Special Article – Scientific Note

THE EUROPEAN PHYSICAL JOURNAL C

Search for a standard model Higgs boson in CMS via vector boson fusion in the $H \rightarrow WW \rightarrow l\nu l\nu$ channel

E. Yazgan^{1,2,a}, J. Damgov^{3,2}, N. Akchurin⁴, V. Genchev³, D. Green², S. Kunori⁵, M. Schmitt⁶, W. Wu², M.T. Zeyrek¹

¹ Middle East Technical University, Ankara, Turkey

² Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, 60510-5011, Illinois, USA

³ Institute for Nuclear Research and Nuclear Energy, Bulgaria Academy of Science, Sofia, Bulgaria

⁴ Texas Tech University, Lubbock, Texas, USA

⁵ University of Maryland, College Park, Maryland, USA

⁶ Northwestern University, Evanston, Illinois, USA

Received: 6 June 2007 / Revised version: 14 November 2007 / Published online: 4 December 2007 – © Springer-Verlag / Società Italiana di Fisica 2007

Abstract. We present the potential for discovering the standard model Higgs boson produced via the vectorboson fusion mechanism. We considered the decay of Higgs bosons to the W^+W^- final state, with both W-bosons subsequently decaying leptonically. The main background is $t\bar{t}$ produced in association with one or more jets. This study is based on a full simulation of the CMS detector. The result is that a signal of 5σ significance can be obtained with an integrated luminosity of 12-72 fb⁻¹ for Higgs boson masses in the range $130 < m_H < 200$ GeV. In addition, the major background can be measured directly to 7% from the data with an integrated luminosity of 30 fb⁻¹. We also suggest a method to determine the Higgs mass using template transverse mass distributions.

PACS. 14.80.Bn

1 Introduction

One of the primary goals of CMS is to search for the Higgs boson. The LEP experiments set a lower limit on the standard model (SM) Higgs boson of 114.4 GeV at a 95% C.L. [1], and unitarity puts an upper limit of about 1 TeV [2,3]. Tighter constraints are obtained from fits to precision electroweak measurements, which limit the mass of a standard model-like Higgs boson to be less than $194\,{\rm GeV}$ [4] at 95% C.L. In extended Higgs sectors, there is often one scalar boson that resembles the Higgs boson of the standard model, and is responsible for electroweak symmetry-breaking. The mass of such a Higgs must also approximately satisfy the constraints mentioned above. In the minimal supersymmetric extension of the standard model (MSSM), there is a more stringent bound which comes from the internal constraints of the theory; the lightest Higgs boson must have mass less than about 135 GeV. For these reasons, we focus on the mass region $120 < m_H < 200 \text{ GeV}.$

The two main decay modes of the standard model Higgs boson in this mass range are $H \rightarrow b\bar{b}$ and $H \rightarrow W^+W^-$. In the latter case, one of the W bosons indicated by W^{*} may be off mass shell. If the Higgs boson is heavier than about

135 GeV, the WW^* branching fraction will dominate, remaining sizable for masses as low as 120 GeV. In this study, we consider the decay $H \to WW^*$ with the subsequent decay of the W-bosons to two charged leptons.

Higgs bosons may be produced in pp collisions when radiated off the virtual W-boson that is exchanged in the t-channel – this is called "vector boson fusion" (VBF). The Feynman diagram for this process is shown in Fig. 1. This channel has good prospects for the discovery of a standard model Higgs boson, especially if it is not too heavy because of the distinctive VBF topology which contains two jets with small angles with respect to the beam axis. Furthermore, when the Higgs decays to two W-bosons, the presence of the HWW vertex both in the production and the decay of the Higgs boson gives a relatively clean determination of the HWW coupling. Given that the Higgs mass completely determines all Higgs properties within the standard model, measurement of the HWW coupling overconstrains the SM. This will be crucial to establishing the origin of electroweak symmetry breaking.

The VBF mechanism was proposed as a potential discovery channel in 2001 [5]. Our initial study of this channel for the CMS detector was carried out in 2002 [6], with a number of simplifications. The conclusion of this previous CMS study was that a convincing signal for a Higgs boson with a mass of 120 GeV would be observed with about

^a e-mail: efe@fnal.gov



Fig. 1. Feynman diagram for Higgs boson production through vector boson fusion. The Higgs boson decays into *W*'s which further decay into electron/muon–neutrino pairs

70 fb⁻¹. In the present study, we repeat the entire analysis in the mass range 120-200 GeV, using the simulation and reconstruction software for CMS in order to verify and improve the 2002 study. A similar study of this channel was performed for the ATLAS detector in 2004 using different event generators and slightly different cuts [7].

The VBF process is characterized by two forward jets with modest transverse momentum, $E_{\rm T} \approx m_W/2$, separated by a large rapidity difference. The Higgs boson is produced at low rapidity, with a pair of clean, isolated leptons and missing energy. The main backgrounds for this channel are irreducible continuum W^+W^- background, and $t\bar{t}$ in which both top quarks decay semi-leptonically. These backgrounds are particularly troublesome when there are additional jets, j, in the event, so particular attention was paid to the generation of W^+W^-jj and $t\bar{t}j$ events.

2 Event generation

The signal process and the W^+W^-jj background have been simulated on the basis of a leading order (LO) matrix-element calculation using MadGraph [8]. For the $t\bar{t}j$ background, we used the LO AlpGen [9] package which correctly simulates spin correlations. We simulated the parton showers using Pythia [10]. The parton distribution functions used in MadGraph and AlpGen simulations are CTEQ6L1 and CTEQ5L1 respectively [8,9]. The minimum transverse momentum cut on jets is 15 GeV,

 Table 1. Production cross-section for the signal and main backgrounds

Channel	Cross-section [pb]	WW BR	$\sigma \times BR \ [pb]$
qqH m = 120	4.549	0.133	0.605
qqH m = 130	4.060	0.289	1.173
qqH m = 140	3.648	0.486	1.773
qqH m = 160	3.011	0.902	2.715
qqH m = 180	2.542	0.935	2.376
qqH m = 200	2.177	0.735	1.600
ttj	736.5	1.0	736.5
WWjj QCD	43.6	1.0	43.6
WWjj EW	0.933	1.0	0.933

and the jet pseudo-rapidity is limited to $|\eta| < 5$. We required a separation of any jet pair, $\Delta R > 0.5$, where $\Delta R = \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2}$.

