Fermi National Accelerator Laboratory

FERMILAB-Conf-95/162-E
CDF

Top Decay to Lepton + Jets: CDF B Tags and Cross Section

J. Incandela
For the CDF Collaboration

Fermi National Accelerator Laboratory
P.O. Box 500, Batavia, Illinois 60510

July 1995

Published Proceedings for the *10th Topical Workshop on Proton-Antiproton Collider Physics*, Fermilab, Batavia, Illinois, May 9-13, 1995

Operated by Universities Research Association Inc. under Contract No. DE-AC02-76CHO3000 with the United States Department of Energy

FERMILAB-CONF-95/162-E
CDF/PUB/TOP/3209
August 1, 1995

Top Decay to Lepton + Jets : CDF B Tags and Cross Section[1]

The CDF Collaboration
J. Incandela
Fermilab, Batavia Illinois, USA

Abstract

The top search in the lepton plus jets channel with the Collider Detector at Fermilab (CDF) is presented. The analysis uses a 67 pb^{-1}data sample of $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV. Soft lepton tagging (SLT) and secondary vertex tagging (SVX) of b jets are used to reduce backgrounds. A significant excess of events over expected backgrounds is seen, and together with the excess of observed events in the dilepton (DIL) channel, firmly establish the existence of the top quark. Combining all channels, the $t\bar{t}$ production cross section is measured to be $\sigma_{t\bar{t}} = 7.6^{+2.4}_{-2.0}$ pb. The Branching of top to Wb is measured to be Br(t $\rightarrow$ Wb) $= 0.87^{+0.13}_{-0.30}(stat)$ $^{+0.13}_{-0.11}(syst)$.

Introduction

Recently, the CDF and D-Zero experiments have published conclusive evidence for the existence of the top quark [1, 2], confirming evidence presented by CDF a year earlier [3]. This paper will discuss the lepton plus jets component of the CDF top signal with an emphasis on b tagging.

The predominant production mechanism for top at the Tevatron is via $q\bar{q}$ annihilation. The event topology is dictated by the decays of the W's. In the lepton plus jets channel only one W decays leptonically as seen in figure 1. Considering only e or μ leptons, this channel represents approximately 35% of all top events as compared to 5% for dileptons. The remainder are in the

[1]Presented at 10th TopicalWorkshop on $p\bar{p}$ Collider Physics - May 10, 1995; Fermilab

purely hadronic and τ channels which have large multijet backgrounds. The lepton plus jets data are the main source for the top cross section measurement. In addition, events which contain four energetic jets can be used to fully reconstruct the top quarks in order to measure the top mass [4].

Top events in the lepton plus jets channels are distinguished by the presence of an energetic and frequently isolated lepton, missing transverse energy ($\not{E}_T$) from the W neutrino, and jets from the b's and the hadronic W decay. By requiring at least three energetic jets, the signal-to-background ratio is $\sim$1/6 for $M_{top} = 180$ GeV. Further background reduction is then afforded by tagging b jets. In this note we discuss the two methods of b tagging used by CDF. As a result of changes for the current run, the secondary vertex (SVX) method dominates our b tagging in both purity and efficiency and hence will be the main emphasis of this paper.

The data were taken in two separate runs of the Fermilab collider. Run 1a took place in 1992-93 and resulted in a total integrated luminosity of 19 pb^{-1}on tape. CDF observed a 2.8 σ excess of top-like events over expected backgrounds in this sample and measured a top mass of $174\pm10^{+13}_{-12}$ GeV and cross section of $\sigma_{t\bar{t}} = 13.9^{+6.1}_{-4.8}$ pb. There have been a number of important changes for run 1b which began at the end of 1993 and is still in progress. CDF has installed a new silicon detector with nearly identical geometry but which is ac-coupled with radiation-hardened electronics. This device has significantly higher signal and lower noise, resulting in an $\sim$ 20% higher average b tagging efficiency. Secondly, the SVX tagging algorithm has been modified, resulting in an additional 25% efficiency increase with no significant change in fake backgrounds. With regard to the SLT tagger, the μ acceptance has been extended to pseudorapidity $\mid \eta^{det} \mid \leq 1.0$. Most importantly, the Fermilab accelerator division has enhanced the performance of the Tevatron to achieve typical instantaneous luminosities of order 10^{31} $cm^{-2}sec^{-1}$ and peak luminosities approaching three times this value. For the results presented here, 48 pb^{-1}of data from run 1b have been combined with the 19 pb^{-1}from run 1a.

Event Selection

The event selection begins with the identification of an isolated e or μ lepton with $\mathrm{E_T}$ or $\mathrm{P_T} > 20$ GeV. Electrons from $\gamma \rightarrow e^+e^-$ conversions are vetoed

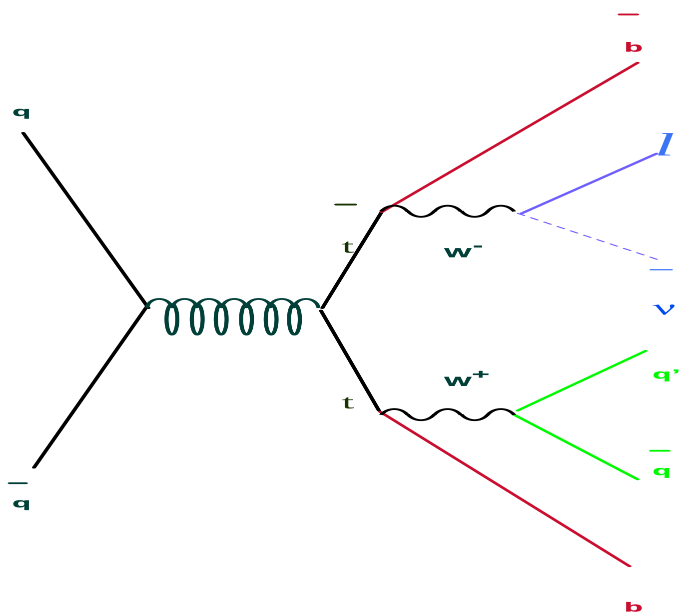


Figure 1: Leading Feynman diagram for top to lepton plus jets.

as are oppositely charged lepton pairs with $75 < M_{l^+l^-} < 105$ GeV. The lepton identification entails some implicit energy isolation. Nevertheless an explicit calorimetric isolation is applied by requiring that the energy in bordering towers not exceed 10 % of the energy in the tower containing the lepton. When these requirements are applied to the total 67 $\mathrm{pb}^{-1}$of data, we find 48000 e and 32000 μ events.

We remove all events with two isolated high energy leptons in order to be complimentary to the dilepton analysis. We then require uncorrected $\not{E}_T > 25$ GeV and three or more jets with $| \eta^{det} | \leq 2$ and uncorrected $\mathrm{E_T} > 15$ GeV. We are left with 203 events of which we expect ~ 30 to be $t\bar{t}$ for the case $M_{top} \sim 180$ GeV. Figure 2 shows the fraction of events as a function of jet multiplicity for direct W + jets production and for 170 GeV top MC. It is seen that the majority of top events lie in our predefined signal region of 3 or more jets. Note that a significant number are also found in the 2 jet bin.

