Alisa Nozdrina, Zach Warner

Double Higgs production

Abstract

In this paper we discuss double Higgs production at the LHC. Specifically we study the process of double Higgs production $pp \to hh$ with $bbll\ \not{E}_T$ final state through the use of a MC particle collision event generator in an attempt to search for background suppression strategies. Here all of our background was considered to be generated from $pp \to t\bar{t}$ production. The best signal significance achieved was $\sigma_{discovery} = 0.49$ and was reached by using a Boosted Decision Tree algorithm. We found that in order to reach the discovery significance obtained with our strategy an integrated luminosity of $\sqrt{L} \approx 10^{21} b^{-1}$would be required at the LHC.

1 Introduction

The Higgs field in the Standard Model (SM) of particle physics is responsible for electroweak symmetry breaking via the Higgs mechanism resulting in the mass generation of $\mathrm{SU}(2)_W \times \mathrm{U}(1)_Y$ gauge bosons. While the discovery of the Higgs field in 2012 by the ATLAS and CMS experiments at the LHC was an important validation of the SM, it also opened the door to test the consistency of the properties of the Higgs with prediction made by the SM. Currently these measurements show no significant deviation from predictions made by the SM [3]. Most of our free paramaters, and therefore ignorance, in the Standard Model are related to the Higgs resulting in the undetaking of new efforts to accurately define the properties of the Higgs; mainly its mass, spin, parity and coupling to itself and the other particles of the Standard Model.

$$\mathcal{L} \supset m_h^2 h^2 + k_3 \lambda_3^{SM} h^3 + \frac{1}{4} k_4 \lambda_4 h^4 \tag{1}$$

One of the properties of the Higgs mechanism is Higgs self-coupling eqn.(1), with knowledge of the self-coupling being vital to understanding electroweak symmetry breaking and the origin of mass in our universe. In standard model scenarios $\lambda_3 = \lambda_4 = \frac{m_h}{2\vartheta^2}$ represents the SM values of Higgs self-coupling where m_h is the SM Higgs mass and ϑ^2 is the Higgs vacuum expectation value. Observations of Higgs self-coupling would update the Higgs mechanism of mass generation, probe the structure of the Higgs potential, and possibly be sensitive to new physics. While measurements of Higgs boson coupling with both fermions and gauge bosons are in good agreement with current SM predictions, Higgs self coupling has been proven difficult to measure experimentally due to its small expected cross section [1]. In order to study Higgs self-coupling in the SM as well as the deviation from the SM values caused by k_3 and k_4 in eqn. (1) it is useful to look at double Higgs production at the LHC, specifically under the High Luminosity (HL) upgrade.

In our paper we look a double Higgs (hh) production at the LHC. Specifically, we study $pp \rightarrow hh$ production at the HL-LHC consisting of a a final state with two b-tagged jets, two leptons, and missing transverse momentum of the lepton neutrinos ($bbll$) by examining distributions of useful kinematics for our signal (hh) and our background events ($t\bar{t}$) . We motivate this analysis by first discussing double Higgs production and the current searches for its signature. We then discuss the analysis of our Monte Carlo (MC) generated particle collision events through our chosen kinematic variables, as well as the application of a Boost Decision Tree (BDT) Machine Learning technique to get the most effective selection criteria of our chosen kinematic variables. Finally, we end with a discussion of the results of our analysis and what it could mean for the required integrated luminosity $\sqrt{L}$ of the upgraded LHC.

1.1 Motivation and Current Experimental Searches

Double Higgs production can be studied through many decay channels like $b\bar{b}bb$, $b\bar{b}\gamma\gamma$, $b\bar{b}\tau\tau$, $b\bar{b}W^+W^-/ZZ$, etc, but for our analysis we will focusing on on $bbll$ final state where we have two b-tagged jets, two leptons, and missing transverse momentum from the lepton neutrinos. The specific signal processs is $(h \to b\bar{b})$ $(h \to W^+W^- \to l\nu_l l'\nu_{l'})$. This decay channels has a large SM background mainly due to top quark pair production $pp \to t\bar{t}$ making it an interesting and difficult case to study. Due to the large difference in production rate the search for signal events require strong background discrimination. The analysis of kinematic distributions for system of two leptons and two b-jets seems to be promising tool [1]. This is based on the event topology where two leptons produced by single higgs decay $h \to$WW$\to l\nu l\nu$ have the same mother particle, while in the $t\bar{t} \to b$wbw$\to bl\nu bl\nu$ they were produced in two different decays.