Next-to-leading order (NLO) cross-sections differ from LO cross-sections by ~ 30% for a 120 GeV Higgs boson and ~ 10% for a 200 GeV Higgs boson [11]. However, since there are no NLO cross-section calculations for the backgrounds, the LO cross-sections are used consistently for both signal and background processes in this study. The cross sections are listed in Table 1. The 'electroweak' (EW) part of the W^+W^-jj process is defined as the subsample with no $\alpha_{\rm s}$ -dependent vertex in the diagrams, and the 'QCD' part is the rest of this process. The EW part is topologically very similar to the signal and hence is almost irreducible.

3 Detector simulation and event reconstruction

We processed the generated events through the CMS detector simulation software which is based on Geant-4 simulation. We simulated pile-up from out-of-time interactions representing the low luminosity LHC running condition $(\sim 2 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1})$. Subsequently, we processed digitized information using the CMS event reconstruction software.

3.1 Trigger

We refer to [12] for the presently planned trigger table. The inclusive single electron trigger has an $E_{\rm T}$ -threshold of 26 GeV, which is too high for our purposes. Therefore we will augment this trigger with the di-electron trigger, which has a threshold of 12 GeV for both electrons. The $p_{\rm T}$ -threshold for the inclusive single muon trigger is 19 GeV, which is well suited to this analysis. Concerning the $e_{-\mu}$ channel, we plan to use the $e + \mu$ di-lepton trigger, which will have a threshold of 10 GeV for each lepton. The efficiency for the L1 + HLT trigger with respect to our offline cuts varies from about 95% to 99% based on [13]. This small inefficiency does not affect the results of our analysis.

Lepton + jet triggers could also be included to give lower lepton thresholds. Since the details for these triggers are not yet available, we have based our study solely on the leptonic triggers.

3.2 Lepton reconstruction and identification

We have used standard packages and selection criteria for muon and electron identification. Below, we describe our estimate of the identification efficiency.

3.2.1 Muons

We use the "global" muon reconstruction, which takes muons found in the muon chambers (drift tubes, cathode strip chambers, and RPC's), and extrapolates them



Fig. 2. $\Delta \eta = |\eta_1 - \eta_2|$ distribution for the forward tagging jets which have $E_{\text{T1}} > 50$ GeV and $E_{\text{T2}} > 30$ GeV for (a) qqH, $m_H = 120$ GeV and backgrounds (b) $t\bar{t}j$, (c) EW WW jj and (d) QCD WW jj. Note that the EW WW jj $\Delta \eta$ distribution is very similar to the signal $\Delta \eta$ distribution

into the silicon tracker to pick up additional hits to better define the kinematics. This extrapolation takes into account the energy lost by the muon as well as multiple scattering.

Muons are found within $|\eta| < 2.4$. The overall muon reconstruction efficiency in this angular range is $\approx 95\%$ for $10 < p_T < 30$ GeV and 97% for $p_T > 30$ GeV.

3.2.2 Electrons

Electrons are reconstructed by combining superclusters [14, 15] and Kalman tracks [16]. The track – supercluster (SC) matching condition is $\Delta R < 0.15$. Such tracks should have at least four hits, and transverse momentum $p_{\rm T} > 5$ GeV. If several tracks satisfy these conditions, then the one having the least difference $|p_{\rm T} - E_{\rm T}|$ is taken. We reject the electron candidates if $E_{\rm T}^{\rm SC} < 10$ GeV or $|\eta^{\rm SC}| > 2.0$. The probability for a generator level electron with $p_{\rm T} > 10$ GeV and $|\eta| < 2.0$ to be reconstructed within $\Delta R < 0.2$ is $\sim 92\% - 98\%$ for $10 < p_{\rm T}(\text{gen}) < 20$ GeV and $\sim 98\% - 99\%$ for $p_{\rm T} > 20$ GeV. These reconstructed electrons are said to be identified if they satisfy $E_{\rm HCAL}/E_{\rm ECAL} < 0.05$, $|\Delta \eta(\text{trk},\text{SC})| < 0.005$, $E^{\rm SC}/p^{\rm trk} > 0.8$ and $|1/E^{\rm SC} - 1/p^{\rm trk}| < 0.06$.

An isolation variable is defined by taking the sum of the $p_{\rm T}$ of all the tracks (except the electron candidate) within a cone of $\Delta R^{\rm SC} < 0.2$, and dividing by $E_{\rm T}^{\rm SC}$. The tracks entering this sum must have at least four hits, $p_{\rm T} > 0.9$ GeV, and $|z^{\rm trk} - z^e| < 0.4$ cm, where z is the position of the track along the beam line. We place the requirement that this isolation ratio be smaller than 0.2. The overall single electron efficiency for electron isolation and identification is $\approx 80\%$ for $10 < p_{\rm T} < 30$ GeV and $\approx 90\%$ for $p_{\rm T} > 30$ GeV. The electron fake rate per jet, which is calculated using the jets from W decay in the associated production and using the forward jets in the qqH sam-



Fig. 3. $\eta_0 = \eta_3 - (\eta_1 + \eta_2)/2$ for the third jet. η of the third jet with respect to the average of the two forward jets. For signal (a) qqH, $m_H = 120$ GeV and backgrounds (b) $t\bar{t}j$, (c) EW WW jj and (d) QCD WW jj

ple, is $\approx 3\%$ for $10 < p_{\rm T}^j < 30$ GeV and less than $\approx 0.1\%$ for $p_{\rm T}^j > 120$ GeV.

3.3 Jet and missing E_{T} reconstruction and correction

The cell-level thresholds are set at least 2σ above the noise level to remove the effects of calorimeter noise fluctuations in jet reconstruction. This is important since we are mainly dealing with quite low- $p_{\rm T}$ jets in the current study.

We reconstructed the jets using the "iterative cone" algorithm, with a cone size of $\Delta R = 0.5$ and a cone seed $E_{\rm T}$ cut of 1 GeV. We removed the jets from an event if they match a reconstructed electron within a cone of $\Delta R < 0.45$.

