Sensitivity and Optimization Studies for Testing CP-Invariance of Higgs-Boson Production in Vector-Boson Fusion Exploiting the Decay $H \to \tau^+\tau^- \to e\mu 4\nu$ at $\sqrt{s} = 13$ TeV with the ATLAS Detector

A Thesis Presented for the Degree of
Master of Science in Physics

Ye Joon Kim

Supervised by

Prof. Dr. Markus Schumacher

Institute of Physics
Albert-Ludwigs-Universität Freiburg
16.11.2022

Declaration of Originality

Erklärung gemäß der geltenden Prüfungsordnung:

Hiermit versichere ich, dass

1. ich die eingereichte Masterarbeit bzw. bei einer Gruppenarbeit meinen entsprechend gekennzeichneten Anteil der Arbeit selbständig verfasst habe,
2. ich keine anderen als die angegebenen Quellen und Hilfsmittel benutzt und alle wörtlich oder sinngemäß aus anderen Werken übernommenen Inhalte als solche kenntlich gemacht habe und
3. die eingereichte Masterarbeit weder vollständig noch in wesentlichen Teilen Gegenstand eines anderen Prüfungsverfahrens war oder ist.

Ort, Datum

Unterschrift

Auszug Prüfungsordnung M. Sc.

Vom 19. August 2005 (Amtliche Bekanntmachungen Jg. 36, Nr. 46, S.269–293) in der Fassung vom 25. September 2020 (Amtliche Bekanntmachungen Jg. 51, Nr. 66, S. 328–337).

§20 (8): Bei der Einreichung hat der/die Studierende schriftlich zu versichern, dass

1. er/sie die eingereichte Masterarbeit beziehungsweise bei einer Gruppenarbeit seinen/ihren entsprechend gekennzeichneten Anteil der Arbeit selbständig verfasst hat,
2. er/sie keine anderen als die angegebenen Quellen und Hilfsmittel benutzt und alle wörtlich oder sinngemäß aus anderen Werken übernommenen Inhalte als solche kenntlich gemacht hat und
3. die eingereichte Masterarbeit weder vollständig noch in wesentlichen Teilen Gegenstand eines anderen Prüfungsverfahrens war oder ist.

Abstract

CP symmetry, which is the combination of charge and parity symmetry, was once thought to be a fundamental symmetry of nature. However, experiments found CP symmetry is violated. The amount of CP-violation described by the SM is not enough to explain the amount of baryon asymmetry in the universe. Therefore, experimental searches of further sources of CP-violation in other sectors are necessary, which could be provided by the coupling of the Higgs-boson to weak gauge bosons. The strength of CP-violation is described by a single parameter $\tilde{d}$. The expected uncertainty on $\tilde{d}$ is determined using two different statistical methods using vector-boson fusion produced Higgs-bosons in the $H \to \tau^+\tau^- \to e\mu 4\nu$ channel. A selection based on neural networks is developed to maximize the sensitivity. Three different CP-odd observables were compared. The expected sensitivity for $\tilde{d}$ using the full Run-2 dataset is $\tilde{d} \in [-0.017, 0.017]$ for $\mathcal{OO}$ and $\mathcal{O}^{Reg}$, and $[-0.025, 0.025]$ for $\Delta\phi_{jj}^{signed}$ using the gauge curve method. For the ML method, this was $\tilde{d} \in [-0.0310, 0.0315]$ for $\mathcal{OO}$, $[-0.0310, 0.305]$ for $\mathcal{O}^{Reg}$, and $[-0.0380, 0.0385]$ for $\Delta\phi_{jj}^{signed}$.

Zusammenfassung

CP-Symmetrie ist die Kombination von Ladungskonjugation und Parität und galt einst als eine fundamentale Symmetrie der Natur. Es wurde aber experimentell bestätigt, dass die CP-Symmetrie verletzt ist. Die Menge an CP-Verletzung in dem SM reicht nicht aus, um die Asymmetrie zwischen Baryonen und Antibaryonen in diesem Universum zu erklären. Deswegen ist es notwendig, eine experimentelle Suche nach weitere Quellen von CP-Verletzung in anderen Sektoren durchzuführen, die in der Kopplung zwischen dem Higgs-Boson und zwei schwachen Eichbosonen vorhanden sein könnten. Die Stärke der CP-Verletzung wird durch einen einzigen Parameter $\tilde{d}$ beschrieben. Die erwarteten Grenzen auf $\tilde{d}$ werden durch zwei statistische Methoden mit VBF-produzierten Higgs-Bosonen in dem $H \to \tau^+\tau^- \to e\mu 4\nu$ Zerfallskanal ermittelt. Eine Selektion durch neuronale Netzwerke wurde entwickelt, um die Sensitivität zu optimieren. Drei verschiedene CP-ungerade Observablen wurden verglichen. Die erwarteten Grenzen an $\tilde{d}$ mit dem vollen Run-2 Datensatz sind $\tilde{d} \in [-0.017, 0.017]$ für $\mathcal{OO}$ und $\mathcal{O}^{Reg}$, und $[-0.025, 0.025]$ für $\Delta\phi_{jj}^{signed}$ mit der Eichkurve-Methode. Mit der ML-Methode, sind die Grenzen auf $\tilde{d}$: $\tilde{d} \in [-0.0310, 0.0315]$ für $\mathcal{OO}$, $[-0.0310, 0.305]$ für $\mathcal{O}^{Reg}$, und $[-0.0380, 0.0385]$ für $\Delta\phi_{jj}^{signed}$.

Contents

1 Introduction

The Standard Model (SM) of particle physics describes elementary particles and the interactions between them. The SM describes the strong, weak nuclear forces and the electromagnetic force. Particles that are described by the SM can be classified into two categories, the fermions with half-integer spin and bosons with integer spin. [56, 84, 90].

Experimental measurements showed that gauge bosons, the W and Z bosons, posses a mass of $m_W = 80.4$ GeV and $m_Z = 91.2$ GeV, respectively [78]. This required an augmentation to the SM with mass terms for the W and Z bosons, which is local gauge invariant. A solution to this problem was the Higgs mechanism, which allowed for mass terms for gauge bosons and fermions while maintaining local gauge invariance [52, 58, 62]. The Higgs mechanism utilizes spontaneous symmetry breaking and introduces the Higgs field. A prediction of the Higgs mechanism is the existence of a spin-0 Higgs boson.

In 2012, both the ATLAS and CMS experiments announced the discovery of a particle with a mass of approximately 125 GeV, which had properties consistent with that of the predicted Higgs boson in the SM [21, 44].

An important property of the Higgs boson is its CP properties. CP symmetry, the combination of charge (C) and parity (P) symmetry is broken in the weak interaction but conserved in the strong and electromagnetic interactions in the SM. This violation of CP symmetry is encoded in the Cabibbo-Kobayashi-Maskawa (CKM) matrix. However, the amount of CP violation present in the SM is not enough to explain the observed baryon asymmetry in the Universe, which is one of the Sakharov conditions for baryon asymmetry [83]. Therefore experimental searches for further sources of CP violation in other sectors of the SM are necessary.

The investigation of the CP properties of the Higgs coupling to two vector bosons (HVV vertex) provides an advantage due to its prevalence in pp collisions in the ATLAS detector. One method in the investigation of CP properties of the HVV coupling is the use of the Optimal Observable method, which is shown to have the largest sensitivity to CP-violating contributions in the Higgs-sector. An analysis on the CP properties of the HVV coupling in the $H \to \tau\tau$ channel by the ATLAS and CMS collaborations showed no significant deviation from SM predictions [28, 41].

This thesis focuses on the CP properties of the HVV coupling using vector boson fusion produced Higgs bosons exploiting the $H \to \tau\tau \to e\mu 4\nu$ decay channel using a dataset of 139 fb^{-1} of recorded using the ATLAS experiment. Using neural networks, signal events are selected from background events. The settings on the neural networks are optimized to maximize the sensitivity using Bayesian hyperparameter optimization. From the distribution of different CP-odd variables, expected limits on $\tilde{d}$ are derived, which is a parameter regulating the strength of CP-violation. One method uses the mean value of the CP odd observable and another is a maximum likelihood fit to the full distribution. The performance of these two statistical methods are optimized and compared by applying them onto an Asimov dataset. The performance of different CP-odd variables such as the $\mathcal{OO}$, $\mathcal{O}^{Reg}$ and $\Delta\phi_{jj}^{signed}$ are compared in addition.

This thesis is structured in the following way: In the second Chapter, the standard model is introduced and CP-odd observables are discussed. In Chapter 3, the LHC and the ATLAS experiment are described. The reconstruction and identification of physics objects in the ATLAS experiment are described in Chapter 4. In Chapter 5, the signal and background processes and defined and described, followed by Chapter 6, which details the estimation of the contribution of background process with jets misidentified as leptons. Chapter 7 describes the event selection requirements that are applied onto the dataset. Chapter 8 describes the theoretical aspects of neural networks and Chapter 9 the optimization of a neural network in discriminating signal from background events. This is followed by Chapter 10, which describes the statistical procedures and their applications in extracting limits on $\tilde{d}$. Lastly, in Chapter 11, the results are summarized and an outlook for further studies are discussed.

2 Theoretical background

2.1 The Standard Model

The Standard Model (SM) of particle physics describes three of the four fundamental forces of nature: electromagnetism, strong force and weak force. It also describes elementary particles and interactions between them. Particles are classified into two categories: *leptons* with half-integer spin and *bosons* with integer spin. These particles can furthermore be classified into the following categories: quarks, leptons, gauge bosons, and the Higgs boson. Quarks constitute hadrons, while leptons do not interact strongly. Gauge bosons act as the intermediary of fundamental forces. Leptons and quarks are each composed of three *generations* defined according the particles' flavor and mass. An overview of the particles and their properties in the SM is shown in Figure 1.

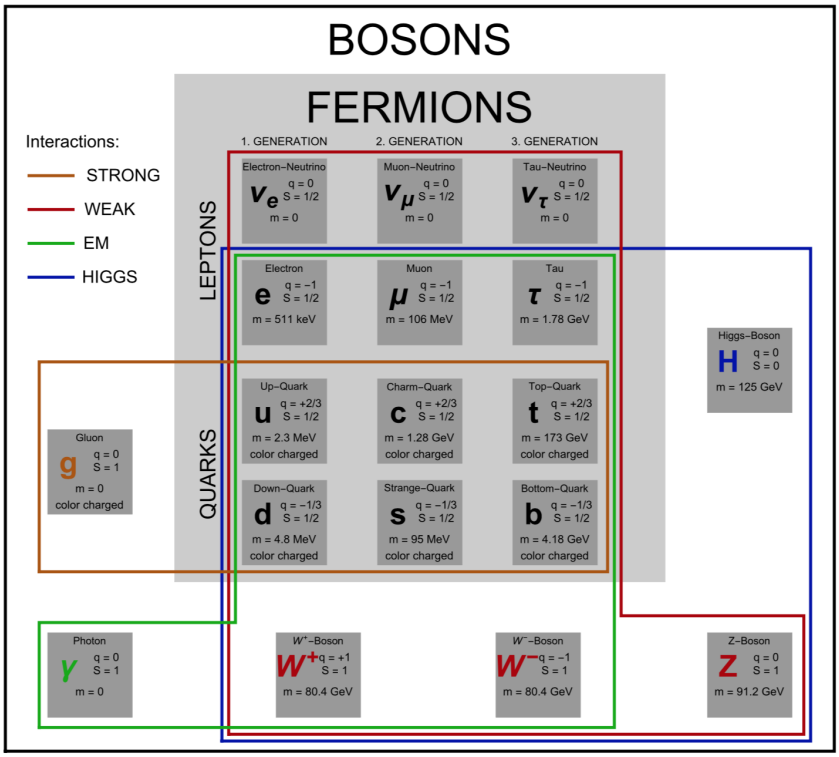


Figure 1: An overview of the particles described by the Standard Model as well as their properties such as mass, charge and spin. Colored frames show which particles are affected by which interactions [77].

The first generation of quarks include the *up* and *down* quark, which comprise the valence quark content of protons and neutrons. The second generation of quarks are the *charm* and *strange* quark and the third generation the *top* and *bottom* quark. Quarks carry a color charge which can have three values: red, blue or green, as well as respective anticolors in the case of quark antiparticles, such that hadrons are overall color neutral (white).

In the case of leptons, the first generation encompasses the electron and electron neutrino, the second generation the muon and muon neutrino and the third generation the tau and

tau neutrino, classified in order of increasing mass of the charged lepton. Neutrinos are, on the other hand, electrically neutral. For each fermion, an antiparticle exists with the same mass but with opposite charge and quantum numbers.

Bosons mediate the fundamental forces between elementary particles and have integer spins. Apart from the Higgs boson, which has a spin of 0, all bosons have a spin of 1. The photon is massless and mediates the electromagnetic force and couples to particles possessing electric charge. W and Z bosons mediate the weak nuclear force and couple to the particles possessing weak isospin. Gluons, which can have 8 different color charge configurations, are massless, and couples to particles possessing color charge. These bosons are collectively called *gauge bosons*. Lastly, the Higgs boson couples to all massive particles, and is detailed in section 2.1.3.

The Standard Model is a relativistic quantum field theory (QFT). Specifically, it is a locally gauge-invariant QFT with the group structure $SU(3)_C \times SU(2)_{I_W} \times U(1)_{Y_W}$, where $SU(3)_C$ is the symmetry group of quantum chromodynamics, representing color charge, $SU(2)_{I_W}$ and $U(1)_{Y_W}$ the symmetry group of the electroweak interaction representing weak isospin and hypercharge respectively. In QFTs the Langrangian describes the dynamics, kinematics and interactions of the theory, which in the case of the Standard Model is also locally gauge invariant. Such a Lagragian is constructed first by defining the symmetry structure of the theory and finding the most general renormalizable Lagrangian for all of its particles, which is not necessarily trivial the Lagrangian is to be invariant under local gauge transformations. A "naive" attempt at an inclusion of mass terms for gauge bosons results in a non-renormalizable theory, which warranted the development of the *Higgs mechanism* to consistently describe masses of gauge bosons and fermions while maintaining local gauge invariance.

2.1.1 Quantum Chromodynamics

Quantum Chromodynamics (QCD) is the theory of the strong interaction and is a non-abelian gauge theory with the gauge structure $SU(3)_C$, which is the generator of the color charge C. Quarks carry a color charge of either red, blue or green, while gluons carry a color and anti-color charge. Although 9 different color combinations are expected, gluons constitute a color-octet, and therefore the $SU(3)$ symmetry group is used to describe QCD. As with the electric charge, the color charge is conserved at each QCD coupling vertex. In addition, the self-coupling of the gluon exists in QCD, contrasting with the photon which has no self coupling since it is charge neutral.

The generators of $SU(3)$ are the *Gell-Mann λ-matrices*, whose commutators define the structure constants f^{abc} of the $SU(3)$ group:

$$[\lambda^a, \lambda^b] = 2if^{abc}\lambda^c \tag{1}$$

The gauge invariant Lagrangian of QCD is:

$$\mathcal{L}_{\text{QCD}} = \bar{\psi}_i(i\gamma^\mu(D\mu)_{ij} - m\delta_{ij})\psi_j - \frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a, \tag{2}$$

where upper and lower indices follow the Einstein summation convention [49]. $\psi_i(x)$ is the quark field, with $i = 1, 2, 3$ representing the color indices of the $SU(3)$ group. D_μ is the covariant derivative and γ^μ the Dirac matrices. $G^a_{\mu\nu}$ represents the gauge-invariant gluon field-strength-tensor, given by $G^a_{\mu\nu} = \partial_\mu \mathcal{A}^a_\nu - \partial_\nu \mathcal{A}^a_\mu + g f^{abc} \mathcal{A}^b_\mu \mathcal{A}^c_\nu$, where $\mathcal{A}^a_\nu$ are the gluon fields, with $a = 1 \dots 8$.

One important aspect of QCD is *color confinement*. The force between the color charges of two quarks remain constant regardless of their distance, and therefore when they are separated the increased energy enables the production quark-antiquark pairs such that the color charges are never isolated. An important consequence of this is the process of *hadronization*, where a scattered quark produces a collimated bunch of hadrons through this process, known as a *jet*. Another consequence of this is that observed composite particles are always color neutral.

2.1.2 Electroweak Interaction

The electroweak interaction is a unified description of electromagnetism and the weak force. Electromagnetism is described by quantum electrodynamics (QED) which has a gauge structure of $U(1)_Q$ with the electric charge Q. A unification of electromagnetic and weak interaction results in a theory with the gauge structure $SU(2)_{I_W} \times U(1)_{Y_W}$. Two new associated parameters are introduced: the weak isospin I_W and hypercharge Y_W, which are generators of $SU(2)$ and $U(1)$, respectively.

The weak isospin is a two-component charge, which is only carried by chiral left-handed fermions. The isospin 1/2 is assigned to the neutrinos and to the up, charm and top quarks. Isospin -1/2 is assigned to the charged leptons and the down, strange and bottom quarks. Antiparticles have the opposite isospin.

Furthermore, in the weak interaction, every weak current stays within its own generation. Hence, in electroweak theory in the SM, there is no violation of lepton number. However, this is not the case for quarks. Transitions between quarks of different generations have been observed, for example in the decay of the Λ baryon [78]:

$$\Lambda(uds) \to p + \pi^-. \tag{3}$$

This is due to the weak interaction being coupled not to the mass eigenstates of quarks, but instead to a linear combination thereof. The transformation from the mass eigenstates to the weak eigenstates of quarks are done via the Cabibbo-Kobayashi-Maskawa (CKM) matrix [78]:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}, \tag{4}$$

where V_{ij} are the entries of the CKM matrix, and $|V_{ij}|^2$ corresponds to the probability of transition from quark flavor i to quark flavor j. The d', s', b' represent the weak eigenstates, and the d, s, b the strong eigenstates, which are equivalent to the mass eigenstates.

The electroweak interaction introduces the weak isospin fields W_1, W_2, W_3 and the hypercharge field B. In the Standard Model, particles associated with the electroweak interaction such as the $W^{\pm}$, Z and the photon are produced by spontaneous breaking of the $SU(2)_{I_W} \times U(1)_{Y_W}$ symmetry and the Higgs mechanism, which is descussed in Section 2.1.3. The bosons emerge through this spontaneous symmetry breaking as linear combinations of the electroweak fields [90].

$$\begin{pmatrix} A \\ Z \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B \\ W_3 \end{pmatrix} \tag{5}$$

Where A and Z are the photon and Z-boson fields, respectively. θ_W is the Weinberg or weak mixing angle.

The charged massive bosons $W^{\pm}$ can be expressed as [90]:

$$W^{\pm} = \frac{1}{\sqrt{2}}(W_1 \mp iW_2) \tag{6}$$

The electric charge can be calculated from the weak hypercharge and the third component of the weak isospin T_3 [90]:

$$Q = T_3 + \frac{1}{2}Y_W \tag{7}$$

2.1.3 The Higgs Mechanism

The Brout-Englert-Higgs-Mechanism, commonly known as the Higgs mechanism was introduced to allow mass terms for fermions and gauge bosons, while maintaining local gauge invariance [52, 63].

The Higgs mechanism introduces an isospin doublet with isospin $I_w = \frac{1}{2}$ and hypercharge $Y = 1$, with charged and neutral fields ϕ^+ and ϕ^0.

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \tag{8}$$

It also introduces the Higgs-potential $V(\phi)$ [62]:

$$V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2. \tag{9}$$

When μ^2 is negative, the Higgs-potential takes the shape as shown in Figure 2. As it can be seen, the Higgs-potential has an infinite number of degenerate ground states located at:

$$\phi_0^\dagger\phi_0 = \frac{\nu^2}{2} = -\frac{\mu^2}{2\lambda}, \tag{10}$$

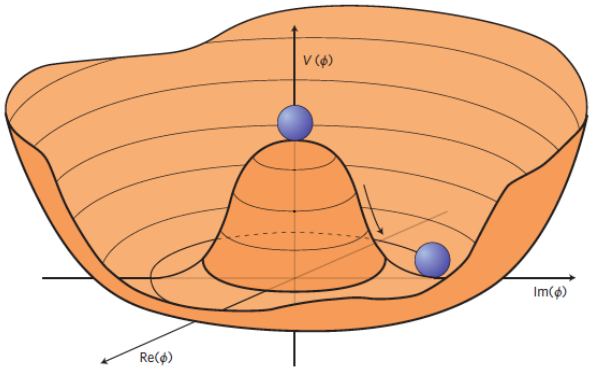


Figure 2: The form of the SM Higgs-potential, also known as the "mexican hat potential". The vacuum state is a random point at the minimum of the potential around the origin [51].

where $\nu \approx 246.22$ GeV is the *vacuum expectation value* [57].

Through spontaneous symmetry breaking one can select a ground state in the unitary gauge:

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu + H(x) \end{pmatrix} \tag{11}$$

Which results in the Higgs-Lagrangian:

$$\mathcal{L}_{Higgs} = (D_\mu \phi)^\dagger (D_\mu \phi) - V(\phi). \tag{12}$$

Here, D_μ is the covariant derivative in $\mathrm{SU}(2)_{I_W} \times \mathrm{U}(1)_{Y_W}$. After local gauge transformation and insertion of ϕ_0, in the unitary gauge and excluding Yukawa coupling terms, Equation 12 becomes:

$$\begin{aligned} \mathcal{L}_H = \frac{1}{2}(\partial_\mu H)(\partial^\mu H) - \lambda\nu^2 H^2 - \lambda\nu H^3 - \frac{1}{4}\lambda H^4 + \frac{1}{4}\lambda\nu^4 \\ + \frac{1}{2}(\frac{\nu g}{2})^2 W^{\mu,+} W^-_\mu + \frac{1}{2}(\frac{\nu g}{2\cos\theta_w})^2 Z^\mu Z_\mu \\ + g(\frac{\nu g}{2}) H W^{\mu,+} W^-_\mu + g(\frac{\nu g}{4\cos^2\theta_w}) H Z^\mu Z_\mu \\ + \frac{g^2}{4} H^2 W^{\mu,+} W^-_\mu + \frac{g^2}{4\cos^2\theta_w} H^2 Z^\mu Z_\mu + const. + \dots \end{aligned} \tag{13}$$

From which a new particle, the Higgs boson, is predicted. The mass of the Higgs boson is given by the term quadratic to the Higgs field, $m_H^2 = 2\lambda\nu^2$. The mass of the Higgs Boson itself is not predicted by the SM.

It can be seen that with the introduction of the Higgs Lagrangian, mass terms for the W and Z bosons can be written as well, namely $m_W = \frac{1}{2} g\nu$ and $m_Z = \frac{1}{2}\frac{g}{\cos\theta_w}\nu$, respectively.

The Higgs Lagrangian also describes the couplings between itself and the W and Z bosons. The terms $\lambda\nu H^3$ and $\frac{1}{4}\lambda H^4$ describe the triple and quadruple self coupling of the Higgs

boson. The terms proportional to $HW^{\mu,+}W_\mu$ and $HZ^\mu Z_\mu$ describe the Higgs coupling to the W and Z bosons.

2.1.4 Yukawa Coupling

The Yukawa coupling allows for the coupling of the left-handed $SU(2)_{I_W}$ fermion doublet to the Higgs doublet, thereby generating fermion masses within the Standard Model. The introduction of the Yukawa coupling was necessary since while in QED, mass terms for fermions are gauge invariant, under $SU(2)_{I_W}$, it is not. This is because left- and right-handed fermion states transform differently, since, for example, an electron the mass term can be written as a product of its chiral states:

$$-m\bar{e}e = -m(\bar{e_L}e_R + \bar{e}_R e_L), \tag{14}$$

where e is the field of the electron, and $e_{R,L}$ the right or left handed chiral states associated with the electron. After electroweak symmetry breaking, one finds for each individual fermion f the Lagrangian density:

$$\mathcal{L}^f_{\text{Yukawa}} = \frac{1}{\sqrt{2}}(\nu(\lambda_f \bar{\psi}^f \psi^f) + (\lambda_f H \bar{\psi}^f \psi^f)), \tag{15}$$

where λ_f is the Yukawa coupling of the Higgs field to the fermion. From this one can write the mass of the fermion as:

$$m_f = \frac{\lambda_f}{\sqrt{2}}\nu \tag{16}$$

The coupling of the Higgs boson to the fermion is proportional to λ_f and the mass of the fermion.

2.2 Higgs-boson Phenomenology

The Higgs Mechanism predicts the existence of a Higgs boson with a spin of 0. However, the mass of the Higgs boson is not predicted by the SM, which is must be known in order to predict other properties of the Higgs boson. In 2012, the ATLAS and CMS collaborations jointly announced the discovery of a new boson with a mass of approximately 125 GeV, whose properties were consistent with those of the Higgs boson [21, 44].

2.2.1 Higgs-boson production modes

The four dominant production modes of the Higgs-boson in pp-collisions at the LHC are gluon fusion (ggF), vector-boson fusion (VBF), Higgs-Strahlung (VH), and Higgs production in association with a pair of top-quarks ($\bar{t}tH$). The predicted Higgs-boson production cross sections as a function of the mass of the Higgs-boson for different production modes are shown in Figure 3. Example leading order Feynman-Diagrams of these processes are shown in Figure 4. The expected production cross section for each of these

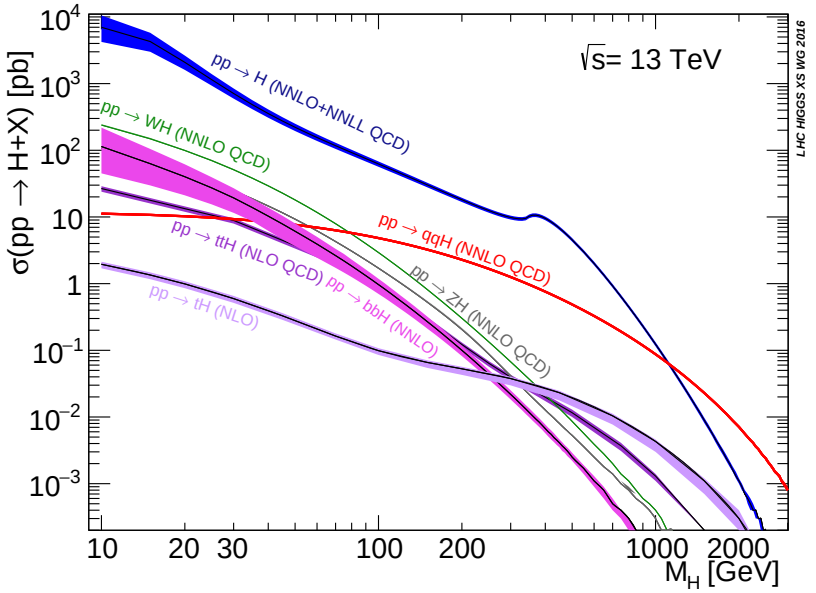


Figure 3: The predicted production cross section of the Higgs-boson as a function of the mass of the Higgs-boson for different production modes [81].

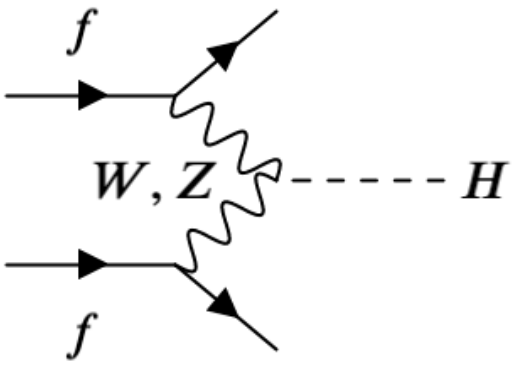
production modes for a center of mass energy of $\sqrt{s} = 13$ TeV and a Higgs-mass of 125.09 GeV can be found in Table 1.

Process		H-Production Cross Section [pb]
VBF		$3.766\ ^{+0.45\%}_{-0.33\%}(\text{scale})\ \pm 2.1\%(\text{PDF}+\alpha_s)$
VH	W^-H	$0.527\ ^{+0.59\%}_{-0.63\%}(\text{scale})\ \pm 2.03\%(\text{PDF}+\alpha_s)$
	W^+H	$0.831\ ^{+0.74\%}_{-0.73\%}(\text{scale})\ \pm 1.79\%(\text{PDF}+\alpha_s)$
	ZH	$0.880\ ^{+3.50\%}_{-2.68\%}(\text{scale})\ \pm 1.65\%(\text{PDF}+\alpha_s)$
ggH		$48.61\ ^{+4.27\%}_{-6.48\%}(\text{theory})\ \pm 1.85\ (\text{PDF})\ ^{+2.59\%}_{-2.62\%}(\alpha_s)$
$\bar{t}tH$		$0.507\ ^{+5.8\%}_{-9.2\%}(\text{scale})\ \pm 3.6\%(\text{PDF}+\alpha_s)$

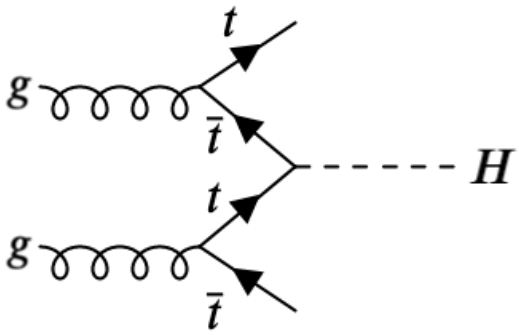
Table 1: The production cross section of the Higgs-boson as predicted by the Standard Model at a Higgs mass of 125.09 GeV at a center of mass energy of 13 TeV. Explanations for the scale and PDF uncertainties are described in [54].

The production cross-section of the Higgs boson for each production mode was experimentally measured to be: 2.7 ± 0.4 (stat) $^{+0.9}_{-0.6}$ (sys) pb for ggH, 0.197 ± 0.028 (stat) $^{+0.032}_{-0.026}$ (sys) pb for VBF, 0.012 ± 0.06 (stat) ± 0.04 (sys) pb for VH and $0.033\ ^{+0.033}_{-0.029}$ (stat) $^{+0.022}_{-0.017}$ (sys) pb for $\bar{t}tH$. The total cross-section of the $pp \to H \to \tau\tau$ process was measured to be 2.94 ± 0.21 (stat) $^{+0.37}_{-0.32}$ (sys) pb [17].

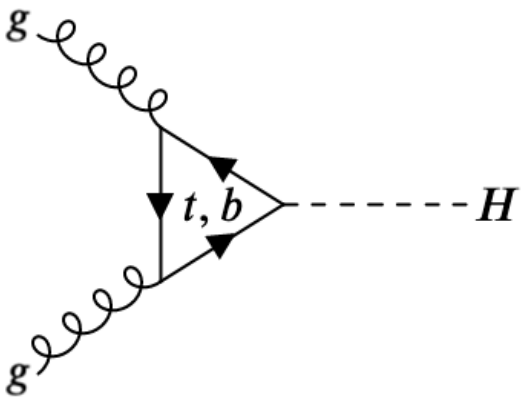
The signal process in this thesis is VBF, where two quark scatter off each other by emitting a W or a Z boson each, which fuse into a Higgs boson. The scattered quarks hadronize into two jets in the forward and backwards region [79].



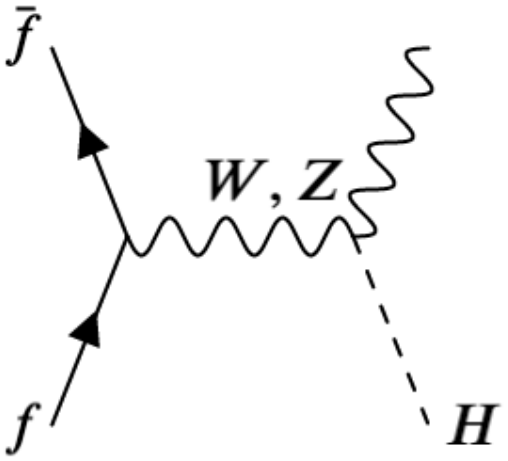
(a) Vector-boson fusion



(b) Associated production with a pair of top quarks



(c) Gluon-fusion



(d) Higgs-Strahlung

Figure 4: Example Feynman diagrams of the different production modes of the Higgs-boson: (a) Vector-boson fusion, (b) Associated production with a pair of top quarks (c) Gluon-fusion (d) Higgs-Strahlung.

2.2.2 Decays of the Higgs-boson

In this thesis the $H \to \tau^+\tau^- \to e\mu 4\nu$ channel is considered. However, since neutrinos are not detected by the ATLAS detector, this process is not distinguishable from the $H \to WW \to e\mu 2\nu$ process. Furthermore, it can be shown that the inclusion of this process increases the sensitivity to CP violation [61]. Therefore, this process is also considered as a signal process.

Branching ratios of the Higgs-boson depend on the mass of the Higgs-boson. The branching ratio of each possible Higgs-boson decay as a function of the mass of the Higgs-boson is shown in Figure 5. The predicted branching ratios of the relevant decays $H \to \tau\tau$ and $H \to WW$ for a Higgs mass of $m_H = 125.36$ GeV are 6.26 ± 0.35 and 22.0 ± 0.9 respectively [18].

The τ-lepton has a very short mean lifetime of $(290.3 \pm 0.5)\times 10^{-15}$ s and therefore only

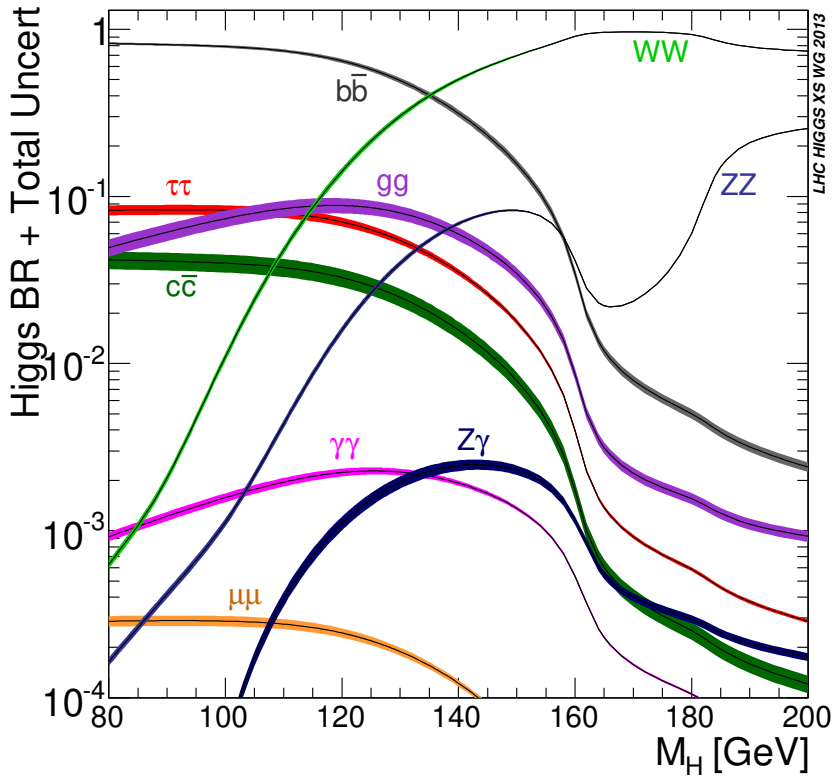


Figure 5: Branching ratios of the different SM Higgs-boson decay modes as a function of the mass of the Higgs-boson [54].

its decay products can be directly measured [78]. The τ-lepton may decay hadronically or leptonically. Considering only the leptonic decay modes, the branching ratio of the $\mu + \bar{\nu_\mu} + \nu_\tau$ final state is 17.39 $\pm$ 0.04% and for $e + \bar{\nu_e} + \nu_\tau$ this is 17.82 $\pm$ 0.04% [78]. Therefore, a ditau system decays into $e + \mu + 4\nu$ in 6.20$\pm$0.03% of cases. A W-boson decays into a $e + \nu_e$ and $\mu + \nu_\mu$ in 10.71$\pm$0.16% and 10.63$\pm$0.15% of cases, respectively [78]. A WW system therefore decays into $e + \mu + 2\nu$ in 2.91$\pm$0.08% of all cases.

