RU-94/E-12
CDF/PHYS/TOP/CDFR/2502
Version 1.0
April 1, 1994

A Search for the Top Quark Using Kinematic Variables

Luc Demortier, Qingfang Wang, Nikos Giokaris

The Rockefeller University

Michael G. Albrow

Fermilab

"What men really want is not knowledge, but certainty."
Bertrand Russell

Abstract

We have made a search for $t\bar{t}$ production in the Run 1A sample of W + jets events using a new variable, H, that equals the scalar sum of the E_T of the lepton, neutrino and jets. By requiring exactly four jets passing relatively loose cuts (uncorrected $E_T \geq 8$ GeV, $|\ \eta_{\rm det}\ | \leq 2.4$), the distribution of H shows signs of a double-bump structure. We can reproduce this with the Monte Carlo programs VecBos (W + 4 jets) and Herwig ($t\bar{t}$, $M_{\rm top} = 180$ GeV) but not with VecBos alone. A fit gives 18.2 ± 6.0 $t\bar{t}$ events in the sample. To further test the hypothesis that there are two distinct classes of events in the sample we study some quantities providing 'orthogonal' information, namely the circularity C, the fourth jet fraction $X_4 \stackrel{\rm def}{=} E_T(\text{jet } 4)/H$, and the probability P_b that the events have a b-tag. Plots of these quantities versus H provide complementary evidence that the structure in the H distribution is caused by two classes of events and that the high-H bump is predominantly $t\bar{t}$.

1 Introduction

As we all know well it is very difficult to extract a good, 'unmistakeable' signal for the top quark with the luminosity we have had so far, 19.6 ± 0.8 pb^{-1} in run 1A. The cleanest channel, dileptons, has just 2 candidates with an expected background of 0.56 events. The fully hadronic 6-jet decay of $t\bar{t}$ is considered almost hopelessly swamped with QCD background. So our best hope seems to be the lepton + $\not{E}_T$+ jets channel, with more statistics than dilepton and less background than 6-jets. However, a simple counting experiment in $W(l\nu)$ + 3 or 4 jets does not show any significant effect, so additional information is needed. The top Phys.Rev.D draft paper describes the use of secondary vertex and soft lepton tagging to preferentially select events with b-jets. Unfortunately both methods have rather low efficiency, at the level of 15–20% per $t\bar{t}$ event [1]. These efficiencies are also very difficult to measure, at least so far. Nevertheless they seem to succeed in increasing the signal-to-background ratio, not quite killing the statistics in the process but putting them on the critical list.

We first describe in section 2 the selection of our event samples, with a leptonic W plus any number of jets, and then in section 3 discuss the distributions of H, the total E_T of leptons and jets. This shows an anomaly at high H. In section 4 we study other event characteristics namely the circularity C, the fourth jet E_T fraction and the fraction of b-tags. In section 5 we compare the data with the Monte Carlos, VecBos and TOP180 or TOP160 and derive a significant fraction of top events and the $t\bar{t}$ cross section. Section 6 will present our conclusions and a "job list" of future studies.

2 Data Samples

The sources of our CDF data sample are the inclusive central electron sample made by Dave Saltzberg [2], and the inclusive central muon sample made by Mark Krasberg [3]. A W-enriched sample is then obtained with the following additional requirements:

1. One central electron or muon with uncorrected p_T above 20 GeV. This lepton must pass the identification, isolation, fiducial and trigger cuts specified for the standard lepton + jets analysis [4]. Z events, cosmic ray muons and conversion electrons are removed by the algorithms described in the same reference.

2. The missing transverse energy must be larger than 20 GeV. Here, the missing E_T is corrected for high p_T muons only.

For this analysis we will need to classify events according to the number of jets they contain. In addition, jets will appear in the calculation of several kinematical variables. For these purposes, we define a jet as a non-electron cluster in the calorimetry, reconstructed with a cone of radius 0.4, with uncorrected $E_T > 8$ GeV and with $| \eta_{det} | \leq 2.4$. For simplicity we will not require different E_T cuts for different jets.

Our background Monte Carlo is VecBos W + 4 jets, with hadronization performed by the Herwig based Herprt module [5], and with QFL for detector simulation. We used the publicly available samples stored on FNALD in the CDF$MONTE_CARLO_DATA:[QFL] area. These samples were generated with the MRSD0 set of structure functions, and with the QCD scale set to the average outgoing parton p_T. Infrared and collinear singularities were regulated by requiring the outgoing partons to have a transverse momentum above 8 GeV/c and a mutual (η, ϕ) separation exceeding 0.4.

For our $t\bar{t}$ Monte Carlo we used the samples available on FNALD in the CDF$TOP_-DATA:[ANA.FITTER_TESTS] area. These samples were generated with Herwig, followed by the Cleo Monte Carlo QQ for the b-decays, and by QFL for the detector simulation.

3 H, the Total E_T of Leptons and Jets

We decided to see, as several others in CDF have done, if some simple kinematical variables, applied to a W + 4 jets sample, would have some discriminating power between $t\bar{t}$ and background. We take it for granted that the top quark is heavy ($M_{top} \geq 150$ GeV/c^2). In this case, even if the top were produced with no p_T, the W and b-jets would themselves have high p_T, above 50 GeV/c. Of course the t and $\bar{t}$ will usually be produced with a large p_T, often enhancing the hardness of the W and b-jets relative to that of radiated gluons in the background processes. One can construct a large number of variables, but in the interest of simplicity we tried $H \stackrel{\text{def}}{=} \Sigma E_T$, where the sum runs over the corrected electron E_T or muon p_T, the QDJSCO corrected jet E_T's, and the $\not{E}_T$ corrected for electron, muons and jets. This variable was suggested by John Yoh [7] last year; he calls it Action. Presumably others have thought of it, it is so simple.

The distribution of H for CDF W events, plotted according to the number of jets, is shown in Fig.1a–f. Notice that the 4 and 5-jet samples contain a total of 73 events (and there are 3 events with at least 6 jets), many more than the 9 4-jet events of the PRD sample with the harder cuts. As one would expect, H increases with the number of jets, the peak in the distribution gaining about 22 GeV for each additional jet. The distribution also broadens.

