Fermi National Accelerator Laboratory

FERMILAB-Conf-95/060-E

CDF

Kinematic Top Analyses at CDF

Hans Grassmann
For the CDF Collaboration

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

March 1995

Submitted to the proceedings of the *Rencontres de Physique de la Vallee' d'Aoste*, La Thuile, Italy, March 1995

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

Disclaimer

Kinematic Top Analyses at CDF

Hans Grassmann, for the CDF Collaboration

Fermilab

Abstract

We present an update of the top quark analysis using kinematic techniques in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV with the Collider Detector at Fermilab (CDF). We reported before on a study which used 19.3 pb^{-1} of data from the 1992-93 collider run, but now we use a larger data sample of 67 pb^{-1}. First, we analyse the total transverse energy of the hard collision in W+$\geq$3 jet events, showing the likely presence of a $t\bar{t}$ component in the event sample. Next, we compare in more detail the kinematic structure of $W+\geq 3$ jet events with expectations for top pair production and with background processes, predominantly direct $W+$ jet production. We again find $W+\geq 3$ jet events which cannot be explained in terms of background, but show kinematic features as expected from top. These events also show evidence for beauty quarks, in agreement with expectations from top, but not compatible with expectations from backgrounds. The findings confirm the observation of top events made earlier in the data of the 1992-93 collider run.

1 Introduction

Kinematic Analysis

At the $p\bar{p}$ collider, the top quark is predominantly produced in $t\bar{t}$ pairs, which in turn decay to W's and beauty quarks. In this paper, we shall use "single lepton" events for our study, that are events where one of the W's decays leptonically and the other one hadronically : $p\bar{p} \rightarrow t\bar{t} \rightarrow W^+bW^-\bar{b} \rightarrow l^{\pm}\nu bjj\bar{b}$; two light quarks from the second W and two beauty quarks can create hadronic jets. There may also be jets from associated gluon radiation. Additional free parameters are for instance the beauty content of the entire event, or of each of its jets. To identify the top particle, one would like to study as many of these free parameters as possible, ideally all of them. With the present data statistics, we have about as many top candidate events as free parameters to be measured. We therefore begin with a study of one parameter, the H analysis, and proceed next to a study involving several parameters, the event structure analysis.

Data sample

The data was collected in two running periods. In 1992-93 (run 1A) we integrated a luminosity of 19.3 pb^{-1}. With the run presently in progress (run 1B) we have collected in addition 48 pb^{-1}. At CDF, both a kinematics study, referred to as "event structure analysis" [1], [2], and a counting experiment [3] were performed with the data from run 1A. The observation of top events, made in these analyses was confirmed by an improved version of the counting experiment, using the full 1A+1B data sample [4]. The D0 Collaboration has also observed the production of top quark pairs in an analysis based in part on kinematic information [5].
Also for this paper we use the data sample available from run 1A+1B. In Section 2 we report on a study of the total transverse energy of W+$\geq$3 jet events, which leads to evidence for a $t\bar{t}$ component in the data. In Section 3 we report the results of the event structure analysis. A detailed description of the CDF detector can be found in ref. [3].

Monte Carlo Predictions

The expected $W+$ jets background is computed with the Vecbos Monte Carlo, which uses the lowest order W+n parton matrix elements, n=1,4 [6]. Since Vecbos results in inclusive predictions, to compare to W+$\geq$ n jet data events W+n parton Vecbos events can be

used. The parton fragmentation process can be simulated either with an independent fragmentation model, or with a Herwig type shower module [7]. For this paper we use the Herwig type fragmentation module. The lowest order matrix elements are sensitive to the choice of the mass scale in the strong coupling constant α_s. We assume $q^2 = {M_W}^2$, that means α_s = constant, as a conservative choice which results in rather hard Et(jet) spectra. Top events are simulated by the Herwig program. Isajet and Pythia simulated top events give for our analysis essentially the same results.
A more detailed desciption of the Monte Carlo programs can be found in ref. [1], [2] and in references therein.

2 H Analysis

Leptonically decaying W's are selected requiring an isolated, high pt lepton and missing transverse energy ($\not{E}_T$). In the H analysis, three W+jet samples are defined :
(a) an exclusive W+3 jet sample consisting of events which contain precisely three jets with Et(jet) > 8 GeV.
(b) a 'low threshold' sample of events with at least four jets with Et(jet) > 8 GeV.
(c) a 'high threshold' sample which is a subsample of (b), requiring in addition that the three leading jets have Et(jet) > 15 GeV.
Sample (a) is used as a background enriched sample since Monte Carlo calculations predict that it should contain a fraction of not more than 1% top events ($M^{top} = 175$ GeV), due to the bias against a fourth jet. The low and high threshold samples,(b) and (c), should be increasingly enriched with top events. We define H as H=ΣEt(final state), where the final state includes the charged lepton, the neutrino ($\not{E}_T$) and all jets. For the event selection neither Et(jet) nor $\not{E}_T$ are corrected for detector effects. However, when calculating H, jet energies and $\not{E}_T$ are corrected by a rapidity and energy dependent factor, which accounts for calorimeter non-linearity and reduced response at detector boundaries [10]. This correction is applied in order to reconstruct the original parton energy as closely as possible. This analysis approach is also discussed in [8]. We note that the D0 collaboration has used a similar variable for their analysis [5].
In figure 1 we compare the data distributions in H to the expectation from direct QCD W+jets production, as calculated by the Vecbos Monte Carlo program. In the W+3 jet (a) sample we find a good agreement between QCD prediction and data. In sample (b) we find an excess of data events at large H; this excess becomes more pronounced in sample (c).

CDF is able to detect b-quarks with two algoritms. One algorithm identifies charged tracks originating from a secondary vertex, separated in space from the primary one ("SVX tag"). The other algorithm identifies electrons or muons in jets (soft lepton tag, or "SLT" [3].) Figure 2 shows the H distibution in sample (c) together with the distribution of the b-tagged events. The 'double peak' structure of the H distribution and the presence of b-tagged events at high values of H suggests very clearly the presence of top events in the sample.

We performed a two-component fit to the H distribution of sample (c) using the predictions from the QCD and top Monte Carlo programs. In figure 3 we show the fit likelihood as a function of the assumed value for M^{top}. One sees that this simple procedure is quite sensitive to M^{top}. Using the rate of b-tagged events (fig. 2) and the SVX tagging efficiency we find that there should be about 35 top events in the sample. A two component fit to the observed H distribution, using the Vecbos and the top predictions, results in 57 (46) top events when we assume $Q^2 = <pt>^2$ ($M_W{}^2$) for the Vecbos calculation.

