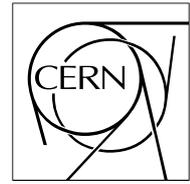


The Compact Muon Solenoid Experiment

CMS Note

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



19 April 2006

Observation of the Standard Model Higgs boson via $H \rightarrow \tau\tau \rightarrow \text{lepton} + \text{jet}$ Channel

C. Foudas, A. Nikitenko, M. Takahashi

Imperial College London, U.K.

Abstract

The potential of observing the Standard Model Higgs boson with a mass below $150 \text{ GeV}/c^2$ via the $H \rightarrow \tau\tau \rightarrow \text{lepton} + \text{jet}$ decay channel with Vector Boson Fusion production process has been evaluated using full simulation of the CMS detector. A number of background rejection methods based on the characteristics of the VBF process have been studied. It is shown that about 10 Higgs boson events are expected to be observed with a significance of 3.9σ (Higgs mass = $135 \text{ GeV}/c^2$) using an integrated luminosity of 30 fb^{-1} .

1 Introduction

The Standard Model Higgs boson decay to a pair of τ leptons ($H \rightarrow \tau\tau$) is an important channel when the Higgs boson mass is below the threshold that permits decays to vector bosons. The CMS detector is able to reconstruct the leptonic and hadronic decays of the τ leptons with high efficiency and precision. The Vector Boson Fusion (VBF) production process provides characteristic signatures of two forward-backward quark jets, which can be used to distinguish the Higgs boson signal from background processes. The production cross section of the VBF process is the second highest at the LHC, and the $H \rightarrow \tau\tau$ decay channel has a moderate branching ratio, which enables the observation of the Higgs boson within a first few years of physics data taking at the LHC.

A particular signature of the VBF events is the rapidity separation of the leading quark jets due to the absence of the colour exchange between the quarks; the quarks from the production process emerge at high pseudorapidity, η , leaving an open space in the centre of the detector enabling observation of isolated Higgs boson decay products. In addition to the high p_T decay products of the Higgs boson, the two outgoing quark jets can be tagged to identify the signal events. The colour coherence between the initial and the final state radiation from each quark leads to a suppression of hadron production in the central region between the two tagging jets [1]. This becomes an important criterion in distinguishing the signal process from the dominant physics backgrounds arising from QCD production of Z bosons, which involves exchange of gluons [2].

The analysis of the $H \rightarrow \tau\tau \rightarrow$ lepton + jet channel via VBF has first been carried out by Rainwater et al. [3, 4] using a parton level simulation. A further study has been carried out using a fast simulation of the CMS detector in the absence of pile-up events (Nikitenko et al. [5]). The studies have introduced a number of selection criteria, and demonstrated that the background contributions can be greatly suppressed in the signal region.

2 Signal and Background Processes and Sample Generation

2.1 Signal $qqH, H \rightarrow \tau\tau \rightarrow lj$

The signal events have been generated using PYTHIA for four different values of the Higgs boson mass: 115, 125, 135 and 145 GeV/c^2 . The Higgs boson is forced to decay to two τ leptons with one τ decaying to leptons and the other τ to hadrons. The TAUOLA package [6, 7] has been used to simulate the τ polarisation.

The Next to Leading Order (NLO) cross sections for the Higgs boson production via VBF and the branching ratios for the $H \rightarrow \tau\tau$ decay channel are listed in Table 1 for the four different mass scenarios. For the calculation of the production cross sections, the CTEQ6M [8] parton distribution function (PDF) is used, and the top quark mass of 175 GeV/c^2 is assumed.

Table 1: The NLO cross sections for the Higgs boson production via VBF and the branching ratios for the $H \rightarrow \tau\tau$ decay channel for different masses of the Higgs boson. (Provided by M. Mangano to the CMS Higgs analysis group)

$M_H \text{ GeV}/c^2$	115	125	135	145
VBF production, σ [pb]	4.65	4.30	3.98	3.70
Branching ratio, $\text{BR}(H \rightarrow \tau\tau)$	7.409×10^{-2}	6.258×10^{-2}	4.532×10^{-2}	2.685×10^{-2}
$\sigma \times \text{BR}(H \rightarrow \tau\tau \rightarrow lj)$ [fb]	157.3	112.9	82.38	45.37

2.2 Background Processes

The irreducible physics background processes to $qqH, H \rightarrow \tau\tau \rightarrow lj$ channel are the QCD and Electroweak (EW) production of two τ leptons from the Z/γ^* with associated jets. The reducible background processes considered in this analysis are the W + multi-jet production and $t\bar{t}$ events, in which one of the jets can be misidentified as a τ -jet. The events have been preselected at generator level (except for $t\bar{t}$), using criteria based on the detector and reconstruction limits as well as on the kinematical cuts introduced by the previous studies [3, 4, 5].

QCD process with $2\tau+2/3j$

The QCD production of $2\tau+2$ jet and $+3$ jet events with the invariant mass of two τ leptons, $M_{\tau\tau} > 70 \text{ GeV}/c^2$, has been generated using Alpgen with CTEQ5L PDF. Given the limit of the detector acceptance and requirements in the course of the event reconstruction, all jets are required to satisfy $p_{Tj} > 20 \text{ GeV}$, $|\eta_j| < 5.0$ and $|\Delta R_{jj}| > 0.5$. Further preselections are applied on the two highest p_T jets ($j1$ and $j2$) reflecting the offline VBF selection cuts: $|\Delta\eta_{j1j2}| > 4.0$, $M_{j1j2} > 600 \text{ GeV}/c^2$.

The ME calculations in Alpgen enable simulation of the jet behaviour in multi-jet events. The 2-jet and 3-jet events have been generated independently, and are processed for hadronisation using PYTHIA. The MLM prescription has been applied to avoid double counting of the events (Ref. [9]). The 2-jet event sample contains strictly 2 jets (exclusive) and the 3-jet event sample allows additional emission of partons which may form extra jets (inclusive). The TAUOLA package is used in PYTHIA to force one τ lepton to decay leptonically and the other hadronically.

EW production of $2\tau+2j$:

The EW production of two τ 's with $M_{\tau\tau} > 70 \text{ GeV}/c^2$ and two jets in the final state has been generated using MadGraph with CTEQ5L PDF. Preselection criteria similar to those used with the QCD Z+jets generation are applied: $|\eta_j| < 5.2$, $|\Delta R_{jj}| > 0.5$, $p_{Tj} > 20 \text{ GeV}$ and $M_{jj} > 500 \text{ GeV}/c^2$. An additional requirement is imposed on the separation of two τ 's, $|\Delta R_{\tau\tau}| > 0.4$. The showering and hadronization of the MadGraph parton level events are carried out using PYTHIA where all decay modes of the τ lepton are open.

W+jets ($W \rightarrow \mu\nu$)

The W+jet events with $W \rightarrow \mu\nu$ decays have been generated using Alpgen with CTEQ5M PDF. The exclusive W+3jet and inclusive W+4jet samples are generated separately, using the same method as with the QCD $2\tau+2/3$ jet samples. In addition to the kinematical cuts on jets used for the QCD $2\tau+2/3$ jet production described above, further preselections are made based on the lepton properties with $|\eta_l| < 3$ and $p_{Tl} > 10 \text{ GeV}/c$.

$t\bar{t} \rightarrow WbWb$ ($W \rightarrow l\nu$)

The $t\bar{t}$ background has been generated using five different event generators, with all leptonic W decays included and without any kinematical preselection applied.

