CDF/PHYS/EXOTIC/CDFR/11122
Version 2.0
October 8, 2014

Search for High Mass ZZ→lljj Resonances with Tight Selection Criteria

E. Andreadaki, V.Giakoumopoulou, N. Giokaris, A. Manousakis-Katsikakis
Univ. of Athens

Abstract

We have searched for ZZ→lljj resonances with invariant mass larger than 320 GeV/c^2 in 6fb^{-1} of CDF data. An excess of rate is found around 350 GeV/c^2, with significance 1.64σ. If this excess is attributed to a non-SM resonance, we estimate its cross section times branching ratio to ZZ to be about 0.6 pb

Contents

1 Introduction

Our original search for high mass ZZ→ll+jets resonances showed an indication for a ZZ→lljj resonance in the 350 - 360 GeV mass region [1]. Our subsequent search for ZZ & WW resonances in the ll+MET channel also showed an excess of events over the Standard Model (SM) background in the same (~350 GeV) mass region [2], [3], [4]. In this note we take another look at the Z (ll) + jets channel using the "tight cuts" selection criteria [1]. In our new approach here we will not assume the presence of a "signal" (as we did in [1]). Instead we will look for a possible excess of events in the high mass Mlljj > 320 GeV/c^2 over the SM expectations. The findings of the present analysis are therefore independent of any theoretical prejudice.

2 Backgrounds

The major background in the Zjj channel is the inclusive Z+jets production. At a much lower rate (a few % of the total) contribute the diboson ZZ, WZ and WW and the ttbar production. We looked for but found not any significant contribution from the W+jets channel. We use PYTHIA [5] to estimate the diboson and the ttbar backgrounds. We will use ALPGEN [6], interfaced with PYTHIA [5] for parton showering and hadronization, to estimate the Z+jets background. The ALPGEN normalization (or the, so called, scale factor SF) is obtained by fitting to the total data yield in the control region defined as the one with Mlljj < 320 GeV/c^2 where M_{lljj} is the invariant mass of the two leptons and the two leading jets in the event. This procedure is described in detail below in subsection 4.1.

3 Datasets

This is an analysis based on GEN6. We analyzed data from periods 0-28 using the electron central 18, z notrack, muon cmup18, muon cmx15 and muon cmx18 triggers. The corresponding integrated luminosity is ~ $6fb^{-1}$. We use the lepton selection described in CDF10144 and select two loose leptons. We apply the DQM v34 good run list emunoSi, which corresponds to an integrated luminosity of 5.998/fb. The principal background in the lljj final state is Drell-Yan Z+jets and for this we use the Alpgen HIGH LUMI datasets bt0sz[09]. We use Pythia WZ, ZZ and top-antitop datasets. We apply Joint Physics trigger efficiencies and livetimes, and lepton identification scale factors measured by us for the WZ and ZZ cross-section measurements (CDF10144) and validated by inclusive Z cross-section measurements (CDF10125).

4 Selection Criteria - Results

4.1 Control region

The selection criteria in the control region are:

1. Two opposite sign leptons (ee's or μμ's) with their invariant mass M_{ll} in the Z-mass window: 76<Mll<106 GeV (Standard CDF Z-mass window).

2. Exactly two jets with:

   - leading jet $E_T > 50$ GeV and $|\eta| < 3$
   - sub-leading jet $E_T > 20$ GeV and $|\eta| < 3$

   (any other jet in the event should have $E_T < 15$ GeV and $|\eta| < 3$)

The reason for selecting events with just two jets with significant E_T is given in the first paragraph of subsection 4.2 of CDF 10637 (at the top of page 12). Another reason is that we like to keep this analysis as close as possible to our ll+MET one [2,3,4]. We remind that in that [ll+MET] analysis we selected events with NO jets with significant E_T.

3. The leptonic Z transverse momentum ZP_T>100GeV/c^2. This cut is justified as it greatly reduces the background while cutting out less that ~20% of a potential signal from a 350 GeV/c^2 resonance decaying to ZZ as shown in Figure 1.

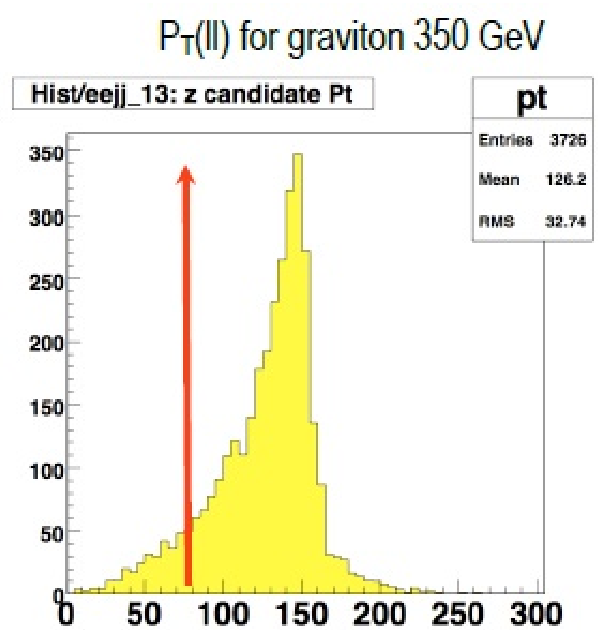


Figure 1: Leptonic Z P_T distribution of a 350 GeV Graviton.

