The following full text is an author's version which may differ from the publisher's version.

For additional information about this publication click this link. http://hdl.handle.net/2066/75613

Please be advised that this information was generated on 2019-04-17 and may be subject to change.
Evidence of $WW + WZ$ production with lepton + jets final states in $pp$ collisions at $\sqrt{s} = 1.96$ TeV

We present first evidence for \(W W + WZ\) production in lepton+jets final states at a hadron collider. The data correspond to 1.07 fb\(^{-1}\) of integrated luminosity collected with the D0 detector at the Fermilab Tevatron in \(pp\) collisions at \(\sqrt{s} = 1.96\) TeV. The observed cross section for \(W W + WZ\) production is 20.2 ± 4.5 pb, consistent with the standard model and more precise than previous measurements in fully leptonic final states. The probability that background fluctuations alone produce this excess is < 5.4 × 10\(^{-6}\), which corresponds to a significance of 4.4 standard deviations.

PACS numbers: 14.70.Fm, 14.70.Hp, 13.85.Ni, 13.85.Qk

The production of vector-boson pairs in \(pp\) collisions (\(WW\), \(WZ\), or \(ZZ\)) provides important tests of the elec-
troweak sector of the standard model (SM). The next-to-leading-order (NLO) cross sections for WW and WZ production in $pp$ collisions at $\sqrt{s} = 1.96$ TeV predicted by the SM are $\sigma(WW) = 12.4 \pm 0.8$ pb and $\sigma(WZ) = 3.7 \pm 0.3$ pb [1]. A discrepancy with this expectation or deviations in the predicted kinematic distributions could signal the presence of new physics, e.g., originating from anomalous trilinear gauge boson couplings [2]. The production of two weak bosons is also relevant to searches for the Higgs boson or for new particles in extensions of the SM. Production of WW and WZ in $pp$ collisions at the Fermilab Tevatron Collider has thus far been observed only in fully leptonic decay modes [3, 4]. Previous searches for WW and WZ in lepton+jets final states [5, 6], which benefit from a higher transverse energy to be $E_T > 20$ GeV; and at least two particles with $|\eta| < 1.1$ (2.0) for electrons (muons); the imbalance transverse momentum $p_T > 20$ GeV and pseudorapidity $|\eta|$ make the WV signal in lepton+jets final states more distinguishable from W+jets background and more accessible to measurement than in the past [5, 6]. This analysis also provides a valuable proving ground for such advanced techniques, now ubiquitous in Higgs searches at the Tevatron.

We analyze 1.07 fb$^{-1}$ of data collected with the D0 detector [8] at a center-of-mass energy of 1.96 TeV at the Tevatron. Candidate $WV\rightarrow \ell
qqqq$ events must pass a trigger based on a single electron or electron+jet(s) requirement that has an efficiency of $98 \pm 3\%$. A suite of triggers for $\mu
qqqq$ candidate events achieves an efficiency of $>95\%$ at $95\%$ confidence level.

To select $WV\rightarrow \ell
qqqq$ candidates, we require: a single reconstructed lepton (electron or muon) [9] with transverse momentum $p_T > 20$ GeV and pseudorapidity $|\eta| < 1.1$ (2.0) for electrons (muons); the imbalance in transverse energy to be $E_T > 20$ GeV; and at least two jets [10] with $p_T > 20$ GeV and $|\eta| < 2.5$. The jet of highest $p_T$ must have $p_T > 30$ GeV. To reduce background from processes that do not contain $\ell$, we require a “transverse” mass [11] of $M_T^T > 35$ GeV. The lepton must be spatially matched to a track reconstructed in the central tracker that originates from the primary vertex. Electrons (muons) must be isolated from other particles in the calorimeter (and central tracker) [12].

