

PS

Measuring Higgs Pair Production in the $b\bar{b}b\bar{b}$ Final State at the HL-LHC with Multivariate Techniques

J. Katharina Behr*†

Deutsches Elektronen-Synchrotron (DESY), Hamburg E-mail: katharina.behr@cern.ch

In this contribution we present a new strategy for measuring Higgs pair production in the $b\bar{b}b\bar{b}$ final state at the Large Hadron Collider (LHC) and its future high-luminosity upgrade, the HL-LHC. This process is of particular interest as it allows for the extraction of the Higgs trilinear coupling and therefore provides a crucial test of electroweak symmetry breaking. Moreover, it is sensitive to effects of physics beyond the Standard Model. The measurement of Higgs pair production is therefore considered one of the key goals of LHC physics programme. In our analysis, we take into account all possible Higgs decay topologies to provide optimal sensitivity over a large kinematic range. Our analysis combines a traditional cut-based approach with multivariate analysis techniques and has been shown to be robust in a high pile-up environment. All relevant backgrounds are taken into account, including the 2b2j component of QCD multi-jet production, which yields a non-negligible contribution due to light- and charm-quark jet mis-identification. We obtain a signal significance of $S/B \simeq 3$ for an integrated luminosity of L = 3 ab⁻¹ and show that, pending various experimental improvements, the $b\bar{b}b\bar{b}$ final state alone may allow for the observation of Higgs pair production at the HL-LHC.

Fourth Annual Large Hadron Collider Physics 13-18 June 2016 Lund, Sweden

*Speaker.

[†]This work was supported by the Rhodes Trust and a Helmholtz Postdoctoral Fellowship (PD-302).

© Copyright owned by the author(s) under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0).

1. Introduction

The measurement of Higgs pair production is one of the central goals of the physics programme at the Large Hadron Collider (LHC) and its future high-luminosity upgrade, the HL-LHC, which is expected to collect around 3 ab⁻¹ of data at or near its design energy of $\sqrt{s} = 14$ TeV [1, 2]. Current measurements of singly produced Higgs bosons only probe the Higgs potential around its minimum value. To determine the full shape of the Higgs potential, a measurement of the Higgs self-coupling is required. Higgs pair production is directly sensitive to the Higgs trilinear coupling λ and will therefore provide a crucial test of the electroweak symmetry breaking mechanism.

The key challenge in measuring this process is its small cross-section. In the Standard Model, the total cross-section for Higgs pair production from gluon fusion, the dominant production mechanism, is approximately 40 fb at next-to-next-to-leading order (NNLO) in QCD at $\sqrt{s} = 14$ TeV. However, many Beyond SM (BSM) processes are expected to enhance this cross-section. The $b\bar{b}b\bar{b}$ final state is of particular interest for the measurement of Higgs pair production as the decay to a $b\bar{b}$ pair has the largest branching ratio of all Higgs boson decay channels, $BR(h \rightarrow b\bar{b}) \approx 0.57$ [3]. At the same time, the overwhelming background from multi-jet production renders it a particularly challening final state. Previous studies of Higgs pair production in this final state [4, 5] concluded that a signal significance $S/\sqrt{B} \simeq 2.0$ could be reached for an integrated luminosity of $\int \mathcal{L} = 3$ ab⁻¹ at $\sqrt{s} = 14$ TeV.

In our recent feasibility study [6], we propose a new strategy for measuring Higgs pair production from gluon fusion in the $b\bar{b}b\bar{b}$ final state that is based on the combination of a traditional cut-based approach and multivariate analysis (MVA). For the first time, all possible Higgs decay topologies are taken into account to optimise the sensitivity over a large kinematic range. We consider all relevant backgrounds, including the 2b2j multi-jet component that has been neglected in previous studies but that we find to yield a non-negligible contribution due to mis-identified light-quark and gluon jets from the parton shower. Our analysis strategy has been optimised for robustness in a high-pile-up environment. We find that the $b\bar{b}b\bar{b}$ alone may allow for observation of Higgs pair production at the HL-LHC and identify keys to further sensitivity improvements.

