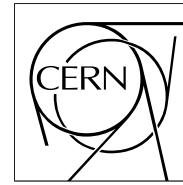**The Compact Muon Solenoid Experiment**

CMS Note

Mailing address: CMS CERN, CH-1211 GENEVA 23, Switzerland

28 Feb 2007

Search for a Standard Model Higgs Boson in CMS via Vector Boson Fusion in the H→WW→ $l\nu l\nu$ Channel

E. Yazgan[1,2], J. Damgov[3,2], N. Akchurin[4], V. Genchev[3], D. Green[2], S. Kunori[5], M. Schmitt[6], W. Wu[2], M.T. Zeyrek[1]

(1) Middle East Technical University, Ankara, Turkey

(2) Fermi National Accelerator Laboratory, Batavia, 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

Abstract

We present the potential for discovering the Standard Model Higgs boson produced by the vector-boson fusion mechanism. We considered the decay of Higgs bosons into the W^+W^- final state, with both W-bosons subsequently decaying leptonically. The main background is $t\bar{t}$ with one or more jets produced. This study is based on a full simulation of the CMS detector, and up-to-date reconstruction codes. The result is that that a signal of 5σ significance can be obtained with an integrated luminosity of $12 - 72$ fb^{-1} for Higgs boson masses between $130 < m_H < 200$ GeV. In addition, the the major background can be measured directly to 7% from the data with an integrated luminosity of 30 fb^{-1}. In this study, we suggested a method to obtain information in Higgs mass using the transverse mass distributions.

1 Introduction

One of the primary goals of CMS is to prove or disprove the existence of the Higgs boson. The LEP experiments set a lower limit on the Standard Model (SM) Higgs boson at 114.4 GeV for a 95% C.L. [1], and unitarity puts an upper limit of about 1 TeV. Even more constraining are the results of fits to precision electroweak measurements, which limit the mass of a Standard Model-like Higgs boson to be less than 194 GeV [2] 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 satisfy these constraints approximately. In the Minimal Supersymmetric extension of the Standard Model (MSSM), there is a more stringent bound coming from the internal constraints of the theory; the lightest Higgs boson must have a mass less than about 135 GeV. For these reasons, we focus on the mass region $120 < m_H < 200$ GeV.

The two main decay modes of the Standard Model Higgs boson in this mass range are $h \to b\bar{b}$ and $h \to W^+ W^-$. In the latter case, one of the W bosons may be off the mass shell. If the Higgs boson is heavier than about 135 GeV, the WW^* branching fraction will dominate, but it can be important 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 production and decay of the Higgs boson gives a relatively clean determination to the hWW coupling. Given the Higgs mass the Standard Model(SM) is completely determined, so that a measure of hWW coupling over-constrains the SM. This will be crucial to establishing the origin of electroweak symmetry breaking.

The VBF mechanism was proposed as a potential discovery channel several years ago [3]. Our initial study of this channel for the CMS detector was carried out in 2002 [4], 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 70 fb^{-1}. In the present study, we repeat the entire analysis in the mass range 120–200 GeV, using the latest simulation and reconstruction software for CMS in order to verify and improve the 2002 study. A similar study of this channel for the ATLAS detector was performed in 2004 using different generators and slightly different cuts [5].

The VBF process is characterized by two forward jets with modest transverse momentum, $E_T \approx m_W/2$, separated by a large rapidity difference. The Higgs boson signature is at low rapidity, with a pair of clean, isolated leptons and missing energy. The main backgrounds for this channel are the irreducible continuum $W^+ W^-$ production, and $t\bar{t}$ in which both top quarks decay semi-leptonically. These backgrounds are particularly troublesome when there are extra jets, j, in the event, so we have taken particular care with 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 matrix-element calculation using MadGraph [6]. For the $t\bar{t}j$ background, we used the AlpGen [7] package which correctly simulates spin correlations. We simulated the parton showers using Pythia [8]. MadGraph and AlpGen calculations are made leading order (LO). The parton distribution functions used by MadGraph and AlpGen are CTEQ6L1 and CTEQ5L1 respectively. The minimum transverse momentum cut on jets is 15 GeV, and the jet pseudo-rapidity is limited to $|\eta| < 5$. We required a separation of any jet pair, namely, $\Delta R > 0.5$, where $\Delta R = \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2}$.

Next-to-leading order (NLO) cross-sections differ from LO cross-sections by $\sim 30\%$ for a 120 GeV Higgs boson and $\sim 10\%$ for a 200 GeV Higgs boson [9]. 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 α_s-dependent vertex in the diagrams, and the 'QCD' part is the rest of this process. Note that the EW part is topologically very similar to the signal and hence is almost irreducible.

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

Channel	cross-section [pb]	WW branching ratio	$\sigma \times$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.	736.5
WWjj QCD	43.6	1.	43.6
WWjj EW	0.933	1.	0.933

3 Detector Simulation and Event Reconstruction

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

3.1 Trigger

We refer to Ref. [10] for the presently planned trigger table. The inclusive single electron trigger has an E_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_T-threshold for the inclusive single muon trigger is 19 GeV, which is well suited to this analysis. Concerning the e-μ 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 Ref. [11]. This presents no significant effect at the current state of our analysis.

There will be lepton+jet triggers that should be very useful for this analysis if lower lepton thresholds are needed. However, since the details for these triggers are not available at this time, 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 assessment 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 into the silicon tracker to pick up additional hits and 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 super-clusters [12, 13] and Kalman tracks [14]. The track – super-cluster (SC) matching condition is $\Delta R < 0.15$. Such tracks should have at least four hits, and transverse momentum $p_T > 5$ GeV. If several tracks satisfy these conditions, then the one having the least difference $|p_T - E_T|$ is taken. We reject the electron candidates if $E_T^{\mathrm{SC}} < 10$ GeV or $|\eta^{\mathrm{SC}}| > 2.0$. The probability for a generator level electron with $p_T > 10$ GeV and $|\eta| < 2.0$ to be reconstructed within $\Delta R < 0.2$ is ~ 92–98% for $10 < p_T(gen) < 20$ GeV and ~ 98–99% for $p_T > 20$ GeV. These reconstructed electrons are said to be identified if they satisfy $E_{\mathrm{HCAL}}/E_{\mathrm{ECAL}} < 0.05$, $|\Delta\eta(\mathrm{trk}, \mathrm{SC})| < 0.005$, $E^{\mathrm{SC}}/p^{\mathrm{trk}} > 0.8$ and $|1/E^{\mathrm{SC}} - 1/p^{\mathrm{trk}}| < 0.06$.

An isolation variable is defined by taking the sum of the p_T of all the tracks (except the electron candidate) within a cone of $\Delta R^{\mathrm{SC}} < 0.2$, and dividing by the E_T^{SC}. The tracks entering this sum must have at least four hits, $p_T > 0.9$ GeV, and $|z^{\mathrm{trk}} - z^e| < 0.4$ 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 $10 < p_T < 30$ GeV and $\approx 90\%$ for $p_T > 30$ GeV. The electron fake rate per jet is $\approx 3\%$ for $10 < p_T^j < 30$ GeV and less than $\approx 0.1\%$ for $p_T^j > 120$ GeV calculated using the jets from W decay in the associated production and using the forward jets in the qqH sample.

3.3 Jet and Missing $\mathbf{E}_T$ 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_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_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 qqH signal 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_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_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 $p_T = 200$ 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 $p_T > 30$ 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 $E_T(\not{E}_T)$ calculated from calorimeter hits. We corrected the $\not{E}_T$ using the sum of the E_T difference between the corrected and uncorrected jets for which the corrected jets have $E_T > 30$ GeV.

4 Event Selection

The strategy of the analysis is not complicated. 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.

4.1 Forward Jet Tagging

The jets are ordered in E_T after the corrections have been applied. The first two tag jets should be energetic, so we require $E_{T1} > 50$ GeV and $E_{T2} > 30$ GeV. Fig. 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 $E_{T3} > 20$ 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 to find a fake jet from pile-up events for low luminosity LHC running is shown in Fig. 4 as a function of the E_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 events for a E_T threshold of 20 GeV is only about 2%.

The effect of the E_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 $p_{T1} > 20$ GeV, and the other, $p_{T2} > 10$ 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. Fig. 6 shows the p_T spectra for electrons in the signal process ($M_H = 120$ 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_T requirements slightly in the di-electron channel. An event is selected if it has two electrons which satisfy:

$$(p_{T1} > 26 \text{ GeV AND } p_{T2} > 10 \text{ GeV})$$
$$\text{OR}$$
$$(p_{T1} > 20 \text{ GeV AND } p_{T2} > 12 \text{ GeV}) \, .$$

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} is the forward-tag jet having higher-rapidity, and η_{lo} is that of the lower-rapidity forward-tag jet, then our requirement can be written $\eta_{\text{lo}} + 0.6 < \eta_\ell < \eta_{\text{hi}} - 0.6$. This condition must be satisfied by both leptons. Fig. 7 shows the distributions of the related quantity, $\eta_\ell' = (\eta_\ell - (\eta_1 + \eta_2)/2) \times 4.2/\Delta\eta$. 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.

First, we require the di-jet mass to be greater than 600 GeV (see Fig. 8). Next, we look at the overall p_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).

When it comes to the leptons, we require a di-lepton mass $M_{\ell\ell} < 80$ GeV (see Fig. 10). This value is lower than the Z-mass, so that leptonic Z-decays do not affect the current analysis. A useful distinction arises in the relative azimuthal angle of the two leptons due to the spin-0 nature of the Higgs boson (see Fig. 11, 12). We take advantage of this discriminant and require $\Delta\phi < 2.4$ radians. Finally, we require that the "WW transverse-mass" be not too high when looking for Higgs bosons with mass below 150 GeV. The cut is that $M_{T,WW} < 125$ GeV, where $M_{T,WW} \equiv \sqrt{(\not{E}_T + p_{T,\ell\ell})^2 - (\vec{\not{E}}_T + \vec{P}_{T,\ell\ell})^2}$. See Fig. 13 and 14 for distributions of this quantity.

4.5   Additional Cuts

Additional cuts may be required in order that $bbjj$ and $\tau\tau jj$ backgrounds not pose a problem. The additional cuts $57.3\Delta\phi(\ell\ell, \not{E}_T) + 1.5p_T^{\text{Higgs}} > 180$ and $12 \times 57.29\Delta\phi(\ell\ell, \not{E}_T) + p_T^{\text{Higgs}} > 360$ (where $\Delta\phi(\ell\ell, \not{E}_T)$ is in radians and p_T^{Higgs} is in GeV units), and also $\not{E}_T > 30$ GeV if $p_T^{\text{Higgs}} < 50$ GeV, are imported from Ref. [3]. Here, p_T^{Higgs} is the vector sum of the transverse energy of tag jets. The distribution of signal events in the $\Delta\phi(\ell\ell, \not{E}_T)$-$p_T^{\text{Higgs}}$ plane is displayed in Fig. 15.