We find it very striking that the distribution for W + 4 jets looks qualitatively different from the 3-jets case. Its average value jumps 61±12 GeV compared with 29±1 GeV for the other increments in n-jet; this is shown in Fig.2. However the position of the peak has moved by the same increment as for the 0-1-2-3 jet cases. It even appears bimodal, although admittedly this is not of great statistical significance. The 5-jet statistics are too low to conclude much, except perhaps that the low-H peak seems to have disappeared and the increment in H may be back to "normal", as shown in Fig.2. A linear extrapolation of <H> through 0, 1, 2 and 3 jets works well but fails to predict the 4-jet value by about 3 standard deviations. That the 3-jet point is slightly high could be due to a "contamination" of this bin also by top events.

If top is dominating the high-H events with 4 jets, it will probably continue to dominate when there are more than four jets, from additional gluon radiation. Therefore we show in Fig. 3 the H distribution for four or more jets; it is the sum of Fig. 1e and 1f together with three events with more than 5 jets. We think it is quite a striking plot even if (of course!) one cries out for more statistics. For some of the following studies we use this sample of 76 events rather than the 57 with exactly 4 jets, but there are no important differences.

4 Other Event Characteristics

We will return later to the shape of the distribution in Fig. 1e, but before we do so we ask whether there are event characteristics (preferably orthogonal to H) that support the hypothesis that most of the events above H = 250 GeV are of a different type than those with lower H. Again, there are many possibilities and we have not done a systematic search for the most promising.

4.1 Circularity

Continuing to work only in the transverse plane, an obvious question is: "How spread out in azimuth are the six E_T vectors?". Are they well spread out in azimuth like a circular "explosion", or does it look more as if three on one side are recoiling against three on the other side? A natural variable to address this question is circularity, C [6]. This is the two-dimensional analog of sphericity, used mostly in e^+e^- collisions and once promoted as a good way to find top in e^+e^- (the event sphericity would increase when the threshold was passed). In hadron-hadron studies, circularity is more natural than sphericity because forward, low E_T but high E jets are irrelevant. Also our detector is uniform in azimuth but it is certainly not isotropic.

The circularity axis of an event is defined in the transverse plane, as the direction along which the sum of the squares of the projected transverse momenta is minimal. This sum itself, when properly normalized, is called circularity:

$$C = \min_{\{n_T\}} \frac{2 \cdot \sum_i (\vec{p}_{Ti} \cdot \vec{n}_T)^2}{\sum_i p_{Ti}^2} \qquad (1)$$

The sum in equation 1 runs over the W p_T and the corrected jet E_T's. The W p_T is calculated as the vector sum of the corrected lepton p_T and the corrected $\not{E}_T$. We prefer to use the p_T of the W rather than the p_T's of its decay products separately; this should give a more direct measurement of C "at production". It can be shown that C is equal to the smallest normalized eigenvalue of the circularity matrix:

$$\begin{pmatrix} \sum_i p_{xi}^2 & \sum_i p_{xi}p_{yi} \\ \sum_i p_{xi}p_{yi} & \sum_i p_{yi}^2 \end{pmatrix}$$

namely:

$$C = 1 - \frac{\sqrt{\left(\sum_i p_{xi}^2 - p_{yi}^2\right)^2 + 4\left(\sum_i p_{xi} p_{yi}\right)^2}}{\sum_i p_{Ti}^2} \tag{2}$$

The circularity ranges from 0 to 1, being 0 for two exactly balancing jets, and 1 for any azimuthally symmetric configuration of more than two equal jets.

When we plot the mean circularity $<C>$ versus H for the 4-or-more jet data there is a very striking fall from about 0.4 to 0.2 followed by a distinct rise or bump (Fig. 4a). This is reminiscent of the expected change in sphericity in e^+e^- machines if they passed the $t\bar{t}$ threshold! We now have recourse to Monte Carlo just to tell us qualitatively what behavior to expect. VecBos W + $\geq$4 jets has C falling smoothly from 0.42 for H = 150 GeV to 0.18 at 400 GeV, see Fig 4c. A Monte Carlo simulation of $t\bar{t}$ with $M_{top} = 180$ GeV shows that above the threshold around H = 200 GeV C is about 0.40 and it remains high, only beginning to fall around 400 GeV, see Fig. 4b. For most of the range 200 GeV $\leq$ H $\leq$ 600 GeV, $<C>$ is about 0.15 higher for $t\bar{t}$ than for VecBos.

This is not a strong discriminator, nevertheless it provides quite independent evidence, since C is orthogonal to H. To labor the point, if the excess of events at large H was just a statistical fluctuation there is *no reason* why those events should look more circular than an extrapolation of the lower H events would predict.

4.2 Fourth Jet E_T Fraction

Another natural question is: "how spread out are the jet E_T vectors in $|\vec{E}_T|$?" A simple way to investigate this is just to study the relative softness of the fourth jet, so we study $X_4 \stackrel{def}{=} E_{T4}/H$. The data plotted as $<X_4>$ versus H are shown in Fig.5a. They decrease linearly with H from about 0.07 (just above 8 GeV at this H), as you would expect for QCD background with the fourth jet tending to stay just above the cut. The VecBos Monte Carlo confirms this expectation, as shown in fig 5c (we discount the odd point at 525 GeV, surely a fluctuation ... this is only Monte Carlo!). But the data appear to change in the direction of higher X_4 around 300 GeV. This is not very significant and we do not claim a signal here. The $t\bar{t}(180)$ Monte Carlo however shows a constant mean X_4, so that E_{T4} scales with H. Hence the variable has discriminatory power, although our statistics do not allow a definite conclusion on this point.

It might be interesting to see, comparing the high-H and low-H events, how well two of the four jets can fit M_W, or how well the whole event fits the full $t\bar{t}$ hypothesis. This is beyond the scope of this note; we encourage our expert kinematic fitters to work on this sample.

4.3 b-Tagging

So far we have not used any b-tagging, because our aim was to see if a top signal could be extracted without that. We believe that the above evidence shows that it can. If we are right, the "H-bump" and "C-bump" will be dominated by top events and they will contain significantly more b-tags than the events with H $<$ 250 GeV. At the time of writing we do not have the b-tag information for most of these events. Of the 52 PRD W + $\geq$3 jet events, only 9 have 4 or more jets. However a sample was studied, as reported in the PRD, where the threshold on just the fourth jet was lowered to 8 GeV and the $\eta_{\rm det}$ range for it opened to $\pm$2.4 (the other three jets were kept above the 15 GeV threshold). This increased the W + $\geq$4 jet sample to 19 events, which were passed through the SVX and SLT tagging algorithms. Seven events were tagged, and you can see from the histogram in Fig. 6 that there is a tendency for them to be at high H. In fact all 7 have H above 250 GeV, where the anomalies in our H and C distributions appear. There are 14 events in this plot with H $>$ 250 GeV and 50% of them have tags. This is even more than we would have expected if *all* these events are top, but the statistics are small. None of the 5 events with H $<$ 250 GeV have tags. We would very much like to see the equivalent plot for the 76 events of Fig.3, but do not wish to delay distribution of this note.

