Available on CMS information server

CMS AN 2008/000

The Compact Muon Solenoid Experiment

Analysis Note

The content of this note is intended for CMS internal use and distribution only

Nov 20, 2008

Search for an intermediate mass SM Higgs boson at CMS using Vector Boson Fusion $H \rightarrow ZZ \rightarrow 2\mu 2\nu$

Yajun Mao, Bo Zhu

Beijing University, Beijing, China

Dan Green, Rick Vidal, Weimin Wu

Fermi National Accelerator Laboratory, Batavia, IL, USA

Chiara Mariotti

INFN, Torino, Italy

Hesheng Chen, Guoming Chen, Jian Wang[1]

Institute of High Energy Physics, Beijing, China

Yves Sirois

Laboratory Leprince-Ringuet, Ecole Polytechnique, CNRS-IN2P3, Palaiseau, France

Halil Gamsizkan, Mehmet Zeyrek

Middle East Tech University, Ankara, Turkey

Mohammad Alwarawrah, Nural Akchurin, SungWon Lee, Youn Roh, Efe Yazgan

Texas Tech University, Lubbock, Texas, USA

Haifeng Pi, Frank Wuerthwein

University of California San Diego, La Jolla, CA, USA

Ugur Akgun, Erhan Gulmez, Alexi Mestvirishvili, Anthony Moeller, Yasar Onel, Taylan Yetkin

University of Iowa, Iowa City, Iowa, USA

Shuichi kunori

University of Maryland, College Park, Maryland, USA

1 Please send comments to wangj@ihep.ac.cn

Abstract

The discovery potential of the CMS experiment to an intermediate mass SM Higgs produced via Vector Boson Fusion mechanism with the channel $H \rightarrow ZZ \rightarrow 2\mu 2\nu$ was investigated. An event selection was chosen to optimize the expected signal significance for Higgs mass of 200GeV. A signal significance of 5σ can be achieved with $26fb^{-1}$ of integrated luminosity.

1 Introduction

The Higgs mechanism is a cornerstone of the Standard Model and its supersymmetric extensions. The direct search in LEP2 experiments yields a lower bound of 114.4GeV on the Higgs mass. To prove or disprove the existence of the Higgs boson is one of the primary goals of the CMS experiment.

The dominant Higgs production mechanism at LHC will be the gluon-gluon fusion. The process with second largest cross section is Vector Boson Fusion. At low mass range, the Higgs boson mainly decays into bb or $\tau\tau$ pairs. For Higgs mass above 135GeV the dominant decay mode is that into WW pair. If the Higgs is heavier than ~ 2 M_Z, the ZZ branching ratio will increase and become also important. In this note, we focus on a 200GeV Higgs produced via Vector Boson Fusion and decaying into a ZZ pair.

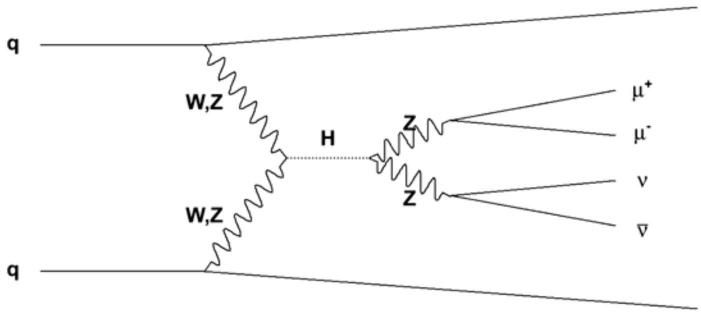

Figure 1: The lowest order Feynman diagram for the considered Higgs boson production and decay channel

The lowest order Feynman diagram for the considered Higgs boson production and decay channel is shown in Figure 1. This process is characterized by two forward jets with modest transverse momentum, separated by a large rapidity gap in which there is a pair of muons and missing energy. The forward jets offer the opportunity to suppress backgrounds significantly.

The present analysis is the first study of the potential to discover a Higgs boson in this channel at intermediate mass range.

