CMS Physics Analysis Summary

2008/07/31

Data-Driven Estimation of the Invisible Z Background to the SUSY MET Plus Jets Search

The CMS Collaboration

Abstract

The large missing transverse momentum (MET) plus jets final state topology has great potential for discovery of new physics involving dark matter candidates early in LHC running, if Standard Model backgrounds can be understood. The irreducible background of high transverse momentum (p_{T}) Z bosons, with $\mathrm{Z}\rightarrow \nu\bar{\nu}$ produced in conjunction with high E_{T} jets, can be estimated from samples of photon plus jets and $\mathrm{W}\rightarrow \mu\nu$ plus jets. These samples have much higher statistics than the previously studied $\mathrm{Z}\rightarrow \mu\mu$ plus jets method. In addition, new physics will likely contribute to these control samples differently; they thus provide useful cross checks in the presence of new particles. For 100 pb^{-1} , the contribution of "invisible Z" events to the MET $>$ 200 GeV region can be estimated with a statistical uncertainty of order 10 % using photon plus jets. Experimental systematic uncertainties obtained via data-driven techniques are expected to be roughly 20 % and are statistically limited. Theoretical uncertainties on photon production are under study and will be reported in an update to this note, along with methods for obtaining the relative normalizations directly from data. For W plus jets, the statistical, systematic and theoretical uncertainties are estimated to be 29 %, 23 %, and 8 % respectively.

1 Introduction

The large missing transverse energy (MET) plus jets topology has great potential for discovery of new physics involving dark matter candidates such as the standard SUSY neutralino. Important observations could come early in LHC running if backgrounds are well understood. A search strategy was outlined in the CMS Physics Technical Design Report (PTDR) [1], where the backgrounds are shown to include an irreducible "invisible Z" background from high p_T Z bosons plus jets, with $Z \to \nu\bar{\nu}$. A data-driven method was also presented for estimating this background based on the clever use of "standard candle" high p_T, $Z \to \mu\mu$ production. The basic idea is to remove the leptons in order to calculate a total MET that models the MET spectrum for $Z \to \nu\bar{\nu}$ plus jets, after correcting for relative branching fractions, efficiencies, and acceptances. Unfortunately, very few $Z \to \mu\mu$ events satisfying all the search criteria are expected in data samples substantially less than 1 fb^{-1}. One typically reduces the number of jets required in the event selection in order to bolster statistics. However, as seen in Fig. 1, the ratio of N+1 to N jets in association with vector bosons (in this case W plus jets) is sensitive to the details of the event kinematics through the event selection criteria, including the p_T of the boson.

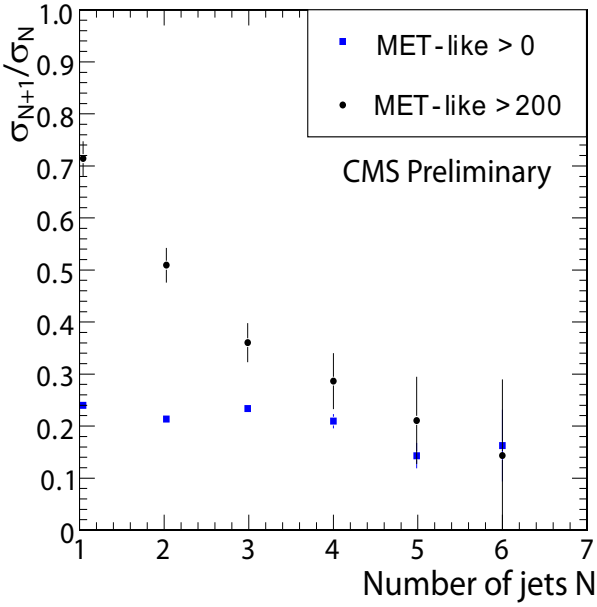

Figure 1: Ratio of events with N+1 to N jets for various values of N in ALPGEN [3] W + jets events. Here a "jet" is defined as clustered calorimetric energy of at least 30 GeV within a cone of radius $R = 0.5$ in $\eta - \phi$ space. The ratios vary with jet definition and also for different regimes of the MET-like quantity. The latter is simply MET in invisible Z events, but it is the combination of MET and lepton or photon p_T in the W or photon case, as described in the text.

Two higher-statistics competitive alternatives to the "standard candle" approach are presented in this note. They correspond to the use of samples containing a high-p_T photon, or W, produced with high-p_T jets. The MET spectrum is obtained by removing the identified photon or lepton, and correcting for residual differences between these events and invisible Z events. The better statistics of these samples allows one to apply all search criteria. These additional samples also provide important cross-checks to each other and to the "standard candle" approach, because they are likely to be impacted differently by new physics.

The differential production cross sections for W, Z or photon plus exactly two additional par-

tons, including all contributing subprocesses, are shown in Fig. 2. The production of W bosons is higher by a factor of three at high-p_T, as expected, while photon production is within 20% of Z production. Above ~150 GeV/c boson p_T these ratios depend mostly on the electroweak characteristics of the events. Stated another way, the hadronic parts of these events are not easily predicted, but to good approximation do not depend upon whether the boson is a Z, W, or photon. The ratios are thus relatively robust to variations in selection criteria, such as number and transverse energies of jets. In the absence of large contributions to these samples from new physics, they therefore have the potential to do a good job of predicting the MET spectrum for invisible Z's at high p_T.

The Monte Carlo (MC) samples that are used for the studies presented here do not take into account theoretical uncertainties associated with such things as Q^2 scale variations, and contributions from uncertainties in parton distribution functions. Initial studies indicate that the former can affect the relative normalization of photon+jets to Z+jets events at a level that is roughly comparable to the other systematic uncertainties, while the latter have much smaller impact. The difference due to collinear photon production is expected to be mitigated by isolation requirements.

In general there is also a difference in the η distribution of photons relative to that of Z bosons, as a result of different phase space factors for massive Z bosons versus massless on-shell photons, and to a lesser extent, due to the different vector and axial couplings. However, at sufficiently high p_T the bosons tend to be found in the central region, which significantly mitigates the difference. The MC distributions for bosons with $p_T > 200$GeV/c is seen in Fig. 3.