The Drell-Yan production of di-lepton pairs, $\gamma^* \to \ell^+\ell^-$, has a large cross-section. In order to reduce this background sufficiently, we impose a di-lepton mass cut $M_{\ell\ell} > 10$ GeV and we require $\not{E}_T > 30$ GeV when the leptons have the same flavor (see Ref. [3]).

Finally, we impose the cut $\Delta\phi(\ell\ell, \not{E}_T) + \Delta\phi(\ell\ell) < 3$ radians, which increases the signal-to-background ratio. Fig. 16 shows distributions of this quantity. The resolution of the quantity $\Delta\phi(\ell\ell, \not{E}_T)$ is improved by the $\not{E}_T$ 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 effect is seperately shown in Tables 4-6. Work is in progress to confirm their effect after full detector simulations.

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.*

channel	accepted cross-sections (fb)			
	e^+e^-	$e^\pm\mu^\mp$	$\mu^+\mu^-$	sum
"low" mass				
$qqH, m_H = 120$ GeV	0.183	0.400	0.253	0.836
$qqH, m_H = 130$ GeV	0.387	0.854	0.601	1.842
$qqH, m_H = 140$ GeV	0.617	1.341	0.955	2.913
$t\bar{t}j$	1.139	2.621	1.065	4.825
W^+W^-jj (EWK)	0.081	0.144	0.092	0.317
W^+W^-jj (QCD)	0.093	0.207	0.119	0.419
all backgrounds				5.561
"high" mass				
$qqH, m_H = 160$ GeV	1.587	3.497	2.102	7.186
$qqH, m_H = 180$ GeV	1.362	3.089	1.837	6.288
$qqH, m_H = 200$ GeV	0.815	1.703	1.087	3.605
$t\bar{t}j$	2.088	4.216	2.024	8.328
W^+W^-jj (EWK)	0.127	0.245	0.165	0.537
W^+W^-jj (QCD)	0.192	0.394	0.252	0.838
all backgrounds				9.703

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 the probability calculated assuming a Poisson distribution with N_B background events to observe equal or greater than a total number of signal and background events ($N_S + N_B$), converted to an equivalent number of sigmas for a Gaussian distribution [15]. The code to calculate S_{cP} is taken from Ref. [16].

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 Section 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. Fig. 17 shows the significance for an integrated luminosity of 30 fb^{-1} as a function of m_H, and Fig. 18 shows the minimum integrated 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 4,5,6 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 $p_T > 30$ GeV. In this analysis, the two tag jets are required to have $Ep_{T1} > 50$ GeV and $E_{T2} > 30$ GeV and we reject additional jets in the central region if their $E_T > 20$ GeV. For the jets with $E_T \sim 20$ GeV, the cross-section uncertainty after jet correction is about 10%. We re-computed all yields 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}.

We also tested our results for the significances to errors in the $\not{E}_T$ scale. Increasing the $\not{E}_T$ scale by 10% decreases the significance by $9 - 11\%$. Decreasing the $\not{E}_T$ 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 also used the Pythia event generator for our signal as an alternative to MadGraph. For $m_H = 120$ GeV, the

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

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

Cut	qqH120	qqH160	ttj	WWjj(EW)	WWjj(QCD)
	5.261	26.97	8617.	10.74	514.3
$E_{T1} > 50, E_{T2} > 30$ GeV	3.742	18.70	6743.	8.838	296.4
$\Delta\eta > 4.2$	1.217	6.067	184.2	2.195	12.22
$\eta_1 \times \eta_2 < 0$	1.215	6.054	183.1	2.193	12.18
$M_{jj} > 600$ GeV	1.073	5.367	147.2	2.071	9.052
$\mathrm{P}_T - balance$ cut	0.653	3.353	54.89	1.021	3.298
Central Jet Veto	0.401	2.309	15.04	0.631	1.490
At least 2 good leptons w opp. charge	0.269	1.915	10.98	0.483	0.695
$p_T > 20, 10$ or $p_T > 26, 12$ GeV	0.250	1.838	10.59	0.475	0.675
$|\Delta R(j,l)| > 0.7$	0.250	1.830	10.33	0.471	0.662
Req. leptons between jets	0.235	1.712	4.990	0.417	0.430
$M_{ll} < 80$ GeV	0.235	1.683	2.386	0.144	0.205
$\Delta\phi_{ll} < 2.4$	0.220	1.587	2.088	0.127	0.192
$M_{T,WW} < 125$ GeV	0.183		1.139	0.081	0.093
$57.29\Delta\phi(ll,\not{E}_T) + 1.5p_T(H) > 180$ &					
$12 \times 57.29\Delta\phi(ll,\not{E}_T) + p_T(H) > 360$	0.161		0.936	0.069	0.073
$M_{ll} > 10 \& \not{E}_T > 30(ee,\mu\mu)$	0.115		0.800	0.053	0.060
$\Delta\phi(ll,\not{E}_T) + \Delta\phi_{ll} < 3$	0.090		0.420	0.031	0.033
High Mass Cuts					
No $M_{T,WW}$ Cut		1.587	2.088	0.127	0.192
$57.29\Delta\phi(ll,\not{E}_T) + 1.5p_T(H) > 180$ &					
$12 \times 57.29\Delta\phi(ll,\not{E}_T) + p_T(H) > 360$		1.501	1.885	0.114	0.172
$M_{ll} > 10\,GeV \& \not{E}_T > 30(ee,\mu\mu)$ GeV		1.303	1.736	0.098	0.152
$\Delta\phi(ll,\not{E}_T) + \Delta\phi_{ll} < 3$		0.862	0.651	0.052	0.046

7

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

Cut	qqH120	qqH160	ttj	WWjj(EW)	WWjj(QCD)
	10.57	53.24	17230.	21.48	1029.
$E_{T1} > 50, E_{T2} > 30$ GeV	7.290	35.54	13320.	17.22	537.1
$\Delta\eta > 4.2$	2.458	12.56	358.5	4.533	24.39
$\eta_1 \times \eta_2 < 0$	2.454	12.55	355.5	4.526	24.25
$M_{jj} > 600$ GeV	2.149	11.08	282.0	4.299	18.28
$P_T - balance$ cut	1.398	7.390	117.4	2.405	8.287
Central Jet Veto	0.879	5.128	32.70	1.502	4.123
At least 2 good leptons w opp. charge	0.670	4.388	25.07	1.186	2.102
$p_T > 20, 10$ GeV	0.544	4.079	23.47	1.131	1.975
$|\Delta R(j,l)| > 0.7$	0.539	4.052	21.71	1.100	1.881
Req. leptons between jets	0.506	3.748	10.60	0.920	1.068
$M_{ll} < 80$ GeV	0.505	3.685	5.014	0.301	0.447
$\Delta\phi_{ll} < 2.4$	0.480	3.497	4.216	0.245	0.394
$M_{T,WW} < 125$ GeV	0.400		2.621	0.144	0.207
$57.29\Delta\phi(ll,\not{E}_T) + 1.5p_T(H) > 180$ &					
$12 \times 57.29\Delta\phi(ll,\not{E}_T) + P_T(H) > 360$	0.329		1.880	0.109	0.153
$\not{E}_T > 30$ GeV if $p_T(H) < 50$ GeV	0.323		1.823	0.105	0.153
$\Delta\phi(ll,\not{E}_T) + \Delta\phi_{ll} < 3$	0.239		0.798	0.066	0.08
High Mass Cuts					
No $M_{T,WW}$ Cut		3.497	4.216	0.245	0.394
$57.29\Delta\phi(ll,\not{E}_T) + 1.5p_T(H) > 180$ &					
$12 \times 57.29\Delta\phi(ll,\not{E}_T) + P_T(H) > 360$		3.105	3.418	0.202	0.334
$\not{E}_T > 30$ GeV if $p_T(H) < 50$ GeV		3.084	3.361	0.199	0.334
$\Delta\phi(ll,\not{E}_T) + \Delta\phi_{ll} < 3$		2.003	1.709	0.107	0.173

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

Cut	qqH120	qqH160	ttj	WWjj(EW)	WWjj(QCD)
	5.133	29.44	8617.	10.77	512.7
$E_{T1} > 50, E_{T2} > 30$ GeV	3.357	18.31	6621.	8.332	232.5
$\Delta\eta > 4.2$	1.271	7.391	178.0	2.365	12.11
$\eta_1 \times \eta_2 < 0$	1.268	7.375	176.7	2.360	12.06
$M_{jj} > 600$ GeV	1.109	6.522	139.7	2.251	8.988
$P_T - balance$ cut	0.854	4.947	55.75	1.585	5.768
Central Jet Veto	0.562	3.523	19.55	1.007	3.139
At least 2 good leptons w opp. charge	0.430	2.891	16.11	0.772	1.472
$p_T > 20, 10$ GeV	0.327	2.605	14.30	0.716	1.324
$|\Delta R(j,l)| > 0.7$	0.319	2.537	11.59	0.680	1.186
Req. leptons between jets	0.290	2.298	5.461	0.556	0.548
$M_{ll} < 80$ GeV	0.290	2.226	2.371	0.190	0.271
$\Delta\phi_{ll} < 2.4$	0.273	2.102	2.024	0.165	0.252
$M_{T,WW} < 125$ GeV	0.253		1.065	0.092	0.119
$57.29\Delta\phi(ll,\not{E}_T) + 1.5p_T(H) > 180$ & $12 \times 57.29\Delta\phi(ll,\not{E}_T) + p_T(H) > 360$	0.200		0.826	0.075	0.095
$M_{ll} > 10\,GeV\,\&\,\not{E}_T > 30(ee,\mu\mu)$ GeV	0.159		0.746	0.060	0.076
$\Delta\phi(ll,\not{E}_T) + \Delta\phi_{ll} < 3$	0.134		0.426	0.051	0.062
High Mass Cuts					
No $M_{T,WW}$ Cut		2.102	2.024	0.165	0.252
$57.29\Delta\phi(ll,\not{E}_T) + 1.5p_T(H) > 180$ & $12 \times 57.29\Delta\phi(ll,\not{E}_T) + p_T(H) > 360$		1.908	1.785	0.147	0.229
$\not{E}_T > 30$ GeV if $p_T(H) < 50$ GeV		1.681	1.678	0.132	0.205
$\Delta\phi(ll,\not{E}_T) + \Delta\phi_{ll} < 3$		1.229	0.746	0.092	0.119

significance obtained with Pythia is higher by 30% for a luminosity of 100 fb^{-1}, while for $m_H = 160$ GeV, it is higher by 10 %.

