Method for the Measurement of the $t\bar{t}$ Cross Section in the Lepton Plus Jets Channel Using Event Kinematics

R. Hughes, R. Marginean, E. Thomson, B. Winer
The Ohio State University

R. Erbacher, R. Roser
Fermi National Accelerator Laboratory

J. Conway
Rutgers University

CDF NOTE 6206

Version 1.1

August 7, 2003

Abstract

This note describes a method for the measurement of the cross section for $t\bar{t}$ production in Run 2 at CDF. Kinematic quantities are used to distinguish the top signal from the backgrounds in the "lepton+jets" channel. This analysis provides a new independent technique to measure the top cross section as well as provide an important cross check of the traditional b-tag analysis. This note describes the most basic analysis - a kinematic fit using the H_T variable. A variety of other kinematic and shape variables are also presented for consistency. Subsequent notes will make use of other energy and shape variables as well as make use of more advanced analysis techniques such as neural nets. Using this method, from a sample of data corresponding to 125.8 pb^{-1} of integrated luminosity, the $t\bar{t}$ cross section is measured at 5.1 $\pm$ 1.8(stat) $\pm$ 2.1(syst)pb in the W+$\geq$ 3 jet mode. In the W+$\geq$4 jet mode, the cross section, as measured in the data, is 7.7 $\pm$ 2.4(stat) $\pm$ 3.0(syst)pb.

1 Changes to the Note

1.1 Version 1.0 to 1.1

- Updated the good run list and in so doing, a few more runs were added.
- Require one and only one good lepton
- Fixed curvature correcton for muons 0.00037 vs 0.000037
- Cleaned up the isolated lepton sample used to measure the QCD bckg.
- Cross check on dphi cut – do MC's predict what we observe in the data
- Comparison of data to MC templates in the W+1,2 and 3 jet bins. We used to perform a KS test. Now we calculate a Kolmogorov distance and use it to determine a goodness of fit.
- Correct MET vs ISO for W-like events and determined a systematic uncertainty on MET vs ISO technique.
- Included all of the different backgrounds into the shapes
- Subdivided the jet energy scale systematic into its individual components to emphasize the different contributions.
- Included a systematic for Q^2
- Included a systematic on PDF

2 Motivation

In Run I, CDF discovered the top quark and began to measure its properties. While a coherent picture of this elusive quark is beginning to take shape, there are still many unanswered questions as well as inconsistencies in our current understanding. The $t\bar{t}$ cross section measurements in the lepton+jets and dilepton channels are quite different. More events are observed in the tagged lepton+jets W+2 jet bin than expected based on our "method 2" background calculations predict. Furthermore, the width of the mass peak in Run I was narrower than our Monte Carlos indicate we should expect. Therefore, it is important to measure the properties of the top quark in as many different ways as possible in order to develop a complete understanding.

This note describes the method for measuring $\sigma_{t\bar{t}}$ by performing a maximum likelihood fit using the kinematic variable H_T (defined as the scalar sum of all reconstructed lepton, jet, and missing E_T in the event) for events with an isolated lepton (e or μ), missing E_T, and three or more hadronic jets. Other kinematic and shape variables are presented here as well to indicate the degree of sensitivity and consistency in this measurement.

This work is the first step of our overall analysis strategy. From this foundation, we will expand this analysis to try to better separate signal from background. This will involve using additional N-jet bins, other kinematic and shape variables as well as b-tagging. Neural network techniques will also be exploited. By making use of as much information as there is available in the lepton+jets $t\bar{t}$ system, we expect to be able to measure the cross section to greater precision than what has been achieved to date.

Time Period	Integrated Lum	
	CEM and CMUP	**CMX**
Mar 02 - Jan 03 (Winter)	71.0 pb^{-1}	54.6 pb^{-1}
Jan 03 - Jun 03 (Summer)	54.8 pb^{-1}	54.8 pb^{-1}
Total	215.8 pb^{-1}	109.4 pb^{-1}

Table 1: Integrated luminosity by run section for this analysis

3 Data Sample

This analysis uses the data described in the official l+jets good run list as its input. The documentation for these runs can be found on the lepton+jets groups' web page http://www-cdf.fnal.gov/internal/physics/top/RunIIWjets/summer03data/index.html. This data sample includes both the 71 pb^{-1} used for the winter conferences (141544-156487) as well as 54.8 pb^{-1} of new data (up to 163527, May 31). Thus, the total sample size is 125.8 ± 7.4 pb^{-1} (109.4 ± 6.4 pb^{-1} for CMX). The offline luminosity is corrected by a factor of 1.019 [2] and the uncertainty on the integrated luminosity is take to be 5.9% [1] . The detailed numbers can be found in table 1.

4 Signal and Background Simulation

Public Monte Carlo samples and TopNtuples for these samples were used in this analysis. They are available on fcdfsgi2 as well as from CAF2 and dCache. The $t\bar{t}$ signal sample was made with both *PYTHIA* and *HERWIG*. The W+N jet background model was generated using *ALPGEN* as the leading order matrix element generator in W+N parton MC. The showering of these ALPGEN samples was handled by Herwig. To eliminate problems with double counting, the W+N parton data sample used for a particular plot matched the desired number of reconstructed jets. For example, a plot of H_T in the reconstructed W+2 jet bin of the heavy flavor background used the ALPGEN W+2 parton MC.

The QCD samples were taken from the data. Two separate samples were used and compared. The first was constructed of non-isolated ($I > 0.2$) electrons and muons. This is the sample used for the central values of the top cross section. The second sample was built from conversion electrons as an alternate sample to set the systematic error. In both cases, all other cuts like lepton ID, $\not{E}_T$, number of jets, etc. were the same.

5 Event Selection

Top-quark events in the lepton+jets channel are characterized by a high-momentum lepton, a substantial amount of missing energy due to the leptonic W decay and, in principle, 4 high energy jets due to the hadronic decay of the second W and the two b quarks in the event.

We use the electron and muon definitions established by the W and Z cross section groups in this analysis. In the muon channel, only CMUP and CMX muons are allowed. For the electron channel, only central electrons are permitted. Given that all of our events must include, by definition, a high-momentum electron or muon, the data sample should only contain events that came in on CEM, CMUP or CMX triggers. However, there are some volunteers. The requirement that the lepton came in on the appropriate trigger is made.

To clean up the sample, electron events in which the primary electron is identified as a conversion are removed. Similarly, muon events containing a cosmic ray are removed. Signal events should contain only one high-p_t lepton. Since the dilepton group has not yet provided a filter with which to cut dilepton events, we have chosen to make our own. This is not quite the same as removing dilepton events and thus our data samples are not strictly orthogonal. In the future, we will implement a true dilepton cut once it becomes available. For consistency with Run I, the same jet algorithm and jet selection criteria are used. However in this analysis, we select events after the jet corrections have been made. For completeness, a summary of all of our selection criteria can be found in table 2. Since we get countless questions on MET, we correct MET for jet corrections, curvature and muon P_T, but nothing else. Details on the event selection criteria such as cosmic ray removal and Z veto can be found on the W+jets web pages.

To implement the above selection strategy, we have used the official TopNtuples produced in offline release 4.11.1.

5.1 $\Delta\phi$ cut

We have imposed one additional cut to those listed in table 2 in order to reduce the size of the QCD background in the signal sample. With the current relatively small non-isolated lepton data sample, it is difficult to construct appropriate templates of the discriminant variable with sufficient statistics in each bin in order to obtain a smooth shape for the QCD background. Therefore, even though QCD is not our dominant background, it is still useful to reduce its contribution in our signal region in order to minimize our sensitivity to a shape that is not measured very well.

In order to motivate the cut, it is useful to look at Figures 1 and 2. These two-dimensional plots show the $\Delta\phi$ between $\not\!\!E_T$ and leading jet versus $\not\!\!E_T$ for the relevant signal and background samples. Figure 1 plots the non-isolated lepton data that we will use to model the QCD background in the 1 jet bin. The top plot of this figure is electron+1 jet data and the bottom is the muon+1 jet data. In Figure 2, the same 2D distributions are shown for $t\bar{t}$ (top) and W+jets(bottom) MC events in the reconstructed W+3 jet bin. From these plots, it is clear that by eliminating events in the region OUTSIDE $0.5 < \Delta\phi < 2.5$ for the $\not\!\!E_T$ band between 20 and 30 GeV, we can reduce the QCD background while having a minimum impact on the signal. In the signal region (W+$\geq$ 3 jets), this cut removes 40% of the QCD background while removing only 5.3 ± 1 % of the $t\bar{t}$ signal.

Table 3 contains the number of observed events in the data as a function of reconstructed jet multiplicity bin. All of the selection criteria described above have been applied. The data is broken out in terms of electron+jets and muon+jets as well as the sum. For comparison, the $t\bar{t}$ fraction as well as the number of expected $t\bar{t}$ events is also presented based on an assumed cross section of 7pb.

5.2 Cross Check on the Dphi Cut

During the preblessing talk, there was some concern that the $\Delta\phi$ cut was doing something strange to the data – namely it was cutting harder than the MC predicted. That is, if one used the efficiency of this cut for background and signal, that number of events observed passing the cut is out of line with what one would expect. In Tables 4 and 5, we present the impact of

Variable	Value
Global event quantities	
Z and cosmic veto	applied
Missing E_T	$>$ 20 GeV
# tight leptons	1
Jets	
E_T	$\geq$ 15 GeV (corrected)
$\lvert\eta\rvert$	$<$ 2
Tight electrons	
Fiducial and CEM	= 1
E_T	$\geq$ 20 GeV
P_T	$\geq$ 10 GeV/c
E/p (if $P_T \leq$ 50 GeV/c)	$<$ 2.0
E_{had}/E_{EM}	$\leq$ 0.055 + 0.00045*E
L_{shr}	$<$ 0.2
$\lvert\Delta z\rvert$	$<$ 3 cm
$\lvert\Delta x\rvert$	$>$ -1.5 and $<$ 3.0 cm
χ^2_{Strip}	$<$ 10
$\lvert\Delta z_0\rvert$	$<$ 60 cm
# axial segments	$\geq$ 3
# stereo segments	$\geq$ 3
Isolation	$<$ 0.1
Not a conversion	$\Delta\Lambda > 0.04$ and $\Delta r > 0.2$
Tight muons	
Region	CMUP or CMX
P_T	$\geq$ 20 GeV/c
E_{EM}	$<$ max(2, 2 + 0.0115*(p- 100))
E_{HAD}	$<$ max(6, 6 + 0.0280*(p- 100))
$\lvert\Delta x\rvert_{CMU}$	$<$ 3.0 cm
$\lvert\Delta x\rvert_{CMP}$	$<$ 5.0 cm
$\lvert\Delta x\rvert_{CMX}$	$<$ 6.0 cm
$\lvert\Delta z_0\rvert$	$<$ 60 cm
# axial segments	$\geq$ 3
# stereo segments	$\geq$ 3
Isolation	$<$ 0.1

Table 2: The "standard" event selection criteria used in the top group.

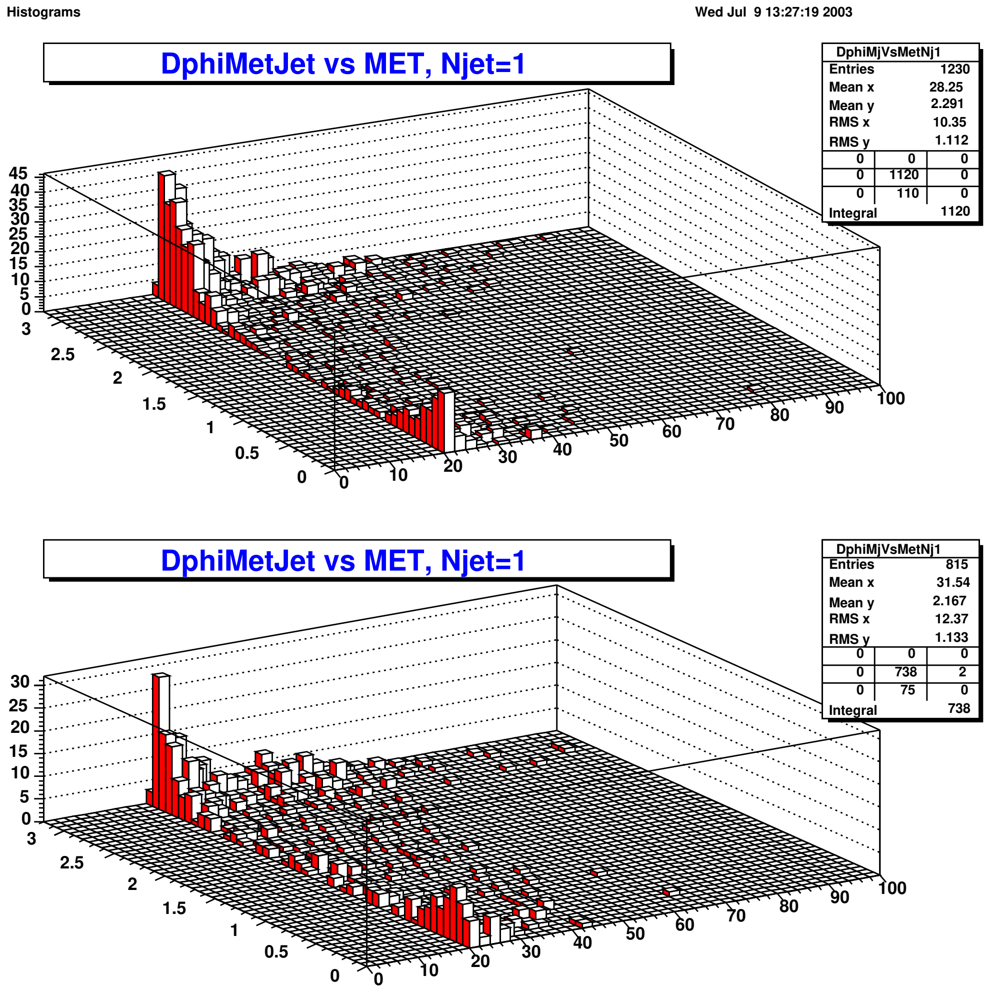


Figure 1: Distribution of $\Delta\phi$ versus $\not\!\!E_T$ for non-isolated leptons in data for the one jet bin where there are reasonable statistics. The top plot is electrons and the bottom muons. This sample is a reasonable model of the QCD background. Note the pile-up of events in the $20 < \not\!\!E_T < 30$ and $0.5 > \Delta\phi > 0$ as well as $2.5 > \Delta\phi > 3.14$. Those are the background events that we are trying to remove.

