Measurement of the b-tagging efficiency using events with jets containing muons

on behalf of Compact Muon Solenoid experiment

Samvel Khalatyan University of Illinois at Chicago

31 May 2011

Outline

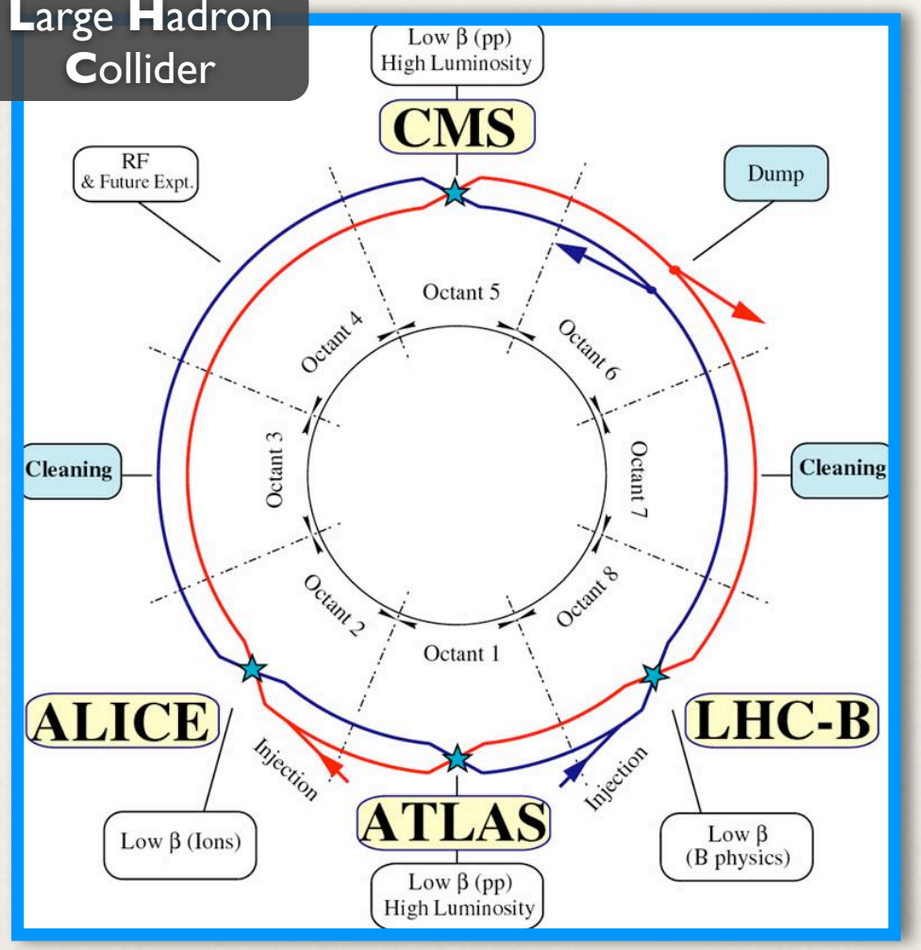
- Introduction
- b-Tagging Methods
- Efficiency Measurements
- Conclusion