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 Σm_T^2, where $m_T^2 = m_{top}^2 + p_T^2$ 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. We found that 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 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 ~ 2.6–3.2 lower than the significance reported in the study for the ATLAS detector [5]. There are several reasons for this difference. First of all, the $t\bar{t}j$ cross-section used in Ref [5] is smaller than the cross-section we use by about a factor of 0.7. Furthermore, the ATLAS study includes the gluon-gluon fusion channel for Higgs production which increases the signal by about 10%.

Another important difference between the two analyses concerns the central jet veto. Our signal simulation generates a larger number of central jets compared to the ATLAS 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 $\sim 5 - 50\%$. 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 ~ 9–14% higher for the same number of signal and background events. If the number of background events is reduced, the apparent improvement in the significance increases more dramatically than for our measure of significance. Thus the uncertainty of ~ 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 Ref. [5] can be understood. Nonetheless, these considerations show that there still are uncertainties in the modelling 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 $M_{\ell\ell} > 110$ GeV – see Fig. 10. For the present discussion we define this as the *signal-free region*. Fig. 19 shows the $M_{\ell\ell}$ distribution computed with looser cuts (no central jet veto, no p_T-balancing cut, $|\Delta\eta| > 3.5$, $\eta_{lo} + 0.3 < \eta_\ell < \eta_{hi} - 0.3$) and the full analysis cuts. The number of events with $M_{\ell\ell} > 110$ GeV is designated by "a" for the distribution with looser cuts and by "c" for the full analysis cuts. The number of events for $M_{\ell\ell} < 80$ GeV is designated by "b" for the distribution with looser cuts and by "d" for the full analysis cuts. The region $80 < M_{\ell\ell} < 110$ GeV is excluded from the calculation in order to avoid any background coming from $Z \to \ell^+\ell^-$. Since $M_{\ell\ell} > 110$ 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 $c/a = 0.097$ and $d/b = 0.098$. 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" works quite well.

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 data-driven 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 distributions, we can differentiate between Higgs boson masses for the cases

of 160, 180 and 200 GeV, and for low masses (120 – 140 GeV). 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}.

6 Conclusions

We have presented an analysis meant to isolate a discovery signal for a Standard Model Higgs boson in the vector-boson 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 generated the main backgrounds, $t\bar{t}j$ and W^+W^-jj, as accurately as is presently possible.

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 luminosity of > 11 fb^{-1} and < 72 fb^{-1} for Higgs masses in the range $130 < m_H < 200$ 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 $M_{T,WW}$ distribution.

7 Acknowledgments

References

[1] LEP Higgs Working Group, Phys. Lett. **B565** (2003) 61

[2] Review of Particle Physics, J. Phys. G: Nucl. Part. Phys. **33** (2006) 1

[3] N. Kauer, T. Plehn, D. Rainwater, D. Zeppenfeld, Phys. Lett. **B503** (2001) 113-120

[4] N. Akchurin, D. Green, S. Kunori, R. Vidal, W. Wu, M.T. Zeyrek, **CMS NOTE 2002/016**

[5] S. Asai *et al.* *Prospects for the search for a standard model Higgs boson in ATLAS using vector boson fusion*, Eur. Phys. J. **C32S2** (2004) 19

[6] F. Maltoni and T. Stelzer, *MadEvent: automatic event generation with MadGraph*, JHEP 0302 (2003) 027

[7] M.L. Mangano, M. Moretti, F. Piccinini, R. Pittau, A. Polosa, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, JHEP **0307** (2003) 001

[8] Torbjorn Sjostrand, Leif Lonnblad, Stephen Mrenna, *PYTHIA 6.2: Physics and Manual*, [hep-ph/0108264]

[9] http://cmsdoc.cern.ch/ anikiten/cms-higgs/PTDR_SMxsect/SM_higgs_cross_sections.dat

[10] https://uimon.cern.ch/twiki/bin/view/CMS/OnSel_06_II_06

[11] G. Davatz, M. Dittmar, and A.-S. Giolo-Nicollerat, **CMS NOTE-2006/047**

[12] E. Meschi, et al. **CMS NOTE 2001/034** (2002)

[13] CMS Data Acquisition and HLT TDR, CERN/LHCC 2002-26 (2002)

[14] W. Adam, et al. **CMS NOTE 2006/041** (2006)

[15] I. Narsky, NIM **A450** (2000) 444; (online at http://cmsdoc.cern.ch/ bityukov/talks/talks.html)

