CDF/PUB/ELECTROWEAK/PUBLIC/9216

Search for WW and WZ production in lepton, neutrino plus jets final states at CDF RUN II

The CDF Collaboration
URL http://www-cdf.fnal.gov
(Dated: February 25, 2008)

We present a search for WW and WZ production in charged lepton, neutrino plus jets final states produced in $p\bar{p}$ collisions with $\sqrt{s} = 1.96$ TeV at the Fermilab Tevatron, using 1.2 fb^{-1} of data accumulated with the CDF II detector. This channel is yet to be observed in hadron colliders due to the large single W plus jets background. However, this decay mode has a much larger branching fraction than the cleaner fully leptonic mode making it more sensitive to anomalous triple gauge couplings that manifest themselves at higher transverse W momentum. Because the final state is topologically similar to associated production of a $Higgs$ boson with a W, the techniques developed in this analysis are also applicable in that search. An Artificial Neural Network has been used for the event selection optimization. The 95% CL upper limit to the cross section is estimated to be

$$\sigma \times Br(W \rightarrow \ell\nu; W/Z \rightarrow jj) < 2.88 \text{ pb}$$

I. INTRODUCTION

We present in the current note the WW and WZ production in the charged lepton, neutrino plus jets decay channel. The Feynman diagrams for this decay mode are shown in Figure 1. The hadronically decaying W ($W \to jj$) can not be differentiated from a hadronically decaying Z ($Z \to jj$) due to the limited jet energy resolution. We therefore study the WW and WZ production together.

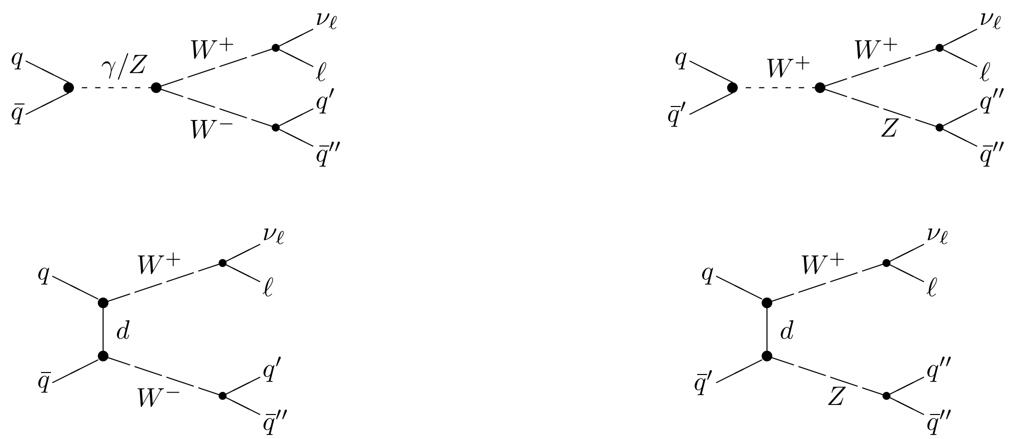

FIG. 1: Feynman diagrams for WW and WZ production in the semi-leptonic decay channel; there are similar diagrams for $W^- \to \ell\nu$ and $W^+ \to q\bar{q}$

This is a decay mode never observed at hadron colliders due to the large $W + jets$ background. The cross section of $W + jets$ production at $\sqrt{s} = 1.96$ TeV is of the order of 300 pb, resulting in a signal over background ratio that is smaller than 1%, making therefore the signal very difficult to observe. However, the branching ratio for this decay mode is very significant, making this channel favorable for aTGC studies.

Given that the signal over background is initially very small, sophisticated statistical techniques need to be used in an attempt to observe this decay mode. This decay mode is topologically similar to associate production of a Higgs boson with a W, therefore techniques that are developed for the WW/WZ searches are also applicable in Higgs searches. This topology is common in many other interesting process (e.g. SUSY signatures and single top), and the WW and WZ are backgrounds for such processes. It is therefore clear that better understanding of the diboson production helps in eliminating significant backgrounds in various other searches.

