Dissertation

Automated optimization of sensitivity in a search for pair production of boosted VBF Higgs bosons in the $b\bar{b}b\bar{b}$ quark final state with the ATLAS detector

For the attainment of the academic degree doctor rerum naturalium

(Dr. rer. nat.) in the subject: Physics

Frederic Renner

Berlin, 07.03.2024

Faculty of Mathematics and Natural Sciences of the Humboldt University of Berlin

1st Supervisor: Dr. Clara Elisabeth Leitgeb

2nd Supervisor: Prof. Dr. Cigdem Issever

(Only after the disputation for publication in the university library according to § 15 of the doctoral regulations enter the names and the date):

Reviewers:

1st:

2nd:

3rd:

Date of the oral examination:

Abstract

I am an abstract.

Contents

Chapter 1

Introduction

Understanding nature through first principles is an intrinsic human endeavor. The Standard Model (SM) of particle physics currently stands as the most precise theory, articulating elementary particles and their interactions through symmetry principles. The landmark discovery of the Higgs boson by the Compact Muon Solenoid (CMS) [1] and A Toroidal LHC Apparatus (ATLAS) [2] collaborations at the Large Hadron Collider (LHC) in 2012, a half-century after its theoretical prediction [3, 4], stands not only as a testament of the scientific method but also filled a pivotal gap in our understanding of the universe's fundamental structure.

Yet, this achievement opens the door to new questions regarding the SMs completeness and consistency. As a Quantum Field Theory (QFT), the SM is subject to loop corrections that demand extreme precision for the observed Higgs mass [5], implying that our current understanding may represent only an effective theory, hinting at the existence of undiscovered physics. Additionally, the conventional potential of the Higgs boson, instrumental in maintaining the SMs consistency, relies on the simplest conceivable form. The true complexity of this potential, the possibility of multiple Higgs bosons, or even the Higgs' status as a fundamental particle remain open questions with any deviation from SM predictions signalling a sign of new physics [6]. These considerations extend to the stability of the universe itself, with current precision of measurements suggesting a meta-stable state [7]. The discovery of the Higgs boson is therefore more than just

a milestone; it acts as a portal to uncharted territories in physics and thus requires more precise measurements [8].

The SM encompasses 26 parameters, 15 of which are determined by the Higgs mechanism, emphasizing its centrality in the model [9]. One prediction is the process of Higgs boson pair production, a phenomenon under rigorous scrutiny by both the ATLAS and CMS collaborations across various decay channels [10]. This thesis concentrates on the search for Higgs boson pairs in a boosted topology, decaying into the four b-quark final state using the ATLAS detector [11].

The advent of machine learning in particle physics introduced powerful tools for classification problems but also challenges in optimizing these tools for the field's unique goals. These tools, often tailored for specific tasks, may not align with the overarching goals of discovering new particles or verifying new theories, leading to suboptimal outcomes [12]. This gap primarily arises from depending on intermediate optimization metrics that disregard systematic uncertainties' significant role in the statistical test's validity for theory confirmation. The innovative NEOS approach [13] introduces a solution to this problem, for which this work demonstrates a first application of of a systematic-aware optimization on sensitivity in a particle physics experiment.

This thesis is structured to first introduce the SM **??** and the ATLAS detector **??**, followed by a comprehensive discussion of analytical methodologies, including the reconstruction of physical objects **??**, analysis strategy and event selection 2, machine learning for systematic-aware neural network training **??**, systematic uncertainties **??** and the framework used for the evaluation of statistical tests **??**. The results section presents a strategy for improved b-quark identification using muons **??** and findings on Higgs pair production cross-section limits using Run 2 data. 3.

Chapter 2

The HH→4b analysis

Investigating the exact shape of the Higgs potential is an interesting endeavor, as it is directly related to Electroweak Symmetry Breaking (EWSB) and fundamental questions about the nature of the universe, as discussed in section **??**. The main Higgs production modes at the LHC are shown in Figure 2.1 and can be understood by the fact that the Higgs boson interacts with fermions via Yukawa couplings from equation **??**. Since Yukawa couplings are directly proportional to the fermion masses the Higgs boson predominantly couples to heavier particles like the top quark or the massive vector bosons. All couplings are scaled relative to their SM values and are denoted as $\kappa_c = c/c_{sm}$. A κ_c value of 1 therforee corresponds to the SM value for some given coupling c.

The first two gluon-gluon fusion (GGF) diagrams 2.1(a) and 2.1(b) have a cross-section of $\sigma^{SM}_{vbf\,HH} = 31.05\,fb$ calculated at a center of mass energy of 13 TeV at next-to-next-to-leading order (NNLO) [14] while the vector-boson fusion (VBF) processes (c), (d) and (e) of figure 2.1 have a production cross-section of $\sigma^{SM}_{vbf\,HH} = 1.73\,fb$ at next-to-next-to-next-to-leading order (N^3LO) [15]. A characteristic of the VBF processes is that the Higgs pair products are accompanied by two additional quarks. The VBF cross section is about 3×10^4 times smaller than the production cross section for single Higgs $\sigma^{SM}_{H} = 48.58\,pb$ at the LHC [16] and underlines the challenge of discovering Higgs pairs in these final states.

Figure 2.2 highlights that an interesting channel in the study of Higgs pair production is the final state with the largest branching fraction, which consists