We conclude that the top content of the sample is subject to significant systematic uncertainties. These are presently under study. Once these studies are completed, the H distribution could result in a measurement of M^{top}.

3 Event Structure Analysis

The event selection and the analysis procedure is as described in [2], except that we now include data events passing a large $\not{E}_T$ trigger, even if the primary lepton did not cause a trigger. In brief, we use tight cuts to select (leptonic) W candidate events, we require $\not{E}_T > 25$ GeV, and we cut on the transverse mass, $M_T > 40$ GeV/c^2 [1]. Next, the events need to contain at least three large transverse energy jets with $E_T(\text{jet}) > 20$ GeV. Jets are reconstructed with a cone of $R = 0.4$ [2]. Jet energies are corrected [10] already before event selection. The three leading jets are required to be separated from each other by $\Delta R \geq 0.7$. We also require $|\eta(\text{jet})| < 2.0$ [3]. This W+$\geq$3 jet event sample contains 158 events from the total available luminosity of 67 pb^{-1}.

Jets from $t\bar{t}$ decay are expected to be produced at larger angles than those from QCD $W+$ jet events. Therefore we select a top enriched sample ("signal sample") of $W+ \geq 3$

[1]The transverse mass is defined as $M_T = [2E_T \not{E}_T(1\text{-}\cos\Delta\phi)]^{1/2}$, where $\Delta\phi$ is the difference in azimuthal angle between the missing energy direction and the direction of the lepton.

[2]$R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$, where $\Delta\phi$ is the cone half-width in azimuth and $\Delta\eta$ is the cone half-width in pseudorapidity.

[3]$\eta = -\ln\tan(\theta/2)$, and θ is the polar angle with respect to the beam.

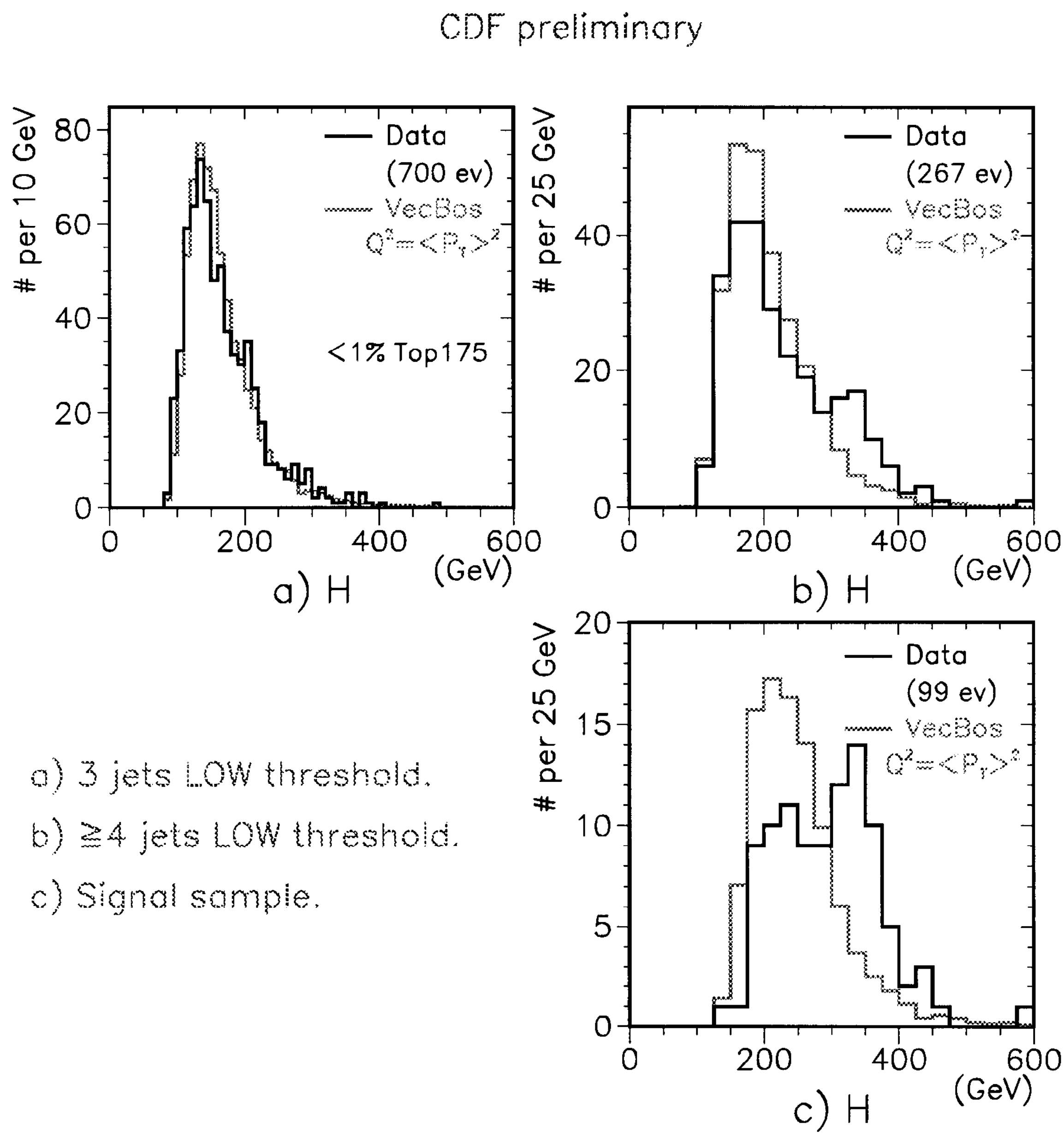

Figure 1: H distribution for data (solid line) and Vecbos events (dotted). Vecbos is normalised to the data. (a) Exclusive W+3 jet sample with E_T(jet) > 8 GeV; (b)W+≥4 jet sample with E_T(jets) > 8 GeV; (c)W+≥4 jet sample with E_T(jet1,2,3) > 15 GeV, E_T(jet4) > 8 GeV.