2.3 Sample Summary and Software Environment

A summary of all the data samples and the software used for the simulation are shown in Table 2. The listed cross sections are calculated after the preselection. The number of simulated events available for each background process corresponds approximately to that expected at an integrated luminosity of 30 fb^{-1} . All generators implement the initial and final state radiation and multiple parton interactions. The pile-up of minimum bias (MB) events expected during the low luminosity phase ($L = 2 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$) is added to the generated events.

For the event reconstruction and analysis, ORCA_8.7.3 [15] has been used with a special Jets package (provided by O. Kodolova, and implemented later in ORCA_8.7.4) which removes fake jets around $\eta \sim 3$.

Table 2: Summary of the data samples used for the analysis listing names of the software used for event generation and detector simulation, whether kinematical preselection has been applied to the events at generation, expected cross section (σ) at the LHC, and the number of events available for the analysis. The * mark on the software indicates that it has been interfaced with PYTHIA for the simulation of hadronisation.

Sample	Event Generator	Preselection	N events	σ [fb]
Signal	PYTHIA [10] (TAUOLA [7])	No	50k / mass	
QCD $2\tau+2/3$ jet	Alpgen [11]* (TAUOLA [7])	Yes	24k (2jet) 61k (3jet)	468.7 (2jet) 1147. (3jet)
EW $2\tau+2$ jet	MadGraph [12]*	Yes	13k	229.
W+3/4jet ($W \rightarrow \mu\nu$)	Alpgen [11]* (TAUOLA [7])	Yes	140k (3jet) 300k (4jet)	4558. (3jet) 9888. (4jet)
$t\bar{t}$ ($W \rightarrow l\nu$)	PYTHIA [10]	No	$\sim 1.9\text{m}$	86×10^3
	TopReX [13]*	No		
	Alpgen [11]*	No		
	CompHEP [14]* MadGraph [12]*	No No		

3 Event Reconstruction

The signatures from the signal $qqH, H \rightarrow \tau\tau \rightarrow lj$ process are a high p_T lepton (electron or muon) and a τ -jet, two energetic forward jets and the total missing E_T of the system. The object reconstruction is carried out using the standard ORCA algorithms. The Higgs boson mass is calculated from the two reconstructed τ leptons, where the τ energies are calculated using collinear approximation of visible part of τ decay products and neutrinos.

3.1 Basic Object Reconstruction

The tracks are reconstructed using the CombinatorialTrackFinder which is based on the Kalman filter technique. The reconstructed tracks are required to contain at least 8 hits out of the total of 14 silicon pixel/strip layers in the inner tracker, and those with $p_T > 1$ GeV/c are used for the analysis. The ‘‘signal’’ vertex in an event, which is a p-p collision point with the highest p_T scattering event (e.g. the Higgs vertex in the signal process), is defined by the track impact parameter (IP) of the reconstructed lepton (which is identified first in our event reconstruction procedure). A track is associated to the signal vertex if their z-IP is within $\Delta z < 0.2$ cm from that of the lepton track.

The jets are reconstructed from the calorimeter towers (ECAL+HCAL) using the iterative cone algorithm. The cone radius of $R = 0.5$ and the seed tower E_T threshold of 0.8 GeV are used. Further cuts are applied to the constituent towers: $E_T > 0.5$ GeV and $E > 0.8$ GeV. For verification purposes, jets are also reconstructed using stable Monte Carlo hadrons originating from the signal vertex. The jet E_T is corrected using MC jet calibration [16].

3.2 Lepton Identification

The lepton from the signal event is identified as the highest p_T electron or muon candidate provided by Offline-ElectronReco or GlobalMuonReconstructor respectively. A muon candidate is defined as a track extending from the central tracking system to the outer muon system, and an electron candidate as an ECAL supercluster with an associated track. The lepton $p_T > 15$ GeV/c is required (which is set by the trigger thresholds). The p_T resolution is 1.4% for the muon tracks, and 2.5% for the electron superclusters.

The muon reconstruction has a very low fake rate, and the impurity is less than 1%. However, the electron candidates suffer from high contamination which comes from soft jets and τ -jets with one charged track. A number of jet rejection criteria are suggested by electron identification studies [16]: $E/P, H/E, 1/E - 1/P, \Sigma P/E$. E is the energy of the supercluster in the ECAL, P is the magnitude of the track momentum, and H is the energy in the HCAL tower behind the ECAL supercluster. The parameter, $\Sigma P/E$, checks the tracker isolation using a sum of all tracks which are within the boundary, $0.01 < \Delta R < 0.2$, around the electron track.

Studies of τ -jet identification [17] has introduced another parameter, the E_T of the highest E_T HCAL tower in the jet, E_T^{hotH} . It has been shown that a cut based on E_T^{hotH} can effectively discriminate against electrons passing the τ -tagging criteria. This method is adapted to the electron ID by using a jet that matches the electron within $\Delta R < 0.3$.

The efficiencies of selecting electrons and discriminating against τ -jets and jets, which are selected by OfflineElectronReco and of $E_T > 15$ GeV, using these variables are summarised in Table 3. The electrons and τ -jets are taken from the signal sample ($M_H = 135$ GeV/c²) and jets are from the background $W+3\text{jet}$ sample. The cut values are those proposed by the corresponding studies. The efficiency, ε_{obj} , is defined as,

$$\varepsilon_{\text{obj}} = \frac{\text{No. passing the selection}}{\text{No. of electron candidates with ID = obj}} \quad (1)$$

where the object ID is obtained via the matched simulated track information. It can be seen that E_T^{hotH} (e) is the single most powerful rejection parameter against both τ -jets and jets. In this analysis, a combination of three parameters (f) is used:

$$E/P > 0.9, \quad E_T^{\text{hotH}} < 2, \quad \Sigma P/E < 0.05$$

E/P (a) is not a powerful discrimination parameter, however, the lower cut is used in order to select ECAL superclusters with a good energy resolution.

Table 3: The efficiencies (ε) of selecting electrons, τ -jets and jets for different discrimination criteria. The ratios of the efficiencies, $\varepsilon_\tau/\varepsilon_e$ and $\varepsilon_j/\varepsilon_e$, are shown in parentheses. The electrons and τ -jets are taken from the signal sample ($M_H = 135 \text{ GeV}/c^2$) and jets are from the background W+3jet sample.

Selection Criteria	electron	τ -jet		jet	
	ε_e %	ε_τ %	$(\varepsilon_\tau / \varepsilon_e)$	ε_j %	$(\varepsilon_j / \varepsilon_e)$
(a) $0.9 < E/P < 2.0$	87.6	32.6	(0.37)	32.7	(0.37)
(b) $H/E < 0.05$	97.1	26.3	(0.27)	13.9	(0.14)
(c) $ 1/E - 1/P < 0.02$	94.4	68.8	(0.73)	68.7	(0.73)
(d) $\Sigma P / E < 0.05$	95.5	36.2	(0.38)	16.5	(0.17)
(e) $E_T^{\text{hotH}} < 2 \text{ GeV}$	91.0	10.3	(0.11)	2.9	(0.03)
(f) $E/P > 0.9 + (d) + (e)$	85.1	5.5	(0.06)	0.9	(0.01)

3.3 τ -jet Identification

The τ -jet identification is seeded from the trigger τ candidates. A jet is searched within a region of radius $R = 0.8$ in the direction of each candidate, and the jet is passed through a series of tagging criteria similar to those used in the HLT [17, 18]. The iterative cone algorithm is used for the τ -jet reconstruction with a cone size of 0.4. For events which are selected by τ related triggers at HLT, τ -jet identification starts from the HLT τ candidate, then descends to the L1 candidates if the HLT candidate does not satisfy all the criteria. If the event has been triggered by single lepton triggers at HLT, then the search starts from the highest E_T L1 τ candidate. If the jet coincides with any of the electron candidates within $\Delta R < 0.3$, then the ones with $E_T^{\text{hotH}} < 2 \text{ GeV}$ are excluded. The highest E_T candidate which passes all the criteria is selected.

