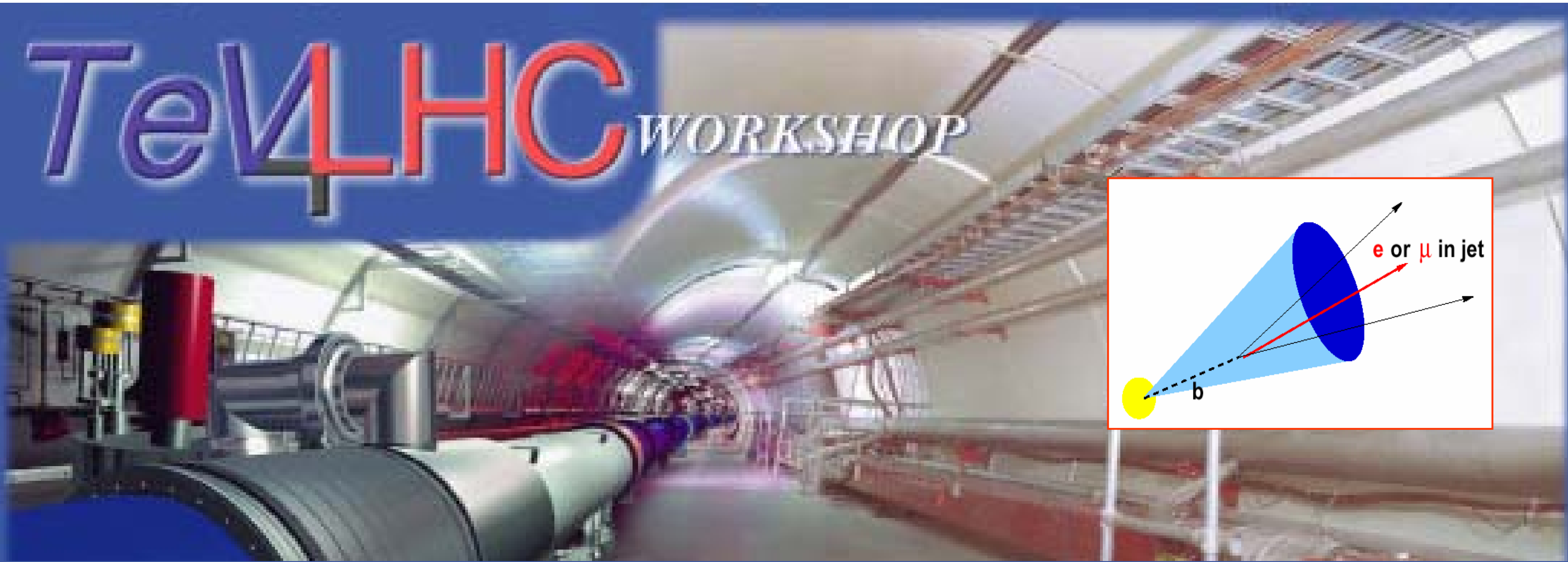


B Tagging at High p_T

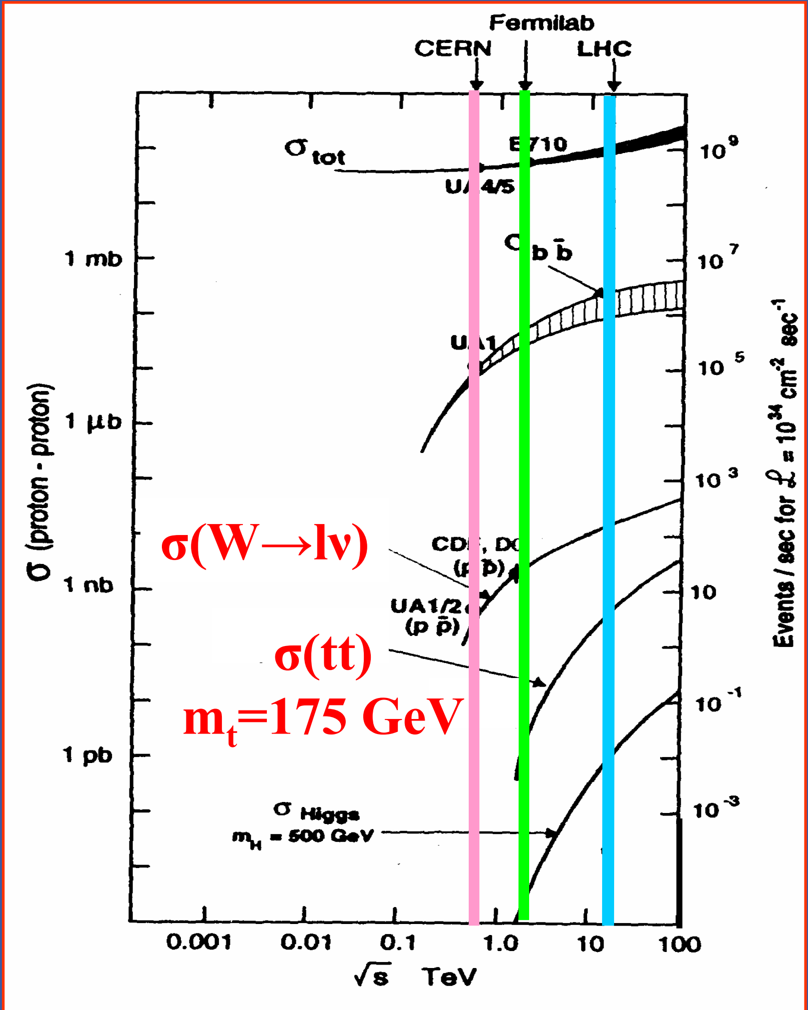
at the Tevatron

TeV4LHC
Feb 5, 2005
Plenary Sessions

Gordon Watts
University of Washington
Seattle

A DØ bias…

Why Tag?



