PROCEEDINGS OF SCIENCE

Measurement of WW and WZ production in the lepton plus heavy flavor jets final state at CDF

Sandra Leone[*†]
INFN Sezione di Pisa
E-mail: sandra.leone@pi.infn.it

We present the CDF measurement of the diboson WW and WZ production cross section in a final state consistent with leptonic W decay and jets originating from heavy flavor quarks, based on the full Tevatron Run II dataset. The analysis of the di–jet invariant mass spectrum allows the observation of 3.7 sigma evidence for the combined production processes of either WW or WZ bosons. The different heavy flavor decay pattern of the W and Z bosons and the analysis of the secondary–decay vertex properties allow to independently measure the WW and WZ production cross section in a hadronic final state. The measured cross sections are consistent with the standard model predictions and correspond to signal significances of 2.9 and 2.1 sigma for WW and WZ production, respectively.

38th International Conference on High Energy Physics
3-10 August 2016
Chicago, USA

[*]Speaker.
[†]On behalf of the CDF Collaboration

PoS(ICHEP2016)684

1. Introduction

The Tevatron Collider provided $p\bar{p}$ collisions at a center-of-mass energy of 1.96 TeV, until it ceased operations in September 2011.

Among other interesting standard model (SM) results, Tevatron experiments participated to the hunt for the Higgs boson. The analysis described here [1] was conceived and developed during the search for the Higgs boson at CDF [2].

The production of a pair of W or Z vector bosons is a process of primary interest at hadron colliders, and is theoretically well known. Measurements of the production of different vector-boson pairs probe the multiple gauge-boson couplings predicted by the SM, therefore a significant excess would open a window on new physics [3]. The study of diboson production provides a benchmark for analyses designed to study lower-cross-section processes sharing the same final states. For instance, in the Higgs boson associated production with a W boson HW, where the W decays leptonically and the Higgs decays to $b\bar{b}$, the final state is the same as the one obtained in the associated WZ production, where the Z boson decays to $b\bar{b}$.

In diboson production, when both vector bosons decay leptonically we have a leptonic final state, which provides a clean signature and has low background contamination, although the branching ratio is low. The semileptonic final state, when one vector boson decays hadronically, is much more challenging, both at the Tevatron and the LHC. It is affected by a large non-resonant background, mostly due to the production of a single vector boson in association with jets. In addition, the di-jet mass resolution is not adequate to separate the W from the Z contribution. For all these reasons, precise measurements of WZ in semi-leptonic final states are very difficult. The analysis described here exploits the WW and WZ heavy-flavor (HF) decays to disentangle the two processes [1]. The full CDF Tevatron Run II dataset, corresponding to 9.4 fb^{-1}, is used.

2. Event selection

The analysis strategy consists in selecting a high-acceptance lepton-plus-two-jets sample, using all the selection tools developed for the single-top quark and WH analyses at CDF [2] [4]. A Support Vector Machine (SVM) [5] discriminant is used to suppress the multi-jet (MJ) background. Secondary-vertex jet tagging is applied to enrich the sample in HF and reduce the $W+$ jets background. After selecting events with one or two secondary-vertex tags, we use a flavor-separator neural network (NN) to distinguish $WW \to \ell\nu + c\bar{s}$ versus $WZ \to \ell\nu + b\bar{b}(c\bar{c})$ [4].

In order to optimize the acceptance, all available trigger paths are used. Triggers designed for forward electrons and for the so called extended muons (based on missing transverse energy and jets information) are added to the standard central high energy electron and high energy muon triggers. The offline lepton selection requires exactly one charged lepton and a significant missing transverse energy. Two central jets are required, with corrected energy greater than 20 GeV, and finally the secondary vertices are required.

3. Background estimate

Main contributions to the background come from: W + HF which is the main source of irreducible background, W + light flavors (LF) mistakenly identified as HF, processes with a real

lepton and HF jets (EWK), and multi-jet (MJ) events giving a boson-like signature and false missing transverse energy. A combination of simulation-based and data-based prescriptions is used to estimate the background rates from the various sources.

As the W + jets event yield is predicted with large uncertainties, we use the data sample prior to applying b-jet-identification requirements (pretag control region) to estimate the total W + jets yield and validate the accuracy of the kinematic modeling of the W + jets simulation. Normalization of the W + jets simulation is determined separately in each lepton category. Templates for EWK and W + jets backgrounds are obtained from simulation while data-driven models are used for MJ background. The MJ and W + jets template normalizations are left free in the fit, while the EWK components are constrained within their uncertainties to the theoretical cross sections. The fit is performed before the application of the SVM selection requirement to leverage the high statistical power of the low SVM-output region to constrain the normalization of the MJ background. Fig. 1 shows the results of the fit to the SVM output distributions for the 4 lepton-classes together with the different selection thresholds.

