Search for WW and WZ production in lepton plus jets final state at CDF

Aaltonen, Timo

2009


http://hdl.handle.net/10138/24399
https://doi.org/10.1103/PhysRevD.79.112011

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.
Search for $WW$ and $WZ$ production in lepton plus jets final state at CDF

We present a search for WW and WZ production in final states that contain a charged lepton (electron or muon) and at least two jets, produced in $\sqrt{s} = 1.96$ TeV $p\bar{p}$ collisions at the Fermilab Tevatron, using data corresponding to 1.2 fb$^{-1}$ of integrated luminosity collected with the CDF II detector. Diboson production in this decay channel has yet to be observed at hadron colliders due to the large single $W$ plus jets background. An artificial neural network has been developed to increase signal sensitivity, as compared with an event selection based on conventional cuts. We set a 95% confidence level upper limit of $\sigma_{WW} \times BR(W \rightarrow \ell\nu, W \rightarrow j) + \sigma_{WZ} \times BR(W \rightarrow \ell\nu, Z \rightarrow j) < 2.88$ pb, which is consistent with the standard model next to leading order cross section calculation for this decay channel of $2.09 \pm 0.12$ pb.

PACS numbers:

---

*Deceased

†With visitors from *University of Massachusetts Amherst,
In the standard model of particle physics (SM), the weak bosons (W, Z) and the photon are the gauge bosons of the local $SU(2) \times U(1)$ symmetry. The spontaneous breaking of this symmetry gives masses to the W and Z bosons, while the gauge symmetry itself defines the interactions among these heavy bosons and the photon. Since the electroweak sector of the SM relies on this mechanism, it is of prime importance to test the boson couplings experimentally. We present in this paper a search for WW and WZ production in the charged lepton (electron or muon), neutrino plus jets decay channel \([a]\). Figure 1 shows the leading order diagrams for the \(p\bar{p} \rightarrow W(\rightarrow \ell\nu)V(\rightarrow \text{jets})\) process, where \(V \equiv W, Z\).

![Leading-order diagrams for WW and WZ production](image)

**FIG. 1:** Leading-order diagrams for WW and WZ production in the semi-leptonic decay channel; there are similar diagrams for the charge conjugates of the processes shown.

The production of WW and WZ could be more sen-

sitive to the triple gauge couplings (TGC) WW(Z/γ), present in the s-channel (Fig. \([a], b\)), and would be enhanced by the presence of nonstandard couplings (anomalous TGC) \([2]\). The hadronically decaying W (W → jets) cannot be differentiated from hadronically decaying Z (Z → jets) due to the limited jet energy resolution of the detector \([3]\). We therefore search for the combined WW and WZ production.

The next to leading order (NLO) SM cross sections times branching ratio for these modes are \(\sigma_{WW} \times BR(W \rightarrow \ell\nu_V, W \rightarrow \text{jets}) = 1.81 \pm 0.12 \text{ pb}\) and \(\sigma_{WZ} \times BR(W \rightarrow \ell\nu_V, Z \rightarrow \text{jets}) = 0.28 \pm 0.02 \text{ pb}\) \([4, 5]\).

The D0 collaboration recently reported the first evidence for the WW and WZ production in lepton plus jets decay mode \([6]\). This decay mode has not been observed yet at hadron colliders due to the large W plus jets background. The cross section for W plus two or more jets at \(\sqrt{s} = 1.96\) TeV is at least two orders of magnitude larger than the total cross section times branching ratio of the signal \([7]\), which translates into an expected signal to background ratio of less than 1%. Given that this diboson production is topologically similar to the associated production of Higgs and W bosons, techniques that are developed for the WW and WZ searches are of key importance for Higgs searches. The lepton plus jets final state is common in other interesting processes, such as top production; thus, diboson production decaying in this channel is a significant background to these processes, and vice-versa.

