B-Tagging Algorithms at the CMS Experiment

Gavril Giurgiu (for the CMS Collaboration)
Johns Hopkins University

DPF-APS Meeting, August 10, 2011

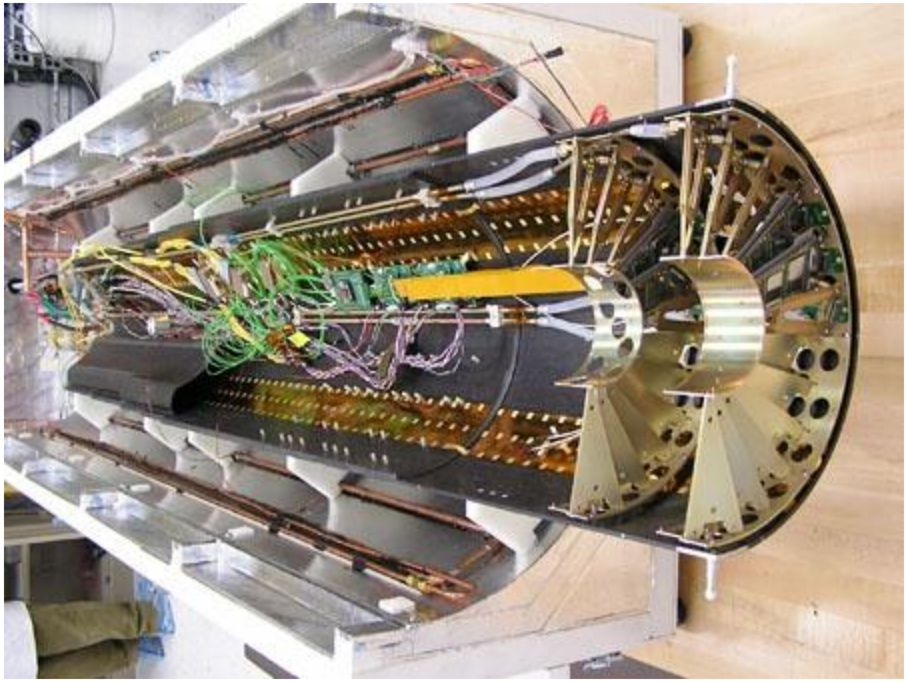
Brown University, Providence, Rhode Island

Introduction

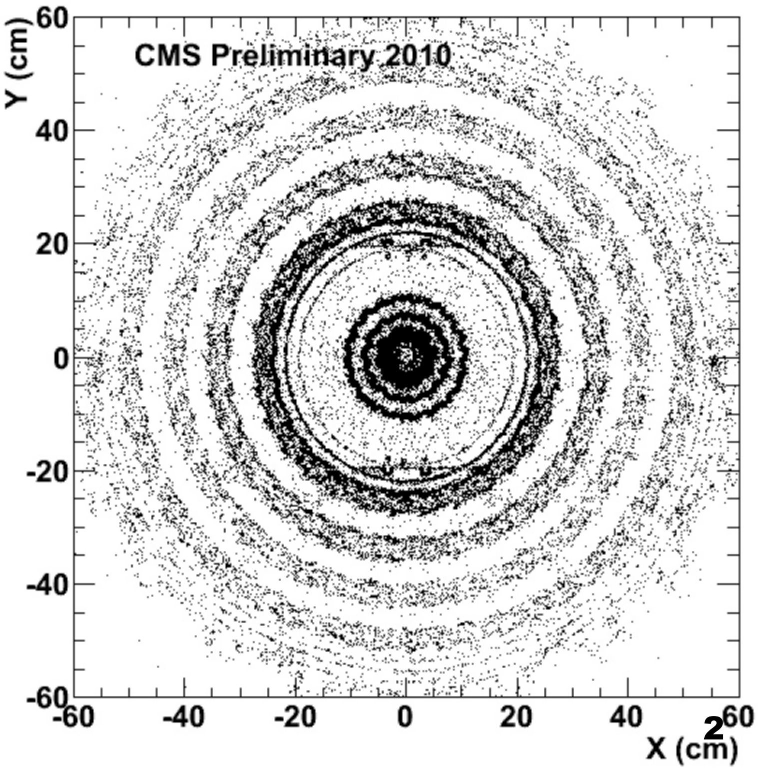
- B-jet identification (b-tagging) crucial for wide range of SM measurements, Higgs and New Physics searches for at LHC
- B-tagging takes advantage of long lifetime, large mass and large semi-leptonic decay fraction of B-hadrons

- Most important detector component for b-tagging is the silicon tracker (pixel + strips)

half pixel detector



tracker radiography from $\gamma \rightarrow e^+ e^-$



Inputs to B-Tagging

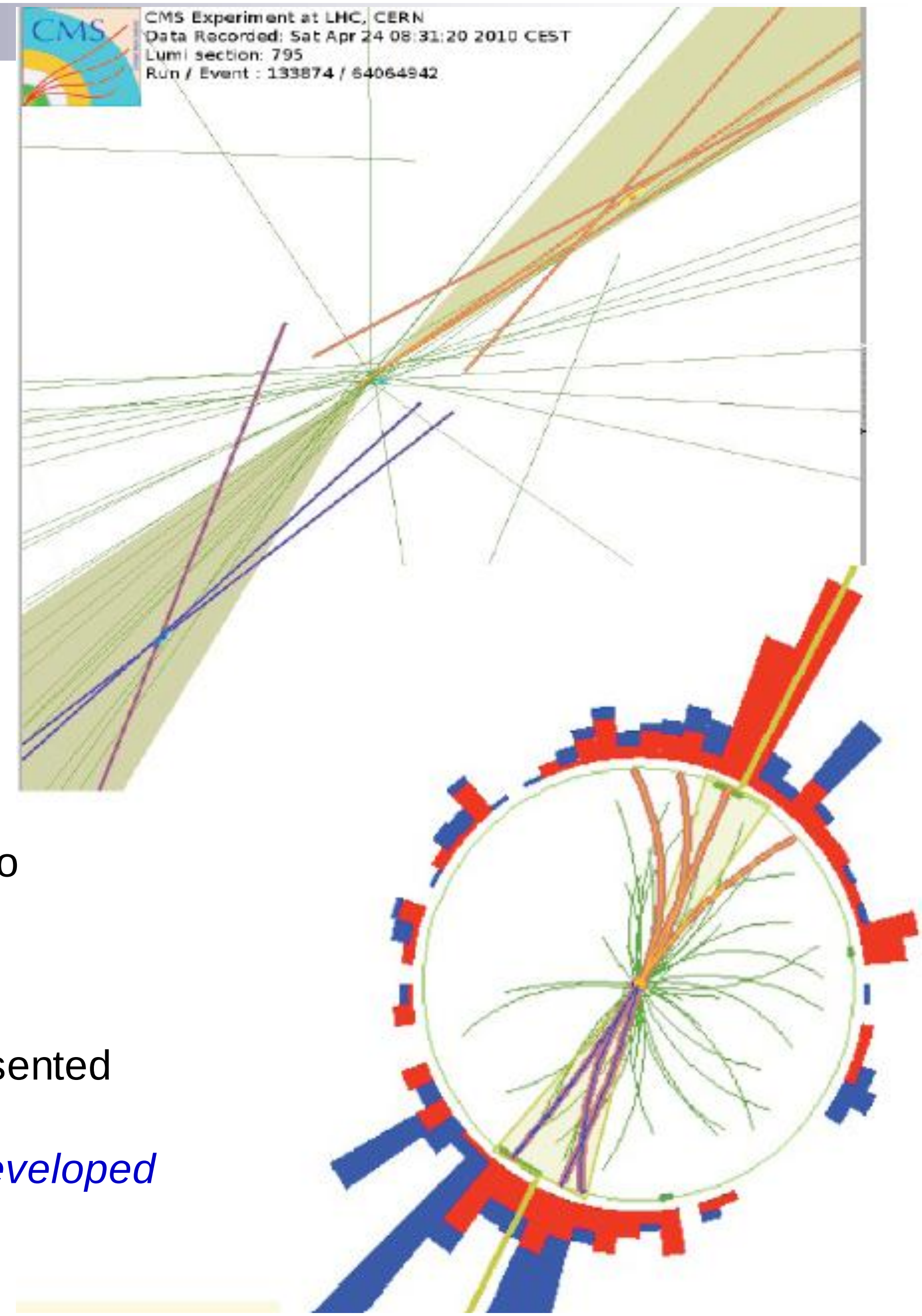
- Jets - iterative Cone with $\Delta R = 0.5$ clustering algorithm (IC05) *CMS PAS JME_07_003, 2007)*

- High quality tracks reconstructed from pixel/strip silicon hits using Kalman Filter

- Primary and secondary vertices
  CMS NOTE 2006_029, 2006,
  CMS Note 2008_033, 2008,
  CMS PAS TRK 10-005

- Muons seeded from muon chambers and linked to tracker tracks (*JINST 3, 2008, S08004*)

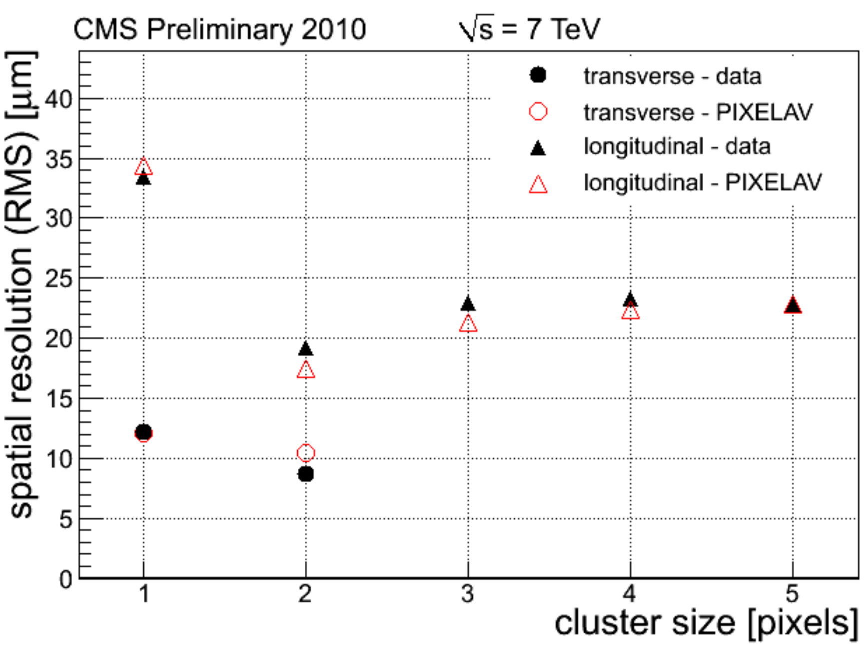
- Performance of b-tagging algorithms on data presented by Saptaparna Bhattacharya, in the next talk:
Efficiency measurement of b-tagging algorithms developed by the CMS experiment

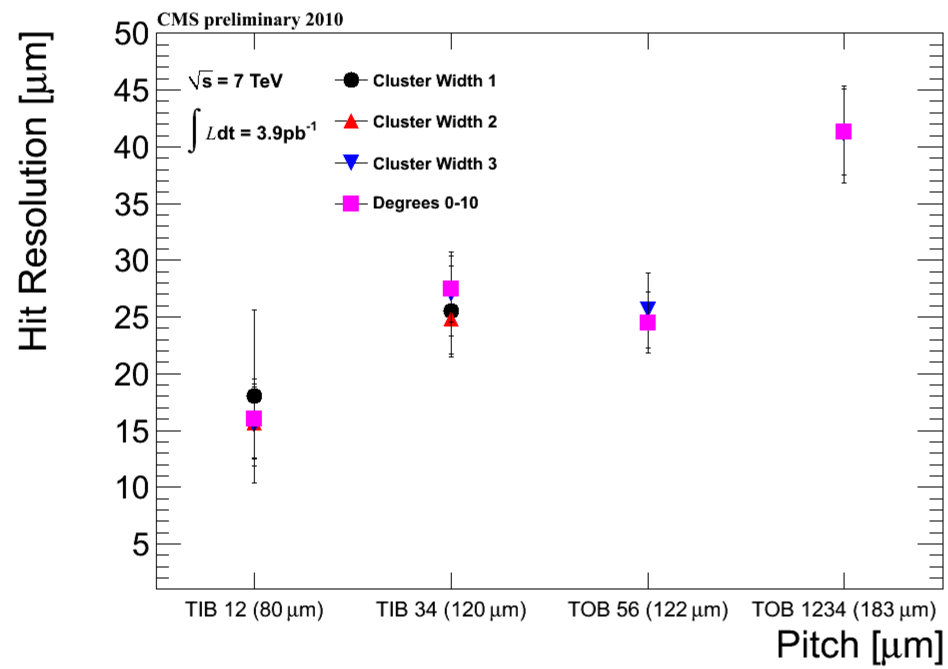


Tracking Hit Resolution

- Accurate reconstruction of tracks and vertices is crucial for b-tagging

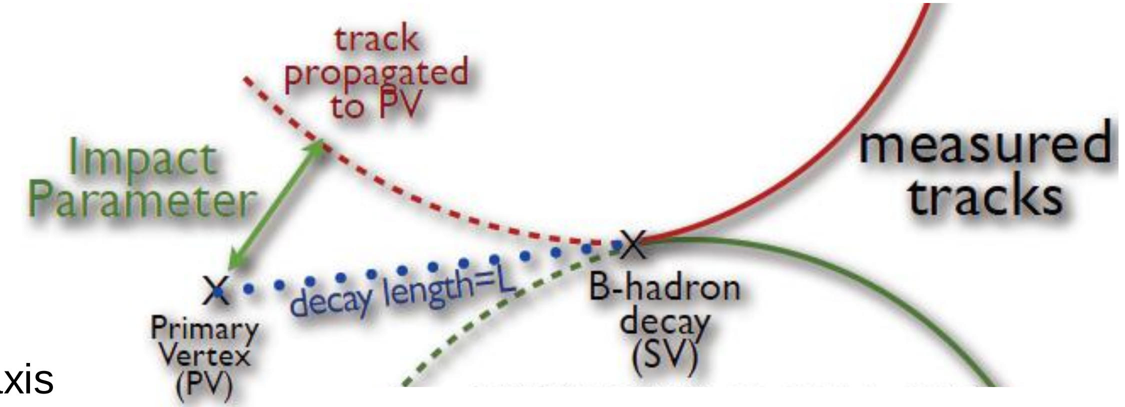
- Resolution of tracking hits:
    - Pixel tracker: ~10 μm along 100 μm pitch direction (transverse)
      ~20-35 μm along 150 μm pitch direction (longitudinal)
      - excellent agreement between data and simulation
    - Strip tracker: 15-40 μm depending on strip pitch

