CMS Physics Analysis Summary

Contact: cms-pag-conveners-exotica@cern.ch 2009/02/02

Search for high mass resonance production decaying into an electron pair in the CMS experiment

The CMS Collaboration

Abstract

This note describes the potential of the CMS experiment for the discovery of high mass resonances decaying into an electron pair, with an integrated luminosity of 100 pb^{-1}. Event selection and high energy electron reconstruction and identification criteria are explained. Using pseudo-experiments, the expected dielectron invariant mass spectrum, in the presence of a new, heavy resonance as well as Drell-Yan production and other backgrounds, is reconstructed. The data are used for checking and determining the selection efficiencies and the backgrounds using the Z peak region and the "control" mass region which is dominated by high mass Drell-Yan production. The backgrounds in the new physics signal region are discussed, and the significance of potential signals are extracted. Finally, the CMS potential is quantified, both for a discovery and by giving upper limits on potential numbers of events from new physics in the absence of an observable signal.

1 Introduction

This note describes procedures that will be used to search for evidence of new resonant structures in the e^+e^- invariant mass spectrum using the CMS detector. The methods described have been evaluated using Monte Carlo simulations assuming an availability of 100 pb^{-1} of integrated luminosity and 7 TeV proton beams. The discovery potential, using representative new physics models, and the potential mass limits in the absence of an observable signal are given.

The procedure uses a part of the Drell-Yan (DY) mass spectrum where no evidence of new physics is expected to be observed as a control region. This is used to determine a differential cross section and show it is consistent with Standard Model (SM) predictions. Having demonstrated that the detector performance and reconstruction procedures are well understood, a search can be made for deviations from SM predictions in the new physics search region.

The approach followed here is to use the data as far as possible to determine the level of non-DY background and all triggering and reconstruction efficiencies. In most cases more than one method has been devised to determine these quantities. In addition, a set of triggers has been designed to ensure that, even in the case of non-optimal detector performance or unexpectedly large trigger rates, the signal data will still be recorded. This strategy also maximizes the overall trigger efficiency. Using the single particle trigger with an E_t threshold of 18 GeV, all of the methods described in this note are accessible, enabling direct cross checks. This trigger explicitly requires that a track be matched to an ECAL cluster and places the strongest requirements on detector performance. It is complemented by two additional robust triggers at E_t thresholds of 80 and 200 GeV where no track matching requirement is made.

In order to check the detector response to high energy electrons, two complete analysis paths are demonstrated in this paper; the first utilizes the large statistics available in the Z resonance to extract efficiency information and allows for consistency checks using the properties of the Z peak in addition to the DY tail. The second method requires that all procedures only make use of information accessible directly at high energy, using the data in an invariant mass spectrum above 200 GeV/c^2. This method therefore does not depend critically on the trigger with an 18 GeV threshold and demonstrates that if the trigger with an 80 GeV E_t threshold, where no track cluster matching requirements are made, is the only one available, the analysis can still be performed.

A brief description of the expectations for new heavy resonance production and for the DY process is given in Section 2. The event selection and the determination of the efficiency with which these events are selected are described in Section 3. In Section 4, the backgrounds and methods for measuring them using the data are described. The SM prediction for the reconstructed e^+e^- spectrum is given in Section 5, where two analyses for determining this spectrum are described. The first utilizes the e^+e^- spectrum from the Z peak up to high masses and the second the DY distribution for $M > 200\,\mathrm{GeV}/c^2$; measures of how well the reconstructed spectra agree with SM expectations are given. Section 6 describes the CMS discovery potential.

2 Model predictions

The potential of the CMS experiment for the discovery of representative sets of massive Z' gauge bosons (spin 1) [1] and of massive gravitons G (spin 2) that arise in the extra-dimension model of Randall and Sundrum (RS) [2], in the channels $Z' \to e^+e^-$ and $G \to e^+e^-$, was studied in ref. [3] for an integrated luminosity of 30 fb^{-1}, with a well understood detector.

Predictions of expected numbers of events for an integrated luminosity of 100 pb^{-1} are given in Table 1, for several resonances with masses between 1.0 and 2.5 TeV/c^2 and for the DY background. Heavy Z' bosons are produced through $q\bar{q} \to Z'$ annihilation and heavy gravitons through the $q\bar{q} \to G$ and $gg \to G$ processes [1]. The resonance masses are taken above the current Tevatron limits, which are of the order of 700-900 GeV/c^2 (depending on the model) for Z' bosons [5] and 900 (300) GeV/c^2 for $c = 0.1$ (0.01) RS gravitons [5, 6].

Model mass (GeV/c^2)	$M = 1000$	$M = 1500$	$M = 2000$	$M = 2500$
SSM Z'				
$\sigma \cdot$ BR (fb)	458	80	20	5.8
nb. ev. for 100 pb^{-1} and 2 el. with $\|\eta\| < 2.5$	38	7.2	1.8	0.54
RS G ($c = 0.1$)				
$\sigma \cdot$ BR (fb)	660	76	14	3.5
nb. ev. for 100 pb^{-1} and 2 el. with $\|\eta\| < 2.5$	62	7.2	1.3	0.32
DY bg. (GeV/c^2)	$M > 600$	$M > 1100$	$M > 1600$	$M > 2100$
cross section (fb)	50	4.4	0.76	0.18
nb. ev. for 100 pb^{-1} and 2 el. with $\|\eta\| < 2.5$	4.0	0.4	0.07	0.02

Table 1: Cross sections (at LO) times e^+e^- decay branching ratio for several SSM Z' bosons and RS extra-dimension gravitons G with coupling $c = 0.1$, and expected numbers of events with the pair of electrons emitted in the tracking detector acceptance, for a luminosity of 100 pb^{-1}; numbers of background DY events in the corresponding mass windows.

Table 2 shows the DY cross sections for e^+e^- invariant masses $M > 40$, 200 and 500 GeV/c^2 and the expected numbers of events for several cuts on the electron E_t. In tables 1 and 2, leading order (LO) cross sections are given.

