Snowmass Letter of Interest, Topic Areas: EF01, EF02, CompF03

Improving Di-Higgs Sensitivity in Hadronic Final States with Machine Learning

Artur Apreseyan[1], Javier Duarte * [2], Sanmay Ganguly † [3], Raghav Kansal[2], Jay Lawhorn[4], Nan Lu[4], Samadrita Mukherjee[5], Cristina Mantilla Suarez[1], Si Xie[4], and Cristián Peña[1]

[1]Fermi National Accelerator Laboratory, Batavia, IL 60510, USA
[2]University of California San Diego, La Jolla, CA 92130, USA
[3]Weizmann Institute of Science, Rehovot 7610001 Israel
[4]California Institute of Technology, Pasadena, CA 91125, USA
[5]Indian Association for the Cultivation of Science, Kolkata 700032 India

ABSTRACT

One of the central goals of the physics program at the future colliders is to elucidate the origin of electroweak symmetry breaking, including precision measurements of the Higgs sector. This includes a detailed study of Higgs boson pair production, which can reveal the Higgs self-interaction strength. Since the discovery of the Higgs boson, a large campaign of measurements of the properties of the Higgs boson has begun and many new ideas have emerged during the completion of this program. One such idea is the use of highly boosted and merged hadronic decays of the Higgs boson ($\mathrm{H} \to \mathrm{b\bar{b}}$, $\mathrm{H} \to \mathrm{WW} \to \mathrm{q\bar{q}q\bar{q}}$) with machine learning methods to improve the signal-to-background discrimination. In this letter of interest, we champion the use of these modes to boost the sensitivity of future collider physics programs to Higgs boson pair production and the Higgs self-coupling.

1 Introduction

Observing the standard model (SM) production of two Higgs bosons (H) and precisely measuring the corresponding Higgs self-coupling λ is a key goal of future colliders such as the high-luminosity LHC (HL-LHC) and Future Circular Collider in hadron mode (FCC-hh). Current projections [1] achieve an expected significance of approximately $4.0\,\sigma$ from CMS and ATLAS combined for the full HL-LHC data set. However, these projections do not include dedicated analyses of highly boosted hadronic final states, which may be especially sensitive to the SM and anomalous Higgs self-couplings [2].

2 Boosted Higgs

In general, the hadronic final states are attractive because of their large branching fractions relative to other channels. While the $\mathrm{b\bar{b}}\gamma\gamma$ "golden channel" has a 0.26% branching fraction, the $\mathrm{b\bar{b}b\bar{b}}$ and $\mathrm{b\bar{b}WW}$ channels have a combined 58.8% branching fraction, which often produce a fully hadronic final state. At low transverse momentum (p_{T}), these final states are difficult to disentangle from the background, but at high-p_{T}, the decay products merge into a single jet, which new machine learning (ML) methods can identify with exceptionally high accuracy. Even with a requirement on the p_{T} of the Higgs boson, the hadronic final states are still appealing in terms of signal acceptance. For example, the boosted $\mathrm{b\bar{b}b\bar{b}}$ ($\mathrm{b\bar{b}WW}$) channel with $p_{\mathrm{T}} > 400$ GeV has X times (Y times) more signal events than the "golden" $\mathrm{b\bar{b}}\gamma\gamma$ channel at the LHC. Given the higher center-of-mass energy of the FCC-hh, the boosted fraction would increase.

Based on preliminary investigations, these boosted channels may be competitive with the $\mathrm{b\bar{b}}\gamma\gamma$ channel ($2.7\,\sigma$ expected significance with the full ATLAS and CMS HL-LHC data set). As such, exploring these additional final states with new methods will be crucial to achieving the best possible sensitivity to the Higgs self-coupling.

3 Machine Learning for Di-Higgs Searches

Emerging ML techniques, including convolutional neural networks (CNNs) and graph neural networks (GNNs), have enabled better identification of these boosted Higgs boson jets while reducing the backgrounds [3–8]. CNNs treat the jet input data as

*jduarte@physics.ucsd.edu
†sanmay.ganguly@weizmann.ac.il

either a list of particle properties or as an image. In the image representation case, CNNs leverage the symmetries of an image, namely translation invariance, in their structure. Deeper CNNs are able to learn more abstract features of the input image in order to classify them correctly. GNNs are also well-suited to these tasks because of their structure. GNNs treat the jet as an unordered graph of interconnected constituents (nodes) and learn relationships between pairs of these connected nodes. These relationships then update the features of the nodes in a *message-passing* [9] or *edge convolution* [10] step. Afterward, the collective updated information of the graph nodes can be used to infer properties of the graph, such as whether it constitutes a Higgs boson jet. In this way, GNNs learn pairwise relationships among particles and use this information to predict properties of the jet.

