GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Fakultät für
Physik

Bachelor's Thesis

Studien über die gleichzeitige Messung der assoziierten Produktion von einzelnen Top-Quarks und Top-Quark-Paaren mit einem Z-Boson in trileptonischen Endzuständen mit ATLAS

Studies of the simultaneous measurement of the associated production of single top quarks and top quark pairs with a Z boson in trileptonic final states with ATLAS

prepared by

Kia-Jüng Yang

from Wolfen

at the II. Physikalischen Institut

Thesis number:	II.Physik-UniGö-BSc-2022/06
Thesis period:	28th March 2022 until 4th July 2022
First referee:	Prof. Dr. Arnulf Quadt
Second referee:	Dr. Joern Große-Knetter

Abstract

Diese Arbeit handelt über die gleichzeitige Messung der assozierten Produktion von einzelen Top-Quarks und Top-Quark-Paaren mit einem Z-Boson. Dafür werden nur Ereignisse in Betracht gezogen, die trileptonische Endzustände haben. Des Weiteren wird in dieser Arbeit ein künstliches Neuronales Netz benutzt, um die einzelnen Top-Quark und die Top-Quark-Paare in Assoziation mit einem Z-Boson und die Untergrund Ereignisse voneinander zu trennen, sodass die Unsicherheiten der Messung verringert werden. Die gemessene Sensitivität für die Produktion von einzelnen Top-Quarks in Assoziation mit einem Z-Boson beträgt dabei $\pm 0.15\,(\text{stat.})\,^{+0.37}_{-0.16}(\text{syst.})$, wobei für die Produktion von Top-Quark-Paaren in Assoziation mit einem Z-Boson die Sensitivität $\pm 0.06\,(\text{stat.}) \pm 0.03\,(\text{syst.})$ beträgt.

Abstract

This thesis deals with the simultaneous measurement of the single top quark and top quark pair in association with a Z boson. Therefore, only events with trileptonic final states are considered. Furthermore, an artificial neural network is used, in order to separate the single top quark and top quark pair in association with a Z boson and background events from each other, so that the uncertainties are decreased. The measured sensitivity for the production of single top quarks in association with a Z boson is $\pm 0.15\,(\text{stat.})\,^{+0.37}_{-0.16}(\text{syst.})$, whereas for the top quark pairs in association with a Z boson, the sensitivity is $\pm 0.06\,(\text{stat.}) \pm 0.03\,(\text{syst.})$.

Contents

1 Introduction

Already since their childhood, people try to observe and understand, how nature and the environment around them work. Let it be just the plate we lift and then drop, just to see that it is accelerated downwards crashing into many pieces. In order to understand the nature and why it behaves as it does, investigations, about what the universe is actually made of, which properties these fundamental constituents have and how they interact with each other, are made.

Particle physics deals with these questions and investigates the fundamental particles and their interaction with each other. In order to study the properties of these particles and their interactions, scattering experiments are used. The best theory, to describe the physics of the elementary particles and their interactions, is currently the Standard Model of elementary particle physics (SM).

The elementary particles within the SM, which are of particular interest for this thesis, is the top quark and the Z boson. This is due to the fact, that the $t\bar{t}Z$ and tZq processes are sensitive to the coupling of the top quark to the Z boson. Furthermore, the coupling of the top quark with the Z boson is interesting, because the third component of the weak isospin can be extracted from this coupling. The third component of the weak isospin is one of the fundamental parameter of the SM. Therefore, the sensitivity of the $t\bar{t}Z$ and tZq processes are studied in this thesis.

However, the experimental challenge of measuring the production of $t\bar{t}Z$ and tZq is, to disentangle the $t\bar{t}Z$ and tZq processes from each other. Therefore, in this thesis a neural network is used, to separate and classify these two processes.

2 The Standard Model of Particle Physics

In today's view, the world is made up of 'elementary particles'. The Standard Model of particle physics (SM) describes the elementary particles and how they can interact through the weak, strong, and electromagnetic force with each other. Furthermore, the SM predicts finite cross sections and decay widths.

The interactions within the SM are based on the $U(1)_Y$, $SU(2)_L$ and the $SU(3)_{Colour}$ gauge groups. Here Y stands for the hyper charge and L stands for the left handedness. The $U(1)_Y \times SU(2)_L$ group describes the electroweak interaction in context of the electroweak unification [1–4]. Furthermore, the $SU(3)_{Colour}$ group describes the strong interaction in the context of QCD [5, 6].

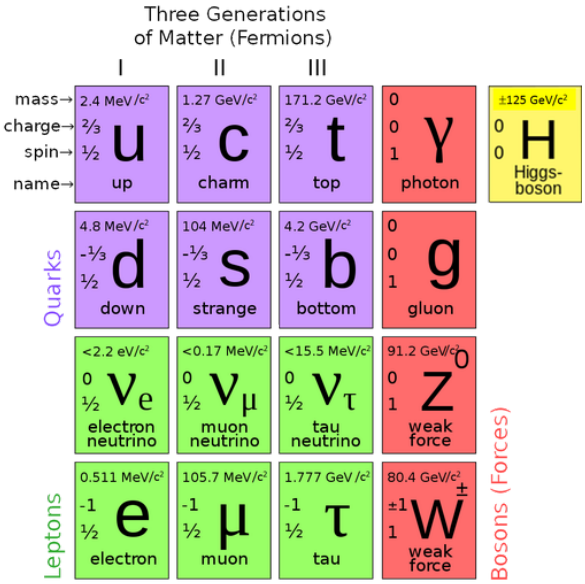

Figure 2.1: Elementary particles of the SM. The fermions are divided into two different types: leptons which are indicated in green and quarks indicated in purple. The bosons are divided into vector bosons, which are marked in red, and scalar bosons in yellow.

Moreover, these gauge groups give rise to the gauge bosons, from which then the physical bosons originate as a superposition of the gauge bosons. The physical bosons are the force carriers, which mediate the different forces between particles. $W^{\pm}$ and Z bosons mediate the weak force, photons mediate the electromagnetic force and the gluons mediate the strong force, which are represented in red in Figure 2.1. These bosons can only couple to particles with the corresponding charge. For the strong interaction this charge is the colour charge. For the electromagnetic interaction this is the electrical charge. The $W^{\pm}$ bosons can only couple to left handed particles, whereas the Z boson can couple to left and right handed particles, but with different coupling strength.

Furthermore, the twelve fermions are grouped into leptons and quarks. They are displayed in Figure 2.1 in green and red, respectively. The difference between quarks and leptons is, that leptons do not carry colour charge while quarks do. Therefore, quarks can interact with each other via the strong interaction and leptons can not.

Furthermore, the up type quarks, carrying an electrical charge of $\frac{2}{3}$ e, are called up (u), charm (c), and top (t). The down type quarks carry an electrical charge of $-\frac{1}{3}$ e and are called down (d), strange (s), and bottom (b). Moreover, the down type leptons all have an electrical charge of 1 e and are called electron (e), muon (μ) and tau (τ), whereas the up type leptons carry no electrical charge and are called electron neutrino (ν_e), muon neutrino (ν_μ) and tau neutrino (ν_τ).

Lepton pairs as well as quark pairs, which are ordered in one column regarding Figure 2.1 belong to one generation. Within the generations, the lepton and quark pairs are represented in their left handed weak isospin doublet. The third component of the weak isospin is $I_3 = \frac{1}{2}$ for up-type quarks and $I_3 = -\frac{1}{2}$ for down-type quarks. Moreover, while all charged fermions can exist as right-handed singlets too, experiments [7] have shown that neutrinos only exist as left-handed particles. Furthermore, the mass of the particles increases with the generation.

Without the mechanism of electroweak symmetry breaking, all SM particles should be massless. Nevertheless, non-zero masses were measured for all charged fermions and the Z and W bosons. This could be solved, by introducing the Higgs mechanism [8–12]. Therefore, massive particles are possible through the process of electroweak symmetry breaking. Furthermore, the higgs mechanism gives rise to the physical Higgs boson, which is the only scalar boson within the SM with spin 0.

3 Top Quark

The top quark was discovered in 1995 by the DØ and CDF experiments [13, 14]. It is the heaviest elementary particle within the SM. The top quark decays nearly in 100 % of the cases to a W boson and a b-quark, due to the suppression of the CKM matrix for other decays [15, 16]. Therefore, the top decay is characterised by the W decay. The W can decay leptonically $W \to \ell\nu_\ell$ or hadronically $W \to q\bar{q}$. The production of a top quark pair or a single top with an additional Z boson is shown in the following.

3.1 $t\bar{t}Z$ and tZq Production

The most important properties of the top quark pair production in association with a Z boson are explained. The leading order Feynman diagrams are shown in Figure 3.1. If the Z boson decays into a lepton pair (here electron and muon), then the decay channel of $t\bar{t}Z$ can be classified into three channels.
The dilepton channel for the $t\bar{t}$ decay is characterised by two b jets, four leptons with opposite signs and missing transverse energy, due to the two neutrinos.
The single lepton channel for the $t\bar{t}$ decay is also characterised by two b jets, two further quark jets, where the sum of the charges is equal to the charge of the W boson, three leptons and missing transverse energy, due to the one neutrino.
The last channel for the $t\bar{t}$ decay is the all-hadronic channel, which is also characterised by two b jets, four quark jets and two leptons with opposite signs.

Moreover, the most important properties of the tZq production are explained. The leading order Feynman diagrams for the tZq production are shown in Figure 3.2. Assuming that the Z boson decays into a lepton pair, then the tZq decay can be classified into two different channels.
One channel is called the single lepton channel, which is characterised by one b jet, one further quark jet, three leptons and missing transverse energy due to the neutrino.
Then there is the all hadronic channel, characterised by one b jet, three quark jets and two leptons with opposite signs.