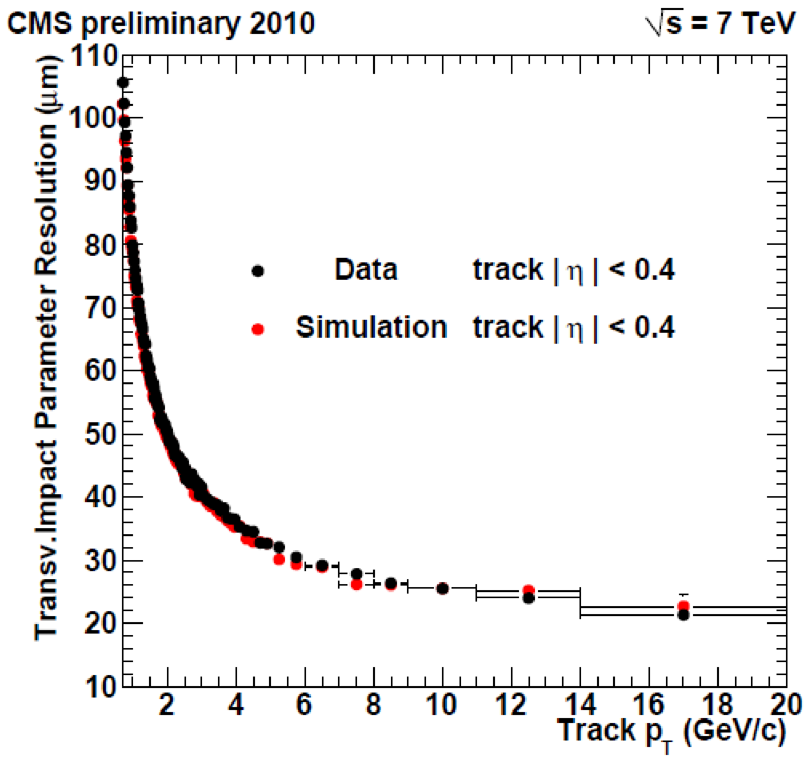


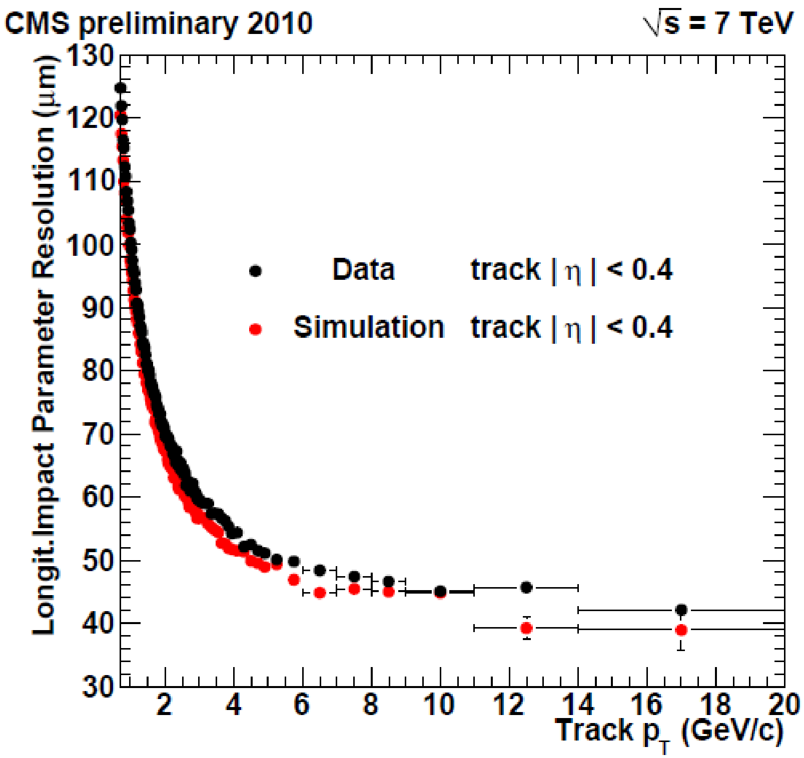


Track Impact Parameter (CMS PAS TRK 10-005)

- Track impact parameter (IP) :
  - most powerful discriminator between b and u,d,s,g flavors
  - signed by the scalar product between the IP segment and jet axis





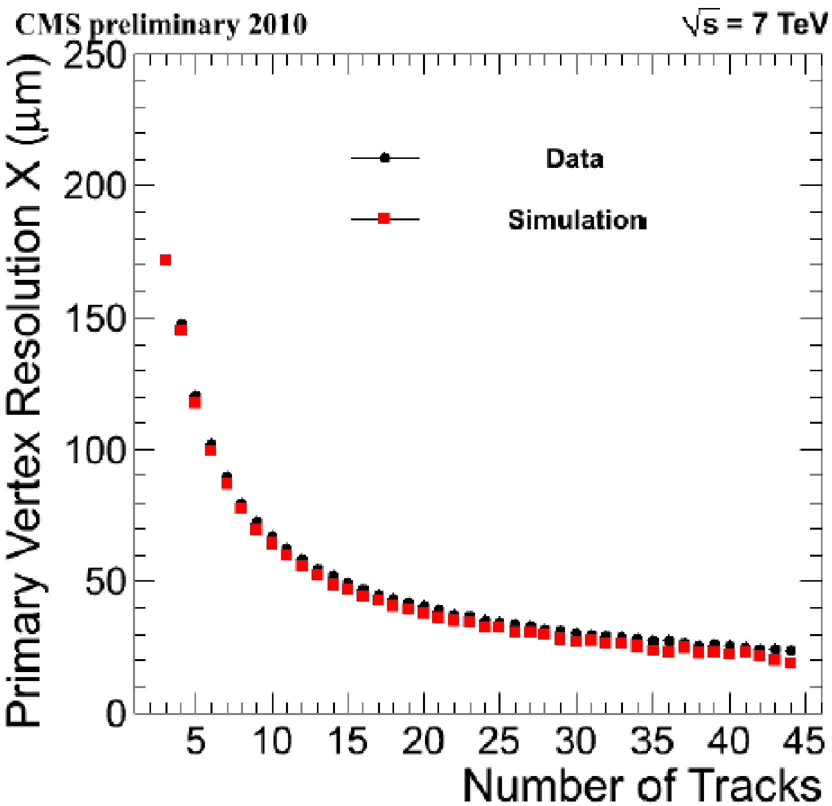


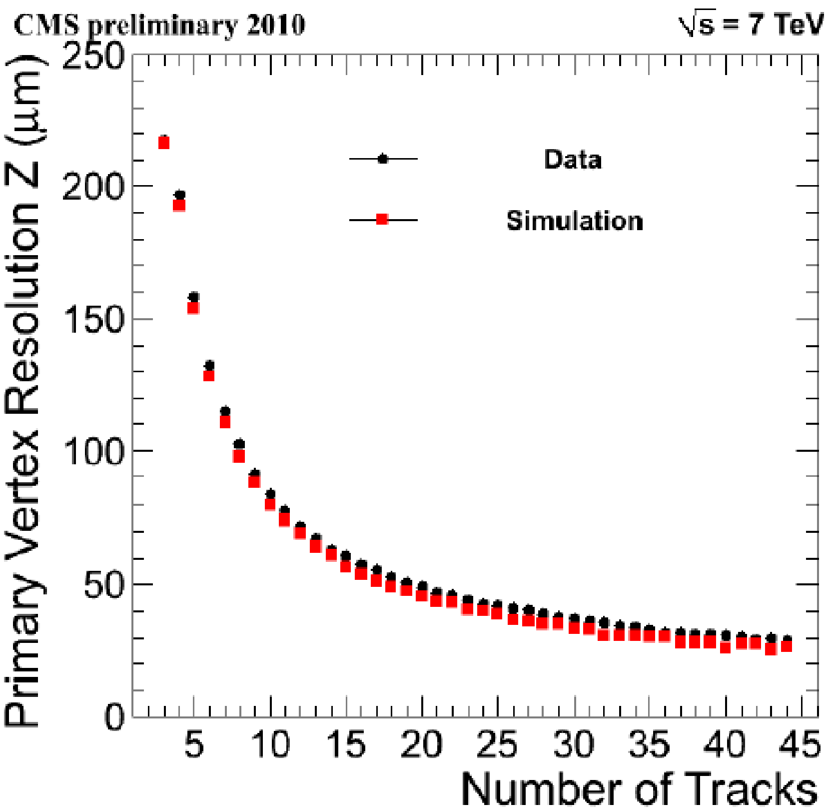
Vertex Resolution (CMS PAS TRK 10-005)

- Improves with number of tracks, similar in transverse and longitudinal directions

- Vertex resolution of ~25 microns for more than 30 tracks

- Very well described by simulation





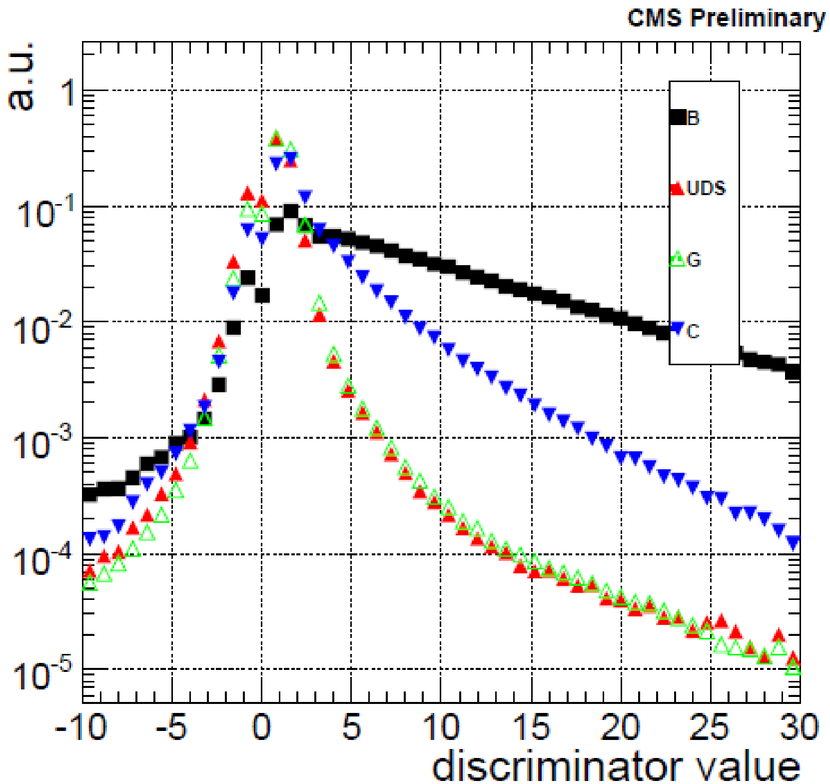
B-Tagging Algorithms

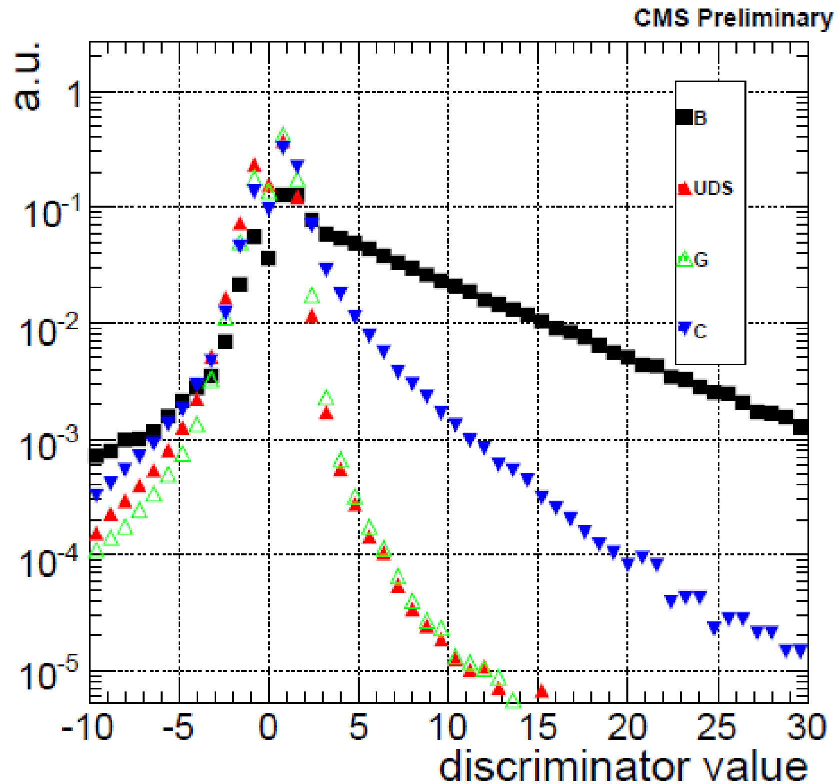
Each b-tagger produces one number, the discriminator between b and u,d,s,g flavors

Track Counting

- Simplest discriminator is the *IP significance = IP / σ_{IP} of the N^{th} most displaced track*

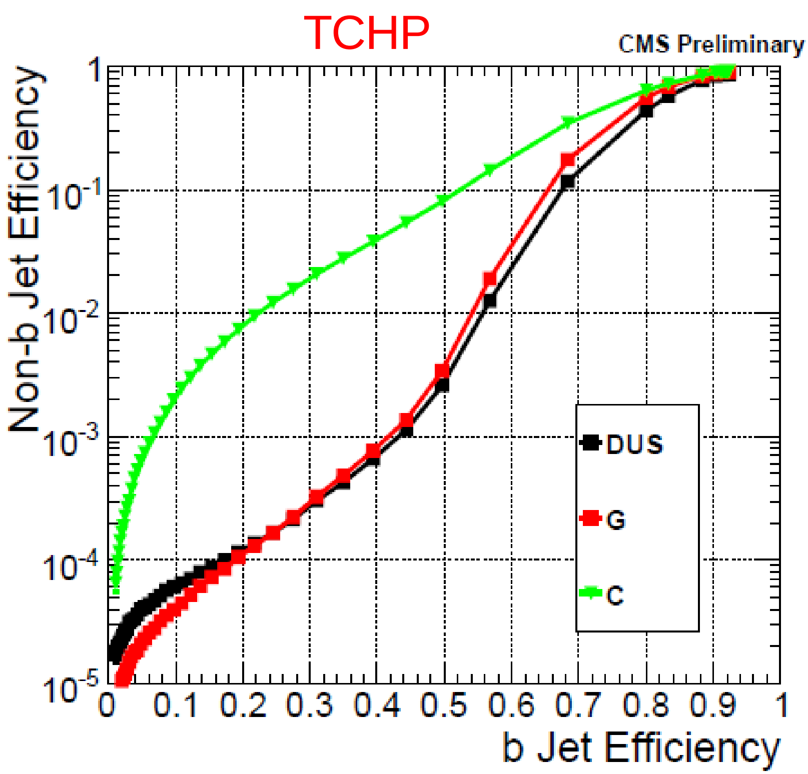
- Use either N = 2 (high efficiency, TCHE) or N = 3 (high purity, TCHP)