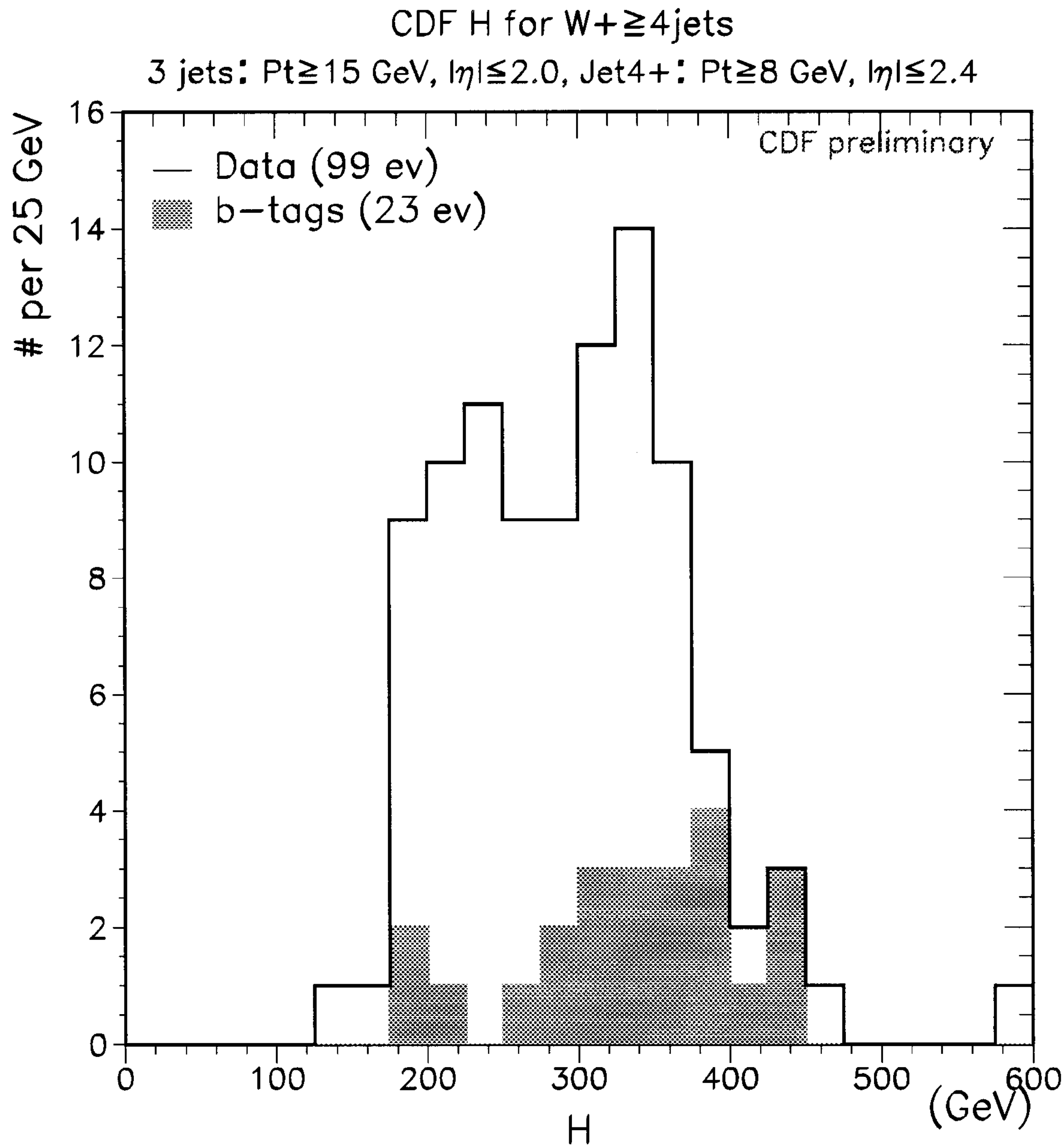

Figure 2: H distribution of events of the high threshold W+$\geq$4 jets sample, and of the associated b-tags (SVX or SLT, shaded).

