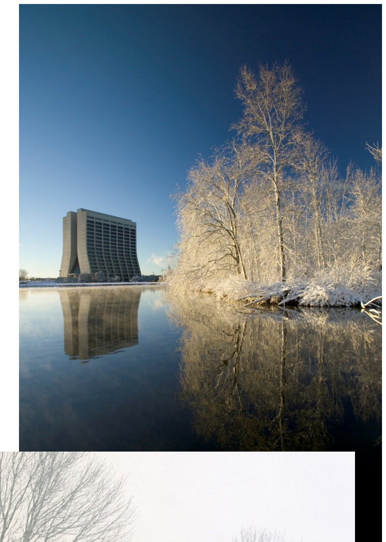
High p_T *b*-tagging at CDF: Measuring Efficiency and Understanding Mistags

Christopher Neu

University of Pennsylvania

on behalf of the CDF Collaboration

Top2006 Workshop
13 January 2006
University of Coimbra, Portugal



Outline:

- **Challenge of *b*-tagging at a hadron machine**
- **Lifetime-based *b*-tagging at CDF**
- **Measuring efficiency in the data**
- **Understanding contribution from non-*b* sources**
- **Other CDF *b*-tagging techniques**
- **Considerations for LHC experiments**
- **Summary**

b-Tagging at the Tevatron

- The **ability to identify jets originating from *b* quark production** is critical for several facets of the Tevatron RunII physics program – top, Higgs, exotic searches, QCD…

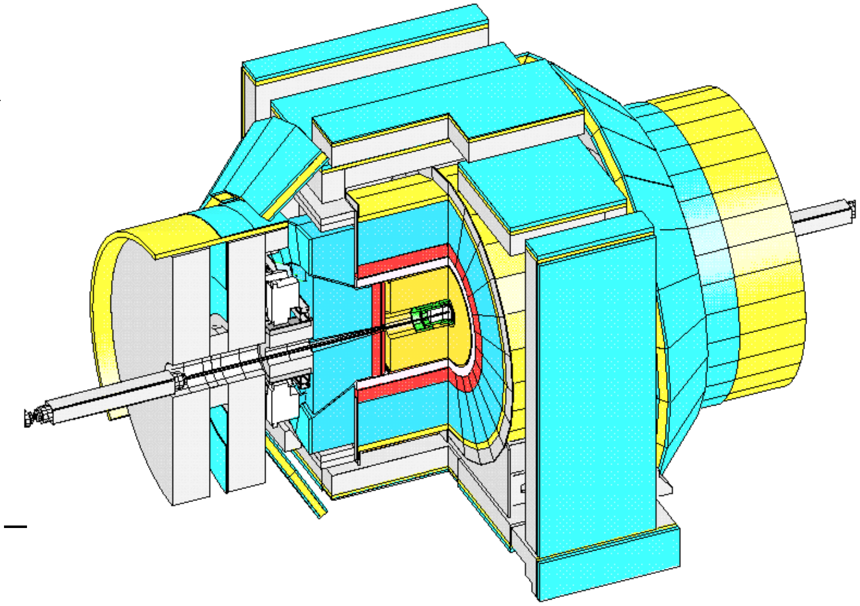
- **Distinguishing jets from *b* quarks** from light flavor and charm:
  - The **long lifetime** of the *b*
  - The **large mass** of *B* hadrons
  - The **energetic semileptonic decay** of *B* hadrons

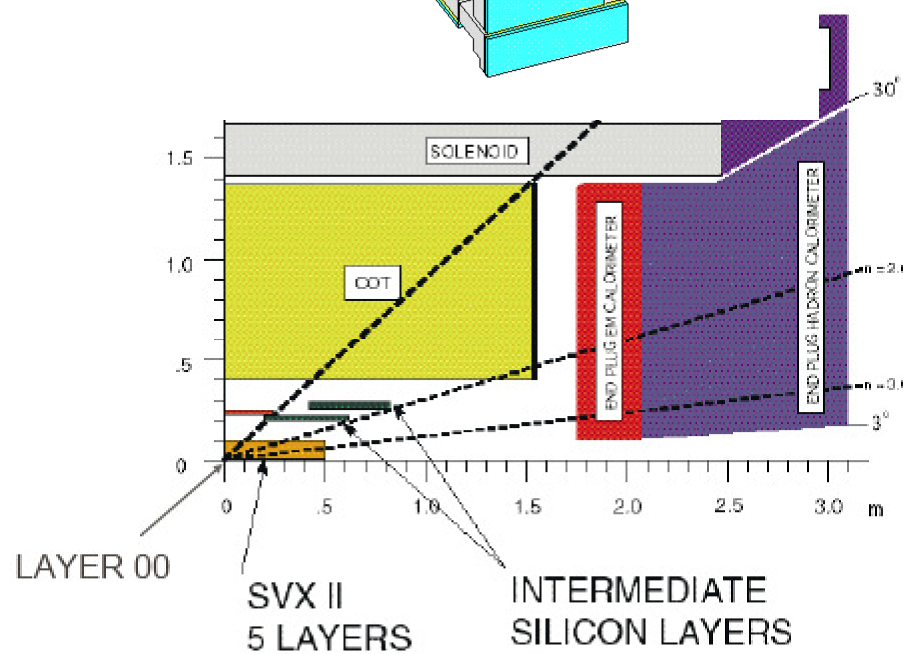
- Given that we have some nice handles – ***b*-tagging sounds easy, right?**

