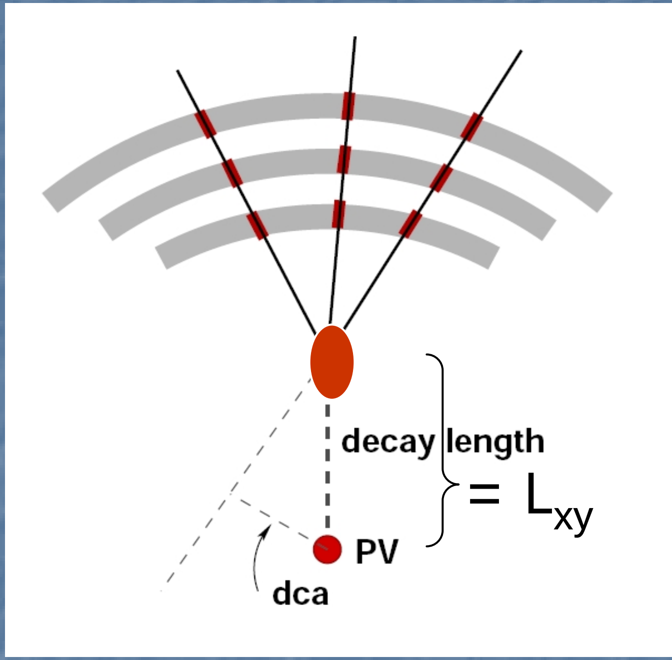
Calibration of b-tagging at Tevatron

1. A Secondary Vertex Tagger
2. Primary and secondary vertex reconstruction
3. Tagger characteristics
4. Determination of ”b-tagging efficiency”
5. (”c-tagging efficiency”)
6. Determination of ”mistag rate”
7. Systematics

Secondary Vertex Tagger algorithm

- Explicitely reconstruct secondary vertices (other options are counting displaced tracks,…)
- Used for most DØ top results $B(t \rightarrow Wb) \sim 1$
- Similar algorithm used by CDF
- Requires the position of the primary interaction (primary vertex or PV)



The Secondary Vertex Tagger Algorithm (SVT)

Three steps

I. Reconstruction and identification of a primary vertex (PV)
II. Reconstruction of track based jets ("track-jets")
III. Secondary vertex finding

Step I: determine PV on a per-event basis

1. Fit all well reconstructed tracks to a common point of origin,
2. Remove tracks with too high χ^2 contributions,
3. Repeat with remaining tracks,
4. Select main PV with p_T distribution least consistent with min. bias (DØ),
5. Select main PV closest to high p_T lepton or PV highest scalar sum of track p_T (CDF).

Step II: track based jets or ”track-jets”

1. Pre-clustering: make precluster in *z* (along beam axis) of tracks that are nearby in *z*. Start from highest p_T tracks.

2. Track selection: associate each precluster to the closest PV, use tracks that have p_T>0.5 GeV, ≥1 hit in the most precise section of the silicon, small *dca* and z_{dca}