2 Event samples

2.1 Signal

The signal events are qqH, $H \rightarrow ZZ \rightarrow 2\mu 2\nu$ with Higgs boson mass equal to 200GeV. 2.6k signal events used in this analysis were generated with PYTHIA6.409 implemented in CMSSW. The Higgs boson was produced with ISUB=123 (WW fusion) and ISUB=124 (ZZ fusion) switched on.

The cross section of qqH at 200GeV is 2.53pb and the branching ratio to a ZZ pair is 26.13%. The branching ratios of a Z boson decaying into two muons or two neutrinos are 3.366% and 20.00% respectively. Therefore the effective cross section for the channel we studied is 2.53pb$\times$26.13%$\times$3.366%$\times$20.00%$\times$2=8.90fb.

The configuration of generator included: initial state radiation, final state radiation, fragmentation, hadronization, multiple parton interaction and underlying event. Full detector simulation based on GEANT4 was used. The simulation and reconstruction of final high level objects such as muons, jets and missing transverse energy were performed via CMSSW_1_6_9.

To match the method producing official CSA07 datasets, we applied four steps in signal sample production: Generation and Simulation, Digitization and DigiToRaw (also L1 Trigger), HLT, Reconstruction.

2.2 Backgrounds

The physics channels that have similar final event topology, two isolated muons of opposite charge together with additional hadronic jets, were considered as background processes:

- *tt + jets*. The cross section for *tt+jets* is expected to be large. The branching ratio of $t \rightarrow Wb$ is close to 100%. If two muons are produced in the decay of the *W*'s, the *tt* can have the same event topology as the signal in the final state. The sample is generated with Madgraph, followed by fast simulation in CMSSW_1_8_4. All jet multiplicities are put together in this sample. We chose to use this sample because the events number corresponds to integrated luminosity of 15fb^{-1}, which is a much larger statistic than any other official *tt* samples.

- *Z + jets* with the *Z* decaying leptonicaly. This process has huge cross section. The decay of *Z* boson into a muon pair results in similar event topology as the signal. Jet energy mis-measurement causes faked missing transverse energy. We used first and second p_T bins' samples for *Z+2j* whereas first p_T bin for *Z+1j* or *3j.* The available sample is small in statistic due to huge cross section.

- *ZZ/WW + jets*. These processes can also fake the signature of the signal. But they have small cross sections, about same order as qqH, and are easy to suppress.

All background samples used in this analysis were taken from the official Monte Carlo production. They are datasets produced during CSA07 except the *tt+jets*.

The details of the signal and backgrounds samples used in this analysis are summarized in Table 1. In this study, we mainly focus on *tt+jets, Z+2/3j* and *ZZ+1/2j*. Contribution of the other backgrounds was found to be negligible.

Table 1: The signal and backgrounds samples used in this analysis

Sample	Total Events	Cross Section	Integrated Luminosity	Dataset Name for official sample
Signal	2.6k	8.90 fb	290 fb^{-1}	
tt + jets	10M	694 pb	15 fb^{-1}	ttnj-madgraph
Z + 1j	950k	940 pb	1.0 fb^{-1}	Z1j-0ptw100-alpgen
Z + 2j	320k	298 pb	1.1 fb^{-1}	Z2j-0ptw100-alpgen, Z2j-100ptw300alpgen
Z + 3j	73k	68 pb	1.1 fb^{-1}	Z3j-0ptw100-alpgen
ZZ + 1j	5.3k	637 fb	8.3 fb^{-1}	zz1j-alpgen
ZZ + 2j	7.3k	247 fb	30 fb^{-1}	zz2j-alpgen
ZZ + 3j	5.4k	239 fb	23 fb^{-1}	zz3j-alpgen
WW +2j	5.9k	4.0 pb	1.5 fb^{-1}	ww2j-alpgen

3 Reconstruction of physics objects

3.1 Muon

We used global muons provided by standard reconstruction algorithm. Global muon algorithm starts from the muon chamber information, adding associated silicon tracker hits and performing a final fit to the track. They also take into account effects of multiple scattering and muon energy losses in the material.

