Multi-Lepton Final States in the Search for New Physics at the Large Hadron Collider

João Carlos Arnauth Pela

Under supervision of Doutor João Manuel Coelho dos Santos Varela
Dep. Physics, IST, Lisbon, Portugal

November 9, 2010

Abstract

The Minimal Universal Extra Dimensions (mUED) models predict the presence of massive Kaluza-Klein particles that produce in their decay standard model particles. In a hadron collider environment multi-lepton final states provide the cleanest signature. In this thesis the ability of the Compact Muon Solenoid (CMS) part of the Large Hadron Collider (LHC) to find 3 and 4 leptons final states is evaluated for 4 compactification radius scales $1/R$.

With the start of the LHC data taking, preliminary results are showed on the studied event selections and developed data driven methods to control backgrounds.

Keywords: Particle Physics, Large Hadron Collider, Compact Muon Solenoid, Universal Extra Dimensions.

1 Introduction

1.1 The Standard Model of Particle Physics

The Standard Model (SM) is one of the most successful theories ever made. It is able to predict and explain a multitude of phenomena to a degree of precision never attained by any theory before. There is currently no experimental evidence of any further internal structure of its particles and to date all experimental tests have agreed with the predictions of the SM.

However, it does not incorporate the physics of general relativity, such as gravitation and dark energy; it does not have any viable candidate to dark matter; the number of generations and the pattern of quark, etc. This are just a few examples
of possible problems with the model. These facts indicate that the SM is a low energy manifestation of a more fundamental theory.

1.2 The Minimum Universal Extra Dimensions Model

The Universal Extra Dimensions (UED) model\cite{UED} is an extension of the sub-millimeter extra dimensions model (ADD)\cite{ADD1}\cite{ADD2} in which all Standard Model (SM) fields, fermions as well as bosons, propagate in the Extra Dimension (ED) 'bulk'. The ED is compactified in a "orbifold" with the size of the compactification radius, \( R \). In the first level excitation mode, KK particles appear with masses below the \( TeV \) scale, accessible at LHC. Important phenomenological characteristic: the first level KK states must be pair produced, the lightest KK particle (LKP) is a neutral and stable KK photon (\( \gamma_1 \)) which is a dark matter candidate, only relevant parameter to define the KK particles masses is \( R^{-1} \).

Precision electroweak data measurements set a lower bound of \( R^{-1} > 600 \, GeV/c^2 \) for a SM Higgs mass of \( m_H = 115\, GeV/c^2 \) and top quark mass of \( m_{top} = 173\, GeV/c^2 \). Assuming a larger Higgs mass \( (m_H \simeq 600\, GeV/c^2) \) this limit can be as low as \( \sim 300\, GeV/c^2 \). Dark matter constraints allow to infer that \( 600 < R^{-1} < 1050\, GeV/c^2 \)\cite{DM}. Experimental limits from the Tevatron are \( R^{-1} > 280\, GeV/c^2 \)\cite{D2}

At LHC the most promising signatures are the ones involving a combination of multiple isolated leptons and jets in the final state, since the predicted cross sections for such processes in mUED is significantly bigger then in SM. In this analysis, we will focus on final states with 3 and 4 leptons, which should allow for good separation from SM background processes.

1.3 The Large Hadron Collider

The Large Hadron Collider (LHC) is at this moment the world’s largest and highest-energy particle accelerator in activity. The LHC is a synchrotron machine designed to collide two opposing particle beams of protons each one with a maximum energy of 7 \( TeV \) per proton. Providing a maximum available energy of 14 \( TeV \) in the center of mass frame for a single proton-proton collision.

All methods and results presented in this thesis refer to machine conditions during the 2010 run at \( \sqrt{s} = 7\, TeV \).