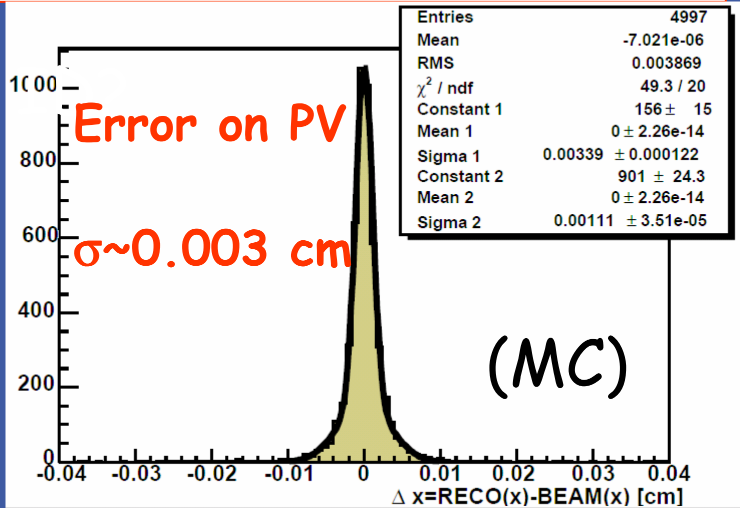
Much of Tevatron High p_T program involves signals with b-quarks

Top, Higgs

S:N can be 50:1 per-jet

TeV4LHC Why Talk About This?

Why is this different from electron



Algorithm Is Complex

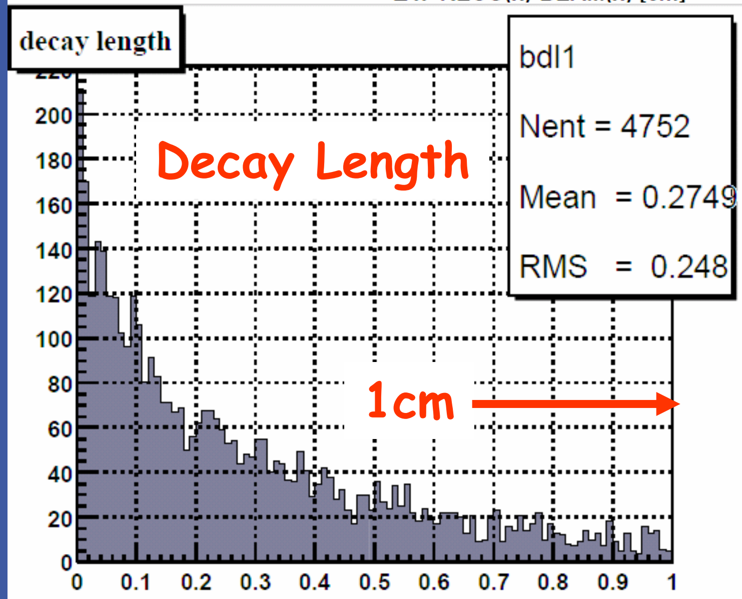
Built on Tracking and Vertexing (primary vertex)

Both must be well characterized!

S:B Determination is Hard

Don't have an equivalent clean sample of Z→ee.

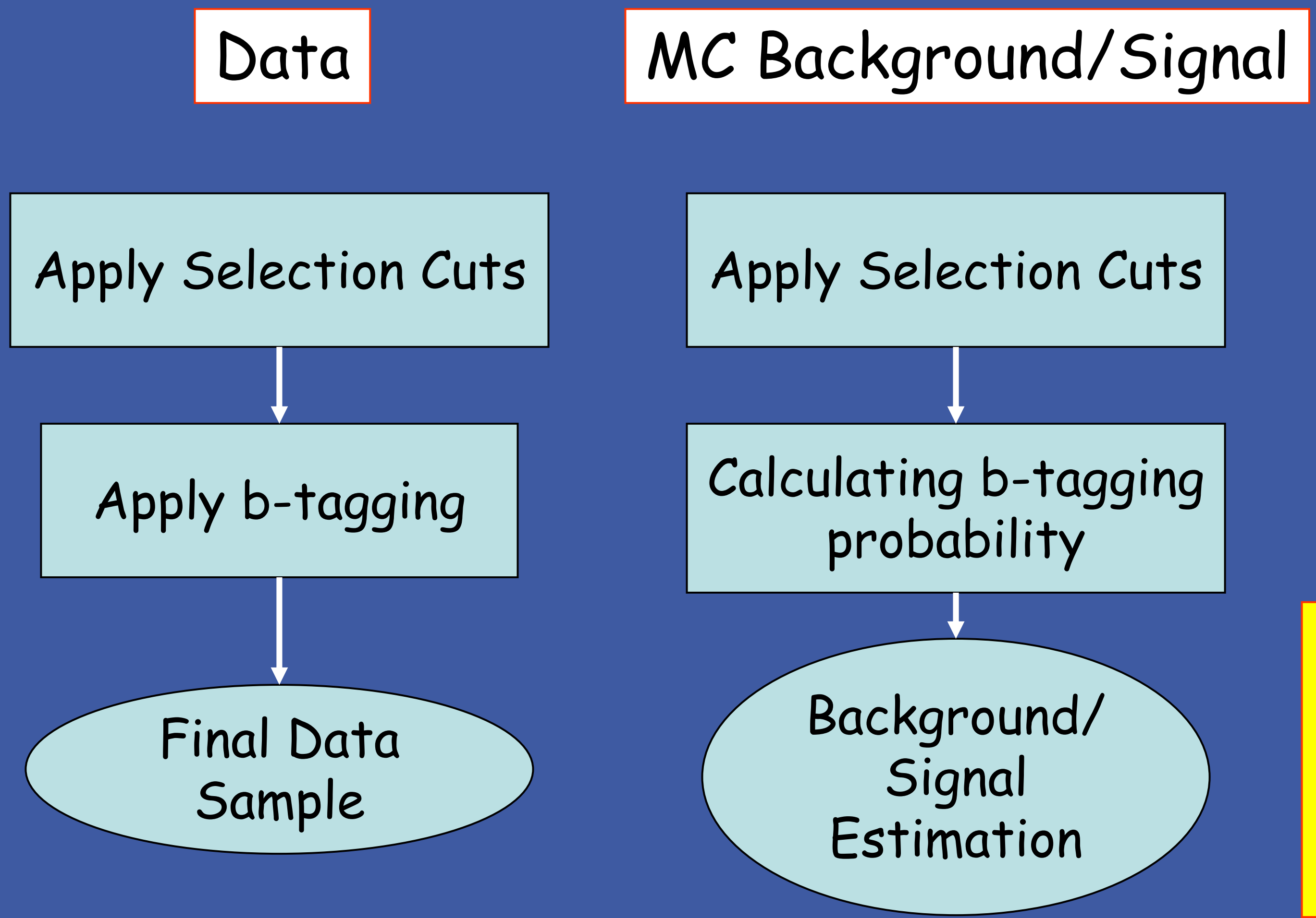
Have to guess b quark content!