3.2 Jet

Jets were reconstructed using the standard Iterative Cone (IC) algorithm with a cone size of $\Delta R = 0.5$. In detail, a seed calorimeter tower is selected and then all objects sufficiently close in (η,ϕ) are used to form a proto-jet. The process of association is iterated until the parameters of the proto-jet have stabilized, and then the associated towers are considered to comprise a jet candidate. The procedure is repeated with the remaining unassociated towers, until no seeding tower with sufficiently high transverse energy remains. The calorimeter tower is defined by combination of ECAL cells and HCAL towers matching in (η,ϕ) space. Multiple reconstructed jets in the same event are ordered in transverse momentum.

The jet energy was corrected using the standard L2+L3 jet correction packages provided by the JetMET group. The jets were first corrected for differences in detector response due to pseudorapidity, and then further corrected for the variation in detector response for p_T.

We matched two highest p_T jets in signal events to generator level jets with $\Delta R < 0.5$ and the matching efficiency was nearly 100% in all (p_T, η) region. Figure 2 shows p_T ratio of jets after corrections to generator level jets in different p_T regions. It can be seen that both jet energy scale and resolution get better as jet p_T gets larger whereas we can not see such difference in different η regions.

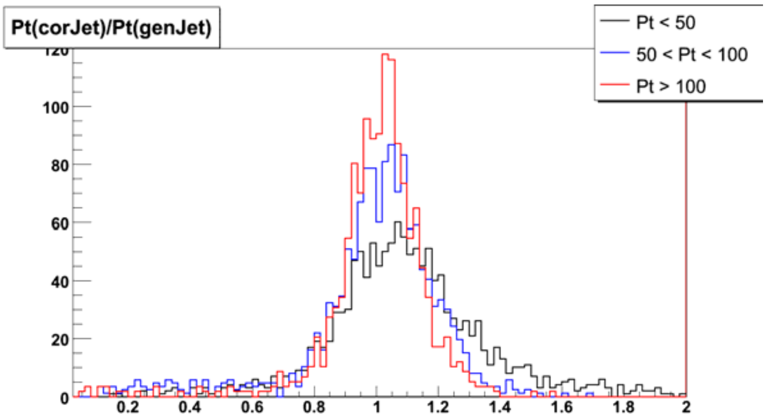

Figure 2: p_T ratio of jets after corrections to generator level jets in different p_T regions

3.3 Missing Transverse Energy

Since the presence of two energetic neutrinos, the missing transverse energy (MET) plays an important role in separating signal events from backgrounds. MET was reconstructed from the vector sum of all ECAL and HCAL tower raw energies. We corrected the MET using the sum of the p_T difference between the corrected and uncorrected jets. The contributions of muons that were not measured in the calorimeter were also corrected. Further details of MET reconstruction performance and analysis will be discussed in Section 4.4.

4 Event selection

4.1 HLT and skimming

The signal is characterized by the presence of two isolated muons and additional jets. Therefore four high level trigger paths were chosen: HLT1MuonIso, HLT1MuonNonIso, HLT2MuonNonIso and HLTXMuonJets. The thresholds for each trigger path in 14TeV physics run at 10^{32}cm^{-2}s^{-1} luminosity are listed in Table 2. A global OR between these HLT paths was chosen to maximize the signal efficiency.

Table 2: Thresholds for each High Level Trigger path used

High level trigger	Thresholds (GeV)
1MuonIso	11
1MuonNonIso	16
2MuonNonIso	(3,3)
XMuonJets	(7,40)

Further suppression of the event flow was obtained via the skimming requiring at least two muons of any sign with p_T > 5GeV. The signal selection efficiency for passing the HLT and skimming was above 75%.

4.2 Preselection

In the signal events, two muons come from the Z boson decay. So the events in which a Z mass peak could not be well reconstructed have no chance to be signal and have to be get rid of. A set of preselection cuts was applied to select the muons truly belonging to the Z boson decay. We required exactly two muons with opposite charge and $p_T > 5$GeV. The pseudorapidity of muons should be in the acceptance of detector, $|\eta| < 2.4$. The invariant mass of the two muons should be in the mass window of 81GeV < $m_{\mu\mu}$ < 101GeV.

The preselection efficiency for *tt+jets* was as low as 0.2%. But it still had 3 orders of magnitude larger cross section than the signal. Figure 3 shows the signal and backgrounds after preselection. We can expect that very tight selection cuts should be used to suppress huge backgrounds and a low signal efficiency may be yielded.