The τ tagging parameters used for this analysis are: search cone size - $R_m = 0.1$, signal cone size - $R_s = 0.07$, isolation cone size - $R_i = 0.45$, and p_T of the leading track - $p_{T\text{tr}} = 6 \text{ GeV}/c$. The reconstructed tracks from the signal vertex, which is identified by the lepton track, are used. The sum of the track charges for the 3-prong τ -jet is required to be ± 1 , and the charge of a τ -jet has to be opposite of the selected lepton charge. The list of τ tagging criteria and their efficiencies for selecting MC τ -jets in the signal sample ($M_H = 135 \text{ GeV}$) using simulated and reconstructed tracks are shown in Table 4. The reconstructed tracks are associated to the MC vertex, hence the results do not include lepton identification efficiency and accuracy of the measured lepton track IP, as it would in the actual τ search. More than 45% of the total τ -jets are lost to the p_T cut and the limitation of the tracker acceptance. A further $\sim 20\%$ reduction comes from the isolation criteria, which depends on the physics of the events.

The energy correction is applied to τ -jets using the calibration constants optimised for τ -jets. A detailed description and the performance of the τ -jet calibration can be found in [16]. The τ -jet candidate is required to satisfy $E_T > 30 \text{ GeV}$ after the calibration. The resolution of the τ -jet E_T is 11.3% for both the 1-prong and 3-prong decays. The impurity of selected τ -jets is 2.7% in the signal events.

Table 4: The efficiency of τ -tagging applied on MC τ -jets using simulated and reconstructed tracks. Numbers shown are of MC τ -jets (normalised to an initial number of 1000) that has passed successive selections, and in () are the efficiency of the individual cuts.

Tagging Criteria	Number of MC τ -jet remained	
	using sim tracks	using rec tracks
[total number]	1000	
tracker acceptance ($ \eta < 2.3$)	834 (83.4)	
τ -jet $p_T > 30 \text{ GeV}$	547 (65.6)	
jet-leading track matching ($R_m = 0.1$)	543 (99.2)	449 (82.2)
1 or 3 tracks within signal cone ($R_s = 0.07$)	517 (95.2)	385 (85.7)
track charge consistency	516 (99.9)	378 (98.2)
leading track with $p_{T\text{tr}} > 6. \text{ GeV}/c$	479 (92.7)	376 (99.5)
no track in isolation cone ($R_i = 0.45$)	382 (79.9)	311 (82.5)
% passing all	38.2%	31.1%
[provided $ \eta_{\tau\text{jet}} < 2.3$ and $p_{T\tau\text{jet}} > 30 \text{ GeV}$]	[69.9%]	[56.8%]

3.4 Forward Jet-pair Identification

In the signal qqH, $H \rightarrow \tau\tau \rightarrow lj$, the forward quarks from the production process are the only high E_T jets in the event. These quark jets are identified as two highest E_T jets; in order to avoid double-counting of electrons and τ -jets, jets within $\Delta R < 0.3$ of the electron and τ -jet candidates are excluded. The jet E_T is required to be greater than 40 GeV and $|\eta| < 4.5$. The η of the two jets are required to have opposite sign, reflecting the kinematics of the signal event: $\eta_{j1} \cdot \eta_{j2} < 0$.

Further restrictions based on jets are imposed to identify the VBF signature, which will be discussed later in connection to the background rejection.

3.5 Missing E_T Reconstruction

The missing E_T (\cancel{E}_T) is measured by taking the opposite vector of the total visible E_T detected. The energies of the calorimeter towers are separated into x and y components based on the tower position, and summed over all ϕ . The p_T of the reconstructed muons is separately subtracted from the total missing E_T , since their energy deposition in the calorimeter is negligible.

The \cancel{E}_T is corrected using the energy correction applied to individual jets (Ref. [19]) using jets with $E_T^{\text{raw}} > 20$ GeV, where “raw” corresponds to the E_T measured in the calorimeters without the calibration. The electrons reconstructed as jets are excluded from the \cancel{E}_T correction. In addition, the τ -jets which require different energy scaling are treated separately. Matching between the jets and the electrons/ τ -jets are made by requiring $\Delta R < 0.3$. The \cancel{E}_T resolution after the correction is 20%; this is the largest contribution towards the resolution of the reconstructed Higgs boson mass.

4 Event Selection and Background Rejection

4.1 Triggers

The following Level-1 (L1) and Higher Level Triggers (HLT) are used:

- L1 – single isolated electron, single muon and a combined e- τ triggers
- HLT – single isolated electron, single muon and combined e- τ and μ - τ triggers

The logical OR of all the triggers are used at each level to select events.

4.2 Kinematical Selection

The kinematical selections introduced in association with the object reconstruction are as follows:

- lepton $p_T > 15$ GeV
- τ -jet $E_T > 30$ GeV
- forward jets $E_T > 40$ GeV

The cut-off value for the lepton reflects the trigger thresholds and that for the τ -jet ensures a good τ -jet identification. The cuts are also effective for reducing the QCD and EW 2τ +jets background events (Fig. 1).

A number of selection criteria based on the dynamics of the event have been suggested by the previous studies [3, 4, 5]. The η and ϕ separation and the invariant mass of the two forward jets are used to characterise the kinematics of the VBF events, which is different to that of the QCD background processes. The transverse mass of the lepton and the missing E_T , $M_T(\text{lep}, \cancel{E}_T)$ is used to discriminate against W boson related backgrounds which form a Jacobian peak in the $M_T(\text{lep}, \cancel{E}_T)$ distribution. The distribution of the four parameters are shown in Figure 2 (a) - (d).

The cuts applied on the four parameters in this analysis are:

- η separation of the forward jets, $\Delta\eta_{j1j2} > 4.2$
- ϕ separation of the forward jets, $\Delta\phi_{j1j2} < 2.2$
- Invariant mass of the forward jets, $M(j1,j2) > 1$ TeV
- Transverse invariant mass of the lepton- \cancel{E}_T system, $M_T(\text{lep}, \cancel{E}_T) < 40$ GeV

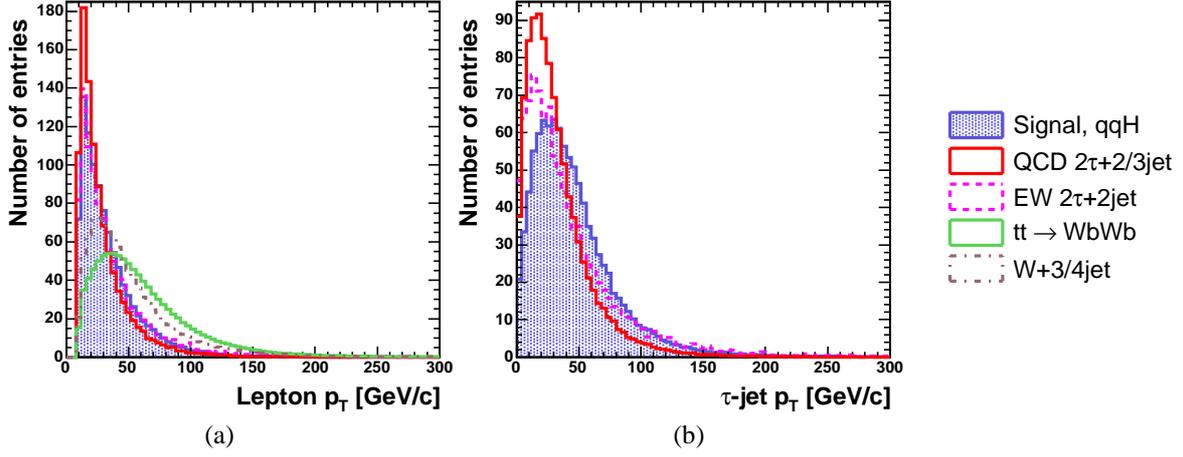


Figure 1: The p_T distribution for the lepton (a) and the τ -jet (b) for the different signal and background events. The total number of entries in each histogram is normalised to 1000 events. A preselection is applied for the leptons at $p_T > 10$ GeV/c.

