GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Fakultät für
Physik

Bachelor's Thesis

Optimierung neuronaler Netze zur Untersuchung effektiver Feldtheorien in der $t\bar{t}\gamma$-Produktion mit dem ATLAS-Experiment

Optimising neuronal networks to study effective field theories in $t\bar{t}\gamma$ production with the ATLAS experiment

prepared by

Gesine Grubert

from Oxford

at the II. Physikalisches Institut

Thesis number:	II.Physik-UniGö-BSc-2020/02
Thesis period:	30th March 2020 until 6th July 2020
First referee:	Prof. Dr. Arnulf Quadt
Second referee:	Prof. Dr. Stan Lai

Abstract

The top quark with its heavy mass is a promising candidate to find hints on physics beyond the Standard Model of Particle Physics. To study the top-photon coupling, decays of top-/antitop quark pairs in association with a photon ($t\bar{t}\gamma$) can be analysed. In this thesis, a neural network, which classifies events according to the photons origin, is applied on photons with non-Standard Model coupling. These samples have been created using an effective field theory approach to modify the photon couplings. Furthermore, a different neural network is trained to distinguish photons obeying Standard Model couplings from Beyond Standard Model photons.

Keywords: top quark, $t\bar{t}\gamma$ decay, neural network, effective field theories

Zusammenfassung

Aufgrund seiner hohen Masse ist das Top-Quark ein vielversprechender Kandidat, Hinweise auf neue Physik zu finden. Um die Top-Photon-Kopplung dahingehend zu untersuchen, kann die Produktion von Top-/Antitop Paaren in Assoziation mit einem Photon ($t\bar{t}\gamma$) genutzt werden. Effektive Feldtheorien beschreiben die modifizierten Top-Photon Kopplungen, welche für die Generation der Samples genutzt werden. Auf diese Samples werden dann neuronale Netze angewandt. Dabei wird ein (auf das Standardmodell trainiertes) neuronales Netz verwendet, um den Ursprung der Photonen zu finden, ein weiteres trainiert, um Standardmodell-Photonen von Photonen anderer Kopplung zu unterscheiden.

Stichwörter: Top-Quark, $t\bar{t}\gamma$-Zerfall, neuronale Netze, effektive Feldtheorien

Contents

1. Introduction

The top-quark is the heaviest particle within the Standard Model of particle physics (SM). The SM describes all phenomena observed at the Large Hadron Collider (LHC) very well but there are observations as for example CP-violation [1], neutrino oscillation [2] or dark matter [3], indicating that the SM is an approximation of an underlying theory. Neutrino oscillation and CP-violation has also been observed at accelerators. Even though, the observed CP-violation is too small to explain reality, which is one indication that there is physics beyond the SM.

In 1995, the top-quark was discovered by the DØ [4] and CDF [5] experiments at the TEVATRON. At the LHC, it is produced mainly in pairs mostly by gluon fusion but also via quark anti-quark annihilation. During the production and decay process, it is possible, that some particle emits a photon. If this photon is emitted by the top-quark, its analysis gives an insight to the top-photon coupling, which offers an opportunity to measure the top-quark charge directly and to test the nature of the coupling, e.g. if it is the vector coupling predicted by the SM or if additional axial-vector components or even tensor components are present. Since the photon can be emitted by any charged particle during the production and decay process of $t\bar{t}$ pairs and the resulting data looks similar, it is not easy to determine the origin of the photon.

ATLAS, as a top-quark factory, delivers huge amounts of data. To optimise the analysis, the aim of the following bachelor thesis will be to further develop the existing neural network, developed and trained by Andreas Kirchhoff within his master thesis [6] and Pascal Herrmann's bachelor thesis [7]. This neural network does already analyse the origin of the photon in $t\bar{t} + \gamma$ processes. The aim of this bachelor thesis will be to adapt the network to effective field theories (EFT) on truth level.

2. The Standard Model of Particle Physics

The SM is a renormalisable quantum-field theory with the following underlying group symmetry: $SU(3) \otimes SU(2) \otimes U(1)$. $SU(3)$ represents the strong interaction and $SU(2) \otimes U(1)$ the electroweak interaction. The (known) elementary particles are described by the SM as well. Gravity is not included in this theory.

2.1. Elementary particles

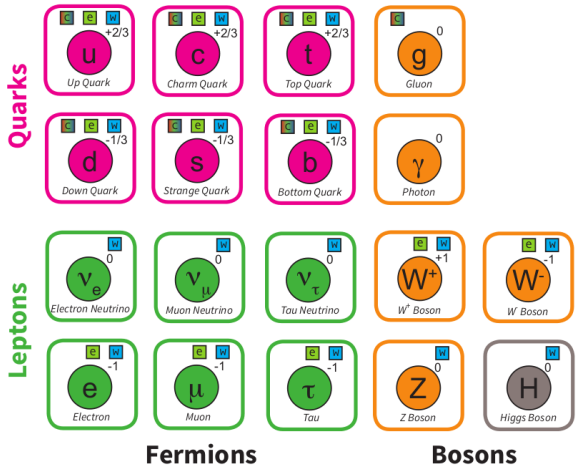

Figure 2.1.: Elementary particles of the SM.

Figure 2.1 shows the elementary particles. There are fermions, having spin $1/2$ and therefore, obeying the Fermi-Dirac-statistics, and bosons with integer spin, obeying the Bose-Einstein-statistics (see also section 2.2).
Fermions are divided into three generations (see Figure 2.1) which differ from each other in their masses. Within each generation, fermions are separated into quarks and leptons. Up-type quarks (first row in Figure 2.1) carry charge $2/3\,e$, where e is the absolute value of the electric charge of an electron. Down-type quarks (second row) carry $-1/3\,e$. Neutrinos

(third row) are neutral, the other leptons (electrons, muons and tauons in the last row) carry the charge $-1\,e$. An antifermion belongs to each fermion that has exactly the same properties as the fermion except for the charge-like quantum numbers, which are inverted. Antiparticles are denoted with a bar.

2.2. Interactions

The SM describes three out of four fundamental interactions, electromagnetism, the weak and the strong interaction. Bosons are the exchange particles transmitting the interactions. Photons are exchanged in electromagnetic processes. All fermions with electric charge may be part of those processes. In a weak interaction process, the $W^{\pm}$ or Z boson couples to the weak isospin. The name isospin refers to the fact that isospin is treated mathematically the same way as the spin. The third component of the isospin I_3 is $1/2$ for up-type quarks and neutrinos, whereas down-type quarks and electrons, muons and tauons have $I_3 = -1/2$. Therefore, all known fermions are able to interact weakly. The weak interaction violates parity symmetry. Neutrinos, as they only interact weakly, are maximum parity violating.
Gluons, the exchange particles of the strong interaction, couple to the colour charge. All quarks carry colour charge. This may be either red, blue, green or their anti colours. No free colour-charged particles exist, therefore, several quarks have to form colourless particles, the hadrons. This is called confinement. Hadrons appear as mesons (two quarks, one with a colour, one with the corresponding anticolour) or baryons (three quarks carrying all different colours or anticolours, respectively). Through interaction with the Higgs field, the particles obtain their mass. The Higgs boson is the excited state of the Higgs field. The SM without the Higgs field or a similar mechanism could not describe mass at all in a renormalisable way and would therefore be no valid theory.

2.3. The top quark

The top quark's existence was predicted in 1973 by Makoto Kobayashi and Toshihide Maskawa, who tried to extend the SM to include theories allowing CP violation [8]. In 1995, it was discovered by CDF [5] and DØ [4] experiments. Its mass $m_{\mathrm{top}} = 173.3 \pm 0.4\,\mathrm{GeV/c^2}$ [9] is, compared to the other quarks, very large. In consequence, its decay width is 1.35 GeV and accordingly its lifetime is $5 \cdot 10^{-25}\,\mathrm{s}$. So, the top quark decays before it is able to hadronise (to form hadrons with other quarks). Hence, a top quark passes its properties directly to its decay products and thus, those properties can be reconstructed

from the properties of the decay products. Because of its special role among the quarks it is a good candidate for testing the SM and theories beyond the SM as done in this thesis.

2.3.1. Production of top quarks

To produce the top quark one needs at least the energy corresponding to the top quark mass and the bottom quark, or, in $t\bar{t}$ decay twice the top quark mass (see Figure 2.2). This energy has only been reached in hadron collisions up to now. Via the exchange of a W Boson (weak interaction), a top and an anti-bottom quark are produced. Another option is the creation of a $t\bar{t}$ pair via the strong interaction. This happens when two gluons or a quark and its anti-quark annihilate. In case of the LHC the gluon fusion happens more often, since at the energy range required to produce a top quark pair, the gluons dominate the parton distribution functions of the proton compared to the fraction of quarks and anti-quarks.

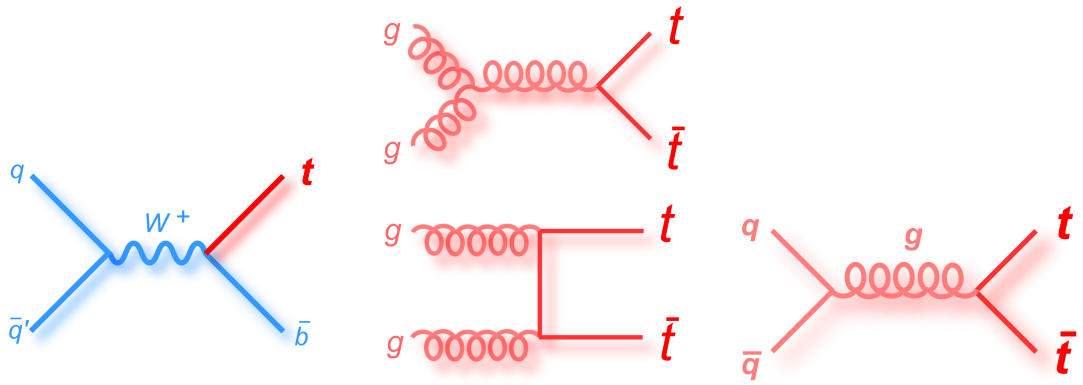

Figure 2.2.: On the left, the single top quark production by the decay of a W^+ boson (here the s channel) is shown. In the middle and on the right, one can see the $t\bar{t}$ production by gluon fusion and $q\bar{q}$ annihilation.

2.3.2. Decay modes of the top quark

The top quark does not build hadrons since it decays before it could form a bound state with other quarks. Top quarks decay via the weak interaction into a W boson and a b quark (see Figure 2.3). These b quarks form jets due to the confinement of hadrons, which can be detected in the hadronic calorimeter. They can be separated from c and light jets using the technique of b-tagging (see also [10]). Since the b quark is long lived on average ($\approx 10^{-12}$ s) it can possibly travel some millimetres in the detector. Therefore, the b jet could be slightly displaced with respect to the interaction point. W bosons decay

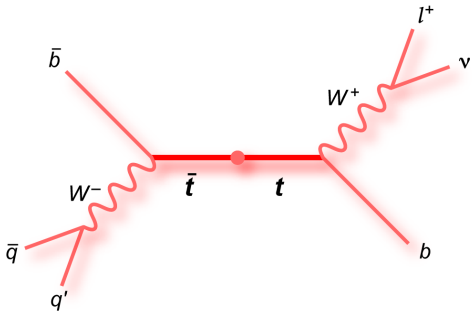

Figure 2.3.: In this Feynman graph, a $t\bar{t}$ pair decays via the weak interaction into $b/\bar{b}$ quarks. Then, the W bosons decay into a $q\bar{q}$ pair and a lepton and its neutrino.

either hadronically into two quark jets or leptonically, one lepton and the corresponding neutrino. Accordingly, a $t\bar{t}$ event may result in either two b-jets and four jets (all hadronic: 45.7 %), two b-jets, two jets, one charged lepton and missing transverse energy (indicating a neutrino) (43.8 %) or two b-jets, two charged leptons and large missing transverse energy (indicating two neutrinos) (10.5 %) [9].

2.3.3. $t\bar{t}\gamma$

Within the process of $t\bar{t}$ production and decay, one of the charged particles can emit a photon. This can be an initial state quark as well as the top quark or any (charged) decay product. The final states are similar to the $t\bar{t}$ processes but with one additional photon. Analysing the properties of the photon gives an insight into the top-photon coupling, since the top-photon vertex is studied directly.

2.4. Effective field theories

Effective Field Theories (EFT) are thought to be embedded in a greater theory being a low energy approximation. Taking the example in Figure 2.4 for a very virtual W boson, the parameter G_F combines the whole interaction without considering each vertex individually [11]. Thus, the 4-fermion Fermi coupling is a low-energy approximation of the weak interaction. Similarly, an EFT approach can be used to extend the SM. New physics at a generally large energy scale could lead to an effective modification of SM couplings at lower scales even though the effect will be larger for larger energies. The Lagrange density, giving the energy density of the space and time dependent fields, is modified. In case of the top-photon coupling, such new physics contributions can be parametrised with three generalised couplings, scaled with the Wilson coefficients and normalised to

an energy scale of Λ^2 as it is shown in Equation 2.1. Usually, $\Lambda^2 \approx 1$ TeV^2 because we expect new physics to appear around this scale. c_{tG} is the coupling corresponding to the top-gluon coupling. In the $t\bar{t}\gamma$ process, top-gluon coupling happens in $t\bar{t}$ production. Therefore, the modified top-gluon coupling should have an influence on the total rate of $t\bar{t}\gamma$ production but not on the top-photon coupling. c_{tW} and c_{tZ} are both electroweak couplings. Therefore they should have an influence on the top-photon coupling. Each Wilson coefficient has a negative and a positive value. In this thesis, ctw_m refers to the negative value of c_{tW} as listed in Table 2.1, whereas ctw_p to the positive one and ctw_0 to set the coefficient equals zero. These values are calculated in the thesis of Sreelakshmi J S [12].

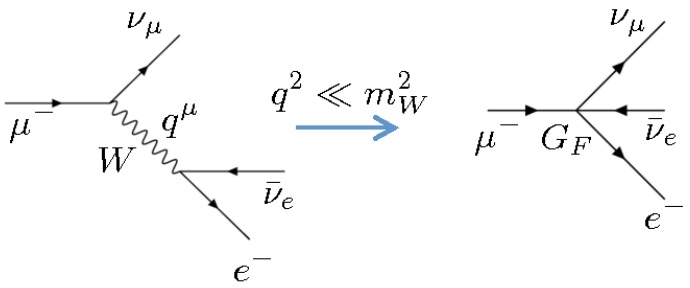

Figure 2.4.: Muon decay as an example for a low-energy approximation of the weak interaction [13].

$$\mathcal{L} = \mathcal{L}_{SM} + \frac{c_{tw}}{\Lambda^2} O + \frac{c_{tz}}{\Lambda^2} O + \frac{c_{tg}}{\Lambda^2} O + ... \tag{2.1}$$

Wilson coefficient	positive value	negative value
c_{tW}	1.0	-1.2
c_{tZ}	1.4	-1.0
c_{tG}	0.4	-0.4

Table 2.1.: Wilson coefficients for each coupling constant established in the thesis of Sreelakshmi J S [12].

