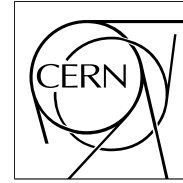


The Compact Muon Solenoid Experiment

# CMS Note

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



19 June 2006

## Vector boson fusion production with $H \rightarrow \gamma\gamma$

M. Dubinin<sup>1)</sup>, V. Litvin<sup>2)</sup>, Y. Ma<sup>2)</sup>, H. Newman<sup>2)</sup>, M. Pieri<sup>3)</sup>

### Abstract

A study of CMS discovery potential for an intermediate mass Higgs boson is presented in this note. Results have been obtained for the vector boson fusion (VBF) production mode,  $pp \rightarrow qqH \rightarrow qq\gamma\gamma$  where the final state includes two high-rapidity jets, during the first years of running at low luminosity. Full detector simulation is used both for the Higgs boson signal and all types of background.

---

<sup>1)</sup> Moscow State University, Moscow, Russia

<sup>2)</sup> California Institute of Technology, Pasadena, CA 91125, USA

<sup>3)</sup> University California San Diego, San Diego, CA, USA

# 1 Introduction

The  $H \rightarrow \gamma\gamma$  process is one of the most promising signals by which to search for the Higgs boson in the intermediate mass region between 115 and 150  $\text{GeV}/c^2$  [1, 2, 3, 14, 15].

Results are obtained for the VBF production mode [16, 17],  $pp \rightarrow qqH \rightarrow qq\gamma\gamma$  where the photon pair is accompanied by two high-rapidity jets. This study uses full detector simulation both for the Higgs boson signal and all types of background.

A special goal of this study is to simulate a large enough sample of QCD hadronic jet background to directly estimate the di-photon misidentification and compare it with contributions from the other types of background. The QCD hadronic jet background cross section is huge ( $\sim 10^9$  pb). Previous studies of QCD hadronic jet background [1, 15] have either been done at the generator level or have obtained estimates for the rate  $r_{\text{jet}}$  at which a jet would be misidentified as a photon. Due to limited computational power, however, these studies estimated the QCD hadronic jet background by simply multiplying the rates together (factor  $r_{\text{jet}}^2$ ). The problem with this method is that the correlations within an event were not taken into account. In addition, the simulation was done with simplified geometry, so non-Gaussian tails in the resolution have not been adequately simulated.

Another important goal of this study is to simulate sufficiently large  $2 \rightarrow 4$  datasets with CompHEP, where all diagrams are taken into account.  $\gamma + 3\text{jets}$  and  $\gamma\gamma + 2\text{jets}$  background samples were calculated with CompHEP and it is shown that the PYTHIA procedure of the  $2 \rightarrow 2$  matrix element convolution with the leading log partonic splitting functions of the radiation from initial/final state, may underestimate the fraction of events with large isolation between the photon and the jet.

It is shown in this note that isolation criteria, based on information from the  $\text{PbWO}_4$  electromagnetic calorimeter and the tracker, are important in separating the signal process from the background processes. An additional key feature is the presence of two energetic forward jets in the final state [16, 17]. It is shown that the forward jet tagging technique is very efficient at further reducing the background.

## 2 Signal and background

### 2.1 Higgs boson signal

Table 1: Cross section for vector boson fusion Higgs boson production process

Higgs mass	$\sigma_{VBF}$	$\text{Br}(H \rightarrow \gamma\gamma)$	$\sigma_{VBF}\text{BR}$
110 $\text{GeV}/c^2$	4.7 pb	$1.90 \times 10^{-3}$	8.93 fb
115 $\text{GeV}/c^2$	4.3 pb	$2.08 \times 10^{-3}$	8.94 fb
120 $\text{GeV}/c^2$	4.1 pb	$2.19 \times 10^{-3}$	8.98 fb
130 $\text{GeV}/c^2$	4.0 pb	$2.23 \times 10^{-3}$	8.92 fb
140 $\text{GeV}/c^2$	3.7 pb	$1.94 \times 10^{-3}$	7.18 fb
150 $\text{GeV}/c^2$	3.4 pb	$1.39 \times 10^{-3}$	4.72 fb

The cross section for the vector boson fusion Higgs boson production as calculated by the program VV2H [6] is shown in Table 1. The most up-to-date next-to-leading order corrections (so-called K factors) were included in the cross section calculation. The branching ratio  $\text{Br}(H \rightarrow \gamma\gamma)$  and the product of the VBF cross section and branching ratio are also shown in Table 1. The branching ratio was calculated with the program HDECAY [25].

Statistical accuracy of the vector boson fusion cross sections is expected at the level of better than 10% [18]. With this statistical precision it is important to know the magnitude of NLO QCD corrections to the total rate and distributions [19]. The distributions in jet  $p_T$  at  $p_T \geq 50$   $\text{GeV}/c$ , jet rapidity and in the difference of jet rapidities  $\Delta y_{jj}$ , essential for this study, are stable with respect to the NLO corrections on the level of a few percent. The NLO effects for the total rate amount to a 5-8% increase, with slight dependence on the method used for selection of forward jets.

### 2.2 Different types of background

The search for the  $H \rightarrow \gamma\gamma$  signal at LHC is affected by four types of backgrounds:

- Prompt di-photon production from the quark annihilation and gluon fusion diagrams, which provides an

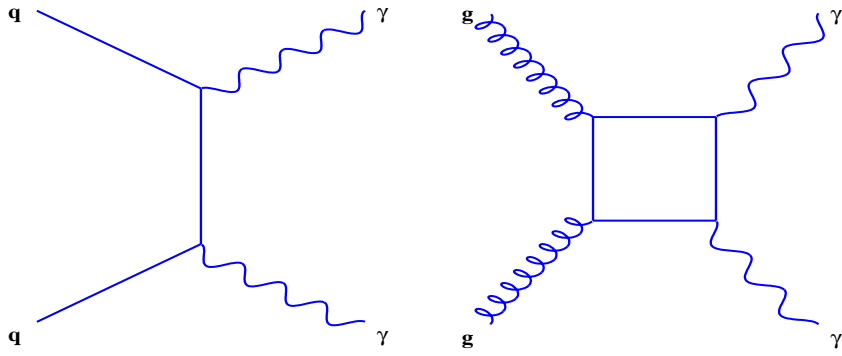


Figure 1: Diagrams of irreducible background: (left) quark annihilation, (right) gluon fusion.

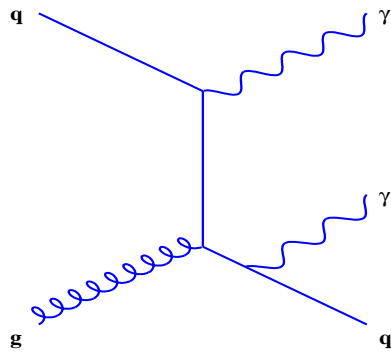


Figure 2: Diagram of bremsstrahlung process.

intrinsic or ‘irreducible’ background (illustrated in Figure 1).

- $\gamma$  + jets production consisting of two parts: i) prompt photon + a second photon coming from the outgoing quark due to bremsstrahlung (illustrated in Figure 2) and ii) prompt photon + neutral hadron (mostly isolated  $\pi^0$ ) in a jet.
- QCD hadronic jets, where an electromagnetic energy deposit results from the decay of neutral hadrons (especially isolated  $\pi^0$ s) in a jet.
- Drell Yan process with  $e^+e^-$  in a final state which could mimic photons when corresponding electron tracks are not assigned to the superclusters during the reconstruction.

Table 2 shows the cross sections of different types of background, along with the corresponding threshold of the transverse momentum of the parton in the hard interaction process  $P_T^{\text{hard}}$  that was used for generation.

Table 2: Cross sections of different types of background, calculated at leading order.

Background process	$P_T^{\text{hard}}$ (GeV/c)	Cross section (pb)
QCD hadronic jets (eg05_jets_2g_pt50up)	50	$2.8 \cdot 10^7$
$\gamma$ + jets (eg03_gamjet)	30	$4.93 \cdot 10^4$
Gluon fusion (eg03_twophoton_box)	20	83
Quark annihilation (eg03_twophoton_born)	20	83
Drell Yan (sm05_dy2e)	15	$4.1 \cdot 10^3$
$\gamma$ + 3 jets, QCD (CompHEP)	–	5970
$\gamma\gamma$ + 2jets, QCD (CompHEP)	–	47.24
$\gamma\gamma$ + 2jets, EW (CompHEP)	–	0.33

In the framework of the CompHEP approach [20], the backgrounds are identified at the level of Feynman diagram sets as:

- The irreducible QCD background  $\gamma\gamma+2$  jets from the 34 partonic  $2 \rightarrow 4$  subprocesses. The main contribution comes from the subprocesses with  $u$ -quark and gluon in the initial state  $ug \rightarrow \gamma\gamma gu$  and  $gu \rightarrow \gamma\gamma gu$ . The amplitude of  $ug \rightarrow \gamma\gamma gu$  is given by the sum of 15 Feynman diagrams shown in Figure 3. Subleading contributions come from  $\bar{u}g \rightarrow \gamma\gamma g\bar{u}$ ,  $g\bar{u} \rightarrow \gamma\gamma g\bar{u}$ ,  $uu \rightarrow \gamma\gamma uu$  and  $gg \rightarrow \gamma\gamma u\bar{u}$ .
- The irreducible electroweak background  $\gamma\gamma+2$  jets from 20 partonic  $2 \rightarrow 4$  subprocesses. The main contribution comes from the subprocesses with  $u, d$ -quarks in the initial state  $ud \rightarrow \gamma\gamma ud$  and  $du \rightarrow \gamma\gamma ud$ .
- The misidentification QCD background in the channel  $pp \rightarrow \gamma+3$  jets from 50 partonic  $2 \rightarrow 4$  subprocesses. The main contribution comes from the subprocesses with  $u$ -quark and gluon in the initial state  $ug \rightarrow \gamma ggu$  and  $gu \rightarrow \gamma ggu$ .
- The misidentification EW background in the channel  $pp \rightarrow \gamma+3$  jets from 24 partonic  $2 \rightarrow 4$  subprocesses. The main contribution is given by  $ud \rightarrow \gamma gud$  and  $du \rightarrow \gamma gud$ . The total cross section for this background is 5.15 pb with the same partonic level cuts as defined above.

Cross sections for the irreducible background are calculated for  $p_T \geq 20$  GeV/c for photons and jets and separation  $\Delta R_{ij} \geq 0.4$ , where  $R = \sqrt{\delta\eta^2 + \delta\phi^2}$ . For the reducible background, an additional preselection of two jets in opposite hemispheres with a rapidity gap greater than 3.5 is introduced. The CompHEP package evaluates precisely the squared amplitudes for all 34(50) partonic subprocesses of the irreducible(misidentification) backgrounds by means of adaptive Monte-Carlo integration and then generates unweighted events at the partonic level followed by shower evolution and hadronization interfaced to PYTHIA. Thus, the main difference for background simulation in the frameworks of the PYTHIA and CompHEP approaches is in the approximations used by these packages for evaluation of the partonic amplitude. For example, the amplitude of the leading QCD partonic subprocess  $ug \rightarrow \gamma\gamma gu$  of the irreducible background (see Figure 3) within the effective approximations can be schematically written as:

$$\begin{aligned}
A_0 = \sum_{i=1}^{15} A_i &= I_g^u \otimes I_u^g \otimes A_{10}(u\bar{u} \rightarrow \gamma\gamma) + I_\gamma^u \otimes I_u^g \otimes [A_2(u\bar{u} \rightarrow \gamma g) + A_5(u\bar{u} \rightarrow \gamma g)] \\
&+ [A_4(ug \rightarrow u\gamma) + A_6(ug \rightarrow u\gamma) + A_9(ug \rightarrow u\gamma) + A_{11}(ug \rightarrow u\gamma)] \otimes F_\gamma^u \otimes F_g^u \\
&+ I_\gamma^u \otimes A_1(ug \rightarrow u\gamma) \otimes F_g^u + I_g^u \otimes A_{12}(ug \rightarrow u\gamma) \otimes F_\gamma^u \\
&+ I_g^g \otimes [A_8(ug \rightarrow u\gamma) + A_{15}(ug \rightarrow u\gamma)] \otimes F_\gamma^u + I_\gamma^u \otimes I_g^g \otimes A_3(ug \rightarrow u\gamma) \\
&+ I_\gamma^u \otimes A_7(ug \rightarrow ug) \otimes F_\gamma^u + [A_{13}(ug \rightarrow ug) + A_{14}(ug \rightarrow ug)] \otimes F_\gamma^u \otimes F_\gamma^u
\end{aligned}$$