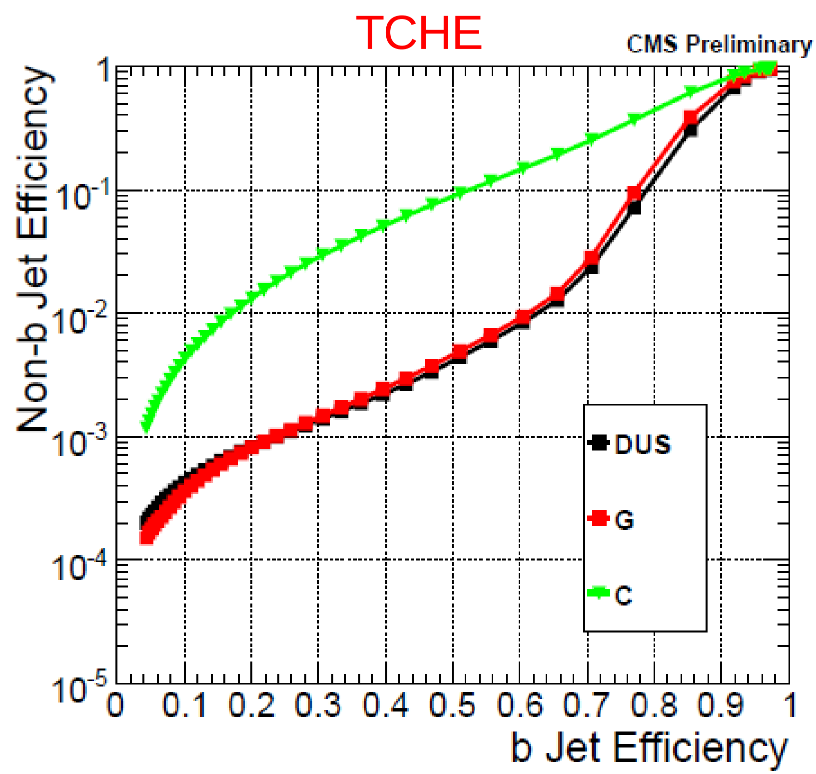




Track Counting Performance

- Use Pythia QCD sample with P_T hat > 80 GeV
- Select jets with PT > 20 GeV and |η| < 2.4
- Efficiency of track counting with high purity (TCHP) less than 100% due to requirement of at least 3 tracks inside jet





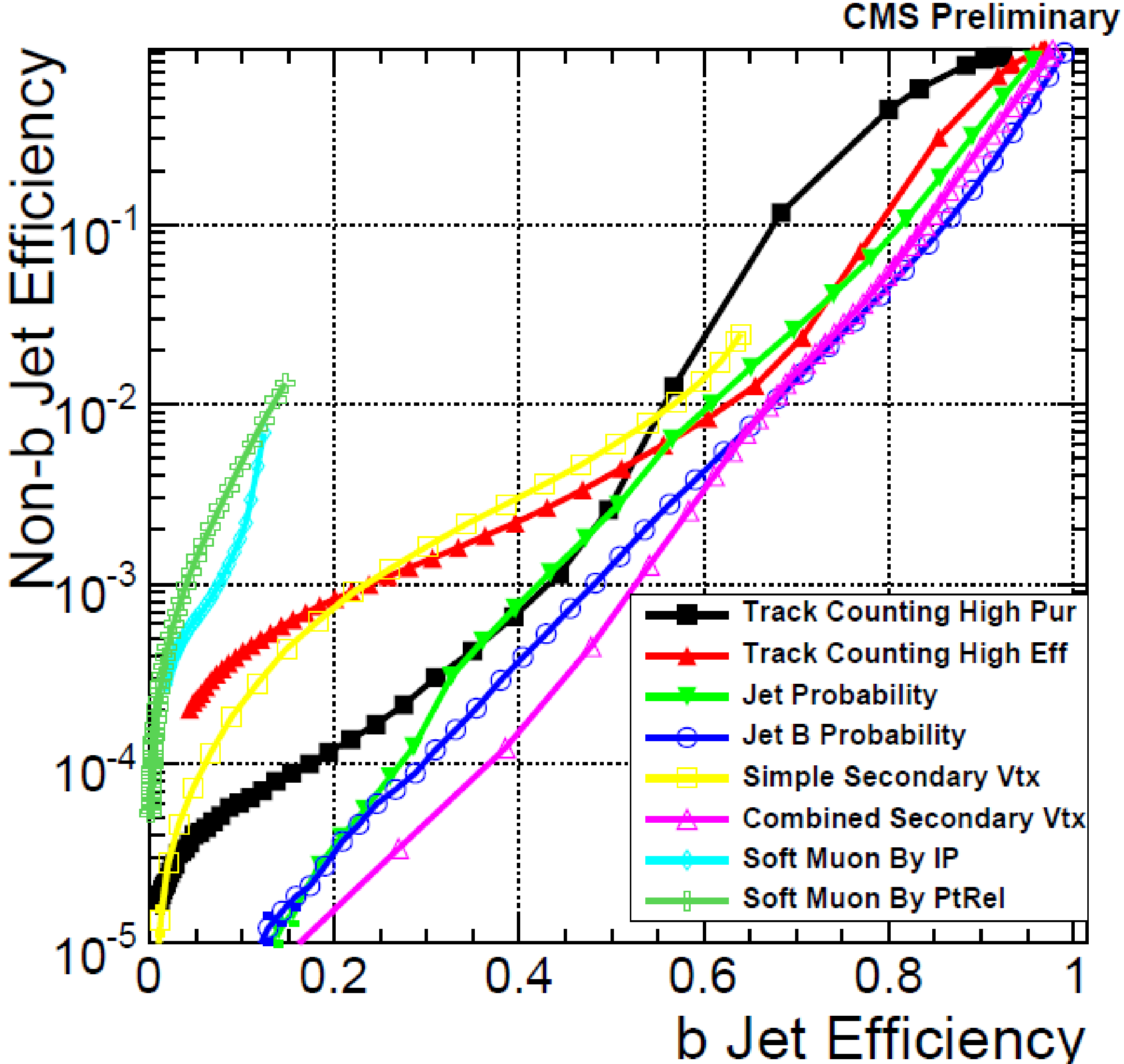
Jet Probability, Muon and Vertex Taggers

- Jet probability tagger:
  - combines info from all selected tracks into jet probability

$$P_{jet} = \Pi \cdot \sum_{j=0}^{N-1} \frac{(-\ln \Pi)^j}{j!} \text{ where } \Pi = \prod_{i=1}^{N} P_{tr}(i)$$

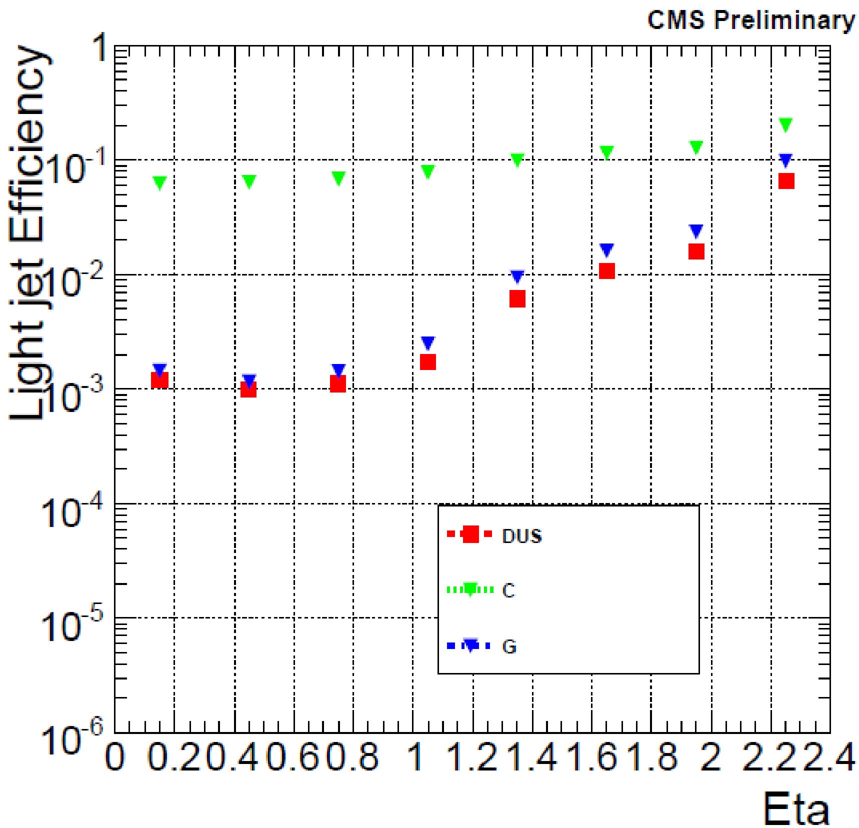
  - or, given average track multiplicity in B decays of ~ 5 and jet tracking efficiency about 80%, overweight the most 4 displaced tracks to form *jet B-probability*
- Lepton taggers :
  - leptons inside jets signal potential semi-leptonic B decays
  - to discriminate b → μ decays from c → μ or sequential b → c → μ decays, use discriminating quantities like:
    - muon impact parameter (if positive) or
    - muon transverse momentum w.r.t. jet axis (PTrel)
- Secondary vertex tagger
  - based on reconstruction of at least one secondary vertex
  - efficiency of 60-70%
  - significance of flight distance from PV to SV used as discriminator
- Combined taggers use most of the available info (flight distance, vertex mass, number of vertex tracks, track IP, …) to form a likelihood discriminator

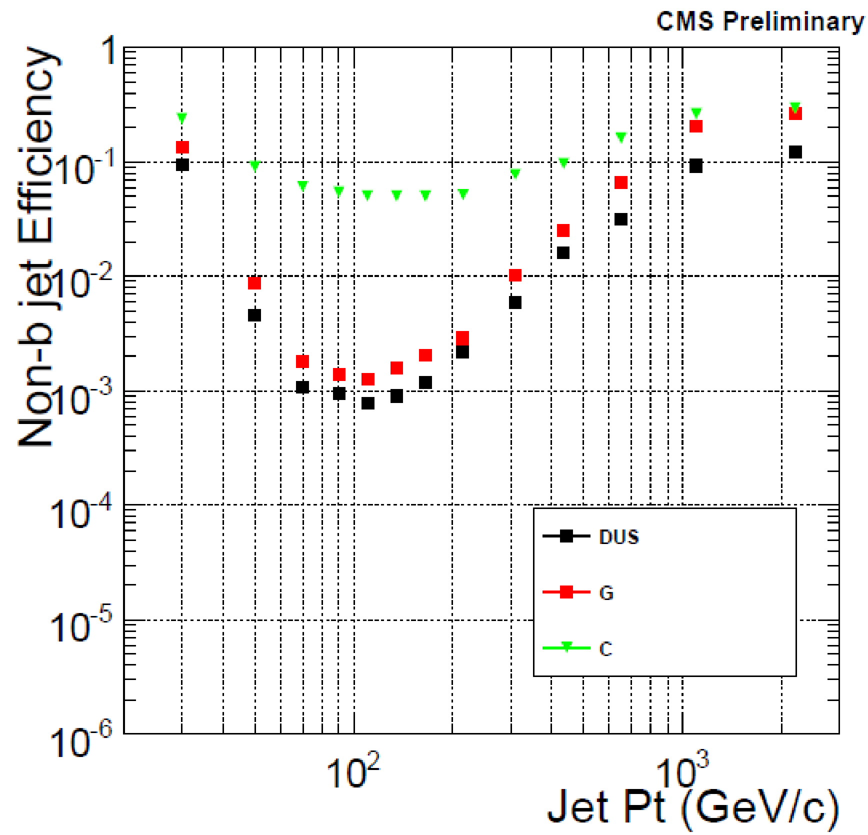
Comparison of B-Tagging Algorithms (simulation)



B-Tagging Fake Rates vs P_T and η (simulation)

- Use TCHP with efficiency at 50% as example

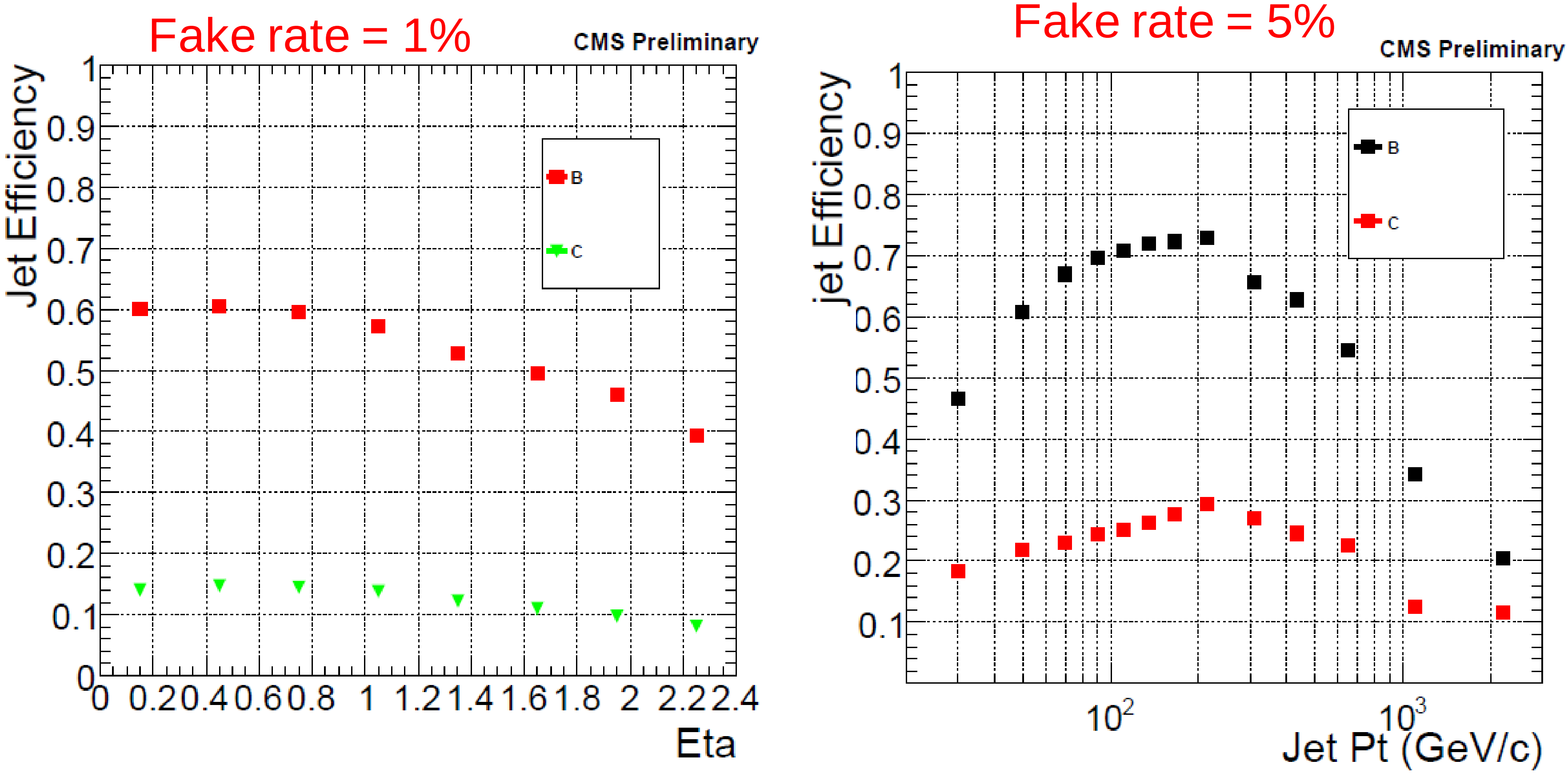




- Note orders of magnitude increase fake rates going from ~100 GeV to ~1 TeV jets

TCHP Efficiency as Function of P_T and η (simulation)

