Investigations to improve the sensitivity to test CP-invariance in Higgs boson production via vector boson fusion exploiting the $H \to WW \to e\mu 2\nu$ decay mode

Bachelor Thesis

presented for the Degree of Bachelor of Science (B.Sc.) in Physics

Submitted by

Damian Lanzenstiel

Supervised by

Prof. Dr. Markus Schumacher

Albert-Ludwigs-Universität Freiburg

Faculty for Mathematics and Physics

February 28, 2023

Erklärung

Hiermit versichere ich, die eingereichte Bachelorarbeit selbstständig verfasst und keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt zu haben. Wörtlich oder inhaltlich verwendete Quellen wurden entsprechend den anerkannten Regeln wissenschaftlichen Arbeitens (lege artis) zitiert. Ich erkläre weiterhin, dass die vorliegende Arbeit noch nicht anderweitig als Bachelorarbeit eingereicht wurde.

Ort, Datum	Unterschrift

Abstract

This thesis presents an investigation performed in order to improve the sensitivity of the test of CP-violation in Higgs boson production via interactions of weak gauge bosons in the vector-boson fusion (VBF) mode using four types of CP-odd observables in the $H \to W^+W^- \to e\mu 2\nu$ final state. The four-momentum vector of the Higgs boson is reconstructed using fully-connected feed-forward regression neural networks trained on simulated samples representing the proton-proton collision data of the full Run 2 data-taking period (2015-2018) of the ATLAS detector at the LHC with a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ and corresponding to an integrated luminosity of $139\,\text{fb}^{-1}$. Improvements have been achieved by the neural-network-based-reconstructions of the Higgs boson compared to an approximated reconstruction (Eff.H) in which the four-momentum vectors of the Higgs boson final state particles were added together. The optimal observables $\mathcal{OO}$ calculated using the neural networks improved the $\mathcal{OO}$ resolution by 0.25% compared to the $\mathcal{OO}$ resolution calculated with the Eff.H reconstruction. Mean values of $\mathcal{OO}$s and two other CP-odd observable, $\Delta\Phi_{jj}^{\text{signed}}$ and $\mathcal{O}_{\text{Reg}}$ are used in a gauge curve fit to estimate the sensitivity of $\tilde{d}$, which parameterises the strength of the CP-violation. The best performing $\mathcal{OO}$ resulted in an expected of $\tilde{d} \in [-5.089 \times 10^{-5}, 5.089 \times 10^{-5}]$ at a 68% confidence level (CL), outperforming $\mathcal{O}_{\text{Reg}}$ and $\mathcal{OO}_{(\text{Eff.H})}$ by 3.80% and 5.71%, respectively. $\Delta\Phi_{jj}^{\text{signed}}$ variable resulted in the weakest constraints with $\tilde{d} \in [-7.82 \times 10^{-5}, 7.82 \times 10^{-5}]$ at an 68% CL.

Zusammenfassung

Diese Arbeit präsentiert eine Investigation zur Verbesserung des Tests von CP verletzenden Wechselwirkungen des Higgs Bosons mit schwachen Eichbosonen im Higgs Boson Produktionsmodus Vektor-Boson Fusion (VBF) mit dem $H \to W^+W^- \to e\mu 2\nu$ Endzustand, bei der vier verschiedene Typen von CP-ungerade Observablen verwendet werden. Die Analyse benutzt "fully-connected feed-forward" Neurale Netze zur Rekonstruktion des Viererimpulses des Higgs Bosons. Die Netzwerke werden mit simulierten Ereignissen trainiert die, dem kompletten Run 2 (2015-2018) des ATLAS Detektors am LHC bei einer Schwerpunktsenergie von $\sqrt{s} = 13\,\text{GeV}$ und einer integrierten Luminosität von 139fb^{-1} entsprechen. Verbesserungen beim Vergleich der durch neuronale Netzwerke rekonstruierten Viererimpules des Higgs Bosons mit einem genäherten Higgs boson (Eff.H), welches durch das vektorielle aufaddieren der Viererimpulse der Teilchen des Higgs Boson Endzustandes berechnet wurden, konnten erzielt werden. Die Optimale Observable ($\mathcal{OO}$) die durch neuronale Netzwerke bestimmt wurde zeigte eine Verbesserung in der Auflösung um 0.25% im Vergleich zu der $\mathcal{OO}$ welche mit Eff.H bestimmt wurde. Die Mittelwerte der $\mathcal{OO}$ werden zusammen mit denen von zwei weiteren CP-ungeraden Observablen $\Delta\Phi_{jj}^{\text{signed}}$ und $\mathcal{O}_{\text{Reg}}$ bei einem Eichkurven Fit genutzt, um die Sensitivität von $\tilde{d}$ zu bestimmen, wobei $\tilde{d}$ die Stärke der CP Verletzung parametrisiert. Die beste $\mathcal{OO}$, resultierte in einem erwarteten Konfidenzintervall von $\tilde{d} \in [-5.089 \times 10^{-5}, 5.089 \times 10^{-5}]$ bei einem Konfidenzlevel von 68%. Damit erreicht es eine Verbesserung bezüglich $\mathcal{O}_{\text{Reg}}$ von 3.80%

und 5.71% bezogen auf $\mathcal{OO}_{(\text{Eff.H})}$. $\Delta\Phi_{jj}^{\text{signed}}$ zeigte die schwächste Beschränkung von $\tilde{d} \in [-7.82 \times 10^{-5}, 7.82 \times 10^{-5}]$ bei einem Konfidenzlevel von 68%.

Contents

Introduction

One goal of physics is to expand our understanding of what holds our world together in its innermost parts. In particle physics, the best-prevailing model is the so-called *Standard Model* (SM) of particle physics. The SM describes the fundamental particles and their interactions which can be split into two groups. Those groups are the odd-integer spin *fermions* and integer-spin *bosons*. The latter act mostly as carriers for the fundamental forces described by the SM [1–3].

Early versions of the SM did not account for massive bosons. However, experiments observed non-zero particle masses. Examples of these are the $W^{\pm}$ bosons mass of $m_{W^{\pm}} = 80.4\,\text{GeV}$ and the Z^0 boson with $m_{Z^0} = 91.2\,\text{GeV}$ [18]. To solve this and include mass terms into the SM, the *Brout-Englert-Higgs* (BEH) mechanism [12–14] was introduced in 1964. The mechanism introduces a scalar field which causes spontaneous symmetry breaking upon interaction for massive particles. This field consequently introduces another massive particle as an excitation of the field called the Higgs boson.

The Higgs Boson was discovered by the ALTAS, and CMS experiments at the *Large Hadron Collider* (LHC) in 2012 [4, 5].

While the SM is a good description of observed data, there are still phenomena unexplained by this theory, like the observed *baryon asymmetry* of the universe [10], which describes the observation that more baryonic matter than anti-baryonic matter is observed. A process called baryogenesis, which produces baryonic and anti-baryonic matter at different rates, is introduced to solve this. The process needs to fulfil three conditions called *Sakharov conditions* [24]. One is the so-called CP-violation described in the SM by the *Cabibbo-Kobayashi-Maskawa* (CKM) matrix. However, the predicted amount of CP-violating processes is insufficient to explain the observations fully. Therefore it is instrumental to look for additional sources of CP-violating processes.

One possible source of CP-violation is presented by the HVV vertex in which a Higgs boson couples to two vector bosons. The *vector boson fusion* (VBF) production mode of the Higgs boson exploiting $H \to \tau\tau$ channel was investigated by the ATLAS, and CSM collaborations in previous analyses [6, 26]. Both studies used a CP-odd observable called *optimal observable* ($\mathcal{OO}$), which showed the best sensitivity to CP-violating processes. No significant deviation from SM predictions was measured.

This thesis focuses on VBF-produced Higgs bosons in the $H \to W^-W^+ \to e\mu 2\nu$ decay channel. The simulated samples correspond to data recorded at a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ with an integrated luminosity of 139fb^{-1} by the ATLAS detector.
Since the $\mathcal{OO}$ takes the four-momentum vector of the Higgs boson and the two tagging jets as inputs, these have to be reconstructed and optimised. For the analysis, neural networks are used to reconstruct the Higgs boson four-momentum vector and

improve the resolution of its components. The neural networks are optimised using a bayesian hyperparameter optimisation called *OPTUNA* [55]. The reconstructed Higgs boson four-momentum vectors are then used to calculate a $\mathcal{OO}$ distribution for each reconstruction. The $\mathcal{OO}$s and two other CP-odd observables $\mathcal{O}_{\text{Reg}}$ and $\Delta\Phi_{jj}^{\text{signed}}$ are used to estimate the sensitivity to constrain measurements of $\tilde{d}$.

This thesis is structured in the following way: In chapter 1, the theoretical background is given. This includes a description of the SM followed by the properties of the Higgs boson at the LHC and a discussion of CP-violation. In chapter 2, the ATLAS experiment at the LHC is introduced. Chapter 3 gives an overview of the machine learning methods used to reconstruct the Higgs boson four-momentum vector. Chapter 4 introduces the samples and explains how the events used in the analysis are selected. Chapter 5 details the reconstruction methods exploiting NN used to predict the four-momentum vector of the Higgs boson. It is followed by chapter 6, which estimates expected constraints on $\tilde{d}$. A conclusion is given in chapter 7.

1 Theoretical Background

This chapter gives a summary of the theoretical background used for this analysis. The first section presents an overview of the Standard Model (SM) of particle physics. Section 1.2 describes the properties of the Higgs boson regarding proton-proton (pp) collisions. The last section describes CP symmetries and how CP-violating processes can be studied.

1.1 The Standart Model of Particle Physics

The SM is based on gauge-invariant renormalizable *quantum field theories* (QFT). Particles arising as excitations of quantum fields and the kinematics are described by the *Lagrangien formalism*. The SM had massive success over the years in predicting the existence of several particles, such as the top-quark [11] or the Higgs boson [4, 5], that were later discovered. Still, it leaves many observations unexplained. Some examples are the baryon asymmetry of the universe [10], dark matter [7], dark energy [8] and neutrino masses [9].

1.1.1 Fermions

The Standard Model contains 12 fermions as shown in Table 1. Each has an antiparticle with the same mass but opposite additive quantum numbers. Fermions have half-integer spin values and obey the Fermi-Dirac statistics. Fermions are split into two categories: *quarks* and *leptons*.
Leptons consist of *electrons* (e), *muons* (μ) and *taus* leptons (τ) with a negative electromagnetic charge. Each has a corresponding neutrino (ν_e), (ν_μ) and (ν_τ) which are el.mag. neutrally charged.
Quarks have fractional el.mag. charges. There are *up* (u), *charm* (c), *top* (t) with an el.mag. charge of $\frac{2}{3}$ and *down* (d), *strange* (s) and *bottom* (b) with a charge of $-\frac{1}{3}$. Quarks also have another property called "colour" or "colour charge". Particles with a colour are subject to the strong force. Quarks appear only in colour-neutral compositions called *hadrons*. They split into two groups. The first group *mesons* are composed of a quark, anti-quark pair. The other group is called *baryons* and is composed of three quarks. Two typical examples are neutrons and protons.

	Generation		Fermion	Q [e]	Mass [MeV]
Leptons	1^{st}	e	Electron	-1	≈ 0.511
		ν_e	Electron neutrino	0	$< 460 \times 10^{-6}$
	2^{nd}	μ	Muon	-1	≈ 105.658
		ν_μ	Muon neutrino	0	< 0.19
	3^{rd}	τ	$\tau-$lepton	-1	≈ 1776.86
		ν_τ	$\tau-$lepton neutrino	0	< 18.2
Quarks	1^{st}	u	Up	$\frac{2}{3}$	≈ 2.16
		d	Down	$-\frac{1}{3}$	≈ 4.67
	2^{nd}	c	Charm	$\frac{2}{3}$	≈ 93.4
		s	Strange	$-\frac{1}{3}$	$\approx 1.27 \times 10^3$
	3^{rd}	t	Top	$\frac{2}{3}$	$\approx 4.18 \times 10^3$
		b	Bottom	$-\frac{1}{3}$	$\approx 172.69 \times 10^3$

Table 1: List of SM fermions with electrical charge and approximate mass with values taken from [18].

1.1.2 Bosons

Bosons are particles with integer spin, following the Bose-Einstein statistics. Bosons are classified by their spin values. All SM bosons except the Higgs boson have a spin of 1 and are called *gauge bosons*. In Table 2, all gauge bosons are listed. Gauge bosons are the mediators of fundamental forces and couple to different charges. *Gluons* (g) are the mediators of the strong force and couple to the colour charge. Similarly, photons (γ) couple to el.mag. charge Q and $W^{\pm}/Z^0$ bosons to the weak isospin I_W and weak hypercharge Y.
The Higgs boson is a scalar boson with a spin of 0. Its corresponding field is needed to explain the emergence of massive particles in the SM. Further details will be described in section 2.1.4.

Gauge boson	Force	Q [e]	Mass [GeV]
Photon (γ)	Electromagnetic	0	0
$W^{\pm}$	Weak	± 1	80.377 ± 0.012
Z^0	Weak	0	91.1876 ± 0.0021
Gluon (g)	Strong	0	0

Table 2: List of the gauge bosons in the SM together with the fundamental interaction they represent, their electric charge and mass with values taken from [18].

1.1.3 Fundamental Interactions

Two theories describe the fundamental interactions in SM. The strong interaction is described by *quantum chromodynamics* (QCD) with symmetry group $SU(3)_C$. It couples to particles carrying a colour charge. Quarks have red, green and blue colours with anti-quarks possessing the corresponding anticolour. Gluons carry eight different colour combinations described by the colour octet. One aspect of colour charges is *confinement*, meaning particles only appear in colour-neutral bound states such as baryons and mesons.
The second theory is the *electroweak* interaction as a unification of *quantum electrodynamics* (QED) describing electromagnetism and the theory describing the weak interaction called *quantum flavourdynamics* (QFD). With the symmetry group $SU(2)_{I_W} \times U(1)_{Y_W}$ the electroweak interaction introduces the weak isospin I_W and hypercharge Y_W as two new charges. They are related to the el.mag. charge by the Gell-Mann–Nishijima formula

$$Q = I_{w_3} + \frac{Y}{2}. \tag{1}$$

Due to the requirement of local gauge invariance, all gauge bosons have to be massless. Experiments show this is not true since $W^{\pm}$ and Z^0 Bosons have a mass as given in Table 2. The mass terms of the gauge bosons $W^{\pm}$ and Z^0 emerge by spontaneous symmetry breaking of the $SU(2)_{I_W} \times U(1)_{Y_W}$ symmetry to $U(1)_Q$ [2] and the Higgs Mechanism.

1.1.4 Brout-Englert-Higgs-Mechanism

The Brout-Englert-Higgs-Mechanism (BEH) or short Higgs-Mechanism, was proposed by F. Englert, R . Brout, P. W. Higgs, G. Guralnik, C. R. Hagen and T. W. B. Kibble [12–14] in 1964. To include mass terms for gauge bosons into the SM, one introduces a complex scalar field ϕ with hypercharge $Y = 1$ and weak isospin $I_W = \frac{1}{2}$. ϕ^+ corresponds to an el.mag. positive charged and ϕ^0 el.mag. neutral charged field.

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{2}$$

The corresponding Lagrangian density of the field can be written as [14]:

$$\mathcal{L}_{\text{Higgs}} = (D_\mu \phi)^\dagger (D_\mu \phi) - \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \tag{3}$$

with the covariant derivatives D_μ. The last two terms describe the Higgs potential:

$$V(\phi) = -\mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \quad \text{with} \quad \lambda > 0 \tag{4}$$

here μ is the mass parameter while λ describes the selfcoupling of the Higgs field. By choosing $\mu^2 < 0$, one gets the minimum of the potential with $\phi_0 = 0$. For $\mu^2 > 0$, the ground state of $V(\phi)$ is degenerate. The ground state can be written as:

$$\phi_0 = \sqrt{\frac{\mu^2}{2\lambda}} = \frac{v^2}{2} \quad \text{with} \quad v = \sqrt{\frac{\mu^2}{\lambda}} \tag{5}$$

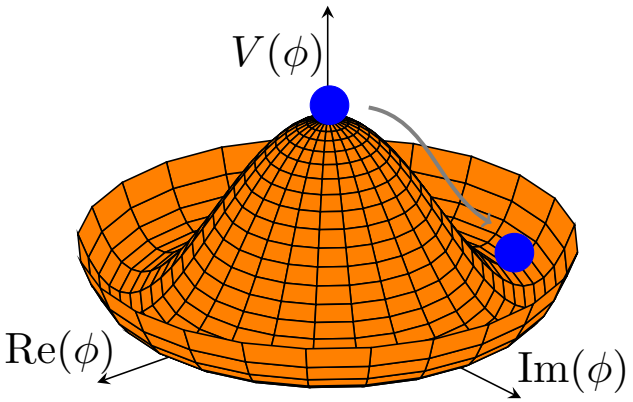


Figure 1: Shape of the Higgs potential $V(\phi)$ with $\mu^2 < 0$. The blue ball rolling down symbolizes breaking symmetry upon choosing a ground state.

where $v \approx 246\,\mathrm{GeV}$ [18] is the vacuum expectation value (vev) of the Higgs field.
The potential with two degrees of freedom is shown in Figure 1.
Choosing one particular ground state with $\phi_1 = \phi_2 = \phi_4 = 0$ leads to

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \tag{6}$$

where the symmetry of the potential is lost. This breaking of the $SU(2)_{I_W} \times U(1)_Y$ symmetry preserves the $U(1)_Q$ symmetry and therefore fulfils the requirement of a massless photon.
The field

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \tag{7}$$

is acquired by parameterization around the vev and imposing unitary gauge to eliminate the resulting massless *Goldstone* bosons [16, 17]. $H(x)$ is a fluctuation around the ground state, which results in the scalar boson called Higgs boson.
The mass terms resulting from the breaking of symmetry are:

$$m_{W^\pm} = \frac{vg}{2}, \tag{8}$$

$$m_{z^0} = \frac{m_{W^\pm}}{cos(\theta_w)}, \tag{9}$$

$$m_H = \sqrt{2\lambda v^2}. \tag{10}$$

Here g (g') is the coupling constant of the weak isospin (hypercharge). The angle $\theta_w = \arctan\left(\frac{g'}{g}\right)$ is the *weak mixing angle* angle [15]. The mass m_H cannot be derived directly from Equation 53 and has to be determined experimentally.

1.2 The Higgs Boson

The Higgs boson, which the Higgs mechanism predicted, is an electrically neutral particle with a spin of 0. It was discovered in 2012 at CERN[1] by the ATLAS and CMS experiments using the data from pp-collisions provided by the LHC at the centre-of-mass energies of 7 TeV and 8 TeV [4, 5]. The mass is measured as $m_H = 125.25 \pm 0.17\,\text{GeV}$ [18]. This section assumes pp-collisions with a Higgs boson mass of $m_H = 125\,\text{GeV}$ and a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$.

Production

The four dominant production modes of Higgs bosons in pp-collision are: gluon-gluon fusion (ggF), vector-boson fusion (VBF), associated production with a pair of top quarks ($t\bar{t}H$) and Higgs-Strahlung (VH). One example of leading-order Feynman diagrams for each of these processes is shown in Figure 2.

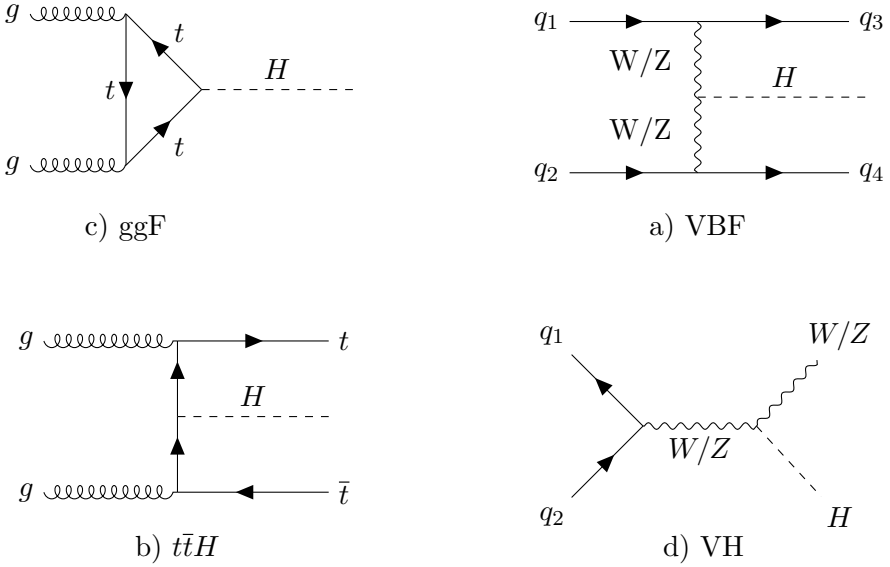


Figure 2: Leading-order Feynman diagrams of the main Higgs boson production modes at the LHC with $\sqrt{s} = 13\,\text{TeV}$: a)VBF vector-boson fusion, b)$t\bar{t}H$ associated production with a top-anti-top quark pair, c)ggF gluon-gluon fusion, d)VH Higgs-Strahlung.

The left side of Figure 3 shows cross-sections for different Higgs boson production modes. The production cross-section of the Higgs boson is highest for ggF. In this process, the Higgs boson couples to the gluons via a loop of heavy quarks. It has a production cross-section of 48.58 pb [19]. The second highest production cross-section is that of the VBF production mode, which this thesis focuses on, with 3.8 pb [19]. In this process, the Higgs boson is produced by the fusion of two W or Z bosons. VBF production is followed by VH and $t\bar{t}H$ processes with cross-sections of 2.3 pb and 0.51 pb respectively [19]. In the HV process, a W or Z boson is created from the annihilation of a quark anti-quark pair, which then radiates a Higgs boson. In the $t\bar{t}H$ process, the Higgs boson is created in association with a pair of top quarks.

[1] Conseil européen pour la recherche nucléaire

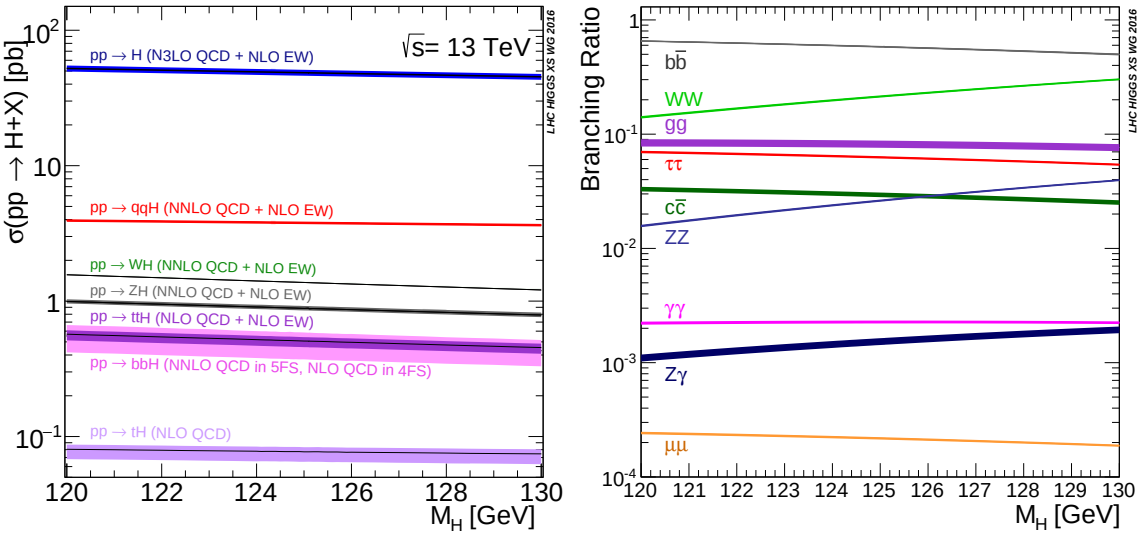


Figure 3: *Left:* Cross-section of Higgs boson production processes in proton-proton collisions at a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ as a function of the Higgs boson mass; *Right:* Branching ratios for different Higgs boson decay modes as a function of its mass [19].