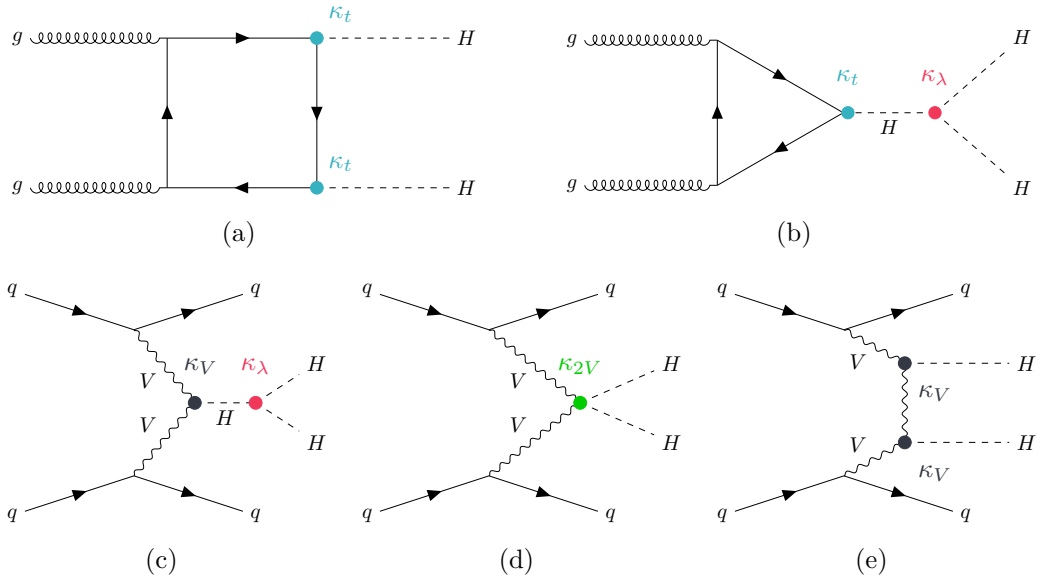


Figure 2.1: Leading Higgs Pair production processes at the LHC. (a), (b) shows GGF and (c), (d), (e) VBF processes. Adopted from [17].

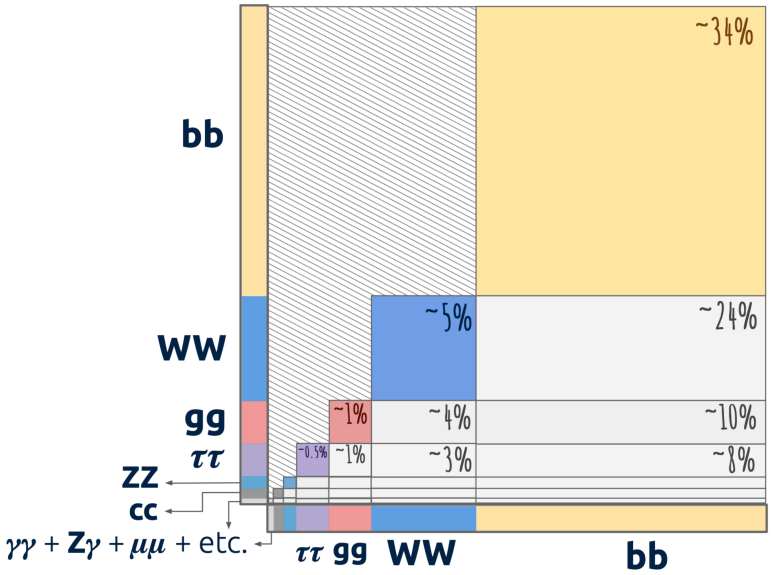


Figure 2.2: Contributions of final states represented by area for a pair of Higgs. Adopted from [18].

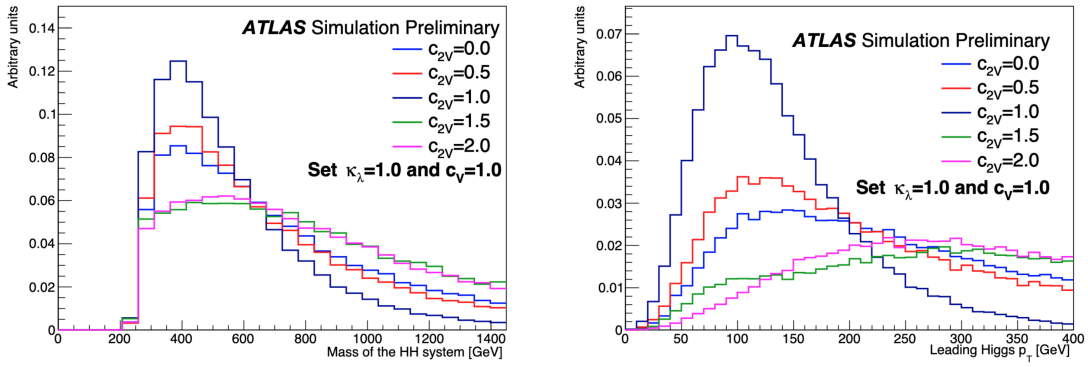


Figure 2.3: Invariant mass of the Higgs pair system and the leading Higgs candidate jet p_{T} reconstructed from simulation for different $\kappa_{2\mathrm{V}}$. Adopted from [19].

of four b quarks and amounts to about 34 %. Thus the SM VBF cross-section is calculated to correspond to the 4b branching ratio by multiplying it with $\mathcal{B}(4b) = 0.3392$. This fully hadronic final state, however, presents the challenge of significant Quantum Chromodynamics (QCD) backgrounds.

This work focuses on the boosted topology of highly energetic jets which do not allow reconstruction of b-jets individually but rather of final states consisting of large-R jets encapsulating two collimated b-jets inside. This approach substantially reduces QCD backgrounds, as highly energetic jets are more likely to originate from heavy particles like b quarks. Additionally, events with jets of large p_{T} are easier to trigger on. While representing a comparatively clean signal, such events are rare and thus have limited statistical power. Despite other decay signatures being more suitable for the discovery of the Higgs pair production process the power of this selection lies in proving the existence of the $\kappa_{2\mathrm{V}}$ coupling shown in figure 2.1(d) to which it is directly sensitive.

The low cross-section for this process is due to the fact that diagrams (d) and (e) in Figure 2.1 exhibit destructive interference for SM values. Conversely, when $\kappa_{2\mathrm{V}}$ deviates from SM values, the production cross-section increases significantly, with $\sigma_{\kappa_{2\mathrm{V}}=0} \approx 20\sigma_{\kappa_{2\mathrm{V}}=1}$ and the decay products exhibit much larger transverse momentum, as illustrated in Figure 2.3.

2.1 Data and Monte Carlo Simulation

This analysis uses the full run 2 data taken by ATLAS between 2015 and 2018. The dataset contains 140.1 fb^{-1} of data good for physics at a center of mass energy of 13 TeV [20].

