EPJ manuscript No.
(will be inserted by the editor)

Fermilab FERMILAB-CONF-05-501-E CDF

B tagging at CDF

Experience, performance, lessons for LHC

Daniel Jeans for the CDF collaboration

CNAF, Bologna, Italy & INFN Roma1

Received: date / Revised version: date

Abstract. We describe the algorithms used to identify b jets in CDF, and discuss various methods used to measure their performance.

PACS. PACS-key discribing text of that key – PACS-key discribing text of that key

1 Introduction

The identification of b jets is fundamental in the study of many interesting physics processes at high energy hadron colliders. Examples are the measurement of the top quark properties, the search for Higgs bosons, and precision tests of QCD.

b jets (jets whose originating parton is a b quark) can be identified in several ways, making use of the distinguishing characteristics of B hadrons with respect to hadrons containing only lighter quarks: their long lifetime ($\sim 1.5ps$), large mass ($\sim 5GeV/c^2$), and large decay fraction into leptons ($\sim 20\%$).

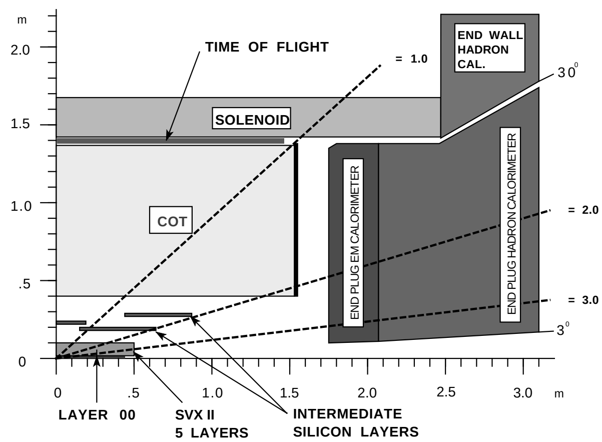

Fig. 1. Schematic of a quarter of the CDF detector

2 Tevatron and CDF

The Tevatron produces $p\overline{p}$ collisions at a centre–of–mass energy of 1.96 TeV, which take place at the centre of the experiments D0 and CDF. The luminous region is large: it is approximately Gaussian, with widths of around $30cm$ along the beam direction, and $26 \rightarrow 32\mu m$ in the plane transverse to the beam, varying along the z (beam) axis.

Bunch crossings occur every 396 ns, and at typical luminosities of $10^{32}cm^{-2}s^{-1}$, the mean number of interactions per bunch crossing is around three.

CDF [1] is a general purpose detector consisting of a high precision charged particle tracking system inside a uniform solenoidal magnetic field of 1.4 Tesla, electronic and hadronic calorimeters, and muon detectors. Some of its components are sketched in figure 1.

The Central Outer Tracker (COT) is a large wire chamber, which covers the region with pseudorapidity $|\eta| < 1$. It measures up to 96 points per track; half the wires are at a small stereo angle to the beam direction, allowing full three dimensional track reconstruction.

Inside the COT are the various components of the silicon tracker. Layer00 is mounted directly on the beam pipe, and is a single sided silicon detector, designed to be radiation hard. Outside L00 lie the five double sided silicon layers of the SVXII, followed by 1 or 2 layers of the Intermediate Silicon Layers (ISL). These layers contain strips parallel to, and at a small stereo angle to, the beam axis, giving full three dimensional information. The silicon tracker covers the region $|\eta| < 2$.

3 Tracking and Primary Vertex finding

Tracks are first reconstructed in the COT. These COT tracks are then extrapolated into the silicon detector, and matching silicon hits are attached to the track. The remaining unassociated silicon hits are then used to search for additional tracks, which are then extrapolated into the COT, and any matching COT hits are added to the track.

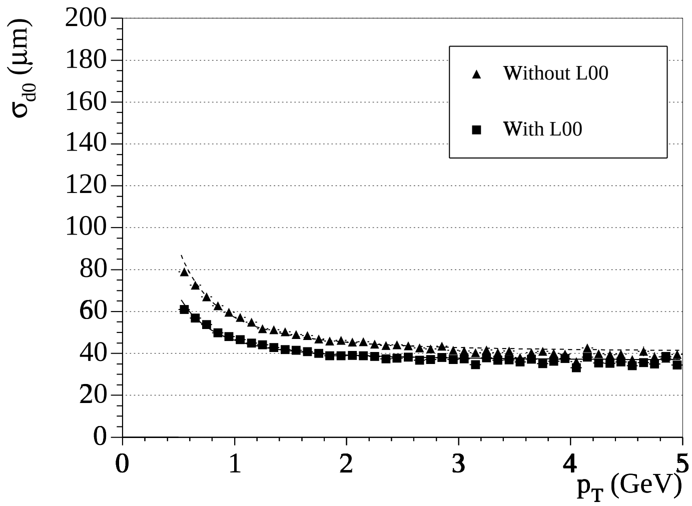

Fig. 2. Impact parameter resolution as a function of track p_T

Typical impact parameter resolution in the plane transverse to the beam for tracks with COT and silicon information is around 40 μm, including a contribution of around 30 μm from the width of the beam, as shown in figure 2.