Decay

The Higgs boson can decay leptonically or hadronically. The branching ratio (BR) is defined as the ratio of the partial decay width to the total decay width. The branching ratio of the most common decay processes as a function of the Higgs boson mass is shown on the right side of Figure 3. A Higgs boson decaying into two bottom quarks has the highest BR of $58.24 \pm 0.65\%$ [19]. The Higgs boson decay into two W bosons has the second highest BR of $21.37 \pm 0.99\%$ [19]. This is followed by the decay into two gluons ($H \to gg$) and two tau leptons ($H \to \tau\tau$) with $8.19^{+3.40\%}_{-3.41\%}$ [19] and $6.27^{+1.17\%}_{-1.16\%}$ [19], respectively. This thesis exploits the decay into two W bosons which can decay leptonically or hadronically. The decay into an electron has a BR of $10.71 \pm 0.16\%$ [18] while the decay into a muon and a tau lepton has a BR of $10.63 \pm 0.15\%$ and $11.38 \pm 0.21\%$ [18], respectively. The hadronic decay of the W boson has a total BR of $67.41 \pm 0.27\%$ [18].

Signal Process

The signal process of this analysis is the VBF process. Two partons radiate heavy bosons which interact and produce a Higgs boson. This boson decays into a positive and a negatively charged W boson. Due two the high mass of the W bosons, at least one has to be off shell. Both W bosons decay fully leptonically, exactly one into an (e, ν_e) pair and the other into a (μ, ν_μ) pair. The final state partons are slightly deflected by the radiation. They are then detected as jets in different hemispheres of the detector. The leading-order Feynman diagram is shown in Figure 4.

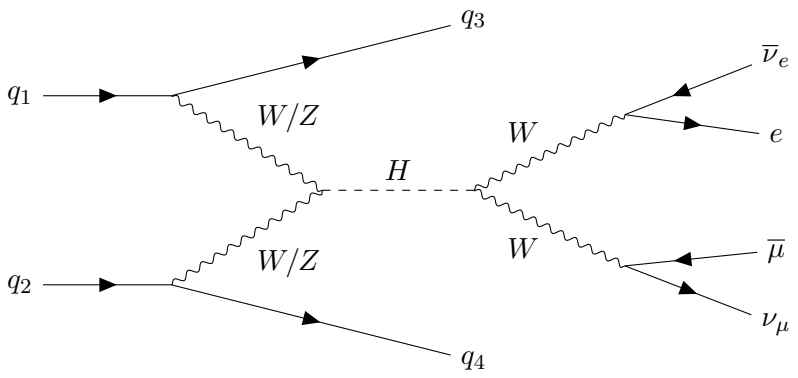


Figure 4: A leading-order Feynman diagram of the signal process for this analysis.

This process has a total BR of 0.506%. For the cross-section of 3.8 pd in the VBF process, one can expect 2661 signal process events at an integrated Luminosity of 139fb^{-1}.

1.3 CP Violation

CP-violating processes are processes in which CP symmetry is not preserved. While the SM predicts CP-violating processes, their amount is insufficient to explain the baryon asymmetry, which describes the observed abundance of baryonic matter over the anti-baryonic matter in the observable universe [10, 24]. In 1967 Sakharov proposed three conditions [24] for baryon-generating interactions to produce matter and antimatter at different rates. These Sakharov conditions are: violation of baryon number conversation, interactions out of thermal equilibrium and C and P violation. The insufficiency of the anti-baryonic matter is the motivation to look for new CP-violating processes beyond the SM. One such interaction of interest is the HVV vertex, in which the Higgs boson couples to two weak gauge bosons. To understand CP-violation, section 1.3.1 describes the C and P symmetries and how CP-violating processes are included in the SM. Section 1.3.2 and 1.3.3 derives the optimal observable and how CP-odd observables can be used to explore CP-violation.

1.3.1 CP Symmetry

Symmetry is a property of a system to remain unchanged under certain transformations. Symmetries play an important role in physics since they are associated with conservation laws as described by Noether's theorem [20].
The CP symmetry combines two distinct and discrete symmetries called charge conjugation and parity symmetry.

The parity transformation is a transformation that changes the sign of spatial coordinates. The parity operator $\mathcal{P}$ is a hermitian operator with eigenvalues ± 1. Its eigenfunctions are called even-parity for an eigenvalue of 1. Otherwise, they are odd-parity.

The charge conjugation symmetry describes the interchanging of a particle by its

antiparticle. That means reversing the additive quantum numbers of a particle but leaving its mass, spin, momentum and energy unchanged. The charge conjugation operator $\mathcal{C}$ has eigenvalues of ± 1. As such, using it two times on a particle leaves the particle unchanged. Only particles that are neutral in their additive quantum numbers are their own antiparticles, such as the photon, the neutral pion π^0 or the Higgs boson.

Parity and charge conjugation are both conserved in all QCD and QED vertices. In the weak interaction, this is not the case. The Wu-experiment showed in 1956 [21] that the conservation of parity is violated for the case of $\beta-$decays of cobalt. J. H. Christenson, J. W. Cronin, V. L. Fitch and R. Turlay discovered CP-violation in 1964 for the decay of neutral kaons [22]. In the SM CP-violation can be described by a single phase in the Cabibbo-Kobayashi-Maskawa-Matrix (CKM-Matrix) [23].

1.3.2 Effective Field Theories

EFTs give a way to describe CP-violating HVV interactions. For this, the SM Lagrangian is augmented by CP-odd operators $\mathcal{O}_{\tilde{B}B}$, $\mathcal{O}_{\tilde{W}W}$ and $\mathcal{O}_{\tilde{B}}$ of mass dimension six to create an effective Lagrangian defined as:

$$\mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{SM}} + \frac{f_{\tilde{B}B}}{\Lambda^2}\mathcal{O}_{\tilde{B}B} + \frac{f_{\tilde{W}W}}{\Lambda^2}\mathcal{O}_{\tilde{W}W} + \frac{f_{\tilde{B}}}{\Lambda^2}\mathcal{O}_{\tilde{B}} \tag{11}$$

The coefficients $f_{\tilde{B}B}$, $f_{\tilde{W}W}$ and $f_{\tilde{B}}$ are dimensionless *Wilson coefficient* and Λ is the energy scale of new physics [25]. The CP-odd operators are constructed from the Higgs doublet Φ and the electroweak gauge fields B^μ and $W^{a,\mu}$ and are given as [25–27]:

$$\mathcal{O}_{\tilde{B}B} = \Phi^\dagger \hat{\tilde{B}}_{\mu\nu}\hat{B}^{\mu\nu}\Phi \tag{12}$$

$$\mathcal{O}_{\tilde{W}W} = \Phi^\dagger \hat{\tilde{W}}_{\mu\nu}\hat{W}^{\mu\nu}\Phi \tag{13}$$

$$\mathcal{O}_{\tilde{B}} = (D_\mu\Phi)^\dagger \hat{\tilde{B}}\mu\nu D_\nu\Phi \tag{14}$$

The last operator $\mathcal{O}_{\tilde{B}}$ contributes to triple gauge boson interactions. This CP-violating coupling is constrained by the LEP experiments [28–30] and thus will not be considered. The effective Lagrangian density in the mass eigenstate bases of the Higgs boson, photon (A), $W^\pm$ and Z^0 bosons after the electroweak symmetry breaking can then be expressed as Ref. [26, 27]:

$$\mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{SM}} + \tilde{g}_{HAA}H\tilde{A}_{\mu\nu}A^{\mu\nu} + \tilde{g}_{HAZ}H\tilde{A}_{\mu\nu}Z^{\mu\nu} + \tilde{g}_{HZZ}H\tilde{Z}_{\mu\nu}Z^{\mu\nu} + \tilde{g}_{HWW}H\tilde{W}^+_{\mu\nu}W^{-\mu\nu} \tag{15}$$

The coupling parameter multiplying the CP-odd operators is given by the coefficients $\tilde{g}_{HVV}$ with $V \in \{H, A, W^\pm, Z\}$. The coupling strengths $\tilde{g}_{HVV}$ can be expressed in terms of two dimensionless couplings $\tilde{d}$ and $\tilde{d}_B$

$$\tilde{d} = -\frac{m_W^2}{\Lambda^2}f_{\tilde{W}W} \tag{16}$$

and

$$\tilde{d}_B = -\frac{m_W^2}{\Lambda^2}\tan^2(\theta_w)f_{\tilde{B}B}. \tag{17}$$

The contributions from $H\gamma\gamma$, $H\gamma Z$, HZZ and HWW to the VBF process cannot be experimentally distinguished. This allows for the arbitrary choice of $\tilde{d} = \tilde{d}_B$ [26, 27]. After that, the $\tilde{g}_{HVV}$ coefficients become:

$$\tilde{g}_{HAA} = \tilde{g}_{HZZ} = \frac{1}{2}\tilde{g}_{HWW} = \frac{g}{2m_W}\tilde{d}, \tag{18}$$

$$\tilde{g}_{HAZ} = 0. \tag{19}$$

With this, the $\tilde{d}$ parameter alone quantities the amount of CP-odd contributions to HVV couplings. This results in the matrix element $\mathcal{M}$

$$\mathcal{M} = \mathcal{M}_{\text{SM}} + \tilde{d} \cdot \mathcal{M}_{\text{CP-odd}}. \tag{20}$$

The SM part is CP-even and the CP-odd part is linear in $\tilde{d}$. The squared matrix element is then given by

$$|\mathcal{M}|^2 = \underbrace{|\mathcal{M}_{SM}|^2}_{\text{CP-even}} + \underbrace{2\tilde{d} \cdot \Re(\mathcal{M}^*_{\text{SM}}\mathcal{M}_{\text{CP-odd}})}_{\text{CP-odd}} + \underbrace{\tilde{d}^2 \cdot |\mathcal{M}_{\text{CP-odd}}|}_{\text{CP-even}}. \tag{21}$$

The first and last terms in Equation 21 are both CP-even and thus cannot be a source of CP-violation. The interference term is CP-odd, which constitutes a possible source of CP-violation. With the quadratic dependence of the third term, changes in event yields can originate from CP-even operators. As such, increased event yields are not a reliable way to test CP-symmetry. Instead, one has to exploit differences in the shape of suitable distributions to search for CP-violation.

1.3.3 CP-odd Observables

A CP-odd observable $\mathcal{O}_{\text{CP-odd}}$ is one which changes its sign under a $\mathcal{CP}$ transformation.

$$\mathcal{CP}\mathcal{O}_{\text{CP-odd}} = -\mathcal{O}_{\text{CP-odd}}. \tag{22}$$

For a CP-symmetric theory, the expectation value of a CP-odd observable vanishes. As such, it can be used to test for CP symmetry since every deviation from a vanishing expectation value indicates a violation of CP invariance.

The Optimal Observable

The Optimal Observable ($\mathcal{OO}$) [31–33] is defined as the ratio of the interference and pure SM terms of Equation 21:

$$\mathcal{OO} = \frac{2\Re(\mathcal{M}^*_{\text{SM}}\mathcal{M}_{\text{CP-odd}})}{|\mathcal{M}_{\text{SM}}|^2}. \tag{23}$$

The Optimal observable uses the complete phase-space information of the VBF Higgs boson production mode, which consists of the Higgs boson and two quarks and condenses this multidimensional phase-space into a one-dimensional variable. It provides

the best sensitivity of CP-violation for small deviations from the SM [82]. The mean of the $\mathcal{OO}$ can be calculated by integrating over the cross-section

$$\langle \mathcal{OO} \rangle = \frac{\int (\mathcal{OO} d\sigma_{\text{SM}} + \mathcal{OO}\tilde{d}\, d\sigma_{\text{CP-odd}} + \mathcal{OO}\tilde{d}^2\, d\sigma_{\text{CP-even}})}{\int (d\sigma_{\text{SM}} + \tilde{d}\, d\sigma_{\text{CP-odd}} + \tilde{d}^2\, d\sigma_{\text{CP-even}})}. \tag{24}$$

The expectation value for CP-odd observable for CP-even cross-section terms and the integral over the CP-odd cross-section both vanish. As such, Equation 24 reduces to

$$\langle \mathcal{OO} \rangle = \frac{\tilde{d} \int \mathcal{OO}\, d\sigma_{\text{CP-odd}}}{\int (d\sigma_{\text{SM}} + \tilde{d}^2\, d\sigma_{\text{CP-even}})}. \tag{25}$$

Equation 25 shows a linear dependence of the expectation value on $\tilde{d}$ for small values of $\tilde{d}$. For larger values of $\tilde{d}$, the term quadratic in $\tilde{d}$ dominates and the expectation value is scaled by $1/\tilde{d}$. Figure 5 shows the $\mathcal{OO}$ distribution for different $\tilde{d}$ values. While the distribution is symmetric for $\tilde{d} = 0$, it becomes asymmetric for $\tilde{d} \neq 0$.

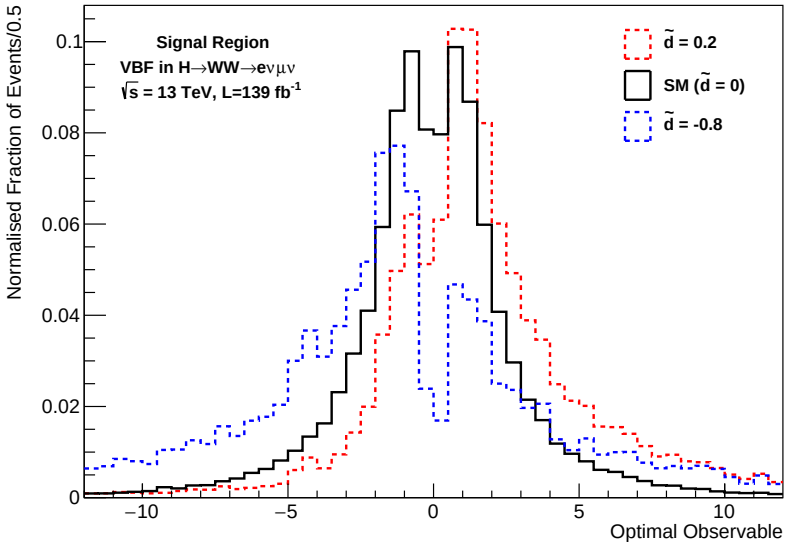


Figure 5: Distribution of $\mathcal{OO}$ for signal region events defined in chapter 4. It shows the optimal observable for three different $\tilde{d}$ values. In black, the SM case while the cases $\tilde{d} = 0.2$ and $\tilde{d} = -0.8$ are drawn in red and blue, respectively. The distributions are normalized to $139\,\text{fb}^{-1}$. Overflow and underflow bins are included.

Other CP-odd variables

Two other CP-odd variables considered to test for CP-violation are $\Delta\Phi_{jj}^{\text{signed}}$ [25] and $\mathcal{O}_{\text{Reg}}$ [34]. The variable $\Delta\Phi_{jj}^{\text{signed}}$ describes the signed difference between the azimuthal angles Φ of the VBF vertex tagging jets. It is defined as

$$\Delta\Phi_{jj}^{\text{signed}} = \phi_{j_1} - \phi_{j_2}, \tag{26}$$

with the jets being sorted by their pseudorapidity such that $\eta_{j_1} > \eta_{j_2}$.
$\mathcal{O}_{\text{Reg}}$ is constructed using neural networks with the *symbolic regression* [34] method to derive a formula directly from matrix-element information. For VBF production, this formula is given by:

$$\mathcal{O}_{Reg} = p_{\text{T}}^{j_1} p_{\text{T}}^{j_2} \sin(\Delta\Phi_{jj}^{\text{signed}}) \tag{27}$$

Here $p_T^{j_1}$ ($p_T^{j_2}$) is the transverse momentum of the leading (subleading) jet of the VBF final state. Distribution for different values of $\tilde{d}$ for $\mathcal{O}_{Reg}$ and $\Delta\Phi_{jj}^{\text{signed}}$ are shown in Figures 7 and 6 respectively. Both show a symmetric distribution for the SM case of $\tilde{d} = 0$ and become asymmetric for $\tilde{d} \neq 0$.

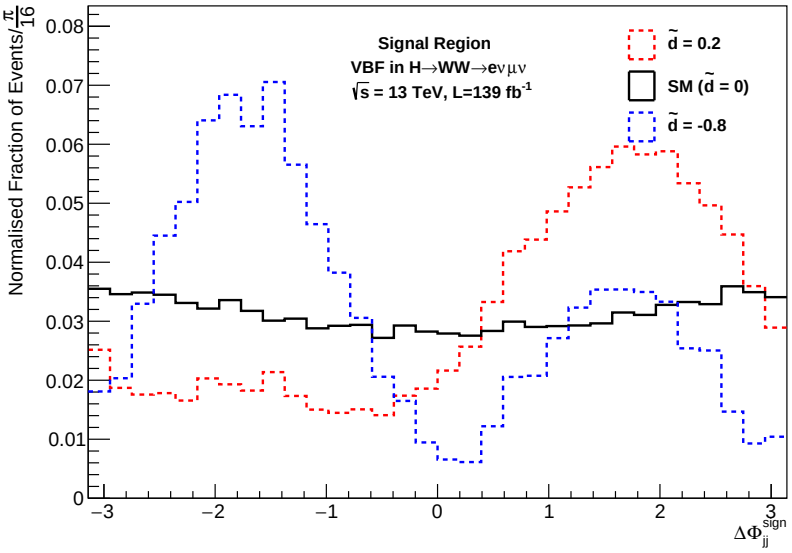


Figure 6: Distribution of $\Delta\Phi_{jj}^{\text{signed}}$ for signal region events defined in chapter 4. It shows $\Delta\Phi_{jj}^{\text{signed}}$ for three different values of $\tilde{d}$. In black the SM case while the cases $\tilde{d} = 0.2$ and $\tilde{d} = -0.8$ are drawn in red and blue, respectively. The distributions are normalized to $139\,\text{fb}^{-1}$. Overflow and underflow bins are included.

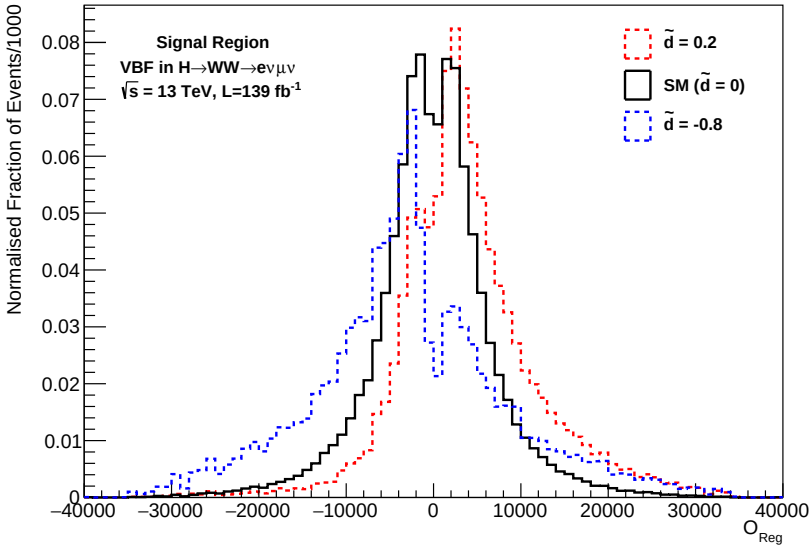


Figure 7: Distribution of $\mathcal{O}_{\text{Reg}}$ for signal region events defined in chapter 4. It shows $\mathcal{O}_{\text{Reg}}$ for three different values of $\tilde{d}$. In black the SM case while the cases $\tilde{d} = 0.2$ and $\tilde{d} = -0.8$ are drawn in red and blue, respectively. The distributions are normalized to $139\,\text{fb}^{-1}$. Overflow and underflow bins are included.

2 The ATLAS Experiment

The ATLAS detector is one of the four large experiments at the Large Hadron Collider (LHC) at CERN[2] in Geneva. This chapter gives a brief overview of the LHC and, in more detail, describes the ATLAS Detector and its components.

2.1 The Large Hadron Collider

The LHC is a circular particle accelerator focusing on *pp*-collisions. It is located at CERN, close to Geneva. The particles are accelerated in a 26.7 km long ring travelling along two beam pipes in opposite directions, kept at ultra-high vacuum. The particles are split into packets, so-called "bunches" and are brought to collision at four different points called interaction points, where the main experiments are located.
Two ways to express the performance of the LHC are the centre-of-mass energy ($\sqrt{s}$) and instantaneous luminosity ($\mathcal{L}$) defined as [35]

$$\mathcal{L} = \frac{f N_1 N_2 N_b}{4\pi\sigma_x\sigma_y} S. \tag{28}$$

Here, N_1 and N_2 are the numbers of particles per bunch, while N_b is the number of bunches. The revolution frequency in which the protons circulate in the accelerator is f. The parameters σ_x and σ_y give the transverse beam width and S is the so-called luminosity reduction factor to account for the crossing angles of the beams at the interaction point. Integrating the instantaneous luminosity over a period of time $\Delta t = t_1 - t_2$ gives the integrated luminosity, defined as

$$L_{\text{int}} = \int_{t_1}^{t_2} \mathcal{L} dt. \tag{29}$$

This is a useful quantity since it quantifies the amount of recorded data. The design of the LHC allows for *pp*-collisions at a centre-of-mass energy of $\sqrt{s} = 14\,\text{TeV}$ and an instantaneous luminosity of $\mathcal{L} = 10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$. During the data-taking period *Run 2* (2015-2018), it operated at $\sqrt{s} = 13\,\text{TeV}$ and had, in the end, a peak instantaneous luminosity of $19 \times 10^{33}\,\text{cm}^{-2}\,\text{s}^{-1}$ [36]. The accumulated data of Run 2 is listed in Table 3 with the years 2015 and 2016 combined.

Year	$\mathcal{L}_{\text{int}}$ (fb^{-1})
2015+2016	36.2 ± 0.8
2017	44.3 ± 1.0
2018	58.5 ± 1.2
Total	139.0 ± 2.4

Table 3: Integrated luminosity $\mathcal{L}_{\text{int}}$ for the different years of Run 2 at the LHC [37].

[2] Conseil européen pour la recherche nucléaire

2.2 ATLAS Detector

The ATLAS (**A** **T**oroidal **L**HC **A**pparatu**S**) [38] detector is a general-purpose detector which is used to explore a wide range of particle physics phenomena. It is cylindrically shaped with the central axis parallel to the beam pipe. The detector has a length of 44 m and a height of 25 m. Detector components are placed in two ways. One as *barrel* cylindrically around the beam pipe while the so-called *end-caps* are placed at the end of the barrel-shaped detectors. Figure 8 shows the main components of the ATLAS detector. The detector is forward-backwards symmetric around the interaction point and weighs approximately 7000 tonnes.

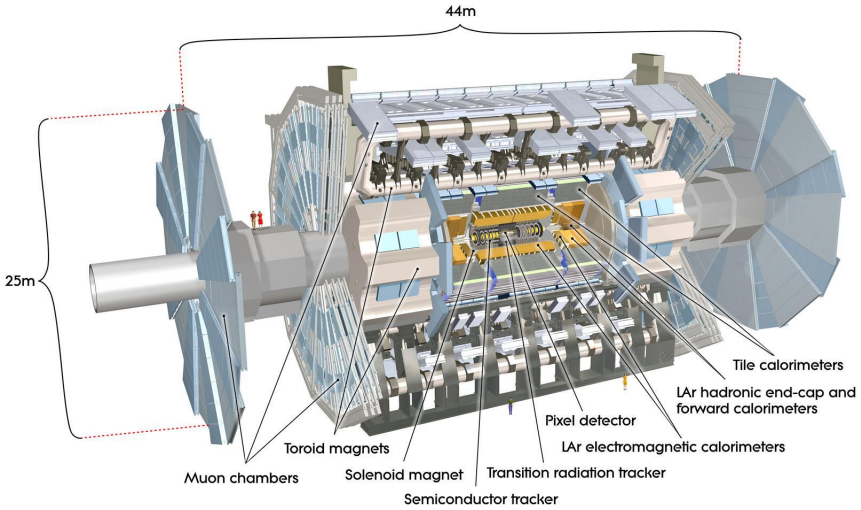


Figure 8: Computer generated image of the ATLAS detector with its main components highlighted [38].