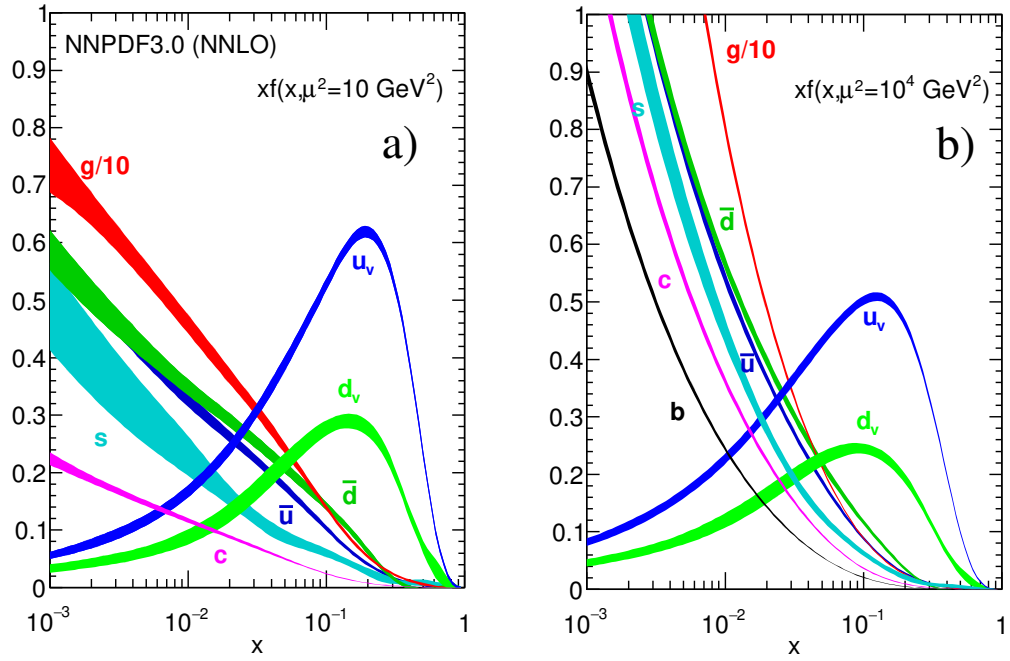
Monte Carlo (MC) generation in ATLAS is typically done in three steps. At first at parton level the matrix element of the process of interest is stochastically simulated with MadGraph (v.2.7.3p3.atlas6) [21]. The cross sectional calculation for proton-proton collisions relies on the factorization theorem [22] which states that contributions from partons participating in the hard scatter event can be factorized. Further partons cannot be observed individually since the approximation of the perturbation ansatz of section **??** breaks down for low energy scales μ^2 as described in section **??**. This is the energy scale for which the approximation would need to hold to describe the partons inside a proton. However parton densities can be studied within QCD using the DGLAP equations [9]. Similar to renormalization, a scaling behavior can be derived from these equations that allows to derive an estimate of the Parton Density Functions (PDFs) by measuring it at some factorization scale μ_F^2 in order to extrapolate it to another. Figure 2.4 exemplifies this for two energy scales from the NNPDF3.0nlo PDF set used in this analysis. Thus for hadrons A, B containing partons a, b and their respective PDFs f_a^A and f_a^B, dependent on the parton's momentum fraction x and factorization scale μ_F^2, the cross-section of a process $A, B \to X$ reads

$$\sigma_{A,B\to X} = \sum_{a,b} \int_0^1 \mathrm{d}x_1 \mathrm{d}x_2 f_a^A(x_1, \mu_F^2) f_b^B(x_2, \mu_F^2) \hat{\sigma}_{a,b\to X}(\alpha_s(\mu_R^2), \mu_R^2). \tag{2.1.1}$$

$\hat{\sigma}_{a,b\to X}$ is the perturbatively calculable part and therefore depends on the strong coupling α_S and renormalization scale μ_R^2.

In a second step the parton shower evolution including hadronization and initial and final state radiation is simulated with Pythia8 [24]. Figure 2.5 illustrates this process.

In a final step the detector response of simulated final state particles is simulated with Geant4 [26]. It models the detector geometry, the particle's path through the magnetic fields and the particle interactions with the detector material, potentially



(a) (b)

Figure 2.4: NNPDF3.0NLO parton distribution functions for two different factorization scales (a) μ_F^2 =10 GeV^2 and (b) μ_F^2 =10 TeV^2 against the momentum fraction x of the particle. Adopted from [23].

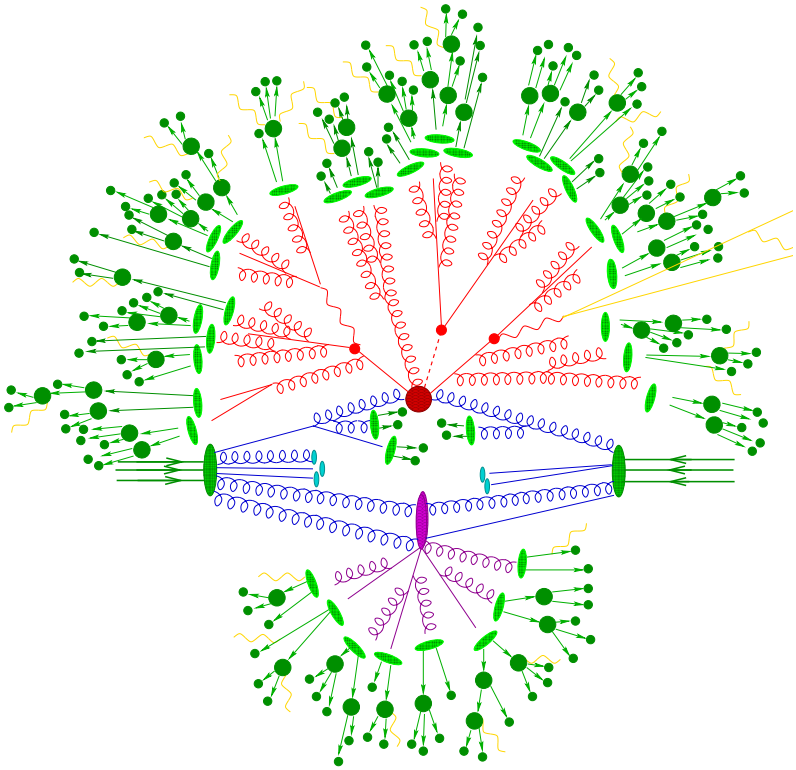


Figure 2.5: Simulation of an evolution of a proton proton collision: The red circle at the center is the hard collision and the purple oval a secondary hard scatter event. Both are surrounded by a tree-like structure of QCD bremsstrahlung interactions simulated with a parton shower. Light green represent hadrons whereas their subsequent decays are shown in dark green. Photons are depicted in yellow. Adopted from [25].

producing new particles or decays. The output of this step are energy deposits in the various subdetectors of ATLAS. Subsequently these are passed on to a process known as digitization which models the readout electronics. The result of this is raw data being no different from that read out in the actual experiment.

