b-tagging @ CDF: experience, performance, lessons for LHC
hadron collider physics symposium 2005

Daniel Jeans, CNAF & Roma1
for the CDF collaboration

- introduction
- detector
- primary vertex
- b–tagging algorithms
- future directions
- conclusions

introduction

for the heaviest, and therefore least studied, SM particles (top & Higgs), theory favours decay to b quarks.

large t/H mass ⇒ decay products get large boost in lab
⇒ high energy hadronic jet containing B hadron

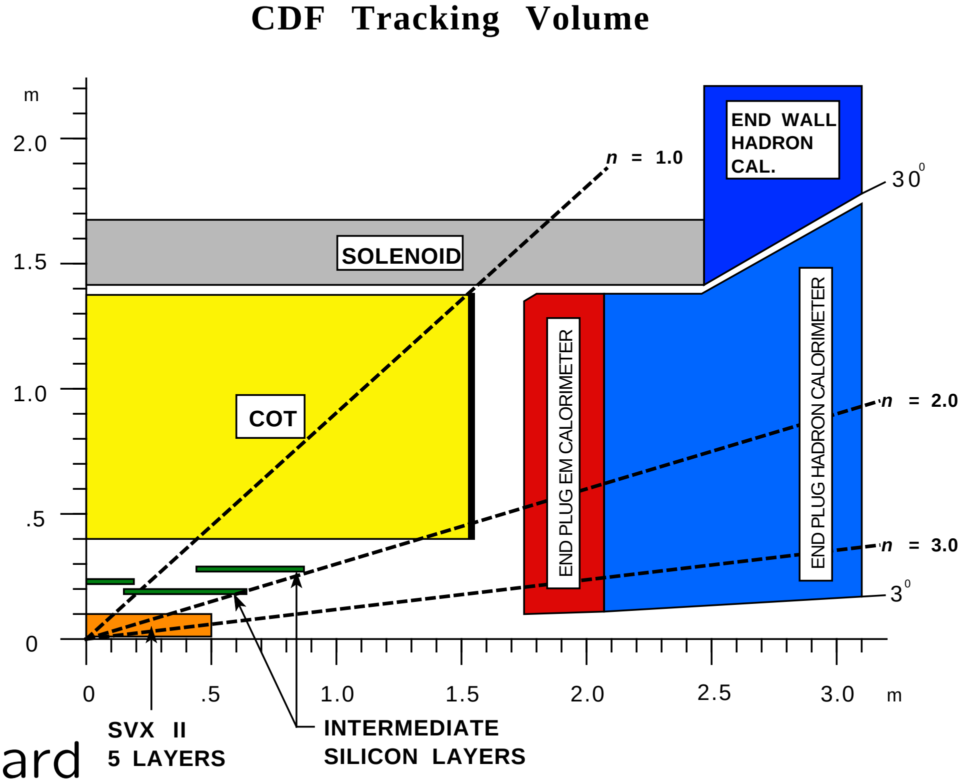
B hadron characteristics:
long mean lifetime ($\sim 1.5ps$)
large mass ($\sim 5.3GeV/c^2$)
large leptonic decay ratio (per lepton flavour:
$\sim$ 11% directly, $\sim$ 20% including daughter decays)

complications:
charmed hadrons also relatively massive & long lived
some strange hadrons have long lifetime: K_s, Λ

TeV

$p\bar{p}$ collisions, 1.96 TeV centre–of–mass energy
large luminous region: Gaussian with widths $\sim 30\mu m \times 30\mu m \times 29cm$
at $\mathcal{L} = 10^{32} cm^{-2} s^{-1}$, expect ~ 3 interactions per bunch crossing