As previously mentioned experimental search for double Higgs production is complicated due to the very small signal cross-section $\sigma_{hhh} = 40{,}7$ fb at $\sqrt{s} = 13$ TeV in comparison with the main background process from two top quarks $\sigma_{t\bar{t}} =$ 953.6 pb. It should be noted that not all events with a double Higgs boson as a final state were produced via Higgs self-coupling, but via heavy-quark Yukawa couplings. Illustration of double Higgs production from a two gluon initial state are provided on Fig. 1.

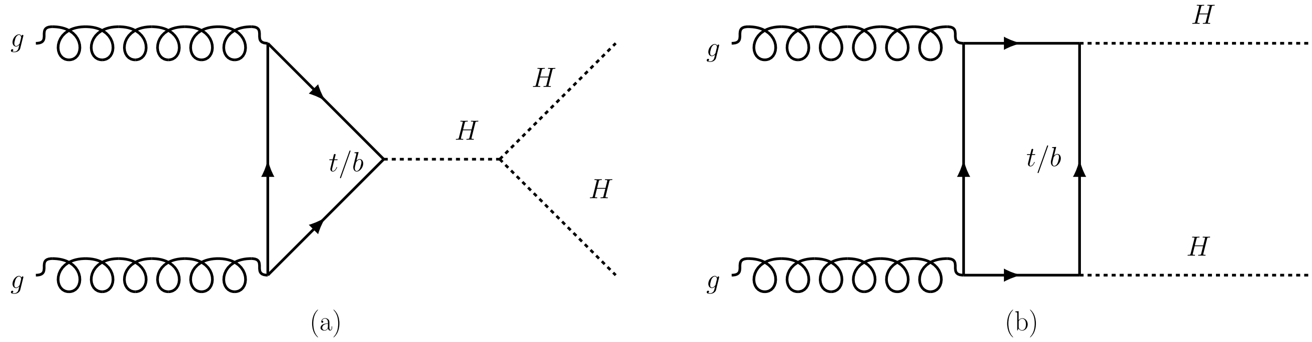

Figure 1: An example of leading-order Feynman diagrams for Higgs boson pair production; (a) via Higgs self-coupling, (b) via heavy-quark Yukawa couplings.

A series of experimental searches for Higgs boson pair production have been conducted with ATLAS data for $\sqrt{s} = 13$ TeV and an integrated luminosity $= 36.1$ fb. The result contains an analysis of 6 different decay channels: $b\bar{b}b\bar{b}$, $b\bar{b}W+W-$, $b\bar{b}\tau\bar{\tau}$, $W+W-W+W-$, $b\bar{b}\gamma\gamma$ and $W+W-\gamma\gamma$ final states [3]. In these searches, no statistically significant excess of events above the Standard Model predictions is found. It should be noted that a similar study was contducted by the CMS experiment with the same data, and they also obtained results within the predictions made by the SM [4]. Illustration of observed number of events in described decay channels are provided on Fig. 1, and the experimental limitation on non-resonant Higgs pairs production cross-section as a function of $k_\lambda = \lambda_{observed}/\lambda_{SM}$ on Fig. 2.

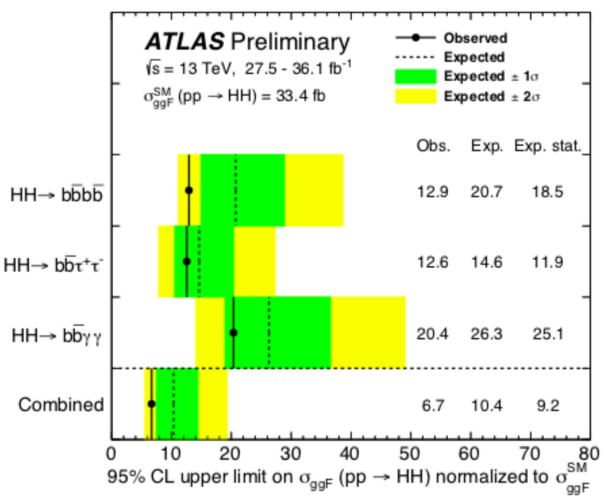

Figure 2: Combination of searches for Higgs boson pairs at ATLAS. Picture taken from [3].

