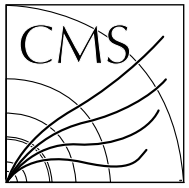
Available on CMS information server

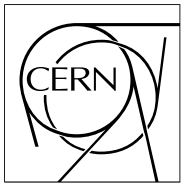
CMS NOTE 2006/092



The Compact Muon Solenoid Experiment

CMS Note

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



May 2006

Search for Standard Model Higgs Boson via Vector Boson Fusion in the $\mathrm{H} \rightarrow \mathrm{W}^{+}\mathrm{W}^{-} \rightarrow \ell^{\pm}\nu\mathrm{jj}$ with $120 < \mathrm{m_H} < 250\ \mathrm{GeV/c^2}$

Haifeng Pi

University of California, San Diego, CA, USA

University of Florida, Gainsville, FL, USA (before 2006)

Paul Avery

University of Florida, Gainsville, FL, USA

James Rohlf

Boston University, Boston, MA, USA

Christopher Tully

Princeton University, Princeton, NJ, USA

Shuichi Kunori

University of Maryland, College Park, MD, USA

Abstract

The feasibility of discovering the Standard Model Higgs boson via Vector Boson Fusion in $\mathrm{H} \rightarrow \mathrm{W}^{+}\mathrm{W}^{-} \rightarrow \ell\nu\mathrm{jj}$ channel is investigated. A comprehensive strategy of Higgs mass ($\mathrm{m_H}$) reconstruction is developed using detector jets, lepton and missing transverse energy ($\mathrm{E_T^{miss}}$). At an integrated luminosity of 30 $\mathrm{fb^{-1}}$, a 5σ discovery can be achieved for $140 < \mathrm{m_H} < 200\ \mathrm{GeV/c^2}$. Several important techniques including forward jet tagging, central jet selection, hadronic and leptonic W reconstruction, $\mathrm{E_T^{miss}}$ selection, lepton-W correlation and lepton isolation, are optimized to increase the signal to background ratio. Data driven analysis methods are developed to further identify the experimental signature of the signal in addition to the reconstructed Higgs mass peak and reduce the effect of a variety of detector systematic uncertainties.

1 Introduction

Vector Boson Fusion (VBF) is the second largest Higgs boson production process at the LHC for Higgs boson mass ($\mathrm{m_H}$) above 100 $\mathrm{GeV/c^2}$. In the VBF process, the Higgs boson is produced with two moderate $\mathrm{E_T}$ jets in forward regions. The central jet activities are suppressed because of color coherence between initial quarks. This is in contrast to most other physics processes involving multiple jet generation where the t-channel color flow between initial quarks enhances the central hadron production. Therefore, a selection strategy based on tagging two forward jets and vetoing central jets could dramatically suppress huge Standard Model (SM) backgrounds whose cross sections are several orders of magnitude higher than that of the Higgs boson signal.

For the SM Higgs boson decaying via $\mathrm{H \rightarrow W^+W^- \rightarrow \ell\nu\ell\nu}$ where it is produced via either gluon-gluon fusion or VBF, a large discovery potential [1, 2] is possible over a wide $\mathrm{m_H}$ mass range because the di-lepton signature can be observed over the SM background. However, the presence of two unobservable neutrinos in the final state prevents a direct measurement of the Higgs boson mass. A precise estimation of the background is extremely important to identify the lepton "excess" if it originated from the Higgs boson signal.

In the medium-high mass range ($\mathrm{m_H > 300\ GeV/c^2}$), the Higgs boson produced via VBF and decaying as $\mathrm{H \rightarrow W^+W^- \rightarrow \ell\nu jj}$, provides another potential route to discovery. The final state is characterized by two high $\mathrm{E_T}$ forward jets, two high $\mathrm{E_T}$ central jets from the W hadronic decay, and one high $\mathrm{p_T}$ lepton and large missing transverse energy ($\mathrm{E_T^{miss}}$) from the W leptonic decay. A high jet $\mathrm{E_T}$ threshold is feasible for both forward and central jets of the signal events, so as to reject SM background with much lower jet $\mathrm{E_T}$ spectra. This channel turns out to have the best discovery potential with $\mathrm{m_H > 600\ GeV/c^2}$, because of the increase of Higgs boson mass width as $\mathrm{m_H}$ goes up and too small cross section of $\mathrm{H \rightarrow ZZ \rightarrow 4\ell}$.

The possibility of extending the use of this channel to the low mass ($\mathrm{m_H < 300\ GeV/c^2}$) is intriguing. For example, in the range of $\mathrm{160 < m_H < 180\ GeV/c^2}$, the $\mathrm{H \rightarrow ZZ^*}$ branching ratio is highly suppressed due to $\mathrm{H \rightarrow W^+W^-}$ resonance. The direct Higgs boson mass reconstruction from $\mathrm{H \rightarrow W^+W^- \rightarrow \ell\nu jj}$ makes valuable physics analysis possible and is complementary to Higgs boson search via $\mathrm{H \rightarrow W^+W^- \rightarrow \ell\nu\ell\nu}$. The branching ratio to $\ell\nu$jj is $\sim$ 5.5 times larger than that of $\ell\nu\ell\nu$, provided only electron and muon are considered. The reconstruction based on the identification of hadronic and leptonic W provides extra capability to suppress the background in addition to forward jet tagging and central jet veto, since not all the background has two Ws intrinsically.

But these advantages come with a variety of physics analysis challenges that must be overcome:

- Many background processes of very large cross section have one lepton and multiple jets in the final states. Simulating the requisite huge number of background events is both a computing and analysis challenge.
- Deep selection cuts and heavy exploitation of physics signal characteristics are necessary to suppress background events and enhance the statistical significance of the signal, but they can lead to a large systematic uncertainty.
- A relatively low Higgs boson mass limits the application of high jet $\mathrm{E_T}$ threshold that would normally be used to suppress background events, in contrast to the situation at high mass.
- Low $\mathrm{E_T^{miss}}$ and low $\mathrm{E_T}$ jets lead to worsened Higgs mass reconstruction.

To meet these challenges, we propose a robust reconstruction and selection strategy for VBF Higgs in $\mathrm{H \rightarrow W^+W^- \rightarrow \ell\nu jj}$ that minimizes systematic uncertainties. The rest of the note is organized as following: the signal and background are discussed in Section 2. Section 3 contains details of the basic detector reconstruction algorithms for those fundamental objects (lepton, jet, and $\mathrm{E_T^{miss}}$). Section 4 describes the Higgs boson reconstruction strategy. In Section 5, a set of general selection cuts is introduced. In Section 6, the intermediate result of general selection cuts is summarized. In Section 7, the optimization of selection cuts and their efficiency are presented. In Section 8, the result of VBF Higgs discovery potential and mass distribution are summarized and discussed. In Section 9, experimental data analysis approaches are described. In Section 10, systematic uncertainties caused by the detector level reconstruction bias and generator level configuration are discussed respectively. The summary is in Section 11.

2 Signal and Background

2.1 Physics Channels

The overall signal process of this analysis is $\mathrm{qqH} \rightarrow \mathrm{qqW^+W^-} \rightarrow \mathrm{qq}\ell^{\pm}\nu\mathrm{jj}$ with Higgs mass range from 120 to 250 GeV/c^2. The cross section and branching ratio for the signal are included in Table 1. The experimental signature of the signal includes:

- One lepton. Only muon and electron are considered in the reconstruction.
- $\mathrm{E_T^{miss}}$ from the unobserved neutrino.
- At least four jets, including two forwards jets, two central jets, and possible extra jets from initial state radiation (ISR) and final state radiation (FSR).

Table 1: VBF Higgs branching ratio with W leptonic decay, cross section, and event simulated in $\ell\nu$jj final states

$\mathrm{m_H}$	BR ($\mathrm{H \rightarrow WW^{(*)}}$)	$\sigma(\ell\nu\mathrm{jj})$ (pb)	Events in 60 $\mathrm{fb^{-1}}$	Generated
120	0.122	0.1789	10734	465.8 %
130	0.279	0.3623	21738	230.0 %
140	0.480	0.5520	33120	150.9 %
150	0.685	0.7037	42222	118.4 %
160	0.918	0.8530	51180	97.70 %
170	0.967	0.8489	50934	98.17 %
180	0.929	0.7639	45834	109.1 %
190	0.778	0.5995	35970	139.0 %
200	0.735	0.5287	31704	157.7 %
210	0.727	0.4895	29370	170.2 %
220	0.719	0.4539	27234	183.6 %
250	0.700	0.3701	22206	225.2 %

Those physics channels that have similar final states in the detector level are considered as background processes:

- $\mathrm{t\bar{t}}$ + jets. The t quark almost exclusively decays to W + b quark. This process is the serious background for many new physics searches because of its large cross section and abundant leptons + multiple jets in the final state that fake the signature of the signal.
- W + jets. Leptons come from W semi-leptonic decay. Based on the signal event's signature, W + 4jets is the main background process. But in the parton level, W + Njets (N=1,2,3,4 ...) can all contribute to detector level lepton + 4jets due to ISR and FSR. In this study, we take parton level W + Njets (N=3,4) as main background processes that are generated by the calculation of tree level matrix elements.

  There is a potential over-estimation of the background due to the higher order correction of W + 3jets (based on ISR and FSR) partially overlapping with W + 4jets. This reflects the complicate situation concerning the event generation, in which the parton level events from a single physics process or several uncorrelated processes are generated in the leading-order (LO) first, and ISR and FSR are used to manifest the higher order correction. But if several physics processes are correlated (e.g. a higher order correction of one process is partially "overlapped" with another process in a similar final state), the proper estimation of the overall background with respect to the higher order prediction is subtle and difficult. A complete treatment of correlated background processes is largely beyond the scope of this analysis and under a separate study. But it should be emphasized that using both W + 3jets and W + 4jets backgrounds is more conservative than using W + 4jets only.
- Z(γ^*) + jets. Z leptonic decay leads to a di-lepton pair but experimentally contributes to one lepton signature (e.g. leptons in the forward region can't be identified, or low $\mathrm{p_T}$ leptons can't be well reconstructed). Especially electrons in the forward region are mis-identified as jets.

  In this analysis, Z + Njets (N=3,4) are considered as the main background processes. The cross section of Z + jets that gives the lepton + jets final state is about two orders of magnitude lower than that of W + jets, but its cross section can be measured precisely and used to interpret W + jets cross section.
- WW + jets. There are two main sources in WW production, electroweak (EW) and QCD. In the EW mode, WW bremsstrahlung comes from initial quark radiation without color flow. In the QCD mode, the W pair

comes from the continuum production with color flow between initial and final partons. Although the cross section of the EW process is much less, it has the very similar signature as that of the signal.

- ZZ + jets and ZW + jets. In these processes, one vector boson decays hadronically and another decays leptonically, thus faking the signature of the signal. The jet energy resolution is not good enough to reject hadronic Z events. Mainly QCD production are considered, because the EW production cross section is too small for these two processes.

- W + t + jets. This process is part of the inclusive W + jets, but t quark decaying to W + b quark can fake the signal signature. In this study, W + $t\bar{b}$ ($\bar{t}b$) + jets is considered as main background process with a significant overall cross section in lepton + Njets (N=2,3,4 ...) final states.

  W + $t\bar{b}(\bar{t}b)$ has the same final state as $t\bar{t}$. In the event generation of W + $t\bar{b}(\bar{t}b)$, the Feynman diagrams that contain $t\bar{t}$ are excluded. The gluon-gluon fusion dominates the cross section of W + $t\bar{b}(\bar{t}b)$, which is about 60 pb in the LO. The fusion of $u\bar{u}$ and $d\bar{d}$ for W + $t\bar{b}(\bar{t}b)$ is negligible. But the interference between W + $t\bar{b}(\bar{t}b)$ processes and $t\bar{t}$ production processes must be considered since they are not in the calculation of $t\bar{t}$ production, which is at the order of 10 pb for gluon-gluon fusion and each of quark-quark fusion. The overall cross section of W + $t\bar{b}(\bar{t}b)$ is estimated as $\sim$ 100 pb, which is still much smaller than that of $t\bar{t}$ + jets and W + jets.

- QCD multi-jets. QCD events with jet faked lepton or non-isolated lepton from heavy flavor decaying can be mis-identified as isolated lepton plus jets events. In Appendix II, we estimate the possible contamination of QCD background in the final result using a factorization model. It shows this background will not influence the discovery potential. But due to its very large cross section, QCD events should remain a serious concern.

Cross sections of above background processes are listed in Table 2. W + jets, Z + jets, W + $t\bar{b}(\bar{t}b)$ + jets and WW + 2jets (EW) have parton level pre-selection, which is explained in section 2.3.

Table 2: Cross section of major backgrounds and fraction of events generated with respect to an integrated luminosity of 60 fb^{-1}

Channels	σ (pb)	Events in 60 fb^{-1}	Fraction Generated
t $\bar{t}$ + jets	840	50.4 million	6.9%
W + t $\bar{b}$ ($\bar{t}b$)	100	6.0 million	57.6%
WW + jets (QCD)	73.1	4.39 million	3.95%
WW + 2jets (EW)	1.26	75600	113.0%
WZ + jets	27.2	1.63 million	15.0%
ZZ + jets	10.7	0.642 million	68.1%
W + 4 jets (W $\rightarrow e/\mu/\tau + \nu$)	677.4	40.7 million	1.95%
W + 3 jets (W $\rightarrow e/\mu/\tau + \nu$)	1689.7	101.3 million	1.04%
Z + 4 jets (Z $\rightarrow ee/\mu\mu$)	44.6	2.68 million	11.2%
Z + 3 jets (Z $\rightarrow ee/\mu\mu$)	112.1	6.73 million	8.91%

2.2 Overview of Background Cross Section Measurement

The cross section of most background processes will be measured in a good statistic precision at LHC, due to their high production rate. A better understanding of physics processes of backgrounds and their kinematic properties play a big role in the search of VBF Higgs boson. Many systematic uncertainties related to next-to-leading order (NLO) prediction and detector efficiency will be resolved experimentally. Measuring the cross section of those background processes is a non-trivial task. Two common issues need to be handled:

1. Multiple background processes have similar final states (e.g., $t\bar{t}$ + jets and W + jets contribute to lepton + jets signature with large cross section). In order to highly suppress certain background processes, some hard cuts are inevitable, which introduce systematic uncertainty in the reconstruction and selection. It is possible to measure the overall cross section of several background processes together and compare to the theoretical prediction. The feasibility needs be investigated.

2. The impact of minimum bias events on jet energy scale is very strong for the low E_T jet. Clearly identifying soft jets in the physics event and faked jets from various detector effects is a technical challenge.

   A reasonable jet E_T threshold is necessary to reduce those systematic effects, but this needs a careful treatment in the analysis. For example, experimental Z + one hard jet event might come from Z + 1jet events or Z

+ multiple jets events in which soft jets can be possibly excluded by the selection. The theoretical prediction for these two types of processes are fundamentally different, although they look the same "experimentally".

In the following, we provide an overview of major background measurement and possible issues in the reconstruction and selection:

- W/Z + jets

  These two processes have large cross section. With appropriate lepton isolation, the contamination of QCD and top background for W + 1jet is at least an order of magnitude lower. For Z + jets, QCD background is negligible. $\mathrm{t\bar{t}}$ + jets is a serious background for W/Z + multi-jets. In certain phase space of observables (e.g. leading di-jet mass), a large excessive W/Z+jets rate can be observed over other backgrounds, which provides good sensitivity for cross section measurement.

- WW/ZZ/ZW + jets

  For ZZ + jets and ZW + jets, the narrow Z mass peak can be reconstructed from Z semi-leptonic decay channels with little background. The associative jet rate in these channels can be reconstructed as well.

  For WW + jets, di-lepton from both W semi-leptonic decay provides a clean signature, but $\mathrm{t\bar{t}}$ with two Ws leptonically decaying is a serious background. Practically the measurement of WW cross section using jet veto can be used to estimate WW + jets cross section.

- $\mathrm{t\bar{t}}$ + jets and W + $\mathrm{t\bar{b}(\bar{t}b)}$ + jets

  In addition to the direct reconstruction of the top quark mass and using its selection efficiency to estimate the cross section $\mathrm{t\bar{t}}$ + jets, di-lepton + Njets (N > 1) provides the promising final state to measure these two background together, since W + $\mathrm{t\bar{b}(\bar{t}b)}$ is an irreducible background to $\mathrm{t\bar{t}}$.

  In the leading-order (LO), $\mathrm{t\bar{t}}$ + 1jet cross section is even bigger than $\mathrm{t\bar{t}}$. Due to a jet $\mathrm{E_T}$ threshold in the reconstruction that ignores low $\mathrm{E_T}$ jets, a careful study on the rate of di-lepton + 2jets, di-lepton + 3jets ... is important to measure and understand the cross section of $\mathrm{t\bar{t}}$ + Njets (N=0,1,2..).

An isolation strategy for the lepton reconstruction is necessary to identify the W or Z leptonic decay and achieve a significant suppression factor on the lepton from heavy flavor decay. These backgrounds can be generally characterized as lepton + associated jet(s). For each process, its background contamination after appropriate selection cuts can be controlled to be at least one or two orders of magnitude lower, so that the statistical uncertainty of measured cross section is several percent or less after the background subtraction. The accuracy can be further improved by cross-checking correlated channels (e.g., Z + jets and W + jets), and fitting the kinematic distribution based on combining multiple background processes due to a large number of events can be taken experimentally.

2.3 Event Generation

In addition to cross section measurement, a tuning of event generation based on the experimental data is a key step to improve the accuracy of theoretical prediction on the kinematic properties of the signal and background events. A lot of important new physics largely relies on a reliable estimation of standard model background and looking for excess of selected objects. The fine tuning of theoretical model and event generator, which involves the calculation of higher order corrections and adjusting the event generation that commonly starts from leading order, is very important. Currently the event generation contributes non-trivial systematic uncertainties. A k-factor (ratio of the cross section of NLO to LO) ranging from 1.0 to 1.2 is expected for those backgrounds that relate to VBF Higgs, which gives a rough quantitative estimation of the uncertainty of event generation at 10-20 %.

The event generation for this analysis is summarized as follows:

- The Higgs boson signal, $\mathrm{t\bar{t}}$ + jets, WW + jets (QCD), WZ + jets, and ZZ + jets were generated with PYTHIA [3] which is implemented in CMKIN [10]. All decay mode of W and Z boson in background were switched on except the signal events with only semi-leptonic mode switched on. The number of events for each process is listed in Table 1 and 2. The configuration of generator includes: ISR, FSR, hadronization, multiple parton interaction and underlying event. CTEQ5M Parton Distribution Function (PDF) set was chosen.

- W + 3jets, W + 4jets, Z + 3jets, and Z + 4jets were generated with ALPGEN [4]. Due to very large cross section of those processes, the parton level pre-selection cuts are implemented based on jet $\mathrm{p_T}$ threshold ($\mathrm{p_T^j}$), jet η range (η^{j}) and minimum jet-jet distance ($\Delta\mathrm{R_{jj}}$). Renormalization and factorization scales were set to $\mu_0 = \mathrm{m_W}$ and CTEQ5L PDF set was chosen. WW + 2 jets (EW) was generated by MADGRAPH [5] with same parton level pre-selection cuts and configuration as ALPGEN. W + $\mathrm{t\bar{b}(\bar{t}b)}$ + jets was generated by COMPHEP [6], no selection cut is applied to t quark, minimum b quark $\mathrm{p_T}$ is set to 15 GeV/c with $|\eta| < 5$.

Events generated by ALPGEN, MADGRAPH, and COMPHEP (called Matrix Element Event Generator, or ME generator) were then processed by PYTHIA for parton showering with the same settings described for "inclusive" event generation except the PDF set was changed to be compatible with the ME generator. A matching technique between the LO matrix element and parton shower can be used to better simulate the hadronic final state for NLO prediction and avoid the double counting between various parton level final states [7]. This technique is not yet used in the current study. The reduction of background due to the matching technique will increase the significance of the signal. The configuration of ALPGEN and MADGRAPH is illustrated in Table 3.

Table 3: The configuration of parton level pre-selection of matrix element event generator (ALPGEN and MADGRAPH)

Generator	Channel	$\mathrm{p_T^j}$ (GeV/c)	η^{j}	$\Delta\mathrm{R_{jj}}$
ALPGEN	Z + 3jets	25	5.0	0.5
ALPGEN	Z + 4jets	25	5.0	0.5
ALPGEN	W + 3jets	25	5.0	0.5
ALPGEN	W + 4jets	25	5.0	0.5
MADGRAPH	WW + 2 jets	25	5.0	0.5

Effects of NLO correction are not generally considered in this analysis because NLO calculations of some backgrounds are not available, e.g., W + Njets (N $\geq$ 3), Z + Njet (N $\geq$ 3), WW/ZZ/WZ + Njets (N $\geq$ 2). The NLO cross section for single W, Z or WW/ZZ/WZ production is available, but they largely overlap with leading order calculation of vector boson + associated jets. For $\mathrm{t\bar{t}}$ process, the NLO is included based on widely used value [8]. The k-factor for VBF Higgs is $\sim$ 1.1 [9], which is generally smaller than that of gluon-gluon fusion Higgs ($\sim$ 1.5-1.8) [9] and some backgrounds' k factor. Due to relatively small NLO correction of signal events, systematic uncertainties in this analysis with respect to NLO mainly come from the backgrounds of W + Njets (N $\geq$ 3) and $\mathrm{t\bar{t}}$ + jets.

3 Detector Simulation and Reconstruction

The full CMS detector simulation based on OSCAR [11] is performed for the signal and background processes including $\mathrm{t\bar{t}}$ + jets, WW + jets (QCD), WZ + jets and ZZ + jets. Fast CMS detector simulation based on FAMOS [13] is performed for background processes including: W + Njets (N = 3,4), Z + Njets (N = 3,4), WW + 2 jets (EW), W + $\mathrm{t\bar{b}}$ (tb) + jets. The pile-up condition is set for low luminosity of LHC ($\mathrm{L} = 2 \times 10^{33}\mathrm{cm}^{-2}\mathrm{s}^{-1}$). The digitization and reconstruction are based on standard CMS software ORCA [12] and FAMOS.

Jets are reconstructed with an iterative cone algorithm with cone size of $\Delta\mathrm{R} = 0.6$ [14]. No off-line threshold on tower constituent is used. The jet energy correction is applied according to the jet energy response based on QCD jets.

$\mathrm{E_T^{miss}}$ is reconstructed from all the calorimeter towers with muon momentum correction applied if muon(s) is present in the event [15]. Jet energy correction for $\mathrm{E_T^{miss}}$ is tested. Because of we considering the low mass Higgs boson in this study, the corrected $\mathrm{E_T^{miss}}$ scale is largely influenced by low $\mathrm{E_T}$ central jets, which causes more bias than that of other background processes (e.g., $\mathrm{t\bar{t}}$ + jets) that have harder jet $\mathrm{E_T}$ spectrum and potentially benefit more from the jet energy correction.

Electrons and muons are reconstructed using standard off-line algorithms [16, 17]. Because of the presence of multiple jets in signal and background final states, a strong calorimeter based isolation is used to identify the leptons from W or Z decay.

The isolation criteria for reconstructed off-line electron object includes:

- $\mathrm{E_T^{Hcal}/E_T^{Ecal}} < 0.05$, where the $\mathrm{E_T^{Hcal}}$ and $\mathrm{E_T^{Ecal}}$ are calculated from the 0.2 isolation cone around electron super-cluster in Hadronic Calorimeter (HCAL) and Electromagnetic Calorimeter (ECAL) respectively.

- $0.9 < \mathrm{E/p} < 1.8$, where E and p are the energy of electron super-cluster measured in ECAL and track momentum measured in Tracker respectively.
- $|\mathrm{E_T^{0.2}} - \mathrm{E_T^e}| < 5.0$ GeV and $|(\mathrm{E_T^{0.2}} - \mathrm{E_T^e})/\mathrm{E_T^e}| < 0.3$, where $\mathrm{E_T^{0.2}}$ is the total $\mathrm{E_T}$ in the 0.2 isolation cone and $\mathrm{E_T^e}$ is the electron super-cluster $\mathrm{E_T}$.
- $\mathrm{E_T^{0.4}/E_T^e} < 0.3$, where $\mathrm{E_T^{0.4}}$ is the sum of $\mathrm{E_T}$ in the 0.2-0.4 isolation cone.