The product of the branching ratio with the VBF Higgs production cross-section for $H \to \tau\tau \to e\mu 4\nu$ and $H \to WW \to e\mu 2\nu$ procceses are therefore 14.6 $\pm$ 1.3 fb and 24.1$\pm$2.3 fb, respectively. Using the relation $N = \mathcal{L}_{\text{int}}\sigma$, where N is the number of events, $\mathcal{L}_{\text{int}} = 139\text{fb}^{-1}$ the integrated luminosity, defined in Chapter 3, and σ the cross section; the expected number of events for $H \to \tau\tau \to e\mu 4\nu$ and $H \to WW \to e\mu 2\nu$ processes are found to be 1950 $\pm$ 170 and 3300 $\pm$ 190 events, respectively.

2.3 CP Violation

There is a large imbalance between the amount of baryons and antibaryons in this universe. One of the Sakharov conditions for baryon asymmetry is CP-violation [83]. CP symmetry is the combination of charge (C) and parity (P) symmetries. A charge transformation transforms particles to their antiparticles and vice versa. Parity transformation

inverts the signs of the components of the spacial coordinates:

$$P : \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} -x \\ -y \\ -z \end{pmatrix}, \tag{17}$$

where P represents the parity operator.

Parity symmetry was thought to be a fundamental symmetry of the universe, however experiments such as the Wu experiment has shown that parity is in fact not conserved. Furthermore, in the weak interaction, parity is maximally violated [91].

The combination of C and P symmetries was also found to be violated in the Cronin and Fitch experiment in Kaon decays [39]. Subsequent experiments have also discovered CP violation in other sectors involving B and D mesons [38, 69].

CP violation is present in the SM through the complex phase of the entries of the CKM matrix. However, the amount of CP violation provided by the CKM matrix is too small to explain the observed baryon asymmetry. Thus, experimental searches for further sources of CP violation in other sectors are necessary.

2.3.1 Effective Field Theories

Effective field theories (EFT) provide a framework to describe physics beyond the SM (BSM). BSM physics may introduce new particles that are too heavy to be produced directly. Utilizing EFT's, one can describe the effect of heavy particles, which may only be produced at high energy scales Λ, at lower energy scale accessible with current technology.

The Standard Model Effective Field Theory (SMEFT) provides a framework to describe new physics using higher-dimensional operators including SM fields [61]. A general form of a Lagrangian in SMEFT is:

$$\mathcal{L} = \mathcal{L}_{SM} + \sum_i \frac{f_i^{(5)}}{\Lambda} \mathcal{O}_i^{(5)} + \sum_i \frac{f_i^{(6)}}{\Lambda^2} \mathcal{O}_i^{(6)} + \ldots, \tag{18}$$

where $\mathcal{L}_{SM}$ is the SM Lagrangian, f_i are called the *Wilson coefficients* [61], $\mathcal{O}^{(D)}$ gauge invariant operators with mass-dimension $D > 4$, and energy scale of new physics is represented by Λ [61].

In this thesis, one would like the investigate the effect of a CP-violating contribution to the HVV coupling. Therefore one considers operators with mass dimension 6, which are CP odd, $\mathrm{U}(1)_Y \times \mathrm{SU}(2)_{I_W,L}$ invariant, and use electroweak and Higgs gauge boson fields. This results in the following Lagrangian:

$$\mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{SM}} + \frac{f_{\tilde{B}B}}{\Lambda^2} \mathcal{O}_{\tilde{B}B} + \frac{f_{\tilde{W}W}}{\Lambda^2} \mathcal{O}_{\tilde{W}W} + \frac{f_{\tilde{B}}}{\Lambda^2} \mathcal{O}_{\tilde{B}}, \tag{19}$$

with operators $\mathcal{O}_i$ are defined as:

$$\mathcal{O}_{\tilde{B}B} = \Phi^+ \hat{\tilde{B}}_{\mu\nu} \hat{B}^{\mu\nu} \Phi \tag{20}$$

$$\mathcal{O}_{\tilde{W}W} = \Phi^+ \hat{\tilde{W}}_{\mu\nu} \hat{W}^{\mu\nu} \Phi \tag{21}$$

$$\mathcal{O}_{\tilde{B}} = (D_\mu \Phi)^+ \hat{\tilde{B}}^{\mu\nu} D_\nu \Phi \tag{22}$$

In these equations, the covariant derivative is given by:

$$D_\mu = \partial_\mu + \frac{i}{2} g' B_\mu + i g \frac{\sigma^a}{2} W_\mu^a. \tag{23}$$

The field strength tensors $\hat{B}^{\mu\nu}$ and $\hat{W}^{\mu\nu}$ of the W and B gauge fields and their dual ones are given by:

$$\hat{B}^{\mu\nu} + \hat{W}^{\mu\nu} = i \frac{g'}{2} B_{\mu\nu} + i \frac{g}{2} \sigma^a W_{\mu\nu}^a = [D_\mu, D_\nu], \tag{24}$$

$$\tilde{V_{\mu\nu}} = \frac{1}{2} \epsilon_{\mu\nu\rho\sigma} V^{\rho\sigma}, \qquad (V = B, W), \tag{25}$$

, with couplings g and g' and $[\cdot, \cdot]$ the commutator operator.

The operator $\mathcal{O}_{\tilde{B}}$ is not considered since it is constrained heavily by measurements at LEP [59].

After electroweak symmetry breaking, the above Lagrangian becomes [61]:

$$\mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{SM}} + \tilde{g}_{HAA} H \tilde{A}_{\mu\nu} A^{\mu\nu} + \tilde{g}_{HAZ} H \tilde{A}_{\mu\nu} Z^{\mu\nu} + \tilde{g}_{HZZ} H \tilde{Z}_{\mu\nu} Z^{\mu\nu} + \tilde{g}_{HWW} H \tilde{W}^+_{\mu\nu} W_-^{\mu\nu}, \tag{26}$$

Anomalous couplings HAA and HAZ, which are not present in the SM. whereas couplings HZZ and HWW are present. The Lagrangian furthermore contains the following coupling strengths:

$$\tilde{g}_{HAA} = \frac{g}{2m_W} (\tilde{d} \sin^2 \theta_W + \tilde{d}_B \cos^2 \theta_W), \qquad \tilde{g}_{HAZ} = \frac{g}{2m_W} \sin 2\theta_W (\tilde{d} - \tilde{d}_B), \tag{27}$$

$$\tilde{g}_{HZZ} = \frac{g}{2m_W} (\tilde{d} \cos^2 \tilde{\theta}_W + \tilde{d}_B \sin^2 \theta_W), \qquad \tilde{g}_{HWW} = \frac{g}{m_W} \tilde{d}, \tag{28}$$

where

$$\tilde{d} = -\frac{m_W^2}{\Lambda^2} f_{\tilde{W}W}, \qquad \tilde{d}_B = -\frac{m_W^2}{\Lambda^2} \tan^2(\theta_W) f_{\tilde{B}B}, \tag{29}$$

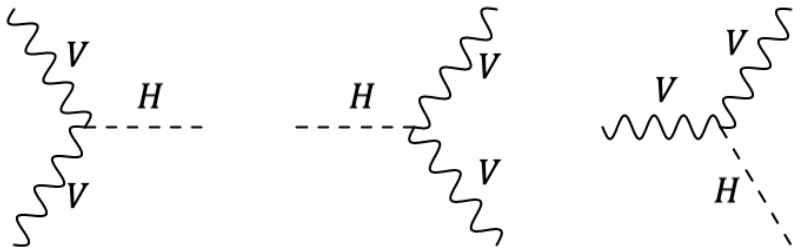


Figure 6: Example Feynman diagrams depicting different possible HVV coupling vertices. From left to right: Vector boson fusion, Higgs decay, and Higgs-Strahlung.

The effect of $\gamma\gamma$, γZ, ZZ and WW to VBF processes cannot be distinguished experimentally [61]. Therefore one can arbitrarily set $\tilde{d} = \tilde{d}_B$, which results in the following coupling strengths:

$$\tilde{g}_{HAA} = \tilde{g}_{HWW} = \frac{1}{2}\tilde{g}_{HWW} = \frac{g}{2m_W}\tilde{d} \text{ and } \tilde{g}_{HAZ} = 0. \tag{30}$$

Therefore, the strength of CP violation in the HVV coupling can be parameterized by a single parameter $\tilde{d}$ [61].

2.3.2 Matrix Element with CP-violating Contributions

Three cases of the HVV coupling are: VBF, Higgs-decay and Higgs-Strahlung. Example Feynman diagrams of these couplings are shown in Figure 6.

Considering the most general tensor structure of the HVV coupling [61]:

$$T^{\mu\nu}(q_1, q_2) = a_1(q_1, q_2)g^{\mu\nu} + a_2(q_1, q_2)[q_1 \cdot q_2 g^{\mu\nu} - q_2^{\mu}q_1^{\nu}] + a_3(q_1, q_2)\eta^{\mu\nu\rho\sigma}q_{1\rho}q_{2\sigma} \tag{31}$$

Where q_1 and q_2 refer to the four-momenta of the two vector-bosons. The first and second terms are CP-even, whereas the last term is CP-odd. In the SM, a_2 and a_3 are equal to zero. Investigations of such BSM contributions are favored at high energies since the second and third terms are proportional to q_1 and q_2.

The Higgs-decay case is therefore not favorable due to the momenta of the vector bosons being limited by the mass of the Higgs boson. The Higgs-Strahlung case is also not favorable due to its low cross section. Therefore, vector boson fusion production of the Higgs boson is advantageous in investigating CP properties of the HVV coupling.

The matrix-element for this process with CP-violating contributions gains an additional term:

$$\mathcal{M} = \mathcal{M}_{\text{SM}} + \tilde{d}\mathcal{M}_{\text{CP-odd}}. \tag{32}$$

The absolute square of $\mathcal{M}$ is:

$$|\mathcal{M}|^2 = \underbrace{|\mathcal{M}_{SM}|^2}_{\text{CP-Even}} + \underbrace{\tilde{d}2\Re\{\mathcal{M}^*_{SM}\mathcal{M}_{CP-odd}\}}_{\text{CP-odd}} + \underbrace{\tilde{d}^2|\mathcal{M}_{CP-odd}|^2}_{\text{CP-Even}} \tag{33}$$

The first term, being the SM-case, and the last term, being quadratic to a CP-odd term, are CP-even. The second interference term, which is linear in $\tilde{d}$, is CP-odd.

2.3.3 Test of CP Invariance

Observables can be constructed that are sensitive to CP-violating contributions of the HVV coupling. CP-odd Observables can be used for investigating the CP properties of the Higgs boson. Any average value of this CP-odd variable incompatible with 0 would therefore imply the existence of CP-violating contributions to the HVV coupling.

The Optimal Observable

For the analysis presented in this thesis, the *optimal observable* is used, which is defined in the following way:

$$\mathcal{OO} = \frac{2\Re\{\mathcal{M}^*_{SM}\mathcal{M}_{CP-odd}\}}{|\mathcal{M}_{SM}|^2}, \tag{34}$$

When comparing Equation 33 and 34,it can be seen that the $\mathcal{OO}$ is the ratio of the CP-odd term in the absolute square of the matrix element in the CP-violating case to that of the SM case.

The matrix elements are calculated using HAWK [48], which takes in the truth information of the Higgs boson, VBF jets and partons as inputs.

The optimal observable combines the full phase-space information into a single observable, and it can be shown that on truth-level it has optimal sensitivity for CP-violating contributions [30]. A histogram showing the distribution of the optimal observable for different $\tilde{d}$ hypotheses is shown in Figure 7.

Other CP-odd variables

Other CP-odd variables can be used in the investigation of CP-properties of the HVV coupling. These include the signed azimuthal angle difference between the tagging jets. Defined as:

$$\Delta\phi_{jj}^{signed} = \phi_{j_1} - \phi_{j_2} \tag{35}$$

where the jets are ordered such that $\eta_{j_1} > \eta_{j_2}$. This variable is P-odd and therefore is also sensitive to CP-violating contributions. Previous studies have shown that $\Delta\phi_{jj}^{signed}$ does not outperform the Optimal Observable [20].

Another a CP-odd variable that can be investigated is constructed using neural networks trained on matrix-element information, where *symbolic regression* is used to construct observables, which are presented as simple human-interpretable formulas [34]. Several

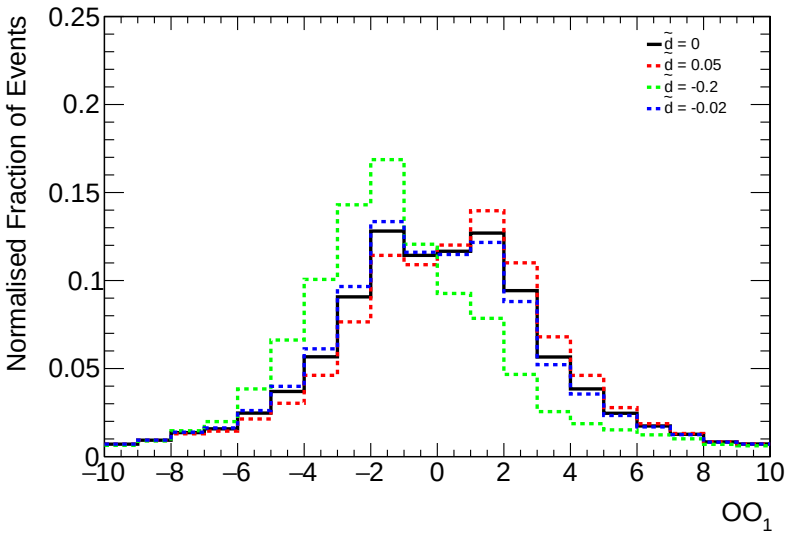


Figure 7: The normalized distribution of the Optimal Observable for different values of $\tilde{d}$, evaluated using a Monte Carlo simulated dataset.

different configurations of such formulas are presented within the study with different performances. The observable with the simplest expression whose performance is not significantly different from the optimum performance is taken. This is defined as:

$$\mathcal{O}^{Reg} = -5.5386 \times 10^{-5} p_T^{j_1} p_T^{j_2} \sin \Delta\phi_{jj}^{signed} \tag{36}$$

3 The ATLAS Experiment

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [53] located at CERN (Conseil Européen pour la Recherche Nucléaire) in Geneva, Switzerland is the most energetic circular particle collider in the world, designed for proton-proton (pp) and heavy ion collisions (Pb-Pb) at center-of-mass energies ($\sqrt{s}$) of up to 14 TeV. The LHC has a circumference of about 27 kilometers and accelerates protons and heavy ions in two parallel pipes. For proton-proton collisions, each beam contains up to 2808 *bunches* with 10^{11} particles each, spaced at an interval of 25 ns.

The beams intersect at four locations, which house particle detectors for the ATLAS [29], CMS [45], LHCb [70] and ALICE [4] experiments. These detectors are used for precise measurements, the largest of them in size being the ATLAS detector.

The amount of collisions and subsequently the amount of recorded data can be quantified by the *luminosity*. The *instantaneous luminosity* ($\mathcal{L}$) is defined to be the ratio of the event rate to the cross-section (σ):

$$\mathcal{L} = \frac{1}{\sigma}\frac{dN}{dt}. \tag{37}$$

In practice, the instantaneous luminosity depends on different collider parameters and can be expressed as

$$\mathcal{L} = n_b \frac{N^2 \gamma_r f_{rev}}{4\pi\beta^* \epsilon_n} F, \tag{38}$$

where N is the number of particles per bunch, n_b the number of bunches per beam, f_{rev} revolution frequency, γ_r the relativistic gamma factor, ϵ_n the normalized transverse beam emittance, β^* the beta-function at the collision point and F the geometric luminosity reduction factor due to the beam crossing angles. The LHC is designed to have an instantaneous luminosity of up to $10^{34}\text{cm}^2\text{s}^{-1}$ [53].

The *integrated luminosity* ($\mathcal{L}_{\text{int}}$) is the integral of the instatenous luminosity with respect to time:

$$\mathcal{L}_{\text{int}} = \int \mathcal{L} dt \tag{39}$$

Data-taking with the LHC began in 2011 with the "Run-1". In 2011, the LHC operated at a center-of-mass energy of 7 TeV, which was subsequently increased to 8 TeV in 2012. During this period, an integrated luminosity of 23.3 fb^{-1} was delivered to the ATLAS detector, of which 20.3 fb^{-1} was used for physics analysis [3, 15]. After a maintenance period of 3 years, 2015 marked the beginning of "Run-2", which lasted until 2018. During this data-taking period, a total integrated luminosity of 147 fb^{-1} was recorded at a center-of-mass energy of $\sqrt{s} = 13$ TeV, 139 fb^{-1} of which is used for physics analyses [27].

Multiple interactions occuring per bunch crossing is referred to as *in-time pile-up* [73]. Another type of pile-up is *out-of-time pile-up* [73], which happens due to the read-out

time of the calorimeters being longer than the bunch interval of 25 ns. The luminosity-weighted distribution of the mean number of interactions per crossing for Run-2 data is shown in Figure 8.

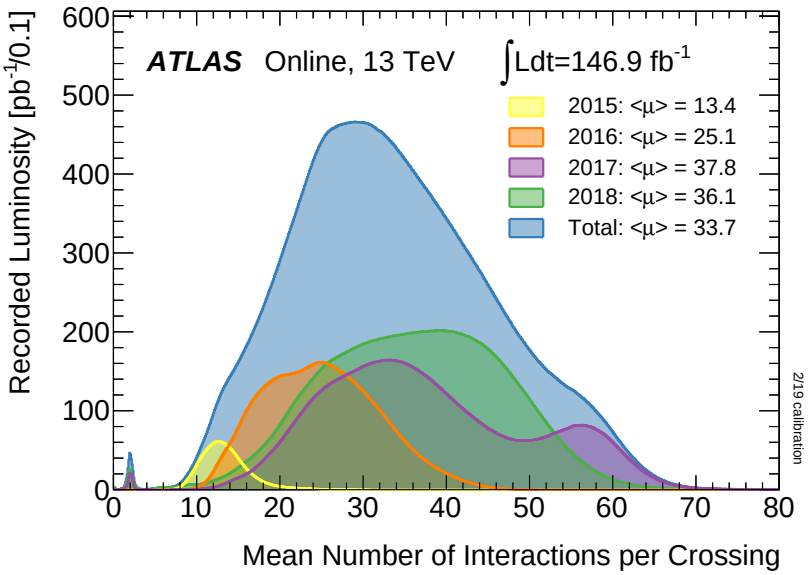


Figure 8: The luminosity-weighted distribution of the mean number of interactions per bunch crossing ($\langle\mu\rangle$) for the full *Run-2 pp* collision data at $\sqrt{s} = 13$ TeV for each year and also combined [25].

3.2 The ATLAS Detector in Run-2

The ATLAS (**A** **T**oroidal **L**HC **A**pparatu**S**) detector is a cylindrical, multi-purpose particle detector with a length of 44 meters and a diameter of 27 meters, centered around the beam axis [29]. It consists of multiple components, each specialized for specific measurements, as well as a magnet system that immerses these units in a strong magnetic field of up to 3.5 T. An illustration of the ATLAS detector is shown in Figure 9. The innermost part closest to the beam axis is the Inner Detector (ID), which determines the tracks and momenta of charged particles. The electromagnetic calorimeters (EM) measure the energy deposited by electrons and photons, whereas hadronic calorimeters measure the energy deposited by hadrons. The muon chambers, which is located at the outermost layer of the detector, provides measurements for the momenta of muons.

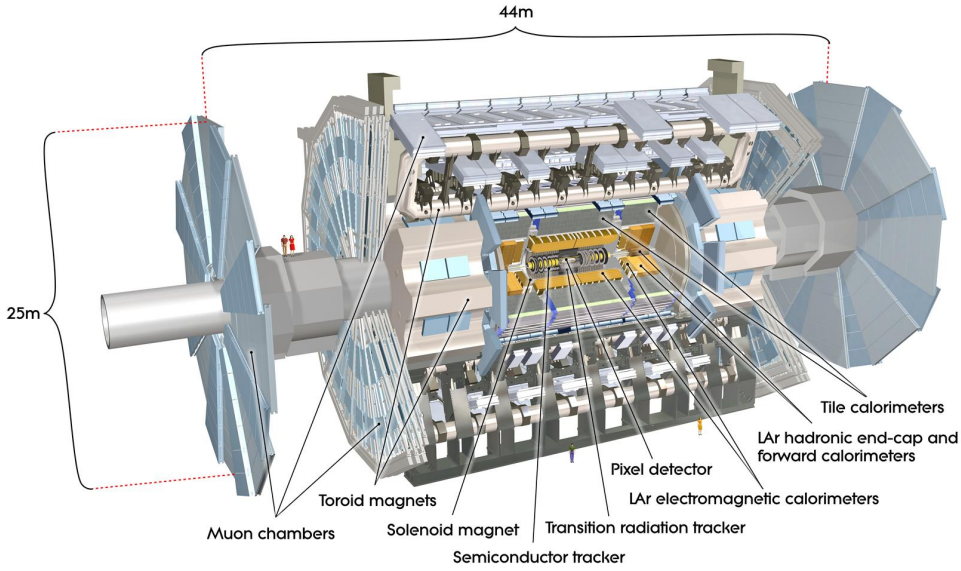


Figure 9: A computer-generated illustration of the ATLAS Detector with its main constituent parts labeled [29].

3.3 Coordinate System of the ATLAS Detector

A cylindrical, right-handed coordinate system is defined with its origin at the nominal interaction point where the protons collide. The beam direction defines the z-axis, where the positive z axis points counter-clockwise when viewing the LHC ring from above. The x-y plane is normal to the beam direction. The positive x direction is defined to be pointing from the interaction point towards the center of the LHC ring, and the positive y direction pointing upwards. The azimuthal angle ϕ is measured around the beam axis and the polar angle θ from the beam axis. The pseudorapidity, η, is defined as

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right], \tag{40}$$

which maps $\theta \in [-\pi, \pi]$ to $\eta \in [-\infty, \infty]$. An illustration of the coordinate system of the ATLAS detector is shown in figure 10.

Another useful quantity is the distance in the pseudorapidity-azimuthal angle space between two objects, defined as:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}, \tag{41}$$

where $\Delta\eta$ and $\Delta\phi$ are the differences in pseudorapidity and azimuthal angle between two objects, respectively.

Transverse observables such as the transverse momentum,

$$p_T = \sqrt{p_x^2 + p_y^2} \tag{42}$$

and the transverse energy,

$$E_T = \sqrt{E_x^2 + E_y^2}, \tag{43}$$

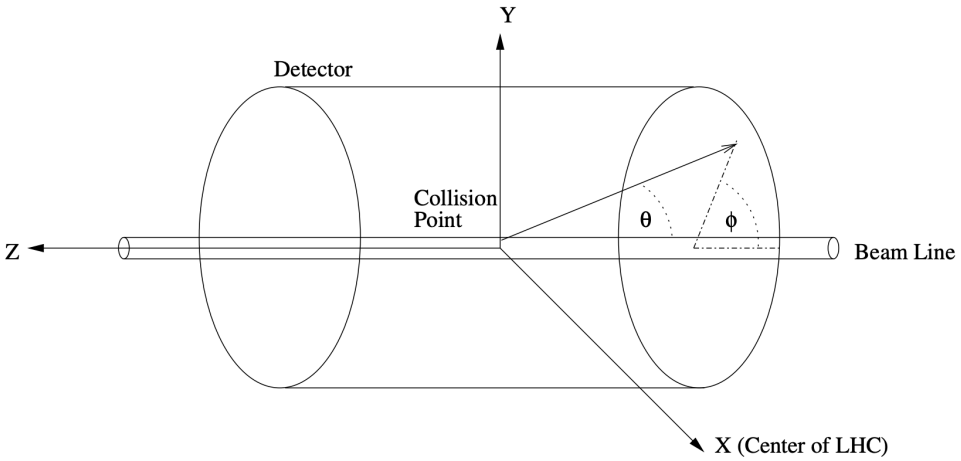


Figure 10: A diagram illustrating the coordinate system of the ATLAS detector [86].

are defined, where $p_{x,y}$ are the x and y components of the momentum vector and $E_{x,y}$ the x and y components of the energy vector.

3.4 Inner Detector

The inner detector provides momentum measurements, tracking and vertex reconstruction of charged particles. The inner detector comprises of the insertable B-layer (IBL), pixel detector, semiconductor tracker (SCT) and the transition radiation tracker (TRT), and is immersed in a magnetic field of 2 T [6].

The insertable B-layer (IBL) is the innermost layer of the pixel detector with a radius of 3.3 cm from the beam axis and compensates for the deterioration of the B-Layer, the first layer of the pixel detector, due to radiation damage [68]. The proximity of the IBL to the interaction point also improves the vertexing and b-tagging (see Section 4) performance. The IBL covers a pseudorapidity range of $|\eta| < 3$ and consists of 14 fiber staves each 2 cm wide and 64 cm long.

The precision tracking detectors consist of the pixel detector and semiconductor tracker, which cover a pseudorapidity range of $|\eta| < 2.5$ [9, 46].

The pixel detectors surround the insertable B-layer and cover a pseudorapidity range of $|\eta| < 2.5$ [9], providing high-precision measurements close to the interaction point. The pixel detectors, as with other components of the ATLAS detector, are split into the barrel region, radially surrounding the beam axis, and the endcap region, which cover the circular area on each side of the barrel region. This system is composed of three barrel layers with mean radii of 5 cm, 8.8 cm and 12.2 cm and length 80.1 cm and six disk layers, with three at each end of the barrel region. The barrel layers contain approximately 67 million pixels, while the endcaps contain approximately 13 million pixels, covering a total active area of about 1.7 m^2.

The SCT surrounds the pixel detector and provides precision measurements allowing the

measurement of momentum, impact parameters and vertex position [46]. It occupies the region 30−52 cm radially from the interaction point with a width of 5.4 m, providing a coverage of up to $|\eta| < 2.5$. The SCT is composed of 4088 silicon strip modules arranged into 4 barrel layers and 9 endcap disks at each end.

The TRT is a straw-tube tracker, which improves the electron identification capabilities and accuracy of the momentum measurement in the ID by providing further hit measurements. [89]. The barrel region covers a radius from 0.5 to 1.1 meters and a pseudorapidity range of $|\eta| < 1$. It consists of 52544 tubes each with a length of 1.5 m and a diameter of 4 mm. The endcaps contains straws with 0.4 m length, arranged radially to the beam axis. Each endcap consists of 122800 straws covering a range length and angle range of 0.8 m$< |z| <$ 2.7 m and $1 < |\eta| < 2$.

The design resolution for p_T (σ_{p_T}) of the inner detector is:

$$\frac{\sigma_{p_T}}{p_T} = 0.05\% p_T \oplus 1\%^1. \tag{44}$$

A cross-sectional drawing of the inner detector with its main components labeled is shown in Figure 11.

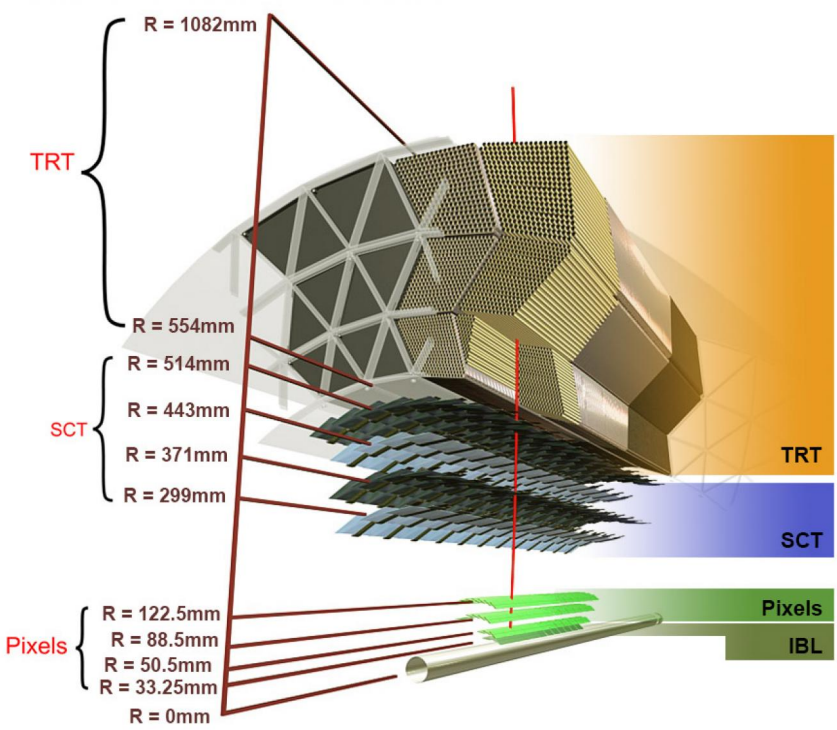


Figure 11: A cross-sectional diagram of the Inner Detector of the ATLAS Detector, with the radial distances (R) of the layers of the Insertable B-Layer (IBL), pixel detector (Pixels), Semiconductor Tracker (SCT) and Transition Radiation Tracker (TRT) from the beam axis provided in units of millimeters (mm)[29].

[1] $a \oplus b = \sqrt{a^2 + b^2}$

3.5 Calorimeters

The calorimeters measure the energies of detectable particles apart from the muon [7]. They consist of two parts: the electromagnetic calorimeter, which measures the energy deposited by electrons and photons, and the hadronic calorimeter, which measures the energy deposited by hadrons. The calorimeters cover a pseudorapidity range of $\eta < |4.9|$.

3.5.1 Electromagnetic Calorimeter

The electromagnetic calorimeter (EM) is a lead-liquid argon (LAr) calorimeter which surrounds the inner detector [7]. The EM calorimeter is a sampling calorimeter with lead as the absorber and liquid argon as the active material. The EM calorimeter is divided into a barrel region ($|\eta| < 1.475$) and two end-cap regions ($1.375 < |\eta| < 3.2$) on each side of the barrel. The barrel region has a length of 6.4 meters, and the end caps have a radius of 2.077 meters. The $|\eta|$ range between 1.37 and 1.52 corresponds to a transition region between the barrel region and end-cap region cryostats and results in a significantly degraded energy resolution. Electrons as well as photons which fall into this region are not used for the analysis. The design energy resolution (σ_E) of the EM calorimeter is:

$$\frac{\sigma_E}{E} = 10\%/\sqrt{E} \oplus 0.7\%. \tag{45}$$

3.5.2 Hadronic Calorimeter

The hadronic calorimeter surrounds the EM calorimeter and measure the energy deposited by hadrons, primarily in the form of jets [29]. It consists of three components: the tile calorimeter, LAr hadronic end-cap calorimeter and LAr forward calorimeter. The tile calorimeter is a sampling calorimeter with a layered structure using steel as the absorber and scintillating tiles as the active material; and covers the range $|\eta| < 1.7$ with an inner radius of 2.28 meters and an outer radius of 4.25 meters. The end-cap hadronic calorimeters are also lead-liquid argon calorimeters and along with the tile calorimeters cover the range $0 < |\eta| < 3.2$ with an inner radius of 0.475 meters and an outer radius of 2.03 meters. The LAr Forward calorimeters provide coverage for pseudorapidity ranges closer to the beam axis ($3.1 < |\eta| < 4.9$). The tile and end-cap calorimeters have a design energy resolution (σ_E) of

$$\frac{\sigma_E}{E} = 50\%/\sqrt{E} \oplus 3\%, \tag{46}$$

while the forward LAr calorimeter has a design energy resolution of

$$\frac{\sigma_E}{E} = 100\%/\sqrt{E} \oplus 10\%. \tag{47}$$

Therefore the energy resolution of the hadronic calorimeter is lower than that of the EM calorimeter. An illustration of both calorimeters with their major components highlighted is shown in Figure 12.

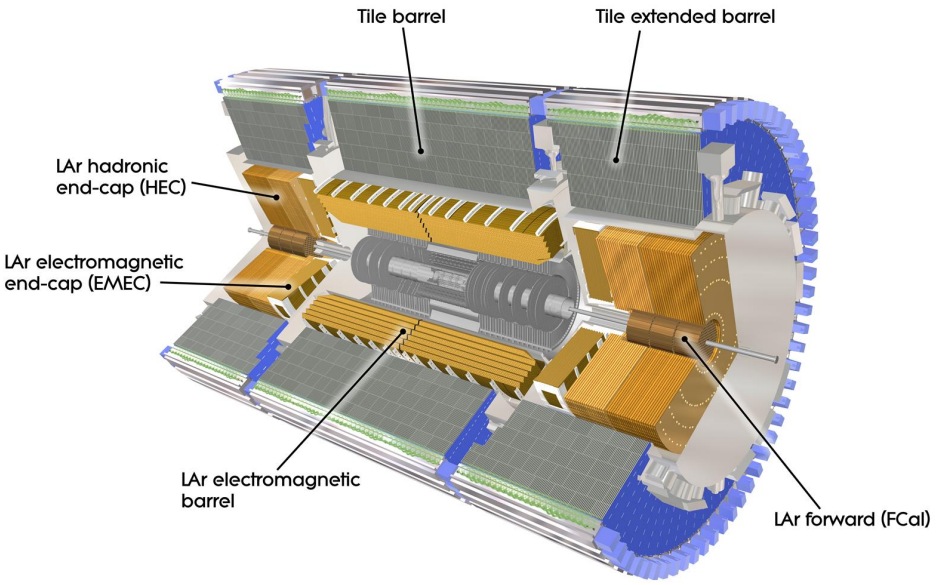


Figure 12: An illustration of the calorimeters of the ATLAS Detector [29].

3.6 Muon Spectrometer

Muons are not stopped by the calorimeters, and therefore it is not possible to precisely measure their energy using calorimetry. The muon spectrometer determines tracks of muons passing through it [8]. It utilizes a strong magnetic field and precise tracking chambers to determine momenta by measuring the curvature of their tracks. For pseudorapidity ranges of $|\eta| < 1.7$ the magnetic field is provided by the large barrel toroid. For pseudorapidity ranges of $|\eta| < 2.7$, the monitored drift tubes; and for ranges of $2.0 < |\eta| < 2.7$, cathode strip chambers provide information for tracking. The design resolution of the transverse momentum measurement ($\sigma_{p_{\mathrm{T}}}$) in the muon spectrometer is:

$$\frac{\sigma_{p_{\mathrm{T}}}}{p_{\mathrm{T}}} = 10\% \tag{48}$$

for a muon with a p_{T} of 1 TeV [29]. An illustration of the muon system is shown in Figure 13.

3.7 Trigger System

Beam bunches at the LHC are delivered at a rate of 40 MHz, and resulting signals from the ATLAS detector create a data volume of approximately 60 million megabytes per second [36]. Therefore, trigger systems are used to select events with distinct characteristics that are relevant for physics analyses. The trigger system of the ATLAS detector is composed of three consecutive parts: Level-1 (L1) trigger, Level-2 (L2) trigger and event filter.