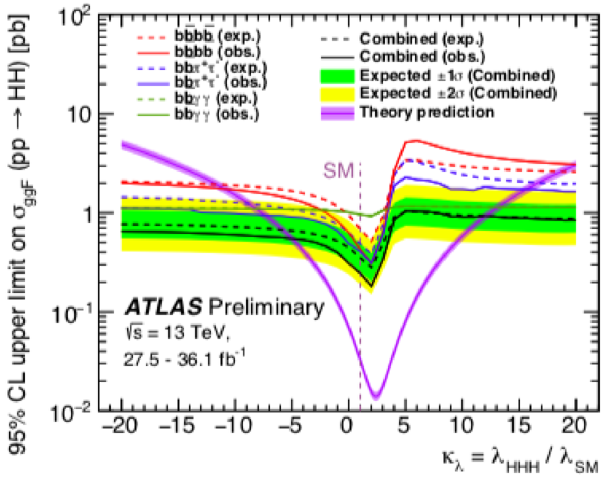

Figure 3: Cross-section limit of the Higgs boson pair production as a function of k_λ. Source [3].

2 Analysis and Kinematic variables

We were using MadGraph5_aMC event generator, Pythia8 for hadronization process and Delphes for detector response simulations. For our analysis we were looking for events contain two b-tagged jets and two leptons as a final state particles. From 100K generated events for signal and background processes only $\approx 10\%$ of the data passed our criteria. For those events we made distributions Fig. 4 for following kinematic variables:

1. M_{ll} which is the invariant mass of two leptons. $M_{ll} = \sqrt{(p^2_{l[1]} + p^2_{l[2]})}$. This is a useful variable based on the event topology since two leptons in signal process were created from a single h decay, while for the background it was produced via two independent W decays.

2. ΔR_{ll} which is angular separation between two leptons. $\Delta R_{ll} = \sqrt{\delta\phi^2 + \delta\eta^2}$

3. Missing Pt - missing transverse momentum for all the particles in the event Pt $=\sum_{i=1}^{n}$ p$_{ti}$

4. M_{bb} which is invariant mass of two b-jets. $M_{bb} = \sqrt{(p^2_{b[1]} + p^2_{b[2]})}$

5. ΔR_{bb} which is angular separation between two b-jets. $\Delta R_{bb} = \sqrt{\delta\phi^2 + \delta\eta^2}$

6. Pt_{ll} the transverse momentum of two leptons $Pt_{ll} = Pt_{l1} + Pt_{l2}$

The number of events expected on LHC with target luminosity L = 3 ab^{-1} could be found by following formula

$$N = L\sigma_{process} \tag{2}$$

where $\sigma_{signal} = 0.245 * 10^{-3}$ pb and $\sigma_{background} = 22.49$ pb were be found by

$$\sigma_{signal} = \sigma_{hh} Br(h \to bb) Br(h \to WW) Br(W \to l\nu) Br(W \to l\nu) \tag{3}$$

$$\sigma_{background} = \sigma_{tt} Br(t \to bW) Br(t \to bW) Br(W \to l\nu) Br(w \to l\nu) \tag{4}$$

where Br$(h \to bb) = 5.809 * 10^{-1}$, Br$(h \to WW) = 2.152 * 10^{-1}$, Br$(t \to bW) \approx 1$, and Br$(W \to l\nu) = 0.1086$ which was multiplied by 2 to take into account e and μ final states. Numbers taken from [5]

For chosen target luminosity number of signal events $N_s = 735$ and number of background events $N_b \approx 6.16$M we found the following efficiency and significance in Table 1.

