Wγ at $\sqrt{s}$ =8 TeV.

Lovedeep[1], Yurii[1], Sachiko[1], Katya[2], Ilya[2], and Yutaro[3]

[1]Kansas State University, Manhattan
[2]University of Nebraska-Lincoln, USA
[3]Carnegie Melon University, USA

February 27, 2015

Abstract

We present a study of Wγ production in proton-proton collisions at $\sqrt{s} = 8$ TeV.

Contents

1 Introduction

The Standard Model (SM) has been proved to be an accurate description of production of elementary particles observed so far. The interaction of W bosons with photons is particularly important as a important test of self coupling of these bosons as predicted by non-Abelian gauge group of electroweak sector. Precise measurements of diboson properties and cross sections are a crucial step towards understanding the production of major backgrounds of Higgs boson searches at LHC.

In this analysis note we report the analysis of inclusive $W\gamma + X$ processes using leptonic decays of $W \to \ell\nu$ where $\ell = e, \mu$. The $W\gamma$ productions at tree level can be represented by Feynman diagrams in Figs. 1 and **??** as three processes: initial state radiation (ISR) where a photon is produced from one of the incoming partons, final state radiation (FSR) where a photon is radiated off one of the charged leptons from the W boson decay, and finally when a photon is produced in $s-$channel via TGC $WW\gamma$ for $W\gamma$ production. The last process is allowed only for $W\gamma$g production in the SM, as there are no neutral TGC in the SM.

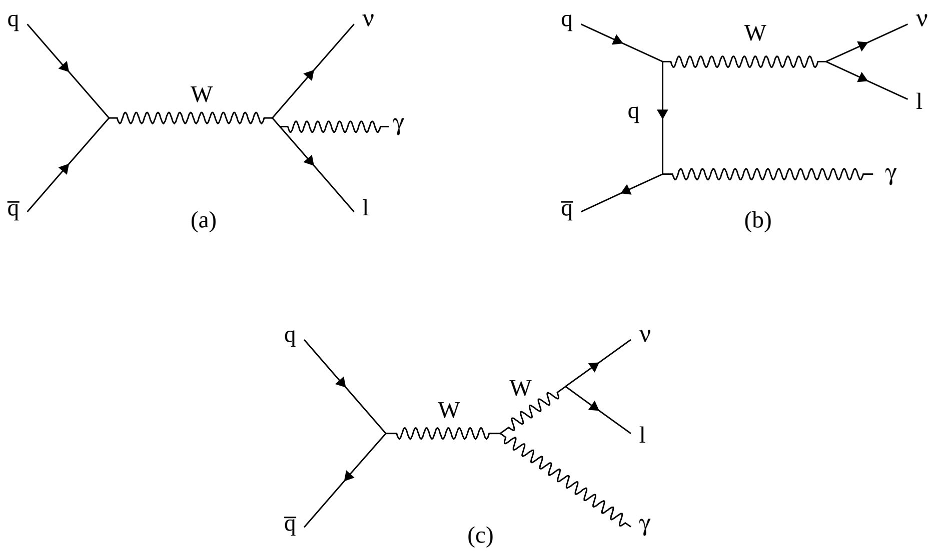

Figure 1: Feynman diagrams of the Wγ production via final (a) and initial (b) state radiation and via WWγ trilinear gauge coupling (c).

2 Data and Monte Carlo samples

2.1 Data sample

The data sample we use in this analysis was recorded by the CMS experiment in 2012. The data is collected by single electron ($p_T > 27$GeV) triggers 1. Only certified runs and luminosity sections are considered, which means that good functioning of all CMS sub-detectors is required. The selection of the validated run and luminosity sections are obtained from the following official JSON file:

- Cert_190456−208686_8TeV_22Jan2013ReReco_Collisions12_JSON.txt

The total statistics analyzed correspond to an integrated luminosity of 19.6 fb^{-1}.

The dataset used for the analysis and the corresponding run ranges are listed in Table 2. All samples have been processed using a `CMSSW_5_3_2` release version.

Dataset	Trigger name	Description
SingleElectron	HLT_Ele27_WP80*	$p_T > 27\ GeV$

Table 1: Analysis triggers for data sample.

