Higgs Boson Pair Production Combination Analysis

Combination of $b\bar{b}\tau^+\tau^-$, $b\bar{b}b\bar{b}$, $b\bar{b}\gamma\gamma$, and multilepton channels

Jesse D. Viola

A honors thesis presented for the degree of
Bachelor of Science
Advised by Professor Thomas Schwarz

Department of Physics
University of Michigan
Ann Arbor, MI
April 2023

ABSTRACT

This thesis discusses Higgs boson pair production (HH) from gluon-gluon fusion (ggF) and the combination analysis used to compute cross-section limits. Searches for pairs of Higgs bosons, in the $b\bar{b}\tau^+\tau^-$, $b\bar{b}b\bar{b}$, $b\bar{b}\gamma\gamma$, and multilepton final states, are included in this combination. The combination results prove to show a higher sensitivity compared with individual channels, as predicted by the Standard Model (SM). A beyond the Standard Model (BSM) interpretation, called the hMSSM, is also explored with no conclusive claims due to minimal data being available at the time of this analysis.

Contents

I. INTRODUCTION

The Standard Model (SM) predicts a rare occurrence of Higgs boson pair production, which can be used as a direct probe to the Higgs boson self-coupling (κ_λ). To study the Higgs mechanism, the Higgs potential is minimized to find the ground state expectation value. The Higgs potential with small fluctuations about this extremum can be defined by the following:

$$V(H) = V_0 + \frac{m_H^2}{2}H^2 + \frac{m_H^2}{2v^2}vH^3 + \frac{m_H^2}{8v^2}H^4 \quad (1)$$

In this expansion, the third term depicts the Higgs boson self-interaction, which dominates when $m_H \gg v$. The leading production of di-Higgs (HH) is from the gluon-gluon fusion $pp \to HH$ process (ggF) **[1]**. This process is exceedingly rare, but some physicists working on models beyond the SM (BSM) predict that these events could happen at a much higher rate than predicted by the SM. Experimentally searching for these di-Higgs events has the potential to provide evidence for BSM theories, thus posing an important field of study. Due to the elusive nature of the Higgs boson, the search for a resonant or non-resonant Higgs boson pair consists of searching for the decay products of the pair. These detectable possible products consist of four bottom quarks ($b\bar{b}b\bar{b}$), two bottom quarks and two taus ($b\bar{b}\tau^+\tau^-$), two bottom quarks and two photons ($b\bar{b}\gamma\gamma$), among other less experimentally relevant decay final states for a Higgs boson pair **[2]**. A Feynman diagram for the production of a di-Higgs via ggF and then the decay into $b\bar{b}\tau^+\tau^-$ is illustrated in **Figure 1**. Through the detection of these fermion and boson constituents, the rate of di-Higgs production can be determined using statistical analysis methods. Analyzing a combination of the theorized decay products of di-Higgs can provide stronger statistical evidence of di-Higgs production, which is studied by the HH Combination effort at ATLAS **[3]**.

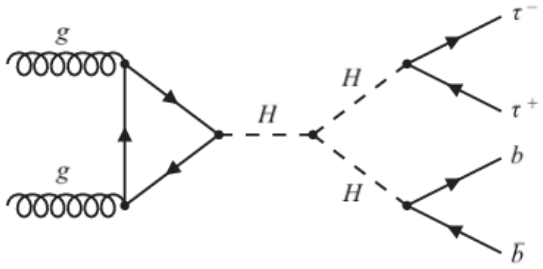

Figure 1. Feynman diagram illustrating Higgs boson production via ggF, and then Higgs self-coupling resulting in a di-Higgs. The di-Higgs then decays into $b\bar{b}\tau^+\tau^-$.

As a member of the University of Michigan HH Analysis group (working for the ATLAS experiment at CERN), my role has been building and debugging the HH Combination framework. This framework consists of the analysis code for the di-Higgs channels ($b\bar{b}\tau^+\tau^-$, $b\bar{b}b\bar{b}$, $b\bar{b}\gamma\gamma$), as well as plotting and other visualization code. The analysis framework combines the channel limits to compute a combined cross-section limit (the measure of the observed probability of Higgs

boson pair production compared to the expected SM probability). A lower limit corresponds to a higher channel sensitivity. The analysis of the channels is mainly written in C++ and Python, using ROOT due to the computational efficiency of this CERN-created language. The input workspaces for each analysis channel are continuously updated when new data is implemented, meaning that the combination analysis must be rerun regularly. The workspace for each channel consists of the latest cross-section limits at each available mass point from Run-2 at CERN, as well as the SM prediction model **[2]**.