3. Experimental setup

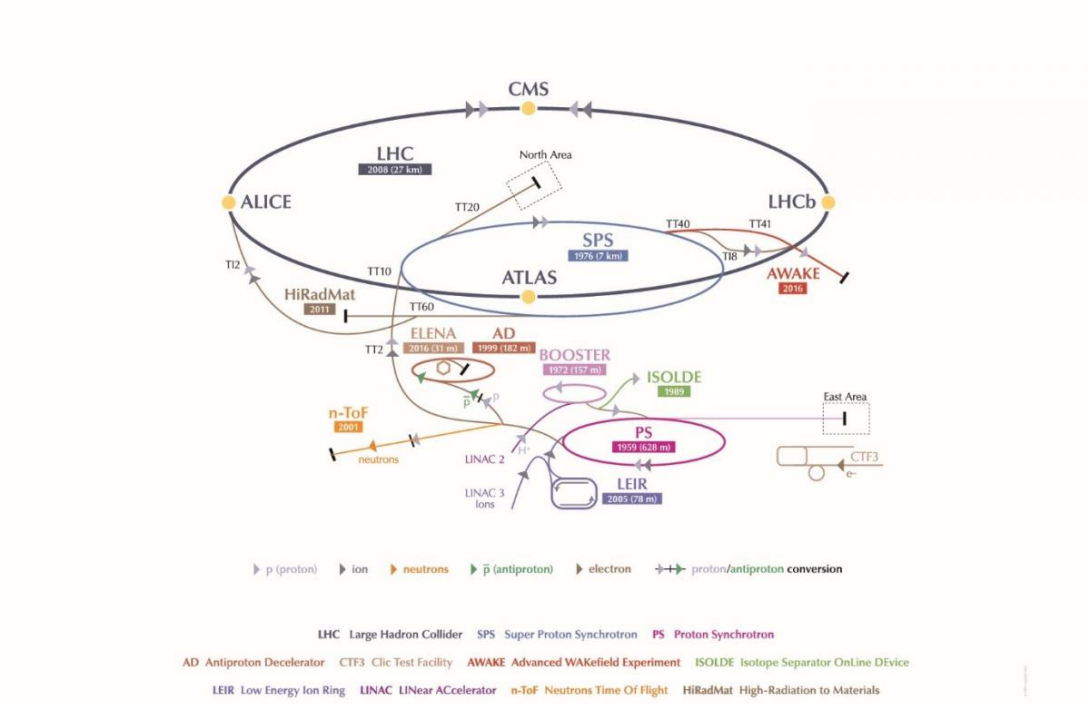

Figure 3.1.: The CERN accelerator complex, located at Geneva, Switzerland ©CERN.

For the moment, the European Organisation for Nuclear Research (CERN) runs the largest hadron collider with 27 km circumference, the Large Hadron Collider (LHC). Figure 3.1 shows an overview of the different accelerators and the experiments. A bottle of hydrogen (replaced twice a year) feeds the accelerators with protons. The protons are separated from the electrons by a strong electric field. They are accelerated in several steps by a linear accelerator, then with circular accelerators: the Synchrotron Booster, the Proton Synchrotron and the Super Proton Synchrotron which increases the beam energy to 450 GeV. Finally, the protons are fed into two different beam pipes. It takes around 20 minutes to accelerate them until each beam reaches its maximum energy of 6.5 TeV [14]. Therefore, the centre of mass energy is 13 TeV, which is nowadays the maximum.

A strong magnetic field keeps the proton beam on track. The protons are brought to collision at four points, where the four major detectors are installed, ATLAS, CMS, ALICE and LHCb. The proton bunches, containing up to 10^{11} protons, are brought to collision at a frequency of 40 MHz. On average 50 to 60 collisions take place in each bunch crossing. ATLAS has a design luminosity of $10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ [15]. The integrated luminosity over the full Run 2 $p\bar{p}$ data sample is 139 fb^{-1} [16]. The Luminosity measures the likelihood of collisions by counting the bunches and protons per bunch, time and area. The high luminosity provides a huge number of events, which requires good detectors and efficient data analysis.

3.1. The Atlas detector

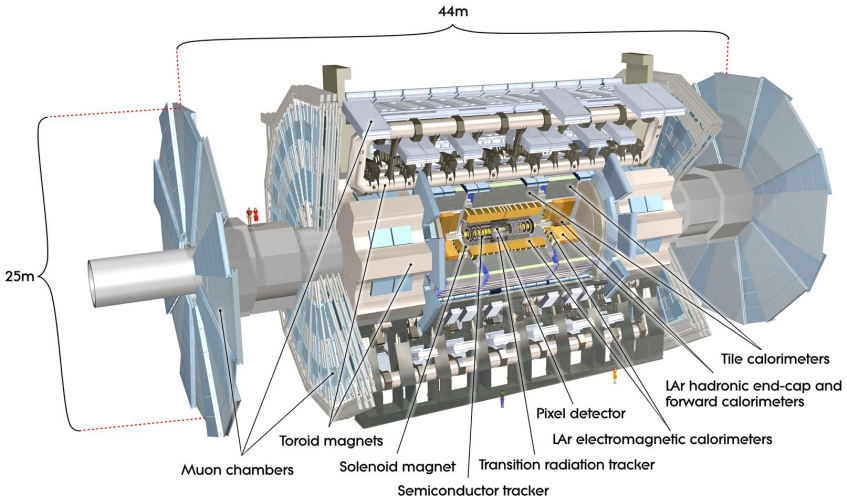

Figure 3.2.: The onion like built ATLAS detector ©CERN.

ATLAS stands for A Toroidal LHC ApparatuS [15]. It is the largest particle detector (in volume). For analysing the data, a (cylindrical) coordinate system is defined. The z axis points along the beam. The angle ϕ is defined to go around the beam and the angle θ is in between the z axis and the scattered particle. Instead of θ, it is more useful to use the pseudorapidity $\eta = -\ln\left(\tan(\theta/2)\right) = 1/2 \ln\left(\frac{|\mathbf{p}|+p_L}{|\mathbf{p}|-p_L}\right)$ [15], where $\mathbf{p}$ is the 3-momentum of the photon and p_L the longitudinal momentum. The differences in pseudorapidity are Lorentz invariant under boosts in z direction. Another property used is the difference in the $\eta - \phi$-plane defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, which is also Lorentz invariant under boosts in z direction.

Inner detector: The inner detector tracks electrically charged particles and measures their momentum and charge using a magnetic field in z direction. In Figure 3.2 the components of the inner detector are shown: The pixel detector (resolution: 14×115 μm^2), the Semiconductor Tracker made of silicon micro-strips and readout strips, measuring the position of charged particles with an accuracy of 17 μm and the Transition Radiation Tracker which basically separates pions from electrons and positrons. The end caps measure particles with a high pseudorapidity, e.g. particles flying almost parallel to the beam [15].

Calorimeter: The calorimeters measure the total energy of the particles by forcing them to deposit their energy in the detector by ionisation, scintillation or Cherenkov radiation. There are electromagnetic calorimeters measuring the energy of electrons, positrons and photons and hadronic calorimeters measuring hadrons. In both calorimeters the energy deposition is achieved by showers, which result from the interaction of incoming particles with the dense calorimeter material. Calorimeters consist of passive material, which induces the showers and active material measuring the signal. The electromagnetic calorimeter uses lead and stainless steel as a passive and liquid Argon as an active material. Hadronic Calorimeters use iron and plastic scintillation tiles.

Muon system: Muon chambers are the outermost layer of the detector. They are gas detectors. Muons pass the other detectors unseen except of the tracking detector. Together with the momentum measurement in the muon chamber, a very good momentum resolution is obtained.

Neutrinos cannot be seen by any of the detectors. Using missing transverse momentum and given the W boson mass, the neutrino momentum can be reconstructedwith a twofold ambiguity.

Trigger systems separate the interesting[1] processes. First, a hardware based trigger system defines so called regions of interest (RoI) in around 2.2 μs and reduces within this process the amount of data to a maximum of ≈ 100 kHz. The High Level Trigger is software based and reduces the data to 1 kHz, this takes around 4 s. The now separated interesting data (and samples of the not interesting ones, for cross-checking) are stored at the CERN computer centre and ready to be analysed [17].

[1]Interesting here means: Processes of high transverse momentum, as this indicates processes containing top quarks, Higgs bosons or (if they exist) new particles like SUSY particles.

4. Neural networks

A neural network (NN) is used for classification. Feedforward NNs are built in layers, consisting of nodes. Each node of a layer is connected to all the nodes in the previous and the following layer, as indicated in Figure 4.1. Each node of the input layer represents one parameter $\mathbf{x_i}$ given to the NN. The nodes in the output layer correspond to the categories in which the data will be divided. High values mean, the sample matches the category with a high probability.

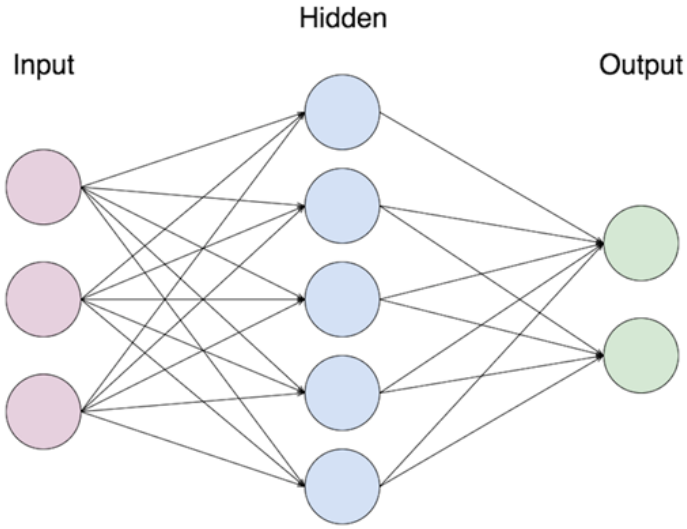

Figure 4.1.: A feed-forward neural network. The input nodes correspond to the (here) three features given to the network, the output layer (with two nodes) gives the answer. The hidden layer has five nodes in this picture. It is not accessible by the user.

Hidden layers are in between the input and the output layer. Each of the 5 nodes in Figure 4.1 carries a value which is the weighted sum of the nodes of the previous layer (Equation 4.1), therefore, the name feedforward NN.

$$f(\mathbf{x}) = \sigma\left(\sum_{i=1}^{N} w_i x_i + C\right) \text{[18]}. \tag{4.1}$$

The σ function squeezes the weighted sum between an upper and a lower bond. tanh or

sigmoid functions are often chosen to stay between -1 and 1 and 0 and 1, respectively. w_i are the weighting factors, corresponding to one connection. C is called bias. The NN learns, when it knows which combination of sub-components corresponds to which output, when bias and weights are well chosen. Choosing weights and bias is done in the training process, where the NN is fed with known data and weights and bias are adjusted. The validation loss function is the function, which has to be minimized by the NN by changing weights and bias.
NNs are especially helpful if there are serveral correlated variables with small but significant separation power. Then, an NN is often able to combine given information and result in a better separation power since it is able to use higher dimensional correlations, which would not be possible to see by simple comparison of the data. There are some functions and values to determine the quality of an NN. The receiver operating characteristic (ROC) curve illustrates the ability of a binary classifier to sort events into the correct category. It shows the background rejection over the signal efficiency. A perfectly working binary classifier would have as a ROC curve a horizontal line at 1.0 (all background is rejected) until a signal efficiency of 1.0 (e.g. total efficiency). The area under [the ROC] curve (AUC) value is therefore a measurement of the ability of the binary classifier to assign correct values to events. Its best value equals one. With aid of this graphs, one can decide how much background one has to accept to have a certain amount of signal within the data. The loss function is important to determine wether it is worth to increase the number of epochs during the training. The training improves further, when the loss function is decreasing. Kolmogorov Smirnov statistic plots test wether the predictions of the training data match the predictions of the testing data. This is useful to check over- and undertraining, which can happen, when the NN gets sensitive to statistical fluctuations. The perfect output would have all signal datapoints at 1 and all background datapoints at zero. Test and training should give almost equal results, where the results from the training are expected to be slightly better, since the NN was trained on them. There are some methods to increase the performance of an NN. Adding dropout layers prevents overfitting (rather fitting to a more simple function than a more complicated, it means technically, that the cost to increase the complexity of a function is rather high). Also, the choice of the working point influences the result. A working point is a cut, where a chosen percentage of events are given a probability to belong to their (true) category. The higher the working point, the more background will also be (mistakenly) classified into this category. Finally, since the NN may end up in a minimum, k-fold cross validation can be used. There, the data set is devided into k parts. Then, one part is trained against the other k-1 parts. This is repeated k times and one is able to check wether all outputs

are similar to each other.

5. Monte Carlo samples

5.1. Description of the samples

This analysis uses samples generated with MADGRAPH 2.6.7 which contain decays of $t\bar{t}$ pairs in association with a photon in the lepton + jets channel with different configurations of the Wilson coefficients (see: section 2.4). Each of the 27 different samples contains 100 000 events with around 94 000 events with full particle content, the other 6 000 do not have full particle content. In the samples, gluons, initial state quarks, top quarks, b quarks and W bosons may be the mother particle of a photon. Here, one needs to pay attention to the assignment of the photon. Physically, photon radiation from a gluon does not exist. The "gluon radiation" corresponds to the emission of a photon from a virtual top quark (see Feynman graph in Figure 2.2 middle, bottom). In the sample generation with MADGRAPH, photons are assigned to the last on-shell particle. Therefore, the photons from off-shell top quarks may have gluons as "mother" particles. The same happens for b quarks. Final state b quarks have to be on-shell (forced by MADGRAPH), in consequence, all radiating b quarks are off-shell and the photon is assigned to the on-shell top quark. Thus, in the analysis photons from on-shell top quarks and from b-quarks are treated as one category. Then, there is final state radiation from the W boson (FSR(W)) and initial state radiation (ISR) coming from initial state light quarks. All samples are normalised to the cross-section of the respective mode. For the analysis, the reference frame is the laboratory frame.

