IL NUOVO CIMENTO **46 C** (2023) 85
DOI 10.1393/ncc/i2023-23085-x

COMMUNICATIONS: SIF Congress 2022

Observation of vector boson scattering W^+W^- pair production in proton-proton collisions at a center-of-mass energy of 13 TeV with the CMS detector

A. C. M. BULLA(1)(2) on behalf of the CMS COLLABORATION

(1) *INFN, Sezione di Milano-Bicocca - Milano, Italy*
(2) *Dipartimento di Fisica, Università di Milano-Bicocca - Milano, Italy*

received 31 January 2023

Summary. — The observation of vector boson scattering (VBS) W^+W^- pair production, with the two W bosons decaying leptonically, has been reported by the CMS Collaboration at the CERN LHC. The data sample analyzed corresponds to an integrated luminosity of 138 fb^{-1} of proton-proton collisions at $\sqrt{s} = 13$ TeV collected by the CMS detector. Events are selected by requiring exactly two opposite-sign leptons (electrons or muons) and two jets with large pseudorapidity separation and high dijet invariant mass. Machine learning techniques are employed to deal with the irreducible background from the QCD-induced production of W bosons and the large background from the production of top quarks pair. The VBS W^+W^- signal is observed with a significance of 5.6 standard deviations (5.2 expected) with respect to the background-only hypothesis. The measured fiducial cross section is 10.2 ± 2.0 fb, in agreement with the standard model prediction of 9.1 ± 0.6 fb.

1. – Introduction

Electroweak scattering of the type $VV' \rightarrow VV'$, with V and V$'$ being W, Z or γ vector bosons, is a crucial tool for investigating the mechanism of electroweak symmetry breaking. By adding additional terms that cancel divergences, the existence of a 125 GeV Higgs boson prevents the violation of unitarity in such vector boson scattering (VBS) processes [1]. Therefore, the CERN LHC's accurate measurement of VBS cross sections may explore the Higgs sector's nature and search for effects beyond the domain of the standard model (BSM).

In order to make the first observation of a VBS process, events with two W bosons with the same electric charge ($W^{\pm}W^{\pm}$) in a fully leptonic final state have been extensively studied by both ATLAS [2] and CMS [3] Collaborations. Here, it is presented the first CMS observation of a purely electroweak production of a pair of opposite-sign ($W^{\pm}W^{\mp}$) bosons which decay in a fully leptonic final state. The data sample analyzed corresponds

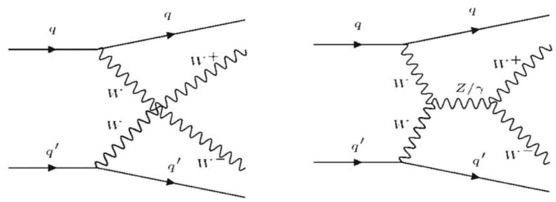

Fig. 1. – Examples of Feynman diagrams α_{EW}^6 for the EW production of $W^{\pm}W^{\mp}$ at LO.

to an integrated luminosity of $138\,\text{fb}^{-1}$ of proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ collected by the CMS detector. Figure 1 shows some examples of Feynman's diagrams at the order α_{EW}^6 that contribute to the signal.

VBS events are characterized by a strong signature that includes, in addition to two vector bosons, two high transverse momenta (p_T) "tagging jets" emitted in opposite directions inside the detector. These jets tag the event and typically have a large separation in pseudorapidity ($|\Delta\eta_{jj}|$) and a high invariant mass (m_{jj}). In addition, the two leptons from the decay of the W bosons (both electrons or muons) are emitted centrally with respect to the tagging jets. This feature is quantified by the Zeppenfeld variable $Z_{ll} = \frac{1}{2}(Z_{l1} + Z_{l2})$ where $Z_l = \eta_l - \frac{1}{2}(\eta_{j1} - \eta_{j2})$.