Furthermore in the $t\bar{t}Z$ and tZq decay channel atleast one lepton pair with opposite sign and same flavour is included, due to the Z boson decaying into two leptons.

In Figure 3.2 on the bottom right and in Figure 3.1 on the right side and the bottom left side, it can be seen, that $t\bar{t}Z$ as well as tZq are sensitive to the weak coupling of the Z boson to the top quark, as the coupling of the Z boson is proportional to $\propto v_f - a_f\gamma^5$, with $v_f = I_{3f} - 2Q_f \sin^2(\theta_W)$ and $a_f = I_{3f}$. Here f is the fermion type, Q_f is the electrical charge of the fermion f, θ_W the Weinberg angle and I_{3f} is the third component of the weak isospin. Therefore, by measuring the production cross section of $t\bar{t}Z$ and tZq, one can probe the third component of the weak isospin of the top quark. The measurement of this parameter is of particular interest, because it is a fundamental parameter of the SM.

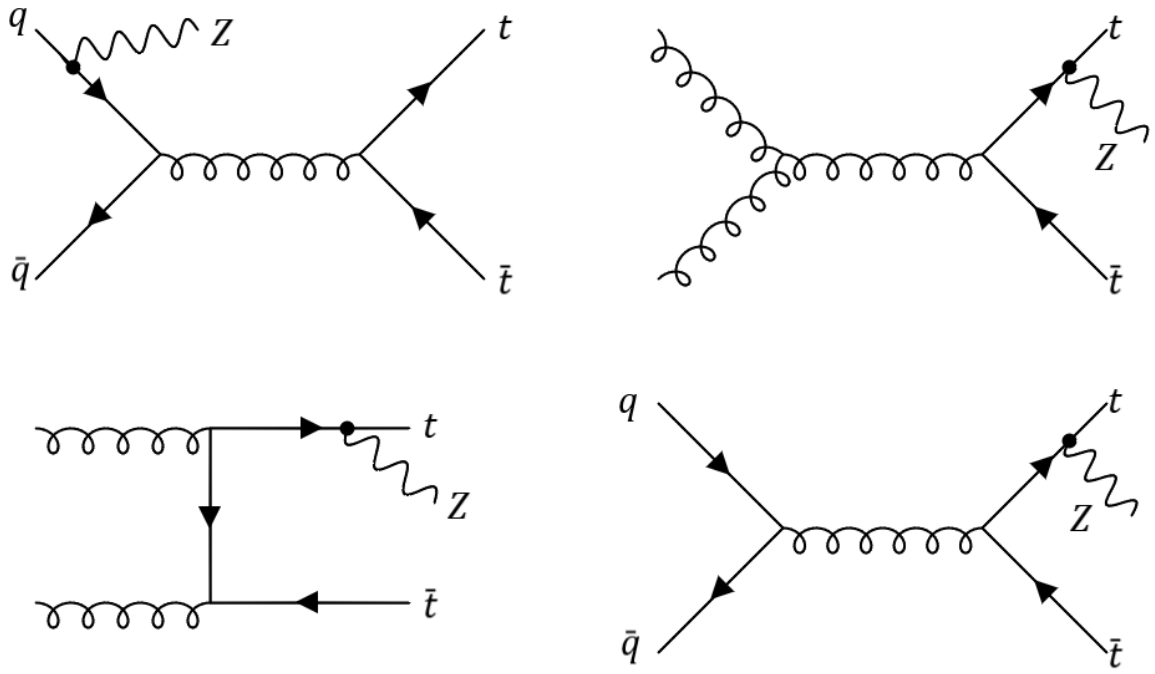

Figure 3.1: Leading order Feynman diagrams for the $t\bar{t}Z$ production. The diagram on the upper left belongs to the initial state radiation and the others to the final state radiation. Only the final state radiations are sensitive to the top - Z coupling.

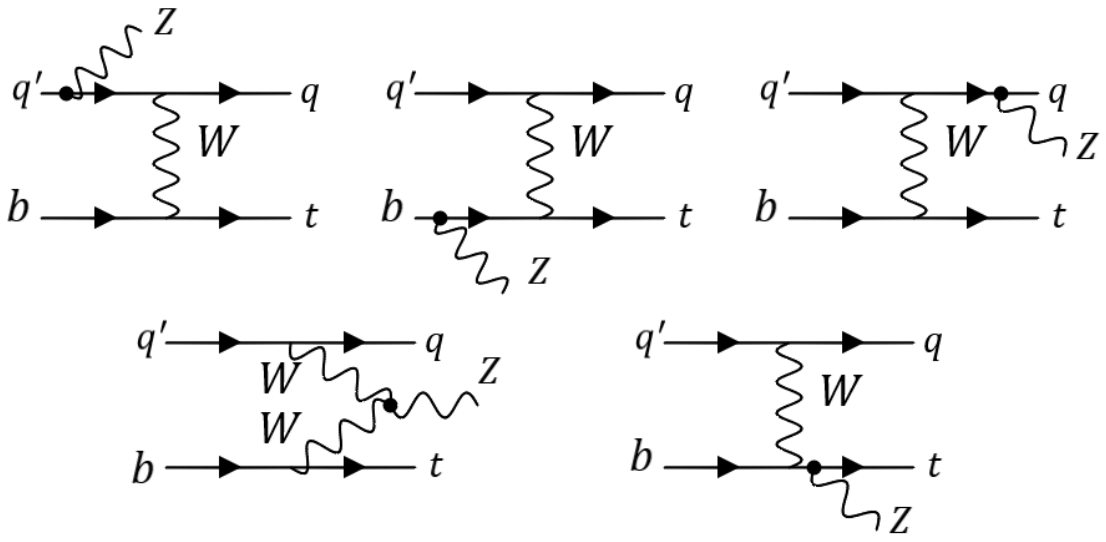

Figure 3.2: Leading order Feynman diagrams for the tZq production. Only the bottom right process is sensitive to the top - Z coupling

4 The ATLAS Experiment

The LHC is a two ring superconducting hadron accelerator and collider with a centre of mass energy of $\sqrt{s} = 13\,\text{TeV}$ for the pp collisions with a circumference of 26.7 km. Furthermore, ATLAS [17] is one of the experiments at the LHC. A cut-away view is shown in Figure 4.1.

The Inner Detector (ID) is immersed in a 2 T magnetic field, which is generated by the superconducting solenoid surrounding the ID cavity. In the ID the trajectories of the particles can be measured through the semiconductor pixels, strip detectors and the use of straw detectors, which cover the region of $|\eta| < 2.5$, with $\eta = -\ln\left(\tan\frac{\Theta}{2}\right)$. Here the polar angle Θ describes the angle from the beam axis. Due to the magnetic field the charged particles trajectories are bent. By measuring the curvature, the transverse momentum p_T is determined. Furthermore, the tracks are used to reconstruct primary and secondary vertices, which are then used for b-tagging.

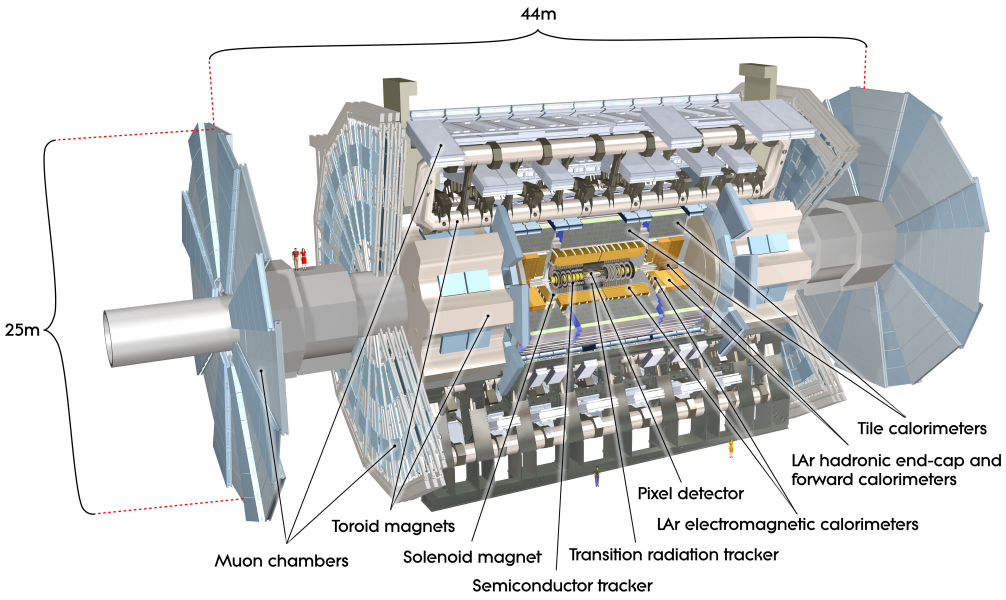

Figure 4.1: Cut-away view of the ATLAS detector, © 2022 CERN.

The calorimeters are surrounded by three large superconducting toroids, arranged in an eight fold azimuthal symmetry. The electromagnetic and the hadronic calorimeters of ATLAS are sampling calorimeters.

The electromagnetic calorimeter is a lead liquid argon (LAr) calorimeter. The electromagnetic calorimeter covers the region $1.375 < |\eta| < 2.5$ and $2.5 < |\eta| < 3.5$. It is used to measure mainly the energy of electrons and photons, by measuring the energy loss fragmentation at each active layer of the sampling calorimeter.