[16] http://cmsdoc.cern.ch/ bityukov
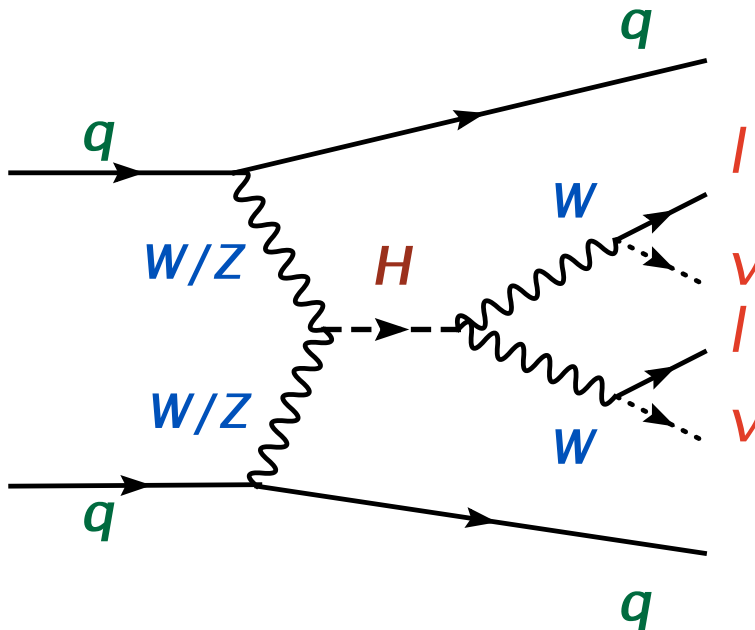
Figure 1: *Feynman diagram for Higgs boson production through Vector Boson Fusion. The Higgs boson decays into W's which further decay into electron/muon-neutrino pairs.*
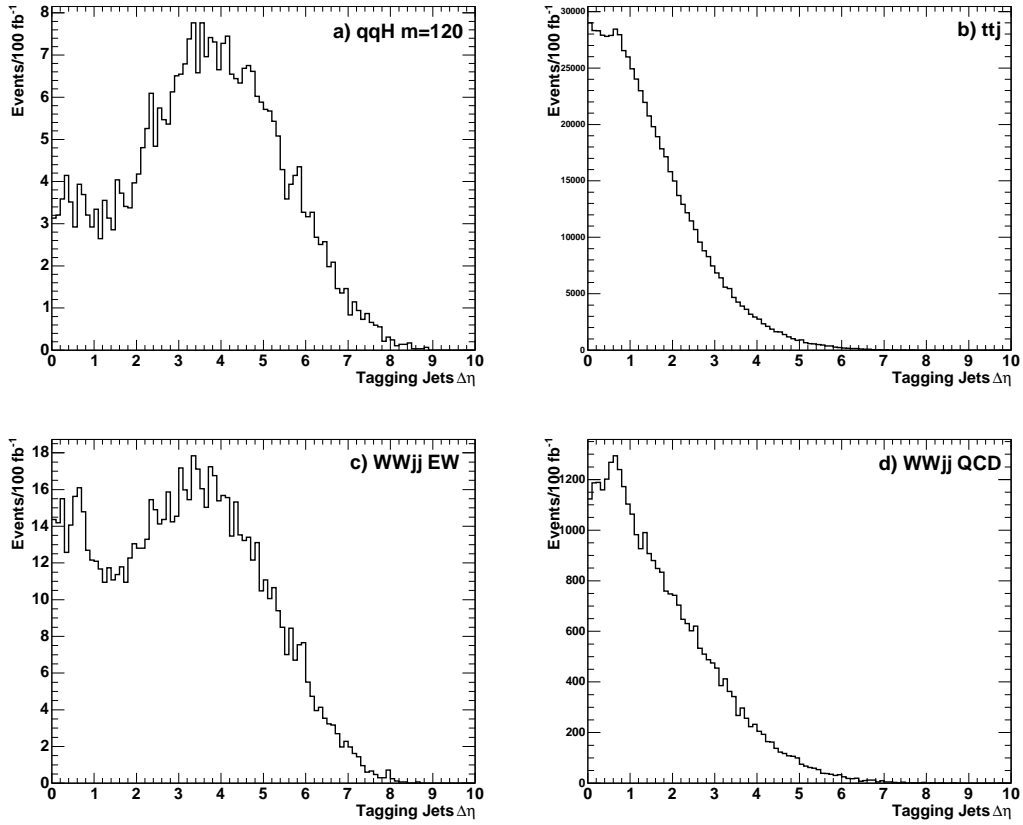
Figure 2: $\Delta\eta = |\eta_1 - \eta_2|$ *distribution for the forward tagging jets which have* $E_{T1} > 50$ GeV *and* $E_{T2} > 30$ GeV *for a) qqH,* $m_H = 120$ *GeV and backgrounds b)$t\bar{t}j$, c) EW WWjj and d) QCD WWjj. Note that the EW WWjj background is basically irreducible.*
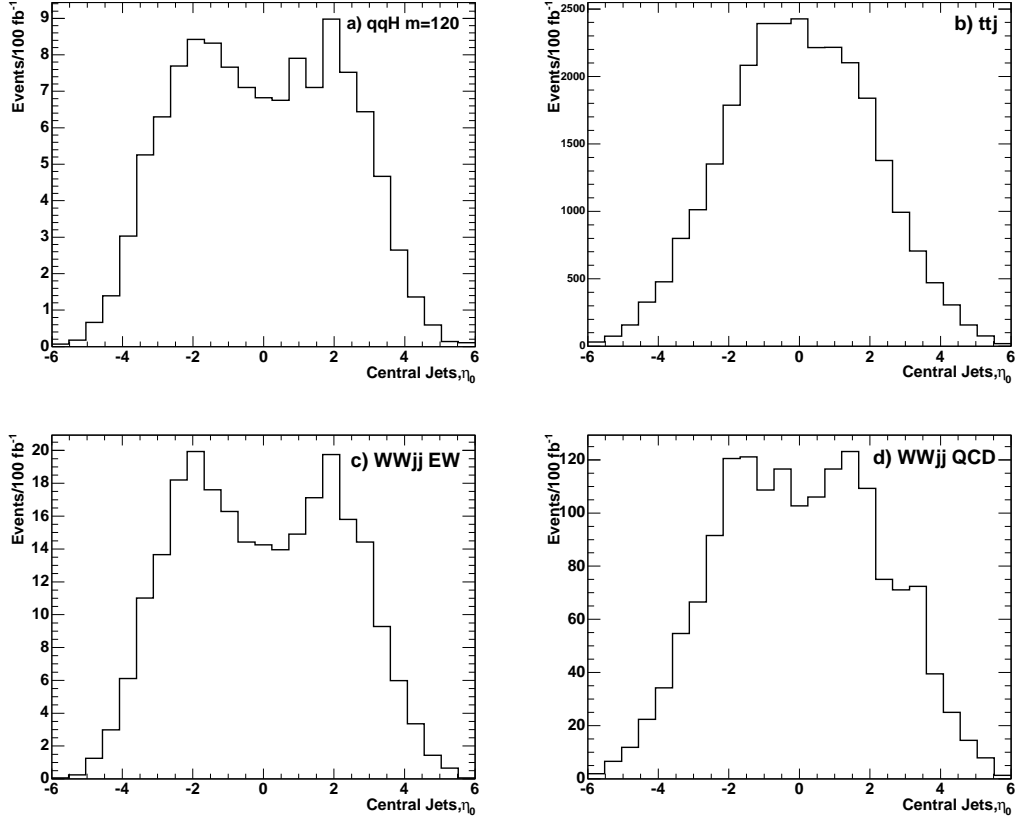
Figure 3: $\eta_0 = \eta_3 - (\eta_1 + \eta_2)/2$ *for the third jet. η of the third jet with respect to the average of the two forward jets. For signal a) qqH, $m_H = 120$ GeV and backgrounds b) $t\bar{t}j$, c) EW WWjj and d) QCD WWjj.*
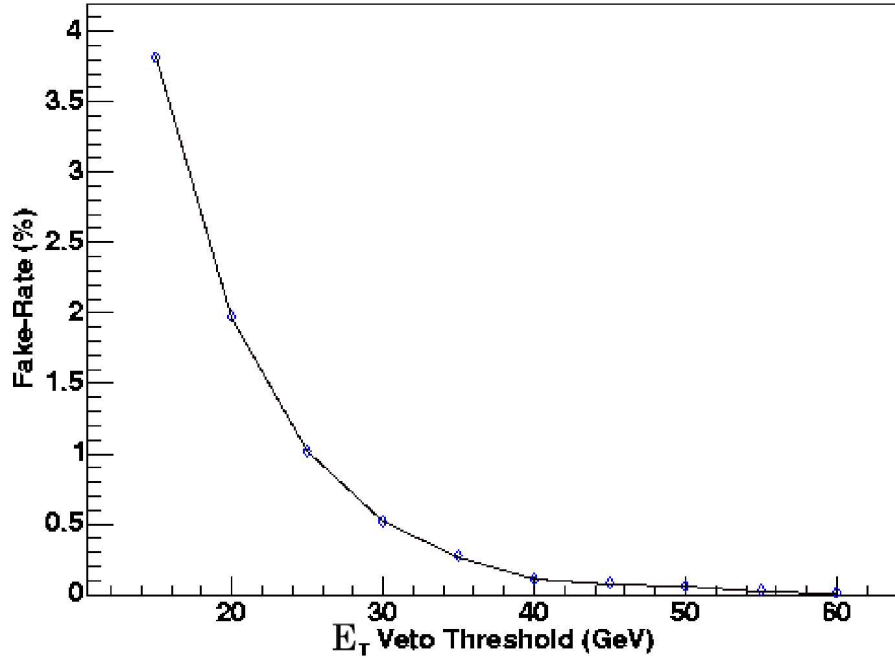
Figure 4: *Fake central jets fraction per event as a function of E_T veto threshold. A fake is defined as the probability to find at least one jet(due to pile-up) satisfying the central jet veto conditions, with no "real" jets satisfying the central jet veto condition in that event.*
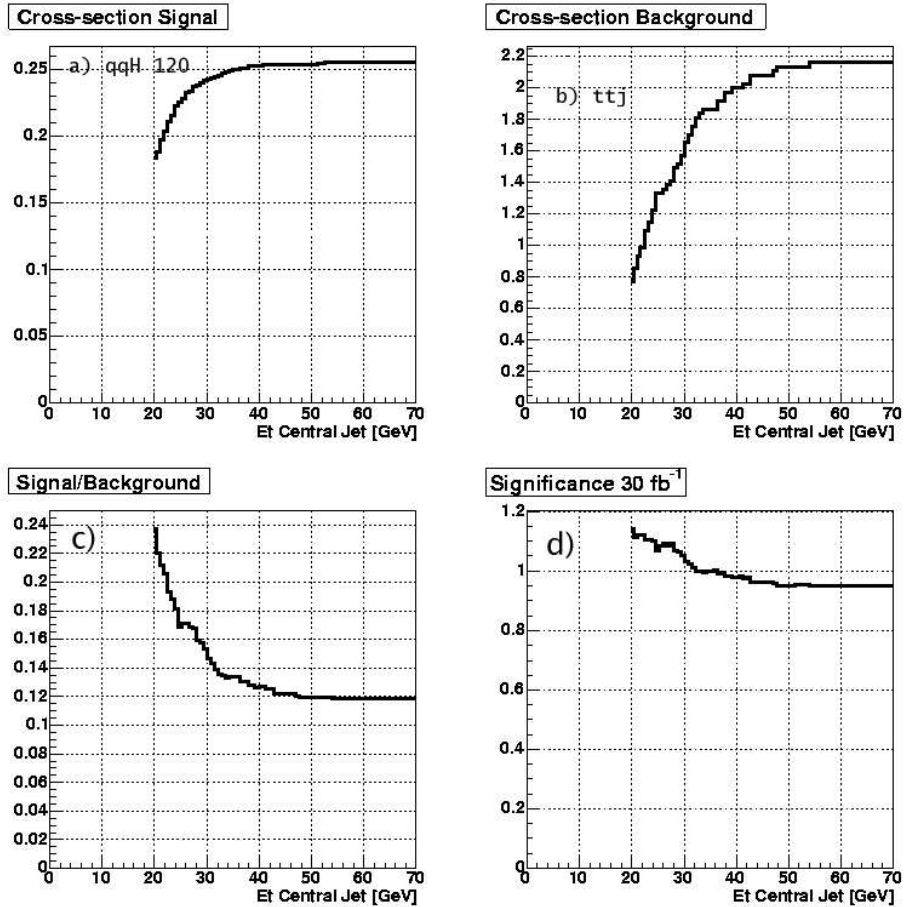
Figure 5: *The effect of the E_T threshold for the central jet veto. For a) signal evernts, qqH with $m_H = 120$ GeV and background events b) $t\bar{t}j$. c) the S/B ratio and d) the significance for a 30 fb^{-1} integrated luminosity.*
