IL NUOVO CIMENTO **45 C** (2022) 124
DOI 10.1393/ncc/i2022-22124-6

Communications: SIF Congress 2021

First observation of the electroweak production of a W^+W^- pair plus two jets in the leptonic channel in $\sqrt{s} = 13$ TeV pp collisions

B. S. Pinolini

Università and INFN, Milano-Bicocca - Milan, Italy

received 28 January 2022

Summary. — The first observation of the electroweak production of a W^+W^- pair in association with two jets, with both W bosons leptonically decaying, is reported. The data sample corresponds to an integrated luminosity of 138 fb^{-1} of proton-proton collisions at centre-of-mass energy of 13 TeV collected by the CMS detector. Events are selected by requiring two leptons (electrons or muons) and two jets with large pseudorapidity separation and high invariant mass. Events are categorised based on the flavour of final state leptons. A signal is observed (expected) with a significance of 5.6 (5.2) standard deviations and the measured fiducial cross section is 10.2 ± 2.0 fb. The results are in agreement with the Standard Model predictions of 9.1 ± 0.6 fb.

1. – Introduction

The Vector Boson Scattering (VBS) happens at the LHC when, in a p-p collision, two vector bosons radiate from the interacting quarks and scatter. The interaction can happen through a variety of mechanisms. Individually, the cross section of these electroweak (EW) processes rises quickly with energy and if considered separately, each of them would violate unitarity. However, their combined effect results in a finite cross section at all energies. Any Beyond-the-Standard-Model addition to the scattering process would alter this delicate balance resulting in changes to the cross section at high scattering centre of mass energy. Therefore, after the discovery of the Higgs boson it is still important to study the VBS, because the Standard Model Higgs boson might not be the only particle contributing to the VBS cross section regularisation. This paper reports the results of the analysis regarding the opposite-sign VBS in the fully leptonic channel [1] (fig. 1). The data used for this analysis have been collected at Run II at centre-of-mass energy of $\sqrt{s} = 13$ TeV with the CMS detector [2].

In the phase space considered, the main backgrounds are the $t\bar{t}$ production, the Drell-Yan and the production of two W bosons involving QCD vertices.

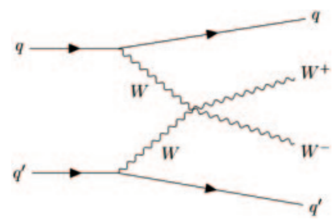

Fig. 1. – Examples of the Feynman diagram for the electroweak production of W^+W^-.

2. – Analysis strategy

The W^+W^- VBS final state is characterised by the presence of two jets from the incoming partons and two opposite-sign leptons and two neutrinos from the W decays.

The kinematic phase space is divided into a signal region (SR), defined such that the signal-to-background ratio is optimal, and two control regions (CRs), used to check the agreement between data and simulation and to constrain the normalisation of the major backgrounds. We define them to be complementary to the SR with respect to a single cut, so that the CRs are orthogonal to the SR, and in likelihood fit the same events are not considered twice. Each region is further categorised according to the charged lepton flavour: ee, $\mu\mu$ or $e\mu$.

To define the SR, a set of preselections has been established taking into account that in VBS events the jet pair has high m_{jj} and high $\Delta\eta_{jj}$, while backgrounds are more isotropic. Moreover, a key feature of the the $t\bar{t}$ process is the presence of two b-quarks in the final state, whose originating jets have peculiar characteristics that can be identified by means of a b-tagging algorithm. Requiring two not b-tagged jets in the SR enhances the $t\bar{t}$ suppression.

The SR is further split into two regions to optimise the signal significance. The categorisation is based on the centrality of the dilepton system with respect to the tagging jets, quantified by the Zeppenfeld variable $Z_{ll} = \frac{1}{2}|Z_{l1} + Z_{l2}|$, where $Z_l = \eta_{lep} - \frac{1}{2}(\eta_{j1} + \eta_{j2})$ [3]. The categories with $Z_{ll} < 1$ are enriched with signal and have less background contamination.

3. – Background estimation and signal extraction

Data driven estimates using CRs are employed for the normalisation of the $t\bar{t}$ and the Drell-Yan backgrounds. For such processes the normalisation is left as a free parameter to be constrained in the fit. In contrast, it is hard to define a CR enriched with events from QCD-induced WW background, so its normalisation is left to float freely but is not constrained by a particular category or region. Other minor backgrounds such as Non-Prompt, Higgs boson production, and multiboson production are estimated entirely from simulation.

3·1. *Deep Neural Network.* – In the $e\mu$ signal category, a feed-forward deep neural network (DNN) is used to separate the VBS signal from the $t\bar{t}$ and QCD-induced WW backgrounds. For optimisation purposes, two different models were built for the $Z_{ll} < 1$ and $Z_{ll} \geq 1$ regions. The two models share the same architecture and are fed with 9 discriminating variables, in table I. The signal extraction procedure is based on a binned maximum likelihood fit of the chosen discriminating variable distribution with signal