Introduction

Large Hadron Collider & Compact Muon Solenoid

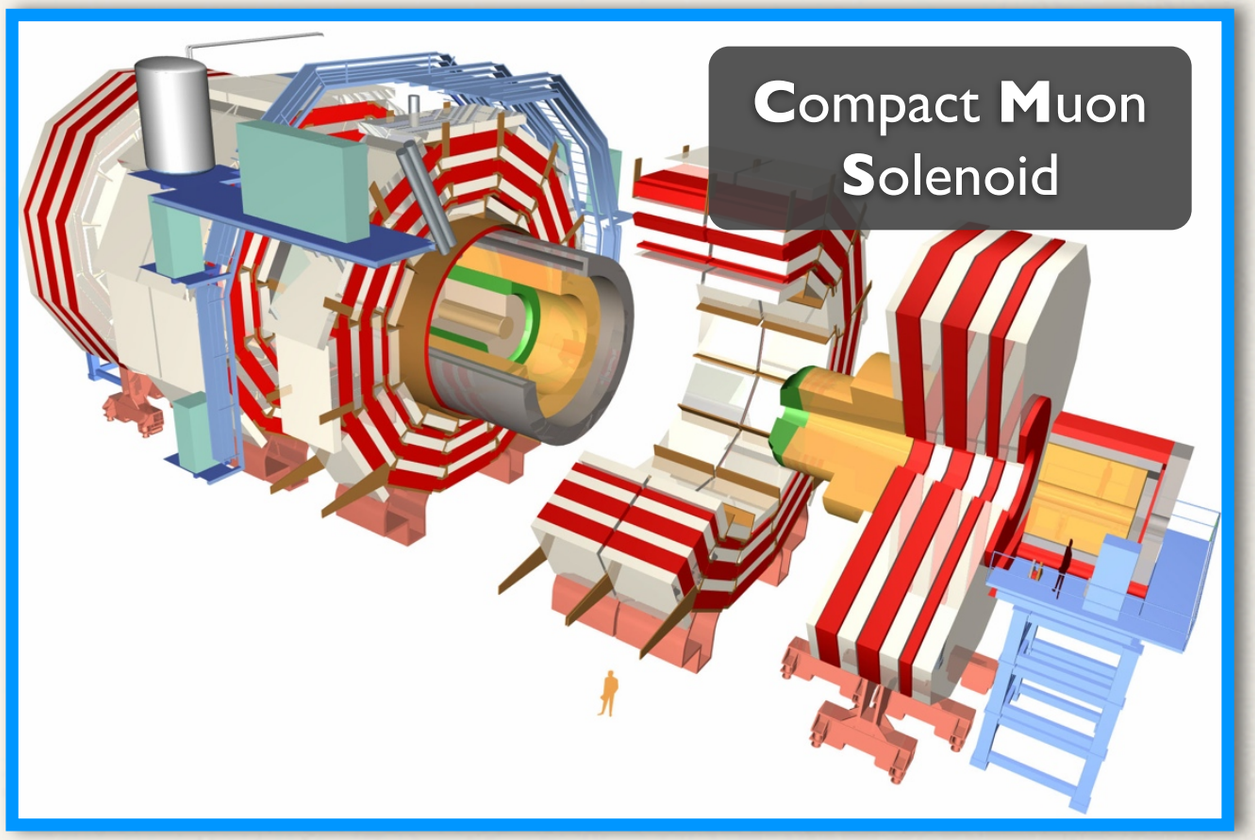
Collisions	proton-proton
2010 Energy √s	7 TeV
CMS Recorded Data	~43.17 pb^{-1}

Large **H**adron **C**ollider



(c) LHC Schematic 2010

Compact **M**uon **S**olenoid



(c) CMS Outreach 2010

strong magnet: 3.8 T

silicon tracker: pixels + strips

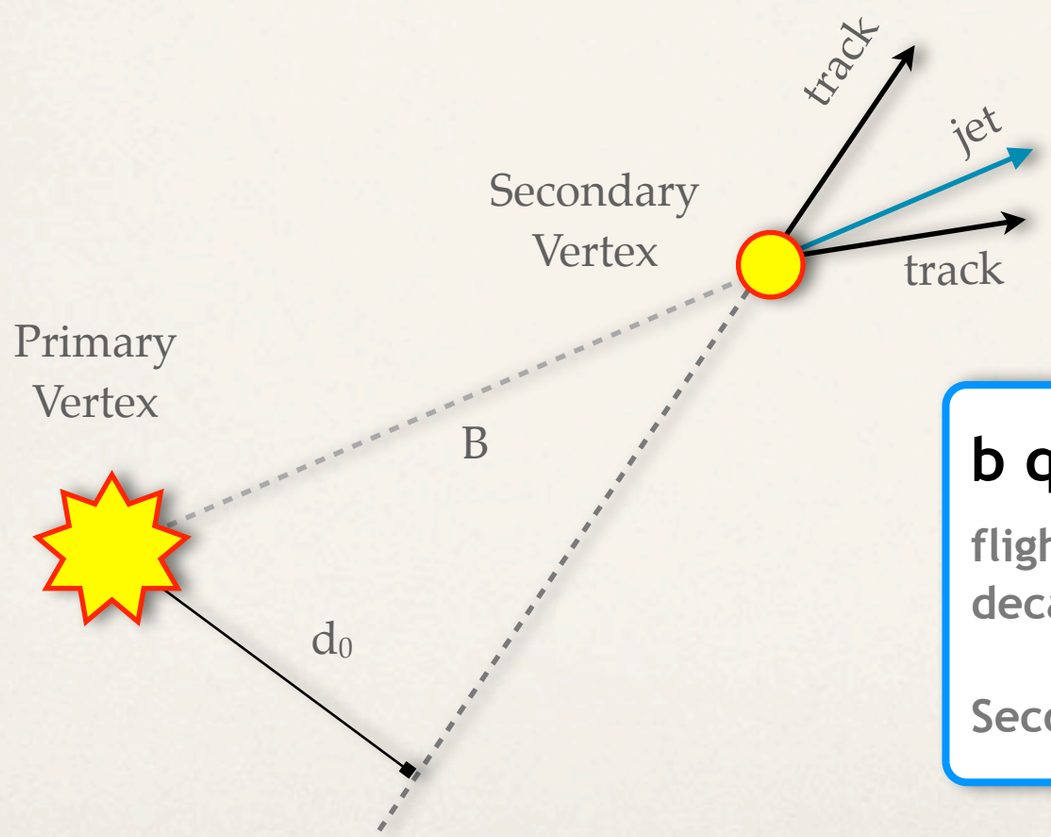
coverage $0 < \phi < 2\pi, |\eta| < 2.5$