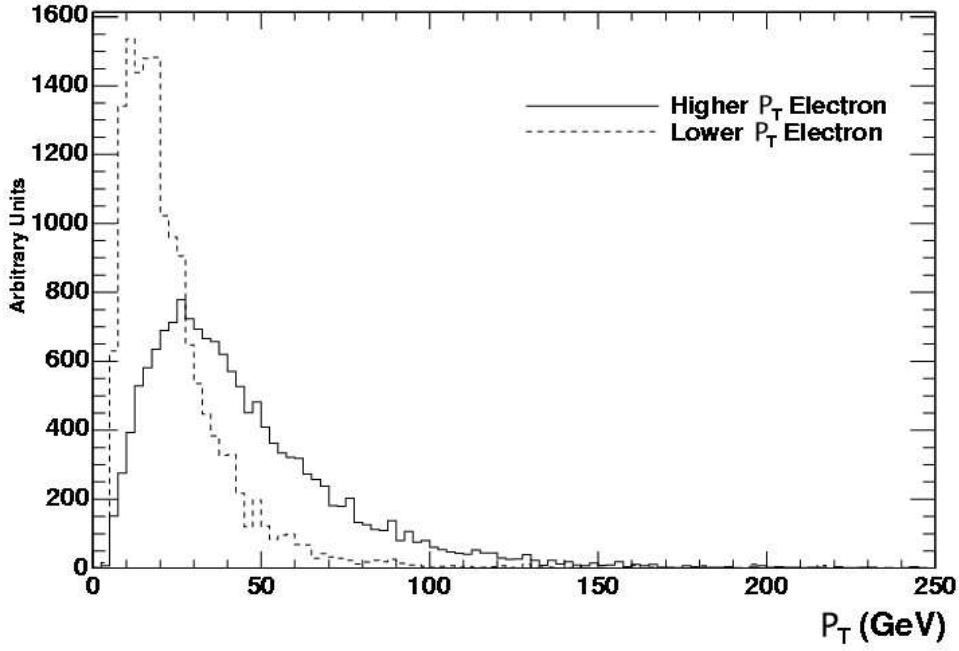
Figure 6: *Electron p_T spectra for the signal process, qqH, when $m_H = 120$ GeV*
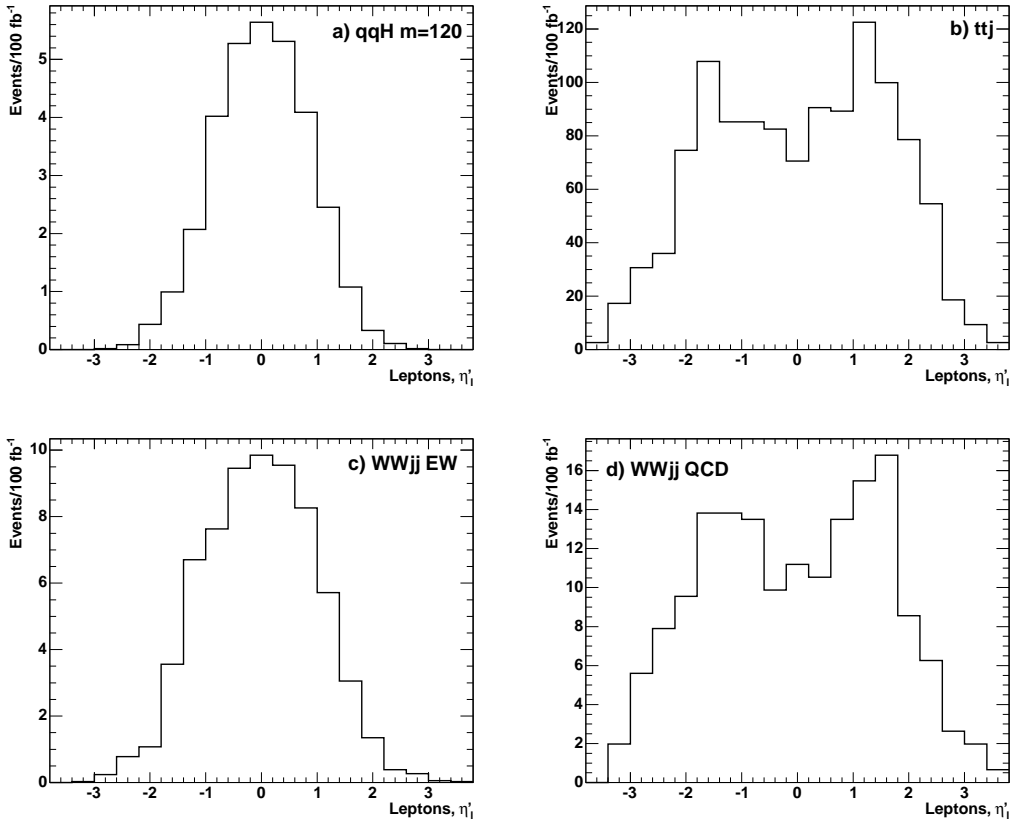
Figure 7: *Centrality of the leptons, using the quantity η_ℓ^l defined in the text for a) qqH, $m_H = 120$ GeV and backgrounds b) $t\bar{t}j$, c) EW WWjj and d) QCD WWjj.*
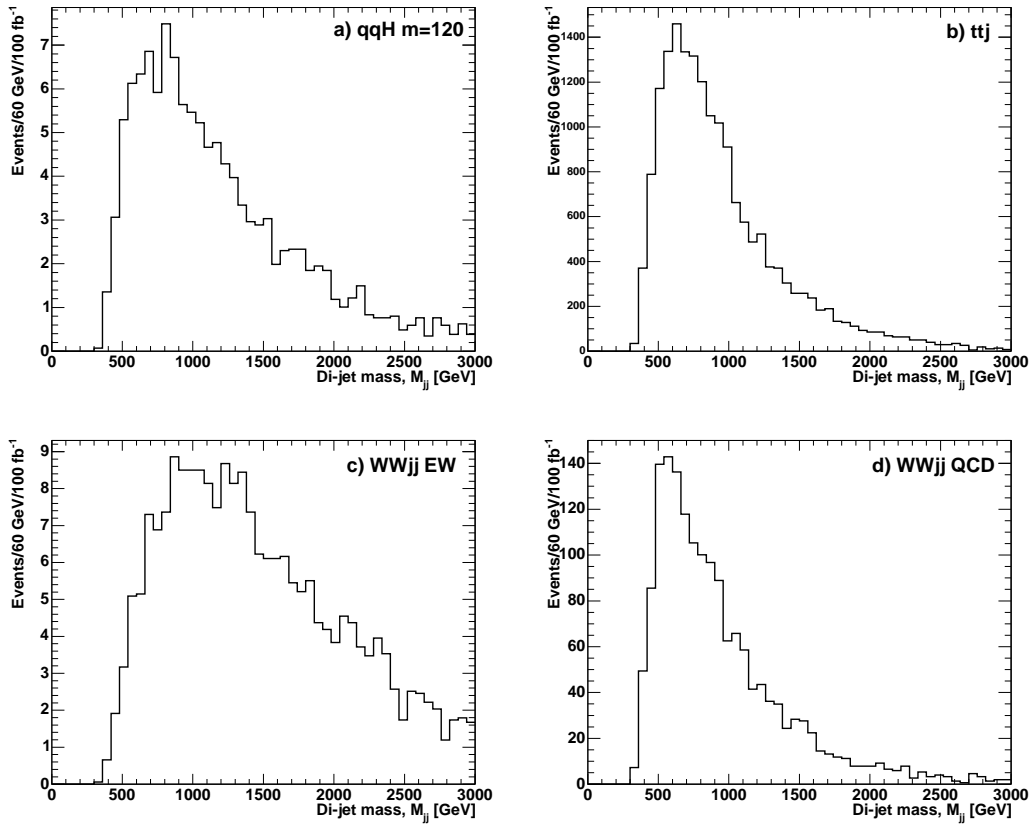
Figure 8: *Invariant mass distributions for the two forward tag jets, for a) qqH, $m_H = 120$ GeV and backgrounds b) $t\bar{t}j$, c) EW WWjj and d) QCD WWjj.*
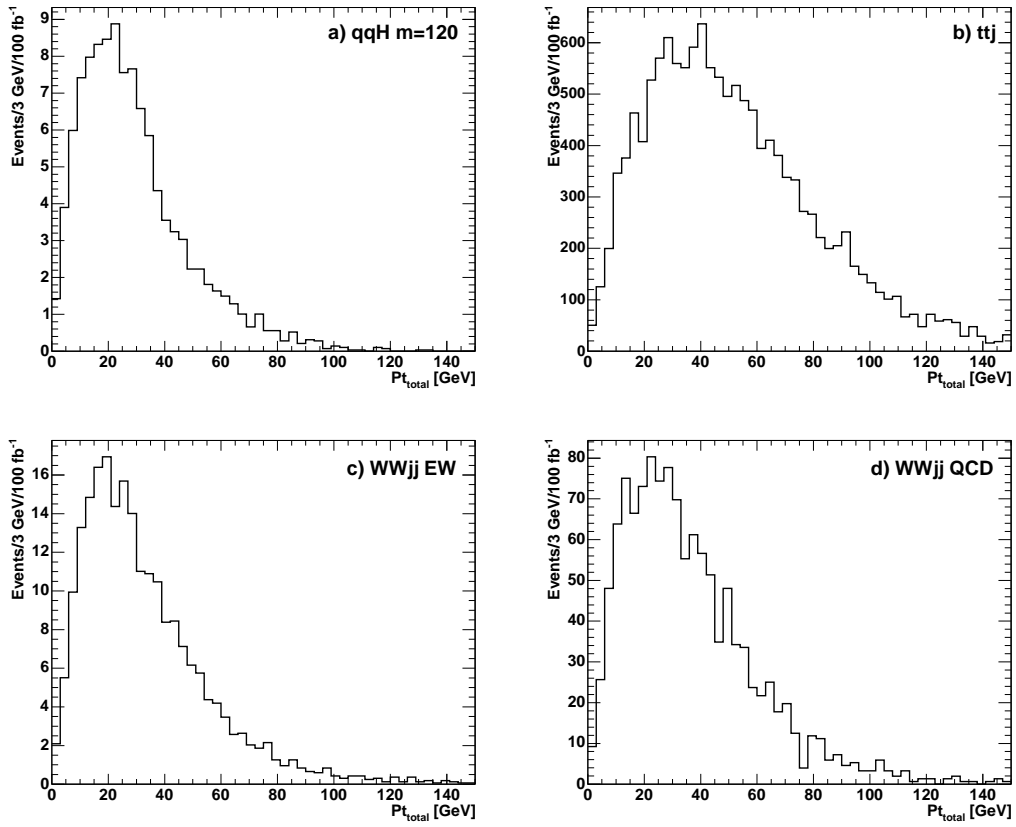
Figure 9: *The overall p_T-balance in the event. See the text for an explanation. for a) qqH, $m_H = 120$ GeV and backgrounds b) $t\bar{t}j$, c) EW WWjj and d) QCD WWjj.*
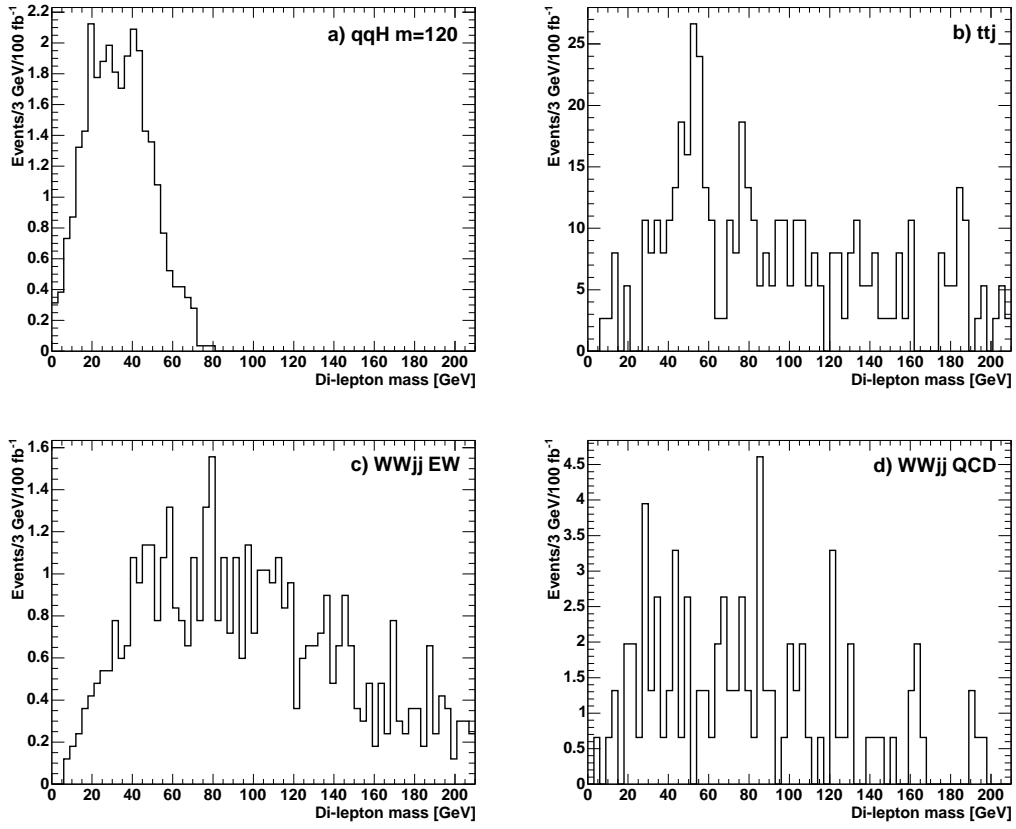
Figure 10: *Di-lepton invariant mass distribution after jet and lepton cuts, for a) qqH, $m_H = 120$ GeV and backgrounds b$t\bar{t}j$, c) EW WWjj and d) QCD WWjj.*
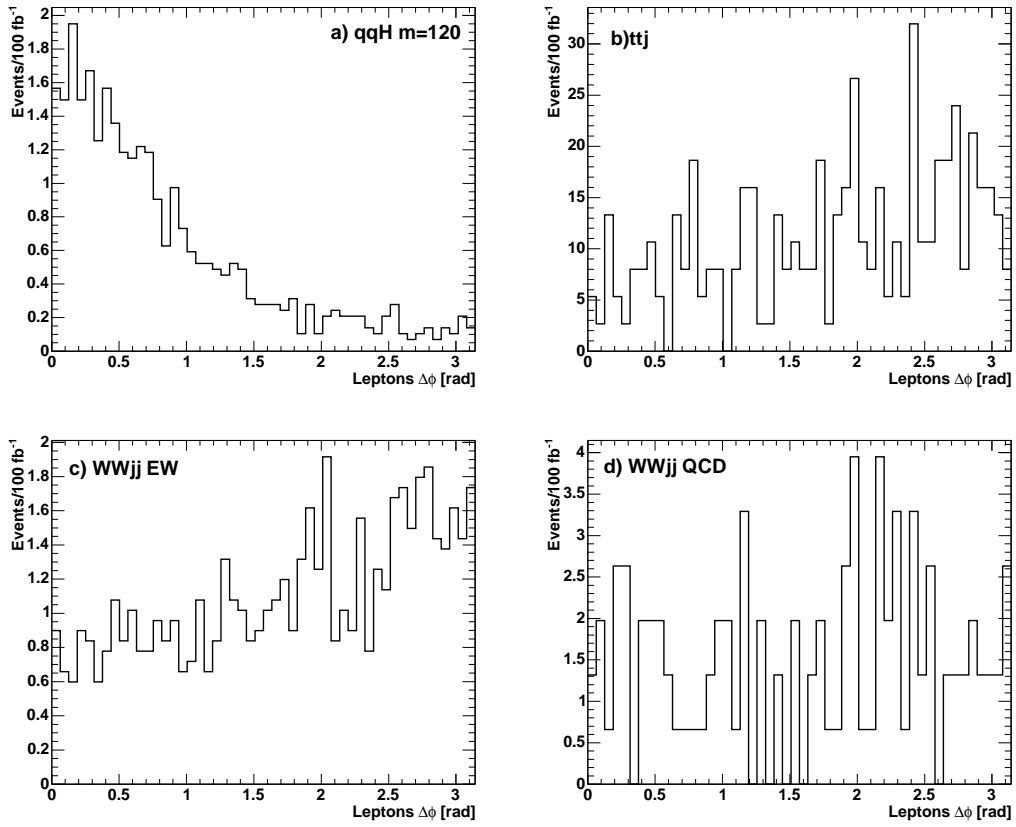
Figure 11: *The distribution of the difference in azimuthal angle between the two leptons, $\Delta\phi$ after jet and lepton cuts, for a) signal events, qqH, $m_H = 120$ GeV and backgrounds b$t\bar{t}j$, c) EW WWjj and d) QCD WWjj.*