- **Challenges** at a **hadron machine**:
  - **Busy environment** in tracking detectors
  - **Multiple interactions** within each crossing
  - **No $Z \rightarrow bb$ peak** with which to calibrate tagging algorithms
  - Calibration samples are available – but **incomplete overlap** with interesting signal spectra
- **Challenges** distinguishing bottom jets **from charm jets**:
  - Charm has **nonzero lifetime**
  - **Intermediate mass** of charmed hadrons
  - **Similar semileptonic decay** spectrum to *B* sector

The CDF Detector: Crucial Components for Tagging

- **Charged particle tracking:**
  - **Solenoid provides a 1.4T magnetic field**
    - Good momentum resolution
  - **Silicon:** several subsystems
    - **SVXII:**
      - 5 layers out to radius of 10.6cm
      - |z| < 45cm
    - **L00:**
      - Directly on beampipe
      - Valuable for improved tracking – 4% increase in tag efficiency
    - **ISL:**
      - Two layers at r = 20,28cm
      - Provides forward silicon tracking
  - **COT:**
    - Open drift chamber
    - Good p_T, spatial resolution
- **Calorimetry –** jets, electrons
- **Muon system -** muons
- **Trigger**
  - Highly efficient for high pT leptons
  - Also collects valuable inclusive lepton, jet samples





CDF *b*-tagging Tools

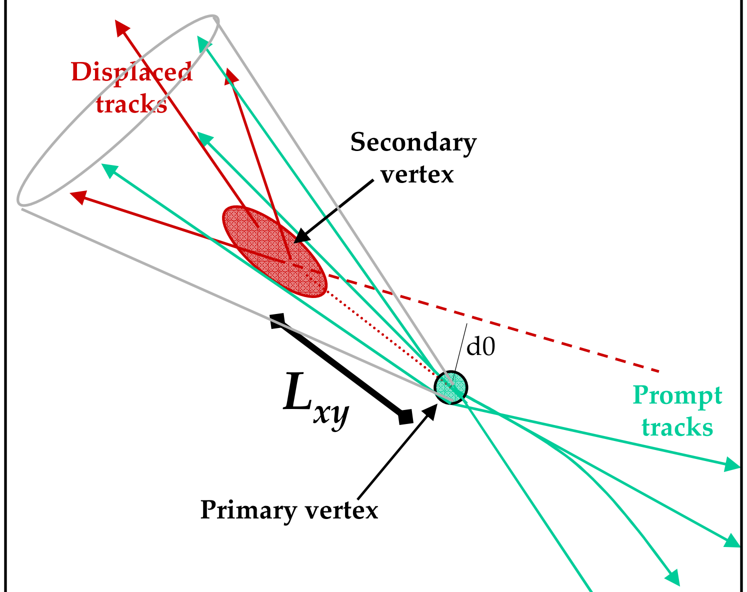
- **Ingredients for a useful tagger:**
  - **Tag efficiency for *b*-jets** in data, MC
  - **Mistag rate** in order to understand **contribution to tagged sample from non-*b* sources** – per-jet mistag probability
  - Efficiency and mistag probability are not single-valued
    - need to be examined as a function of jet- and event-level quantities

- CDF has **several tagging tools** in use/development for RunII analyses:
  - **Identification of jets with a secondary vertex – SECVTX:** ← Main focus of this talk
    - Exploits the long lifetime of the *b* quark
    - Additional handle one can use is the *mass of the reconstructed secondary vertex*
  - **Jet Probability:** incorporates lifetime, mass information
    - Assigns a per-jet probability that the jet was consistent with coming from a prompt source
  - **Soft lepton tagging:** looks for energetic electron or muon within a jet
  - **NN tagging algorithms:**
    - Simultaneous incorporation of lifetime, mass, semileptonic decay information along with event level quantities
    - Two versions under development
      - One that attempts to increase purity within SECVTX selected sample
      - Another that looks for tags in generic jet sample

Secondary vertex *b*-tagging at CDF

- **SECVTX algorithm:** attempt to **construct a secondary vertex** among **large impact parameter (d_0) tracks** using a two-pass scheme
  - **Pass1:**
    - Starts with construction of **2-track "seed" vertex**
    - **Attach** all remaining tracks that are consistent with seed.
    - Construct the multitrack vertex, iteratively **pruning** away the attached tracks if they spoil vertex fit.
    - **Resulting candidate vertex required to have 3 or more tracks**
  - **Pass2:** tighter track d_0 significance requirement
    - Attempt to **vertex all** these tracks to a common point.
    - **Remove** any track that spoils the vertex fit, re-vertexing after each removal.
    - **Resulting candidate vertex required to have 2 or more tracks**
  - Apply **vertex quality cuts**
    - removal of K_s,Λ **vertices**
    - Removal of vertices in the **material portion of CDF** (beampipe, silicon ladders)
  - **If the vertex survives, the jet is "tagged" –**
    - **sign of transverse displacement** of secondary vertex wrt interaction point, L_{xy}, determines **positive tag** or **negative tag**.

