GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Fakultät für
Physik

[q,p]=iℏ

Master's Thesis

Studien zur Klassifikation prompter Photonen im $t\bar{t}\gamma$-Prozess

Studies to classify prompt photons in the $t\bar{t}\gamma$-process

prepared by

Andreas Kirchhoff

from Wipperfürth

at the II. Physikalischen Institut

Zusammenfassung

Der $t\bar{t}\gamma$-Wirkungsquerschnitt wurde intensiv bei 7, 8 und 13 TeV mit dem ATLAS-Experiment untersucht. Um die Top-Photon-Kopplung und ihre Stärke näher zu studieren, ist es notwendig, die Ereignisse zu suchen, bei denen das Photon vom Top-Quark kommt (Top-Strahlung). Diese Prozesse können nicht ereignisbasiert gefunden werden. Es werden multivariate Analysetechniken benötigt. Diese Arbeit ist der erste Schritt um Photonen, die von einem Top-Quark ausgesendet wurden von anderen prompten Photonen, die von den Quarks im Anfangszustand (ISR) oder den Zerfallsprodukten des Top-Quarks kommen (b Quark (FSR(b)) und W Boson und dessen Zerfallsprodukte (FSR(W))), zu trennen. Dies geschieht mit Hilfe neuronaler Netze. Dafür wurden spezielle Datensätze mit einer modifizierten MadGraph-Version erzeugt. Die am stärksten separierenden Variablen waren die invarianten Massen der Top-Quarks, entweder unter Berücksichtigung des Photons oder ohne das Photon. Auf der Ebene der Partonen konnte eine gute Trennung zwischen Top-Strahlung und ISR auf der einen Seite und FSR(b) und FSR(W) auf der anderen Seite erreicht werden. Die Trennung von Top-Strahlung und ISR war der limitierende Faktor. Auf Rekonstruktionsebene verringerte sich die gute Trennungskraft der invarianten Massen aufgrund dessen, dass der Rekonstruktionsalgorithmus für das Top-Quark das Photon nicht mitbetrachtet. Als Konsequenz sind die Resultate nicht so gut wie auf der Ebene der Partonen. Ein weiterer limitierender Faktor auf der Rekonstruktionsebene war die geringe Statistik. Dennoch konnte diese Arbeit zeigen, dass Photonen, die vom Top-Quark abgestrahlt werden, auf einer statistischen Basis von denen separiert werden können, die nicht vom Top-Quark stammen.

Abstract

The $t\bar{t}\gamma$ cross section was studied intensively at 7, 8, and 13 TeV with the ATLAS experiment. To study the top-photon coupling and its strength more deeply, it is necessary to search for events where the photon comes from the top quark (top radiation). These processes cannot be found on an event-by-event basis. Multivariate analysis techniques are needed. This thesis is the first step to separate photons emitted by a top quark from other prompt photons coming from initial state quarks (ISR) or the decay products of the top quark (b quark (FSR(b)) and W bosons or its decay products (FSR(W))) via neural networks. Therefore, dedicated samples with a modified version of MadGraph were produced. The most separating variables are the invariant masses of the top quarks including or excluding the photon in their reconstruction. On parton level, a good separation between top radiation and ISR on the one hand and FSR(b) and FSR(W) on the other hand could be achieved. Separating ISR from top radiation was the limiting factor. On reconstruction level the good separation power of the invariant mass variables was reduced due to the fact that the reconstruction algorithm does not take the photon into account when reconstructing the top quark. As a consequence the results are not as good as on parton level. Another limiting factor on reconstruction level were the low statistics. Nevertheless, this thesis proved that photons coming from the top quark can be separated on a statistical basis from those not coming from the top quark.

Contents

Contents

1. Introduction

Since the discovery of the electron in 1897 by J.J. Thomson [1] and the proof that the atom is mostly empty [2, 3], it is known that the world is not made of continuous matter[1]. The development of quantum mechanics in the 1920's and 1930's and the discovery of new particles, like the positron [5], made particle physics a field in physics in its own right. The ordering given to the different baryons and mesons by Gell-Mann and Ne'emann [6, 7] (named the eightfold way by Gell-Mann), led to the postulation of the existence of quarks [8, 9]. All experimental evidence for this model was achieved by different fixed-target experiments (some of them are mentioned in Section 2.6). The need for higher energies made it necessary to develop collider experiments. The results published by these different experiments led to the development of the Standard Model (SM) of particle physics [10], explaining three of the four fundamental forces and the origin of mass [11–15]. With the Large Hadron Collider (LHC) at CERN, a new era in particle physics began. The discovery of the Higgs boson in 2012 completed the SM [16, 17]. The huge amount of data made it possible to probe this model to an unprecedented level of accuracy, at energies that were never reached before. But there are experimental facts that cannot be explained by the SM. Therefore, probing the SM, as extensively as possible in the search for deviations or new particles, has become the major goal in particle physics. One area for this is top quark physics.

The top quark is the heaviest known elementary particle and was discovered in 1995 at the Tevatron accelerator by the CDF and DØ experiments [18, 19]. Because of its high mass, it might play a crucial role in physics beyond the Standard Model (BSM physics). Hence, its properties are studied intensively. One of those is the gauge coupling to the photon. It depends on two properties. One property is the electric charge of the top quark, which is $2/3e$ according to the SM and defines the strength of the coupling and has direct influence on the cross section of processes involving a top-photon coupling. The other property is related to the structure of the coupling. The SM predicts a vector coupling. But other couplings, like a tensor coupling, cannot be excluded yet [20]. This can be investigated by angular distributions of the emitted photons. To probe these properties, the process

---

[1]See for example J.J. Thomsen's "plum pudding" model [4].

where a $t\bar{t}$-pair is produced in association with a prompt photon is a good testing ground. The difficulty lies in the fact that not only the top quark radiates prompt photons, but also initial state quarks or the decay products of the top or anti-top quark. Until now, there has been no strong effort to suppress these types of prompt photon background. The first step for this are developed in this thesis.

2. The Standard Model of particle physics

2.1. Fundamental particles and interactions

2.1.1. Particle content of the Standard Model

The Standard Model (SM) of particle physics contains two classes of particles: fermions with spin 1/2 and gauge bosons with spin 1 [1]. A graphical representation of the SM particles is shown in Figure 2.1. In addition to those there is also the Higgs boson, which is a scalar particle with spin 0. The corresponding Higgs field allows particles to possess mass (see Section 2.2). The fermions are divided into two groups: leptons and quarks. Each group comprises three generations, containing two quarks and two leptons. The quarks in the first generation are the up and the down quark. The up quark has charge $2/3e$ and the down quark charge $-1/3e$. The two leptons of the first generation are the electron (charge $= -1e$) and the electron neutrino (no electric charge). As protons and neutrons are made up of up and down quarks, nature around us is formed only by those two quarks and the electron. The other two generations are heavier copies of the first and the constituents of both generations only appear at high energies and are unstable, decaying into the known stable particles. Neutrinos behave a bit differently. They are all stable particles but can oscillate between the different flavours [21].
The second generation contains the charm quark ($2/3e$), the strange quark ($-1/3e$), the muon ($-1e$) and the muon neutrino (no charge). The corresponding particles in the third generation are named top quark, bottom quark, tau lepton and tau neutrino. For each of these twelve particles there is also an antiparticle with opposite charge as a consequence of the Dirac equation. In case of the neutrinos, the anti-neutrinos also have no charge. All these particles, as they are fermions, behave according to the Dirac equation as long

---

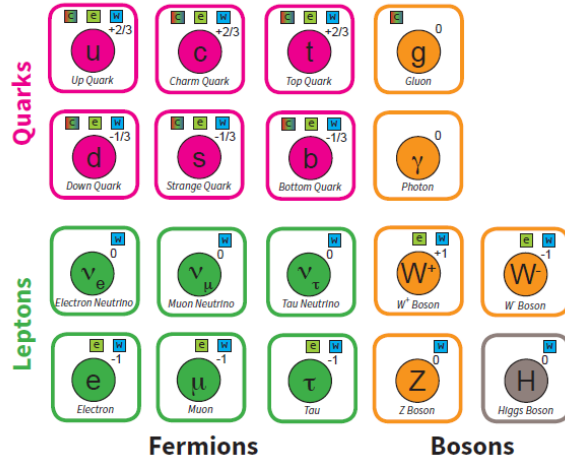[1]In this thesis natural units are used which means that $\hbar = c = 1$.

Figure 2.1.: Graphical representation of all particles in the SM. The number in the upper right corner gives the electric charge and the small boxes indicate how the particles couple (c=colour, e=electromagnetic, w=weak). Copyright: B. Lemmer.

as no interactions with other particles take place. The Dirac equation is given by [22]:

$$(i\gamma^\mu \partial_\mu - m)\psi = 0 \ , \tag{2.1}$$

where ψ is the particle spinor, m the mass of the particle and γ^μ are the Dirac matrices.

2.1.2. Interactions

Interactions between particles are mediated by the four fundamental forces: gravity, electromagnetic, weak, and strong force. Gravity is described by general relativity, which belongs to classical physics, as trajectories are determined. Although a quantum version of gravity is expected to exist, none of the proposed theories can predict observable effects self-consistently. The other three forces are incorporated in the SM as quantum field theories (QFTs). The forces are mediated via the interchange of gauge bosons between the fermions.

The electromagnetic force

In case of the electromagnetic force, the gauge boson is the photon. Combining the Dirac equation and the interaction term between a fermion and a photon, the following equation
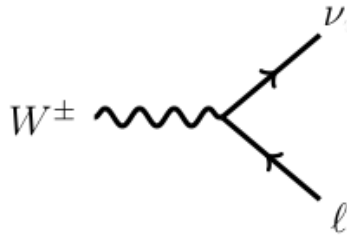
Figure 2.2.: The absorption or emission of a W boson by a lepton, converting itself into a neutrino. The time axis is the vertical axis.

is gauge invariant under local U(1) transformation:

$$i\gamma^{\mu}(\partial_{\mu} + iqA_{\mu})\psi - m\psi = 0 \,, \tag{2.2}$$

where q is the charge of the particle and A_{μ} the gauge field (here: the photon field). A local U(1) transformation means that the spinors ψ, which describe fermions, are transformed according to

$$\psi(x) \rightarrow \psi'(x) = \exp(iq\chi(x))\psi(x) \,. \tag{2.3}$$

Here, $\chi(x)$ is an arbitrary function and U(1) is an Abelian group. No direct couplings between photons are possible, and the photon couples only to charged particles.

The weak force

The presence of the weak force can be observed in nuclear β-decay. The Lagrangian of the weak force is invariant under SU(2) transformations. As a consequence, three gauge bosons are needed corresponding to the generators of SU(2). Two of the three gauge fields combine to the observable $W^{\pm}$ bosons. The third gauge field gives rise to a neutral gauge boson. The weak force couples to a particle property called weak isospin, more precisely to the third component of the weak isospin T_3. All left-handed particles form so called isospin doublets. The two quarks and the two leptons of each generation form such a doublet (see Figure 2.1). The up-type quarks and the neutrinos have $T_3 = +1/2$ and down-type quarks and charged leptons have $T_3 = -1/2$. The W boson can convert a particle into its corresponding isospin partner. For example an electron can be converted into an electron neutrino by "absorbing" or "emitting" a W boson (see Figure 2.2). The neutral gauge boson cannot convert particles into others, but allows scattering interactions between different particles (no flavour-changing neutral currents (FCNC) at leading order (LO)). It is observed that the W boson interacts only with left-handed particles and right-handed
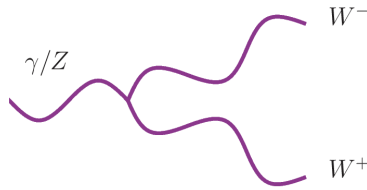
Figure 2.3.: A photon or a Z boson splitting into a pair of W bosons. The time axis is the horizontal axis[2].

antiparticles, which causes maximal parity violation [23, 24]. The actual coupling is called a V−A coupling. As a consequence, the neutral weak gauge boson should couple in the same way. In case of neutrinos this is true, but the weak neutral interaction also couples to right-handed charged leptons. To explain this behaviour, the Glashow-Weinberg-Salam (GWS) mechanism is needed, also called electroweak unification [25, 26]. In fact the U(1) symmetry of QED emerges from the direct product of the SU(2) symmetry of the weak force and another U(1) symmetry. The corresponding gauge field couples to a particle property called hypercharge Y. To match experimental results it turns out that electric charge, weak isospin and hypercharge are related by

$$Y = 2(Q - T_3) \,. \tag{2.4}$$

This gauge field and the neutral gauge field of the weak force mix to form the photon field and the Z field:

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos(\theta_W) & \sin(\theta_W) \\ -\sin(\theta_W) & \cos(\theta_W) \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^{(3)} \end{pmatrix} \,. \tag{2.5}$$

The resulting bosons are the photon and the Z boson, whereas the original gauge fields are called B_μ and $W_\mu^{(3)}$. The mixing is fixed by the electroweak mixing angle (also called Weinberg angle) θ_W. The SU(2) group is not Abelian and therefore interactions between gauge bosons are possible. For example a photon or a Z boson can produce a W^+ and a W^- (see Figure 2.3). The electroweak unification can be described as the direct product of U(1)×SU(2).

The non-diagonal CKM matrix [27] describes the possibility of mixing quark generations within the W decay (see Figure 2.4). It was a development of Cabbibo's idea [28] to relate

---

[2]Figure 2.3 shows a triple gauge coupling. Quartic gauge couplings are also possible.
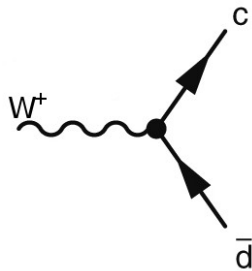
Figure 2.4.: A W$^+$ boson decaying into a charm quark (2nd generation) and an anti-down quark (1st generation). The time is on the horizontal axis.

the mass eigenstates of the down-type quarks with their corresponding weak eigenstates:

$$
\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} .
\tag{2.6}
$$

The postulation of a third generation gave the possibility to introduce CP-violation into the SM, due to a complex phase being part of the CKM matrix. Before, the GIM mechanism (Glashow-Iliopoulos-Maiani mechanism) explained the suppression of FCNC at higher orders by postulating the existence of the charm quark [29].

The strong force

The gauge group of the strong force is the SU(3) and it is related to a property called colour charge. There are three colours (red, green and blue). The only fermions that carry colour are the quarks and as a consequence only they interact via the strong force. In case of anti-quarks, they carry anti-colour. The SU(3) group has eight generators which means that eight gauge bosons are needed. They are called gluons and carry colour and anti-colour. As QCD is not an Abelian theory, self interactions between gluons are possible (triple and quartic gauge couplings). In contrast to particles that have no colour charge, particles with colour charge cannot be observed as free propagating particles. This is due to confinement. The strong force only allows colourless objects to propagate freely through time and space. These colourless objects are called hadrons, subdivided into baryons (containing three quarks or three anti-quarks) and mesons (containing one quark and one anti-quark). More exotic hadrons called tetraquarks and pentaquarks have also been observed [30]. With respect to the age of the Universe, the only stable hadron is the proton. Inside the atomic nucleus, the neutron is also stable. If a quark or a gluon is produced, it hadronises to form these colourless objects. Those can be observed as

particle jets in the detectors. The Lagrangian of QCD also allows the implementation of CP-violation, but experimental data shows that the corresponding parameter is extremely small (only upper limits) [31], and therefore it is assumed that the strong force conserves CP-symmetry.

2.2. The BEH mechanism and the Higgs boson

The mass terms of the massive W and Z boson would break gauge invariance. This can be solved by introducing a complex scalar isospin doublet

$$\phi = \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{2.7}$$

and the following Lagrangian

$$\mathcal{L} = \overline{D_\mu \phi} D^\mu \phi - \frac{m_h^2}{2v^2} \left(\bar{\phi}\phi - \frac{v^2}{2} \right)^2 , \tag{2.8}$$

where m_h is the Higgs boson mass and v is the vacuum expectation value. D_μ is the covariant derivative including the electroweak gauge fields. The potential has the shape of a Mexican hat as shown in Figure 2.5. Expanding the potential from the non-zero minimum, one can obtain a massive particle and three massless particles, called Goldstone bosons. Using the so-called unitary gauge, these massless particles turn out to be the longitudinal degrees of freedom of the gauge bosons, and interaction terms between the new massive field and the gauge bosons are found, where the coupling is given by the mass of the gauge bosons. These terms make the Lagrangian in Equation 2.10 gauge invariant. The mass of the Higgs boson itself is generated by self interaction.

The mass terms for the fermions are gauge invariant but the masses have to be introduced in an ad-hoc manner. The coupling between the Higgs field gives also an explanation for the origin of the fermion masses via a Yukawa coupling.

The masses can now be introduced via the following terms:

$$\mathcal{L}_{Mass} = -\frac{\sqrt{2}}{v}\left[(\bar{\nu}_L,\bar{e}_L)\phi M^e e_R + \bar{e}_R \bar{M}^e \bar{\phi}\begin{pmatrix}\nu_L\\e_L\end{pmatrix}\right] \qquad \text{(electron, muon and tau mass term)}$$

$$-\frac{\sqrt{2}}{v}\left[(\bar{u}_L,\bar{d}_L)\phi M^d d_R + \bar{d}_R \bar{M}^d \bar{\phi}\begin{pmatrix}u_L\\d_L\end{pmatrix}\right] \qquad \text{(mass term of down-type quarks)}$$

$$-\frac{\sqrt{2}}{v}\left[(-\bar{d}_L,\bar{u}_L)\phi^* M^u u_R + \bar{u}_R \bar{M}^u \bar{\phi}^T\begin{pmatrix}-d_L\\u_L\end{pmatrix}\right] , \qquad \text{(mass term of up-type quarks)}$$

$$(2.9)$$

where the bar denotes the adjoint spinor and the $\bar{M}$ gives the mass parameters. The summation over all three generations is not indicated explicitly.
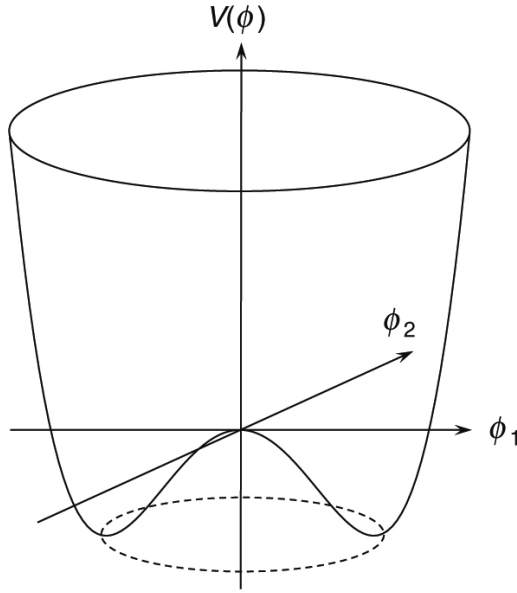

Figure 2.5.: The shape of the Higgs-potential for one complex scalar field.

2.3. The SM-Lagrangian

All the three forces are mathematically described by quantum field theories. The classical Lagrangian of the SM is given by:

$$
\begin{aligned}
\mathcal{L} = & -\frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{8}tr(\mathbf{W}_{\mu\nu}\mathbf{W}^{\mu\nu}) - \frac{1}{2}tr(\mathbf{G}_{\mu\nu}\mathbf{G}^{\mu\nu}) && \text{(dynamical terms of gauge fields)} \\
& + (\bar{\nu}_L, \bar{e}_L)\tilde{\sigma}^\mu i D_\mu \begin{pmatrix} \nu_L \\ e_L \end{pmatrix} + \bar{e}_R\sigma^\mu i D_\mu e_R + \bar{\nu}_R\sigma^\mu i D_\mu \nu_R + (h.c) && \text{(dynamical terms of lepton fields)} \\
& - \frac{\sqrt{2}}{v}\left[(\bar{\nu}_L, \bar{e}_L)\phi M^e e_R + \bar{e}_R \bar{M}^e \bar{\phi} \begin{pmatrix} \nu_L \\ e_L \end{pmatrix} \right] && \text{(electron, muon and tau mass term)} \\
& + (\bar{u}_L, \bar{d}_L)\tilde{\sigma}^\mu i D_\mu \begin{pmatrix} u_L \\ d_L \end{pmatrix} + \bar{u}_R\sigma^\mu i D_\mu u_R + \bar{d}_R\sigma^\mu i D_\mu d_R + (h.c) && \text{(dynamical terms of quark fields)} \\
& - \frac{\sqrt{2}}{v}\left[(\bar{u}_L, \bar{d}_L)\phi M^d d_R + \bar{d}_R \bar{M}^d \bar{\phi} \begin{pmatrix} u_L \\ d_L \end{pmatrix} \right] && \text{(mass term of down-type quarks)} \\
& - \frac{\sqrt{2}}{v}\left[(-\bar{d}_L, \bar{u}_L)\phi^* M^u u_R + \bar{u}_R \bar{M}^u \bar{\phi}^T \begin{pmatrix} -d_L \\ u_L \end{pmatrix} \right] && \text{(mass term of up-type quarks)} \\
& + \overline{D_\mu \phi}D^\mu\phi - \frac{m_h^2}{2v^2}\left(\bar{\phi}\phi - \frac{v^2}{2} \right)^2 , && \text{(Higgs dynamics and potential)}
\end{aligned}
$$

$$(2.10)$$

where D_μ are the corresponding covariant derivatives, the bar over the spinors means the adjoint spinor, σ^μ is the vector with the Pauli matrices and $\tilde{\sigma}^\mu$ obtains negative signs in the spatial components of the vector. The theory is invariant under local gauge symmetry of $U(1)_Y \times SU(2)_L \times SU(3)_C$, where Y refers to hypercharge, L to left-handed isospin doublets and C to colour charge.

To calculate cross sections of different processes it is necessary to know the matrix element. To approximate it, perturbation theory is used. A graphical representation of the different terms in the series expansion of the coupling constants is given by Feynman diagrams. In calculations at higher orders than leading order, divergences appear. These divergences can be treated with a mathematical formalism called renormalisation. A reliable quantum field theory has to be renormalisable. The SM is renormalisable [32–34]. As a consequence the coupling constants for the different forces are in fact not constant. They vary with the energy. This behaviour is called *running*. The fine structure constant ($\alpha \approx 1/137$) increases for higher energies (at 90 GeV (Z boson mass) to 1/128). In case of the weak and the strong coupling the coupling strength decreases. This is especially important for QCD. At low energies the coupling constant α_s is of the order of 1. Hence perturbation theory
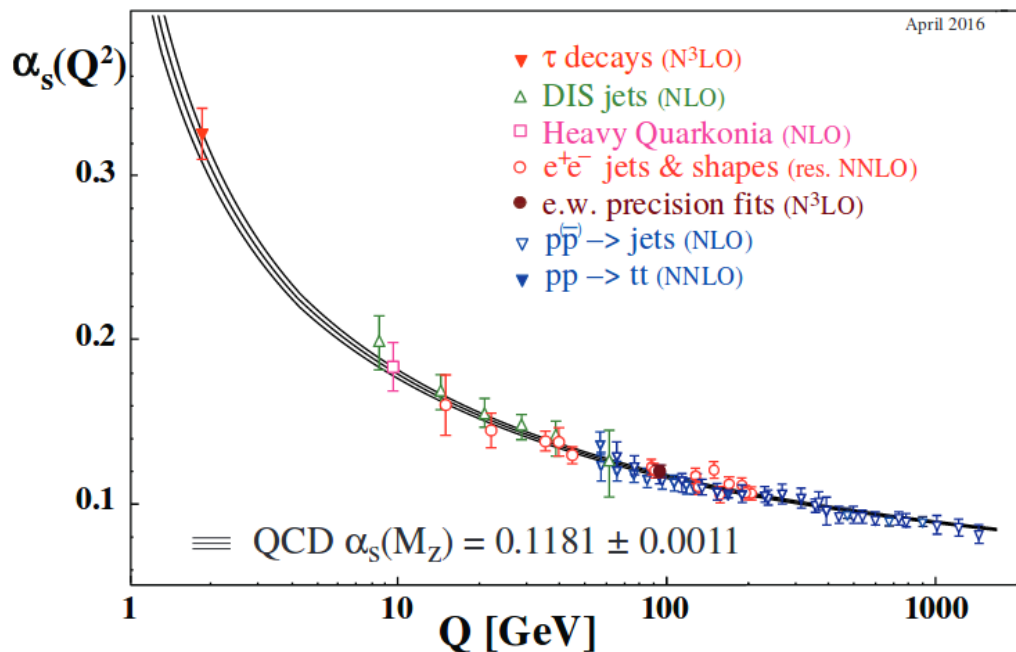
Figure 2.6.: Measurements of α_s at different energies and the theoretical prediction [38].

cannot be applied and confinement is observed. This is the reason why the hadronisation cannot be calculated by first principle calculations. Perturbation theory works only above 1 GeV approximately, when α_s has become small enough. This is called asymptotic freedom [35–37]. The measurements of α_s in comparison to theoretical prediction is presented in Figure 2.6. At the LHC this becomes important. Due to asymptotic freedom, quarks and gluons inside the protons behave like free particles and cross sections can be calculated without taking the interaction between them into account. They are negligible. The SM as given by Equation 2.10 has 19 free parameters. These are the nine masses (or their Yukawa couplings to the Higgs field) of all fermions, except neutrinos, three coupling constants for electromagnetic, weak, and strong coupling, the mass and the vacuum expectation value (see next chapter) of the Higgs boson, the four parameters of the CKM matrix (three angles and one complex phase) and the CP-violation phase in QCD.

2.4. Physics Beyond the Standard Model

Not everything that is observed in nature can be explained by the SM[3]. First of all, gravity is not included. Due to dimensional reasons, it cannot be quantised consistently in a

---

[3]And not all problems will be presented here.

renormalisable way using standard QFT. Another problem is that under normal circumstances quantum-gravitational effects would become measurable at energies at the Planck scale. All theories that predict signals at today's accessible energies have failed so far. Another unanswered question concerns the nature of dark matter. We can observe effects [39] of its presence but no direct detection has been made so far. One theory that could solve the problem would be supersymmetry (SUSY). The original motivation of SUSY although was different. It introduced a symmetry between fermions and bosons to overcome the fact that matter is formed by fermions and interactions are related to bosons. The decay of SUSY particles into its lightest particle, which might be stable, could solve the dark-matter problem. Still, no direct observation of such particles could be made at the LHC[4]. In addition, SUSY could also solve the hierarchy problem.

Another problem that has to be solved in order to explain why the Universe looks exactly the way it looks, is the problem of baryogenesis, closely related to CP-violation. As far as one can observe, only matter is found in the Universe. But the Hot Big Bang should have created an equal amount of matter and antimatter. Most of it annihilated but a small amount of matter could not be annihilated [40], due to CP-violation. The established CP-violation, described by the CKM matrix, is not strong enough to account for what is observed. Thus, more sources of CP-violation must exist. A further shortcoming of the SM is the number of free parameters. Although this could be considered as an aesthetic problem, physicists have the impression that all these fine-tuned parameters need some explanation.

The last question mentioned here, concerns the nature of neutrinos. Since neutrino oscillation was established, we know that neutrinos have mass [21]. How exactly to incorporate neutrino masses into the standard model is still an open question. The Higgs mechanism would work if there were so called sterile right-handed neutrinos that would only interact with the Higgs field. But, as the masses of neutrinos are many orders of magnitude lighter than the electron, it is believed that neutrino masses are generated differently. The most common solution is the seesaw mechanism [41]. It assumes that neutrinos are a mixture of Majorana and Dirac fermions. But so far there is no observation to confirm the Majorana nature of neutrinos.

Explaining most measurements in particle physics, the SM fails to describe some key observations made in experiments. Hence, it is known that the SM cannot be the full theory. Since the discovery of the Higgs boson, particle physics has become mainly a search for unknown phenomena and deviations from the SM to set conditions that theories beyond the SM have to fulfill.

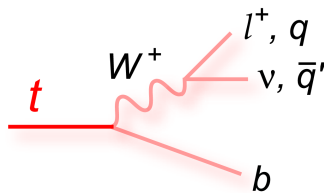[4]There are a lot of different SUSY-theories but so far no prediction could be observed.

Figure 2.7.: Feynman diagram of hadronic or leptonic top quark decay, depending on the W decay.

2.5. The Top Quark in the Standard Model and in BSM physics

The existence of the top quark was postulated by Kobayshi and Maskawa in 1973 [27]. They introduced a third generation of quarks to extend the Cabbibo-matrix to a 3×3 matrix. As a consequence, a complex phase could be introduced that would allow CP-violation as observed by Christenson, Cronin, Fitch and Turley [42]. The down-type quark of the third generation was called bottom quark (b quark) and was found in 1977 [43]. But 18 years were needed to find the isospin partner of the b quark as the top quark was heavier than expected [18, 19]. The current value for its mass is 173.1±0.6 GeV [38]. It is the heaviest elementary particle known today. But not only its mass makes it so interesting, but rather its decay width. The theoretical prediction of the width at NNLO-QCD is 1.33 GeV for a mass of 172.5 GeV [44, 45]. The most accurate direct measurement gives a value of 1.76±0.33 (stat.)$^{+0.79}_{-0.68}$ (syst.) GeV [46]. Such a decay width implies an expected lifetime of the order of 10^{-25} s. Hence, the top quark decays before it can hadronise. This would happen on a time scale of 10^{-23} s [47] and results in the fact that its properties, like spin, are transferred in a predictable way to its decay products and do not vanish by forming hadrons. It is thus possible to study a "bare" quark.

The top quark decays almost exclusively into a b quark and W^+ bosons, whereas the anti-top quark decays into an anti-b quark and a W^- boson. As the W boson can decay hadronically or leptonically, the decay of a top quark results in either three jets (one of them is a b jet) or one b jet, a charged lepton, and a neutrino (observable through missing transverse energy $\not{E}_T$) respectively (see Figure 2.7). Top quarks are produced in pairs (top and anti-top) via the strong force, or singly via the weak force (see Feynman diagrams in Figures 2.8 and 2.9). The following description will only focus on $t\bar{t}$-pair production. In Figure 2.10 measured cross sections at different energies are shown, including measurements at the Tevatron where proton-antiproton collisions were used (for more information about cross section calculations see Section 2.6). The strong production
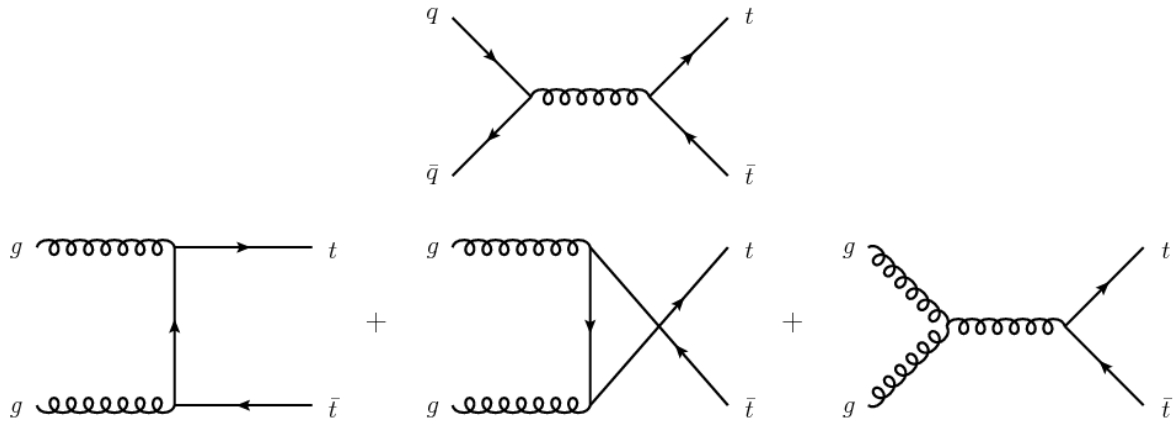
Figure 2.8.: Feynman diagrams for all possible top quark pair production mechanisms. The diagram at the top represents the quark anti-quark annihilation. At the bottom t-,u- and s-channel diagrams for gluon-gluon fusion are shown.
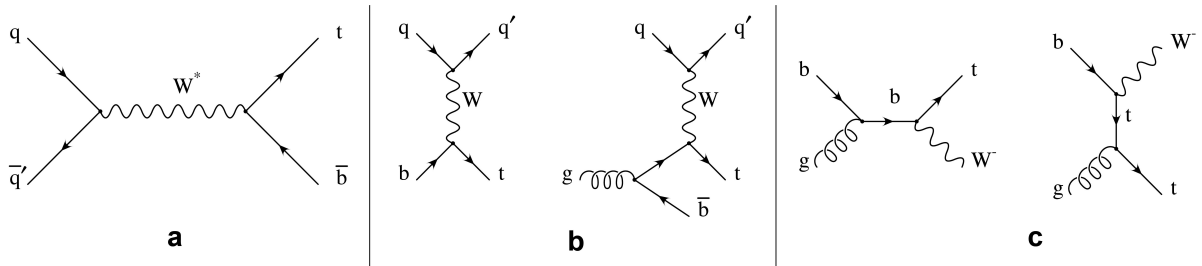

Figure 2.9.: Feynman diagrams for single top quark production via the weak force in the s-channel (a), in the t-channel (b), and for the W-associated production (c).
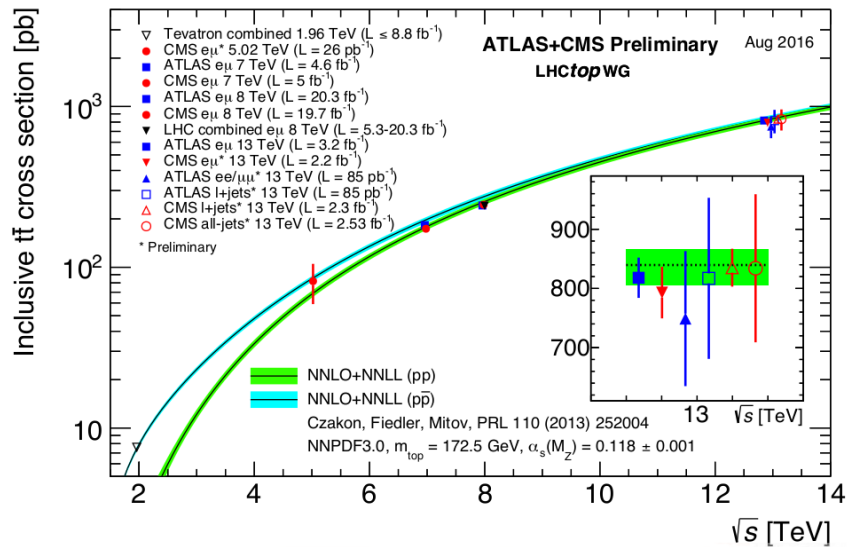

Figure 2.10.: $t\bar{t}$ cross section measurements made at the Tevatron and the LHC [48].
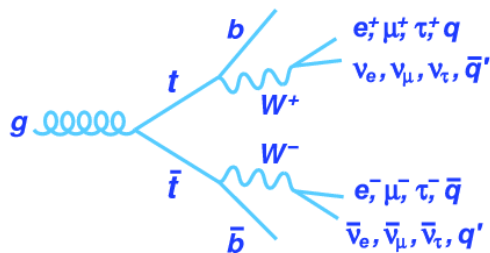
Figure 2.11.: Feynman diagram for all different possibilities in W decay (16 permutations), leading to the definition of three channels.

of top quark pairs is possible via $q\bar{q}$-annihilation or via gluon-gluon fusion. The latter production mechanism is, with about 90%, the dominant one at the LHC at $\sqrt{s} = 13$ TeV (see Section 2.6).

To find events with top quark pairs one can look for three different signatures (see Figure 2.11). The most probable one is that both W bosons decay into quarks and therefore six jets will be observed, where two of them are b jets (all-hadronic channel). This happens in 45.7% of the cases [38]. The advantage of this channel is the large branching ratio, but the reconstruction of top quark momenta is very difficult. b-tagging might help to assign the jets correctly. Another difficulty is the differentiation between signal and background. The major background are multijet events.

The next signature contains one W boson decaying hadronically and the other leptonically, assuming lepton universality. This leads to four jets, where two of them are b jets, a charged lepton and $\not{E}_T$, due to the invisible neutrino. This channel occurs in 43.8% of the time [38], but normally events containing hadronically decaying taus are ignored. Therefore only 29% of all events belong to this channel. Despite providing less events, the clear lepton signature makes it better to identify and the reconstruction is easier. A bit problematic is the neutrino momentum, which cannot be reconstructed unambiguously, due to a quadratic equation. The trade-off between statistics and reconstruction efficiency makes this channel the best for statistically limited analyses.

The dilepton channel where both W bosons decay into leptons is by far the rarest one (10.5% [38]). If hadronically decaying taus are ignored, the fraction reduces to approximately 4.7%. As only two b jets, two leptons with opposite sign, and a large amount of $\not{E}_T$ are expected, this channel is the easiest one to identify. The full reconstruction of the top quark is however problematic, as two neutrinos are present in the final state. In Figure 2.12 all the different possibilities are shown. This thesis will focus only on the

---

[5]The given numbers differ a bit from those given in the text, due to rounding or neglected mass terms.

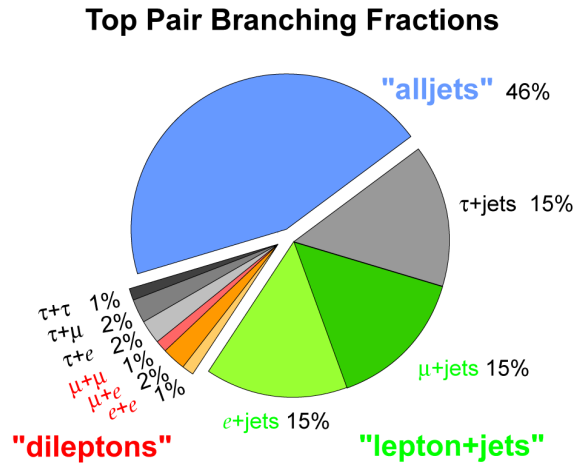**Top Pair Branching Fractions**

Figure 2.12.: All different possibilities that could appear in the decay of $t\bar{t}$-events, due to the different possibilities in the decay of both W bosons[5]

dilepton and lepton+jets channel. For the first one, the main background with prompt leptons is due to Wt-production, but also diboson+jets and Z+jets are non-negligible. Events where a jet or a non-prompt lepton is misidentified as a prompt lepton, must be considered as well. The main contributions here are due to $t\bar{t}$ (one W decaying hadronically), W+jets and t-channel single top. In the lepton+jets channel the main background is mostly the same, but QCD multijet background becomes non-negligible.

Important measurements of top quark properties have been cross section measurements (e.g [49]), mass measurements (e.g. [50]), width measurements (e.g. [46]) and W-helicity measurements (e.g [51]) amongst others. Whereas the coupling to W bosons and gluons can be studied directly in the production and decay of the top quark, its couplings to the other two gauge bosons (Z boson and photon) and the Higgs boson have to be studied in separate processes, where the top quark radiates off one of these bosons. Until now, all measurements show good agreement with the SM.

In many BSM theories the top quark is the window to study these extensions to the SM. For example, interactions to unknown particles would increase the decay width. In some SUSY models the *stop* quark would be the lightest SUSY quark. Its decay could include pairs of top quarks (see Figure 2.13). Due to other SUSY-particles, there would be an excess for $t\bar{t}$-events with a large amount of $\not{E}_T$. Another possibility is is related to the top quark mass. If there are additional heavier Higgs bosons, these would couple primarily to the top quark [52]. So, a heavier Higgs boson could decay into top quarks producing a peak in the invariant mass spectrum of top quark pairs.

Studying the top quark and measuring its properties is not only necessary as it has special properties (for instance no hadronisation), but also to look for BSM physics. Both reasons
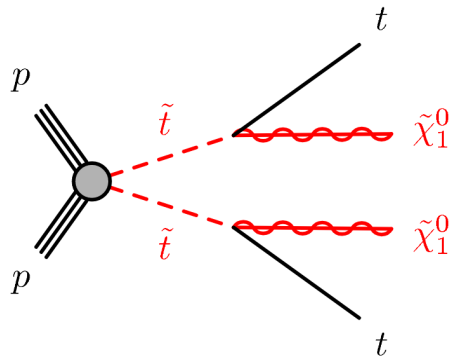
Figure 2.13.: Diagram representing the production and decay of *stop* quarks, including neutralinos.

make top quark physics an important pillar of elementary particle physics and especially in the ATLAS and CMS physics programme.

2.6. The proton structure

The LHC (see Section 3.1) is a proton-proton collider. To understand the collisions, it is necessary to understand the structure of the proton very precisely, as the colliding particles are the partons (quarks and gluons) inside the proton. To calculate cross sections, the momentum distributions of the partons are needed and the factorisation theorem is used, dividing the process into a perturbative and non-pertubative part [53]. The cross section of a particular process $\sigma(\sqrt{s})_{pp \to X+Y}$ is given by

$$\sigma(\sqrt{s})_{pp \to X+Y} = \sum_{a,b \,\in\, \{\text{partons}\}} \int_0^1 \mathrm{d}x_a \int_0^1 \mathrm{d}x_b \, f_a(x_a, \mu_f^2) f_b(x_b, \mu_f^2) \cdot \sigma(\sqrt{\hat{s}}, \mu_f, \mu_R)_{ab \to X+Y} \; , \tag{2.11}$$

where f_a and f_b are the parton distribution functions for partons a and b (PDFs, see below), μ_f the factorisation scale, μ_R the renormalisation scale and $\sqrt{\hat{s}}$ the effective centre-of-mass energy of the colliding partons.

