Towards a Measurement of the Inclusive $W \rightarrow e\nu$ and $\gamma^*/Z \rightarrow e^+e^-$ Cross Sections in pp Collisions at $\sqrt{s} = 10$ TeV.

The CMS Collaboration

Abstract

We investigate methods for an early measurement of the inclusive $W$ and $\gamma^*/Z$ production cross sections in the electron decay channel, assuming 10 pb$^{-1}$ of pp collision data at $\sqrt{s} = 10$ TeV. A simple and robust electron selection is used. This selection is intended to be tolerant of the anticipated imperfections in calibration and alignment of the CMS detector in early data. Data driven methods have been used to measure the selection efficiencies, to tune the selection cuts, and to subtract the background in the $W$ case. The electroweak boson signal yields can be measured with a statistical precision of less than 2.0% in each case, while the systematic biases in our methods have been estimated from simulation studies to be 4.0% and 2.4% for $W$ and $Z$, respectively. Initial precision in the integrated luminosity measurement of 10% is expected to dominate the precision of the cross section measurements.
1 Overview of the measurement

The signature of high transverse momentum leptons from W and Z decays is very distinctive in the environment of hadron collisions. As such, the decays of W and Z bosons into leptons provide a clean experimental measurement of their production rate. Such events are expected to play a major role in the physics commissioning of CMS and in the understanding of leptons with the first data. The measurement of W and Z production cross sections is likely to be one of the first physics measurements to be made at LHC.

The study described demonstrates the methods that may be used to measure the inclusive W and Z production cross-sections in the electron decay channel. Data driven methods have been used to measure the selection efficiencies, to tune the selection cuts, and to subtract the background in the W case. Sources of systematic error have been investigated in some detail and the dominant sources identified.

The $W \rightarrow e \nu$ cross section can be calculated using the following formula (a similar formula can be used for the $\gamma^*/Z \rightarrow e^+e^-$ cross section):

$$\sigma_W \times BR(W \rightarrow e\nu) = \frac{N_{\text{pass}}^W - N_{\text{bkgd}}^W}{A_W \times \epsilon_W \times \int L dt}$$

where $N_{\text{pass}}$ is the number of candidates selected from the data, $N_{\text{bkgd}}$ represents the expected background events and $A$ is the acceptance defined as the fraction of these decays satisfying the geometric constraints of the detector and the kinematic constraints of the imposed selection criteria. The $\epsilon_W$ is the selection efficiency for W decays falling within the acceptances and $\int L dt$ is the integrated luminosity.

2 Data samples

Monte Carlo data samples have been used to perform the study described. They were produced with PYTHIA [1]. Events were fully simulated using Geant4 [2], digitized without pile-up, and reconstructed with CMS reconstruction software. The following processes have been included as sources of background: di-jets and $\gamma$+jets in different $p_T$ bins, $Z \rightarrow \tau\tau$, $t\bar{t}$, and $W \rightarrow \tau\nu$.

To allow the simulation of electron background with an equivalent integrated luminosity of 10pb$^{-1}$ the di-jets were preselected at generation time to be enriched in events likely to pass electromagnetic triggers.

3 Selection of W and Z samples

The $W \rightarrow e\nu$ and $\gamma^*/Z \rightarrow e^+e^-$ samples are selected from events that pass a single electron High Level Trigger [3]. We require one (for $W \rightarrow e\nu$) or two (for $\gamma^*/Z \rightarrow e^+e^-$ ) high-$p_T$ electrons formed from the association of a high $E_T$ supercluster in the PbWO$_4$ crystal electromagnetic calorimeter (ECAL) and a high $p_T$ track [4]. An ECAL supercluster gathers the energy deposited in a region around the main energy cluster in an attempt to recover most of the energy of Bremsstrahlung photons emitted along the electron trajectory. The momentum of the electron candidate is fitted along its trajectory using a Gaussian-Sum Filter algorithm (GSP) [5] dealing with the possible emission of hard Bremsstrahlung photons in the scattering layers of the Tracker. The electron(s) should fall within the ECAL fiducial region ($|\eta| < 2.5$, excluding the Barrel-Endcap transition region). The ECAL supercluster(s) should have a transverse energy $E_T > 20.0$ GeV for the Z selection and $E_T > 30.0$ GeV for the W selection.
Since the electron(s) from the W(Z) decay is(are) isolated, we require low particle (charged and/or neutral) activity in a cone around the electron candidate. The isolation variables consisted of sums of transverse components of energy deposits in ECAL and HCAL and track $p_T$ within regions of $\Delta R < 0.4$. The regions were centred on the supercluster position for the calorimetric isolation variables, and on the track direction at the vertex for the track isolation. In all three cases the possible track or energy footprint of the electron is removed by excluding an inner cone, and in the case of the ECAL sum also a narrow strip in the $\phi$ direction.