The hadronic calorimeter consists of a LAr sampling calorimeter, which is surrounded by layers of steel and plastic scintillating tiles. The hadronic calorimeter covers the region $|\eta| < 4.9$. It is used to measure mainly the energy of the hadrons, analogue to the energy measurement of the electrons and photons.

The muon detector is based on the magnetic deflection of muon tracks in the large superconducting air-core-toroid magnets with tracking chambers. The muon detector covers the region $|\eta| < 2.7$. Therefore, the trajectory and also its momentum is measured. Since the muon is a minimal ionising particle, it can not be stopped, neither by the electromagnetic, nor by the hadronic calorimeter. Therefore, the muon chamber is needed and located at the outermost layer, to improve the muon identification.

5 Object Reconstruction and Signal Region

In order to analyse the physics collision data, the kinematic reconstruction of particles is necessary, because the final states characterises these processes. Furthermore, the reconstruction of the kinematic properties of the particles is necessary to extract distributions, which are sensitive to the previously mentioned SM parameters. In this chapter, the object reconstruction is discussed.

The primary vertex is identified as the point, whose sum of associated tracks has the highest transverse momenta and has in total a $p_{\mathrm{T}} > 400\,\mathrm{MeV}$ [18]. The primary vertex is used for the reconstruction of the leptons or the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$.

Electrons are reconstructed by matching the energy measured in the EM calorimeter with the reconstructed tracks from the inner detector. Moreover, they have to be within the regions $0 < |\eta| < 1.37$ and $1.57 < |\eta| < 2.47$, satisfying the condition of $E_{\mathrm{T}} > 20\,\mathrm{GeV}$ and being isolated. At last a likelihood-based discriminant, using a set of variables, is constructed. This discriminant is used to improve the electron identification, which is further described in [19].

Muons are reconstructed by matching the track from the inner detector with that from the muon detector. Furthermore, they have to be in the region $|\eta| < 2.5$ and have to have $p_T > 20\,\mathrm{GeV}$. In order to decrease misidentification rates of muons, requirements described in [20] are applied.

Jets are reconstructed by using the anti-$\mathrm{k_T}$ algorithm with a cone width of 0.4 [21]. Moreover, the reconstructed jets have to satisfy the condition $p_{\mathrm{T}} > 35\,\mathrm{GeV}$ and have to be in the region $|\eta| < 4.5$. In order to decrease the jets, which come from pileup effects, a likelihood-based discriminant called jet vertex tagger is introduced, on which a further selection criterion is set if the jet has $p_{\mathrm{T}} < 120\,\mathrm{GeV}$ and $|\eta| < 2.5$ [22]. For the identifica-

tion of b-jets, a multivariate algorithm called DL1r b-tagging algorithm is used [23].

The missing transverse energy is reconstructed as the negative vector sum of the transverse momenta of all reconstructed objects [24].

In the following, the production of tZq and $t\bar{t}Z$ is analysed. Therefore a signal region is constructed, in order to filter the background processes, while at the same time, trying to preserve as many signal processes as possible. Therefore the channel, where the Z boson decays into two leptons and exactly one W boson, which originated from a top decay and decays leptonically, is taken into consideration. Hence in total three leptons (trileptonic) are expected. Moreover, further cuts are applied, regarding the reconstructed objects, in order to construct the regions with purer $t\bar{t}Z$ and tZq contribution. This region is defined as the SR-3ℓ region, whose event selections are shown in Table 5.1.

Here m_{ll} is the minimal invariant mass of a lepton pair and `hasOSSFPair` describes, if there is a lepton pair with opposite charge and same flavor. Moreover, tight leptons refers to the reconstructed leptons, but with tight conditions on the likelihood based discriminant [25]. The event yields for the SR-3ℓ region is shown in Table 5.2.

Table 5.1: Requirements for the trileptonic signal region.

Selection requirement	SR-3ℓ
Lepton & tight lepton multiplicity	$= 3$
Z boson candidates multiplicity	$= 1$
Jet multiplicity	≥ 2
|mz1 - m_Z|	$\leq 10\,\text{GeV}$
`hasOSSFPair`	$\surd$
minimal m_{ll}	$> 12\,\text{GeV}$
Leading lepton p_T	$> 25\,\text{GeV}$
Sub-leading lepton p_T	$> 15\,\text{GeV}$
Sub-Sub-leading lepton p_T	$> 10\,\text{GeV}$

Table 5.2: Event yields for the tZq, $t\bar{t}Z$ and background region.

Event	background region
tZq	243.6 ± 1.2
tWZ	74.0 ± 0.2
$t\bar{t}Z$	782.1 ± 1.8
$t\bar{t}(X)$	25.6 ± 0.4
WZ_l	350.8 ± 2.1
WZ_c	490.2 ± 2.3
WZ_b	240.3 ± 1.4
$VV(V)$	175.0 ± 2.1
NPL	476.2 ± 17.5
Total	2857.6 ± 18.1

6 Deep Neural Network

In this thesis, the sensitivity of the $t\bar{t}Z$ and tZq production are measured. This is done by a simultaneous $t\bar{t}Z$ and tZq binned likelihood fit. In order to decrease the uncertainties of the fitted parameter, $t\bar{t}Z$, tZq, and background events need to be separated. Therefore, a multi class deep neural network is used.

A deep neural network (DNN) consists of many layers, represented as one column of circles in figure 6.1. The circles are called neurons or nodes. The first layer is called the input layer, where one node represents one input parameter. Therefore the number of nodes at the input layer correspond to the number of input parameters. Furthermore, these input parameters represent the properties of the objects, which are used for classification. Moreover, the nodes of the last layer represent the output and are therefore called output layer. As the number of nodes in the output layer depends on the number of outputs, for a binary output, only one node is needed. If there are more than two outputs, the number of outputs will be equal to the number of nodes. Moreover, the layers between the input and output layer are called hidden layers. A sketch of the DNN is shown in Figure 6.1.

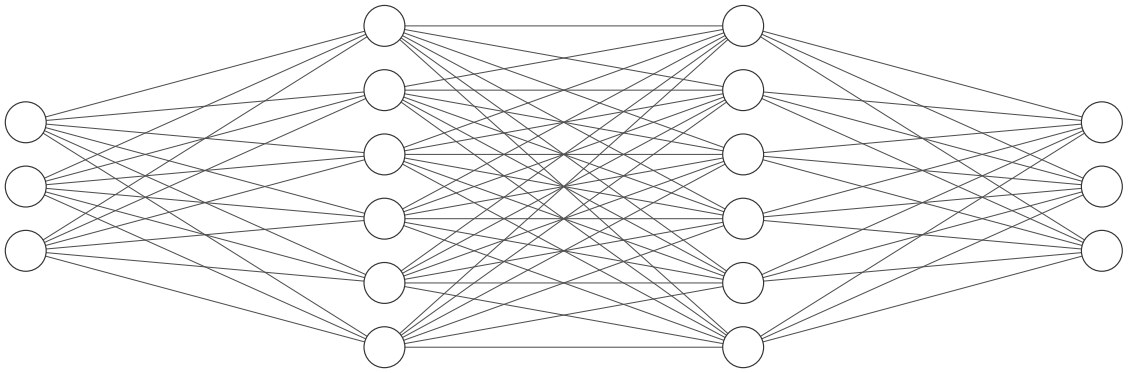

Figure 6.1: Sketch of a deep neural network with 3 nodes as the input and output layer and 6 nodes in each of the two hidden layers.

Every node of each layer takes the value of each node of the previous layer, which can be represented as a vector $\vec{x}$. Then for each node, $\vec{x}$ is multiplied component wise by the weight vector $\vec{w}$. Each component of the weight vector describes the connection to the respective node of the previous layers. Then the weighted vector $\vec{x}_{weight}$ needs to be transformed into a scalar with the transformation $T : \mathbb{R}^n \to \mathbb{R}$, $\vec{x}_{weight} \mapsto \vec{x} \cdot \vec{w}$, which is then used as an input variable for a function $f : \mathbb{R} \to \mathbb{R}$, $\vec{x} \mapsto f(\vec{x} \cdot \vec{w} + b)$. Here n is the number of nodes in the previous layer and b is the bias which represents an offset. Moreover, f is an activation function, which determines how the weighted sum of $\vec{x}$ with a bias is transformed into the node value. This is how the input values are propagated through the layers, until the output layer is reached. The values of the nodes of the output layer are not weighted, and take directly the vector $\vec{x}$, which represents the previous layer and uses this as input for its activation function $f : \mathbb{R}^n \to \mathbb{R}$, $\vec{x} \mapsto f(\vec{x})$.

In order to quantify the "goodness" of the DNN, a loss function is introduced. Using this loss function, the DNN can then be trained, by adjusting the weights and biases from the former layers, so that the loss function is optimised. These adjustments of the weights and biases are done by the method called backpropagation [26]. The value of the loss function therefore represents, how reliable the DNN classifies the objects, based on their input parameters.

Further, the DNN also needs data on which it can train and learn how the input values of the different objects differ, in order to classify them. Nevertheless, a problem of DNNs can be, that they overfit the data, which means, that they have a strong predictive power on the data the DNN trained on, but perform worse on any other data. In order to use the whole data, to train the DNN and not just a sub set, the k-fold method is introduced. Therefore, the whole data set is divided into k equal and statistically independent subsets, where only k-1 sets are identified as the training set and the last one is identified as the test set. From the training set, p percent are set aside and used as the validation set. Therefore, only $1 - p$ percent of the training set is used for the training of the DNN in each fold. Further, not the whole training set is passed to the DNN at once, but rather small shares of the training set, called batches. Passing all batches once through the DNN is called one epoch. During these epochs, the DNN adjusts the weights and biases, where the learning rate represents the amount of changes of the weights and biases in each epoch. Then the validation set is used, to validate the performance of the DNN for each epoch.