- Use TCHP with b-tagging as example

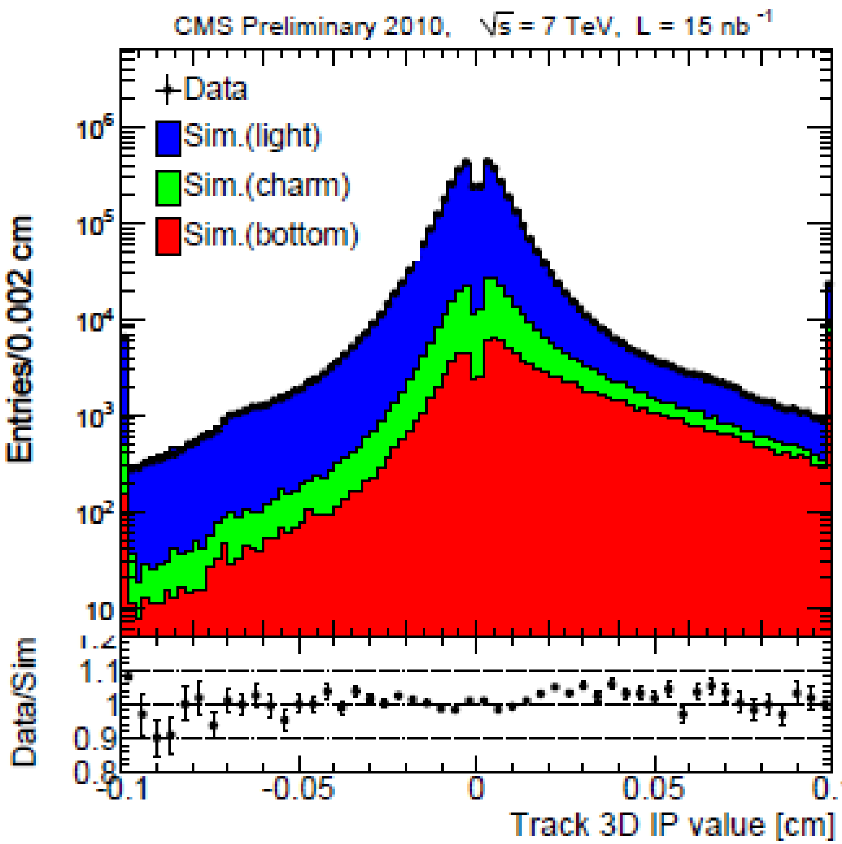


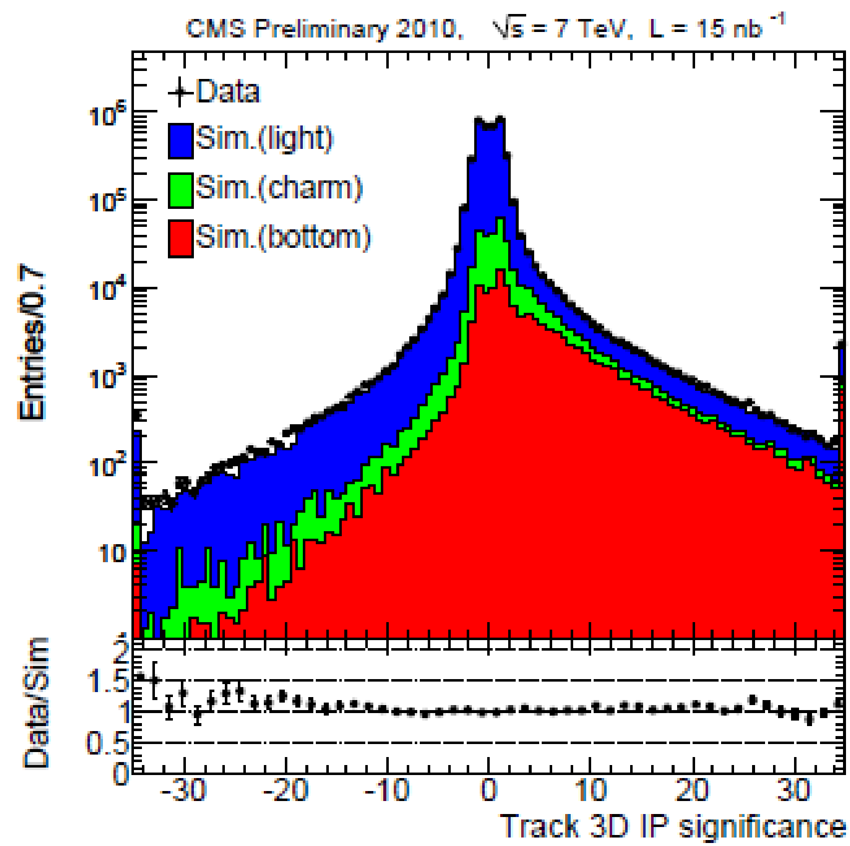
- Note degrading b-tagging efficiency going from ~100 GeV to ~1 TeV jets
- Degradation of b-tagging performance at high P_T due to tracking difficulties in dense jet environments (merged pixel/strip hits, large combinatorics, secondary interaction)

Validation of Track Impact Parameter with 7 TeV Data

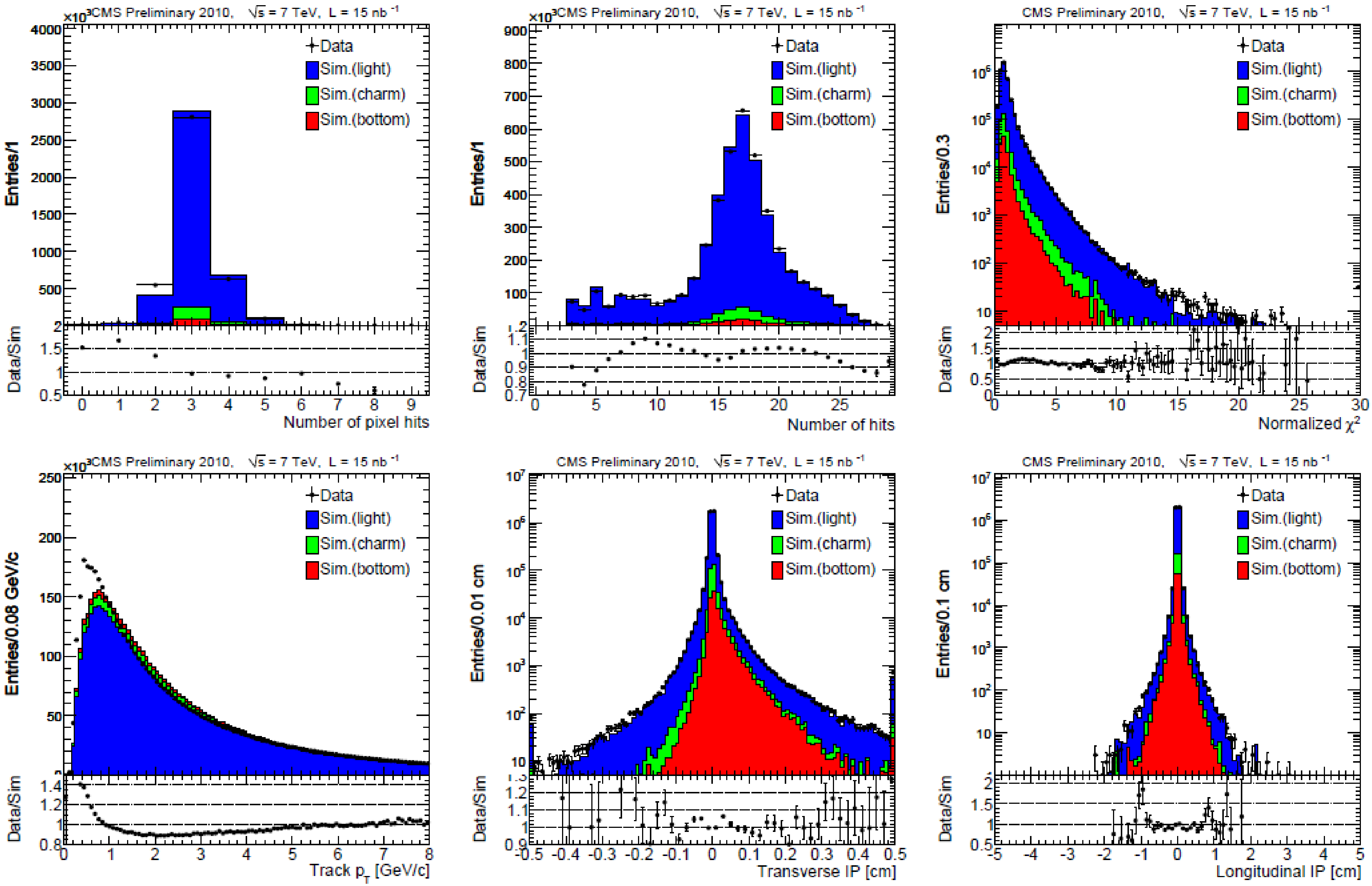
- Comparison between 8-15/nb of data and Pythia simulation performed for b-tagging inputs

- Very good agreement between data and simulation

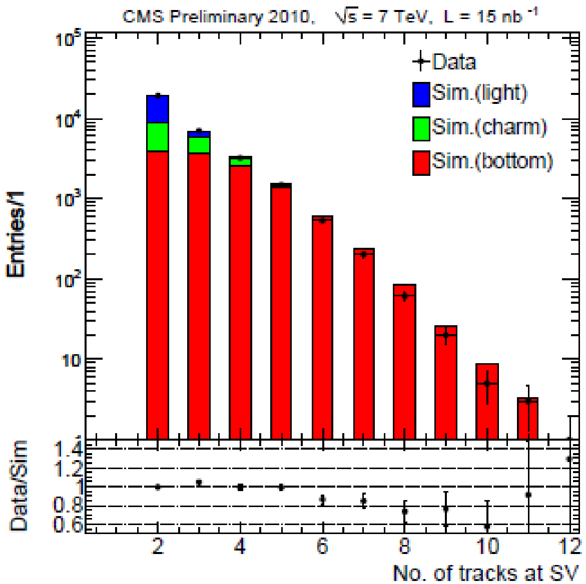


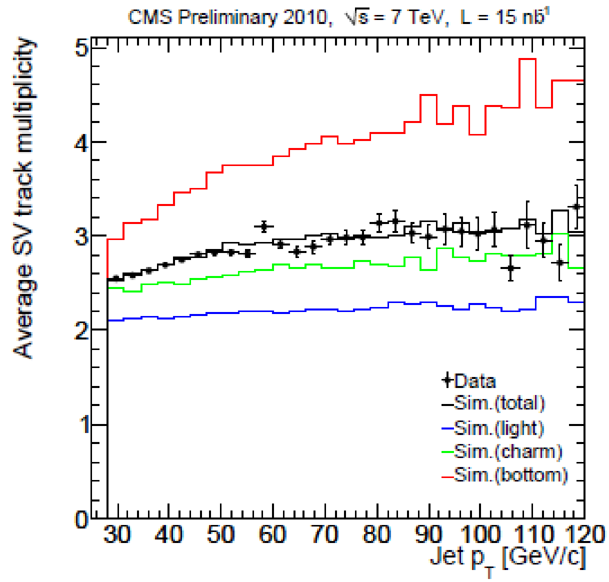


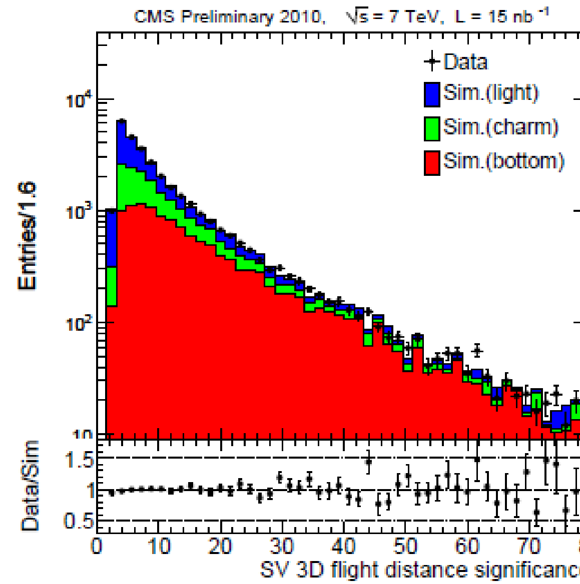
Validation of Other Tracks Quantities with 7 TeV Data

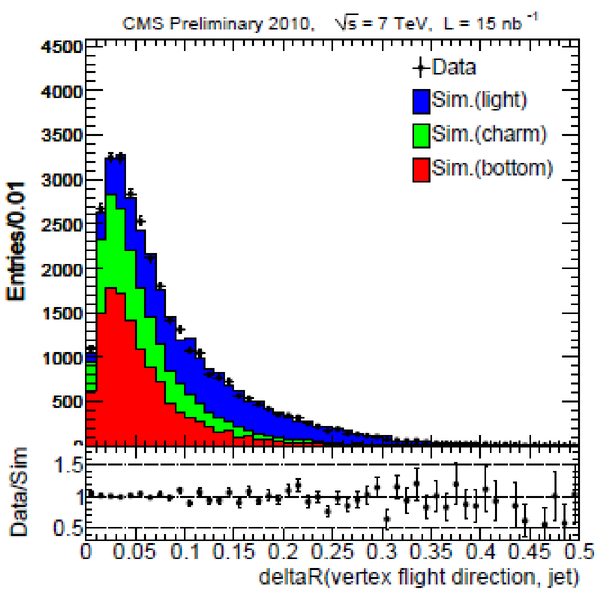


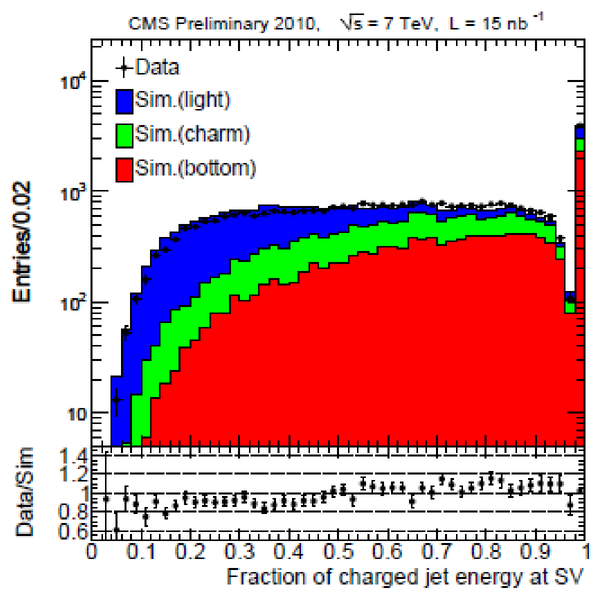
Validation of Secondary Vertices with 7 TeV Data