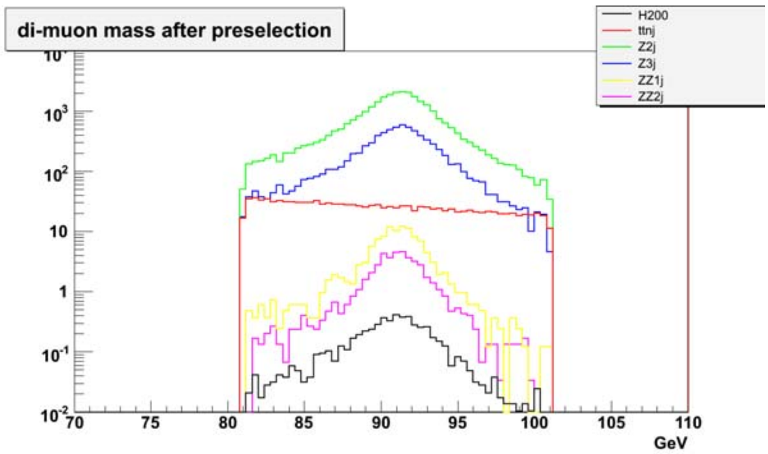

Figure 3: $m_{\mu\mu}$ distribution for the signal and backgrounds after preselection

4.3 Forward jets tagging

The presence of two forward jets is the distinctive signature of VBF Higgs signal. In production, the incoming quarks have high energy and are then scattered by the emission of W or Z bosons. The final state quarks usually have relatively large energy and modest transverse momentum, so the scattering angle with respect to the beam line is small.

At the reconstructed level, tag jets were searched for over the full calorimeter coverage. In our study, tag jets were defined as the two jets with highest transverse momentum. This choice has a high efficiency for correctly identifying the tag jets. For our signal, 60.1% of the events have both tag jets matching with the initial forward partons with $\Delta R < 0.5$ whereas 35.7% have one of the tag jets matching with the forword partons (Figure 4). These efficiencies are higher than those in other alternative definitions of tag jets, such as, jet with highest transverse momentum in each hemisphere, two highest transverse momentum jets above a certain energy threshold, two jets with highest energy, etc. Further more, after all tag jets cuts, the definition used in this analysis gives a better result.

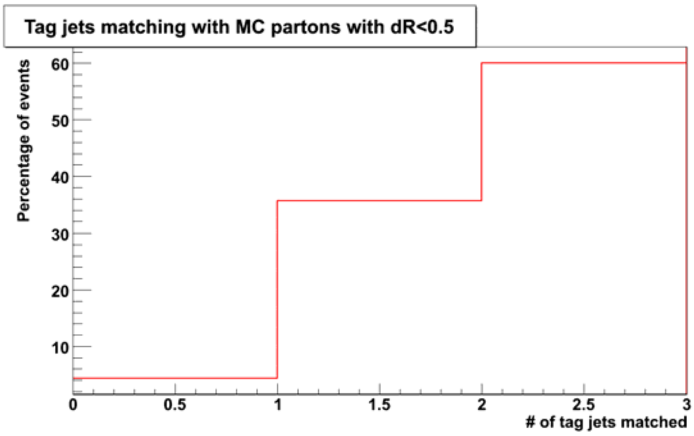

Figure 4: number of tag jets matching with MC partons with $\Delta R < 0.5$

The jet multiplicity distribution for signal and backgrounds events is shown in Figure 5. The p_T threshold of the jets was set 30GeV. That was intended to make as the most possible signal events to peak at the third bin, i.e. exactly two jets have p_T exceed 30GeV.