b-Tagging

Introduction

Motivation

- often signal contains b quarks while background does not.
- high b-tagging efficiency at low mis-tag rate may significantly reduce background and therefore reduce jets combinatorics that is often used in the analysis



b quark

flight distance	~1.8mm *(at 20 GeV)* before decay
decay	weak, mostly into c-quarks *(~20% into muons)*
Secondary Vertex	high tracks multiplicity, displacement

b-Tagging

Track Counting algorithm

1. order tracks associated with jet by Interaction Point significance

2. use N^{th} track Interaction Point significance value as the b-Tag discriminant

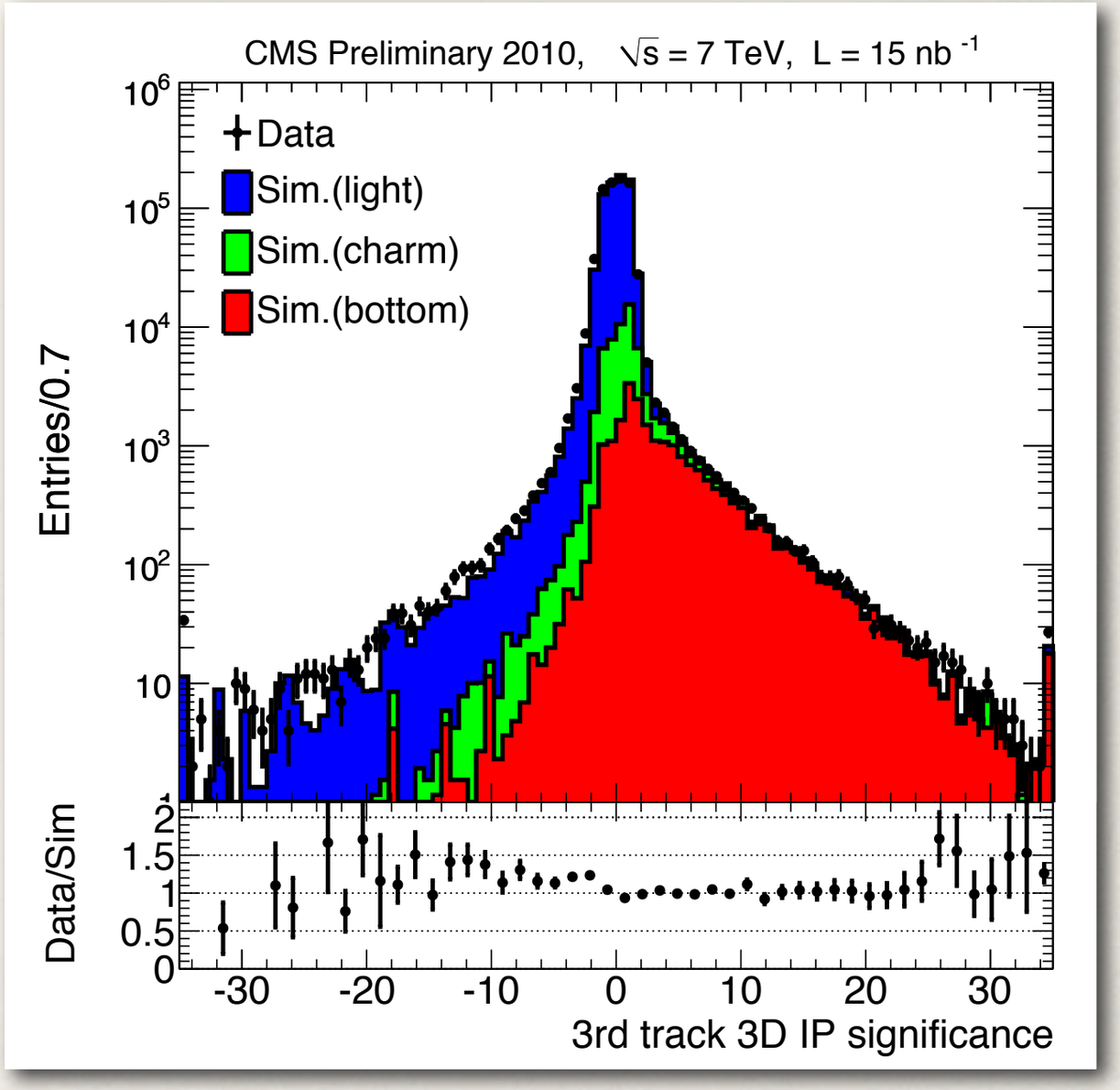