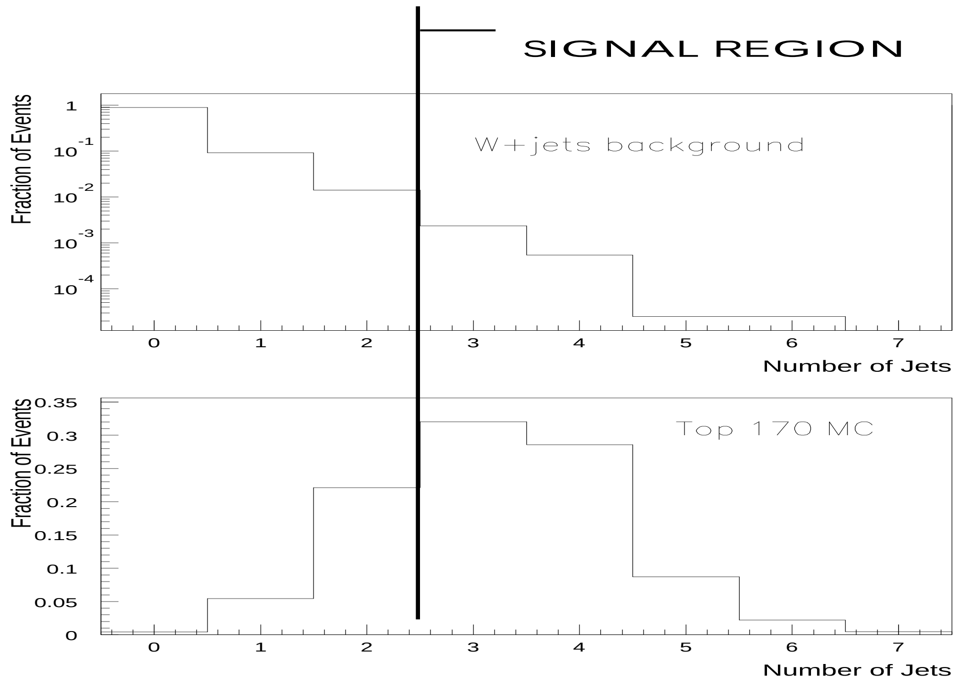


Figure 2: Jet multiplicities for W plus jets data and 170 GeV top MC.

Soft Lepton Analysis

The basis of this method is the fact that $t\bar{t}$ events contain 2 b's, 2 - 3 c's, and sometimes a τ from W decay, all of which have significant branching to e and μ leptons. On average, one expects approximately 0.8 additional e or μ leptons per event. Unlike leptons from W decay however, these will be lower P_T and generally occur inside of jets. The SLT analysis requires the lepton to have E_T or $P_T > 2$ GeV.

A good source of non-isolated electrons are those from photon conversions which can be identified cleanly in CDF. J/ψ decays provide an analogous source of low P_T, non-isolated μ's. We use these to define and study our identification criteria and to measure their efficiencies. The variables we use to identify electrons and muons are discussed in reference [3] and include many of the standard CDF variables, modified to reduce implicit isolation requirements. Explicit isolation requirements are avoided.

In order to determine false tagging rates, we parametrize the tag rate per track as a function of P_T using dijet samples. We have checked our tag

Table 1: b tagging results in the lepton plus jets sample

Subsample N jets	SLT est. bkg.	SLT Tags (Evts)	SVX est. bkg.	SVX Tags (Evts)
$W\ +\ 1\ Jet$	159 ± 24	163 (161)	50.4 ± 12.4	40 (40)
$W\ +\ 2\ Jets$	47 ± 7	55 (54)	21.1 ± 6.5	34 (30)
$W + \geq\ 3\ Jets$	15.4 ± 2.0	23 (22)	6.7 ± 2.1	27 (21)

rate parametrization on different samples (e.g. jets from different jet trigger samples or from photon samples) and find numbers of tags in good statistical agreement with the numbers we predict. We estimate that roughly 25% of dijet tags are due to real heavy flavor. Tag rates in b-enriched samples significantly exceed fake rate predictions as desired.

To determine our top tagging efficiency we use top MC samples. We proceed by identifying tracks associated to leptons from b, c or τ decays using generator level information. We then use the $\mathrm{P_T}$ of the track and our cut efficiencies to determine the probability that this lepton will be identified. The density of tracks in the vicinity of the lepton is also taken into account. We find that the efficiency for tagging top via an e or μ from heavy flavor or τ decay is $\epsilon_{slt}\ =\ 20\ \pm\ 2$ %. There is also a roughly 10% probability that a top event is tagged via a track not associated to a lepton. The probability of tagging a W plus 3 or more jets event is roughly 8% for comparison. The higher "mistag" probability for top events is due to the higher track multiplicity.

The results for the SLT analysis are presented in table 1. Note that the estimated numbers of tags in the 1 and 2 jet bins are consistent with the number observed, while a small excess is seen in the three or more jets region.We determine the probability that this excess is due to a background fluctuation to be $P\ =\ 6 \times 10^{-2}$ ($\sim 1.9\sigma$).

SVX Analysis

Bottom quarks from heavy top decay are very energetic. For $M_{top}\ \sim$ 180 GeV the mean transverse displacements of the b and sequential c decay vertices are $\sim 3.5\ mm$ and $\sim 5.3\ mm$, respectively. Tracks from these decays generally have large impact parameters relative to the primary vertex. The SVX analysis uses the excellent hit resolution of the silicon vertex detector, figure 3, to resolve these tracks and to use them to construct displaced vertices. The impact parameter resolution of the silicon detector is $\sigma_d \sim\ 16\ (1\ +\ (0.8/\mathrm{P_T})^2)\ \ \mu m$.

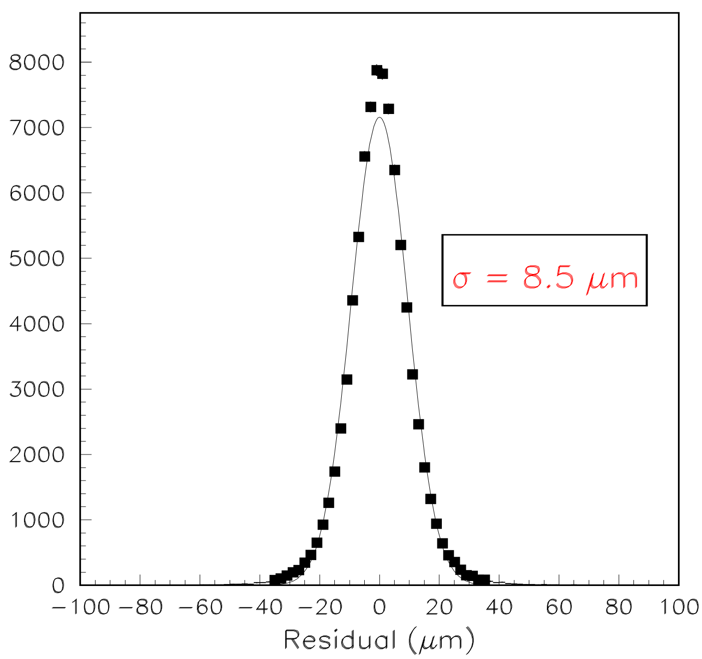


Figure 3: Tracking residuals for all 4 layers of the new radiation hard silicon microstrip detector (SVX′). The inner and outer radii of the detector are 2.8 and 7.9 cm, respectively.

Tagging Algorithm

The SVX tagging algorithm begins by selecting displaced tracks in a cone of radius $\sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.4$ about the axis of the selected jet. The minimal requirement for tracks used in vertexing are : (i) impact parameter significance $S_d \equiv d/\sigma_d \geq 2.5$, (ii) $\mathrm{P_T} \geq 0.5$ GeV, and (iii) track quality cuts. The track quality cuts pertain to the number of hits in the Central Tracking Chamber (CTC) and the number and quality of hits in the SVX. The fewer the number of SVX hits, the more likely the track is erroneous (e.g. combines hits from several particles). Our track quality cuts and our vertexing algorithm take this into account by making additional requirements for tracks with fewer SVX hits and by not using tracks with fewer than 3 SVX hits to form vertices with only two tracks.

The vertexing algorithm is comprised of two passes. The first uses "seed vertexing" which takes advantage of the fact that vertices with 3 or more tracks have much lower fake background than those with 2 tracks and thus do not require as strict track selection requirements. The algorithm starts by forming a 2 track "seed" vertex from a quality-ordered list of displaced tracks. It then loops over the remaining tracks in the list to see if any can be associated to the seed (impact parameter within 3 σ). If at least three tracks are vertexed, the algorithm stops and a positive decay length tag is formed if the transverse displacement of the vertex, L_{xy}, satisfies $L_{xy} > 3 \cdot \sigma_{L_{xy}}$. The sign of the vertex is given by the sign of the dot product of the vector from the primary to the secondary vertex and the jet direction. A negative decay length tag can occur as a result of resolution smearing or pattern recognition mistakes. If no track can be associated to the seed, a new seed is formed from the tracks in the list and the process is repeated. If all pairs of tracks are tried and no 3 or more track vertex can be formed, a second pass algorithm is initiated. For this case, since we are looking for a two track vertex, we tighten all of our track requirements: $S_d > 4$ (3) and $\mathrm{P_T} > 1.5$ (1.0) GeV for the first (second) track. The addition of seed vertexing has led to a 25% increase of tagging efficiency ϵ_{svx} relative to our previous algorithm with no increase in fake tag rate.