We calibrated the reconstructed jets using the qqHsignal sample. Reconstructed jets are first matched to generator level jets within a cone of $\Delta R < 0.12$. We fit the jet response to second-order polynomials as a function of generator-level jet $E_{\rm T}$ for 20 different η regions covering $\eta = 0$ to $\eta = 4$ in bins of $\Delta \eta = 0.2$. The difference between the corrected and uncorrected responses varies by 10% to 30% depending on the jet $E_{\rm T}$ and η values. When applying the correction to jets with $|\eta| > 4$, we used the correction parameters for the last interval $|\eta| = 3.8$ –4.0. The polynomial extrapolation is unreliable beyond $p_{\rm T} = 200$ GeV, so we fixed the corrections above 200 GeV to those obtained at 200 GeV. The response to jets in the QCD di-jet sample is lower than the response to jets in the qqH sample. This produces different correction functions. However, in the current study, VBF tag jets are at high η and have at least $p_{\rm T} > 30$ GeV and for this part of phase space the differences between responses (or equivalently, the jet correction functions) are very small.

In the analysis, we used missing $E_{\rm T}$ ($\not\!\!E_{\rm T}$) calculated from calorimeter hits. We corrected the $\not\!\!E_{\rm T}$ using the sum of the $E_{\rm T}$ difference between the corrected and uncorrected jets for which the corrected jets have $E_{\rm T} > 30$ GeV.



Fig. 4. Fake central jets fraction per event as a function of $E_{\rm T}$ veto threshold. A fake is defined as the probability to find at least one jet (due to pile-up) satisfying the central jet veto conditions, with no "real" jets satisfying the central jet veto condition in that event



Fig. 6. Electron $p_{\rm T}$ spectra for the signal process, qqH, when $m_H = 120~{\rm GeV}$



Fig. 5. The effect of the $E_{\rm T}$ threshold for the central jet veto. For (a) signal events, qqH with $m_H = 120$ GeV and background events (b) $t\bar{t}j$. (c) the S/B ratio and (d) the significance for a 30 fb⁻¹ integrated luminosity



Fig. 7. Centrality of the leptons, using the quantity η'_{ℓ} defined in the text for (a) qqH, $m_H = 120$ GeV and backgrounds (b) $t\bar{t}j$, (c) EW WW jj and (d) QCD WW jj

4 Event selection

The strategy of this analysis is straightforward. We select events with two forward jets separated by a large rapidity difference, veto any event with additional central jets, and demand two energetic, isolated leptons in the central region. Finally, we apply additional cuts on the kinematics and the event topology to reduce backgrounds.

4.1 Forward jet tagging

The jets are ordered in $E_{\rm T}$ after the corrections have been applied. The first two tag jets should be energetic, so we require $E_{\rm T1} > 50$ GeV and $E_{\rm T2} > 30$ GeV. Figure 2 shows the rapidity separation $|\Delta \eta|$ between these two most energetic jets, for the signal (a) and the backgrounds (b–d). It is clear that the jets for signal events are well separated in rapidity, and we apply the cut $|\Delta \eta| > 4.2$. We also make sure that they fall in opposite laboratory hemispheres by requiring $\eta_1 \cdot \eta_2 < 0$.

4.2 Central jet veto

In the signal process, there is no color exchange between the protons, and consequently any additional jets will tend to be radiated in the forward direction. In contrast, the backgrounds will tend to have additional jets in the central region, especially the $t\bar{t}j$ process. We take advantage of this distinction by vetoing events with additional jets in the central region. In particular, we consider any jet with $E_{\rm T3} > 20$ GeV and compute the rapidity with respect to the average of the two forward jets: $\eta_0 = \eta_3 - (\eta_1 + \eta_2)/2$. We veto the event if $|\eta_0| < 2$. See Fig. 3 for distributions of both signal and background.

The probability of finding a fake jet from pile-up events for low luminosity LHC running is shown in Fig. 4 as



Fig. 8. Invariant mass distributions for the two forward tag jets, for (a) qqH, $m_H = 120$ GeV and backgrounds (b) $t\bar{t}j$, (c) EW WWjj and (d) QCD WWjj

a function of the $E_{\rm T}$ threshold for the central jet veto. The fake rate is defined as the rate for pile-up jets satisfying the central jet veto condition in an event where there are no real jets satisfying those conditions. Therefore, the fake rate is just the rate of events mistakenly rejected due to pile-up. The loss of signal events for an E_T threshold of 20 GeV is only about 2%.

The effect of the $E_{\rm T}$ threshold for the central jets on the final cross sections and significances for the 120 GeV signal and for the background are displayed in Fig. 5. Here, the significance is defined as S/\sqrt{B} , where S and B represent the numbers of signal and background events.

4.3 Lepton kinematics

We require two opposite-sign leptons in an event. The most energetic lepton must have $p_{T1}^l > 20 \text{ GeV}$, and the other, $p_{T2}^l > 10 \text{ GeV}$. The p_T -threshold for the second lepton must be low since one of the two W's in the Higgs

decay is off the mass shell for low Higgs masses. Figure 6 shows the $p_{\rm T}$ spectra for electrons in the signal process $(m_H = 120 \text{ GeV})$. We reject events with more than two leptons. The two leptons must be well separated from all jets with $\Delta R_{\ell j} > 0.7$.

In light of the thresholds for the electron triggers, we modified our $p_{\rm T}$ requirements slightly in the di-electron channel. An event is selected if it has two electrons which satisfy:

or

$$(p_{T1}^l > 26 \text{ GeV} \text{ and } p_{T2}^l > 10 \text{ GeV})$$

$$(p_{\rm T1}^l > 20 \,{\rm GeV} \text{ and } p_{\rm T2}^l > 12 \,{\rm GeV})$$
.

Since the leptons come from the W's that come from the centrally-produced Higgs boson, we require them to be central. If η_{hi}^{j} is the forward-tag jet having higher-rapidity, and η_{lo}^{j} is that of the lower-rapidity forward-tag jet, then our requirement can be written as $\eta_{lo}^{j} + 0.6 < \eta_{\ell} < \eta_{hi}^{j} - 0.6$.