Signal and background processes containing charged leptons are modeled via Monte Carlo (MC) simulation. The signal includes all possible W and Z decays, including their decays to leptons. The diboson signal (WW and WZ) is generated with PYTHIA [13] using CTEQ6L parton distribution functions (PDFs). The fixed-order matrix element (FOME) generator ALPGEN [14] with CTEQ6L1 PDFs is used to generate W+jets, Z+jets, and $t\bar{t}$ events to leading order at the parton level. The FOME generator COMPHEP [15] is used to produce single top-quark MC samples. ALPGEN and COMPHEP are interfaced to PYTHIA for subsequent parton showering and hadronization. All simulated events undergo a GEANT-based [16] detector simulation and are reconstructed using the same programs as used for D0 data. The MC samples are normalized using next-to-leading-order (NLO) or next-to-next-to-leading-order predictions for SM cross sections, except W+jets which is scaled to the data.

The probability for multijet events with misidentified leptons to pass all selection requirements is small; however, because of the copious production of multijet events, the background from this source cannot be ignored. For $\mu
qqqq$, the multijet background is modeled with data that fail the muon isolation requirements, but pass all other selections. The normalization is determined from a fit to the $M_T^T$ distribution. For $e
qqqq$, the multijet background is estimated using a “loose-but-not-tight” data sample obtained by selecting events that pass loosened electron quality requirements, but fail the tight electron quality criteria [9]. This sample is normalized by the probability for a jet that passes the “loose” electron requirements to also pass the tight requirement. Both $\mu
qqqq$ and $e
qqqq$ multijet samples are corrected for contributions from all processes modeled through MC.

Accurate modeling of the selected events is vital. The dominant background is W+jets, and the modeling of ALPGEN W+jets and sources of uncertainty are therefore studied in great detail. Comparison of ALPGEN with other generators and with data shows discrepancies [17] in jet $\eta$ and dijet angular separation. Data are used to correct these quantities in the ALPGEN W+jets and Z+jets samples. The possible bias in this procedure from the presence of the diboson signal in data is small, but is nevertheless taken into account via a systematic uncertainty. Systematic effects on the differential distributions of the ALPGEN W+jets and Z+jets MC events from changes of the renormalization and factorization scales and of the parameters used in the MLM parton-jet matching algorithm [18] are also considered. Uncertainties on PDFs, as well as uncertainties from object reconstruction and identification, are evaluated for all MC samples. We consider the effect of systematic uncertainty both on the normalization and on the shape of differential distributions for signal and backgrounds [19].

The signal and the backgrounds are further separated using a multivariate classifier to combine information from several kinematic variables. This analysis uses a
FIG. 1: (a) The RF output distribution from the combined $evqq$ and $\mu vqq$ channels for data and MC predictions following the fit of MC to data. (b) A comparison of the extracted signal (filled histogram) to background-subtracted data (points), along with the ±1 standard deviation (s.d.) systematic uncertainty on the background. The residual distance between the data points and the extracted signal, divided by the total uncertainty, is given at the bottom.

FIG. 2: (a) The dijet mass distribution from the combined $evqq$ and $\mu vqq$ channels for data and MC predictions following the fit to the RF output. (b) A comparison of the extracted signal (filled histogram) to background-subtracted data (points), along with the ±1 standard deviation (s.d.) systematic uncertainty on the background. The residual distance between the data points and the extracted signal, divided by the total uncertainty, is given at the bottom.

TABLE I: Measured number of events for signal and each background after the combined fit (with total uncertainties determined from the fit) and the number observed in data.

<table>
<thead>
<tr>
<th></th>
<th>$evqq$ channel</th>
<th>$\mu vqq$ channel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diboson signal</td>
<td>436 ± 36</td>
<td>527 ± 43</td>
</tr>
<tr>
<td>W+jets</td>
<td>10100 ± 500</td>
<td>11910 ± 590</td>
</tr>
<tr>
<td>Z+jets</td>
<td>387 ± 61</td>
<td>1180 ± 180</td>
</tr>
<tr>
<td>tt + single top</td>
<td>436 ± 57</td>
<td>426 ± 54</td>
</tr>
<tr>
<td>Multijet</td>
<td>1100 ± 200</td>
<td>328 ± 83</td>
</tr>
<tr>
<td>Total predicted</td>
<td>12460 ± 550</td>
<td>14370 ± 620</td>
</tr>
<tr>
<td>Data</td>
<td>12473</td>
<td>14392</td>
</tr>
</tbody>
</table>

Random Forest (RF) classifier [20, 21]. Thirteen well-modeled kinematic variables [19] that demonstrate a difference in probability density between signal and at least one of the backgrounds, such as dijet mass and $E_T$, are used as inputs to the RF. The RF is trained using half of each MC sample. The other halves, along with the multijet background samples, are then evaluated by the RF and used in the measurement.

