PROCEEDINGS OF SCIENCE

Higgs self-coupling measurements in the HL-LHC era: new approaches for the $hh \rightarrow 4b$ final state.

Santiago Paredes Sáenz[a,†,*]

[a] *IIHE(ULB-VUB), Université libre de Bruxelles, ULB, Brussels, Belgium*

E-mail: santiago.paredes@cern.ch

Searches for pairs of Higgs bosons will be, in all likelihood, the best tools to precisely measure the Higgs boson self-coupling λ_{hhh} in future colliders. We study various strategies for the $hh \rightarrow b\bar{b}b\bar{b}$ search in the HL-LHC era with focus on constraining λ_{hhh}. We implement a machine-learning-based approach to separate signal and background and apply recent advances in machine learning interpretability, compare the traditional 4 b-jet reconstruction to final states with 1 or 2 large-radius jets, and test scenarios with different top-quark Yukawa couplings, among other factors.

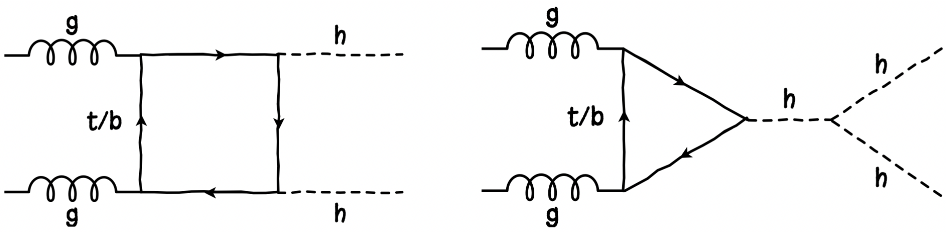
**** The European Physical Society Conference on High Energy Physics (EPS-HEP2021), ****
**** 26-30 July 2021 ****
**** Online conference, jointly organized by Universität Hamburg and the research center DESY ****

[*] Speaker
[†] On behalf of the authors of [1]

PoS(EPS-HEP2021)608

1. Introduction

Detecting the production of a pair of Higgs bosons is one of the current principal goals of the Large Hadron Collider (LHC) experiments, as well as future high-energy physics experiments. This process, often referred to as di-Higgs or hh production, opens a direct experimental probe of the Higgs self-coupling λ_{hhh}, a key parameter in the Standard Model (SM) of particle physics. The value of λ_{hhh} has a profound impact on the dynamics of electroweak symmetry breaking, and Beyond the Standard Model (BSM) physics can modify this coupling. Higgs boson pair events are most frequently produced via gluon-gluon fusion at the LHC through the diagrams shown in figure 1, with the Triangle diagram being sensitive to the Higgs boson self-coupling.



(a) Box **(b)** Triangle

Figure 1: Di-Higgs production modes via gluon-gluon fusion.

1.1 The $hh \to b\bar{b}b\bar{b}$ channel

With around 58%, the $h \to b\bar{b}$ is the most likely decay of the SM Higgs boson, making $b\bar{b}b\bar{b}$ the most likely hh decay. This final state, however, is obscured by formidable multi-jet backgrounds from QCD and top-quark processes, complicating the search. This study applies various strategies used by the LHC experiments to search for $hh \to b\bar{b}b\bar{b}$ and explores novel extensions to these techniques focusing on the channel's sensitivity to the Higgs self-coupling.

We study three separate final state topologies that depend on the transverse momenta of the Higgs bosons. In the *resolved* channel, with the lowest Higgs boson momentum, the signal has four distinct small-radius jets each product of a b-decay. At the other end of the spectrum, the decay products of highly energetic Higgs bosons obtain a high enough Lorentz boost that individual small-radius jets would overlap. The *boosted* channel then uses large-radius jets to capture the full Higgs decay product, and jet substructure is used to infer the presence of a b-decay within them. These two cases are shown in the diagram in figure 2. A third *intermediate* channel covers the region between the two above cases, where one Higgs is captured by a large-radius jet, and the other is reconstructed with two distinct small-radius jets.

