

Bachelor's Thesis

Studium des $t\bar{t} + \gamma$ Produktionsprozesses mit dem ATLAS-Experiment am LHC

Studies of the $t\bar{t} + \gamma$ production process with the ATLAS experiment at the LHC

prepared by

Pascal Herrmann

from Hannover

at the II. Physikalischen Institut

Thesis number: II.Physik-UniGö-BSc-2018/06

Thesis period: 3rd April 2018 until 10th July 2018

First referee: Prof. Dr. Arnulf Quadt

Second referee: Prof. Dr. Stan Lai

Zusammenfassung

Messungen der Eigenschaften des Top Quarks sind ein wichtiger Bestandteil der aktuellen Forschung in der Teilchenphysik. Diese Arbeit beschäftigt sich mit der Untersuchung des $t\bar{t}\gamma$ Produktionsprozesses. Hierbei wird unter anderem die Kopplung zwischen dem Top Quark und dem Photon untersucht. Das Photon kann hierbei direkt von dem Top Quark oder seinen Zerfallsprodukten abgestrahlt werden, oder von den Quarks im Anfangszustand. Aktuelle Forschungsergebnisse berücksichtigen diese Unterscheidung nicht, welche aber für sehr präzise Messungen, insbesondere der Top-Photon-Kopplung, nötig ist. Daher wurde im Rahmen dieser Arbeit ein Werkzeug entwickelt, welches in der Lage ist, diese beiden Prozesse zu unterscheiden. Dafür wurden Simulationen durchgeführt und Variablen untersucht, in welchen die beiden Prozesse unterscheidbar sind. Diese wurden dann genutzt, um ein neuronales Netzwerk zu trainieren. Ein Area Under Curve (AUC)-Wert von 0.8323 wurde erreicht, was bedeutet, dass etwa 83% der Prozesse richtig klassifiziert wurden.

Abstract

The measurement of the properties of the top quark is an important part of current research in particle physics. This thesis involves the $t\bar{t}\gamma$ analysis, where the coupling between the photon and the top quark can be examined. The photon can be radiated off the top quark or its decay products, or the quarks in the initial state. The current analysis does not distinguish between those cases. But this is necessary for high precision measurements, in particular of the top-photon-coupling. In this thesis, a tool is developed to separate events where the photon is radiated off the top quark or the quarks in the initial state. Events were simulated and discriminant variables were investigated. These variables were then used to train a neural network. An Area Under Curve (AUC) value of 0.8323 is achieved which means that about 83% of the events were classified correctly.

Contents

1. Introduction	1
2. Theory	3
2.1. The Standard Model	3
2.2. The top quark	5
2.3. The $t\bar{t}\gamma$ process	6
3. Experimental setup	9
3.1. LHC	9
3.2. ATLAS	10
4. Neural Networks	15
4.1. General concepts	15
4.2. Machine learning tools	18
5. Development of a tool to distinguish between FSR and ISR photons	19
5.1. Discriminating Variables	19
5.2. Training of a neural network	23
6. Conclusion	27
A. Appendix	29
A.1. Changes in MadGraph	29
A.2. AUC values of trained neural networks	30

1. Introduction

After the discovery of the top quark in 1995 at the TEVATRON [1, 2] the research on the top quark progressed a lot. At first it was only a rare signal process. The theory of the Standard Model (SM) and more detail about the top quark are described in Chapter 2. Now in the Large Hadron Collider (LHC) [3] era (see Chapter 3), the top quark production has become a well known process which is relatively easy to observe. A lot of different properties of the top quark were measured and all measurements are so far in agreement with the SM. But one measurement which is still not done yet is the coupling strength between the top quark and the photon. By doing that, the electric charge of the top quark can be determined. This is one of the goals of the $t\bar{t}\gamma$ analysis.

Current measurements are done but one aspect is not taken into account. The photon can not only be radiated off the top quark but also the quarks in the initial state of the $t\bar{t}$ production process can emit a photon. Currently both processes are counted as signal but for high precision measurements of the coupling only the events where the photon was emitted by the top quark are relevant. The goal of this thesis is to develop a tool which is able to distinguish these events.

Neural networks are used for this task. The theory behind them as well as the tools which were used in this thesis are described in Chapter 4.

In order to develop a tool that can distinguish between the processes, events were simulated using MadGraph [4] and discriminant variables were investigated. After that a neural network was trained to distinguish the events (see Chapter 5).

The final results and an outlook for future analysis are shown in Chapter 6.

2. Theory

2.1. The Standard Model

The Standard Model (SM) of particle physics describes fundamental particles and the interactions between them. These fundamental particles are called fermions and are spin- $\frac{1}{2}$ particles. Currently, 12 fermions of different flavour are known which are grouped in three generations of quarks and leptons. Particles of different generations only differ in their masses which increases with higher generations.

Furthermore, the quark and the lepton sector are each divided into up-type and down-type particles which differ by one unit in the third component of the weak isospin I_3 . Up-type fermions have $I_3 = +\frac{1}{2}$ and down-type fermions have $I_3 = -\frac{1}{2}$.

The SM describes three of the four fundamental forces which are the electromagnetic force, described by Quantum Electro Dynamics (QED), the weak force, and the strong force, described by Quantum Chromo Dynamics (QCD). Gravity is not described by the SM. Actually, as shown by Glashow [5], Weinberg [6] and Salam [7], the electromagnetic force and the weak force both come from the same physical phenomenon which is described in the electroweak unification.

In the SM, these forces get transmitted via gauge bosons which are spin-1 particles. The electromagnetic force is transmitted by massless photons, the strong force is transmitted by massless gluons and the weak force is transmitted by massive electrically charged W^\pm -bosons and neutral Z -bosons.

However, these bosons can only couple to the different charges of a fermion. Photons couple to the electric charge, which all quarks and their antiparticles and all charged leptons and their antiparticles have. Neutrinos do not have an electric charge and hence do not interact electromagnetically. The W^\pm -bosons have an electric charge as well so they also interact electromagnetically. Gluons couple to the colour charge which only quarks and the gluon itself has, which means that gluons can couple to themselves. The charge of the weak force is the third component of the isospin which all particles have as mentioned before. This means that all fermions can interact via the weak force.

In the electroweak theory, the weak hypercharge [6] is defined as $Y = 2(Q - I_3)$, where Q

2. Theory

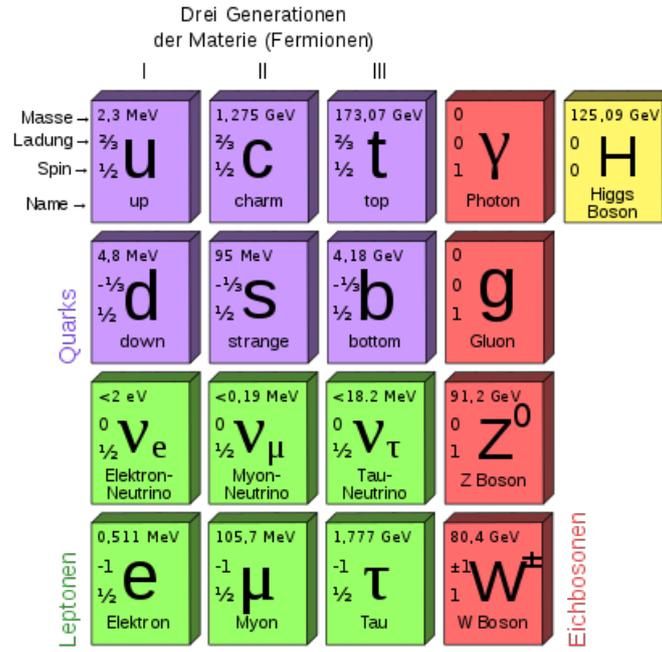


Figure 2.1.: All components of the Standard Model.

is the electric charge. Therefore, the photon, the W^\pm -boson and the Z -boson couple to the weak hypercharge.

The last part of the SM is the Higgs boson. According to the mathematical description of the SM all gauge bosons should be massless. But experiments show that the W^\pm -boson has actually a mass of $m_W = (80370 \pm 19) \text{ MeV}$ [8] and the Z -boson a mass of $m_Z = (91187.6 \pm 0.0021) \text{ MeV}$ [9]. Mathematically, the SM is described as a local gauge and Lorentz invariant Lagrangian density function which is fully renormalisable [10]. To solve the problem with the gauge boson masses, it is not possible to just add a mass term to the Lagrangian density function because it would violate the local gauge invariance. The solution to this problem was found by Higgs [11], Englert and Brout [12] and Guralnik, Hagen and Kibble [13]. Instead of a mass term a scalar field was added. This spontaneously breaks the electroweak symmetry which allows the bosons to acquire a mass while the local gauge invariance is still satisfied. This scalar field was later called Higgs field and is also responsible for the masses of most fermions (except neutrinos). One consequence of this scalar field is the existence of a spin-0 scalar boson which would be a result of the excitation of the field. This particle, called Higgs boson, was discovered in 2012 at the LHC by ATLAS [14] and CMS [15], therefore completing the SM of elementary particle physics.

All the components of the SM are summarised in Figure 2.1.

The SM is very successful in describing the smallest parts of the known universe. It

makes precise predictions which were probed by many different experiments. However, it is known that the SM is an incomplete description of the universe. For example, as it was already mentioned, gravity is not included in the SM. Furthermore, dark matter or dark energy are not explained so only about 5% of the energy distribution of the universe is included in the SM. Also, neutrinos do not interact with the Higgs field which means that they should be massless, but that is a contradiction to experimental results [16]. This shows that further investigations and high precision measurements are necessary to explain these aspects of our universe.