<u>Operating Points</u>

HE *High Efficiency*

use 2^{nd} track IP significance

HP *High Purity*

take 3^{rd} track IP significance



(c) CMS PAS BTV-10-001, http://goo.gl/4S207

Samvel Khalatyan - samvel@uic.edu

b-Tagging

efficiency measurements

b-Tagging algorithms are applied to **all** jets: heavy and light flavor.

Measure the efficiency of the b-jets identification (**b-Tag efficiency**) and light-jets (g, u, d, s) misidentification (**mis-Tag rate**) in data instead of MC due to differences in:

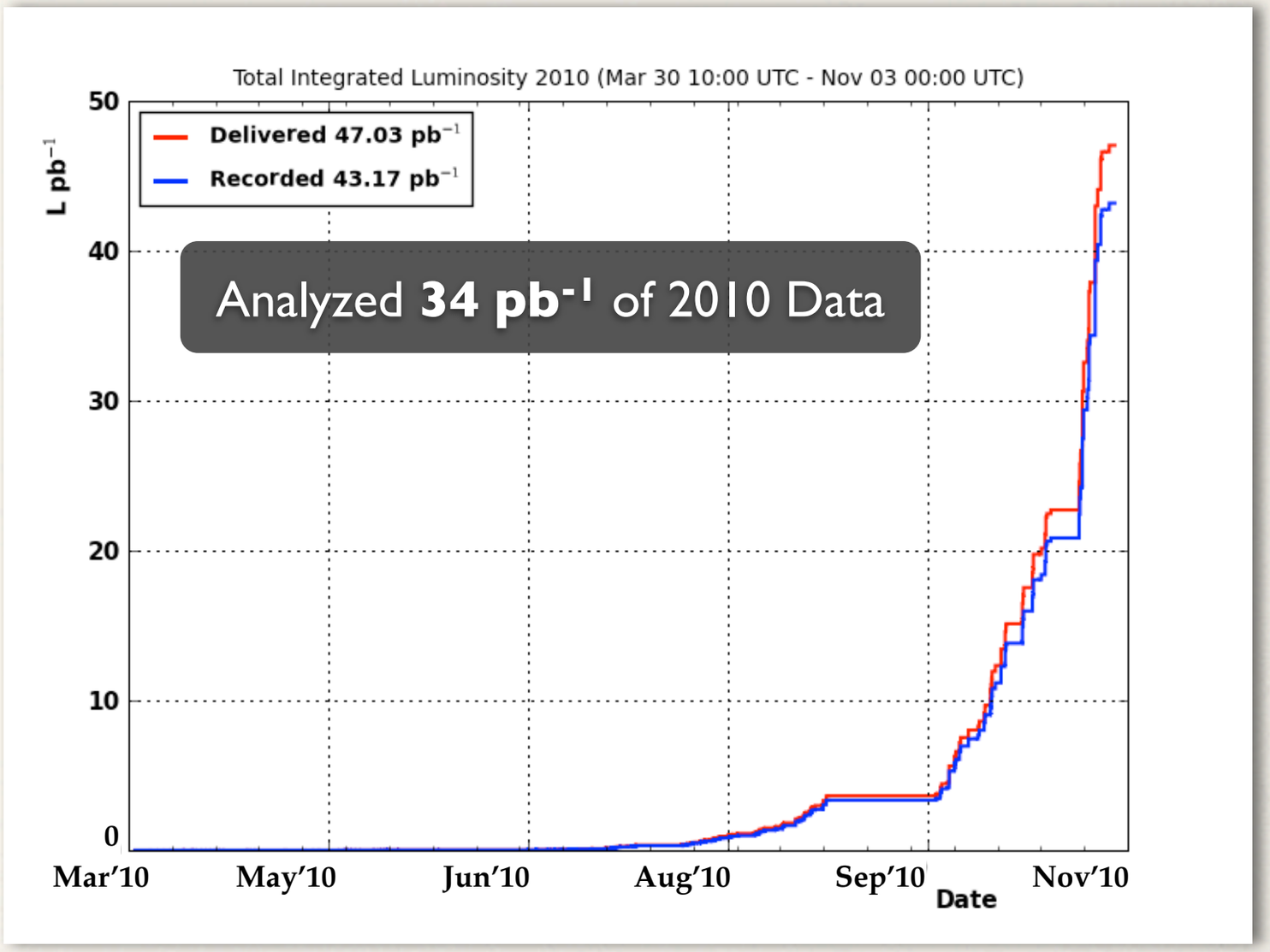
- hit resolutions
- readout system (present in Data)
- calibration effects

