Fermi National Accelerator Laboratory

FERMILAB-Conf-95/176-E
CDF

$t\bar{t}$ Kinematics in W+ $\geq$ 3 Jet Events

Morris Binkley
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

Disclaimer

CDF/PUB/TOP/GROUP/3188
July 12, 1995
Version 2.1

$t\bar{t}$ Kinematics in W+ $\geq$ 3 Jet Events

Morris Binkley
(for the CDF Collaboration Fermilab)

Presented at the 10th Topical Workshop on $p\bar{p}$ Collider Physics
Fermilab, May 9-13, 1995

Abstract

The objective of this talk is to compare the CDF W+ $\geq$ 3 Jets data with Monte Carlo predictions using the standard model with the top quark. The data is seen to be consistent with these predictions using the the $t\bar{t}$ production rate indicated by the SVX b-tagging and the previously reported top mass.[1]

1 Introduction

$t\bar{t}$ kinematics in W+ $\geq$ 3 jet events where the W decays leptonically is presented using 67 pb^{-1} of data from the CDF detector. In the first section, distributions of directly measured quantities such as jet or lepton energies and directions are shown. The second section examines $t\bar{t}$ production variables requiring mass fits to make jet assignments.

For all of the studies in this talk, the $t\bar{t}$ simulation is done with Herwig, a parton shower Monte Carlo using leading order QCD for the hard process. The non-top QCD W+ $\geq$ 3 Jets is simulated with VECBOS, a parton level Monte Carlo; Herwig is used to perform the shower evolution of both initial and final state partons.[1] Only the shapes of distributions are compared; no absolute normalization is used.

2 Directly Measured Kinematics

The goal of this section is to show that the W+ $\geq$ 3 Jets data agrees with Monte Carlo predictions for several variables using data sets with $t\bar{t}$ content varying from about 20% to 75%. This will test our understanding of both $t\bar{t}$ and QCD W+ $\geq$ 3 Jets production. Several variables are studied in order to see many projections of the complicated W+ $\geq$ 3 Jets system. In addition, using a large number of variables and data sets will minimize biases due to statistical fluctuations.

After the overview, results for two variables studied in more detail will be presented: H, the sum of all transverse energies, and the Relative Likelihood, a function of the transverse energies of the 2nd and 3rd highest energy jets.

2.1 Overview: The Data Sets

For the overview, four data sets have been chosen. All data sets require a charged lepton (electron or muon) with $P_T > 20$ GeV and missing transverse energy (MET), i.e. neutrino, greater than 20 GeV. The jets are ordered in transverse energy (E_T) and cuts are made on the jets. $E_T(2)$ refers to the transverse energy of the 2nd highest E_T jet. The four data sets have the following names and characteristics:

Data Set I:	Standard 3 Jet	3 Jets $E_T > 15$ GeV	170 events
Data Set II:	Mass Fitter	Jet4 $E_T > 8$ GeV	80 events
Data Set III:	High Thresh	Jet4 $E_T > 15$ GeV	30 events
Data Set IV:	SVX b-tags	Data Set I Cuts	16 events

The first 3 jets are also required to have a detector $\eta < 2.0$ and any 4th jet is required to have $\eta < 2.4$. In addition, Data Set III has a minimum dijet separation requirement in $\triangle\eta\triangle\phi$ space of 0.6. The Data Set I cuts are essentially the same as those used for the CDF b-tag analysis.[1]

Figure 1 shows the distribution of events in the variable $E_T(3) + E_T(4)$, the sum of transverse energies of the 3rd and 4th jet, for Data Set 2. This variable is expected to have good discriminating power between $t\bar{t}$ and QCD W+Jets events. Shown are the data distribution and Monte Carlo predictions for VECBOS alone and for a mixture of 30% $t\bar{t}$ and 70% VECBOS. Monte Carlo predictions are normalized to the data. The 30% $t\bar{t}$ is estimated from the number of events with an SVX b-tag in Data Set I subtracted for background and corrected for efficiency. Figure 1 shows that the data is consistent with the 30-70 mixture and is a poor match to VECBOS alone.

2.2 Overview: Definition of Quantities Plotted

The purpose of this overview is to give a relatively complete and balanced picture of the data. To facilitate examining many variables for the four data sets, a type of integral plot is used that shows deviations of the data from Monte Carlo predictions in units of statistical uncertainty.

A variable is chosen and a series of cuts are made on this variable. The horizontal axis is the fraction of TOP170 Monte Carlo events passing the cut, i.e. the efficiency of the cut for TOP170. This makes it easy to compare different variables for the same $t\bar{t}$ cut efficiency and does not overemphasize the tails of the distributions.

The vertical axis is the deviation of the fraction of events above the cut from the prediction of a VECBOS template. This deviation is in units of expected statistical uncertainty assuming the data is distributed like the VECBOS template. If the data is assumed distributed like the indicated mixture of $t\bar{t}$ and VECBOS, the expected statistical uncertainty is typically 1.0 to 1.5 vertical units (shown by error bars on selected data points). Remember that these are a type of integral plot, so the errors on adjacent points are correlated.[5]

Figure 2 is a pure Monte Carlo plot comparing the predictions for several variables for Data Set II assuming that there are 80 events of which 33% are TOP170. All the curves start at 0.86 because that is the efficiency of the Data Set II cuts for TOP170 (with respect to Data Set I). The solid circles are Aplanarity (ratio of sum P_2 out of the plane to sum P_2

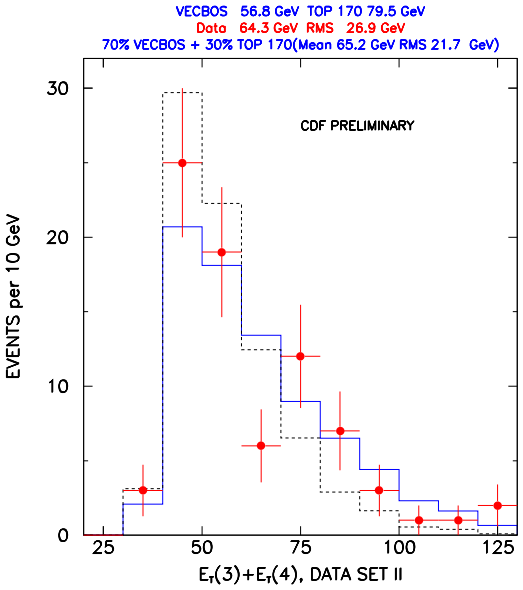

Figure 1: Distribution of events in $E_T(3)+E_T(4)$. Points are the data, dashed is VECBOS, and solid is 30% TOP170 and 70% VECBOS.