The theoretical cross sections for this decay mode are ([1], [2])

$$\sigma_{WW} \times Br(W \to \ell\nu_\ell, W \to jj) = (12.4 \pm 0.8)\ \text{pb} \times 0.146 = 1.81 \pm 0.12\ \text{pb}$$

$$\sigma_{WZ} \times Br(W \to \ell\nu_\ell, Z \to jj) = (4.0 \pm 0.3)\ \text{pb} \times 0.07 = 0.28 \pm 0.02\ \text{pb}$$

The total cross section in the Standard Model of the signature we are looking for is therefore

$$\sigma_{WW+WZ} \times Br(W \to \ell\nu_\ell, W/Z \to jj) = (2.09 \pm 0.14)\ \text{pb}$$

II. EVENT SELECTION

The decay channel of this analysis contains three objects: the *charged lepton* (electron or muon), the *neutrino* (missing transverse energy, $\not\!\!E_T$), and the *jets*.

There are several processes that result in the same final state topology (1 charged lepton, electron or muon, $\not\!\!E_T$ and jets) as the diboson production, and thus are backgrounds to this search. The background processes that are taken into account are:

- $p\bar{p} \to W(\to \ell\nu_\ell) + jets;\ \ell = e,\ \mu$
- $p\bar{p} \to W(\to \tau\nu_\tau) + X$

- $p\bar{p} \rightarrow Z(\rightarrow \ell\ell) + X;\ \ell = e,\ \mu$
- $QCD\ processes$
- $p\bar{p} \rightarrow t\bar{t} + X,\ t \rightarrow Wb$

The largest of these backgrounds is the $(W \rightarrow \ell\nu)$+jets. The large cross section of this process leads by itself, but even more with the rest of the backgrounds, to a very poor Signal over Background ratio ($S/B < 1\%$ initially).

In order to study the semileptonic decays of WW and WZ production, we select events that have:

- Exactly one lepton, central electron or central muon. The lepton description is given at [3].
- $\not{E}_T$ >25 GeV.
- At least two jets with p_T >15 GeV each. The jets are corrected at Level 7 [4], in order to achieve reconstruction at parton level. In order to achieve good data to Monte Carlo agreement, an additional cut has been set at $\Delta\eta(J1, J2)$ <2.5. The signal signature doesn't have significant contribution above this point.
- One additional cut applied to ensure good data to Monte Carlo agreement is on the leptonic W transverse mass: 30 GeV $< \mathrm{M_T} <$ 120 GeV. The good agreement in this region will be demonstrated in the following chapter.

III. ANALYSIS DESCRIPTION

Using the Monte Carlo description in the signal region we find that the signal (S) over background (B) ratio (S/B) is very poor. Also very poor are the signal fraction ($S/(S+B)$) and the $significance$, defined as the ratio of the signal yield S to the total statistical error (in the gaussian approximation) ($significance = S/\sqrt{(S+B)}$) (Table I). In order to increase the probability of seeing a signal these ratios need improvement, and in this perspective is based the methodology which has been chosen.

Before ANN Cut		$\frac{S}{S+B}$	$\frac{S}{\sqrt{S+B}}$
SIG	716	0.024	4.2
BGR	29093		

TABLE I: Signal and background yields for invariant mass [45,160]GeV

In principle, direct cuts could be applied in various kinematic variables where the signal and the background are separated. However, there is the possibility to exploit correlations between several variables by using some *multivariate technique*. In this analysis, the multivariate technique we use is the *Artificial Neural Network*.

The Artificial Neural Network (ANN) is trained using variables where the signal and background are well separated in shape. The output is a global variable where the signal and background are as widely separated as possible. This global variable is used for the maximum significance gain point definition. We perform a lower cut at this point. In the data above this point the signal and background have the 'maximum possible' significance, within the context of the chosen Neural Network.

The training of the ANN has been performed using dimensionless variables (angles and angle related ones). Performing cuts in these variables does not significantly change the dijet invariant shape, since these variables are less correlated with it. Several such variables have been used in the input of the ANN, but the variables giving the best discrimination between signal and background (best significance improvement) are the six shown in 2. The training is done in the signal region only, and both the background and signal descriptions are given by the MC. The signal description is given by the combination of the WW and WZ expectations, as given by the Pythia inclusive MC samples. The ANN has been trained with the electrons and muons combined. In order to do that, we verify that the electron and muon signal and background shapes are similar.