5.2. Description of the variables

To comprehend the different properties of a photon emitted under modified coupling, one has to analyse the behaviour of the variables. The studied variables are explained in the following list:

- $p_T(\gamma)$ is the transverse momentum of the photon
- $E(\gamma)$ is the energy of the photon

- $\Delta\eta(\gamma)$ is the difference in pseudorapidity of the photon and the nearest top quark as defined in section 3.1
- $\Delta R(\gamma$, top quark) is the distance in the $\eta - \phi$ plane between the photon and the closest top quark as defined in section 3.1
- $\Delta R(\gamma$, b quark) is the distance in the $\eta - \phi$ plane between the photon and the closest b quark as defined in section 3.1
- $\cos(\theta)(\gamma)$ is the angle between the z axis and the scattered particle (see: section 3.1
- $m_{\text{max}}(\text{Wb}\gamma)$ is the maximum invariant mass of a W boson, a b quark and the photon (photon comes from either ISR, top quark or one of its decay products).
- $m_{\text{min}}(\text{Wb}\gamma)$ is the minimum invariant mass of a W boson, a b quark and the photon
- $m_{\text{max}}(\text{Wb})$ is the maximum invariant mass of a W boson and a b quark
- $m_{\text{min}}(\text{Wb})$ is the minimum invariant mass of a W boson and a b quark.

The variable distributions of $p_T(\gamma)$, $E(\gamma)$, $\Delta R(\gamma$, top quark) and $m_{\text{max}}(\text{Wb}\gamma)$ are plotted in Figure 5.1. One can see the distribution of the SM sample as well as the one for EFT sample $\text{ctw}_m\text{ctz}_p\text{ctg}_m$[1], and the ratio between both. This sample is chosen, because it shows the highest deviation from the SM sample[2]. In Figure 5.1 the black histogram corresponds to SM simulation and the pink one to EFT simulation. The difference in the shape of the samples is for $p_T(\gamma)$, $E(\gamma)$ and $m_{\text{max}}(\text{Wb}\gamma)$ a slightly broader distribution for the EFT sample, what can be nicely observed, looking at the ratio plot. This is in agreement with the expectations. As explained in section 2.4, an EFT parametrises the influence of new physics at high energy scales. Therefore, the influence of modified couplings is higher, the higher the energy and photons obeying an EFT may be noticed easier the higher the energy of the emitted photon. Taking this into account, all kinetic variables should show the same behaviour. Indeed, the EFT distribution of $E(\gamma)$ is slightly shifted towards higher energies. Since for higher energies, the average opening angles are larger, $\Delta R(\gamma$, top quark) and $\Delta R(\gamma$, b quark) should also be shifted towards higher values,

[1]Remark on notation: *ctw* refers to the Wilson coefficient of the operator representing the top-W boson coupling, *ctz* represents the top-Z coupling and *ctg* the top-gluon coupling. m, p and 0 refer to the negative and positive values listed in Table 2.1 or to set the coupling equals zero respectively.

[2]The ratio plots for the other variable distributions as well as the ones for a different sample may be found in Figure A.1 and Figure A.3

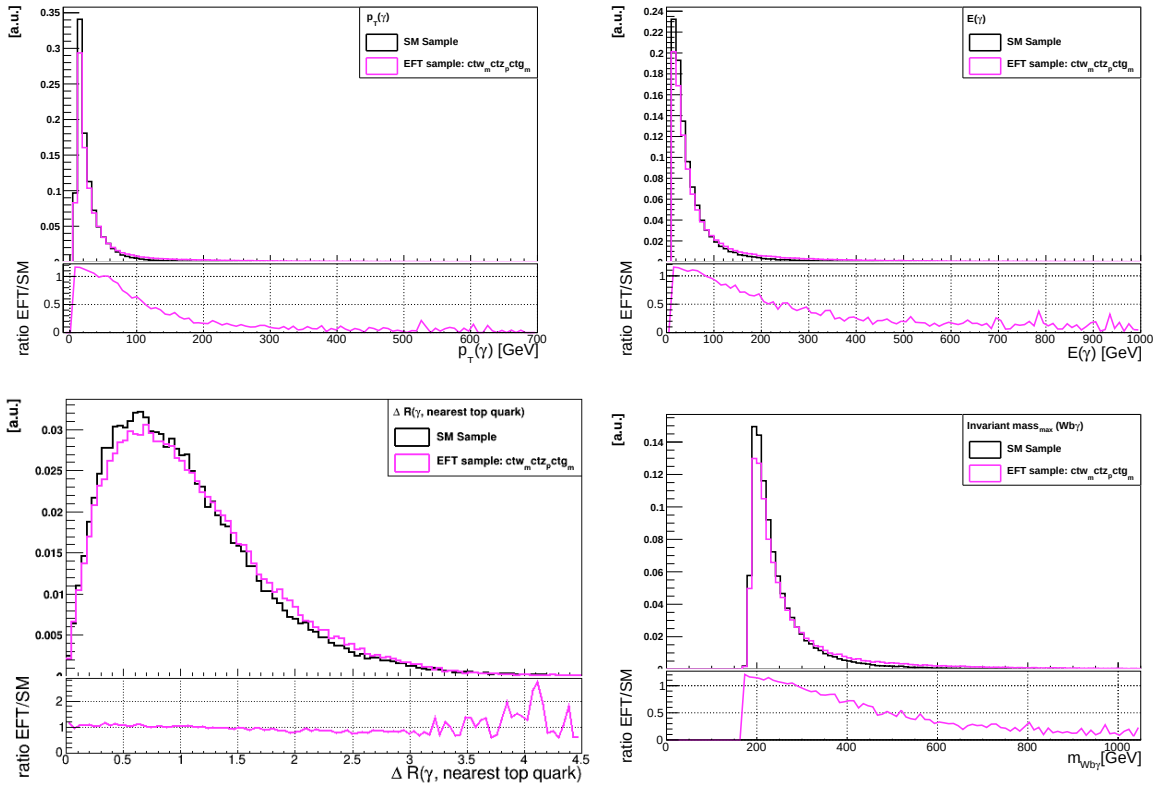

Figure 5.1.: The ratio plots for the variable distributions for $p_T(\gamma)$, $E(\gamma)$, $\Delta R(\gamma$, top quark) and $m_{\max}(\mathrm{Wb}\gamma)$ for the SM and for $\mathrm{ctw}_m\mathrm{ctz}_p\mathrm{ctg}_m$.

but this is not the case, the fluctuations in the ratio plots are due to less data points for higher energies. For the invariant mass, the reasoning is different. If the coupling is modified (with respect to SM) the distribution of mother particles of the photon may be different and hence, the invariant mass.

5.3. Separation power

The more the variable distributions (see: section 5.2) differ between EFT and SM samples, the better the NN will be able to separate the variables and sort them correctly into those categories. To compare the separation power of the variables, one can define the separation S:

$$S = \frac{1}{2}\sum_{i=0}^{N}\frac{(s_i - b_i)^2}{s_i + b_i}.$$

Here, s_i is the normalised signal contribution and b_i the normalised background contribution. The sum is over all bins. Table 5.1 shows the separation power for each variable for

three different samples. $\mathrm{ctw}_m\mathrm{ctz}_p\mathrm{ctg}_m$ and $\mathrm{ctw}_m\mathrm{ctz}_p\mathrm{ctg}_0$ are the samples with the highest and second highest separation power. One further sample has separation powers in the range of those two: $\mathrm{ctw}_m\mathrm{ctz}_p\mathrm{ctg}_p$ (separation power of $p_T(\gamma) = 5.64\%$). The third sample in Table 5.1, $\mathrm{ctw}_p\mathrm{ctz}_p\mathrm{ctg}_p$, has a separation power alike the ones of the other 24 samples[3]. Naively, one could imagine that this sample (together with $\mathrm{ctw}_m\mathrm{ctz}_m\mathrm{ctg}_m$) has the highest separation power since all operators are oriented in the same direction and so, would have the highest influence on the sample distributions. But this is not the case.

Variable\Sample	$\mathrm{ctw}_m\mathrm{ctz}_p\mathrm{ctg}_0$ [%]	$\mathrm{ctw}_m\mathrm{ctz}_p\mathrm{ctg}_m$ [%]	$\mathrm{ctw}_p\mathrm{ctz}_p\mathrm{ctg}_p$ [%]
$p_T(\gamma)$	6.75	7.88	0.03
$E(\gamma)$	5.15	6.15	0.03
$\eta(\gamma)$	0.06	0.06	0.01
$\Delta\eta(\gamma$,top quark)	0.06	0.09	0.02
$m_{\max}$(Wb)	0.10	0.20	0.01
$m_{\min}$(Wb)	2.27	2.86	0.02
$m_{\max}$(Wbγ)	5.71	5.74	0.02
$m_{\min}$(Wbγ)	7.49	7.51	0.02
ΔR(b quark)	0.22	0.27	0.03
ΔR(top quark)	0.55	0.64	0.04
$\cos(\theta)$	0.07	0.09	0.03
$\eta_{p_T,\,\mathrm{highest}}$	0.06	0.06	0.03
$\eta_{p_T,\,2^{\mathrm{nd}}\,\mathrm{highest}}$	0.03	0.03	0.02
$\eta_{\mathrm{smallest}\,\lvert\eta\rvert}$	0.04	0.05	0.04
$\eta_{2^{\mathrm{nd}}\,\mathrm{smallest}\,\lvert\eta\rvert}$	0.04	0.04	0.03
event(top)	5.36	6.42	0.01
event(isr)	4.45	5.34	0.02
event(W)	2.92	3.64	0.02
event(b)	2.90	3.62	0.02

Table 5.1.: Separation power of three different samples, the meaning of the variables is explained in section 5.2.

5.4. Analysis of EFT samples

There are three samples having a significantly higher separation power in $p_T(\gamma)$ than the other ones. These are $\mathrm{ctw}_m\mathrm{ctz}_p\mathrm{ctg}_0$ ($S = 6.75\%$), $\mathrm{ctw}_m\mathrm{ctz}_p\mathrm{ctg}_m$ ($S = 7.88\%$) and $\mathrm{ctw}_m\mathrm{ctz}_p\mathrm{ctg}_p$ ($S = 5.64\%$). Apparently, $\mathrm{c_{tG}}$ has a small influence on the variable distributions. The same can be observed for other sample combinations where $\mathrm{c_{tG}}$ is modified. That makes perfectly sense, since $\mathrm{c_{tG}}$ has influence only on the top-gluon coupling and

[3]The separation power of those other samples can be found in Appendix B.

therefore on the production of $t\bar{t}$ but not the decay. Working with the separation power values[4] for $p_T(\gamma)$, one can see, that c_{tW} differs a lot from the SM if set to -1.2 (see sample $ctw_0ctz_pctg_p$ ($S = 0.55\%$), $ctw_mctz_pctg_p$ ($S = 5.64\%$) and $ctw_pctz_pctg_p$ ($S = 0.03\%$)). However, for the samples $ctw_0ctz_mctg_m$ ($S = 0.39\%$), $ctw_mctz_mctg_m$ ($S = 0.06\%$) and $ctw_pctz_mctg_m$ ($S = 2.17\%$), the graduation in separation power is reversed. Obviously, the separation power does not depend only on the value of some coupling but also on interferences between the different couplings. Varying the last coupling, c_{tZ}, also leads to changes in the separation from the SM sample. Looking for example at the two sample sets $ctw_mctz_pctg_m$ ($S = 7.88\%$), $ctw_mctz_0ctg_m$ ($S = 1.63\%$) and $ctw_mctz_mctg_m$ ($S = 0.06\%$) and $ctw_pctz_pctg_p$ ($S = 0.03\%$), $ctw_pctz_0ctg_p$ ($S = 0.37\%$) and $ctw_pctz_mctg_p$ ($S = 1.82\%$), one can see that the differences between one sample set is higher than for c_{tG}, so this coupling has a higher influence on physical behaviour than c_{tG}. Here, the variation with the negative value for c_{tZ} has the highest separation power whereas in the other sample it is the positive value. That means one cannot say wether a positive or negative value has a bigger impact. The interaction between c_{tZ} and c_{tW} seems to be responsible for high differences between the EFT and the SM sample. Opposite signs for two couplings within a sample lead to higher separation from the SM. This explains why neither $ctw_mctz_mctg_m$ nor $ctw_pctz_pctg_p$ have high separation values. Naively, one could expect that those two could have the biggest difference from SM but apparently this is not true.

[4]See: Appendix B

6. Application of a neural network to separate the origin of the photon to EFT samples

For the analysis an already existing neural network (NN) is used. It is able to separate top quark radiation from initial state radiation (ISR), final state radiation from b quarks (FSR(b)) and FSR(W). This NN will be referred to as $\mathrm{NN}_{t\bar{t}\gamma}$. It is more sensitive to the correct classification of top radiation since the top events were overrepresented (with respect to the cross-section) in the training. $\mathrm{NN}_{t\bar{t}\gamma}$ has 3 layers with 16, 8 and 4 nodes. The most important variables for the correct classification of the photon's origin are the invariant masses [6].