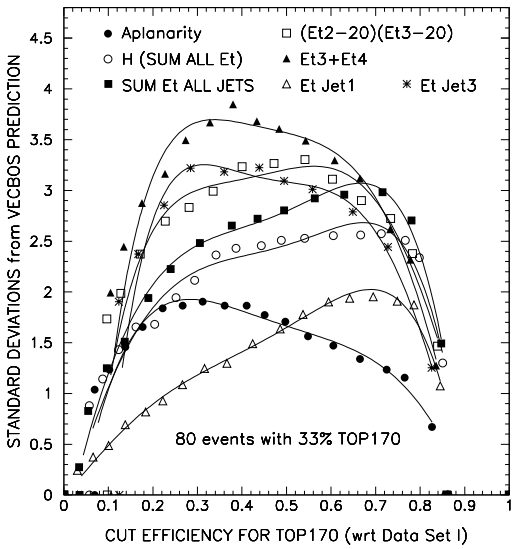

Figure 2: Deviation Plot comparing variables for Data Set II. This is a purely Monte Carlo plot using 33% TOP170.

in the plane for the plane with maximum sum P_2); this variable doesn't contain the energy scale and its discriminating power is modest. The open triangles are $E_T(1)$ and it is also modest in discriminating power. Somewhat better is H (Sum of all E_T's in the event with the neutrino and charged lepton added separately) which is the open circles. The solid squares are the sum of the E_T of all jets above threshold. The stars are $E_T(3)$. The open squares are $(E_T(2)-20)*(E_T(3)-20)$, a product of $E_T(2)$ and $E_T(3)$ minus roughly the threshold energy, which does slightly better than $E_T(2)+E_T(3)$. Finally, the solid triangles are $E_T(3)+E_T(4)$ which has the best predicted discriminating power. The conclusion to draw from this plot is that variables with the energy scale do better and those that emphasize $E_T(2)$, $E_T(3)$, and $E_T(4)$ do best. Adding in the charged lepton, the neutrino, and the $E_T(1)$ tends to dilute the discriminating power.

2.3 Overview: Description of the Data Plots

Data is shown in figures 3 to 8. Each figure compares one variable for the four data sets. The solid points with selected error bars are the data. The hatched band is the expectation for no $t\bar{t}$; the hatched width shows the limits of the VECBOS predictions for two extreme q^2 scales for $\alpha(s)$: $\langle P_T \rangle^2$ and M_W^2. The shaded band is the expectation with the indicated mixture of TOP170. The width of the shaded band represents the uncertainty in the q^2 scale for VECBOS plus any indicated uncertainty in the $t\bar{t}$ percentage. The striped band is the same expectation for TOP190.

The comparison for $(E_T(2)-20)*(E_T(3)-20)$ is shown in figure 3. The plot for Data Set I represents 170 events. The SVX b-tag rate predicts that 18% of these events are $t\bar{t}$. The band widths are due only to the uncertainty in q^2 scale for VECBOS; nothing is added for the

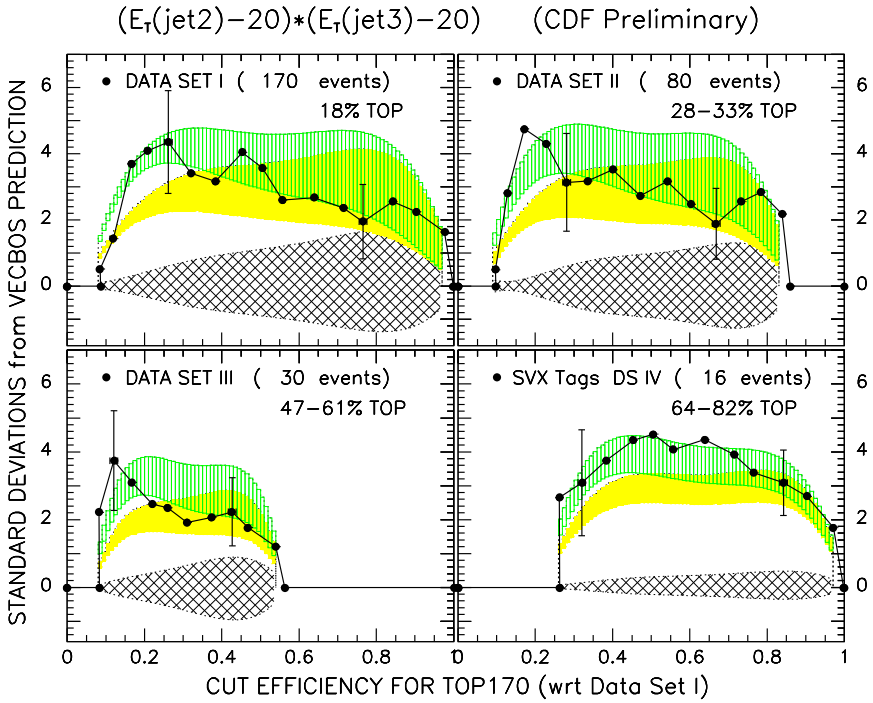

Figure 3: Deviations for $(E_T(2)–20)*(E_T(3)–20)$ in units of statistical uncertainty. Points are data and bands are MC predictions: hatched is pure VECBOS, shaded is the TOP170 mixture, and striped is the TOP190 mixture.

uncertainty in the 18%.

Data Set II has 80 events. Extrapolating from the 18% for Data Set I, the $t\bar{t}$ content should be 28-33% (depending on the extrapolation method and top mass). The band widths are due to the uncertainties in this $t\bar{t}$ extrapolation and the q^2 scale for VECBOS. Again nothing is added for the uncertainty in the 18%. Data Set III has 30 events with an expected $t\bar{t}$ content of 47-61% extrapolating from Data Set I. The band width contributions are the same as for Data Set II.

There are 16 SVX b-tagged events in Data Set IV. The $t\bar{t}$ content is 64-82% based on the calculated b-tag background and its error. This is the full uncertainty in $t\bar{t}$ content and is included with the q^2 scale systematics in the band widths. Note the power of the SVX b-tagging to reduce systematic errors: the band widths are narrower even though they include the full uncertainty in $t\bar{t}$ content.