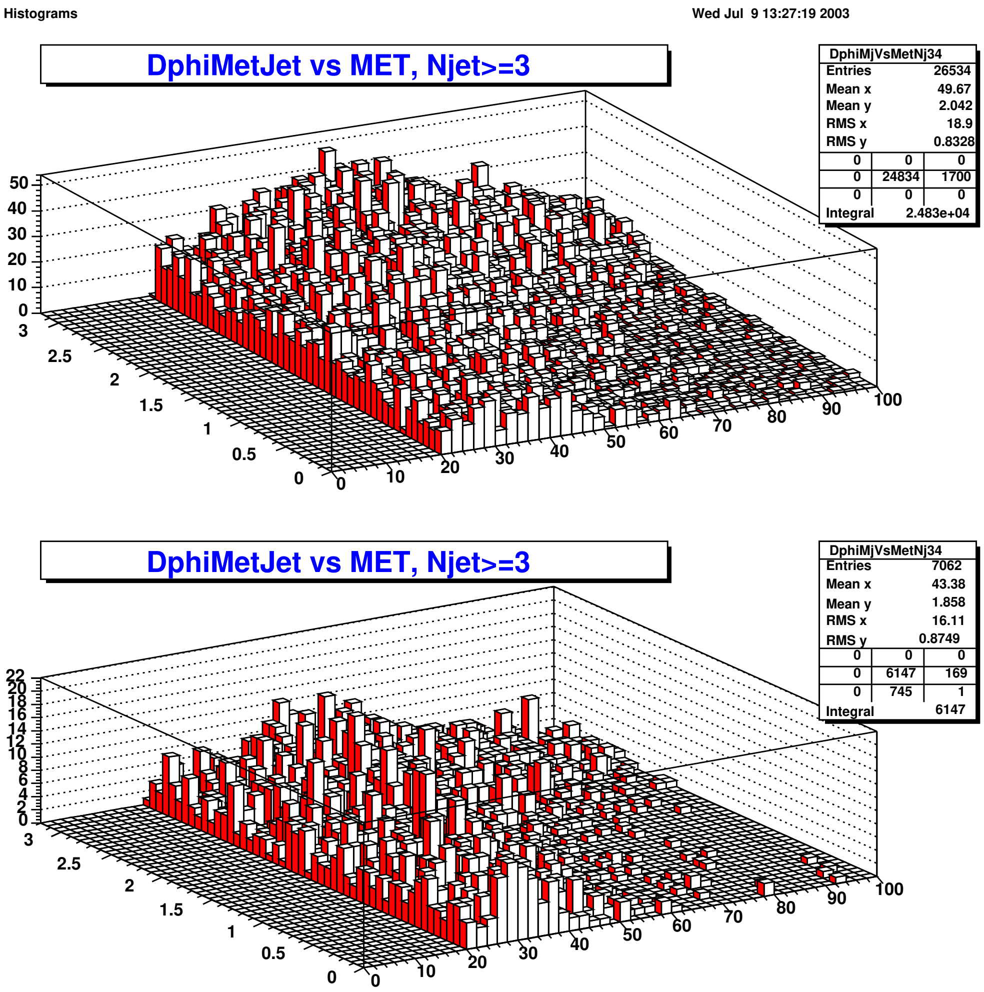


Figure 2: Distribution of $\Delta\phi$ versus $\not\!\!E_T$ in $t\bar{t}$ MC (above) and ALPGEN W+3 parton MC (below) in the $\geq$3 jet bin (signal region). Here the cut $0.5 < \Delta\phi < 2.5$ has very little impact on either $t\bar{t}$ or the dominant W+jet background.

Njets	**electrons**	**Muons**	**total**	$t\bar{t}$ **fraction**	$t\bar{t}$ **Events**
0	59655	44737	104392	0.2%	0.2
1	6105	4474	10579	4.0%	3.4
2	913	704	1617	18.5%	16.1
3	153	80	233	34.5%	29.7
4	39	18	57	32.0%	27.6
≥5	14	4	18	10.6%	9.1

Table 3: The observed and expected number of events as a function of N jets for 125.8 pb^{-1} of data. A trigger is required. For $t\bar{t}$ fraction, a cross section of 7 pb is assumed.

the dphi cut on electrons and muons separately as well as our ability to predict the outcome. The table shows the number of events before and after the dphi cut is made for each jet multiplicity bin. Electrons and muons are listed separately. The column Expected background is the percentage of background found using MET vs ISO *prior* to making the cut. Finally, the Expected results post cut, we calculate the number of events we would expect based on the formula detailed in the table caption. Thus, from the results in Tables 4(Muon) and 5(electron), we are able to predict within the statistical precision of the data sample (using the MC tools) the impact that the dphi cut has on the data. Based on this study, the cut is behaving exactly as we expect.

Njets	**After Cut**	**Before Cut**	**Exp Bckg**	**Expected post Cut**
1	4474	4976	5.1%	4469
2	704	785	4.0%	635
3	80	103	5.6%	92.1
≥4	22	26	5.6%	23.7

Table 4: A test of the dphi cut on the MUON data as a function of Njet. The "before" column shows the number of events that pass all the cuts PRIOR to the dphi cut. The "After" column imposes strictly the dphi cut on the events in the "Before" column. The "Exp Bckg" is the percentage of background expected (prior to making the cut using the MET vs ISO technique. Finally, the "Expected post cut" category is determined from the equation: (exp back)$*$ (Before Cut) $*$ dphi_eff_bckg + (1-Exp Bckg) $*$ (Before Cut)$*$ dphi_eff_W. The dphi_eff_bckg comes from the predicted background (MET vs ISO) taken as the ratio of before and after the dphi cut is made. The dphi_eff_W is the dphi efficiency taken from the ALPGEN W+jets sample.

6 Top kinematic quantities

Due to the large mass of the top quark, $t\bar{t}$ events have very large total energy, leading to large H_T compared with the non-top background processes. The main background to $t\bar{t}$ production in the "lepton+jets" channel comes from W plus multiple hadronic jets from QCD processes, called simply W+jets here. Figure 3 (upper left) shows a comparison of the shape of the

Njets	After Cut	Before Cut	Exp Bckg	Expected post Cut
1	6105	6893	9.6%	6027
2	913	1092	12.6%	944
3	153	188	15.0%	161
≥4	53	60	15.0%	52.8

Table 5: A test of the dphi cut on the ELECTRON data as a function of Njet. The "before" column shows the number of events that pass all the cuts PRIOR to the dphi cut. The "After" column imposes strictly the dphi cut on the events in the "Before" column. The "Exp Bckg" is the percentage of background expected (prior to making the cut using the MET vs ISO technique. Finally, the "Expected post cut" category is determined from the equation: (exp back)$*$ (Before Cut) $*$ dphi_eff_bckg + (1-Exp Bckg) $*$ (Before Cut)$*$ dphi_eff_W. The dphi_eff_bckg comes from the predicted background (MET vs ISO) taken as the ratio of before and after the dphi cut is made. The dphi_eff_W is the dphi efficiency taken from the ALPGEN W+jets sample.

distribution of H_T for $t\bar{t}$ and W + 3 parton samples in events with three or more "tight" jets, where a tight jet has $\eta < 2.0$ and $E_T > 15$ GeV. The $t\bar{t}$ signal has, on average, much more total transverse energy than the background W+jets signal as one would expect.

While the choice of H_T can certainly be motivated by first principles, it is not necessarily the only choice for a kinematic variable. Other choices such as the sum of all the jet energies in the event, the sum of the two leading jet E_T, $E_{T12} \equiv E_T(j1) + E_T(j2)$ and the minimum dijet invariant mass $m_{jj}^{min} \equiv \min(m_{jj})$ could also prove useful. CDF Note 6078 looked at a number of such discriminating variables in Run I. The final set of variables we use in this analysis is based on that study as well as our own investigations.

In order to decide which variable to use as the discriminant in this cross-section analysis, a single input neural network was used to discriminate $t\bar{t}$ from the dominant background of W+jets. A wide range of variables were studied. Table 6 lists some of the most promising – some are based on E_T and the rest are based on shape. The numbers in the table represent the percentage of correct classifications that are possible with the particular variable in question. For example, using the E_T of the third leading jet, it is possible to correctly classify 74.4% of a sample of W+jets and $t\bar{t}$ while the sphericity variable can only obtain the proper classification 60% of the time.

Based on the results in Table 6, there are several E_T based variables that have very similar discriminating power. This analysis could have used E_T of jet 3, H_T, $E_{T2} + E_{T3}$, and E_T of jet 2. All four of these variables have very similar separation power. In the future, the final choice of variables will be motivated based on the combined systematic and statistical uncertainty as estimated from pseudo experiments. For now, we will choose the historical favorite from Run I; the kinematic variable H_T.

While table 6 indicates the ability of a given variable to discriminate $t\bar{t}$ from the W+jets backgrounds, it is useful to look at the shape of the variables to see the degree of separation. Figures 3, 4, and 5 contain the shapes of each of the variables in table 6 for the signal region of W+$\geq$ 3 jet bin. The points represent the W+3 parton ALPGEN MC. The red histogram is $t\bar{t}$ MC. The plots are normalized to equal area.

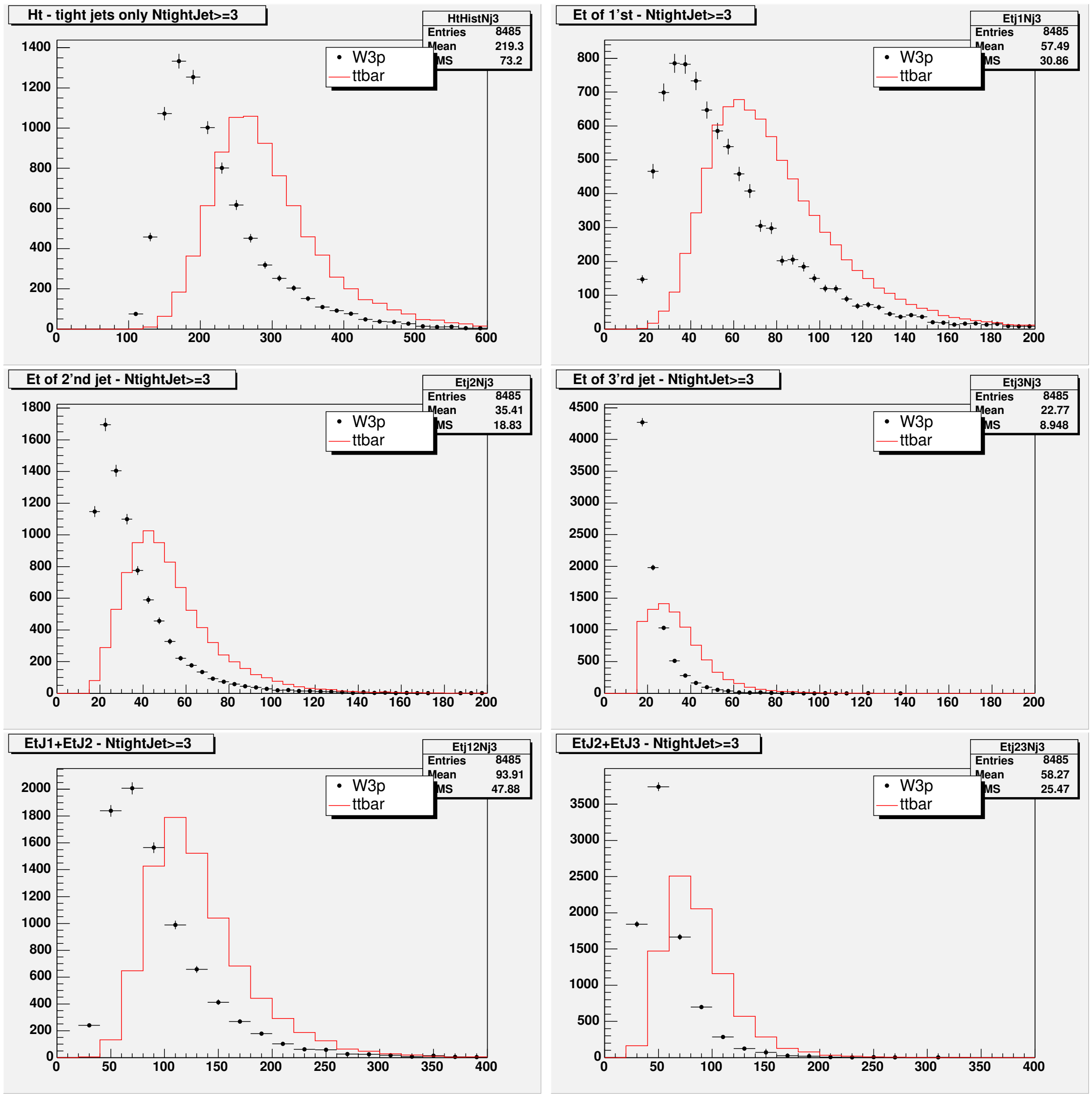


Figure 3: Comparison plots of kinematic variables using ALPGEN W+3 parton (black points) and $t\bar{t}$ (red curve) for a variety of kinematic variables in the reconstructed W+$\geq$ 3 jet bin. The curves are normalized to equal area.

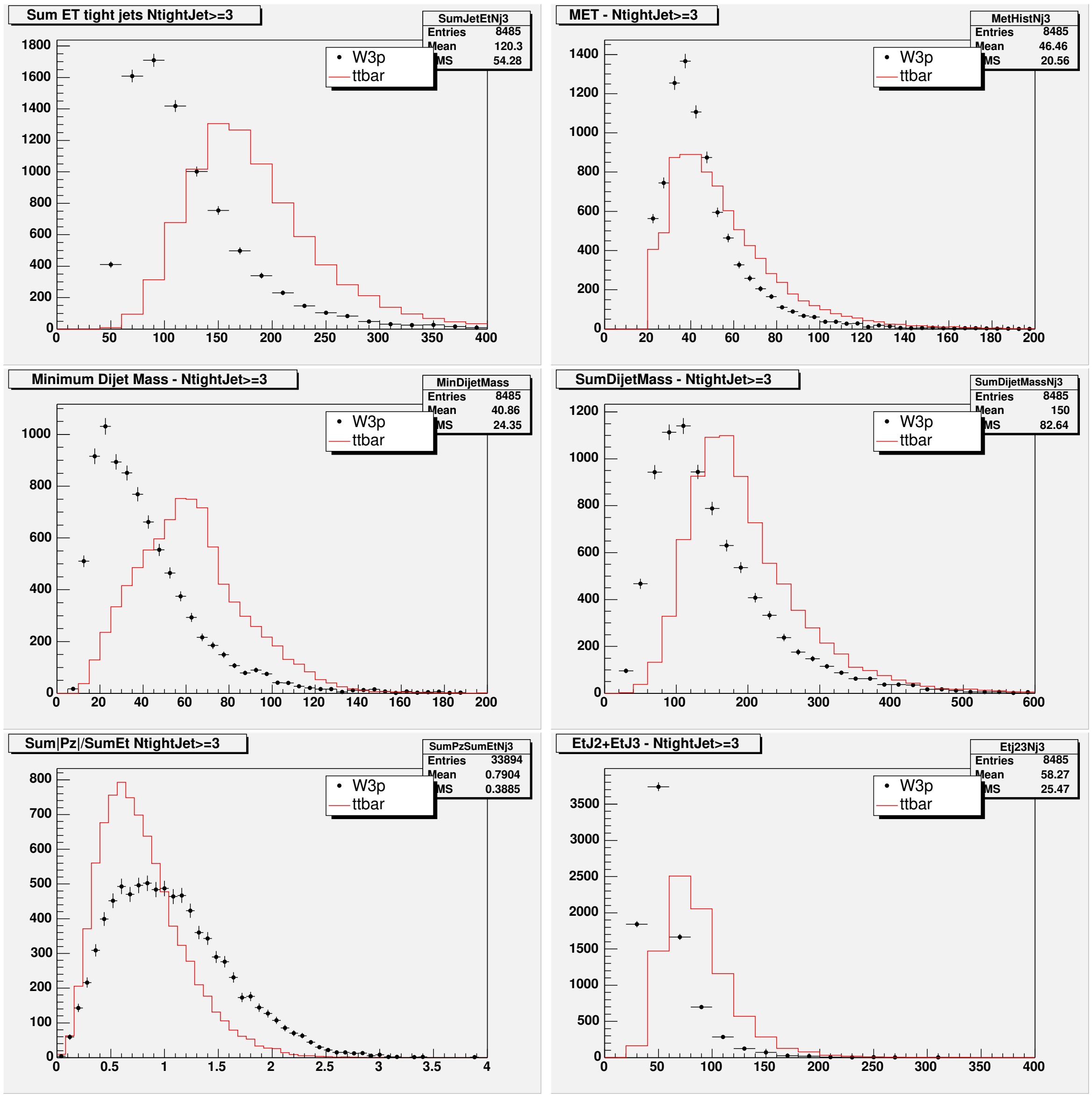


Figure 4: Comparison plots of kinematic variables using ALPGEN W+3 parton (black points) and $t\bar{t}$ (red curve) for a variety of kinematic variables in the reconstructed W+$\geq$ 3 jet bin. The curves are normalized to equal area.