The signal cross section is determined from a fit of signal and background RF templates to the data by minimizing a Poisson $\chi^2$ function with respect to variations in the systematic uncertainties [7]. The magnitude of systematic uncertainties is effectively constrained by the regions of the RF distribution with low signal over background. A Gaussian prior is used for each systematic uncertainty. Different uncertainties are assumed to be mutually independent, but those common to multiple samples or lepton channels are assumed to be 100% correlated.

The fit simultaneously varies the $WV$ and $W+\text{jets}$ contributions, thereby also determining the normalization factor for the $W+\text{jets}$ MC sample. This obviates the need for using the predicted ALPGEN cross section, and provides a more rigorous approach that incorporates an unbiased uncertainty from $W+\text{jets}$ when extracting the $WV$ cross section. The normalization factor from the fit for the $W+\text{jets}$ component is 1.53 ± 0.13, similar to the expected ratio of NLO to LO cross sections [22]. The measured yields for signal and each background are given in Table I. Table II contains...
Fitted signal

Observed p-value (significance) II, although less precise, is

Expected p-value (significance) than previous independent measurements of WW and

events at a hadron collider. The result is more precise

large as observed in data is $5.4 \times 10^{-6}$, corresponding
certainties discriminant.

standard deviations when using the RF output (dijet mass)

X2 fit with respect to variations in the systematic un-

tation, which includes the relative data/MC uncertainty

in each channel (p-values), separately and combined, are

by the SM). The probabilities that background fluctu-

ations could produce the expected and observed signal

in channel (p-values), separately and combined, are shown

in Table II, although less precise, is consistent with that obtained using the RF output.

The significance of the measurement is obtained via

fits of the signal+background hypothesis to pseudo-data

cases drawn from the background-only hypothesis [23].

The observed (or expected) significance corresponds to the fraction of outcomes that yield a WW cross section

at least as large as that measured in data (as predicted by the SM). The probabilities that background fluctu-

ations could produce the expected and observed signal in each channel (p-values), separately and combined, are shown in Table II, along with their corresponding significance (equivalent one-sided Gaussian probabilities). The $\chi^2$ fit with respect to variations in the systematic uncertainties [7] results in an improvement of the expected significance of the result from 2.4 (1.6) to 3.6 (2.9) standard deviations when using the RF output (dijet mass) discriminant.

In summary, we measure $\sigma(WV) = 20.2 \pm 4.5$ pb (with $V=W$ or $Z$) in pp collisions at $\sqrt{s} = 1.96$ TeV. The probability that the backgrounds fluctuate to give an excess as large as observed in data is $< 5.4 \times 10^{-6}$, corresponding to a significance of 4.4 standard deviations. This represents the first evidence for WW production in lepton+jets events at a hadron collider. The result is more precise than previous independent measurements of WW and WZ yields in fully leptonic final states [3, 4] and consis-
tent with the SM prediction of $\sigma(WV) = 16.1 \pm 0.9$ pb [1].

This work clearly demonstrates the ability of the D0 experiment to isolate a small signal in a large background in a final state of direct relevance to searches for a low mass Higgs, and thereby validates the analytical methods used in searches for Higgs bosons at the Tevatron [24].

We thank the staffs at Fermilab and collaborating institutions, and acknowledge support from the DOE and NSF (USA); CEA and CNRS/IN2P3 (France); FASI, Rosatom and RFBR (Russia); CNPq, FAPERJ, FAPESP and FUNDUNESP (Brazil); DAE and DST (India); Colciencias (Colombia); CONACyT (Mexico); KRF and KOSEF (Korea); CONICET and UBACyT (Argentina); FOM (The Netherlands); STFC (United Kingdom); MSMT and GACR (Czech Republic); CRC Program, CFI, NSERC and WestGrid Project (Canada); BMBF and DFG (Germany); SFI (Ireland); The Swedish Research Council (Sweden); CAS and CNSF (China); and the Alexander von Humboldt Foundation (Germany).