To analyse the performance of an NN, one uses confusion matrices, the one in Figure 6.1 shows the performance of $\mathrm{NN}_{t\bar{t}\gamma}$ for SM simulation. The rows indicate the true event category and the columns the classification of $\mathrm{NN}_{t\bar{t}\gamma}$. The probability in each field gives the percentage of true events, classified into a certain category, hence, the sum over each row is equals one (deviations come from rounding). The perfect confusion matrix, where all events are classified into their correct category would have only ones on the diagonal and zeros elsewhere. Some methods have been tried to further improve the performance of $\mathrm{NN}_{t\bar{t}\gamma}$. First, the variables have been rescaled such that they are all in the same range for the training and the NN does not get biased by the size of the variables. This confusion matrix is shown on the left in Figure 6.2. One can see, that top events are more often classified as top events as without the rescaling. Anyway, also ISR events are more often classified as top events. Changes of some percent are seen in almost all other categories. The confusion matrix on the right in Figure 6.2 is for training, where input variables are weighted according to their cross-section, e.g. the background got weighted higher since top events where overrepresented in the samples. The value for top and ISR classifications for ISR events has changed significantly. The other values show as well only changes of a few percent. For the correct classification of top events, both methods do improve the result. Even, if there is a higher correct classification on top events, this comes with

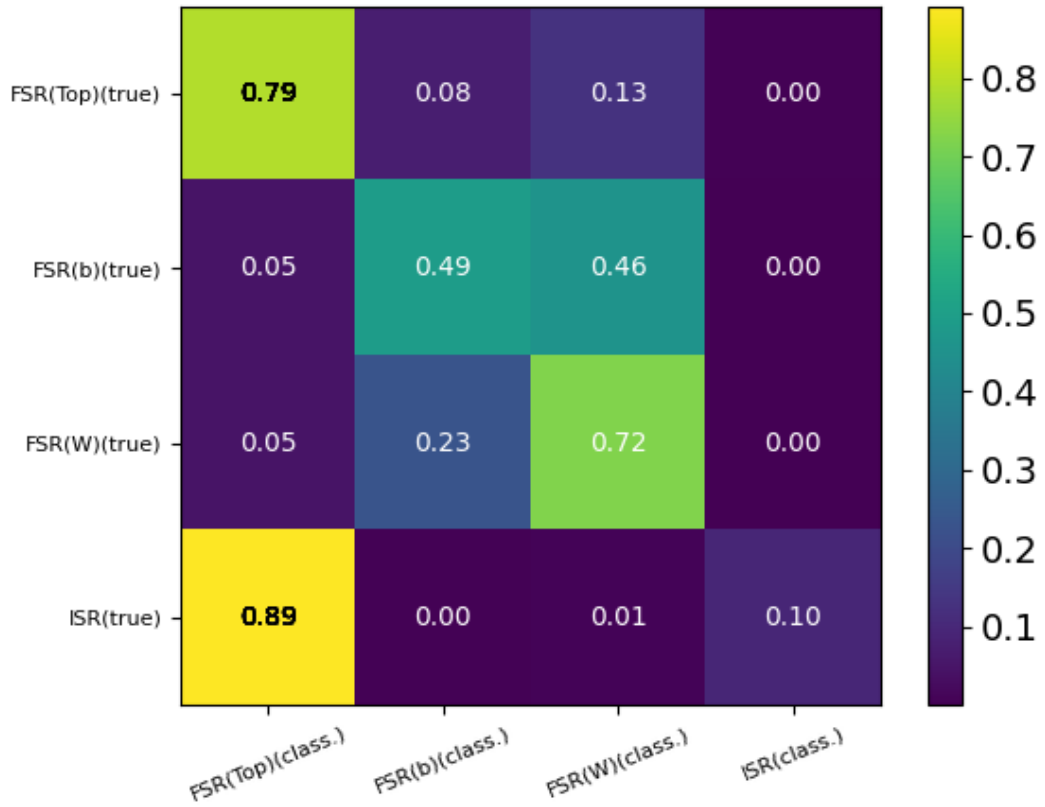

Figure 6.1.: Confusion matrix for training data of an NN with 3 layers (16/8/4 nodes) [6].

higher background and is therefore, no improvement of $\mathrm{NN}_{t\bar{t}\gamma}$.

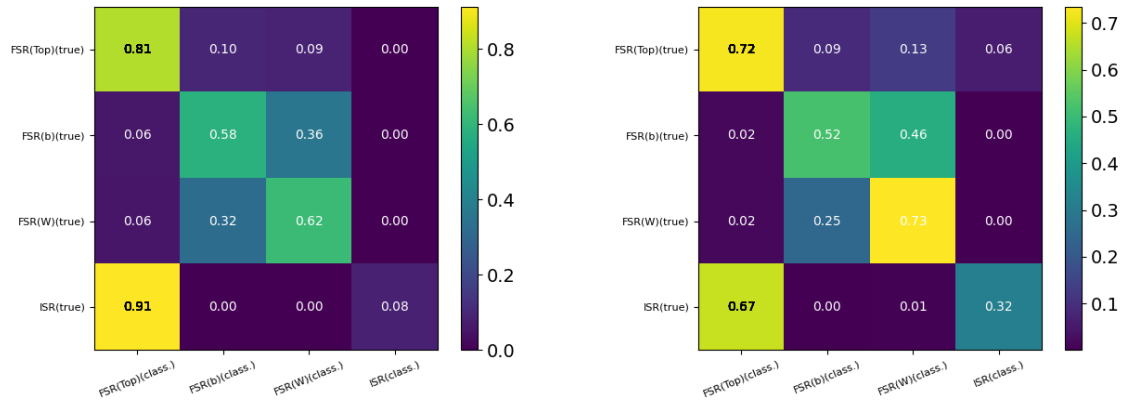

Figure 6.2.: Confusion matrix for training data of an NN with 3 layers (16/8/4 nodes) for rescaled variables (left) and reweighted (right) training. The non-modified confusion matrix is in Figure 6.1

6.1. How the neural network sorts the EFT samples

In Figure 6.3 the four classifiers of photon radiation for SM events are shown. The plots show the probability given by $\mathrm{NN}_{t\bar{t}\gamma}$ to belong to a certain category, namely FSR(top) (up left), ISR (up right), FSR(b) (bottom left) and FSR(W) (bottom right). In the plots one still sees some edges. For our studies, the top probability is the most important one. Since the four probabilities, that $\mathrm{NN}_{t\bar{t}\gamma}$ assigns to each event, need to add up to one, it may happen, that one of the less important categories (e.g. background) gets the rest probability and therefore those edges seen in the plots may occur (as in Figure 6.3 bottom right plot for W boson around 0.6).

To figure out wether $\mathrm{NN}_{t\bar{t}\gamma}$ is able to distinguish between the categories, even when applying it on EFT samples, the classifiers for the four categories are plotted in Figure 6.4. Each plot contains 8 histograms, four show the output for SM simulation, four the output for the EFT sample $\mathrm{ctw}_m\mathrm{ctz}_p\mathrm{ctg}_m$. This sample is chosen, since the separation power of this sample is the highest. For this reason one expects the highest deviation also within the application of $\mathrm{NN}_{t\bar{t}\gamma}$ (see: section 5.3). Taking the upper left plot as an example, the signal contribution are top events, the background are FSR(b), ISR and FSR(W), in the upper right plot, the signal contribution would be ISR events and the others the background. In all plots, one can see fluctuations between the samples for ISR, FSR(W) and FSR(b) events. Between the top curves (black and grey) differences are observed. In the top classifier, the black line is shifted towards the right (with respect to the grey one). The EFT top events are not as well classified as the SM events. For the other classifiers, the black histograms are shifted towards the left. The probabilities for the SM top events

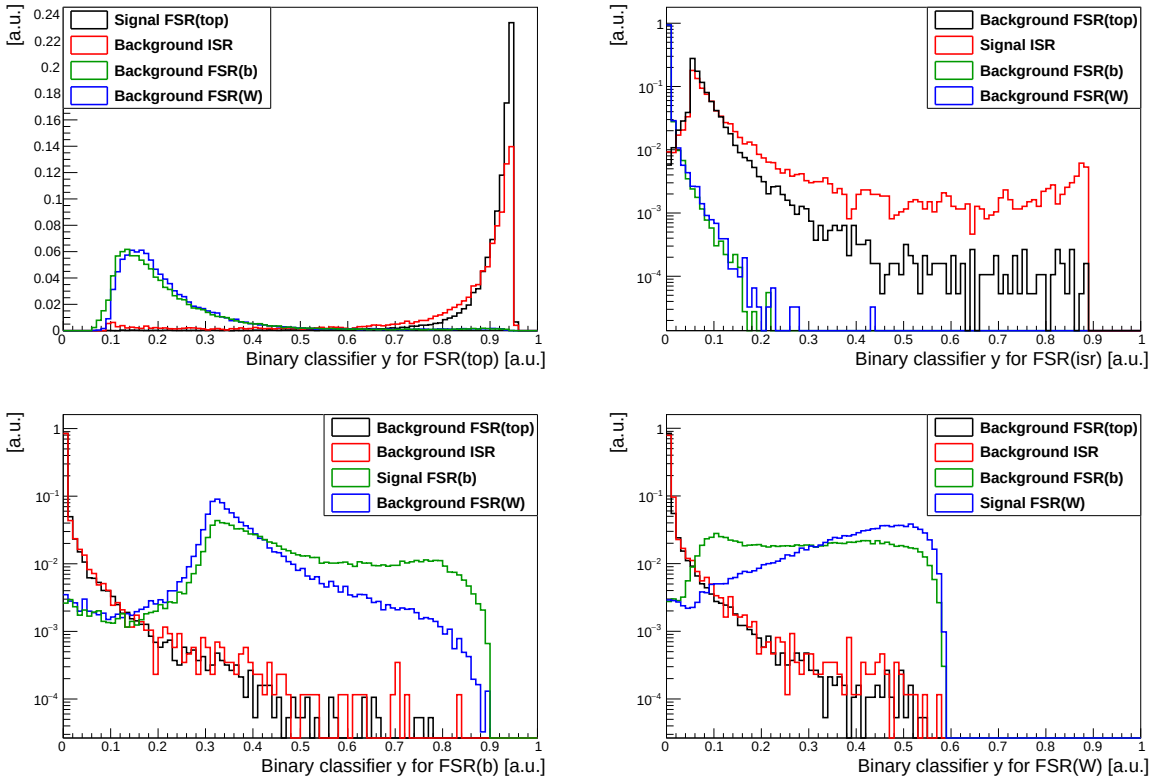

Figure 6.3.: Distributions for the classifiers for FSR(top) (up left), ISR (up right), FSR(b) (bottom left) and FSR(W) (bottom right). $\mathrm{NN}_{t\bar{t}\gamma}$ has 3 layers with 16, 8 and 4 nodes respectively.

to belong to the background are smaller than the ones for EFT top events.
$\mathrm{NN}_{t\bar{t}\gamma}$ does not sort (the most deviating) sample as well as the SM sample into the four different categories, but still may distinguish between FSR(top), ISR and FSR(b), FSR(W). This was expected, since $\mathrm{NN}_{t\bar{t}\gamma}$ was trained with SM events.

The next question to answer is, wether $\mathrm{NN}_{t\bar{t}\gamma}$ helps distinguishing EFT and SM events (for top radiation). This would be possible, if the classifier distributions for category top look differently for EFT and SM events. The ratio plots in Figure 6.5 and Figure 6.6 examine the effect of different samples on the perfomance of $\mathrm{NN}_{t\bar{t}\gamma}$. Figure 6.5 shows the performance of the top classifier on events belonging to different samples and different categories. Each plot shows one classifier, FSR(top) (up, left), ISR (up, right), FSR(b) (bottom, left) and FSR(W) (bottom, right). Each plot contains five histograms, one is the SM top classifier. The second is the SM classifier of this specific plot FSR(top), ISR, FSR(b) or FSR(W), respectively. The other three histograms are the classifiers for three different EFT samples, where one coupling is varied. Here, c_{tZ} is varied, this was chosen, since the two samples with the largest deviation from the SM are within this three (see:

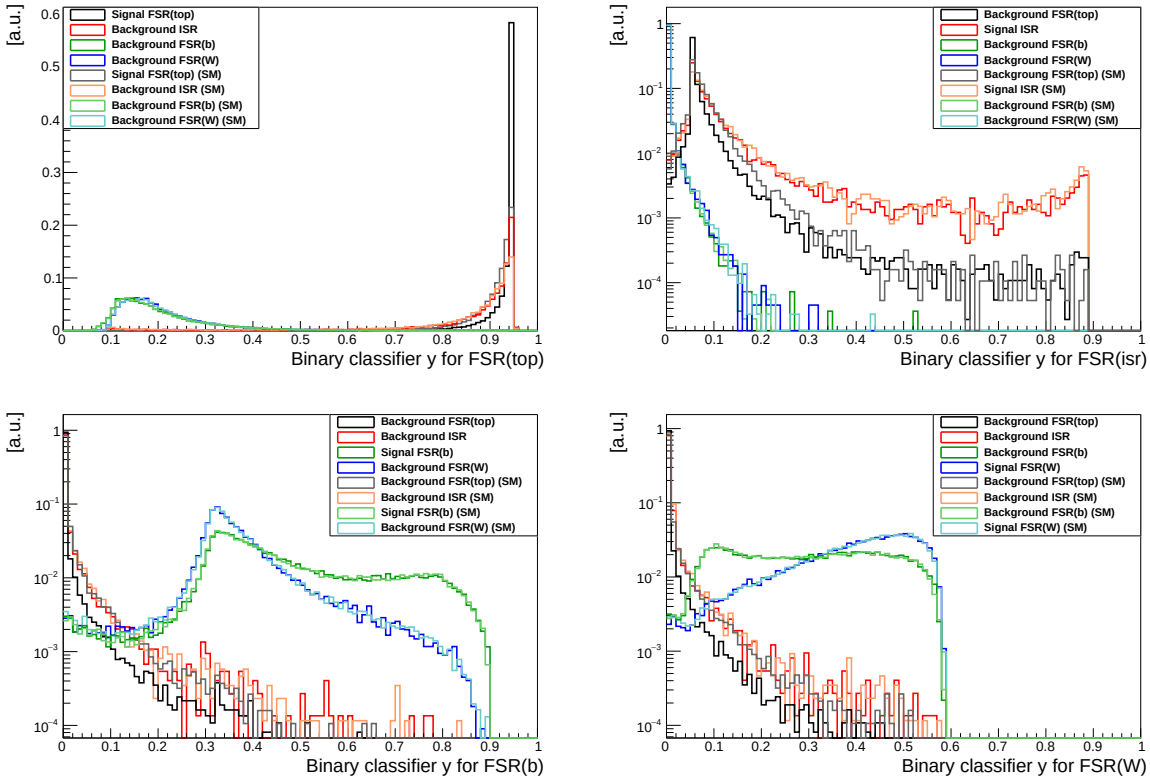

Figure 6.4.: The probability distributions $\mathrm{NN}_{t\bar{t}\gamma}$ assigned to the events to be FSR(top) (up, left), ISR (up, right), FSR(b) (bottom, left) and FSR(W) (bottom, right) applied on SM simulation and EFT sample $\mathrm{ctw}_m\mathrm{ctz}_p\mathrm{ctg}_m$.

section 5.3).

Figure 6.6 shows the effect of different EFT samples on different classifiers for top events. In each plot one can see the distribution of probability to be a photon coming from the top quark for SM simulation and the three samples described above. The upper left plot shows the FSR(W) classifier distribution. On the right is the FSR(b) classifier and on the bottom the ISR classifier. The distribution of the classification of each plot differs from the SM. For example, EFT top events are far less often classified with a high probability to be ISR than SM events.