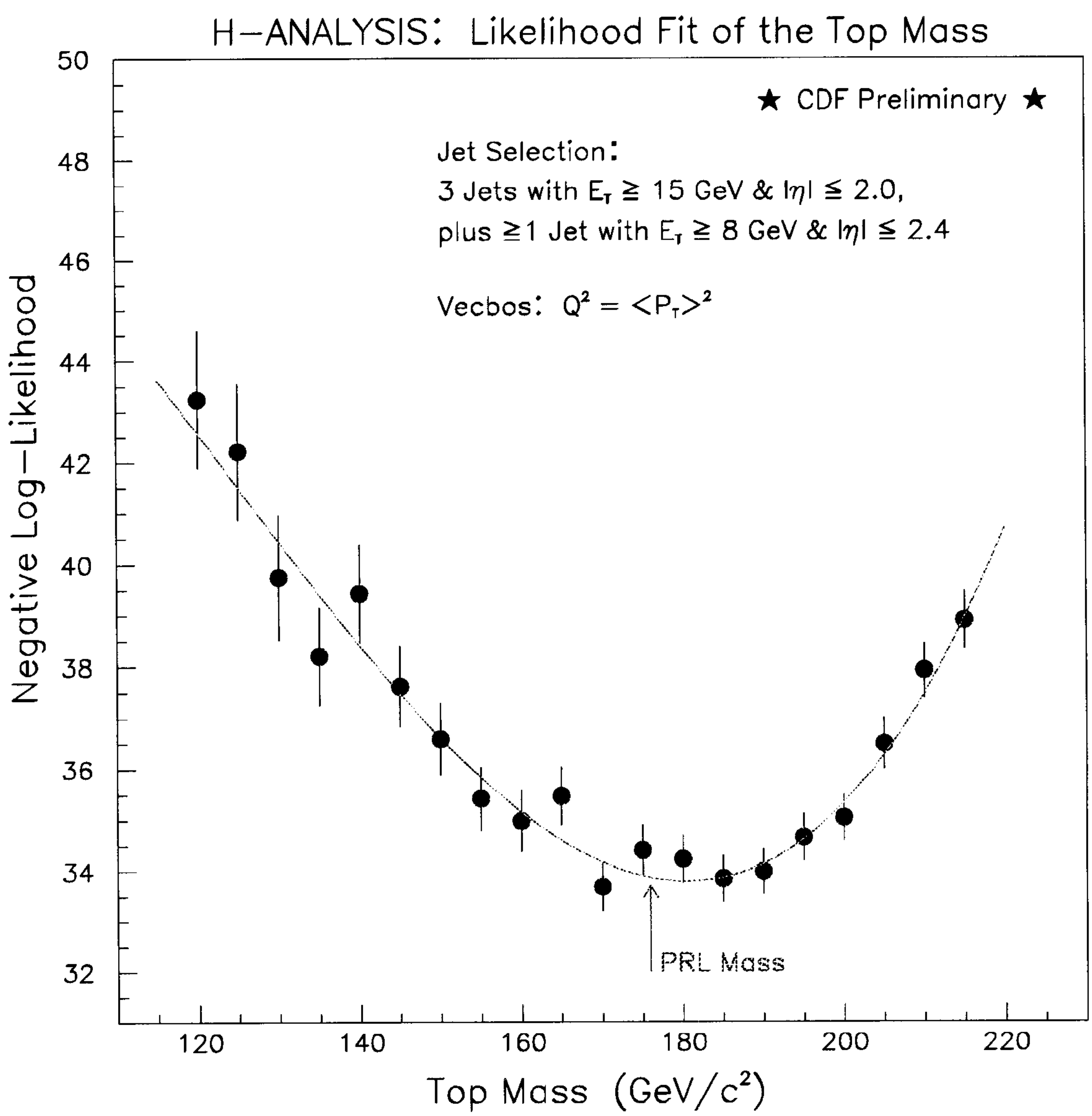

Figure 3: Least-squares fit of a cubic polynomial through a set of negative log-likelihood points versus the top mass. Each point was obtained by fitting a linear combination of Vecbos W+$\geq$4 jets and Herwig $t\bar{t}$ to CDF data passing the high threshold selection (sample (c)). The arrow indicates the value of M^{top} reported in ref.[4].

jet events by requiring the three jets with highest E_T to have $|\cos\theta^*(\text{jet})| < 0.7$, where θ^* is the jet polar angle in the rest system of the event. The system is defined as the sum of the charged lepton, the transverse missing E_T and the jets with $E_T > 15$ GeV. The remaining events in which at least one of the jets has $|\cos\theta^*(\text{jet})| > 0.7$ form a complementary background enriched sample ("control sample"). This procedure was suggested in ref. [9].
Monte Carlo calculations predict that both the cuts $\Delta R > 0.7$ and $|\cos\theta^*(\text{jet})| < 0.7$ would enrich the event sample with top events. This is because QCD W+jets events often show the structure which is characteristic of final state gluon radiation (jets close together) or initial state radiation (jets in the forward direction). One might also use the events at $\Delta R < 0.7$ as a control sample. However, jets which are close together are not yet studied very well. There might be large experimental and theoretical uncertainties. Also, our jet correction routines do not explicitely take into account the overlap of jet energy flow when jets are close together in space.
In figures 4, 5 we show how Monte Carlo simulated QCD and top events and data events are distributed in E_{T2}, E_{T3} with the cuts of the signal sample and control sample, respectively [4]. In the signal sample the data is not in agreement with the Vecbos prediction alone. Figures 4(d), 5(d) display the b–tagged data events. The b–tagged events of the signal sample are mostly found in the region where top events are expected.

For a quantitative comparison of data and Vecbos we calculate a "relative likelihood" for each event, as a measure of whether the event is more "top-like" or more "QCD-like". The relative likelihood is defined in terms of the Monte Carlo predicted jet E_T distributions $d\sigma/dE_T$ of the second and third highest E_T jets, for $t\bar{t}$ (M^{top}=170 GeV) and direct W+jets production. The cross sections are normalised to 1.

$$L = [(\frac{1}{\sigma}\frac{d\sigma}{dE_{T2}}) \times (\frac{1}{\sigma}\frac{d\sigma}{dE_{T3}})]^{t\bar{t}} / [(\frac{1}{\sigma}\frac{d\sigma}{dE_{T2}}) \times (\frac{1}{\sigma}\frac{d\sigma}{dE_{T3}})]^{QCD} \quad (1)$$

In figure 6(a),(b) we show the signal sample ln(L) distributions for Monte Carlo events and data events, respectively. There are 25 events at $\ln(L) < 0$ and 22 events at $\ln(L) > 0$. The distribution of events between positive and negative ln(L) is similar to that observed in the data of Run 1A [1]. The QCD Monte Carlo predicts that not more than 22±5% of QCD $W+$ jet events will be at $\ln(L) > 0$. We have evaluated other backgrounds, such as non-W and WW events. The estimated number of these events in the signal sample is 6.9±1.8. These background events are expected to have softer jet E_T distributions for

[4] Jets are ordered in E_T(jet) with $E_{T1} > E_{T2} > E_{T3}$.

