b reconstruction, efficiency and fake rates in CMS

Jyothsna Komaragiri
on behalf of CMS collaboration

29th September 2010

3rd International Workshop on
Prospects for Charged Higgs Discovery at Colliders
Uppsala

Outline

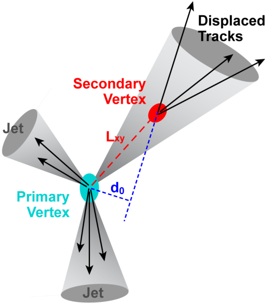
- Introduction
- *b*-tagging algorithms
- Validation of *b*-tagging inputs and discriminators
- Efficiency estimation (Scale factor=Data/Simulation)
- Mis-tag rate estimation (Scale factor=Data/Simulation)
- Conclusions

Reference: CMS Physics Analysis Summary BTV-10-001.

Introduction

Properties of b-hadrons

- The lifetime ~1.5 ps ($c\tau$ = 450 μm)
- Decay multiplicity of ~ 5 charged tracks.
- Hard *b*-fragmentation function $\Longrightarrow$ High p_T of decay products, relative to the flight direction of *b*-hadrons.
- The semi-leptonic decays, branching fraction of ~11%



Life time and Lepton signatures

- *b*-tagging is one of the main ingredients of the high p_T physics program at LHC.
  - Very pure top quark sample
  - Higgs searches
  - SUSY Higgs searches (charged and neutral)
  - Many exotic scenarios

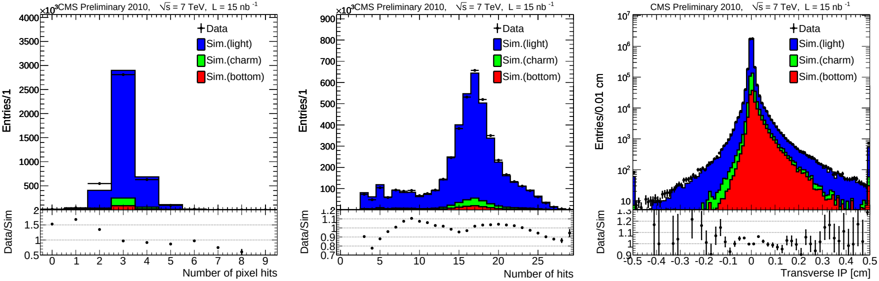
Ingredients

- Jets:
  Particle Flow jets reconstructed by anti-k_T clustering algorithm with radius parameter ΔR of 0.5
- Tracks:
  - Track quality selection cuts
    1. Number of pixel hits ≥ 2
    2. Number of total tracker + pixel hits ≥ 8
    3. Transverse Impact Parameter $|2D\ IP| < 0.2cm$
    4. Transverse momentum $p_T > 1GeV$
    5. Normalized $\chi^2 < 5$
    6. Longitudinal Impact Parameter: $|long\ IP| < 17cm$
    7. Distance to jet axis < 0.07
    8. Decay length < 5
  - Tracks are associated to the jet if $\Delta R(track, jet) < 0.5$
- Reconstructed Primary vertex

Track selection cuts

Applying all track selection cuts, except on the quantity shown

Pixel hits (Left)   Total hits (Middle)   Transverse IP (Right)

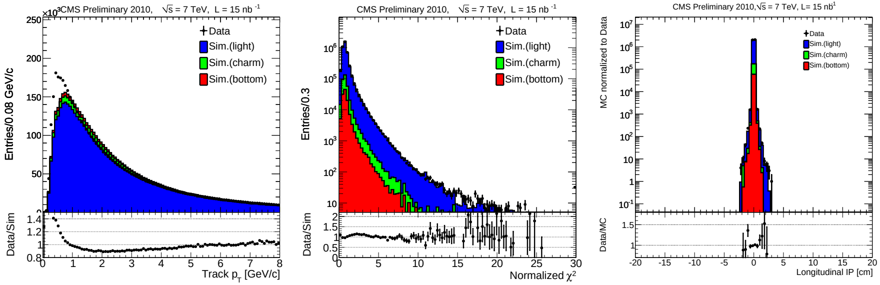


Legend:: red: b , green: charm, blue: light (u,d,s,g)