2. Data sets

Several simulations were produced with various values of λ_{hhh}, labelled by different κ_λ values, where $\kappa_\lambda = \lambda_{hhh}/\lambda_{\mathrm{SM}}$ represents the ratio of the tested coupling λ_{hhh} to its SM expectation λ_{SM}. Figure 3a shows the di-Higgs invariant mass in several of these samples. Samples were also

PoS(EPS-HEP2021)608

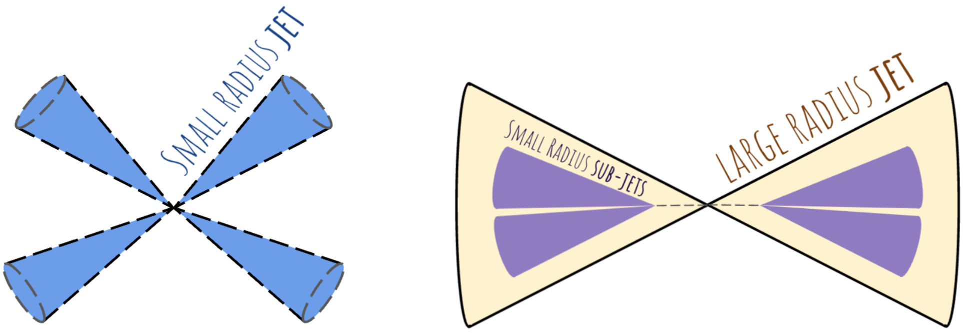
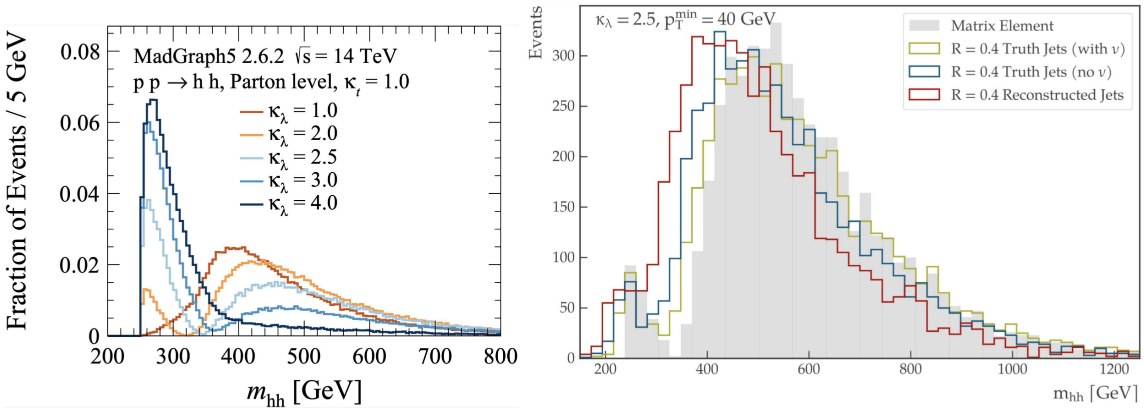


Figure 2: Diagrams depicting the *resolved* (left) and *boosted* (right) topologies of the $hh \to 4b$ decay. In the *intermediate* topology, not shown here, on Higgs boson is reconstructed as two small radius jets, and the other as a single large radius jet.



(a) m_{hh} calculated from several of the κ_λ variations of the signal.

(b) m_{hh} from the $\kappa_\lambda = 2.5$ sample after applying a 40 GeV requirement on the reconstructed jets.

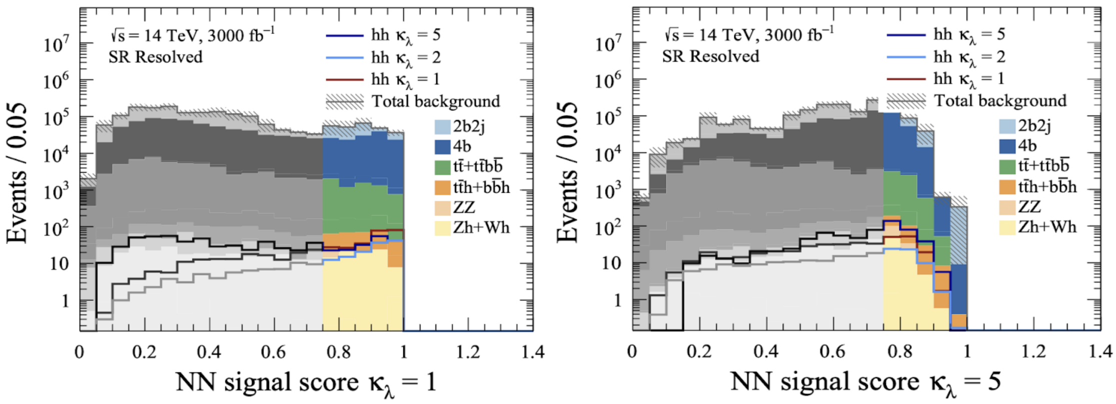
Figure 3: Invariant mass spectrum of the di-Higgs system. The structure seen on some of the signal variations in 3a is affected by the minimum jet p_T requirement as well as some effects of jet reconstruction. The various histograms in 3b show this distribution at different stages of jet reconstruction, and with/without including the contribution of neutrinos.