Fig. 9. The overall $p_{\rm T}$ -balance in the event. See the text for an explanation. For (a) qqH, $m_H = 120$ GeV and backgrounds (b) $t\bar{t}j$, (c) EW WW jj and (d) QCD WW jj

This condition must be satisfied by both leptons. Figure 7 shows the distributions of the related quantity, $\eta'_{\ell} = (\eta_{\ell} - (\eta^j_{\rm hi} + \eta^j_{\rm lo})/2) \times 4.2/\Delta\eta$, where $\Delta\eta$ is the η difference between the forward and backward-tag jets. This quantity is sensitive to the η distribution of leptons with respect to the forward tag jets.

4.4 Further kinematic requirements

After the forward-jet tag, the central jet veto, and the lepton kinematics cuts, we are left with a sample which still has a large contamination from background processes. We can further reduce this contamination with some additional kinematic cuts.

For the jets, we require the di-jet mass to be greater than 600 GeV (see Fig. 8). Next, we look at the overall $p_{\rm T}$ -balance in the event, by computing the vector sum of the transverse momenta of the two leading jets, the leptons, and the missing energy. The magnitude of that sum should be less than 40 GeV (see Fig. 9).

4.5 Additional cuts

Additional cuts may be required to reduce bbjj and $\tau\tau jj$ backgrounds. These cuts imported from [5] are



Fig. 10. Di-lepton invariant mass distribution after jet and lepton cuts, for (a) qqH, $m_H = 120$ GeV and backgrounds (b) $t\bar{t}j$, (c) EW WW jj and (d) QCD WW jj

 $\begin{array}{l} 57.3\Delta\phi(\ell\ell,\not\!\!\!E_{\rm T})+1.5p_{\rm T}^{\rm Higgs}>180~{\rm and}~12\times57.29\Delta\phi(\ell\ell,\not\!\!\!E_{\rm T})\\ +p_{\rm T}^{\rm Higgs}>360~({\rm where}~\Delta\phi(\ell\ell,\not\!\!\!E_{\rm T})~{\rm is~in}~{\rm rad}~{\rm and}~p_{\rm T}^{\rm Higgs}~{\rm is}\\ {\rm in}~{\rm GeV}~{\rm units}),~{\rm and}~{\rm also}~\not\!\!\!E_{\rm T}>30~{\rm GeV}~{\rm if}~p_{\rm T}^{\rm Higgs}<50~{\rm GeV}\\ {\rm where}~p_{\rm T}^{\rm Higgs}~{\rm is}~{\rm th}~{\rm vector}~{\rm sum}~{\rm of}~{\rm th}~{\rm transverse}~{\rm energy}~{\rm of}~{\rm tag}~{\rm jets}.~{\rm Th}~{\rm distribution}~{\rm of}~{\rm signal}~{\rm events}~{\rm in}~{\rm th}\\ \Delta\phi(\ell\ell,\not\!\!\!\!E_{\rm T})-p_{\rm T}^{\rm Higgs}~{\rm plane}~{\rm is}~{\rm displayed}~{\rm in}~{\rm Fig}.~{\rm 15}. \end{array}$

The Drell–Yan production of di-lepton pairs, $\gamma^* \rightarrow \ell^+ \ell^-$, has a large cross-section. In order to reduce this background sufficiently, we impose a di-lepton mass cut $M_{\ell\ell} > 10 \text{ GeV}$ and we require $E_T > 30 \text{ GeV}$ when the leptons have the same flavor (see [5]).

Finally, we impose the cut $\Delta \phi(\ell \ell, E_{\rm T}) + \Delta \phi(\ell \ell) < 3$ rad, which increases the signal-to-background ratio. Figure 16 shows distributions of this quantity. The resolution of the quantity $\Delta \phi(\ell \ell, E_{\rm T})$ is improved by the $E_{\rm T}$ correction. The additional cuts imposed after the transverse mass cut were determined for generator level analysis. Therefore, we did not include these cuts in the significance, background or mass estimation and their effects are separately shown in Tables 4–6. Work is in progress to confirm their effects after full detector simulations.

5 Results

The total accepted signal cross-sections range from about 0.8 fb up to 7.2 fb, depending on the Higgs mass. They are listed in Table 2. The contributions from the e^+e^- and $\mu^+\mu^-$ channel are very similar, and the $e^\pm\mu^\mp$ channels are twice as large due to branching ratios. The total efficiency is 3%-6%, depending on m_H . The background cross-sections are somewhat larger, and there are two background values corresponding to the "low-mass" and the "high-mass" cuts – see Table 2.

We computed the significance S_{cP} of an excess of events over the $t\bar{t}j$ and W^+W^-jj backgrounds, assuming an integrated luminosity of $\mathcal{L} = 10$, 30 and 100 fb⁻¹. S_{cP} is



Fig. 11. The distribution of the difference in azimuthal angle between the two leptons, $\Delta \phi$ after jet and lepton cuts, for (a) signal events, qqH, $m_H = 120$ GeV and backgrounds (b) $t\bar{t}j$, (c) EW WWjj and (d) QCD WWjj

the probability calculated assuming a Poisson distribution with $N_{\rm B}$ background events to observe equal or greater than a total number of signal and background events ($N_{\rm S} + N_{\rm B}$), converted to an equivalent number of sigmas for a Gaussian distribution [17]. The code to calculate $S_{\rm cP}$ is taken from [18].

The background uncertainty is included in the calculation. This uncertainty comes from the statistical error in the background estimation and amounts to about 12% at 10 fb^{-1} , 7% at 30 fb⁻¹ and 4% at 100 fb⁻¹. See Sect. 5.1 for a discussion of the background estimation.

The results are summarized in Table 3. Even for a Higgs mass as low as 130 GeV, a 5σ signal can be obtained with a reasonable integrated luminosity. For higher Higgs masses, a very strong signal would be expected, and prospects for a measurement of the cross section for $pp \rightarrow qqH$ become more promising. Figure 17 shows the significance for an integrated luminosity of 30 fb⁻¹ as a function of m_H , and Fig. 18 shows the minimum inte-

grated luminosity needed for a 5σ signal also as a function of m_H . The individual cut efficiencies with respect to the starting cross-section for 120 and 160 GeV Higgs bosons and the backgrounds are shown in Tables 4–6 for each channel.

