Introduction

Amongst the properties that have to be investigated after the discovery of the Higgs Boson, the scalar potential of the Higgs field that is responsible for the electroweak symmetry breaking has to be fully defined. This potential can be written as follows:

\[ V_H = \mu^2 (\phi^+ \phi) + \frac{\lambda}{2} (\phi^+ \phi)^2 \quad \text{with} \quad \lambda = \frac{M_H^2}{v^2} \quad \text{and} \quad \mu^2 = -\frac{1}{2} M_H^2 \]

where \( v \) is the electroweak breaking scale and is known to be 246GeV.

If we rewrite this potential in terms of a physical Higgs Boson, we obtain the trilinear Higgs self-coupling \( \lambda_{\text{HHH}} \), which in the Standard Model is only related to the mass of the Higgs Boson:

\[ \lambda_{\text{HHH}} = \frac{3 M_H^2}{v} \]

This self-coupling can be observed when one Higgs resonance is created via the usual production channel, and then decays into two real Higgs bosons. Here, we will study the most probable production mode at LHC, which is the gluon fusion process.

![Figure 1: Feynman diagram of the double higgs production via gluon fusion](image)

To study the Higgs self coupling, we also have to choose the best decay channel. Here, we consider that one of the Higgs boson decays into two photons and the other one into a quark b and an antiquark \( \bar{b} \). The advantage of this approach is that we benefit from the good resolution of the \( H \to \gamma\gamma \) channel, and in the meantime we keep a satisfactory branching ratio with the \( H \to bb \) process.

The goal of this project is to evaluate the possibility to measure this Higgs self-coupling at LHC with a center of mass energy of 14TeV and a luminosity of 3000 fb\(^{-1}\).

First of all, the backgrounds which will be considered for this process can be separated in two categories:

- The QCD background made of \( \gamma\gamma+2\text{jets}, \gamma+\text{jet(fake photon)}+3\text{jets} \)
- The peaking backgrounds with production of a Higgs : ZH, WH, ggH and t\( \bar{t} \)H
Figure 2: Feynman diagrams for the W/ZH, ggH and ttH processes

The cross-sections of these processes at 8TeV and 14TeV are given below:

<table>
<thead>
<tr>
<th>Process</th>
<th>ZH</th>
<th>ttH</th>
<th>ggH</th>
<th>WH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross-section at 8TeV (pb)</td>
<td>0.41</td>
<td>0.13</td>
<td>19.12</td>
<td>0.7</td>
</tr>
<tr>
<td>Cross-section at 14TeV (pb)</td>
<td>0.88</td>
<td>0.61</td>
<td>49.9</td>
<td>1.5</td>
</tr>
</tbody>
</table>

In this study we use MC and Data samples from CMS at 8 TeV and 19.2 fb⁻¹, and then extrapolate at 14TeV and 3000 fb⁻¹ where it becomes more relevant, given the number of expected events.

**Photon and Jet Selection**

The selections applied on the photons are the ones that are commonly used in the H→γγ analysis. First, the photon of leading momentum is required to have $P_T$ above 40GeV and the one of subleading momentum to have $P_T$ above 25GeV. We also require the ratio of the PT of the first photon over $m_{\gamma\gamma}$ (the reconstructed mass of the two first photons) to be above ½, and the ratio of the $P_T$ of the second photon over $m_{\gamma\gamma}$ to be above 25/120. Besides, the selected photons are only the ones which have the absolute value of Eta < 2.5 and which pass the CiC selection (a cut based discriminator that enables to identify photons).

To study the jets, an improvement of 10% in $P_T$ resolution is emulated before proceeding to the jet selection. This is motivated by the work on b-jet energy regression which will lead to such an improvement. For the selection, a kinematic cut is applied by requiring the momentum of the two leading jets to be above 40 GeV and the absolute value of Eta to be under 2.1. To try to attack the ttH background, the jet multiplicity for the main peaking backgrounds is studied and it appears that a selection of the events with 2 or 3 jets enables to discriminate between the ttH events and the others, as shown below:

Figure 3: Stack Histogram of the number of jets for the signal and peaking backgrounds
Now, the jets out of which the mass of the dijet will be computed have to be chosen. Usually, the jets are ordered in PT and the mass is computed from the two leading jets. Here, the goal is to select the jets which have the highest probability to be b-jets. To do so, the jets have been ordered according to their probability of being a b-jets using the csv discriminator. In figure 4, it can be seen that the two leading jets actually contain most of the true b-jets (the b-matched jets). Nevertheless, by computing the invariant mass of the dijet for the two first jets in csv, one loses less true b-matched events than by selecting the first two jets in PT.