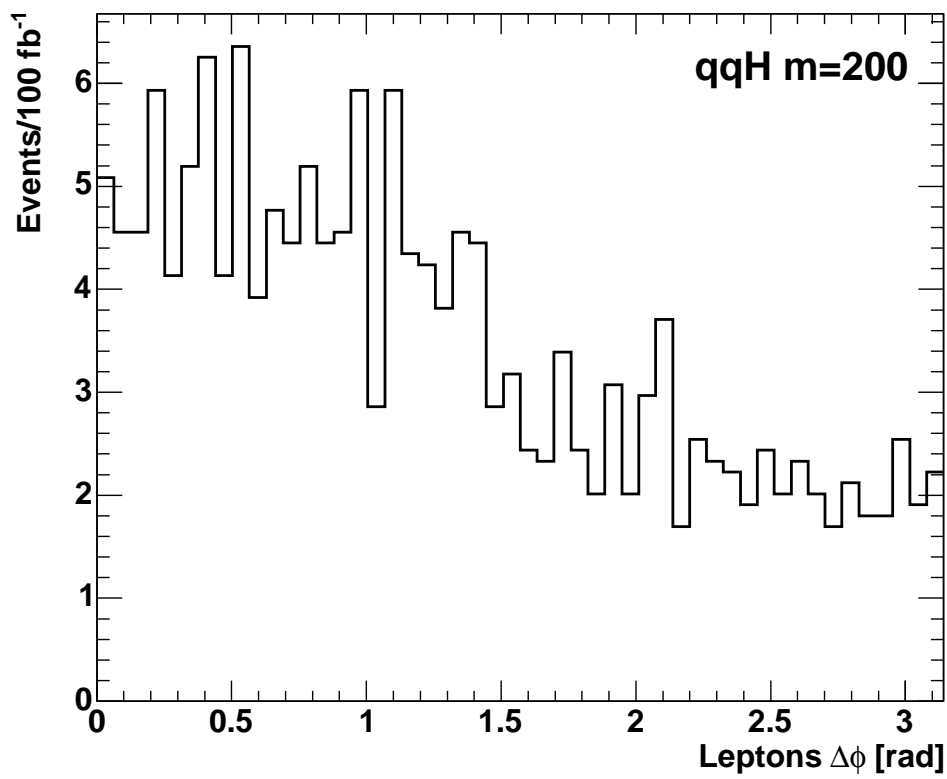
Figure 12: *The $\Delta\phi$ distribution between the two leptons after jet and lepton cuts for qqH, m_H = 200 GeV*
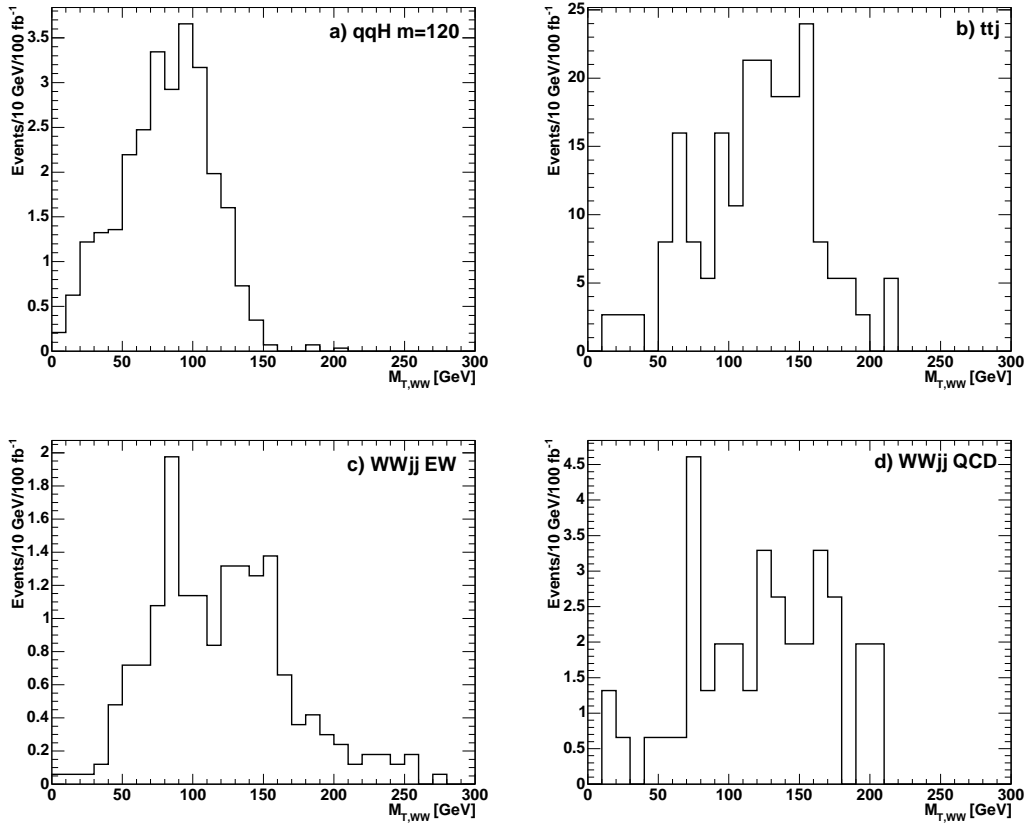
Figure 13: *The transverse mass of the two W bosons, $M_{T,WW}$, for a) signal events, qqH, $m_H = 120$ GeV and backgrounds $bt\bar{t}j$, c) EW WWjj and d) QCD WWjj.*
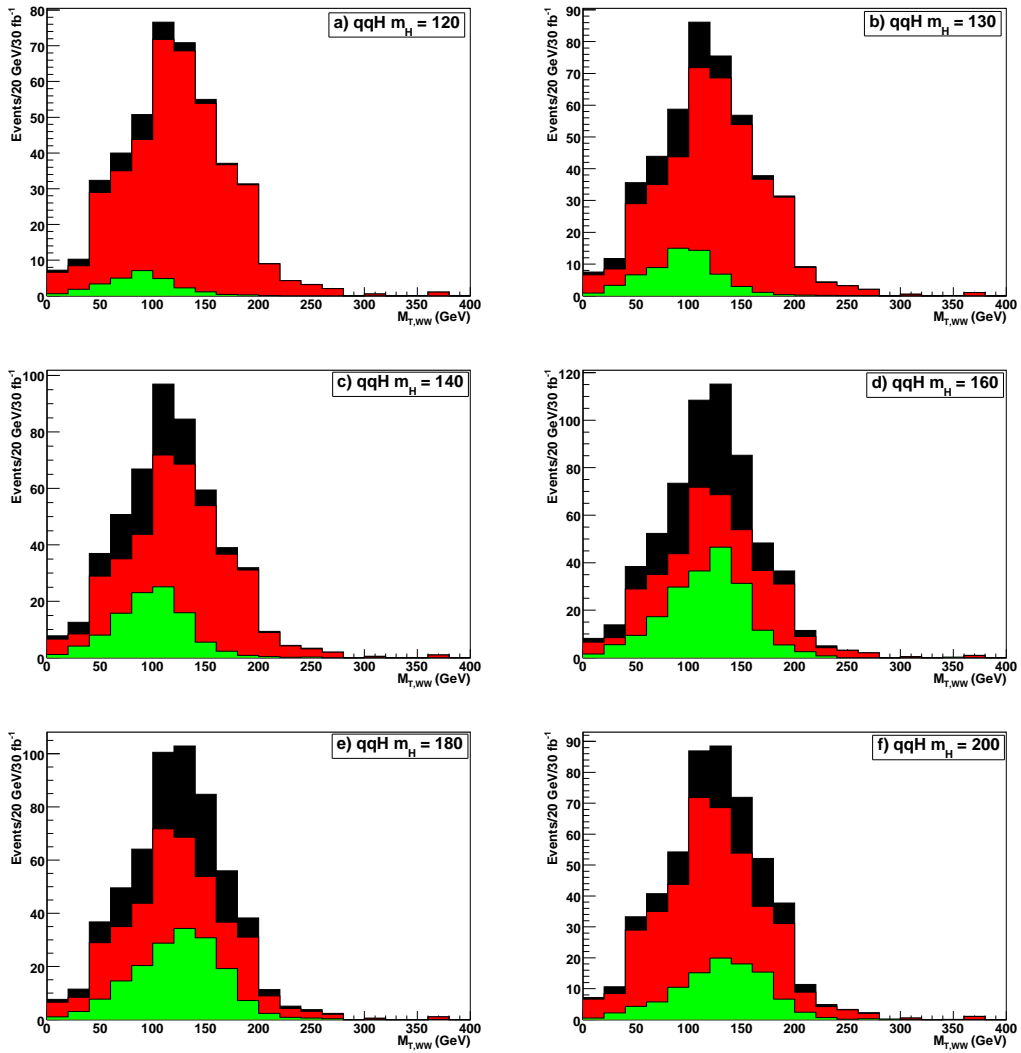
Figure 14: *The transverse mass, $M_{T,WW}$, distributions for signal and background, with Higgs mass = 120, 130, 140, 160, 180 and 200 GeV respectively shown in a),b),c),d),e),f). The Lower plot (light grey) is the signal, the middle plot(dark grey) is the background, and the black histogram is the sum.*
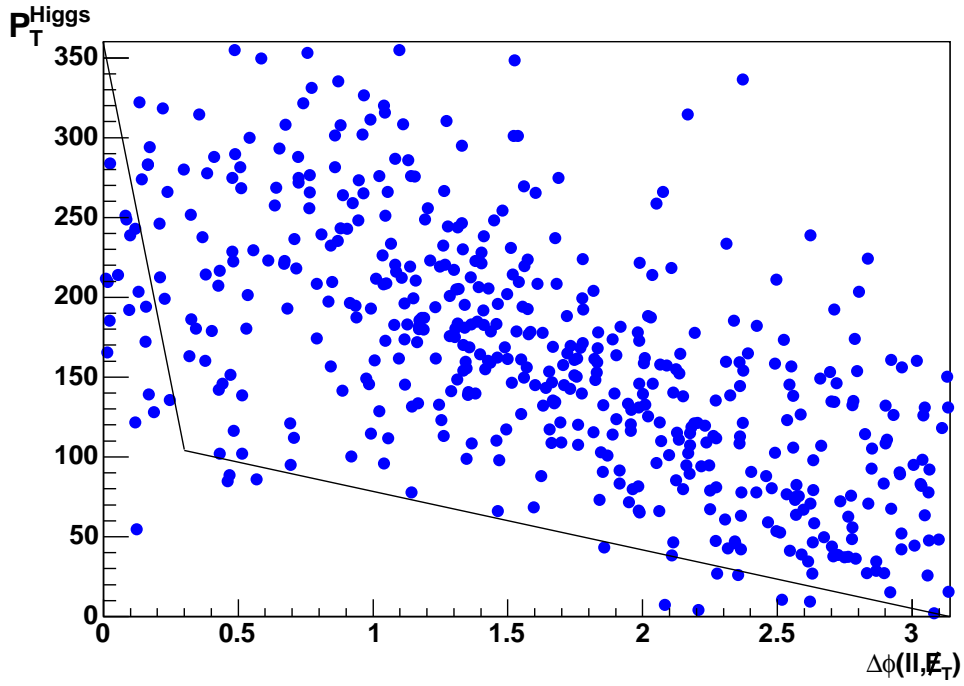
Figure 15: *The azimuthal angle difference in radians between the dilepton momentum vs. the missing E_T vs p_T^{Higgs} for qqH with m_H = 120 GeV. The lines correspond to the cuts $57.29\Delta\phi(ll, E_T) + 1.5 p_T^{Higgs} > 180$, $12 \times 57.29\Delta\phi(ll, E_T) + p_T^{Higgs} > 360$.*
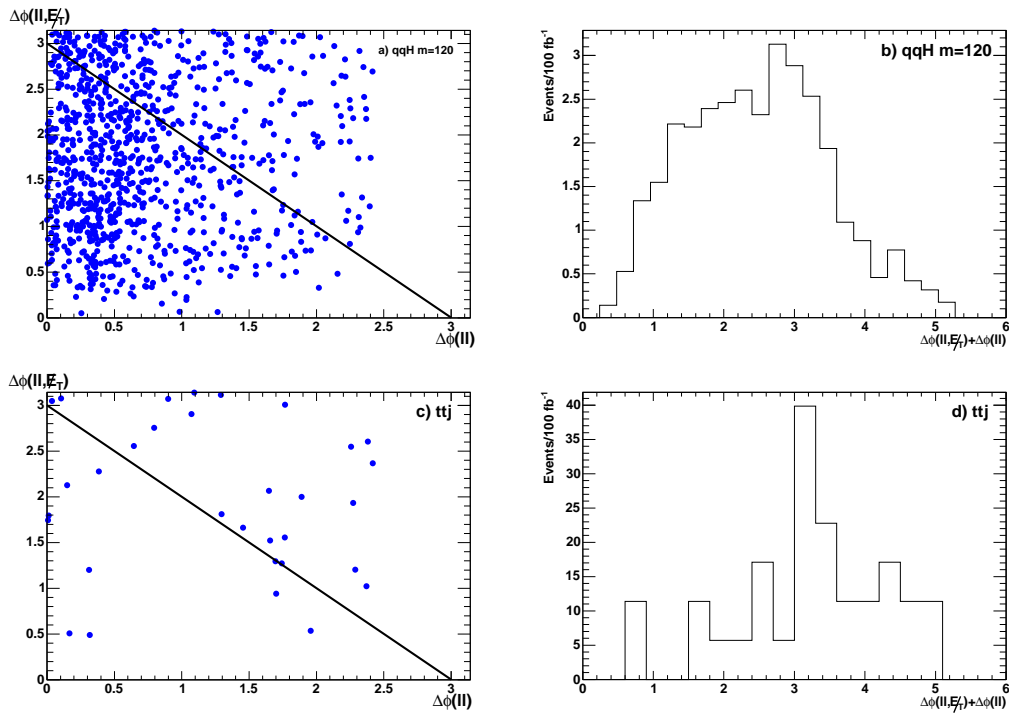
Figure 16: *a)* $\Delta\phi(ll, E_T)$ *vs.* $\Delta\phi_{ll}$ *and b) the som of* $\Delta\phi_{ll}$ *and* $\Delta\phi(ll, E_T)$ *for qqH* m_H = 120 *and for ttj bacground c) and d). The lines correspond to* $\Delta\phi_{ll} + \Delta\phi(ll, E_T) = 3$ *radians.*

