b-tagging in CMS

Caroline Collard
IPHC (Strasbourg)

On behalf of the CMS Collaboration

Outline of the talk

- Identification of b-jets : variety of algorithms (offline and online)

  → Used in Top, SM (bb, V+bb, V+cc) and Higgs (H->bb) studies, and in 3rd generation in SUSY and BSM searches (W', Z', T', b', T_5/3, ...) + in other analyses with veto against top background.

- Performance Measurements and Scale Factor determination

- French contribution to the b-tagging group

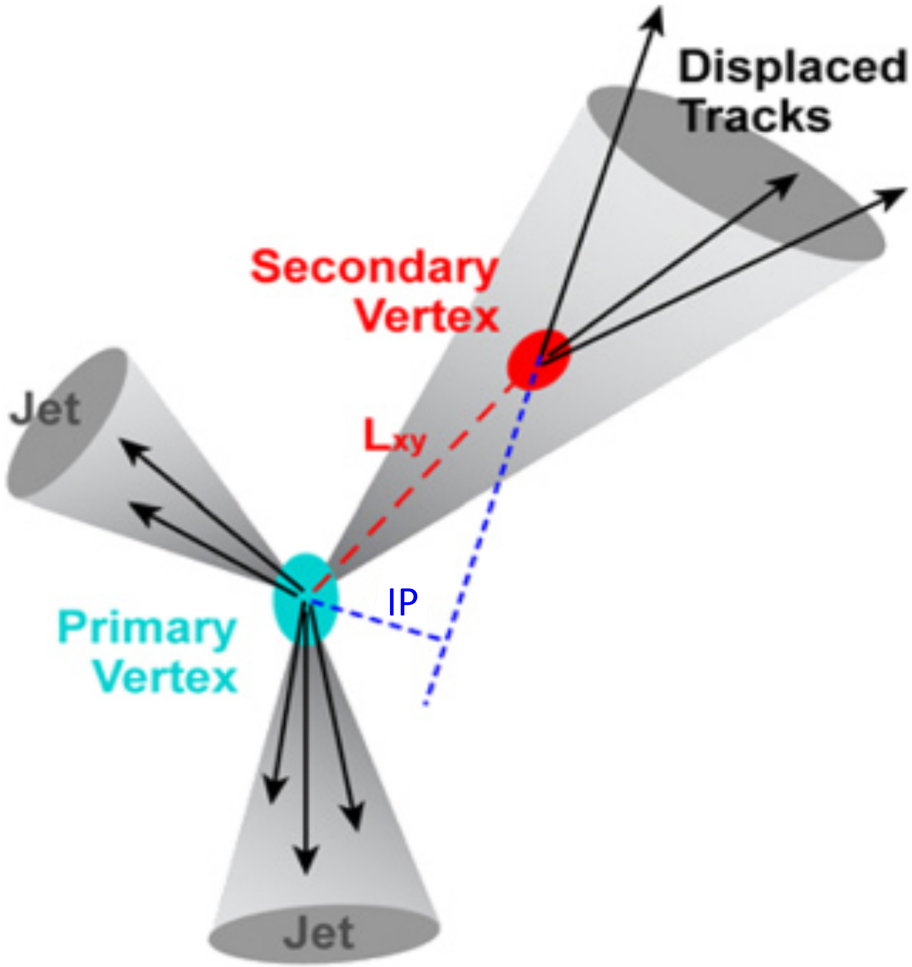
Identification of b-quark based on B-hadron properties

b-quark hadronize in B-hadrons, forming jets.

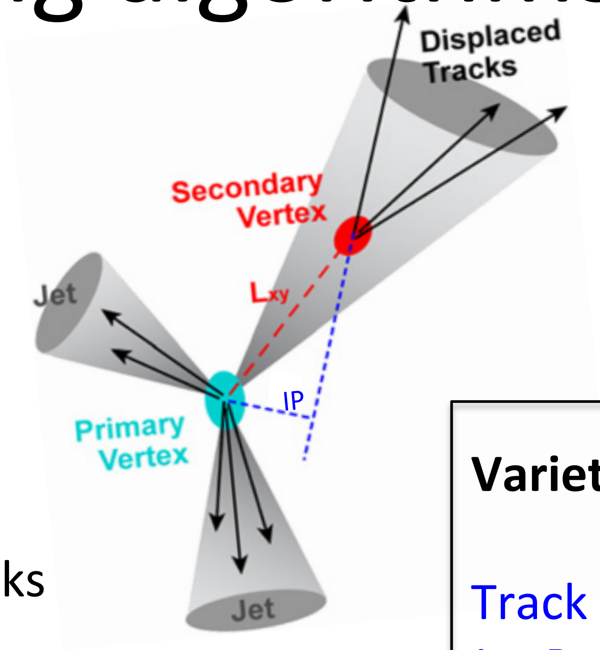
B-lifetime → displaced tracks
→ **large and positive Impact Parameter (IP)**

B-lifetime → **secondary vertex (SV)**

B-decay into leptons (BR in e, μ ~40%)



b-tagging algorithms in CMS



B-lifetime → displaced tracks
→ **large and positive Impact Parameter (IP)** →

B-lifetime → **secondary vertex (SV)** →

B-decay into leptons (BR in e, μ ~40%) →

Variety of algorithms in CMS :

Track Counting
Jet Probability

Simple Secondary Vertex
Inclusive Vertex Finder

Soft Lepton taggers

Combined taggers

Algorithms based on track IP

IP typical size: few 100 μm,

Uncertainty on IP: can be large as 10-100 μm

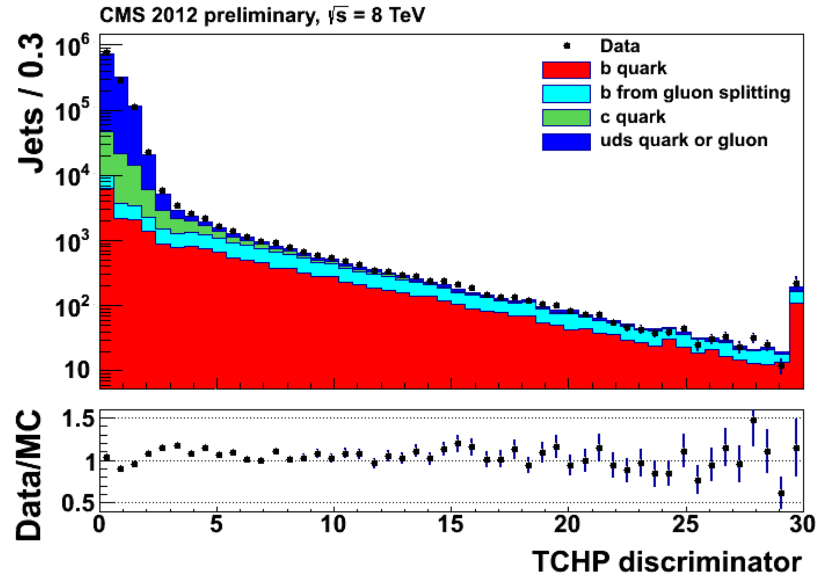
→ Use IP Significance : IP/σ

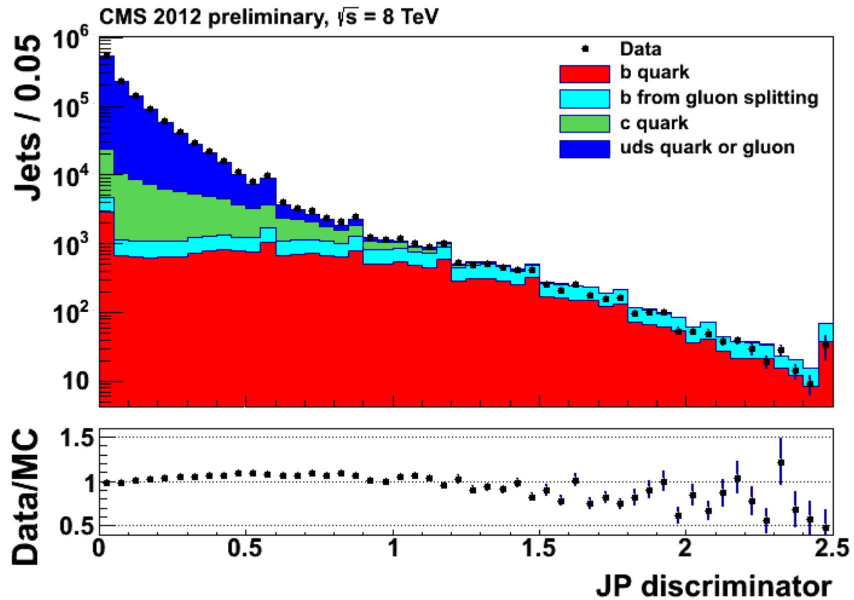
- Track counting:

Order tracks by IP significance, and look at the 2[nd] [TCHE] or 3[rd] [TCHP] ranked track.

