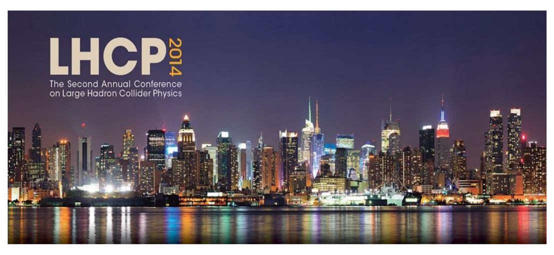
Performance of b-tagging at CMS with pp collision data at $\sqrt{s}$ = 8 TeV

Nitish Dhingra[1,2] (on the behalf of CMS collaboration)

[1] Brown University, USA

[2] Panjab University, India

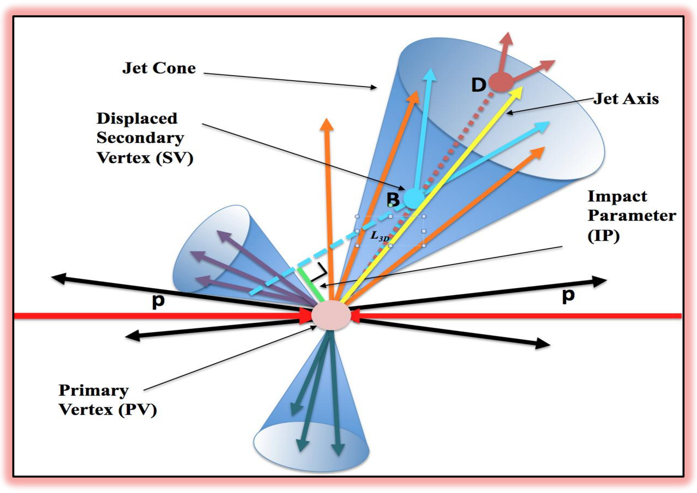


Abstract: The identification of jets originating from b-quarks is crucial both for the searches for new physics and for the measurement of standard model processes. CMS has developed a variety of algorithms to select b-quark jets based on variables such as impact parameter of charged particle tracks, properties of reconstructed decay vertices, and the presence or absence of a lepton, or combinations thereof. The studies on these algorithms and their performances are presented, using multijet events and ttbar events recorded in proton-proton collision data at √s=8 TeV with the CMS detector in the LHC Run 1.

Introduction

- Identification of jets originating from b-quarks is the "key" for new physics searches and for the measurements of standard model processes
  - Higgs→ bbar searches, Analyses involving top quarks, SUSY….
  - Also a crucial tool to suppress ttbar background in many physics channels
- b-tagging or b-jet tagging is a reconstruction technique to determine whether the jet originated from a bottom quark

- Relies on distinct properties of b-hadrons
  - Large proper lifetime (τ≈1.5 ps, cτ ~450 μm, mean path length ≈ 5 mm@50 GeV)
  - Relatively large mass (~ 5 GeV)
  - Leptons from semi-leptonic b-hadron decays have large transverse momentum w.r.t. jet axis
  - Decay to final states having high charged track multiplicities (~5 on the average)



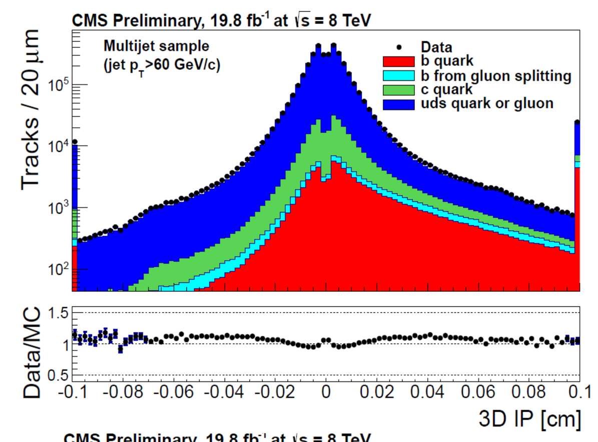
- b-tagging can be done on the basis of
  - Track information
  - Secondary vertex information
  - Soft lepton information
  - Some combination of the above

