Development of a multivariate reconstruction algorithm for hadronic $t\bar{t}$ systems in $t\bar{t}Z$ events with $Z \rightarrow ee/\mu\mu$ using the ATLAS detector

Ludwig-Maximilians-Universität Munich
Faculty of Physics

Master's Thesis

Nina Wenke

Munich, 8th May 2019

Entwicklung eines multivariaten Algorithmus zur Rekonstruktion von hadronischen t$\bar{\text{t}}$-Systemen in t$\bar{\text{t}}$Z-Ereignissen mit Z $\rightarrow$ ee/$\mu\mu$ für den ATLAS-Detektor

Ludwig-Maximilians-Universität München
Fakultät für Physik

Masterarbeit

vorgelegt von

Nina Wenke

München, den 8. Mai 2019

Abstract

The associated production of $t\bar{t}$ pairs with Z bosons has just been discovered in 2015 and is regarded as an important process in top-quark physics ever since. It is sensitive to the tZ coupling, thus the measurement of the $t\bar{t}$Z cross section is a direct probe of the weak coupling of the top quark. The latter might be modified in the presence of physics beyond the Standard Model, so a precise measurement of this coupling is essential. On top of that, the $t\bar{t}$Z process is also an important background in the measurement of the $t\bar{t}$H process in the multi-lepton final state. In this context, it is advantageous to have a precise measurement of the $t\bar{t}$Z cross section as well.

This thesis focuses on the reconstruction of $t\bar{t}$(Z) systems in which the top and the antitop quark decay hadronically and the Z boson decays either in an electron-positron or muon-antimuon pair. The studies are based on Monte Carlo simulated data and are conducted in the context of the full Run-2 (140 fb^{-1}) analysis.

First, the standard reconstruction technique, the minimum-χ^2 method, is applied and challenges are identified. Moreover, so-called 'improvement' cuts have been defined to address some of them. As an alternative approach, a multivariate reconstruction method using a Neural Network is developed and investigated. Comparison of their performance yields a better result for the χ^2 reconstruction. Nevertheless, the multivariate reconstruction has been found promising and should not be rejected in general. The main reasons for the lower performance are that, on the one hand, the chosen Neural Network is too simple to account for the complex $t\bar{t}$ system and, on the other hand, it was not able to extract sufficient useful information during training. A more sophisticated and well-performing multivariate method has however high potential to become an alternative or an even better reconstruction approach.

Contents

Chapter 1

Introduction

Physics aims at providing effective mathematical models of systems and phenomena on the smallest as well as on the largest scales. On the (sub-)atomic scale, one triumph of modern physics is the quantum-mechanical description of observable phenomena. Based on this, the Standard Model of particle physics describing the elementary particles and their interactions has been developed.

The heaviest elementary particle in the Standard Model is the top quark, which was discovered in 1995 by two experiments at the Tevatron [1, 2]. Since then, powerful accelerators like the Large Hadron Collider (LHC) have allowed to measure its properties more and more precisely. Recently, in 2015, the production of top and anti-top quarks ($t\bar{t}$) in association with a Z boson was discovered by CMS after evidence had been found by ATLAS. Currently, physicists working on these experiments aim to measure its properties, in particular the cross section, even more precisely using the full Run-2 data of the LHC. This is important, as the cross section of $t\bar{t}Z$ production is related to the coupling of the top quark to the Z boson, which has experimentally not been well constrained yet. Therefore, deviations might hint at some process related to physics beyond the Standard Model. Furthermore, the $t\bar{t}Z$ process is also an irreducible background of the $t\bar{t}H$ process. In order to conduct precise measurements also in this channel, it is advantageous to know the precise $t\bar{t}Z$ cross section as well.

To achieve a precise cross-section measurement a reliable and truthful reconstruction of the $t\bar{t}(Z)$ system is of utmost importance. This is a challenging task as the correct decay products of this system need to be selected from the multiple physical objects produced at the high luminosities and the high energies at the LHC. In order to assure a good reconstruction of the $t\bar{t}$ system for $t\bar{t}Z$ events, numerous hypotheses need to be considered. That is why dedicated procedures have been developed and established: the standard approach is the minimum-χ^2 method which is conducted and studied in detail in this thesis with the aim of gaining insights into reconstruction challenges. As an alternative new approach, a multivariate method based on a Neural Network is developed to check for a better reconstruction performance. Both methods are tasked with reconstructing $t\bar{t}$ systems in the 2LOS channel of the $t\bar{t}Z$ analysis using Monte-Carlo simulated data. In the 2LOS channel, a hadronic decay of the $t\bar{t}$ system is considered, while the Z boson decays into $\mathcal{2}$ Leptons

with same flavor and opposite charge (*O*pposite *S*ign). Subsequently, a comparison of the approaches regarding their performance is conducted in order to find the best performing one.

In this thesis, first, the theoretical background and the experimental setup are briefly described in Chapter 2. Chapter 3 covers the simulation of proton-proton collisions and the object definitions. It is followed by the presentation of Neural Network basics in Chapter 4. Next, the event selection used in this thesis is presented and the background of the 2LOS-t$\bar{t}$Z channel is estimated in Chapter 5. In Chapter 6 the minimum-χ^2 method is conducted and investigated, before an alternative multivariate reconstruction method using a Neural Network is developed in Chapter 7. The latter is compared to the χ^2 method regarding its performance in Chapter 8. Finally, the conclusion and an outlook are laid out in Chapter 9.

In this thesis, natural units ($\hbar = 1, c = 1$) are used. All masses and transverse momenta are given in GeV.

Chapter 2

Theoretical background and experimental setup

2.1 The top quark in the Standard Model of particle physics

The Standard Model (SM) of particle physics provides a unified picture of the elementary particles, i.e. particles without substructure, and their interactions. Developed mainly in the 1960s and 1970s [3–7], it yields an extremely precise description of experimental data up to today.
(based on [8, 9])

2.1.1 The Standard Model of particle physics

The Standard Model of particle physics describes two types of fundamental particles: fermions and bosons.
Fermions, characterized by carrying half-integer spin, are the elementary building blocks of matter and obey Fermi-Dirac statistics. The interactions between the fermions are mediated by (gauge) bosons, characterized by their integer spin and described by Bose-Einstein statistics.
In the following, the different elementary particles and their interactions will be described briefly. Further information can be found in various textbooks, e.g. [8–10].
(based on [8, 9])

Particle content

The fermions, i.e. the matter particles, are further subdivided into leptons and quarks, which both come in different types called 'flavors'. For the quarks, these are up (u),

down (d), charm (c), strange (s), top (t) and bottom (b), while the lepton flavors are electron (e), muon (μ) and tau (τ) and their associated neutrinos. They are arranged in pairs to form three generations according to their increasing masses as displayed in table 2.1.

	1st gen.	2nd gen.	3rd gen.	$\frac{Q}{|e|}$	color
Leptons	e^-	μ^-	τ^-	-1	-
	ν_e	ν_μ	ν_τ	0	-
Quarks	u	c	t	$+\frac{2}{3}$	r,g, b
	d	s	b	$-\frac{1}{3}$	r, g, b

Table 2.1: Overview of the particle content of the Standard Model of particle physics.

All leptons carry integer electric-charge, while the quarks carry a fractional charge (either $+\frac{2}{3}|e|$ or $-\frac{1}{3}|e|$). Moreover, the quarks have an additional internal degree-of-freedom called 'color'. Each flavor of quarks comes in three different colors: red (r), blue (b) and green (g). Due to the nature of strong interactions (see below), quarks cannot be observed as free particles and can only be found in colorless bound states called 'hadrons'. This phenomenon is referred to as 'confinement'.

Furthermore, both the leptons and the quarks carry isospin: the left-handed (subsection 2.1.2) particles are arranged in weak-isospin doublets of the total weak-isospin $I_W = \frac{1}{2}$, thus one of them carries $I_3 = +\frac{1}{2}$ (referred to as up-type) and the other one $I_3 = -\frac{1}{2}$ (referred to as down-type). The right-handed particles are placed in weak-isospin singlets with $I_W = 0$ and thus also $I_3 = 0$.

In addition, every one of the presented fermions has a corresponding antiparticle. They have the same mass and lifetimes as their matter counterparts, but opposite charge and magnetic moment. In particular, the antiquarks carry one unit of anticolor: anti-red ($\bar{r}$), anti-blue ($\bar{b}$) or anti-green ($\bar{g}$). The antiparticles are denoted either by their charge for the leptons or by a bar over the particle symbol for the quarks, e.g. positron e^+ and anti-up quark $\bar{u}$.
(based on [8, 9])

Interactions

All in all, there are three[1] different fundamental interactions embodied in the SM which are mathematically described by a quantum field theory. Ordered by decreasing strength, they are: the strong interaction responsible for holding together the quarks in the proton/neutron of nuclei, which is formally described by Quantum Chromodynamics (QCD); the electromagnetic interaction responsible for most phenomena outside of nuclear physics formally described by Quantum Electrodynamics (QED) and the weak interaction responsible for nuclear beta-decay formally described by Quantum Flavor Dynamics (QFD). The latter can also be described unified with the electromagnetic interaction in the electroweak theory.
Within these quantum field theories, the forces are described by the exchange of gauge bosons resulting mathematically from the requirement of local gauge invariance: this concept takes into account the arbitrary choice of the phase of the fermion field in quantum mechanics and requires that changing the phase locally in space-time should not change the underlying physics. The resulting gauge boson of the electromagnetic interaction is the massless photon. For QCD, there are eight massless gluons. In contrast, the weak interaction is mediated by the massive $W^{\pm}$ and Z^0 bosons.

The individual strength of each of the three fundamental interactions is determined by the associated coupling strength g which is the corresponding charge. As the quantum-mechanical probability for a certain process to occur is proportional to g^2 though, it is advantageous to introduce an associated dimensionless constant $\alpha \sim g^2$. Strictly speaking, the coupling constants are so-called 'running coupling constants', i.e. the coupling constants depend on the energy scale q^2 at which the measurement is conducted. This is a consequence of the renormalization property of the QFTs which is related to the requirement that the amplitudes (probabilities) of different interaction processes should be 'well-behaved', i.e. they should not diverge at high energies or to high orders of the coupling constant in perturbation theory. The general dependence of the coupling constants on the momentum transfer is described by the 'renormalization group equations'. From these it can be derived that the effective coupling constant of the electromagnetic interaction increases (very slowly) with energy-momentum transfer. The strong coupling constant α_S decreases with increasing q^2 in contrast. At low q^2 scales, α_S becomes very large and thus the perturbation expansion of QCD becomes meaningless. This is believed to be connected to the confinement of quarks mentioned above.

The nature of the three interactions is determined by the properties of the associated mediators and their interaction with the fermions: A particle can only couple to a gauge boson if it carries the conserved charge of the interaction: this is the electric charge for QED, the color (charge) for QCD and the weak isospin for the weak interaction.

---

[1]In general, there is a fourth fundamental interaction: gravity. As it is by far the weakest, it has only a vanishingly small impact in the discussion of particle interactions and is thus neglected. It is also not incorporated in the SM.

In general, QCD is formally characterized by the $SU(3)_C$ gauge symmetry group with the subscript C indicating that the conserved charge is color. A peculiarity of the strong interaction is that not only quarks but also the gluons carry color. As the latter carry one unit of color and one unit of anti-color, there should be $3^2 = 9$ gluon states. However, one of them is the colorless singlet state $\frac{1}{\sqrt{3}}(r\bar{r} + b\bar{b} + g\bar{g})$ which cannot be observed. This leaves 8 different gluons as mentioned above. Based on the fact that the gluons carry color, they also interact with themselves which contributes to the behavior of α_S and thus to the 'confinement' phenomenon.

The weak and the electromagnetic interaction are described in a unified way by the electroweak theory developed by Weinberg, Salam and Glashow (GSW theory).
Formally, this theory is based on four massless mediating boson fields: three of the bosons denoted $\vec{W}_\mu = W_\mu^{(1)}, W_\mu^{(2)}, W_\mu^{(3)}$ are the components of an I = 1 triplet of the group $SU(2)_L$ while the fourth B_μ is an isoscalar I = 0 belonging to the $U(1)_Y$ group of weak hypercharge. The weak hypercharge is defined by $Y = Q - I_3$ with Q the electric charge and I_3 the third component of the weak isospin. Thus, this theory is characterized by an $SU(2)_L \times U(1)_Y$ gauge symmetry group. The two different field types have coupling constants g for the $W_\mu^{(i)}$ and g' for B_μ, which are related by the Weinberg angle Θ_W: $\frac{g'}{g} = \tan(\Theta_W)$. The physical bosons are then combinations of these fields: while the W$^\pm$ bosons are mixtures of $W_\mu^{(1)}$ and $W_\mu^{(2)}$, the Z boson and the photon are mixtures of $W_\mu^{(3)}$ and B_μ. They acquire mass due to a process called spontaneous symmetry breaking.
(based on [8–10])

Limitations of the Standard Model of particle physics

Even though the Standard Model of particle physics provides an extremely precise description of the properties of the elementary particles and their interactions, there are still some open questions which the SM cannot answer.
Astronomical results for example exhibit that the particles incorporated in the Standard Model describe only about 5% of the mass/energy content of the universe. The rest consists of Dark Matter (about 20%) and Dark Energy (about 75%). One promising theory providing particle candidates for Dark Matter is called Supersymmetry (SUSY) in which every fermion gets a bosonic partner and vice versa.
(based on: [9, 10])

2.1.2   The top quark

The top quark is the heaviest of the elementary particles described by the SM. Discovered in 1995 by the CDF and D0 experiments at Fermilab's Tevatron [1, 2] after two decades of searches, precise measurements of its properties have been conducted at the LHC due to

the high luminosities and energies available there. Currently, its coupling to the Z boson
is measured among others.

As the mass of the top quark, which is related to the Yukawa coupling to the Higgs boson, is
a free parameter, it has been measured very precisely and was found to be 173.0 ± 0.4 GeV
[11]. The decay width of the top quark was measured to be $\Gamma = \frac{\hbar}{\tau} = 1.41^{+0.19}_{-0.15}$ GeV [11],
from which a very short lifetime $\tau_t \sim 10^{-25}$ s can be deduced. This is a peculiarity as
it is shorter than the time scale required for hadronization and to decorrelate the spin
configuration of the decay products. Thus, the top-quark polarization is not diluted by
hadronization effects and can be calculated reliably within perturbation theory. Exper-
imentally it can be analyzed using the angular distribution of its decay products. An
observable sensitive to this is the helicity angle of the W boson (subsection 2.1.2).
(based on [12])

Production

In general, top quarks can either be produced via the electroweak interaction as 'single
top-quarks' involving the Wtb vertex or in pairs ($t\bar{t}$) via the strong interaction.
As this thesis focuses on $t\bar{t}$ pairs, only the pair production via the strong interaction is
discussed in the following.

The two dominant production mechanisms of top-quark pairs in hadronic collisions are
quark-antiquark annihilation and gluon fusion. The corresponding Feynman diagrams of
lowest-order (LO) processes can be found in figure 2.1.
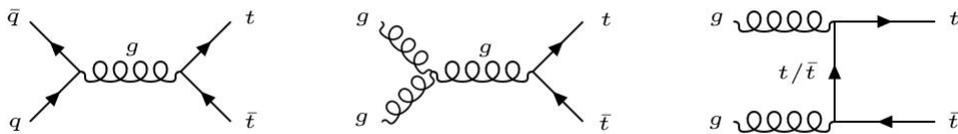
Figure 2.1: Lowest-order Feynman diagrams for the production of $t\bar{t}$ pairs by the two dom-
inant production mechanisms quark-antiquark annihilation (left) and gluon fusion (middle
and right).

While at the proton-antiproton accelerator Tevatron the quark-antiquark annihilation was
the dominant process with 85% [13], at the LHC most top-quark pairs are produced via
gluon fusion with 90% [13].
Considering the structure of the proton, this difference can be explained: The proton
consists principally of three valence quarks (u, u, d) which are bound together by gluons.
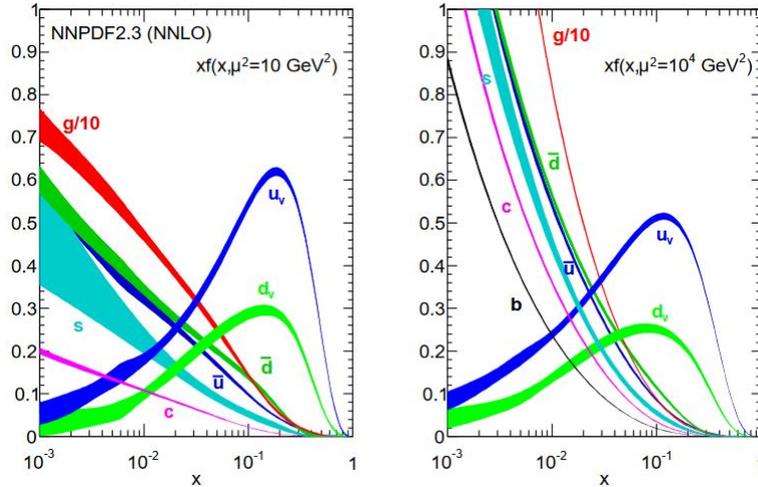
Figure 2.2: Parton distribution functions (PDFs) for two different momentum-transfer scales μ^2. [14]

In addition, so-called sea quarks can be found which result from vacuum fluctuations. The probability of finding a parton i inside a hadron H with a momentum fraction between x and $x + dx$ of the mother hadron is given by the so-called parton distribution functions (PDFs) $f_{i/H}(x, \mu_F)$ considered in the cross-section calculations. The factorization scale μ_F refers to the scale at which in higher-order calculations the initial-state singularities are factorized into the parton distribution functions. These PDFs are provided by different groups like e.g. the NNPDF group. Figure 2.2 depicts the parton distribution functions, which are used for the parton shower in this thesis, for the different center-of-mass energies or rather the different momentum-transfer scales of the Tevatron (left) and the LHC (right).