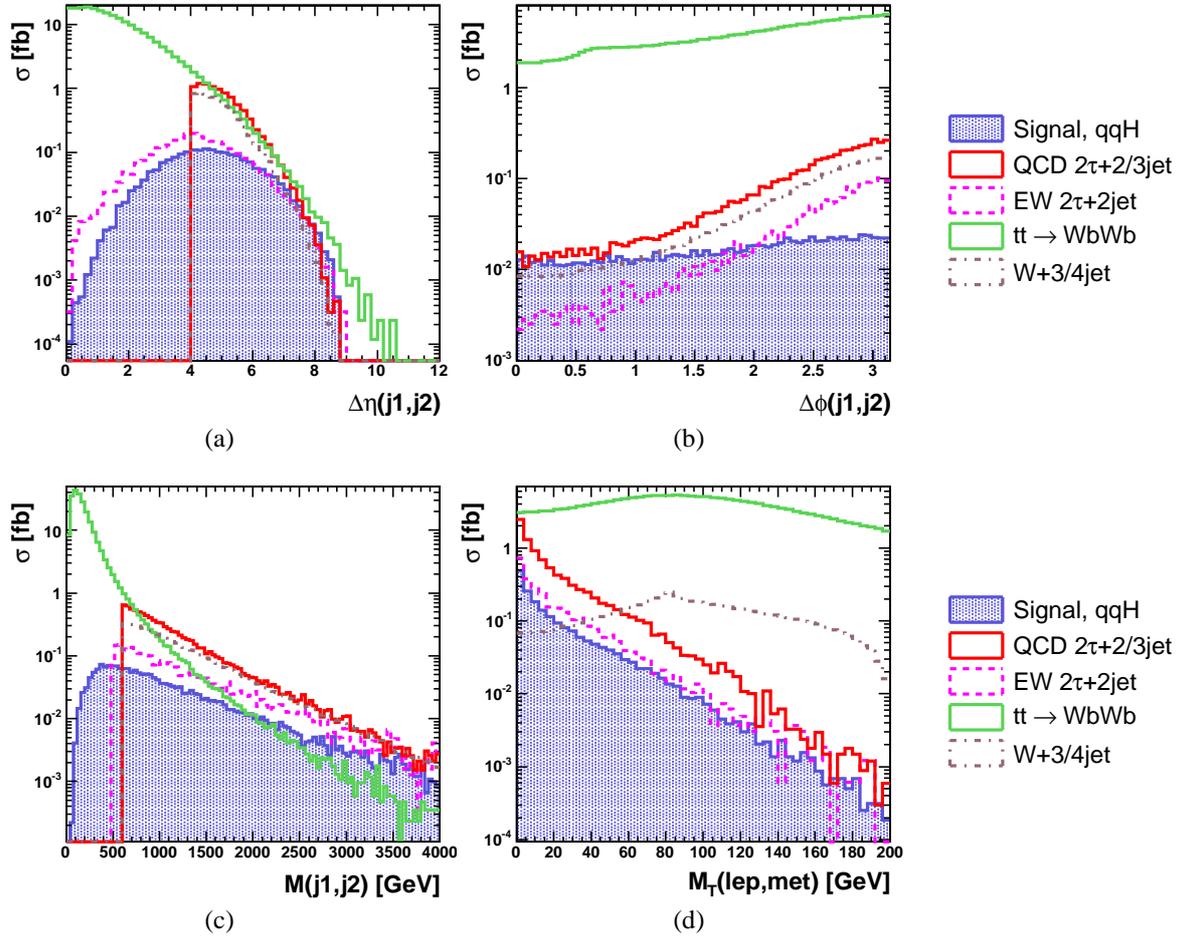


Figure 2: The distribution of $\Delta\eta$ (a), $\Delta\phi$ (b) and invariant mass (c) of the two forward jets, and (d) transverse mass of the lepton- \cancel{E}_T system, for the different signal and background events. The total number of entries in each histogram is normalised to the cross section expected after the lepton, τ -jet, and forward jets identification and the corresponding p_T cuts. The cut-off shown in (a) and (b) for the QCD/EW $2\tau+2/3jet$ and $W+3/4jet$ samples corresponds to the preselection applied.

4.3 Central Jet Veto

It has been mentioned in Introduction that the hadronic activity is heavily suppressed in the EW production of Z and Higgs bosons whereas in QCD processes the additional jets appear in the central region of the detector. A veto on such central jets has been proven a successful rejection method against the QCD background events, as demonstrated by the previous studies [3, 4, 5].

Additional Central Jets

An additional jet, *3rd jet*, is the jet which lies between the rapidity gap of the two highest E_T forward jets,

$$(\eta_{\min} + 0.5) < \eta_{j3} < (\eta_{\max} - 0.5) \quad (2)$$

where subscripts min and max correspond to the forward jet with the smaller and the larger value of η respectively. A margin of 0.5 is required to avoid counting soft radiation near high E_T jets. This is automatically satisfied in the iterative cone algorithm with cone size 0.5, used for jet reconstruction in this analysis. The jets that coincide with the selected electron or τ -jet within $\Delta R < 0.3$ are ignored in the search. An event is rejected if there exists such a 3rd jet above certain threshold E_T , $E_{Tj3} > E_{Tth}$.

The efficiencies of the veto for the different background data samples are plotted versus the signal efficiency ($M_H = 135 \text{ GeV}/c^2$ sample) in Figure 3, with a varying value of E_{Tth} . The efficiency is evaluated using MC jets, and defined as:

$$\text{efficiency} = \frac{\text{events containing a 3rd jet with } E_T > E_{Tth}}{\text{all events passing the VBF criteria}} \quad (3)$$

where the VBF criteria used here are: $|\eta| < 4.5$, $\eta_{j1} \cdot \eta_{j2} < 0$, $E_{Tj} > 40 \text{ GeV}$ and $\Delta\eta > 4.2$. An optimal threshold would be $E_{Tth} \sim 20 \text{ GeV}$, where the signal process has a nearly 80% efficiency while all the background processes are suppressed below 50%. The threshold value has been chosen at an equivalent point in terms of raw E_T in this analysis, with a value, $E_{Tth}^{\text{raw}} = 10 \text{ GeV}$.

The rapidity distribution of the 3rd jets with respect to the two forward jets are shown in Figure 4(a) for the signal and the background samples. The η position is measured relative to those of the forward jets, defined as,

$$\eta_{j3}^* = \eta_{j3} - (\eta_{\min} + \eta_{\max}) \quad (4)$$

The η_{j3}^* distribution for the EW processes – the VBF in the signal sample and the EW 2τ +jet sample – show the double-peak nature, whereas the 3rd jets are central with respect to the forward jets for the other QCD background samples. This agrees with the behaviour shown in the parton-level study [2].