We note that the above cut on ZP_T is pretty similar to the MET>100GeV cut in the signal region of our ll+MET analysis [2,3,4], as here we can use another variable (the M_{lljj} one) to define and separate the control region from the signal one.

4. Additional cuts

   a. $70 < M_{jj} < 120$ GeV, where M_{jj} is the invariant mass of the two jets

   b. |leptonic ZP_T - hadronic ZE_T| < 25 GeV

   c. $120 < P(Z) < 170$ GeV, where $P(Z)$ is the full momentum of the leptonic Z

All three of the above cuts reduce further the background while having no significant effect on a possible resonance with a mass of 350 GeV. Specifically the upper bound (170 GeV) on cut c is used to cut out events with very high energy QCD jets for which ALPGEN (used to simulate the main Z+jets background) could have trouble in simulating them correctly. After the imposition of the rest of the cuts there are very few events with P(Z) > 170 GeV anyway.

5. $M_{lljj} < 320$ GeV, where M_{lljj} is the invariant mass of the two leptons and the two jets.

From Figure 2 we see that the contribution of a possible 350 GeV resonance in the $M_{lljj} < 320$ GeV region is minimal.

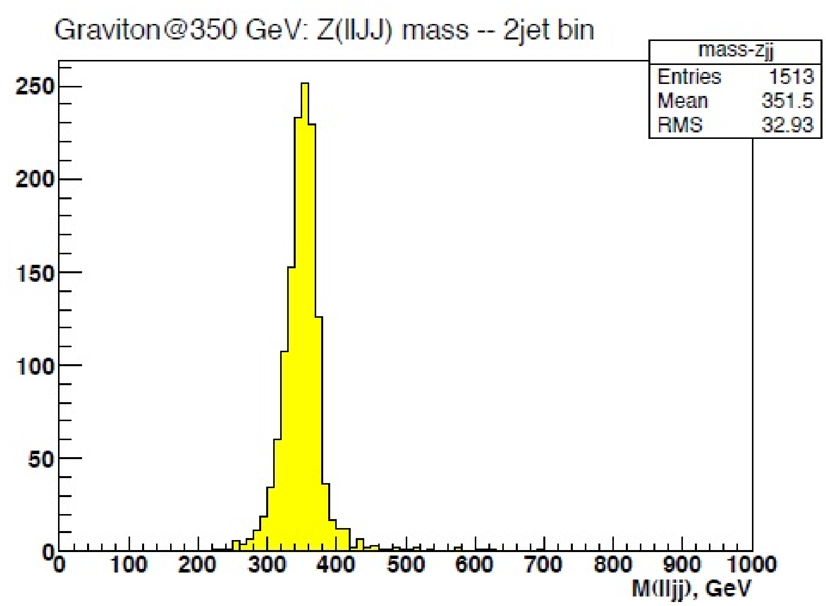


Figure 2: Invariant mass of the two leptons and the two leading jets for a 350 GeVGraviton.

Table 1 lists the control region selection cireria:

cut	value
M_{lljj}	< 320 GeV
Mll	> 76 and < 106 GeV
leading jet E_T	> 50 GeV
sub-leading jet E_T	> 20 GeV
any other jet E_T	< 15 GeV
all jets $\|\eta\|$	< 3
$P_T Z(ll)$	> 100 GeV
M_{jj}	> 70 and < 120 GeV
$\|$leptonic ZP_T – hadronic $ZE_T\|$	< 25 GeV
$P(Z(ll))$	> 120 and < 170 GeV

Table 1: List of **control region** criteria.

In Figures 3 and 4 the M_{lljj} data distributions of the ee+2jets and μμ+2jets respectively are shown with the control region cuts as summarized in Table 1.