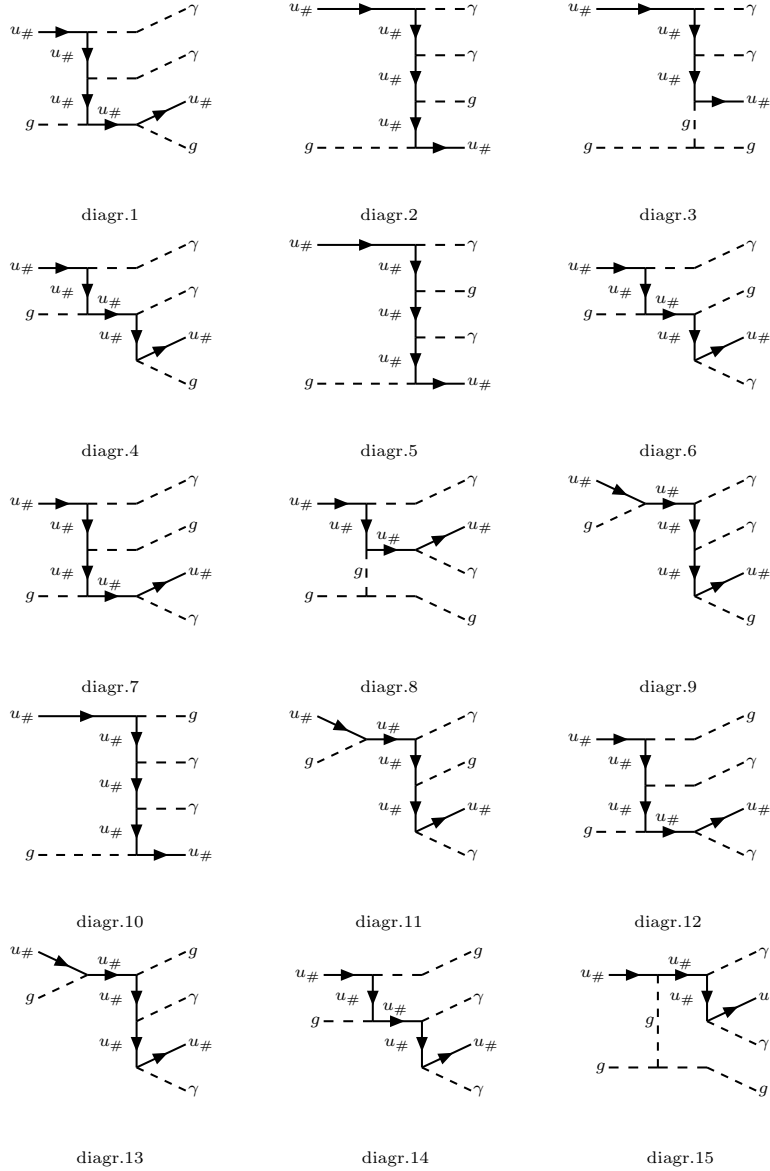


Figure 3: Complete set of  $2 \rightarrow 4$  tree level diagrams for the partonic subprocess No.7  $u\#g \rightarrow \gamma\gamma gu\#$ .

where, for example, the first term  $I_g^u \times I_u^g \otimes A_{10}(u\bar{u} \rightarrow \gamma\gamma)$  denotes the ladder diagram No.10 in Figure 3 evaluated as the convolution of the  $2 \rightarrow 2$  subprocess  $u\bar{u} \rightarrow \gamma\gamma$  amplitude (given by  $\text{MSUB}(18)=1$  in PYTHIA encoding) with the two splitting functions  $I_g^u(z)$  and  $I_u^g(z)$  which define in the collinear approximation the initial state radiative corrections. The final state radiative corrections are denoted by  $F_j^i(z)$  (for the partonic species  $j$  radiated from  $i$  with the momentum fraction  $z$  calculated as the leading log splitting function). While the agreement of total rates with such approximations is usually rather good, large difference between distributions and event orientation may take place. As an illustration we generated various distributions (see Figure 4) using the CompHEP

unweighted event sample for  $pp \rightarrow \gamma\gamma + 2 \text{ jets}$  and PYTHIA event sample for  $MSUB(18)=1$ ,  $q\bar{q} \rightarrow \gamma\gamma$ , with partons reconstructed by PYCLUS routine. Jets appear in PYTHIA from the gluon radiation in the initial state of the  $2 \rightarrow 2$  subprocess  $q\bar{q} \rightarrow \gamma\gamma$  (e.g. see the first term in  $A_0$ ), while in CompHEP they evolve from two partons which are in the final state of various  $2 \rightarrow 4$  channels. Only the events with two jets at  $p_T \geq 20 \text{ GeV}/c$  reconstructed by PYCLUS are accepted, while the multijet configurations (with 4-5 jets and even more) are rejected. Note that we do not require the jets to be in the opposite hemispheres. One can observe that CompHEP distributions in  $p_T$  decrease slower than PYTHIA distributions, calculated in the effective approximation. Much jet activity in the central region is naturally expected with various diagrams of  $s$ -channel topologies (see e.g. diagrams 9,11,13 in Figure 3). On the other hand, in the case of very forward jets the QCD corrections to ladder diagrams like diagram 10 in Figure 3 are large and the rate at high rapidities may be underestimated with CompHEP event generation. Diagrams beyond the  $t$ -channel quark exchange  $q\bar{q} \rightarrow \gamma\gamma$  lead to broader distributions in  $\Delta y_{\gamma\gamma}$  and  $\Delta y_{jet,jet}$ , the latter is especially important in view of the following procedure for the Higgs signal separation. The  $s$ -channel topologies in PYTHIA can be included by using  $MSUB(29)=1$  to switch on  $gq \rightarrow \gamma q$ , however this  $2 \rightarrow 2$  partonic level amplitude includes only one final photon, then the second photon is generated from ISR or FSR and simulated in the effective collinear approximation. Although direct numerical matching of matrix elements in PYTHIA and CompHEP approaches is extremely difficult, it follows from such simple examples that on the qualitative level one can expect at the partonic level substantial enhancement of CompHEP backgrounds in comparison with PYTHIA

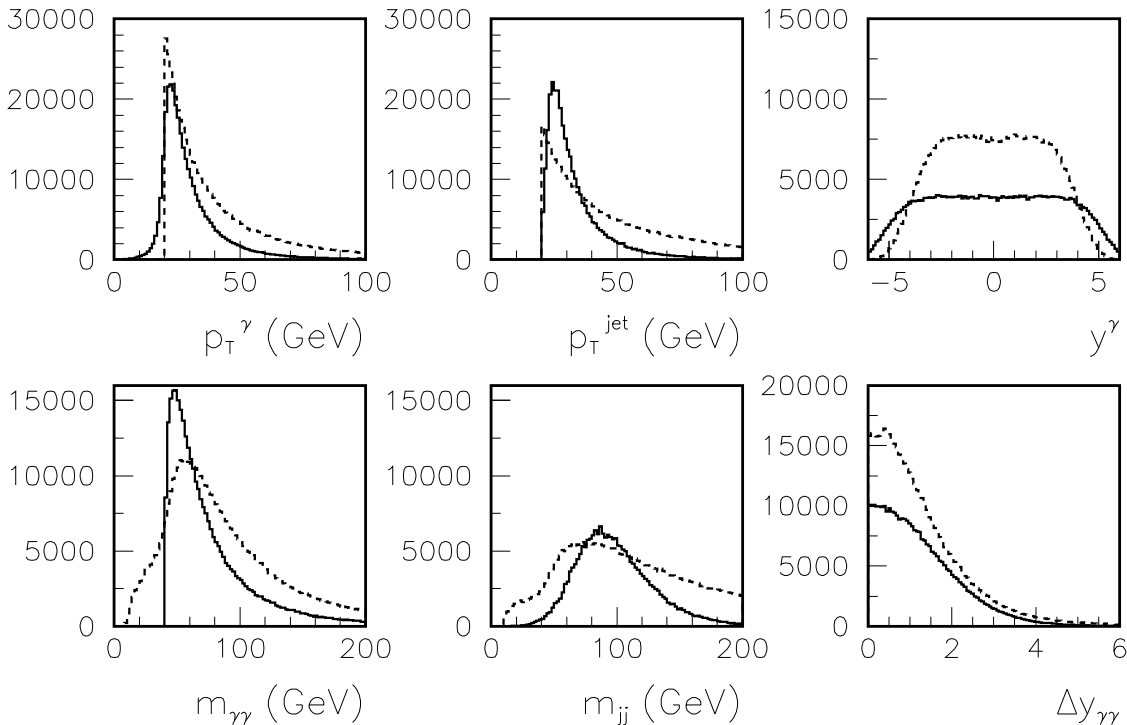


Figure 4: PYTHIA (solid line) and CompHEP (dashed line) partonic level distributions for the irreducible background  $\gamma\gamma + 2\text{jets}$

datasets at the  $p_T \geq 30 \text{ GeV}/c$ , at high invariant masses ( $m_{\gamma\gamma} \geq 100 \text{ GeV}/c^2$ ) and large rapidity gaps ( $\Delta y \sim 2-3$ ). The CompHEP and PYTHIA approaches to the evaluation of the amplitude use quite different approximations, so careful analysis is appropriate, generally speaking, for their applications in specific kinematical configurations. Some other comparisons of PYTHIA results with various generators for complete tree-level sets of diagrams at LEP2 energies can be found in [22].

### 2.3 Generator level preselection for QCD hadronic jets and CompHEP background samples $\gamma + 3\text{jets}$ and $\gamma\gamma + 2\text{jets}$

A selection based on the generated particles was devised, aimed at selecting events which could produce in the detector two electromagnetic showers consistent with isolated photons. In order to apply cuts on the invariant mass of the two candidates, an attempt to estimate lower and upper limits to the energy of the candidates that will be

reconstructed after the simulation was done.

The idea of this preselection, is to pick up events that will give rise to energy depositions in ECAL large enough and isolated enough to be important for this analysis. The preselection algorithm gets all particles which might deposit electromagnetic energy in ECAL, and looks around each particle in a narrow cone to find another, possibly less energetic particle which will make deposits in ECAL as well, and will potentially be reconstructed as one supercluster. In addition to that, loose tracker isolation was applied:  $N_{tkconemax}$  charged particles were required in a cone of  $\Delta R = 0.2$  around the “supercluster candidate”, described above, per one “supercluster candidate”, and no more than  $N_{tksummax}$  per two first most energetic candidates. The most energetic candidate, for the purposes of this note, means the candidate with highest  $p_T$ .

After that some other cuts were applied for the “supercluster candidates” as well -  $p_T > 37.5$  GeV/c for most energetic one and  $p_T > 22.5$  GeV/c for the second most energetic one. The invariant mass of the first most energetic and second most energetic “supercluster candidates” should be more than 90 GeV/c<sup>2</sup> for the purpose of this analysis.

Firstly, all seed candidates were identified. Only the following particles, that could produce in the detector electromagnetic showers consistent with photons, were considered:  $\gamma$ ,  $e^\pm$ ,  $\pi^0$ ,  $\eta$ ,  $\eta'$ ,  $\rho$ ,  $\omega$ .

The seeds should satisfy the cuts:

- $P_{tseed} > P_{tseedmin}$
- $|\eta_{seed}| < \eta_{seedmax}$

In order to take into account the fact that superclusters have a sizeable spread in  $\eta$  and  $\phi$  and that more than one particle can contribute to the reconstructed energy, these seeds were not used as photon candidates. All final state  $\gamma$  and  $e^\pm$  around their direction were added. The photon candidate was defined as the vector sum of the momenta of the  $\gamma$  and  $e^\pm$  that satisfy:

- $P_{t\gamma} > P_{t\gamma min}$
- $|\eta_\gamma| < \eta_{\gamma max}$
- $P_{te} > P_{te min}$
- $|\eta_e| < \eta_{e max}$
- $\Delta R < \Delta R_{\gamma-seed}$  or ( $\Delta\eta < \Delta\eta_{\gamma-seed}$  and  $\Delta\phi < \Delta\phi_{\gamma-seed}$ )

It is also possible to include  $K_{long}^0$  as seeds and as components of the candidate with the cuts:

- $P_{tK_{long}^0} > P_{tK_{long}^0 min}$
- $|\eta_{K_{long}^0}| < \eta_{K_{long}^0 max}$

In addition, the  $\gamma$  and  $e^\pm$  within a narrow cone around the seed were added in order to estimate the lowest possible energy that could be reconstructed for the candidate. These should satisfy:

- $P_{t\gamma} > P_{t\gamma min}$
- $|\eta_\gamma| < \eta_{\gamma max}$
- $P_{te} > P_{te min}$
- $|\eta_e| < \eta_{e max}$
- $\Delta R < \Delta R_{narrow}$

This energy will be referred to as “narrow cone energy”.

In order to evaluate the effect of the tracker isolation, stable or long lived charged particles around the direction of the photon candidates are counted using the following cuts:

- $P_{t_{tk}} > P_{t_{tkmin}}$
- $\Delta R < \Delta R_{tkmax}$
- $|\eta_{tk}| < \eta_{tkmax}$

Finally, it was required that at least a pair of candidate photons in the event satisfy:

- $P_{t_1} > P_{t_{1min}}$
- $P_{t_2} > P_{t_{2min}}$
- $N_{tkcone1}, N_{tkcone2} \leq N_{tkconemax}$
- $N_{tkcone1} + N_{tkcone2} \leq N_{tksummax}$
- $|\eta_1|, |\eta_2| < \eta_{candmax}$
- Invariant mass of the pair  $> M_{\gamma\gamma min}$
- Invariant mass of the pair (using the narrow cone energies)  $< M_{\gamma\gamma max}$

Table 3: Cuts for the QCD and CompHEP backgrounds preselection

Variable	QCD cut value	CompHEP cut value
$P_{t_{seedmin}}$ (GeV)	5.	5.
$\eta_{seedmax}$	2.6	2.6
$P_{t_{\gamma min}}$ (GeV)	0.0	0.0
$\eta_{\gamma max}$	2.8	2.8
$P_{t_{emin}}$ (GeV)	2.0	2.0
$\eta_{emax}$	2.8	2.8
Use $K_{long}^0$	Yes	Yes
$P_{t_{K_{long}^0 min}}$ (GeV)	1.0	1.0
$\eta_{K_{long}^0 max}$	2.8	2.8
$\Delta\phi_{\gamma-seed}$	0.4	0.4
$\Delta\eta_{\gamma-seed}$	0.15	0.15
$\Delta R_{\gamma-seed}$	0.2	0.2
$\Delta R_{narrow}$	0.02	0.02
$P_{t_{tkmin}}$ (GeV)	1.6	1.6
$\eta_{tkmax}$	2.2	2.2
$\Delta R_{tkmax}$	0.2	0.2
$P_{t_{1min}}$ (GeV)	37.5	37.5
$P_{t_{2min}}$ (GeV)	22.5	22.5
$\eta_{candmax}$	2.6	2.6
$N_{tkconemax}$	1	2
$N_{tksummax}$	2	4
$M_{\gamma\gamma min}$ (GeV)	90	90
$M_{\gamma\gamma max}$ (GeV)	14000	14000

The main difference between the two sets of preselection cuts in Table 3 is the stricter cuts for tracker isolation in the case of PYTHIA than CompHEP. The reason was that we know the PYTHIA background much better from our previous studies [23] so less restrictive preselection cuts in the case of CompHEP insures that we will not lose anything important here.

