Measurement of b-tagging efficiency using $t\bar{t}$ events

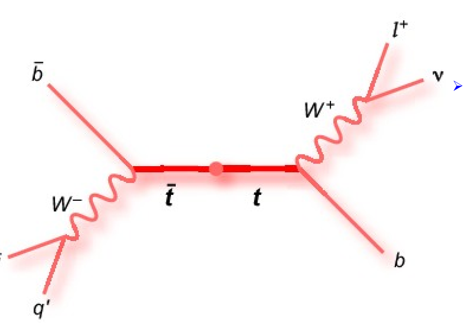
Michael Segala on behalf of the CMS Collaboration

Jet flavour tagging is one of the key ingredients of the diverse physics program of the CMS experiment. To extract the efficiencies for tagging the b-quark, and c-quark jets, we exploit the $t\bar{t}$ samples collected at √s=7 TeV running of the LHC during the first half of 2011 where a total of 2.18 fb^{-1} of data was collected. Different techniques based on a collection of high purity b-flavored jets, or on requiring the consistency between the observed and expected number of tags in the events are used to extract the performance of the heavy flavor algorithms directly from data and measure the differences between data and simulation.

Event Selection

Lepton + Jets Analysis

- Trigger
  - Isolated Single Muon Triggers
- Leading isolated, prompt muon
  - Global and tracker muon
  - P_T > 30 GeV, |η| < 2.5
  - Veto events with 2nd lepton
  - Jet-lepton cleaning (ΔR > 0.3) with respect to selected jets
- At least 4 jets reconstructed using Particle Flow algorithm
  - P_T > 30 GeV, |η| < 2.5
  - Apply Jet Energy Corrections



Dilepton Analysis

- Trigger
  - Inclusive Dilepton Triggers
- 2 isolated, prompt leptons of opposite sign (ee,μμ,eμ)
  - P_T > 20 GeV, |η| < 2.1(e) / 2.4 (μ)
  - Dilepton invariant mass M > 12 GeV/c^2
  - Z-veto for ee and μμ M < 76 GeV/c^2 or M > 106 GeV/c^2
- At least 2 jets reconstructed using Particle Flow algorithm
  - P_T > 30 GeV, |η| < 2.5
  - Apply Jet Energy Corrections
- Missing Transverse Energy > 30 (ee,μμ)

Event Yields

Lepton+jets Analysis

	no b-tagging	≥ 1 b-tagged jets	≥ 2 b-tagged jets
tt	7741 ± 1161	6514 ± 977	2939 ± 441
Single top	437±105	350± 83	108± 23
W+Jets	5673 ± 1701	1247 ± 374	153 ± 46
Z+Jets	422 ± 127	72 ± 22	12 ± 4
QCD	22 ± 7	8 ± 2	0 ± 0
Sum MC	14295 ± 2066	8191 ± 1049	3210 ± 444
Data	13935	8205	3246

Dileptons Analysis

Processes	Channel ee	Channel $\mu\mu$	Channel eμ
Without b-tagging requirement			
$t\bar{t}$ signal	932 ± 141	1223 ± 184	3313 ± 500
$t\bar{t}$ background	11.0 ± 5.5	3.2 ± 1.6	22.7 ± 11.3
Single top	46.7 ± 14.0	60.2 ± 18.1	157.0 ± 47.1
Di-bosons	21.4 ± 6.4	28.0 ± 8.4	47.4 ± 14.2
Z+jets	409 ± 204	545 ± 273	200 ± 100
W+jets	12.0 ± 6.0	0.0	11.4 ± 5.7
Sum MC	1432.1 ± 249.1	1860.4 ± 329.5	3751.2 ± 512.8
Data	1442	1773	3898
With ≥ 1 b-tagged jets			
Sum MC	1056.3 ± 165.2	1388.0 ± 212.1	3259.6 ± 456.3
Data	1080	1364	3375
With ≥ 2 b-tagged jets			
Sum MC	514.6 ± 71.7	677.7 ± 94.8	1757.2 ± 253.3
Data	554	686	1854

b-tagging

Algorithms: Several b-tagging algorithms have been developed to determine the flavor of a jet. These algorithms exploit the long lifetime of the b-quark and hence rely on the reconstruction and identification a large displaced secondary vertex from the primary vertex.