The first systematic studies of the proton structure with deep-inelastic electron-proton scattering at the Stanford Linear Accelerator Center (SLAC) at centre-of-mass energies from a few GeV to 20 GeV [54–59] suggested that the proton is made of point-like spin 1/2 constituents, identified as the quarks (see Chapter 1) or partons [60, 61]. The parton model predicted the observed Bjorken scaling [62] and the Callan-Gross relation [63]. Going to higher energies, it was seen that there are also interactions between the quarks inside the proton (scaling violations) [64]. At even higher energies, higher order QCD effects

become important (sea quarks). A graphical representation of the proton structure can be seen in Figure 2.14. To understand the kinematic behaviour of protons, it is necessary to know the parton distribution functions (PDFs) that describe the momentum distribution within the proton between valence quarks, gluons, and sea quarks. These functions give the probability to find a certain parton with a certain fraction of the proton momentum at a certain momentum transfer. As QCD is a quantum theory, the PDFs are only of statistical nature. Unfortunately, QCD does not predict the PDFs a priori. QCD only predicts the evolution of the PDFs via the so-called DGLAP evolution equations [65–67]. The PDFs were measured extensivly at the HERA-collider at DESY (Hamburg) [68], but also other sets of PDFs exist [69–72]. In Figure 2.15 one can see an example set of PDFs.
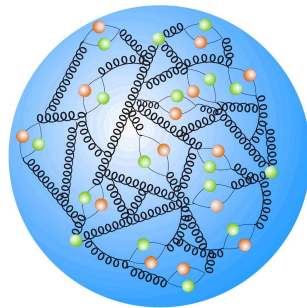
Figure 2.14.: A graphical illustration of interactions expected to occur within the proton, demonstrating the complicated structure. The three single green dots are the valence quarks, emitting and absorbing gluons (springs) which split into quark-antiquark pairs (sea quarks).
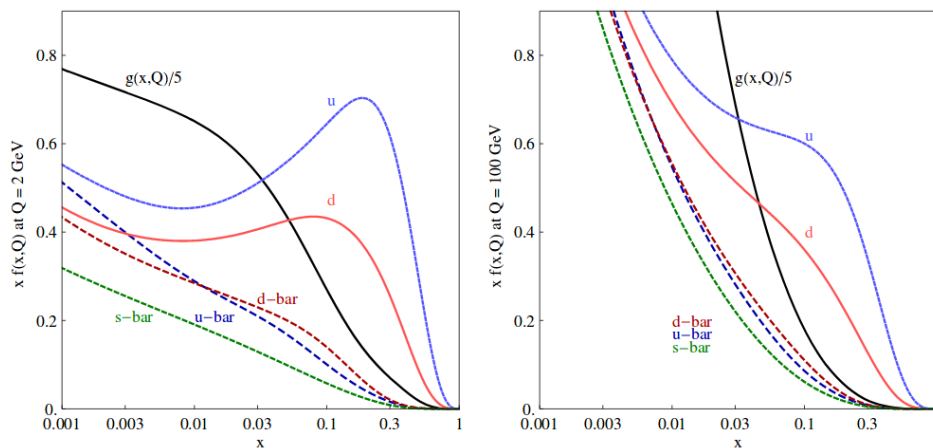

Figure 2.15.: The PDFs for u-, d- and s-quarks and their antiquarks (in case of sea-quarks) at two different scales of momentum transfer (2 GeV and 100 GeV). For the s and s̄, the PDF is the same. The gluon-PDF is scaled down by a factor of 5 [72].

3. The ATLAS detector at the LHC

3.1. The Large Hadron Collider

The idea of a hadron collider at the TeV range in the LEP-tunnel came up at the end of the 1970's, even before LEP was approved. The possible reuse of the LEP tunnel also influenced its design [73]. The idea of such a high-energy hadron collider became more concrete in 1984 [74]. Many discussions followed until LHC-approval in 1996[1] [76]. During the years the SM became more and more established. By the time of LHC approval, only two particles were missing: the τ neutrino (found in 2001 [77]) and the Higgs boson. By the start of the LHC in 2008, the Higgs boson was still not observed, despite the efforts made at the Tevatron and LEP. In addition to that, no new physics phenomena beyond the SM had been observed[2]. So the questions directed at the LHC did not change over the years. The LHC, as it is installed in the LEP-tunnel, has a circumference of nearly 27 km [78]. Its design centre-of-mass energy is 14 TeV (7 TeV for each beam). As protons are collided, synchrotron radiation is not a limiting factor. Essentially the bending magnets limit the energy. They are superconducting, using NbTi as material, and reach a field strength of 8.33 T. Only positively charged particles are accelerated, so the magnets must have opposite alignment which results in the necessity of two beam pipes. A structure was developed, where both beam pipes are enclosed in one magnet structure, sharing the return yoke and the cryostat. This is shown in Figure 3.1, where one can see the field lines of the dipole field. The holes that are near to the edge are used for the liquid helium that cools down the magnets to the temperature regime where they are superconducting. The working temperature lies below 2 K [78]. The general magnet alignment is a so called "FODO"-structure [78]. In addition to the dipole and quadrupole magnets, magnets of higher order are used for small scale corrections, needed to achieve good beam quality. The design luminosity is 10^{34} cm^{-2} s^{-1} [78]. This is achieved by the good focusing abilities and the fact that the LHC is a pp-collider, in contrast to the Tevatron which was a p$\bar{\text{p}}$-collider where the $\bar{\text{p}}$-beam intensity was limited [78].

---

[1]A short review on the history of the LHC and other hadron colliders can be found in this book: [75].
[2]With relevance for the LHC physics programme.
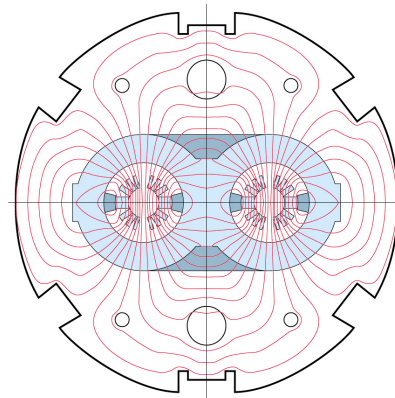
LHC dipole

Computed magnetic flux map at B$_0$=10 Tesla

CERN AC - HE110 - 10/10/95

Figure 3.1.: The magentic flux of the LHC dipole magnet at 10 T. Copyright: CERN.

As the $t\bar{t}$ production cross section is dominated by gluon-gluon fusion (at LHC energies), the advantage of higher cross sections[3] using $\bar{p}$, vanishes (see Figure 2.10). For Run 2, the magnets are not used at full capacity and the beam energy is kept at 6.5 TeV. In comparison to the huge number of magnets (1232 dipole magnets) only 16 accelerator units (radio-frequency (RF) cavities) are installed. They have a voltage of 2 MV. The protons are injected into the LHC with a beam energy of 450 GeV. It takes about 20 minutes to ramp them up to their final energy. The LHC is divided into eight sectors. Each sector is defined as the arch structure between two straight segments. At four of them, the beams are brought to collision. The four major experiments are located there. ATLAS [79] is located at interaction point 1 (IP 1), ALICE [80] at IP 2, CMS [81] at IP 5 and LHCb [82] at IP 8. IP 2 and IP 8 are also used for beam injection. The other four straight segments are called interaction region (IR). At IR 3 and IR 7 collimators are installed to assure beam quality. IR 4 was used to put the RF-cavities and at IR 6 the beam dump is located. Beside the four big experiments, there are smaller ones. TOTEM [83] measures elastic and diffractive cross sections, MOEDAL [84] searches for magnetic monopoles and LHCf [85] looks at hadron interactions to test models used in cosmic ray analyses. All this is shown in Figure 3.2. As mentioned, the protons are injected at an energy of 450 GeV into the LHC. To get this energy, they are accelerated in
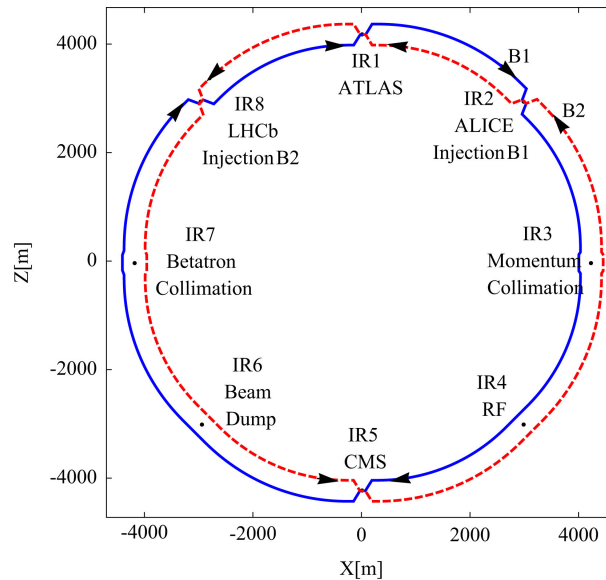
---

[3]Especially at low energies.

Figure 3.2.: The LHC and its structure [86].

a chain of pre-accelerators. The CERN accelerator chain for Run 2 is as follows (see also Figure 3.3): a bottle of hydrogen serves as primary source. The electrons are stripped off by a strong electric field. The remaining protons are accelerated in the LINAC 2 to 50 MeV. Afterwards, they are injected into the Proton-Synchrotron Booster (PSB) where they reach 1.4 GeV. From there, they are moved into the Proton-Synchrotron (PS) and are accelerated to 25 GeV. The last step before the LHC is the Super Proton Synchrotron (SPS) where the protons reach 450 GeV. The beam structure is such that the LHC is filled with bunches of protons. Each bunch contains about $1.1 \cdot 10^{11}$ protons. The LHC is ideally filled with 2808 bunches. The time difference between two bunches is 25 ns and protons are moving very close to the speed of light. It takes[4] 16 minutes to fill the LHC. Heavy ions are also collided at the LHC. Ion collisions are investigated at ALICE, whereas all other experiments focus mainly on pp-collisions. The accelerator chain for ions differs a bit from the one used for protons. Beside the LHC and its experiments other experiments are carried out at CERN, hence the whole structure of accelerators is even more complex, which can be seen in Figure 3.3

3.2. The ATLAS detector

ATLAS is one of the two multipurpose experiments at the LHC. The collaboration was founded in 1992 by merging two collaborations with similar ideas. The ATLAS detector

---

[4]This is the design value as published in [79].
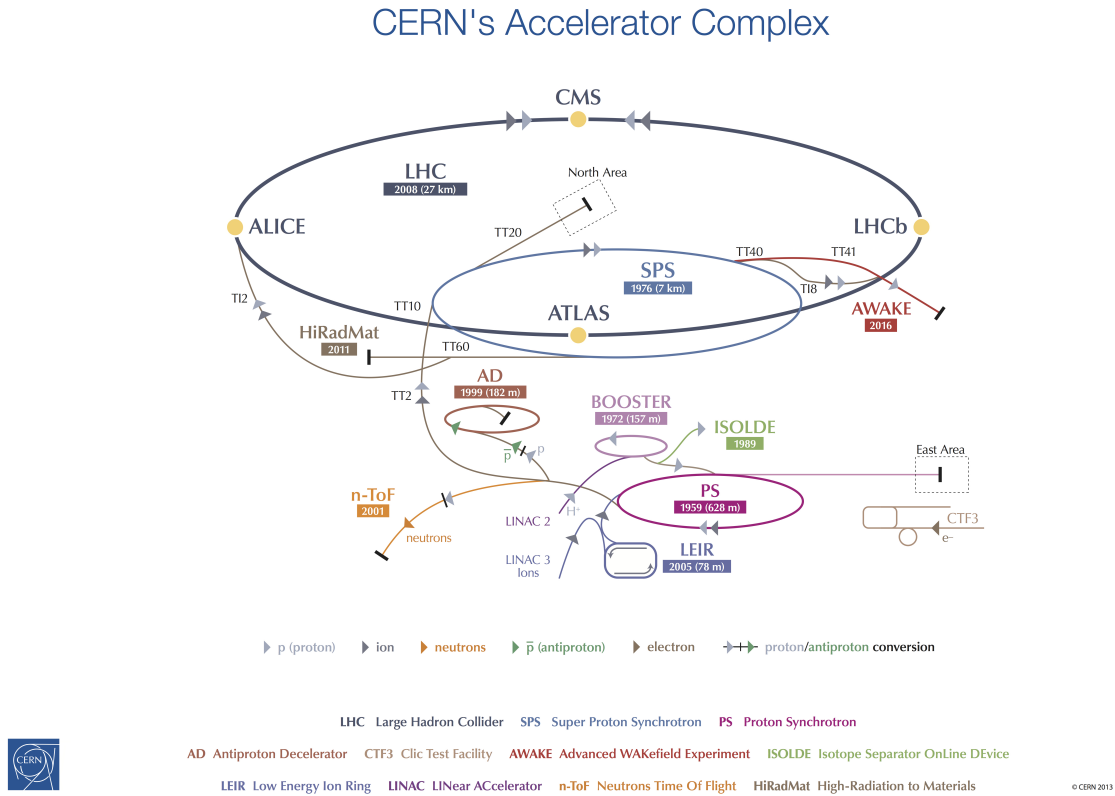
CERN's Accelerator Complex

Figure 3.3.: The accelerator complex at CERN. Copyright: CERN.

is the biggest of all four detectors. It is about 46 m long and 25 m high. It has the typical onion structure, covering nearly the full solid angle around the collision point. Each detector component is divided into a barrel part and an end cap part. Nearest to the beam pipe is the tracking detector, divided into several subsystems. The tracking detector is enclosed by a solenoid magnet. The next part is the calorimeter. There are two calorimeters, first the electromagnetic and than the hadronic. The outermost part of the detector is the muon spectrometer, equipped with a toroidal magnet [79]. The coordinate system for the detector has its origin in the collision point. The z-axis is the beam line. The transversal plane is adjusted such that the x-axis points to the centre of the LHC and the y-axis points upwards in a right handed system. As the ATLAS detector has a cylindrical shape, cylindrical coordinates are used. The polar angle ϕ is measured around the beam pipe with respect to the x-axis. The azimuthal angle θ is often replaced by the pseudorapidity $\eta = -\ln\tan\left(\frac{\theta}{2}\right)$ [79]. The advantage is that differences in η are Lorentz invariant[5] for boosts in z-direction. In Figure 3.4 the whole detector is shown.

---

[5]If masses can be neglected.
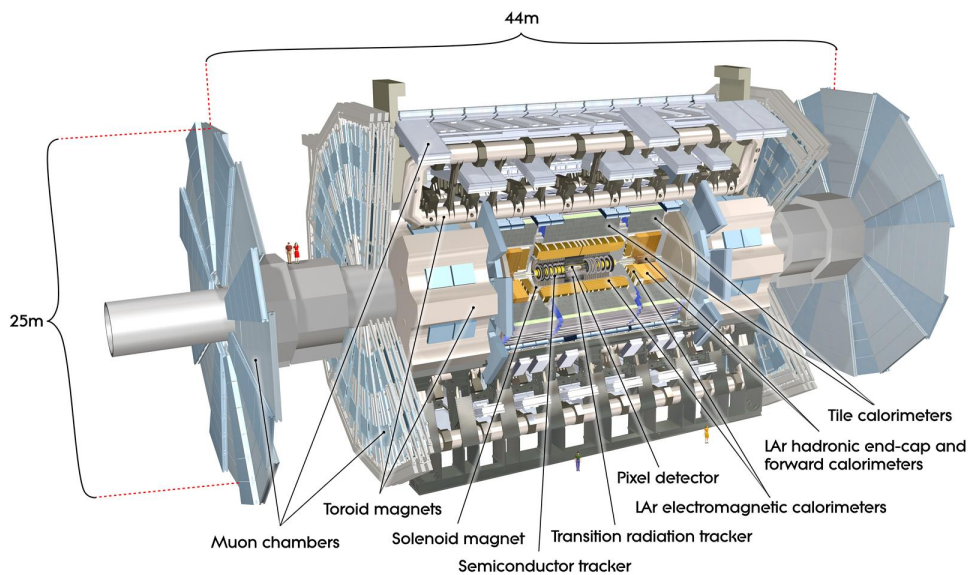
Figure 3.4.: The full ATLAS detector with all its subsystems [79].

3.2.1. Inner detector

The inner detector (ID) is put as close as possible to the beam pipe [79]. It measures the tracks of charged particles and their momentum. To measure the momentum, the ID is surrounded by a solenoid magnet of 2 T. As a consequence, the magnetic field is parallel to the beam pipe. In the transverse plane, charged particles are bent by the Lorentz force which acts perpendicularly on them. The bending radius of the circular trajectory is used to measure the transverse momentum p_T. By reconstructing the particle tracks, the polar angle ϕ and the pseudorapidity η are measured. Combining the transverse momentum and the angular information one reconstructs the 3-momentum of the particle. Tracks are reconstructed by exploiting the hits that they leave in the three innermost parts of the ID[6]. These are the insertable b-layer (IBL), the silicon pixel detector and the semiconductor tracker (SCT). The momentum resolution is proportional to the particle momentum. The overall resolution (including TRT) is $\sigma_{p_T}/p_T = 0.05\% \, p_T$. The IBL is located 3.3 cm away from the beam pipe (see Figure 3.5) [87]. It was inserted after Run 1 to improve b-tagging. It is made of silicon pixels of a size of $250 \times 50 \, \mu\mathrm{m}^2$. A pseudorapidity range of $|\eta| \leq 2.9$ is covered. The next part is the silicon pixel detector [79]. It consists of three layers at 50.5 mm (first layer), 88.5 mm (second layer), and 122.5 mm (third layer) from the beam pipe (see Figure 3.5). The pixel size is $50 \times 400 \, \mu\mathrm{m}^2$. In the end caps three disks are installed on each side. They are 495, 580, and 650 mm away from the centre (see Figure 3.6). The

---

[6]The ID has a fourth part, the transition radiation tracker (TRT) for particle identification.
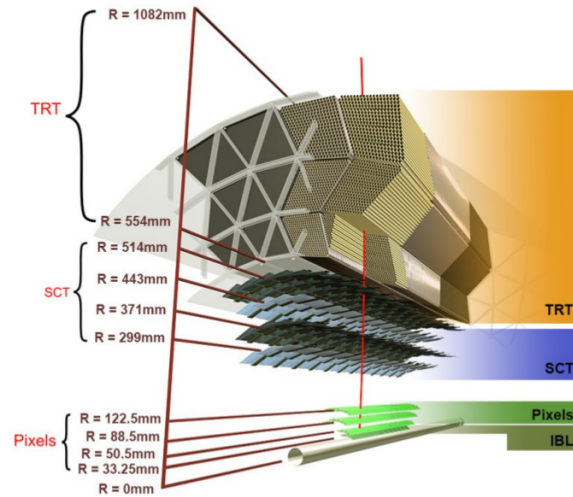
Figure 3.5.: The barrel part of the ID of ATLAS. Copyright: CERN.
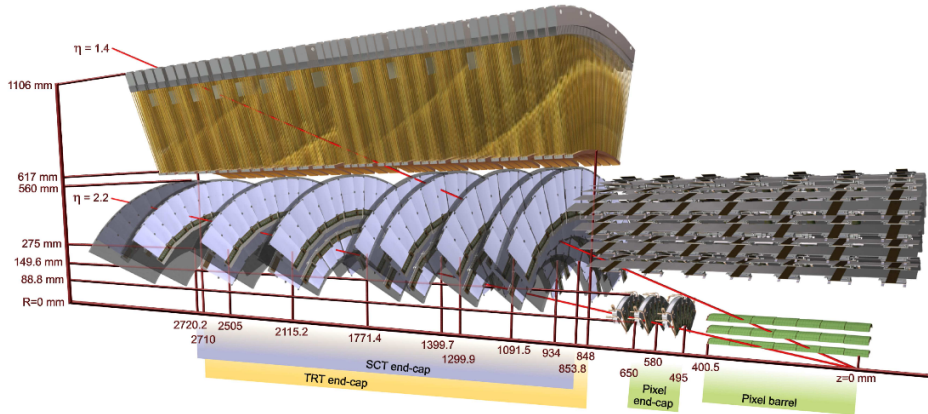

Figure 3.6.: The end cap of the ID of ATLAS with all subsystems. The two red lines indicate particles moving through the ID at different η [79].

pixel size is the same. The spatial resolution is 10 μm in the transverse plane and 115 μm in z/R direction for barrel and end cap, respectively. Signals within $\eta = \pm 2.5$ are detected. Next comes the SCT [79], which is a silicon strip detector. In the barrel part there are four layers between 299 and 514 mm from the beam pipe (see Figure 3.5). In the end caps one finds nine layers on each side, extending the structure up to approximately 2.7 m on each side of the collision point. In contrast to the IBL and silicon pixel detector, the SCT has to deal with a lower track density. This means the granularity does not need to be as high as before to separate the different tracks. Hence, the strips have a pitch of 80 μm. The spatial resolution is 17 μm in the transverse plane and 580 μm in z/R direction for barrel and end cap, respectively. To get a better 2D resolution, strips are located back to back with a stereo angle of 40 mrad between them. The last part of the ID

is the transition radiation tracker (TRT) [79]. It helps to identify particles. Electrons can be identified in the range of $|\eta| \leq 2.0$ which is complementary to the identification by the calorimeter. In contrast to the other parts of the ID, the TRT is a gaseous detector, using Xe and CO_2. It is made of polyimide straw tubes of 4 mm diameter, containing the gas. In the middle of each straw tube is a tungsten wire. Between the straws, radiator material is placed. If particles pass by, they will emit transition radiation when transversing the boundary surface. This radiation ionises the gas, which can be measured and helps to identify electrons and pions. The spatial resolution is 130 μm. In the barrel part the straws are 554 to 1082 mm away from the beam pipe. In the end caps they are located outside the SCT end cap. The whole ID is put into a vessel of about 7 m length and a radius of 1.15 m. The detector operates at a temperature of -5 to -10 °C for the silicon parts and at room temperature for the TRT.

3.2.2. Calorimeter systems

The calorimeter is responsible for the measurement of the particle energy. Therefore, the particles are stopped and the energy deposits are measured. As one differentiates between electromagnetic interacting particles (electrons and photons) and strongly interacting particles (hadrons), the calorimeter is divided into two subsystems. Behind the solenoid magnet the electromagnetic calorimeter is located. Electrons and photons are mostly stopped here. Hadrons and muons will pass through. Behind the electromagnetic calorimeter is the hadronic calorimeter. Using denser material, it stops nearly all hadrons and their decay products. Muons will still pass through into the muon chamber. In contrast to the momentum measurement, the relative uncertainty in the energy measurement improves with increasing energy. After noise subtraction, it is given by

$$\frac{\Delta E}{E} = \sqrt{\left(\frac{a}{\sqrt{E}}\right)^2 + b^2} \, , \tag{3.1}$$

where a and b are constants with predefined values that are set a priori in order to fulfill physics requirements. For the EM-calorimeter $a = 10\%$ and $b = 0.17\%$. In the hadronic calorimeter a and b differ for different particles. The same holds for the forward calorimeter. In both cases, electrons are measured substantially better than other particles[7].

Electromagnetic Calorimeter
The electromagnetic calorimeter (EM calorimeter) [79] consists of two parts, the barrel

---

[7]a from 20 to 30 % for electrons and above 80% for pions.
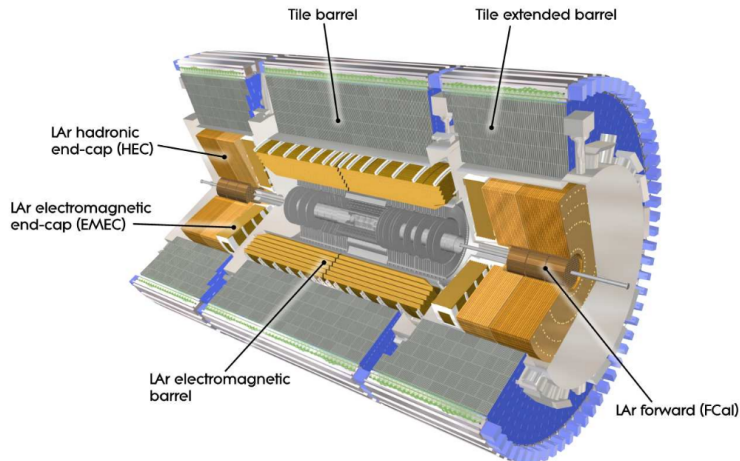
Figure 3.7.: The different parts of both calorimeters of the ATLAS detector [79].

and the end cap part. The structure can be seen in Figure 3.7. They use liquid argon as active material and lead as passive material. The barrel part has an accordion geometry. The barrel part is divided into two half-barrels separated by a small gap. The $|\eta|$ coverage extends to 1.475. The two end cap wheels cover the pseudorapidity of $1.375 \leq |\eta| \leq 3.2$. In the range of $|\eta| \leq 2.5$ the granularity is substantially finer than in the forward region, as the barrel part is used for precision measurements. The barrel part has a thickness of more than 22 radiation lengths[8]. The end cap extends to more than 24 radiation lengths[9].

Hadronic Calorimeter

The hadronic calorimeter [79] consists of three parts. Again, it has a barrel and an end cap section. Additionally, there is also the so called forward calorimeter, located directly around the beam pipe and enclosed by the hadronic end cap calorimeter. The barrel part is split into two sections. One section is located behind the EM-calorimeter (tile barrel) and the other above the end cap wheels (tile extended barrel) (see Figure 3.7). Both barrel structures are made of steel (absorber) and scintillating tiles (active material). The tile barrel covers $|\eta| \leq 1$, whereas the tile extended barrel covers $0.8 \leq |\eta| \leq 1.7$. The barrel part extends to 7.4 interaction lengths. In contrast to the barrel part, the end cap wheels use liquid argon and copper as active material and absorber, respectively. It covers $|\eta|$ within 1.5 to 3.2. The thickness is ten interaction lengths. The last part is the forward calorimeter. It uses liquid argon as its active material. The first module uses copper as its absorber[10], but the other two use tungsten. Pseudorapidity is covered from $|\eta| \geq 3.1$ to $|\eta| \leq 4.9$. The spatial extension corresponds to ten interaction lengths.

---

[8]After one radiation length the energy of a particle is reduced to a fraction of 1/e.

[9]Quoted interaction lengths do not include the material used for the outer support of the calorimeters.

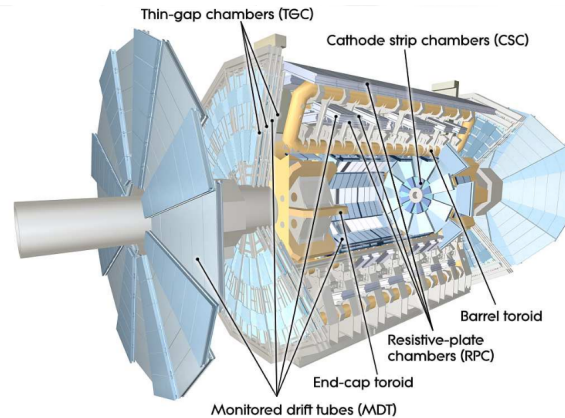[10]It shall measure EM-interactions as no forward EM-calorimeter exists.

Figure 3.8.: The muon chamber and its different subsystems [79].

3.2.3. Muon chamber

The muon chamber [79] (see Figure 3.8) is the largest detector system. It measures the momentum of muons which are not absorbed in the calorimeters. The magnetic field for bending the muon trajectory is toroidal. Its field strength varies in the barrel part from 0.15 T to 2.5 T and in the end cap from 0.2 T to 3.5 T. It uses different gaseous detectors. The tracking is done by monitored drift tubes in the barrel ($|\eta| \leq 1.4$) and the end cap part ($1.6 \leq |\eta| \leq 2.7$). In the end cap, cathode strip chambers are used to support tracking ($2.0 \leq |\eta| \leq 2.7$). Triggering is done by resistive plate chambers in the barrel part and with thin gap chambers in the end cap. The resolution is proportional to the momentum (as in the ID) and is approximately 10% of the muon p_T (for $p_T = 1$ TeV) .

3.2.4. Trigger

As the time between two bunch crossings is 25 ns, the collision rate is 40 MHz. Due to pile-up, up to 35 proton-proton collisions happened on average per bunch crossing in 2015/2016[11] [88]. The resulting amount of data cannot be stored. To filter the interesting events a trigger is used. In Run 1, the trigger had three levels, reducing the recorded data to approximately 200 events per second [79]. For Run 2, the trigger system has two levels [88]. The level 1 trigger (L1) has access to information from small cluster cells within the calorimeters and the muon chamber. It is hardware based. From the information, the trigger builds regions of interest (RoI) and decides to pass these on to the second level. This decreases the rate to 100 kHz. Each decision has to be made within 2.5 μs. The second level is called the High Level Trigger (HLT) and is software based. It has full

---

[11]So approximately 1.4 billion proton-proton collisions happened each second. In 2017 the pile-up rate was even higher with 40 to 50 proton-proton collisions per bunch crossing.
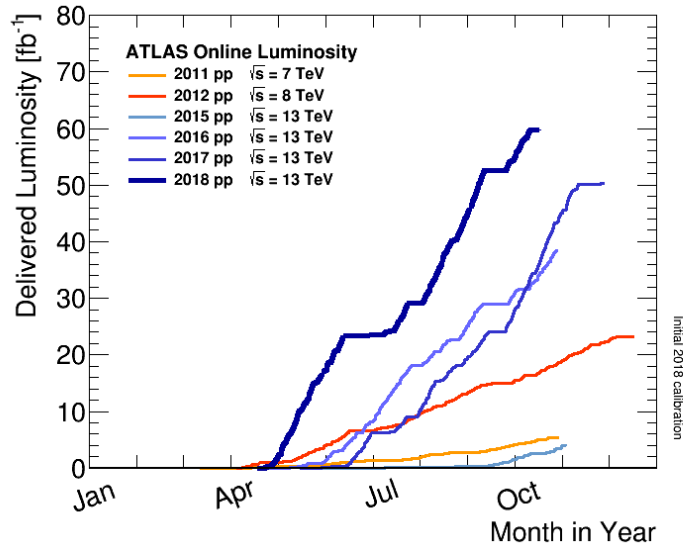
Figure 3.9.: Yearly time development of integrated luminosity recored from 2011 -2018.

access to the data from the calorimeter. Based on this and the RoIs, it decides to keep an event or discard it. The rate of events stored is about 1 kHz. The decision is made in 300 ms.

3.2.5. Luminosity

The LHC was designed to produce a beam with excellent focusing, short bunch crossing time and high proton density within one bunch. This provides the conditions to get high luminosity. The instantaneous luminosity is given as

$$\mathcal{L} = f \frac{n_b n_1 n_2}{4\pi \sigma_x \sigma_y} \ , \tag{3.2}$$

where f is the frequency of bunch crossings, n_b the number of bunches in the LHC, and $n_{1,2}$ the density of particles in each bunch. $\sigma_{x,y}$ is the width of the beam in both transverse directions, assuming a 2D-Gaussian beam shape in the transverse plan. The number of events N is then given by integrating $\mathcal{L}$ over time (integrated luminosity) and multiplying with the cross section σ of a certain process:

$$N = \sigma \int \mathcal{L} \mathrm{d}t \ . \tag{3.3}$$

So a higher luminosity means more data. Especially for rare processes, where the statistical uncertainty is still dominant, this is important. The data ATLAS recorded over the past years is depicted in Figure 3.9. In Run 2 nearly 150 fb^{-1} were recorded.

4. Measuring the $t\bar{t}\gamma$-process with the ATLAS detector

4.1. Summary of the analysis strategy and the motivation to measure the $t\bar{t}\gamma$-process

As described in Section 2.5, top quarks can be produced in pairs (top and anti-top) via the strong force. In the production and decay process, the standard model predicts the possibility that a photon is emitted by one of the participating particles. Besides the top quark emitting a photon, such a photon from the hard process (prompt photon) can also originate from initial state quarks (initial state radiation; ISR) or from the decay products of the top quark (b quark, W boson, decay products of the W boson; radiative decay in general) (see Figures 4.1 to 4.4). Events where the photon is radiated off the top quark can be used to study the coupling between the two particles. Additionally, the top quark charge can be measured relatively model independently. Until now every measurement indicates that the top quark has its predicted charge of $+2/3e$. But the measurements carried out so far (e.g. [89]), rely on the classification of b jets as originating from b quarks or anti-b quarks. So far there is no measurement of the top quark charge using the coupling between the top quark and the photon directly. The analyses from ATLAS at 7 and 8 TeV [90, 91] and the Run 2 analysis [92][1] do not differentiate between the sources of different prompt photons within the $t\bar{t}\gamma$-process. They measured the cross section requiring the photon just to be a prompt one from the $t\bar{t}\gamma$-process. The results are compatible with the SM predictions.

The measurements for 7 and 8 TeV were done in the lepton+jets channel. One of the main backgrounds were the so called hadronic fakes. This background arises when either photons are produced in the jets or hadrons are misidentified as photons. To suppress them, an isolation cut on the photon candidate is applied. Other background sources are electrons misidentified as photons and prompt photons from processes that look similar

---

[1]Paper in submission process.
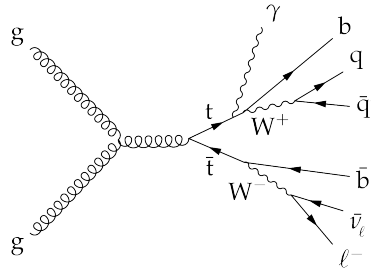
Figure 4.1.: Feynman diagram for the $t\bar{t}\gamma$-process (s-channel), where a real top quark radiates a photon. t-channel production is also possible but not shown here.
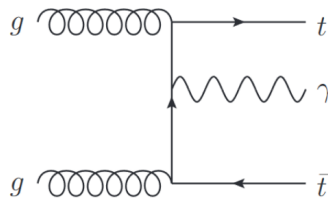
Figure 4.2.: Feynman diagram for the $t\bar{t}\gamma$-process, where a virtual top quark radiates a photon. The top quark decay is not shown.
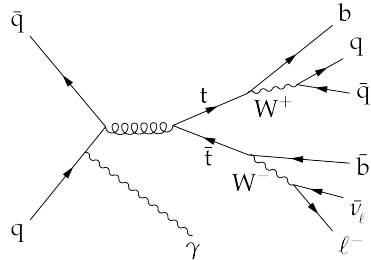

Figure 4.3.: Feynman diagram for the $t\bar{t}\gamma$-process, where an initial state quark radiates a photon (ISR).
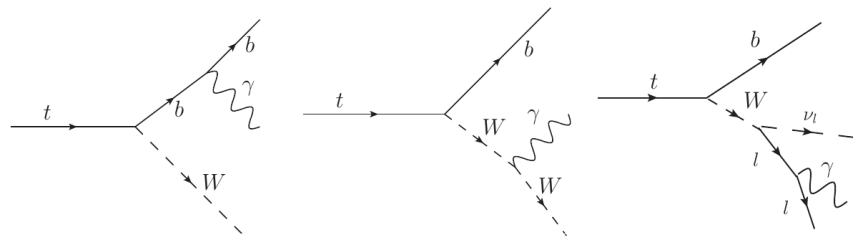

Figure 4.4.: Feynman diagrams for the top quark decay where one of the decay products radiates a photon (radiative decay). Top quark production is not shown here.

to $t\bar{t}\gamma$, for instance the multijet background in association with a prompt photon. For both types of background, a likelihood fit was made to extract the expected number of signal events. The generation of the templates used in the fit was done differently (simulation vs. data driven techniques) for the two. The p_T^{iso} variable was chosen as the best discriminating variable between signal and background. It is defined as the sum of all p_T in a cone of $\Delta R < 0.2$ around the photon. But to really measure the top quark charge, or the top-photon coupling, the uncertainties for the results were still too large.

For 13 TeV, the analysis strategy has changed. Most important may be the fact that now the dilepton channel is also included into the analysis. To suppress hadronic fakes, the so-called Prompt Photon Tagger (PPT) is used [93]. It is a neural network (NN) trained to separate between hadronic fakes and prompt photons. This tool is coupled to another NN, the event level discriminator (ELD), to finally separate out the prompt-photon background. To suppress also prompt photons from the radiative decay of the top quarks, a ΔR-cut[2] between the prompt photon and the lepton is applied. ISR is still not considered to be background and even the photons from radiative decay are not strongly suppressed. To extract the cross section, a likelihood fit on the ELD distribution was done. Furthermore differential cross sections were extracted.

4.2. Extending the analysis

So far, the measurement of the $t\bar{t}\gamma$-process is "just" a cross section measurement, where every photon from the hard process is considered to be signal. But the most dominant contributions are from the radiative decay (b quarks, W bosons or its decay products radiating off the photon). If one wants to go beyond a pure cross section analysis and attempt to investigate the nature and/or the strength (charge) of the coupling between the top quark and the photon, it is necessary to know which particle in the hard process actually emitted the photon. On an event-by-event basis this is not possible, but the kinematic behaviour, although similar, is not identical for the different sources. The differences seen in the different variables (see Section 7.1) can be used for multivariate analyses (MVAs) to differentiate events on a statistical level, and hence purify the data sample to contain mostly events where the top quark is the mother particle of the photon. This strategy is investigated in this thesis.

---

[2]$\Delta R = \sqrt{(\phi_1 - \phi_2)^2 + (\eta_1 - \eta_2)^2}$, where the indices refer to the two corresponding particles.

Category	Description	Name	*Loose*	*Tight*
Acceptance	$|\eta| < 2.37$, with $1.37 < |\eta| < 1.52$ excluded	-	✓	✓
Hadronic leakage	Ratio of E_T in the first sampling layer of the hadronic calorimeter to E_T of the EM cluster (used over the range $|\eta| < 0.8$ or $|\eta| > 1.37$)	R_{had1}	✓	✓
	Ratio of E_T in the hadronic calorimeter to E_T of the EM cluster (used over the range $0.8 < |\eta| < 1.37$)	R_{had}	✓	✓
EM middle layer	Ratio of 3×7 $\eta \times \phi$ to 7×7 cell energies	R_η	✓	✓
	Lateral width of the shower	$\omega_{\eta2}$	✓	✓
	Ratio of 3×3 $\eta \times \phi$ to 3×7 cell energies	R_ϕ		✓
EM strip layer	Shower width calculated from three strips around the strip with maximum energy deposit	ω_{s3}		✓
	Total lateral shower width	$\omega_{s\,tot}$		✓
	Energy outside the core of the three central strips but within seven strips divided by energy within the three central strips	F_{side}		✓
	Difference between the energy associated with the second maximum in the strip layer and the energy reconstructed in the strip with the minimum value found between the first and second maxima	ΔE		✓
	Ratio of the energy difference associated with the largest and second largest energy deposits to the sum of these energies	E_{ratio}		✓

Table 4.1.: Discriminating variables used for loose and tight photon identification. Table taken from [94].

4.3. Photon reconstruction and identification at ATLAS

To measure the $t\bar{t}\gamma$-process, a crucial part is to reconstruct and identify photons. This is done in two steps [94]. The first step is the reconstruction. It is done in parallel to the electron reconstruction, as both particles have similar signatures in the EM-calorimeter. The reconstruction is based on several steps. The first step is to look for EM-calorimeter cells which build a cluster of energy deposits. Afterwards, tracks from the ID are loosely matched to the clusters in order to find electrons or converted photons. Then seed clusters are matched to possible conversion vertices. In the end an algorithm, using all the tracks,

momentum, and energy information, decides whether a signal is an electron, a photon, or a converted photon. After the reconstruction of a signal the identification process is applied. This refers to checking whether the photon is a prompt photon or a background photon. Therefore, so-called shower shape variables are used (see Table 4.1). These variables can describe, for example, how narrow a shower in the EM-calorimeter is. Prompt photons are known to have narrower showers than background photons. Other information used is leakage information, which indicates how much energy a photon has deposited in the hadronic calorimeter. Hadronic fakes are more likely to have a large leakage into the hadronic calorimeter. Several other discriminating variables are used for the photon identification. There are two possible categories, loose and tight. Loose identification means that only a few variables are used. The loose category is optimised to give a good prompt photon efficiency. In case of the tight selection, more variables are taken into account, which reduces the efficiency, but increases the purity. As the calorimeters have a different granularity and geometry in different η regions, the photon identification is optimised differently for seven η intervals.

5. Introduction to machine learning and neural networks

Experiments in particle physics produce an enormous amount of data, which is good as statistical uncertainties become small and even rare processes can be observed. Rare processes are often difficult to distinguish from their backgrounds, especially if the background processes have higher cross sections. To help to distinguish them, machine learning can be applied. This application of machine learning is called classification, which is the most common use of machine learning in particle physics. Other applications (not necessarily physics applications) are pattern or speech recognition, regression or the reproduction of patterns. Machine learning can be done in different ways. In particle physics, supervised machine learning is commonly used. It means that the training algorithm knows the labels (categories) of the data. In contrast to that, unsupervised learning uses data without any labels and the algorithm generates its own labels. This can be used to search for unknown structures within the data. A third way is semi-supervised learning. It means that from a fraction of the data the labels are known and from another part they are unknown. There are several methods of machine learning. All of them combine the small differences of various variables to get a better overall separation. Here the focus lies on neural networks (NN). The first models of machine learning were linear models such as [95]

$$f(\mathbf{x}) = \sum_{i=1}^{N} \omega_i x_i + C \; , \tag{5.1}$$

where the vector $\mathbf{x}$ is called a sample, each entry x_i is called a feature and C is a constant bias term. The weights for each feature are given by ω_i. This function maps the multidimensional space of the features onto the real numbers $\mathbb{R}$. Choosing good weights might help to classify samples into different categories according to their value $f(\mathbf{x})$. The first so-called perceptron was introduced in 1958 [96]. A perceptron is a model that can learn to adjust the weights based on labels given to the samples. This model was not able to learn the *exclusive or* [97]. The problem was solved by the introduction of the backpropagation algorithm [98].
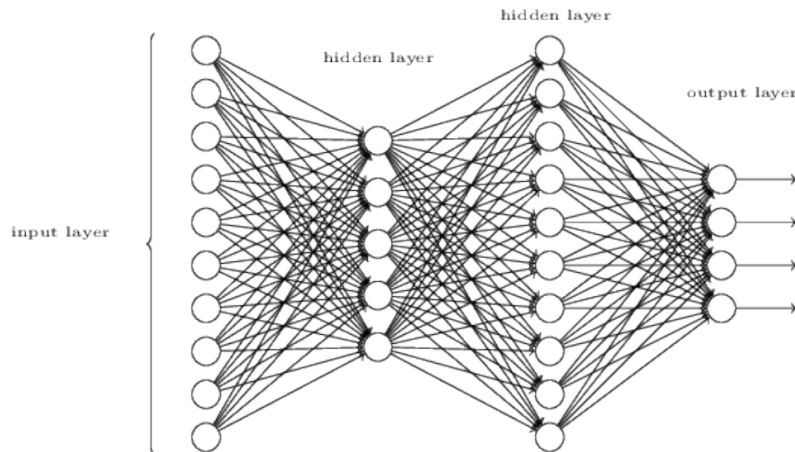
Figure 5.1.: A feedforward NN with two hidden layers. The input layer has 10 nodes, the first hidden layer 5, the second 10 and the output layer has 4 nodes. In this case it is a *complete* multipartite graph.

There are different types of NNs. As feedforward NN are used in this analysis, the focus lies on them. The structure of a NN network is relatively simple (see Figure 5.1). It has at least two layers, which are called input and output layers. Additionally it is possible to add so-called hidden layers in between. In Figure 5.1, two hidden layers are depicted. Each layer consists of a certain number of nodes (perceptrons). There are no connections between the nodes of each layer. The nodes are only connected to nodes (not necessarily all of them) of the previous and consecutive layer. In terms of graph theory a feedforward NN is a multipartite graph. Mathematically, each perceptron works as follows: using the weight vector $\boldsymbol{\omega}$, Equation 5.1 assigns a value for each node to each sample. Each node has a different weight vector $\boldsymbol{\omega}$. The actual output o of each node is calculated by a so-called activation function. It is a nonlinear function h. For instance, $\tanh(x)$ can be such a function. Other activation functions are the softmax function, sigmoid function or the rectifier function. It follows that

$$o = h(\mathbf{x}) \ . \tag{5.2}$$

The input information for the nodes of the input layer is the feature vector $\mathbf{x}$. In Figure 5.1 each node of the input layer assigns a value to a sample. These ten values are forwarded to all the nodes in the second layer. Each node in the hidden layer has its own weight vector $\boldsymbol{\omega}$ and uses the ten values from the input layer as the feature vector $\mathbf{x}$. This procedure is repeated until each sample is processed through the whole NN.