Significantly, it has been shown that these ML methods can identify several classes of boosted jets better than previous methods. For instance these methods have been used to search for highly boosted $\mathrm{H}(\mathrm{b}\overline{\mathrm{b}})$ [11] and $\mathrm{VH}(\mathrm{c}\overline{\mathrm{c}})$ [12] in CMS. We intend to study the impact of the use of these ML algorithms in future colliders like the FCC-hh in a variety of boosted HH final states, including $\mathrm{b}\overline{\mathrm{b}}\mathrm{b}\overline{\mathrm{b}}$, $\mathrm{W}(\mathrm{q}\overline{\mathrm{q}})\mathrm{W}(\mathrm{q}\overline{\mathrm{q}})\mathrm{b}\overline{\mathrm{b}}$, $\mathrm{W}(\mathrm{q}\overline{\mathrm{q}})\mathrm{W}(\ell\nu)\mathrm{b}\overline{\mathrm{b}}$, and $\mathrm{b}\overline{\mathrm{b}}\gamma\gamma$. The study will cover the sensitivity of different jet reconstruction algorithms in the previously mentioned final states. When estimating the signal sensitivity, we will need to consider uncertainties on the signal efficiency, which may be constrained in data using a other SM processes as a proxy for the signal. We will also investigate the implications of these experimental considerations on optimal analysis and detector design. An important deliverable is to understand the sensitivity improvement from considering these channels on both the HH signal strength and the self-coupling.

4 Outlook

Higgs pair production is a crucial process to characterize and measure precisely at future colliders. In order to do so with the best precision possible, it is important to exploit all possible production and decay modes. This includes the high-p_{T} hadronic final states, whose sensitivity can be improved with ML methods. Quantifying the impact of these final states on the ultimate sensitivity achievable (and how these considerations may impact optimal detector design) is the target of this study.

References

1. A. Dainese et al., eds., "Report on the Physics at the HL-LHC,and Perspectives for the HE-LHC", volume 7. 7/2019 of *CERN Yellow Reports: Monographs*. CERN, 2019. doi:10.23731/CYRM-2019-007, ISBN 9789290835493,
2. F. Kling et al., "Maximizing the significance in Higgs boson pair analyses", *Phys. Rev. D* **95** (2017) 035026, doi:10.1103/PhysRevD.95.035026, arXiv:1607.07441.
3. J. Lin, M. Freytsis, I. Moult, and B. Nachman, "Boosting $H \to b\bar{b}$ with Machine Learning", *J. High Energy Phys.* **10** (2018) 101, doi:10.1007/JHEP10(2018)101, arXiv:1807.10768.
4. H. Qu and L. Gouskos, "ParticleNet: Jet tagging via particle clouds", *Phys. Rev. D* **101** (2020) 056019, doi:10.1103/PhysRevD.101.056019, arXiv:1902.08570.
5. E. A. Moreno et al., "JEDI-net: a jet identification algorithm based on interaction networks", *Eur. Phys. J. C* **80** (2020) 58, doi:10.1140/epjc/s10052-020-7608-4, arXiv:1908.05318.
6. E. A. Moreno et al., "Interaction networks for the identification of boosted $H \to b\bar{b}$ decays", *Phys. Rev. D* **102** (2020) 012010, doi:10.1103/PhysRevD.102.012010, arXiv:1909.12285.
7. E. Bernreuther et al., "Casting a graph net to catch dark showers", arXiv:2006.08639.
8. CMS Collaboration, "Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques", *J. Intrum.* **15** (2020) P06005, doi:10.1088/1748-0221/15/06/P06005, arXiv:2004.08262.
9. J. Gilmer et al., "Neural Message Passing for Quantum Chemistry", in *ICML*. 2017. arXiv:1704.01212.
10. Y. Wang et al., "Dynamic graph CNN for learning on point clouds", *ACM Trans. Graph.* **38** (2019) doi:10.1145/3326362, arXiv:1801.07829.
11. CMS Collaboration, "Inclusive search for highly boosted Higgs bosons decaying to bottom quark-antiquark pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV", arXiv:2006.13251.
12. CMS Collaboration, "A search for the standard model Higgs boson decaying to charm quarks", *J. High Energy Phys.* **03** (2020) 131, doi:10.1007/JHEP03(2020)131, arXiv:1912.01662.