2. Modelling of signal and background processes

Higgs pair production is simulated at leading order (LO) with MADGRAPH5_AMCATNLO [7] with a dedicated model for double Higgs boson production via gluon-fusion [8]. Mass effects from the exact form factors for top-quark triangle and box loops are taken into account [9]. The simulation is performed in the four-flavour scheme ($n_f = 4$). The renormalisation and factorisation scales are chosen to be $\mu_R = \mu_F = H_T/2$. The NNPDF 3.0 $n_f = 4$ LO PDF set [10] with $\alpha_s(m_Z^2) = 0.118$, as provided in LHAPDF6 [11], is used. The simulated cross-section is rescaled to the total inclusive cross-section calculated at NNLO with corrections from soft-gluon resummation up to next-to-next-to-leading logarithmic accuracy (NNLL) [12, 13]. The parton-level events are showered using PYTHIA8 [14, 15] v8.201 with the Monash tune [16] and the NNPDF 2.3 LO PDF set [17, 18].

The backgrounds from QCD multi-jet production are generated at LO with SHERPA [19] v2.1.1 with the same PDF set and scales as used for the signal processes. We consider QCD 4b multi-jet production, as well as QCD 2b2j and 4j production, and top quark pair produc-

J. Katharina Behr

tion with fully hadronic final states. The simulated cross-sections for the 4b, 2b2j, and 4j processes are rescaled to NLO precision based on results obtained with MADGRAPH5_AMCATNLO (4b and 2b2j) and BLACKHAT [20] (4j). The simulated cross-section for the $t\bar{t}$ component is rescaled to NNLO+NNLL precision [21]. The values for the k-factors are listed in Ref. [6]. Other background processes, such as single Higgs production in the $Z(\rightarrow b\bar{b})h(\rightarrow b\bar{b})$, and $t\bar{t}h(\rightarrow b\bar{b})$ channels, and electroweak backgrounds, such as e.g $Z(\rightarrow b\bar{b})b\bar{b}$, have been simulated with MAD-GRAPH5_AMCATNLO but found to yield a significantly smaller contribution in the signal regions than the OCD backgrounds and are therefore not included in the analysis.

Pile-up (PU) is simulated by overlaying a certain number n_{PU} of Minimum Bias events, generated with PYTHIA8, on each signal and background event. The SOFTKILLER method [22], as implemented in FASTJET [23, 24], is used to subtract PU contaminations at the event level. Two scenarios are explored, one with $n_{PU} = 80$ and one with $n_{PU} = 150$. The combined signal significances are similar for both scenarios and the former is adopted as a baseline scenario.

3. Event selection and reconstruction

The final state particles obtained after the parton shower are clustered with jet reconstruction algorithms implemented in FASTJET v3.1.0. *Small-R jets* are reconstructed using the anti- k_t algorithm [25] with size parameter R = 0.4. Only small-R jets with transverse momentum $p_T^{\text{jet}} > 40 \text{ GeV}$ and pseudorapidity $|\eta^{\text{jet}}| < 2.5$ are considered in the analysis. *Large-R jets* are reconstructed with the same algorithm with R = 1.0. A trimming [26] procedure with parameters $R_{\text{trim}} = 0.2$ and $p_T^{\text{frac}} = 0.05$ is applied to mitigate the effects of PU on the large-R jets properties. Large-R jets are required to satisfy the requirements $p_T^{\text{jet}} > 200 \text{ GeV}$ and $|\eta^{\text{jet}}| < 2.0$. In addition, they must satisfy the BDRS mass-drop tagger (MDT) [27] conditions with parameters $\mu_{\text{mdt}} = 0.67$ and $y_{\text{mdt}} = 0.09$.

The identification of jets from *b*-quarks (*b*-tagging) plays a key role in achieving a high signal purity in the $b\bar{b}b\bar{b}$ final state. A small-*R* jet is *b*-tagged with a probability of $f_b = 0.80$ if there is at least one *b*-quark with $p_T > 15$ GeV among its constituents. A small-*R* jet with no *b*-quark but at least one *c*-quark with $p_T > 15$ GeV among its constituents is mis-tagged as a *b*-jet with probability $f_c = 0.10$. The mis-tag probability for jets from light quarks (*d*, *u*, *s*) and gluons is assumed to be $f_l = 0.01$. Only jets that have at least one constituent with $p_T > 15$ GeV can be *b*-tagged and only the four leading small-*R* jets in an event are considered for *b*-tagging to reduce the background due to mis-tagged jets. Large-*R* jets are considered *b*-tagged if they have at least two matching anti- k_t R = 0.3 subjets that are *b*-tagged by the same criteria as used for small-*R* jets [28].