Drell-Yan events (GeV/c^2)	$M > 40$	$M > 200$	$M > 500$
nb. events for 100 pb^{-1}	179,500	252	9.9
AND 2 electrons with $\|\eta\| < 2.5$	81,475	148	7.5
AND $E_t(1,2) > 30$ GeV	50,475	147	7.5
AND $E_t(1,2) > 80$ GeV	215	110	7.4

Table 2: Numbers of Drell-Yan events (LO cross sections) with $M > 40, 200$ and 500 GeV/c^2, for a luminosity of 100 pb^{-1} and $|\eta|$ and p_t cuts.

3 Electron identification and event selection

This section describes the triggering algorithms, the electron identification criteria, final event selection criteria, the efficiency of the algorithms and the determination of these efficiencies from the data. The electron identification criteria described here are used everywhere in this paper.

[1] The SSM Z', RS G and DY processes were generated at LO with the PYTHIA program [4] using the CTEQ5L parton distribution functions. EW and QCD initial and final state radiations off incoming partons and outgoing leptons were included. The full $\gamma^*/Z^0/Z'$ interference structure was taken into account.

3.1 Triggers

The Level 1 trigger for electrons requires a minimum energy deposit in the electromagnetic (em) ECAL calorimeter, with additional isolation criteria for em objects with $E_t < 64$ GeV. The L1 trigger efficiency for high E_t electrons is found to be close to 100%, using the L1 emulator.

For the high level trigger (HLT), three trigger paths have been designed. The Relaxed Single Electron trigger requires an energy deposit where $E_t > 18$ GeV in the ECAL calorimeter, direction and energy matching between the ECAL deposit and a track in the tracking detector, and imposes isolation requirements in the ECAL, the HCAL hadronic calorimeter and the tracker. The High E_t trigger requires an ECAL deposit where $E_t > 80$ GeV and loose isolation requirements in the HCAL and the tracker; no track matching the ECAL energy deposit is required, which makes this trigger very robust against detector alignment and calibration problems. The Very High E_t trigger only requires an ECAL energy deposit where $E_t > 200$ GeV. No other conditions are imposed. More stringent selection criteria are imposed for the trigger with a lower E_t threshold in order to reject the potentially large background from QCD events where a jet fakes an electron. The event rates for the three triggers are estimated to be 9.1, 0.8 and 0.14 Hz, respectively, for a luminosity of 10^{32} cm^{-2}s^{-1}.

The global efficiency of the L1 trigger and of the combined HLT triggers for events containing a pair of electrons emitted within the tracker acceptance ($|\eta| < 2.5$) is found to be 94% for DY production with mass $M > 200$ GeV$/c^2$, 98% for DY with $M > 500$ GeV$/c^2$, and 99% for SSM Z' bosons with $M = 1$ or $M = 4$ TeV$/c^2$.

3.2 Electron identification and efficiencies

The electron identification criteria have been designed to ensure high efficiencies for high energy electrons, whilst efficiently rejecting the QCD background. They are required to be insensitive to the quality of the detector calibration and therefore useful under startup conditions.

High Energy (HE) electrons are selected within the tracker acceptance ($|\eta| < 2.5$), using the following criteria.

- **Electron candidates** Electron tracks are reconstructed offline from initial track segments in the inner pixel detector that are required to be compatible with the ECAL energy deposit (named "supercluster" - SC) in direction and p_t; loose preselection cuts are imposed on the SC shape and the energy ratio H/E is required to be less than 0.2, H/E being computed as the energy deposit in the HCAL within a cone of radius $\Delta R < 0.1$ centred on the electron SC, divided by the ECAL SC energy.
- **ECAL fiducial region** In the ECAL barrel, electrons are required to lie within areas completely covered by ECAL modules, rejecting, in a conservative way, the inter-module separation regions.
- **Shower shape and track matching criteria** Cuts are imposed on the spread of the electron shower in η and on the matching of the direction of the track, measured in the inner tracker layer, with the ECAL SC position (the matching parameters being scaled according to the resolution).
- **Isolation criteria in the ECAL and HCAL calorimeters and in the tracker** The ECAL isolation implies a cut on the sum of the E_t deposits measured in all ECAL cells contained within a cone of radius $\Delta R = 0.3$ around the SC barycentre, except for the cells belonging to the electron candidate SC. The HCAL isolation cut applies on the sum of the E_t deposits in all HCAL cells contained between two cones of radii $\Delta R = 0.1$ and 0.3 around the SC direction; for $\eta > 1.560$ (endcap region)

the longitudinal segmentation of the HCAL is also used. For the tracker isolation, cuts are applied on the sum of the p_t of tracks with $p_t > 1.5$ GeV/c, reconstructed between two cones of radii $\Delta R = 0.02$ and 0.2 around the electron track.

The values of the cuts on the shower shape, track matching and isolation variables are chosen differently for electrons emitted in the barrel ($|\eta| < 1.442$) and in the endcap ($1.560 < |\eta| < 2.5$) regions, because of the different detector geometries and activities in the isolation cones.

Saturation occurs in the ECAL electronics for very high energy deposits in a single ECAL crystal ($\gtrsim 1.7$ TeV for the barrel and $\gtrsim 3.0$ TeV for the endcaps). The energy in the saturated crystal can be reconstructed, with a resolution of about 7%, using the energy deposit distribution in the surrounding crystals (see [7]).

The global efficiencies of HE electron reconstruction and identification, with the alignment and calibration conditions corresponding to an integrated luminosity of 100 pb^{-1}, are presented in Fig. 1 as a function of the electron E_t, separately for the ECAL barrel and endcaps. The efficiency in the barrel region includes the effect of the fiducial cuts. The decrease in the selection efficiency for electrons emitted in the endcap at low E_t is a reflection of the limited η coverage of the border of the pixel detector, which affects the reconstruction of electron candidates at large η.

The overall reconstruction and identification efficiencies of HE electrons with $E_t \approx 100(500)$ GeV, a value typical for high mass DY (resonance production), and emitted toward the ECAL barrel or endcaps and within the tracker acceptance are 80(80)% and 82(85)%, respectively. Note that, for barrel electrons, this efficiency includes a 9% loss due to the ECAL barrel fiducial requirement.

