GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737

Fakultät für Physik

Bachelor's Thesis

Untersuchungen zur gleichzeitigen Messung der Top-Quark-Paarproduktion mit einem off-shell Z-/W-Boson in trileptonischen Endzuständen mit ATLAS

Studies of simultaneous measurements of top quark pair production associated with an off-shell Z-/W-boson in the trileptonic channel with ATLAS

prepared by

Ireas Tom Raschke

from Finsterwalde

at the II. Physikalisches Institut

Thesis number: II.Physik-UniGö-BSc-2022/02

Thesis period: 28th March 2022 until 3rd July 2022

First referee: Prof. Dr. Arnulf Quadt

Second referee: Prof. Dr. Ariane Frey

CERN-THESIS-2022-252 //2022

Zusammenfassung

Diese Bachelorarbeit studiert die gleichzeitige Messung von Top-Quark-Paar Ereignissen mit einem Z/W-Boson in Regionen außerhalb der Massenschale. Diese Prozesse sind schwer zu unterscheiden, da deren messbare Signale ähnlich sind. Es wird ausschließlich der trileptonische Zerfallskanal untersucht, da dieser klarere Signaturen als hadronische oder einzel-leptonische Zerfälle bietet. Für die Klassifizierung werden zwei verschiedene Ansätze verwendet. Der erste Ansatz basiert auf Selektionen, die aus den Verteilungen der verschiedenen Prozesse abgeleitet werden. Für den anderen Ansatz wird ein neuronales Netzwerk implementiert und trainiert, sodass dieses abstrakte Klassenwerte definiert, welche zur Klassifizierung der Prozesse verwendet werden. Zusätzlich wird die EFT Sensitivität analysiert, denn es wird erwartet, dass für die Regionen außerhalb der Massenschale die SM-EFT Beiträge für höhere Energieskalen steigen.

Stichpunkte: Teilchenphysik, Bachelorarbeit, ATLAS, Massenschale, Z-/W-Boson, Neuronale Netzwerke, SM-EFT

Abstract

This Bachelor's thesis studies the simultaneous measurements of top quark pair events with a Z/W-boson in off-shell regions. These processes are difficult to distinguish because their measured signals are similar. The trileptonic channel is exclusively analysed since it provides a cleaner event signature in comparison to hadronic or single leptonic channels. For the classification, two separate approaches are used. One is based on cuts to several variables which are derived from event distributions. For the other approach, a neural network is implemented and trained to define abstract class scores which are used for classification. Additionally, EFT sensitivity is analysed in these off-shell regions because SM-EFT contributions are expected to increase for higher energy scales.

Keywords: Particle physics, Bachelor's thesis, ATLAS, off-shell, Z-/W-boson, neural networks, SM-EFT

Contents

1 Introduction **1**

2 The Standard Model of Particle Physics **3**

2.1 Elementary Particles 4

2.2 The Top Quark 4

2.3 $t\bar{t}$ Production in Association with a Z/W-Boson 5

2.4 Effective Field Theory 6

3 Experimental Setup **8**

4 Region Definition **11**

4.1 Preselection 11

4.2 Cut And Count Based Region Definition 11

4.3 Neural Network Based Region Definition 12

5 Neural Networks and Fitting Algorithm **15**

5.1 Deep Neural Networks 15

5.2 Neural Network Architecture 16

5.3 Fitting Algorithm 19

6 Experimental and Theoretical Systematic Uncertainties **21**

6.1 Experimental Uncertainties 21

6.2 Theoretical Uncertainties 22

7 Measurements of the Inclusive Cross-Section **24**

8 EFT Sensitivity **29**

8.1 Cross-Section Ratio 29

8.2 Separation Power 30

9 Conclusion **32**

1 Introduction

As early as 400 BC, humankind was philosophising about the fundamental structure of nature. The Greek philosopher Democritus is said to have formulated the idea, that matter can not be infinitely divisible. Thus, there has to be a smallest, indivisible part which would be called atom (from Greek *atomos*='indivisible'). This idea was not experimentally tested but built purely on philosophical thoughts, however, it represents the beginning of the search for a fundamental description of the universe.

Over 2000 years later, in 1808, the chemist John Dalton refined the idea of Democritus by introducing the concept of elements. Each element represents a unique, indivisible atom which differ by their mass and size. This refined model allowed Dalton to explain the apparent conservation of mass.

In the late 19th century, the physicist Sir Joseph John Thomson developed an atomic model. This model describes atoms as positive charged spheres which contain lighter negative charged particles, so called electrons. The complete atom is electrically neutral. The atom itself was no longer the smallest, most fundamental particle any more. The model was based on his experiments using a hot, radiating cathode that emits electrons. Hence, the Thomson-model explained the observed cathode electron beam.

In 1911, Ernest Rutherford conducted an experiment using an α-particle beam that was focused onto a thin gold foil. The results of the experiments suggest that most of the mass in an atom is contained in a very small volume which is the positively charged nucleus. It is surrounded by negatively charged electrons that orbit and empty space that makes up most of the atom.

Two years later, in 1913, Niels Bohr refined the model. The refinement describes the motion of electrons around the nucleus like planets around the sun. He postulated that electrons move on circular orbits without emitting radiation. Additionally, the radii of those orbits can only be certain quantised values. This marks the beginning of quantum mechanics.

Electrons are able to change their orbits by emitting or absorbing photons of quantised energies. Therefore, the model made it possible to explain the photoelectric effect. In 1916, the Bohr-model was expanded by Arnold Sommerfeld who generalised the electron

orbits and allowed an elliptical shape. Thus, the Bohr-Sommerfeld-model explained, why there are multiple spectral lines very close to each other.

In 1926, Werner Heisenberg and Erwin Schrödinger independently developed a mathematical description of quantum mechanics. Requiring that electrons fulfil the quantum mechanical Schrödinger equation led to the atomic orbital model. The orbits are space regions with high probability of containing electrons which obey specific selection rules.

While the concept of an atomic nucleus was first formulated by Rutherford, it was first thought to only be made up of positively charged protons. However, this model was expanded in 1932 by James Chadwick, who introduced neutral particles, neutrons, that are also part of the nucleus.

Nowadays, the research of the fundamental structure of nature goes on and particle physics has superseded nuclear physics. Modern particle physics' task is to find a general description of fundamental phenomena. It focuses on formulating and expanding an underlying theory which is known as the Standard Model of particle physics.

This Bachelor's thesis will cover the underlying basis of the Standard Model in Chapter 2. Following the theoretical basis, Chapter 3 covers the experimental setup of the ATLAS-experiment. Two approaches for region definition are explained in Chapter 4 and an introduction into neural networks as well as the used algorithms is given in Chapter 5. Chapter 6 goes into more detail about the implemented uncertainties. The results of the fitting process and the studies of possible EFT sensitivities are given in Chapter 7 and 8, respectively. Chapter 9 will conclude this thesis with an outlook for potential improvements.

2 The Standard Model of Particle Physics

The Standard Model of particle physics (SM) is the most successful description of the fundamental particles and their interactions to this day. It covers all known elementary particles and their antiparticles as well as three of the four known fundamental forces: the strong, weak and electromagnetic force. Gravity is not described by the SM.

The SM is a renormalisable quantum field theory which is defined by an internal SU(3)⊗SU(2)⊗U(1) gauge symmetry. The SU(3) group symmetry describes the strong interactions, originating from quantum chromodynamics (QCD) [1], and the SU(2)⊗U(1) group symmetry corresponds to the electroweak interactions. The later being the unification of the electromagnetic force, originating from quantum electrodynamics (QED) [2–4], and the weak force, originating from quantum flavour dynamics (QFD) [5].

In 1967, the Brout-Englert-Higgs mechanism was incorporated into the electroweak interaction [6–8]. The mechanism adds a quantum Higgs-field that induces a spontaneous symmetry breaking in the SU(2)⊗U(1) symmetry. Initially all bosons were massless, but this symmetry breaking causes all bosons interacting with the Higgs-field to become massive. Fermions also obtain their masses by interacting with the Higgs-field. The massive gauge bosons of the electroweak interaction are known as the Z/W-boson. Additionally, one of the degrees of freedom introduced by the Higgs-field is not needed by the Brout-Englert-Higgs mechanism, thus becoming the scalar Higgs-boson H [6].

Even though the SM can predict most observations with remarkable precision, there are phenomena which cannot be explained using the SM. An example for this is neutrino oscillations [9], which describes flavour changing neutrinos. This oscillation requires at least two massive neutrinos, however, the SM predicts all neutrinos to be massless. Hence, the SM cannot explain the neutrino oscillation. Another example is dark matter [10]. The measured rotational velocity for large spiral galaxies requires a much larger mass than the visible matter provides. These astronomical observation provide evidence of a stable form of matter that does not interact electromagnetically, making it invisible to direct optical observations. No particle in the SM can explain these measurements.

2.1 Elementary Particles

The elementary particles are divided into fermions and bosons. Fermions have half-integer spin and make up matter, while bosons have integer spin and are responsible for mediating the strong, weak and electromagnetic force. All electrically charged particles also have an antiparticle partner, which has the same mass but opposite electric charge.

Figure 2.1 shows the different SM-particles. Furthermore, it shows that fermions are subdivided into quarks and leptons, which both are separated into three generations. Particles corresponding to a higher generation have significantly increased masses.

The six quarks are split into three up-type (up, charm, top) and three down-type (down, strange, bottom) quarks. The up-type quarks have an electric charge of $+\frac{2}{3}\,e$ and down-type quarks of $-\frac{1}{3}\,e$. All quarks also have a colour charge, thus participating in strong interactions. Fermions also have a weak isospin T of which the third component T_3 is conserved in weak interactions [4].

Left-handed fermions, and quarks in particular, have a weak isospin of $T = \frac{1}{2}$ and can be grouped into doublets with $T_3 = \pm\frac{1}{2}$. The positive sign corresponds to up-type quarks and the negative sign to down-type quarks, respectively. Right-handed fermions form singlets and the electroweak W-bosons do not interact with them.

Leptons on the other hand are split into electrically charged leptons (e, μ, τ) and their corresponding neutrinos $(\nu_e, \nu_\mu, \nu_\tau)$. The charged leptons all have an electric charge of $-1\,e$, while neutrinos are electrically neutral. Left-handed leptons can also be grouped into doublets with $T_3 = \pm\frac{1}{2}$, where neutrinos are the up-type and charged leptons the down-type particles.

2.2 The Top Quark

The top (t) quark was discovered in 1995 at Fermilab by the experiments CDF [11] and DØ [12]. It is assumed to be the partner of the bottom (b) quark with a mass of $m_t = 172.76 \pm 0.30\,\mathrm{GeV}$ [13], which makes it the heaviest of all known quarks and has a very short lifetime of $\tau_t \approx 5 \cdot 10^{-25}\,\mathrm{s}$ [13]. The top quark decays via the weak interaction into a W-boson and a down-type quark before it hadronises. The probability of decaying into a particular down-type quark is given by the respective CKM-matrix element [14]. Due to the CKM-matrix having vanishing off-diagonal contributions, top quarks predominantly decay into bottom quarks. The W-boson further decays either into a charged lepton and neutrino, which results in one jet from the bottom quark and one measurable lepton, or into a quark-antiquark-pair which results in three produced jets.