A similar isolation criterion was applied to reconstructed off-line muon object:

- $|\mathrm{E_T^{0.2}} - \mathrm{p_T^\mu}| < 9.0$ GeV and $\mathrm{E_T^{0.2}/p_T^\mu} < 0.3$, where $\mathrm{E_T^{0.2}}$ is the total $\mathrm{E_T}$ in the 0.2 isolation cone and $\mathrm{p_T^\mu}$ is the muon transverse momentum measured in Tracker.
- $\mathrm{E_T^{0.4}/p_T^\mu} < 0.3$, where $\mathrm{E_T^{0.4}}$ is the sum of $\mathrm{E_T}$ in the 0.2-0.4 isolation cone.

Detailed kinematic distributions concerning above selection cuts for reconstructed electron and muon are summarized in Appendix I. The overall off-line electron and muon reconstruction efficiencies for VBF Higgs sample are shown in Fig. 1. It can be seen that intensive jet activities cause overlapping between jets and leptons that results in the loss of lepton efficiency, which is lower than the efficiency of benchmark physics channels (e.g., leptonic decays of W or Z events) in which jet activities are negligible. $\mathrm{t\bar{t}}$ + jets events were used to check the performance of the isolation. The purity of the isolated lepton with $\mathrm{p_T} > 30$ GeV/c is 99.73% and 99.88% for electron and muon respectively.

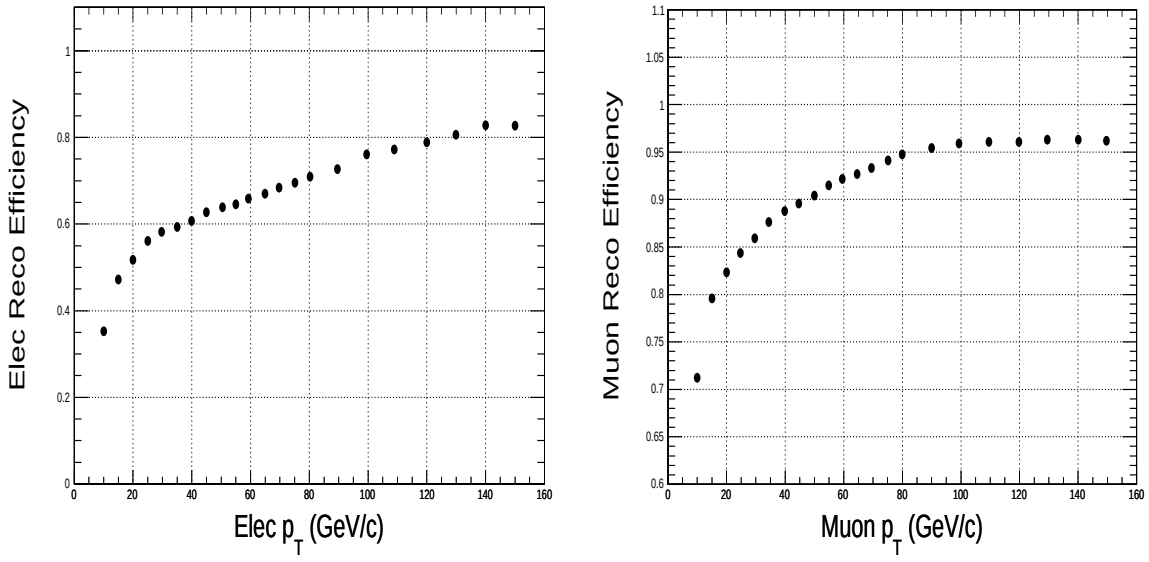


Figure 1: The overall reconstruction and selection efficiency of electron (left) and muon (right) in VBF Higgs events

It might be argued that the lepton isolation efficiency can be further optimized by using track based isolation methods (e.g., counting number of tracks around the lepton track in an isolation cone, setting a threshold for the sum of transverse momentum from tracks nearby, and using vertex information to further suppress leptons from b-quark decays), but this approach is limited in predicting the neutral energy of jets. In general, the optimization of the isolation efficiency will not increase the signal to background ratio, because we mainly quantitatively evaluate those backgrounds that have W or Z boson leptonic decay with isolated lepton intrinsically. To minimize the influence of QCD background, a tight calorimeter-based isolation criterion is necessary to veto those jet-faked or jet-induced leptons. So the track-based isolation criterion does not necessarily lead to a looser calorimeter-based isolation criterion because of the presence of abundant jets in the final state.

4 Higgs Boson Reconstruction and Selection Strategy

The VBF Higgs boson production mechanism and its signature drives our reconstruction algorithm emphasizing lepton selection, identification and tagging of two forward jets, hadronic W reconstruction using two central jets

and leptonic W reconstruction using $\mathrm{E_T^{miss}}$ and the isolated lepton. From the point view of results and their sensitivity to analysis techniques, the reconstruction of those basic objects (forward jets, central jets, lepton, $\mathrm{E_T^{miss}}$) can be classified in two weakly coupled groups: lepton + $\mathrm{E_T^{miss}}$ system and jet system, which mainly influence the reconstruction of leptonic and hadronic W respectively.

4.1 Offline Lepton Selection Strategy

Only one isolated high $\mathrm{p_T}$ lepton from W decay is in the final state of the signal events. Some consideration is needed for the events that have one high $\mathrm{p_T}$ lepton and one or more low $\mathrm{p_T}$ leptons.

- If the extra lepton comes from heavy flavor decay, it is largely within the consideration of one lepton + Njets final states for both signal and background.

  Technically it is difficult to apply the isolation criterion to the low $\mathrm{p_T}$ lepton due to the significant reconstructed calorimeter energy from jet activities, pileup and underlying events, in which the performance of isolation is more sensitive to those factors and has large systematic uncertainty.

- If the extra lower $\mathrm{p_T}$ lepton comes from W or Z decays that is not the same vector boson that gives the highest $\mathrm{p_T}$ lepton, its effect is negligible since the cross section of producing two vector bosons is several orders of magnitude lower than that of one vector boson. Furthermore if both vector boson decay leptonically, the probability for passing the multiple jet selection criteria is much lower than events that have one vector boson decaying into two jets.

- If extra lepton comes from Z decay, it is negligible because the overall Z leptonic decay rate from Z + jets is at least two orders of magnitude lower than that of W + jets. The lepton $\mathrm{p_T}$ spectrum for Z + jets is shown in Fig. 2.

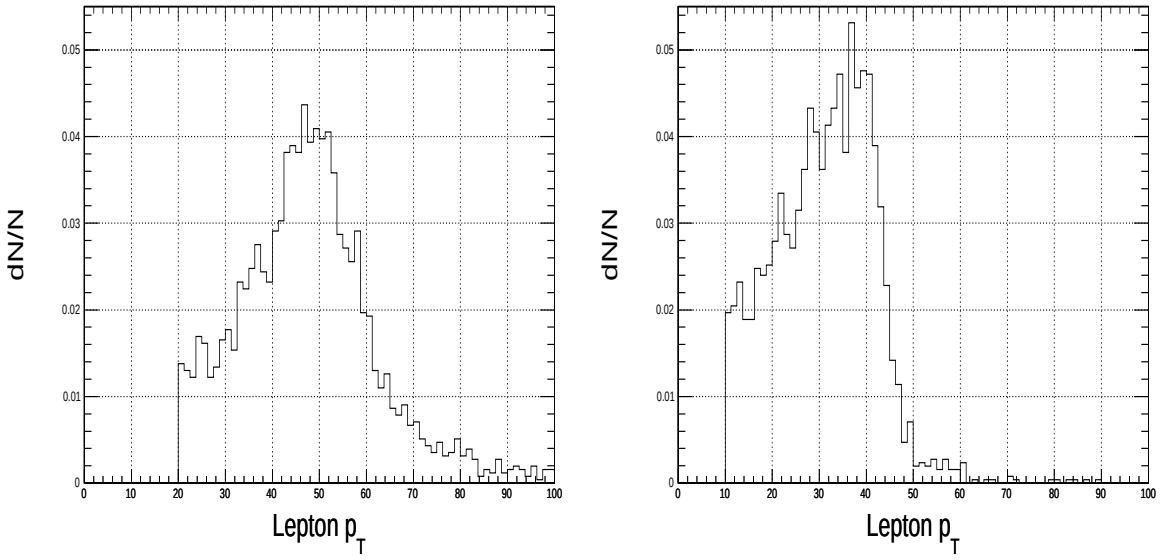


Figure 2: Lepton $\mathrm{p_T}$ spectrum for the highest $\mathrm{p_T}$ lepton (left) and the second highest $\mathrm{p_T}$ lepton (right) in the Z+jets sample with Z leptonic decay

As discussed above, the presence of one or more low $\mathrm{p_T}$ lepton in addition to an isolated high $\mathrm{p_T}$ doesn't jeopardize the W selection. We adopt a robust lepton selection strategy in this analysis, which is less influenced from various physics and detector systematic effects:

- After the lepton trigger selection, the calorimeter based isolation is applied off-line to the lepton with $\mathrm{p_T} >$ 10 GeV/c.

- The lepton with $\mathrm{p_T} <$ 10 GeV/c isn't counted.

- The lepton with $\mathrm{p_T} > 10$ GeV/c but fails isolation selection isn't counted.
- The veto of the event is based on whether there is only one isolated lepton.
- The isolated lepton $\mathrm{p_T}$ is required to be above 30 GeV/c, of which the threshold is optimized separately.

4.2 Properties of Multiple Jet System

The understanding of parton behavior and its corresponding jet activities in the VBF Higgs boson signal events are very important to look for appropriate selection strategies for tagging forward jets and hadronic W reconstruction. In order to analyze the properties of jet system, the detector jet is matched the parton with angular distance ($\Delta \mathrm{R} = \sqrt{\Delta\eta^2 + \Delta\phi^2}$) less than 0.3. The matched jets are called quark-jet. The reconstructed quark-jet efficiency as a function of jet $\mathrm{E_T}$ threshold is shown in Fig. 3.

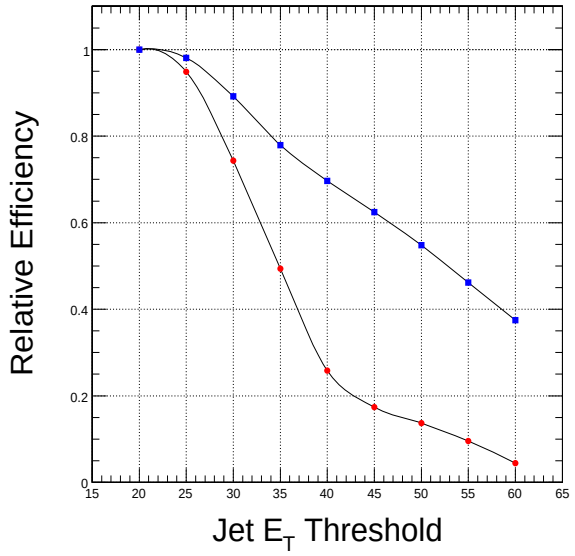


Figure 3: Quark-jet relative matching efficiency as a function of jet $\mathrm{E_T}$ threshold for valence quark (square) and quark from W hadronic decay (circle) in VBF Higgs events with $\mathrm{m_H} = 170\ \mathrm{GeV/c^2}$. The efficiency is normalized to 1.0 for jet $\mathrm{E_T}$ threshold of 20 GeV.

4.2.1 Forward Jet Tagging

In this section, detector jets are matched with two valence quarks using the signal events, the major target of forward jet tagging. The η distance ($\Delta\eta = |\eta_1 - \eta_2|$) and the di-jet mass ($\mathrm{m_{qq}}$) of two jets that correspond to two valence quarks in the signal events ($\mathrm{m_H} = 170\ \mathrm{GeV/c^2}$) are shown in Fig. 4. There is a peak around 5.0 in $\Delta\eta$ distribution, so the plausible range for minimum $\Delta\eta$ cut is below 5.0, otherwise the signal efficiency will decrease dramatically. The minimum $\mathrm{m_{qq}}$ can be set above 1000 $\mathrm{GeV/c^2}$, which need be optimized with the signal and background efficiency.

Extra detector jet from ISR and FSR or detector effects that has higher $|\eta|$ might cause mis-identification in the forward jet tagging. For example, in those signal events that two valence quarks don't have wide enough η distance, extra jets can significantly enhance the chance of those events to pass the forward jet tagging, but this effect is largely reduced by a higher jet $\mathrm{E_T}$ threshold as shown in Fig. 5. Although this effect doesn't influence the forward jet tagging efficiency, it increases the chance of mis-identification of central jets in hadronic W reconstruction.

A high jet $\mathrm{E_T}$ threshold can be used to remove those extra jets in forward regions as shown in Fig. 6. For a jet $\mathrm{E_T}$ threshold below 35 GeV, there is a much stronger dependency of forward jet tagging efficiency on the jet $\mathrm{E_T}$ threshold, which can be explained by the intensive soft jet activities. Due to this fact, the systematic uncertainty of jet energy scale will be significantly enhanced in forward jet tagging for $\mathrm{E_T}$ threshold below 35 GeV, which should be considered in the optimization of the selection cuts.

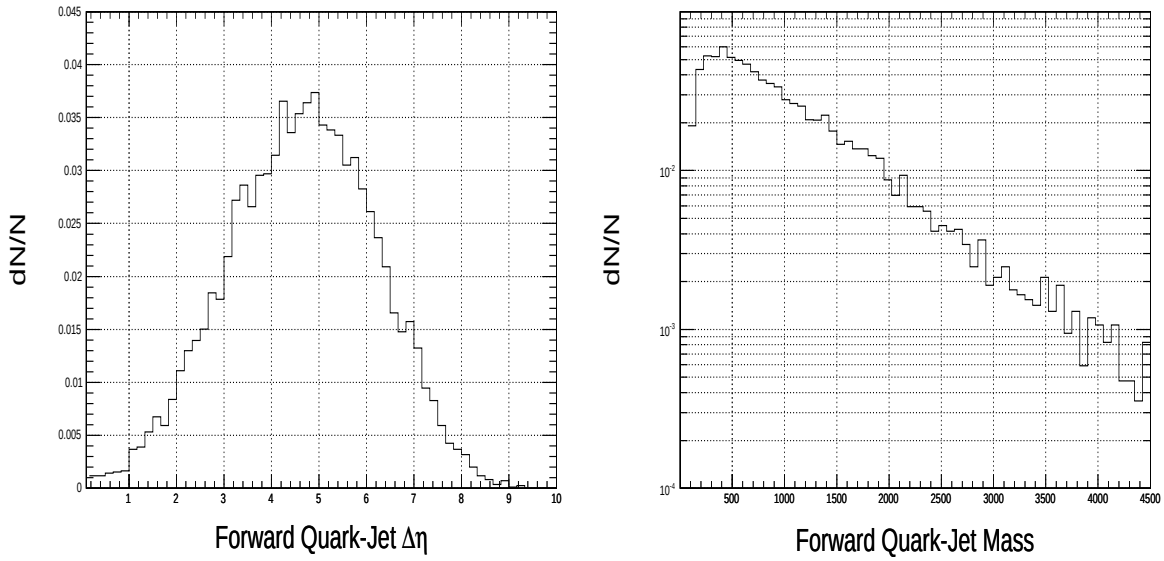


Figure 4: Two forward quark-jet properties: $\Delta\eta$ distribution (left) and $\mathrm{m_{qq}}$ distribution (right)

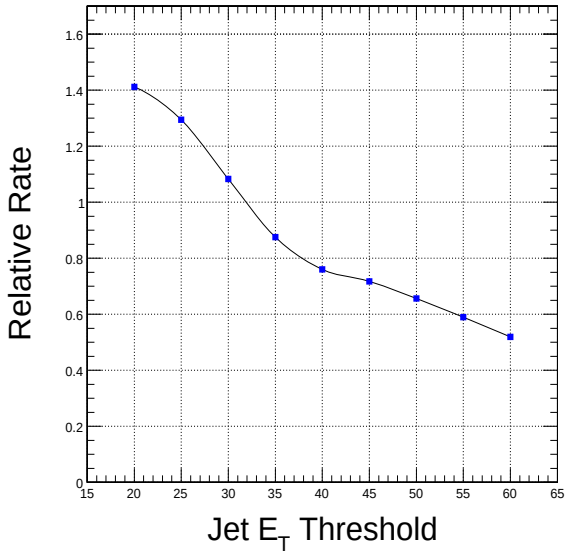


Figure 5: The relative rate of VBF Higgs events ($\mathrm{m_H} = 170\ \mathrm{GeV/c^2}$) that pass forward jet tagging by extra jets (but quark-jet fail) to those events that quark-jet passes tagging as a function of jet $\mathrm{E_T}$ threshold. Intensive ISR and FSR largely enhanced the forward jet tagging efficiency, especially for the $\mathrm{E_T}$ threshold below 35 GeV

The increase of jet $\mathrm{E_T}$ threshold and η distance threshold causes the reduction of tagging efficiency as shown in Fig. 3. With a fixed quark-jet $\mathrm{E_T}$ threshold, the increase of η distance will reduce the mis-identification rate, but it also results in the reduction of overall tagging efficiency, as shown in Fig. 7.

4.2.2 Hadronic W Reconstruction

In the discussion of this section, jets matching with two quarks from W hadronic decay are used to study the detector selection strategy. The detector W mass reconstructed from quark-jet is shown in Fig. 8, which provides a basic estimation of hadronic W mass resolution of $\sim 14.8\ \mathrm{GeV/c^2}$. The reconstruction efficiency is very sensitive to jet $\mathrm{E_T}$ threshold as shown in Fig. 3. A threshold higher than 30 GeV will have serious impact on the signal selection efficiency.

The di-jet mass scale and resolution are sensitive to the jet cone size. The result of average reconstructed W mass

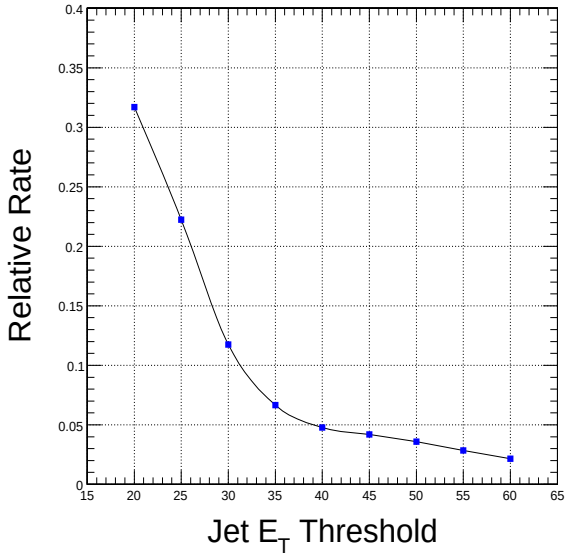


Figure 6: The rate of VBF Higgs events with $\mathrm{m_H} = 170\ \mathrm{GeV/c^2}$ that contain extra jet(s) outside the range of two jets matched with the valence quark with η distance bigger than 3.8 as a function of jet $\mathrm{E_T}$ threshold. The rate increases significantly as jet $\mathrm{E_T}$ threshold goes below 35 GeV, which indicates a strong enhancement of the soft jet activities via ISR/FSR and detector effects.

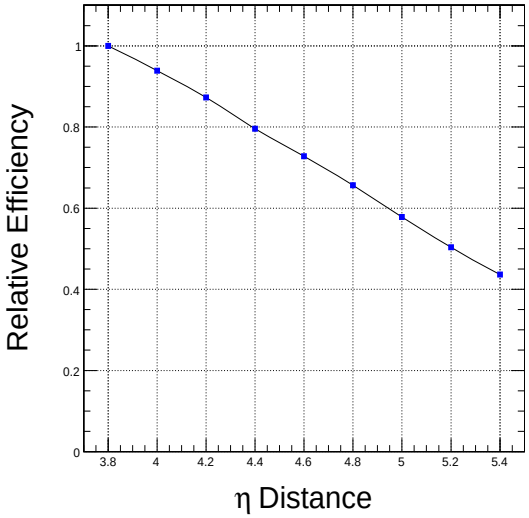


Figure 7: Forward Jet Tagging efficiency for different threshold of η distance in VBF Higgs events with $\mathrm{m_H} = 170$ $\mathrm{GeV/c^2}$

($< \mathrm{m_W} >$) and W mass resolution ($\sigma(\mathrm{m_W})$) for three cone sizes (0.4, 0.6, and 0.8) is summarized in Table 4. It shows 0.6 cone jet provides a better W mass scale and resolution, which allows a symmetric di-jet mass selection window with respect to the true W mass.

Table 4: Reconstructed W mass resolution with various jet cone. Real W mass ($81.2\ \mathrm{GeV/c^2}$) is used to scale the reconstructed W mass ($\mathrm{m_W}$), which leads to a scaled $\sigma_\mathrm{s}(\mathrm{m_W}) = \mathrm{m_W}/81.2 \cdot \sigma(\mathrm{m_W})$

Cone Size	$< \mathrm{m_W} >\ \mathrm{GeV/c^2}$	Detector $\sigma(\mathrm{m_W})\ \mathrm{GeV/c^2}$	Scaled $\sigma_\mathrm{s}(\mathrm{m_W})\mathrm{GeV/c^2}$
0.4	55.1	11.52	16.8
0.6	82.3	14.75	14.4
0.8	90.27	17.25	15.4

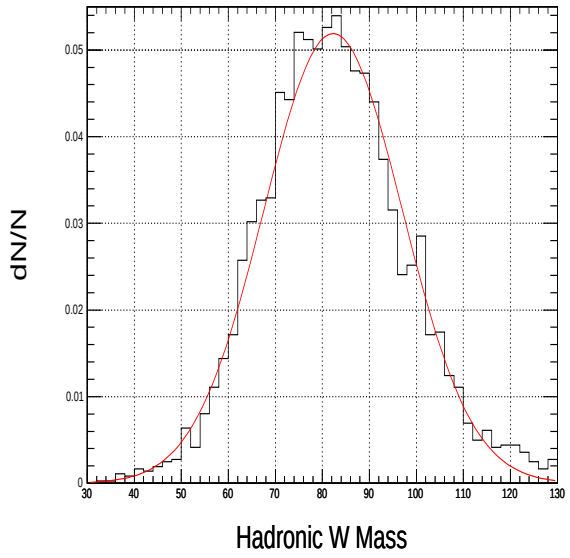


Figure 8: Reconstructed $\mathrm{m_W}$ using quark-jet that two quarks are identified from hadronic W decay in VBF Higgs events with $\mathrm{m_H} = 170\ \mathrm{GeV/c^2}$. The distribution is fitted by a Gaussian with σ = $\sim$ 14.8 GeV/$\mathrm{c^2}$.

Multiple central jets cause a combinatorial problem. In the following, an overview of several possible selection strategies are provided:

- The pair of jets with the least error to the true W mass is used for hadronic W reconstruction. The selection criterion about extra jets in the central region can be optimized with respect to the signal and background.
- The second approach is to require exact two jets in the central region and veto those events that have extra jets, so there is no ambiguity in combining two jets for W reconstruction. This method results in a large reduction of signal and background selection efficiency and makes further optimization hard to proceed because of low statistics of our data samples. Fig. 9 shows about 60% of the VBF Higgs events have extra jets with $\mathrm{E_T} > 20$ GeV in detector level.

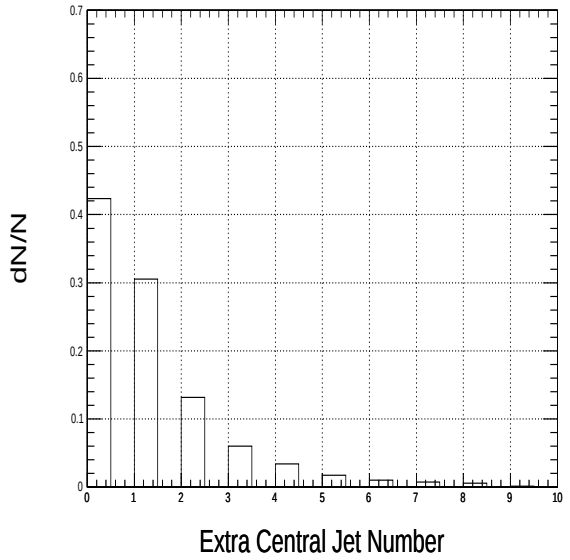


Figure 9: Number of extra jets in the central region excluding the quark-jet from forward jet tagging and hadronic W reconstruction in VBF Higgs events with $\mathrm{m_H} = 170\ \mathrm{GeV/c^2}$. A jet $\mathrm{E_T}$ threshold of 20 GeV is used.