Signal sample

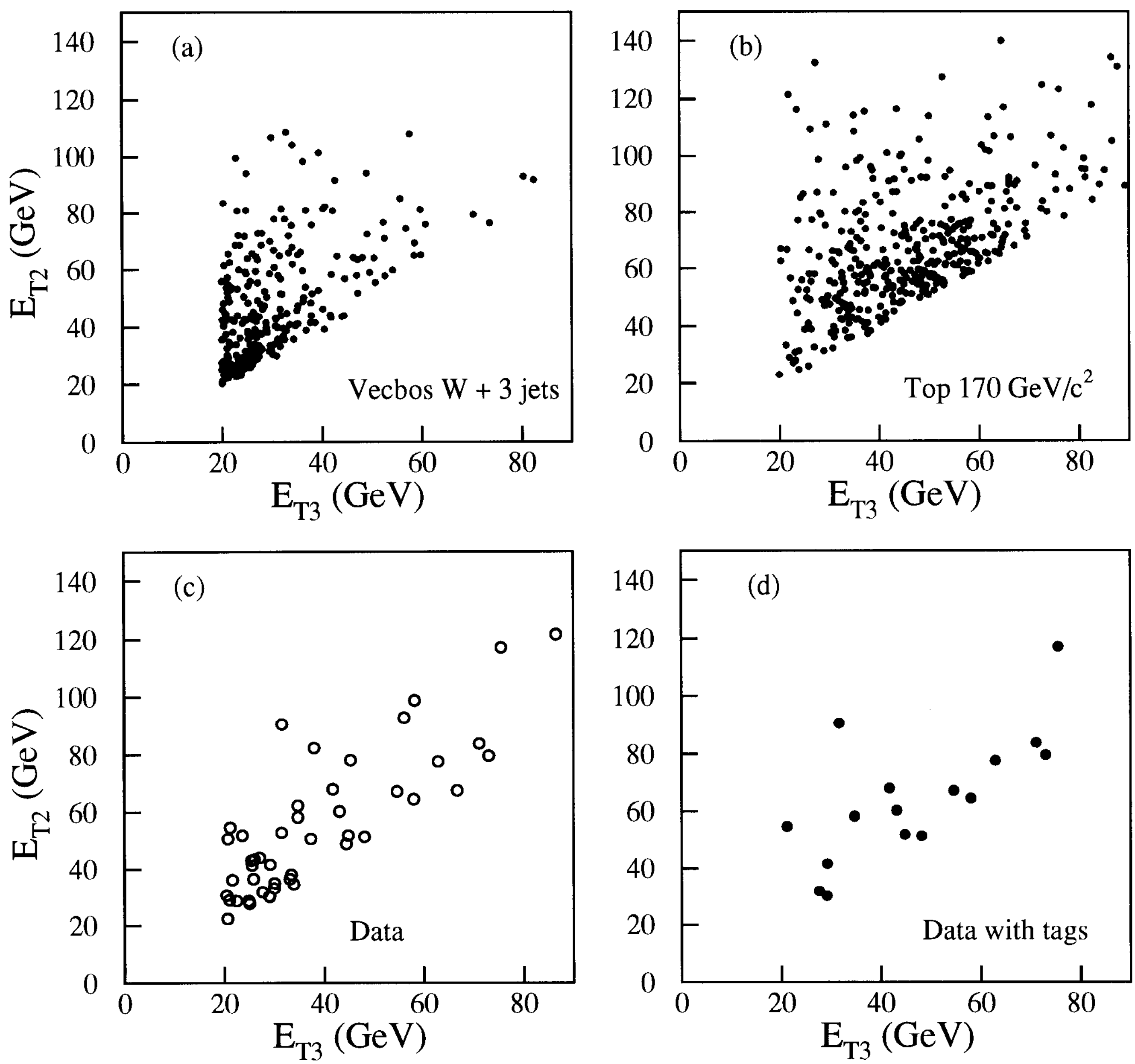

Figure 4: Signal sample : distribution of events in E_{T2} and E_{T3} for (a) direct QCD W+jets, (b) top (M^{top}=170 GeV) Monte Carlo events and (c) for all data events. Data events with a b–tag are shown separately in (d). There is one overflow at E_{T2}, E_{T3} = 121 GeV, 118 GeV in (c).

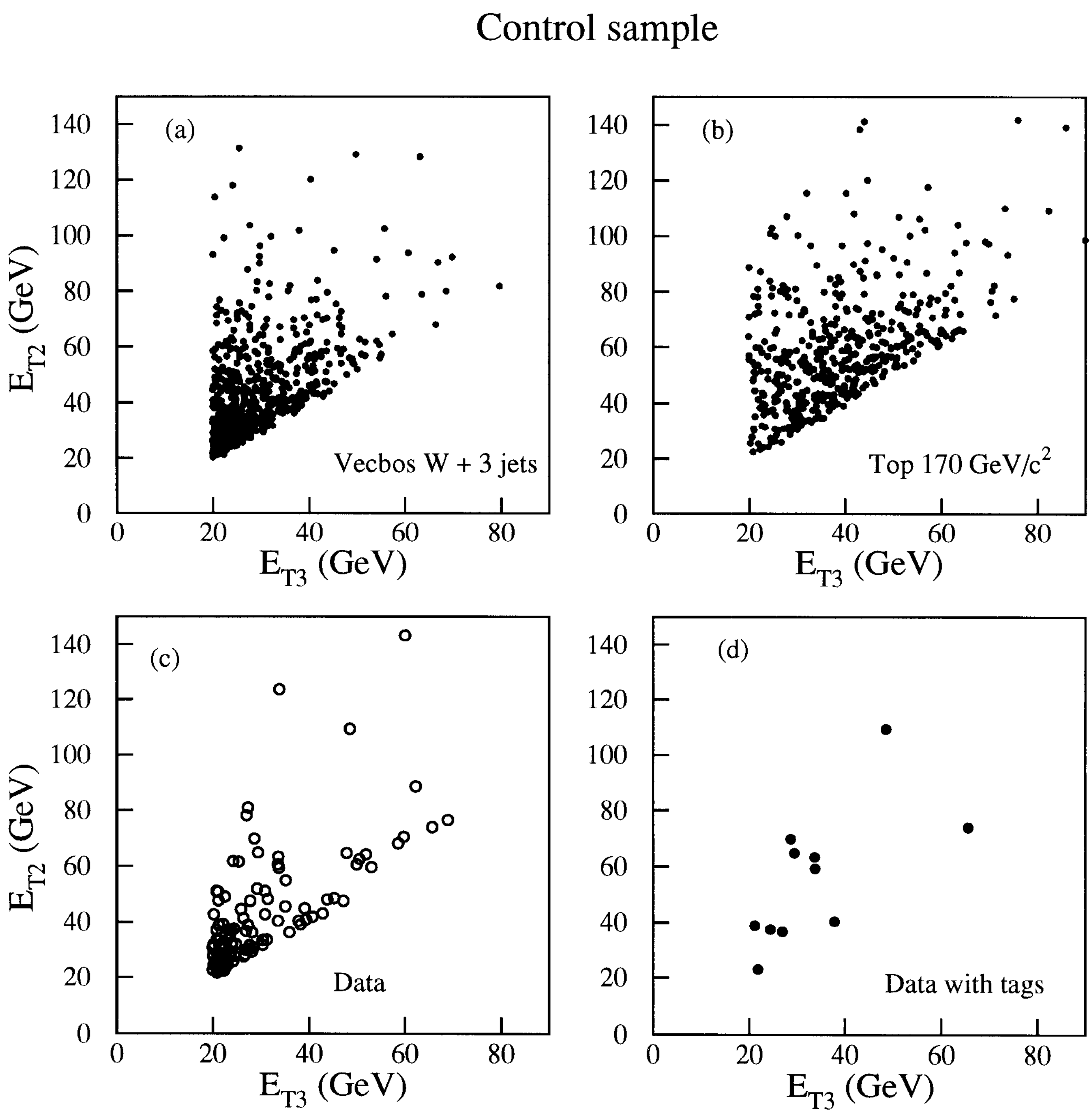

Figure 5: Control sample : distribution of events in E_{T2} and E_{T3} for (a) QCD, (b) top (M^{top}=170 GeV) Monte Carlo events and (c) for all data events. Data events with a b–tag are shown separately in (d).

the second and third highest E_T jet than the Vecbos prediction for direct QCD $W+$ jets production. As a result, also this background is expected primarily at ln(L) < 0. Conservatively, we take the QCD background shape as predicted by Vecbos to represent the shape of all background.