The L1 trigger is a hardware-based trigger using electronics attached to the detectors. It utilizes information from the calorimeters and the muon spectrometer to select events with

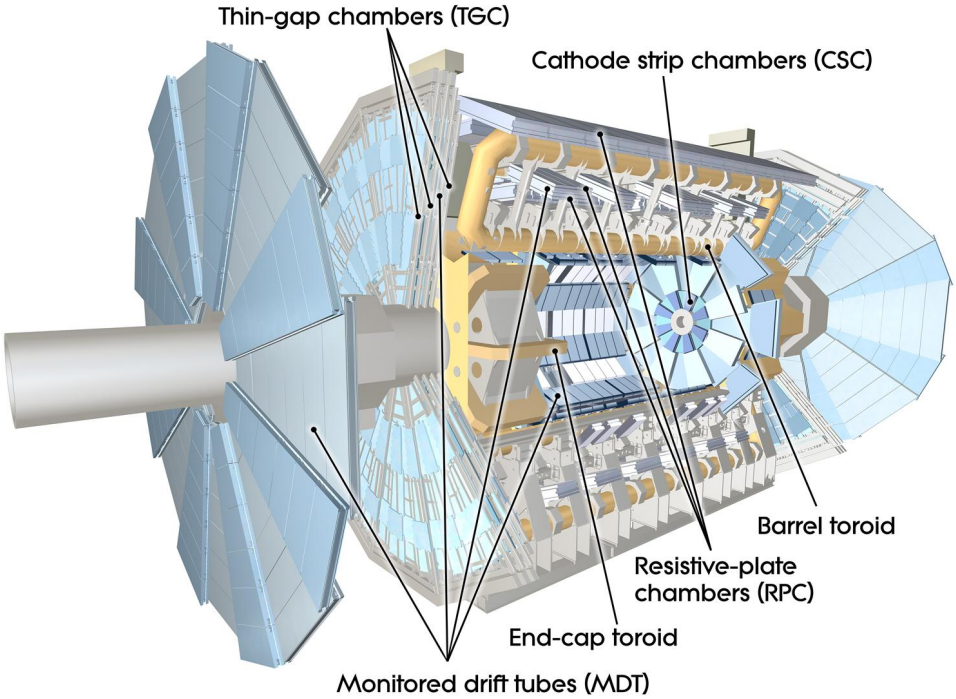


Figure 13: An illustration of components of the Muon Spectrometer of the ATLAS Detector [29].

muons, electrons, photons, jets and hadronically decaying τ leptons with high transverse momentum, as well as a large total and missing transverse energy [36]. This decision is made in less than 2.5 microseconds, and reduces the data rate to approximately 75 kHz. In addition, the L1 trigger defines regions-of-interest (RoI) which marks regions in the $\eta - \phi$ plane where interesting features have been identified. This information is then used by the L2 trigger [29].

The L2 trigger and the event filter collectively make up the High-Level Trigger (HLT) [40]. The L2 trigger is a software-based trigger, which uses the full detector information in the RoI defined by L1 trigger and further reduces data rate to approximately 3.5 kHz with an average event processing time of around 40 ms. The event filter is also a software-based trigger, which further select events down to a rate of 200 Hz.

4 Reconstruction and Identification of Physics Objects

Raw signals recorded by the ATLAS detector are reconstructed and identified as specific physics objects using various methods and algorithms. This section describes the identification and reconstruction of electrons, muons, jets, τ leptons, missing transverse energy and the Higgs boson momentum.

4.1 Electrons

Electrons are reconstructed from topological energy clusters in the EM calorimeter that are matched with tracks reconstructed in the ID. Energy is measured from energy deposits in the EM calorimeter, while energy-loss effects such as Bremsstrahlung are considered through a calibration after this matching is done [11]. The track of the electron is reconstructed using information provided by the ID. The electron reconstruction efficiency increases from 97% to 99% over an E_{T} range of 15-50 GeV [11].

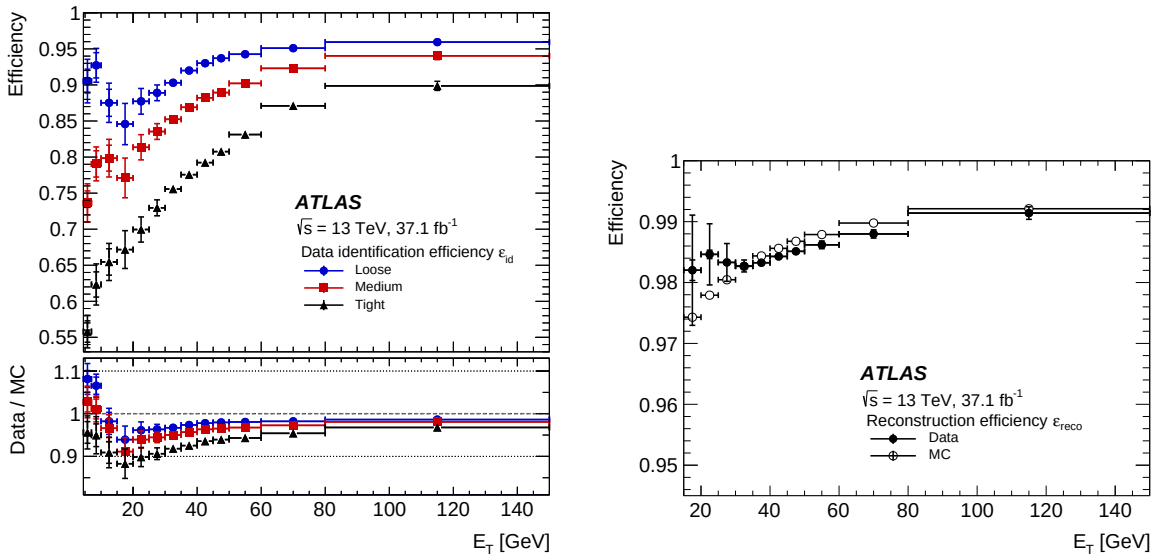
Electrons in the transition region in the $|\eta|$ range between 1.37 and 1.52 are excluded. Furthermore, one only considers electrons with $|\eta| < 2.47$. Electrons are also required to fulfill a transverse momentum of at least 15 GeV.

Identification of electrons is based on a likelihood-based discriminant [11]. This discriminant is computed from information provided by the ID and EM calorimeter. By applying a set of requirements on this discriminant, four operating points are defined: *Veryloose*, *Loose*, *Medium* and *Tight*, given in order of decreasing signal efficiency (the ratio of correctly identified electrons to the total number of electrons), and increasing background rejection (the ratio of the number of other particles not identified as electrons to their total number.) [11]. These correspond to efficiencies of 93%, 88%, and 80% for identifying a prompt electron with a p_{T} of 40 GeV, for the *Loose*, *Medium* and *Tight* working points, respectively. Therefore Tight electrons have the highest signal purity, but suffer from low signal efficiency. Electron reconstruction and identification efficiencies as a function of the transverse energy is shown in Figure 14.

The isolation of a particle generally refers to the amount of deposited energy and momentum in a cone of radius R around the particle. Applying conditions on the isolation can limit the contributions from non-prompt background processes and misidentified jets. Isolation criteria can be calorimeter-based, track-based or both. For this analysis the calorimeter-based isolation criteria, the calorimeter isolation energy $E_{\mathrm{T}}^{\mathrm{cone0.2}}$ is used, which is defined as the sum of the transverse energy deposited in calorimeter in a radius of $R = 0.2$ around the electron candidate [11].

For track-based isolation criteria, similarly to the calorimeter-based isolation, the sum of transverse momenta measured within a cone, $p_{\mathrm{T,\,var}}^{\mathrm{isol}}$, but with a variable radius depending on the p_{T} of the electron, determined by the equation:

$$R = \min\left(\frac{10\ \mathrm{GeV}}{p_{\mathrm{T}}[\mathrm{GeV}]}, R_{\mathrm{max}}\right), \tag{49}$$



(a) (b)

Figure 14: The electron identification efficiencies (ϵ_{id}) for the *Loose*, *Medium* and *Tight* criteria (a) and reconstruction efficiencies (ϵ_{reco}) (b) for data and Monte-Carlo simulated samples as a function of the transverse energy of the electron (E_{T}). Both statistical and systematic uncertainties are shown [11].

where for electrons $R_{\text{max}} = 20$ GeV is used. The working point *FCLoose* is defined such that the ratio of the $E_{\text{T}}^{\text{cone0.2}}$ and $p_{\text{T, var}}^{\text{isol}}$ to the electron's transverse momenta is less than 0.2 and 0.15, respectively [22].

For this analysis, the baseline definition of an electron has *Medium* ID requirements, with *FCLoose* isolation [11].

4.2 Muons

Muons distinguish themselves from other particles in that they deposit little energy in the calorimeters and leave clear signatures in the inner detector and the muon spectrometer. Tracks in the muon spectrometer are reconstructed using hit information in different components of the spectrometer. After the tracks are built, muon reconstruction is performed by dedicated algorithms, using information from the ID, calorimeters and muon spectrometer [19].

Similar to electrons, identification working points are defined for muons based on a likelihood-based discriminant. The *Loose*, *Medium* and *Tight* identification working points are defined, corresponding to efficiencies of 98%, 97% and 93% for identifying a muon with a $p_{\text{T}} > 100$ GeV [20].

The isolation working point *FCTightTrackOnly* is defined such that the ratio of $p_{\text{T, var}}^{\text{isol}}$ to the transverse momenta is less than 0.06, with a R_{max} of 0.3. No requirement on the calorimeter-based isolation is set [22].

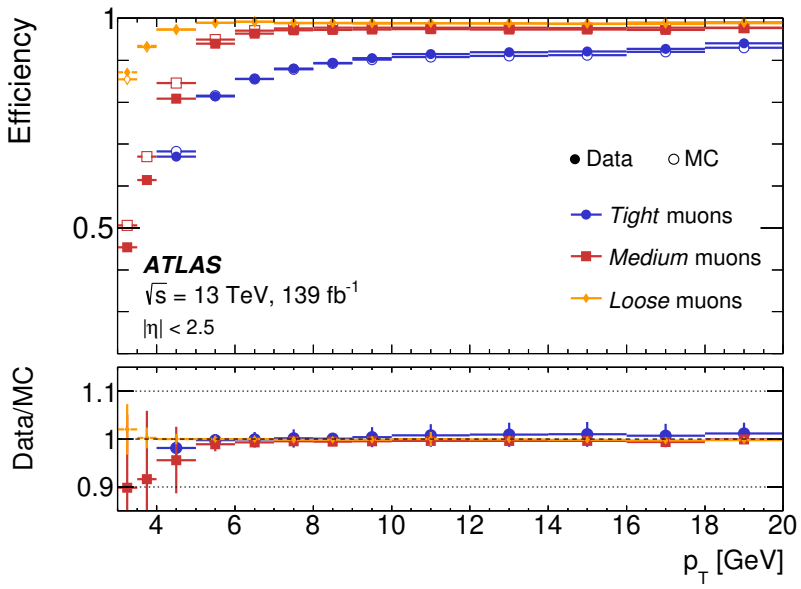


Figure 15: The combined muon identification and reconstruction efficiency as a function of the muon transverse momentum for the *Tight*, *Medium* and*Loose* muon identification operating points for both data and monte-carlo simulated samples. Errors represent both statisitical and systematic uncertainties [19].

In this thesis, the baseline definition of a muon has *Loose* ID requirements, $p_{\mathrm{T}} > 10$ GeV, $|\eta| < 2.47$ and *FCTightTrackOnly* isolation. A plot of the combined muon identification and reconstruction efficiency as a function of the transverse energy is shown in Figure 15.

4.3 Jets

Jets are reconstructed from particle flow objects, which combine signals from both the tracker and the calorimeter [14], and topological energy clusters using the anti-k_t algorithm [37] with a radius parameter of $R = 0.4$. Jets which survive the overlap removal (see Section 4.7) are subjected to additional requirements from the Jet Vertex Tagger (JVT) tool [26] in order to suppress jets originating from pile-up vertices (see Section 3.1).

Jets are required to have $p_{\mathrm{T}} > 20$ GeV, $|\eta| < 4.5$, pass the *LooseBad* [10] quality criteria and *Tight* JVT [26] requirements.

Jets originating from b-hadrons, also known as b-jets, have different topologies, which can be exploited to determine whether a jet originates from a b-hadron. Hadrons containing bottom quarks have a relatively longer lifetime which enables them to travel a certain distance before decaying, causing the point of decay to be shifted from the nominal pp collision point. Furthermore, the large mass of the bottom quark causes the decay products to have large transverse momenta. Such characteristics are used by the DL1r algorithm [74] to assign each jet with $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.5$, a score corresponding to the probability that this jet is a b-jet. This identification process is called *b-tagging*. For this analysis, the working point corresponding to a tagging efficiency of 85% is used.

4.4 Missing Transverse Energy

Neutrinos are not detected by the ATLAS detector, and therefore their properties must be investigated indirectly. The momentum carried away by the neutrinos results in an overall imbalance in the measured transverse momenta of the entire system. This can be estimated by calculating the negative vectorial sum of the transverse momenta of all the reconstructed physics objects and of tracks associated to the primary vertex but not associated with any of the reconstructed objects, known as the *soft term* [24].The missing transverse energy ($\vec{E}_{\mathrm{T}}^{miss}$), which is then defined as:

$$\vec{E}_{\mathrm{T}}^{miss} = \vec{E}_{\mathrm{T}}^{miss,e} + \vec{E}_{\mathrm{T}}^{miss,\gamma} + \vec{E}_{\mathrm{T}}^{miss,\tau} + \vec{E}_{\mathrm{T}}^{miss,jet} + \vec{E}_{\mathrm{T}}^{miss,soft} + \vec{E}_{\mathrm{T}}^{miss,\mu}. \tag{50}$$

The magnitude of the missing transverse energy vector is denoted as E_{T}^{miss}, and it computed as:

$$E_{\mathrm{T}}^{miss} = \sqrt{\left(E_{\mathrm{T},x}^{miss}\right)^2 + \left(E_{\mathrm{T},y}^{miss}\right)^2} \tag{51}$$

where $E_{\mathrm{T},x/y}^{miss}$ refers to the x and y components of the missing transverse energy vector.

4.5 Hadronic τ lepton decays

The τ lepton can decay leptonically into a lepton and two neutrinos, or hadronically. The charged leptonic decay products are reconstructed as electrons or muons, whereas neutrinos are accounted for via the missing transverse energy. In the hadronic decay mode, τ leptons decay into a τ neutrino along with one or more hadrons, predominantly in the form of pions. Depending on the number of pion tracks, a hadronic τ lepton decay can be classified into one- or multi-prong (usually 3). The visible decay products of the hadronic τ lepton decay are denoted as $\tau_{\text{had-vis}}$. Although not directly investigated in the scope of this thesis, a veto on $\tau_{\text{had-vis}}$ objects is applied.

$\tau_{\text{had-vis}}$ candidates are reconstructed using the anti-k_t algorithm with a radius parameter of $R = 0.4$. $\tau_{\text{had-vis}}$ candidates may be hard to distinguish from jets, and therefore multivariate techniques exploiting characteristics of hadronic τ decays such as their narrow calorimeter clusters, isolation and their charged track multiplicity are used [71] . For this, Boosted Decision Trees (BDTs) were trained to reject quark- or gluon-initiated background processes. Three working points: *Loose*, *Medium* and *Tight*, are defined with signal efficiencies for one-prong (multi-prong) $\tau_{\text{had-vis}}$'s of 70% (65%), 60% (55%), and 40% (35%) respectively [71]. In this thesis, the $\tau_{\text{had-vis}}$ objects must satisfy the *Medium* working point and furthermore have one or three associated tracks with a $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.47$, excluding the pseudorapidity region $1.37 < |\eta| < 1.52$.

4.6 Invariant Mass of the Higgs-Boson Candidate

The invariant mass of the ditau system is an important discriminant in the selection of Higgs events from background events. However, since the leptonic decay of the τ lepton

involves the production of two neutrinos whose momenta is not directly measured by the detectors, the invariant mass of the ditau system cannot be directly calculated. Therefore the invariant mass of the Higgs boson is reconstructed with two different methods.

4.6.1 Collinear Approximation

The collinear approximation is based on the assumption that the decay products of the τ lepton is collinear with the τ lepton itself [13]. This assumption holds since the mass of the Higgs boson is much greater than the mass of two τ leptons and therefore one expects the τ leptons to have large momenta. Another assumption is that the missing transverse energy is solely due to neutrinos from the τ lepton decay. The invariant mass of the ditau system in the collinear approximation (m_{coll}) can be written as:

$$m_{coll} = \frac{m_{vis}}{\sqrt{x_1 x_2}}, \tag{52}$$

where m_{vis} is the invariant mass of the visible decay products of the ditau system and $x_{1(2)}$ the momentum fractions of each τ lepton carried away by the visible decay products, i.e.:

$$\vec{p}_{vis,1(2)} = x_{1(2)}\vec{p}_{\tau 1(2)} \tag{53}$$

The momentum fractions in the collinear approximation can be written as:

$$x_{1(2)} = \frac{p^x_{vis,2}p^y_{vis,1} - p^y_{vis,2}p^x_{vis,1}}{p^x_{vis,2}p^y_{vis,1} + (-)E^{miss}_{\mathrm{T},x}p^x_{vis,1(2)} - p^y_{vis,2}p^x_{vis,1} - (+)E^{miss}_{\mathrm{T},y}p^x_{vis,1(2)}}, \tag{54}$$

where $p^{x,y}_{vis,1(2)}$ are the x and y components of $\vec{p}_{vis,1(2)}$.

This approximation works well for when the ditau mass is boosted, i.e. has a large momentum. However, events with back-to-back topology, $\Delta\phi(\tau_1, \tau_2) \approx \pi$, Equation 54 results in unphysical solutions, and this approximation cannot be used. Many $H \to \tau\tau$ events, which have a back-to-back topology, must therefore be discarded. Additionally, the collinear approximation is sensitive to E^{miss}_{T} reconstruction resolutions, therefore resulting in long tails in the reconstructed ditau mass distribution [50], thereby reducing the discrimination capacity.

4.6.2 The Missing Mass Calculator

The Missing Mass Calculator (MMC) [50] aims to improve upon the shortcomings of the collinear approximation method and allows for a complete reconstruction of event kinematics of the ditau system.

The full reconstruction of the ditau kinematics in the dilepton channel involves finding 8 unknown variables: x, y, and z components of the τ neutrinos for each of the τ-leptons and the invariant masses of the two neutrinos from each leptonic decay. However, there are only four independent equations, which relate these unknowns:

$$E_{\mathrm{T},x}^{miss} = p_{miss,1}\sin(\theta_{miss,1})\cos(\phi_{miss,1}) + p_{miss,2}\sin(\theta_{miss,2})\cos(\phi_{miss,2}) \tag{55}$$

$$E_{\mathrm{T},y}^{miss} = p_{miss,1}\sin(\theta_{miss,1})\sin(\phi_{miss,1}) + p_{miss,2}\sin(\theta_{miss,2})\sin(\phi_{miss,2}) \tag{56}$$

$$\begin{aligned} m_{\tau_1}^2 &= m_{miss_1}^2 + m_{vis,1}^2 + 2\sqrt{p_{miss,1}^2 + m_{miss,1}^2}\sqrt{p_{vis,1}^2 + m_{vis,1}^2} \\ &\quad - 2p_{vis,1}p_{miss,1}\cos(\theta_{vis,1} - \theta_{miss,1}) \end{aligned} \tag{57}$$

$$\begin{aligned} m_{\tau_2}^2 &= m_{miss_2}^2 + m_{vis,2}^2 + 2\sqrt{p_{miss,2}^2 + m_{miss,2}^2}\sqrt{p_{vis,2}^2 + m_{vis,2}^2} \\ &\quad - 2p_{vis,2}p_{miss,2}\cos(\theta_{vis,2} - \theta_{miss,2}) \end{aligned} \tag{58}$$

where the index $miss, 1(2)$ refers to the neutrinos from the decay of $\tau_{1(2)}$ and vis, i the visible part of the decay of $\tau_{1(2)}$, which are the charged leptons, while $m_{miss_{1(2)}}$ refers to the invariant mass of the neutrinos from the lepton decays of $\tau_{1(2)}$.

Since it is impossible to find a unique solution with more unknowns than constraints, the MMC compares the likeliness of different solutions by considering the characteristics of τ lepton decays. This is done by scanning a four dimensional parameter space of all solutions, corresponding to the degrees of freedom of the system of equations, and assigning an event probability to each phase space point based on the probability density functions for τ lepton decay properties. The solution with the highest probability is then used to calculate the final reconstructed ditau mass m_{MMC}.

The MMC allows for the calculation of the reconstructed ditau invariant mass for any configuration, as opposed to the collinear approximation method. It additionally results in a better resolution and accuracy of the ditau mass in comparison with the collinear approximation method [50].

4.7 Overlap Removal

It is possible that multiple physics object fulfill the identification and reconstruction requirements. In order to resolve this ambiguity, an overlap removal procedure is applied. One reconstructed object is compared with another, and rejected if it fulfills a certain criteria. Through a 9-step procedure, every reconstructed object type is compared with another, allowing in the end only one reconstructed object type. The overlap removal criteria for the *Standard* of the *AssociationUtils* package [17], which is used for this thesis, can be seen in Table 2.

Step	Object to remove	Object to keep	Criteria
1	electron	electron	If they share the same track, the electron with the highest transverse momentum is kept.
2	$\tau_{\text{had-vis}}$	electron	If $\Delta R < 0.2$, the electron is kept
3	$\tau_{\text{had-vis}}$	muon	If $\Delta R < 0.2$, the muon is kept
4	electron	muon	If they share a track, the electron is removed if the muon is associated with a signature in the muon spectrometer, otherwise the muon is removed.
5	jet	electron	Any jet within $\Delta R < 0.2$ of an electron is removed.
6	jet	muon	Any jet within $\Delta R < 0.2$ of an electron is removed if it has fewer than three associated tracks.
7	electron	jet	Any electron within $\Delta R < 0.4$ of a jet is removed.
8	muon	jet	Any muon within $\Delta R < 0.4$ of a jet is removed.
9	jet	$\tau_{\text{had-vis}}$	Any jet within $\Delta R < 0.2$ of a $\tau_{\text{had-vis}}$ is removed.

Table 2: The steps of the overlap removal in the *Standard* working point using the *AssociationUtils* package. Table is taken from [17].

5 Signal and Background Processes

Various processes occur in the ATLAS detector arising from pp-collisions. Many of these processes are simulated using Monte-Carlo simulations (MC) for use in the analysis. This section aims to define the signal processes, which is the process of interest to be analyzed as well as other processes, which may leave similar signatures in the detectors as those of signal processes, constituting the background processes of the signal process.

5.1 Signal Processes

This analysis considers the Vector Boson Produced (VBF) $H \to \tau\tau \to e\mu 4\nu$ as the signal process. One also considers $H \to WW \to e\mu 2\nu$ as a signal process as well since it can also be used to increase the sensitivity to CP violation in the HVV vertex [28].

The final reconstructed topology of VBF produced Higgs decay into $\tau\tau \to e\mu 4\nu$ is characterized by two charged leptons of opposite sign with missing transverse energy, associated with at least 2 jets with a large angle separation. A Feynman diagram of this process can be seen in Figure 16.

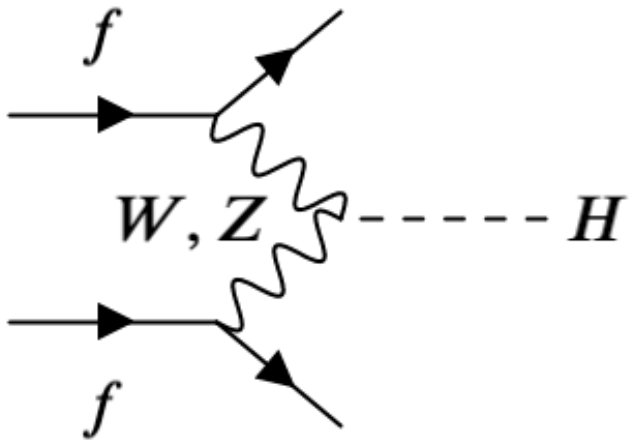


Figure 16: An example Feynman diagram of Vector Boson Fusion production of the Higgs boson.

5.2 Background Processes

Many processes occur within the ATLAS detector originating from pp collisions, some of which can mimic the signature of signal processes. These are termed *background* processes. Background processes can be classified into the *reducible* and *irreducible* backgrounds. Reducible backgrounds have different final states from the signal process and therefore can be separated effectively. Irreducible backgrounds on the other hand have the same final state as the signal process, making separation challenging. Furthermore,

detector effects and limitations in identification and reconstruction can also cause reducible backgrounds to appear like they have the same final state as signal processes. Background processes include $Z \to \tau\tau$, $Z \to ll$, $(l = e, \mu)$, diboson production, top-quark background, misidentified leptons, and other Higgs boson production modes. Gluon-Gluon Fusion, Higgs-Strahlung and production in association with a pair of top quarks are grouped into the *Other Higgs Boson production modes* category. Apart from contributions from misidentified leptons, all background processes are estimated through MC simulations, which are discussed below. The contribution of misidentified leptons are estimated through a data-driven method called the Matrix Method.

5.2.1 $Z \to \tau\tau$

The $Z \to \tau\tau$ process represents the most dominant background contribution, since it produces two τ-leptons as in the case of the signal process. Similarly to the Higgs boson, the Z boson can also be produced via vector-boson fusion, thereby also resulting in two VBF jets. This processes is therefore an irreducible background process. An example Feynman diagram of a VBF produced $Z \to \tau\tau$ process can be seen in figure 17

The contribution of such processes can be suppressed by applying requirements on the invariant mass of the decay products, due to the differing mass between the Z boson and the Higgs Boson.

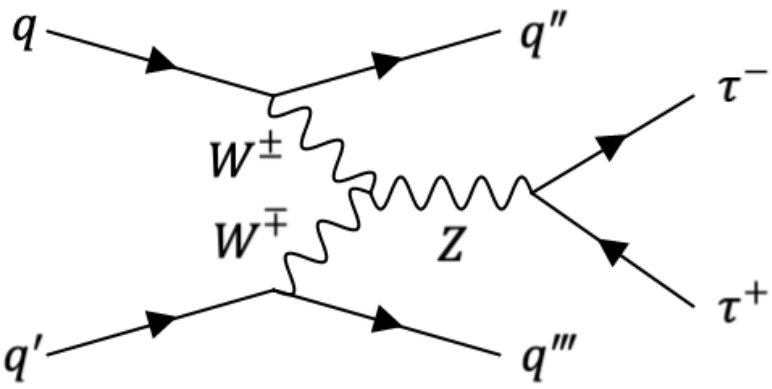


Figure 17: Example Feynman diagram of VBF $Z \to \tau\tau$ process.

5.2.2 Top-quark Production

Single top and top-quark pair productions are collectively referred to as the Top background, whose example Feynman diagrams are shown in Figure 19. Both can result in a topology that is similar to that of the signal process, since a top-quark almost always decays into a bottom-quark and a W boson. Therefore in the case of top quark pair production, both of these top quarks can produce two jets, two leptons of opposite sign and missing transverse energy. An example Feynman diagram of top-quark pair production can be seen in Figure 18. In single-top quark production, top-quarks can be produced in

three channels, called the t, s and tW-channels, whose Feynman diagrams can be seen in Figure 19. Additional jets may be produced via final state radiation, associated with two leptons and missing transverse energy.

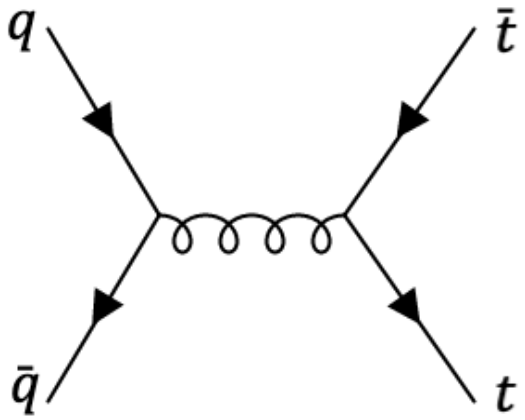
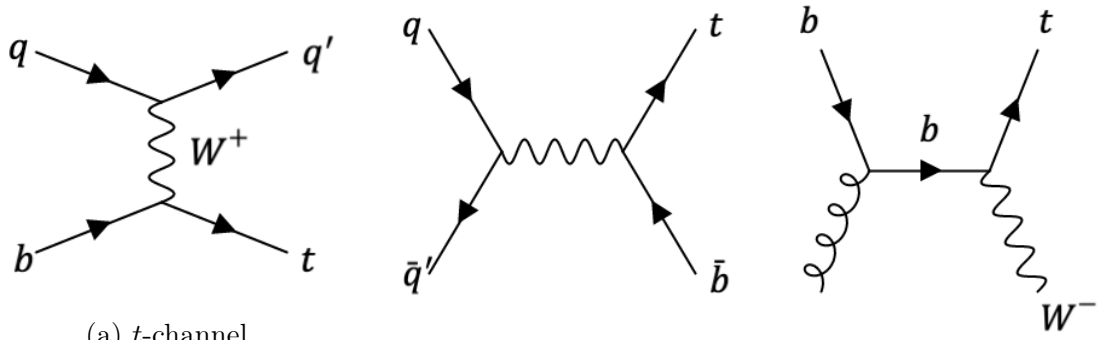


Figure 18: Top-quark pair production.



(a) t-channel. (b) s-channel. (c) tW-channel.

Figure 19: Example Feynman diagrams of different single top production channels.

This process can be suppressed by applying a veto on b-tagged jets. However, it nonetheless represents a sizable background contribution due to the limited efficiency of b-jet identification. Furthermore, since the nature of the jet production associated with top background are different from that of VBF processes, VBF processes can still be discriminated from top processes through jet properties such as the transverse momentum.

5.2.3 $Z \to ll$

$Z \to ll$ refers collectively to the $Z \to ee$, and $Z \to \mu\mu$ processes processes. An example Feynman diagram of the $Z \to ll$ process is shown in Figure 20. The figure shows how a Z boson could be produced through VBF, just like the Higgs boson in the signal process, resulting in two jets and two leptons. Another possibility is the fusion of a quark-antiquark pair into a Z boson which then decays into a lepton-antilepton pair.

Two additional jets may be produced by initial state radiation, resulting in two jets and two leptons of opposite charge. This process is more common than the VBF production mode. The $Z \to ee$ and $Z \to \mu\mu$ processes have the largest cross section among the processes considered in this analysis, however these processes by themselves cannot produce two leptons of different flavors, and therefore its contribution as a background process is limited. It is nonetheless possible that $Z \to ee$ or $Z \to \mu\mu$ processes mimic signal processes through detector effects, such as lepton misidentification. These processes also do not produce missing transverse momentum. However, a possible incorrect reconstruction of the missing transverse momentum could produce an apparent E_{T}^{miss}.

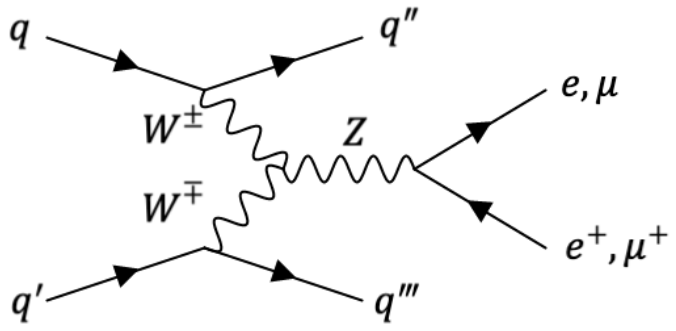
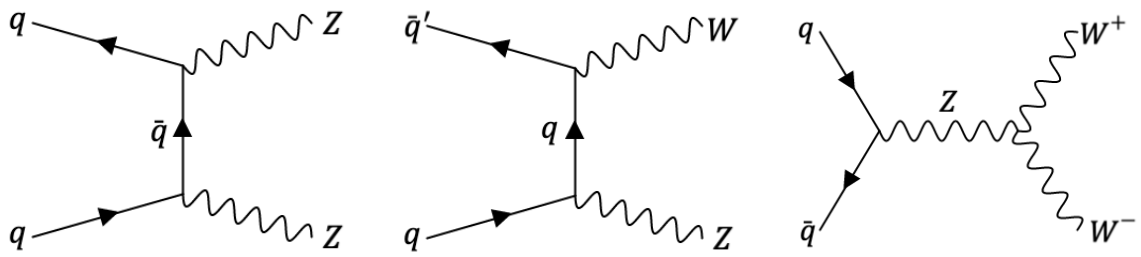


Figure 20: Example Feynman diagram of VBF produced $Z \to ee$ or $Z \to \mu\mu$

5.2.4 Diboson Production

Diboson production (VV) refers to the production of a pair of vector bosons, i.e. ZZ, WZ or WW production. Example Feynman diagrams of diboson production can be seen in figure 21. In the case of Figures 21a and 21b the Z boson can decay into two leptons of the same flavor, associated with two jets and no missing transverse energy. However, similarly to $Z \to ll$, it is still possible for this process to mimic the signature of signal processes through detector effects. The nature of jets produced from such processes are different from VBF processes, since in VBF the two jets are produced from scattering of quarks while in processes shown in Figures 21a and 21b, they are produced the decay of a Z boson. Therefore the contribution of such processes can be suppressed by examining jet properties such as angle and transverse momentum. The process shown in Figure 21b shows a case where a Z boson decays into two W bosons, which then can decay into two different-flavor leptons. Two jets can also be produced through initial state radiation of two gluons. In such a case two opposite sign leptons of different flavors are produced along with missing transverse momentum and two jets. This process also has jet properties different from VBF processes, such as the invariant mass. The invariant mass of the system of leptons and missing transverse momenta would also correspond to the mass of the Z boson, instead of the Higgs boson. Therefore VV processes can be suppressed through consideration of such kinematic variables.



(a) ZZ production. (b) WZ production. (c) WW production.

Figure 21: Example Feynman diagrams of (a), ZZ production, (b), WZ production and (c), WW production, which may produce a final state similar to that of signal processes.

5.2.5 W Boson Production

W bosons can also be produced in pp-collisions, however, its contribution in this analysis is limited, since a W boson by itself cannot produce two different-flavored leptons with opposite charge. It is however nonetheless possible, that a jet is misidentified as a lepton, therefore creating two apparent leptons. An example Feynman diagram of W boson production in association with two jets can be seen in Figure 22.

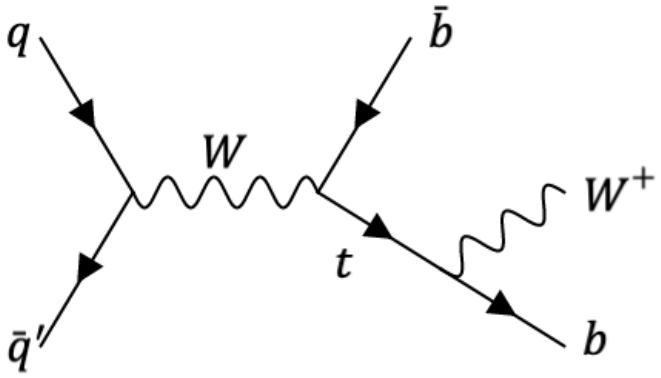


Figure 22: An example Feynman diagram of W boson in association with two jets.

Events from W boson production are not directly used for the background estimation, since these overlap with events originating from misidentified leptons. However, for certain calculations in the estimation of contribution from misidentified leptons, they are considered.