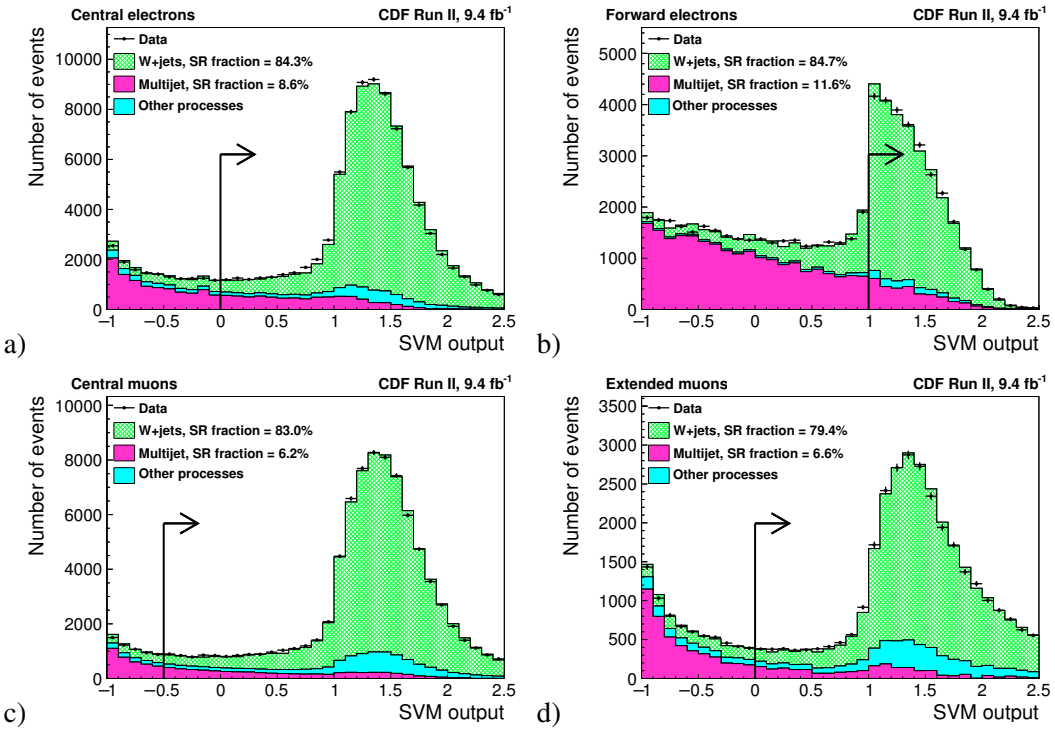

Figure 1: SVM output distributions for pretag control-region data (points), W+jets (green light hatched filling), MJ (dark magenta uniform filling), and other processes (light blue uniform filling) along with the MJ and W+jets event fractions after the signal region selection marked by the black arrows. The MJ and W+jets event yields are determined by fitting backgrounds to the data. The figure shows (a) central electrons, (b) forward electrons, (c) central muons, and (d) extended muons categories.

The production of W bosons in association with HF quarks represents the main background process in the single and double-tagged signal regions. A crucial part of this analysis is the extraction of correction factors to account for the W + HF yield difference in data with respect to the

prediction, which results from the interplay of matrix element generation, parton-shower matching scheme, and the strategy used to avoid HF-parton double-counting across samples. The heavy-flavor fraction correction is derived from a $W+1$ jet control region for both $W+b\bar{b}$ plus $W+c\bar{c}$, and $W+c$ processes simultaneously, studying the flavor-separator-NN output distribution. The final HF correction factors are $K_{bb;cc} = 1.24 \pm 0.25$ and $K_c = 1.0 \pm 0.3$.

4. Cross section extraction

After the HF-tag requirement the expected signal-to-background ratio is less than 0.03. Additional sensitivity comes from the study of the distribution of the invariant mass of the two jets in the event, m_{jj}, where signal is expected to cluster in a narrow resonance structure over a smooth non resonant background. Fig. 2 shows the best fit to the data di-jet invariant mass distribution for events with one b-tagged jet (on the left) and 2 b-tagged jets (on the right). Rate and shape systematic uncertainties of signal and background processes are treated as nuisance parameters. The lower part of the plots shows the signal excess.