![Image](image_url)

**Figure 4**: Histograms showing the repartition of the Pt ranking for all the jets (bright violet) and for the true jets only (dark violet) for the first jet in CSV (upper left), the second jet in CSV (upper right), the third jet in CSV (lower left) and the fourth jet in CSV (lower right).

Having completed these selections, the strategy of the analysis will be the following:

- First, a cut will be applied on the dijet mass according to the optimization which will be described later
- Secondly, the γγ peak will be fitted and the known peaking backgrounds will be subtracted

The data at 8 TeV is first normalized to the number of events in the signal region (between 123 and 127 GeV) in the mγγ spectrum, and then used to estimate the QCD background in the dijet mass. All the events that pass the previous selection and which are not in the region of the signal in the mγγ spectrum are considered to be part of this QCD background. In other words, the region with 123 < mγγ < 127 GeV is excluded.

We also use the sidebands of mγγ to fit the background and find the number of QCD events in the signal region.
Figure 5: $M_{\gamma\gamma}$ for the data between 100 and 180 GeV with fit and excluded region between 123 and 127 GeV

Considering the QCD background, the shape of the dijet mass after the previous selection and some additional cuts explained later is shown below:

Figure 6: Di-jet mass after selection and b-tagging (T-M) with the QCD background renormalized to the $M_{\gamma\gamma}$ window

Here, it is visible that the signal and all the peaking backgrounds are dominated by the QCD background, but as this background will be fitted in the $M_{\gamma\gamma}$ spectrum, it won’t be taken into the optimization.

The principle of this optimization is to maximize the significance of the signal over peaking background. The parameters that will be optimized will be the lower mass limit used to compute the significance, the upper mass limit, and the b-tagging of the jets all-together.

To do so, the significance is plotted as a function of the lower integration limit and the upper integration limit. This plot is done for 6 combinations of B-tagging selection (using the csv discriminator) as shown in figure 7. The csv cuts are referred as Loose working point for csv>0.244, Medium working point for csv>0.679 and Tight working point for csv>0.898. The 6 combinations studied here are Loose-Loose, Medium-Loose, Medium-Medium, Tight-Loose, Tight-Medium and Tight-Tight.
Figure 7: 2D histogram showing the significance as a function of the lower integration limit (x axis) and of the upper integration limit (y axis), for the B-tagging combination T-M

After optimization, the value of the best significance is 0.075 and it is got with the Tight-Medium combination and for the mass window from 108 to 138 GeV.

Results

This mass window is then used to cut on the dijet mass, the following spectrum for m\gamma\gamma is obtained:

![Diphoton mass with CiC≥4,Bcuts : Medium-Tight, 108<di-jet mass<138, at 8TeV and 19.2 fb⁻¹](image)

Here, the peak contains 37% of signal, 41% of ttH events and 22% of ZH events. Nevertheless, to become relevant, this analysis has to be extrapolated at higher luminosity. Thus, a collision energy of 14TeV and an integrated luminosity of 3000fb⁻¹ is assumed. After multiplication by the proper ratio of cross section times luminosity, the yields are the following:

<table>
<thead>
<tr>
<th>Processes</th>
<th>Yields at 14 TeV and 3000 fb⁻¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>HH</td>
<td>7</td>
</tr>
<tr>
<td>ttH</td>
<td>9</td>
</tr>
<tr>
<td>ZH</td>
<td>2</td>
</tr>
</tbody>
</table>
Conclusion

In this analysis, samples and selections at 8 TeV and 19.2 fb-1 were used and then extrapolated to 14 TeV and 3000 fb$^{-1}$. The current analysis strategy consists of cutting on $m_{jj}$, fitting on $m_{\gamma\gamma}$ peak and subtracting the peaking backgrounds. It leads to measurable number of events but to go further, suitable Monte Carlo samples with proper collision energy and higher statistics are needed. Besides, the sensitivity of the search could be improved by using categories (2 btagged jets, 1 btagged jet) and the b-jet energy regression could be used to improve PT resolution. Although the statistics dominate, the systematic uncertainties should be studied and other discriminating variables should be investigated.