It is Still New

Certainly for DØ, less so for CDF.

Tagging Analysis



Analysis applies b-tagging two ways…

Driven by measuring efficiency on data, not MC!!

TeV4LHC

Operating Points

Different Analyses are Optimal with different S:B Ratios

An analysis requiring 2 tags may be able to do two "loose" tags rather than two "tight"

Operating Points

Tight: 0.3% Fake Rate (ε ~ 45%)
Medium: 0.5% (ε ~ 50%)
Loose: 1.0% (ε ~ 60%)

Some analyses have requested even larger background rates

Many tagging algorithms provide a continuous variable. Operating points cut on these variables

TeV4LHC

What is Tagging?

Address http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/BTAG/btag.html

Google b tagging Search Web PageRank AutoFill Options

Type "b tagging" into Google…

Flavour-tagging Working Group

ATLAS b-tagging Group
First hit!

- Agenda of the next meeting
  - 24 February 2004 9:00 at CERN
- Previous meetings:
  - 9 December 2004, 23 September 2004, 27 May 2004, 4 March 2004, 4 Dec 2003, 25 Sep 2003, 6 Mar 2003, 21 Nov 2002, 13 Sep 2002 Before Sep 2002 .

Flavour tagging Performance sessions during Physics weeks and ATLAS weeks:

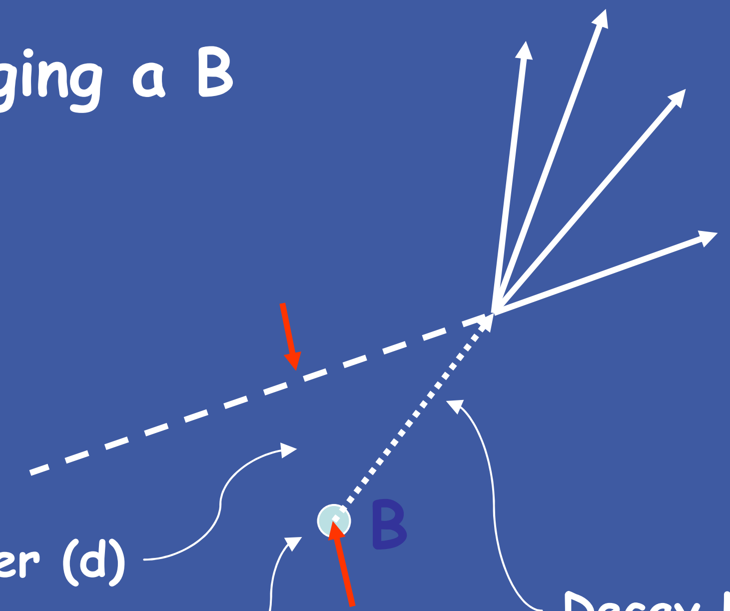
- 4 November 2004 Physics week at CERN
- 4 October 2004 ATLAS overview week at Fribourg

Flavour tagging workshop 13-14 December 2004 at Marseille :

- Agenda
- Draft of the list of actions of the workshop

Tagging a B

Vertex Tagging a B



- B Mesons
  - Has finite life time
  - Travels some distance from the vertex before decaying
    - ~ 1mm
    - With charm cascade decay, about 4.2 charged tracks

Several algorithms under active development

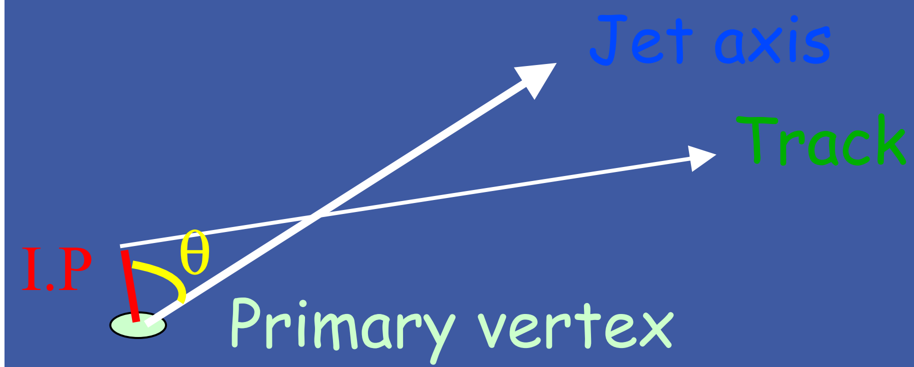
Impact Parameter Resolution	$d/\sigma(d)$
Decay Length Resolution	$L_{xy}/\sigma(L_{xy})$