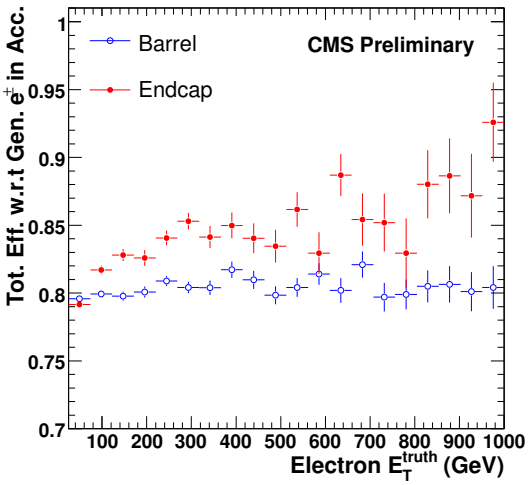

Figure 1: Global HE electron reconstruction and identification efficiencies, for the ECAL barrel ($|\eta| < 1.442$) (open points) and endcap ($1.560 < |\eta| < 2.5$) (full points) electrons as a function of the electron η.

3.3 Event selection

The final event selection sample contains events with two HE electrons passing the criteria defined in Section 3.2, emitted outside the ECAL barrel-endcap gap regions ($1.442 < |\eta| < 1.560$) and within the tracker acceptance ($|\eta| < 2.5$). No charge requirement is imposed.

In the following, where the spectrum considered includes the Z resonance, both of the electrons are required to have an E_t greater than 30 GeV. Where the spectrum considered is confined to events with an invariant mass greater than 200 GeV then both electrons are required to have an E_t greater than 80 GeV. No additional cuts are applied. The electrons are not required to be of opposite charge, and no cuts on the global event structure (number of jets, missing transverse energy, ...) are applied.

3.4 Efficiency checking using the data

The electron identification efficiency can not be directly determined at the energies in the region where searches for new physics will take place. The electron identification has therefore been designed to have an efficiency that is flat as a function of E_t. The efficiency can then be determined from the data in regions of E_t with significant statistics and, given reasonable agreement with Monte Carlo simulations, used to give the efficiency in the potential signal region.

Both electron identification and trigger efficiencies are determined from the data. Two approaches are described here, the first utilizing the large number of events at the Z resonance, but dependent upon Monte Carlo simulation information to extrapolate to the region of interest and the second using higher energy events in the DY tail utilizing events closer to the energies of interest, but having limited statistics and having to suppress sources of fake electrons. Both studies below use pseudo-experiment Monte Carlo samples corresponding to 100 pb^{-1}. The overall detector geometric acceptance is determined using Monte Carlo simulations.

The efficiencies are estimated using the so-called "tag and probe" method [8].

3.4.1 Efficiency checking using low mass to high mass extrapolation

Utilizing the large number of di-electron events in the mass region of the Z resonance, a hybrid data/MC strategy is used to estimate the signal efficiency in the high mass region. The strategy is outlined below:

- That the Monte Carlo simulation describes the CMS detector's response to electrons well, is validated. This is done by using the tag and probe method to construct "$N-1$ distributions" [2] for each criterion used in the triggering, reconstruction, and identification of the electrons. Any scale factors (SF) necessary for agreement between Monte Carlo simulation and data are determined.
- Next, the high energy behaviour of the computed SF is examined. Using data in a mass window around the Z peak allows the range to be extended approximately to 100 GeV in E_T. If the SF is flat, or approaches an asymptotic value at high E_T, a single number is extracted for the SF for each of the triggering, reconstruction, and identification efficiencies. The statistical error on each SF is dominated by the number of events in the Z peak. The significantly more complicated situation where the SF actually depends on E_T while tractable, may not be practical and the deficiencies in the simulation may need to be addressed directly.
- Given the set of SF, the efficiency derived from MC simulation of high E_T electron events can be corrected to the efficiency expected to be observed in high E_T electron data. This assertion is checked independently by computing the electron efficiency in the high mass control region ($200 < M < 800$ GeV/c^2) from the data itself as described in Section 3.4.2 and comparing it to the SF corrected MC efficiencies obtained from the simulation. The difference between these efficiencies is taken as a measure

[2] For these plots, all but the N^{th} reconstruction criteria (the one being examined) are applied.

of the systematic uncertainty with which the high E_T efficiency can be established by the simulation.

- Finally, the SF corrected MC efficiencies are obtained from the simulation for the high-mass signal region ($M > 800$ GeV/c^2) in a manner analogous to that which was used for the high-mass control region. The systematic uncertainty is taken to be that determined in the previous step giving the desired signal efficiencies.

This procedure is applied separately for the online (trigger) and offline reconstruction and detection efficiencies, with the following results: $\varepsilon^{EB}_{offline}$ ($\varepsilon^{EE}_{offline}$) $= 0.84 \pm 0.02$ (stat.) ± 0.03 (syst.) (0.83 ± 0.03 (stat.) ± 0.03 (syst.)), $\varepsilon^{EB}_{online}$ ($\varepsilon^{EE}_{online}$) $= 0.95 \pm 0.01$ (stat.) ± 0.01 (syst.) (0.97 ± 0.014 (stat.) ± 0.01 (syst.)).

3.4.2 Efficiency checking at high mass

The efficiencies here are estimated using higher energy events in the DY tail ($M > 200$ GeV/c^2). The detector related electron reconstruction and identification efficiencies are factorized as the product of two contributions, implying the two main parts of the selection: $\varepsilon_{offline} = \varepsilon_{\rm cand} \times \varepsilon_{\rm id}$, $\varepsilon_{\rm cand}$ being the efficiency for an electron to be reconstructed as an electron candidate (thus the association of a SC and a track, first condition in Section 3.2), and $\varepsilon_{\rm id}$ the efficiency for a candidate to pass the other HE identification cuts of section 3.2.

To reduce background contamination, only events with $M > 200$ GeV/c^2, both electrons having $E_t > 80$ GeV and missing $E_t < 50$ GeV are retained (the first two conditions are efficient against W+jet, γ+jet and QCD multijet backgrounds, the third against the $t\bar{t}$ and W+jet backgrounds). In addition, selection criteria tighter than for standard HE identification are imposed on the tag electron, to decrease the jet backgrounds. For 100 pb^{-1}, approximately 140 probe electrons will be available.

The $\varepsilon_{\rm cand}$ efficiency is estimated to be 0.94 ± 0.02 (stat.) $^{+0.03}_{-0.07}$ (syst.). This number is obtained after correction for the background contamination among probe electrons, estimated to be 13% using the MC simulations described in section 4. The systematic error is obtained by varying the amount of background by factors 2 and 0.5.