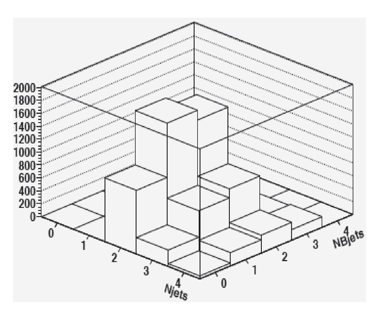
- TCHE: Track Counting High Efficiency
- TCHP: Track Counting High Purity
- JP: Jet Probability
- JBP: Jet B Probability
- SSVHE: Simple Secondary Vertex High Efficiency
- SSVHP: Simple Secondary Vertex High Purity
- CSV: Combined Secondary Vertex

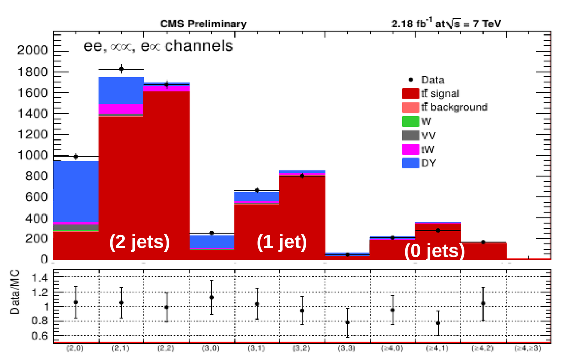
	Loose	Medium	Tight
TCHE	1.7	3.3	10.2
TCHP	1.19	1.93	3.41
JP	0.275	0.545	0.790
JBP	1.33	2.55	3.74
SSVHE	-	1.74	3.05
SSVHP	-	-	2.0
CSV	0.244	0.679	0.898

Working Points: Standard operating points are established by choosing three values for each discriminant, "loose", "medium", and "tight", being the value at which the acceptance of light flavor jets is estimated from simulation to be 10%, 1%, and 0.1% respectively.

Profile Likelihood Ratio method (PLR)

- Fit of jet multiplicity vs. b-tagged jet multiplicity 2D distribution





Likelihood in bin *i*, channel *j*

$$L_{i,j}(SF_b, N^{obs}_{i,j}, \{U_k\}) = \mathcal{P}oisson(N^{obs}_{i,j}, \mu_{i,j}(SF_b, \{U_k\})) \times \prod_k Gauss(U_k, 0, 1)$$

- $N^{obs}_{i,j}$: number of observed events (data)
- $\mu_{i,j}$: number of expected events (MC)
- SF_b: scale factor of the b-tagging efficiency (free parameter to fit)
- U_k : nuisance parameters (systematics fitted during the minimization)
- Likelihood in channel *j*: $L_j(SF_b, N^{obs}_{i,j}, \{U_k\}) = \prod_i L_{i,j}(SF_b, N^{obs}_{i,j}, \{U_k\})$
- Overall Likelihood: $L(SF_b, N^{obs}_{i,j}, \{U_k\}) = \prod_j L_j$

References

- CMS Collaboration, "Performance of the b-jet identification in CMS", CMS Physics Analysis Summary CMS-PAS-BTV-11-001 (2011).
- CMS Collaboration, "Measurement of the b-tagging efficiency using $t\bar{t}$ events", CMS Physics Analysis Summary CMS-PAS-BTV-11-003 (2011).
- CMS Collaboration, "b-Jet identification in the CMS Experiment", CMS Physics Analysis Summary CMS-PAS-BTV-11-004 (2011).
- CMS Collaboration, "Combination of top pair production cross section measurements", CMS Physics Analysis Summary CMS-PAS-TOP-11-024 (2011).