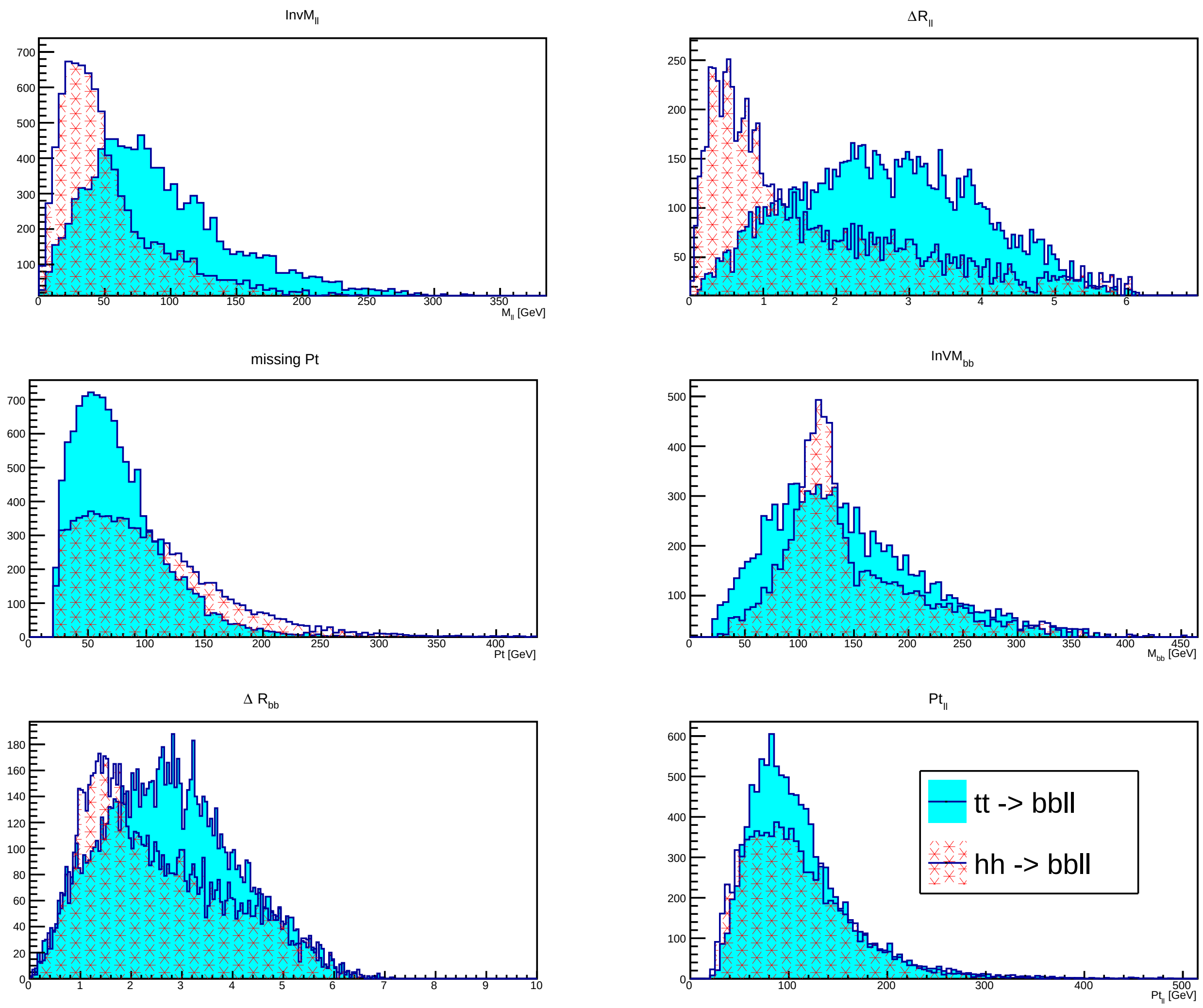

Figure 4: Kinematic variable distributions for signal and background processes.

It was determined that additional conditions on any other kinematic variables from the list above would not have given any significant improvement on signal/background ratio, even making it worse in some cases.

2.1 Machine learning application

In order to get the most effective selection criteria based on chosen kinematic variables we used a BDT (Boosted Decision Tree) algorithm. BDT is a machine learning technique which requires a tagged training data set with which it builds an optimal sequence of conditional statements (if – else statements). After the training the constructed model could be exported and used on untagged data.