Track selection cuts

Applying all track selection cuts, except on the quantity shown

Track p_T (Left)   Normalized χ^2 (Middle)   Long IP (Right)

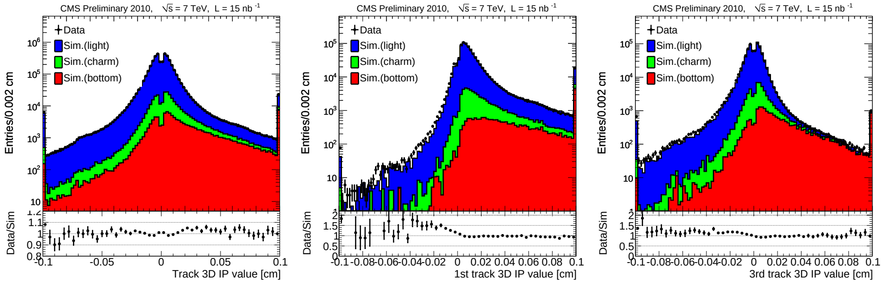


The distributions in simulation agree well with that of data.

Impact Parameter

3D Impact Parameter Value:

All tracks (Left), 1^{st} track (Middle), 3^{rd} track (Right)

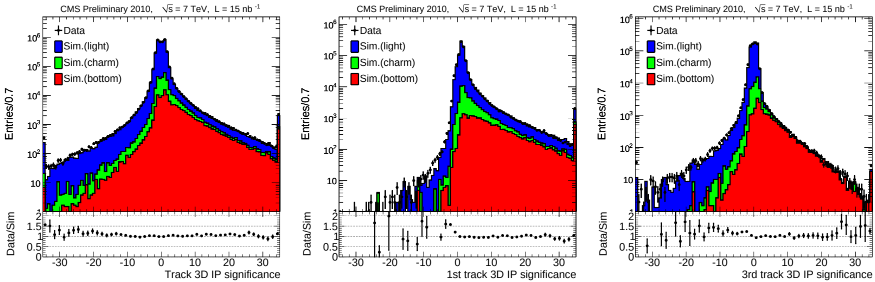


Tracks are ordered in the decreasing 3D Impact Parameter significance.

Impact Parameter

3D Impact Parameter Significance:

All tracks (Left), 1^{st} track (Middle), 3^{rd} track (Right)



The distributions in simulation agree well with that of data.

b-tagging algorithms

- **Simple taggers** are covered in this talk.
- Suitable with early data as no calibration is required.
- The discriminator for these simple taggers consists of
  - reconstructed object:
    Tracks
    Secondary Vertex
  - derived from single observable:
    Impact Parameter Significance
    3D Flight Distance Significance

Impact parameter based algorithms

Track Counting

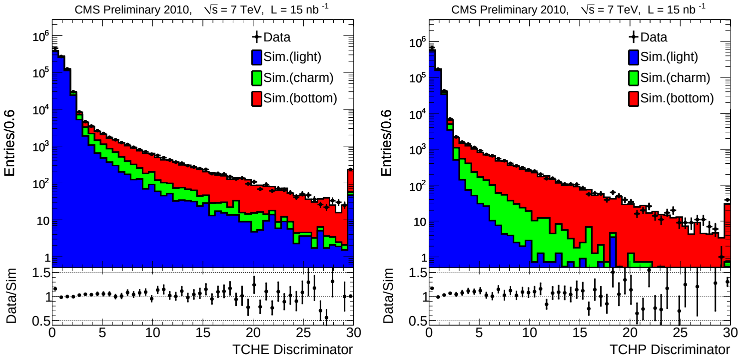
- Tags a b-jet if it finds at least N tracks with the impact parameter significance greater than a threshold, S.
- The discriminating variable is the value of S for the N^{th} track. Tracks are ordered based on decreasing significance.
  - Track Counting High Efficiency :
    Discriminator associated with N=2.
  - Track Counting High Purity :
    Discriminator associated with N=3.

Jet probability

- The jet probability algorithms combine information from all selected tracks in a jet.
- A variant of the above “JetBProbability” considers the case of four displaced tracks.