- MC simulation known not to completely reproduce b-tagging performance in data
  - Apply Data/MC scale factors to simulated events to take differences into account

b-Tagging Observables

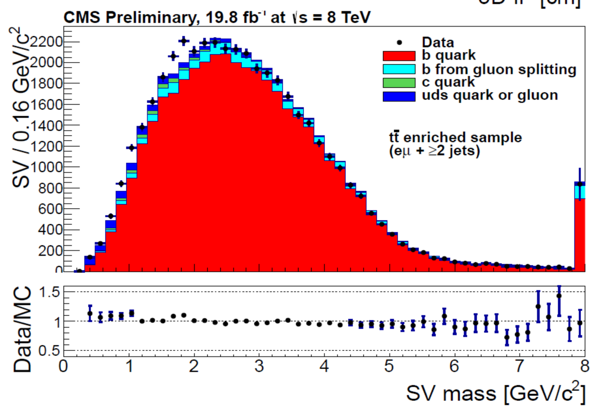
Impact Parameter (IP)

- The distance between the primary interaction vertex (PV) and track at the point of closest approach
- IP is calculated in 3 dimensions – thanks to good x-y-z resolution provided by the pixel detector
- IP significance = IP/σ(IP)
  - ✓ Positive, if $\vec{IP} \cdot \vec{jet} > 0$
  - ✓ Negative, if $\vec{IP} \cdot \vec{jet} < 0$
- IP significance is expected to be symmetric for light quarks (u, d, s) or gluons (g) while more positive for weakly decaying b-hadrons
- Use negative of IP significance distribution to extract the probability density function for tracks not originating from b/c-jets



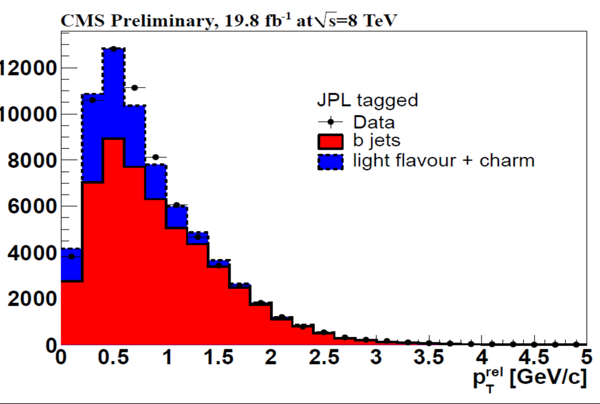
Secondary Vertex (SV)

- High resolution of CMS tracking system allows the reconstruction of the Secondary Vertex (SV) – the point where b-hadron decays



Lepton

- Muon or electron coming from semi-leptonic b-hadron decay can also be used to tag b-jets



b-Tagging Algorithms

Track Counting Algorithm

- Identifies a b-jet if there are at least "N" tracks each with an IP significance above a given threshold
- Tracks in a jet are sorted by decreasing values of IP significance and discriminator = IP sig. ("N[th]" track)
- "High Efficiency (TCHE)" & "High Purity (THCP)" versions use IP sig. of 2[nd] & 3[rd] track, respectively

Jet Probability (JP) Algorithm

- As a natural extension, JP algorithm combines information from all the selected tracks in the jet
- Computes a "likelihood" for tracks to originate from the PV
- Jet B Probability (JBP) works in a similar way but giving more weight to the four most displaced tracks
  - ✓ Matches the average number of reconstructed charged particles from b-hadron decays