Flavor tag Consistency Method (FtCM)

- Consistency between expected and observed number of b-tagged jets
  - Simultaneous measurement of b-jet tagging efficiencies and σ_{tt}
  - ML fit: enforce consistency between observed and expected distributions of b-tagged jets in the events
- Expected Number of Tagged Jets

Free parameters

$$\langle N_{0,1,2,3,4}\rangle = f(\epsilon_b, \epsilon_c, \epsilon_l, \epsilon_{sel}, lumi, \sigma_{t\bar{t}}, F_{ijk})$$

Fixed parameters

  - Fraction of events with "i" b-jets, "j" c-jets and "k" light-jets F_{ijk} (From MC truth)
  - Selection efficiency ε_{sel} taken from MC
  - Tagging efficiency for c and light jets ε_c, ε_l : from MC, with SF
  - Events with 4 – 7 jets considered
- For expected events =

$$\langle N_n\rangle = \mathcal{L}\cdot\sigma_{t\bar{t}}\cdot\varepsilon\cdot\sum_{i,j,k} F_{ijk}\sum_{\substack{i'\le i, j'\le j, k'\le k\\ i'+j'+k'=n}}[C_i^{i'}\varepsilon_b^{i'}(1-\varepsilon_b)^{(i-i')}C_j^{j'}\varepsilon_c^{j'}(1-\varepsilon_c)^{(j-j')}C_k^{k'}\varepsilon_l^{k'}(1-\varepsilon_l)^{(k-k')}].$$

- Likelihood to measure $\boxed{\epsilon_b \;, \sigma_{t\bar{t}}}$ $\boxed{\mathcal{L} = -2\log\prod_n \mathcal{P}oisson(N_n, \langle N_n\rangle)}$

Flavor tag Matching Method (FtMM)

- Matching between expected and observed number of b-tagged jets
  - Dilepton channel
- Expected number of events with *n* b-tagged jets <N_n>: $\langle N_n\rangle = \sum_{\text{k jets}=2}^{\text{all jets}} n_k \cdot P_{n,k}$
  - n_k : observed number of events with *k* jets
  - $P_{n,k}$: probability to count *n* b tags in a *k*-jet event:
    - $P_{n,2,i}$: probability that *k* b tags are observed in an event with two jets of which *i* jets come from $t\bar{t}$ decays.

$$P_{n,2} = \sum_{\substack{\text{i jets}=0\\ \text{from top decay}}}^{2} \alpha_i \cdot P_{n,2,i}$$

$P_{2,2,0} = \varepsilon_q^2$	if no jets are from $t\bar{t}$ decays
$P_{2,2,1} = 2^2\varepsilon_b\varepsilon_q$	if 1 jet is from $t\bar{t}$ decays
$P_{2,2,2} = \varepsilon_b^2$	if 2 jets are from $t\bar{t}$ decays

    - $P_{n,2,i}$: expressed as function of ε_b, ε_q and combinatorics.

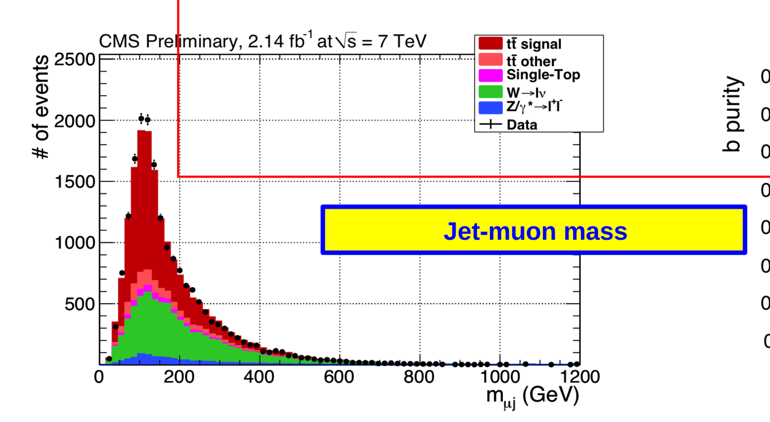
α_i : mis-assignment probabilities
probability *i* jets from $t\bar{t}$ decay have been reconstructed and selected
α_i : probability that both b jets from the $t\bar{t}$ decay have been selected
Measured on data to account for ISR, background, reco effect, etc...