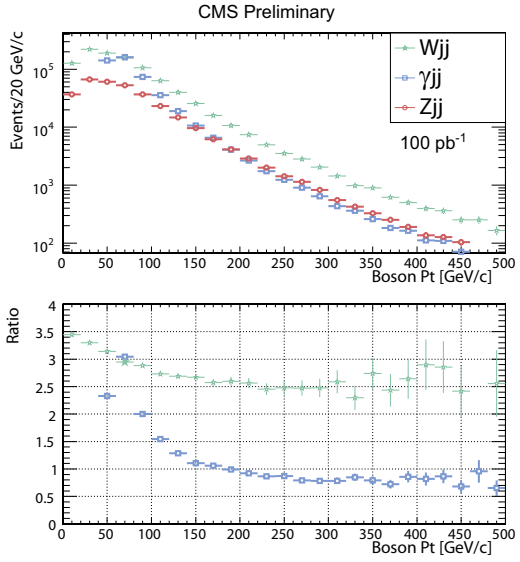

Figure 2: Top: differential event yield as a function of boson p_T, for the processes pp $\rightarrow$ boson + 2 partons (boson = W, γ, or Z) as evaluated by MadGraph [2] at generator level for 100 pb^{-1} integrated luminosity. Bottom: ratios of the yields of W relative to Z and γ relative to Z. The γ to Z ratio levels out at a value that is simply predicted by the differences in the couplings of Z's versus photons to up-like and down-like quarks.

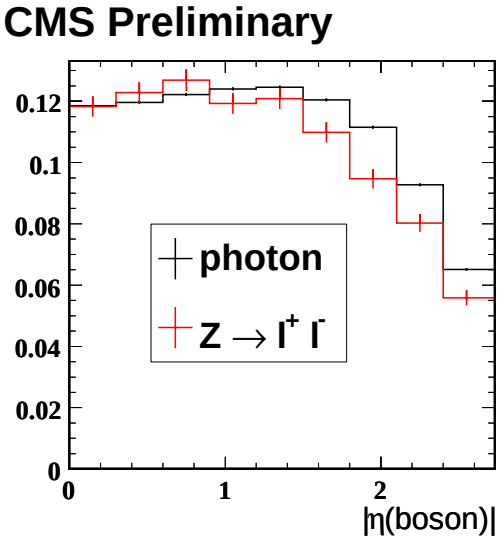

Figure 3: Differential event yield as a function of pseudorapidity $\eta(boson)$, for bosons with $p_T > 200$ GeV/c produced in conjunction with inclusive numbers of jets, as obtained using fully simulated ALPGEN [3] events. The histograms are normalized to unit area to facilitate comparison of shapes.

As noted above, an advantage of using a variety of samples to study the invisible Z background is that they allow for very important cross-checks if there is new physics in the data. In particular, if there is a massive neutral and stable dark matter candidate, it is expected to be part of a larger spectrum of new particles. If produced at sufficient rates at the LHC, some of these new-particle decays could contribute to, and distort, the background control samples. One can however expect the impacts to differ for the W, Z and photon samples. For example, in the Minimal SUSY Standard Model (MSSM) there are many scenarios in which there are many events produced containing leptons, but not photons, while the converse is often true in Gauge Mediated SUSY Breaking (GMSB) models. Consistency checks between W, Z and photon samples thus provide an important element in the understanding of any new signal and potentially greater confidence in the modeling of the Standard Model background.

In this note, results are presented from studies performed on MC samples to develop data-driven methods for employing samples of high p_{T} photon, or W, plus jets events to estimate the irreducible "invisible Z" contribution to the MET plus jets topological signature.

2 Object and Event Selection

For events with large MET and significant jet activity, the trigger critera are:

- Level 1 trigger: $H_{\mathrm{T}} >$ 300 GeV where H_{T} is defined as the summed E_{T} of all jets. The jets are defined using the transverse energy sums in 12x12 calorimeter trigger tower windows, where the trigger towers are defined as 5x5 crystals in the ECAL to match a single HCAL tower.
- High Level Trigger (HLT): $H_{\mathrm{T}} >$ 350 GeV and uncorrected MET $>$ 65 GeV. The H_{T} uses all jets in the event with uncorrected $E_{\mathrm{T}} > 5$ GeV. The jets are clustered using the iterative cone algorithm with a radius of 0.5 in $\eta - \phi$ space.

The photon + jets measurement can only be performed on a high-energy EM trigger because these events have no intrinsic MET and will typically fail the MET requirement. The translation to obtain a $\mathrm{Z} \to \nu\bar{\nu}$ prediction is performed by correcting the photon events back up to 100% of the cross-section, then correcting down to the estimated efficiency of $\mathrm{Z} \to \nu\bar{\nu}$ events to pass the SUSY search trigger.

For the $\mathrm{W} \to \mu\nu$ measurement, on the other hand, the SUSY search trigger can be used to collect events because the high p_{T} muon is very nearly invisible to the calorimeter triggers. This enables events in the $\mathrm{W} \to \mu\nu$ + jets channel to pass the SUSY search trigger with high efficiency. Moreover, there is no need for additional corrections to translate to the $\mathrm{Z} \to \nu\bar{\nu}$ prediction. Events are nevertheless collected using the single muon trigger for background studies purposes.

The offline selection criteria [1] are:

- pass basic beam halo clean-up and have at least one reconstructed primary vertex;
- uncorrected MET $>$ 200 GeV;
- at least three clustered jets, with uncorrected $E_{\mathrm{T}} > 30$ GeV and $|\eta| < 3$;
- leading jet $E_{\mathrm{T}} > 180$ GeV and $|\eta| < 1.7$, second leading jet $E_{\mathrm{T}} > 110$ GeV;
- passing angular cuts between the MET and jets (for QCD suppression);
- $H_{\mathrm{T}} > 500$ GeV, where H_{T} is defined as the sum of MET and the E_{T} of the 2^{nd}, 3^{rd} and (if present) 4^{th} leading jets.

The γ + jets and W + jets analyses presented below are targeted to model the spectrum of $Z \to \nu\bar{\nu}$ events that satisfy the above selections.