Dataset name	Run range
/SingleElectron/Run2012A-22Jan2013-v1/AOD	190456-193621
/SingleElectron/Run2012B-22Jan2013-v1/AOD	193833-196531
/SingleElectron/Run2012C-22Jan2013-v1/AOD	198022-203746
/SingleElectron/Run2012D-22Jan2013-v1/AOD	203777-208686

Table 2: Summary of data samples used and run ranges of applicability.

2.2 Monte Carlo samples

All MC samples considered in this analysis come from the official "Summer12_53X". Events from all samples were reconstructed making use of a `CMSSW_5_3_X` release version. The simulated samples are reweighted to represent the distribution of the number of pp interactions per bunch crossing (pile-up), as measured in the data.

Information on Monte Carlo samples used for the analyses is given in Table 3 for signals and backgrounds. The corresponding leading order (LO) and next-to-leading order (NLO) cross sections are also listed in this Table.

Table 3: Summary of Monte Carlo background samples used.

Process (Fall11)	σ, pb	Dataset Name (AODSIM data tier)
$W\gamma \to l\nu\gamma$	553.92 (NLO)	`/WGToLNuG_TuneZ2star_8TeV-madgraph-tauola`
$W \to l\nu + jets$	36257.2 (NNLO)	`/WJetsToLNu_TuneZ2Star_8TeV-madgraph-tarball`
$Z \to ll + jets$	3503.71	`/DYJetsToLL_M-50_TuneZ2Star_8TeV-madgraph-tarball`
$t\bar{t} + jets + 1l$	99.44 (NNLO)	`/TTJets_SemiLeptMGDecays_8TeV-madgraph`
$t\bar{t} + jets + 2l$	23.83	`/TTJets_FullLeptMGDecays_8TeV-madgraph`
$t\bar{t} + \gamma$	1.444	`/TTGJets_8TeV-madgraph`
$Z\gamma \to ee\gamma$	159.120	`/ZGToLLG_8TeV-madgraph`

3 Object selection

In this Section we document the electron, muon, and photon identification and isolation criteria, MET criteria, and provide the results of comparing Monte Carlo simulation with data. We use cut-based selection of lepton + photon + MET where lepton is either muon or electron. Photon and MET selection are a little bit different for different channels. The second lepton veto is applied. dR(lep,pho)>0.7 where $dR = \sqrt{(d\phi^2 + d\eta^2)}$ is applied to avoid divergence of ISR and FSR contributions. No restrictions applied on number of photons in the event but the candidate with the hardest photon is selected in particular event among those photon which passed all the other cuts including dR. dR is also part of phase space selection.

3.1 Electron selection

In $W\gamma$ analysis we consider electrons with $p_T > 30$ GeV and passing the Medium-identification and isolation optimized by EGamma-POG for 2012 analyses [REFERENCE]. We summarize electron identification and isolation requirements in Table ??.

The ECAL fiducial region is defined in terms of barrel and endcap sections with pseudorapidity ranges of $|\eta| < 1.4442$ and $1.566 < |\eta| < 2.5$, respectively. An electron is considered to be within this ECAL acceptance if its associated SuperCluster (SC) is within the ECAL acceptance. Data/MC scale factors are applied.

3.2 Photon selection

Photon candidates are reconstructed as SuperClusters with $E_T > 15$ GeV in the fiducial volume of the ECAL detector: barrel (EB) with $|\eta| < 1.4442$ and endcap (EE) with $1.566 < |\eta| < 2.5$. To reduce copious background objects from jets misidentified as photons we apply the Medium identification and isolation selections as recommended by EGamma-POG [REFERENCE]. Data/MC scale factors are applied.

3.3 Muon selection

The muon selection includes the kinematics cuts $P_T > 26$ GeV and $|\eta| < 2.1$ and the muon ID cut recommended by POG [REFERENCE]. If there is the second reconstructed muon candidate with $P_T > 10$ GeV and $|\eta| < 2.4$ is found in the event, then the whole event is vetoed. No muon ID requirements on the muon to be vetoed are applied.