One other point to note: below the cut efficiency where the VECBOS template predicts one event (0.5 event for Data Set IV), all data points and bands are set to zero. This happens at a larger cut efficiency for variables with more discriminating power.

2.4 Overview: Summary

Figures 3 to 9 are comparison plots for 7 variables: $(E_T(2)-20)*(E_T(3)-20)$, Aplanarity, $E_T(1)$, $E_T(3)+E_T(4)$, H (Sum All E_T), Mass W+4Jets, and $P_T(electron)$. The variables have

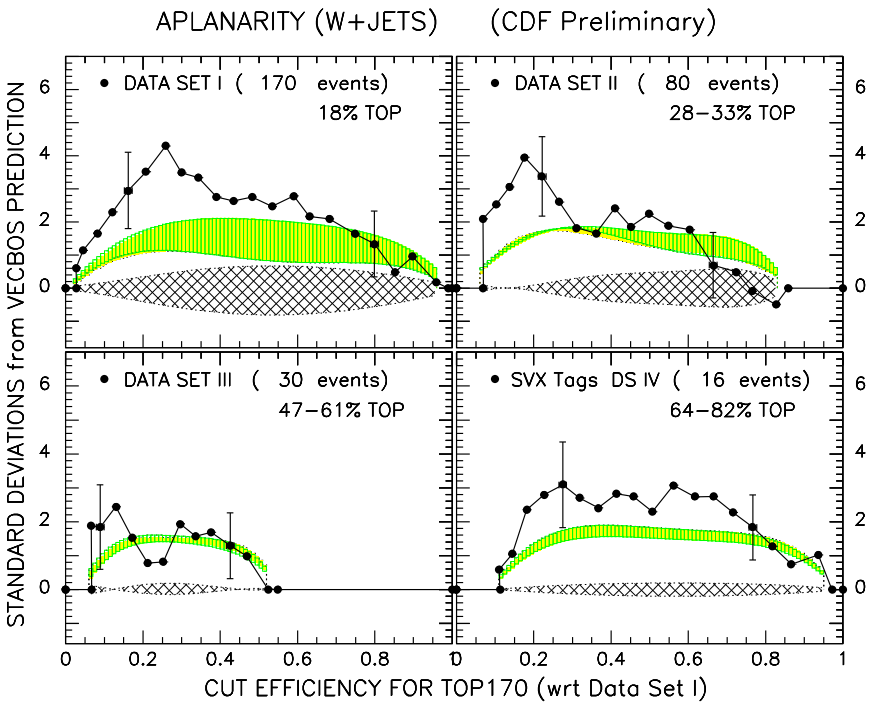

Figure 4: Deviations for Aplanarity in units of statistical uncertainty. Points are the data; bands are Monte Carlo predictions.

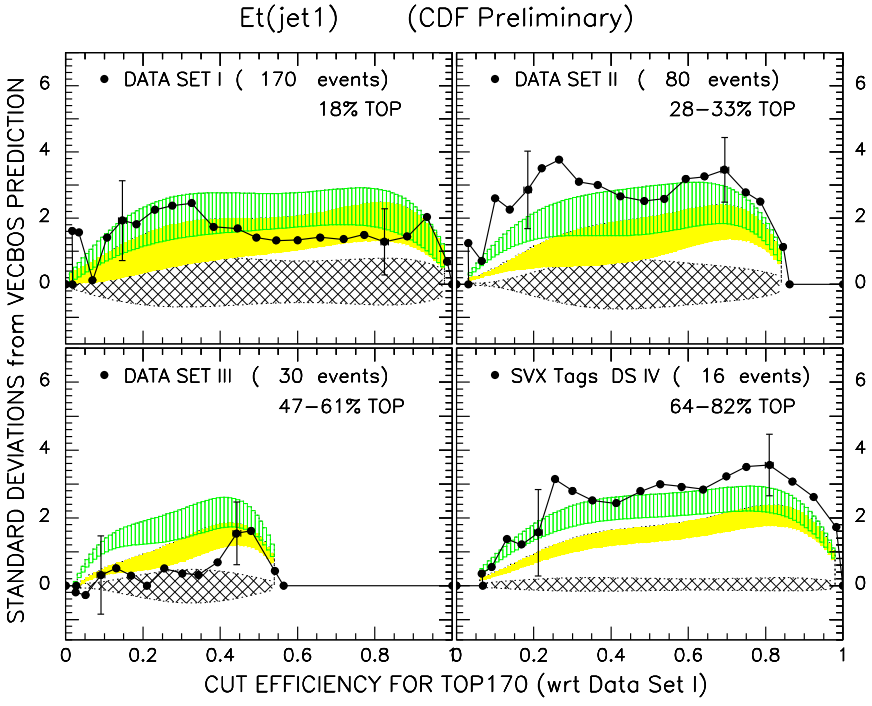

Figure 5: Deviations for $E_T(1)$ in units of statistical uncertainty. Points are the data; bands are Monte Carlo predictions.

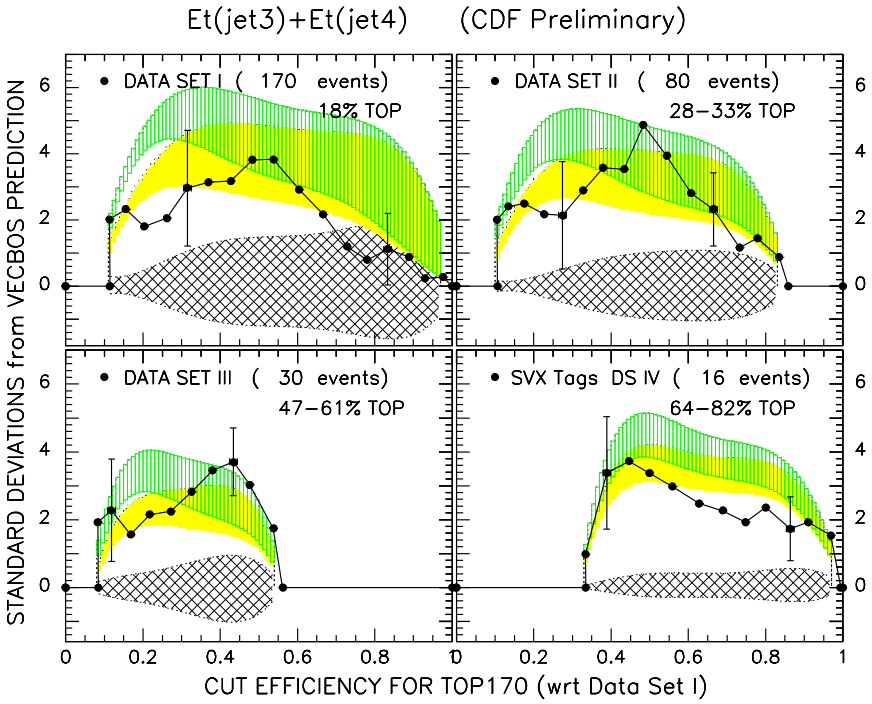

Figure 6: Deviations for $E_T(3) + E_T(4)$ in units of statistical uncertainty. Points are the data; bands are Monte Carlo predictions.