Track Counting Discriminator

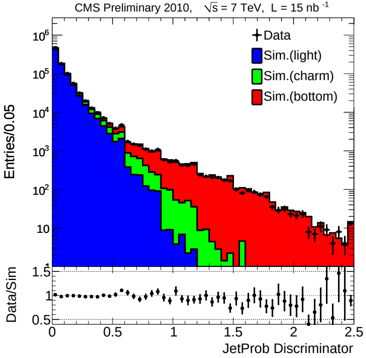
High efficiency(Left) High purity(Right)

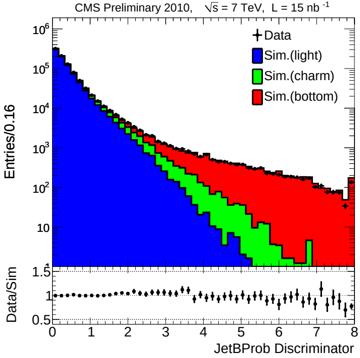


Simulation and Data agreement on 5-10% level

Jet Probability Discriminator

combines IP significance of all tracks(Left), with emphasis on four displaced tracks(Right)





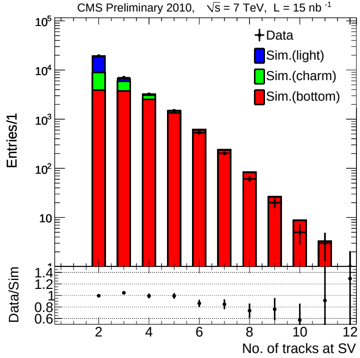
same level of Simulation/Data agreement

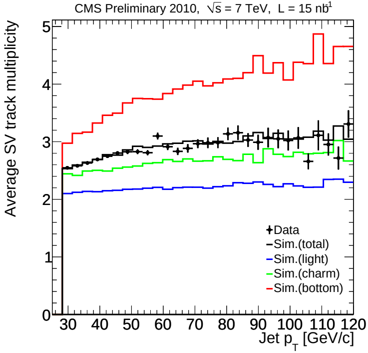
Simple Secondary Vertex algorithm

- Tags a *b*-jet if a Secondary Vertex (SV) is reconstructed with at least 2 (High Efficiency) or 3 (High Purity) tracks.
- Reconstructed SV has various properties
  - Track multiplicity at SV
  - Vertex mass: Invariant mass of tracks attached to the vertex
  - 3D flight distance significance
  - Angular separation between the jet axis and SV flight direction
- Discriminator is calculated from the 3D flight distance significance.

Secondary vertex properties

Number of tracks at SV (Left)   Average no. of tracks vs jet p_T (Right)

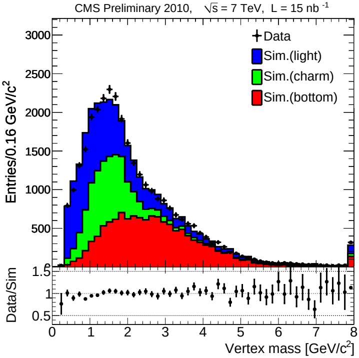


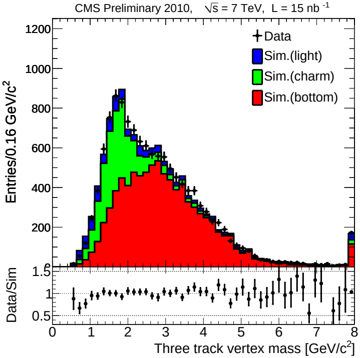


The distributions in simulation agree well with that of data.

Secondary vertex properties

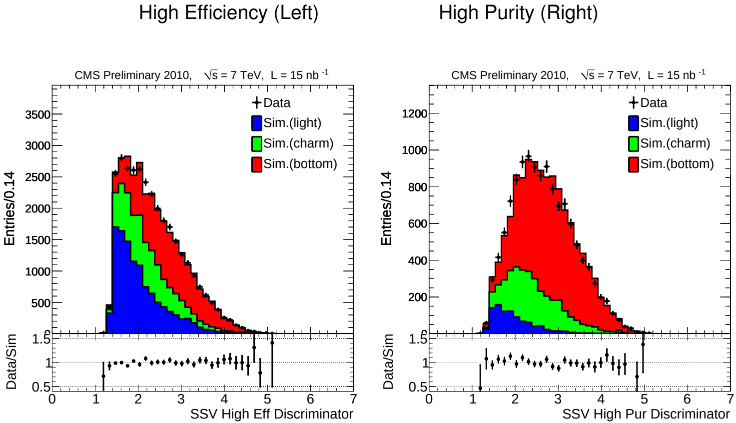
Vertex mass: $\geq$ 2 tracks (Left) $\geq$ 3 tracks (Right)