CSIP Algorithm

Counting Signed Impact Parameter

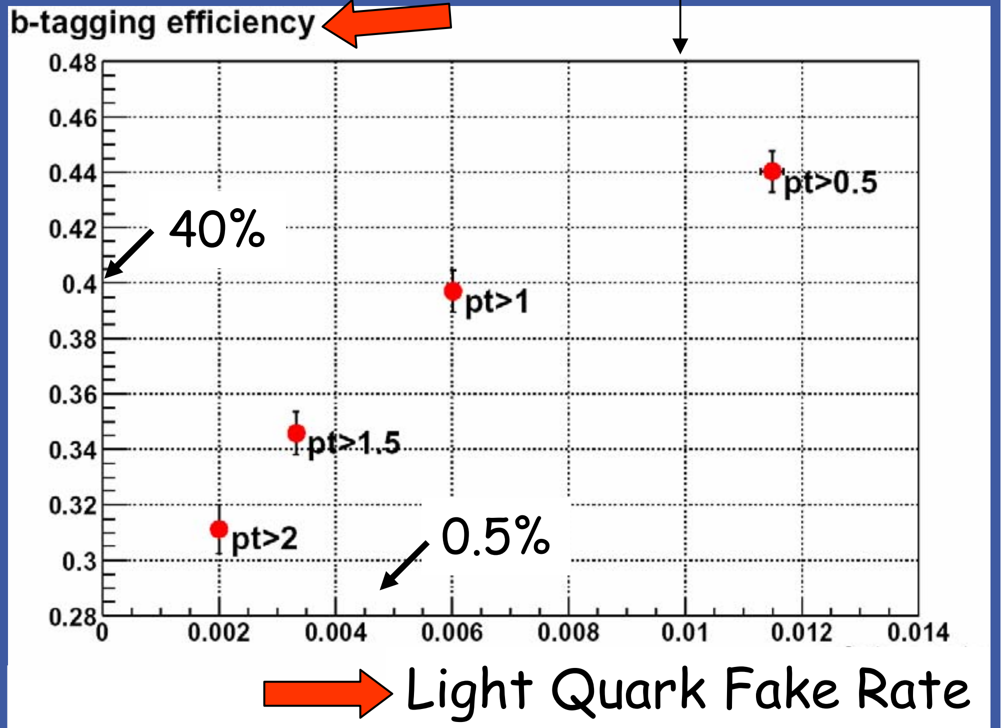
Based on Impact Parameter Significance

$S(IP) = IP/\sigma(IP)$



Requirements to tag a jet:
- at least 2 tracks with S(IP) > 3
- or at least 3 tracks with S(IP) > 2

Per Taggable Jet Rates Measured in Data!