II. THE ATLAS DETECTOR

The ATLAS detector is one of the four detection experiments of the Large Hadron Collider (LHC) at CERN, designed to detect elementary particles at incredibly high energy scales. The detector itself is a cylinder with a diameter of 25 meters and a length of 46 meters, weighing 7,000 tonnes and is located in a cavern 100 meters below ground. The ATLAS detector consists of six different detecting subsystems wrapped concentrically in layers around the collision point to record the trajectory, momentum, and energy of particles, allowing them to be individually identified and measured. A schematic of the detector can be seen in **Figure 2**.

The six subsystems of the ATLAS detector are the inner detector, electromagnetic calorimeter, hadronic calorimeter, muon spectrometer, forward calorimeter, and forward muon spectrometer. The inner detector is composed of pixels, micro-strip detectors, and a transition radiation tracker. The electromagnetic calorimeter measures the energy of electrons and photons. The hadronic calorimeter measures the energy of hadrons. The muon spectrometer measures the momentum of muons. The forward calorimeter measures the energy of particles that travel close to the beam pipe. The forward muon spectrometer measures the momentum of muons that travel close to the beam pipe.

The ATLAS detector is one of nine detector experiments along the LHC that uses various data acquisition and analysis techniques to explore particles produced by collisions. The trigger and data acquisition (TDAQ) system is responsible for selecting relevant events from the millions of collisions that occur every second. The ATLAS detector has been extensively tested and debugged, and any relevant results are able to be cross-checked with other detectors at the LHC such as CMS **[3]**.

For the HH Combination effort, the ATLAS detector is used to search for Higgs boson pair production using the various decay channels outlined in the introduction above ($b\bar{b}\tau^+\tau^-$, $b\bar{b}b\bar{b}$, $b\bar{b}\gamma\gamma$). The data collected by the ATLAS detector is absolutely vital for the cross-checking of simulated cross-section results for each decay mode.

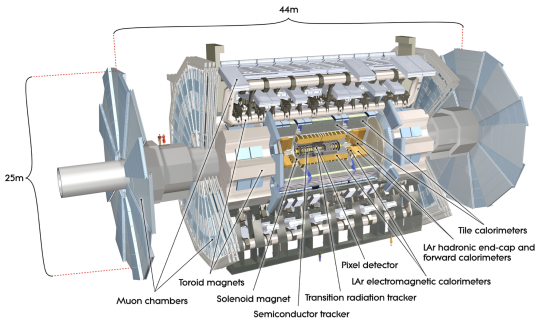

Figure 2. Cut-away view of the ATLAS detector **[3]**.

III. HIGGS BOSON PAIR PRODUCTION THEORY

As briefly discussed in the introduction, the leading production of HH production is from the gluon-gluon fusion (ggF) process. This process accounts for over 90% of non-resonant HH production. There are two leading order Feynman diagrams that represent ggF, which can both be seen in **Figure 3** (triangle diagram) and **Figure 4** (box diagram). Due to the destructive interference of these diagrams, the SM cross-section for HH production is much smaller than the cross-section of single Higgs boson production. The SM cross-section for HH production via ggF is predicted to be $\sigma^{\text{SM}}_{\text{ggF}}(HH) = 31.05^{+6\%}_{-23\%}(\text{scale} + m_{\text{top}}) \pm 3.0\%(\text{PDF} + \alpha_s)$ fb for a Higgs boson mass of $m_H = 125$ GeV and $\sqrt{s} = 13$ TeV. The "scale" uncertainty is due to the finite order of the quantum chromodynamics (QCD) calculations, the "m_{top}" uncertainty is due to the top-quark mass scheme, and the "PDF + α_s" uncertainty is due to the effects of the strong coupling constant and parton distribution functions **[2]**.

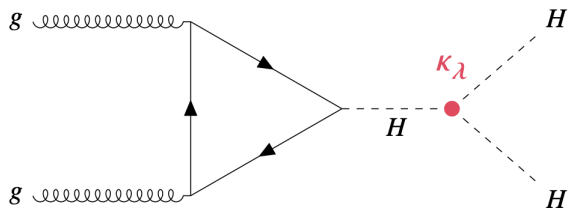

Figure 3. Triangle diagram of HH production via ggF.