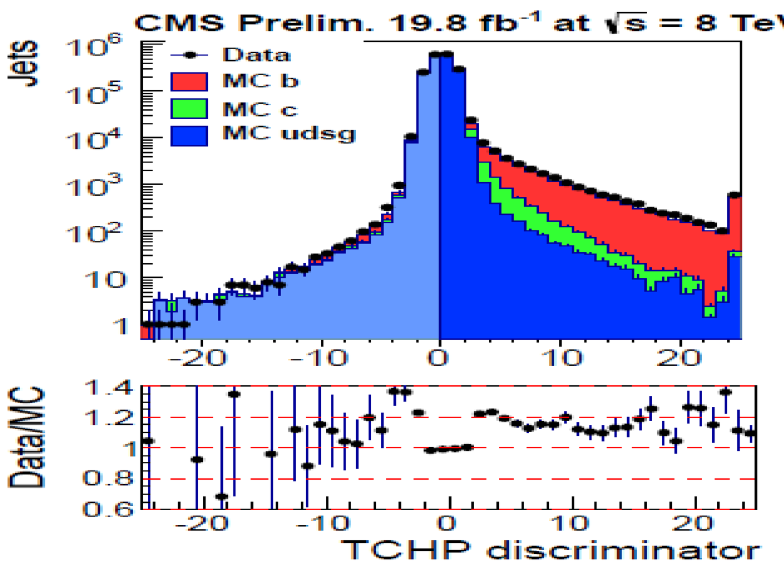
Soft-Lepton Tagging Algorithms

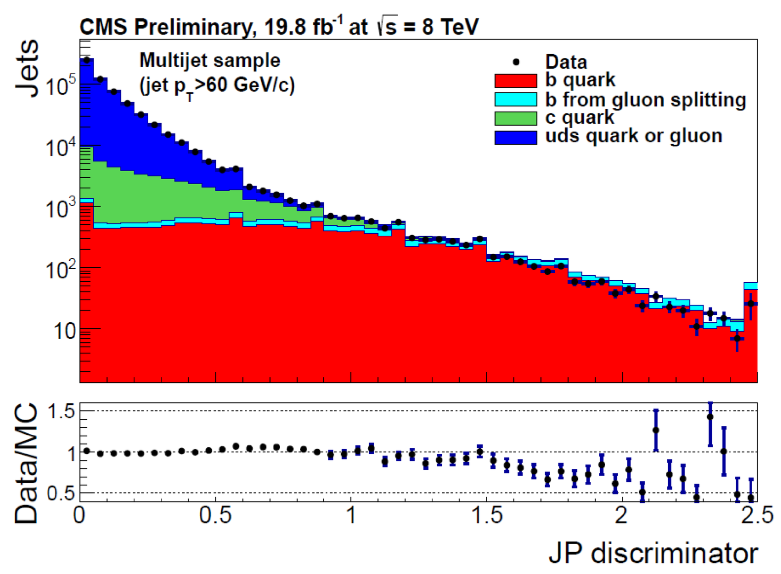
- Rely on the properties of muons or electrons from semileptonic b-hadron decay