3 Photon+Jets Channel

3.1 Event Selection in the Photon+Jets Channel

The event selection in the photon plus jets channel starts by collecting prompt photon candidates with $E_T > 100$ GeV. Photons are rejected if their superclusters are also reconstructed as electrons; those that pass this "electron veto" are required to pass tracker and calorimeter isolation cuts as described below.

The tracker isolation variable is computed as the sum p_T of tracks in a hollow cone (inner radius 0.02, outer radius 0.3 in $\eta - \phi$ space) about the vector from the nominal interaction point to the electromagnetic calorimeter (ECAL) supercluster centroid, and is required to be less than 1 GeV/c. For the most part, reconstructed tracks for conversion electrons and positrons do not contribute to the p_T sum. This is because they are generally not reconstructed by the standard track reconstruction sequence or they are found within the inner veto cone.

A relative calorimeter isolation requirement is used in order to achieve a more uniform photon efficiency as a function of the photon transverse energy. In particular, the sum of transverse energy in the ECAL and hadronic calorimeter (HCAL) is computed in a cone of radius 0.3 around the ECAL supercluster centroid, omitting the E_T of the photon supercluster, and required to be less than $1\,\text{GeV} + 0.02 \cdot E_T(\gamma)$. The efficiency of isolation cuts is estimated using a tag-and-probe method with $Z \to ee$ events, as discussed in Section 3.3.

The SUSY MET + jets search criteria, (Section 2) must undergo a few modifications before they can be applied to the photon + jets sample. First, there can be no explicit requirement on MET, which is purely instrumental in photon + jets events. Rather, the vector sum of event MET and photon p_T is computed; the transverse component of this "MET-like" quantity is directly comparable to the MET in $Z \to \nu\bar{\nu}$ events. The QCD-suppression cuts on the angles between MET and jets are also defined using the azimuthal angle of this "MET-like" quantity.

In Table 1 the numbers of signal and QCD background events are shown for various stages in the event selection.

	Number of signal events per $100\,\text{pb}^{-1}$	Number of QCD events per $100\,\text{pb}^{-1}$	S/B	$S/\sqrt{S+B}$
1 iso. photon $E_T > 100\,\text{GeV}$	55000	26000	2.1	193
number of jets ≥ 3	5400	865	6.1	68
leading jet cuts	1200	94	13	33
second-leading jet cuts	791	57	14	27
QCD angular cuts	580	46	13	23
$H_T > 500\,\text{GeV}$ + MET-like >200	124	2.7	46	11

Table 1: Signal and QCD background events expected for 100 pb^{-1} , at various steps in the selection. The final five steps are designed for the hadronic SUSY search (jets + MET channel), and thus are not optimized for the selection of photons.

3.2 Estimation of Backgrounds to Photon+Jets

The main backgrounds to prompt photons are photons from mesons produced in QCD jets, and electrons for which no electron track is reconstructed. The background to the full photon selection is estimated from $t\bar{t}$, W, Z, and QCD jet production, generated with ALPGEN [3] and fully simulated with GEANT [4]. The expected contribution of these backgrounds to the final photon sample are each less than 3%.

Secondary photon background Neutral pions and etas decaying to pairs of photons are produced copiously in hadronic jets. For pions with $E_{\mathrm{T}} > 100$ GeV, the "secondary" photon pairs are sufficiently collimated to be generally indistinguishable from prompt photons in calorimetric response. However, isolation is very effective in rejecting these secondary photons, providing about a factor 100 in rejection of secondary photons while maintaining more than 90% efficiency for prompt photons collected in the high E_{T} EM cluster trigger stream.

Background from electrons faking photons At high p_{T}, the ECAL supercluster of an electron is indistinguishable from that of a photon and so, if no track is reconstructed for an electron, the electron will be identified as a photon. Thus $t\bar{t}$, W, and Z production are all potential backgrounds. A data-driven method to estimate the electron-to-photon fake rate, as well as the corresponding contamination of the photon selection, are presented in Section 3.2.2.

3.2.1 Data Driven Estimation of the QCD background to Photon+jets

To confirm the expectation of a small secondary-photon background to the photon + jets selection, a data-driven background estimate is foreseen that exploits the production of conversion electrons in the tracker. Single high-E_{T} pions are reconstructed as single isolated photons instead of photon pairs because the pion decay products are highly collimated. However, to good approximation each secondary photon has the same probability to convert in the tracker as a single prompt photon. Furthermore, the individual photons from a pion decay do not carry all of the pion momentum. When a conversion occurs, the momenta of the two conversion tracks will tend to sum to an energy that is much lower than that of the clustered ECAL supercluster energy which captures both photons. On the other hand, for conversions of single photons the summed p_{T} of the two tracks is expected to be consistent with the E_{T} of the ECAL supercluster, up to measurement uncertainties, which are dominated by the p_{T} resolution of the conversion tracks.

Fig. 4 shows the summed p_{T} of the tracks belonging to conversions [5] divided by the ECAL supercluster E_{T} of the photon candidate that seeded the track reconstruction, for prompt and secondary photons when the ECAL cluster has E_{T} of at least 100 GeV. The differences seen in the distributions for prompt and secondary photons can can be used to estimate the background contribution from real data. The first step would be to obtain a template distribution for secondary photons alone. This is achieved using photon candidates obtained with an inverted HCAL isolation cut. A particularly simple approach to estimating the background in the final sample would then be to look at the number of events in the region below 0.5 and assume this is all background and use this together with the background template distribution to estimate the total contribution of seconday photons. The background is over-estimated by this simple approach, since the region below 0.5 also contains signal. The over-estimation of the background fraction increases as the signal to background ratio increases.