Preselection inefficiency will be less than 3% to preselect  $\gamma + jet$  events and less than 5% to preselect QCD hadronic jet events. This 3-5% inefficiency for preselected events will be propagated to the 3-5% of corresponding cross-section uncertainties.

## 2.4 $\gamma + 3jets$ and $\gamma\gamma + 2jets$ CompHEP produced backgrounds

Partonic level events generation was done with CompHEP version 4.2p1 for  $pp \rightarrow \gamma\gamma + 2jets$  subprocess with the preselection cuts:

- $p_T^\gamma > 20 \text{ GeV}/c$
- $p_T^j > 20 \text{ GeV}/c$
- $\Delta R_{ij} > 0.4$

34 partonic subprocesses for the QCD background and 20 partonic subprocesses for the electroweak background were taken into account.

For the subprocess  $pp \rightarrow \gamma + 3jets$  (50 partonic subprocesses, QCD misidentification background) preselection cuts were:

- $p_T^\gamma > 20 \text{ GeV}/c$
- $p_T^j > 20 \text{ GeV}/c$
- $\Delta R_{ij} > 0.4$
- two jets in opposite hemispheres with a rapidity gap greater than 3.5

The parameter values are  $\alpha_e = 1/137$  (i.e.  $e = 0.3029$ ) and  $\alpha_s = 0.1298$  ( $g_s = \sqrt{4\pi\alpha_s} = 1.2772$ ),  $m_Z = 91.19 \text{ GeV}/c^2$ ,  $\sin\theta_w = 0.481$ ,  $\Gamma_Z = 2.44 \text{ GeV}$ ,  $\Gamma_W = 2.03 \text{ GeV}$ . The PDF set was taken to be CTEQ5L, and the factorization scale 50 GeV.

Hadronization was done by PYTHIA and generator-level preselection was applied for both  $\gamma + 3jets$  and  $\gamma\gamma + 2jets$  datasets as it was described (see Table 3). Rejection factors are 2.5 for  $\gamma\gamma + 2jets$  dataset and 7.8 for  $\gamma + 3jets$  dataset.

Because of the large cross section,  $2 \rightarrow 4$  subprocess with four jets in final state, was not studied here.

After that, the usual selection from Section 4 was applied.

## 3 Simulation tools and Monte Carlo data samples

For the Higgs boson background study the following data samples were used:

Table 4: Number of generated and simulated events for different types of background.

Background process	Number of generated events	Rejection rate	Number of simulated events	$L_{\text{intg}}$ ( $\text{fb}^{-1}$ )
QCD hadronic jets	$31.2 \times 10^9$	6048	4.5M	$\sim 1$
$\gamma + jets$	$1.3 \times 10^8$	20.0	5.5M	$\sim 2$
Gluon fusion	$2.25 \times 10^6$	2	1M	$\sim 52$
Quark annihilation	$3.0 \times 10^6$	2.7	1M	$\sim 55$
Drell Yan	$1.0 \times 10^6$	1	1M	0.25
$\gamma + 3jets$ QCD	$0.3 \times 10^6$	7.8	40k	0.05
$\gamma\gamma + 2jets$ QCD	$0.5 \times 10^6$	2.56	200k	6
$\gamma\gamma + 2jets$ EW	$41 \times 10^3$	1	41k	120

- Signal:
  - $H \rightarrow \gamma\gamma$  at five different Higgs boson masses ( $m_H = 115, 120, 130, 140$  and  $150 \text{ GeV}/c^2$ ):
    - $115 \text{ GeV}/c^2$  - 16000 events,
    - $120 \text{ GeV}/c^2$  - 192920 events,
    - $130 \text{ GeV}/c^2$  - 10700 events,

- 140 GeV/c<sup>2</sup> - 20000 events,
- 150 GeV/c<sup>2</sup> - 20000 events,
- Background:
  - 4.5M preselected QCD hadronic jet events (QCD),
  - 5.5M preselected  $\gamma$  + jets events (brem),
  - 1M preselected quark annihilation events (born),
  - 1M preselected gluon fusion events (box),
  - 1M preselected Drell Yan events,
  - 40k preselected QCD  $\gamma$  +3jets CompHEP events,
  - 200k preselected QCD  $\gamma\gamma$  +2jets CompHEP events,
  - 41k preselected electroweak  $\gamma\gamma$  +2jets CompHEP events,

All events were generated with PYTHIA 6.227 [4] at  $\sqrt{s}=14$  TeV with the CTEQ5L parton density structure functions. The simulation of particles propagated through the detector is done using OSCAR version 3.6.5 [9].

The simulated events were digitized using the reconstruction and analysis program ORCA 8.7.3 [10]. Pileup events, corresponding to the luminosity of  $2 \times 10^{33}/\text{cm}^2/\text{s}$  were added.

The numbers of generated and fully simulated events are shown in Table 4 for different types of background. In the last column the corresponding equivalent integrated luminosity is shown.

## 4 Selection and reconstruction

A detailed description of the CMS detector can be found elsewhere [7, 8]. The simulation package OSCAR version 3.6.5 [9] based on GEANT4 was used to describe the detector geometry and materials. This package also handles the particle propagation and interactions with the detector. The reconstruction is done with the CMS object-oriented reconstruction package ORCA version 8.7.1 [10]. This package handles all reconstruction tasks as well as the simulation of the detector response, the Level-1 trigger [7] and High Level Trigger [8]. All events have been simulated with low luminosity pile-up.

### 4.1 Photon reconstruction

The hybrid and island algorithms [24] were used to reconstruct basic clusters in the barrel and the endcap, respectively. These basic clusters were in turn clustered into superclusters to recover the energy radiated by electrons and positrons from photon conversion, which falls outside the seed basic cluster [23]. A photon candidate is associated to each reconstructed supercluster.

Energy deposited by electromagnetic showers in the ECAL are constructed into basic clusters using the island clustering algorithm [24]. The minimum value of the transverse energy for the seed crystal of a basic cluster is 0.5 GeV. These basic clusters are in turn clustered into superclusters to recover the energy radiated by electrons or positrons from photon conversions, which fall outside the seed shower cluster [24]. The hybrid algorithm is used to reconstruct the supercluster in the barrel, and the island algorithm is used in the endcap [24]. The reconstructed supercluster is associated with a photon candidate.

The following photon selection was applied, which includes the trigger, kinematical and geometrical acceptance cuts. The 12 GeV Level-1 double-isolated electron trigger [26] was applied to both the signal and background data samples. Both photon candidates had to match Level-1 trigger objects, such that, the distance  $R$  ( $R = \sqrt{\delta\eta^2 + \delta\phi^2}$ ) between the photon candidate and trigger object be less than 0.5. The transverse energies of the two photon candidates were required to be greater than 40 GeV/c and 25 GeV/c [2] respectively. The fiducial volume in rapidity was restricted to  $|\eta| < 1.4442$  in the barrel and  $1.566 < |\eta| < 2.5$  in the endcap for both photon candidates [2, 27]. Efficiencies of 12 GeV L1 double-isolated trigger, L1 matching procedure and photon candidates transverse energy cuts with invariant mass in the range 110 - 170 GeV/c<sup>2</sup> can be found in Table 5).

Table 5: Efficiencies of 12 GeV L1 double-isolated trigger and L1 matching procedure with invariant mass in the range 110 - 170 GeV/ $c^2$ .

	signal	gluon fusion	quark annihilation	QCD jets	$\gamma$ +jets	Drell Yan	$\gamma$ +3jets QCD	$\gamma\gamma$ +2jets QCD	$\gamma\gamma$ +2jets EW
L1 trigger	92.2%	80.1%	79.9 %	1.7 %	15.2 %	13.2%	20.6%	60%	67.7%
L1 match	92.0 %	78.9 %	78.6 %	1.2%	13.9%	13.0%	20.2%	59.5%	67.5%

## 4.2 Higgs boson mass reconstruction

### 4.2.1 Energy correction

The difference between the photon measured energy and the true generated energy,  $E_{meas} - E_{true}$ , was considered as a function of the ratio S1/S9, where S9 is the energy in a 3×3 crystal matrix, centered around most energetic crystal, S1 is the energy of the most energetic crystal in the matrix. This dependence on S1/S9 shows the influence of the photon impact point with respect to the crystal center. The measured energy  $E_{meas}$  was corrected for this difference.

### 4.2.2 Higgs boson vertex reconstruction

The bunch length at LHC,  $\sigma_Z(\text{bunch}) \approx 75$  mm, results in a longitudinal spread of interaction vertices:  $\sigma_Z(\text{interaction}) \approx 53$  mm. If the mean longitudinal position is used the effective mass of a narrow two-photon state, such as the  $H \rightarrow \gamma\gamma$ , is smeared by about 1.5 GeV/ $c^2$ .

The vertex of the Higgs boson can be found with the help of additional tracks in the same event [28]. The Higgs boson  $p_T$  is balanced by the rest of the particles in the event and therefore the tracks associated with a Higgs boson event are harder than the tracks of a minimum-bias event. Therefore the vertex can be identified by the hardest tracks of the bunch crossing. In addition in the events with converted photons, the position of the primary vertex could also be defined by using the tracks from photon conversion (if any). This improvement was not considered in this paper.

Three different algorithms were studied.

- **$P_T$  balance** - the  $P_T$  balance for charged particle tracks along the reconstructed Higgs boson direction is defined as  $P_T^B = -\sum P_{Ti} \cos\theta_i$ , where  $\theta_i$  is the angle between the Higgs boson and track  $i$  direction in the transverse plane
- **Maximal  $P_T$** : the primary vertex is selected as the vertex with the track of highest  $P_T$
- **Number of charged particle tracks** above  $P_T$  cutoff in pixel vertex. The primary vertex is selected as the vertex with a largest number of tracks.

To compare different vertex reconstruction algorithms, the number of events reconstructed in a 5 GeV/ $c^2$  mass window are determined. The  $P_T$  balance and Maximal  $P_T$  algorithms give exactly the same number of events, while track counting algorithm gives a few percent less efficiency. The left plot in Figure 5 shows the reconstructed Higgs boson mass for two different cases: i) corrected vertex by Maximal  $P_T$  algorithm with energy correction; and ii) no knowledge of the vertex, but still with energy correction applied. The vertex correction improves the resolution of the reconstructed Higgs boson mass. The Higgs boson efficiency in 5 GeV/ $c^2$  mass window is improved by 15%.

The right plot in Figure 5 shows the Higgs boson mass resolution, when the vertex correction, using the Maximal  $P_T$  algorithm, is applied for  $m_{\gamma\gamma}=130$  GeV/ $c^2$ . The typical mass resolution is 0.7% in 115-160 GeV/ $c^2$  mass region.

## 4.3 Isolation

Isolation requirements were found to be essential to reject the backgrounds.

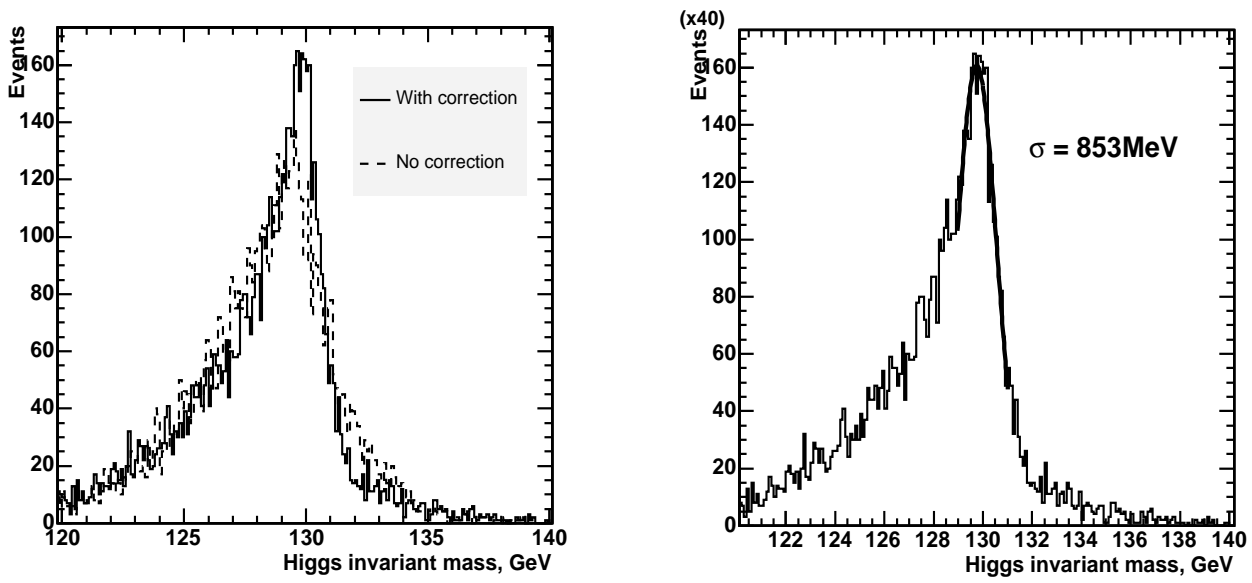


Figure 5: Left plot - the reconstructed Higgs boson  $130 \text{ GeV}/c^2$  invariant mass for two different cases: i) corrected vertex by present algorithm and energy correction; and ii) - no knowledge of the vertex, but with energy correction applied. Right plot - the Higgs boson mass resolution, when the vertex correction, using the present algorithm, is applied.