2.2.1 The Coordinate System

The coordinate system describing the detector is right-handed, originating at the interaction point. The z-axis lies along the beam direction, with the positive z-axis pointing counterclockwise. The x-y plane lies transverse to the beam axis. The positive x-axis points to the LHC ring's centre, and the y-axis upwards. Alternatively, using the cylindrical coordinates, the polar angle θ is measured from the beam axis, and the azimuthal angle ϕ is measured around the beam axis. Figure 9 shows the described coordinate system.
The polar angle is often given in the form of the pseudorapidity η [38] defined as:

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]. \tag{30}$$

The transverse energy E_{T} and momentum p_{T} of the particles as projections in the x-y plane:

$$E_{\mathrm{T}} = \sqrt{E_x^2 + E_y^2}, \tag{31}$$

$$p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2}. \tag{32}$$

To measure the distance in the $\eta - \phi$ plane one defines:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}, \tag{33}$$

with $\Delta\eta$ and $\Delta\phi$ being the difference in pseudorapidity and azimuthal angle, respectively.

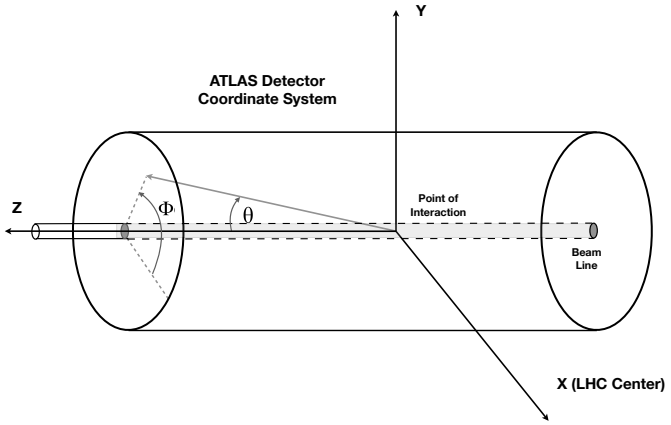


Figure 9: Coordinate system of the ATLAS detector with the interaction point as the centre.

2.2.2 Inner Detector

The inner detector (ID) is the innermost layer detecting particle collisions. It has a length of 6.2 m and radius of 1.1 m immersed in a magnetic field of 2 T created by the solenoidal magnets. It provides particle trajectory and vertex information for charged particles. Due to the high track density, the granularity of the detectors in this region has to be very fine to achieve a high momentum and spatial resolution. The ID consists of silicon pixel detectors [41] with an *Insertable B-layer* (IBL) [40], *Semiconductor Trackers* (SCT) [38] and *Transition Radiation Trackers* (TRT) [39]. This design is illustrated in Figure 10. The innermost layer is the IBL at a 33 mm radius. Around it is the pixel detector with three barrels and three end-caps on each side. It is followed by the SCTs covering a range of $|\eta| < 2.5$. It comprises four layers of silicon strip detectors in the barrel region and nine layers in the end-caps. The TRT is the outermost part of the ID. It is a straw-tube tracker that also improves particle identification, e.g. distinguishing between electrons, charged pions and other charged particles. The ID has a design resolution for $\sigma_{p_{\mathrm{T}}}$ of:

$$\frac{\sigma_{p_{\mathrm{T}}}}{p_{\mathrm{T}}} = 0.05\%_{p_{\mathrm{T}}} \oplus 1\%^{3}. \tag{34}$$

[3]The $\oplus$-notation is defined as $a \oplus b = \sqrt{a^2 + b^2}$.

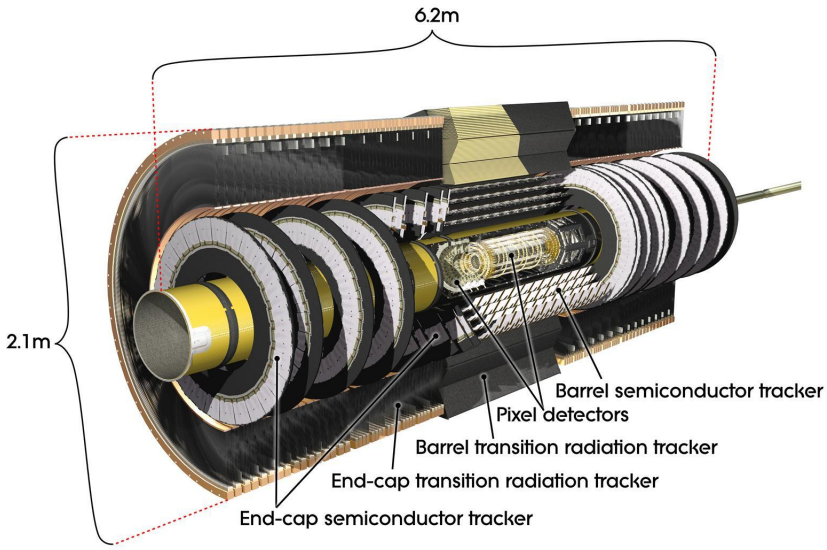


Figure 10: Computer generated image of the ATLAS inner detector with its various components highlighted [41].

2.2.3 Calorimeters

Around the inner detector, the calorimeter system is built. It consists of two main sub-systems. The inner part is the *electromagnetic calorimeter* (ECal) with the barrel region covering $|\eta| < 1.475$ and end-caps with a range of $1.375 < |\eta| < 3.2$. It gives information about the direction and energy of electrons and photons. The ECal has a layered structure alternating between an active part created from liquid argon (LAr) and an absorber made from lead. The design energy resolution is given as

$$\frac{\sigma_{E_{\mathrm{T}}}}{E} = {}^{10\%}/_{\sqrt{E}} \oplus 0.7\%. \tag{35}$$

Particles leaving the ECal move into the *hadronic calorimeter* (HCal) located outside of the ECal, as shown in Figure 11. They measure the energy deposited by hadrons and jets. The barrel regions are instrumented with scintillating tiles instead of liquid argon as active material and use steel as the absorber. It has central and extended parts covering regions of $|\eta| < 1.0$ and $0.8 < |\eta| < 1.7$, respectively. The end-caps use LAr like the ECal, but the absorber material is copper instead of lead covering $1.5 < |\eta| < 3.2$ region. The design energy resolution of the tile calorimeter and the end-caps are given by

$$\frac{\sigma_{E_{\mathrm{T}}}}{E} = {}^{50\%}/_{\sqrt{E}} \oplus 3\%. \tag{36}$$

For more forward η region $3.1 < |\eta| < 4.9$, the *forward calorimeter* (FCal) is used. It uses LAr as its active part, while the absorber is made out of copper for the inner layer, followed by two tungsten-based layers and has a design energy resolution of

$$\frac{\sigma_{E_{\mathrm{T}}}}{E} = {}^{100\%}/_{\sqrt{E}} \oplus 10\%. \tag{37}$$

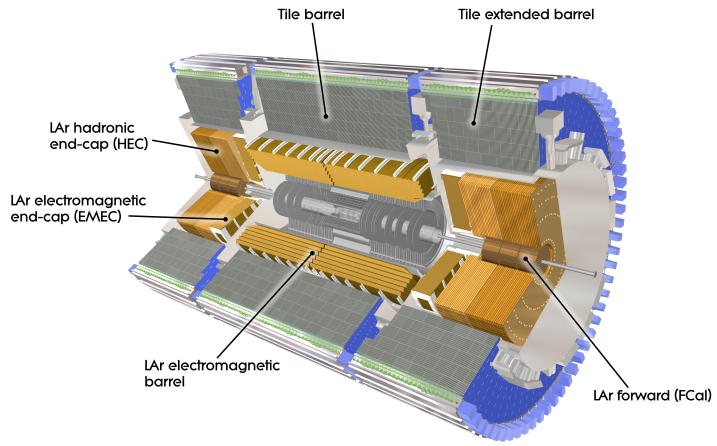


Figure 11: Computer generated image of the ATLAS calorimeter with its components highlighted [42].

2.2.4 Muon System

The outermost part of the ATLAS detector is occupied by the *Muon System* (MS). It is used for precision tracking and momentum measurement of muons. The measurement is achieved through a combination of toroidal magnetic fields that curve the particle tracks and high-precision muon chambers that detect them. There are four types of muon chambers. For momentum resolution, *Monitored Drift Tubes* (MDT) and *Cathode-Strip Chambers* (CSC) are used. MDTs cover a range of $|\eta| < 2.7$, where the CSCs are used in the end-cap regions of $2.0 < |\eta| < 2.7$. Fast trigger chambers complement these. In the central region at $|\eta| < 1.05$ *Resitive Plate Chambers* (RPCs) are used and the end-caps $1.05 < |\eta| < 2.4$ use *Thin Gap Chambers* (TGC). Figure 12 shows the described MS, an essential part of the trigger system. The design resolution for muons with $p_{\mathrm{T}} = 1\,\mathrm{TeV}$ is:

$$\frac{\sigma_{p_{\mathrm{T}}}}{p_{\mathrm{T}}} = 0.05\%_{p_{\mathrm{T}}}. \tag{38}$$

2.2.5 Trigger System

The LHC has a collision rate of 40 MHz but only a fraction of this can be stored. A two-level trigger system reduces the incoming events by selecting those of interest. The first trigger is called *level-1* trigger and is a hardware-based trigger. It selects muons, electrons, photons, jets, hadronically decaying τ leptons with high transverse momentum and events with a high total energy deposit or high missing transverse energies. Within 2.5 μs for each decision, it reduces the event rate to 75 kHz [43]. It also defines regions-of-interest (RoI) in the ϕ-η plane. Events selected by the level-1 trigger are then sent to the *High-Level Trigger* [44]. This software-based trigger comprises a level-2 trigger combined with an event filter. It reduces the data rate to around 1 kHz with an average decision time of 200 ms.

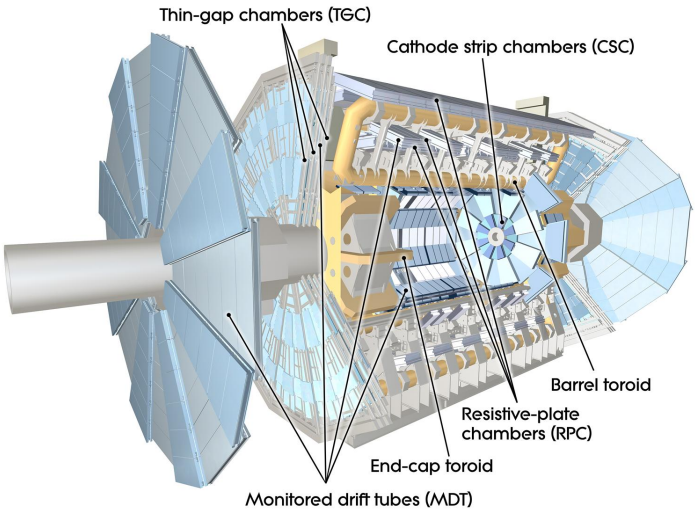


Figure 12: Computer generated image of the ATLAS MS with various components highlighted [45].

3 Neural Networks

The ability of systems to acquire knowledge by extracting patterns from raw data is known as machine learning [46]. Neural Networks (NN) are part of machine learning and give a robust framework for supervised learning. *Supervised learning* is a strategy using labelled data, where one tries to find a function mapping data to the correct label [47]. Typical examples of supervised learning are classification and regression problems. This thesis focuses on fully-connected *feed-forward regression networks*. As such, they will be the focus of this chapter. They are implemented in *Python* using the open source library *Keras* [52], which is based on *Tensorflow* [53]. The first section 3.1 will give an idea of the general design used to create NNs. The sections 3.2 and 3.3 describe how a network is trained. Section 3.4 describes the *Early Stopping Algorithm*, and section 3.5 gives an overview of a method to optimise *hyperparameters*.

3.1 General Design

The goal of a feed-forward network is to approximate a function $y = f'(\vec{x})$, which maps some input $\vec{x}$ to a target y. By defining a function f($\vec{x}$,θ) and learning the optimal values for θ the network approximates the real function $f'(\vec{x})$ [46]. The universal approximation theorem states, that a NN can approximate any function arbitrarily closely, given the correct values θ, but it gives no information about which values are optimal [48]. An NN comprises nodes (neurons) ordered in different layers as basic elements. Each node i contains a weight vector $\vec{w}_i$ and a bias b_i to be optimised. All J layers of a network together, build the composite function $f(\vec{x}) = (f^{(J)} \circ f^{(J-1)} \circ ... f^{(1)})(\vec{x})$, where the first layer represented by $f^{(1)}$ is called *input layer*. It is followed by the *hidden layers* while the final layer is called *output layer*. In a feed-forward network, each node of layer $(j-1)$ is connected to every node in layer j, and no feedback loop is created. This structure can be seen in Figure 13.

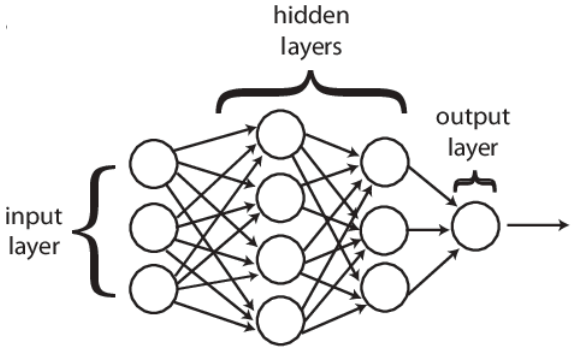


Figure 13: Basic feed-forward neural network design with nodes divided into different layers and the output of one layer feeding into the next [47].

This structure means a node gets a vector $\vec{x} = (x_1, x_2, ..., x_d)$, which contains the outputs of all nodes in the previous layer. The number of input features d defines the number of nodes in the input layer features, while the output layer is defined by the number of features the network is trained to predict. To calculate the scalar output

a_i of a node, one uses:

$$a_i = \sigma_i(\vec{w}_i \cdot \vec{x} + b_i) \tag{39}$$

where σ_i is a non-linear activation function, the most common is ReLU (*rectified linear unit*) [49]:

$$\sigma_{\text{ReLU}}(x) = \max(0, x) \tag{40}$$

Regression networks usually use the identity as the output layer since one is interested in the numerical values.

3.2 Trainings Process

The dataset used to train the NN is split into one for training and one for testing. The training dataset is used to train and, as such, find the optimal network parameters. On the other hand, the testing dataset is used to test the network's performance on unseen data, giving information about how well the network model generalises.

To find the function $f(\vec{x}, \theta)$, which best approximates $f'(\vec{x})$, the network has to optimise the weights and biases in each node. For this, one has to define a *loss function* which describes the differences between true target y and the predicted value of the target $\hat{y}$. A typical loss function for regression networks is *Mean Square Error*:

$$L(\vec{w}) = \frac{1}{n}\sum_{i}^{n}(y_i - \hat{y}_i(\vec{w}))^2 \tag{41}$$

with n the number of data points. The goal now is to minimise the loss function. A *gradient descent* [50] based method is chosen to find the optimal values for weights and bias. For NNs, a technique called *backpropagation* [51] is used, which leverages their structure to efficiently calculate the gradients $\frac{\partial L}{\partial b}$ and $\frac{\partial L}{\partial w}$. The first one defines the output of the ith neuron in the jth layer as:

$$a_i^j = \sigma \underbrace{\left(\sum_k w_{ik}^j a_k^{j-1} + b_i^j\right)}_{:=z_i^j} \tag{42}$$

Due to the network's structure, this equation depends on the bias and weights of the neuron as well as the output of neurons in the layer $(j-1)$. For the backpropagation algorithm, one starts by calculating a_i^j and z_i^j for every neuron going through the network layer by layer, starting with the input layer. At the output layer, one uses:

$$\Delta_i^j = \frac{\partial L}{\partial z_i^j} = \frac{\partial L}{\partial a_i^j}\sigma'(z_i^j) \tag{43}$$

to calculate the error Δ_i^j for each node. In Equation 43 $\sigma'(z_i^j)$ denotes the derivative of the non-linear function. The following equation:

$$\Delta_i^j = \frac{\partial L}{\partial z_i^j} = \left(\sum_k \Delta_k^{j+1} w_{ki}^{j+1}\right)\sigma'(z_i^j) \tag{44}$$

is what gives the algorithm its name. Knowing the error Δ_i^j of the output layer, we can backpropagate it again layer by layer using Equation 44 to get the error Δ_i^j for every other node. Finally, with the information of the errors, Δ_i^j one defines

$$\frac{\partial L}{\partial b_i^j} = \frac{\partial L}{\partial b_i^j}\frac{\partial b_i^j}{\partial z_i^j} = \frac{\partial L}{\partial z_i^j} = \Delta_i^j \tag{45}$$

and

$$\frac{\partial L}{\partial w_{ik}^j} = \frac{\partial L}{\partial z_i^j}\frac{\partial z_i^j}{\partial w_{ik}^j} = \Delta_i^j a_k^{j-1} \tag{46}$$

to calculate the gradients for the bias and weights. This algorithm requires only a forward and backward pass to calculate all derivatives [47].

3.3 Gradient Decent

To train a NN, one has to combine the backpropagation algorithm with a gradient descent method to calculate derivatives. The most basic version of this is *batch gradient descent* [50]:

$$\vec{w} = \vec{w} - \eta \times \nabla_{\vec{w}} L(\vec{w}) \tag{47}$$

Here the parameter $\vec{w}$ will be updated in the opposite direction of the loss function's gradient. This process is done multiple times. The number of these updates is called *epochs*. The factor η in Equation 47 is the *learning rate*, modifying how much the $\vec{w}$ changes in each epoch. There are many other ways to implement a gradient descent method. In this thesis, the *Adam-Optimizer* is used [54].

3.4 Early Stopping Algorithm

The number of epochs has a significant influence on the network performance. Training a NN with too few epochs will result in not learning enough patterns from the data called *underfitting*. On the other hand, using too many epochs can result in *overfitting*. Overfitting describes the effect of a network becoming better at finding patterns on the training dataset but performing worse regarding the testing dataset. The Early Stopping Algorithm can be used to choose the number of epochs. This algorithm stops the training process when the performance on the dataset used for testing gets worse. The parameter `patience` decides how many epochs without any improvements are accepted before the training process is stopped.

3.5 OPTUNA

One problem with training a NN is that many parameters are not changed during the training process but can affect performance. These parameters are called hyperparameters. The hyperparameters optimised in this analysis are the learning rate, the batch size, the number of layers and nodes in each hidden layer. Other hyperparameters not optimised here are the number of epochs, activation functions and gradient decent method. Finding a good set of hyperparameters that gives a well-performing

network for a specific task can be difficult.
The python package OPTUNA [55] gives a framework to automate the search for optimal hyperparameters. For this, OPTUNA uses a bayesian optimisation which assumes the existence of a mapping between the input hyperparameters and the loss function called the *objective function*. The goal is to find the minimum of this function. While one does not know this function, it can be probed by training NNs with a set of hyperparameters and sampling points of the objective function. A probabilistic *surrogate function* which includes uncertainties, is derived based on the sampled hyperparameter configurations. A third function, the *acquisition function*, is created with it. It decides where to sample next to have the highest probability of improving. Finding an optimal value of the acquisition function is an optimisation problem, but evaluating the surrogate function is much faster than training a NN model [56]. This sampling process is done multiple times, and the set of hyperparameters which gave the lowest objective function value is chosen.

4 Simulated Event Samples and Event Selection

This chapter introduces the simulated event samples used in this analysis (section 4.1) and how the generated events are weighted in order to represent best the collected data (section 4.2). Section 4.3 describes the reconstruction of physical objects and 4.4 defines the region containing the events used in the analysis called the signal region.

4.1 Dataset

The dataset consists of samples corresponding to the full Run 2 data measured at the ATLAS detector between 2015 and 2018 corresponding to an integrated Luminosity of $139\,\text{fb}^{-1}$ at $\sqrt{s} = 13\,\text{TeV}$. The samples are created via *Monte Carlo* (MC) simulations using the *POWHEG v2* [68–71] generator. The protons are modelled with *parton distribution function* (PDF) while the *Underlying Event Parton Shower* (UEPS) model of the *Pythia 8* [72] generator is used to model the parton shower and hadronisation process. The samples are created from the VBF $H \to W^-W^+ \to e\nu_e\mu\nu_\mu$ signal process for which the Feynman diagram is shown in Figure 4. The analysis does not consider any background processes. The samples include the true four-momentum vectors of the Higgs boson and outgoing partons of the VBF production mode and will be referred to as truth-level information. They are used as target to train NNs. The four-momentum vectors of the two jets with the highest transverse momentum p_T called leading jets and the muon and the electron are also given at the reconstruction-level. Reconstruction-level means the reconstructed information after the detector simulation. The neutrinos cannot be reconstructed since they leave the detector undetected. However, some information can be accessed by assuming that the transverse momenta of incoming partons are zero. Since the two neutrinos are undetected, adding the transverse momenta of all detected particles leads to a non-zero value. The negative of this value represents the missing transverse energy E_T^miss

4.2 Event Weights

The MC-generated samples are created in larger numbers than the measured data. For this reason, the Monte Carlo events have to be rescaled to the number of expected events $N_\text{Exp}^\mathcal{L}$ in $\mathcal{L} = 139\text{fb}^{-1}$. The number of simulated events N_sim is weighted to calculate the number of expected events

$$N_\text{Exp}^\mathcal{L} = \frac{N_\text{sim}}{N_\text{all}} \cdot \mathcal{L} \cdot x_\text{sec} \cdot \text{K} \cdot \epsilon_\text{filter} \tag{48}$$

Here, $\mathcal{L}$, the luminosity of the data. x_sec is the cross-section of the signal process is 0.0437%. The filter efficiency ϵ_filter is 100% and the K is the so-called K-factor which encapsulates next to leading-order information is 1. The number of simulated events and the integrated luminosity depends on the different campaigns of Run two, shown in Table 4.

Campaign	N_{all}	$\mathcal{L}\,[\text{fb}^{-1}]$
mc16a	71473517.2	36.207
mc16d	9327222.6	44.307
mc16e	9366519.6	58.450

Table 4: List of parameters used to calculate weights which scale the MC events to an integrated luminosity of $139\,\text{fb}^{-1}$.

4.3 Object Selection

From the raw data recorded by the ATLAS experiment, physics objects like electrons, muons and jets are reconstructed and identified using specific algorithms.
Electrons are reconstructed by matching the inner detector tracks with the energy deposits in the ECal [73].Electron are required to fulfill $p_T > 15\,\text{GeV}$ and $|\eta| < 2.47$. Events where electrons are reconstructed in the transition region of $1.37 < |\eta| < 1.52$ are excluded. Furthermore, electrons are required to fulfil the criteria for *FCLoose* [57] working point that uses a combination of calorimeter- and track-based isolation variables. The particle identification is done using a likelihood-based identification method [57] with electrons having to pass the *Medium* identification working point defined by it.
The muon reconstruction is done by matching muon tracks in the ID to the signals from the MS [74]. Muons with $|\eta| < 2.5$ and $p_T > 10\,\text{GeV}$ are selected. For identification and isolation, they are required to satisfy the *Loose* [58], and *FCTight-TrackOnly* [64] working points, respectively.
Jet objects are created with a *particle-flow* algorithm [60]. The anti-k_t algorithm [59] is then applied to reconstruct jets with a radius parameter of $R = 0.4$. In addition, the *Jet Vertex Tagger* (JVT) [61] and *Forward Jet Vertex Tagger* (fJVT) [62] tools are employed. JVT removes pile-up jets with $p_{\text{T}} < 60\,\text{GeV}$ and $|\eta| < 2.5$ while fJVT does the same for jets in the forward region of $|\eta| > 2.5$. Jets with $p_{\text{T}} > 20\,\text{GeV}$, $|\eta| < 4.5$ and passing the *LooseBad* [65] quality criterion are used. Since hadrons stemming from the hadronisation bottom quarks have a relatively long lifetime, jets originating from them can be traced back to a secondary vertex relative to the pp-collision interaction point. Using the DL1r b-tagging algorithm [63], these jets are identified. This analysis uses a fixed 85% efficiency working point to veto b-tagged jets.

4.4 Event Selection

The event selection requirements define the signal region. It is split into two parts: The more general Preselection tightens some object definitions and suppresses the background processes. The VBF Topology criteria select events closely resembling the signal process. The signal region selection is based on the previous publications investigating Higgs boson productions via VBF in the $H \to WW$ [66] and $H \to \tau\tau$ [67] decay modes. In Table 5, all applied criteria are listed.