Precise measurements of the SM and the search for physics beyond the SM are part of the research at the LHC. This includes for example the search for supersymmetric particles or heavy resonances (W', Z'). But these particles have not been observed at the LHC.

2.2. The top quark

An important part of current research at the LHC is the top quark. With a mass of $m_t = 173.0 \pm 0.4$ GeV [9], it is currently the heaviest known elementary particle. Because of that it has a high decay width and hence a short lifetime of $\tau = 0.5 \times 10^{-24}$ s [17]. This time is shorter than the interaction time of the strong force which means that the top quark decays before it interacts strongly or hadronizes. This gives a rare opportunity to study a "bare" quark. Normally, quarks have to be in bound states due to confinement. This also gives restrictions on how a top quark decays. Since it does not interact strongly it only decays via the weak force almost always into a W^+ -boson and a bottom quark. Decays to other down-type quarks are possible but highly suppressed because of the structure of the CKM matrix [18]. The top antiquark also decays via the weak force into a W^- -boson and a bottom antiquark. The W^\pm -bosons then decay further both into quarks (full hadronic decay), both into leptons (dileptonic decay) or one into quarks and one into leptons (semileptonic decay). Knowing these decay modes is important to identify a top-antitop pair in the detector.

Some measurements which were already done include the mass difference between the top quark and the top antiquark [19], forward-backward asymmetry [20] and charge asymmetry [21], spin correlation [22], the $|V_{tb}|$ element of the CKM matrix [23] and the search for flavour-changing neutral currents [24]. All those measurements are currently consistent with the SM but more precise measurements could give clues about physics beyond the SM.

Another aspect of current research, which is the topic of this thesis, is the measurement of the coupling between the top quark and photons by measuring the associated production

2. Theory

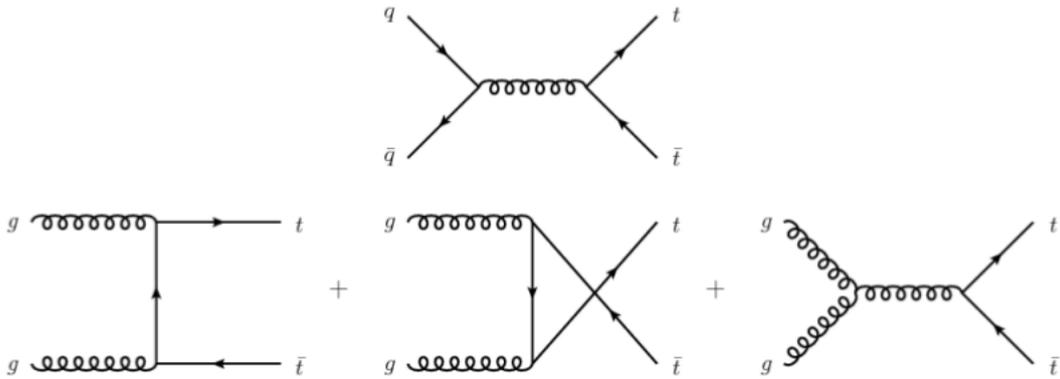


Figure 2.2.: All possible production channels of a $t\bar{t}$ pair at leading order. The diagram on top shows quark-antiquark annihilation and the three diagrams at the bottom show gluon-gluon fusion.

of a top quark-antiquark pair and a photon. Any deviation of the measured cross section value from the SM prediction would be an indication of beyond SM physics. Even though recent measurements are in agreement with the SM predictions [25, 26].

The LHC is a proton-proton collider, so the main production process for a $t\bar{t}$ pair is gluon fusion with about 90% [27]. The other 10% are quark-antiquark annihilation. In Figure 2.2 all possible $t\bar{t}$ production processes at leading order are shown.

2.3. The $t\bar{t}\gamma$ process

The $t\bar{t}\gamma$ production process gives a good opportunity to directly study the coupling strength between the photon and the top quark. Current cross section measurements at ATLAS and CMS yield a cross section of $\sigma_{sl} = 139 \pm 18$ fb [25] and $\sigma_{sl} = 127 \pm 27$ fb [28], respectively, in the semileptonic channel at an energy of $\sqrt{s} = 8$ TeV. In both measurements, all prompt photons (photons originating from the hard collision) are counted as signal. This means, that not only photons radiated off the top quark are taken into account, but also photons radiated off the partons in the initial state and photons radiated off the decay products of the top quarks. Since only photons radiated off the top quark are of interest for these coupling measurements, one must distinguish between the initial state radiation (ISR) and final state radiation (FSR). The final state radiation can further be separated into photons from the top quarks, photons from the bottom quarks and photons from the W bosons. In this thesis only the separation between FSR from the top quark (which will only be called FSR in the following) and ISR is discussed. The processes are shown in Figure 2.3.

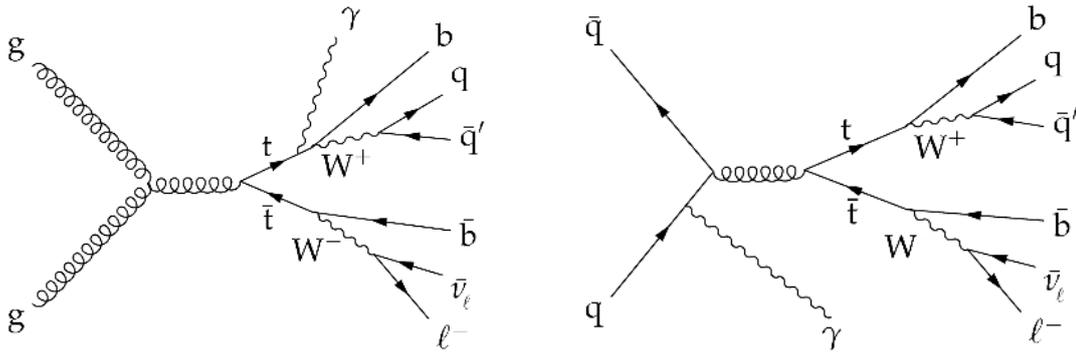


Figure 2.3.: The two processes which will be discussed in this thesis are shown. The left figure shows a photon from the final state radiation (FSR) process from the top quark. In this case the top quarks are created by gluon fusion. The right figure shows a photon from the initial state radiation (ISR) process.

Since the dominant production process for $t\bar{t}\gamma$ is gluon fusion and gluons can not radiate off photons, ISR has only a small contribution to the cross section. But for high precision measurements, the separation between FSR and ISR is still necessary.

3. Experimental setup

3.1. LHC

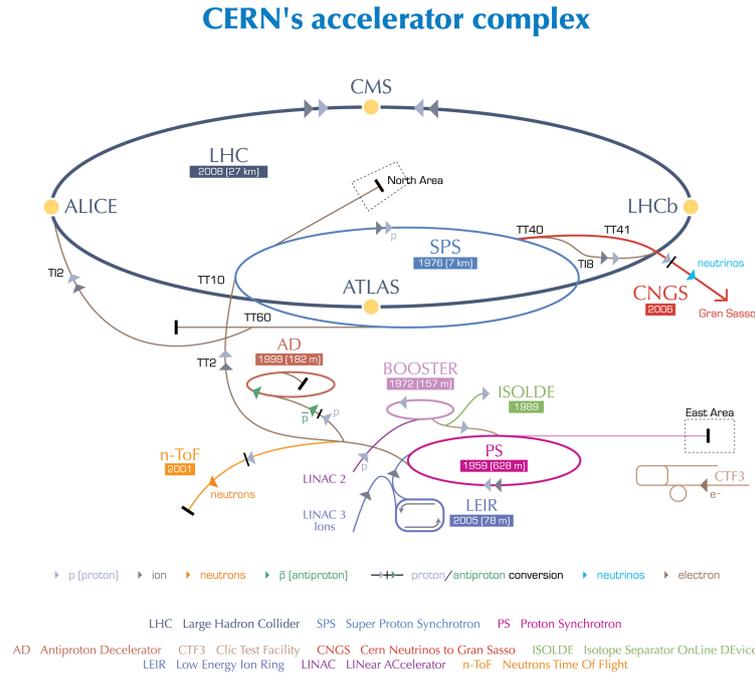
The Large Hadron Collider (LHC) [3] at CERN started operating in 2009. It is installed about 100 m underground in the tunnel previously used for the Large Electron-Positron Collider (LEP) near Geneva, Switzerland. The LHC has a circumference of about 27 km. It is designed for a maximum centre-of-mass energy of $\sqrt{s} = 14$ TeV. Because the LHC is a particle-particle collider, two separate beam pipes with magnetic fields pointing in opposite directions are needed, unlike particle-antiparticle colliders which only need one pipe. The particles inside the beam are grouped into bunches of up to 10^{11} protons and both beam pipes contain 2808 bunches each. The LHC has three modes of operations. In one mode two proton beams are used, in the second mode two beams of lead ions are used and the third mode collides protons with lead ions. A high magnetic field strength is necessary to bend the high energetic particle beams to a circular orbit. For that, superconductive NbTi magnets are used which operate at a temperature below 2 K using liquid helium. In the 7 TeV run a field strength up to $|B| = 8.33$ T was used.

The LHC has four experiments: ATLAS [29], CMS [30], ALICE [31] and LHCb [32]. The first two are general-purpose experiments which aim at a peak luminosity of $\mathcal{L} = 10^{34}$ cm⁻² s⁻¹ for proton operation. The ATLAS experiment is going to be discussed in more detail in Section 3.2. LHCb is a low luminosity experiment specialised on *b*-physics at a peak luminosity of $\mathcal{L} = 10^{32}$ cm⁻² s⁻¹ for proton collisions. ALICE is a heavy ion experiment aiming at a peak luminosity of $\mathcal{L} = 10^{27}$ cm⁻² s⁻¹ for lead-lead ion operation.