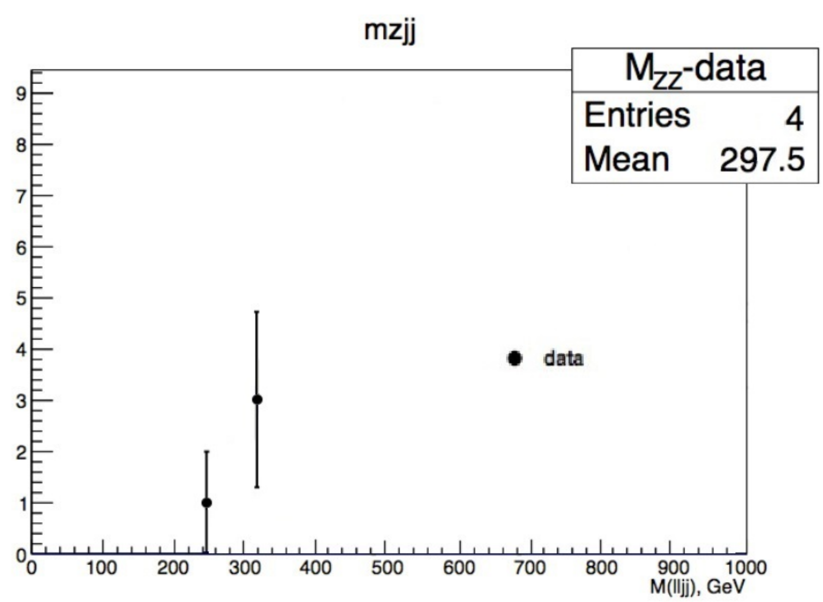


Figure 3: The ee+2jets M_{lljj} data distribution with the **control region** cuts.

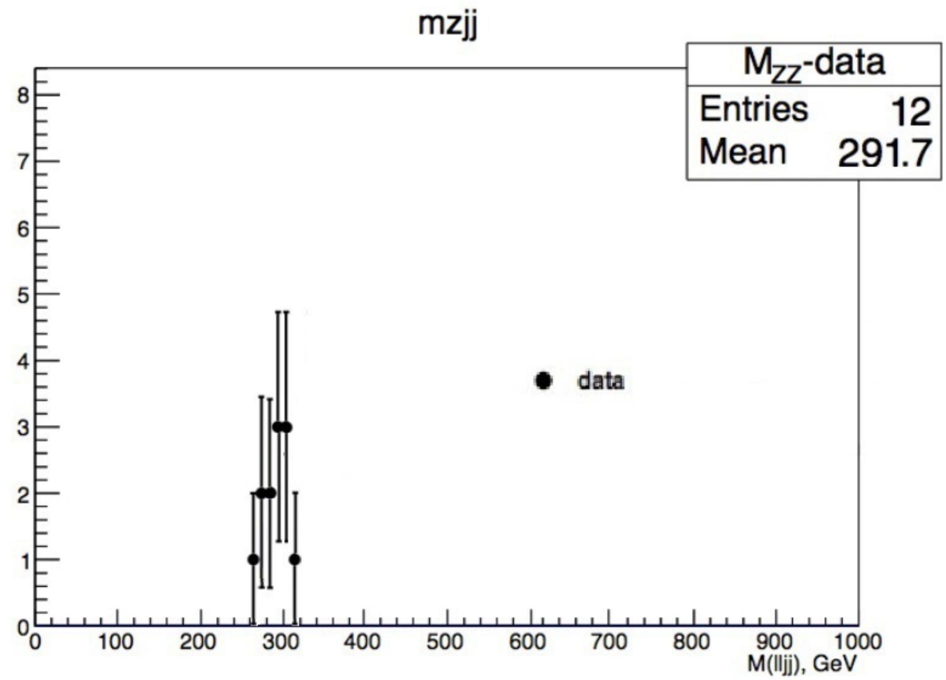


Figure 4: The μμ+2jets M_{lljj} data distribution with the **control region** cuts.

From Figures 3 and 4 we see that there are only 16 ee+μμ data events (4 ee and 12 μμ) in the M_{lljj} < 320 GeV region (i.e. what we have defined as our control region). For this reason (i.e. for the low statistics of data events in our control region) form now on we will sum up the two samples (ee+μμ). The ee+μμ M_{lljj} data distribution with the control region cuts, is shown in Figure 5.