5 Monte Carlo Predictions

Up to now we have only used Monte Carlo predictions to provide a qualitative guide as to the expected behavior of variables as a function of H. We think our claim that the data show a distinct population of events, with general characteristics not unreasonable for top, stands up independently of Monte Carlo simulations. Now we shall relax and see what the Monte Carlos predict for the H-distribution with our cuts, to see whether the data are accounted for and if so with what top quark mass and production cross section. Figs. 7 and 8 show an enlargement of fig. 1e, with the data shown as points with ($\sqrt{N}$) errors (not that the number of events observed has any error, but we follow tradition). The solid histogram is the best fit linear combination of VecBos and $t\bar{t}$(160) (fig.7), or $t\bar{t}$(180) (fig.8).

There certainly are other backgrounds besides W+jets production and which we have ignored, but it is clear that VecBos fails to reproduce the high H data. The overall χ^2/dof for a normalization-free fit without any top contribution is 15.3/12. This "topless" fit to the shape has only 7.2 of the 57 events above H = 250 GeV while the data has 21 events there. These are very significant numbers, but of course there are other non-Vecbos backgrounds such as fake lepton + jet events. This is an on-going study, but indications are that at least this background is negligible. Of course we cannot totally exclude QCD having some bizarre behavior coming in only when the W is accompanied by 4 jets and not by 3 jets! That might be a more important discovery than the top quark! We should study other backgrounds in terms of H and C, but we would be astonished if any background can reproduce the observed behaviour. When we include a contribution from $t\bar{t}(160)$ or $t\bar{t}(180)$, allowing the fraction of W+jets and $t\bar{t}$ to float so that their sum fits the data, we get an improved χ^2/dof of 9.7 ($t\bar{t}160$), Fig.7, or 6.9 ($t\bar{t}180$), Fig.8. For the better 180 GeV fit the fraction of top events in the whole sample is 0.24 ± 0.08 and 65% of the events with $H > 250$ GeV are $t\bar{t}$. Clearly one could try to optimize M_{top} from this kind of fit, but the uncertainty in the shape of the background does not warrant it. The fit with $t\bar{t}(180)$ has better χ^2 than $t\bar{t}(160)$ or $t\bar{t}(170)$. The values are, with 11 d.o.f., 9.68 (160 GeV), 8.41 (170 GeV) and 6.88 (180 GeV).

We now make a rough estimate of the $t\bar{t}$ production cross-section.

$$\sigma(t\bar{t}) = \frac{\text{Number of 4jet events fit}}{\int \mathcal{L}\,dt \times BR \times \epsilon(W) \times \epsilon(4jet)} \tag{3}$$

The fit in Fig.8 for $t\bar{t}(180)$ gives 18.2 ± 6.0 top events. The $t\bar{t}(180)$ Monte Carlo predicts that 40% of all W + jets events will result in exactly four jets with our cuts, with most of the others in the 3- and 5-jets classes. So we take $\epsilon(4jet) = 0.40$. The BR to the lepton (e or μ) + jets class is 0.30, and we take $L = 19.6 \pm 0.8\ pb^{-1}$, and $\epsilon(W) = 0.90$ from the Draft PRD. These numbers give $\sigma(t\bar{t}) = 8.6 \pm 3.0$ pb. This is lower than the PRD cross section for 180 GeV top (PRD Fig. 44) but only by one standard deviation.

6 Future Work

We are well aware that there are many studies to do, and that some of them will be much more efficiently and quickly done by you. We would welcome such collaboration. A partial list of obvious jobs follows:

1. Apply our standard b-tagging algorithms to the 76 events with 4 or more jets, and study the b-tagging probability vs H.

2. Do full kinematic fits on the high-H and high circularity events. How does the χ^2 for the fit to a hadronic W, and for the fit to equal mass $t\bar{t}$ kinematics, vary with H?

3. Include other backgrounds in the fits.

4. Investigate the 6-jet events using both H and C, without and with asking for 1 or 2 b-tags.

5. Study errors and probabilities more carefully.

6. The obvious.

7 Conclusions

We think that the Run 1A data sample of W + jets shows anomalous behaviour in the simple kinematic variable H, the summed E_T of the lepton, $\not{E}_T$ and jets, for four or more jets above (only) 8 GeV. There is an excess of events with large H compared with expectations from 1,2,3-jets or from VecBos. Independent evidence that the large H events are anomalous comes from the mean event circularity which, rather than decreasing monotonically with H has a rise above H = 250 GeV. The magnitude of this rise is reasonable for a population changing from predominantly VecBos to predominantly top, with mass around 180 GeV (We do not claim to disfavor lighter masses like 160 GeV). A fit of the H distribution to VecBos + $t\bar{t}(180)$ leads us to conclude that about 2/3 of the events with H > 250 GeV may be $t\bar{t}$.

If this method is successful, as we believe it is, that is because the top quark is as heavy as 160-180 GeV/c^2. A much lighter top quark, say 120-140 GeV/c^2, might not stand out so much, but we have not made any quantitative study of this.

8 Acknowledgements

Several people have provided specific help for which we are grateful. We will try to name them in a later version. Right now (Nikos is incommunicado) we would surely miss some, so prefer to give none. But everyone in CDF past or present has obviously contributed.

Also we apologize in advance for not giving specific references to other attempts to extract top signals from kinematic properties. The understanding of this paper does not require such referrals, and we did not wish to risk being incomplete.

References

[1] Draft PRD tables 18 and 22.

[2] David Saltzberg, "High-E_T Central Electron Datasets", CDF note 2213 (August 30, 1993)

[3] Mark Krasberg, "High-P_T Central Muon Datasets", CDF note 2326 (October 25, 1993)

[4] Dan Amidei *et al.*, "Sample Selection for the Lepton + Jets Channel Top Search", CDF note 2065 (December 6, 1993)

[5] José Benlloch *et al.*, "On Transforming Partons into Jets: HERPRT, a New Interface between ME MC and Fragmentation Models", CDF note 1823 (October 1, 1992)

[6] T. Åkesson *et al.*, "The Dominance of Jets at Large Transverse Energy in a Full Azimuth Hadron Calorimeter at ISR Energies", Phys.Lett. 128B (1983) p.354

[7] John Yoh, "Proposal for Calibratable Variable to Discriminate between Top and Background", version 0.5, October 25, 1993 (privately circulated).