2.2 Linear combination of samples

This analysis is interested in constraining the couplings $\kappa_{\mathrm{V}}, \kappa_\lambda, \kappa_{2\mathrm{V}}$ associated to the VBF processes shown in figure 2.1. As computing resources are limited and MC generation is unfortunately computationally expensive only a few hypothesis can be simulated. However by exploiting the properties of the differential cross-sectional calculation, samples of any hypothesized coupling value can be created through linear combination of samples [27]. It is illustrative to consider the two GGF diagrams with the κ_λ and κ_t couplings when calculating the differential cross-section

$$\frac{d\sigma(\kappa_\lambda,\kappa_t)}{dm_{\mathrm{HH}}} = |A(\kappa_t,\kappa_\lambda)|^2 = |\kappa_\lambda\kappa_t M_\triangle(m_{\mathrm{HH}}) + \kappa_t^2 M_\square(m_{\mathrm{HH}})|^2 \tag{2.2.1}$$

$$= \kappa_\lambda^2\kappa_t^2|M_\triangle(m_{\mathrm{HH}})|^2 \tag{2.2.2}$$

$$+ \kappa_\lambda\kappa_t^3[M_\triangle^*(m_{\mathrm{HH}})M_\square(m_{\mathrm{HH}}) + M_\square^*(m_{\mathrm{HH}})M_\triangle(m_{\mathrm{HH}})] \tag{2.2.3}$$

$$+ \kappa_t^4|M_\square|^2 \tag{2.2.4}$$

$$= \kappa_\lambda^2\kappa_t^2 a_1(m_{\mathrm{HH}}) + \kappa_\lambda\kappa_t^3 a_2(m_{\mathrm{HH}}) + \kappa_t^4 a_3(m_{\mathrm{HH}}). \tag{2.2.5}$$

When setting κ_t to its SM value 1 the equation reduces to

$$\frac{d\sigma(\kappa_\lambda)}{dm_{\mathrm{HH}}} = \kappa_\lambda^2 a_1(m_{\mathrm{HH}}) + \kappa_\lambda a_2(m_{\mathrm{HH}}) + a_3(m_{\mathrm{HH}}). \tag{2.2.6}$$

The parameters a_i depend non-trivially on m_{HH}. However for three given hypotheses of κ_λ a linear system of equations with variables a_i can be solved and thus $\frac{d\sigma(\kappa_\lambda)}{dm_{\mathrm{HH}}}(\kappa_\lambda)$ is a function of κ_λ only.

In complete analogy the squared expansion of the cross-sectional formula involving the three VBF couplings $\kappa_{\mathrm{V}}, \kappa_\lambda, \kappa_{2\mathrm{V}}$ samples can be combined to produce

any hypotheses from 6 simulated hypotheses

$$\begin{aligned}
\frac{d\sigma}{dm_{\mathrm{HH}}}(\kappa_{2\mathrm{V}},\kappa_\lambda,\kappa_\mathrm{V}) = \\
\left(\frac{68\kappa_{2V}^2}{135} - 4\kappa_{2V}\kappa_V^2 + \frac{20\kappa_{2V}\kappa_V\kappa_\lambda}{27} + \frac{772\kappa_V^4}{135} - \frac{56\kappa_V^3\kappa_\lambda}{27} + \frac{\kappa_V^2\kappa_\lambda^2}{9}\right) \times \frac{d\sigma}{dm_{\mathrm{HH}}}(1,1,1) \\
+ \left(-\frac{4\kappa_{2V}^2}{5} + 4\kappa_{2V}\kappa_V^2 - \frac{16\kappa_V^4}{5}\right) \times \frac{d\sigma}{dm_{\mathrm{HH}}}\left(\frac{3}{2},1,1\right) \\
+ \left(\frac{11\kappa_{2V}^2}{60} + \frac{\kappa_{2V}\kappa_V^2}{3} - \frac{19\kappa_{2V}\kappa_V\kappa_\lambda}{24} - \frac{53\kappa_V^4}{30} + \frac{13\kappa_V^3\kappa_\lambda}{6} - \frac{\kappa_V^2\kappa_\lambda^2}{8}\right) \times \frac{d\sigma}{dm_{\mathrm{HH}}}(1,2,1) \\
+ \left(-\frac{11\kappa_{2V}^2}{540} + \frac{11\kappa_{2V}\kappa_V\kappa_\lambda}{216} + \frac{13\kappa_V^4}{270} - \frac{5\kappa_V^3\kappa_\lambda}{54} + \frac{\kappa_V^2\kappa_\lambda^2}{72}\right) \times \frac{d\sigma}{dm_{\mathrm{HH}}}(1,10,1) \\
+ \left(\frac{88\kappa_{2V}^2}{45} - \frac{16\kappa_{2V}\kappa_V^2}{3} + \frac{4\kappa_{2V}\kappa_V\kappa_\lambda}{9} + \frac{152\kappa_V^4}{45} - \frac{4\kappa_V^3\kappa_\lambda}{9}\right) \times \frac{d\sigma}{dm_{\mathrm{HH}}}\left(1,1,\frac{1}{2}\right) \\
+ \left(\frac{8\kappa_{2V}^2}{45} - \frac{4\kappa_{2V}\kappa_V\kappa_\lambda}{9} - \frac{8\kappa_V^4}{45} + \frac{4\kappa_V^3\kappa_\lambda}{9}\right) \times \frac{d\sigma}{dm_{\mathrm{HH}}}\left(1,-5,\frac{1}{2}\right).
\end{aligned} \tag{2.2.7}$$

A validation of the method can be found in [28].

2.3 Analysis strategy

This section describes the event selection and analysis strategy. A detailed description of reconstructed physical objects used is described in chapter **??**.

2.3.1 Trigger

As outlined in section **??** events need to be preselected. The high level trigger (HLT) applied in this analysis selects events with a large transverse energy E_T large-R jet. The definition slightly changed over the data taking years as can be seen in table 2.1. Previous studies have shown that they become fully efficient at about $p_\mathrm{T} > 420\,\mathrm{GeV}$ [18, 28].

Table 2.1: Trigger selections per data taking year and minimum requirements on transverse energy E_{T} and mass m on the large R jet.