Region	Requirements
Preselection	1 electron (e) & 1 muon (μ) with opposite charge
	$p_{\mathrm{T}}(e) > 15$ to 27 GeV
	$p_{\mathrm{T}}(\mu) > 10$ to 27.3 GeV
	Electron Id.: $p_{\mathrm{T}}(e) < 25$ Medium, $p_{\mathrm{T}}(e) > 25$ Tight
	Muon Id.: Tight
	Electron Iso.: FCLoose
	Muon Iso.: Tight
	$m_{\tau\tau}^{coll} > \lvert m_Z - 25\,\mathrm{GeV}\rvert$
	$M_{e\mu} > 10\,\mathrm{GeV}$
	b-Jet veto: DL1r with 85% efficiency working point
	$E_{\mathrm{T}}^{\mathrm{miss}} > 20\,\mathrm{GeV}$
	JVT > 0.5 and fJVT < 0.5
	Leading (Subleading) jet $p_{\mathrm{T}} > 40\,(30)\,\mathrm{GeV}$
VBF Topology	$N_{\mathrm{jets}} \geq 2$
	$M_{jj} > 350\,\mathrm{GeV}$
	Lead(Sublead) lepton $p_{\mathrm{T}} > 22(15)\,\mathrm{GeV}$
	Outside lepton veto

Table 5: Summary of signal region selection split into Preselection and VBF topology categories.

4.4.1 Preselection

The Preselection starts by requiring two leptons of different flavours, i.e. one electron and one muon. Due to the Higgs boson having a neutral charge, the leptons are required to have an opposite charge. A combination of single lepton and dilepton triggers are used. Each event must pass at least one of these triggers to be included. Leptons must pass a p_{T} threshold depending on the trigger response. For electrons, this threshold lies in the range of 15 to 27 GeV and 10 to 27.3 GeV for muons. Electrons identification is tightened to the *Tight* working point only for $p_{\mathrm{T}} > 25$. Muons require *FCTight* for isolation and *Tight* for identification. To reduce $Z \to \tau\tau$ background, events within the Z boson mass window $m_{\tau\tau}^{\mathrm{coll}} > \lvert m_Z - 25\,\mathrm{GeV}\rvert$ are excluded. With $m_{\tau\tau}^{\mathrm{coll}}$ being the missing invariant mass calculated with a collinear approximation. The dilepton invariant mass $M_{e\mu}$ must be higher than 10 GeV to suppress background processes originating from resonances such as J/ψ and Y mesons. A b-jet veto is applied to reduce background processes involving top quarks. The leading (subleading) jets are required to have $p_{\mathrm{T}} > 40\,(30)$ GeV. For the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}} > 20$ GeV is required.

4.4.2 VBF Topology Cuts

Each event is required to have at least two jets since the two outgoing partons from VBF production mode are hadronised into jets. The jets are also required to have an invariant mass of $M_{jj} > 350\,\text{GeV}$. For the leading (subleading) lepton a selection of $p_{\text{T}} > 22\,(15)\,\text{GeV}$ and an *Outside Lepton Veto* (OLV) is applied. The OLV is defined as

$$\text{OLV} = 2 \cdot \left| \frac{\eta_l - \overline{\eta}}{\eta_{j_1} - \eta_{j_2}} \right|, \tag{49}$$

where $\overline{\eta} = {}^{(\eta_{j_2}+\eta_{j_1})}/_{2}$ is the average pseudorapidity of the two leading jets and η_l is the pseudorapidity of the leptons. As such, leptons are required to have rapidities that fall into the rapidity gap between the two jets.

After all, cuts are applied, the number of expected events is reduced to 47.26 from 451.33 correspondingly, the number of simulated events is reduced from 2599185 to 270748 (see Table 6).

Region	N_{sim}	N_{exp}
All Events	2599185	451.33
Signal Region	270748	47.26

Table 6: Number of simulated and expected events before and after applying the signal region selection.

5 Reconstruction of the Higgs Boson Four-Momentum Vector

The four-momentum vectors of the VBF tagging jets and the Higgs boson are needed to calculate the optimal observable. In order to improve on the resolution of the optimal observable, in this thesis, one of its inputs, the reconstruction of the Higgs boson four-momentum, is investigated. This chapter explores how neural networks can be used to reconstruct the Higgs boson four-momentum. The resolutions of the four-momentum vector components are compared to the resolution of the Higgs boson reconstruction following the approximation described in section 5.1. Section 5.2 gives details of the NN setups. In section 5.3, training and optimisation of the fully connected feed-forward networks are described. In section 5.4, the resolutions of the components of the Higgs boson four-momentum vector are reconstructed using different methods are compared.

5.1 Higgs Boson Approximation

One way to reconstruct the Higgs boson is to add the four-momentum vectors of particles coming from the Higgs boson decay. While the neutrino momentum in the z-axis cannot be accessed, the Higgs boson can still be approximated by adding the remaining four-momentum vectors together:

$$p^{\mu}(H) = p^{\mu}(\mu) + p^{\mu}(e) + p^{\mu}(E_{\mathrm{T}}^{\mathrm{miss}}) \tag{50}$$

Here $p^{\mu}(\mu)$, $p^{\mu}(e)$ are the four-momentum vectors of the leptons. The four-momentum vector of the missing transverse energy $p^{\mu}(E_{\mathrm{T}}^{\mathrm{miss}})$ is created by using its magnitude and azimuthal angle. At the same time, the pseudorapidity component is set to zero. This approximated reconstruction of the Higgs boson will be referred to as *effective Higgs boson* (Eff.H) reconstruction.

5.2 The Network Setups

In addition to the effective Higgs boson, eight combinations of networks will be used to reconstruct the Higgs boson four-momentum vector or its individual components.

5.2.1 Neural Network 1

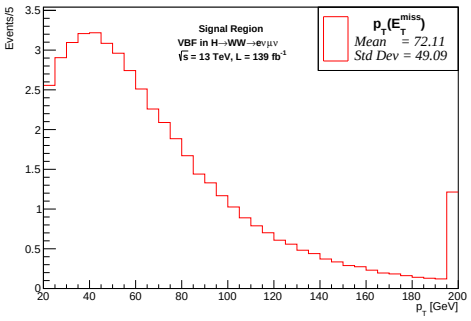
The first neural network (NN_1) calculates the Higgs bosons four-momentum vector using the same information as the Eff.H reconstruction. That means it used the full four-momentum vectors of the electron and muon as input variables in spherical coordinates p_{T}, η, ϕ and E. Since the sum of the neutrinos cannot be fully reconstructed, the missing transverse energy's magnitude and azimuthal angle components are used. The ϕ component are given to the network as $cos(\phi)$ and $\sin(\phi)$. The input variables are shown in Figure 14-16. The targets are the Higgs bosons four-momentum vector's components in cartesian coordinates p_{x}, p_{y}, p_{z} and E. The loss function used for this network is Mean Square Error (MSE), defined as:

$$\mathrm{MSE}(\vec{w}) = \frac{1}{n}\sum_{i}^{n}(p_x^i - \hat{p}_x^i(\vec{w}))^2 + (p_y^i - \hat{p}_y^i(\vec{w}))^2 + (p_z^i - \hat{p}_z^i(\vec{w}))^2 + (E^i - \hat{E}^i(\vec{w}))^2. \tag{51}$$

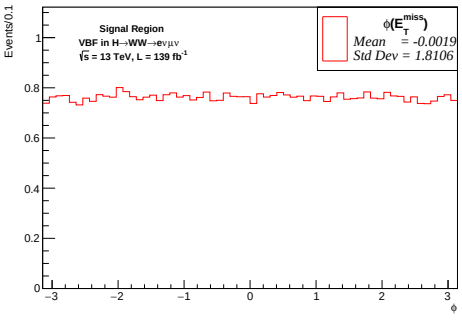
Here n is the number of events. The variable with a hat in each term of the sum represents the prediction of the network. The corresponding variable without a hat is the true target. The network setup for NN_1 is shown in Table 7

Neural Network	Input Variables		Target Variables	Loss Function
NN_1	Electron:	p_T, η, ϕ, E	p_x(H, truth) p_y(H, truth) p_z(H, truth) E(H, truth)	MSE
	Muon:	p_T, η, ϕ, E		
	$E_\mathrm{T}^\mathrm{miss}$:	p_T, ϕ		

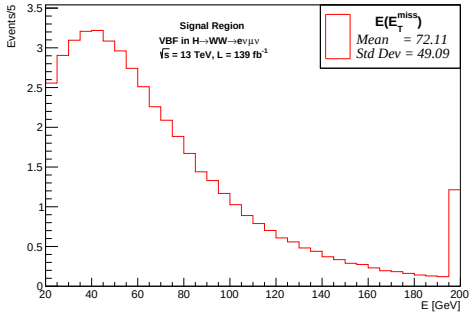
Table 7: List of input variables, targets and loss function defining NN_1.



(a) Electron p_T

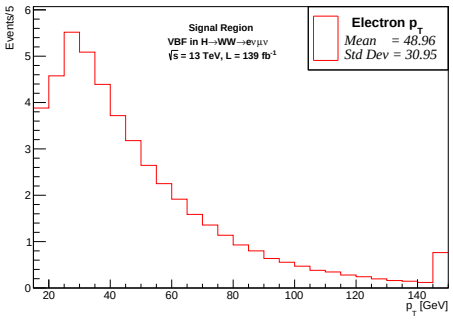


(b) Electron ϕ

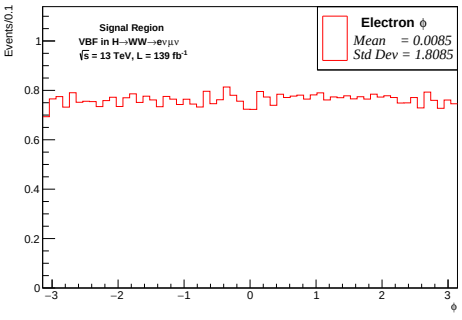


(c) Electron E

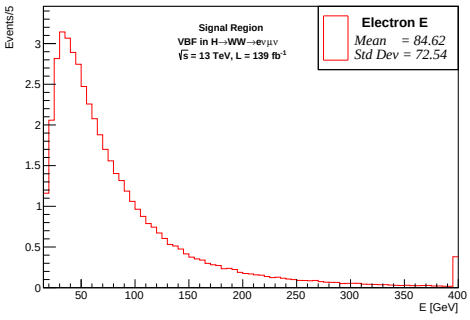
Figure 14: Signal region distributions of $E_\mathrm{T}^\mathrm{miss}$ input variables. All distributions are normalised to 139 fb^{-1}. Underflow and overflow bins are included.



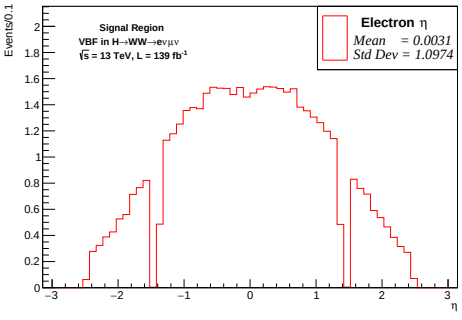
(a) Electron p_{T}



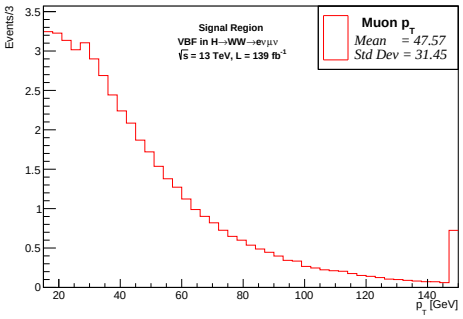
(b) Electron ϕ



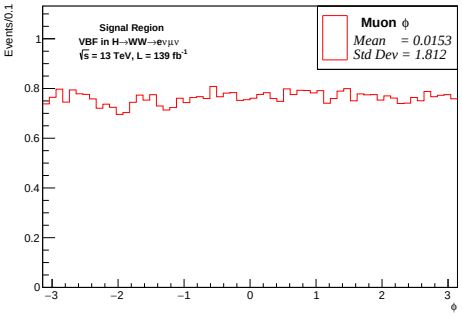
(c) Electron E



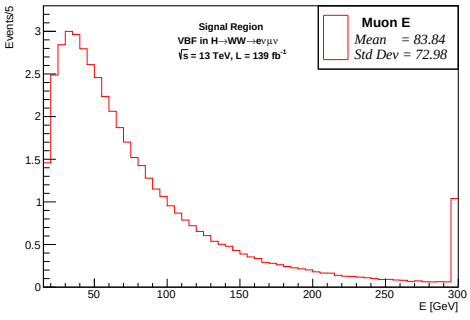
(d) Electron η



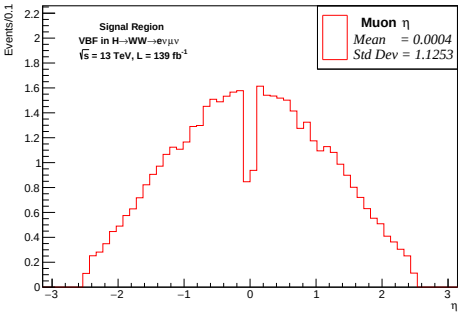
(e) Muon p_{T}



(f) Muon ϕ

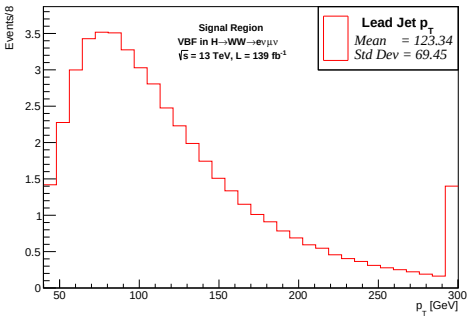


(g) Muon E

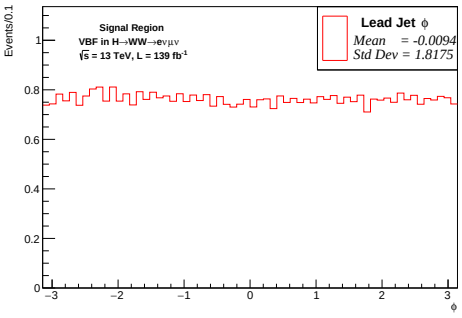


(h) Muon η

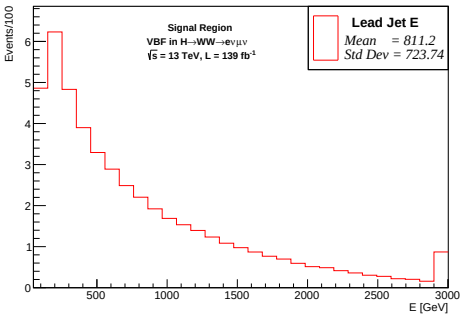
Figure 15: Signal region distributions of input variables. The plots a), b), c) and d) show the distributions for the electron while e), f), g) and h) show them for the muon. All distributions are normalised to 139 fb^{-1}. Underflow and overflow bins are included.



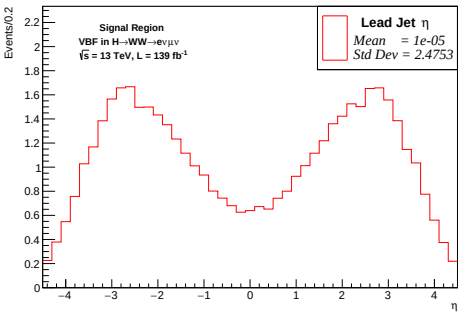
(a) Leading jet p_{T}



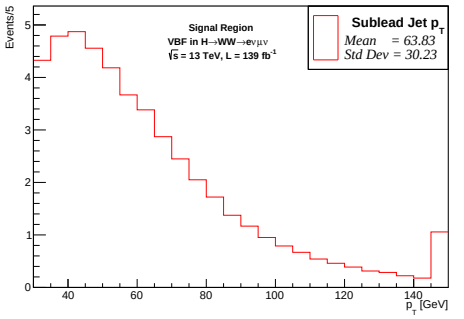
(b) Leading jet ϕ



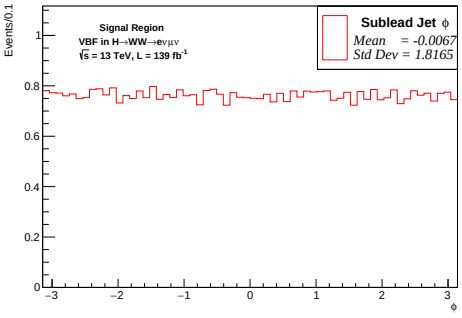
(c) Leading jet E



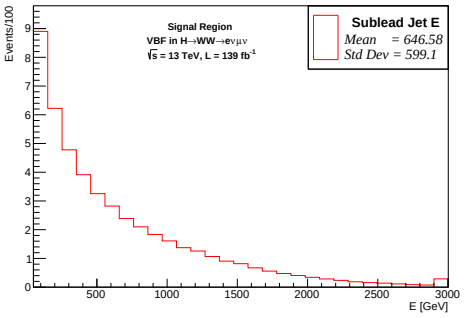
(d) Leading jet η



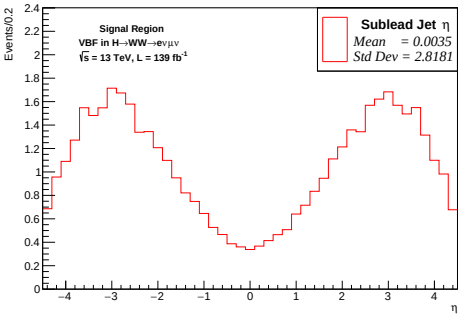
(e) Subleading jet p_{T}



(f) Subleading jet ϕ



(g) Subleading jet E



(h) Subleading jet η

Figure 16: Signal region distributions of input variables. The plots a), b), c) and d) show the distributions for the leading jet while e), f), g) and h) show them for the subleading jet. All distributions are normalised to 139 fb^{-1}. Underflow and overflow bins are included.

5.2.2 Neural Network 2

In addition to the variables used by NN_1, the second network (NN_2) also uses information from the leading and subleading jets. They are also provided to the network as spherical coordinates p_T, η, ϕ and E. The targets and loss function remains unchanged. The setting used in NN_2 is shown in Table 8.

Neural Network	Input Variables		Target Variables	Loss Function
NN_2	Electron:	p_T, η, ϕ, E	p_x(H, truth)	MSE
	Muon:	p_T, η, ϕ, E	p_y(H, truth)	
	(Sub)Lead Jet:	p_T, η, ϕ, E	p_z(H, truth)	
	$E_\mathrm{T}^\mathrm{miss}$:	p_T, ϕ	E(H, truth)	

Table 8: List of input variables, targets and loss function used in NN_2.

5.2.3 Neural Network 3 and 4

The third network (NN_3) and fourth network (NN_4) use the same input and target variables as NN_1 and NN_2, respectively; however, the loss function is a modified version of MSE. An additional term is added, leading to the new loss function:

$$\mathrm{MSE}_\mathrm{Mass} = \mathrm{MSE} + \frac{1}{n}\sum_{i}^{n}(125 - \hat{M}^i_\mathrm{Higgs}(\vec{w}))^2 \tag{52}$$

Since the mass of the Higgs boson is reliably well measured, this network is incentivised to predict a Higgs boson four-momentum vector corresponding to the observed Higgs boson mass of 125 GeV. The variable $\hat{M}^i_\mathrm{Higgs}$ is the Higgs boson invariant mass predicted by the NN. It is calculated from its prediction for the four-momentum components by using the following:

$$\hat{M}^i_\mathrm{Higgs} = \sqrt{(\hat{E}^i)^2 - (\hat{p}^i_x)^2 - (\hat{p}^i_y)^2 - (\hat{p}^i_z)^2}. \tag{53}$$

The Higgs boson four-momentum vector reconstruction using NN_1 (NN_2) will be referred to as NN_3 (NN_4). Both network setup parameters are shown in Table 9.

Neural Network	Input Variables		Target Variables	Loss Function
NN_3	Electron:	p_T, η, ϕ, E	p_x(H, truth)	MSE_{Mass}
	Muon:	p_T, η, ϕ, E	p_y(H, truth)	
	E_T^{miss} :	p_T, ϕ	p_z(H, truth)	
			E(H, truth)	
NN_4	Electron:	p_T, η, ϕ, E	p_x(H, truth)	MSE_{Mass}
	Muon:	p_T, η, ϕ, E	p_y(H, truth)	
	(Sub)Lead Jets:	p_T, η, ϕ, E	p_z(H, truth)	
	E_T^{miss} :	p_T, ϕ	E(H, truth)	

Table 9: List of input variables, targets and loss function defining NN_3 and NN_4.

5.2.4 Neural Networks 5 to 10

The four NN already discussed each predicted the complete four-momentum vector of the Higgs boson. The following six networks only predict parts of the Higgs boson four-momentum vector. Their predictions will be combined to give four complete sets of the Higgs boson four-momentum vector's components.

The networks 4 to 8 are called NN_5, NN_6, NN_7 and NN_8 and correspond to former networks setups NN_1, NN_2, NN_3 and NN_4, respectively. The difference lies in the target variables since only the p_z and E components will be predicted by the four networks.
The missing component p_x (p_y) is instead reconstructed using the network NN_9 (NN_{10}). The input features of NN_9 and NN_{10} are the four-momentum vectors of the leptons, leading (subleading) jet and E_T^{miss}. The ϕ components are given as $\cos(\phi)$ and $\sin(\phi)$ like for the networks above. NN_9 and NN_{10} use the loss function MSE as defined in Equation 51. The setups for the network 5 to 10 are shown in Table 10
To make this approach work for the modified loss function MSE_{mass} used by NN_7 and NN_8, one has to modify the calculation of $\hat{M}^i_{Higgs}$. The reason is that NN_7 (NN_8) does not provide any predictions for p_x (p_y) to be used in the calculation. Instead, the predictions of NN_9 and NN_{10} are input to the loss function.

The predictions of these six networks are combined to reconstruct the following four complete Higgs boson four-momentum vectors: $NN_{(5,9,10)}$, $NN_{(6,9,10)}$, $NN_{(7,9,10)}$ and $NN_{(8,9,10)}$, where the numbers represent the number of the networks used in the reconstruction.

Neural Network	Input Variables		Target Variables	Loss Function
NN_5	Electron:	p_T, η, ϕ, E	p_z(H, truth)	MSE
	Muon:	p_T, η, ϕ, E	E(H, truth)	
	E_T^{miss} :	p_T, ϕ		
NN_6	Electron:	p_T, η, ϕ, E		MSE
	Muon:	p_T, η, ϕ, E	p_z(H, truth)	
	(Sub)Lead Jets:	p_T, η, ϕ, E	E(H, truth)	
	E_T^{miss} :	p_T, ϕ		
NN_7	Electron:	p_T, η, ϕ, E	p_z(H, truth)	MSE_{Mass}
	Muon:	p_T, η, ϕ, E	E(H, truth)	
	E_T^{miss} :	p_T, ϕ		
NN_8	Electron:	p_T, η, ϕ, E		MSE_{Mass}
	Muon:	p_T, η, ϕ, E	p_z(H, truth)	
	(Sub)Lead Jets:	p_T, η, ϕ, E	E(H, truth)	
	E_T^{miss} :	p_T, ϕ		
NN_9	Electron:	p_T, η, ϕ, E		MSE
	Muon:	p_T, η, ϕ, E	p_z(H, truth)	
	(Sub)Lead Jets:	p_T, η, ϕ, E	E(H, truth)	
	E_T^{miss} :	p_T, ϕ		
NN_{10}	Electron:	p_T, η, ϕ, E		MSE
	Muon:	p_T, η, ϕ, E	p_z(H, truth)	
	(Sub)Lead Jets:	p_T, η, ϕ, E	E(H, truth)	
	E_T^{miss} :	p_T, ϕ		

Table 10: List of input variables, targets and loss function defining the networks NN_5, NN_6, NN_7, NN_8, NN_9 and NN_{10}.