A significance curve is plotted as a function of the ANN output. Several sets of input variables have been used, with different numbers of input nodes and different minimization algorithms, and the set that leads to the best significance gain has been chosen as the one to be used. This is how the 6 variables shown in Figure 2 have been chosen. Those variables give the ANN output that is shown in Figure 3. A summary of the signal and background yields after the ANN cut is given in Table II.

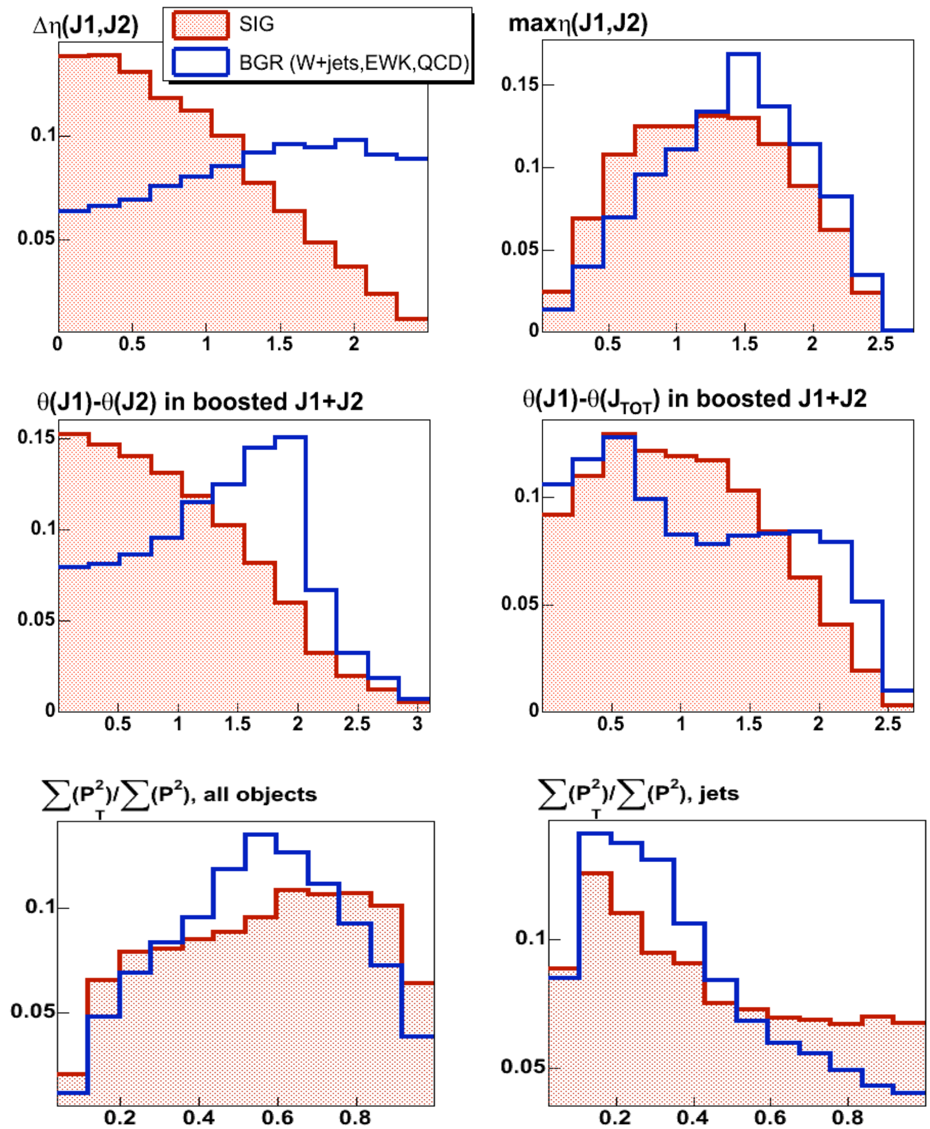

FIG. 2: Neural Network input variables. The ANN is trained with events in the signal region only.

After ANN Cut		$\frac{S}{S+B}$	$\frac{S}{\sqrt{S+B}}$
SIG	554	0.036	4.5
BGR	14481		

TABLE II: Signal and background yields using Monte Carlo, for invariant mass [45,160]GeV, after the ANN cut.