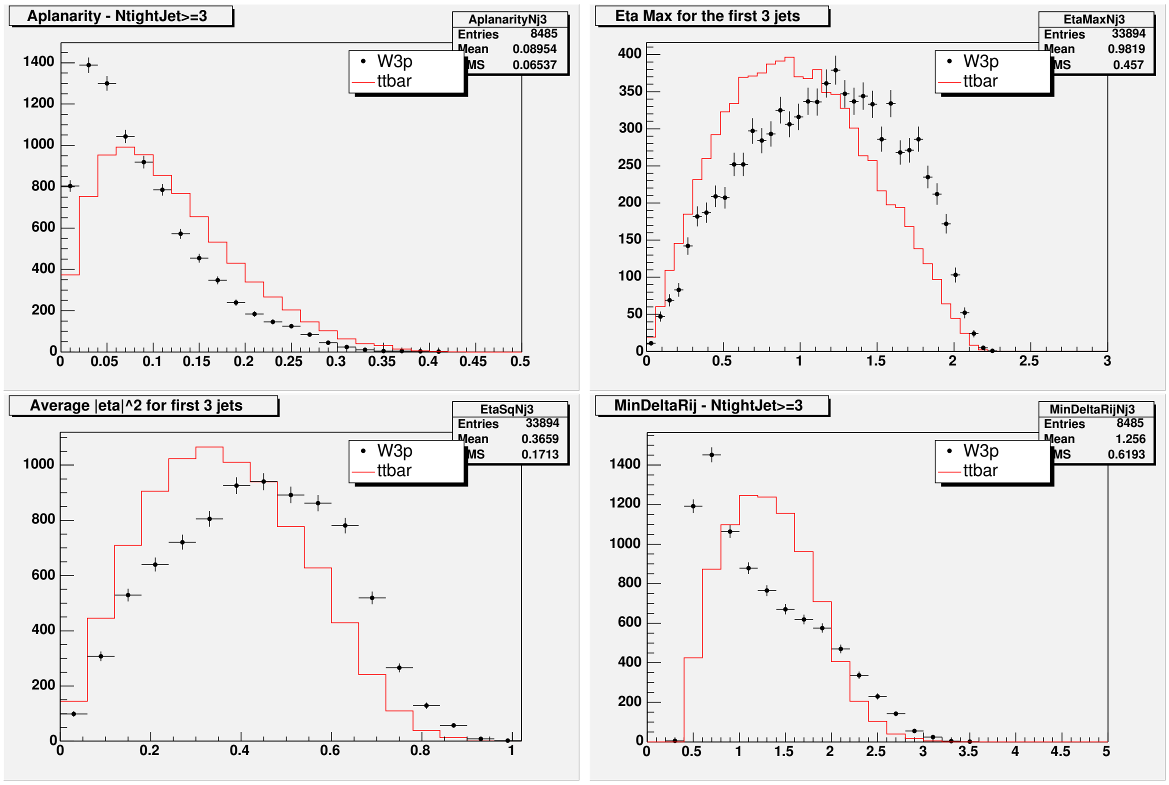


Figure 5: Comparison plots of shape variables using ALPGEN W+3 parton (black points) and $t\bar{t}$ (red curve) for a variety of kinematic variables in the reconstructed W+$\geq$ 3 jet bin. The curves are normalized to equal area.

	Herwig $t\bar{t}$ ALPGEN W+jets	**Pythia $t\bar{t}$ ALPGEN W+jets**
Et jet3	74.4% ± 0.5%	73.7% ± 0.5%
H_T	73.9% ± 0.5%	72.5% ± 0.5%
Et(2+3)	73.9% ± 0.5%	73.4% ± 0.5%
Et jet2	73.7% ± 0.5%	71.8% ± 0.5%
Min Dijet Mass	72.3% ± 0.6%	71.0% ± 0.6%
Et (1+2)	71.0% ± 0.6%	71.0% ± 0.6%
Event Inv. Mass	70.8% ± 0.6%	70.1% ± 0.6%
Et jet4	69.4% ± 0.6%	69.5% ± 0.6%
Et jet1	67.7% ± 0.6%	67.1% ± 0.6%
Sum Dijet Mass	66.6% ± 0.6%	64.9% ± 0.6%
Sum Ez/SumEt	63.3% ± 0.6%	64.0% ± 0.6%
Aplanarity	59.2% ± 0.6%	60.2% ± 0.6%
Min Jet Sep.	60.3% ± 0.6%	60.6% ± 0.5%
Sphericity	60.0% ± 0.6%	59.6% ± 0.6%
$\not{E}_T$	57.7% ± 0.6%	55.7% ± 0.6%

Table 6: Using a single input neural net to discriminate $t\bar{t}$ from W+jets, the table gives the percentage of correct classifications that are possible with a given kinematic or shape variable. The errors are statistical only.

7 Expected Sensitivities

In order to determine the expected precision of the H_T variable as well as the others presented above on the $t\bar{t}$ cross section, a series of pseudo-experiments were performed. We used the acceptance numbers based on our cuts, 125 pb^{-1} of data, and an assumed $t\bar{t}$ cross section of 7 pb. The number of QCD background events was determined from MET vs ISO in the non-isolated lepton sample. The $\Delta\phi$ cut was imposed. (The process of determining the size of the QCD sample will be explained in detail in a later section on backgrounds.) The number of W+jet events is derived from the number of observed data events minus the number of events expected from QCD and $t\bar{t}$ production. Each of the three numbers (QCD, $t\bar{t}$, and W+jets) correspond to the mean of a Poisson. The H_T is used as our primary discriminating variable but a number of other variables are also checked. Each pseudo-experiment samples the H_T distribution from $t\bar{t}$ MC, ALPGEN W+3 parton MC, and the non-isolated lepton data sample for QCD.

The results of these pseudo-experiments can be found in figure 6. The top plot in this figure contains the expected sensitivity for a variety of kinematic and shape variables in the W+$\geq$3 jet bin. The bottom plot in the same figure is for the W+$\geq$4 jet bins. The data point for each variable indicates the mean returned uncertainty from fitting the pseudo experiments. The error bars represent the one sigma range of possible returned uncertainties from the fit. Just to belabor the point and be perfectly clear, the smaller the median value and the smaller the error bars the better. From the plots, one can see, as in the neural net studies above, that H_T is not the only choice for a kinematic fit. Others like E_T of the 3rd jet and sum E_T of all the jets in the event do equally well. For the H_T variable in the $\geq$ 3 jet bin, we expect a fractional uncertainty of 25% in the $\geq$ 3 jet bin and 35% in the $\geq$4 jet bin.

8 Analysis method

In this analysis, the $t\bar{t}$ cross section comes from a binned likelihood fit of the observed distribution of H_t to a combination of

- Signal: This is modeled using HERWIG $t\bar{t}$, using a top mass value of 175 GeV/c^2.
- W+Jets Background: When fitting the W+$\geq$ 3 jets sample, the W+3 parton ALPGEN sample is used. When fitting the W+$\geq$ 4 jets sample, the W+4 parton ALPGEN sample is used.
- non-W Background: This is modeled using the non-isolated lepton sample.

Other background sources (WW, WZ, $W \to \tau\nu$, $Z \to ee/\mu\mu/\tau\tau$) are estimated from a combination of data and Monte Carlo to account for of order ~10% of the total observed events in our sample. The non-W background is constrained to the value expected using the MET versus Isolation technique as described in the next section. The result of this fit is the number of $t\bar{t}$ events (along with a statistical error) preferred by the fit. To turn this into a cross section, we need the luminosity (section 2), and the acceptance (section 10).

In the final measurement the main systematic uncertainties come from uncertainty in the jet energy scale, the choice of top mass, lepton ID efficiency (including PDF's), luminosity, choice of Q^2 scale used in the simulations, and imperfect modeling of the shapes of the kinematic distributions.

To accomplish this, we first need to determine the non-W fraction in the data sample (using the standard MET vs ISO technique) and validate the background models for QCD and W+jets. Once templates are made that describe the data well, a fit to the signal region is performed. In this case, we will present information in the W+3 jet bin, W+$\geq$ 3 jet bin and W+$\geq$4 jet bin.

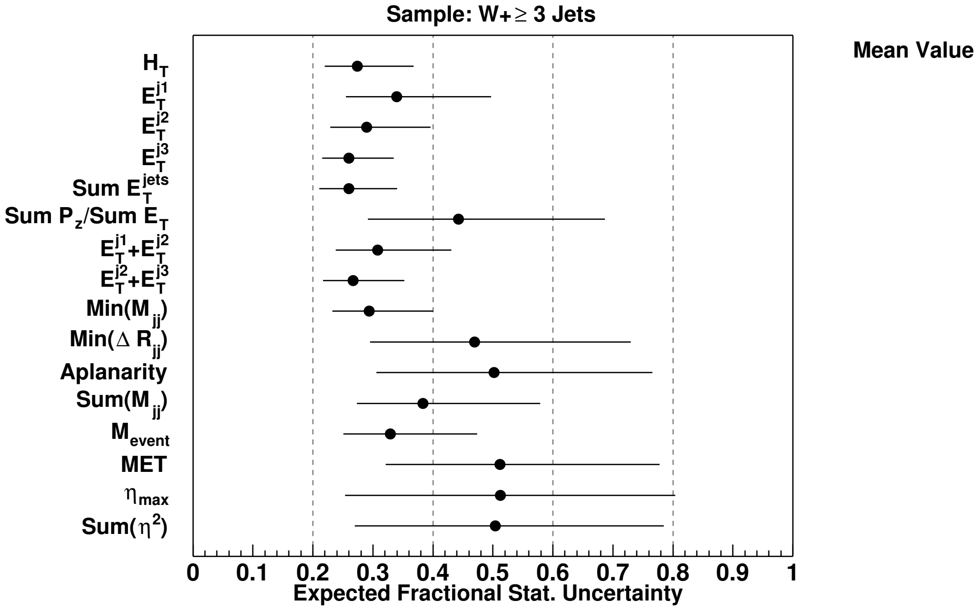
9 Backgrounds

The dominant background for this kinematic analysis is W+light quarks. This background comprises approximately 80% of the total. The remainder is made up of non-W/Z, dibosons and QCD fakes. A combination of data and MC samples were used to generate the background templates for each of these classes of backgrounds.

9.1 Validating the MC

Since this is a "shape" analysis and not a counting experiment, it is imperative that the MC's model the distributions properly. As stated before, the dominant background to $t\bar{t}$ is W+jets which we model with ALPGEN W+N parton MC. To validate the MC, we need to look in a region that is top-depleted.

The top plot in figure 7 shows the observed W+jet data(red) along with a prediction for the amount of $t\bar{t}$ (blue) one might expect as a function of jet multiplicity. The theoretical cross section used for the blue curve ($t\bar{t}$) is 7 pb. The lower plot in the same figure is the ratio of the red and blue curves. From the ratio, one can easily see that very little top is expected in the W+1 and W+2 jet bins and so these are good choices to test the background MC shapes with data. Furthermore, even in the exclusive W+3 jet bin, the amount of top expected is only



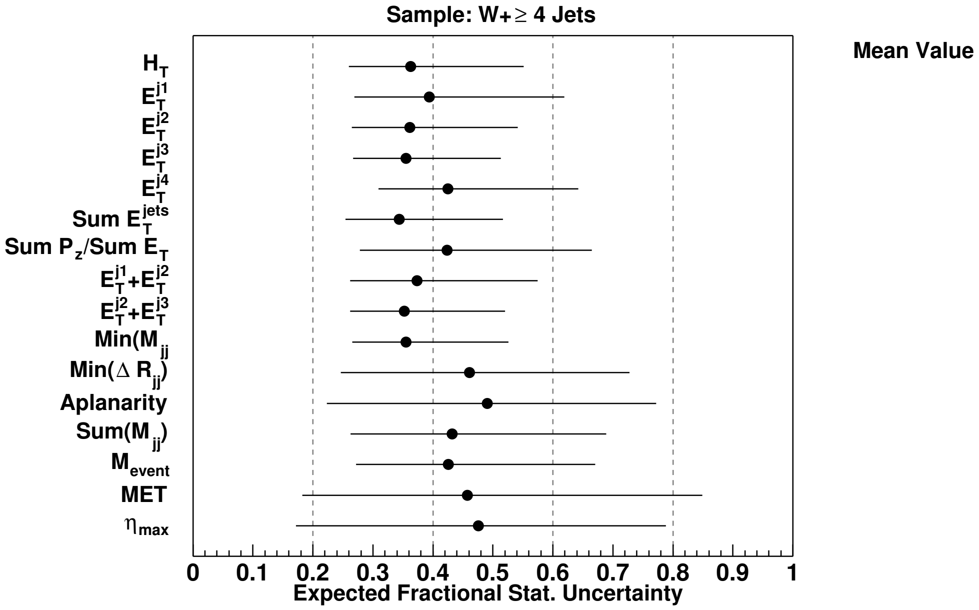


Figure 6: Expected fractional statistical uncertainty as determined by pseudo-experiments based on our current expected data sample size. The points represent the mean returned fractional uncertainty from fitting the pseudo-experiments and the error bar represents the 1 sigma range of possible returned uncertainties from the fit. The $\Delta\phi$ cut was imposed. The inputs to the PE's were $t\bar{t}$ MC, ALPGEN W+3,4 parton MC and QCD as drawn from the non-isolated lepton sample. The top plot is for $\geq$3 jet bin and the bottom is the $\geq$4 jet bin.

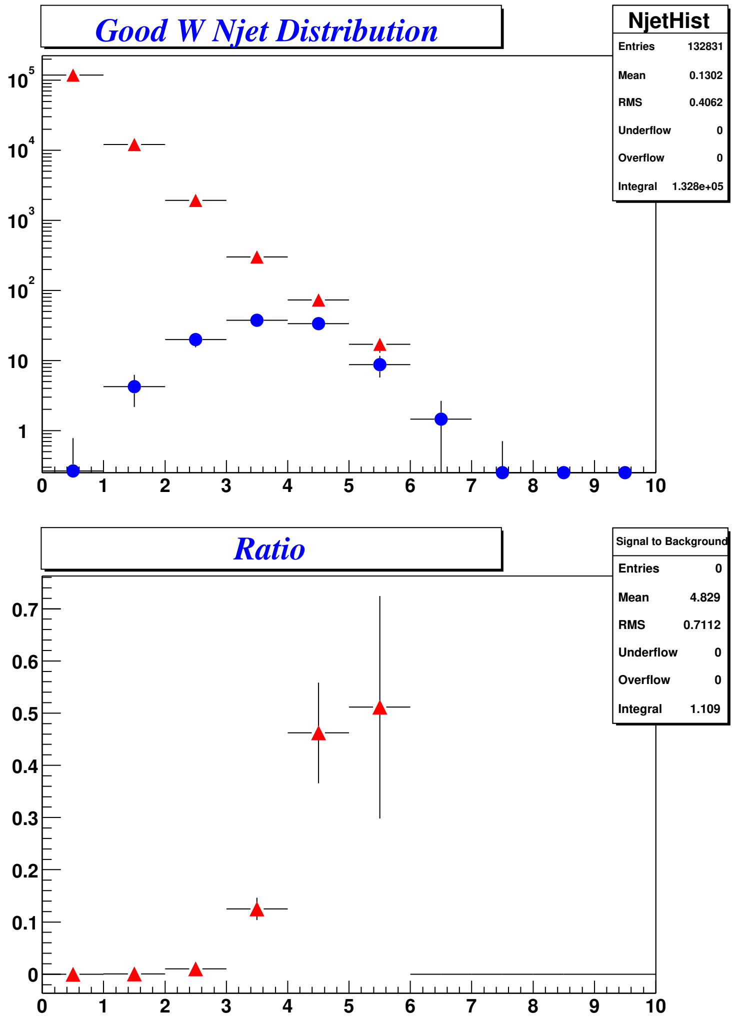


Figure 7: W+njet distributions and the comparison to $t\bar{t}$ MC.

about 12% of the total and thus can also be used as a cross check. Where we expect the bulk of the signal ($\geq$ 4 jet bin), the top to W ratio is approximately 1:1.