The $\varepsilon_{\rm id}$ efficiency is estimated to be 0.88 ± 0.02 (stat.) $^{+0.05}_{-0.04}$ (syst.). The background, dominated by the W+jet channel, is measured from the events where the two electron candidates have the same charge; the small charge mis-measurement in the DY signal ($\sim$5%) and the charge correlation between the quark and the lepton in the W+jet events are taken into account. The systematic error is obtained by varying the background contribution within its statistical uncertainty and by modifying by $\pm$30% the effect of the charge correlation.

4 Backgrounds

The dominant sources of background in the high mass DY region are: $t\bar{t}$, QCD, W+jet, γ+jet and $\gamma\gamma$. They can be divided into two classes: two real electrons from $t\bar{t}$ events, which is the dominant background, and at least one jet faking an electron in W+jet, γ+jet, $\gamma\gamma$ and QCD multijet events.

The contributions of these different backgrounds obtained using data-driven estimates are described in Sections 4.1 and 4.2. In the present analysis exercise (up to Section 6), LO cross sections are used for the signal, DY and all backgrounds, except for the $t\bar{t}$ production where the LO cross section is used with an additional K factor of 1.8.

4.1 $t\bar{t}$ background measurements from the data

Two different methods have been devised to determine the $t\bar{t}$ background from the data: the b-tag and the $e\mu$ methods. The two methods use entirely independent data samples and have largely different sources of systematic error.

4.1.1 b-tag method

To estimate the $t\bar{t}$ contribution, the HE selected sample itself is used. As top quarks decay to bW with a branching fraction close to 100%, the background from $t\bar{t}$ events to the HE selected sample is estimated by counting events with one (or two) b-tag, knowing the efficiency with which a jet originating from a b quark is reconstructed and tagged, ε_b. Monte Carlo studies show that these samples are very purely due to $t\bar{t}$ events. The value of ε_b is thus estimated from the measured ratio of the number of events with two b tags, divided by the number of events with one b tag. The ratio of the probabilities for exactly one and exactly two b quarks to be produced in the selected detector acceptance from $t\bar{t}$ production is determined from Monte Carlo simulation.

Using the events in the HE selected sample with $M > 40$ GeV/c^2, the ε_b efficiency is thus estimated to be 0.20 ± 0.09, which is in good agreement with Monte Carlo expectations. The events satisfying $70 < M < 110$ GeV/c^2 have been excluded to suppress the contamination from $Z + b\bar{b}$ events. The latter events, while they contain real b jets, are nonetheless suppressed for the b tagging performance estimation since they have different acceptance corrections.

The background distribution obtained from the b-tag method is shown in Fig. 2(a) and compared to the genuine distribution of the top background (shaded histogram). As an example, using the single tag sample, an estimate of 105 ± 50 events above a reconstructed di-electron invariant mass of 140 GeV/c^2 is found, compared to the true value of 95 events, for a luminosity of 100 pb^{-1}.

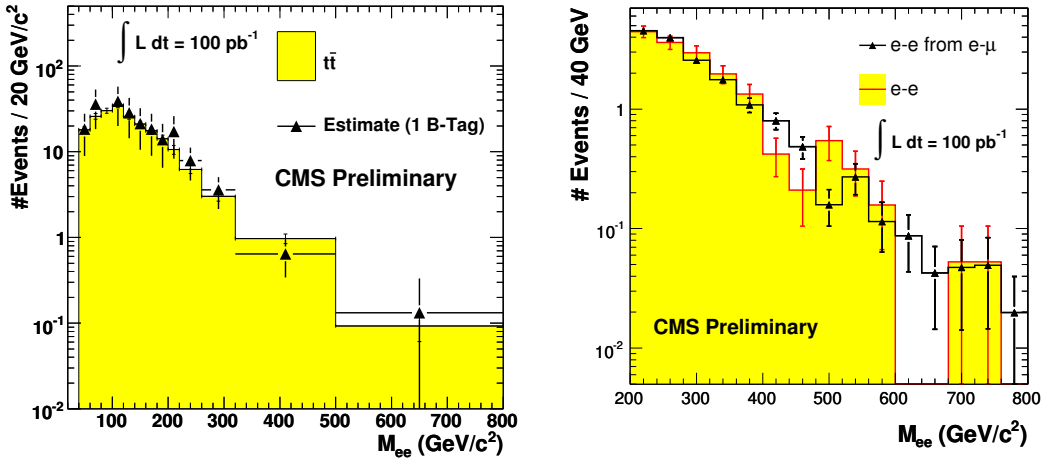

Figure 2: Estimation of the $t\bar{t}$ background to the Drell-Yan spectrum, computed from the data using (a) the b-tag method for electron $E_t > 30$ GeV and $40 < M < 800$ GeV/c^2; (b) the electron-muon method for electron $E_t > 80$ GeV and $200 < M < 800$ GeV/c^2. The triangles are the background estimates obtained from the data, the dashed histogram is the genuine distribution of the top background. The Monte Carlo statistics have been scaled to correspond to an integrated luminosity of 100 pb^{-1}.

4.1.2 $e\mu$ method

The $t\bar{t}$ background is estimated directly from data using di-lepton events with different lepton flavours, i.e. one electron and one muon ($e\mu$ events), coming from the two W boson decays. The event kinematics are identical to those of dielectron decays so that, when correcting for the differences in acceptance and selection efficiencies between muons and electrons, the number of selected $t\bar{t} \to e\mu$ events should be twice that of $t\bar{t} \to ee$ events.

The method takes into account a contribution to the $e\mu$ events coming from W+jet production, where either the muon originates from the W and the electron is a fake, or vice versa. This contribution to the $e\mu$ sample is estimated from Monte Carlo simulations to be 5%, to which a 100% systematic uncertainty is assigned.

In Fig. 2 (b), the $t\bar{t} \to ee$ spectrum determined using $e\mu$ events is compared to the true $t\bar{t} \to ee$ distribution (shaded histogram). Good agreement is observed, which confirms the validity of the method. For a luminosity of 100 pb^{-1}, the expected number of $t\bar{t} \to ee$ events with both electrons passing the HE selection criteria is 16.1 events for electron $E_t > 80$ GeV and $200 < M < 800$ GeV/c^2. This background will be determined from an expected sample of 42.5 $e\mu$ events, taking into account W+jet production, with a statistical precision of 15% and a systematic uncertainty of 5% coming mainly from the imperfect knowledge of the W+jet contamination.