If we make the conservative assumption that all events at ln(L) < 0 are background and normalize the expected background distribution to the observed events with ln(L) < 0, then we would expect 7.2$\pm$2.1 events at ln(L) > 0 compared to the 22 observed. We obtain a probability of $< 0.26\%$ that the 47 events would be distributed with at least 22 events at ln(L) > 0. For this limit we have also taken into account uncertainties of the Vecbos prediction, in the fragmentation modeling and in the reconstruction of primary parton energies from observed jet energies [2]. Figure 7(a) shows the control sample ln(L) distributions for Monte Carlo $t\bar{t}$ and QCD $W+$ jet events. In figure 7(b) we show the ln(L) distribution of the 111 data events in this sample. From Monte Carlo studies, this sample is expected to contain about the same number of top events as the signal sample. In figure 7(b) there are 79 events at ln(L) < 0 and 32 at ln(L) > 0. We have performed a two component fit to the ln(L) distributions using the predictions of the $t\bar{t}$ and QCD Monte Carlo simulations. For Vecbos we use both the predictions based on $Q^2 = M_W^2$ (harder E_T(jet) spectra) and $Q^2 = < P_T >^2$ (softer E_T(jet) spectra). For the signal sample the fit yields 18.0$\pm$5.5 (18.8$\pm$5.4) top events; for the control sample we get 0.8$\pm$8.1 (14.5$\pm$8.1) top events for $Q^2 = M_W^2$ ($Q^2 = < P_T >^2$). Because of the larger background, we find that the estimated top content of the control sample varies significantly when the Vecbos parameters are varied.

In figure 6(b), 7(b) the shaded areas indicate the b–tagged events. The darker areas indicate events with more than one SVX or SLT tag. We observe that a large number of events in the signal enriched event sample are b–tagged. We find a total of 13 SVX tags (in 8 events) compared to 2.80$\pm$ 0.35 SVX tags expected from background alone [11]. We observe that all 13 SVX tags are associated with events with ln(L) > 0, namely in the region where we expect most top events. The SLT b-tag algorithm gives consistent information, but has a much larger background. We observe 11 SLT tags with an expected background of 5.6$\pm$0.8. In the control sample we observe 5 SVX tags (in 4 events) compared to a background expectation of 4.10$\pm$0.44 , and 9 SLT tags with a background of 8.1$\pm$1.2. All SVX tags are at ln(L)> 0.

We compare the number of observed SVX tagged events in the signal and control sample with that expected based on the top content of the samples. The top content is estimated from the two component fits to the ln(L) distributions. Multiplying this by the SVX

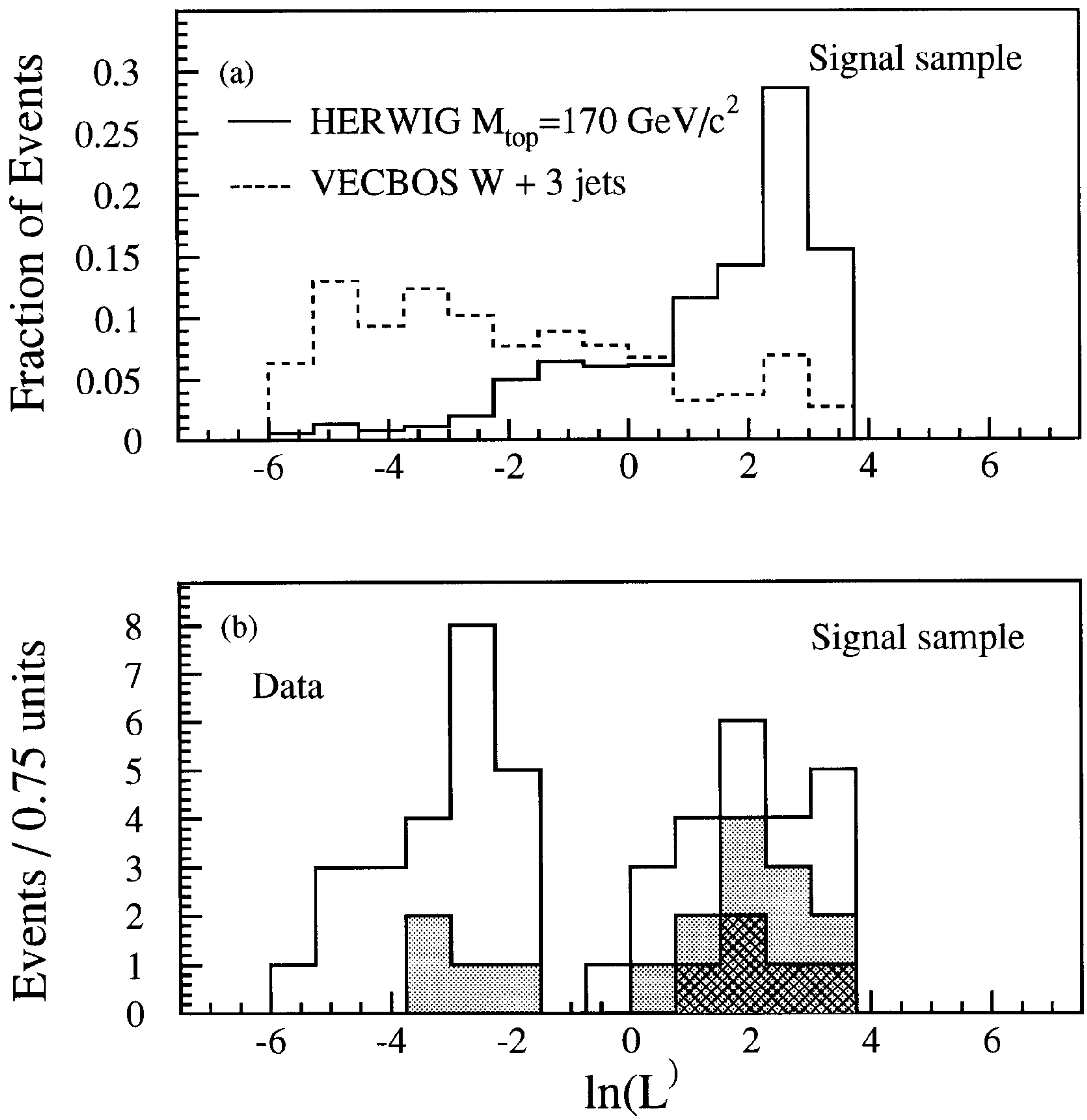

Figure 6: (a) Vecbos QCD and Herwig (M_{top} = 170 GeV/c^2) top Monte Carlo predicted distributions for the $W+ \geq 3$ jet signal sample. Both distributions are normalized to one;$Q^2 = M_W^2$ is used in the Vecbos calculation; (b) Data;The shaded area indicates the b–tagged events from SVX and SLT; The darker area indicates events with more than one SVX or SLT tag.