- Jet Probability:

Likelihood to estimate the probability of jet tracks to come from the primary vertex [JP], JBP gives more weight to highest IP significance tracks. Calibrated with Data.





Algorithms based on Secondary Vertices

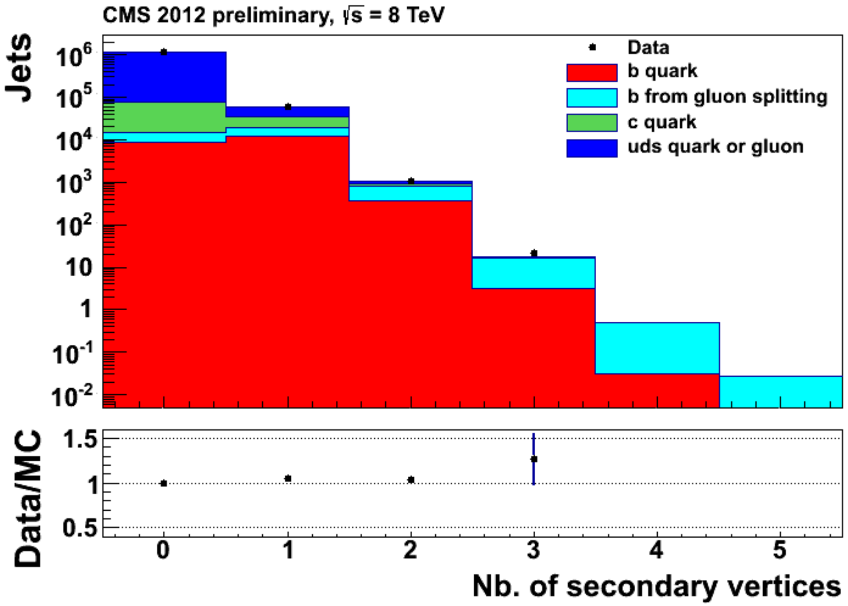
Reconstruct secondary vertex using dedicated algorithm → limited to efficiency : max 65%

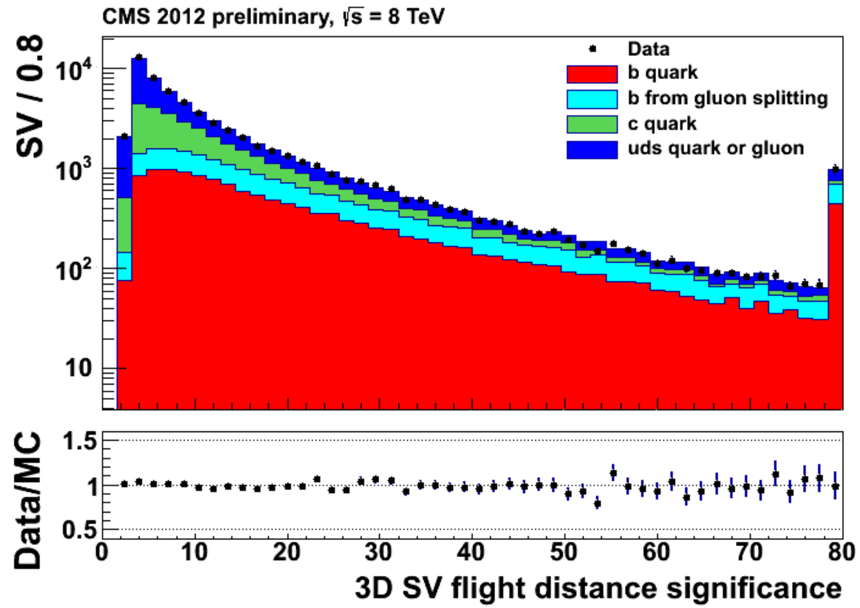
- Simple SV:

Look at the flight distance significance (L/σ between PV and SV) if there is ≥ 2 [SSVHE] or ≥3 [SSVHP] tracks @SV

- Inclusive vertex finding [IVF]:

use another vertex finding algorithm, independent of jets





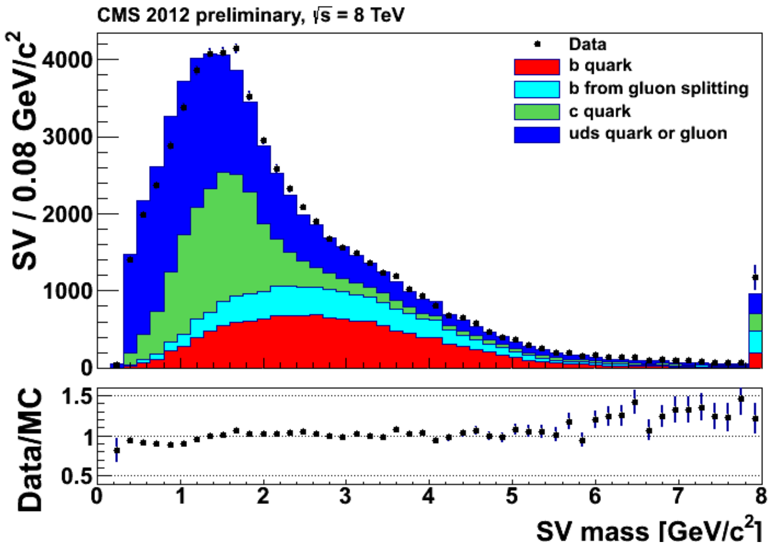
Combined Algorithms

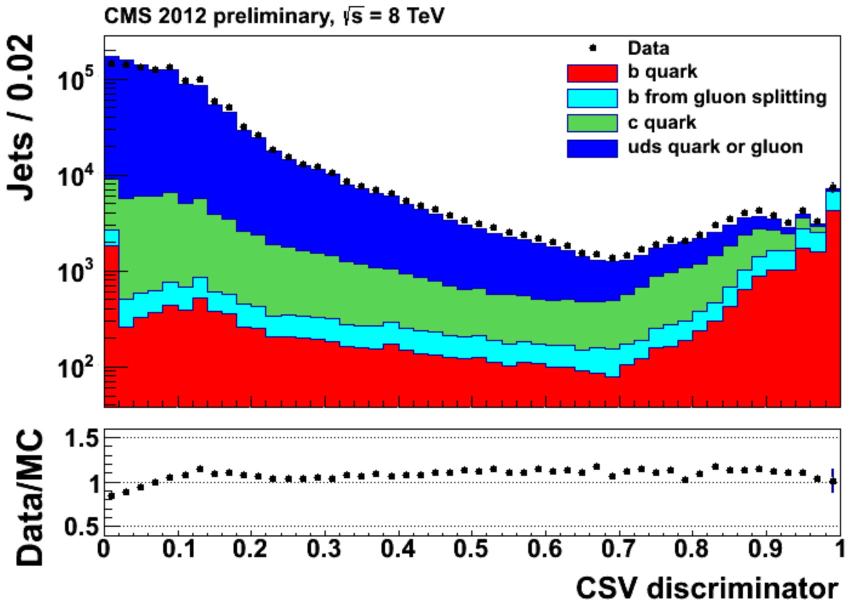
Aim: Try to recover the cases when no SV is found

- "real vertex" → SV-based info
  (vertex mass, 2D flight distance,
  energy ratio, …)
- "pseudo vertex" : combination of tracks with an IP significance > 2
  → SV-based info
- "no vertex"
  → # of tracks, IP significances

- Combined Secondary Vertex:

Using a likelihood [CSV]
or a boost decision tree or neural network [CSVMVA]



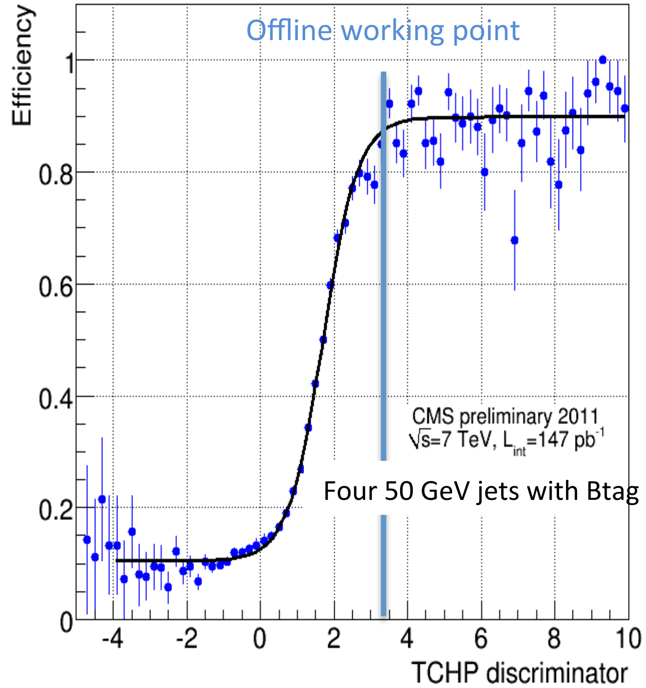


Online b-tagging

Triggers using b-tagging information @HLT :