The electrons are also required to satisfy electron identification criteria using three simple variables. A shower shape variable which is a measure of the width of the cluster in the $\eta$ direction (the direction unaffected by showering in the tracker material) and the differences between the energy-weighted position in $\phi$, $\eta$ of the supercluster and the $\phi$, $\eta$ of the GSF track at the innermost point extrapolated to the ECAL.

The isolation and electron identification cuts applied to both electrons in the Z selection are looser than those applied to the single electrons in the W selection since the demand of two electrons in the final state significantly reduces the background. Especially for the selection of W candidates, events with a second electron candidate with $E_T > 20.0$ GeV are removed.

The neutrino from the W boson decay does not give a signal in the detector, resulting in events with unbalanced energy in the transverse plane ($E_T$). The reconstructed $E_T$ distribution for the signal and the various backgrounds for events passing the above $W \rightarrow e\nu$ selection is shown in Fig. 1. As can be seen, the most important backgrounds are the QCD backgrounds, which consist of the di-jets and $\gamma + jet$ events.

![Figure 1: The $E_T$ distribution for the $W \rightarrow e\nu$ signal together with the considered backgrounds after selection cuts applied for 10 pb$^{-1}$ of integrated luminosity.](image)

The reconstructed $M_{ee}$ distribution for the signal and the various backgrounds for events passing the $\gamma^*/Z \rightarrow e^+e^-$ selection is shown in Fig. 2.

The number of events for both $W \rightarrow e\nu$ and $\gamma^*/Z \rightarrow e^+e^-$ are normalized to an integrated
Figure 2: The $M_{ee}$ distribution for the $\gamma^*/Z \rightarrow e^+e^-$ signal together with the considered backgrounds after selection cuts applied for 10 pb$^{-1}$ of integrated luminosity.

luminosity of 10 pb$^{-1}$. Mis-calibration and mis-alignment expected for the initial data taking was applied up to the High Level Trigger, while ideal conditions are assumed for the offline reconstruction.

4 Determination of selection efficiencies from the data

The efficiencies of the selection criteria used in the measurement of the W and Z cross sections will be measured from the data using a tag and probe methodology [6]. The method relies upon $\gamma^*/Z \rightarrow e^+e^-$ decays to provide an unbiased, high-purity, electron sample with which to measure reconstruction and selection efficiencies. One of the electrons, the tag, is required to pass stringent electron identification criteria whilst the other electron, the probe, is required to make an effective mass close to $M_Z$ and satisfy a subset of criteria depending on the efficiency under study.

5 Backgrounds studies

The electroweak background in the W sample consists mostly of $\gamma^*/Z \rightarrow e^+e^-$ events (4.5% of the selected W candidates) with one electron escaping detection, and W and Z decays to $\tau$'s followed by a $\tau$ decay to an electron (1.3% of the selected W candidates). Since these backgrounds are small, and because they arise from electroweak cross sections that can be computed reliably, they can be estimated with adequate precision from simulation.

The hadronic background in the $W \rightarrow e\nu$ sample consists mostly of QCD events with two hard $p_T$-balanced jets (di-jet events), in which one jet is misidentified as an electron while missing transverse energy results from mismeasurements. The size of the di-jet background depends on the probability of a jet faking an electron. This is hard to estimate and control from simula-
tion, and therefore the QCD background must be measured from data. Signal and background efficiencies for requiring events with $E_T > 30$ GeV are estimated from the data, and once efficiencies are known, the signal and background yields can be solved for algebraically in terms of the total observed event yields above and below the $E_T$ requirement.

For di-jet and $\gamma + jet$ backgrounds (their sum will be called QCD background from now on), we have studied events which have passed the electron selection, with the offline track isolation requirement inverted.

To obtain the $W \rightarrow e \nu$ $E_T$ template, we use $\gamma^* / Z \rightarrow e^+ e^-$ candidates with one electron removed from the $E_T$ calculation. Firstly, we ensure that one of the electrons in the $\gamma^* / Z \rightarrow e^+ e^-$ events will satisfy the conditions imposed on the electron in the $W \rightarrow e \nu$ selection. The neutrino is then emulated by making the vector sum over the calorimeter towers, but excluding the energy of the supercluster of the second electron. In order to account for the difference in kinematics between $W \rightarrow e \nu$ and $\gamma^* / Z \rightarrow e^+ e^-$ events, the transverse momentum of the $\gamma^* / Z$ is subtracted from the ersatz $E_T$, which is then scaled by the ratio of the $M_W / M_Z$ before the $\gamma^* / Z \ p_T$ is added back on. As shown in Fig. 3, $\gamma^* / Z \rightarrow e^+ e^-$ candidates with one electron removed, provide a reasonable representation of the $E_T$ distribution in $W \rightarrow e \nu$ events.