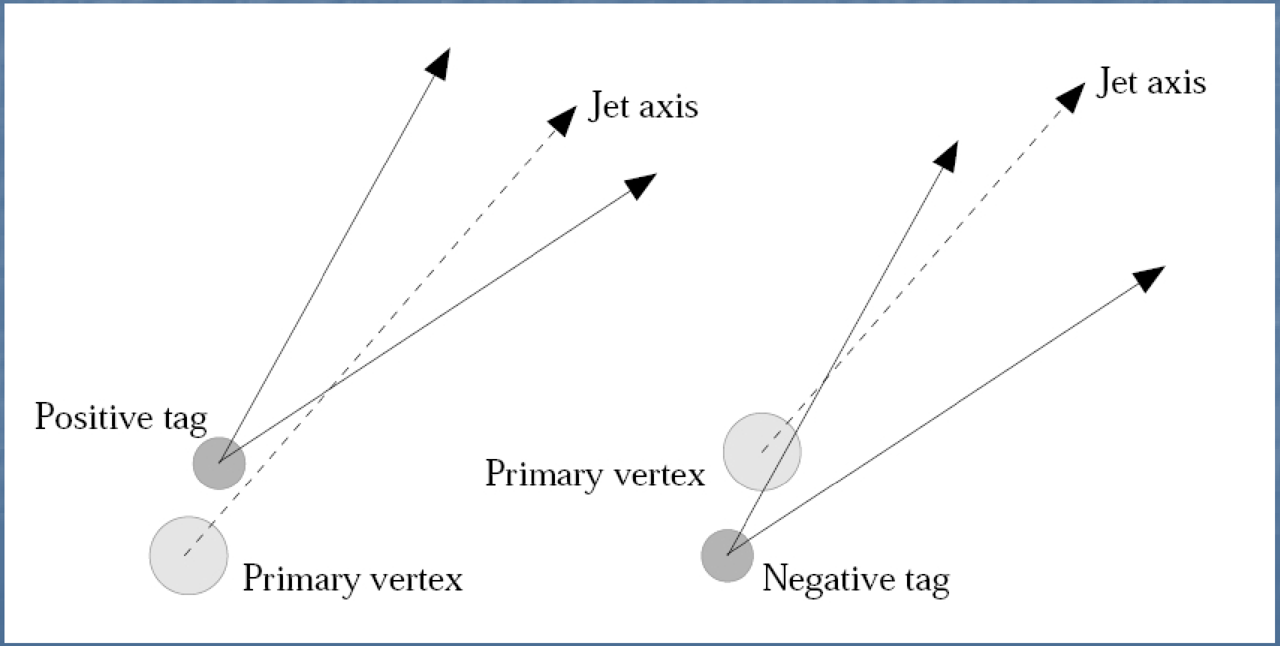
3. From the preclusters, the tracks are clustered with a simple cone algorithm, with a track seed of p_T>1 GeV.

Track-jets useful in many other situations…

Step III: Secondary vertex finding

1. Start from seed vertices in each track-jet (i.e. all pairs of tracks)
2. Add tracks to seed vertices if there χ^2 contribution is not too large
3. Select vertices with ≥2 tracks, $|L_{xy}|<2.6$ cm (within first silicon layer!), $L_{xy} > n \times \sigma(L_{xy})$, (adjust n to required rejection), χ^2,... (2 steps in CDF.)

”b-tagged” = there is ≥1 SV within $\Delta R=0.5$ of the calorimeter jet.



Tagger characteristics

- Probability to tag a b-jet =
  ”b-tagging efficiency”

- Probability to tag a light jet (g, u, d, s)
  ”mistag rate”

- Probability to tag a c-jet
  ”c-tagging efficiency”

These parameters are in general functions of the jet pT and η,
Could also be dependent on the PV position, the luminosity, run range …

Probability to tag a b-jet

To decouple from detector issues, define (CDF and D∅)

I. Taggable jets, (experiment wide definition)
II. Tagged jets

A calorimeter jet is taggable if:

1. E_T>15 GeV, $|\eta|<2.5$, (i.e. Jet energy scale is defined!, detector dependent)
2. If it contains a track-jet within $\Delta R<0.5$
3. Some quality requirements on the track-jet.

Later on, derived in 3 regions of z_{PV} (D∅)

Taggability:

$$\text{Taggability}(E_T,\eta) = \frac{\#\ \text{taggable jets}\ (E_T,\eta)}{\#\ \text{jets}(E_T,\eta)}$$

Different parameters In CDF

Taggability

Taggability must be derived from ”generic QCD” data

Use same trigger than the signal sample, to incorporate luminosity, run number dependences…

For example ttbar *l+4 jet* signature, take the events passing the lepton+1 jet trigger