3.4 MET

W transverse mass cut is applied. $M_T > 50$ GeV for muon channel and $M_T > 70$ GeV for electron channel where $M_T = \sqrt{(2 \cdot P_T^{Lep} \cdot P_T^{MET} \cdot (1 - \cos{(\phi^{lep} - \phi^{MET})}))}$. MET (missing transverse enery) reconstructed with particle flow algorithm was used [REFERENCE]. TODO: write something about MET corrections and MET smearing

3.5 Pile Up reweighting

applied

3.6 Blinding

This is a policy of SMP group that all the analyses must be performed in a blinded way. We discussed how to implement blinding in our case with the statistics committee [REFERENCE] and were given the following recipe:
”We believe we have cooked up a possible procedure to blind your data in a way which prevents you from biasing your results (as the SMP-VV group wants) while not making it appreciably harder for you to analyze the data.
The idea is to produce a filter which removes at random a certain fraction of your data with pt(photon) above some threshold, such that the region where you are sensitive to new physics (say above 40 GeV) is artificially decreased with respect to background predictions. This allows you access to high-pT data while preventing you to be influenced by good or bad agreement in that region.
In addition, you should make a second dataset containing just a flat X% of the data, irrespective of photon pT. This second dataset allows you to verify that your modeling is correct, while leaving room for surprises. X should be chosen such that the resulting statistical uncertainty in the signal-sensitive region is at least three times as large as the systematic uncertainty; and in any case it should not exceed 10%.
In practice this means:
1) decide a pTg threshold above which there starts to be sensitivity to new physics effects. We probably want this to be safe, so maybe P_T^{γ} >40 GeV could be the right ballpark;
2) have somebody decide decreasing ”acceptance factors” S(P_T^{γ} >) for the bins above threshold, unknown to you. In the graph I see three bins above 40, so they could be: in the region 40-60 and 60-80 GeV something around 0.8; in the region 80-200 GeV something around 0.6; in the region above 200 GeV something around 0.4.
3) the same person should produce data ntuples only containing data passing the requirement gRandom→Uniform()¡S(P_T^{γ}), and pass them to you for the analysis. For the Monte Carlo simulations, no such random removal is to be applied.
4) The same person could produce data ntuples containing a flat X% of all data, using the same random procedure (such that you do not get, e.g., only data from the first part of the run). This

factor X need not be hidden from you, as you will want to try and match your MC prediction to the data to see that things are in good order.
5) You can then do the analysis using the two reduced samples, checking high-pt data as well as verifying the agreement at the high-pt end with the X% of the sample.
6) Unblinding just means removing the filter and doing the full analysis on the whole data.
Note that as you get rescaled data but the MC is not rescaled, you cannot make any inference from the (dis)-agreement of high-ptg data. You still can, however, check other distributions (like muon pt or eta or whatever) by doing separate graphs for the separate ptg bins, and suitably renormalizing them." [Tommaso Dorigo for the Statistics Commitee]
We decided on P_T threshold of 40 GeV and, relying on systematic error of 10%, quoted by 7 TeV measurement [REFERENCE], on X%=5%. Thus, for blinding we are using either 5% of data or are just looking at data with $P_T < 40$ GeV

3.7 Vjets/$V\gamma$ overlap removal

To prepare plots data vs MC, the overlap removal between inclusive and exclusive samples was done. [TODO: add the explanation here or to appendix? Explain studies?]

4 Background Subtraction

Background subtraction

4.1 $jets \rightarrow \gamma$ background estimation and subtraction

The main background to Wγ W+jets which is XX% of selected events and we don't see any way to significantly reduce it without reducing Wγ as well. Photon ID helps to reduce W+jets somewhat but there is still significant number of jets which are reconstructed as photons and pass all the photon ID criteria. DY+jets is another source of $jets \rightarrow \gamma$ background but this one is significantly supressed with W transverse mass cut [REFERENCE to subsection].