CDF detector



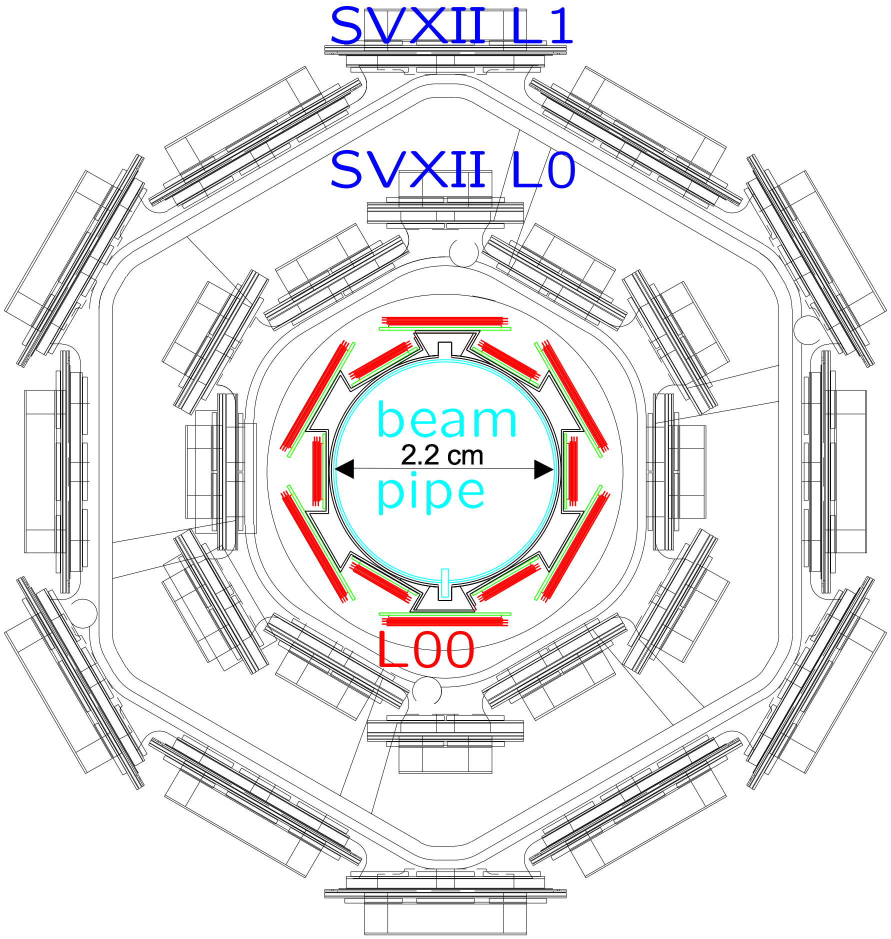
3–d tracking in 1.4 T magnetic field

COT: wire chamber: 8 super–layers
96 wire planes in total

silicon:
L00: radius of 1.2 cm
95 cm long
SVXII: 2.5 → 10.6 cm
3 "barrels" in z, each 29 cm long
ISL: 20 → 28 cm
1 layer in central region, 2 in forward

Lepton ID: muon chambers, EM calorimeter

silicon detectors



L00: radiation tolerant
single sided

SVXII: 5 double sided layers
L0: $R\phi$, Rz
L1: $R\phi$, Rz
L2: $R\phi$, small angle (1.2°)
L3: $R\phi$, Rz
L4: $R\phi$, small angle (1.2°)
3 barrels in z,
separated by "bulkheads" with
electronics, mechanical support...

ISL helps in forward region not covered by COT

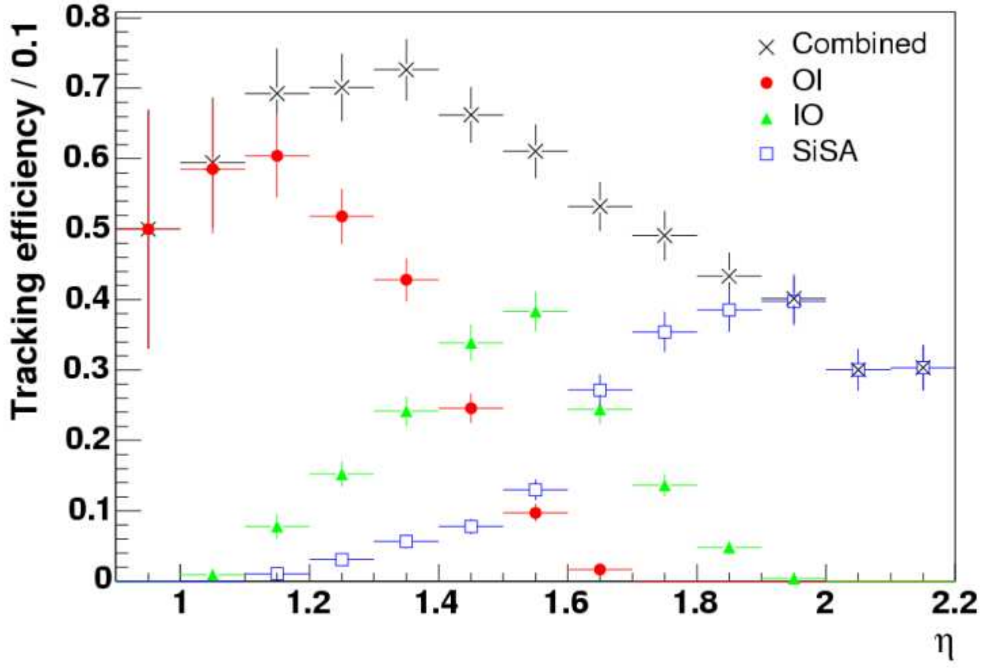
tracking detectors aligned *in situ* using tracks

tracking

tracking algorithms:

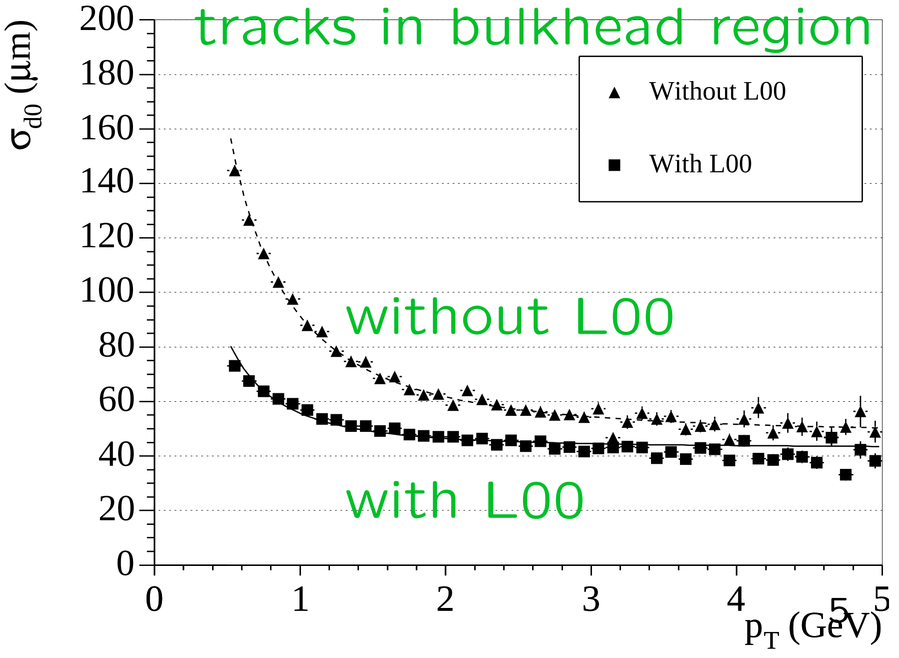
- find track segments in COT layers, combine layers to make tracks
- extrapolate COT tracks into silicon detectors, add silicon hits
- look for tracks using unused silicon hits
- extrapolate silicon tracks into COT, attach COT hits

forward tracking eff in $Z \to ee$ events



typical d_0 resolution

(includes width of beam)



tracks in bulkhead region

without L00

with L00

primary vertex measurement

since the luminous region is large, can get more precise determination of interaction point by reconstructing the primary interaction vertex

identify seed position in z: make histogram of track z_0, look for peak

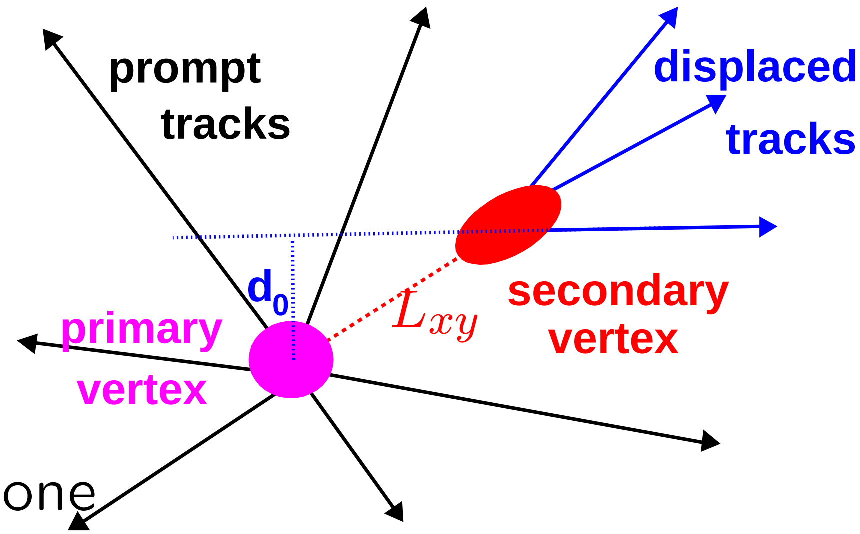
consider tracks in window around seed ($\delta z < 1cm$, $|d_0|/\sigma_{d0} < 3.0$)
(track z_0: z position where track is closest to beamline,
track d_0: 2–d impact parameter)
fit tracks to a vertex with beamline constraint,
exclude tracks which give large χ^2 contribution to fit