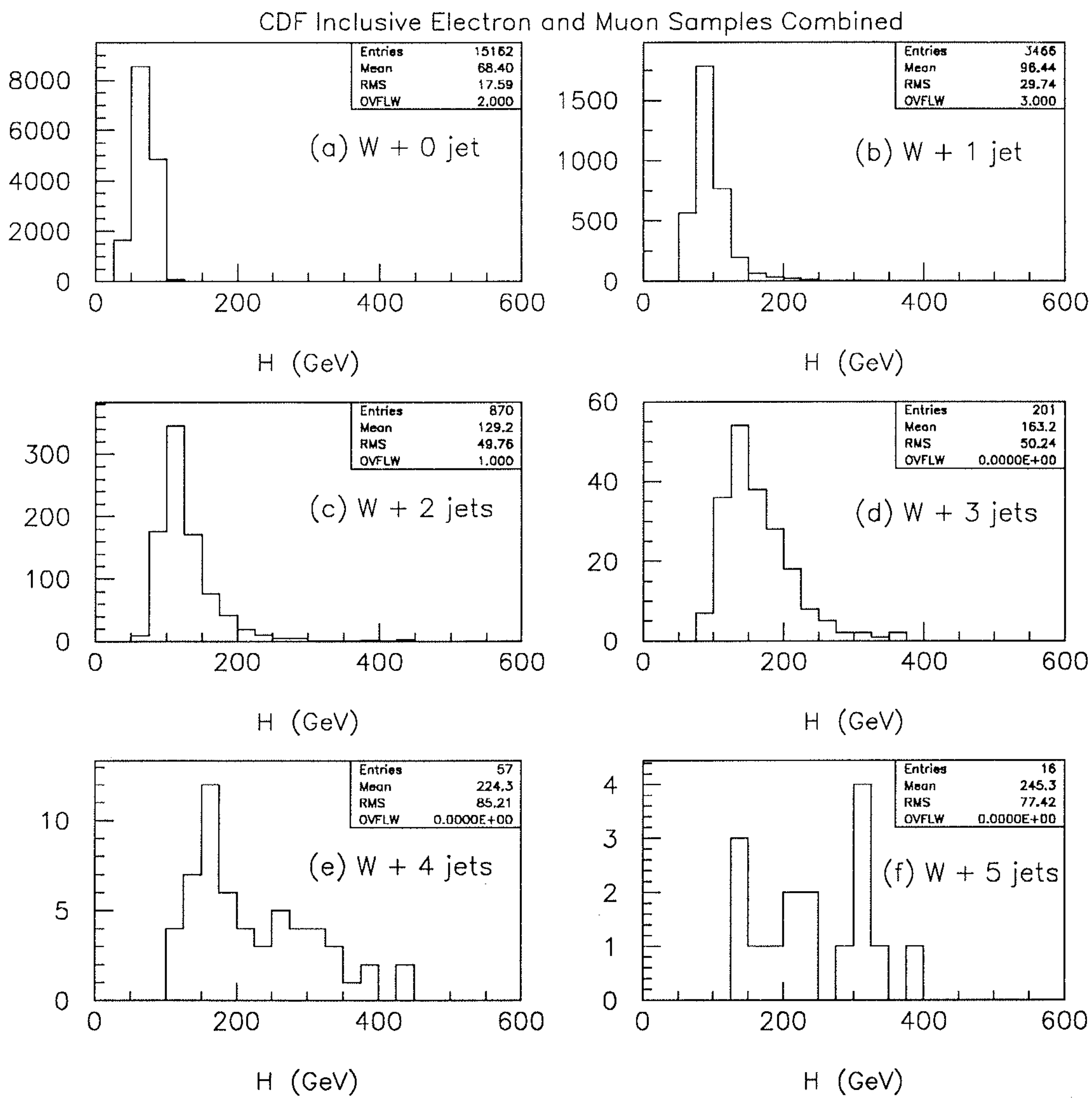

Figure 1: CDF data: distribution of H for W events with (a) 0 jets, (b) 1 jet, (c) 2 jets, (d) 3 jets, (e) 4 jets and (f) 5 jets. The jets all have $E_T^{\mathrm{uncorrected}} \geq 8$ GeV and $|\ \eta_{\mathrm{det}}\ | \leq 2.4$.