The search for WW and WZ production is performed using data corresponding to 1.2 fb\(^{-1}\) of integrated luminosity collected with the CDF II detector from pp collisions at \(\sqrt{s} = 1.96\) TeV at the Fermilab Tevatron. The CDF II detector is a general-purpose, multilayered detector designed to study many aspects of particle physics. It combines precise tracking systems with calorimeters and muon detectors \([8]\). A tracking system is positioned closest to the beamline to provide accurate momentum determination of charged particles. The tracking system is immersed in a 1.4 T uniform magnetic field, produced by a superconducting solenoid and aligned along the proton direction. Calorimeters located outside the tracking volume provide energy measurement of electrons, photons, and jets. The geometrical coverage of the calorimeters is maximized to measure the energy flow of all particles produced in a collision and indirectly detect the neutrinos by the presence of missing transverse energy \(E_T\) \([9]\). Muon chambers are located on the outer part of the CDF II detector.

The selection of the signal events proceeds as follows. The trigger system selects events with leptons of central pseudorapidity \(|\eta| < 1\); electron candidates with transverse energy \(E_T > 18\) GeV or muon candidates with transverse momentum \(p_T > 18\) GeV/c. Events that are reconstructed offline are required to contain one electron candidate with \(E_T > 25\) GeV or one muon candidate...
with $p_T > 25 \text{ GeV/c}$. The sample is enriched in events containing a neutrino by requiring that the $E_T$, corrected for the calorimeter energy leakage and the presence of muons, satisfies $E_T > 25 \text{ GeV}$.

The jets are reconstructed in the calorimeter using the JETCLU cone algorithm [12] with cone radius $R = \sqrt{\Delta\phi^2 + \Delta\eta^2} = 0.4$. Starting from seed locations corresponding to calorimeter towers with $E_T > 1 \text{ GeV}$, all nearby towers with $E_T > 0.1 \text{ GeV}$ are used to search for stable cones. To resolve ambiguities due to overlaps, cones sharing an energy fraction greater than 0.75 are merged into a single jet. The measured energy deposition in the detector is corrected for effects that distort the true jet energy [3]. Such effects include the non-linear response of the calorimeter to the particle energy, uninstrumented regions of the detector, spectator interactions, and energy radiated outside the cone. We select events that contain two or more jets with $E_T^{\text{jet}} > 15 \text{ GeV}$.

To enhance the signal selection, events are rejected if the difference in pseudorapidity between the two leading jets, $\Delta\eta(\text{Jet1, Jet2})$, is greater than 2.5.

The $WW$ and $WZ$ production in the lepton plus jets event signature is simulated using PYTHIA v6.3 [10] Monte Carlo generator, followed by GEANT-based [11] CDF detector simulation. We search for diboson production in the region of $[45,160] \text{ GeV/c}^2$ in the dijet invariant mass, that is constructed taking the two leading jets into account. Using the signal Monte Carlo description, we define a signal region of $[60,100] \text{ GeV/c}^2$. It contains approximately 80% of the reconstructed hadronically decaying $W$ bosons. Outside of the signal region we define a lower sideband region of $[45,60] \text{ GeV/c}^2$ and a higher sideband region of $[100,160] \text{ GeV/c}^2$. We enhance the $W$ event selection by rejecting events if the transverse mass $M_T$ of the lepton and $E_T$ system is not within the interval $30 \text{ GeV/c}^2 < M_T < 120 \text{ GeV/c}^2$.

The most significant background to the $WW$ and $WZ$ search in the lepton plus jets decay channel consists of $W$ plus jets events where the leptonically decaying $W$ boson is produced in association with jets that mimic a hadronically decaying $W$ or $Z$. The $W$ plus jets background is simulated using the ALPGEN v1.3 [13] Monte Carlo generator, followed by PYTHIA Monte Carlo generator for the parton shower and fragmentation, and full GEANT detector simulation. Other, less significant backgrounds originate from a tau lepton that is detected as an electron or a muon; events with large transverse energy due to the Drell-Yan process, where one of the two leptons is not reconstructed; and QCD events with a jet misidentified as a lepton. The QCD background is derived from the data, while the other background processes are simulated using PYTHIA Monte Carlo events.

Based on the Monte Carlo simulation, in the dijet invariant mass region of $[45,160] \text{ GeV/c}^2$ we predict 716 signal ($S$) events and 29,093 background ($B$) events. The estimated signal fraction ($S/(S + B)$) is small (0.024).

The statistical significance ($S/\sqrt{(S + B)}$) is equal to 4.1. In order to increase the sensitivity to $WW$ and $WZ$ in the lepton plus jets final state, more sophisticated techniques beyond event counting are required. Correlations between kinematic quantities are exploited using a feed-forward artificial neural network (ANN) [14].