9.2 QCD Background

The sample that we are using to model the background from non-W is the non-isolated lepton sample. This sample requires that the isolation of the lepton be > 0.2, but otherwise, all of the other cuts, both on the lepton as well as the MET and Jet cuts, are identical. This sample presumably includes fake leptons, as well as $b\bar{b}$ in which one of the b's decays semi-leptonically to an electron or muon.

With the $\not\!\!E_T$ vs Isolation technique that has been used since the Run I "evidence" paper, we can determine the expected QCD background fraction. Table 7 contains the results of this fit on the isolated-lepton sample with the $\Delta\phi$ cut applied. We see that in the signal region, we expect approximately $7.0 \pm 1.7\%$ of the background to be attributed to QCD. The electron+jets sample has substantially more (factor of 3) QCD background than that in the muon sample.

This makes sense since there are substantial numbers of conversions in the electron sample that are not present in the muon sample.

The 7% background in the W+$\geq$ 3 jet bin is however an over estimate of the total QCD background. We need to correct for "W-like" processes that contaminate our sample. In order to correct for this, we determine the fraction of W-like events which populate each of the MET vs ISO regions. Events which we considered include We+3 partons, WW+1parton, Zee+2partons, Ztautau+2 partons and Wtau+3 partons. We+3 partons contributes the most in each bin at about 65% and the remaining account for about 10% in total. Folding all of this in, we determine the corrected QCD background to be 6.3 $\pm$ 1.6%. The relative contributions to this background is shown in the last row of Table 7.

To measure the size of the impact of the $\Delta\phi$ cut on our data, it is instructive to perform the same technique on the isolated lepton sample – except this time without making the $\Delta\phi$ cut. Those results are listed in table 8. Hence in our data sample, the $\Delta\phi$ cut reduces the QCD fraction by anywhere from 40 to 50%.

W+N jets	**Ele Bckg**	**Muo Bckg**	**Total Bckg**
1	3.8 $\pm$ 0.2%	2.9 $\pm$ 0.2%	3.4 $\pm$ 0.3%
2	6.1 $\pm$ 0.5%	2.0 $\pm$ 0.2%	4.3 $\pm$ 0.5%
$\geq$3	8.9 $\pm$ 1.5%	4.0 $\pm$ 0.9%	7.0 $\pm$ 1.7%
$\geq$3 corrected	7.7 $\pm$ 1.3%	3.1 $\pm$ 0.9%	6.3 $\pm$ 1.6%

Table 7: QCD Background fraction as a function of Njet. Numbers were determined using the MET vs ISO technique. The isolated lepton sample was used and the $\Delta\phi$ cut was applied. The *corrected* row means that the QCD background calculation was corrected for W-like event contamination in our isolated lepton sample. The value of 6.3 $\pm$ 1.6 % will be the QCD background fraction used in the fits.

W+N jets	**Ele Bckg**	**Muo Bckg**	**Total Bckg**
1	9.6 $\pm$ 0.3%	5.1 $\pm$ 0.2%	7.6 $\pm$ 0.4%
2	12.6 $\pm$ 0.8%	4.0 $\pm$ 0.4%	8.7 $\pm$ 0.9%
$\geq$3	15.0 $\pm$ 1.9%	5.6 $\pm$ 1.2%	11.7 $\pm$ 2.2%

Table 8: QCD Background fraction as a function of Njet. Numbers were determined using the MET vs ISO technique. The isolated lepton sample was used and the $\Delta\phi$ cut was NOT applied.

9.3 Testing the MC models

Unlike the traditional counting experiment analyses that have been done to measure the top cross section, this is a shape analysis. It is not sufficient that the MC get the production thresholds correct. Rather, the MC calculations must also get the production shapes correct. Hence, we must show that the data and MC shapes agree. Fortunately, there is a way to demonstrate the level of agreement. One can look at the shape of a variety of variables and see whether the predicted shapes match that in the data in a region in which there is no $t\bar{t}$ signal.

Based on the bottom plot of figure 7 and the discussion in the MC validation subsection, the W+1, 2 and 3(only) jet bins are logical choices for where to look.

In the W+1 jet sample, we perform a fit using two different combinations of backgrounds. The first includes ALPGEN W+1 parton MC plus non-isolated leptons (+1jet). The second contains the same ALPGEN sample but this time conversions are used rather than the non-isolated leptons to model the QCD fakes. In the W+2 jet bin, we use the same technique except the ALPGEN sample contains the appropriate W+2 partons. In the W+3 jet bin, there is a non negligible amount of $t\bar{t}$ – of order 12% as described above. Therefore, we also add in Herwig $t\bar{t}$ MC to the mix of ALPGEN and QCD shapes. When $t\bar{t}$ is introduced in this 3 jet bin, we do not fix its amount but rather allow it to float. For simplicity, we did not impose the $\Delta\phi$ cut in these background fits. When it comes time to fit the signal regions (W+$\geq$ 3) jets, we will impose the $\Delta\phi$ cut.

9.3.1 W+1 Jets

Looking at the H_T distribution in the W+1 jet data as shown in the top plot of figure 8, we see a somwhat poor fit. The fit undershoots H_T in the low end of the curve and over shoots H_T at the high end. Rather than just state that the resulting fit is poor by eye, we tried to quantify the difference by using the Kolmogorov-Smirnov statistic. We first performed a binned KS test but people were unhappy with this approach. It is misleading (skewed) at large KS values. Instead, we tried to flatten the distribution by essentially making an integral plot. That is, we used Igor Volobouev's suggestion and calculated the Kolmogorov distance D_d. Precisely, we fit the data to the MC in order to determine the Kolmogorov distance D_d. (This distance is calculated in ROOT using the TH1::KolmogorovTest function with option "M".) We generage pseudo-experiments according to the MC fit to the data. We then fit the pseudo experiments and determine the Kolmogorov distance D_MC between the random sample and its fit. We kept track of whether the D_MC was larger or smaller than D_d for the 1000 pseudo-experiments performed. The probability that the data agree with the MC is equal to the number of times that D_d< D_MC.

Table 9 lists the results of the Kolmogorov distance measurement described above in the W+1 jet bin where the background shape is taken from the non-isolated leptons. Based on both the plots and the table, we see that the H_T and jet Et are not well modeled. Surprisingly, neither is the lepton Et.

	Probability of Fit
H_T	0.0%
lepton Et	0.0%
$\not{E}_T$	0.0%
Et Jet1	11.4%

Table 9: Probability of data fit as determined using the Kolmogorov distance and a series of pseudo experiments for a fit of ALPGEN W+1 parton MC and non-isolated lepton+1 jet sample to the W+1 jet data for a variety of kinematic and shape variables.

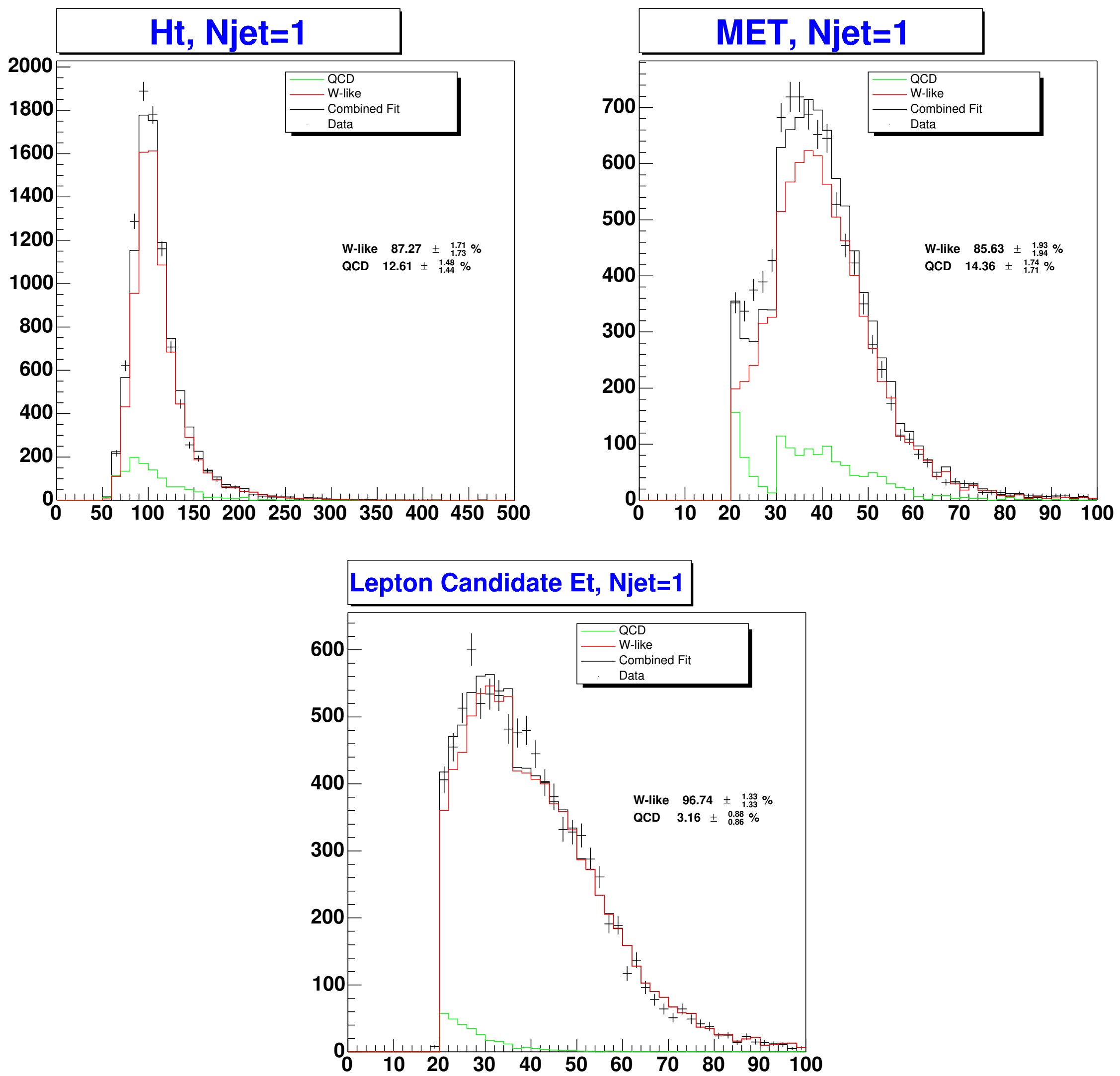


Figure 8: Monte Carlo to Data comparison of the H_T variable (top), lepton Pt (middle) and Et of the highest jet(bottom) in the W+1 jet bin. We use the summer conference data sample and a fit of W+1 parton MC(red curve) and non-isolated leptons for the QCD Fake background (green curve). The points represent the data. The black curve is the combined fit of the red and green curves and should be used to compare to the data points.

9.3.2 W+2 Jets

Similarly, we can make the same comparisons in the W+2 jet bin that were just shown above in the W+1 jet bin. Table 10 lists the KS probabilities for a set of kinematic and shape variables in the W+2 jet bin. The QCD background shape for this table is taken from the non-isolated lepton sample. In this jet multiplicity bin, the $\not\!\!E_T$ has the lowest probability of fit but it is still reasonable. All other Et and shape variables are quite compatible with the expected mix of ALPGEN W+2 parton MC and non-isolated background. Figure 9 illustrates the level of agreement for 3 of the kinematic variables, H_T(top), Et of the tight jet(middle), and lepton pt(bottom).

	Probability of Fit
H_T	42.4%
lepton Et	65.8%
$\not\!\!E_T$	24.0%
Et Jet1	14.1%
Et Jet2	90.0%

Table 10: Probability of data fit as determined using the Kolmogorov distance and a series of pseudo experiments for a fit of ALPGEN W+2 parton and non-isolated lepton+2 jet to the reconstructed W+2 jet data for a variety of kinematic and shape variables.

9.3.3 W+3 Jet, Exclusive

In the W+3 jet bin (exclusive), we expect 12% of the data to contain $t\bar{t}$. Therefore, unlike the comparisons above in the 1 and 2 jet bins, in this bin we will add in the fit $t\bar{t}$ to the ALPGEN W+3 parton sample and either the conversion+3 jets or non-isolated lepton+3 jet. The $t\bar{t}$ contribution is allowed to float in the fit. Table 11 lists the result of the KS test for a subset of variables in the W+3 and only 3 jet bin. The background used here is non-isolated leptons to model the QCD fakes. In this region, all of the kinematic and shape variables are compatible with the expected mix of ALPGEN W+3 parton MC, and background ($t\bar{t}$ and non-isolated leptons).

	Electrons+muons
H_T	47.6%
lepton Et	71.4%
$\not\!\!E_T$	86.3%
Et Jet1	45.1%
Et Jet2	98.9%
Et Jet3	74.5%

Table 11: Probability of data fit as determined using the Kolmogorov distance and a series of pseudo experiments for a fit of ALPGEN W+3 parton and non-isolated lepton+3 jets and $t\bar{t}$ to the reconstructed W+3 jet data (exclusive) for a variety of kinematic and shape variables.