Year	E_{T}	m
2015	> 360	0
2016	> 420	0
2017	> 420	> 35
2018	> 420	> 40

2.3.2 Large Radius Jets

To fully capture the boosted Higgs pair topology two large $R = 1.0$ jets clustered with the Anti-k_t algorithm from Track CaloClusters (TCCs) are used as described in section **??**. These enclose the two boosted collimated b-jets in each of them to form the Higgs candidates. If there are several large-R jets the two with the highest p_{T} are chosen. To be fully efficient on the trigger the leading large-R jet is required to have $p_{\mathrm{T}} > 450\,\mathrm{GeV}$. For decay products to be inside a jet holds approximately $R \approx 2m/p_{\mathrm{T}}$ with the mass m and transverse momentum of the parent particle [29]. For a Higgs mass of 125 GeV to be contained inside a large-R jet the Higgs candidate therefore must have $p_{\mathrm{T}} \gtrsim 250\,\mathrm{GeV}$ and is thus chosen as the p_{T} requirement on the sub-leading Higgs candidate. If there are several large-R jets the two leading p_{T} jets are selected. Additionally both Higgs candidates have a mass requirement $m > 50\,\mathrm{GeV}$ to reduce QCD background. The $X \to bb$ tagger described in **??** is used to identify b-jets within the selected large-R jets. The top fraction f_{top} is set to 0.25 and the 60 % Higgs efficiency working point (WP) is required. Studies with the more inclusive 70 % WP displayed slightly worse limit results [28].

2.3.3 Small Radius Jets

Two small radius $R = 0.4$ jets are required for the VBF signature and are referred to as VBF jets in the following. They are also reconstructed with the anti-k_t algorithm and as Particle Flow Objects (PFOs) as described in **??**. The tight WP for the jet vertex tagger (JVT) and the LooseBad WP for the event cleaning are applied both described in **??**. Small-R jets j are selected for $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and

$|\eta| < 4.5$ and are required to be outside of the Higgs candidate large-R jets J by imposing $\Delta R(J, j) > 1.4$. Further cuts applied on the VBF jet system optimized on significance are $|\Delta\eta(j, j)| > 3$ and $m_{jj} > 1\,\text{TeV}$.

2.3.4 Kinematic Regions

Signal Region (SR), Validation Region (VR) and Control Region (CR) are explored and optimized in previous analyses [17, 28] in the m_{H1}, m_{H2} plane and are defined as

$$SR = X_{hh} = \sqrt{\left(\frac{m_{H1} - 124\,\text{GeV}}{1500/m_{H1}}\right)^2 + \left(\frac{m_{H2} - 117\,\text{GeV}}{1900/m_{H2}}\right)^2} < 1.6, \tag{2.3.1}$$

$$VR = \sqrt{\left(\frac{m_{H1} - 124\,\text{GeV}}{0.1\ln(m_{H1})}\right)^2 + \left(\frac{m_{H2} - 117\,\text{GeV}}{0.1\ln(m_{H2})}\right)^2} < 100, \tag{2.3.2}$$

and

$$CR = \sqrt{\left(\frac{m_{H1} - 124\,\text{GeV}}{0.1\ln(m_{H1})}\right)^2 + \left(\frac{m_{H2} - 117\,\text{GeV}}{0.1\ln(m_{H2})}\right)^2} > 100\ \&\ < 170. \tag{2.3.3}$$

Figure 2.6 depicts the regions in the m_{H1}, m_{H2} plane on the SM signal sample.

2.3.5 Background Estimation

Since the final state of this analysis is hadronic it remains a challenging task to estimate the contributions from the plethora of QCD processes that contribute to backgrounds via misidentification of light quarks as heavy (b, t)-quarks. Therefore the well established ABCD method is employed to derive a data-driven background estimate [30, 31]. It is based on the idea to use two independent variables e.g. f and g to define four orthogonal regions A, B, C and D as illustrated in figure 2.7 so that for some combination of the ratio of the event yields in the regions hold

$$\frac{N_A}{N_B} = \frac{N_C}{N_D}. \tag{2.3.4}$$

By rearranging the equation for the unknown N_A an estimate for the background

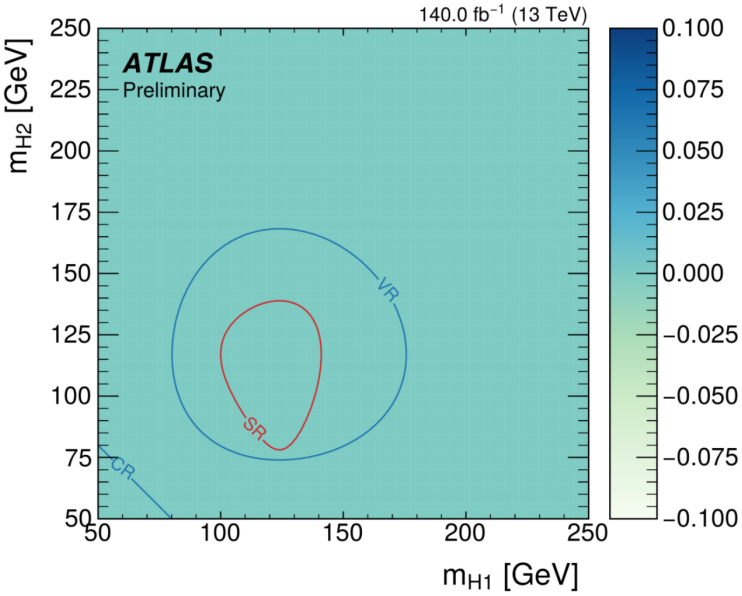


Figure 2.6: REDO

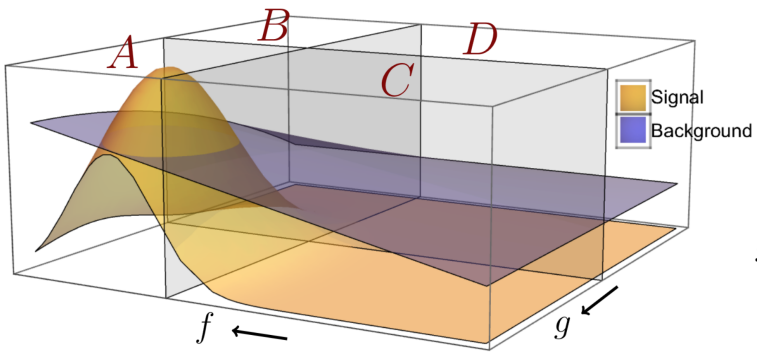


Figure 2.7: Illustration of four orthogonal regions A,B,C and D defined by two variables f and g in the horizontal and signal and background yields in the vertical dimension. Adopted from [31].

of the signal region can be derived from the other known quantities that lie in the regions dominated by the background. This approach relies on the assumption that the shape of the background in figure 2.7 does not vary greatly between C to D and A to B. Therefore the method must always be tested in a different region to determine its reliability.