Except for the division of the data into validation and training sets, another method called early stopping is used, in order to prevent overfitting. The idea behind early stopping is, if the loss starts to only change by a small amount, which is smaller than Δ_{min}, then the DNN stops the training and moves to the next fold. However, the DNN does not stop immediately after the change of the learning rate is smaller than Δ_{min}, but rather if it happens for a certain number of epochs called patience. This prevents that the DNN adapts the weights and biases to the small statistical fluctuations of the training data sets, which would make the generalisation of the DNN worse.

In order to detect overfitting, calculating the loss of the DNN for the validation and the test set is needed. If the validation accuracy and loss are far worse than that of the test set, the DNN has been overfitted.

6.1 $t\bar{t}Z$-tZq-Background Multi-Class Classifier

In this analysis of the $t\bar{t}Z$ and tZq production, a multi-class DNN is used, in order to separate the classes $t\bar{t}Z$, tZq and the background from each other. A DNN with 4 hidden layers with 20, 30, 30, 20 nodes is chosen. The activation functions for the hidden layers are chosen to be ReLU, which is defined as

$$f(x) = \begin{cases} 0 & x < 0 \\ x & x \geq 0 \end{cases},$$

whereas the activation function of the output layer is chosen to be the softmax function, which is defined as

$$\sigma(\vec{x})_i = \frac{\exp(x_i)}{\sum_{j=1}^{K} \exp(x_j)}.$$

Further, $\vec{x}$ represents the values of the nodes in the previous layer, $\sigma(\vec{x})_i$ is the i-th value of the output node and K is the number of nodes in the previous layer. Due to the normalisation, the output nodes have values between zero and one, which adds up to one. Moreover, the output value of the DNN can be interpreted as probability, regarding how likely it is, that the input object belongs to one of the classes. Furthermore the loss function g is chosen to be the categorical cross-entropy, which is defined as follows:

$$g = -\sum_{i=1}^{n} t_i \cdot \log(y_i).$$

Assume, that the j-th output node represents the actual class of the object, which is classified by the DNN. Then n is the number of classes, t_i is the i-th element of the vector t, which represents the actual class, by setting one as value at the j-th spot and anywhere else zero. Furthermore y_i represents the value of the i-th output.

The DNN uses Nadam [27–29] as optimiser, has a learning rate of 0.0005 and uses as metric accuracy. Furthermore the number of folds is chosen to be 3, the validation size is chosen to be 25 %, the batch size is chosen to be 5000 and the epoch number is chosen to be 2000. For this DNN, also early stopping is implemented. Therefore the patience is chosen to be 100 and Δ_{min} is chosen to be 0.0005. The data, used for the DNN are all within the SR-3ℓ region and the input variables used to discriminate between the processes $t\bar{t}Z$, tZq and the background is shown in Table 6.1.

Table 6.1: Discriminating variables for the DNN for separation of the process $t\bar{t}Z$, tZq and background. NPLS correspond to events with at least one non-prompt lepton.

Variables	Reason
trailing lepton p_{T}	NPLS vs Signal
trailing jet p_{T}	ttZ/tWZ/ttX vs all other
$p_{\mathrm{T}}^{\mathrm{miss}}$	VVV/NPLS vs all other
central jets with $p_{\mathrm{T}} > 35$ multiplicity	tZq vs all other
forward jets with $p_{\mathrm{T}} > 35$ multiplicity	tZq vs all other
jet multiplicity	NPLS vs ttZ
η of leading forward jets	tZq/ttZ vs all other
η of leading central jets	ttZ/ttX/tWZ vs all other
b-jet multiplicity at working point of 60/70/77/85	WZ vs ttZ/tZq
trileptonic invariant mass	NPLS/ttX vs all other
b-Tag working point of (sub-)leading jet	ttZ/tZq/ttX/WZ+b vs all other
invariant transverse mass of the W boson	tZq/Wz+l vs all other

Here the trailing lepton/jet p_{T} refers to the three highest transverse momenta of the leptons/jets. Furthermore the b-tag working point refers to, how tight the requirement are, so that it has an efficiency 60/70/77/85 %. The output for each class and its separation of the other classes is shown in Figure 6.2.

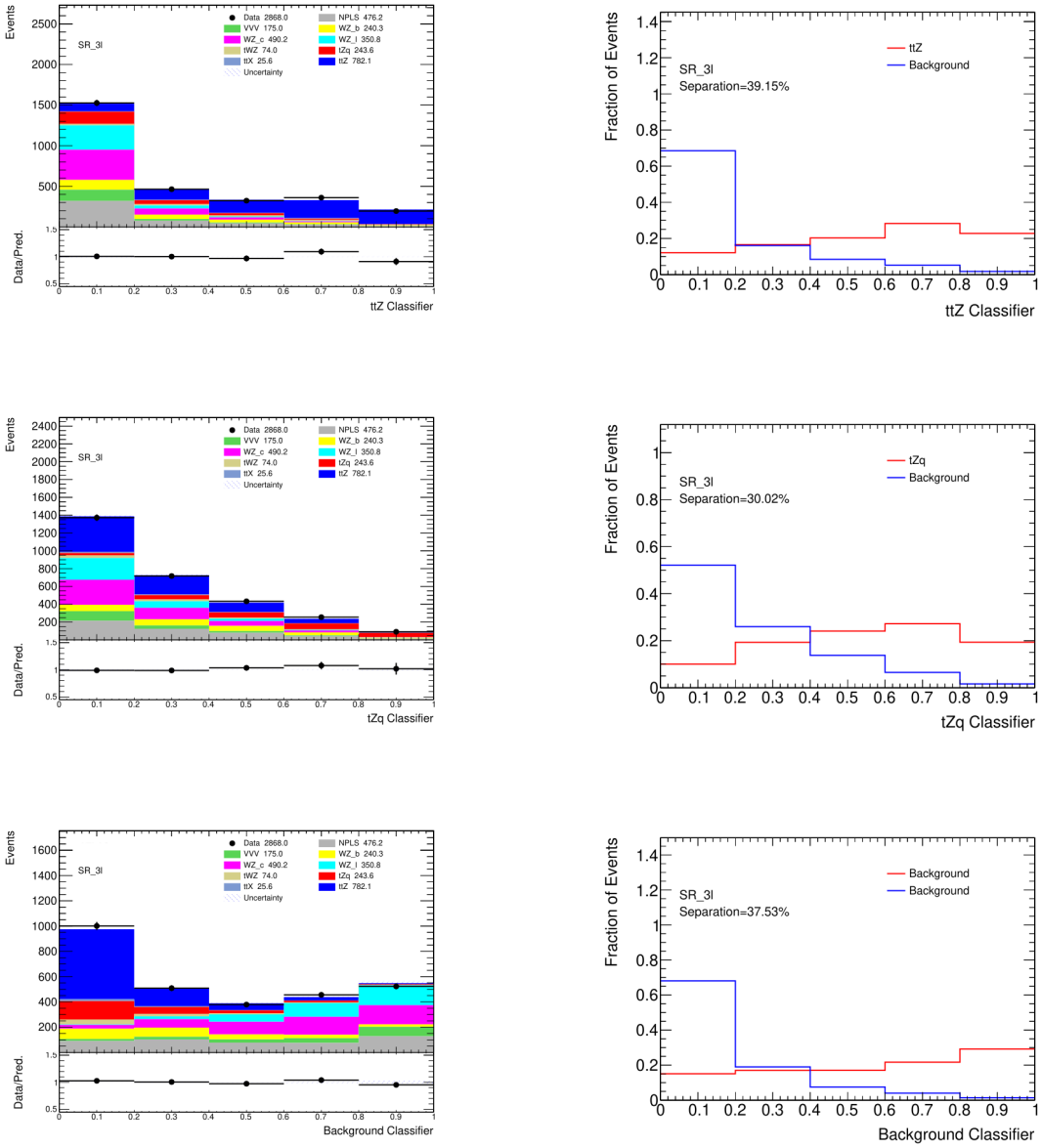

Figure 6.2: Plots for the DNN classification and the respective separation plots. The upper plot is the class $t\bar{t}Z$, in the middle the class is tZq and on the bottom is the class background. The class background shown as the red curve in the bottom plot corresponds to every background events, excluding tZq and $t\bar{t}Z$ events.

The DNN was able to separate the $t\bar{t}Z$ processes the best, then the background processes and the worst separation was achieved regarding the tZq processes. This can be seen in Figure 6.2 with their respective separation plots. In order to evaluate the classification performance, the confusion matrix is introduced, shown in Figure 6.3

The diagonal elements of Figure 6.3 shows the probability, with which the background, $t\bar{t}Z$ and tZq is classified correctly. Based on the Figure 6.3, the $t\bar{t}Z$ classification is the best, then the tZq classification and the worse performance regarding the classification were the background processes. Furthermore, the loss for each epoch is shown in Figure 6.4

As mentioned in Chapter 6, a problem of DNNs is, that they could overfit the data. Therefore it needs to be ensured, that the DNN, used in this analysis did not overfit the data. In order to do so, it needs to be checked, whether the loss performance on the validation set and the training set have high deviations. The loss for each epoch is shown in Figure 6.4. It can be seen, that in neither of the three folds, the loss of the training set has high deviations from the validations set. However, at the tail of the loss plot, the validation loss and the train loss starts to deviate. Nevertheless, due to early stopping, the DNN training was stopped.