Concerning systematics, we have first considered the impact of the jet energy scale. The expected jet energy scale uncertainty in CMS is about 3%. For the $t\bar{t}j$ background the scale uncertainty after correction is about 5% for $p_{\rm T} > 30$ GeV. In this analysis, the two tag jets are required to have $Ep_{\rm T1} > 50$ GeV and $E_{\rm T2} > 30$ GeV and we reject additional jets in the central region if their $E_{\rm T} > 20$ GeV. For the jets with $E_{\rm T} \sim 20$ GeV, the cross-section uncertainty after scaling the raw jet energies up and down by 10%. In general, signal and background yields correlate, so the impact on the significance with a 10% jet energy scale uncertainty is less than $\sim 8\%$ -10% at 30 fb⁻¹.



Fig. 12. The $\Delta \phi$ distribution between the two leptons after jet and lepton cuts for qqH, $m_H = 200$ GeV

We also tested our results for the significances to errors in the $\not{E}_{\rm T}$ scale. Increasing the $\not{E}_{\rm T}$ scale by 10% decreases the significance by 9%–11%. Decreasing the $\not{E}_{\rm T}$ scale by 10% increases the significance by 0.3%–3.4% depending on m_H . This is a systematic uncertainty on the signal cross section.

We found that the production cross-section depends on the choice of scale (renormalization scale × factorization scale) for the $t\bar{t}j$ background. The $t\bar{t}j$ cross-section is 736.5 pb as reported in Table 1, with the definition of the scale $\Sigma m_{\rm T}^2$, where $m_{\rm T}^2 = m_{\rm top}^2 + p_{\rm T}^2$ and the sum is over *final state light partons*. However, if we change the definition of the above sum to include all the final state partons including the heavy quarks, then the cross-section decreases to 530 pb. These two definitions of scale are the defaults in AlpGen 1.3.3 and 2.0.x respectively. The choice of scale does not affect the kinematics of $t\bar{t}j$ at all. Moreover, the cross-section and kinematics of the qqH process are not affected by the choice of scale. The significance with the new



Fig. 13. The transverse mass of the two W bosons, $M_{T,WW}$, for (a) signal events, qqH, $m_H = 120$ GeV and backgrounds (b) $t\bar{t}j$, (c) EW WW jj and (d) QCD WW jj



Fig. 14. The transverse mass, $M_{T,WW}$, distributions for signal and background, with Higgs mass = 120, 130, 140, 160, 180 and 200 GeV respectively shown in (**a**-**f**). The *lower plot* (*light gray*) is the signal, the *middle plot* (*dark gray*) is the background, and the *black histogram* is the sum

scale choice is $\sim 18\%$ higher. Therefore, the uncertainties in the computed $t\bar{t}j$ background make it very important to measure the background directly in the experiment.

It should be pointed out that the statistical significance of our analysis is generally a factor of $\sim 2.6-3.2$ lower

than the significance reported in the study for the ATLAS detector [7]. There are several reasons for this difference. First of all, the $t\bar{t}j$ cross-section used in [7] is smaller than the cross-section we use by about a factor of 0.7. Furthermore, the ATLAS study includes the gluon–gluon fusion



Fig. 15. The azimuthal angle difference in radians between the dilepton momentum vs. the missing $E_{\rm T}$ vs. $p_{\rm T}^{\rm Higgs}$ for qqH with $m_H = 120$ GeV. The *lines* correspond to the cuts $57.29\Delta\phi(ll, E_{\rm T}) + 1.5p_{\rm T}^{\rm Higgs} > 180$, $12 \times 57.29\Delta\phi(ll, E_{\rm T})$ $+ p_{\rm T}^{\rm Higgs} > 360$

channel for Higgs production which increases the signal by about 10%.

Another important difference between the two analyzes concerns the central jet veto. Our signal simulation generates a larger number of central jets compared to the AT-LAS study, which used the PYTHIA Monte Carlo event generator. When we compare the signal efficiency after all cuts using PYTHIA instead of MadGraph, we find a difference of $\sim 10\%$ for 160 GeV Higgs boson in the ee channel, and 15% in the $e\mu$ channel. For 120 GeV Higgs boson, the difference is 47% in the ee channel an 43% in the $e\mu$ channel. Moreover, when these differences are propagated to the significance calculated with $S_{\rm cP}$, we find that PYTHIA gives 50% higher significance for $120\,\text{GeV}$ Higgs boson for 10 fb^{-1} , and 30% for 100 fb^{-1} when compared to MadGraph results. The significance difference is less for 160 GeV Higgs boson; 10 to 7% higher for the 160 GeV Higgs boson for 10 and $100 \,\mathrm{fb}^{-1}$ respectively. Finally, the very definition of significance (S_{cP}) differs between the two studies. The ATLAS study used a definition which gives a value which is $\sim 9\%$ -14% higher for the same number of signal and background events. Moreover,



Fig. 16. (a) $\Delta \phi(ll, \not\!\!\!E_T)$ vs. $\Delta \phi_{ll}$ and (b) the sum of $\Delta \phi_{ll}$ and $\Delta \phi(ll, \not\!\!\!E_T)$ for $qqH m_H = 120$ and for ttj background (c) and (d). The *lines* correspond to $\Delta \phi_{ll} + \Delta \phi(ll, \not\!\!\!E_T) = 3$ rad

Table 2. Summary of accepted cross sections, in fb. A series of assumed Higgs boson masses is shown, as well as the backgrounds for the "low-mass" and "high-mass" cuts

	Accepted cross-sections (fb)				
Channel	e^+e^-	$e^{\pm}\mu^{\mp}$	$\mu^+\mu^-$	Sum	
	"Low" ma	SS			
$qqH, m_H = 120 \text{ GeV}$	0.18	0.40	0.25	0.84	
$qqH, m_H = 130 \text{ GeV}$	0.39	0.85	0.60	1.84	
$qqH, m_H = 140 \text{ GeV}$	0.62	1.34	0.96	2.91	
$t\bar{t}j$	1.14	2.62	1.06	4.82	
W^+W^-jj (EWK)	0.08	0.14	0.09	0.32	
W^+W^-jj (QCD)	0.09	0.21	0.12	0.42	
all backgrounds				5.56	
	"High" ma	SS			
$qqH, m_H = 160 \text{ GeV}$	1.59	3.50	2.10	7.19	
$qqH, m_H = 180 \text{ GeV}$	1.36	3.09	1.84	6.29	
$qqH, m_H = 200 \text{ GeV}$	0.81	1.70	1.09	3.61	
$\bar{t}\bar{t}j$	2.09	4.22	2.02	8.33	
W^+W^-jj (EWK)	0.13	0.25	0.17	0.54	
W^+W^-ii (QCD)	0.19	0.39	0.25	0.84	
all backgrounds				9.70	