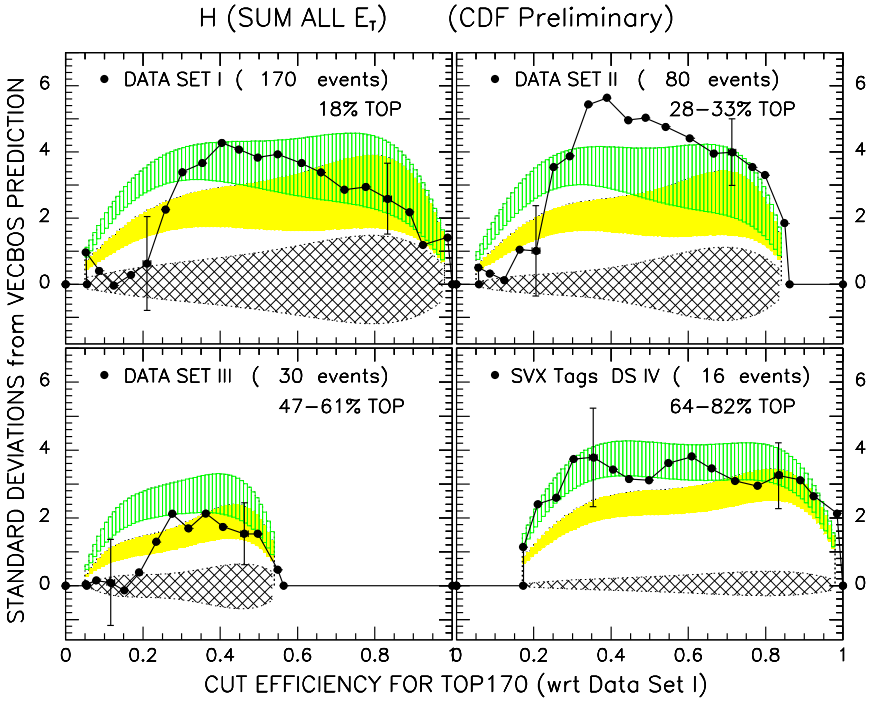

Figure 7: Deviations for H (Sum All E_T) in units of statistical uncertainty. Points are the data; bands are Monte Carlo predictions.

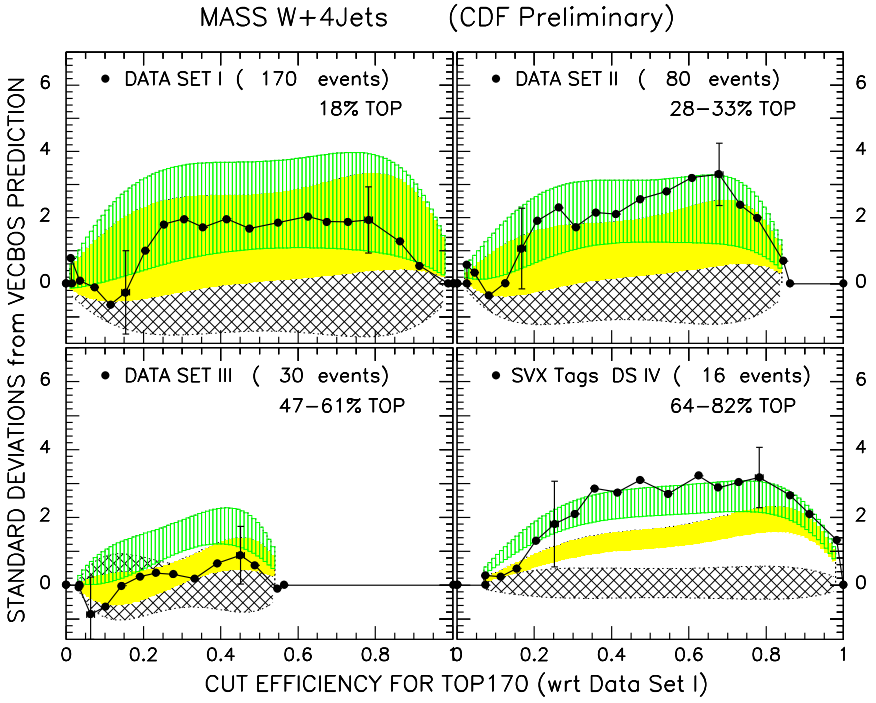

Figure 8: Deviations for MASS W+4Jets in units of statistical uncertainty. Points are the data; bands are Monte Carlo predictions.