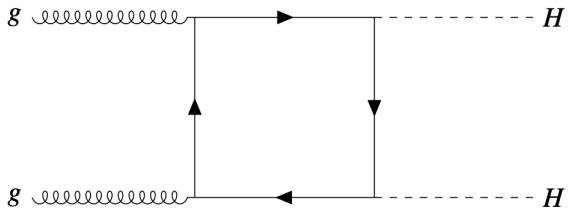

Figure 4. Box diagram of HH production via ggF.

The next leading production of HH production is from Vector Boson Fusion (VBF), which contributes to 5% of SM HH production. VBF production is omitted from all of the combination analysis described in this paper, meaning that only the ggF production mode is considered **[2]**.

Most of my contributions to this analysis resided within SM predictions, except for completing some BSM work on the habemus Minimal Supersymmetric Standard Model (hMSSM). The hMSSM is a parameterization of the Minimal Supersymmetric Standard Model (MSSM). The hMSSM is a special case of the MSSM in which the mass of the lightest Higgs boson is automatically set to the measured value of 125 GeV by adjusting the supersymmetric particle

spectrum to provide the required amount of radiative corrections to the Higgs boson mass. The hMSSM approach simplifies the Higgs sector by reducing the number of inputs to only two, $\tan\beta$ and m_A, and eliminates the need for SUSY parameters **[5]**.

The hMSSM approach is relevant within di-Higgs analysis due to it having been revisited for Higgs self-couplings. The dominant coupling of the hMSSM is used to compare with the full MSSM result to disentangle the deviations due to the hMSSM approximation of the coupling from those originating from momentum-dependent contributions. The hMSSM approach is used as a benchmark scenario for the interpretation of experimental results. The hMSSM approach simplifies the MSSM parameter space, making it easier to investigate **[5]**.

IV. DATA AND SIMULATION

For the non-resonant ggF HH Monte Carlo (MC) samples, the POWHEG BOX v2 generator was used with next-to-leading order (NLO) accuracy in QCD (using the PDF4LHC15 parton distribution function set). The samples were generated with κ_λ values equal to 1 and 10. After the samples were generated, a reweighting method was used to determine the signal yield at a given value of κ_λ **[2]**.

For the spin-0 resonant ggF HH Monte Carlo (MC) samples (which follow the decay of $pp \to X \to HH$), MADGRAPH5_aMC@NLo v2.6.1 was used with leading order (LO) accuracy (using the NNPDF2.31o PDF set). Any interference with non-resoannt HH production was neglected **[2]**.

V. METHODS

The bulk of the di-Higgs combination analysis work revolve around the HH combination framework, which is a framework used for the preprocessing and combination of workspaces, the extraction of limits, exploring BSM interpretations (such as EWK singlet and hMSSM), as well as profile ratio scans. This framework was build to work for Run 2 combination efforts at the time of my contributions on this research. The general workflow of the HH combination framework consists of: regularization, rescaling, and then combination of single channels. The entire framework can be found on a GitLab repository, and is mostly written in Python **[6]**.

After using the HH combination framework to combine single channels, the rest of the analysis work consists of plotting and plot analysis, which I completed using Python.

In both the framework and the plotting code, the computer language ROOT (and its Python constituent, PyROOT) was used throughout due to its efficient data processing. ROOT is a software framework and toolset that provides a suite of C++ libraries for data processing, anal-

ysis, and visualization in high-energy physics developed at CERN. The ROOT framework is designed to handle large datasets efficiently, with support for parallel processing and distributed computing. It includes an object-oriented programming model that enables users to create complex data structures and algorithms in a modular and reusable manner. ROOT also provides a set of tools for data storage and retrieval, with support for multiple file formats and database management systems **[7]**.

VI. MY CONTRIBUTIONS

As I concurrently worked on the last stages of the combination effort, which required the running of the combination framework and plotting, I would employ the most recent workspaces to compute results. This process consisted of augmenting the input ROOT files containing the limit data in order to prevent issues in the combination, as well as updating the framework code itself. Once the combination was complete, I focused my resources to update the plotting code to keep up with changing data input.