1.4 The Compact Muon Solenoid Experiment

The Compact Muon Solenoid Experiment (CMS) is a general purpose experiment that is an integrating part of the LHC machine. It was designed to study the collisions of two intersecting proton beams in its center. This detector was planned with the intention of studying a broad spectrum of physics processes and is made of several subsystems, each one designed to take advantage of some characteristics of the particles produced in order to measure their energy, momentum and charge. The
detector has classical onion shape with several layers, starting from the collision point outwards we have: Tracker System, Electromagnetic Calorimeter, Hadronic Calorimeter, Magnet, Muon System and Return Yoke.

2 Standard Model backgrounds

Multi-lepton backgrounds are processes were real SM are produced and/or objects as mis-reconstructed as leptons prompt isolated leptons. The most relevant SM processes to multi-lepton studies are: $WZ$, $t\bar{t}$, $Z\bar{b}$, $ZZ$ and Drell-Yan (DY).

3 Monte Carlo simulations

The CMS Software (CMSSW) was used in interface with several physics process generators, was used to generate signal and background samples using a full detector, electronics and reconstruction simulation.

3.1 The mUED signal samples characterization

As the parameter $1/R$ increases the percentage of $g_1 g_1$ (4 jets) and $q_1 q_1$ (3 jets) events decreases while $q_1 q_1$ (2 jets) increases.

All 3 and 4 leptons mUED signal channels have similar lepton $p_T$ distributions. The first lepton peaks between 20 and 30 $GeV$ for $1/R$ 300 and 900 respectively. For last lepton all distributions reduce significantly their average. For $1/R = 300$ $GeV$ the distribution peaks around 2 $GeV$ while $1/R = 900$ $GeV$ peaks around 20 $GeV$. This represent a major experimental difficulty since lepton reconstruction efficiency especially for electrons but also for muons fall rapidly for $p_T < 5$ $GeV$.

In terms of $\eta$, leptons for mUED events are produced preferentially in the central rapidity area.

4 Discrimination of signal from Standard Model backgrounds

Two 2 event selection aimed at the discrimination of both 3 and 4 mUED final states we developed in order to establish the existence or not of such signal and provide grounds for claiming discovery or exclusion. This selections therefore aim for suppressing SM events while enhancing mUED signal and are based in simple yet effective cuts on several key variables.

4.1 Trigger Selection

Since we are interested in multi-lepton events with final states that range from 1 to 4 leptons of a given flavor, we choose the lowest threshold un-prescaled single and double lepton triggers available. Which are the following for the HLT1E31 menu: HLT_Ele20_SW_L1R, HLT_Mu9, HLT_DoubleEle10_SW_L1R and HLT_DoubleMu3.

A trigger efficiency study was preformed over the Monte Carlo signal samples. It shows that the current trigger selection has an efficiency higher then 90% for all $1/R$ points in both 3 and 4 lepton final states with the exception of the 3 lepton
final state for $1/R = 300$ which has an efficiency of 82.2%.

4.2 Offline selection

Two sets of selection cuts are proposed, a 4 lepton selection optimized to select events from mUED to 4 leptons ($pp \to Z_1Z_1 \to 4l$) and a 3 lepton selection, optimized to select events from mUED to 3 leptons ($pp \to W_1Z_1 \to 3l$). Both selections attempt to capitalize on the kinematical and topological features of the events they are aimed to select. We will attempt to select high quality isolated prompt leptons (and a reasonable $p_T$ jet in the 3 lepton selection) by using robust identification and isolation definition based on simple variables.

5 Monte Carlo expectations for 1 fb$^{-1}$

We can now apply the event selections defined in previous chapters to our simulated Monte Carlo events. Event yields were produced and a summary of the results for $1/R = 300$ GeV can be found on table 1.

For the 3 lepton selection for all channels the dominating backgrounds are $t\bar{t}$ and $Z(ll)$, while all other studied background channels represent then 10% of total background. Still, for $1/R = 300$ GeV we can obtain a $S/B$ higher then 1.4 for all channels (best channel is $3\mu$ 2.6).