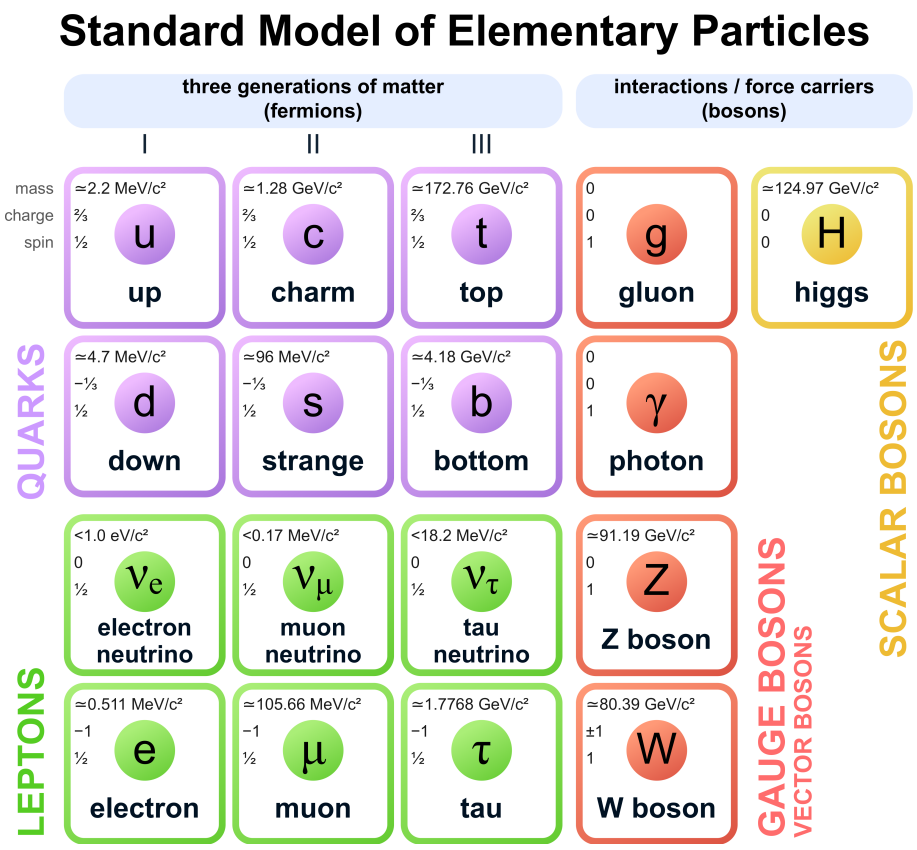

Figure 2.1: The elementary particles of the Standard Model subdivided into fermions (left) and bosons (right). The fermions are further split into quarks (top) and leptons (bottom), which are each divided into three generations. The bosons are divided into four vector bosons and one scalar boson.

Producing top quarks requires high energies due to the mass of the top quark. These energies can be achieved in hadrons colliders. Figure 2.2 shows two example processes for top quark production in $t\bar{t}$-processes. Each of the two top quarks decays as described before, therefore the measured signal depends on the W-boson decay mode.

2.3 $t\bar{t}$ Production in Association with a Z/W-Boson

When producing $t\bar{t}$ pairs, an additional Z/W-boson can be produced. Figure 2.2 shows example processes for $t\bar{t}$-production via gluon-gluon fusion and quark-antiquark annihilation with an additional Z/W-boson. Since the LHC is a proton-proton collider, the gluon-gluon fusion process is dominating. A selection of possible signals from $t\bar{t}Z/t\bar{t}W$ events is summarised in Table 2.1.

The $t\bar{t}Z$ events are of interest as they allow studying the coupling of top quarks to neutral currents of the weak interaction through final state radiation of a Z-boson in $t\bar{t}$

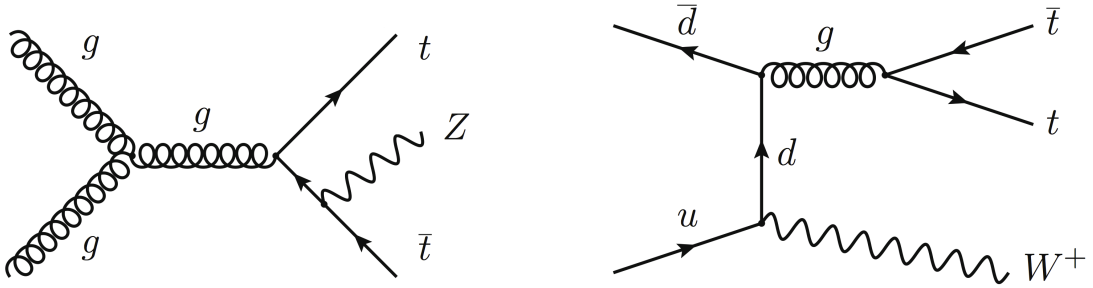

Figure 2.2: The Feynman diagrams for an example $t\bar{t}Z$ process by gluon-gluon fusion (left) in which the produced $t\bar{t}$ pair emits an additional Z-boson and a $t\bar{t}W$ process by quark-antiquark annihilation (right) emitting the W-boson as initial state radiation.

events. Therefore, $t\bar{t}Z$ events allow analysing the third component of the weak isospin of top quarks.

The $t\bar{t}W$ events are not sensitive to the coupling of top quarks to charged currents of the weak interaction, because, as seen in Fig 2.2, the W-bosons are only emitted as initial state radiation. However, the $t\bar{t}W$ processes are sensitive to the initial parton distribution functions (PDF) and thus measurements on these processes probe those PDFs.

Analysing $t\bar{t}Z/t\bar{t}W$ events together at the same time is particularly interesting because both can generate similar signals and thus are important background processes for each other. Especially studying these events in off-shell regions gives rise to a high rate of background from the respective other process. Some kinematic variables in $t\bar{t}Z/t\bar{t}W$ off-shell events like invariant mass and transverse momentum are also expected to be sensitive to SM-EFT contributions. Furthermore, the sensitivity of these SM-EFT contributions is expected to increase for higher boson masses [15].

2.4 Effective Field Theory

For the analysis of effects beyond the Standard Model, one can use an Effective Field Theory (EFT) approach such as the Standard-Model Effective Field Theory (SM-EFT) [16]. This theory assumes that the current SM is an effective low-energy approximation which holds for interactions up to an energy scale Λ and can be expanded using higher dimensional interactions [17]. The transition into the Standard Model happens via decoupling of heavy particles at energies higher than Λ. Therefore, higher-dimensional operators Q, which are suppressed by powers of Λ, are used in a perturbation expansion [17]

$$\mathcal{L}_{\text{SM-EFT}} = \mathcal{L}_{\text{SM}} + \frac{1}{\Lambda}\sum_k C_k^{(5)} Q_k^{(5)} + \frac{1}{\Lambda^2}\sum_k C_k^{(6)} Q_k^{(6)} + \mathcal{O}\left(\frac{1}{\Lambda^3}\right) . \tag{2.1}$$

Table 2.1: Possible leptonic signals from $t\bar{t}Z/t\bar{t}W$ events which provides a cleaner measurement in comparison to the hadronic decays. At least one of the top quarks or the Z/W-boson decays into leptons. The number in brackets shows the number of not measurable neutrinos.

Signal	Reaction	Jets	Leptons
$t\bar{t}W$-Monoleptonic	$t\bar{t}W \to (bq\bar{q}')\,(\bar{b}q\bar{q}')\,(\ell\bar{\nu}_\ell)$	6	1(+1)
$t\bar{t}W$-Dileptonic	$t\bar{t}W \to (bq\bar{q}')\,(\bar{b}\ell\bar{\nu}_\ell)\,(\ell\bar{\nu}_\ell)$	4	2(+2)
$t\bar{t}W$-Trileptonic	$t\bar{t}W \to (b\bar{\ell}\nu_\ell)\,(\bar{b}\ell\bar{\nu}_\ell)\,(\ell\bar{\nu}_\ell)$	2	3(+3)
$t\bar{t}Z$-Dileptonic	$t\bar{t}Z \to (bq\bar{q}')\,(\bar{b}q\bar{q}')\,(\ell\bar{\ell})$	6	2(+0)
$t\bar{t}Z$-Trileptonic	$t\bar{t}Z \to (bq\bar{q}')\,(\bar{b}\ell\bar{\nu}_\ell)\,(\ell\bar{\ell})$	4	3(+1)
$t\bar{t}Z$-Tetraleptonic	$t\bar{t}Z \to (b\bar{\ell}\nu_\ell)\,(\bar{b}\ell\bar{\nu}_\ell)\,(\ell\bar{\ell})$	2	4(+2)

In this perturbation expansion $\mathcal{L}_{\mathrm{SM}}$ describes the common SM-Lagrangian, $Q_k^{(n)}$ the n-dimensional interaction operators and $C_k^{(n)}$ their corresponding Wilson coefficients.

For a better understanding of $t\bar{t}Z$ and $t\bar{t}W$ production, their relevant operators can be analysed. Since the Z-boson is a combination of the W^0- and B-bosons originating the Brout-Englert-Higgs mechanism, the operator Q_{tZ} for the tZ-coupling is given by [6]

$$Q_{\mathrm{tZ}} = \cos(\Theta_W) Q_{\mathrm{tW}} - \sin(\Theta_W) Q_{\mathrm{tB}} \,. \tag{2.2}$$

Hence, the operators for the tW- and the tB-coupling and their complex coefficients C_{tW} and C_{tB} are of particular interest for this analysis [16]. To measure the sensitivity, the separation power

$$S = \frac{1}{2} \sum_{\mathrm{Bins}} \frac{(\mathrm{EFT}_i - \mathrm{SM}_i)^2}{\mathrm{EFT}_i + \mathrm{SM}_i} \tag{2.3}$$

is used. The fraction inside the sum is calculated for each bin for a given normalised distribution. SM stands for the Standard Model prediction and EFT includes a specific variation on one or multiple Wilson coefficients as in Equation 2.1. Studies regarding the sensitivity of several variables for these coefficients are summarised in Chapter 8.

3 Experimental Setup

For collecting data of $t\bar{t}Z$ and $t\bar{t}W$ events, a high energy particle collider is needed. Furthermore, a setup for the detection of signals as well as appropriate reconstruction algorithms are necessary.

The 27 km long Large Hadron Collider (LHC) is a proton-proton collider. The centre-of-mass energy reached by the LHC was around 7-8 TeV during Run I and 13 TeV during Run II [18]. During Run III the centre-of-mass energy is expected to reach 13.6 TeV [19]. Those high-energy collisions produce multiple particles like top quarks, which in turn decay into more particles. Detecting events requires a calibrated detector system which measures the properties of the particles such as charge, momentum and energy. Furthermore, the event reconstruction from measured signals uses several algorithms to reconstruct for example the particles trajectories and collision vertices.

The ATLAS-detector [20] is a multi-purpose particle detector that consists of different layers for detecting various particles. These layers are built in concentric cylinders around the beam axis. Measurements of the tracks and deposited energies of the decay products allow the reconstruction of the initial particles. Figure 3.1 shows a schematic view of those detector layers. Each layer is described in the following.

The Inner Detector

The task of the Inner Detector is to precisely measure the tracks of various charged particles. Because of the 2 T magnetic field, which is induced by a central solenoid, the moving charged particles are deflected via the Lorentz force. Measurements of the curvature allows precise measurements of their momentum and reconstructing the particle track. Thus, the tracking system is crucial for b-tagging because it makes it possible to identify secondary vertices, which are indications for b-jets [20]. Because of the high density of particles in the detector close to the collision point, high-precision measurements use Pixel-, Strip- and Transition Radiation Trackers.

The Pixel Tracker is the first point of detection and allows tracking near the collision point using small Pixel-modules [20]. It is built from silicon to withstand strong radiation

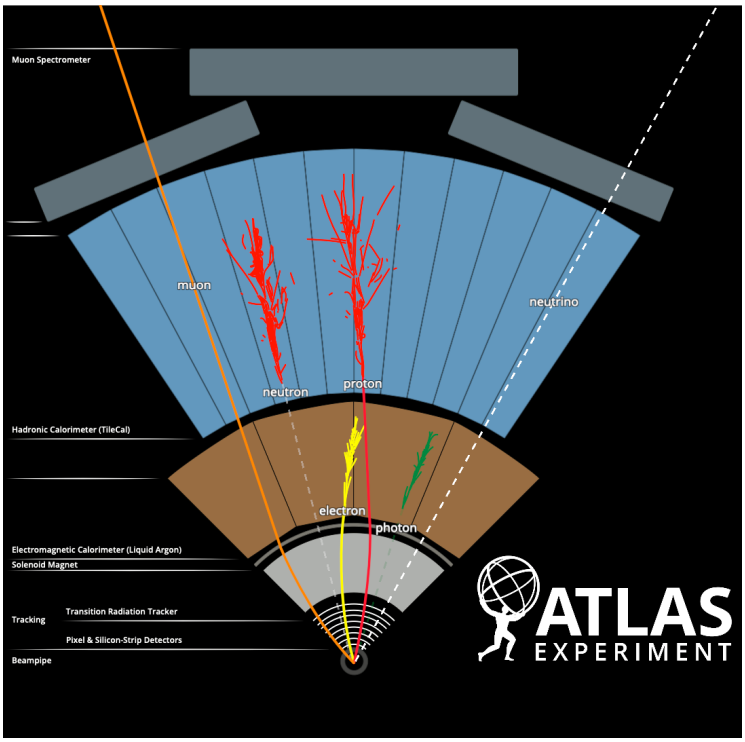

Figure 3.1: Cross-section of the concentric layers inside the ATLAS Detector (© CERN).

from the collision. The semiconductor Strip Tracker (SCT) consists of longer strips that also contribute to the tracking by offering larger scale measurement.