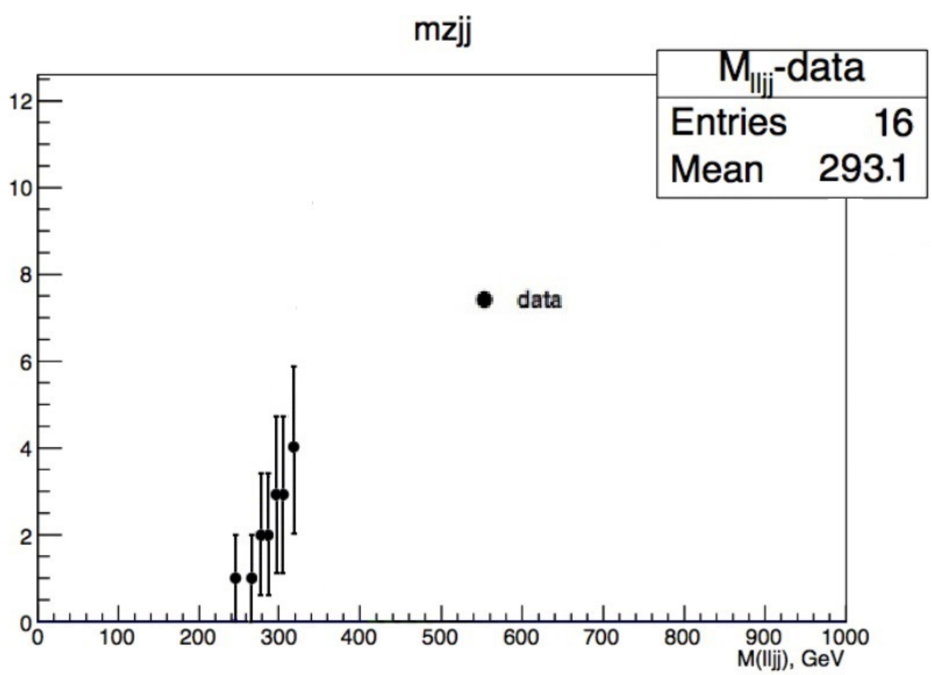


Figure 5: The ee+μμ M_{lljj} data distribution with the **control region** cuts.

In Figure 6 the M_{lljj} ee+μμ distribution in the control region is shown for the data and the total background. A Z+jets scale factor of 1.04 ± 26% has been used so as to equalize the total background to the data. This quoted error on the Z+jets scale factor is statistical and is dominated by the small number of data events (16) in the control region.

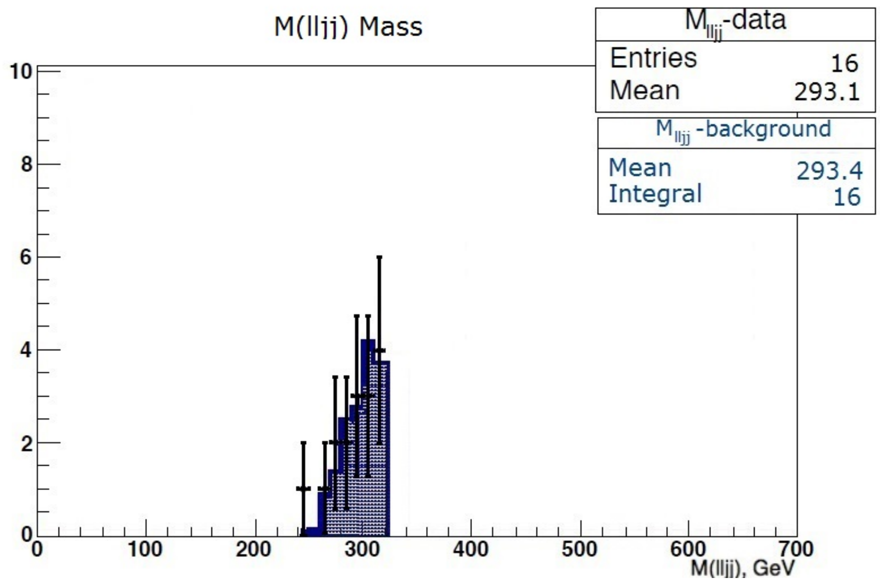


Figure 6: The ee+μμ M_{lljj} data (black crosses) and total background (blue) distributions in the control region. The Z+jets ALPGEN background has been multiplied by a scale factor equal to 1.04.

In Table 2 the number of the background and the data events in the control region are listed. The quoted errors are just statistical.

source	[Z(ee)+2jets] + [Z(μμ)+2jets]
ZZ, WZ & ttbar	0.5 ± 0.05
Z + jets (= Z+1jet + Z+2jets + Z+3jets + Z+4jets)	15.5 ± 4.03
Total Background	16.0 ± 4.03
Data	16

Table 2: List of [(Z(ee)+2jets) + (Z(μμ)+2jets)] data and background events in the **control region**. The Z+jets ALPGEN background has been multiplied by a scale factor equal to 1.04.