- A third possible approach is to look for two highest $\mathrm{E_T}$ jets in the central region to reconstruct the hadronic

W. Due to the low W $\mathrm{p_T}$ of the signal, the hadronic W decay don't necessarily provide the highest two $\mathrm{E_T}$ jets as shown in Fig. 10. So this strategy will cause mis-identification.

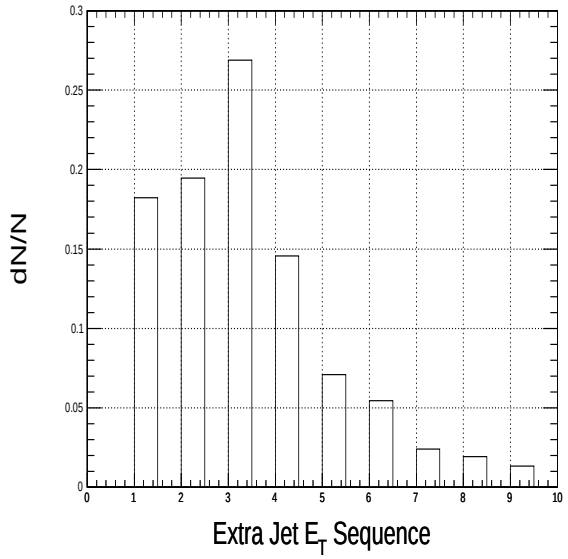


Figure 10: The ID of extra jet, which is numbered based on jet $\mathrm{E_T}$ from highest to lowest in VBF Higgs events with $\mathrm{m_H} = 170\ \mathrm{GeV/c^2}$. The quark-jet from forward jet tagging are excluded. If two highest $\mathrm{E_T}$ central jets are required for W reconstruction, the mis-identification rate is high, because extra jets are $\sim 17\%$ (19 %) of the highest (second highest) $\mathrm{E_T}$ jets in the central region.

The optimized central jet selection strategy used in this study combines the first and second methods into a modified central jet veto scheme by looking for a di-jet mass with least error to the true W mass and controlling the maximum number of central jets in an event, so that the combinatorial effect is reduced and physics nature of the real jet from W decay can be manifested.

4.2.3 Jets $\mathrm{E_T}$ System

A proper $\mathrm{E_T}$ threshold is important in tagging forward jet, reconstructing the hadronic W and optimizing the selection criterion for extra jet. In the reconstruction, extra jets are those detector jets which are not used for forward jet tagging and hadronic W reconstruction. The primary interest in extra jets relates to the jets that are within the η range of two tagged forward jets, so the extra jets are counted only in this η range.

In the reconstruction of forward jet and hadronic W, we avoid to use highest jet $\mathrm{E_T}$ selection criterion (e.g., using two highest $\mathrm{E_T}$ jets for the forward jet tagging and/or the rest two highest $\mathrm{E_T}$ jets for the hadronic W), which largely reduce the systematic effect of jet energy response and calibration bias between different η region of the calorimeter. For example, jet energy response is quite different between the central and forward region. The jet energy scale is sensitive to the jet $\mathrm{E_T}$ spectrum, which inevitably causes systematic bias. The approach based on highest $\mathrm{E_T}$ selection also shows significant mis-identification rate (as previously shown in Fig. 10) and loss of the true efficiency.

For the forward jet tagging, a robust strategy is used for this analysis that is based on the threshold of jet $\mathrm{E_T}$, di-jet $\Delta\eta$ and di-jet mass. The jet η can be measured in good precision due to the fine granularity of CMS HCAL. A similar strategy is used for hadronic W as discussed in previous section.

The major constraint on optimizing jet $\mathrm{E_T}$ threshold relates to a large number of low $\mathrm{E_T}$ jets coming from various detector effects. A 25 GeV threshold on jet $\mathrm{E_T}$ is choosen, so that those jets below this threshold will not be counted, which largely prevents the analysis and results from various detector effects and systematic uncertainties. Although it is anticipated that the average detector jets $\mathrm{E_T}$ will be lower than the quark because of ISR and FSR, we need to keep a reasonable $\mathrm{E_T}$ threshold to make the result less affected by generator level pre-selection cuts.

Fig. 11 shows the multiple jet selection efficiency (requiring at least 4 jets in an event) for various samples as a function of jet $\mathrm{E_T}$ threshold. The curve of W + 3jets is more sensitive to the threshold (as the threshold goes

down, the passing rate of W + jets increases significantly), since soft jets from ISR and FSR plays a stronger role in making W + 3jets pass the 4 jet selection criterion than other samples. For the threshold around 25 GeV, the efficiency curves of various samples have almost the same slope, which indicates the ratio of signal to background will be less affected by the systematic effects of jet energy scale and intrinsic features of various physics processes.

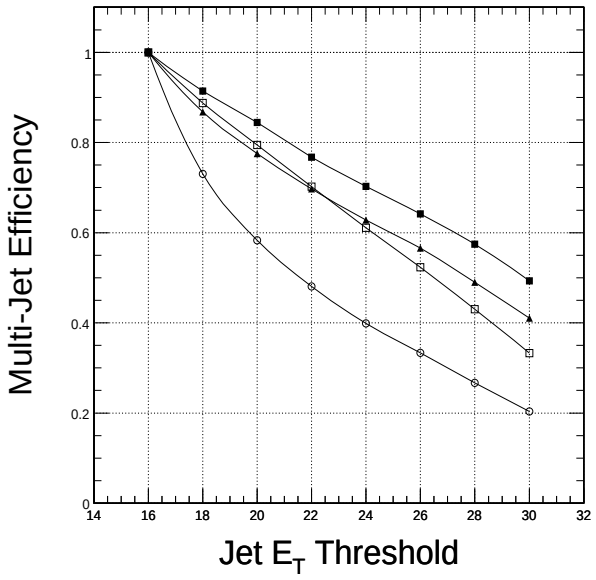


Figure 11: Multiple jet selection efficiency (requiring at least 4 jets in an event) as a function of jet $\mathrm{E_T}$ threshold. The efficiency is normalized to the rate with jet $\mathrm{E_T}$ threshold of 16 GeV for each sample. The physics channels include: $\mathrm{t\bar{t}}$ + jets (solid square), W + 3jets (open circle), W + 4 jets (solid triangle), and VBF Higgs with $\mathrm{m_H}$ = 170 GeV/$\mathrm{c^2}$ (open square)

5 Basic Event Selection

In this section, a number of kinematic cuts are applied to make a basic filtering of the events before the optimized selection. The selection strategy attempts to properly manifest the true nature of VBF Higgs events and minimize the systematic effect in the reconstruction and selection. The selection cuts are introduced according to their sequence in the reconstruction chain.

The analysis mainly assumes both W from the Higgs boson decay are on-shell, especially for $160 < \mathrm{m_H} < 180$ $\mathrm{GeV/c^2}$. There is small difference in the cuts for $\mathrm{m_H} < 160$ $\mathrm{GeV/c^2}$, which is described in the end.

5.1 Level-1 and High-Level Trigger for Electron or Muon (Trigger)

The lepton trigger is the only effective trigger for low mass VBF Higgs. The single electron and muon trigger in CMS Level-1 and high level trigger (HLT) streams provide the first filtering of the events [18]. This step is performed with standard algorithm and criterion. The HLT threshold for single electron (muon) at low luminosity is 29 (19) GeV/c. Due to staged muon detector, the muon trigger is limited to $|\eta| < 2.1$, but off-line muon can be reconstructed up to $|\eta| = 2.4$.

5.2 Lepton Selection (L-S)

One isolated lepton in the central detector region is the most important object to suppress hadronic events and "leptonic" events with lepton from heavy flavor decay. In addition to the lepton isolation criterion in the off-line reconstruction described in previous section, a lepton-jet isolation is used: $\Delta\mathrm{R}_{\ell-\mathrm{j}} > 0.5$, where $\Delta\mathrm{R}_{\ell-\mathrm{j}}$ is the distance in η-ϕ space between the reconstructed lepton and the nearest jet with $\mathrm{E_T} > 25$ GeV.

Lepton $\mathrm{p_T}$ selection is defined by a selection window between 30 and 120 GeV/c, because VBF Higgs boson events have a relatively higher fraction of lepton rate in this range as illustrated in Fig. 12. Especially in low $\mathrm{p_T}$ region, background processes have much higher lepton rate, while the signal lepton rate falls quickly. This fact

explains that a lower p_T threshold in lepton threshold will not improve VBF Higgs discovery potential in the $\ell\nu$jj final state.

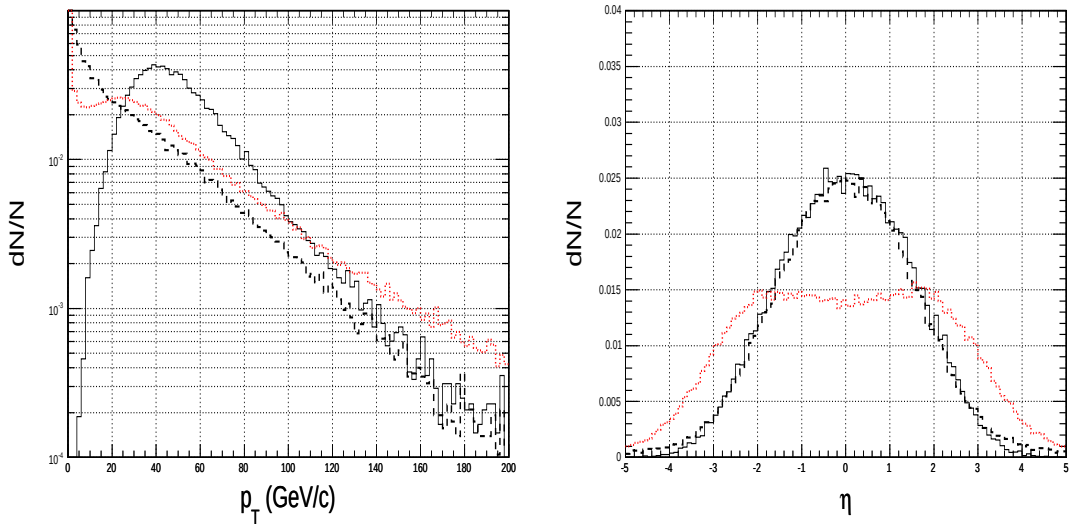


Figure 12: Normalized lepton p_T distribution (left) and normalized lepton η distribution (right) of signal with $m_H = 170\ GeV/c^2$ (solid), $t\bar{t}$ + jets (dash), and W + 4jets (dot) respectively

5.3 Event Selection of Jet Counting and E_T^{miss} (E-S)

A minimum number of four jets with $E_T > 25$ GeV are required for forward jet tagging and hadronic W reconstruction. The jets below the threshold will not be treated as a reconstruction object due to a large number of detector level jets that actually come from the fluctuation of electronic noise, pileup and underlying event.

The E_T^{miss} is required to be above 30 GeV. No jet energy correction is used for E_T^{miss} because of the significant difference in the generator level E_T^{miss} spectrum and detector jet E_T spectrum between the signal and major backgrounds as shown in Fig. 13. This issue will be further discussed in the summary of result.

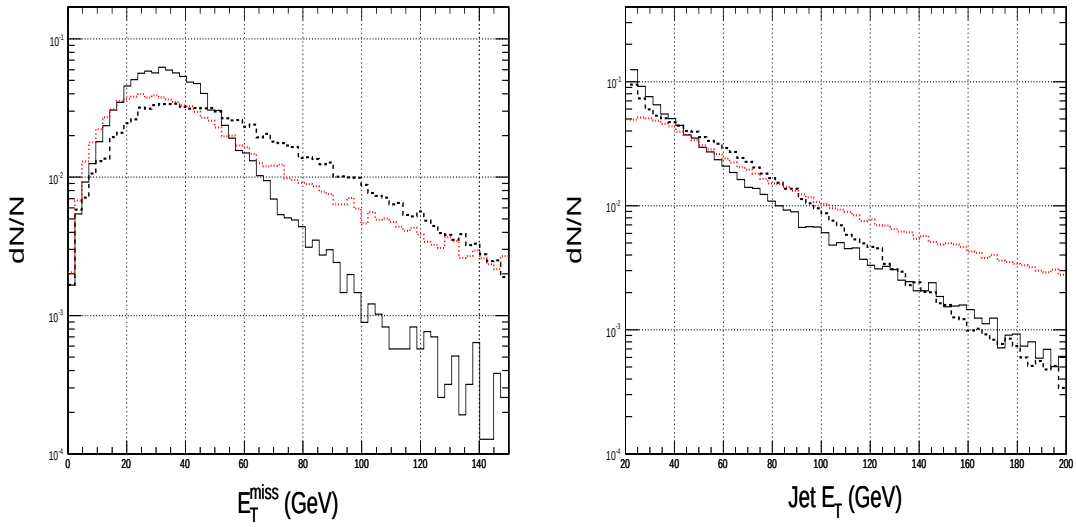


Figure 13: Normalized E_T^{miss} distribution (left) and normalized Jet E_T distribution (right) of signal with $m_H = 170\ GeV/c^2$ (solid), $t\bar{t}$ + jets (dash), and W + 4jets (dot) respectively

5.4 Forward Jet Tagging (FJT)

Forward jet tagging (FJT) leads to a strong suppression factor on those background channels that don't have a similar nature to the signal. Following cuts are used:

- Two jets with $\mathrm{E_T} > 30$ GeV, no other jets with $\mathrm{E_T} > 30$ GeV in the further forward region
- $\eta_1 \cdot \eta_2 < 0$
- $|\eta_1 - \eta_2| > 3.8$
- $\mathrm{m_{qq}} > 800\ \mathrm{GeV/c^2}$, where $\mathrm{m_{qq}}$ is the invariant mass of two forward jets

5.5 Hadronic W Reconstruction (H-W)

Any two central jets pairs with jet $\mathrm{E_T} > 25$ GeV are tested as a candidate of hadronic W. The di-jet invariant mass must be within 25 GeV window of true W boson mass. If multiple pairs satisfy the criterion, the one with least error to true W mass is selected.

5.6 Leptonic W Reconstruction (L-W)

Lepton momentum and $\mathrm{E_T^{miss}}$ are used for the leptonic W reconstruction:

$$\begin{aligned} \mathrm{p_x^\nu} &= \mathrm{E_x^{miss}} \\ \mathrm{p_y^\nu} &= \mathrm{E_y^{miss}} \\ \mathrm{p_z^\nu} &= \frac{\mathrm{A} \cdot \mathrm{p_z^\ell} \pm \sqrt{\mathrm{A^2 (p_z^\ell)^2 - B}}}{(\mathrm{p_x^\ell})^2 + (\mathrm{p_y^\ell})^2} \end{aligned} \tag{1}$$

where $\mathrm{E_{x(y)}^{miss}}$ is the x(y)-component of $\mathrm{E_T^{miss}}$, $\mathrm{p_{x(y,z)}^\nu}$ and $\mathrm{p_{x(y,z)}^\ell}$ are the $\mathrm{p_{x(y.z)}}$ of neutrino and lepton from W decay, A = $\frac{\mathrm{m_W^2}}{2} + \mathrm{p_x^\ell p_x^\nu} + \mathrm{p_y^l p_y^\nu}$, B = $[(\mathrm{p_x^\ell})^2 + (\mathrm{p_y^\ell})^2][(\mathrm{p^\ell})^2(\mathrm{p_t^\nu})^2 - \mathrm{A}^2]$. The known W boson mass is used to calculate the z-component of neutrino's momentum.

Since the Higgs boson mass is very close to di-W mass ($\sim 160.8\ \mathrm{GeV/c^2}$), two Ws are nearly static in the rest frame of Higgs boson and flying in almost the same direction in the experimental frame. This characteristics can be used to resolve the ambiguity of neutrino's momentum in z direction. The $\Delta\mathrm{R}$ between each of two leptonic W candidates and the hadronic W is computed and compared. The one with smaller $\Delta\mathrm{R}$ is selected as leptonic W.

5.7 Selection Criterion for Higgs Boson Mass below 160 $\mathrm{GeV/c^2}$

Most of the selection is the same as the case for $\mathrm{m_H} \geq 160\ \mathrm{GeV/c^2}$, except the hadronic W mass selection window is between 30 and 90 $\mathrm{GeV/c^2}$.

6 Summary of Intermediate Results

The hadronic and leptonic Ws are two crucial objects to reconstruct Higgs boson. Using VBF Higgs boson signal events, the resolution of the detector W with respect to the generated W is illustrated by two quantities: average W $\mathrm{p_T}$ error ($= p_T^{det} - p_T^{gen}$) and $\Delta\mathrm{R}$ between the detector and generator W (Fig. 14 and 15). The limited $\mathrm{E_T^{miss}}$ resolution causes the worse quality of the leptonic W than that of the hadronic W.

The reconstructed leptonic W $\mathrm{p_T}$ has been used to evaluate the possibility of applying the jet energy correction for $\mathrm{E_T^{miss}}$ (Fig. 16). As the $\mathrm{m_H}$ goes up, the W $\mathrm{p_T}$ error shifts from positive to negative. The positive value of $\mathrm{p_T}$ error represents over-measured $\mathrm{E_T^{miss}}$, a common feature of intrinsic low $\mathrm{E_T^{miss}}$ events (e.g., QCD events) that various detector effects randomly enhance the $\mathrm{E_T^{miss}}$. In this case, jet energy correction will not work for $\mathrm{E_T^{miss}}$. Due to the low Higgs boson mass and induced low $\mathrm{E_T^{miss}}$ spectrum studied in this analysis, jet energy correction is not applied.

For $\mathrm{m_H} > 200\ \mathrm{GeV/c^2}$, the W $\mathrm{p_T}$ error turns negative, which shows the effect of low jet energy response in the detector that causes the under-measurement of $\mathrm{E_T^{miss}}$. This is the common feature of high $\mathrm{E_T^{miss}}$ events. In this

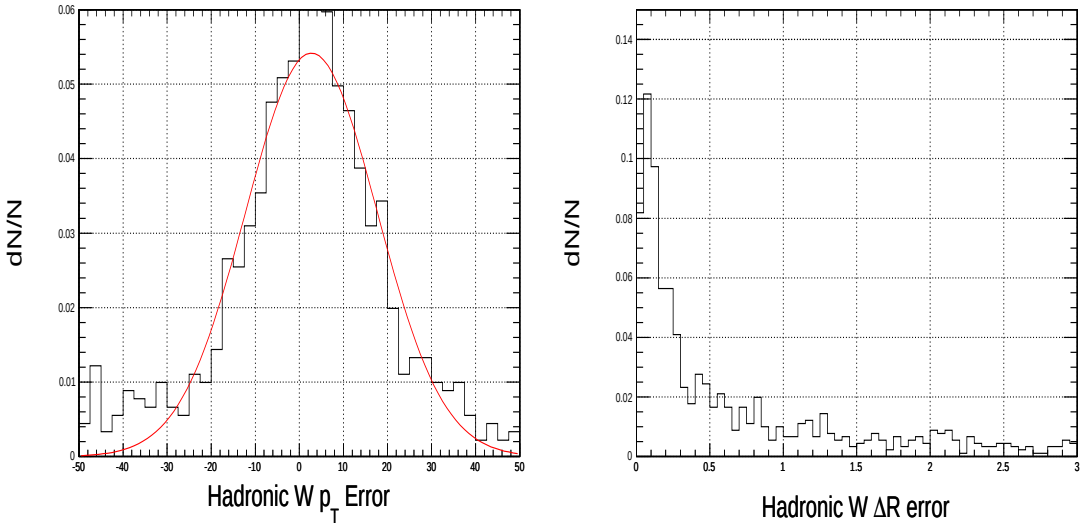


Figure 14: Hadronic W properties: $\mathrm{p_T}$ error (left) and $\Delta\mathrm{R}$ (right) between the detector and generator level hadronic W. The $\mathrm{p_T}$ error is fitted by a Gaussian with $\sigma \sim 15.1$ GeV/c.

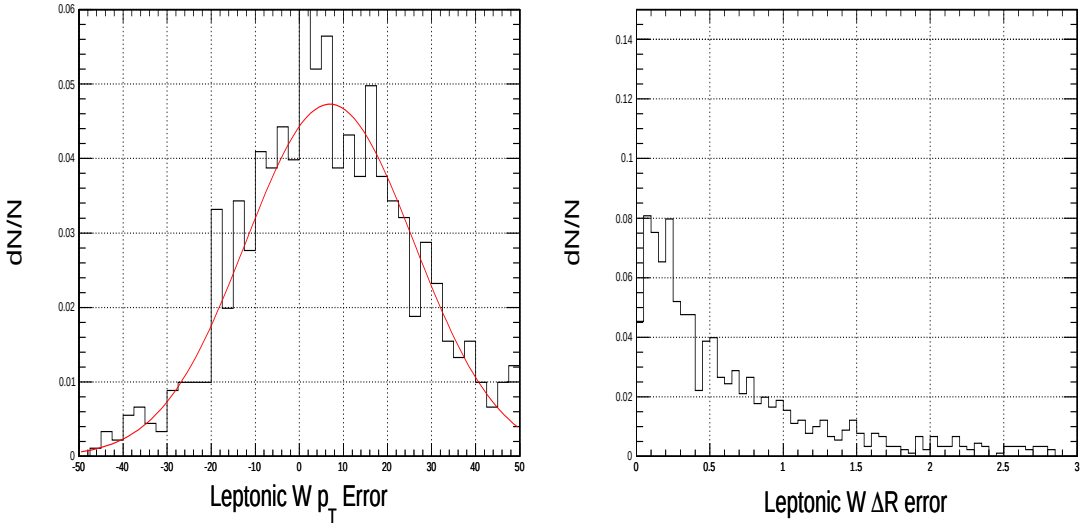


Figure 15: Leptonic W properties: $\mathrm{p_T}$ error (left) and $\Delta\mathrm{R}$ (right) between the detector and generator level leptonic W. The $\mathrm{p_T}$ error is fitted by a Gaussian with $\sigma \sim 19.5$ GeV/c.

case, random detector effects will only deteriorate the $\mathrm{E_T^{miss}}$ resolution but not change the scale of $\mathrm{E_T^{miss}}$. The restoration of $\mathrm{E_T^{miss}}$ scale will need jet energy correction. Roughly at $\mathrm{m_H}$ = 250 $\mathrm{GeV/c^2}$, average W $\mathrm{p_T}$ error is $\sim$ 0 after applying jet energy correction.

But for both cases, there is no significant difference in W $\mathrm{p_T}$ resolution between the "corrected" and "uncorrected" $\mathrm{E_T^{miss}}$ (note that the correction using muonmentum is applied for both cases).