Each output node stands for one label or category in which the data shall be divided.

The higher the output value is, the more likely the sample belongs to that category. In case of binary classification only one output node is needed. The second node would be redundant. If the activation function for the output layer is chosen to be the softmax function, then the output will be between 0 and 1 and can be interpreted as a posterior probability, as the values of all output nodes add up to 1. The best collection of weights for each node is found in the training process.

For training purposes, all samples are divided into groups. Such a group is called a batch. After each batch is processed by the NN, the weights are adjusted. Instead of adjusting the weights after all samples have been processed, this procedure is used to speed up the learning process. The goal of training a NN is to minimise the loss function l. In case of binary classification, the output can be interpreted as a conditional probability $P(Y = 1|\mathbf{x})$ for the label $Y = 1$ for a certain sample $\mathbf{x}$. Consequently $1 - o = P(Y = 0|\mathbf{x})$. Combining both expressions gives

$$P(y|\mathbf{x}) = o^y (1 - o)^{1-y} \ , \tag{5.3}$$

where y stands for signal ($y = 1$) or background ($y = 0$). The loss function is defined as the negative log-likelihood and is called binary cross entropy. It is given by

$$l = -\sum_{i=1}^{N} [y_i \log(o_i) + (1 - y_i) \log(1 - o_i)], \tag{5.4}$$

where the sum loops over all training samples and tends to go to 0 if the samples are classified with a high probability correctly. In case of multiclass classification, the function has to be adopted appropriately and is then called categorical cross entropy. To adjust the weights and minimise the loss function backpropagation is used. It is an algorithm based on partial derivatives and the chain rule.

After all batches are processed by the NN an epoch is finished. Training NNs may need few or many epochs depending on the model. To quantify the quality of the training, the receiver-operator-characteristic curve (ROC-curve) and especially the area under the curve (AUC) are good measures. An example is shown in Figure 5.2. A ROC-curve plots the background rejection versus the signal efficiency. To decide if an event is background or signal, one cuts on the NN output value. All events with a higher output value than some predefined value are accepted. The signal efficiency gives the fraction of signal events that were accepted and the background rejection gives the fraction of background that was not accepted. The best ROC-curve would have an AUC of 1. The worst result would be a line from the top left to the bottom right with an AUC=0.5. This would be the same
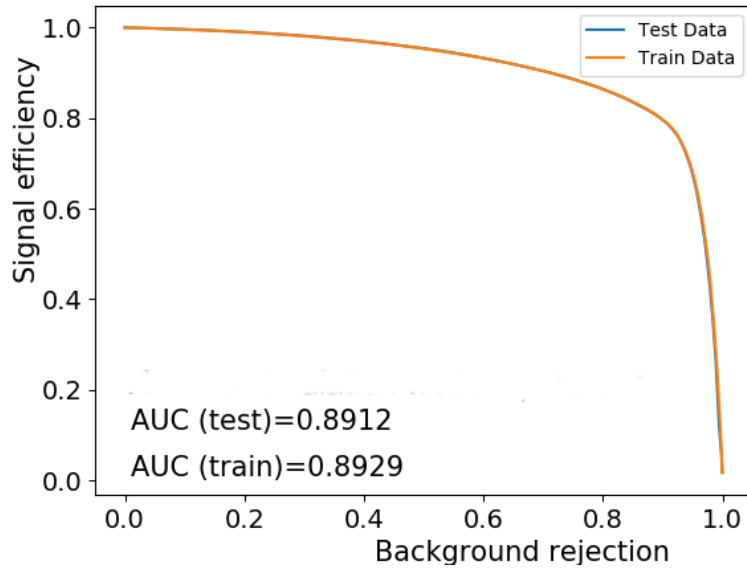
Figure 5.2.: An example for a ROC-curve with nearly 90% AUC for both, testing and training data. No overtraining is observed.

as tossing a coin to decide whether an event is signal or background[1]. This procedure is made for training and testing data.

The training is done only on a subset[2] of all labelled data. The trained NN is then applied to the rest of the data for testing. Testing is necessary, because the training subset contains statistical fluctuations. If the network is trained for too many epochs, it can get sensitive to those fluctuations. That would result in a worse outcome for the testing data. This problem is called overtraining. If no overtraining happens then the ROC-curve for testing data should be the same or only give a slightly worse result than the training data.

A further approach to check the consistency of a NN is k-fold cross validation. This means that the whole data set is split into k subsets of equal size. Then the network gets trained k times. Each time $k-1$ subsets are used for training and one set for testing. All ROC-curves should look similar and differences between test and training curve should be minimal. If not, the network might be overtrained.

---

[1]If signal and background are equally likely.
[2]For Figure 5.2, 80% of the data were used for training.

6. Event generation

6.1. Matrix element generation with MadGraph

The $t\bar{t}\gamma$ analysis at 13 TeV uses MadGraph_aMC@NLO [99] as its Monte-Carlo generator to generate events for the hard process. MadGraph is a next-to-leading order (NLO) generator. As NLO event generation for the $t\bar{t}\gamma$ process would be too CPU-intense, events were generated at leading order.

Looking into the Les-Houches-Event-files (LHE-files) [100], the truth record for the photon gives the following four scenarios for the mother particle of the photon:

- A gluon is the mother particle of the photon. Physically this is senseless but those events must be interpreted as events where an off-shell top quark emitted the photon.

- If light quarks are the mother particle, it is an ISR photon or an off-shell top quark.

- If a W boson is named to be the mother particle, the W boson itself or one of its decay products emitted the photon (FSR(W)).

- A top quark is named to be the mother particle of the photon. These events mix two categories which are the following ones:

  - An on-shell top quark radiated the photon.

  - An off-shell b quark radiated the photon (FSR(b)).

The second and the last category, mixing ISR and off-shell top radiation, as well as mixing radiation from on-shell top quarks and off-shell b quarks, make it necessary to implement small changes in MadGraph. These changes are necessary because otherwise one cannot differentiate between the two components in the relevant categories named before. As described in Section 4.2, to measure the top-photon coupling it is necessary to know which photon was emitted by the top quark. To understand the kinematics of each relevant category (ISR, radiation from top quark, radiation from b quark and from a

W boson or its decay products) one needs the possibility to differentiate between them[1]. This is achieved with the modifications described in the next section.

6.2. Modifying MadGraph

The MadGraph package was downloaded as a standalone version from the MadGraph homepage[2]. To separate off-shell top radiation from ISR and on-shell top radiation from FSR(b) some changes were applied to the MadGraph code (see Appendix A). These changes give the coupling between the top quark and the photon its own name. This does not allow to produce a single sample, where all categories named in Section 6.1 can be differentiated further than in the nominal MadGraph version, but it can be used to produce one sample with events where the photon comes only from the top quark and a sample where all other processes (ISR and radiative decay) are considered[3]. For both samples, one million events were produced in the dilepton and lepton+jets channel. In a leptonic decay of a W boson only electrons and muons were allowed as charged leptons[4].

6.3. Closure test of modified MadGraph

The production of the samples described before leads to a problem. The Feynman diagrams for ISR and for top radiation interfere. As both categories treated separately, one neglects interference terms. In this section it is shown that, nevertheless, one can work with those samples as long as the chosen variables are not sensitive to the neglected interference terms and variable distributions agree within statistical uncertainties. To prove that, dedicated samples were produced, where the production mechanism for the $t\bar{t}$-pair was limited to $q\bar{q}$ annihilation. This was done because only for $q\bar{q}$-annihilation ISR is possible. A sample allowing gluon-gluon-fusion as well, would have given lower statistics for $q\bar{q}$ initiated events, since gluon-gluon-fusion is by far the dominant production mechanism at $\sqrt{s} = 13$ TeV. Hence, fewer events, where the negligence of the interference terms could be of any relevance, would have been generated. For both categories, where top radiation is mixed with ISR or FSR(b), control plots were made for all variables that were later used in the training of NNs. The two other categories are not affected by the

---

[1]Only those events where a top quark emitted the photon are considered to be signal. All other categories will also be referred to as background.

[2]http://madgraph.phys.ucl.ac.be/

[3]This is achieved by requiring the exact number of QCD, QED and TGA (the new top-photon coupling) vertices in the event generation.

[4] The run card and parameter card (defining model and physics parameter) were chosen to be the default ones given by MadGraph itself.

changes and therefore no control plots were produced. All compared distributions agree within statistical uncertainties (example distributions are shown in Figures 6.1 and 6.2), but the ΔR distribution in Figure 6.3 shows deviations of about 10% around the peak. So far no other way is known to differentiate the categories mentioned in Section 6.1 and, as all other distributions behave well, it was decided to accept this small deviation. Further deviations can be seen in the tails of some distributions, but those have large uncertainties and therefore do not cause problems. The distributions for the modified version of MadGraph are produced by combining the distributions from the sample where only top quarks radiate photons ("top-only sample") and the sample where no photon is radiated by a top quark ("no-top sample") according to the respective cross sections. As stated before, interference terms are neglected, so scaling the samples produced with the modified MadGraph version does not reflect the exact relative importance of both samples, but at the moment it is the best way to combine them. The possible systematic uncertainty introduced by this procedure is expected to be small. This is concluded from the following observation: the cross section for the nominal MadGraph sample is 0.1994 pb, for the no-top sample 0.1831 pb, and for the top-only sample 0.02977 pb. Adding the latter two cross sections gives 0.2129 pb. The difference to the nominal MadGraph cross section is 0.0135 pb or approximately 7%. This has to be taken into account later on in a complete analysis, maybe as a systematic uncertainty. Figures 6.1 and 6.2 show some control plots. Further plots for more variables can be found in Appendix B.

6.4. Reconstructed Samples

The samples produced with the modified version of MadGraph were used for several studies on parton level (see Chapter 7) including the training of NNs. The results help in understanding the kinematic behaviour of the different categories of prompt photons, but they are not useful to analyse real data. Therefore it is necessary to understand the simulated events on reconstruction level[5]. Reconstruction level means that physics objects, like electrons or jets, are reconstructed from the electric signals in the detector. For Monte-Carlo samples the detector response is simulated.

To reconstruct physics objects it is necessary to define what their detector response looks like. There are several algorithms that use the detector signals to build so called candi-

---

[5]To get the information, it was requested to get TOPQ1 derivations. Therefore, the parton level information was used to shower the events and simulate the detector response. The output from the detector simulation was then used to produce the derivation. 200,000 events were requested for each sample. In case of the top-only sample 165,000 events could be processed and for the no-top sample 160,000.
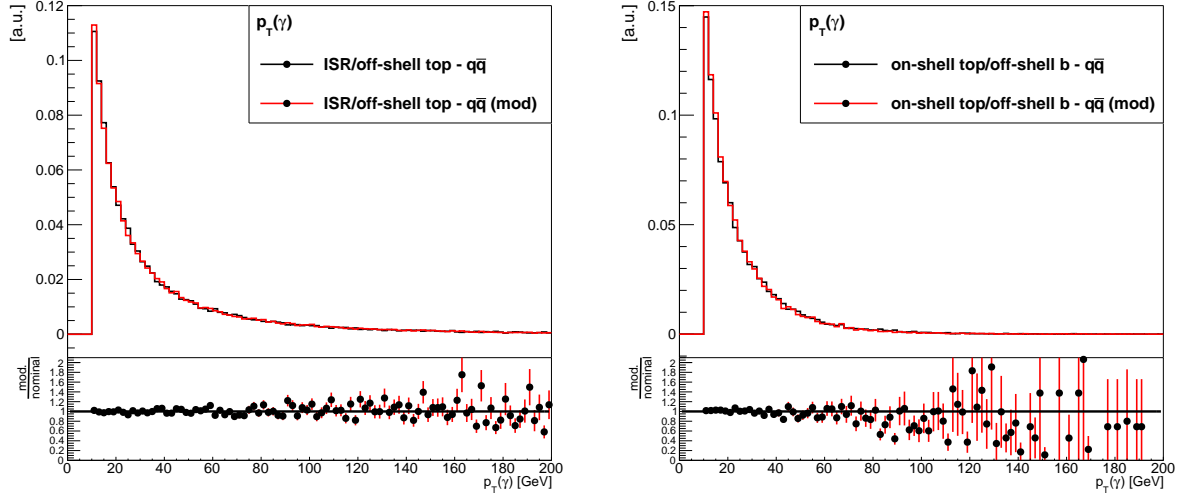
Figure 6.1.: Normalised distributions for the transverse momentum of the photon for nominal (black) and modified (red) MadGraph version. Shown are the categories where the origin of the photon is unclear in the nominal MadGraph sample.
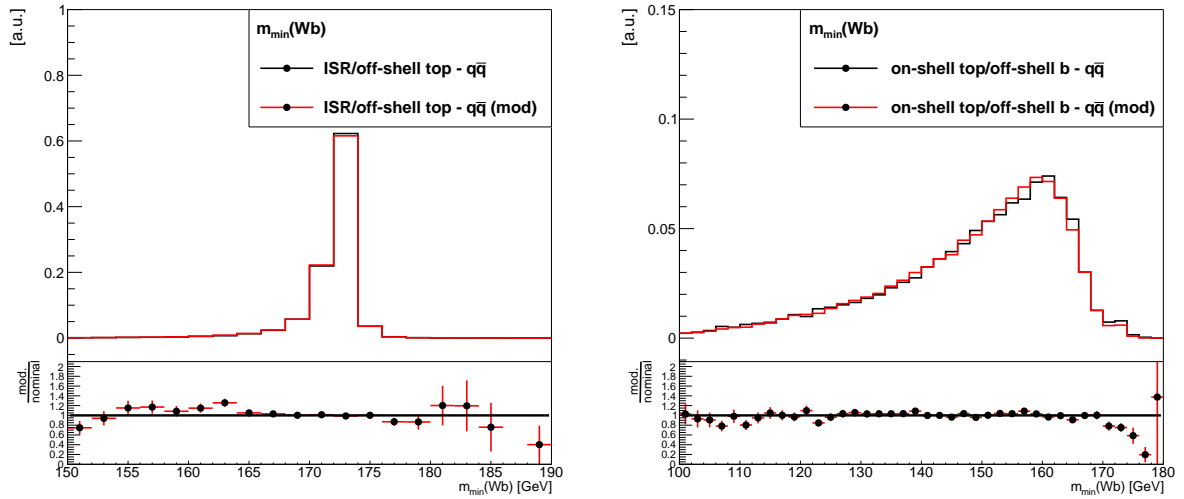
Figure 6.2.: Normalised distributions for the smaller of both invariant mass distributions for the Wb system for nominal (black) and modified (red) MadGraph version. Shown are the categories where the origin of the photon is unclear in the nominal MadGraph sample.
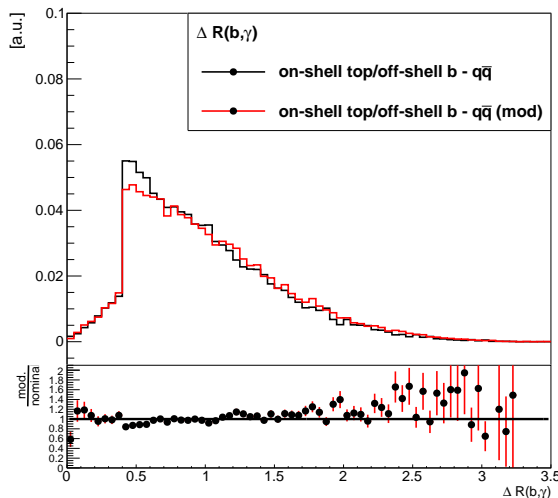
Figure 6.3.: Normalised distributions for ΔR between photon and nearest b quark for nominal (black) and modified (red) MadGraph version. Shown is the category where on-shell top quarks and off-shell b quarks as mother particles of the photon cannot be differentiated in the nominal MadGraph sample. At the peak differences are observed.

dates. For the reconstruction those candidates must pass following object definitions to be used:

- Electrons: Electron candidates have to pass the tightLH identification and Gradient isolation [101].

- Muons: Muons candidates have to pass medium identification and also Gradient isolation [102].

- Jets: Jet candidates are reconstructed with the anti-k_t algorithm using topological energy clusters [103, 104]. The radius parameter is $R = 0.4$ in the $\phi - \eta$ plane. The Jet-Vertex-Tagger was used to suppress jets from pile-up [105].

- b-jets: To tag b-jets the MV2c10 algorithm was used and b-jets were tagged at different working points (77% and 85%) [106, 107].

- Photons: No ID or isolation requirements were set. Every photon candidate was accepted [94].

Beside these object definitions kinematic cuts were also applied. Two different sets of cuts were applied to the samples[6]. The first set was the same as used by the ATLAS $t\bar{t}\gamma$

---

[6]On these datasets, selection cuts were applied using the AnalysisTop sofware framework.

Sample	top-only	no-top
Total event number	165,000	160,000
Events passed selection 1	61,475	42,524
Relative event yields	37.26%	26.58%
Events passed selection 2	32,115	18,501
Relative event yields	19.46%	11.21%

Table 6.1.: Event yields for the top-only sample and the no-top sample.

group. The cuts are the following ones for the lepton+jets channel:

- Exactly one electron or muon with $p_T > 25$ GeV

- At least two jets with $p_T > 25$ GeV

- At least one photon with $p_T > 15$ GeV and $|\eta| \leq 2.37$

In the dilepton channel the cuts were:

- Exactly two charged leptons with $p_T > 25$ GeV and opposite sign

- Invariant mass of both charged leptons above 15 GeV

- At least one photon with $p_T > 15$ GeV and $|\eta| \leq 2.37$

These cuts were used for comparison studies with the official $t\bar{t}\gamma$ samples (see Section 8.1). The second set of cuts, which is for the usage of NNs on reconstruction level (see below), only considers the lepton+jets channel and the jet requirement was changed to:

- At least four jets with $p_T \geq 20$ GeV

These cuts are relatively loose so that statistics can be kept at a level still useful for a NN training. Loose cuts are sufficient here, as the sample contains only signal events in the sense that just Feynman diagrams for the hard $t\bar{t}\gamma$-process were considered. The requirement of at least four jets is motivated to reconstruct the full $t\bar{t}$ kinematics. In Table 6.1 the yields are presented[7]. In the training of the NN on parton level (see Chapter 7) some variables related to the 4-vectors of the top quarks were used. As these variables have a high impact on the quality of the training, they should be kept at reconstruction level. This makes it necessary to perform top quark reconstruction which was carried out with KLFitter [108]. KLFitter is a kinematic likelihood fitter to reconstruct both top quarks in the lepton+jets channel. It tries to find the best permutation of assignments of

---

[7]Cutflow plots are shown in Appendix C.

the jets to final state quarks. As four jets are expected in a lepton+jets channel event, there are 4!=24 permutations. But interchanging the two jets related to the hadronically decaying W boson reduces the permutations to 12, as both permutations would give the same result. The best permutation is defined by maximising the following likelihood function and taking the permutation with the highest likelihood value:

$$
\begin{aligned}
\mathcal{L} = &B\left\{m(\mathrm{q_1 q_2}) \mid m_W, \Gamma_W\right\} \cdot B\left\{m(\ell\nu) \mid m_W, \Gamma_W\right\} \cdot \\
&B\left\{m(\mathrm{q_1 q_2 b_{had}}) \mid m_{top}, \Gamma_{top}\right\} \cdot B\left\{m(\ell\nu b_{lep}) \mid m_{top}, \Gamma_{top}\right\} \cdot \\
&W\left(\tilde{E}_{\mathrm{jet}_1} \mid E_{b_{had}}\right) \cdot W\left(\tilde{E}_{\mathrm{jet}_2} \mid E_{b_{lep}}\right) \cdot W\left(\tilde{E}_{\mathrm{jet}_3} \mid E_{\mathrm{q1}}\right) \cdot W\left(\tilde{E}_{\mathrm{jet}_4} \mid E_{\mathrm{q2}}\right) \cdot \\
&W\left(\tilde{E}_{\mathrm{x}}^{\mathrm{miss}} \mid p_{x,\nu}\right) \cdot W\left(\tilde{E}_{\mathrm{y}}^{\mathrm{miss}} \mid p_{y,\nu}\right) \cdot
\begin{cases}
W\left(\tilde{E}_\ell \mid E_\ell\right) \ , \ e+ \ jets \ \text{Kanal} \\
W\left(\tilde{p}_{\mathrm{T},\ell} \mid p_{\mathrm{T},\ell}\right) \ , \ \mu+ \ jets \ \text{Kanal}
\end{cases}
\end{aligned}
\tag{6.1}
$$

B gives the Breit-Wigner functions of the mass distributions and the top quark mass is fixed at 172.5 GeV and the W boson mass at 80.4 GeV. W are the transfer functions. They describe, on a statistical level, how the measured values of the 4-vector of each particle correspond to the parton level values of the leading order decay. To improve the performance, b-tagging information can be taken into account [108]. In this case, the 77% working point for b-tagging was chosen, which means that 77% of all b-jets are tagged as such. The resulting samples could then be used for the studies presented in Chapter 8. In difference to the samples used at parton level, at reco level only lepton+jets events could be taken into account since top quark reconstruction for the dilepton events would have been too time consuming for the given time scale of this analysis.

7. Parton level studies

7.1. Variable distributions

7.1.1. Variable distributions in the laboratory frame

The goal of this thesis is to classify hard photons according to their respective mother particle (initial state quark, top quark, b quark, W boson and its decay products). The first step is the search for variables with differences in their distributions that can be exploited. Following variables were considered using the two samples described in Chapter 6:

- The transverse momentum p_T of the photon.

- The energy E of the photon.

- The pseudorapidity η of the photon.

- The distribution of $\cos(\theta)$ of the photon.

- ΔR between the photon and the nearest b quark.

- ΔR between the photon and the nearest top quark.

- $|\Delta\eta|$ between the photon and the nearest top quark.

- The minimum invariant mass of both systems of W boson and b quark.

- The maximum invariant mass of both systems of W boson and b quark.

- The minimum invariant mass of both systems of W boson, b quark and photon.

- The maximum invariant mass of both systems of W boson, b quark and photon.

The first four variables were chosen to look for differences in the kinematic behaviour of the photon itself. The next three variables describe kinematic characteristics of the photon with respect to other particles in the process. The use of the invariant mass variables is

motivated by the expectation that they are different if the photon was emitted after the decay of the top quark or before.

The variable distributions can be seen in Figures 7.1 to 7.11. The presented categories are those explained in Section 6.1. Due to the usage of the samples produced with the modified MadGraph version, a clear categorisation without any ambiguities is possible. This means that FSR(b) and ISR are separated from top radiation. FSR(W) events stay untouched. It can be seen (Figures 7.1 and 7.2) that photons from ISR and from the top quark are expected to be harder than those from the radiative decay (FSR(W) and FSR(b)). This behaviour is expected as the particles in the decay of the top quark have less energy than the top quark itself (energy conservation). Therefore the respective phase space to emit hard photons becomes smaller. The b quark also emits softer photons than the W boson. This is due to the higher mass of the W boson which gives it higher energy (relativistic energy-momentum relation). The fact that ISR has the hardest spectrum is expected because the energy of the initial state quarks is higher than those of the produced top quarks, as most top quark pairs are produced in highly asymmetric collisions of two gluons (divergent behaviour of gluon PDFs for low x; see Figure 2.15). Due to the asymmetry of those collisions the centre-of-mass energy is smaller and hence the energy to radiate a photon as well. Another behaviour that is observed in the angular distributions of the photon (Figures 7.3 and 7.4) is that ISR photons tend to be more collinear to the beam pipe, which means they have higher η-values or $\cos(\theta)$ near 1 or -1, respectively. Photons from the other three categories are distributed more narrowly around the central part of the detector, with photons from top quarks as the most central ones.

Both ΔR variables also show different behaviour for the different categories. In case of ΔR between the photon and the nearest b quark (Figure 7.5), there is a clear peak at low values for events where a b quark emitted the photon. Intuitively this matches expectations. In case a top quark or a W boson emitted the photon, the peak is shifted to higher values. ISR photons have the broadest spectrum. This is due to the fact that in this case, photon emission is independent of the kinematics of the decay of the top quark. A similar behaviour can be seen for ΔR between the photon and the nearest top quark (Figure 7.6). Interestingly, if the top quark emits the photon, the distribution for this category is not the one with the peak at lowest values. FSR(b) and FSR(W) have, on average, a smaller ΔR distance between the photon and the nearest top quark. Again, ISR has the broadest spectrum. The same behaviour can be seen for $|\Delta\eta|$ between the photon and the top quark (Figure 7.11). The reason for this is that the photons radiated by off-shell top quarks are not related to the kinematics of the decay of the top quark and therefore behave similarly to ISR photons. A look into the LHE-file of the top-only
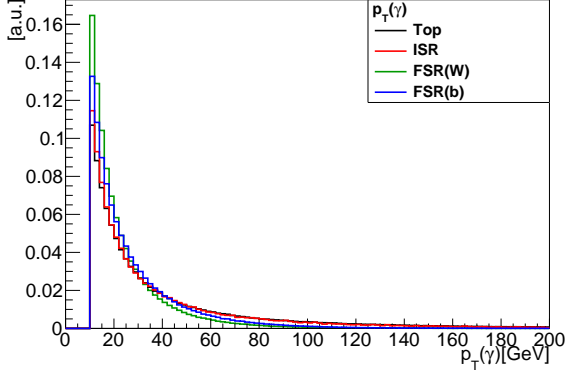
Figure 7.1.: The distributions of photon p_T for the four different categories.
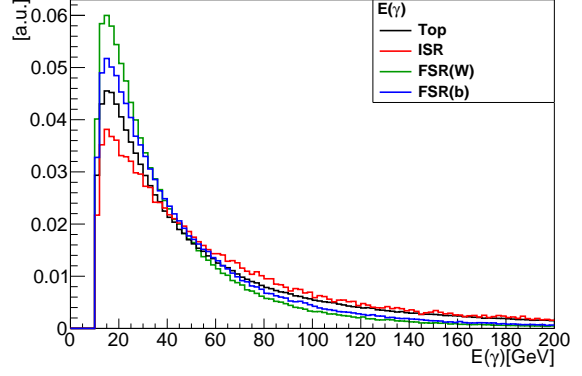
Figure 7.2.: The energy distribution of the photon for the four different categories.
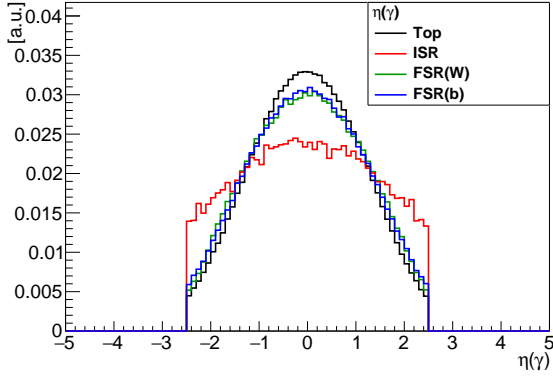
Figure 7.3.: The distributions of photon η for the four different categories.
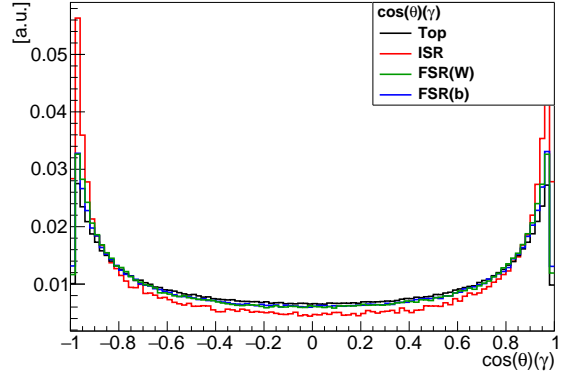
Figure 7.4.: The $\cos(\theta)$ distribution of the photon for the four different categories.
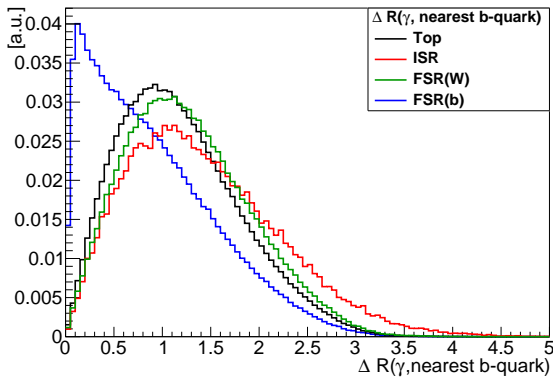
Figure 7.5.: The ΔR distributions between the photon and the nearest b quark for the four different categories.
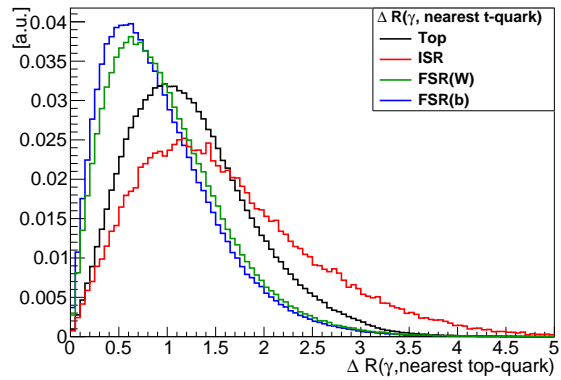
Figure 7.6.: The ΔR distributions between the photon and the nearest top quark for the four different categories.
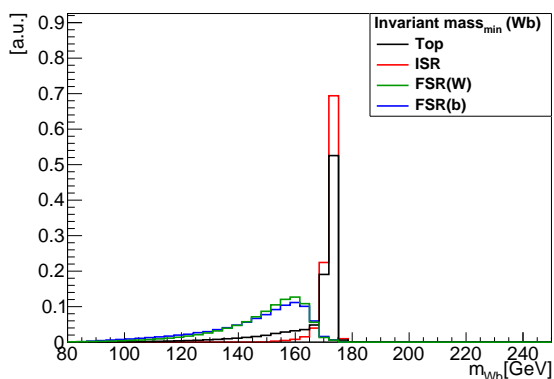
Figure 7.7.: The distributions for the lower invariant mass of both Wb systems for the four different categories.
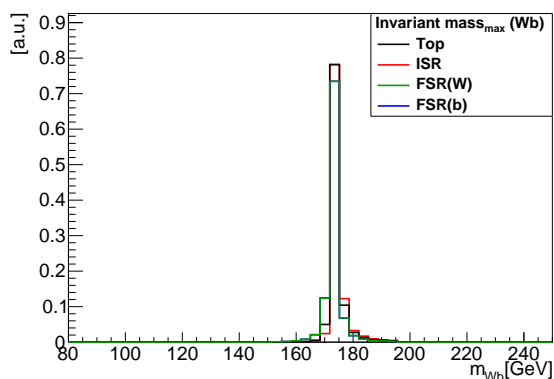
Figure 7.8.: The distributions for the higher invariant mass of both Wb systems for the four different categories.
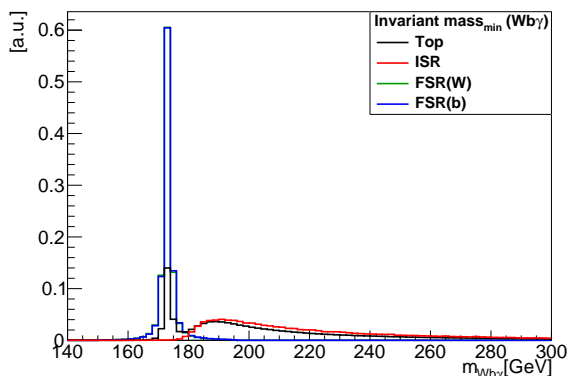
Figure 7.9.: The distributions for the lower invariant mass of both Wbγ systems for the four different categories.
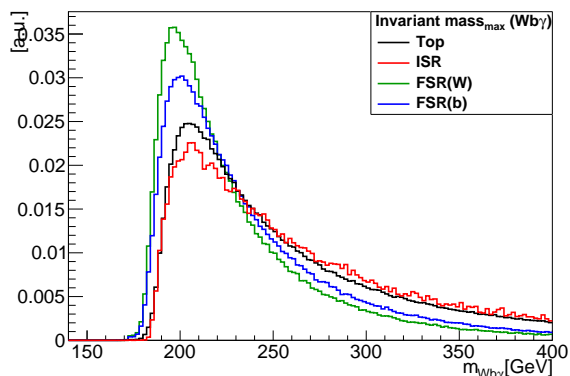
Figure 7.10.: The distributions for the higher invariant mass of both Wbγ systems for the four different categories.
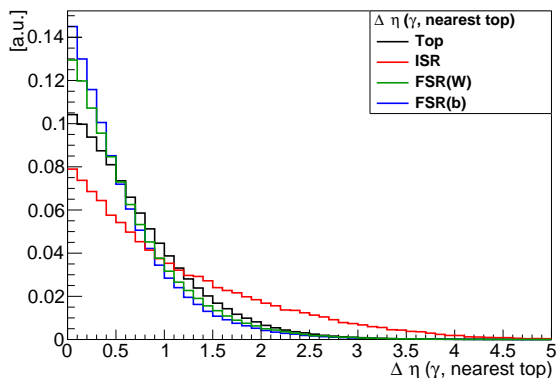
Figure 7.11.: The distribution for $|\Delta\eta|$ between the photon and the nearest top quark for the four different categories.
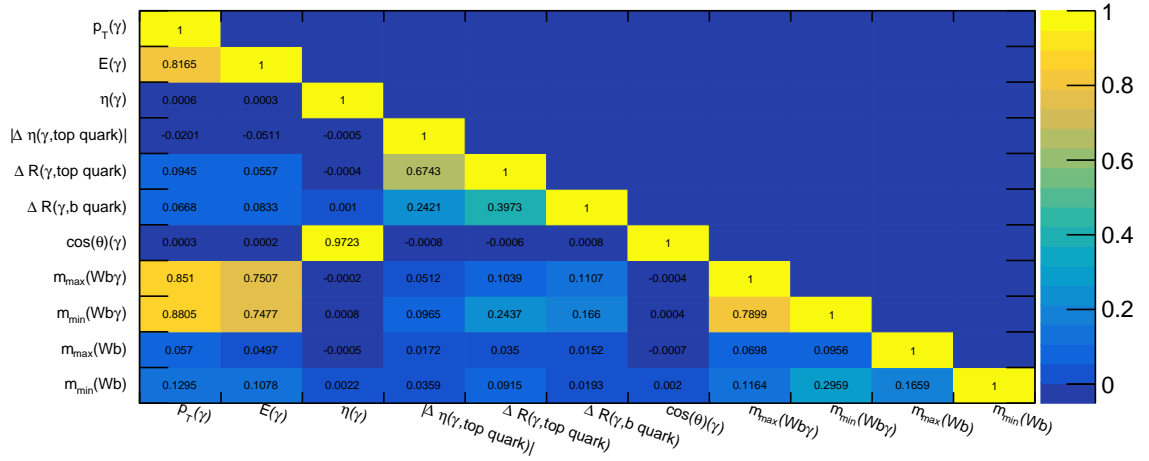
Figure 7.12.: Linear correlation coefficients for the variables in events where the top quark radiated the photon.
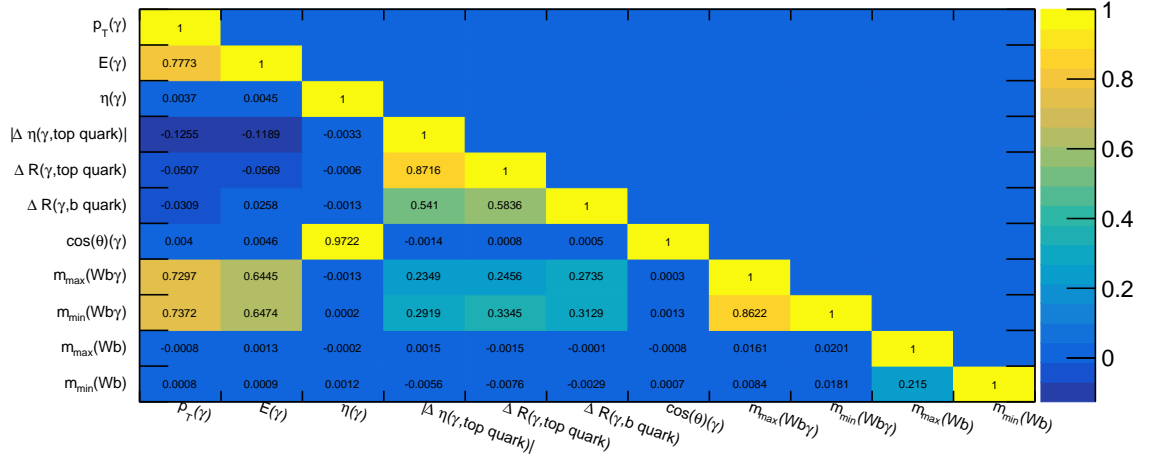
Figure 7.13.: Linear correlation coefficients for the variables in case of ISR photons.
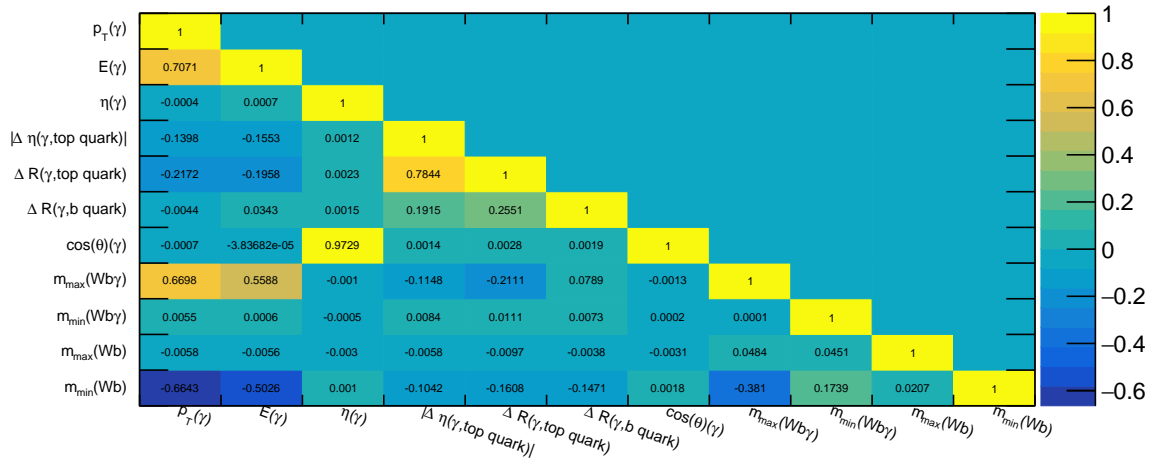
Figure 7.14.: Linear correlation coefficients for the variables in events where the W boson or its decay products (FSR(W)) radiated the photon.
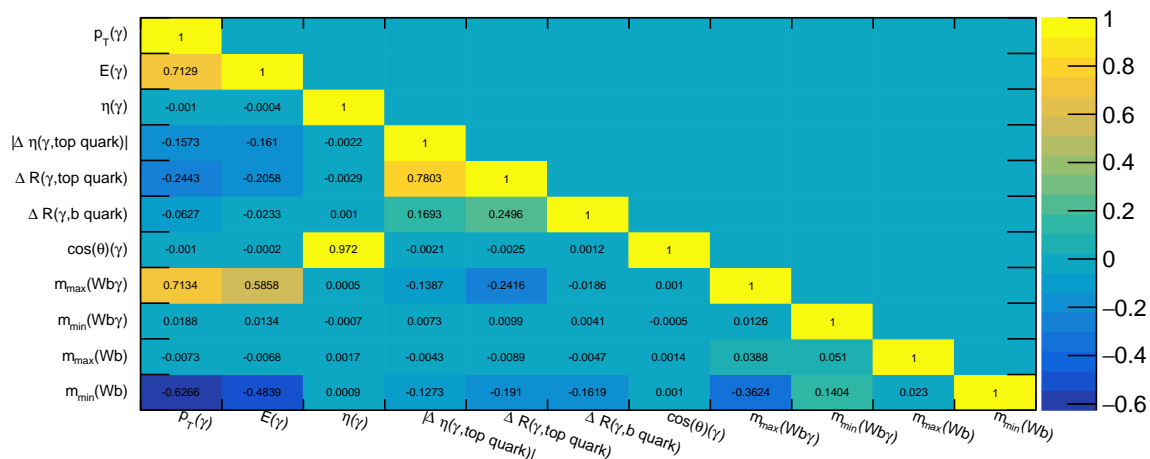
Figure 7.15.: Linear correlation coefficients for the variables in events where the *b* quark (FSR(b)) radiated the photon.

sample shows that most photons coming from top quarks come from off-shell top quarks. The distribution is not as broad as in case of ISR events, due to events with on-shell top quarks emitting a photon. In case of the invariant mass distributions (Figures 7.7 - 7.10), the differences between the categories, in which the events shall be classified, are the biggest. This is related to the fact that if a *W* boson or a *b* quark radiate the photon, its four-vector would be needed to reconstruct the top mass. This is not the case for ISR photons or photons from off-shell top quarks. The slight differences between ISR and top radiation come again from the fact that on-shell top quarks radiate photons and these events exhibit a similar behaviour as photons from FSR(W) and FSR(b).

To assess how useful the different variables are, the linear correlation between them was calculated for each category (see Figures 7.12 - 7.15). Most variables are uncorrelated, which is ideal for later use in NNs. The correlations that are observed are expected a priori. The pattern of correlations looks quite similar for top radiation and ISR, on the one hand, and on the other hand FSR(W) and FSR(b) have similar correlations between the single variables. Again, this shows that top radiation behaves, in most cases, like ISR, which can be related to the fact that most photons coming from top quarks come from off-shell top quarks.

Another feature that is desirable for these events, is high separation power. The more the distributions differ for the respective categories, the easier and more efficient the training of a NN. The separation S is defined as

$$S = \frac{1}{2} \sum_{i=1}^{N} \frac{(s_i - b_i)^2}{s_i + b_i} \ , \tag{7.1}$$

Variables	Top/ISR	Top/FSR(W)	Top/FSR(b)	Top/all
p_t	0.15	8.43	5.29	5.42
E	0.55	5.26	2.96	2.97
η	3.15	0.22	0.19	0.31
$\Delta\eta$	6.77	0.65	1.59	0.84
$\Delta R(b,\gamma)$	3.32	0.48	7.25	2.08
$\Delta R(t,\gamma)$	5.03	4.95	7.70	5.05
$\cos(\theta)$	3.14	0.23	0.19	0.31
$m_{min}(Wb)$	5.13	54.62	54.06	42.37
$m_{max}(Wb)$	0.90	3.02	3.05	2.50
$m_{min}(Wb\gamma)$	11.42	33.36	32.65	47.68
$m_{max}(Wb\gamma)$	0.41	9.60	5.23	5.68

Table 7.1.: The separation power (in percent) of the different variables in the laboratory system. The separation was calculated for top radiation (Top) versus the other categories. The last column gives the separation with respect to all background categories combined according to cross sections.