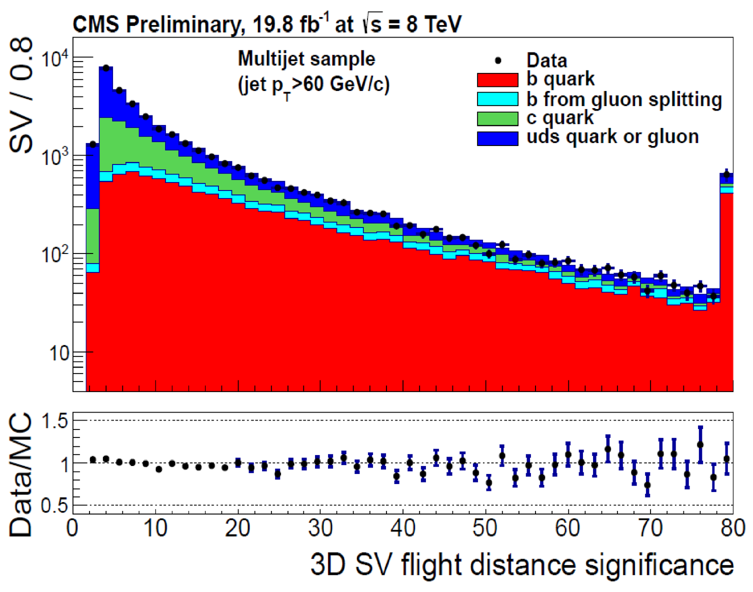
Secondary Vertex Tagging Algorithms

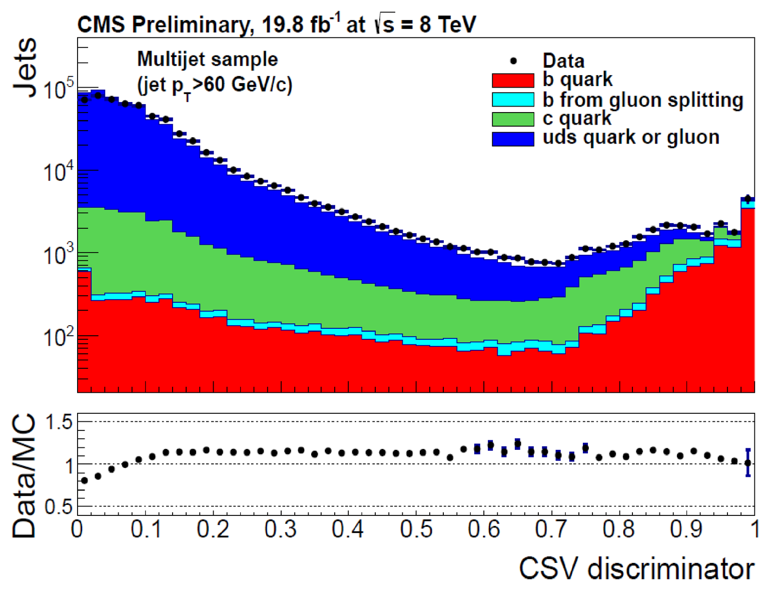
- **Simple Secondary Vertex (SSV)**
  - ✓ Relies on the reconstruction of ≥1 SV and uses 3D flight distance sig. as a discriminating variable
  - ✓ Two versions: "High Efficiency" (SSVHE with N_{tracks} ≥2) & "High Purity" (SSVHP with N_{tracks} ≥3)
- **Combined secondary vertex (CSV)**
  - ✓ Based on secondary vertex and track-based lifetime information
  - ✓ Provides discrimination even when no secondary vertices are found
  - ✓ Purity of sample is determined from the mass of reconstructed charged particles associated to the SV

b-tagging significantly enhances the sensitivity of jet substructure techniques for boosted topologies





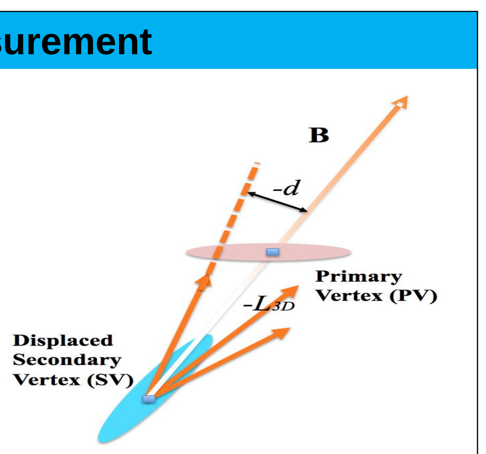




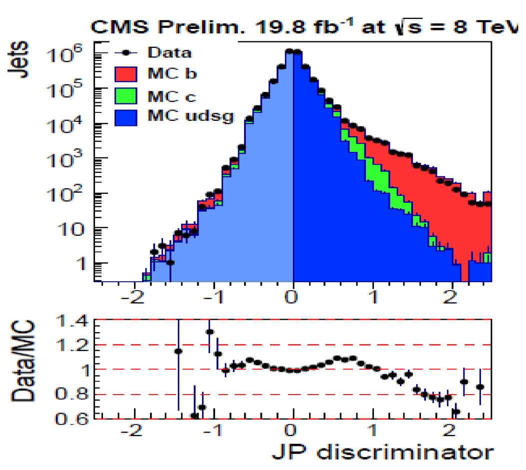
Misidentification Probability Measurement