The Transition Radiation Tracker (TRT) is built from thin and long polyimide straws [20]. These straw tubes are filled with a gas mixture and while they offer a worse resolution, they allow for a relatively cheap way to cover a greater volume and hence, contribute to a precise measurement. It is used to distinguish electrons $e^{\pm}$ and charged pions $\pi^{\pm}$.

The Calorimeters

As seen in Figure 3.1, the calorimeters are built around the Inner Detector and the solenoid magnet. In general, the calorimeters measure the energy of particles by stopping them. The calorimeter system used in the ATLAS-detector is subdivided into the electromagnetic and the hadronic calorimeter. Both are sampling calorimeters that have an absorber material, which induces particle showers and an active medium, which is used for measuring the signals [21].

The electromagnetic calorimeter measures all particles that interact electromagnetically. Those charged particles and photons each create an electromagnetic shower which can be used to reconstruct the initial particles. It offers a precise measurement of the deposited

energy and the location of the energy-deposition. It is built from Lead as the absorber and liquid Argon as active medium [21].

The hadronic calorimeter absorbs particles that pass the electromagnetic calorimeter and interact via the strong force. The hadronic showers created can produce additional electromagnetic sub-showers which makes them more difficult to reconstruct and less precise. It consist of steel as an absorber and plastic scintillating tiles [21].

The Muon Spectrometer

The Muon Spectrometer is the outermost layer and follows a similar approach as the Inner Detector to measure the momentum of muons which usually pass the Inner Detector and Calorimeter undetected. It consists of four different subsystems [20].

For the muon triggering, the Resistive Plate Chambers (RPC) and Thin Gap Chambers (TGC) are used in different regions of the detector. Furthermore, they are used in the coordinate measurement. The Cathode Strip Chambers (CSC) offer a more precise coordinate measurements at the ends of the detector. Lastly, the Monitored Drift Tubes (MDT) track the curved trajectories of the muons. The magnetic field responsible for deflecting the muons is generated by outer toroidal magnets [20].

4 Region Definition

This investigation of simultaneous measurements utilises the trileptonic decay channel because of its cleaner event signature in comparison to hadronic or single leptonic decays. Sensitive regions for $t\bar{t}Z$ and $t\bar{t}W$ are defined to be used in fitting process. Since the trileptonic channel is used, non-prompt electrons (fakes) are also a dominant background. Thus, ideally a signal region for $t\bar{t}Z$ and two separate control regions for $t\bar{t}W$ and fakes are defined.

4.1 Preselection

The analysis only includes events with exactly three electrons or muons, three or more jets of which at least one needs to be b-tagged and an off-shell Z-boson in the final state. A Z-boson is labelled as off-shell if the reconstructed mass differs from the SM prediction by at least 10 GeV. Because of the low statistics in the off-shell region, the 85% working point for b-tagging is chosen. Additionally, these events need to have a lepton pair of opposite sign and same flavour (OSSF) which is expected to originate from the Z-boson. This OSSF-pair is required to have a reconstructed mass of at least 10 GeV to suppress background contributions from low-mass resonances. All three leptons have to pass the tight selection and need to satisfy $p_{\mathrm{T},1} \geq 27\,\mathrm{GeV}$, $p_{\mathrm{T},2} \geq 20\,\mathrm{GeV}$ and $p_{\mathrm{T},3} \geq 15\,\mathrm{GeV}$ respectively. These preselections are applied to all events before region definition. For the binning of variables, the same algorithm as described in Section 5.3.1 of Reference [22] is used and was tested for several numbers of bins of which 6 bins were found to be optimal.

4.2 Cut And Count Based Region Definition

To define sensitive regions, several cuts on kinematic variables are analysed. The choice of variables is motivated by the underlying physical processes as described in Section 2.3.

The best performance was achieved by using the missing energy $E_{\mathrm{T}}^{\mathrm{Miss}}$ and the ΔR

Table 4.1: The selections for the (top) cut and count based and (bottom) neural network based region definition.

Region	Selection
SR-ttZ-CC	$\Delta R_{\text{OSSF}} < 1.8$
CR-ttW-CC	$\Delta R_{\text{OSSF}} \geq 1.8$; $E_{\text{T}}^{\text{Miss}} \geq 78\,\text{GeV}$
CR-Fakes-CC	$\Delta R_{\text{OSSF}} \geq 1.8$; $E_{\text{T}}^{\text{Miss}} < 78\,\text{GeV}$
SR-ttZ-NN	ttW score<0.4, Fakes score<0.4
CR-ttW-NN	ttW score≥ 0.4
CR-Fakes-NN	ttW score<0.4, Fakes score≥ 0.4

variable which is defined as

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} \tag{4.1}$$

between two objects. ΔR_{OSSF} between the OSSF-pair is of particular interest. The reason for that is that the leptons from the Z-boson decay in $t\bar{t}Z$ events are expected to be more bundled and hence closer in ΔR compared to other lepton combinations. Also, $t\bar{t}W$ should have higher values for $E_{\text{T}}^{\text{Miss}}$ than $t\bar{t}Z$, since $t\bar{t}W$ processes have three neutrinos in their final state instead of only one for $t\bar{t}Z$ events. Furthermore, these two variables offer fine region tuning, since they are continuously distributed unlike for example jet multiplicities.

The cuts are set by evaluating the separation plots in Figure 4.1 and choosing the initial cut-off at the values for which the $t\bar{t}W$ background becomes larger than the $t\bar{t}Z$ signal. Several configurations were tested by varying those cuts. A lower cut value for $E_{\text{T}}^{\text{Miss}}$ and a higher value for ΔR_{OSSF} were found to have the best performance. These can also be seen in Figure 4.1. The cut and count (CC) based region definition is summarised in Table 4.1 and the corresponding yields can be found in Table 4.2.

4.3 Neural Network Based Region Definition

To improve the quality of the defined regions, a neural network is implemented. Its task is to define three classes, each corresponding to either $t\bar{t}Z$, $t\bar{t}W$ or fakes events. The class scores are based on several input variables which are expected to be sensitive to these processes. The architecture of the neural network and the used input variables are further explained in Section 5.2.

For the placement of the class score cuts, the 2D separation plots in Figure 4.2 are used. These plots show the two dimensional projection of the event distribution in respect to the three class scores. Since the neural network uses the *Softmax* activation for the output

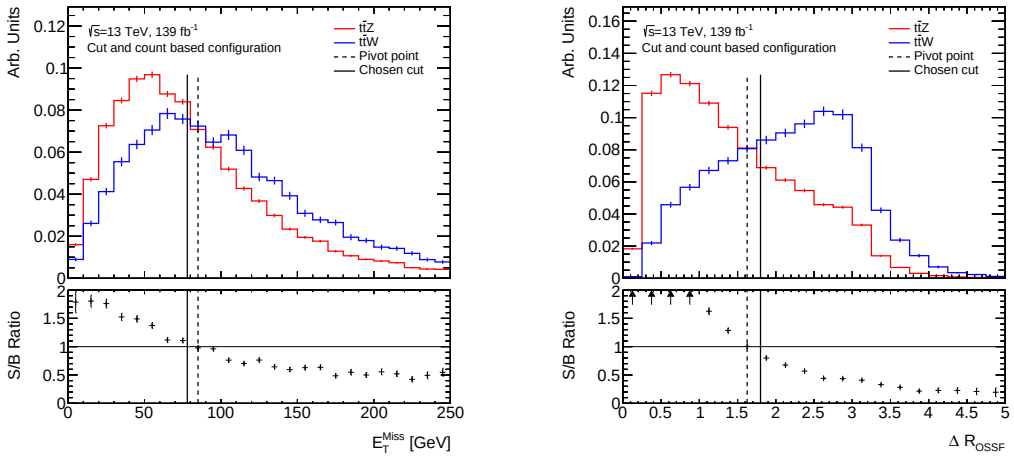

Figure 4.1: The separation plots for $E_{\mathrm{T}}^{\mathrm{Miss}}$ (left) and ΔR_{OSSF} (right) used in the cut and count based configuration. The dashed lines represent the point at which the $t\bar{t}W$ background becomes higher than the $t\bar{t}Z$ signal. The solid line shows the chosen cut for the best performance.

layer, the sum of all class scores are normalised to one. Thus, all events lie on a three dimensional plane. Two linear cuts were chosen to separate $t\bar{t}W$ and fakes from the $t\bar{t}Z$ events. Figure 4.2 also shows these cut placements.

The neural network (NN) based region definition is also summarised in Table 4.1 and the corresponding yields are listed in Table 4.2.

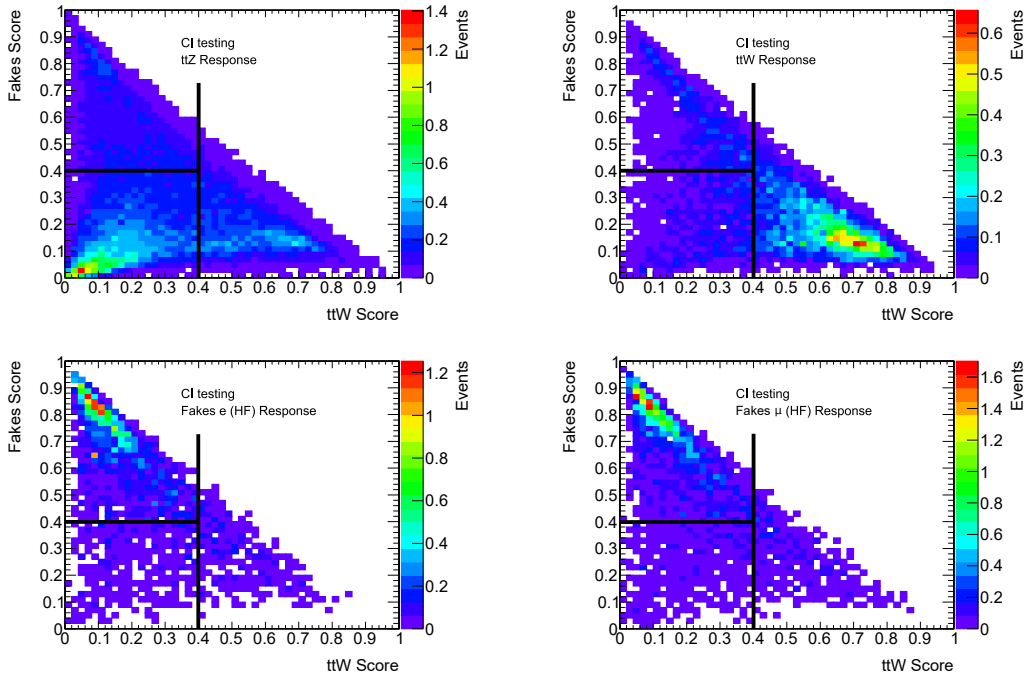

Figure 4.2: 2D projection plots of four selected samples showing the distribution of events in respect to the defined class scores. All events lie on a three dimensional plane, thus, the $t\bar{t}Z$ score is perpendicular to both the $t\bar{t}W$ and fakes score. The black lines represent the chosen cuts which are based on these plots.

Table 4.2: The yields for the cut and count based (left) and the neural network based (right) region definition.