Measure the b-Tag Efficiency in data with "**System8**"

Input Data

b-Tagging

1. Select data with Muon Trigger
2. Use reconstructed Muons with $pT > 5\ GeV/c,\ |\eta| < 2.4$
3. Work with reconstructed Jets with $pT > 20\ GeV/c,\ |\eta| < 2.4$

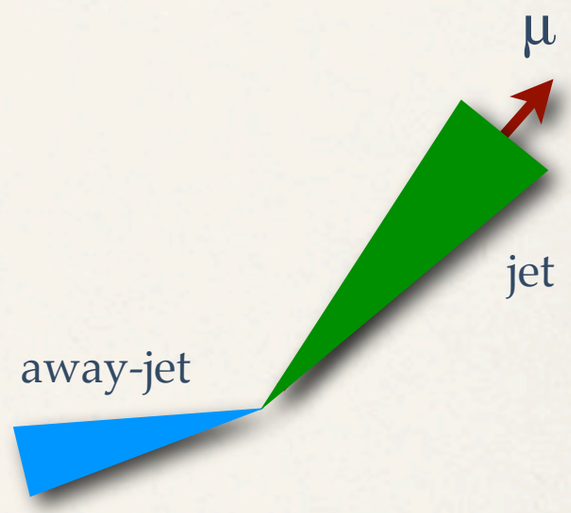


System8

b-Tagging efficiency

Topology

System8 measures the b-Tag efficiency in **b → μX** channel in a sample with μ in jet and away jet



Samples

(n) **μ** in **jet** + **away jet**

(p) **μ** in **jet** + **tagged away jet**

Taggers

probe b-Tag discriminant for the jet with μ inside

tag p_T^{rel} cut is applied to the "μ in jet" in order to enrich sample with b's

$$\vec{p}_T^{\,rel} = \frac{\vec{p}_\mu \times \vec{p}_{jet}}{p_{jet}}$$

S8 has minimal dependency on MC: only correlation coefficients are extracted from MC. Other System8 inputs are obtained from Data.

System8

b-Tagging efficiency

$$
\begin{aligned}
n &= n_b + n_{cl} && (1)\\
p &= p_b + p_{cl} && (2)\\
n^t &= \epsilon_b^t\ n_b + \epsilon_{cl}^t\ n_{cl} && (3)\\
p^t &= \beta\ \epsilon_b^t\ p_b + \alpha\ \epsilon_{cl}^t\ p_{cl} && (4)\\
n^\mu &= \epsilon_b^\mu\ n_b + \epsilon_{cl}^\mu\ n_{cl} && (5)\\
p^\mu &= \delta\ \epsilon_b^\mu\ p_b + \gamma\ \epsilon_{cl}^\mu\ p_{cl} && (6)\\
n^{t,\mu} &= \kappa_b\ \epsilon_b^t\ \epsilon_b^\mu\ n_b + \kappa_{cl}\ \epsilon_{cl}^t\ \epsilon_{cl}^\mu\ n_{cl} && (7)\\
p^{t,\mu} &= \kappa_b\ \beta\ \delta\ \epsilon_b^t\ \epsilon_b^\mu\ p_b + \kappa_{cl}\ \alpha\ \gamma\ \epsilon_{cl}^t\ \epsilon_{cl}^\mu\ p_{cl} && (8)
\end{aligned}
$$