The distance between the hadronic and leptonic W ($\Delta\mathrm{R_{Di-W}}$) defined in Eq. 2 plays an important role in the reconstructed Higgs boson mass. The error on the distance between the detector and generator level can fitted by a Gaussian distribution with a σ of $\sim$ 0.25 (Fig. 17), which leads to $\sim$ 20 GeV variance in reconstructed Higgs mass. The long tail is due to two factors: the wrong identification of jets in the hadronic W reconstruction, limited $\mathrm{E_T^{miss}}$ resolution in the leptonic W reconstruction.

$$\Delta\mathrm{R_{Di-W}} = \sqrt{\Delta\eta^2_{\mathrm{Di-W}} + \Delta\phi^2_{\mathrm{Di-W}}} \tag{2}$$

The selection efficiency for the signal and background with respect to two reconstruction scenarios of $\mathrm{m_H} \geq 160$

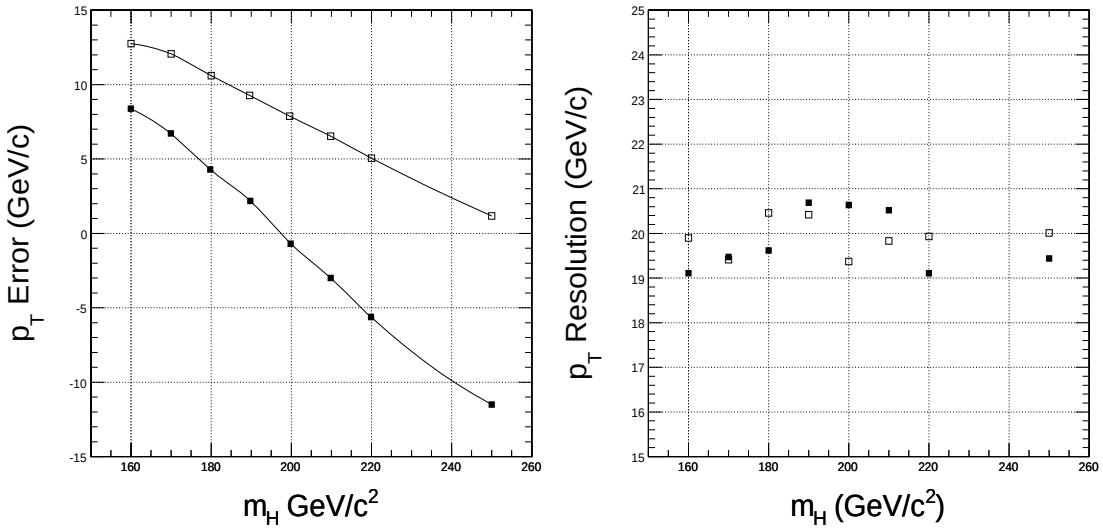


Figure 16: Leptonic W properties as a function of $\mathrm{m_H}$: average $\mathrm{p_T}$ error (left) and $\mathrm{p_T}$ resolution (right) between the detector and generated leptonic W with uncorrected $\mathrm{E_T^{miss}}$ (solid square) and corrected $\mathrm{E_T^{miss}}$ (open square)

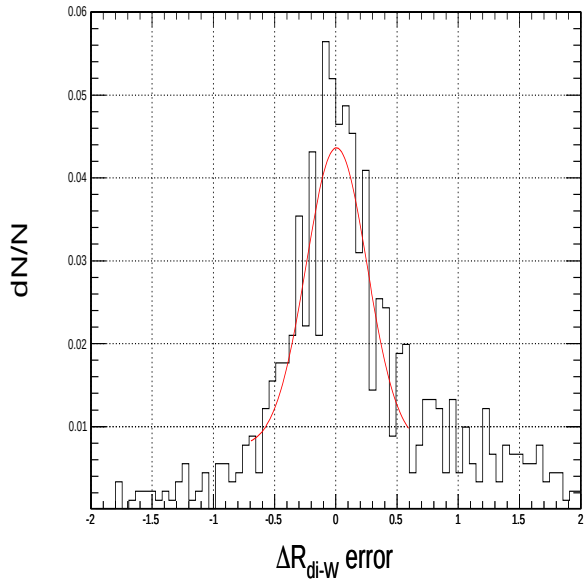


Figure 17: Di-W $\Delta\mathrm{R}$ error between detector and generator level in VBF Higgs events with $\mathrm{m_H}$ = 170 GeV/$\mathrm{c^2}$. The distribution is fitted by a Gaussian with $\sigma = 0.25$

$\mathrm{GeV/c^2}$ and $\mathrm{m_H} < 160$ $\mathrm{GeV/c^2}$ are included in Table 5 and 6 respectively. The summary of the selection cuts is included in Table 7.

Higgs boson mass distribution of backgrounds and VBF Higgs signal ($\mathrm{m_H} = 170$ $\mathrm{GeV/c^2}$) under $\mathrm{m_H} \geq 160$ $\mathrm{GeV/c^2}$ scenario is shown in Fig. 18 and 19 respectively.

Table 5: Selection efficiency for signal and background events with scenario of $\mathrm{m_H} \geq 160\ \mathrm{GeV/c^2}$

Channels	Trigger	L-S	E-S	FJT	H-W	L-W	σ (fb)
VBF Higgs ($\mathrm{m_H}$=160)	0.594	0.519	0.346	0.347	0.798	0.642	16.15
VBF Higgs ($\mathrm{m_H}$=170)	0.607	0.539	0.372	0.353	0.795	0.552	15.99
VBF Higgs ($\mathrm{m_H}$=180)	0.618	0.571	0.383	0.348	0.810	0.559	16.28
VBF Higgs ($\mathrm{m_H}$=190)	0.629	0.586	0.400	0.366	0.809	0.542	14.16
VBF Higgs ($\mathrm{m_H}$=200)	0.644	0.596	0.413	0.374	0.821	0.535	13.78
VBF Higgs ($\mathrm{m_H}$=210)	0.652	0.603	0.424	0.370	0.810	0.549	13.43
VBF Higgs ($\mathrm{m_H}$=220)	0.664	0.608	0.443	0.383	0.814	0.528	13.35
VBF Higgs ($\mathrm{m_H}$=250)	0.682	0.610	0.411	0.383	0.835	0.542	10.71
$\mathrm{t\bar{t}}$ + jets	0.422	0.310	0.465	0.063	0.816	0.568	1494.2
WW + jets (QCD)	0.227	0.539	0.078	0.048	0.718	0.393	9.27
WW + jets (EW)	0.252	0.530	0.417	0.319	0.768	0.458	7.88
ZZ + jets	0.147	0.289	0.097	0.051	0.758	0.594	1.00
ZW + jets	0.177	0.464	0.098	0.057	0.777	0.631	7.23
$\mathrm{W + t\bar{b}(\bar{t}b)}$	0.422	0.123	0.428	0.056	0.706	0.452	92.8
W + 4j ($\mathrm{W \to e/\mu/\tau + \nu}$)	0.553	0.360	0.303	0.136	0.451	0.502	1110.8
W + 3j ($\mathrm{W \to e/\mu/\tau + \nu}$)	0.479	0.330	0.116	0.107	0.406	0.523	701.8
Z + 4j ($\mathrm{Z \to ee/\mu\mu}$)	0.793	0.193	0.334	0.149	0.523	0.455	82.3
Z + 3j ($\mathrm{Z \to ee/\mu\mu}$)	0.750	0.220	0.160	0.105	0.455	0.513	72.4

Table 6: Selection efficiency for signal and background events with scenario of $\mathrm{m_H} < 160\ \mathrm{GeV/c^2}$

Channels	Trigger	L-S	E-S	FJT	H-W	L-W	σ (fb)
VBF Higgs ($\mathrm{m_H}$=120)	0.460	0.465	0.206	0.311	0.741	0.705	1.28
VBF Higgs ($\mathrm{m_H}$=130)	0.492	0.485	0.230	0.355	0.767	0.747	4.03
VBF Higgs ($\mathrm{m_H}$=140)	0.523	0.496	0.256	0.347	0.787	0.713	7.12
VBF Higgs ($\mathrm{m_H}$=150)	0.561	0.510	0.288	0.343	0.802	0.659	11.01
$\mathrm{t\bar{t}}$ + jets	0.422	0.310	0.465	0.063	0.807	0.570	1483.0
WW + jets (QCD)	0.227	0.122	0.078	0.048	0.744	0.397	9.70
WW + jets (EW)	0.252	0.530	0.417	0.319	0.781	0.454	7.94
ZZ + jets	0.147	0.289	0.097	0.051	0.758	0.565	0.954
ZW + jets	0.177	0.464	0.098	0.057	0.804	0.745	7.45
$\mathrm{W + t\bar{b}(\bar{t}b)}$	0.422	0.123	0.428	0.056	0.741	0.471	101.5
W + 4j ($\mathrm{W \to e/\mu/\tau + \nu}$)	0.553	0.360	0.303	0.136	0.457	0.488	1110.7
W + 3j ($\mathrm{W \to e/\mu/\tau + \nu}$)	0.479	0.330	0.116	0.107	0.430	0.534	758.0
Z + 4j ($\mathrm{Z \to ee/\mu\mu}$)	0.793	0.193	0.334	0.149	0.510	0.445	81.3
Z + 3j ($\mathrm{Z \to ee/\mu\mu}$)	0.750	0.220	0.160	0.105	0.443	0.513	70.0

Table 7: Summary of basic event selection cuts

Selection	Configuration
Lepton selection (L-S)	Electron: $\mathrm{E_T^{Hcal}/E_T^{Ecal} < 0.05}$ $\mathrm{0.9 < E/p < 1.8}$ $\mathrm{\lvert E_T^{0.2} - E_T^e \rvert < 5.0\ GeV}$ $\mathrm{\lvert (E_T^{0.2} - E_T^e)/E_T^e \rvert < 0.3}$ $\mathrm{E_T^{0.4}/E_T^e < 0.3}$ Muon : $\mathrm{\lvert E_T^{0.2} - p_T^\mu \rvert < 9.0\ GeV}$ $\mathrm{E_T^{0.2}/p_T^\mu < 0.3}$ $\mathrm{E_T^{0.4}/p_T^\mu < 0.3}$ $\mathrm{30 < p_T < 120\ GeV/c}$ $\mathrm{\Delta R_{\ell-j} > 0.5}$
Event selection (E-S)	$\mathrm{N_{jet} > 4}$ jets with $\mathrm{E_T > 25\ GeV}$ $\mathrm{E_T^{miss} > 30\ GeV}$
Forward jet tagging (FJT)	$\mathrm{E_T > 30\ GeV}$ $\eta_1 \cdot \eta_2 < 0$ $\lvert \eta_1 - \eta_2 \rvert > 3.8$ $\mathrm{m_{qq} > 800\ GeV/c^2}$
Hadronic W reco (H-W)	$\mathrm{\Delta m_W < 25\ GeV/c^2}$ ($\mathrm{m_H \geq 160\ GeV/c^2}$) $\mathrm{30 < \Delta m_W < 90\ GeV/c^2}$ ($\mathrm{m_H < 160\ GeV/c^2}$) select di-jet with least $\mathrm{\Delta m_W}$
Leptonic W reco (L-W)	select leptonic W candidates of smaller ΔR with hadronic W

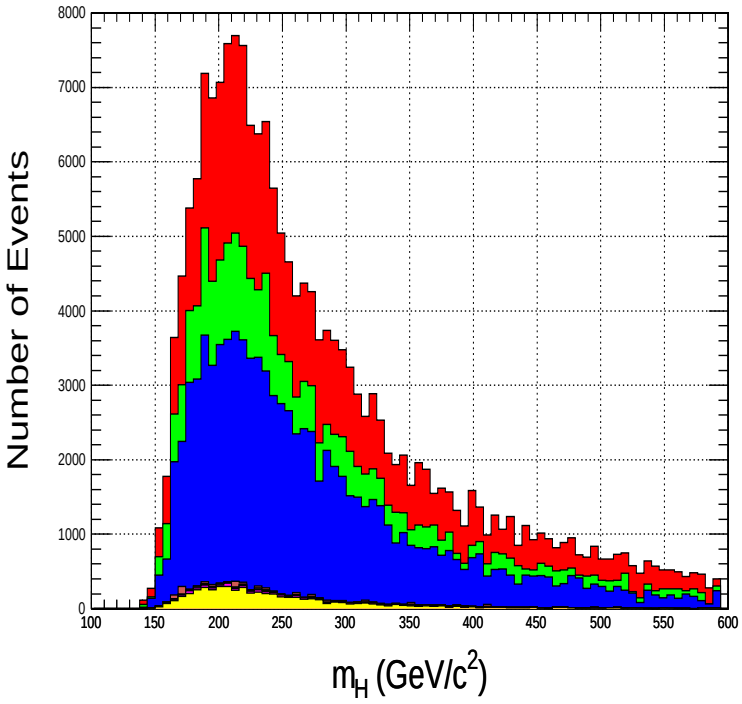


Figure 18: VBF Higgs mass reconstructed from background events under high-mass scenario. Major background include W + 4jets (red), W + 3jets (green), $\mathrm{t\bar{t}}$ + jets (blue), and W + $\mathrm{t\bar{b}}$ ($\mathrm{\bar{t}b}$)(yellow).

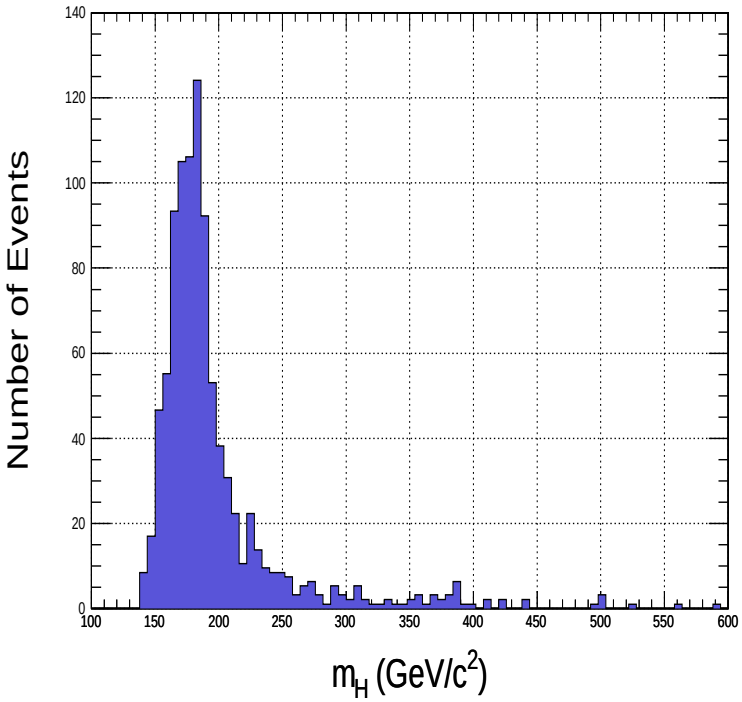


Figure 19: VBF Higgs mass reconstructed from signal events with $\mathrm{m_H} = 170\ \mathrm{GeV/c^2}$

7 Selection Optimization

In this section, more and optimized selection cuts are introduced to increase the significance of the signal ("temporarily" defined as $S/\sqrt{S+B}$) and the signal to background ratio (S/B), where S and B are the number of signal and background events respectively after the selection and scaled to an integrated luminosity of 60 fb^{-1}. At the starting point, S/B is less than 0.5%. The optimized selection is conducted in three steps with multiple selection cuts for each step.

Two scenarios are used to estimate the W/Z + jets background:

- Conservative scenario (c). W + 3jets and W + 4jets are both considered as the background with respect to their cross section in LO. An over-estimation of the background occurs due to overlap between the high order correction of W + 3jets and W + 4jets. So the significance and S/B based on this scenario can be taken as conservative. The same strategy is used for Z + 3jets and Z + 4jets.
- Optimistic scenario (o). Only W + 4jets are considered, which lead to an optimistic estimation of the background.

7.1 Optimization of Forward Jet Selection (Step-1)

Forward jet tagging is optimized with four related parameters: the higher and lower E_T threshold (E_T^{FH} and E_T^{FL}), the η distance ($\Delta\eta$), and the invariant mass (m_{qq}) of two tagged jets.

Fig. 20 shows the $\Delta\eta$ distribution of the background and signal (m_H =170 GeV/c^2). Fig. 21 shows the S/B with respect to different $\Delta\eta$ thresholds (minimum $\Delta\eta$ cut).

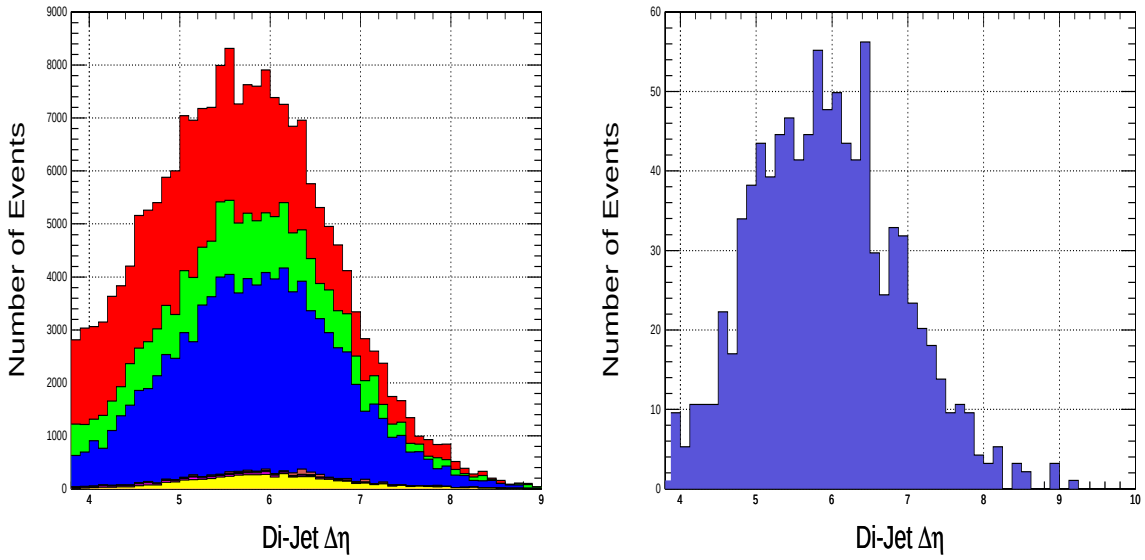


Figure 20: $\Delta\eta$ distribution of backgrounds (left) and signal (right) with m_H = 170 GeV/c^2. Major background processes include W + 4jets (red), W + 3jets (green), $t\bar{t}$ + jets (blue), and W + $t\bar{b}$ ($\bar{t}b$)(yellow).

Fig. 22 shows the m_{qq} distribution of the background and VBF Higgs signal (m_H = 170 GeV/c^2) with a $\Delta\eta$ threshold of 4.3. Fig. 23 shows the S/B with respect to different m_{qq} thresholds (minimum m_{qq} cut).

The S/B with respect to jet E_T threshold are tested under $m_{qq} > 1200$ GeV/c^2 and $\Delta\eta > 4.3$ (Table 8), which shows that a higher jet E_T threshold increases the S/B, but there is a significant loss of the signal efficiency for $E_T^{FL} > 30$ GeV.

In this step, the overall background is still several hundred times bigger than the signal. The loss of signal efficiency is significant with modest increase of S/B. A higher jet E_T threshold is preferred to reduce the systematic error of various detector effects as part of the optimization. Following configuration of the cuts is used for the event selection:

- $E_T^{FH} > 45$ GeV, $E_T^{FL} > 35$ GeV, $\Delta\eta > 4.2$, and $m_{qq} > 1200$ GeV/c^2.

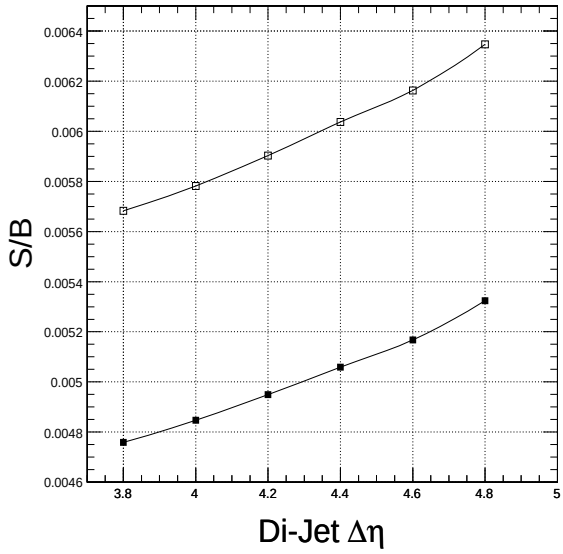


Figure 21: S/B with respect to different $\Delta\eta$ threshold for Conservative (solid square) and Optimistic Scenario (open square)

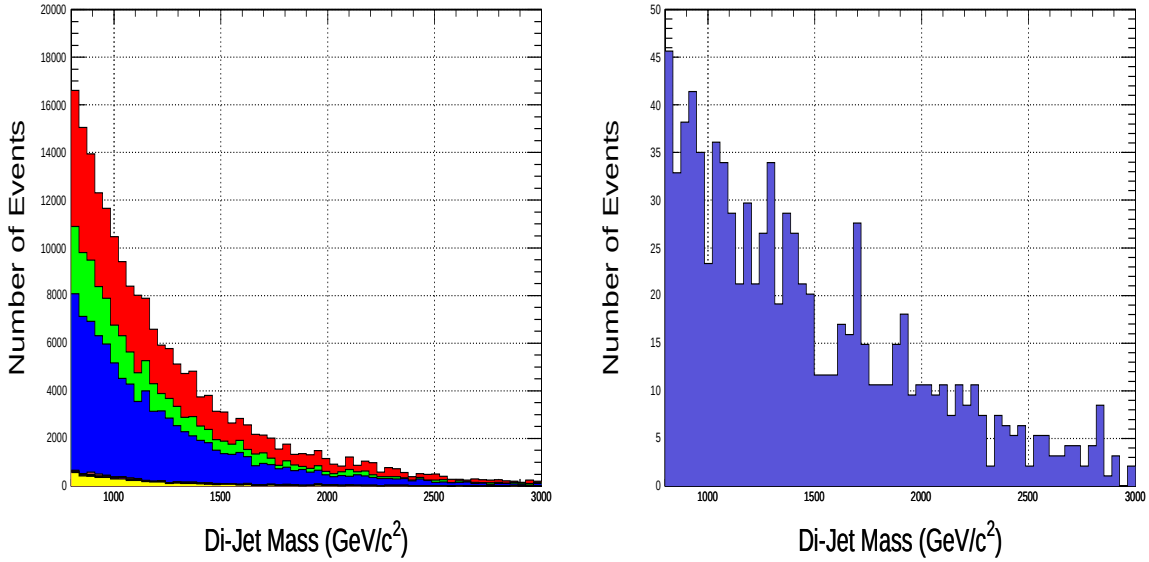


Figure 22: $\mathrm{m_{qq}}$ distribution of backgrounds (left) and signal (right) with $\mathrm{m_H} = 170\ \mathrm{GeV/c^2}$. Major background processes include W + 4jets (red), W + 3jets (green), $\mathrm{t\bar{t}}$ + jets (blue), and W + $\mathrm{t\bar{b}}$ ($\mathrm{\bar{t}b}$)(yellow).

Table 8: Forward jet tagging efficiency with various jet $\mathrm{E_T}$ threshold for Conservative (c) and Optimistic Scenario (o)

$\mathrm{E_T^{FH}}$	$\mathrm{E_T^{FL}}$	S/B (c)	S/B (o)	Signal Efficiency
35	30	0.0064	0.0073	0.612
40	30	0.0065	0.0074	0.603
45	30	0.0067	0.0077	0.601
40	35	0.0070	0.0080	0.600
45	35	0.0072	0.0082	0.506
45	40	0.0078	0.0090	0.504

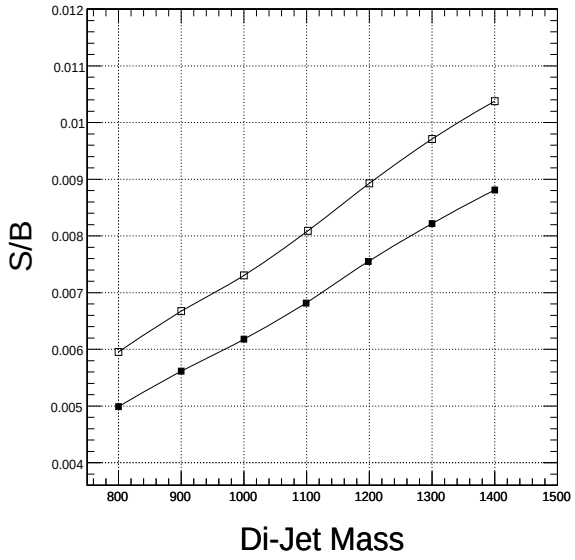


Figure 23: S/B with respect to different $\mathrm{m_{qq}}$ thresholds for Conservative (solid square) and Optimistic Scenario (open square)

For $\mathrm{m_H} < 160\mathrm{GeV/c^2}$, the selection cuts are modified with slightly lower jet $\mathrm{E_T}$ threshold:

- $\mathrm{E_T^{FH}} > 40$ GeV, $\mathrm{E_T^{FL}} > 30$ GeV, $\Delta\eta > 4.2$, and $\mathrm{m_{qq}} > 1200\ \mathrm{GeV/c^2}$.

After this step of the selection cuts, the overall S/B ratio with respect to various VBF Higgs boson mass is shown in Fig. 24.

