Cross Section Determination in Higgs Decaying to Two Photons

Zhiyu Liu, *Department of Physics and Astronomy, University of Manchester*

This lab report aims to obtain the cross section of the Higgs boson decaying to two photons process by analyzing the invariant mass distribution of the diphoton system. The data used in the analysis were obtained from the ATLAS Open Data 13 TeV. The selection cuts used to eliminate the background were determined by optimising the signal-to-background ratios. Final signal entries were determined by subtracting the sidebands fitted background from the real data in the signal region. Statistical uncertainties were obtained from evaluating the Poisson errors. Systematic uncertainties were determined by varying different background fitting functions. The final result of the cross section is 46.19 ± 27.25 (stat.) $\pm$ 3.34 (syst.) $\pm$ 0.00093 (lumi.) pb and is consistent with standard model prediction 55.6 ± 2.8 pb .

1 Introduction

The Higgs mechanism, first proposed in 1962, explains how bosons acquire mass. The Higgs particle was discovered in 2012 at the LHC and has mass 125.35 ± 0.15 GeV [1]. The Higgs can be produced via pp collisions. The most dominant production modes for the Higgs particle include gluon fusion (ggF) and vector boson fusion (VBF). The Higgs particle can undergo several decay modes, including quark-antiquark pairs and two photons. Examples of these processes are depicted in the following diagrams.

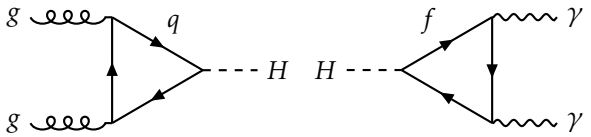

Figure 1: Leading order diagrams for ggH production (*left*) and Higgs decaying to two photons (*right*).

The number of events of a certain process in a collider experiment is $N = \varepsilon\sigma L$, where ε is the detector's efficiency, L is the luminosity representing the number of interactions in the collision, and the cross section σ is a measure of the probability of the process. This experiment's measurements were obtained using the ATLAS detector at the LHC, which consists of an inner detector for tracking charged particles and a calorimeter for measuring the energies. When two photons with momenta greater than 35 and 25 GeV are identified, an event is triggered for reconstruction. However, there are various physical processes that can produce two photons in the final state, including the Z boson decay, and additional particles, such as jets, may be misidentified as photons in the calorimeter, which consist of the background in the diphoton decay channel. The physical quantity used to identify the entries was the invariant mass of the diphoton system, which is a conserved quantity in all physical processes, given by $m_{\gamma\gamma} = \sqrt{E_{\gamma\gamma}^2 - \mathbf{p}_{\gamma\gamma}^2}$. The expected result of the diphoton invariant mass distribution is a continuum of background with a bump in the region [120, 130] GeV, representing the resonance decay peak of the Higgs.

Due to the extremely low production rate of the Higgs boson, we must increase the signal-to-background ratio for improved discrimination between the signal and the background, as well as reducing statistical fluctuations. This was accomplished by applying selection cuts to various physical observables, such as the transverse momenta of the two photons. The expected signal events passing the cuts were determined by Monte Carlo simulation, while the number of background events was obtained by fitting the data outside the signal region [120, 130] GeV.

Finally, a combination of the optimized selection cuts was used to determine the number of signal events from the data, which were then used to calculate the cross section for the Higgs boson decaying into two photons.

2 Signal and background modelling

The Monte Carlo simulated signal can be modelled by several functions, among which the double sided crystal ball function [2], defined in Eq. 1, has the best performance, as shown in Fig. 2. The modelling focuses in the range [105, 140] GeV. The background is modelled by the sidebands data, defined as $100 < m_{\gamma\gamma} < 120$ GeV and $130 < m_{\gamma\gamma} < 160$ GeV, using a quartic function. The integral of these modelling functions in the signal region [120, 130] GeV will be used to determine the signal-to-background ratios.