- Track Counting → used in studies on Top (single top) and Higgs (MSSM Higgs+b, HV)
- Combined Secondary Vertex → used in studies on Higgs (VBF)
- Soft Lepton → For Commissioning and performance studies

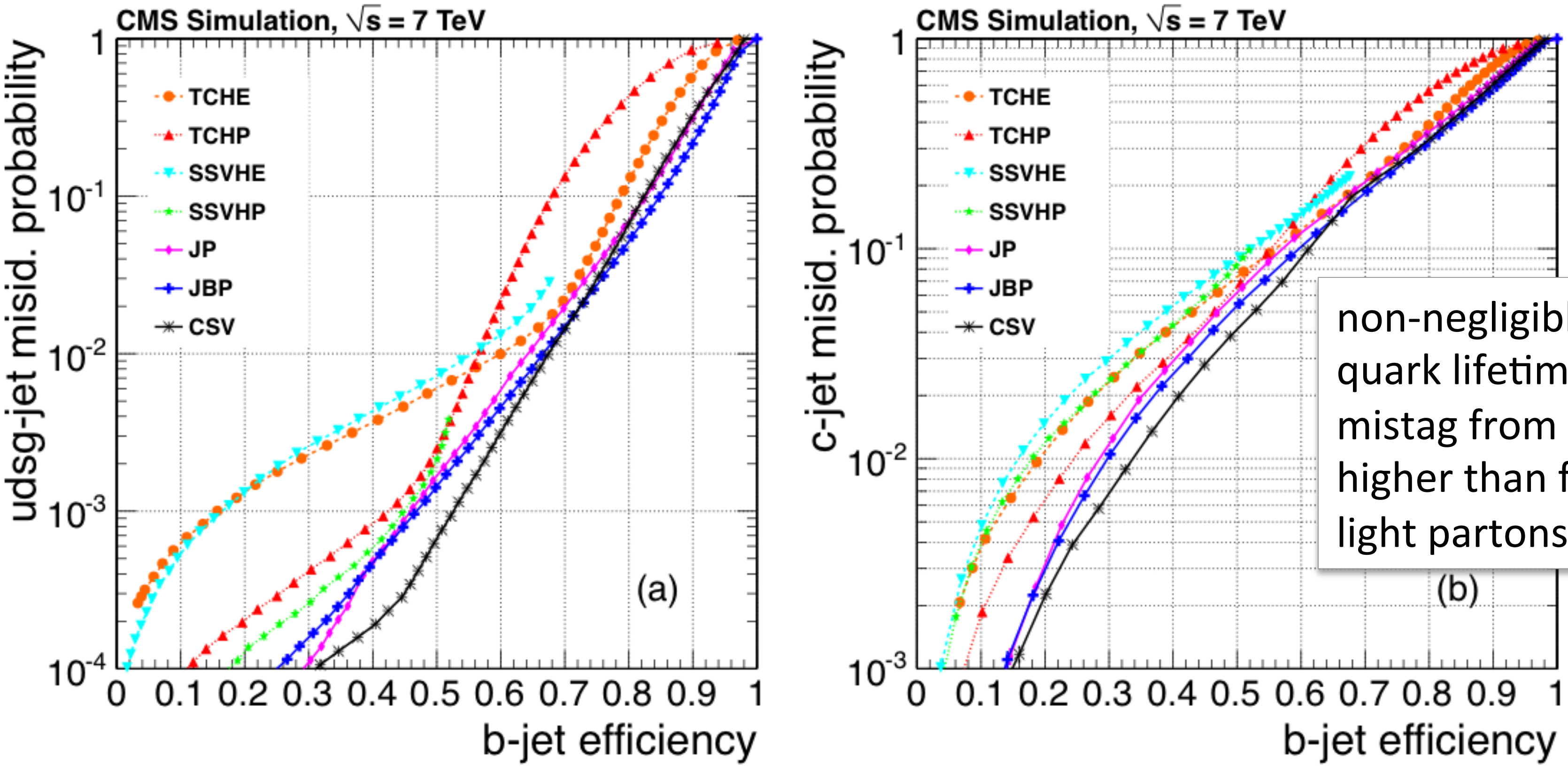
Allows to decrease the threshold on other objects while keeping an acceptable rate.



Important : Always apply a looser selection online than in the offline analysis

Expected performance in simulation

Scans when varying the discriminator cuts:



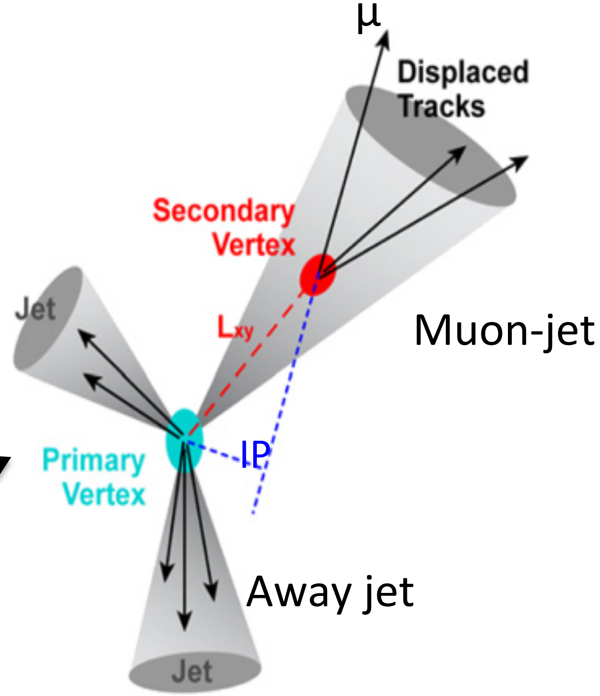
non-negligible c-quark lifetime → mistag from c higher than from light partons

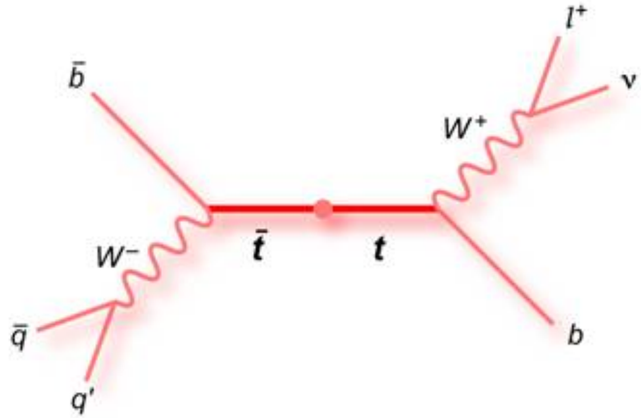
Selection	b-jet efficiency	Mistag for light partons	Best tagger
Loose	80-85%	10%	JBP
Medium	60-70%	1%	CSV
Tight	45-55%	0.1%	CSV

Performance Measurements

Need to measure b-tag and mistag efficiency (versus P_T and η) in data to be used in physics analyses like top, Higgs, Susy, etc.

- b-tagging efficiency: measurement performed in jet samples enriched in b jets
  - Multijet events with a jet containing a soft muon (muon-jet events)
  - ttbar events
- Mistag efficiency



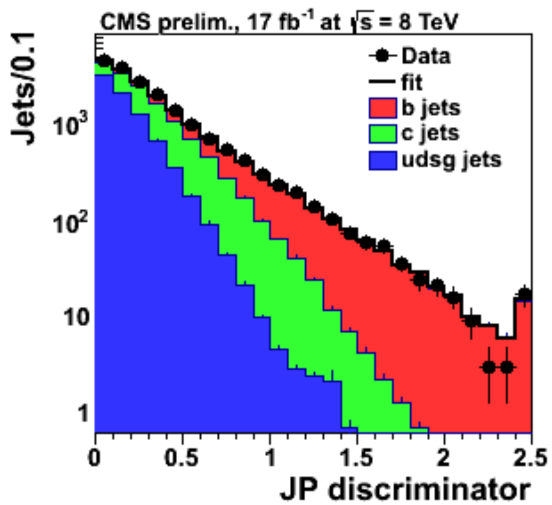


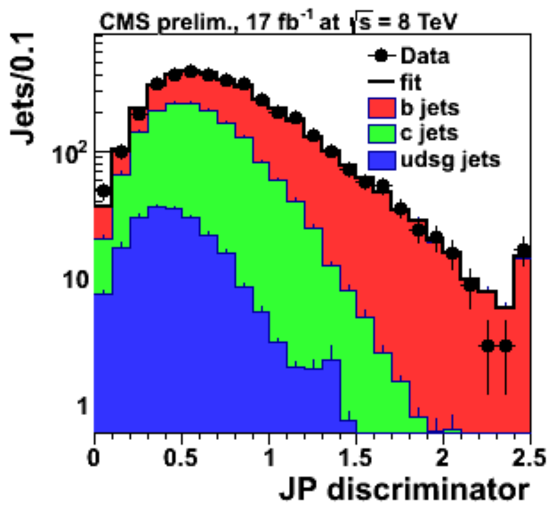
→ Derived scale factors ($\varepsilon^{Data}/\varepsilon^{MC}$) for recommended taggers : JP and CSV @ Loose, Medium and Tight working points and TCHP with Tight selection

b-tagging efficiency in muon-jet events

Different methods:

Method	P_T range
Fit of muon kinematics : • P_T of the muon relative to the jet axis (P_T^{rel}) • Fit of the IP of the muon (IP3D)	$20 \leq P_T \leq 120$ GeV $120 \leq P_T \leq 800$ GeV
Solve a 8-equation system with 8 unknowns (System8) with 3 different tagging (P_T^{rel}, tag of muon-jet, tag of away jet)	$20 \leq P_T \leq 120$ GeV
Fit of Jet Probability (LT method)	$20 \leq P_T \leq 800$ GeV





(left) before and

(right) after applying the CSVM tagger

[here jet P_T>600 GeV]

$\varepsilon_b^{tag} = N_{b\text{-}jet}^{tagged} / N_{b\text{-}jet}^{total}$

b-tagging efficiency in muon-jet events

Combination of all the methods with the BLUE method

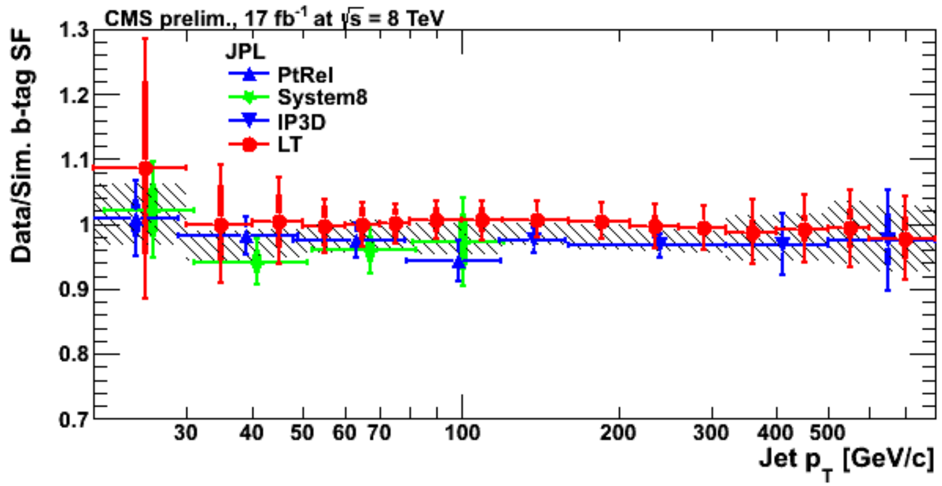
Treatment of systematics:

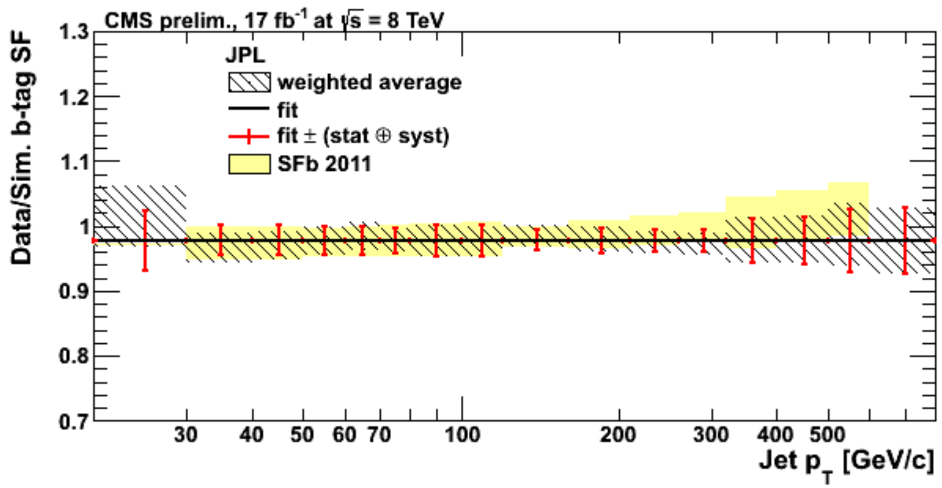
- Common (PU, gluon splitting, P_T^{μ}) or for 2 of them (away-jet tagger)
  → 100% correlated or anti-corr
- Other specific to 1 method: uncorr

The event overlap has been taken into account in the combination.

Results:

- SF_b parameterized vs P_T
- $\sigma[SF_b] \approx$ 3-5% for P_T< 320 GeV and $\sigma[SF_b] \approx$ 9% for P_T> 320 GeV @ 7 TeV
- Similar or a bit better @ 8 TeV



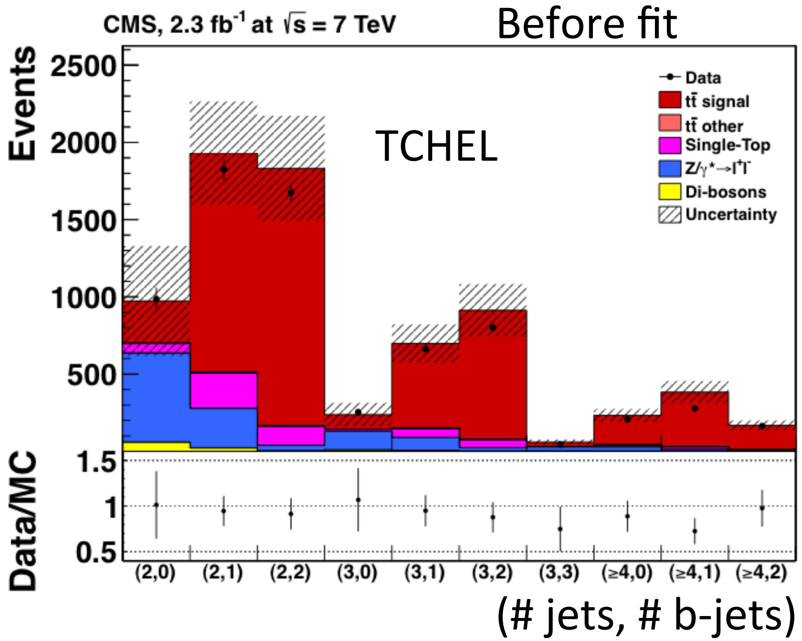


b-tagging efficiency in ttbar events

Different methods (event Selection follows closely the x-section analyses):

Dilepton channel

@7TeV only: Profile Likelihood Ratio (PLR)



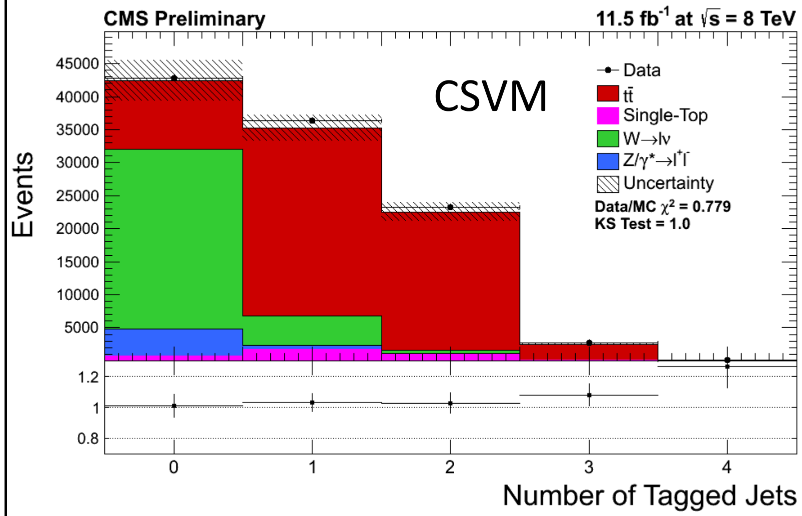
new @8 TeV: LT ttbar

Flavour Tag Matching (FTM)

l+jets channel

Flavour Tag Consistency (FTC)

After likelihood fit of σ_{ttbar} and ε_b



b-enriched Sample (bSample)

b-tagging efficiency in ttbar events

Combination with the BLUE method:

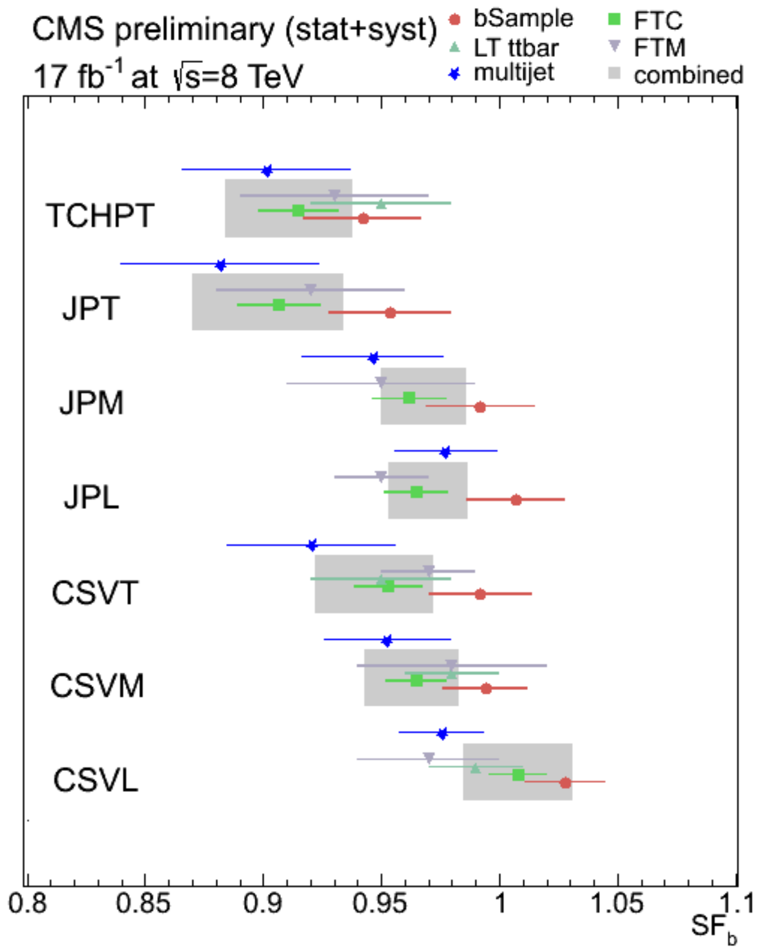
- @ 7 TeV: only PLR (dilepton) and FTC (lepton+jet)
- @ 8 TeV: 4 available methods

Systematics: mostly common to all methods (dominant: JES, factorisation, ttbar x-sect, …)

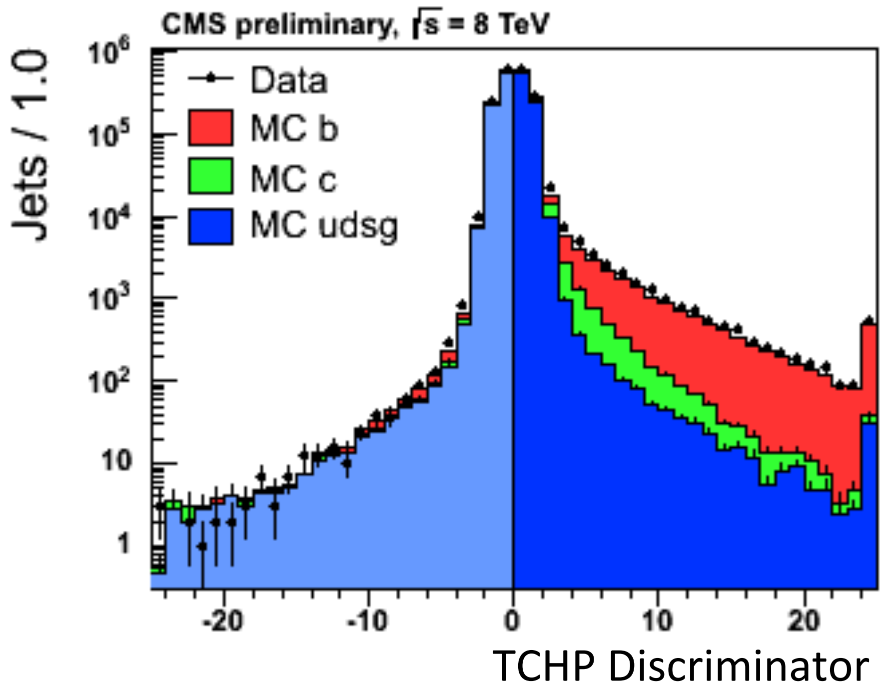
→ Treated as fully correlated

Results : $\sigma[SF_b] \approx \pm 3\%$ @7 TeV

Compared to muon-jet measurements, in the P_T range of ttbar events: good agreement!



Mistag Measurement



Mistag = Misidentification probability for light-parton jets

Measurement on multi-jet events enriched in light flavours, when using a "negative tagger" (i.e. inverted tagging algorithm)

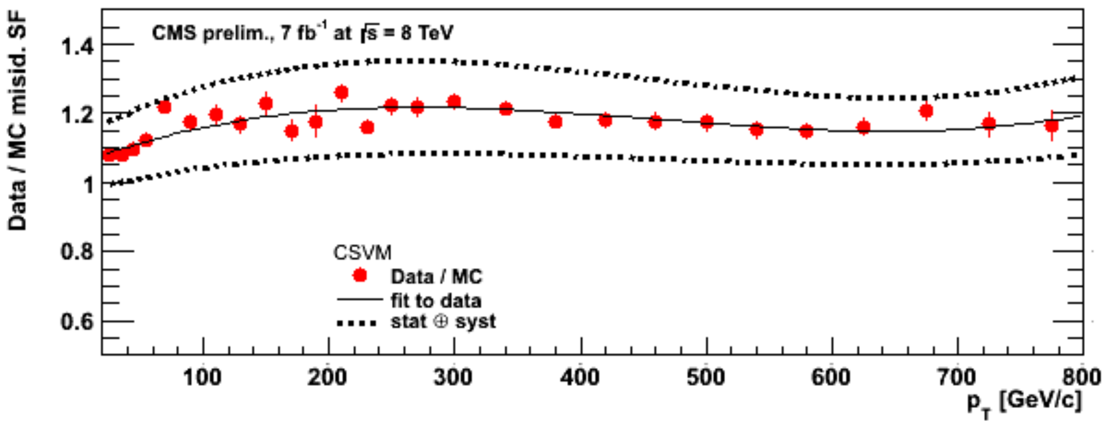
Correction of ε^{Mistag} in data = $\varepsilon^{-} * R_{light}$

with $R_{light} = \varepsilon^{Mistag}$ in MC / ε^{-} in MC

Systematics: event topology, b/c fractions, Long-lived K^0_S and Λ decays, Sign flip, etc

Results:

- vs P_T and η
- $\sigma[SF_{light}] \approx$ 9-17% @7 TeV



French contribution

~15% of people working in the CMS b-tagging group

French activities on :

- Commissioning of b-tagging algorithms, Data/MC comparisons with multijets and ttbar events
- Calibration of the JP tagger
- Validation of Data reprocessing and CMS software releases
- Performance measurements in data :
  - b-tagging efficiencies in muon jets events (LT)
  - b-tagging efficiencies in ttbar events (PLR @7 TeV)
  - Mistag measurements
- Starting works:
  - high P_T jets
  - algorithm optimization for the HLT trigger

Conclusions

- Variety of algorithms developed by CMS to be used in analysis
  - @ start : simple and robust techniques (TC and SSV)
  - Now : algorithm with better performance (JP and CSV)
  - High degree of agreement data/simulation + robustness against variations in the running conditions.
- Performance measurements in data
  - Extraction of SF_b by different methods in multijet and ttbar events, overall good agreement between them!
  - SF_{light} measured by applying inverted tagging algorithms to the multijet events
  - Measurements over a large range in P_T, versus P_T → σ[SF_B] ≈ 2-3% in the Top and Higgs kin. regimes and ~5-10% up to P_T~800 GeV for new physic searches
- b-tagging used in >50 published CMS papers (even for Heavy Ions).