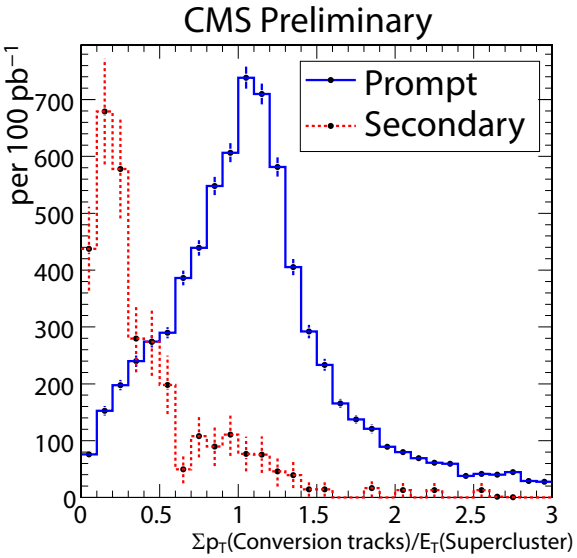

Figure 4: Summed p_T of the tracks belonging to a conversion, divided by the ECAL supercluster E_T of the photon that seeded the conversion reconstruction. Error bars correspond to Monte Carlo statistics.

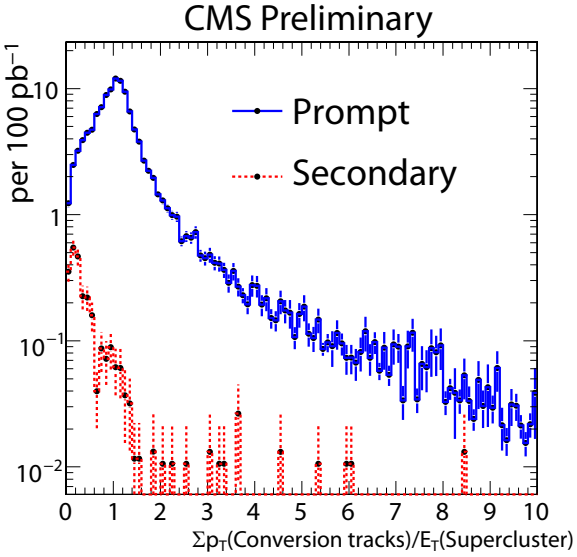

Figure 5: Expectation for the ratio of summed p_T of conversion tracks to ECAL supercluster energy, distributions for 100 pb^{-1} of data, after applying full event selections. The histograms are extracted from all events passing the initial photon selection, as in figure 4, and then scaled to the expectation for events passing the final selection (see table 1). Error bars correspond to Monte Carlo statistics.

As previously noted, MC studies indicate that secondary photons will be a very small fraction of the total number of events passing the photon + jets selection. Therefore, to avoid a substantial over-estimate of the background, the secondary-normalization procedure described above is applied to data at an earlier stage in the event selection where the signal to background ratio is much lower. In particular, the E_T (γ) $>$ 100GeV stage of the event selection as seen in Figure 4 is chosen and the impacts of the remaining selection criteria are followed through with MC datasets. The estimation method assigns a systematic uncertainty of 100% to this QCD background, which however corresponds to a systematic uncertainty of only 1.5% on the number of primary photons. As an illustration, Figure 5 contains the same distributions as Figure 4, but with relative contributions scaled to MC expectations for the full set of event selection criteria. The MC predicted abundance of signal in the normalization region below 0.5 makes it clear that one cannot normalize the background at this stage without also assuming a hypothesis for the prompt photon shape.

3.2.2 Data Driven Estimate of Background from Electrons Faking Photons

In cases where track reconstruction fails, or there is an early occurrence of hard bremsstrahlung, sufficiently energetic and isolated electrons such as those from decays of primary W or Z bosons have a high probability to be misidentified as photons. This constitutes a background to the photon + jets selection that is similar in size to the estimated QCD background.

A tag and probe method applied to $Z \rightarrow e^+e^-$ is used to estimate the probability for an electron to fake a photon. Electrons (photons) are selected if they pass the same pseudorapidity and isolation criteria used for the photons in the photon + jets selection described above. Events

with candidate ee (eγ) pairs are selected if and only if the pair with the highest Σp_T have a tag electron passing tight electron identification criteria. The pair must also have invariant mass within 10% of the Z mass.

The ee events selected by the above procedure are very clean, with less than a fraction of a percent background in the $p_T(\text{probe}) > 100$ GeV range. The eγ pair selection is the more delicate measurement, but already has background less than 7% in the region of interest, with the above cuts. Moreover, the largest class of backgrounds, where both the electron and photon are real, can be estimated by modeling them with (tag μ)-(probe γ) pairs in $Z \rightarrow \mu\mu$ data. This number, which involves a correction factor for relative electron-to-muon selection efficiencies, is used to set limits on the background systematics of the measurement.

The results are shown in Fig. 6 in comparison to higher statistics MC studies. The per-electron fake rate is about 7% in the barrel and 14% in the endcaps.

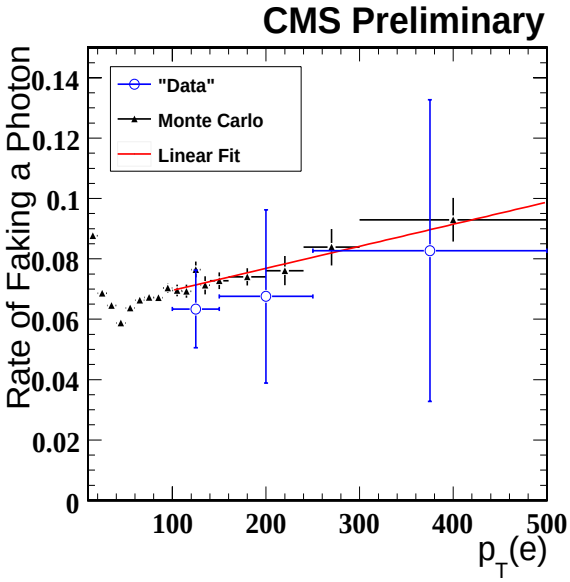

Figure 6: The per-electron rates to fake a photon. The "data" error bars shown are binomial errors appropriate for the 100 pb^{-1} data scenario. The points in the high-statistics "Monte Carlo" data fit well to a straight line.

The rate for electrons to fake photons is employed to obtain an estimate of the electron background to the photon + jets selection. A by-product of this method is an estimate for the number of photon + jets events that are lost because they are reconstructed as electrons.