Here, the PDF of the gluons is found to grow rapidly with decreasing x, while the PDFs of the valence quarks u and d peak towards higher values of x.
In order to produce a $t\bar{t}$ pair, x needs to be roughly $\frac{350\text{GeV}}{1.96\text{ TeV}} \sim 0.2$ for $\sqrt{s} = 1.96$ TeV (Tevatron), while it needs to be only $\frac{350\text{ GeV}}{14\text{TeV}} \sim 0.03$ for $\sqrt{s} = 14$ TeV (LHC). For these x values, more gluons (valence-quarks) can be found in the right (left) plot of figure 2.2 and hence gluon fusion (quark-antiquark annihilation) is the dominant production mode at the LHC (at the Tevatron).
(based on [12, 15, 16])

Decay

In the SM, the top quarks decay via the charged weak interaction into a W boson and a down-type quark. The probabilities for the three decay modes t → Wd, t → Ws and t → Wb are proportional to the squared elements of the Cabibbo-Kobayashi-Maskawa (CKM) matrix which indicates the coupling strength of the weak charged interaction be-

tween up- and down-type quarks.

The experimentally derived magnitudes of the matrix elements related to the three decay modes are: $|V_{td}| = (8.1 \pm 0.5) \times 10^{-3}$ [11], $|V_{ts}| = (39.4 \pm 2.3) \times 10^{-3}$ [11] and $|V_{tb}| = 1.019 \pm 0.025$ [11]. From this it can be deduced that the top quark decays almost exclusively to a W boson and a b quark.

The W boson can decay either leptonically into a (antilepton) lepton and its (neutrino) anti-neutrino ($\sim 33\%$) or hadronically into an up-type (antiquark) quark and a down-type (quark) antiquark ($\sim 67\%$). According to the decay products of the W boson, three decay channels of top-quark pairs can be identified: (all-)hadronic (both W bosons decay hadronically), dileptonic (both W bosons decay leptonically) and semileptonic (one W boson decays leptonically, the other one decays hadronically).

(based on [12, 15, 16])

Spin information of the top quark: the helicity angle of the W boson

As previously mentioned, one peculiarity of the top quark is its short decay time. Due to this, its polarization and spin information is transferred onto its decay products. One experimental observable sensitive to this and in general to the Wtb-vertex structure is the so-called helicity angle.

As its name indicates, it refers to the underlying concept of the helicity of particles which is defined as the normalized component of the particle's spin $\vec{s}$ along the flight/momentum direction $\vec{p}$: $h = \frac{\vec{s} \cdot \vec{p}}{|\vec{s}| \cdot |\vec{p}|}$ [17]. For a fermion, i.e. a particle with spin $\frac{1}{2}$, there are two possible states: the projected spin can be parallel to the flight direction referred to as 'right-handed' or it can be antiparallel referred to as 'left-handed'. For a spin-1 particle like the W boson, there is an additional possible outcome as there are three possible spin states. This additional state is called 'longitudinal' referring to a configuration in which the spin is perpendicular to the momentum direction.

Helicity, is however not Lorentz invariant for massive particles moving with $v < c$ as they can always be transformed into another reference frame in which their direction of flight would be reversed, but not their spin direction. The related Lorentz-invariant concept is called chirality. Unfortunately, this is not directly related to any observable and can thus not be measured. The concepts of helicity and chirality are indistinguishable in the limit $E \gg m$ and for massless particles moving at the speed of light. Massive left-handed fermions consist mainly of a left-handed chiral state. The right-handed chiral state is suppressed by $\frac{m}{E}$ or $\sqrt{1 - \beta}$ with $\beta = \frac{v}{c}$ [17]. The analogous situation is also found for massive right-handed antifermions.

The concept of helicity is particularly important in the context of the weak interaction as the latter exhibits a so-called V-A structure. The name 'V-A' refers to the transformation properties (V for vector and A for axial vector) of the weak interaction under spatial reflection. This combination takes into account the observation that only (chiral) left-handed

particles or (chiral) right-handed antiparticles participate in the weak charged-current interaction (mediated by the $W^{\pm}$ bosons). This is not as strict for the weak neutral-current, i.e. the weak interactions mediated by the Z^0 boson due to its B_{μ}-field content.

To investigate the helicity states of the W boson in the top-quark decay, the system is transformed into the rest frame of the top quark. In consequence, the b quark and the W boson are emitted back-to-back from the top quark. As the W boson is a spin-1 particle, it can in principle be either right-handed, left-handed or longitudinal. Taking into account that the top quark as well as the b quark are both fermions, i.e. spin-$\frac{1}{2}$ particles, and that the spin is a conserved quantity, the three scenarios depicted in figure 2.3 can be found. Here, the thin arrows indicate the momentum direction of the particles while the bold arrows indicate the spin direction of the particles. The length of the spin arrows is related to the spin value.
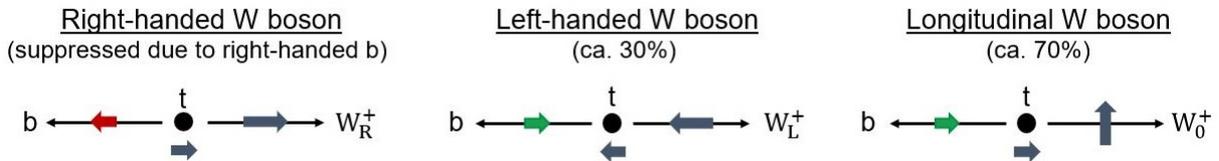
Figure 2.3: Sketch of the helicity states of the W boson in the top-quark decay. The thin arrows indicate the momentum direction while the bold arrows indicate the spin direction.

From these sketches, it can be deduced that for the right-handed W boson, the b quark also needs to be right-handed. This is highly suppressed and even strictly forbidden in the approximation that the mass of the b quark is negligible in comparison to the top-quark mass. Hence, almost no right-handed W bosons take part in the top-quark decay. In the other two helicity states of the W boson, left-handed and longitudinal, the b quark is left-handed. Thus, left-handed and longitudinal W bosons can both be found in top-quark decays.
For the anti-top quark decaying into a $\bar{\text{b}}$ quark and a W boson, the situation is reversed: the left-handed W boson needs to be excluded as the $\bar{\text{b}}$ quark needs to be right-handed. In exchange, the right-handed W boson contributes.

The probability for the different W-boson helicity states is quantified by the helicity fractions. They are defined using the partial decay widths of the three helicity states and dividing them by the total decay width of the W boson:

$$F_{\text{L, R, 0}} = \frac{\Gamma_{\text{L,R,0}}}{\Gamma} \text{ with } F_{\text{L}} + F_{\text{R}} + F_0 = 1 \text{ [18]}$$

Theoretically, they have been calculated to be $F_0 = 0.687$ [18], $F_{\text{L}} = 0.311$ [18] and $F_{\text{R}} = 0.0017$ [18] at next-to-next-to-leading order (NNLO) including electroweak effects and a finite b-quark mass.
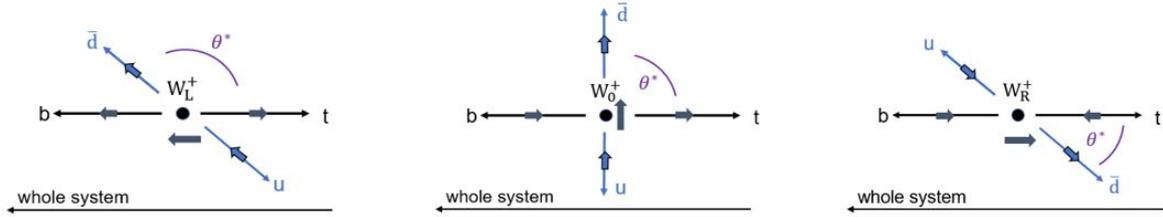
Figure 2.4: Sketches visualizing the definition of the helicity angle for the different helicity states of the W boson.

Experimentally, the helicity fractions can be measured studying the cosine of the helicity angle $\cos(\Theta^*)$ which is related to the angular distributions of the top-quark decay products. The helicity angle Θ^* is defined as the angle between the direction of the down-type fermion, i.e. the down-type quark in the hadronic $t\bar{t}$ decay, originating from the W-boson decay and the reversed direction of the b quark, both in the rest frame of the W boson [19]. This definition is visualized for a left-handed/longitudinal/right-handed W boson in the left-most/center/right-most sketch in figure 2.4. The sketches arose by considering spin conservation and the helicities of particles/antiparticles.

It is observed that the helicity angle for the left-handed W boson lies in the range between 90° and 180°, while for the longitudinal W boson the helicity angle is mainly 90°. The right-handed W bosons, though strongly suppressed, have in principle helicity angles between 0° and 90°. For the anti-top quark the sketches of the left- and right-handed W bosons would be switched, thus altogether the same distribution is found.
Furthermore, it can be deduced from these sketches, that the down-type quarks from left-handed W bosons are emitted preferentially in the opposite direction of the whole system. Thus, they tend to have lower momentum and are closer to the b quark from the top-quark decay as compared to down-type quarks from longitudinal or right-handed W bosons.

Experimentally, the distributions for the cosine of the helicity angle $\cos(\Theta^*)$ depend on the helicity fractions $F_{L/0/R}$:

$$\frac{1}{\Gamma}\frac{d\Gamma}{d\cos\Theta^*} = \frac{3}{8}F_L(1 - \cos\Theta^*)^2 + \frac{3}{4}F_0(1 - \cos^2\Theta^*) + \frac{3}{8}F_R(1 + \cos\Theta^*)^2 \text{ [19]}$$

Here, Γ indicates the decay width of the W boson.
The resulting, normalized distributions are depicted in figure 2.5 separately for each contribution and also for the combined SM expectation.
(based on [8, 17, 19, 20])
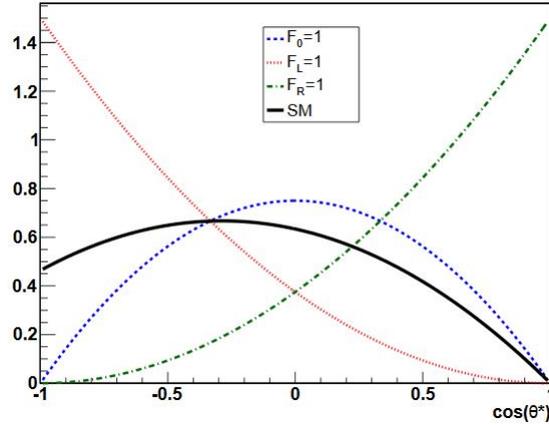
Figure 2.5: Predicted $\cos(\Theta^*)$ distributions for the different helicity fractions F_L, F_R and F_0 of the W boson according to the Standard Model of particle physics. In addition also their sum is depicted. (from [21])

2.2   Associated production of $t\bar{t}$ pairs and a Z boson

Recently, another process related to $t\bar{t}$ production has been discovered: the associated production of a Z boson with a $t\bar{t}$ pair. In this context, the Z boson is e.g. emitted from the top quarks as depicted in the two lowest-order Feynman diagrams in figure 2.6.
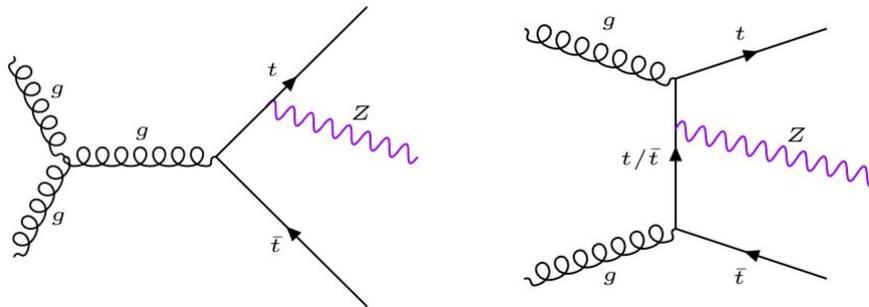
Figure 2.6: Two lowest-order Feynman diagrams of the associated production of a Z boson and a $t\bar{t}$ pair.

The $t\bar{t}$ pair of this system can decay dileptonically, semi-leptonically or all-hadronically. The Z boson can decay into a quark-antiquark pair, two leptons with opposite sign or to an invisible signature of two neutrinos. The measured branching ratios can be found in the table 2.2.

Decay mode	Z → q$\bar{q}$	Z → l^+l^-	Z → invisible ($\nu\bar{\nu}$)
Branching ratio	$(69.911 \pm 0.056)\%$	$(3.3658 \pm 0.0023)\%$	$(20.000 \pm 0.055)\%$

Table 2.2: Branching ratios related to the decay of the Z boson. (from [11])

In general, this process is a very rare process: on the one hand, the coupling of the weak interaction is weaker than the one of strong interaction reducing the probability of this process to occur. Thus, high luminosities are required for the observation of this process. On the other hand, high energies are required to kinematically allow for this process to occur as both the top quark and the Z boson are very heavy. As a consequence, the cross section of the t$\bar{t}$Z process is very small in comparison to other SM processes. This is depicted in figure 2.7.
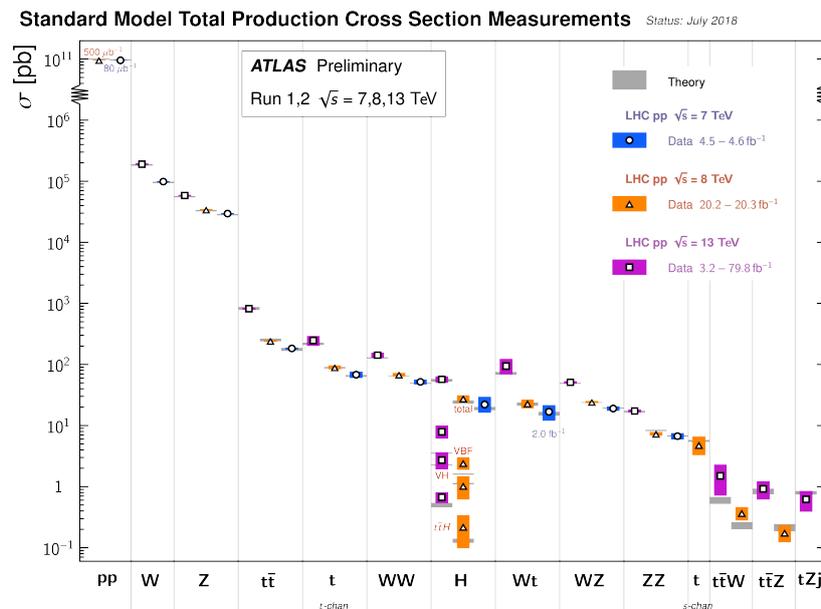
Figure 2.7: Total production-cross-section measurements for center-of-mass energies $\sqrt{s} = 7, 8, 13$ TeV for standard-model processes using the ATLAS detector. (from [22])

Here, the cross section of different SM processes is plotted. It combines the results from theory calculations (grey) and from measurements conducted at different center-of-mass energies (7 TeV in blue, 8 TeV in orange and 13 TeV in purple) with the ATLAS detector. It can be seen clearly that the cross section of the t$\bar{t}$Z process is orders of magnitudes smaller compared to e.g. the t$\bar{t}$ process representing one of the main backgrounds.

Due to this, the t$\bar{t}$Z process has been discovered only recently by CMS in 2016 [23] in the

multi-lepton final state with 6.2σ after evidence had been declared by ATLAS in 2015 with 4.2σ [24].

One of the current goals of CMS and ATLAS in this context is the precise measurement of the cross section of this system. The theoretical prediction including NLO QCD and electroweak corrections for a center-of-mass energy of $\sqrt{s} = 13$ TeV was calculated to be $\sigma_{t\bar{t}Z}^{\text{theo.}} = 0.88^{+0.09}_{-0.11}$ pb [25]. A first measurement at a center-of-mass energy of $\sqrt{s} = 13$ TeV using 77.5fb^{-1} of integrated luminosity was conducted by CMS with the following result: $\sigma_{t\bar{t}Z} = 1.00^{+0.06}_{-0.05}(\text{stat.})^{+0.07}_{-0.06}(\text{syst.})$ pb [26]. Comparing this to the theory SM prediction, a good agreement can be deduced. The measurement is currently repeated using the full Run-2 dataset.

A precise measurement of the cross section is important as it tests predictions of the SM-electroweak coupling of the top quark to the Z boson. Experimentally, this has not been well constrained yet and its value varies significantly in different models describing physics beyond the SM like e.g. vector-like quarks, strongly coupled Higgs bosons or technicolor. Furthermore, $t\bar{t}Z$ also represents an important irreducible background in measurements of the $t\bar{t}H$ process. This process is important for constraining the Yukawa coupling of the Higgs boson to the top quark which is one of the free parameters of the SM. Deviations of the production cross section of this process could also indicate processes of physics beyond the SM. [27]

For the measurement of the $t\bar{t}Z$ cross section, the multi-lepton final states, i.e. $t\bar{t}$ systems together with a leptonically decaying Z boson, are used. Even though they have the smallest branching ratio, they also provide the cleanest signature as leptons can be quite easily identified in the detector. During the last years sufficient data were collected to perform a cross-section measurement in these channels despite the small branching ratio.

The focus in this thesis lies on an analysis channel which was among the observation channels of $t\bar{t}Z$ production: the '2LOS'-channel. Its name refers to the fact that the final state has exactly 2 leptons originating from the Z boson with same flavor and opposite sign. The $t\bar{t}$ system is required to decay hadronically. This channel is characterized by a large background rate, mainly from the associated production of a Z boson and additional jets (Z+jets) and dileptonic $t\bar{t}$ processes due to similar signatures [25].