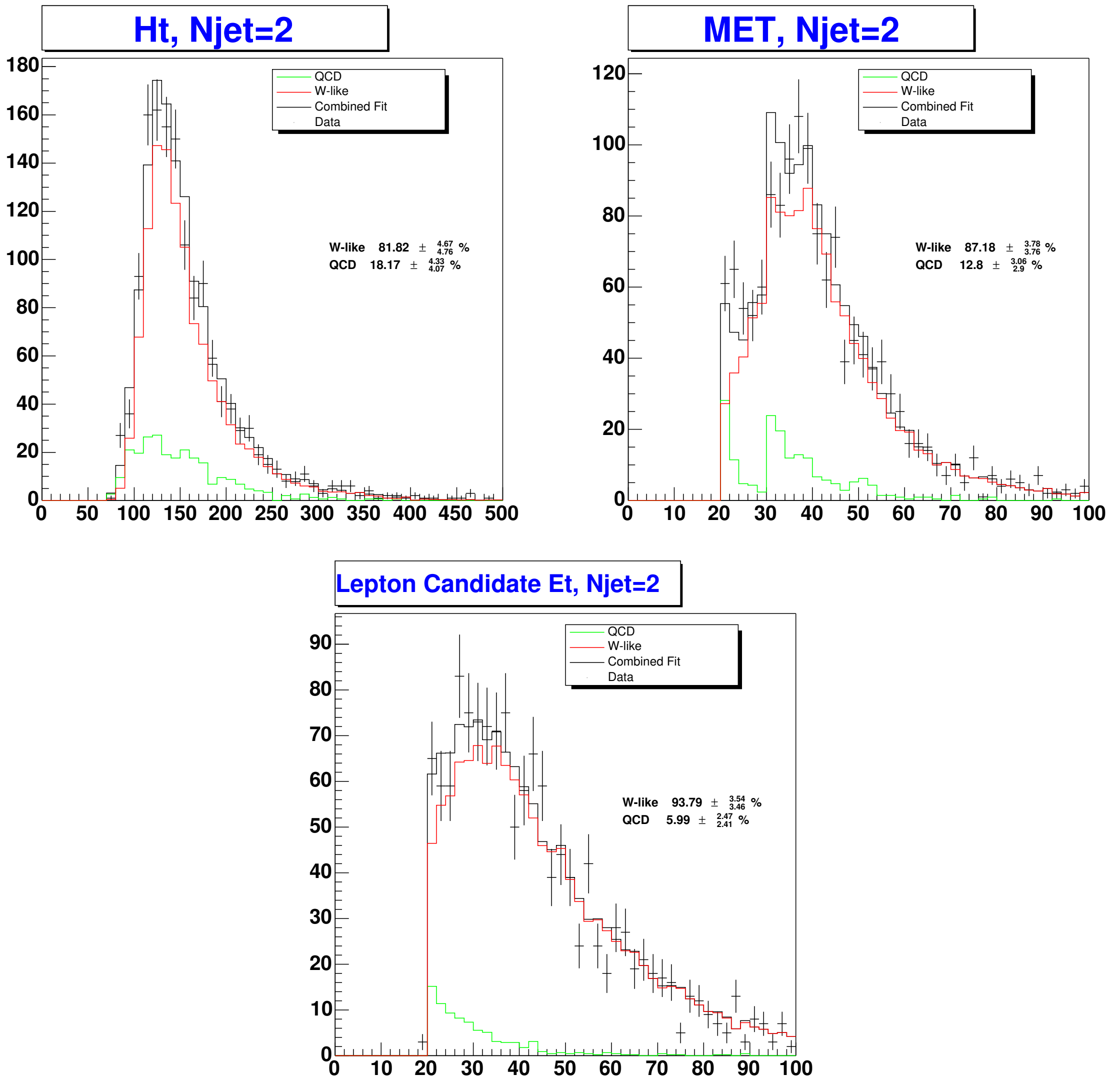


Figure 9: Monte Carlo to Data comparison of the H_T variable (top), lepton Pt (middle) and Et of the highest jet(bottom) in the W+2 jet bin. We use the summer conference data sample and a fit of W+2 parton MC(red curve) and non-isolated leptons for the QCD Fake background (green curve). The points represent the data. The black curve is the combined fit of the red and green curves and should be used to compare to the data points.

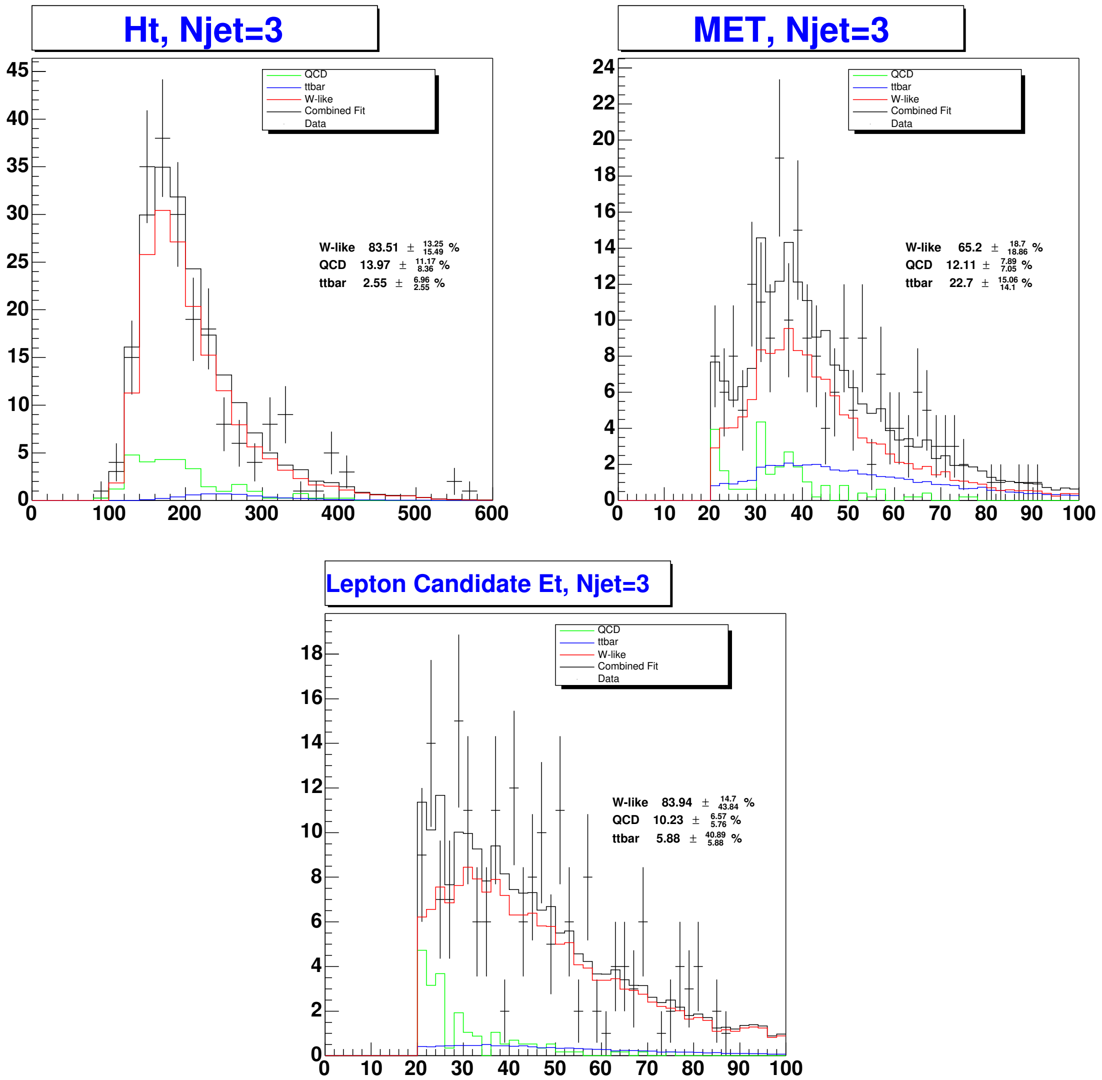


Figure 10: Monte Carlo to Data comparison of the H_T variable (top), lepton Pt (middle) and Et of the highest jet(bottom) in the W+3 jet bin. We use the summer conference data sample and a fit of W+3 parton MC(blue curve) to the backgrounds of non-isolated leptons for the QCD Fake background (green curve) and $t\bar{t}$ (red). The points represent the data. The black curve is the combined fit of the blue, red and green curves and should be used to compare to the data points.

9.4 Conclusions of MC and Data Comparisons

In summary, our MC background shapes model our data in the W+2 and W+3 jet bin quite well across a wide range of kinematic and shape variables. To some extent, the comparisons indicate reasonable agreement even in the W+1 jet bin. The ALPGEN W+N parton sample along with the non-isolated lepton sample agree sufficiently well with our current level of statistical precision that they can be used to determine a signal fraction in the ≥ 3 and ≥ 4 jet bins.

9.5 Other Backgrounds

Up to this point in the note, we have assumed that in our signal region, the background is comprised solely of W+jets and QCD fakes. This is not quite true. There are other backgrounds such as WW, WZ, and W $\rightarrow \tau\nu$. The processes Z$\rightarrow$ ee, $\mu\mu$, and $\tau\tau$ also contribute. The sum of these backgrounds are small in our signal region. Taken as a group, they are similar in contribution to the QCD fakes.

In figure 11, the H_T variable for the backgrounds listed above(red) are compared to the ALPGEN W+3 parton MC (points). While the agreement between the two are not perfect, they are very close. As a result, we could have just used the W+jets background. However, since the MC files for these backgrounds were available we included the shape of each contribution into the overall templates.

Table 12 contains all of the information on each of the MC samples that we ran on. For each MC sample in the table, we determined both a shape for each kinematic variable as well as an acceptance value. Armed with this information, we then constructed a weighted "W+jets" shape based on the ALPGEN W+jets sample as well as each of the backgrounds listed in the table. It is this constructed template that will be used in all of our final fits.

10 Fitting the Signal Region

Based on pseudo-experiments, we expect our best measurement of the signal to come from the W+$\geq$3 jet bin. To check for consistency, we also fit the W+$\geq$4 jet bin which has a better signal/noise ratio but fewer statistics. For the QCD background, we have chosen the non-isolated lepton sample. From the background section, the amount of QCD in our fits is set to 6.3%. The W+jets background is taken from ALPGEN W+3 and W+4 parton MC(for the two separate signal regions). Herwig $t\bar{t}$ is used to model the signal region. For our final answer, we will use H_T but we will investigate the behavior of the fit fraction for the other kinematic variables as well.

Figures 12 and 14 show the resulting $t\bar{t}$ fit fraction in the W+$\geq$3 and W+$\geq$4 jet bins respectively for four separate kinematic and shape variables. The $\Delta\phi$ cut is used and the QCD shape is modeled using non-isolated leptons. A summary of the fit fraction results in both signal bins can be found in figures 13 and 15. In all of these plots, it is clear that the data prefers a substantial fraction of $t\bar{t}$, in both the W+$\geq$3 and W+$\geq$4 samples. In addition, there is general agreement within each sample for the approximate fraction. The W+$\geq$3 sample prefers about 15% $t\bar{t}$, while the W+$\geq$4 sample prefers about 50%. Using the H_T variable, the $t\bar{t}$ fit fraction in the W+$\geq$3 jet bin is 0.1475 $\pm$ 0.052(stat). In the W+$\geq$4 jet bin, the resulting fit fraction is 0.5104 $\pm$ 0.164(stat).

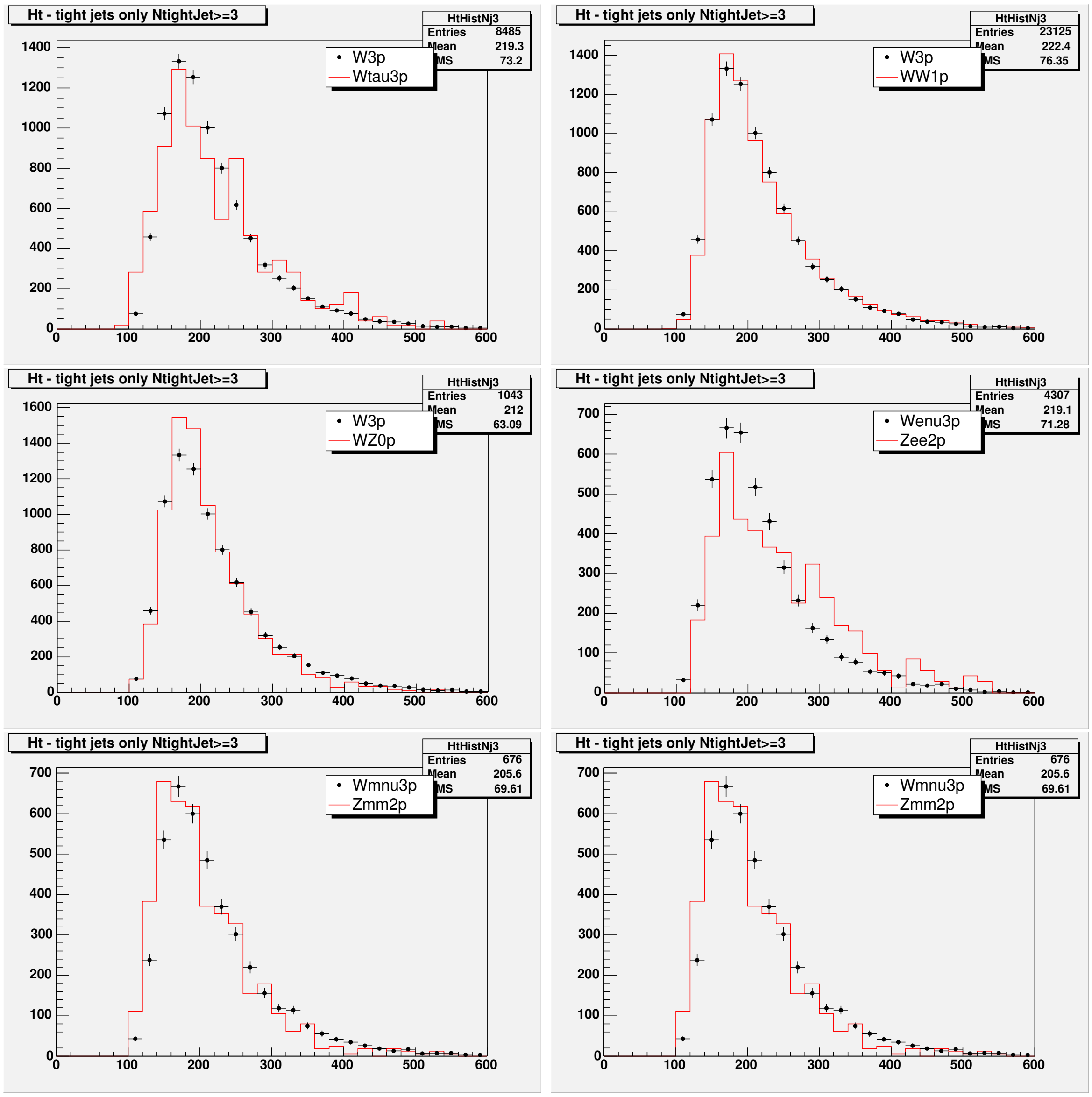


Figure 11: H_T shape comparison between ALPGEN W+3 parton MC (points) and 6 other backgrounds(red) listed in the text.