- It is the probability or rate of light-flavour quarks and gluons to get misidentified as b-jets

- Evaluated with negative tagging algorithms



- To first order expect symmetric discriminator values b/w negative and positive taggers for light-flavour jets
  - ✓ Derive mistag rate "ϵ^-" from negative -tagged jets in inclusive jet data sample



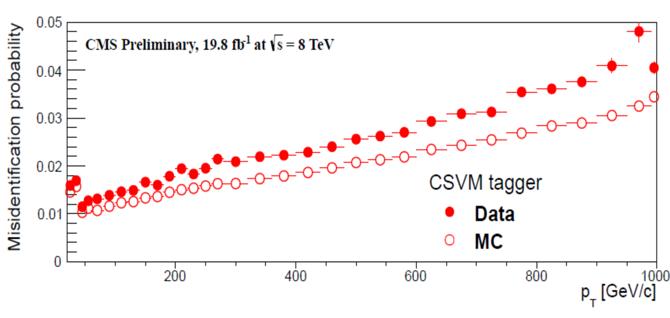
- Evaluate a correction factor from simulation to account for second-order asymmetries in negative & positive tag rates of light -flavour quarks & gluon jets and for heavy flavour contribution to the negative taggers

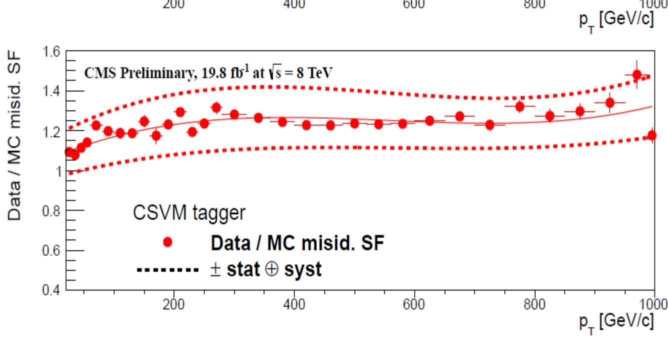
- Get the mistag rates as:

$$\varepsilon_{data}^{mistag} = \varepsilon_{data}^{-} \cdot \varepsilon_{MC}^{mistag} / \varepsilon_{MC}^{-}$$

80 GeV< jet p_T <120 GeV

b tagger	misidentification probability	SF_{light}
JPL	0.0944 ± 0.0004	$1.03 \pm 0.01 \pm 0.07$
CSVL	0.0990 ± 0.0004	$1.10 \pm 0.01 \pm 0.05$
JPM	0.0105 ± 0.0002	$1.10 \pm 0.02 \pm 0.20$
CSVM	0.0142 ± 0.0002	$1.17 \pm 0.02 \pm 0.15$
TCHPT	0.0026 ± 0.0001	$1.27 \pm 0.06 \pm 0.27$
JPT	0.0013 ± 0.0001	$1.11 \pm 0.07 \pm 0.31$
CSVT	0.0016 ± 0.0001	$1.26 \pm 0.07 \pm 0.28$





b-tagging Efficiency Measurement

- Measured from data using a control sample of jets enriched with heavy flavour content (muon-jets)
  - Require a soft-muon inside the jet within a cone ΔR <0.4 around jet-axis
  - Require the event to have another tagged-jet ("away tag") to reduce the light-flavour background
  - ttbar events can also be used but due to event kinematics they cannot cover efficiency in high p_T range