3.3 Estimation of the Photon Selection Efficiency

The photon selection depends on photon isolation to select a prompt photon signal sample. To estimate the photon isolation efficiency with the data, however, no pure sample of prompt photons can be selected at the startup of the LHC without using isolation requirements. As an alternative, a tag-and-probe method that once again uses electron pairs from Z-boson decays to mimic photons is employed. The differences between electrons and photons traversing the CMS tracker will induce some differences in isolation properties for both objects. In an attempt to approximate the isolation properties of photons, the subclasses of "golden" and "narrow" electrons [6] are studied in addition to all reconstructed electrons.

Events are selected that pass the high-EM trigger with a relaxed isolation requirement, and

	$\epsilon(p_T = 80\,\mathrm{GeV}/c)$	$\epsilon(p_T = 500\,\mathrm{GeV}/c)$
All electrons	$0.884\ ^{+0.010}_{-0.010}$	$0.932\ ^{+0.067}_{-0.077}$
Golden+narrow electrons	$0.907\ ^{+0.016}_{-0.016}$	$0.945\ ^{+0.068}_{-0.078}$

Table 2: Expected central values and statistical uncertainties for efficiency measurements on data.

that contain one reconstructed and isolated electron (the tag) with $p_T > 20\,\mathrm{GeV}/c$, and another electron (the probe) with $p_T > 80\,\mathrm{GeV}/c$. In addition to these cuts, backgrounds are further suppressed by requiring $|m_{Z^{\mathrm{reco}}} - m_Z| < 9\,\mathrm{GeV}/c^2$ and $p_{T,Z^{\mathrm{reco}}} > 50\,\mathrm{GeV}/c$, where the Z boson is reconstructed from the two electrons with the largest p_T. This selection suppresses the backgrounds to a negligible level.

To estimate the photon isolation efficiency, tracker and relative calorimeter isolation requirements are applied to the probe electrons. The efficiency from golden+narrow electrons is observed to indeed be closer to the actual expectation from prompt photons than the all-electrons case. It is also observed that the p_T dependence of the isolation efficiency is found to be quasi linear in the considered range $80\,\mathrm{GeV}/c < p_T < 500\,\mathrm{GeV}/c$. This linearity was found to hold true in a wide variety of different isolation criteria, showing potential robustness for the actual behaviour of the isolation in data.

The proposed method for the measurement of the photon isolation efficiency with the data starts from a linear fit to the isolation efficiency for all the probe electrons. Correcting with a scale factor between the two electron classes derived from MC simulations, the fit results are then used to constrain the efficiency fit to the golden+narrow electrons, leaving only the overall scale free to be set. The uncertainties on the efficiency estimations are shown in Fig. 7 in the case of a data sample with an integrated luminosity of $100\,\mathrm{pb}^{-1}$. Using pseudo-experiments the expected statistical uncertainty from the linear fit was estimated. In Table 2 the results from this estimation are summarized for both all-electron and golden+narrow cases.

The measurement of the efficiency on the golden+narrow electron sample is further corrected for the remaining bias with respect to the prompt photon efficiency from simulation. This bias can then be taken as an estimate of the systematic uncertainty as a function of p_T, making this uncertainty dependent on the actual performance of the isolation in data. Alternatively the difference between the all-electrons and the golden+narrow measurements can be assigned as systematic uncertainty, which was checked to be robust against large variations in isolation requirements. The maximum of the two estimates is taken as the final systematic uncertainty.

The expected performance of the method is summarized in Fig. 8. The statistical, systematic and total uncertainties are shown, along with a linear fit to the bias from the golden+narrow measurement with respect to the prompt photon expectation.

3.4 Estimation of $Z \longrightarrow \nu\bar{\nu}$ MET Spectrum from Photons

Once the photon p_T spectrum is measured for events passing the SUSY MET + jets selection criteria, the transverse component of the vector sum of photon E_T and event calorimeter MET is computed, and this "MET-like" quantity is corrected for the photon isolation efficiency and the $Z \to \nu\bar{\nu}$ branching ratio. In Fig. 9 this spectrum is compared to the actual MET spectrum in MC $Z \to \nu\bar{\nu}$ events. In the high MET region, the residual differences in the two distributions are predominantly due to the differences in electroweak couplings of photons and Z's to up-

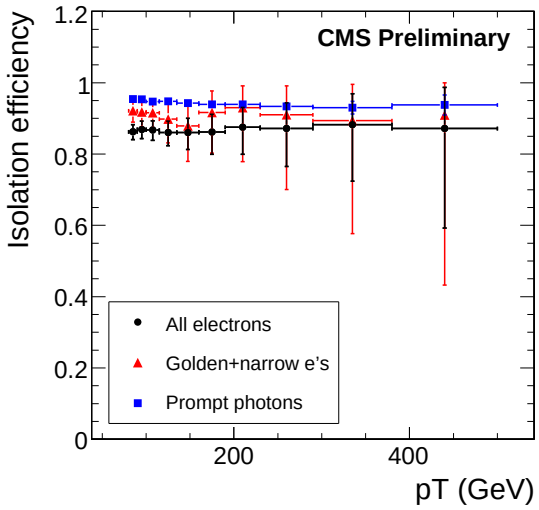

Figure 7: Photon and electron isolation efficiency. Uncertainties correspond to $100\,\mathrm{pb}^{-1}$.

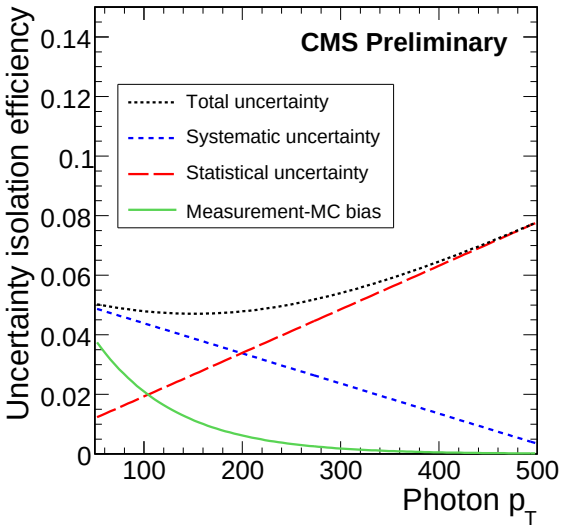

Figure 8: Summary of the expected uncertainties for an integrated luminosity of $100\,\mathrm{pb}^{-1}$. Also the bias between the efficiency on golden+narrow electrons and the actual prompt photon expectation is shown.

like and down-like quarks at high p_T. At low MET there are also differences resulting from the masslessness of the photon relative to the large mass of the Z. As noted earlier, theoretical uncertainties associated with Q^2 scale, parton distribution functions, and collinear photons are not included and will be presented in a future version of this note. Preliminary studies indicate that these effects contribute uncertainties comparable to the systematic uncertainties presented here. Table 6 lists the MC-truth estimate of $Z \to \nu\bar{\nu}$ events passing the model SUSY selection, the estimate obtained using photon events, and corresponding statistical and systematic errors.