System8

b-Tagging efficiency

$$n = n_b + n_{cl} \quad (1)$$

$$p = p_b + p_{cl} \quad (2)$$

$$n^t = \epsilon_b^t\, n_b + \epsilon_{cl}^t\, n_{cl} \quad (3)$$

$$p^t = \beta\, \epsilon_b^t\, p_b + \alpha\, \epsilon_{cl}^t\, p_{cl} \quad (4)$$

$$n^\mu = \epsilon_b^\mu\, n_b + \epsilon_{cl}^\mu\, n_{cl} \quad (5)$$

$$p^\mu = \delta\, \epsilon_b^\mu\, p_b + \gamma\, \epsilon_{cl}^\mu\, p_{cl} \quad (6)$$

$$n^{t,\mu} = \kappa_b\, \epsilon_b^t\, \epsilon_b^\mu\, n_b + \kappa_{cl}\, \epsilon_{cl}^t\, \epsilon_{cl}^\mu\, n_{cl} \quad (7)$$

$$p^{t,\mu} = \kappa_b\, \beta\, \delta\, \epsilon_b^t\, \epsilon_b^\mu\, p_b + \kappa_{cl}\, \alpha\, \gamma\, \epsilon_{cl}^t\, \epsilon_{cl}^\mu\, p_{cl} \quad (8)$$

8 inputs

(from Data)

System8

b-Tagging efficiency

$$n = n_b + n_{cl} \quad (1)$$

$$p = p_b + p_{cl} \quad (2)$$

$$n^t = \epsilon_b^t \, n_b + \epsilon_{cl}^t \, n_{cl} \quad (3)$$

$$p^t = \beta \, \epsilon_b^t \, p_b + \alpha \, \epsilon_{cl}^t \, p_{cl} \quad (4)$$

$$n^\mu = \epsilon_b^\mu \, n_b + \epsilon_{cl}^\mu \, n_{cl} \quad (5)$$

$$p^\mu = \delta \, \epsilon_b^\mu \, p_b + \gamma \, \epsilon_{cl}^\mu \, p_{cl} \quad (6)$$

$$n^{t,\mu} = \kappa_b \, \epsilon_b^t \, \epsilon_b^\mu \, n_b + \kappa_{cl} \, \epsilon_{cl}^t \, \epsilon_{cl}^\mu \, n_{cl} \quad (7)$$

$$p^{t,\mu} = \kappa_b \, \beta \, \delta \, \epsilon_b^t \, \epsilon_b^\mu \, p_b + \kappa_{cl} \, \alpha \, \gamma \, \epsilon_{cl}^t \, \epsilon_{cl}^\mu \, p_{cl} \quad (8)$$

6 coefficients

(from Monte-Carlo)

System8

b-Tagging efficiency

$$n = n_b + n_{cl} \quad (1)$$

$$p = p_b + p_{cl} \quad (2)$$

$$n^t = \epsilon_b^t\, n_b + \epsilon_{cl}^t\, n_{cl} \quad (3)$$

$$p^t = \beta\, \epsilon_b^t\, p_b + \ldots \quad (4)$$

$$n^\mu = \epsilon_b^\mu\, n_b + \epsilon_{cl}^\mu\, n_{cl} \quad (5)$$

$$p^\mu = \delta\, \epsilon_b^\mu\, p_b + \gamma\, \epsilon_{cl}^\mu\, p_{cl} \quad (6)$$

$$n^{t,\mu} = \kappa_b\, \epsilon_b^t\, \epsilon_b^\mu\, n_b + \kappa_{cl}\, \epsilon_{cl}^t\, \epsilon_{cl}^\mu\, n_{cl} \quad (7)$$

$$p^{t,\mu} = \kappa_b\, \beta\, \delta\, \epsilon_b^t\, \epsilon_b^\mu\, p_b + \kappa_{cl}\, \alpha\, \gamma\, \epsilon_{cl}^t\, \epsilon_{cl}^\mu\, p_{cl} \quad (8)$$

$$\beta = \frac{\epsilon_b^t\ from\ (p)\ sample}{\epsilon_b^t\ from\ (n)\ sample}$$

(n) μ in jet + away jet
(p) μ in jet + tagged away jet

coefficients represent correlation between (n) and (p) samples for different jet flavors and cuts applied.

System8

b-Tagging efficiency

$$n = n_b + n_{cl} \quad (1)$$

$$p = p_b + p_{cl} \quad (2)$$

$$n^t = \epsilon_b^t \, n_b + \epsilon_{cl}^t \, n_{cl} \quad (3)$$

$$p^t = \beta \, \epsilon_b^t \, p_b + \alpha \, \epsilon_{cl}^t \, p_{cl} \quad (4)$$

$$n^\mu = \epsilon_b^\mu \, n_b + \epsilon_{cl}^\mu \, n_{cl} \quad (5)$$

$$p^\mu = \delta \, \epsilon_b^\mu \, p_b + \gamma \, \epsilon_{cl}^\mu \, p_{cl} \quad (6)$$

$$n^{t,\mu} = \kappa_b \, \epsilon_b^t \, \epsilon_b^\mu \, n_b + \kappa_{cl} \, \epsilon_{cl}^t \, \epsilon_{cl}^\mu \, n_{cl} \quad (7)$$

$$p^{t,\mu} = \kappa_b \, \beta \, \delta \, \epsilon_b^t \, \epsilon_b^\mu \, p_b + \kappa_{cl} \, \alpha \, \gamma \, \epsilon_{cl}^t \, \epsilon_{cl}^\mu \, p_{cl} \quad (8)$$