Vertex mass is used in “inclusive b” analysis to extract *b*-purity, CMS BPH-10-009

Secondary vertex discriminators



The distributions in simulation agree well with that of data.

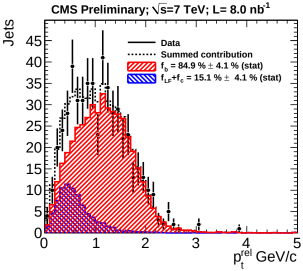
b-tagging efficiency measurements

- Select b-enriched (semileptonic decays of b-hadrons) jets
- The transverse momentum of the muon with respect to the jet direction, μp_T^{rel} is calculated.
- Hard b fragmentation $\Longrightarrow$ high μp_T^{rel} values for b-jets
- Divide the data into tagged and non-tagged samples
- From the shapes of b and non-b (c + light flavor jets), one may extract their fractions (f_b^{tag} , f_b^{untag}) with a maximum likelihood fit.
- The fractions and the total yields (N_{tag}^{data}, N_{untag}^{data}) are used to calculate the efficiency:

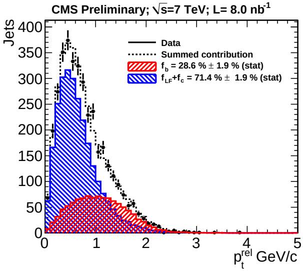
$$\epsilon_b^{\text{data}} = \frac{f_b^{\text{tag}} \cdot N_{data}^{\text{tag}}}{f_b^{\text{tag}} \cdot N_{data}^{\text{tag}} + f_b^{\text{untag}} \cdot N_{data}^{\text{untag}}}$$

Fits of μp_T^{rel}

Pass tagging (Left)



Fail tagging (Right)



Fits of the μp_T^{rel} distributions to *b* and *light* flavor templates for jets containing muons that pass (left) or fail (right) for Track Counting High Purity Tagger

Data/Simulation scale factors are compatible with 1.

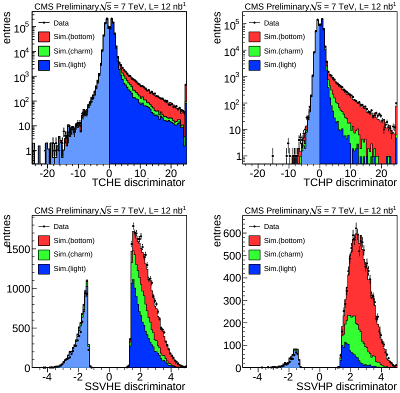
Mis-tag rate: Negative taggers

- The mis-tag rate is obtained from tracks with negative impact parameters or secondary vertices with negative decay lengths.
- The negative Imapct Parameters are ordered from the most negative upwards. Ordering on the positive side remains unchanged.
- The mis-tag rate is evaluated as:

$$\varepsilon_{data}^{mistag} = \varepsilon_{data}^{-} \cdot R_{light}$$

ε_{data}^{-} is the negative tag rate in data and $R_{light} = \varepsilon_{Sim}^{mistag} / \varepsilon_{Sim}^{-}$ is the ratio between the light flavor mis-tag rate and negative tag rate of all jets in the simulation.

Mis-tag rate: Negative taggers



Data/Simulation scale factors are compatible with 1.

Conclusions

- Good agreement between data and simulations is seen for the *b*-tagging observables indicating that the *b*-tagging algorithms are well understood.
- Very good results considering the fact that this is the very first iteration without any tuning/optimisation.
- Data driven techniques are exercised to estimate efficiency and mis-tag rate to get the Data/Simulation scale factors.
- Stay tuned for the updated results with more data!!