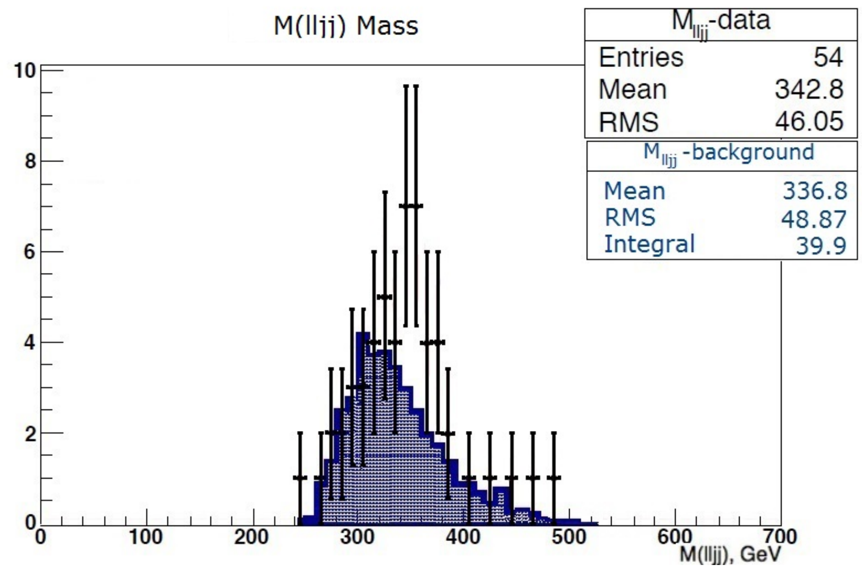
4.2 Signal region

We define as full region the region with the criteria listed in Table 1 but without the cut on M_{lljj} and as the signal region the one with the criteria listed in Table 3 (i.e. the same ones as in the control region but, now, M_{lljj} > 320 GeV).

cut	value
M_{lljj}	> 320 GeV
Mll	> 76 and < 106 GeV
leading jet E_T	> 50 GeV
sub-leading jet E_T	> 20 GeV
any other jet E_T	< 15 GeV
all jets $\|\eta\|$	< 3
$Z(ll)P_T$	> 100 GeV
M_{jj}	> 70 and < 120 GeV
$\|$leptonic ZP_T – hadronic $ZE_T\|$	< 25 GeV
$P(Z(ll))$	> 120 and < 170 GeV

Table 3: List of **signal region** criteria.

In Figure 7 the ee+μμ M_{lljj} data and total background distribution in the full region is shown. A Z+jets scale factor of 1.04 (same as in the control region) has been applied.



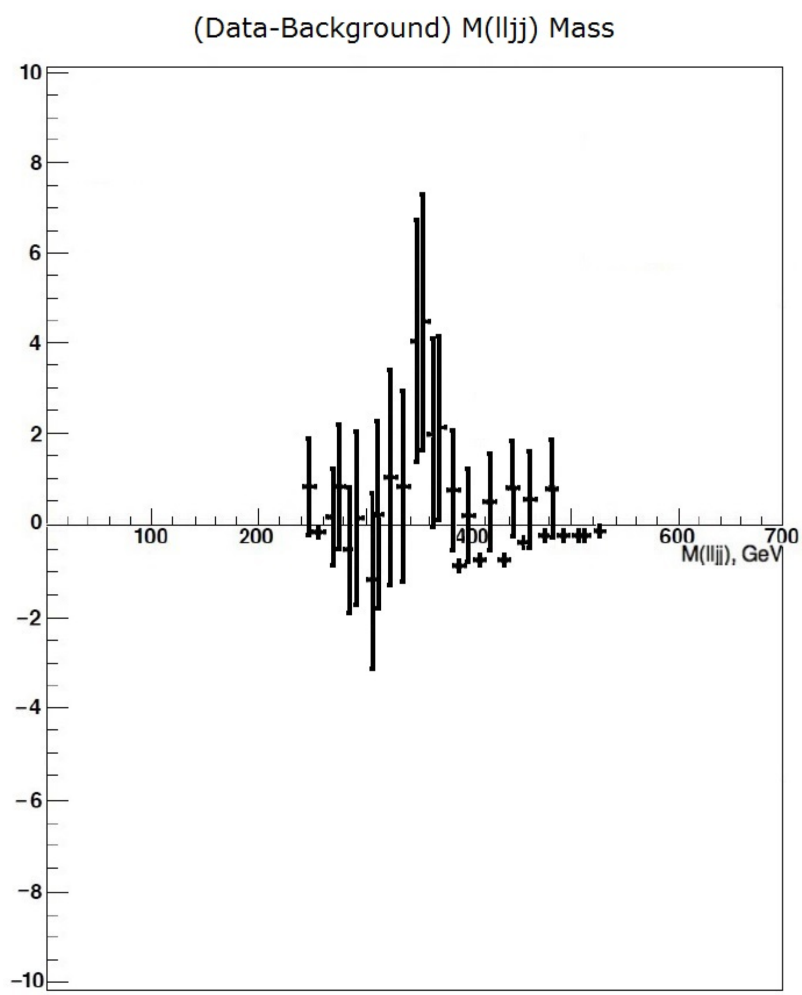


Figure 7: The ee+μμ M_{lljj} data (black crosses) and total background (blue) distributions in the **full region** (top) and the bin by bin (Data-Background) Mlljj distribution also in the full region (bottom). The Z+jets ALPGEN background has been multiplied by a scale factor equal to 1.04.