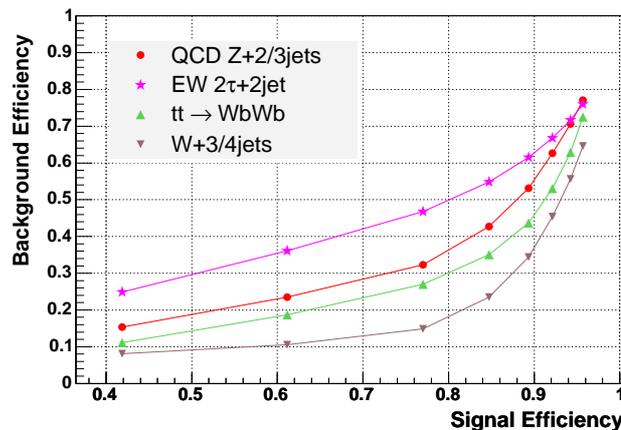


Figure 3: The efficiency of the veto for background versus that for the signal ($M_H = 135 \text{ GeV}/c^2$ sample), for the different thresholds, $E_{Tth}^{\text{MC}} = 10$ (the lowest signal efficiency point), 15, 20, ... 45 GeV (the highest signal efficiency point).

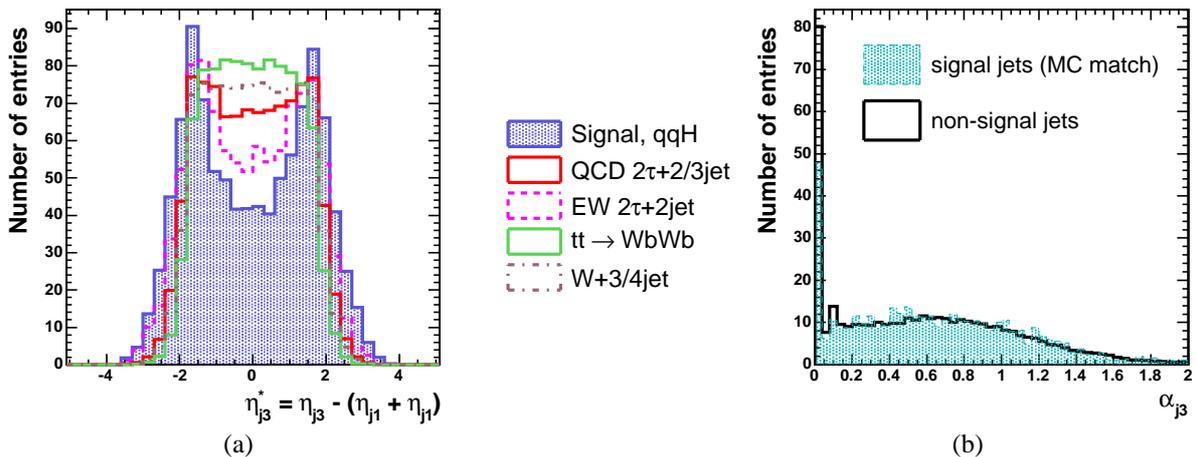


Figure 4: (a) The η distribution of the 3rd jet with respect to the two forward jets, $\eta_{j3}^* = \eta_{j3} - (\eta_{\min} + \eta_{\max})$. (b) The distribution of the parameter, α_{3j} , which is used to match the jets to the signal vertex. The total number of entries in each histogram is normalised to 1000 events.

Matching Jets to the Signal Vertex

In order to avoid considering jets from pile-up events for a veto, jets are associated to the signal vertex using tracks. Studies of jet veto for the decay channel, qqH , $H \rightarrow WW$, has introduced a parameter which is defined for every 3rd jet (Ref. [16]):

$$\alpha_{j3} = \Sigma p_{T\text{trk}} / E_{Tj3} \quad (5)$$

where $p_{T\text{trk}}$ is the p_T of the tracks originating from the signal vertex which are within the jet reconstruction cone, and E_{Tj3} is the raw E_T of the jet measured in the calorimeter. The sum is made over all tracks which lie within the $R = 0.5$ cone around the jet axis. The value of α_{3j} depends on the fraction of neutral objects contained in the jet, however, it should always be zero for a jet originated from a vertex different from the signal vertex.

The α_{3j} distribution for the reconstructed jets with $E_T^{\text{raw}} > 10$ GeV is shown in Figure 4(b). The reconstructed tracks associated to the MC vertex are used. Jets from the signal vertex and those from other pile-up vertices are distinguished by the presence/absence of MC jets with $E_T > 20$ GeV within $\Delta R < 0.3$. The largest difference is seen in the first few bins corresponding to the lowest α_{3j} , hence jets with $\alpha_{3j} > 0.1$ is considered for the veto. Jets from the signal vertex may have the value of $\alpha_{3j} = 0$ due to the limited tracker acceptance range. In some cases, the calorimeter jets are comprised of purely neutral hadrons within the reconstruction cone; these jets also find no associating track.

5 Results

5.1 Selection Efficiency

The efficiency of each reconstruction and selection step and the cumulative cross section expected at the LHC are given in Table 5. For the W+3/4jet samples, the efficiencies of some selection cuts have been obtained from factorisation of cuts, as explained later in the section.

For signal events, the main reduction arises from the online triggers, τ -jet identification and the VBF related cuts. About 15% of τ -jets is lost because they are outside the tracking region. The lepton identification efficiency is $\sim 95\%$ since the tracker acceptance requirement is already satisfied by the lepton triggers. The VBF cuts serve as good background rejection methods, given that three of the background samples (QCD Z+2/3jet, EW $2\tau+2$ jet and W+3/4jet) are preselected at the generator level to satisfy relaxed cuts. The total selection efficiencies (excluding the factor from the Higgs decay branching ratio) are, 0.32%, 0.34%, 0.42%, 0.39%, for the simulated Higgs boson masses of 115, 125, 135 and 145 GeV/ c^2 respectively.

For the background samples, the efficiencies vary depending on the preselection and the type of physics involved. The τ -jet identification efficiency is minimal for the W+3/4jet sample, which has no real τ -jet, and for the $tt \rightarrow WbWb$ sample. The $M_T(\text{lep}, \cancel{E}_T)$ cut works effectively, rejecting more than a half of the W+3/4jet events and 65% of the $t\bar{t}$ events. A successful reduction of the QCD background events is achieved by the central jet veto.

Table 5: Cumulative cross sections in the units of fb after successive selection cuts. The entry, “valid mass”, corresponds to the fraction remained after the calculation of the di- τ mass; some events are lost due to the negative reconstructed neutrino energies. The efficiency (%) of the individual cut is listed inside the brackets. For the W+3/4jet samples, the efficiency is obtained from factorisation of cuts for the steps after the lepton identification and p_T cut, and the number of events at 30 fb^{-1} is calculated for all leptonic decay modes, $W \rightarrow l\nu$. The τ -jet ID efficiency for the W+3/4jet samples (indicated by the * mark) includes the initial τ -jet candidate selection, E_T cut, and tagging efficiencies. The VBF ID efficiency for the same samples (indicated by **) includes efficiencies from forward jet candidate identification with $\Delta\eta$ and E_T cuts.