It can be concluded, that a slight indication of overfitting can be observed, However, due to early stopping, this could be prevented, as the deviation of the loss is still very small.

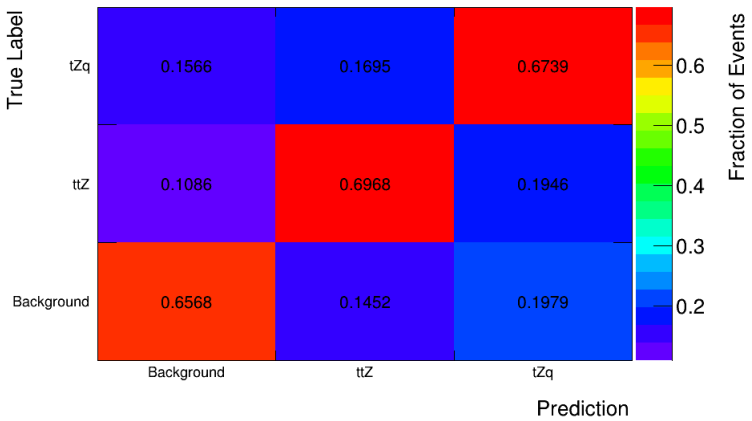

Figure 6.3: Confusion matrix for the DNN with the classes $t\bar{t}Z$, tZq and background.

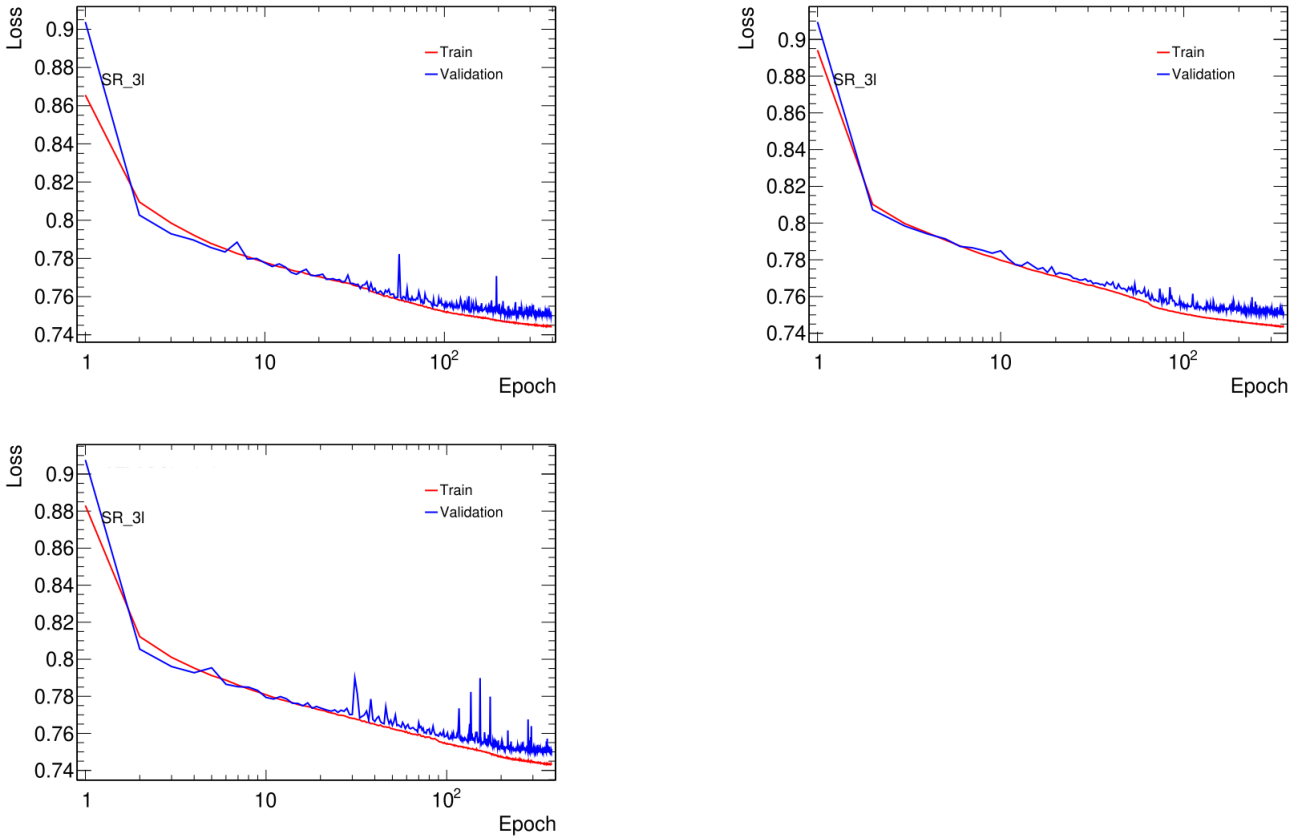

Figure 6.4: Accuracy and loss for each epoch. The upper left plot belongs to the first fold, then the upper right belongs to the second fold and the plot on the bottom belongs to the third fold.

7 Binned Likelihood Fit

As mentioned in Chapter 6, the binned likelihood fit is performed, to extract the signal strength of the processes tZq and $t\bar{t}Z$. The signal strength is defined as

$$\mu = \frac{\sigma_{exp.}}{\sigma_{SM}}.$$

In that equation, $\sigma_{Exp.}$ is the experimental production cross section of a certain process and σ_{SM} is the theoretically predicted production cross section of the same process. Therefore the experimental cross section can be extracted from the signal strength, which on the other hand can be extracted through a binned likelihood fit.

For the binned likelihood, assume N measurements, where each measurement outcome belongs to one of in total k bins. The measured observable is $\vec{n}$, where $\vec{n}$ contains the number of outcomes within the bin j with $j \leq k$. In the frame of this thesis, the measured data are resulting from a counting experiment, where every event is independent from any other. The likelihood for the binned observables, in the case that the measurements are independent, is then

$$L = \prod_{j=1}^{k} f_j(n_j|\theta).$$

Here the f_j represents the conditional probability, that n_j of the outcome has a value, which is within the j-th bin, under the condition that the parameters are chosen to be θ. Furthermore $f_j(n_j|\theta)$ follows a Poisson distribution, so that the likelihood is then

$$L = \prod_{j=1}^{k} \frac{\mu_j^{n_j}(\theta)}{n_j!} \exp\left(-\mu_j(\theta)\right).$$

μ_j is the expectation value for the number of events within the j-th bin and is defined as

$$\mu_j = N \cdot \int_{\{n_j\}} p(x|\theta)dx.$$

$p(x|\theta)$ describes in this case the probability, that the measurement outcome is x, under

the condition, that the parameters are θ. Moreover, in the binned likelihood fit, the likelihood is maximised, by optimising the parameters θ.

7.1 Definition of Binned Likelihood Fit Regions

In order to decrease the uncertainties in a simultaneous binned likelihood fit with the signal strength of tZq and $t\bar{t}Z$ as parameter of interest (POI) fit parameters, histograms enriched only in $t\bar{t}Z$ and tZq are needed. Furthermore, variables with high separation power are needed, in order to decrease the uncertainties.

Therefore the chosen variables are the output values of the DNN. The three orthogonal regions, one for each class of the DNN, are constructed. The cuts on the output value for the tZq region are $tZq \geq 0.4$ and background < 0.4. For the $t\bar{t}Z$ region the cuts are $tZq < 0.4$ and background < 0.4. The last region is the background, where the cut is only set on the background output with background < 0.4.

Furthermore, one more region is constructed, but without the usage of the outputs of the DNN. This region is not orthogonal to the upper three regions, as the histogram for this region is used for another independent binned likelihood fit. The fit result is then used to compare the performance of the method with DNN and the method, were only variables are used, which are already in the ntuples, called cut and count method. As the central jet multiplicity is the chosen variable, the region is called the nCJets region. For the nCJets region, the cuts of the SR-3ℓ region is implemented. Moreover, a cut on the b-tag working point is set to be equal or bigger than 77. The event yields for these four regions are shown in Table 7.1.

Table 7.1: Event yields for the tZq, $t\bar{t}Z$ and background region.

Event	background region	$t\bar{t}Z$ region	tZq region	nCJets region
tZq	43.3 ± 0.6	40.2 ± 0.6	160.1 ± 1.0	209.9 ± 1.1
tWZ	16.5 ± 0.1	38.5 ± 0.2	19.0 ± 0.1	61.6 ± 0.2
$t\bar{t}Z$	89.0 ± 0.6	533.5 ± 1.5	159.5 ± 0.8	691.0 ± 1.7
$t\bar{t}(X)$	3.9 ± 0.2	16.4 ± 0.3	5.2 ± 0.2	22.8 ± 0.4
WZ_l	324.1 ± 2.1	10.4 ± 0.3	16.2 ± 0.4	22.5 ± 0.6
WZ_c	393.6 ± 2.1	42.9 ± 0.6	53.8 ± 0.8	128.4 ± 1.3
WZ_b	86.1 ± 1.0	68.1 ± 0.6	86.1 ± 0.8	195.8 ± 1.2
$VV(V)$	138.0 ± 2.0	15.1 ± 0.3	21.8 ± 0.7	56.0 ± 1.1
NPL	282.0 ± 17.3	82.9 ± 1.8	111.2 ± 2.5	329.6 ± 14.2

8 Systematic Uncertainties

Before performing the fit, systematic uncertainties need to be taken into account by implementing these as nuisance parameter (NP) for the likelihood. The systematic uncertainties arise from an imperfect knowledge of the detector and the underlying theoretical modelling of physics processes These systematic uncertainties, which are included, are categorised in the following.