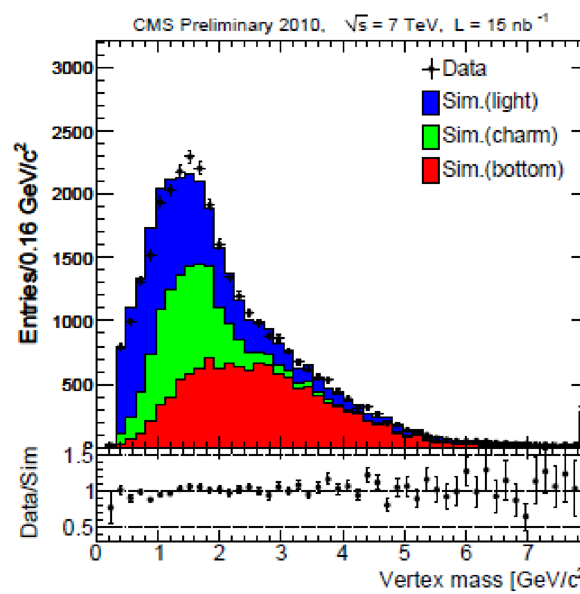








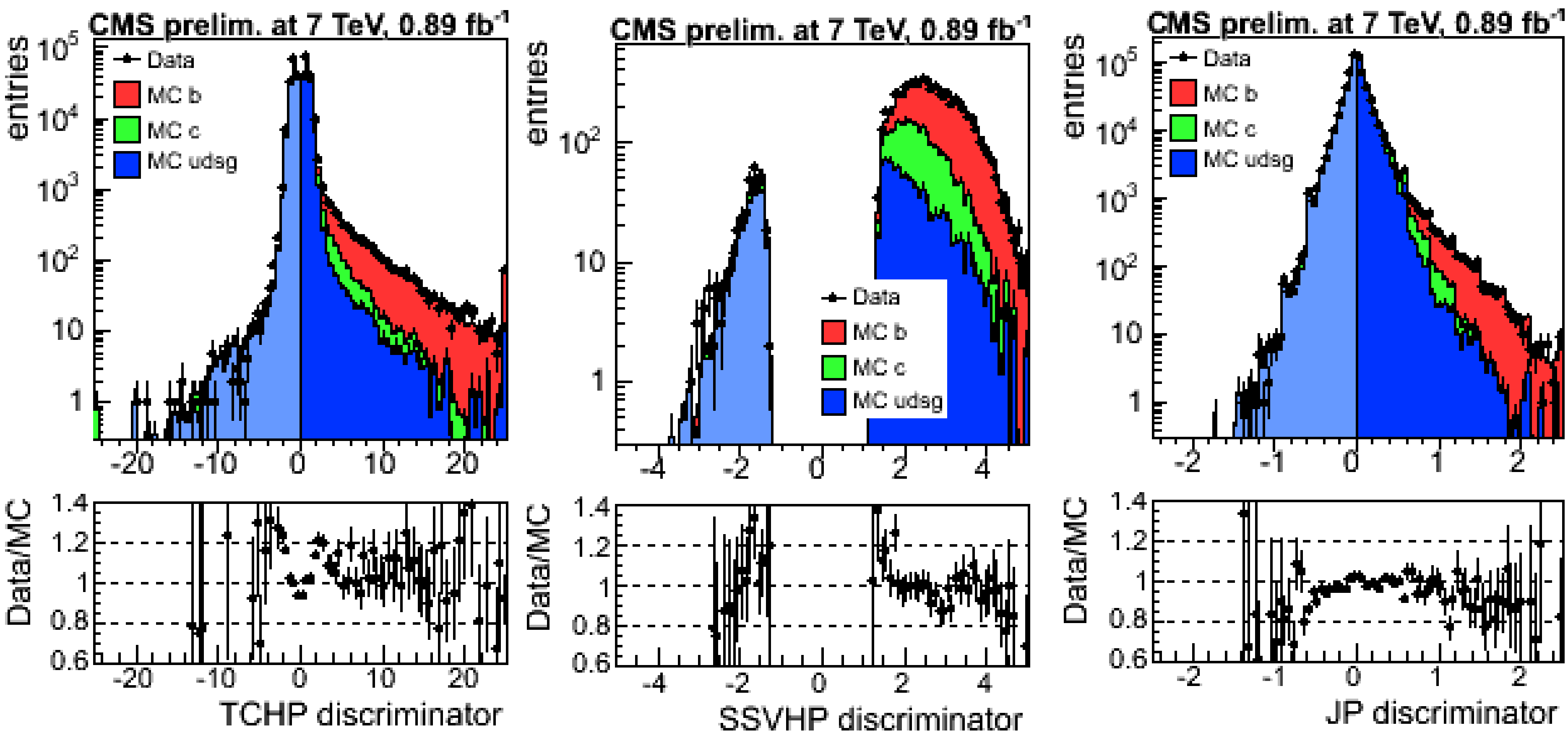




Validation of Displacement Based Discriminators

- Simulation reproduces most track and vertex quantities within 5-10%

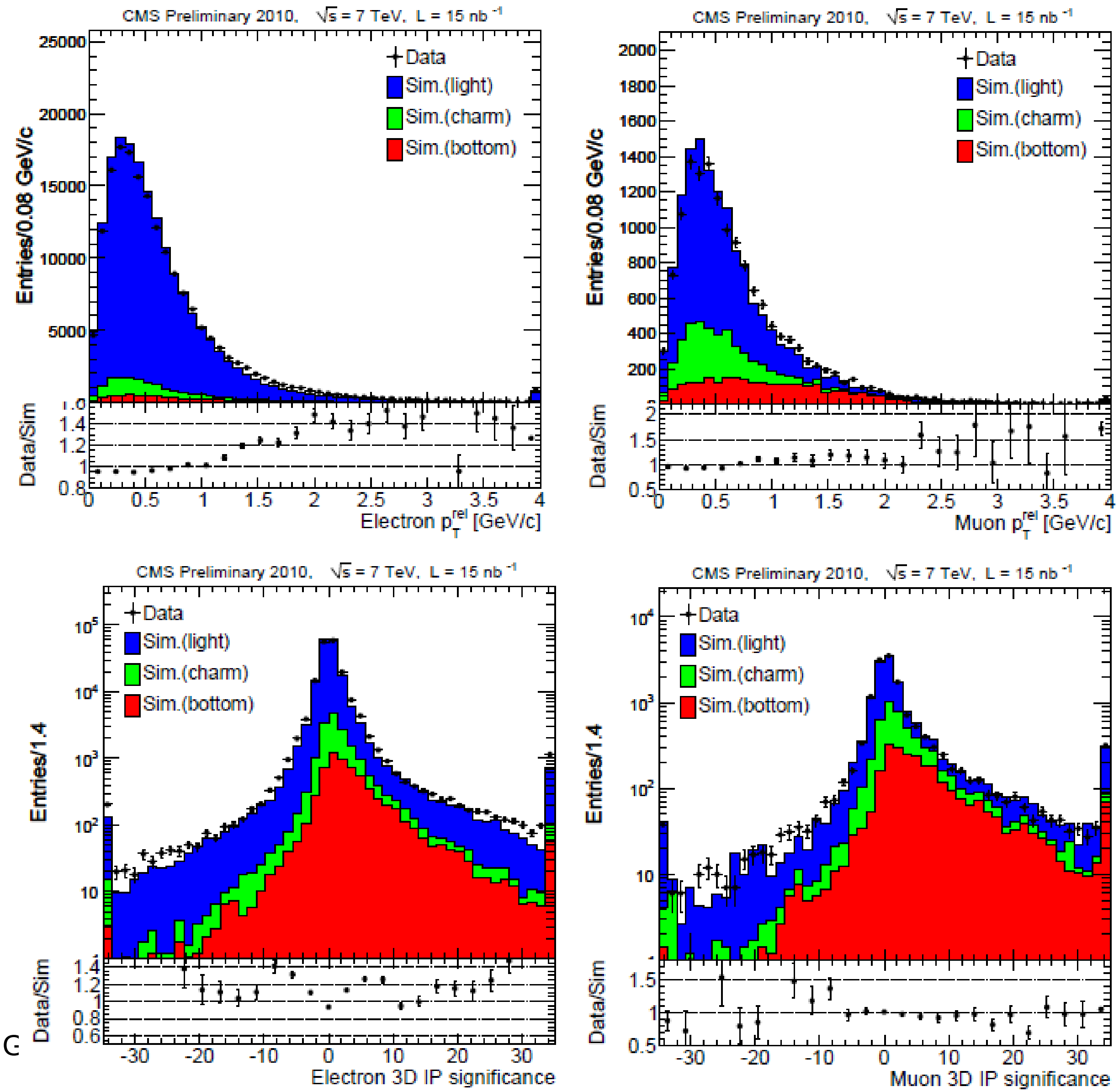
- B-tagging algorithms can be safely used in physics analyses



Validation of Leptons with 7 TeV Data

- Lepton transverse momentum w.r.t. jet axis

- Lepton impact parameter w.r.t. primary vertex

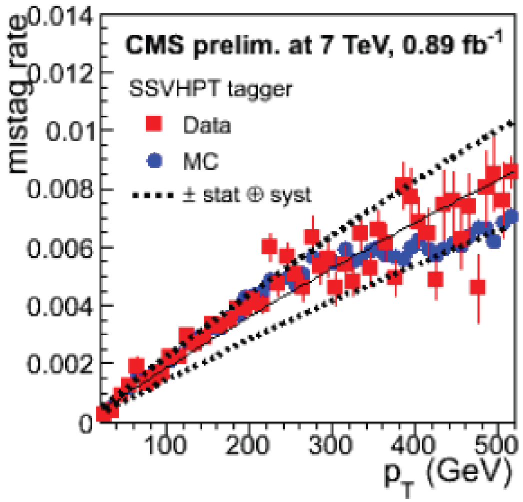


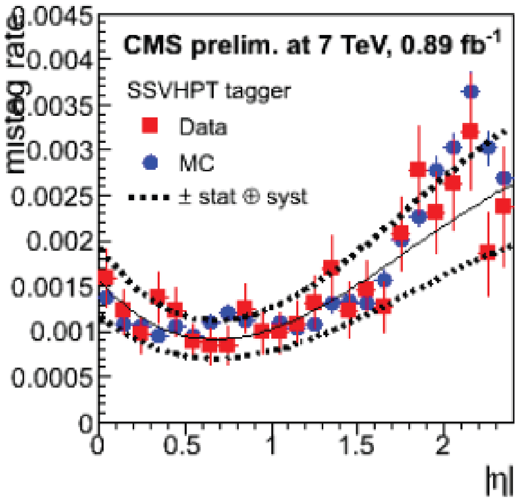
Measuring Tagging Efficiency and Fake Rates in Data
- see next talk by Saptaparna Bhattacharya for details

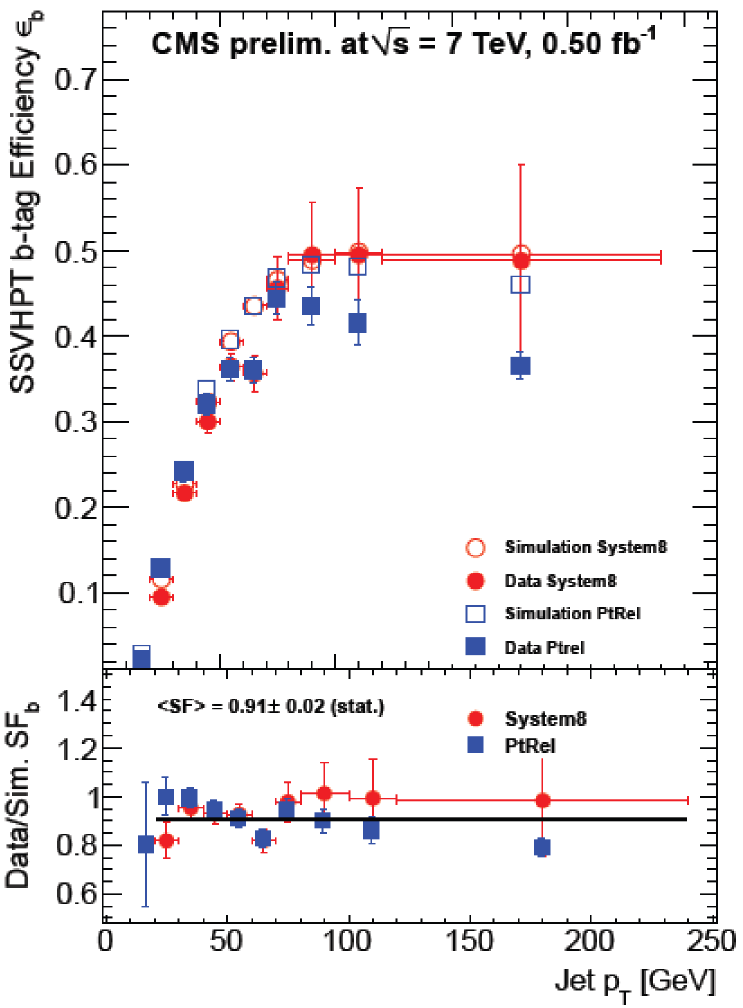
- Efficiency measured as function of jet P_T in data using lepton P_T^{rel} and System8 methods

- Fake rates measured using tags with negative impact parameter or negative decay length

- Figures show mistag rate as function of P_T and η and efficiency vs P_T for "Track Counting High Efficiency"