final resolution in $t\bar{t}$ events is $10 \rightarrow 32\mu m$ in (x, y),
depends on number of tracks used, z position (silicon bulkheads)

significantly better than using just the beam position
(beam width $= \sim 26 \rightarrow 32\mu m, z$ dependent)

tagging algorithms

SecVtx



identify vertices displaced from
the primary vertex (lifetime tagger)

consider COT+silicon tracks inside jet cone
remove tracks: identified as daughters of K_S, Λ, γ
consistent with primary vertex (in x, y)
too far from primary vertex ($|d_0| < 0.15$ cm, $\delta z < 2$ cm)
make 2 attempts to find vertex:

- vertex with at least 3 tracks
- harder track cuts, accept 2 track vertices

resolution on 2-d primary–secondary vertex separation typically 190 μm

require a vertex well separated from primary in 2–d, on correct side of PV, reasonable χ^2; veto 2 track vertices in the material regions

one version optimised for higher efficiency, another for better purity (different requirements on tracks, fit χ^2, primary–secondary separation)

JetProbability

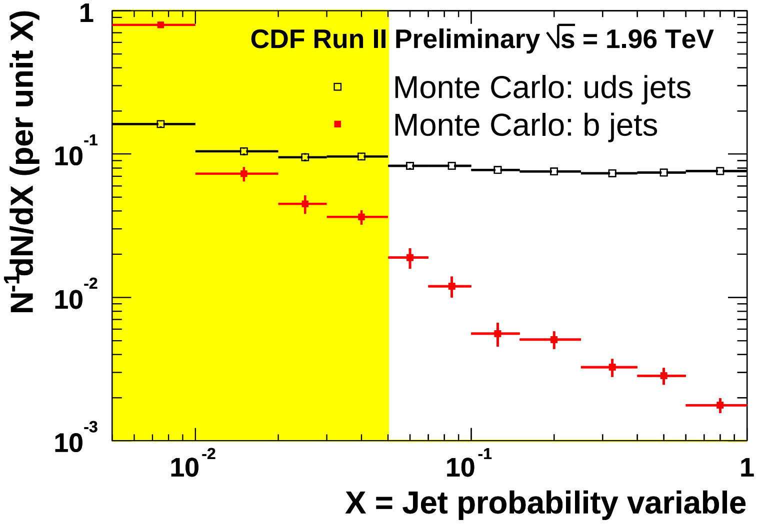
identify jets whose tracks are unlikely to have all come from primary

sign d_0 with respect to the jet direction:
Long lived particle decay daughters tend to have $d_0 > 0$
mis-measured tracks tend to have symmetric distribution

in jet data, parameterise negative side distribution of tracks' d_0 significance S_{d0} (split tracks into various "quality" classes)

then consider positive d_0 tracks in jet:
$\longrightarrow$ per track, calculate probability that
light flavour track has larger S_{d0}
combine probabilities for all tracks
inside jet $\longrightarrow$ per jet probability
by construction, flat for light,
peaked towards 0 for long–lifetime

"tag" jets with probability < 1 or 5%



data/MC efficiency "scale factor"

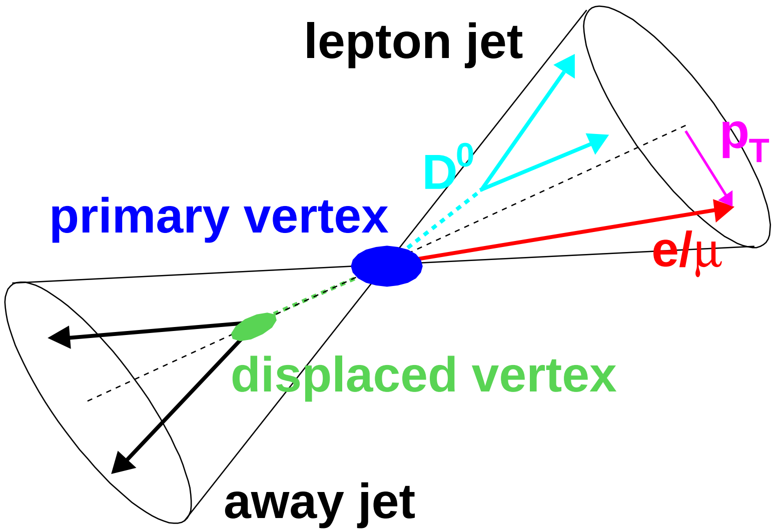
physics processes have different distributions of b jets in E_T, η
use MC simulation to account for these differences

however, simulation is never perfect:
tracking efficiency & resolution, B hadron decay models, ...