	SR-ttZ-CC	CR-ttW-CC	CR-Fakes-CC	SR-ttZ-NN	CR-ttW-NN	CR-Fakes-NN
$t\bar{t}Z$	121 ± 4	25 ± 2	31 ± 1	95 ± 19	48 ± 10	35 ± 7
$t\bar{t}W$	26 ± 2	28 ± 2	19 ± 1	13 ± 6	46 ± 20	13 ± 6
$t\bar{t}H$	46 ± 4	19 ± 2	16 ± 2	33 ± 4	26 ± 3	22 ± 2
WZ+jets	73 ± 19	10 ± 3	21 ± 6	36 ± 9	23 ± 6	45 ± 12
ZZ+jets	20 ± 6	1 ± 1	9 ± 3	8 ± 3	3 ± 2	18 ± 6
tZq	13 ± 4	3 ± 1	6 ± 2	6 ± 2	8 ± 2	8 ± 3
tWZ	8 ± 2	3 ± 1	3 ± 1	6 ± 2	5 ± 2	3 ± 1
Fakes e (HF)	30 ± 2	16 ± 1	20 ± 1	8 ± 1	8 ± 1	51 ± 7
Fakes e (Other)	46 ± 3	25 ± 1	30 ± 3	22 ± 3	33 ± 4	46 ± 6
Fakes μ (HF)	41 ± 3	23 ± 1	25 ± 1	11 ± 2	12 ± 2	65 ± 8
Fakes (Other)	19 ± 2	8 ± 1	11 ± 2	10 ± 2	10 ± 2	18 ± 3
Others	8 ± 1	8 ± 1	5 ± 1	6 ± 1	11 ± 1	3 ± 1
Total	450 ± 24	168 ± 6	193 ± 9	254 ± 14	230 ± 14	327 ± 17
$t\bar{t}Z$/Total	0.27 ± 0.02	0.15 ± 0.02	0.16 ± 0.01	0.37 ± 0.08	0.21 ± 0.05	0.11 ± 0.03
$t\bar{t}W$/Total	0.06 ± 0.01	0.17 ± 0.02	0.10 ± 0.01	0.05 ± 0.03	0.20 ± 0.09	0.04 ± 0.02
Fakes/Total	0.30 ± 0.02	0.43 ± 0.02	0.45 ± 0.03	0.20 ± 0.03	0.27 ± 0.03	0.55 ± 0.05

5 Neural Networks and Fitting Algorithm

The implementation of a machine learning approach such as a deep neural network is expected to increase the quality of region definition. Unlike classical algorithms, neural networks can be programmed to adapt to a given dataset. This adaptation simulates a learning process and hence provides the possibility to train neural networks for certain tasks. Machine learning approaches are widely used in several modern fields of studies such as image recognition [23], identification of biological structures [24] and user-analysis on social media [25].

5.1 Deep Neural Networks

In the mid 20th century, Warren S. McCulloch and Walter Pitts formulated the first attempt to simulate an artificial network using logical structures similar to the structure of human neurons in the brain [26]. In 1957, a machine called Mark 1 Perceptron based on this idea was built by Frank Rosenblatt [27]. The machines' network consists of an input layer using an array of photocells and an output layer for the electrical signal. The connections between those layers were weighted by electric potentiometers and could be tweaked by electric motors to simulate a learning process. The performance of this network was limited, since it was an analogue single-layered network.

Modern neural networks are algorithms which simulate multi-layered deep neural networks (DNN) with several hidden layers between the input and output layer. Figure 5.1 shows a so-called feedforward neural network with one hidden layer. In a feedforward DNN each node of a layer receives a weighted input from every node of the previous layer and emits an output, which will be used in the next layer [28]. The output of the hidden layers are usually not stored and thus invisible.

This idea can be described mathematically using a vector $\vec{v}$ for input and $\vec{u}$ for output. The length of these vectors correspond to the number of nodes in the first and last layer [28]. The number of hidden layers and their respective multiplicity of nodes are determined

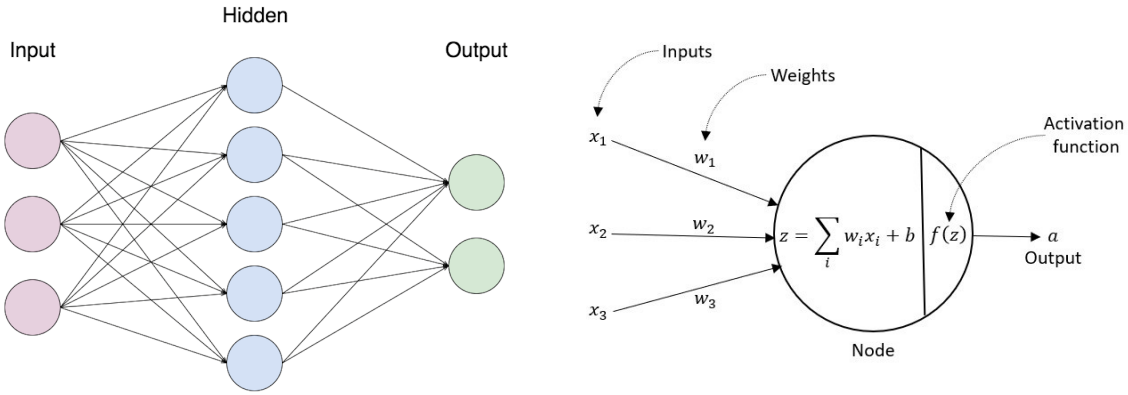

Figure 5.1: A schematic visualisation of a feedforward DNN (left) using one hidden layer. Each node can be represented (right) with multiple weighted inputs and an activation function f.

by the neural network architecture. Each node has n inputs x_i, corresponding weights w_i and an offset b which are all used to calculate the raw activation

$$z = \sum_{i}^{n} w_i x_i + b\,. \tag{5.1}$$

The raw activation z is used in a chosen activation function f to produce an output a [28]. This process per node is shown in Figure 5.1. Common choices for the activation function $f(z)$ are *tanh*, *Sigmoid*, *ReLU* and *Softmax* [29].

To train a DNN, the given dataset is usually divided into a training and a validation set. The first set is used in the training process to teach the neural network. Each step, a certain batch of samples are used for training the DNN. After the whole training set is used, an epoch is completed and the performance of the neural network is tested on the statistically independent validation set. The improvement for a given epoch is determined by the loss which needs to be minimised [30]. If the neural network trains for too many epochs on a given training dataset, it begins to overfit. This means, the DNN improves its performance on the training set, but worsens on validations sets [30]. Overfitting can be prevented by including some regularisation technique such as early stopping.

5.2 Neural Network Architecture

For the analysis, a multi-class approach is used. As mentioned before, the task of the DNN will be to define three class scores that are sensitive to the $t\bar{t}Z$, $t\bar{t}W$ and fakes respectively. The fakes score includes all fake samples with no further subdivision. These scores are used for region definition as described in Section 4.3.

The framework used for the neural network is based on KERAS [31], TENSORFLOW [32] and SCIKIT-LEARN [33]. To adjust the neural network, several DNN parameters were individually varied and optimised.

The number of hidden layers and their respective multiplicity of nodes can be chosen to provide enough degrees of freedom to the network [34]. For this analysis, 4 layers with 20, 25, 20 and 20 nodes, respectively, were chosen. Additionally, the activation function for the nodes of each layer [29] must be chosen depending on the desired output. The hidden layers use the $ReLU$ activation and the output layer uses *Softmax* activation since it normalises the class scores and thus allows for an easier cut placement. In order to achieve a learning process, the neural network must be trained for a given number of epochs [34]. Each epoch, the neural network trains on the training set using the results from previous epochs. The maximum number of epochs is set to 300, however, this value is not reached, since early stopping is implemented and occurs at around 150 epochs. For training, the optimiser NADAM [35] is implemented.

The relative size of the validation set must be selected sufficiently large to be representative, however, a smaller validation set allows for a bigger training set. Thus a compromise between those two is chosen carefully and is set to 20%. To better utilise the given dataset, k-fold cross validation [36] is implemented with 2 folds. Furthermore, the batch size [34] as well as the learning rate [28] are optimised in dedicated studies resulting in a batch size of 128 events per batch and a learning rate of 0.0005.

As input for the DNN, several kinematic variables were tested. The chosen variables are expected to be sensitive to either of the three processes and thus the choice is motivated by the underlying physics processes. The DNN uses jet and b-jet multiplicity, the missing energy $E_{\mathrm{T}}^{\mathrm{Miss}}$, ΔR of the OSSF-pair and the three highest p_{T} of jets and leptons as input.

The choice to implement jet multiplicity is motivated by the expected signals as explained in Section 2.3 and the reasons for choosing $E_{\mathrm{T}}^{\mathrm{Miss}}$ and ΔR_{OSSF} are discussed in Section 4.2. Including the transverse momentum p_{T} for both jets and leptons is motivated by the fakes background, which is expected to be sensitive to the low regions of the third lepton's p_{T} in particular. The additional p_{T} variables were tested and found to improve the DNN performance.

Moreover, implementing the sum of lepton charges is also expected to improve the DNN performance, because, as explained in Section 2.3, the $t\bar{t}W$ processes happen as initial state radiation. Thus, an asymmetry in the sum of lepton charges is expected for the $t\bar{t}W$ events but not for $t\bar{t}Z$ or fakes processes. However, this variable is not included in the analysis, since the used DNN-framework does not support sums of variables as input. Further inputs were tested and found to be non-beneficial for the DNN performance.

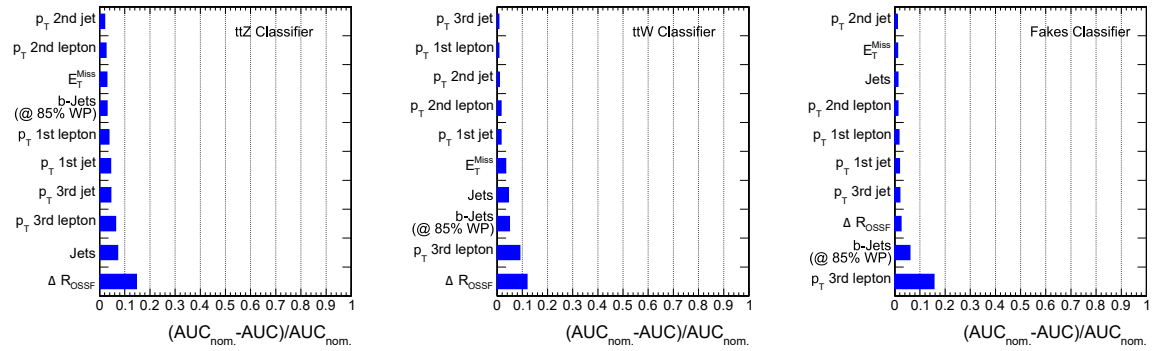

Figure 5.2: Histograms showing the impact of input variables for the $t\bar{t}Z$ (left), $t\bar{t}W$ (centre) and fakes (right) class. The variables are sorted by their impact for each classifier.

The impact of the input variable is analysed for each region and the results are summarised in Figure 5.2. As expected, ΔR_{OSSF} has the most significant impact for the $t\bar{t}Z$ score. Similarly, for the fake classification, the third lepton's p_{T} is found to have the highest impact. Both of these variables also provide approximately equally high separation performance for the $t\bar{t}W$ classifier. The missing energy has a lower impact than expected. Nevertheless, the $E_{\text{T}}^{\text{Miss}}$ impact is highest for the $t\bar{t}W$ classifier.

To further evaluate the performance of the DNN, confusion matrices for each of the three classifiers are plotted as shown in Figure 5.3. Each confusion matrix shows the classification of events per sample. The $t\bar{t}Z$ classifier shows that high fractions of around 70% of $t\bar{t}Z$ events have a $t\bar{t}Z$ score of less than 0.5. Additionally, these plots show that the separation performance between $t\bar{t}W$ and the *Others* sample is relatively low. This is expected, since one of the processes included in the *Others* sample is $t\bar{t}WW$ which has a similar event signature. Since the *Others* sample has low yields in comparison to $t\bar{t}W$, as seen in Table 4.2, this will not have a significant impact on the DNN performance. The fake classifier works best on the heavy flavour (HF) fake samples. For the other two fake samples around $\sim 60\%$ of events score lower than 0.5 for the fake classifier. Also, the WZ/ZZ+jets events achieved relatively high fake class scores. These results are supported by the observed yields for the neural network based configuration in Table 4.2.