The 2LOS-analysis channel is further subdivided into 3 different signal regions. The focus in this thesis is put on the 6j2b signal region requiring ≥ 2 b jets and ≥ 6 jets. The signature of the considered processes consists thus of 6 jets: 2 b jets and 4 light jets. All of them are high transverse momentum (p_T) jets (subsection 2.3.2) due to the large mass of the top quarks. Furthermore, there are 2 leptons of opposite sign and same flavor originating from the Z boson and thus also exhibiting high p_T. In this thesis, leptons refer to electrons and muons if not stated otherwise. A Feynman graph of a 6j2b-2LOS-$t\bar{t}Z$ process is shown in figure 2.8.

(based on: [12, 28, 29])
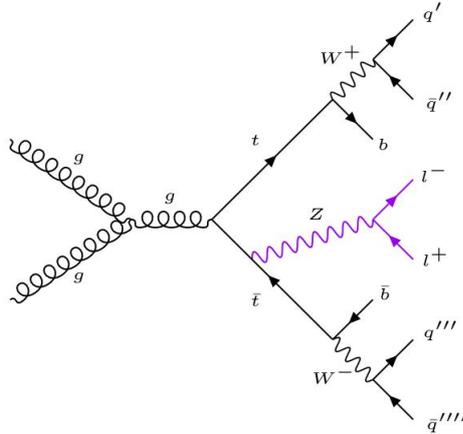
Figure 2.8: Feynman diagram of a process targeted by the analysis in this thesis: the associated production of a $t\bar{t}$ system with a Z boson with the $t\bar{t}$ pair decaying hadronically and the Z decaying leptonically, i.e. either in an electron-positron or muon-antimuon pair.

2.3 The ATLAS experiment at the LHC

To study the elementary particles and to search for new physics, protons are accelerated to almost the speed of light in the currently most powerful accelerator, the Large Hadron Collider (LHC). They are then brought to collision in beam crossings at the detectors of different experiments like ATLAS. Collision data were recorded in the 'Run-1' period from 2009 - 2013 and recently in the 'Run-2' period from 2015-2018.
In this section a brief description of the LHC and the detector of the ATLAS Experiment is given. Further information can be found in [30–33].
(based on: [34–38])

2.3.1 The Large Hadron Collider (LHC)

The LHC is part of the particle physics research center CERN situated at the Franco-Swiss boarder near Geneva. It accelerates mainly protons[2] to center-of-mass energies of up to 14 TeV (13 TeV in Run-2) in the 26.7 km [37] long tunnel built initially for the Large Electron-Positron collider LEP. Situated between 45m and 170m below the surface of the earth, it represents the last chain of the accelerator complex (Linac2, PS Booster, Proton Synchrotron (PS) and Super Proton Synchrotron (SPS)) receiving the pre-accelerated proton bunches with a spacing of 25 ns [37]. Radio-frequency (RF) cavities, i.e. electromagnetic resonators, then accelerate the particles first to the desired energy and then to compensate for the synchrotron radiation losses.

---

[2]Also heavy ions, in particular lead nuclei, are accelerated to 2.8 TeV [37] per nucleon and collided in special runs.

To keep the protons on their nearly circular orbit, different types of superconducting magnets are used. Among them are dipole magnets to bend the beam and quadrupole magnets to focus the beam especially before the collision points to maximize the number of interactions.

To explore rare events not only high energies, but also high luminosities are required. Using the measured (instantaneous) luminosity L, the number of events/interactions per second generated in the LHC collisions can be determined: $N_{event} = L \cdot \sigma$ with σ referring to the cross section, i.e. the quantum-mechanical probability, of the process under study. At the design luminosity of the LHC of $10^{34} \mathrm{cm}^{-2} \mathrm{s}^{-1}$ [37], the bunches collide about 40 million times per second [34].

These collisions take place in the detectors of the four main experiments ATLAS, CMS, LHCb and ALICE shown in figure 2.9. ATLAS (A ToroidaL ApparatuS) and CMS (Compact Muon Solenoid) are the two high-luminosity multi-purpose experiments covering a wide range of physics from precision measurements to searches of evidence for physics beyond the SM. Situated at diametrically opposite sites, they differ only in their technical realization. The two low luminosity experiments focus on special topics: LHCb was designed to investigate the matter-antimatter asymmetry in interactions of B particles and ALICE (A Large Ion Collider Experiment) studies the properties of quark-gluon plasma in lead-ion collisions.
(based on: [8, 34–38])
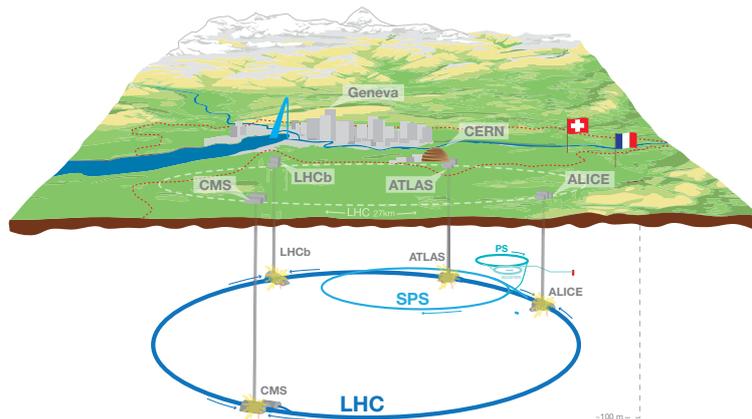
Figure 2.9: Schematic picture of the LHC and its four main experiments ATLAS, CMS, ALICE and LHCb. (from [39])

2.3.2 The ATLAS Experiment

ATLAS, one of the main experiments at the LHC, has a height of 25 m and a length of 44 m which makes it the detector with the largest volume ever constructed [36].

It is described by a cartesian coordinate system depicted in figure 2.10 with an origin situated at the interaction point. The z-axis is defined as the beam direction, the positive x-axis points from the interaction point to the center of the LHC ring and the positive y-axis points upwards. Thus, the x-y plane defines a plane transverse to the beam direction. In this plane also transverse properties like the transverse momentum, p_T, are defined.

To take advantage of the cylindrical form of the ATLAS detector, also the azimuthal angle ϕ and the polar angle Θ are defined as the angle around the beam axis and as the angle from the beam axis respectively. As Θ is not Lorentz invariant, the pseudo-rapidity is defined as $\eta = -\ln\left(\tan\left(\frac{\Theta}{2}\right)\right)$ [34]. In case of massive objects the rapidity $y = \frac{1}{2}\ln\left[\frac{(E+p_z)}{(E-p_z)}\right]$ [34] is used. The distance in the pseudorapidity-azimuthal angle space is then defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ [34].
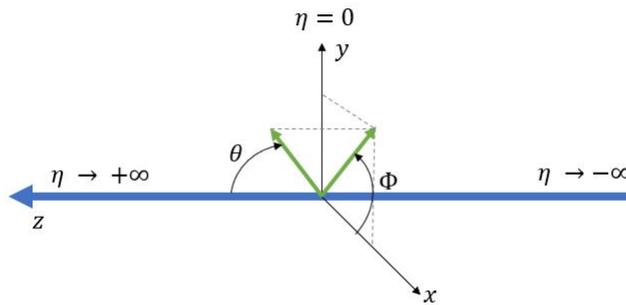
Figure 2.10: Visualization of the ATLAS coordinate system. (leaned on [40])

The cylindric detector is forward-backwards symmetric with respect to the interaction point and consists of a barrel part and two end caps. Three sub-systems exploiting different technologies are used to detect and measure the properties of the particles produced: the Inner Detector (ID), the calorimeters (ECAL, HCAL) and the Muon Spectrometer (MS).

The detector situated closest to the interaction point is the Inner Detector (ID) which tracks the charged-particle trajectories and interaction vertices. Its acceptance range is the pseudorapidity range $|\eta| \leq 2.5$ [34]. At inner radii, multiple layers of two high-resolution detectors, the semiconductor Pixel detector (barrel layer: $|\eta| \leq 1.7$ [31]) and the semiconductor tracking detector (SCT) (barrel part: $|\eta| \leq 1.4$ [31]) are used. Outside, straw-tube tracking detectors (TRT) (barrel axial straws: $|\eta| \leq 0.7$ [31]) are used to provide quasi-continuous tracking. Traversing charged particles leave a hit in each layer from which

the track can be reconstructed. All these three detectors are penetrated by a solenoidal magnet field of 2 T [34] provided by a superconducting solenoid. This enables momentum measurements of the particles $\vec{p}$ by determination of the bending radius R of the particle tracks: $|\vec{p}| = q \cdot |\vec{B}| \cdot R$ with q the charge of the particle and $\vec{B}$ the magnetic field strength.
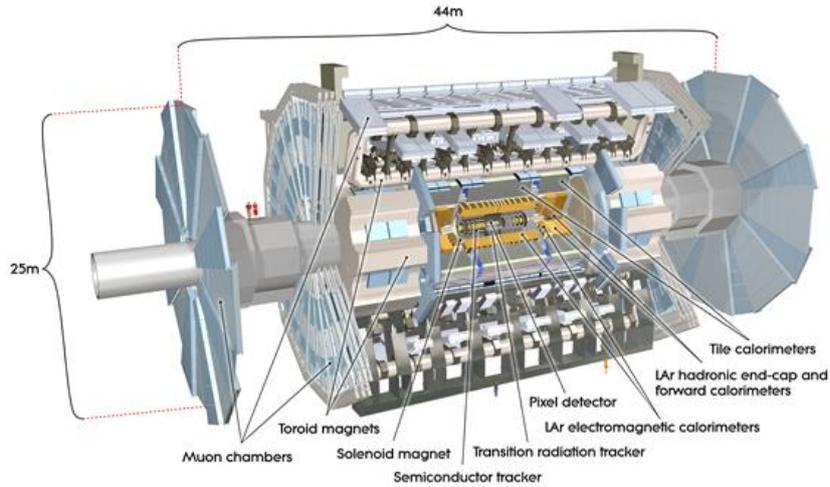
Figure 2.11: Schematic picture showing the different sub-systems of the ATLAS Experiment. (from [41])

This tracking volume is surrounded by two calorimeters: the Electromagnetic Calorimeter (ECAL) and the Hadronic Calorimeter (HCAL). Their task is to measure the energies of the particles in the range $|\eta| < 4.9$ [34] using so-called sampling detectors consisting of alternating layers of absorbing and active material. In the absorbing material, cascades of particles ('showers') based on the electromagnetic interaction in the form of Bremsstrahlung and pair production (ECAL) or strong interaction (HCAL) develop. Their energies are then measured in the active material.

The ECAL is an accordion-shaped lead liquid-argon (LAr) calorimeter covering the pseudo-rapidity range $|\eta| < 3.2$ [34] (barrel part $|\eta| < 1.475$ [34] and two end caps $1.375 < |\eta| < 3.2$ [34]). The HCAL, placed outside the ECAL, uses steel as absorbing material and scintillating tiles as the active material in the barrel ($|\eta| < 1.7$ [34]). In the end caps, ($1.5 < |\eta| < 3.2$ [34]) copper plates are combined with LAr gaps.

To extend the coverage of the calorimeter to $|\eta| = 4.9$ [34], forward calorimeters provide both electromagnetic and hadronic energy measurements using copper/tungsten in combination with LAr.

Outside the calorimeters, the muon spectrometer can be found. It records the trajectories of muons, the only measurable particles penetrating the other sub-systems, in Monitored Drift Tubes (MDTs) and Cathode Strip Chambers (CSCs) within $|\eta| < 2.7$ [34] and $2 < |\eta| < 2.7$ [34] respectively. The tracks are bent by large superconducting toroid

magnets with a bending power of 1 to 7.5 Tm [34] allowing for the measurement of the muon momenta.

As not all the measured data can be recorded with current technology, a trigger system is used to decide which events to keep for later analysis. It consist of a hardware-based level-1 trigger determining regions-of-interest (RoIs) and reducing the event rate from about 30 MHz [38] to 100 kHz [38]. The RoIs are then transferred to the software-based high-level trigger (HLT) reducing the event rate further to approximately 1 kHz [38]. In the Run-2 data taking period about 140 fb^{-1} have been recorded.
(based on [8, 31–34, 38])

Chapter 3

Simulation and object definitions

To compare theoretical predictions of the Standard Model to data taken in experiments like ATLAS, Monte-Carlo simulations of proton-proton (pp) collisions are used. As the output of the simulation is comparable to the measurements provided by the detector, the physical objects used in the analysis have to be reconstructed first.

In this chapter it is described how proton-proton collisions are simulated by Monte-Carlo event generators. Subsequently, it is explained how the objects used in this thesis have been reconstructed based on the simulation output. More information on these topics can be found in the sources given.

3.1 Monte-Carlo simulation of proton-proton collisions

To validate the predictions of the Standard Model, simulations of proton-proton collisions are used. They provide further insights by demonstrating how the physical input is distorted step-by-step and by providing detailed descriptions of the final states. The simulated information can then be used to predict experimental observables and subsequently compare them to experimental data.

Usually, so-called Monte-Carlo (MC) event generators are used to model the quantum-mechanical probabilities of the various stages of the collision process based on (pseudo-) random numbers and the Monte-Carlo method.
To be able to provide reliable predictions, a large number of simulated events is required as only their average can describe the expected probability distributions and thus provide a sufficiently realistic description of the average behavior and its fluctuations.

3.1.1 Process of a proton-proton collision

Most event generators divide the process of a proton-proton collision, called event, into the five steps described below. Four of them are visualized in figure 3.1.
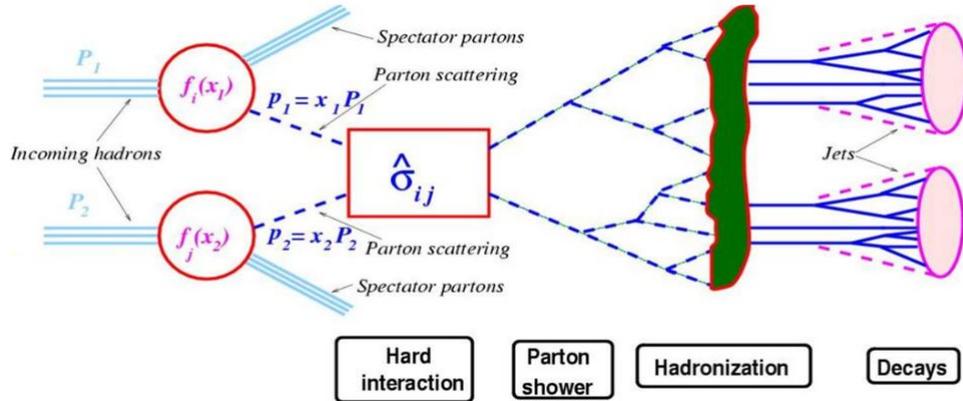
Figure 3.1: Visualization of the process of a proton-proton collision. (adapted from [42])

First, the *hard interaction* is simulated in which two protons consisting of quarks and gluons approach each other and collide. This step can be calculated in perturbation theory using the probability density functions (PDFs) of the partons in the proton. If short lived particles like the top quark or Z boson are produced, their decay is also included in this step as e.g. their spin correlations are transferred from the production to the decay states. The Lorentz vectors of these final-state particles/partons are saved in the so-called truth information.

Subsequently, a *parton shower* develops originating from the produced partons. As accelerated color-charged partons are included in the collision, Bremsstrahlung in the form of gluon emission occurs. Emissions associated with the two incoming colliding partons are referred to as Inital-State Radiation (ISR,) while emissions associated to outgoing partons are referred to as Final-State Radiation (FSR). As gluons carry color themselves and therefore also couple to each other, emitted gluons will give rise to new radiation as well. This leads to a cascade of particles modeled as an evolution downwards in momentum scale. The modeling stops at the point at which perturbation theory breaks down. This is due to the fact that the coupling constants are actually running coupling constants (subsection 2.1.1). In fact, the coupling constant of QCD, α_S, increases towards lower momentum scales and thus perturbation theory developed in powers of α_S can no longer be used for calculations.

That is why, in the next step called *hadronization*, different models have to be taken into account to simulate the confinement of a system of partons into hadrons. The two most well-known models are the string model [43] and the cluster model [44]. This step is important as colored partons cannot be observed freely and thus only hadrons can be observed

in the detector. Subsequently, also the *decays of unstable hadrons* have to be considered. So far, only two partons were extracted from the incoming protons to undergo the hard interaction. However, the proton is made up of a multitude of further partons. These so-called beam remnants are still colored and thus hadronize and interact as well. As a result, soft hadrons arise which overlay and contaminate the simulated hard process. These secondary interactions of the beam remnants are addressed in the simulation step called *underlying event.*

Having simulated the observable particles, finally also the detector response needs to be simulated in order to allow for comparison between simulation output and the data taken in LHC pp collisions. This is done for the ATLAS detector using Geant4 [45]. As this is quite costly in terms of CPU, there is also a fast simulation which reduces CPU costs by approximating the calorimeter simulation [46].
(based on [47, 48])

3.1.2   Event generators

The described interaction process can be split up in two parts according to the energy scale used: The hard process is simulated by generators like Madgraph5_aMC@NLO [49] (the new version of both Madgraph and aMC@NLO) or Powheg [50] (short for POsitive Weight Hardest Emission Generator) which calculate the matrix elements. These are then combined with so-called general-purpose generators like Pythia [51] using the Lund string model for hadronization. To assure that no double-counting of processes occurs, different approaches have been developed and incorporated in the different generators. For example, MC@NLO subtracts the parton shower approximation from the NLO calculation which leads to un-physical negative weights. Powheg in contrast provides only positive event weights.
Alternatively, proton-proton collisions can be simulated using the Sherpa [52] generator alone. Here, the hadronization is simulated based on the cluster model [52].