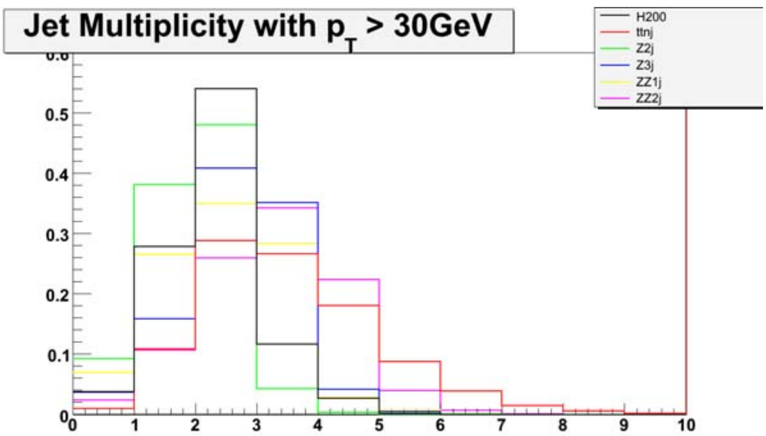

Figure 5: Jet multiplicity with jet $p_T > 30$GeV

The difference in pseudorapidity ($\Delta\eta$) between the two tag jets for the signal and backgrounds is shown in Figure 6. In signal the tag jets are widely separated in pseudorapidity. This shows that a cut $\Delta\eta$ for the tag jets will be effective:

$$\Delta\eta = |\eta_{j1} - \eta_{j2}| > 4.2$$

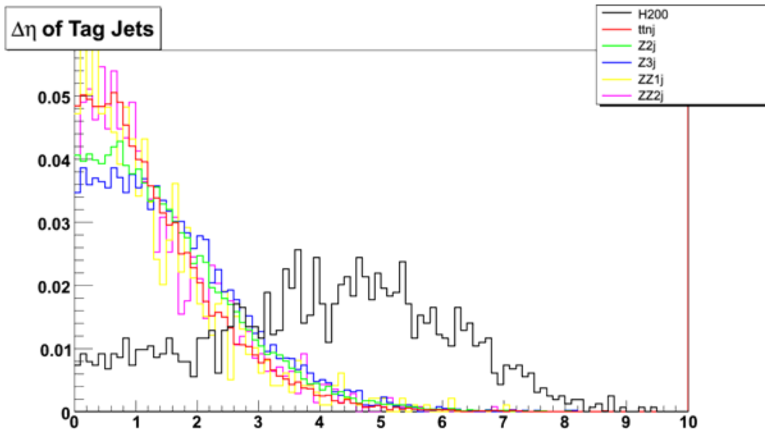

Figure 6: $\Delta\eta$ distribution for the signal and backgrounds

A further consequence of the tag jets being high energy and widely separated is that the invariant mass of the tag jets will be relatively large compared with the background processes (Figure 7). Therefore a cut was applied on the minimum invariant mass of the tag jets:

$$M_{jj} > 800GeV$$

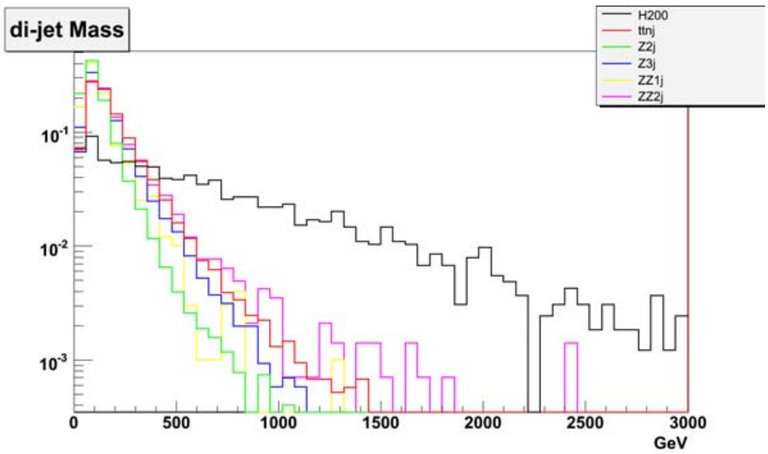

Figure 7: the invariant mass of tag jets for the signal and backgrounds

At last both muons were required to lie in between the tag jets in pseudorapidity and to be separated from the jets:

$$\eta_{jet}^{low} + 0.7 < \eta_{\mu} < \eta_{jet}^{high} - 0.7$$

4.4 MET correlation with jets

After forward jets tagging cuts, *Z+jets(2/3j)* still have a significant contribution because of huge cross section. It should be further suppressed by the MET difference with signal events.