Since the luminous region where the collisions take place is large, and the track reconstruction precision is good, improved information on the position of the interaction can be obtained by reconstructing the event primary vertex. First a seed position in z is identified by looking where an event's tracks approach closest to the beamline. Tracks displaced from this vertex by less than 1 cm in z, and with a two–dimensional impact parameter significance with respect to the beam position of less than three are used to fit a vertex constrained to lie inside the beamline. Tracks giving a large contribution to the vertex χ^2 are excluded from the vertex. Typical resolution on the primary vertex position is in the range $10 \rightarrow 32 \mu m$ in the plane transverse to the beam direction, depending mostly on the number of tracks used in the fit. This resolution is significantly smaller than the width of the beam.

4 Lifetime tagging algorithms

4.1 SecVtx

The SecVtx algorithm searches for track vertices inside a jet displaced from the primary vertex position, making use of the long lifetime of B hadrons.

Tracks lying inside the jet cone are considered; they are required to have both COT and silicon hits associated to them, and to satisfy various quality requirements. Tracks are required to lie within 2 cm of the primary vertex in z (to remove tracks from possible multiple interactions), and to have an impact parameter significance of at least 2.5 (to remove tracks produced at the primary vertex). In order to reduce the effects of particle interactions in the detector material, tracks with an impact parameter greater than 0.15 cm are rejected. Tracks identified as coming from K_S, Λ decays, or from photon conversions, are also rejected.

The remaining tracks are then used to search for a vertex: in a first pass a vertex made of at least three tracks is required; if such a vertex is not found, vertices with only two tracks (with more stringent track quality requirements) are accepted. The resolution on the separation of the primary and secondary vertices is typically $190 \mu m$. To identify a jet as a b jet, the significance of the separation between the primary and secondary vertices is required to be significant, and the χ^2 of the vertex fit reasonable; the vertex is required to lie on the "correct" side of the primary vertex with respect to the jet axis. Two track vertices reconstructed inside the detector material are rejected.

Two versions of this algorithm are in use, one optimised for higher efficiency ("loose"), the other for higher purity ("tight"); the precise requirements on track quality, vertex separation, and vertex χ^2 are different in the two versions.

More details can be found in [2].

4.2 JetProbability

The JetProbability algorithm also makes use of the long lifetime of the B hadron to tag b jets, by identifying jets whose tracks are unlikely all to have been produced at the primary vertex.

The impact parameter of tracks is signed with respect to the jet direction in such a way that tracks from long–lived particle decays are more likely to have positive impact parameters, while tracks from the primary vertex have equal positive and negative contributions.

The method is calibrated in generic jet data. Tracks are classified according to various quality criteria. In each track class, the negative side of the signed impact parameter significance distribution (dominated by tracks produced at the primary vertex) is parameterized.

To tag a jet, only tracks with a positive impact parameter are used. For a given track, the appropriate parameterization is used to calculate the probability that a track from the primary vertex would have a larger impact parameter significance. Using all tracks in the jet (after removal of identified K_S, Λ and conversion tracks), the per–track probabilities are combined to produce a per–jet probability. This is constructed in such a way that light–flavour jets have a flat probability distribution between 0 and 1, while jets containing long–lived particles tend to have a small probability. The JetProbability distributions for b and light quark jets are shown in figure 3.

b jets are typically tagged by requiring that the JetProbability is less than 1 or 5 %, depending on the efficiency & purity required by the analysis.

4.3 Data/simulation scale factor

The various processes to which b tagging is applied have b jets with different energy and η distributions. To understand the efficiency for correctly tagging an event, Mon-

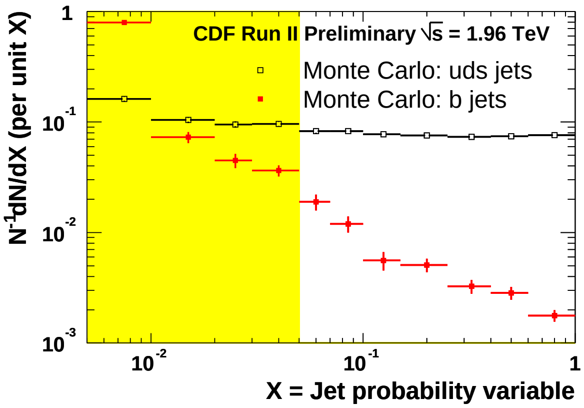

Fig. 3. JetProbability distribution

teCarlo simulation is used to take account of these differences.