There is good evidence of an excess of data over the estimated Standard Model (SM) background in the region 320 < M_{lljj} < 380 GeV.

In Table 4 the number of the background and data events in the signal region are listed. The quoted errors are just statistical.

source	[Z(ee)+2jets] + [Z(μμ)+2jets]
ZZ, WZ & ttbar Z + jets (= Z+1jet + Z+2jets + Z+3jets + Z+4jets)	4.7 ± 0.19 19.2 ± 4.99
Total Background	23.9 ± 4.99
Data	38

Table 4: Number of data and background events in the **signal region**.

In Figures 8 the ee and μμ invariant mass distributions for the data in the full region are shown. We see that both distributions have means very close to the nominal Z-mass.

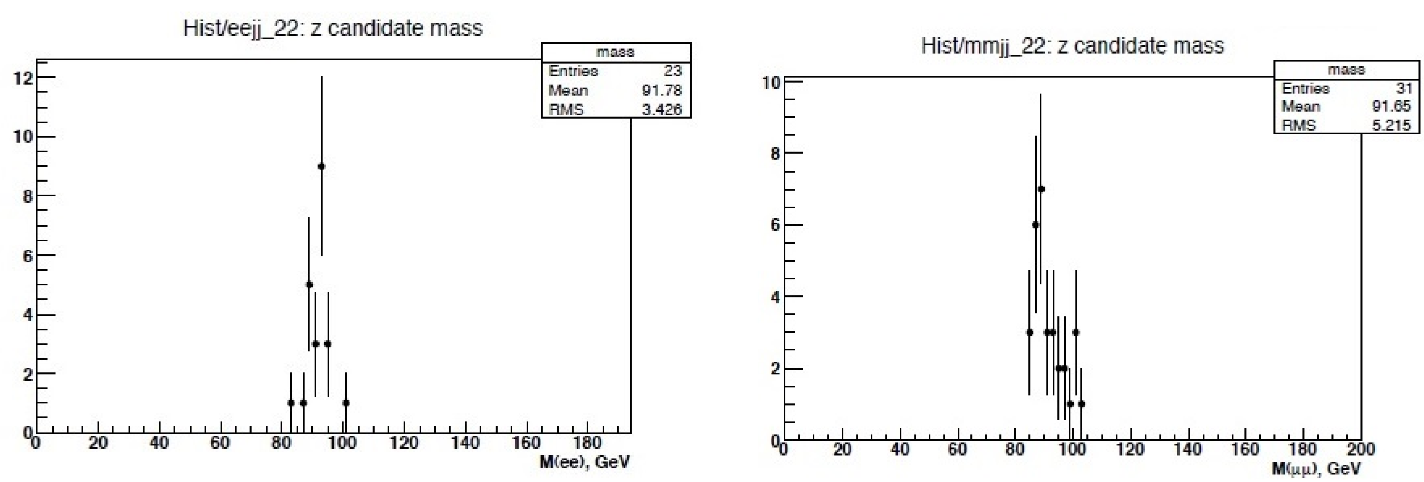


Figure 8: ee (left) and μμ (right) invariant mass data distributions in the **full region**.

In Figure 9 the two leading jets invariant mass M_{jj} distribution for the data and background for ee +μμ in the full region is shown. A good evidence for an excess of events in the 90 GeV region (Z-mass) is seen.