In this analysis the two orthogonal variables are defined via the amount of $X \to bb$ Higgs tagged large-R jets denoted as Xbb and the kinematic regions of the SR and CR defined in 2.3.4. This gives the four orthogonal regions shown in

Table 2.2: Four orthogonal region definitions for the ABCD method

2 Xbb in CR	2 Xbb in SR
1 Xbb in CR	1 Xbb in SR

table 2.2. Hence, the background in the SR is estimated with a weight extracted from the CR

$$N_{\text{SR}}^{\text{2Xbb}} = \frac{N_{\text{CR}}^{\text{2Xbb}}}{N_{\text{CR}}^{\text{1Xbb}}} N_{\text{SR}}^{\text{1Xbb}} = w_{\text{CR}} N_{\text{SR}}^{\text{1Xbb}} = 0.0081 \times N_{\text{SR}}^{\text{1Xbb}}. \tag{2.3.5}$$

The method is validated in the variable radius (VR) within statistical uncertainties as displayed in figure 2.8. It is noted that the previous analysis also studied a binned transfer-factor without but did not see any improvement [28].

2.3.6 Event Classification

After the selection of events, a deep feed-forward neural network is employed to construct the final histogram for the statistical test. This neural network's training utilizes a novel approach NEOS, which is thoroughly discussed in Chapter **??** and optimizes on the CL_s quantity detailed in section **??**. Inputs to the neural network include 20 features, which are the four vectors $(p_{\text{T}}, \eta, \phi, m)$ of the Higgs boson pair system, the individual Higgs candidates, and the two VBF jets.

The network's architecture features three fully connected layers, each comprising 100 nodes, and concludes with a singular output node. The hidden layers are followed by a rectified linear unit activation function whereas the output node employs a sigmoid activation for classification. The architecture is thus defined as

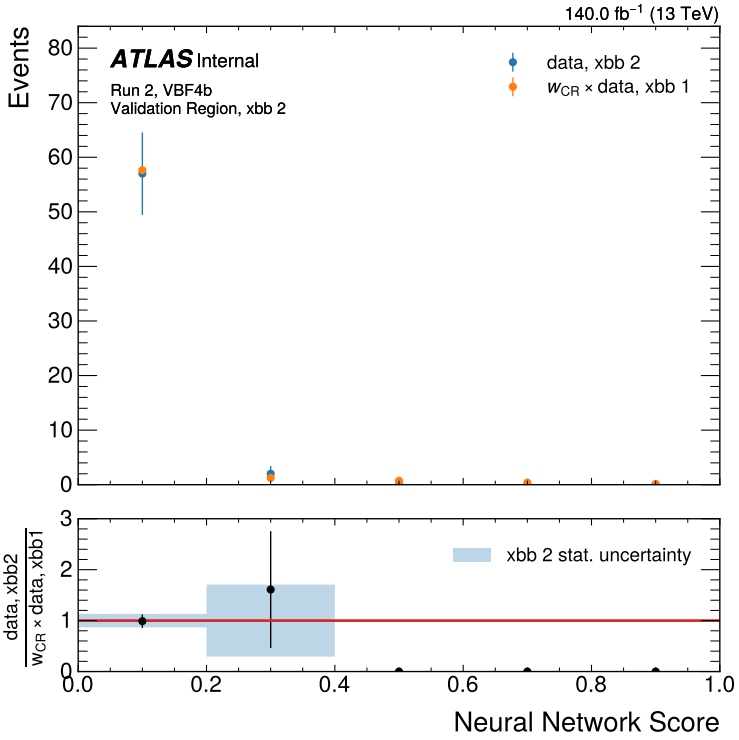


Figure 2.8: The background estimated in the VR from data with 1 xbb tag, agrees with data in the VR with 2 xbb tags within statistical uncertainties.

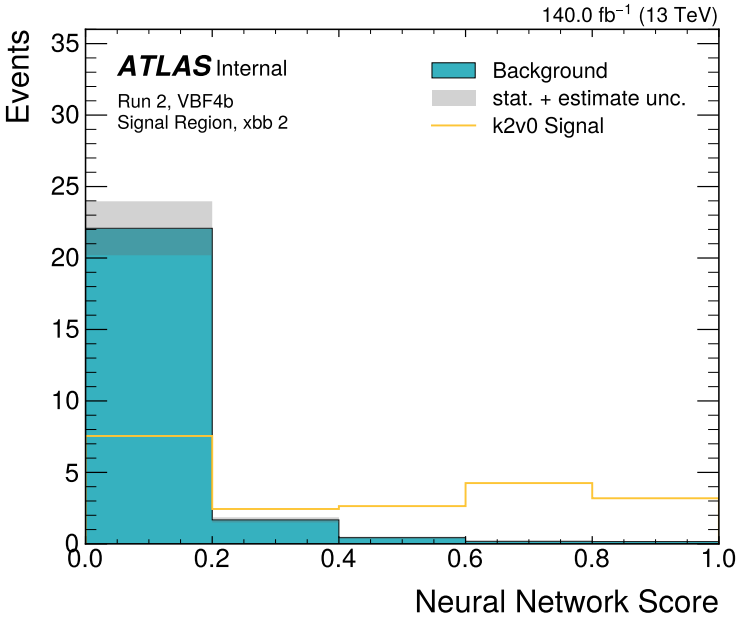


Figure 2.9: Expected histogram the data-driven background estimate and a signal hypothesis with $\kappa_{2V} = 0$ and other couplings set to their SM value. add other nominal hypotheses

[20,100,100,100,1], indicating the sequence of layers from input to output. Figure 2.9 displays the nominal expected histogram for this analysis.

Chapter 3

$HH \rightarrow 4b$ Results

unblinded plot fit plots ranking scan

how to frame it, show improvement comparison to naive training?, whole scan or k2v0 enough

Pulls allow one to estimate how well a model fits the data. A pull is a value computed for each data bin. It is given by (observed - predicted) / standard-deviation. If the model is correct, the expectation value of each pull is zero and its variance is one in the asymptotic limit of infinite samples. Under these conditions, the chi-square statistic is computed from the sum of pulls squared has a known probability distribution if the model is correct. It therefore serves as a goodness-of-fit statistic.