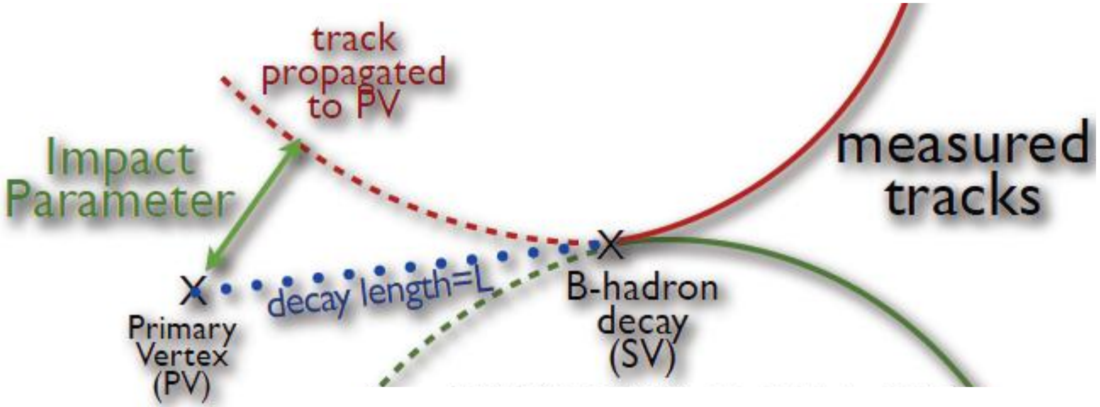
Conclusions

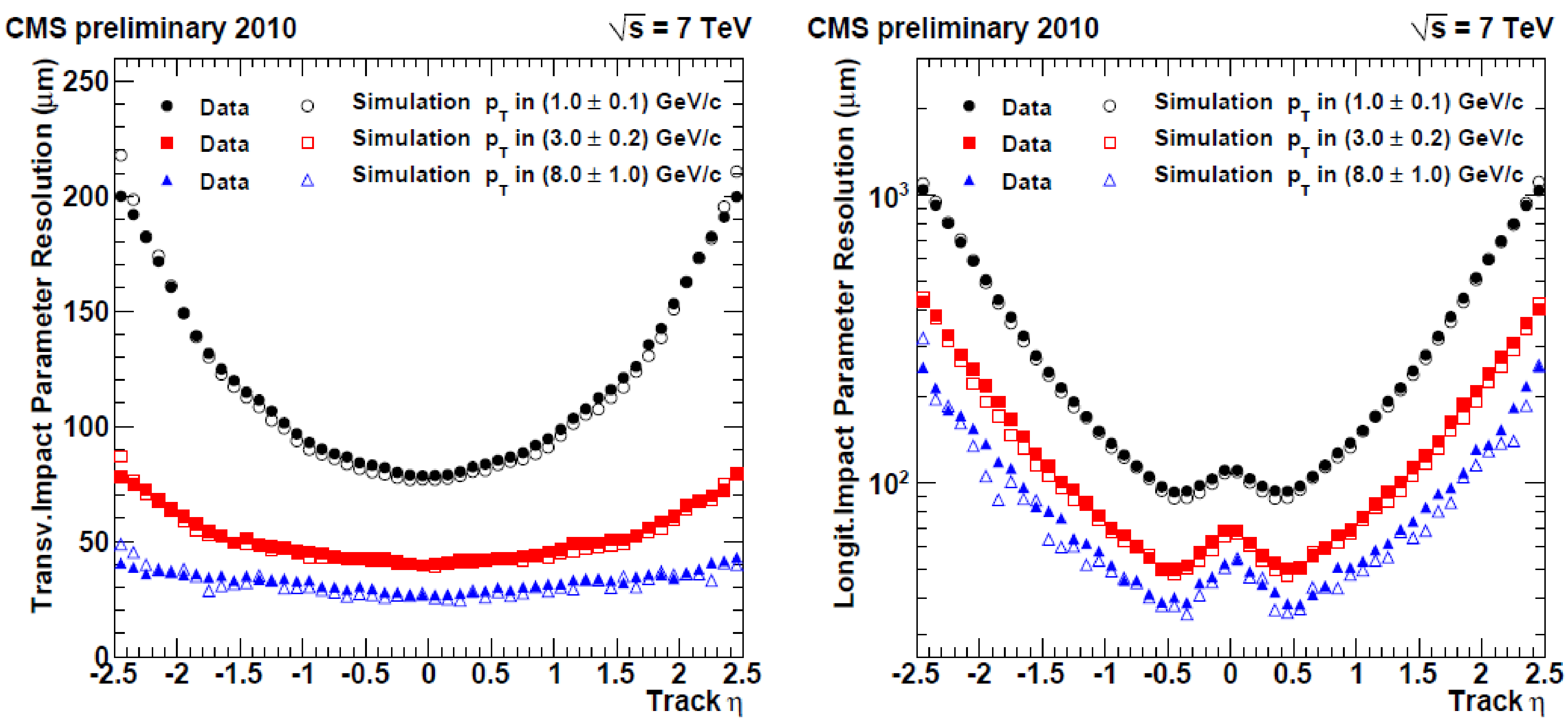
- CMS deploys a wide range of tagging algorithms, from simple and robust ones (TC, SV) to more sophisticated ones (CVS, JP) which use optimally all available b-tagging information

- Taggers successfully commissioned with 7 TeV collision data and used in physics measurents

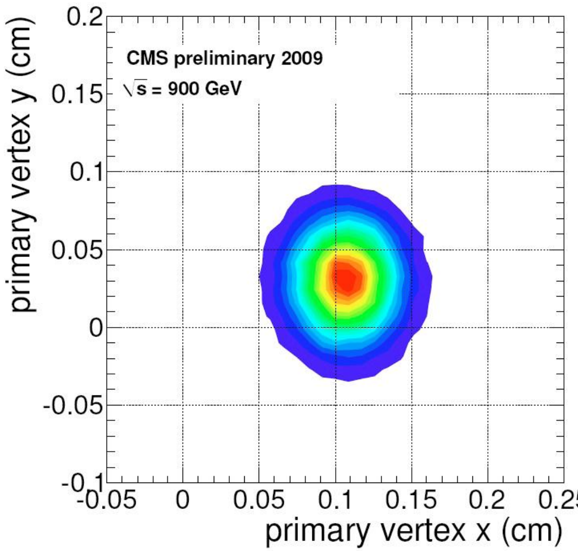
- Performance of b-tagging algorithms on data presented by Saptaparna Bhattacharya, in the next talk: *Efficiency measurement of b-tagging algorithms developed by the CMS experiment*

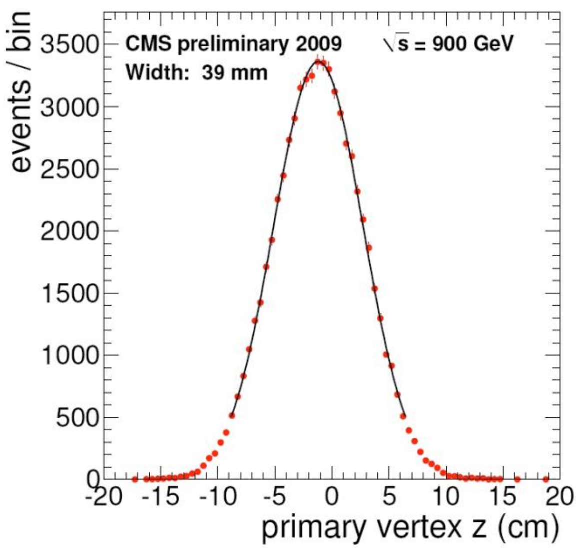
Track Impact Parameter (CMS PAS TRK 10-005)

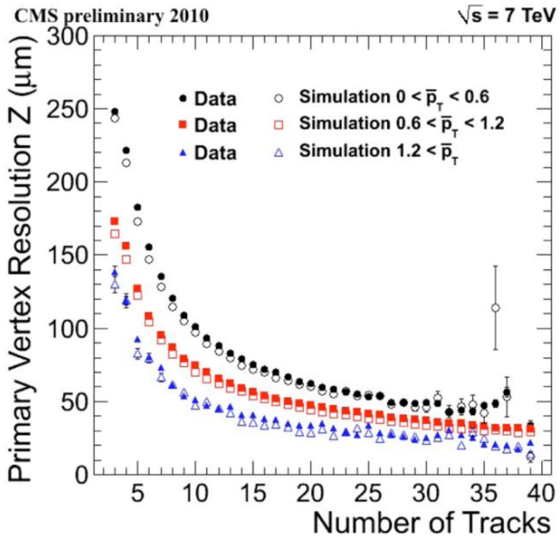


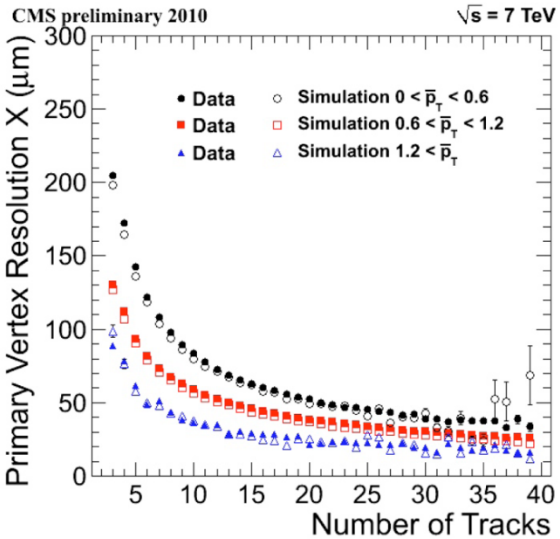


Primary Vertex







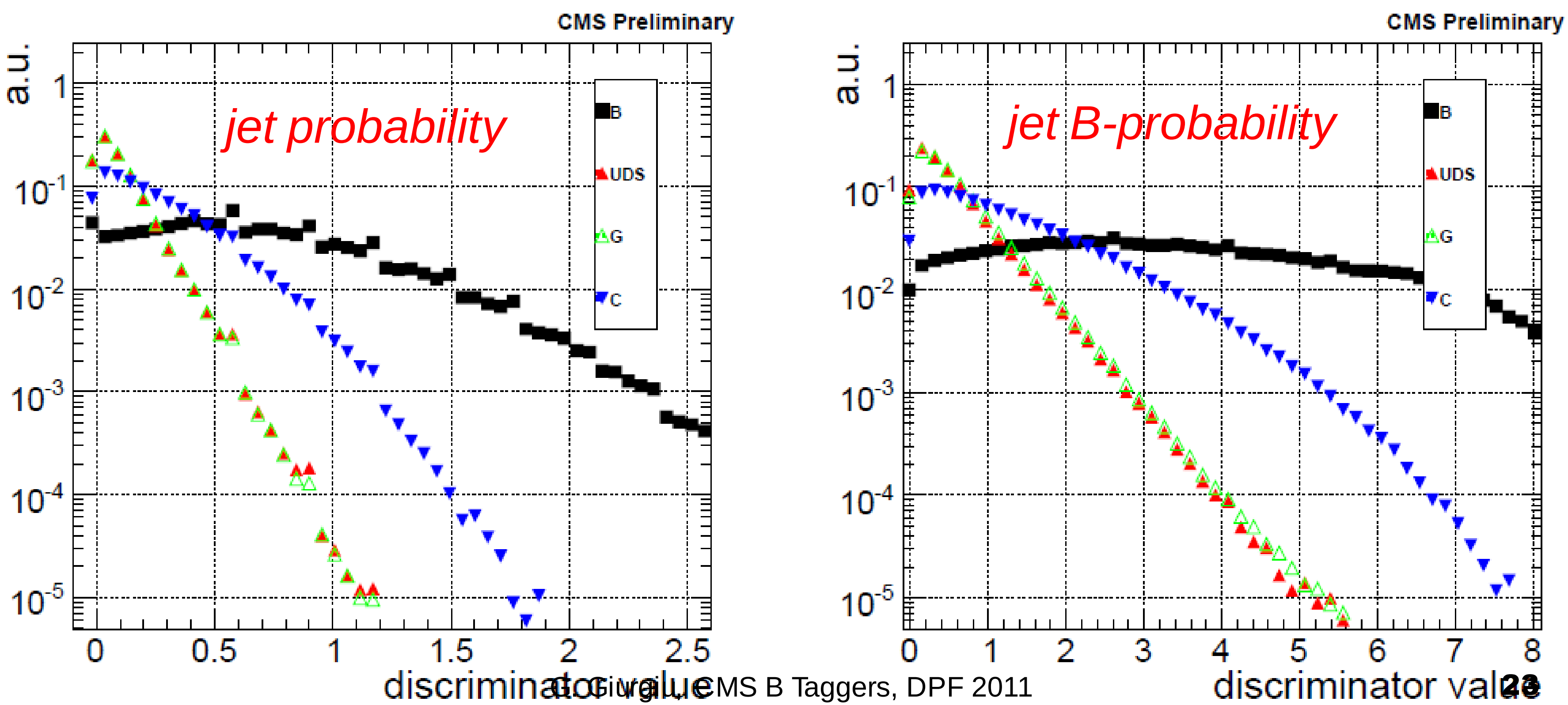


Jet Probability

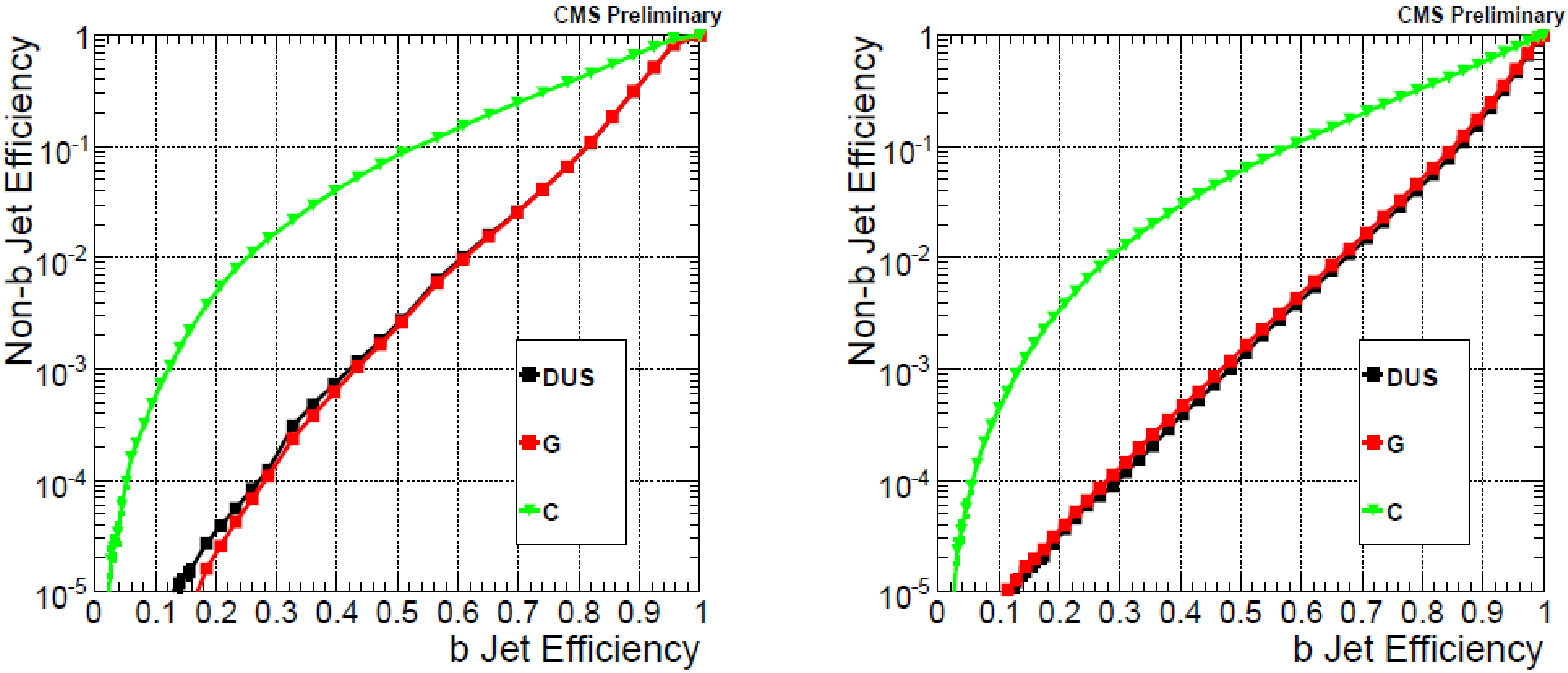
- Combines info from all selected tracks into *jet probability*

$$P_{jet} = \Pi \cdot \sum_{j=0}^{N-1} \frac{(-\ln \Pi)^j}{j!} \text{ where } \Pi = \prod_{i=1}^{N} P_{tr}(i)$$

- Or, given average track multiplicity in B decays of ~ 5 and jet tracking efficiency about 80%, could only select the most 4 displaced tracks to form *jet B-probability*