4.2 Jet background - fake rate method

Backgrounds arising from jets falsely identified as electrons are due to QCD multijet events, where two jets pass the HE electron selection requirements (events with two fake electrons), W+jet events (events with a real electron from the W decay and a fake electron) and γ+jet events (events with a γ reconstructed as an electron and a jet faking an electron). From MC studies, the contributions of these backgrounds are expected to be small. Nevertheless, the production cross section of these processes have large uncertainties and a direct estimation of these background contributions from data is important, especially at startup.

The "jet fake-rate", which is the probability for a jet to pass the full HE electron selection criteria, including isolation conditions, is measured with data using jet triggered events. It is computed in an E_t bin as the number of jets passing the electron selection requirements divided by the total number of jets, and is estimated to be $0.04 \pm 0.01 (0.3 \pm 0.1)\ 10^{-3}$ in the barrel (endcap) region for a jet with E_t around 200 GeV.

To describe the invariant mass spectrum of the di-jet events faking electrons, a second sample is constructed with the events containing, in addition to a jet faking an electron, a second jet and no other electron candidate (in order to minimize DY contributions). Events in this sample are then weighted by the fake rate corresponding to the E_T of the non-faking jet. The invariant mass of the two jets is computed, with the energy of the non-faking jet corrected to the energy it would have if reconstructed as an identified electron. This provides an estimate of the number of di-jet events in each mass bin.

This method is used in the following sections to estimate the jet contributions, as shown in Figs. 3, 4 and 5. As an example, the estimated number of jet background events with $M > 200$ GeV/c^2 where both electrons have E_t above 80 GeV is measured to be 0.84 ± 0.31, i.e. a statistical precision of $\sim 40\%$. The systematic error on the method is estimated to be of the order of 50%.

5 Dielectron spectrum

In order to show that the detector performance and particle reconstruction is understood the di-electron spectrum spectrum is reconstructed including low masses where no new physics is expected. In this control region data driven estimates of the non-DY backgrounds are made allowing the DY distribution to be determined. Demonstrating that this agrees with the SM expectation constitutes the principal control test in this analysis.

Descriptions of two procedures, the first considering the reconstructed di-electron spectrum including the Z resonance peak and the second using the spectrum where $M > 200$ GeV/c^2, are given in the following sections. Both of these procedures are possible using the Relaxed Single Electron trigger, and the second is in addition possible using only the High E_t trigger.

5.1 Drell-Yan spectrum from the Z peak

In order to illustrate that the background estimates described above are adequate to perform the intended analysis with the first usable CMS data, we assemble a sample of pseudo-data containing simulated events from the Drell-Yan process and from all significant background processes ($t\bar{t}$, QCD, W+jet, γ+jet, $\gamma\gamma$), in the proportion expected for 100 pb^{-1} integrated luminosity. The sample is run through the event selection described above, including a simulation of the 18 GeV single electron trigger. This sample is then analysed as if it were real data. It is divided into low-mass and high-mass control regions, and the data-driven techniques discussed above are used on the control regions to estimate background contributions to the signal region. This is performed both with and without the addition of a 1 TeV/c^2 Z' signal.

The distribution of this 100 pb^{-1} of "pseudo-data" is presented in Fig. 3, together with the data-derived background estimates for $t\bar{t}$ (using the b-tag method), QCD jets, W+jet, γ+jet (using the fake-rate method) and $\gamma\gamma$ [3] production. The Drell-Yan estimate is obtained by taking the shape from MC, normalized to the Z peak in the low-mass control region. Taken together, these form the total estimated SM background (shown by the dashed line in the figure). Good agreement is found between the observed spectrum and the total SM prediction estimate in the high-mass control region. For example, a χ^2 test performed in the mass range ($200 < M < 600$ GeV/c^2) returns a $\chi^2/d.o.f.$ of 1.4, equivalent to a p-value of 17%. Were this real data, this would give us confidence to examine the signal region.

Figure 4 shows the same plot with a 1 TeV/c^2 Z' resonance added to the pseudo-data. A clear signal above background is "observed".

5.2 High mass Drell-Yan cross section

The expected mass spectrum from SM processes only for high-mass di-electron events, is shown in Fig. 5(a) for electrons with $E_t > 80$ GeV and $200 < M < 800$ GeV/c^2. It includes DY production and $t\bar{t}$, W+jet, γ+jet and QCD multijet backgrounds; these backgrounds are small ($\lesssim 20\%$) compared to DY production in the whole mass range. The background distributions are determined by imposing the event selection criteria on simulated samples, except for the QCD background where the fake rate method is used to estimate the QCD contribution. The spectrum is scaled to correspond to an integrated luminosity of 100 pb^{-1}. The expected numbers of DY, $t\bar{t}$, γ+jet, W+jet and QCD events are respectively 64.0 ± 0.5, 16.1 ± 1.0, 0.63 ± 0.15, 2.3 ± 1.3 and 0.73 ± 0.67 events.

Figure 5(b) shows the expected spectrum, obtained from the sum of DY and other background pseudo-experiments, each corresponding to an integrated luminosity of 100 pb^{-1}. Also shown

[3] Since the diphoton contribution is expected to be quite small, it is estimated from Monte Carlo simulation.

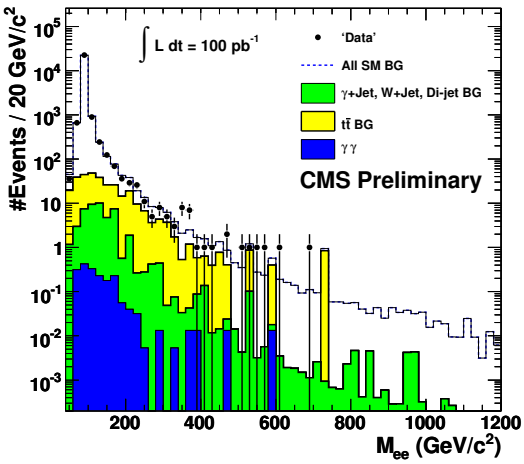

Figure 3: Dielectron invariant mass spectrum for a 100 pb^{-1} pseudo-experiment, compared to SM background estimates for the Drell-Yan process, $t\bar{t}$, QCD dijets, W+jet, γ+jet and $\gamma\gamma$.