to estimate the effect of these imperfections, measure an efficiency scale factor in a large, independent dataset.

then apply correction to Monte Carlo simulation of all other processes

scale factor: efficiency measurement

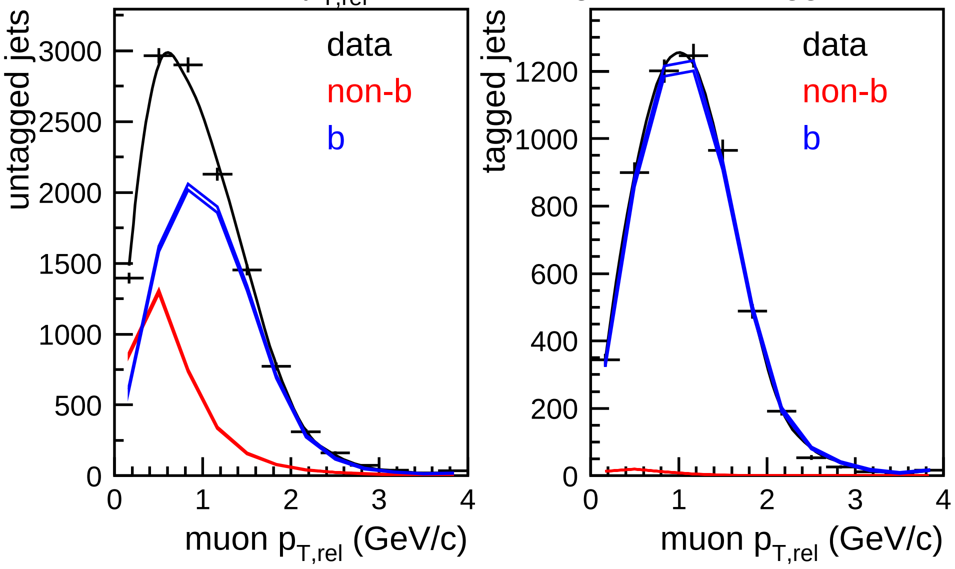


get sample of jets with large b content;
count how many are tagged by algorithm

events with jet containing an e/μ
estimate b fraction in jet:
  fit muon p_T with respect to the jet axis
  count $e + D^0$ or $e + \mu$ events
count how many of these b jets get tagged
rather sensitive to b fraction

$\Rightarrow$ increase b fraction in lepton jet
require e/μ jet $\sim$ back–to–back to
  another jet, require "away jet"
  to be tagged
less light flavour contamination,
less sensitive to b fraction