5.3 Event Generation

Apart from the contribution from misidentified leptons, all background processes are estimated through Monte Carlo (MC) simulations. For each process, a Monte Carlo generator is used for the matrix element, using a parton distribution function (PDF) set, which can be estimated at various degrees of leading orders. The parton shower and

hadronization are modeled by the Underlying Event Parton Shower (UEPS) model. In this thesis, the Powheg-Box v1 [80], v2 [75][76][55][5] and Sherpa 2.2.1 [64] generators are used. The following PDF sets are used: PDF4LHC1[60][35], PDF4LHC15 NNLO [32], NNPDF2.3 NLO [72], and NNPDF3.0NNLO [31] are used. For the UEPS model, Pythia 8 [88] and Sherpa 2.2.1 are used. Detector effects are simulated using GEANT4 [1]. A summary of the MC generators used for the considered processes can be seen in table 3, and a table listing the measured cross sections of the considered processes can be seen in table 4.

Process	Generator	PDF Set	UEPS Model
VBF	Powheg-Box v2	PDF4LHC15 NLO	Pythia 8
ggH	Powheg-Box v2	PDF4LHC15 NNLO	Pythia 8
$t\bar{t}H$	Powheg-Box v2	PDF4LHC15 NNLO	Pythia 8
VH	Powheg-Box v2	NNPDF2.3 NLO	Pythia 8
Z	Sherpa 2.2.1	NNPDF3.0 NNLO	Sherpa 2.2.1
W	Sherpa 2.2.1	NNPDF3.0 NNLO	Sherpa 2.2.1
VV	Sherpa 2.2.1	NNPDF3.0 NNLO	Sherpa 2.2.1
Single top	Powheg-Box v1	NNPDF2.3 NLO	Pythia 8
Top pair	Powheg-Box v2	NNPDF2.3 NLO	Pythia 8

Table 3: An overview of the Monte Carlo generators used to generate samples for processes considered in this thesis.

Process	Cross section (× BR) [pb]	Ref.
W^--production ($\times BR_{lep}$)	11.83 ± 0.02 (stat) ± 0.32 (sys) ± 0.25 (lumi) $\times 10^3$	[16]
W^+-production ($\times BR_{lep}$)	8.79 ± 0.02 (stat) ± 0.24 (sys) ± 0.18 (lumi) $\times 10^3$	[16]
Z-production ($\times BR_{lep}$)	1.981 ± 0.007 (stat) ± 0.038 (sys) ± 0.042 (lumi) $\times 10^3$	[16]
Top quark pair production	781 ±7 (stat) ± 62(sys) ± 20(lumi)	[43]
Single top production	219.0 ± 1.5(stat)± 13.0(sys)	[42]
WW production	115±5.8(stat)±5.7(exp)±6.4(theo)±3.6(lumi)	[65]
tW production	63.1 ± 1.8(stat) ± 6.4(sys) ± 2.1(lumi)	[87]
WZ production	$40.9 \pm 3.4(\text{stat})^{+3.1}_{-3.3}(\text{sys}) \pm 0.4(\text{theo}) \pm 1.3(\text{lumi})$	[65]
ZZ production	$14.6^{+1.9}_{-1.8}(\text{stat})^{+0.5}_{-0.3}(\text{sys}) \pm 0.2(\text{theo}) \pm 0.4(\text{lumi})$	[65]

Table 4: The cross section of different background processes ordered by their respective cross section. For certain processes the product of the cross section and leptonic branching ratio is listed. Uncertainties are explained in their respective references.

6 Estimation of Misidentified and Non-Prompt Leptons

Background contribution from misidentified jets and non-prompt leptons, also known as fakes, mainly originate from W boson production in association with jets and multi-jet QCD production, which refer to events where several jets are strongly produced. In contrast with other background processes, fake processes are poorly modeled in MC simulations [12]. Therefore data-driven methods often are used to estimate the contribution from fake events. In this thesis, the *matrix method* is used to estimate the yield of fake events.

6.1 The Matrix Method

The contribution of real and fake leptons cannot be directly measured experimentally. Therefore one has to determine another property, which can be measured experimentally, and relate that with the realness and fakeness of leptons to find the number of fake events. For this, of *tight* (T) and *loose* (L) criteria are defined per lepton flavor, such that leptons which pass tight selection requirements have a higher chance of being a real lepton than those that pass loose requirements. The definition of *tight* and *loose* are exclusive from each other, meaning that a loose lepton can also be interpreted as being *not-tight*, while passing baseline requirements. In this analysis, tight leptons are required to fulfill the same isolation and identification requirements as leptons in the signal region (See Section 7). For electrons this is Medium ID requirement with FCLoose isolation, and for muons medium ID requirements with FCTightTrackOnly isolation. In the loose region, the isolation requirements for both lepton flavors are removed, and for electrons, the ID requirement is also loosened to Loose. The removal of the isolation requirements is motivated by the fact that processes such as misidentification of hadrons as leptons and semileptonic decays of heavy quarks are characterized by activity in a large ΔR surrounding the candidate object. On the other hand, prompt leptons are characterized by little activity in the calorimeter and inner detector surrounding the reconstructed object in ΔR [12].

The probability that a real or fake lepton is classified as a tight lepton is then referred to as the real (ϵ_r) and fake efficiencies (ϵ_f). They can be calculated in dedicated control regions, which are dominated by real or fake leptons, in the following way:

$$\epsilon_{r,f} = \frac{N^T}{N^{\bar{L}}} \tag{59}$$

where N^T is the number of tight events in their respective control region and $N^{\bar{L}}$ the number of events which fulfill the baseline requirements. $\bar{L}$ therefore refers to the set of events passing baseline requirements, i.e. the set of tight or loose events.

In the single lepton case, one can use Equation 59 to derive the following equations:

$$N^T = N_r\epsilon_r + N_f\epsilon_f \tag{60}$$

$$N^L = N_r(1-\epsilon_r) + N_f(1-\epsilon_f) \tag{61}$$

where N_r and N_f are the number of real and fake leptons, respectively. In Equation 61, the $(1-\epsilon_{r,f})$ term originates from the fact that the definition of *loose* is orthogonal to that of *tight*.

Above equations can be generalized to the two lepton case through the following system of equations, expressed as a product of matrices:

$$\begin{bmatrix} N_{TT} \\ N_{TL} \\ N_{LT} \\ N_{LL} \end{bmatrix} = \begin{bmatrix} \epsilon_r\epsilon_r & \epsilon_r\epsilon_f & \epsilon_f\epsilon_r & \epsilon_f\epsilon_f \\ \epsilon_r\bar{\epsilon_r} & \epsilon_r\bar{\epsilon_f} & \epsilon_f\bar{\epsilon_r} & \epsilon_f\bar{\epsilon_f} \\ \bar{\epsilon_r}\epsilon_r & \bar{\epsilon_r}\epsilon_f & \bar{\epsilon_f}\epsilon_r & \bar{\epsilon_f}\epsilon_f \\ \bar{\epsilon_r}\bar{\epsilon_r} & \bar{\epsilon_r}\bar{\epsilon_f} & \bar{\epsilon_f}\bar{\epsilon_r} & \bar{\epsilon_f}\bar{\epsilon_f} \end{bmatrix} \begin{bmatrix} N_{rr} \\ N_{rf} \\ N_{fr} \\ N_{ff} \end{bmatrix} \tag{62}$$

where $\bar{\epsilon} = 1-\epsilon$ and N_{XY} the number of events in lepton categories X or Y, ordered by p_{T}. The first and second efficiency term in the product always refers to the first and second lepton, respectively. The matrix containing the product of efficiencies is termed the efficiency matrix M.

By inverting the matrix, one can estimate the number of events with at least one fake lepton (N_{rf}, N_{fr}, N_{ff}) from the number of events with tight and/or loose leptons. The inverse of the matrix was found to be:

$$M^{-1} = \frac{1}{(\epsilon_{r,1}-\epsilon_{f,1})(\epsilon_{r,2}-\epsilon_{f,2})} \begin{bmatrix} \bar{\epsilon}_{f,1}\bar{\epsilon}_{f,2} & \bar{\epsilon}_{f,1}\epsilon_{f,2} & -\epsilon_{f,1}\bar{\epsilon}_{f,2} & \epsilon_{f,1}\epsilon_{f,2} \\ -\bar{\epsilon}_{f,1}\bar{\epsilon}_{r,2} & \bar{\epsilon}_{f,1}\epsilon_{r,2} & -\epsilon_{f,1}\bar{\epsilon}_{r,2} & -\epsilon_{f,1}\epsilon_{r,2} \\ -\bar{\epsilon}_{r,1}\bar{\epsilon}_{f,2} & \bar{\epsilon}_{r,1}\epsilon_{f,2} & -\bar{\epsilon}_{r,1}\bar{\epsilon}_{f,2} & -\epsilon_{r,1}\epsilon_{f,2} \\ \bar{\epsilon}_{r,1}\bar{\epsilon}_{r,2} & -\bar{\epsilon}_{r,1}\epsilon_{r,2} & -\epsilon_{r,1}\bar{\epsilon}_{r,2} & \epsilon_{r,1}\epsilon_{r,2} \end{bmatrix}, \tag{63}$$

where here the numerical index of the efficiencies refer to either the first or second lepton.

The contribution of events with at least one fake lepton in the nominal region where both leptons are tight is then given by the sum of their projections into the region, where both leptons are tight.

$$N_{TT}^{fake} = \epsilon_{r,1}\epsilon_{f,2}N_{rf} + \epsilon_{f,1}\epsilon_{r,2}N_{fr} + \epsilon_{f,1}\epsilon_{f,2}N_{ff}, \tag{64}$$

which can be rewritten in following way by plugging in the inversion of the matrix and gathering terms:

$$
\begin{aligned}
N_{TT}^{fake} &= \epsilon_{r,1}\epsilon_{f,2}N_{rf} + \epsilon_{f,1}\epsilon_{r,2}N_{fr} + \epsilon_{f,1}\epsilon_{f,2}N_{ff} \\
&= \underbrace{\alpha[\epsilon_{r,1}\epsilon_{f,2}(\epsilon_{f,1}-1)(1-\epsilon_{r,2}) + \epsilon_{f,1}\epsilon_{r,2}(\epsilon_{r,1}-1)(1-\epsilon_{f,2}) + \epsilon_{f,1}\epsilon_{f,2}(1-\epsilon_{r,1})(1-\epsilon_{r,2})]}_{w_{TT}} N_{TT} \\
&+ \underbrace{\alpha\epsilon_{f,2}\epsilon_{r,2}[\epsilon_{r,1}(1-\epsilon_{f,1}) + \epsilon_{f,1}(1-\epsilon_{r,1}) + \epsilon_{f,1}(\epsilon_{r,1}-1)]}_{w_{TL}} N_{TL} \\
&+ \underbrace{\alpha\epsilon_{f,1}\epsilon_{r,1}[\epsilon_{f,2}(1-\epsilon_{r,2}) + \epsilon_{r,2}(1-\epsilon_{f,2}) + \epsilon_{f,2}(\epsilon_{r,2}-1)]}_{w_{LT}} N_{LT} \\
&- \underbrace{\alpha\epsilon_{f,1}\epsilon_{f,2}\epsilon_{r,1}}_{w_{LL}}\epsilon_{r,2}N_{LL} \\
&= w_{TT}N_{TT} + w_{TL}N_{TL} + w_{LT}N_{LT} + w_{LL}N_{LL},
\end{aligned}
$$

where $\alpha = ((\epsilon_{r,1}-\epsilon_{f,1})(\epsilon_{r,2}-\epsilon_{f,2}))^{-1}$. Therefore it can be seen that the final fake estimation amounts to weighting events in different tightness regions and then summing them up.

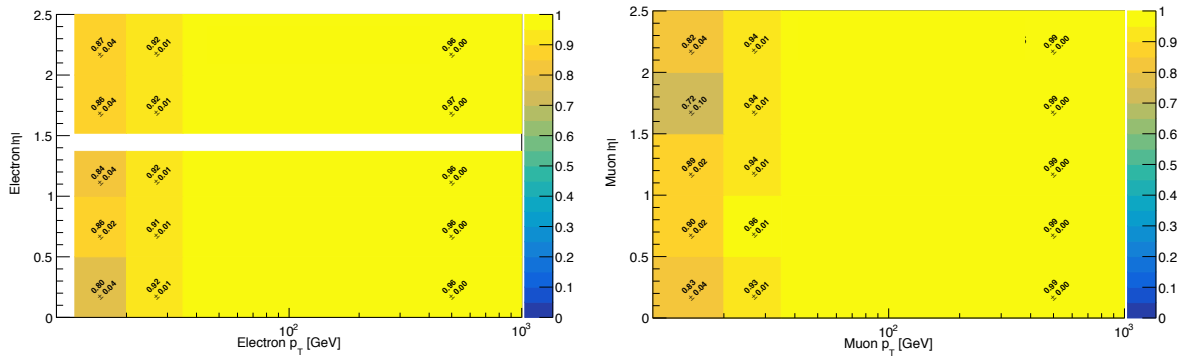
6.1.1 Real Efficiencies

The real efficiencies of each lepton flavor are measured in a control region enriched in real leptons and is calculated using leptons matched to generator-level particles, referred to as truth-matched leptons. The real CR is defined by having one electron and muon with opposite charge. The invariant mass of the dilepton system is limited to a range of ± 10 GeV within the Z-mass, in order to select $Z \to ll$ events. At least one jet with a transverse momentum of at least 40 GeV is required and a b-jet veto is applied. The samples considered for the calculation of the real efficiencies are: Diboson production, $Z \to ll$,$Z \to \tau\tau$, top production and W production.

The real efficiency is calculated as a function of the p_{T} and $|\eta|$, separately per lepton flavor. For muons, 5 equidistant bins were used in an $|\eta|$ range of 0 to 2.5. For electrons, the $|\eta|$ region between 1.37 and 1.52 is excluded since this region is excluded in the definition of an electron. For electrons, p_{T} bins with edges at 12, 20, 35 and 1000 GeV are used. For muons, bins with edges at 10, 20, 35, and 1000 GeV are used. The lowest bin edge corresponds to the baseline p_{T} requirement of the respective leptons. The p_{T} binning is set in such a way to ensure enough statistics in each bin, such that no bin content is negative. Furthermore, as the lepton p_{T} increases, one expects an increase in the real efficiency, since the identification performance of leptons improve as a function of p_{T} [11]. One expects this efficiency to plateau after a certain point, which is considered by the larger width of the last p_{T} bin.

A plot of the electron and muon real efficiencies as a function of their p_{T} and $|\eta|$ values is shown in figure 23. A figure comparing the real efficiencies of the muon and electron as a function of the p_{T} is shown in figure 24.

It can be seen that the efficiencies generally increase with increasing p_{T}, which is to be expected. It can also be seen that the muon real efficiencies are also generally higher



(a) Electrons (b) Muons

Figure 23: Real efficiencies of (a), electrons and (b), muons as a function of the lepton transverse momentum and absolute value of the pseudorapidity of the lepton. All uncertainties are statistical.

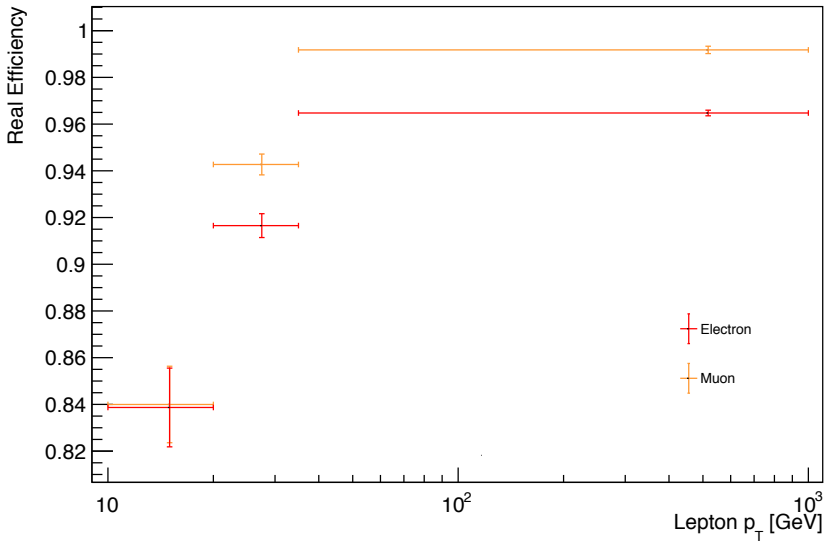


Figure 24: The electron (red) and muon (orange) real efficiencies as a function of the lepton transverse momentum. Uncertainties in y direction are statistical.

than those of electrons. This is due to the muon identification efficiencies being better than those of the electrons.

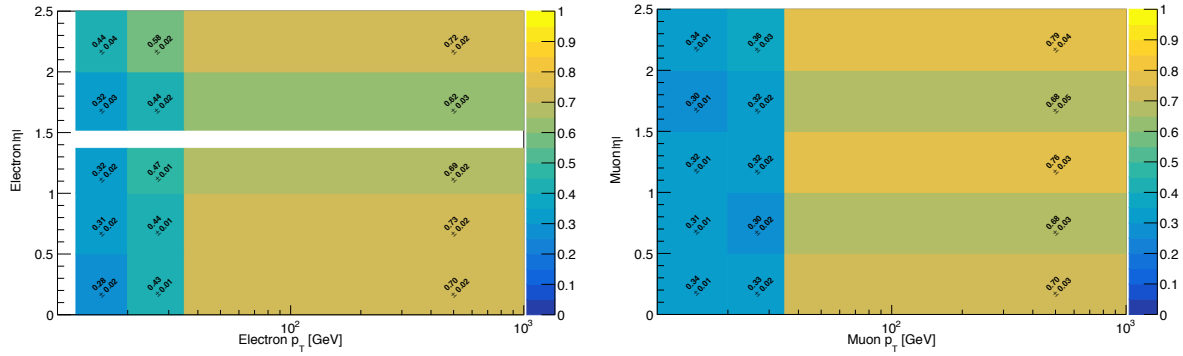
6.1.2 Fake Efficiency

As with the real efficiencies, the fake efficiencies are also measured in a control region, albeit enriched in fake leptons. The fake CR is defined as having one muon and one electron with the same charge, at least one jet with $p_{\mathrm{T}} > 40$ GeV and a veto on b-jets. Due to contamination of real leptons in the fake CR, truth-matched lepton and charge flipped electron contributions are subtracted from data. Charge flipped electrons are defined to be electrons whose MC truth charge is opposite in sign to their reconstructed charge. The efficiency can therefore be calculated as:

$$\epsilon_f = \frac{N_{Tight}^{Data} - N_{Tight}^{MC} - N_{Tight}^{CF}}{N_{Loose}^{Data} - N_{Loose}^{MC} - N_{Loose}^{CF}}, \tag{65}$$

where N^{Data}, N^{MC} the number of data events and truth-matched MC generated events, respectively and N^{CF} the number of events charge-flipped electrons. The lower index refers to whether these numbers are counted in the tight or loose region.

Processes considered for the calculation of the fake efficiency are: Diboson production, $Z \to ll$, $Z \to \tau\tau$, top production and W production. Fake efficiencies are calculated as a function of the p_{T} and $|\eta|$ for each lepton flavor, separately. The binning in $|\eta|$ and p_{T} are same as the real efficiencies. The electron and muon fake efficiencies as a function of the lepton transverse momentum and absolute value of the pseudorapidity are shown in Figure 25. The electron and muon fake efficiencies are compared as a function of the lepton p_{T} in Figure 26.



(a) Electron (b) Muon

Figure 25: Fake efficiencies of (a), electrons and (b), muons as a function of the lepton transverse momentum and absolute value of their pseudorapidity. Uncertainties are statistical.

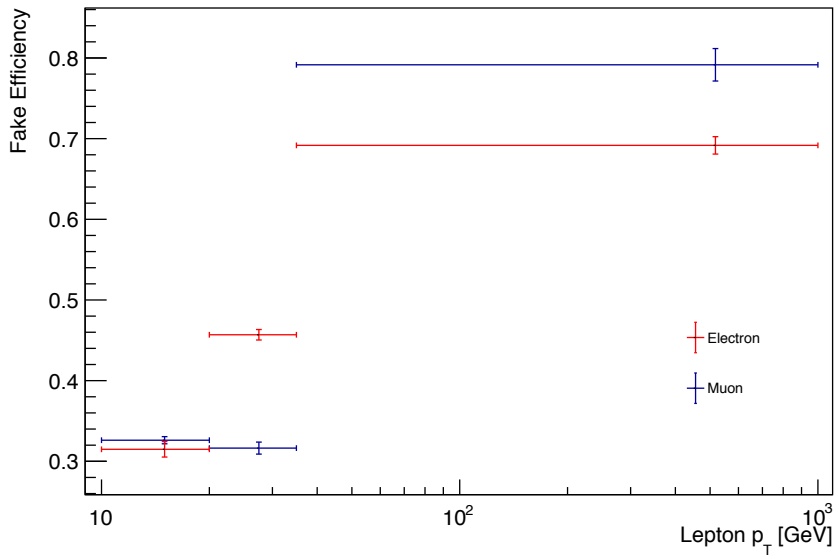
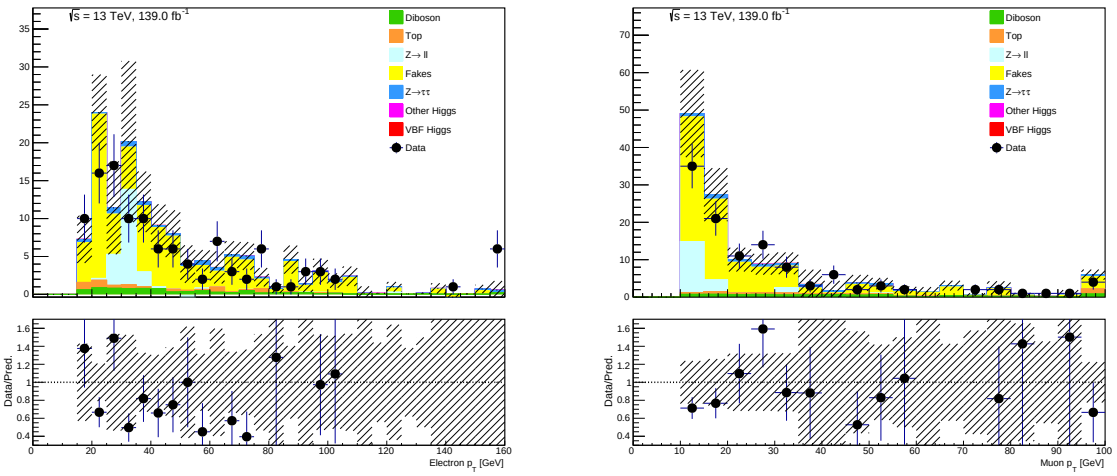


Figure 26: The electron (red) and muon (orange) real efficiencies as a function of the lepton transverse momentum. Uncertainties are statistical.

As with the real efficiencies, one sees an increase in the fake efficiency of the electron as a function of the transverse momentum. This trend is not as pronounced in the case of the muon, where the real efficiencies in the first and second bins are not significantly different.

6.2 Validation of the Fake Estimation

After estimating the final contribution of fake events, the agreement between the predicted distribution of variables with those of data is evaluated. In order to validate the fake estimation, a validation region is defined orthogonal to that of the signal region, and the agreement between data and prediction is investigated. For this, the opposite charge requirement on the two leptons is reversed so that both leptons have the same charge. This region is expected to be dominated by fake events. Example distributions of the electron and muon p_{T}'s in this region are shown in Figure 27.



(a) Electron p_{T} (b) Muon p_{T}

Figure 27: Distributions of (a) the electron and (b) muon p_{T}'s in the same-sign fake validation region. Bottom is a *ratio-plot* showing the data-to-prediction ratio. Error bars represent statistical uncertainties.

It can be generally seen that there is good agreement between the prediction and data, thereby supporting the validity of the fake estimation.

Plots of further kinematic variables in this same-sign validation region can be seen in the appendix A.1.

7 Event Selection

Event selection requirements, also known as *cuts*, are applied to data and simulated samples in order to suppress the contribution of background processes, while maintaining signal events. These requirements are chosen according to [17].

7.1 Trigger Selection

Different triggers are initially used to select which events to discard or further analyze. Each trigger requires a p_{T} threshold that the leptons must fulfill. Different triggers are used in different data-taking periods. In the $\tau_e\tau_\mu$ channel a combination of single-lepton and dilepton triggers are used, and an event is required to pass at least one of these triggers. A summary of the triggers used for this analysis and their p_{T} thresholds are listed in Table 5.

Trigger	Data-Taking Period	Name	p_T threshold [GeV]
Single Electron	2015	*HLT_e24_lhmedium_L1EM20VH*	$p_T^e > 24$
		HLT_e60_lhmedium	$p_T^e > 60$
		HLT_e120_lhloose	$p_T^e > 120$
	2016-2018	*HLT_e26_lhtight_nod0_ivarloose*	$p_T^e > 26$
		HLT_e60_lhmedium_nod0e	$p_T^e > 60$
		HLT_e140_lhloose_nod0	$p_T^e > 140$
Single Muon	2015	*HLT_mu20_iloose_L1MU15*	$p_T^\mu > 20$
		HLT_mu50	$p_T^\mu > 50$
	2016-2018	*HLT_mu26_ivarmedium*	$p_T^\mu > 26$
		HLT_mu50	$p_T^\mu > 50$
Dilepton	2015	*HLT_e17_loose_mu14*	$p_T^e > 17$ $p_T^\mu > 14$
	2016-2018	*HLT_e17_lhloose_nod0_mu14*	$p_T^e > 17$ $p_T^\mu > 14$

Table 5: List of triggers used in this analysis and their respective p_T thresholds. Explanations for triggers can be seen in [23, 27].

7.2 Preselection and VBF Topology Cuts

Preselection and VBF topology cuts are applied onto events passing the trigger cuts to increase the signal-to-background ratio. The former aims to suppress the contribution of different background processes, while whereas the latter attempts to capture the topology of VBF-produced Higgs boson decays. A summary of the preselection and VBF topology cuts are listed in Table 6.

Preselection	1 Medium ID e, 1 Medium ID μ $\Delta\eta_{ll} < 1.5$, $\Delta R_{ll} < 2.5$ $p_T^{j_0} > 40$ GeV, $p_T^{j_1} > 30$ GeV $E_T^{\text{miss}} > 20$ GeV $m_{ll}^{\text{coll}} > M_Z - 25$ GeV ≈ 66 GeV $p_T^{\mu} > 10$ to 27.3 GeV, $p_T^{e} > 15$ to 27 GeV $30 < m_{ll}^{vis} < 100$ GeV $N_{\text{b-jets=0}}$ $0.1 < x_1 < 1.0$, $0.1 < x_2 < 1.0$
VBF Topology	$N_{\text{Jets}} \geq 2$, $M_{jj} > 300$ GeV, $\lvert\Delta\eta_{jj}\rvert \geq 3$, $\eta_{j_0} \times \eta_{j_1} < 0$, Lepton Centrality

Table 6: A summary of the event selection requirements applied to MC and data samples. ll and jj refers to the system of the leading and subleading p_{T} leptons and jets, respectively.

7.2.1 Preselection

One electron and one muon with opposite charge are required since the Higgs boson is neutral. Electrons must fulfill the Medium ID requirement and FCLoose isolation, while muons must fulfill Medium ID requirements and FCTightTrackOnly isolation. Furthermore requirements on the angle difference of the two leptons are applied, such that $\Delta\eta_{ll} < 1.5$ and $\Delta R_{ll} < 2.5$, in order to select leptons originating from decays of high-p_{T} particles, which especially suppresses contributions from Z bosons that are produced with small transverse momenta at the LHC. The leptons must also pass a p_{T} threshold depending on which trigger is fired, this can range from 10 to 27 GeV for the muon and 15 to 27 GeV for the electron. Furthermore, p_{T} requirements are applied onto jets to select jets with properties consistent with those expected in VBF, namely that the leading p_{T} jet must have a transverse momenta of at least 40 GeV, and that of the subleading p_{T} jet 30 GeV. A requirement of at least 20 GeV is applied on the missing transverse energy to select events with neutrinos, thereby significantly limiting contribution from $Z \to ll$ decays. A requirement on the collinear mass is applied to ensure orthogonality with other analyses investigating $H \to WW$ decays, such that $m_{ll}^{\text{coll}} > 66.1876$ GeV. The invariant mass of the two leptons are limited to between 30 and 100 GeV to further suppress $Z \to ll$ events. Events are also required to have no b-tagged jets in order to suppress the contribution of top background. Lastly, the momentum fractions of the decay leptons are required to be between 0.1 and 1 in order to make sure the direction of the missing transverse energy is consistent with those expected in ditau decays.

7.2.2 VBF Topology Cuts

VBF topology cuts define the VBF region. Every event is required to have at least 2 jets, since VBF involves two partons which scatter off and hadronize into two *VBF jets*.

Furthermore, a threshold on the invariant mass of the two leading p_{T} jets is set to 300 GeV. The two jets are required to be in opposite hemispheres with a η separation of at least 3, due to the high energy of the scattered partons. Furthermore lepton centrality is required, meaning that the visible decay products of the τ leptons need to be between the VBF jets in η.

A the event yields of each process after all cuts are applied is shown in Table 7.

Number of Events				
Process	No Cuts		VBF Cuts	
VBF $H \to \tau\tau$	160.77	$\pm$ 0.33	40.82	$\pm$ 0.16
$H \to WW$	88.43	$\pm$ 0.28	16.92	$\pm$ 0.12
Other Higgs	1092.4	$\pm$ 2.1	21.64	$\pm$ 0.30
Top Background	99 712	$\pm$ 73	182.5	$\pm$ 3.4
Diboson	6288	$\pm$ 18	66.3	$\pm$ 1.4
$Z \to \tau\tau$	66 524	$\pm$ 160	894.8	$\pm$ 9.3
$Z \to ll$	5000	$\pm$ 150	12.0	$\pm$ 6.7
Fakes	31 200	$\pm$ 300	213	$\pm$ 21
Total Background	210 140	$\pm$ 270	1448	$\pm$ 12
Data	449 240		1311	

Table 7: Expected number of different signal and background events before and after application of VBF region selection requirements.

The expected number of signal events in the VBF region is 40.82 ± 0.16 events for VBF $H \to \tau\tau$ process and 16.92 ± 0.12 for the $H \to WW$ process. In the VBF region, the most dominant background process is the $Z \to \tau\tau$ process which accounts for more than half of all background events. After the application of preselection and VBF topology cuts, the signal-to-background ratio increases from 0.0012 to 0.04, and the significance from 0.54 to 1.49. There is appoximately a 10% excess in prediction compared to data..

8 Artificial Neural Networks

Several methods can be used to select events that are relevant for the analysis. One approach is the utilization of event selection requirements on variables, which can discriminate signal processes from background events. In this thesis, such a cut-based approach is initially used to increase the signal-to-background ratio. However, the small signal yield and persistent background contributions make it difficult to increase the signal sensitivity using a traditional cut-based method. Another limitation of such a method is that each kinematic variable is considered independently, and correlations between different kinematic variables, which may be optimized better when considered together, can not be combined effectively.

To address such problems, machine learning (ML) methods can be utilized to further discriminate signal from background events. This chapter introduces the basic principles of machine learning and specifically, neural networks.

8.1 Machine Learning

The field of machine learning aims to build and improve models based on data. Two major approaches in machine learning are *supervised* and *unsupervised* learning. Supervised learning utilizes labeled data, which associate inputs with their desired outputs, to build a model that maps inputs to outputs as accurately as possible. Unsupervised learning works without labels to find an overall structure or pattern in the presented data. Two major tasks of machine learning algorithms are *classification* and *regression*. Classification tasks aim to accurately assign class labels to data points, while regression predicts the numerical value of a continuous variable.

In this analysis, supervised learning is used for the classification of events into signal and background processes.

8.2 Artificial Neural Networks

An artificial neural network (NN) is composed of an interconnected collection of units called *nodes*.

A node can receive one or multiple inputs and applies a non-linear transformation on them to produce a single output value. This non-linear transformation is is achieved by the *activation function*. A *bias* can be added to the argument of the activation function as a further degree of freedom. The output of a node can therefore be computed as:

$$x_i^j = f\left(\sum_k w_k x_k^{j-1} + b_i^j\right), \tag{66}$$

where k runs over the number of nodes in the previous layer, x_i^j the output of the i-th node in the j-th layer, w_i the weight applied to the output of the i-th node, b_i^j the bias

applied to the output of the i-th node in the j-th layer and f the activation function. In this thesis, two activation functions are considered. The *rectified linear unit*, (ReLU)

$$f_{\text{ReLU}}(x) = \max(0, x) \tag{67}$$

is used for the hidden layers and softmax, which is used for the output layer, defined as:

$$f_{\text{Softmax}}(x_i) = \frac{e^{x_i}}{\sum_k e^{x_k}}, \tag{68}$$

where k runs over the number of nodes in the last layer and x_i the input to the i-th node in the output layer. Softmax normalizes the output of the layers into a probability distribution, such that the outputs sum up to one. Therefore the output of each output node can be interpreted as the probability that a certain event corresponds to a process represented by that output node.

In this thesis, a feed-forward neural network is used, where the connections do not form a cycle. The nodes are organized into layers, where nodes from each layer are connected with nodes of other layers. The first layer is referred to as the *input layer*, and the last layer the *output layer*. Each node in the input layer takes the value of one input variable as its input, while the output of the output layer provides the prediction of the network. The layers between the input and output layers are called the *hidden layers*. An illustration of a feed-forward neural network is shown in Figure 28.

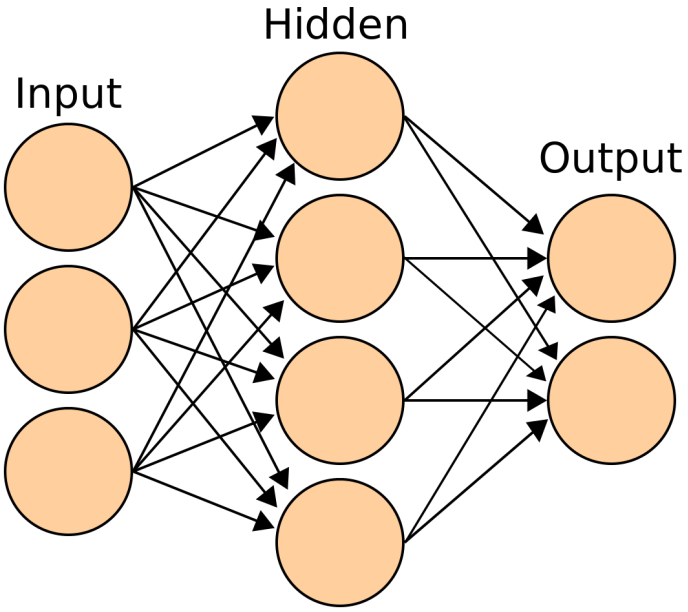


Figure 28: An illustration of an example feed-forward neural network with one hidden layer [33].

