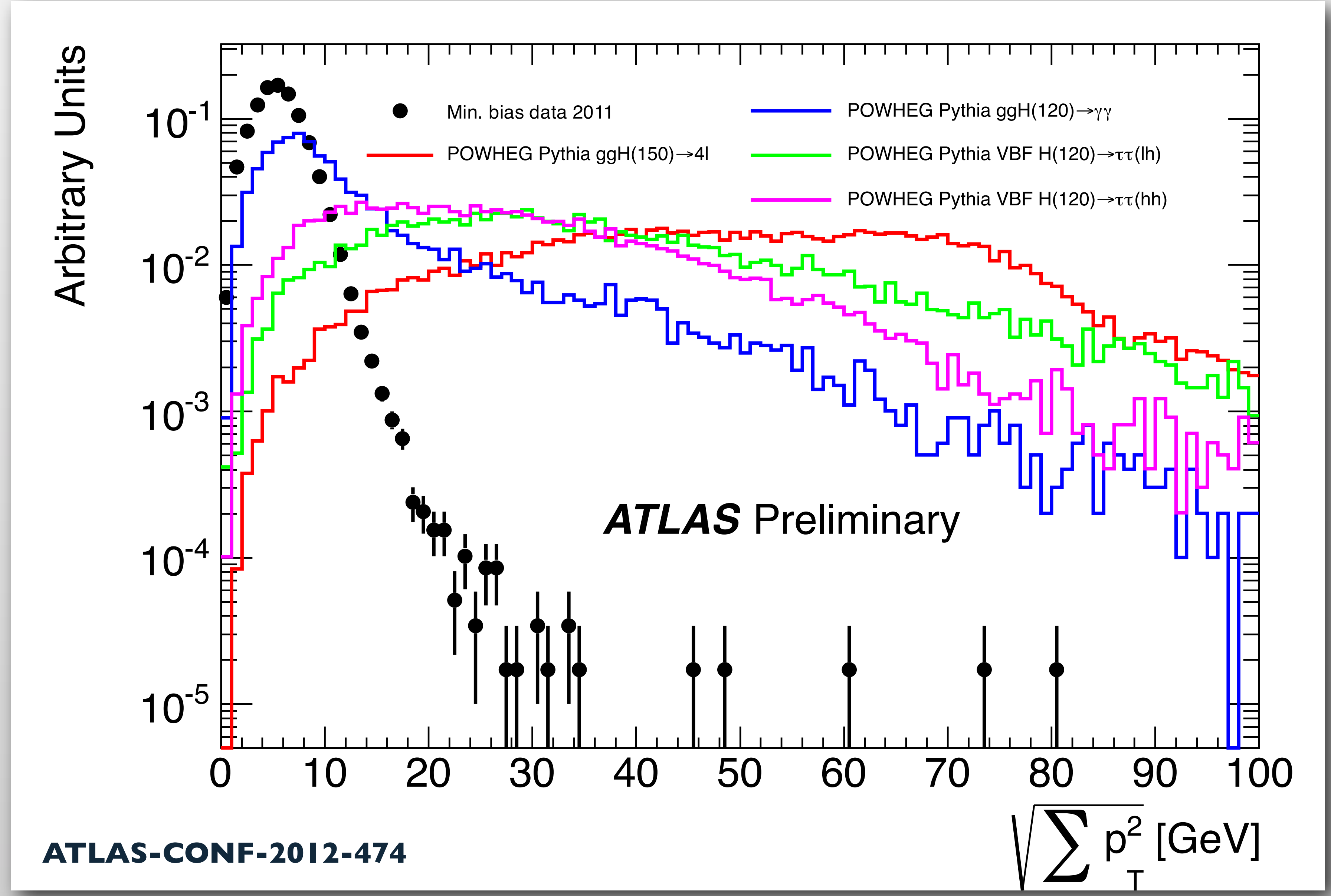


Excellent vertexing

- Efficiency ~**95%**
- Resolution (vertices with 70 tracks)
  - transverse: ~**30 µm**
  - longitudinal: ~**50 µm**