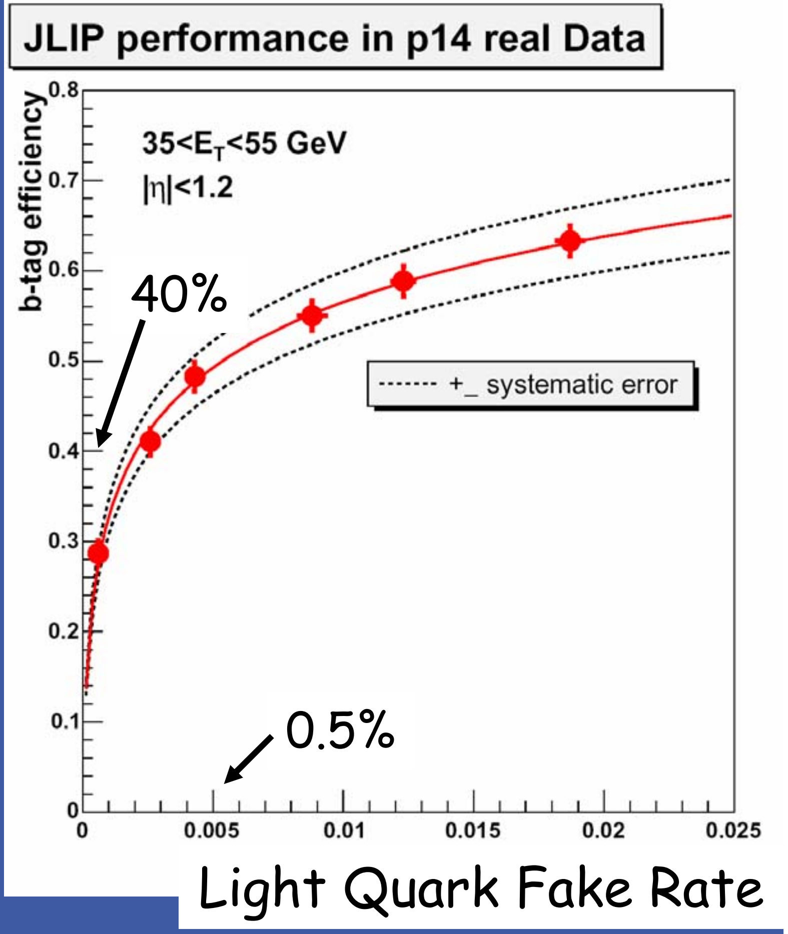
JLIP Algorithm

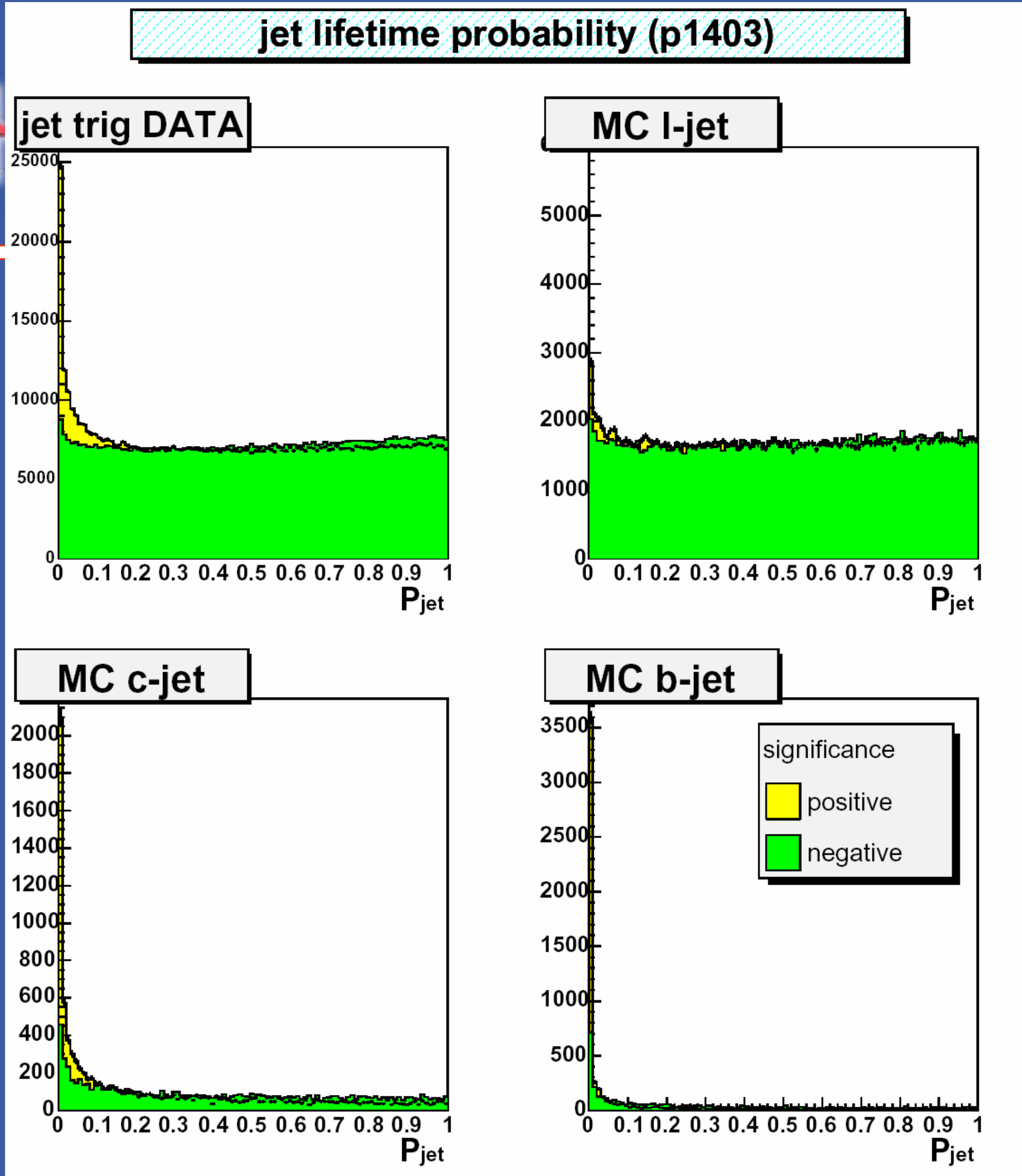
Jet Lifetime Impact Parameter

Based on Impact Parameter Significance

Probability distributions
- P(Track from PV)
- Defined for each class of tracks
  - # of SMT Hits, p_T, etc.

Each jet assigned P(light quark)



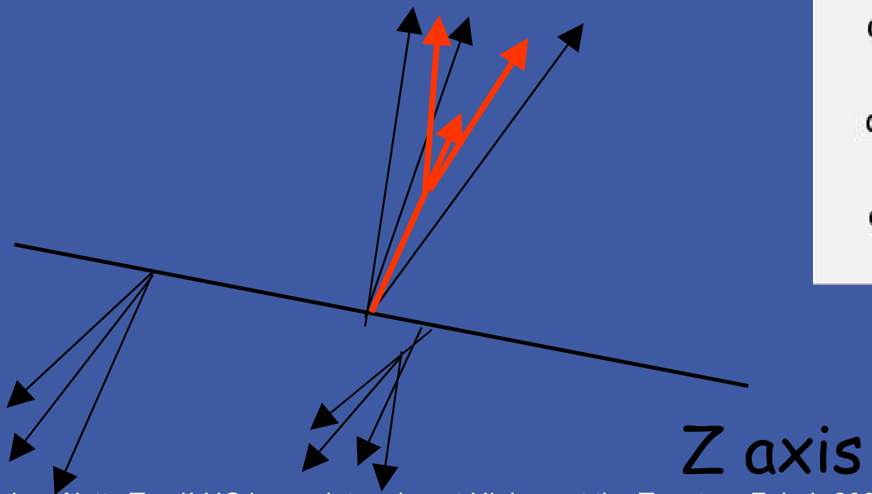

jet lifetime probability (p1403)
jet trig DATA
MC l-jet
MC c-jet
MC b-jet
significance
positive
negative
P_jet

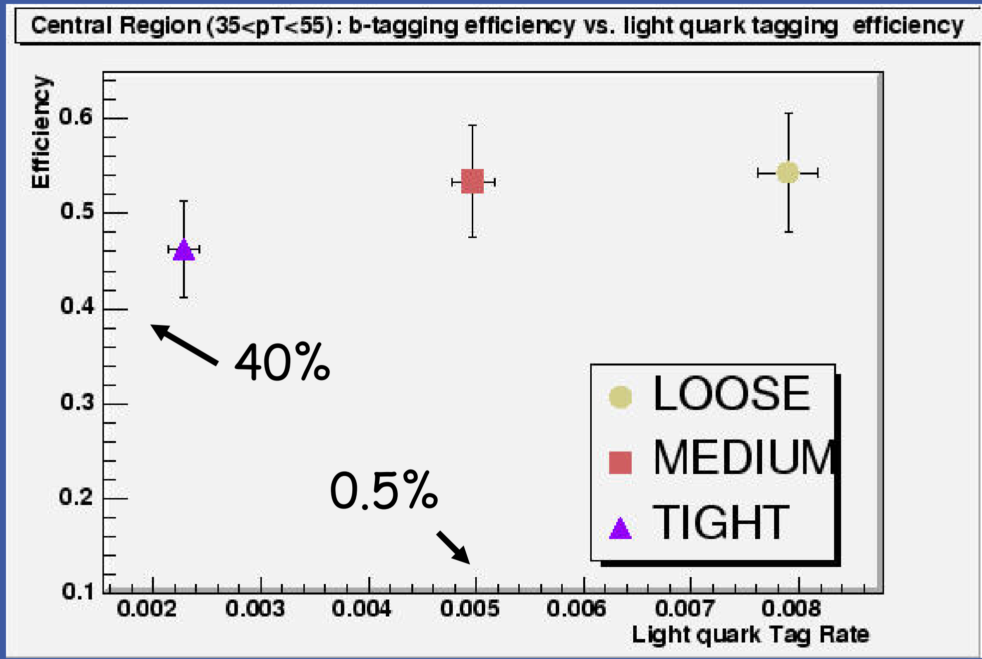
SVT Algorithm

Secondary Vertex Tagger

- Reconstruct Vertices using displaced tracks
- Cut on Decay Length Significance

$S(L_{xy}) = L_{xy}/\sigma(L_{xy})$.





TeV4LHC

Why 3?