24
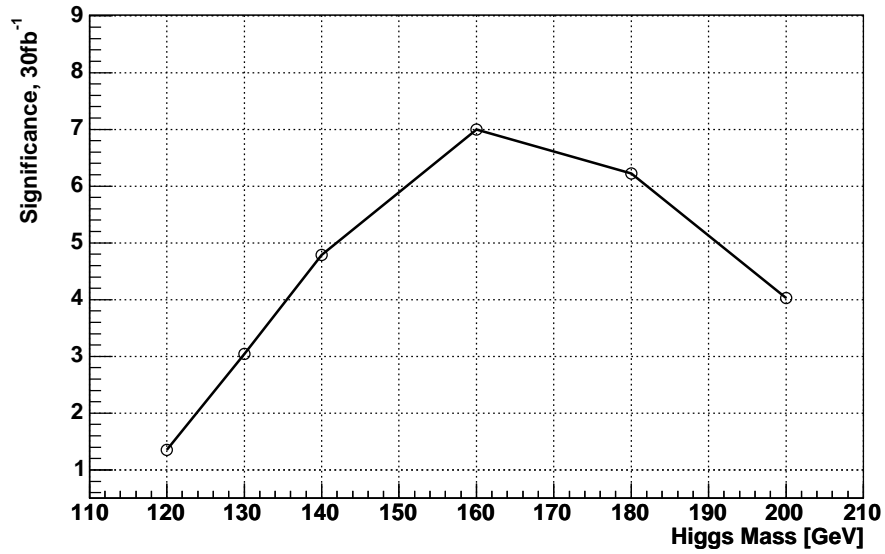
Figure 17: *Significance of the Higgs signal as a function of Higgs mass for a 30 fb^{-1} integrated luminosity.*
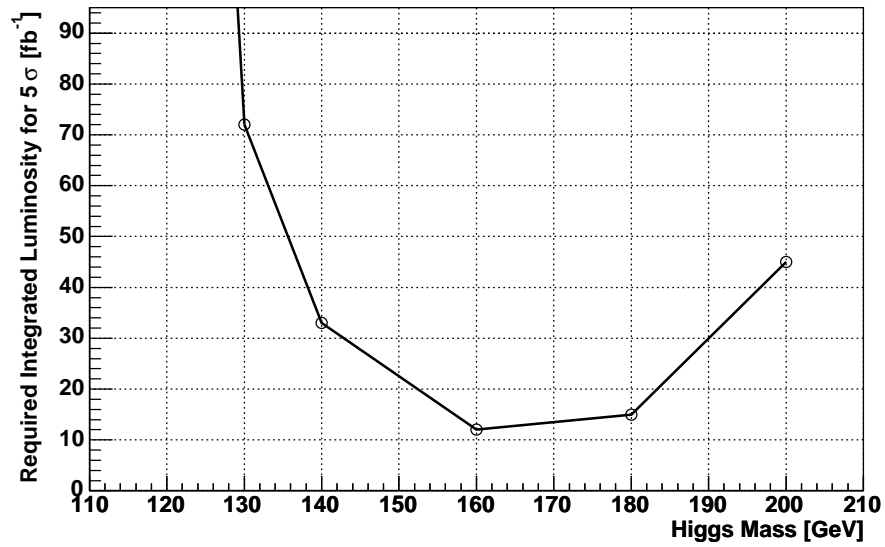
Figure 18: *Minimum integrated luminosity (fb^{-1}) needed to obtain a 5σ excess over the $t\bar{t}j + W^+W^-jj$ background as a function of the Higgs mass.*
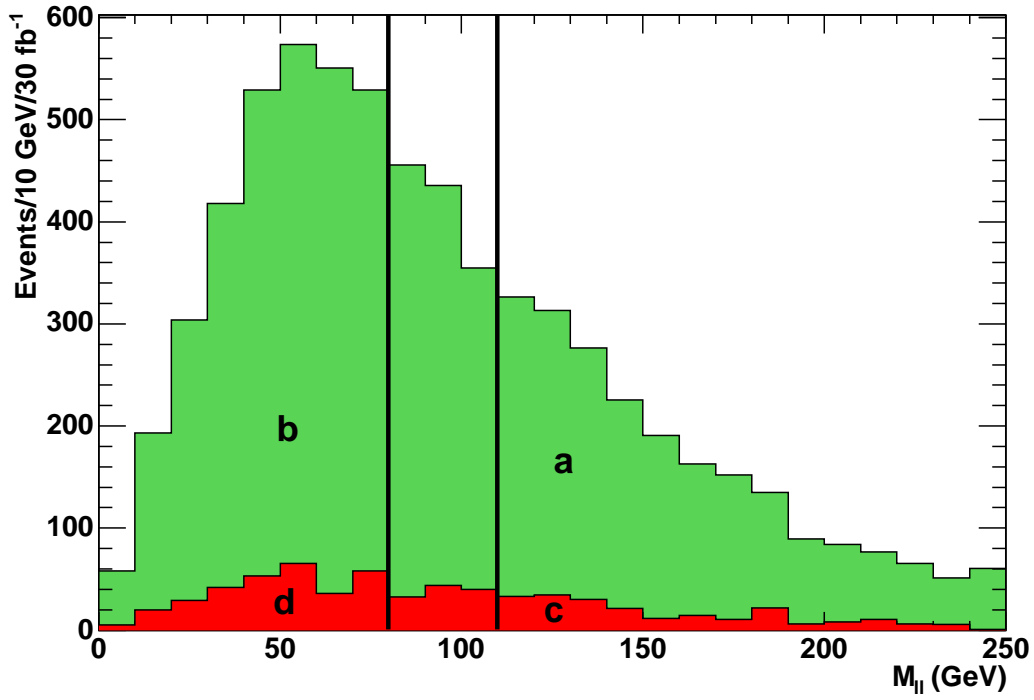
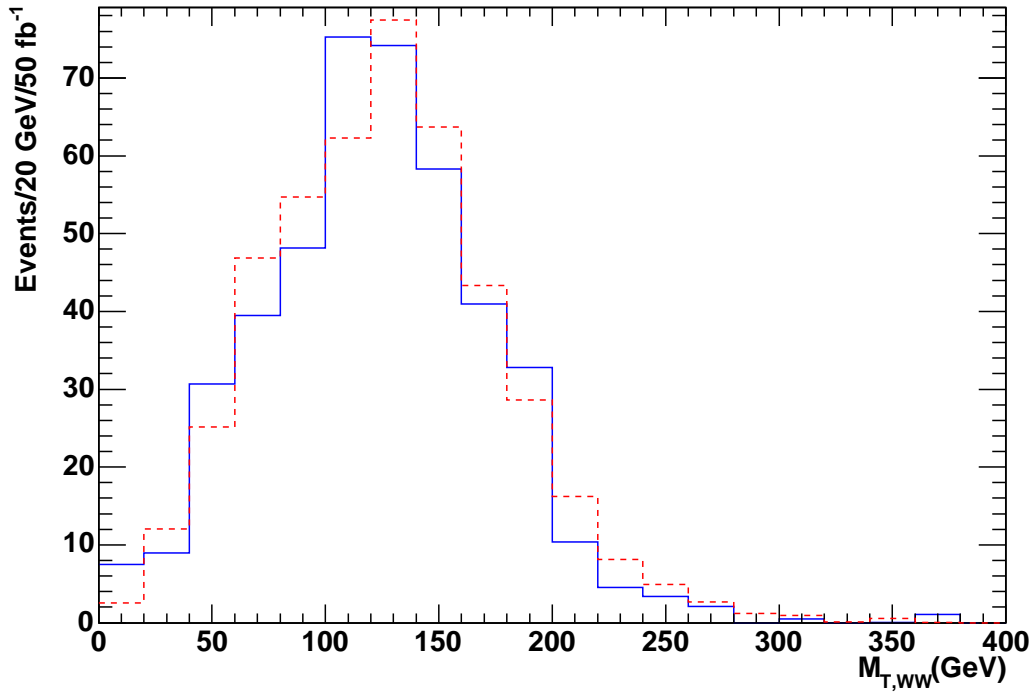Figure 19: *The $M_{\ell\ell}$ distribution computed with looser cuts and full analysis cuts.*