The result of plotting the combination analysis was a visual representation of the observed and expected limits for each analysis channel. This was completed for both the non-resonant and resonant cases. The resonant case occurs when a resonance exists, meaning a peak in energy in which the differential cross-section is a maximum. The cross-section limits for non-resonant and resonant decay are compared to SM predictions. While many of the early versions of the workspaces were blinded, meaning that the observed limits were hidden, the later versions included both, thus giving the ability to compare observed and expected results. For the resonant case, the combination analysis was completed for both the resolved and boosted cases, which was dependent on what range of mass points (m_S) were available for each channel. The boosted case includes higher mass points due to the increased energy from a Lorentz boost **[3]**.

The protocol for running the combination framework on each channel consists of first processing each channel and producing processed output files. Processing the ROOT files organizes the data for the combination analysis and plotting code to run properly. Once all of the necessary channels are processed, a combination command is run on the corresponding channels, creating combination files that the plotting code is able to extract from. A combination of any number of the analysis channels can be completed, as specified in the combination command. Finally, plotting can be completed from both the combination files as well as the individual processed channel files.

One BSM interpretation of the resonance limits is the habemus Minimal Supersymmetric Standard Model (hMSSM), which predicts the existence of a heavy CP-even scalar particle

decaying into SM di-Higgs. This BSM model can be illustrated by plotting the relationship between the mass of this heavy particle (m_A), and the ratio of the vacuum expectation values of the two Higgs bosons ($\tan\beta$) **[4]**. The observed m_A is compared to the expected m_A to look for deviation from the SM. I also took part in working on the framework of the hMSSM model, by completing data interpolation. For this model I focused solely on the $b\bar{b}\tau^+\tau^-$ channel, where I analyzed both observed and expected limits using various data interpolation procedures in Python.

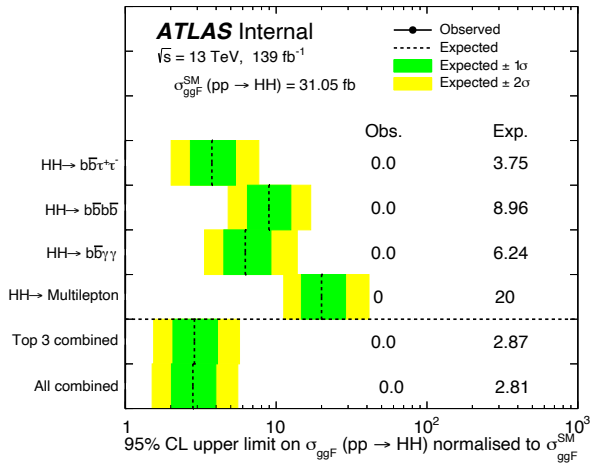

Figure 5. Blinded observed and expected 95% CL upper limit on the signal strength for SM HH production in the non-resonant $b\bar{b}\tau^+\tau^-$, $b\bar{b}b\bar{b}$, $b\bar{b}\gamma\gamma$, multilepton channels, and their statistical combination.

VII. RESULTS

For the non-resonant case, **Figure 5** illustrates the 95% confidence level (CL) upper limits on the cross-section for each of the channels as well as the combined case.

Since a lower cross-section limit corresponds to a higher rate of production (the limit is normalized to the SM cross-section), the $b\bar{b}\tau^+\tau^-$ channel predictably shows a higher sensitivity of di-Higgs production with an expected limit of 3.54. Most importantly, the combination of the top 3 channels ($b\bar{b}\tau^+\tau^-$, $b\bar{b}b\bar{b}$, and $b\bar{b}\gamma\gamma$) and the combination of all of the channels give the expected limit of 2.87 and 2.81 respectively, thus proving the combination effort to be worthwhile. In the resonant case, **Figure 6** shows the upper limit on the cross-section for various channels at different mass points in a range from 251-2000 GeV. The mass points from 251-1000 GeV are considered the resolved case, while the mass points above 1000 GeV are a part of the boosted case.