Drawing of the transverse plane of a single-top event – forward jet escapes down beampipe:



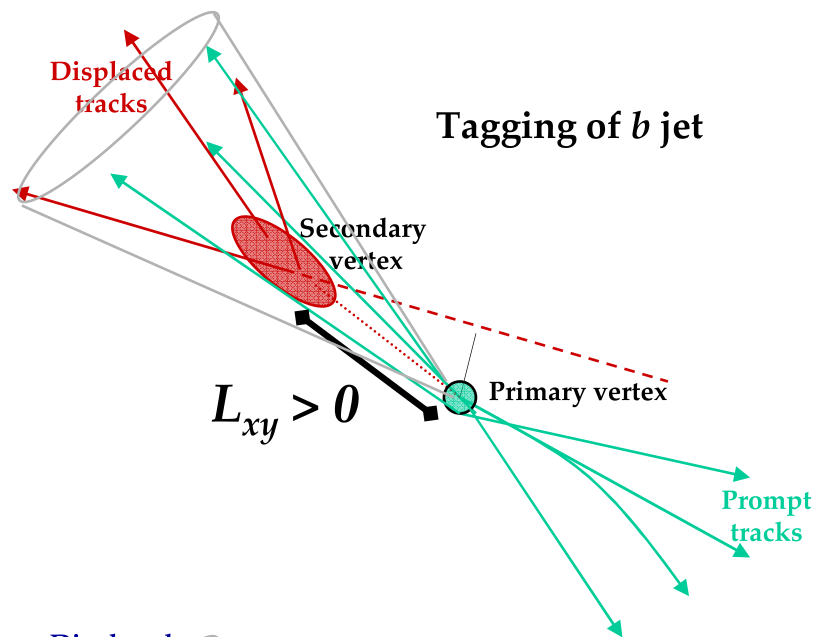
Here – **positive L_{xy} tag**.

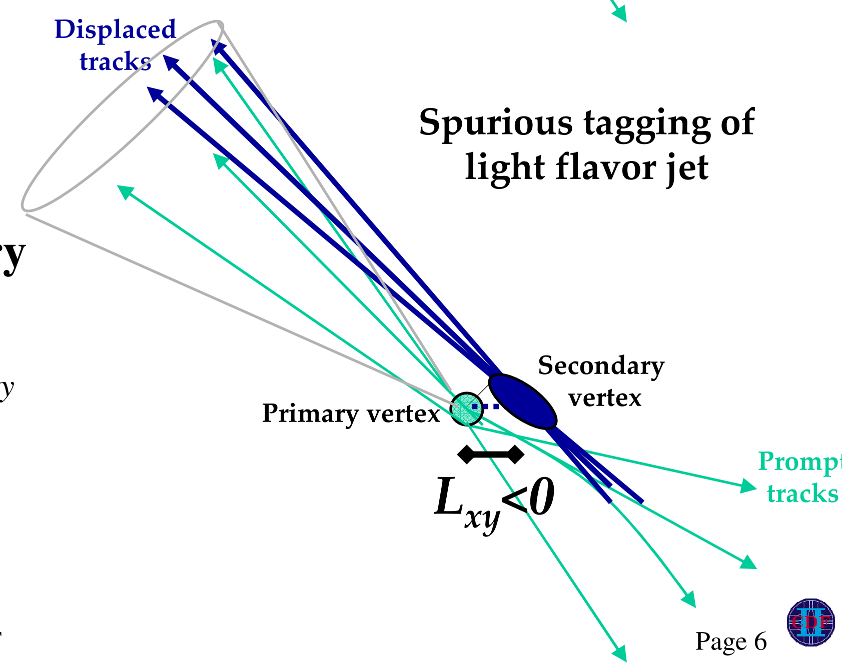
Contribution to *b*-Tag Sample from Light Flavor Jets

- The **flight direction a *B* hadron** travels in during its lifetime is **correlated to the jet direction**

- **Light flavor jets** should be consistent with **zero lifetime**
  - However **fake tracks within a jet** with large impact parameter can help satisfy vertex requirements
- **Sources** of fake tracks:
  - **Limited detector resolution**
  - **Long-lived light particle decays** (Λ, K_s)
  - **Material interactions**

- Fake tracks within a jet from limited detector resolution should be **symmetric about the primary interaction point**
  - Therefore light flavor vertices symmetric in L_{xy}
- This allows one to ***use the ensemble of negatively tagged jets*** as a ***prediction to the light flavor contribution to the positive tag rate*** **(aka mistags)**