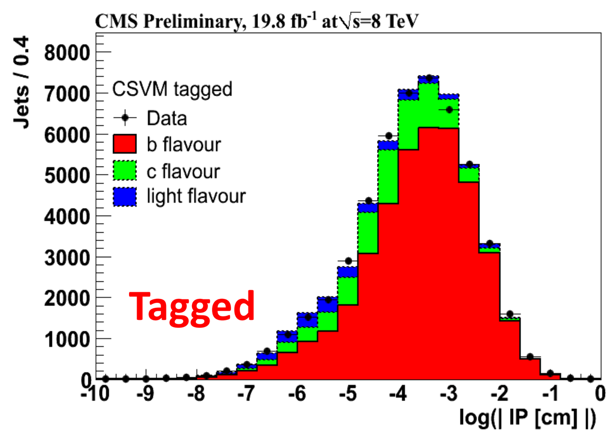
Method based on kinematic properties of muon-jets

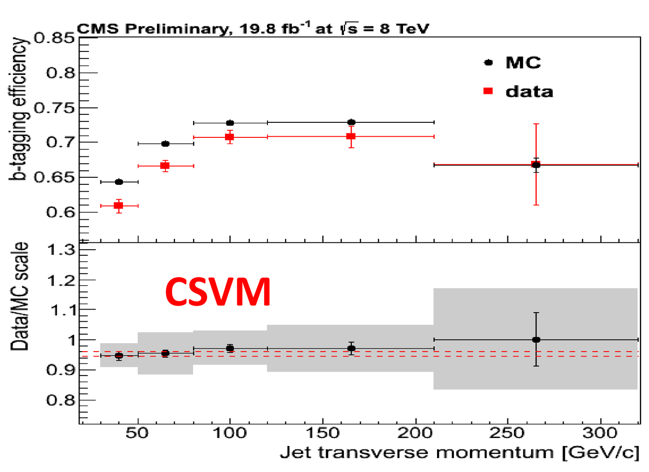
- Separate muon-jets into tagged & untagged subsamples by a discr. working point whose efficiency is to be measured
- For the two subsamples separately, fit the spectra of muon-jets p_T^{rel} or IP3D using templates of b, c and udsg jets derived from simulation or inclusive jet data. The p_T^{rel} is the transverse momentum of muon relative to the jet axis
- Calculate fraction of b-jets in the two subsamples: f_b^{tag}, f_b^{untag} and efficiency as:

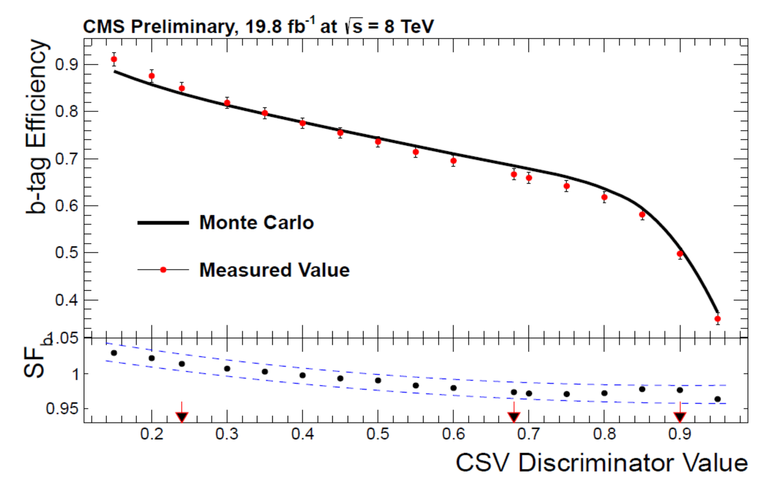
$$\varepsilon_b^{tag} = \frac{f_b^{tag} \cdot N_{data}^{tag}}{f_b^{tag} \cdot N_{data}^{tag} + f_b^{untag} \cdot N_{data}^{untag}}$$

In addition, Lifetime Tagging (LT) method and System8 methods are also implemented