The MC samples used in this thesis are the following ones:

The **t$\bar{\text{t}}$Z** sample was generated with MADGRAPH5_aMC@NLO [49] version 2.33.p0 interfaced to Pythia8 [51] v8.210. The NNPDF3.0NLO PDF set [53] was used in the matrix-element (ME) computation. In the parton shower, the A14 [54] set of tuned MC parameters was used together with the NNPDF2.3LO PDF set [55].

The **dileptonic-t$\bar{\text{t}}$** sample was generated with Powheg-Box [56–59] v2 (r3026) using the NNPDF3.0NLO PDF set. The parton shower and the underlying event were generated by Pythia8 [51] v8.210 using the NNPDF2.3LO PDF set with the A14 [54] tune.

Heavy-flavor hadron decays involving b and c quarks were modeled in both samples using

EVTGen [60] v1.20.

The sample containing the associate production of a Z boson with additional jets (**Z+jets**) was generated by Sherpa 2.21 [52,61] using the NNPDF3.0NNLO PDF set.

The detector simulation was performed based on Geant4 [45] for all three samples.

3.2   Object definitions

To reconstruct the physical objects used in this thesis from the output of the detector simulation, different criteria are applied. These are described briefly in the following for the primary vertex, electrons, muons and jets. Furthermore, the identification of b jets (b-tagging) and the determination of the jet charge are explained. Finally, also the overlap-removal procedure is presented.

3.2.1   Primary vertex

The primary-vertex position, i.e. the spatial coordinate of the hard scatter, is reconstructed using the charged-particle tracks of the Inner Detector with a transverse momentum > 400 MeV within the pseudorapidity range $|\eta| < 2.5$ having a certain number of hits in the different layers of the Inner Detector.
As in the proton-proton collision additional low transverse-momentum collisions occur (pile up), several vertices might be reconstructed. In this case, the one with the highest square root of the sum of the squared transverse momenta of the associated tracks is selected to be the primary vertex.
(based on [62])

3.2.2   Electrons

Electrons (including positrons) are charged particles and hence leave tracks in the Inner Detector as well as energy deposits in the electromagnetic calorimeter. Based on these, electron candidates in the central region of the ATLAS detector ($|\eta| < 2.47$) are defined. To further distinguish the signal-like electron candidates from background-like ones, e.g. converted photons, certain identification and isolation criteria are applied. For electron identification, shape properties of the electromagnetic showers in the calorimeter, tracking information and the track-to-cluster-matching quantities are combined using a likelihood method. Based on the resulting discriminant, three different working points are defined. In order of increasing background rejection, they are 'loose', 'medium' and 'tight'. For this thesis, the medium working point has been chosen.
In addition, further criteria are applied requiring the electron to be isolated from other

activity in the calorimeter/Inner Detector using two discriminating variables: the calorimetric isolation energy $E_T^{\mathrm{cone0.2}}$ and the track isolation $p_T^{\mathrm{varcone0.2}}$. $E_T^{\mathrm{cone0.2}}$ is defined as the sum of the transverse energies of topological clusters within a cone of $\Delta\mathrm{R} = 0.2$ around the candidate-electron cluster. For this sum only clusters with reconstructed positive calibrated energy are considered and the energy around the electron-cluster barycenter is subtracted. $p_T^{\mathrm{varcone0.2}}$ is defined as the sum of the transverse momenta of all tracks satisfying certain quality requirements within a cone of $\Delta\mathrm{R} = \min(0.2, 10\frac{\mathrm{GeV}}{\mathrm{E_T}})$ around the candidate electron track and originating from the reconstructed primary vertex of the hard collision. Electron associated tracks, i.e. the electron track, and additional tracks from converted Bremsstrahlung photons are excluded. The used 'fixed-cut-tight' isolation sets cut values of 0.06 on both $\frac{\mathrm{E_T^{cone0.2}}}{\mathrm{E_T}}$ and $\frac{\mathrm{p_T^{varcone0.2}}}{\mathrm{E_T}}$.
(based on [63])

3.2.3 Muons

Muons leave signals in the Inner Detector and the muon spectrometer, where reconstruction is performed independently at first. Most muons considered in this thesis are first reconstructed in the muon spectrometer and the tracks are then extrapolated to match a track in the Inner Detector. This procedure cannot be conducted in the pseudorapidity range $2.5 < |\eta| < 2.7$ as it is not covered by the Inner Detector. Here, the muon trajectory is reconstructed solely in the muon spectrometer and is tested for compatibility with the interaction point.

To suppress non-prompt muons, mainly originating from pion and kaon decays, muon identification is performed. In this context, again three different working-points are defined referred to as 'loose', 'medium' and 'tight'. The medium muons used in this thesis are selected by applying quality requirements which include a minimum number of hits in the muon chamber and compatibility between Inner-Detector and Muon-Spectrometer momentum measurements among others.

Furthermore, it is exploited that muons originating from the decay of Z bosons are often produced isolated from other particles. This is done by measuring the detector activity around a muon candidate and choosing an appropriate isolation working point. In this thesis, the fixed-cut-tight-track-only isolation is chosen, which requires $\frac{p_T^{varcone30}}{p_T^\mu} < 0.06$. $p_T^{varcone30}$ is in this context defined as the scalar sum of the transverse momenta of tracks with $\mathrm{p_T} > 1$ GeV in a cone of size $\Delta\mathrm{R} = \min(10\frac{\mathrm{GeV}}{\mathrm{p_T^\mu}}, 0.3)$ around the muon with transverse momentum p_T^μ. In this procedure, the muon track itself is excluded.
(based on [64])

3.2.4   Jets

Due to confinement quarks cannot be observed free but only in the form of a collection of collimated sprays of hadrons referred to as 'jets' [65]. They are detected as energy deposits in the calorimeters.

The energy deposits are first gathered into so-called 'topological clusters' using a sequence of seed-and-collect steps which are repeated until all topologically appropriate connected cells have been used.

The topological clusters with positive energy are further sequentially recombined using the 'anti-k_t' jet-clustering algorithm [66] with a radius parameter of R = 0.4. This algorithm has the important properties of infrared and collinear safety which refer to the fact that modifying an event by introducing collinear splittings or adding soft emissions will not change the set of hard jets in the event.

To suppress the impact of pile-up jets, a variable called jet-vertex tagger is constructed based on tracking information.

The reconstructed jets can then be related to tracks by a method called ghost association [67], in which the tracks are treated as infinitesimally soft, low-p_T particles. This is done by setting their p_T to 1 eV. They are then added to the list of inputs for the jet-finding algorithm. Due to the low-p_T-value, the tracks do not affect the reconstruction of calorimeter jets, but nevertheless allow to identify which tracks are clustered into which objects afterwards.

(based on [66–72])

3.2.5   B-tagging

In general, identifying the flavor of the quark which initiated a jet is not a trivial task. However, for b-initiated jets ('b jets'), some characteristics can be exploited: b quarks hadronize into B hadrons which are contained in the detected jets. These are very long-lived and thus typically travel some millimeters before decaying into multiple charged particles. The distance of closest approach between the track and the primary vertex is referred to as the so-called impact parameter. Moreover, the large mass of the b quark results in an emission of the decay products at relatively large angles to the original b-quark direction. This results in the formation of a secondary vertex from the B-hadron decay which is displaced from the primary vertex.

To identify b-initiated jets, different algorithms exploit the impact parameter, the secondary vertex and the topological structure of weak B- and c-hadron decays. Their output is combined with a boosted decision tree (BDT) in a multivariate discriminant called MV2. In this thesis, the MV2c10 version is used. The 10 in the name indicates that the c jet (light jet) share in the training of the BDT was 10% (90%) [73]. To obtain a b-tagging efficiency of 77% as used in this thesis, a cut value of 0.6459 on the BDT output is set [73]. The c-jet/light-jet/tau rejection is given in the form of benchmark numbers of 6/134/22 [73]

referring to a misidentification rate of $\frac{1}{6} \sim 16\%/\frac{1}{134} \sim 1\% \, / \, \frac{1}{22} \sim 5\%$ [74].
(based on: [8, 73])

3.2.6 Jet charge

Due to hadronization, the electric charge of the initial quarks/gluons cannot be measured. This information is, however, embedded in the detected jets in the form of the jet charge. One definition of the latter uses ghost-associated tracks to calculate the jet charge Q_J of a jet J using a transverse-momentum weighting scheme:

$$Q_J = \frac{1}{p_T^J} \cdot \kappa \sum_{i \in \text{tracks}} q_i \cdot (p_{T,i})^\kappa \ [65]$$

with q_i the charge of the track i with associated transverse momentum $p_{T,i}$. p_T is the transverse momentum of the jet J and κ is a free regularization parameter controlling the sensitivity of the jet charge to soft radiation. For $\kappa > 0$, the definition is infrared safe, but still not Lorentz invariant. For low values of κ, the contribution of low-p_T particles is enhanced. Choosing an infinitely-high value for κ, however, reduces the contributing tracks to the one with the highest p_T. Theoretical predictions show that a value of 0.5 would be the most sensitive to the charge of the quark initiating the jet. As only κ values of 0.3 and 0.7 were available, the 0.3 option has been selected for this thesis.
(based on [65])

3.2.7 Overlap removal

In order to avoid that one physical object is reconstructed as two different objects (duplication) or that two separate but close-by objects (isolation) are treated appropriately, the so-called overlap removal is conducted.
In this context, electron candidates are removed from the event in a first step if they overlap with a muon candidate within a cone of $\Delta R < 0.2$ or if they share a track with a muon. In a next step, jet candidates are removed if they are situated within a cone of $\Delta R < 0.2$ to the electron candidate. Subsequently, the remaining electron/muon candidates are rejected if they exhibit a distance of $0.2 < \Delta R < 0.4$ to a jet. Finally, further muon candidates are rejected if the distance between the jet and the muon candidate is smaller than $\Delta R = 0.4$ and the muon candidate has fewer than two associated tracks.
(based on [75])

3.3   Scale factors

In order to take into account the different efficiencies for identification, isolation as well as for the impact-parameter determination in simulation and data, so-called scale factors/weights are calculated. These are defined as the ratio of the efficiency/mistag-rate in data ϵ_{data} to the one obtained in simulation ϵ_{sim} [76]. These scale factors exist for b-tagging, lepton definitions and jet-vertex tagging (JVT).
Furthermore, there is another weight adapting the pile-up profile of the generated events to the one of data.
Moreover, also the specific MC-generator weight has to be applied to assure a physical output of the simulation.
Taking all these factors into account, the overall event weight used in this thesis is defined as follows:

$$
\begin{aligned}
\text{event weight} = & \text{mc weight} \cdot \text{pileup weight} \cdot \text{jvt weight} \cdot \text{b-tag weight} \cdot \\
& \cdot \text{lepton weight} \cdot \frac{\text{cross section} \cdot \text{lumniosity}}{\text{total number of weighted events}}
\end{aligned}
$$

The aim of the last term is to adapt the simulated luminosity to the measured one.

Chapter 4

Neural Network basics

In order to develop an alternative, multivariate reconstruction approach for $t\bar{t}$ systems in $t\bar{t}Z$ events, a classifier is needed to distinguish between well-reconstructed (signal) and badly-reconstructed (background) $t\bar{t}$ systems.

The traditional way for classification is the cut selection, i.e. the sequential application of requirements to the individual observables of one event. To date, more sophisticated and powerful approaches have been developed: So-called multivariate algorithms (MVA) combine the information of the different observables of an event into a single output variable. Rather than assigning the event to a definite class, they use this output to assign a certain probability to an event which indicates how likely this event belongs to a certain category. One way to approximate this multi-dimensional decision function in the observable space is to use machine-learning techniques. In a subfield of machine learning, called *supervised learning*, the algorithm learns from examples which consist of pairs of input and output values. After this learning process, referred to as training, the learned function approximation is used for classification of events with yet unknown class probabilities.

The technique used for classification in this thesis is a feed-forward Neural Network (NN), also called multi-layer perceptron or artificial Neural Network. It is applied in the context of supervised learning.

In this section, the structure and the basic working principle of feed-forward Neural Networks (NNs) are discussed and important challenges are outlined. Towards the end, some performance-evaluation options for the training process are presented. More information can be found in [77].

The whole chapter is based on [77–82].

4.1 Feed-forward Neural Networks

Feed-forward Neural Networks (NN) are machine-learning techniques which in their structure loosely resemble the human brain. In this way, they achieve to learn underlying

features of the example data. As the name already suggests, information flows only in the forward direction through the network, meaning from the input to the output. Thus no feedback of information is allowed. If feedback connections are allowed, the network is called Recurrent Neural Network (RNN).

4.1.1   Single- and multi-layer perceptrons

The basic components of the human brain are *neurons*. They pass on information if the input values exceed a certain threshold. Their output is then transferred to neighboring neurons through connections which are strengthened/weakened based on how often they are used.

Artificially, this idea is implemented in the form of a so-called (single-layer) *perceptron* [83]. Its functionality is visualized in figure 4.1.
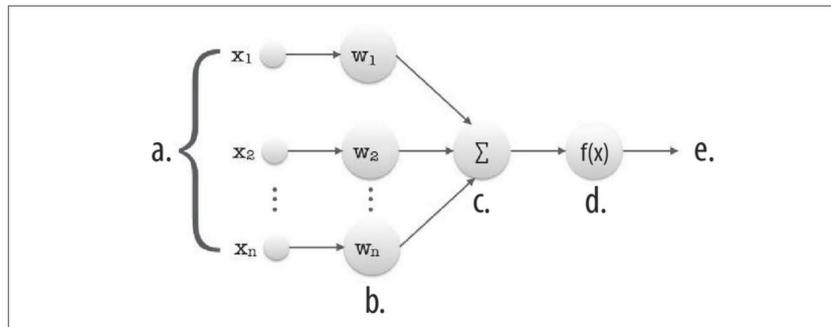
Figure 4.1: Schematic structure and work flow of a single-layer perceptron as described in the text. (adapted from [80])

Analogous to the biological neuron, the perceptron/artificial neuron, takes multiple inputs $x_1, x_2, ..., x_n$ (step a)). The *connection strength* is given by weights $w_1, w_2, ..., w_n$ associated to the inputs as shown in step b). Within the perceptron, the input values are added up in the form of a weighted sum $\sum = \sum_{j=1}^{n} w_j x_j + w_0$ in step c). Not shown in the sketch is w_0 representing the intercept value, also called *bias*, which makes the model more general. The weighted sum is then used as an input for the so-called *activation function* f in step d) which models the threshold of the biological neurons. It can be interpreted as the extent of activation of the neuron. Usually it is chosen to be one of the following:

- **sigmoid**: $f(x) = (1 + e^{-x})^{-1}$ ranging from 0 to 1

- **hyperbolic tangent**: $f(x) = \tanh(x) = \frac{(e^x - e^{-x})}{(e^x + e^{-x})}$ with a range from -1 to 1

- **rectified linear unit (ReLu)**: $f(x) = \max(0, x)$

- **leaky ReLu**:

$$f(x) = \begin{cases} a \cdot x & x \leq 0; \ a = \text{const.} \\ x & x \geq 0 \end{cases}$$

These activation functions are visualized in figure 4.2. Here, one can see that the sigmoid and the hyperbolic tangent have both S-shaped distributions and squeeze the output either in the range of 0 to 1 or -1 to 1. The ReLu function is only non-zero for values above 0. This gives rise to the problem that for negative x values no gradient can be determined, which is important later on. The leaky ReLu takes care of that by providing a slope a, in this case $a = 0.01$, for negative x values.
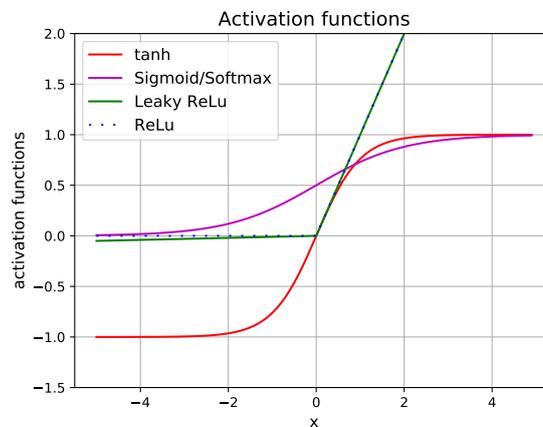
Figure 4.2: Visualization of the activation-function distributions.

For classification, a probability value is desired as an output and hence a sigmoid function is used to calculate the output-value.

Finally, the output information of the perceptron passed on in step e) is consequently of the form $y = f(\vec{x} \cdot \vec{w} + w_0)$, where f is the activation function, $\vec{x}$ is the vector containing all input data, $\vec{w}$ is the vector containing weights corresponding to the inputs and the bias w_0. In general, the output of the perceptron defines a (hyper-)plane in a n-dimensional space and thus can be used for multivariate linear fits.