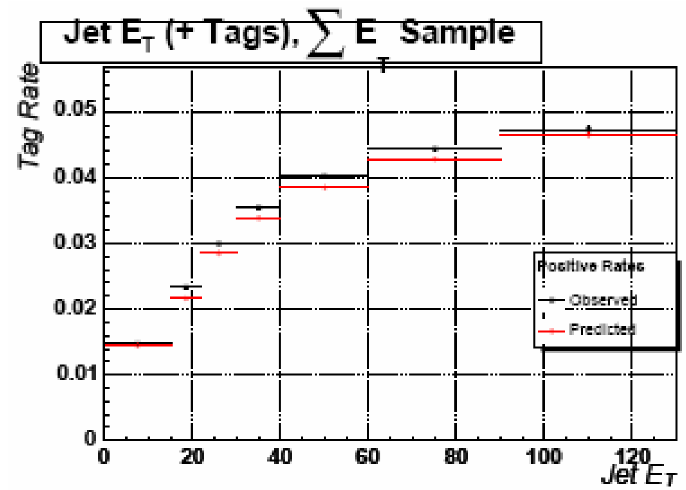


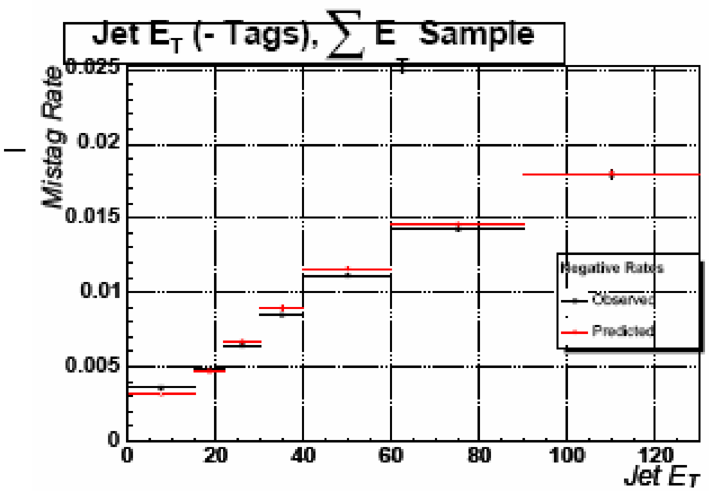
Contribution to *b*-Tag Sample from Light Flavor Jets

- However what is needed is an **a priori prediction of the light flavor content** of the positively tagged jets in the signal data sample

- **Procedure:**
  - For *b*-tagging based top physics analyses, the focus is the ***W*+jets data sample**
  - Use **inclusive jet sample for calibration** of mistags
  - Determine **per-jet mistag probability** in a number of different variables –
    - Jet E_T, $|\eta|$, φ
    - Jet track multiplicity
    - ΣE_T^{jets}
  - Use calibration jet samples to determine parameterization – then **apply to signal data sample**

- **Sources of systematic error:**
  - Extrapolation from calibration sample to signal sample
  - Uncertainty on ΣE_T^{jets}
  - Trigger bias
- Result: can predict **mistag contribution to 8%**





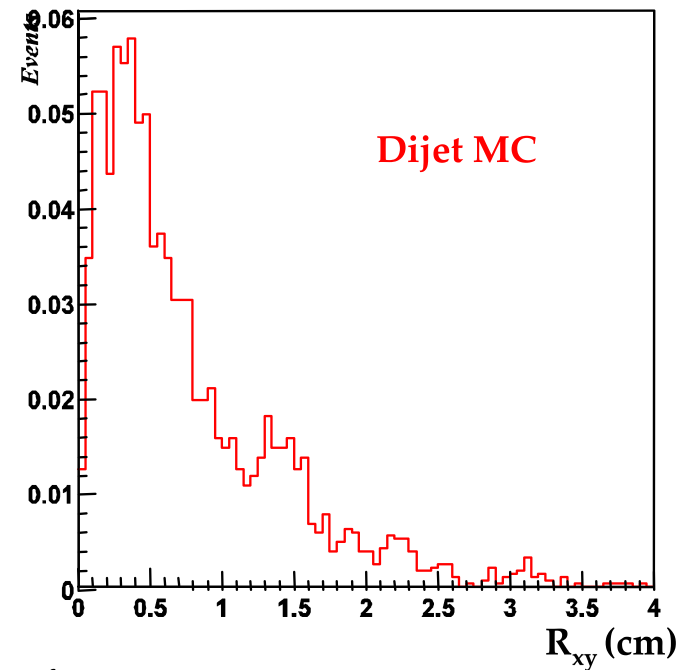
Light Flavor Jet Tag Asymmetry

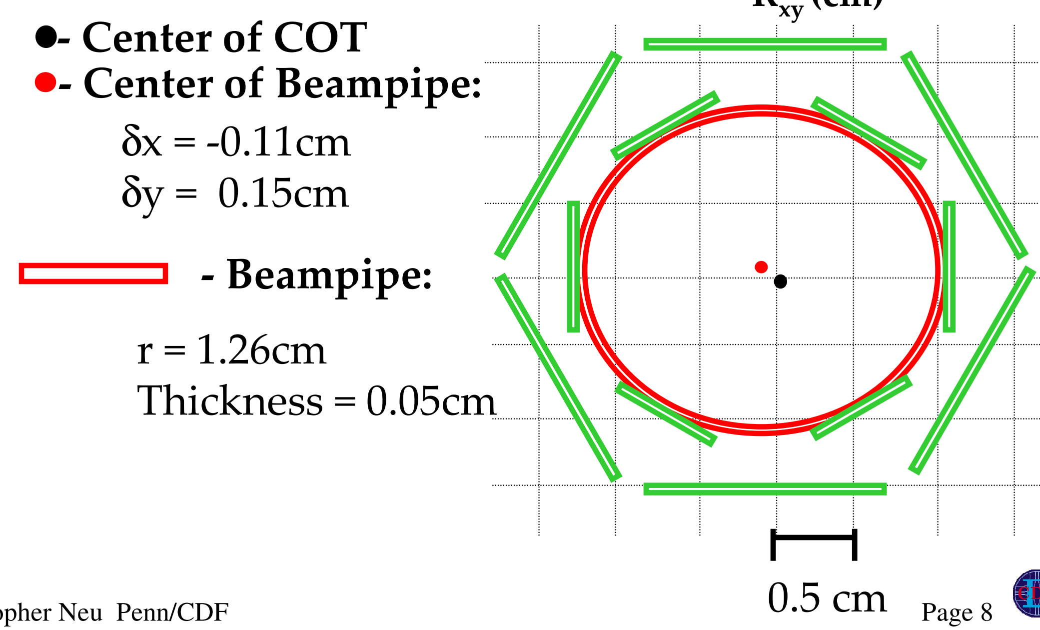
- The **mistag parameterization** only accounts for limited detector resolution source of the mistag sample