To account for imperfections in the simulation (arising from, for example, imperfect description of silicon detector efficiency, tracking efficiency and resolution, or of the B hadron decay), the efficiencies measured in the simulation must be corrected by a "scale factor" to apply them to real data. This scale factor is measured in a large, independent dataset, and then applied to the simulation of the physics channel of interest.

First the efficiency of the btagging algorithm being considered is measured: the basic idea is to find a sample of jets with a high b content, and measure how many of the b jets are tagged by the algorithm.

To do this, events with a jet containing an identified lepton are selected; these jets have an enhanced heavy flavour content with respect to generic jets. To further enhance the HF fraction, the jet is required to be balanced by a second jet, which is required to be tagged as a heavy flavour jet.

The heavy flavour fraction in the jet containing the lepton (the "lepton jet") is estimated using several techniques: a fit of the distribution of the muon p_t relative to the jet axis (which is different for heavy and light quark jets), as illustrated in figure 4; the number of jets with both an identified electron and muon of opposite sign (one coming from the primary B hadron decay, the second from the decay of a charmed hadron from the B decay); or the number of lepton jets which contain an identified D^0 meson in addition to the charged lepton.

By using these techniques to estimate the number of heavy flavour jets in the lepton jet sample before and after applying the b tagging algorithm, the efficiency of the algorithm can be measured.

A sample of MonteCarlo data is then produced to simulate the lepton jet sample, and the same technique is used to measure the b–tagging efficiency in this sample. The comparison of the efficiencies measured in data and simulation gives the scale factor. It is typically in the range $82 \to 93 \pm 6\%$, depending on the tagger being considered.

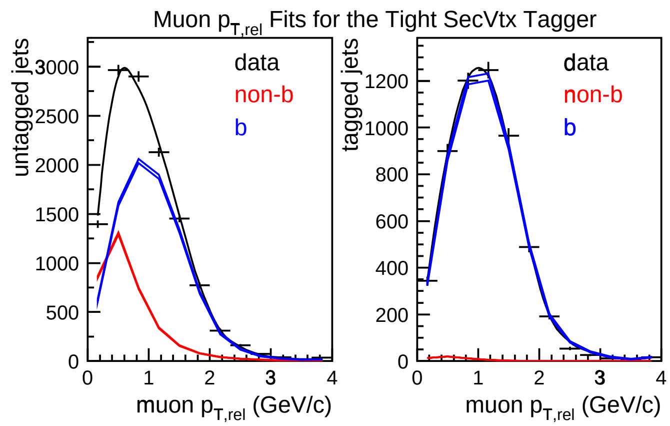

Fig. 4. Fits to μ relative p_T distributions

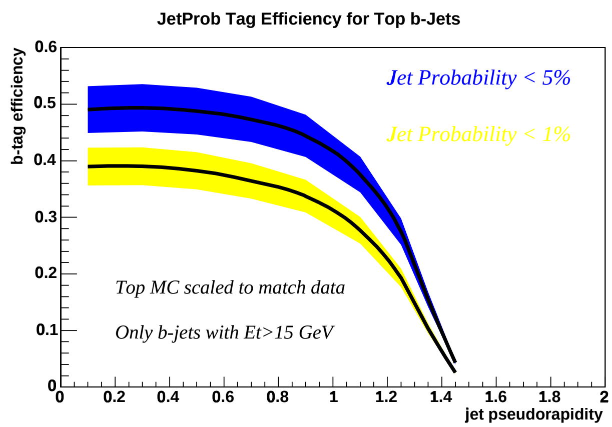

Fig. 5. JetProbability efficiency as a function of η

Figure 5 shows the efficiency of the JetProbability algorithm in $t\bar{t}$ MonteCarlo as a function of jet pseudorapidity. The efficiency has been corrected by the scale factor.

4.4 Mistagging probability

As well as understanding the efficiency of a tagging algorithm, it is also important to understand the mistagging probability: the fraction of light–flavoured jets which are incorrectly tagged as being b jets.

A first–order approximation of the mistagging rate is given by the negative tag rate. In the case of SecVtx, a negative tag is defined as when the identified vertex is well separated from the primary vertex, but lies on the "wrong" side of the primary vertex with respect to the jet direction. Such vertices are usually due to finite tracking resolution (or incorrect hit assignments), and are therefore assumed to be symmetrical positive–negative.

In the case of JetProbability, the "negative JetProbability" is the probability measured using only tracks with a negative signed impact parameter (with respect to the jet direction).