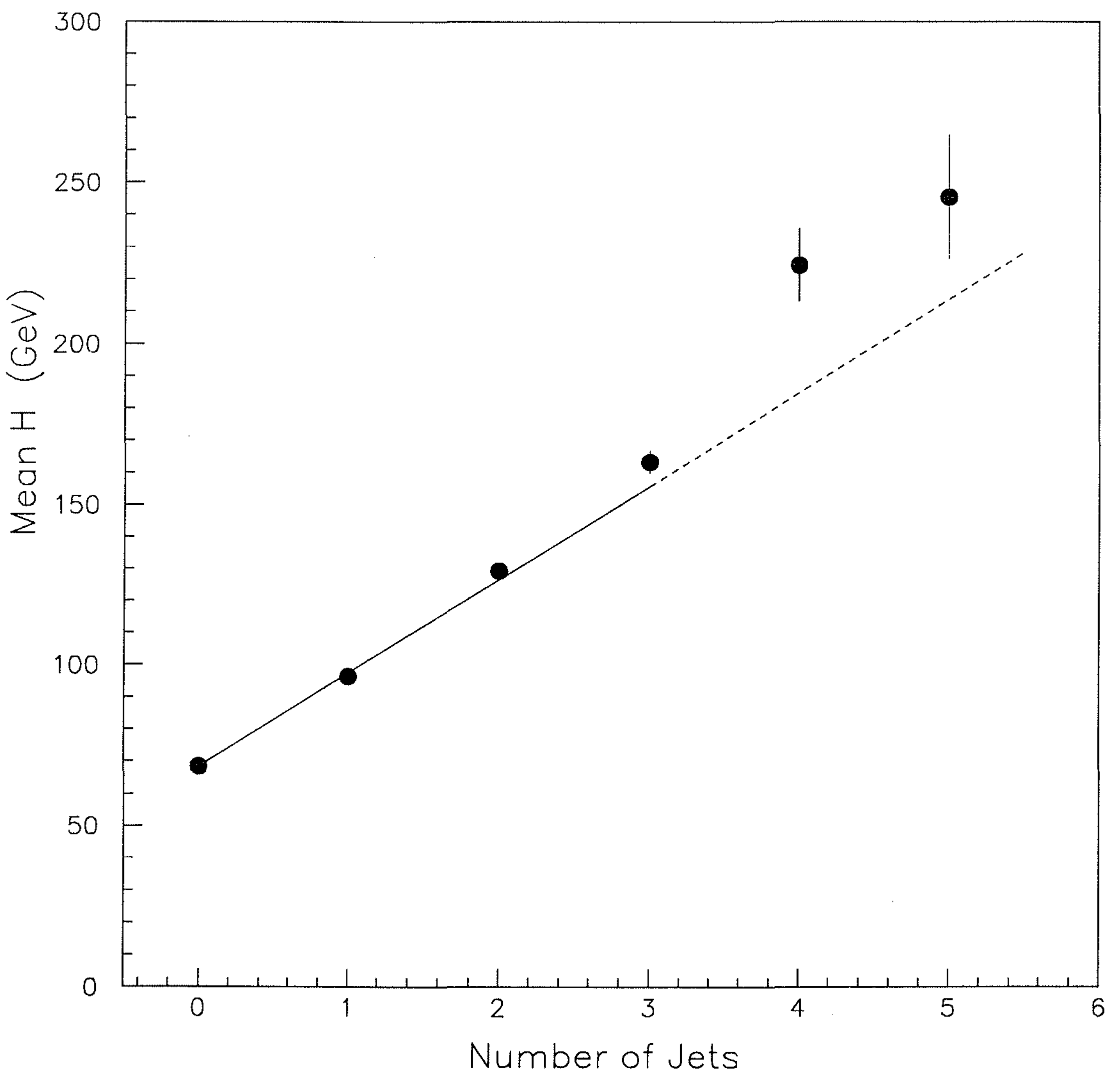

Figure 2: CDF data: average value of H versus the number of jets. The straight line is a fit to the first four points, 0–3 jets.

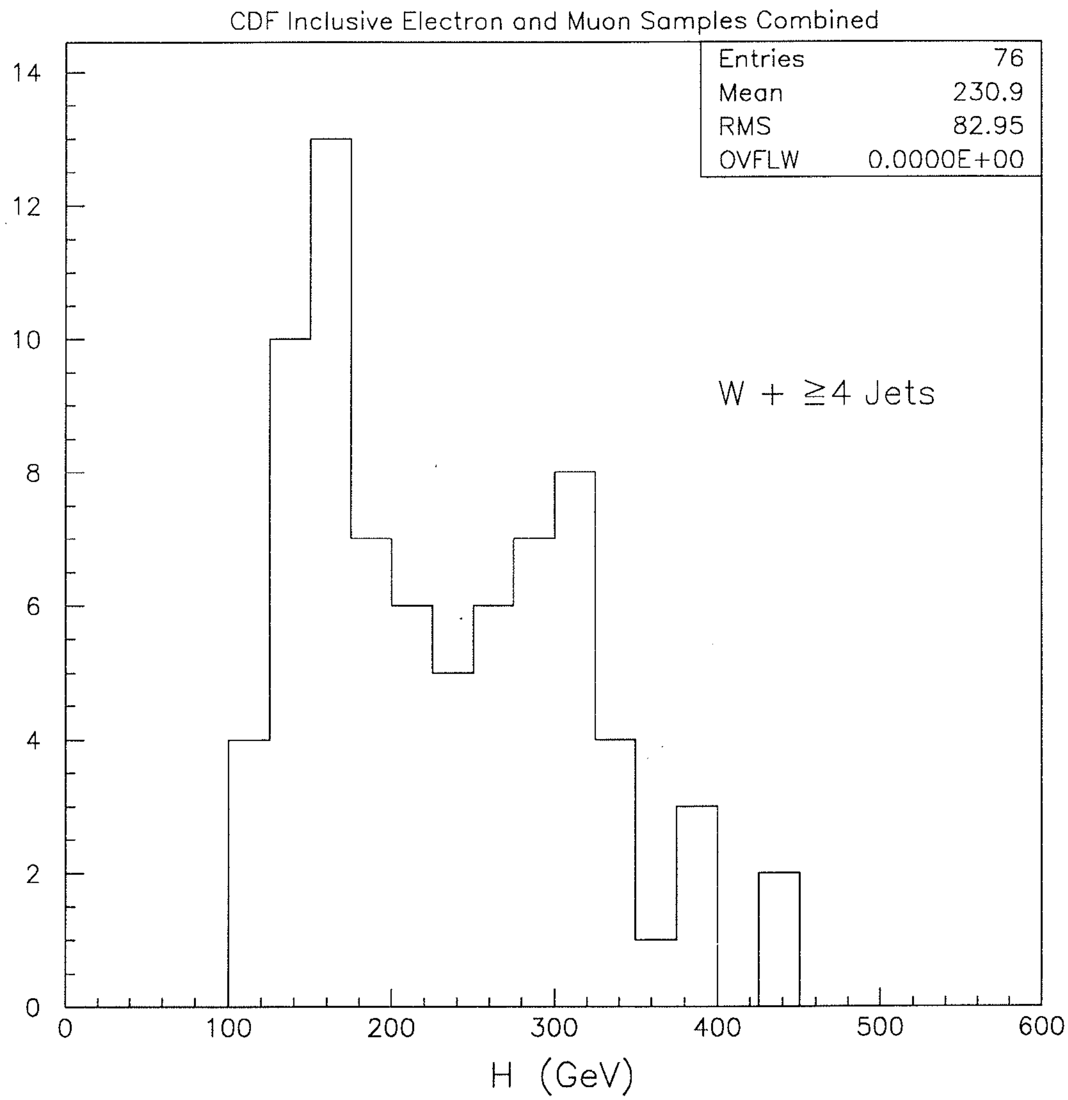

Figure 3: CDF data: distribution of H for W events with four or more jets.