- Material interactions within the jet decay bias the distribution to positive L_{xy} values – introducing a **light flavor jet tag asymmetry**

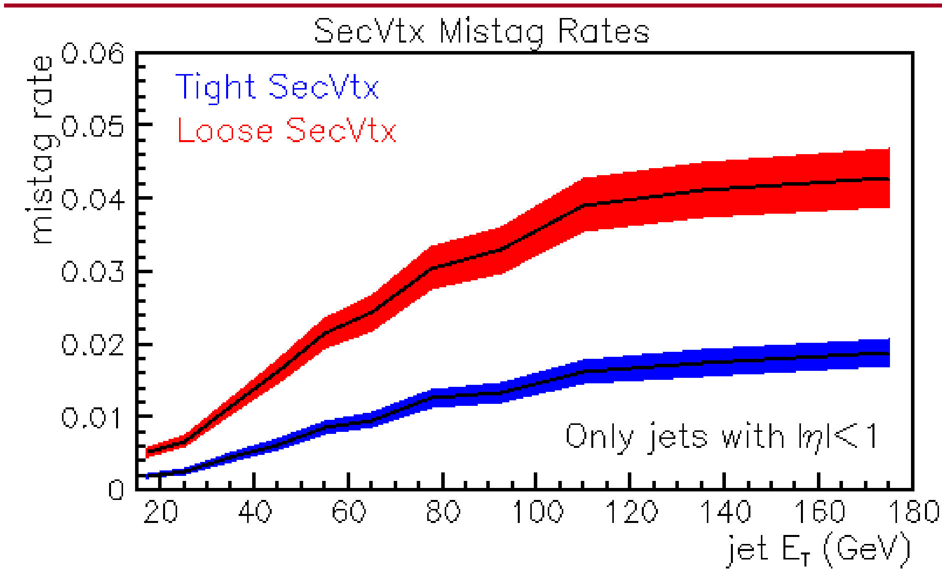
- Asymmetry can be measured
  - MC templates of pseudo-$c\tau$ for *b*, *c*, and light flavor jets
  - Fit to pseudo-$c\tau$ distribution from generic jet sample

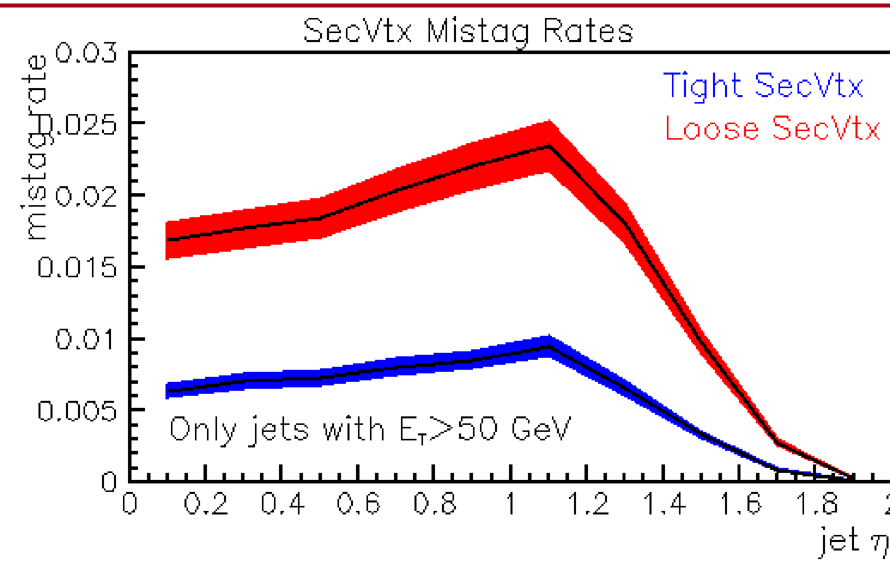
$N_{light}^{+} / N^{-} = 1.27 \pm 0.13$





Summary: Mistags





- **Mistag studies:**
  - Data from inclusive jet samples
  - Two SECVTX operating points – **Tight and Loose**
    - Different points in **efficiency-versus-purity space**
- **Loose operating point** is similar to proposed LHC taggers
  - Relaxed track requirements wrt Tight SECVTX – larger mistags
  - For a **central E_T = 40 GeV jet**, the SECVTX mistag rate is ~1 %