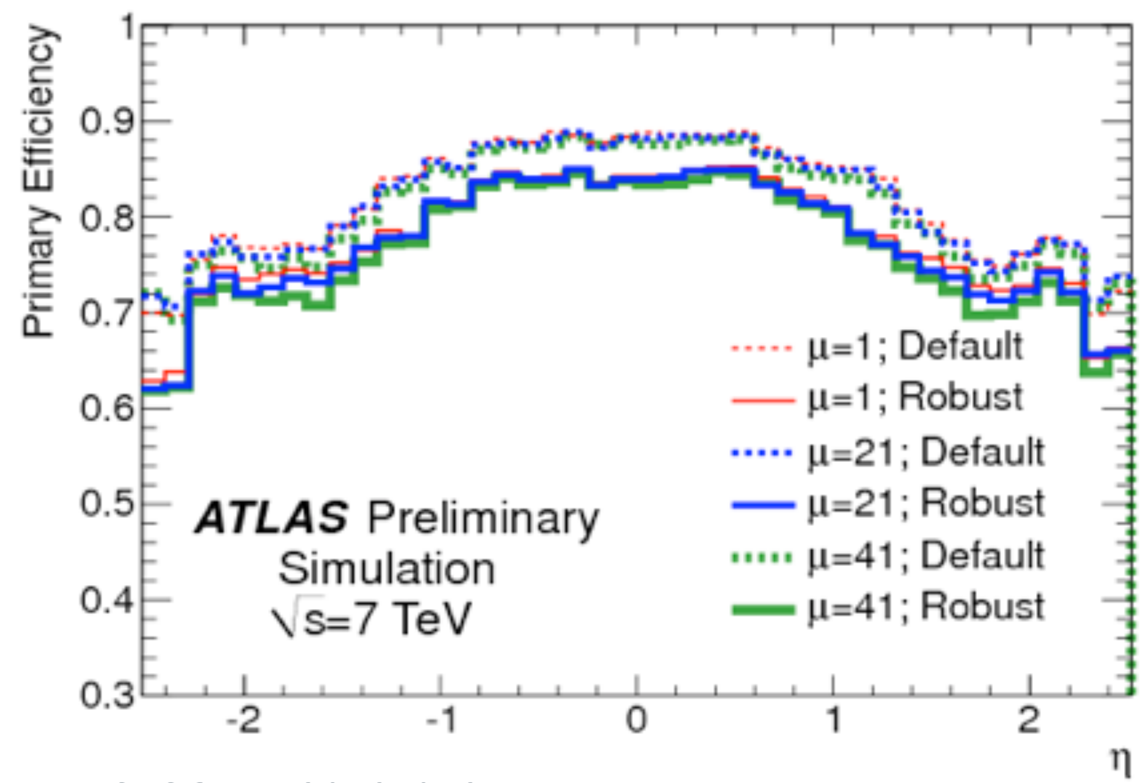
Primary vertex choice

Discriminant between hard and non-hard scatter vertex



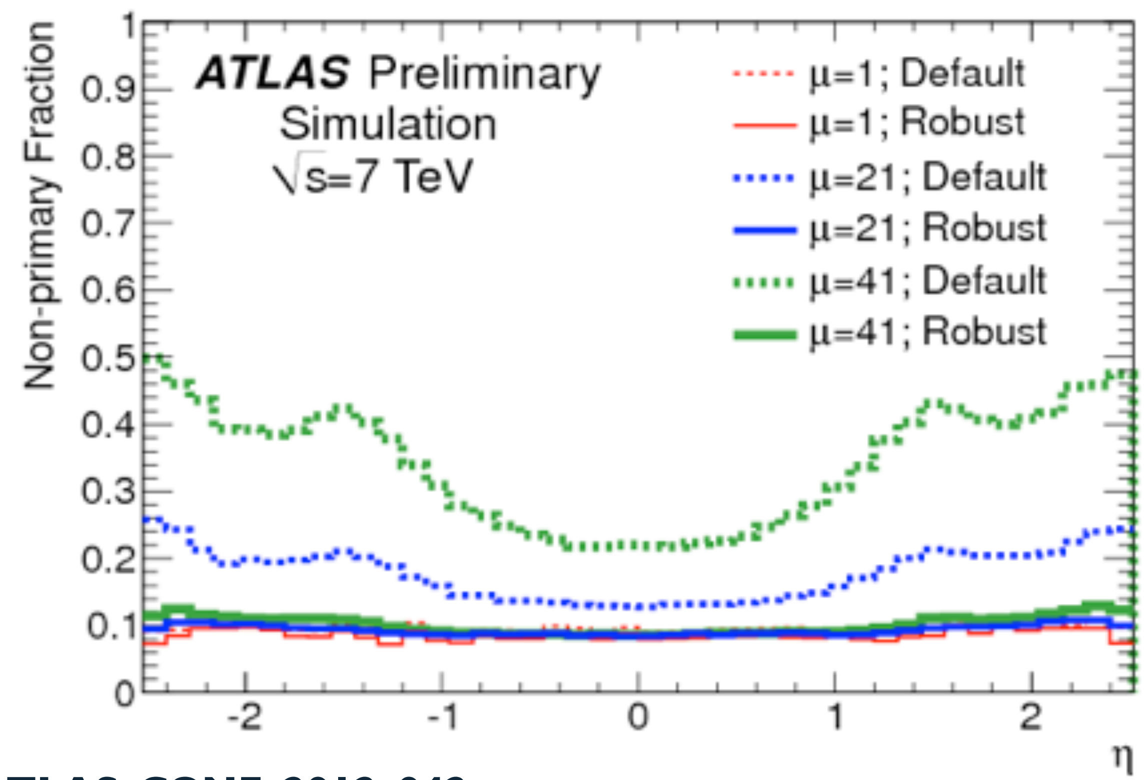
Backup - Tracking efficiency

Track reconstruction efficiency in minimum bias MC samples



ATLAS-CONF-2012-042

Fraction of fake tracks in 3 pile-up configurations

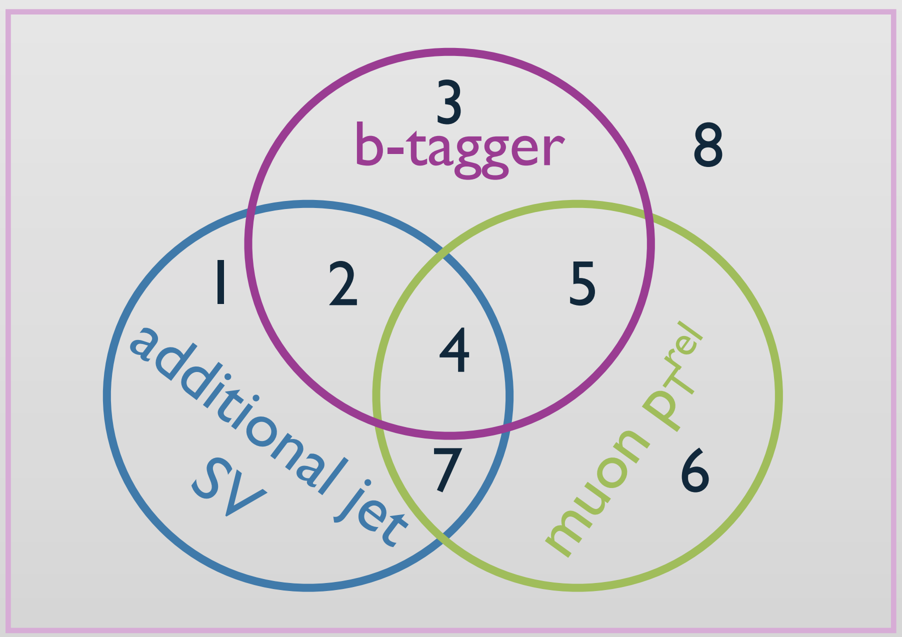


ATLAS-CONF-2012-042

Backup - Historically used methods

In sample with muons

- multijet events with soft muons
- exploit characteristics of the muons associated with jets
- **System8** method - applies 3 independent criteria do data



- Build the system of 8 equation between observed and expected counts

In enriched $t\bar{t}$ samples

- statistically significant at 8 TeV!
- $t \rightarrow Wb$ → enriched with b
- Apply **kinematic selections** to enhance very high purity

1. **Count** fraction of tagged jets in jet multiplicity bins
2. **Extract b-tag efficiency** by counting the fraction of tagged events in data and in simulation.