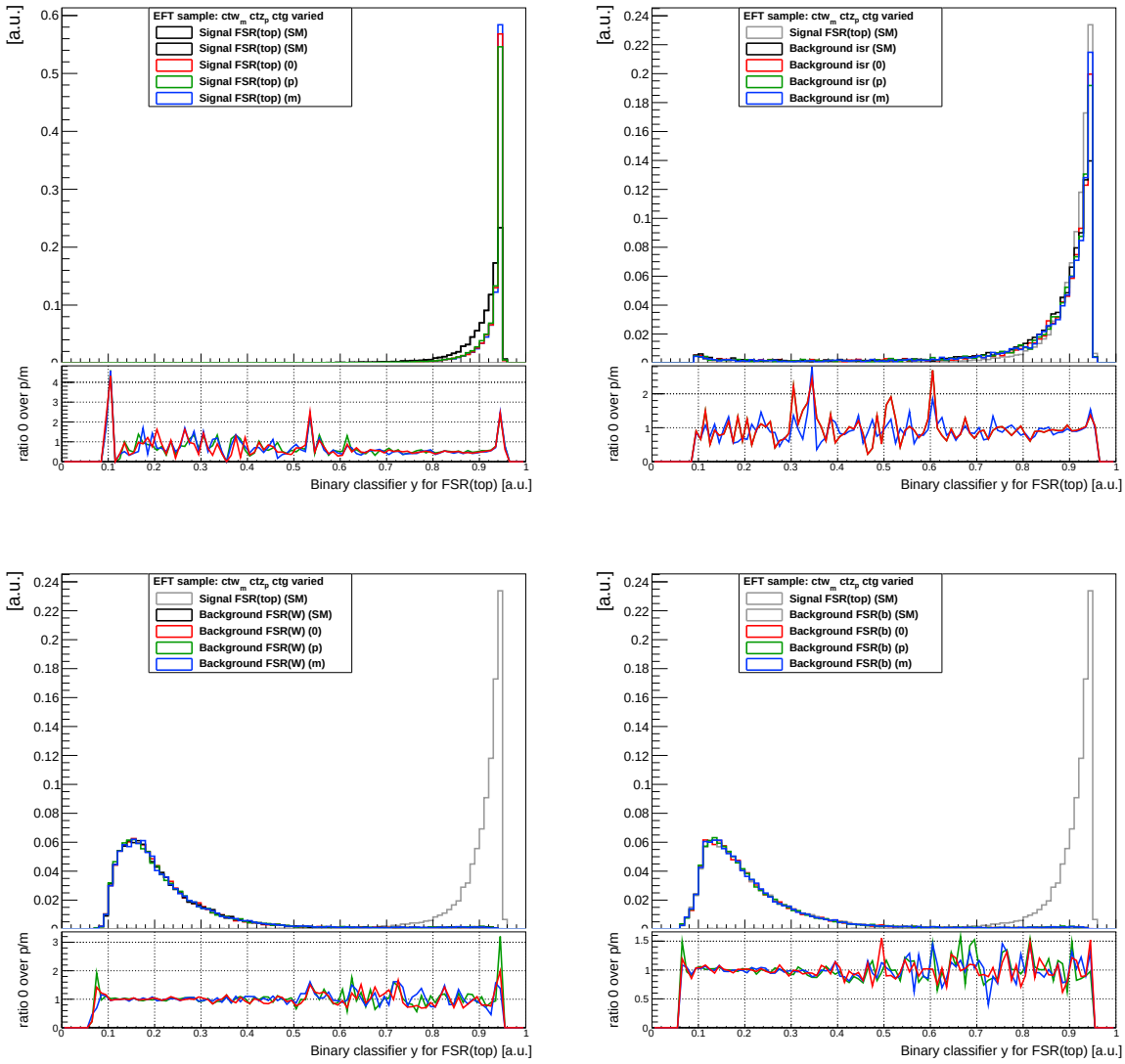

Figure 6.5.: The distributions for the classifiers for the four different categories. The NN has 3 layers with 16, 8 and 4 nodes respectively. ctw_m and ctz_p remain the same, ctg is varied. Each variation is shown with its ratio plot with respect to the SM. 0, p and m refer to the corresponding Wilson coefficients in Table 2.1.

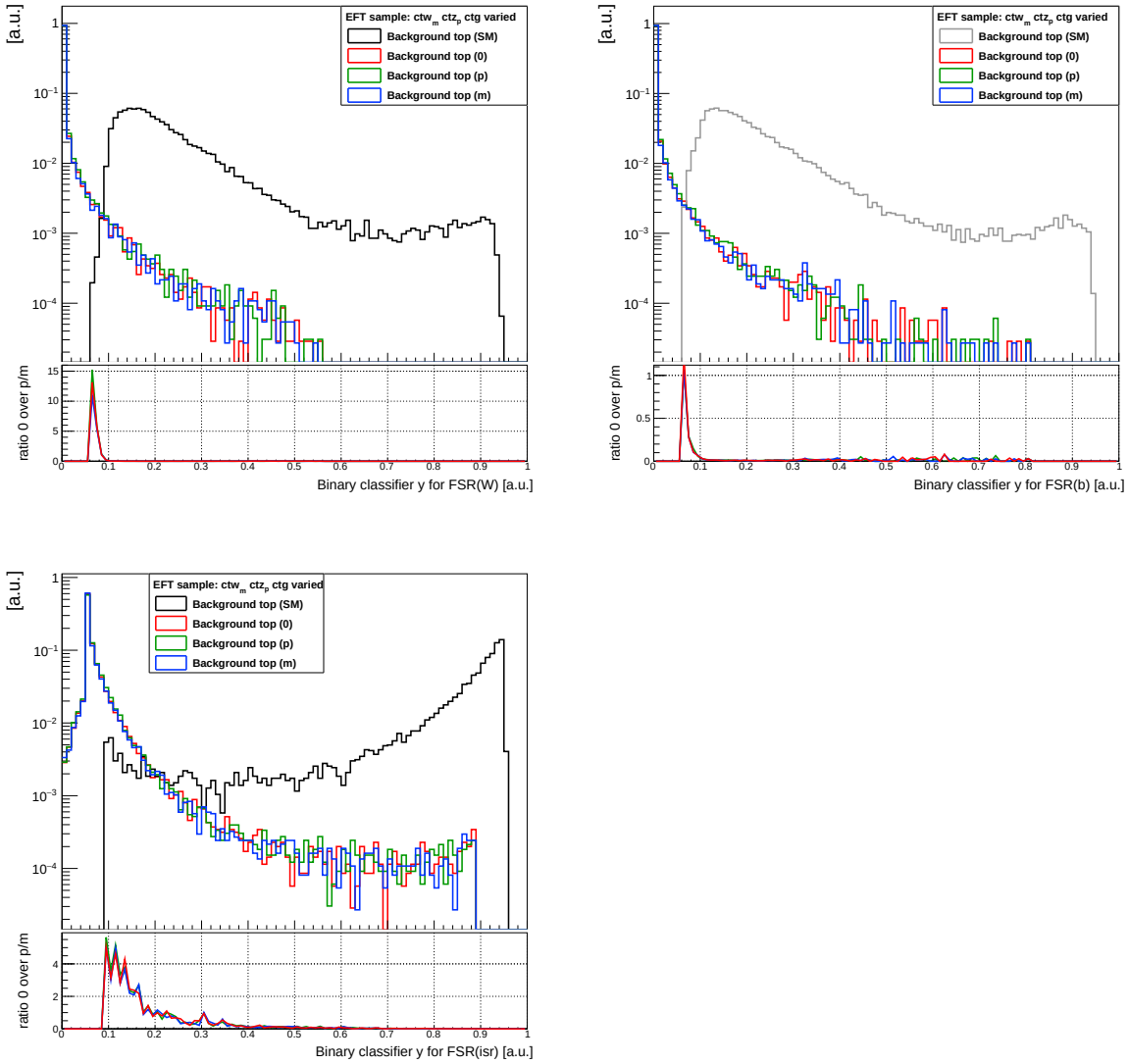

Figure 6.6.: The distributions for the four classifiers for top events. The NN has 3 layers with 16, 8 and 4 nodes respectively. ctw_m and ctz_p remain the same, ctg is varied. Each variation is shown with its ratio plot with respect to the SM. 0, p and m refer to the corresponding Wilson coefficients in Table 2.1.

7. Training of a neural network to distinguish EFT and SM events

7.1. Correlation matrices

Since it is not possible to distinguish SM and EFT samples with $\mathrm{NN}_{t\bar{t}\gamma}$ with the shape of the classifer, a new network was trained to explicitely differentiate SM and EFT events. This binary NN will be referred to as $\mathrm{NN}_{\mathrm{EFT}}$. In order to find out which variable is useful for the training, one can calculate the correlation between variables. If the correlation between two variables is high, the NN does not learn additional information when using both variables instead of one. To train a network, the best would be serveral variables with less correlation and high separation power. An NN might also be able to use higher dimensional correlations, not seen in the correlation matrices or correlations appearing in for example EFT variables but not in SM variables. The variables considered for training are listed in section 5.2 and eight more variables in the following list. Furthermore, the distributions for the variables as well as the ratio plots with respect to the SM can be found in Figure 5.1, Figure A.1 and Figure 6.6.

- $\eta_{p_T,\,\mathrm{highest}}$ is the $\eta-$distribution of the quark with the highest p_T
- $\eta_{p_T,\,2^{\mathrm{nd}}\,\mathrm{highest}}$ is the $\eta-$distribution of the quark with the second highest p_T
- $\eta_{\mathrm{smallest}\,|\eta|}$ of η is the $\eta-$distribution of the quark with the smallest absolute value
- $\eta_{2^{\mathrm{nd}}\,\mathrm{smallest}\,|\eta|}$ is the $\eta-$distribution of the quark with the second smallest absolute value of η
- event top is the output for the top quark classifier for $\mathrm{NN}_{t\bar{t}\gamma}$ discussed in chapter 6
- event ISR is the output for the ISR classifier for $\mathrm{NN}_{t\bar{t}\gamma}$ discussed in chapter 6

- event b is the output for the b quark classifier for $\mathrm{NN}_{t\bar{t}\gamma}$ discussed in chapter 6
- event W is the output for the W boson classifier for $\mathrm{NN}_{t\bar{t}\gamma}$ discussed in chapter 6

In Figure 7.1 and Figure 7.2 the correlation matrices for the EFT sample $\mathrm{ctw}_m\mathrm{ctz}_p\mathrm{ctg}_m$ and SM sample are shown. High correlations can be seen between $p_T(\gamma)$, $E(\gamma)$, $m_{\max}(\mathrm{Wb}\gamma)$

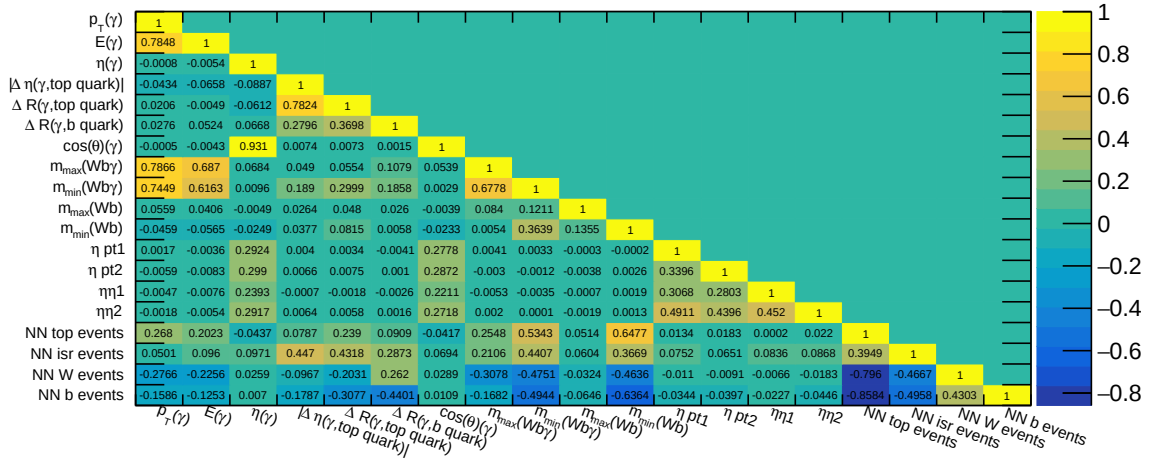

Figure 7.1.: Correlation matrix for the SM sample.

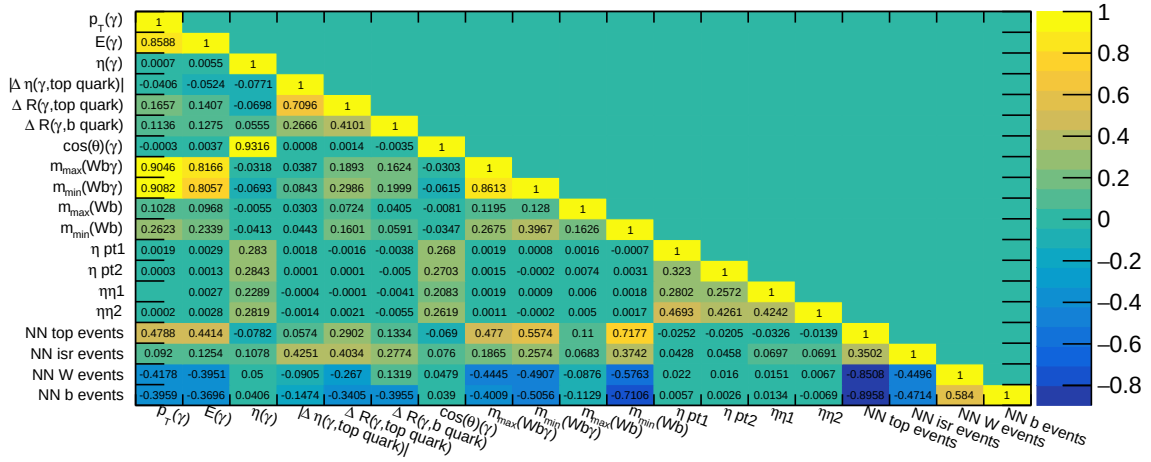

Figure 7.2.: Correlation matrix for the EFT sample $\mathrm{ctw}_m\mathrm{ctz}_p\mathrm{ctg}_m$.

and $m_{\min}(\mathrm{Wb}\gamma)$. This is due to the photon 4-vector entering the calculus of these variables. High anticorrelation values are found for the output of $\mathrm{NN}_{t\bar{t}\gamma}$ in top events with W and b events, respectively. This is also not astonishing since those events already have been sorted by $\mathrm{NN}_{t\bar{t}\gamma}$.