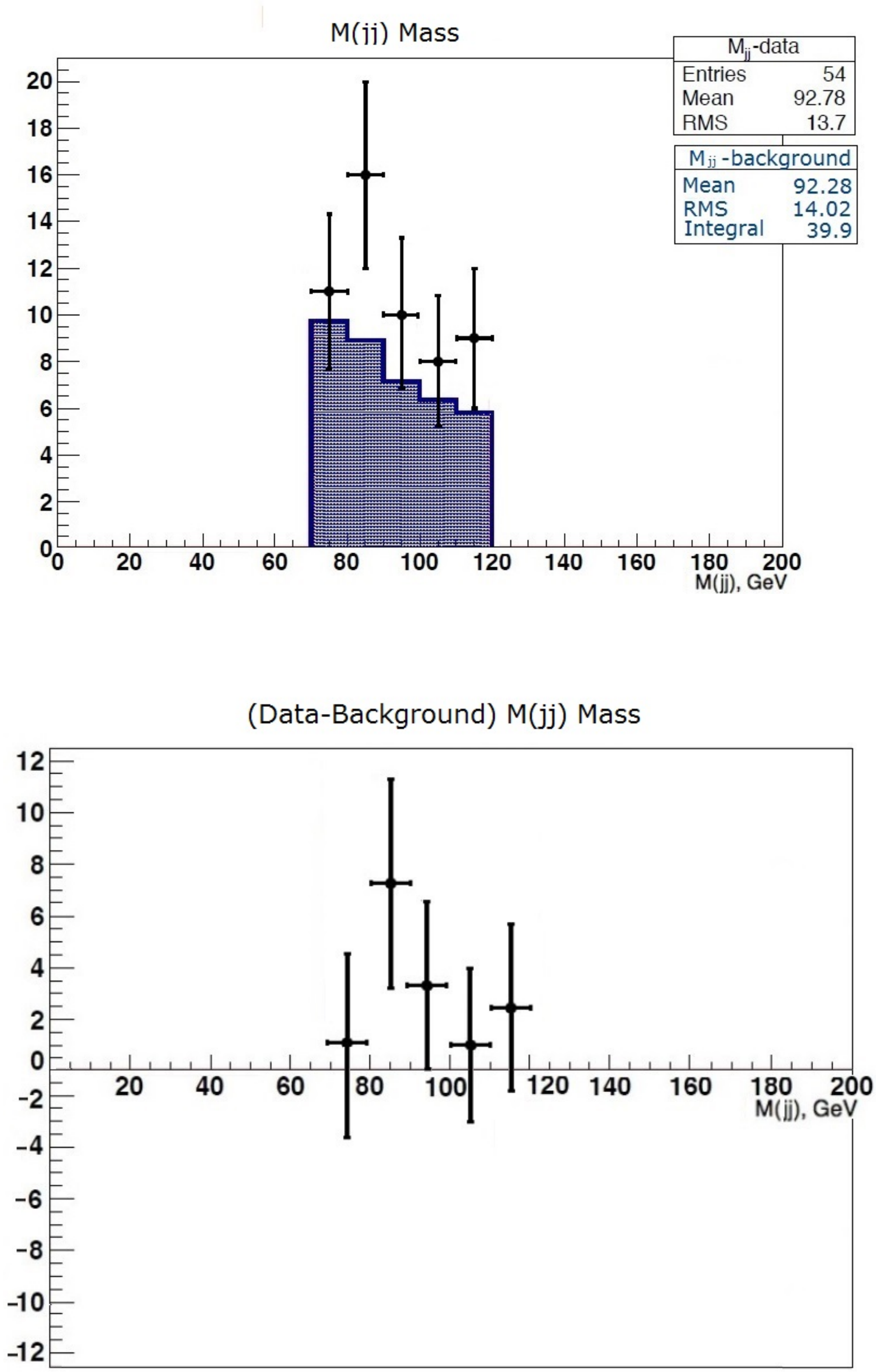


Figure 9: The ee+μμ data (black crosses) and total background (blue) M_{jj} distributions in the **full region** (top) and the bin by bin (Data-Background) Mjj distribution also in the full region (bottom). The Z+jets ALPGEN background has been multiplied by a scale factor equal to 1.04.

From Figure 7 we observe that the excess of data events over the background is in the 320→380 GeV M_{lljj} region and is centered around 350 GeV. In this region the number of data and background events is given in Table 5.

source	[Z(ee)+2jets] + [Z(μμ)+2jets]
Total Background	18.0 ± 3.91
Data	31

Table 5: Total background and data events in the 320 < M_{lljj} <380 GeV region.

From Figure 7 we also observe that the center M_{lljj} value of the excess of events is at 350 GeV whereas the SM background peaks at about 300 GeV, i.e. the event excess peaks at more than 15% over the background peak. This large shift between the data and the SM background peaks cannot be due to data vs MC scale shifts as both the lepton and the jet $P_T|E_T$ scales are equal to better than 2% as can be seen from the means of the relevant distribution (like M_{ll}, leading lepton PT|ET , subleading lepton $P_T|E_T$, jjE_T, leading jE_T) in Table 3 of CDF 10637 [1].

5 Systematic Uncertainties

There are the following sources of systematic uncertainties:

1. Total integrated luminosity: It is taken to be 6%.

2. Lepton-ID: It is taken to be 2% [2].

3 Trigger-Efficiency: Is it taken to be 1% [2].

4. AlPGEN Z+jets Scale Factor: It is taken to be 10%. It represents a possible variation of the SF from going from the control region of M_{lljj}<320GeV to the signal region of M_{lljj}>320 GeV.