generated with varied coupling of the Higgs to the top quark κ_t. See [1] for more details on the simulated samples.

We observe that the double peak structure in the invariant mass spectrum seen for some of the signal samples in 3a, particularly at $\kappa_\lambda = 2.5$, is affected significantly by the lower threshold on transverse momentum p_T imposed on the jets, as well as various factors of jet reconstruction, as shown in figure 3b. We also observed that this effect is mitigated by lowering the 40 GeV threshold.

3. Analysis strategy

The analysis proceeds for all three regions (resolved, intermediate, and boosted) by applying several requirements on various kinematic properties of the jets and also event-level properties such



(a) **(b)**

Figure 4: Neural network signal score for the resolved analysis. The networks were trained with a signal sample of $\kappa_\lambda = 1$ (4a) and $\kappa_\lambda = 5$ (4b). The gray area represents the events rejected from the final signal region.

as missing transverse energy (see [1] for details). A signal region is defined for each channel in the m_{hh} variable, which are the final selection for the *baseline* analysis. Events in these signal regions are used in a simple statistical analysis to evaluate the sensitivity of each channel.

A second version of the analysis named the *Deep Neural Network (DNN)* analysis is run in parallel after the initial event selection. In this analysis, deep neural networks are trained for each channel to separate signal from the two main backgrounds (multijet QCD and $t\bar{t}$). The performance of two of these networks is shown in figure 4. Interestingly, the network trained with $\lambda_{hhh} = \lambda_{\mathrm{SM}}$ classified a significant fraction of some of the BSM signals as background (see 4a), which suggests that analyses optimized exclusively on SM signal could loose sensitivity to BSM signals. The final signal region of the DNN analyses is also defined in m_{hh} but only selecting events with a neural network signal score above 0.75 on the DNN trained with a signal of $\kappa_\lambda = 5$.

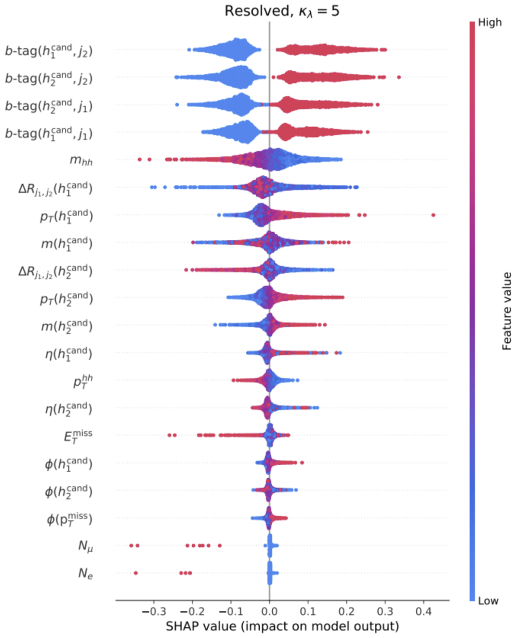
The importance of each variable used for the DNN training was studied using the SHAP values framework [6]. Figure 5a shows the inputs used to train the resolved $\kappa_\lambda = 5$ neural network, ranked by their impact on the signal score as measured by their SHAP value. The presence of b-tagged jets and low m_{hh} were among the most important features to classify an event as signal, while high m_{hh} and the presence of missing transverse momentum E_t^{miss} were important to classify events as background.