8.3 Training

The goal of the training process is to find a combination of weights and biases that minimizes the error between the predicted and true labels. The error as a function of the

weights and biases is called the *loss function* (L). In this thesis, the *categorical cross-entropy* loss function is used, which is typical for classification problems. The categorical cross-entropy [47] is defined as:

$$L = -\sum_i y_i \cdot \log \hat{y}_i, \tag{69}$$

where $\hat{y}_i$ is the prediction and y_i the true value for output node i, which represent a class. A commonly used algorithm to find such a minimum is Stochastic Gradient Descent (SGD) [66]. In SGD, a random initial combination of weights and biases, represented by the vector $\hat{w}$, is selected. Then the gradient of the loss function with respect to the weights and biases ($\nabla_{\hat{w}} L(\hat{w})$) is calculated using a randomly selected set of data. A vector proportional to the calculated gradient is then added to the initial weight vector, effectively shifting $\hat{w}$ closer to a minimum of the loss function. The size of this increment is termed *learning rate*, η, and the amount of data used to compute the gradient for one iteration is called the *batch size*. The weight vector for each iteration i, is therefore given by:

$$\hat{w}_i = \hat{w}_{i-1} - \eta \nabla_{\hat{w}} L \tag{70}$$

Computing the gradient of the loss function with respect to the weights and biases can be computationally intensive, since it must use the chain rule to calculate $\frac{L}{\partial w_{jk}^l}$, where w_{jk}^l represents the weights between layer $l-1$ and l, connecting the k-th node in layer $l-1$ and j-th node in layer l, which results in duplicate calculations. Therefore, the *backpropagation* algorithm [82] is used to efficiently calculate this gradient. For this, backpropagation calculates the gradient for individual layers with regards to the input of each layer beginning from the last layer. The gradient computed from this layer is then used to compute the gradient for the previous layer. One can repeat this recursively until the first layer is reached. This makes the computation of the gradient much more efficient than in the naive case where one uses the chain rule from the first to the last layer, since when calculating the gradient at layer l one does not need to compute gradients on all layers thereafter.

In this thesis, a modified version of SGD is used, namely the Adaptive Moment Esimation (Adam) optimizer [67]. It utilizes first and second moment vectors m and v, defined as the running averages of the gradient and square gradient of the loss function, respectively. With Adam, each iteration is defined by the following set of equations:

$$\begin{aligned}
m_w^{(t+1)} &\leftarrow \beta_1 m_w^{(t)} + (1-\beta_1)\nabla_w L^{(t)} \\
v_w^{(t+1)} &\leftarrow \beta_2 v_w^{(t)} + (1-\beta_2)(\nabla_w L^{(t)})^2 \\
\hat{m}_w^{(t)} &= \frac{m_w^{(t+1)}}{1-\beta_1^{t+1}} \\
\hat{v}_w^{(t)} &= \frac{v_w^{(t+1)}}{1-\beta_2^{t+1}} \\
w^{(t+1)} &\leftarrow w^{(t)} - \eta \frac{\hat{m}_w^{(t)}}{\sqrt{\hat{v}_w^{(t)}} + \epsilon},
\end{aligned}$$

where $m_w^{(t)}$ and $v_w^{(t)}$ are the first and second moment vectors at iteration t for the weight w respectively, and $\hat{m}_w$ and $\hat{v}_w$ the bias-corrected first and second moment estimates respectively. L is the loss function. The initial moments, m_w^0 and v_w^0 are set to 0. Parameters β_1, β_2 and ϵ are set to 0.99, 0.999 and 10^{-8} respectively [67]. Adam is preferred for its computational efficiency and low memory requirements, which makes it suitable for problems involving large amounts of data [67].

8.4 Validation and Overtraining

Parameters which are set constant during the training process, such as the learning rate and number of layers, are called hyperparameters. Furthermore, epoch refers to a complete iteration over the training dataset.

After each epoch, the trained model is evaluated using a dataset orthogonal to the training dataset, called the *validation dataset*. The trained model may be able to give accurate predictions for the training dataset, however, it may not be able to generalize this performance to a previously unseen dataset. This could manifest itself in a significantly higher value of the loss function when computed using the validation dataset, compared to that of using the training dataset. This phenomenon is referred to as *overtraining*. Several methods exist to mitigate overtraining, such as regularization, dropout layers, learning rate decay and cross-validation.

8.4.1 L2 Regularization

In machine learning, regularization refers to the general process of the simplification of a model. One common method is the use of L2 regularization, also known as ridge regression [85]. L2 regularization penalizes networks with large weights by adding to the loss function a term quadratic in the weights.

$$L_{L2} = L(w_1, \ldots, w_n, b_1, \ldots, b_n) + \lambda \sum_{i=1}^{n} w_i^2, \tag{71}$$

where L is the original loss function and λ a hyperparameter smaller than one. L2 regularization prevents overtraining by penalizing overly complicated and flexible networks, which are modelled too closely to the training dataset. Such models are characterized by large weights, and cannot accurately generalize to data it has not seen. In this thesis, L2 regularization parameters, λ, between 10^{-6} and 10^{-3} are considered.

8.4.2 Dropout

The Dropout method can also be utilized to avoid overtraining. Nodes in the dropout layers randomly set their inputs to zero at a given rate, which prevents the network from being too dependent on a few input variables, and allows the network to learn from a larger set of the input variables, thereby improving generalizability.

In this thesis, dropout rates between 0.2 and 0.8 are considered.

8.4.3 Learning Rate Decay

Learning rate decay refers to the reduction of the learning rate after every iteration. This limits how much the weight vector can be incremented every iteration and allows for a faster convergence onto the minimum of the loss function. At the same time it decreases the possibility of excessively large weights. The learning rate at iteration i is modulated according to the following equation:

$$\eta_{i+1} = \eta_i \frac{1}{1 + \lambda_d \cdot i}, \tag{72}$$

where the decay rate, λ_d is set to a constant of 0.001.

8.4.4 k-fold Cross-Validation

To increase the generalizability of the trained model, k-fold cross-validation can be used. The entire dataset is randomly split into k equal parts. For each fold, $k-2$ parts are used for training, and one part each to validation and testing (see next section). After each fold, the training, validation and testing samples are permuted such that; after k folds, the network has done training validation and testing on all k parts at least once. In this thesis, 5-fold cross-validation is used. An illustration of 5-fold cross validation is shown in figure 29.

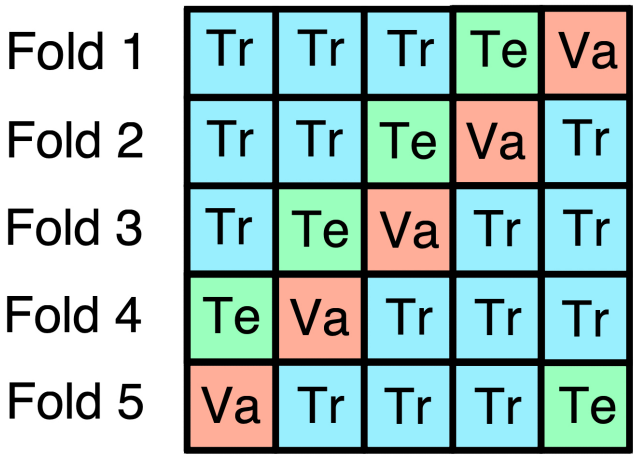


Figure 29: An illustration of how parts of the dataset is assigned to training (Tr), testing (Te) or validation (Va) tasks a in 5-fold mixed cross validation setup.

8.5 Hyperparameter Optimization

Hyperparameters remain constant during the duration of the training. They can be optimized to improve the performance of the neural network, which is evaluated using a dataset orthogonal to both the training and validation datasets, called the *testing* dataset. In this thesis, hyperparameter optimization is done using the the *Optuna* Framework [2], which utilizes the tree-structured parzen estimator (TPE), a Bayesian hyperparameter optimization algorithm.

8.5.1 Tree-Structured Parzen Estimator

The Tree-Structured Parzen Estimator (TPE) is a Bayesian optimization algorithm [85]. In contrast to grid searches, where hyperparameters are scanned in a certain range at regular step sizes, and random searches, where random combinations of hyperparameters are evaluated, Bayesian optimization utilizes past evaluations to select which combination of hyperparameters are selected. For this, Bayesian optimization constructs a probabilistic model $P(y|x)$ relating the hyperparameters, x, to a score of an *objective function*, y, which one wishes to minimize or maximize. This probabilistic model is referred to as the *surrogate function* for the objective function. The hyperparameters, which perform best according to the surrogate function are selected and evaluated using the objective function. The result of the evaluation is then used to update the surrogate model. These steps are repeated until a predefined maximum amount of iterations or time is reached. The set of hyperparameters at each iteration is chosen according to the *selection function*, which is commonly chosen to be the *expected improvement* [85]:

$$\mathrm{EI}_{y^*}(x) = \int_{-\infty}^{y^*} (y - y^*) P(y|x) dy, \tag{73}$$

where y^* is a threshold value of the objective function. The hyperparameters maximizing the expected improvement with the surrogate function $P(y|x)$ are selected.

In Tree-Structured Parzen estimators, the surrogate function $P(y|x)$ is built based on Bayes' Rule [85]:

$$P(y|x) = \frac{P(x|y) \cdot P(y)}{P(x)}. \tag{74}$$

where $P(x)$ and $P(y)$ are the probability distribution functions of hyperparameters and the objective function, respectively and $P(x|y)$ represents the probability of having a certain hyperparameter combination x given a score y and is expressed in TPE as:

$$P(x|y) = \begin{cases} l(x) & \text{if } y < y^* \\ g(x) & \text{if } y \geq y^* \end{cases} \tag{75}$$

where $l(x)$ is the PDF formed by the observations resulting in scores less than the threshold value y^* and $g(x)$ the density formed by the remaining observations. The threshold

value y^* is chosen to be a quantile of observed y values Therefore, two different distributions of the hyperparameters are constructed depending on whether the score passes this threshold or not. By doing so, the algorithm can focus on and optimize in a specific region depending on whether the score should be minimized or maximized.

8.6 Evaluation of the Neural Network Performance

In order to evaluate the ability of the neural network in identifying signal events from background events, one uses the output of the signal node, also known as the neural network score. The output of the signal node corresponds to the probability that an event is of a signal process. A threshold can be applied, such that only events above a certain score are kept for the analysis. This is determined by calculating the expected significance, defined by:

$$\text{Sign} = \frac{s}{\sqrt{s+b}}, \tag{76}$$

where s and b correspond to the number of signal and background events which pass a given score, respectively, for different score thresholds and finding the threshold, which results in the highest significance. The maximum significance is then used as a measure of the performance of the neural network, which is maximized by Optuna for its hyperparameter optimization.

Another method of evaluating the performance of the neural network is the evaluation of the loss and the accuracy. The loss refers to the average of the loss function after every epoch. In this thesis, the categorical accuracy (Acc) is used as a metric for the accuracy, which is defined to be:

$$\text{Acc} = \frac{N_{correct}}{N_{tot}} \tag{77}$$

where $N_{correct}$ and N_{tot} are the number of correct predictions and total number of predictions, respectively.

In contrast with the significance, which is evaluated after the entire training processes. The loss and the accuracy can be evaluated during the training process after each epoch, and therefore is used to determine the presence of overtraining.

9 Final Event Selection Using Neural Networks

This section describes the process of hyperparameter optimization, with the goal of maximizing the selection performance of signal processes. The neural networks were trained on Monte Carlo simulated samples, considering the following processes: VBF Higgs production, top background, $Z \to \tau\tau$, $Z \to ll$, diboson production and other Higgs production modes, as well as the data-driven estimation of contributions from fake leptons. This step is applied after an initial selection though event selection requirements since the training of neural networks is a computationally-intensive process, and therefore one wishes to optimize the amount of data that is used for the training the neural network. The training is done in the VBF region, which is defined to contain events which pass both preselection and VBF topology cuts.

Before the training of the neural networks, the data is pre-processed such that each input variable is scaled down so that their averages are zero and their standard deviation is one. This prevents certain variables from biasing the neural network and thereby making sure the network is able to extract more information from all variables. Furthermore, an input feature with a large variance may result in large gradient values of the loss function, causing the shift in weights to be large, thus making the training process unstable.

Events with negative weights are excluded from the training due to technical limitations [17]. However, for validation and testing steps, they are not excluded.

Events corresponding to each output class is weighted so that the sum of weights for all events of each output class is the same. This allows the neural network to treat all output classes equally without biasing the network to focus more on one class over the other.

During the optimization of hyperparameters, following hyperparameters and ranges are considered:

- Batch Size $\in \{128, 512, 1024, 2048, 4096, 8192\}$
- Number of layers $\in \{2, 3, 4, 5, 6, 7, 8, 9, 10\}$
- Number of nodes per layer $\in \{128, 256, 512\}$
- L2 Regularization parameter $\in [10^{-6}, 10^{-3}]$
- Learning rate $\in [10^{-6}, 10^{-2}]$

A total of a 100 trials are used.

9.1 Input Variable Optimization

The goal of input feature optimization is to keep the neural network simple and therefore computationally efficient, while at the same time maintaining an optimal performance. Therefore variables which do not significantly improve the performance of the neural

network are removed. This is done through the calculation of *permutation importance*. For this, a neural network is initially trained with a given set of variables and the resulting significance is calculated. Then, for every input feature, values are randomly shuffled for each event, and the significance is calculated again, thereby decoupling the correlation between this variable and other variables. The permutation importance (I) of an input feature is defined as the decrease in the significance when that feature is shuffled randomly, and can be computed as:

$$I = \frac{\mathrm{Sig}^{\mathrm{orig}}}{\mathrm{Sig}^{\mathrm{perm}}}, \tag{78}$$

where $\mathrm{Sig}^{\mathrm{orig}}$ is the significance computed in the original unpermuted case, and $\mathrm{Sig}^{\mathrm{perm}}$ for the case where the variable is permuted.

The initial input variable set is composed of the following variables and is mainly based on the findings of previous analyses [17, 28]:

- $p_{\mathrm{T}}^{j_{0,1}}$, the transverse momentum of the leading (j_0)/subleading (j_1) p_{T} jet.
- $\eta^{j_{0,1}}$, the pseudorapidity of the leading/subleading p_{T} jet.
- $\phi^{j_{0,1}}$, the azimuthal angle of the leading/subleading p_{T} jet.
- m_{ll}^{MMC}, invariant dilepton mass reconstructed using the missing mass calculator.
- m_{ll}^{vis}, visible invariant mass of the dilepton system
- m_{ll,j_0}, invariant mass of the dilepton and leading p_{T} jet system.
- $m_T(l_{0,1}, E_{\mathrm{T}}^{miss})$, transverse mass of the leading (l_0)/sub-leading (l_1) lepton candidate and missing transverse energy, defined as:

$$\sqrt{2p_{\mathrm{T}}^{l_{0,1}} E_{\mathrm{T}}^{miss}(1 - \cos\Delta\phi(l_{0,1}, E_{\mathrm{T}}^{miss}))}. \tag{79}$$

- ΔR_{ll}, angular distance between the two leptons
- $\Delta\eta_{ll}$, pseudorapidity difference of the two leptons
- $\Delta\phi_{ll}$, azimuthal angle difference between the two leptons
- $p_{\mathrm{T}}(l_0) + p_{\mathrm{T}}(l_1)$, scalar sum of the leptons' transverse momenta
- $p_{\mathrm{T}}(ll)$, transverse momenta of the dilepton system
- $p_{\mathrm{T}}(l_0, l_1, E_{\mathrm{T}}^{miss})$, transverse momenta of the Higgs boson candidate.
- $p_{\mathrm{T}}(l_0)/p_{\mathrm{T}}(l_1)$, ratio of the transverse momenta of the leading p_{T} lepton to the subleading p_{T} lepton.
- $(p_{\mathrm{T}}(l_0) - p_{\mathrm{T}}(l_1))/(p_{\mathrm{T}}(l_0) + p_{\mathrm{T}}(l_1))$ ratio of the transverse momentum difference between the two leptons to the scalar sum of their transverse momenta.

- $C(\phi(E_{\mathrm{T}}^{miss}))$, azimuthal centrality of E_{T}^{miss}, defined as:

$$C(\phi(E_{\mathrm{T}}^{miss})) = \sqrt{2}\frac{r+s}{\sqrt{r^2+s^2}} \tag{80}$$

where

$$r = \frac{\sin(\phi_{E_{\mathrm{T}}^{miss}} - \phi_{l_0})}{\sin(\phi_{l_1} - \phi_{l_0})}, s = \frac{\sin(\phi_{l_1} - \phi_{E_{\mathrm{T}}^{miss}})}{\sin(\phi_{l_1} - \phi_{l_0})}. \tag{81}$$

- $E_{\mathrm{T}}^{miss}/p_{\mathrm{T}}(l_{0,1})$, ratio of the missing transverse energy and the transverse momentum to the leading/sub-leading lepton.
- $\Delta\eta_{jj}$, the pseudorapidity difference between the two leading p_{T} jets
- m_{jj}, invariant mass of the two leading jets
- $\eta_{j_0} \times \eta_{j_1}$, signed product of the pseudorapidity of the leading and sub-leading jets.
- $p_{\mathrm{T}}(l_0, l_1, j_0, j_1, E_{\mathrm{T}}^{miss})$, the transverse momentum of the two leading p_{T} leptons, two leading p_{T} jets and missing transverse energy system.
- $(\sum p_{\mathrm{T}})_{scalar}$, the scalar sum of the p_{T} of all objects.
- p_{T}^{total} the p_{T} of the system of all objects
- $C_{jj}(l_{0,1})$, the centrality of the leading/sub-leading p_{T} lepton candidate with respect to the two leading p_{T} jets, defined as:

$$C_{jj}(l) = \exp\left[\frac{-4}{(\eta_{j_0} - \eta_{j_1})^2}\left(\eta_{l_{0,1}} - \frac{\eta_{j_0} + \eta_{j_1}}{2}\right)^2\right] \tag{82}$$

Each variable is permuted randomly 10 times and the resulting significances are averaged. This is then compared with the significance calculated with the set of all initial variables unpermuted.

To measure the permutation importance of these variables, a network was trained with baseline setup defined with the following settings:

- 3 hidden layers
- 50 nodes per layer
- L2 parameter of 10^{-6}
- Learning rate of 0.001
- Decay rate: 0.001
- 50 Epochs
- Batch size of 512

With all variables unpermuted, the network resulted in an expected significance of 3.31, with an optimum neural network score threshold of 0.82. The importance of all of the investigated kinematic variables is shown in Figure 30.

Variables whose performance was not significantly better than that of the baseline setup were removed from the list. These are: $E_{\mathrm{T}}^{miss}/p_{\mathrm{T}}(l_0)$, $E_{\mathrm{T}}^{miss}/p_{\mathrm{T}}(l_1)$, $C(\phi(E_{\mathrm{T}}^{miss}))$, ϕ_{j_0}, ϕ_{j_1}, η_{j_0}, η_{j_1}, $\Delta\eta_{ll}$, $\Delta\phi_{ll}$.

As expected, variables such as the azimuthal angle of jets do not significantly improve the performance of the neural network since their distributions are uniform.

These 9 variables were removed and the neural network was trained again with the same hyperparameters to observe the influence of the lack of the removed variables. A 3% decrease in the expected significance, from 3.31 to 3.22 was observed.

Using the selected variable set, a hyperparameter optimization was performed. A maximum significance of 3.39 was reached at an optimum neural network score threshold of 0.88, with a learning rate of 0.002, batch size of 1024, 10 layers with 128 nodes each and a L2 parameter of 1.18×10^{-6}. A plot of the loss and accuracy curve of this optimized network is shown in Figures 31a and 31b. A distribution of the neural network score is shown in figure 32. From the loss curve one can observe the presence of overtraining.

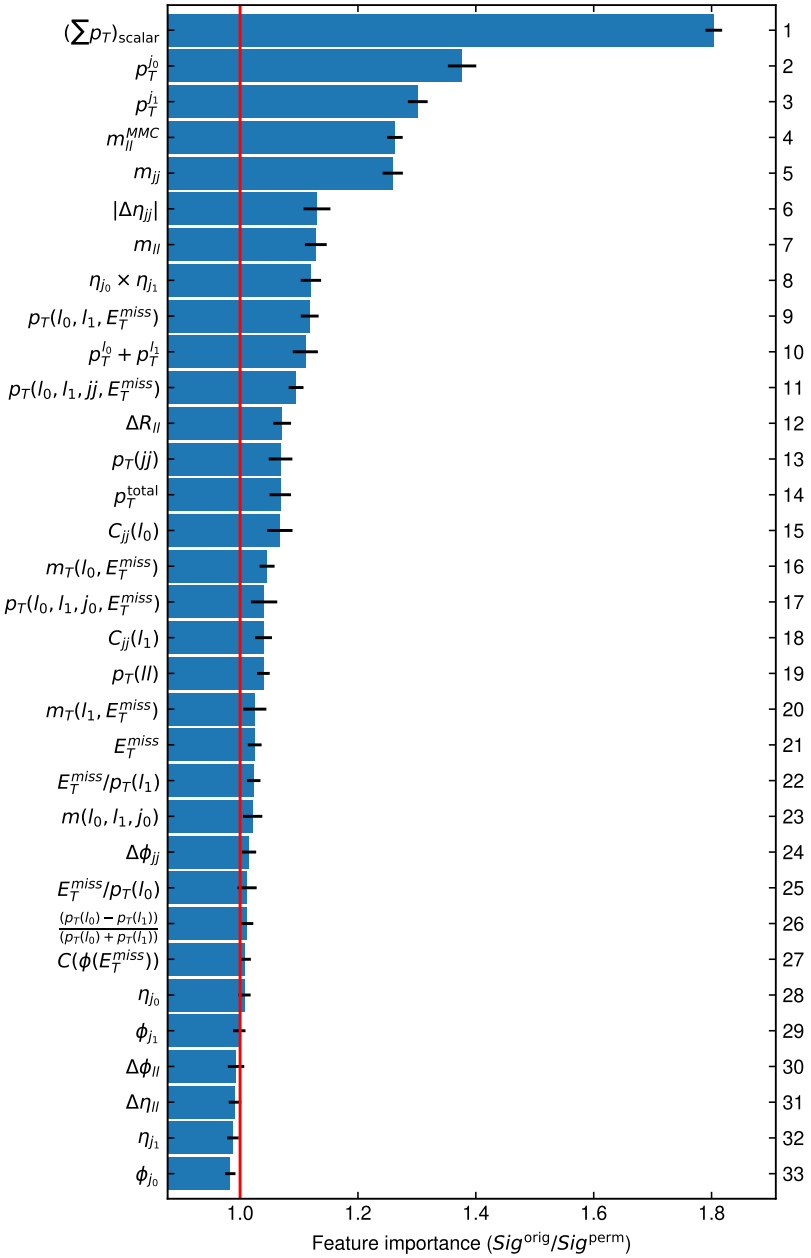
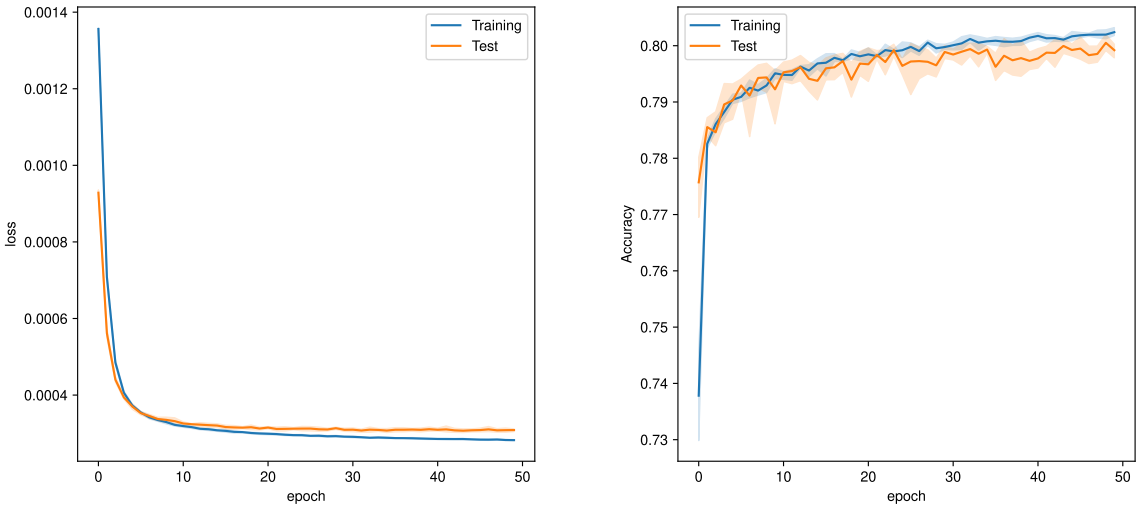


Figure 30: A ranking plot of the measured permutation importance of all variables determined using the baseline neural network setup, as described in section 9.1. Uncertainties are statistical uncertainties on the average values. The red line represents the performance of the original network.



(a) Loss as a function of the epoch.

(b) Accuracy as a function of the epoch.

Figure 31: The training and validation loss and accuracy as a function of the epoch for the optimized network utilizing selected input variables.

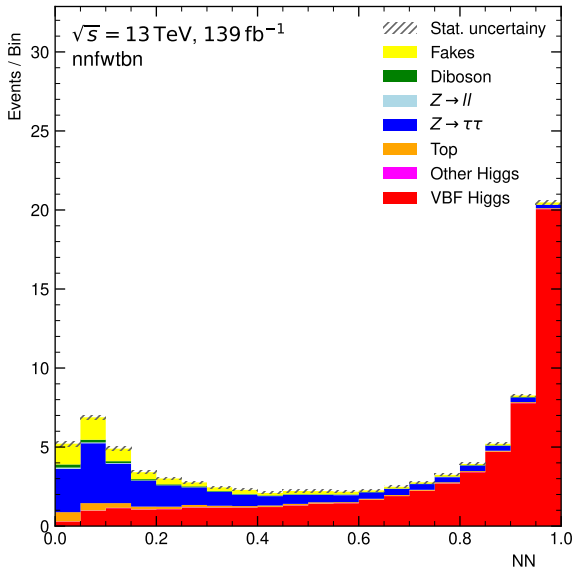


Figure 32: The distribution of the neural network score for each simulated processes stacked.

The performance of the network optimized using selected input variables are compared with the performance of networks utilizing input variable sets used for training Boosted Decision Trees in previous analyses [17, 28]. The input variable set that was used in the ATLAS analysis [17], is termed set 1, and that of the ATLAS analysis [28] is termed set 2. Full hyperparameter optimization was done separately with both input variables sets and the performance of the best performing neural networks are taken. A table comparing the performances of neural networks utilizing different variable sets is shown in table 8.

- Input variable set 1: m_{ll}^{MMC}, m_{ll}^{vis}, m_{jj}, $m_T(l_0, E_{\mathrm{T}}^{miss})$, ΔR_{ll}, $p_{\mathrm{T}}(ll, j_0, j_1, E_{\mathrm{T}}^{miss})$, $p_{\mathrm{T}}(j_2)$, $C_{jj}(l_0)$, $C_{jj}(l_1)$, $E_{\mathrm{T}}^{miss}/p_{\mathrm{T}}(l_0)$, $E_{\mathrm{T}}^{miss}/p_{\mathrm{T}}(l_1)$.
- Input variable set 2: p_{T}^{Hjj}, $\eta_{j_0} \times \eta_{j_1}$, p_{T}^{jj}, $\Delta\eta_{jj}$, $\Delta\phi_{jj}$, m_{jj}, $p_{\mathrm{T}}^{j_1}$

Input Variable Set	Maximum Significance	NN Score Threshold	s	b
Selected	3.39	0.88	29.9	50.9
1	3.22	0.94	19.54	18.96
2	2.51	0.74	27.76	96.46

Table 8: Table outlining the expected significance and the optimal neural network score threshold as well as the number of signal (s) and background (b) events passing this threshold. The selected input variable set refers to the set that was obtained via the permutation importance. Set 1 is taken from [17] and set 2 from [28].

The selected set outperforms the other variable sets, which is to be expected since the selected set utilizes a larger number of input variables, thereby allowing the greater exploitation of correlations between each input feature.

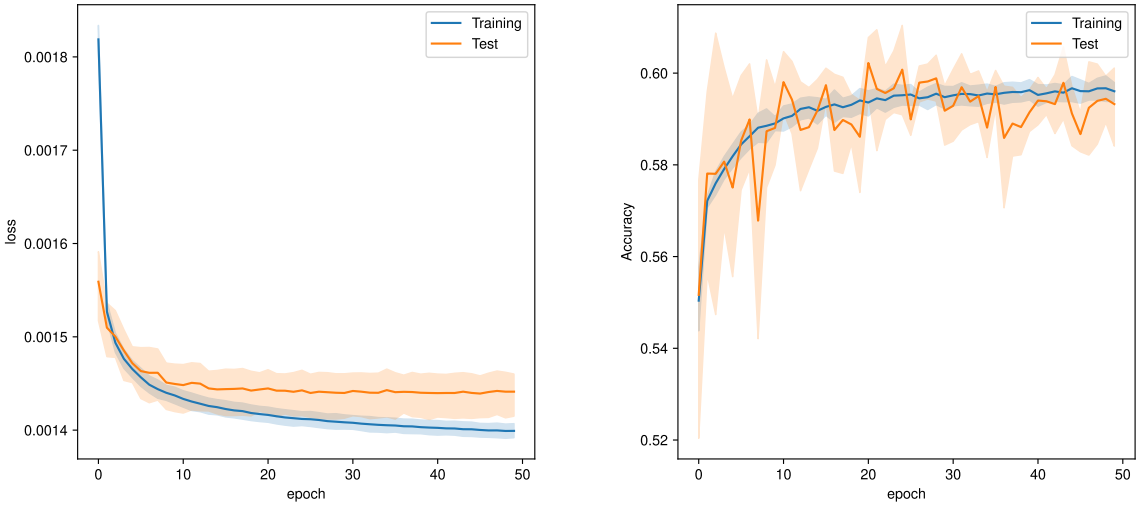
9.2 Output Node Diversification

As an alternative to having two output nodes, which correspond to one output for signal and one output for all background processes combined, one can increase the amount of output nodes to consider some dominant background processes individually. By doing so one can train the network to specifically identify the signatures of major background contributions and thereby also increase the discrimination of signal events from background events.

The two largest background contributions in the $H \to \tau\tau \to e\mu 4\nu$ channel are the $Z \to \tau\tau$ and the top background processes. A network with four output nodes is set up, corresponding to the following processes: signal, $Z \to \tau\tau$, top and other background processes. The definition of other background class is redefined to not contain $Z \to \tau\tau$ and top events. The selected input variables, as described in Section 9.1 is used for the training. Hyperparameter optimization is done again with this setting and compared with the case with two output nodes. The optimized four-output network has the following hyperparameters: Learning rate of 0.0034, batch size of 128, 5 layers with 256 nodes each and an L2 parameter of 2.36×10^{-6}. A table comparing the performance of the optimized

	$s/\sqrt{s+b}$	Threshold	s	b
2 Outputs	3.39	0.88	29.9	50.9
4 Outputs	3.49	0.68	28.1	39.4

Table 9: The significance and optimum neural network score threshold for the optimized 2-output and 4-output networks, as well as the number of signal (s) and background (b) event yields, which pass this threshold.



(a) Loss as a function of the epoch.

(b) Accuracy as a function of the epoch.

Figure 33: The training and validation loss and accuracy as a function of the epoch for the optimized network utilizing four output nodes.

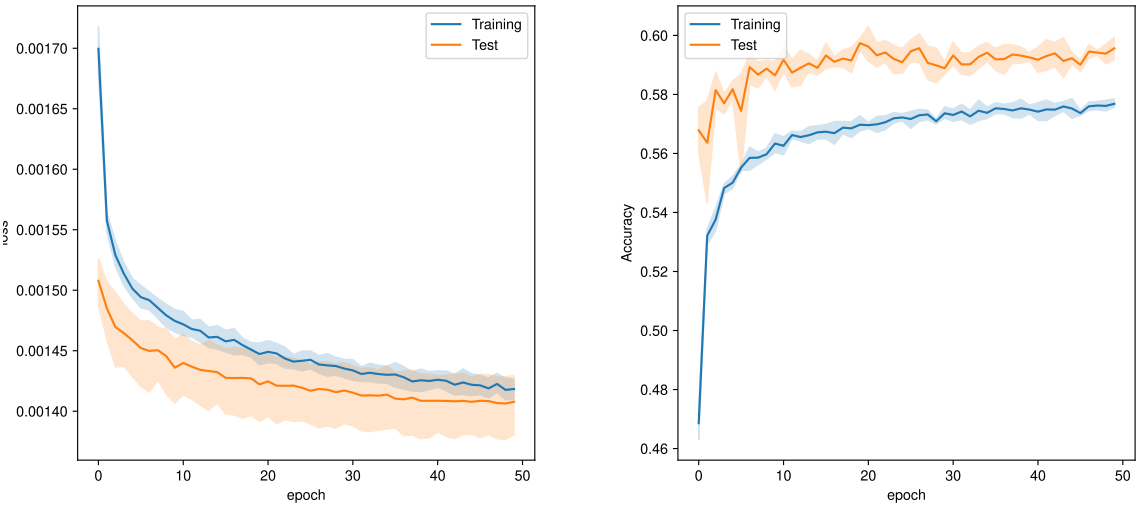
two-output and four-output node networks is shown in Table 9. Figures of the loss and accuracy of the network as a function of the epoch are shown in Figures 33a and 33b.

It can be seen from Figure 33a that overtraining is present, which was also observed in the optimized 2-output network. The two additional output nodes for $Z \to \tau\tau$ and top background processes increased the maximum significance by 0.1. Furthermore, one observes the shift of the optimum neural network score threshold to a lower value, implying an improved identification of background processes.

9.3 Dropout Layers

An alternative method of reducing overtraining apart from L2 regularization is the use of dropout layers. The performance of the utilization of dropout layers is compared with L2 regularization using the optimized 4-output network using selected variables. Dropout layers are inserted between each hidden layer with a variable dropout rate from 0.2 to 0.8. The L2 regularization in each hidden layer is removed. Hyperparameter optimization is done with the dropout rate as an additional variable. The resulting network has a learning rate of 0.003, batch size of 2048, 2 layers with 256 nodes each and a dropout rate of 0.69.

The best significance is found to be 3.38 with an optimum neural network score threshold of 0.66. A plot of the loss and accuracy of the neural network as a function of the epoch is shown in Figures 34a and 34b.



(a) Loss as a function of the epoch.

(b) Accuracy as a function of the epoch.

Figure 34: The training and validation losses and accuracies as a function of the epoch for the optimized network utilizing dropout layers.

In both curves one can observe that the test loss is lower than that of the training loss and vice versa for the accuracy. This is due to the fact that the dropout is only applied during training and not during validation. The optimized network with L2 regularization has a significance of 3.49, which is 0.11 higher than that of the method with dropout layers. Therefore, L2 regularization was chosen over the dropout layers method.

9.4 Final Optimized Neural Network Setup

The neural network, whose configurations resulted in the best significance, was selected to be the the following:

- Input variables: $p_{\mathrm{T}}^{j_{0,1}}$,$\eta^{j_{0,1}}$, $\phi^{j_{0,1}}$, m_{ll}^{MMC}, m_{ll}^{vis}, $m_T(l_0, E_{\mathrm{T}}^{miss})$, m_{l_0,l_1,j_1}, $m_T(l_1, E_{\mathrm{T}}^{miss})$, ΔR_{ll}, $p_{\mathrm{T}}(l_0)$+$p_{\mathrm{T}}(l_1)$, $p_{\mathrm{T}}(ll)$, $p_{\mathrm{T}}(l_0, l_1, E_{\mathrm{T}}^{miss})$, $p_{\mathrm{T}}(l_0, l_1, j_0, E_{\mathrm{T}}^{miss})$, $p_{\mathrm{T}}(l_0)/p_{\mathrm{T}}(l_1)$, $\frac{(p_{\mathrm{T}}(l_0)-p_{\mathrm{T}}(l_1))}{(p_{\mathrm{T}}(l_0)+p_{\mathrm{T}}(l_1))}$, $\Delta\eta_{jj}$, m_{jj}, $\eta_{j_0} \times \eta_{j_1}$, $p_{\mathrm{T}}(l_0, l_1, j_0, j_1, E_{\mathrm{T}}^{miss})$, $C_{jj}(l_0)$, $C_{jj}(l_1)$.
- Learning Rate: 0.0034
- Batch size: 128
- 5 layers.
- 256 Nodes per layer.
- L2 regularization with $\lambda = 2.36 \times 10^{-6}$.