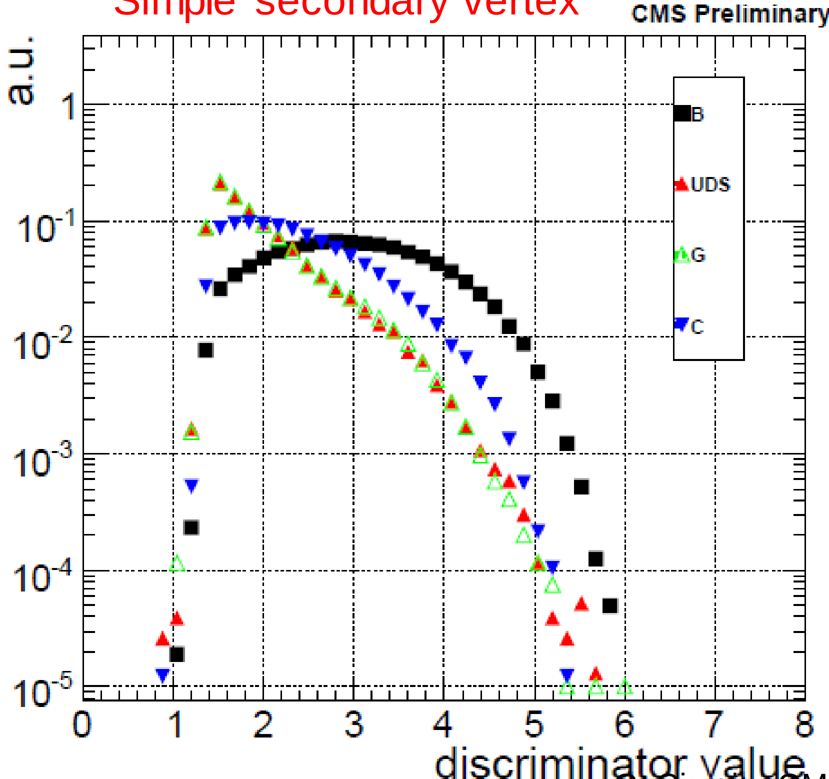
Jet Probability Performance



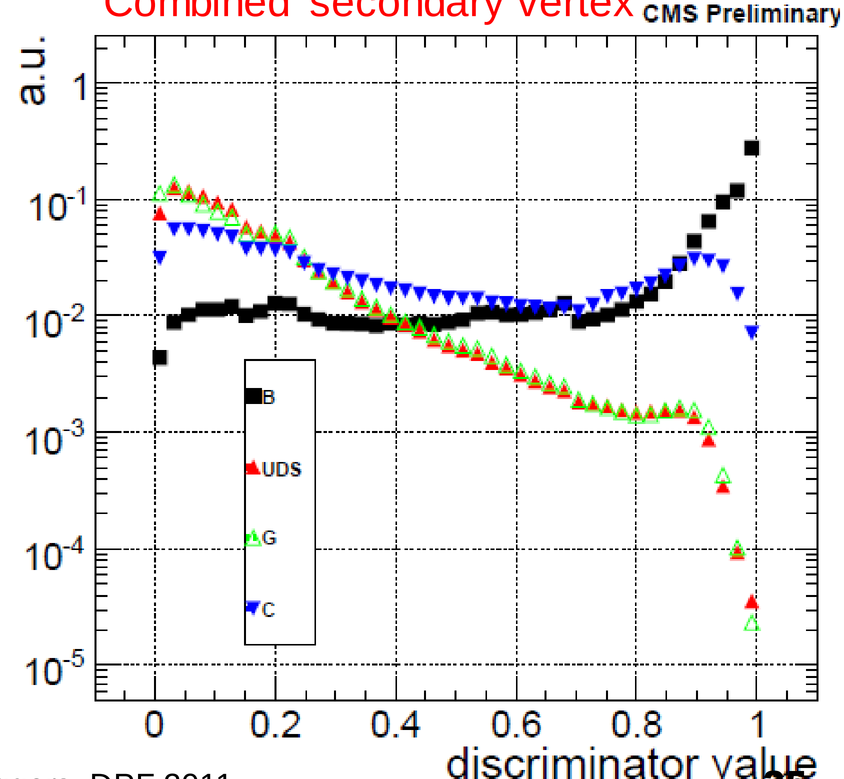
Secondary Vertex Taggers

- Based on reconstruction of at least one secondary vertex
- Efficiency of 60-70%
- Significance of flight distance from PV to SV used as discriminator
- Combined taggers use most of the available info (flight distance, vertex mass, number of vertex tracks, track IP, …) to use a likelihood discriminator

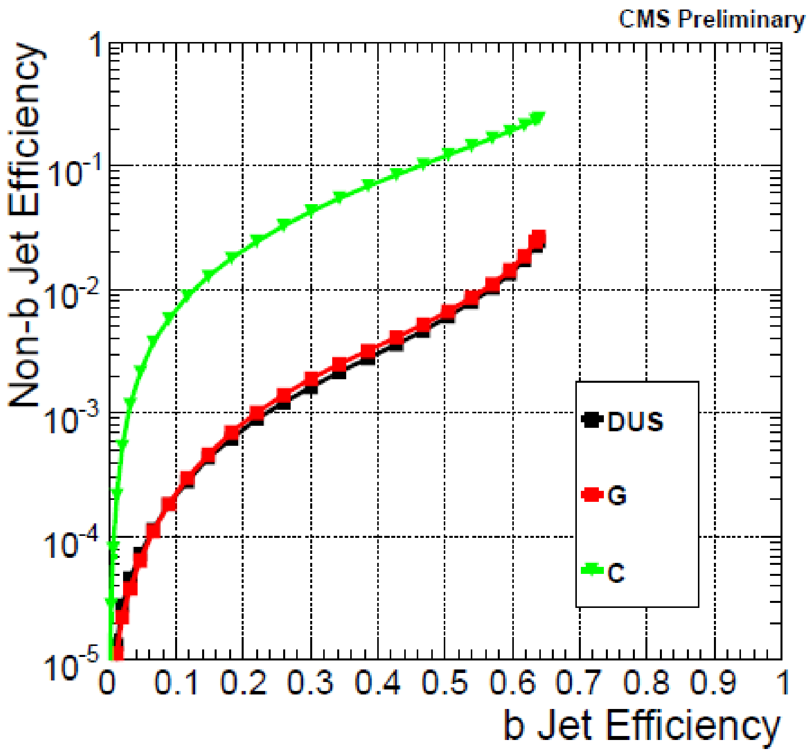
Simple secondary vertex

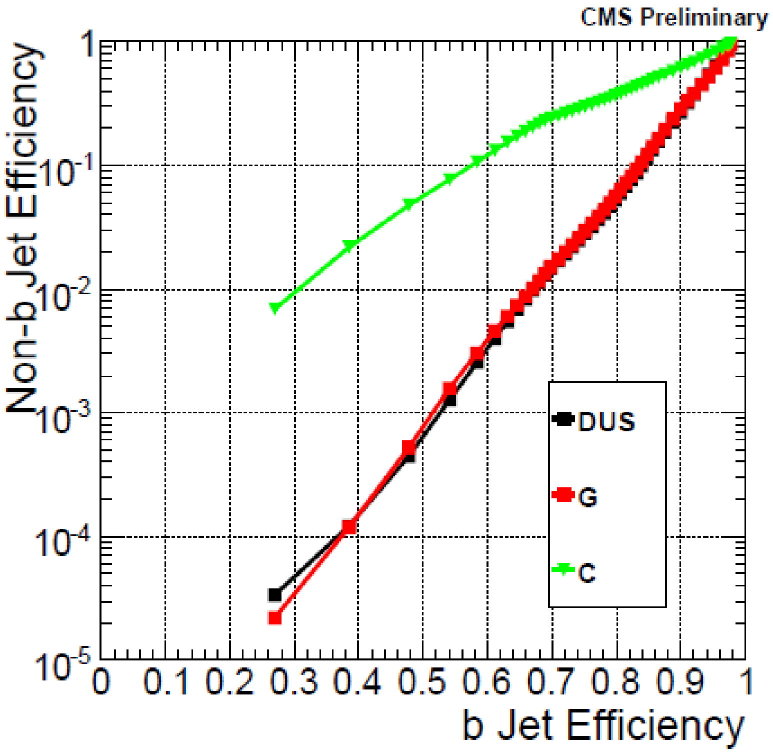


Combined secondary vertex



Performance of Secondary Vertex Taggers

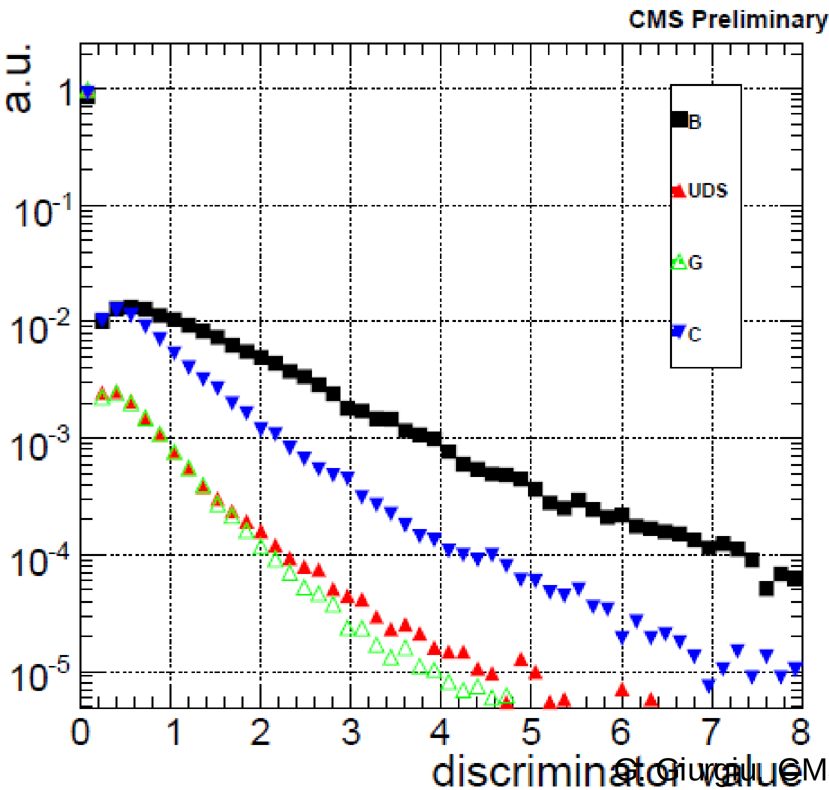




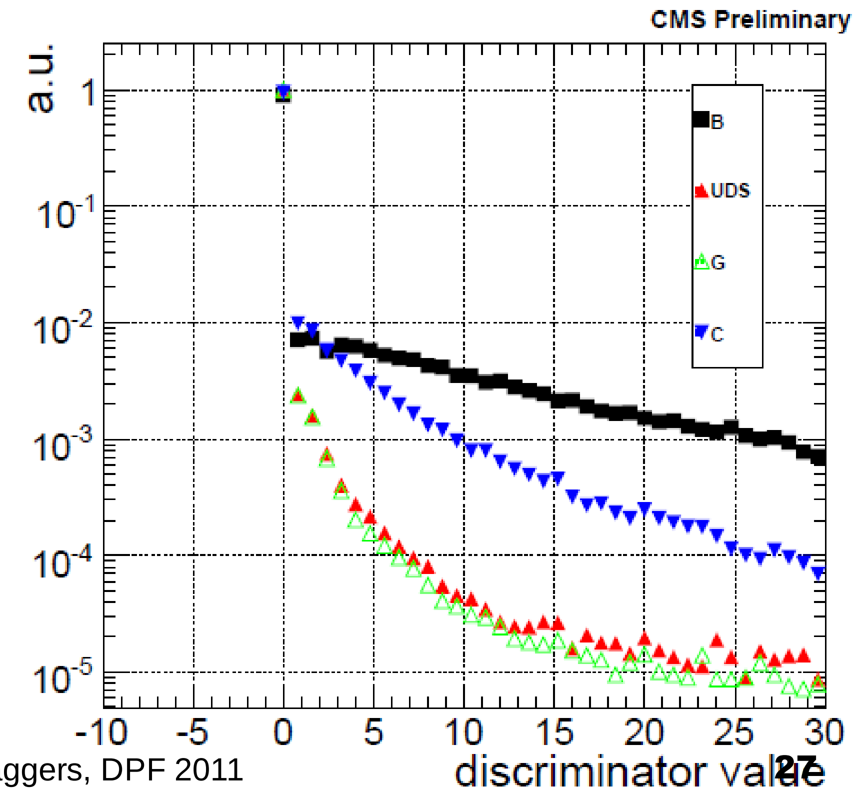
Muon Taggers

- Muons inside jets signal potential semi-leptonic B decays
- To discriminate b → μ decays from c → μ or sequential b → c → μ decays, use discriminating quantities like:

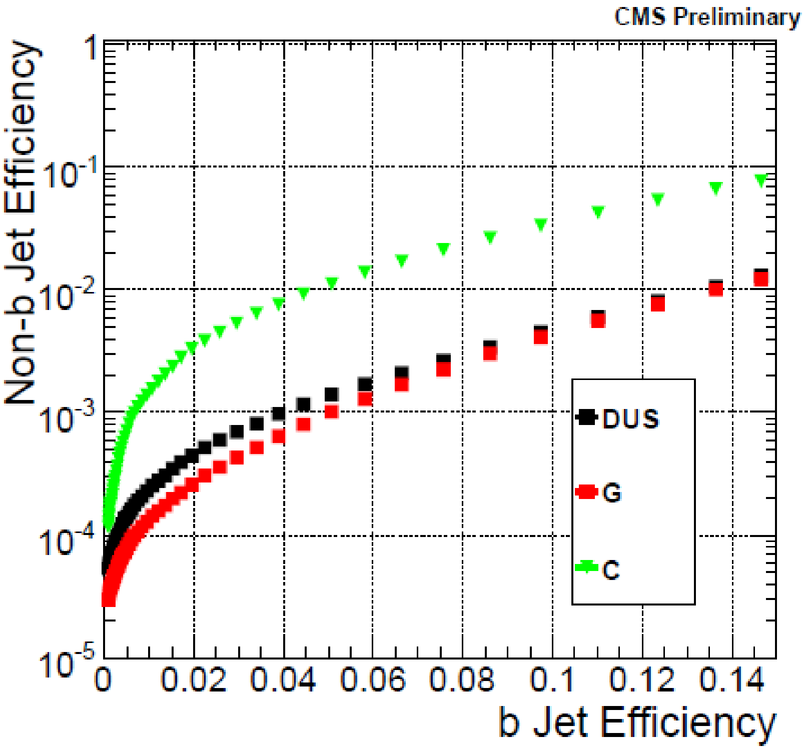
muon impact parameter significance (if positive)

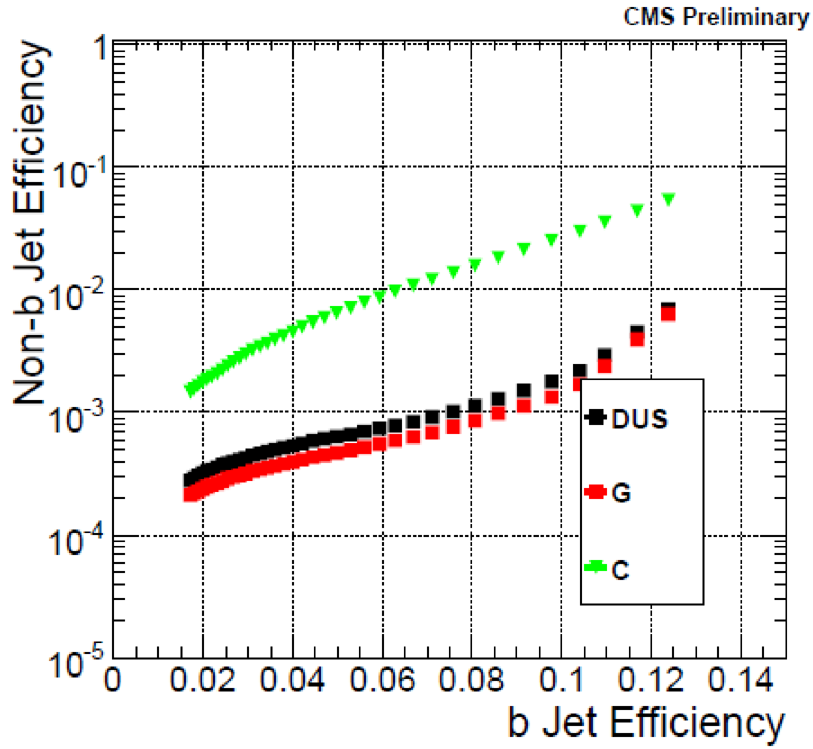


muon transverse momentum w.r.t. jet axis P_T^{rel}



Performance of Muon Taggers





Comparison of B-Tagging Algorithms (simulation)