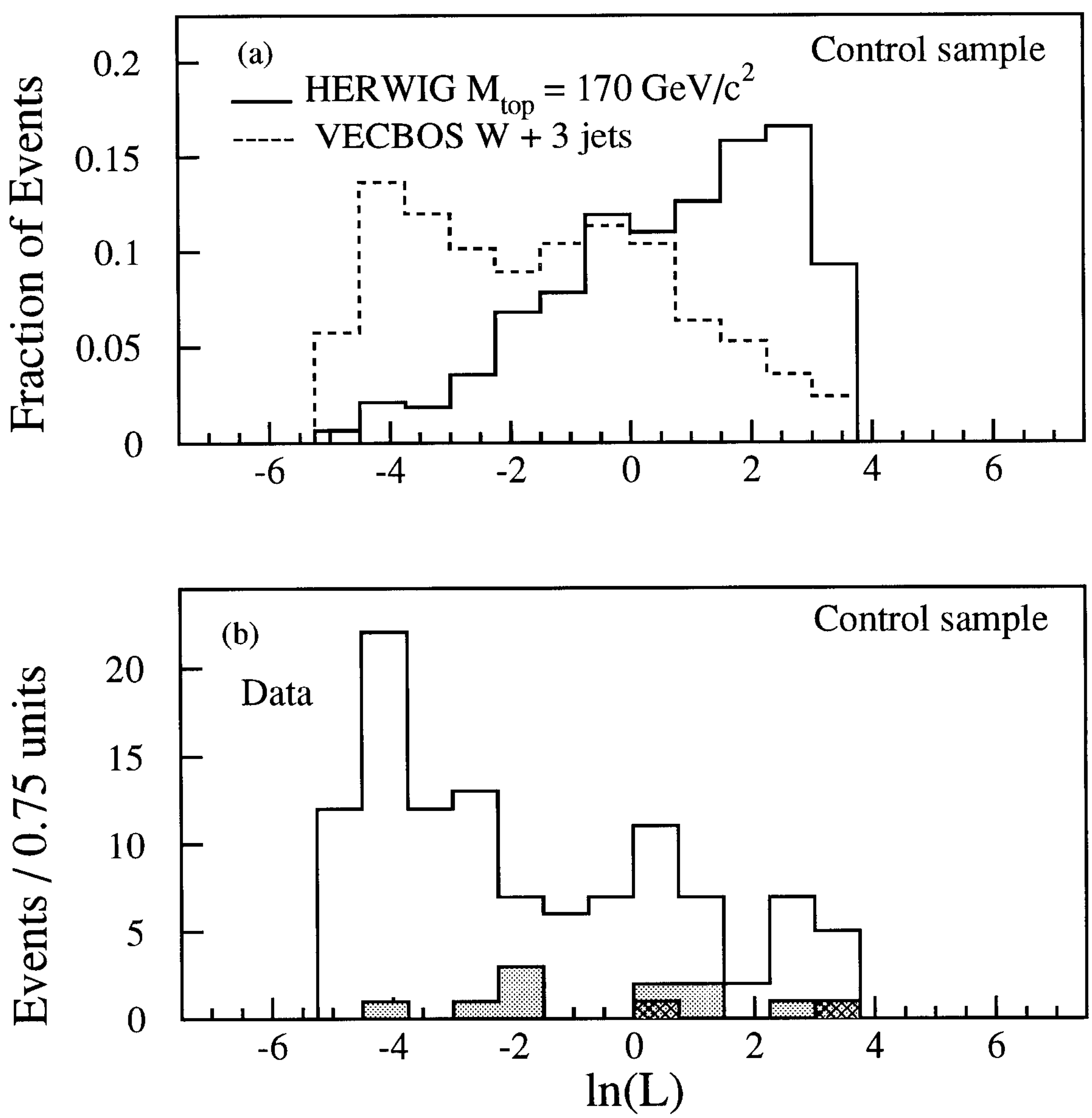

Figure 7: The same as in figure 5, for the control sample.

tagging efficiency[12] and adding the expected tags from background yields an estimate of the number of SVX tagged events. The results are shown in Table 1. Comparisons are shown for the two different Q^2 choices used to simulate the Vecbos background shapes for the two component fits. The agreement between number of expected and observed tagged events is good. We note however that the number of SVX tags observed in the control sample may indicate that the top fraction in this sample is lower than expected from $t\bar{t}$ Monte Carlo.

We evaluate the probability that the observed excess of b-tags in the signal sample over

Sample	$t\bar{t}$ events from fit	Exp. number of SVX tags from backg.	Exp. number of SVX tags from $t\bar{t}$ + backg.	Observed SVX tagged events
Signal ($q^2=M_W^2$)	18.0±5.5	1.7±0.2	9.6±2.6	8
Signal ($q^2=<P_T>^2$)	18.8±5.4	1.7±0.2	10.0±2.6	8
Control ($q^2=M_W^2$)	0.8±8.1	4.10±0.4	4.4±3.0	4
Control ($q^2=<P_T>^2$)	14.5±8.1	3.6±0.4	8.8±3.0	4

Table 1: Comparison of number of observed SVX tagged events in the signal and control sample with the expected b–tags. The top content is estimated from a two component fit of top and QCD background to the ln(L) distributions. The expected number of tagged events from background is modified based on the estimated $t\bar{t}$ content of the sample. Comparisons are shown for two different Q^2 choices used to simulate the Vecbos background shapes.

non-top background expectations is due to a statistical fluctuation. Conservatively we consider the SVX tagged events only and ignore the fact that many events contain more than one b-tag. We find that the probability to observe 8 or more events when 2.80±0.35 are expected is 0.96×10^{-2}. All SVX b–tags occur in events with $\ln(L) > 0$, that is the kinematic region where most top events are expected to be found. At $\ln(L) > 0$ we expect only 1.37±0.17 SVX tags from background, compared to an observation of 8 events. The probability that this observation be due to a statistical fluctuation of background tags is $< 1.2\times10^{-4}$.

4 Outlook

The primary purpose of these analyses was to prove the existence of a new particle, which in the context of the Standard Model must be the top quark. To this purpose, a 'top-like' deviation of the data from expectation in only one parameter would be sufficient, either a kinematical or a b–tag parameter. We combined several parameters (E_{T2}, E_{T3}, cosΘ, ΔR, b–tag) in order to increase the sensitivity of the search. With the top quark (or a close relative) found, one should now proceed to a more detailed study of its properties, investigating as many parameters as possible. One can do this by studying the events of the signal sample, and especially its subsample of events at ln(L) > 0. As an example we investigate the distribution ΔR_{min} (ΔR_{min} = min(ΔR(jet i,j),i,j=1,3), which is used in our analysis to apply a cut on. In figure 8(a) we show the ΔR_{min} distribution for data and QCD Monte Carlo events, under the cuts of the signal sample (except the cut ΔR_{min} > 0.7) for the 'QCD-like' region at ln(L) < 0. QCD events show a peak at small values of ΔR_{min}. For the most part the data agree qualitatively with the QCD prediction, but it is possible that the QCD Monte Carlo does not predict enough events with jets close to each other. In figure 8(b) we show ΔR_{min} for the data events of the 'top-like' region at ln(L) > 0, together with the predictions for top and QCD W+jets events. The data are in good agreement with the prediction for top. There is no indication of the 'peak structure' at small ΔR which is expected for QCD background.
In a similar way we are planning to study all the other kinematical distributions of interest in the near future.