5. Jet Energy Scale (JES): It is taken to be 5%. It has been estimated by the procedure described in section 5 of point 6 of [2]. It is a conservative estimate as the up-down variation of the 50 GeV (leading jet E_T threshold), 20 GeV (subleading jet E_T threshold) and 15 GeV (E_T threshold for NO any additional jets) thresholds have been taken to be twice as large as the ones indicated by the data vs MC differences of the relevant variables from Table 4 of [1].

In Table 6 the systematic uncertainties are summarized.

Source	**Systematic Uncertainty (%)**
Luminosity	6
Lepton_ID	2
Trigger Efficiency	1
ALPGEN Z+jets Scale Factor	10
Jet Energy Scale (JES)	5

Table 6: List of systematic uncertainties.

Adding in quadrature the 5 (1,2,3,4 & 5) uncertainties we find the total systematic background uncertainty in the signal region to be ± 12.9% or its absolute value ±3.08 events.

There is a question if an additional systematic could arise from uncertainties in the shape of the simulated distributions (like Mlljj and Mjj) as a result that ALPGEN (the generator used to simulate the Z+jets background which is actually the main one) could be in trouble when predicting jets at large momenta. The answer is that there is NO any indication for such a possibility for the following reasons: a) ALPGEN is describing pretty accurately the leading and subleading jet ET distributions (including the tails) in the control region of CDF 10638 as shown in Figures 8 and 9 of that note. The ET of jets in the full region of this analysis are higher but not by much - their highest value is ~ 150 GeV. The upper cut of 170 GeV on P(Z(ll)) is exactly preventing the acceptance of events with jets with very high momenta, b) Lets look at the tail (over 380 GeV) of the Mlljj distribution of Figure 7 of this analysis (it is actually the distribution used to extract the significance of the excess of data events over the SM background estimate). There are 7 data events there and the ALPGEN background (so normalized as to equal the data in the control region defined as Mlljj < 320 GeV) is just over 6 events i.e. in pretty good, within statistics, agreement with the data, and c) As it can be seen from Figure 16 of [1] (ee+μμ M_{lljj} full region distribution with standard selection criteria) there is very good agreement between data and the background estimate for Mlljj>400 GeV all the way up to ~ M_{lljj}=700 GeV.

6 Significance

The total number of Z(ee)+2jets and Z(μμ)2jets data events in the signal region is 38. The total number of SM background events in the same region is: 23.9 ± 4.99 (stat.) ± 3.08 (syst.). The significance of the excess of the observed events over the expected background was estimated by the use of the routine "pln2" provided by Luc Demortier [7]. The program gives a p-value of $5.06x10^{-2}$ which corresponds to a significance of 1.64 sigma. The inputs to the program are: the # of the observed data events (38), the total expected SM background (23.9) and the total (statistical and systematic uncertainties added in quadrature) relative background uncertainty of 24.5%.

We get a local significance by focusing in the 320 < Mlljj <380 GeV region (see Table 5). The program [7] gives for this region a p-value of 3.53x10-2 which corresponds to a significance of 1.81 sigma. The inputs to the program are: the # of the observed data events (31), the total expected SM background (18 events) and the total (statistical(3.91 events) and systematic (2.32 events) uncertainties added in quadrature) relative background uncertainty of 25.26%

Additional significance is provided by the fact that the excess of events seen in the two jets invariant mass M_{jj} region of 70 - 120 GeV has a maximum very close to the Z-mass value.

7 Conclusions

We have compared to SM expectations the mass of the Zjj system M(lljj) in Z → ll events accompanied by two jets. To discriminate against QCD production of associated jets we have defined a data sample by a number of suitable cuts requesting large PT jets. After accounting for statistical and systematic rate uncertainties an excess of events with a significance 1.64 σ is found at M(lljj)> 320 GeV/c^2. The significance of the excess becomes 1.81 σ at the 320<M(lljj)< 380 GeV/c^2 mass region.

References

[1] CDF Collaboration: "Search for high-mass Z_{jj} resonances", CDF 10637 (Version 2.0), October 2014.

[2] CDF Collaboration: "Search for high-mass Zννbar resonances", CDF 10638 (Version 8.0), October 2014.

[3] CDF Collaboration: "Search for high-mass WW→lνlν resonances", CDF 10959 (Version 3.0), October 2014.

[4] CDF Collaboration: "Search for high-mass ZZ & WW resonances in the ll+MET channel", CDF 11111 (Version 2.0), October 2014.

[5] T. Sjostrand et al,. Comput. Phys. Commun. 135, 238 (2001).

[6] E. Gerchtein and M. Paulini, CHEP-2003-TUMT005.

[7] Luc Demortier - private communication.