To learn more complicated functions, using not only one, but rather the interaction of multiple neurons is key. This idea is implemented in the *multi-layer perceptron* [84–87], which is also called feed-forward Neural Network.
Within the human brain, the neurons are connected to each other in a way that they use the output of the previous neurons as input. This can be thought of arranging them in *layers*, so that the information flows from one layer to the next.
Taking this concept as archetype, an artificial 'Neural Network' can be constructed. Its structure is sketched in figure 4.3. It should however be mentioned at this point, that the

goal of Neural Networks is not to perfectly model the brain, but rather to occasionally draw some insights from neuroscience.
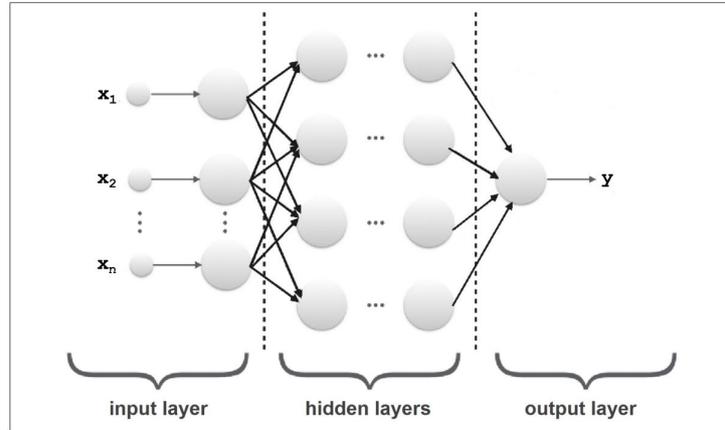
Figure 4.3: Schematic structure of a multi-layer perceptron/Neural Network: It consists of an input layer with the dimension of the input variables, an output layer and hidden layers inbetween. (adapted from [80])

The first layer, the *input layer*, pulls in the input data $\vec{x}$. The last layer, the *output layer*, contains the output node and computes the final answer. The layers inbetween are called *hidden layers*. The number of nodes in a hidden layer defines the width of the model, while the number of hidden layers defines the depth of the network.

Thus, the input data propagate from the input layer to the first hidden layer consisting of a specified number of neurons. In each of these neurons, the weighted sum of the inputs is calculated and transformed by the activation function. The output-value, the 'activation' of each neuron, z_h, propagates then in forward direction to the next, $(h+1)^{\text{th}}$ layer where it is used as the input $z_{h+1}^j = f(\sum_{i, j} v_{ih}^j z_{ih}^j + v_{h+1}^j)$ with weights v_{ih} for all connections i coming from the previous layer h for every neuron j in the $(h+1)^{\text{th}}$ layer. v_{h+1}^j refers to the bias of the neuron j for the considered layer. This is repeated until the output layer is reached.

If the hidden layer outputs were only linear, the concept of hidden layers would be of no use as linear combinations of linear combinations stay linear combinations. By applying the nonlinear activation functions, the output becomes a linear combination of the nonlinear function values computed by the hidden units instead. The more hidden layers the model has, the more complex functions can be approximated.

4.1.2   Learning: backpropagation and optimizer

Classification can formally be described by a multi-dimensional decision function F in the observable space which maps an input x to an output y: $y = F(x)$. With the training

examples at hand, this function can only be sampled though. Thus, the goal of the Neural Network is to approximate F with $y^* = F^*(x, w)$. This is done by learning the value of the parameters $\vec{w}$ that result in the best approximation during a process called training.

In the training phase, a large number of examples consisting of input-output pairs is presented to the NN in the case of supervised learning. Thus, each example x is accompanied by a label $y = F(x)$.
Starting from random initial weights, the weights are iteratively modified in order to minimize the discrepancy between the 'real' output y and the calculated output y^*. This discrepancy, or in other words the error or loss, is calculated using a so-called *loss function*. In the setup of this thesis, the cross entropy is used as the loss function:

$$\text{loss} = -\sum_t y^{(t)} \log(y^*)^{(t)} + \left(1 - y^{(t)}\right) \log\left(1 - (y^*)^{(t)}\right) \quad [82]$$

The loss describes the sum over all training examples t with the desired output $y^{(t)}$ and the output of the neuron $(y^*)^{(t)}$ for the example considered. $y^{(t)}$ can either take the value of 0 (background) or 1 (signal).

Having processed a sufficient number of examples, the NN is expected to have found the best possible weights, i.e. weights that minimize the loss over all the training examples. In consequence, it is assumed that the best possible approximation F^* has been found and that the NN is consequently well-performing for the task it has been trained for. To check this, one option is to calculate the *accuracy* which is defined as the ratio of correct classified examples to all examples.

Due to the nonlinearity of the activation function, one cannot simply solve a linear system of equations to determine the best choice of weights. Instead an *optimizer* has to be used to maximize the performance of the NN by iteratively adapting its parameters. Well-established optimizers are the *(stochastic) gradient descent method* [88–90] and the *adam optimizer* [91].
This weight optimization can either be done separately for each example or for a subset called *batch*. In the latter case, the optimization changes are accumulated and applied after all the examples of this subset have been used. The number of examples in the considered subset is also referred to as *batch size*.
The weights can however not be adapted appropriately within one single step, rather multiple steps are necessary to reach the minimum and thus the optimal weight values. For each of these steps, the whole training dataset is used to calculate the gradient. One pass through the whole training dataset is in this context called an *epoch*.

Both optimizers are based on calculating the error gradient with respect to the individual weights. The idea is that, based on the gradient, a direction of steepest descent can be determined. If a step in this direction is taken in every iteration, the point of minimum error will be reached eventually. For this process, a *learning rate* ϵ defining the step size and a

decay parameter taking into account the flatter surface at the minimum are determined. These are both multiplied to the gradient. Picking an appropriate learning rate is crucial in this context as an ϵ value which is too small will result in a very long training process and an ϵ value chosen too large will increase the risk to diverge away from the minimum at some point.

In order to determine the gradient used by the optimizer, the *backpropagation algorithm* [92] has to be used as errors can only be determined for the output layer, but not for the hidden layers. This is why the error has to be propagated backwards, meaning from the output layer in the direction of the input layer.

4.1.3   Testing and its challenges: overtraining and undertraining

The fundamental requirement to NNs is a good performance not only for the training examples, but also for new, previously unseen data. To ensure this, the whole example dataset is divided into a training and a testing subset. The former is used to optimize the weights and the latter is then used to test the performance on unseen data, not used for training. The goal is that both, the training error itself as well as the difference to the testing error, should be as low as possible.

These two requirements are related to two central challenges in machine learning: *overtraining* and *undertraining*. These are visualized in figure 4.4. Here, the points represent the training examples with their respective labels indicated by the two colors. The shaded areas with different colors represent the approximated classification function of the Neural Network.

Overtraining means that too many details are taken into account in order to determine F^* as depicted in the left picture of figure 4.4. Here, almost every single point is described correctly by the learned/approximated function. Thus the model used is too complex and too adapted to the training data. This subsequently leads to a large difference between the training and the testing error when applied to new data.

Undertraining means that the model is not complex enough to describe the underlying properties of the input data appropriately as shown in the right picture of figure 4.4. Models suffering from undertraining usually have quite a high training error.

The central picture of figure 4.4 presents a visualization of a good model, which is just complex enough to take into account the underlying features of the input data.

In the framework used in this thesis, two methods are used to prevent the NN from overtraining: The first one is the so-called *dropout* [93]. Here, at each weight update, only a random subset of neurons is used for training as shown in figure 4.5. The size of the subset is defined in the form of a fraction p of neurons which stay active. As a consequence the network is forced to perform well also in the absence of certain information which prevents the network from becoming too dependent on some neurons. As for testing all neurons are used, every active neuron has its output divided by the fraction p during training before

the value is propagated to the next layer. This is essential in order to stay consistent.
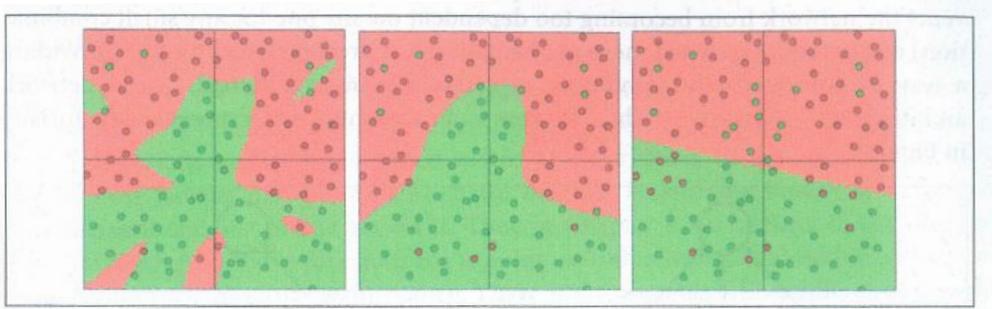
Figure 4.4: Visualization of a good model (center) as well as an overtrained (left) and an undertrained model (right). (adapted from [78])
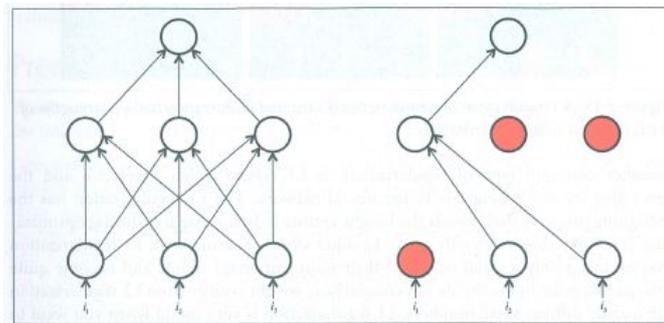
Figure 4.5: Visualization of dropout. (from [78])

The second method to prevent overtraining is *regularization* which describes any modification to a learning algorithm that is intended to reduce its testing error. In this thesis, the L1 regularizer [94] is used which adds a penalty term of the form $\lambda|w|$ for every weight w to the loss function in order to penalize large weights. This regularizer has the property that it leads to weight vectors staying close to zero during optimization. This is particularly useful in order to understand exactly which features are contributing to a decision. Choosing an appropriate value for λ is crucial in this context as $\lambda = 0$ implies that no measures to prevent potential overtraining are taken, while a λ value which is chosen too large will prioritize to keep the weights as small as possible over trying to find the parameter values that perform well on the training set.

4.1.4 Hyperparameters

In order to define an appropriate NN for the classification task, some parameters have to be specified by the user beforehand. These are called *hyperparameters*. The Neural Network

used in this thesis is defined by the following hyperparameters:

Hyperparameter	Description
neurons per layer	number of neurons in a hidden layer; the number of layers is the number of comma-separated numbers
ncycles	maximal number of epochs used for training
batch size	batch size used for updating the weights
dropout	dropout fraction
optimizer	either stochastic gradient descent or adam optimizer
activation	activation function, either sigmoid, leaky ReLu or tanh
initializer	random-number distribution used to initialize the weigths
regularizer	value for λ
class weight	weight to balance the number of signal/background events
learning rate	value for ϵ
decay	value for the decay parameter

Table 4.1: Description of the hyperparameters used in this thesis.

4.2   Evaluation of training/testing performance

To evaluate the performance of the training and testing processes of the NN, multiple approaches exist to visualize the separation power regarding signal and background as well as the degree of overtraining. Those used in this thesis are presented in the following.

4.2.1   Train-test plot

In the train-test plot the output score or Event-Probability-Discriminator (EPD) is plotted against the normalized number of events as displayed in figure 4.6.
In general, the output-score lies within a range of 0 (background) and 1 (signal). Thus, an event situated close to 1 is 'signal-like' and an event situated close to 0 is 'background-like'. The distributions showing the results of the training process are indicated by the filled areas. Blue (red) represents the training distributions of the background (signal). The results of the testing process are visualized using points in blue for background events and in red for the signal events. Furthermore, the ratio of signal to background is depicted per bin in the bottom panel. The bins for which no point is visible exhibit a signal-to-background ratio greater than 2.

Using this plot, the following two conclusions can be drawn:

1. **Performance of signal-background separation**: For a perfect separation, all the background events are situated close to 0 and all the signal events are situated close to 1. Consequently, the overlap between the distribution should be close to 0. However, one has to be cautious and make sure that the signal-to-background ratio has an increasing tendency towards the signal bin. The separation power in figure 4.6 seems to be quite perfect.

2. **Overtraining**: In this plot, also the amount of overtraining is visualized in the form of the distance between the testing distributions (points) and the training distributions (filled areas). The less they agree, the more overtraining has been conducted. In figure 4.6 one can see that the training and testing distributions match very well and thus no sign of overtraining is observed. In order to check for overtraining, also the scores of the Kolmogorov-Smirnov (KS) test [95, 96] can be considered which are given in the upper right corner. This is a statistical test to check for the difference in the empirical cumulative distributions of two samples, in this case the test and training subset [97]. The KS scores given are both very good [74].
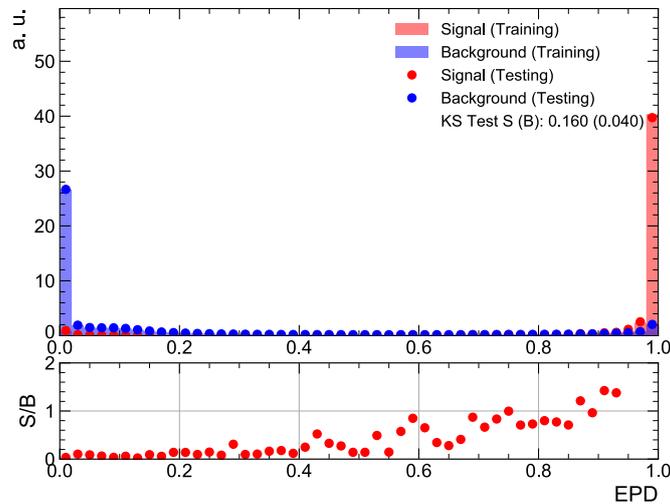
Figure 4.6: Example of a train-test plot. Details can be found in the text.

4.2.2 ROC curve and AUC value

Another performance indicator is the *Receiver-Operating-Characteristic curve* (ROC curve) which shows the background rejection, defined as one minus the background efficiency, depending on the signal efficiency. A perfectly performing NN has a rectangular ROC shape, going from (0,1) to (1,1) and then to (1,0), while a random decision would be

indicated by a diagonal, i.e. going from (0,1) directly to (1,0). In order to be able to evaluate the performance of the NN, usually the *Area-Under-Curve value* (AUC value) is calculated. An AUC value of 1 indicates a perfectly performing NN regarding signal-background separation, while a value of 0.5 would indicate that the performance of the NN is not better than a random decision.

This is depicted in figure 4.7 for the training (blue) and the testing (red) process.
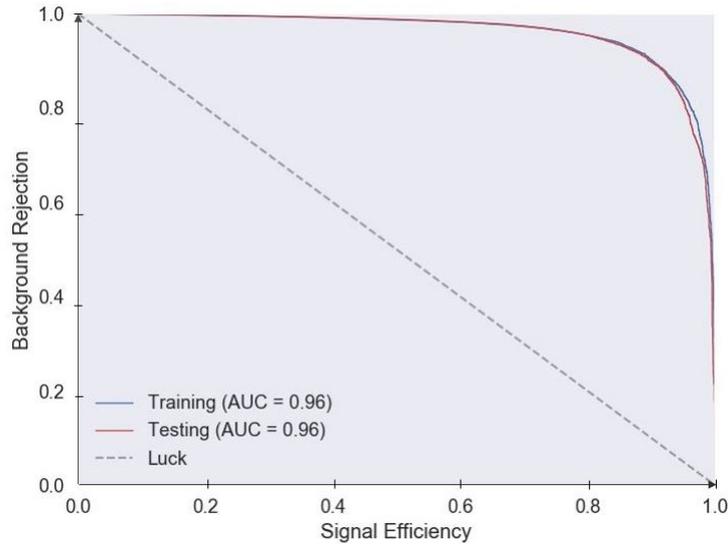
Figure 4.7: Example of a ROC-curve plot.

Here, both ROC curves have almost rectangular shapes and AUC values close to 1. This indicates a well-performing NN regarding its separation power between signal and background.

From this plot the degree of overtraining can be estimated by looking at the discrepancy of the two ROC curves. In figure 4.7 the ROC curves of training and testing almost completely superimpose, which indicates a small amount of overtraining.

4.2.3 Confusion matrices

Another way of visualizing the separation power between signal and background and the degree of overtraining of the NN are confusion matrices. They have the form displayed in table 4.2.

Here, the shares of events for which the predicted and the true labels agree, i.e. the so-called true positives and true negatives, are displayed on the diagonal. The off-diagonal entries exhibit the shares of events having different predicted and true labels, i.e. the false positive and false negative events. Ideally this matrix is the unity matrix. Deviations mean that the signal-background separation is not perfect: The higher the deviation, the worse

	classifier signal (EPD $\geq$ 0.5)	classifier background (EPD $<$ 0.5)
true signal	share of true positives	share of false negatives
true background	share of false positives	share of true negatives

Table 4.2: Visualization of the form of a confusion matrix.

the separation power of the NN.

This matrix is usually plotted separately for the training and the testing process. Comparing the shares between them gives an estimation for the amount of overtraining: the larger the deviation, the higher the degree of overtraining.

4.2.4 Variable correlations and ranking

To investigate which observables contribute the most during training, the linear correlations between the different observables as well as the linear correlation between the input variables and the output-score EPD is calculated. For this purpose, the linear correlation coefficient is used which is defined as

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \cdot \sigma_Y} \ [98]$$

In this definition, the covariance (cov) of the two considered observables X and Y is calculated. σ_X and σ_Y refer to the standard deviations of the two observables. The results ranging from -1 to 1 can then be displayed in form of a matrix, separately for signal and background events.

In general, a strong correlation of the observable to the output-score is desired. This indicates that the observable is important for the NN in the learning process. Furthermore, a low correlation among the input variables is desired as this indicates that a lot of new information is provided by the different input variables.

To gain further insights, the input variables can also be ranked. For that purpose, different metrics, like e.g. the sum of the absolute weights between the input layer and the first hidden layer, exist.