For the 4 lepton selection the dominant background for $4\mu$ and $2e2\mu$ is once more $Z(ll)$, but for the channel $4e$, dominant background are $ZZ$ and $Zb\bar{b}$. However all channels in the 4 leptons selection have low signal statistics despite the better $S/B$ that can reach 8.9 for $2e2\mu$, making it difficult to use them in real data.

5.1 Invariant mass distributions

The invariant mass and maximum invariant mass are key variables in the analysis. We can start by using the dilepton invariant mass spectrum to evaluate the $(Z/\gamma^*)$ shape. Then can start tightening our requirements by requesting 2 isolated + 1 non-isolated leptons and 3 isolated leptons. By doing so we can start suppressing SM background gradually allowing us to see the shape changes and gain confidence on the Monte Carlo description of the SM backgrounds.

Finally, we can make plots the 3 lepton selection $M_{ll}^{max}$ plots (example for the $3\mu$ on fig. 1) for each signal channel (events with 3 leptons + 1 Jet).

Finally we can do the same for the 4 lepton selection. The channel where shapes are better defined is $2e2\mu$ (fig. 2) since in this case we don’t have to cope with combinatorial background.

A very good signal to background separation is archived in all plots, showing that this is a very useful variable in the mUED studies context.

6 Uncertainties

An overall estimated uncertainty of 10% (i.e. 2.5% for each lepton) is assumed in order to cover a pos-
The systematic uncertainty due to the luminosity measurement is estimated at 10% and it is common to both signal and background. In order to account for the small number of events which will be used with early data to determine the background contribution in the signal region, a systematic uncertainty of 100% is estimated.

### 7 Discovery potential and exclusion limits

Exclusions limits to 95% confidence level and the discovery potential for 3 and 5 \( \sigma \) significances were calculated for \( 1/R = 300 \text{ GeV} \). It was considered 100% background uncertainty, since errors are heavily dominated by statistics and cross section uncertainties.

Exclusion can only be attained for luminosities under 1 \( \text{fb}^{-1} \) of data (which is the limit for the 7 TeV run) for 2e2\( \mu \) and 4\( \mu \) channels, at luminosities of 310 and 600 \( \text{pb}^{-1} \) respectively. While 3 \( \sigma \) discovery would only be possible for the 2e2\( \mu \) channel, while no channel can reach 5 \( \sigma \) discovery.

### 8 Data selection

When dealing with real data some filtering must be done in order to select the good quality events. A list of the corrections and filters used follows.

- **Physics trigger bit**: Requirement that each event is marked at trigger level as a physics event.
- **No scraping filter**: Rejecting events with a large multiplicity of low quality tracks.
- **Vertex filter**: Requesting that events have good quality and well defined primary vertex.
- **HB HE Noise Filter**: Correct/Reject events with problems with noise in the Hadronic Calorimeter.
- **ECAL Misalignment correction**: Correction to Electromagnetic Calorimeter alignment.

### 9 Results with early data

Currently all triggers used in the Monte Carlo analysis are available in real data. For this reason all
Figure 1: $M_{ll}^{max}$ of 3 isolated muons. Using the definitions of the 3 lepton selection.

data results present in this thesis will be over the runs where all triggers from Monte Carlo analysis are available, from run 141956 to run 143336, this interval represents $1.064 \, pb^{-1}$ of data.

10 Event Yields

The results of both selections show only one event passing in all 7 possible signal channels, specifically in the 3 muons channel. This channel is affected by combinatorial background, so it is bound to have some events until the combinatorial background subtraction method starts working (depends on shape definition).