The fraction of jets with a negative tag is measured in jet data, as a function of the jet transverse energy, azimuthal angle and pseudorapidity, the number of tracks

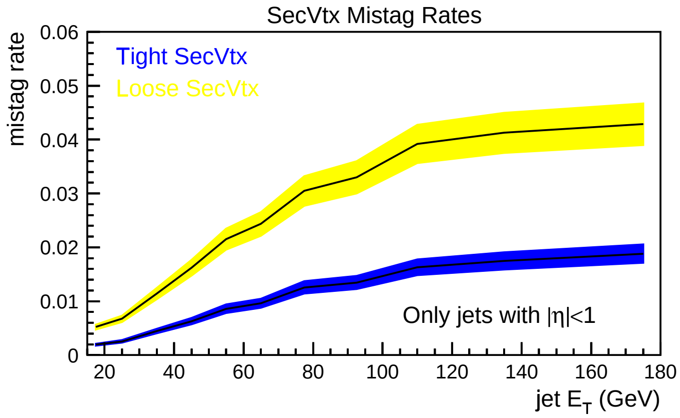

Fig. 6. SecVtx mistag rate as a function of jet E_T

inside the jet, and the sum of the transverse energies of all jets in the event. To estimate the mistagging rate, the negative tag rate is corrected for effects due to interactions in the detector material, unidentified long–lived strange hadrons (mostly K_S and Λ), and the b content of negative tags in the jet data.

Figure 6 shows the mistag rate for the SecVtx algorithm as a function of the jet transverse energy.

5 Soft Muon tagger

In total, around 20 % of B hadrons decay into muons, 11 % directly, and the remainder via a charmed hadron. These muons are non–isolated, and have a relatively soft p_T distribution. These properties preclude the use of the calorimeter for muon identification, and induce significant multiple scattering to the muon as it passes through the detector material. A dedicated muon identification algorithm has been developed to identify these muons, based on the matching of tracks to muon chamber track segments.

After muons identified as decay products of the $J/\psi, \Upsilon$ or Z bosons have been rejected, a jet is regarded as tagged if it contains an identified muon.

The efficiency of this algorithm is measured directly in data, by looking at "second legs" of J/ψ and Z events; it is measured to be between 70 and 90%, depending in the p_T of the muon. Since the muon tracks in these samples tend to be more isolated than those found in b jets, the efficiency measurement is cross–checked in $b\bar{b}$ events. The measured identification efficiency for muons in the central region is shown in figure 7.

The fake rate due to the positive identification of non–muon tracks is measured in generic jet data, which, after the removal of tracks due to $J/\psi, \Upsilon$ and Z decays, have a rather small true muon content. The fraction of remaining tracks identified as a muon is parameterised as a function of track p_T, azimuthal angle and pseudorapidity. Values of the fake rate are typically in the range $0.6 \rightarrow 0.9\%$. More details can be found in [3].

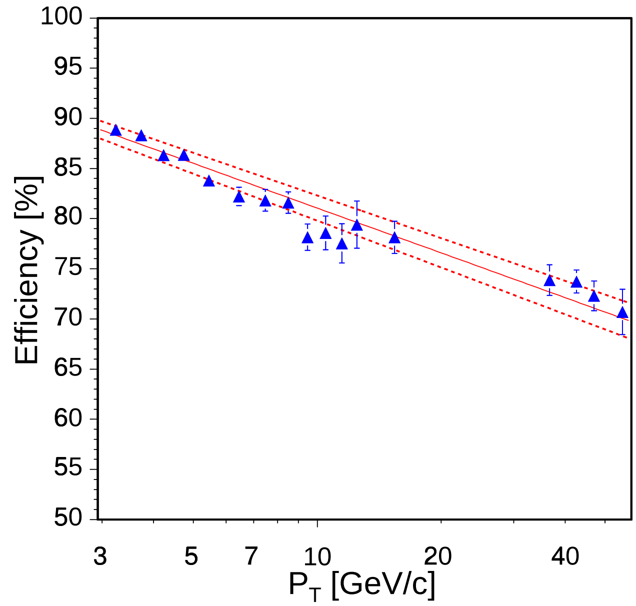

Fig. 7. μ ID efficiency as a function of μ p_T

6 Conclusions and plans for improvements

At CDF, a number of stable and well understood tools are used for the identification of b jets. The performance of the algorithms is measured in data, and correction factors to be applied to simulation are calculated.

At present, various new tagging techniques are being studied, including the identification of electrons inside jets, and more sophisticated tagging algorithms, based on Neural Networks, which make use of more information contained inside the jet.

References

1. D. Acosta et al., Phys. Rev. **D71**, (2005) 032001
2. D. Acosta et al., Phys. Rev. **D71**, (2005) 052003
3. D. Acosta et al., Phys. Rev. **D72**, (2005) 032002