$$f^{data}_{btag} = \epsilon_b f^{sim}_{b-jets} + \sum_{non-b} \epsilon^{sim.}_{non-b} f^{sim}_{non-b}$$

fraction of tagged jets in data

fraction of real b tagged

fraction of non-b (mis)tagged

Backup - PDF based calibration method

Main difference in respect to the other methods (2 jet case)

- Information if the second jet in the event can be used

$$f_{\text{tagged}} = f_b \epsilon_b + (1 - f_b)\epsilon_j \rightarrow$$

$$f_{2\ \text{tags}} = f_{bb}\epsilon_b^2 + f_{bj}\epsilon_j\epsilon_b + (1 - f_{bb} - f_{bj})\,\epsilon_j^2$$

$$f_{1\ \text{tag}} = 2f_{bb}\epsilon_b\,(1-\epsilon_b) + f_{bj}\left[\epsilon_j\,(1-\epsilon_b) + (1-\epsilon_j)\,\epsilon_b\right] + (1 - f_{bb} - f_{bj})\,2\epsilon_j\,(1-\epsilon_j)\,,$$

- In the case of N bins in kinematic variables, N^2 possible combinations for 2 jet channel -> 2 x N^2 non-linear equations (N eff for b, N for on b).
- Instead, model the system using unbinned likelihood (can be extended to an arbitrary binning in any jet kinematic quantity).