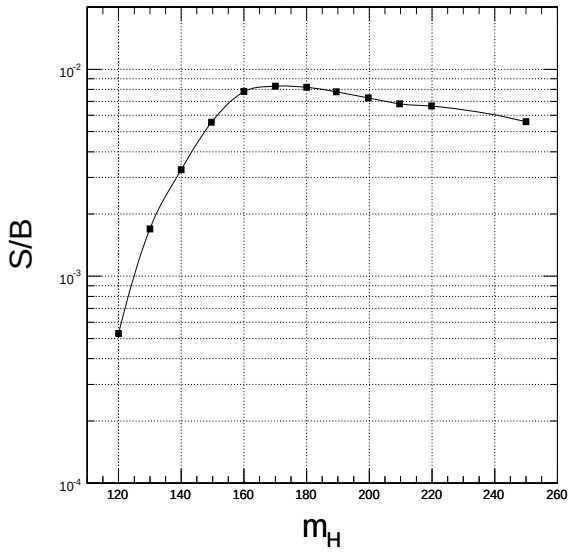


Figure 24: S/B with respect to various VBF Higgs mass by using the Conservative Scenario

7.2 Optimization of Central Jet Selection (Step-2)

The central jet selection is optimized with four related parameters: a higher and lower jet $\mathrm{E_T}$ threshold ($\mathrm{E_T^{CH}}$ and $\mathrm{E_T^{CL}}$) for the jets used for hadronic W reconstruction, hadronic W mass ($\Delta\mathrm{m_W}$) selection window, and number of extra jets ($\mathrm{N_{extra}}$).

The N_{extra} distribution of backgrounds and signal ($m_H = 170\ GeV/c^2$) is shown in Fig. 25. The results of S/B with respect to different selection cuts on the maximal number of N_{extra} is summarized in Table 9, that shows a large increase of S/B by requiring fewer extra central jets.

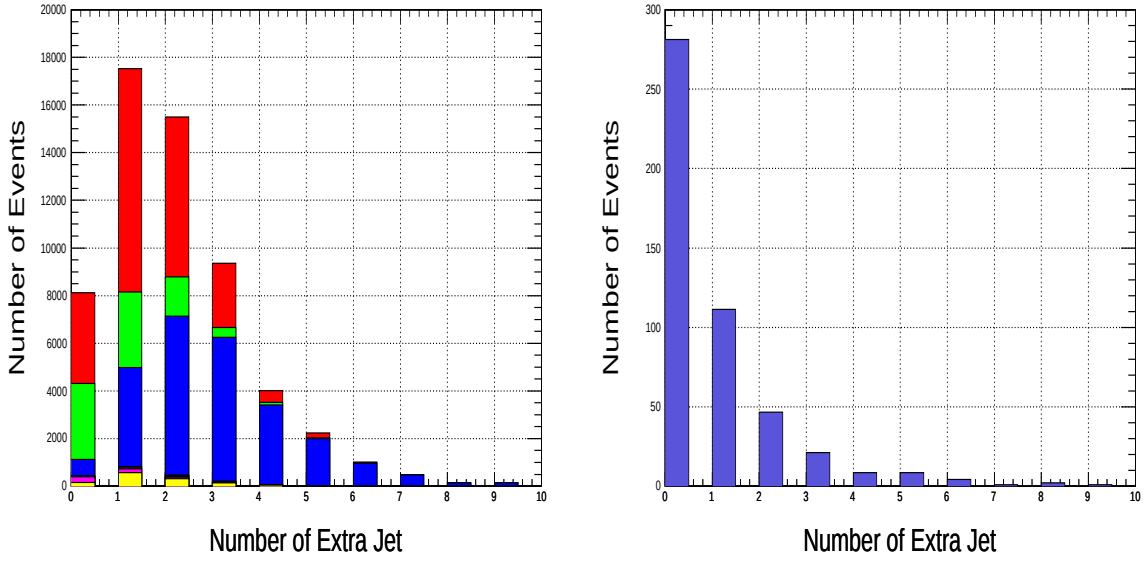


Figure 25: N_{extra} distribution of backgrounds (left) and signal (left) with m_H =170 GeV/c^2. Major background processes include W + 4jets (red), W + 3jets (green), $t\bar{t}$ + jets (blue), and W + $t\bar{b}$ ($\bar{t}b$)(yellow)

Table 9: Selection efficiency with various maximal number of extra jet for Conservative (c) and Optimistic Scenario (o)

MAX(N_{extra})	S/B (c)	S/B (o)	Signal efficiency
4	0.0074	0.0085	0.889
3	0.0079	0.0091	0.871
2	0.0092	0.0110	0.828
1	0.0133	0.0167	0.732
0	0.0337	0.0517	0.517

The m_W distribution of backgrounds and signal ($m_H = 170\ GeV/c^2$) is shown in Fig. 26. Using N_{extra} <2, the results of S/B with respect to different selection cuts on E_T^{CH} and E_T^{CL} are summarized in Table 10, that shows an insensitiveness of S/B with respect to jet E_T threshold.

Table 10: Selection efficiency with various jet E_T threshold for Conservative (c) and Optimistic Scenario (o)

E_T^{CH}	E_T^{CL}	S/B (c)	S/B (o)	Signal efficiency
30	25	0.0131	0.0164	0.707
30	30	0.0127	0.0156	0.498
35	25	0.0125	0.0155	0.649
35	30	0.0125	0.0153	0.480
35	35	0.0111	0.0133	0.310

In this step, the overall background is reduced to about 80 times of the signal. The loss of signal efficiency is modest with increase of S/B. A significant loss the signal efficiency with higher central jet E_T threshold for hadronic W reconstruction is observed because the signal has lower jets E_T than that of the background. The control of N_{extra} provides a large increase of S/B. Two schemes are defined with respect to N_{extra}:

- $E_T^{CH} > 30\ GeV$, $E_T^{CL} > 25\ GeV$, $\Delta m_W < 20\ GeV/c^2$, and $N_{extra} < 2$. In this scheme (called Loose Extra Jet Veto), one extra jet is allowed, which is more "inclusive" to the configuration of ISR and FSR in the event generation.
- $E_T^{CH} > 30\ GeV$, $E_T^{CL} > 25\ GeV$, $\Delta m_W < 20\ GeV/c^2$, and $N_{extra} < 1$. In this scheme (called Extra

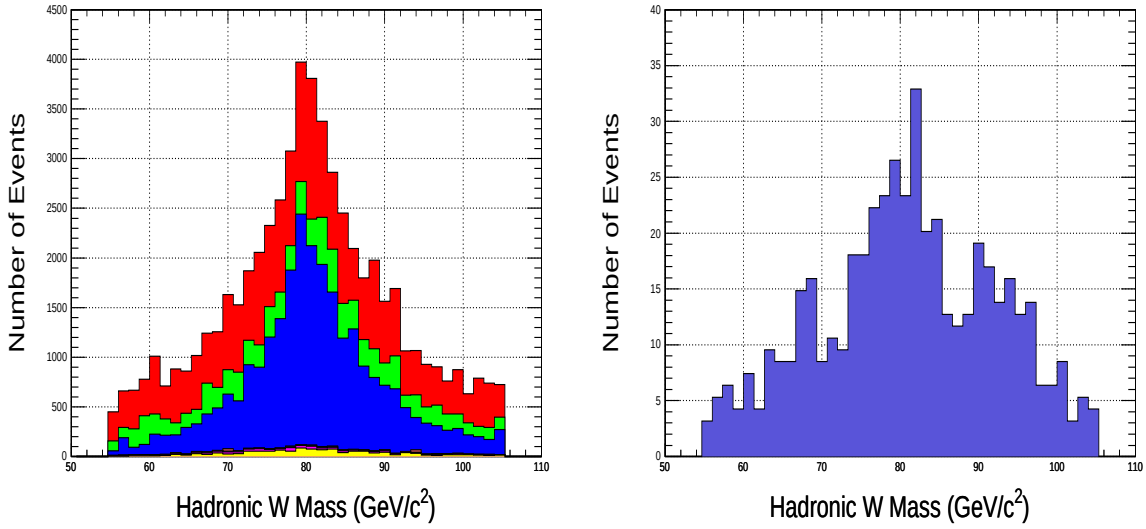


Figure 26: m_W distribution of backgrounds (left) and signal (right) with $m_H = 170\ GeV/c^2$. Major background processes include W + 4jets (red), W + 3jets (green), $t\bar{t}$ + jets (blue), and W + $t\bar{b}$ ($\bar{t}b$)(yellow).

Jet Veto), no extra jet is allowed, which provides stronger suppression of background (especially $t\bar{t}$ + jets events) and achieves higher S/B.

Due to limited statistics of background events, a loose forward jet tagging cut is used: $m_{qq} > 1000\ GeV/c^2$.

For $m_H < 160\ GeV/c^2$, the selection cuts of W mass selection is modified:

- For Loose Extra Jet Veto, $E_T^{CH} > 30$ GeV, $E_T^{CL} > 25$ GeV, $30 < m_W < 90\ GeV/c^2$, and $N_{extra} < 2$
- For Extra Jet Veto, $E_T^{CH} > 30$ GeV, $E_T^{CL} > 25$ GeV, $30 < m_W < 90\ GeV/c^2$, and $N_{extra} < 1$.

  A loose forward jet tagging cut is used: $m_{qq} > 1000\ GeV/c^2$.

Using optimized selection cuts, the overall S/B and significance with respect to various VBF Higgs boson mass is shown in Fig. 27.

7.3 Optimization of qqWW System (Step-3)

In this step, physics nature of the signal events is exploited in order to highly suppress the background processes:

- E_T^{miss} of di-W and forward jet system (called qqWW system, defined in Eq. 3).

$$E_T^{miss} = |\vec{E_T^H} + \vec{E_T^{J1}} + \vec{E_T^{J2}}| \tag{3}$$

where $\vec{E_T^H}$ is the transverse energy vector of Higgs, $\vec{E_T^{J1}}$ and $\vec{E_T^{J2}}$ are the transverse energy vectors of forward jets. qqWW system is the key part of signal events which should contain small E_T^{miss}. For background events, the existence of extra jets, extra leptons missed in the detector reconstruction, and low p_T leptons or low E_T jets without being counted, will make more significant E_T^{miss} in qqWW system.

The E_T^{miss} in qqWW system of backgrounds and signal ($m_H = 170\ GeV/c^2$) is shown in Fig. 28. The S/B and significance with respect to the maximum E_T^{miss} cut in qqWW system is shown in Fig. 29. A large increase of S/B and significance is achieved due to fundamental difference in the E_T^{miss} distribution between the signal and background.

- ΔR between the lepton and hadronic W.

  A significant difference in ΔR distribution between backgrounds and signal ($m_H = 170\ GeV/c^2$) is shown in Fig. 30. The S/B and significance as a function of ΔR cut combined with $E_T^{miss} < 40$ GeV is shown in

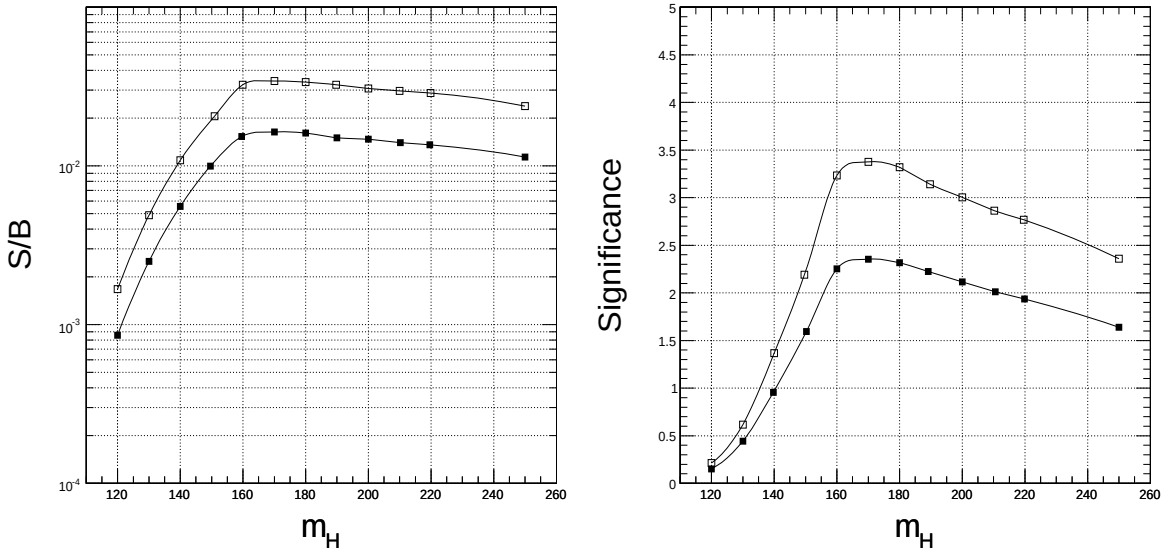


Figure 27: S/B verse VBF Higgs mass (left) and significance verse VBF Higgs mass (right). The higher (lower) S/B and significance curves correspond to Extra Jet Veto (Loose Extra Jet Veto) Scheme respectively

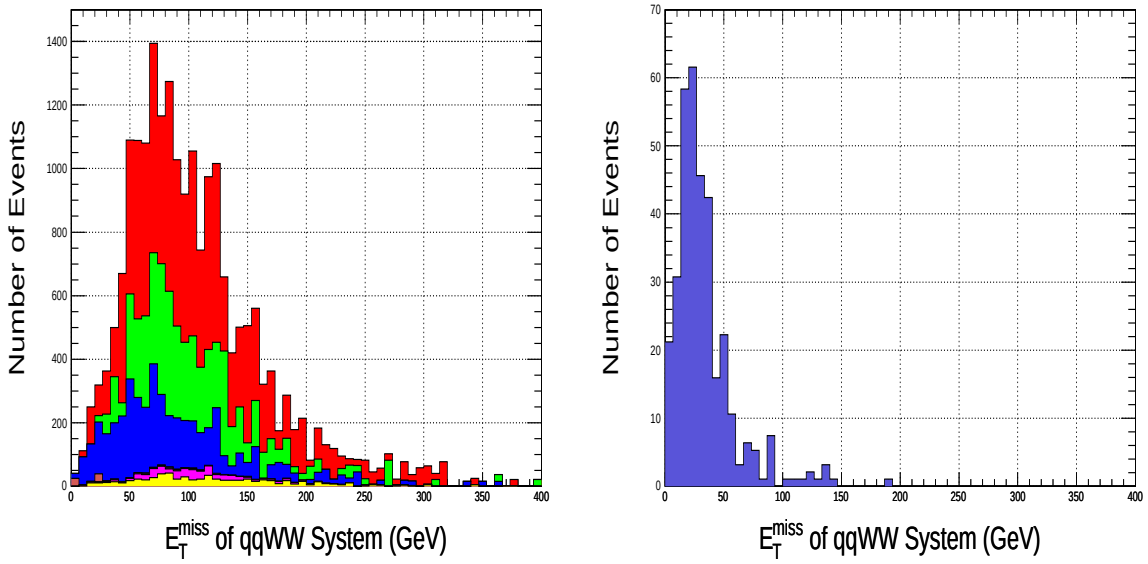


Figure 28: $\mathrm{E_T^{miss}}$ of qqWW system of backgrounds (left) and signal (right) with $\mathrm{m_H} = 170\ \mathrm{GeV/c^2}$. Major background include W + 4jets (red), W + 3jets (green), $\mathrm{t\bar{t}}$ + jets (blue), and W + $\mathrm{t\bar{b}}$ ($\mathrm{\bar{t}b}$)(yellow).

Fig. 31. A large increase of S/B and significance is achieved with a low threshold of $\Delta\mathrm{R}$ (maximum $\Delta\mathrm{R}$ cut).

- $\Delta\mathrm{R}$ between the hadronic and leptonic W.

  In the reconstruction of semi-leptonic W, a smaller $\Delta\mathrm{R}$ with the hadronic W is selected to remove the ambiguity caused by neutrino momentum in z-direction. For low mass Higgs boson, this parameter can also provide a strong suppression of background events as illustrated in Fig. 32. Signal events populate in $\Delta\mathrm{R} < 1.0$, while background has a much longer tail. A threshold of 1.0 for $\Delta\mathrm{R}$ is implemented with little loss of Higgs boson efficiency. However for high mass Higgs, the $\Delta\mathrm{R}$ is not small.

After this step of the selection, the overall background is reduced to about the same level of the signal with $\mathrm{m_H} = 170\ \mathrm{GeV/c^2}$. Several effective selections make a significant increase of significance and S/B. This effect is more apparent combined with Extra Jet Veto Scheme ($\mathrm{N_{extra}} < 1$). Because the background is reduced to a very low

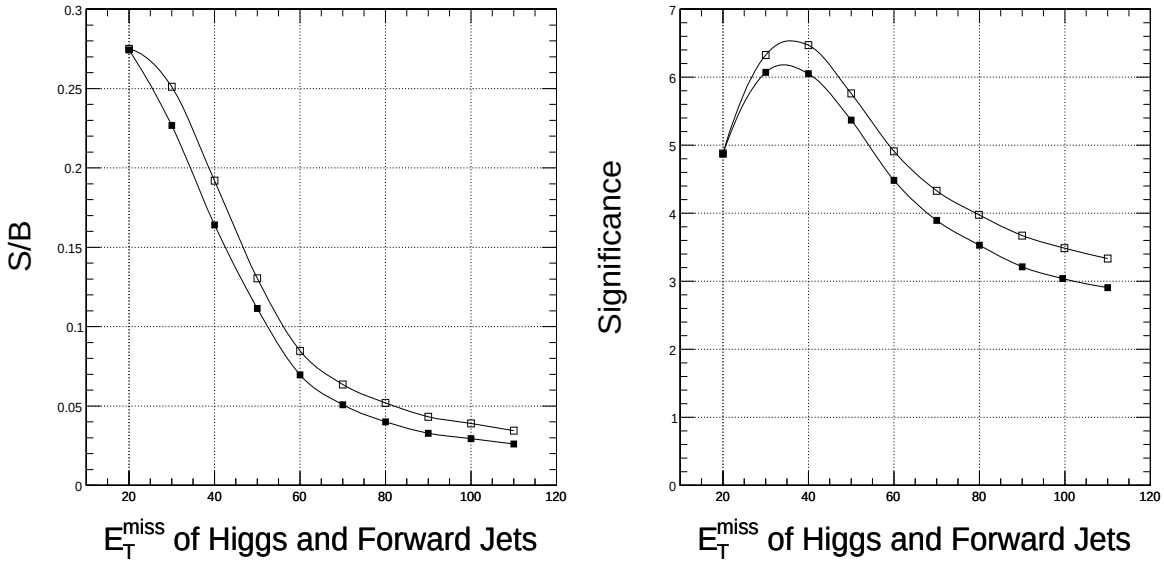


Figure 29: S/B verse $\mathrm{E_T^{miss}}$ cut in qqWW system (left) and significance at 60 $\mathrm{fb^{-1}}$ verse $\mathrm{E_T^{miss}}$ cut in qqWW system (right). The higher (lower) S/B and significance curves correspond to optimistic (conservative) scenario respectively

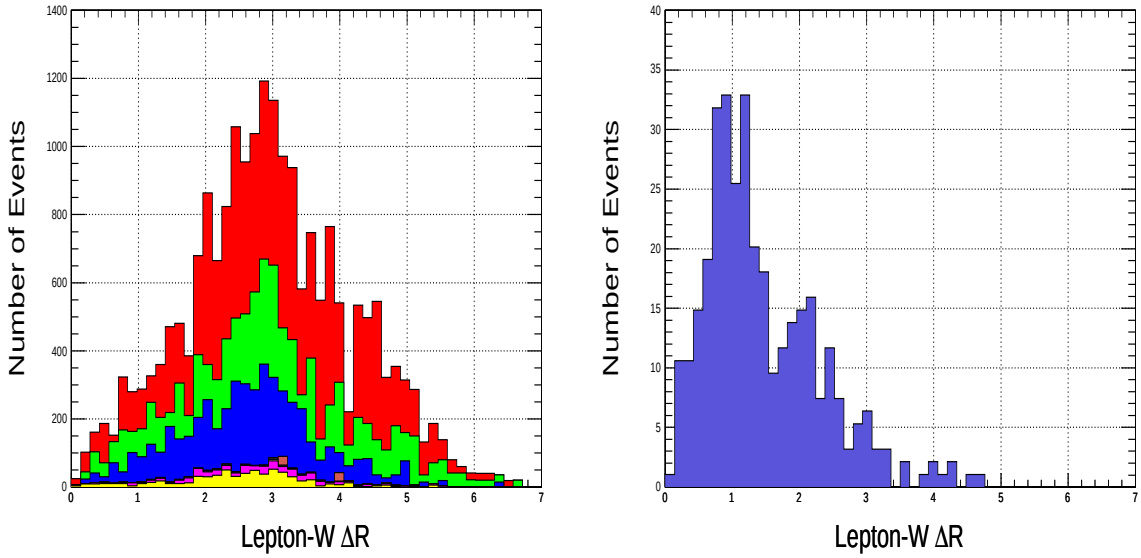


Figure 30: $\Delta\mathrm{R}$ between leptonic and hadronic W of backgrounds (left) and signal (right) with $\mathrm{m_H} = 170\ \mathrm{GeV/c^2}$. Major background include W + 4jets (red), W + 3jets (green), $\mathrm{t\bar{t}}$ + jets (blue), and W + $\mathrm{t\bar{b}}$ ($\mathrm{\bar{t}b}$)(yellow). In these plots, Loose Extra Jet Veto Scheme in Step-2 is used.

statistics, there is a large statistical uncertainty. Two schemes of selection cuts according the extra jet selection schemes are adopted:

1. For Loose Extra Jet Veto, $\mathrm{E_T^{miss}}$(qqWW) < 40 GeV, ΔR(lepton-W) < 1.6, and ΔR(Di-W) < 1.0.
2. For Extra Jet Veto, $\mathrm{E_T^{miss}}$(qqWW) < 40 GeV, ΔR(lepton-W) < 2.0, and ΔR(Di-W) < 1.0.

   In this Scheme, a stronger suppression of background and improvement of the significance and S/B can be achieved. In order to get enough statistics to estimate the significance and S/B, several cuts are loosened.

For $\mathrm{m_H} < 160\ \mathrm{GeV/c^2}$, the selection cuts are the same in this step.

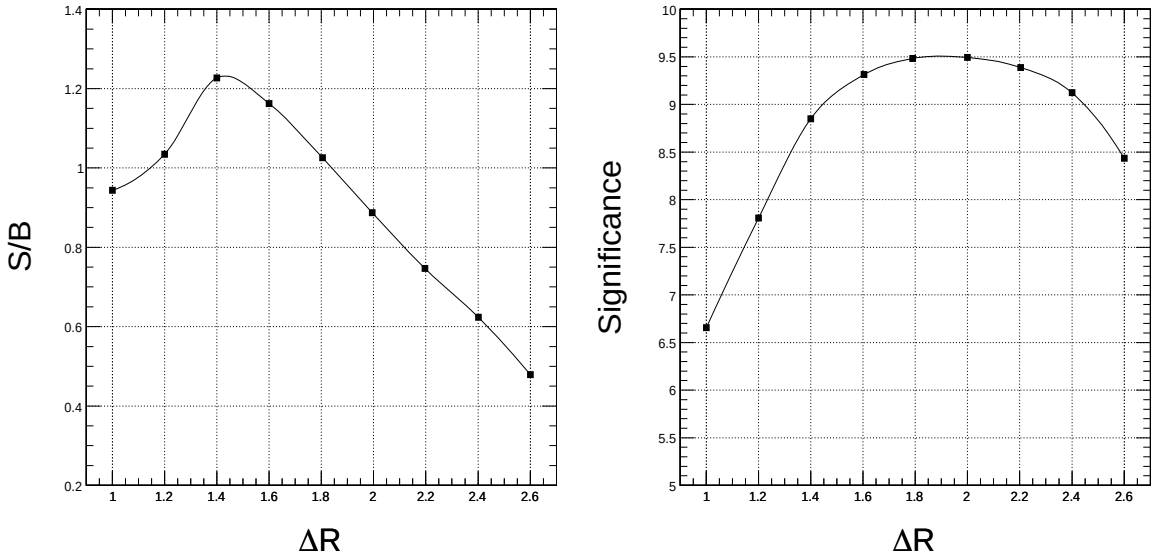


Figure 31: S/B verse ΔR cut (left) and significance verse ΔR cut (right). Both use $\mathrm{E_T^{miss}} < 40$ GeV for qqWW system and loose Extra Jet Veto Scheme in Step-2. Due to strong suppression of the W + 3jets background from combining ΔR and $\mathrm{E_T^{miss}}$ cuts, the difference between Conservative and Optimistic Scenario is negligible.