Appendices

Appendix A

Acronyms

CERN Organisation européenne pour la recherche nucléaire

ATLAS A Toroidal LHC Apparatus

CMS Compact Muon Solenoid

SM Standard Model

QFT Quantum Field Theory

QCD Quantum Chromodynamics

QED Quantum Electrodynamics

EW Electroweak

EWSB Electroweak Symmetry Breaking

VEV Vacuum Expectation Value

CKM Cabibbo-Kobayashi-Maskawa

EM electromagnetic

IP impact parameter of tracks

ML Machine Learning

neos neural end-to-end-optimized summary statistics

HEP High Energy Physics

LHC Large Hadron Collider

HL-LHC High Luminosity LHC

ID Inner Detector

SCT semiconductor tracker

TRT transition radiation tracker

IBL insertable *b*-layer

HLT high level trigger

L1 Level-1

PFO Particle Flow Object

TCC Track CaloCluster

UFO Unified Flow Object

JES Jet Energy Scale

JER Jet Energy Resolution

JMR Jet Mass Resolution

GGF gluon-gluon fusion

VBF vector-boson fusion

LO leading order

NLO next-to-leading order

NNLO next-to-next-to-leading order

N^{3}LO next-to-next-to-next-to-leading order

SR Signal Region

VR Validation Region

CR Control Region

KDE Kernel Density Estimation

bKDE binned Kernel Density Estimation

MC Monte Carlo

PDF Parton Density Function

PV primary vertex

JVT jet vertex tagger

NN Neural Network

ANN Artificial Neural Network

DL1 Deep Learning based heavy-flavour tagger

WP working point

VR variable radius

DIPS Deep Impact Parameter Sets

SMT Soft Muon Tagger

ROC Receiver Operator Characteristic

SHAP SHapley Additive exPlanations

MCP Muon Combined Performance

Appendix B

Cutflow

TODO, also fine like that?

Selection	Event	Fraction [%]	Total Fraction [%]
Initial	16854036422.000		
Preselections (MNT + Jet Cleaning)	670573995.000	100.000	100.000
PassTrigBoosted	63944638.000	9.536	9.536
PassTwoFatJets	57510800.000	89.938	8.576
PassTwoHbbJets	12875.000	0.0223	<0.001
PassVBFJets	5762.000	44.753	<0.001
PassFatJetPt	3902.000	67.720	<0.001
PassVBFCut	314.000	8.047	<0.001

Table B.1: Cut-flow table for data before signal region cut

Selection	Event	Fraction [%]	Total Fraction [%]
Initial	1475.226		
Preselections (MNT + Jet Cleaning)	547.960	100.000	100.000
PassTrigBoosted	20.926	3.819	3.819
PassTwoFatJets	14.141	67.576	2.581
PassTwoHbbJets	5.353	37.852	0.977
PassVBFJets	2.243	41.903	0.409
PassFatJetPt	1.408	62.793	0.257
PassVBFCut	0.148	10.539	0.027
PassSR	0.097	65.484	0.018
OverlapRemoval	0.059	61.200	0.011

Table B.2: Cut-flow table for DSID = 600463

Bibliography

[1] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett.*, B716:30–61, 2012. doi:10.1016/j.physletb.2012.08.021.

[2] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys.Lett.*, B716:1–29, 2012. doi:10.1016/j.physletb.2012.08.020.

[3] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964. doi:10.1103/PhysRevLett.13.321. URL `https://link.aps.org/doi/10.1103/PhysRevLett.13.321`.

[4] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964. doi:10.1103/PhysRevLett.13.508. URL `https://link.aps.org/doi/10.1103/PhysRevLett.13.508`.

[5] Michael E Peskin. On the trail of the higgs boson. *Annalen der Physik*, 528 (1-2):20–34, 2016. doi:10.1002/andp.201500225.

[6] Pankaj Agrawal, Debashis Saha, Ling-Xiao Xu, Jiang-Hao Yu, and C.-P. Yuan. Determining the shape of the higgs potential at future colliders. *Phys. Rev. D*, 101:075023, Apr 2020. doi:10.1103/PhysRevD.101.075023. URL `https://link.aps.org/doi/10.1103/PhysRevD.101.075023`.

[7] Dario Buttazzo, Giuseppe Degrassi, Pier Paolo Giardino, Gian F. Giudice, Filippo Sala, Alberto Salvio, and Alessandro Strumia. Investigating the near-criticality of the Higgs boson. *JHEP*, 12:089, 2013. doi:10.1007/JHEP12(2013)089.

[8] Sally Dawsona, Patrick Meadeb, Isobel Ojalvoc, and Caterina Vernierid. Report of the topical group on higgs physics for the energy frontier: The case for precision higgs physics. *arXiv preprint arXiv:2209.07510*, 2022. URL `https://arxiv.org/abs/2209.07510`.

[9] Mark Thomson. *Modern particle physics*. Cambridge University Press, 2013.

[10] Maxime Gouzevitch and Alexandra Carvalho. A review of higgs boson pair production. *Reviews in Physics*, 5:100039, 2020. ISSN 2405-4283. doi:https://doi.org/10.1016/j.revip.2020.100039. URL `https://www.sciencedirect.com/science/article/pii/S2405428320300022`.

[11] The ATLAS Collaboration. Search for pair production of higgs bosons in the $$ b\overline{b}b\overline{b} $$final state using proton-proton collisions at $$ \sqrt{s}=13 $$tev with the atlas detector. *Journal of High Energy Physics*, 2019(1):30, 2019. doi:10.1007/JHEP01(2019)030. URL `https://doi.org/10.1007/JHEP01(2019)030`.

[12] Dan Guest, Kyle Cranmer, and Daniel Whiteson. Deep learning and its application to lhc physics. *Annual Review of Nuclear and Particle Science*, 68(1):161–181, 2018. doi:10.1146/annurev-nucl-101917-021019. URL `https://doi.org/10.1146/annurev-nucl-101917-021019`.

[13] Nathan Simpson and Lukas Heinrich. neos: End-to-end-optimised summary statistics for high energy physics. *Journal of Physics: Conference Series*, 2438 (1):012105, feb 2023. doi:10.1088/1742-6596/2438/1/012105. URL `https://dx.doi.org/10.1088/1742-6596/2438/1/012105`.