#### 4.3.1 Isolation based on the tracker

The tracker isolation criteria are based on the number of charged particle tracks with  $p_T$  greater than a  $p_T$  threshold,  $p_T^{\text{thresh}}$ , calculated in a cone  $R$  ( $R = \sqrt{\delta\eta^2 + \delta\phi^2}$ ) around the photon candidate. The algorithm contains three parameters:

- The size of the cone  $R$  around the photon candidate, wherein the number of charged tracks is counted.
- The  $p_T$  threshold,  $p_T^{\text{thresh}}$ . Only charged particle tracks with  $p_T$  greater than  $p_T^{\text{thresh}}$  are considered in isolation calculations.
- The ‘number of tracks’ threshold  $N^{\text{thresh}}$ . If the number of charged particle tracks in cone  $R$  with  $p_T$  greater than the chosen  $p_T^{\text{thresh}}$  is greater than  $N^{\text{thresh}}$ , then the photon candidate is considered non-isolated, otherwise isolated.

The jet rejection factor is very sensitive to the ‘number of tracks’ threshold,  $N^{\text{thresh}}$ . By increasing  $N^{\text{thresh}}$  from 0 to 1, the Higgs boson signal efficiency is improved by 6-10%, but the jet rejection factor drops by a factor of  $\sim 2$ . Therefore, the parameter  $N^{\text{thresh}}$  was fixed to zero. The cone size  $R = 0.30$  and  $p_T^{\text{thresh}} = 1.5 \text{ GeV}/c$  were used in this study. The details and the optimization of the isolation method are discussed in Ref. [23].

#### 4.3.2 Isolation based on the electromagnetic calorimeter

The isolation criteria are based on the sum of transverse energies deposited in basic clusters in some cone  $R$  ( $R = \sqrt{\delta\eta^2 + \delta\phi^2}$ ) around a photon candidate. The basic clusters that belong to the photon candidate’s supercluster are not counted as part of the sum. The algorithm contains four parameters:

- The size of the cone  $R$  around the photon candidate wherein the transverse energies deposited in the basic clusters are summed.
- The transverse energy sum threshold  $E_T^{\text{thresh}}$ . If the sum of transverse energies is below this threshold, the photon candidate is considered isolated, otherwise non-isolated.
- The ratio,  $r$ , of the transverse energy sum in all surrounding basic clusters to the transverse energy of the most energetic super cluster.

- The ratio,  $hoe$ , of the super-cluster energy deposited in HCAL to the energy of this super-cluster, deposited in ECAL.

There is no strong dependence of the QCD hadronic jet rejection factor on the cone size  $R$ , though slightly better rejection factors are empirically obtained for a cone size  $R = 0.30 - 0.35$ . The cone size  $R = 0.30$  is used in this study. The transverse energy sum thresholds,  $E_T^{\text{thresh}}$ , were chosen to be 1.2 GeV in the barrel and 1.6 GeV in the endcap. The ratio,  $r = 0.01$ , of the transverse energy sum in all surrounded basic clusters to the transverse energy of the most energetic super cluster was used for this study. The HCAL/ECAL ratio of the energies of the super-cluster, deposited in HCAL over the super-cluster energy deposited in ECAL was set to 0.1. The details and the optimization of the isolation method are discussed in [23].

## 5 Forward jet tagging

The presence of two forward hard jets is the distinctive signature of vector boson fusion Higgs boson production. These tagging jets provide an additional powerful tool to reduce the background significantly, since the energetic jets at high rapidity, although present in the backgrounds, are separated by smaller rapidity gaps (compared to the signal) and can be suppressed by the cut on  $\Delta\eta_{jet}$ .

Jet tagging was done based on the reconstructed jets by  $E_t$  recombination scheme based on EcalPlusHcalTowers with 0.5 GeV threshold, ConeCut  $\Delta R$  variable was set to 0.7 and JetEtCut variable was set to 10 GeV. No additional Monte Carlo calibration was done.

A basic set of selection criteria for forward jet tagging is similar to [16]. The selection criteria for the forward tagged jet are

$$p_T^{jet} > 20 \text{ GeV}/c, \quad |\eta_{jet}| \leq 4.5, \quad \Delta R_{\gamma jet} \geq 0.5$$

$$\Delta\eta_{jets} = |\eta_{jet1} - \eta_{jet2}| \geq 4.0, \quad \eta_{jet1} \times \eta_{jet2} < 0$$

Each jet should have a transverse momentum greater than 20 GeV/c and rapidity well inside the CMS hadron calorimeter. The distance between the photon candidate and any hadronic jet,  $\Delta R_{\gamma jet}$ , should be greater than 0.5. In addition a large separation in rapidity between two tagged jets is used, and the two jets should be in opposite hemispheres.

The difference in rapidity between the two tagged jets is shown on the left plot of Figure 6 for Higgs boson signal. On the right plot of Figure 6 the rapidity of the selected forward jets is shown for the Higgs boson signal. The distinctive feature that the jets go in the forward/backward region with a peak at  $|\eta| \approx 3$  is clearly seen.

Three additional cuts were applied to the already selected two forward jets in order to reduce the background further:

- $p_t^{jet1} > 50 \text{ GeV}/c$ , where  $p_t^{jet1}$  is the transverse momentum of the first most energetic forward jet, selected by forward jet tagging procedure, described above.
- $p_t^{jet2} > 25 \text{ GeV}/c$ , where  $p_t^{jet2}$  is the transverse momentum of the second most energetic forward jet, selected by forward jet tagging procedure, described above.
- $m_{j1j2} > 500 \text{ GeV}/c^2$ , where  $m_{j1j2}$  is the invariant mass of the two most energetic forward jets, selected by forward jet tagging procedure, described above.

To be able to reduce the total number of events passed through all cuts from CompHEP background sample  $\gamma\gamma + 2jets$  the cut on the transverse momentum of the second most energetic forward jet, selected by forward jet tagging procedure, was adjusted:

- $p_t^{jet2} > 35 \text{ GeV}/c$ , where  $p_t^{jet2}$  is the transverse momentum of the second most energetic forward jet, selected by forward jet tagging procedure, described above.

To suppress the CompHEP background further, an additional cut was introduced:

$$- \min[\eta_{jet1}, \eta_{jet2}] + 0.7 < \eta_{\gamma,2} < \max[\eta_{jet1}, \eta_{jet2}] - 0.7$$

This cut will be named as “geom. cut” in Tables 6 - 10.

The last two cuts will be applied only in the case when CompHEP background samples will be analyzed.

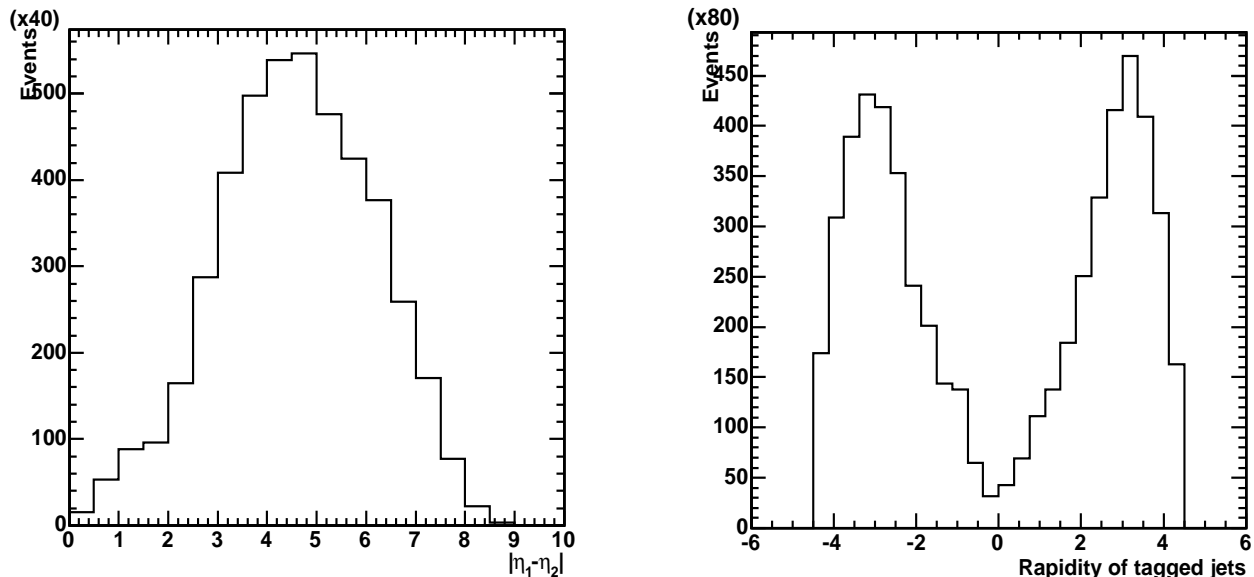


Figure 6: Left plot -  $\Delta\eta$  distribution for the forward tagged jets for the Higgs boson signal ( $130 \text{ GeV}/c^2$  invariant mass). Right plot - The rapidity distribution of the forward tagged jets for the Higgs boson signal after the forward jet selection is applied. Plots were normalized to  $30 \text{ fb}^{-1}$  and multiplied by factor of 40 (left plot) and 80 (right plot).

Figures 7 and 8 show other variables, which are useful for the discrimination between the signal and PYTHIA background. The transverse momentum of tagged jets in Higgs boson signal sample is much harder than the transverse momentum of tagged jets in the background samples, on average (Figure 7). The transverse momentum of the Higgs boson, reconstructed from the photon energies and directions, and the invariant mass of the two forward jets (Figure 8) also tend to peak at larger values for the Higgs boson signal compared with different types of background.

Figures 9 and 10 show other variables, which are useful for the discrimination between the signal, two CompHEP backgrounds  $\gamma + 3\text{jets}$  and  $\gamma\gamma + 2\text{jets}$ , and gluon fusion background. The transverse momentum of tagged jets in Higgs boson signal sample is much harder than the transverse momentum of tagged jets in the background samples, on average (Figure 9), but CompHEP distributions have larger tails and contains more higher energy particles than PYTHIA distributions of the same variables. The transverse momentum of the Higgs boson, reconstructed from the photon energies and directions, and the invariant mass of the two forward jets (Figure 10) also tend to peak at larger values for the Higgs boson signal compared with different types of background. This effect is also smaller than the same effect in pure PYTHIA signal and backgrounds.

## 6 Results

Tables 6 - 10 contain the number of preselected events, left after the each selection cut, for the Higgs boson signal with different masses and different types of the background, with mass window  $5 \text{ GeV}/c^2$  around each mass of Higgs boson.

It is clear from these Tables that the forward jet tagging technique is very efficient in rejecting the background. The irreducible background (quark annihilation and gluon fusion processes) is much smaller than the Higgs boson signal after forward jet tagging technique is applied. The two CompHEP backgrounds  $\gamma + 3\text{jets}$  and  $\gamma\gamma + 2\text{jets}$  are reduced as well.

The statistics of the QCD hadronic jet and  $\gamma + \text{jets}$  background samples as well as two CompHEP backgrounds samples  $\gamma + 3\text{jets}$  and  $\gamma\gamma + 2\text{jets}$  are not sufficient to fully explore and optimize all of the forward jet tagging cuts.

Because of the limited statistics, two different approaches to get the significance of the signal for  $30 \text{ fb}^{-1}$  will be considered in this study.