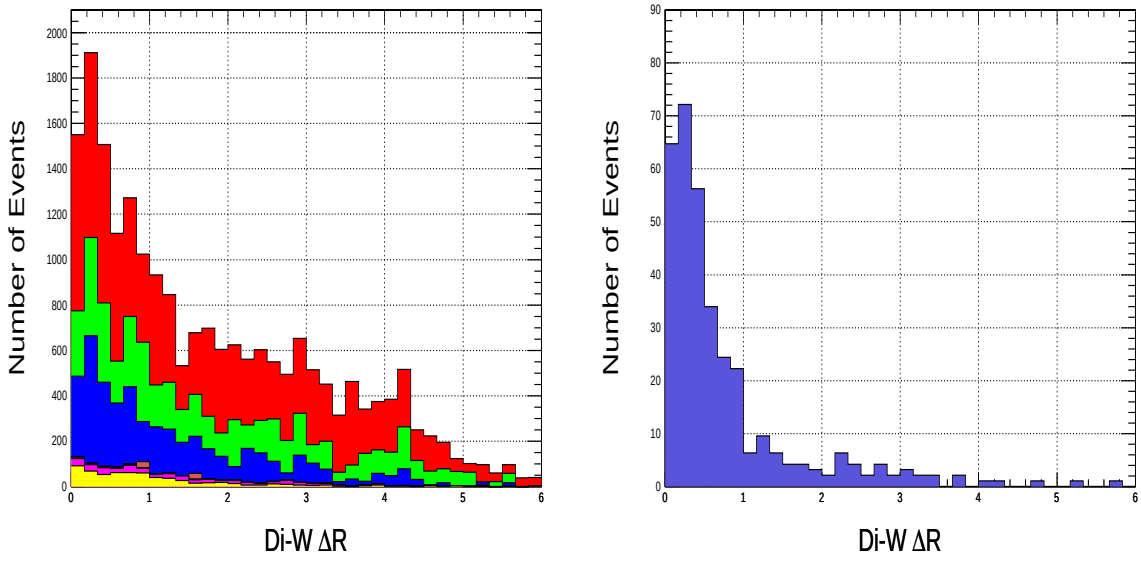


Figure 32: ΔR between semi-leptonic and hadronic W of backgrounds (left) and signal (right) with $\mathrm{m_H} = 170$ $\mathrm{GeV/c^2}$. Major background include W + 4jets (red), W + 3jets (green), $\mathrm{t\bar{t}}$ + jets (blue), and W + $\mathrm{t\bar{b}}$ ($\mathrm{\bar{t}b}$)(yellow). In these plots, Loose Extra Jet Veto Scheme in Step-2 is used.

It is anticipated, the best selection efficiency and significance will be achieved by combining the Loose Extra Jet Veto and Extra Jet Veto with $\mathrm{m_{qq}} < 1200$ $\mathrm{GeV/c^2}$ in Step-1, $\mathrm{N_{extra}} < 1$ in Step-2, and ΔR(lepton-W) < 1.6 in Step-3.

8 Summary of the Optimization Selection Results

8.1 Discovery Potential

After optimized selection cuts, the overall S/B and significance with respect to various Higgs boson masses based on the conservative scenario are shown in Fig. 33 which are scaled to an integrated luminosity of 30 $\mathrm{fb^{-1}}$. Several

general methods based on the number of signal and background events after the final selection are used to compute the significance:

- S_c = S/$\sqrt{B}$.
- S_{cP} = $\sqrt{2\ln Q}$, where Q=$(1 + S/B)^{S+B}e^{-S}$. S and B are expected average signal and background events at a given luminosity condition [19].

Since the overall number of background events is limited, there is roughly 50% statistical uncertainty for major backgrounds' efficiency in the final step of selection (Section 7.3). If we combine several main backgrounds together (treating them the same under the first order approximation), the statistical uncertainty is roughly 35%. To account this factor into the significance calculation, we use two schemes:

- Using the predicted B to makes a "normal" estimation of the significance.
- Using a scaled B of 50% more than the predicted one to makes a "pessimistic" estimation of the significance. The 50% takes into account the statistical uncertainty in the predicted average number of background events.

In the calculation of the significance based on S_{cP}, the detector systematic uncertainty needs to be considered (Section 10 summarizes the detector systematic uncertainties). Another type of statistical uncertainty that relate to stochastic effect of event counting for the experiment measurement needs to be considered and has been taken into account inside the algorithm[19].

8.2 Selection Efficiency

The optimized selection cuts are summarized in Table 11. Using these cuts, the efficiency for signal and backgrounds with respect to two reconstruction scenarios of $m_H \geq 160\ GeV/c^2$ and $m_H < 160\ GeV/c^2$ are included in Table 12 and 13 respectively.

Table 11: Summary of optimization cuts for $m_H \geq 160\ GeV/c^2$ ($m_H < 160\ GeV/c^2$)

Selection	Loose Extra Jet Veto	Extra Jet Veto
Step-1 (L-S)	$E_T^{FH} > 45(40)$ GeV $E_T^{FL} > 35(30)$ GeV $\Delta\eta > 4.2$ $m_{qq} > 1200\ GeV/c^2$	$E_T^{FH} > 45(40)$ GeV $E_T^{FL} > 35(30)$ GeV $\Delta\eta > 4.2$ $m_{qq} > 1000\ GeV/c^2$
Step-2 (E-S)	$E_T^{CH} > 30$ GeV $E_T^{CL} > 25$ GeV $60 < m_W < 100\ GeV/c^2$ $(30 < m_W < 90\ GeV/c^2)$ $N_{extra} < 2$	$E_T^{CH} > 30$ GeV $E_T^{CL} > 25$ GeV $60 < m_W < 100\ GeV/c^2$ $(30 < m_W < 90\ GeV/c^2)$ $N_{extra} < 1$
Step-3 (FJT)	E_T^{miss}(qqWW)< 40 GeV ΔR(lepton-W)< 1.6 ΔR(Di-W)< 1.0	E_T^{miss}(qqWW)< 40 GeV ΔR(lepton-W)< 2.0 ΔR(Di-W)< 1.0

Due to very low statistics, some background processes get zero efficiency in the final step of selection, the estimation of the upper limit of the efficiency for those processes has been made:

- For WW + jets (QCD), ZZ + jets, and ZW + jets, the selection efficiency is estimated for a reduction factor of 100 in Step-3, because those backgrounds with non-zero efficiency have a reduction factor ranging from 100 to 300.
- For WW + jets (EW) and $W + t\bar{b}(\bar{t}b)$, we estimate the upper limit of the selection efficiency by assuming one event passing Step-3, because the number of events of those processes that corresponds to over 50% of the luminosity in 60 fb^{-1} have been produced as shown in Table 2.
- For W + 3jets (Z + 3jets), the selection efficiency of W + 4jets (Z + 4jets) in Step-3 is used. It can be seen that W + 3jets contributes a large uncertainty in the final result.

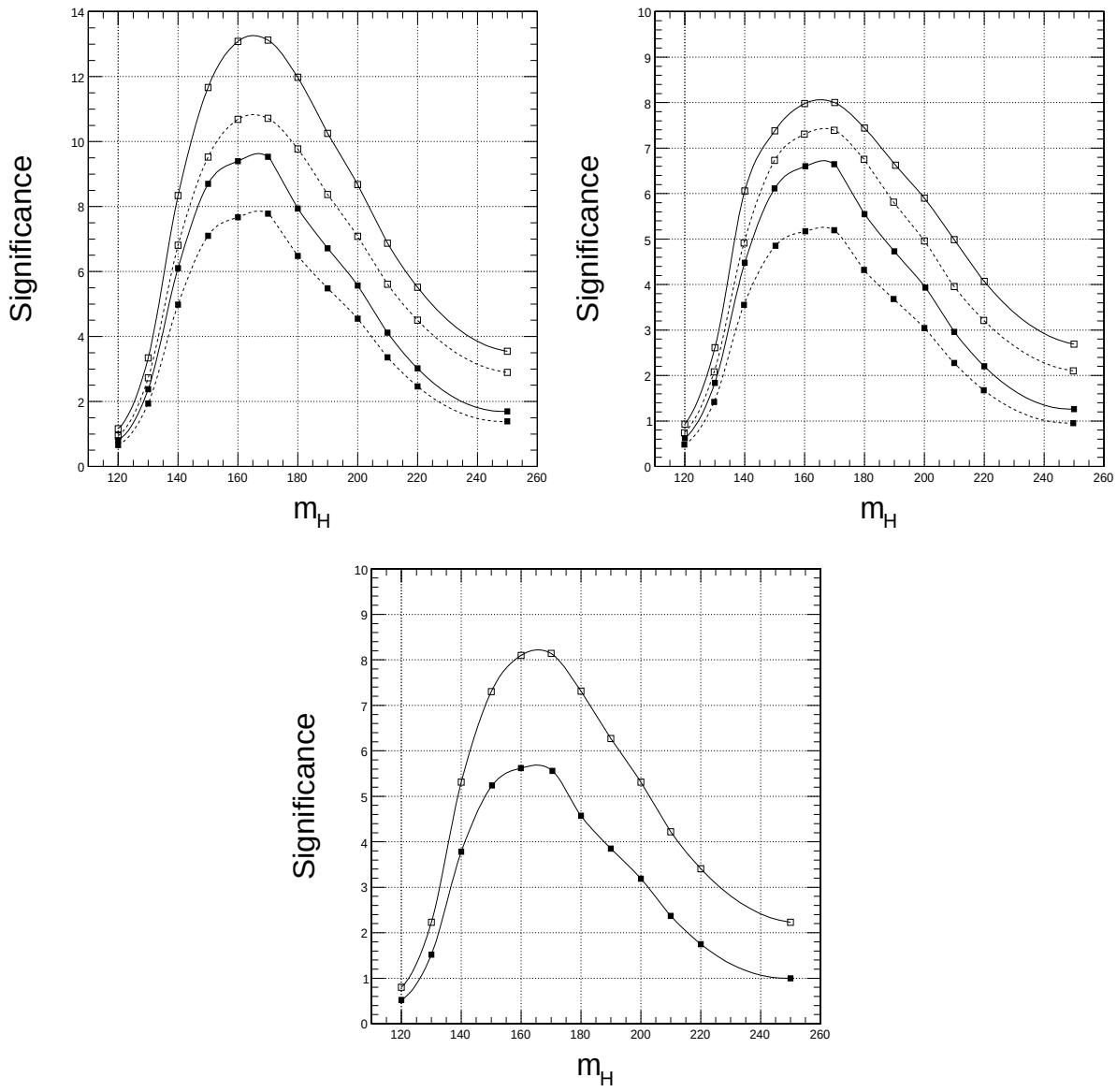


Figure 33: significance verse $\mathrm{m_H}$ based on $\mathrm{S_c}$ at 30 fb^{-1} (upper left), significance verse $\mathrm{m_H}$ based on $\mathrm{S_{cP}}$ at 30 fb^{-1} (upper right) with estimated 10% detector systematic uncertainty taken into the $\mathrm{S_{cP}}$ calculation, and significance verse $\mathrm{m_H}$ based on $\mathrm{S_{cP}}$ at 30 fb^{-1} (bottom) with conservatively estimated 16% detector systematic uncertainty taken into the $\mathrm{S_{cP}}$ calculation. The high (low) solid curve corresponds to Extra Jet Veto (Loose Extra Jet Veto) Scheme. The dashed curves correspond to Extra Jet Veto and Loose Extra Jet Veto Scheme with using 50% more average background events ("pessimistic" estimation of the background).

Table 12: Cross section (fb) of the signal and background under optimized selection cuts with $\mathrm{m_H} \geq 160\ \mathrm{GeV/c^2}$ for Extra Jet Veto (E) and Loose Extra Jet Veto Scheme (L)

Channels	S-1 (L)	S-2 (L)	S-3 (L)	S-1 (E)	S-2 (E)	S-3 (E)
VBF Higgs ($\mathrm{m_H}$=160)	7.639	5.482	2.564	9.531	4.580	2.989
VBF Higgs ($\mathrm{m_H}$=170)	8.099	5.730	2.600	9.814	4.828	3.006
VBF Higgs ($\mathrm{m_H}$=180)	8.006	5.635	2.165	9.916	4.711	2.738
VBF Higgs ($\mathrm{m_H}$=190)	7.365	5.256	1.831	9.363	4.294	2.340
VBF Higgs ($\mathrm{m_H}$=200)	6.963	5.145	1.520	8.626	4.341	1.983
VBF Higgs ($\mathrm{m_H}$=210)	6.467	4.794	1.122	8.211	4.080	1.571
VBF Higgs ($\mathrm{m_H}$=220)	6.655	4.847	0.824	8.227	4.128	1.259
VBF Higgs ($\mathrm{m_H}$=250)	5.463	3.982	0.463	6.900	3.426	0.810
$\mathrm{t\bar{t}}$ + jets	413.1	67.496	1.478	626.5	16.751	1.232
WW + jets (QCD)	0.843	0.843	< 0.008	1.265	0.422	< 0.008
WW + jets (EW)	7.747	6.170	0.0277	9.683	4.454	< 0.0277
ZZ + jets	0.171	0.098	< 0.001	0.269	0.0245	< 0.001
ZW + jets	1.668	0.667	< 0.001	2.335	0.223	< 0.001
$\mathrm{W + t\bar{b}(\bar{t}b)}$	20.745	10.821	0.05787	35.21	4.427	< 0.05787
W + 4j ($\mathrm{W \rightarrow e/\mu/\tau + \nu}$)	388.5	176.8	0.6463	583.0	72.066	0.323
W + 3j ($\mathrm{W \rightarrow e/\mu/\tau + \nu}$)	142.8	86.1	< 0.3147	228.2	68.633	< 0.3147
Z + 4j ($\mathrm{Z \rightarrow ee/\mu\mu}$)	32.804	8.250	0.012	43.443	1.650	0.0122
Z + 3j ($\mathrm{Z \rightarrow ee/\mu\mu}$)	22.507	8.629	< 0.0095	32.383	3.263	< 0.0094
Sum of Background	1029.7	365.87	2.232	1562.3	171.92	1.567

Table 13: Cross section (fb) of signal and background under optimized selection cuts with $\mathrm{m_H} < 160\ \mathrm{GeV/c^2}$ for Extra Jet Veto (E) and Loose Extra Jet Veto Scheme (L)

Channels	S-1 (L)	S-2 (L)	S-3 (L)	S-1 (E)	S-2 (E)	S-3 (E)
VBF Higgs ($\mathrm{m_H}$=120)	0.711	0.447	0.184	0.951	0.363	0.231
VBF Higgs ($\mathrm{m_H}$=130)	2.280	1.306	0.536	3.004	1.125	0.664
VBF Higgs ($\mathrm{m_H}$=140)	4.405	2.898	1.380	5.520	2.369	1.656
VBF Higgs ($\mathrm{m_H}$=150)	6.556	4.224	1.965	8.345	3.505	2.317
$\mathrm{t\bar{t}}$ + jets	555.2	84.49	0.739	859.5	20.94	0.493
WW + jets (QCD)	2.951	0.422	< 0.004	4.215	0.422	< 0.004
WW + jets (EW)	8.770	7.110	0.0277	11.21	5.395	< 0.0277
ZZ + jets	0.294	0.0979	< 0.001	0.465	0.0979	< 0.001
ZW + jets	2.557	0.900	< 0.01	3.781	0.334	< 0.01
$\mathrm{W + t\bar{b}(\bar{t}b)}$	33.187	16.03	0.0868	54.37	6.799	< 0.0289
W + 4j ($\mathrm{W \rightarrow e/\mu/\tau + \nu}$)	520.0	264.7	0.6463	778.5	118.9	0.667
W + 3j ($\mathrm{W \rightarrow e/\mu/\tau + \nu}$)	218.6	146.5	< 0.343	346.6	113.9	< 0.3147
Z + 4j ($\mathrm{Z \rightarrow ee/\mu\mu}$)	40.45	10.68	0.0122	54.99	1.810	0.0222
Z + 3j ($\mathrm{Z \rightarrow ee/\mu\mu}$)	30.98	14.94	0.0186	44.24	4.942	< 0.01333
Sum of Background	1414.2	546.5	1.532	2157.9	273.5	1.181

8.3 Higgs Boson Mass and Distribution in Signal Events

After the selection, the Higgs mass distribution of signal events is illustrated in Fig. 34. Using projected background, the overall reconstruction results are illustrated in Fig. 35.

Under the Conservative Scenario and Loose Extra Jet Veto Scheme, the number of background events at $60\mathrm{fb}^{-1}$ is estimated as ~ 133. An estimation of the background shape in the mass distribution is performed by using a loose $\mathrm{E_T^{miss}}$ cut of qqWW system ($\mathrm{E_T^{miss}} < 125$ GeV) in the third step of optimization, so as to get more statistics from the background. Other selection cuts are the same.

A loose $\mathrm{E_T^{miss}}$ cut instead of other cuts (e.g. ΔR between lepton and hadronic W, ΔR between hadronic and leptonic W) is used because $\mathrm{E_T^{miss}}$ is highly related to extra jet activities with little impact on the reconstructed di-W system. So the $\mathrm{E_T^{miss}}$ cut will not significantly change the reconstructed Higgs boson mass distribution for background and signal events. Loosing the forward jet tagging criterion (e.g., di-jet mass and di-jet η distance) also provides a way to get more statistics without heavily influencing the di-W system.

Due to the change of the selection criterion, the projection should be taken as a first-order approximation of the background distribution. A signal-like bump in the background can be observed.

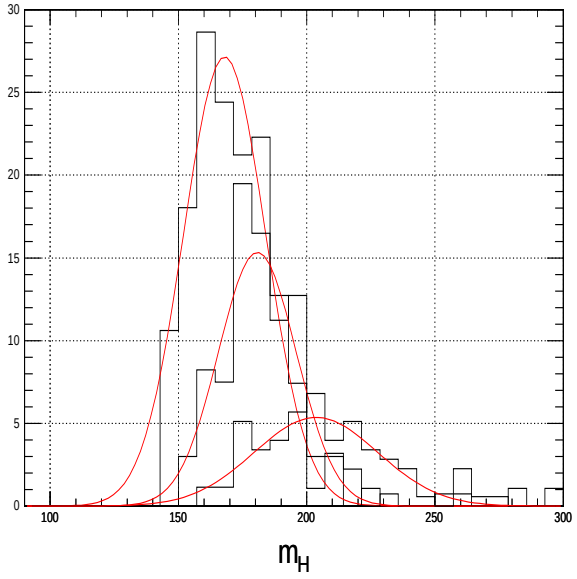


Figure 34: Fitting of VBF Higgs mass distribution using signal events only. From left to right $\mathrm{m_H}$ of 160 (left), 190 (middle), and 220 (right) $\mathrm{GeV/c^2}$ give σ of 14.1, 15.5, and 23.9 $\mathrm{GeV/c^2}$ respectively.

8.4 Background Shape in Higgs Mass Distribution

In the following, a quantitative estimation of the background shape in reconstructed Higgs boson mass distribution as a function of $\mathrm{E_T^{miss}}$ is performed by using four regions of reconstructed $\mathrm{m_H}$:

- Region A: $\mathrm{m_H} < 186\ \mathrm{GeV/c^2}$
- Region B: $186 < \mathrm{m_H} < 203\ \mathrm{GeV/c^2}$
- Region C: $203 < \mathrm{m_H} < 228\ \mathrm{GeV/c^2}$
- Region D: $228 < \mathrm{m_H} < 262\ \mathrm{GeV/c^2}$

For a given $\mathrm{E_T^{miss}}$ cut, total numbers of events in these regions and fraction of events in each region are calculated for the signal ($\mathrm{m_H} = 170\ \mathrm{GeV/c^2}$) and backgrounds respectively (Fig. 36):

- For the signal, the shape of the Higgs boson mass distribution is very stable, because the fraction of events in each region does not changes with $\mathrm{E_T^{miss}}$ cut. Region A dominates the mass distribution which accounts for $\sim$ 2/3 of total events.
- For background events, the change mainly occurs with $\mathrm{E_T^{miss}} < 80$ GeV. There is a roughly 10% difference in the fraction of events for Region A and C between the low and high $\mathrm{E_T^{miss}}$ cut. Tightening the $\mathrm{E_T^{miss}}$ cut increases (decreases) the relative contribution of Region C (A). Between Region A and C, Region B with width of 17 GeV ($\sim$ Higgs boson mass resolution of the signal events) gets very little influence from the $\mathrm{E_T^{miss}}$ cut.

We conclude that the background events are more widely distributed in different regions. A much lower bump from the background than the signal events is expected, which mainly lies on the tail of the signal's peak distribution.

9 Experimental Identification of VBF Higgs Boson Signature

This section addresses an issue in the experimental reconstruction and analysis: how to identify the existence of Higgs boson signal and make the result less dependent on the uncertainty of reconstruction efficiency for different background processes. The major result of the reconstruction is the Higgs mass distribution. Once a peak is reconstructed experimentally, it is non-trivial to confirm that it is a real signature of the signal and not the faked

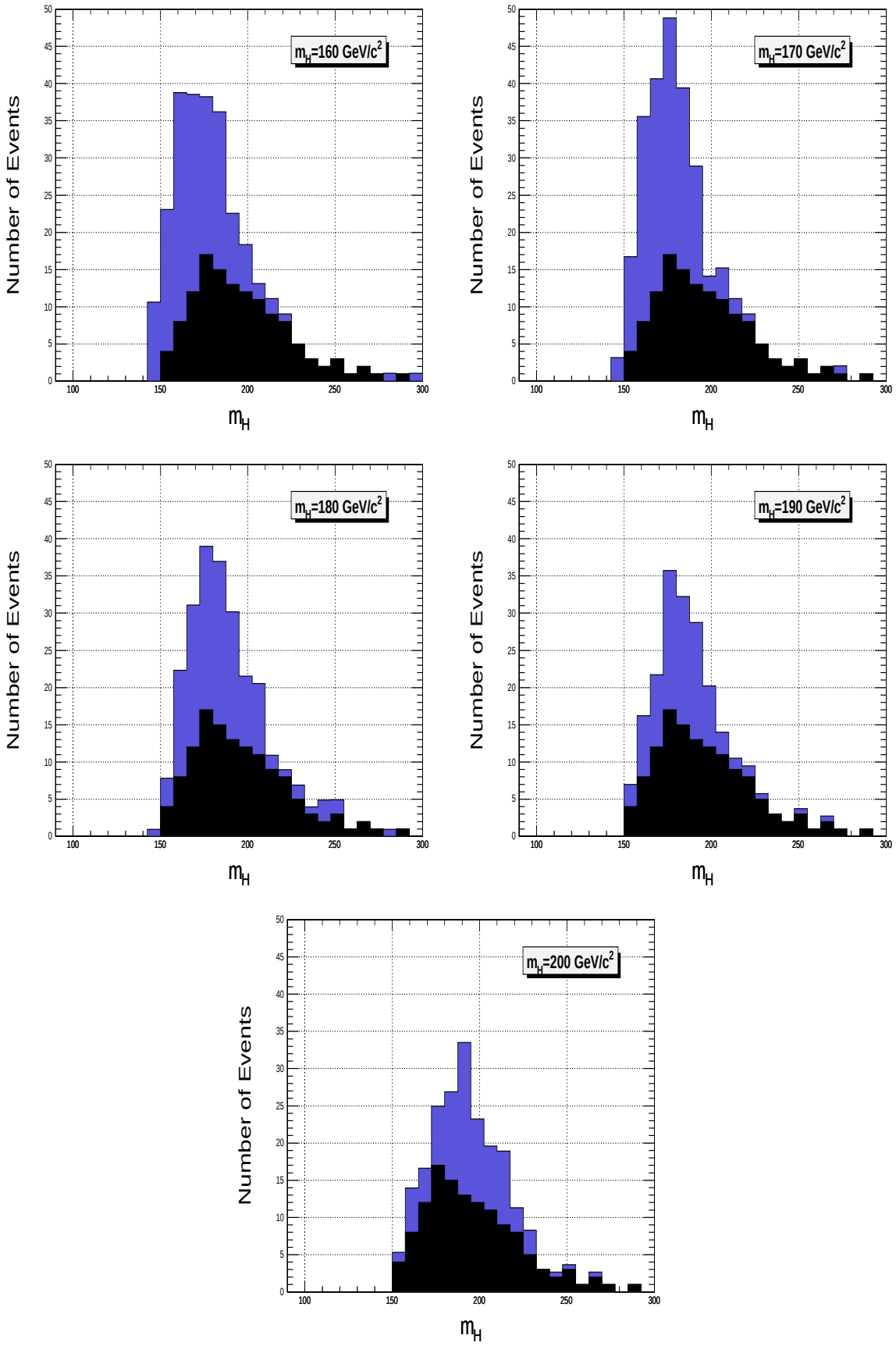


Figure 35: Reconstructed Higgs mass distribution from signal (blue) and projected background (black) for $\mathrm{m_H} = 160$ (upper left), $\mathrm{m_H} = 170$ (upper right), $\mathrm{m_H} = 180$ (middle left), $\mathrm{m_H} = 190$ (middle right), and $\mathrm{m_H} = 200$ $\mathrm{GeV/c^2}$ (bottom)

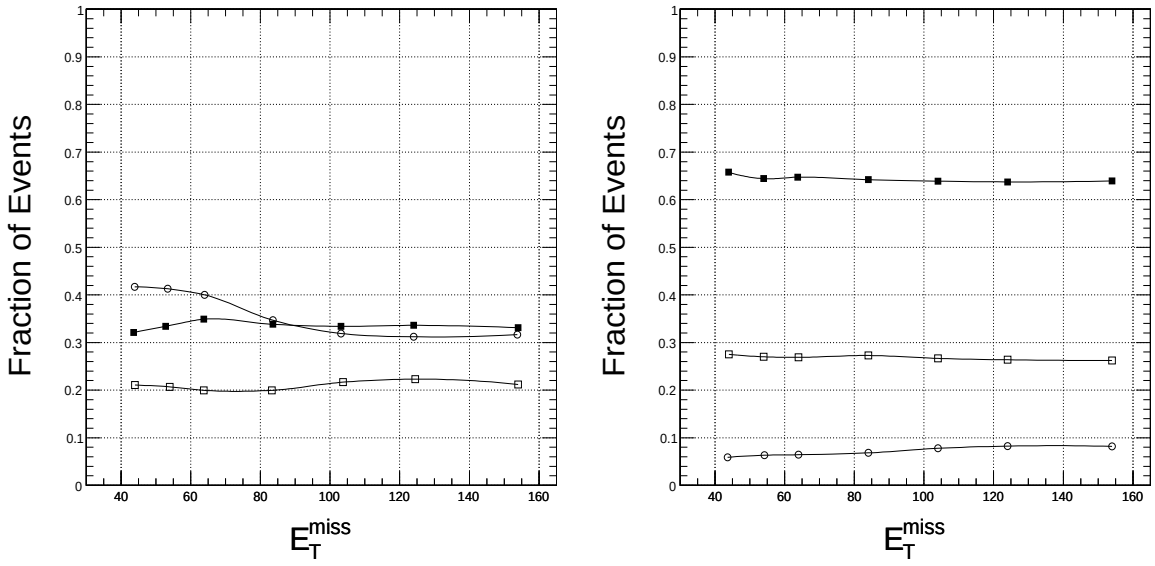


Figure 36: The fraction of events in different regions for the overall background (left) and VBF Higgs signal (right) as a function of $\mathrm{E_T^{miss}}$ cuts. Region A (close square), Region B (open square), and Region C (open circle).

signature of backgrounds due to the "fluctuation" of the selection efficiency or under-estimation of background cross section. In general, the background cross section can be measured in good precision (discussed in Section 2), so the former factor has a bigger impact on the final result. A long selection chain and hard kinematic cuts contain potentially large systematic uncertainties.