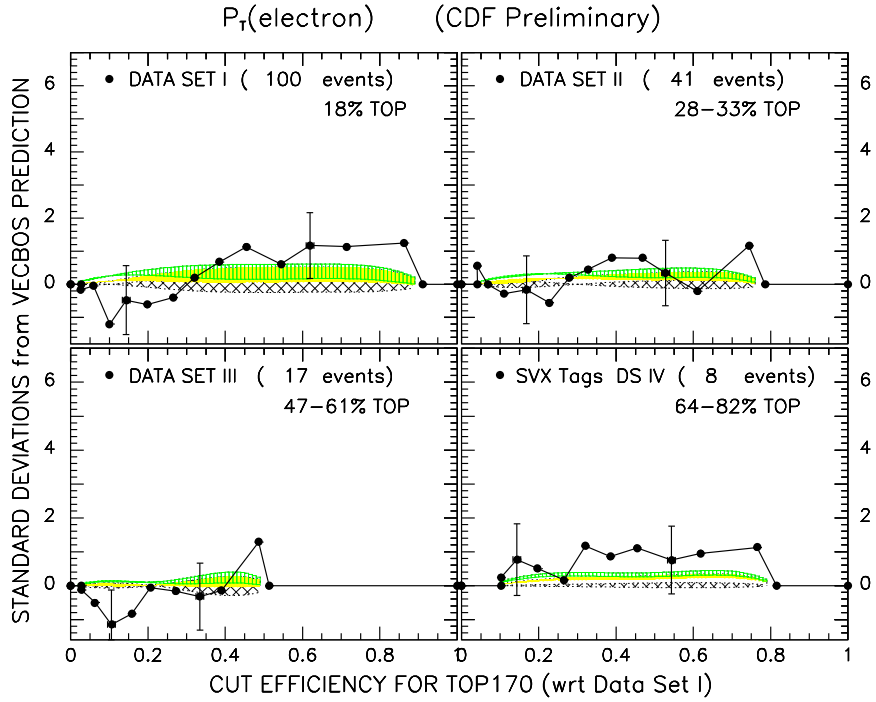

Figure 9: Deviations for P_T of the electron in units of statistical uncertainty. Points are the data; bands are Monte Carlo predictions.

been chosen to give a balanced picture of the kinematics: $E_T(1)$ for the higher energy jets, $(E_T(2)-20)*(E_T(3)-20)$ for the intermediate energy jets, and $E_T(3)+E_T(4)$ for the lower energy jets. In addition the H variable sums all E_T's in the event, Aplanarity is an event shape variable, Mass W+4Jets includes longitudinal energy, and $P_T(electron)$ is a variable with little dependence on the jet energy scale. Most of the data points are within one sigma of the expectation of VECBOS plus $t\bar{t}$ with top mass of 170-190 GeV; the rest are within two sigma.

Mass W+4Jets (figure 8), nominally the $t\bar{t}$ invariant mass, is the only variable with longitudinal energy. Since VECBOS type events can have large longitudinal energies with large uncertainties, the bands for this variable are wide and its discriminating power between VECBOS and $t\bar{t}$ is diluted.

Aplanarity (figure 4) is the only variable without direct dependence on the energy scale. For this reason, the Monte Carlo predictions for the two different top masses are almost identical. Another point of interest is that the distribution of $t\bar{t}$ events in aplanarity is closer to the distribution of VECBOS events with $q^2 = \langle P_T \rangle^2$ than to the distribution of VECBOS events with $q^2 = M_W^2$. This means that the upper edges of the bands in figure 4 correspond to the $q^2 = \langle P_T \rangle^2$ prediction while in the rest of the plots the upper edges of the bands correspond to $q^2 = M_W^2$.

Note that Data Set III, which is rich in $t\bar{t}$ events and avoids the larger systematic uncertainties associated with lower energy jets and less well separated jets, has good agreement with Monte Carlo predictions. And Data Set IV, the SVX b-tags, which is richer $t\bar{t}$ events and minimizes the uncertainties due to the q^2 scale for VECBOS also has agreement with Monte Carlo predictions.

Many other variables have been studied. These include: $E_T(2)$, $E_T(4)$, MET, $P_T(muon)$, $cos(\theta_J^*)_{max}$, $P_T(W)$, Circularity, M_{JJ}^{min}, η_J^{max}, and $\sum E_T(jets)$. They are all consistent with the Monte Carlo expectations. The conclusion is that the data agrees with Monte Carlo predictions using the $t\bar{t}$ content indicated by the SVX tagging rate.

2.5 Detailed Analysis Using the H Variable

H, the sum of all the E_T's in the event, is a simple, global variable that is approximately the transverse mass of the $t\bar{t}$ system. Its good correlation with the Top Mass provides a check on the top mass and number of $t\bar{t}$ events. The analysis includes a thorough study of systematics.[2]

Data Set II cuts, requiring $E_T(4) > 8$ GeV, define a signal sample. There are 99 events, a few more than for the Overview Plots, primarily because an extra trigger path is allowed and there is slightly different background removal. Control samples, with low $t\bar{t}$ content for studying systematics, are defined by vetoing events having $E_T(4) > 8$ GeV. The low threshold control sample requires the $E_T > 8$ GeV for the first three jets; the high threshold control sample has the same 15 GeV requirement as the signal sample.

For the low threshold control sample, VECBOS with $q^2 = \langle P_T \rangle^2$ reproduces the turn-on of H better than VECBOS with $q^2 = M_W^2$; however, systematic uncertainties in the jet energy scale could explain much of the difference. For the high threshold control sample and signal sample, the agreement is equally good using the two different q^2 scales for VECBOS.

The distribution of the 99 signal sample events is shown in figure 10. Also shown is the best

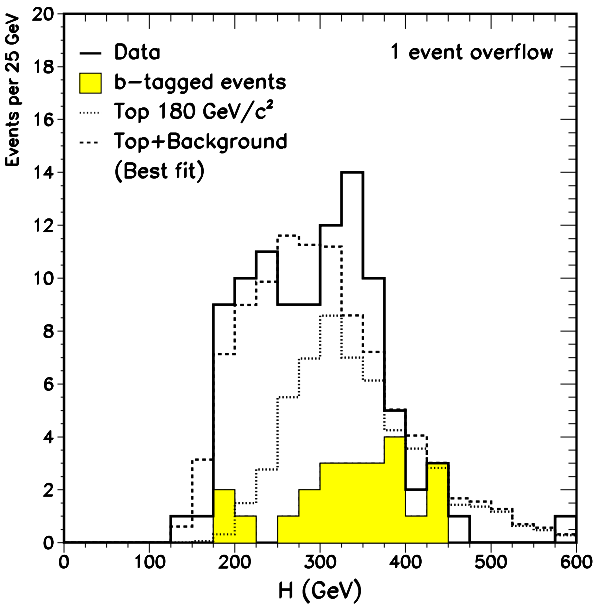

Figure 10: Distributions in H: solid is data, dashed is best fit, and dotted is TOP180 component of best fit.