Chapter 5

Event selection and background estimation

Having gathered the necessary theoretical knowledge in the last chapters, the focus is now set on the development of a multivariate reconstruction alogrithm for $t\bar{t}$ systems in the 6j2b-2LOS channel of the $t\bar{t}Z$ analysis. The events used in the following are required to pass the following criteria:

- only events listed in the *GoodRunsLists* of 2015 and 2016 data [99] [100] are considered

- *GoodCalo*: LAr and Tile Calorimeters need to work properly for the event [74] [101]

- events need to pass various single electron and muon *triggers* [74]

- *JetCleaning:* 'bad jets' need to be excluded as they are not associated to real energy deposits in the calorimeter, but rather arise from other sources like hardware problems or cosmic ray showers [74] [102]

- the events considered need to have a *primary vertex*

- *leptons*: In general, muons/taus need to be situated within $|\eta| < 2.5$ and electrons need to be situated within $|\eta| < 2.47$. The truth electrons and muons are only required to have a transverse momentum > 7 GeV (25 GeV for taus).
  Within this analysis, the following additional cuts for leptons are set (analogous to [25]): The event has to contain exactly 2 leptons with same sign and opposite flavor, i.e. 2 electrons and no muons or taus and analogous for the other two leptons on truth level. The 2 leptons need to have a transverse momentum of 30 GeV and 15 GeV respectively and their invariant mass has to be close to the mass of the Z boson: $|M_Z - M_{ll}| < 10$ GeV. The lepton cuts listed up to here are part of the cuts referred to as 'analysis cuts' later on. If not specified, leptons will refer to only electrons and muons in the following studies.

- *jets*: If not stated otherwise, the events considered need to contain ≥ 6 jets and ≥ 2 b jets (analogous to [25]) with a transverse momentum of > 25 GeV. These requirements are also included in the term 'analysis cuts' in the following.

- to assure that only the *signal process* is considered, i.e. a hadronic $t\bar{t}$ decay and the Z boson decaying to ee/$\mu\mu$, further signal-region (SR) cuts are set: the Monte-Carlo sample including the $Z \rightarrow \tau\tau$ decay is excluded and only events containing hadronically decaying $t\bar{t}$ systems on truth level are accepted.

As mentioned in section 2.2, the 2LOS-channel is characterized by its large background rate. This results mainly from the associated production of a Z boson and additional jets (Z+jets) as well as the dileptonic $t\bar{t}$ decay, as both of them exhibit a similar signature as the signal $t\bar{t}Z$ process [25]. The impact of the background is estimated in the following based on the event yields after application of the mentioned cuts without the signal-region cuts. These are loosened as there should be no hadronic $t\bar{t}$ systems in the background processes. The denoted uncertainty is the square root of the sum of the squared, weighted event yields over all bins. The relative shares are visualized in a pie chart scaled to 140 fb^{-1} corresponding to the amount of data taken during the full Run-2 period. The event yields and the pie chart can be found in figure 5.1.
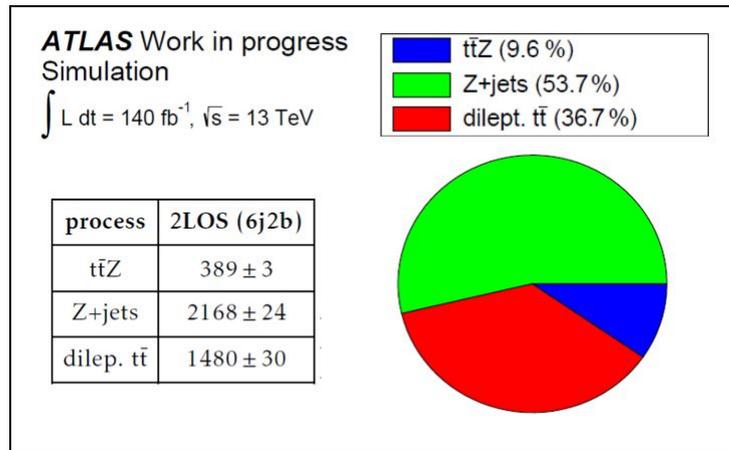
Figure 5.1: Visualization of weighted event yields along with the relative shares of the signal $t\bar{t}Z$ process and the two major backgrounds of the $t\bar{t}Z$-2LOS-6j2b region (section 2.2), the Z+jets process and dileptonic $t\bar{t}$ decays [25].

The $t\bar{t}Z$ process includes in this case all decay modes of the $t\bar{t}$ system and a Z boson decaying leptonically, thus either to ee, $\mu\mu$ or $\tau\tau$ [1]. The same decay modes for the Z boson are simulated within the Z+jets sample. As the name already indicates, the dileptonic $t\bar{t}$

---

[1]The hadronic Z boson decay is excluded due to its negligible impact. Z decaying to $\nu\nu$ could not be considered due to problems with the simulation dataset.

sample contains exclusively dileptonic $t\bar{t}$ decays.

The $t\bar{t}$Z events represent the smallest share, with 385 weighted events and a corresponding relative contribution of 9.6%. The largest share (53.7%) is represented by the Z+jets process with 2151 weighted events. The event yield of the dileptonic $t\bar{t}$ decay is with 1468 weighted events, corresponding to a relative share of 36.7%, situated between the other two. This visualizes the challenge of the $t\bar{t}$Z process mentioned in section 2.2: This process is very rare and has a very small cross section, so the 'contamination' with background processes is quite high [2].

---

[2]To further reduce the background contamination, the official $t\bar{t}$Z analysis cuts also imply a cut on the Boosted Decision Tree (BDT) output value. As within the scope of this thesis, most studies are conducted only for the $t\bar{t}$Z process, this is not used.

Chapter 6

Reconstruction studies using the minimum-χ^2 reconstruction

Since the discovery of the top quark at the Tevatron, steadily increasing center-of-mass energies and integrated luminosities allowed for investigation of signatures at ever higher mass scales. This also led to the discovery of the associated production of a top-antitop-quark system with a Z boson ($t\bar{t}Z$ system) in 2015. [23, 24]
As a basis for conducting precision measurements of its properties such as the cross section, a good reconstruction of these systems is indispensable. This implies selecting the 6 jets originating from the $t\bar{t}$ system among the multiple physical objects produced in a high number of interactions. As this is quite a challenging task and a lot of different hypotheses have to be considered in order to find the one with a well-reconstructed $t\bar{t}$ system, dedicated procedures have been developed and established.

One of them is the minimum-χ^2 reconstruction algorithm which will be described and investigated in the following sections using MC-simulated data. If not stated otherwise, the histograms shown in this chapter contain raw, i.e. un-weighted, events.

6.1 Standard reconstruction technique: the minimum-χ^2 method

In this section, the standard minimum-χ^2 reconstruction is described and applied to the $t\bar{t}$ system in $t\bar{t}Z$ decays. To evaluate its discrimination power regarding background processes, this method is also applied to the two major backgrounds of the 2LOS-$t\bar{t}Z$-analysis channel for the 6j2b signal region: the Z+jets and dileptonically decaying $t\bar{t}$ systems [25]. To quantify the performance for $t\bar{t}Z$ events, the χ^2-reconstructed $t\bar{t}$ systems are compared to the ones on truth level in a procedure called truth matching. Furthermore, they are also compared to all hypotheses considered during reconstruction to get more detailed in-

formation about the performance of the minimum-χ^2 algorithm.

6.1.1   The minimum-χ^2 reconstruction algorithm

The standard reconstruction method for $t\bar{t}$ systems is the so-called minimum-χ^2 method. There, the jets of one event are grouped together in groups of 6 jets under the condition that exactly 2 of them have a b-tag. Then, for each of these groups, every combination of these 6 jets to obtain a $t\bar{t}$ system is formed. This is done by iteratively considering all possibilities to place the selected jets on the positions of the 6 decay products of the $t\bar{t}$ system as visualized in figure 6.1. By name, these are the b position, the $\bar{b}$ position, the position of the first/second decay product of the W boson originating from the top quark (W1t/W2t) and the position of the first/second decay product of the W boson originating from the anti-top quark (W1$\bar{t}$/W2$\bar{t}$). On truth level, the W1$\bar{t}$ quark and the W2t quark are down-type quarks while the W2$\bar{t}$ and W1t quarks are up-type ones.
Subsequently, only combinations with a b-tagged jet placed on the b/$\bar{b}$ position, highlighted in grey in the table in figure 6.1, are considered.

b	$\bar{b}$	W1t	W2t	W1$\bar{t}$	W2$\bar{t}$
jet 1	jet 2	jet 3	jet 4	jet 5	jet 6
jet 2	jet 1	jet 3	jet 4	jet 5	jet 6
...	...	...	...	...	...
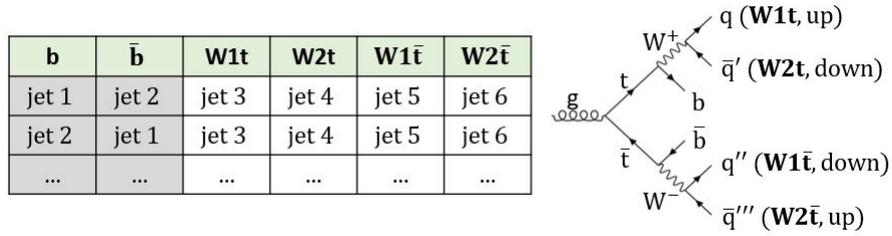
Figure 6.1: Visualization of the minimum-χ^2 procedure described in the text. In this context, W1t (W2t) is the first (second) decay product of the W boson originating from the top quark and W1$\bar{t}$ (W2$\bar{t}$) is the first (second) decay product of the W boson originating from the anti-top quark. On truth level, W1t and W2$\bar{t}$ are up-type quarks while W2t and W1$\bar{t}$ are down-type quarks.

For each of these combinations/hypotheses within one group of 6 jets and for all different groups, a χ^2 value is calculated which is defined as

$$\chi^2 = \left(\frac{m_{b,W1t,W2t} - m_t}{\sigma_t}\right)^2 + \left(\frac{m_{\bar{b},W1\bar{t}W2\bar{t}} - m_t}{\sigma_t}\right)^2 + \left(\frac{m_{W1t,W2t} - m_W}{\sigma_W}\right)^2 + \left(\frac{m_{W1\bar{t},W2\bar{t}} - m_W}{\sigma_W}\right)^2$$

Here, m_i represents the invariant mass of the objects i, m_t is the literature value of the top-quark mass of 173.1 GeV [103], σ_t is the width of the top-quark mass of 1.41 GeV [103], m_W is the mass of the W boson of 80.4 GeV [103] and σ_W is the width of the W-boson mass of 2.1 GeV [103].

Finally, the combination/hypothesis which has the overall minimal χ^2 value is decided to be the reconstructed $t\bar{t}$ system. This procedure is repeated for all the events.

6.1.2 Signal-background discrimination using the minimum-χ^2 reconstruction

In order to evaluate the discrimination power of the minimum-χ^2 method regarding the signal and the background processes, it is applied not only to the $t\bar{t}Z$ process, but also to the two major backgrounds of the 2LOS-6j2b region, the dileptonic $t\bar{t}$ decay and the Z+jets process [25]. In this context, the same event selection is applied as for the pie chart in chapter 5.

In a first step, the χ^2 value is plotted in figure 6.2 for the $t\bar{t}Z$ process (blue), for the Z+jets process (red) and for the dileptonic $t\bar{t}$ decay (yellow) scaled to a luminosity of 140 fb^{-1}. The normalized results are depicted with event weights applied.
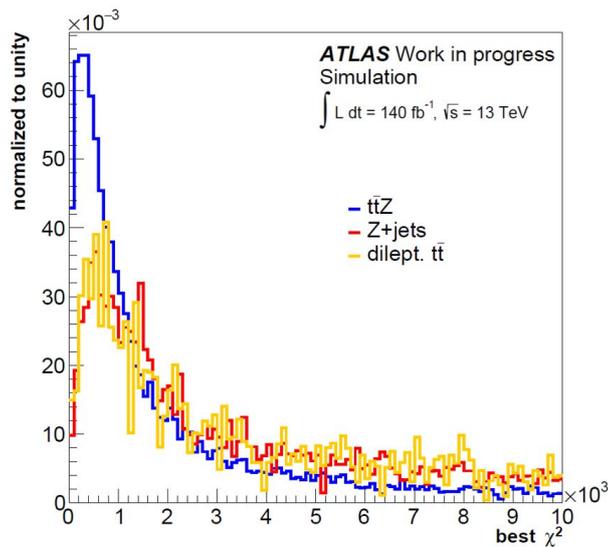
Figure 6.2: Normalized and weighted χ^2-value distribution of the best hypotheses selected by the minimum-χ^2 algorithm for the $t\bar{t}Z$ process (blue), for Z+jets (red) and for the dileptonic $t\bar{t}$ decay (yellow).

In this plot, all distributions have a peak at low χ^2 values and a long tail towards higher values. The $t\bar{t}Z$ process has a significantly higher and slightly sharper peak with a narrower tail compared to the other two. The background processes show the same basic shape within statistical fluctuations. All in all, there are more entries at low χ^2 values for the $t\bar{t}Z$ process which implies a better reconstruction performance as expected.

Furthermore, the reconstructed top-quark and W-boson masses are compared in order to check the reproduction performance of physical properties. The weighted and normalized results are scaled to 140 fb^{-1} and are depicted in figure 6.3 for the three different processes in the same colors as above.

(a)   (b)
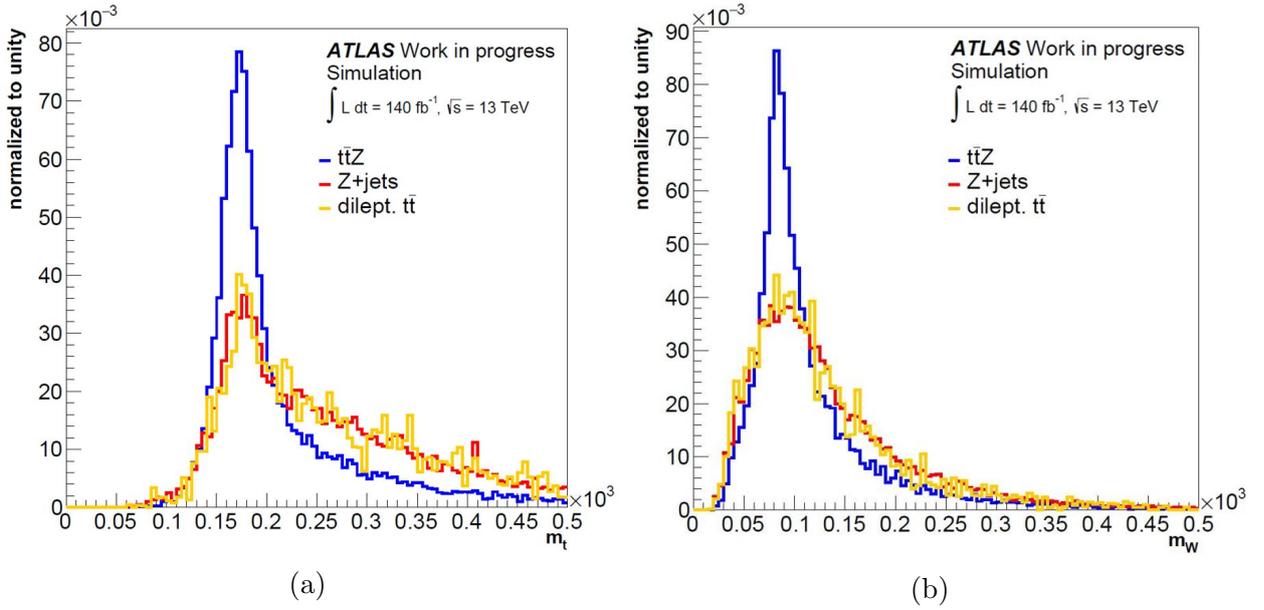
Figure 6.3: Normalized and weighted mass distributions of the reconstructed top-quark masses (a)) and the W-boson masses (b)) of the best χ^2 hypothesis for the signal $t\bar{t}Z$ process (blue) as well as for the two major backgrounds Z+jets (red) and the dileptonic $t\bar{t}$ decay (yellow).

In figure 6.3a the distributions of the reconstructed top-/anti-top-quark mass are displayed, while figure 6.3b depicts the reconstructed W-boson masses. For both plots only the best χ^2 hypothesis of each considered event has been used.
The highest and sharpest peaks situated at a value of about 170 GeV in figure 6.3a (80 GeV in figure 6.3b) with a small tail towards higher mass values, belongs to the $t\bar{t}Z$ process. The two background processes exhibit comparably lower peaks with broader tails. The tails of the $t\bar{t}Z$ process are caused by either choosing wrong jet combinations for the reconstruction of the top quark/W boson or jets that do not originate from these objects. In the two background processes the broad tails arise due to the non-existence of hadronic $t\bar{t}$ systems. Thus, the jets used for reconstruction do not originate from a real top quark/W boson which leads to the observed tails.
The fact that all three processes have a significant peak at a value of about 170 GeV/80 GeV might seem a bit surprising at first sight. However, taking a closer look at the definition of the χ^2 value, it can be observed that only the differences of the reconstructed masses to the literature values are used in the formula. That is why the resulting mass distributions

are pulled towards these literature values and thus all distributions show a peak there. Nevertheless, the best result is clearly achieved by the t$\bar{\text{t}}$Z (signal-)process as expected. Compared to the top-quark mass distributions, the shapes of the W-boson masses are a bit broader. This is a consequence of the increased number of combinations for light jets compared to the b jets. For the latter, the b-tagging requirement needs to be fulfilled in addition which reduces the number of possible combinations.

Concluding, it can be stated that the χ^2 reconstruction procedure performs better for the t$\bar{\text{t}}$Z process than for the two major backgrounds as expected. Thus, the χ^2 value can be used as a variable for discriminating the signal t$\bar{\text{t}}$Z process from the major backgrounds.