- 4 output nodes: Signal, Top, $Z \to \tau\tau$, and other background processes.

The network resulted in a maximum significance of 3.49, with an optimized neural network score threshold of 0.68. Therefore, events in the final signal region must fulfill the requirement that their neural network score is greater than 0.68.

Figures of the loss and accuracy of this optimized network as a function of the epoch are shown in Figures 33a and 33b. Furthermore, the distribution of the output of the signal node for the validation dataset is shown in Figure 35.

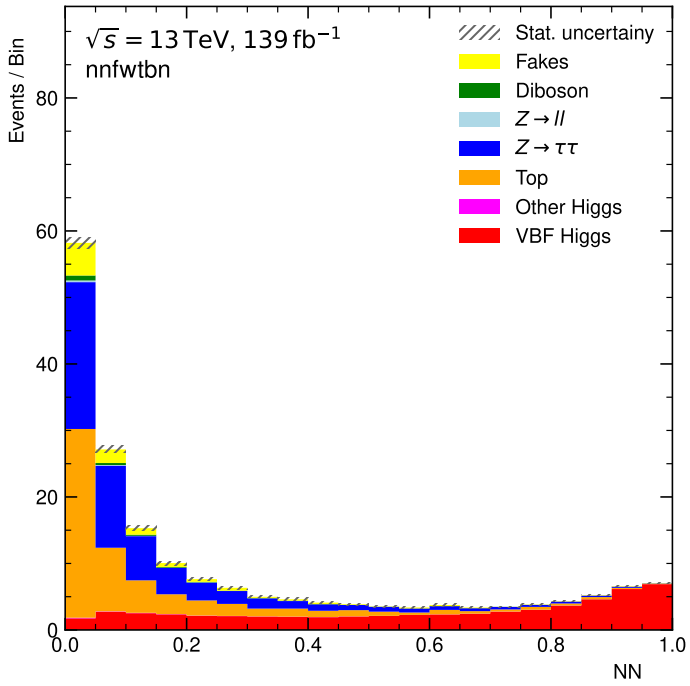
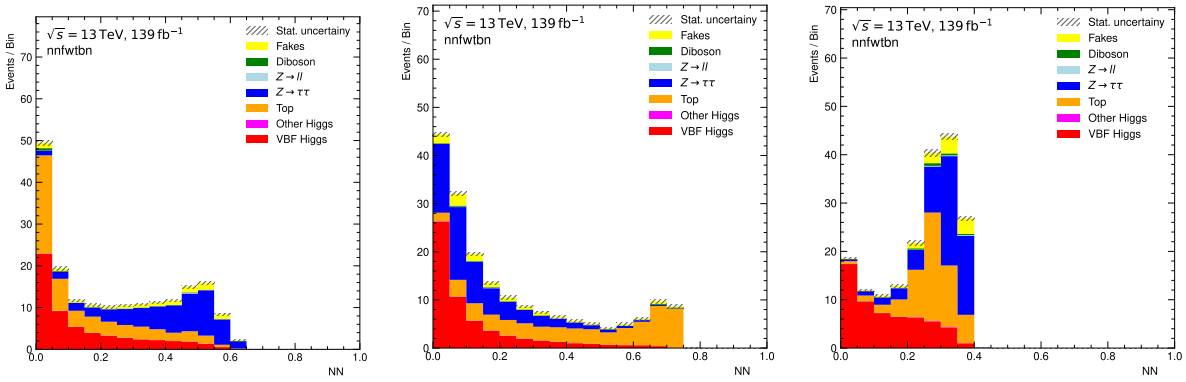


Figure 35: The output of the signal node of the final optimized network, evaluated using the validation dataset. Processes corresponding to each output node are scaled such that they are normalized to each other.

As it can be seen, the signal events are generally at the higher end of the NN-score distribution, while background processes are generally at the lower end of the distribution. Although not used in the calculation of the significance, the distributions of outputs of the the other output nodes can also be compared. These are shown in Figure 36.

It can also be seen from these distributions that for each output node, their corresponding background process generally skew towards the higher end of the neural network score distribution. This shows that the network is able to learn the characteristics of such specific processes. One phenomenon that can be noticed is that the neural network score range is limited to a certain threshold below the maximum value of one. For $Z \to \tau\tau$ this is approximately at 0.7, for top background 0.8 and for other backgrounds around 0.45. This implies that the neural network is more uncertain about the classification of such processes, which may be due to limited statistics, especially for the other background processes since they only take up 7% of the total expected number of events.



(a) $Z \to \tau\tau$ (b) Top Background (c) Other Background Processes

Figure 36: The distribution of the outputs of the output nodes for (a) the $Z \to \tau\tau$, (b) top and (c) other background processes, evaluated using the validation dataset. Histograms of processes corresponding to each output node are scaled such that they are normalized to the signal process.

9.5 Validation of Input Features

The samples used for the training of the neural networks are simulated using Monte Carlo simulations, except for fakes which are estimated through data-driven methods. Since the same neural network will have to be applied on data, it is necessary to make sure that input variables used in the training of the neural network are modeled well.

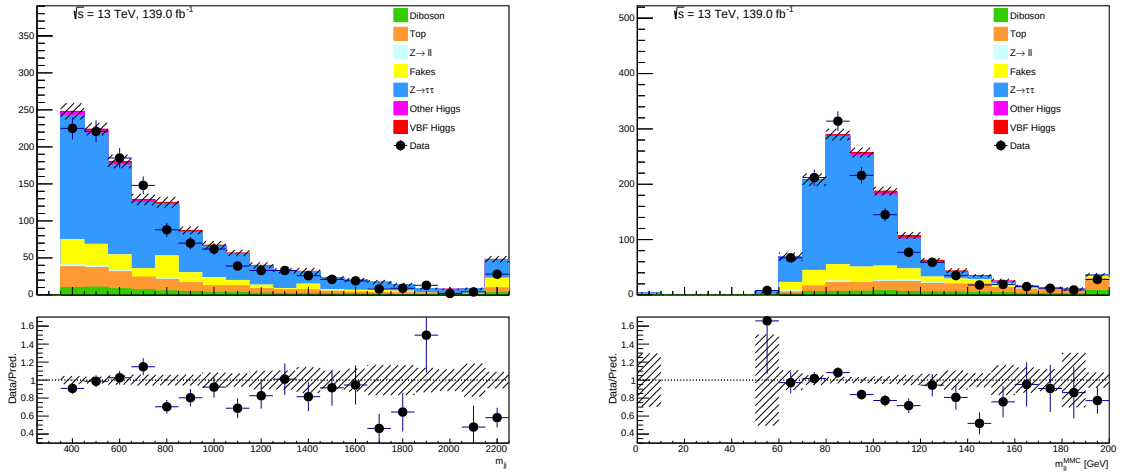
The analysis is done blinded in order to avoid bias, i.e. data in the signal region is not available. Therefore input validation is done in a region orthogonal to the signal region, called the low-NN score region. The neural network score requirement is inverted i.e. $NN < 0.68$ and the expected and observed distributions of input variables used in the training of the neural network are compared. As an example the distributions of m_{jj} and the m_{ll}^{MMC} are shown in Figure 37a and Figure 37b, respectively. The distribution of the other input variables can be seen in the appendix A.2.

It can been in Figure 37a, that for certain kinematic regions, the data consistently underestimates the prediction, namely between approximatly 750 GeV and 950 GeV as well as between approximately 1050 GeV and 1200 GeV. A similar phenomenon can be seen in Figure 37b between 90 GeV and 120 GeV. However, in other distributions, one sees no significant inconsistency between the expected and data distributions.

9.6 Event Yields

The neural network score requirement of > 0.68 is applied in addition the VBF topology cuts. The predicted event yield before and after the NN score cut is shown in Table 10.

After the neural network score cut, the signal-to-background ratio increased from 0.04 to 0.72. The significance is increased from 1.52 to 3.45. Large reductions in dominant



(a) m_{jj} (b) m_{ll}^{MMC}

Figure 37: The predicted and observed distribution of the invariant mass of the dijet system and the reconstructed invariant mass of the Higgs boson using the MMC in the low NN-score region. Furthermore, below, the ratio of the observed yield to the predicted yield for each bin is shown. All uncertainties are statistical and the last bin is the overflow bin.

	VBF H	VBF $H \to WW$	Top	Diboson	$Z \to \tau\tau$	$Z \to ll$	Other Higgs	Fakes	$\sum s$	$\sum b$
VBF Top. Cuts	40.69 ± 0.16	17.03 ± 0.13	182.37 ± 3.49	66.13 ± 1.43	893.06 ± 9.45	12.00 ± 6.69	21.63 ± 0.30	213.88 ± 21.04	57.72 ± 0.20	1389.06 ± 12.18
NN Score Cut	21.65 ± 0.12	6.63 ± 0.08	5.17 ± 0.62	2.40 ± 0.31	23.76 ± 1.37	0.29 ± 0.20	2.65 ± 0.10	4.82 ± 3.80	28.28 ± 0.14	39.09 ± 1.55

Table 10: The expected number of events for each signal and background process after the VBF topology (VBF Top.) and NN score cuts. The last two columns show the sum of the signal ($\sum s$) and background ($\sum b$) yields. Uncertainties are statistical.

background processes are observed. For example the top and $Z \to \tau\tau$ backgrounds were reduced by approximately 97%. Overall, background events saw a reduction of 97.2%, while for signal events, this was 51.0%. This is a demonstration of the effectiveness of the utilization of machine learning methods for the discrimination of signal from background events.

10 Test of CP Invariance

Using CP-odd variables such as the Optimal Observable, one can investigate CP properties of the HVV vertex. The asymmetry in the distributions of CP-odd variables when $\tilde{d} \neq 0$ can be exploited to estimate limits on $\tilde{d}$. Two statistical methods of estimating limits on $\tilde{d}$ are examined and compared. Both methods rely on generating expected BSM distributions with CP-odd contributions and comparing these with that of measured data.

In this thesis, the analysis is performed blinded. Therefore data events in the signal region are not available. These methods are therefore applied onto an *Asimov* dataset, which is generated from simulated samples corresponding to the SM case of $\tilde{d}$, in order to extract expected limits on $\tilde{d}$.

10.1 Signal Prediction for $\tilde{d} \neq 0$

In order to determine limits on $\tilde{d}$, the distributions of CP-odd variables are predicted for different $\tilde{d}$. To achieve this, SM samples for signal processes are reweighted such that their overall distribution corresponds to a given $\tilde{d} \neq 0$ case. The applied weights (w) are calculated as the ratio of the square of the matrix element of the CP-mixed BSM case to that of the CP-even SM case, which can be rewritten in the following way:

$$w(\tilde{d}) = \frac{|\mathcal{M}|^2}{|\mathcal{M}_{\text{SM}}|^2} \tag{83}$$

$$= 1 + \tilde{d}\frac{2\Re(\mathcal{M}^*_{\text{SM}}\mathcal{M}_{\text{CP-odd}})}{|\mathcal{M}_{\text{SM}}|^2} + \tilde{d}^2\frac{|\mathcal{M}_{\text{CP-odd}}|^2}{|\mathcal{M}_{\text{SM}}|^2} \tag{84}$$

$$= 1 + \tilde{d}w_{\text{lin}} + \tilde{d}^2 w_{\text{quad}}, \tag{85}$$

where at equation 84, the expression for the square of the matrix element with CP-violating contributions according to Equation 33 is inserted. The terms w_{lin} and w_{quad} are calculated for every event and are called the linear and quadratic weights. These weights are computed using matrix elements reported by HAWK [48], which takes as input truth-level information.

10.2 Gauge Curve Method

CP-odd variables such as the $\mathcal{OO}$ and the $\Delta\phi_{jj}^{signed}$ show asymmetry in the presence of a CP-violating contribution in the HVV coupling. The amount of asymmetry of the distribution depends on the strength of CP-violation, which is modulated by $\tilde{d}$. Therefore, the average value of the CP-odd observable distribution shifts as a function of $\tilde{d}$. The curve given by the average value of a CP-odd observable as a function of $\tilde{d}$ is called the *gauge curve*. An example gauge curve of $\mathcal{OO}$ is shown in Figure 38.

For small values of $\tilde{d}$, the relation between $\langle\mathcal{OO}\rangle$ can be approximated by a linear function. Therefore, a linear fit can be performed in a range $\tilde{d} \ll 1$ with the following function:

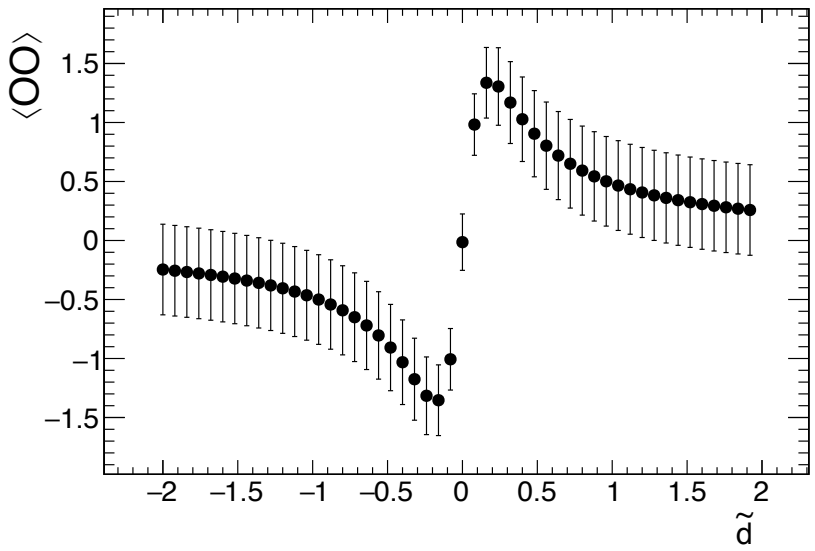


Figure 38: The average value of the $\mathcal{OO}$ as a function of the $\tilde{d}$, also known as the gauge curve. Uncertainties are statistical.

$$\langle \mathcal{OO} \rangle = a \cdot \tilde{d}, \tag{86}$$

where a represents the slope of the linear fit.

This results in the following for the $\tilde{d}$:

$$\tilde{d} = \frac{\langle \mathcal{OO} \rangle}{a}, \tag{87}$$

The error on $\tilde{d}$ ($s_{\tilde{d}}$) can be calculated using Gaussian error propagation:

$$s_{\tilde{d}} = \sqrt{\left(\frac{\partial \tilde{d}}{\partial a}\right)^2 s_a^2 + \left(\frac{\partial \tilde{d}}{\partial \langle \mathcal{OO} \rangle}\right)^2 s^2_{\langle \mathcal{OO} \rangle}} \tag{88}$$

$$= \sqrt{\left(\frac{\langle \mathcal{OO} \rangle}{a^2}\right)^2 s_a^2 + \left(\frac{1}{a}\right)^2 s^2_{\langle \mathcal{OO} \rangle}} \tag{89}$$

$$= \sqrt{\left(\frac{\tilde{d}}{a}\right)^2 s_a^2 + \left(\frac{1}{a}\right)^2 s^2_{\langle \mathcal{OO} \rangle}} \tag{90}$$

$$= \left|\frac{1}{a}\right| \sqrt{\tilde{d}^2 s_a^2 + s^2_{\langle \mathcal{OO} \rangle}}, \tag{91}$$

$$\tag{92}$$

,where s_a and $s_{\mathcal{OO}}$ are the errors on the fitted slope and the average value of the $\mathcal{OO}$ distribution.

Since each event is weighted, the standard error on the average value of the optimal observable is calculated in the following way:

$$s_{\langle \mathcal{OO} \rangle} = \frac{\sigma_{\mathcal{OO}}}{\sqrt{N_{\text{eff}}}}, \tag{93}$$

where $\sigma_{\mathcal{OO}}$ is the standard deviation of the Optimal Observable and the effective number of events, N_{eff}, is defined as the following:

$$N_{\text{eff}} = \frac{(\sum w)^2}{\sum w^2} \tag{94}$$

where the sum runs over the weight of every event.

A χ^2 fit is performed, which is a modified least-square method [47], in which the residual for each point is computed using the error on that point. The following value is minimized:

$$\chi^2 = \sum_i \left(\frac{y(i) - f(x(i))}{e(i)} \right)^2. \tag{95}$$

In Equation 95, $y(i)$ and $x(i)$ are y and x values for point i, $e(i)$ the error in the y direction of point i, and $f(x)$ the function to be fitted.

10.2.1 Selection of Fit Range

The linear approximation is not valid for larger values of $\tilde{d}$. Furthermore, it is also important to use a sufficient amount data points for the fit itself in order to decrease the error on the fit parameters. Although one expects the error of the fit parameter to be small in comparison with of the CP-odd observable distribution, optimizing the settings of the linear fit presents a possibility to improve the sensitivity.

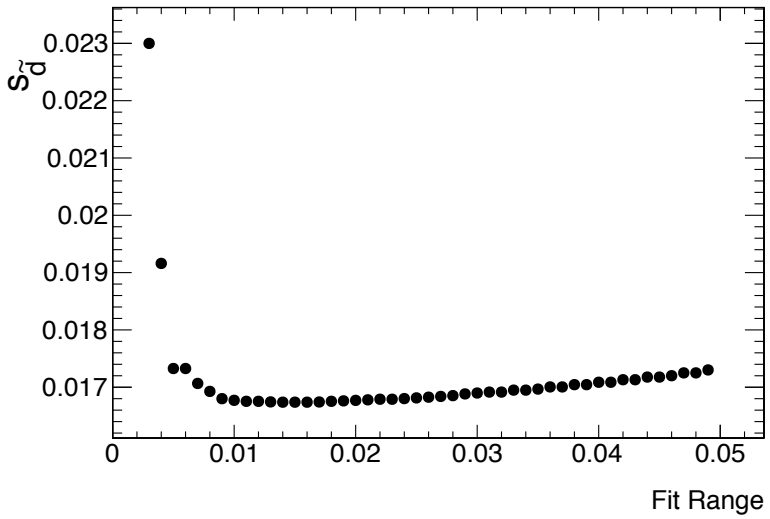
For this, the average value of the CP-odd observable is evaluated for 100 evenly spaced $\tilde{d}$ values between -0.05 and 0.05. 50 linear fits are performed within the range $[-a, a]$ where the value of a is scanned from 0.001 to 0.05 in steps of 0.001. The signal samples are reweighted so that for every $\tilde{d}$ the signal yield is equal to the SM case. For each of these fits, the error on the $\tilde{d}$ is computed, and the range resulting in the smallest error is found. The $\tilde{d}$ itself and its error are computed using an Asimov dataset corresponding to a $\tilde{d} = 0$. For $\mathcal{OO}$, the optimal range is found to be [-0.014, 0.014], resulting in an error on $\tilde{d}$ of 0.01673. For $\Delta\phi_{jj}^{signed}$ the range is chosen as was [-0.027, 0.027], which resulted in an uncertainty of 0.0250. For $\mathcal{O}^{Reg}$ a range of [-0.017, 0.017] is decided, which gave an uncertainty on $\tilde{d}$ of 0.0166. The error on $\tilde{d}$ as a function of the fit range for different CP-odd variables is shown in Figure 39.

The resulting estimations of $\tilde{d}$ for each CP-odd variable, evaluated using an Asimov dataset with $\tilde{d} = 0$, with 100 data points, are shown in Table 11.

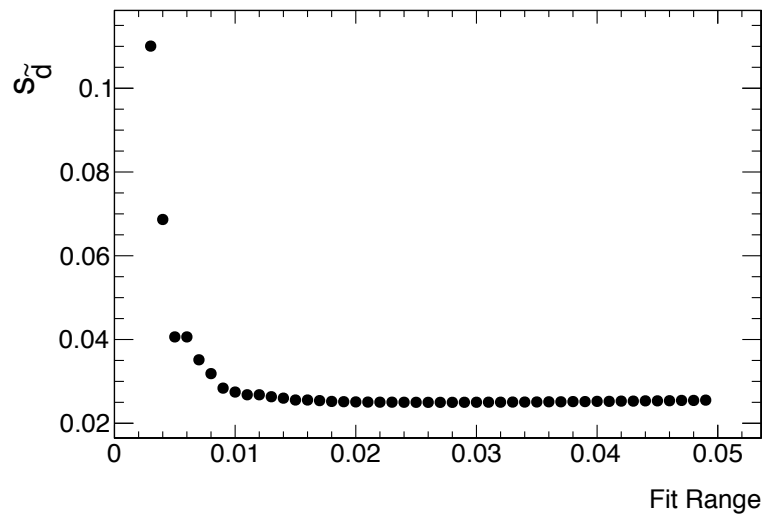
CP-Odd Observable	a	$\tilde{d}$
$\mathcal{OO}$	13.1± 2.7	0.003 ± 0.017
$\mathcal{O}^{Reg}$	-1.83 ± 0.31	0.004 ± 0.017
$\Delta\phi_{jj}^{signed}$	3.46 ± 0.56	0.014 ± 0.025

Table 11: The fitted slope of the linear regression performed in their respective optimized fit ranges, for all three CP-odd observables. The resulting estimation of $\tilde{d}$ and its error is shown.

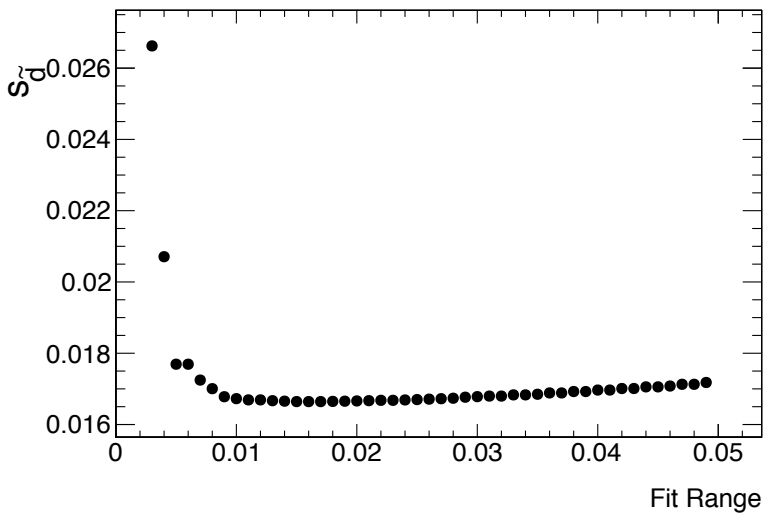
It can be seen, in all cases, the calculated value of $\tilde{d}$ does not significantly deviate from 0, which is to be expected since it was applied onto an Asimov dataset with $\tilde{d} = 0$. It can also be seen that limits on $\tilde{d}$ are similar for $\mathcal{OO}$ and $\mathcal{O}^{Reg}$. Both limits outperform the limits achieved by using $\Delta\phi_{jj}^{signed}$. This is expected since $\mathcal{OO}$ and $\mathcal{O}^{Reg}$ are both designed to be maximally sensitive to CP violation, while $\Delta\phi_{jj}^{signed}$ is not. In theory, $\mathcal{OO}$ should result in the best sensitivity, since it can be mathematically shown, that at the truth-level $\mathcal{OO}$ has the optimal sensitivity [30]. However, this may be obscured at the reconstruction-level due to the finite resolution of detectors which could result in a similar performance as that of $\mathcal{O}^{Reg}$.



(a) $\mathcal{OO}$



(b) $\Delta\phi_{jj}^{signed}$

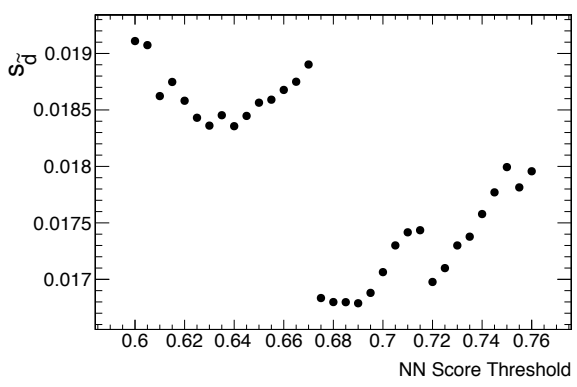


(c) $\mathcal{O}^{Reg}$

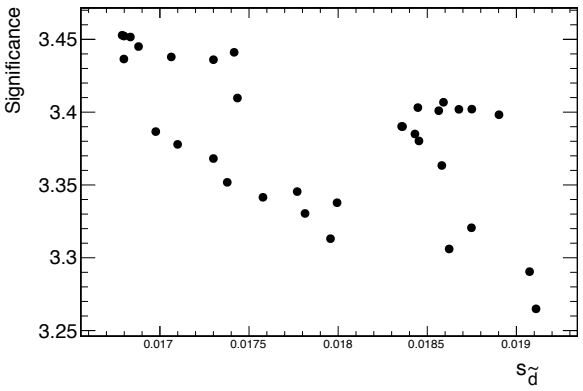
Figure 39: The uncertainty on $\tilde{d}$ as a function of the fit range, computed using three different CP-odd observables.

10.2.2 Influence of the Threshold on the Neural Network Score on the Sensitivity on $\tilde{d}$

The threshold selection on the neural network score, which defines the signal region, is determined such that the significance is maximal. This allows for a balance between good background rejection and large signal statistics. This section presents a study conducted in order to validate the selection of this neural network score threshold. To achieve this, different neural network score thresholds are scanned from 0.6 to 0.76, corresponding to range of $\pm$ 0.08 around the nominal threshold, 0.68, in steps of 0.005. Linear fits are then performed separately for each threshold value for each CP-odd variable using the optimized settings described in Section 10.2.1. The resulting errors on $\tilde{d}$ as a function of the NN score threshold for each CP-odd variable are shown in Figures 40c to 40i. Furthermore, change in the significance $s/\sqrt{s+b}$ as a function of the error on $\tilde{d}$ are also shown.

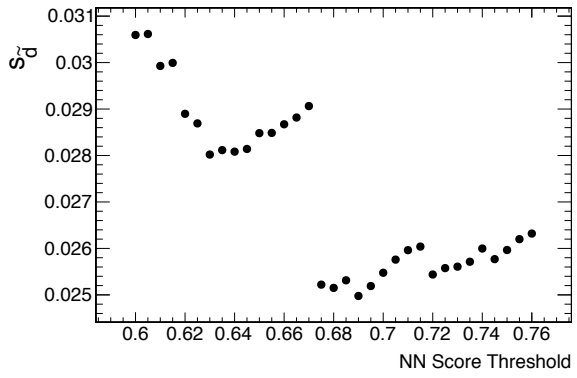


(a) $s_{\tilde{d}}$ as a function of the NN score threshold.

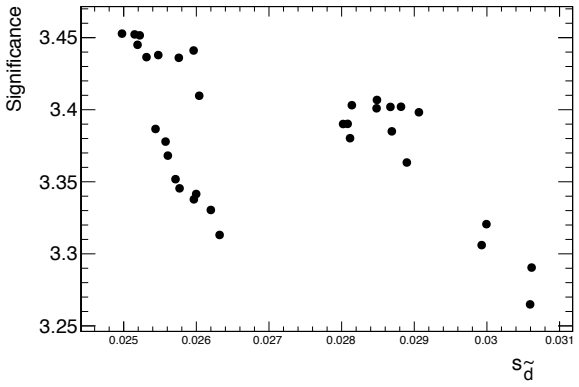


(b) Significance as a function of $s_{\tilde{d}}$.

(c) $\mathcal{OO}$

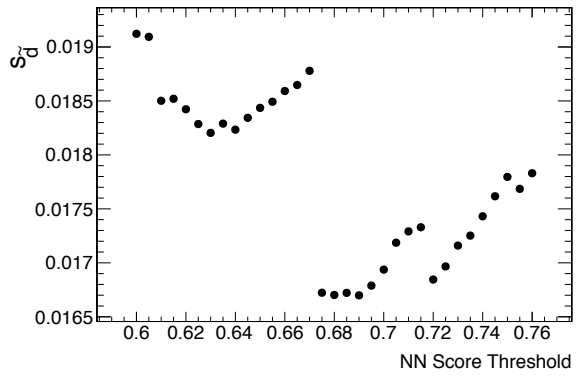


(d) $s_{\tilde{d}}$ as a function of the NN score threshold.

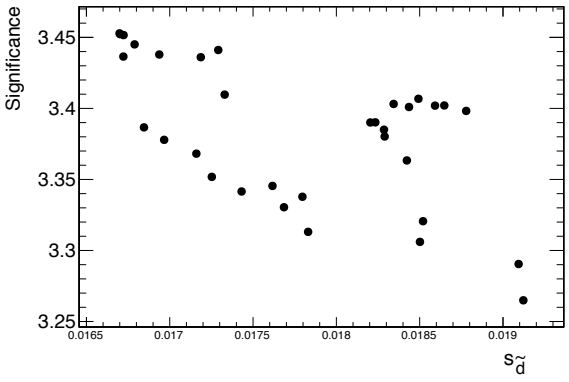


(e) Significance as a function of $s_{\tilde{d}}$.

(f) $\Delta\phi_{jj}^{signed}$



(g) $s_{\tilde{d}}$ as a function of the NN score threshold.



(h) Significance as a function of $s_{\tilde{d}}$.

(i) $\mathcal{O}^{Reg}$

Figure 40: The error on $\tilde{d}$ as a function of the NN score threshold applied onto the dataset, computed using different CP-odd variables, as well as the significance as a function of $s_{\tilde{d}}$

In all three cases, the curve is characterized by a minima approximately around the nominal NN score threshold of 0.68. The threshold resulting in the absolute minimum error deviates from this nominal threshold by no more than 0.01. Large drops in the curve are due to individual events with large weights being removed at certain NN score thresholds.

In the scatter plots it can be observed that a general negative correlation exists between these two variables, i.e. the higher the significance is, the smaller the limits are on $\tilde{d}$.

10.2.3 Validation of the Gauge Curve Method

In order to demonstrate the validity of the gauge curve method, the fit procedure is repeated using Asimov datasets with different values of $\tilde{d}$ in order to see whether the linear fit is able to recover the original value of $\tilde{d}$. Asimov $\tilde{d}$ values from -0.5 to 0.5 in steps of 0.05 are probed for all three CP-odd variables separately. This range was chosen since it goes beyond the range of linear behavior (see Figure 38). The resulting values for $\tilde{d}$ are shown in Tables 12, 13 and 14.

$\tilde{d}$	
Asimov	Estimated
-0.5000	-0.0579 ± 0.0269
-0.4500	-0.0631 ± 0.0272
-0.4000	-0.0689 ± 0.0275
-0.3500	-0.0752 ± 0.0278
-0.3000	-0.0818 ± 0.0282
-0.2500	-0.0878 ± 0.0284
-0.2000	-0.0915 ± 0.0282
-0.1500	-0.0893 ± 0.0270
-0.1000	-0.0755 ± 0.0241
-0.0500	-0.0440 ± 0.0195
-0.0000	0.0030 ± 0.0168
0.0500	0.0505 ± 0.0201
0.1000	0.0831 ± 0.0251
0.1500	0.0983 ± 0.0282
0.2000	0.1016 ± 0.0296
0.2500	0.0989 ± 0.0298
0.3000	0.0936 ± 0.0297
0.3500	0.0876 ± 0.0293
0.4000	0.0817 ± 0.0289
0.4500	0.0762 ± 0.0285

Table 12: Comparison of the expected and estimated $\tilde{d}$ values from the gauge curve method applied, for different values of $\tilde{d}$ using $\mathcal{OO}$.

It can be seen that the gauge curve method is only able to retrieve the original $\tilde{d}$ values when $\tilde{d}$ is small, i.e. between -0.1 and 0.1. Below and above this region, the fit generally

$\tilde{d}$	
Asimov	Measured
-0.5000	-0.0496 ± 0.0272
-0.4500	-0.0548 ± 0.0274
-0.4000	-0.0606 ± 0.0276
-0.3500	-0.0669 ± 0.0279
-0.3000	-0.0733 ± 0.0282
-0.2500	-0.0792 ± 0.0284
-0.2000	-0.0827 ± 0.0285
-0.1500	-0.0802 ± 0.0281
-0.1000	-0.0657 ± 0.0271
-0.0500	-0.0335 ± 0.0256
-0.0000	0.0141 ± 0.0251
0.0500	0.0619 ± 0.0268
0.1000	0.0945 ± 0.0289
0.1500	0.1095 ± 0.0303
0.2000	0.1125 ± 0.0307
0.2500	0.1094 ± 0.0305
0.3000	0.1039 ± 0.0302
0.3500	0.0977 ± 0.0298
0.4000	0.0917 ± 0.0295
0.4500	0.0860 ± 0.0291

Table 13: The measured $\tilde{d}$ using the gauge curve method applied onto asimov datasets for different values of $\tilde{d}$ using $\Delta\phi_{jj}^{signed}$.

underestimates the $\tilde{d}$. This demonstrates and confirms that the linear approximation is only valid for small values of $\tilde{d}$, and cannot be generalized to a larger range.

$\tilde{d}$	
Asimov	Measured
-0.5000	-0.0572 ± 0.0258
-0.4500	-0.0624 ± 0.0259
-0.4000	-0.0682 ± 0.0261
-0.3500	-0.0745 ± 0.0262
-0.3000	-0.0810 ± 0.0263
-0.2500	-0.0870 ± 0.0262
-0.2000	-0.0907 ± 0.0258
-0.1500	-0.0885 ± 0.0246
-0.1000	-0.0746 ± 0.0222
-0.0500	-0.0430 ± 0.0186
-0.0000	0.0040 ± 0.0167
0.0500	0.0515 ± 0.0192
0.1000	0.0841 ± 0.0231
0.1500	0.0993 ± 0.0257
0.2000	0.1026 ± 0.0270
0.2500	0.0998 ± 0.0274
0.3000	0.0945 ± 0.0275
0.3500	0.0885 ± 0.0273
0.4000	0.0826 ± 0.0271
0.4500	0.0771 ± 0.0269

Table 14: The measured $\tilde{d}$ using the gauge curve method applied onto asimov datasets for different values of $\tilde{d}$ using $\mathcal{O}^{Reg}$.

10.3 Maximum Likelihood Fit Method

The gauge curve method makes use of the average value of the entire $\mathcal{OO}$ distribution, which eliminates degrees of freedom that could otherwise be utilized to improve the estimation of $\tilde{d}$. Instead of reducing the entire distribution information into a single value, the Maximum-Likelihood method (ML) can utilize the full distribution information available.

The ML method compares bin-by-bin the predicted and measured values for different $\tilde{d}$ hypotheses and finds the value for which both distributions are most compatible. This compatibility is given by a product of Poisson distributions, and defined by the binned likelihood function ($\mathcal{L}$) [47]:

$$\mathcal{L}(\tilde{d}) = \prod_{i=0}^{N} \text{Pois}(n_i; n_i^{pred}(\tilde{d})) \tag{96}$$

where the Poisson distribution probability density function (PDF) is given by $\text{Pois}(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$. The likelihood function is summed over all N bins of a histogram with observed and predicted bin contents n_i and $n_i^{pred}(\tilde{d})$, for bin i respectively. In this analysis, the bin content $n_i = s_i + b_i$ is simply the sum of the signal and background event yield in that bin.