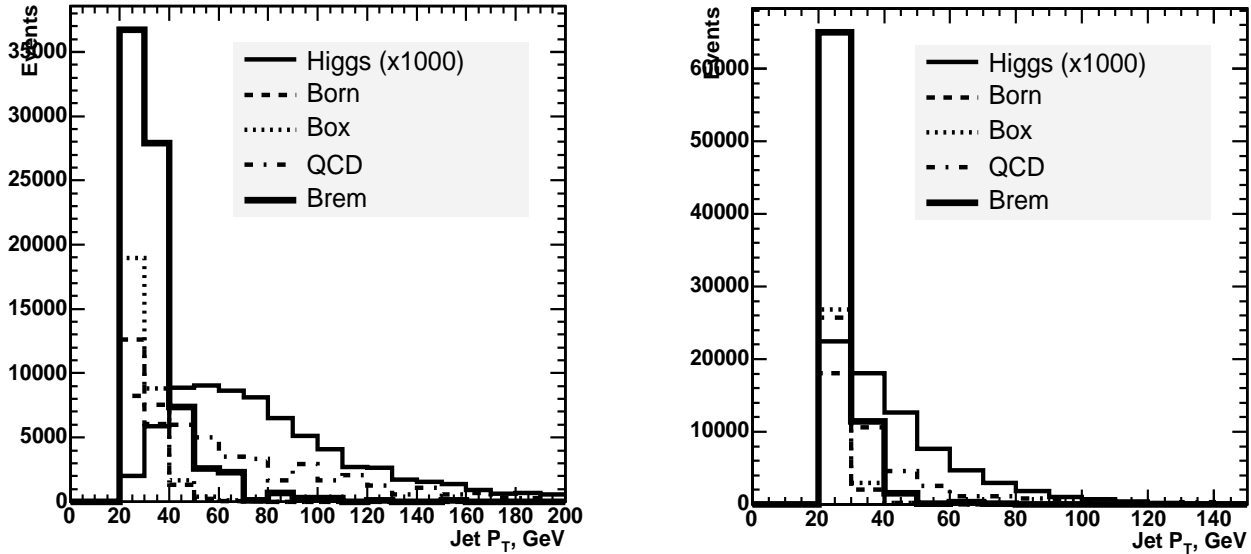


Figure 7: Transverse energy distributions of forward tagged jets after L1 match is applied. Left plot - the distribution for the more energetic jet. Right plot - the distribution for the less energetic jet. Plots were normalized to  $30 \text{ fb}^{-1}$  and Higgs signal was multiplied by 1000 ( $130 \text{ GeV}/c^2$  invariant mass). Description of the backgrounds can be found in Section 3.

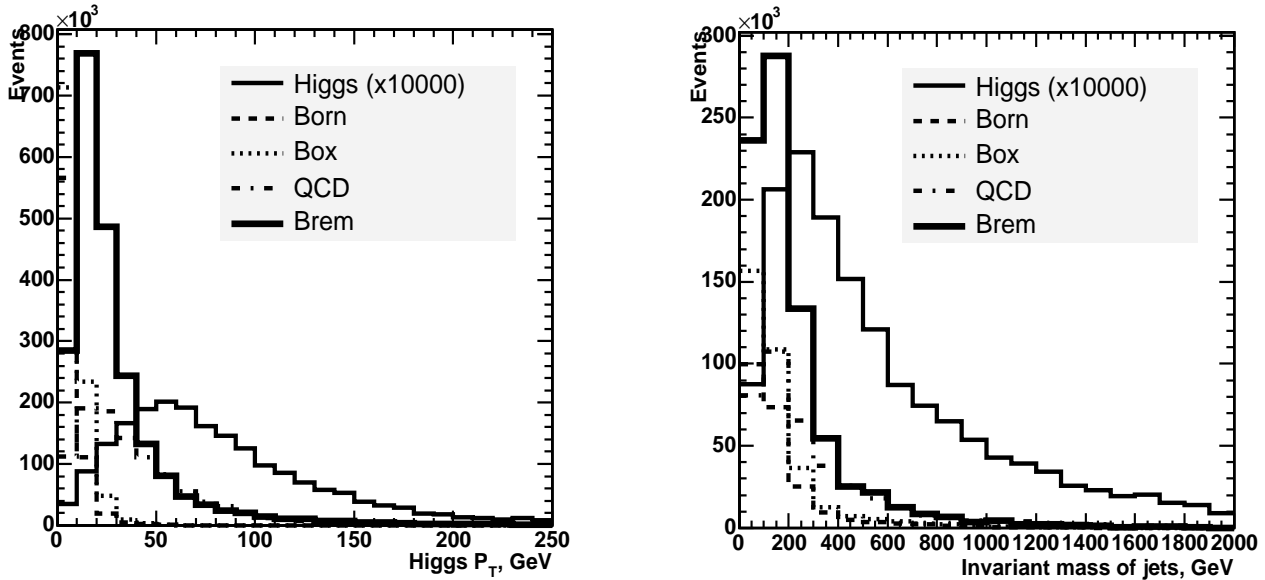


Figure 8: Left plot - the transverse momentum of the Higgs boson, reconstructed from the photon energies and directions. Right plot - the invariant mass of two tagged jets. In both cases L1 match and forward jet tagging were applied. Plots were normalized to  $30 \text{ fb}^{-1}$  and Higgs signal was multiplied by 10000 ( $130 \text{ GeV}/c^2$  invariant mass). Description of the backgrounds can be found in Section 3.

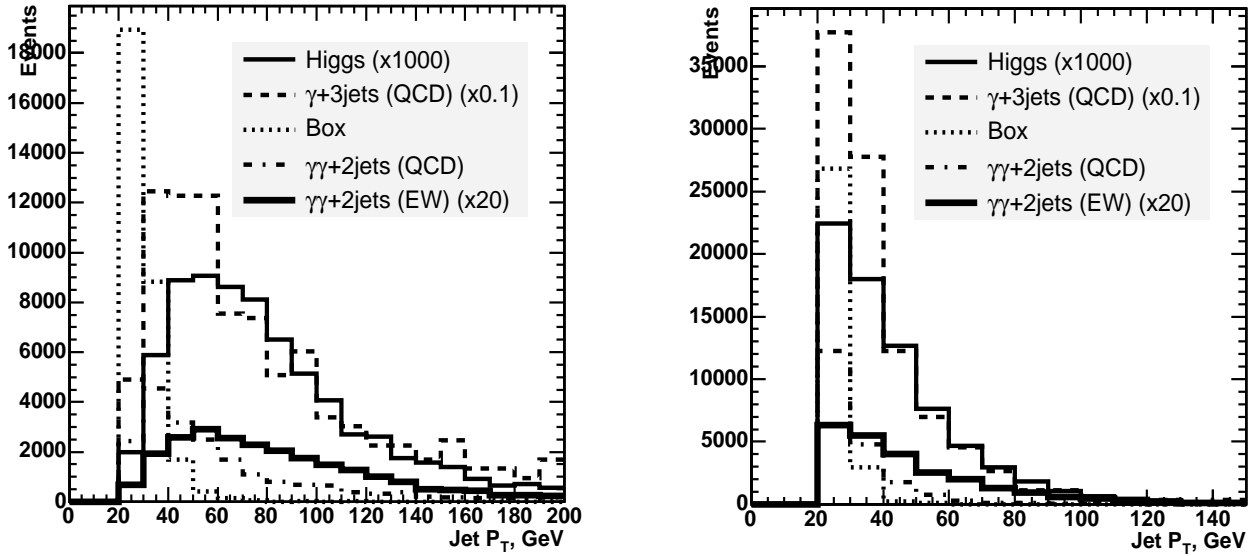


Figure 9: Transverse energy distributions of forward tagged jets for two CompHEP background samples, gluon fusion background sample and Higgs signal sample, after L1 match is applied. Left plot - the distribution for the more energetic jet. Right plot - the distribution for the less energetic jet. Plots were normalized to  $30 \text{ fb}^{-1}$  and Higgs signal was multiplied by 1000 ( $130 \text{ GeV}/c^2$  invariant mass). Description of the backgrounds can be found in Section 3.

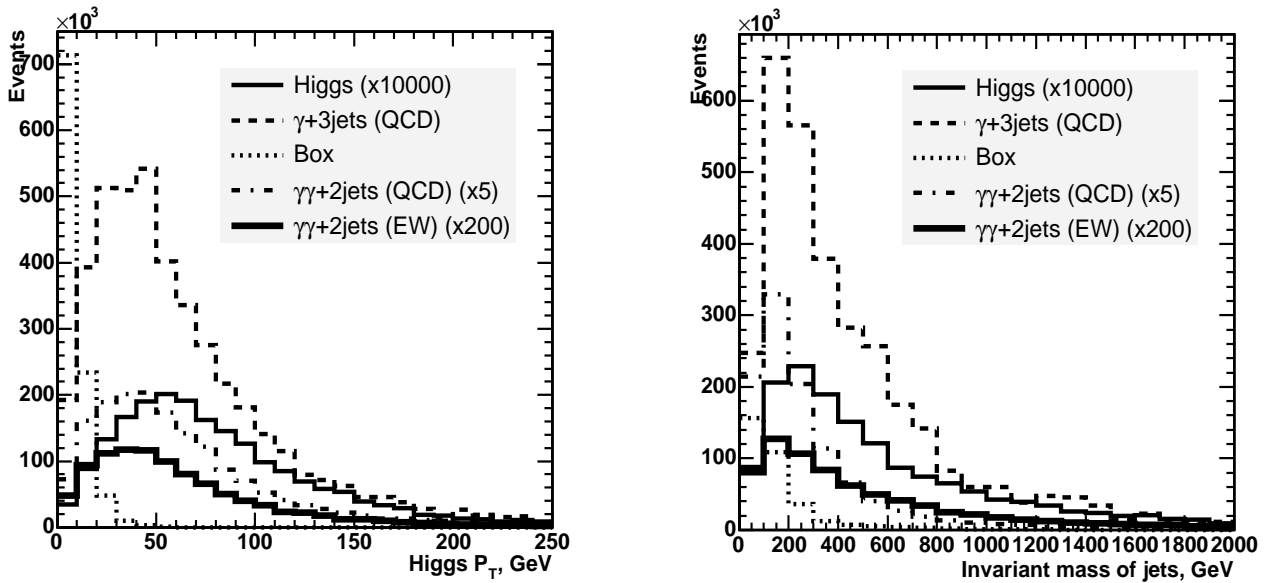


Figure 10: Two CompHEP background samples, gluon fusion background sample and Higgs signal sample are shown. Left plot - the transverse momentum of the Higgs boson, reconstructed from the photon energies and directions. Right plot - the invariant mass of two tagged jets. Plots were normalized to  $30 \text{ fb}^{-1}$  and Higgs signal was multiplied by 10000 ( $130 \text{ GeV}/c^2$  invariant mass). Description of the backgrounds can be found in Section 3.

Table 6: The fraction of the number of events in corresponding mass window left out of preselected events for the Higgs boson ( $m_H = 115 \text{ GeV}/c^2$ ) and all backgrounds after each selection cut is applied (mass window  $5.0 \text{ GeV}/c^2$ , ECAL threshold  $r = 0.01$ ), (1) - These two cuts are not applied in case of PYTHIA samples. First column for Higgs signal corresponds to PYTHIA study, when those two cuts are not applied, second column corresponds to CompHEP study, when those two cuts are applied.

Selection cuts	Higgs signal	Quark annihil.	Gluon fusion	$\gamma + \text{jets}$	QCD jets	DY	$\gamma + 3\text{jets}$	$\gamma\gamma + 2\text{jets}$	EW $\gamma\gamma + 2\text{jets}$
Cross section in mass window, fb	6.2	$4.5 \cdot 10^2$	$3.8 \cdot 10^2$	$5 \cdot 10^4$	$16.5 \cdot 10^4$	$1.5 \cdot 10^4$	$4.2 \cdot 10^4$	$2.4 \cdot 10^3$	5.6
L1 trig in mass window	91.9%	77%	79%	3.9%	1.5%	17.6%	20.8%	59.4%	67.7%
L1 match	90.9%	76.8%	78.8%	3.56%	1.1%	14%	15.5%	48.5%	52.3%
photon $E_t$ cuts	82.8%	50.2%	52.9%	2.5%	0.75%	11.6%	8.0%	37%	41.7%
both track iso	64.9%	37%	37.8%	0.3%	0.075%	0%	0.46%	22.4%	23.8%
both ecal iso	52%	28.5%	27.8%	0.14%	0.03%	0%	0.1%	16.5%	18.3%
$P_T^H > 50 \text{ GeV}/c$	36.6%	0%	0.15%	0.013%	0.009%	0%	0%	8.1%	10.7%
$HCAL/ECAL < 0.1$	36.6%	0%	0.15%	0.013%	0.008%	0%	0%	8.0%	10.7%
Jet tagging	15.2%	0%	0.06%	0%	0%	0%	0%	0.40%	4%
$P_T^{\text{jet}1} > 50 \text{ GeV}/c$	9.4%	0%	0%	0%	0%	0%	0%	0.13%	3.7%
$P_T^{\text{jet}2} > 35 \text{ GeV}/c$ (1)	9.4%/6.6%	0%	0%	0%	0%	0%	0%	0.04%	3.1%
Jet mass $> 500 \text{ GeV}/c^2$	8.4%/6.1%	0%	0%	0%	0%	0%	0%	0.02%	2.8%
Geom. cut(1)	8.4%/5.3%	0%	0%	0%	0%	0%	0%	0.02%	2.7%

## 6.1 Scenario I: Rescaled PYTHIA

Keeping in mind that there is a limited statistics for  $\gamma + \text{jets}$  and QCD hadronic jets background samples, but the number of events passed through all selection cuts without mass window restriction is small - one event survived through the cuts from QCD hadronic jets background sample and two events survived through the cuts from  $\gamma + \text{jets}$  background sample.

Next logical assumption here to suppose that this small number of events passed through the analysis, can be found in each mass window, surrounding each Higgs boson invariant mass with equal probability which is equal the whole mass interval ( $110 \text{ GeV}/c^2 - 170 \text{ GeV}/c^2$ ) divided by mass window ( $5 \text{ GeV}/c^2$ ). This probability will be equal to  $1/12$  for the purposes of this study.