To demonstrate how one might correct for the residual differences associated with couplings to quarks, a final correction was calculated via the ratio of Z plus 3 parton to photon plus 3 parton generator level events obtained with ALPGEN. The ratio is flat at high E_T, as expected. This correction, along with all other corrections, is applied to fully simulated photon plus jet events selected with a subset of the full selection criteria. In particular, only the first 3 jet requirements are applied in order to be better aligned with the correction which is derived from MC samples with 3 additional partons and no explicit or MET-related cuts.

The resulting spectrum is found to be in excellent agreement with that of the invisible Z events in the MET region above 200 GeV as seen in Fig. 10. For this exercise, all the corrections were evaluated and applied in the barrel and endcap separately.

MC-truth $Z \to \nu\bar{\nu}$ background	35
estimate of $Z \to \nu\bar{\nu}$ background	29
statistical uncertainty	2.7
systematic photon isolation	4.8
systematic QCD background	0.75
systematic electron background	1.5

Table 3: MC-truth $Z \to \nu\bar{\nu}$ background, estimate provided by data-driven methods, and associated errors

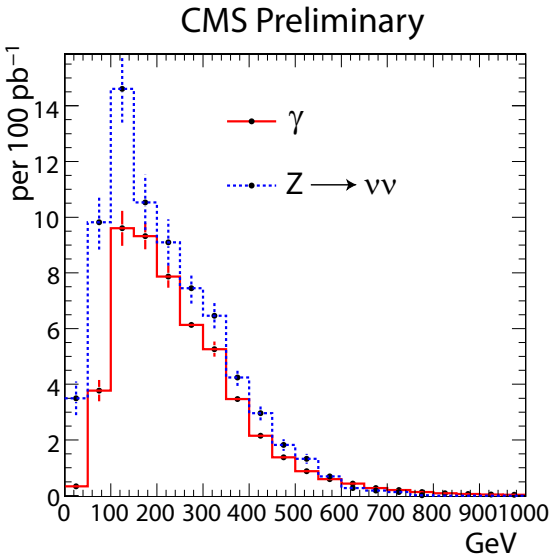

Figure 9: The spectrum of the transverse component of MET and photon E_{T}, corrected for photon isolation efficiency and scaled according to the $Z \to \nu\bar{\nu}$ branching ratio, along with the quantity it is intended to predict, the MET spectrum in $Z \to \nu\bar{\nu}$ events. The residual difference in the region of MET > 200 GeV is largely due to differences in the couplings to quarks.

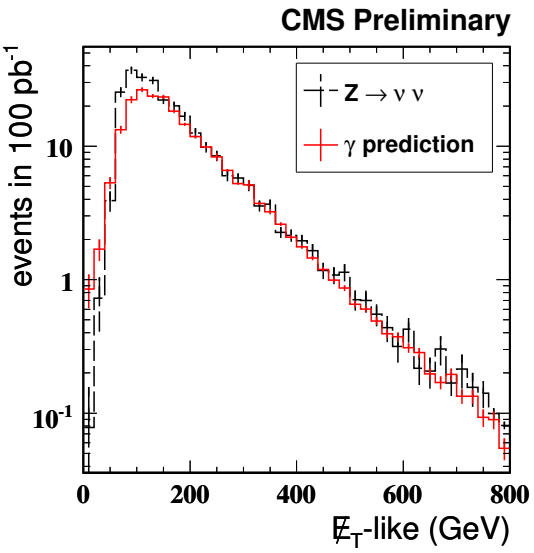

Figure 10: Demonstration of the inclusion of theoretical corrections obtained from generator level MC samples, and applied to full simulation samples for which a subset of the full selection is applied as described in the text. The photon spectrum is corrected for photon selection efficiencies, then scaled down to the $Z \to \nu\bar{\nu}$ branching ratio, and finally corrected for the theoretical differences with $Z \to \nu\bar{\nu}$ events as extracted from an ALPGEN generator-level-only calculation of the relative cross-sections of photon + 3 partons and Z + 3 partons. All corrections are evaluated and applied in the barrel and endcap separately.

4 W+Jets Channel

Another estimation of the $Z \to \nu\bar{\nu}$ background to the MET plus jets search can be obtained from the $W \to \mu\nu$ sample. In performing this study the event selection has again been kept as close as possible to the one adopted for the MET plus jets search. The selection differs in the following points:

- Events are required to pass the Level-1 single muon trigger selection with a p_{T} threshold of 7 GeV/c and the HLT relaxed single-muon trigger with a threshold of $p_{\mathrm{T}} > 16$ GeV/c. The W+jets study is performed with this trigger and not using the SUSY trigger stream (HT+MET) since the QCD control sample is unlikely to pass the HT+MET trigger with sufficient rate. Nevertheless the HT+MET trigger can be evaluated directly on the selected events because the muon will not deposit sufficient energy in the calorimeter to be "seen" by the HT/MET algorithms.
- Events are required to have one and only one isolated muon with $p_{\mathrm{T}} > 20$ GeV/c. The muon isolation variable is defined as $\mu_{Iso} = \sum_{trk}^{\Delta R} p_T^{trk}$ (sum of the transverse momenta of the tracks in a cone of ΔR=0.3, excluding the muon track itself). A muon is considered isolated when $\mu_{\mathrm{Iso}} < 3$ GeV/c.