A larger value of the likelihood function implies a greater compatibility since the Poisson distribution has its maximum at λ. Therefore, one would like to find the value of $\tilde{d}$ for which the likelihood function is maximal. Instead the negative log-likelihood function is used as a convention:

$$-\ln \mathcal{L}(\boldsymbol{n}; \tilde{d}) = \text{NLL}(\tilde{d}) = -\ln\left(\prod_{i=0}^{N} \text{Pois}(n_i; n_i^{pred}(\tilde{d}))\right) \tag{97}$$

$$= \sum_{i=0}^{N} \left(n_i \ln n_i^{pred} - n_i^{pred} - \ln n_i!\right) \tag{98}$$

Finding the best-fit value therefore involves finding the value of $\tilde{d}$ for where the NLL function is minimal, NLL_{min}. The advantage of the negative log-likelihood method is the simplicity in determining *Confidence Intervals* (CI's). The confidence intervals at 68% and 95% confidence level can be determined by calculating the values of $\tilde{d}$ which correspond to a $\Delta\text{NLL} = \text{NLL} - \text{NLL}_{\text{min}}$ of 0.50 and 1.92 [47].

10.3.1 Settings for the Negative Log-Likelihood Method

When working with expected distributions, the signal yield is scaled so that for all $\tilde{d}$ values it stays the same as the SM prediction.

For testing the performance of different fit settings, the fit is applied onto an Asimov distribution, constructed using the predicted observable distribution for a $\tilde{d} = 0$.

For each observable, different fit ranges are used. For the optimal observable, this is ± 15, for the signed $\Delta\phi_{jj}$ this is ± 3.2, and for the observable obtained through symbolic regression, this is ± 2, where the range is chosen such that the shape of the distribution is captured. The binning of these observables can be chosen freely, with either uniform bin widths or variable bin widths. One expects the performance of the fit to be dependent on the choice of the binning of the histogram used, which is investigated in the next section.

10.3.2 Binning Studies

Different choice of binnings of the observables are investigated to see their impact on the sensitivity to $\tilde{d}$. For all three CP-odd observables, histograms with 10, 20, 40 and 100 bins are used for the fit and their resulting sensitivities are compared. Furthermore, the same study is repeated with using only signal samples to validate that the same trend is observed with distributions that are in theory maximally sensitive to $\tilde{d}$. These values are shown in Table 15 and 16.

In the case of the signal-only ML fits, a larger bin numbers are also investigated, which is shown in Table 17. For fits having both signal and background events, 200 equidistant values of $\tilde{d}$ are scanned between -0.05 and 0.05, while for the signal-only fits, a range between -0.04 and 0.04 is used

N_{bins}	5	10	15	20	25
$\mathcal{OO}$	$^{+0.0390}_{-0.0395}$	$^{+0.0345}_{-0.0340}$	0.0340	$^{+0.0335}_{-0.0320}$	0.0330
$\mathcal{O}^{Reg}$	$^{+0.0375}_{-0.0380}$	0.0340	0.0335	-	-
$\Delta\phi_{jj}^{signed}$	$^{+0.0455}_{-0.0460}$	$^{+0.0435}_{-0.0440}$	$^{+0.0425}_{-0.0430}$	0.0420	0.0415

Table 15: The limits on $\tilde{d}$ for different choices of binning. In case of asymmetric limits, the upper and lower limits are distinguished.

N_{bins}	5	10	15	20	25
$\mathcal{OO}$	0.0264	0.0244	0.0240	0.0240	0.0240
$\mathcal{O}^{Reg}$	0.0272	0.0252	0.0248	0.0248	0.0248
$\Delta\phi_{jj}^{signed}$	0.0300	0.0288	0.0284	0.0284	0.0284

Table 16: The limits on $\tilde{d}$ for different choices of binning using only signal samples.

In all three cases, the nominal value for the calculated $\tilde{d}$ was not significantly different from zero. In some cases, finer binning caused certain bins in the distribution to have a zero or negative bin content, whose negative log-likelihood value cannot be computed. These are represented by the dashes in Table 15. This is also the reason why in Table 15, a bin number limit of 25 was chosen, since beyond this bin number, for all variables, the fit failed.

N_{bins}	10	20	40	100
$\mathcal{OO}$	0.0246	0.0240	0.0240	0.0240
$\mathcal{O}^{Reg}$	0.0252	0.0249	0.0246	0.0246
$\Delta\phi_{jj}^{signed}$	0.0288	0.0284	0.0282	0.0282

Table 17: The limits on $\tilde{d}$ put using an asimov dataset for $\tilde{d} = 0$, for different binnings of ML method, using only signal samples.

The reason why negative bins have not appeared in signal-only fits, despite smaller bin sizes, is because large negative weight events are predominantly found to be present in background samples, especially among fake events. Therefore, for certain ranges of $\tilde{d}$, other processes would often not be able to compensate the negative weight of such events.

It can be seen in both Tables 15 and 16 that increasing the number of bins improves the error on $\tilde{d}$. However, the improvement is not as pronounced beyond 20 bins. In order to optimize the sensitivity, the number of bins should be large as possible but not large enough to cause zero or negative bin contents. For this, an algorithm is implemented, which adjusts the bins of a histogram with uniform binning into a histogram with variable bin width. These new bin widths are defined so that each bin yield is above a predefined positive threshold value. This is done by looping over the bins and adding their contents from the center bin outwards, while at the same time merging them. When the sum reaches the threshold, it is set to 0, and a bin edge is set at that point.

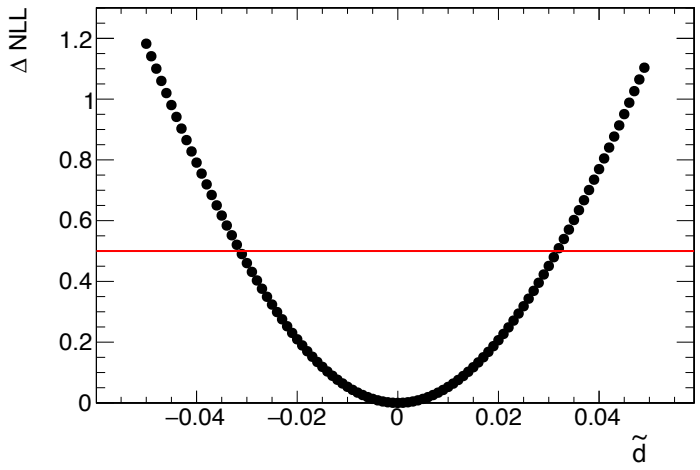
For each variable, a histogram with 100 bins is initialized. The minimum bin content threshold is increased in steps of 0.05 starting from zero until the ML fitting procedure no longer fails. For the Optimal Observable, this point was 0.1, and for $\Delta\phi_{jj}^{signed}$ and $\mathcal{O}^{Reg}$ at 0.05. The resulting bin edges for each variable is listed in Table 22 in the appendix.

The resulting 68% confidence intervals when using $\mathcal{OO}$, $\mathcal{O}^{Reg}$ and $\Delta\phi_{jj}^{signed}$ observables are [-0.0310, 0.0315], [-0.0310, 0.0305] and [-0.0380, 0.0385], respectively. In all cases, the nominal value of $\tilde{d}$ is compatible with 0. With all three variables, an improvement in the sensitivity is observed in comparison with the case with uniform bin widths. An improvement of 5%, 8% and 8% in the range of the CI's is achieved when using $\mathcal{OO}$, $\mathcal{O}^{Reg}$ and $\Delta\phi_{jj}^{signed}$ variables, respectively. As expected, the setup using $\Delta\phi_{jj}^{signed}$ underperforms compared both $\mathcal{OO}$ and $\mathcal{O}^{Reg}$, while the latter two perform similar to each other. The value of ΔNLL as a function of $\tilde{d}$, also known as the *NLL curve*, is shown in in Figure 41.

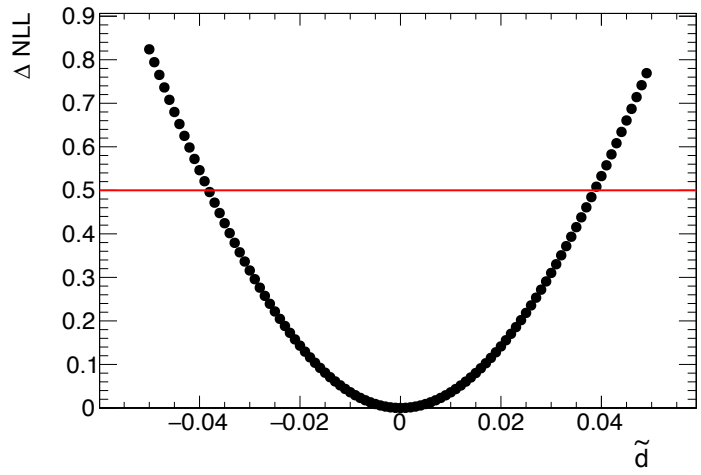
10.3.3 Investigation on the Influence of the Threshold on the Neural Network Score on the Sensitivity to CP Violation

As with the gauge curve method, the impact of the NN score threshold on the limits on $\tilde{d}$ is investigated by varying the threshold from 0.6 to 0.76. Since in this case, asymmetrical uncertainties on $\tilde{d}$ are possible, the measure of the sensitivity was evaluated by taking half of the width of the confidence interval, effectively taking the average of the upper and lower boundaries. The average error on $\tilde{d}$ as a function of the NN score threshold is

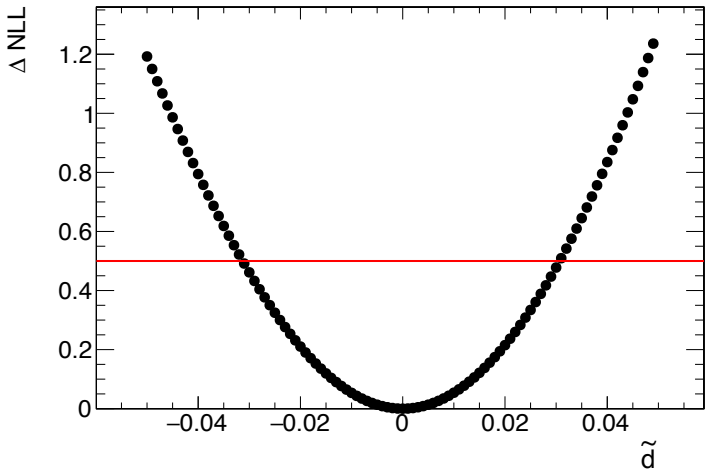
shown in Figures 42c to 42i. Change in the significance as a function of the average error on $\tilde{d}$ is shown in Figures 42c to 42i.



(a) $\mathcal{OO}$

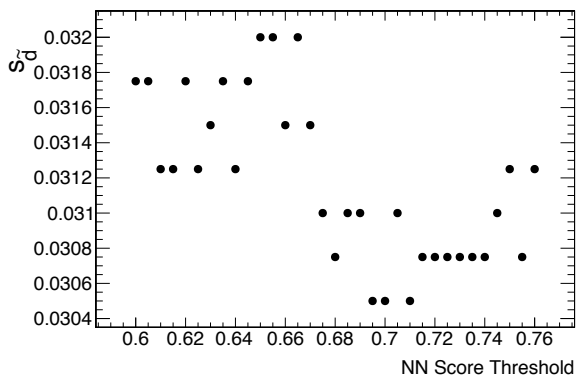


(b) $\Delta\phi_{jj}^{signed}$

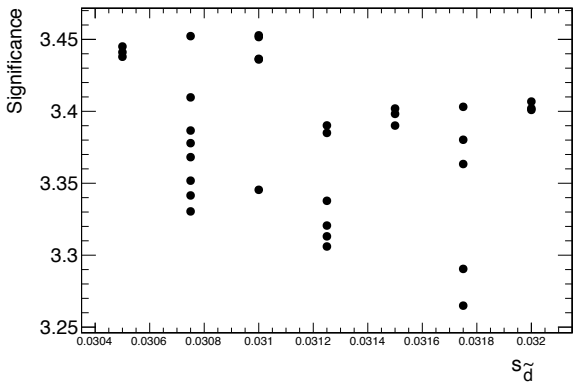


(c) $\mathcal{O}^{Reg}$

Figure 41: The value ΔNLL as a function of $\tilde{d}$ for three different CP-odd variables, using settings optimized in Section 8.3.2. A total of 100 points between ± 0.05 are evaluated. Marked in red is the value $\Delta\text{NLL} = 0.5$, corresponding to the 68% confidence interval.

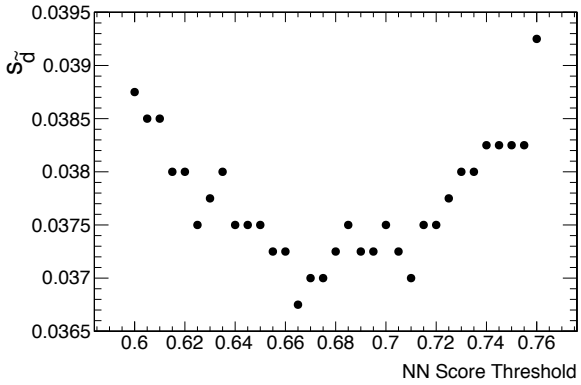


(a) $s_{\tilde{d}}$ as a function of the NN score threshold.

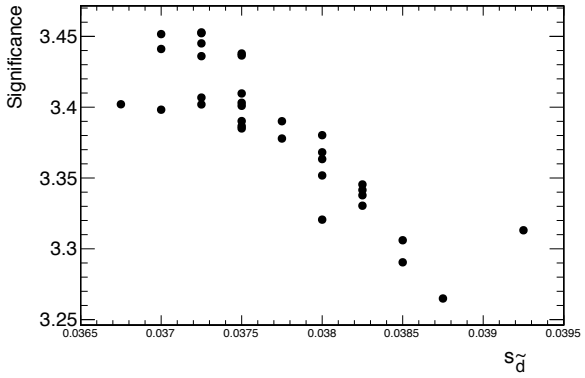


(b) Significance as a function of $\langle s_{\tilde{d}} \rangle$.

(c) $\mathcal{OO}$

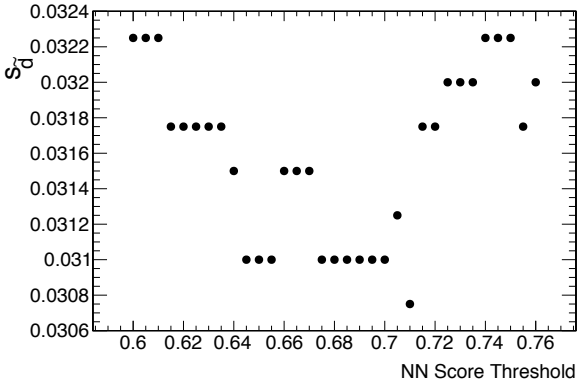


(d) $s_{\tilde{d}}$ as a function of the NN score threshold.

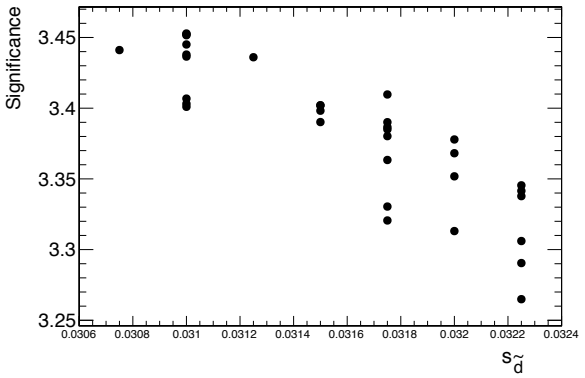


(e) Significance as a function of $\langle s_{\tilde{d}} \rangle$.

(f) $\Delta\phi_{jj}^{signed}$



(g) $s_{\tilde{d}}$ as a function of the NN score threshold.



(h) Significance as a function of $\langle s_{\tilde{d}} \rangle$.

(i) $\mathcal{O}^{Reg}$

Figure 42: The average error on $\tilde{d}$ as a function of the NN score threshold, computed using different CP-odd variables using the ML method, the significance as a function on the average error.

When using $\Delta\phi_{jj}^{signed}$ and $\mathcal{O}^{Reg}$, the curves show a clear minima around the threshold 0.68, while this is not as apparent in the case of $\mathcal{OO}$. There it seems to minimize at around 0.7. Furthermore, one does not observe any significant jumps in the curve, which was the case with the gauge curve method originating from events with large weights. This is due to the fact that ML utilizes the information of the entire distribution.

When observing the relationship between the significance and the average error on $\tilde{d}$, as with the gauge curve method, a general negative correlation between the two variables is noticeable. For $\Delta\phi_{jj}^{signed}$ and $\mathcal{O}^{Reg}$ observables this correlation can be seen clearly, while for $\mathcal{OO}$ this is not as clear.

10.3.4 Validation of the ML method

Asimov $\tilde{d}$ values from -0.2 to 0.2 in steps of 0.05 are investigated, for all three CP-odd variables. Larger ranges of $\tilde{d}$, as in the gauge curve method, are not investigated since larger $\tilde{d}$ values result in larger event weights, causing the fits to become unstable due to negative bin contents. For the same reason, the minimum bin content threshold is also raised to 0.2. Furthermore, the scan range was increased from 0.05 to 0.25, and to 0.4 when using $\Delta\phi_{jj}^{signed}$. The resulting values of $\tilde{d}$ are shown in Tables 18 to 20

Asimov $\tilde{d}$	Measured $\tilde{d}$
-0.3000	$-0.3000^{+0.1025}_{-0.2150}$
-0.2500	$-0.2500^{+0.0850}_{-0.1600}$
-0.2000	$-0.2000^{+0.0700}_{-0.1150}$
-0.1500	$-0.1500^{+0.0600}_{-0.0875}$
-0.1000	$-0.1000^{+0.0450}_{-0.0675}$
-0.0500	$-0.0500^{+0.0325}_{-0.0450}$
-0.0000	$0.0000^{+0.0325}_{-0.0300}$
0.0500	$0.0500^{+0.0450}_{-0.0325}$
0.1000	$0.1000^{+0.0650}_{-0.0450}$
0.1500	$0.1500^{+0.0875}_{-0.0600}$
0.2000	$0.2000^{+0.1275}_{-0.0750}$
0.2500	$0.2500^{+0.1750}_{-0.0900}$
0.3000	$0.3000^{+0.2350}_{-0.1075}$

Table 18: The measured $\tilde{d}$ using the ML method applied onto asimov datasets for different values of $\tilde{d}$ using $\mathcal{OO}$.

As shown, the ML method is able to retrieve the original value of $\tilde{d}$ from the Asimov dataset between -0.3 and 0.3. It is also observed that for extreme values of $\tilde{d}$, the width of the confidence interval increases as well.

Asimov $\tilde{d}$	Measured $\tilde{d}$
-0.3000	$-0.3000\ ^{+0.1760}_{-0.3120}$
-0.2500	$-0.2500\ ^{+0.1560}_{-0.2640}$
-0.2000	$-0.2000\ ^{+0.1200}_{-0.2240}$
-0.1500	$-0.1500\ ^{+0.0840}_{-0.2040}$
-0.1000	$-0.1000\ ^{+0.0520}_{-0.1600}$
-0.0500	$-0.0500\ ^{+0.0400}_{-0.0560}$
-0.0000	$0.0000\ ^{+0.0360}_{-0.0360}$
0.0500	$0.0500\ ^{+0.0560}_{-0.0400}$
0.1000	$0.1000\ ^{+0.1720}_{-0.0520}$
0.1500	$0.1500\ ^{+0.2080}_{-0.0800}$
0.2000	$0.2000\ ^{+0.2280}_{-0.1200}$
0.2500	$0.2500\ ^{+0.2640}_{-0.1560}$
0.3000	$0.3000\ ^{+0.3240}_{-0.1880}$

Table 19: The measured $\tilde{d}$ using the ML method applied onto asimov datasets for different values of $\tilde{d}$ using $\Delta\phi_{jj}^{signed}$.

Asimov $\tilde{d}$	Measured $\tilde{d}$
-0.3000	$-0.3000\ ^{+0.1075}_{-0.2225}$
-0.2500	$-0.2500\ ^{+0.0900}_{-0.1650}$
-0.2000	$-0.2000\ ^{+0.0725}_{-0.1200}$
-0.1500	$-0.1500\ ^{+0.0625}_{-0.0925}$
-0.1000	$-0.1000\ ^{+0.0450}_{-0.0725}$
-0.0500	$-0.0500\ ^{+0.0325}_{-0.0475}$
-0.0000	$0.0000\ ^{+0.0300}_{-0.0300}$
0.0500	$0.0500\ ^{+0.0475}_{-0.0325}$
0.1000	$0.1000\ ^{+0.0700}_{-0.0475}$
0.1500	$0.1500\ ^{+0.0925}_{-0.0625}$
0.2000	$0.2000\ ^{+0.1275}_{-0.0750}$
0.2500	$0.2500\ ^{+0.1700}_{-0.0875}$
0.3000	$0.3000\ ^{+0.2350}_{-0.1075}$

Table 20: The measured $\tilde{d}$ using the ML method applied onto asimov datasets for different values of $\tilde{d}$ using $\mathcal{O}^{Reg}$.

10.4 Comparison of Fit Methods

In both the gauge curve and ML method, the settings are optimized to improve the sensitivity on $\tilde{d}$. An overview of the computed values of $\tilde{d}$, evaluated using an Asimov dataset with $\tilde{d} = 0$ is shown in Table 21.

CP-Odd Observable	Gauge Curve Method	ML Method
$\mathcal{OO}$	0.003±0.017	$0.000\ ^{+0.0315}_{-0.0310}$
$\mathcal{O}^{Reg}$	0.004±0.017	$0.000^{+0.0305}_{-0.0310}$
$\Delta\phi_{jj}^{signed}$	0.014±0.025	$0.000^{+0.0385}_{-0.0380}$

Table 21: Estimated $\tilde{d}$ values using the gauge curve and ML methods for three different CP-odd variables.

In none of the cases is the estimated $\tilde{d}$ found to be significantly different from 0. In both gauge curve and ML methods, $\Delta\phi_{jj}^{signed}$ underperformed relative to both $\mathcal{OO}$ and $\mathcal{O}^{Reg}$. This is expected since both $\mathcal{OO}$ and $\mathcal{O}^{Reg}$ are specifically optimized for their sensitivity on $\tilde{d}$. The performance of $\mathcal{OO}$ and $\mathcal{O}^{Reg}$ were found to be similar in both methods.

When comparing the limits on $\tilde{d}$ evaluated using the gauge curve and ML methods, it is observed that the gauge curve method performs better than the ML method. This is unexpected since the ML method uses the full distribution information and therefore is able to exploit more degrees of freedom, while the gauge curve method relies on a single averaged value, thereby possibly discarding important information.

11 Conclusion

In this thesis, an investigation of the CP properties of the HVV vertex in Higgs-boson production via VBF in the $H \to \tau\tau \to e\mu 4\nu$ decay channel is presented. For this, the full Run-2 dataset corresponding to an integrated luminosity of 139.1 fb^{-1} is used. After an initial selection procedure to increase the signal-to-background ratio, two different methods on estimating the expected uncertainty on $\tilde{d}$ are compared.

Firstly the data-driven matrix method was used to estimate the contribution of background processes where jets is misidentified as leptons, also known as fakes. Fake electrons and muons mostly originate from multi-jet QCD production and W-boson production. For this, real and fake efficiencies were measured in their respective control regions. This was done separately for electrons and muons as a function the lepton p_{T} and $|\eta|$. This information was then used to calculate weights to be applied onto data events, which yields an estimate of the total fake contribution. The resulting fake estimation is then validated in a validation region defined such that the two leptons have the same signed charge.

A cut-based analysis was then employed to initially increase the signal to background ratio. These event selection requirements exploit the kinematic differences between signal and background events in order to suppress the contribution of background processes. These were split into the preselection cuts, which targets and reduces background processes, and the VBF topology cuts, which specifically targets VBF Higgs-boson production events. These selection requirements define the VBF region, in which neural networks are used to further improve the signal-to-background ratio.

After the application of event selection requirements, neural networks were trained to discriminate signal events from background events. The neural networks were trained on simulated data, which was split according to 5-fold mixed cross validation. Hyperparameters, such as the batch size, number of layers were optimized using Optuna, which uses a Bayesian optimization algorithm called the Tree-Structured Parzen Estimator. The performance of the neural network was evaluated by varying the neural network score thereshold, and finding for which threshold the maximum significance $s/\sqrt{s+b}$, is reached. Firstly, the input variable set was determined by computing the permutation importance on a large set of variables and removing those which do not contribute to the improvement of the performance. Furthermore, the number of output nodes were varied as well to differentiate different background processes. Lastly, the performance of the use of dropout layers as an alternative to L2 regularization was investigated. The optimized neural network has a learning rate of 0.0034, batch size of 128, 5 layers with 256 nodes per layer, L2 regularization parameter of 2.36×10^{-6}, with 4 output nodes considering: signal, $Z \to \tau\tau$, top and other background processes. This resulted in an maximum significance of 3.49, with a neural network score threshold of 0.68. The application of the neural network resulted in an increase of the signal-to-background ratio from 0.07 to 0.72. The selection on the neural network score then defines the signal region. Previous analyses also showed no signed of CP violation and the expected constraints on $\tilde{d}$ were $\tilde{d} \in [-0.035, 0.033]$ at the 68% confidence level [28].

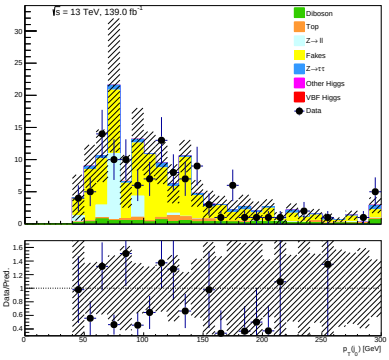
Two different statistical methods were applied onto an Asimov dataset of $\tilde{d} = 0$ to extract

expected limits on $\tilde{d}$ in the signal region. For this, three CP-odd variables, $\mathcal{OO}$, $\Delta\phi_{jj}^{signed}$, and $\mathcal{O}^{Reg}$ were investigated. One such method utilizes the gauge curve, which assumes a linear relation between the average value of the CP-odd variable and $\tilde{d}$ for small $\tilde{d}$ values. Settings such as the fit range and the cut on the NN score were optimized to yield the smallest expected uncertainty on $\tilde{d}$. It was observed that the optimal observable and the observable obtained from symbolic regression have similar performances, and both outperformed the signed azimuthal angle difference between the tagging jets. The expected sensitivity on $\tilde{d}$ using the full Run-2 dataset is $\tilde{d} \in [-0.017, 0.017]$ for $\mathcal{OO}$ and $\mathcal{O}^{Reg}$, and $[-0.025, 0.025]$ for $\Delta\phi_{jj}^{signed}$ using the gauge curve method. Another statistical method is the maximum likelihood fit to the full distribution of the CP-odd observable, which in contrast to the gauge curve method, utilizes the full distribution information to estimate limits on $\tilde{d}$. As with the linear method, settings such as the binning were optimized to yield the best performance. For the ML method, The expected sensitivity on $\tilde{d}$ are $\tilde{d} \in [-0.0310, 0.0315]$ for $\mathcal{OO}$, $[-0.0310, 0.305]$ for $\mathcal{O}^{Reg}$, and $[-0.0380, 0.0385]$ for $\Delta\phi_{jj}^{signed}$. It was observed with the ML method, that $\mathcal{OO}$ and $\mathcal{O}^{Reg}$ resulted in similar performances, while at the same time outperforming $\Delta\phi_{jj}^{Signed}$. Furthermore it was observed that the gauge curve method consistently outperforms the NLL method, which is unexpected, since the NLL method takes the entire distribution information instead of one value. The exact reason for this phenomenon is unknown, and therefore further investigations are necessary. Nonetheless, this thesis presents the first expected limits on $\tilde{d}$ using a simulated data corresponding to an integrated luminosity of 139 fb^{-1} in the VBF $H \to \tau\tau \to e\mu 4\nu$ channel.

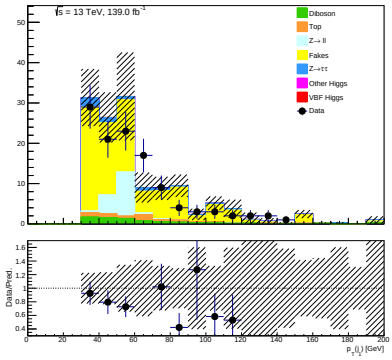
Further improvements and studies are possible. One assumption made when applying the matrix method for the estimation of the fake background is that the lepton efficiencies are independent of the jet multiplicity and the data taking period. A study could have been done to validate this assumption by measuring and comparing the efficiencies for different years and jet multiplicities. Another point of improvement regarding the efficiency measurements pertains to truth matching. The current official truth matching criteria for ATLAS analyses utilizes the categorization scheme given by the International Fake Forum (IFF). However, in the samples used for this thesis, information regarding the IFF truth categorization was not present, and therefore those of the MCTruthClassifier were used, which are no longer officially used. Therefore, the use of IFF categories could have yielded more accurate measurements of the electron and muon efficiencies. Lastly, in a full analysis, one would also consider the effects of systematic uncertainties on the obtained expected limits on $\tilde{d}$. However, due to time and technical constraints, investigations on the effect of systematic uncertainties were not made. If studies on systematic uncertainties had been made, the thesis could have possibly encompassed a full analysis on the CP properties of the HVV coupling, applied onto measured data instead of an Asimov dataset. These limits can then be compared with previous analyses.

A Appendix

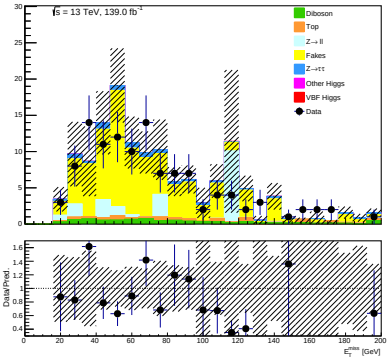
A.1 Same-Sign Fake Validation Region Distributions



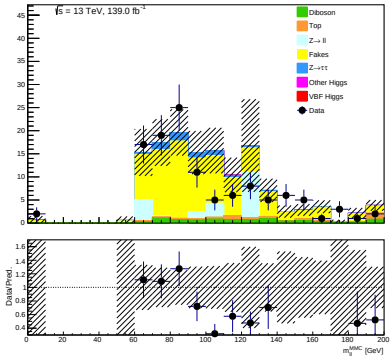
(a) $p_{\mathrm{T}}(j_0)$



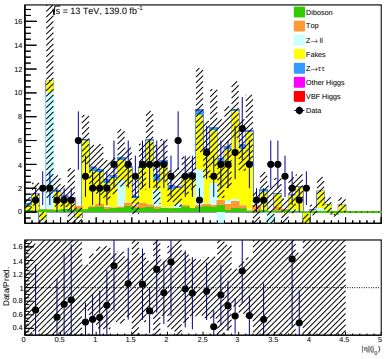
(b) $p_{\mathrm{T}}(j_1)$



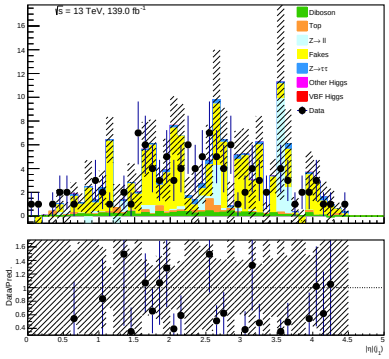
(c) E_{T}^{miss}



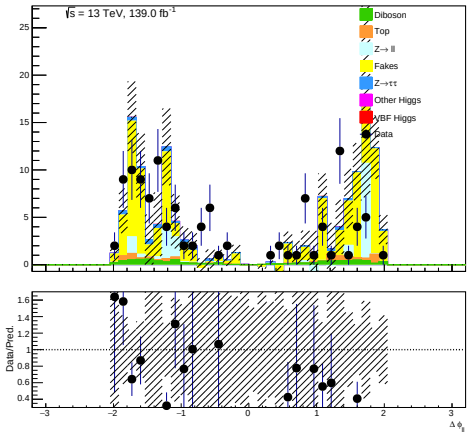
(d) m_{ll}^{MMC}



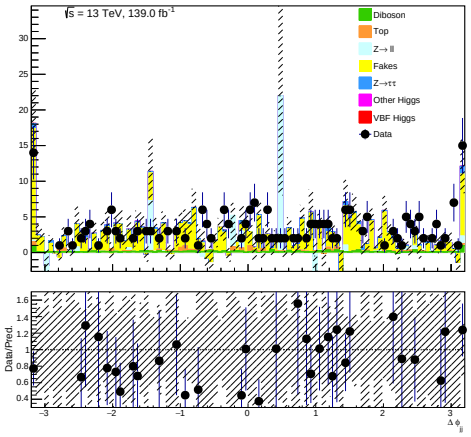
(e) $|\eta|(j_0)$



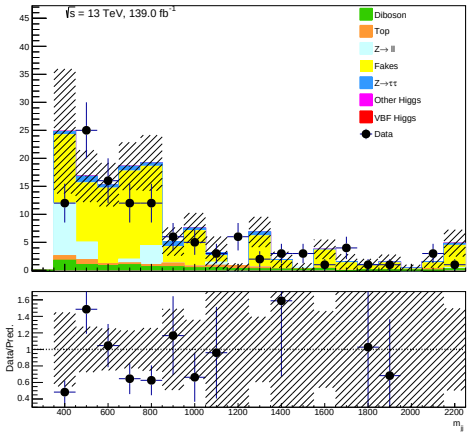
(f) $|\eta|(j_1)$



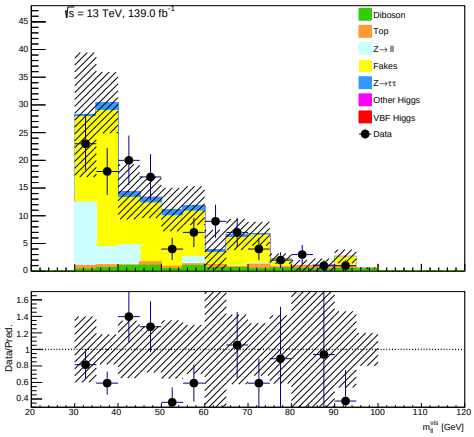
(g) $\Delta\phi_{ll}$



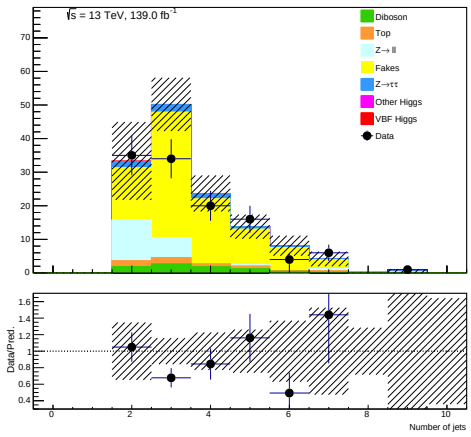
(h) $\Delta\phi_{jj}$



(i) m_{jj}



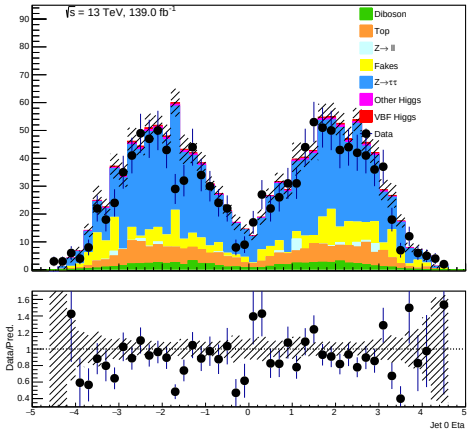
(j) m_{ll}



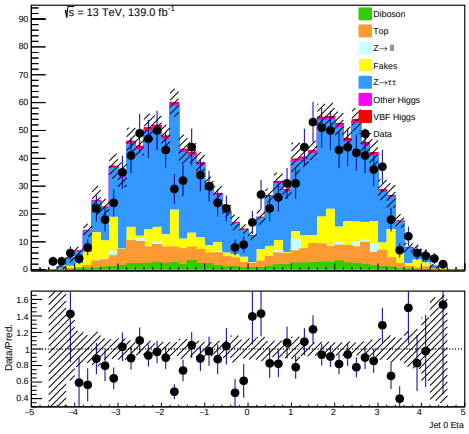
(k) N_{jet}

Figure 44: Expected and data distributions in the same-sign fake validation region of different kinematic variables. Error bars represent statistical errors.