4. Results

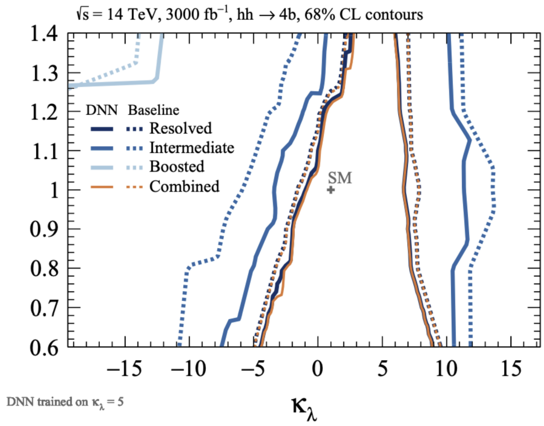
The final signal regions of both DNN and baseline analyses of each of the resolved, intermediate and boosted channels are then analyzed. A χ^2 test is used to compare the sensitivity of each, with the following definition:

$$\chi^2 = \sum_i \left[\frac{(S - S_{\mathrm{SM}})^2}{S + B + (\zeta_b B)^2 + (\zeta_s S)^2} \right]_i . \qquad (1)$$

where B is the total background rate, S_{SM} is the SM signal rate, and S is the signal rate to be distinguished from the SM counterpart. Figure 5b shows the $\chi^2 = 1$ contours obtained for the



(a) Inputs used to train the resolved $\kappa_\lambda = 5$ neural network, ranked by their impact on the signal score as measured by their SHAP value [6].



(b) $\chi^2 = 1$ contours obtained for the resolved, intermediate and boosted categories. The results of the baseline analyses are shown in dotted lines and the DNN results in solid lines.

Figure 5: DNN training variables ranked by their impact on the signal score (5a) and $\chi^2 = 1$ contours of the baseline and DNN analyses (5b).

resolved, intermediate and boosted categories for the baseline (DNN) analysis in dotted (solid) lines in the $\kappa_\lambda - \kappa_t$ plane. The resolved category showed the highest constraining power, followed by intermediate and then boosted. The DNN analysis shows tighter constraints, with notable improvement for the negative κ_λ values in the intermediate category. This plot also shows the constraints on κ_λ change considerably with changes of κ_t, which suggests that the precision of the measurement of the coupling of the Higgs boson to the top quark will affect the constraint of the Higgs self coupling.

5. Conclusions

New techniques for searches of $pp \to hh \to b\bar{b}b\bar{b}$ events, extending some of the methods used in recent works by ATLAS [2] and CMS [3–5], were tested in this analysis. The use of deep neural networks for signal-background discrimination provided noticeable improvement, and, interestingly, the best results came from networks trained on non-SM signal events. This suggests that searches optimized only to search for SM Higgs boson pairs could be sub-optimal for constraining λ_{hhh}. We also showed that one of the main discriminating variables of these analyses, m_{hh}, is highly sensitive to experimental effects from triggering and reconstruction. Despite the challenges of using the $b\bar{b}b\bar{b}$ channel to probe λ_{hhh}, it has moderate constraining power among the hh decay channels, and it can provide independent information as part of a statistical combination.

References

[1] J. Amacker et al. *Higgs self-coupling measurements using deep learning in the $b\bar{b}b\bar{b}$ final state*, *JHEP* **12** (2020) 115.

[2] ATLAS Collaboration, *Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using proton–proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, *Phys. Rev.* **D94** (2016), no. 5 052002, [arXiv:1606.04782].

[3] CMS Collaboration, *Search for nonresonant Higgs boson pair production in the $b\bar{b}b\bar{b}$ final state at $\sqrt{s}$ = 13 TeV*, *JHEP* **04** (2019) 112, [arXiv:1810.11854].

[4] CMS Collaboration, *Search for production of Higgs boson pairs in the four b quark final state using large-area jets in proton-proton collisions at $\sqrt{s}$ = 13 TeV*, *JHEP* **01** (2019) 040, [arXiv:1808.01473].

[5] CMS Collaboration, *Search for a massive resonance decaying to a pair of Higgs bosons in the four b quark final state in proton-proton collisions at $\sqrt{s}$ = 13 TeV*, *Phys. Let.* **B781** (2018) 244–269, [arXiv:1710.04960].

[6] Lundberg, Scott M and Lee, Su-In, *A Unified Approach to Interpreting Model Predictions*, *Advances in Neural Information Processing Systems* **30** (2017), [arXiv:1705.07874].