[a] Visitor from Augustana College, Sioux Falls, SD, USA.
[b] Visitor from Rutgers University, Piscataway, NJ, USA.
[c] Visitor from The University of Liverpool, Liverpool, UK.
[d] Visitor from H. Physikalisches Institut, Georg-August-
University, Göttingen, Germany.
[e] Visitor from Centro de Investigacion en Computacion -
IPN, Mexico City, Mexico.
[f] Visitor from ECFM, Universidad Autonoma de Sinaloa,
Culiacán, Mexico.
[g] Visitor from Helsinki Institute of Physics, Helsinki, Finland.
[h] Visitor from Universitat Bern, Bern, Switzerland.
[i] Visitor from Universitat Zürich, Zürich, Switzerland.
[j] Deceased.

[1] J. M. Campbell and R. K. Ellis, Phys. Rev. D 60, 113006 (1999). Cross sections were calculated with the same parameter values given in the paper, except with $\sqrt{s} = 1.96$ TeV.
[10] G. C. Blazey et al., arXiv:hep-ex/0005012 (2000). The seeded cone algorithm with radius 0.5 was used.
[19] See attached supplemental material.
Supplemental Material:

SYSTEMATIC UNCERTAINTIES

Table III gives the % systematic uncertainties for Monte Carlo simulations and multi-jet estimates. We consider the effect of systematic uncertainty both on the normalization and on the shape of differential distributions for signal and backgrounds. Although Table III lists an uncertainty for the W+jets simulation, this uncertainty is not used when measuring the diboson signal cross section, for which the W+jets normalization is a free parameter. However, the size of the uncertainty must be specified for generating the pseudo-data used in the estimation of significance. Also in the table is the contribution of each systematic uncertainty to the total systematic uncertainty of 3.6 pb on the measured cross section, $\sigma_{\text{meas}(W^+)}$. This total systematic uncertainty is obtained from the systematic uncertainties on the parameter in the fit to the Random Forest (RF) output, $\sigma_{\text{meas}(W^+)} / \sigma_{\text{th}(W^+)}$, by multiplying each contribution by the theoretical cross section $\sigma_{\text{th}(W^+)}$. The additional uncertainty on the integrated luminosity for data (6.1%) is therefore considered separately.

INPUT VARIABLES TO THE RANDOM FOREST CLASSIFIER

The 13 kinematic variables used in the RF classifier are listed below, and their distributions are shown in Fig. 3. The variables are derived from characteristics of objects reconstructed from observables in each event and can be loosely classified into three categories: (i) variables based on the kinematics of individual objects, (ii) variables based on the kinematics of multiple objects, and (iii) variables based on the angular relationships among objects. Several variables are calculated using the four-momentum of the dijet system or the leptonic W candidate ($W^{\ell\nu}$). The dijet system is defined as the four-momentum sum of the jets with highest $p_T$ (Jet_1) and second highest $p_T$ (Jet_2). $W^{\ell\nu}$ is reconstructed from the charged lepton and the $\vec{E}_T$. The neutrino from the $W \rightarrow \ell\nu$ decay is assigned the transverse momentum defined by $\vec{E}_T$ and a longitudinal momentum that is calculated assuming the mass of the W for $\ell\nu$ ($M_W = 80.4$ GeV). Of the two possible solutions, we choose the one that provides the smaller total invariant mass of all objects in the event.

• Kinematics of Individual Objects:

1. The imbalance in transverse energy ($\Delta E_T$), which is defined by the imbalance in transverse momentum as determined from the summing of products of energies and cosines of polar angles of calorimeter cells relative to the center of the detector (corrected for transverse momenta of muons and energy scales for jets and electrons in the event).