- Jets with an isolated muon within a cone of radius $\Delta R = 0.5$ are discarded. The jet is removed because it is not really a jet, but rather a muon whose deposit was clustered as a jet by the jet-clustering algorithm. About 1.7 % of jets above 30 GeV/c are removed by this procedure.
- The p_T of the W boson is reconstructed using uncorrected MET. This quantity will again be referred to as "MET-like". The signal region is then defined by the MET-like $>$ 200 GeV requirement.
- In order to reduce $t\bar{t}$ contamination, a b-veto requirement is applied as explained below.

Irreducible backgrounds considered are those containing one real muon. These include $W \rightarrow \tau\nu$, $t\bar{t}$, single-top, and multiboson production. Relative normalizations of $W \rightarrow \tau\nu$ and $Z \rightarrow \mu\mu$ to $W \rightarrow \mu\nu$ are estimated from MC. The impact of the di-boson and single-top backgrounds was found to be negligible. For $t\bar{t}$, a data driven technique is employed. Multi-jet events in which a muon arises from meson decay within or near to a jet are also considered, and a data driven approach is used to evaluate the event yield after full selection.

Table 4 shows the yields for the $W \rightarrow \mu\nu$ signal and backgrounds at different stages of the event selection.

cut level	Sig	QCD	$t\bar{t}$	$Z \rightarrow \mu\mu + W \rightarrow \tau\nu$
L1+HLT1MuonNonIso, 1 muon $p_T > 20$ GeV/c, $>=$ 3 jets, MET-like $>$ 200 GeV	102	5	88	17
+ b-veto	87	5	43	17
+ leading jet and next leading jet cuts	40	1	18	9
+ QCD angular cuts	29	0	14	9
+ H_T	24	0	13	9

Table 4: Signal and background events expected for 100 pb^{-1}, for various steps of the cuts

After applying the W-boson selection on the signal and background samples, a sizable fraction of $t\bar{t}$ events having genuine missing energy pass the selection. This is by far the dominant background. The data-driven technique that was developed for $t\bar{t}$ relies on the presence of two b-jets in every $t\bar{t}$ event. Among the b-jet identification algorithms (b-taggers) available in CMS, the track probability algorithm [7] proves to be the most robust on simulated events with regard to the misidentification efficiency of high-energy jets. The tagger is applied at two different working points, one with high tagging efficiency and higher mis-tag rate, one with lower tagging efficiency and low mis-tag rate. The two working points will be referred to as "medium" and "tight", respectively. By requiring at least one tight tag, it is possible to obtain an enriched $t\bar{t}$ sample. Alternatively, an enriched W+jets sample is selected by vetoing the presence of at least one medium b-tagged jet. One then obtains the $t\bar{t}$ MET-like spectrum from the $t\bar{t}$-enriched selection, and correct it for the efficiencies and acceptances for the medium-b-veto relative to the tight-b-tag to obtain an estimate of the residual $t\bar{t}$ contamination in the signal region. With the selection described above, 12 events are observed in the $t\bar{t}$-enriched sample with MET-like quantity above 200 GeV for 100 pb^{-1} of integrated luminosity. The corresponding statistical uncertainty is 29%. Considering a 15% uncertainty on the tagging efficiency at 10 pb^{-1} [8], a combined statistical and systematic uncertainty of 37% on the $t\bar{t}$ yield is expected.

Multi-jet QCD events are generally expected to not pass the MET-like $>$ 200 GeV cut, while they dominate the distribution in the region MET-like $<$ 200 GeV. Nevertheless, a limit on this process needs to be derived from data. The MET-like distribution of the multi-jet background is modeled using a QCD-enriched subset of events where the muon is required to be less isolated;

$10 < \mu_{Iso} < 15$ GeV/c. The shape of the MET-like distribution is found to be independent of the isolation criteria applied to the muon. The other criteria are left unchanged, except the bVeto, leading jet, and next-to-leading jet criteria, which were removed in order to maintain sufficient statistics. The QCD distribution is normalized in the region between 0 and 100 GeV after subtracting the W events in this region. The contribution of the latter is estimated using $Z \rightarrow \mu\mu$ events, after correcting for relative efficiencies, acceptances, cross sections and branching fractions. Given the shape of the QCD distribution from the enriched multi-jet sample and the normalization in the low MET-like region, only 1 QCD event is predicted to survive in the MET-like $>$ 200 GeV region. QCD thus appears to have negligible contribution, and is expected to be further decreased after applying the leading jet and next-to leading jet selections. The size of this decrease is not known. If one were to assume that no QCD events make it into the final sample, this could result in an underestimation of the W plus jets cross-section, and hence also that of the invisible Z. In view of this, the prediction of 1 QCD event at this intermediate stage is used as a systematic uncertainty on the QCD contamination.

Beyond Standard Model physics is often characterized by multi-jet events. When a real muon is also present, such events may pass the W + jets selection. Here the CMS SUSY Low Mass (LM) benchmark points are considered as examples of potential new physics backgrounds to W+jets. MC simulations show that the LM points and W+jets have comparable numbers of events in the high MET-like tail, after all selection criteria are applied. Even if these additional events are assumed to be W+jets, a conservative upper bound on invisible Z plus jets can still be obtained.

Finally, in order to convert from the measured W spectrum to the MET spectrum of $Z \rightarrow \nu\bar{\nu}$, various differences need to be taken into account.

- The muon kinematic and geometrical acceptance ($p_{\mathrm{T}} > 20$ GeV/c and $|\eta| < 2.1$) is estimated to be 75% at the generator level in events selected with W boson $p_{\mathrm{T}} > 200$ GeV/c and three jets with $p_{\mathrm{T}} > 60$ GeV/c and $|\eta| < 3$ (to match the requirement used in this analysis of 3 jets with raw $p_{\mathrm{T}} > 30$ GeV/c). The uncertainty associated with the scale dependance and PDF variations is estimated to be 2%.
- A correction for the muon isolation efficiency must be applied. The muon isolation efficiency will be measured from data using Z+jets events with a tag and probe method. In MC studies we estimate that this isolation criterion keeps 85% of signal events in an environment with at least 3 jets. The expected uncertainty at 100 pb^{-1} is 1.8 %.
- A correction for the single muon trigger efficiency will be measured with a tag and probe method, where the tag muon is isolated and selected by Level1 and HLT trigger criteria and the efficiency is the fraction of events in which the second muon (probe) meets the trigger requirement. The efficiency for the single non-isolated muon trigger in the fiducial region $|\eta| < 2.1$ and for $p_{\mathrm{T}} > 20$ GeV/c is 85 % with respect to the offline selection criteria and will be known with 2 % uncertainty [9].
- A correction for the relative cross sections and branching ratios of W and Z due to the differences in their masses and couplings to quarks is also required. This amounts to a correction factor of 1.5. The uncertainty associated to the scale dependance and the PDF is within 3 %.
- A correction to account for the b-veto efficiency (50%) is taken into account. We consider 4% as mistag efficiency with a relative uncertainty of 10%.