Table 3. Significance of an excess as a function of Higgs mass, for three assumed integrated luminosities. The last column shows the minimum luminosity required for a 5σ excess

Higgs mass (GeV)	$10{ m fb}^{-1}$	Significance $30 \mathrm{fb}^{-1}$	$100{\rm fb}^{-1}$	$\mathcal{L}_{ ext{min}}^{5\sigma}\ (ext{fb}^{-1})$
120 130 140 160 180 200	$\begin{array}{c} 0.72 \\ 1.77 \\ 2.68 \\ 4.54 \\ 3.95 \\ 2.31 \end{array}$	$1.35 \\ 3.04 \\ 4.79 \\ 7.00 \\ 6.22 \\ 4.03$	$2.60 \\ 5.85 \\ 8.33 \\ 13.0 \\ 11.6 \\ 6.99$	$340 \\ 72 \\ 33 \\ 12 \\ 15 \\ 45$

when the number of background events is reduced, the apparent improvement in the significance increases more dramatically than our measure of significance. Thus the uncertainty of $\sim 9\%$ -14% should be taken as a lower limit for this particular factor. Considering all of the above, the differences between our results and those reported in [7] can be understood. Nonetheless, these considerations show that there still are uncertainties in the modeling of this channel which should be investigated by both experiments.

5.1 Background estimation from the data

For the Higgs masses considered here, there is practically no signal with $M_{\ell\ell} > 110 \text{ GeV}$ – see Fig. 10. For the present discussion we define this as the *signal-free region*. Figure 19 shows the $M_{\ell\ell}$ distribution computed with looser cuts (no central jet veto, no $p_{\rm T}$ -balancing cut, $|\Delta \eta| > 3.5$, $\eta_{\rm lo} + 0.3 < \eta_{\ell} < \eta_{\rm hi} - 0.3$) and the full analysis cuts. The



Fig. 17. Significance of the Higgs signal as a function of Higgs mass for a 30 fb^{-1} integrated luminosity



Fig. 18. Minimum integrated luminosity (fb⁻¹) needed to obtain a 5σ excess over the $t\bar{t}j + W^+W^-jj$ background as a function of the Higgs mass

number of events with $M_{\ell\ell} > 110 \text{ GeV}$ is designated by "a" for the distribution with looser cuts and by "c" for the full analysis cuts. The number of events for $M_{\ell\ell} < 80 \text{ GeV}$ is designated by "b" for the distribution with looser cuts and by "d" for the full analysis cuts. The region 80 < $M_{\ell\ell} < 110\,{\rm GeV}$ is excluded from the calculation in order to avoid any background coming from $Z \to \ell^+ \ell^-$. Since $M_{\ell\ell} > 110 \text{ GeV}$ represents the signal-free region, we can use the numbers a, c and b to estimate the number of background events in the region where we expect the signal (i.e., d). Using the simulations, we find that c/a = 0.097and d/b = 0.098. The error on this estimation is dominated by the statistical uncertainty which is $\sqrt{c}/c \approx 7\%$. In order to obtain the background distribution in $M_{T,WW}$, we take the distribution obtained with the looser cuts and scale it by the factor of 0.098. A comparison of the real and rescaled background distributions is given in Fig. 20 which indicates that this "data driven" method works quite well.

Table 4. Accepted signal (for $m_H = 120, 160 \text{ GeV}$) and major background cross-sections in fb for the $ee\nu\nu$ final state

Cut	qqH	qqH	ttj	WWjj	WWjj
	120	160	-	EW	QCD
	5.26	27.0	8610	10.7	514
$E_{\rm T1} > 50, E_{\rm T2} > 30 {\rm GeV}$	3.74	18.7	6740	8.84	296
$\Delta \eta > 4.2$	1.22	6.07	184	2.20	12.2
$\eta_1 imes \eta_2 < 0$	1.21	6.05	183	2.19	12.2
$M_{jj} > 600 \mathrm{GeV}$	1.07	5.37	147	2.07	9.05
$P_{\rm T}$ -balance cut	0.65	3.35	54.9	1.02	3.30
Central jet veto	0.40	2.31	15.0	0.63	1.49
$\geq 2 \text{ good leptons w opp. charge}$	0.27	1.92	11.0	0.48	0.70
$p_{\rm T} > 20, 10 \text{ or } p_{\rm T} > 26, 12 {\rm GeV}$	0.25	1.84	10.6	0.48	0.68
$ \Delta R(j,l) > 0.7$	0.25	1.83	10.3	0.47	0.66
Req. leptons between jets	0.24	1.71	4.99	0.42	0.43
$M_{ll} < 80 {\rm GeV}$	0.24	1.68	2.39	0.14	0.21
$\Delta \phi_{ll} < 2.4$	0.22	1.59	2.09	0.13	0.19
$M_{{\rm T},WW} < 125 \; {\rm GeV}$	0.18		1.14	0.08	0.09
$57.29\Delta\phi(ll, \not\!\!\!E_{\mathrm{T}}) + 1.5p_{\mathrm{T}}(H) > 180 \&$					
$12 \times 57.29 \Delta \phi(ll, E_{\rm T}) + p_{\rm T}(H) > 360$	0.16		0.94	0.07	0.07
$M_{ll} > 10\& E_{\rm T} > 30(ee, \mu\mu)$	0.12		0.80	0.05	0.06
$\Delta \phi(ll, \not\!\!\!E_{\rm T}) + \Delta \phi_{ll} < 3$	0.09		0.42	0.03	0.03
High mass cuts					
No $M_{T,WW}$ cut		1.59	2.09	0.13	0.19
$57.29\Delta\phi(ll, E_{\rm T}) + 1.5p_{\rm T}(H) > 180$ &					
$12 \times 57.29 \Delta \phi(ll, E_{\rm T}) + p_{\rm T}(H) > 360$		1.50	1.89	0.11	0.17
$M_{ll} > 10 { m GeV} \& E_{ m T} > 30 (ee, \mu\mu) { m GeV}$		1.30	1.74	0.10	0.15
$\Delta \phi(ll, \not\!\!\!E_{\rm T}) + \Delta \phi_{ll} < 3$		0.86	0.65	0.05	0.05