Understanding and Simulating the Silicon Detector

Since the SVX detector plays such a crucial role in b tagging we expended some effort to understand signal formation and noise in the detector and to develop an accurate simulation. Our primary motivation was an $\sim 30\%$ discrepancy between data and MC b tag efficiencies in run 1a.

In simple terms, we identified three main contributions to the formation of signal clusters, defined as strings of contiguous microstrips with charge above threshold. These are (i) primary signal due to the -dE/dx of the ionizing particle forming electron hole pairs in the 300 μm bulk of the silicon, (ii) secondary ionization such as δ rays, and (iii) noise. We parametrized these components and their correlations to one another and to primary track parameters in order to model the signal. The model results in good agreement in charge distributions and multiplicity of strips as a function of track parameters, as seen in figures 4 and 5, without parametrizing these quantities directly. The simulation also reproduces the detector position resolutions as a function of the number of strips in the clusters. These are 13, 11 and 19 μm for 1, 2 and 3 strip clusters, respectively.

As it turns out, the new detector simulation, has only $\sim 5\%$ effect on the MC b tag efficiency in run 1b. This is because the signal-to-noise ratio of the new detector is sufficiently high that hits on tracks are formed and detected with nearly 100 % efficiency in active silicon regions while noise clusters are virtually non-existent. The 30% discrepancy in run 1a is now understood to be due to reduced hit finding efficiency and more excessive resolution degradation in the previous detector as a result of radiation damage to the readout electronics. The radiation damage resulted in a large increase in noise and a loss in amplifier gain.

Tagging in b-Enriched Data and MC Control Samples

The most important data-to-MC comparisons are those for b jets. To make meaningful comparisons one must understand how reasonable variations in MC generation of b jets affect their experimentally observable characteristics. With regard to data, it is necessary to isolate a b-enriched sample and to make a reasonably accurate estimate of the b content.

For MC systematics we studied observable quantities relevant to b tagging such as the number of tracks in b-jets and their P_T, impact parameter

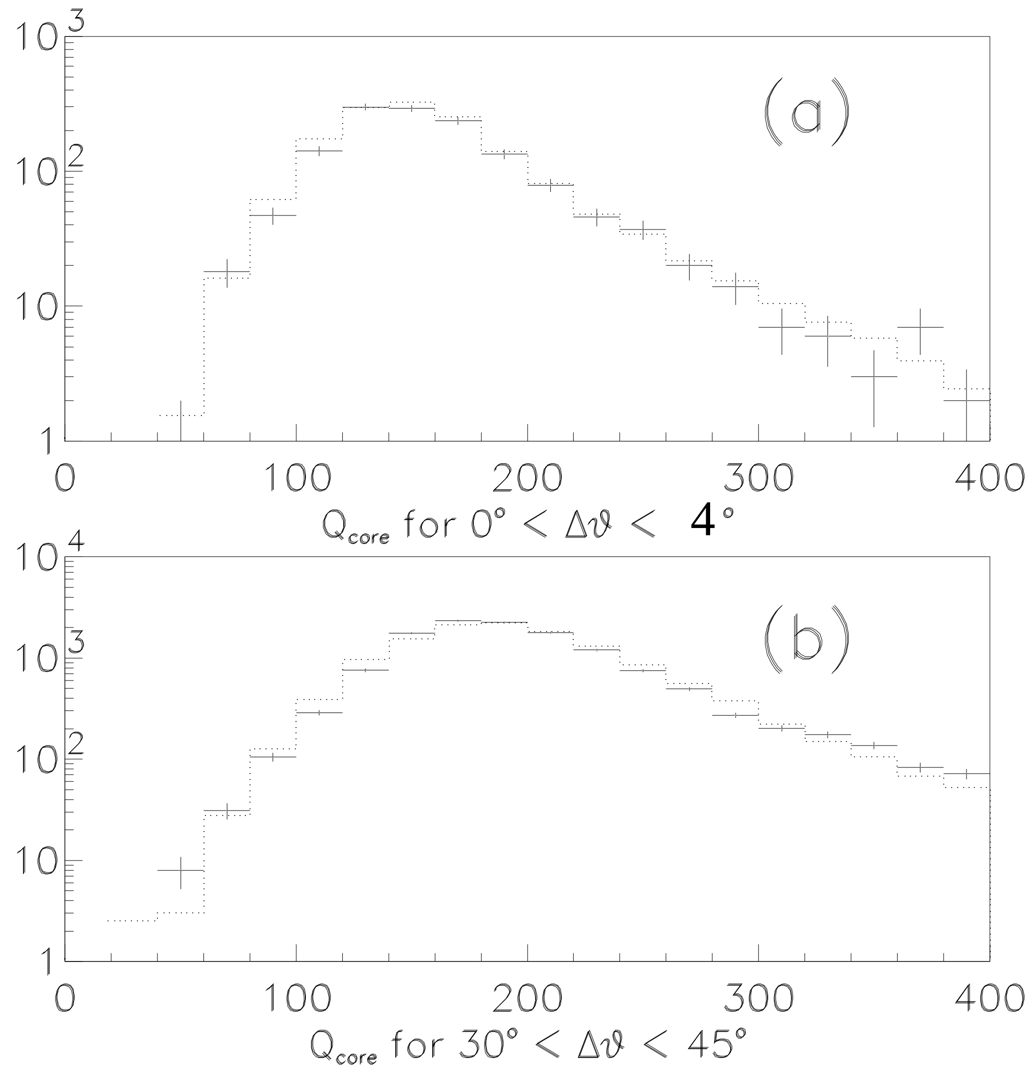


Figure 4: Charge distributions from MC (dotted) and data (solid) for tracks with incident polar angle θ (a) near to and (b) far from the silicon wafer normal.