In additional to an accurate event generation and detector simulation, some extra signature of singal events can be extracted from data, which is the major task of the experimental approach to confirm the existence of VBF Higgs signal without requiring an accurate knowledge of reconstruction and selection efficiency. It is expected that intensively using detector data to estimate the reconstruction and selection efficiency will play an important role for VBF Higgs study.

Two types of signatures of VBF Higgs are discussed in the following, which help experimentally resolve the ambiguity of the origin of the reconstructed Higgs mass peak. The general strategy is based on using two uncorrelated selection cuts to which both signal and background show good sensitivity. We can establish a two-dimensional phase space with these two selection cuts and divide the space into several non-overlapped regions. For most regions, background events dominate (called background region), while for some regions, signal events will be significant (called signal region). A significant excess of events in the signal region provides an experimental evidence of the existence of VBF Higgs. The background normalization method can be heavily exploited, since detector reconstruction and selection efficiency can be determined and measured in background region and extrapolated to the signal region if the hypothesis of the signal leads to a large deviation from the predicted number based on background only.

9.1 Signature of $\mathrm{E_T^{miss}}$ in qqWW System

The selection cut based on $\mathrm{E_T^{miss}}$ in qqWW system shown in previous section provides a strong suppression of background. The distinct distribution of $\mathrm{E_T^{miss}}$ in qqWW system between the signal and backgrounds make it a valuable experimental signature for this analysis. To identify the excess of signal events, a phase space is formed by $\mathrm{E_T^{miss}}$ of qqWW system with three regions:

- Region A: $\mathrm{E_T^{miss}} < 40$ GeV.
- Region B: $40 < \mathrm{E_T^{miss}} < 55$ GeV
- Region C: $55 < \mathrm{E_T^{miss}} < 65$ GeV

The numbers of events in each region are defined as $\mathrm{N_A}$, $\mathrm{N_B}$ and $\mathrm{N_C}$ respectively. A loose selection cuts based on $\Delta\mathrm{R}$ between the lepton and hadronic W is applied:

- $\Delta R = 0.6$, 0.9, 1.2, 1.5, 1.8, 2.1, 2.4, 2.7, 3.0, 3.5, 4.0, 5.0, and 6.0.
- Other selection cuts except E_T^{miss} and ΔR are the same as those optimized ones described in previous section with Loose Extra Jet Veto Scheme.

Using the phase space defined above, we can investigate the correlation of the number of events in each region with respect to various values of ΔR cut. This correlation is mainly affected by whether there is the existence of VBF Higgs signal (called Scenario of Signal + Background) or just the background (called Scenario of Background Only). The selection efficiency will play a less important role in the correlation. Region A will be influenced by VBF Higgs signal, while the rest two regions get much less influence.

Fig. 37 shows the ratio of the number of events between each of two regions ($R_{AB} = N_A/N_B$, $R_{AC} = N_A/N_C$, and $R_{BC} = N_B/N_C$). R_{AB} and R_{AC} are highly affected by the existence of VBF Higgs between two scenarios, while there is almost no change in R_{BC}.

For R_{AC} as shown Fig. 37, the value from Background Only Scenario is about 0.8, but the Signal + Background Scenario can give a much larger ratio which can be identified easily. The scale of the ratios (defined as R^{s+b}/R^b, where R^{s+b} is the ratio of Signal + Background Scenario, R^b is the ratio of Background Only Scenario) between two scenarios are illustrated in Fig. 38, which shows the excess in the low E_T^{miss} (Region A) causes significant increase of the scale as ΔR goes lower. In general, ratios between different regions defined by E_T^{miss} provide a good probe of VBF Higgs signature.

The experimental measurement of these quantities combined with Higgs boson mass peak can also be compared to the prediction of Monte Carlo simulation and reconstruction.

9.2 Signature of Lepton-W ΔR

In this method, two parameters (selection cuts): ΔR between lepton and hadronic W and di-jet mass (m_{qq}) in the forward jet tagging selection, are used to define the two-dimensional phase space. Other selection cuts are the same as those of optimized ones with Loose Extra Jet Veto Scheme.

- Region A (background dominates): $\Delta R < 1.8$ and $800 < m_{qq} < 1200\ GeV/c^2$
- Region B (signal + background): $\Delta R < 1.8$ and $m_{qq} > 1200\ GeV/c^2$
- Region C (background dominates): $1.8 < \Delta R < 2.6$ and $800 < m_{qq} < 1200\ GeV/c^2$
- Region D (background dominates): $1.8 < \Delta R < 2.6$ and $< m_{qq} > 1200\ GeV/c^2$

The numbers of events in each region (defined as N_A, N_B, N_C, and N_D) under two scenarios: Signal + Background and Background Only are included in Table 14.

Table 14: Summary of number of events in region A, B, C, and D with respect to Signal + Background and Background Only scenarios ($m_H = 170\ GeV/c^2$)

Region	N^{s+b}	N^b	Number of signal events
A	551	482	69
B	387	213	174
C	644	619	25
D	353	304	49

Using N_A, N_C and N_D with Signal + Background Scenario, the projected N_B is $\sim$ 302 events and we "observe" 387 events, which makes an excess of 85 events with significance of $85/\sqrt{302} = 4.89$. This shows the evidence of the possible existence of VBF Higgs signal in Region B.

10 Estimation of Selected Systematic Uncertainties

10.1 Detector Systematic Uncertainty

Systematic uncertainties in the jet reconstruction and selection are important issues in the analysis. These factors also largely determines the systematic uncertainty of E_T^{miss}. Here we mainly consider two systematics about jets:

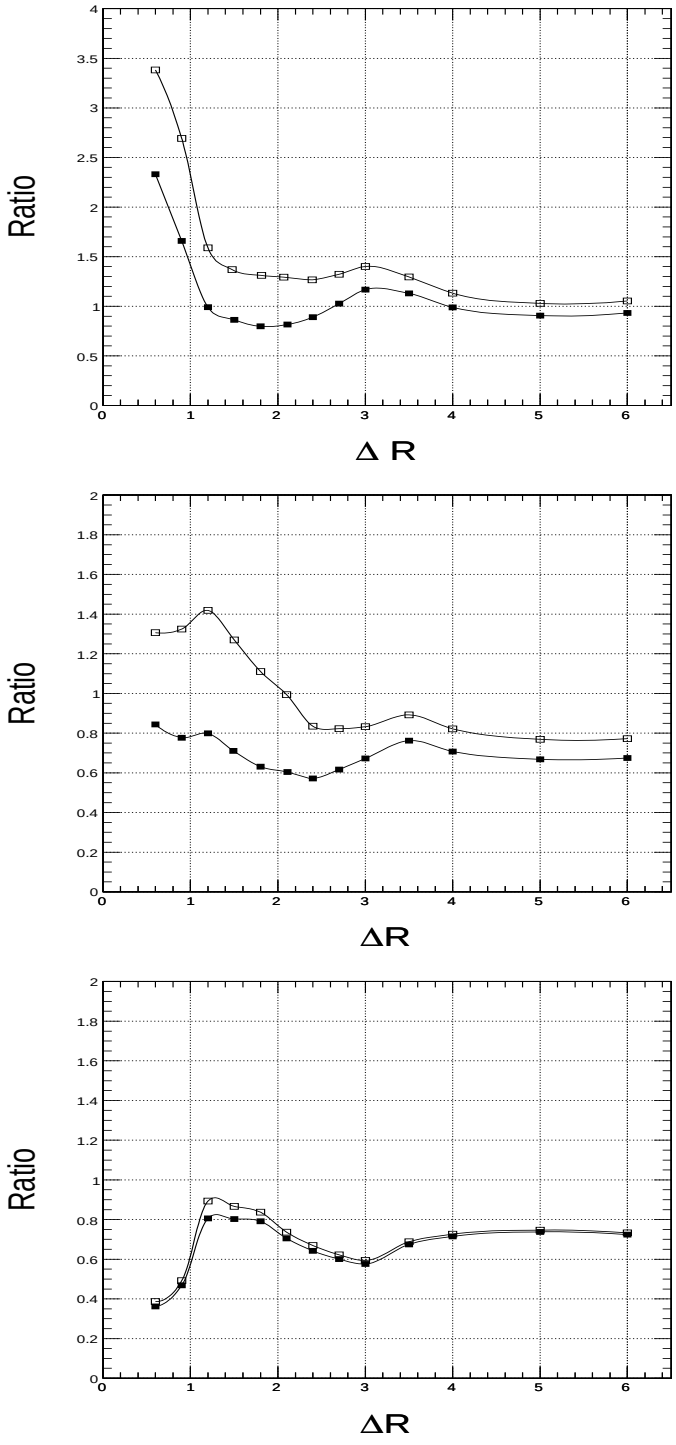


Figure 37: The ratio of number of events as a function of ΔR: Region A to Region B (upper), Region A to Region C (middle), and Region B to Region C (bottom). Two scenarios are illustrated: Signal + Background (open square) and Background Only (solid square) respectively.

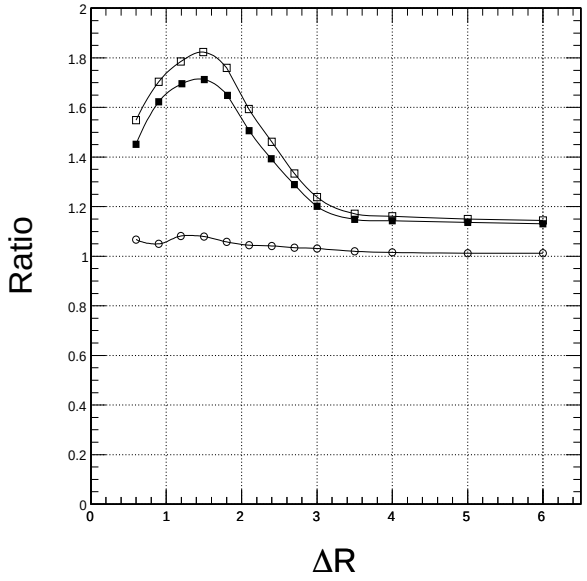


Figure 38: The ratio of Signal + Background Scenario to Background Only Scenario as a function of ΔR for Region A to Region B (open square), Region A to Region C (solid square), and Region B to Region C (open circle)

jet energy scale and jet energy resolution. The first one can be determined by using several dedicated physics channel (e.g., γ + jet and Z + jet) which leads to less than 3 % uncertainty for jet $\mathrm{E_T} > 50$ GeV due to plenty of events can be taken experimentally. The bias of jet energy scale can be offset by tuning the offline jet energy threshold. We changed the jet energy scale by ± 0.03 for jet $\mathrm{E_T} > 50$ GeV and $\pm[0.1 - 0.07(\mathrm{E_T} - 20)/30)$ for jets with $\mathrm{E_T}$ between 20 and 50 GeV (based on the jet energy scale has a systematic uncertainty of $\sim$10% for $\mathrm{E_T} \sim$ 20 GeV). We found the S/B changes in $\pm 6\%$ and B(S) rate changes in $\sim \pm 14.5(10.6)\%$.

In the following, the discussion focuses on the second factor, especially if the experimental jet energy resolution is not as good as the simulated one, an estimation of the sensitivity of VBF Higgs discovery potential to jet energy resolution is critical. It should be emphasized that the absolute jet energy resolution in the off-line reconstruction directly relates to the intrinsic detector resolution. We introduce a smearing factor ($\mathrm{F_{smear}}$) worsening jet $\sigma(\mathrm{E_T})$ to study the impact of the jet energy resolution on the reconstruction and selection efficiency (Eq. 4). The $\mathrm{F_{smear}}$ of 0.14, 0.28, 0.41, 0.50, 0.60, and 0.67 corresponds to worsening $\sigma(\mathrm{E_T})$ by 2%, 5%, 10%, 15%, 20% and 25% respectively. The $\mathrm{E_T^{miss}}$ is re-calculated event by event accordingly.

$$\mathrm{E_T^{smear}} = \mathrm{E_T}(1 + \mathrm{Gaus}(0.0, \mathrm{F_{smear}} * \sigma(\mathrm{E_T}))) \tag{4}$$

where Gaus(0.0, σ) is a randomly thrown sampling of a normal distribution with a mean of zero and a width of σ. The impact of the smearing is illustrated in Fig. 39. The $\mathrm{t\bar{t}}$ + jets process significantly benefits from a worse jet energy resolution, because low $\mathrm{E_T}$ jets in $\mathrm{t\bar{t}}$ events can be mis-measured and increase the selection efficiency based on a fixed jet $\mathrm{E_T}$ threshold. In signal events, the jet activities are suppressed, so the influence of jet energy resolution is much smaller. Although it shows that a large unexpected uncertainties can be introduced by jet energy resolution, extra jet veto technique largely suppresses the background events with soft jets.

Current detector simulation is consistent with the results from Test Beam data, it is anticipated that the ultimate detector jet energy resolution largely determined by stochastic effect of hadron response in the HCAL, will be close to that of the simulation ($< 10\%$), which introduce 2-3 % systematic uncertainties in the S/B.

We studied the standalone systematic effects of $\mathrm{E_T^{miss}}$ scale uncertainty by changing $\mathrm{E_T^{miss}}$ with $\pm 5\%$. The S/B varies within $\pm 2.5\%$ and B(S) rate changes in $\pm 1.7\%$. The systematic error of lepton isolation is studied similar to that of jet energy scale, a smaller change of B(S) rate $\pm 1.3(1.4)\%$ and S/B $\pm 0.5\%$ are observed when varies the isolation cone energy by $\pm 5\%$.

In a summary, we predict that jet and $\mathrm{E_T^{miss}}$ measurement causes $\sim$ 5.9% systematic uncertainty in S/B, $\sim$ 14.8% in

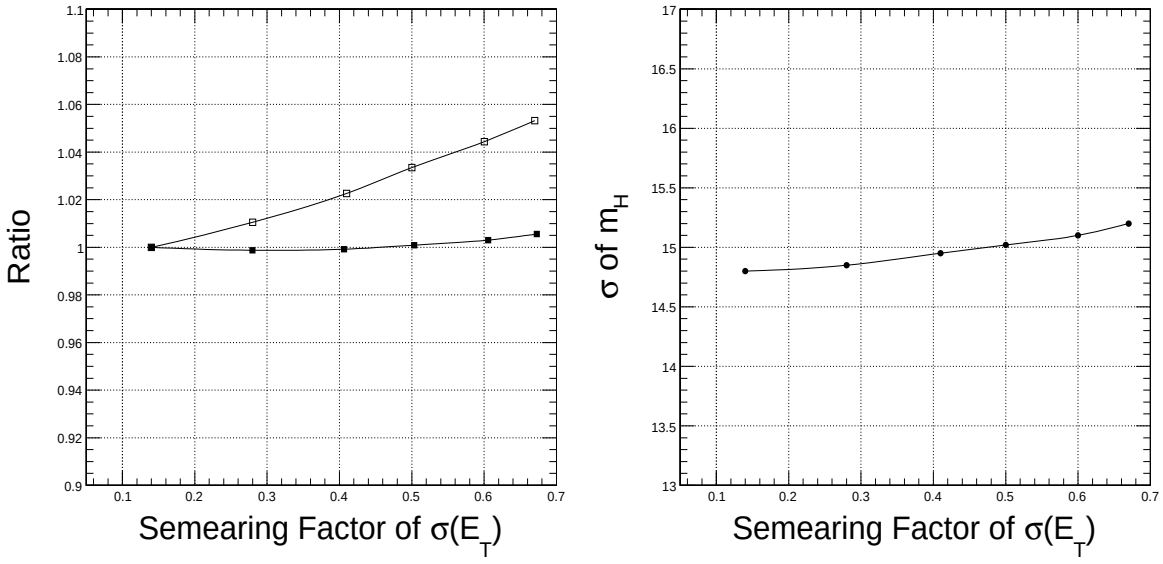


Figure 39: Basic selection efficiency (normalized to non-smeared rate of signal and background) as a function of jet resolution factor (left) and reconstructed Higgs boson mass resolution from signal events after basic filtering as a function of jet resolution factor (right). $\mathrm{t\bar{t}}$ + jets background (open square) and VBF Higgs signal (solid square)

absolute rate of B and $\sim 11\%$ in absolute rate of S (Table 15). If considering other detector systematic uncertainties of luminosity, high level trigger, and lepton $\mathrm{p_T}$ measurement. The total uncertainty is 6.5% in S/B, 16% in B and 6.3% in S/B.

Table 15: Systematic Uncertainties due to Jet and $\mathrm{E_T^{miss}}$

Source	S	B	S/B
Jet energy scale	10.6%	14.5%	5.2%
Jet energy resolution	0.1%	2.0%	2.0%
$\mathrm{E_T^{miss}}$	2.5%	1.2%	1.7%
Lepton isolation	1.4%	1.3%	0.5%

The uncertainty of jet energy scale leads to significant change in the absolute rate. For an off-line strategy that needs to control the overall systematic error, a loose cut can be used to find the targeted rate which is dominated by the background instead of using a targeted jet $\mathrm{E_T}$ threshold. In early days of the experiment we expect that the jet $\mathrm{E_T}$ threshold is less reliable due to uncertainties of the calibration and luminosity. However as a far as S/B is stable for a range of jet $\mathrm{E_T}$ threshold, we can use the measured event rate to adjust the threshold and normalize the MC prediction, so that the final result will be much less dependent on the systematic effect of jet energy scale. If the lowest jet $\mathrm{E_T}$ threshold can be "measured" from data (in our study, a 25 GeV $\mathrm{E_T}$ threshold is used for jet counting and rejecting extra jet), the total variance in the absolute rate of background due to the uncertainty of jet energy scale is $\sim 5.5\%$. So the data driven method can potentially reduce the total detector systematic uncertainty from 16% to 10%. In the calculation of significance, we use 10% and 16% (a conservative estimation) for the total detector systematic uncertainty respectively.

10.2 Systematic Uncertainty in Event Generation

In the event generation, systematic uncertainties will arise from various configurations of parton distribution function, QCD normalization and factorization scale, initial state radiation (ISR) and final state radiation (FSR), underlying event (UE) and etc.

This section summarizes several outstanding systematic issues in the theoretical side (generator level): the effect of ISR and FSR, effect of underlying event (UE) model. W + 3 jets and W + 4 jets are mainly used for this study as benchmark processes.

Following scenarios in the event generation other than the standard one are considered:

- UE configuration with two $\mathrm{p_T}$ thresholds for the multiple parton scattering cutoff: 1.9 GeV/c (default) and 2.9 GeV/c .
- ISR is switched off.
- FSR is switched off.
- ISR and FSR are switched off.

Other unspecified parameters in the event generation are the same as standard ones. Different configuration scenarios causes significant changes in the rate of events passing various selection cuts in the basic filtering (discussed in Section 5): E-S, FJT and H-W. Based on the results from standard samples, the number of events for those scenarios are compared with its ratio to the standard one summarized in Table 16 and 17.

Table 16: Selection efficiency of W + 3jets with various configuration scenarios to the standard one

Selection Cut	UE	No ISR	No FSR	No ISR and FSR
E-S	0.847	0.840	0.425	0.187
FJT	0.755	0.659	0.599	0.216
H-W	0.808	0.759	0.367	0.208

Table 17: Selection efficiency of W + 4jets with various configuration scenarios to the standard one

Selection Cut	UE	No ISR	No FSR	No ISR and FSR
E-S	0.965	0.980	1.065	1.021
FJT	0.917	0.886	1.470	1.254
H-W	0.890	0.922	0.651	0.544

Some interesting effects are found in W + 4jets when the ISR and/or FSR are switched off: the event selection rate in some step even get enhanced, but the overall efficiency after hadronic W reconstruction receives a significant reduction. This effect is because FSR normally smears the jet $\mathrm{E_T}$ spectrum and make less events pass the $\mathrm{E_T}$ threshold. If FSR is switched off, a harder jet $\mathrm{E_T}$ spectrum cause more events pass the threshold. But switching off FSR also results in lower probability of getting a pair of jets with invariant mass within the W mass selection window, which turns out to make a stronger effect.

The scalar sum of total transverse energy ($\Sigma\mathrm{E_T}$) is another detector observable highly related to the UE, ISR/FSR configuration, which also directly influence the lepton isolation cut efficiency and jet energy scale. Average $\Sigma\mathrm{E_T}$ for W + 3jets and W + 4jets with different configurations are included in Table 18.

Table 18: Average $\Sigma\mathrm{E_T}$ of W + 3jets and W + 4jets with different configuration scenarios

Channel	UE	No ISR	No FSR	No ISR and FSR	Standard
W + 3jets	399.3	483.1	482.5	465.4	498.4
W + 4jets	537.9	620.4	608.9	595.4	634.5

Among different configurations, the variance in $\Sigma\mathrm{E_T}$ is $\sim$ 100 GeV, which roughly corresponds to 0.2-0.5 (0.5-1.0) GeV of total transverse energy in a 0.2 (0.2-0.4) isolation region, the influence on the isolation efficiency of 30 GeV/c lepton is less than 3%. For jet energy scale, the fluctuation of jet $\mathrm{E_T}$ with 0.6 cone size is $\sim$ 1-2 GeV, which is 5-10 % systematic effects in the jet energy scale for jet $\mathrm{E_T} < 50$ GeV. After LHC takes data, all these effects will be well measured and will not have significant impact on the selection efficiency after tuning the selection cuts with experimental data.