Detector Startup

JLIP and SVT are strongly dependent on excellent tracking.
CSIP is designed to be simple to understand

Not 100% correlated

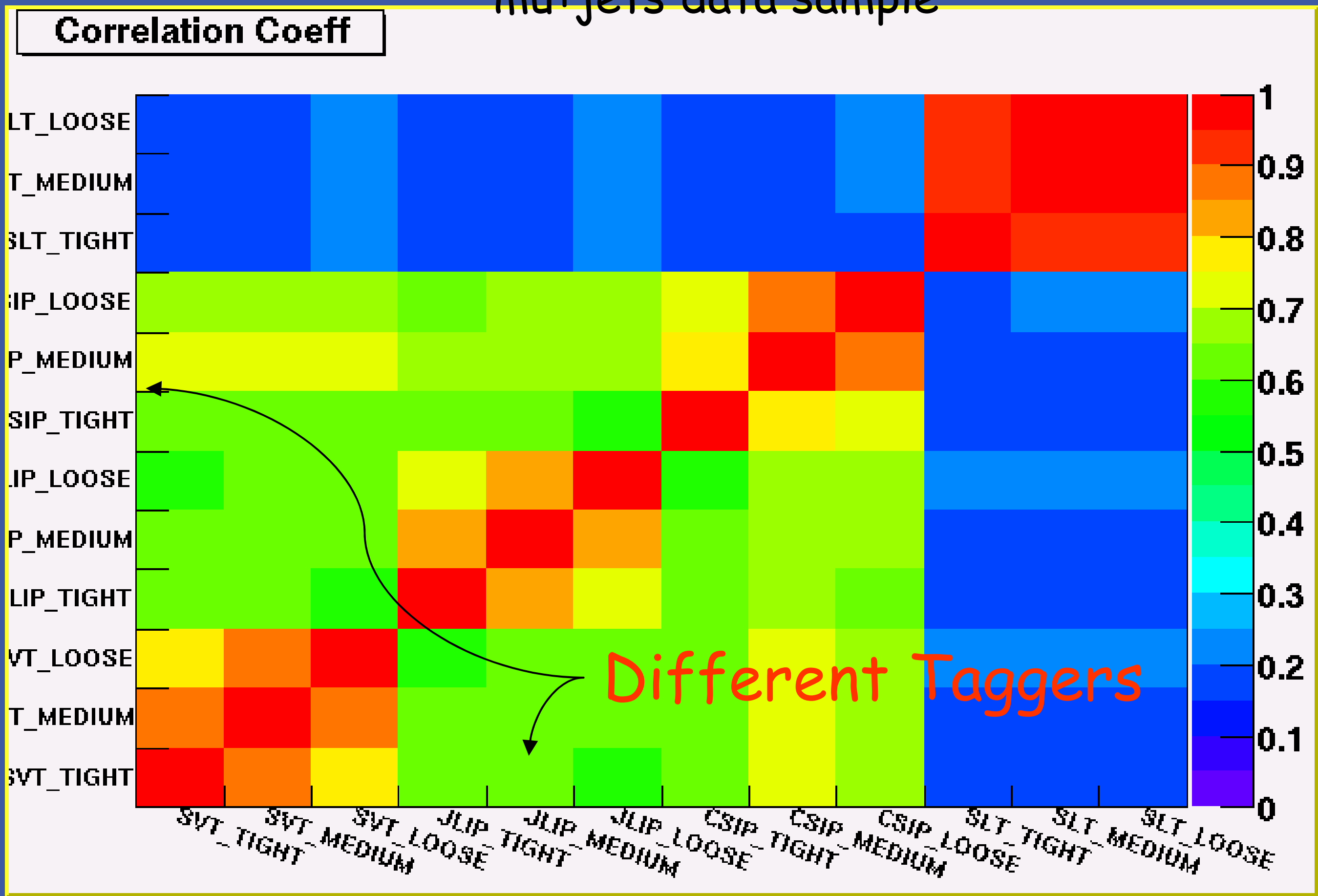
Possibility to combine and increase S:B separation.

Group Strength

Competition
But important they render results in similar format
Ian-style speech…

Correlation

mu+jets data sample



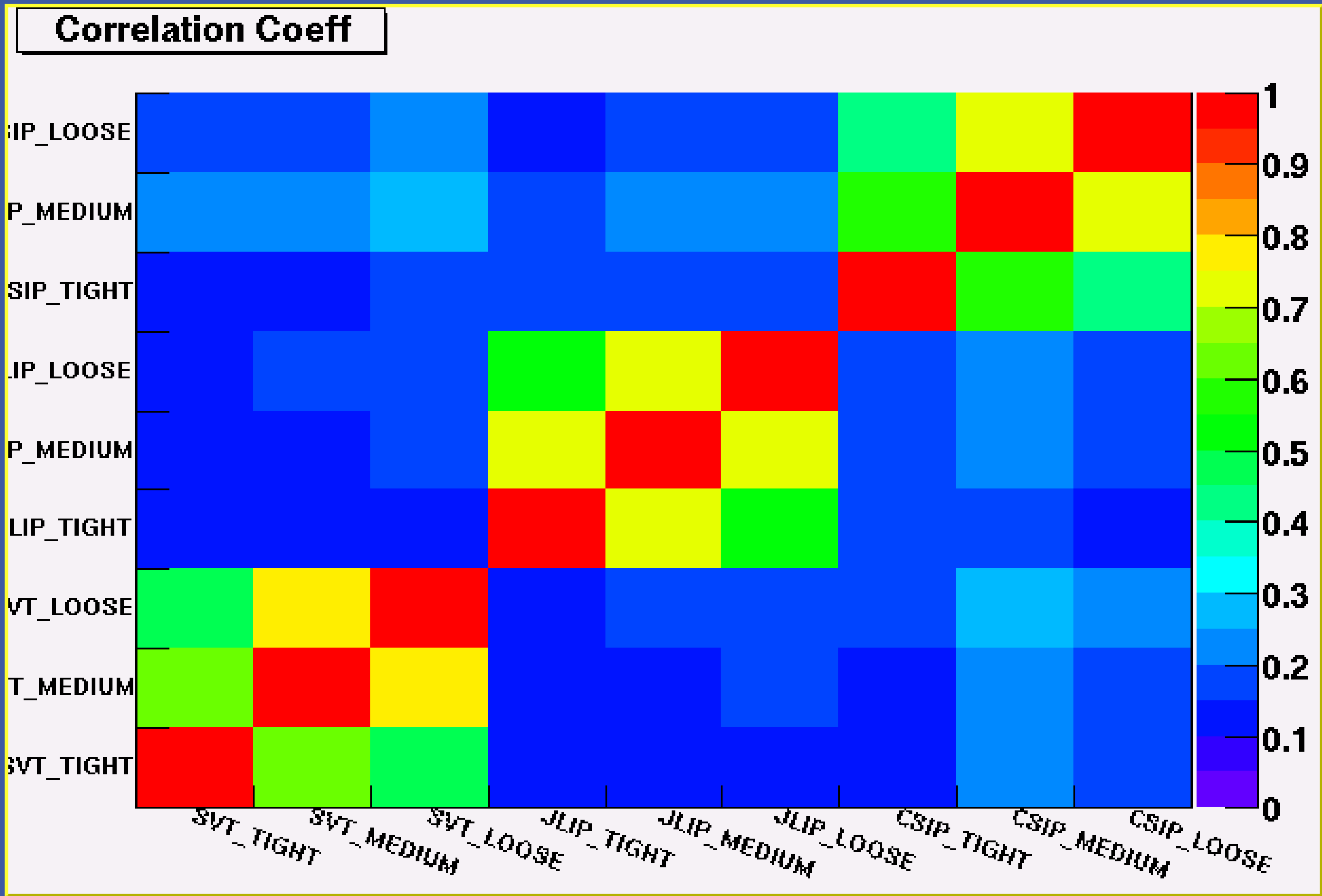
70% Correlated (signal)