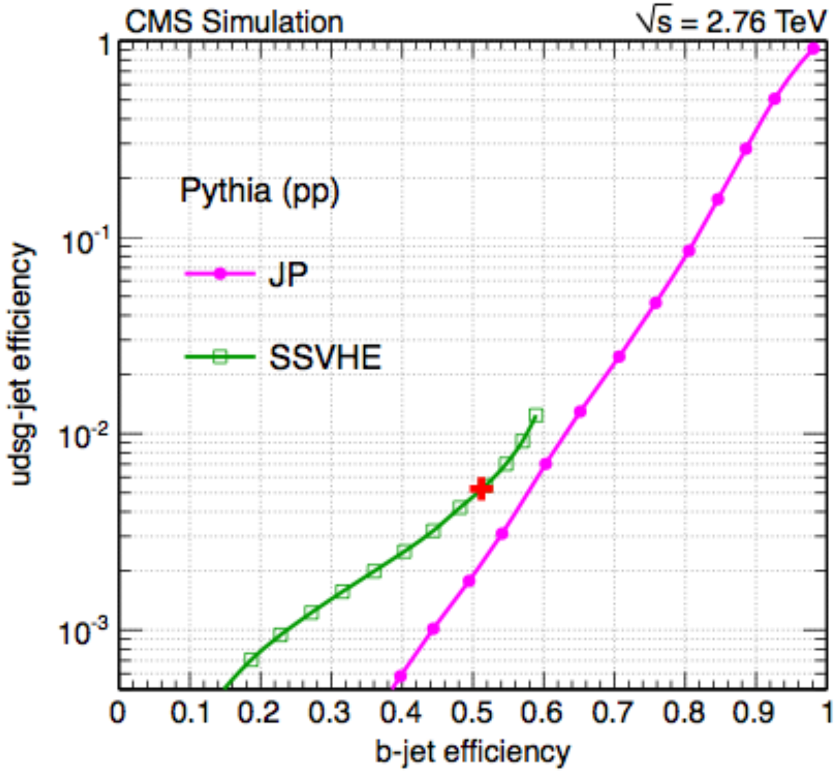
The French community is well involved in this effort, which is a key element for top studies, as well as Higgs and New Physics search.

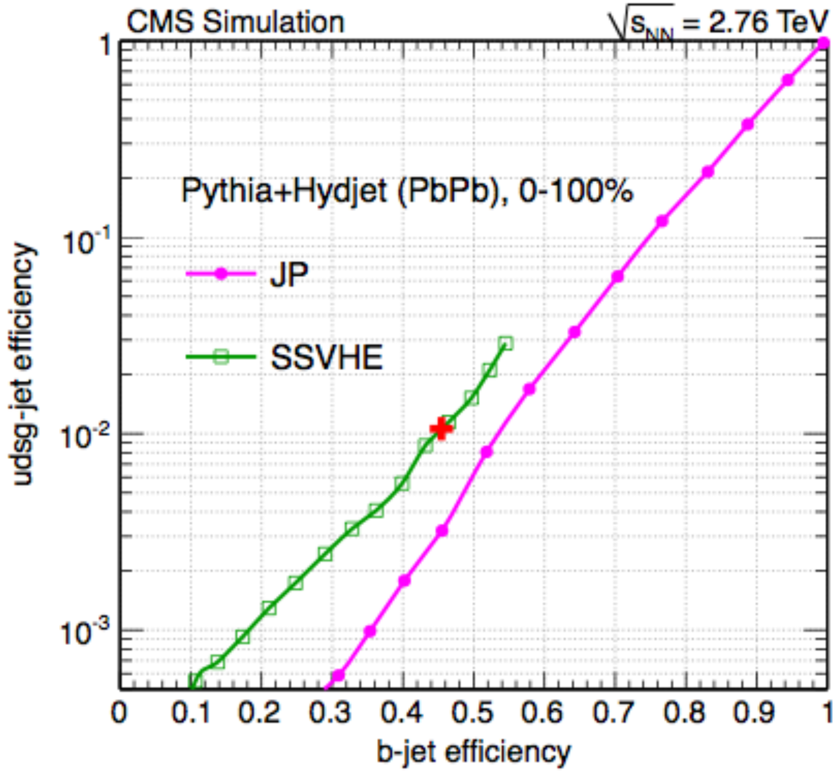
References

- https://twiki.cern.ch/twiki/bin/view/CMSPublic/TWikiBTV_Moriond2013 @8 TeV
- Identification of b-quark jets with the CMS experiment, CMS Collab., CERN-PH-EP-2012-262, arXiv:1211.4462. (accepted by JINST) @7 TeV

b-tagging in Heavy Ions

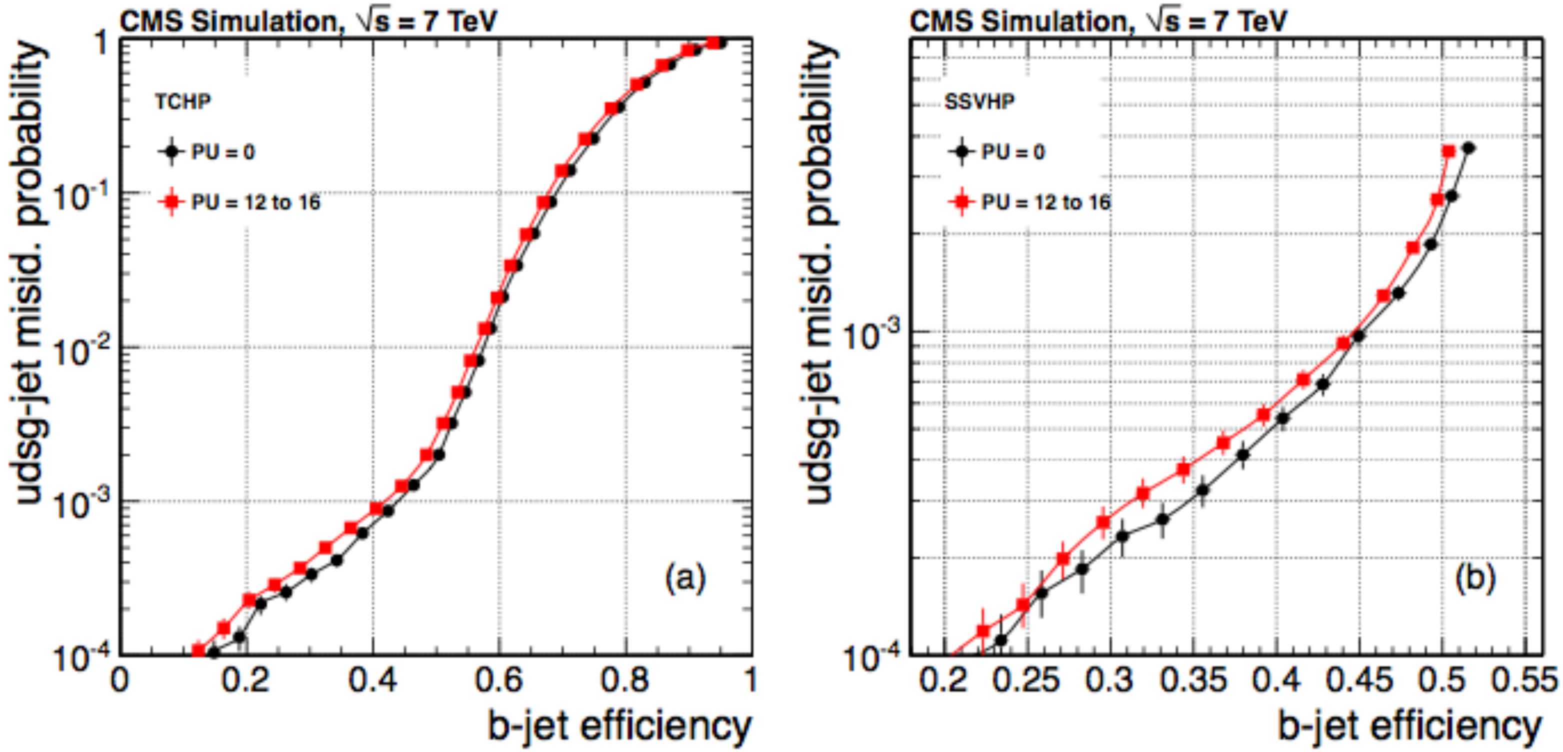
- CMS HIN-12-003 : Measurement of the b-jet to inclusive jet ratio in PbPb and pp collisions at $\sqrt{s_{NN}}$ = 2.76 TeV with the CMS detector





Running conditions

- Misalignment studies: With the current estimated accuracy of the positions of the active elements, no significant deterioration is observed with respect to a perfectly aligned detector.
- PU: The changes on performance are small and concentrated in the regions of very high purity.