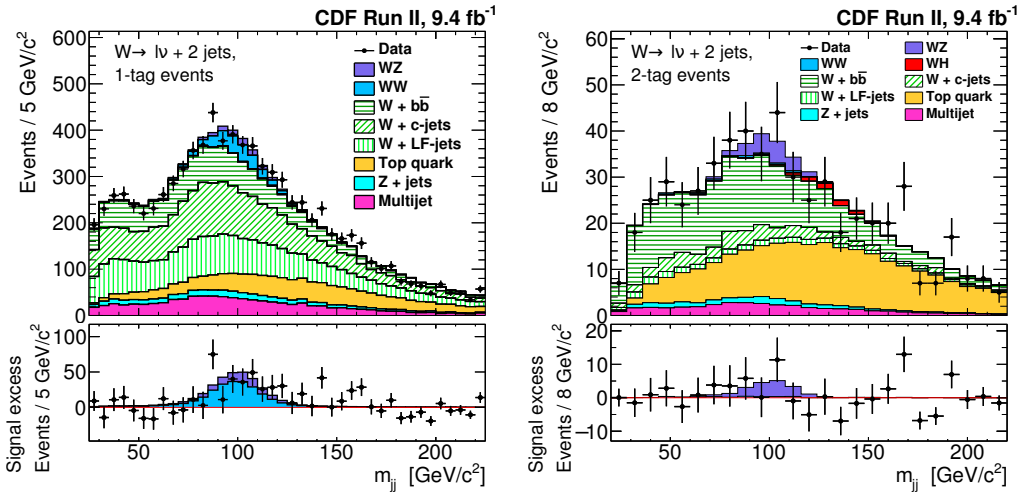

Figure 2: Distributions of m_{jj} for the one-tag candidates (left) and for the two-tag candidates (right) where events from all lepton categories are added together. The best fit to the data is shown.

For events with one tagged b-jet, in addition to the m_{jj} distribution, the flavor-separator NN distribution is used, to achieve b-to-c-jet separation. The flavor-separator NN uses the information from the secondary-decay vertex to assign an output score between -1 and 1, depending on the jet being more LF-like or b-like. Jets originating from c quarks are likely to obtain negative scores, clustering around the NN output value of -0.5.

Fig. 3 shows the m_{jj} distribution on the right for events with a high value of the flavor-separator NN (therefore enriched in b-jets), and on the left for events with a low flavor-separator value (b-suppressed region).

The m_{jj} distribution and the flavor-separator NN output (divided in six bins) are combined in a two-dimensional distribution. This improves the separation of the WW and WZ signals in the single-tag signal region.

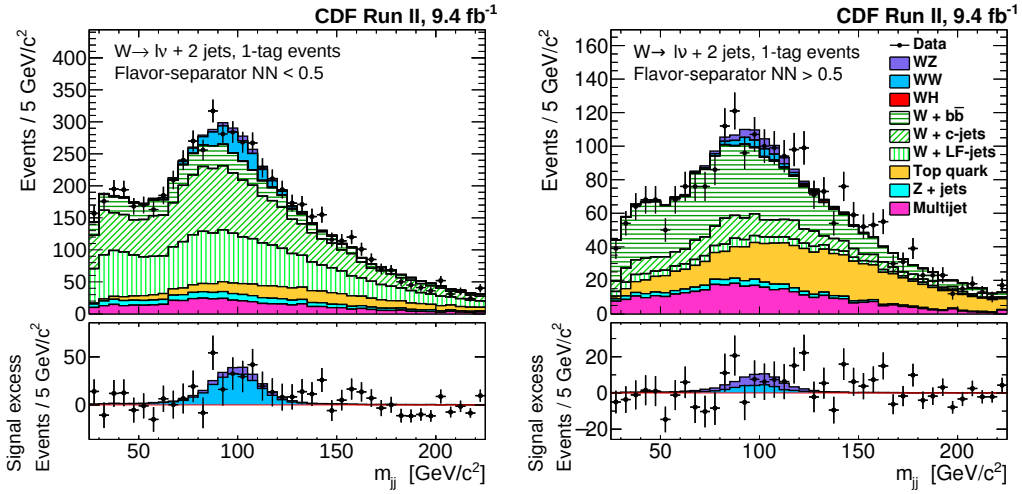

Figure 3: Distributions of m_{jj} for the one-tag candidates; for events with flavor-separator-NN output< 0.5 (left), and flavor-separator-NN output> 0.5 (right). Events from all lepton categories are added together. The best fit to the data is shown.

In total, eight regions are used for the signal extraction: four lepton subsamples (central electrons, central muons, forward electrons, extended muons) times two HF-tag prescriptions (one tag with flavor-separator NN and two tags). A likelihood function is built from the observed numbers of events in each bin, and the estimated signal and background distributions, assuming Poisson statistics. The prior probabilities of background and signal templates are included in the likelihood, together with all rate and shape systematic uncertainties, which are treated as nuisance parameters. The signal cross-section is obtained by marginalizing the posterior probability distribution over the nuisance parameters and studying the resulting posterior distribution for the signal yield.

For the combined $WW+WZ$ measurement, a one-dimensional uniform signal prior is used with relative rates for the WW and WZ processes given by the SM. Fig. 4 (left) shows the resulting posterior distribution of the $WW+WZ$ cross section together with the 68.3% and 95.5% Bayesian credibility intervals. A combined cross section of $\sigma_{WW+WZ} = 13.7 \pm 3.9$ pb is measured, in agreement with SM predictions.