As Figure 8 shows, the signal events have large true MET in generator level due to presence of two energetic neutrinos. For *Z+jets* events, MET is primarily a detector effect. So a minimum MET threshold would benefit the signal. However, the threshold could not be high enough to remove the MET tail in *Z+jets*, otherwise the signal efficiency would be largely reduced.

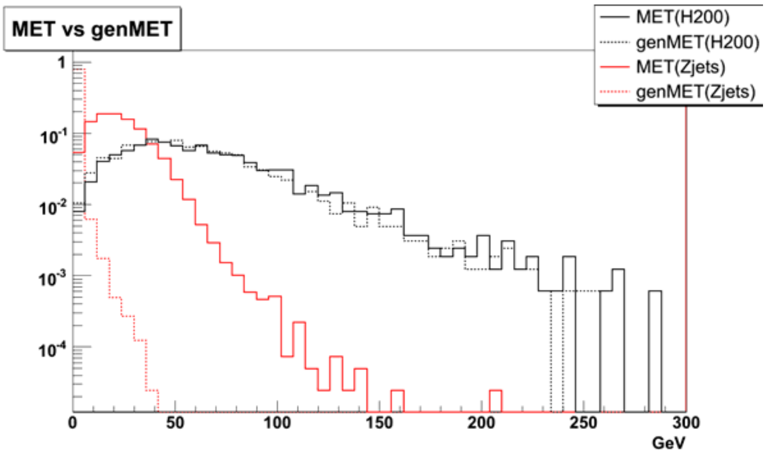

Figure 8: MET and generator level MET distribution for the signal and *Z+jets*

By Projecting MET to the direction of leading jet, which means the highest p_T jet, a more distinct difference could be found. The signal events usually have large negative components in leading jet direction. In *Z+jets*, a perfect Gaussian distribution could be seen just as expected for a detector effect. Remember that in Figure 2 the mean value of jet energy scale is larger than 1 especially when p_T is not large. That causes that we get a bigger chance to over-measure a jet than under-measure it. So the MET component in jet direction has a mean value smaller than 0, just as Figure 9 shows.

A cut on MET component in leading jet direction is very powerful to discriminate the signal and *Z+jets*. We required

$$MET \cos\phi < -40GeV$$

ϕ is the azimuth angle between MET and the leading jet.

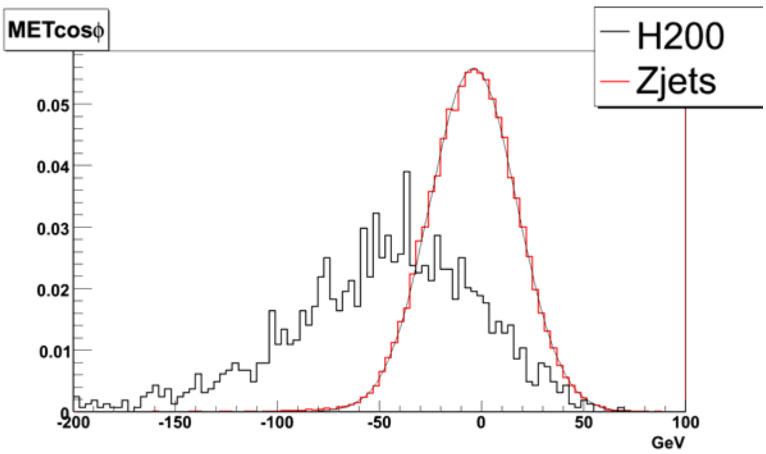

Figure 9: MET component in leading jet direction

5 Results

The results presented in terms of accepted cross sections after application of each selection for the signal and backgrounds are summarized in Table 3. The total accepted signal cross section is 0.72fb and the total signal efficiency is 8.1%.