A feed-forward ANN can be thought of as a single-valued function of input vectors. The function has many parameters, the values of which determine the output for a given input vector. Usually the output is a continuous distribution in the range 0 to 1. The training of the network is equivalent to a minimization procedure. The aim is to reduce the error function, which is the sum of the squared deviations of the neural network output from the desired output (usually 1) and background (usually 0). When the trained network with its optimised parameters is used with real events, the network output for each event is used to define if the event is selected or not [15].

The ANN we developed for this analysis is trained using six input variables that can discriminate the signal from the background. The output of the ANN is a variable where the signal and background are well separated. We perform the ANN training using angles and event shape quantities and ensure that cutting on the output of this ANN does not introduce significant bias on the signal and background dijet invariant mass shapes.

The quantities used in the ANN training are shown in Fig. 2. We used the difference between the pseudorapidity $\eta$ of the leading jets ($\Delta\eta(\text{Jet1, Jet2})$); the maximum value of the pseudorapidity $\eta$ of the two jets (max $\eta(\text{Jet1, Jet2})$); the fraction $\sum p_T^2 / \sum p^2$, where the sum is over all objects, leptons and jets (for the neutrino the $E_T$ is used in the denominator); the fraction $\sum p_T^2 / \sum p^2$, where the sum is over the two leading jets; and finally the quantities $\Delta\theta_{1,2} = \theta_{\text{Jet1}} - \theta_{\text{Jet2}}$ and $\Delta\theta_{\text{dijet},1}$, both calculated in the rest frame of the dijet system. The quantity $\Delta\theta_{\text{dijet},1}$ is given by the expression $\Delta\theta_{\text{dijet},1} = |\theta_{\text{Jet1}} - \theta_{\text{dijet}}|$, if $\theta_{\text{Jet1}} \cdot \theta_{\text{dijet}} > 0$; otherwise, it is given by $\Delta\theta_{\text{dijet},1} = |\pi - \theta_{\text{Jet1}} - \theta_{\text{dijet}}|$, where $\theta_{\text{Jet1}}$ and $\theta_{\text{dijet}}$ are calculated in the rest frame of the dijet system.

The training is performed using the variables in the signal region only. Both signal and background distributions are given by Monte Carlo simulations. The ANN has been trained for the electron and the muon channels combined.

The ANN output is shown in Fig. 3. A cut is applied at the point where the statistical significance is maximized. After applying this cut in Monte Carlo simulated events, we estimate within the dijet invariant mass region of $[45,160] \text{ GeV/c}^2$ an expected number of 554 signal events and 14,481 background events. The signal fraction is 0.037, improved by 53% with respect to the value before the ANN cut was applied. The statistical significance is 4.5, an improvement of about 10%.
A comparison between the data and the Monte Carlo simulated events of the ANN output shape in the sideband regions is shown in Fig. 4. For the ANN output in the sidebands, the data are well described by the Monte Carlo simulation.

We measure the signal fraction in the data by performing a likelihood fit on the dijet invariant mass distribution. The shape of the dijet invariant mass parameterization for the events that pass the cut in the ANN output. The overall parameterization consists of the signal and background descriptions, with the signal fraction, which is then converted into the number of events.

The signal fraction measured from the data over a dijet invariant mass region of $[45,160]$ GeV/$c^2$, and for the events that pass the ANN cut, is $f_S = 0.027 \pm 0.014$. Given a total of 15,016 events, this signal fraction corresponds to $410 \pm 213$ signal events. The uncertainty is statistical, obtained from the fit, and accounts for the poisson fluctuations of the total number of events measured on the data.

The likelihood fit from the data is shown in Fig. 3. The overall fit result (signal plus background) and the measurement of the signal fraction is maximized, in the context of this specific ANN output.

FIG. 2: Neural Network input variables. The ANN is trained with events in the signal region only and for electrons and muons simultaneously. Both signal and background descriptions are given by Monte Carlo simulated events.

FIG. 3: ANN Output, interpreted as a function associating the output, from 0.0 to 1.0, to each event. The distributions for signal and background samples are shown. A cut is applied at the value 0.46. This is the value where the statistical significance is maximized, in the context of this specific ANN output.