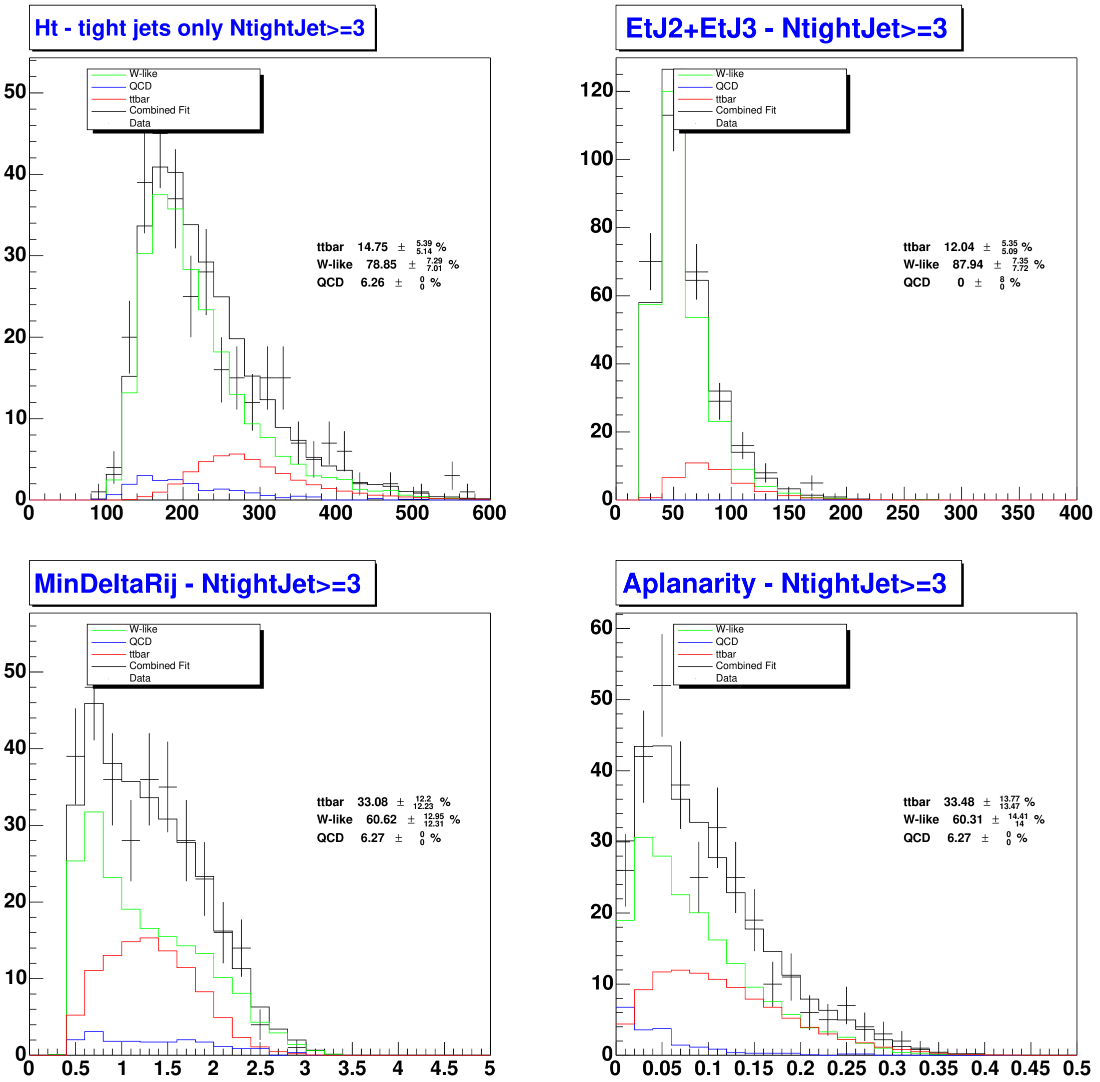


Figure 12: Fits of the data to W+jets and QCD background templates for a range of kinematic and shape variables in the W+$\geq$ 3 jet bin

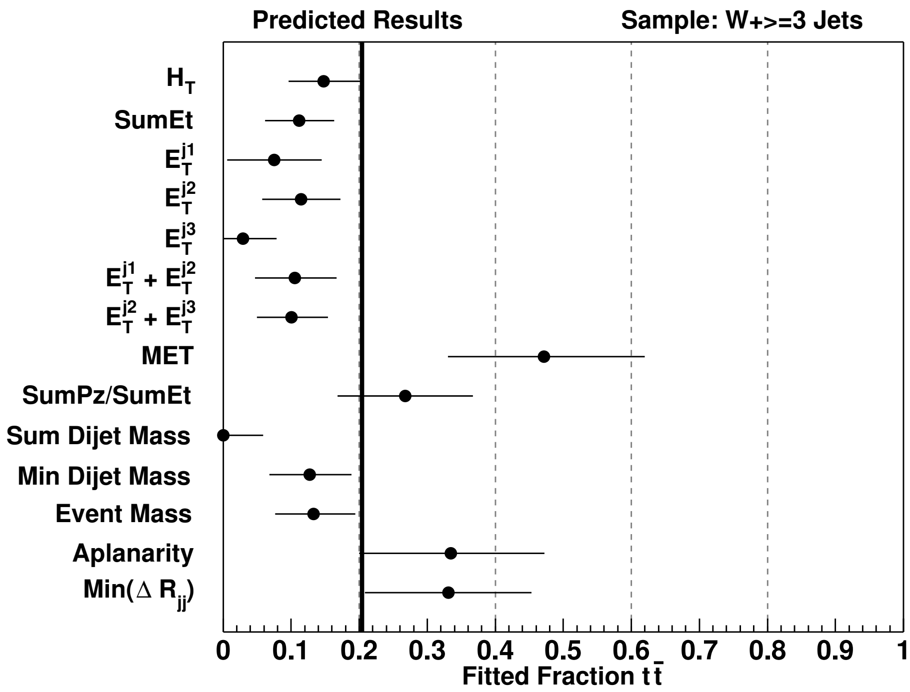


Figure 13: Summary plot of the fit fraction of $t\bar{t}$ results for a variety of kinematic and shape variables in the W+≥3 jet bin. Non-isolated lepton sample was used as the QCD model and the $\Delta\phi$ cut was applied.

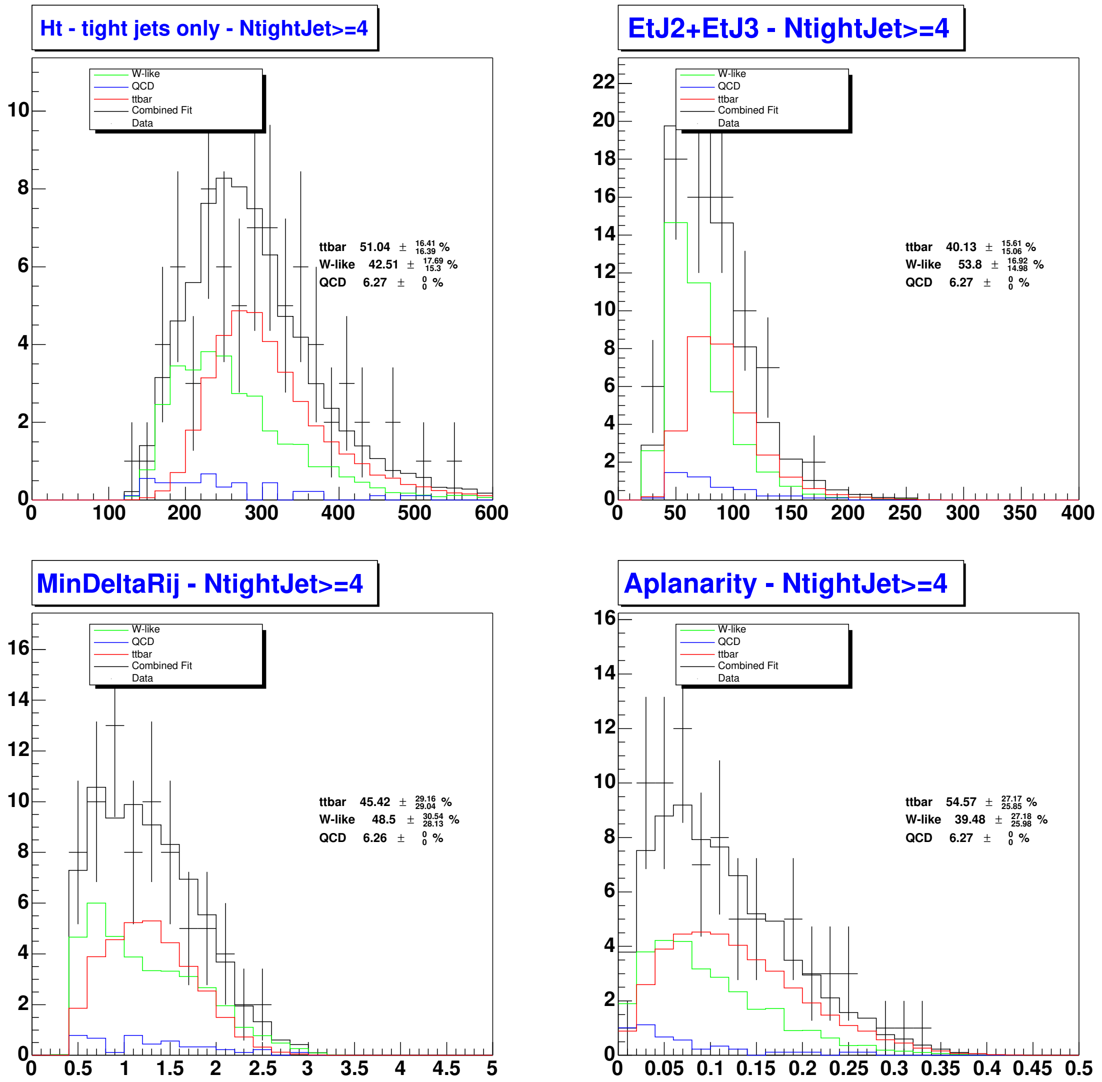


Figure 14: Fits of the data to W+jets and QCD background templates for a range of kinematic and shape variables in the W+$\geq$ 4 jet bin.

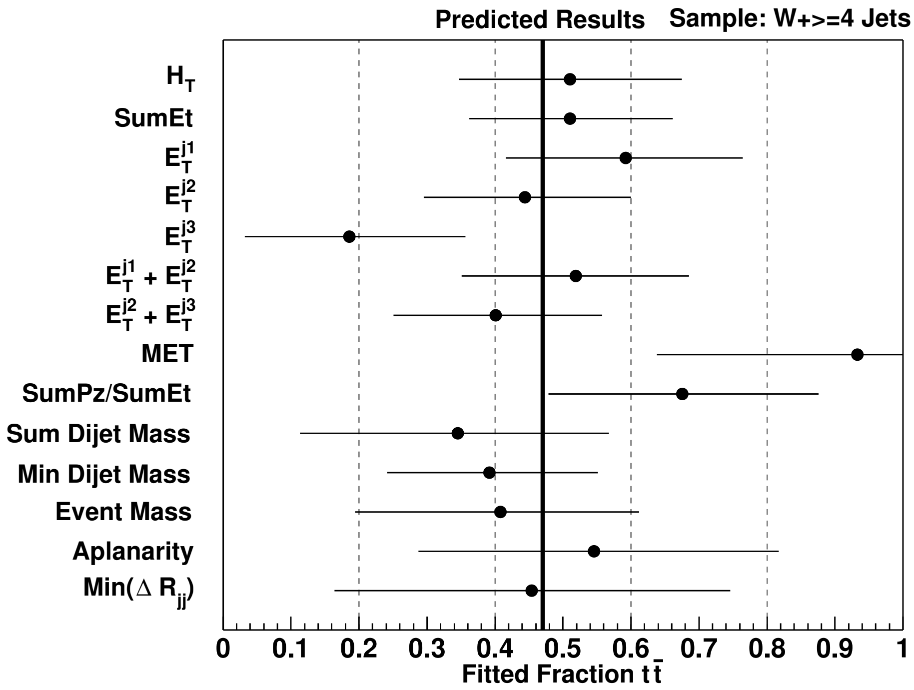


Figure 15: Summary plot of the fit fraction of $t\bar{t}$ results for a variety of kinematic and shape variables in the W+≥4 jet bin. Non-isolated lepton sample was used as the QCD model and the $\Delta\phi$ cut was applied.

MC Sample	MC σ	ℓ_{type}	N$_{gen}$	N$_{Acc}$	$F^{acc}_{data/mc}$	L_{data}	Total
W($e\nu$)+3p	89.9±0.6	CEM	268682	4311	0.955 ± 0.016	125.8	173.3
W($\mu\nu$)+3p	89.9±0.6	CMUP	293814	2724	0.773 ± 0.027	125.8	81.1
		CMX		1270	0.973 ± 0.023	109.4	41.4
W($e\nu$)+bb+1p	1.11±0.01	CEM	236005	2806	0.955 ± 0.016	125.8	1.6
W($\mu\nu$)+bb+1p	1.11±0.01	CMUP	171610	1114	0.773 ± 0.027	125.8	0.7
		CMX		544	0.973 ± 0.023	109.4	0.3
W($\tau\nu$)+3p	89.9±0.6	CEM	261917	219	0.955 ± 0.016	125.8	9.0
		CMUP		128	0.773 ± 0.027	125.8	5.3
		CMX		62	0.973 ± 0.023	109.4	2.3
WZ+0p	2.37±0.01	CEM	191011	554	0.955 ± 0.016	125.8	0.8
		CMUP		332	0.773 ± 0.027	125.8	0.4
		CMX		153	0.973 ± 0.023	109.4	0.2
WW+1p	4.38±0.00	CEM	200000	2926	0.955 ± 0.016	125.8	7.7
		CMUP		1720	0.773 ± 0.027	125.8	3.7
		CMX		712	0.973 ± 0.023	109.4	1.7
Z(ee)+2p	23.3±0.1	CEM	287734	335	0.955 ± 0.016	125.8	3.3
Z($\mu\mu$)+2p	23.3±0.1	CMUP	292446	578	0.773 ± 0.027	125.8	4.5
		CMX		280	0.973 ± 0.023	109.4	2.4
Z($\tau\tau$)+2p	23.3±0.1	CEM	296823	283	0.955 ± 0.016	125.8	2.7
		CMUP		172	0.773 ± 0.027	125.8	1.3
		CMX		76	0.973 ± 0.023	109.4	0.7
ttbar	7	CEM	398037	17969	0.955 ± 0.016	125.8	37.4
		CMUP		11076	0.773 ± 0.027	125.8	18.9
		CMX		4207	0.973 ± 0.023	109.4	7.9
single top	3	CEM	415364	2527	0.955 ± 0.016	125.8	2.2
		CMUP		1485	0.773 ± 0.027	125.8	1.0
		CMX		627	0.973 ± 0.023	109.4	0.5

Table 12: Other Background quantities for >=3Jet backgrounds.

10.1 Cross Check – Letting the QCD Bckg Float

Given the limited statistics available for this analysis, it is difficult to be certain that we are modeling the QCD shape properly. Monte Carlo's offer little guidance in this fake background and extracting information from the data is tenuous. Therefore, even after all the other checks have been done on the QCD shape and agreement in the W+1, 2 and 3 jet bins, we perform this last cross check. Here, we redo the fits and allow the QCD fraction to float to whatever value it prefers. The results of the fits are found in Figure 16. The top plots show the fits to the H_T variable in the ≥ 3 and $\geq$4 jet bins respectively. The bottom plots provide a summary in each jet bin over a wide variety of kinematic and shape variables. One sees excellent agreement between these results and those in which the QCD background was fixed to 6.3%. In the ≥ 3 jet bin, the $t\bar{t}$ fraction when the QCD was constrained was 14.75% compared to 15.42%.

Similarly, in the ≥ 4 jet bin, the fraction is 52.1% (floating) compared to 51.04% (fixed).