11 Invariant mass distributions

This final states are dominated by Drell-Yan and $Z$ decays into leptons and should allow us to estimate this kind of background shapes as well as possible contaminations from other processes like heavy flavor decays to leptons. We now present invariant mass distributions for di-electrons (fig. 3) and di-muons (fig. 4):

We can see a good agreement in the $Z$ description, in both distributions, but in the di-electron case there is a significant disagreement in the description of DY despite the reasonable description of the di-muons channel. This result indicates that electron channels will be more difficult to handle and will probably require enhanced definitions of
Figure 3: $M_{ll}$ of 2 isolated electrons for $1.064 \text{ pb}^{-1}$. Using the definitions of the 4 lepton selection.

Figure 4: $M_{ll}$ of 2 isolated muons for $1.064 \text{ pb}^{-1}$. Using the definitions of the 4 lepton selection.

isolation and/or additional requirements in order to suppress backgrounds.

12 Data driven Methods to estimate background using data

12.1 Fake rate method

With this method we have the objective of estimating the fraction of Z+Jets events that have 3/4 isolated leptons according to our lepton definitions for each event selection. The isolated leptons in the final state come from the leptonic decay of the $Z^0$ boson (resulting in two opposite sign same flavor leptons) plus 1 or 2 fake leptons from jets.

12.1.1 Determination of fake rate functions

First we need to evaluate the probability of a jet “faking” a lepton directly from data. We choose to use events that passed the trigger path HLT_{Jet15U}, which triggers on jet candidates with an uncorrected minimum transverse momentum of 15 GeV. To avoid bias, trigger objects are matched to reconstructed jets using distance between objects of $\Delta R < 0.5$.

Now, we can define the fake rate as a function parametrized in $p_T$ and $\eta$ which are its main dependences. Since we are interested in evaluating the probability of a jet being mis-reconstructed has as lepton, our denominator will be the number of isolated leptons that are matched to jets, and our
numerator is the total number of jets found. In this case an isolated lepton is considered matched to a jet if the are at $\Delta R < 0.3$.

Distributions are more or less plain in $\eta$ with the exception of the transition between barrel and endcaps, for both electrons and muons. In the $p_T$ distributions we can see that the fake increases with $p_T$ but converges for a constant level in all distributions, electrons converge to around $10^{-2}$ and muons converge around $10^{-3}$ to $10^{-4}$.

12.1.2 Application to QCD background estimation

By doing a simple set of assumptions we can make a reasonable prediction of how many events with 3 and 4 leptons will be generated by QCD jets alone. To do so we just need to assume that the Fake Rate is independent of the jet multiplicity in an event in order to apply the average fake rate to jets present in multi-jet events. We also need to know what is the number of event with at least 4 jets above minimum $p_T$ to fake a lepton that would pass our event selections. We can obtain:

- Expected Fake at $1 fb^{-1}$ for $3e = 12.3$
- Expected Fake at $1 fb^{-1}$ for $3\mu = 0.0047$
- Expected Fake at $1 fb^{-1}$ for $4e = 0.27$
- Expected Fake at $1 fb^{-1}$ for $4\mu = 4.62 \times 10^{-6}$

We can see from the obtained results that at most at $1 fb^{-1}$ we will have 1 purely fake event and it will be in the $3e$ channel. Of all channels this is the one that is most probable to have a higher fake rate since electrons have a higher probability of being faked by jets and this is the lowest multiplicity all electron channel considered.

12.1.3 Application to $(Z/\gamma^*)+\text{Jets background estimation}$

After determination of the fake rate function we can now apply it to the jets present in a sample of Z+Jets candidates. We chose events with 2 isolated leptons (and a jet for 3 leptons final states) with lepton pair invariant mass of $76 < M_{ll} < 106 GeV$.

We calculate the probability of a Z candidate faking a 4 lepton event by running over all possible combinations of jet pair in the event and calculating the probability of those 2 jets faking 2 leptons. Repeat the process for all Z candidates and average the probability of one of Z candidate faking a 4 lepton events.

For 3 lepton final states we require that the Z candidate to be associated with a $30 GeV$ jet. We then compute the probability of the remaining jets faking a third leptons. If there are 2 or more jets passing the $30 GeV$ requirement all jets are considered in the search for a third lepton.