Events are selected into three mutually exclusive categories based on the Higgs decay topology. First, events with at least two selected large-*R* jets, the leading two of which are taken as the Higgs candidates, are assigned to the *boosted* category. Events with exactly one selected large-*R* jet, taken as the leading Higgs candidate, and at least two *b*-tagged small-*R* jets with an angular separation $\Delta R > 1.2$ from the large-*R* jet are assigned to the *intermediate* category. The remaining events may be classified into the *resolved* category if they contain at least four *b*-tagged small-*R* jets. In this case, the two Higgs candidates are chosen to be the two dijet combinations with the smallest mass difference. In all categories, only events for which the invariant mass of each Higgs candidate lies within a symmetric mass window of width 80 GeV around 125 GeV are considered. The selection criteria are deliberately loose as the final selection is determined by the MVA.

4. Multivariate analysis

The events selected in the cut-based analysis are processed with a multi-layer feed-forward artificial neural network (ANN), known as a *perceptron* or *deep neural network*, to optimise the separation between signal and background. This is done separately for the three event categories. The ANN architecture is given by $N_{\text{Var}} \times 5 \times 3 \times 1$, where N_{Var} denotes the number of input variables. Input variables include kinematic properties of the reconstructed Higgs boson candidates as well as a number of substructure variables in the case of the intermediate and boosted categories. The full list of input variables, along with their relevance for the discrimination between signal and background, as obtained from the trained ANNs in a fully automated way, are listed in Ref. [6]. Events are classified as signal or background based on a cut, y_{cut} , on the ANN output, which is chosen such as to optimise the signal significance in a given category.

5. Results

The number of signal and background events obtained in the three categories after the selection cut on the ANN output are given in Table 1, along with the corresponding signal significances S/\sqrt{B} for an integrated luminosity of $\int \mathcal{L} = 3 \text{ ab}^{-1}$. The signal significance for the combination of the three categories is derived by adding those for the individual categories in quadrature.

Category			signal	background		C / D	C / D
		ycut	N _{ev}	N _{ev} ^{tot}	$N_{\rm ev}^{\rm 4b}$	$S/\sqrt{D_{tot}}$	$S/\sqrt{D_{4b}}$
Boosted	no PU	0.80	290	$1.2 \cdot 10^4$	$8.0 \cdot 10^{3}$	2.7	3.2
	PU80+SK+Trim	0.80	290	$3.7 \cdot 10^4$	$1.2 \cdot 10^{4}$	1.5	2.7
Intermediate	no PU	0.75	130	$3.1 \cdot 10^{3}$	$1.5 \cdot 10^{3}$	2.3	3.3
	PU80+SK+Trim	0.75	140	$5.6 \cdot 10^3$	$2.4 \cdot 10^{3}$	1.9	2.9
Resolved	no PU	0.50	630	$1.1 \cdot 10^{5}$	$5.8 \cdot 10^{4}$	1.9	2.7
	PU80+SK	0.60	640	$1.0 \cdot 10^{5}$	$7.0 \cdot 10^{4}$	2.0	2.6
Combined	no PU					4.0	5.3
	PU80+SK+Trim					3.1	4.7

Table 1: Number of signal and background events in the three event categories after the selection requirement on the ANN output with cut value y_{cut} for an integrated luminosity of $\int \mathscr{L} = 3 \text{ ab}^{-1}$. The number of total background events, $N_{\text{ev}}^{\text{tot}}$, is given along with the number of events from the irreducible 4*b* background, N_{ev}^{4b} . The corresponding signal significances, $S/\sqrt{B_{\text{tot}}}$ and $S/\sqrt{B_{4b}}$ are also given. The results are quoted for the scenarios without PU and for the baseline analysis with $n_{\text{PU}} = 80$ (PU80+SK+Trim).