The LHC makes use of the accelerating facilities at CERN because the magnet system of the LHC alone is not capable to cover the complete acceleration process. A scheme of the acceleration complex is shown in Figure 3.1.

The starting point of the proton injection chain is the linear accelerator LINAC 2. There the protons get accelerated up to 50 MeV after which they reach the BOOSTER. Here, the protons get further accelerated and reach the Proton Synchrotron (PS) with an energy of 1.4 GeV. They leave it with an energy of 25 GeV where they are injected into the Super Proton Synchrotron (SPS) and finally reach the LHC with an energy of 450 GeV. There

3. Experimental setup



European Organization for Nuclear Research | Organisation européenne pour la recherche nucléaire

© CERN 2008

Figure 3.1.: The acceleration facilities at CERN. © CERN.

they get accelerated up to the desired energy. The LHC has four interaction points where the two beam pipes cross. At these crossings, the four detector experiments are installed to observe the final states of the collision.

The ion acceleration process is almost the same except that instead of LINAC 2 and BOOSTER the accelerators LINAC 3 and LEIR are used. For lead ions, a total centre-of-mass energy of 1.15 PeV (2.76 TeV per nucleon) is achieved.

3.2. ATLAS

The ATLAS detector is a general purpose detector [29] designed for more precise measurements of the SM and physics beyond that. A sketch of the detector and its main components is shown in Figure 3.2.

The ATLAS detector is placed symmetrically around the beam pipe and covers almost a solid angle of 4π to detect as many particles as possible.

The most inner part of the detector is the tracking system. It is used for pattern recognition, momentum and vertex measurement and electron identification. The tracking system itself is composed of three layers: the semiconductor pixel tracker and silicon microstrip tracker (SCT) and the Transition Radiation Tracker (TRT). The inner detector

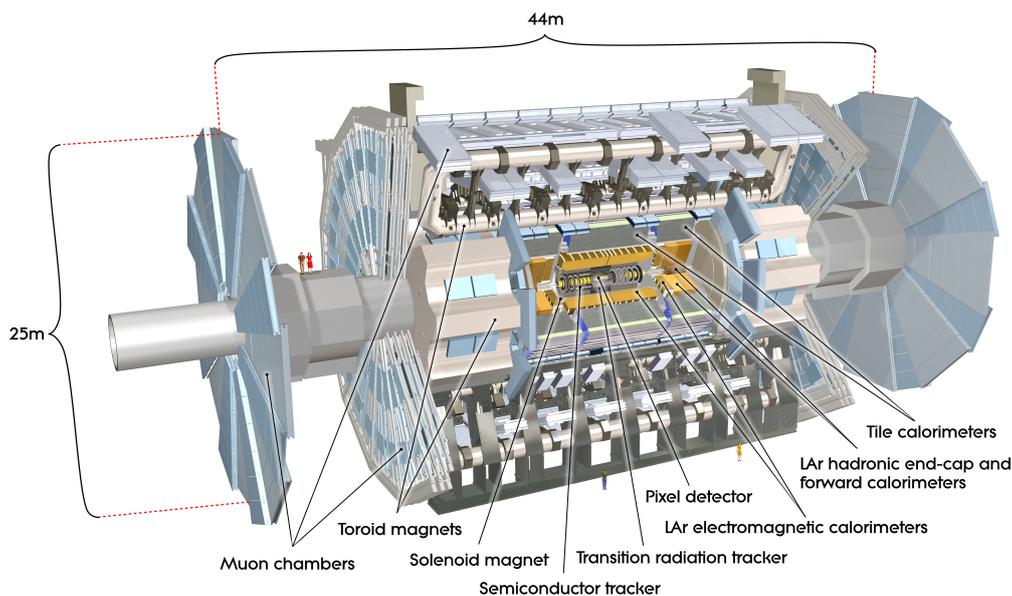


Figure 3.2.: A schematic picture of the ATLAS detector and its main components.
© CERN.

is immersed in a 2 T magnetic field pointing in beam direction produced by a solenoid magnet.

The calorimeter system of ATLAS is built around the tracking system. The task of the calorimeter system is to measure the energy of the particles produced in the crossing area as well as missing transverse energy which normally indicates the production of neutrinos. The calorimeter system consists of two parts which are the electromagnetic (EM) calorimeter and the hadronic calorimeter.

The EM calorimeter is ideally suited to measure particles which interact mostly by the electromagnetic force. That is the reason why the EM calorimeter comes before the hadronic calorimeter. Particles like photons and electrons deposit all of their energy in the EM calorimeter while particles like hadrons pass through it. Nevertheless, hadrons lose some energy in the EM calorimeter but only in the hadronic calorimeter they get stopped completely. An exception to this are muons which are too heavy to lose energy via Bremsstrahlung as a main process. So they pass through both calorimeter types and only get detected in the muon chambers which will be discussed later.

The ATLAS experiment uses sampling calorimeters which means they consist of alternating layers of an absorber material and an active material. The absorber material slows the particle down but is not capable to do measurements. The EM calorimeter uses lead for this purpose. The active material, which is liquid argon for the EM calorimeter, is used

3. Experimental setup

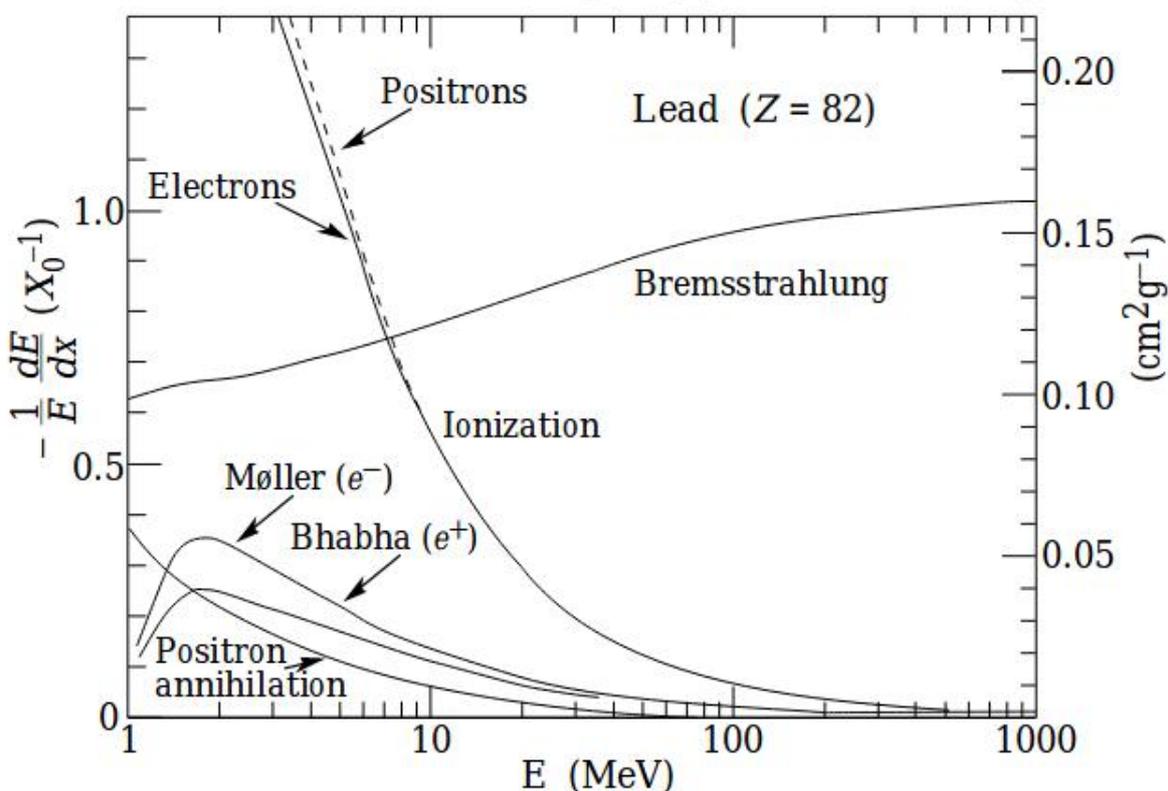


Figure 3.3.: The fractional energy loss in lead as a function of electron or positron energy is shown. The left y-axis shows the energy loss in units of inverse radiation length and the right y-axis in units of cm^2/g . The critical energy E_C is the energy where the ionisation curve and the Bremsstrahlung curve cross. Taken from [33].

to actually detect a signal. The hadronic calorimeter uses steel as the absorber material and scintillator tiles as the active material.

The EM calorimeter measures the energy of a particle as follows [33]: the main process for high energetic electrons/positrons to lose energy in matter is by emitting a high energetic photon. This radiation is called Bremsstrahlung. This photon loses its energy in matter again by electron-positron pair productions and these particles again emit photons. An electromagnetic cascade is generated. This process happens until the electrons/positrons and photons do not have enough energy for pair production or Bremsstrahlung. At this point these particles lose their energy rather through ionisation. The behaviour how a charged particle loses its energy in matter is described by the Bethe-Bloch equation. The energy where the energy loss by Bremsstrahlung/pair production is the same as the energy loss by ionisation is called the critical energy E_C which is shown in Figure 3.3.