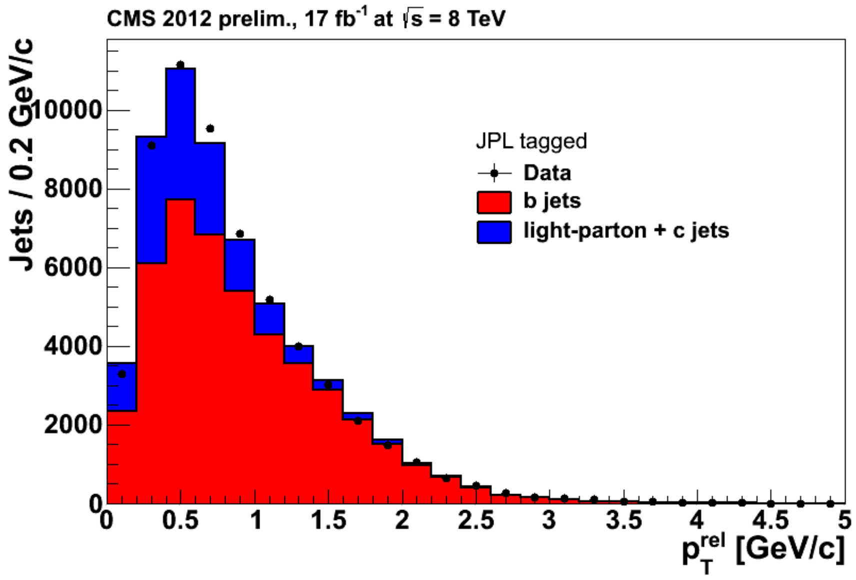
CSV information

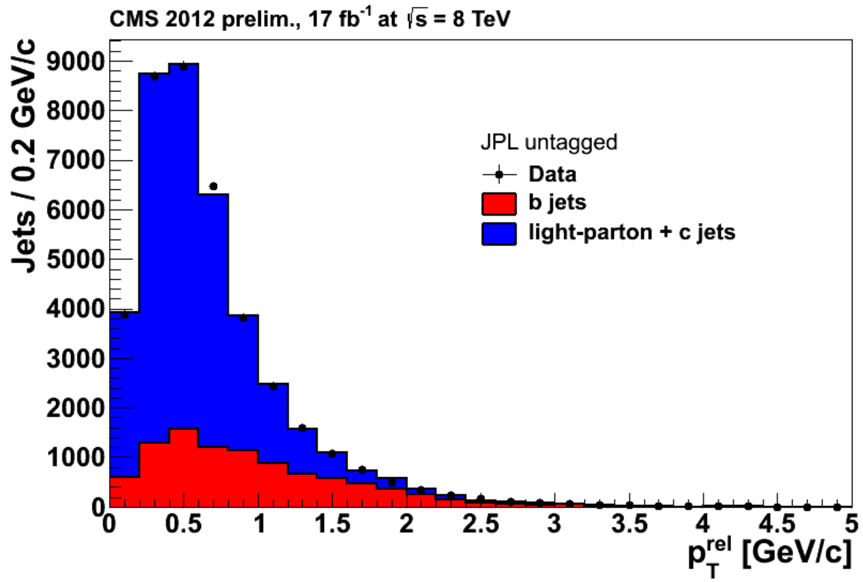
The following set of variables with high discriminating power and low correlations is used (in the "no vertex" category only the last two variables are available):

- the vertex category (real, "pseudo," or "no vertex");
- the flight distance significance in the transverse plane ("2D"); the vertex mass; the number of tracks at the vertex; the ratio of the energy carried by tracks at the vertex with respect to all tracks in the jet; the η of the tracks at the vertex with respect to the jet axis; the 2DIP significance of the first track that raises the invariant mass above the charm threshold of 1.5GeV/c2 (tracks are ordered by decreasing IP significance and the mass of the system is recalculated after adding each track);
- the number of tracks in the jet; the 3D IP significances for each track in the jet.

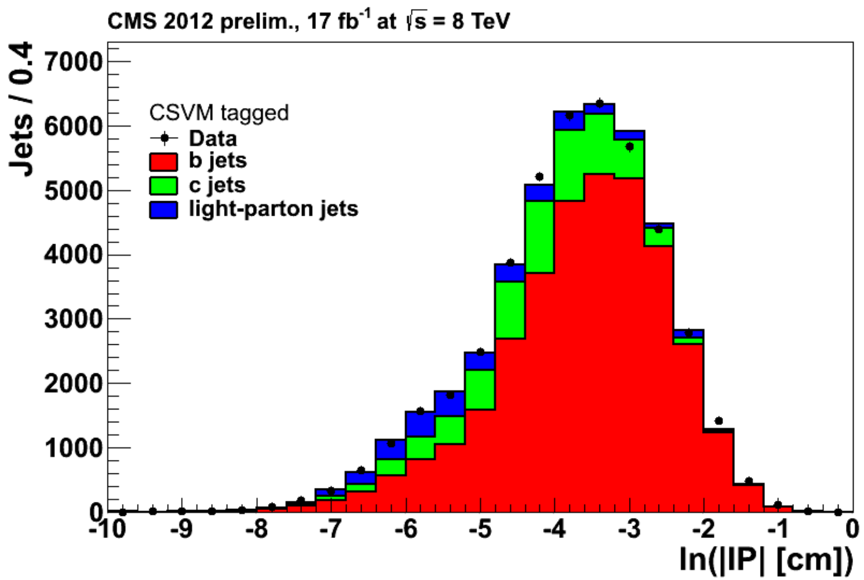
Two likelihood ratios are built from these variables. They are used to discriminate between b and c jets and between b and light-parton jets. They are combined with prior weights of 0.25 and 0.75, respectively.

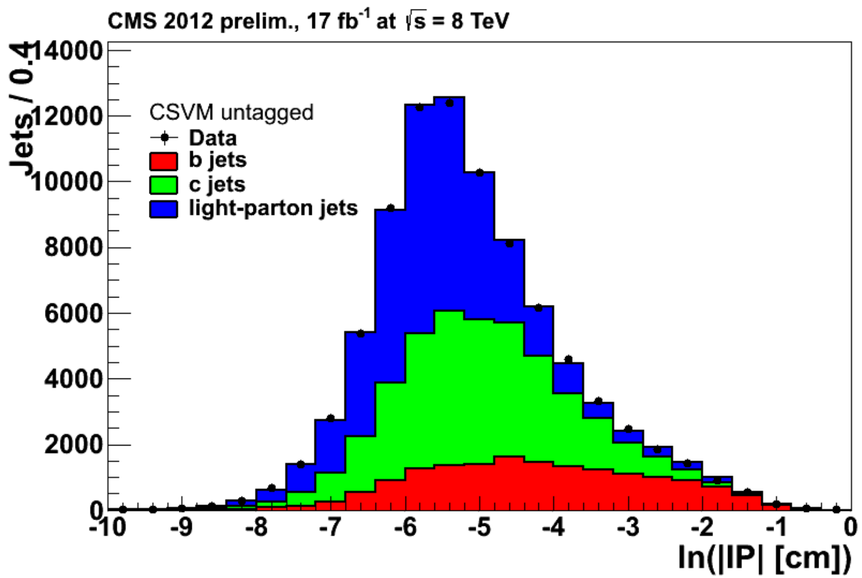
Performance methods: PtRel





Performance methods: IP3D





Performance methods: System8

$$n = n_b + n_{cl}$$ muon-in-jet + away-jet

$$p = p_b + p_{cl}$$ muon-in-jet + tagged-away-jet

$$n^{\mathrm{tag}} = \varepsilon_b^{\mathrm{tag}} n_b + \varepsilon_{cl}^{\mathrm{tag}} n_{cl}$$

$$p^{\mathrm{tag}} = \beta\, \varepsilon_b^{\mathrm{tag}} p_b + \alpha\, \varepsilon_{cl}^{\mathrm{tag}} p_{cl}$$

apply “probe” tagger

$$n^{\mathrm{pTrel}} = \varepsilon_b^{\mathrm{pTrel}} n_b + \varepsilon_{cl}^{\mathrm{pTrel}} n_{cl}$$

$$p^{\mathrm{pTrel}} = \delta\, \varepsilon_b^{\mathrm{pTrel}} p_b + \gamma\, \varepsilon_{cl}^{\mathrm{pTrel}} p_{cl}$$

apply “tag” tagger

$$n^{\mathrm{tag,pTrel}} = \kappa_b\, \varepsilon_b^{\mathrm{tag}} \varepsilon_b^{\mathrm{pTrel}} n_b + \kappa_{cl}\, \varepsilon_{cl}^{\mathrm{tag}} \varepsilon_{cl}^{\mathrm{pTrel}} n_{cl}$$

$$p^{\mathrm{tag,pTrel}} = \kappa_b\, \beta\, \delta\, \varepsilon_b^{\mathrm{tag}} \varepsilon_b^{\mathrm{pTrel}} p_b + \kappa_{cl}\, \alpha\, \gamma\, \varepsilon_{cl}^{\mathrm{tag}} \varepsilon_{cl}^{\mathrm{pTrel}} p_{cl} .$$

apply “tag” and “probe” taggers

LHS: observables, GREEK: correlation factors from MC

Systematic uncertainties on ε_b [@7TeV]