8 outputs

(System8 results)

System8

b-Tagging efficiency

$$n = n_b + n_{cl} \quad (1)$$

$$p = p_b + p_{cl} \quad (2)$$

$$n^t = \epsilon_b^t\, n_b + \epsilon_{cl}^t\, n_{cl} \quad (3)$$

$$p^t = \beta\, \epsilon_b^t\, p_b + \alpha\, \epsilon_{cl}^t\, p_{cl} \quad (4)$$

$$n^\mu = \epsilon_b^\mu\, n_b + \epsilon_{cl}^\mu\, n_{cl} \quad (5)$$

$$p^\mu = \delta\, \epsilon_b^\mu\, p_b + \gamma\, \epsilon_{cl}^\mu\, p_{cl} \quad (6)$$

$$n^{t,\mu} = \kappa_b\, \epsilon_b^t\, \epsilon_b^\mu\, n_b + \kappa_{cl}\, \epsilon_{cl}^t\, \epsilon_{cl}^\mu\, n_{cl} \quad (7)$$

$$p^{t,\mu} = \kappa_b\, \beta\, \delta\, \epsilon_b^t\, \epsilon_b^\mu\, p_b + \kappa_{cl}\, \alpha\, \gamma\, \epsilon_{cl}^t\, \epsilon_{cl}^\mu\, p_{cl} \quad (8)$$

b-Tag efficiency

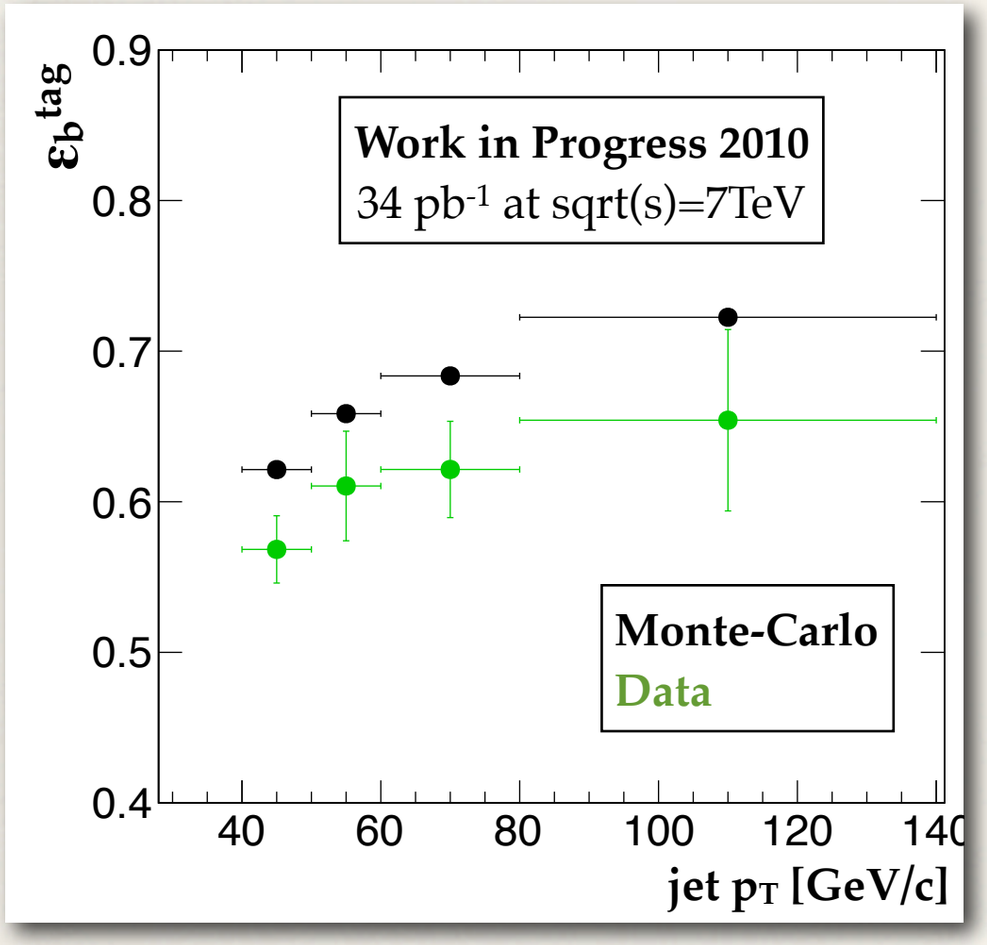
($b \to \mu$ channel)