Results can be seen in Table 12 where each Higgs boson invariant mass has the same value assigned - one event normalized to  $30 \text{ fb}^{-1}$  and multiplied by aforementioned probability  $1/12$ .

In this case, the Higgs boson signal and significance, normalized to  $30 \text{ fb}^{-1}$ , are shown in Table 12. Significance was calculated, based on the formula:  $S = 2(\sqrt{N_s + N_b} - \sqrt{N_b})$ .

The significance, vs Higgs boson invariant mass, normalized to  $30 \text{ fb}^{-1}$ , for this scenario can be found with label "rescaled PYTHIA" on Figure 12 (left plot). This curve has error bars which reflected the systematic uncertainties based on the hard scale uncertainties, PDF uncertainties, generator-level preselection uncertainties, jet scale and tracking efficiency uncertainties, described in the Appendix.

The integrated luminosity needed to reach  $5\sigma$  discovery, vs Higgs boson invariant mass is shown as a second line in Table 14 and can be found as a curve with label "rescaled PYTHIA" on Figure 12 (right plot). This curve has error bars which reflected the systematic uncertainties based on the hard scale uncertainties, PDF uncertainties, generator-level preselection uncertainties and jet scale and tracking efficiency uncertainties, described in the Appendix.

Results Obtained are in a good agreement with results from previous study, which were  $40\text{-}45 \text{ fb}^{-1}$  for  $5\sigma$  discovery reach in Higgs boson invariant mass range  $110 - 140 \text{ GeV}/c^2$ .

Table 7: The fraction of the number of events in corresponding mass window left out of preselected events for the Higgs boson ( $m_H = 120 \text{ GeV}/c^2$ ) and all backgrounds after each selection cut is applied (mass window  $5.0 \text{ GeV}/c^2$ , ECAL threshold  $r = 0.01$ ). (1) - These two cuts are not applied in case of PYTHIA samples. First column for Higgs signal corresponds to PYTHIA study, when those two cuts are not applied, second column corresponds to CompHEP study, when those two cuts are applied.

Selection cuts	Higgs signal	Quark annihil.	Gluon fusion	$\gamma + \text{jets}$	QCD jets	DY	$\gamma + 3\text{jets}$	$\gamma\gamma + 2\text{jets}$	EW $\gamma\gamma + 2\text{jets}$
Cross section in mass window, fb	6.3	$4.0 \cdot 10^2$	$3.0 \cdot 10^2$	$4 \cdot 10^4$	$15.6 \cdot 10^4$	$1.2 \cdot 10^4$	$3.8 \cdot 10^4$	$2.2 \cdot 10^3$	5.6
L1 trig in mass window	91.7%	77.4%	80%	4.2%	1.4%	15.5%	20.3%	60.6%	68.2%
L1 match	90%	77.2%	79.8%	3.9%	1.1%	12.5%	15.1%	48%	48.8%
photon $E_t$ cuts	83.6%	50.8%	56%	2.9%	0.68%	10.5%	7.5%	37.8%	36.7%
both track iso	65%	38.1%	42.1%	0.29%	0.06%	0.35%	0.7%	22.6%	20.1%
both ecal iso	52.5%	29.2%	32%	0.11%	0.019%	0.35%	0.35%	16.2%	14.6%
$P_T^H > 50 \text{ GeV}/c$	37.2%	0%	0.18%	0.011%	0.011%	0%	0.12%	8.4%	9.3%
$HCAL/ECAL < 0.1$	37.1%	0%	0.18%	0.0088%	0.0075%	0%	0.12%	8.3%	9%
Jet tagging	16%	0%	0%	0.003%	0%	0%	0.12%	0.43%	3.4%
$P_T^{\text{jet}1} > 50 \text{ GeV}/c$	10.0%	0%	0%	0%	0%	0%	0.12%	0.19%	3.3%
$P_T^{\text{jet}2} > 35 \text{ GeV}/c$ (1)	10.0%/7.2%	0%	0%	0%	0%	0%	0%	0.15%	2.44%
Jet mass $> 500 \text{ GeV}/c^2$	9.0%/6.5%	0%	0%	0%	0%	0%	0%	0.09%	3.3%
Geom. cut(1)	9.0%/5.7%	0%	0%	0%	0%	0%	0%	0.06%	1.8%

## 6.2 Scenario II: CompHEP

The next logical step is to use CompHEP generated background samples to double check the PYTHIA generated samples.

The usual  $2 \rightarrow 2$  subprocesses in PYTHIA contain part of NLO contributions if ISR and FSR are switched on, but this contribution is adequate only in cases when isolation of the emitted photon or gluon from quark or gluon is not very large. If this isolation is required to be large enough, PYTHIA underestimates this contribution. For the purposes of this study, this part of NLO contribution will be called ‘‘PYTHIA NLO’’.

CompHEP datasets have  $2 \rightarrow 4$  subprocesses (see Section 2.2 for details) with further FSR and ISR switched on in PYTHIA during the hadronization. These CompHEP complete set of diagrams can be considered as NLO for the usual PYTHIA  $2 \rightarrow 2$  subprocesses and called ‘‘PYTHIA NNLO’’ in terms of previous paragraph. Approximations are not used in the calculation of NNLO given by complete  $2 \rightarrow 4$  tree level sets of diagrams (see section 2.2).

Keeping in mind that CompHEP cannot calculate box-like diagrams, and based on matching study from Section 2, three background samples - quark annihilation,  $\gamma + \text{jets}$  and QCD hadronic jets - can be replaced by four CompHEP calculated background samples -  $\gamma + 3\text{jets}$  and  $\gamma\gamma + 2\text{jets}$  (QCD and electroweak). This replacement will change the discovery picture substantially.

### 6.2.1 $\gamma + 3\text{jets}$ QCD and electroweak backgrounds estimation

Due to the limited statistics, rescaling and factorization will be applied to reduce statistical errors. The rescaling procedure was described in Scenario I.

The factorization procedure was done as follows (see Table 11). All analysis was split on three different subsets of cuts:

- 1) all preliminary cuts were done to make sure that we have two photon candidates and two jets in event, passed through the requirements : to have two L1 electromagnetic isolated trigger, to have more than two photons matched with L1 trigger objects, to have more than two jets and to pass through jet tagging procedure. This subset of cuts will be mentioned as ‘‘prep. cuts’’ in Table 11.
- 2) isolation cuts - all cuts including photon  $E_t$  cuts, tracker and ECAL isolation for both photons,  $P_T^H > 50 \text{ GeV}/c$ ,  $HCAL/ECAL < 0.1$ . All these cuts will be mentioned in Table 11 as ‘‘isol. cuts’’.

Table 8: The fraction of the number of events in corresponding mass window left out of preselected events for the Higgs boson ( $m_H = 130 \text{ GeV}/c^2$ ) and all backgrounds after each selection cut is applied (mass window  $5.0 \text{ GeV}/c^2$ , ECAL threshold  $r = 0.01$ ). (1) - These two cuts are not applied in case of PYTHIA samples. First column for Higgs signal corresponds to PYTHIA study, when those two cuts are not applied, second column corresponds to CompHEP study, when those two cuts are applied.

Selection cuts	Higgs signal	Quark annihil.	Gluon fusion	$\gamma + \text{jets}$	QCD jets	DY	$\gamma + 3\text{jets}$	$\gamma\gamma + 2\text{jets}$	EW $\gamma\gamma + 2\text{jets}$
Cross section in mass window, fb	6.4	$3.0 \cdot 10^2$	$2.1 \cdot 10^2$	$3.3 \cdot 10^4$	$13.3 \cdot 10^4$	$8.8 \cdot 10^3$	$3.6 \cdot 10^4$	$1.9 \cdot 10^3$	4.2
L1 trig in mass window	91.6%	77.7%	82%	4.3%	1.7%	13.5%	20.6%	60.3%	67.5%
L1 match	88.9%	77.6%	81.8%	3.9%	1.1%	9.3%	15.4%	48.5%	51.6%
photon $E_t$ cuts	84.1%	57%	63%	3.0%	0.85%	7.5%	9.1%	40.1%	41.4
both track iso	65.2%	42.2%	47.4%	0.36%	0.07%	0%	0.42%	23.6%	22.7%
both ecal iso	53%	31.7%	36.1%	0.17%	0.018%	0%	0.1%	17.2%	17.1
$p_T^H > 50 \text{ GeV}/c$	38.2%	0.03%	0.27%	0.017%	0.006%	0%	0%	9.4%	10%
$HCAL/ECAL < 0.1$	38.1%	0.03%	0.27%	0.017%	0.004%	0%	0%	9.3%	10%
Jet tagging	16.5%	0.03%	0%	0%	0%	0%	0%	0.45%	4.1%
$p_T^{\text{jet}1} > 50 \text{ GeV}/c$	10.6%	0%	0%	0%	0%	0%	0%	0.06%	3.6%
$p_T^{\text{jet}2} > 35 \text{ GeV}/c$ (1)	10.6%/7.0%	0%	0%	0%	0%	0%	0%	0.03%	2.6%
Jet mass $> 500 \text{ GeV}/c^2$	9.3%/8.6%	0%	0%	0%	0%	0%	0%	0.03%	2.6%
Geom. cut(1)	9.3%/7.3%	0%	0%	0%	0%	0%	0%	0.015%	2.6%

3) jet cuts - all cuts including  $p_T^{\text{jet}1} > 50 \text{ GeV}/c$ ,  $p_T^{\text{jet}2} > 35 \text{ GeV}/c$ ,  $M_{j_1 j_2} > 500 \text{ GeV}/c^2$ , and geometry cut:  $\min[\eta_{\text{jet}1}, \eta_{\text{jet}2}] + 0.7 < \eta_{\gamma 1,2} < \max[\eta_{\text{jet}1}, \eta_{\text{jet}2}] - 0.7$ .

After this splitting, we will run all events through first subset of cuts (preliminary cuts, described above) and will pass through last two subset separately to increase the statistics of events passed through the cuts.

Based on QCD samples  $\gamma+3\text{jets}$  and  $\gamma\gamma+2\text{jets}$  we can estimate rejection factors for the photon to be a photon and for jet to be a photon.

From QCD  $\gamma\gamma+2\text{jets}$  sample:  $(r_{\gamma \rightarrow \gamma}^{\text{QCD}})^2 = 847/192227$ . From  $\gamma+3\text{jets}$  QCD sample:  $(r_{\gamma \rightarrow \gamma}^{\text{QCD}}) * (r_{\text{jet} \rightarrow \gamma}^{\text{QCD}}) = 16/37559$ .

From these two equations we have got next values for  $r_{\gamma \rightarrow \gamma}^{\text{QCD}}$  and  $r_{\text{jet} \rightarrow \gamma}^{\text{QCD}}$ :

$$r_{\gamma \rightarrow \gamma}^{\text{QCD}} = 0.066, \quad r_{\text{jet} \rightarrow \gamma}^{\text{QCD}} = 6.45 * 10^{-3}$$

Keeping this in mind, we can get  $r_{\gamma \rightarrow \gamma}^{\text{EW}}$  value from  $\gamma\gamma + 2\text{jets}$  electroweak CompHEP background sample:  $(r_{\gamma \rightarrow \gamma}^{\text{EW}})^2 = 1000/41000$ , hence  $r_{\gamma \rightarrow \gamma}^{\text{EW}} = 0.156$ .

To be able to estimate electroweak part of  $\gamma+3\text{jets}$  ComHEP background, we need to know  $r_{\text{jet} \rightarrow \gamma}^{\text{EW}}$ .

**The main assumption here is  $r_{\text{jet} \rightarrow \gamma}^{\text{EW}} = r_{\text{jet} \rightarrow \gamma}^{\text{QCD}} = 6.45 * 10^{-3}$ .**

The rest of the jet cuts give us additional factor at the level from 0.084 to 0.15. We will use  $r_{\text{jet}}^{\text{EW}} = 0.1$ .

Now we can calculate an average number of events in mass window 5 GeV (rescaling probability will be again 1/12 as above in Scenario D) around each invariant mass of Higgs boson. Keeping in mind that total cross section of  $\gamma+3\text{jets}$  CompHEP background will be 5.15 pb, and normalizing to  $30 \text{ fb}^{-1}$ , we can count the events:

$$N_{ev} = 5150 \text{ fb} * 30 \text{ fb}^{-1} * r_{\text{jet} \rightarrow \gamma}^{\text{EW}} * r_{\gamma \rightarrow \gamma}^{\text{EW}} * r_{\text{jet}}^{\text{EW}} * 1/12 = 5000 \text{ fb} * 30 \text{ fb}^{-1} * 0.156 * 6.45 * 10^{-3} * 0.1 * 1/12 = 15/12 \approx 1.25 \text{ ev}$$

So, in a total, we have an upper limit on the number of events in 5 GeV mass window of 1.25 events.

Based on this digit, we will estimate the upper limit on the number of events in  $\gamma + 3\text{jets}$  QCD CompHEP background. Looking at the Table 13, we will make **second assumption - the ratio of events between  $\gamma\gamma+2\text{jets}$  QCD**

Table 9: The fraction of the number of events in corresponding mass window left out of preselected events for the Higgs boson ( $m_H = 140 \text{ GeV}/c^2$ ) and all backgrounds after each selection cut is applied (mass window  $5.0 \text{ GeV}/c^2$ , ECAL threshold  $r = 0.01$ ). (1) - These two cuts are not applied in case of PYTHIA samples. First column for Higgs signal corresponds to PYTHIA study, when those two cuts are not applied, second column corresponds to CompHEP study, when those two cuts are applied.