In the calorimeter, the length of the electromagnetic cascade N in units of radiation

length X_0 is measurable. The radiation length X_0 is defined as the mean distance in which an electron loses $1/e$ of its initial energy. The radiation length for photons is almost the same ($\frac{7}{9}X_0$). The following equation applies for the length of the electromagnetic cascade:

$$N \approx \frac{E}{E_C}, \quad (3.1)$$

where E is the energy of the initial particle. So by measuring N and calculating E_C from theory the energy of the initial particle can be calculated.

The processes in the hadronic calorimeter are very similar but a bit more complicated [34]. If quarks are produced in particle collisions they are not allowed to be free particles due to confinement. They rather form baryons and mesons through hadronisation. By this a parton shower is produced. Heavier hadrons are unstable and decay into lighter hadrons like pions or η -mesons creating a hadronic cascade. These hadrons decay further into photons which then form electromagnetic cascades like discussed before, transforming the hadronic into an electromagnetic cascade. Hadronic cascades are described by λ_I which is the nuclear interaction length. It describes the mean length in which the number of relativistic charged particles is reduced by $1/e$. Then again, by measuring λ_I the energy of the initial particle can be calculated.

The last and outermost part of the ATLAS detector is the muon system. As the name suggests, it is used to do high precision measurements of the momentum of the muons. Since muons can pass through both calorimeters they have to be measured last. They are the only particles which leave a track in the muon chamber because all the other particles get stopped in the calorimeter or are not detectable like neutrinos. The measurement is done by deflecting the paths of the muons with a magnetic field generated from large superconducting toroid magnets. The field is mostly orthogonal to the direction in which the muons move. The momentum of the muon is then calculated by measuring the curvature of its path.

Because of the layered structure in principal every particle can be identified. Figure 3.4 shows which particle leaves which track.

3. Experimental setup

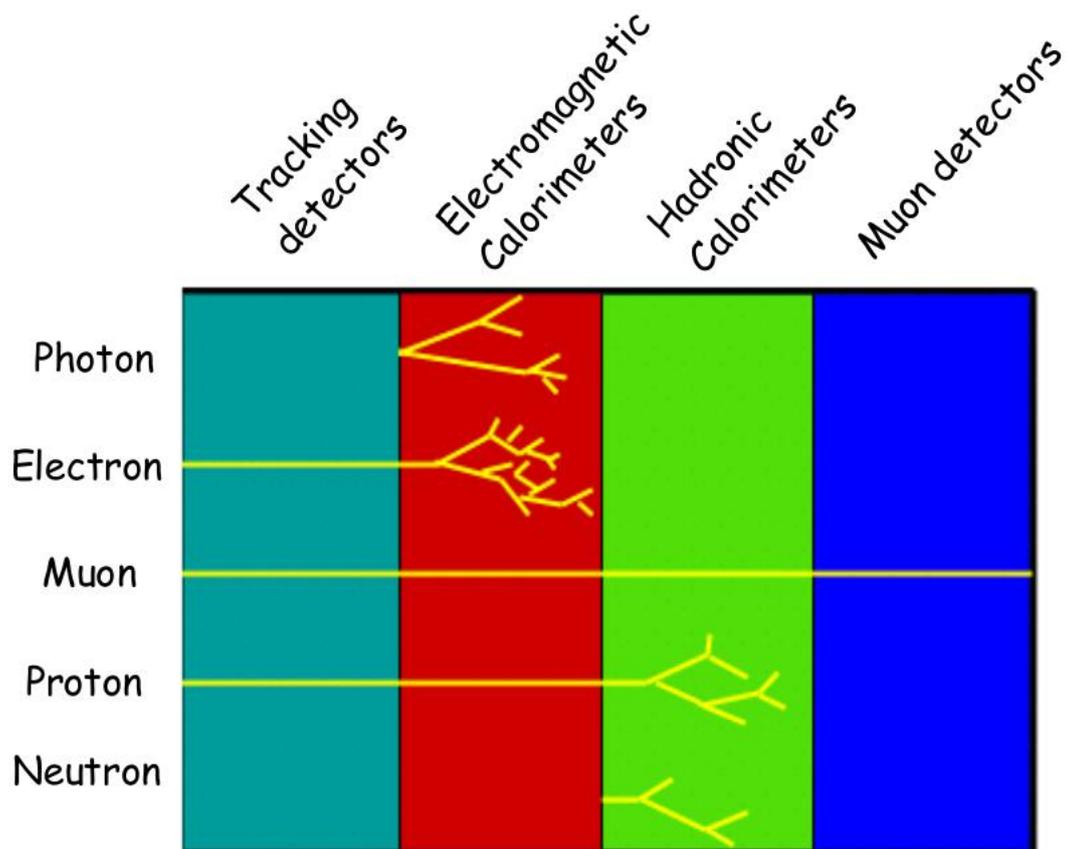


Figure 3.4.: It is shown which particle leaves which track in the different layers of ATLAS.

4. Neural Networks

4.1. General concepts

In order to distinguish between photons emitted by top quarks and other photons, machine learning algorithms are used. The most common algorithms used in physics are Boosted Decision Trees (BDTs) and neural networks (NNs). But only neural networks are going to be discussed here.

Machine learning is inspired by how the brain works with its synapses and neurons. The first attempts in machine learning were done by Pitts and McCulloch in 1943 [35] by introducing linear models with the following equation:

$$f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} = \sum_i w_i x_i, \quad (4.1)$$

each entry x_i of the vector \mathbf{x} is called a feature and each vector summarises the features of one sample of data. Every feature i is also weighted by the factor w_i , which are the parameters of the model. Every sample gets a value assigned by $f(\mathbf{x})$. If for a specific input a specific output is desired, all values for w_i had to be set by hand accordingly.

Since samples could become arbitrarily large, it was not feasible to adjust all weights by hand. So in 1958, Rosenblatt [36] introduced the perceptron which was able to learn the values for the weights by using labelled samples in the algorithm. But in this simple form, the algorithm was not capable to do simple tasks like the logic exclusive OR operation. This problem was solved by Rumelhart, Hinton and Williams [37] in 1988 by introducing backpropagation. For that, several layers were introduced to the algorithm which also calculated the weights by themselves.

One important concept of machine learning is the classification problem. The task of the algorithm is to classify the data in different categories. It can be distinguished between the binary classification problem where only two classes are available and the multi-class classification where more classes are possible. Discriminating between signal and background or photons from top quarks and other photons falls in the binary category and will be part of this thesis.

4. Neural Networks

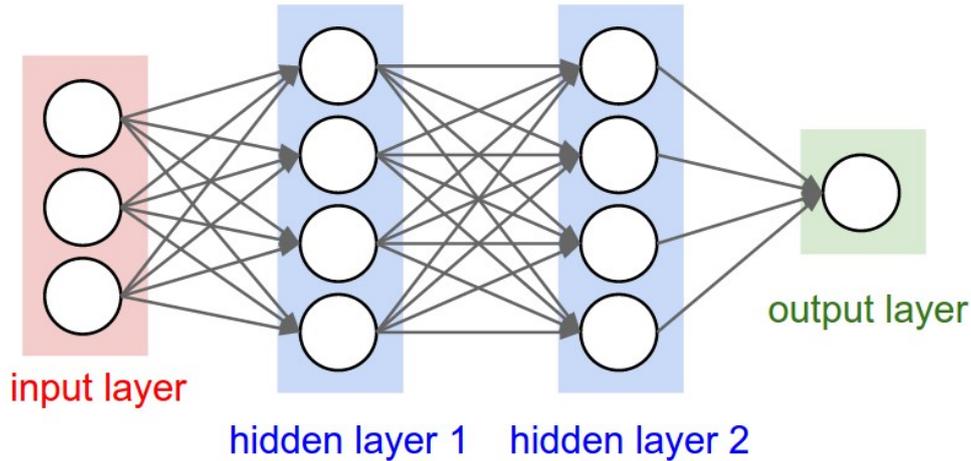


Figure 4.1.: The general scheme for a neural network is shown. Every neural network consists of one input layer, one output layer and in this case two hidden layers.

In the approach of supervised learning, the algorithm uses a set of data where the true label ($\hat{\mathbf{y}}$) is known. The machine learning algorithm can then be thought of as a function $f(\mathbf{x})$ with parameters w_i which assigns a label \mathbf{y} to the presented data \mathbf{x} . In the binary classification problem, \mathbf{y} is scalar and every value above a specific value belongs to one class and every value below it to the other.

During the learning period the weights w_i are adjusted in a way that for every sample the calculated label ($f(\mathbf{x}_i) = \mathbf{y}_i$) is as close as possible to the true label $\hat{\mathbf{y}}$. To measure how well the algorithm performs, a scalar loss function C is defined. The weights w_i are calculated in a way that the loss function C gets minimised which corresponds to calculating $\nabla_w C|_{\mathbf{w}_0} = 0$. Calculating this gradient can be very complicated and hence concepts like backpropagation are used.

The training period is usually followed by a testing period where labelled data is used in the algorithm as well. The difference to the training period is that the weights do not get adjusted anymore. Also, the data used for testing was not used for training. This is important to check if the algorithm is able to classify unseen data correctly.

In particle physics, these labelled data sets often come from Monte Carlo simulations.

A NN is one type of machine learning algorithm. They have one input layer in which the data gets read in, an output layer where the resulting label gets displayed and usually one or more hidden layers. Each layer consists of neurons which are a type of perceptrons which were already mentioned. The connections between the neurons of these layers is again described by the function $f(\mathbf{x})$ which was already discussed. A scheme for a neural network is shown in Figure 4.1.