Muon $p_{T,rel}$ Fits for the Tight SecVtx Tagger

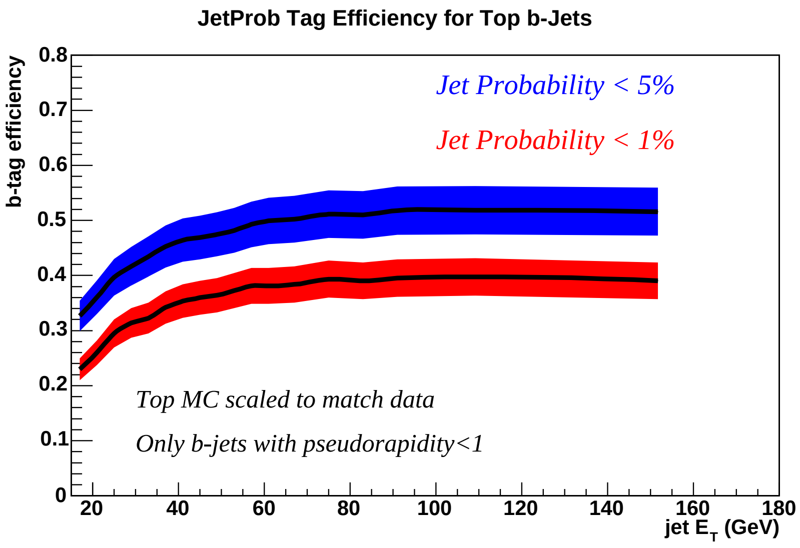


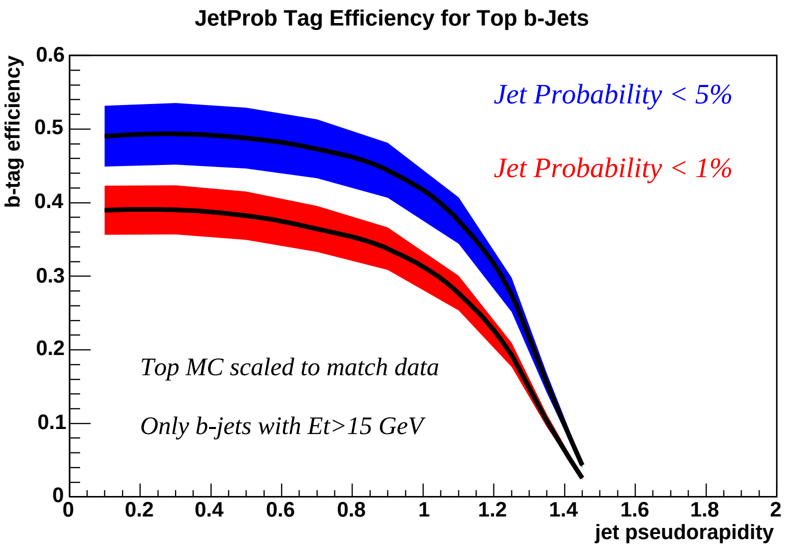
measure tagging efficiency in data and corresponding MC
then Scale Factor $= \epsilon_{\text{data}}/\epsilon_{\text{MC}}$

need to correct for some effects:
– leptonic b decays have lower multiplicity than generic decays
– take care about E_T dependence: not many high E_T jets in sample

scale factor $\sim (82 \rightarrow 93) \pm 6\%$, depending on tagger

JetProbability b jet tagging efficiency in MC $t\bar{t}$ events,
corrected by "scale factor"





mistagging probability

"negative tag":

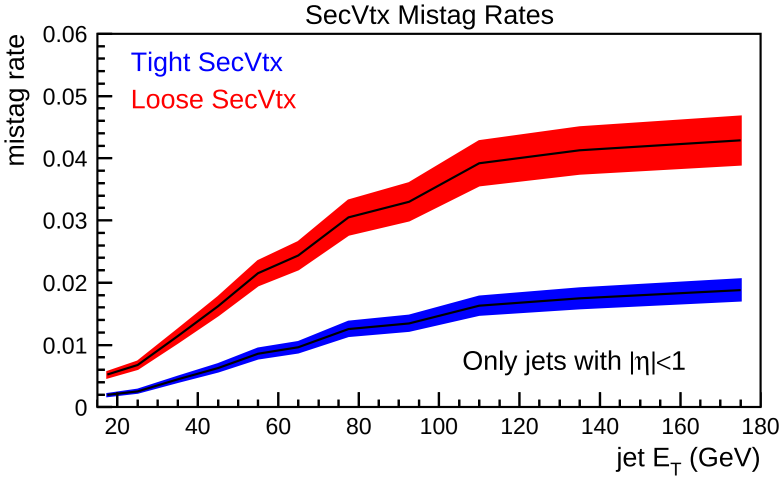
SecVtx: displaced vertex on wrong side of the primary vertex with respect to the jet direction

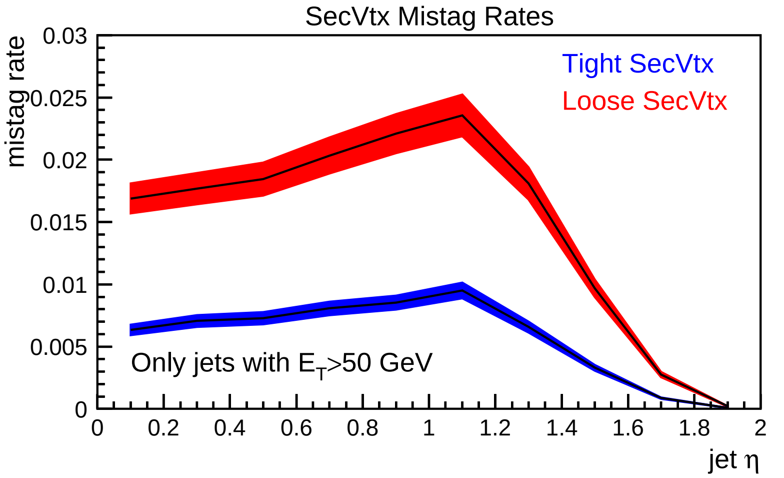
JetProbability: probability calulated using tracks with $d_0 < 0$

assume fake tags due to finite tracking resolution are positive/negative symmetrical, then estimate mistagging probability by using negative tag rate

parameterize the mistag rate as a function of jet E_T, ϕ, η, # tracks, event $\sum E_T$