30% Correlated (fake)

Uncorrelated on background, correlated on signal

Working on combining taggers

Correlation



Background

Estimating The Signal/Background

Measure b-tagging efficiency on data, but wish to apply in MC or other non-b-quark data sample.

Method 1 Determine efficiency vs. jet p_T, η, etc., on data, and use as lookup table in MC. (top)

Method 2 Determine MC-to-data tagging ratio vs. jet p_T, η, etc., on data, and use as lookup table applied to tags found in MC. (Wbb)

⇨ Both require determining, in Data, the eff of a standard b-jet and matching it to MC jets

Method 3 Use data sample with same flavor content of sample you are interested in, and derive tagging function (p_T, η, etc.) and apply directly. (hbb)

Not Easy

Populating these tagging functions in p_T, η, etc. takes millions of events of a known b-content!

Systematic Errors: b content, MC sizes, JES (MC/Data)

TeV4LHC Object Certification

(e, μ, jet, etc.)

- Attempt to avoid different physics cuts coming up with a slightly different "electron"
- Reduce duplication of effort
- Make standard objects to speed use in new analyses.

Certified objects are reviewed by a committee

- Selection Cuts
- ε on standard samples
- Background Rates
- Documentation

TeV4LHC bID Certification

The algorithms, operating points, *tag-rate-functions*

Determining these things takes about
2 months with experienced people

B-tag based analyses need this before they can proceed with background estimation

⇨ Critical Path! (conference deadlines, etc.)

Once certified, other analyses can quickly include b-tagging

Win: Centralize processing of those millions of events, determining fake rate and efficiency.

b-ID Certification

b-ID p14 Pass 2 Certification Guidelines

The BID Group

This list is currently a copy of the guidelines that were used for p14. Everyone should review these guidelines with an eye to modification for the next round of certification.

1. Provide operating points – efficiency vs light quark mistag rates at or near 0.3%, 0.5%, and 1% mistag rates.
   - These are the integral rates -- over all pT and eta.
   - Mistag means taking into account the MC scale factors.
2. Provide code – including documentation and examples on how to use the tagger in a user code based on BOTH TMB and TMBTrees. This has to be released in CVS.
3. Performance on Data (mujets, background)
   - Efficiency – computed using various methods
     - Ptrel fits to single tagged muon jets
     - Double tagged muon jets with ptrel cut >1 geV
     - Double tagged muon jets using ptrel fits
     - System 8
   - Mistag rates on emqcd and jet data
   - Both efficiency and mistag rates have to parameterized as a function of pT and eta of the jet.
4. Performance on Monte Carlo
   - Z samples, ttbar, and QCD pt>40 Sample.
   - Parameterize efficiency as a function of pT and eta of the jet
     Here we have to address the question on stability of the assumption that the pT and eta are uncorrelated while making these parameterizations. This assumption starts to fail for tight cuts.

13 Steps…

b-ID Certification

Certification Frequency

- Tracking
- Jet Energy Scale
- Primary Vertex Finding
- b-tagging Algorithm itself