The overall performance for the defined DNN can be assessed using ROC curves [37]. The area under the curve (AUC) is calculated for each of the folds and is a measure of the DNN performance. The computed integral can be in the range of $[0, 1]$, values close to 0.5 indicate a bad performance that is comparable to guessing. Very performant DNN would achieve values close to one or zero [37]. The calculated ROC curves are summarised in Figure 5.4 and show, that the $t\bar{t}W$ classifier achieves the best performance.

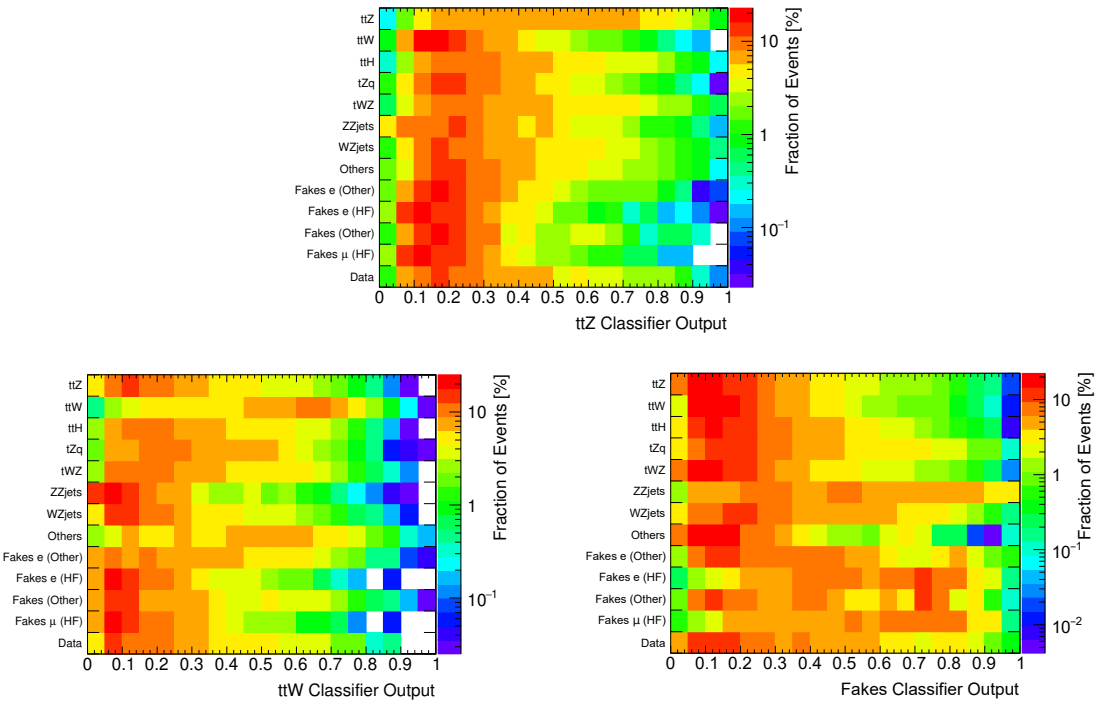

Figure 5.3: The confusion matrices showing the distribution of outputs separated into each sample for the $t\bar{t}Z$ (top), $t\bar{t}W$ (bottom left) and fakes (bottom right) classifier.

5.3 Fitting Algorithm

For the fitting process, a binned profiled likelihood fit [38] is performed. In a likelihood fit, a probability density function $L(\mu, \theta)$ that depends on parameters of interest (PoI) μ and nuisance parameters (NP) θ is optimised to describe the observed events as the most likely. The profiled likelihood fit reduces the number of dimensions by expressing the nuisance parameters through the PoI. For the fitting process, a dataset in which the model parameters are set to their expected value is generated and used instead of the data set. The value of the PoI used in the fit is set to one.

Furthermore, for each NP with a Gaussian distribution of mean μ and width σ, the pull

$$g = \frac{x - \mu}{\sigma}$$

can be defined and will be distributed as a standard Gaussian. A deviation of the pulls' mean from zero, is an indicator for possible forms of bias. Uncertainties different to ± 1 represent constraints for a given NP [39].

This analysis fits the signal strength μ_{ttZ} and two normalisation factors N_{ttW}, N_{Fakes} as

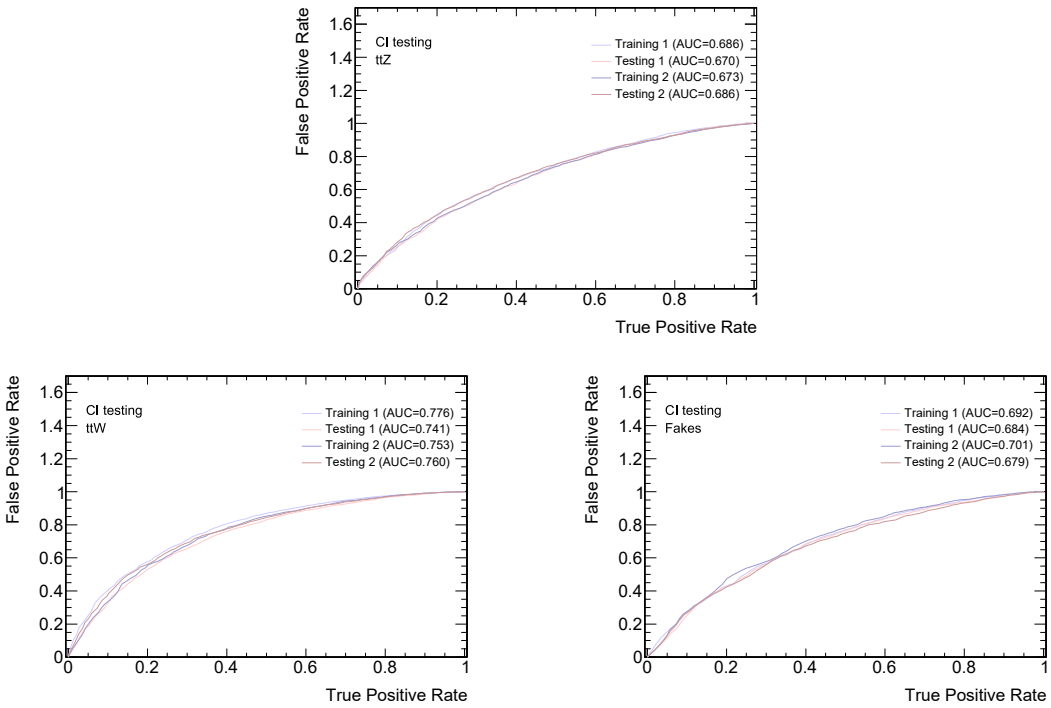

Figure 5.4: The ROC curves showing the performance for the $t\bar{t}Z$ (top), $t\bar{t}W$ (bottom left) and fakes (bottom right) class. The training and testing number refers to the k-fold.

PoI simultaneously. The included uncertainties described in Chapter 6 are implemented as NP. The results of the fitting process are given in Chapter 7.

6 Experimental and Theoretical Systematic Uncertainties

The study includes several uncertainties on the used event samples to consider possible deviations from the predicted values. The included uncertainties can be subdivided into experimental uncertainties, which cover the identification and reconstruction of objects inside the detector, and theoretical uncertainties, which cover modelling and estimation variability. In the fitting process, these uncertainties are implemented as nuisance parameters. Up-to-date recommendations of several performance groups are used to determine the uncertainty size.

6.1 Experimental Uncertainties

The uncertainty on the luminosity of $139\,\mathrm{fb}^{-1}$ for the combined Run II dataset is measured to be 1.7% by the official luminosity working group using the LUCID-2 detector [40] and scans for x-y beam-separation by following a method similar to that described in Reference [41]. This nuisance parameter is applied to all MC samples.

Uncertainties related to pile up in the detector [42] are implemented to match the data and MC distributions by rescaling the pile-up correction factors during the reweighting. This nuisance parameter is symmetrised using the average of two variations for the upper and lower uncertainty. The total impact of these effects is around 0.2%-1.0% per sample.

For the lepton selection, NP for several efficiencies such as trigger, reconstruction, identification and signal isolation as well as momentum scale and resolution are implemented [43, 44]. These uncertainties are all symmetrised using two variations and are subdivided into NP for electrons and muons. Each uncertainty individually is below 1%. Overall there are 51 separate uncertainties for electrons and 37 for muons.

For the missing energy, two symmetric NP for the resolution and one NP for the scaling are applied [45]. Also, three more NP are used for the electron-photon identification and isolation [43]. Each of these six uncertainties has an impact less than 1%.

Analogous to the leptons, jet uncertainties are implemented for the energy scale (JES),

energy resolution (JER) and the vertex tagging (JVT) [46]. These were derived from test-beam and collision data, as well as MC simulations. In total, 46 uncertainties are applied, each less than 2%.

Further uncertainties are applied for the flavour tagging uncertainties. These are separated for b-tagging [47], c-tagging [48] and light quark tagging [49]. Measurements for those are done using control samples in data and MC to calculate correction factors to correct the rates of flavour tagging in the simulations. In total, 85 uncertainties each less than 2% are applied for flavour tagging. Each of them is also symmetrised by averaging an up- and down-variation.

6.2 Theoretical Uncertainties

For the theoretical prediction of the $t\bar{t}Z$ events, several uncertainties are implemented. The evaluation of the renormalisation scale μ_{r} and factorisation scale μ_{f} uncertainty is done by varying Pythia 8 samples. This nuisance parameter is not symmetric. Another uncertainty is added for the A14 tune of Pythia 8 [50] and has an impact of $\sim$2%. Two additional uncertainties from the choice of PDF are set according to the PDF4LHC recommendation [51], each having an impact less than 0.5%. The systematic uncertainty for the modelling of the parton shower and the hadronisation is evaluated using alternative samples generated with AMC@NLO [52] which is interfaced to Herwig 7 [53] instead of Pythia 8. The impact from the shower uncertainty is up to 6.3% and is symmetrised using one variation.

Most theoretical NP cover the diboson WZ+jets and ZZ+jets processes. These uncertainties are taken from the most recent $t\bar{t}Z$ publication [54] in which the uncertainties were determined by comparing data to MC for Zb/Zc events and taking heavy flavour jet fraction between Z+jets and ZZ/WZ+jets into account. Therefore, a flat cross-section uncertainty of 30% is applied to the ZZc, WZc and WZl events. For ZZl an uncertainty of 15% is used and for ZZb as well as WZb an uncertainty of 50% is used. Further uncertainties for these processes related to the renormalisation and factorisation scale as well as the PDF choice are derived using the same methods as for the signal $t\bar{t}Z$ process. Analogously to $t\bar{t}Z$, neither nuisance parameters on PDF scale, nor the renormalisation and factorisation are symmetrised.

For the tWZ background, in addition to the diboson NP, an uncertainty for the interference between $t\bar{t}Z$ and tWZ processes is implemented, which is calculated using the DR2 model compared to the DR1 model as described in Reference [55]. It is evaluated by calculating their modelling difference of uncertainties in the signal regions. This un-

certainty is applied symmetrically to the tWZ events. Moreover, a flat normalisation uncertainty of 10% is applied.

Following the results of the measurements in Reference [56], a flat 14% normalisation uncertainty is applied for the tZq cross-section. PDF and factorisation related uncertainties are again obtained as in the previous samples. Additional uncertainties for the tZq showering and radiation are motivated by the measurements in Reference [57] and [58].

For $t\bar{t}H$ events, a flat asymmetric uncertainty of $+6.8\%$ and -9.9% is applied. The implementation of this nuisance parameter covers the uncertainties from QCD scale and PDF choice [59].

For additional negligible processes like $t\bar{t}t\bar{t}$ or $t\bar{t}WW$ a single flat normalisation uncertainty is applied. These minor processes contribute less than 1% to the total yield as seen in 4.2 and thus, a 50% flat uncertainty is assumed to be sufficient.

7 Measurements of the Inclusive Cross-Section

As explained before, to evaluate the simultaneous measurements of $t\bar{t}Z$ and $t\bar{t}W$, the signal strength μ_{ttZ} and two normalisation factors N_{ttW} and N_{Fakes} for the respective samples are fitted. The normalisation factor N_{Fakes} does not separate the four fakes samples but treats them as one group. The fitting process uses a binned profile likelihood fit as described in Section 5.3 and is done for both region definitions which are summarised in Chapter 4.