Table 3: accepted cross sections in fb after application of each selection for the signal and backgrounds

Selection Cut	Signal	*tt+jets*	*Z+2j*	*Z+3j*	*ZZ+1j*	*ZZ+2j*
	8.90	694k	298k	68k	637	247
preselection	5.67	1.28k	29.4k	7.83k	120	47.7
jets multiplicity	3.06	369	14.1k	3.19k	41.9	12.3
jets $\Delta\eta$	1.74	3.66	259	93.0	0.24	0.36
di-jet mass	1.31	0.80	43.6	16.4	0	0.13
mu between jets	1.14	0.20	20.0	8.18	0	0.10
MET	0.72	0.13	0	0	0	0

5.1 Z+jets estimation

Tight selection cuts were used due to huge backgrounds. After all selection cuts applied, no *Z+jets* events survived. The *Z+jets* sample we used has very low statistics (~ 1.1fb^{-1}). Since ~ 30 events could be found before the final selection, i.e. MET cut, we loosed the MET cut step by step and fit the number of surviving events versus the threshold of cut.

As we mentioned, for *Z+jets* the MET component in leading jet direction has a Gaussian distribution (Figure 9). So the number of surviving events after MET cut could be well fitted by the cumulative distribution function of Gaussian function:

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\, dt$$

σ is the standard deviation and μ is the mean value, which both could be obtained from the original Gaussian distribution. Fitting with the above function, we expected 0.22fb in final cross section for *Z+jets* (Figure 10).

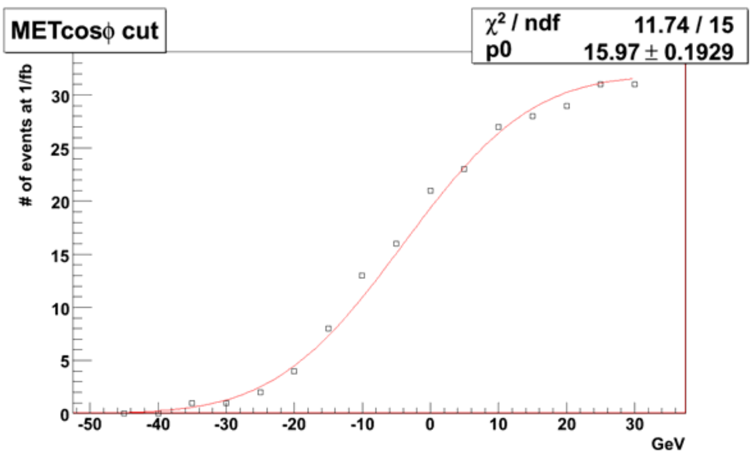

Figure 10: *Z+jets* estimation using number of events v.s. MET cut

When it comes to *Z+1j* or *ZZ+3j*, they have similar selection efficiencies compared to *Z+2/3j* or *ZZ+1/2j* respectively, except that a significant larger reduction factor could be seen in jet multiplicity cut. As a result, their contributions could be neglected.

5.2 Higgs transverse mass

We lack sufficient information to reconstruct the invariant mass of the Higgs boson because of presence of neutrinos. Instead, a transverse mass of Higgs is defined:

$$M_T = \sqrt{(\sqrt{P_{T,\mu\mu}{}^2 + M_{\mu\mu}{}^2} + \sqrt{P_{T,\nu\nu}{}^2 + M_Z{}^2})^2 - (\vec{P}_{T,\mu\mu} + \vec{P}_{T,\nu\nu})^2}$$

As Figure 11 shows, The signal transverse mass distribution is of Gaussian shape. However, as a result of small statistics of background events in the final result, it is difficult to conserve a distinct background shape.