TABLE I. – *Set of variables used as inputs to the DNN.*

Variable	Description
m_{jj}	Invariant mass of the two VBS jets pair
$\Delta\eta_{jj}$	Pseudorapidity gap between the two VBS jets
$p_T^{j_1}$	p_T of the highest-p_T jet
$p_T^{j_2}$	p_T of the second highest-p_T jet
$p_{T,ll}$	p_T of the lepton pair
$\Delta\phi_{\ell\ell}$	Azimuthal angle between the two leptons
Z_{l1}	Zeppenfeld variable of the highest-p_T lepton
Z_{l2}	Zeppenfeld variable of the second highest-p_T lepton
$m_T^{W_1}$	Transverse mass of the ($p_T^{\ell_1}$, p_T^{miss}) system

and background templates, performed simultaneously in all the SR categories. CRs are included as single-bin templates where the number of events is fit to data. In the $e\mu$ signal categories, the binned DNN outputs are chosen as the discriminating variables. In the ee and $\mu\mu$ signal categories, different discriminating variables are chosen in different m_{jj} and $\Delta\eta_{jj}$ regions. For $m_{jj} > 500\,\text{GeV}$ and $\Delta\eta_{jj} > 3.5$, where the signal-to-background ratio is highest, m_{jj} is used. In the remaining phase space the number of events is the discriminating variable, and it is divided into three regions:

- $300 < m_{jj} < 500\,\text{GeV}$ and $2.5 < \Delta\eta_{jj} < 3.5$
- $m_{jj} > 500\,\text{GeV}$ and $2.5 < \Delta\eta_{jj} < 3.5$
- $300 < m_{jj} < 500\,\text{GeV}$ and $\Delta\eta_{jj} > 3.5$

Figures 2 and 3 show the observed post-fit distributions for the full data set in bins of DNN output for the $e\mu$ category and in bins of m_{jj} and $\Delta\eta_{jj}$ for ee and $\mu\mu$ categories, where histograms related to different processes are stacked on top of each other. The contributions from background (stacked histograms) and signal (superimposed violet histograms) processes are shown. Systematic uncertainties are plotted as dashed gray bands.

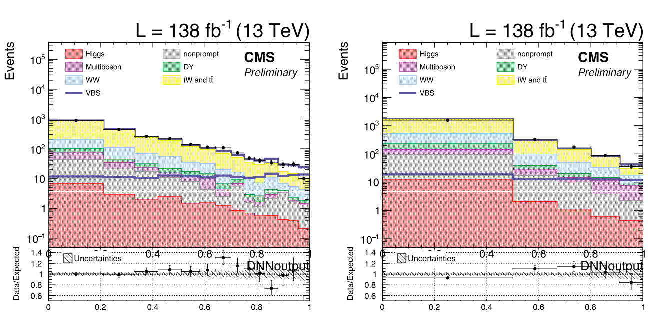

Fig. 2. – Post-fit DNN output distributions in the $e\mu$ SR for $Z_{ll} < 1$ (left) and $Z_{ll} \geq 1$ (right).

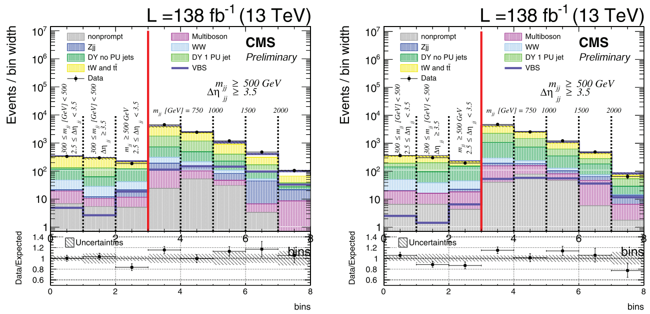

Fig. 3. – Post-fit m_{jj} distribution for the SRs in same-flavour categories (ee and $\mu\mu$ combined) for $Z_{ll} < 1$ (left) and $Z_{ll} \geq 1$ (right).

4. – Results

Fit results are combined in all categories. Expected results are predicted using a toy Asimov dataset [4]. We quantify the statistical significance of the signal by means of a p-value, converted to an equivalent Gaussian significance, which corresponds to the probability of the background-only hypothesis under the asymptotic approximation [4]. The observed (expected) significance for the signal is 5.6 (5.2) standard deviations.

The EW WW production cross section is measured in a fiducial volume defined by the same requirements of the preselection region transposed to generator-level variables. The measured fiducial cross section is $10.2 \pm 2.0\,\text{fb}$, while the LO theoretical prediction is $9.1 \pm 0.6\,\text{fb}$.

5. – Conclusion

The first observation of the opposite-sign VBS in the fully leptonic channel is reported.

The measured signal corresponds to an observed (expected) significance of 5.6 (5.2) standard deviations with respect to the only-background hypothesis. The EW WW measured production cross section is $10.2 \pm 2.0\,\text{fb}$. The result is dominated by the statistical uncertainty, so the analysis is expected to benefit from the larger Run III dataset.

REFERENCES

[1] CMS Collaboration, *First observation of the electroweak production of a leptonically decaying W^+W^- pair in association with two jets in $\sqrt{s} = 13$ TeV pp collisions*, CMS-PAS-SMP-21-001.

[2] CMS Collaboration, *JINST*, **3** (2008) S08004.

[3] Rainwater D. L., Szalapski R. and Zeppenfeld D., *Phys. Rev. D*, **54** (1996) 6680.

[4] Cowan Glen, Cranmer Kyle, Gross Eilam and Vitells Ofer, *Asymptotic formulae for likelihood-based tests of new physics*, arXiv:1007.1727.