Selection	Cumulative Cross Section [fb] (% from previous step)									
	signal samples: qqH, $H \rightarrow \tau\tau \rightarrow lj$				background samples					
	$M_H=115$	$M_H=125$	$M_H=135$	$M_H=145$	QCD $2\tau+2/3\text{jet}$		EW $2\tau+2\text{jet}$	tt \rightarrow WbWb ($W \rightarrow l\nu$)	W+3/4jet ($W \rightarrow \mu\nu$)	
					$2\tau+2\text{jet}$	$2\tau+3\text{jet}$			W+3jet	W+4jet
production σ	4.65×10^3	4.30×10^3	3.98×10^3	3.70×10^3	-	-	-	86×10^3	-	-
$\times \text{BR}(H \rightarrow \tau\tau \rightarrow lj)$	157.3 (3.4)	112.9 (2.9)	82.38 (2.1)	45.37 (1.2)	-	-	-	-	-	-
preselection	-	-	-	-	468.7	1147.	299.0	-	4558.	9888.
L1	81.81 (52.0)	60.76 (53.8)	46.50 (56.5)	26.36 (58.1)	132.6 (28.3)	411.3 (35.9)	179.8 (60.1)	71.39×10^3 (83.0)	2815. (61.8)	6371. (64.4)
L1 + HLT	41.46 (50.7)	31.39 (51.7)	24.60 (52.9)	14.19 (53.8)	52.53 (39.6)	148.7 (36.2)	58.81 (32.7)	55.43×10^3 (77.6)	2138. (76.0)	4472. (70.2)
Lepton ID	39.46 (95.2)	29.95 (95.4)	23.34 (94.9)	13.46 (94.9)	49.44 (94.1)	138.0 (92.8)	50.67 (86.2)	54.08×10^3 (97.6)	2119. (99.1)	4430. (99.1)
Lepton p_T	39.12 (99.1)	29.71 (99.2)	23.16 (99.3)	13.34 (99.1)	49.17 (99.4)	136.4 (98.9)	49.13 (97.0)	53.54×10^3 (99.0)	2118. (99.9)	4425. (99.9)
τ -jet ID	12.70 (32.5)	10.36 (34.9)	8.276 (35.7)	4.888 (36.7)	10.60 (21.6)	29.04 (21.3)	10.49 (21.3)	5056. (9.4)	(0.07)*	(0.31)*
τ -jet E_T	9.014 (71.0)	7.564 (73.0)	6.422 (77.6)	3.858 (78.9)	6.092 (57.5)	18.16 (62.5)	7.360 (70.2)	3215. (63.6)	-	-
valid mass	6.113 (67.8)	5.042 (66.7)	4.462 (69.5)	2.649 (68.7)	3.866 (63.5)	10.62 (58.5)	4.232 (57.5)	848.6 (26.4)	(25.0)	(13.3)
VBF ID (η , E_T)	2.718 (44.4)	2.192 (43.5)	1.949 (43.7)	1.081 (40.8)	1.679 (43.4)	7.462 (70.3)	2.944 (69.6)	222.9 (26.3)	(68.5)**	(52.1)**
VBF: $\Delta\eta_{jj}$	1.498 (55.1)	1.231 (56.1)	1.104 (56.6)	0.588 (54.4)	1.230 (73.3)	4.417 (59.2)	1.012 (34.4)	13.11 (5.9)	-	-
VBF: $\Delta\phi_{jj}$	1.174 (78.4)	0.928 (75.4)	0.806 (73.0)	0.427 (72.5)	0.723 (58.7)	2.481 (56.2)	0.460 (45.5)	9.380 (71.6)	(16.3)	(30.4)
VBF: M_{jj}	0.771 (65.7)	0.634 (68.4)	0.545 (67.7)	0.283 (66.4)	0.312 (43.2)	1.353 (54.5)	0.391 (85.0)	2.738 (29.2)	(50.8)	(65.6)
$M_T(l, \cancel{E}_T)$	0.620 (80.4)	0.476 (75.1)	0.423 (77.6)	0.207 (73.1)	0.254 (81.3)	1.128 (83.3)	0.322 (82.4)	0.942 (34.4)	(34.3)	(30.2)
CJV	0.503 (81.2)	0.382 (80.1)	0.344 (81.3)	0.175 (84.6)	0.254 (100.)	0.301 (26.7)	0.230 (71.4)	0.224 (23.8)	(80.1)	(21.7)
Events at 30 fb^{-1}	15.1	11.4	10.3	5.3	16.6		6.9	6.7	1.5 ($W \rightarrow l\nu$)	

Factorisation

The W+3/4jet sample has been generated with a number of events equivalent to that expected at an integrated luminosity of 30fb^{-1} . When all the selection cuts are applied successively, only one MC event for each sample has remained, which demonstrates that the event selection criteria are powerful enough to reduce the W+jet contribution in the signal region to a negligible level at 30fb^{-1} . In order to obtain a better estimate of the efficiencies of the background rejection cuts, the selection criteria are factorised in two separate streams after events have passed initial preselections. The initial preselections are the same as the standard procedure up to the lepton p_T cut. A pair of VBF forward quark jet candidates and a τ -jet candidate are then selected, with the following criteria:

- forward jet candidates – two highest E_T jets (excluding the electron) satisfying η and $\Delta\eta$ requirements
- τ -jet candidate – the highest E_T L1/HLT τ candidate (excluding the electron and the forward jet candidates) with $E_T > 30\text{ GeV}$

Two streams, A and B, are then carried out in parallel:

Stream A	Stream B
1. VBF p_{Tj}	1. central jet veto
2. VBF $\Delta\phi$	2. τ tagging
3. VBF M_{jj}	3. mass calculation
4. $M_T(\text{lep}, \cancel{E}_T)$	

The final cross section is calculated from the total efficiencies of stream A and B,

$$\sigma_f = \sigma_0 \times \varepsilon_A \times \varepsilon_B \quad (6)$$

where σ_0 is the cross section at the end of the initial preselections. Using this factorization, there are 2k and 15k MC events remained for W+3jet and W+4jet samples respectively after the initial preselection, and at least \sim tens of events are used to calculate the efficiency of each selection cut.

5.2 Reconstructed Mass

The distributions of the invariant mass of the reconstructed τ 's, $M_{\tau\tau}$, are shown in Figure 5 for the signal sample with $M_H = 135\text{ GeV}/c^2$ and for the background samples. The number of entries is normalised to the expected number of events at an integrated luminosity of 30 fb^{-1} . In order to understand the shapes of the background $M_{\tau\tau}$ distributions, some selection cuts have been relaxed. For the QCD and EW production of 2τ +jets the histogram is made with the distribution without the central jet veto. A further relaxation of cuts has been applied for the reducible background samples, hence the mass distribution is extracted immediately after the identification of the forward jets in the standard stream.

A gaussian distribution (F_H) is used to fit the signal distribution. For the resonance peak from the Z/γ^* decays, the Breit-Wigner function (F_Z) is used, and for the reducible background events a second order polynomial function (F_R) is used. Due to the over-corrected \cancel{E}_T , the calculated $M_{\tau\tau}$ is over-estimated. Both the Z boson and the Higgs boson masses are greater than the simulated values by $\sim 5\text{ GeV}/c^2$. The resolution of the Higgs boson mass obtained from the fit is 9.1% for the simulated Higgs boson mass of $135\text{ GeV}/c^2$.

5.3 Significance

The significance of the observed signal events, N_S , is estimated using the method, S_{cP} [20]. The algorithm uses a Poisson distribution with the mean = N_B to calculate the probability to observe $\geq N_B + N_S$ events, where N_B is the number of background events expected within a chosen mass range. The program allows for the systematic uncertainty to be included in the calculation in the form of Gaussian smearing of N_B with an input σ_B .