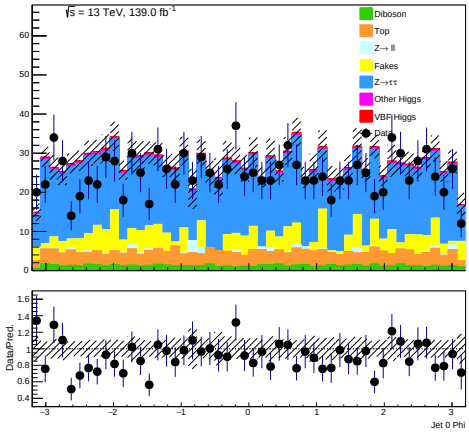
A.2 Low Neural Network Score Region Distributions.



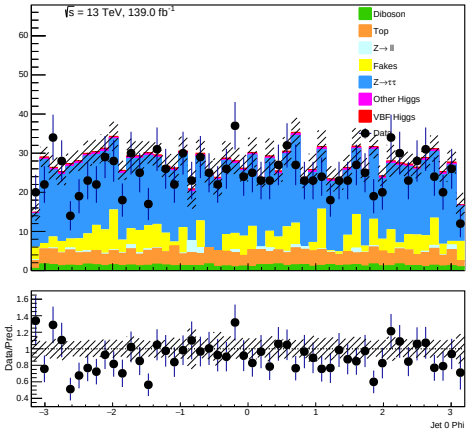
(a) η^{j_0}



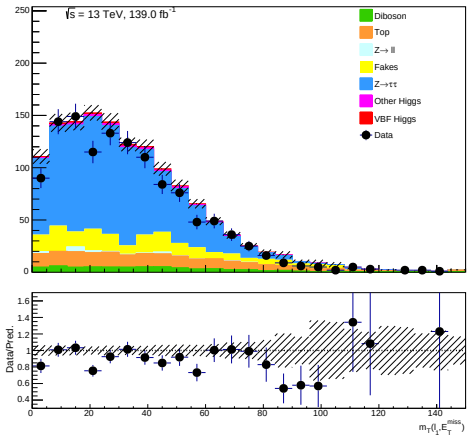
(b) η^{j_1}



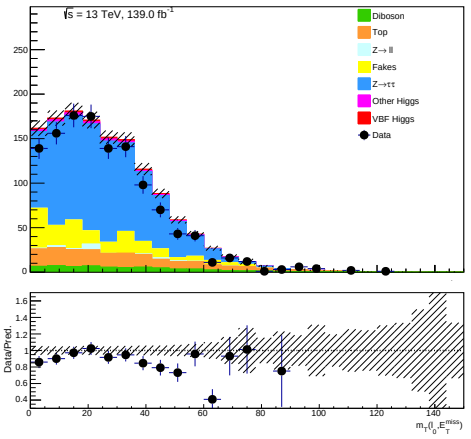
(c) ϕ^{j_0}



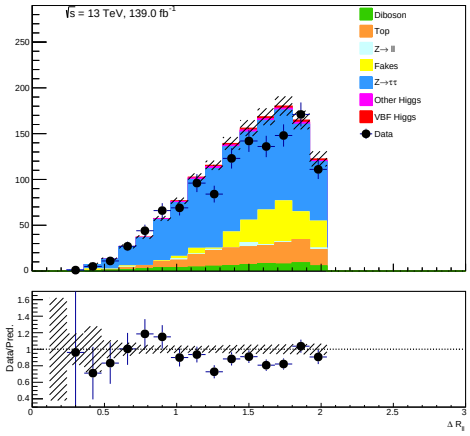
(d) ϕ^{j_1}



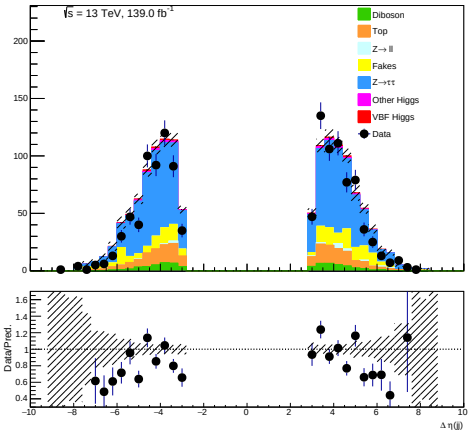
(e) $m_T(l_0, E_{\mathrm{T}}^{miss})$



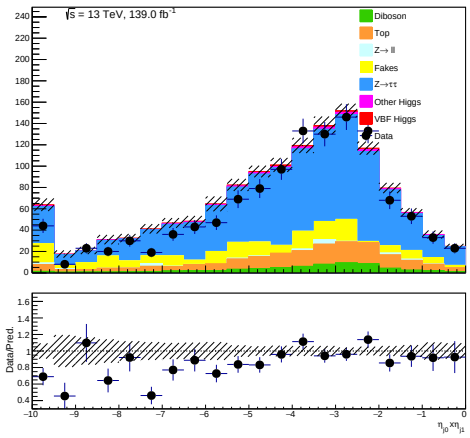
(f) $m_T(l_1, E_{\mathrm{T}}^{miss})$



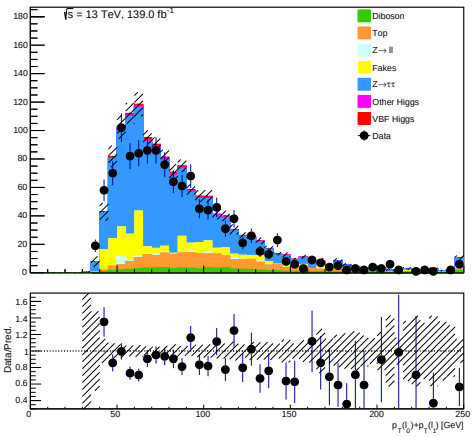
(g) ΔR_{ll}



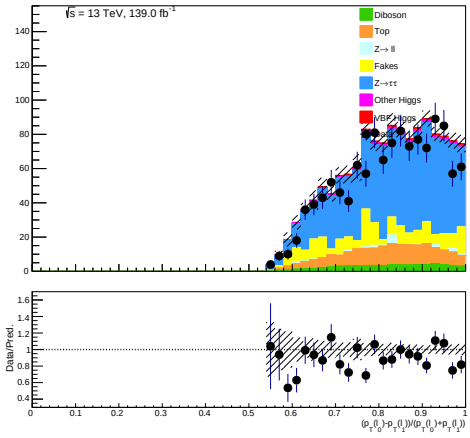
(h) $\Delta\eta_{jj}$



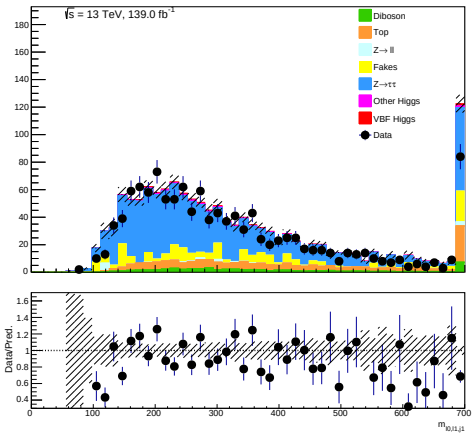
(i) $\eta_{j_0} \times \eta_{j_1}$



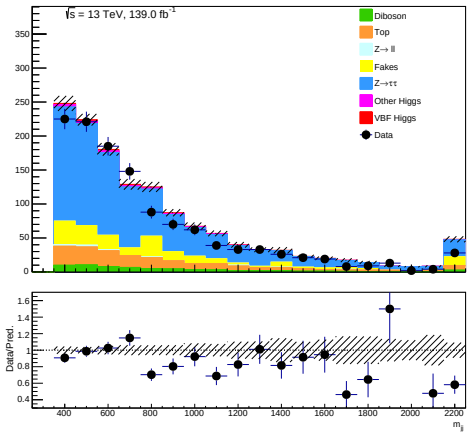
(j) $p_{\mathrm{T}}(l_0) + p_{\mathrm{T}}(l_1)$



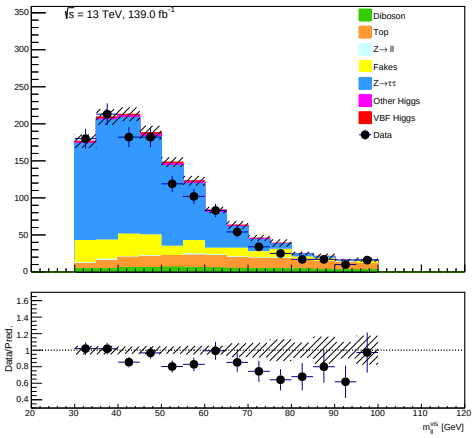
(k) $\frac{(p_{\mathrm{T}}(l_0)-p_{\mathrm{T}}(l_1))}{(p_{\mathrm{T}}(l_0)+p_{\mathrm{T}}(l_1))}$



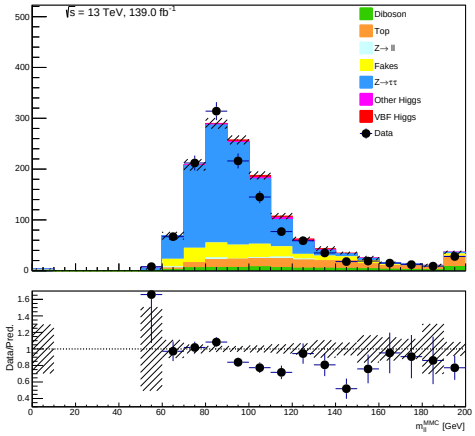
(l) m_{l_0,l_1,j_1}



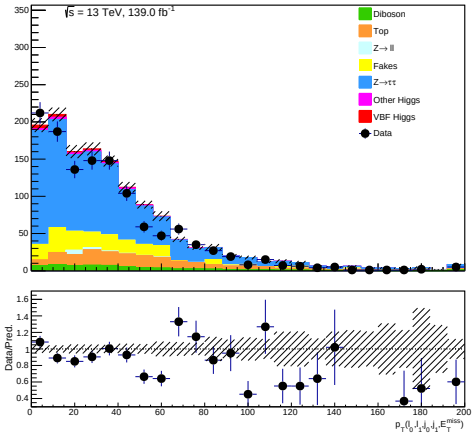
(m) m_{jj}



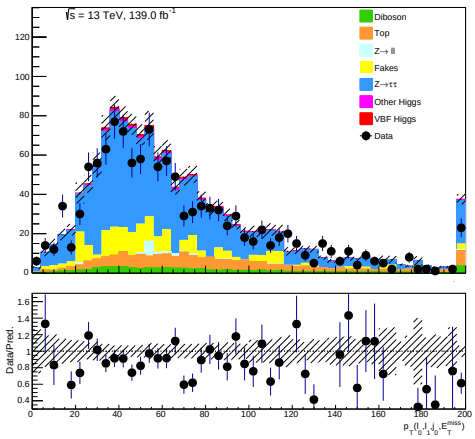
(n) m_{ll}^{vis}



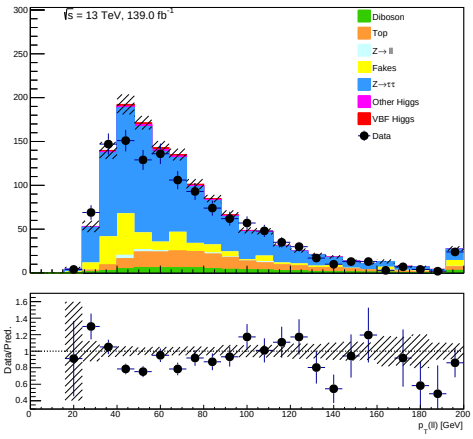
(o) m_{ll}^{MMC}



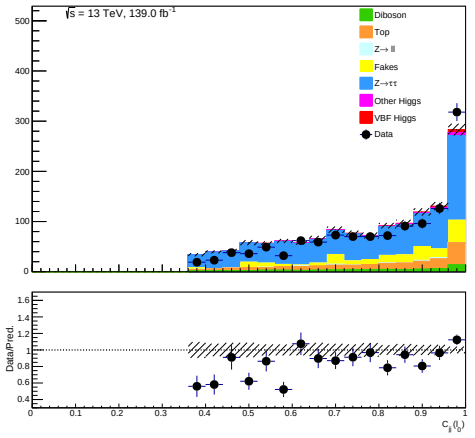
(p) $p_{\mathrm{T}}(l_0, l_1, j_0, j_1, E_{\mathrm{T}}^{miss})$



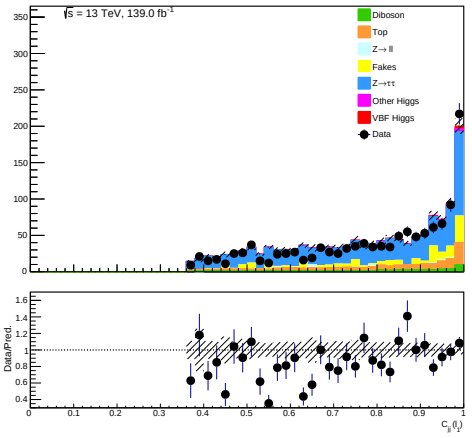
(q) $p_{\mathrm{T}}(l_0, l_1, j_1, E_{\mathrm{T}}^{miss})$



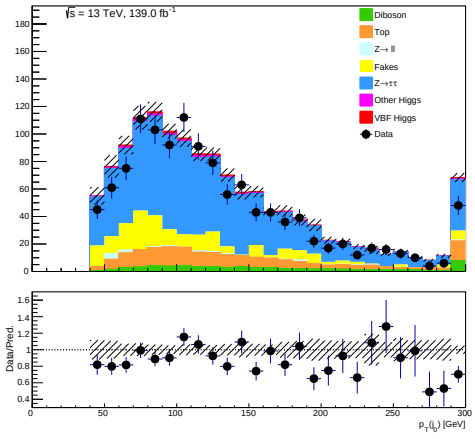
(r) $p_{\mathrm{T}}(ll)$



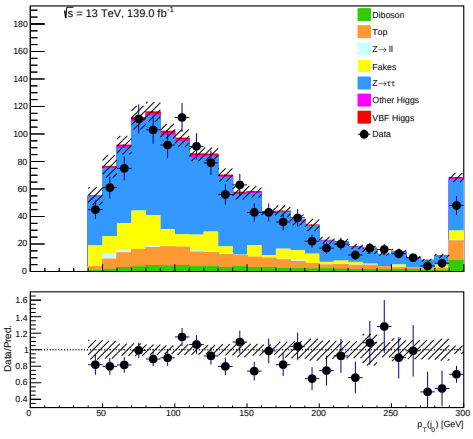
(s) $C_{jj}(l_0)$



(t) $C_{jj}(l_1)$



(u) $p_{\mathrm{T}}^{j_0}$



(v) $p_{\mathrm{T}}^{j_1}$

Figure 48: Expected and data distributions of the optimized input variable set in the low NN score region. Error bars represent statistical errors.

A.3 Optimized Binnings of CP-odd Observables for the ML Method

Variable	Bin Edges
$\mathcal{OO}$	$[-15., -11.4, -10.8, -9.9, -9.6, -8.7, -8.1, -7.8, -7.5, -7.2, -6.9 - 6.6, -6.3, -6.$ $-5.7, -5.4, -5.1, -4.8, -4.5, -4.2, -3.9, -3.6, -3.3, -3. - 2.7, -2.4, -2.1, -1.8, -1.5,$ $-1.2, -0.9, -0.6, -0.3, 0., 0.3, 0.6, 0.9, 1.2, 1.5, 1.8, 2.1, 2.4, 2.7, 3., 3.3,$ $3.6, 3.9, 4.2, 4.5, 4.8, 5.1, 5.4, 5.7, 6., 6.3, 6.9, 7.5, 8.1, 8.4, 9., 9.3, 9.6, 10.2$ $, 11.1, 12.3, 15.]$
$\mathcal{O}^{Reg}$	$[-2., -1.68, -1.6, -1.52, -1.4, -1.36, -1.32, -1.24, -1.2, -1.16, -1.12, -1.08,$ $-1.04, -1., -0.92, -0.88, -0.84, -0.8, -0.76, -0.72, -0.68, -0.64, -0.6, -0.56,$ $-0.52, -0.48, -0.44, -0.4, -0.36, -0.32, -0.28, -0.24, -0.2, -0.16, -0.12, -0.08,$ $-0.04, 0., 0.04, 0.08, 0.12, 0.16, 0.2, 0.24, 0.28, 0.32, 0.36, 0.4,$ $0.44, 0.48, 0.52, 0.56, 0.6, 0.64, 0.68, 0.72, 0.76, 0.8, 0.84, 0.88,$ $0.92, 0.96, 1., 1.04, 1.08, 1.12, 1.16, 1.2, 1.32, 1.4, 1.44, 1.48,$ $1.6, 1.64, 2.]$
$\Delta\phi_{jj}^{signed}$	$[-3.2, -3.008, -2.944, -2.88, -2.816, -2.752 - 2.688, -2.624, -2.56, -2.496,$ $-2.432, -2.368, -2.304, -2.24, -2.176, -2.112, -2.048, -1.984, -1.92, -1.856,$ $-1.792, -1.728, -1.664, -1.6, -1.536, -1.472, -1.408, -1.344, -1.28, -1.216,$ $-1.152, -1.088, -1.024, -0.96, -0.896, -0.832, -0.768, -0.64, -0.576, -0.512,$ $-0.448, -0.384, -0.32, -0.256, -0.192, -0.128, -0.064, 0., 0.064, 0.128,$ $0.192, 0.256, 0.32, 0.384, 0.448, 0.512, 0.576, 0.64, 0.704, 0.768,$ $0.832, 0.896, 0.96, 1.024, 1.088, 1.152, 1.216, 1.28, 1.344, 1.408,$ $1.472, 1.536, 1.6, 1.664, 1.728, 1.792, 1.856, 1.92, 1.984, 2.048,$ $2.112, 2.176, 2.24, 2.304, 2.368, 2.432, 2.496, 2.56, 2.624, 2.688,$ $2.752, 2.816, 2.88, 2.944, 3.008, 3.072, 3.2]$

Table 22: The bin edges of the histogram used for the NLL method, optimized such that for each variable, the bin contents fulfill a threshold.

References

[1] S. Agostinelli et al., *Geant4—a simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250, ISSN: 0168-9002, URL: https://www.sciencedirect.com/science/article/pii/S0168900203013688.

[2] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A Next-generation Hyperparameter Optimization Framework", *Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.

[3] R Alemany-Fernandez et al., *Operation and Configuration of the LHC in Run 1*, (2013), URL: https://cds.cern.ch/record/1631030.

[4] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, J. Instrum. **3** (2008) S08002, URL: https://doi.org/10.1088/1748-0221/3/08/s08002.

[5] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, J. High Energy Phys. **2010** (2010), ISSN: 1029-8479, URL: http://dx.doi.org/10.1007/JHEP06(2010)043.

[6] ATLAS Collaboration, *ATLAS inner detector: Technical Design Report, 1*, Technical design report. ATLAS, CERN, 1997, URL: https://cds.cern.ch/record/331063.

[7] ATLAS Collaboration, *ATLAS liquid-argon calorimeter: Technical Design Report*, Technical design report. ATLAS, CERN, 1996, URL: https://cds.cern.ch/record/331061.

[8] ATLAS Collaboration, *ATLAS muon spectrometer: Technical Design Report*, Technical design report. ATLAS, CERN, 1997, URL: http://cds.cern.ch/record/331068.

[9] ATLAS Collaboration, *ATLAS pixel detector electronics and sensors*, J. Instrum. **3** (2008) P07007, URL: https://doi.org/10.1088/1748-0221/3/07/p07007.

[10] 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, URL: https://cds.cern.ch/record/1277678.

[11] 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), ISSN: 1434-6052, URL: http://dx.doi.org/10.1140/epjc/s10052-019-7140-6.

[12] ATLAS Collaboration, *Estimation of non-prompt and fake lepton backgrounds in final states with top quarks produced in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, tech. rep., CERN, 2014, URL: https://cds.cern.ch/record/1951336.

[13] ATLAS Collaboration,
Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics, 2009, URL: https://arxiv.org/abs/0901.0512.

[14] ATLAS Collaboration,
Jet reconstruction and performance using particle flow with the ATLAS Detector, Eur. Phys. J. C **77** (2017),
URL: https://doi.org/10.1140%2Fepjc%2Fs10052-017-5031-2.

[15] ATLAS Collaboration, *LuminosityPublicResults*, 2012, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults.

[16] ATLAS Collaboration, *Measurement of $W^{\pm}$ and Z-boson production cross sections in pp collisions at $\sqrt{s}$=13 TeV with the ATLAS detector*, Phys. Lett. B **759** (2016) 601, ISSN: 0370-2693, URL: https://www.sciencedirect.com/science/article/pii/S0370269316302763.

[17] 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, J. High Energy Phys. **2022** (2022),
URL: https://doi.org/10.1007%2Fjhep08%282022%29175.

[18] ATLAS Collaboration,
Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s}$ = 7 and 8 TeV in the ATLAS experiment, Eur. Phys. J. C **76** (2016),
URL: https://doi.org/10.1140%2Fepjc%2Fs10052-015-3769-y.

[19] ATLAS 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,
URL: https://doi.org/10.1140/epjc/s10052-021-09233-2.

[20] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$ =13 TeV*, Eur. Phys. J. C **76** (2016), ISSN: 1434-6052, URL: http://dx.doi.org/10.1140/epjc/s10052-016-4120-y.

[21] ATLAS Collaboration, *Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector*, Phys. Lett. B **784** (2018) 173, ISSN: 0370-2693, URL: https://www.sciencedirect.com/science/article/pii/S0370269318305732.

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

[23] ATLAS Collaboration,
Performance of electron and photon triggers in ATLAS during LHC Run 2, Eur. Phys. J. C **80** (2020),
URL: https://doi.org/10.1140%2Fepjc%2Fs10052-019-7500-2.

[24] ATLAS Collaboration,
Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13$ TeV, Eur. Phys. J. C **78** (2018),
URL: https://doi.org/10.1140%2Fepjc%2Fs10052-018-6288-9.

[25] ATLAS Collaboration, *Public ATLAS Luminosity Results for Run-2 of the LHC*, 2022, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2.

[26] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, tech. rep., CERN, 2014, URL: https://cds.cern.ch/record/1700870.

[27] 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 s=13TeV with the ATLAS detector*, Phys. Lett. B **805** (2020) 135426, ISSN: 0370-2693, URL: http://dx.doi.org/10.1016/j.physletb.2020.135426.

[28] 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 $\sqrt{s}$=13TeV with the ATLAS detector*, Phys. Lett. B **805** (2020) 135426, ISSN: 0370-2693, URL: https://www.sciencedirect.com/science/article/pii/S0370269320302306.

[29] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, J. Instrum. **3** (2008) S08003, URL: https://doi.org/10.1088%2F1748-0221%2F3%2F08%2Fs08003.

[30] 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** (7 1992) 2405, URL: https://link.aps.org/doi/10.1103/PhysRevD.45.2405.

[31] R. D. Ball et al., *Parton distributions for the LHC run II*, J. High Energy Phys. **2015** (2015), ISSN: 1029-8479, URL: http://dx.doi.org/10.1007/JHEP04(2015)040.

[32] R. D. Ball et al., *Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO*, Nucl. Phys. B **855** (2012) 153, ISSN: 0550-3213, URL: http://www.sciencedirect.com/science/article/pii/S0550321311005463.

[33] C. M. Burnett, *File:Artificial neural network.svg*, 2006, URL: https://commons.wikimedia.org/wiki/File:Artificial_neural_network.svg.

[34] A. Butter, T. Plehn, N. Soybelman, and J. Brehmer, *Back to the Formula – LHC Edition*, 2021, URL: https://arxiv.org/abs/2109.10414.

[35] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, J. Phys. G **43** (2016) 023001, ISSN: 1361-6471, URL: http://dx.doi.org/10.1088/0954-3899/43/2/023001.

[36] W Buttinger, *The ATLAS Level-1 Trigger System*, tech. rep., CERN, 2012, URL: https://cds.cern.ch/record/1456546.

[37] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-kt jet clustering algorithm*, J. High Energy Phys. **2008** (2008) 063–063, ISSN: 1029-8479, URL: http://dx.doi.org/10.1088/1126-6708/2008/04/063.

[38] CDF Collaboration, *Observation of $B_s^0 - \overline{B}_s^0$ Oscillations*, Phys. Rev. Lett. **97** (24 2006) 242003, URL: https://link.aps.org/doi/10.1103/PhysRevLett.97.242003.

[39] 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** (4 1964) 138, URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.138.

[40] P. Clarke, *The L2 trigger for the ATLAS detector at the LHC*, NIM-A **368** (1995) 175, Proceedings of the Third International Workshop on B-Physics at Hadron Machines, ISSN: 0168-9002, URL: https://www.sciencedirect.com/science/article/pii/016890029500890X.

[41] 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), URL: https://doi.org/10.1103%2Fphysrevd.104.052004.

[42] CMS Collaboration, *Measurement of the single top quark and antiquark production cross sections in the t channel and their ratio in proton-proton collisions at $\sqrt{s}$=13TeV*, Phys. Lett. B **800** (2020) 135042, ISSN: 0370-2693, URL: https://www.sciencedirect.com/science/article/pii/S0370269319307646.

[43] CMS Collaboration, *Measurement of the top quark pair production cross section in dilepton final states containing one τ lepton in pp collisions at $\sqrt{s} = 13$ TeV*, J. High Energy Phys. (2020).

[44] 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, ISSN: 0370-2693, URL: https://www.sciencedirect.com/science/article/pii/S0370269312008581.

[45] CMS Collaboration, *The CMS experiment at the CERN LHC*, J. Instrum. **3** (2008) S08004, URL: https://doi.org/10.1088/1748-0221/3/08/s08004.

[46] E Coniavitis, *ATLAS silicon microstrip detector operation and performance*, J. Instrum. **6** (2011) C01054, URL: https://doi.org/10.1088/1748-0221/6/01/c01054.

[47] G. Cowan, *Statistical Data Analysis*, Oxford science publications, Clarendon Press, 1998, ISBN: 9780198501558, URL: https://books.google.de/books?id=ff8ZyW0nlJAC.

[48] 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) 161–171, ISSN: 0010-4655, URL: http://dx.doi.org/10.1016/j.cpc.2015.04.021.

[49] A. Einstein, *Die Grundlage der allgemeinen Relativitätstheorie*, Ann. Phys. (Berl.) **354** (1916) 769, eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.19163540702, URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.19163540702.

[50] A. Elagin, P. Murat, A. Pranko, and A. Safonov,
A new mass reconstruction technique for resonances decaying to,
Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **654** (2011) 481–489,
ISSN: 0168-9002, URL: http://dx.doi.org/10.1016/j.nima.2011.07.009.

[51] J. Ellis, M. K. Gaillard, and D. V. Nanopoulos,
"A Historical Profile of the Higgs Boson",
The Standard Theory of Particle Physics, chap. Chapter 14 255, URL:
https://www.worldscientific.com/doi/abs/10.1142/9789814733519_0014.

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

[53] L. Evans and P. Bryant, *LHC Machine*, J. Instrum. **3** (2008) S08001.

[54] D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, 2016, arXiv: 1610.07922 [hep-ph].

[55] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*,
J. High Energy Phys. **2007** (2007) 070–070, ISSN: 1029-8479,
URL: http://dx.doi.org/10.1088/1126-6708/2007/11/070.

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

[57] P. D. Group, *Review of Particle Physics*, Phys. Rev. D **98** (3 2018) 030001,
URL: https://link.aps.org/doi/10.1103/PhysRevD.98.030001.

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

[59] K. Hagiwara, S. Ishihara, R. Szalapski, and D. Zeppenfeld,
Low energy effects of new interactions in the electroweak boson sector,
Phys. Rev. D **48** (5 1993) 2182,
URL: https://link.aps.org/doi/10.1103/PhysRevD.48.2182.

[60] K. Hamilton, P. Nason, and G. Zanderighi, *MINLO: multi-scale improved NLO*,
J. High Energy Phys. **2012** (2012), ISSN: 1029-8479,
URL: http://dx.doi.org/10.1007/JHEP10(2012)155.

[61] V. Hankele, G. Klämke, D. Zeppenfeld, and T. Figy,
Anomalous Higgs boson couplings in vector boson fusion at the CERN LHC,
Phys. Rev. D **74** (9 2006) 095001,
URL: https://link.aps.org/doi/10.1103/PhysRevD.74.095001.

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

[63] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*,
Phys. Rev. **145** (4 1966) 1156,
URL: https://link.aps.org/doi/10.1103/PhysRev.145.1156.

[64] S. Höche, F. Krauss, M. Schönherr, and F. Siegert,
QCD matrix elements + parton showers. The NLO case,
J. High Energy Phys. **2013** (2013), ISSN: 1029-8479,
URL: http://dx.doi.org/10.1007/JHEP04(2013)027.

[65] L. L. Iglesias, *Diboson production at the LHC*, PoS **LHCP2016** (2016) 022,
URL: https://cds.cern.ch/record/2276624.

[66] J. Kiefer and J. Wolfowitz,
Stochastic Estimation of the Maximum of a Regression Function,
Ann. Math. Stat. **23** (1952) 462, ISSN: 00034851,
URL: http://www.jstor.org/stable/2236690.

[67] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization.",
ICLR (Poster), ed. by Y. Bengio and Y. LeCun, 2015,
URL: http://dblp.uni-trier.de/db/conf/iclr/iclr2015.html#KingmaB14.

[68] A. La Rosa, *The ATLAS Insertable B-Layer: from construction to operation*,
(2016), URL: https://arxiv.org/abs/1610.01994.

[69] LHCb Collaboration, *Observation of D^0–$\overline{D}^0$ Oscillations*,
Phys. Rev. Lett. **110** (10 2013) 101802,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.110.101802.

[70] LHCb Collaboration, *The LHCb Detector at the LHC*,
J. Instrum. **3** (2008) S08005,
URL: https://doi.org/10.1088/1748-0221/3/08/s08005.

[71] C. Limbach, *Reconstruction and Identification of Tau Leptons in ATLAS*,
Nucl. Part. Phys. Proc. **260** (2015) 195, The 13th International Workshop on Tau Lepton Physics, ISSN: 2405-6014, URL:
https://www.sciencedirect.com/science/article/pii/S2405601415000991.

[72] G. Luisoni, P. Nason, C. Oleari, and F. Tramontano,
$HW \pm/HZ$ + 0 and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO, J. High Energy Phys. **2013** (2013),
ISSN: 1029-8479, URL: http://dx.doi.org/10.1007/JHEP10(2013)083.

[73] Z. Marshall and the ATLAS Collaboration,
Simulation of Pile-up in the ATLAS Experiment,
J. Phys. Conf. Ser. **513** (2014) 022024,
URL: https://dx.doi.org/10.1088/1742-6596/513/2/022024.

[74] K. Mochizuki,
b-jet identification algorithms and performance in the ATLAS experiment,
Nucl. Part. Phys. Proc. **273-275** (2016) 2536, 37th International Conference on High Energy Physics (ICHEP), ISSN: 2405-6014, URL:
https://www.sciencedirect.com/science/article/pii/S2405601415009396.

[75] P. Nason,
A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms,
J. High Energy Phys. **2004** (2004) 040–040, ISSN: 1029-8479,
URL: http://dx.doi.org/10.1088/1126-6708/2004/11/040.

[76] P. Nason and C. Oleari, *NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG*, J. High Energy Phys. **2010** (2010),
ISSN: 1029-8479, URL: http://dx.doi.org/10.1007/JHEP02(2010)037.

[77] K. Olive, *Review of Particle Physics*, Chin. Phys. C **38** (2014) 090001, URL: https://doi.org/10.1088/1674-1137/38/9/090001.

[78] Particle Data Group, *2018 Particle Physics Booklet*, 2018.

[79] M. Rauch, *Vector-Boson Fusion and Vector-Boson Scattering*, 2016, arXiv: 1610.08420 [hep-ph].

[80] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, Eur. Phys. J. C **71** (2011), ISSN: 1434-6052, URL: http://dx.doi.org/10.1140/epjc/s10052-011-1547-z.

[81] T. Rias, *File:HiggsBR.svg*, URL: https://commons.wikimedia.org/wiki/File:HiggsBR.svg.

[82] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Learning representations by back-propagating errors*, Nature **323** (1986) 533, URL: https://doi.org/10.1038/323533a0.

[83] A. D. Sakharov, *Violation of CP invariance,C asymmetry, and baryon asymmetry of the universe*, Phys.-Uspekhi **34** (1991) 392, URL: https://doi.org/10.1070/pu1991v034n05abeh002497.

[84] A. Salam, "Weak and Electromagnetic Interactions", *Elementary particle theory, Relativistic groups and analyticity*, Proceedings of the Eighth Nobel Symposium (Aspenäsgarden, Lerum, May 19–25, 1968), ed. by N. Svartholm, Almquist & Wiksell, 1968 367.

[85] J. Schmidhuber, *Deep learning in neural networks: An overview*, Neural Netw **61** (2015) 85, URL: https://doi.org/10.1016%2Fj.neunet.2014.09.003.

[86] M. Schott and M. Dunford, *Review of single vector boson production in pp collisions at* $\sqrt{s} = 7$ *TeV*, Eur. Phys. J. C **74** (2014), ISSN: 1434-6052, URL: http://dx.doi.org/10.1140/epjc/s10052-014-2916-1.

[87] A. M. Sirunyan et al., *Measurement of the production cross section for single top quarks in association with W bosons in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, J. High Energy Phys. **2018** (2018), ISSN: 1029-8479, URL: http://dx.doi.org/10.1007/JHEP10(2018)117.

[88] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159–177, ISSN: 0010-4655, URL: http://dx.doi.org/10.1016/j.cpc.2015.01.024.

[89] A Vogel, *ATLAS Transition Radiation Tracker (TRT): Straw Tube Gaseous Detectors at High Rates*, tech. rep., CERN, 2013, URL: https://cds.cern.ch/record/1537991.

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

[91] 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** (4 1957) 1413, URL: https://link.aps.org/doi/10.1103/PhysRev.105.1413.

Acknowledgments

I would like the thank Prof. Dr. Markus Schumacher for allowing me the opportunity to be able to work on this interesting project, as well as for his supervision and support during the entire process.

I would also like to thank Oğul Öncel for supervising us for the entire process as well. He has given his utmost effort to provide us with the supervision that we need despite his busy schedule as a Post-doc, for which I am very thankful and grateful.

I also thank Dr. Valerie Lang, Helena Moyano Gomez and Daniel Bahner for the help I have received during the process of this master thesis.

I would lastly like to thank the members of the group for the fun time I had in this AG.