FIG. 4: Comparison of the experimental data with Monte Carlo simulation in the dijet invariant mass sidebands for the ANN output. The $\chi^2$ probability the agreement between the experimental data and the Monte Carlo simulation is $\approx 30\%$. 

$\frac{\alpha - \beta x^2}{\sqrt{\text{Poission fluctuations}}}$

$\frac{1}{2}$
ured background shape are displayed. Figure 6 shows the signal shape that is measured on the data, obtained by subtracting the background from the data. The signal shape is compared to the expected signal shape, normalized according to the measured signal fraction.

![Signal Region: [60,100] GeV/²](image)

FIG. 5: Likelihood fit on data (solid line). The dashed line shows the background estimation, as given by the data. The inset provides a close-up in the signal region.

![Signal shape for fraction $f_s$](image)

FIG. 6: The signal shape measured on data. The plot is made by subtracting the measured background shape from the data. The errors are statistical only. The solid line corresponds to the Monte Carlo simulated events shape for the measured signal fraction. There is a good agreement between the Monte Carlo events shape and the shape seen in the data.

The significance of the result is evaluated using the likelihood ratio, $Q = \mathcal{L}_S + \mathcal{B}(\alpha, \beta, f_s)/\mathcal{L}_B(\alpha, \beta)$, as a test statistic. We test the SM signal plus background hypothesis and the background-only hypothesis by analyzing a set of simulated experiments, as done in data. We perform a fit for the three free parameters, and each time calculate the likelihood ratio. Using Monte Carlo simulated experiments, we estimate a $\approx 2.5\sigma$ statistical significance for the expected signal given its SM cross section. From the data, we measure a $1.9\sigma$ statistical significance. The data are thus compatible with SM expectations, and we estimate an upper limit of the WW and WZ cross section.

The cross section times branching ratio that corresponds to the measured number of signal events is estimated using the formula $\sigma \times BR = N_{signal}/\alpha \cdot \epsilon \cdot \mathcal{L}$, where $N_{signal}$ is the measured number of signal events; $\alpha$ is the signal acceptance, derived from the Monte Carlo simulated events; $\epsilon$ is the global efficiency factor that includes vertex, tracking and trigger efficiencies; and $\mathcal{L}$ is the total integrated luminosity of the data we used. The product of the acceptance times efficiency is estimated separately for electrons and muons.

To assess the effects of systematic uncertainties on the measurement, we address separately two kinds of systematic uncertainties: those that affect the signal fraction, and those that affect the acceptance. The signal fraction uncertainties define the uncertainty in the significance of the measurement.

The dominant systematic uncertainty in the signal fraction come from the background shape parameterization. The background shape is fit to the form PDF$_{BGR} \propto \exp(\alpha x + \beta x^2)$, as already described, which has two parameters $\alpha$ and $\beta$. This form gives an adequate fit to both the Monte Carlo simulation and the data. In order to quantify the size of the systematic uncertainty associated with the background shape, fits with additional parameters in the exponent (from three to six) were carried out. The variations obtained were used to assign the systematic uncertainty.

Other systematic uncertainties in the signal fraction include those originating from the energy scale of the jets (JES). The effect in the signal fraction is quantified by varying the parameters of the signal shape to account for the $\pm 1\sigma$ variations of the JES. We generate simulated experiments with the new parameterizations and fit them with the standard signal parameterization. The difference in the signal fraction from the different fits determines the systematic uncertainty.

The result also depends on the dijet invariant mass resolution which in turn depends on the jet energy resolution. To estimate the systematic uncertainty due to this effect, we introduce an additional Gaussian smearing relative to the jet energy. Other systematic uncertainties that affect the signal shape but have smaller effects on the result are the initial and final state radiation effects.

A summary of all systematic uncertainties on the signal fraction and their effect on the measurement is given in...
The systematic uncertainties affecting the acceptance are evaluated by counting the number of events that pass the selection cuts, after varying the various uncertainty sources. The sources that have been taken into account, as well as the actual effect on the acceptance are listed in Table I

The overall uncertainty on the cross section is given by taking into account the uncertainties in the signal fraction, the acceptance and the luminosity. The total effect is estimated to be 26%. The total uncertainty in the measurement is given by the statistical and systematic uncertainties added in quadrature. Taking into account the systematic uncertainties, the significance of the measurement is 1.7σ.