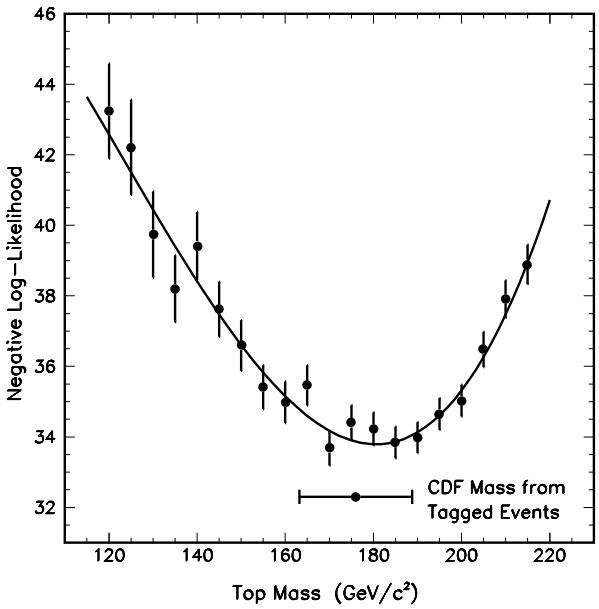

Figure 11: Negative log-likelihood for the fit of the data to VECBOS plus $t\bar{t}$ vs top mass.

fit to the data of a mixture of $t\bar{t}$ and VECBOS with $q^2 = \langle P_T \rangle^2$. The Data Set II overview plot for H in figure 7 shows the same behavior: the data rises above the Monte Carlo prediction as the cut in H is hardened (i.e., as cut efficiency decreases) and then falls below the prediction for the hardest cuts (the high H tail).

Figure 11 shows the negative log-likelihood for the two component fit as a function of the top mass. The best fit is $180 \pm 12(stat.)^{+19}_{-15}(sys.)$ GeV top mass and $56 \pm 10 \pm 5$ $t\bar{t}$ events. Using $q^2 = M_W^2$ for VECBOS gives $184 \pm 15^{+19}_{-15}$ GeV and $45 \pm 11 \pm 5$ events. The masses are consistent within errors to the reported top mass of 176 GeV[1]; the number of $t\bar{t}$ events is consistent with the $34 \pm 10 \pm 5$ expected from the SVX b-tag rate in Data Set I.

2.6 Kinematic Separation Using the Relative Likelihood

The Relative Likelihood (L_{rel}) variable is a function of $E_T(2)$ and $E_T(3)$ with good discriminating power. L_{rel} was used for previously published Run 1A results and it doesn't use $E_T(4)$ which has larger systematic uncertainties.[3][4] It is similar to the overview variable $(E_T(2)-20)*(E_T(3)-20)$ described earlier. The exact functional form of L_{rel} is as follows:

$$L_{rel} = \left[\left(\frac{1}{\sigma}\frac{d\sigma^{t\bar{t}}}{dE_T(2)}\right) \Big/ \left(\frac{1}{\sigma}\frac{d\sigma^{QCD}}{dE_T(2)}\right)\right] * \left[\left(\frac{1}{\sigma}\frac{d\sigma^{t\bar{t}}}{dE_T(3)}\right) \Big/ \left(\frac{1}{\sigma}\frac{d\sigma^{QCD}}{dE_T(3)}\right)\right]$$

Events with $ln(L_{rel}) < 0$ are more likely to be QCD $W + Jets$ and those with $ln(L_{rel}) > 0$ are more likely to be $t\bar{t}$.

The data set for this analysis is similar to Data Set I, except the jet E_T threshold is a

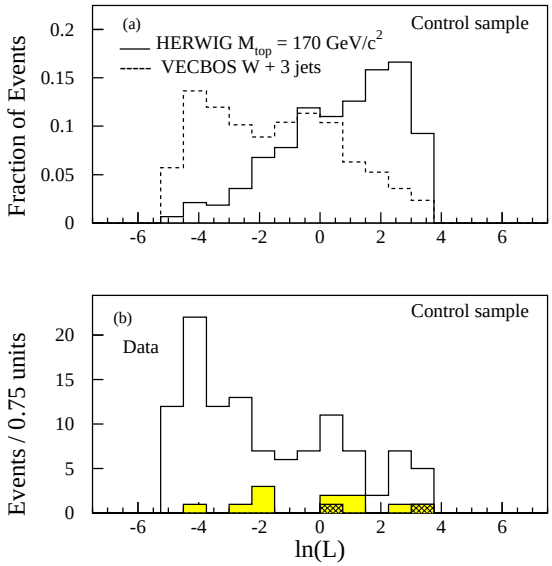

Figure 12: Control Sample: The top plot shows Monte Carlo expectations and the bottom plot the data (b-tags are shaded).

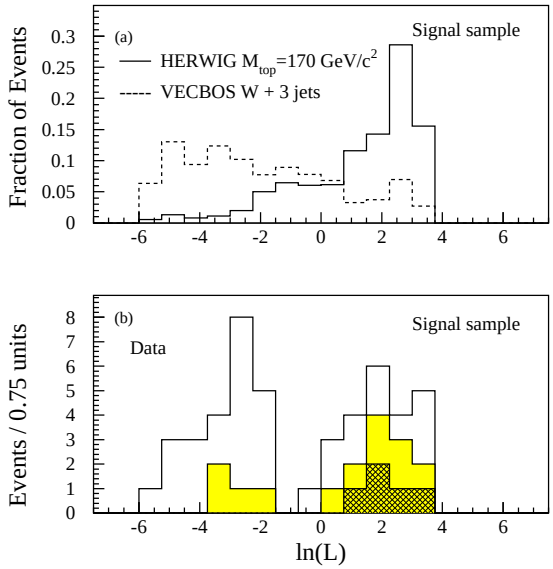

Figure 13: Signal Sample: The top plot shows Monte Carlo expectations and the bottom plot the data (b-tags are shaded).

little lower giving a longer lever arm and more resolving power. Also there is a minimum dijet separation requirement and differences in the MET cut. The result is a total of 158 events.

The maximum $cos(\theta)$ in the center of mass for the three highest E_T jets divides the data set into a 'Signal Sample' and a 'Control Sample'. The 'Signal Sample' with $cos(\theta_j^*)_{max} < 0.7$ is more central and richer in $t\bar{t}$. Figure 12 shows the 111 event 'Control Sample' with $cos(\theta_j^*)_{max} > 0.7$. It should have about 75% of the VECBOS type events (QCD W+Jets) and 50% of the $t\bar{t}$ events. The top plot shows the Monte Carlo predictions and the bottom plot shows the data with the b-tags shaded. The 'Control Sample' should be mostly VECBOS type events and the data shape is consistent with this.