Comparisons between data and MC in the sidebands is shown in Figure 4 (Invariant mass comparisons) and in Figure 5 (ANN output comparison). We conclude that the data are well described by the MC for what concerns the ANN output in the sidebands (the χ^2 probability for the Data-MC agreement is $\approx 30\%$).

We use the dijet invariant mass to continue the analysis. We parameterize the shape of the dijet invariant mass after the Neural Network cut is performed. The shape of the signal is given by the Pythia Monte Carlo, while the shape of the background is motivated by Monte Carlo and contains two free parameters. The overall parameterization includes the signal and background descriptions, with the signal fraction being a free parameter. A likelihood function

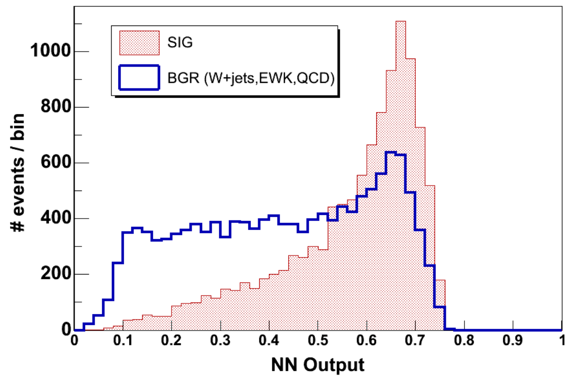

FIG. 3: ANN Output.

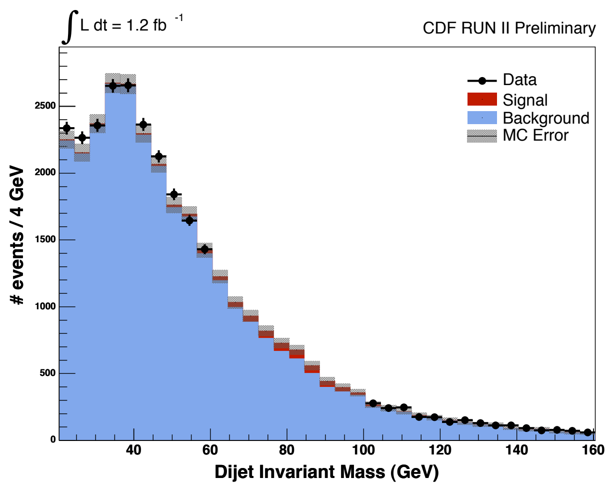

FIG. 4: DATA-MC comparison after the ANN cut (sidebands only). The χ^2 probability for the Data-MC agreement is ≈60%.

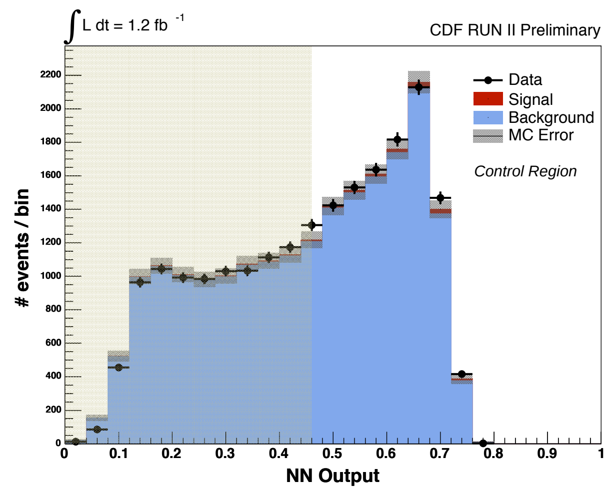

FIG. 5: ANN Output - Data-MC comparison in the dijet invariant mass sidebands.

is constructed using this parameterization, and a fit is performed on the data. The fit will give us the parameters of the background as well as the signal fraction, that is interpreted as a number of events.

The result is evaluated using the likelihood ratio as a test statistic. We test the signal plus background hypothesis and the background-only hypothesis with pseudoexperiments and we deal with them exactly as we deal with the data: we perform a fit for the three free parameters, and each time calculate the likelihood ratio. The likelihood ratio distributions will answer the question of how much of the background could fake the signal, and vice versa. Using Monte Carlo, we estimate an $\approx 2.5\sigma$ statistical significance for the expected signal.