![Figure 3](image-url)

Figure 3: Missing transverse energy in the ECAL Barrel as reconstructed in $W \rightarrow e \nu$ (shown as blue line), compared with the missing transverse energy calculated by excluding the supercluster energy of an electron in $\gamma^* / Z \rightarrow e^+ e^-$ (shown as dashed red line), after corrections for the kinematics of leptons from the decay.

Having checked, for both signal and background, that the data-driven estimation of $E_T$ efficiency is adequately unbiased with respect to the direct MC prediction, one can extract a background-subtracted $W$ yield via solution of two equations relating the two unknown signal and background yields to the observed yields and $E_T$ efficiencies for signal and background. This calculation results in a $W$ signal yield of $37500 \pm 453$ (stat) events. This number is to be compared with the true $W$ yield in this signal/background cocktail of $37221$ events. The estimated systematic uncertainty on the $W$ yield from the modelling of the QCD background and the $W$
signal $E_T$ distributions is 2.1% while the systematic uncertainty related to the accuracy of the prediction of the electroweak backgrounds remains to be estimated.

The electroweak background in the $\gamma^*/Z \to e^+e^-$ channel is expected to be small and can be estimated adequately using simulation. The hadronic background in the $\gamma^*/Z \to e^+e^-$ channel results from one or both leptons originating from jets. The predominant production mechanisms include di-jet production where the jet fragments are misidentified as leptons, and $W, Z +$ jet production where one of the leptons is from electroweak boson production and the other is from a jet. As can be seen from Fig. 2, the background expected after the $\gamma^*/Z \to e^+e^-$ selection is quite small, totalling to 0.35% of the signal in the $70 < M_{ee} < 110$ GeV region.

6 Results

Tables 1,2 give a summary of the results for the $W \to e\nu$ and the $\gamma^*/Z \to e^+e^-$ cross section measurement. The offline single electron reconstruction and selection efficiency, $\varepsilon_{\text{offline}}$, and the single electron trigger efficiency, $\varepsilon_{\text{online}}$, were estimated using the tag and probe technique. The offline electron reconstruction and selection efficiency is measured with respect to superclusters that fall in the ECAL fiducial region and pass the transverse energy requirement. The single electron trigger efficiency is measured with respect to electrons that pass the offline selection.

For $W \to e\nu$, the event efficiency is written as

$$\varepsilon_{\text{total}} = \varepsilon_{\text{offline}} \times \varepsilon_{\text{online}}$$

while for $\gamma^*/Z \to e^+e^-$, the event efficiency is written as

$$\varepsilon_{\text{total}} = \varepsilon_{\text{offline}}^2 \times \left[ 1 - (1 - \varepsilon_{\text{online}})^2 \right]$$

since both electrons should pass the offline selection and the trigger could be fired by one or both electrons. The selected number of events for the $\gamma^*/Z \to e^+e^-$ case is given after a $70 < M_{ee} < 110$ GeV cut is applied. This selection criterion is included in the acceptance calculation.

The uncertainties quoted in the Tables 1,2 are of statistical nature only. The tag and probe efficiencies were estimated without the presence of background. Subtraction of background in the probe sample is expected to increase the statistical uncertainty in the single electron efficiency by a factor of 1.07 (1.31) for the $W$ ($Z$) electron selection under the assumption that the background can be subtracted without additional systematic uncertainties. Systematic uncertainty estimations have been explicitly performed in some cases. The systematic uncertainty from the theoretical uncertainties on the acceptance calculation is estimated to be 2.5% for the $W \to e\nu$ and 2.3% for the $\gamma^*/Z \to e^+e^-$ final state. The estimate is based on the methods described in [7, 8], recalculated with 10 TeV center-of-mass energy and MSTW 2008 PDFs [9]. For both channels a systematic uncertainty of 10% [10] arises from the measurement of the integrated luminosity. Methods for estimating the systematic uncertainties of the signal yields and their efficiencies have not yet been fully developed; their potential magnitude can be estimated from the deviation of the results from the true values observed in the simulation. For $\gamma^*/Z \to e^+e^-$, the absence of a background subtraction in the selected signal yield induces a bias of +0.35%, and the efficiency estimate is biased with respect to the true value by -0.35%. When background subtraction is introduced for the electron efficiency measurement, it is expected to result in an additional statistical uncertainty for the cross section of 0.5%. The sum in quadrature of these
biases, including the acceptance, is 2.4%. For $W \to e \nu$, the background subtraction procedure introduces a systematic uncertainty of 2.1% in the signal yield, primarily due to correlations in backgrounds between the discriminating variables used. The efficiency is biased by two effects, the inefficiency of the veto on dielectron events (+0.9%), and the electron efficiency estimate from $\gamma^*/Z \to e^+e^-$ (-2.2%). When background subtraction is introduced for the W electron efficiency measurement, it is expected to result in an additional statistical uncertainty for the cross section of 0.2%. The sum in quadrature of these biases, including acceptance, is 4.0%. For both W and Z, the efficiency- and acceptance-corrected signal yields have statistical uncertainties and systematic biases well below the expected uncertainty in the cross section from the integrated luminosity measurement.

