

FIRST EVIDENCE OF $WW/WZ \rightarrow \ell\nu q\bar{q}$ AT THE TEVATRON

JOSEPH HALEY

(For the D0 Collaboration)

Princeton University Department of Physics

Princeton, NJ, USA



We present the first evidence from a hadron collider of $WW + WZ$ production with semi-leptonic decays. The data were recorded by the D0 detector at the Fermilab Tevatron and correspond to 1.07 fb^{-1} of integrated luminosity obtained in proton-antiproton collisions at $\sqrt{s} = 1.96 \text{ TeV}$. The cross section observed for $WW + WZ$ production is $20.2 \pm 4.5 \text{ pb}$ with a significance of 4.4 standard deviations.

1 Introduction

There are many reasons for studying $WW/WZ \rightarrow \ell\nu q\bar{q}$ at the Tevatron. From the electroweak prospective, diboson production provides a probe of self-interactions of vector bosons. Deviations from the Standard Model (SM) of these trilinear gauge boson coupling would affect the cross sections and event kinematics of diboson production¹. The cross sections for diboson production at the Tevatron had previously only been measured for the fully leptonic final states^{2,3}, so this analysis provides a compliment to the previous measurements.

Reconstruction of WW and WZ events in semi-leptonic final states represents a challenge in separating signal from the dominant background of a W boson produced in association with jets. This is a challenge shared by many Higgs boson searches, *e.g.* $WH \rightarrow \ell\nu b\bar{b}$, making this measurement a benchmark for these similar Higgs boson searches. Furthermore, this analysis provides a proving ground for the multivariate event-classification schemes and the accompanying statistical techniques⁴ that are used for the Tevatron Higgs boson searches in the entire mass range allowed by the SM.

2 Event Selection

To select candidate events for $p\bar{p} \rightarrow WW/WZ \rightarrow \ell\nu q\bar{q}$, we required a single reconstructed lepton (electron or muon)⁶ with transverse momentum $p_T > 20 \text{ GeV}$ and $|\eta| < 1.1$ (for electrons) or

$|\eta| < 2$ (for muons), an imbalance in transverse momentum $\cancel{E}_T > 20$ GeV, and at least two jets⁷ with $p_T > 20$ GeV and $|\eta| < 2.5$. The leading jet (*i.e.* with the highest p_T) was also required to have $p_T > 30$ GeV. To reduce background from processes that do not contain $W \rightarrow \ell\nu$, we required a transverse W mass of $M_T^W > 35$ GeV, where $M_T \equiv \sqrt{(E_T)^2 - (\vec{p}_T)^2}$ ⁸. The electron or muon trajectories were required to be isolated from other objects in the calorimeter, and had to match a track reconstructed in the central tracking system that originated from the primary vertex. Also, the muon had to be reconstructed as an isolated track in the central tracking system. The resulting kinematic distributions are shown in Fig. 1.