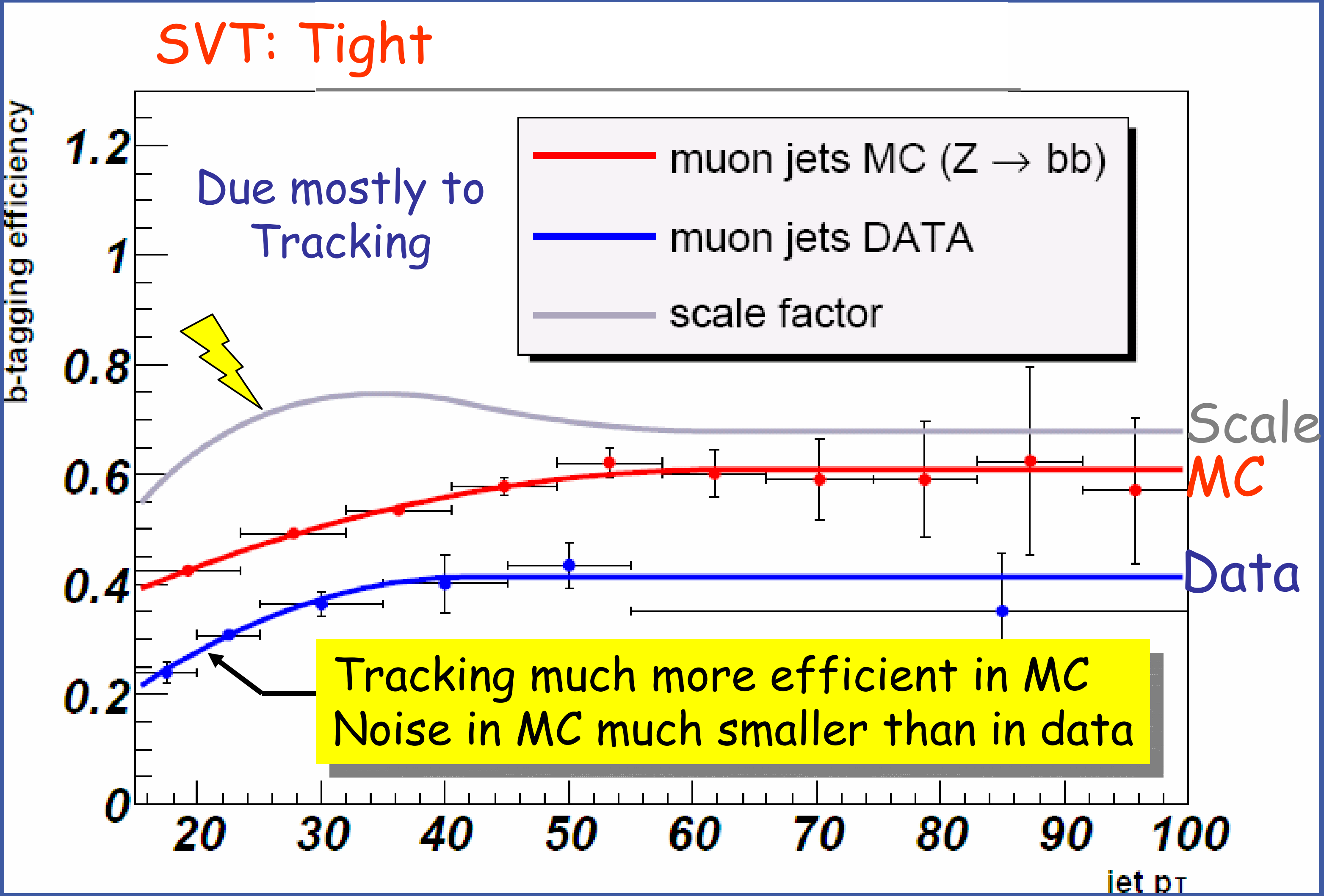
⇒ Force a re-certification

Practically: 1.5 times a year

Needs to be automated

N.B. One of the things that slows us down the most is waiting for good quality data samples

Data/MC Scale Factor



Charm Tagging Rate

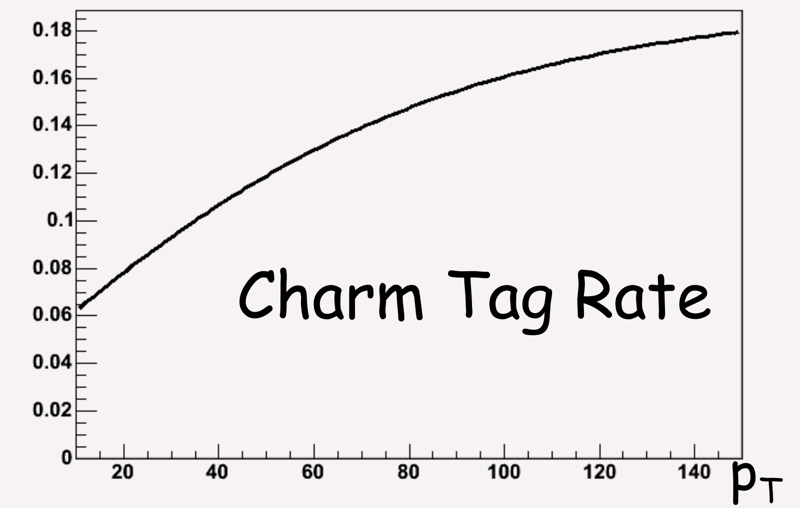
Use MC Ratio

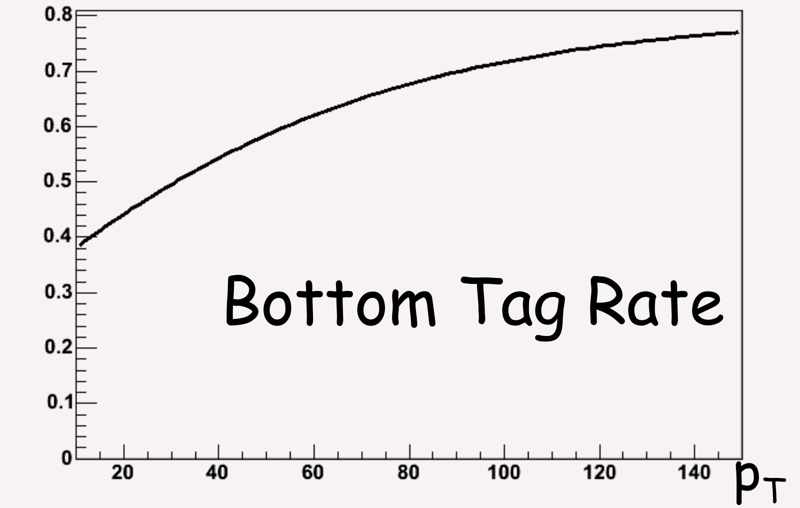
$$T_C = T_B^{Data} \frac{T_C^{MC}}{T_B^{MC}}$$

Systematic Errors caused are directly proportional to MC size and simulation!

Some physics channels can have distributions changed by charm tag rate.

Need to determine directly





Conclusions

- Backbone of may of the most important analyses at the Tevatron
- Determining the algorithm is 50% of the work the first time, 20% for subsequent certifications!
- Three tagging algorithms
  - Differences provide for competition and cross checks.
- Using b-tagging in an analysis is complex
  - Systematic Error treatment (not discussed)
  - b-tagging group supplies the TRFs and algorithms
  - Fast pickup by other analyzers
- Proper Efficiency Determination Difficult
  - Lack of clean bottom sample. Top @ LHC?