One systematic uncertainty category is the theoretical uncertainties. The systematic uncertainties included are the choice of renormalisation and factorisation scale, parton distribution functions (PDF) and the modelling uncertainties. These uncertainties arise, due to the fact, that the scaling factor or the showers are just approximations to what is observed. That is why Monte Carlo samples are taken into account, which are generated with different scaling factors or alternative shower algorithms. Here the default shower algorithm is Pythia8 and the alternative shower algorithm is Herwig7. Furthermore, the PDFs can only be measured experimentally. As these PDFs are needed and included within the theoretical calculations, this leads to further theoretical uncertainties.

Additionally there are the object reconstruction uncertainties. The systematics, which were included in this category, are the lepton identification, trigger and isolation. Moreover, the b-tagging uncertainties for the jets are also included, but no further jet systematics uncertainties are included in this analysis. These systematic uncertainties arise, due to the fact, that the resolution of the detector for the energy or the reconstruction of the track of any object is not perfect.

Moreover, there is the integrated luminosity uncertainty. The whole uncertainty for the integrated luminosity for 2015-2018 is 1.7 % [30].

The last uncertainties that have been included, are the pile up uncertainties. These are taken into consideration, by varying the reweighing of the Monte Carlo simulations.

9 Results

9.1 The DNN Method

For the DNN method, a DNN was trained, in order to classify the events into three classes. One class were the $t\bar{t}Z$ processes, the second were the tZq processes and the third were the background processes. After that, the $t\bar{t}Z$, tZq and the background region were constructed, by setting further cuts on the DNN output value within the SR-3ℓ region. For each of these regions, one histogram with systematic uncertainties is created, which can be seen in Figure 9.1. Furthermore, the pull plots for the systematic uncertainties for the three histograms is shown in Figure 9.2 and 9.3. The pull plots show, how much the NPs have changed after the fit. In the pull plots, created in this thesis, the change is zero, due to the fact, that these are Asimov fits. Asimov fits describe a fit, which is performed not on real data, but on simulated once. Therefore the MC samples are fitted on itself in the frame of an Asimov fit. Furthermore the pull plots show the uncertainties for the NPs, represented by the black bars.

On the histograms, shown in Figure 9.1, an Asimov fit is performed, where the Monte Carlo simulation is used as pseudo data. The resulting signal strengths are $\mu_{t\bar{t}Z} = 1 \pm 0.06\,(\text{stat.}) \pm 0.03\,(\text{syst.})$ and $\mu_{tZq} = 1 \pm 0.15\,(\text{stat.})\,^{+0.37}_{-0.16}(\text{syst.})$.

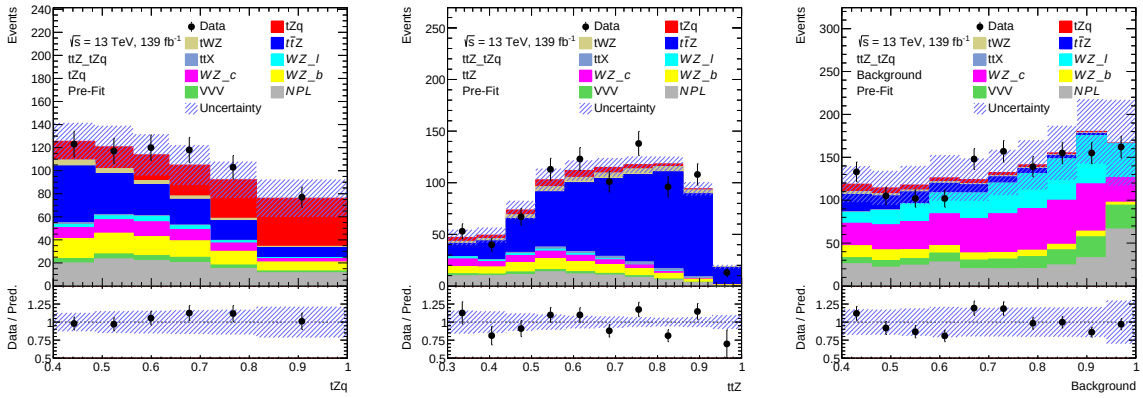

Figure 9.1: Distributions for the tZq, $t\bar{t}Z$ and background region. On the left, the tZq, in the middle the $t\bar{t}Z$ and on the right the background region is shown. The horizontal is chosen to be the output of the DNN for the corresponding class.

9.2 The Cut and Count Method

The second method, which is used to investigate the sensitivity for the production of $t\bar{t}Z$ and tZq, is the cut and count method. This method is mainly used, in order to compare, if a DNN really improves the sensitivity, or if the cut and count method is just as good. For the cut and count method, a fourth region, called the nCJets region was constructed. This region uses only cuts on variables. The chosen variable is the multiplicity of the central jets, as this gives the best separation between $t\bar{t}Z$ and tZq from all investigated variables. The histogram is shown in Figure 9.4. Furthermore, the pull plots for the systematic uncertainties for the nCJets histogram is shown in Figure 9.5 and 9.6.

On the distribution, shown in Figure 9.4, an asimov fit is performed, where the Monte Carlo simulation is used as pseudo data. The resulting signal strengths are $\mu_{t\bar{t}Z} = 1 \pm 0.08\,(\text{stat.}) \pm 0.02\,(\text{syst.})$ and $\mu_{tZq} = 1 \pm 0.25\,(\text{stat.}) \pm 0.48\,(\text{syst.})$.

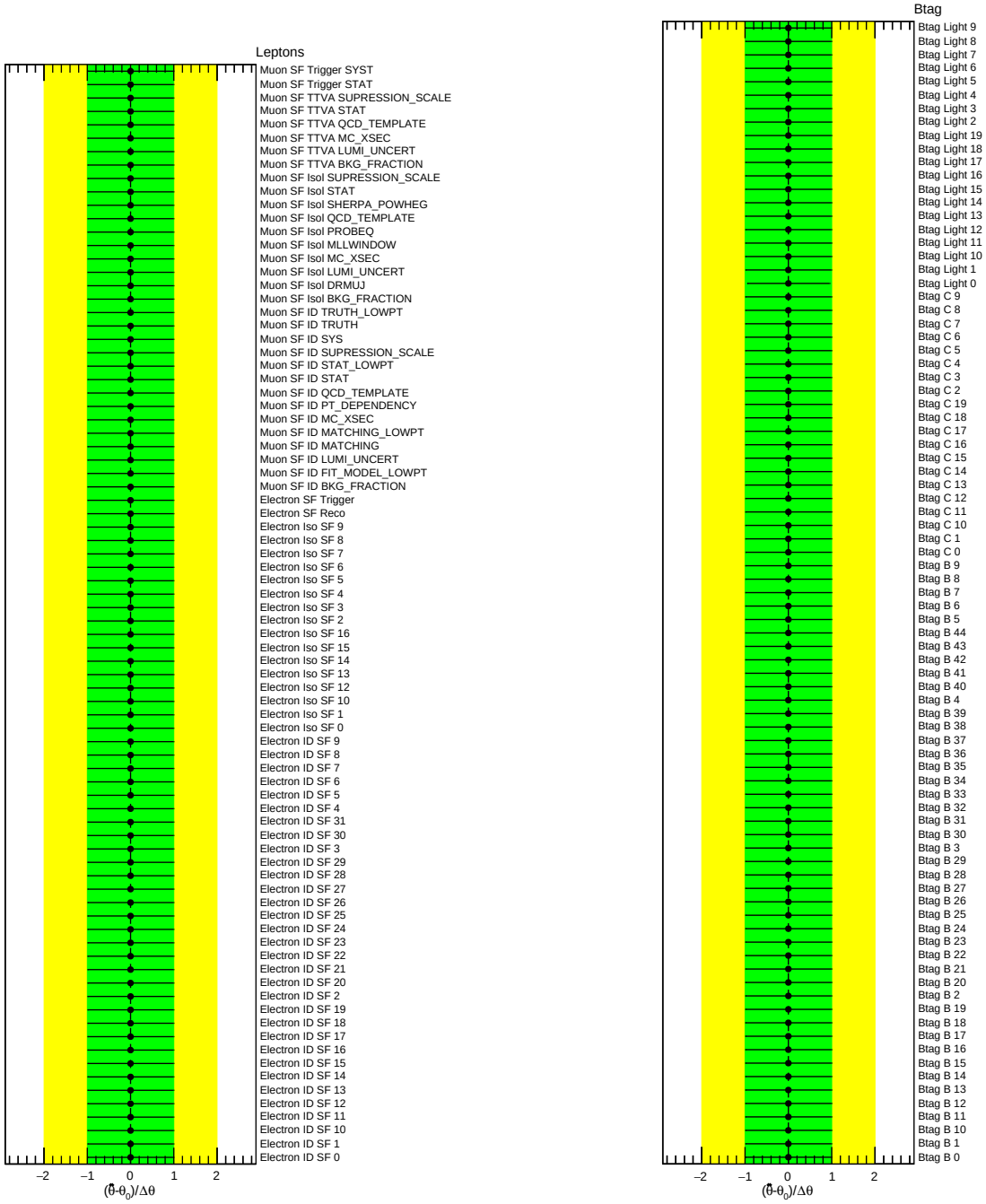

Figure 9.2: Pull plots for the lepton and b-tag systematic uncertainties.