Variables	Top/ISR	Top/FSR(W)	Top/FSR(b)	Top/all
p_t	0.10	10.17	7.01	6.69
E	3.49	8.72	6.39	5.02
η	12.67	0.47	0.22	0.74
$\Delta\eta$	0.29	3.49	4.43	0.29
$\Delta R(b,\gamma)$	2.70	1.12	6.45	2.00
$\Delta R(t,\gamma)$	0.48	6.98	9.30	6.89
$\cos(\theta)$	12.71	0.47	0.22	0.73

Table 7.2.: The separation power of the different variables in the rest frame of the $t\bar{t}$ system. The separation was calculated for top radiation versus the other categories. The last column gives the separation with respect to all background categories combined according to cross sections.

where s_i means the normalised signal contribution in bin i and b_i the normalised background contribution in bin i. The separation for each variable is listed in Table 7.1. The invariant masses give by far the best separation power.

7.1.2. Variable distributions in the $t\bar{t}$ rest frame

All variables, but the invariant masses, change their shape if one boosts into the rest frame of the $t\bar{t}$ system. That means both top quarks are in a back-to-back configuration. Such a transformation can help to increase the efficiency of NNs because they do not need to learn it. The corresponding plots can be seen in Figures 7.16 to 7.22. The changed separation values are listed in Table 7.2. Relevant changes with respect to the values
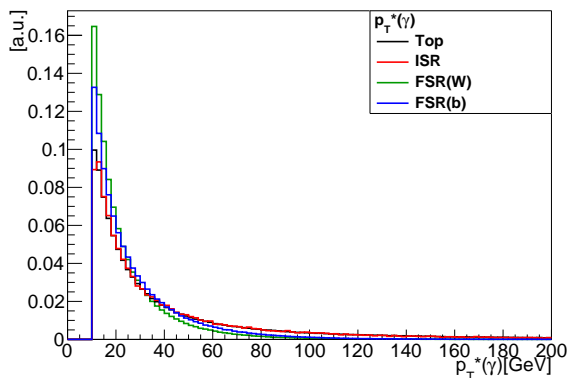
Figure 7.16.: The distributions of photon p_T for the four different categories.
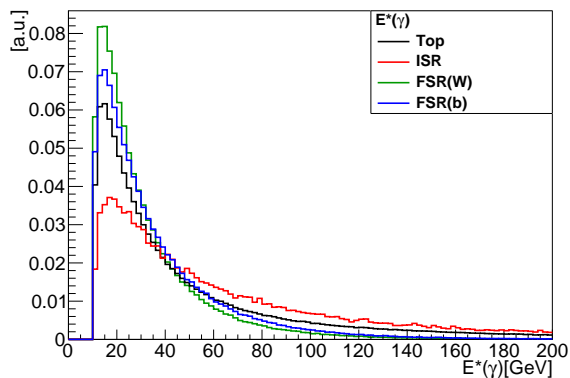
Figure 7.17.: The energy distribution of the photon for the four different categories.
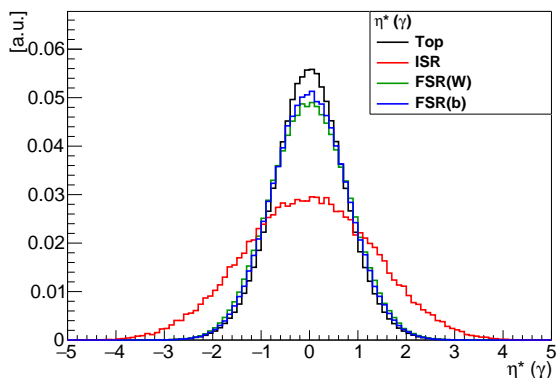
Figure 7.18.: The distributions of photon η for the four different categories.
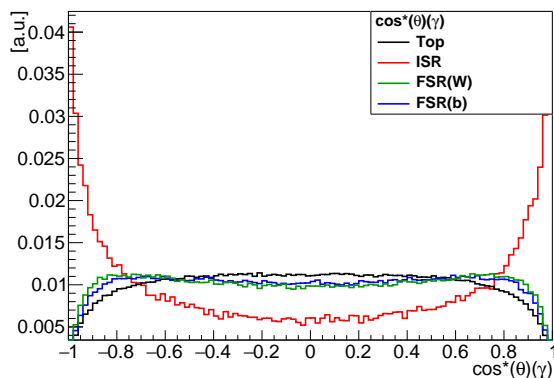
Figure 7.19.: The $\cos(\theta)$ distribution of the photon for the four different categories.
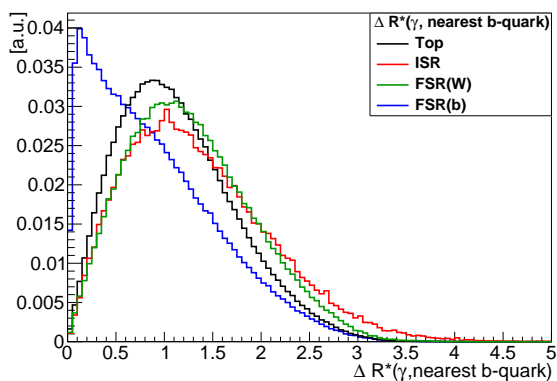
Figure 7.20.: The ΔR distributions between the photon and the nearest b quark for the four different categories.
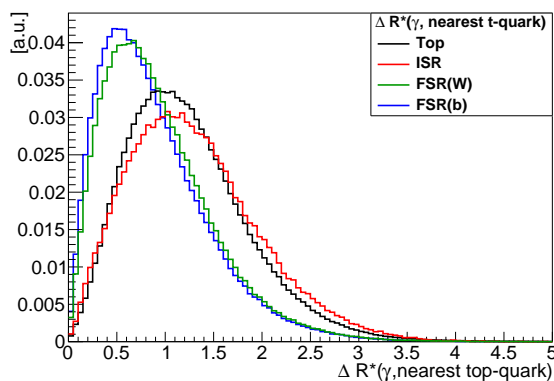
Figure 7.21.: The ΔR distributions between the photon and the nearest top quark for the four different categories.
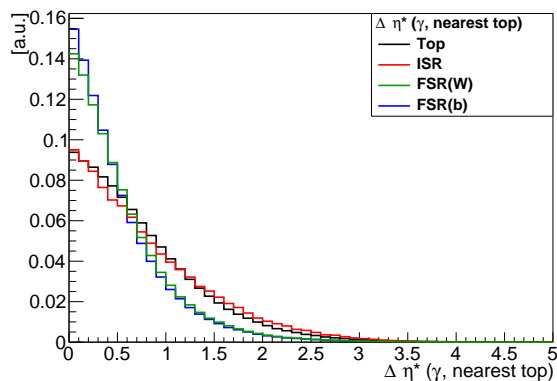
Figure 7.22.: The distribution for $|\Delta\eta|$ between the photon and the nearest top quark for the four different categories.

presented in Table 7.1 are found for several variables. For the transverse momentum, separation between top radiation and FSR(b)/FSR(W) becomes better. For the energy, this also happens for the separation between top radiation and ISR. For η and $\cos(\theta)$ a strong improvement in the separation is observed. The trade off for the improvement is that the separation values for top radiation versus ISR in the remaining three variables become smaller. For the radiative decay, the values in some cases get better and in others slightly worse. In general these separation values are a signal for a possible better performance of the NN using the rest frame instead of the lab frame. The correlation plots show some changes. They can be found in Figures 7.23 to 7.26. But those changes are, in most cases, quite small. Some relevant changes can be observed for correlations between E and p_T.

7.1.3. Conclusion

The studies of the variables, related to properties of the production and decay of top quark pairs in association with a photon, have shown differences between the relevant categories. That should allow the usage of multivariate analyses techniques (MVAs) to enhance the purity in a $t\bar{t}\gamma$ sample.

7.2. Training neural networks

The separations presented in Section 7.1 give good reasons to try multivariate anlyses techniques (MVA). From the pool of different possible methods, it was decided to use NNs. They are more efficient than other MVAs, but this comes with a disadvantage. The neural network output is difficult to interpret, as the learning process is very complex.
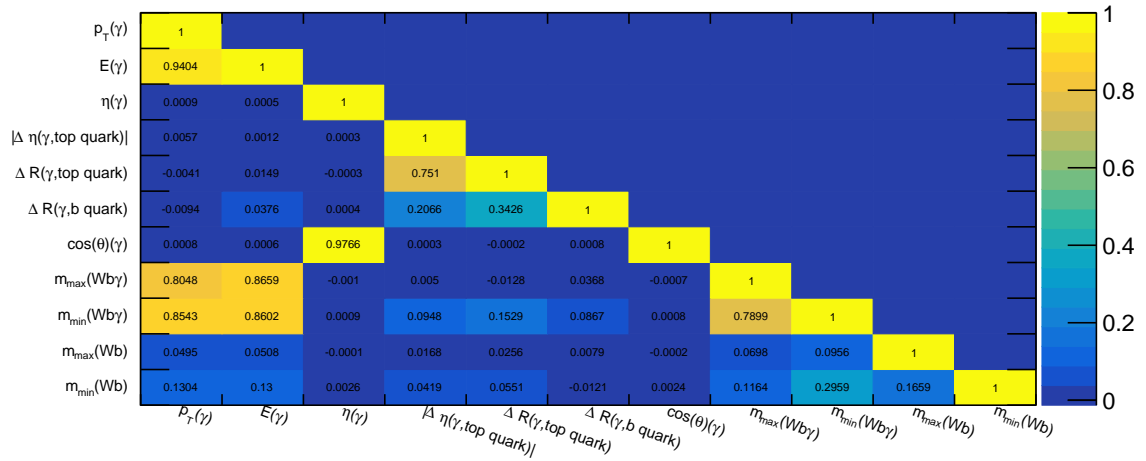
Figure 7.23.: Linear correlation coefficients for the variables in events where the top quark radiated the photon.
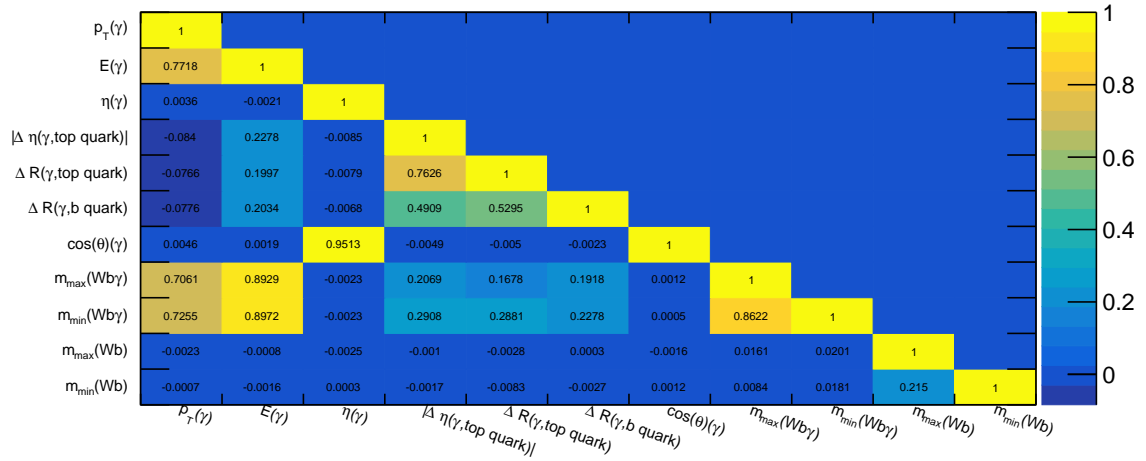
Figure 7.24.: Linear correlation coefficients for the variables in case of ISR photons.
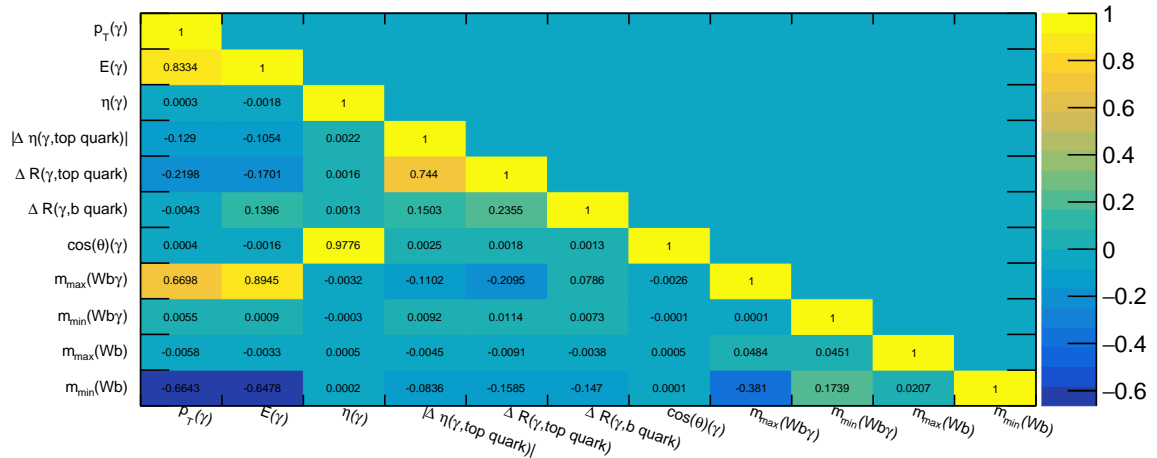
Figure 7.25.: Linear correlation coefficients for the variables in events where the W boson or its decay products (FSR(W)) radiated the photon.
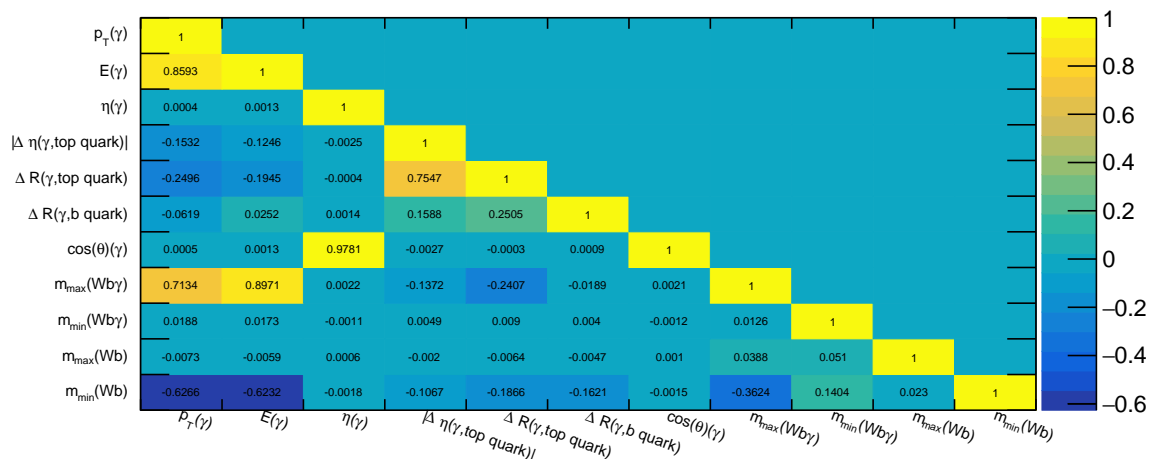
Figure 7.26.: Linear correlation coefficients for the variables in events where the b quark (FSR(b)) radiated the photon.

Nevertheless, the expected performance was the reason to favour them over other MVA methods.

A lot of different software libraries can be used to construct NNs. The ATLAS analysis of the $t\bar{t}\gamma$-process at 13 TeV uses a tool called Prompt Photon Tagger (PPT) which was developed by Benedikt Völkel [93]. The framework of this tool was modified to fit the purposes of this analysis. Hence, the library Keras [109] is used with Theano [110] as its backend. As shown in Chapter 7, there are differences between the background categories. It was considered to account for this by doing multiclass classification. In this case, four categories are differentiated. It means that the NN needs four output nodes. Each output node is a classifier for one category. For the last layer, with its four nodes, the softmax function was chosen as the activation function. This allows the interpretation of the output classifier as a posterior probability.

Both samples used for training (top-only and no-top samples) contain nearly one million usable events[1] each, which does not reflect the cross sections[2]. The total number of events in each category can be found in Table 7.3. To train and test NN architectures, all 1903616 events were split into two random samples, one with 80% for the training and 20% for the testing. A priori there is no way to say which NN architecture will give the best result. Therefore, different architectures with different numbers of nodes and layers were tested. The whole list of models can be seen in Appendix D. At first, 16 models were tested, using

---

[1] Some events do not contain both top quarks and or both W bosons. This is due to a generator cut on the Breit-Wigner function of both particles.

[2] Top radiation events are over represented which will bias the NN towards a good classification efficiency for top radiation.

Category	Number of Events
Top	958832
ISR	75618
FSR(b)	477031
FSR(W)	392135

Table 7.3.: Number of events for the different categories used in training and testing NN architectures.
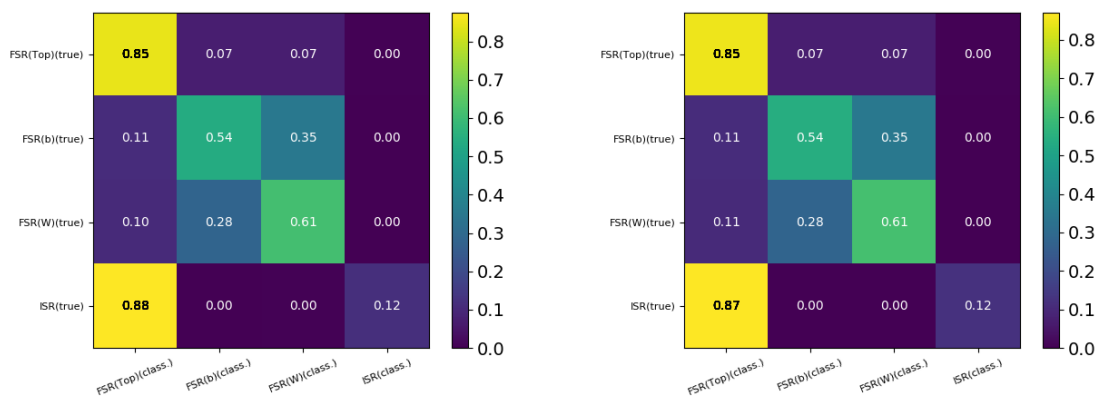
Figure 7.27.: Confusion plots for training (left) and testing (right) data of a NN with 3 layers (16/8/4 nodes)

architectures from three to six layers. For each number of layers, two different models were tested in both reference frames. To quantify the NN performance, it was decided to use the following approach: as each event has four output values assigned, where each output value refers to the posterior probability for one of the four categories, the highest output value was chosen to build a confusion matrix. The vertical axis of such a confusion matrix gives the true category of an event and the horizontal axis gives the category with the highest of all four output values for a given event. Numbers give the relative amount of events from one category labeled as a certain category. Therefore the numbers in each row add up to 1 (small differences due to rounding). This holds for all confusion plots that will be shown in this thesis. In Figure 7.27, one can see the confusion plots for testing and training a NN with three layers. The first layer had 16 output nodes, the second eight and the last one four. All eleven variables presented in Section 7.1 were used with their $t\bar{t}$ rest frame values. The training was done for 500 epochs and a batch size of 1000. The optimal confusion plot would have entries of 1 on the diagonal and 0 elsewhere. It can be observed that events from top radiation (here FSR(Top)) are classified quite well with over 80% of events classified correctly. For events that belong to the radiative decay (FSR(b) and FSR(W)) the performance is worse. Those two categories are very similar,
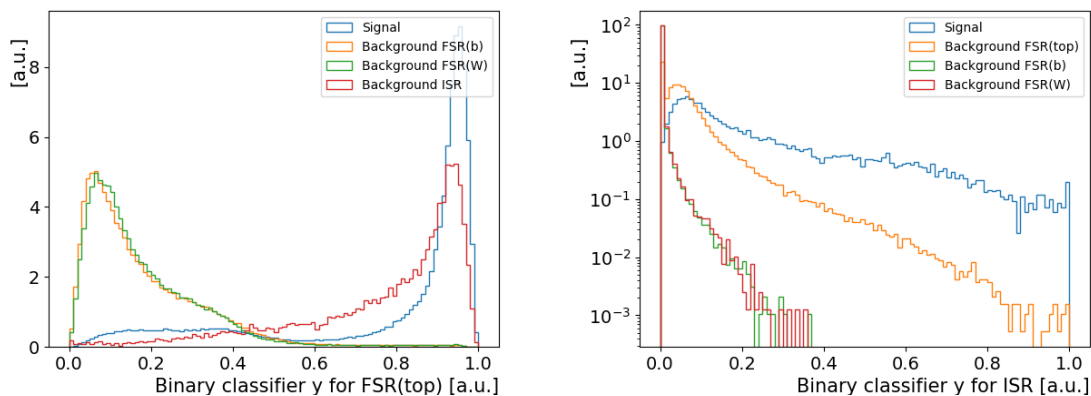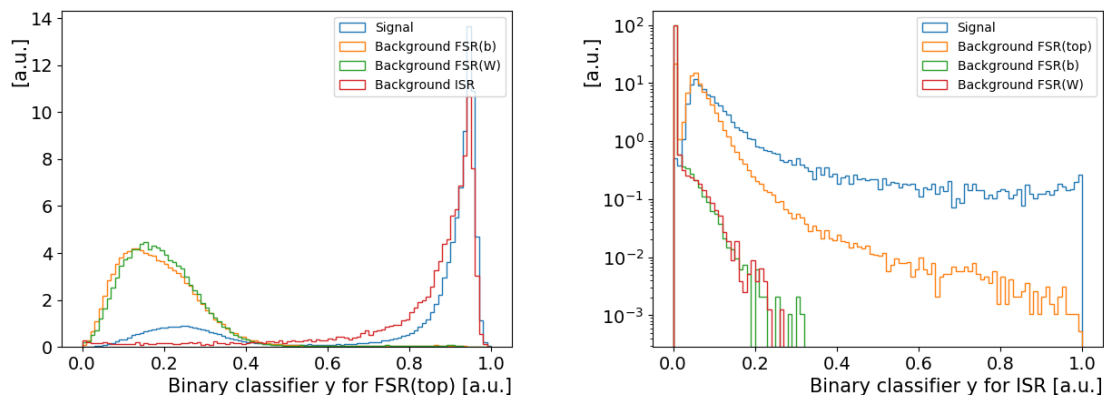
Figure 7.28.: The distributions for the classifier for top radiation (left) and ISR (right) for the different categories. The NN had 3 layers with 16, 8 and 4 nodes respectively and was trained with samples in the $t\bar{t}$ rest frame.
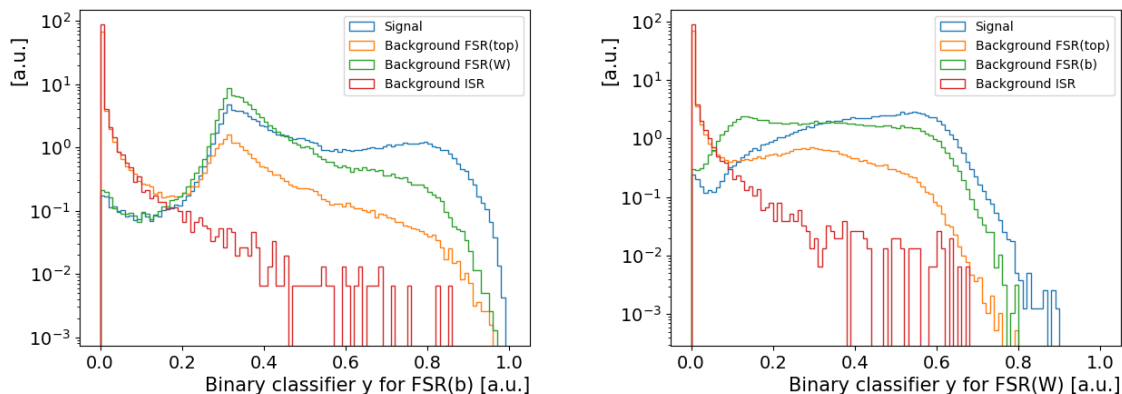
Figure 7.29.: The distributions for the classifier for FSR(b) (left) and FSR(W) (right) for the different categories. The NN had 3 layers with 16, 8 and 4 nodes respectively and was trained with samples in the $t\bar{t}$ rest frame.

so that a lot of events from one category are classified to be from the other. This is not a serious problem, as both categories are well separated from the signal. In contrast, the classification of ISR events exhibits low performance. Most ISR events are classified as top radiation events[3].

In Figures 7.28 and 7.29, one can see the distributions for the four different classifiers for the results of the testing procedure. The results in the training look similar which is proven with the ROC-curves (Figures 7.30 to 7.33). This also shows that no overtraining is present[4]. In the plots, showing the distributions for the classifiers, one can see again

---

[3]The reasons will be presented in Section 7.3
[4]More results of NN will be presented later. Their ROC-curves show no indication for overtraining as well and therefore will not be presented.
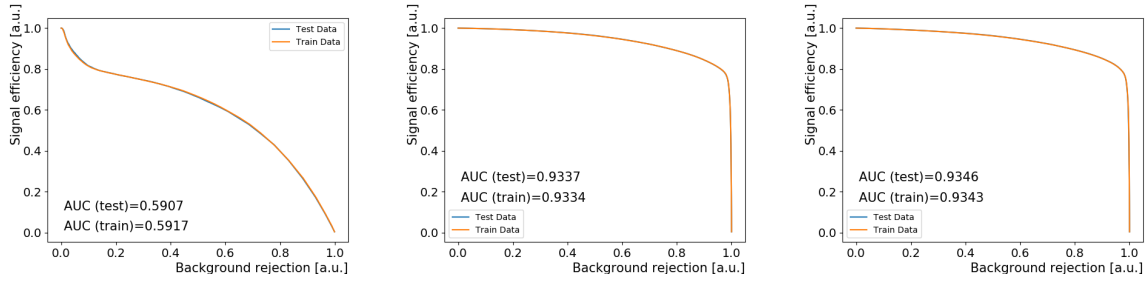
Figure 7.30.: The ROC-curves for the classifier for top radiation. On the left with respect to ISR, in the middle with respect to FSR(b) and on the right with respect to FSR(W).
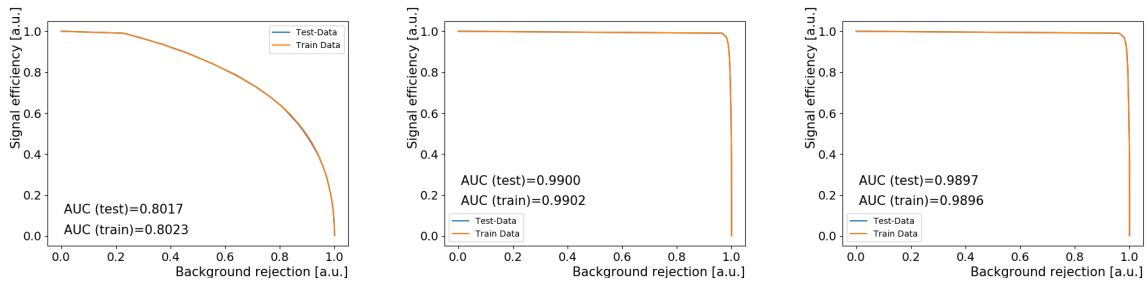
Figure 7.31.: The ROC-curves for the classifier for ISR. On the left with respect to top radiation, in the middle with respect to FSR(b) and on the right with respect to FSR(W).
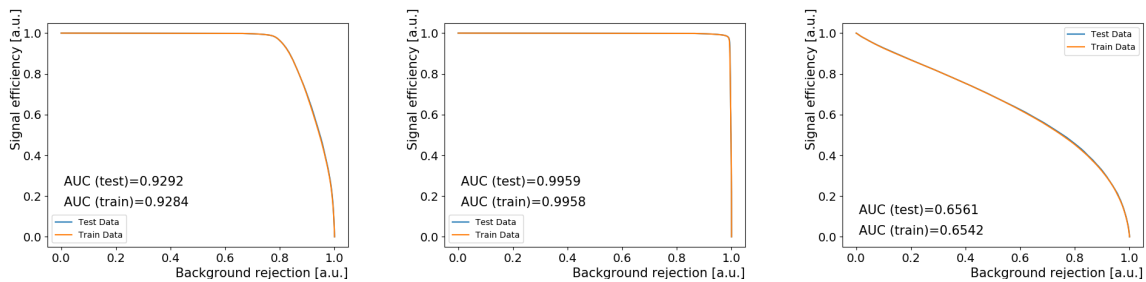
Figure 7.32.: The ROC-curves for the classifier for FSR(b). On the left with respect to top radiation, in the middle with respect to ISR and on the right with respect to FSR(W).
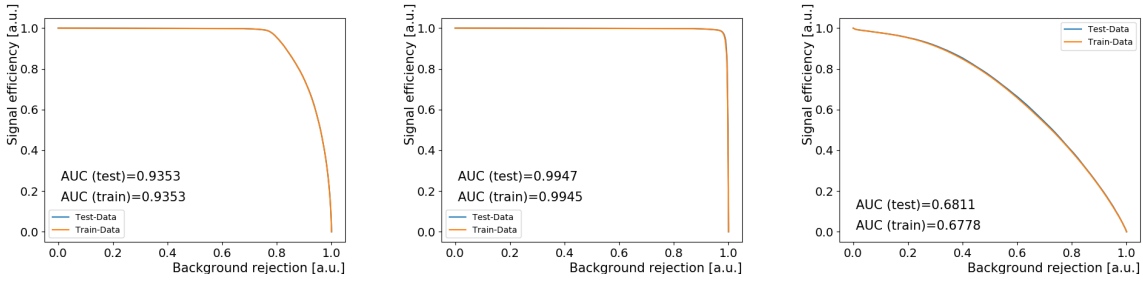
Figure 7.33.: The ROC-curves for the classifier for FSR(W). On the left with respect to top radiation, in the middle with respect to ISR and on the right with respect to FSR(b).
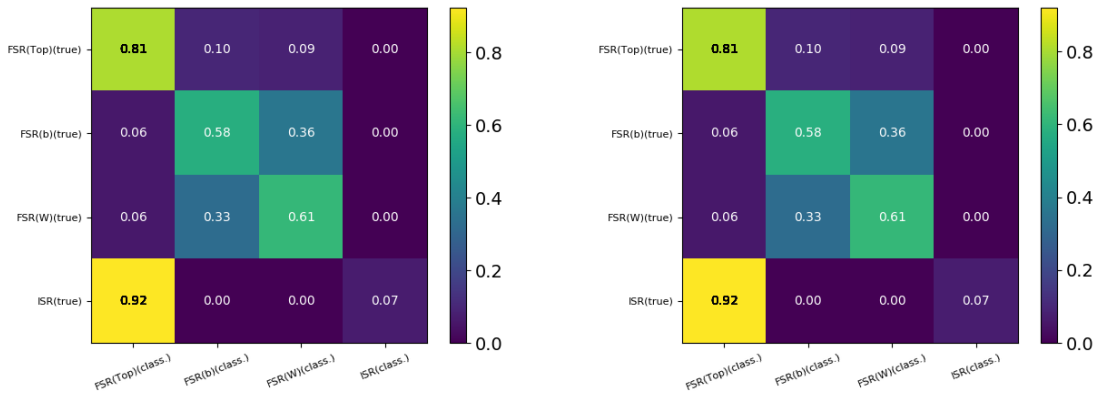
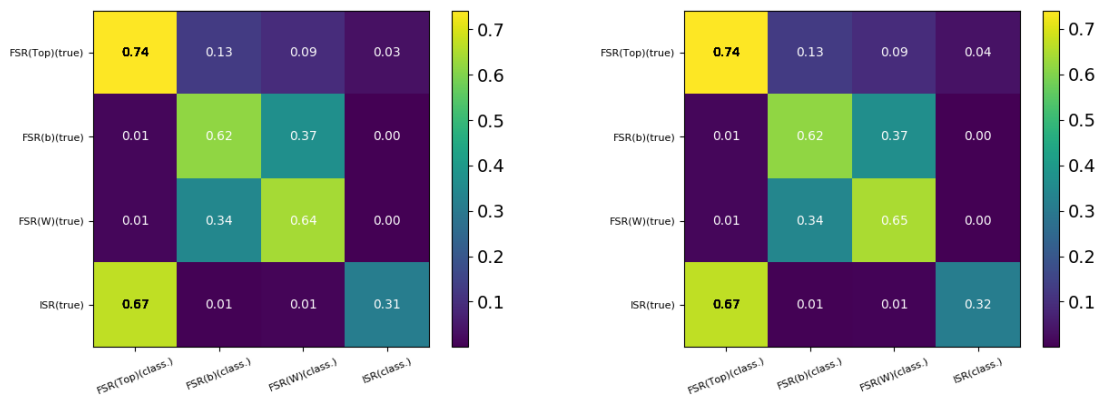***Figure 7.34.:*** Confusion plots for training (left) and testing (right) data of a NN with 3 layers (16/8/4 nodes).

that FSR(b) and FSR(W) events behave similarly, and ISR and top radiation are also similar. But there is one additional fact only slightly visible in the confusion plots. A small fraction of top radiation events behave similarly to FSR(b) and FSR(W). This is also seen in Figure 7.28. Those events are the ones where an on-shell top quark radiated the photon. This conclusion can be drawn based upon the fact that on-shell top quarks radiating photons appear a lot less in the top-only sample, and that in those events the photon, the *W* boson and the *b* quark are needed to get the top quark mass, whereas for off-shell decays the *W* boson and *b* quark are sufficient. As the invariant mass observables are the ones that have the highest separation power, they will dominate the training. The results for the other NNs (see Appendix D) are basically the same, choosing the same inertial frame. The results for training the NNs with laboratory frame variables is worse. In the confusion plots for the same NN, but using the laboratory frame variables (Figure 7.34), one can see that correctly classified top radiation events decreased by 5 percentage points, as well as for ISR events. This shows that the Lorentz transformation was not

Figure 7.35.: The distributions for the classifier for top radiation (left) and ISR (right) for the different categories. The NN had 3 layers with 16, 8 and 4 nodes respectively and was trained with samples in the laboratory frame.
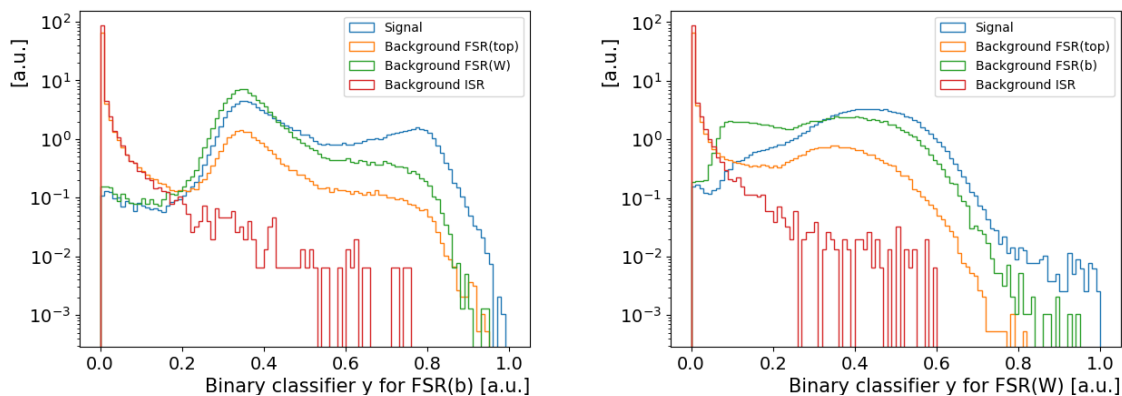
Figure 7.36.: The distributions for the classifier for FSR(b) for the different categories. The NN had 3 layers with 16, 8 and 4 nodes respectively and was trained with samples in the laboratory frame.

learned by the NN during the training process and that defining the variables in the $t\bar{t}$ rest frame increases the performance of the NN.

The distributions of the four classifiers in the testing procedure are presented in Figures 7.35 and 7.36. There are nearly no differences to the distributions for the training procedure. The systematic training of the first eight NN architectures in Table D.1 did not show any hint how to improve the performance significantly. To choose completely different numbers of nodes in each layer did not help as well[5]. The same outcome was observed by adding some options like batch normalization or by changing activation functions. As stated before, the number of events in each category do not reflect the cross section. Due to the high number of top radiation events and a small number of ISR events, the training

---

[5]The tested architectures were 3l_60_100_4 and 6l_60_100_150_100_50_4 (see Table D.1).

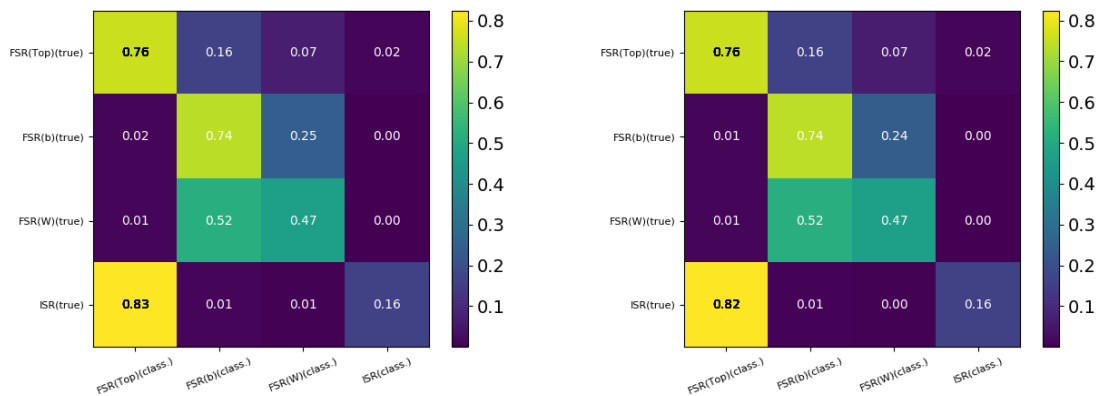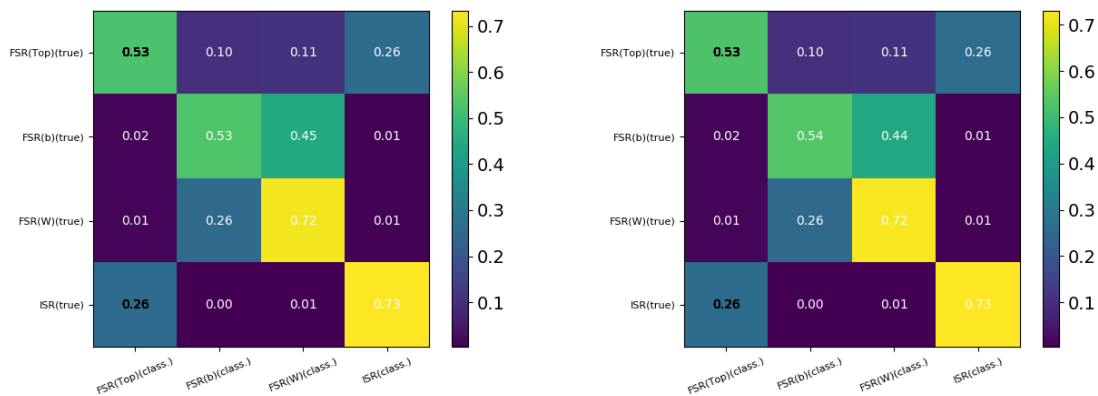**Figure 7.37.:** Confusion plots for training (left) and testing (right) data of a NN with 3 layers (16/8/4 nodes).

is biased in such a way that high values for ISR events in the top-classifier do not allow the loss function increase considerably. This problem can be corrected with weights that can be given to certain events. As all networks so far showed similar behaviour,it was decided to test weights with the 3l_16_8_4 architecture (meaning 3 layers with 16, 8 and 4 nodes respectively), because it is fast to train. Two different ways of weighting were applied. One method reflects the cross section and therefore the expected relative contribution of the four categories to all events. The other method was to put the weights so that multiplying the number of events with the weight gives the number of events in the top radiation category. The latter case is justified by the idea not to bias the decision of the network by the influence of the relative contribution that each category has in the training and test sample. For the first case the confusion plots are shown in Figures 7.37 ($t\bar{t}$ rest frame) and 7.38 (lab frame), for the latter in Figures 7.39 ($t\bar{t}$ rest frame) and 7.40 (lab frame).

From these results one can say that weights affect the classification drastically, as expected. The weights influence the contribution of each event to the loss function. So if an event is assigned a high weight, the loss function will be minimal if the event is classified correctly. In principal, one could give weights to the events so that in the training one category would always be classified correctly. To do that, the weight just needs to be high enough. So applying weights needs a physical justification, as given above. In both cases where weights were applied the classification of ISR events gets better, but this comes with a decreasing performance for top radiation events. A comparison between the three possibilities gives the following picture: the fraction of correctly classified events stays at a relatively high level of approximately 75%, in comparison to up to 85% in the unweighted case. The misclassification of ISR events improves from over 90 % to approximately 82%

Figure 7.38.: Confusion plots for training (left) and testing (right) data of a NN with 3 layers (16/8/4 nodes).

Figure 7.39.: Confusion plots for training (left) and testing (right) data of a NN with 3 layers (16/8/4 nodes).
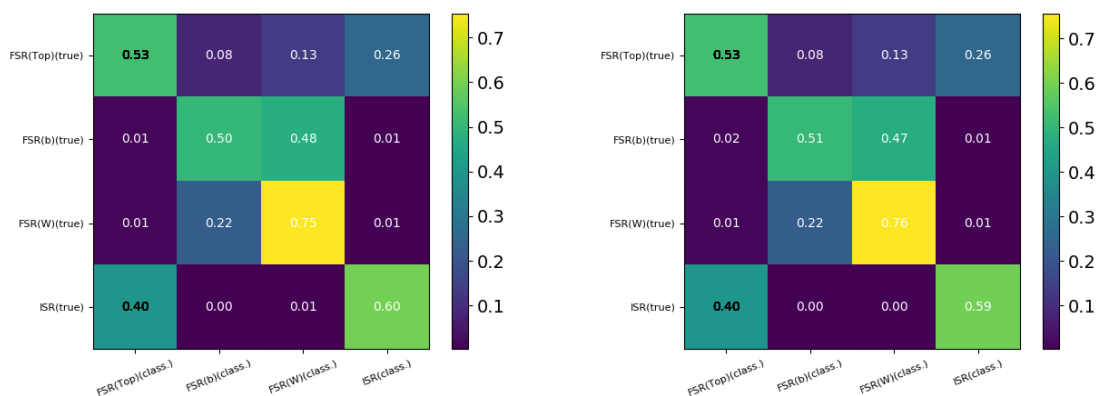
Figure 7.40.: Confusion plots for training (left) and testing (right) data of a NN with 3 layers (16/8/4 nodes).

when the laboratory frame is chosen. In case of the $t\bar{t}$ rest frame, the improvement is even better. The misclassification rate lowers from 87% to 67%. For the second weighting option, the ISR misclassification goes down to 40% in the lab frame and even 26% in the $t\bar{t}$ rest frame, but the fraction of correctly classified top radiation events decreases to nearly 50%. This increases the signal to background ratio, which is roughly 4:1 if all top radiation and ISR events are taken[6]. But the improvement would come with a high loss in statistics. If such a big loss of signal is justifiable, needs dedicated studies that could not be carried out here.