2. – Events selection

Data is selected by requiring events to pass some HLT selections, which triggers the presence of both one and two isolated leptons with minimum energy thresholds that vary between 8 and 23 GeV. Candidate signal events are then pre-selected according to the following requirements:

- Two opposite-sign leptons, with invariant mass $m_{ll} > 50\,\text{GeV}$ and a transverse momentum $p_T^{ll} > 30\,\text{GeV}$. The most energetic lepton is required to have $p_T > 25\,\text{GeV}$, while the threshold for the second one is $p_T > 13\,\text{GeV}$. Events with an additional lepton with $p_T > 10\,\text{GeV}$ are rejected to suppress multiboson backgrounds.
- Missing transverse momentum $p_T^{miss} > 20\,\text{GeV}$.
- At least two jets with $p_T > 30\,\text{GeV}$, an invariant mass $m_{jj} > 300\,\text{GeV}$, and a pseudorapidity gap $|\Delta\eta_{jj}| > 2.5$.

The final state with different lepton flavors ($e\mu$) is characterized by a better signal-to-noise ratio than the same flavor (SF) channel (ee,$\mu\mu$). This is due to the Drell-Yan (DY) processes, which heavily contaminate the SF channels, whereas in the $e\mu$ category only a residual DY contribution (from $\tau\tau$ production) survives. Other backgrounds for the analysis are the $t\bar{t}$ production, where a t quark decays into a b quark and a W boson, the QCD production of a W pair, and the non-prompt events. Non-prompt events are the only reducible backgrounds for the analysis, since they arise from inefficiencies of the detector.

3. – Analysis strategy

The phase space is divided into different regions through cuts on the main kinematic variables. The signal regions (SRs) are optimized to have the best signal-to-background ratio, while the control regions (CRs) are used to check the agreement between data and simulation and to constrain the normalization of the main backgrounds with a data-driven method.

SRs are defined requiring that no b-jets (according to loose working point of the DeepJet algorithm [4]) are present in the event. In this way, it is possible to reduce the contribution of the $t\bar{t}$ background in the analysis. Furthermore, in the SF category, additional $m_{ll} > 120\,\mathrm{GeV}$ and $p_T^{miss} > 60\,\mathrm{GeV}$ requirements are needed, while in the DF final state, the transverse mass $m_T = \sqrt{2p_T^{ll}p_T^{miss}(1-\cos\Delta\phi)}$ is required to be above 60 GeV. Every SR is then split into two regions to optimize the signal significance, according to the $Z_{ll} \gtrless 1$ value. The categories with $Z_{ll} < 1$ are enriched with signal and have less background contamination.

DY CR for the SF category is defined requiring the invariant mass of the two leptons to be close to the Z boson mass peak, or $|m_{ll} - m_Z| < 15\,\mathrm{GeV}$ and zero b-jets in the event. For the DF final state, m_T cut is reversed with respect to the SR to preserve orthogonality between different regions and a $50\,\mathrm{GeV} < m_{ll} < 80\,\mathrm{GeV}$ window is selected.

Finally, the $t\bar{t}$ CRs are defined by inverting the b veto condition and thus requiring the presence of at least one b-jet with $p_T > 20\,\mathrm{GeV}$ in the final state.

4. – Signal extraction

The signal extraction procedure is carried out through a binned maximum likelihood fit on the most discriminating variable distribution with signal and background templates, mainly generated with MadGraph5_aMC@NLO interfaced with PYTHIA. The fit is performed simultaneously in all the SRs. The normalization of $t\bar{t}$ and DY backgrounds is estimated including the CRs as single-bin templates, where the number of events is fit to the data. Systematics uncertainties are considered as nuisance parameters in the maximum likelihood fit.

In the $e\mu$ category SR, a feed-forward deep neural network (DNN) is trained to separate the signal from $t\bar{t}$ and QCD WW backgrounds. The DNN is trained with 9 variables including jet and lepton kinematic and angular properties. For optimization purposes,