Selection cuts	Higgs signal	Quark annihil.	Gluon fusion	$\gamma + \text{jets}$	QCD jets	DY	$\gamma + 3\text{jets}$	$\gamma\gamma + 2\text{jets}$	EW $\gamma\gamma + 2\text{jets}$
Cross section in mass window, fb	6.0	$2.4 \cdot 10^2$	$1.6 \cdot 10^2$	$2.7 \cdot 10^4$	$11.5 \cdot 10^4$	$8.4 \cdot 10^3$	$3.1 \cdot 10^4$	$1.6 \cdot 10^3$	4.2
L1 trig in mass window	92.2%	78.4%	82.1%	4.3%	1.7%	10.8%	20.9%	61.8%	69.9%
L1 match	92.0%	78.3%	82.0%	3.9%	1.3%	6.0%	14%	48.2%	51.5
photon $E_t$ cuts	88.1%	58.8%	62.2%	3%	1%	5%	7.1%	40%	40.6%
both track iso	69.1%	43.1%	44.6%	0.28%	0.058%	0%	0.28%	22.7%	25.1%
both ecal iso	56.0%	32.8%	33.9%	0.12%	0.019%	0%	0.14%	16.9%	19.5%
$P_T^H > 50 \text{ GeV}/c$	39.6%	0.23%	0.24%	0.02%	0.009%	0%	0.14%	10.3%	13.2%
$HCAL/ECAL < 0.1$	39.6%	0.15%	0.24%	0.02%	0.009%	0%	0%	10.2%	12.6%
Jet tagging	17.8%	0.038%	0%	0%	0%	0%	0%	0.35%	6%
$P_T^{\text{jet}1} > 50 \text{ GeV}/c$	11.8%	0%	0%	0%	0%	0%	0%	0.25%	5.5%
$P_T^{\text{jet}2} > 35 \text{ GeV}/c$ (1)	11.8%/10.2%	0%	0%	0%	0%	0%	0%	0.09%	4.3%
Jet mass $> 500 \text{ GeV}/c^2$	10.6%/9.2%	0%	0%	0%	0%	0%	0%	0.074%	4.1%
Geom. cut(1)	10.6%/8.1%	0%	0%	0%	0%	0%	0%	0%	3.6%

and EW will be the same as  $\gamma+3\text{jets}$  QCD and EW. This assumption is a bit more questionable than the previous one, but in this case we will again get larger numbers of events through this estimation than if we would produce enough statistics, because, as can be seen in Tables 6-10,  $\gamma+3\text{jets}$  QCD background will be suppressed remarkably well and our procedure, likely, overestimates the number of events passed through the analysis.

Based on this assumption, we will have :

- $N_{ev}(115 \text{ GeV}, \gamma+3\text{jets QCD}) = 1.25 * 5.4/5 \approx 1.35$
- $N_{ev}(120 \text{ GeV}, \gamma+3\text{jets QCD}) = 1.25 * 6.6/3.5 \approx 2.35$
- $N_{ev}(130 \text{ GeV}, \gamma+3\text{jets QCD}) = 1.25 * 4.9/3.5 \approx 1.75$
- $N_{ev}(140 \text{ GeV}, \gamma+3\text{jets QCD}) = 1.25 * 4.4/5.5 \approx 1.0$
- $N_{ev}(150 \text{ GeV}, \gamma+3\text{jets QCD}) = 1.25 * 2.3/1 \approx 2.9$

The data itself will be used to estimate and cross-check the total background at very high energy. If Higgs boson is discovered, the number of events in the side-bands of the Higgs boson and their mass dependence will be used to estimate the number of background events under the peak, provided there is enough statistics in the side-bands. In this approach, the uncertainties on the background cross-sections, the PDF and on the luminosity measurement are highly reduced. For Higgs boson production the two regions left and right of the Higgs boson can be used. We estimate a reduction of the signal significance of the order of 10-20%. The significance, vs Higgs boson invariant mass, for side-band method of background estimation, normalized to  $30 \text{ fb}^{-1}$ , can be found on Figure 11 (left plot).

The integrated luminosity needed to reach  $5\sigma$  discovery, vs Higgs boson invariant mass, for side-band method of background estimation, can be found on Figure 11 (right plot).

Higgs boson signal and significance, normalized to  $30 \text{ fb}^{-1}$ , are shown in Table 13. Significance was calculated, based on the formula:  $S = 2(\sqrt{N_s + N_b} - \sqrt{N_b})$ .

The significance, vs Higgs boson invariant mass, normalized to  $30 \text{ fb}^{-1}$ , for this scenario can be found as the lower curve with label "CompHEP" on Figure 12 (left plot).

The integrated luminosity needed to reach  $5\sigma$  discovery, vs Higgs boson invariant mass, is shown as a third line in Table 14 and can be found as a higher curve with label "CompHEP" on Figure 12 (right plot).

Table 10: The fraction of the number of events in corresponding mass window left out of preselected events for the Higgs boson ( $m_H = 150 \text{ GeV}/c^2$ ) and all backgrounds after each selection cut is applied (mass window  $5.0 \text{ GeV}/c^2$ , ECAL threshold  $r = 0.01$ ). (1) - These two cuts are not applied in case of PYTHIA samples. First column for Higgs signal corresponds to PYTHIA study, when those two cuts are not applied, second column corresponds to CompHEP study, when those two cuts are applied.

Selection cuts	Higgs signal	Quark annihil.	Gluon fusion	$\gamma$ + jets	QCD jets	DY	$\gamma$ + 3jets	$\gamma\gamma$ + 2jets	EW $\gamma\gamma$ + 2jets
Cross section in mass window, fb	3.5	$2.1 \cdot 10^2$	$1.2 \cdot 10^2$	$2.3 \cdot 10^4$	$10 \cdot 10^4$	$6.2 \cdot 10^3$	$2.7 \cdot 10^4$	$1.3 \cdot 10^3$	3.3
L1 trig in mass window	92.8%	79.7%	78%	4.4%	1.7%	9.8%	23.1%	59.4%	68%
L1 match	92.6%	79.5%	77.8%	4%	1.4%	6.7%	15.2%	47.5%	48.9%
photon $E_t$ cuts	89.2%	62.5%	64.6%	3.4%	1%	4.2%	10%	40.7%	41.4%
both track iso	70.9%	45.1%	47.6%	0.25%	0.11%	0%	0.17%	21.3%	20.4%
both ecal iso	58.6%	35.5%	35.9%	0.14%	0.02%	0%	0.17%	15.9%	16.3%
$P_T^H > 50 \text{ GeV}/c$	42.3%	0%	0.5%	0.02%	0.005%	0%	0.17%	10.3%	9.4%
$HCAL/ECAL < 0.1$	42.3%	0%	0.5%	0.02%	0.0025%	0%	0.17%	10.2%	9.4%
Jet tagging	18.7%	0%	0.15%	0%	0%	0%	0.17%	0.33%	2.3%
$P_T^{\text{jet}1} > 50 \text{ GeV}/c$	16.1%	0%	0%	0%	0%	0%	0%	0.15%	1.9%
$P_T^{\text{jet}2} > 35 \text{ GeV}/c$ (1)	16.1%/11.2%	0%	0%	0%	0%	0%	0%	0.076%	1.4%
Jet mass $> 500 \text{ GeV}/c^2$	11.1%/10.2%	0%	0%	0%	0%	0%	0%	0.038%	1.2%
Geom. cut(1)	11.1%/8.9%	0%	0%	0%	0%	0%	0%	0.019%	0.8%

Table 11: Factorization procedure for all CompHEP-produced background samples, no mass window

	$\gamma\gamma+2\text{jets EW}$	$\gamma\gamma+2\text{jets QCD}$	$\gamma+3\text{jets QCD}$
Total Nev	41000	192227	37559
Prep.cuts	4938	7359	1558
Isol cuts	1000	847	16
Jet cuts	2019	617	228

It would be useful to emphasize here, that the largest CompHEP background sample  $\gamma + 3\text{jets}$  was suppressed remarkably well and future study is needed to reduce statistical errors.

## 7 Conclusion

A study of CMS Higgs discovery potential using Higgs boson decays into two photons for the vector boson fusion production mode, is presented.

Isolation cuts, based on the tracker and ECAL information, are important tools to reject the background, while keeping the Higgs boson signal efficiency reasonably high.

The forward jet tagging technique is very efficient in further reducing the background in the vector boson fusion production case. With this technique it is possible to reduce the remaining background events to a level well below the number of expected Higgs boson events.

There are two scenarios studied in this note - based on PYTHIA produced background samples and background samples, produced with CompHEP. PYTHIA significantly underestimates QCD and electroweak backgrounds for this particular channel.

Further improvement can be achieved with more Monte Carlo events produced for backgrounds estimation, especially in the CompHEP case.

Additional improvements might be achieved by using photon conversion and pizero rejection, but these techniques are not used in this note.

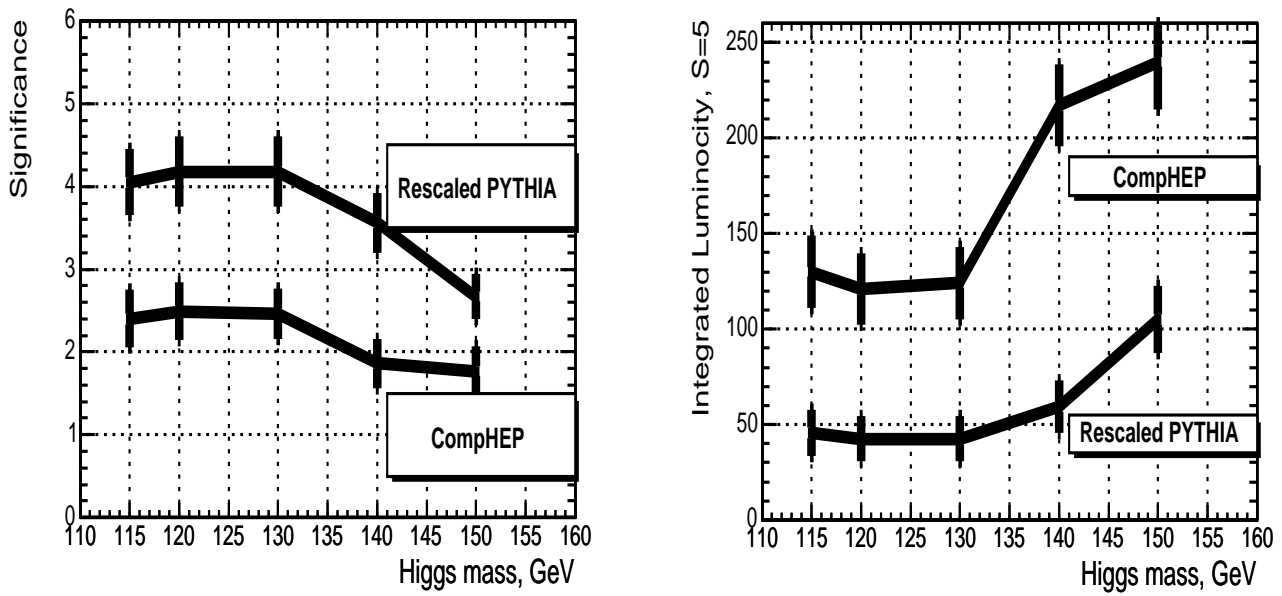


Figure 11: Significance vs Higgs boson invariant mass (left), normalized to  $30 \text{ fb}^{-1}$  and integrated luminosity, needed to reach  $5\sigma$  significance for the two different scenarios described in Section 6. Error bars correspond to side-band method of uncertainties estimation.

Table 12: Higgs boson signal for different masses and significance, normalized to  $30 \text{ fb}^{-1}$  (mass window  $5.0 \text{ GeV}/c^2$ ). Scenario I: Rescaled PYTHIA.

	$m_H = 115$ $\text{GeV}/c^2$	$m_H = 120$ $\text{GeV}/c^2$	$m_H = 130$ $\text{GeV}/c^2$	$m_H = 140$ $\text{GeV}/c^2$	$m_H = 150$ $\text{GeV}/c^2$
$N_s$	12.0	12.5	12.5	10.1	7.0
quark annihilation	0	0	0	0	0
gluon fusion	0	0	0	0	0
$\gamma$ +jets	1.8	1.8	1.8	1.8	1.8
QCD hadronic jets	2.0	2.0	2.0	2.0	2.0
Drell Yan	0	0	0	0	0
$N_b$	3.8	3.8	3.8	3.8	3.8
$S$	4.05	4.18	4.18	3.56	2.67

Table 13: Higgs boson signal for different masses and significance, normalized to  $30 \text{ fb}^{-1}$  (mass window  $5.0 \text{ GeV}/c^2$ ). Scenario II: CompHEP.

	$m_H = 115$ $\text{GeV}/c^2$	$m_H = 120$ $\text{GeV}/c^2$	$m_H = 130$ $\text{GeV}/c^2$	$m_H = 140$ $\text{GeV}/c^2$	$m_H = 150$ $\text{GeV}/c^2$
$N_s$	10.11	10.55	9.55	7.84	5.62
gluon fusion	0	0	0	0	0
$\gamma$ +3jets (QCD)	1.35	2.35	1.75	1.0	2.9
$\gamma$ +3jets (EW)	1.25	1.25	1.25	1.25	1.25
$\gamma\gamma$ + 2jets (QCD)	5.4	6.62	4.93	4.44	2.3
$\gamma\gamma$ + 2jets (EW)	5	3.5	3.5	5.5	1.0
Drell Yan	0	0	0	0	0
$N_b$	13.0	13.12	10.7	14.10	7.45
$S$	2.40	2.49	2.46	1.86	1.77

Table 14: Integrated luminosity needed to reach  $5\sigma$  discovery limit for two scenarios described in Section 6.