Table I: The systematic uncertainties and their effect in the signal fraction, the acceptance and finally the cross section. All uncertainties are added in quadrature. 

<table>
<thead>
<tr>
<th>Source</th>
<th>Signal Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jet Energy Scale</td>
<td>10%</td>
</tr>
<tr>
<td>Jet Resolution</td>
<td>10%</td>
</tr>
<tr>
<td>Background Radiation</td>
<td>20%</td>
</tr>
<tr>
<td>Initial State Radiation</td>
<td>5%</td>
</tr>
<tr>
<td>Final State Radiation</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>Source</td>
<td>Acceptance</td>
</tr>
<tr>
<td>Jet Energy Scale</td>
<td>3%</td>
</tr>
<tr>
<td>Jet Resolution</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>Initial State Radiation</td>
<td>2%</td>
</tr>
<tr>
<td>Final State Radiation</td>
<td>3%</td>
</tr>
<tr>
<td>Efficiency Factor</td>
<td>3%</td>
</tr>
<tr>
<td>Source</td>
<td>Cross Section</td>
</tr>
<tr>
<td>Total Signal Fraction</td>
<td>25%</td>
</tr>
<tr>
<td>Total Acceptance</td>
<td>5%</td>
</tr>
<tr>
<td>Luminosity</td>
<td>6%</td>
</tr>
<tr>
<td>Total Effect in Cross Section</td>
<td>26%</td>
</tr>
</tbody>
</table>

Taking into account both the statistical and systematic error in the number of signal events, we measure σ_{WW} × BR(W → ℓντ, W → jets) + σ_{WZ} × BR(W → ℓντ, Z → jets) = 1.47±0.77(stat)±0.38(sys) pb, which is consistent with the SM theoretical prediction for the cross section, 2.09 ± 0.12 pb. We set a 95% confidence level (C.L.) upper limit for the measured cross section. Given that the uncertainties follow a Gaussian distribution, the 95% C.L. limit can be set by the estimated value plus 1.65 standard deviations [16]. The 95% C.L. upper limit set for the cross section is σ × BR < 2.88 pb.

In summary, we have used 1.2 fb$^{-1}$ of CDF II data to search for WW and WZ production in the lepton plus jets final state. We use an ANN to discriminate the signal from the background. This technique improves the expected statistical significance by ≈ 10% and the expected signal fraction by 50%, as compared with an event selection based on conventional cuts. We find no evidence for anomalous WW and WZ production in the lepton plus jets final state and we set a 95% C.L. upper limit for the cross section at σ × BR < 2.88 pb.

We thank the Fermilab staff and the technical staffs of the participating institutions for their vital contributions. This work was supported by the U.S. Department of Energy and National Science Foundation; the Italian Istituto Nazionale di Fisica Nucleare; the Ministry of Education, Culture, Sports, Science and Technology of Japan; the Natural Sciences and Engineering Research Council of Canada; the National Science Council of the Republic of China; the Swiss National Science Foundation; the A.P. Sloan Foundation; the Bundesministerium für Bildung und Forschung, Germany; the Korean Science and Engineering Foundation and the Korean Research Foundation; the Science and Technology Facilities Council and the Royal Society, UK; the Instituto Nacional de Fisica Nucleaire et Physique des Particules/CNRS; the Russian Foundation for Basic Research; the Ministerio de Ciencia e Innovación, and Programa Consolider-Ingenio 2010, Spain; the Slovak R&D Agency; and the Academy of Finland.

[9] We use a cylindrical coordinate system along the proton direction in which θ (ϕ) is the polar (azimuthal) angle. We define η = −ln(tan(θ/2)), p_T = p sin θ, E_T = E sin θ, \vec{E}_T = −[\sum_i E_i \vec{n}_{T,i}], where \vec{n}_{T,i} is a unit vector in the transverse plane that points from the beamline to the i\textsuperscript{th} calorimeter tower, and M_T = [2 · p_T · \vec{E}_T · (1 − \cos \Delta φ(\text{lepton},\vec{E}_T))^{1/2}].
[14] Neural Networks for Pattern Recognition, Christopher M. Bishop, Oxford Science Publications.