Cut and Count Based Configuration

Figure 7.1 shows the post-fit distributions for each region of the cut and count based configuration. As mentioned in Section 4.1, the binning is automatically done by an algorithm. The underflow and overflow are included in the first and last bin, respectively. None of the discrepancies between MC data is greater than 3σ. The fitting results for the parameters of interest are

$$\mu_{t\bar{t}Z} = 1.00 \pm 0.22 = 1.00 \pm 0.19(\mathrm{stat.}) \pm 0.12(\mathrm{syst.}) \tag{7.1}$$

$$N_{t\bar{t}W} = 1.00 \pm 0.81 = 1.00 \pm 0.71(\mathrm{stat.}) \pm 0.39(\mathrm{syst.}) \tag{7.2}$$

$$N_{\mathrm{Fakes}} = 1.00 \pm 0.20 = 1.00 \pm 0.18(\mathrm{stat.}) \pm 0.09(\mathrm{syst.}) \tag{7.3}$$

whereby the systematic uncertainty is calculated from the statistical and total uncertainty. The ranking plots, which list the most important nuisance parameters, are shown in Figure 7.2. Many of the diboson uncertainties are impactful, since the WZ/ZZ events have yields comparable to $t\bar{t}W$, as seen in Table 4.2. Since the analysed channel has low number of events, the luminosity and JET pile up NP are significant for the $\mu_{t\bar{t}Z}$ uncertainty. For the $N_{t\bar{t}W}$ uncertainties, two of the JER related nuisance parameters have an significant impact. The missing energy is crucial for neutrino reconstruction and thus is important for the $t\bar{t}W$ events, resulting in the significant impact. The ranking for N_{Fakes} shows that one of the jet tagging uncertainties has the second highest impact. The correct

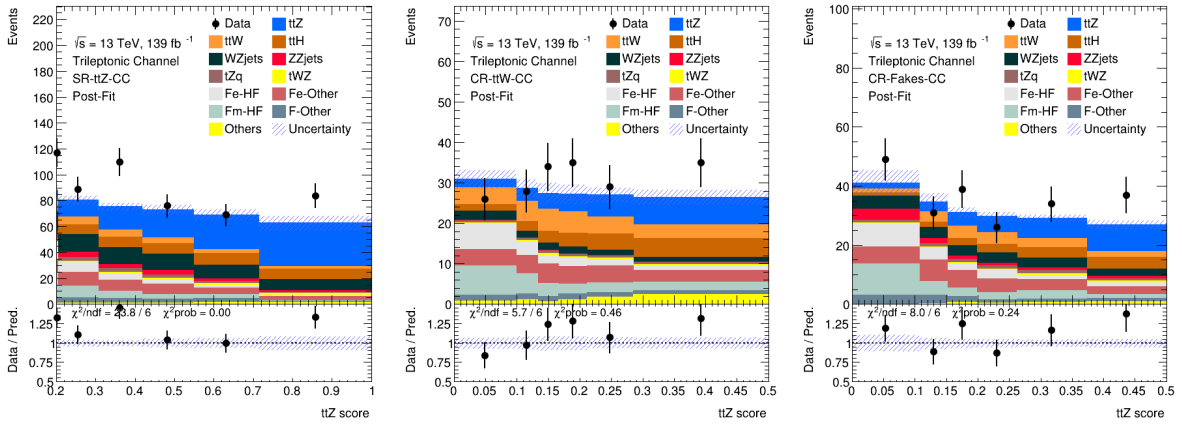

Figure 7.1: The post-fit distributions for the SR-ttZ-CC (left), CR-ttW-CC (centre), CR-Fakes-CC (right) region of the cut and count based configuration.

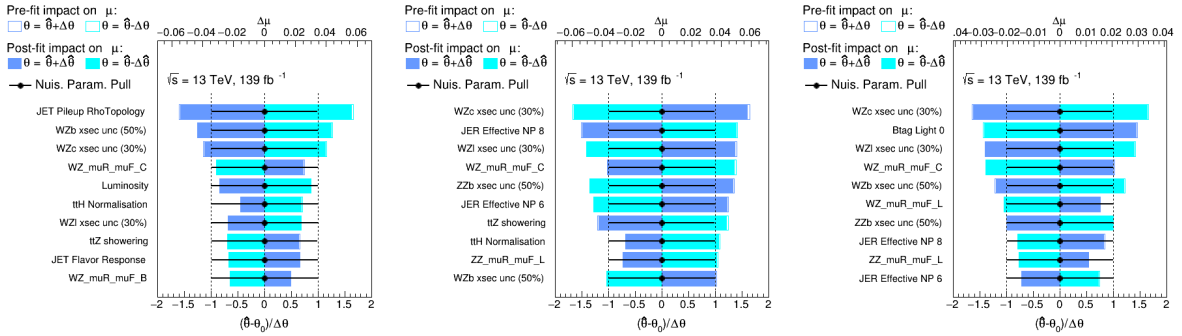

Figure 7.2: The ranking plots for $\mu_{t\bar{t}Z}$ (left), $N_{t\bar{t}W}$ (centre) and N_{Fakes} (right) using the cut and count based regions.

identification for light jets is important for fakes because these non-prompt leptons have relatively low energies and thus can be misidentified as light jets. All nuisance parameters are also analysed using pull plots in Figure 7.5 and are found to have no significant pull or constrain.

Neural Network Based Configuration

Analogous to the cut and count based configuration, Figure 7.3 shows the post-fit distributions for the neural network based configuration. Again, the discrepancies between

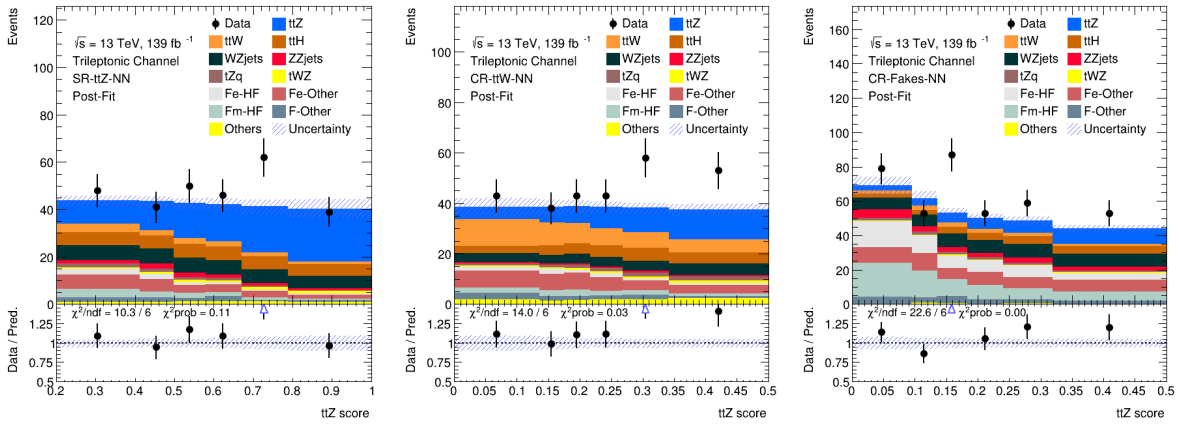

Figure 7.3: The post-fit distributions for the SR-ttZ-NN (left), CR-ttW-NN (centre), CR-Fakes-NN (right) region of the neural network based configuration.

MC and data are less than 3σ for all bins. The fitting results are given by

$$\mu_{t\bar{t}Z} = 1.00 \pm 0.24 = 1.00 \pm 0.18(\text{stat.}) \pm 0.15(\text{syst.}) \tag{7.4}$$

$$N_{t\bar{t}W} = 1.00 \pm 0.44 = 1.00 \pm 0.42(\text{stat.}) \pm 0.14(\text{syst.}) \tag{7.5}$$

$$N_{\text{Fakes}} = 1.00 \pm 0.14 = 1.00 \pm 0.12(\text{stat.}) \pm 0.07(\text{syst.}) \,. \tag{7.6}$$

The ranking plots in Figure 7.4 show similar results to the ranking plots from the cut and count based configuration. Most NP are diboson uncertainties for the same reasons as mentioned before. Surprisingly, the jet tagging uncertainty, which is also important for N_{Fakes}, is the third most impactful NP for $\mu_{t\bar{t}Z}$. This is most likely caused by some neural network classification that is sensitive to this nuisance parameter. These plots also show, that many of the NP, which are impactful on $\mu_{t\bar{t}Z}$, are also impactful on $N_{t\bar{t}W}$. This is expected since these two processes have similar event signatures and are difficult to separate. Moreover, Figure 7.5 shows the pull plots for the neural network based configuration. As before, no significant constrains or pulls are observed.

Comparison

Comparing the yields for both configurations in Table 4.2 shows that the neural network based configuration has cleaner regions. The relative number of $t\bar{t}Z$ events in the signal region has increased by 10%, while the relative number of $t\bar{t}W$ and fake events has decreased by 1% and 10%, respectively. Similar results can be observed for the control regions with the exception, that the neural network based $t\bar{t}W$ control region has 6% more $t\bar{t}Z$ events compared to the cut and count based one.

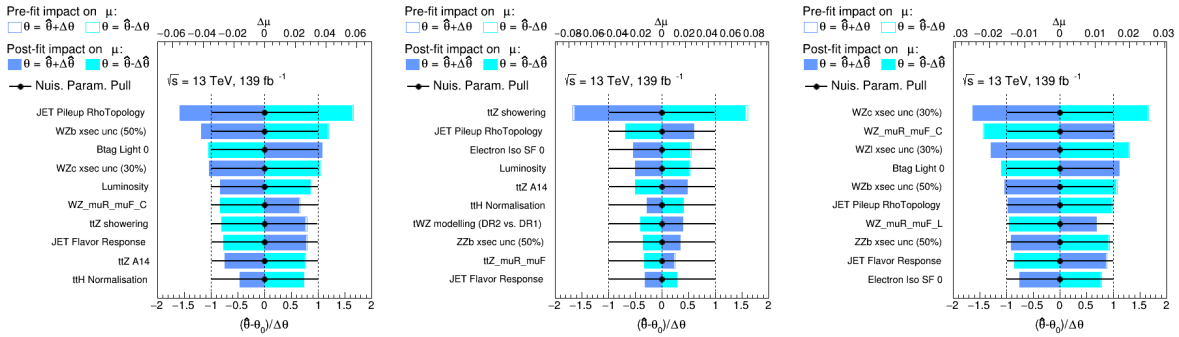

Figure 7.4: The ranking plots for $\mu_{t\bar{t}Z}$ (left), $N_{t\bar{t}W}$ (centre) and N_{Fakes} (right) using the neural network based regions.

The uncertainty on the PoI for normalisation has decreased by 37% for $t\bar{t}W$ and 6% for fakes. Both the statistical and systematic uncertainty is halved for $N_{t\bar{t}W}$ using the neural network based configuration. For N_{Fakes}, the improvements are mostly for statistical uncertainties. The uncertainty of $\mu_{t\bar{t}Z}$, however, is increased by 2%. This increase is most likely caused by the low statistics in the neural network based signal region. Comparing the total values in Table 4.2 shows, that the SR-ttZ-NN has 44% less events compared to SR-ttZ-CC.

Although the neural network has refined the region definition, the limited statistics for analysed channel restrict the improvements on the signal strength $\mu_{t\bar{t}Z}$ in the fitting processes. Hence, if only the signal strength $\mu_{t\bar{t}Z}$ is analysed, the cut and count based configuration should be used. However, for a simultaneous analysis, it is recommended to use the neural network based configuration because of its overall performance.

To reduce the uncertainties due to low statistics, more events need to be measured. The High-Luminosity LHC upgrade after Run III is planned to provide an integrated luminosity of $250\,\text{fb}^{-1}$ per year [60]. This could also improve the DNN training since a greater dataset for training can be used.