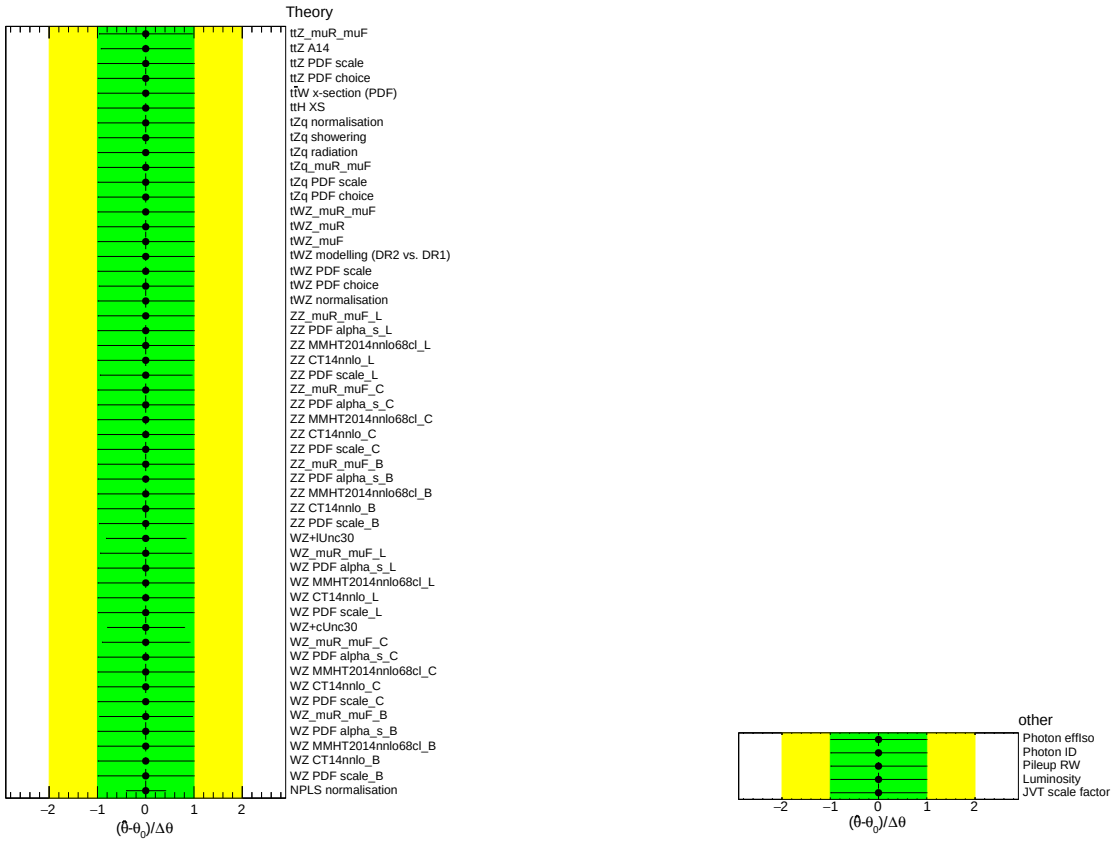

Figure 9.3: Pull plots for the theory, photon, integrated luminosity, jet vertex tagger and pile up uncertainties.

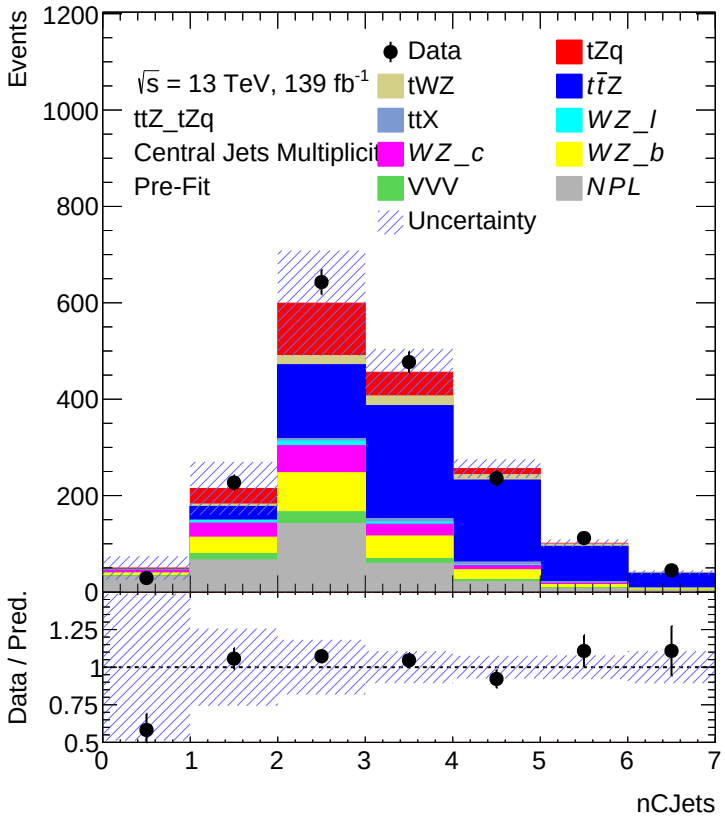

Figure 9.4: Distribution for the nCJets region.

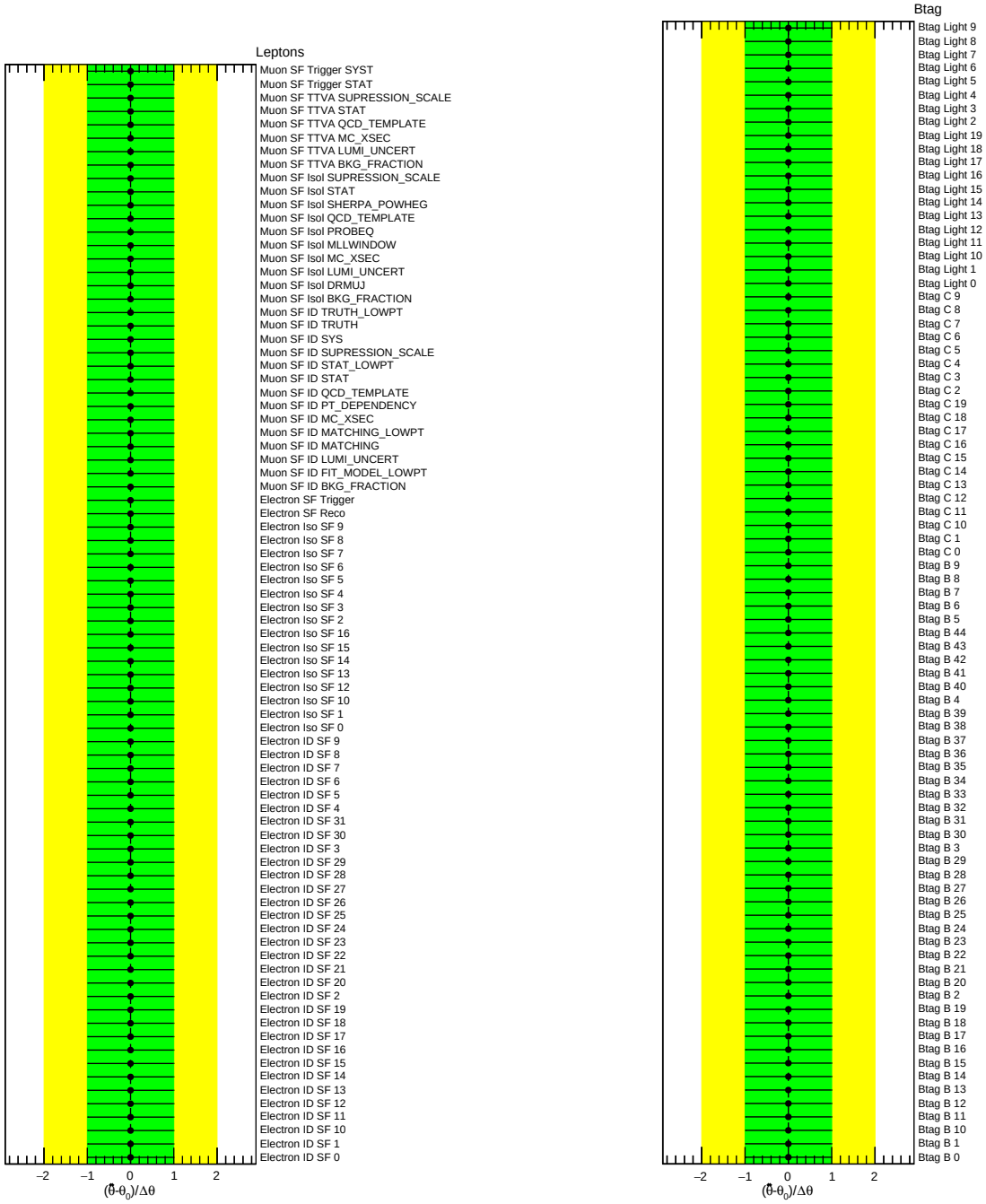

Figure 9.5: Pull plots for the lepton and b-tag systematic uncertainties.