5.3 Hyperparameter Optimization with OPTUNA

With the ten network setups defined, one has to specify the hyperparameters. While the network optimisation is done for each network individual, the search space and other hyperparameters are fixed.
All networks use 400 epochs and the early stopping algorithm with a `patience` of 5 to prevent overfitting. Adam-optimiser [54] is used as a gradient descent method. The activation function of the hidden layers is ReLU [49] while the output layer uses the identity. The learning rate, batch size, number of layers and number of layers in each individual layer are optimised by OPTUNA [55]. For this, the search space of each hyperparameter is defined as follows: The learning rate is set to be between 10^{-5} and 10^{-1}. The batch size is set to be between 400 and 2400 in steps of 200. The number of hidden layers is between 2 and 7; each hidden layer can have 20 to 100 nodes in steps of 10. The hyperparameters to be optimised are listed in Table 11. For

Hyperparameter	Range
learning rate	10^{-5} to 10^{-1}
batch size	400 to 2400 in steps of 200
layers	2 to 7
nodes per layer	20 to 100 in steps of 10

Table 11: List of hyperparameters optimised with OPTUNA and the ranges from which OPTUNA can choose.

the training and optimisation, simulated samples are split into 70% of total events for the training and the remaining 30% for testing. The input features are standardised to a mean of zero and a variance of one. Standardising the input features to a similar range is done so that the network does not prioritise inputs with a more extensive range due to their larger effect on the loss function. The loss calculated on the testing samples was chosen as an objective function, which OPTUNA tries to estimate. The networks are optimised in 100 iterations called *trials*, where each trial represents a sampling of the objective function. To illustrate this, Figure 17 shows the optimisation process of NN_2. Blue dots represent each trial's validation loss or objective value, while the red line shows the best objective value throughout the optimisation. Most of the improvement takes place during the first 20 trials. Afterwards, the number of improvements gradually decline. This implies that the choice of 100 trials is enough to optimise the network. Figures for the optimisation process of the other networks are shown in Appendix A.

After the optimisation, the model with the set of hyperparameters resulting in the best objective value is used to predict the components of the Higgs boson four-momentum vector. The hyperparameters determined for each network are summarised in Table 12.
The learning rate for all networks is in a similar range, with NN_1 having the largest learning rate of 0.0036 and NN_9 having the lowest learning rate of 0.0002. The batch

sizes vary strongly among networks, with NN_8 hitting the search space's upper and NN_2 and NN_{10} the lower limit. Similarly, the number of nodes varies, with the largest node located mostly in the first layer. The exceptions are NN_4 and NN_{10}. These two networks also have the largest number of hidden layers, with NN_{10} reaching the upper limit of 7 layers.

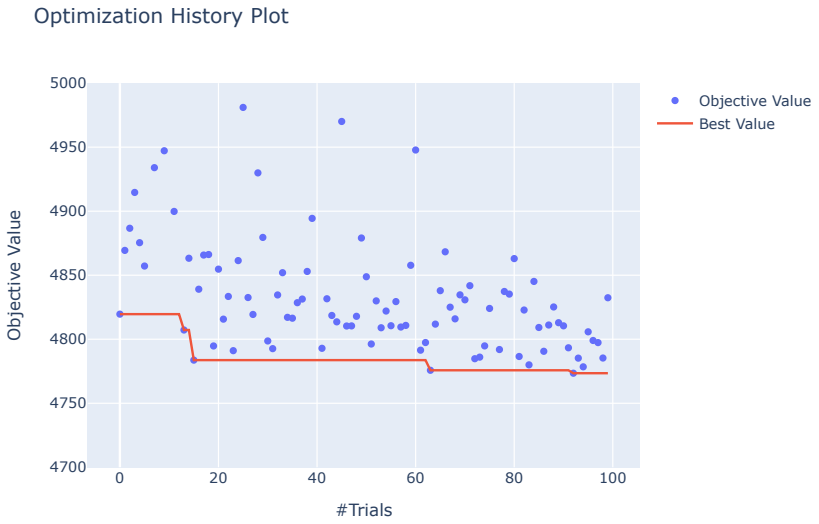


Figure 17: OPTUNA optimisation process of NN_2. The x-axis shows the number of trials (#Trials), and the y-axis is the objective value. Blue dots represent the objective value in each trial, while the red curve shows the best objective value. Some trial numbers are missing since their objective value is too large and outside the range.

Hyperparameter		NN_1	NN_2	NN_3	NN_4	NN_5	NN_6	NN_7	NN_8	NN_9	NN_{10}
Batch Size		1200	400	600	600	2000	1000	1800	2400	600	400
Learning Rate		0.0036	0.0010	0.0007	0.0003	0.0013	0.0005	0.0006	0.0015	0.0002	0.0010
Layers		4	3	2	6	2	5	3	2	3	7
Nodes in Layer:	1	100	80	60	20	90	90	50	100	90	90
	2	70	70	20	90	80	80	20	60	80	100
	3	40	30	-	70	-	70	30	-	60	30
	4	70	-	-	40	-	30	-	-	-	20
	5	-	-	-	20	-	60	-	-	-	70
	6	-	-	-	40	-	-	-	-	-	100
	7	-	-	-	-	-	-	-	-	-	20

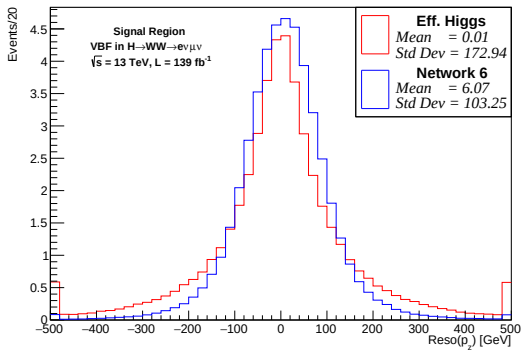
Table 12: List of the optimised hyperparameter for each NN used in the analysis.

5.4 Comparison of Resolutions for Components of the Higgs Boson Four-Momentum Vectors

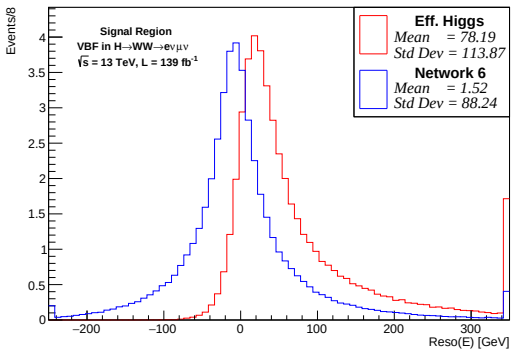
After training, the ten networks are used to predict the Higgs boson four-momentum vector's components. This results in eight different versions of the Higgs boson four-momentum vector. Their performance is compared by calculating the resolution according to Equation 54.

$$\text{Res}((p^{\mu}_{H,\text{NN}_i})_j) = (p^{\mu}_{H,\text{truth}})_j - (p^{\mu}_{H,\text{NN}_i})_j \tag{54}$$

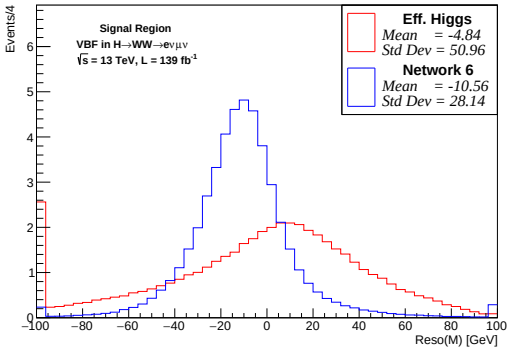
Here $(p^{\mu}_{H,\text{NN}_i})_j$ corresponds to the Higgs boson four-momentum vector's component j predicted by Network NN_i and $(p^{\mu}_{H,\text{truth}})_j$ to its truth-level counterpart. The resolution of the Eff.H (see section 5.1) is calculated analogue to this. In addition to the resolutions of the Higgs boson vector's components p_x, p_y, p_z and E, the resolution for the Higgs boson mass M is also calculated using Equation 53. Resolutions of predicted variables for NN_1, NN_6, NN_9 and NN_{10} are shown in Figure 18-20, the remaining resolutions are shown in Appendix B. For each component, the standard deviation of the resolution is listed in Table 13.



(a) Resolution of p_z

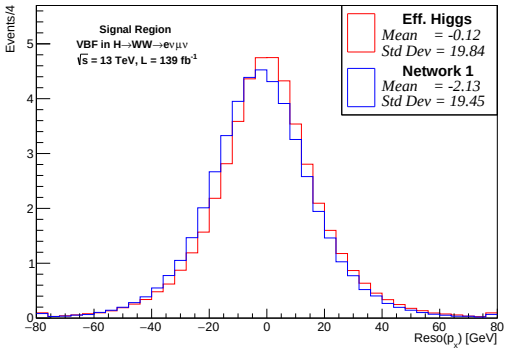


(b) Resolution of E

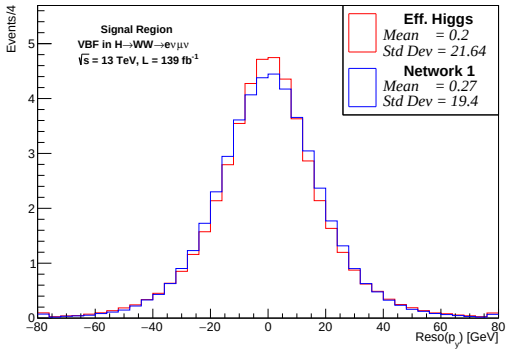


(c) Resolution of M

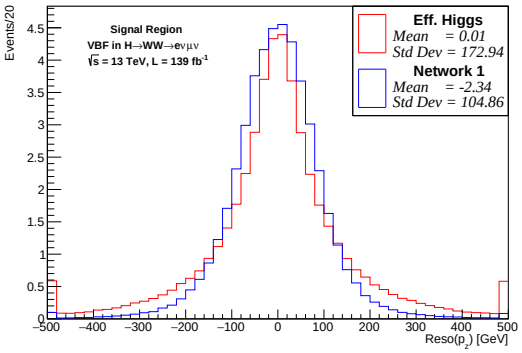
Figure 18: Resolution of different components of the Higgs boson four-momentum vector reconstructed with NN_6 in blue and red the resolution of the Eff.H reconstruction components. All resolutions are normalised to 139fb^{-1}. Overflow and underflow bins are included.



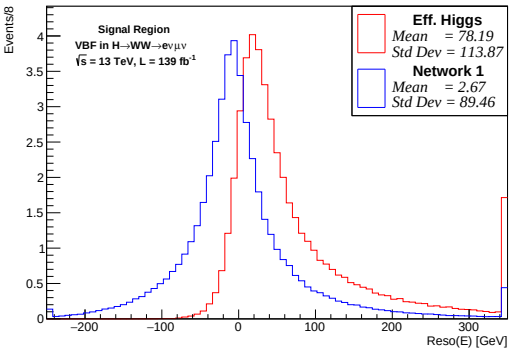
(a) Resolution of p_{x}



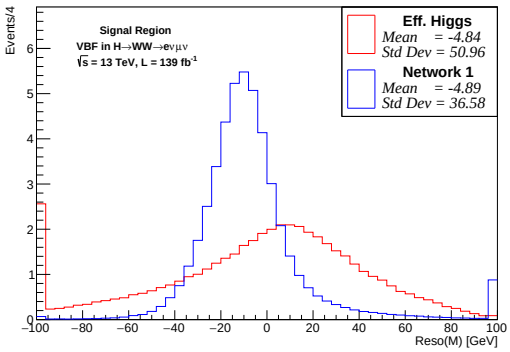
(b) Resolution of p_{y}



(c) Resolution of p_{z}

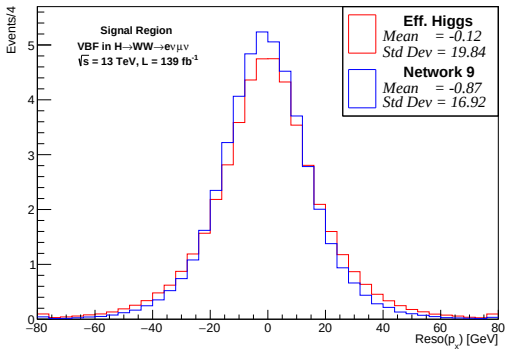


(d) Resolution of E

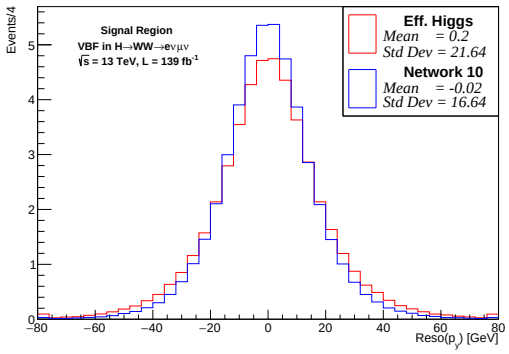


(e) Resolution of M

Figure 19: Resolutions of different components of the Higgs boson four-momentum vector reconstructed with NN_1 in blue and red the resolution of the Eff.H reconstruction components. All resolutions are normalised to 139fb^{-1}. Overflow and underflow bins are included.



(a) Resolution of p_{x} predicted by NN_9



(b) Resolution of p_{y} predicted by NN_{10}

Figure 20: Resolution of the p_{x} and p_{y} components of the Higgs boson four-momentum vector reconstructed with NN_9 and NN_{10} in blue and red the resolution of the Eff.H reconstruction components. All resolutions are normalised to 139fb^{-1}. Overflow and underflow bins are included.

By comparing the standard deviations of NN_1 to the Eff.H reconstruction, one can see that the resolution of all parameters improved. Especially p_{z}, M and E components improved by 39.37%, 28.21% and 21.28%, respectively. Relatively small improvements on p_{x} and p_{y} are expected since the Eff.H reconstruction possesses all the needed information regarding the transversal plane. In NN_2 adding jet information leads to further improvements of around $\approx 1.5\%$ for p_{z} and E while M improved by 11.34% compared to NN_1. Changing the loss function to $\mathrm{MSE}_{\mathrm{mass}}$ (NN_3 and NN_4) worsened the resolution in all parameters except p_{z} and M compared to Eff.H, with p_{z} still performing worse than for NN_1 and NN_2. The predicted mass M, on the other hand, outperformed all the predictions and gained 99.61% compared to the Eff.H reconstruction. This is because the loss calculated by the additional term in Equation 52 is much larger than the loss for MSE. As such, the network prioritised minimising the loss function for the added term and neglected the MSE term. The networks which only predicted the p_{z} and E components (NN_5, NN_6, NN_7 and NN_8) did not gain any significant improvements compared to NN_2, but the trend of networks using jet information gaining $\approx 1.5\%$ in p_{z} and E compared to their counterpart without jets continued. The network NN_9 (NN_{10}) improved p_{x} (p_{y}) by 11.92% (14.49%) compared to NN_2. The best-performing reconstruction method is $\mathrm{NN}_{(6,9,10)}$. It reached $\sigma_{p_x} = 16.92\,\mathrm{GeV}$ and $\sigma_{p_y} = 16.64\,\mathrm{GeV}$ via NN_9 and NN_{10}, respectively. The E and p_{z} components predicted by NN_6 achieved a standard deviation of $\sigma_E = 88.24\,\mathrm{GeV}$ and $\sigma_{p_z} = 103.25\,\mathrm{GeV}$. Using them to predict M achieved a standard deviation of $\sigma_M = 28.14\,\mathrm{GeV}$.

Reconstruction	σ_{p_x} [GeV]	σ_{p_y} [GeV]	σ_{p_z} [GeV]	σ_E [GeV]	σ_M [GeV]
Eff.H	19.84	21.64	172.94	113.87	50.96
NN_1	19.45	19.44	104.86	89.46	36.58
NN_2	19.21	19.46	103.27	88.27	32.43
NN_3	109.97	109.16	164.74	151.82	0.20
NN_4	111.21	113.77	185.62	158.27	0.20
$NN_{(5,9,10)}$	16.92	16.64	104.78	89.81	46.24
$NN_{(6,9,10)}$	16.92	16.64	103.25	88.24	28.14
$NN_{(7,9,10)}$	16.92	16.64	104.99	90.00	32.23
$NN_{(8,9,10)}$	16.92	16.64	103.29	88.51	41.60

Table 13: Standard deviations of the different Higgs boson four-momentum vector's component resolutions for the different reconstruction methods.

6 Estimation of the Sensitivity to Constrain $\tilde{d}$

The mean value of the distribution of a CP-odd observable can be used to constrain $\tilde{d}$, which parameterises the strength CP-violation in the HVV vertex. By calculating the mean value of a CP-odd observable for the different assumed values of $\tilde{d}$, one can derive a *gauge curve*. In the approximate linear regime of the SM case ($\tilde{d} = 0$), the gauge curve can be approximated by the linear relationship $\langle \text{CP-odd Obs.} \rangle = a + b \cdot \tilde{d}$. $\tilde{d}$ can then be estimated from the measured value of $\langle \text{CP-odd Obs.} \rangle$. Section 6.1 details how to calculate the different CP-odd observables. Section 6.2 explains how BSM distributions can be predicted by reweighting events. Section 6.3 describes how gauge curve fits are used to estimate the sensitivity of $\tilde{d}$ for the CP-odd observables.

6.1 CP-odd Observables

The optimal observable $\mathcal{OO}$ is calculated using the Python module *vbfcprw* [75], which uses *Fortran* routines extracted from *HAWK* [76–80]. HAWK is a MC-generator for the Higgs boson production channels VFB and VH. The module requires the four-momentum vectors of the VBF tagging jets and the Higgs boson. As jets, the two leading jets given in the samples are used, while the Higgs boson four-momentum vectors are estimated in chapter 5. As such, nine $\mathcal{OO}$s can be calculated. One using the Eff.H reconstruction $\mathcal{OO}_{\text{Eff.H}}$ and one for each NN reconstruction labelled $\mathcal{OO}_i$ with i the network used for the reconstruction. The truth-level observable $\mathcal{OO}_{\text{truth}}$ is also used. At truth-level, access to the parton information is possible. As such, these partons are used instead of jets for $\mathcal{OO}_{\text{truth}}$. Only events with $-12 < \mathcal{OO} < 12$ are considered in this analysis to reduce the effect of outliers. This requirement is applied individually to each $\mathcal{OO}$ distribution. Consequently, the number of expected events differs slightly for each choice of calculated $\mathcal{OO}$.
Resolution are calculated by subtracting $\mathcal{OO}_{\text{truth}}$ from each $\mathcal{OO}$. For the resolutions, the requirement of $-12 < \mathcal{OO} < 12$ is only applied to the reconstruction-level $\mathcal{OO}$. As an example, Figure 21 shows the resolutions for $\mathcal{OO}_{(6,9,10)}$ and $\mathcal{OO}_{\text{Eff}}$. Similar plots for the other observables are provided in Appendix C. The standard deviations for all $\mathcal{OO}$ resolutions are listed in Table 14. The resolution of the $\mathcal{OO}_{\text{Eff.H}}$ could be improved by 0.25% using $\mathcal{OO}_{(8,9,10)}$ with a standard deviation of 3.5709 regarding the resolution. The networks using the modified MSE (NN_3 and NN_4) had the worst resolutions.

Resolution of	$\mathcal{OO}_{\text{Eff.H}}$	$\mathcal{OO}_1$	$\mathcal{OO}_2$	$\mathcal{OO}_3$	$\mathcal{OO}_4$	$\mathcal{OO}_{(5,9,10)}$	$\mathcal{OO}_{(6,9,10)}$	$\mathcal{OO}_{(7,9,10)}$	$\mathcal{OO}_{(8,9,10)}$
St. Dev.	3.5797	3.5764	3.5779	3.6931	3.6117	3.5770	3.5760	3.5729	3.5709

Table 14: List of standard deviations of the resolution corresponding to different $\mathcal{OO}$s.

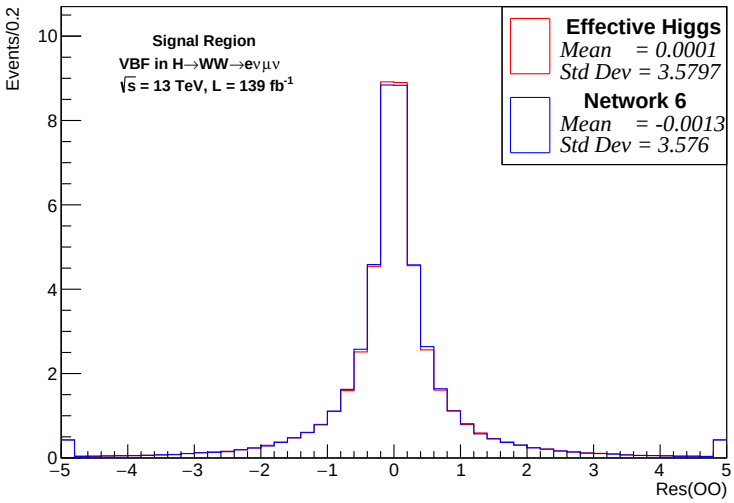


Figure 21: Resolutions of $\mathcal{OO}_{\text{Eff}}$ and $\mathcal{OO}_{(6,9,10)}$. The reference value for both distributions is the truth-level $\mathcal{OO}$. Distributions are normalized to $139\,\text{fb}^{-1}$. Overflow and underflow bins are included.

In addition to the $\mathcal{OO}$, $\Delta\Phi_{jj}^{\text{signed}}$ and $\mathcal{O}_{\text{Reg}}$ are also compared. $\Delta\Phi_{jj}^{\text{signed}}$ is defined in Equation 26 and requires the ϕ components of the leading and subleading jets. The symbolic regression observable $\mathcal{O}_{\text{Reg}}$ is defined in Equation 27 and uses $\Delta\Phi_{jj}^{\text{signed}}$ and the transverse momenta of the jets. To align them closer to the event numbers of the $\mathcal{OO}$s, they are required to fulfil the $-12 < \mathcal{OO}_{\text{Eff.H}} < 12$ requirement. Details of CP-odd observables are discussed in section 1.3.3.

6.2 Prediction for BSM Distributions

In order to create gauge curves, simulated distributions of the CP-odd observables have to be produced for the BSM cases ($\tilde{d} \neq 0$). The Monte Carlo simulated samples are generated only for SM ($\tilde{d} = 0$) case and thus cannot be used for this purpose. The BSM samples are produced by reweighting the events of the SM samples. The weights w used for reweighting are calculated by dividing the squared matrix element given in Equation 21 by $|\mathcal{M}_{\text{SM}}|^2$:

$$w(\tilde{d}) = \frac{|\mathcal{M}|^2}{|\mathcal{M}_{\text{SM}}|^2} \tag{55}$$

$$= 1 + \tilde{d} \cdot \frac{2\Re\left(\mathcal{M}_{\text{SM}}^{*}\mathcal{M}_{\text{CP-odd}}\right)}{|\mathcal{M}_{\text{SM}}|^2} + \tilde{d}^2 \cdot \frac{|\mathcal{M}_{\text{CP-odd}}|^2}{|\mathcal{M}_{\text{SM}}|^2} \tag{56}$$

$$= 1 + \tilde{d} \cdot w_{\text{lin}} + \tilde{d}^2 \cdot w_{\text{quad}} \tag{57}$$

The w_{lin} and w_{quad} are the linear and quadratic weights terms, repsectievely. They are calculated similarly to $\mathcal{OO}$, using the vbfcprw module for every event.