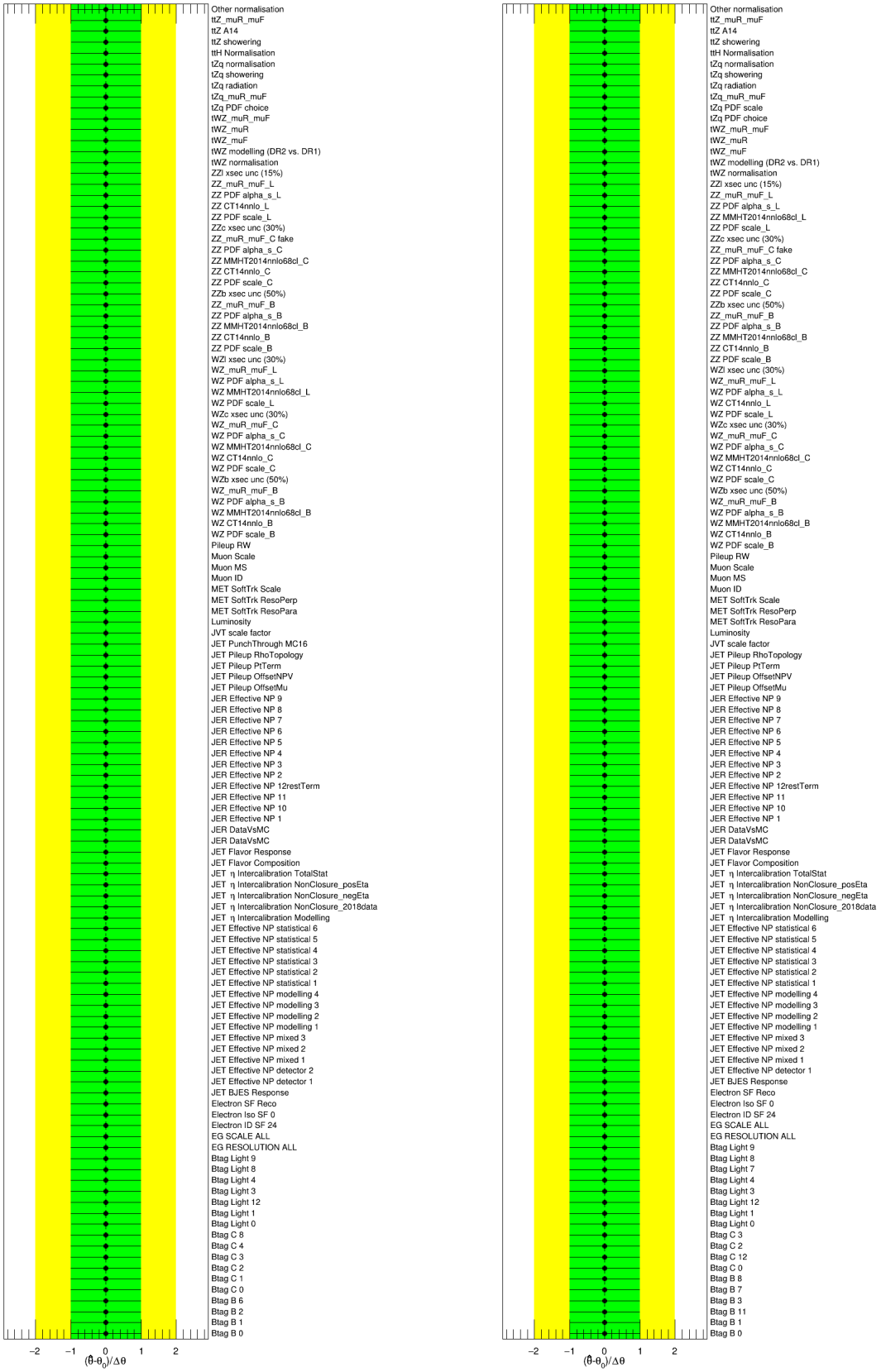

Figure 7.5: Pull plots for the cut and count based (left) and neural network based (right) configuration showing all uncertainties and possible pulls or constraints.

8 EFT Sensitivity

As explained in Section 2.3, effects from SM-EFT contributions can be investigated by studying off-shell $t\bar{t}Z$ events. For that, the cross-section ratio and separation power of MC simulations using variations on Wilson coefficients compared to the SM predictions is calculated. Since the operator of interest Q_{tZ} is a linear combination of Q_{tB} and Q_{tW}, as formulated in Equation 2.2, cross-section ratios and separation power are calculated for the Wilson coefficients C_{tB} and C_{tW}.

8.1 Cross-Section Ratio

For the calculation of the cross-section ratio, several distributions for different kinematic variables are used. The set of analysed variables for the Z-boson consists of the cosine of the deflection angle $\cos(\theta^*_{\mathrm{Z}})$, the reconstructed mass m_{Z} and the transverse momentum $p_{\mathrm{T,Z}}$. For m_{Z} and $p_{\mathrm{T,Z}}$ there is an alternative method of reconstruction, which is based on ΔR_{OSSF} and is indicated by the subscript ΔR.

The plotted distributions in Figure 8.1 show the analysed variations on C_{tB} and C_{tW} in the SR-ttZ-NN region for $\cos(\theta^*_{\mathrm{Z}})$. These studies are continued for the all mentioned variables and also for the other regions of the neural network based configuration, resulting in the averaged cross-section ratios which are summarised in Table 8.1.

The results show that the highest sensitivity is seen by analysing the variables $\cos(\theta^*_{\mathrm{Z}})$ and m_{Z}. For C_{tB}, the alternative ΔR_{OSSF} based variables show no significant difference to the non-ΔR_{OSSF} based ones. However, the results for C_{tW} show noticeable differences between m_{Z} and $m_{\mathrm{Z},\Delta R}$. Furthermore, the change in cross-section is higher in variations for the imaginary parts of the Wilson coefficients. This sensitivity to the imaginary part can introduce a non-flat offset, which can be seen in Figure 8.1. It should be noted, that the variations on the real part of C_{tW} also seem to introduce a non-flat offset. Since the distribution of the Wilson coefficient variation has non-flat offsets in comparison to the SM prediction, a non-zero separation is expected [61].

Table 8.1: The cross-section ratios calculated for several variations on the Wilson coefficient C_{tB} (top) and C_{tW} (bottom), separated into real and imaginary part. All values are averaged over all three regions defined by the neural network based configuration. The highest sensitivities are marked.

	$\cos(\theta^*_{\mathrm{Z}})$	m_{Z}	$m_{\mathrm{Z},\Delta R}$	$p_{\mathrm{T,Z}}$	$p_{\mathrm{T,Z},\Delta R}$
ctBRe +0.3	1.00 ± 0.00	1.01 ± 0.01	1.00 ± 0.01	1.00 ± 0.01	1.00 ± 0.01
ctBRe −0.3	1.01 ± 0.00	1.01 ± 0.01	1.01 ± 0.01	1.00 ± 0.01	1.00 ± 0.01
ctBRe +0.9	1.03 ± 0.01	1.02 ± 0.01	1.02 ± 0.01	1.02 ± 0.01	1.01 ± 0.01
ctBRe −0.9	1.04 ± 0.01	1.04 ± 0.01	1.04 ± 0.01	1.03 ± 0.01	1.02 ± 0.01
ctBIm +0.6	1.03 ± 0.01	1.03 ± 0.01	1.03 ± 0.01	1.02 ± 0.01	1.02 ± 0.01
ctBIm −0.6	1.02 ± 0.01	1.01 ± 0.01	1.01 ± 0.01	1.01 ± 0.01	1.01 ± 0.01
ctBIm +1.8	1.17 ± 0.01	1.14 ± 0.01	1.14 ± 0.01	1.10 ± 0.01	1.09 ± 0.01
ctBIm −1.8	1.17 ± 0.01	1.14 ± 0.01	1.14 ± 0.01	1.10 ± 0.01	1.09 ± 0.01

	$\cos(\theta^*_{\mathrm{Z}})$	m_{Z}	$m_{\mathrm{Z},\Delta R}$	$p_{\mathrm{T,Z}}$	$p_{\mathrm{T,Z},\Delta R}$
ctWRe +0.7	1.03 ± 0.01	1.02 ± 0.01	1.02 ± 0.01	1.02 ± 0.01	1.01 ± 0.01
ctWRe −0.7	1.03 ± 0.01	1.03 ± 0.01	1.03 ± 0.01	1.02 ± 0.01	1.02 ± 0.01
ctWRe +1.1	1.07 ± 0.01	1.04 ± 0.01	1.05 ± 0.01	1.04 ± 0.01	1.04 ± 0.01
ctWRe −1.1	1.07 ± 0.01	1.07 ± 0.01	1.06 ± 0.01	1.04 ± 0.01	1.04 ± 0.01
ctWIm +0.6	1.02 ± 0.01	1.02 ± 0.01	1.00 ± 0.01	1.01 ± 0.01	1.01 ± 0.01
ctWIm −0.8	1.04 ± 0.01	1.03 ± 0.01	0.99 ± 0.01	1.03 ± 0.01	1.03 ± 0.01
ctWIm +1.2	1.10 ± 0.01	1.08 ± 0.01	0.99 ± 0.01	1.06 ± 0.01	1.06 ± 0.01
ctWIm −1.4	1.13 ± 0.01	1.10 ± 0.01	0.99 ± 0.01	1.08 ± 0.01	1.08 ± 0.01

8.2 Separation Power

The separation power is analysed for both Wilson coefficients, subdivided into the real and imaginary part analogous to the cross-section ratio. The same set of variables is analysed. Using Equation 2.3 and Gaussian error propagation, the separation power is calculated individually for each region of the neural network based configuration. The averaged separation power for the summed distributions C_{tB} are in the order of magnitude of 10^{-4}–10^{-7} and for C_{tW} in the order of 10^{-2}–10^{-6}. Both having dominating errors in the order of magnitude of 10^{-1}–10^{-2}.

However, certain regions such as for $\cos(\theta^*_{\mathrm{Z}})$ are expected to have a significant separation introduced by the non-flat offset. This prediction can not be supported, since these regions also have high uncertainties due to the low statistics. To further study the separation in these regions, higher statistics are needed. As mentioned in Chapter 7, the High-Luminosity upgrade will provide significantly more data and thus will be useful to this analysis.

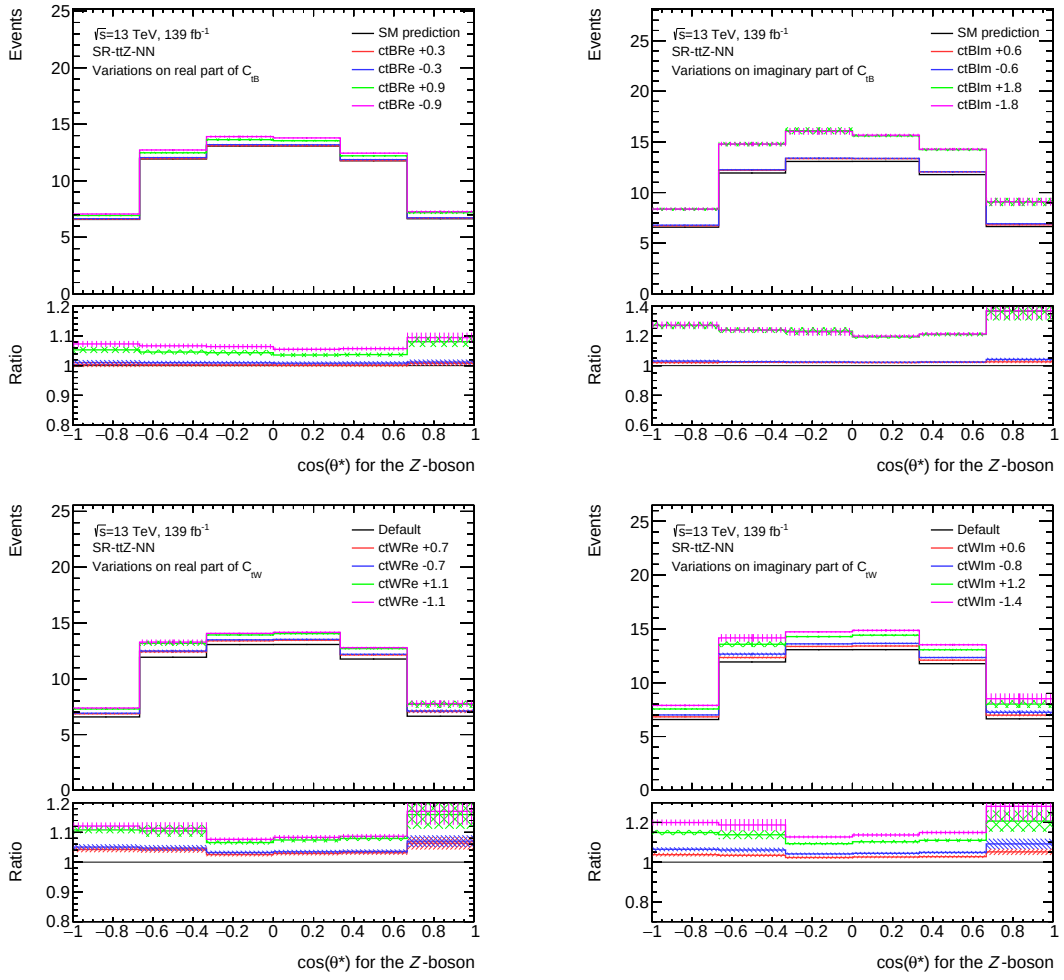

Figure 8.1: The distribution plots showing several variations on the C_{tB} (top) and C_{tW} (bottom) Wilson coefficient plotted against the $\cos(\theta^*_{\mathrm{Z}})$ variable in the SR-ttZ-NN region. These are further separated into the real (left) and imaginary (right) part.