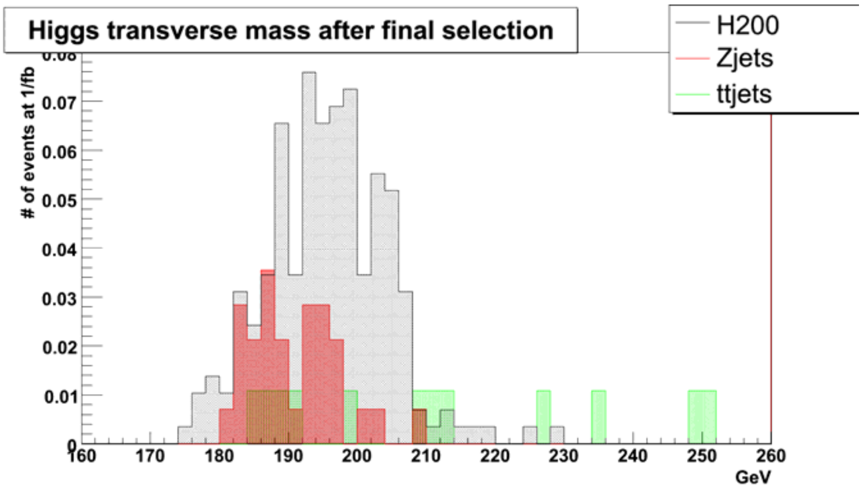

Figure 11: Higgs transverse mass distribution after all selection

5.3 Significance and uncertainties

The significance was obtained based on the likelihood ratio,

$$S_L = \sqrt{2\ln Q}$$

Q is the likelihood ratio. As to the simple approach of the counting method,

$$S_{cL} = \sqrt{2((s+b)\ln(1+\frac{s}{b}) - s)}$$

s and b are the expected numbers of the signal and background events at a given luminosity condition. In this analysis, we got $S_{cL} = 5$ at 26fb^{-1} and $S_{cL} = 3$ at 10fb^{-1}.

The most important systematic errors for this analysis are the uncertainties of jet energy scale and missing energy scale. According to the recommended treatment for the jet energy systematic in [7], we changed jets energy by $\pm$10% for jet p_T < 20GeV, $\pm$3% for jet p_T > 50GeV and $\pm$(0.1-0.07(p_T-20)/30) for jet p_T between 20 and 50 GeV. We found 3% change is significance. The systematic uncertainty of MET is correlated with that of jet energy. We also changed MET by $\pm$10%. This caused about 10% changes in the significance. For integrated luminosity our analysis based on, it is assumed that the uncertainty of luminosity measurement is 3%. This affects the result by 2%. So the total detector systematic uncertainty is 11%.

Additional improvements might be achieved if more Monte Carlo events available. Muon isolation and Central Jet Veto techniques were carefully studied as well, but turned out to be not useful to improve the significance. For higher Higgs mass, a stronger signal would be expected.

6 Conclusions

We have presented a new analysis of Higgs production via Vector Boson Fusion and $H \rightarrow ZZ \rightarrow 2\mu 2\nu$ at CMS. Forward jets tagging is a powerful tool to reject backgrounds. The results of our study are promising. An excess signal with a significance of 5σ can be achieved for Higgs mass of 200GeV after data from CMS corresponding to an integrated luminosity of 26fb^{-1} have been taken. Our analysis indicates that this channel can contribute to the discovery of an intermediate mass Higgs boson.

7 Acknowledgments

We are grateful to Weimin Wu and Haifeng Pi for their valuable comments and suggestions on this study.

References

[1] T. Sjostrand et al., PYTHIA physics and manual., JHEP 0605, (2006) 026

[2] P. Schieferdecker et al., Peformance of Jet Algorithms in CMS, CMS AN 2008/001

[3] R. Harris et al., MC Truth L2&L3 Factorized Jet Corrections at CMS, CMS AN 2008/003

[4] H. Pi et al., Measurement of Missing Transverse Energy With the CMS Detector at the LHC, CMS NOTE 2006/035

[5] S. Esen et al., MET Performance in CMS, CMS AN 2007/041

[6] CMS Collaboration, CMS Physics Technical Design Report Volume I: Detector Performance and Software, CERN/LHCC 2006-001, 2006

[7] CMS Collaboration, CMS Physics Technical Design Report Volume II: Physics Performance, CERN/LHCC 2006-026, 2006

[8] V. Bartsch et al., Expected Signal Observability at Future Experiments, CMS Note 2005/004

[9] The CMS Offline Guide, https://twiki.cern.ch/twiki/bin/view/CMS/SWGuide

[10] DBS data discovery page, https://cmsweb.cern.ch/dbs_discovery/

[11] M.Pierini et al., Study of V+jets in bins of boson Pt and jet multiplicity for 2007 1fb^{-1} Monte Carlo production at CMS, CMS IN 2007/031