Each neuron has an activation function

$$\sigma(x) = \begin{cases} 0 & x \leq b \\ 1 & x > b, \end{cases} \quad (4.2)$$

where b is the bias of the neuron. This can be thought of as a threshold which the input signal needs to overcome to create an output. The values are again weighted with parameters w_i . Often it is not enough to distinguish between signal and no signal. For a continuous activation, some commonly used functions are shown in Table 4.1.

Function	Definition
Softmax	$\sigma_{soft}(\vec{x})_i = \frac{e^{x_i}}{\sum_{j=1}^J e^{x_j}}$
relu	$\sigma_{relu}(x) = \begin{cases} 0, & \text{if } x < 0 \\ x, & \text{otherwise} \end{cases}$
sigmoid	$\sigma_{sig}(x) = \frac{1}{1+e^{-x}}$
selu	$\sigma_{selu}(\alpha, x) = \lambda \begin{cases} \alpha(e^x - 1), & \text{if } x < 0 \\ x, & \text{otherwise} \end{cases}$
Softplus	$\sigma_{softplus}(x) = \ln(1 + e^x)$

Table 4.1.: Examples of activation functions. Relu is short for rectifier linear unit. Selu is short for scaled exponential linear unit. $\lambda = 1.0507$, $\alpha = 1.67326$ [38].

All these activation functions are used to train a neural network in Section 5.2. There are also different layers. One for example is the so-called batch normalisation layer [39]. It is used to scale the output of each neuron in a layer in a way that the distribution of the outputs of all neurons has a mean of 0 and a standard deviation of 1. This is used to keep the parameters of a NN at the same order of magnitude. In the training phase of the neural network the weights and the biases need to be adjusted. When calculating the loss function, a commonly used function for binary classification problems is the binary cross entropy. It is defined as:

$$C_{BCE}(y, \hat{y}) = -\sum_{i \in samples} [\hat{y}_i \log y_i + (1 - \hat{y}_i) \log(1 - y_i)]. \quad (4.3)$$

Here, \hat{y}_i is the true label and y_i is the predicted label for the sample i . In this thesis \hat{y}_i can have values between 0 and 1.

4.2. Machine learning tools

In order to use these theoretical considerations about neural networks, they have to be implemented in software. Two libraries which support machine learning algorithms are Theano [40] and TensorFlow [41]. For an easier use of these libraries, Keras [42] was developed. Keras is a high-level API which runs on top of Theano or TensorFlow and is written in python. In this thesis, Theano is used as backend. With Keras, it is very easy to define new neural networks and train them and so it was used in this thesis.

The variables which are used to train the neural network (see Section 5.2) are saved in root files. But these files cannot be used directly in Keras. For that, the root2kerasPipeline [43] tool is used. First, the root-files generated by MadGraph (see Section 5.1) are converted into a single root file, which contain the variables of interest. The number of FSR events and ISR events have to be similar to avoid the risk of overtraining. Overtraining means that the neural networks becomes highly specific to the applied events, which means it would be not very good at classifying unseen events. This has to be avoided in an analysis. After saving the variables into a single root file, the file has to be converted into a hdf5 file using the root2kerasPipeline. In this step, the features (variables) for the neural network are selected as well as the label of the variables which is either FSR or ISR.

The hdf5 file can then be used as the input to train a neural network which is also done using the root2kerasPipeline tool.

5. Development of a tool to distinguish between FSR and ISR photons

5.1. Discriminating Variables

In order to distinguish between photons from FSR and ISR events, it is necessary to find variables in which these processes differ. To do that, two data samples were created by using the event generator MadGraph [4]. But MadGraph normally does not distinguish between FSR and ISR photon events. Therefore, it had to be modified. A new coupling was introduced which is equivalent to QED. This coupling was then set to be the coupling between photons and the top quark. By that, two data samples could be created, from which one only contains events where a photon was radiated off the top quark. The other sample contains events where a photon was either radiated off the quarks in the initial state or the decay products of the top quark. Two samples are necessary because in the calculation of the matrix element the FSR photon from the top quark and ISR photon categories are connected via interference terms. This interference is neglected by using two samples. The detailed changes are shown in Appendix A.1.

In both samples a cut on the transverse momentum of the photon is applied. Only events with $p_T(\gamma) > 10$ GeV are selected.

After the generation of the samples, different variables were investigated. The most discriminant variables are described in Table 5.1 and were used in the training of a neural network.

These variables are shown in Figure 5.1 and Figure 5.2.

5. Development of a tool to distinguish between FSR and ISR photons

Variable	Description
$\cos(\theta)(\gamma)$	Cosine of the angle between the photon and the beam axis in the lab system.
$\Delta R(\gamma, \text{nearest } b)$	Calculated ΔR for the b quark and the b anti quark and took the minimum.
$\Delta R(\gamma, \text{nearest } W)$	Calculated ΔR for the W^+ and W^- and took the minimum.
$E(\gamma)$	Energy distribution of the photon.
$E(W)_{max}$	Calculated energy for W^+ and W^- and took the maximum.
$\eta(\gamma)$	Pseudorapidity η distribution of the photon.
$\eta(b)_{min}$	Calculated pseudorapidity η distribution for the b quark and the b anti quark and took the minimum.
$\eta(b)_{max}$	Calculated pseudorapidity η distribution for the b quark and the b anti quark and took the maximum.
$\eta(W)_{min}$	Calculated pseudorapidity η distribution for the W^+ and the W^- and took the minimum.
$\eta(W)_{max}$	Calculated pseudorapidity η distribution for the W^+ and the W^- and took the maximum.
$m_{Wb\gamma, max}$	Calculated the invariant mass of the W^+ , b and γ system and of the W^- , \bar{b} and γ system and took the maximum.
$m_{Wb\gamma, min}$	Calculated the invariant mass of the W^+ , b and γ system and of the W^- , \bar{b} and γ system and took the minimum.
$p_T(b)_{max}$	Calculated the transverse momentum of the b quark and b anti quark and took the maximum.
$p_T(W)_{max}$	Calculated the transverse momentum of the W^+ boson and W^- boson and took the maximum.

Table 5.1.: The variables used to train a neural network.

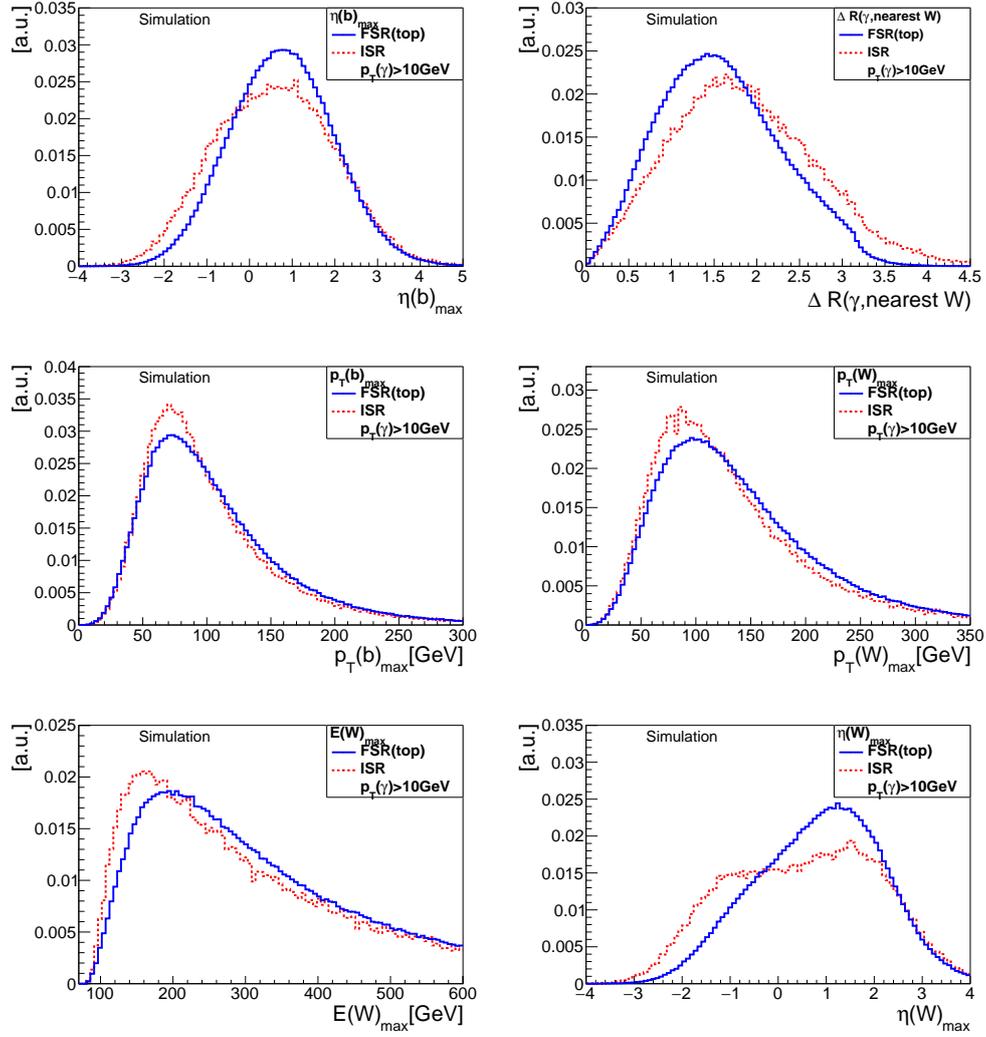


Figure 5.1.: Variables from the $t\bar{t}\gamma$ production process which were used to train a neural network. Description of the variables in Table 5.1

