Measurements of the hadronic activity and the electroweak production in events with a Z boson and two jets in proton-proton collisions with the CMS experiment

Paolo Azzurri, for the CMS Collaboration

INFN Pisa, Largo Pontecorvo 3, 56127 Pisa, Italy

Abstract

The observation of the electroweak production of a Z boson with two jets in pp collisions at \( \sqrt{s} = 8 \text{ TeV} \) with the CMS experiment at the CERN LHC is presented, based on a data sample with an integrated luminosity of 19.7 fb\(^{-1}\). The cross section measurement, combining the muon and electron channels, is in agreement with the theoretical expectations. Radiation patterns of selected Z plus two jets events, and the hadronic activity in the rapidity interval between the jets are also measured. These results are of substantial importance in the more general study of vector boson fusion processes, of relevance for Higgs boson searches and for measurements of electroweak gauge couplings and vector boson scattering.

1. Introduction

In proton collisions at the LHC Vector Boson Fusion (VBF) happens when a valence quark of each one of the colliding protons radiates a W or Z boson that subsequently interact or “fuse”. For both valence quark radiating the weak bosons a t-channel four-momentum with \( Q^2 \sim m_Z^2, m_W^2 \) is exchanged. In this way the two valence quarks are typically scattered away from the beam line and inside the detector acceptance, where they can be revealed as hadronic jets. The distinctive signature of VBF is therefore the presence of these two energetic hadronic jets (tagging jets), roughly in the forward and backward direction with respect to the proton beam line.

The VBF production has a great prominence at the LHC for its importance for the measurements of the Higgs sector couplings [1, 2]. The study of the VBF production of Z or W bosons is therefore an important benchmark to cross-check and validate Higgs VBF measurements [3], but serves further as a probe of triple-gauge-boson couplings [4], for searches for physics beyond the standard model [5, 6], and as a precursor to the measurement of elastic vector boson pair scattering.

On the other hand the VBF production of Z or W bosons has some intriguing differences with respect to the Higgs VBF productions. When focusing on VBF Z/W production, the observed final state is composed of a pair of fermions (f), either quarks or leptons, from the Z/W decay, associated with a pair of quarks (qq) from the VBF production mechanism; but in this context there is a large number of non-VBF diagrams that lead to identical \( ffqq \) final states that can’t be neglected [7].

Considering only the classes of diagrams with purely electroweak (EW) interactions, (like the VBF one), shown in Figure 1 and no QCD interactions, the additional diagrams have strong negative interferences with the VBF productions. These large negative interference effects are in fact related to well-known non-abelian gauge cancellations that preserve the scattering unitarity and the electroweak model theory renormalizability [8]. This situation makes the VBF Z/W channel more complicated but also more interesting.

Another main scope of selecting “VBF-like” Z plus two jets events, is to study the event hadronization properties connected with the peculiar VBF production color structure. In VBF processes and more in general also for the contributing electroweak processes with identical final states, there is no t-channel color exchange. This leads to the expectation of a “rapidity gap” of
suppressed hadronic activity between the two tagging jets that is a very peculiar feature, in particular in the case of a large rapidity separation between the two tagging jets [9][10][11]. Measurements of the additional hadronic activity in the rapidity gap provides a precious validation of the Monte Carlo models simulations and benchmark results for the use of rapidity gap observables, like jet vetoes, in independent VBF event productions (e.g. for Higgs selections).

At the LHC, the EW Z plus two jets process was first measured by the ATLAS experiment using pp collisions at √s = 7 TeV [12], and more recently by both the ATLAS and CMS experiments with √s = 8 TeV data [13][14]. The results presented here will focus on the most recent CMS results using pp collisions collected at √s = 8 TeV and corresponding to an integrated luminosity of 19.7 fb⁻¹ [14]. Different methods have been used to confirm the presence of the signal: two multivariate analyses methods (A) and (B) as developed for the 7 TeV analysis [12], and new method (C) with a data-driven model of the main Drell-Yan (DY) background.

2. Event reconstruction and simulation

A detailed description of the CMS detector can be found in Ref. [15]. The central feature of the CMS apparatus is a superconducting solenoid of 6 m internal diameter providing a field of 3.8 T. Within the field volume are a silicon pixel and strip tracker, a crystal electromagnetic calorimeter (ECAL), and a brass/scintillator hadron calorimeter (HCAL) providing coverage for pseudorapidities |η| < 3. The forward calorimeter modules extend the coverage of hadronic jets up to |η| < 5.

Electrons are reconstructed from clusters of energy depositions in the ECAL that match tracks extrapolated from the silicon tracker [16]. Muons are reconstructed by fitting trajectories based on hits in the silicon tracker and in the outer muon system [17].