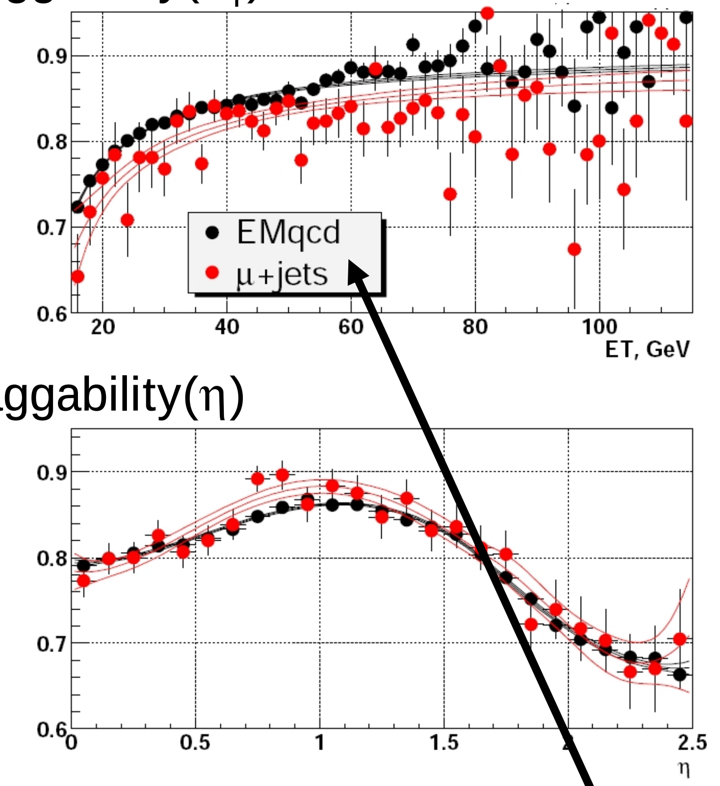
Signal fraction is ~10^{-4} : so no bias.

Compute taggability in bins of η and p_T

Taggability(E_T,η) ≈ k × Taggability(E_T) × Taggability(η)

Taggability(E_T)

Taggability(η)



Sample dependence:
-Low MET passing EM trigger
-μ + jet + high MET sample

Taggability and jet flavor

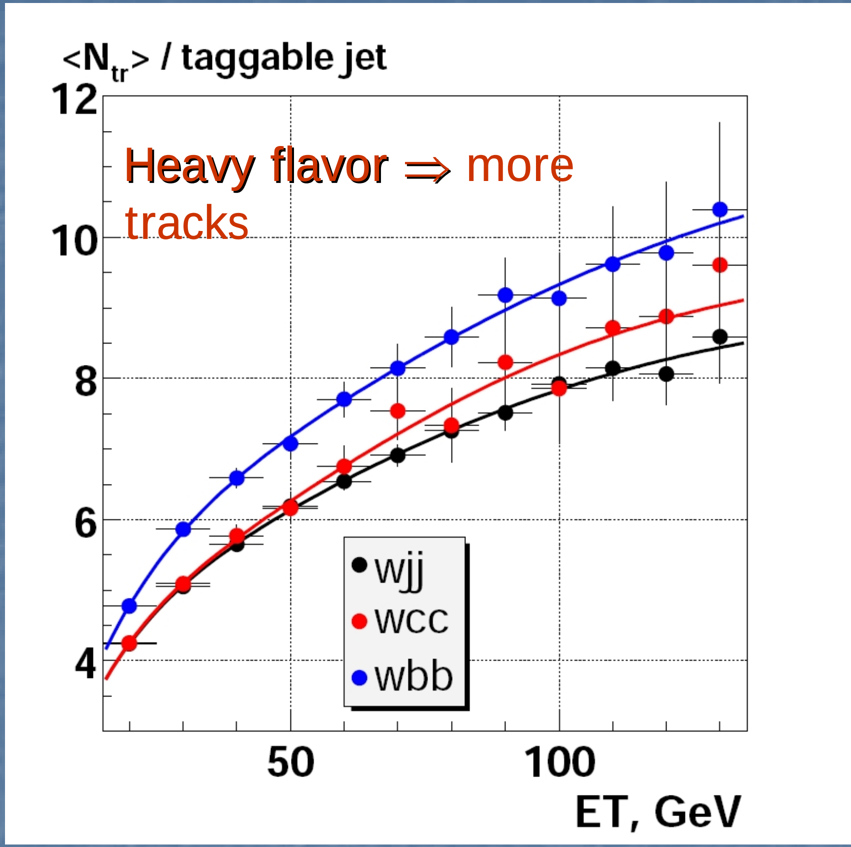
Taggability derived from data is valid for light flavor jets ONLY.