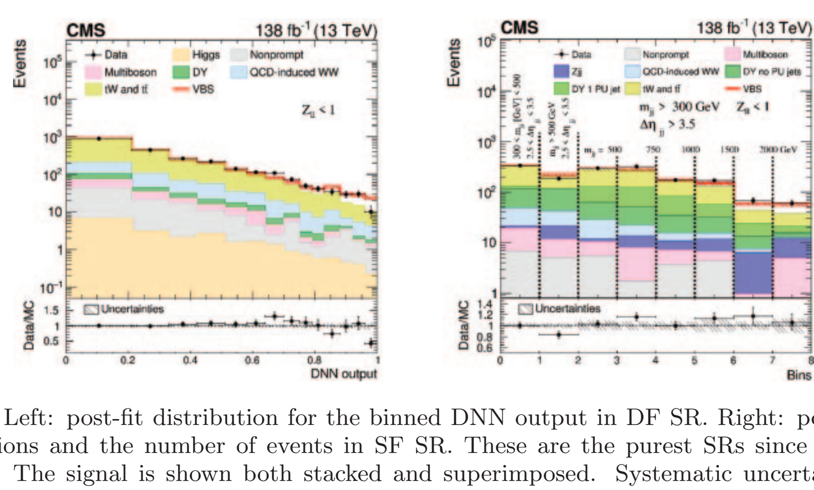

Fig. 2. – Left: post-fit distribution for the binned DNN output in DF SR. Right: post-fit m_{jj} distributions and the number of events in SF SR. These are the purest SRs since they have $Z_{ll} < 1$. The signal is shown both stacked and superimposed. Systematic uncertainties are plotted as grey bands in the ratio plot.

TABLE I. – *Cuts applied to define the fiducial volume closed to the signal region.*

Objects	Requirements
Leptons	$e\mu$,ee,$\mu\mu$ (not from τ decay), opposite charge $p_T^{dressed\ l} = p_T^l + \sum_i p_T^{\gamma_i}$ if $\Delta R(l,\gamma_i) < 0.1$ $p_T^{l_1} > 25\,\text{GeV}$, $p_T^{l_2} > 13\,\text{GeV}$, $p_T^{l_3} < 10\,\text{GeV}$, $\lvert\eta\rvert < 2.5$, $p_T^{ll} > 30\,\text{GeV}$, $m_{ll} > 50\,\text{GeV}$ $p_T^{miss} > 20\,\text{GeV}$
Jets	At least 2 jets, no b-jets, $p_T^j < 30$, $\Delta R(j,l) > 0.4$, $\lvert\eta\rvert < 4.7$, $m_{jj} > 300\,\text{GeV}$, $\lvert\Delta\eta_{jj}\rvert > 2.5$

two identical DNNs are actually trained in the low $Z_{ll} < 1$ and high $Z_{ll} \geq 1$ regions. The binned DNN outputs are chosen as discriminating variables for the fit.

In the SF category SR, events are divided into four $m_{jj} - |\Delta\eta_{jj}|$ regions. The discriminating variables used in the fit are chosen to optimize the signal significance and depend on the region. For $m_{jj} > 500\,\text{GeV}$ and $|\Delta\eta_{jj}| > 3.5$, the chosen variable is the binned m_{jj} distribution. In the other cases, a single bin number of events is chosen. All these regions are then placed side by side for a better graphic display. The fitting procedure is applied for all these binned regions, in each $Z_{ll} < 1$ and $Z_{ll} \geq 1$ regions and for each ee, $\mu\mu$ final state. Post-fit distributions are shown in fig. 2.

5. – Results and conclusions

Expected results are obtained through an Asimov dataset [5]. The observed (expected) significance for the signal is 5.6 (5.2) standard deviations with respect to the background-only hypothesis.

The cross-section is measured in two fiducial volumes; in the more inclusive one, the cross-section is $99 \pm 20\,\text{fb}$ (89 ± 5 fb expected), whereas in the one closer to the reconstructed signal region (cuts reported in table I) the measured cross-section is $10.2 \pm 2.0\,\text{fb}$ ($9.1 \pm 0.6\,\text{fb}$ expected) confirming the expectations of the SM.

REFERENCES

[1] BALLESTRERO A. *et al.*, *Rev. Phys.*, **3** (2018) 44.
[2] THE ATLAS COLLABORATION, *Phys. Rev. Lett.*, **123** (2019) 161801.
[3] THE CMS COLLABORATION, *Phys. Rev. Lett.*, **120** (2018) 081801.
[4] BOLS E., KIESELER J., VERZETTI M., STOYE M. and STAKIA A., *JINST*, **15** (2020) P12012.
[5] COWAN G., CRANMER K., GROSS E. and VITELLS O., *Eur. Phys. J. C*, **71** (2011) 1554.