6.3 Gauge Curves

This section describes the gauge curve fit setup for the $\mathcal{OO}$. It is done analogue for the other CP-odd observables $\Delta\Phi_{jj}^{\text{signed}}$ and $\mathcal{O}_{\text{Reg}}$ mentioned above. As such, following this description, results are provided for all the CP-odd observables in the end. The gauge curves show the mean of the observables as a function of $\tilde{d}$. Since the distributions are weighted, the weighted mean for n events

$$\langle \mathcal{OO} \rangle = \frac{\sum_{i=1}^{n} w_i \cdot \mathcal{OO}_i}{\sum_{i=1}^{n} w_i} \tag{58}$$

is used. The error is given by the standard error of the mean:

$$s_{\langle \mathcal{OO} \rangle} = \frac{\sigma_{\mathcal{OO}}}{\sqrt{N_{\text{eff}}}} \quad \text{with} \quad N_{\text{eff}} = \frac{(\sum_{i=1}^{n} w_i)^2}{\sum_{i=1}^{n} w_i^2}. \tag{59}$$

Here, N_{eff} is the effective number of events and $\sigma_{\mathcal{OO}}$ is the standard deviation of the distribution of the $\mathcal{OO}$. Figure 22 shows the gauge curve for $\mathcal{OO}_{(6,9,10)}$. The

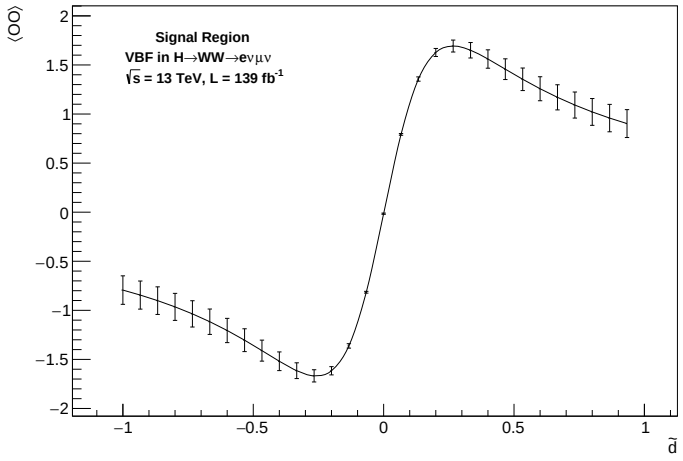


Figure 22: Gauge Curve for $\mathcal{OO}_{(6,9,10)}$ in the range of $[-1, 1]$. Error bars represent the standard error of weighted mean on $\langle \mathcal{OO} \rangle$.

reweighting with $\tilde{d} \neq 0$ leads to a weighted mean value different from zero. For values close to $\tilde{d} = 0$, an approximately linear relation between $\langle \mathcal{OO} \rangle$ and $\tilde{d}$ is visible. The region $\tilde{d} = [-0.05, 0.05]$ is used to perform a linear fit using

$$\langle \mathcal{OO} \rangle = a \cdot \tilde{d} + b, \tag{60}$$

where a is the slope and b is the y-intercept. Since one expects the gauge curve to be zero for $\tilde{d} = 0$, the y-intercept should be compatible with zero. Solving Equation 60 results in the following expression for $\tilde{d}$:

$$\tilde{d} = \frac{\langle \mathcal{OO} \rangle - b}{a}. \tag{61}$$

Here, $\langle \mathcal{O}\mathcal{O} \rangle$ is the weighted mean for $(\tilde{d})$. The error $\sigma_{\tilde{d}}$ of $\tilde{d}$ is calculated using *Gaussian error propagation*:

$$\sigma_{\tilde{d}} = \sqrt{\left(\frac{\partial \tilde{d}}{\partial a}\right)^2 s_a^2 + \left(\frac{\partial \tilde{d}}{\partial b}\right)^2 s_b^2 + \left(\frac{\partial \tilde{d}}{\partial \langle \mathcal{O}\mathcal{O} \rangle}\right)^2 s_{\langle \mathcal{O}\mathcal{O} \rangle}^2} \tag{62}$$

$$= \frac{1}{|a|}\sqrt{\tilde{d}^2 s_a^2 + s_b^2 + s_{\langle \mathcal{O}\mathcal{O} \rangle}^2}. \tag{63}$$

Here, s_a and s_b are the uncertainties of the fit parameters while $s_{\langle \mathcal{O}\mathcal{O} \rangle}$ is the standard error of the SM distributions. By assuming that the standard error $s_{\langle \mathcal{O}\mathcal{O} \rangle}$ is dominant this can be further simplified to

$$\sigma_{\tilde{d}} = \frac{s_{\langle \mathcal{O}\mathcal{O} \rangle}}{|a|} \tag{64}$$

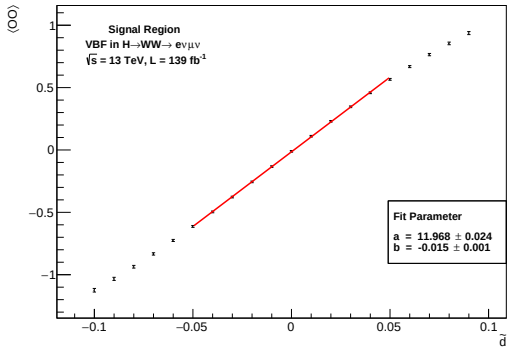
To better compare the uncertainties of $\tilde{d}$, the standard error of $\sigma_{\tilde{d}}$ is estimated and denoted $\Delta\sigma_{\tilde{d}}$. For this, the estimator for the variance of the sample variance [81]

$$V[s^2] = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\mu_2^2\right), \tag{65}$$

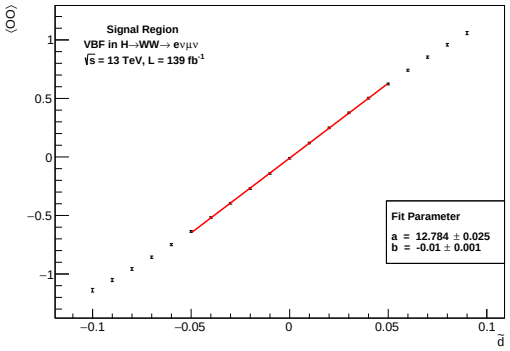
is used. Here, μ_k is the k^{th} central moment. Assuming a Gaussian distribution, i.e. $\mu_4 = 3\mu_2^2$ and $n \gg 3$, the estimator for the error on the standard error is given by

$$\Delta\sigma_{\tilde{d}} = \frac{s_{\tilde{d}}}{\sqrt{2n_{\mathrm{eff}}}}. \tag{66}$$

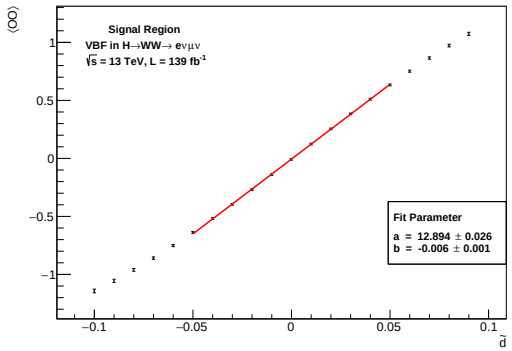
Figure 23 shows fits to the gauge curve for $\mathcal{O}\mathcal{O}_1$, $\mathcal{O}\mathcal{O}_{\mathrm{truth}}$, $\Delta\Phi_{jj}^{\mathrm{signed}}$ and $\mathcal{O}_{\mathrm{Reg}}$ in the range of $\tilde{d} = [-0.05, 0.05]$. The fits of the remaining observables are provided in Appendix D.



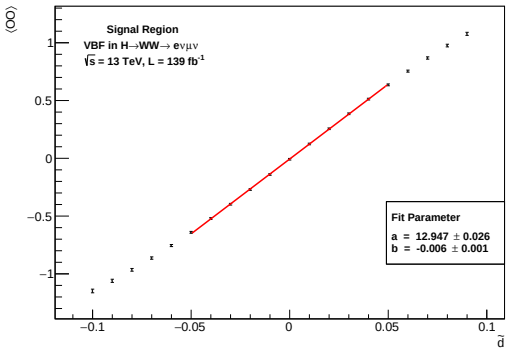
(a) $\mathcal{OO}_{\text{truth}}$



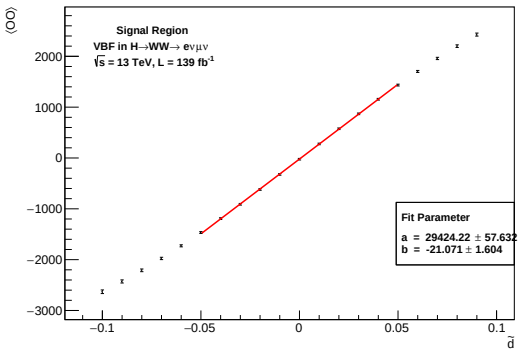
(b) $\mathcal{OO}_{\text{Eff}}$



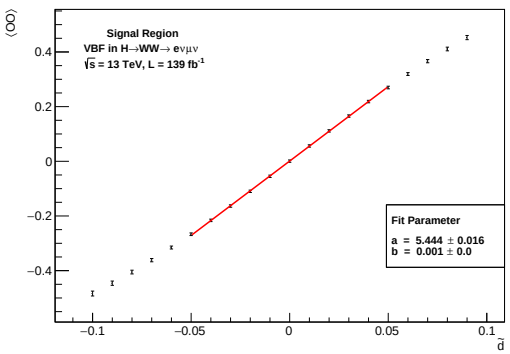
(c) $\mathcal{OO}_{1}$



(d) $\mathcal{OO}_{(6,9,10)}$



(e) $\Delta\Phi_{jj}^{\text{signed}}$



(f) $\mathcal{O}_{\text{Reg}}$

Figure 23: Gauge curves of ⟨CP-odd Obs.⟩ against $\tilde{d}$ using different CP-odd observables. The fit is performed in the $\tilde{d} = [-0.05, 0.05]$ range. The parameters of the fits are shown in the box in the bottom left corner of each figure.

The parameters used to calculate $\tilde{d}$ and their uncertainties are given in Table 15 together with the results. The last column shows the number of expected events (N_{exp}), which slightly differs between the observables.
All estimated values for $\tilde{d}$ are compatible with zero. However, it is unexpected that b is incompatible with zero. The error on the weighted mean of the CP-odd observables is

Observables	a	$b\,[10^{-2}]$	$\langle$CP-odd Variable$\rangle$	$\tilde{d}\,[10^{-5}]$	$\sigma_{\tilde{d}}\,[10^{-5}]$	$\Delta\sigma_{\tilde{d}}\,[10^{-5}]$	N_{exp}
$\mathcal{OO}_{\text{truth}}$	11.968 ± 0.024	-1.488 ± 0.066	-0.0111 ± 0.0064	3.1668	5.3832	0.0085	46.67
$\mathcal{O}_{\text{Reg}}$	29424.2 ± 57.6	-2107.1 ± 160.4	-23.65 ± 15.48	-0.8774	5.2893	0.0084	46.22
$\Delta\Phi_{jj}$	5.444 ± 0.016	0.10 ± 0.06	0.0006 ± 0.0042	-0.5863	7.8231	0.0124	46.22
$\mathcal{OO}_{\text{Eff.H}}$	12.783 ± 0.025	-0.965 ± 0.071	-0.0112 ± 0.0069	-1.2556	5.3968	0.0085	46.22
$\mathcal{OO}_1$	12.947 ± 0.026	-0.621 ± 0.072	-0.0080 ± 0.0070	-1.3851	5.4063	0.0085	46.21
$\mathcal{OO}_2$	12.911 ± 0.026	-0.584 ± 0.072	-0.0078 ± 0.0070	-1.5025	5.4198	0.0086	46.19
$\mathcal{OO}_3$	12.617 ± 0.024	-1.211 ± 0.066	-0.0130 ± 0.0064	-0.7059	5.0921	0.0081	46.66
$\mathcal{OO}_4$	12.579 ± 0.024	-1.458 ± 0.066	-0.0153 ± 0.0064	-0.6066	5.0885	0.0080	46.67
$\mathcal{OO}_{(5,9,10)}$	12.957 ± 0.026	-0.765 ± 0.072	-0.0094 ± 0.0070	-1.3541	5.4071	0.0085	46.20
$\mathcal{OO}_{(6,9,10)}$	12.894 ± 0.026	-0.609 ± 0.072	-0.0079 ± 0.0070	-1.4422	5.4078	0.0086	46.20
$\mathcal{OO}_{(7,9,10)}$	12.854 ± 0.026	-0.628 ± 0.071	-0.0081 ± 0.0069	-1.4177	5.3640	0.0085	46.27
$\mathcal{OO}_{(8,9,10)}$	12.782 ± 0.025	-0.735 ± 0.070	-0.0092 ± 0.0068	-1.4423	5.3561	0.0085	46.28

Table 15: List of fits parameters a and b, weighted mean of the CP-odd observables for $\tilde{d} = 0$, N_{exp} and estimation of $\tilde{d}$ including $\sigma_{\tilde{d}}$ and $\Delta\sigma_{\tilde{d}}$ for each observable.

the dominant contribution to the overall uncertainty. Due to Equation 64, larger slope values and smaller errors on the observable weighted mean yield a higher sensitivity for $\tilde{d}$.

The best result is achieved by $\mathcal{OO}_4$ constraining $\tilde{d}$ at a expected 68% *confidence level* (CL) to a confidence interval of $\tilde{d} \in [-5.089 \times 10^{-5}, 5.089 \times 10^{-5}]$. It improved by 0.07% compared to the second best performing observable $\mathcal{OO}_3$. The smallest sensitivity to $\tilde{d}$ was obtained from $\Delta\Phi_{jj}^{\text{signed}}$ with the confidence interval $\tilde{d} \in [-7.82 \times 10^{-5}, 7.82 \times 10^{-5}]$. This is expected since it is not as sensitive to CP-violation as $\mathcal{OO}$ and $\mathcal{O}_{\text{Reg}}$ [34, 83]. $\mathcal{O}_{\text{Reg}}$ limited $\tilde{d}$ with an expected CL of 68% to a confidence interval of $\tilde{d} \in [-5.289 \times 10^{-5}, 5.289 \times 10^{-5}]$. As such $\mathcal{OO}_4$ performed 3.79% better than $\mathcal{O}_{\text{Reg}}$. The $\mathcal{OO}$ with the worst sensitivity is $\mathcal{OO}_2$ performing 6.11% worse than $\mathcal{OO}_4$. The sensitivities of $\mathcal{OO}_{(5,9,10)}$, $\mathcal{OO}_{(6,9,10)}$, $\mathcal{OO}_1$, $\mathcal{OO}_{\text{Eff.H}}$ and $\mathcal{OO}_{\text{truth}}$ are all compatible with each other. The reconstruction-level observables are not expected to perform poorer than the truth-level $\mathcal{OO}_{\text{truth}}$. Still, all observables reconstructed with the networks using the modified loss function $\mathcal{OO}_3$, $\mathcal{OO}_4$, $\mathcal{OO}_{(7,9,10)}$ and $\mathcal{OO}_{(8,9,10)}$ outperformed $\mathcal{OO}_{\text{truth}}$. With $\mathcal{OO}_4$ gaining 5.47% on $\mathcal{OO}_{\text{truth}}$.

7 Conclusion

In this thesis, the sensitivity of several CP-odd observables to constrain $\tilde{d}$, which parameterises the strength of CP-violation in the HVV vertex, is investigated. The Higgs boson production via VFB with a subsequent decay $H \to WW^- \to e\mu 2\nu$ is considered. The simulated signal corresponds to the full Run 2 data-taking period by the ATLAS detector at a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ with an integrated luminosity of $139\,\text{fb}^{-1}$. The investigation exploits fully-connected feed-forward regression networks to reconstruct the Higgs boson four-momentum vector to construct the $\mathcal{OO}$.

The Higgs boson four-momentum vector is reconstructed with nine different methods. The first method approximates the Higgs bosons four-momentum vector from the lepton four-momenta and the missing transverse energy. This method is labelled as Eff.H. The other Higgs boson four-momentum vector reconstructions are done via neural networks, with the first four targeting the complete four-momentum vector of the Higgs bosons. The first network (NN_1) used the same input vectors as the Eff.H reconstruction and is trained with the mean-square-error loss function to predict all four-momentum vector components of the Higgs boson. The second network (NN_2) added the four-momentum vectors of the leading and subleading jets to the list of input variables. The third and fourth networks (NN_3 and NN_4) used a modified version of mean-square-error, constraining the reconstructed mass to the measured value of 125 GeV. The target and input features of NN_3 (NN_4) are same as that of NN_1 (NN_2). The last four reconstructions of the Higgs boson use multiple networks to predict different components of the Higgs boson four-momentum vector. These reconstruction methods are labelled $\text{NN}_{(5,9,10)}$, $\text{NN}_{(6,9,10)}$, $\text{NN}_{(7,9,10)}$ and $\text{NN}_{(8,9,10)}$ with the numbers referring to the networks used in each method. The common networks setups NN_9 and NN_{10} predict the p_x and p_y components, respectively. The other two components p_z and E were predicted simultaneasly with NN_5, NN_6, NN_7 and NN_8 which used the loss function and input features of NN_1, NN_2, NN_3 and NN_4, respectively.

The hyperparameters of each network are optimised with OPTUNA [55]. The optimised hyperparameters are the learning rate, the batch size, the number of layers and the number of nodes in each hidden layer. The networks are trained using the optimal set of hyperparameters and utilised to predict the Higgs boson four-momentum vector's components. Their resolutions are compared to the resolution of the four-momentum vector of the Eff.H reconstruction. In this comparison NN_9 (NN_{10}) improved the resolution of p_x (p_y) by 14.72% (23.11%). With NN_6, the resolutions for p_z and E could be improved by 22.51% and 40.30%, respectively. The networks NN_3 and NN_4, while improving M by 99.61%, performed worse on p_x, p_y, p_z and E components in comparison to the Eff.H reconstruction. The best performance is given by a combination of NN_6, NN_9 and NN_{10}. The E and p_z components predicted by NN_6 achieved a standard deviation of $\sigma_E = 88.24\,\text{GeV}$ and $\sigma_{p_z} = 103.25\,\text{GeV}$ while NN_9 and NN_{10} achieved a standard deviation of $\sigma_{p_x} = 16.92\,\text{GeV}$ and $\sigma_{p_y} = 16.64\,\text{GeV}$, respectively.

The $\mathcal{OO}$s are calculated for every reconstructed four-momentum vector of the Higgs

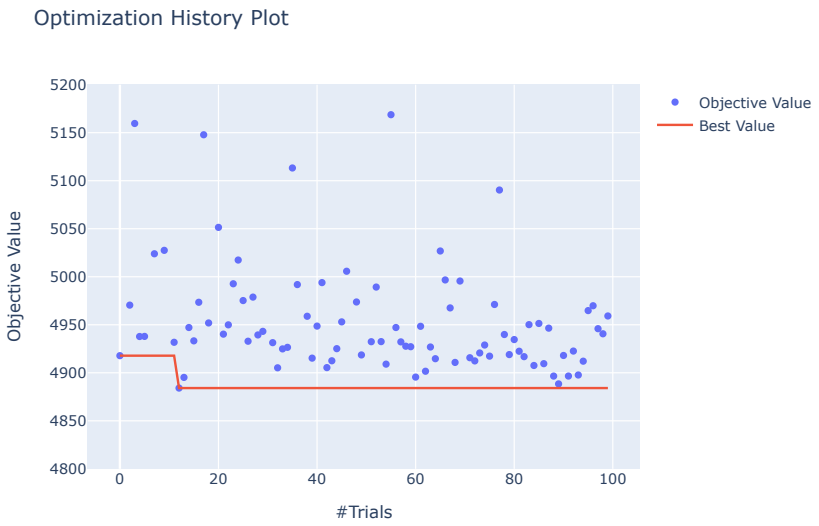
boson. Only events in the range of $-12 < \mathcal{OO} < 12$ were considered. Comparing the $\mathcal{OO}$ resolutions using Eff.H reconstruction to the ones using NNs, the largest improvement of 0.25% was achieved with NN_8.

A gauge curve method was utilised to estimate the sensitivity on $\tilde{d}$. The method assumes a linear regime between the mean of the CP-odd observable and $\tilde{d}$ for small values of $\tilde{d}$. Using a linear fit in the range of $\tilde{d} = [-0.05, 0.05]$, the sensitivity of $\tilde{d}$ was estimated. In addition to the reconstruction-level $\mathcal{OO}$ calculated with Eff.H and the NN predictions, the $\mathcal{OO}_{\mathrm{truth}}$ using truth-level information, the signed azimuthal angle difference between the tagging jets $\Delta\Phi_{jj}^{\mathrm{signed}}$ and the symbolic regression observable $\mathcal{O}_{\mathrm{Reg}}$ were considered.

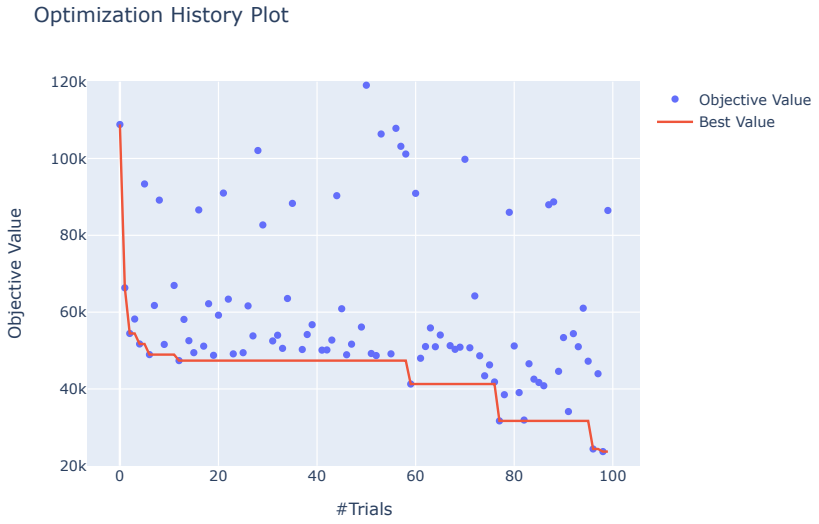
The best sensitivity for $\tilde{d}$ was achived by $\mathcal{OO}_4$ limiting $\tilde{d}$ to an interval of $\tilde{d} \in [-5.089 \times 10^{-5}, 5.089 \times 10^{-5}]$, outperforming the use of $\mathcal{O}_{\mathrm{Reg}}$ by 3.80% and the use of $\mathcal{OO}_{\mathrm{Eff.H}}$ of the Eff.H reconstruction by 5.71%. The smallest sensitivity to $\tilde{d}$ was obtained from $\Delta\Phi_{jj}^{\mathrm{signed}}$ with the expected 68% CL limiting $\tilde{d}$ to the confidence interval of $\tilde{d} \in [-7.82 \times 10^{-5}, 7.82 \times 10^{-5}]$. The constrains of $\tilde{d}$ for $\mathcal{OO}_{\mathrm{truth}}$ are compatible with the constrains achieved by $\mathcal{OO}_{\mathrm{Eff.H}}$. It is unclear why the $\mathcal{OO}$ on truth-level gets outperformed by reconstruction-level $\mathcal{OO}$ calculated by the predictions of networks using the modified MSE loss function.

Appendix

A Optimisation Process of NNs

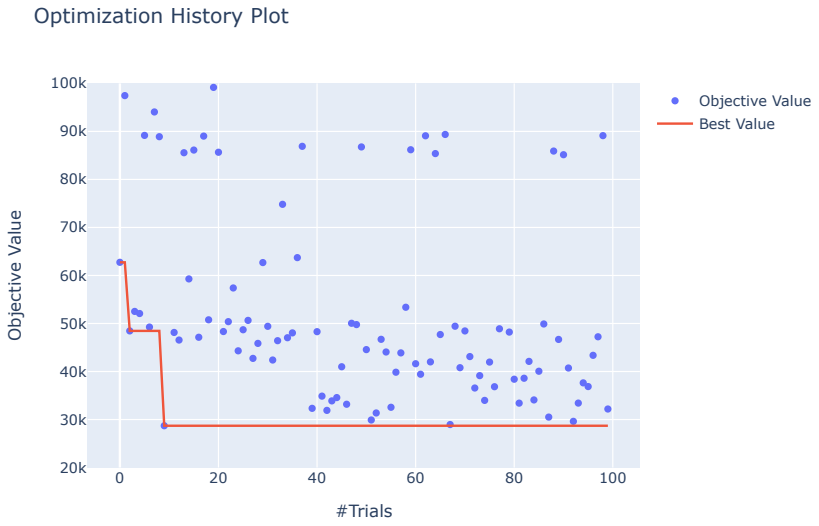


(a) NN_1

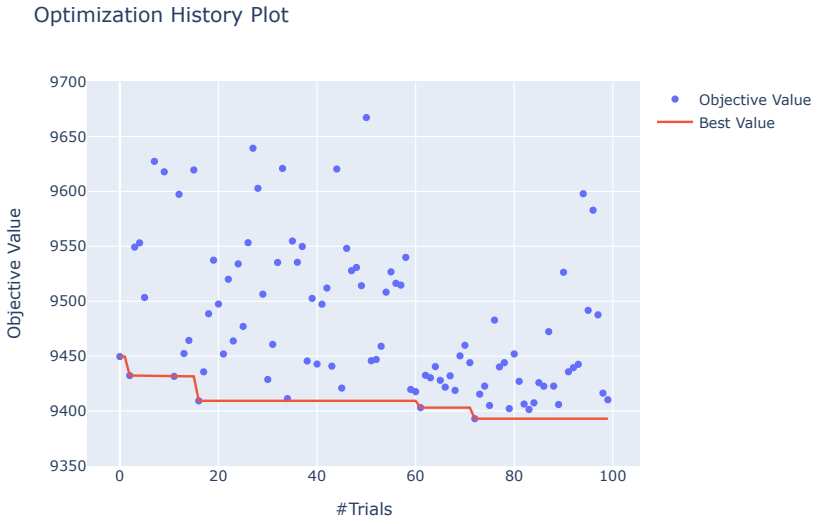


(b) NN_3

Figure 24: OPTUNA optimisation process of a) NN_1 and b) NN_3. The x-axis shows the number of trials (#Trials) and the y-axis the objective value corresponding to the network's loss function. Blue dots represent the objective value in each trial, while the red curve shows the best objective value. Some trial numbers are missing since their objective value is too large and outside the range.