Jets are clustered using the anti-kt algorithm [18] with a distance parameter of 0.5. Two different types of jets are used in the analysis: jet-plus-track (JPT) and particle-flow (PF) jets. The JPT jets are reconstructed calorimeter jets whose energy response and resolution are improved by incorporating tracking information according to the JPT algorithm [19]. The CMS particle flow algorithm [20][21] combines the information from all relevant CMS sub-detectors to identify and reconstruct particle candidates in the event, and PF jets are reconstructed clustering particles identified by the particle flow algorithm.

The signal is defined as the pure EW production of ℓℓjj final states in the kinematic region defined by dilepton mass M_ℓℓ > 50 GeV, parton transverse momentum p_Tj > 25 GeV, parton pseudorapidity |η| < 5, diparton mass M_jj > 120 GeV.

Signal events are simulated at leading order (LO) using the MADGRAPH Monte Carlo (MC) generator [22][23], interfaced to PYTHIA (v6.4.26) [24] for parton showering (PS) and hadronisation. The underlying event is modeled with the so-called ZZ* tune [25]. The predicted signal cross section is σ_{LO}(EW ℓℓjj) = 208 ± 16 fb, for a single lepton flavor.

Background DY ℓℓ events are also generated with MADGRAPH using a LO matrix element (ME) calculation that includes up to four partons generated from quantum chromodynamics (QCD) interactions, and interfaced to PYTHIA for PS. The ME-PS matching is performed following the ktMLM prescription [26][27].

Possible LO interference effects between the EW signal and DY processes have been evaluated making use of MADGRAPH, comparing the differential distributions of samples with (i) pure signal, (ii) pure DY plus two partons, and (iii) both signal and DY together.

Other residual backgrounds from ditop (t̅t̅) and diboson (VV) productions are generated with MAD-
relative lepton isolation is defined as CMS algorithms for electrons \[16\] and muons \[17\]. A ≈ is incorporated by simulating additional interactions (both in-time and out-of-time with the collision) with a multiplicity that matches the one observed in data. The average number of pileup events in the 8 TeV data is estimated as ≈21 interactions per bunch crossing.

3. Event selection and Drell-Yan background model

Opposite sign lepton pairs are selected with validated CMS algorithms for electrons \[16\] and muons \[17\]. A relative lepton isolation is defined as \(I = \frac{\sum p_T}{p_T}\), where the sum includes all reconstructed PF objects inside a cone of \(\Delta R = \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2} < 0.4\) around the lepton. Each lepton is required to have a transverse momentum in excess of 20 GeV, and a relative isolation \(I\) smaller than 0.10 and 0.12 for electrons and muons respectively. The invariant mass \(M_\ell\ell\) of selected same flavor leptons is finally required to satisfy \(|M_Z - M_\ell\ell| < 15\) GeV, where \(M_Z\) is the nominal Z-boson mass.

Analyses (A) and (B) make respectively use of PF and JPT reconstructed jets in the selected events, and both rely on MC simulations to predict the main DY plus jets background. Analysis (C) makes use of PF jets, as analysis (A), but uses a model of DY plus jets derived from photon plus jets data events \[14\], where the requirement \(p_T(Z/\gamma) > 50\) GeV is applied to ensure a good photon purity. It is further verified that the data-driven method works correctly with simulated events.

For the rapidity gap and signal measurements events are required to have two PF or JPT jets within \(|\eta| \leq 4.7\), with \(p_T > 50, 30\) GeV and with a dijet invariant mass \(M_\eta > 200\) GeV for the \(p_T\)-leading and subleading jets.

4. Event jet radiation patterns

The selected Z plus jets event “radiation patterns” are studied, and for this, according to the prescriptions in Ref. 31, only PF jets with \(p_T > 40\) GeV are considered. The investigated observables are: (i) the number of jets, \(N_j\), (ii) the total scalar sum of the transverse momenta of jets reconstructed within \(|\eta| < 4.7, H_T\), (iii) \(\Delta \eta_j\) between the two jets which span the largest pseudorapidity gap in the event, and (iv) the cosine of the azimuthal angle difference, \(\cos \Delta \phi_j\), for the two jets with criterion (iii).

Figure 2 shows the average number of jets and the average \(\cos \Delta \phi_j\) as a function of the total \(H_T\) and \(\Delta \eta_j\). The plots indicate that the MADGRAPH + PYTHIA (ME+PS) predictions are in good agreement with the data, even in the regions of largest \(H_T\) and \(\Delta \eta_j\).

5. Hadronic activity in the dijet rapidity gap

The rapidity gap activity is studied in events with a Z and two VBF “tagging” PF jets with \(p_T > 50, 30\) GeV, making use of charged tracks only, and with additional PF jets in a region of higher signal purity.