5. Development of a tool to distinguish between FSR and ISR photons

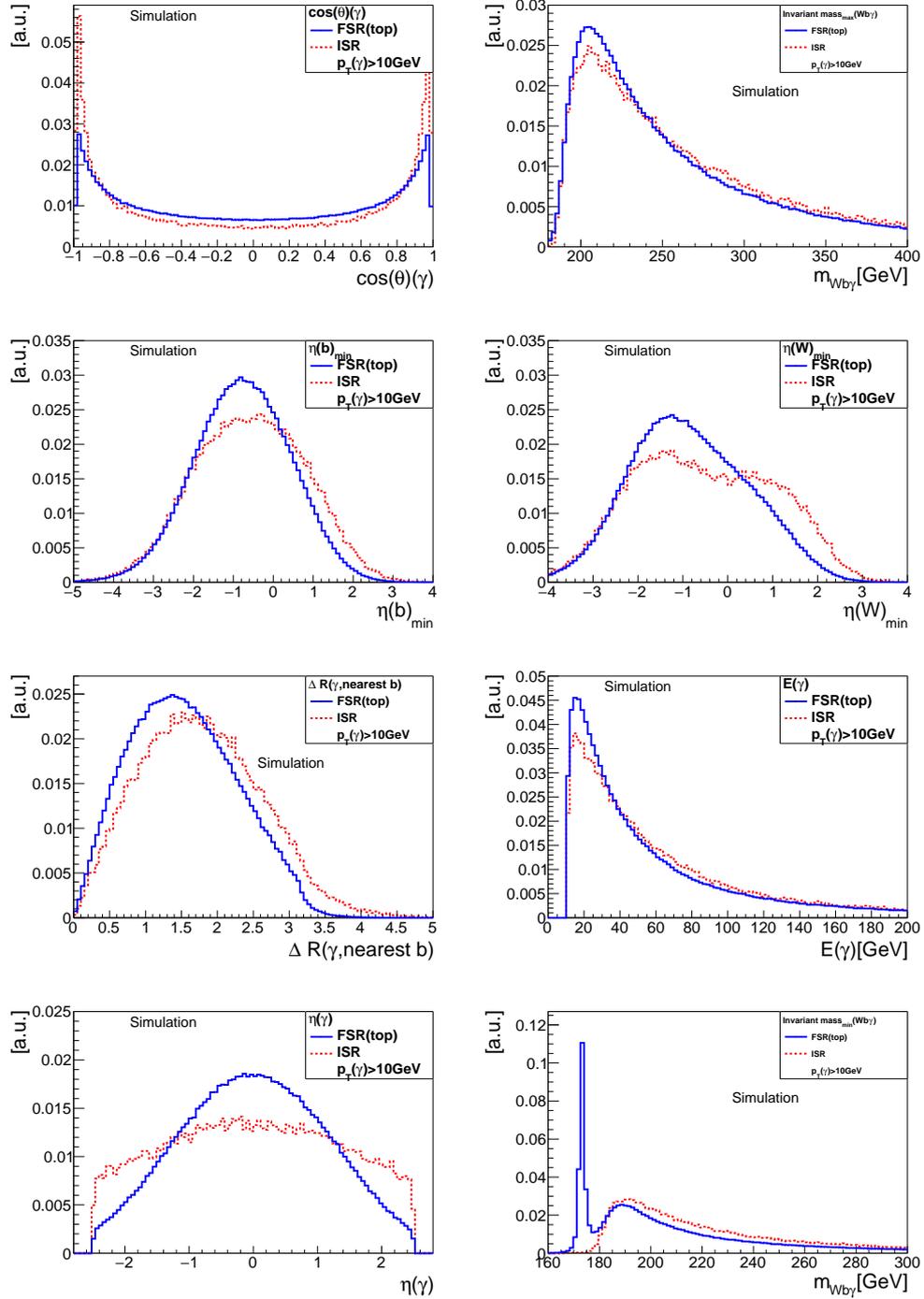


Figure 5.2.: Variables from the $t\bar{t}\gamma$ production process which were used to train a neural network. Description of the variables in Table 5.1

The main reason the variables look different is due to energy differences. In the case of ISR, the photon is emitted before the $t\bar{t}$ pair is produced. Therefore, less energy is available for the production and for the decay products of the top quark as well. This explains the differences in the variables of the decay products of the top quark. The difference in the angle variables come from the way the photon was emitted. ISR photons are more likely to have larger $|\eta|$ values because the initial quarks are moving in the beam direction ($|\eta| = \infty$). The top quarks on the other hand move in directions of smaller $|\eta|$ values and therefore the FSR photons move in this direction as well.

One additional comment is necessary for the variable of the minimum of the invariant mass of the W , b and γ system $m_{Wb\gamma}$. In the ISR case, basically the energy distribution of the photon added to the invariant mass of the top quark is shown. The photon was emitted before the production of the top quark. Therefore, the invariant mass of the W and b is equal to the invariant mass of the top quark. By adding the energy of the photon, this distribution is produced.

In the FSR case, the invariant mass of the W , b and γ system should be equal to the invariant mass of the top quark because the photon is emitted after the production of the top quark pair. But in addition to the invariant mass, this case also shows an energy distribution. The reason for that is, that two top quarks are produced and both can radiate off a photon. The goal of calculating the minimum of the invariant mass of the system was to find the top quark which actually emitted the photon. But sometimes the photon was matched to the other top quark and therefore a part of the distribution looks similar to the ISR case.

5.2. Training of a neural network

After the variables are selected, a neural network is trained to distinguish between the FSR and ISR events. Therefore, the data was split into two disjoint sets of training data and test data. 80% of the data was used for training and 20% for testing. The training data is used to actually train the neural network and the test data to evaluate if the neural network is able to classify unseen data correctly.

For evaluation, the ROC (Receiver Operating Characteristic) curves and the AUC (Area Under Curve) values are used. The ROC curves describe the background rejection in dependence of signal efficiency. These curves are plotted for both test and training data. If the neural network is not overtrained, these curves should be aligned. Hence, the ROC curves can indicate if the neural network is overtrained or not.

5. Development of a tool to distinguish between FSR and ISR photons

The AUC values are the integral of the ROC curves from 0 to 1. They can take a value between 0 and 1 and indicate how many events were classified correctly. A AUC value of 1 would indicate a perfect classifier and a AUC value of 0.5 means that the classifier is no better than tossing a coin.

The neural networks, which will be discussed in the following are trained on 210 973 FSR events and 75 618 ISR events. The FSR events had to be reduced down from 958 832 events in the conversion (see Section 4.2) and were randomly selected.

The samples are further separated into batches. Batches are used to reduce the computational effort for calculating the gradient of the loss function. With batches, the sum in Equation 4.1 does not take all samples into account but only as many as the batch size. After calculating the loss function, the gradient of it is calculated and all weights and biases get adjusted. Then the loss function is calculated for the next batch until all samples are used. Adjusting the weights and biases for all samples is called an epoch.

The advantage of using batches is that the computational effort gets reduced. The disadvantage is that for smaller batches the gradient of the loss function becomes less accurate. So a balance must be found when choosing the batch size. Here, a batch size of 10 000 was selected.

All models were trained on 100 epochs. ISR events were weighted by a factor of 2 to slightly increase their impact on the loss function. All neural networks have a feed-forward structure using the adam optimiser [44] and the binary cross-entropy loss function. The number of layers mentioned in the following are always referring to the number of hidden layers. In addition to that, every neural network has an input layer which is defined by the number of features and an output layer with only one node. The output layer uses the sigmoid activation function and therefore yield values between zero and one, where zero indicates an ISR event and one an FSR event.

In the first steps of the analysis, neural networks were trained without a real strategy just to get a feeling of how different architectures behave. The result was that the impact of the number of nodes and layers on the ROC curves and AUC values seem arbitrary. Neural networks with only one layer sometimes yield better results than neural networks with many layers. But this is because for every added layer, the dimension of the neural network increases. And with more dimensions, more samples are needed to describe that space. But since the number of samples is unchanged the accuracy for neural networks with many layers can go down. The result from that part was that it is always better to use BatchNormalization layers after every normal layer as shown in Table 5.2.

Then, a more systematic approach was made. At first, only one layer followed by a BatchNormalization layer was used.

Layers	Nodes (activation function)	AUC(train)	AUC(test)
1	64 (relu)	0.7517	0.7506
2	64 (relu), BN	0.7991	0.7987
1	64 (softmax)	0.7750	0.7739
2	64 (softmax), BN	0.8074	0.8083
1	46 (relu)	0.7714	0.7710
2	46 (relu), BN	0.8010	0.8027
1	46 (softmax)	0.7462	0.7411
2	46 (softmax), BN	0.8024	0.8002

Table 5.2.: Comparison of different architectures with or without BatchNormalization (BN) layers.

For that layer, different numbers of nodes and different activation functions were tested. After that, the best one-layered neural network was used to add another layer followed by a BatchNormalization layer. The AUC values of all trained neural networks are shown in Appendix A.2.

Since only neural networks which are not overtrained are useful, the difference of the AUC value for training and the AUC value for testing is calculated. Neural networks with differences of less than 0.002 were considered as not overtrained. The AUC values of the best neural networks are shown in Table 5.3.

The relevant value is the AUC (test) value because it indicates the performance of the neural network for unseen data. Therefore the neural network with 7 layers yields the best classification with an AUC (test) value of 0.8323. The output and the ROC curve for this neural network is shown in Figure 5.3.

This neural network can be used in future $t\bar{t}\gamma$ analyses to separate between the FSR and ISR events. Thus, the cross section measurements could become more precise since the ISR events which are background are no longer taken into account.