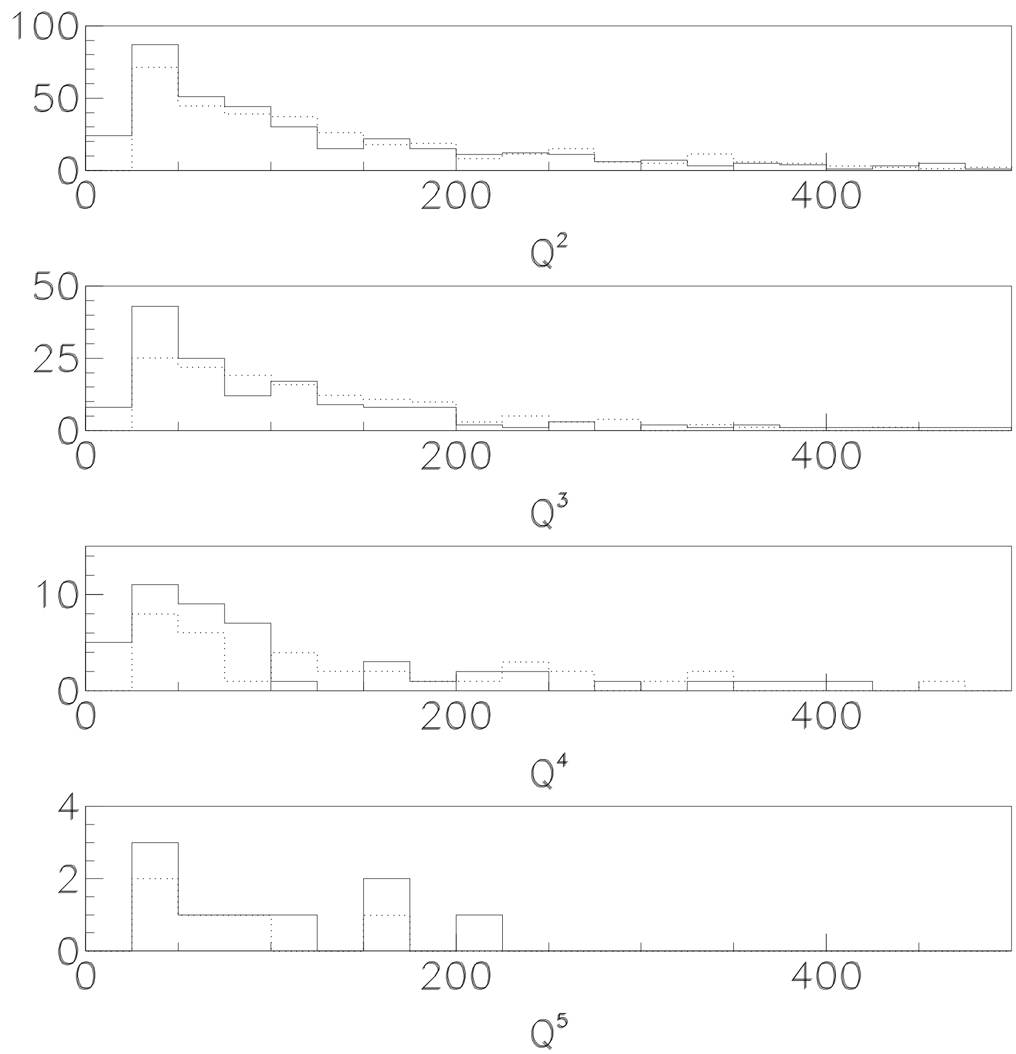


Figure 5: Comparison between data (solid) and MC (dotted) charge distributions on the second through fifth additional strips in long SVX clusters due to secondary ionization such as δ rays. The data and MC samples are chosen from the same number of tracks and are not otherwise normalized. The frequency of occurence and the shapes of the distributions are well-reproduced.

significance (S_d), and momentum relative to the jet axis. We compared distributions of these quantities for a wide variety of MC types and configurations (see [5] and references therein). We use PYTHIA with the CLEO b-decay tables and the full CDF detector simulation as our default with b-fragmentation tuned to match that measured at LEP ($x(E)_B = 0.70 \pm 0.01$). For our study of PYTHIA, we varied fragmentation by three times the LEP uncertainty, turned off initial state radiation, varied structure function sets and used the JETSET decay tables. The effects on b tagging quantities were negligible. We further compared our PYTHIA default to HERWIG and ISAJET defaults, again with no significant effect on quantities relevant to b tagging.

We then used the default PYTHIA to generate $b\bar{b}$ events for which at least one b decayed semileptonically to an electron with $E_T > 10$ GeV. We compared characteristics of tagged electron jets and tagged jets opposite the electron to their analogues in our data. For the latter we used inclusive electron samples. We estimate a b-purity of order 90% for tagged electron jets. Figure 6 compares data and MC distributions for tagged electron jets. The agreement is good.

We also use these samples (and analogous μ samples) to measure the MC and data b tagging efficiencies as a function of jet E_T. The efficiency is determined by the methods described in [3]. The most accurate method is to select the subsample of events for which the jet opposite the electron jet is tagged. This sample is predominantly ($\sim$ 95%) $b\bar{b}$ in our data. We then measure the fraction of tagged electron jets and compare this to MC as shown in figure 7. We find that the data and MC efficiencies match well and predict a b tagging efficiency in excess of 40% for b jets above $\sim$ 30 GeV as is typical in heavy top decays. The ratio of data to MC efficiencies is 0.96 ± 0.07.

Results

The new b tag algorithm was chosen via an optimization in which we compared the calculated probabilities of observing 3,4 or 5 σ excesses of 170 GeV top signal over background in 50 and 100 pb^{-1}samples for 3 different sets of track selection criteria. The 3 cases considered spanned what we felt was the acceptable range of mistag rates. The discovery potential of the 3 cases was very similar and we chose the one with both the highest b efficiency and mistag rate. For this case we calculated a 50% probability of observing a 4

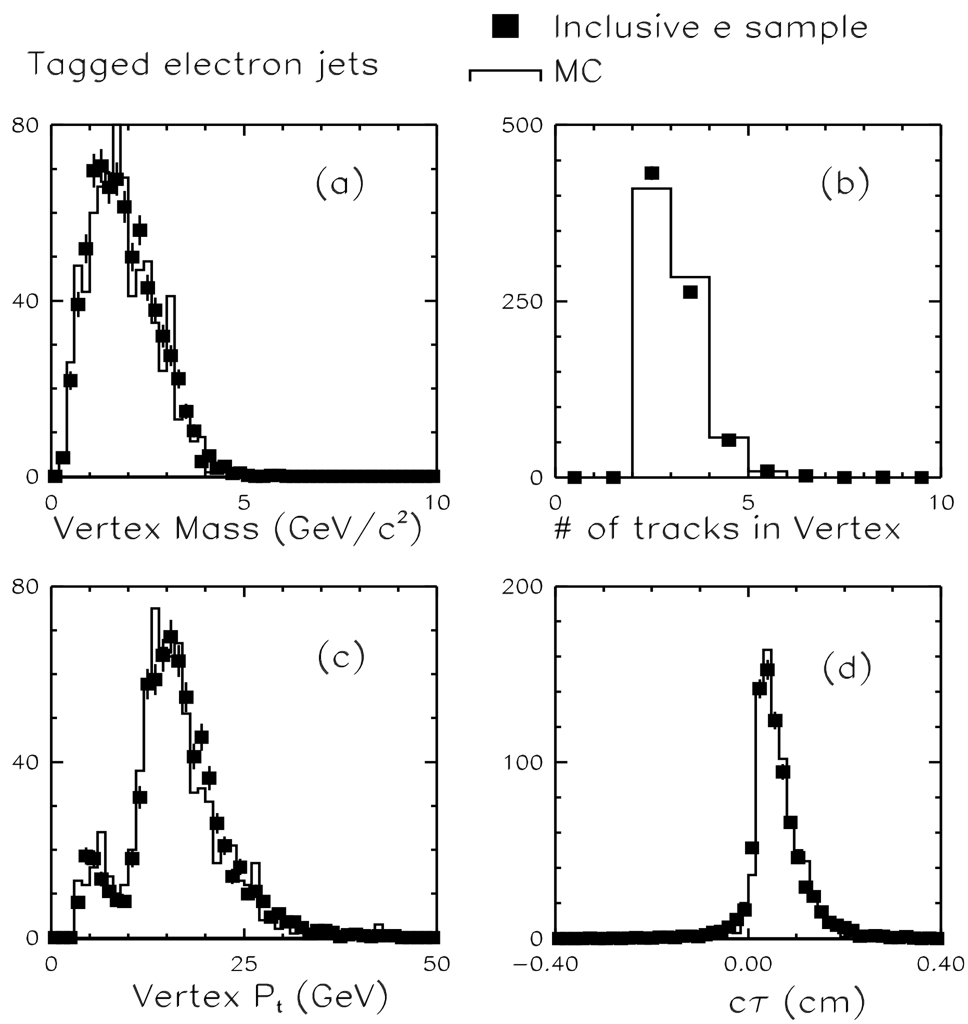


Figure 6: Properties of tagged electron jets in inclusive electron data and $b\bar{b}$ MC.