4.1.1 Template Method Description

The template method is used to estimate $jets \rightarrow \gamma$ background. The idea of the template method is the following. Suppose we have two varibles which are independent (at least for background events) V_{fit} and $V_{sideband}$. We consider V_{fit}=$\sigma_{i\eta i\eta}^{photon}$ and $V_{sideband}$=I_{ch-SCR}^{photon} and vice versa. We apply all the selection criteria on data except $\sigma_{i\eta i\eta}^{photon}$ and I_{ch}^{photon}. $\sigma_{i\eta i\eta}^{photon}$ is part of photon ID. I_{ch-SCR}^{photon} is not part of photon ID but it is strongly correlated with I_{ch}^{photon} which is part of photon ID. Then one needs to prepare the templates which would describe the distribution of true photons T_{true} and of fake photons (e.g. jets reconstructed as photons) T_{fake}. In case of V_{fit}=$\sigma_{i\eta i\eta}^{photon}$ the true template is taken from FSR events of Zγ-selected dataset, in case of V_{fit}=I_{ch-SCR}^{photon}, random cone isolation method was used. Nominal cut of $V_{sideband}$ applied to prepare true-γ templates. To prepare the fake template, distribution of V_{fit} in sideband range of $V_{sideband}$ is taken. Given that V_{fit} and $V_{sideband}$ are assumed to be independent, the V_{fit} distribution of fake photons in nominal and sideband ranges of $V_{sideband}$ is the same. The leakage of true photons to sideband range of $V_{sideband}$ is subtracted based on signal-MC sample normalized to data luminocity.
Then the nominal cut of $V_{sideband}$ is applied on data and the distribution of V_{fit} is fitted with $F(V_{fit}) = N_{true} \cdot T_{true}(V_{fit}) + N_{fake} \cdot T_{fake}(V_{fit})$ where N_{true} and N_{fake} are fit parameters corresponding to fractions of true and fake photons in fitted data sample. To extract yield from the N_{true}, efficiency of the V_{fit} is applied on the extracted yield. The efficiency is estimated from signal MC. [TODO: estimate it from Zg FSR events]. In case if V_{fit} is variable used in photon ID and there is a bin boundary exactly at cut value, the yield can be extracted by summing over the true γ template. It is possible if V_{fit}=$\sigma_{i\eta i\eta}^{photon}$ for most bins. However, the binning of template may be changed in case if there are not enough statistics in true or fake template and then it's possible that cut value of V_{fit} will not coincide with any template bin boundary and

thus the method of extracting yields by summing over histogram will become impossible and one will have to apply the efficiency of V_{fit}. In case of V_{fit}=I^{photon}_{ch-SCR}, the variable for fit is charged isolation computed with particle flow algorithm with super cluster removal [REFERENCE] but the variable used in the photon ID is computed without super cluster removal. These I^{photon}_{ch-SRR} and I^{photon}_{ch} are strongly correlated but they are not the same and thus to extract yield from the fit output, one has to apply efficiency of I^{photon}_{ch} used for photon ID.

There were several studies performed to check whether our $\sigma^{photon}_{i\eta i\eta}$ and I^{photon}_{ch-SCR} are independent and the studies showed that there is dependence between these two variables and the result vary significantly on sideband definition.

TODO: include all our chiso vs sihih, chiso vs phopt and sihih vs phopt plots.

We studied carefully whether this problem had place in approved CMS Zγ analysis [REFERENCE] and found out that the problem was there too and we decided to follow their method how to deal with that. The idea of the approach is not to assume that the V_{fit} and $V_{sideband}$ are independent but to find the range of $V_{sideband}$ which would give the correct distribution for V_{fit}. We first do fits on the MC-mixture which mimics data and select the sideband range which would give the true and fake γ yields the same as predicted by MC-truth information. Then we apply the same sideband range to data and perform fits on data. The fits are performed separately for barrel and endcap photons and separately for each pt bins. The systematic error according to this method [WILL BE] computed the same as it was done in Zγ analysis mentioned above and WILL BE documanted in section [REFERENCE (systematics)].

4.1.2 Photon I_{ch-SCR} fits

Template method with I_{ch-SCR} fits is our primary method for $jets \to \gamma$ background estimation. The signal template was obtained with Random Cone Isolation method [REFERENCE] and the background template was obtained from sideband of $\sigma^{photon}_{i\eta i\eta}$. TODO: add here table for each with fit ranges and binning, sideband ranges, percent of signal to sideband leakage and whether the templates for different pt bins were combined, separately for tru and fake templates. The table must be produced automatically by script in latex format. Add 2D colorful plots for sideband variation for MC-mixture and data fits (probably, just one pt bin here and all bins to appendix) and plot the black dot on MC-mixture plot where the output coincide to MC-truth

4.1.3 Photon $\sigma_{i\eta i\eta}$ fits

$\sigma_{i\eta i\eta}$ fits show bad performance probably due to especially high impact of the first bin in I_{ch-SCR} distribution which is always thrown from the sideband. TODO: but we can document here all our studies, which are pretty much the same as for I_{ch-SCR} fits. Also we can add Zγ FSR selection here.