IV. SYSTEMATICS

For the measurement we performed, we need to address separately two kinds of systematics: those that affect the significance of the measurement, and those that affect the cross section. In the first category the uncertainties is the signal fraction measurement are interpreted. These uncertainties can affect the conclusion about the significance. The uncertainties in the signal fraction, together with the uncertainties in the acceptance and the luminosity are giving the overall uncertainty in the cross section.

A summary of all systematic uncertainties the affect the result of this analysis is given in Table III.

Source	Effect in Signal Fraction
Jet Energy Scale	10%
Jet Resolution	10%
Background Shape	20%
ISR	5%
FSR	$< 1\%$
Source	**Effect in Acceptance**
Jet Energy Scale	3%
Jet Resolution	$< 1\%$
ISR	2%
FSR	3%
W leptonic	3%
Source	**Effect in Cross Section**
Total Signal Fraction	25%
Total Acceptance	5%
Luminosity	6%
Total Effect in Cross Section	26%

TABLE III: The systematic uncertainties and their effect in the signal fraction, the acceptance and finally the cross section.

V. RESULTS

After unblinding the data a simultaneous, for the background parameters and signal fraction, fit is performed in the dijet invariant mass range [45,160]GeV. The signal fraction is measured to be $f_S = 0.027 \pm 0.014$. We measured 15016 total number of events, that correspond to 410 ± 212 signal events. From the Standard Model MC studies we were expecting to measure 15035 ± 123 total events, out of which 554 ± 24 signal events.

The likelihood fit on the data, as well as the background estimation, is shown in Figure 6. Figure 7 shows the signal shape measured with the data. The expected shape and the shape we see with the data agree well. The statistical significance of the signal is 1.9σ and interpreting the systematic error into this (using the assumption of a result that follows a gaussian distribution and considering as a total error the statistical and systematic errors added in quadrature), the significance of the measurement becomes 1.7σ.

Taking into account only the statistical and systematic error in the number of signal events we measure

$$\sigma_{WW/WZ} \times Br(W \to \ell\nu, W/Z \to jj) = 1.47 \pm 0.77(stat) \pm 0.38(sys)\ pb$$

that is compatible with the theoretical prediction for the cross section

$$\sigma^{theory}_{WW+WZ} \times Br(W \to \ell\nu_\ell, W/Z \to jj) = (2.09 \pm 0.14)\ \text{pb}$$

We will set a 95% CL upper limit for the measured cross section. Given that the measurement follows a gaussian distribution, the 95% CL limit can be set by the estimated value plus 1.64 5 standard deviations [5]. A total error (statistical and systematic added in quadrature) is taken into account. Therefore, the 95% CL upper limit we set for the cross section is

$$\sigma \times BR < 2.88\ pb$$

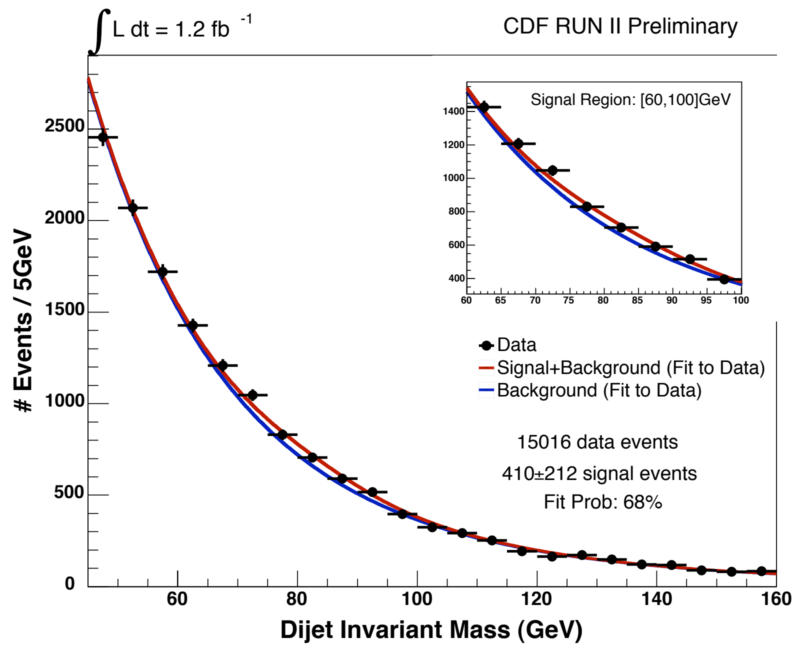

FIG. 6: Likehood fit on data (red line). The blue line shows the background estimation.