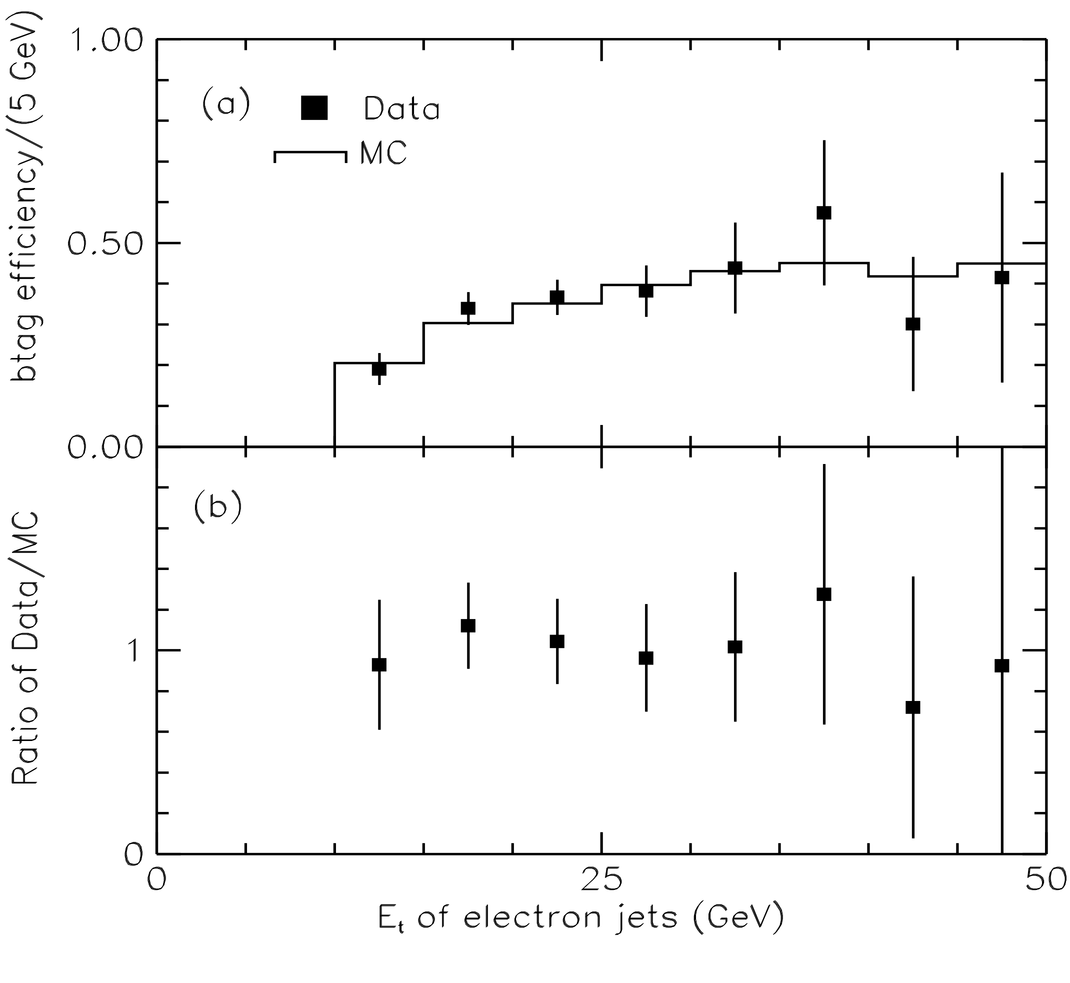


Figure 7: a) B tag efficiency and b) ratio of efficiencies vs. e jet E_T.

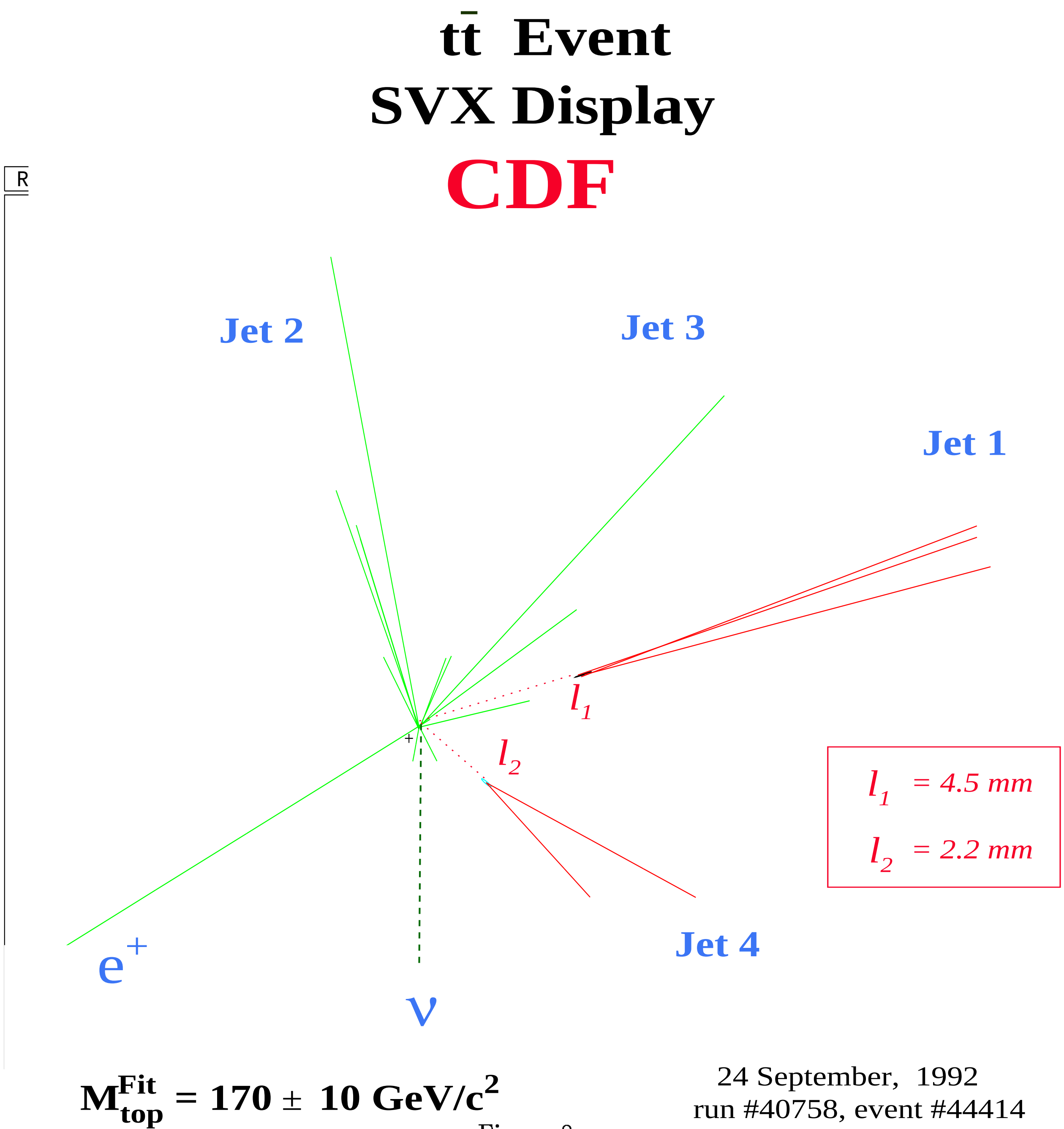
σ excess of signal in 67 $pb^{-1}$of data. In 170 GeV top MC we determined the top tagging efficiency of the chosen algorithm to be $\epsilon_{svx} = 42 \pm 5$ %.

For top in the mass range from 160 to 180 GeV the production cross section [6] varies between 8.2 and 4.2 pb. We expect between 20 and 10 tagged top events in 67 $pb^{-1}$and observe 27 tagged jets in 21 events as indicated in table 1. The first tagged event in run 1a is shown in figure 8. The 3 track tag (jet 1) is the result of seed vertexing. The event has all of the expected characteristics of top and is fit to a top mass of 170 ± 10 GeV while the invariant mass of the untagged jets is $\sim$ 79 GeV and consistent with W decay. Figure 9 shows the transverse mass distribution for the lepton and neutrino in the tagged events and compares it to 170 GeV top MC. The data are seen to be consistent with the MC prediction.