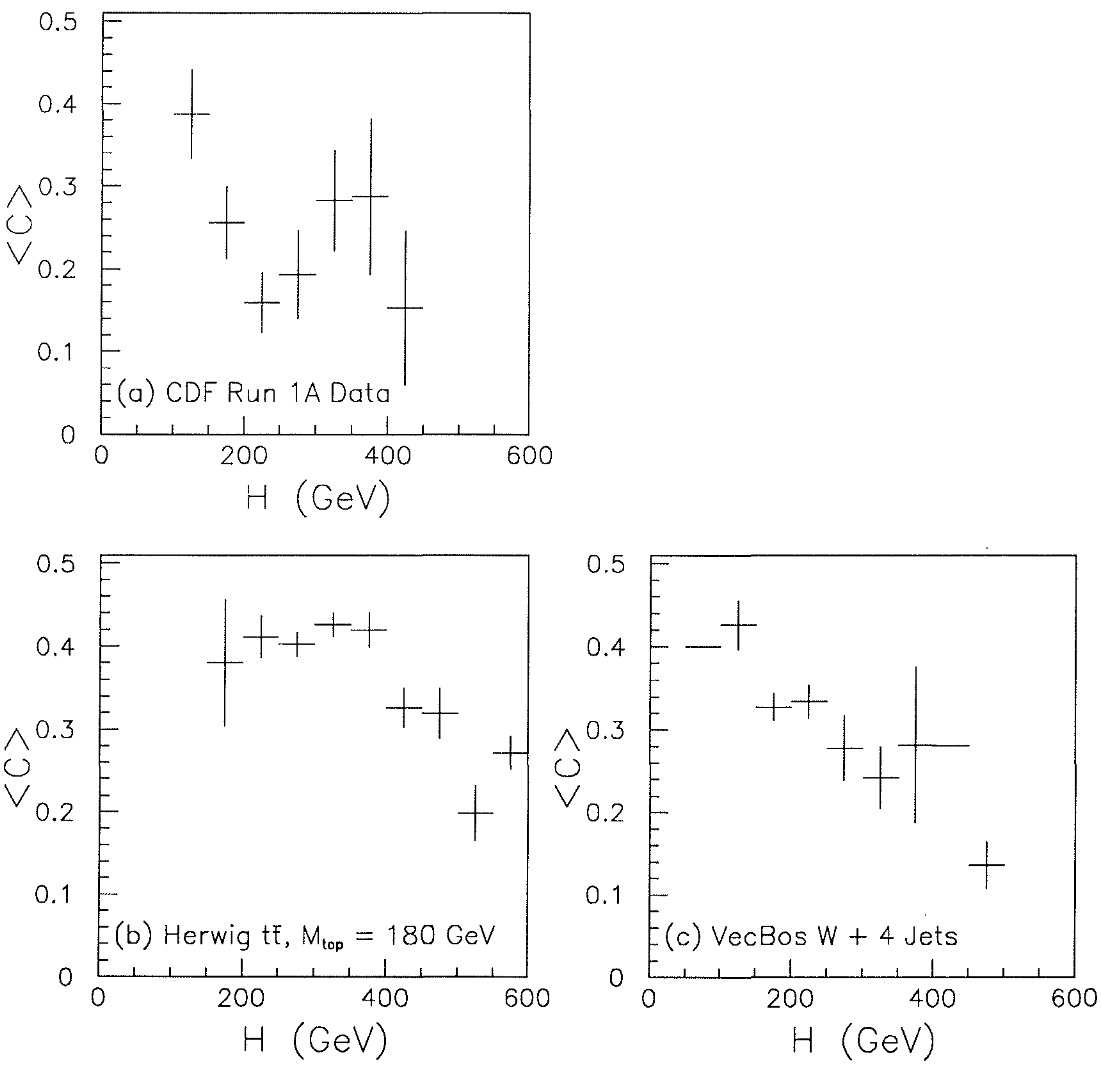

Figure 4: Average circularity of events with W + 4 or more jets, versus H. (a) CDF Data; (b) Herwig $t\bar{t}$, M_{top} = 180 GeV/c^2; (c) VecBos W+jets Monte Carlo.