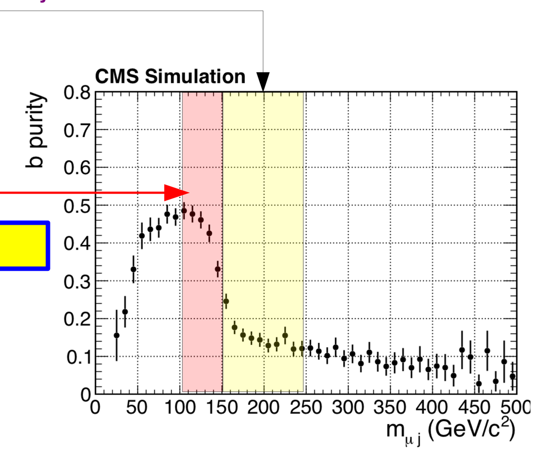
Measurement from a b-enriched sample (b-Sample)

- SF measured from a b-jet-enriched sample
  - Contamination from light jets estimated from data and subtracted
- Lepton+Jet events
- χ^2 : for each jet-parton combination
  - 4 jets: 12 combinations to assign jets in $t\bar{t}$ decay
  - Mean, width from tt MC with correct assignment
- Use combination with lowest χ^2
  - Cut at χ^2 < 90

$$\chi^2 = \left(\frac{m_{bqq}-m_t}{\sigma_t}\right)^2 + \left(\frac{m_{qq}-m_W}{\sigma_W}\right)^2$$

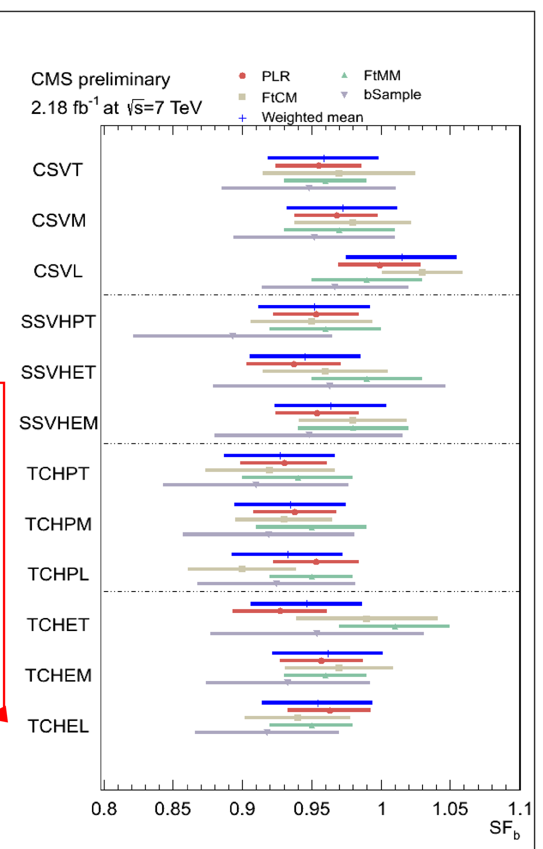
- b-jet sample: jet from leptonic t → Wb decay
- Define b-enriched , b-depleted subsamples:
  - jet-muon mass, $m_{\mu j}$: invariant mass of leptonic b jet and muon