In general, one can say that the contribution of events to the $t\bar{t}\gamma$ signal, where the photon was emitted from decay products, can be clearly distinguished from most of the signal events (top radiation) by using a NN. However, the ISR contribution is difficult to distinguish. Some reasons for this will be investigated in the next section.

Which neural network gives the best performance must be studied finally on reconstruction level. The studies undertaken here were done to show that separating the defined classes is in principle possible and to understand the kinematics of the processes.

7.3. Improving ISR separation

It was shown that top radiation from off-shell top quarks is difficult to separate from ISR. The goal is to get a pure sample of events, where the top quark is the mother particle of the photon. Thus one also needs to separate these two processes. Several studies were done to understand the reasons for the difficult separation, as there are enough differences in the respective distributions to expect a better separation.

7.3.1. Cut on ISR classifier

Figures 7.28 and 7.35 show a clear difference between the distributions of ISR events and top radiation events for the ISR classifier. This behaviour was not only seen for the NNs related to these figures but for all tested networks, regardless of the input ($t\bar{t}$ rest frame or laboratory frame) and the parameters. The differences allow a cut to be placed on the NN output. The idea is to classify all events that have a value for the ISR classifier above a certain threshold as ISR. For all other events, the previous classification method (choosing the highest value) is used. This means that training and ROC-curves are not affected. Only the confusion plots will change. A priori it was clear that this method costs signal efficiency. The goal was to know how big the loss is. The cuts were tested

---

[6]The cross section of the top-only sample is 0.41 pb and the cross section of the no-top sample is 1.35 pb and the relative ISR contribution is about 7.5%. So the ratio is $0.41/(1.35 \cdot 0.075) \approx 4$.

7. Parton level studies
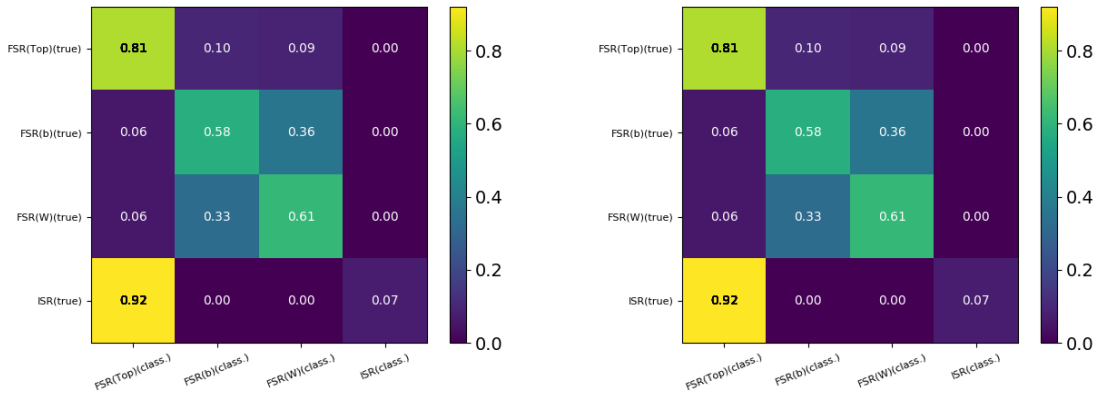
Figure 7.41.: Confusion plot for the testing procedure for the NN 3l_16_8_4 using a cut on the ISR classifier of 0.8 (left) and 0.6 (right).
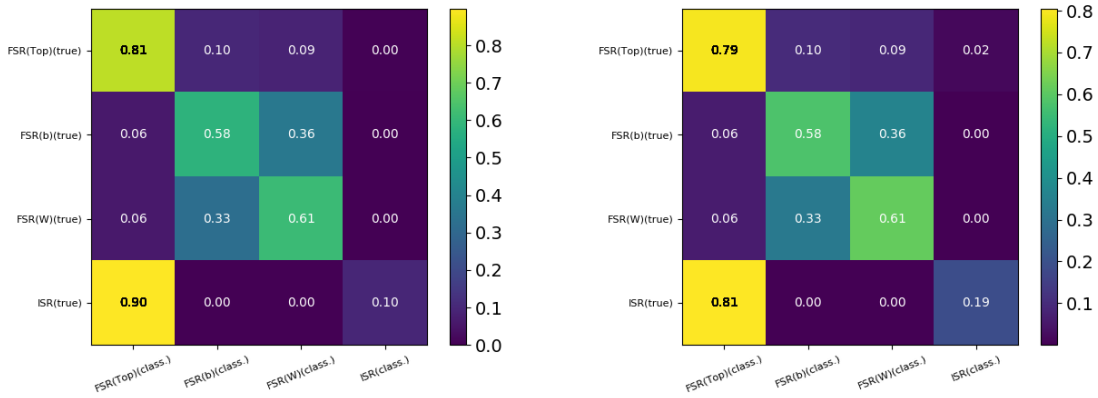
Figure 7.42.: Confusion plot for the testing procedure for the NN 3l_16_8_4 using a cut on the ISR classifier of 0.4 (left) and 0.2 (right).

systematically for the following values: 0.2, 0.4, 0.6, 0.8. To apply cuts, the network with three layers (16, 8 and 4 nodes) and for all sets of weights (none, xsec, amount) was used. Observable distributions related to the laboratory frame were used as input. Networks with more layers were not tested because their behaviour was very similar to the three layer network. The advantage of the three layer network is the relatively short training time. This was the reason to test the ISR cut for this NN. Despite the fact that the results were better in the $t\bar{t}$ rest frame, the laboratory rest frame was chosen because at reconstruction level it is more likely to use data in the laboratory frame. Applying a Lorentz boost to the data would introduce uncertainties, which might be higher than the gain in efficiency. In Figures 7.41 and 7.42, the confusion plots for the different cuts are presented where no weights were applied. In comparison to the right plot in Figure 7.34, where no cut was applied, it can be seen that a good improvement was made only for a cut

value of 0.2. The signal efficiency was reduced by two percentage points, whereas the ISR misclassification rate was reduced by eleven percentage points. This shows that the lower the cut value the better the separation, although at some point it will get worse again. Although 0.2 is not the optimal cut value, the results show that applying such a cut makes the results only slightly better, as still more than 80% of ISR events are wrongly classified. For the cases where weights were applied, the conclusions change. This is due to the point in the distributions of the ISR classifier where the lines of top radiation events and ISR events cross each other. This crossing point changes the optimal cut value. So in case of the weights related to the cross section, 0.4 as a cut value gives better results than 0.2. In the first case, the efficiency for top radiation events as well as ISR misclassification are 73%. The values for the latter case are 38% and 28% respectively. In the case without any weights (see right plot in Figure 7.34) the values are 81% and 92%. In case where the weights were set to account for the total number of events in each category, the results at a cut of 0.8 and 0.6 are basically the same as the ones for no cut (see Figure 7.40). At a cut value of 0.4, the results reduce to 16% and 9%, for 0.2 even to 1% and below 1%. So an optimal cut value must be somewhere between 0.4 and 0.6. In general one can say that this method can be used to improve results, but it does not allow to distinguish between ISR and top radiation as well as between the two categories and the radiative decay categories (FSR(b) and FSR(W)).

7.3.2. Combining reference frames

Looking at the different variable distributions shown in Section 7.1.1 and 7.1.2 one can observe that some of them have better separation for ISR versus top radiation in the laboratory frame and others in the $t\bar{t}$ rest frame. The idea is to use those variables that yield higher separation and combine them to use the advantages of both reference frames. For that, the following variables were chosen from the laboratory frame:

- $|\Delta\eta|$ between photon and nearest top quark.

- ΔR between photon and nearest top quark.

- ΔR between photon and nearest b quark.

All other variables were chosen from the $t\bar{t}$ rest frame[7]. From the tested networks (see Table D.2) the best network was the one with six layers and 64, 128, 32, 16, 8 and 4 nodes, respectively. The confusion matrix is shown in Figure 7.43. In comparison with Figure

---

[7]For the invariant mass distributions the question for the reference frame is irrelevant.
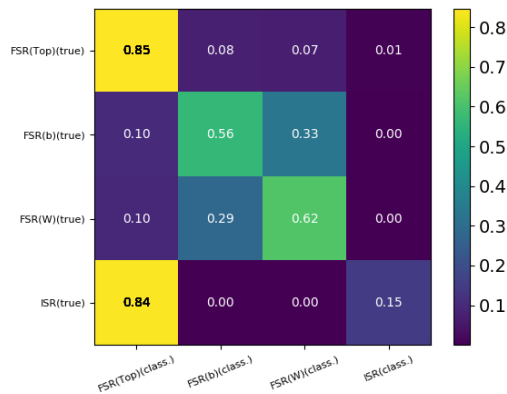
Figure 7.43.: Confusion plot for the testing procedure for the NN 6l_64_128_32_16_8_4 using variable distributions from both reference frames, laboratory frame and $t\bar{t}$ rest frame.

7.27 the efficiency of correctly classified top radiation events stayed constant, whereas a slight improvement for the ISR events was made. The rate of ISR events misclassified as top radiation therefore went down by three percentage points.

7.3.3. Adding variables to the training process

So far eleven different variables were considered that are related to different aspects of the $t\bar{t}\gamma$ process. They allow good separation between top radiation and events where the photon was emitted in the top quark decay. Therefore variables are needed that allow better separation between ISR events and top radiation events. One crucial difference is that ISR events are restricted to the production mechanism of $q\bar{q}$-annihilation. In contrast to that, the main production mechanism for top radiation is gluon-gluon fusion, which is the dominant production mechanism for $t\bar{t}$ pairs (see Section 2.5). From the PDFs (see Figure 2.15) one expects that events initiated by gluon-gluon fusion are most likely to have asymmetric energies in the initial state gluons, which results in a boost for the $t\bar{t}$ system and hence the decay of the top quarks is expected to leave its signature in the forward regions of the detector. However, events initiated by $q\bar{q}$-annihilation are expected to be more central because the incoming quarks are quite likely to have energies of the same order of magnitude. These differences in the production, leading to kinematic differences in the decay, were investigated. Various distributions were checked, which are the following:

- H_T; the scalar sum of the transverse momenta of all jets.

- H_T in the opposite hemisphere of the photon. This means that the angle between

the 3-momenta of the photon and the jet is between 90° and 270°.

- H_T in a cone of $\Delta R \geq 1$ around the photon.

- η distributions of the jets (ordered in decreasing p_T).

- η distributions of the jets (ordered in increasing $|\eta|$).

The second and third variable try to use the feature of ISR that the η distribution for the photon is broader than in other cases. This could result in smaller energy deposits around the photon, although for top radiation the jets are expected to be in the forward regions of the detector and the photons more in the central part (see third variable). For the second variable, one expects the reverse behaviour. If there is not much energy from jets around the photon it must be, to some extent, in the opposite hemisphere. The distributions can be seen in Figures 7.44 to 7.55. In some cases it was necessary to distinguish between dilepton and lepton+jets channels due to the different numbers of expected jets. Looking at the distributions, one observes minor differences between the four categories. It is important that in the η-distributions (Figures 7.50 to 7.53) the jets in ISR events have a broader spectrum than those from top radiation. This is contrary to the expectation. One reason could be that the emission of ISR photons forces the collision of a quark and an anti-quark to be more asymmetric and therefore the event is not expected in the central region of the detector, especially as photons are required to be quite central (generator cut on absolute photon η at 2.5), whereas jets can be in the range of $|\eta| \leq 5.0$. Due to time constraints these deviations from the expectations were not investigated more deeply. Also the explanations for the differences in the other distributions do not hold. More studies on the event kinematics might be useful. To use these variables in a NN, it is necessary to split the channels, as dilepton events and lepton+jets events differ fundamentally. Since the ATLAS 13 TeV $t\bar{t}\gamma$ analysis follows a channel independent approach, it was decided to continue without splitting the channels. The distributions that were used are those in Figures 7.50, 7.51, 7.54 and 7.55. Including the variables to the training of the 3l_16_8_4 network gives a small improvement of the separation of top radiation versus ISR. The confusion plots for training and testing are shown in Figure 7.56. The input events were weighted according to the cross section. The efficiency of correctly classified top radiation events is the same but the misclassification rate for ISR events lowers from 83% or 82% (training and testing) to 77% in both.
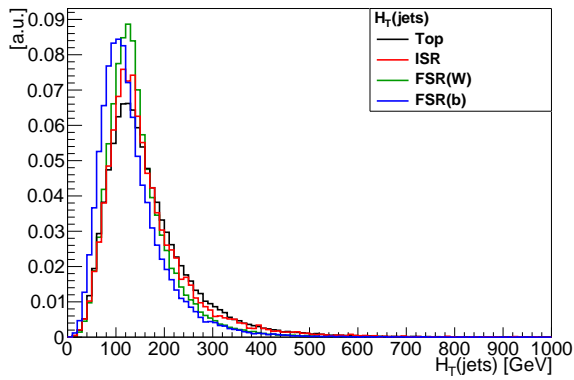
Figure 7.44.: The distributions for the scalar sum of all jet p_T in the dilepton channel.
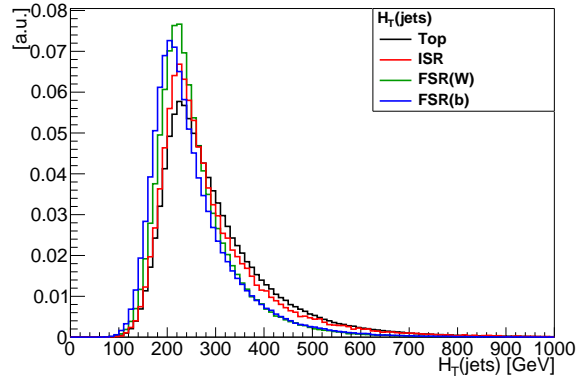
Figure 7.45.: The distributions for the scalar sum of all jet p_T in the lepton+jets channel.
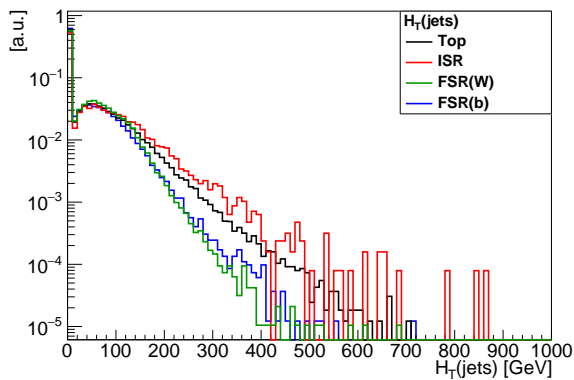
Figure 7.46.: The distributions for the scalar sum of all jet p_T in the opposite hemisphere of the photon in the dilepton channel.
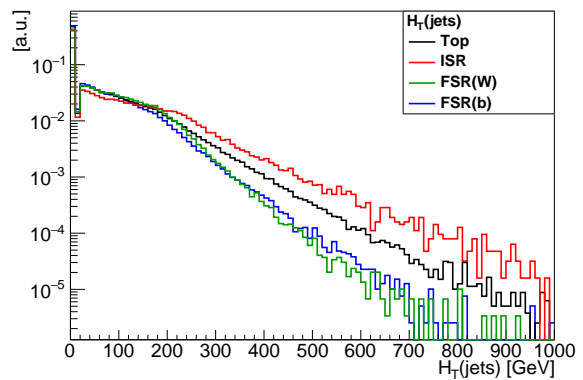
Figure 7.47.: The distributions for the scalar sum of all jet p_T in the opposite hemisphere of the photon in the lepton+jets channel.
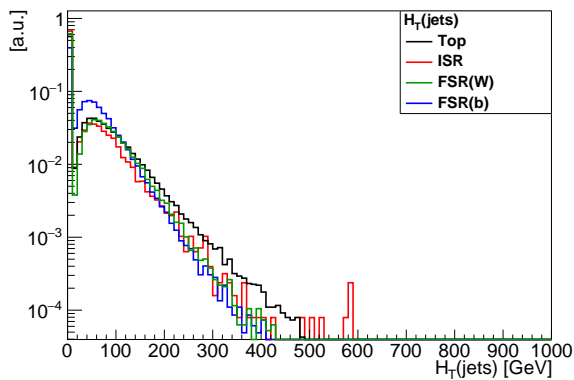
Figure 7.48.: The distributions for the scalar sum of all jet p_T in a ΔR cone of 1 around the photon in the dilepton channel.
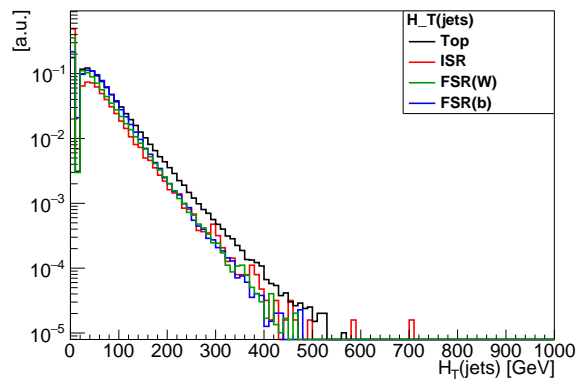
Figure 7.49.: The distributions for the scalar sum of all jet p_T in a ΔR cone of 1 around the photon in the lepton+jets channel.
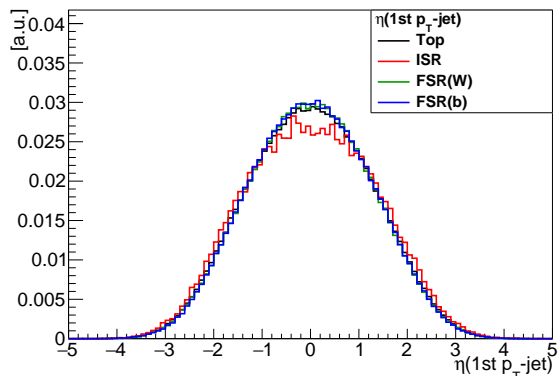
Figure 7.50.: The η distributions for the highest p_T-jet in both channels.
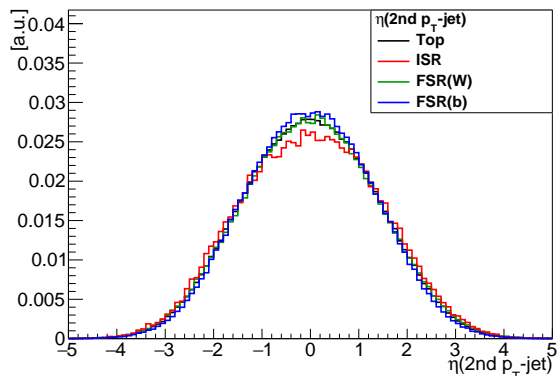
Figure 7.51.: The η distributions for the 2nd highest p_T-jet in both channels.
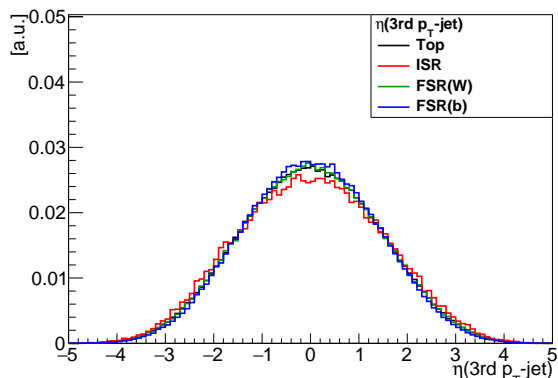
Figure 7.52.: The η distributions for the 3rd highest p_T-jet in the dilepton-channel.
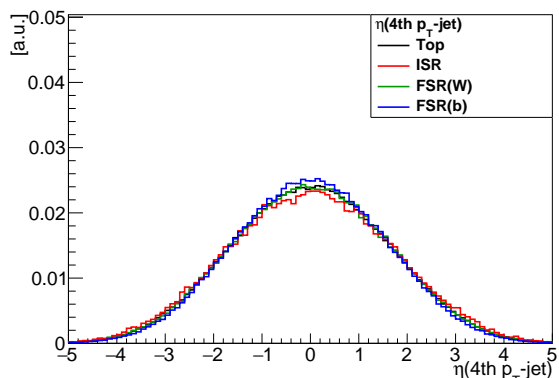
Figure 7.53.: The η distributions for the 4th highest p_T-jet in the dilepton-channel.
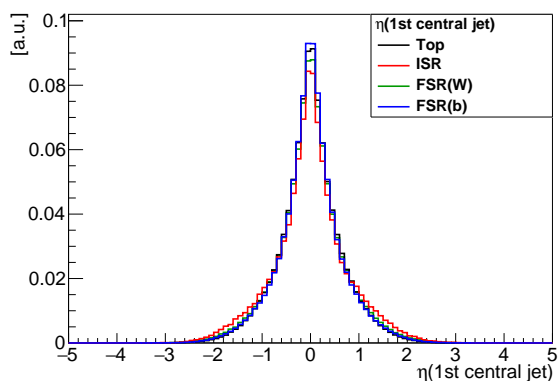
Figure 7.54.: The η distributions for the jet with lowest $|\eta|$ in both channels.
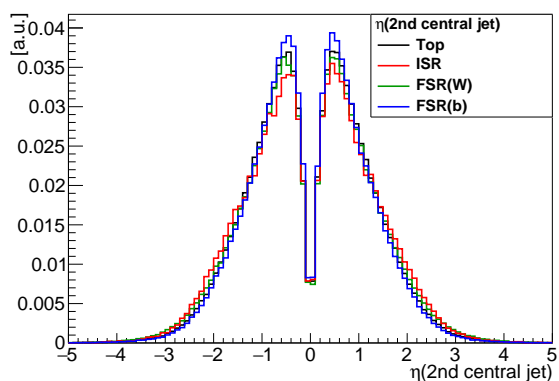
Figure 7.55.: The η distributions for the jet with 2nd lowest $|\eta|$ in both channels.
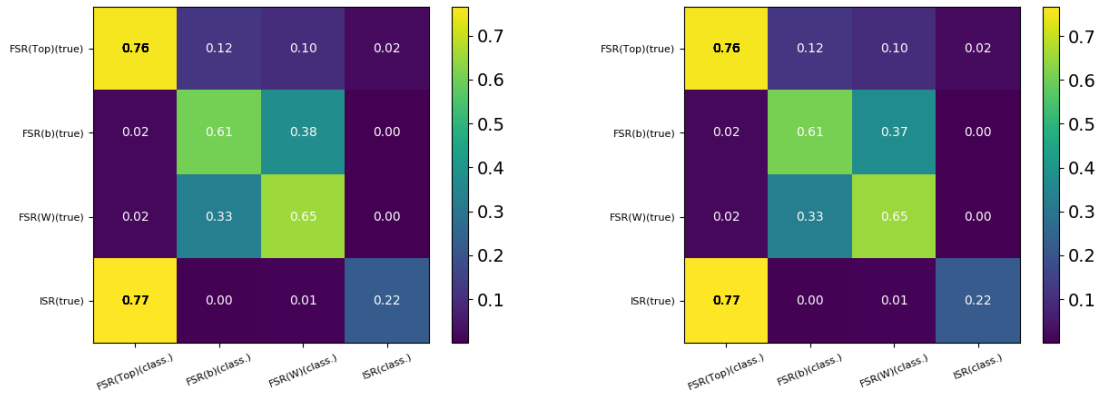
Figure 7.56.: Confusion plots for training (left) and testing (right) data of a NN with 3 layers (16/8/4 nodes).
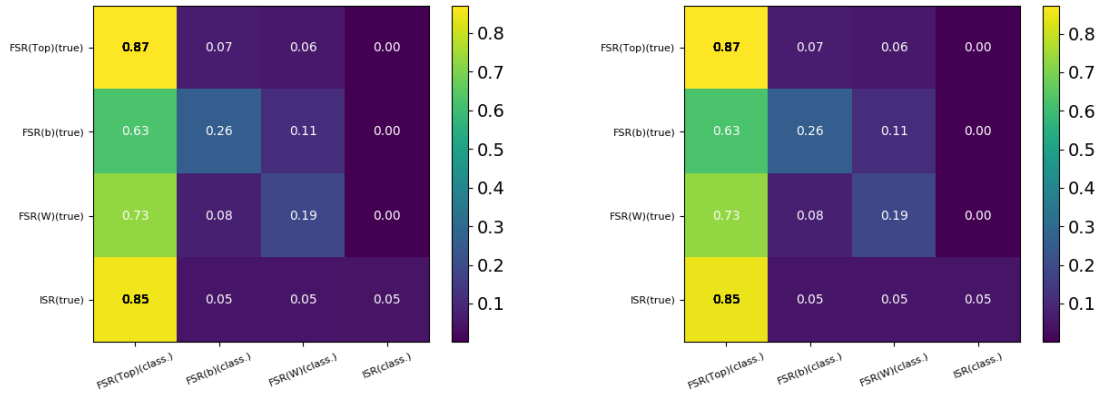
Figure 7.57.: Confusion plots for training (left) and testing (right) data of a NN with 3 layers (16/8/4 nodes).

7.3.4. Conclusions

Only a small improvement but no clear separation was achieved between top radiation and ISR events. The reason for this seems to be the separation power of the invariant masses for top radiation events versus FSR(W)/FSR(b). In fact they are so much higher than the separation of all other variables that they completely dominate the classification process. It was demonstrated by training the 3l_16_8_4 network without the invariant mass variables which left seven variables. The training was done for the laboratory frame and the confusion plots can be seen in Figure 7.57. No weighting was applied. The confusion plots show small improvement for top radiation versus ISR separation (compare Figure 7.34), but the separation between top radiation and FSR(W)/FSR(b) is lost. The second reason is that when accounting for cross sections, it is about four times more likely to have

a top radiation event than an ISR event. This impacts the classification. An ISR event must differ much from a top radiation event to overcome this and be classified correctly. This was shown by applying weights to simulate that each event is equally likely (see 7.39 and 7.40), which led to a drastic loss in signal efficiency. To get a similar separation for top radiation and ISR than for top radiation and FSR(W)/FSR(b), a variable with similar separation power as the invariant masses (for top radiation versus FSR(W)/FSR(b)) is needed. Such a variable was not found. Another way might be to separate first between top radiation/ISR and FSR(b)/FSR(W) and afterwards between top radiation and ISR. This procedure looks promising due to the results found here and in [111].

8. Studies on reconstruction level

8.1. Comparison studies

8.1.1. Comparison studies on reconstruction level

To prove that the generated samples do not differ from the official $t\bar{t}\gamma$ samples, comparison studies were done. In Section 6.1 the different categories of prompt photons that MadGraph differentiates were introduced (top radiation, ISR, FSR(b) and FSR(W)). At reconstruction level more categories appear. Besides the known categories, charged leptons and hadrons appear as the mother particles of photons. As these photons do not correspond to one of the categories present at parton level, it was decided to ignore them to assure a good training. The p_T cut for the photon was placed at 15 GeV and events with more than one photon were allowed to pass the selection. The hardest[1] photon was taken as the candidate photon of interest. Only if this photon corresponds to one of the two additional categories, the second hardest photon in the event (if present) that corresponds to one of the categories that can be connected to parton level categories, was considered.

Due to the fact that the official $t\bar{t}\gamma$ sample was produced with a nominal MadGraph version, it is not always possible to compare category by category, as the meaning of some categories differ between the modified samples and the nominal $t\bar{t}\gamma$ sample. For the events, where the mother particle was labelled a W boson or a gluon (see Section 6.1), a direct comparison is possible and no differences in the respective distributions are expected. In case of the other two categories, a comparison for both subcategories in each category (see Section 6.1) was done. Certain differences are therefore expected and prove again that the new samples provide the possibility to distinguish the categories also at reconstruction level.

The investigated variables were the p_T spectrum of the photon, the η spectrum of the photon, and the ΔR spectrum between the photon and the nearest b quark. In the samples on parton level the b quark could be identified unambiguously. At reconstruction

---

[1]Ordered in p_T.

75
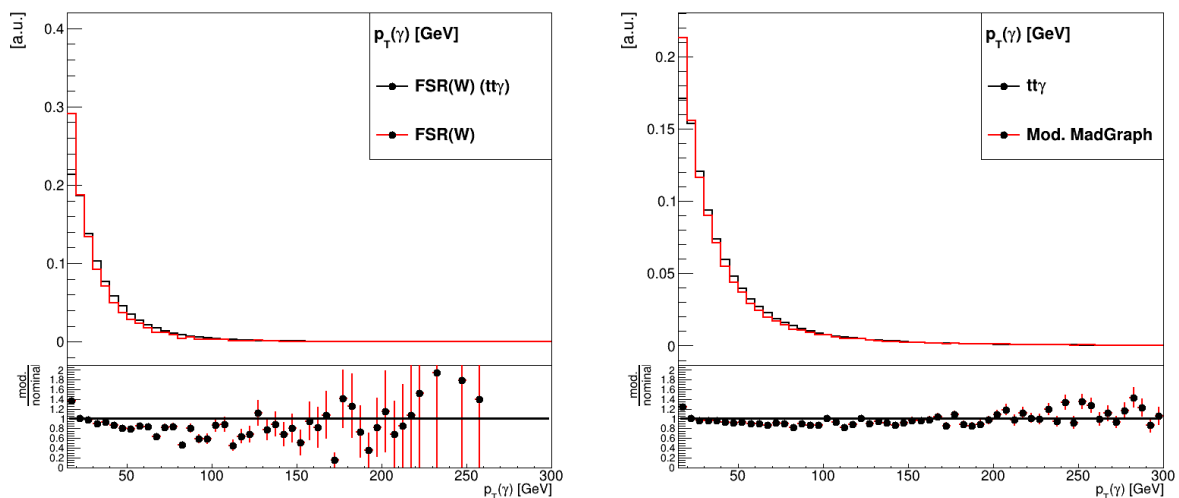
Figure 8.1.: Comparison between the p_T spectra for the modified MadGraph samples and the official $t\bar{t}\gamma$ samples for the FSR(W) category (left) and the hardest photon (right).

level b-tagging is used. To label a jet as a b-jet, the 85% working point was chosen and the event needs to have at least one b-jet. The loose working point was chosen to maximise event yields and a tighter working point would have cut away too many events and rendering a comparison not very meaningful.

The results agree with the expectations, but the p_T spectrum of the modified samples is much softer. This disagreement is observed over all categories. An example plot of this is found on the left in Figure 8.1. As seen there, the discrepancy is mainly due to the first bin. This holds for all categories. Therefore, the full p_T spectra of the hardest photons were also compared. The same discrepancy can be observed (see Figure 8.1 (right)). An examination of the second and third hardest photon also showed that there is a discrepancy especially in the first bin, but it becomes smaller. As expected, for the FSR(W) and the off-shell photon categories, no differences were observed for η and ΔR.

At parton level, events where the photon comes from a b quark or an on-shell top quark, are labelled the same in the nominal MadGraph. A comparison of this category with both subcategories from the modified samples shows clear differences for η and ΔR. The distributions can be seen in Figures 8.2 and 8.3. For the ΔR distributions one sees that if the nominal MadGraph distribution is compared to the distribution where all photons come from on-shell top quarks, the latter distribution peaks at higher values which is expected, as photons emitted by a b quark should be in general near to the b quark. If instead the nominal MadGraph distribution is compared to FSR(b) events both peak at
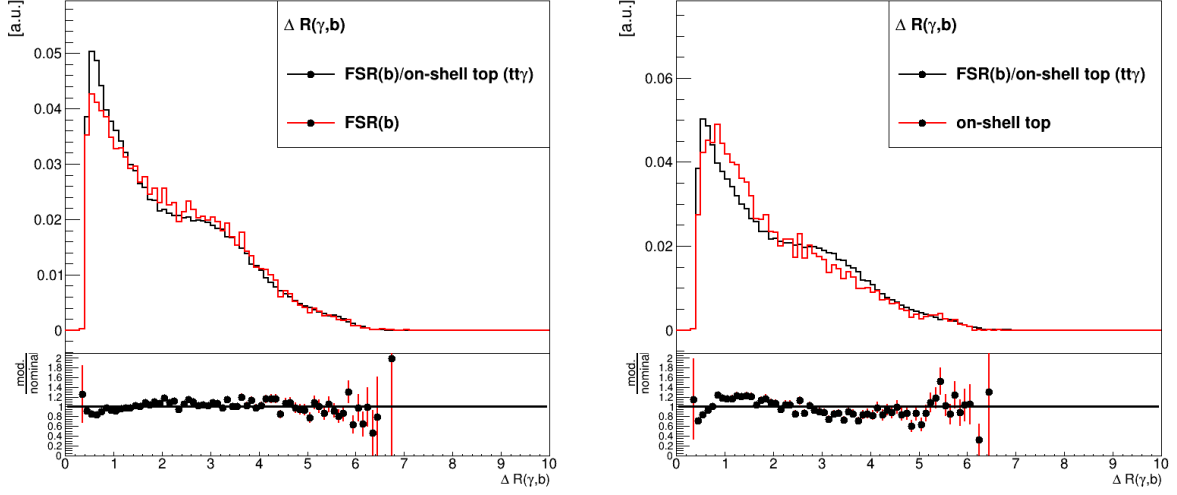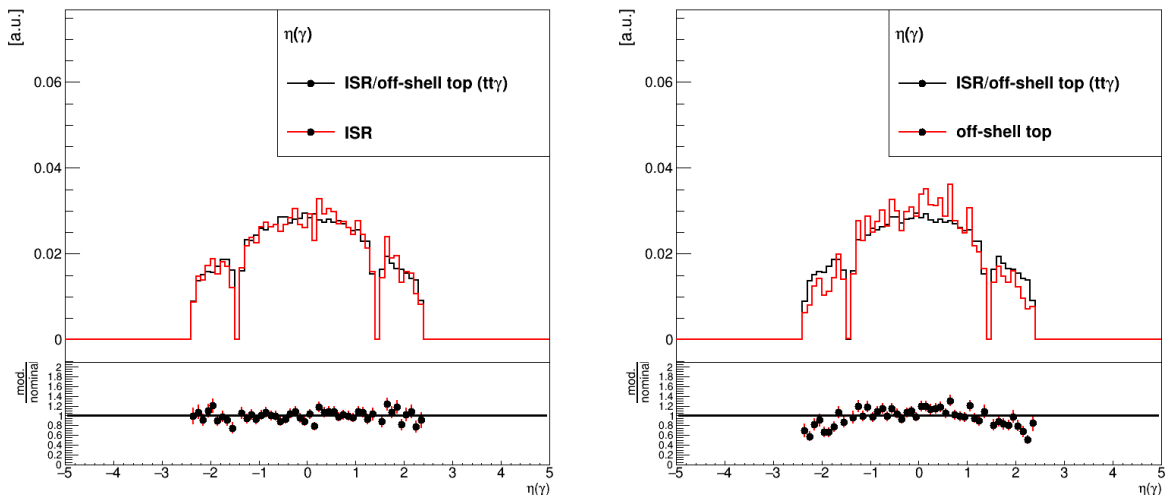
Figure 8.2.: Comparison between the ΔR spectra for the modified MadGraph samples (left only FSR(b); right on-shell top quarks) and the official $t\bar{t}\gamma$ samples (FSR(b) and on-shell top quarks).
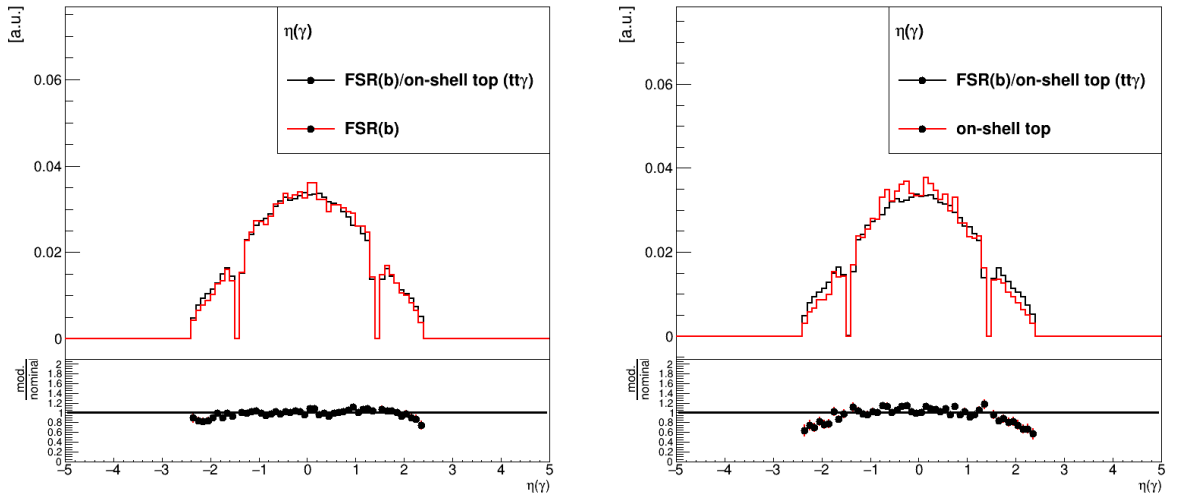
Figure 8.3.: Comparison between the η spectra for the modified MadGraph samples (left only FSR(b); right on-shell top quarks) and the official $t\bar{t}\gamma$ samples (FSR(b) and on-shell top quarks).

77
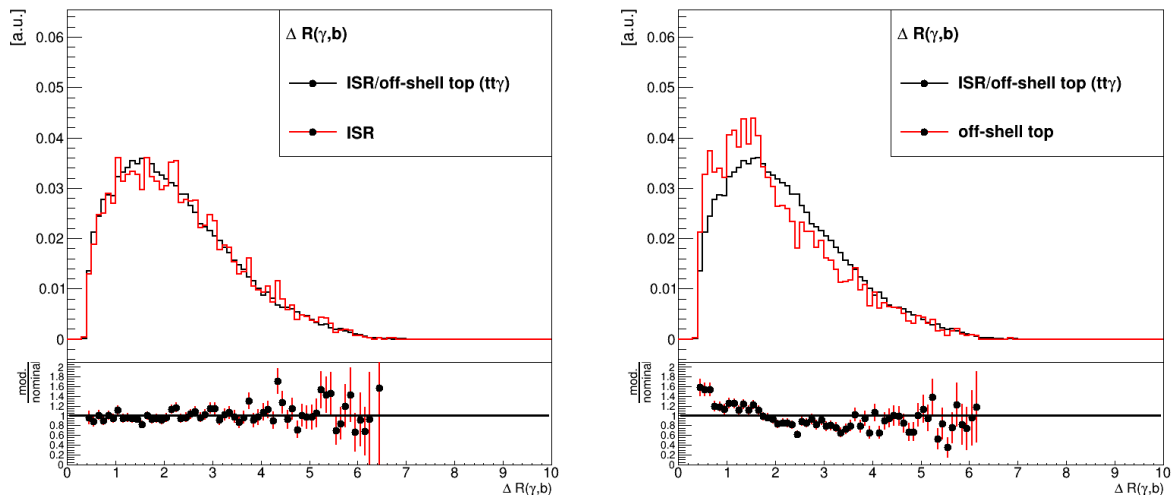
Figure 8.4.: Comparison between the ΔR spectra for the modified MadGraph samples (left only ISR; right off-shell top quarks) and the official $t\bar{t}\gamma$ samples (ISR and off-shell top quarks).

the same values but the peak value for the distribution of the modified samples is lower. This is expected to be caused by the difference seen in Figure 6.3. In case of the η distributions, the nominal MadGraph distribution is broader in comparison to photons emitted by on-shell top quarks. This is somewhat expected as on parton level the η-distribution for FSR(b) is broader than for top radiation[2] (see Figure 7.3 to compare).

The opposite effect is not observed. This is due to the fact that most events in the nominal MadGraph sample are actually FSR(b) events and not related to on-shell top quarks. Therefore, the effect of a narrower nominal η distribution is not visible with the given statistics. A comparison between the ISR/off-shell top quark category (from the nominal MadGraph) and ISR and off-shell top quark category individually (modified MadGraph) shows a visible difference only for ISR/off-shell top quark versus off-shell top quarks (see Figures 8.4 and 8.5). The η distribution for photons from off-shell top quarks is narrower than for ISR photons, as expected (compare Figure 7.3). In case of the ΔR distributions, the events where the photon comes from the off-shell top quark peak at lower values. It is expected, as the photon and the b-quark from the top decay are relatively near to each other, whereas for ISR this is not the case, as the photon trajectory is not related to the top quarks (compare Figure 7.5).

To show that these differences are not related to systematic differences between the samples (like in the p_T distributions in Figure 8.1.), η and ΔR distributions for the hardest

---

[2]Although in Figure 7.3 on-shell and off-shell top quarks are combined.

Figure 8.5.: Comparison between the η spectra for the modified MadGraph samples (left only ISR; right off-shell top quarks) and the official $t\bar{t}\gamma$ samples (ISR and off-shell top quarks).

photon are shown in Figure 8.6. In both cases, the nominal distribution and the one from the modified samples agree within statistical uncertainties. The conclusion is that the discrepancy seen at reconstruction level is (besides small other effects) an effect of the p_T cuts. Smaller deviations are due to different options used in the generation of the samples and there is no reason not to use the modified samples, although the differences might be of relevance when it comes to estimate systematic uncertainties.