The significance for the different masses of Higgs boson are calculated using the number of events, estimated from the fits to individual samples, within a mass window which slides in steps of $5\text{ GeV}/c^2$. An optimum window position which maximises the significance is chosen for each mass. The width of the mass window is fixed as $40\text{ GeV}/c^2$, which corresponds to about $\pm 1.5\sigma$ of the reconstructed Higgs mass distribution. The calculated values of S_{cP} for the different signal samples are listed in Table 6 for an integrated luminosity of 30 fb^{-1} and 60 fb^{-1} . Systematic uncertainties of 7.8% for 30 fb^{-1} and 5.9% for 60 fb^{-1} are included. These arise from the fitting procedure, which will be explained in the following section.

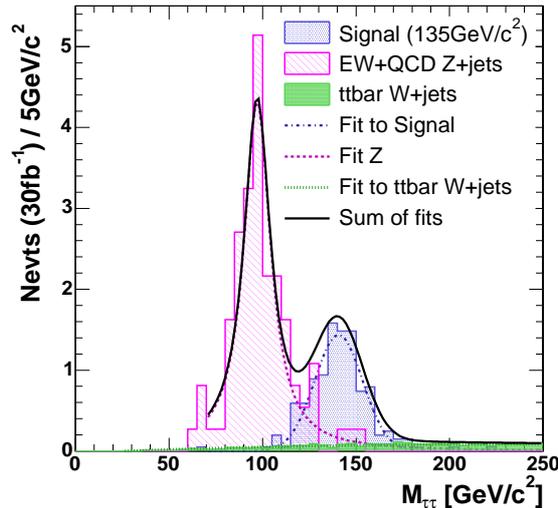


Figure 5: The invariant mass of the two reconstructed τ 's. The number of entries is normalised to the expected number of events at an integrated luminosity of 30 fb^{-1} .

The significance is the highest at $M_H = 115$ and $135 \text{ GeV}/c^2$. The branching ratio of the $H \rightarrow \tau\tau$ decay is high at low mass. However, the event selection efficiency increases and the contribution from the background events decreases with the Higgs mass. Given the presence of the Z mass peak at $M_{\tau\tau} \sim 91 \text{ GeV}/c^2$, the observation of the Higgs boson signals in the $115 \text{ GeV}/c^2$ region would be challenging without a further accumulation of the luminosity and a good understanding of the shape of the Z distribution. The factor of two decrease in the significance for the $M_H = 145 \text{ GeV}/c^2$ sample compared to the other masses is due to the fact that the Higgs boson branching ratio to two τ 's decreases rapidly as the bosonic decay channels start to dominate in this mass range. The values of S_{CP} exceed 5σ for masses below $140 \text{ GeV}/c^2$ at 60 fb^{-1} .

M_H [GeV]	115	125	135	145
N_S (30fb^{-1})	10.47	7.79	7.94	3.63
N_B (30fb^{-1})	3.70	2.21	1.84	1.42
S_{CP} at 30fb^{-1} (no uncertainty)	4.04	3.71	3.98	2.19
S_{CP} at 30fb^{-1} ($\sigma_B = 7.8\%$)	3.97	3.67	3.94	2.18
S_{CP} at 60fb^{-1} ($\sigma_B = 5.9\%$)	5.67	5.26	5.64	3.19

Table 6: The expected number of signal and background events and the significance for four different masses of the Higgs boson.

5.4 Fitting Procedure and Uncertainty

It is envisaged that the shapes of the two background distributions, F_Z and F_R , will be extracted experimentally from the LHC data in a region unaffected by the signal contribution, using some relaxation of selection cuts. For the reducible backgrounds, some of the VBF cuts and the $M_T(\text{lep}, \cancel{E}_T)$ cut may be removed or altered in order to regain the statistics. An order of magnitude increase in the statistics can be achieved while suppressing the contribution from the Z/γ^* decays. The shape of the Z/γ^* resonance peak is sensitive to some of the selection cuts; a relaxation of the central jet veto is considered a safe way to increase the statistics by a factor of two. The shape functions are then used to re-fit the total data distribution. Since the number of background events in the signal region will be estimated with real data from the fit of the $M_{\tau\tau}$ distribution, the fitting procedure is the only contribution to the uncertainty which needs to be considered for the significance estimate.

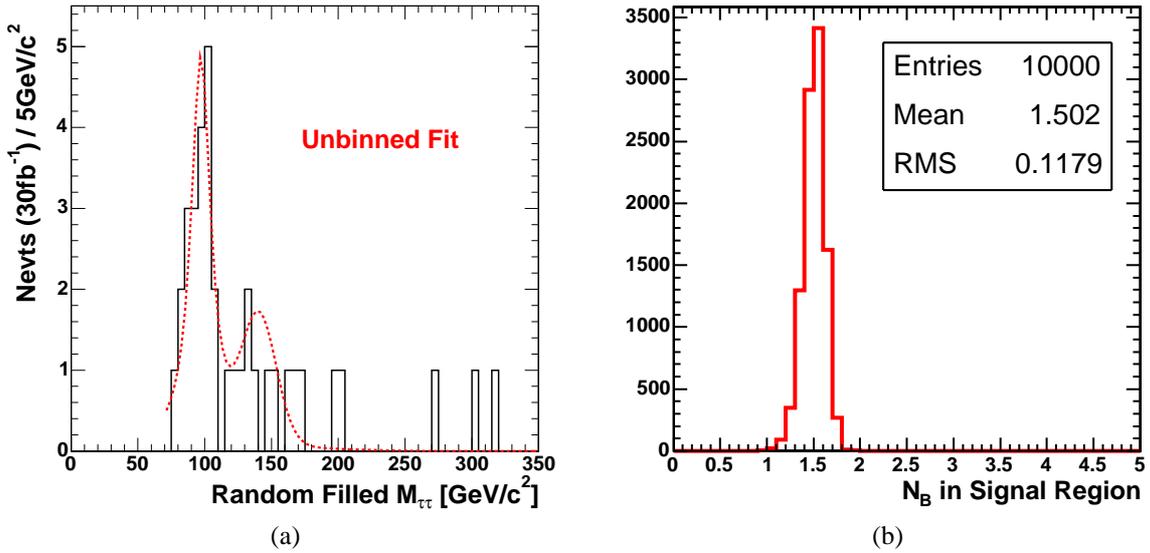


Figure 6: (a) An example of the MC distribution containing 30 fb^{-1} worth of events randomly selected from the sum of the functions used to fit the signal and background samples. The dotted line corresponds to the unbinned fit to the distribution using Equation 8. (b) The distribution of the number of background events estimated from 10k MC trials.

Systematic Uncertainty from Fitting Procedure

The uncertainty in the background events arising from the fitting procedure is evaluated by performing MC trials. At each trial, an “MC data distribution” is made with values taken randomly from the sum of the fit functions,

$$F_{\text{tot}}^0 = F_H + F_Z + F_R \quad (7)$$

The total number of entries is equal to the number of events expected at 30 fb^{-1} , N_{tot} . An example of the MC distribution is shown in Figure 6(a). The range used for the distribution is set between 70 - 350 GeV/c^2 ; the lower bound comes from the preselection of the Z background samples, and the upper bound is chosen below the point at which the polynomial approximation of the reducible background distribution fails.

The “MC data distribution” is fitted with a combination of the same three functions, where the shape parameters are fixed while each function is allowed to be scaled freely,

$$F_{\text{tot}}^{\text{fit}} = p_h \cdot F_H + p_z \cdot F_Z + p_r \cdot F_R \quad (8)$$

The three scale factors are restricted by N_{tot} , hence Equation 8 has only two free parameters. Since the statistics of events at each MC trial is limited, the MC distribution is treated as a data array rather than a histogram, and the unbinned fit method is used. The number of background events in the signal region is estimated from this total fit function, $F_{\text{tot}}^{\text{fit}}$. The uncertainty of the fit is calculated as the spread of the number of background events extracted from repeated trials.