4.1.4 Transfer function and fake distributions

[Would be good if Yutaro describes here the method where he generates all the distributions with transfer function and if we had all the plots generated with this method instead of or in addition to data vs MC plots. Given that sihih fits show problems for all three of us, he may want to construct this transfer function and construct samples with true and fake rates which Katya and Lovedeep obtain with Random Cone Isolation method]

4.2 $e \to \gamma$ background estimation and subtraction

MC-based, not separated from true γ background at the moment. The template method describe above doesn't distinguish between true γ and $e \to \gamma$ contributions because electrons and photons reconstructed in ECAL would have the same shower shape and isolation distributions and thus $e \to \gamma$ proceeds as part of true γ component given by template method. For the electron channel, Zγ is the main source of $e \to \gamma$ background. We managed to significantly supress the background by introducing Photon Pixel Veto, Photon conversion veto and Matching-GSF electron veto cuts.

TODO: add quantitative effects of each cut to the table and dut to MC-based estimation this background doesn't exceed [QUESTION] X% of signal. QUESTION: Does our overlap removal procedure takes care of removing this background from DYjets? For the muon channel, Zγ is fully true γ background and doesn't contribute to $e \to \gamma$. WW, WZ, ZZ, WWγ etc. processes can have muon and electron misidentified as photon in final state and thus contribute to $e \to \gamma$ background in muon channel. Overall contrubution from these sources for muon channel doesn't exceed X% of signal. TODO: add table with all MC-based contributions, try to split between true-gamma and e to gamma.

4.3 True γ background estimation and subtraction

MC-based. Main source of true γ background are Zγ, W$\gamma \to \tau\nu\gamma$, WWγ

4.4 $jets \to lepton$ and $jets \to lepton + jets \to \gamma$

Negligible (show from gamma+jets, diphoton and QCD MC samples)

5 Detector Resolution Unfolding

Performed with D'Agostini method using RooUnfold package [REFERENCE] Checked that D'Agostini and simple inversion matrix give the same result (yields are the same, errors are different). Also checked that privetly implemented unfolding gives the same result as RooUnfold. Also the MC closure test is routinely performed every time when unfolding is applied on data. TODO: switch to more recent version of RooUnfold TODO: present table: data yields; unfolded data yields; MC yeidls; unfolded MC yields; gen MC yields TODO: think whether additional eff reco is needed here Unfolding recommendations for SMP [1] https://indico.cern.ch/event/322577/ [2] https://twiki.cern.ch/twiki/bin/view/CMS/TwikiSMP-GENRecommendations#Unfolding_How_to

6 Acceptance X Efficiency

6.1 MC-based computation of Acceptance X Efficiency

Computed as combined value; Based on signal MC
Numerator:
For total cross section: selected signal MC yields with PU weight applied
For differential cross section: selected signal MC yields with PU weight applied in phoEt GEN bins 15-20-25... Transition to phoEt RECO bins performed with unfolding
Denominator:
For each event in unskimmed signal MC file, photon and lepton (two leptons for Zg) which refer to Wg (Zg) and determine dR GEN and phoEt GEN
For total CS: calculate number of events for which $dR_{gen} > 0.7$ and $phoEt_{gen} > 15$ GeV
For differential CS: calculate number of events for which $dR_{gen} > 0.7$ in 15-20-25... $phoEt_{gen}$ bins
Computed together, bin-by-bin

6.2 Electron, photon and muon data/MC scale factors

From POG

6.3 M_T^W scale factors

TODO (or not TODO)

7 Systematic Uncertainties

TODO: make list of sources

7.1 Uncertanties due to $jets \rightarrow \gamma$ background estimation

Following approach used in Zγ analysis. ================================ http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2013/280 ================================ In Zg, results of Ipho and sihih template fits are treated as separate measurements and combined after cross sections for both of them as computed ================================ Section 7.1.4 in [REFERENCE](Uncertainty due to Background Templates):