After applying the above corrections, an overall correction factor of 1.45 is applied to the 24 W events predicted in Fig. 11, giving an estimate of 35 $Z \to \nu\bar{\nu}$ events. All uncertainties are summarized in table 5 and include a statistical uncertainty of 10 events, an experimental systematic uncertainty of 8.25 events, and a theoretical uncertainty of 3 events.

Estimated number of $Z \to \nu\bar{\nu}$	35
statistical uncertainty	10
muon trigger and isolation efficiency	4.1
bVeto efficiency on W	0.35
$t\bar{t}$ systematics uncertainty	7.01
QCD systematics	1.45
muon acceptance	2.4
$Z \to \mu\mu$	0.09
W/Z ratio	0.4

Table 5: Errors associated to the $Z \to \nu\bar{\nu}$ background based on $W \to \mu\nu$ +jets measurement

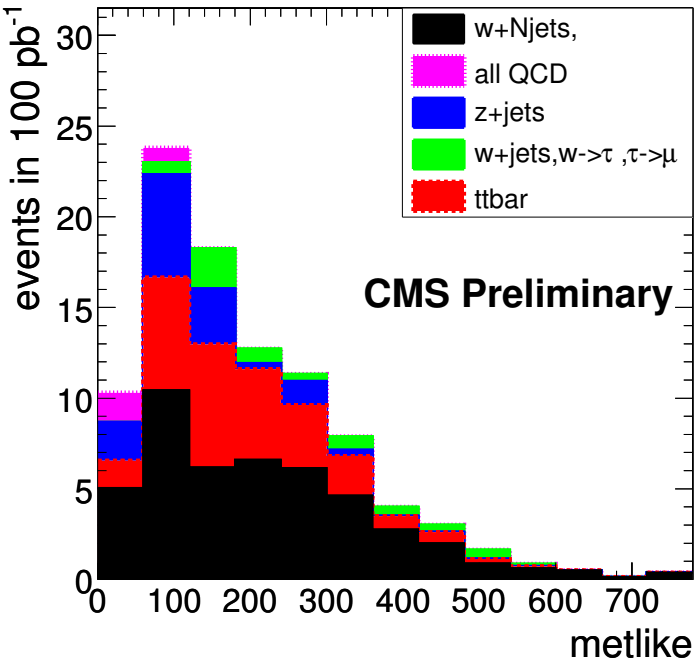

Figure 11: Expected MET-like spectrum for a 100 pb^{-1} luminosity scenario after all the selections are applied, as in table 4.

5 Conclusions

For early LHC data-taking, such as the first 100 pb^{-1}, it has been shown in this note that the photon and W plus jets samples provide independent sources for the estimation of the "invisible Z" background. Table 6 compares the MC truth number of $Z \to \nu\bar{\nu}$ + jets events, to the estimates obtained using γ, or W, plus jets events. At the moment, some theoretical uncertainties for the photon plus jets sample have not been considered. These are under study along with a complete prescription for how to minimize these uncertainties in a purely data-driven way by comparing events in control regions.

$Z \to \nu\nu$ background estimate (100 pb^{-1})	
MC-truth	35
From γ+jets	29 ± 3 (stat) ± 5 (sys)
From W+jets	35 ± 10 (stat) ± 8 (sys) ± 3 (theory)

Table 6: Estimates for Z $\to \nu\bar{\nu}$ + jets background for 100 pb^{-1} integrated luminosity. For photons, the quoted uncertainties are for statistics (124 events in 100 pb^{-1}) and experimental systematics, respectively, and do not include theoretical uncertainties. The photon-based estimate does not include the upward correction for differences in couplings to quarks. For W's (24 events in 100 pb^{-1}) all uncertainties are included. By comparison, after the same selection criteria are applied, one expects roughly 4 $Z \to \mu\mu$ events in 100 pb^{-1}.

As an additional benefit, it has been noted that use of γ, W and Z samples provides useful information in an environment where there is not only one, but a spectrum of new particles which could be contributing to control samples.

References

[1] G. L. Bayatian *et al.* [CMS Collaboration], "CMS technical design report, volume II: Physics performance," J. Phys. G **34**, 995 (2007).

[2] J. Alwall *et al.*, "MadGraph/MadEvent v4: The New Web Generation," JHEP **0709**, 028 (2007) [arXiv:hep-ph/0706.2334].

[3] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau and A. D. Polosa, "ALPGEN, a generator for hard multiparton processes in hadronic collisions," JHEP **0307**, 001 (2003) [arXiv:hep-ph/0206293].

[4] S. Agostinelli *et al.* [GEANT4 Collaboration], "GEANT4: A simulation toolkit," Nucl. Instrum. Meth. A **506**, 250 (2003).

[5] N. Marinelli, "Track finding and identification of converted photons", CMS-NOTE-2006-005.

[6] S. Baffioni *et al.*, "Electron reconstruction in CMS," Eur. Phys. J. C **49**, 1099 (2007).

[7] G. L. Bayatian *et al.* [CMS Collaboration], "CMS physics: Technical design report, volume I"

[8] C.Gerber *et al.* "Performance measurement of b-tagging algorithms using data containing muons within jets", CERN-CMS-PAS-BTV-07-001.

[9] J.Alcarez *et al.* "Towards a measurement of the inclusive W $\rightarrow \mu\nu$ and Z $\rightarrow \mu\mu$ cross section in pp collisions at $\sqrt{s} = 14$ TeV", CERN-CMS-PAS-EWK-07-002.