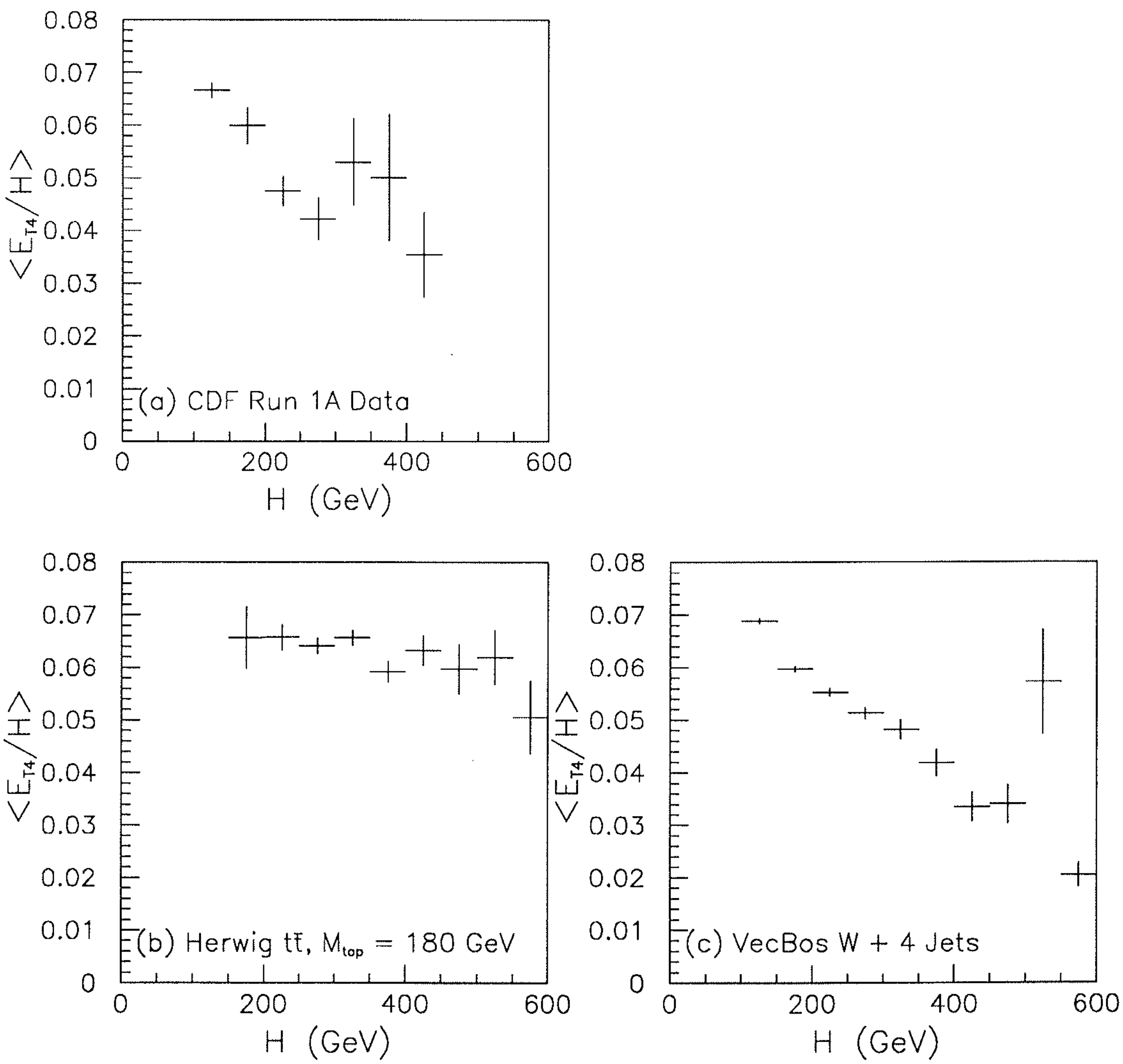

Figure 5: Average E_{T4}/H of events with W + 4 or more jets, versus H. (a) CDF Data; (b) Herwig $t\bar{t}$, $M_{top} = 180$ GeV/c^2; (c) VecBos W+jets Monte Carlo.

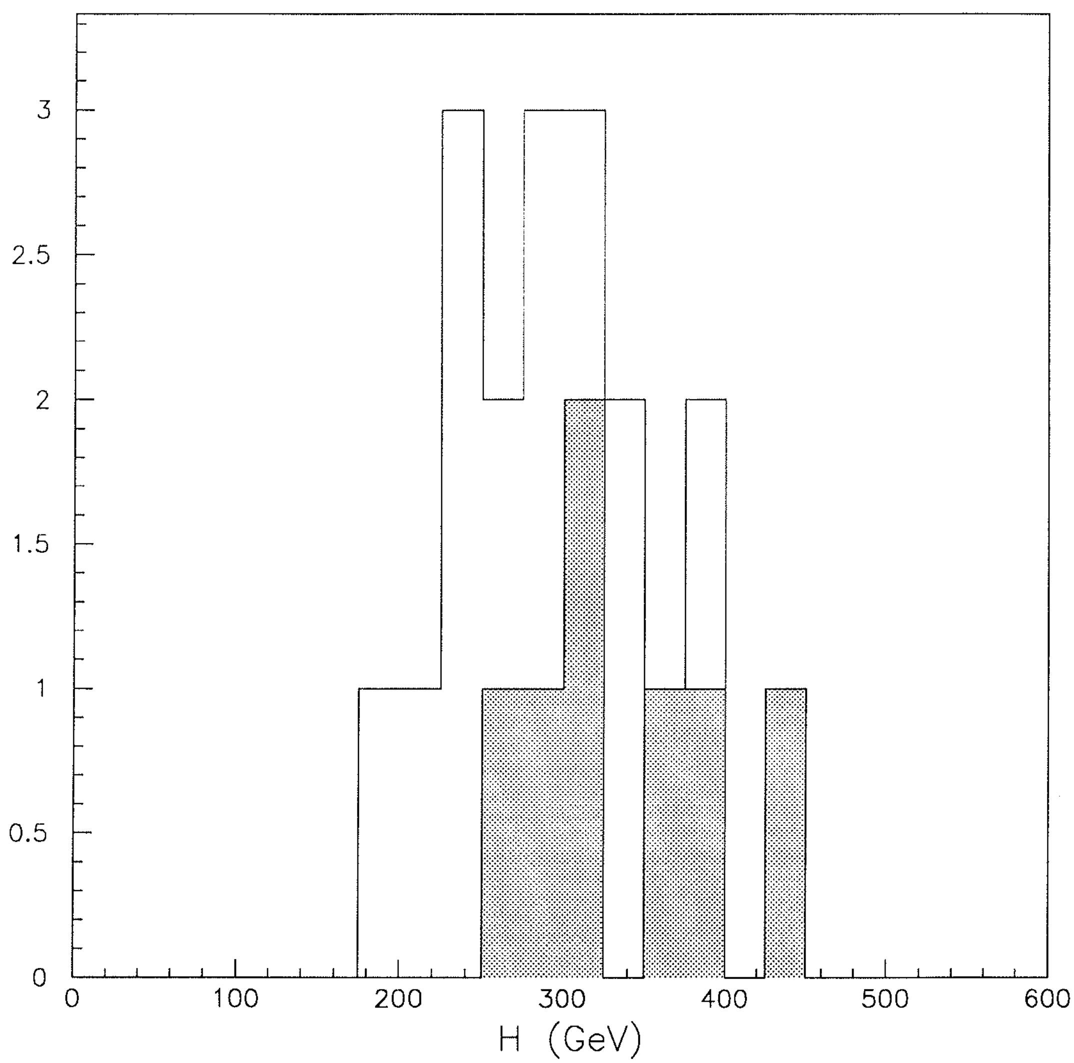

Figure 6: CDF data: H-distribution of the 19 events with 4 or more jets from the PRD sample with loosened jet-4 cuts. The shaded histogram shows the subsample of 7 events with an SVX or SLT b-tag.