2. The jet with second highest $p_T$: $p_T$(Jet_2).

• Kinematics of Multiple Objects:

1. The “transverse W mass” reconstructed from the charged lepton and the $\vec{E}_T$: $M_W^T = \sqrt{2 p_T^\ell E_T (1 - \cos(\Delta\phi(\ell, \vec{E}_T)))}$.

2. The $p_T$ of the $W^{\ell\nu}$ candidate.

3. The invariant mass of the dijet system.

4. The magnitude of the leading jet momentum perpendicular to the plane of the dijet system: $|\vec{p}_{T\ell}(\text{Jet}_1 + \text{Jet}_2)| / |\vec{p}_T(\text{Jet}_1 + \text{Jet}_2)|$, where “x” represents the usual vector cross product. This variable is calculated in the rest frame of the $W^{\ell\nu}$ candidate and is denoted $p_T^{\ell\nu}(\text{DiJet}, \text{Jet}_1)$.

5. The magnitude of the second-leading jet momentum perpendicular to the plane of the dijet system: $|\vec{p}_{T\ell}(\text{Jet}_1 + \text{Jet}_2)| / |\vec{p}_T(\text{Jet}_1 + \text{Jet}_2)|$. This variable is calculated in the laboratory frame and is denoted $p_T^{\ell\nu}(\text{DiJet}, \text{Jet}_2)$.

6. The angular separation between the two jets of highest $p_T$, weighted by the ratio of the transverse momentum of the second-leading jet and the $W^{\ell\nu}$ candidate: $\Delta R(\text{Jet}_1, \text{Jet}_2) = p_T(\text{Jet}_2) / p_T(\text{Jet}_1)$.

7. The “centrality” of the charged lepton and jets system, defined as the scalar sum of transverse momenta divided by the sum of energies of the charged lepton and all jets in the event.

• Angular Relationships of Objects:

1. The azimuthal separation between the charged lepton and the $\vec{E}_T$ vector: $\Delta\phi(\vec{E}_T, \text{lepton})$.

2. The cosine of the angle between the dijet system and the leading jet in the laboratory frame: $\cos(\angle(\text{DiJet}, \text{Jet}_1))$.

3. The cosine of the angle between the dijet system and the second-leading jet in the laboratory frame: $\cos(\angle(\text{DiJet}, \text{Jet}_2))$.

4. Cosine of the angle between leading jet and the $W^{\ell\nu}$ candidate: $\cos(\angle(W^{\ell\nu}, \text{Jet}_1))$, evaluated in the rest frame of the dijet system.
TABLE III: The % systematic uncertainties for Monte Carlo simulations and multijet estimates. Uncertainties are identical for both lepton channels except where otherwise indicated. The nature of the uncertainty, i.e., whether it refers to a differential dependence (D) or just normalization (N), is also provided. The values for uncertainties with a differential dependence correspond to the maximum amplitude of fluctuations in the RF output. Also provided is the contribution of each source to the total systematic uncertainty of 3.6 pb on the measured cross section, which does not include the additional uncertainty of 6.1% for the luminosity.