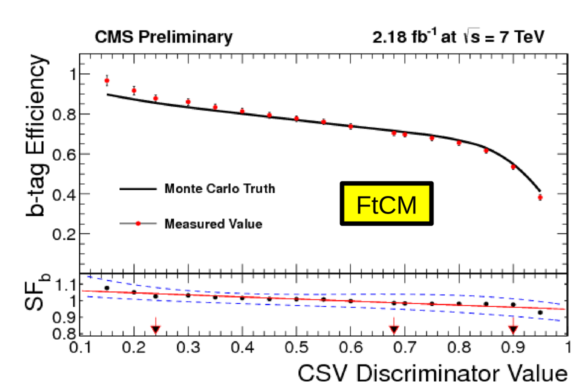


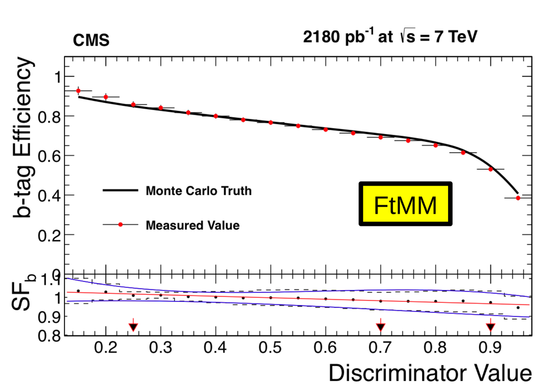
Results

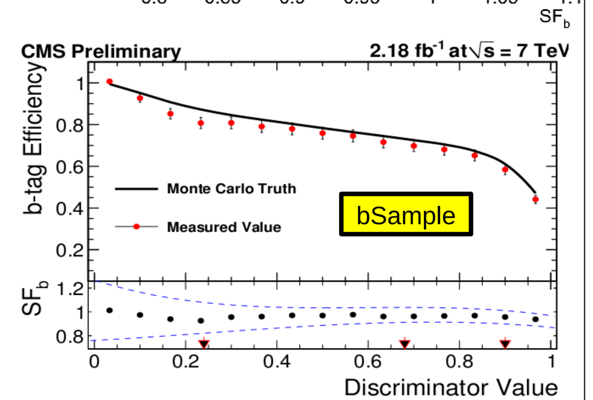
Point	Dijet	PLR	FtCM	FtMM	bSample	WM
TCHEL	0.95±0.095	0.96 ± 0.03	0.94 ± 0.04	0.95 ± 0.03	0.92 ± 0.05	0.95± 0.04
TCHEM	0.94±0.094	0.96 ± 0.03	0.97 ± 0.04	0.96 ± 0.03	0.93 ± 0.06	0.96± 0.04
TCHET	-	0.93 ± 0.03	0.99 ± 0.05	1.01 ± 0.04	0.95 ± 0.08	0.95± 0.04
TCHPL	-	0.95 ± 0.03	0.90 ± 0.04	0.95 ± 0.03	0.92 ± 0.06	0.93± 0.04
TCHPM	0.91±0.091	0.94 ± 0.03	0.93 ± 0.04	0.95 ± 0.04	0.92 ± 0.06	0.93± 0.04
TCHPT	0.88±0.088	0.93 ± 0.03	0.92 ± 0.05	0.94 ± 0.04	0.91 ± 0.07	0.93± 0.04
SSVHEM	0.95±0.095	0.95 ± 0.03	0.98 ± 0.04	0.98 ± 0.04	0.95 ± 0.07	0.96± 0.04
SSVHET	-	0.94 ± 0.03	0.96 ± 0.05	0.99 ± 0.04	0.96 ± 0.08	0.95± 0.04
SSVHPT	0.90±0.090	0.95 ± 0.03	0.95 ± 0.04	0.96 ± 0.04	0.89 ± 0.07	0.95± 0.04
CSVL	1.00±0.100	1.00 ± 0.03	1.03 ± 0.03	0.99 ± 0.04	0.97 ± 0.05	1.01± 0.04
CSVM	0.96±0.096	0.97 ± 0.03	0.98 ± 0.04	0.97 ± 0.04	0.95 ± 0.06	0.97± 0.04
CSVT	0.94±0.094	0.95 ± 0.03	0.97 ± 0.05	0.96 ± 0.03	0.95 ± 0.06	0.96± 0.04



Weighted Mean (WM): PLR (dilepton) + FtCM (l+jets)
Best in each channel - Statistically uncorrelated
The average standard deviation of the 4 measurements, 0.02, as been used to compute the uncertainty on the the WM.







Final function = Linear fit to FtCM + offset to fit the Medium point of the Weighted Mean