2 dim PDF for [$p_{T,1}$, $p_{T,2}$] for flavour combination *bj (bb,ljj*

PDF for the b-tagging discriminant for b(j) jet

$$\begin{aligned}\mathcal{L}\left(p_{\mathrm{T},1}, p_{\mathrm{T},2}, w_1, w_2\right) = \big[& f_{bb}\mathcal{P}_{bb}\left(p_{\mathrm{T},1}, p_{\mathrm{T},2}\right)\mathcal{P}_b\left(w_1|p_{\mathrm{T},1}\right)\mathcal{P}_b\left(w_2|p_{\mathrm{T},2}\right) \\ & + f_{bj}\mathcal{P}_{bj}\left(p_{\mathrm{T},1}, p_{\mathrm{T},2}\right)\mathcal{P}_b\left(w_1|p_{\mathrm{T},1}\right)\mathcal{P}_j\left(w_2|p_{\mathrm{T},2}\right) \\ & + f_{jj}\mathcal{P}_{jj}\left(p_{\mathrm{T},1}, p_{\mathrm{T},2}\right)\mathcal{P}_j\left(w_1|p_{\mathrm{T},2}\right)\mathcal{P}_j\left(w_2|p_{\mathrm{T},2}\right) \\ & + 1 \leftrightarrow 2 \,\big]/2,\end{aligned}$$

Maximizing using Minuit

P_{bb} and Pjj symmetrised - reduces stat fluctuations in MC

Backup - PDF based calibration method

Systematics breakdown

p_T interval [GeV]	20-30	30-40	40-50	50-60	60-75	75-90	90-110	110-140	140-200	200-300
SF	0.968	0.979	0.986	0.985	0.971	0.980	0.965	1.000	0.989	1.008
Total error [%]	6.5	3.4	2.8	2.5	2.3	1.8	2.1	2.2	3.6	8.4
Stat. error [%]	2.3	1.2	1.0	1.0	0.9	1.0	1.0	1.0	1.4	3.2
Syst. error [%]	6.1	3.1	2.7	2.3	2.1	1.5	1.9	2.0	3.3	7.6
Systematic Uncertainties [%]										
Hadronisation ($t\bar{t}$)	1.0	0.6	1.5	1.4	1.1	0.5	0.8	0.3	1.0	2.0
Modelling ($t\bar{t}$)	1.1	0.4	1.0	1.1	1.0	0.5	0.7	0.9	0.7	1.7
Top p_{T} reweighting ($t\bar{t}$)	0.2	0.3	0.3	0.2	0.2	0.1	0.1	0.4	1.4	4.6
More/less PS ($t\bar{t}$)	0.5	0.6	0.9	0.8	0.9	1.0	0.9	0.8	1.4	1.9
More/less PS (single top)	0.2	0.0	0.1	0.1	0.2	0.1	0.2	0.2	0.0	0.0
Modelling (Z+jets)	0.8	0.3	0.2	0.5	0.3	0.2	0.3	0.3	0.9	2.4
Modelling (dibosons)	0.7	0.7	0.6	0.6	0.6	0.6	0.7	0.8	1.3	3.1
Norm. single top	0.5	0.4	0.3	0.2	0.2	0.2	0.2	0.2	0.3	0.0
Norm. Z+jet	0.9	0.6	0.9	0.4	0.7	0.5	0.6	0.7	1.1	1.7
Norm. Z+b/c	0.1	0.1	0.2	0.1	0.1	0.0	0.1	0.0	0.1	0.2
Norm. lepton fakes	0.3	0.3	0.2	0.3	0.2	0.2	0.3	0.3	0.3	0.4
Pile-up reweighting	0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.1
Electron eff./res./scale	0.1	0.1	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Muon eff./res./scale	0.2	0.3	0.2	0.3	0.1	0.2	0.1	0.1	0.0	0.0
$E_{\mathrm{T}}^{\mathrm{miss}}$ soft-terms	0.1	0.1	0.2	0.0	0.2	0.0	0.1	0.2	0.3	0.5
Jet energy scale	4.1	2.2	1.2	0.7	0.7	0.3	0.7	0.8	1.2	2.6
Jet energy resolution	2.6	1.0	0.3	0.3	0.1	0.2	0.2	0.0	0.2	0.2
Jet vertex fraction	0.8	0.1	0.0	0.1	0.1	0.1	0.2	0.2	0.2	0.2
Mis-tag rate	2.8	1.1	0.5	0.4	0.3	0.2	0.3	0.4	0.5	1.1