11 Summary

The signal topology of Higgs boson with $\mathrm{H} \rightarrow \mathrm{W^+W^-} \rightarrow \ell\nu\mathrm{jj}$ via vector boson fusion was studied for $\mathrm{m_H}$ from 120 to 250 $\mathrm{GeV/c^2}$. In the mass range between 140 and 200 $\mathrm{GeV/c^2}$, a 5 σ significance of Higgs boson discovery can be achieved at integrated luminosity of 30 $\mathrm{fb^{-1}}$ using extra jet veto selection scheme.

We carried out a comprehensive study of calorimeter-based lepton isolation and jet selection strategies. The reconstruction and selection chain was optimized to reduce the detector systematic effects and enhance the Higgs boson

signal. The estimation of background includes correlated processes of W + jets and Z + jets, making the discovery potential conservative. Better significance can be achieved with more statistics of the major background channels ($\mathrm{t\bar{t}}$ + jets and W + jets), so that the optimized selection cuts can be applied more effectively with fewer systematic uncertainties.

We found that several selection criteria were highly effective in suppressing backgrounds: forward jet tagging, extra jet veto, $\mathrm{E_T^{miss}}$ in qqWW system, $\Delta\mathrm{R}$ between lepton and hadronic W, and $\Delta\mathrm{R}$ between di-W. The detector $\mathrm{E_T^{miss}}$ resolution and jet energy resolution were determined to be critical to the quality of reconstructed Higgs boson. Because the low value of $\mathrm{m_H}$ in this analysis leads to low $\mathrm{E_T^{miss}}$ and jet $\mathrm{E_T}$ spectra, an effective background suppression relies on heavy exploitation of the physics signature of the signal.

We carried out a data analysis approach using $\mathrm{E_T^{miss}}$ in qqWW system and $\Delta\mathrm{R}$ between lepton and hadronic W. In addition to Higgs boson mass as the major signature, extra signatures can be effectively extracted and used to resolve the ambiguity of Higgs boson mass peak without requiring very accurate knowledge of selection efficiency.

Most of the detector systematic uncertainties relate to the jet energy scale and resolution. The total detector systematic uncertainty is expected as 10-16%. However, since S/B > 1 was achieved for the most interesting Higgs boson mass range, the estimated systematic effect will not significantly influence the discovery potential.

12 Acknowledgments

We would like to thank Albert De Roeck, Alexandre Nikitenko, Dan Green, and Yves Sirois for those wonderful discussion and suggestions on this analysis that lead to a better quality of the paper and physics result.

References

[1] M.Zeyrek, W.Wu and etc., *120-180 GeV Higgs in qq -> qqH Channel at CMS*, CMS NOTE 2002/066

[2] M.Dittmar and H.Dreiner, *LHC Higgs Search with l+nul-nubar Final States*, CMS NOTE 1997/083

[3] T.Sjostrand, Comput.Phys.Commun. **135**, 238 (2001)

[4] M.Mangano, M.Moretti, F.Piccinini, R.Pittau and A.Polosa, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, hep-ph/0206293, (2002)

[5] F.Maltoni and T.Stelzer, *MADEVENT: AUTOMATIC EVENT GENERATION WITH MADGRAPH*, hep-ph/0208156, (2002)

[6] A.Pukhov, E.Boos, M.Dubinin, V.Edneral, V.Ilyin, D.Kovalenko, A.Kryukov, V.Savrin, S.Shichanin, and A.Semenov, *CompHEP - a package for evaluation of Feynman diagrams and integration over multi-particle phase space*, hep-ph/9908288, (1999)

[7] S.Catani, F.Krauss, R.Kuhn and B.Webber, *QCD Matrix Elements + Parton Showers*, hep-ph/0109231, (2001)

[8] N. Kidonakis and and R. Vogt, *Next-to-next-to-leading order soft-gluon corrections in top quark hadroproduction*, Phys. Rev. D **68**, 114014 (2003)

[9] D. Rainwater, M. Spira, and D. Zeppenfeld, *Higgs Boson Production at Hadron Colliders: Signal and Background Processes*, hep-ph/0203187, (2002)

[10] CMS Collaboration, *CMS Kinematics Interface Package*, http://cmsdoc.cern.ch/cmsim/cmsim.html

[11] CMS Collaboration, *Object oriented Simulation for CMS Analysis and Reconstruction*, http://cmsdoc.cern.ch/oscar

[12] CMS Collaboration, *Object oriented Reconstruction for CMS Analysis*, http://cmsdoc.cern.ch/orca

[13] CMS Collaboration, *CMS Fast Simulation*, http://cmsdoc.cern.ch/famos

[14] A. Heister, O. Kodolova, V. Konoplianikov, S. Petrushanko, J. Rohlf, C. Tully and A. Ulyanov, *Measurement of Jets with the CMS Detector at the LHC*, CMS Note 2006/036

[15] H.Pi, P.Avery, D.Green, J.Rohlf, and C.Tully, *Measurement of Missing Transverse Energy with the CMS Detector at the LHC*, CMS Note 2006/035, (2006)

[16] S. Baffioni, C. Charlot, F. Ferri, D. Futyan, P. Meridiani, I. Puljak, C. Rovelli, R. Salerno and Y. Sirois, *Electron reconstruction in CMS*, CMS Note 2006/040

[17] E.James, Y. Maravin, M. Mulders and N. Neumeister, *Muon Identification in CMS*, CMS Note 2006/010

[18] CMS Collaboration, *The TriDAS Project Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger*, CERN/LHCC 2002/26

[19] S.Bityukov et al., http://cmsdoc.cern.ch/∼bityukov

Appendix 1: Lepton Isolation Strategy

Isolation cuts for electrons:

- $E_T^{Hcal}/E_T^{Ecal} < 0.05$, where the E_T^{Hcal} and E_T^{Ecal} are calculated from the 0.2 isolation cone around electron super-cluster in Hadronic Calorimeter (HCAL) and Electromagnetic Calorimeter (ECAL) respectively (Fig. 40).
- $0.9 < E/p < 1.8$, where E and p are the energy of electron super-cluster measured in ECAL and track momentum measured in Tracker (Fig. 41).
- $|E_T^{0.2} - E_T^e| < 5.0$ GeV and $|(E_T^{0.2} - E_T^e)/E_T^e| < 0.3$, where $E_T^{0.2}$ is the total E_T in the 0.2 isolation cone and E_T^e is the electron super-cluster E_T (Fig. 42 and 43).
- $E_T^{0.4}/E_T^e < 0.3$, where $E_T^{0.4}$ is the sum of E_T in the 0.2-0.4 isolation cone (Fig. 44).

Isolation cuts for muons:

- $|E_T^{0.2} - p_T^\mu| < 9.0$ GeV and $E_T^{0.2}/p_T^\mu < 0.3$, where $E_T^{0.2}$ is the total E_T in the 0.2 isolation cone and p_T^μ is the muon transverse momentum measured in Tracker (Fig. 45 and 46).
- $E_T^{0.4}/p_T^\mu < 0.3$, where $E_T^{0.4}$ is the sum of E_T in the 0.2-0.4 isolation cone (Fig. 47).

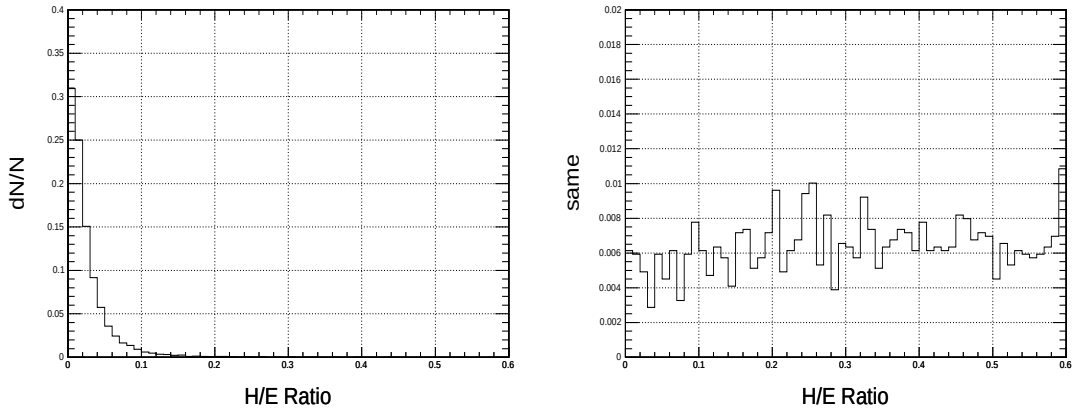


Figure 40: E_T^{Hcal}/E_T^{Ecal} of true electron (left) and faked electron (right) in the signal sample with $m_H = 170$ GeV/c^2

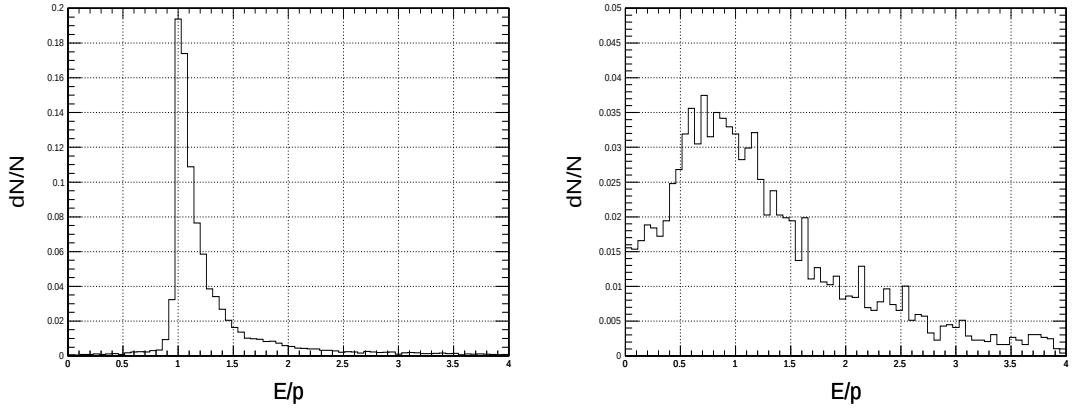


Figure 41: E/p of true electron (left) and faked electron (right) in signal sample with $m_H = 170$ GeV/c^2

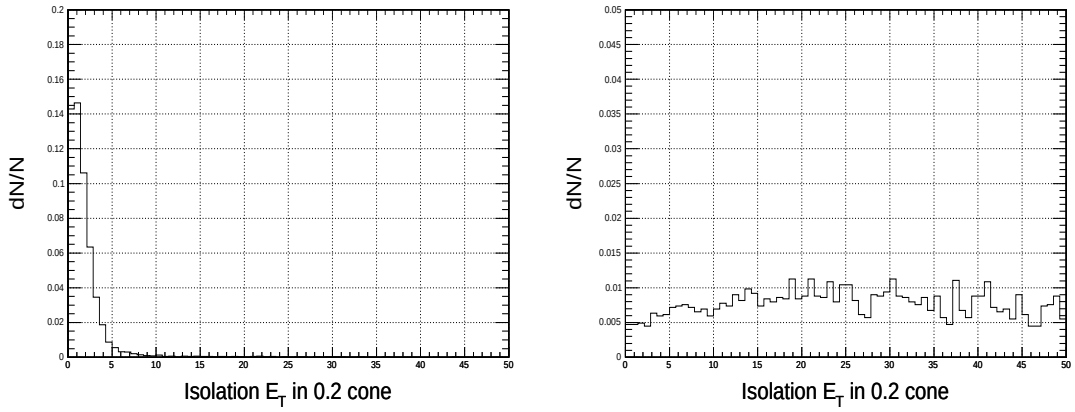


Figure 42: $|\mathrm{E}_{\mathrm{T}}^{0.2} - \mathrm{E}_{\mathrm{T}}^{\mathrm{e}}|$ of true electron (left) and faked electron (right) in the signal sample with $\mathrm{m_H} = 170$ $\mathrm{GeV/c^2}$

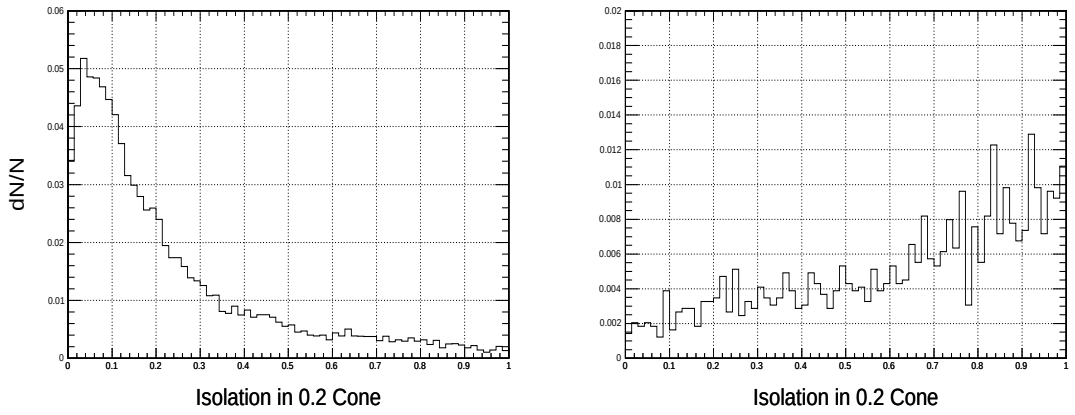


Figure 43: $|(\mathrm{E}_{\mathrm{T}}^{0.2} - \mathrm{E}_{\mathrm{T}}^{\mathrm{e}})/\mathrm{E}_{\mathrm{T}}^{\mathrm{e}}|$ of true electron (a) and faked electron (b) in the signal sample with $\mathrm{m_H} = 170$ $\mathrm{GeV/c^2}$

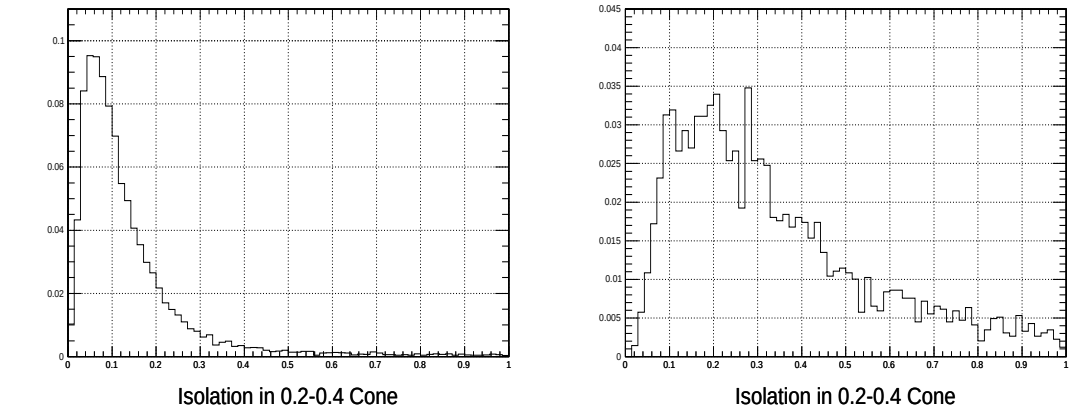


Figure 44: $|(\mathrm{E}_{\mathrm{T}}^{0.4}/\mathrm{E}_{\mathrm{T}}^{\mathrm{e}}|$ of true electron (left) and faked electron (right) in the signal sample with $\mathrm{m_H} = 170\,\mathrm{GeV/c^2}$

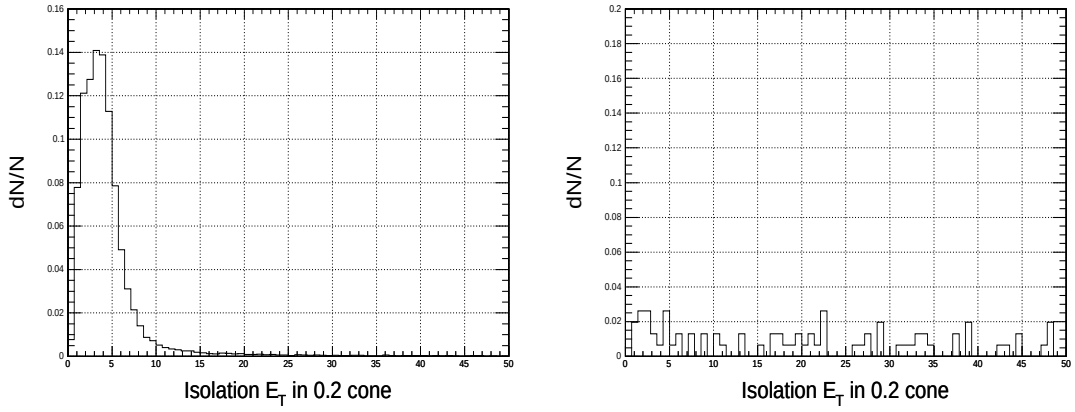


Figure 45: $|\mathrm{E}_{\mathrm{T}}^{0.2} - \mathrm{E}_{\mathrm{T}}^{\mathrm{e}}|$ of true muon (left) and faked muon (right) in the signal sample with $\mathrm{m_H} = 170\ \mathrm{GeV/c^2}$

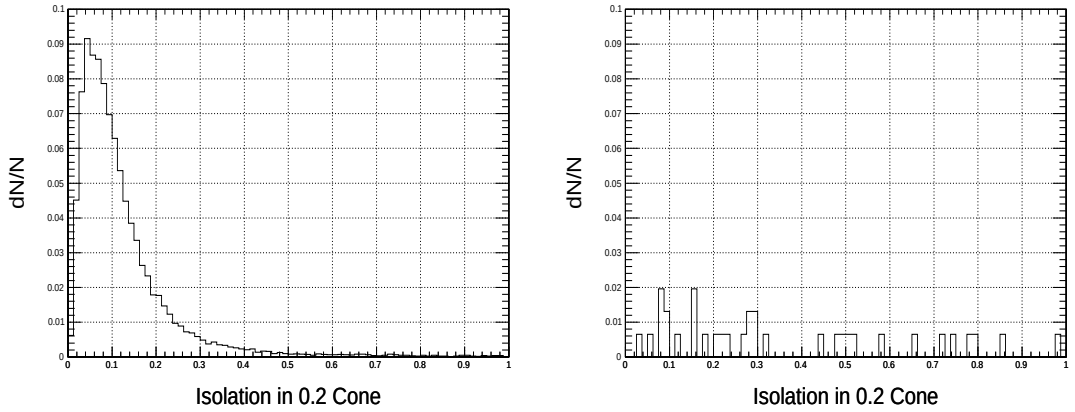


Figure 46: $|(\mathrm{E}_{\mathrm{T}}^{0.2} - \mathrm{E}_{\mathrm{T}}^{\mathrm{e}})/\mathrm{E}_{\mathrm{T}}^{\mathrm{e}}|$ of true muon (left) and faked muon (right) in the signal sample with $\mathrm{m_H} = 170\ \mathrm{GeV/c^2}$

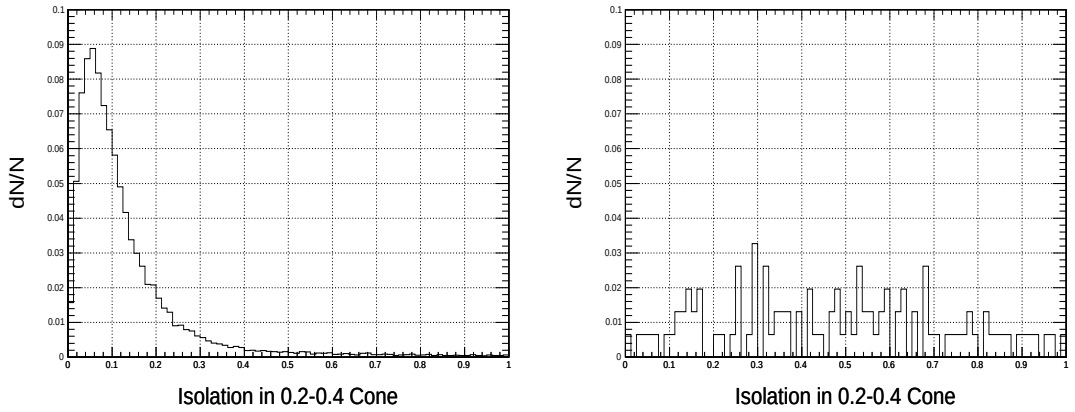


Figure 47: $|(\mathrm{E}_{\mathrm{T}}^{0.4}/\mathrm{E}_{\mathrm{T}}^{\mathrm{e}}|$ of true muon (left) and faked muon (right) in the signal sample with $\mathrm{m_H} = 170\ \mathrm{GeV/c^2}$

Appendix 2: Estimate QCD Event Selection Efficiency

In order to estimate the contamination of QCD background in the final selection, the overall selection chain are grouped into three parts: leptonic-W chain, hadronic-W chain and optimization chain. The selection cuts for leptonic-W and hadronic-W chain are the same as basic selection cuts in Section 5, and the optimization chain is same as those in Section 7. The leptonic-W chain and hadronic-W chain are independent, while optimization chain will need to be conducted after Leptonic-W and Hadtonic-W chains.

The basic method is to factorize the selection efficiency of QCD events for each selection chain:

- Leptonic-W chain includes: lepton trigger, lepton isolation, leptonic-W selection. The selection efficiency for QCD events is $\mathrm{E_L}$.
- Hadronic-W chain includes: jet topology cuts ($\geq$5 jets, $\mathrm{E_T^{miss}}$ cut), forward jet tagging, hadronic-W selection. The reason we require 5 jets in the topology cut is because we need save one jet faking an isolated e/μ ($\mathrm{p_T}$ $\geq$30 GeV/c), then rest 4 jets used for two forward jets and central jets of hadronic-W reconstruction and selection. The selection efficiency for QCD events is $\mathrm{E_H}$.
- For QCD events, since there is no way we can get a single event passing two previous selection chains, so we directly use W+jets efficiency (=0.0005) to estimate the efficiency of QCD events. In another word, we treat those QCD events passing basic selection as leptonic background. The selection efficiency for QCD events is $\mathrm{E_O}$.

The overall selection efficiency for QCD events is: E = $\mathrm{E_L} \times \mathrm{E_H} \times \mathrm{E_O}$. The $\mathrm{E_L}$ and $\mathrm{E_H}$ is directly measured from QCD samples. Following is the summary of selection efficiency for QCD events:

Table 19: Factorized QCD Event Selection Efficiency

QCD $\hat{\mathrm{p}}_\mathrm{T}$	L_W Chain	H_W Chain	Opt Chain	Evt in 60 fb^{-1}
qcd_20_30	0	6.3E-4	5.0E-4	
qcd_30_50	0	1.3E-3	5.0E-4	
qcd_50_80	0	3.4E-3	5.0E-4	
qcd_80_120	<3.8E-6	6.7E-3	5.0E-4	<2.57
qcd_120_170	5.0E-6	1.3E-2	5.0E-4	1.1
qcd_170_230	5.6E-6	2.2E-2	5.0E-4	0.44
qcd_230_300	1.1E-5	3.3E-2	5.0E-4	0.31
qcd_300_380	2.4E-5	3.4E-2	5.0E-4	0.19
qcd_380_470	2.6E-5	3.9E-2	5.0E-4	0.08
qcd_470_600	1.6E-5	4.3E-2	5.0E-4	0.02

The lepton-W selection efficiency for low $\hat{\mathrm{p}}_\mathrm{T}$ samples is very little due to the isolation cuts and lepton $\mathrm{p_T}$ threshold.

The number of "expected" events at 60 fb^{-1} for each QCD $\hat{\mathrm{p}}_\mathrm{T}$ range is summarized in Table 19. The overall QCD contamination is estimated as $\sim$ 2-5 events for an upper limit, which causes possible a few percent increase of background (total $\sim$ 100 background events after the final selection at 60 fb^{-1}). There is almost no change in the significance (we have $\sim$ 100 signal events after the final selection in the most interested $\mathrm{m_H}$ range).