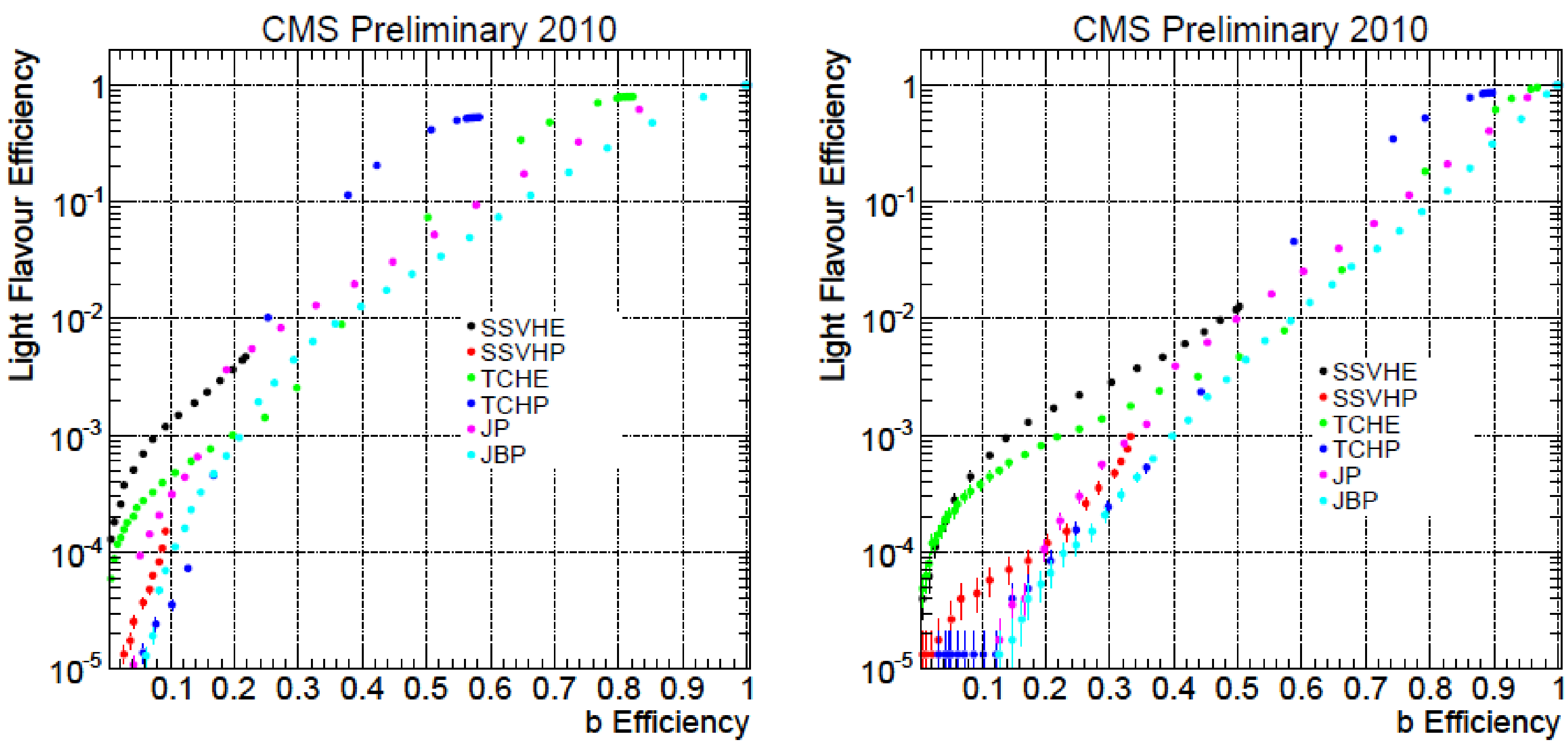
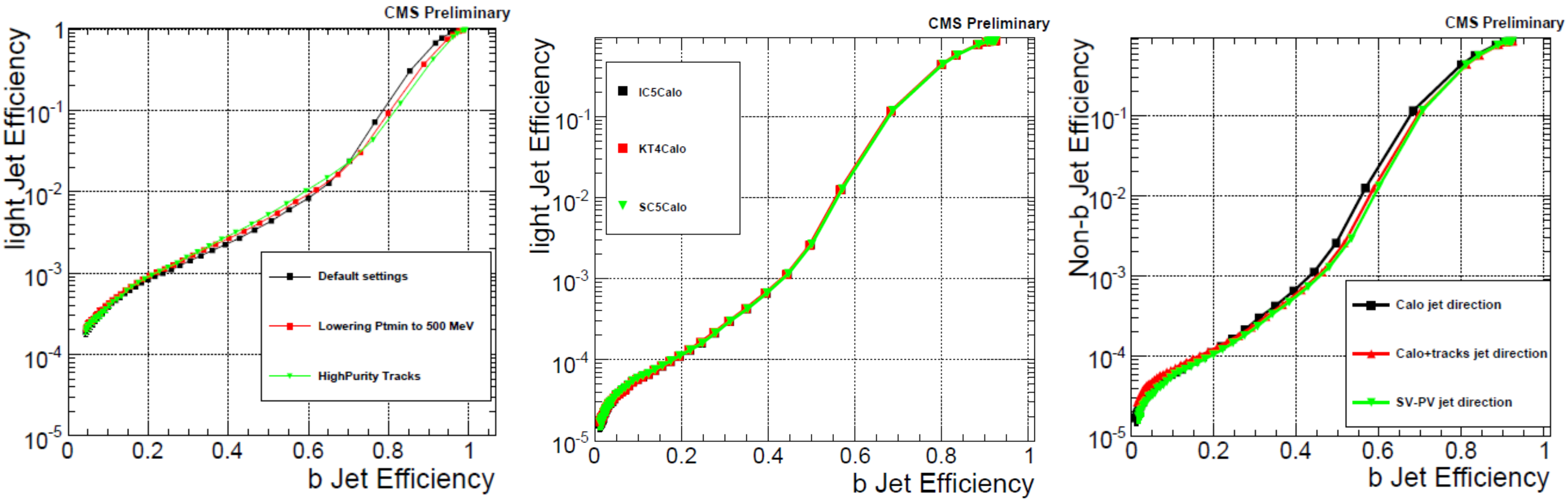


Figure 1: Efficiency for obtaining a b tag of a non-b vs true b jet for each of the tagging algorithms. The methods of [1] have been extended to the very low jet p_T bins (left) $10 < p_t < 30$ and (right) $30 < p_t < 50$.

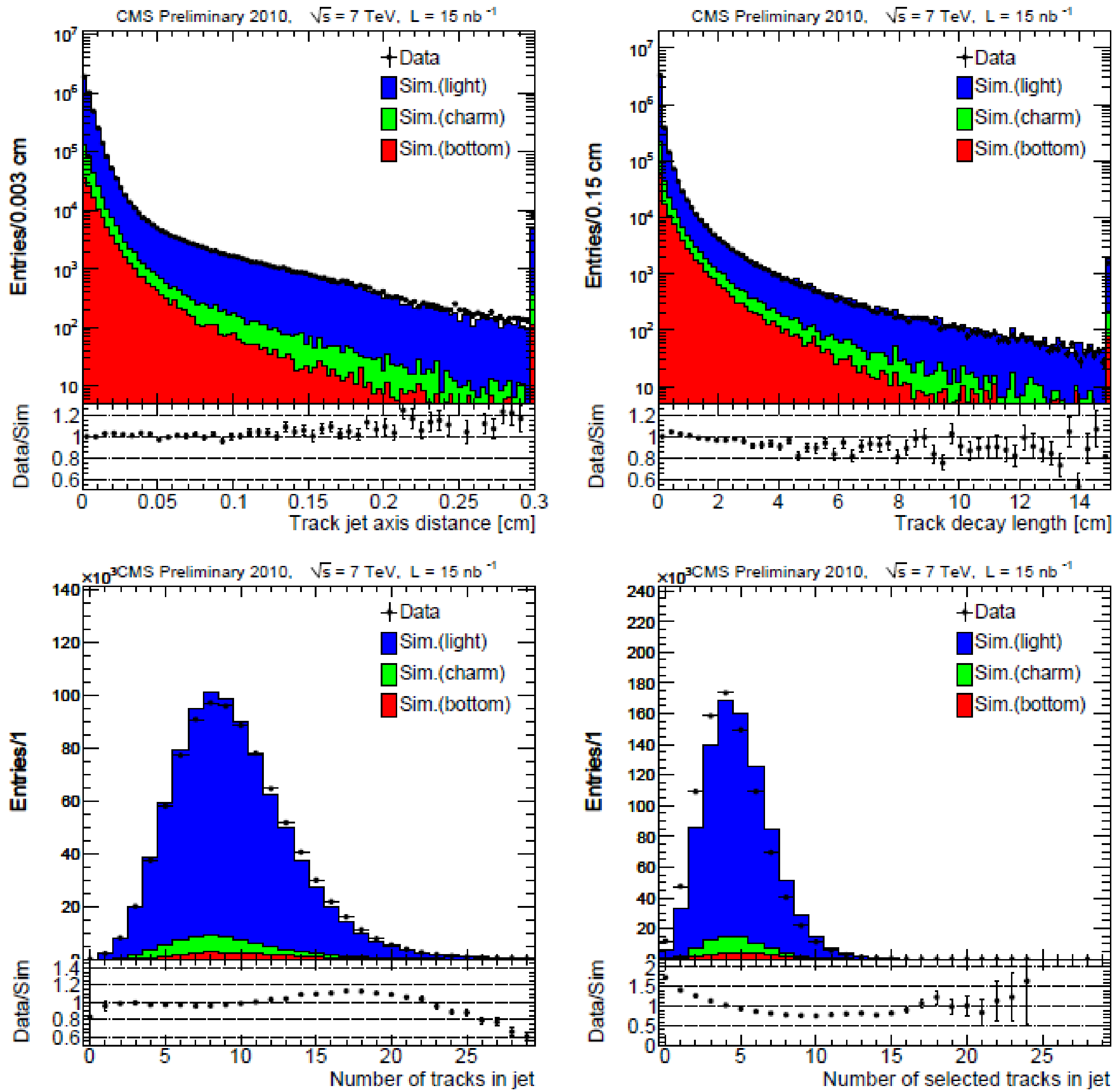
Additional Studies

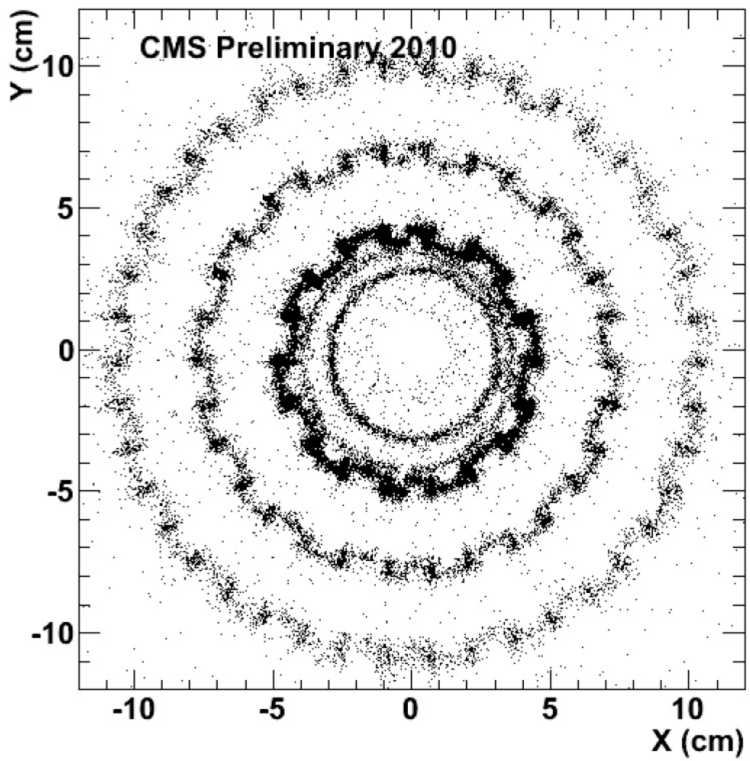
- Vary input quantities to check b-tagging robustness
  - vary minimum track P_T and track quality selection
  - change jet direction definition
  - explore different jet reconstruction algorithms

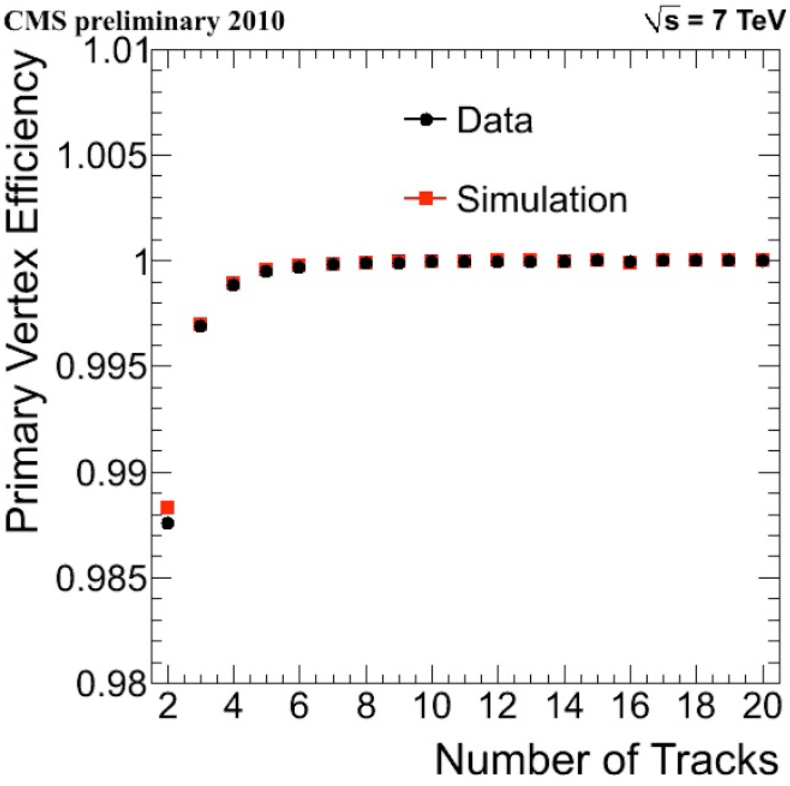
- No significant differences are observed



Validation of Tracks with 7 TeV Data (cont)




CMS Preliminary 2010
Y (cm)
X (cm)
10
5
0
-5
-10
-10
-5
0
5
10


CMS preliminary 2010
√s = 7 TeV
Primary Vertex Efficiency
Data
Simulation
Number of Tracks
1.01
1.005
1
0.995
0.99
0.985
0.98
2
4
6
8
10
12
14
16
18
20