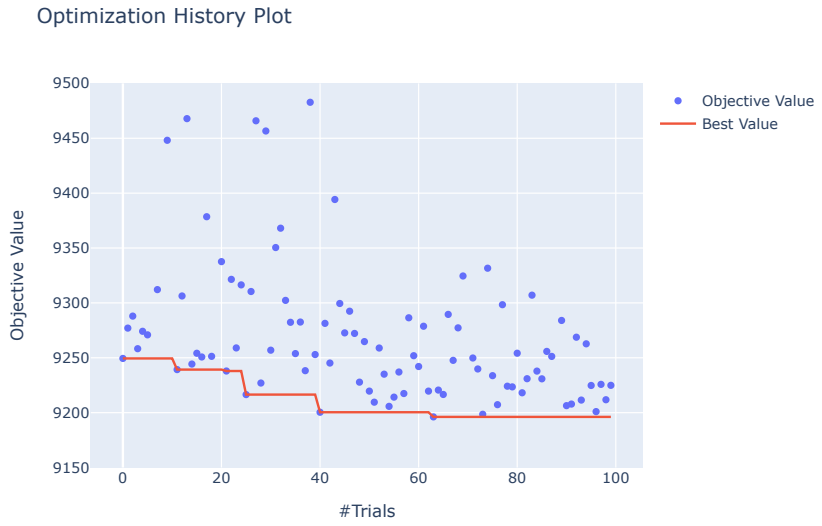


a) NN_4

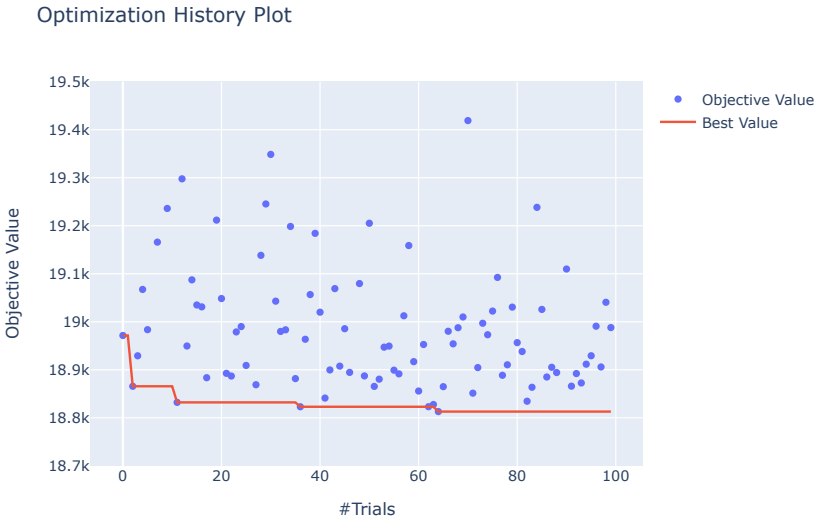


b) NN_5

Figure 25: OPTUNA optimisation process of a) NN_4 and b) NN_5. The x-axis shows the number of trials (#Trials) and the y-axis the objective value corresponding to the network's loss function. Blue dots represent the objective value in each trial, while the red curve shows the best objective value. Some trial numbers are missing since their objective value is too large and outside the range.

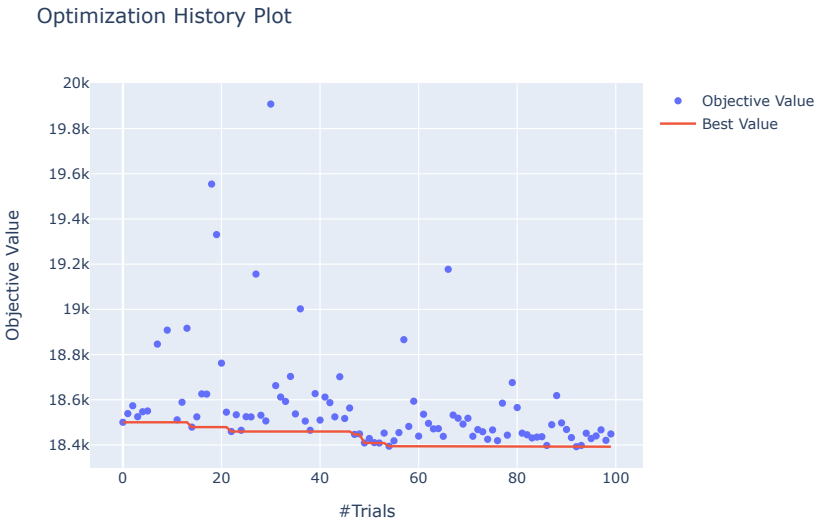


a) NN_6

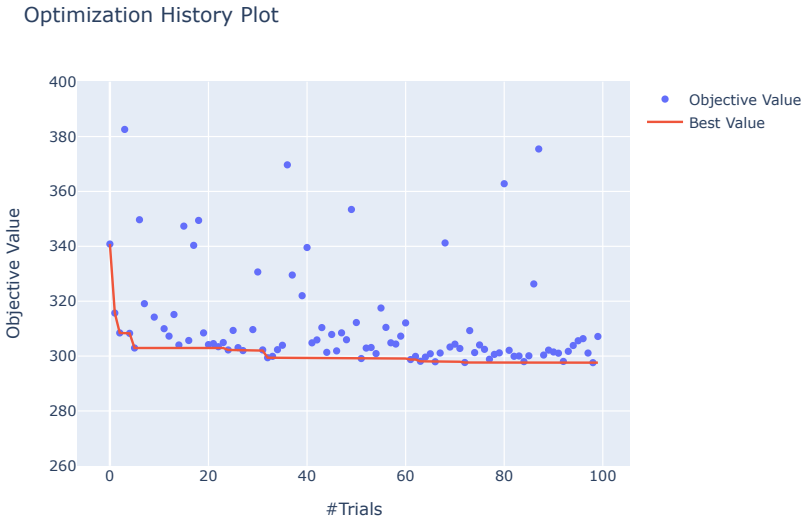


b) NN_7

Figure 26: OPTUNA optimisation process of a) NN_6 and b) NN_7. The x-axis shows the number of trials (#Trials) and the y-axis the objective value corresponding to the network's loss function. Blue dots represent the objective value in each trial, while the red curve shows the best objective value. Some trial numbers are missing since their objective value is too large and outside the range.



a) NN_8



b) NN_9

Figure 27: OPTUNA optimisation process of a) NN_8 and b) NN_9. The x-axis shows the number of trials (#Trials) and the y-axis the objective value corresponding to the network's loss function. Blue dots represent the objective value in each trial, while the red curve shows the best objective value. Some trial numbers are missing since their objective value is too large and outside the range.

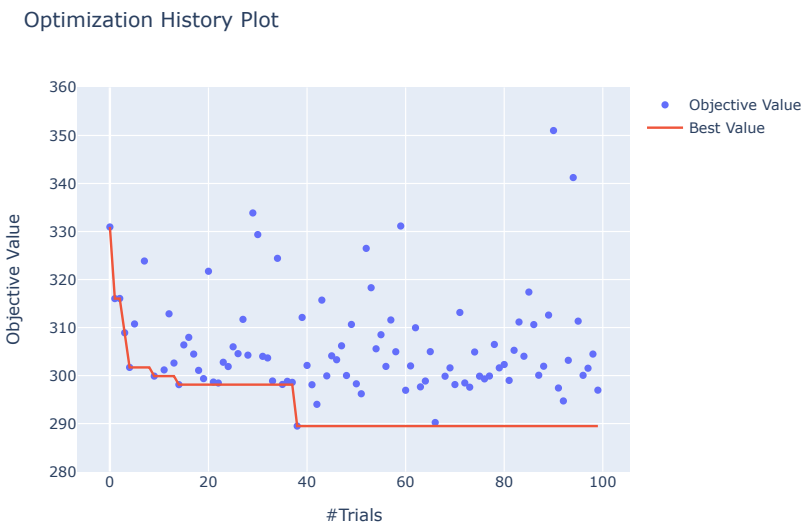
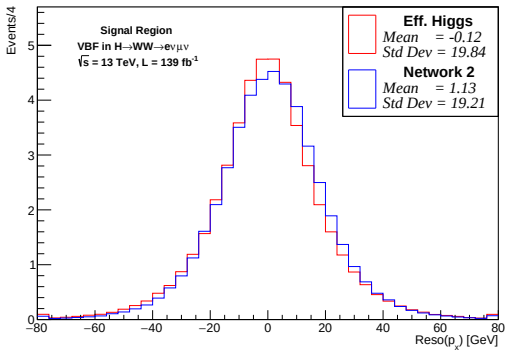


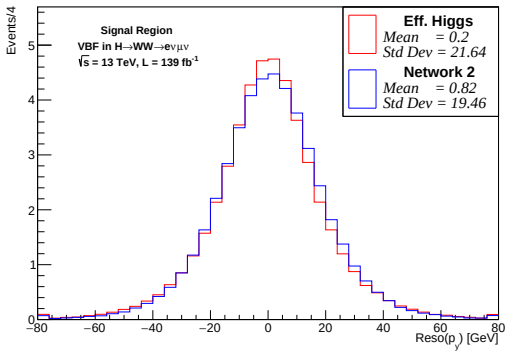
Figure 28: OPTUNA optimisation process NN_{10}. The x-axis shows the number of trials (#Trials) and the y-axis the objective value corresponding to the network's loss function. Blue dots represent the objective value in each trial, while the red curve shows the best objective value. Some trial numbers are missing since their objective value is too large and outside the range.

B Resolutions of Higgs Reconstructions

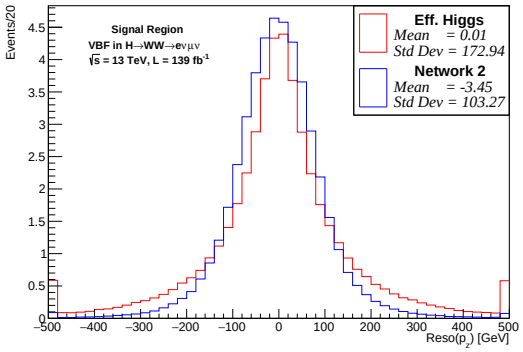
B.1 Network 2



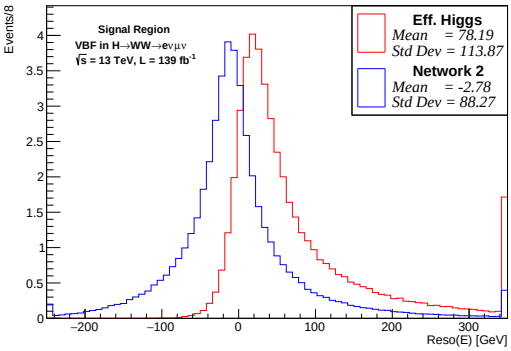
(a) Resolution of p_{x}



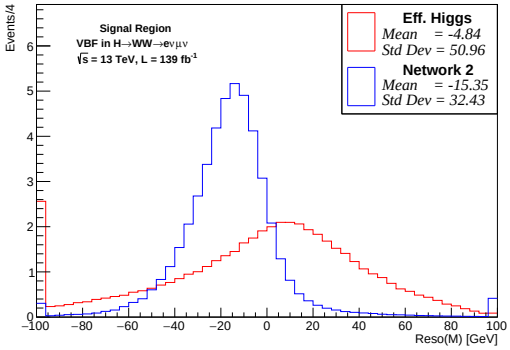
(b) Resolution of p_{y}



(c) Resolution of p_{z}



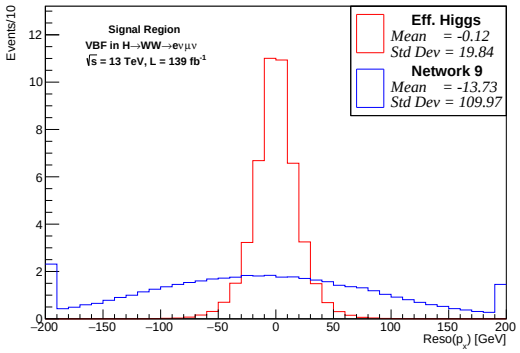
(d) Resolution of E



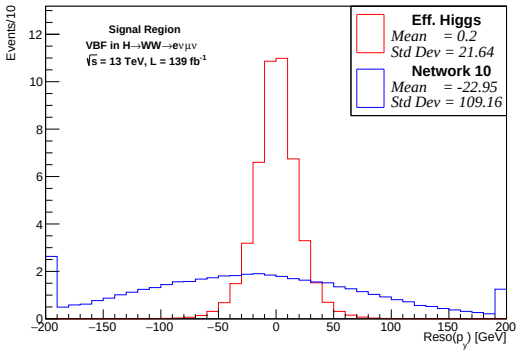
(e) Resolution of M

Figure 29: Resolutions of different components of the Higgs boson four-momentum vector reconstructed with NN_2 in blue and red the resolution of the Eff. H reconstruction components. All resolutions are normalised to 139fb^{-1}. Overflow and underflow bins are included.

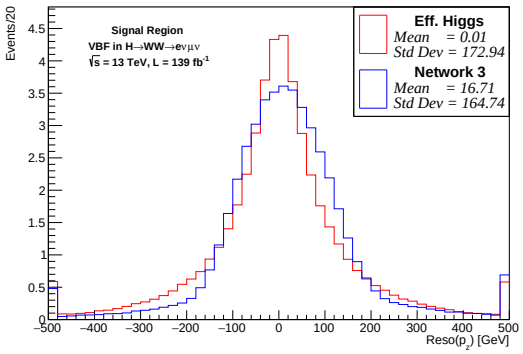
B.2 Network 3



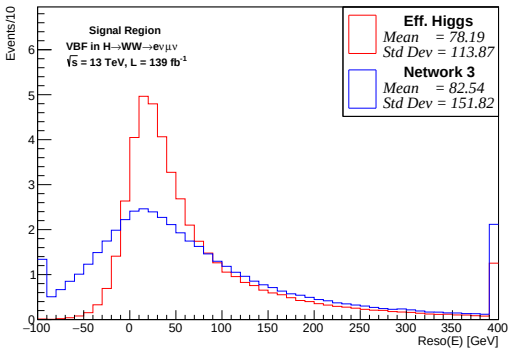
(a) Resolution of p_{x}



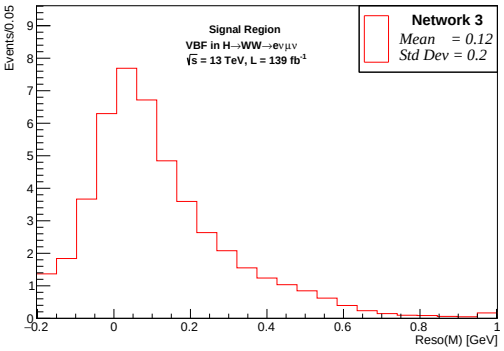
(b) Resolution of p_{y}



(c) Resolution of p_{z}



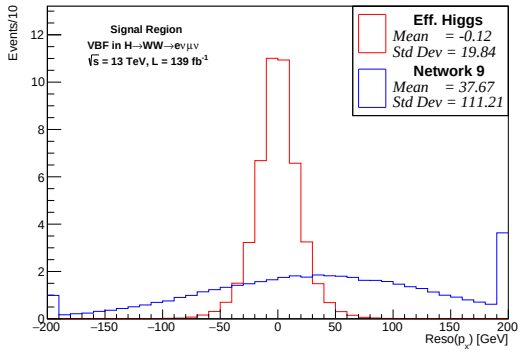
(d) Resolution of E



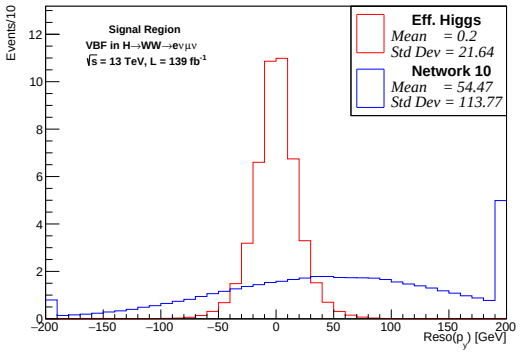
(e) Resolution of M

Figure 30: Resolutions of different components of the Higgs boson four-momentum vector reconstructed with NN_3 in blue and red the resolution of the Eff. H reconstruction components. All resolutions are normalised to 139fb^{-1}. Overflow and underflow bins are included.

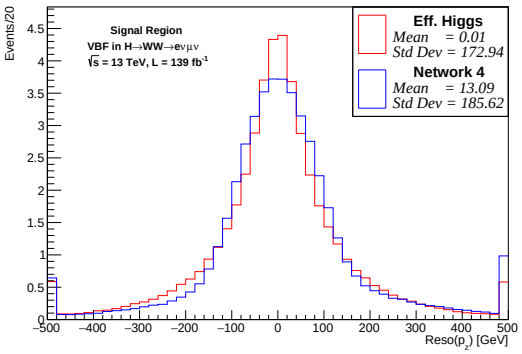
B.3 Network 4



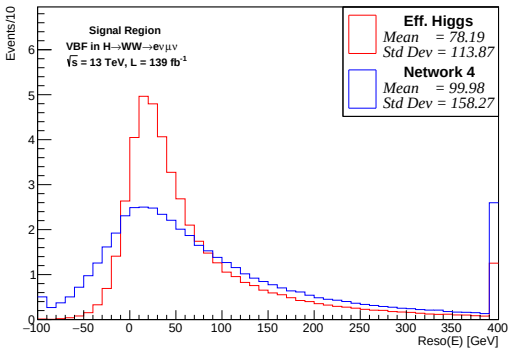
(a) Resolution of p_{x}



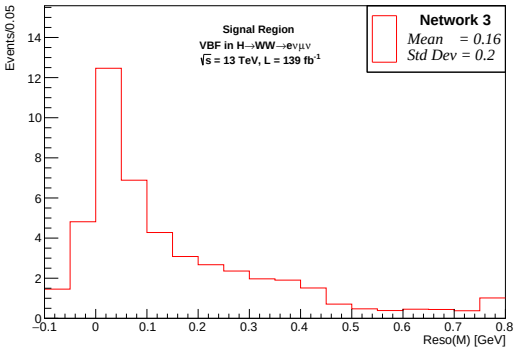
(b) Resolution of p_{y}



(c) Resolution of p_{z}



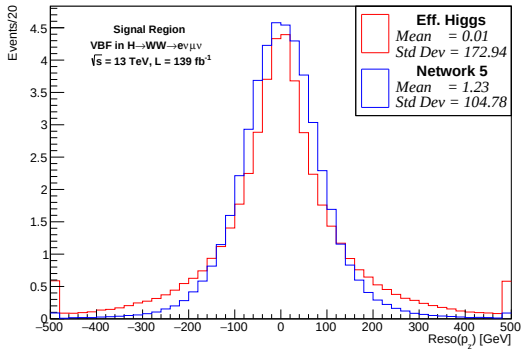
(d) Resolution of E



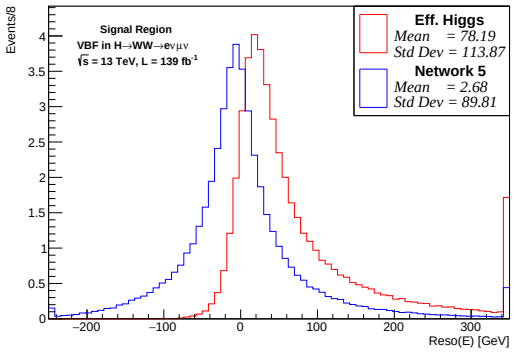
(e) Resolution of M

Figure 31: Resolutions of different components of the Higgs boson four-momentum vector reconstructed with NN_4 in blue and red the resolution of the Eff. H reconstruction components. All resolutions are normalised to 139fb^{-1}. Overflow and underflow bins are included.

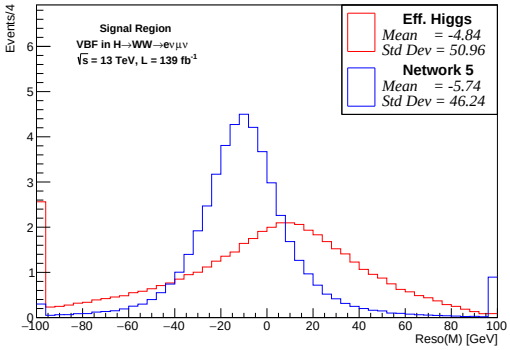
B.4 Network 5



(a) Resolution of p_z



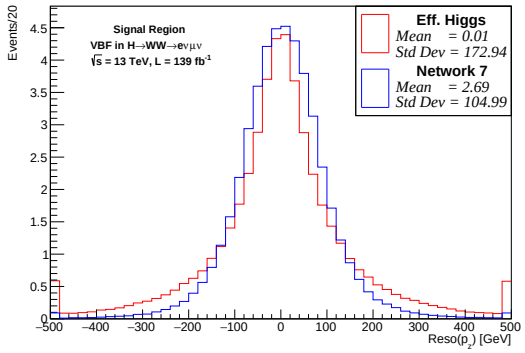
(b) Resolution of E



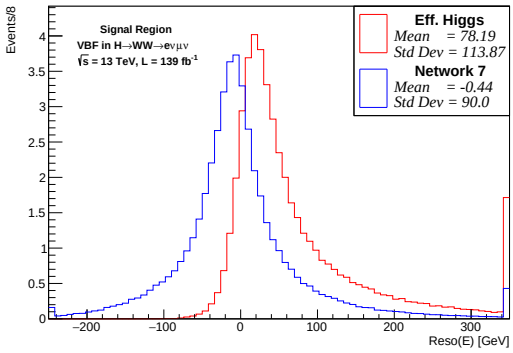
(c) Resolution of M

Figure 32: Resolutions of different components of the Higgs boson four-momentum vector reconstructed with NN_5 in blue and red the resolution of the Eff. H reconstruction components. All resolutions are normalised to 139fb^{-1}. Overflow and underflow bins are included.

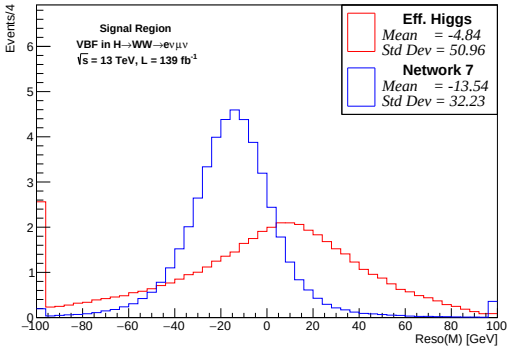
B.5 Network 7



(a) Resolution of p_z



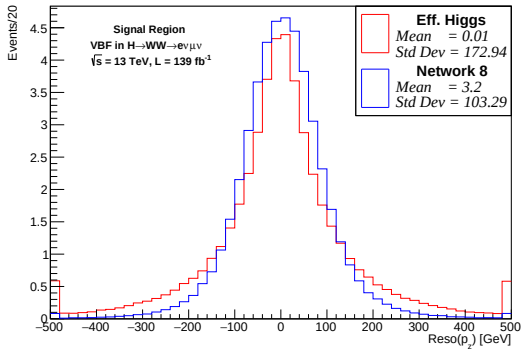
(b) Resolution of E



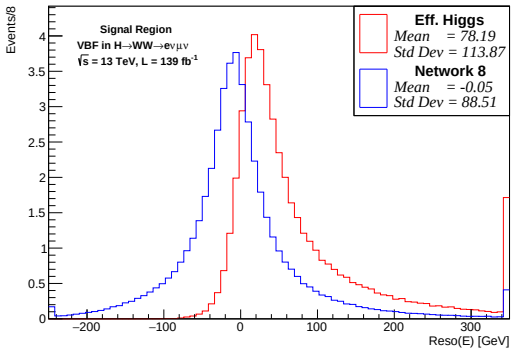
(c) Resolution of M

Figure 33: Resolutions of different components of the Higgs boson four-momentum vector reconstructed with NN_7 in blue and red the resolution of the Eff. H reconstruction components. All resolutions are normalised to 139fb^{-1}. Overflow and underflow bins are included.