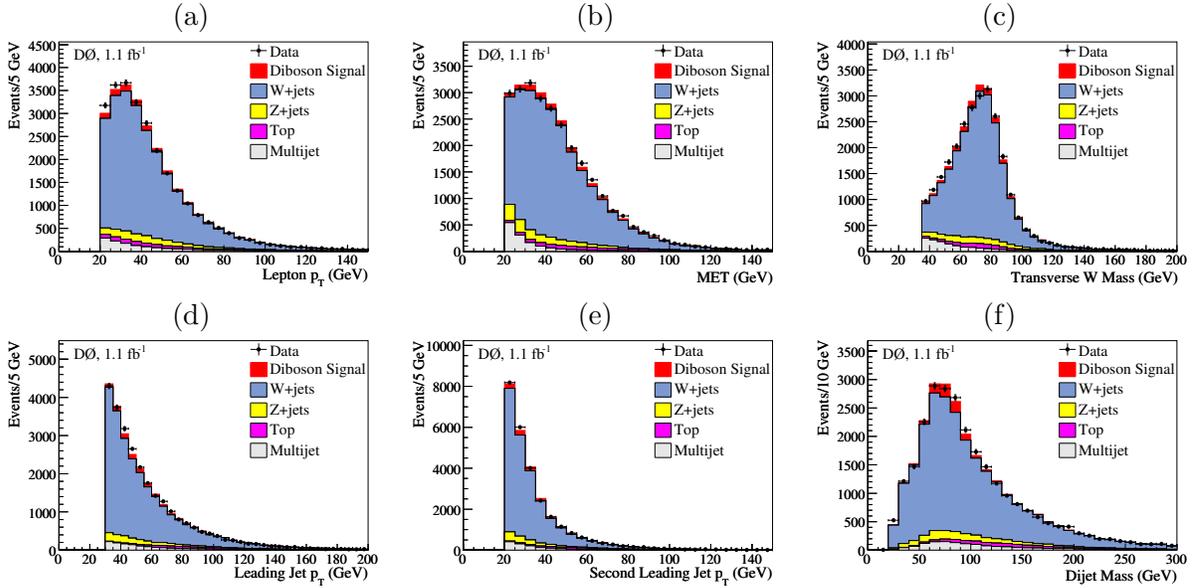


Figure 1: Kinematic distributions after all selection requirements: (a) p_T of lepton; (b) \cancel{E}_T ; (c) transverse W mass; (d) p_T of leading jet; (e) p_T of second-leading jet; (f) dijet mass.

3 Data Sample

The data were collected with the D0 detector⁵ at the Fermilab Tevatron Collider at a center-of-mass energy of $\sqrt{s} = 1.96$ TeV. The events studied in this analysis correspond to 1.07 fb^{-1} of integrated luminosity collected during Run IIa (2002-2006). To be considered for analysis, events in the $e\nu q\bar{q}$ channel were required to pass at least one single electron or electron+jet(s) trigger. The resulting trigger efficiency was $98_{-3}^{+2}\%$. A suite of triggers was used for the $\mu\nu q\bar{q}$ channel resulting in a trigger efficiency of nearly 100%.

4 Signal and Background Estimations

Monte Carlo generators were used to simulate the signal and background samples that contained a charged lepton in the final state. Signal events were generated with PYTHIA⁹ using CTEQ6L parton distribution functions (PDF). ALPGEN¹⁰ with CTEQ6L1 PDFs was used to generate W +jets, Z +jets, and $t\bar{t}$ events and COMPHEP¹¹ with CTEQ6L1 PDFs was used to simulate single-top events. All ALPGEN and COMPHEP events used PYTHIA for parton showering and hadronization. After generation, the events underwent a GEANT-based¹² detector simulation before being reconstructed with the same programs as the data.

With the exception of W +jets, all background MC samples were normalized to next-to-leading-order (NLO) or next-to-next-to-leading-order SM predictions. The W +jets normaliza-

Table 1: Measured number of events for signal and each background after the combined fit of the RF distribution (with total uncertainties determined from the fit) and the observed number of selected events.

	$e\nu q\bar{q}$ channel	$\mu\nu q\bar{q}$ channel
Diboson signal	436 ± 36	527 ± 43
W +jets	10100 ± 500	11910 ± 590
Z +jets	387 ± 61	1180 ± 180
$t\bar{t}$ + single top	436 ± 57	426 ± 54
Multijet	1100 ± 200	328 ± 83
Total predicted	12460 ± 550	14370 ± 620
Data	12473	14392

tion was determined simultaneously with the signal cross section by a fit to data, as discussed later.

The probability for a multijet event to mimic a lepton and pass all selection cuts was quite small; however, because the cross section for multijet production is so large, the background from multijet events had to be accounted for. For the $\mu\nu q\bar{q}$ channel, the multijet background was modeled with “anti-isolated” data corresponding to events that failed the muon isolation requirements, but passed all other selections. The kinematic distributions of the anti-isolated data were corrected for contributions from processes already modeled via MC. The normalization of the multijet background in the muon channel was determined from a fit to the transverse W mass distribution of the $\mu + \nu$ system.

For the $e\nu q\bar{q}$ channel, the multijet background was estimated using a “loose-but-not-tight” (LNT) data sample obtained by selecting events that passed a loosened electron-quality requirement, but did not pass the electron-quality requirement of the final selection⁶. To estimate the correct rate for multijet events, a weight was applied to each LNT event based on the probability for a jet to mimic an electron. Also, the contribution from events that were already modeling via MC was subtracted.