6.1.3 Truth matching

In order to quantify how well the χ^2 reconstruction works for t$\bar{\text{t}}$Z compared to the parton-level truth information, the selected jets of the best χ^2-reconstructed t$\bar{\text{t}}$ system are matched with the truth-quark positions.
On parton/jet level, a match is present if $\Delta R(\text{jet,quark}) < 0.3$ as shown in figure 6.4. Otherwise the jet is referred to as unmatched.
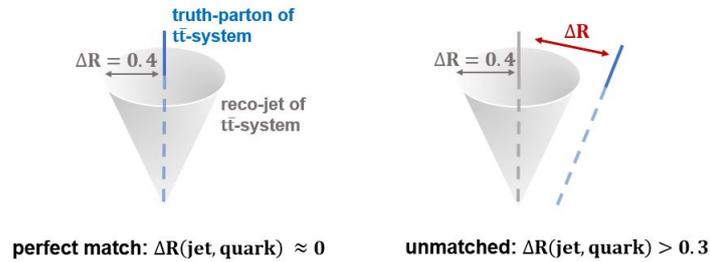
Figure 6.4: Visualization of the definition of a matched jet. In this context, the jets are approximated to be cone-shaped. The truth partons are depicted as blue lines.

Based on this definition three different categories for the whole t$\bar{\text{t}}$ system are defined: 'matched', 'combinatorial' and 'non-matched'. They are visualized in figure 6.5 and are described below:

- **matched:** The reconstructed t$\bar{\text{t}}$ system belongs to the 'matched' category if two conditions are fulfilled: First, every χ^2-reconstructed jet has to match a truth parton of the t$\bar{\text{t}}$ system, i.e. $\Delta R(\text{jet, parton}) < 0.3$ for all 6 χ^2-reconstructed jets is required. Furthermore, the assigned/matched parton to a particular jet also needs to be the corresponding one. For example, if the considered jet is the b jet, then the matched parton must be the b quark. If both these criteria are fulfilled, the t$\bar{\text{t}}$ system is assigned to this category. This situation is visualized in sketch 6.5a.
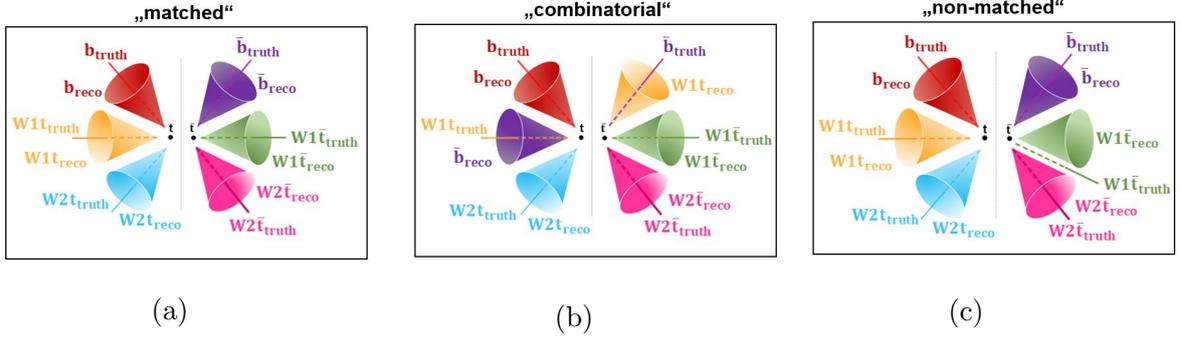
Figure 6.5: Sktech of the three different truth-matching categories, 'matched' (a), 'combinatorial' (b) and 'non-matched' (c). In this context, the jets are approximated to be cone-shaped. The truth partons are depicted as lines. See text for more details.

- **combinatorial:** In order for the $t\bar{t}$ system to be assigned to this category, every χ^2-reconstructed jet needs to match a truth parton, i.e. all $\Delta R(\text{jet, parton})$ values need to be smaller than 0.3. However, the assigned parton is, in at least one case, not the corresponding one. For example, the $\bar{b}$ jet is assigned to the W1t parton and vice versa as depicted in sketch 6.5b.

- **non-matched:** In this category, at least one jet of the χ^2-reconstructed $t\bar{t}$ system cannot be matched, i.e. $\Delta R(\text{jet, parton}) \geq 0.3$. This is visualized in sketch 6.5c, in which the W1$\bar{t}$ jet is unmatched.

Applying this truth-matching procedure to the $t\bar{t}Z$ events fulfilling all requirements of chapter 5 yields the following result:

matched	combinatorial	non-matched
1.1 %	2.9 %	96.0 %

The low portion of 1.1% of reconstructed $t\bar{t}$ systems in the matched category and the extremely high share of 96% of non-matched events, show very clearly that something goes wrong. Apparently, the reconstruction implies some challenges which have not been considered in the reconstruction procedure. To identify and understand these, more detailed investigation studies need to be conducted.

6.1.4 Performance investigation of the minimum-χ^2 reconstruction for $t\bar{t}Z$ events

Having found in the previous chapter that the χ^2 reconstruction is not working properly, the performance of the χ^2 reconstruction is evaluated in more detail in this subsection.

To reduce the probability that jets not belonging to the $t\bar{t}$ system are erroneously considered for reconstruction, a cut on exactly 6 jets and exactly 2 b jets is set in addition to the other requirements presented in chapter 5.

Compared to the event yields found in figure 5.1, the stricter cut on exactly 6 jets and exactly 2 b jets implies a reduction by 56 %. The constraint on the signal process, thus a hadronically decaying $t\bar{t}$ system with a Z boson decaying to ee/$\mu\mu$, implies a reduction of 18.4%. As a consequence, this means that the following studies based on a combination of these cuts are conducted using 34% of the event yield presented in figure 5.1.

As the following studies only consider the $t\bar{t}Z$ process, the histograms in this section contain raw events, i.e. no event weights have been applied. This is allowed as only the $t\bar{t}Z$ process, i.e. one physical process, is considered.

At first, the χ^2 values of the selected best hypotheses are compared to the values of all considered hypotheses. Figure 6.6 depicts the resulting normalized distributions for the best χ^2 hypotheses (blue) and all hypotheses (red).
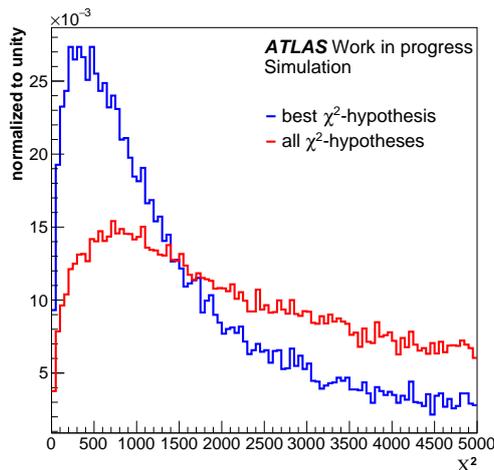
Figure 6.6: Normalized distributions of the χ^2 value for the best χ^2 hypotheses (blue) and all considered hypotheses (red).

In this plot, a clear separation between the two distributions can be identified. While the distribution of the best χ^2 hypotheses has a relatively sharp peak at low values and a comparably small tail towards higher values, the other distribution shows a small peak at low values followed by a broad shoulder. Thus, it can be concluded that, indeed, the χ^2 reconstruction selects 'better' hypotheses, i.e. those with lower χ^2 values.

To get further insights on how well physical observables are reproduced, the masses of the

reconstructed top quark and the W boson are investigated as well. Figure 6.7 shows the normalized distributions for the best χ^2 hypotheses (blue) and all hypotheses (red).
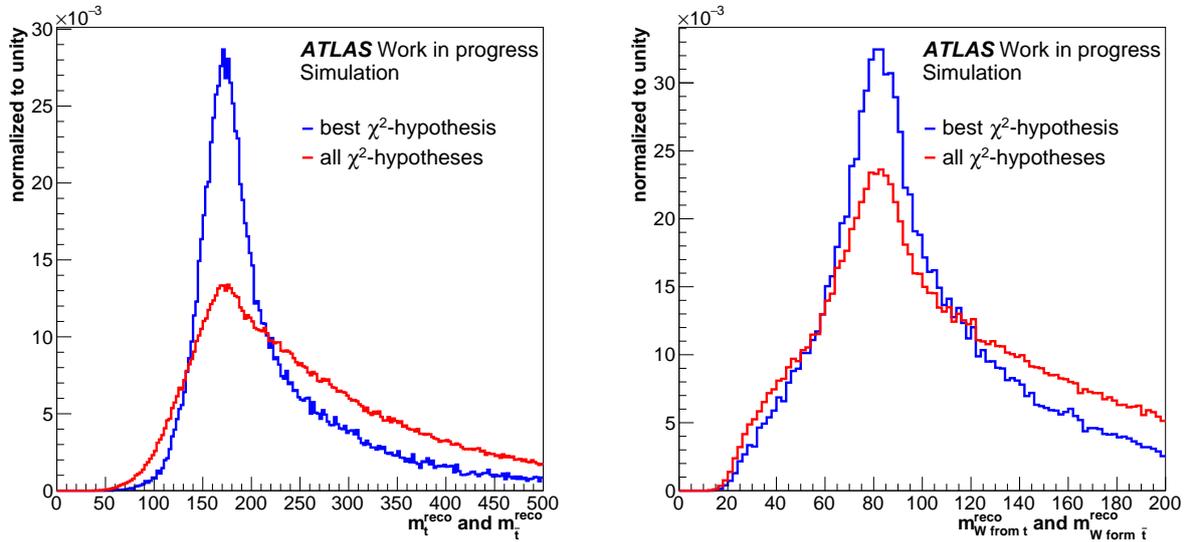
Figure 6.7: Normalized distributions of the reconstructed top-quark mass (a)) and the reconstructed W-boson mass (b)) for the best χ^2 hypothesis (blue) and the non-selected hypotheses (red).

Figure 6.7a shows the normalized distributions of the top-quark mass. The selected hypotheses have a well-defined, sharp peak situated at a value of about 170 GeV and exhibit a small tail towards higher values resulting from wrongly selected combinations. The distribution which represents all hypotheses has also a small peak at the same value due to the nature of the χ^2 definition which pulls the distribution towards the literature value. Furthermore, the distribution of all hypotheses shows a broader tail towards both lower and higher values in comparison to the one representing the best hypotheses. The broader tail towards higher values is a result of choosing a wrong combination of jets for the reconstructed $t\bar{t}$ system. The tail towards lower values is caused by choosing jets for the reconstruction which do not originate from the $t\bar{t}$ system but instead e.g. from other processes/pile up/underlying event.

In figure 6.7b the reconstructed W-boson masses are shown. Here, the same trends can be found as for the top-quark masses. However, the difference between the normalized distributions of the selected and all hypotheses is reduced significantly.

The fact, that in both cases the tails are broader, shows that the selection of the correct jets as well as placing them on the corresponding positions works better for the best χ^2 hypotheses.

As a conclusion, it can be stated that the physical properties are better reproduced for the selected χ^2 hypotheses. However, the separation is, especially for the W-boson mass, not

perfect. This indicates that there seem to be some challenges with which the reconstruction algorithm has to cope. This becomes obvious by applying the truth matching for the event selection presented in chapter 5 with the stricter cut on exactly 6 jets and exactly 2 b jets. It leads to the following result:

matched	combinatorial	non-matched
1.6 %	4.3 %	94.1 %

This displays that the matching share is improved by focusing on the signal process by applying a stricter cut on the number of jets/b jets. However, a share of 1.6% of well-reconstructed $t\bar{t}$ systems is still an extremely bad result.

In this context, also the overall matchable share, i.e. the ratio $\frac{\text{matched}}{\text{matched+combinatorial}}$, can be considered. The assumption is that only 5.9% of the events contain a fully reconstructable $t\bar{t}$ system as every jet was able to be matched according to the category definition. Thus, $\frac{1.6\%}{5.9\%} = 27\%$ of those events have been correctly reconstructed. However, this share is still quite low.

6.2 Investigation studies of non-matched events

As shown in the previous section, the reconstruction of hadronically decaying $t\bar{t}$ systems in $t\bar{t}Z$ events is not a trivial task and faces some challenges. In order to obtain a deeper understanding of those, detailed investigation studies are conducted in this section.

Analogously to the previous subsection, a cut on exactly 6 jets and exactly 2 b jets is set on top of the requirements presented in chapter 5. This aims at reducing the probability that jets not originating from the $t\bar{t}$ system are erroneously considered for reconstruction. As the following studies only consider the $t\bar{t}Z$ process, the histograms contain only raw events, i.e. no event weights have been applied.

To identify the challenges in reconstructing hadronic $t\bar{t}$ systems within $t\bar{t}Z$ events, the χ^2-reconstructed $t\bar{t}$ systems of the non-matched category are used to conduct diverse investigation studies.

In this context, at first, the question of a jet-type depending matching performance is targeted. Furthermore, flavor and ΔR studies are conducted and it is studied, if the choice of the ΔR value of 0.3 used in the truth matching is justified. Finally, the impact of phase-space adaption between truth and reco level as well as events containing partons situated within $\Delta R < 0.8$ to each other are studied.

6.2.1 Jet-type depending matching performance

In a first step, it is investigated if the reconstruction works better for a certain type of jet. This is done using the events in which the χ^2-reconstructed $t\bar{t}$ system is assigned to the

non-matched category during truth matching. Using these, it is counted how often for which of the 6 jets the ΔR(jet, parton) is ≥ 0.3. The result is shown in figure 6.8.
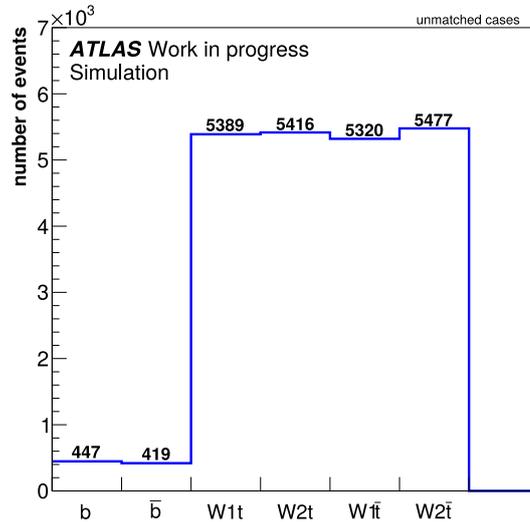
Figure 6.8: Number of events in which one of the jets shown on the x-axis cannot be matched (ΔR < 0.3) to a truth parton considering events of the non-matched category (subsection 6.1.3).

The resulting number for the b jets is with about 500 events roughly a factor 10 lower than for the light jets (W1t, W2t, W1$\bar{t}$, W2$\bar{t}$) with about 5500 events. As a lower number of unmatched b jets implies a larger number of matched ones, it can be concluded that the reconstruction performance is better for the b jets than for the light jets. The reason is, that the jets placed on the b-jet positions by the minimum-χ^2 method were required to have a b-tag. Hence, more information is used for the b jets during reconstruction which in consequence leads to a better matching/reconstruction performance. The slight differences between the light jets do not have a physical reason, but rather result from statistical fluctuations.

6.2.2 ΔR and flavor studies

In a second step, a better understanding of the ΔR distance between the χ^2-selected reco jets and the truth partons is developed and also the flavors of the selected reco jets are studied in the context of ΔR and flavor studies.
For that purpose the focus is shifted to the particular jets of the χ^2-reconstructed t$\bar{t}$ system in non-matched events which exhibit a ΔR(jet, parton) ≥ 0.3 to the assigned truth parton. If the particular jet is e.g. the b jet, this is in the following referred to as an unmatched b case. The following studies are then conducted for the respective unmatched cases, e.g.

studying the flavor of the b jet is performed for unmatched b cases and so on.

At first, the ΔR distance between the reco jet and the corresponding truth parton is investigated. Figure 6.9 depicts the resulting normalized distributions for the different reco jets W1t (blue), W2t (red), W1$\bar{\text{t}}$ (yellow), W2$\bar{\text{t}}$ (green), b (purple) and $\bar{\text{b}}$ (black) for the respective unmatched cases of the non-matched events.
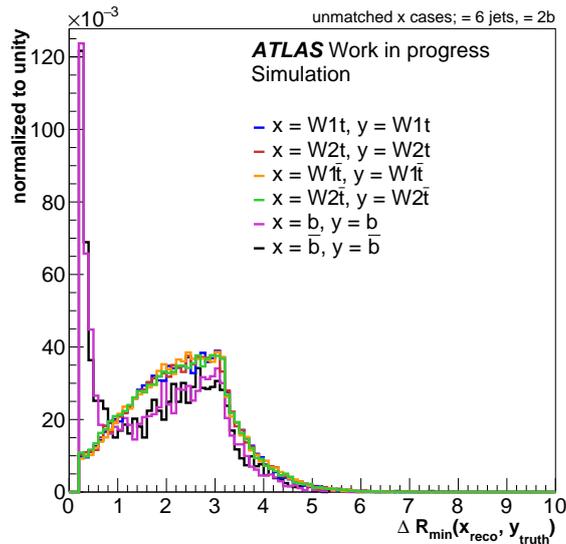
Figure 6.9: Normalized distributions of ΔR between the reco-jet x and their corresponding truth-parton y for the respective unmatched cases of the events of the non-matched category (subsection 6.1.3); x either represents the W1t jet, the W2t jet, the W1$\bar{\text{t}}$ jet, the W2$\bar{\text{t}}$ jet, the b jet or the $\bar{\text{b}}$ jet.