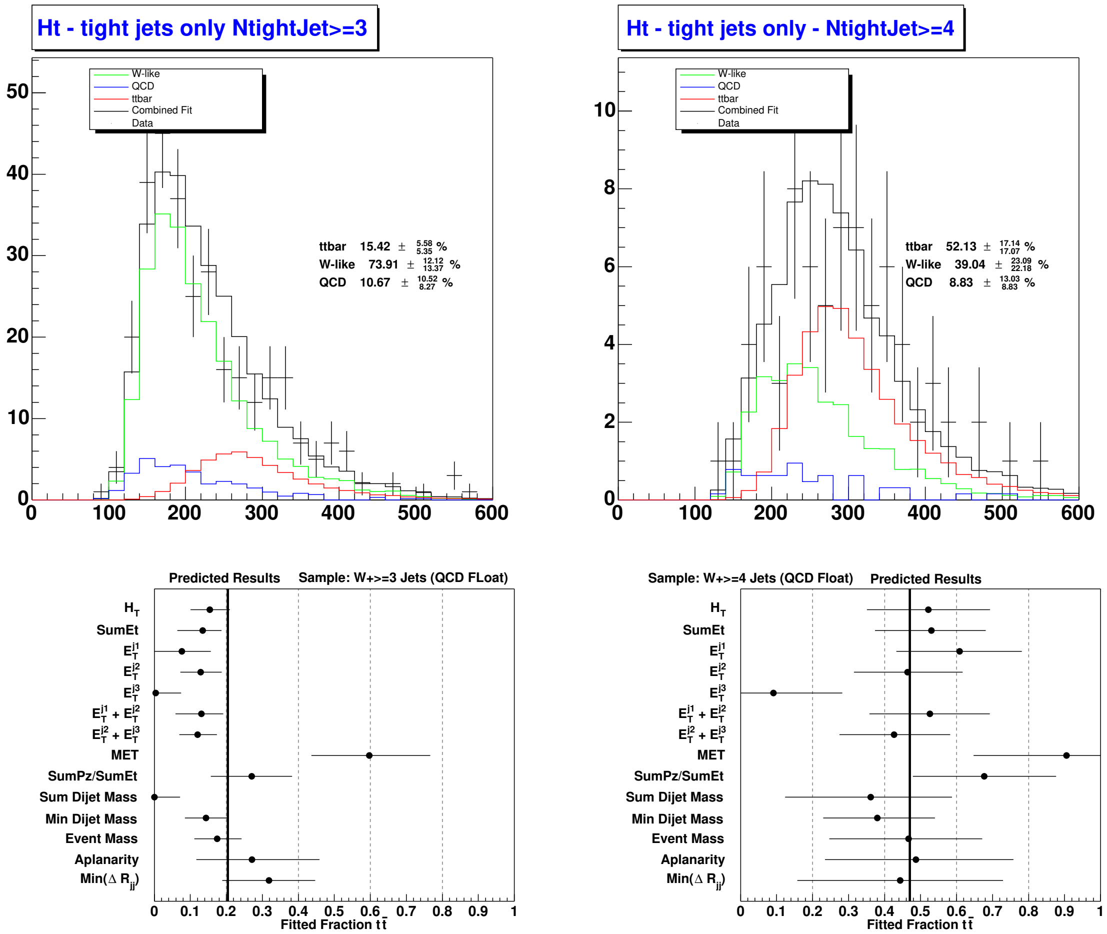


Figure 16: Fits of the data to W+jets and QCD background templates for a range of kinematic and shape variables in the W+$\geq$ 3(left) and $\geq$ 4(right) jet bins respectively. In these plots, the QCD backgrounds are allowed to FLOAT. The top plots are the fits to H_T in the two signal regions, the bottom plots are a summary of the fit fractions for a variety of variables.

Hence, this result is quite stable to the QCD background fraction. This result is true not just for H_T but for the other kinematic and shape variables as well.

10.2 Cross Check – Fitting the Signal region with and without top

A second check that we can perform is to fit our signal region with and without $t\bar{t}$ and then look at the KS test, as described earlier to see how well we can do. This gives us two pieces of information. It tells us how consistent our nominal fit is with the data and it tells us whether the fit really prefers top or not. From the Table 13, it is clear from the KS test that the fits in both the ≥ 3 and the $\geq$4 jet bins, that the best fit is with a fixed amount of QCD and with $t\bar{t}$. In that case, the KS test is 19%. If QCD is allowed to float, the KS test is slightly worse in the 3 jet bin and slightly better in the 4 jet bin. Letting QCD float has no appreciable impact on the $t\bar{t}$ fraction that we measure. Any fit to the data WITHOUT including $t\bar{t}$ yield a KS test of

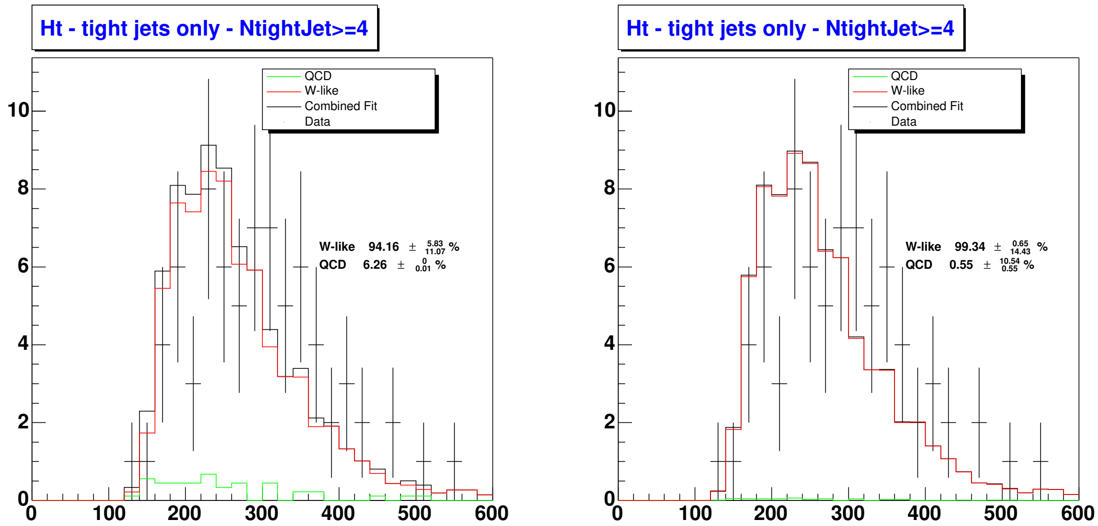


Figure 17: Fits in the W+4 jet bin without ttbar. The top plot has a fixed amount of QCD and the bottom plot has a floating QCD fraction. In both cases, the KS test is less than 1%.

less than 1% and shows that the fit is incompatible with the data. This can be seen by eye as well in Figure 17

Case	W+≥3	W+≥4
With $t\bar{t}$, Fixed QCD (nominal)	19%	60.6%
With $t\bar{t}$, QCD Float	17.9%	62.2%
No $t\bar{t}$, Fixed QCD	1.4%	0.1%
No $t\bar{t}$, QCD Float	1.0%	0.4%

Table 13: KS test in the fit regions for our nominal fit (With ttbar and fixed amount of QCD) as well as when we allow the QCD to float and do not include $t\bar{t}$

11 Deriving the Cross Section

The $t\bar{t}$ cross section is calculated based on the following formula.

$$\sigma(t\bar{t}) = \frac{\mathcal{F} \cdot N}{\mathcal{L} \cdot A}$$

Where $\mathcal{F}$ is the $t\bar{t}$ fit fraction, N is the number of observed events in the data, $\mathcal{L}$ is the integrated luminosity, and A is the acceptance.

The acceptance can be broken down into the following form:

$$\text{Acceptance} = [\varepsilon_{cmx}(t\bar{t}) \cdot \mathcal{L}_{cmx} + \varepsilon_{cmu}(t\bar{t}) \cdot \mathcal{L}_{cmu} + \varepsilon_{cem}(t\bar{t}) \cdot \mathcal{L}_{cem}] \cdot \varepsilon_{common} \cdot \varepsilon_{d\phi} \cdot \varepsilon_{jet}$$

The values for $\varepsilon_{cmx}(t\bar{t})$, $\varepsilon_{cmu}(t\bar{t})$, $\varepsilon_{cem}(t\bar{t})$, and ε_{common} come from CDF Note 6574 [5]. The numbers for each of the quantities in the equations above are listed for reference in Table 14.

Putting all of the numbers together, the final results for the cross section (with statistical errors only) can be found in table 15.

Variable	Value
$\mathcal{L}_{cmx}$	109.44 ± 6.44 pb^{-1}
$\mathcal{L}_{cmu,cem}$	125.82 ± 7.41 pb^{-1}
$\varepsilon_{d\phi}$	0.947 ± 0.001
ε_{jet}**(n≥3)**	1.0
ε_{jet}**(n≥4)**	0.554 ± 0.002
$\varepsilon_{cmx}(t\bar{t})$	0.0091 ± 0.004
$\varepsilon_{cmu}(t\bar{t})$	0.0244 ± 0.0008
$\varepsilon_{cem}(t\bar{t})$	0.0466 ± 0.0009
ε_{common}	0.951 ± 0.062

Table 14: Acceptance and luminosity numbers used in deriving the cross section.

Sample	Observed Evts	$t\bar{t}$ **fraction**	$\sigma(t\bar{t})$
W+≥3 Jets	308	0.147 ± 0.053	5.1 ± 1.8 pb
W+≥4 Jets	75	0.510 ± 0.164	7.7 ± 2.4 pb

Table 15: The $t\bar{t}$ production cross section in the W+≥ 3 and W+≥4 jet bins as well as the fraction of $t\bar{t}$ events found in our data.

12 Systematics

Inspection of the above cross section formula shows that we will be sensitive to systematic effects from the following sources:

- The uncertainty on the luminosity $\mathcal{L}$ of 5.9%[1].
- Systematic effects on the acceptance A. These fall into sub-categories of those that affect the lepton ID, those that affect the efficiency for passing the MET and Jet E_T cuts.
- Systematic effects on the determination of the $t\bar{t}$ fraction $\mathcal{F}$ in the fitting of the H_T distribution. These are shape systematics. These include effects that the modify the H_T distribution from $t\bar{t}$ and W+jets. In addition, there are systematics associated with our background model. We use non-isolated leptons as our primary non-W background model, and we check our sensitivity to this by using the conversion electron sample to model the non-W background.

Note that some effects will show up in more than one category. For example, the uncertainty in the Jet energy scale will lead to both a change in the shape of the H_T distribution, as well as a change in the $t\bar{t}$ acceptance (as more or fewer jets will pass the Jet E_T cut).

12.1 Determination of shape systematics

To determine the shape-based systematics, we use a pseudo-experiment (PE) technique. We will use the determination of the systematic due to the uncertainty of the top quark mass as an example.

12.1.1 Shape systematic from variation of the assumed top quark mass

In fitting our data, we assume that the top quark mass is 175 GeV/c^2. Our default model for fitting the $W+ \geq 3$ jet sample then includes 3 samples:

- A PYTHIA $t\bar{t}$ sample, with the top quark mass set at 175 GeV/c^2. PYTHIA was used rather than the standard HERWIG since we had samples of PYTHIA at different top quark masses readily available. As is shown later, there is very little difference in H_T shape between PYTHIA and HERWIG.
- An ALPGEN $W + 3$ parton sample.
- A non-isolated sample, obtained from the data.

However, since we only know the top mass to within about 5 GeV/c^2, it is entirely possible that the true top quark mass is really 170 GeV/c^2, or 180 GeV/c^2. The systematic is determined using the following procedure:

1. We generate PE's using a PYTHIA 170 (or 180) GeV/c^2 $t\bar{t}$ sample, the standard ALPGEN $W + 3$ parton sample, and the data non-isolated sample.
2. We fit each PE twice:
   - Using the SAME model as applied to the data (that is the 175 GeV/c^2 $t\bar{t}$ sample, the standard ALPGEN $W + 3$ parton sample, and the data non-isolated sample).
   - Using the APPROPRIATE model for this PE (that is the 170 GeV/c^2 $t\bar{t}$ sample, the standard ALPGEN $W + 3$ parton sample, and the data non-isolated sample).
3. We take the difference of the two fits, and plot the result for many PE's. The systematic is the mean of this distribution.

The plots of this difference are shown in Figure 18. The plot on the left shows the difference in the fitted fraction between using the standard templates (with a top mass of $175 GeV/c^2$, and using the generated templates with a top mass of $170 GeV/c^2$. The $t\bar{t}$ fraction used in the pseudoexperiments was 0.19, so the mean of -0.022 corresponds to a -11.6% systematic on the fraction. The plot on the right shows the difference in the fitted fraction between using the standard templates (with a top mass of $175 GeV/c^2$, and using the generated templates with a top mass of $180 GeV/c^2$. Using this higher top mass yields a mean difference of +0.0084, which corresponds to a -4.4% systematic on the fraction. We symmetrize this result and obtain a systematic of ±8% on the fit fraction due to varying the top quark mass.

12.1.2 Shape systematic from variation of the E_T scale

Our dependence on the E_T scale is significant, since this will come into play in both the jet E_T's, as well as in the determination of the MET. Jet E_T's are corrected using Level 4, and the method used is provided by the JetUser package (Version3_4_NewSyst). To determine the systematics associated with the Jet corrections, we use a procedure outlined on the Jet Corrections web page (along with a code snippet), which can be found on:
http://www-cdf.fnal.gov/internal/physics/top/jets/jetcorr-systematics.html

In this procedure, only the Monte Carlo Jet E_T's are shifted due to the systematic uncertainty on the E_T scale. In our case, this means we generate what we refer to as "shifted templates" for both $t\bar{t}$ and $W+ \geq 3$ jets Monte Carlo. There will be shifted templates for both an upper systematic uncertainty shift, as well as a separate set for a lower systematic

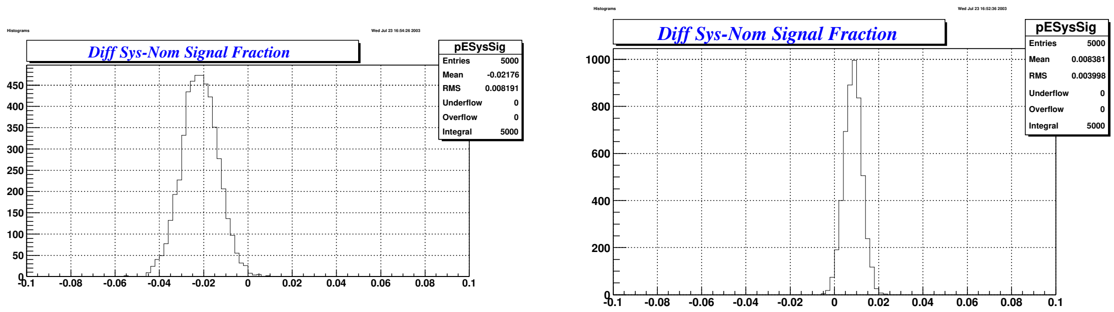


Figure 18: These plots show the difference in the fitted fraction between using the standard templates (with a top mass of $175GeV/c^2$, and using the generated templates with a top mass of $170GeV/c^2$ (left) and $180GeV/c^2$ (right).

uncertainty shift. We then generate PE's using the upper shifted MC, and fit the templates with both the upper shifted templates (plus QCD shape from the data) and the default (unshifted) templates. The difference is the systematic effect. For the upper systematic, we find a shift in the fitted fraction of +30% (meaning we systematically find 30% more $t\bar{t}$ if the E_T is higher than our default). For the lower systematic, we find a shift in the fitted fraction of -26% (meaning we systematically find 26% less $t\bar{t}$ if the E_T scale is lower than our default). We symmetrize this result and obtain a systematic of ±28% on the fit fraction due to varying the Jet E_T scale. This systematic increases if we use the $W+\geq 4$ sample to measure the cross section. In this case, we obtain a systematic of ±33%.

Since this systematic is the primary uncertainty for this analysis, we tried to break it down and understand which piece(s) of the jet energy scale process dominates. To do this, we used the above outlined process and applied each correction (syscode) separately. The $t\bar{t}$ fraction is then measured from the new PE and compared to the nominal value of 0.2042. The answer was determined separately for a shift up and a shift down. What is reported in Table 16 is just the average shift. That of course is not the entire story. We need to add in the acceptance systematic as well since events move in and out of our signal bin as a result of this energy scale change. The column labeled "Shape+Accept" is a straight addition of these effects. The bottom line by adding all of the "Shape+Accept" column in quadrature is that the overall systematic is 30.2%. The dominant component is the scale correction followed in close second with the relative correction, absolute correction and out of cone.