The 47 events in the 'Signal Sample' should contain about 50% of the $t\bar{t}$ events and only 25% of the VECBOS type events. The shape of the data shown in figure 13 is much flatter than that predicted by VECBOS and is consistent with significant $t\bar{t}$ content. The tagged events are clustered at $ln(L_{rel}) > 0$ confirming that the L_{rel} variable is good at identifying $t\bar{t}$ events. There are 22 events with $ln(L_{rel}) > 0$ and the probability that this is due to a statistical fluctuation of the VECBOS distribution is conservatively estimated at 0.0026. This probability was calculated using the worst case $q^2 = M_W^2$ for VECBOS and the worst case jet energy scale.

3 Kinematics Requiring a Mass Fit

With the existence of the top quark established, $t\bar{t}$ production properties are of great interest. To study them, jet assignments are made using a mass fit. The constraints of the mass fit also improve the resolution of kinematic quantities. However, the mass fit is not perfect because of combinatorics and gluon radiation, so the data must be compared to Monte Carlo events that have undergone the same analysis.

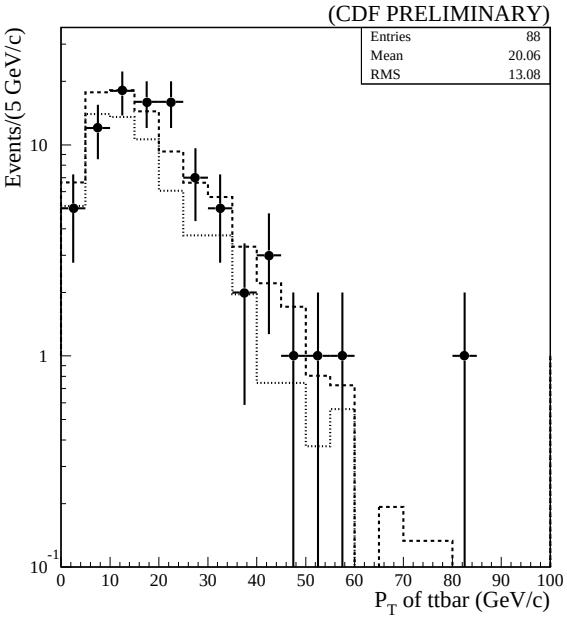

Figure 14: P_T of the $t\bar{t}$ system. Points are the 88 data events, dotted is VECBOS, and dashed is VECBOS+TOP175.

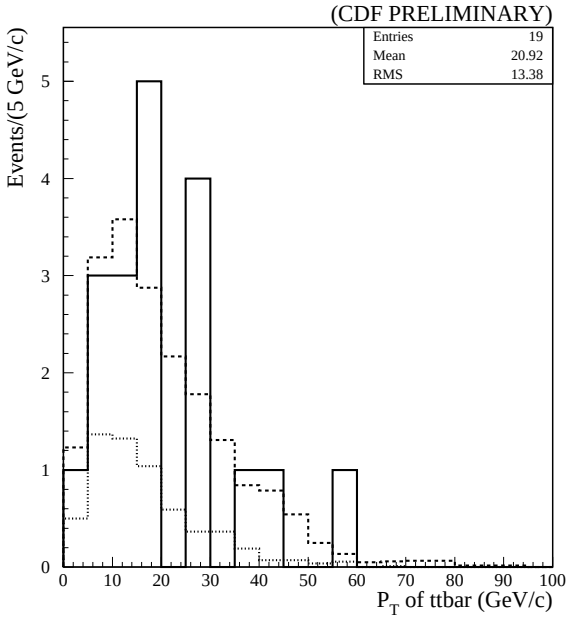

Figure 15: P_T of the $t\bar{t}$ system. Solid histogram is the 19 b-tags, dotted is VECBOS, and dashed is VECBOS+TOP175.

The data set is same as that used for the H analysis (Data Set II cuts). There are 99 events and 88 of these have a mass fit with a satisfactory chisquare$<$ 10. Requiring either an SVX or a 'soft lepton' b-tag leaves 23 event of which 19 have a satisfactory mass fit.

Figure 14 shows the P_T of the $t\bar{t}$ system for the 88 events with a good mass fit. Figure 15 shows the same distribution for the 19 b-tagged events. For the b-tagged events, the tagged jet is constrained to be a b-jet. The agreement is good for both plots.

3.1 Mass of $t\bar{t}$ System.

A quantity that has generated more interest is the mass of the $t\bar{t}$ system. Figure 16 shows the $t\bar{t}$ mass for the 88 events with a good mass fit. Fig 17 shows the $t\bar{t}$ mass for the 19 b-tags. These plots are consistent with Monte Carlo predictions within the statistical accuracy of the data.

There are theoretical papers about possible structure in the $t\bar{t}$ mass. One conjecture is that there is a technicolor Z' that decays to $t\bar{t}$ that would help explain why the measured cross-section is larger that the standard model prediction.[6] Also some people have noticed a small, but not statistically significant, excess of events in the region of 500 GeV in figure 17.

By constraining the top mass in the mass fit to the best measured value of 176 GeV, the resolution of the $t\bar{t}$ mass is improved in the region of a peak by about a factor of 2 according to Monte Carlo studies. Even if the top mass of 176 GeV is not correct, there is still an improvement in resolution with the $t\bar{t}$ mass peak shifted by twice the error in the top mass. Figure 18 shows the prediction for the $t\bar{t}$ mass spectrum with the top mass constraint for

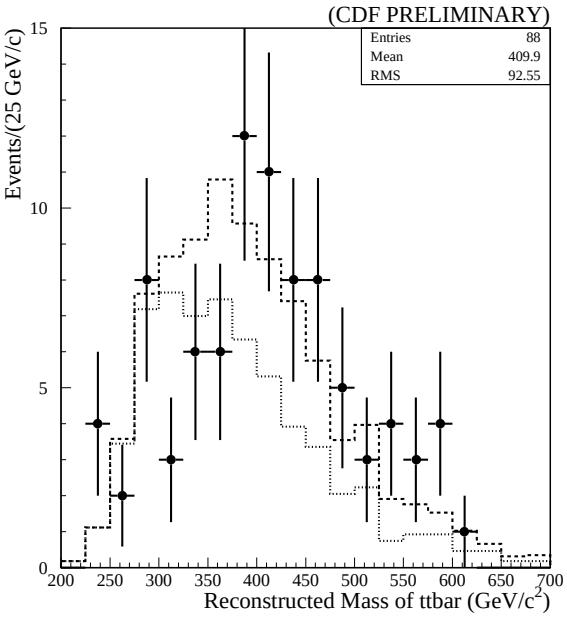

Figure 16: The $t\bar{t}$ mass. Points are the 88 data events, dotted is VECBOS, and dashed is VECBOS+TOP175.