5 MC Corrections and Systematic Uncertainties

As one can see from Table 1, contributions to the selected events was dominated by the background from W +jets. Therefore, accurate modeling of the W +jets background was of particular importance. We performed detailed studies of the ALPGEN W +jets MC sample and associated sources of uncertainty. Comparison with other generators and data showed discrepancies between the modeling of jet η and ΔR between jets¹³. Therefore, the data were used to correct these quantities in the ALPGEN W +jets and Z +jets samples. The effect of the diboson signal on the derived corrections was small, but nonetheless taken into account via a systematic uncertainty assigned to the procedure. The ALPGEN W +jets sample was also assigned systematic uncertainties for variations of the renormalization (and factorization) scale and jet-parton matching parameters¹⁴. PDF uncertainties were evaluated for all of the MC samples, as were uncertainties from object reconstruction and identification. A full list of the systematic uncertainties and the magnitude of each is given in Table 2. We considered systematic uncertainties that affected both normalization and the shapes of kinematic distributions.

6 Multivariate Classification

Improved separation between the signal and the backgrounds was achieved using a multivariate classification technique to combine information from several kinematic variables. The technique used was a random forest (RF) classifier^{15,16} from the STATPATTERNRECOGNITION¹⁶ software

Table 2: The % systematic uncertainties for Monte Carlo simulations and multijet estimates. Uncertainties are identical for both lepton channels except where indicated otherwise. The nature of the uncertainty, i.e., whether it had a differential dependence (D) or just normalization (N), is also provided. The values for uncertainties with a differential dependence correspond to the RMS amplitudes in the RF output distribution. Also provided is the contribution of each source to the total systematic uncertainty of 3.6 pb on the measured cross section.

Source of systematic uncertainty	Diboson	W +jets	Z +jets	Top	Multijet	$\Delta\sigma$ (pb)
Trigger efficiency, $e\nu q\bar{q}$ channel	+2/ - 3	+2/ - 3	+2/ - 3	+2/ - 3		< 0.1
Trigger efficiency, $\mu\nu q\bar{q}$ channel	+0/ - 5	+0/ - 5	+0/ - 5	+0/ - 5		< 0.1
Lepton identification	± 4	± 4	± 4	± 4		< 0.1
Jet identification	± 1	± 1	± 1	$\pm < 1$		0.3
Jet energy scale	± 4	± 9	± 9	± 4		1.9
Jet energy resolution	± 3	± 4	± 4	± 4		< 0.1
Cross section		± 20	± 6	± 10		1.1
Multijet normalization, $e\nu q\bar{q}$ channel					± 20	0.9
Multijet normalization, $\mu\nu q\bar{q}$ channel					± 30	0.5
Multijet shape, $e\nu q\bar{q}$ channel					± 6	< 0.1
Multijet shape, $\mu\nu q\bar{q}$ channel					± 10	< 0.1
Diboson signal NLO/LO shape	± 10					< 0.1
Parton distribution function	± 1	± 1	± 1	± 1		0.2
ALPGEN η and ΔR corrections		± 1	± 1			< 0.1
Renormalization and factorization scale		± 3	± 3			0.9
ALPGEN parton-jet matching parameters		± 4	± 4			2.4

package. The RF algorithm creates many decision tree classifiers, which are basically a series of optimized binary splits to separate signal from background. The RF is then formed by taking the average of all of the decision trees. The key to the RF is that each decision tree uses only a subset of the input variables (selected randomly for each tree) and is trained on a bootstrap replica¹⁵ of the full training set. This results in each of the trees generalizing differently to unseen data because each tree was trained with differently. The net effect of then averaging all the trees is an accurate and stable classifier.

The inputs to the RF were thirteen well-modeled kinematic variables that demonstrated a difference in probability density between signal and at least one of the backgrounds. A RF for each channel was trained using one half of each MC sample. The other halves, along with the multijet background samples, were used to evaluate the RF output distributions for comparison to the data. These RF output distributions were then used to measure the excess of events in the data consistent with the kinematics of WW and WZ production (over that expected from multijet and other SM processes).

7 Cross Section Measurement

The cross section for $WW + WZ$ production was determined from a fit of signal and background RF templates to the data by minimizing a Poisson χ^2 function within variations of the systematic uncertainties⁴. The systematic uncertainties were treated as Gaussian-distributed uncertainties on the expected numbers of signal and background events in each bin of the RF distribution. Each individual uncertainty was treated as 100% correlated between channels, samples, and from bin to bin. Different sources of uncertainty were assumed to be independent.