8.1.2. Comparison studies at parton level

To investigate the differences in the p_T spectra, the parton level samples were checked (see Figure 8.7). There is still a discrepancy in the first bin, but it has become smaller. The reason why it is smaller might be due to the fact that on generator level the p_T cut on the photon between the modified and the official samples is different. For the official sample the photon p_T cut was at 15 GeV and for the modified samples at 10 GeV. As a photon p_T cut of 15 GeV is the same for all samples in the event selection, events with photons of more than 15 GeV before the detector simulation might be below 15 GeV (due to resolution effects) afterwards and cut away. The same happens in the case of the modified samples but there are also events that enter the detector simulation with photons with less than 15 GeV and come out with a photon p_T above 15 GeV due to those resolution effects. In this case, the number of events with a photon above 15 GeV is kept
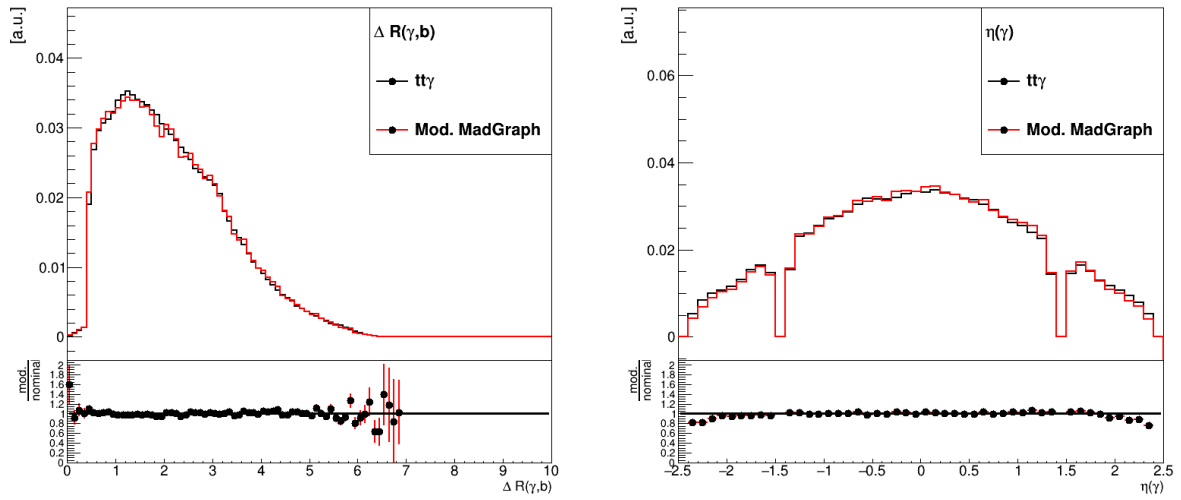
Figure 8.6.: Comparison of ΔR spectra (left) and the η spectra (right) for the modified MadGraph samples and the official $t\bar{t}\gamma$ samples for the hardest photon.
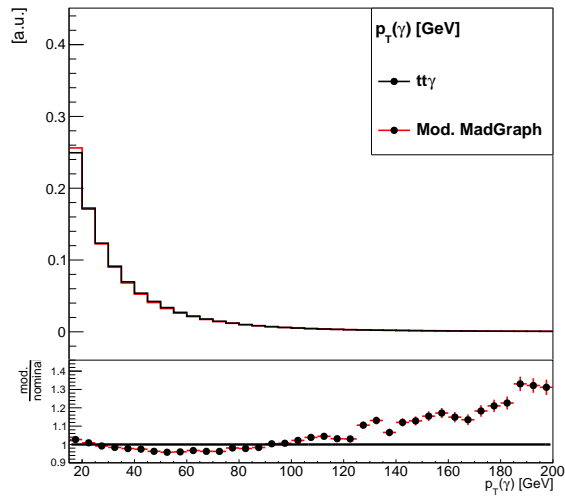
Figure 8.7.: Comparison of p_T spectra for the modified MadGraph samples and the official $t\bar{t}\gamma$ samples at parton level.
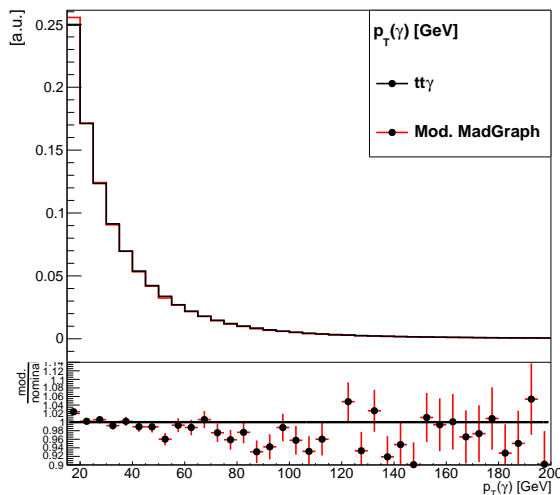
Figure 8.8.: Comparison of p_T spectra for the modified MadGraph samples and the official $t\bar{t}\gamma$ samples at parton level with basically the same set of options. The ratio in the first bin ($>$1) explains why there are more bins with ratios lower than 1 than with ratios higher than 1.

more or less equal. Therefore the photon p_T spectrum is softer at low energies than the official samples. There are also significant deviations in the tail of the p_T distribution in Figure 8.7. The discrepancies seen in Figure 8.7 must come from the different generator options used. A sample with basically the same options[3] as used for the official samples was produced in the standalone version of MadGraph that was used for the modified samples. The result (see Figure 8.8) still shows the discrepancy in the first bin that has basically not changed, but the tail shows better agreement. In general, the differences are within +5% and -10%.

8.2. Variable distributions at reconstruction level

The variables that were used at parton level shall also be used at reconstruction level to train a NN. The performance depends heavily on the differences of the variable distributions for the different categories. The hadronisation process, the detection of the particles, the reconstruction of the different objects, and the effort to match them to parton level particles introduces many effects that lead to a smearing of the parton level distributions. To check the effect of this smearing, the distributions for all 15 variables were plotted and can be seen in Figures 8.9 to 8.23. It was decided to use the laboratory

---

[3]Not all options could be adopted easily.

frame at reconstruction level, as the reconstruction of the top quarks contains systematic uncertainties. In a Lorentz Boost these uncertainties would be transfered to variables that are independent of top quark reconstruction, and the reconstruction is expected to be biased because KLFitter is designed to reconstruct top quark pairs and not $t\bar{t}\gamma$ events, where the photon must be taken into account. The number of events is slightly lower than the yields found in Section 6.4. This is due to various requirements: at least one b-tag is required; the events where the mother particle of the photon does not belong to one of the categories found in the LHE files are discarded; and also KLFitter cannot reconstruct every event. Therefore only 30701 events are used[4] for the Top category and the three background categories accumulate 17059 events (ISR: 2154; FSR(W): 5697; FSR(b): 9208). Comparing the distributions to those at parton level (Figures 7.1 to 7.11 and 7.50,7.51,7.54 and 7.55) most of them exhibit a similar behaviour, but the smearing is clearly visible. Unfortunately, the distributions also suffer from low statistics which results in clearly visible fluctuations. For the η distributions of the leading and second leading jet (in p_T), the small difference between ISR and the other categories observed at parton level is not visible at all. The distributions that differ much from those at parton level are the invariant mass distributions. The distributions where only the mass of Wb systems are plotted (Figures 8.15 and 8.16) do not show any difference at all. Basically all events are centered around the top quark mass. This huge discrepancy with respect to the parton level distributions (Figures 7.7 and 7.8) is related to KLFitter. Its design goal, as described in this section, forces the mass of the fitted top quarks to be around the pole mass without taking the photon into account[5]. If one includes the photon in the calculation of the invariant mass a difference between the categories becomes visible (see Figures 8.17 and 8.18). But this difference is not related to the fact that the photon is emitted at different stages of the top quark pair production and decay, but to the fact the p_T and E distributions of top radiation and ISR are harder than those of FSR(b)/FSR(W) (see Figures 8.9 and 8.10). Nevertheless the invariant mass distributions which take the photon into account show useful differences.

8.3. Training neural networks

Due to the cuts in the reconstruction process, the samples and therefore the variable distributions changed from parton level to reconstruction level. So the optimal NN architecture

---

[4]If there is more than one photon in an event belonging to one of the usable categories, the hardest one is taken.

[5]Even if the top mass is allowed to float in the fitting, the distributions do not reproduce a shape similar to the one seen at parton level.
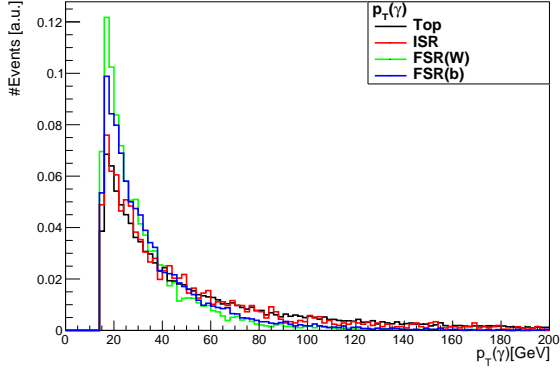
Figure 8.9.: The distributions of photon p_T for the four different categories.
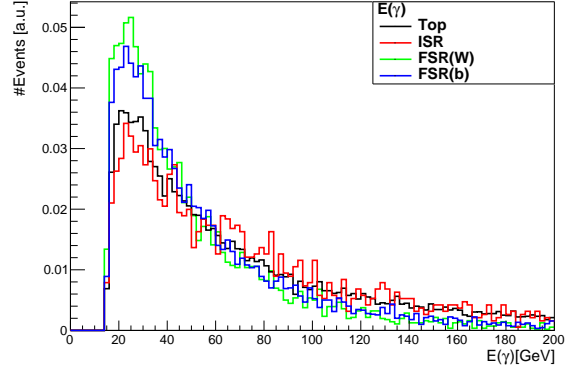
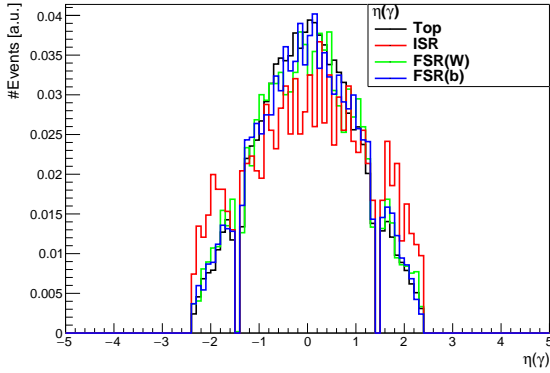**Figure 8.10.:** The energy distribution of the photon for the four different categories.

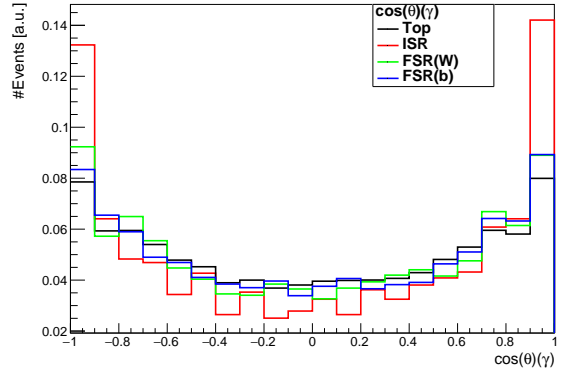**Figure 8.11.:** The distributions of photon η for the four different categories.

Figure 8.12.: The $\cos(\theta)$ distribution of the photon for the four different categories.
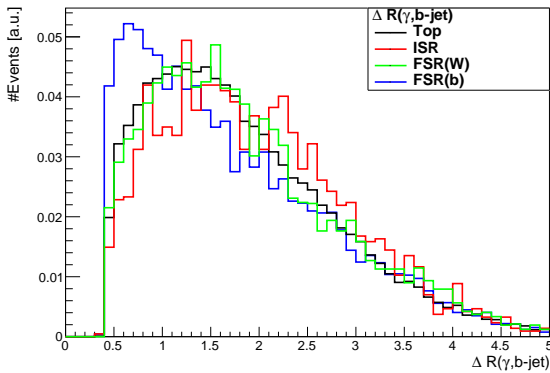

Figure 8.13.: The ΔR distributions between the photon and the nearest b quark for the four different categories.
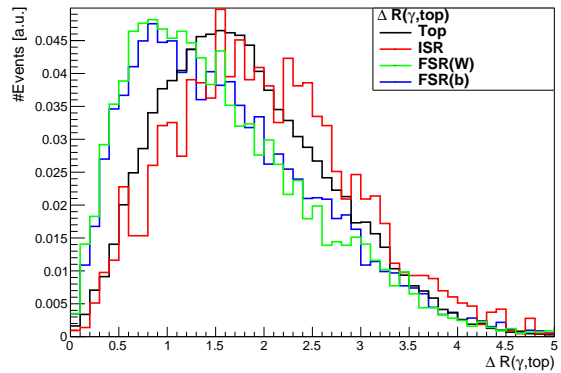
Figure 8.14.: The ΔR distributions between the photon and the nearest top quark for the four different categories.
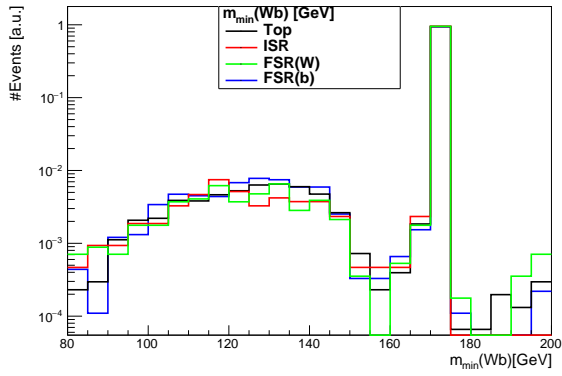
Figure 8.15.: The distributions for the lower invariant mass of both Wb systems for the four different categories.
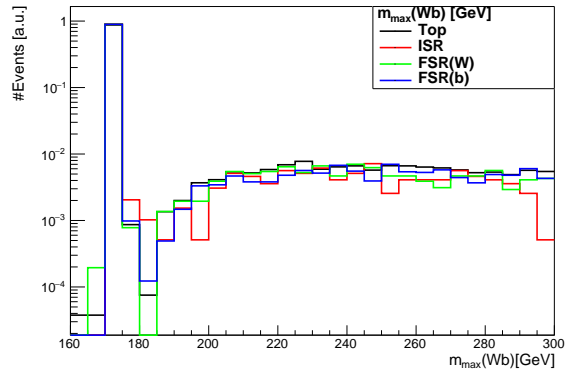
Figure 8.16.: The distributions for the higher invariant mass of both Wb systems for the four different categories.
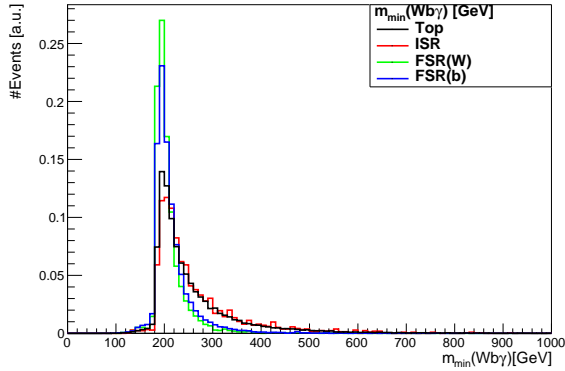
Figure 8.17.: The distributions for the lower invariant mass of both $Wb\gamma$ systems for the four different categories.
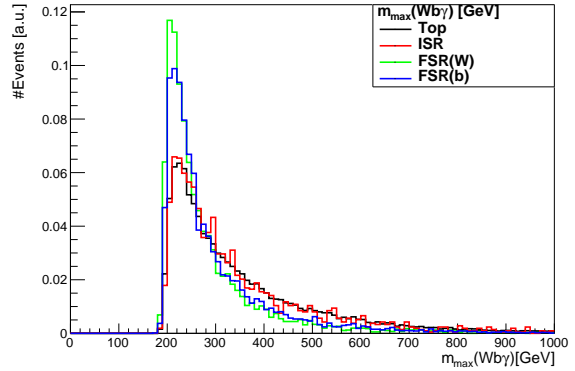
Figure 8.18.: The distributions for the higher invariant mass of both $Wb\gamma$ systems for the four different categories.
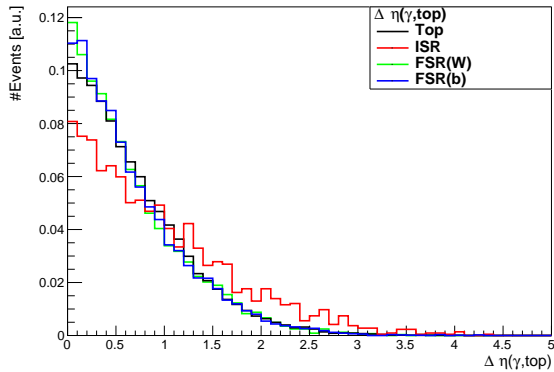
Figure 8.19.: The distribution for $|\Delta\eta|$ between the photon and the nearest top quark for the four different categories.
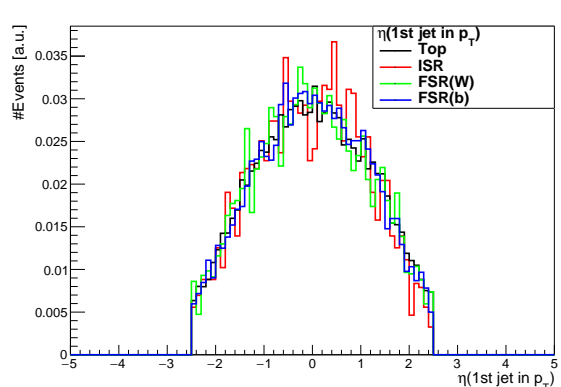
Figure 8.20.: The η-distributions for the hardest jet in p_T for the four different categories.
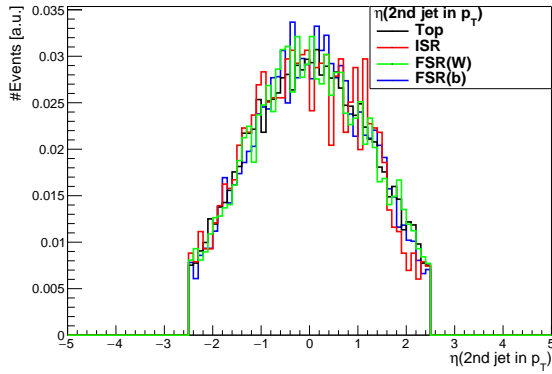
Figure 8.21.: The η-distributions for the second hardest jet in p_T for the four different categories.
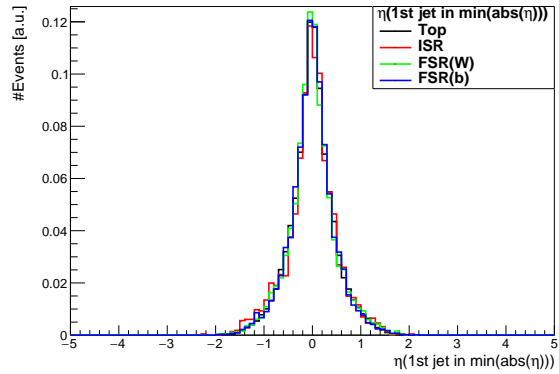
Figure 8.22.: The η-distributions for the jet with lowest $|\eta|$ for the four different categories.
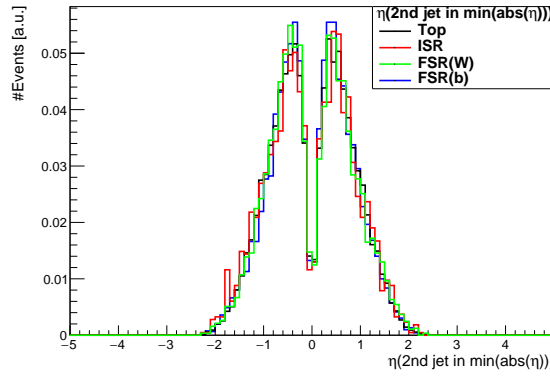
Figure 8.23.: The η-distributions for the jet with second lowest $|\eta|$ for the four different categories.

must be found again. But as most distributions have similar shapes at reconstruction level and at parton level, the same basic architectures as at parton level were trained. They are listed in Table 8.1. The weights were set to account for the different cross sections and all 15 variables were used. The results of the networks differ only within a few percentage

Name	No. of layers	Nodes per layer	Batch size	Epoch	Frame	Weights	Further information
3l_16_8_4	3	16/8/4	1000	500	Lab	xsec	-
3l_8_16_4	3	8/16/4	1000	500	Lab	xsec	-
3l_60_100_4	3	60/100/4	1000	500	Lab	xsec	-
4l_32_16_8_4	4	32/16/8/4	1000	500	Lab	xsec	-
4l_16_32_8_4	3	16/32/8/4	1000	500	Lab	xsec	-
5l_64_32_16_8_4	5	64/3216/8/4	1000	500	Lab	xsec	-
5l_32_64_16_8_4	5	32/64/16/8/4	1000	500	Lab	xsec	-
6l_128_64_32_16_8_4	6	128/64/32/16/8/4	1000	500	Lab	xsec	-
6l_64_128_32_16_8_4	3	64/128/32/16/8/4	1000	500	Lab	xsec	-
6l_60_100_150_100_50_4	6	60/100/150/100/50/4	1000	500	Lab	xsec	-

Table 8.1.: Basic NN architecture trained with the samples at reconstruction level.
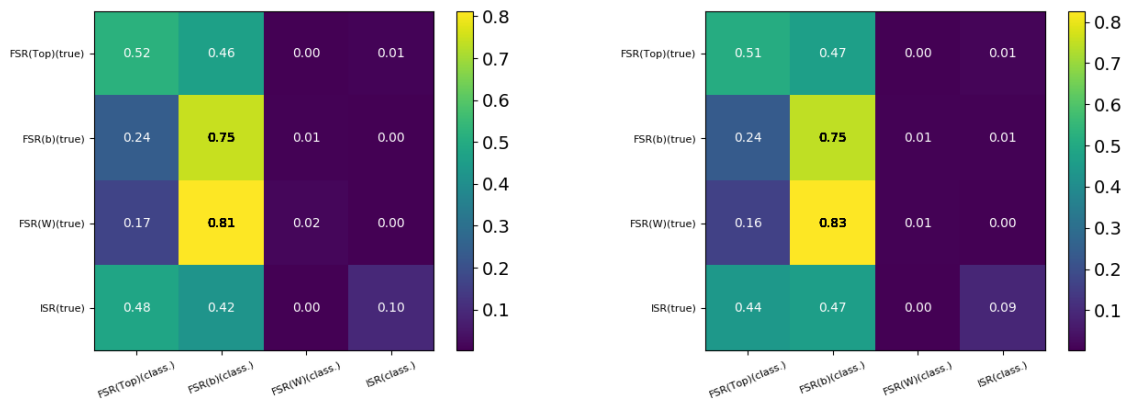
Figure 8.24.: Confusion plots for training (left) and testing (right) data of the NN with 4 layers (16/32/8/4 nodes).

points. To define the best model the following ratio was calculated

$$\frac{0.41 \cdot p_{top} N_{top}}{1.35(p_b N_b + p_W N_W + p_{ISR} N_{ISR})} \ . \tag{8.1}$$

p gives the percentage that an event of a certain category is classified as top radiation (from testing procedure), N gives the total number of events and the factors 0.41 and 1.35 are the cross sections in pb given by MadGraph. In case two NN architectures had similar ratios, it was checked how much the statistics would decrease and how well ROC-curves for testing and training agree. Using no neural network but just the raw numbers from Section 8.2 gives a ratio of 0.55. The best architecture had four layers and the following sequence of nodes per layer: 16-32-8-4. The ratio was 1.17. The confusion plots for training and testing are in Figure 8.24. Obviously the data allows the NN basically to differentiate only between two categories. The reason for this is the similarity between photons from top radiation and ISR on the one hand and FSR(b) and FSR(W) on the other hand. Taking into account low statistics, the result is that the network is not able to categorise more than two categories properly. The distributions for the four output values for the testing procedure of the NN are shown in Figures 8.25 and 8.26. The ROC-curves (see Figures 8.27 to 8.30) also show mostly good behaviour, which indicates that no overtraining happened. Some of them show deviations between training and testing of the order of 1 or 2 percentage points. Due to the low statistics available for training and testing, this is acceptable. In contrast to Section 7.3.1 there are no obvious starting points to improve the separation between the categories.

One idea was to use methods from hypothesis testing. Therefore, in each output variable the three background distributions were unified to one. The signal and the background
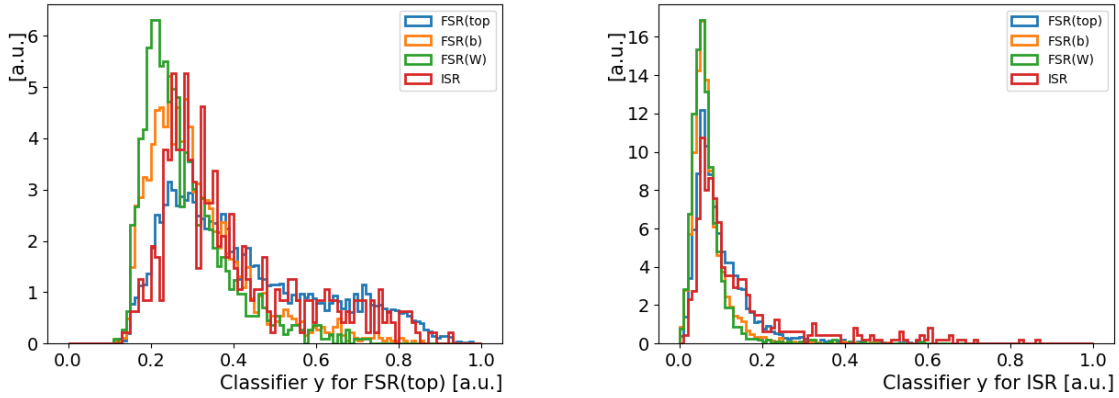
Figure 8.25.: The distributions for the classifier for top radiation (left) and ISR (right) for the different categories. The NN had 4 layers with 16, 32, 8 and 4 nodes respectively.
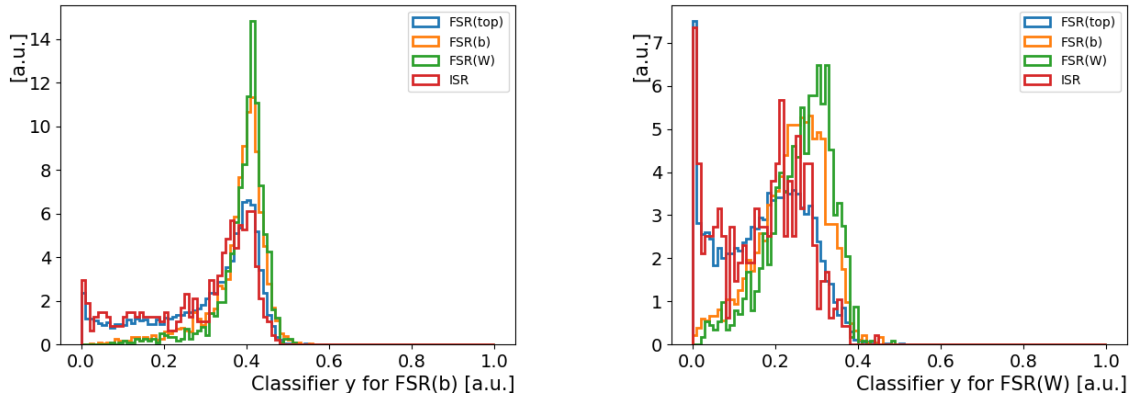

Figure 8.26.: The distributions for the classifier for FSR(b) (left) and FSR(W) (right) for the different categories. The NN had 4 layers with 16, 32, 8 and 4 nodes respectively.

distribution can be used as a test statistic to calculate the significance α and the power $1 - \beta$, where the null hypothesis H_0 is that the event is background and the alternative hypothesis H_1 is that the event is signal. As a weighting method the ratio r of both values is calculated:

$$r = \frac{\alpha}{1 - \beta} = \frac{\int_{t_c}^1 g(t|H_0)}{\int_{t_c}^1 g(t|H_1)} \ . \tag{8.2}$$

t_c is the output value of the neural network for the corresponding classifier. The same neural network with the same weights as before was used to test this procedure. If one plots the distributions of r one can see that signal events tend to have lower values than background. This is expected as for a background event α should be small and for a signal
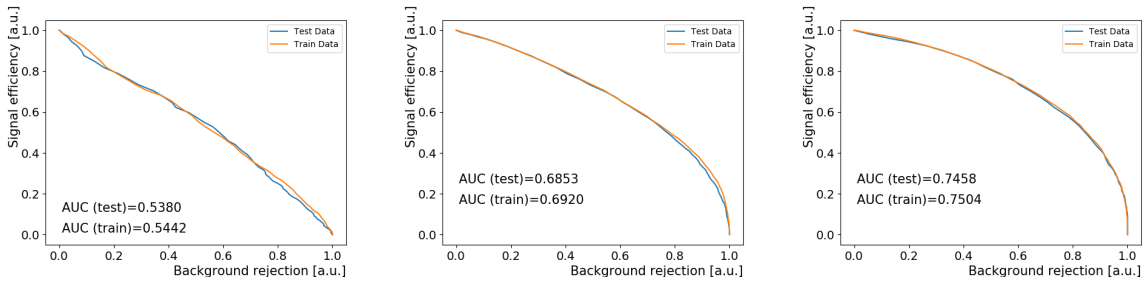
Figure 8.27.: The ROC-curves for the classifier for top radiation. On the left with respect to ISR, in the middle with respect to FSR(b) and on the right with respect to FSR(W) (model: 4 layers and 16/32/8/4 nodes).
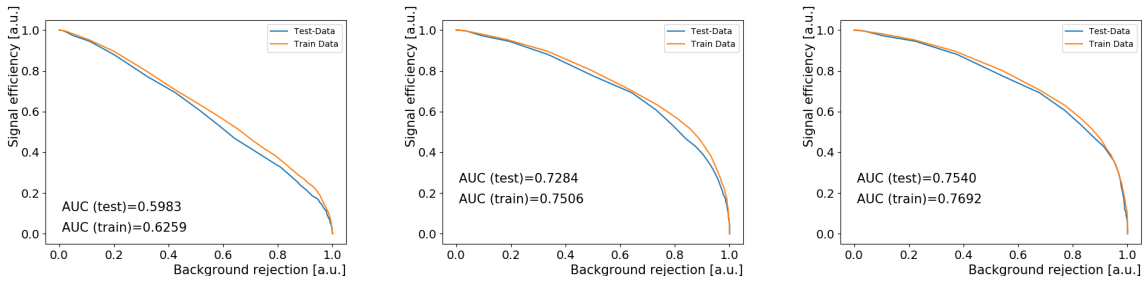

Figure 8.28.: The ROC-curves for the classifier for ISR. On the left with respect to top radiation, in the middle with respect to FSR(b) and on the right with respect to FSR(W) (model: 4 layers and 16/32/8/4 nodes).
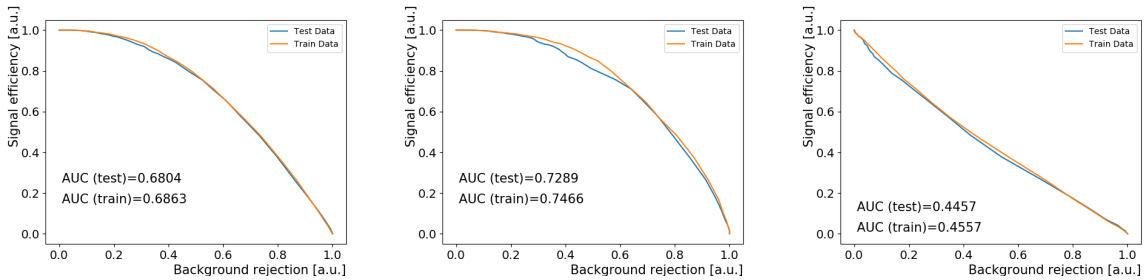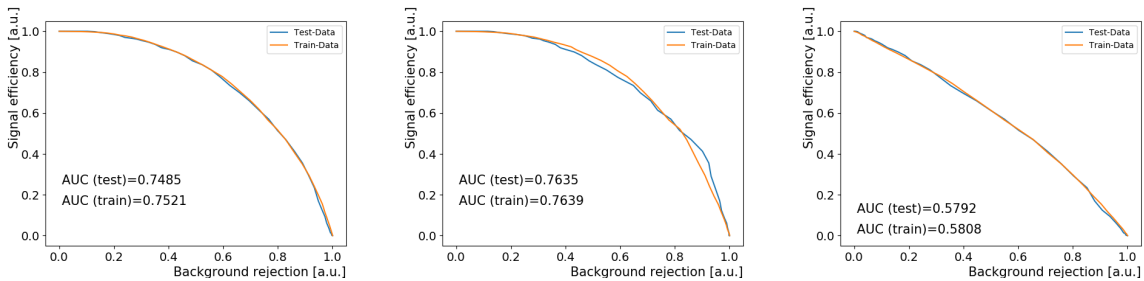

Figure 8.29.: The ROC-curves for the classifier for FSR(b). On the left with respect to top radiation, in the middle with respect to ISR and on the right with respect to FSR(W) (model: 4 layers and 16/32/8/4 nodes).

Figure 8.30.: The ROC-curves for the classifier for FSR(W). On the left with respect to top radiation, in the middle with respect to ISR and on the right with respect to FSR(b) (model: 4 layers and 16/32/8/4 nodes).
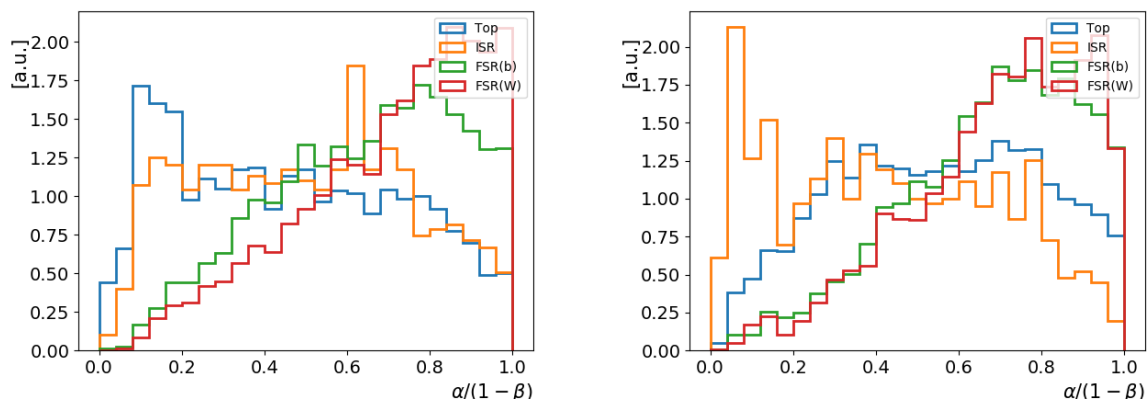
Figure 8.31.: The distributions for $r = \alpha/(1-\beta)$ for top radiation (left) and ISR (right) for the different categories. The NN had 4 layers with 16, 32, 8 and 4 nodes respectively.
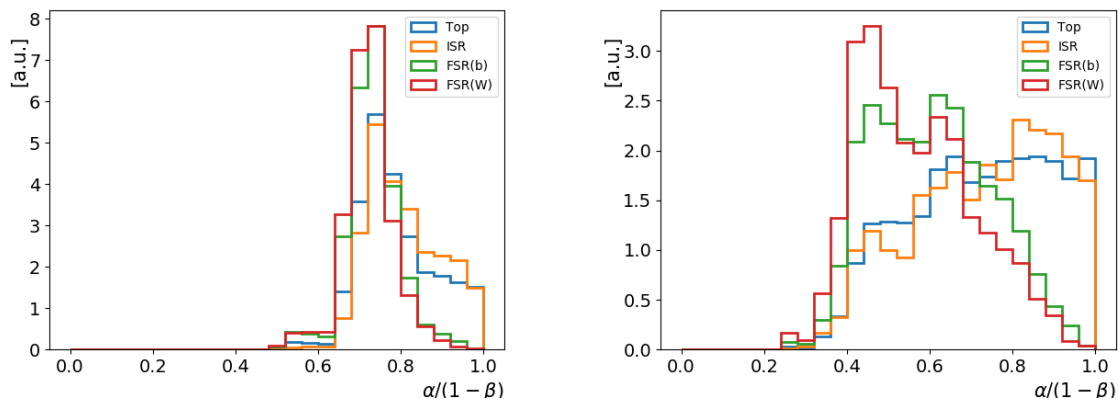

Figure 8.32.: The distributions for $r = \alpha/(1-\beta)$ for FSR(b) (left) and FSR(W) (right) for the different categories. The NN had 4 layers with 16, 32, 8 and 4 nodes respectively.

event $1-\beta$ should be large. The distributions for the four output classifiers (training) can be seen in Figures 8.31 and 8.32. To decide how an event is classified, the lowest value of the four r values is taken. The corresponding confusion plots are in figure 8.33 for training and testing. It can be seen that most events are classified as top radiation (here FSR(Top)). Although events are now classified to all four categories, the value calculated with Equation 8.1 is smaller than before. It decreased from 1.17 to 0.81. Therefore this method, although appearing better at a first glimpse, does not improve the results.

In order to be coherent with the studies at parton level, it was also decided to change the weights so that each category is equally important[6]. In this case the best model was the model with three layers and 60, 100 and 4 nodes per layer. The results (confusion

---

[6]The product of weight and number of events is the same for each category.
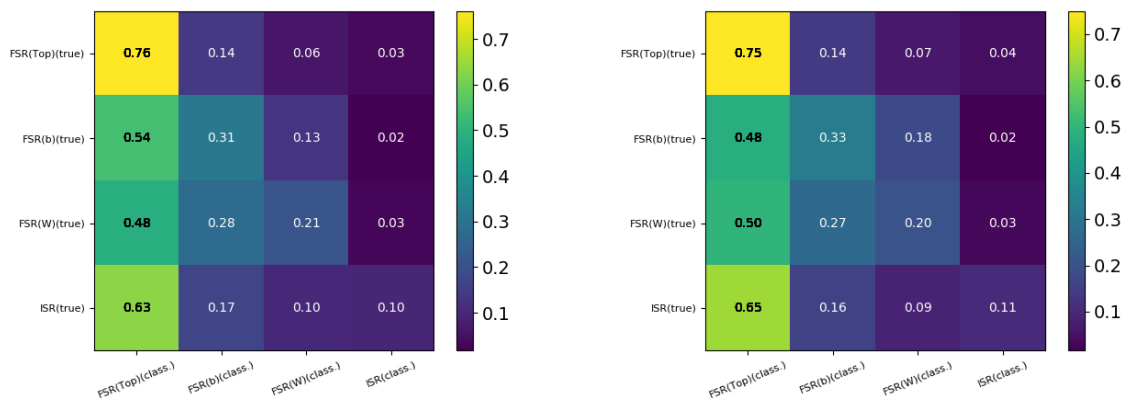
Figure 8.33.: Confusion plots for training (left) and testing (right) data of the NN with 4 layers (16/32/8/4 nodes) in case that $r = \alpha/(1-\beta)$ was used to classify events.

plots, output distributions and ROC-curves) from this model are presented in Figures 8.34 to 8.40. It can be seen that the confusion plots now show that the network can classify the events into the four categories, although the classification often fails. Only 30 to 40% of the events are classified correctly. In comparison with the results from parton level, this is much worse. One reason for this is the low statistics which results in larger fluctuations in the sample distributions. Another reason is the difference in the invariant mass distributions between truth and reconstruction level. The separation power given by these distributions at parton level is drastically reduced at reconstruction level. The ROC-curves are slightly worse than before, but in all cases it involves ISR events. Therefore the deviations between the ROC-curves for training and testing are mostly due to the low statistics and not due to overtraining.