Higher taggability for heavy flavor jets

Derive correction from MC

$$C_{taggabiliy}(\text{flavor}) = \frac{\text{Taggability}(E_T,\eta,\text{flavor})\text{ in MC}}{\text{Taggability}(E_T,\eta,\text{light jets})\text{ in MC}}$$

Cross check ratio of heavy-enhanced-to-light taggability in data and MC, agreement better than 2% level



$$\text{Taggability}(E_T,\eta,\text{flavor}) = C_{taggability}(\text{flavor}) \times \frac{\#\text{ taggable jets }(E_T,\eta)}{\#\text{ jets}(E_T,\eta)}$$

b-tagging efficiency

b-tagging efficiency is defined by

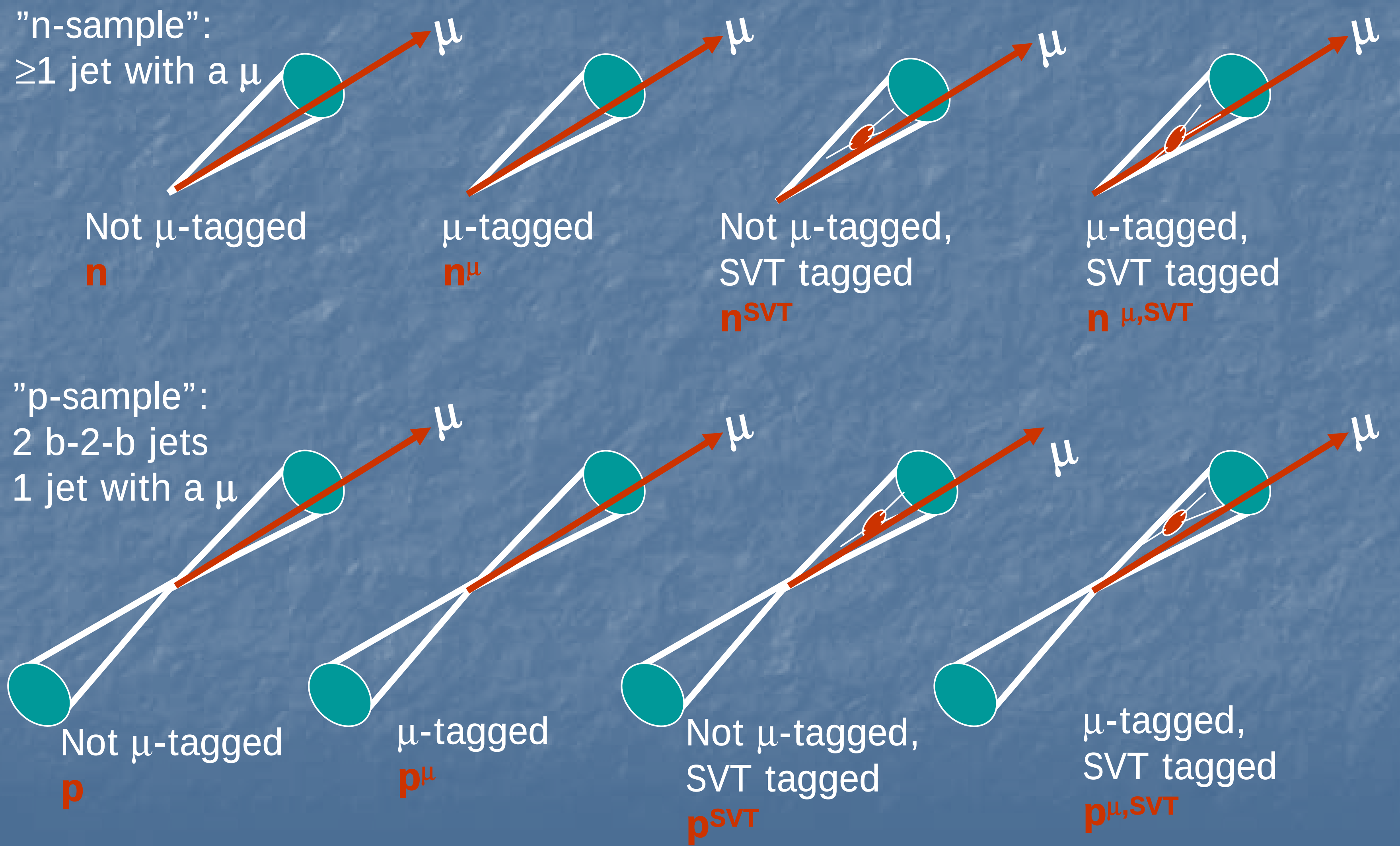
$$\varepsilon^b(E_T,\eta) = \frac{\#\ \text{tagged jets}\ (E_T,\eta)}{\#\ \text{taggable jets}\ (ET,\eta)}$$

Derive this quantity from data using a sample enhanced in heavy flavor.