$$\mathcal{S}^{\mathrm{DSCB}}\left[\mu, \sigma, s = \frac{m_{\gamma\gamma}-\mu}{\sigma}; \alpha_{\mathrm{low}}, n_{\mathrm{low}}, \alpha_{\mathrm{high}}, n_{\mathrm{high}}; m_{\gamma\gamma}\right]$$
$$= N \begin{cases} e^{-s^2/2} & -\alpha_{\mathrm{low}} \le s \le \alpha_{\mathrm{high}} \\ \left(\frac{n_{\mathrm{low}}}{|\alpha_{\mathrm{low}}|}\right)^{n_{\mathrm{low}}} e^{-\alpha_{\mathrm{low}}^2/2} \left(\frac{n_{\mathrm{low}}}{|\alpha_{\mathrm{low}}|} - |\alpha_{\mathrm{low}}| - s\right)^{-n_{\mathrm{low}}} & s < -\alpha_{\mathrm{low}} \\ \left(\frac{n_{\mathrm{high}}}{|\alpha_{\mathrm{high}}|}\right)^{n_{\mathrm{high}}} e^{-\alpha_{\mathrm{high}}^2/2} \left(\frac{n_{\mathrm{high}}}{|\alpha_{\mathrm{high}}|} - |\alpha_{\mathrm{high}}| + s\right)^{-n_{\mathrm{high}}} & s > \alpha_{\mathrm{high}} \end{cases} \quad (1)$$

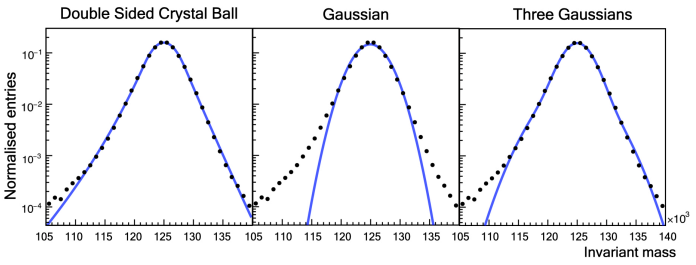

Figure 2: Fitting results of the signal Monte Carlo simulation.

Table 1: Signal-to-background ratio comparisons for various selection cuts. $P_T\text{cone}^{\text{rel}}$ and $E_T\text{cone}^{\text{rel}}$ values were determined beforehand and are optimal when used separately.

Photon number	P_T^{min} [GeV]	$\eta \notin [1.37, 1.52]$	Balance	TightID	$P_T\text{cone}^{\text{rel}}$	$E_T\text{cone}^{\text{rel}}$	$S/\sqrt{B}$	Alias
—	—	—	—	—	—	—	0.8593	
2	—	—	—	—	—	—	0.8595	
—	[40, 30]	—	—	—	—	—	0.9407	
—	—	✓	—	—	—	—	0.8349	
—	—	—	✓	—	—	—	0.9613	
—	—	—	—	✓	—	—	2.0142	
—	—	—	—	—	0.033	—	1.8448	
—	—	—	—	—	—	0.03	2.0484	
2	[40, 30]	✓	✓	✓	0.033	0.03	2.8861	Comb.1
2	[40, 30]	✓	✓	✓	0.055	0.055	3.0454	Comb.2
—	[40, 30]	✓	✓	✓	0.055	0.055	3.0239	Comb.3
2	[35, 25]	✓	✓	✓	0.055	0.055	3.0514	Comb.4
—	[35, 25]	✓	✓	✓	0.055	0.055	3.0529	Comb.5
—	[35, 25]	✓	—	✓	0.055	0.055	2.9443	Comb.6

3 Determination of selection cuts

3.1 *An overview of candidate selection cuts*

To ensure that the final state contains precisely two photons, a cut can be applied based on the number of photons detected. Other sources, such as neutral mesons, can produce photons with lower momenta. Therefore, a lower limit on the transverse momenta can be chosen as an additional cut. A pseudorapidity cut is employed to exclude the region $1.37 < |\eta| < 1.52$. This cut is due to the detector being less controllable in the transition region between the barrel and the endcap of the calorimeter. If the photons are produced from the decay of a Higgs, they tend to be isolated. Two isolation variables, $P_T\text{cone}$ and $E_T\text{cone}$, are defined as the sum of momenta/energies within a cone of half-width $\Delta R = 0.3/0.2$ around the identified photon, where $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$. As photons with higher energies tend to be surrounded by particles of higher energies, the isolation variables are divided by the transverse momentum of their corresponding photons, giving the relative isolation variables $P_T\text{cone}^{\text{rel}}$ and $E_T\text{cone}^{\text{rel}}$. When particles interact with the calorimeter, they create a shower of particles. The photons have a characteristic shower shape, which is defined by the variable TightID. To distinguish between γj and jj production, the cut $E_T/m_{\gamma\gamma} > 0.35$ and 0.25 for leading and subleading photons is required based on statistics and is referred to as the balance condition.