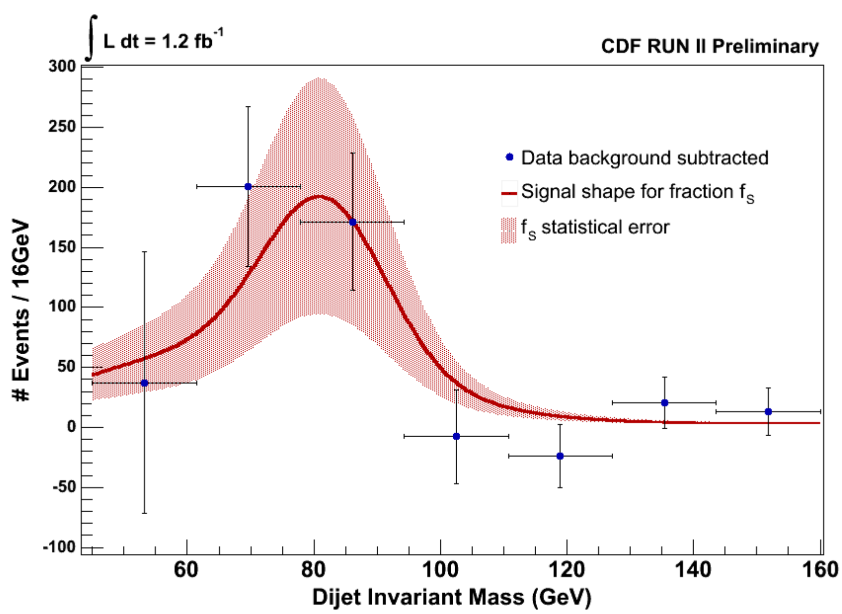

FIG. 7: The signal shape measured on data. The plot is made by subtracting the measured background shape from the data. The red line corresponds to the MC shape for the measured signal fraction. There is a good agreement between the MC shape and the shape seen in the data.

VI. SUMMARY

A search for WW plus WZ production in lepton-neutrino plus dijet final state, using 1.2 fb^{-1} of data, has been presented. A ANN has been used for the significance optimization. MC studies showed an $\approx 2.5\sigma$ expected effect, but on data we saw a statistical 1.9σ effect. Including systematic uncertainties, the significance of the measurement is 1.7σ. The theoretical prediction for the cross section is

$$\sigma_{WW/WZ}^{theory} \times Br(W \to \ell\nu, W/Z \to jj) = 2.09 \pm 0.14\ pb$$

We measured

$$N_{Signal} = 410 \pm 212(stat) \pm 102(sys)\ signal\ events$$

that correspond to a cross section

$$\sigma_{WW/WZ} \times Br(W \to \ell\nu, W/Z \to jj) = 1.47 \pm 0.77(stat) \pm 0.38(sys)\ pb$$

The 95% CL upper limit to the cross section is set at

$$\sigma \times Br < 2.88\ pb$$

[1] *Review of Particle Physics*, W.-M.Yao et al, Journal of Physics G 33. 1 (2006).

[2] *An update on vector boson pair production at hadron colliders*, J.M. Campbell and R.K. Ellis, Theory Department, Fermilab, hep-ph/9905386, May 1999.

[3] *Observation of WZ Production*, CDF Collaboration, Phys. Rev. Lett. 98, 161801 (2007).

[4] *Generic Jet Energy Corrections at CDF*, Jet Energy and Corrections Public Web Page, Conveners: Monica D'Onofrio, Kenichi Hatakeyama, Andrew Mehta:
http://www-cdf.fnal.gov/physics/new/top/2004/jets/cdfpublic.html

[5] *Statistical Data Analysis, Chapter 9.3: Confidence interval for a Gaussian distributed estimator*, Glen Cowan, Oxford Science Publications.