To determine the significance of the signal, we perform a test by comparing the data with expectations under the null hypothesis of contributions from backgrounds only. The probability of a background fluctuation to produce a signal strength equal or greater than the observed signal strength corresponds to evidence for $WW+WZ$ production in the lepton-neutrino-HF final-state with a significance of 3.7 sigma. The result is compatible with the expected significance of 3.9 sigma, obtained from pseudoexperiments generated under the SM hypothesis.

The WW and WZ cross sections are also measured separately by exploiting the differing decay patterns of the W and Z bosons, which result in differing signal fractions in the one and two-tag signal regions and in different distributions of the flavour-separator NN. When measuring the WW and WZ cross sections separately, a two-dimensional uniform prior is used. The Bayesian analysis is repeated leaving the σ_{WW} and σ_{WZ} parameters free to vary. Fig. 4 (right) shows the resulting Bayesian posterior distribution, with integration contours at 68.3% and 95.5% credibility levels.

Integrating the two-dimensional-Bayesian posterior with respect to the σ_{WZ} or to the σ_{WW}

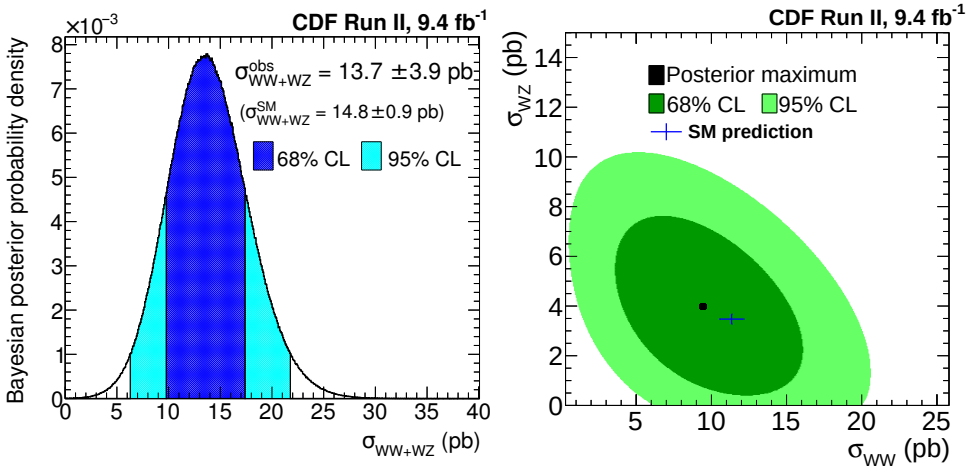

Figure 4: Left: Bayesian posterior distribution of the $WW + WZ$ cross section. Right: Two-dimensional posterior probability distribution in the σ_{WW} and σ_{WZ} plane. The cross shows the SM predictions and the corresponding uncertainties. In both plots the dark and light areas represent the smallest intervals enclosing 68.3% and 95.5% of the posterior integrals, respectively.

variable respectively, we find $\sigma_{WW} = 9.4 \pm 4.2$ pb and $\sigma_{WZ} = 3.7^{+2.5}_{-2.2}$ pb, in agreement with the SM predictions, and corresponding to the most precise measurement of the WZ-production cross section in a semileptonic final state to date.

The separate significances of the WW and WZ signals are evaluated as done for the combined signal. Simulated experiments are generated under the null hypothesis for both WW and WZ signals and significances of 2.9 sigma and 2.1 sigma respectively for WW and WZ are found.

5. Conclusion

We search for WW and WZ diboson production in a semi-leptonic final state enriched in HF jets using the full CDF Tevatron Run II data set. The di-jet invariant mass and a flavour-separator NN have been used to extract the total and separate WW and WZ signal cross sections. The total $WW + WZ$ cross section is measured with a precision of about 30%, comparable with other experiment measurements in semi-leptonic final-states. Separate WW and WZ cross sections are measured with a precision of 45% and 65% respectively, with WZ measurement being the most precise in this final state. Five years after the shutdown of the Tevatron, we are still digging out interesting results from the Tevatron data.

References

[1] T. Aaltonen *et al.* (CDF Collaboration) Phys. Rev. **D 94**, 032008 (2016).

[2] T. Aaltonen *et al.* (CDF Collaboration), Phys. Rev. Lett. **109**, 111804 (2012).

[3] C. Degrande *et al.* Ann. Phys. **335**, 21 (2013).

[4] T. Aaltonen *et al.* (CDF Collaboration), Phys. Rev. **D 82**, 112005 (2010).

[5] F. Sforza and V. Lippi, Nucl. Instrum. Methods Phys. Res., Sec. A **722**, 11 (2013).