<table>
<thead>
<tr>
<th>Source of systematic uncertainty</th>
<th>Diboson signal</th>
<th>W+jets</th>
<th>Z+jets</th>
<th>Top</th>
<th>Multijet</th>
<th>Nature</th>
<th>$\Delta\sigma$ (pb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trigger efficiency, $evqq$ channel</td>
<td>$+2/\pm 3$</td>
<td>$+2/\pm 3$</td>
<td>$+2/\pm 3$</td>
<td>$+2/\pm 3$</td>
<td>N</td>
<td>$&lt; 0.1$</td>
<td></td>
</tr>
<tr>
<td>Trigger efficiency, $\mu qq$ channel</td>
<td>$+0/\pm 5$</td>
<td>$+0/\pm 5$</td>
<td>$+0/\pm 5$</td>
<td>$+0/\pm 5$</td>
<td>D</td>
<td>$&lt; 0.1$</td>
<td></td>
</tr>
<tr>
<td>Lepton identification</td>
<td>$\pm 4$</td>
<td>$\pm 4$</td>
<td>$\pm 4$</td>
<td>$\pm 4$</td>
<td>N</td>
<td>$&lt; 0.1$</td>
<td></td>
</tr>
<tr>
<td>Jet identification</td>
<td>$\pm 4$</td>
<td>$\pm 4$</td>
<td>$\pm 4$</td>
<td>$\pm 4$</td>
<td>D</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>Jet energy scale</td>
<td>$\pm 4$</td>
<td>$\pm 9$</td>
<td>$\pm 9$</td>
<td>$\pm 9$</td>
<td>D</td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>Jet energy resolution</td>
<td>$\pm 3$</td>
<td>$\pm 4$</td>
<td>$\pm 4$</td>
<td>$\pm 4$</td>
<td>N</td>
<td>$&lt; 0.1$</td>
<td></td>
</tr>
<tr>
<td>Cross section</td>
<td>$\pm20^a$</td>
<td>$\pm 6$</td>
<td>$\pm 10$</td>
<td></td>
<td>N</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>Multijet normalization, $evqq$ channel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\pm 20$</td>
<td>N</td>
</tr>
<tr>
<td>Multijet normalization, $\mu qq$ channel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\pm 30$</td>
<td>N</td>
</tr>
<tr>
<td>Multijet shape, $evqq$ channel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\pm 6$</td>
<td>D</td>
</tr>
<tr>
<td>Multijet shape, $\mu qq$ channel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\pm 10$</td>
<td>D</td>
</tr>
<tr>
<td>Diboson signal NLO/LO shape</td>
<td>$\pm 10$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>D</td>
</tr>
<tr>
<td>Parton distribution function</td>
<td>$\pm 1$</td>
<td>$\pm 1$</td>
<td>$\pm 1$</td>
<td></td>
<td></td>
<td>D</td>
<td>0.2</td>
</tr>
<tr>
<td>ALPGEN $\eta$ and $\Delta R$ corrections</td>
<td>$\pm 1$</td>
<td>$\pm 1$</td>
<td></td>
<td></td>
<td></td>
<td>D</td>
<td>$&lt; 0.1$</td>
</tr>
<tr>
<td>Renormalization and factorization scale</td>
<td>$\pm 3$</td>
<td>$\pm 3$</td>
<td></td>
<td></td>
<td></td>
<td>D</td>
<td>0.9</td>
</tr>
<tr>
<td>ALPGEN parton-jet matching parameters</td>
<td>$\pm 4$</td>
<td>$\pm 4$</td>
<td></td>
<td></td>
<td></td>
<td>D</td>
<td>2.4</td>
</tr>
</tbody>
</table>

$^a$The uncertainty on the cross section for $W+$jets is not used in the diboson signal cross section measurement (the $W+$jets normalization is a free parameter); however, it is needed for generating pseudo-data to estimate the significance of the observed signal.

CORRELATION BETWEEN DIJET INVARIANT MASS AND RF OUTPUT

There is a high degree of correlation between the dijet invariant mass and the RF output. This can be observed in the dijet invariant mass distributions for events with low, intermediate and high values for the RF output shown in Fig. 4. As expected, events in the low region of the RF output correspond to the background dominated sidebands of the dijet invariant mass distribution and events in the high region of the RF output correspond to the signal resonance region of the dijet invariant mass. The purity of the signal in the dijet invariant mass distribution is enhanced for high values of the RF output because a substantial fraction of the background events in the dijet invariant mass signal region has been moved to the intermediate region of the RF output.
FIG. 3: Distributions of the RF input variables for the combined $e\nu\nu\bar{q}$ and $\mu\nu\nu\bar{q}$ channels comparing the data with the MC predictions, following the fit of MC to data.
FIG. 4: Distributions of the dijet invariant mass for the combined $e\nu q\bar{q}$ and $\mu\nu q\bar{q}$ channels comparing the data with the MC predictions for events in three regions of the RF output: (a) $0 \leq \text{RF output} \leq 0.33$, (b) $0.33 < \text{RF output} \leq 0.66$ and (c) $0.66 < \text{RF output} \leq 1$. 