For the baseline scenario, PU80+SK+Trim, with all backgrounds included, we obtain a signal significance of $S/\sqrt{B_{tot}} \approx 3.1$ for the combination of all categories, which would be enough to claim evidence for Higgs pair production. A signal significance of $S/\sqrt{B_{tot}} \approx 4.7$, close to the threshold for claiming observation, is found if all backgrounds but the irreducible 4*b* component are neglected, indicating that a reduction of the mistag rates for jets from light and charm quarks is of key importance. Further significance improvements can be achieved through more effective PU mitigation techniques, as a comparison of the baseline significance with that obtained for a scenario

without PU shows. We have also found that the sensitivity of the analysis depends strongly on the Higgs mass resolution. Hence jet energy and mass resolution improvements, especially in high-PU environments should be another key objective for future analyses.

6. Conclusion

We have presented a novel strategy for measuring Higgs pair production in the $b\bar{b}b\bar{b}$ final state that combines a traditional cut-based analysis with MVA techniques. All possible Higgs decay topologies and all relevant backgrounds are taken into account. PU effects are also considered. The resulting signal significance of $S/\sqrt{B_{\text{tot}}} \approx 3.1$ represents a notable improvement over previous results, which did not include PU effects. We also show that, pending a number of experimental improvements, the $b\bar{b}b\bar{b}$ final state alone could allow for observation of Higgs pair production.

References

- [1] ATLAS Collaboration, arXiv:1307.7292 [hep-ex].
- [2] CMS Collaboration, arXiv:1307.7135.
- [3] S. Dittmaier et al., arXiv:1201.3084 [hep-ph].
- [4] D. Wardrope et al., Eur. Phys. J. C 75 (2015) no.5, 219 [arXiv:1410.2794].
- [5] D. E. Ferreira de Lima et al., JHEP 1408 (2014) 030 [arXiv:1404.7139].
- [6] J. K. Behr et al., Eur. Phys. J. C 76 (2016) no.7, 386 [arXiv:1512.08928].
- [7] J. Alwall et al., JHEP 1407 (2014) 079 [arXiv:1405.0301].
- [8] F. Maltoni et al., JHEP 1411 (2014) 079 [arXiv:1408.6542].
- [9] T. Plehn et al., Nucl. Phys. B 479 (1996) 46 [hep-ph/9603205].
- [10] R. D. Ball et al. [NNPDF Collaboration], JHEP 1504 (2015) 040 [arXiv:1410.8849].
- [11] A. Buckley et al., Eur. Phys. J. C 75 (2015) 132 [arXiv:1412.7420].
- [12] D. de Florian and J. Mazzitelli, Phys. Rev. Lett. 111 (2013) 201801 [arXiv:1309.6594].
- [13] D. de Florian and J. Mazzitelli, JHEP 1509 (2015) 053 [arXiv:1505.07122].
- [14] T. Sjostrand et al., Comput. Phys. Commun. 178 (2008) 852 [arXiv:0710.3820].
- [15] T. Sjostrand et al., Comput. Phys. Commun. 191 (2015) 159 [arXiv:1410.3012].
- [16] P. Skands et al., Eur. Phys. J. C 74 (2014) no.8, 3024 [arXiv:1404.5630].
- [17] R. D. Ball et al., Nucl. Phys. B 867 (2013) 244 [arXiv:1207.1303].
- [18] R. D. Ball et al. [NNPDF Collaboration], Nucl. Phys. B 877 (2013) 290 [arXiv:1308.0598].
- [19] T. Gleisberg et al., JHEP 02 (2009) 007, [arXiv:0811.4622].
- [20] Z. Bern et al., Phys.Rev.Lett. 109 (2012) 042001, [arXiv:1112.3940].
- [21] M. Czakon et al., Phys. Rev. Lett. 110 (2013) 252004, [arXiv:1303.6254].
- [22] M. Czakon et al., Eur. Phys. J. C75 (2015), no. 2 59, [arXiv:1407.0408].
- [23] M. Czakon et al., Eur.Phys.J. C72 (2012) 1896, [arXiv:1111.6097].
- [24] M. Cacciari et al., Phys. Lett. B641 (2006) 57âÅŞ61, [hep-ph/0512210].
- [25] M. Cacciari et al., JHEP 0804 (2008) 063, [arXiv:0802.1189].
- [26] D. Krohn et al., JHEP 02 (2010) 084, [arXiv:0912.1342].
- [27] J. M. Butterworth et al., Phys.Rev.Lett. 100 (2008) 242001, [arXiv:0802.2470].
- [28] ATLAS Collaboration, Eur. Phys. J. C 75 (2015) no.9, 412 [arXiv:1506.00285].