Typically back-to-back dijet events with various taggers: SVT,
soft muon or electron tagger
(D∅ = a muon inside a jet with p_T^{rel}>0.7 GeV, CDF electron inside a jet)

Method introduced in D∅ by LEP folks…

b-tagging efficiency from data



Solve system of 8 equations, with 8 unknowns (in bins of η and p_T)

$$
\begin{aligned}
n &= n_b + n_{cl} \\
p &= p_b + p_{cl} \\
n^{\mu} &= \varepsilon_b^{\mu} n_b + \varepsilon_{cl}^{\mu} n_{cl} \\
p^{\mu} &= \varepsilon_b^{\mu} p_b + \varepsilon_{cl}^{\mu} p_{cl} \\
n^{SVT} &= \varepsilon_b^{JLIP} n_b + \varepsilon_{cl}^{JLIP} n_{cl} \\
p^{SVT} &= \beta\, \varepsilon_b^{JLIP} p_b + \alpha\, \varepsilon_{cl}^{JLIP} p_{cl} \\
n^{\mu,SVT} &= \varepsilon_b^{\mu} \varepsilon_b^{JLIP} n_b + \varepsilon_{cl}^{\mu} \varepsilon_{cl}^{JLIP} n_{cl} \\
p^{\mu,SVT} &= \beta\, \varepsilon_b^{\mu} \varepsilon_b^{JLIP} p_b + \alpha\, \varepsilon_{cl}^{\mu} \varepsilon_{cl}^{JLIP} p_{cl}.
\end{aligned}
$$

events that are c- or light- jets

ε^{μ}_{b} (E_T,η) efficiency of μ-tagger
ε^{SVT}_{b}(ET,η) efficiency of 2nd tagger

b- contribution c/light contribution

⇒ Extract: sample composition and efficiency of the taggers

⇒ Makes a number of assumptions… → systematic errors

System 8 assumptions and systematics

- Decorrelation of the 2 taggers:

$\varepsilon^{\mu,SVT} = c \times \varepsilon^{\mu} \times \varepsilon^{SVT}$, assume c=1 (MC gives c=1.01±0.01)

- Assume that the μ-tagger has same efficiency for c- and light-jets, ok because p_T^{rel} has similar shape for c- and light-jets at Tevatron energy.
  - Compare p_T^{rel} templates from several generators
- Assume that c- and light-jet backgrounds can be lumped together, this is characterised by a factor α (varied for systematics)
- Solve the system for various values of p_T^{rel} cut 0.3 - 1.5 GeV.
- β~1 takes into account correlations b/w p and n samples (varied foe systematics)

b-tagging efficiency

From data we can only extract b-tagging efficiency for muonic b-jets

$$\varepsilon^{b\to\mu,\ data}(E_T,\eta)$$

We need the b-tagging efficiency for ”all kinds of b-jets”

$$\varepsilon^{b}(E_T,\eta)$$

$$P^{b}_{tag}(E_T,\eta) = \frac{\varepsilon^{b,MC}(E_T,\eta)}{\varepsilon^{b\to\mu,MC}(E_T,\eta)} \times \varepsilon^{b\to\mu,\ data}(E_T,\eta) \times \text{Taggability}(E_T,\eta) \times C_{taggability}(b)$$