Table 5. Accepted signal (for $m_H = 120, 160 \text{ GeV}$) and major background cross-sections in fb for the $e\mu\nu\nu$ final state

Cut	qqH	qqH	ttj	WWjj	WWjj
	120	160		\mathbf{EW}	QCD
	10.6	53.2	17200	21.5	1030
$E_{\rm T1} > 50, E_{\rm T2} > 30 {\rm GeV}$	7.29	35.5	13300	17.2	537
$\Delta \eta > 4.2$	2.46	12.6	359	4.53	24.4
$\eta_1 imes \eta_2 < 0$	2.45	12.6	356	4.53	24.3
$M_{ij} > 600 \mathrm{GeV}$	2.15	11.1	282	4.30	18.3
$P_{\rm T}$ -balance cut	1.40	7.39	117	2.41	8.29
Central jet veto	0.88	5.13	32.7	1.50	4.12
$\geq 2 \text{ good leptons w opp. charge}$	0.67	4.39	25.1	1.19	2.10
$p_{\rm T} > 20, 10 {\rm GeV}$	0.54	4.08	23.5	1.13	1.98
$ \Delta R(j,l) > 0.7$	0.54	4.05	21.7	1.10	1.88
Req. leptons between jets	0.51	3.75	10.6	0.92	1.07
$M_{ll} < 80 {\rm GeV}$	0.50	3.69	5.01	0.30	0.45
$\Delta \phi_{ll} < 2.4$	0.48	3.50	4.22	0.25	0.39
$M_{{\rm T},WW} < 125\;{\rm GeV}$	0.40		2.62	0.14	0.21
$57.29\Delta\phi(ll,\not\!\!\!E_{\rm T}) + 1.5p_{\rm T}(H) > 180 \ \&$					
$12 \times 57.29 \Delta \phi(ll, E_{\rm T}) + P_{\rm T}(H) > 360$	0.329		1.88	0.11	0.15
$\not\!$	0.32		1.82	0.11	0.15
$\Delta\phi(ll, \not\!$	0.24		0.80	0.07	0.08
High mass cuts					
No $M_{\mathrm{T},WW}$ Cut		3.50	4.22	0.25	0.39
$57.29\Delta\phi(ll,\not\!\!\!E_{\rm T}) + 1.5p_{\rm T}(H) > 180 \ \&$					
$12 \times 57.29 \Delta \phi(ll, E_{\rm T}) + P_{\rm T}(H) > 360$		3.11	3.42	0.20	0.33
$E_{\rm T} > 30 {\rm GeV}$ if $p_{\rm T}(H) < 50 {\rm GeV}$		3.08	3.36	0.20	0.33
$\Delta\phi(ll, \not\!\!\!E_{\rm T}) + \Delta\phi_{ll} < 3$		2.00	1.71	0.11	0.17

Cut	qqH	qqH	ttj	WWjj	WWjj
	120	100		EW	QUD
	5.13	29.4	8620	10.8	513
$E_{\rm T1} > 50, E_{\rm T2} > 30 {\rm GeV}$	3.36	18.3	6620	8.33	233
$\Delta \eta > 4.2$	1.27	7.39	178	2.37	12.1
$\eta_1 imes \eta_2 < 0$	1.27	7.38	177	2.36	12.1
$M_{jj} > 600 \mathrm{GeV}$	1.11	6.52	140	2.25	8.99
$P_{\rm T}$ -balance cut	0.85	4.95	55.8	1.59	5.77
Central jet veto	0.56	3.52	19.6	1.01	3.14
$\geq 2 \text{ good leptons w opp. charge}$	0.43	2.89	16.1	0.77	1.47
$p_{\rm T} > 20, 10 {\rm GeV}$	0.33	2.61	14.3	0.72	1.32
$ \Delta R(j,l) > 0.7$	0.32	2.54	11.6	0.68	1.19
Req. leptons between jets	0.29	2.30	5.46	0.56	0.55
$M_{ll} < 80 {\rm GeV}$	0.29	2.23	2.37	0.19	0.27
$\Delta \phi_{ll} < 2.4$	0.27	2.10	2.02	0.17	0.25
$M_{T,WW} < 125 \text{ GeV}$	0.25		1.07	0.09	0.12
$57.29\Delta\phi(ll, E_{\rm T}) + 1.5p_{\rm T}(H) > 180 \ \&$					
$12 \times 57.29 \Delta \phi(ll, E_{\rm T}) + p_{\rm T}(H) > 360$	0.20		0.83	0.08	0.10
$M_{ll} > 10 \text{GeV} \& E_{\mathrm{T}} > 30(ee, \mu\mu) \text{GeV}$	0.16		0.75	0.06	0.08
$\Delta\phi(ll,\not\!\!\!E_{\rm T}) + \Delta\phi_{ll} < 3$	0.13		0.43	0.05	0.06
High mass cuts					
No $M_{\mathrm{T},WW}$ cut		2.10	2.02	0.17	0.25
$57.29\Delta\phi(ll, \not\!\!\!E_{\mathrm{T}}) + 1.5p_{\mathrm{T}}(H) > 180 \ \&$					
$12 \times 57.29 \Delta \phi(ll, E_{\rm T}) + p_{\rm T}(H) > 360$		1.91	1.79	0.15	0.23
$E_{\rm T} > 30 {\rm GeV}$ if $p_{\rm T}(H) < 50 {\rm GeV}$		1.68	1.68	0.13	0.21
$\Delta \phi(ll, E_{\rm T}) + \Delta \phi_{ll} < 3$		1.23	0.75	0.09	0.12

Table 6. Accepted signal (for $m_H = 120, 160 \text{ GeV}$) and major background cross-sections in fb for the $\mu\mu\nu\nu$ final state



Fig. 19. The $M_{\ell\ell}$ distribution computed with looser cuts and full analysis cuts



Fig. 20. The transverse mass, $M_{T,WW}$, distribution for estimated (*dashed*) and real (*solid*) background

5.2 Sensitivity to the Higgs mass

The above significance estimates are for a pure "counting experiment". We can, in addition, use the information contained in the distribution of $M_{T,WW}$ with regard to the Higgs mass. We infer the mass of the Higgs boson from the observed distribution in $M_{T,WW}$ by subtracting the datadriven estimate of the background $M_{T,WW}$ distribution from the distribution obtained with the full set of analysis cuts. The estimated and real $M_{T,WW}$ distributions for signal events are shown in Fig. 21 for several different Higgs boson masses. The inferred and the real mean values and shapes approximately agree.