[14] M. Grazzini, G. Heinrich, S. Jones, S. Kallweit, M. Kerner, J. M. Lindert, and J. Mazzitelli. Higgs boson pair production at NNLO with top quark mass effects. *Journal of High Energy Physics*, 2018(5), may 2018. doi:10.1007/jhep05(2018)059. URL https://doi.org/10.1007%2Fjhep05%282018%29059.

[15] Frédéric A. Dreyer and Alexander Karlberg. Vector-boson fusion higgs pair production at n^3LO. *Phys. Rev. D*, 98:114016, Dec 2018. doi:10.1103/PhysRevD.98.114016.

[16] Daniel de Florian, D Fontes, J Quevillon, M Schumacher, FJ Llanes-Estrada, AV Gritsan, E Vryonidou, A Signer, P de Castro Manzano, D Pagani, et al. *arXiv: Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*. Number arXiv: 1610.07922. Cern, 2016.

[17] ATLAS Collaboration. Search for nonresonant pair production of higgs bosons in the b b^- b b^- final state in p p collisions at s= 13 tev with the atlas detector. *Phys. Rev. D*, 108(5):052003, 2023. doi:10.1103/PhysRevD.108.052003.

[18] Dale Charles Abbott, William Keaton Balunas, Lucas Santiago Borgna, Alexander Emerman, James Frost, Sean Joseph Gasiorowski, James Cameron Grundy, Nicole Michelle Hartman, Shota Hayashida, Todd Brian Huffman, Cigdem Issever, Michael Kagan, Yu Nakahama, Santiago Rafael Paredes Saenz, Attilio Picazio, Jana Schaarschmidt, Todd Seiss, Mel Shochet, Beojan Stanislaus, Maximilian J Swiatlowski, Rafael Teixeira De Lima, Stephane Willocq, Anna Goussiou, Nikolaos Konstantinidis, Sau Lan Wu, Chen-Hsun Chan, Chen Zhou, Rui Zhang, Christopher Gubbels, Marta Maja Czurylo, Raif Rafideen Bin Norisam, Teng Jian Khoo, Arely Cortes-Gonzalez, Daniel Guest, Liaoshan Shi, Iza Veliscek, Marco Valente, Alessandra Betti, Christopher Don Milke, and Katharine Leney. Supporting Document: The Search for Resonant HH Production Decaying to the 4b Final State Using the Full Run-2 Data and the Boosted Analysis Channel. Technical report, CERN, Geneva, 2020. URL https://cds.cern.ch/record/2708599.

[19] Validation of signal Monte Carlo event generation in searches for Higgs boson pairs with the ATLAS detector. Technical report, CERN, Geneva, 2019. URL `https://cds.cern.ch/record/2665057`. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2019-007.

[20] ATLAS Collaboration. Luminosity determination in *pp* collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC. 2022. doi:10.48550/ARXIV.2212.09379.

[21] Johan Alwall, R Frederix, S Frixione, V Hirschi, Fabio Maltoni, Olivier Mattelaer, H-S Shao, T Stelzer, P Torrielli, and M Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, 2014(7):1–157, 2014.

[22] Francis Halzen, A Martin, and Leptons Quarks. An introductory course in modern particle physics. *John and Wiley*, 1984.

[23] M. Tanabashi et al. Review of particle physics. *Phys. Rev. D*, 98:030001, Aug 2018. doi:10.1103/PhysRevD.98.030001.

[24] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159, 2015. doi:10.1016/j.cpc.2015.01.024.

[25] Stefan Höche. Introduction to parton-shower event generators. In *Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders*, pages 235–295, 2015. doi:10.1142/9789814678766_0005.

[26] GEANT4 Collaboration, S. Agostinelli, et al. GEANT4 – a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250, 2003. doi:10.1016/S0168-9002(03)01368-8.

[27] Constraints on the Higgs boson self-coupling from the combination of single-Higgs and double-Higgs production analyses performed with the ATLAS experiment. Technical report, CERN, Geneva, 2019. URL http://cds.cern.ch/record/2693958. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2019-049.

[28] Sau Lan Wu, Ashutosh Kotwal, Arely Cortes Gonzalez, Michael Kagan, Shu Li, Maximilian J Swiatlowski, Liaoshan Shi, Janna Katharina Behr, Valentina Cairo, Thomas Andrew Schwarz, Sebastien Rettie, Yanlin Liu, Rui Zhang, Rachel Jordan Hyneman, Sanmay Ganguly, Dilia Maria Portillo Quintero, Kunlin Ran, Marco Valente, Mohamed Belfkir, Rafael Teixeira De Lima, Zhen Wang, Daariimaa Battulga, Jem Aizen Mendiola Guhit, Yuwen Ebony Zhang, Russell Bate, Karl Ver Hage Falb, Salah Nasri, Hemza Azri, and Marcus Vinicius Gonzalez Rodrigues. Search for resonant and non-resonant boosted Higgs boson pair production in *bbbb* final state via vector-boson-fusion (VBF) production using the full Run 2 data with ATLAS detector. Technical report, CERN, Geneva, 2023. URL https://cds.cern.ch/record/2848140.

[29] ATLAS Collaboration. Measurement of the ATLAS Detector Jet Mass Response using Forward Folding with $80\,\text{fb}^{-1}$ of $\sqrt{s} = 13\,\text{TeV}$ pp data. ATLAS-CONF-2020-022, 2020. URL https://cds.cern.ch/record/2724442.

[30] Will Buttinger. Background estimation with the abcd method. URL https://twiki.cern.ch/twiki/pub/Main/ABCDMethod/ABCDGuide_draft18Oct18.pdf. Last accessed: 2023-11-21.

[31] Gregor Kasieczka, Benjamin Nachman, Matthew D. Schwartz, and David Shih. Automating the abcd method with machine learning. *Phys. Rev. D*, 103:035021, Feb 2021. doi:10.1103/PhysRevD.103.035021. URL https://link.aps.org/doi/10.1103/PhysRevD.103.035021.

Statutory Declaration - Eidesstattliche Erklärung

I declare that I have authored this thesis independently, that I have not used other than the declared sources/ resources and that I have explicitly marked all materials which has been quoted either literally or by content form the used sources.

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Berlin, 07.03.2024

Frederic Renner