7.2. Training

For the training, the sample $\mathrm{ctw}_m\mathrm{ctz}_p\mathrm{ctg}_m$ was used, since its variable distributions differ the most from the SM (see: section 5.3). The following variables were used for training: $p_T(\gamma)$, $E(\gamma)$, $m_{\min}(\mathrm{Wb}\gamma)$, $m_{\max}(\mathrm{Wb})$, $m_{\min}(\mathrm{Wb})$, event top, event ISR, event W and event b. They are explained in section 5.2 and section 7.1. Those variables have been chosen because they all have separation power higher than 1%. The others were sorted out to reduce the number of variables. This is appreciated because less variables will need less time for training. Also, understanding the NN is easier with less variables, as well as applying performance checks. Serveral methods have been tried to improve the NN:

- different architectues: combinations of between one and nine layers and one and 150 nodes
- three different scalers, to transform the simulation set into a more suitable representation of the simulation for further processing. The first scaler subtracts the mean and scales to unit variance, the second scales each feature to the range between 0 and 1 and the third scales each feature by its maximum value.
- full and reduced (see above) variable content
- weighting the samples according to the number of training events
- use of batch normalization and dropout layers

There are two neural networks with a good performance. The signal corresponds to EFT samples, the background are SM events. The separation power of the one with one layer of 64 nodes (and the output layer) for training is 7.35% and for testing 7.31%, remember, the separation power of $p_T(\gamma)$ is 7.88%, discussed in section 5.3. Kolmogorov-Smirnov plot, ROC and loss function are shown in Figure 7.3. For this NN, the cross validation plot is shown (see: Figure 7.3, left plot), which indicates, that the NN did not find a local minimum but the global one. The second NN has another architecture. It has five layers with 64, 40, 40 and eight nodes (plus the one node output layer). Its separation power is for training 7.53% and for testing 7.22%.
Even though many options were tried or modified to improve the performance of $\mathrm{NN}_{\mathrm{EFT}}$, the separation power is less than the one of the variables (see: section 5.3).

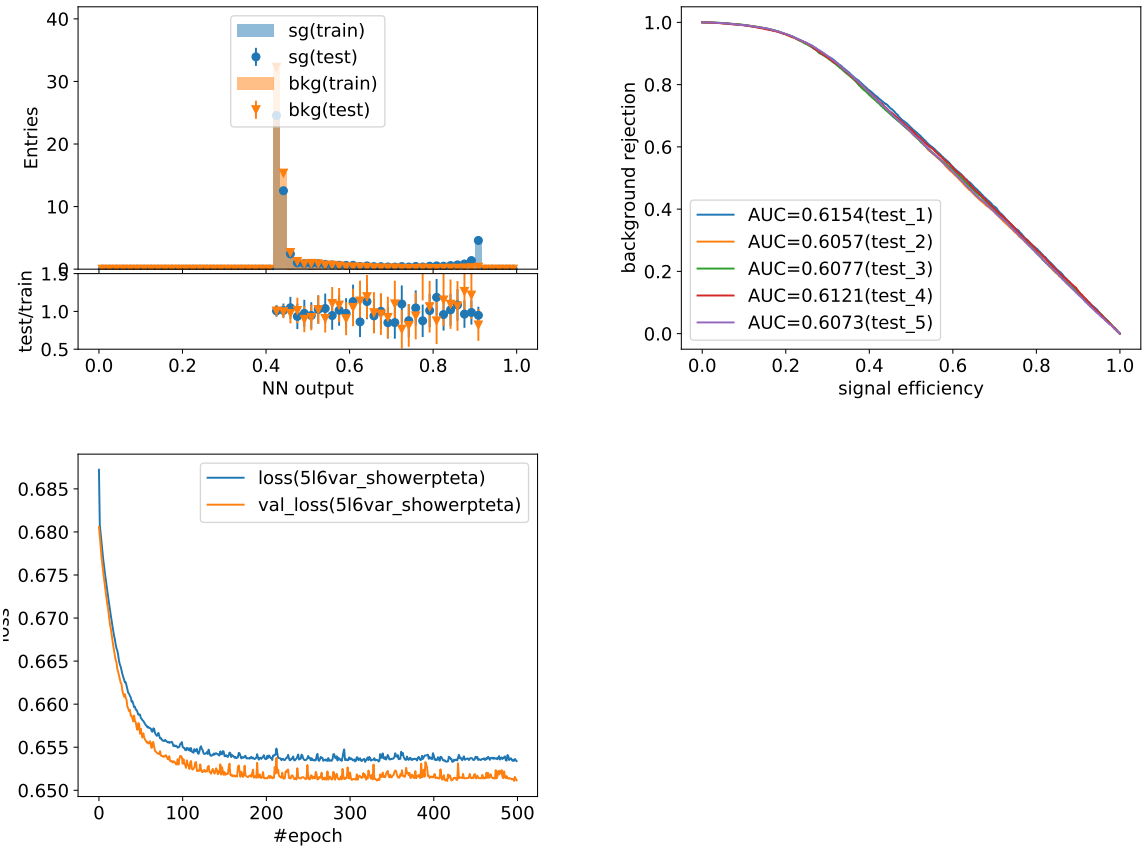

Figure 7.3.: NN output for a binary NN with one layer of 64 nodes.

7.3. Analysis of the performance of $\mathrm{NN_{EFT}}$

7.3.1. Analysis of events classified as EFT

The Kolmogorov-Smirnov plots in Figure 7.3 and Figure 7.4 show blue bumps around 0.9. This is the probability given by $\mathrm{NN_{EFT}}$ to belong to the category "EFT sample". For this analysis, a probability higher than 0.8 for an EFT event is called a "correctly" classified event. For an SM event, the probability to be "correctly" classified would be less then 0.2. "Wrongly" classified events are SM events with a probability higher than 0.8 or EFT events with a probability less than 0.2. To understand where the peaks in the figures come from, one may have a look at Figure 7.5. Plots of the variable distributions for the SM (pink) and three plots corresponding to the EFT sample are shown, whereas the blue line shows all events, the green one all events with a probability given by $\mathrm{NN_{EFT}}$[1] higher than 0.8 (so the blue-bump events) and the black one shows events lower than 0.8[2]. one can

[1] $\mathrm{NN_{EFT}}$'s architecture had 64/40/40/8/1 nodes
[2] The distributions for the other variables are shown in Appendix C.

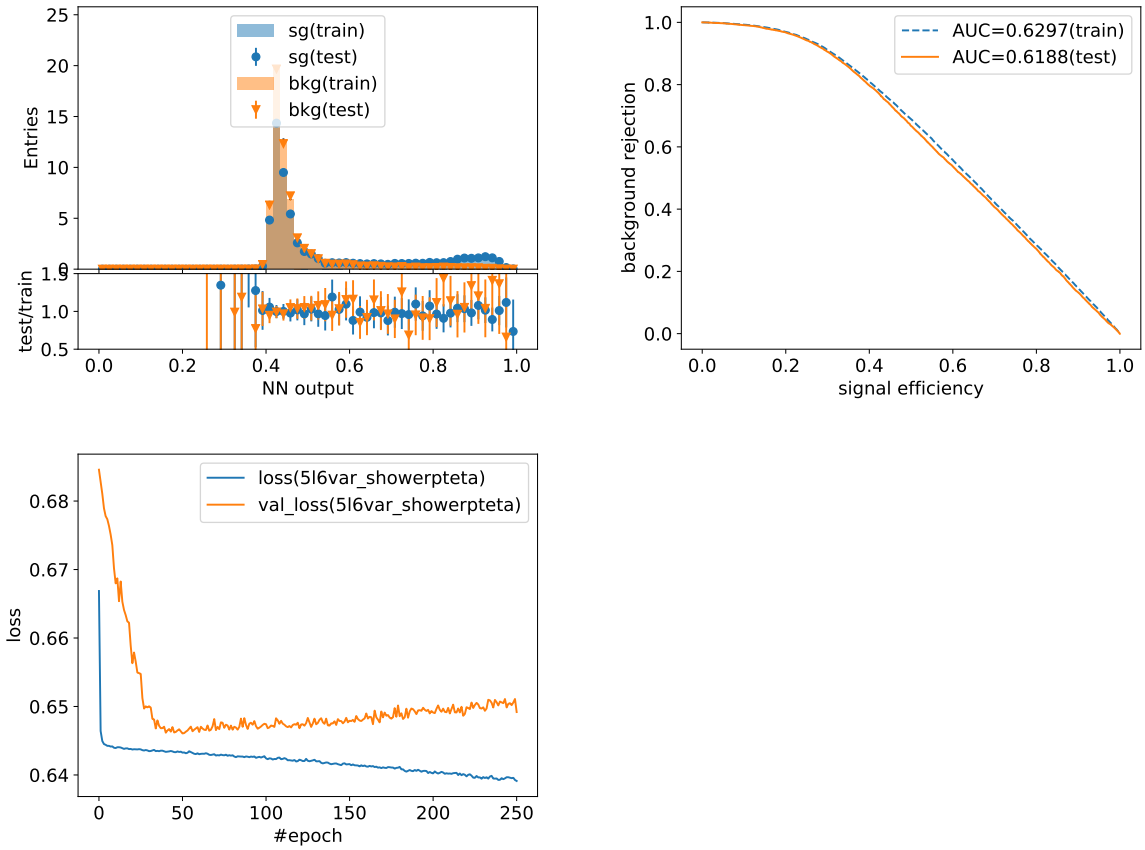

Figure 7.4.: NN output for a binary NN with five layers of 64, 40, 60, 8 and one node respectively.

see, that as expected from the variable distributions in section 5.2, the NN understands high $p_T(\gamma)$, $E(\gamma)$, $m_{\max}(\mathrm{Wb}\gamma)$ values as EFT events. In the $p_T(\gamma)$ distribution, there is a small overlap of the green and the pink line but above $\approx 160\,\mathrm{GeV}$, all EFT events are declared as EFT events. Therefore, there are wrongly declared SM events in this region as well.

7.4. How the neural network works on other samples

Even though $\mathrm{NN_{EFT}}$ is trained on EFT sample $\mathrm{ctw}_m\mathrm{ctz}_p\mathrm{ctg}_m$, it is able to separate also some EFT events of other EFT samples from the SM. In Figure 7.6, ratio plots are drawn to analyse the performance of $\mathrm{NN_{EFT}}$ on different EFT samples. In each ratio plot, two couplings remain the same and one is varied in the three variations explained in Table 2.1. The ratio plots are constructed with respect to the SM sample, drawn are only the true EFT events. The left ratio plots show the output of $\mathrm{NN_{EFT}}$ when two couplings are set to

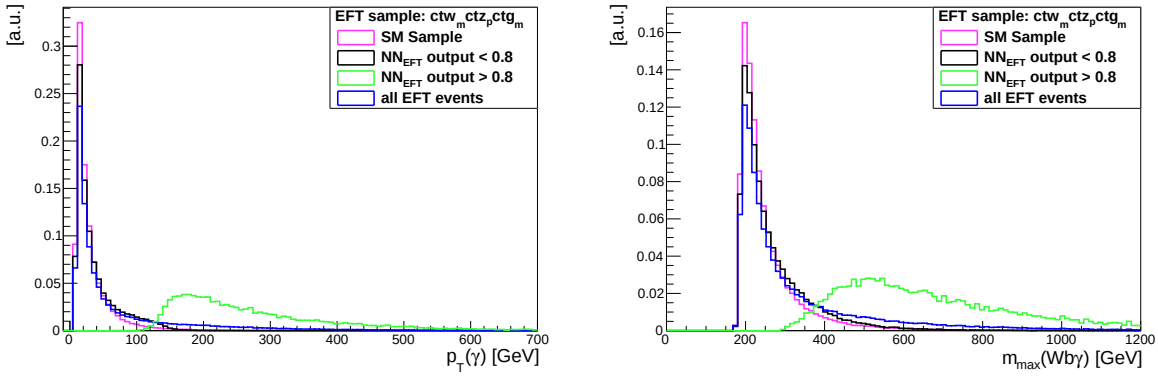

Figure 7.5.: Variable distributions of events with a high given probability to be EFT. The distributions for transverse momentum and the maximum invariant mass (Wbγ) are plotted.

zero, so the closest samples to the SM. The right side shows samples with high separation power. In the first row, the first coupling c_{tW} is varied, in the second one c_{tZ} and in the third, c_{tG}.

Note, that samples with high separation power (see: section 5.3 and section 5.4) are separated well by NN_{EFT}. Also the ratio plots show, that c_{tG} does not have a huge impact on the variables since the output of NN_{EFT} is either almost identical to SM or for all three variations equal. For the other two couplings, it looks like the ctw_m and ctz_p variations have the highest impact on the sample.

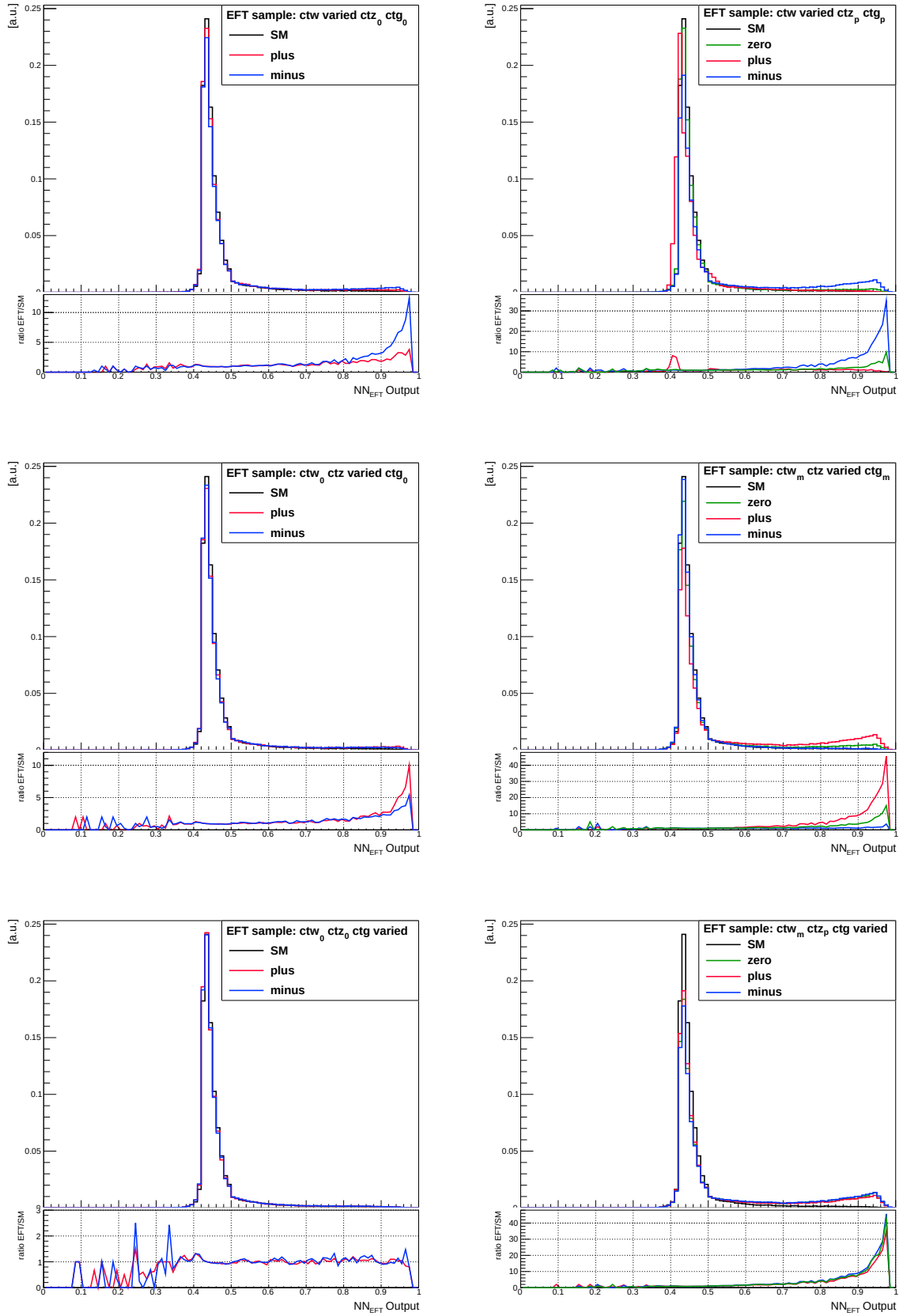

Figure 7.6.: Ratio Plots of the NN_{EFT} output on different samples with respect to the SM sample. On top of the legend the samples are declared.