SecVtx mistag rates

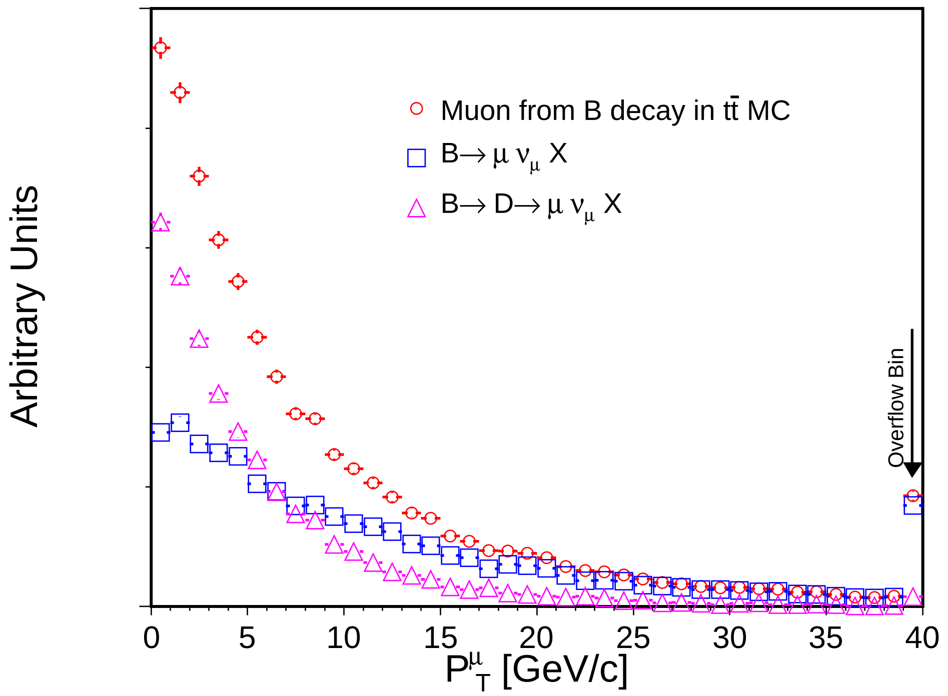




need to account for effect of additional positive contributions from material interactions, unidentified K_S, Λ (increases rate by around 25 %)

soft muon tagger

p_T distribution of μ in $t \to B \to \mu$



$\sim$ 20% of B hadrons decay to μ;
these are non-isolated,
relatively low p_T
→ can't use M.I.P. characteristics
in calorimeters
→ multiple scattering significant

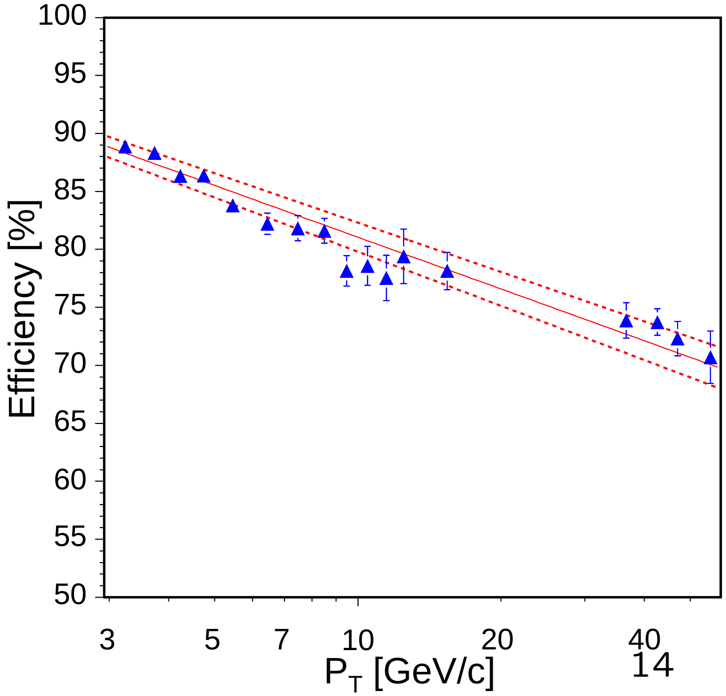
dedicated muon ID:
match tracks close to the jet to muon chamber track segments
remove tracks from J/ψ, Υ, Z decays

soft muon tagger performance

mistag rate from the probability to tagging tracks in generic jet data.
exclude J/ψ, Υ, Z $\Rightarrow$ number of true muons is expected to be small
dominated by fake μ: punch–through, decays in flight
parameterise fake rate as a function of track p_T, η and ϕ
typically 0.6 → 0.9%