Backgrounds and Significance of Excess

The background events in the sample are predominantly tagged, direct W plus jets events where the tagged jet is either real heavy flavor (e.g. $Wb\bar{b}$, $Wc\bar{c}$, Wc as shown in figure 10) or a mistagged light quark jet. Other much smaller backgrounds include $b\bar{b}$, WW, WZ and $Z \rightarrow \tau\tau$. To determine the background in our sample we first parametrize the negative decay length tag rate in inclusive dijet data as a function of jet E_T, track multiplicity, and event total transverse energy (ΣE_T). We apply this to the jets in our lepton plus jets sample prior to tagging to estimate the expected number of mistagged events. The parametrization is compared in numerous ways to dijet samples obtained with other trigger thresholds as well as photon triggers and found to be in good agreement ($\pm$15%) in all cases. For W plus heavy flavor we use MC samples to calculate the fraction of events expected from various background sources in each of the jet multiplicity bins and their tagging efficiencies. The number of background events is then obtained as the product of these quantities determined for a given jet multiplicity, times the observed number of events with this number of jets. The results are shown in table 1. Figure 11 shows the same information graphically.

The probability that the excess of tagged 3 or more jet events is the result of a background fluctuation was conservatively estimated to be 2×10^{-5} corresponding to $\sim$ 4 σ for a Gaussian probability distribution. Combining this with the results from the SLT analysis and the Dilepton analysis [7] the combined probability for the excess seen in all channels to be the result of a



$t\bar{t}$ Event
SVX Display
CDF
Jet 2
Jet 3
Jet 1
l_1
l_2
l_1 = 4.5 mm
l_2 = 2.2 mm
Jet 4
e^+
ν
M_{top}^{Fit} = 170 ± 10 GeV/c^2
24 September, 1992
run #40758, event #44414

Figure 8:

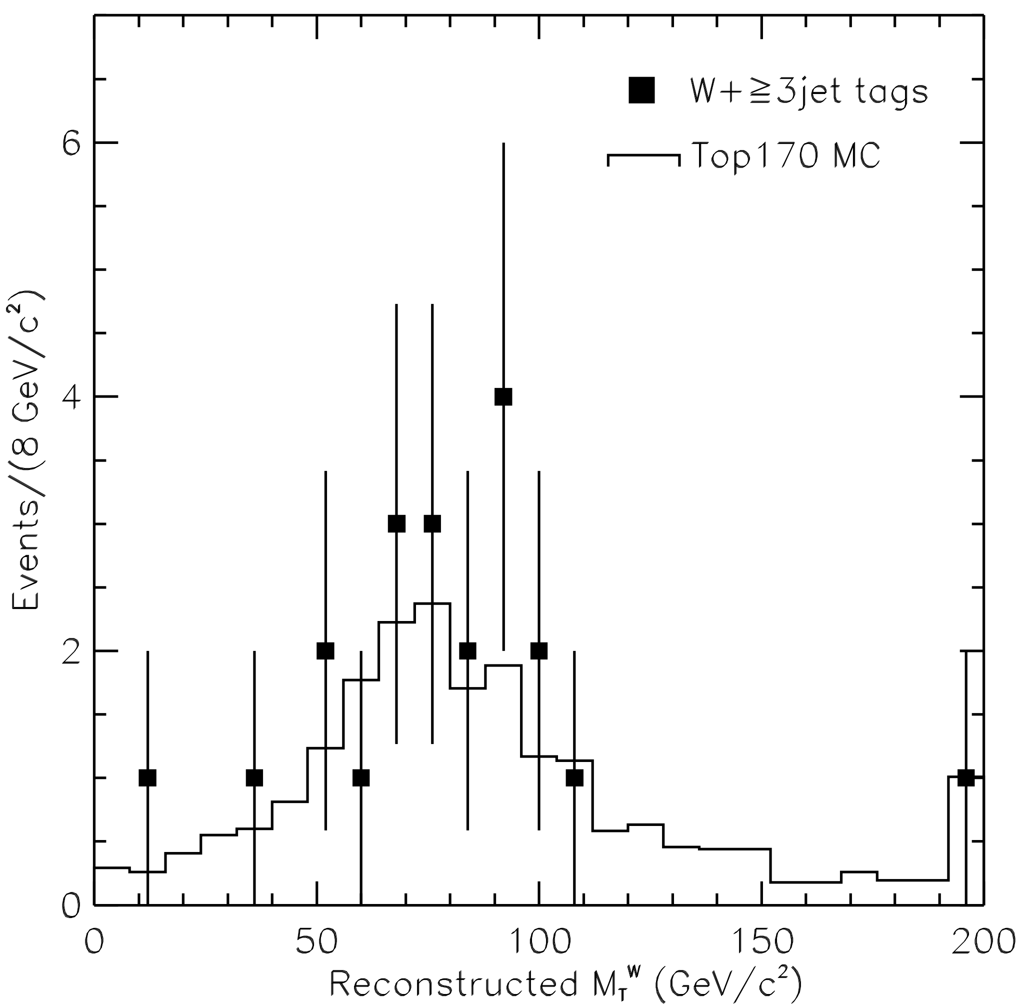


Figure 9: Transverse mass of lepton and $\not{E}_T$ for SVX-tagged data and top MC. (The last bin includes overflows.)

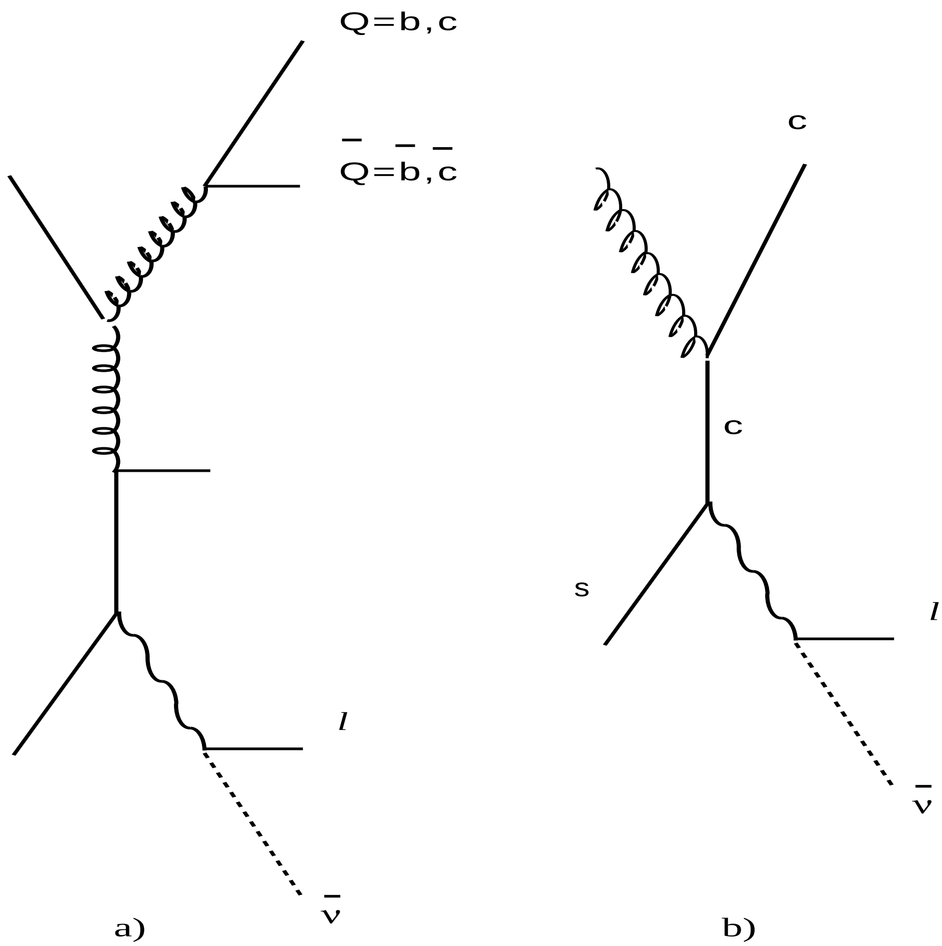


Figure 10: Feynman Diagrams for a) W plus $b\bar{b}$ or $c\bar{c}$ and b) W plus c.