8. Conclusion and outlook

$NN_{t\bar{t}\gamma}$ separates the origin of the photon within EFT samples almost as well as within SM samples. EFT events may not be separated from SM events by the shape of $NN_{t\bar{t}\gamma}$'s output distributions. Therefore, a second network was trained
The second network is not better than the separation power of the single variable $p_T(\gamma)$. Nevertheless, it is interesting to notice, that one network may also separate different EFT samples.

To further improve the results of this thesis, a deeper understanding of the EFT samples is nessecary. Therefore, I suggest to look more detailed into the EFT samples. Analyse, why the ΔR distributions do not show broader distributions in case of EFT samples even if one might expect this due to the relation between higher energies and larger opening angles. Also one should analyse, which operator has which effects on the variable distributions and find out, what kind of interferences are in between the operators. This may be done via the generation of more EFT parameter points, for example with larger variations. c_{tG} has little influence on the output (as analysed in section 5.3). That makes sense considering, that the top-gluon coupling is only important for the production of $t\bar{t}$ pairs, not for their decay. c_{tW} and c_{tZ} are both electroweak couplings and therefore have an electroweak component, which results in the modification of the photon behaviour.
To use $NN_{t\bar{t}\gamma}$ on real data, it is nessecary to adapt the network to reconstruction level, that means to reconstruct the W boson and b quark variables from their decay products and jets. As discussed in chapter 6, the invariant masses are the crucial factor for $NN_{t\bar{t}\gamma}$. Since the invariant masses do have different distributions for EFT and SM samples, training of a new NN on reconstruction level on EFT samples may be considered. Then, even EFT photons may be classified correcly. It would be interesting to see if the performance of an NN, trained on EFT samples may work for SM simulation.
For both networks, but especially for NN_{EFT}, feature importance is a nice technique to figure out the functioning of an NN. Here, one shuffles the entries of one variable such that the values do not belong anymore to their event. In the later training, one can see, depending on the output, whether a variable has been important for the training. Feature

importance can help as well, to figure out, if it is even possible to further improve the performance of an NN. If then, an output has a better result, one should think about systematic uncertainities of the NN. Here, this is not important since the NN output is not satisfying[1].

The aim was, that serveral variables with separation power in the range of 5% would have a higher separation power, when combined in an NN. Maybe, $\mathrm{NN_{EFT}}$ can be improved focussing on the use of little correlation variables. Here, variables with high separation power have been used, unfortunately, the correlation was checked afterwards and it turned out to be high for the used variables. Thus, using different variables could improve the network, even when having little separation power. In either way, I recommend to check the correlation before deciding, which variables shall be used for training. In this thesis, feedforward networks were used. Different, more advanced types of NN could maybe give a better result as for example recurrent neural networks, if they find higher order correlations between variables.

To conclude: The found separation power of the trained $\mathrm{NN_{EFT}}$ is not better than the one of the single variable $p_T(\gamma)$, which has the highest separation power. However, if $\mathrm{NN_{EFT}}$ shall be applied on real data, it has to work on reconstruction level. Then the separation power on reconstruction level can be found in fitting both, $\mathrm{NN_{EFT}}$ output and the $p_T(\gamma)$ distribution on real data. Based on these fits, limits for the Wilson coefficients can be found. Here, then is the question wether the limits for the single variable distribution or $\mathrm{NN_{EFT}}$ are better e.g. with smaller allowed regions for the EFT parameters.

[1] means here: better than the separation power of a single variable

A. Variable distributions - Comparison of EFT and SM sample

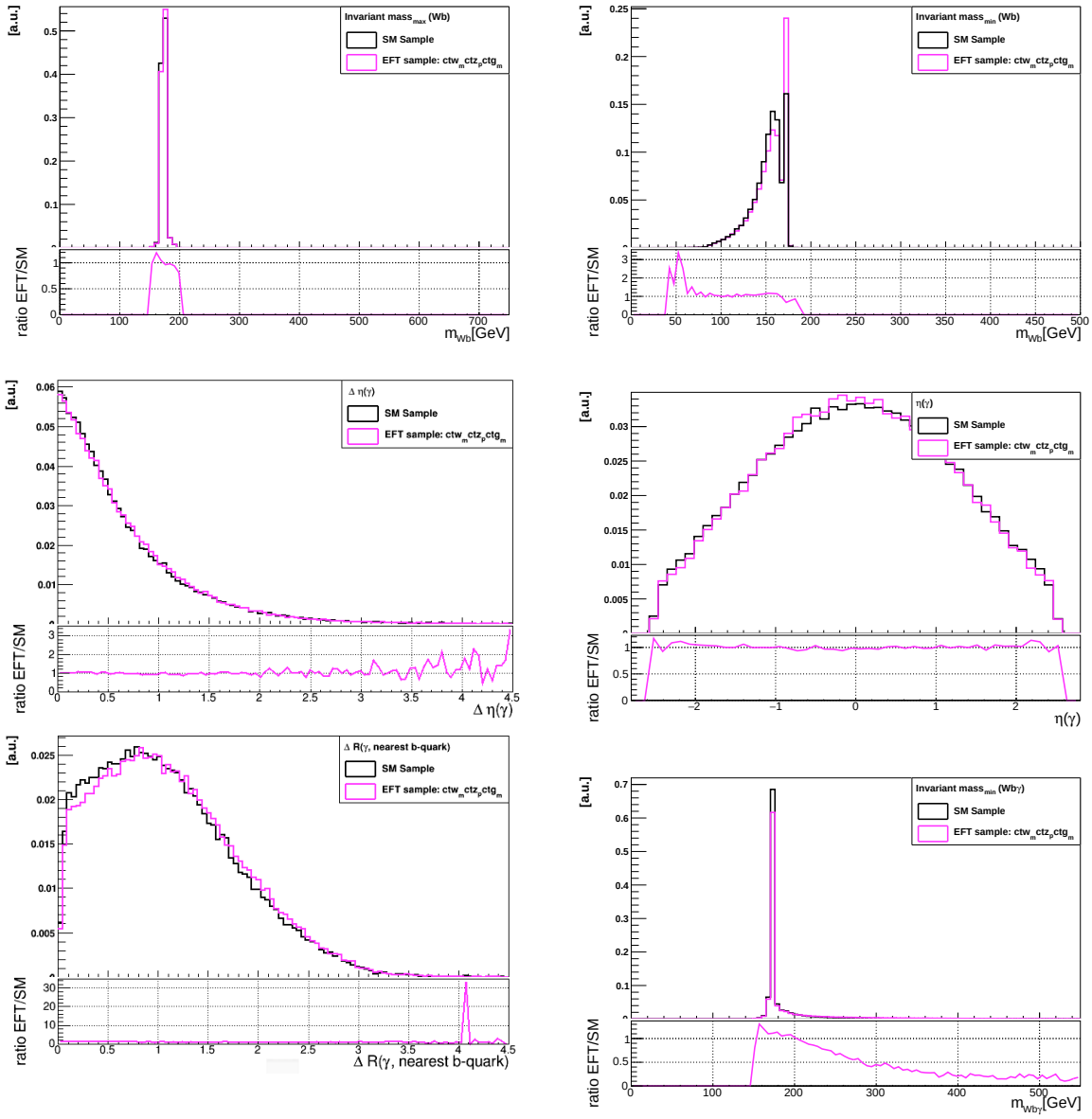

Figure A.1.: Ratio plots of the variable distributions for the variables listed in section 5.2. For each variable the SM and the $\mathrm{ctw}_m\mathrm{ctz}_p\mathrm{ctg}_m$ sample histogram is plotted.

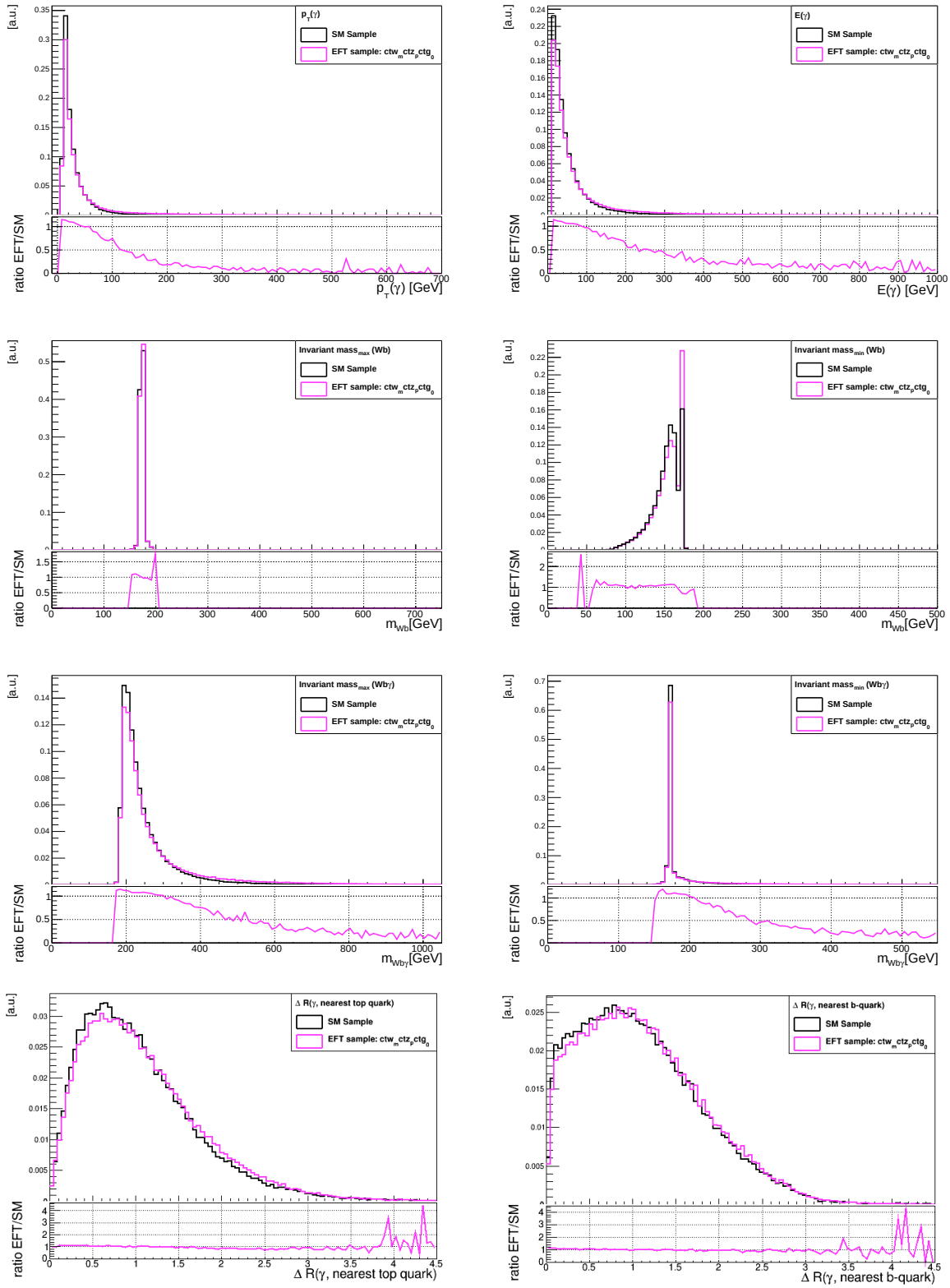

Figure A.2.: Ratio plots for the variable distributions for the variables listed in section 5.2. For each variable the SM and the $\mathrm{ctw}_m\mathrm{ctz}_p\mathrm{ctg}_0$ sample histogram is plotted, it is the one with the second highest separation power.

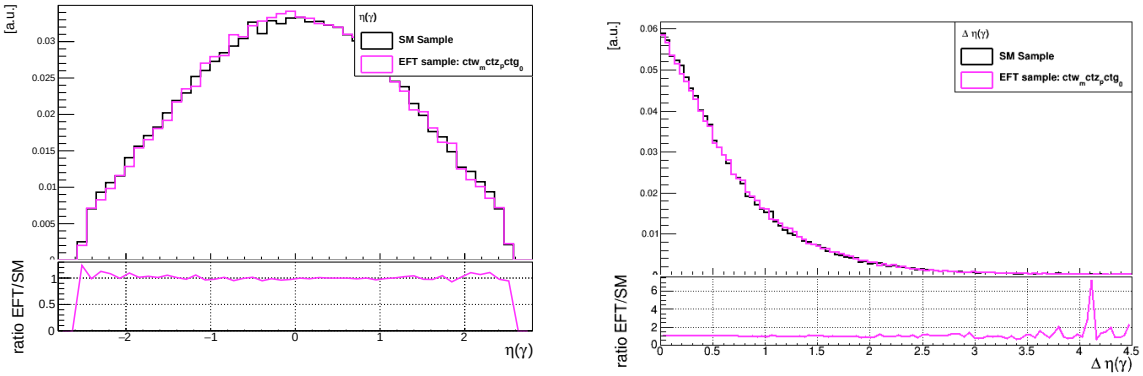

Figure A.3.: Ratio plots for the variable distributions for the variables listed in section 5.2. For each variable the SM and the $\mathrm{ctw}_m\mathrm{ctz}_p\mathrm{ctg}_0$ sample histogram is plotted, it is the one with the second highest separation power.