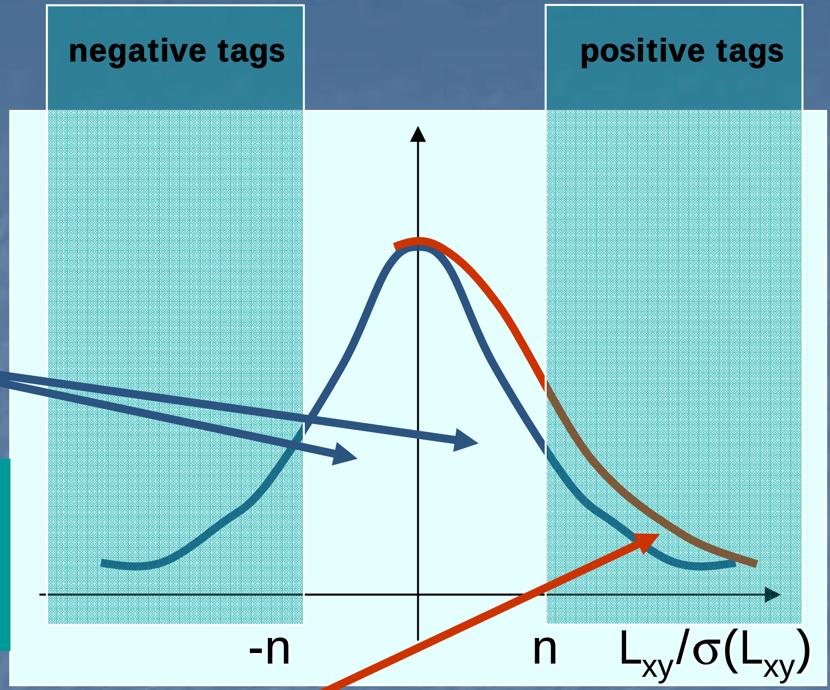
Transform semi-muonic b-tag efficiency into inclusive one

Mistag rate

Extract from data as much as possible
Similar approach in CDF and D∅

The number of negative tags gives information on resolution effects

$$\varepsilon\_^{data}(E_T,\eta) = \frac{\#\ \text{negatively tagged jets}\ (E_T,\eta)}{\#\ \text{taggable jets}\ (E_T,\eta)}$$

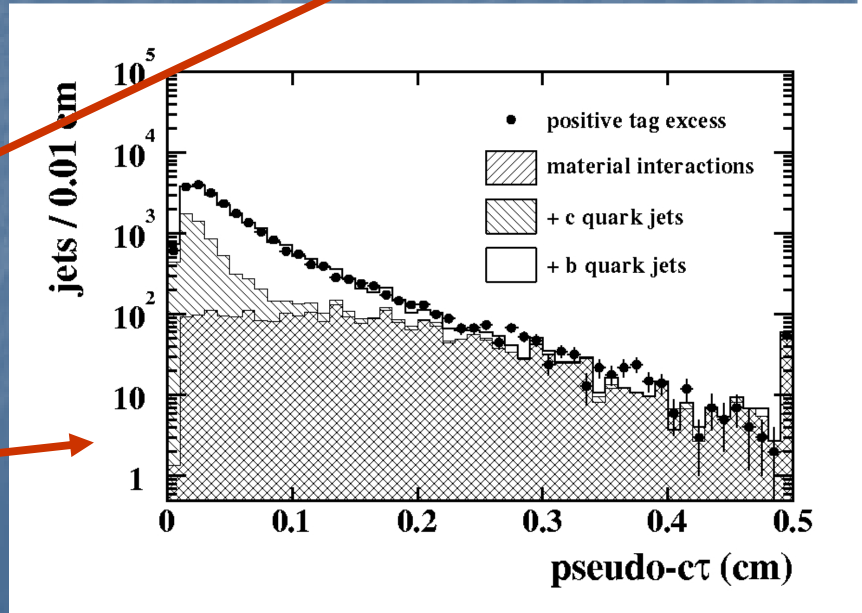


Must take into account:

1) Long-lived particles in light jets are not completely removed by V^0 filter: contribute to positive tags in light jets

2) Contamination of negative tags data by heavy flavor (2% b-jets and ~4% c-jets)

CDF fits contributions to observed pseudo-cτ



$\varepsilon\_^{data}(E_T,\eta)$ extracted in data passing e+jet trigger, with MissingET<10 GeV

Validation:

1. Alternative parametrization derived from single electron trigger
2. Compare predicted and observed number of negative tags in high MissingET region

Correct for long-lived particles in light-jet sample:

SF_{ll} = #negative tags/#positive tags in light-flavor QCD Monte Carlo

Correct for the fraction of heavy flavor in the low MissingET electron sample

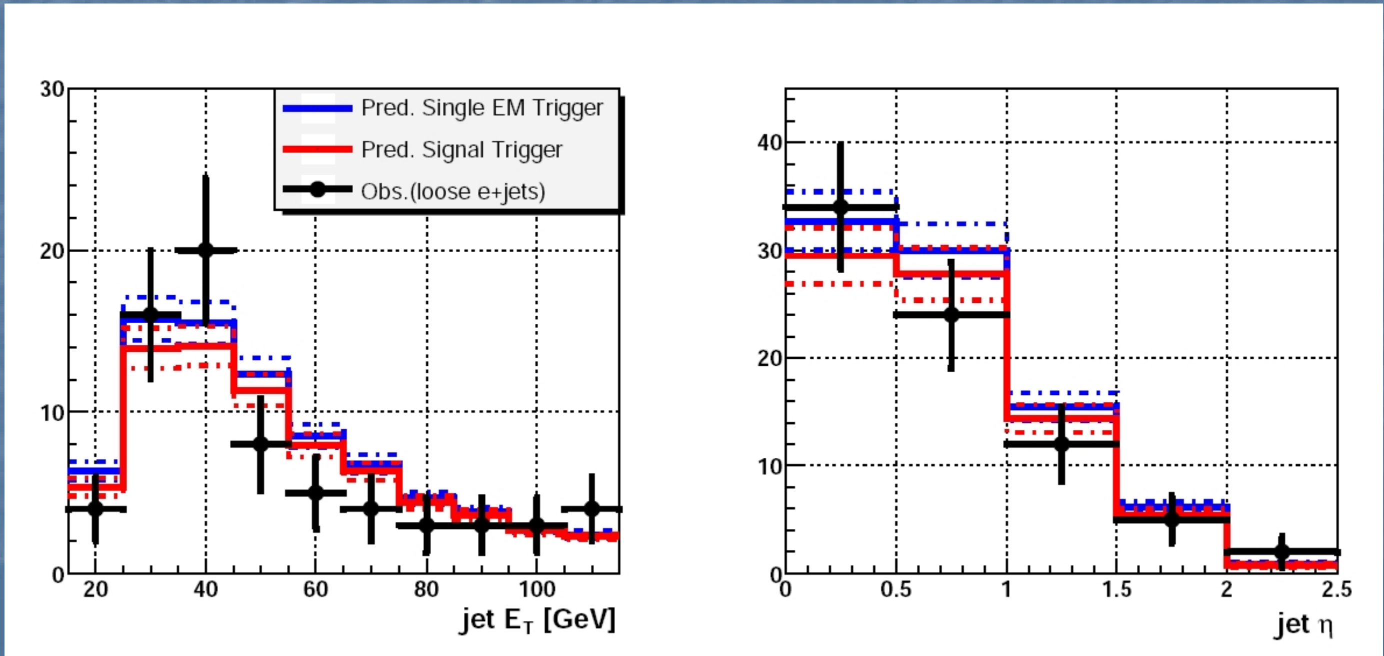
SF_{hf} = #positive tag from light flavor / # positive tag from all flavors

$$\varepsilon^{light}(E_T,\eta) = \varepsilon\_^{data}(E_T,\eta) \times SF_{hf} \times SF_{ll}$$

Negative tag rate validation

Alternative parametrizations derived:
- from single electron trigger
- (instead of e+jets)

Compare predicted and observed number of negative tags in high MissingET region



Systematic uncertainties

I. **Taggability**
   1. Statistical error on parametrization from data
   2. Variation on the parametrization by changing sample
   3. Difference b/w predicted and observed # taggable jets at high Njet
   4. Flavor dependence of taggability: MC dependence

II. **b-tagging efficiency**
   1. Statistical error on semi-muonic b-tagging parametrization from data
   2. System-8 assumptions
   3. Ratio of semi-muonic to inclusive b-tagging efficiency in MC (statistical+sample dependence)

III. **c-tagging efficiencies**

IV. **Mistag rate**
   1. Negative tag rate, data statistics
   2. Negative tag rate, sample dependance
   3. Heavy flavor contamination
   4. Negative to positive tag ratio for light flavor jets