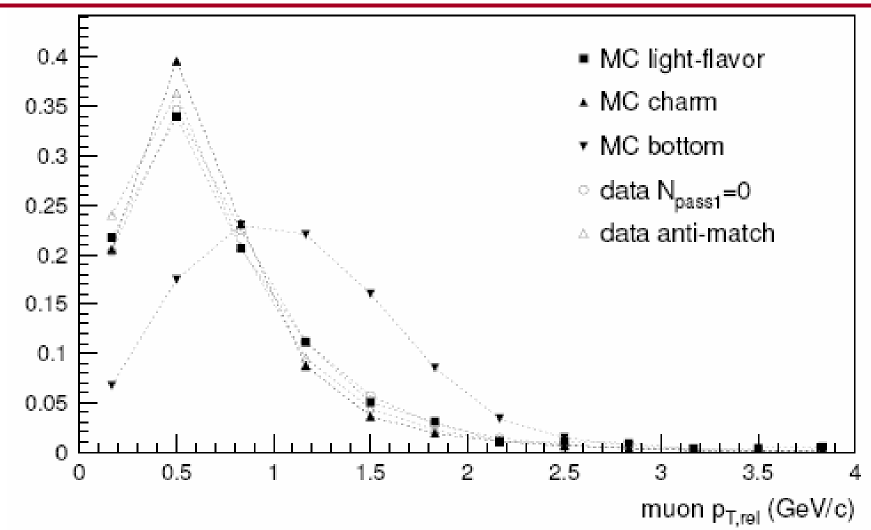
Efficiency Measurement in the Data

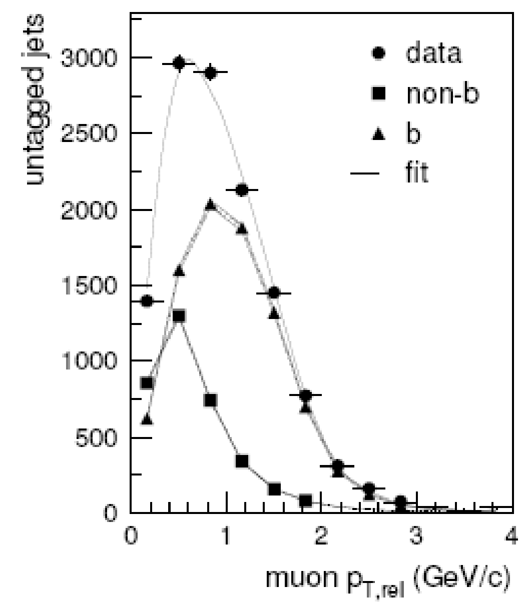
- **Understanding the tag efficiency in the Monte Carlo is simple**
- But what one really seeks is **the efficiency for tagging b-jets in the data**

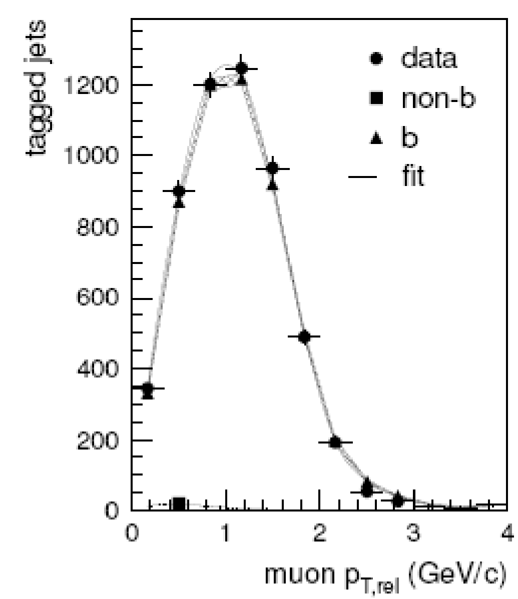
- **Strategy:**
  - Measure the tag efficiency in data in a sample that is enriched in real b-jets
  - Measure the tag efficiency in MC in a sample that models this HF-enriched data sample
  - Calculate a ***b*-tagging scale factor** = Ratio of data tag efficiency / MC tag efficiency
    - Scale factor is a measure of how the MC differs from reality

- **Two techniques currently employed at CDF:**
  - Both use samples of dijets
  - Enrich the HF content:
    - One jet demanded to have a lepton – so-called "lepton-jet" – indicative of semileptonic B decay
    - Other jet – recoil or "away-jet" – demanded to be tagged
  - One method relies on **"muon-jets"** and fits the b- and non-b content using **templates of the relative p_T of the muon wrt jet axis = p_T^{rel}**
  - One method considers **double tags** in events where the away jet is paired with an **"electron-jet" that is also tagged**

b-Tag Efficiency: Muon p_T^{rel} Method







- **p_T^{rel} templates drawn from MC**
  - Charm template very similar to that of light-flavor jets
  - *b* template similar for tagged and untagged *b*-jets
- **Used to fit for *b* and non-*b* content** in untagged and tagged data sample

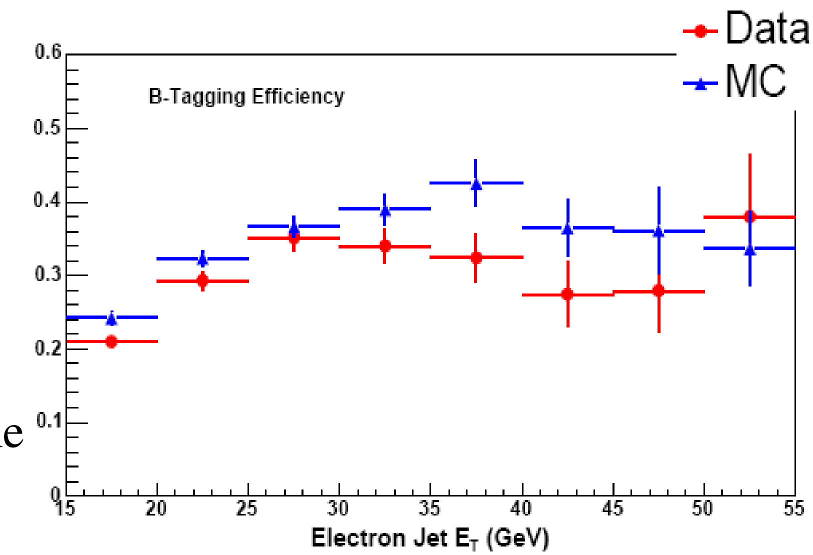
Statistical errors only

Pretag b-fraction	0.779 +- 0.009 +- 0.015
Tagged b-fraction	0.990 +- 0.016 +- 0.002
Data tag efficiency	0.392 +- 0.007 +- 0.008
MC tag efficiency	0.4278 +- 0.0019