	$m_H = 115$ GeV/ $c^2$	$m_H = 120$ GeV/ $c^2$	$m_H = 130$ GeV/ $c^2$	$m_H = 140$ GeV/ $c^2$	$m_H = 150$ GeV/ $c^2$
Scenario I: Rescaled PYTHIA	45.7	42.5	42.5	59.2	105
Scenario II: CompHEP	130	121	124	217	240

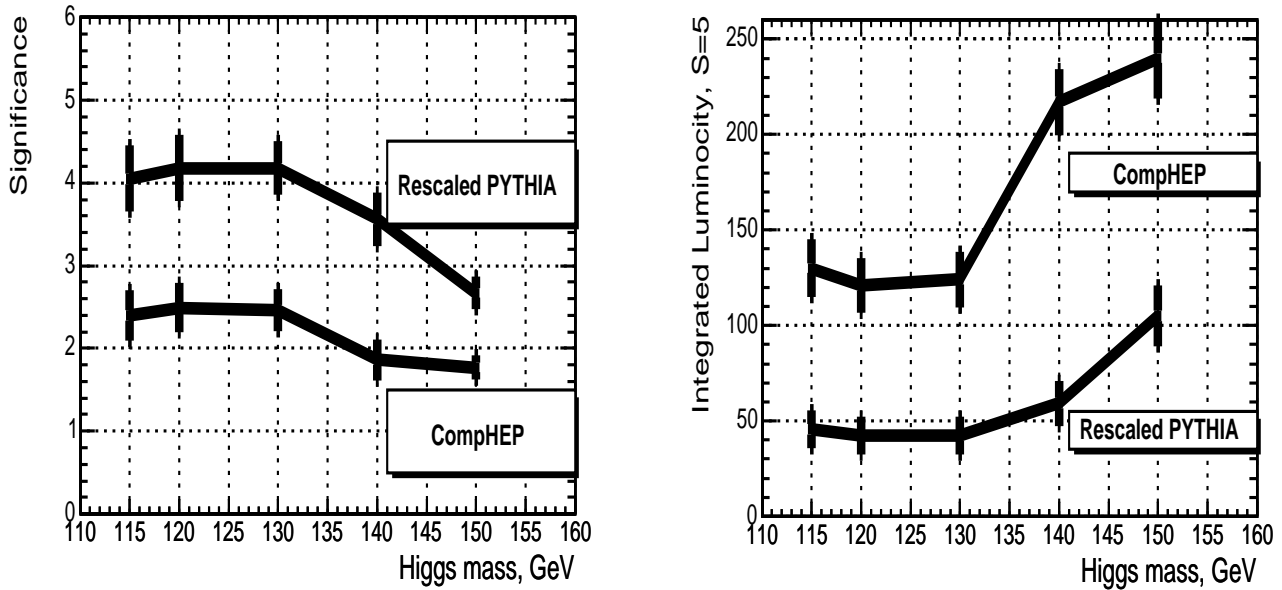


Figure 12: Significance vs Higgs boson invariant mass (left), normalized to  $30 \text{ fb}^{-1}$  and integrated luminosity, needed to reach  $5\sigma$  significance for the two different scenarios described in Section 6. Error bars correspond to all four sources of uncertainties, described in Appendix B.

## 8 Acknowledgements

We would like to thank A. Nikitenko and T. Sjostrand for useful discussions.

## 9 Appendix A

The LHAGLUE interface [12] included from the most recent LHAPDF versions simplifies the use of the Les Houches accord PDF in PYTHIA by the switches  $MSTP(52) = 2$ ,  $MSTP(51) = LHAPDF_{id}$ , where  $LHAPDF_{id}$  is documented in [13].

There are 40 fits for CTEQ6M with  $MSTP(51) = 10050$  is equivalent for the main fit and  $MSTP(51) = 10051 - 10090$  for 40 additional weights. Lets  $X(S)$  be any variable which depends on PDFs. We are calculating the PDF “best fit”,  $X_0 = X(S_0)$ , and the rest of the fits  $X_k^\pm = X(S_k^\pm)$ , where

$$D_k = X_k^+ - X_k^-, k = 1, \dots, d, \quad D_k = (X_k^+ - X_0) - (X_k^- - X_0)$$

$$R_j, j = 1, \dots, 2d, \quad R_1 = X_1^+ - X_0, R_2 = X_1^{-1} - X_0, R_3 = X_2^+ - X_0, R_4 = X_2^- - X_0, \dots$$

There are two major equations to calculate uncertainties [11]:

$$\Delta X = \frac{1}{2} \sqrt{\sum_{i=1}^{N_{PDF}} [X_{2i} - X_{2i-1}]^2}, \quad \Delta X_C = \frac{1}{2} \sqrt{\sum_{i=1}^{2N_{PDF}} [X_i - X_0]^2}$$

## 10 Appendix B : Cross section uncertainties

### 10.1 Hard process scale

Usually, the hard process defines the  $Q^2$  scale, which directly influences the parametrization of PDF structure functions and  $\alpha_s$  which, in turn, influences the cross-section.

The dependence of the observables on the choice of the  $Q^2$  scale is unphysical and should be treated as one of the contributions to the total uncertainty of the theoretical predictions.

The usual choice for the hard process scale in  $2 \rightarrow 1$  processes is often  $\hat{s}$ , which is the default in PYTHIA. In order to study the effect of hard process scale variation for  $2 \rightarrow 1$  processes, two scenarios were used, corresponding to  $0.25\hat{s}$  and  $4.0\hat{s}$  [11]. Results for the most important background and signals are presented in Table 15.

Table 15: Hard scale uncertainties for the signal samples as well as for the background samples.

	box	born	brem	QCD	DY
$4\hat{s}$	-12.7%	-21.3%	-0.6%	-3.2%	-24.3%
$0.25\hat{s}$	+17.0%	+19.1%	+0.26%	+3.4%	+23%
	115 GeV/c <sup>2</sup>	120 GeV/c <sup>2</sup>	130 GeV/c <sup>2</sup>	140 GeV/c <sup>2</sup>	150 GeV/c <sup>2</sup>
$4\hat{s}$	-2.98%	-2.73%	-0.79%	-0.98%	-0.22%
$0.25\hat{s}$	+2.55%	+3.13%	+0.34%	+2.24%	+0.32%

### 10.2 PDF uncertainties

The parton distribution functions of interacting particles describe the probability density for partons undergoing hard scattering at the hard process scale  $Q^2$  and taking a certain fraction  $x$  of the total particle momentum.

Various approaches are currently available to quote the PDFs of the proton, but CTEQ6M and MRST PDFs seem to be well suited for use in Monte Carlo simulations for the LHC [11] (see also Appendix A).

$\Delta X_C$  estimator of PDF uncertainties (see Appendix A) will be used to estimate the cross-section uncertainties in this analysis.

PDF uncertainties for all signals and PYTHIA backgrounds can be found in Table 16

Table 16: PDF uncertainties for the selected signal samples as well as for selected background samples.

Signal	115 GeV/c <sup>2</sup>	120 GeV/c <sup>2</sup>	130 GeV/c <sup>2</sup>	140 GeV/c <sup>2</sup>	150 GeV/c <sup>2</sup>
CTEQ6M, $\Delta X_C/X_0$	5.37%	7.1%	5.19%	5.96%	5.51%
Backgrounds	born	box	brem	DY	QCD
CTEQ6M, $\Delta X_C/X_0$	4.76%	7.28%	3.69%	6.42%	4.1%

### 10.3 Generator-level preselection uncertainties

Preselection inefficiency will be less than 3% to preselect  $\gamma + jet$  events and less than 5% to preselect QCD hadronic jet events. This 3-5% inefficiency for preselected events will be propagated to the 3-5% of corresponding cross-section uncertainties.

### 10.4 Other uncertainties

Taking into account other uncertainties like jet scales, tracking efficiency, alignment for the ECAL and Tracker, specified in Physics TDR Vol I, total uncertainty due to these reasons will be less than 4% for QCD and  $\gamma + 3jets$  CompHEP background samples and less than 2.5% for quark annihilation, gluon fusion,  $\gamma\gamma + 2jets$  CompHEP background samples.

Jet scale uncertainties and tracker efficiency uncertainties were taken into account and estimated at the level of 5%.

## References

- [1] **CMS TN/92-56**, C. Seez and J. Virdee, “*Detection of an intermediate mass Higgs boson at LHC via its two photon decay mode*”.
- [2] **CERN/LHCC 97-33**, CMS Collaboration, “*The Electromagnetic Calorimeter Project - Technical Design Report*”.
- [3] **CMS Note 1997/101**, M.N. Dubinin et. al., “*Light Higgs Boson Signal at LHC in the Reactions  $pp \rightarrow \gamma\gamma + jet$  and  $pp \rightarrow \gamma\gamma + lepton$* ”.
- [4] T. Sjöstrand, *Comput. Phys. Commun.* **82** (1994) 74;  
T. Sjöstrand et al., *Comput. Phys. Commun.* **135** (2001) 238, LU TP 00-30, hep-ph/0010017.
- [5] CTEQ Collaboration, H.L. Lai et al., “*Global QCD Analysis of Parton Structure of the Nucleon: CTEQ5 Parton Distributions*”, *Eur. Phys. J.* **C12** (2000) 375.
- [6] M.Spira, A.Djouadi, D.Graudenz, P.Zerwas, *Nucl.Phys.* **B453** (1995) 17.
- [7] CMS Collaboration, **CERN/LHCC 2000-038**, **CMS TDR 6.1**, “*CMS The Tridas Project Design Report, Volume 1: The Trigger Systems*”.
- [8] CMS Collaboration, **CERN/LHCC 2002-02**, **CMS TDR 6.2**, “*CMS The Tridas Project Technical Design Report, Volume 2: Data acquisition and High-Level Trigger*”.
- [9] CMS Collaboration, “*Object oriented Simulation for CMS Analysis and Reconstruction*”, <http://cmsdoc.cern.ch/oscar/>.
- [10] CMS Collaboration, “*CMS OO Reconstruction*”, <http://cmsdoc.cern.ch/orca/>.
- [11] P. Bartalini, R.Chierici, A De Roeck, “*Guidelines for the estimation of theoretical uncertainties at the LHC*”, CMS Note 2005/013, 2005.
- [12] D.Bourilkov, “*The Les Houches Accord PDFs (LHAPDF) and Lhaglu*”, [hep-ph 0508110]
- [13] H. Baer et al [hep-ph 0403045]. Code and documentation available at <http://durpdg.dur.ac.uk/lhapdf>
- [14] **CMS Note 2001/022**, M. Dubinin, “*Higgs Boson Signal in the Reactions  $pp \rightarrow \gamma\gamma + 2forwardjet$* ”.

- [15] V. Tisserand, “*The Higgs to two photon decay in the ATLAS Detector*”, ATLAS Internal Note, **PHYS-NO-90** (1996).
- [16] D. Rainwater and D. Zeppenfeld, **JHEP** **12** (1997) 5.
- [17] D. Rainwater, D. Zeppenfeld and K. Hagiwara, Phys. Rev. **D 59** (1999) 014037.
- [18] D.Zeppenfeld, R.Kinnunen, A.Nikitenko, E.Richter-Was, Phys.Rev. **D62** (2000) 013009
- [19] T.Figy, C.Oleari, D.Zeppenfeld, Phys.Rev. **D68** (2003) 073005
- [20] E.Boos et.al., Nucl.Instr.Meth. **A534** (2004) 250 (hep-ph/0403113)  
A.Pukhov et.al., INP report INP-MSU-98-41-542 (hep-ph/9908288)  
( <http://theory.sinp.msu.ru/comphep>)
- [21] E. E. Boos, V. A. Ilyin and A. N. Skachkova, JHEP 0005 (2000) 052 (hep-ph/0004194)
- [22] D. Bardin et.al., in:Physics at LEP2, vol.2, CERN report 96-01A (hep-ph/9709270)  
M. Grunewald et.al., in:Reports of the working groups for precision calculations for LEP2 physics, CERN report 2000-09-A (hep-ph/0005309)
- [23] **CMS Note 2002/030**, V. Litvin, H. Newman, S. Shevchenko, N. Wisniewski, “*The rejection of background to the  $H \rightarrow \gamma\gamma$  process using isolation criteria based on information from electromagnetic calorimeter and tracker*”.
- [24] **CMS Note 2001/034**, E. Meschi *et al.*, “*Electron Reconstruction in the CMS Electromagnetic Calorimeter*”.
- [25] A.Djouadi, J.Kalinowski and M.Spira, Comp. Phys. Commun. **108** (1998) 56.
- [26] **CMS Note 1993/092**, C. Seez and T.S. Virdee, “*Using Tracks to Locate the two Photon Vertex*”.
- [27] **CMS Note 1993/115**, C. Seez, “*An Algorithm Using Tracks to Locate the two Photon Vertex*”.
- [28] **CMS Note 1995/115**, D.J. Graham, “*An algorithm using tracks to locate the two photon vertex at high luminosity*”.