Cuts	Signal [pb]	ϵ_{hh}	Background [pb]	ϵ_{tt}	$\sigma_{discovery}$	S/B
before cuts	$0.245 * 10^{-3}$	–	22.49	–	0.28	$1.09 * 10^{-4}$
$M_{ll} < 65$ GeV	$0.18 * 10^{-3}$	74 %	7.65	34 %	0.36	$0.25 * 10^{-3}$
$\Delta R_{ll} < 1.5$	$0.13 * 10^{-3}$	74 %	3.8	55 %	0.38	$0.35 * 10^{-4}$
MLP	$0.098 * 10^{-3}$	75 %	0.95	25 %	0.44	$0.35 * 10^{-3}$
BDT	$0.1 * 10^{-3}$	78 %	0.84	22 %	0.49	$0.4 * 10^{-3}$

Table 1: Efficiency of cuts and statistical significance

Specifically we were using the BDT classifier from python sklearn library.

As an input we were using the set of 6 kinematic variables presented in the previous section. For each generated event which passed the basic reconstruction we made a vector of correspondent input features. All of the signal events were tagged as 1 and all of the background events were tagged as 0. Thus we had $\approx$ 20K tagged events, 80% of which we were using for training and the remaining 20% we were using to test the accuracy of our classifier. Before beginning the training, the data was randomly mixed to avoid possible implicit correlations.

After the training we ran BDT on the testing data set. The returned mean accuracy based on given labels was equal to 0.78. This mean that 78% of signal events would be classified correctly. Thus, the efficiency of BDT discrimination for signal events is equal to 78% and for backgrounds it is 22%. This would give us a significance $\sigma_{discovery} = 735 * .78/\sqrt{(6.16 * 10^6 * .22)} = 0.49$.

Fig. 5 presents the importance of the input features which correlates to the predictive power of the input kinematic variables. It can be seen that the main impact is the M_{ll} variable. This is intuitive since M_{ll} shows the biggest separation for signal and background distribution. Here the test data set contains $\approx$ 2K events. The BDT algorithm associated each event with a number from 0 to 1 in accordance with their input features, which we would call the returned probability. Events with returned probability $>$ than 0.5 have been classified as background and the ones < 0.5 as signal. Distribution of all of the returned probability for test data set are presented on the Fig. 6. In our case the distribution is almost discrete due to the small number of input parameters. Therefore, the BDT algorithm could be interpreted as a function with a finite range of values. As the amount of input features increases the distribution will become more and more continuous, eventually taking a shape that is more common in machine learning reviews and articles.

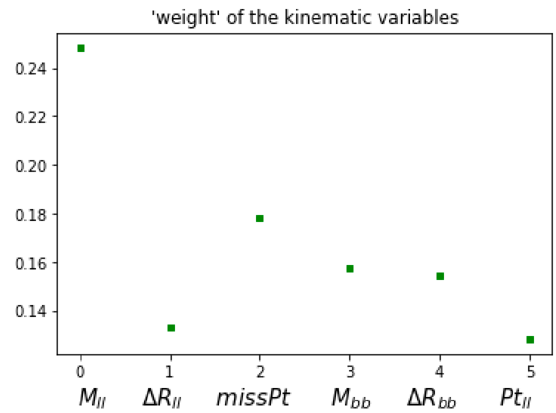

Figure 5: Importance of input kinematic variables

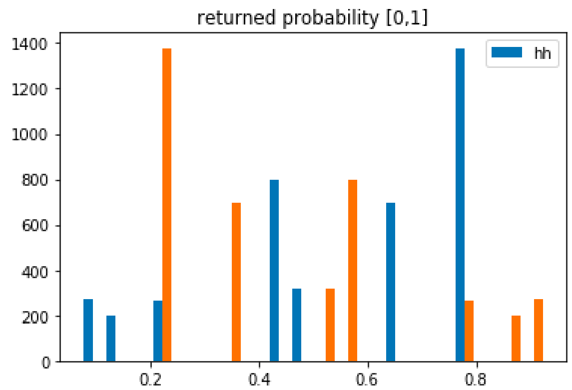

Figure 6: Distribution of returned probabilities

We also tried to apply artificial neural net algorithm for a background discrimination using exactly the same input features and training and test data sets. We were using MLP (multi-layer perceptron) classifier from sklearn python library. This algorithm required significantly longer time for training and returned accuracy was qual 0.75. This is slightly worse than BDT accuracy resulting in us using BDT for our final results.