9 Conclusion

To conclude this thesis, a brief summary and possible improvements for future studies are presented.

This study analysed the simultaneous measurements of $t\bar{t}Z$ and $t\bar{t}W$ events in off-shell regions. This was done by defining sensitive regions, which were then used in the fitting process of the inclusive cross-section. Additionally, this study further separated the fakes background, which is dominating in the trileptonic decay channel.

The region definition used two different approaches, as described in Chapter 4, to define a signal region for $t\bar{t}Z$ and two control regions for $t\bar{t}W$ and fakes. The cut and count based configuration is based on manually set cuts on $E_{\mathrm{T}}^{\mathrm{Miss}}$ and ΔR_{OSSF}, which were studied and optimised. For the neural network based configuration a DNN was trained to identify $t\bar{t}Z$, $t\bar{t}W$ and fakes events by using several variables as inputs. This DNN defined class scores for each process, which were then used to define the three regions. The comparison of both approaches showed that the DNN based regions are cleaner in comparison to the cut and count based regions. However, the number of events in the NN based signal region decreased by 44%, which is problematic for the analysis, because the off-shell region provides low statistics. These results were reflected in the fitting process.

A binned likelihood fit was conducted including one PoI for the signal strength $\mu_{t\bar{t}Z}$ and two PoI for the normalisation of $t\bar{t}W$ and fakes as well as several NP for the uncertainties. As expected from the region definition, the uncertainties for the normalisation factors decreased using the NN based configuration. For $N_{t\bar{t}W}$ the uncertainties decreased by 37%, for N_{Fakes} by 6%, whereby the uncertainties on $\mu_{t\bar{t}Z}$ increased by 2%.

Thus, for studies in the off-shell region, where $t\bar{t}Z$ or $t\bar{t}W$ events are analysed, it is recommended to use a neural network based region definition, as it improves the overall results and especially the of the different processes. The improvements in the normalisation are expected to outweigh the increased uncertainty of the signal strength.

As discussed before, in Chapter 7, the most impactful improvement is expected to be the usage of a greater dataset, since the statistical uncertainty is higher than the systematic uncertainty for all PoI. Further improvements could be achieved by modifying the used DNN framework to allow the sum of variables to be used as input. By studying the chosen

cuts for the NN based configuration in more detail, including non-linear cuts, additional improvements could be realised.

Furthermore, the SM-EFT sensitivity was analysed in Chapter 8 by comparing the SM-prediction to several samples which include variations on Wilson coefficients. For the comparison, the cross-section ratios and the separation power was calculated for several variables. The results of the cross-section ratio showed that the studied variables have higher sensitivity to the imaginary parts of the analysed variations. The separation power provided no further information because of the dominating uncertainties due to the low amount of events.

As for the fitting results, a greater dataset is expected to reduce the uncertainties and allow for more precise calculations. Furthermore, to increase the sensitivity to possible EFT contributions further, a separation into high-Z-mass and low-Z-mass regions could be useful. This, as mentioned in Section 2.3, is due to the SM-EFT contributions increase for higher boson masses.

Bibliography

[1] M. Gell-Mann, *The Eightfold Way: A Theory of strong interaction symmetry*, California Institute of Technology (1961)

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

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

[4] S. L. Glashow, *Partial Symmetries of Weak Interactions*, J. Phys. G. **22** (1961)

[5] C. Jarlskog, *Quantum flavour dynamics*, Phys. Scrs **24(5)** (1981)

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

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

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

[9] SNO Collaboration, *Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory*, Phys. Rev. Lett. **89** (2002)

[10] F. Zwicky, *Die Rotverschiebung von extragalaktischen Nebeln*, Helv. Phys. Acta **6** (1933)

[11] CDF Collaboration, *Observation of top quark production in $\bar{p}p$ collisions*, Phys. Rev. Lett. **74** (1995)

[12] DØ Collaboration, *Observation of the top quark*, Phys. Rev. Lett. **74** (1995)

[13] ATLAS Collaboration, *Measurement of the top quark mass in the $t\bar{t} \to$ lepton+jets channel from $\sqrt{s} = 8$ TeV ATLAS data and combination with previous results*, Eur. Phys. J. C **79(4)** (2019)

[14] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (1963)

[15] A. Denner et al., *NLO QCD corrections to off-shell top-antitop production with leptonic decays at hadron colliders*, JHEP **10** (2012)

[16] D. Barducci et al., *Interpreting top-quark LHC measurements in the standard-model effective field theory* (2018), 1802.07237

[17] B. Grzadkowski et al., *Dimension-Six Terms in the Standard Model Lagrangian*, JHEP **10** (2010), 1008.4884

[18] L. R. Evans, *The Large Hadron Collider (LHC)*, Eur. Phys. J. C **34** (2004)

[19] N. Karastathis et al., *LHC Run 3 Configuration Working Group Report* (2019)

[20] ATLAS Collaboration, *The ATLAS experiment at the CERN large hadron collider*, JINST **3(S08003)** (2008)

[21] ATLAS Collaboration, *ATLAS Calorimeter: Run 2 performance and Phase-II upgrades*, PoS **EPS-HEP2017** (2017)

[22] T. P. Calvet, *Search for the production of a Higgs boson in association with top quarks and decaying into a b-quark pair and b-jet identification with the ATLAS experiment at LHC* (2017), URL https://cds.cern.ch/record/2296985

[23] D. Ciregan et al., *Multi-column deep neural networks for image classification*, IEEE Transactions on Neural networks (2012)

[24] A. Zhavoronkov et al., *Deep learning enables rapid identification of potent DDR1 kinase inhibitors*, Nat. Biotechnol. **37(9)** (2019)

[25] H. Lin et al., *User-level psychological stress detection from social media using deep neural network*, Journal of the ACM (2014)

[26] W. S. McCulloch and W. Pitts, *A logical calculus of the ideas immanent in nervous activity*, The bulletin of mathematical biophysics **5(4)** (1943)

[27] F. Rosenblatt, *The perceptron: a probabilistic model for information storage and organization in the brain*, Psychological review **65(6)** (1958)

[28] H. Larochelle et al., *Exploring strategies for training deep neural networks*, J. Mach. Learn. Res. **10(1)** (2009)

[29] S. Sharma, S. Sharma and A. Athaiya, *Activation functions in neural networks*, Towards Data Science **6(12)** (2017)

[30] D. Nitish, *A simple way to prevent neural networks from overfitting*, J. Mach. Learn. Res. **(15)** (2014)

[31] A. Gulli and S. Pal, *Deep learning with Keras*, Packt Publishing Ltd. (2017)

[32] J. V. Dillon et al., *TensorFlow Distributions* (2017), `1711.10604`

[33] F. Pedregosa et al., *Scikit-learn: Machine Learning in Python*, J. Mach. Learn. Res. **12** (2011)

[34] F. H. Leung et al., *Tuning of the structure and parameters of a neural network using an improved genetic algorithm*, IEEE Transactions on Neural networks **14(1)** (2003)

[35] T. Dozat, *Incorporating nesterov momentum into adam*, ICLR (2016)

[36] T. Fushiki, *Estimation of prediction error by using K-fold cross-validation*, Stat. Comput. **21(2)** (2011)

[37] Luzia Gonçalves et al., *ROC curve estimation: An overview*, REVSTAT **12(1)** (2014)

[38] S. R. Cole, H. Chu and S. Greenland, *Maximum Likelihood, Profile Likelihood, and Penalized Likelihood: A Primer*, Am. J. Epidemiol. **179(2)** (2013)

[39] Li-Gang Xia, *Study of constraint and impact of a nuisance parameter in maximum likelihood method*, J. Phys. G **46(8)** (2019)

[40] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, JINST **13(07)** (2018)

[41] ATLAS Collaboration, *Improved luminosity determination in pp collisions at $\sqrt{s}$ = 7 TeV using the ATLAS detector at the LHC*, Eur. Phys. J. C **73(8)** (2013)

[42] ATLAS Collaboration, *Pile-up subtraction and suppression for jets in ATLAS*, ATLAS-CONF-2013-083 (2013)

[43] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data*, JINST **14(12)** (2019)

[44] 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(7)** (2021)

[45] 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(11)** (2018)

[46] ATLAS Collaboration, *Jet energy scale and resolution measured in proton-proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Eur. Phys. J. C **81(8)** (2021)

[47] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with* $t\bar{t}$ *events in pp collisions at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **79(11)** (2019)

[48] ATLAS Collaboration, *Measurement of b-tagging efficiency of c-jets in* $t\bar{t}$ *events using a likelihood approach with the ATLAS detector*, ATLAS-CONF-2018-001 (2018)

[49] ATLAS Collaboration, *Calibration of light-flavour b-jet mistagging rates using ATLAS proton-proton collision data at* $\sqrt{s} = 13$ *TeV*, ATLAS-CONF-2018-006 (2018)

[50] A. Buckley, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021 (2014)

[51] J. Rojo, *PDF4LHC recommendations for Run II*, PoS **DIS2016** (2016), 1606.08243

[52] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **2014(7)** (2014)

[53] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, Eur. Phys. J. C **76(4)** (2016)

[54] ATLAS Collaboration, *Measurements of the inclusive and differential production cross sections of a top-quark–antiquark pair in association with a Z boson at* $\sqrt{s}$*= 13 TeV with the ATLAS detector*, Eur. Phys. J. C **81(8)** (2021), 2103.12603

[55] F. Demartin et al., *tWH associated production at the LHC*, Eur. Phys. J. C **77(1)** (2017)

[56] ATLAS Collaboration, *Observation of the associated production of a top quark and a Z boson in pp collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, JHEP **07** (2020)

[57] ATLAS Collaboration, *Measurement of the production cross-section of a single top quark in association with a Z boson in proton–proton collisions at 13 TeV with the ATLAS detector*, Phys. Lett. B **780** (2018)

[58] CMS Collaboration, *Measurement of the associated production of a single top quark and a Z boson in pp collisions at* $\sqrt{s}$*= 13 TeV*, Phys. Lett. B **779** (2018)

[59] LHC Collaboration, *Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector* (2016)

[60] L. Rossi et al., *Progress with the high luminosity LHC project at CERN*, JACoW (2019)

[61] J. D'Hondt et al., *Learning to pinpoint effective operators at the LHC: a study of the* $t\bar{t}b\bar{b}$ *signature*, JHEP **2018(11)** (2018)

Erklärung nach §13(9) der Prüfungsordnung für den Bachelor-Studiengang Physik und den Master-Studiengang Physik an der Universität Göttingen:
Hiermit erkläre ich, dass ich diese Abschlussarbeit selbständig verfasst habe, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen wurden, als solche kenntlich gemacht habe.
Darüberhinaus erkläre ich, dass diese Abschlussarbeit nicht, auch nicht auszugsweise, im Rahmen einer nichtbestandenen Prüfung an dieser oder einer anderen Hochschule eingereicht wurde.

Göttingen, den 5. Dezember 2022

(Ireas Tom Raschke)