Figure 20: *The transverse mass, $M_{T,WW}$, distribution for estimated(dashed) and real(solid) background.*
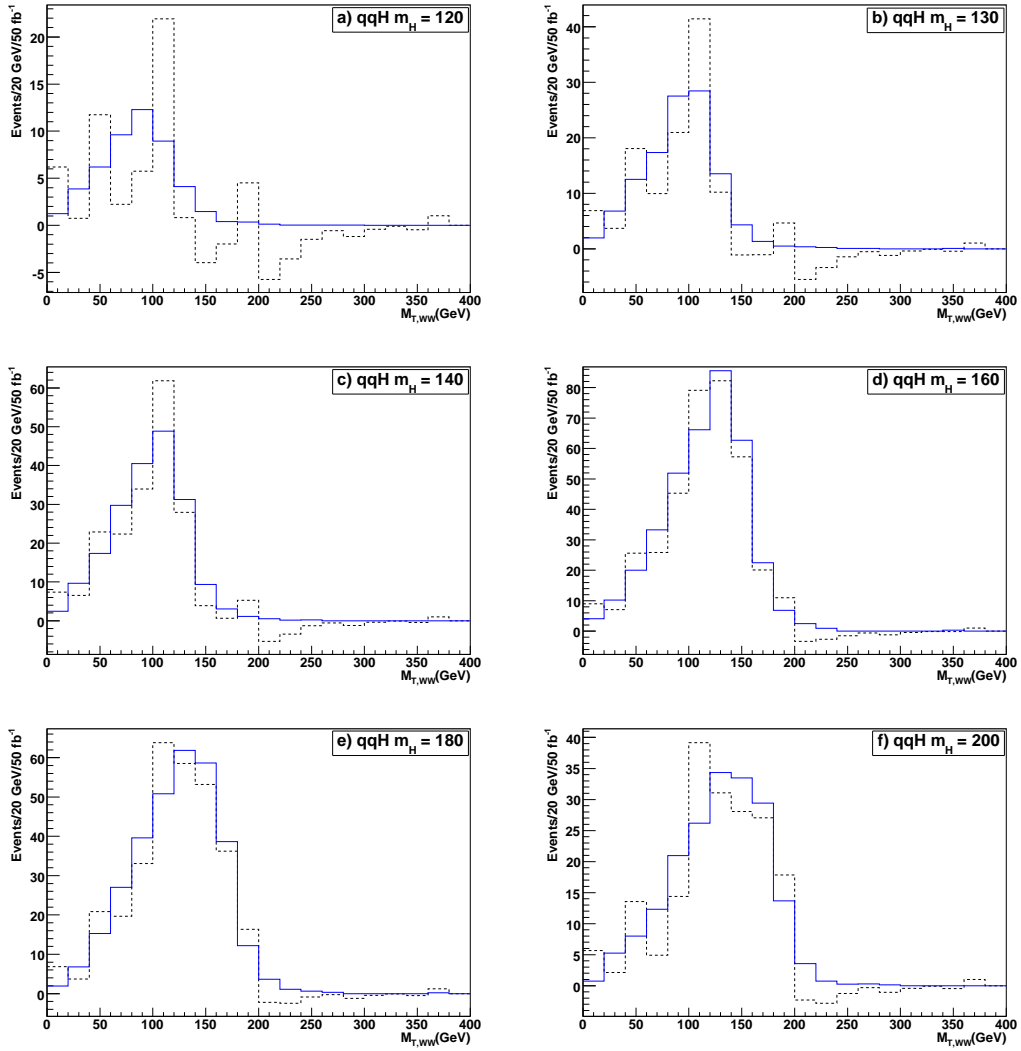
Figure 21: *Estimated(dashed) and real(solid) $M_{T,WW}$ distributions for signal events, with Higgs mass of 120,130,140,160,180 and 200 GeV shown in a),b),c),d),e) and f) respectively.*
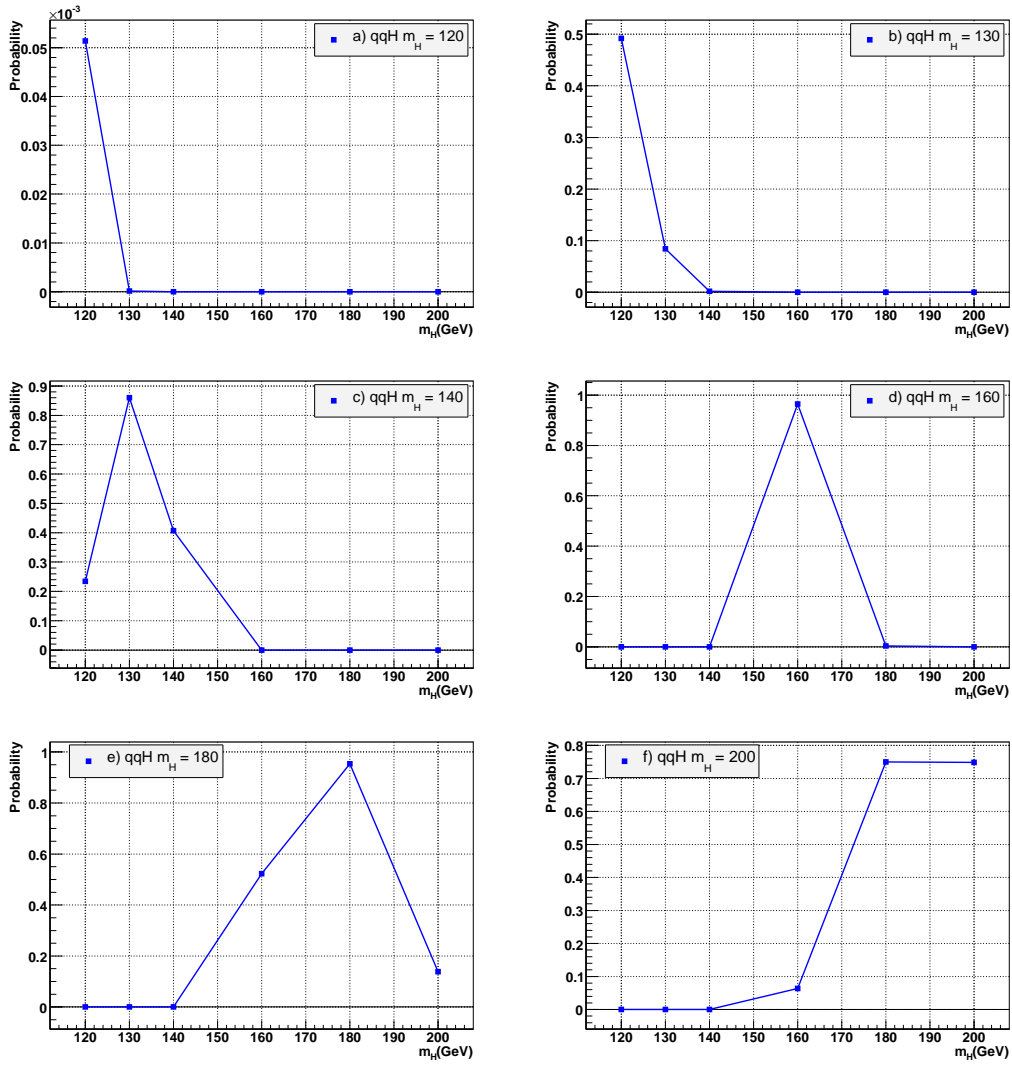
Figure 22: *Kolmogorov test function for estimating the Higgs boson mass for Higgs masses of 120,130,140,160,180 and 200 GeV shown in a),b),c),d),e) and f) respectively.*