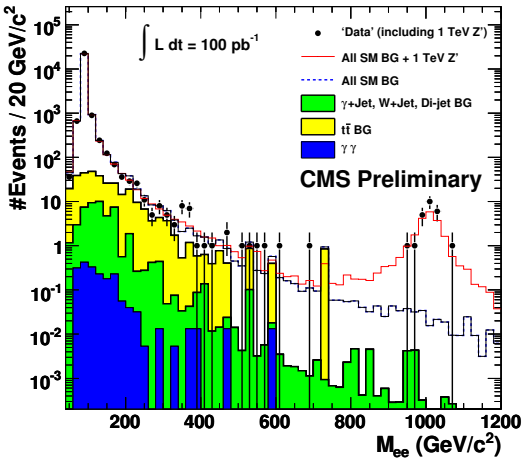

Figure 4: Dielectron invariant mass spectrum for a 100 pb^{-1} pseudo-experiment including a 1 TeV/c^2 Z' signal, compared to SM background estimates for the Drell-Yan process, $t\bar{t}$, QCD dijets, W+jet, γ+jet and $\gamma\gamma$.

are the $t\bar{t}$ background (shaded histograms) estimated from the pseudo-experiment using the $e\mu$ method described in Section 4.1.2, and the jet backgrounds, estimated using the fake rate method (see Section 4.2). The number of observed events is 83. The $t\bar{t}$ background is estimated to be 19.6 ± 2.7 (stat) ± 0.8 (syst) from the $e\mu$ method, and a jet background of 0.8 ± 0.5 is estimated using the fake-rate method (see Section 4.2).

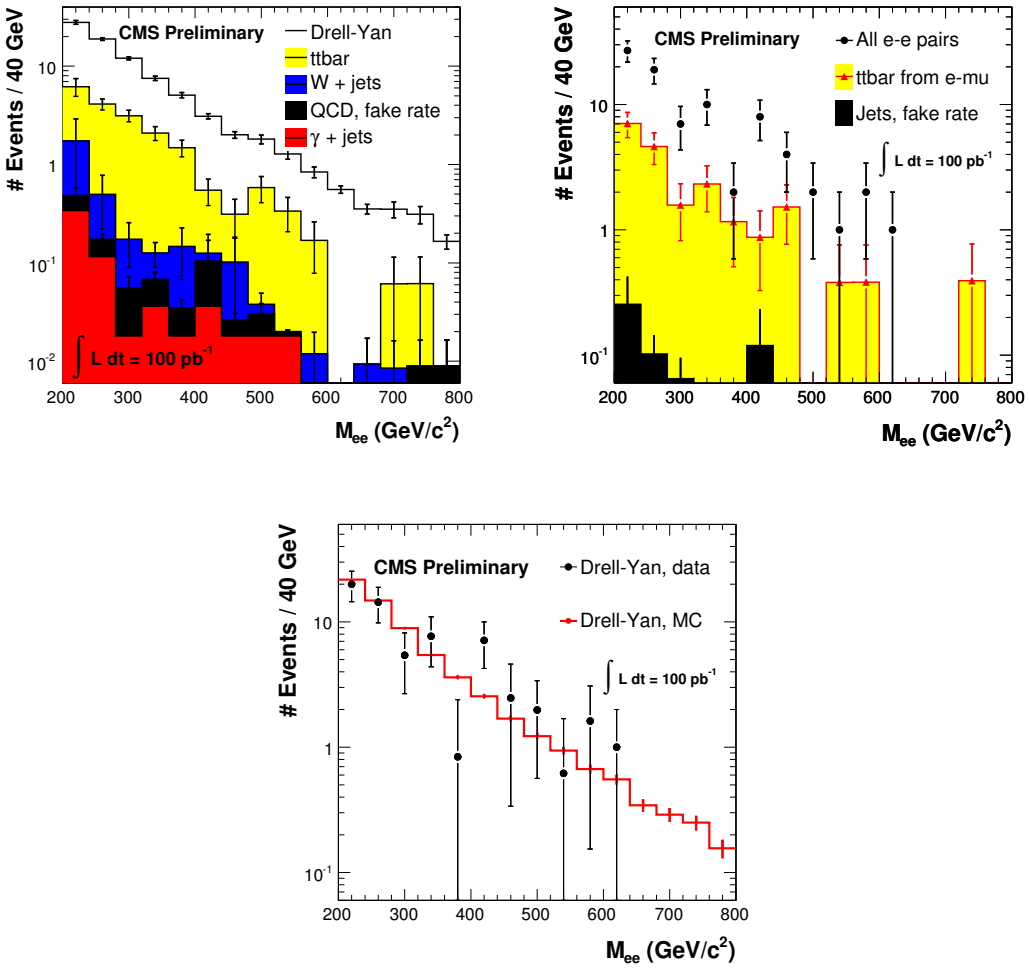

Figure 5: (a) Expected dielectron mass spectrum, containing the DY and $t\bar{t}$, W+jet, γ+jet and QCD multi-jet contributions, with two electron candidates passing the HE criteria with $E_t > 80$ GeV and $200 < M < 800\,\mathrm{GeV}/c^2$ (the histograms are stacked, and scaled to correspond to an integrated luminosity of 100 pb^{-1}); (b) expected spectrum for 100 pb^{-1}, obtained from pseudo-experiments; the $t\bar{t}$ background, estimated by the $e\mu$ method, is shown as shaded histograms, and the jet background, estimated using the fake-rate method, is shown as dark histograms; (c) measured Drell-Yan spectrum (same as (b), after background subtraction); the superimposed histogram is the SM prediction.

The background subtracted, DY distribution is shown in Fig 5(c), where the SM prediction is represented by the histogram. Within statistical precision, general agreement is observed, both in shape and normalization. The probability of a χ^2 fit between the measurement and the expectation, when data are distributed in 3 variable-size bins in invariant mass, is 31 %. The number of DY events at $M > 200\,\mathrm{GeV}/c^2$ is then measured to be 62.6 ± 9.4 (stat.) ± 3.1 (syst.).