In an effort to obtain a quantitative measure of m_H , we can use signal $M_{T,WW}$ distributions as templates to be compared to the observed distribution. The comparison is done using the Kolmogorov–Smirnov test, and the results are shown in Fig. 22. A value close to one indicates a good match between the shapes. Comparing the means and shapes of the observed and template distri-



Fig. 21. Estimated (*dashed*) and real (*solid*) $M_{T,WW}$ distributions for signal events, with Higgs mass of 120, 130, 140, 160, 180 and 200 GeV shown in (**a**-**f**) respectively

butions, we can differentiate between Higgs boson masses for the cases of 160, 180 and 200 GeV, and for low masses (120–140 GeV). A simple Monte Carlo study indicates that statistical error in the Kolmogorov test is ~ 11 GeV for 160 GeV and ~ 58 GeV for a 120 GeV Higgs boson. The statistical errors of the Kolmogorov test are estimated by simulating measured mass and template distributions. The template distributions are constructed for 50 to 250 GeV Higgs bosons with 10 GeV steps. The rms of the Higgs mass values predicted by 1000 Kolmogorov tests are taken as the statistical uncertainty of the tests. The errors for each Higgs mass value is tabulated in Table 7. To differentiate between the cases of 120, 130 and 140 GeV Higgs mass, we must reduce the $t\bar{t}j$ background more or we must have data corresponding to an integrated luminosity greater than 50 fb⁻¹.



Fig. 22. Kolmogorov test function for estimating the Higgs boson mass for Higgs masses of 120, 130, 140, 160, 180 and 200 GeV shown in $(\mathbf{a}-\mathbf{f})$ respectively

6 Conclusions

We have presented an analysis with prospects for a discovery signal for a standard model Higgs boson in the vectorboson fusion channel. We utilize the final state in which both W bosons decay to electrons or muons. Our study is based on a full simulation of the CMS detector and an up-to-date version of the reconstruction codes. Furthermore, we have simulated the main backgrounds, $t\bar{t}j$ and W^+W^-jj using state-of-the-art generators, and optimized to reduce these backgrounds.

The results of our study are encouraging, and indicate that an excess signal with a statistical significance of over 5σ can be obtained with an integrated luminos-

Table 7.	The	statistical	errors	of	the	Kolmogorov	test	esti-
mated by	the N	Aonte Carle	o study					

Higgs mass (GeV)	Uncertainty of the test (GeV)					
120	57.5					
130	27.8					
140	14.6					
160	10.8					
180	13.0					
200	19.6					

ity of $> 11 \,\mathrm{fb}^{-1}$ and $< 72 \,\mathrm{fb}^{-1}$ for Higgs masses in the range $130 < m_H < 200 \,\mathrm{GeV}$. Our analysis also shows that the background can be measured to 7% accuracy directly from the data. This uncertainty is dominated by statistics for 30 fb^{-1} . Finally, we suggest a method to obtain information on the Higgs mass using the shape of the $M_{\mathrm{T},WW}$ distribution.

Acknowledgements. We are grateful to A. Nikitenko for his valuable help and useful comments. We would like to thank M. Zielinsky for his assistance in using tower thresholds for jets and Y. Gerstein for his assistance in electron identification and selection. We also would like to thank N. Hadley, P. Bloch, R. Vidal and Albert DeRoeck for their comments, suggestions and criticisms. Special thanks to Sharon Hagopian for reading the manuscript and her comments.

References

- 1. LEP Higgs Working Group, Phys. Lett. B ${\bf 565},\, 61\;(2003)$
- B.W. Lee, C. Quigg, H.B. Thacker, Phys. Rev. Lett. 38, 883 (1977)
- B.W. Lee, C. Quigg, H.B. Thacker, Phys. Rev. D 16, 1519 (1977)
- W.-M. Yao et al., Review of Particle Physics, J. Phys. G: Nucl. Part. Phys. 33, 1 (2006)
- N. Kauer, T. Plehn, D. Rainwater, D. Zeppenfeld, Phys. Lett. B 503, 113 (2001)
- N. Akchurin, D. Green, S. Kunori, R. Vidal, W. Wu, M.T. Zeyrek, CMS NOTE 2002/016
- 7. S. Asai et al., Eur. Phys. J. C **32S2**, 19 (2004)
- 8. F. Maltoni, T. Stelzer, JHEP 0302, 027 (2003)
- 9. M.L. Mangano, M. Moretti, F. Piccinini, R. Pittau, A. Polosa, JHEP 0307, 001 (2003)
- 10. Torbjørn Sjøstrand, Leif Lonnblad, Stephen Mrenna, hep-
ph/0108264 $\,$
- 11. http://cmsdoc.cern.ch/~anikiten/cms-higgs/ PTDR_SMxsect/SM_higgs_cross_sections.dat
- 12. https://uimon.cern.ch/twiki/bin/view/CMS/ OnSel_06_II_06
- G. Davatz, M. Dittmar, A.-S. Giolo-Nicollerat, CMS NOTE-2006/047
- 14. E. Meschi et al., CMS NOTE 2001/034 (2002)
- CMS Data Acquisition and HLT TDR, CERN/LHCC 2002-26 (2002)
- 16. W. Adam et al., CMS NOTE 2006/041 (2006)
- I. Narsky, Nucl. Instrum. Methods A 450, 444 (2000) (online at http://cmsdoc.cern.ch/~bityukov/talks/talks.html)
- 18. http://cmsdoc.cern.ch/~bityukov