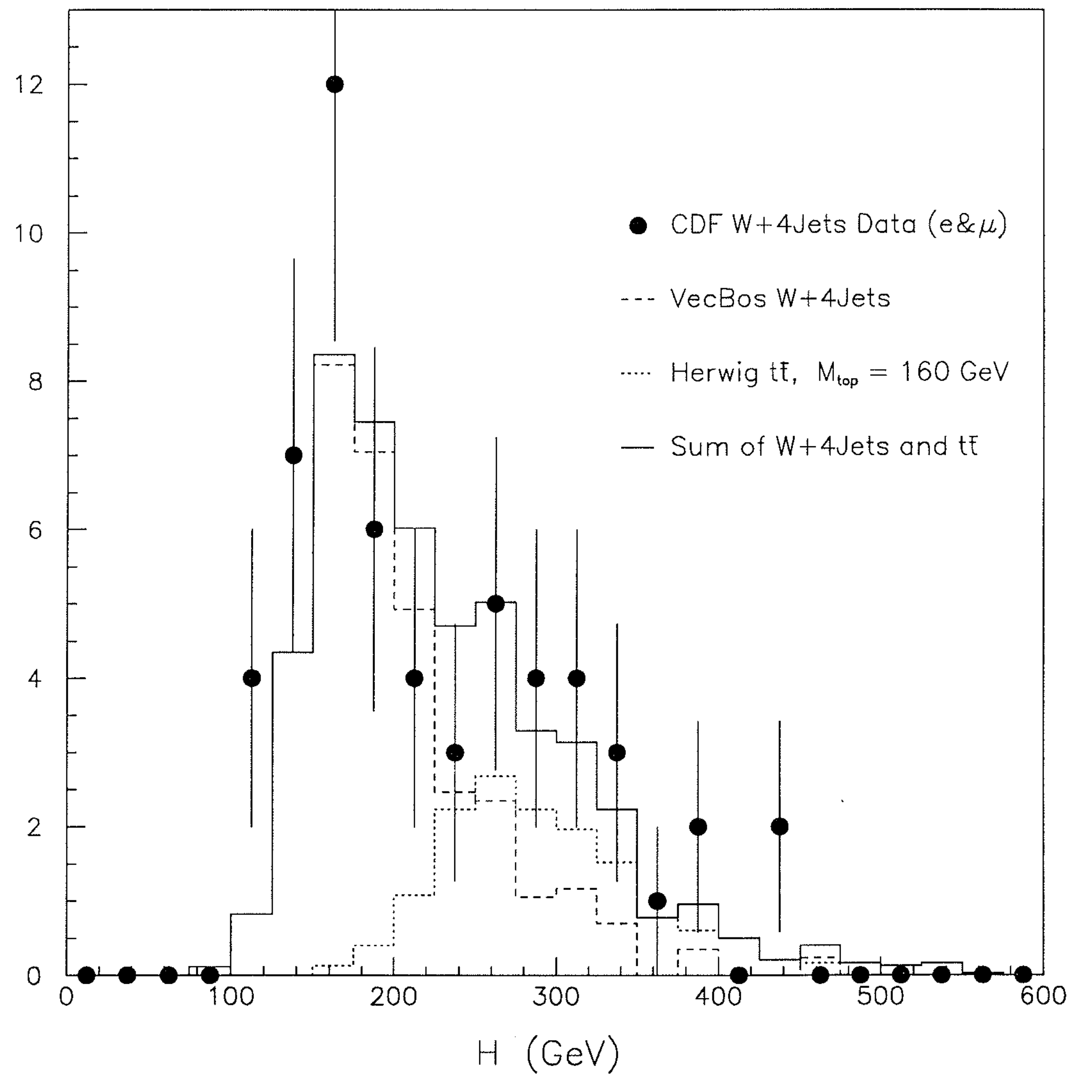

Figure 7: H distribution of CDF data (points with $\sqrt{N}$ error bars), compared with a fit to VecBos W+jets (dashed histogram) plus Herwig $t\bar{t}(160)$ (dotted histogram). The χ^2 for this fit is 9.68 for 11 dof.

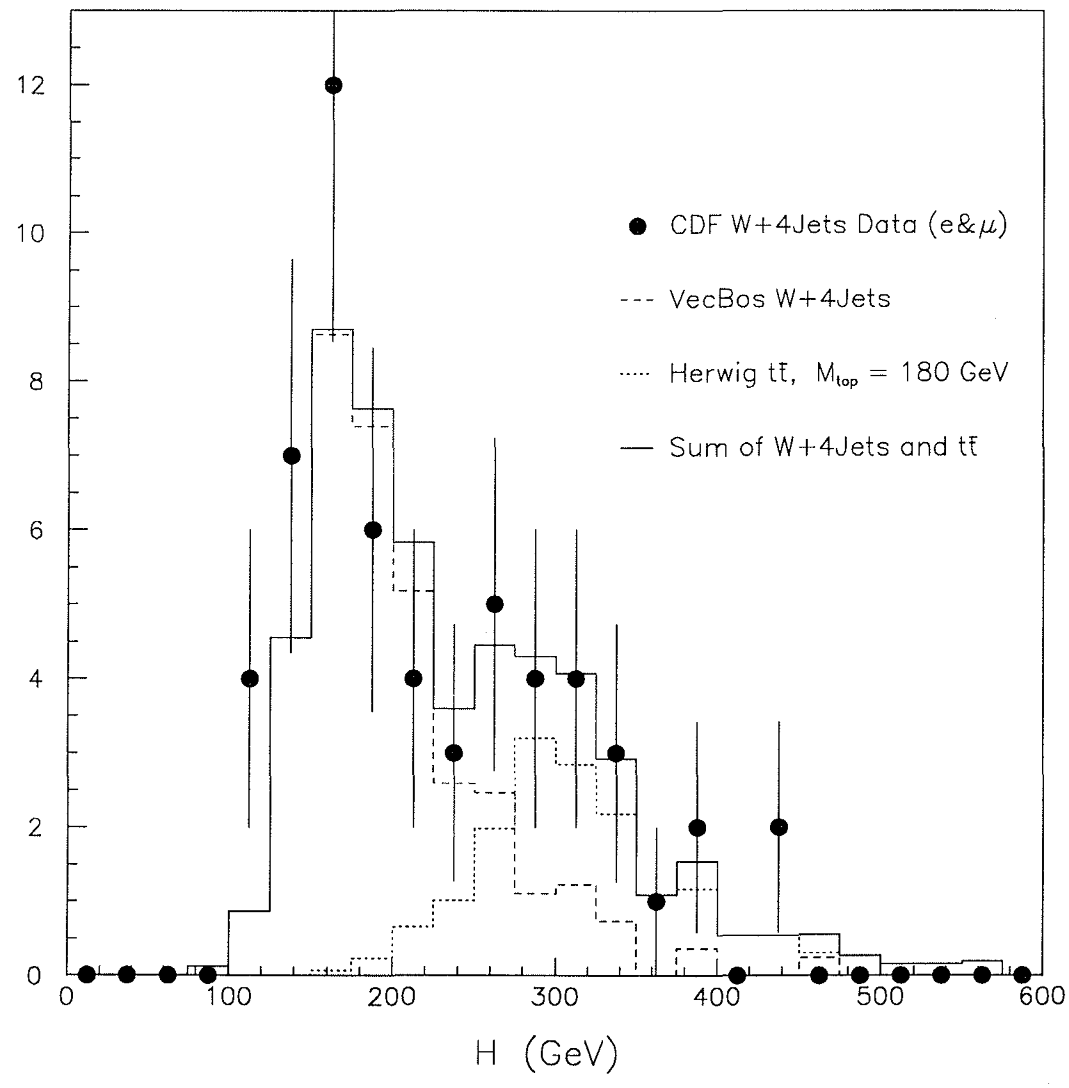

Figure 8: H distribution of CDF data (points with $\sqrt{N}$ error bars), compared with a fit to VecBos W+jets (dashed histogram) plus Herwig $t\bar{t}(180)$ (dotted histogram). The χ^2 for this fit is 6.88 for 11 dof.