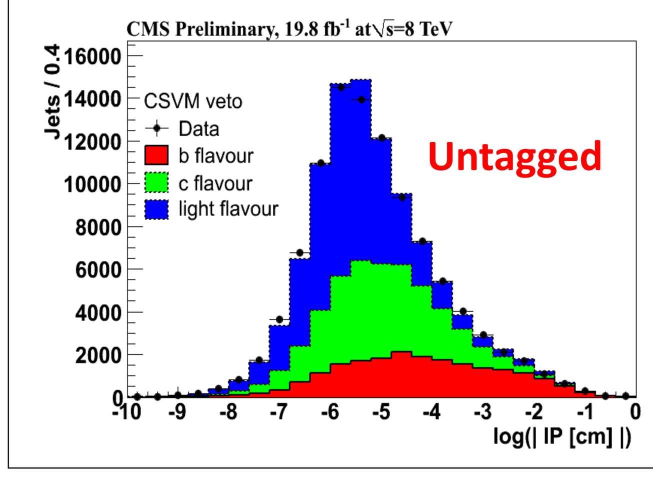
- LT method can be performed on both inclusive jets and muon jet samples
- System8 method is based on muon-jet sample but it makes a minimal use of MC information

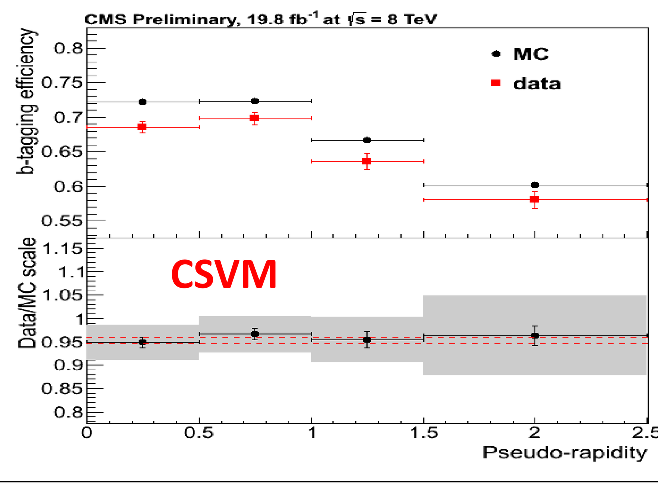


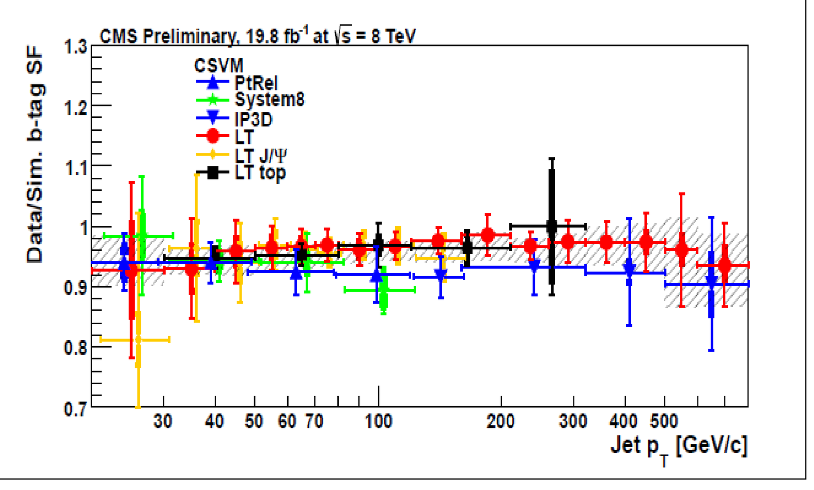




160 GeV< muon-jet p_T <320 GeV







References:

[1] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", JINST 8 P04013 (2013)

[2] CMS Collaboration, "Performance of b tagging at √s=8TeV in multijet, ttbar and boosted topology events", CMS PAS BTV-13-001 (2013)

[3] CMS Collaboration, "Measurement of Tracking Efficiency", CMS PAS TRK-10-002 (2010)