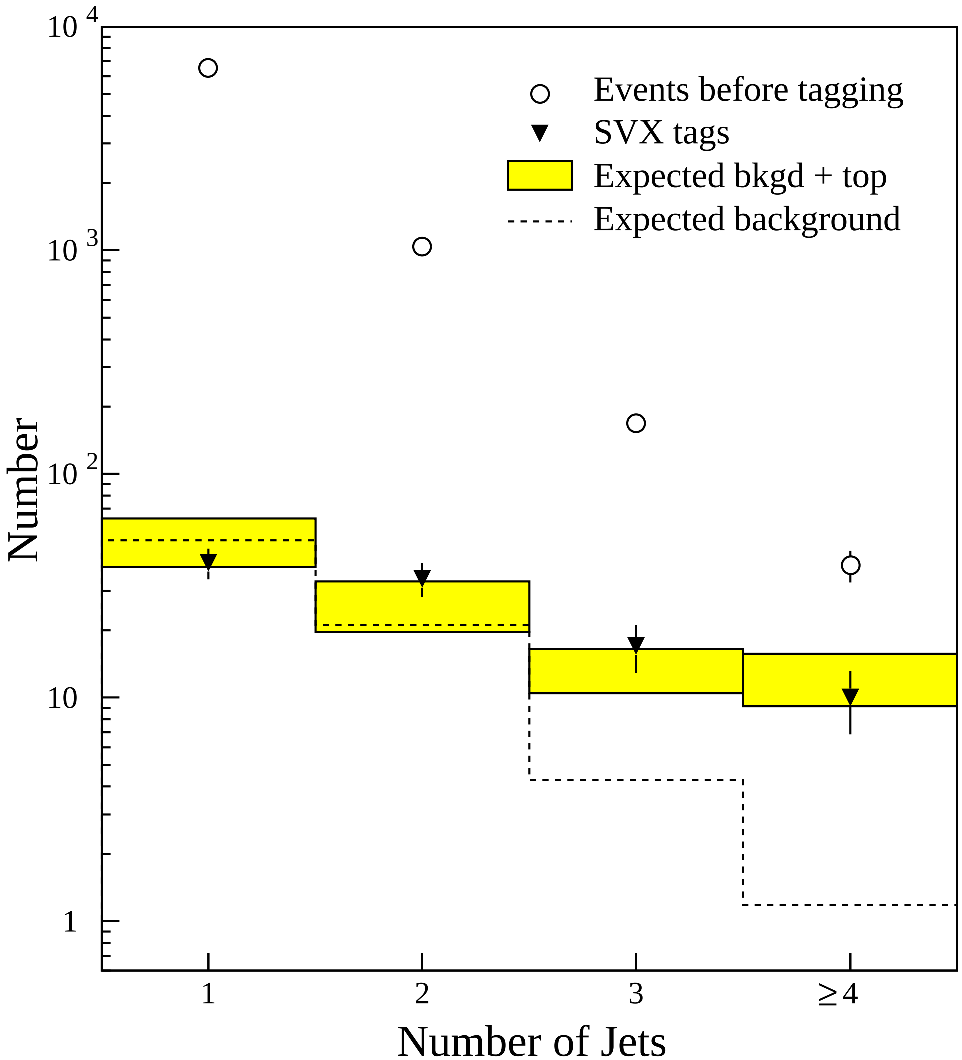


Figure 11: The number of observed tagged jets and the number expected from background in the absence of, and including, a contribution from 170 GeV top.

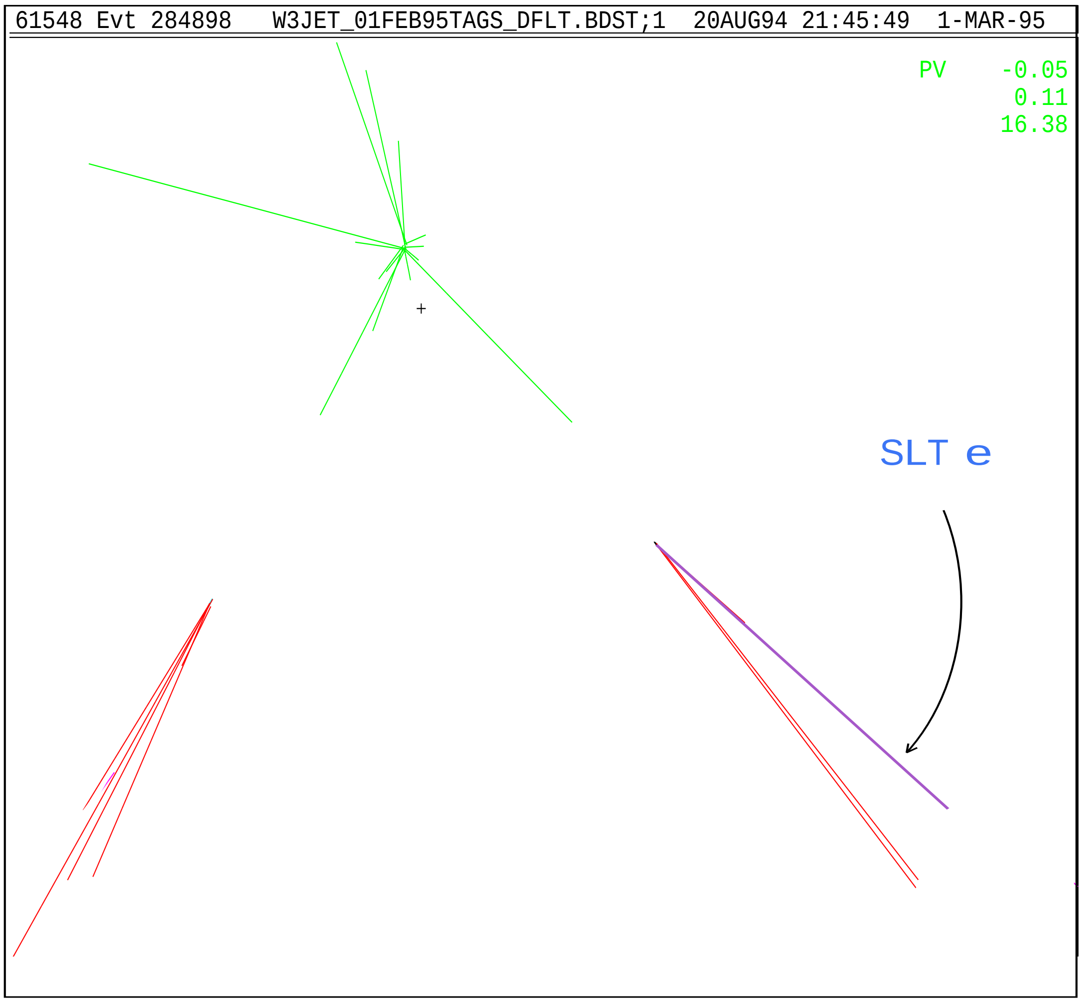


Figure 12: Triple-tagged event display $\sim 1.0\ \text{cm}^2$ about the primary vertex.

background fluctuation is 1×10^{-6} corresponding to $\sim 4.8\ \sigma$.

A number of crosschecks can be performed to validate this result. For instance, we observe 6 events with two SVX tagged jets, which is consistent with an expectation of ~ 4 for 170 GeV top plus background but significantly higher than the expectation of < 1 for background alone. Similarly, there are 6 events with both SVX and SLT tags (usually in the same jet) and three of these have 2 SVX tags. This is also consistent with top. Figure 12 shows the vertex region display for one of the triple-tagged events.