B. Separation power

variable\sample	00p[%]	00m[%]	0m0[%]	0mp[%]	0mm[%]	0p0[%]	0pp[%]	0pm[%]
$p_T(\gamma)$	0.01	0.03	0.44	0.36	0.39	0.65	0.55	0.89
$E(\gamma)$	0.01	0.03	0.28	0.23	0.24	0.35	0.29	0.53
$\eta(\gamma)$	0.01	0.01	0.02	0.02	0.01	0.02	0.02	0.02
$\Delta\eta(\gamma$,topquark)	0.01	0.04	0.02	0.03	0.04	0.04	0.03	0.04
$m_{\max}$(Wb)	0.01	0.01	0.01	0.02	0.02	0.01	0.01	0.01
$m_{\min}$(Wb)	0.01	0.01	0.11	0.07	0.12	0.08	0.06	0.11
$m_{\max}$(Wbγ)	0.01	0.03	0.25	0.18	0.2	0.31	0.25	0.44
$m_{\min}$(Wbγ)	0.01	0.03	0.35	0.29	0.3	0.41	0.33	0.65
ΔR(b quark)	0.02	0.03	0.03	0.03	0.03	0.03	0.04	0.03
ΔR(top quark)	0.02	0.04	0.04	0.04	0.05	0.05	0.03	0.04
$\cos(\theta)$	0.01	0.03	0.03	0.03	0.03	0.04	0.03	0.02
$\eta_{p_T,\,\text{highest}}$	0.02	0.03	0.05	0.04	0.04	0.04	0.40	0.04
$\eta_{p_T,\,2^{\text{nd}}\,\text{highest}}$	0.03	0.03	0.03	0.02	0.03	0.03	0.03	0.02
$\eta_{\text{smallest}\,\lvert\eta\rvert}$	0.01	0.03	0.03	0.03	0.03	0.05	0.03	0.03
$\eta_{2^{\text{nd}}\,\text{smallest}\,\lvert\eta\rvert}$	0.01	0.02	0.03	0.02	0.03	0.04	0.03	0.03
event(top)	0.01	0.01	0.25	0.19	0.25	0.28	0.21	0.43
event(isr)	<0.00	0.02	0.19	0.14	0.2	0.20	0.16	0.32
event(W)	<0.00	0.01	0.13	0.09	0.16	0.10	0.07	0.15
event(b)	0.01	0.01	0.14	0.09	0.15	0.10	0.08	0.15

Table B.1.: Separation power of different EFT samples. The meaning of the variables is explained in section 5.2. 00p refers to EFT sample $\text{ctw}_0\text{ctz}_0\text{ctg}_p$.

variable\sample	m00[%]	m0p[%]	m0m[%]	mm0[%]	mmp[%]	mmm[%]	mpp[%]
$p_T(\gamma)$	1.22	0.92	1.63	0.03	0.03	0.06	5.64
$E(\gamma)$	0.74	0.51	1.02	0.04	0.03	0.04	4.31
$\eta(\gamma)$	0.02	0.02	0.03	0.02	0.02	0.02	0.03
$\Delta\eta(\gamma$,topquark)	0.03	0.03	0.04	0.03	0.04	0.04	0.05
$m_{\max}$(Wb)	0.02	0.01	0.03	0.01	<0.00	0.01	0.12
$m_{\min}$(Wb)	0.20	0.12	0.33	0.02	0.02	0.02	1.74
$m_{\max}$(Wbγ)	0.61	0.45	0.93	0.03	0.02	0.04	3.87
$m_{\min}$(Wbγ)	0.92	0.64	1.29	0.05	0.02	0.05	5.19
ΔR(b quark)	0.05	0.04	0.06	0.04	0.03	0.03	0.14
ΔR(top quark)	0.07	0.04	0.10	0.04	0.03	0.04	0.38
$\cos(\theta)$	0.03	0.03	0.03	0.03	0.02	0.03	0.04
$\eta_{p_T,\,\text{highest}}$	0.03	0.03	0.05	0.04	0.03	0.04	0.05
$\eta_{p_T,\,2^{\text{nd}}\,\text{highest}}$	0.03	0.04	0.03	0.03	0.03	0.04	0.04
$\eta_{\text{smallest}\,\lvert\eta\rvert}$	0.03	0.03	0.04	0.03	0.03	0.04	0.04
$\eta_{2^{\text{nd}}\,\text{smallest}\,\lvert\eta\rvert}$	0.03	0.04	0.03	0.04	0.03	0.03	0.04
event(top)	0.68	0.45	0.97	0.02	0.01	0.03	4.37
event(isr)	0.50	0.33	0.74	0.02	0.02	0.03	3.57
event(W)	0.25	0.14	0.43	0.01	0.01	0.02	2.27
event(b)	0.25	0.15	0.42	0.01	0.01	0.02	2.27

Table B.2.: Separation power of different EFT samples. The meaning of the variables is explained in section 5.2. m00 refers to EFT sample $\text{ctw}_m\text{ctz}_0\text{ctg}_0$.

variable\sample	p00[%]	p0p[%]	p0m[%]	pm0[%]	pmp[%]	pmm[%]	pp0[%]	ppm[%]
$p_T(\gamma)$	0.34	0.37	0.37	1.94	1.82	2.17	0.05	0.02
$E(\gamma)$	0.22	0.24	0.27	1.38	1.24	1.56	0.03	0.05
$\eta(\gamma)$	0.02	0.01	0.02	0.01	0.01	0.01	0.02	0.02
$\Delta\eta(\gamma$,topquark)	0.03	0.03	0.03	0.04	0.05	0.04	0.03	0.03
$m_{\max}$(Wb)	0.01	0.01	0.01	0.04	0.05	0.04	0.01	0.01
$m_{\min}$(Wb)	0.09	0.06	0.09	0.56	0.57	0.76	0.03	0.02
$m_{\max}$(Wbγ)	0.18	0.20	0.22	1.22	1.09	1.39	0.02	0.03
$m_{\min}$(Wbγ)	0.28	0.27	0.30	1.81	1.68	2.06	0.03	0.02
ΔR(b quark)	0.03	0.04	0.03	0.06	0.06	0.06	0.03	0.03
ΔR(top quark)	0.04	0.04	0.04	0.12	0.14	0.13	0.04	0.04
$\cos(\theta)$	0.02	0.02	0.03	0.02	0.03	0.03	0.02	0.03
$\eta_{p_T,\,\text{highest}}$	0.04	0.03	0.04	0.03	0.04	0.04	0.04	<0.00
$\eta_{p_T,\,2^{\text{nd}}\,\text{highest}}$	0.04	0.02	0.03	0.04	0.04	0.03	0.03	<0.00
$\eta_{\text{smallest}\,\lvert\eta\rvert}$	0.04	0.04	0.04	0.04	0.03	0.04	0.03	<0.00
$\eta_{2^{\text{nd}}\,\text{smallest}\,\lvert\eta\rvert}$	0.03	0.03	0.02	0.03	0.04	0.03	0.03	<0.00
event(top)	0.20	0.20	0.20	1.33	1.28	1.59	0.02	0.01
event(isr)	0.16	0.16	0.16	1.04	1.02	1.29	0.02	0.01
event(W)	0.11	0.08	0.12	0.74	0.73	0.95	0.01	0.01
event(b)	0.11	0.08	0.12	0.73	0.72	0.94	0.01	0.02

Table B.3.: Separation power of different EFT samples. The meaning of the variables is explained in section 5.2. p00 refers to EFT sample $\text{ctw}_p\text{ctz}_0\text{ctg}_0$.

C. Analysis of events classified as EFT

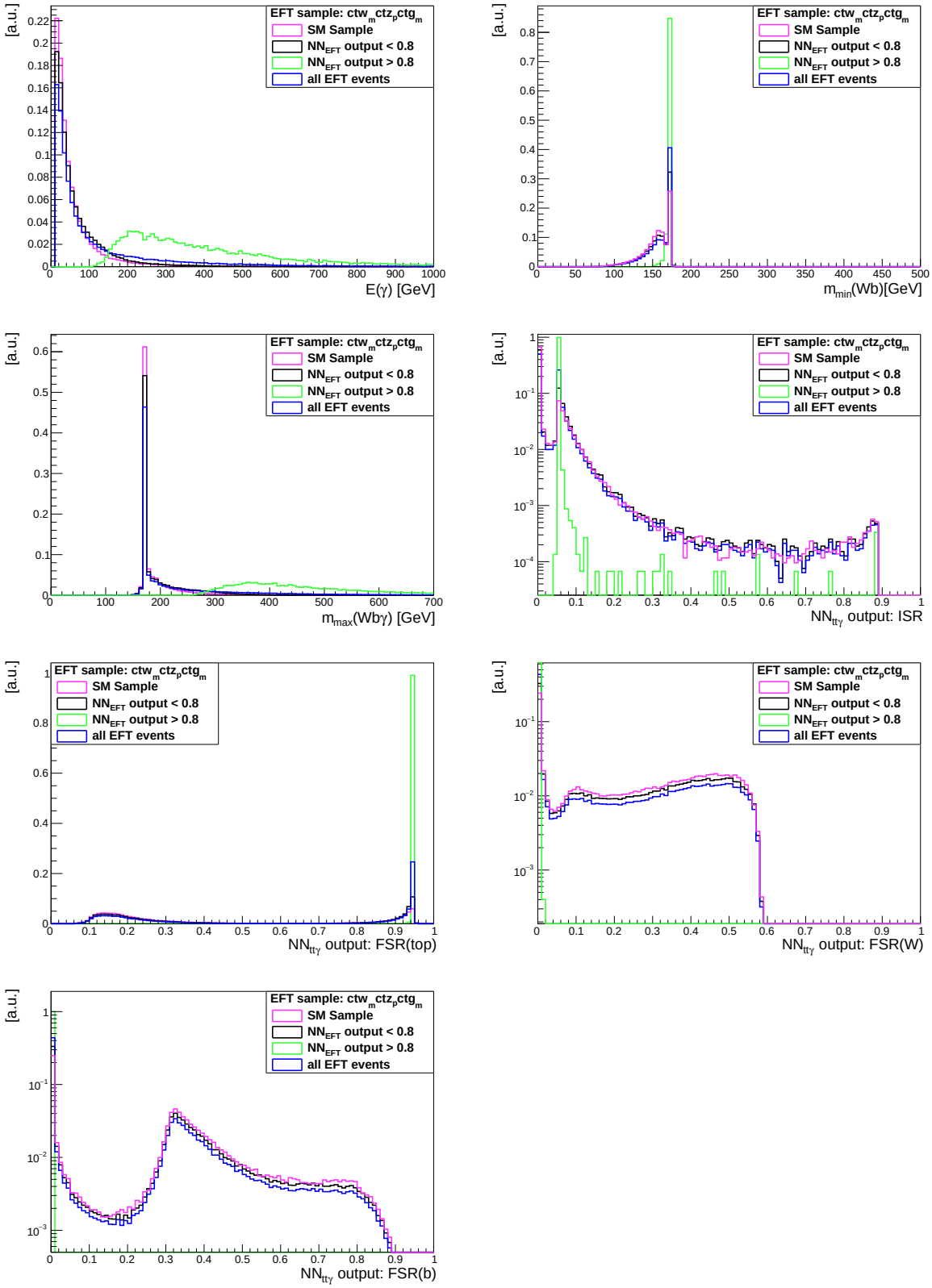

Figure C.1.: Distributions for the classifiers for SM or EFT sample $\mathrm{ctw}_m\mathrm{ctz}_p\mathrm{ctg}_m$. The green line shows the distribution of the events classified as EFT with a probability higher than 0.8. The black line shows the distribution for a probability below 0.8 and the other two the distribution for SM and EFT respectively. Here, one tries to understand which events $\mathrm{NN}_{\mathrm{EFT}}$ declares as EFT events.

Bibliography

[1] A. D. Sakharov, *Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe*, Pisma Zh. Eksp. Teor. Fiz. **5**, 32 (1967), [JETP Lett.5,24(1967); Sov. Phys. Usp.34,no.5,392(1991); Usp. Fiz. Nauk161,no.5,61(1991)]

[2] Y. Fukuda, et al., *Evidence for Oscillation of Atmospheric Neutrinos*, Phys. Rev. Lett. **81**, 1562 (1998)

[3] L. Bergström, *Non-baryonic dark matter: observational evidence and detection methods*, Rep. Prog. Phys. **63**, 793 (2000)

[4] S. Abachi, et al. (DØ), *Observation of the Top Quark*, Phys. Rev. Lett. **74**, 2632 (1995)

[5] F. Abe, et al. (CDF), *Observation of Top Quark Production in $p\bar{p}$ Collisions with the Collider Detector at Fermilab*, Phys. Rev. Lett. **74**, 2626 (1995)

[6] A. Kirchhoff, *Studies to classify prompt photons in the $t\bar{t}\gamma$-process*, II.Physik-UniGö-MSc-2018/04

[7] P. Herrmann, *Studies of the $t\bar{t}\gamma$ production process with the Atlas experiment at the Lhc*, II.Physik-UniGö-BSc-2018/06

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

[9] M. Tanabashi, et al. (Particle Data Group), *Review of Particle Physics*, Phys. Rev. D **98**, 030001 (2018)

[10] G. Aad, et al. (Atlas), *Performance of b-jet identification in the Atlas experiment*, JINST **11**, P04008 (2016)

[11] O. Bylund, et al., *Probing top quark neutral couplings in the Standard Model Effective Field Theory at NLO in QCD*, JHEP **05**, 052 (2016)

[12] Sreelakshmi J S, *Interpretation of the Top-Antitop-Photon Production Rate in Effective Field Theories* (2020), Indian Institute of Science Education and Research, Pune, India and Georg August Universität Göttingen, Germany, to be published

[13] D. J. Marsh, *Axion cosmology*, Phys. Rep. **643**, 1 (2016)

[14] *High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1* **4/2017** (2017)

[15] G. Aad, et al. (ATLAS), *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3**, S08003 (2008)

[16] *Luminosity determination in pp collisions at* $\sqrt{s} = 13$ *TeV using the ATLAS detector at the LHC*, Technical Report ATLAS-CONF-2019-021, CERN, Geneva (2019)

[17] M. zur Nedden, *The LHC Run 2 ATLAS trigger system: design, performance and plans*, JINST **12**, C03024 (2017)

[18] W. McCulloch, W. Pitts, *A logical calculus of the ideas immanent in nervous activity*, Bull. Math. Biophys. **5**, 115 (1943)

Danksagung

I would like to thank Prof. Arnulf Quadt for welcoming me warmly in his working group. Furthermore, I would like to thank Thomas and Andreas, always answering my questions in short time and taking the time to answer in detail. It is a pity, that we never had the opportunity to drink a coffee together, but they did a very good job, creating a cosy athmosphere on Mattermost.
I want to thank my dear and beloved friend Elias, who is always answering my calls.
And there are all people engaging themselfes in fights for climate justice. Who are not waiting until their bachelor is finished, since likewise, climate change is not waiting.
Thank you.

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 7. August 2020

(Gesine Grubert)