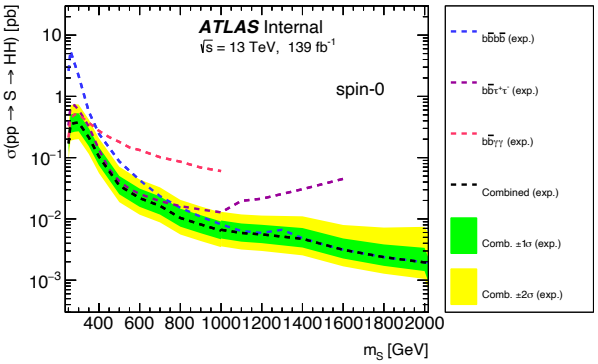

Figure 6. Expected 95% CL upper limit on the signal strength for SM HH production in the resonant $b\bar{b}\tau^+\tau^-$, $b\bar{b}b\bar{b}$, $b\bar{b}\gamma\gamma$, multilepton channels, and their statistical combination.

Similar to the non-resonant case, the combined channels have a limit consistently lower

than the individual channels.

For the hMSSM interpretation, the results consisted of a plot relating $\tan\beta$ at different mass point values (m_A) above 180 GeV. Using the input data, multiple interpolation methods were compared and a cubic interpolation using the Scipy function Griddata proved to be the most effective. Once this was determined, a contour plot could be created to illustrate $\tan\beta$ at a range of mass points, as seen in **Figure 7** (the contour lines for m_H can be seen as the blue dashed lines).

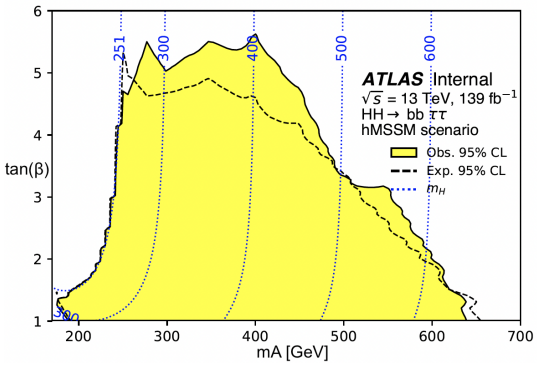

Figure 7. Observed and expected 95% CL upper limits on the signal strength for hMSSM HH production in the $b\bar{b}\tau^+\tau^-$ channel.

This plot includes both the expected and observed results in the form of contour lines, which share a similar shape.

VIII. CONCLUSIONS

In the non-resonant and resonant cases, the combination results are statistically a more effective analysis for the di-Higgs production, as they show a higher sensitivity compared with individual channels, as predicted by the SM. From the unblinded observed results analyzed, there was no evidence of di-Higgs production beyond the SM, as all differences between the expected and observed results are within the range of statistical fluctuations.

With the hMSSM interpretation, the next steps will be to overlay the results from each of the three channels in order to reduce the areas of low sensitivity. For example, $b\bar{b}b\bar{b}$ may have a higher sensitivity at a low $\tan\beta$ and a high m_A. With more data incoming from the ATLAS experiment at CERN, the results will be continually updated and the analysis techniques will be able to be refined. This in turn will produce results with higher confidence intervals, and therefore more conclusive claims.

IX. ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my research mentor and thesis advisor, Professor Thomas Schwarz, for his unwavering support and guidance throughout this project and my undergraduate physics career in general.

I would also like to acknowledge the post-doctoral student and now professor, Professor Yanlin Liu, for his incredible assistance and instruction throughout this research.

X. REFERENCES

[1] S. Playfer. Lecture 17 – The Higgs Boson. The University of Edinburgh.

[2] ATLAS Collaboration. (2021). Combination of searches for non-resonant and resonant Higgs boson pair production in the $b\bar{b}\gamma\gamma$, $b\bar{b}\tau^+\tau^-$ and $b\bar{b}b\bar{b}$ decay channels using pp at $\sqrt{s} = 13$ TeV with the ATLAS detector. ATLAS-CONF-2021-052.

[3] ATLAS Collaboration. (2010). Studies of the performance of the ATLAS detector using cosmic-ray muons. European Physical Journal C. 71. 1-36. 10.1140/epjc/s10052-011-1593-6.

[4] ATLAS Collaboration. (2020). Combination of searches for Higgs boson pairs in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. ArXiv:1906.02025v2 [hep-ex].

[5] S. Liebler el al. (2019). The hMSSM approach for Higgs self-couplings revisit. ArXiv:1810.10979v2 [hep-ph].

[6] https://gitlab.cern.ch/atlasHBSM/atlas-phys-higgs-dihiggs-combteam/hh_combination_fw

[7] https://root.cern