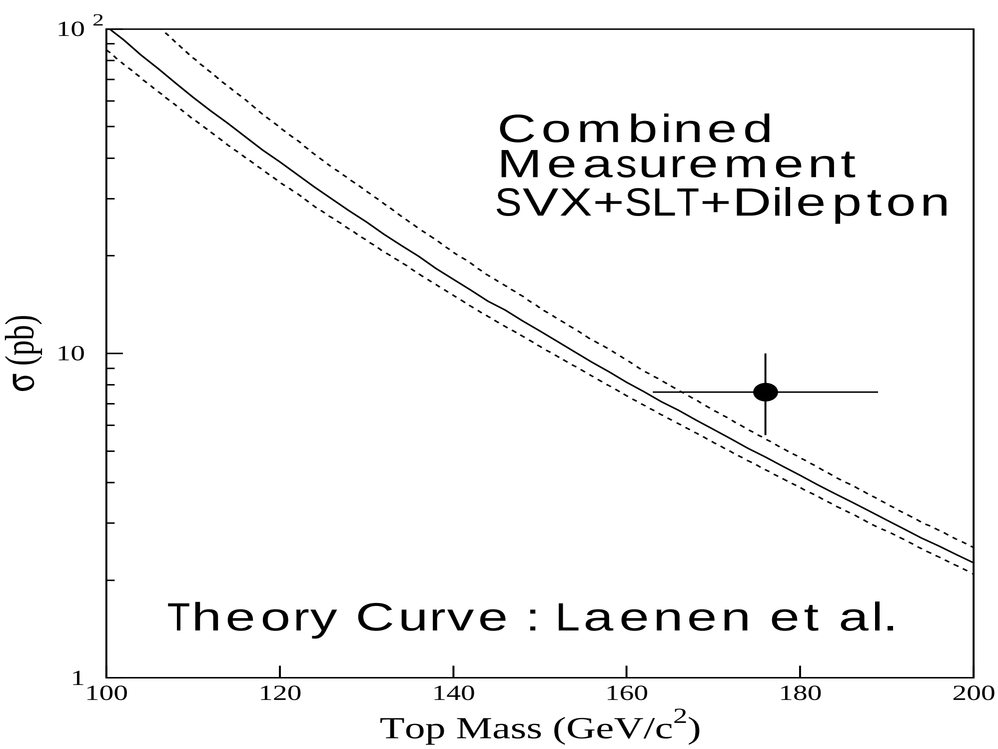


Figure 13: CDF cross-section and mass measurements with expectation of theory.

Finally, if we were to see significantly more tagged Z plus jet events than we expect, it would indicate that we are probably underestimating our W plus heavy flavor background. For 1,2 and 3 or more jets, the predicted numbers of tagged Z plus jets events are 17.5, 4.2 and 1.5 and we observe 15, 3 and 2, respectively. The good agreement gives us confidence that our estimates for W plus jet backgrounds are reasonable.

Cross Section and Branching Fraction

From the various counting experiments we are able to calculate the top production cross section in a -Log(likelihood) minimization using the standard expression

$$\sigma = \frac{n-b}{\epsilon \times \int L dt} \tag{1}$$

The total efficiency ϵ, total number of events n, and estimated number of background events b, are decomposed in the fit process into factors which are or are not common to the two runs and the various channels being combined. Our best result is obtained by combining all of the counting experiments (SVX, SLT and DIL) to obtain $\sigma_{t\bar{t}} = 7.6^{+2.4}_{-2.0}\ pb$. Figure 13 plots the CDF measured mass and cross section point along with the theoretical expectation [6]. The measured value is seen to be consistent with theory.

One can also use our data to make a direct measurement of the branching fraction $Br(t \rightarrow Wb)$ which is expected to be essentially 100% in the standard model. In practice one uses the acceptance-corrected numbers of events with 0,1 or 2 b tags in the dilepton signal sample, and 1 or 2 b tags in the lepton plus jets sample in a maximum likelihood estimator. The basic idea is that $Br(t \rightarrow Wb)$ values significantly smaller than 100% would have a noticable effect on the relative number of multiply tagged top events since it would imply a significant reduction in the number of top events containing two b quarks. We find

$$Br(t \rightarrow Wb) = 0.87^{+0.13}_{-0.30}(stat)^{+0.13}_{-0.11}(syst) \tag{2}$$

Work in Progress

We continue to collect and analyze new data. Initial studies of roughly 25 pb^{-1}of new lepton plus jets data indicate that the rate of tags in the 3 or

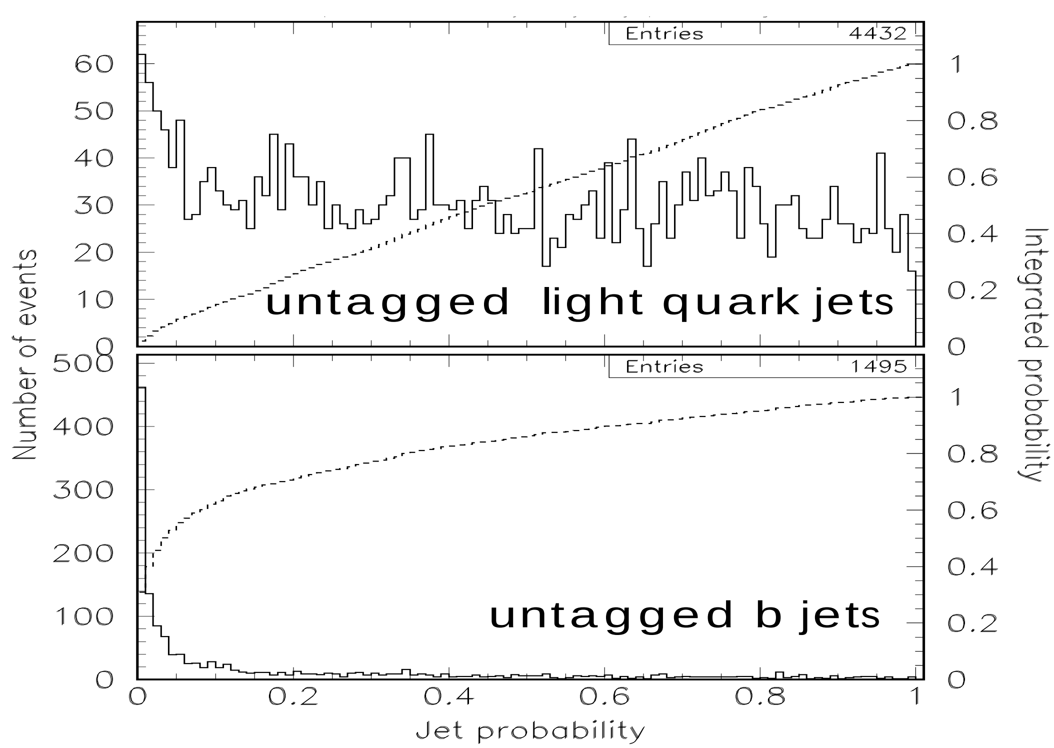


Figure 14: Jet probability distributions for untagged light quark jets and untagged b jets in SVX-tagged top events from 175 GeV top MC.

more jet region is consistent with what we have presented here. In other areas we are interested in using SVX track information to help distinguish untagged b jets from light quark jets in our tagged top events. By identifying the second b in the event, one severely limits the number of possible combinations to be considered in reconstructing the event while also facilitating the identification of the hadronic W decay jets which could improve our kinematic studies and mass measurement. To this end we have employed the jet probability algorithm [8] to study untagged jets in tagged top events in 175 GeV top MC. The algorithm uses track impact parameters to define a probability which has a flat distribution for light quark jets but is peaked at low probability values for heavy quark jets. Figure 14 shows the distributions for untagged light quark jets and untagged b jets in 175 GeV MC events which are tagged by the SVX b tag algorithm. It is seen that this variable affords significant discrimination between light and heavy quark jets. It does not however distinguish as strongly between b and c jets which results in a $\sim 20\%$ reduction in purity for identification of the second b jet.

References

[1] F. Abe et al. "The CDF Collaboration", *Observation of Top Quark Production in $p\bar{p}$ Collisions with the Collider Detector at Fermilab,* Phys. Rev. Lett. **74** (1995) 2626

[2] S. Abachi et al. "The D-Zero Collaboration", *Observation of the Top Quark,* Phys. Rev. Lett. **74** (1995) 2632

[3] F. Abe et al. "The CDF Collaboration", *Evidence for Top Quark Production in $p\bar{p}$ Collisions at $\sqrt{s} = 1.8$ TeV,* Phys. Rev. **D50** (1994) 2966

[4] S. Vejcik these proceedings.

[5] C.Chiou et al. *A Generator Level Study of b Jets from Top Decay,* CDF internal note. 2568 (1994)

[6] E. Laenen, J. Smith, and W. L. van Neerven, Phys. Lett. B**321**(1994)254

[7] R. Roser these proceedings.

[8] D. Buskulic *et al.*, "The ALEPH Collaboration", Phys. Lett. **B 313**, (1993) 535.