This is **NOT** inclusive b-Tag efficiency: correct with scale factor from MC

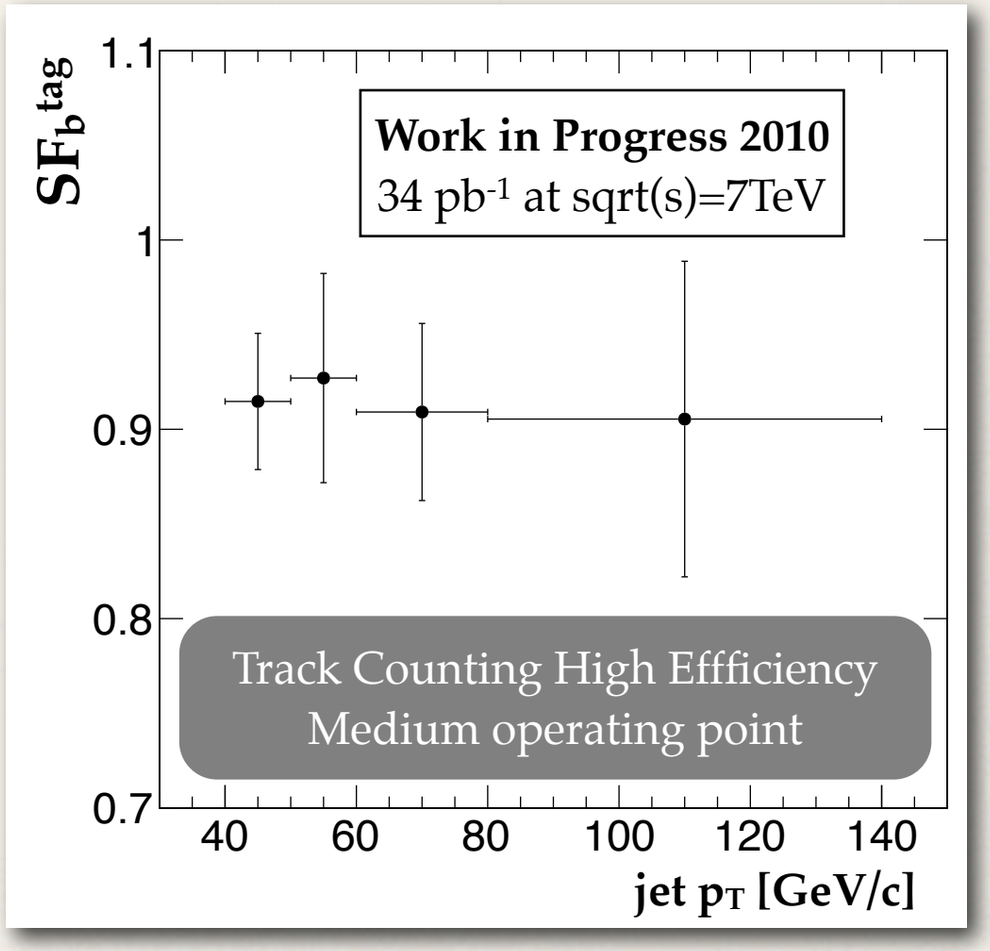
System8 Efficiency

b-Tagging efficiency

b-Tagging Efficiency



b-Tagging Efficiency Scale Factor (Data/MC)



Only **statistical** error is shown

System8 Systematic Uncertainty

b-Tagging efficiency

Source	Default Value	Procedure
away jet tagger	TCHPL	*vary away jet tagger*
μ p_T cut	5 GeV/c	*increase cut to 7, 10 GeV/c*
p_T^{rel} cut	0.8 GeV/c	*change cut to ± 10%*
coefficients	inclusive QCD	*extract correlation coefficients from different MC sample*
gluon splitting		*enhance gluon splitting*
closure		*estimate difference in the MC true b-Tag efficiency and S8 running on MC*
		total ~15%

Conclusion

b-Tag methods

1. Track Counting b-Tag algorithm uses track impact parameter significance as discriminant
2. System8 weakly depends on Monte-Carlo
3. 2010 data was processed
4. Scale Factors (Data/MC) are ~90% in a wide range of jet p_T
5. Systematic error is ~15%