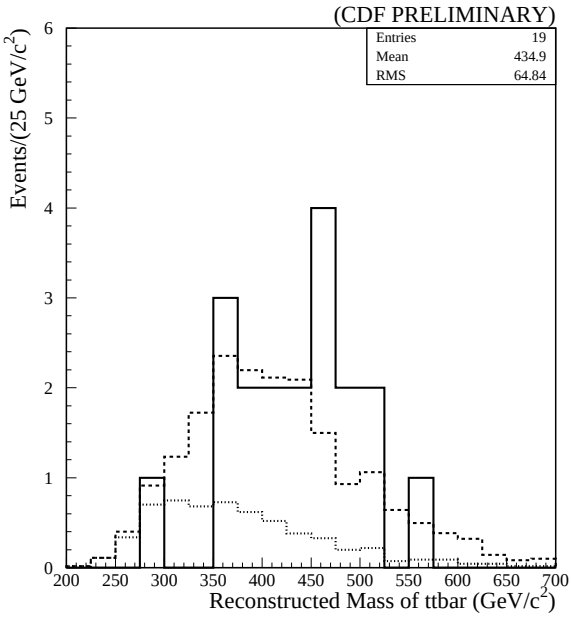

Figure 17: The $t\bar{t}$ mass. The solid histogram is the 19 b-tags, dotted is VECBOS, and dashed is VECBOS+TOP175.

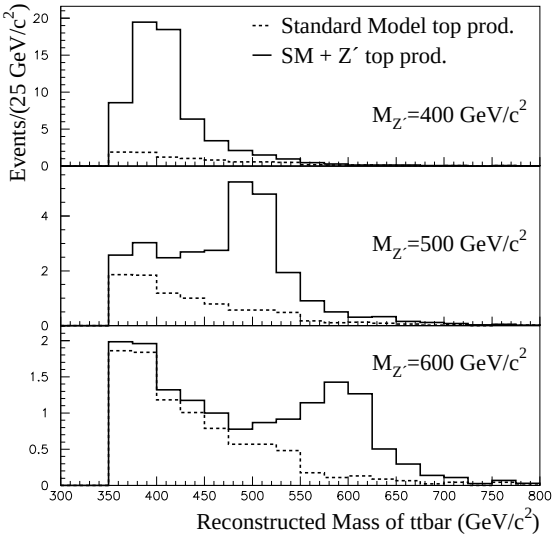

Figure 18: Predicted $t\bar{t}$ mass distributions reconstructed with top mass constraint. Standard model and Z' top production are shown (175 GeV top).

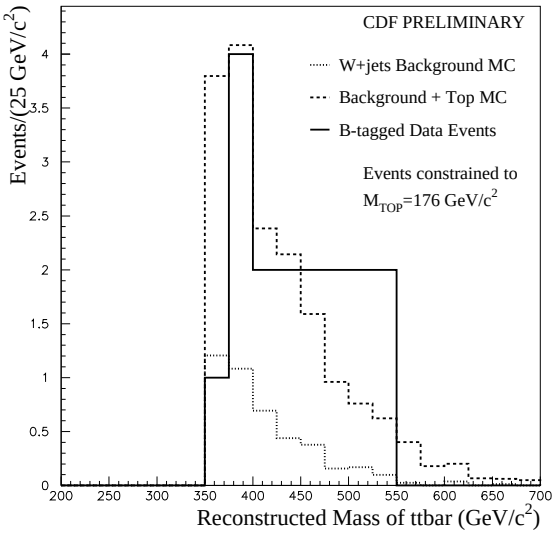

Figure 19: The $t\bar{t}$ mass with the top mass constrained to 176 GeV. The solid histogram is the 17 b-tags, dotted is VECBOS, and dashed is VECBOS+TOP175.

standard model top alone and for a mixture with 400, 500, and 600 GeV Z'. The contribution of the QCD W+Jets background is not included but should only be about 25%. The 17 events with a good mass fit using a 176 GeV top mass constraint are compared to the expected standard model contributions in figure 19. The data is still consistent with the standard model.

Another look at the $t\bar{t}$ mass is given by figure 8. In particular, Data Set III has the advantage of being rich in $t\bar{t}$ and of having only events with better jet separation and higher $E_T(4)$. Although the mass resolution is worse, better defined jets and additional $t\bar{t}$ events not in figure 19 make Data Set III interesting. Note that a 500 GeV $t\bar{t}$ mass corresponds to a 0.11 cut efficiency for Data Set III (increasing to 0.18 for Data Set I). Again, the agreement of the data in figure 8 with the predictions for standard model top is good.

4 Summary

Currently the data seems consistent with the standard model with the top quark. By the end of the present run we should accumulate another 60 pb^{-1} of data. We should have a better understanding of the systematics and the data will provide more constraints on this understanding. The emphasis will change to studying the production variables for top. We should soon be showing variables such as $P_T(top)$, $\eta(top)$, and the polarization of the W.

References

[1] CDF Collaboration, F. Abe *et al.*, PRD **43**, 664 (1994). CDF Collaboration, F. Abe *et al.*, PRL **74**, 2626 (1995).

[2] CDF Collaboration, F. Abe *et al.*, "Study of $t\bar{t}$ Production in $p\bar{p}$ Collisions Using Total Transverse Energy" (submitted to PRL).

[3] CDF Collaboration, F. Abe *et al.*, PRD **51**, 4623 (1995).

[4] CDF Collaboration, F. Abe *et al.*, "Identification of Top Quarks at CDF using Kinematic Variables" (submitted to PRL).

[5] The vertical axis is $(f_d - f_{vb})/\sigma$ where f_d and f_{vb} and the fractions of data and VECBOS above the cut and σ is $\sqrt{(f_{vb} + 1/n)(1 - f_{vb} + 1/n)/n}$ where n is the number of events. In the limit of large n, σ becomes the more familiar $\sqrt{f_{vb}(1 - f_{vb})/n}$.

[6] C. T. Hill, Phys.Let. **B345**, 489 (1995). K. Lane and E. Eichten, Fermilab-Pub-95/052-T (1995).