Layers	Nodes (activation function)	AUC(train)	AUC(test)
7	19 (relu), BN, 95 (softmax), BN, 29 (softmax), BN, 11 (softmax)	0.8307	0.8323
3	24 (relu), BN, 64 (softmax)	0.8269	0.8270
3	24 (relu), BN, 64 (relu)	0.8273	0.8271
5	24 (relu), BN, 64 (relu), BN, 48 (softmax)	0.8255	0.8252
4	64 (softplus), BN, 46 (softmax), BN	0.829	0.8282
4	64 (softplus), BN, 46 (selu), BN	0.8282	0.8286

Table 5.3.: The AUC values for the best neural networks. BN refers to a BatchNormalization layer.

5. Development of a tool to distinguish between FSR and ISR photons

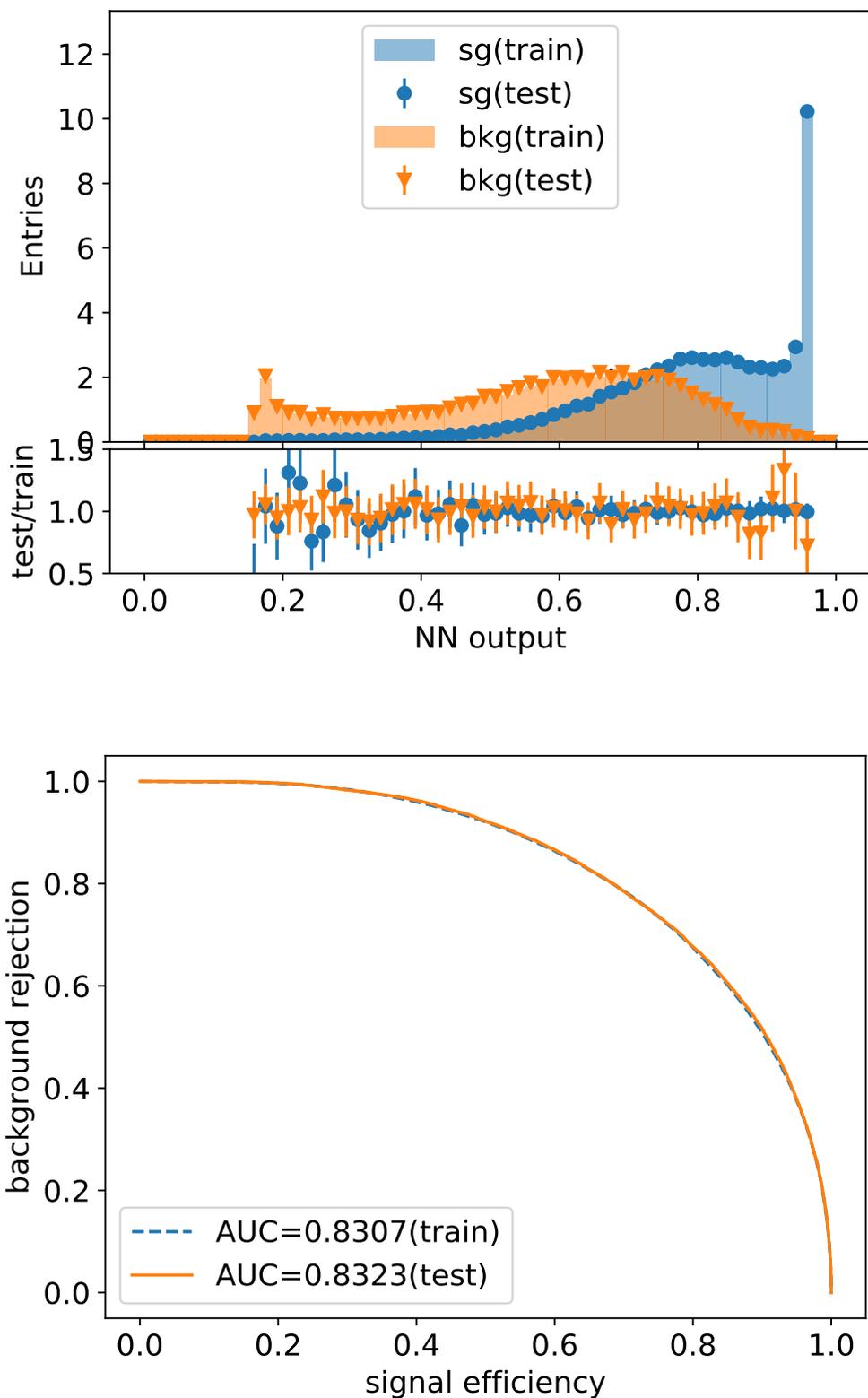


Figure 5.3.: The output and the ROC curve for the best neural network (7 layers).

6. Conclusion

A tool to distinguish between photons from FSR events and ISR events at parton level was successfully developed. An Area Under Curve (AUC) value, for the test sample, of 0.8323 is achieved which means that about 83% of all the input events were classified correctly. But further research is still necessary. The correlation between the variables should be checked because correlated variables have only a small effect on the classification compared to uncorrelated ones. By adding new uncorrelated variables the performance of the neural network can be increased. It is also necessary to run the neural network on real reconstructed data since the neural network was only trained and tested on simulated samples.

Furthermore, as it can be seen in Appendix A.2, an AUC value of about 0.8400 seems to be the maximal value. For better classification it is necessary to either look at new variables or use more advanced neural network algorithms than the ones used in this thesis.

But still, a separation between photons from FSR and ISR is possible and future analysis can use these results to get more precise measurements in the $t\bar{t}\gamma$ research.

A. Appendix

A.1. Changes in MadGraph

The MadGraph package was downloaded from the MadGraph homepage. The changes had to be applied in the following directory: `~/MadGraph/models/sm`. Three files had to be modified. The exact changes in each of these files are given below:

- In `couplings.py` define at the end of the file:

```
GC_109 = Coupling(name = ' GC_109',  
                  value = ' 2 * ee * complex(0,1))/3.',  
                  order = ' TGA' : 1)
```

- In `coupling_orders.py` added:

```
TGA = CouplingOrder(name = ' TGA',  
                    expansion_order = 99,  
                    hierarchy = 2)
```

- In `vertices.py` change the lines corresponding to `V_134` as follows:

```
V_134 = Vertex(name = ' V_134',  
               particles = [P.t__tilde__, P.t, P.a],  
               lorentz = [L.FFV1],  
               couplings = (0,0) : C.GC_109)
```

A.2. AUC values of trained neural networks

The tables show all the neural networks that were trained. The number in front of the l in the model name is referring to the number of dense layers in the network. The entries in parentheses refers to the number of nodes per layer and the activation function in that layer. 'soft' is the softmax activation function and 'sig' is the sigmoid activation function. 'with BN' means that the neural network has a BatchNormalization layer after every dense layer.

Model	AUC(train)	AUC(test)
3l(64relu,40soft,52soft) with BN	0.8311	0.8258
3l(64relu,40relu,52relu)	0.8053	0.8018
3l(64relu,40relu,52relu) with BN	0.8306	0.8263
4l(19relu,95relu,29relu,11relu) with BN	0.8119	0.8088
4l(19relu,95soft,29soft,11soft) with BN	0.8307	0.8323
1l(8soft)	0.7355	0.7335
1l(128relu)	0.7901	0.7871
1l(8relu)	0.7358	0.7324
2l(24relu,64soft) with BN	0.8269	0.8270
2l(24relu,64relu) with BN	0.8273	0.8271
3l(24relu,64relu,48soft) with BN	0.8255	0.8252
3l(24relu,64sig,48sig) with BN	0.8198	0.8198
4l(64relu,152relu,60relu,50relu) with BN	0.8376	0.8209
2l(24sig,64soft) with BN	0.8191	0.8207
3l(24soft,64sig,48sig) with BN	0.8117	0.8141
3l(24sig,64sig,48sig) with BN	0.8174	0.8157
4l(64relu,40relu,52relu,48soft) with BN	0.8315	0.8241
4l(64relu,40relu,52relu,82sig) with BN	0.8317	0.8266
4l(64relu,48soft,52relu,74sig) with BN	0.8182	0.8175
4l(64relu,48sig,52relu,74soft) with BN	0.8335	0.8292
5l(64relu,48sig,52relu,74soft,68relu) BN	0.8304	0.8159

A.2. AUC values of trained neural networks

Model	AUC(train)	AUC(test)
4l(78relu,62sig,66relu,88soft) with BN	0.8354	0.8253
1l(64relu) with BN	0.7991	0.7987
1l(64sig) with BN	0.8137	0.8111
1l(64soft) with BN	0.8074	0.8083
1l(46relu) with BN	0.8010	0.8027
1l(46sig) with BN	0.8145	0.8147
1l(46soft) with BN	0.8024	0.8002
1l(82relu) with BN	0.8064	0.8072
1l(82sig) with BN	0.8174	0.8129
1l(82soft) with BN	0.7730	0.7761
1l(128sig) with BN	0.8153	0.8173
2l(128sig,46soft) with BN	0.8310	0.8251
2l(128sig,46relu) with BN	0.8319	0.8231
2l(128sig,46sig) with BN	0.8258	0.8229
2l(128sig,64soft) with BN	0.8278	0.8248
2l(128sig,64relu) with BN	0.8294	0.8230
2l(128sig,64sig) with BN	0.8203	0.8202
2l(128sig,82soft) with BN	0.8312	0.8250
2l(128sig,82relu) with BN	0.8280	0.8242
2l(128sig,82sig) with BN	0.8230	0.8193
3l(128sig,1sig,46soft) with BN	0.8129	0.8133
3l(128sig,1sig,128sig) with BN	0.8138	0.8080
2l(82sig,46relu) with BN	0.8301	0.8245
3l(128sig,82relu,46soft) with BN	0.8341	0.8204
1l(46elu) with BN	0.8174	0.8136
1l(46lin) with BN	0.7357	0.7304
1l(46selu) with BN	0.8171	0.8128
1l(46softplus) with BN	0.8133	0.8154
1l(46softsign) with BN	0.8100	0.8103
1l(46tanh) with BN	0.8144	0.8102
1l(64elu) with BN	0.8143	0.8160
1l(64selu) with BN	0.8139	0.8122