It is not possible to make a prediction for channels $3e$ and $1\mu 2e$ since no dilepton events is found in the Z mass region. Predictions are dominated by statistical error since we are extrapolating to a factor of 1000 more data. Yet while prediction for 4 lepton final states seem to converge to values close to the MC predictions.
For double fake lepton to pass our selection it needs to have opposite charge, and we are not requiring any charge on the evaluated jets, this effect is not considered thus adding a fact of 2 to calculated values. Due to this factor and cross section corrections from NLO, the predicted values seem reasonable, but require further study to confirm validity.

12.2 SM Candle method

The standard model candle method is based in the use of the Z peak in the $M_{ll}^{\text{max}}$ distribution as a handle to extrapolate our Monte Carlo prediction to the signal region. This region is signal signal free that we call the control region. We are assuming that our background is similar in shape between our predictions and data.

Unfortunately, there isn’t enough collected integrated luminosity at this point to have enough events passing our selection so we have some events in both regions. Thus, we cannot apply this method yet.

12.3 Relaxed final states

This method relies in using the Z peak again, as an handle to normalize our distributions but instead of using Monte Carlo predictions we use distributions of lower lepton multiplicities. Since our signal will manifest in 3 and 4 leptons $M_{ll}^{\text{max}}$ distributions, we will use 2 lepton $M_{ll}$ has a signal free distribution and use the signal and background scheme as explained in the SM Candle method. By doing so, and since $M_{ll}$ is dominated by Drell-Yan and $Z^0$ we are assuming that our 3 and 4 leptons $M_{ll}^{\text{max}}$ distribution have shapes essentially dominated by this same SM process.

Once more, since this method is based in using the Z peak for normalizing our distributions we cannot apply it yet to data.

12.4 Charge/Flavor method

This method is based on using lepton and/or charge combinations that are unaccessible to signal in order to estimate the background. Since our signal is generated by the decay of $Z_1Z_1$ and $Z_1W_1$ the possible combinations of charge/flavor are limited.

To apply this method we need to have some counts in at least some forbidden channels. Unfortunately, at this point there are only a few event passing selection cuts, not allowing us to make any significant prediction.

13 Conclusion

A study of the discovery potential for extra dimensions within the context of the MUED model at the CMS detector was presented. The processes with three ($3e, 2e\mu, 1e2\mu$ and $3\mu$) and four leptons ($4e, 4\mu$ and $2e2\mu$) in the final state were studied using full detector simulation and event reconstruction at four points of the parameter space ($R^{-1} = 300, 500, 700, 900 \text{ GeV}/c^2$ and $m_H = 120 \text{ GeV}/c^2, \Lambda R = 20$). A center of mass energy for
the LHC of 7 TeV was assumed.

A trigger study was performed showing efficiency higher than 90% for all 1/R points in both 3 and 4 lepton final states with the exception of the 3 electron final state for 1/R = 300 which has an efficiency of 82.2%.

Several background sources were considered. Two event selections were presented each one optimized for each signal multiplicity under study. Both selections are built over robust yet simple variables, that try to capitalize on the kinematical and topological properties of mUED signal events. Good signal to background separation is obtained and all studied channels show S/B higher than 1.4 for 1/R = 300 GeV. Exclusions limits are calculated for 1/R = 300 showing that the only channels able to exclude this point during the 7 TeV run are 2e2µ and 4µ, which could be achieved around 300 pb⁻¹.

Preliminary results over first pb⁻¹ of data are presented including results on 2 data driven methods for background estimation. This preliminary study reveals good agreement between Monte Carlo prediction and data for 2 and 3 muon final states while showing some difficulties and need of further study for electron final states. Data driven methods already reveal that QCD only events will not be a problem on multi-lepton selections, while (Z/γ⁺)+Jets background is estimated and assumes the role of dominant background in this analysis.

References