- For each I_{ch} range fit MC mixture and determine $b_i = f_i^{MC}/f^{MC}$, where f^{MC} is MC-truth signal fraction and f_i^{MC} - signal fraction determined from fit for the i-th Ich range
- Determine $U_i = f_i^d - b_i f^d$ where f^d is signal fraction determined from data fit with fake template obtained from I_{ch} range selected from MC closure test and f_i^d is signal fraction determined from data fit with fake template obtained from i-th I_{ch} range
- Fit U_i distribution with Gaussian and assign sigma as systematic uncerntainty

Appendix C. in [REFERENCE] (Extracted Signal Fraction in Jet Data and QCD MC):

- Select all possible Ich ranges where QCD MC sihih template is in good agreement with MC-truth fake template taken from Drell-Yan MC (”good” means Kolmogorov-Smirnov test probability ¿ 30%)
- Apply each of selected Ich ranges on jet-enriched dataset, use it as fake sihih template, fit data and get signal yields
- Fit distribution of obtained signal yields and assign sigma as syst ucerntainty

Same approach used for Ipho template fits with sihih ranges

NOT simply assigning difference in yields estimating from different sideband ranges. TODO

7.2 Uncertanties due to Unfolding

Unfolding recommendations for SMP [1] https://indico.cern.ch/event/322577/ [2] https://twiki.cern.ch/twiki/bin/view GENRecommendations#Unfolding_How_to TODO: follow what was written in the e-mail which Ilya sent

7.3 Uncertanties due to limited MC statistics

TODO

7.4 Uncertanties due to constants used

TODO: some errors quoted for luminocity, scale factors, maybe something else

8 Cross Section

8.1 Theoretical Cross Section

The theoretical cross section was computed with MCFN [REFERENCE] in NLO and is for the dedicated signal MC sample $\sigma(sample)$=553.92 pb. The MC sample was generated with MADGRAPH [REFERENCE] and the cross section in our selected phase space was computed as $\sigma(phasespace) = \sigma(sample) \cdot N_{phasespace}/N_{sample}$ where $N_{phasespace}$ and N_{sample} are numbers of events falling into selected phase space and generated in the whole MC sample respectively. For the differential cross section, $N_{phasespace}$ is number of events falling into specific pt bin and to compute $d\sigma/dP_T^\gamma$, we divide over the bin width.

8.2 Cross Section Measurement

The cross section calculation is performed with the following analysis flow

- cut-based event selection
- $jets \rightarrow \gamma$ data driven background subtraction
- $e \rightarrow \gamma$ and true γ MC-based background subtraction
- TODO: perform data-driven $e \rightarrow \gamma$ background estimation for electron channel or prove that effect is small
- TODO: show that true photon + fake lepton background is neglibgible (just on MC)
- detector resolution unfolding correction (for differential cross section only)
- acceptance x efficiency correction (bin-by-bin)
- data/MC electron and photon scale factors correction
- TODO: apply muon scale factors and photon scale factors for muon channel
- TODO: if other syst errors are not very high, compute and apply WMt scale factor
- divide over luminocity (19.5 fb^{-1})
- divide over bin width (for differential cross section only)
- estimate systematic errors

A APPENDIX: Comments about our code

A.1 ggNtuples

Prepared by Central Taiwan University and Kansas State University [REFERENCE]

A.2 SUSY ntuples

A.3 Katya's framework

based on ggNtuples

A.4 Lovedeep's framework

based on ggNtuples

A.5 Yutaro's framework

based on SUSY ntuples

A.6 Cross Checks

We are doing cross checks between different analysts and Zγ check. We cross checked muon selection based on completely indepedent frameworks based on different ntuples (ggNtuples and SUSY ntuples). We cross checked electron selection based on the same ntuples (ggNtuples) but different analyzers. We are going through the template fit procedures for $jets \rightarrow \gamma$ background estimation which are implemented by different people. To compare selection, we are using event-by-event approach. However, this check is TODO (doesnt pass all the tests). It doesnt give the same result for $\sigma_{i\eta i\eta}$ fits on MC mixture (exactly the same MC-mixture was used). And it doesnt give the same result for I ch fits for electron channel if use different (but cross-checked) selection trees. Also, we are doing Zγ check: applying all the same methods and procedures on Zγ selected events and are cross-checking with approved CMS Zγ measurement [REFERENCE].