A. Appendix

Model	AUC(train)	AUC(test)
1l(64softplus) with BN	0.8173	0.8158
1l(64softsign) with BN	0.8040	0.8034
1l(64tanh) with BN	0.8182	0.8152
1l(82elu) with BN	0.8155	0.8160
1l(82selu) with BN	0.8172	0.8137
1l(82softplus) with BN	0.8149	0.8159
1l(82softsign) with BN	0.8038	0.8027
1l(82tanh) with BN	0.8165	0.8147
2l(64softplus,46relu) with BN	0.8307	0.8260
2l(64softplus,46sig) with BN	0.8111	0.8133
2l(64softplus,46elu) with BN	0.8223	0.8225
2l(64softplus,46selu) with BN	0.8293	0.8243
2l(64softplus,46soft) with BN	0.8290	0.8282
2l(64softplus,46softplus) with BN	0.8091	0.8058
2l(64softplus,46softsign) with BN	0.8282	0.8254
2l(64softplus,46tanh) with BN	0.8149	0.8136
2l(64softplus,64relu) with BN	0.8302	0.8242
2l(64softplus,64sig) with BN	0.8087	0.8077
2l(64softplus,64soft) with BN	0.8289	0.8253
2l(64softplus,64elu) with BN	0.8229	0.8233
2l(64softplus,64selu) with BN	0.8282	0.8286
2l(64softplus,64softsign) with BN	0.8259	0.8228
2l(64softplus,64tanh) with BN	0.8250	0.8214
2l(64softplus,64softplus) with BN	0.8166	0.8172
2l(64softplus,82relu) with BN	0.8292	0.8263
2l(64softplus,82sig) with BN	0.8182	0.8205
2l(64softplus,82soft) with BN	0.8321	0.8280
2l(64softplus,82elu) with BN	0.8173	0.8182
2l(64softplus,82selu) with BN	0.8181	0.8160
2l(64softplus,82softplus) with BN	0.8144	0.8122
2l(64softplus,82softsign) with BN	0.8279	0.8253
2l(64softplus,82tanh) with BN	0.8184	0.8191
3l(64softplus,82soft,46relu) with BN	0.7997	0.7955

Bibliography

- [1] F. Abe, et al. (CDF), *Observation of top quark production in $\bar{p}p$ collisions*, Phys. Rev. Lett. **74**, 2626 (1995)
- [2] S. Abachi, et al. (D0), *Search for high mass top quark production in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV*, Phys. Rev. Lett. **74**, 2422 (1995)
- [3] L. Evans, P. Bryant, *LHC Machine*, JINST **3**, S08001. 164 p (2008)
- [4] 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 **07**, 079 (2014)
- [5] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22**, 579 (1961)
- [6] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19**, 1264 (1967)
- [7] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. **C680519**, 367 (1968)
- [8] ATLAS Collaboration, *Measurement of the W -boson mass in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Eur. Phys. J. **C78(2)**, 110 (2018)
- [9] M. Tanabashi, et al. (Particle Data Group), *Review of Particle Physics*, Phys. Rev. D **98**, 030001 (2018)
- [10] G. 't Hooft, M. J. G. Veltman, *Regularization and Renormalization of Gauge Fields*, Nucl. Phys. **B44**, 189 (1972)
- [11] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13**, 508 (1964)
- [12] F. Englert, R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13**, 321 (1964)
- [13] G. S. Guralnik, C. R. Hagen, T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13**, 585 (1964)

Bibliography

- [14] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. **B716**, 1 (2012)
- [15] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. **B716**, 30 (2012)
- [16] S. F. King, *Neutrino mass*, Contemp. Phys. **48**, 195 (2007)
- [17] A. Quadt, *Top quark physics at hadron colliders*, Eur. Phys. J. **C48**, 835 (2006)
- [18] M. Kobayashi, T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49**, 652 (1973)
- [19] ATLAS Collaboration, *Measurement of the mass difference between top and anti-top quarks in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, Phys. Lett. **B728**, 363 (2014)
- [20] D.-W. Jung, P. Ko, J. S. Lee, *Mass dependent top forward backward asymmetry in the effective Lagrangian approach: Addendum to ‘Model independent analysis of the forward-backward asymmetry of top quark production at the Tevatron’*, Phys. Lett. **B708**, 157 (2012)
- [21] CMS Collaboration, *Measurement of the charge asymmetry in top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV using a template method*, Phys. Rev. **D93(3)**, 034014 (2016)
- [22] ATLAS Collaboration, *Observation of spin correlation in $t\bar{t}$ events from pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, Phys. Rev. Lett. **108**, 212001 (2012)
- [23] T. Aaltonen, et al. (CDF), *Observation of Single Top Quark Production and Measurement of $|V_{tb}|$ with CDF*, Phys. Rev. **D82**, 112005 (2010)
- [24] CMS Collaboration, *Search for the flavor-changing neutral current interactions of the top quark and the Higgs boson which decays into a pair of b quarks at $\sqrt{s} = 13$ TeV*, JHEP **06**, 102 (2018)
- [25] ATLAS Collaboration, *Measurement of the $t\bar{t}\gamma$ production cross section in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, JHEP **11**, 086 (2017)
- [26] ATLAS Collaboration, *Observation of top-quark pair production in association with a photon and measurement of the $t\bar{t}\gamma$ production cross section in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, Phys. Rev. **D91(7)**, 072007 (2015)

- [27] G. Cortiana, *Top-quark mass measurements: Review and perspectives*, Rev. in Phys. **1**, 60 (2016)
- [28] CMS Collaboration, *Measurement of the semileptonic $t\bar{t}+\gamma$ production cross section in pp collisions at $\sqrt{s} = 8$ TeV*, JHEP **10**, 006 (2017)
- [29] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3**, S08003 (2008)
- [30] CMS Collaboration, *The CMS Experiment at the CERN LHC*, JINST **3**, S08004 (2008)
- [31] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, JINST **3**, S08002 (2008)
- [32] LHCb Collaboration, *The LHCb Detector at the LHC*, JINST **3**, S08005 (2008)
- [33] D. E. Groom, S. R. Klein, *Passage of particles through matter*, Eur. Phys. J. **C15(1)**, 163 (2000)
- [34] C. W. Fabjan, F. Gianotti, *Calorimetry for particle physics*, Rev. Mod. Phys. **75**, 1243 (2003)
- [35] W. S. McCulloch, W. Pitts, *A logical calculus of the ideas immanent in nervous activity*, Bull. Math. Biophys. **5(4)**, 115 (1943)
- [36] F. Rosenblatt, *The Perceptron: A Probabilistic Model for Information Storage and Organization in The Brain*, Psychol. Rev., 65 (1958)
- [37] D. E. Rumelhart, G. E. Hinton, R. J. Williams, *Learning Representations by Back Propagating Errors*, Nature **323**, 533 (1986)
- [38] G. Klambauer, et al., *Self-Normalizing Neural Networks*, CoRR (2017)
- [39] S. Ioffe, C. Szegedy, *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*, CoRR (2015)
- [40] R. Al-Rfou, et al. (Theano Development Team), *Theano: A Python framework for fast computation of mathematical expressions*, arXiv e-prints (2016)
- [41] M. Abadi, et al., *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, <https://www.tensorflow.org> (2015)

Bibliography

- [42] F. Chollet, et al., *Keras*, <https://keras.io> (2015)
- [43] B. Völkel, *Studies of the discrimination between prompt photons and hadron fakes using neural networks*, Masters thesis **II.Physik-UniGö-MSc-2017/07** (2017)
- [44] D. P. Kingma, J. Ba, *Adam: A Method for Stochastic Optimization*, CoRR (2014)

Acknowledgements

First of all, I want to thank Prof. Dr. Arnulf Quadt for the opportunity to write my thesis in his group and that he agreed to be my supervisor and first referee. I also want to thank Prof. Dr. Stan Lai that he agreed to be the second referee for this thesis.

Next, I would like to thank Dr. Royer Ticse Torres for his continuous support and who always helped me with the next steps in this thesis. And I also want to thank him for proofreading.

Then, I want to thank Joshua Wyatt Smith for helpful discussions about neural networks and also for proofreading.

Many thanks go to Andreas Kirchhoff who helped me a lot with getting started with root and also with keras. He saved me a lot of time and made especially the start of this thesis very enjoyable.

I also want to thank everyone in the Göttingen Top Group for helpful comments and discussions during the weekly status reports.

Next, I want to thank Noreen and Maximilian, who shared an office with me, for nice conversations and helpful tips.

Finally, I want to thank my family for their continuous support.

Special thanks go to Karo who was always there when I needed her and who helped me through many stressful days.

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 nichtbestandenenen Prüfung an dieser oder einer anderen Hochschule eingereicht wurde.

Göttingen, den 10. Oktober 2018

(Pascal Herrmann)