12.1.3 Shape systematic from ISR and PDF's

To determine this systematic, we break it up into two pieces; one which handles the ISR shape and the second piece to determine the PDF shape systematic. The details follow.

- The model for initial state radiation (ISR) used in the generation of the top quark can also have an impact on the shape of the resulting H_T distribution. To study this, we use the PYTHIA Monte Carlo, with ISR turned on (the default) and ISR turned off. Just as described above, we generate PE's using the ISR-off Monte Carlo, but fit with the default (in this case PYTHIA with ISR on) templates. The systematic is taken to be half of the result, or 0.56%.
- In order to obtain a systematic for the PDF, we followed the prescription of the top mass

Data/MC	Sys Code	Shift	Shape+Accept
Data	1(relative)	0.01838	0.09882
Data	2(cal stab)	0.00930	0.04942
Data	3(scale)	0.02810	0.14113
MC	1(relative)	0.02109	0.12059
MC	2(cal stab)	0.0	0.0
MC	3(Scale Cor)	0.03290	0.18058
MC	4(mult int)	0.0	0.0
MC	5(absolute)	0.02151	0.12011
MC	7(out of cone)	0.02579	0.04592
MC	8(splash out)	0.00987	0.07256

Table 16: List of individual contributions to the jet energy scale systematic. The $t\bar{t}$ input fraction in the pseudo-experiments is 0.2042. The shifts quoted are all relative to this input fraction. MC Syscodes 0-4 are determined using Level 4 corrections. MC syscodes 5-8 use Level 7 corrections. In the "Shape + Acceptance" column, this takes into account the acceptance systematic as a result of the shift as a straight add to the "shift" column.

group. The cross section systematic for this PDF is measured using Herwig MC samples with three of the 20 fit parameters in CTEQ6M PDF varied within their uncertainties ($\pm1\sigma$). These three parameters were chosen because their corresponding uncertainties dominate the uncertainties in the dijet cross section at high Q^2. To be specific, the parameters chosen were eigenvectors 10, 15 and 19. Eigenvector 15 for instance controls the amount of glue-glue in ISR. Using pseudo-experiments, we determine a cross section after varying each parameter by +1 and -1 σ. We divide the difference by two and then add them in quadrature. The results are shown in Table 17.

The final systematic for PDF in ≥ 3 jets is 0.24 pb and in the ≥ 4 jets of 0.30 pb. In terms of top fraction, the PDF systematic for W+$\geq$3, $\geq$4 jets translates into 3.5% and 4.3% respectively.

Eigenvector	njet	-1 σ	+1 σ	$\Delta\sigma/2$
15	≥ 3	7.28 $\pm$ 0.02pb	7.39 $\pm$ 0.02pb	0.06 pb
10	≥ 3	7.36 pb	7.15 pb	0.11 pb
19	≥ 3	7.39 pb	7.00 pb	0.20 pb
15	≥ 4	6.86 $\pm$ 0.04pb	7.04 $\pm$0.04pb	0.09 pb
10	≥ 4	6.99 pb	6.69 pb	0.15 pb
19	≥ 4	6.80 pb	6.32 pb	0.24 pb

Table 17: Shifts in the ttbar cross section as determined using pseudo-experiments for $\pm1\sigma$ shifts in the specified PDF Eigenvector parameters.

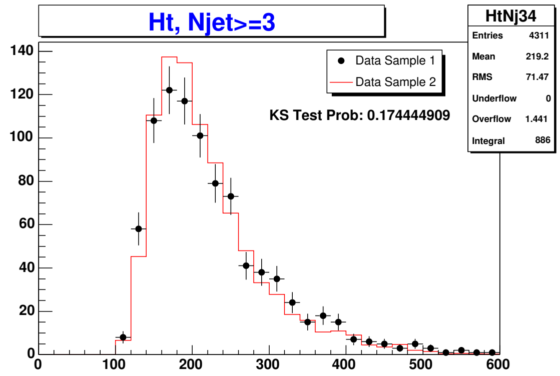


Figure 19: H_T for two different ALPGEN samples, one in which Q^2 = M_W^2, and one in which Q^2 = $M_W^2 + \Sigma {P_T}^2(jet)$ (nominal).

12.2 Determination of the Q^2 Shape Systematic

The choice of Q^2 scale in our MC's can have a significant impact on the shape of the H_T distribution. To measure our sensitivity to this effect, we performed pseudo-experiments using two different ALPGEN MC samples. One sample was our nominal one which uses for a Q^2 value of $M_W^2 + \Sigma P_T^2(jet)$. The second ALPGEN sample was generated using Q^2 = M_W^2. Pseudo-experiments were generated by selecting events from the second template and then fitting the result to the nominal template. For the W+$\geq$ 3 jet events, the difference between the two over the full range of Q^2 is 28%. Taking half the difference, we use 14% as the Q^2 systematic. For the W+$\geq$ 4 jet bin, the fraction of signal to background is much larger. Consequently, we are less sensitive to the Q^2 scale in the W+jets. In this bin, the systematic is 0.62%. The H_T shape for the two ALPGEN samples is shown in Figure 19.

12.2.1 Shape systematic from variation of the top quark generator

To test the dependence of our result on the generator used for $t\bar{t}$, we examine the difference in the H_T shape obtained using HERWIG (the default) and PYTHIA. Once again, the systematic is taken to be half of the result, or 0.60%.

12.3 Determination of acceptance systematics

Most of the systematics listed above also impact the acceptance. Unfortunately, each of these are completely correlated with the shape systematic. This can be seen by again examining the cross section formula:

$$\sigma(t\bar{t}) = \frac{\mathcal{F}\cdot N}{\mathcal{L}\cdot A}$$

For example, the if the true top mass is 170 GeV/c^2, the above shape systematic discussion indicates that we would determine a fit fraction $\mathcal{F}$ which is -11.6% smaller than its true value on average (since we are using an H_T shape determined using a top mass of 175 GeV/c^2). This would result in a cross section that is -11.6% smaller than its true value. From the Monte Carlo, we find that the acceptance A for events using a mass of 170 GeV/c^2 is approximately 5% smaller than for a mass of 175 GeV/c^2. Again, since we are assuming the top mass is 175

GeV/c^2 and don't know it is really 170 GeV/c^2, we use the larger acceptance and obtain a cross section which is 5% too small. These two effects go in exactly the same direction, unfortunately, and as a result, the overall effect of varying the top mass is the sum of these two effects. This result, along with the acceptance systematics due to E_T scale, ISR, and top generator are shown in tables 20 and 21, for the $W+\geq 3$ and $W+\geq 4$ samples, respectively.

The systematic due to lepton ID effects is also listed in these tables, and is taken from CDF Note 6574 [5]. The systematic is smaller than listed in that note, since ISR and E_T scale effects were included, and we have redone those systematics for this analysis.

12.4 Determination of the QCD background fraction using MET vs ISO systematic

In order to estimate the systematic uncertainty of the QCD background using this MET vs ISO technique, we varied the location of the upper isolation and lower $\not\!\!E_T$ cuts. Each of these values are somewhat arbitrary. We varied Isolation boundary between 0.17 and 0.50. (0.2 is our standard) in steps of 0.03 to 0.05. We varied the MET boundary between 5 and 15 GeV in 2 GeV steps (10 is our standard). The results of these studies can be found in Tables 18 and 19. Based on the fluctuations in these numbers, we take as our background uncertainty to be 10%.

To determine a systematic, we use pseudo-experiments and shift the fraction of QCD normalization by ± 10%. In so doing, we find that the change in cross section is negligible. So, in order to quote a systematic, we will take the expected QCD fraction of 6.7%, double it and halve it in pseudo experiments, the nominal cross section goes from 7.0pb to 6.86 ± 0.02 pb (doubling it) and 7.10 ± 0.02 pb(halving it). We will take as the systematic then half the difference between these two extremes. In so doing, the systematic uncertainty on the MET vs ISO bckg calculation is 0.12pb or 1.7%.

MET Boundary	**Bckg Fraction**	**Iso Boundary**	**Bckg Fraction**
0.17	9.7 ± 1.6%	5 GeV	8.6 ± 1.7 %
0.18	9.4 ± 1.5%	7 GeV	8.9 ± 1.6 %
0.19	9.1 ± 1.5%	9 GeV	8.9 ± 1.5%
0.20	8.9 ± 1.6%	10 GeV	8.9 ± 1.5%
0.21	8.9 ± 1.6%	11 GeV	8.9 ± 1.5%
0.25	8.0 ± 1.5%	13 GeV	9.6 ± 1.6%
0.30	8.5 ± 1.7%	-	-
0.40	8.9 ± 1.9%	-	-
0.45	9.2 ± 2.2%	-	-

Table 18: QCD background fractions for the electron+jet data as determined using MET vs ISO in the non-isolated lepton sample. The boundary regions are varied and the resulting fluctuations in the background fraction used to determine the QCD systematic

MET Boundary	Bckg Fraction	Iso Boundary	Bckg Fraction
0.17	4.1 ± 1.1%	5 GeV	4.5 ± 1.3 %
0.18	4.1 ± 1.0%	7 GeV	4.0 ± 1.1 %
0.19	4.0 ± 1.0%	9 GeV	4.0 ± 1.0 %
0.20	4.0 ± 1.0%	10 GeV	4.0 ± 1.0%
0.21	4.1 ± 1.0%	11 GeV	4.1 ± 1.0%
0.25	3.8 ± 1.0%	13 GeV	4.2 ± 1.0%
0.30	3.7 ± 0.9%	15 GeV	4.1 ± 0.9%
0.40	3.3 ± 0.9%	-	-
0.45	3.4 ± 1.0%	-	-

Table 19: QCD background fractions for the muon+jet data as determined using MET vs ISO in the non-isolated lepton sample. The boundary regions are varied and the resulting fluctuations in the background fraction used to determine the QCD systematic

12.5 Determination of background model systematic

The sample that we are using to model the background from non-W is the non-isolated lepton sample. This sample requires that the isolation of the lepton be > 0.2, but otherwise, all of the other cuts, both on the lepton as well as the MET and Jet cuts, are identical. This sample presumably includes fake leptons, as well as $b\bar{b}$ in which one of the b's decays semi-leptonically to an electron or muon. What we don't know is how the kinematics (and composition) of this sample changes as the isolation cut is changed from $I < 0.1$ to $I > 0.2$.

One thing we can do is to see what happens if we change the composition of the sample so that it is dominantly made up of only fakes. A convenient sample for this is the conversion electron sample. This sample is obtained easily, since these events are included in the standard tight lepton sample, and are already flagged as conversions. We simply require that the electron is an identified conversion (as opposed to removing these, which is the default procedure). Conversion electrons are real electrons, but are definitely non-W in origin. In addition, since the isolation cut *is* applied, this tests some of the assumptions we make in using the non-isolated sample.

The size of this systematic also provides a very good argument for reducing the overall size of the non-W background. As should be obvious, the smaller the non-W background, the smaller the systematic associated with lack of knowledge of the non-W H_T shape.

We perform pseudo-experiments for two different cut sets: without the $\Delta\phi$ cut and with the $\Delta\phi$ cut. The results are the following:

- Without the $\Delta\phi$ cut: the expected precision on the cross section is approximately 28%, and the systematic due to the uncertainty in modeling the background is 32%.
- With the $\Delta\phi$ cut: the expected precision on the cross section is approximately 28%, and the systematic due to the uncertainty in modeling the background is 16%.

So applying the $\Delta\phi$ does not change the expected precision on the cross section, but it does drop the systematic due to background modeling by a factor of 2. This is the primary motivation for making this cut, and we quote the systematic as ±16%.

Effect	Shape	Acceptance	Total
Et Scale	28%	5.1%	30.2%
Generator	0.60%		0.60%
Top Mass	8.0%	5.0%	13%
Lepton ID (includes PDF)	3.5%	5.3%	8.8%
ISR	0.56%	0.78%	1.3%
Q^2	14%		14%
QCD	1.7%		1.7%
Luminosity			5.9%
Background model	16%		16%
Total			40.6%

Table 20: Relative systematic uncertainty in the fitted value of the $t\bar{t}$ cross section in the W+≥3 jet sample, coming from various sources, as discussed in the text.

Effect	Shape	Acceptance	Total
Et Scale	20.4%	13%	31.5%
Generator	0.60%		0.60%
Top Mass	8.0%	5.0%	13%
Lepton ID (includes PDF)	4.3	5.3%	9.6%
ISR	0.56%	0.78%	1.3%
Q^2	0.6 %		0.6%
QCD	3.3%		3.3%
Luminosity			5.9%
Background model	16%		16%
Total			39.5%

Table 21: Relative systematic uncertainty in the fitted value of the $t\bar{t}$ cross section in the W+≥4 jet sample, coming from various sources, as discussed in the text.

12.6 Total systematic uncertainty

The total systematic uncertainty, including effects from luminosity, acceptance and shape, and background modeling, are shown in Tables 20 and 21, for the $W+\geq 3$ and $W+\geq 4$ samples, respectively. In the cases where both a shape and acceptance systematic are listed, the total is given by the staight sum of the two. The final systematic including all effects is obtained by adding in quadrature the quantities listed in the column labeled $Total$. The systematic in the W+≥ 3 jet sample is 40.6%, while in the W+≥4 jet sample it is 39.5%.

13 Summary

We have measured the $p\bar{p} \rightarrow t\bar{t}$ cross section in the lepton+jets channel using a kinematic analysis. The H_T variable was chosen to separate signal from background. The background fractions were determined from a combined fit H_T templates based on "W-like" backgrounds, QCD fakes and $t\bar{t}$. The $t\bar{t}$ and W background shapes were determined from MC simulations. The QCD fakes template was derived from a data sample of non-isolated leptons and then checked with conversions. Based on the 125.8 pb^{-1} of data collected for the summer conferences, 5.1 $\pm$ 1.8(stat) $\pm$ 2.1(syst)pb in the W+$\geq$ 3 jet sample. In the W+$\geq$4 jet sample, the cross section is measured to be 7.7 $\pm$ 2.4(stat) $\pm$ 3.0(syst)pb. These numbers should be compared to the latest theoretical prediction of $6.7^{+0.71}_{-0.88}$ pb [6].

References

[1] D. Acosta *et al.*, *A First Look at the CLC Luminosity Measurements*, CDF Note 6052.

[2] S. Klimenko, J. Konigsberg, T. Liss, *Averaging of the inelastic cross-section measured by the CDF and E811 experiments*, CDF Note 6314.

[3] V. Drollinger, M. Gold *et al.*, *A Neural Network Measurement of the Top Quark Mass*, CDF note 6078.

[4] Andy Beretvas, Morris Binkley and Luc Demortier, *Binned KS Probability*, CDF Note 4373.

[5] Ken Bloom, Koji Ebina, Evelyn Thomson, Kohei Yorita, *ttbar Event Selection and Detection Efficiency for Summer 2003 Lepton+Jets Analyses*, CDF Note 6574.

[6] R. Bonciani, S. Catani, M.L. Mangano and P. Nason, *Nucl. Phys. B 529* (1998) 424, as updated by M. Cacciari, S. Frixione, M.L. Mangano, P. Nason, G. Ridolfi, *The t-tbar cross-section at 1.8 and 1.96 TeV: a study of the systematics due to parton densities and scale dependence*, CERN-TH/2003-054