The distribution of the number of background events expected in the signal region, N_B , over 10k trials is shown in Figure 6(b). The signal region is set to the mass window used for the signal sample with $M_H = 135 \text{ GeV}/c^2$, between 125 - 165 GeV/c^2 . The mean of the N_B distribution is under-estimated compared to the value obtained from the original background shape functions. This is due to the fact that the statistics is too low above $M_{\tau\tau} \sim 200 \text{ GeV}/c^2$, and the polynomial fit in this region is not optimally carried out, as shown by the fit curve in Figure 6(a). The uncertainty in N_B is 7.8%, which is calculated as the RMS divided by the mean of the N_B distribution. The uncertainty of 5.9% has been obtained for the case of 60 fb^{-1} of data, using the same method.

Higgs Boson Mass Estimate

The estimation of the background uncertainty from the fit has been evaluated assuming a fixed mass for the signal distribution, however, in reality the Higgs boson mass is an unknown parameter. The fit is repeated for different hypotheses of the Higgs mass, m_H , which is used as the mean of the gaussian function, F_H . The χ^2 of the fit is

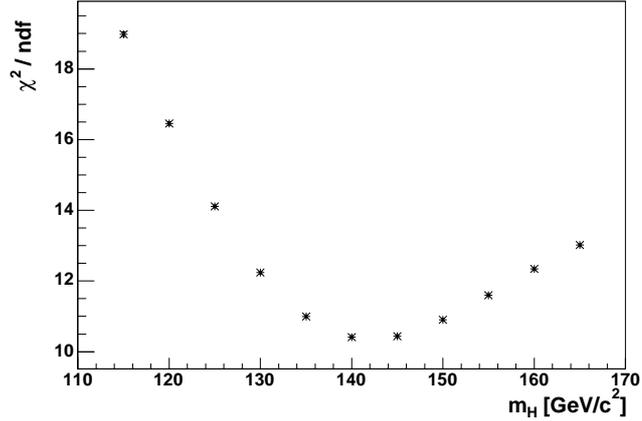


Figure 7: The χ^2/ndf for the fit to the MC distribution as a function of m_H , which is the mean of the gaussian function used to fit the Higgs signal distribution.

minimised for every fit. The mean of the $\chi^2_{\text{min}}/\text{ndf}$ distribution obtained from the fits to 1000 MC trial distributions for each of the different mass hypotheses is shown in Figure 7, where the value of m_H is varied in 5 GeV/c² steps. The $\chi^2_{\text{min}}/\text{ndf}$ distribution has a minimum at ~ 140 GeV/c², which is consistent with the value of Higgs boson mass used in the MC experiments.

6 Conclusions

The possibility of observing the Standard Model Higgs boson via the $qqH, H \rightarrow \tau\tau \rightarrow \text{lepton} + \text{jet}$ channel has been evaluated for four different masses of the Higgs boson in the low mass range ($M_H = 115, 125, 135, 145$ GeV/c²). Two types of background processes have been considered, (a) the irreducible background arising from QCD and EW Z boson production with 2 or 3 associated jets, (b) the reducible background comprising of W production with 3 or 4 associated jets and $t\bar{t}$ production. An effective rejection of these processes can be achieved by using the properties of the VBF process: the wide rapidity separation of the two leading quark jets and the suppressed hadronic activity in the central region due to the absence of colour exchange between the forward quark jets.

The efficiency of selecting the Higgs boson events is of the order of 10^{-3} , and the number of signal events expected at 30 fb^{-1} is ~ 10 . The statistical significance for discovering the Higgs boson is 3.9σ for a Higgs mass of $M_H = 135$ GeV/c². This result includes an experimental systematic uncertainty of 7.8% which arises from the fitting procedure.

7 Acknowledgments

The authors would like to thank the referees, L. Lista, J. Mans and R. Tenchini and the CMS Higgs Analysis group for their valuable discussions and comments.

References

- [1] J. D. Bjorken “*Rapidity Gap and Jets and a New-physics Signature in Very-high-energy Hadron-hadron Collisions*”, Phys. Rev. D47 (1993) 101
- [2] D. Rainwater, R. Szalapski, D. Zeppenfeld “*Probing Color-singlet Exchange in Z+2-jet Events at the CERN LHC*”, Phys. Rev. D54 (1996) 6680
- [3] D. Rainwater, D. Zeppenfeld, K. Hagiwara “*Searching for $H \rightarrow \tau\tau$ in weak boson fusion at the CERN LHC*”, Phys. Rev. D59 (1998) 014037
- [4] D. Rainwater “*Intermediate-Mass Higgs Searches in Weak Boson Fusion*”, hep-ph/9908378

- [5] Higgs Working Group “*The Higgs Working Group: Summary Report*”, hep-ph/0203056
- [6] S. Jadach, J. H. Kuhn, Z. Was “*TAUOLA - a library of Monte Carlo programs to simulate decays of polarized τ leptons*”, Comp. Phys. Comm. 64 (1991) 275
- [7] S. Jadach, Z. Was, R. Decker, J. H. Kuhn “*The τ decay library TAUOLA, version 2.4*”, Comp. Phys. Comm. 76 (1993) 361, Nucl.Phys.Proc.Suppl. 98 (2001) 96-102, hep-ph/0011305
- [8] H.L. Lai et al. “*Global QCD analysis and the CTEQ parton distributions*”, Phys. Rev. D51 (1995) 4763
- [9] M. Mangano “*Merging multijet matrix elements and shower evolution in hadronic collisions*”
<http://mlm.home.cern.ch/mlm/talks/lund-alpgen.pdf>
- [10] T. Sjostrand, et al., “*PYTHIA 6.3: Physics and manual*”, hep-ph/0308153
<http://www.thep.lu.se/torbjorn/Pythia.html>
- [11] M.L. Mangano et al. “*ALPGEN, a generator for hard multiparton processes in hadronic collisions*”, hep-ph/0206293
- [12] F. Maltoni, T. Stelzer “*MadEvent: Automatic event generation with MadGraph*”, hep-ph/0208156
- [13] S.R. Slabospitsky, L. Sonnenschein “*TopReX generator (version~3.25). Short manual*”, Comp. Phys. Comm. 148 (2002) 87
- [14] A.S. Beyaev et al. “*CompHEP-PYTHIA interface: integrated package for the collision events generatoin based on exact matrix elements*”, hep-ph/0101232
- [15] ORCA: CMS Reconstruction Package,
<http://cmsdoc.cern.ch/orca>
- [16] CMS Collaboration, “*CMS Physics Technical Design Report Volume I: Detector Performance and Software*”, CERN/LHCC 2006-001 CMS TDR 8.1
- [17] S. Gennai et al. “*Tau jet reconstruction and tagging at High Level Trigger and off-line*”, CMS NOTE 2006/028
- [18] CMS Collaboration, “*The TriDAS Project Volume II - Data Acquisition and High-Level Trigger, Technical Design Report*”, CERN/LHCC 2000-016 CMS TDR 6.2
- [19] A. Nikitenko, S. Kunori, R. Kinnunen “*Missing Transverse Energy Measurement with Jet Energy Corrections*”, CMS Note 2001/040
- [20] S. Bitukov, Program for calculations of the significance, S_cP
The software can be downloaded from, <http://cmsdoc.cern.ch/bitukov/>
S.Bitukov, N.Krasnikov, Mod.Phys.Lett.A 13 (1998) 3235