The systematic error only includes the error on the number of non-DY events subtracted.

The acceptance due to the geometrical requirements (electrons emitted within the tracker acceptance $|\eta| < 2.5$ and outside the ECAL barrel-endcap transition region) and to the requirement of minimum electron E_t values ($E_t > 80$ GeV), is found to be 0.40. This is determined using a Monte Carlo Simulation. The overall event reconstruction efficiency extracted from the simulation is 0.66. This value is consistent with that obtained using the data-driven method described in Section 3.4.2, which gives an estimate of 0.68 ± 0.14.

The measured cross section obtained from a pseudo-experiment corresponding to an integrated luminosity of 100 pb^{-1} is 2370 ± 350 (stat) ± 120 (syst) fb. The systematic error is obtained by varying the background contribution within its statistical uncertainty. The measurement is in agreement with the SM LO DY cross section of 2524 fb, as used for the Monte Carlo simulation.

The consistency of the measurement (with a 15 % statistical error and a 5% systematic error) with the SM predictions in this high mass range provides an important indication that the detector response to large p_t electrons is understood and that the experimental techniques are under control.

6 BSM signals and CMS potentials

This section presents the signal resonance mass spectrum for Z' production at different masses. The CMS discovery potential, as as well as the upper limit in case of the absence of a signal, is obtained using the procedure described in Section 3.3.4 of Ref. [3]. LO cross sections plus a K factor of 1.35 are used for the Z' and DY production.

6.1 Signal resonance and cross section estimation

Fig. 6 presents an expected e^+e^- mass spectrum for $M > 400$ GeV/c^2, for an integrated luminosity of 100 pb^{-1}, obtained from pseudo-experiments. It contains the contributions of (a) a SSM Z' boson with mass 1 TeV/c^2 and width 30 GeV, and (b) a ψ Z' resonance with mass 1 TeV/c^2 and width of 5.5 GeV, in addition to Drell-Yan (interfering with Z' production), and $t\bar{t}$ backgrounds; other backgrounds are negligible in the high mass region.

Peaks are observed in the distributions at 1 TeV/c^2, of which about 55% of the events are formed of pairs with both electrons detected in the ECAL barrel. In this region, in a mass window of 80 GeV/c^2 corresponding to slightly more than twice the SSM Z' resonance width, 0.1 Drell-Yan events are expected. Monte Carlo studies using alignment and calibration conditions corresponding to an integrated luminosity of 100 pb^{-1} (see ref. [9]), indicate that, in this region, the mass resolution is close to 2%.

6.2 Five σ discovery

The procedure to determine the expected integrated luminosity required to observe a given model with 5σ significance follows the methods described in [10]. The search for new resonances is performed using an unbinned likelihood fit to the e^+e^- invariant mass spectrum over a range which includes the DY continuum as well as a possible peak. The fit takes as input the presumed signal and background shapes, where the signal is modeled as a Breit-Wigner function convoluted with a Gaussian and the background is described by an exponential multiplied by a powerlaw. The natural width and mass of the signal template are fixed to the model parameters and the Gaussian width is set to the expected mass resolution assuming the 100 pb^{-1} alignment and calibration scenario. The parameters for the background shape are set to val-

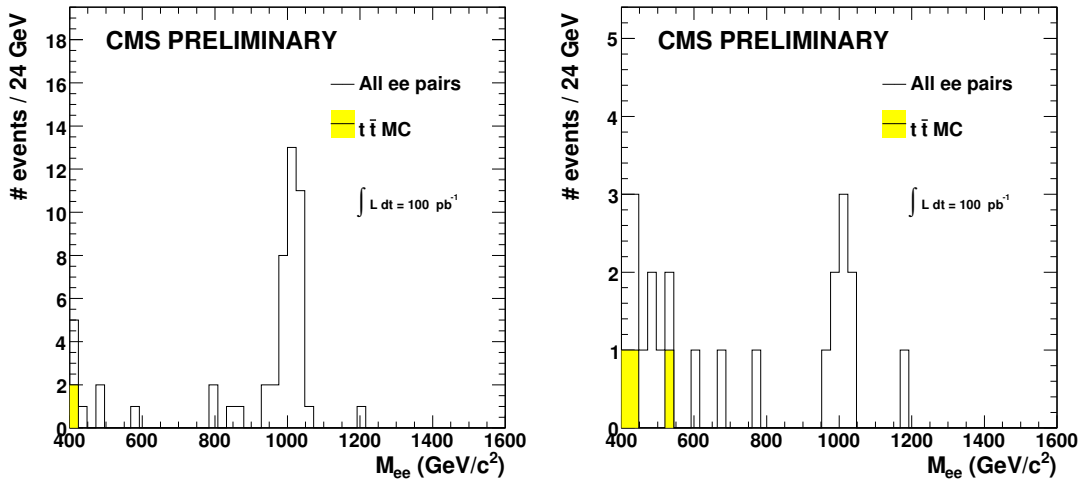

Figure 6: Dielectron invariant mass spectrum after final selections for $M > 400$ GeV/c^2, obtained from pseudo-experiments with an integrated luminosity of 100 pb^{-1}, for the Drell-Yan and $t\bar{t}$ (shaded histograms) backgrounds for (a) a SSM Z'; (b) a ψ Z resonance, with masses of 1 TeV/c^2.

ues determined by a fit to a large event sample DY simulation reaching an invariant mass of 1.8 TeV. The data are then fitted with a sum of signal and background, where the signal fraction is a free parameter. For the present study, backgrounds other than DY have been neglected in the fitting procedure (see Figure 6).

To establish a discovery, the S>5 convention is used, with the likelihood ratio estimator $S = \sqrt{2\ln(L_{s+b}/L_b)}$, where L_{s+b} is the maximum likelihood value obtained in the full signal plus background fit, and L_b is the maximum likelihood from the background only fit.

The integrated luminosity needed to reach a 5σ significance is given in Figure 7(a) as a function of the new resonance mass, for two Z' models. With a luminosity of 100 pb^{-1}, a new heavy SSM Z' or ψ Z' resonance can be discovered at 5σ significance up to masses of 1.5 and 1.2 TeV/c^2, respectively. These estimates of signal significance do not incorporate systematic uncertainties.