The normalizations of the RF templates for the signal and the W +jets background were unconstrained in the fit; allowing the fit to simultaneously measured the signal cross section and determine the normalization of the dominant background. This approach eliminated the need to use the W +jets cross section predicted by ALPGEN and provided an unbiased uncertainty for the normalization of the dominant background. As a check of the procedure, the fit yielded an effective k-factor of 1.53 ± 0.13 that needed to be applied to the ALPGEN cross section to

best match the data, which is close to what one would expect from the ratio of NLO to LO predictions for the W +jets cross section.

Table 3 contains the results of the fit in the $evq\bar{q}$, $\mu\nu q\bar{q}$, and the combined channels. The combined distribution of the RF output after the combined fit and the same plot with the background subtracted are shown in Fig. 2. Also in Fig. 2 is the background-subtracted plot for the dijet mass distributions showing the resonant dijet signal peak observed in data. The common behavior of each fit indicates a $WW + WZ$ cross section consistent with, though somewhat larger than, the expected SM value of $\sigma(WW + WZ) = 16.1$ pb¹⁷. The combined lepton channel cross section fit yielded a total value of $20.2 \pm 2.5(\text{stat}) \pm 3.6(\text{sys}) \pm 1.2(\text{lum})$ pb, which is slightly less than one standard deviation from expectation.

Table 3 also provides the result from performing the measurement using only the dijet mass distribution. As expected, the measurement from the dijet mass distribution was less precise than from the RF because the RF was better at discriminating signal from background.

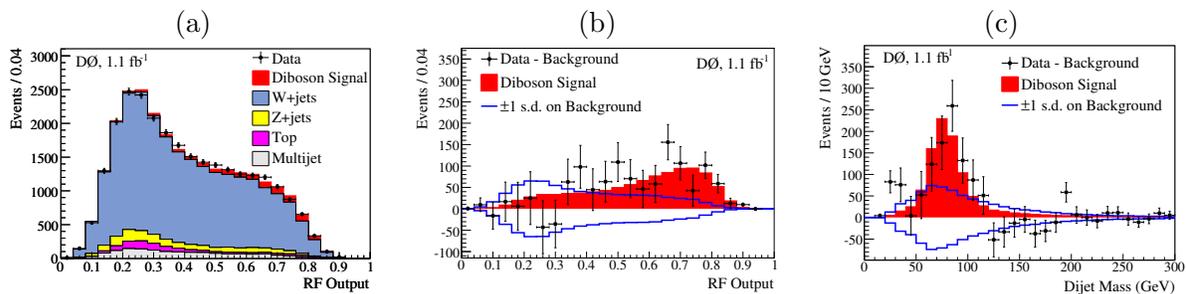


Figure 2: The distributions after cross section fit of the RF distribution: (a) RF output; (b) RF output with background subtracted; (c) dijet mass with background subtracted.

Table 3: The signal cross section determined from a simultaneous fit to the data of the $WW + WZ$ cross section and the normalization factor for W +jets.

Channel	Fitted signal σ (pb)
$evq\bar{q}$ RF Output	$18.0 \pm 3.7(\text{stat}) \pm 5.2(\text{sys}) \pm 1.1(\text{lum})$
$\mu\nu q\bar{q}$ RF Output	$22.8 \pm 3.3(\text{stat}) \pm 4.9(\text{sys}) \pm 1.4(\text{lum})$
Combined RF Output	$20.2 \pm 2.5(\text{stat}) \pm 3.6(\text{sys}) \pm 1.2(\text{lum})$
Combined Dijet Mass	$18.5 \pm 2.8(\text{stat}) \pm 4.9(\text{sys}) \pm 1.1(\text{lum})$

8 Significance

Arguably just as important as the cross sections measurement is the significance of the measurement. The expected and observed significances were obtained via fits of the signal plus background hypothesis to MC events drawn from the background-only hypothesis¹⁸. The pseudo-data samples were generated from random Poisson trials seeded by the predicted number of background events smeared within the systematic uncertainties. A measurement of the signal cross section was performed on each of the background-only pseudo-data distributions just as for the data. The *expected* significance corresponds to the fraction of outcomes that yielded a cross section at least as large as the SM prediction for $WW + WZ$ production. The *observed* significance was determined by the fraction of outcomes above the measured cross section.