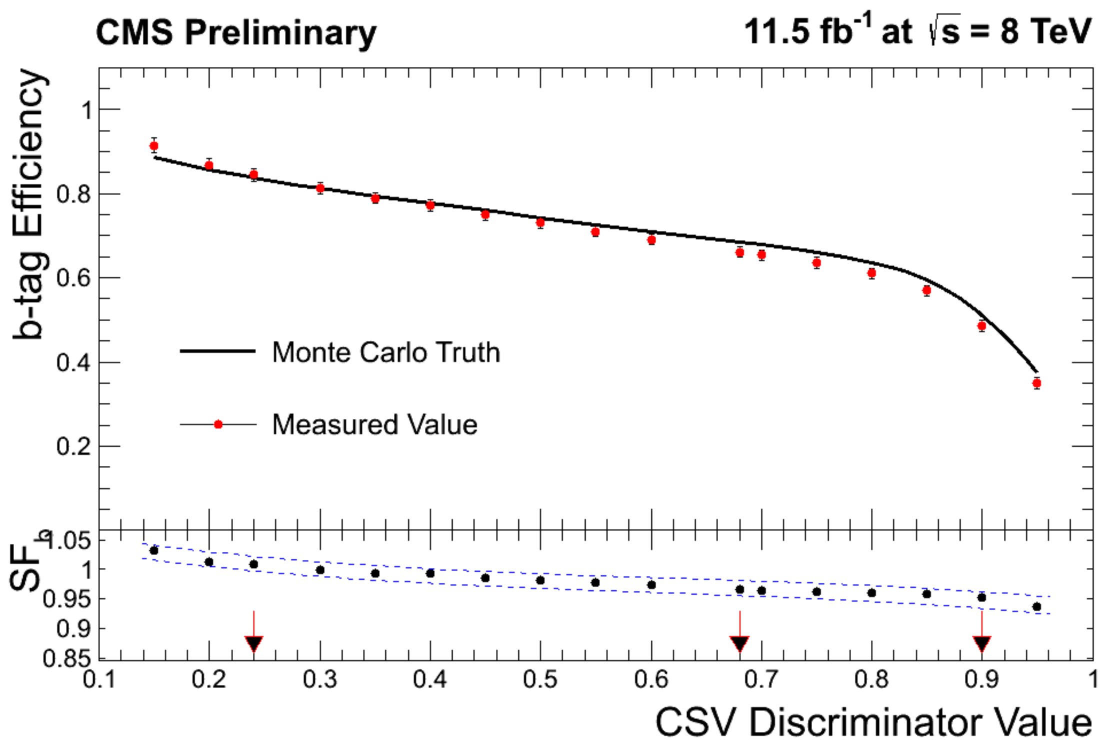
List of systematic uncertainties :

- Common for all methods (PtRel, IP3D, System8, LT): Pileup, Gluon splitting, $P_T(\mu)$
- Common for PtRel, IP3D and System8 methods: Away-jet tagger
- For PtRel only: Ratio of light-parton to charm jets in simulation
- For System8 only: Selection on prel, MC closure test
- For LT only: Fraction of b jets with JP information, Difference between muon jets and inclusive jets, Bias for the JP and JBP taggers.

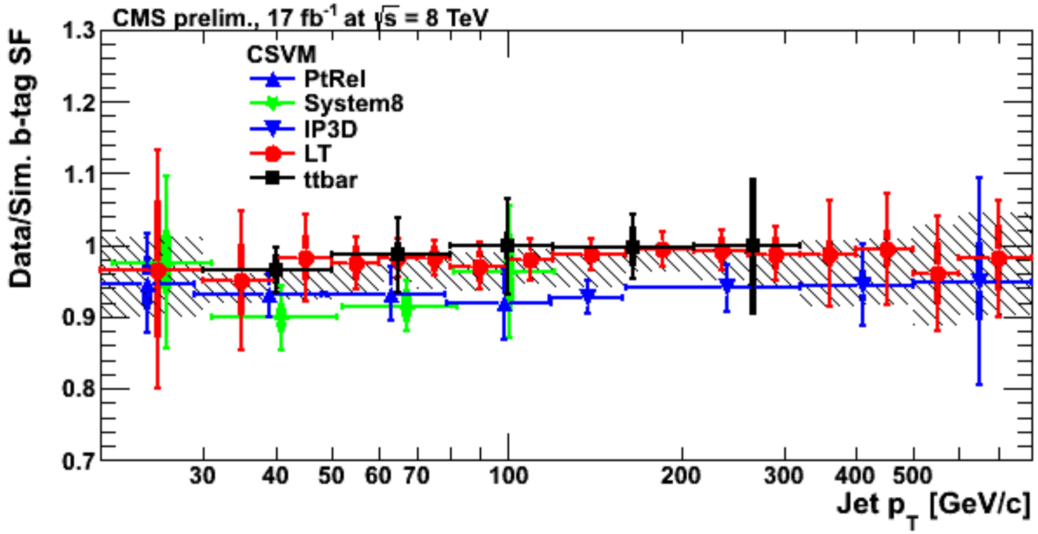
Average systematic uncertainty on the data/MC scale factors in a dedicated P_T bin:

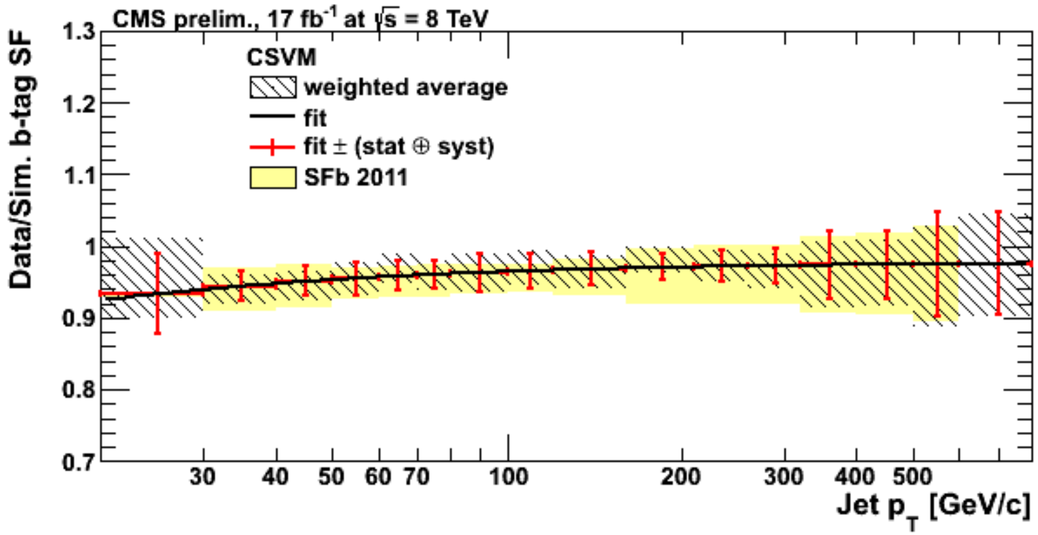
- average uncertainty is about 3% for the PtRel method ($80 < P_T < 120$ GeV/c)
- 6–10% for the System8 method ($80 < P_T < 120$ GeV/c)
- 3–4% for the IP3D method ($160 < P_T < 320$ GeV/c)
- 2–7% for the LT method ($160 < P_T < 320$ GeV/c)

b-tag eff. SF_b in ttbar with FTC



b-tag eff. SF_b vs jet p_T





A brief glossary of the used acronyms is given below:

bSample Method to measure the b-tagging efficiency in $t\bar{t}$ events from a b-enriched lepton+jets sample

CSV Combined Secondary Vertex algorithm, based on secondary vertex and track-based lifetime informations

CSVL, CSVM, CSVT CSV algorithm at the loose, medium, tight operating points

FTC Flavour Tag Consistency method for the measurement of the b-tagging efficiency in $t\bar{t}$ events with lepton+jets

FTM Flavour Tag Matching method for the measurement of the b-tagging efficiency in $t\bar{t}$ events with dileptons

IP3D Method for the measurement of the b-tagging efficiency in multijet events based on the 3D impact parameters of muons

JP Jet Probability algorithm, based on the likelihood of tracks to come from the primary vertex (using the impact parameter significance values)

JPL, JPM, JPT Jet Probability algorithm at the loose, medium, tight operating points

LT Lifetime Tagging method for the measurement of the b-tagging efficiency in multijet events, based on template fits to the JP or CSV distributions

PtRel Method for the measurement of the b-tagging efficiency in multijet events based on the transverse momenta of muons w.r.t. the jet axis

System8 Method for the measurement of the b-tagging efficiency in multijet events with a muon, solving a system of equations

TCHP Track Counting High Purity algorithm, based on the impact parameter significance of the third track, ordered in decreasing values.

TCHPT Track Counting High Purity algorithm at the tight operating point