6.3 Upper limits in the absence of signal

A similar likelihood ratio based discriminant is used to set confidence limits in the case of the absence of a signal. An extended likelihood formulation is used to fit the actual numbers of signal and background events, and the resonance mass is varied as a free parameter in the fits. The method follows that described in [10]. The expected limit for a SSM Z' resonance is given in Figure 7(b).

6.4 Systematics

Systematic errors on the high mass e^+e^- spectrum measurement will mainly come from uncertainties affecting the detector performance, and from theoretical uncertainties. Concerning the detector performance, the main uncertainties are expected to affect the electron identification, the efficiency and background determinations, and the electron energy reconstruction. The estimation of these uncertainties will be performed using the tools described in previous sections. The uncertainty on $\mathrm{d}\sigma/\mathrm{d}M_{ee}$ due to the PDF uncertainties, as evaluated following the CTEQ6.1 set, are from 6% at $M \approx$ 200 GeV/c^2 to 4% around 1 TeV/c^2. The NLO QCD corrections induce

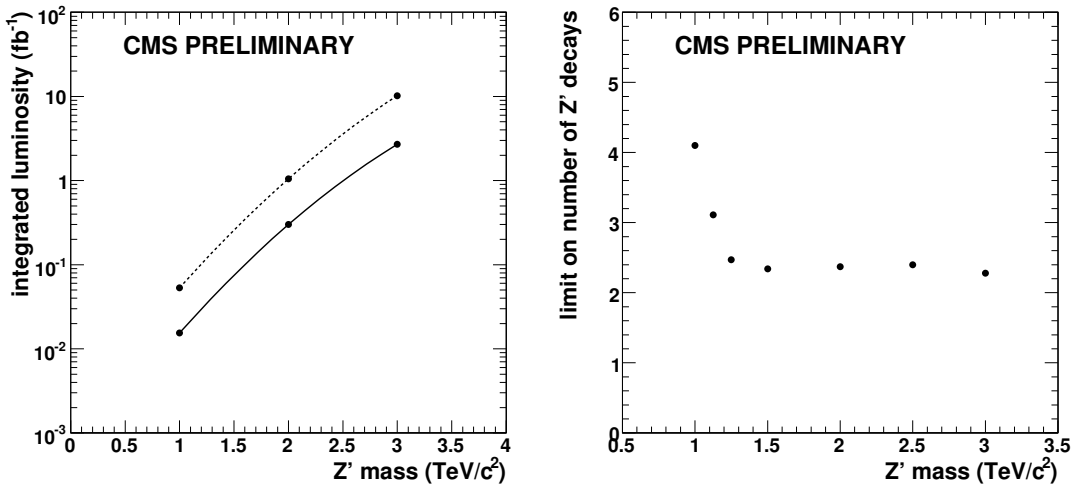

Figure 7: (a) Integrated luminosity needed to reach 5σ significance as a function of the mass, for a $\psi Z'$ (higher points) and a SSM Z' resonance (lower points); (b) expected 95% CL limits on number of events as a function of the resonance mass.

an increase of the cross section of some 15%, and the EW corrections a decrease of -5 to -10%, depending on the mass. This results in an increase of 10% (for masses $\approx$ 200 GeV/c^2) to 5% (for masses around 1 TeV/c^2).

7 Conclusions

This note describes the potential of the CMS experiment to search for new physics using high invariant mass e^+e^- pair data, where the focus is on the conditions expected for an integrated luminosity of 100 pb^{-1}. The electron reconstruction and identification and the final event selection are described. Methods to extract background and efficiency estimates directly from the data reduce the dependence on Monte Carlo simulations. This ultimately enables direct comparisons of Monte Carlo simulation and data to be used to demonstrate that the detector performance and event reconstruction is understood. For many elements of the analysis two independent methods have been devised, allowing for cross checks and estimates of systematic uncertainties. The study takes into account a comprehensive set of background processes, the most important of which are discussed in detail. The results show that the prospects for the discovery of a heavy resonance with the CMS detector are excellent even at low integrated luminosities and with suboptimal detector performance.

References

[1] M. Cvetic and S. Godfrey, "Discovery and identification of extra gauge bosons," arXiv:hep-ph/9504216.

[2] L. Randall and R. Sundrum, "A large mass hierarchy from a small extra dimension," *Phys. Rev. Lett.* **83** (1999) 3370–3373, arXiv:hep-ph/9905221. doi:10.1103/PhysRevLett.83.3370.

[3] **CMS** Collaboration, G. L. Bayatian et al., "CMS technical design report, volume II: Physics performance," *J. Phys.* **G34** (2007) 995–1579. doi:10.1088/0954-3899/34/6/S01.

[4] T. Sjostrand, S. Mrenna, and P. Skands, "A Brief Introduction to PYTHIA 8.1," *Comput. Phys. Commun.* **178** (2008) 852–867, arXiv:0710.3820. doi:10.1016/j.cpc.2008.01.036.

[5] **CDF** Collaboration, T. Aaltonen et al., "Search for new physics in high mass electron-positron events in $p\bar{p}$ collisions at $\sqrt{s}$ = 1.96 TeV," *Phys. Rev. Lett.* **99** (2007) 171802, arXiv:0707.2524. doi:10.1103/PhysRevLett.99.171802.

[6] **D0** Collaboration, V. M. Abazov et al., "Search for Randall-Sundrum gravitons with 1 fb^{-1} of data from $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV," *Phys. Rev. Lett.* **100** (2008) 091802, arXiv:0710.3338. doi:10.1103/PhysRevLett.100.091802.

[7] B. Clerbaux, S. Elgammal, T. Mahmoud, and P. Marage, "Saturation and energy corrections for TeV electrons and photons," *CMS NOTE,* **2006-149** (2006).

[8] **CMS** Collaboration, "Measuring Electron Efficiencies at CMS with Early Data," *CMS PAS,* **EGM-07-001** (2007).

[9] P. Adzic et al., "Intercalibration of the barrel electromagnetic calorimeter of the CMS experiment at start-up," *CMS NOTE,* **2008-018** (2008).

[10] **CMS** Collaboration, "Search for new high mass resonances decaying to muon pairs in the CMS experiment," *CMS PAS,* **SBM-07-002** (2007).