Table 4 gives the probability (p-value) and Gaussian significance (number of standard deviations for the corresponding Gaussian confidence level) for expected and observed outcomes corresponding to the measurements in Table 3. Again one can see the merit of the multivariate

classifier. While the observed significance using the dijet mass was found to be 3.3 standard deviation, the RF had an observed significance of 4.4 standard deviations.

Table 4: Expected and observed p-values obtained by comparing the measurement with background-only pseudo-experiments and the corresponding significance in number of standard deviations (s.d.) for a one-sided Gaussian integral.

Channel	Expected p-value (significance)	Observed p-value (significance)
$evq\bar{q}$ RF Output	6.8×10^{-3} (2.5 s.d.)	3.2×10^{-3} (2.7 s.d.)
$\mu\nu q\bar{q}$ RF Output	1.8×10^{-3} (2.9 s.d.)	5.2×10^{-5} (3.9 s.d.)
Combined RF Output	1.5×10^{-4} (3.6 s.d.)	5.4×10^{-6} (4.4 s.d.)
Combined Dijet Mass	1.7×10^{-3} (2.9 s.d.)	4.4×10^{-4} (3.3 s.d.)

9 Conclusions

Using semi-leptonic decay channels, we measured $\sigma(WW + WZ) = 20.2 \pm 4.5$ pb in proton-antiproton collisions $\sqrt{s} = 1.96$ TeV. This is consistent with the SM prediction of $\sigma(WW + WZ) = 16.1 \pm 0.9$ pb as well as with previous measurements of WW and WZ in the fully leptonic final states^{2,3}. The significance of the measurement is 4.4 standard deviations about the background, indicating the first direct evidence for $WW + WZ$ production with semi-leptonic decays at a hadron collider. Finally, this analysis demonstrates the ability to measure a small signal in a large background for a final state of direct relevance to searches for a low mass Higgs boson and provides a validation of the analytical methods used in searches for Higgs bosons at the Tevatron¹⁹.

Acknowledgments

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 and the Royal Society (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).

References

1. K. Hagiwara, S. Ishihara, R. Szalapski and D. Zeppenfeld, Phys. Rev. D **48** (1993).
2. D0 Collaboration: V. M. Abazov *et al.*, Phys. Rev. Lett. **94**, 151801 (2005); Phys. Rev. D **76**, 111104(R) (2007); Phys. Rev. Lett. **101**, 171803 (2008).
3. CDF Collaboration: D. Acosta *et al.*, Phys. Rev. Lett. **94**, 211801 (2005); A. Abulencia *et al.*, Phys. Rev. Lett. **98**, 161801 (2007); T. Aaltonen *et al.*, Phys. Rev. Lett. **100**, 201801 (2008).
4. W. Fisher, FERMILAB-TM-2386-E (2006).
5. B. Abbott *et al.* (D0 Collaboration), Nucl. Instrum. Methods Phys. Res. A **565**, 463 (2006).
6. V. M. Abazov *et al.* (D0 Collaboration), Phys. Lett. B **626**, 45 (2005).
7. G. C. Blazey *et al.*, arXiv:hep-ex/0005012 (2000). The seeded cone algorithm with radius 0.5 was used.

8. J. Smith, W. L. van Neerven, and J. A. M. Vermaseren, *Phys. Rev. Lett.* **50**, 1738 (1983).
9. T. Sjöstrand *et al.*, *Comput. Phys. Commun.* **135**, 238 (2001). Verison 6.3 was used.
10. M. L. Mangano *et al.*, *JHEP* **0307**, 001 (2003). Version 2.05 was used.
11. A. Pukhov *et al.*, arXiv:hep-ph/9908288 (2000).
12. R. Brun, F. Carminati, CERN Program Library Long Writeup W5013 (1993).
13. J. Alwall *et al.*, *Eur. Phys. C* **53**, 473 (2008).
14. S. Höche *et al.*, arXiv:hep-ph/0602031 (2006).
15. L. Breiman, *Machine Learning* **45**, 5 (2001).
16. I. Narsky, arXiv:physics/0507143 [physics.data-an] (2005).
17. 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.
18. V. M. Abazov *et al.* (D0 Collaboration), *Phys. Rev. D* **78**, 012005 (2008).
19. TEVNPH Working Group, for the CDF Collaboration and D0 Collaboration, arXiv:0804.3423 [hep-ex] (2008).