μ identification efficiency
central muon chambers

efficiency measured by looking at
"second leg" of J/ψ, Z events
typically 90 → 70% for
muon $p_T = 3 \rightarrow 60\ GeV/c$
these muons tend to be more isolated
than those in a b jet:
$\Rightarrow$ cross-check in $b\bar{b}$ events



performance in $t\bar{t}$ events

for $t\bar{t}$ events satisfying the kinematic selection:

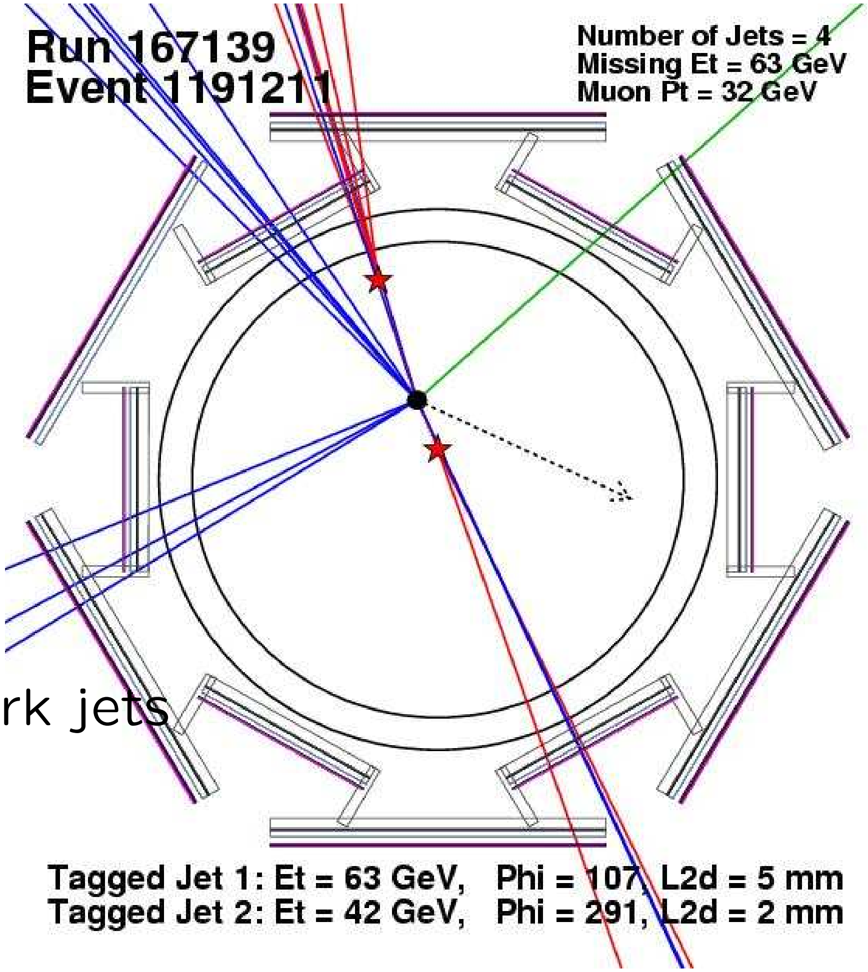
$\sim$ 60% have ≥ 1 jet tagged by SecVtx (tight version)

$\sim$ 16% have ≥ 2 jets tagged

$\sim$ 15% have ≥ 1 jet tagged by the soft muon tagger

backgrounds from mis–tagged light quark jets account for $\sim 12\%$ of the single SecVtx tagged $t\bar{t}$ data sample with ≥ 3 jets

double tagged $t\bar{t}$ candidate



work in progress

tag soft electrons inside jets

use more information about identified vertices:

- invariant mass of tracks
- relation between vertex & jet momenta
- fraction of jet tracks in vertex
- charge of vertex tracks
- …

combine information from various techniques in optimal way

→ artificial neural network

conclusions

a number of stable, well understood b tagging tools have been developed

techniques for measuring performance directly from data

more sophisticated tagging tools are being developed
⇒ larger, cleaner samples for top & QCD physics studies
& increased probability of discoveries