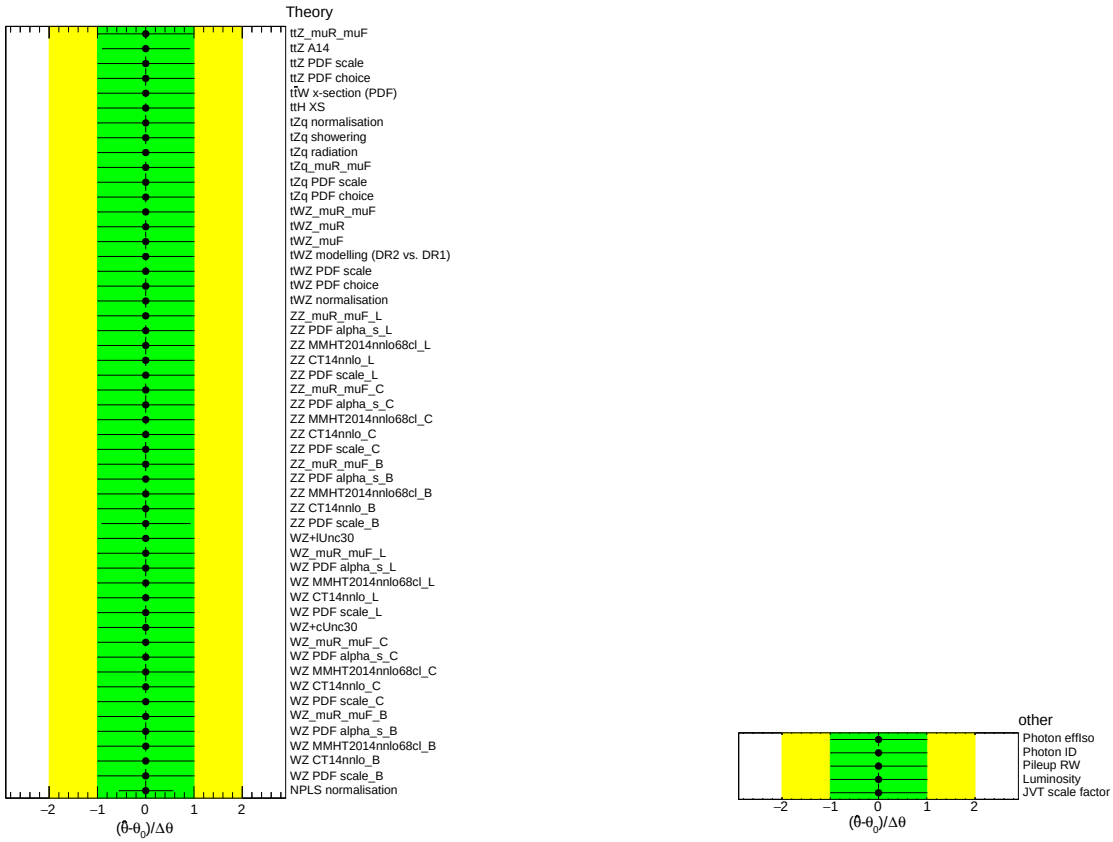

Figure 9.6: Pull plots for the theory, photon, integrated luminosity, jet vertex tagger and pile up uncertainties.

10 Discussion and Outlook

10.1 Performance with and without DNN

In this analysis, a DNN was used, to decrease the uncertainties. This can be achieved, by creating histograms, using the output of the DNN. However, it needs to be discussed, if there is really a difference between using a DNN or not and if the DNN might even perform worse than just using the cut and count method. Comparing the uncertainties of the signal strength of tZq for both methods, it can be seen, that the maximal total uncertainty for the DNN method is 52 %, whereas the overall total uncertainty for the cut and count method is 73%. Therefore the usage of the DNN improves the precision by 21 %, which is a lot, as the total uncertainty for the last tZq measurement was only 14 % [31]. For the $t\bar{t}Z$ uncertainty, an improvement of 1 % could be achieved, by using the DNN method instead of the cut and count method. However, this improvement is negligible, as it is just a slight improvement.

Therefore it can be concluded, that the usage of the DNN methods makes a difference and can significantly decrease the uncertainties.

10.2 Uncertainties Comparison with Previous Paper

In this thesis, the signal strength for $t\bar{t}Z$ and tZq is simultaneously measured via a binned likelihood fit. In the following it will be discussed, how big the uncertainties in this analysis were, compared to the uncertainties in the signal strengths previously measured at ATLAS, from which the cross section is then extracted. However, the value of the fit results in this thesis will not be discussed, as the Monte Carlo simulations was set as the data, on which the parameters were fitted. Therefore the cross section, extracted from the signal strength, will be identical to the theoretically predicted cross section.

The most recent cross section measurement at ATLAS for $t\bar{t}Z$ had a total uncertainty of 10 % [32], whereas the most recent tZq cross section measurement had a total uncer-

tainty of 14 % [31]. In comparison to the uncertainties in Chapter 9.1, the most recent $t\bar{t}Z$ measurement by ATLAS has a higher uncertainty, whereas for tZq the most recent measurement by ATLAS has a lower uncertainty.

Furthermore it can be observed, that the statistical uncertainty for tZq in this thesis is already worse, than the total uncertainty, in the most recent measured cross section by ATLAS. However, the simultaneous measurement does not lead to the high tZq uncertainty of this thesis, as the fit was also performed as tZq single parameter, which did not improve the uncertainty.

An assumption for the bad performance of the tZq fit would be, that the tZq classification of the DNN was not very good, due to the used input parameters. If the input parameters of the most recent ATLAS tZq measurement is compared to the input parameters of this thesis, it can be seen, that the highest ranking ones of the ATLAS paper are not included in the training of the DNN in this analysis.

10.3 Outlook

This analysis was able to improve the uncertainties for $t\bar{t}Z$ compared to the most recent ATLAS $t\bar{t}Z$ cross section measurement. However, the uncertainties for tZq in this thesis, compared to the most recent ATLAS tZq cross section measurement, are way worse. The next step is to check if the tZq uncertainties can be reduced, by training a DNN with the variables, used in the tZq ATLAS paper [31].

Moreover, regions with higher tZq yields could be constructed, in order to reduce the tZq uncertainties further, although these regions may not be as pure, as the SR-3ℓ region defined in this thesis. However, regions with high tZq purity can be constructed by using a DNN. Therefore the uncertainties, arising from the low tZq yields can be avoided and at the same time the DNN is given more tZq data, to enhance the correct classification rate.

Bibliography

[1] S. Weinberg, *A model of leptons*, Phys. Rev. Lett. **19(21)** (1967)

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

[3] S. L. Glashow, *Partial-symmetries of weak interactions*, Nucl. Phys. **22(4)** (1961)

[4] S. L. Glashow, J. Iliopoulos, and L. Maiani, *Weak interactions with lepton-hadron symmetry*, Phys. Rev. D **2(7)** (1970)

[5] H. D. Politzer, *Asymptotic freedom: An approach to strong interactions*, Phys. Rept. **14(4)** (1974)

[6] W. Marciano and H. Pagels, *Quantum chromodynamics*, Phys. Rept. **36(3)** (1978)

[7] M. Goldhaber, L. Grodzins, and A. W. Sunyar, *Helicity of neutrinos*, Phys. Rev. **109(3)** (1958)

[8] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, Phys. Rev. Lett. **13** (1964)

[9] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, Phys. Rev. Lett. **13** (1964)

[10] P. W. Higgs, *Spontaneous symmetry breakdown without massless bosons*, Phys. Rev. **145** (1966)

[11] G. S. Guralnik et al., *Global conservation laws and massless particles*, Phys. Rev. Lett. **13** (1964)

[12] D. J. Gross and F. Wilczek, *Ultraviolet behavior of non-abelian gauge theories*, Phys. Rev. Lett. **30** (1973)

[13] DØ Collaboration, *Observation of the top quark*, Phys. Rev. Lett. **74(14)** (1995)

[14] CDF Collaboration, *Observation of top quark production in $p\bar{p}$ collisions with the Collider Detector at Fermilab*, Phys. Rev. Lett. **74(14)** (1995)

[15] N. Cabibbo, *Unitary symmetry and leptonic decays*, Phys. Rev. Lett. **10(12)** (1963)

[16] M. Kobayashi, T. Maskawa, *CP-violation in the renormalizable theory of weak interaction*, Prog. Theor. Phys. **49(2)** (1973)

[17] Atlas Collaboration, *The Atlas experiment at the Cern large hadron collider*, JINST **3(S08003)** (2008)

[18] Atlas Collaboration, *Vertex Reconstruction Performance of the ATLAS Detector at* $\sqrt{s} = 13$ *TeV*, Technical report, ATL-PHYS-PUB-2015-026 (2015)

[19] Atlas Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data*, JINST **14(12)**, P12006 (2019)

[20] Atlas Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at* $\sqrt{s} = 13\,\text{TeV}$, Eur. Phys. J. C **76(5)** (2016)

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

[22] Atlas Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at* $\sqrt{s} = 8\,\text{TeV}$ *using the ATLAS detector*, Eur. Phys. J. C **76(11)** (2016)

[23] Atlas Collaboration, *Optimisation of the ATLAS b-tagging algorithms for the 2017-2018 LHC data-taking*, Technical report, ATL-COM-PHYS-2017-1472 (2017)

[24] Atlas Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at* $\sqrt{s} = 13\,\text{TeV}$, Eur. Phys. J. C **78(11)** (2018)

[25] Atlas Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at* $\sqrt{s} = 13\,\text{TeV}$, Eur. Phys. J. C **79(8)** (2019)

[26] H. J. Kelley, *Gradient theory of optimal flight paths*, Ars Journal **30(10)** (1960)

[27] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, Proceedings of the 3rd International Conference on Learning Representations (2014)

[28] S. J. Reddi, S. Kale, and S. Kumar, *On the convergence of adam and beyond*, arXiv preprint arXiv:1904.09237 (2019)

[29] I. Sutskever, et al., *On the importance of initialization and momentum in deep learning*, in S. Dasgupta, D. McAllester, editors, *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 1139–1147, PMLR, Atlanta, Georgia, USA (2013)

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

[31] ATLAS Collaboration, *Observation of the associated production of a top quark and a Z boson in pp collisions at* $\sqrt{s} = 13\,\text{TeV}$ *with the ATLAS detector*, JHEP **2020(7)**, 1 (2020)

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

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 23. Dezember 2022

(Kia-Jüng Yang)