3.2 *Signal-to-background ratio comparison*

The signal-to-background ratios are compared for different candidate selection cuts, either used alone or in combination, as shown in Table 2. It is found that the values $P_T\text{cone}^{\text{rel}} = 0.033$ and $E_T\text{cone}^{\text{rel}} = 0.030$ have the highest $S/\sqrt{B}$ when used alone, whereas the values 0.055 have the highest $S/\sqrt{B}$ when the isolation cuts are used together with other essential cuts. The different optimal values

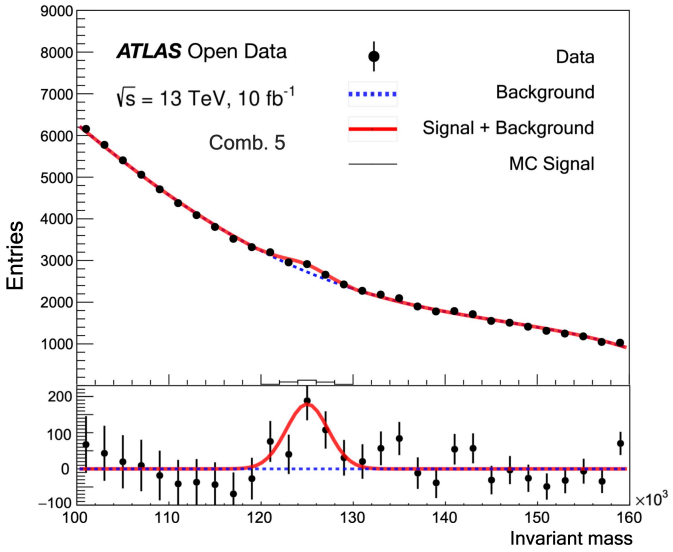

Figure 3: The data passing selection cut combination 5 with a clear peak in the signal region, indicates the production of the Higgs boson. The dotted blue line represents the quartic sidebands background fitting, and the red line superimposed on the background represents the fitted signal. The data subtracted by the background fitting are shown in the lower part of the plot. The errorbars are the Poisson errors from the original data.

for the isolation variables are a result of correlations between selection cuts. The determination of these values are omitted due to limited pages. In total, six combinations of cuts are shown in Table 1. As their signal-to-background ratios are comparable, but the sets of events they contain can be quite different, the cross sections of three combinations are calculated in the following section for comparison. The final result passing the combination 5 is shown in Figure Fig. 3, where a clear bump in the signal region is visible.

Table 2: Cross section determination and statistical errors for data passing chosen cut combinations. Fitted results with χ^2/ndf $\in [0.8, 1.2]$ are preserved and results with χ^2/ndf $\in [0.9, 1.1]$ are colored in red except the outlier.

Cut comb.	Function Alias	χ^2/ndf	N_{selected}	N_{bkg}	N_{signal}	$N^{\text{err}}_{\text{selected}}$	$N^{\text{err}}_{\text{bkg}}$	σ	Stats. Uncer.
Comb.1	`poly4`	0.855	9987	9814.24	172.76	99.93	54.17	46.98	27.251
Comb.1	`poly5`	0.994	9987	9833.41	153.59	99.93	55.33	41.77	27.252
Comb.1	`poly6`	0.938	9987	9803.78	183.22	99.93	56.01	49.83	27.253
Comb.2	`poly5`	0.854	14155	13900.70	254.30	118.97	65.79	69.16	32.430
Comb.2	`expPoly3`	1.122	14155	13921.42	233.58	118.97	117.99	63.53	32.490
Comb.2	`expPoly4`	1.039	14155	13980.74	174.26	118.97	118.24	47.39	32.490
Comb.5	`poly4`	1.097	14208	13776.59	431.41	119.20	66.50	117.33	32.494
Comb.5	`expPoly2`	0.954	14208	14044.59	163.41	119.20	88.69	44.44	32.519
Comb.5	`expPoly3`	1.017	14208	14059.91	148.09	119.20	118.57	40.27	32.553

Table 3: Results of cross sections with uncertainties of each combination of cuts.