8.4. Conclusions

Although the good results from parton level were not reproduced, one can say that despite low statistics and less separation within the variables, the purity of the sample can be enhanced. This comes with a loss in statistics. How much the sensitivity to the top-photon coupling is improved, must be studied, taking the background processes to the $t\bar{t}\gamma$-process into account, as well as leptonically decaying τ-leptons. Also necessary is to increase statistics to train the NNs and an algorithm that is directly dedicated to $t\bar{t}\gamma$ reconstruction, because KLFitter is biased.
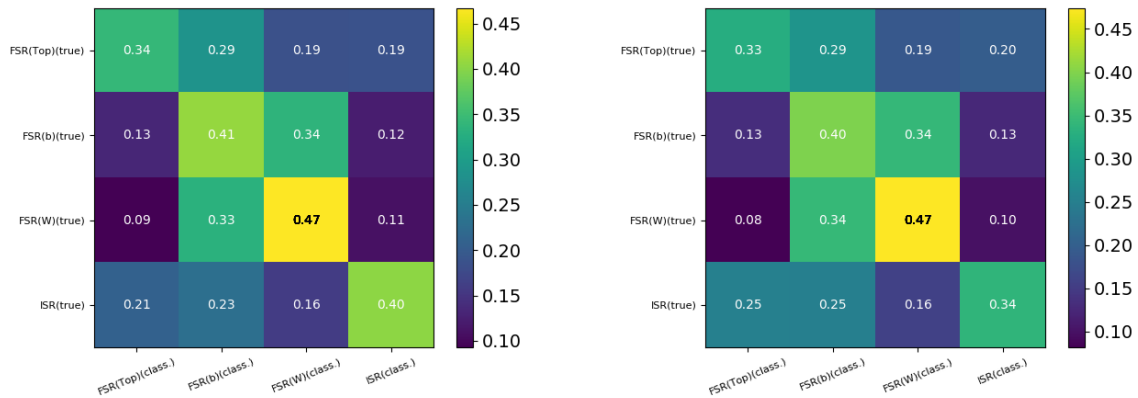
Figure 8.34.: Confusion plots for training (left) and testing (right) data of the NN with 3 layers (60/100/4 nodes).
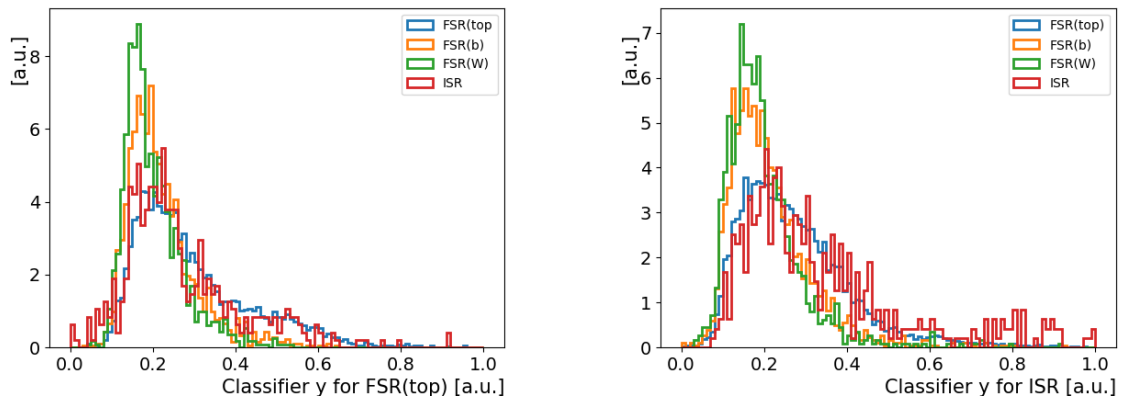

Figure 8.35.: The distributions for the classifier for top radiation (left) and ISR (right) for the different categories. The NN had 3 layers with 60, 100 and 4 nodes respectively.
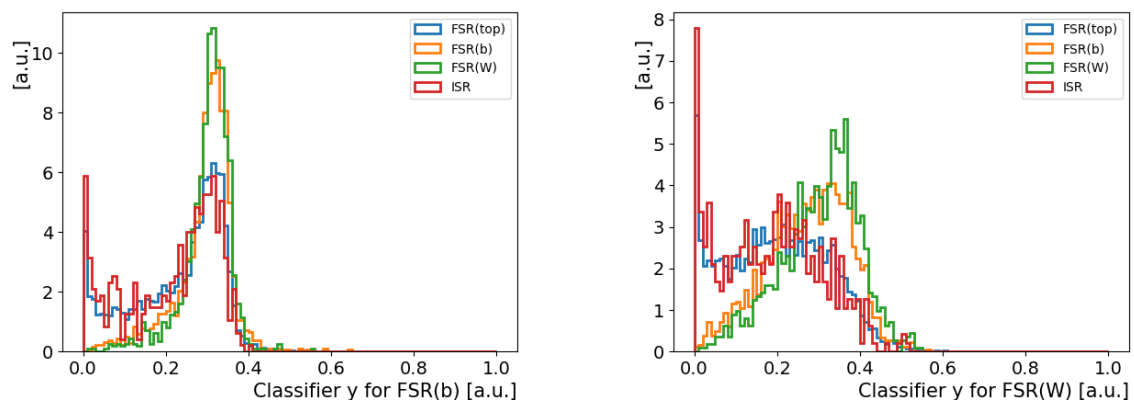

Figure 8.36.: The distributions for the classifier for FSR(b) (left) and FSR(W) (right) for the different categories. The NN had 3 layers with 60, 100 and 4 nodes respectively.
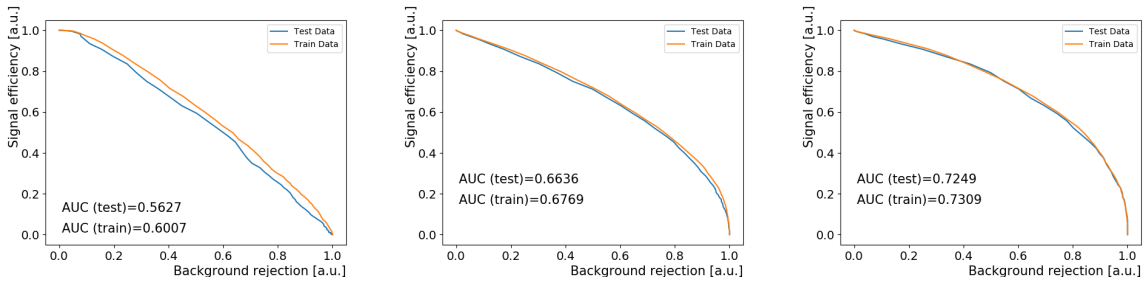
Figure 8.37.: The ROC-curves for the classifier for top radiation. On the left with respect to ISR, in the middle with respect to FSR(b) and on the right with respect to FSR(W) (model: 3 layers and 60/100/4 nodes).
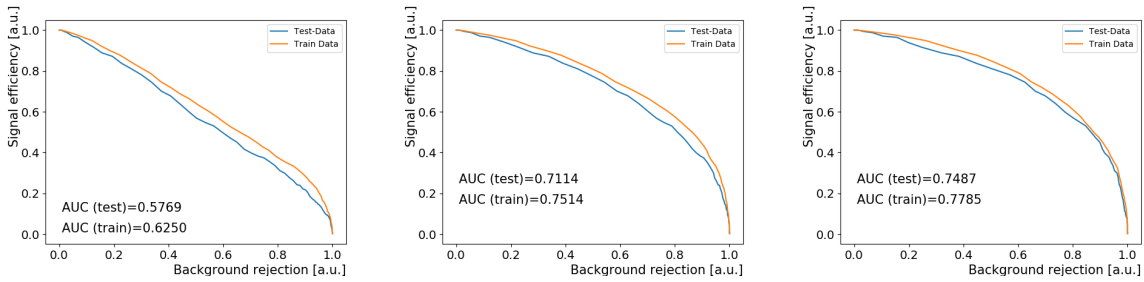

Figure 8.38.: The ROC-curves for the classifier for ISR. On the left with respect to top radiation, in the middle with respect to FSR(b) and on the right with respect to FSR(W) (model: 3 layers and 60/100/4 nodes).
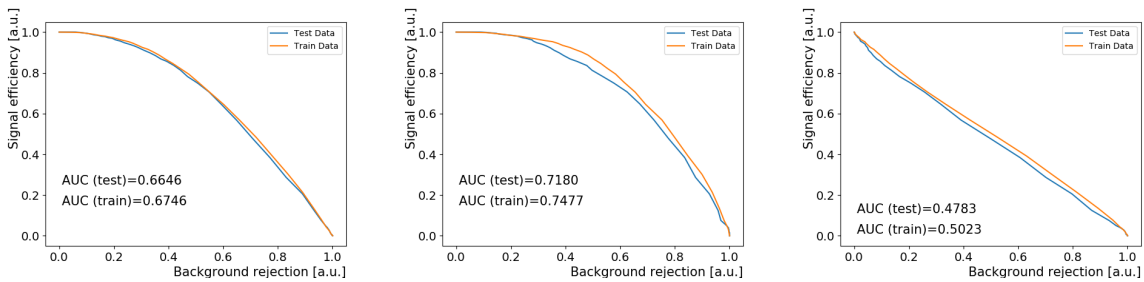

Figure 8.39.: The ROC-curves for the classifier for FSR(b). On the left with respect to top radiation, in the middle with respect to ISR and on the right with respect to FSR(W) (model: 3 layers and 60/100/4 nodes).
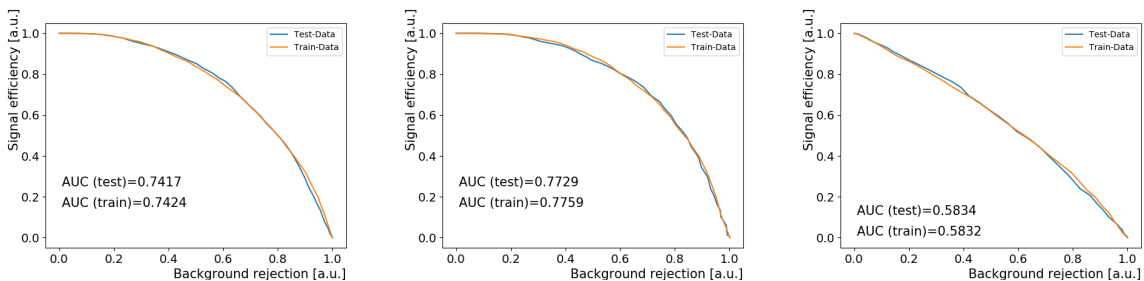

Figure 8.40.: The ROC-curves for the classifier for FSR(W). On the left with respect to top radiation, in the middle with respect to ISR and on the right with respect to FSR(b) (model: 3 layers and 60/100/4 nodes).

9. Summary, conclusion and outlook

At the ATLAS experiment at the LHC, the $t\bar{t}\gamma$-process is studied because it gives access to the top-photon coupling which still needs to be investigated more deeply. For the AT-LAS detector it is not possible to tag the emitting particle of a photon 100% correctly for each event individually. Thus, the measurement of the top-photon coupling relies on the statistical methods to get a sample as pure as possible. Nowadays, usually multivariate analysis techniques (MVA) are used to do this task. One of them is neural networks (NNs), as used in this analysis. They use different input variables to optimise the separation between different categories by minimising a certain function, the loss function. Therefore it is absolutely necessary to know the labels, which means to know which event belongs to which category. This was the first problem that needed to be addressed in this analysis. MadGraph, which is also the event generator for the $t\bar{t}\gamma$-analysis within ATLAS, does not directly allow to give such labels to the events. As described in Chapter 6, changes had to be implemented to obtain a sample only with events where the top quark radiates the photon. A second sample was produced, where no top quark radiates a photon. Such an artificial separation is only possible by neglecting interference. This fact has to be taken into account when interpreting the results. It was also necessary to check whether the samples produced with the modified MadGraph version were reliable and that variables used in the further analysis are not affected through the neglected interference terms. In general, it was concluded that the samples are reliable, although a slight deviation was detected for the ΔR distance between the photon and the nearest b quark in case that an on-shell top quark or an off-shell b quark were responsible for the photon emission. Nevertheless, this variable was later used in the training of NNs, since no deviations for this variable were seen for other mother particles of the photon. But a more concrete study of this effect might be useful, which could then lead to an exclusion of this variable for future NN trainings.

The first steps in analysing the sample were shape comparisons of different variables to look for those which allow a good separation between the four defined categories (photons from top quarks (top radiation), initial state radiation (ISR), radiative decay from b quarks (FSR(b)) and from W bosons or their decay products (FSR(W)). The most im-

pressive result was the good separation between events from top radiation and ISR on the one hand and FSR(b) and FSR(W) on the other hand by calculating invariant masses. This effect is related to when the photon is emitted (before or after the top decay) as it affects the particles needed to reconstruct the top quark. The strong separation of around 50% then also dominated the trainings of NNs at parton level. This led to a very good separation between top radiation and ISR versus FSR(b) and FSR(W). But as a consequence, those categories could not be separated well. In case of FSR(b) and FSR(W), this is not a problem, as both processes are considered to be background with respect to top radiation. Separating the ISR events from the signal different strategies improved the results but only within a few percentage points. To obtain a better separation between these two categories, a variable with separation similar to the invariant masses is needed. Such a variable was not found.

Although the results of the parton level studies give good insight into the kinematic behaviour of the $t\bar{t}\gamma$-process, they are not directly transferable when it comes to analysing real data. The samples must be studied at reconstruction level. Due to limited computational resources, only 165,000 events for the top-only sample and 160,000 for the no-top sample were available. To check consistency they were compared to the nominal $t\bar{t}\gamma$-samples. Basically, the results matched the expectations but a discrepancy was detected in the photon p_T distributions. It could be explained with a different p_T cut in the sample production with MadGraph that was applied for my samples and the official $t\bar{t}\gamma$-samples. The results obtained at reconstruction level cannot compete with those on parton level, but they clearly show that even on reconstruction level it is possible to increase the purity of the selected events with respect to top radiation events.

How can these results be interpreted and used in the future? First of all it could be shown that, although the samples were produced in an unphysical way (neglecting interference terms), they are usable for physics analysis, but in a limited way. As long as variable distributions are not sensitive to the interference terms, those distributions can be used in MVAs. This effect must be cross-checked for every single variable. The procedure to do so is not completely free from further assumptions. One has to combine both samples, which is done by adding the distributions using weights, reflecting the individual cross sections. The values used for these weights were the ones given by MadGraph. But these were produced neglecting the interference terms. Therefore, adding them up does not gives the cross section weight of a physical sample. It is therefore necessary that the impact of the interference term is small, so that the physical samples are approximated as much as possible. In this analysis, the distributions of the physical sample and those of the combined unphysical ones do agree within statistical uncertainties. As long as this

is the case, one should be able to claim that neglecting the interference terms in order to get the independent distributions of the processes that interfere is acceptable. This includes the assumption that correlations between variables do not change.

In the training it is very important that no overtraining occurs. To show this, ROC-curves are used. If no overtraining is observed, the next question is how the output values of the MVA method are used? There is no unique way to do so and the optimal one has to be found. In this analysis, the classification was done by assigning the category with the highest output value. But also modifications were presented. The relevant plot in the end is the confusion plot, which shows how many events from one category were classified to be from which category. The optimal interpretation depends also on the trade-off between statistics and improvement of the signal to background ratio. Only if all these steps are done precisely, are the results reliable.

This analysis was the first step to show that it is possible to separate processes whose Feynman diagrams do interfere. Before a MVA (in this case a NN) can be implemented into the full physics analysis, further studies need to be done. First of all higher statistics on reconstruction level are necessary. Furthermore, the used samples neglected τ-decays. The presence of those decays must be implemented as well. The usage of more variables suggests that the training should be done separately for dilepton and lepton+jets channel as the event topology differs. This would help, as more specific variables could be used. The most powerful variables at parton level need top quark reconstruction to use them on reconstruction level. KLFitter is optimised for $t\bar{t}$ pairs and does not account for the photon. Therefore, an algorithm to reconstruct $t\bar{t}\gamma$ events is needed. Changing the separation strategy might also help. One could first separate FSR(b) and FSR(W) from the other two categories and then do separation between top radiation and ISR. And last, but not least, one needs to quantify the systematic uncertainties, which includes independent cross check of the NN with samples from other generators and other showering algorithms. To use a neural network to measure the top-photon coupling it is also necessary to quantify the sensitivity of such a network to variations of the top-photon coupling. This means to study the outcome of the neural network with samples where the top-photon coupling was modified. When the network is then applied to real data one can perform hypothesis tests to see which top-photon coupling fits the results best.

As a final conclusion one can say that the first step to separate top radiation from the other prompt processes that so far are considered to be signal, is made, but further studies are necessary.

A. Modifications in MadGraph

Three files were modified[1]. The exact changes in each of these files are given below:

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

$$GC\_109 = Coupling(name =' GC\_109',$$
$$value =' (2 * ee * complex(0, 1))/3.',$$
$$order = 'TGA' : 1)$$

  - This change introduces a copy of the QED coupling under the new label TGA.

- In coupling_orders.py add:

$$TGA = CouplingOrder(name =' TGA',$$
$$expansion\_order = 99,$$
$$hierarchy = 2)$$

  - This change tells MadGraph that a label named TGA exists.

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

$$V\_134 = Vertex(name =' V\_134',$$
$$particles = [P.t\_\_\_tilde\_\_\_, P.t, P.a],$$
$$color = ['Identity(1, 2)'],$$
$$lorentz = [L.FFV1],$$
$$couplings = (0, 0) : C.GC\_109)$$

  - This change defines that TGA is the coupling between the top quark and the photon.

---

[1]The changes were applied in the following directory: $\sim /MadGraph/models/sm.$

B. Control plots for modified MadGraph
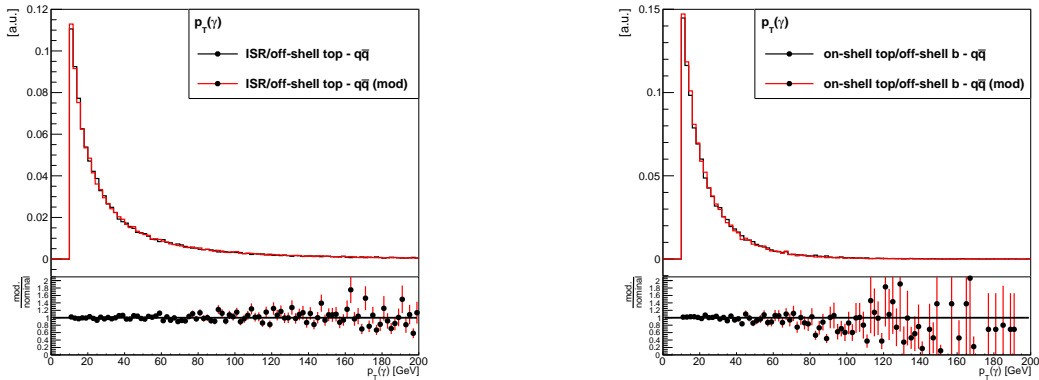
Figure B.1.: p_T-distributions of the photon for nominal (black) and modified (red) Mad-Graph version for both categories containing ambiguities. All distributions are scaled. The hight of each bin gives the relative amount of events in percent.
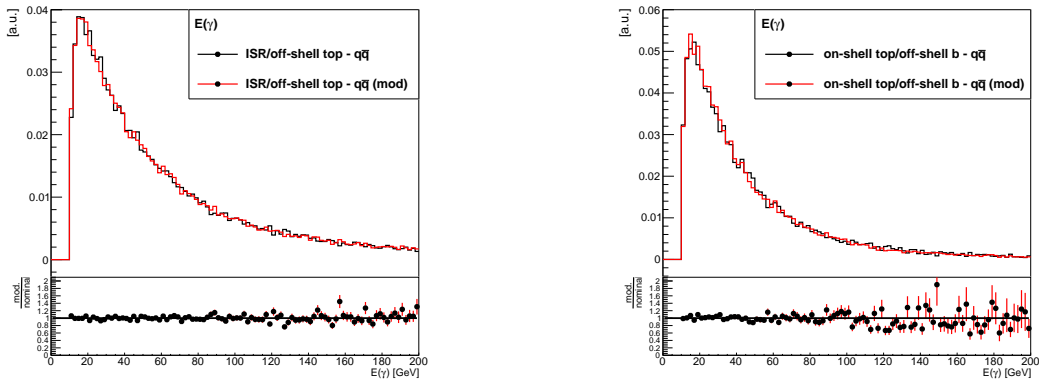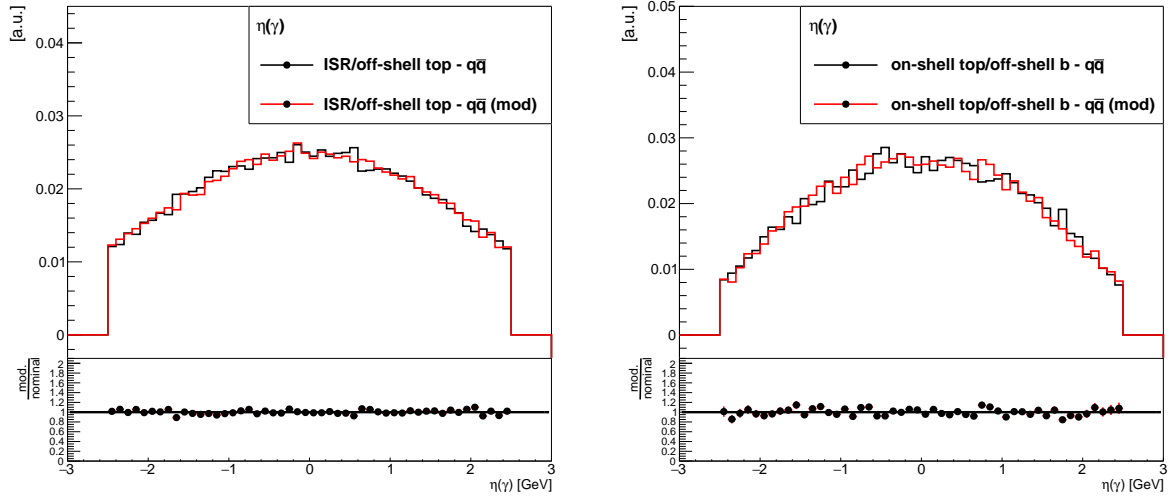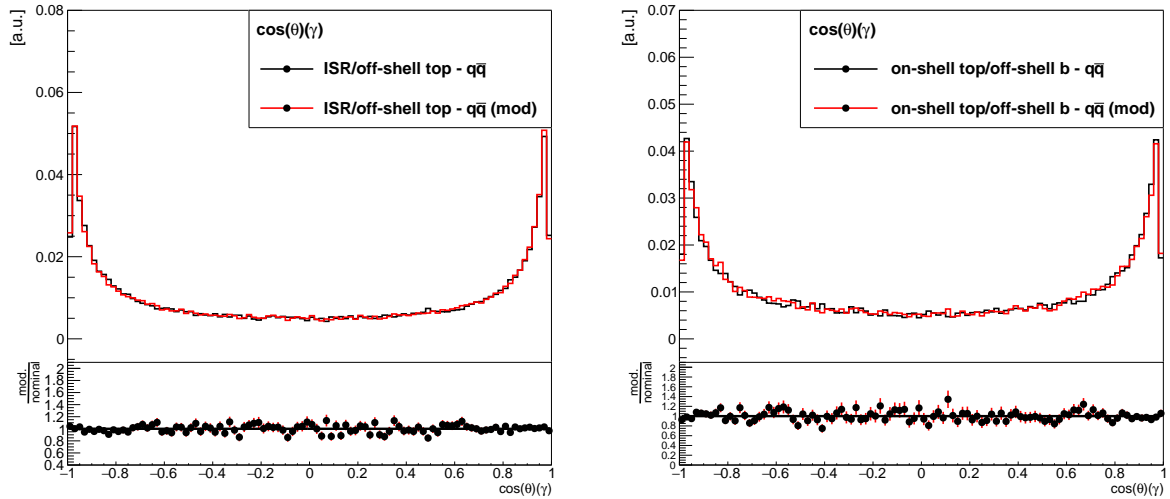

Figure B.2.: Energy distributions of the photon for nominal (black) and modified (red) MadGraph version for both categories containing ambiguities. All distributions are scaled. The hight of each bin gives the relative amount of events in percent.
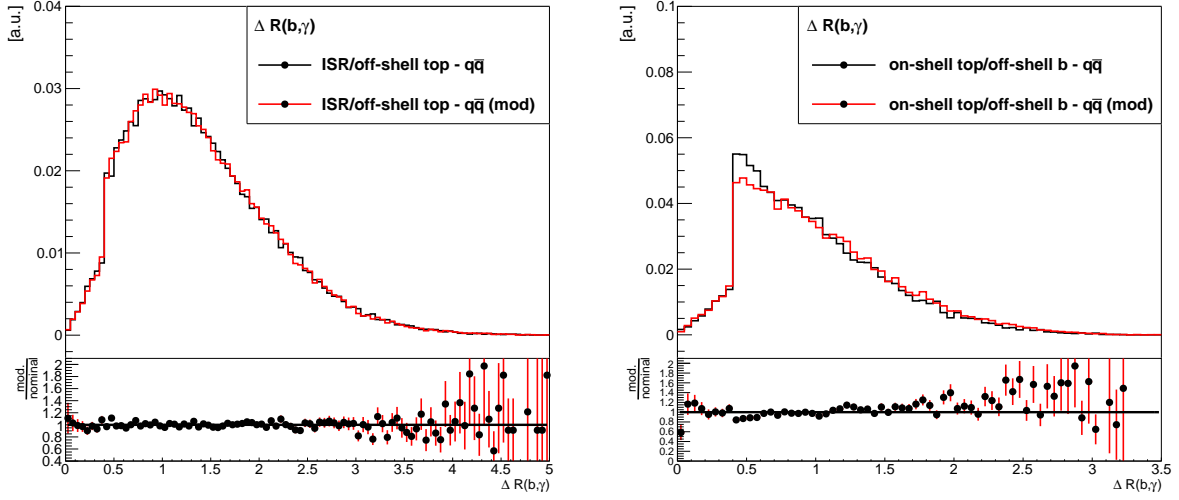
Figure B.3.: Distributions for the pseudorapidity of the photon for nominal (black) and modified (red) MadGraph version for both categories containing ambiguities. All distributions are scaled. The hight of each bin gives the relative amount of events in percent.
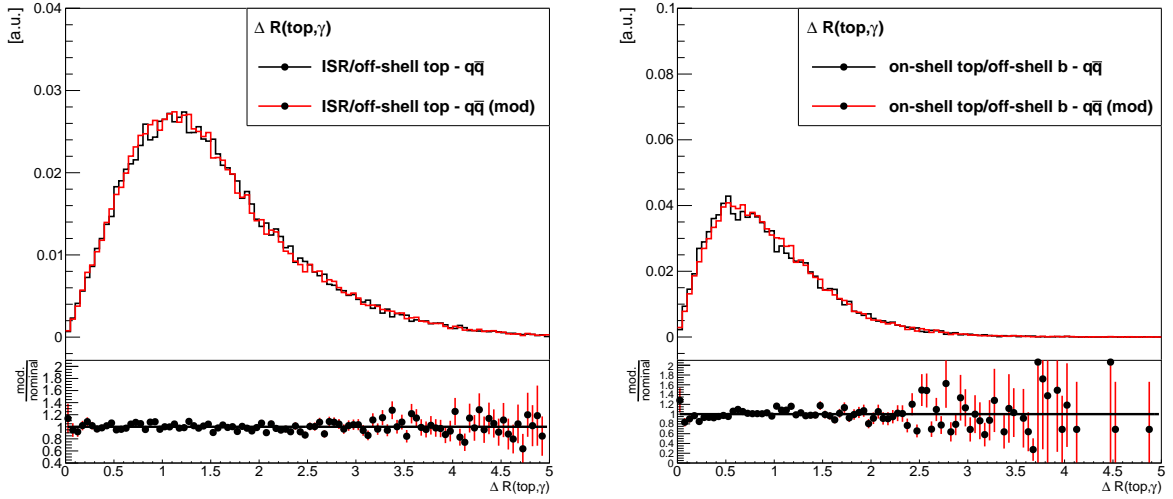

Figure B.4.: Distributions for $\cos(\theta)$ of the photon for nominal (black) and modified (red) MadGraph version for both categories containing ambiguities. All distributions are scaled. The hight of each bin gives the relative amount of events in percent.

Figure B.5.: Distributions for ΔR between photon and nearest b quark for nominal (black) and modified (red) MadGraph version for both categories containing ambiguities. All distributions are scaled. The hight of each bin gives the relative amount of events in percent.

Figure B.6.: Distributions for ΔR between photon and nearest top quark for nominal (black) and modified (red) MadGraph version for both categories containing ambiguities. All distributions are scaled. The hight of each bin gives the relative amount of events in percent.
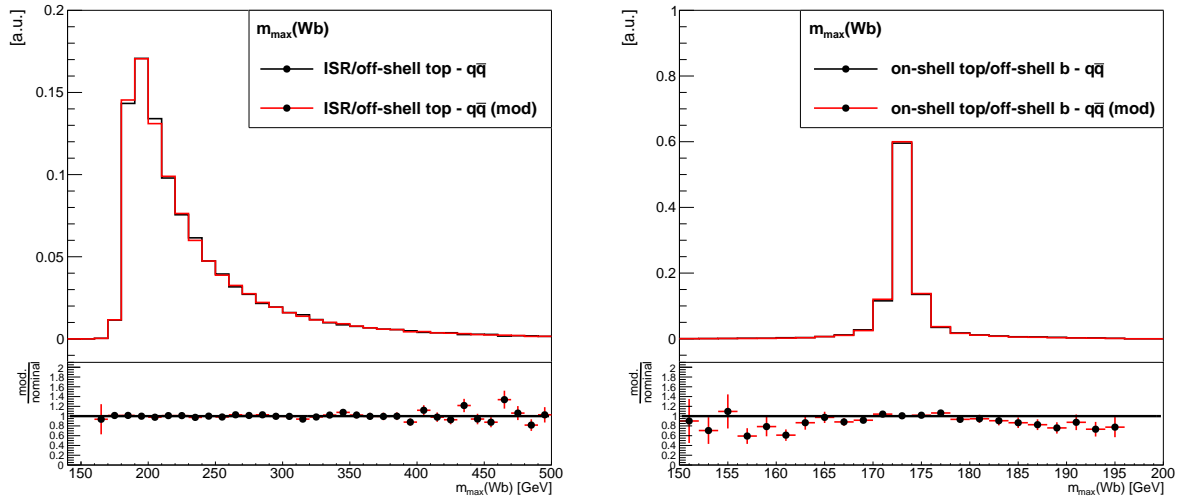
Figure B.7.: Distributions for the lower invariant mass of both Wb systems for nominal (black) and modified (red) MadGraph version for both categories containing ambiguities. All distributions are scaled. The hight of each bin gives the relative amount of events in percent.
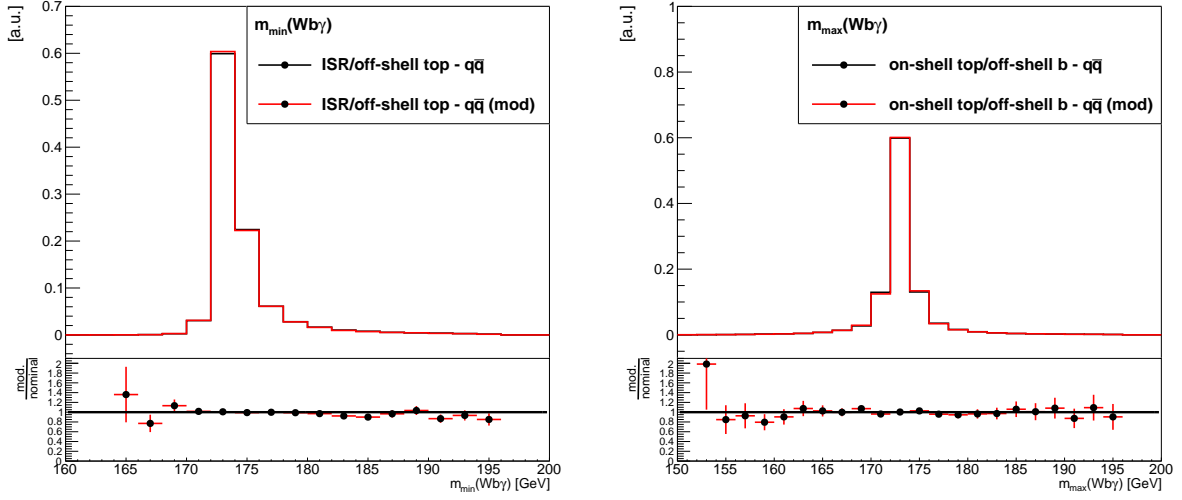

Figure B.8.: Distributions for the higher invariant mass of both Wb systems for nominal (black) and modified (red) MadGraph version for both categories containing ambiguities. All distributions are scaled. The hight of each bin gives the relative amount of events in percent.

Figure B.9.: Distributions for the lower invariant mass of both Wbγ systems for nominal (black) and modified (red) MadGraph version for both categories containing ambiguities. All distributions are scaled. The hight of each bin gives the relative amount of events in percent.
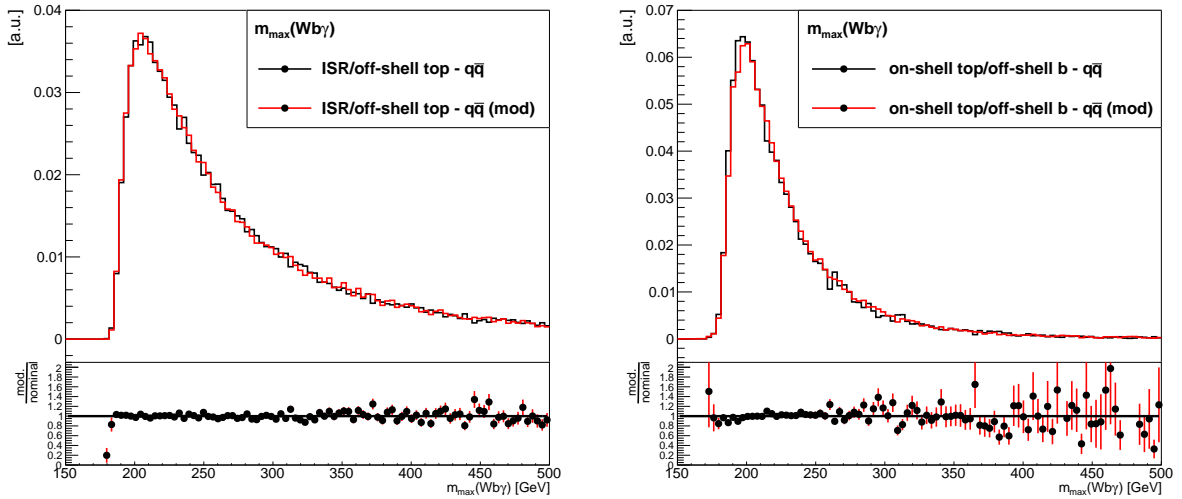

Figure B.10.: Distributions for the higher invariant mass of both Wbγ systems for nominal (black) and modified (red) MadGraph version for both categories containing ambiguities. All distributions are scaled. The hight of each bin gives the relative amount of events in percent.

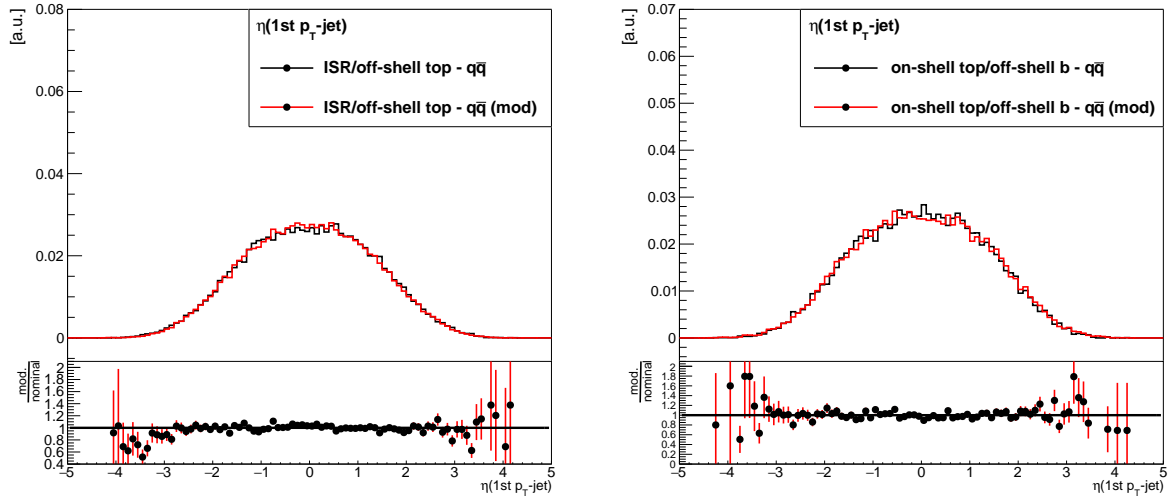**Figure B.11.:** η distributions for the leading jet in p_T for nominal (black) and modified (red) MadGraph version for both categories containing ambiguities. All distributions are scaled. The hight of each bin gives the relative amount of events in percent.

Figure B.12.: η distributions for the 2nd-leading jet in p_T for nominal (black) and modified (red) MadGraph version for both categories containing ambiguities. All distributions are scaled- The hight of each bin gives the relative amount of events in percent.
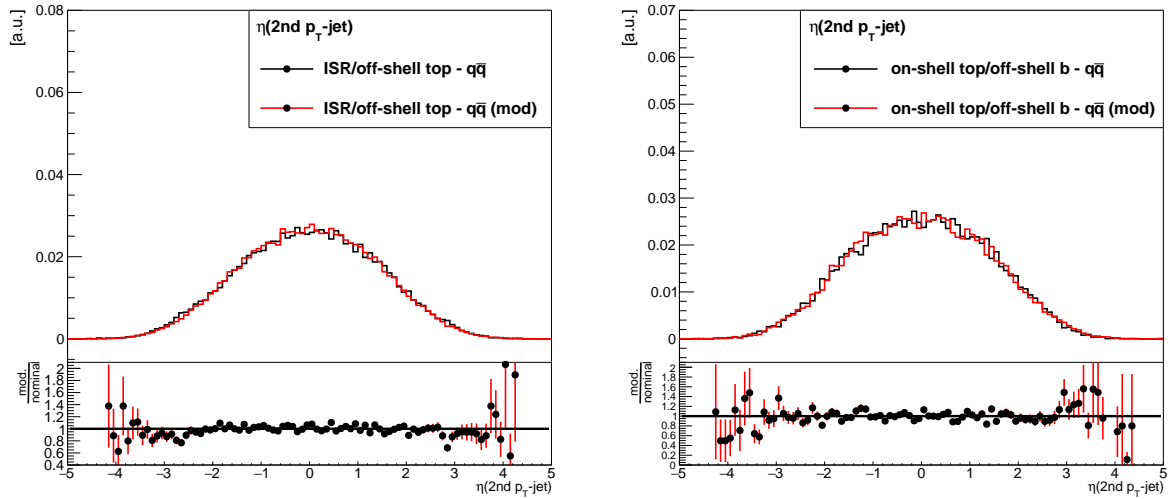
Figure B.13.: η distributions for the most central jet for nominal (black) and modified (red) MadGraph version for both categories containing ambiguities. All distributions are scaled. The hight of each bin gives the relative amount of events in percent.
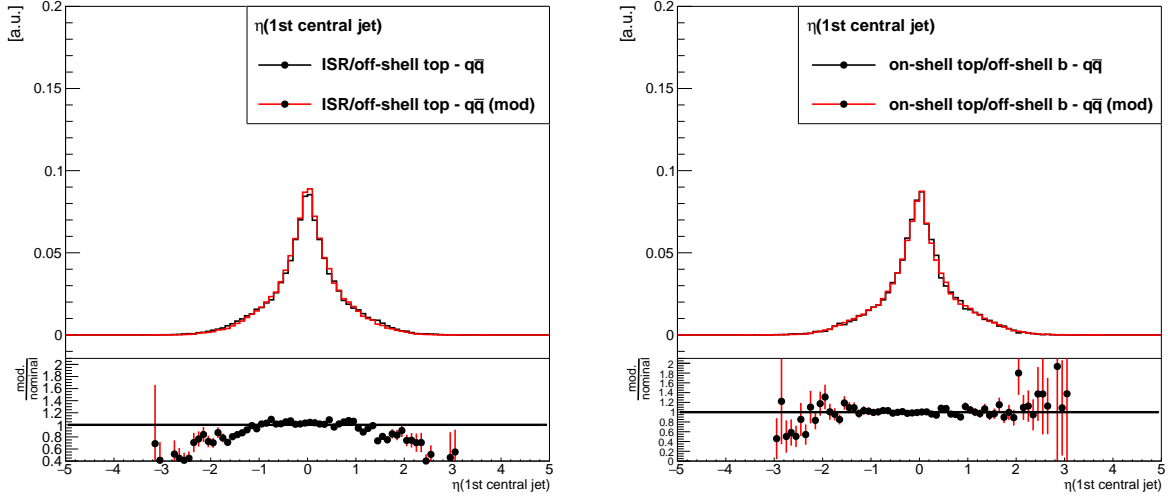

Figure B.14.: η distributions for the 2nd most central jet for nominal (black) and modified (red) MadGraph version for both categories containing ambiguities. All distributions are scaled. The hight of each bin gives the relative amount of events in percent.
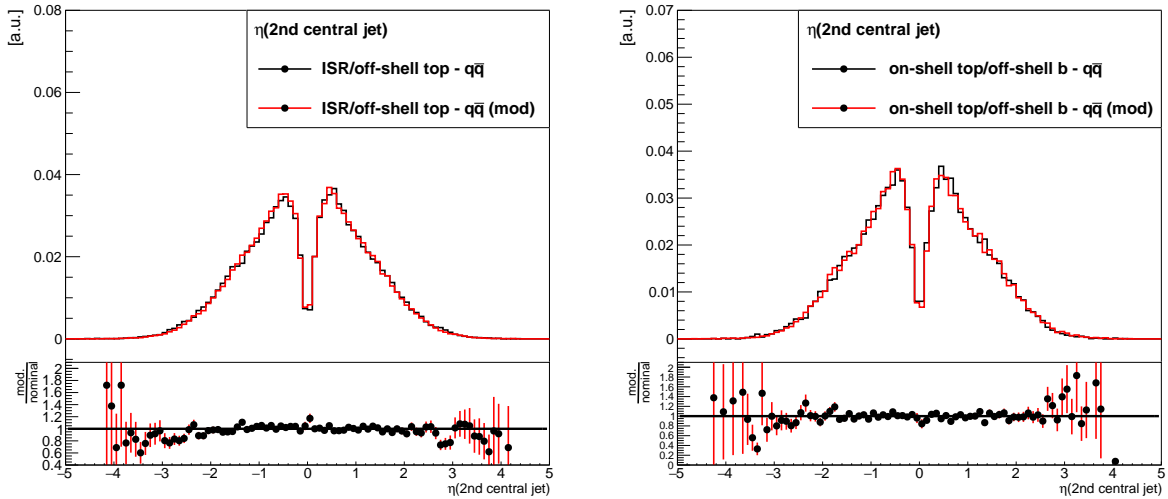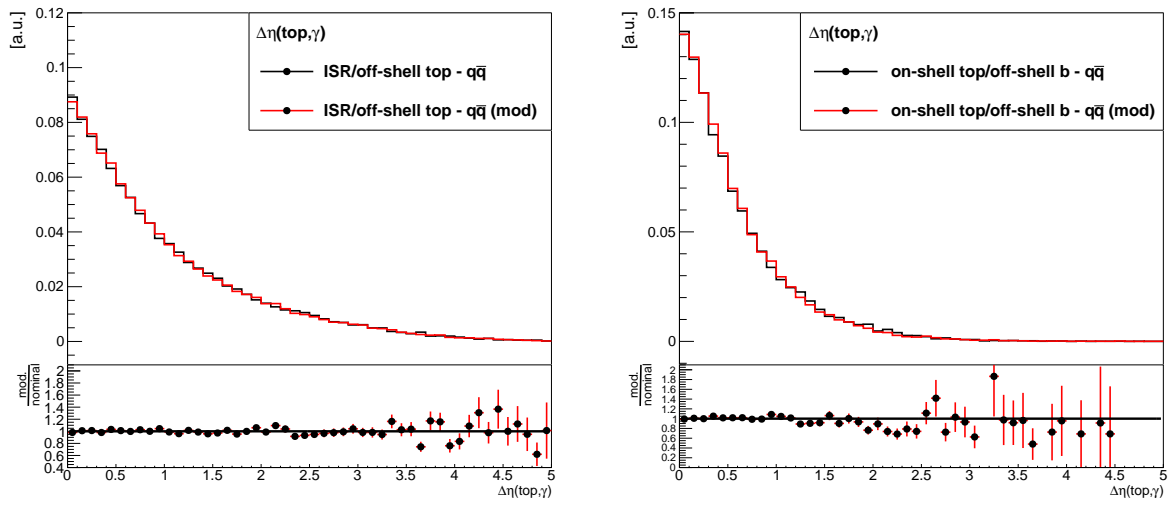
Figure B.15.: Distributions for $\Delta\eta$ between photon and nearest top quark for nominal (black) and modified (red) MadGraph version for both categories containing ambiguities. All distributions are scaled. The hight of each bin gives the relative amount of events in percent.

C. Cutflow

Figure C.1.: Cutflow plots for the e+jets channel for 2015 (left) and 2016 (right) for the top-only sample and the first set of AnalysisTop cuts.

Figure C.2.: Cutflow plots for the μ+jets channel for 2015 (left) and 2016 (right) for the top-only sample and the first set of AnalysisTop cuts.

C. Cutflow

Figure C.3.: Cutflow plots for the eμ channel for 2015 (left) and 2016 (right) for the top-only sample and the first set of AnalysisTop cuts.

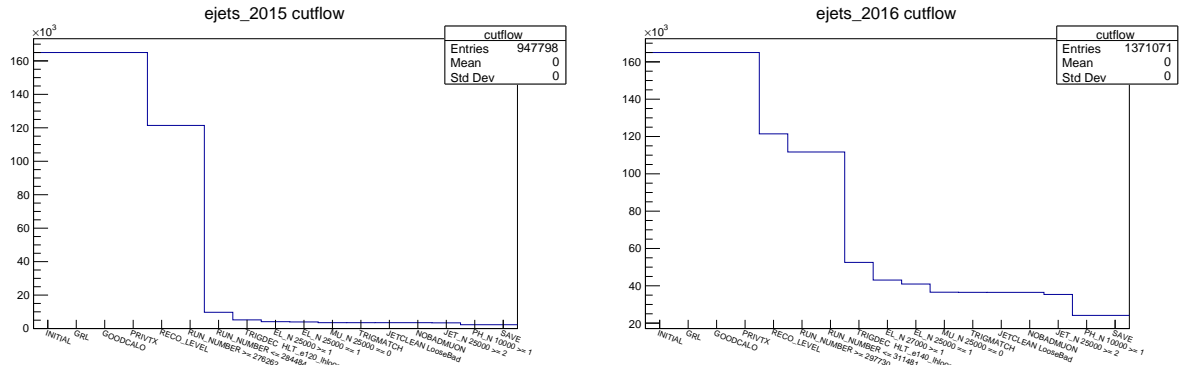***Figure C.4.:*** Cutflow plots for the ee channel for 2015 (left) and 2016 (right) for the top-only sample and the first set of AnalysisTop cuts.

Figure C.5.: Cutflow plots for the μμ channel for 2015 (left) and 2016 (right) for the top-only sample and the first set of AnalysisTop cuts.