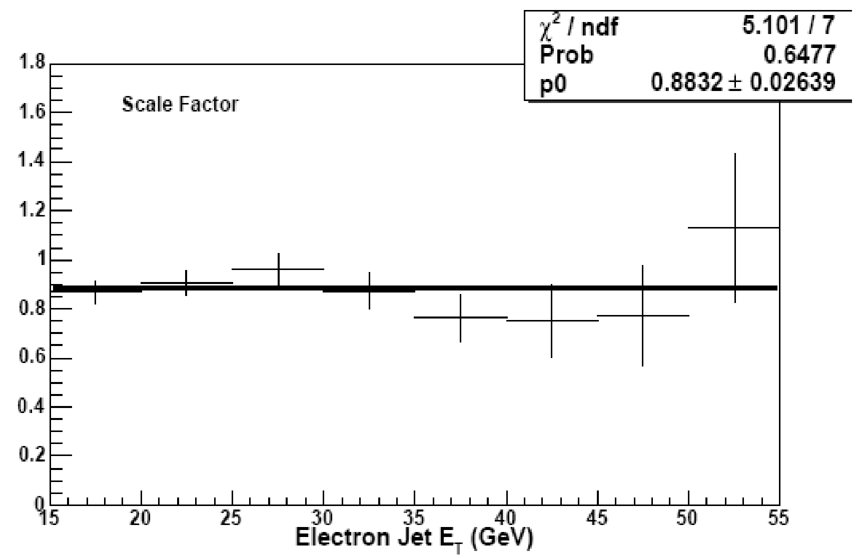
- Systematic errors: main source is **extrapolation to higher jet E_T**
- Result: **SF = 0.915 +- 0.017(stat) +- 0.060(sys)**

b-Tag Efficiency: Electron Method and Comparison

- HF-enriched electron-jet sample contains both semileptonic B decays and conversions
  - Use single tag rate in electron jet to **algebraically solve** for HF content of untagged sample
  - **Conversions** provide a complementary sample with similar topology with which one can understand the real HF content of the away-jet tagged sample



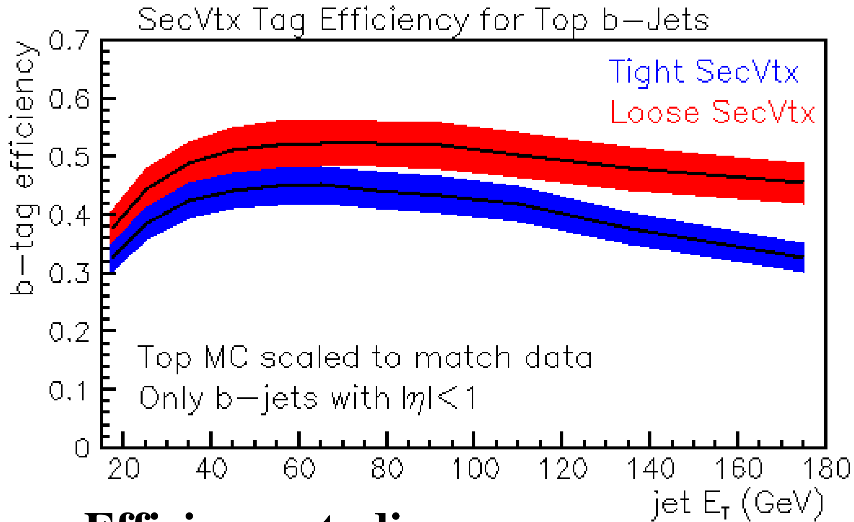
- Main sources of systematic error: **extrapolation** to higher jet E_T , ***b,c* fraction** in electron jets
- Result: **SF = 0.890 +- 0.028(stat) +- 0.072(sys)**

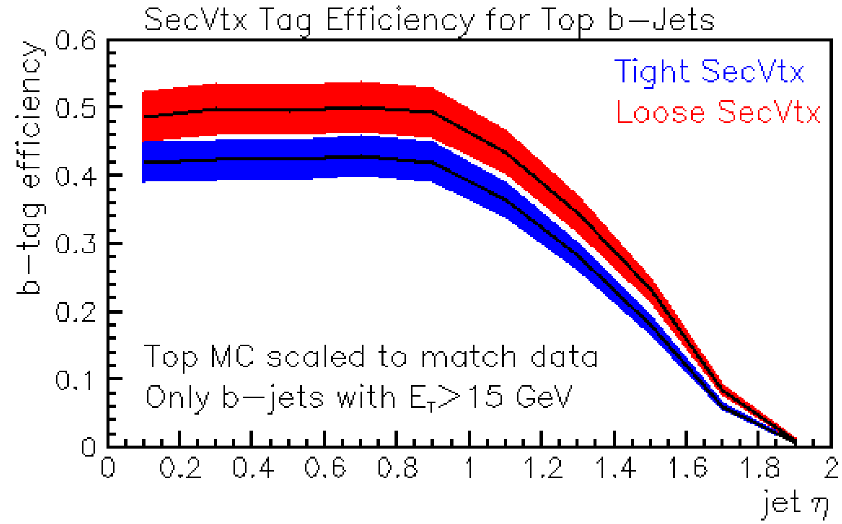


- Combination of electron and muon methods:

$SF_{combined}$ = 0.909 +- 0.060(stat+sys)

Summary: Efficiency



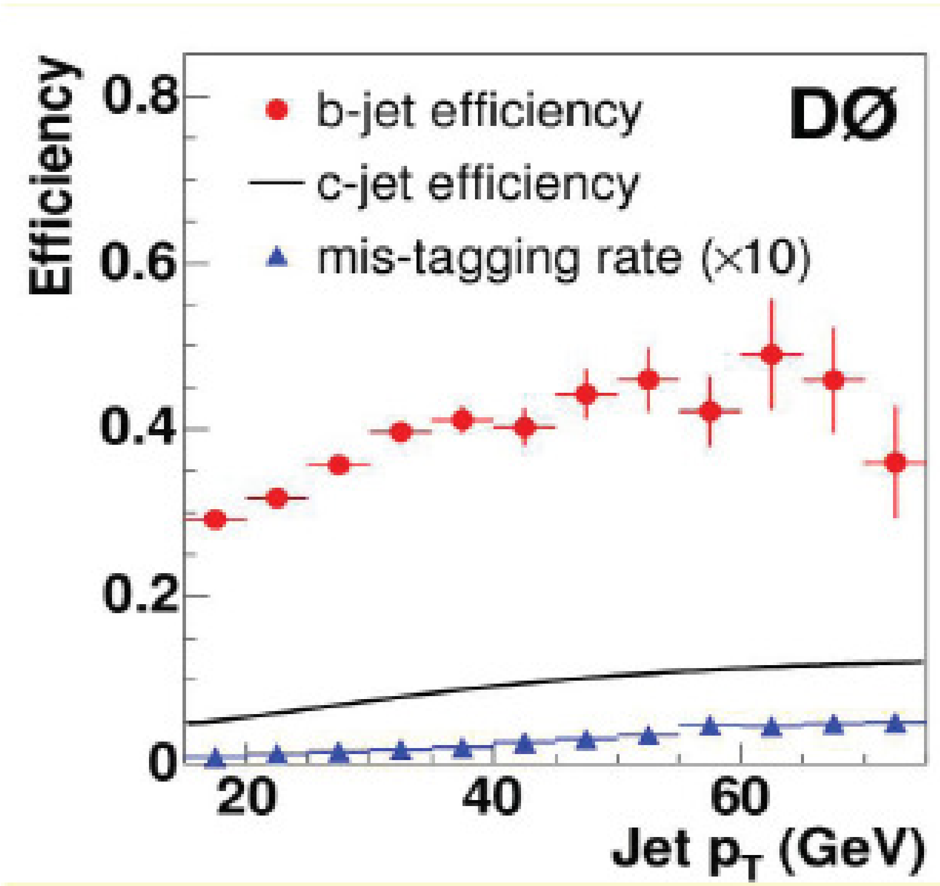


- **Efficiency studies:**
  - ttbar Pythia MC studies
  - *b*-tagging SF has been applied
- **Loose SECVTX operating point** – used in several top complete/ongoing top analyses
  - For a **central E_T = 60 GeV b-jet in top decay**, the Loose SECVTX tag efficiency is **~52%**
  - **Efficiency decrease at large |η|** is due mostly to tracking efficiency in the forward region – which are currently seeking to improve
- **Charm efficiency:**
  - Measured in MC, similar SF
  - Efficiency ranges from 5-10% as a function of jet E_T

b-Tagging at D0

- **D0 in RunII also has secondary vertex b-tagging in RunII**
- **Benchmarks:**
  - Efficiency for a 60 GeV b-jet is ~45%
  - Mistag rate for 40 GeV jet is ~0.3%

- This is best compared to the CDF **SECVTX Tight operating point**:
  - CDF Tight SECVTX efficiency for a 60 GeV b-jet is ~45%
  - CDF Tight SECVTX mistag rate for 40 GeV jet is ~0.4% for central jets –



CDF and D0 tagging algorithms have similar efficiency and mistag rates.

Looking Ahead to *b*-Tagging at LHC Experiments

- **Good amount of experience** has been gained at the Tevatron experiments
- Fairly successful *b*-tagging tools have been developed

- **This is not to mean however that all the problems are easy to solve**

- There are **many issues that deserve attention** for the future experiments:
  - **Alignment** of the silicon tracking detector
  - Understanding of the **charge deposition models** for particles as they traverse the silicon detector
  - Understanding the **material content around the interaction point**
  - **Tracking simulation** and its relation to reality
  - **Trigger effects** – ensure that enough calibration data is collected at appropriate ET, η range for the physics one wants to do

Summary

- Several critical portions of the Tevatron RunII physics program rely on the ability to identify jets originating from b quark production

- CDF has several b-tagging tools in use, including the secondary vertex tagger discussed here in particular

- With any *b*-tagging tool it is important to understand and quantify
  - Efficiency for tagging b-jets in the data
  - The rate at which non-b jets are tagged
- CDF has made progress in understanding these issues

- Tagger development for the LHC experiments can build upon the knowledge we have developed at the Tevatron

Backup

Backup – Muon Method Jet ET Dependence