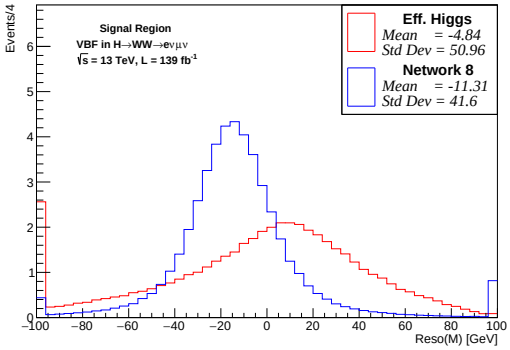
B.6 Network 8



(a) Resolution of p_z



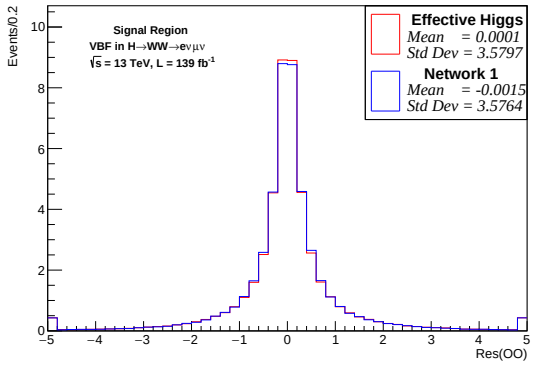
(b) Resolution of E



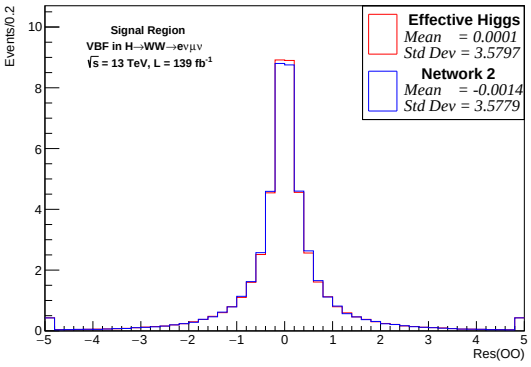
(c) Resolution of M

Figure 34: Resolutions of different components of the Higgs boson four-momentum vector reconstructed with NN_8 in blue and red the resolution of the Eff. H reconstruction components. All resolutions are normalised to 139fb^{-1}. Overflow and underflow bins are included.

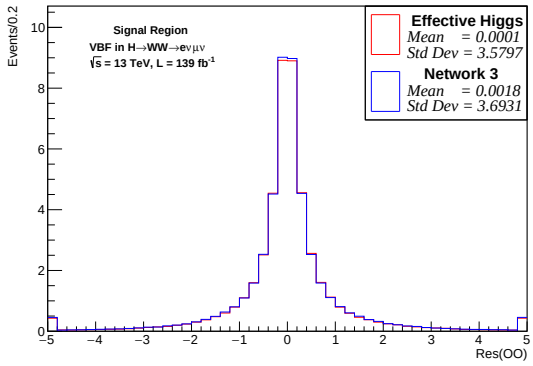
C $\mathcal{OO}$ Resolutions



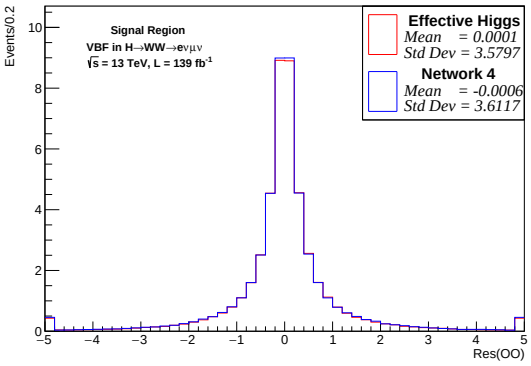
(a) Resolution of $\mathcal{OO}_1$



(b) Resolution of $\mathcal{OO}_2$

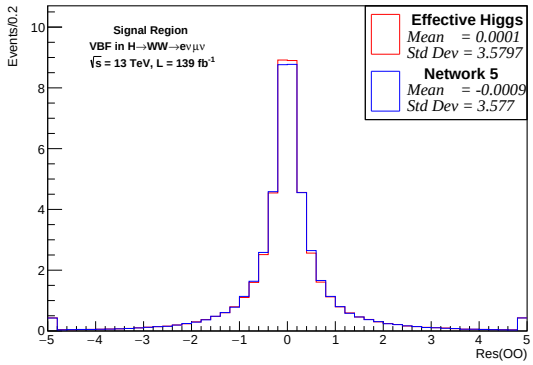


(c) Resolution of $\mathcal{OO}_3$

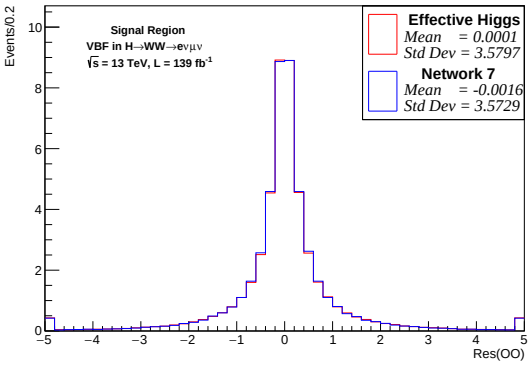


(d) Resolution of $\mathcal{OO}_4$

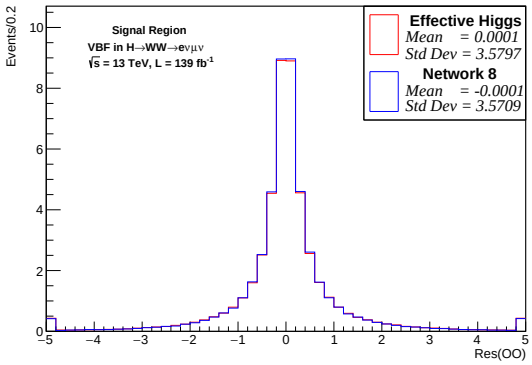
Figure 35: Resolutions of a) $\mathcal{OO}_1$, b) $\mathcal{OO}_2$, c) $\mathcal{OO}_3$ and $\mathcal{OO}_4$, all are compared to $\mathcal{OO}_{\text{Eff}}$. The reference value for the resolutions is the truth-level $\mathcal{OO}$. Resolutions are normalised to 139 fb^{-1}. Overflow and underflow bins are included.



(a) Resolution of $\mathcal{OO}_{(5,9,10)}$



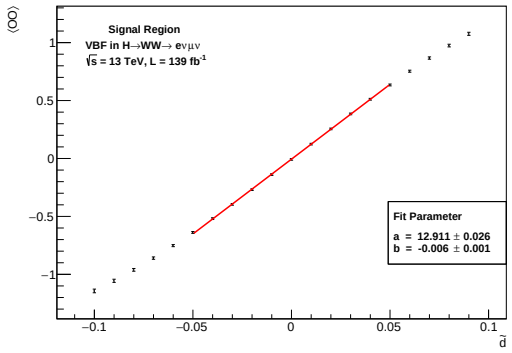
(b) Resolution of $\mathcal{OO}_{(7,9,10)}$



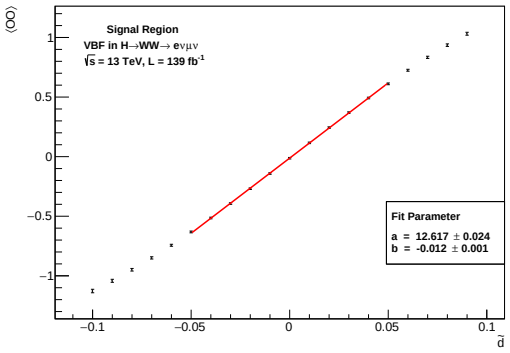
(c) Resolution of $\mathcal{OO}_{(8,9,10)}$

Figure 36: Resolutions of a) $\mathcal{OO}_{(5,9,10)}$, b) $\mathcal{OO}_{(7,9,10)}$ and c) $\mathcal{OO}_{(8,9,10)}$, all are compared to $\mathcal{OO}_{\text{Eff}}$. The reference value for the resolutions is the truth-level $\mathcal{OO}$. Resolutions are normalised to 139 fb^{-1}. Overflow and underflow bins are included.

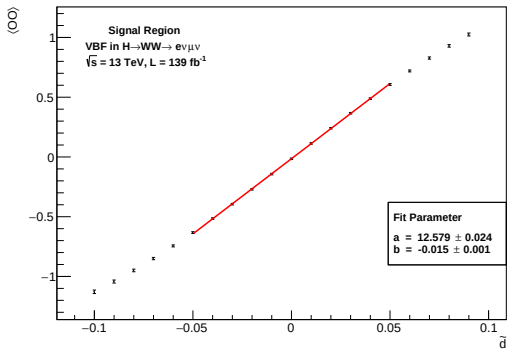
D CP-odd Observable Fits



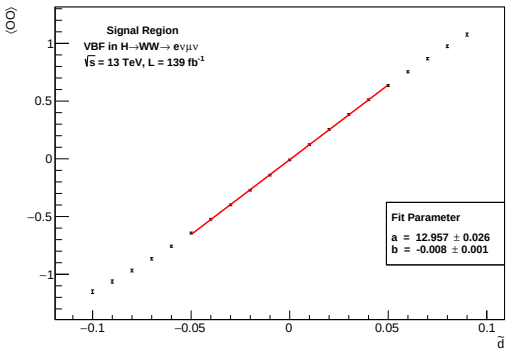
(a) $\mathcal{OO}_2$



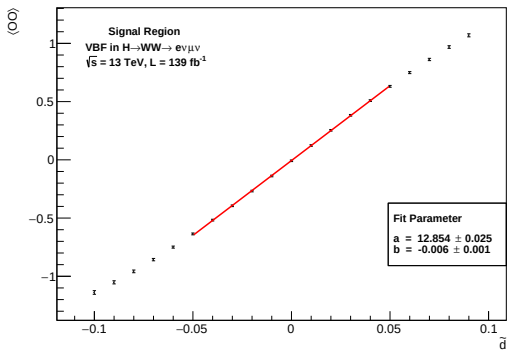
(b) $\mathcal{OO}_3$



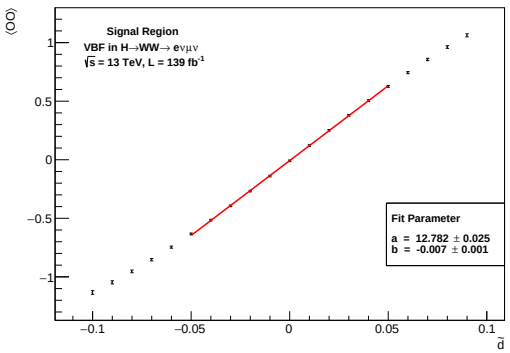
(c) $\mathcal{OO}_4$



(d) $\mathcal{OO}_{(5,9,10)}$



(e) $\mathcal{OO}_{(7,9,10)}$



(f) Fit for $\mathcal{OO}_{(8,9,10)}$

Figure 37: Gauge curves of ⟨CP-odd Variable⟩ against $\tilde{d}$ using different CP-odd observables. The fit is performed in the range of $\tilde{d} = [-0.05, 0.05]$. The parameters of the fits are shown in the box in the bottom left corner of each figure.

References

[1] S. L. Glashow, *Partial-symmetries of weak interactions*, Nucl. Phys. **22** (1961) 579.

[2] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264.

[3] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. C **680519** (1968) 367.

[4] 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. B **716** (2012) 1.

[5] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B **716** (2012) 30.

[6] CMS Collaboration, *Constraints on anomalous Higgs boson couplings to vector bosons and fermions in its production and decay using the four-lepton final state*, Phys. Rev.D **104** (2021).

[7] J. M. Diego et al., *A free-form lensing grid solution for A1689 with new multiple images*, Mon. Notices Royal Astron. Soc. **446** (2014) 683.

[8] P. J. E. Peebles and B. Ratra, *The cosmological constant and dark energy*, Rev. Mod. Phys. **75** (2003) 559.

[9] R. N. Mohapatra and G. Senjanović, *Neutrino Mass and Spontaneous Parity Nonconservation*, Phys. Rev. Lett. **44** (1980) 912.

[10] Planck Collaboration, *Planck 2018 results. VI. Cosmological parameters*, 2018, A&A **641** A6 (2020) 67.

[11] DØ Collaboration, *Observation of the Top Quark*, Phys. Rev. Lett. **74** (1995) 2632.

[12] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508.

[13] G. Guralnik, C. Hagen, and T. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. **13** (1964) 585.

[14] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321.

[15] J. Bleck-Neuhaus, *Elementare Teilchen*, Springer-Verlag Berlin Heidelberg (2013), ISBN: 978-3-642-32579-3.

[16] J. Goldstone, *Field theories with Superconductor solutions*, Il Nuovo Cim. **19** (1961) 154.

[17] Y. Nambu, *Quasi-Particles and Gauge Invariance in the Theory of Superconductivity*, Phys. Rev. **117** (1960) 648.

[18] R.L. Workman et al. (Particle Data Group), *Review of Particle Physics*, PTEP (2022) 083C01.

[19] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, CERN Yellow Reports: Monographs **Volume 2** (2017).

[20] E. Noether, *Invariante Variationsprobleme* GDZ Göttingen, Mathematisch-Physikalische Klasse (1918) 235.

[21] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, *Experimental Test of Parity Conservation in Beta Decay*, Phys. Rev. 105 (1957) 1413.

[22] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, *Evidence for the* 2π *Decay of the* K_2^0 *Meson* Phys. Rev. Lett. **13** (1964) 138.

[23] M. Kobayashi and T. Maskawa, *CP-Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652.

[24] A. D. Sakharov, *Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe*, JETP Lett. **5** (1967) 24.

[25] T. Figy, V. Hankele, G. Klamke, D. Zeppenfeld *Anomalous Higgs boson couplings in vector boson fusion at the CERN LHC* Phys. Rev. D **74** (2006) 095001.

[26] ATLAS Collaboration, *Test of CP invariance in vector-boson fusion production of the Higgs boson in the* $H \to \tau\tau$ *channel in proton–proton collisions at* $\sqrt{s} =$ 13 *TeV with the ATLAS detector*, Phys. Lett. B **805** (2020) 135426.

[27] ATLAS Collaboration, *Test of CP invariance in vector-boson fusion production of the Higgs boson using the Optimal Observable method in the ditau decay channel with the ATLAS detector*, Eur. Phys. J. C **76** (2016) 658.

[28] OPAL Collaboration, *Measurement of W boson polarisations and CP-violating triple gauge couplings from* W^+W^- *production at LEP*, Eur. Phys. J. C **19** (2001) 229.

[29] DELPHI Collaboration, *Study of W boson polarisations and Triple Gauge boson Couplings in the reaction* $e^+e^- \to W^+W^-$ *at LEP 2*, Eur. Phys. J. C **54** (2008) 345.

[30] ALEPH Collaboration, *Improved measurement of the triple gauge-boson couplings* γWW *and* ZWW *in* e^-e^+ *collisions*, Phys. Lett. B **614** (2005) 7.

[31] M. Diehl and O. Nachtmann, *Optimal observables for the measurement of three gauge boson couplings in* $e^-e^+ \to W^-W^+$, Z. Phys. C**62** (1994) 397.

[32] M. Diehl, O. Nachtmann, and F. Nagel, *Triple gauge couplings in polarized* $e^-e^+ \to W^-W^+$ *and their measurement using optimal observables*, Eur. Phys. J. C**27** (2003) 375.

[33] D. Atwood and A. Soni, *Analysis for magnetic moment and electric dipole moment form-factors of the top quark via* $e^-e^+ \to t\bar{t}$, Phys. Rev. D45 (1992) 2405.

[34] A. Butter, T. Plehn, N. Soybelman and J. Brehmer, *Back to the Formula – LHC Edition*, arXiv:2109.10414.

[35] W. Herr and B. Muratori, *Concept of luminosity*, CERN Yellow Reports: School Proceedings (2003) 361.

[36] ATLAS Collaboration, *Luminosity determination in pp collisions at* $\sqrt{s}$ = 13 *TeV using the ATLAS detector at the LHC*, arXiv:2212.09379.

[37] ATLAS Collaboration, *Luminosity determination in pp collisions at* $\sqrt{s}$ = 13 *TeV using the ATLAS detector at the LHC*, ATLAS-CONF-2019-021 (2012).

[38] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[39] D. V. Krasnopevtsev, *Tracking properties of the ATLAS Transition Radiation Tracker (TRT)*, J. Phys.: Conf. Ser. **798** (2017) 012150.

[40] F. Hügging, *The ATLAS Pixel Insertable B-layer (IBL)*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment Volume **650** (2011) 45.

[41] ATLAS Collaboration, *ATLAS pixel detector electronics and sensors*, JINST **3** (2008) P07007.

[42] J. Pequenao, *Computer Generated image of the ATLAS calorimeter*, 2008, https://cds.cern.ch/record/1095927.

[43] W. Buttinger, *The ATLAS Level-1 Trigger System*, J. Phys.: Conf. Ser. **396** (2012) 012010.

[44] P. Clarke, *The L2 trigger for the ATLAS detector at the LHC*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment Volume **368** (1995) 175.

[45] J. Pequenao, *Computer generated image of the ATLAS Muons subsystem*, 2008, https://cds.cern.ch/images/CERN-GE-0803017-01.

[46] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, MIT Press, 2016.

[47] P. Mehta, M. Bukov, C. H. Wang, A. G. R. Day, C. Richardson, C. K. Fisher and D. J. Schwab, *A high-bias, low-variance introduction to Machine Learning for physicists*, Phys. Report V.**810** (2019) 1.

[48] K. Hornik, M. Stinchcombe and H. White *Multilayer feedforward networks are universal approximators*, Neural Networks Vol.**2** (1989) 359.

[49] A. F. Agarap, *Deep learning using rectified linear units (ReLU)*, arXiv:1803.08375.

[50] S. Ruder, *An overview of gradient descent optimization algorithms*, arXiv:1609.04747.

[51] D. E. Rumelhart, G. E. Hinton and R. J. Williams, *Learning representations by back-propagating errors*, Nature **323** (1986) 533.

[52] F. Chollet et al., *Keras*, https://keras.io, 2015.

[53] M. A. et al., *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, https://www.tensorflow.org/ (2015).

[54] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, arXiv:1412.6980.

[55] T. Akiba, S. Sano, T. Yanase, T. Ohta and M. Koyama, *Optuna: A Next-generation Hyperparameter Optimization Framework*, arXiv:1907.10902.

[56] C. Wallace and M. Beck, *Exploring Multi-Objective Hyperparameter Optimization*, Clouderafastforwardlabs (blog), [accessed 9.February-2023].

[57] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015-2017 LHC proton-proton collision data*, JINST **14** (2019) P12006.

[58] ATLAS Collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at* $\sqrt{s} = 13$ *TeV*, Eur. Phy. J. C **81** (2020) 578.

[59] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-*k_t *jet clustering algorithm*, JHEP **04** (2008) 063.

[60] ATLAS Collaboration, *Jets reconstruction and performance using particle flow with the ATLAS Detector*, Eur. Phy. J. C **77** (2016) 581.

[61] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at* $\sqrt{s} = 8$ *TeV using the ATLAS detector*, Eur. Phys. J. C **76** (2016) 581.

[62] ATLAS Collaboration, *Identification and rejection of pile-up jets at high pseudorapidity with the ATLAS detector*, Er. Phys. J. C **77** (2017) 580.

[63] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with* $t\bar{t}$ *events in pp collisions at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **79** (2019) 970.

[64] ATLAS Collaboration, *Official Isolation Working Points*, 2022, https://twiki.cern.ch/twiki/bin/view/Sandbox/MatthewKleinSandbox.

[65] ATLAS Collaboration, *Data-Quality Requirements and Event Cleaning for Jets and Missing Transverse Energy Reconstruction with the ATLAS Detector in Proton-Proton Collisions at a Center-of-Mass Energy of*$\sqrt{s} = 7$ *TeV*, tech. rep. CERN, 2010.

[66] ATLAS Collaboration, *Measurements of Higgs boson production by gluon-gluon fusion and vector-boson fusion using $H \to WW \to e\nu_e\mu\nu_\mu$ decays in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, arXiv:2207.00338.

[67] ATLAS Collaboration, *Measurements of Higgs boson production cross-sections in the $H \to \tau^-\tau^+$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **08** (2022) 175.

[68] P. Nason and C. Oleari, *NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG*, JHEP **02** (2010) 37.

[69] P. Nason, *A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms*, JHEP **11** (2004) 40.

[70] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, JHEP **11** (2007) 70.

[71] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 43.

[72] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159.

[73] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at $\sqrt{s} =$ 13 TeV*, Eur. Phys. J. C **79** (2019) 639.

[74] TLAS Collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **81** (2021) 578.

[75] Dr. Michael Böhler, *Software package: vbfcprw*, Private communication, 2023, [Online; accessed 4-February-2023].

[76] A. Denner, S. Dittmaier, S. Kallweit and A. Mück, *HAWK 2.0: A Monte Carlo program for Higgs production in vector-boson fusion and Higgs strahlung at hadron colliders*, Comput.Phys.Commun. **195** (2015) 161.

[77] M. Ciccolini, A. Denner and S. Dittmaier, *Strong and electroweak corrections to the production of Higgs+2 jets via weak interactions at the LHC*, Phys. Rev. Lett. **99** (2007) 161803.

[78] M. Ciccolini, A. Denner and S. Dittmaier, *Electroweak and QCD corrections to Higgs production via vector-boson fusion at the LHC*, Phys. Rev. D **77** (2008) 013002.

[79] A. Denner, S. Dittmaier, S. Kallweit and A. Mück, *Electroweak corrections to Higgs-strahlung off W/Z bosons at the Tevatron and the LHC with HAWK*, JHEP **75** (2012).

[80] A. Denner, S. Dittmaier and J. N. Lang, *Renormalization of mixing angles*, JHEP **104** (2018).

[81] G. Cowan, *Statistical data analysis*, Oxford University Press Inc 1998, ISBN 0-19-8501560.

[82] D. Atwood and A. Soni, *Analysis for magnetic moment and electric dipole moment form factors of the top quark via* $e^+e^- \to tt^-$, Phys. Rev. D **45** (1992) 2405.

[83] ATLAS Collaboration, *Test of CP Invariance in vector-boson fusion production of the Higgs boson using the Optimal Observable method in the ditau decay channel with the ATLAS detector*, Eur. Phys. J. C**76** (2016) 658.

Acknowledgements

I would like to take this opportunity to thank all who supported me in the three months of creating my Bachelor thesis. I enjoyed the time a lot while learning a lot about various subjects.

First and foremost, thanks go to Prof. Dr. Markus Schumacher. Not only for allowing me to work on my thesis in such an interesting field of science but also for the time he spent supporting me.

Special thanks go to my supervisor Oğul Öncel. His support, advice and efforts in helping me along were indispensable.

I would also like to thank Daniel Bahner for his help and explanations. I enjoyed our talks during the lunch breaks immensely. Thanks also to Michael Boehler and Benjamin Rottler for the help regarding my thesis's programming and machine learning parts.

Finally, a huge thank you to the research group in general. I am very grateful for the friendly environment where I was allowed to research and write my thesis.