5 Conclusion

The excess of W+jets events at large total transverse energy, H, fits well the top hypothesis. It is being studied in particular as a means to measure M^{top}.
The event structure analysis expands the study to the use of several different parameters. We expect top events to feature high E_T jets produced at large angles relative to the beam and well separated from each other. Indeed we find events with these characteristics in excess of what we expect from non-top background. This excess has a probability of less than 0.26% to be due to a statistical fluctuation, but is in good agreement with what we expect from Standard Model top production. The kinematically top-like events also show a large content of beauty quark candidates, in agreement with expectations from top. The probability that the SVX b–tags are due to a statistical fluctuation of non-top

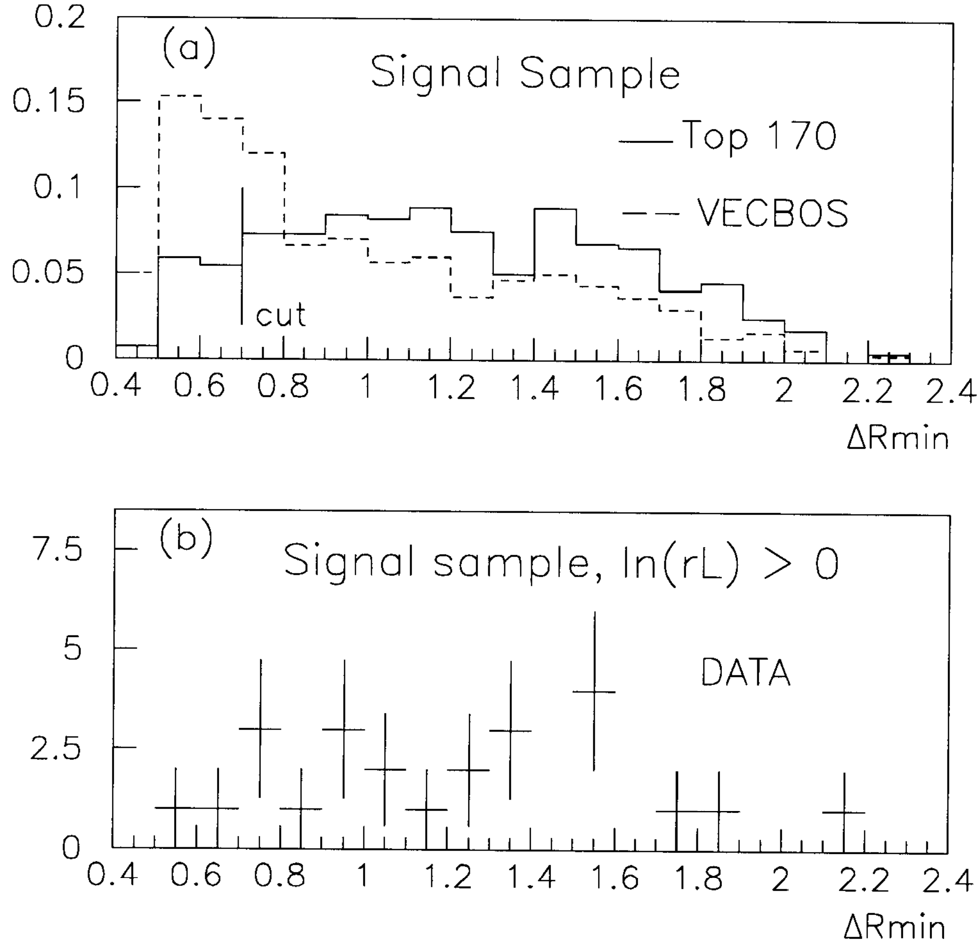

Figure 8: ΔR_{min} = min(ΔR(jeti,j) i,j=1,3) for the events of the signal sample (|cosΘ^* (jet1,2,3) |< 0.7), but without the cut at $\Delta R_{min}>$ 0.7. (a) Monte Carlo predicted distribution for QCD background (Vecbos) and data events for the QCD-like region ln(L) < 0. (b) data and Monte Carlo events from the 'top-like' region of ln(rL) > 0 (M^{top} = 170 GeV). The Monte Carlo events are normalised to the number of data events in each figure.

backgrounds is less than 1.2×10^{-4}.

In conclusion, we have observed that a new physics process contributes to the final state with W+$\geq$3 jets. In the context of the Standard Model this process can only be top. Furthermore we have also shown, that this new process indeed agrees with top expectations with respect to the 5 different parameters which we studied, namely E_{T2}, E_{T3}, ΔR(jets), cosΘ^*(jets) and the beauty content.

References

[1] M.Cobal, PhD Thesis, Feb.1994, Univ. Pisa, INFN-PI/AE 94/004.

[2] F.Abe et al., "*Kinematic Evidence for top pair production at the Tevatron collider*", FERMILAB-PUB-94/411-E, Phys.Rev.D, 51.

[3] F. Abe et al., Phys. Rev. Lett. **73** (1994), 225; F. Abe et al., Phys. Rev. D **50** (1994), 2966.

[4] F. Abe et al., Phys. Rev. Lett. 74, 2626 (1995).

[5] S.Abachi et al., Phys.Rev.Lett. 74, 2632 (1995).

[6] F.A.Berends, W.T.Giele, H.Kuif, B.Tausk, Nucl.Phys. B357, 32 (1991).
W.Giele, PhD.Thesis, Leiden.

[7] G.Marchesini and B.R.Webber, Nucl.Phys.B310, 461(1994).
G.Marchesini et al., Comput.Phys.Comm.67, 456 (1992).

[8] N.Giokaris, RU 95/E-21 (Report Rockefeller University).

[9] M.Cobal, H.Grassman, S.Leone, Il Nuovo Cimento 107 A, 75 (1994).

[10] F.Abe et al., Phys.Rev. D 45,1448 (1992).
F.Abe et al.,Phys. Rev. D 47, 4857 (1993).

[11] We use the conservative 'method 1' b–tag background calculation described in ref.[3]. The probability of finding a b-tag is predominantly derived from tags found in multijet QCD events.

[12] The SVX tagging efficiency is estimated to be 44% for the signal sample and 36% for the control sample.