Cut	σ	Uncertainties		
		Stat.	Syst.	Lumi.
Comb.1	46.19	27.25	3.34	0.00093
Comb.2	60.03	32.47	9.23	0.00093
Comb.5	42.36	32.54	2.08	0.00093

4 Cross section and uncertainties

The cross section for this process is calculated using the formula shown in Eq. 2.

$$\sigma = \frac{N_{\text{selected}} - N_{\text{bkg}}}{\varepsilon \int L\mathrm{d}t} = \frac{N_{\text{signal}}}{\varepsilon \int L\mathrm{d}t}. \tag{2}$$

N_{selected} is the total number of data entries that pass the selection cuts in the signal region. N_{bkg} is determined by the integral of the background fitting function in the signal region. $\int L\mathrm{d}t = 10.064$ pb^{-1} and $\varepsilon = 0.36536$ are given in the script.

The final results are shown in Table 2. The functions used to fit the backgrounds are shown in the second column of the table. The function is either a polynomial or an exponential polynomial. For example, the function `expPoly2` is defined as $\exp\left(m^2_{\gamma\gamma} + m_{\gamma\gamma} + C\right)$. The results within a desired reduced chi-squared range of $[0.8, 1.2]$ are preserved. Fits with reduced chi squared in the range $[0.9, 1.1]$ are colored in red. An result with a cross section of 117.33 is considered to be an outlier and will be excluded in the following analysis. The imperfection of the ROOT algorithm is believed to be the cause of this outlier. The mean value of the cross sections is determined as the final result in each combination.

The error propagation formula for Eq. 2 is shown in Eq. 3. The uncertainties in the number of selected events arise from the Poisson counting error in the discrete bins. The combined statistical errors in the signal region can be obtained using the ROOT function `TH1::GetIntegralAndError()`. The uncertainties in the number of background events come from the fitting errors and can be obtained using the ROOT function `TF1::IntegralError()`. These two sources of uncertainties are the statistical errors and the values are shown in the last column of the table. The mean values are used as the statistical uncertainties of the final results in Table 3.

$$[\text{Err.}(\sigma)]^2 = \left(\frac{1}{\varepsilon \int L\mathrm{d}t}\right)^2 [\text{Err.}(N_{\text{selected}})]^2 + \left(\frac{1}{\varepsilon \int L\mathrm{d}t}\right)^2 \left[\text{Err.}(N_{\text{bkg}})\right]^2 + \left(\frac{1}{\varepsilon \int L\mathrm{d}t}\right)^4 \left[\text{Err.}\left(\int L\mathrm{d}t\right)\right]^2. \tag{3}$$

The uncertainty in the luminosity is often computed separately, which is due to the limitations of the measurement apparatus and is given as $10.064 \times 1.7\%$. As different fitting functions yield different signal entries, various fitting functions are tested. The largest difference between each cross section and the mean value is determined as the systematic uncertainty. The combined results are shown in Table 3, which are in consistent with the standard model prediction 55.6 ± 2.8 pb [3].

5 Conclusion and outlook

This experiment consists of three parts: signal and background modelling, determining selection cuts and calculating the cross section. The validation for selection cuts require huge amount of time and can be performed better on the lower limits of momenta and the isolation variables. The systematic uncertainties can be further optimised by using different ranges of the sidebands. Only a small amount of ATLAS data was used in this experiment, so the final results are not accurate and have large uncertainties.

6 Refernces

[1] The CMS collaboration. A measurement of the Higgs boson mass in the diphoton decay channel. 2019.

[2] S. Menary. Higgs cross section measurements at $\sqrt{s} = 13$ tev using the atlas detector. pages 134,135, 2019.

[3] The ATLAS collaboration. Measurement of the total and differential Higgs boson production cross-sections at $\sqrt{s} = 13$ TeV with the ATLAS detector by combining the $H \to ZZ^* \to 4l$ and $H \to \gamma\gamma$ decay channels . 2022.