For the first study we use tracks associated with the main event primary vertex (PV), defined as the PV with the largest \(\sum p_T^2\) of the tracks used to fit the vertex, and exclude tracks associated with the two leptons or with the tagging jets. A collection of “soft track-jets” is defined by clustering the selected tracks using the anti-\(k_T\) algorithm with \(R = 0.5\). The use of track jets repre-
$\Delta \eta_3$ is shown in Fig. 3 and good agreement is observed between data and the simulation in all ranges.

The rapidity gap interval has also been studied using PF jets with $p_T > 15$ GeV, in the $M_3 > 1250$ GeV region with higher signal purity, to examine possible evidence of the color exchange suppression for the EW signal component. Results are shown in Fig. 4 for the additional jet multiplicity in the dijet rapidity gap, where the data, in agreement with the MC expectations, indicates the presence of the EW signal with a suppressed third jet emission compared to the background-only prediction.

6. Signal measurements

The three analyses make use of multivariate boosted decision tree (BDT) discriminators to acquire the best expected separation of signal and background sources. The BDTs make mostly use of the dijet and $Z$ boson kinematics, with the addition of a quark/gluon ($q/g$) jet discriminator $f_{C2}$ as both signal jets are originated from quarks while the jets in background events are more probably initiated by gluons emitted from QCD processes.

BDT output distributions for the different analyses are shown in Fig. 5, where a good overall agreement is observed between the data and the MC predictions.

To measure the signal cross-section, each analysis builds a binned likelihood based on the BDT output distributions that is used to fit strength modifiers for both the signal and the main DY background. Nuisance parameters are added to modify the expected rates and shapes according to the estimate of the systematic uncertainties affecting each analysis. Possible interference effects between the signal and the DY background processes is taken into account in the fit, with a parameterisation derived from MADGRAPH, as a function of the $M_3$ variable. The statistical methodology used follows what used in CMS Higgs analysis $^{34}$ using asymptotic formulas $^{35}$.

A summary of fitted signal strengths is reported in Table 1, together with the breakdown of all relevant uncertainties. The signal strength obtained from the combined fit of two channels in analysis A is $\sigma = 0.84 \pm 0.07$ (stat) $\pm 0.19$ (syst) corresponding to a measured signal cross section

$$\sigma(\text{EW } \ell jj) = 174 \pm 15 \text{(stat)} \pm 40 \text{(syst)fb},$$

with the background-only hypothesis excluded with a significance greater than 5$\sigma$.

References

[1] D. Zeppenfeld, R. Kannum, N. Nikitenko, E. Richter-Was, Measuring Higgs boson couplings at the CERN LHC, Phys. Rev. D 62 (2000) 013009. arXiv:hep-ph/0002036 doi:10.1103/PhysRevD.62.013009

[2] M. Duhrssen, S. Heinemeyer, H. Logan, D. Rainwater, G. Weiglein, D. Zeppenfeld, Extracting Higgs boson couplings from CERN LHC data, Phys. Rev. D 70 (2004) 113009. arXiv:hep-ph/0406323 doi:10.1103/PhysRevD.70.113009

[3] V. A. Khoe, M. G. Ryskin, W. J. Stirling, P. H. Williams, A $Z$-monitor to calibrate Higgs production via vector boson fusion with rapidity gaps at the LHC, Eur. Phys. J. C 26 (2003) 429. arXiv:hep-ph/0207365 doi:10.1140/epjc/s2002-01069-2
Figure 5: Distributions of the BDT outputs for (left) dielectron events in analysis A, (center) dimuon events in analysis B, and all dilepton events in analysis C, with a dijet invariant mass $M_{jj} > 750$ GeV. The bottom panels show the differences between data or the expected signal contribution, with respect to the expected background (BG).

Table 1: Measured signal strength $\mu$ in the different analyses and channels, with statistical and relevant systematic uncertainties.

|                      | Analysis A | Analysis B | Analysis C |
|----------------------|------------|------------|------------|
|                      | ee         | $\mu$      | ee + $\mu$ |
| Luminosity           | 0.03       | 0.03       | 0.03       |
| Lepton trigger/selection | 0.04       | 0.04       | 0.04       |
| Jet Energy Scale     | 0.06       | 0.05       | 0.05       |
| Jet Energy Resolution| 0.02       | 0.02       | 0.02       |
| DY background         | 0.07       | 0.05       | 0.07       |
| Signal acceptance    | 0.03       | 0.04       | 0.04       |
| DY/EW interference    | 0.14       | 0.14       | 0.14       |
| Systematic uncertainty| 0.19       | 0.18       | 0.19       |
| Statistical uncertainty| 0.11       | 0.10       | 0.07       |
| $\mu = \sigma/\sigma_{th}$ | 0.82       | 0.86       | 0.84       |