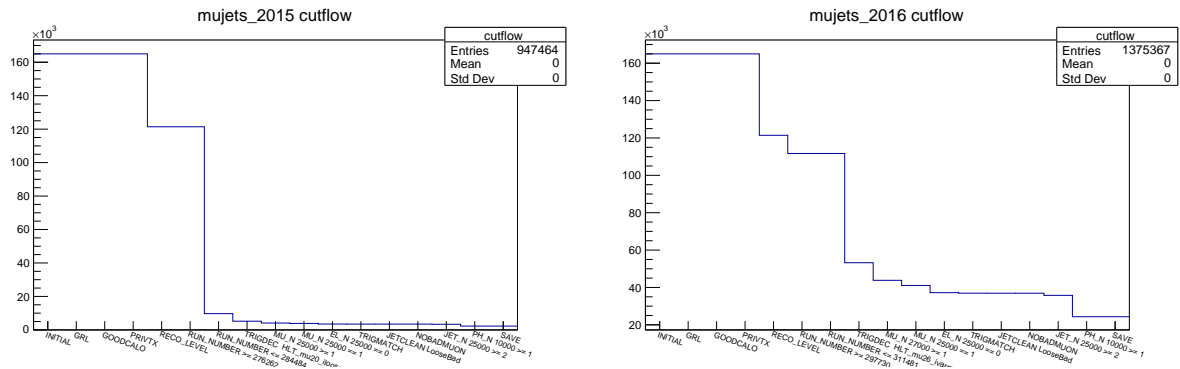***Figure C.6.:*** Cutflow plots for the e+jets channel for 2015 (left) and 2016 (right) for the no-top sample and the first set of AnalysisTop cuts.

Figure C.7.: Cutflow plots for the μ+jets channel for 2015 (left) and 2016 (right) for the no-top sample and the first set of AnalysisTop cuts.
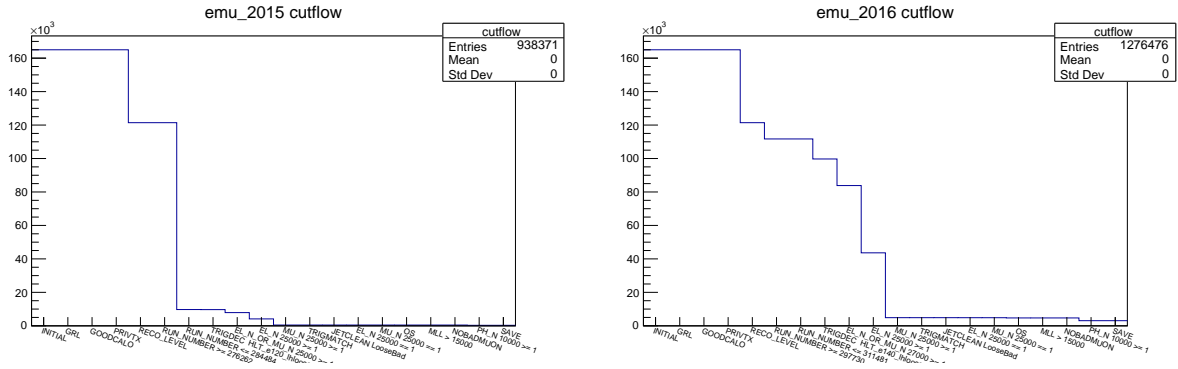

Figure C.8.: Cutflow plots for the eμ channel for 2015 (left) and 2016 (right) for the no-top sample and the first set of AnalysisTop cuts.
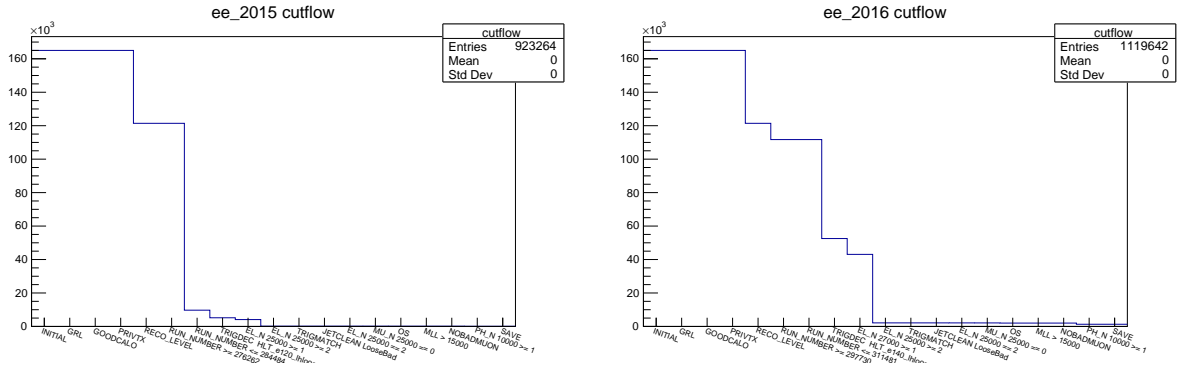
Figure C.9.: Cutflow plots for the ee channel for 2015 (left) and 2016 (right) for the no-top sample and the first set of AnalysisTop cuts.
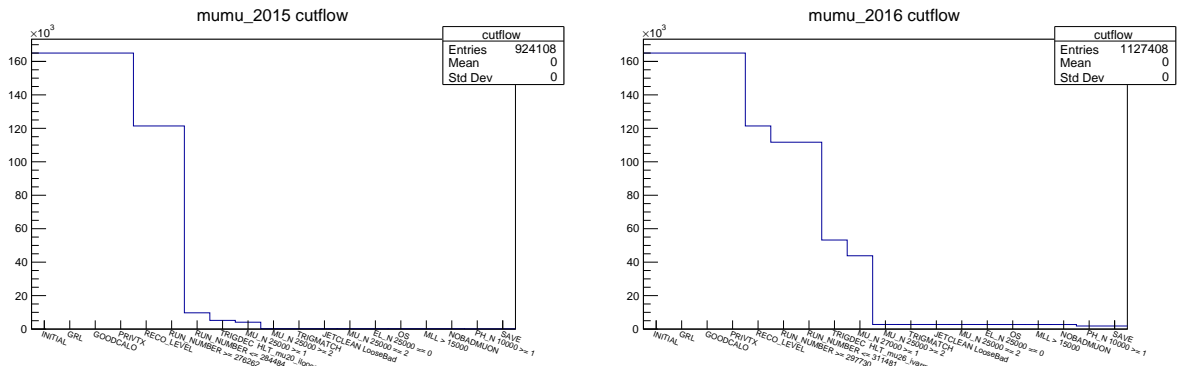

Figure C.10.: Cutflow plots for the $\mu\mu$ channel for 2015 (left) and 2016 (right) for the no-top sample and the first set of AnalysisTop cuts.

Figure C.11.: Cutflow plots for the e+jets channel for 2015 (left) and 2016 (right) for the top-only sample and the second set of AnalysisTop cuts.

Figure C.12.: Cutflow plots for the μ+jets channel for 2015 (left) and 2016 (right) for the top-only sample and the second set of AnalysisTop cuts.
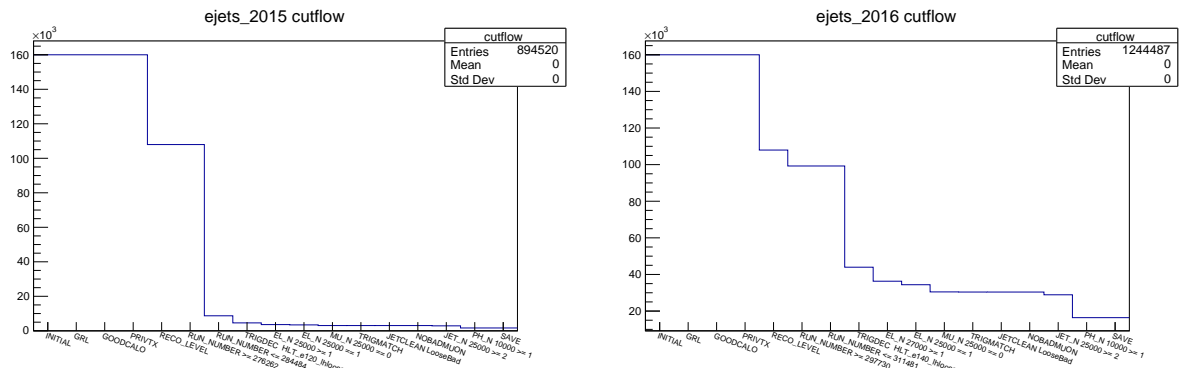

Figure C.13.: Cutflow plots for the e+jets channel for 2015 (left) and 2016 (right) for the no-top sample and the second set of AnalysisTop cuts.

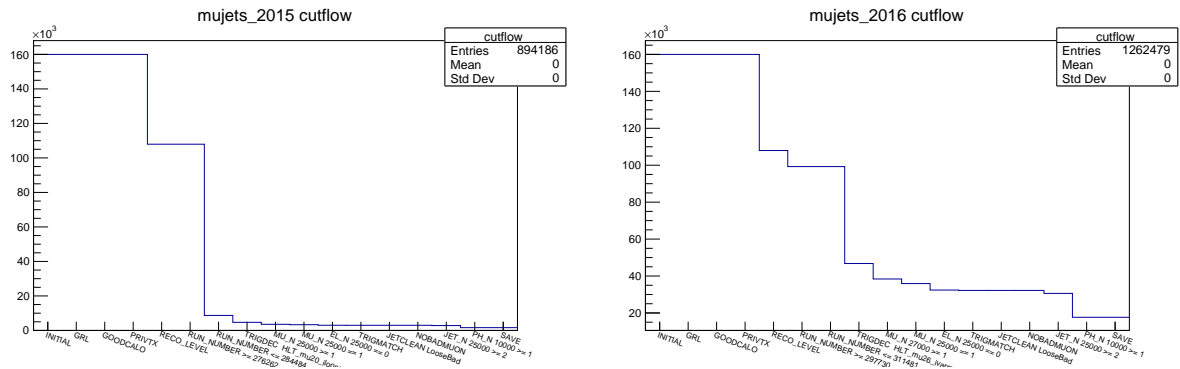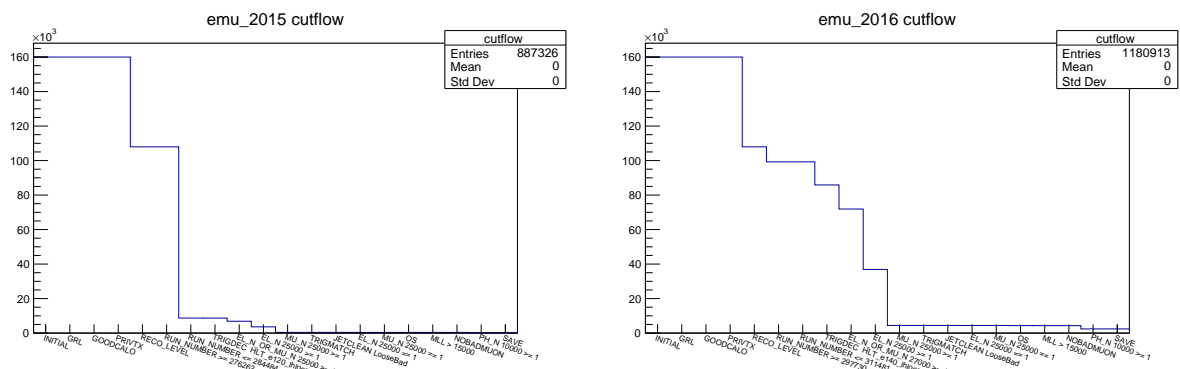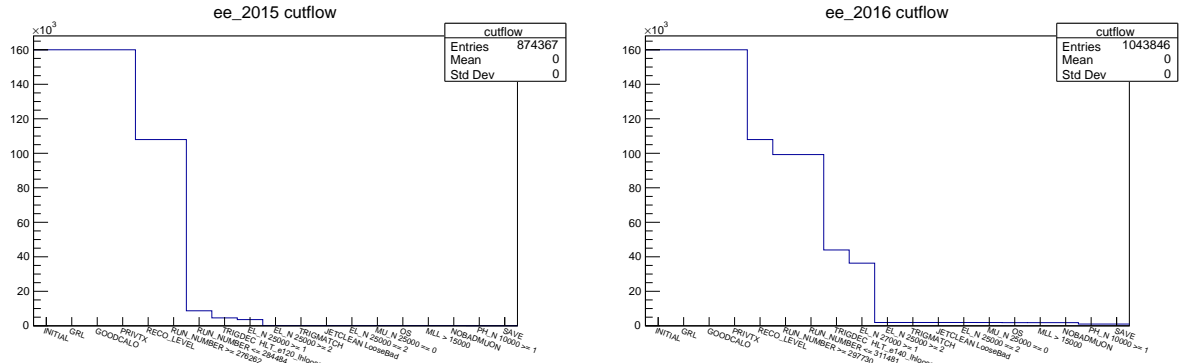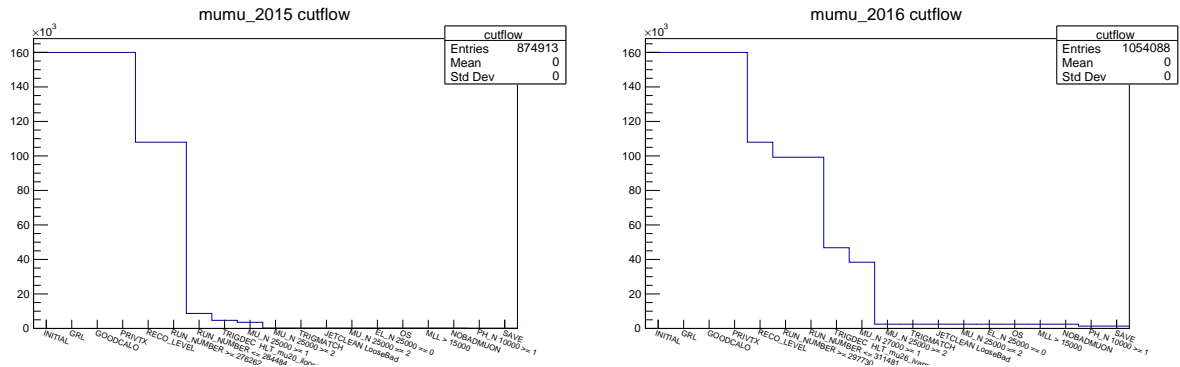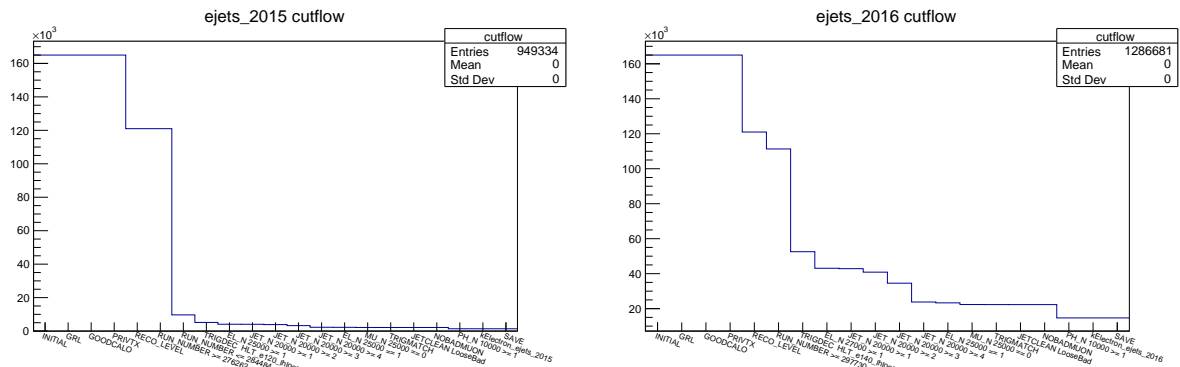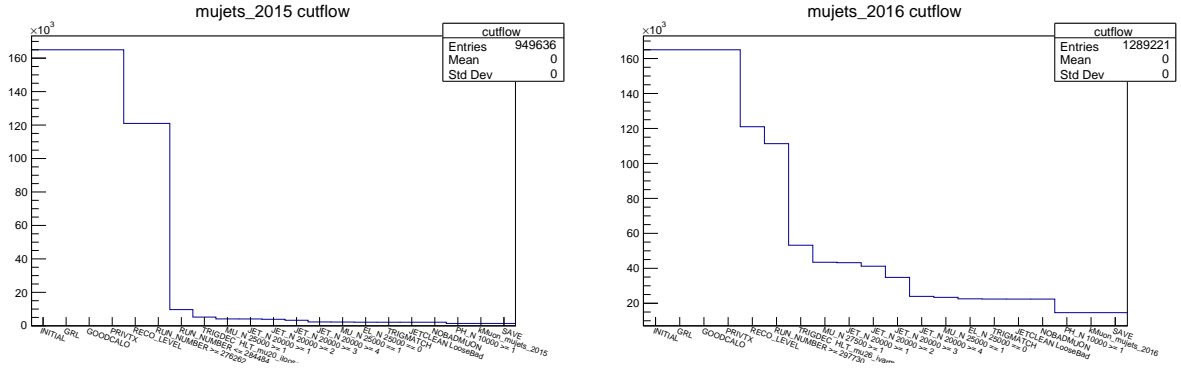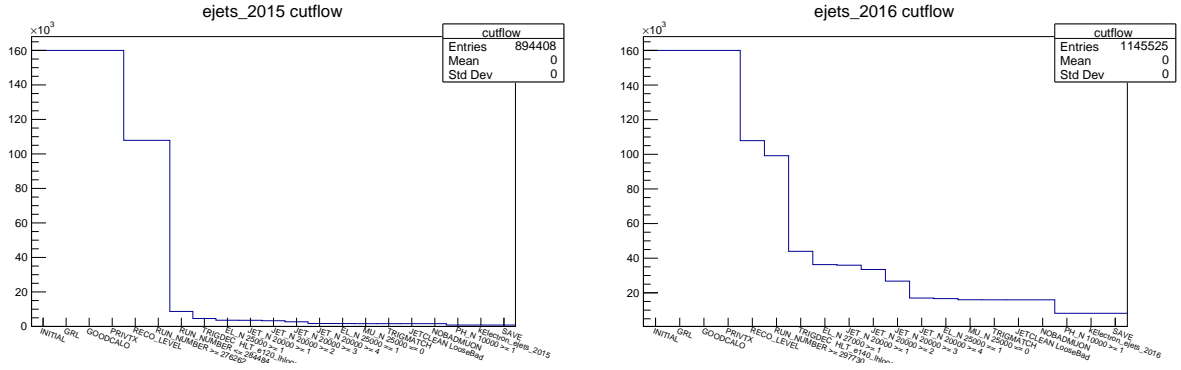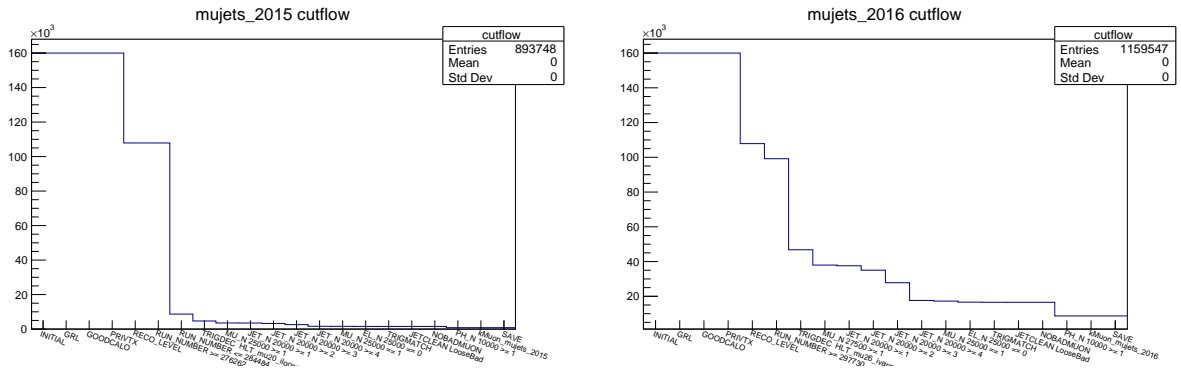***Figure C.14.:*** Cutflow plots for the μ+jets channel for 2015 (left) and 2016 (right) for the no-top sample and the second set of AnalysisTop cuts.

D. Tested neural network architectures

Name	No. of layers	Nodes per layer	Batch size	Epochs[1]	Frame	Weights	Further information
3l_16_8_4	3	16/8/4	1000	500	$t\bar{t}$/Lab	none	-
3l_8_16_4	3	8/16/4	1000	500	$t\bar{t}$/Lab	none	-
4l_32_16_8_4	4	32/16/8/4	1000	500	$t\bar{t}$/Lab	none	-
4l_16_32_8_4	4	16/32/8/4	1000	500	$t\bar{t}$/Lab	none	-
5l_64_32_16_8_4	5	64/32/16/8/4	1000	500	$t\bar{t}$/Lab	none	-
5l_32_64_16_8_4	5	32/64/16/8/4	1000	500	$t\bar{t}$/Lab	none	-
6l_128_64_32_16_8_4	6	128/64/32/16/8/4	1000	500	$t\bar{t}$/Lab	none	-
6l_64_128_32_16_8_4	6	64/128/32/16/8/4	1000	500	$t\bar{t}$/Lab	none	-
3l_16_8_4	3	16/8/4	1000	500	$t\bar{t}$	amount[2]	-
3l_16_8_4	3	16/8/4	1000	500	$t\bar{t}$	xsec[3]	-
3l_16_8soft_4	3	16/8/4	1000	500	$t\bar{t}$	none	activation function for second last layer set to softmax[4]
3l_16_8soft_4	3	16/8/4	1000	500	$t\bar{t}$	amount	activation function for second last layer set to softmax
3l_16_8soft_4	3	16/8/4	1000	500	$t\bar{t}$	xsec	activation function for second last layer set to softmax
3l_60_100_4	3	60/100/4	1000	500	$t\bar{t}$	none	-
6l_60_100_150_100_50_4	6	60/100/150/100/50/4	1000	500	$t\bar{t}$	none	-
3l_16_8_4bn	3	16/8/4	1000	500	$t\bar{t}$	none	batch normalization used between layers
6l_12_4_8_16_8_4bn	6	16/8/4	1000	500	$t\bar{t}$	none	batch normalization used between layers
3l_16_8_4	3	16/8/4	1000	500	Lab	amount	-
3l_16_8_4	3	16/8/4	1000	500	Lab	xsec	-

Table D.1.: Information about the different NN architectures that were tested. The first eight rows are the networks that were tested systematically with a basic set of options. For the other networks some changes of the basic options were tested.

[1]If during 100 epochs no improvement was archived, the training was stoped. The number presented here refers to the maximum number of possible training epochs.

[2]Amount means that the background events (N is the total number of events) got weights w so that $w \cdot N = 958832$, which is the number of events for the signal (top radiation).

[3]xsec means that background events were weighted with 1.35 and signal events with 0.4. These numbers represent the calculated cross section in the samples. It has to be reminded that these cross sections are calculated for samples that neglect interference terms. So they are just approximations to describe the ratio between signal and background

[4]Normally only the last layer used softmax as its activation function. All other layers used the rectifier linear unit (relu)

Name	No. of layers	Nodes per layer	Batch size	Epochs	Frame	Weights	Further information
3l_16_8_4	3	16/8/4	500	1000	Lab	none	cut at 0.2 for ISR classifier
3l_16_8_4	3	16/8/4	500	1000	Lab	none	cut at 0.4 for ISR classifier
3l_16_8_4	3	16/8/4	500	1000	Lab	none	cut at 0.6 for ISR classifier
3l_16_8_4	3	16/8/4	500	1000	Lab	none	cut at 0.8 for ISR classifier
3l_16_8_4	3	16/8/4	500	1000	Lab	xsec	cut at 0.2 for ISR classifier
3l_16_8_4	3	16/8/4	500	1000	Lab	xsec	cut at 0.4 for ISR classifier
3l_16_8_4	3	16/8/4	500	1000	Lab	xsec	cut at 0.6 for ISR classifier
3l_16_8_4	3	16/8/4	500	1000	Lab	xsec	cut at 0.8 for ISR classifier
3l_16_8_4	3	16/8/4	500	1000	Lab	amount	cut at 0.2 for ISR classifier
3l_16_8_4	3	16/8/4	500	1000	Lab	amount	cut at 0.4 for ISR classifier
3l_16_8_4	3	16/8/4	500	1000	Lab	amount	cut at 0.6 for ISR classifier
3l_16_8_4	3	16/8/4	500	1000	Lab	amount	cut at 0.8 for ISR classifier
3l_16_8_4	3	16/8/4	1000	500	mixed	none	-
3l_8_16_4	3	8/16/4	1000	500	mixed	none	-
4l_32_16_8_4	4	32/16/8/4	1000	500	mixed	none	-
4l_16_32_8_4	4	16/32/8/4	1000	500	mixed	none	-
5l_64_32_16_8_4	5	64/32/16/8/4	1000	500	mixed	none	-
5l_32_64_16_8_4	5	32/64/16/8/4	1000	500	mixed	none	-
6l_128_64_32_16_8_4	6	128/64/32/16/8/4	1000	500	mixed	none	-
6l_64_128_32_16_8_4	6	64/128/32/16/8/4	1000	500	mixed	none	-
6l_60_100_150_100_50_4	6	60/100/150/100/50/4	1000	500	mixed	none	-
3l_16_8_4	3	16/8/4	500	1000	Lab	xsec	using 15 variables
3l_16_8_4	3	16/8/4	500	1000	Lab	none	using 7 variables; no invariant masses

Table D.2.: Information about the different NN architectures that were used to improve separation between ISR and top radiation. The horizontal lines refer to the different purposes of the networks in the order of the sections 7.3.1 to 7.3.4.

Bibliography

[1] J. J. Thomson, *Cathode Rays*, Phil. Mag. **44(269)**, 293 (1897)

[2] H. Geiger, E. Marsden, *On the Scattering of α-Particles by Matter*, Proc. Roy. Soc. Lond. A **82(557)**, 495 (1909)

[3] E. Rutherford, *The Scattering of α and β Particles by Matter and the Structure of the Atom*, Phil. Mag. **21**, 669 (1911)

[4] J. J. Thomson, *On the structure of the atom: an investigation of the stability and periods of oscillation of a number of corpuscles arranged at equal intervals around the circumference of a circle; with application of the results to the theory of atomic structure*, Phil. Mag. **7(39)**, 237 (1904)

[5] C. D. Anderson, *The Positive Electron*, Phys. Rev. **43**, 491 (1933)

[6] M. Gell-Mann, *Symmetries of Baryons and Mesons*, Phys. Rev. **125**, 1067 (1962)

[7] Y. Ne'eman, *Derivation of strong interactions from a gauge invariance*, Nuc. Phys. **26(2)**, 222 (1961)

[8] M. Gell-Mann, *A Schematic Model of Baryons and Mesons*, Phys. Lett. **8**, 214 (1964)

[9] G. Zweig, *An SU_3 model for strong interaction symmetry and its breaking; Version 1*, Technical Report CERN-TH-401, CERN, Geneva (1964)

[10] S. Weinberg, *The Making of the Standard Model*, Eur. Phys. J. C **34**, 5 (2003)

[11] P. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. **12(2)**, 132 (1964)

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

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

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

[15] T. W. B. Kibble, *Symmetry Breaking in Non-Abelian Gauge Theories*, Phys. Rev. **155**, 1554 (1967)

[16] The 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. B **716(1)**, 1 (2012)

[17] The CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B **716(1)**, 30 (2012)

[18] The CDF Collaboration, *Observation of Top Quark Production in $\overline{p}p$ Collisions with the Collider Detector at Fermilab*, Phys. Rev. Lett. **74**, 2626 (1995)

[19] The D0 Collaboration, *Search for High Mass Top Quark Production in $p\overline{p}$ Collisions at $\sqrt{s} = $ 1.8 TeV*, Phys. Rev. Lett. **74**, 2422 (1995)

[20] The 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 **2017(11)**, 86 (2017)

[21] The SNO Collaboration, *Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory*, Phys. Rev. Lett. **89**, 011301 (2002)

[22] P. A. M. Dirac, *The Quantum Theory of the Electron*, Proc. Roy. Soc. Lond. A **117(778)**, 610 (1928)

[23] C. S. Wu, et al., *Experimental Test of Parity Conservation in Beta Decay*, Phys. Rev. **105**, 1413 (1957)

[24] M. Goldhaber, L. Grodzins, A. W. Sunyar, *Helicity of Neutrinos*, Phys. Rev. **109**, 1015 (1958)

[25] S. L. Glashow, *Partial-symmetries of weak interactions*, Nuc. Phys. **22(4)**, 579 (1961)

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

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

[28] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10**, 531 (1963)

[29] S. L. Glashow, J. Iliopoulos, L. Maiani, *Weak Interactions with Lepton-Hadron Symmetry*, Phys. Rev. D **2**, 1285 (1970)

[30] LHCb Collaboration, *Observation of J/ψ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays*, Phys. Rev. Lett. **115**, 072001 (2015)

[31] C. A. Baker, et al., *Improved Experimental Limit on the Electric Dipole Moment of the Neutron*, Phys. Rev. Lett. **97**, 131801 (2006)

[32] G. 't Hooft, *Renormalizable Lagrangians for massive Yang-Mills fields*, Nuc. Phys. B **35(1)**, 167 (1971)

[33] G. 't Hooft, M. Veltman, *Regularization and renormalization of gauge fields*, Nuc. Phys. B **44(1)**, 189 (1972)

[34] G. 't Hooft, M. Veltman, *Combinatorics of gauge fields*, Nuc. Phys. B **50(1)**, 318 (1972)

[35] D. J. Gross, F. Wilczek, *Ultraviolet Behavior of Non-Abelian Gauge Theories*, Phys. Rev. Lett. **30**, 1343 (1973)

[36] H. D. Politzer, *Reliable Perturbative Results for Strong Interactions?*, Phys. Rev. Lett. **30**, 1346 (1973)

[37] H. D. Politzer, *Asymptotic freedom: An approach to strong interactions*, Phys. Rep. **14(4)**, 129 (1974)

[38] C. Patrignani et al. (Particle Data Group), *Review of Particle Physics*, Chin. Phys. C **40**, 100001 (2016 and 2017 update)

[39] V. C. Rubin, N. Thonnard, W. K. Ford, Jr., *Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 /R = 4kpc/ to UGC 2885 /R = 122 kpc/*, Astrophys. J. **238**, 471 (1980)

Bibliography

[40] A. D. Sakharov, *Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe*, Phys. Uspekhi **34(5)**, 392 (1991)

[41] P. Minkowski, $\mu \to e\gamma$ *at a rate of one out of 109 muon decays?*, Phys. Lett. B **67(4)**, 421 (1977)

[42] J. H. Christenson, et al., *Evidence for the 2π Decay of the K_2^0 Meson*, Phys. Rev. Lett. **13**, 138 (1964)

[43] S. W. Herb, et al., *Observation of a Dimuon Resonance at 9.5 GeV in 400-GeV Proton-Nucleus Collisions*, Phys. Rev. Lett. **39**, 252 (1977)

[44] A. Czarnecki, K. Melnikov, *Two-loop QCD corrections to top quark width*, Nuc. Phys. B **544(3)**, 520 (1999)

[45] K. G. Chetyrkin, et al., *Second order QCD corrections to* $\Gamma(t \to Wb)$, Phys. Rev. D **60**, 114015 (1999)

[46] The ATLAS Collaboration, *Direct top quark decay width measurement in the $t\bar{t}$ lepton+jets channel at $\sqrt{s} = 8$ TeV with the ATLAS experiment*, Eur. Phys. J. C **78(2)**, 129 (2018)

[47] I. Bigi, et al., *Production and decay properties of ultra-heavy quarks*, Phys. Lett. B **181(1)**, 157 (1986)

[48] A. Hoecker, *Physics at the LHC Run-2 and Beyond*, CERN Yellow Reports: School Proceedings **5(0)**, 153 (2017)

[49] The ATLAS Collaboration, *Measurement of the production cross-section using $e\mu$ events with b-tagged jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Lett. B **761**, 136 (2016)

[50] The ATLAS Collaboration, *Measurement of lepton differential distributions and the top quark mass in $t\bar{t}$ production in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, Eur. Phys. J. C **77(11)**, 804 (2017)

[51] The ATLAS Collaboration, *Measurement of the W boson polarisation in $t\bar{t}$ events from pp collisions at $\sqrt{s} = 8$ TeV in the lepton + jets channel with ATLAS*, Eur. Phys. J. C **77(4)**, 264 (2017)

[52] D. Roy, *Looking for the charged Higgs Boson*, Mod. Phys. Lett A **19(24)**, 1813 (2004)

118

[53] J. Collins, D. Soper, *The Theorems of Perturbative QCD*, Annual Review of Nuclear and Particle Science **37(1)**, 383 (1987)

[54] L. W. Mo, Y. S. Tsai, *Radiative Corrections to Elastic and Inelastic* ep *and* up *Scattering*, Rev. Mod. Phys. **41**, 205 (1969)

[55] E. D. Bloom, et al., *High-Energy Inelastic $e-p$ Scattering at 6° and 10°*, Phys. Rev. Lett. **23**, 930 (1969)

[56] M. Breidenbach, et al., *Observed Behavior of Highly Inelastic Electron-Proton Scattering*, Phys. Rev. Lett. **23**, 935 (1969)

[57] G. Miller, et al., *Inelastic Electron-Proton Scattering at Large Momentum Transfers and the Inelastic Structure Functions of the Proton*, Phys. Rev. D **5**, 528 (1972)

[58] A. Bodek, et al., *The ratio of deep-inelastic e-n to e-p cross sections in the threshold region*, Phys. Lett. B **51(4)**, 417 (1974)

[59] E. M. Riordan, et al., *Extraction of $R = \frac{\sigma_L}{\sigma_T}$ from Deep Inelastic $e-p$ and $e-d$ Cross Sections*, Phys. Rev. Lett. **33**, 561 (1974)

[60] J. D. Bjorken, E. A. Paschos, *Inelastic Electron-Proton and γ-Proton Scattering and the Structure of the Nucleon*, Phys. Rev. **185**, 1975 (1969)

[61] R. P. Feynman, *Photon-Hadron Interactions*, W. A. Benjamin, New York (1972)

[62] J. D. Bjorken, *Asymptotic Sum Rules at Infinite Momentum*, Phys. Rev. **179**, 1547 (1969)

[63] C. G. Callan, D. J. Gross, *High-Energy Electroproduction and the Constitution of the Electric Current*, Phys. Rev. Lett. **22**, 156 (1969)

[64] E. Riordan, et al., *Tests of scaling of the proton electromagnetic structure functions*, Phys. Lett. B **52(2)**, 249 (1974)

[65] Y. L. Dokshitser, *Calculation of structure functions of deep-inelastic scattering and $e^{+-}e$ annihilation by perturbation theory in quantum chromodynamics*, JETP **73**, 1216 (1977)

[66] V. N. Gribov, L. N. Lipatov, *Deep inelastic e p scattering in perturbation theory*, Sov. J. Nucl. Phys. **15**, 438 (1972)

Bibliography

[67] G. Altarelli, G. Parisi, *Asymptotic Freedom in Parton Language*, Nucl. Phys. B **126**, 298 (1977)

[68] H. Abramowicz, et al. (ZEUS, H1), *Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data*, Eur. Phys. J. C **75(12)**, 580 (2015)

[69] J. Rojo, *Parton Distributions based on a Maximally Consistent Dataset*, Nuclear and Particle Physics Proceedings **273-275**, 2122 (2016), 37th International Conference on High Energy Physics (ICHEP)

[70] J. Rojo, et al., *The PDF4LHC report on PDFs and LHC data: results from Run I and preparation for Run II*, J. Phys. G **42(10)**, 103103 (2015)

[71] L. A. Harland-Lang, et al., *Parton distributions in the LHC era: MMHT 2014 PDFs*, Eur. Phys. J. C **75(5)**, 204 (2015)

[72] S. Dulat, et al., *New parton distribution functions from a global analysis of quantum chromodynamics*, Phys. Rev. D **93**, 033006 (2016)

[73] L. Maiani, L. Bonolis, *The LHC timeline: a personal recollection (1980–2012)*, Eur. Phys. J. H **42(4)**, 475 (2017)

[74] M. Jacob, editor, *ECFA-CERN Workshop on large hadron collider in the LEP tunnel, Lausanne and CERN, Geneva, Switzerland, 21-27 Mar 1984: Proceedings. 1.* (1984)

[75] T. Schoerner-Sadenius, editor, *The Large Hadron Collider : Harvest of Run 1*, Springer International Publishing, Cham (2015)

[76] CERN Council, *Resolution Concerning the Construction of the LHC and the Funding of the Organization* (1996)

[77] K. Kodama, et al., *Observation of tau neutrino interactions*, Phys. Lett. B **504(3)**, 218 (2001)

[78] L. Evans, P. Bryant, *LHC Machine*, JINST **3(08)**, S08001 (2008)

[79] The ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3(08)**, S08003 (2008)

[80] The ALICE Collaboration, *The ALICE experiment at the CERN LHC*, JINST **3(08)**, S08002 (2008)

[81] The CMS Collaboration, *The CMS experiment at the CERN LHC*, JINST **3(08)**, S08004 (2008)

[82] The LHCb Collaboration, *The LHCb Detector at the LHC*, JINST **3(08)**, S08005 (2008)

[83] The TOTEM Collaboration, *The TOTEM Experiment at the CERN Large Hadron Collider*, JINST **3(08)**, S08007 (2008)

[84] J. Pinfold, et al., *Technical Design Report of the MoEDAL Experiment*, Technical Report CERN-LHCC-2009-006, MoEDAL-TDR-001, CERN (2009)

[85] The LHCf Collaboration, *The LHCf detector at the CERN Large Hadron Collider*, JINST **3(08)**, S08006 (2008)

[86] P. Hermes, et al., *Measured and simulated heavy-ion beam loss patterns at the CERN Large Hadron Collider*, Nucl. Instrum. Meth. A **819**, 73 (2016)

[87] A. Miucci, *The ATLAS Insertable B-Layer project*, JINST **9(02)**, C02018 (2014)

[88] M. zur Nedden, *The LHC Run 2 ATLAS trigger system: design, performance and plans*, JINST **12(03)**, C03024 (2017)

[89] The ATLAS Collaboration, *Measurement of the top quark charge in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, JHEP **2013(11)**, 31 (2013)

[90] The 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. D **91**, 072007 (2015)

[91] The 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 **2017(11)**, 86 (2017)

[92] The ATLAS Collaboration, *Measurements of inclusive and differential cross-sections of $t\bar{t}\gamma$ production in leptonic final states in a fiducial volume at $\sqrt{s} = 13$ TeV in ATLAS*, Technical Report ATLAS-CONF-2018-048, CERN (2018)

[93] B. Völkel, *Studies of the discrimination between prompt photons and hadron fakes using neural networks*, , II.Physik-UniGö-MSc-2017/07 (2017)

[94] The ATLAS Collaboration, *Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run-1 data*, Eur. Phys. J. C **76(12)**, 666 (2016)

[95] W. McCulloch, W. Pitts, *A logical calculus of the ideas immanent in nervous activity*, Bull. Math. Biophys. **5(4)**, 115 (1943)

[96] F. Rosenblatt, *The Perceptron: A Probabilistic Model for Information Storage and Organization in The Brain*, Psychol. Rev. page 65 (1958)

[97] M. Minsky, S. Papert, *Perceptrons*, MIT Press, Cambridge (1969)

[98] D. Rumelhart, G. Hinton, R. Williams, *Learning representations by back-propagating errors*, Nature **323**, 533 (1986)

[99] J. Alwall, et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **2014(7)**, 79 (2014)

[100] J. Alwall, et al., *A standard format for Les Houches Event Files*, Comput. Phys. Commun. **176(4)**, 300 (2007)

[101] The ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data*, Technical Report ATLAS-CONF-2016-024, CERN, Geneva (2016)

[102] The Atlas Collaboration, *Muon reconstruction performance of the ATLAS detector in proton-proton collision data at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. **C76(5)**, 292 (2016)

[103] M. Cacciari, G. P. Salam, G. Soyez, *The anti- k t jet clustering algorithm*, JHEP **2008(04)**, 063 (2008)

[104] The Atlas Collaboration, *Properties of Jets and Inputs to Jet Reconstruction and Calibration with the ATLAS Detector Using Proton-Proton Collisions at $\sqrt{s} = 13$ TeV*, Technical Report ATL-PHYS-PUB-2015-036, CERN, Geneva (2015)

[105] The Atlas Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, Technical Report ATLAS-CONF-2014-018, CERN, Geneva (2014)

[106] The Atlas Collaboration, *Expected performance of the ATLAS b-tagging algorithms in Run-2*, Technical Report ATL-PHYS-PUB-2015-022, CERN, Geneva (2015)

[107] The Atlas Collaboration (ATLAS Collaboration), *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*, Technical Report ATL-PHYS-PUB-2016-012, CERN, Geneva (2016)

[108] J. Erdmann, et al., *A likelihood-based reconstruction algorithm for top-quark pairs and the KLFitter framework*, Nucl. Instrum. Meth. A **748**, 18 (2014)

[109] F. Chollet, et al., *Keras*, `https://keras.io` (2015)

[110] Theano Development Team, *Theano: A Python framework for fast computation of mathematical expressions*, arXiv e-prints **abs/1605.02688** (2016)

[111] P. Herrmann, *Studies of the $t\bar{t} + \gamma$ production process with the ATLAS experiment at the LHC*, , II.Physik-UniGö-BSc-2018/06 (2018)

Danksagung

Bedanken möchte ich mich bei allen Kolleg*innen am II. Physikalischen Institut für die tolle Arbeitsatmosphäre, vor allem bei Tobias und Ben mit denen ich mir ein Büro geteilt habe, bei Josh, Royer, Knut und Thomas, die mir bei Fragen jederzeit weitergeholfen haben und bei Herrn Prof. Dr. Arnulf Quadt für die Überlassung des Themas, die fachliche und persönliche Begleitung und das Verständnis für meine besondere Lebenslage.

Danken möchte ich auch meiner Familie, vor allem meinem Vater, der mich gerade während des Umzugs unglaublich unterstützt hat. Meinem Sohn Camilo gilt auch ein besonderer Dank. Seit er bei mir wohnt, hat er mein Leben im positiven Sinne auf den Kopf gestellt. Ich freue mich schon auf die Zukunft. Danken möchte ich ebenfalls meiner Mutter, die das Ende meines Studiums leider nicht mehr erleben kann. Auch sie hat mich immer und überall unterstützt.

Des Weiteren gilt ein Dank all meine Freund*innen, die ich hier jetzt nicht alle aufzählen kann. Zeit mit euch zu verbringen, sei es bei politischen Aktivitäten, auf dem Basketballplatz oder einfach irgendwo, hat immer viel Spaß gemacht und man konnte gut die alltäglichen Sorgen hinter sich lassen.

Ein Danke, aber vor allem mein Mitgefühl, geht an meine nicaraguanischen Freund*innen, die gerade schwere Zeiten durchleben.

Ein weiterer Dank geht an alle, die sich für den Hambacher Forst bisher so erfolgreich eingesetzt haben. Leider war es mir in letzter Zeit nicht möglich euch aktiv zu unterstützen.