3 Summary and conclusion

The goal of our work was to study the process of double Higgs production $pp \to hh$ at the LHC with $bbll \not{E}_T$ final state and to use this study to search for background suppression strategies.

1. 100K events $pp \to hh \to bbll\nu\nu$ and 100K events $pp \to tt \to bbll\nu\nu$ were generated by using a MC event generator.

2. Preliminary analysis was made by picking out events with the correct signature – two b-jets and two leptons in the final state

3. For events which passed primary reconstruction we made the set of distributions of selected kinematic variables, the results of which are presented on Fig 4.

4. Based on the obtained distributions, selection criteria were used that maximize signal to background ratio. For the chosen criteria, the cut efficiencies were then calculated.

5. A theoretical estimation of the number of expected events was made, and by using found cut efficiencies the expected statistical significance of the signal was obtained. The main results of our work is presented in Table 1. The best signal significance achieved = 0.49 and was reached by using Boosted Decision Tree algorithm.

During the analysis, it turned out that the returned accuracy of the algorithm increases with the number of events in the training set. Thus, in order to improve the existing result, the first step would be to generate a larger number of events. Also, the BDT classifier decides which class the event belongs to based on the value of returned probability. By increasing the threshold of the returned probability it is possible to improve signal to background ratio.

In order to achieve the obtained discovery level significance with our analysis strategy we found a required integrated luminosity $\sqrt{L} = \sigma_{discovery}$

References

[1] Portraying Double Higgs at the Large Hadron Collider - Kim, Jeong Han et al. JHEP 1909 (2019) 047 arXiv:1904.08549 [hep-ph]

[2] Probing the Triple Higgs Self-Interaction at the Large Hadron Collider - Kim, Jeong Han et al. Phys.Rev.Lett. 122 (2019) no.9, 091801 arXiv:1807.11498 [hep-ph]

[3] Combination of searches for Higgs boson pairs in pp collisions at $\sqrt{s}$ =13 TeV with the ATLAS detector - ATLAS Collaboration (Aad, Georges et al.) arXiv:1906.02025 [hep-ex] CERN-EP-2019-099

[4] Search for resonant and nonresonant Higgs boson pair production in the $b\bar{b}$ $l\nu l\nu$ final state in proton-proton collisions at =13 TeV - CMS Collaboration (Sirunyan, Albert M et al.) JHEP 1801 (2018) 054 arXiv:1708.04188 [hep-ex] CMS-HIG-17-006, CERN-EP-2017-168

[5] `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageBR`

[6] D. de Florian et al. [LHC Higgs Cross Section Working Group], doi:10.23731/CYRM-2017-002 arXiv:1610.017922 [hep-ph].

[7] J. Grigo, K. Melnikov and M. Steinhauser, Nucl. Phys. B 888, 17 (2014) doi:10.1016/j.nuclphysb.2014.09.003 [arXiv:1408.2422 [hep-ph]].

[8] M. Czakon, P. Fiedler and A. Mitov, Phys. Rev. Lett. 110, 252004 (2013) doi:10.1103/PhysRevLett.110.252004 [arXiv:1303.6254 [hep-ph]].

[9] Revisiting the non-resonant Higgs pair production at the HL-LHC - Adhikary, Amit et al. JHEP 1807 (2018) 116 arXiv:1712.05346 [hep-ph] LAPTH-062-17, IPPP-17-100, LAPTH-062/17, IPPP/17/100

[10] Projected performance of Higgs analyses at the HL-LHC for ECFA 2016 - CMS Collaboration CMS-PAS-FTR-16-002

[11] Double Higgs boson production and Higgs self-coupling extraction at CLIC - P. Roloff et al. (2019), arXiv:1901.05897v1 [hep-ex].

[12] Probing the Higgs self-coupling through double Higgs production in vector boson scattering at the LHC. - E. Arganda et al. (2019) doi:10.1016/j.nuclphysb.2019.114687 arXiv:1807.09736 [hep-ph].