Table 1: Results for the $W \to e \nu$ cross section measurement. The uncertainties quoted are of statistical nature only. Systematic uncertainties are discussed in the text.

<table>
<thead>
<tr>
<th>$N_{selected} - N_{bkgd}$</th>
<th>$37500 \pm 453$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tag&amp;Probe $\varepsilon_{off line}$</td>
<td>$74.4 \pm 0.6 %$</td>
</tr>
<tr>
<td>Tag&amp;Probe $\varepsilon_{trigger}$</td>
<td>$97.2 \pm 0.3 %$</td>
</tr>
<tr>
<td>Tag&amp;Probe $\varepsilon_{off line \times trigger}$</td>
<td>$72.3 \pm 0.6 %$</td>
</tr>
<tr>
<td>Acceptance</td>
<td>$36.6 \pm 0.1 %$</td>
</tr>
<tr>
<td>Int. Luminosity</td>
<td>$10 , pb^{-1}$</td>
</tr>
<tr>
<td>$\sigma_W \times BR(W \to e \nu)$</td>
<td>$14200 \pm 200 , pb$</td>
</tr>
</tbody>
</table>

Table 2: Results for the $\gamma^*/Z \to e^+e^-$ cross section measurement. The uncertainties quoted are of statistical nature only. Systematic uncertainties are discussed in the text.

<table>
<thead>
<tr>
<th>$N_{selected}$</th>
<th>$4273 \pm 65$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_{bkgd}$</td>
<td>assumed 0.0</td>
</tr>
<tr>
<td>Tag&amp;Probe $\varepsilon_{off line}$</td>
<td>$90.4 \pm 0.3 %$</td>
</tr>
<tr>
<td>Tag&amp;Probe $\varepsilon_{trigger}$</td>
<td>$99.88 \pm 0.02 %$</td>
</tr>
<tr>
<td>Tag&amp;Probe $\varepsilon_{total}$</td>
<td>$81.6 \pm 0.5 %$</td>
</tr>
<tr>
<td>Acceptance</td>
<td>$40.4 \pm 0.2 %$</td>
</tr>
<tr>
<td>Int. Luminosity</td>
<td>$10 , pb^{-1}$</td>
</tr>
<tr>
<td>$\sigma_{Z/\gamma^<em>} \times BR(Z/\gamma^</em> \to e^+e^-)$</td>
<td>$1300 \pm 20 , pb$</td>
</tr>
<tr>
<td>$\sigma_{Z/\gamma^<em>} \times BR(Z/\gamma^</em> \to e^+e^-)$ (MCtruth)</td>
<td>$1296 , pb$</td>
</tr>
</tbody>
</table>
7 Summary

We have presented a strategy for the early measurement of the inclusive W and $\gamma^* / Z$ production cross sections in the electron decay channel, assuming 10 pb$^{-1}$ of $pp$ collision data. A simple and robust electron selection is used, resulting in roughly forty thousand W and four thousand Z candidates expected. This selection is intended to be easily tuned and uses variables expected to be largely insensitive to the anticipated imperfections in the calibration and alignment of the CMS detector in early data. A tag and probe technique was used to determine the reconstruction and selection efficiency from the data and a method to estimate the QCD background in the W sample was developed.

The electroweak boson signal yields can be measured with a statistical precision of less than 2% in each case, after accounting for backgrounds. Using purely data-driven techniques, the efficiency of these selections can be determined to 1% statistical accuracy. Systematic biases in our methods have been estimated from simulation studies to be (summed in quadrature) 4.0% and 2.4% for W and Z, respectively. Initial precision in the integrated luminosity measurement of 10% is expected to dominate the precision of the cross section, even for samples significantly smaller than 10 pb$^{-1}$. Alternatively, this precision makes these measurements attractive as estimators of integrated luminosity for CMS.

References