For the b and the $\bar{\text{b}}$ jets, the ΔR distance to the corresponding parton has a sharp peak just above the matching threshold of 0.3. Moving towards higher values, a broad shoulder with another maximum at about a ΔR value of 3 can be found. For the light jets, this can be observed as well. However, in comparison to the b/$\bar{\text{b}}$ distributions, there is no sharp peak just above the threshold value. In general, all distributions have a maximum ΔR less than 6.
This clearly displays that the unmatched b/$\bar{\text{b}}$ jets are closer to the corresponding truth-parton positions and thus a better reconstruction performance of b jets compared to light jets is indicated[1].

The above observations raise the question about the physical scenarios causing these

---

[1]Within the b-tagging procedure a cut on the ΔR between the jet and the b-vertex direction might have already been included. This could have had an impact on the analyzed distribution as well as on other distributions in this chapter.

shapes. In the next step, it is therefore taken into account that the χ^2 reconstruction algorithm has selected hypotheses in which the jets are placed on wrong positions, i.e. the b jet is assigned to the $\bar{\text{b}}$ position and vice versa and the light jets are switched among each other. Figure 6.10 displays the resulting normalized distributions of the minimum ΔR between the b/$\bar{\text{b}}$ jet and the b or $\bar{\text{b}}$ truth parton as well as the minimum ΔR between the light jets and all light truth partons. This is conducted for the respective unmatched cases of the non-matched events and depicted in the same color-coding as in figure 6.9.
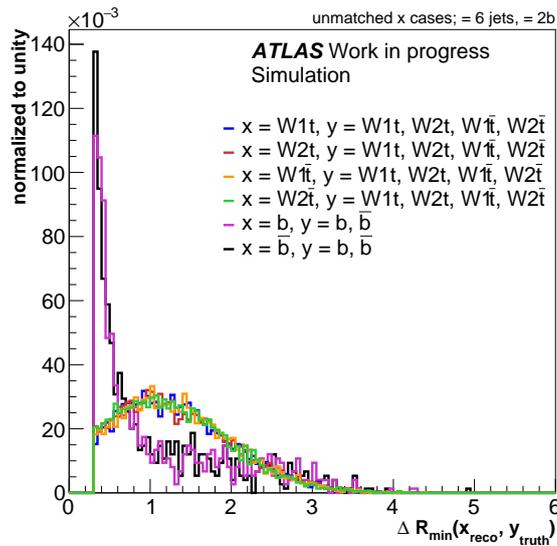
Figure 6.10: Normalized distribution of the minimum ΔR between a certain reco jet x and a truth-parton y for the respective unmatched reco-jet cases for events of the non-matched category (subsection 6.1.3); x represents either the W1t jet, the W2t jet, the W1$\bar{\text{t}}$ jet, the W2$\bar{\text{t}}$ jet, the b jet or the $\bar{\text{b}}$ jet while y represents either the light partons (W1t, W2t, W1$\bar{\text{t}}$, W2$\bar{\text{t}}$) or the b/$\bar{\text{b}}$ parton as indicated in the legend of the plot.

For the b-/$\bar{\text{b}}$-jet distributions, once again, a high, sharp peak just above the threshold can be observed which is not present in the light-jet distributions. While the b-jet distributions show a broad tail towards high ΔR values, a shoulder with a maximum value of about 1.5 can be found in the light-jet distributions.

In comparison to figure 6.9, all distributions are in general narrower. This manifests itself in a lower maximum value of 3.5 instead of 5 and a peak value for the light-jet distributions situated at about a ΔR value of 1.5 instead of 3. Moreover, also the significant shoulder found in figure 6.9 in the b/$\bar{\text{b}}$ cases vanishes and instead becomes a tail.

Thus, from these observations it can be concluded that one challenge in reconstruction is the switching among the light jets and the b/$\bar{\text{b}}$ jet respectively. Reducing this is not trivial as the mass value of the reconstructed t$\bar{\text{t}}$ system, the reconstructed top-quark mass as well as the reconstructed W-boson mass exhibit only minor changes for these switched cases.

A potential solution could be to include not only a b-tagging condition, but also jet charges

or conditions on other sensitive variables in the reconstruction procedure.

To push further, also the minimum distance between the reco jet at hand and all truth quarks is plotted for the respective unmatched cases for events of the non-matched category. The normalized distributions are shown in figure 6.11 in the same color-coding as in figure 6.9.
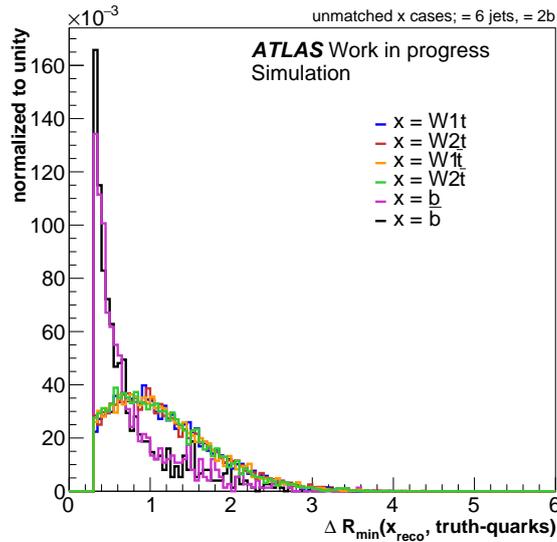
Figure 6.11: Normalized distribution of the minimum ΔR between a reco-jet x and all truth partons originating from the $t\bar{t}$ system for the respective unmatched cases of events of the non-matched category (subsection 6.1.3); x represents either the W1t jet, the W2t jet, the W1$\bar{t}$ jet, the W2$\bar{t}$ jet, the b jet or the $\bar{b}$ jet.

The same trends can be observed as in the previous scenario shown in figure 6.10: The distributions of b and $\bar{b}$ have a sharp peak at low ΔR values, just above the matching threshold of 0.3. At larger ΔR values a tail ending at a value of circa 3 can be observed. The light jets have a broader distribution with a peak at a value of about 1 and reach up to a maximum value of circa 3.
Compared to figure 6.10, the distributions are a lot sharper and end at lower values. Obviously, the corresponding parton or one of the same type is not always the best-fitting one, i.e. the one exhibiting the lowest ΔR(jet, parton) value.

To gain further information about this best-fitting parton, the absolute value of the pdgId[2] of the truth quark which gives the minimum ΔR between the reco jet at hand and all truth

---

[2]The pdgId gives the flavor of a particle with values according to a Monte-Carlo particle numbering scheme which can be found in [104]. Antiparticles have the same values, but with a negative sign.

partons in figure 6.11 is investigated. Figure 6.12 depicts the normalized result for the different unmatched cases for the events of the non-matched category in the same color-coding as in figure 6.9.
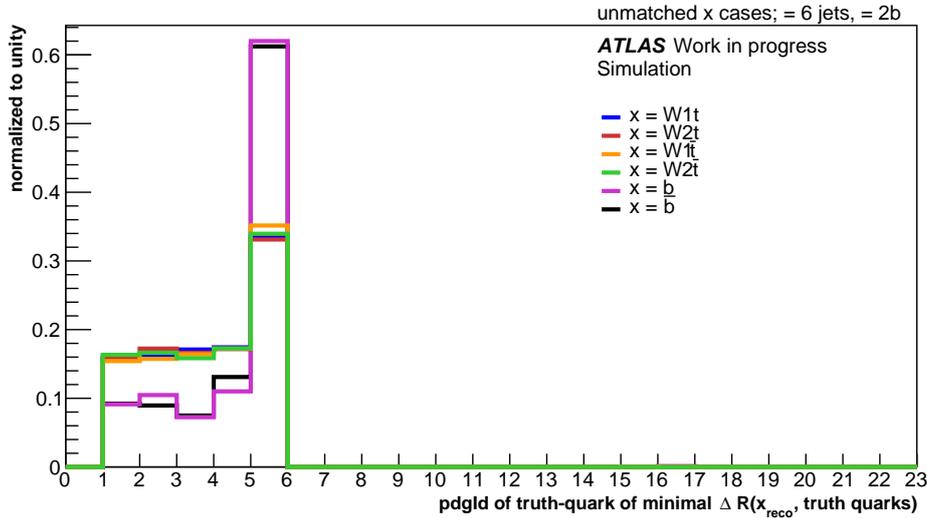
Figure 6.12: Normalized distribution of the absolute value of the pdgId of the truth quark which yields the minimum ΔR between a reco-jet x and all the truth quarks for the unmatched x cases for events of the non-matched category (subsection 6.1.3); x refers either to the W1t jet, the W2t jet, the W1$\bar{\text{t}}$ jet, the W2$\bar{\text{t}}$ jet, the b jet or the $\bar{\text{b}}$ jet.

For both, b and $\bar{\text{b}}$, the best-fitting truth partons are mostly those with an absolute value of the pdgId of 5, hence b quarks. Charm quarks (pdgId of 4) are a bit more often the best-fitting partons to the b-tagged jet positions compared to up, down and strange quarks (pdgID 1, 2, 3). This can be explained by considering misidentification during which non-b-jets are erroneously labeled with a b-tag. At the b-tagging working point used, this is the case for 16% of the charm quarks (subsection 3.2.5). This corresponds well to the value of about 12% in the plot. For the light jets the misidentification rate of 1% (subsection 3.2.5) is negligible and cannot explain the observed contribution of about 10%. It is more likely that the b and the light partons lie close together. In consequence, the b jet has a slightly smaller ΔR to one of the light partons than to the b quark. This situation will be referred to as b-light merging in the following.

Focusing on the light jets, a large contribution of best-fitting light partons can be found. As already outlined, switching among the light jets is quite frequent which results in equalized contributions of up- and down-type quarks (1, 2, 3, 4). Furthermore, the share of 35% of b quarks which fit best to the light jets (W1t, W2t, W1$\bar{\text{t}}$ and W2$\bar{\text{t}}$) is eye-catching. This is once again attributable to the phenomenon of b-light merging.

Concluding, it seems that b-light merging represents an additional challenge in the recon-

struction of hadronic $t\bar{t}$ systems in $t\bar{t}Z$ events.

This assumption is verified by considering the minimal ΔR value between the b and light truth partons for the different unmatched cases for events of the non-matched category. Figure 6.13a depicts the normalized result in the same color-coding as in figure 6.9.

(a)

(b)
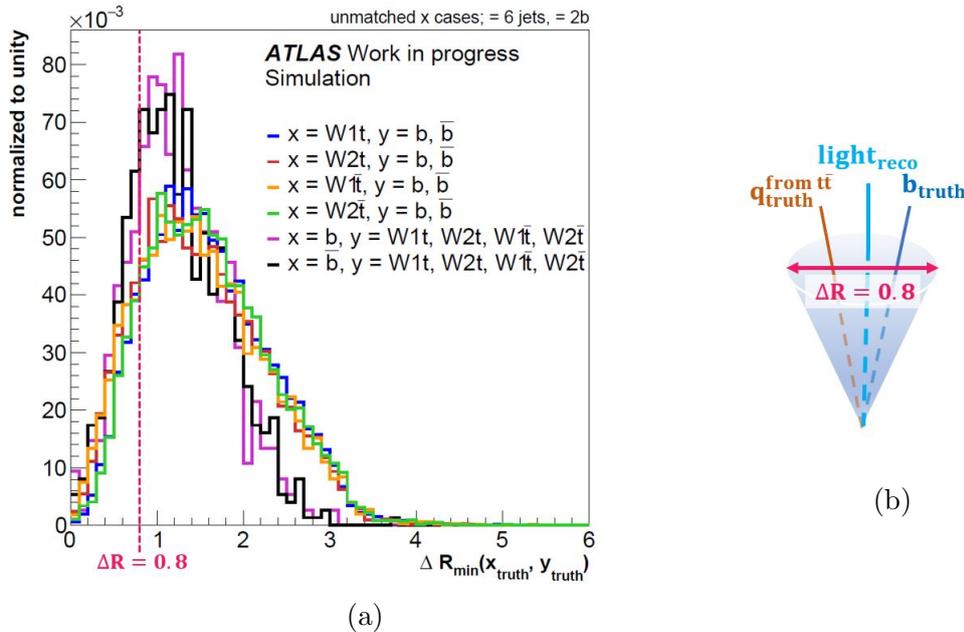
Figure 6.13: Normalized distribution of the minimum ΔR between a truth-parton x and a truth-parton y for the corresponding unmatched x reco-jet case of the events of the non-matched category (subsection 6.1.3) (a) as well as a sketch of b-light merging (b) corresponding to entries below a value of $\Delta R = 0.8$; x in this case can either be the W1t parton, the W2t parton, the W1$\bar{t}$ parton, the W2$\bar{t}$ parton, the b parton or the $\bar{b}$ parton with y being either the b/$\bar{b}$ partons or the W1t/W2t/W1$\bar{t}$/W2$\bar{t}$ partons as indicated in the legend of the plot. In the sketch, the jets are approximated to be cone-shaped. The truth partons are depicted as lines.

It can be found that a high portion of b/$\bar{b}$ truth partons are indeed situated relatively close to the light partons (W1t, W2t, W1$\bar{t}$, W2$\bar{t}$). In this context, 'close' refers to a ΔR distance smaller than 0.8 between the b/$\bar{b}$ and the light truth partons. Taking into account the fact that this is the diameter[3] of the anti-k_t algorithm used to cluster the jets on reco level, ambiguities in reconstruction and matching can arise in these cases. This is depicted in the sketch in figure 6.13b.

---

[3]Actually, the resulting jets of the anit-k_t algorithm are only very roughly cone-shaped. The idea of an exactly cone-shaped jet with a certain diameter is just used as an approximation here.

To obtain further information about the selected jets, their flavors are investigated for the cases in which the reco jet at hand cannot be matched. There are two techniques for deriving partonic flavor labels for reco jets in Monte-Carlo event simulations [105].

One approach is to assign each jet the label of a parton by using a certain value of ΔR(jet, label). This method is used to determine the so-called 'jet-truthflavor' of the reconstructed jets by checking whether, within a radius of $\Delta R < 0.3$, the jet can be associated to a B (value of 5) or a c (value of 4) hadron with $p_T > 5$ GeV. Furthermore, also a check for taus is conducted. The B hadron has then priority over c hadrons and over taus. The jets which could not be assigned to either of these types are labeled with 0. This value is thus a collection of up, down, strange, gluon and pile-up jets. [74] [105]

Figure 6.14 displays the normalized jet-truthflavor distributions for the 6 jets originating from the reconstructed $t\bar{t}$ system for the respective unmatched cases for events of the non-matched category again using the same color-coding as in figure 6.9.
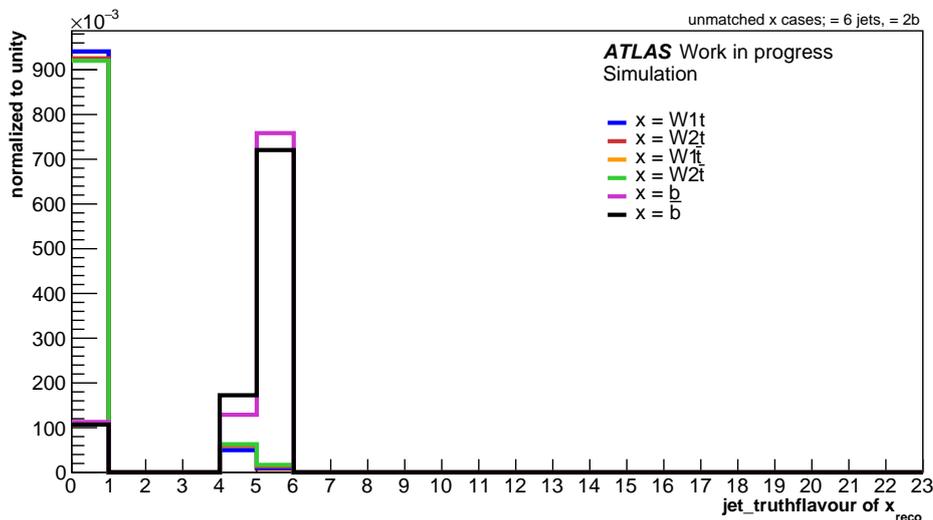
Figure 6.14: Normalized distributions of the jet-truthflavor of a reco-jet x for the respective unmatched cases for events of the non-matched category (section 6.1.3); x refers either to the W1t jet, the W2t jet, the W1$\bar{t}$ jet, the W2$\bar{t}$ jet, the b jet or the $\bar{b}$ jet.

For the b jets, a large share can be assigned to a B hadron. In about 16% of the cases, they are close to c hadrons which corresponds to the misidentification rate of charm jets in b-tagging (section 3.2.5). There is also a share of about 10% of jets which cannot be assigned to a B/c hadron or a tau and thus are light, gluon and pile-up jets. This might be a result of b-light merging.

For the light jets, one can see that a majority of them are not close to B/c hadrons and taus as expected. In about 1% of the cases, they are assigned to B hadrons which is a result of misidentification during b-tagging (subsection 3.2.5).

Another technique to determine the flavor of the reconstructed jet is to use ghost association

(subsection 3.2.4) by clustering the low-p_T partons into jets. The flavor is then assigned in the form of the pdgId of the hardest, ghost-associated parton inside the jet as this parton is assumed to have initiated the jet. This flavor information is contained in the so-called 'jet-truthPartonLabel' [74] [105].

Compared to the jet-truthflavor information, this has the advantage that jets can also be labeled as gluon jets and as all the distinct quark flavors. The bin corresponding to a value of -1 represents the jets from pile up.

Figure 6.15 displays the results of the jet-truthPartonLabel variable for the 6 jets of the reconstructed $t\bar{t}$ system for the respective unmatched cases for the events of the non-matched category in the same color-coding as used in figure 6.9.
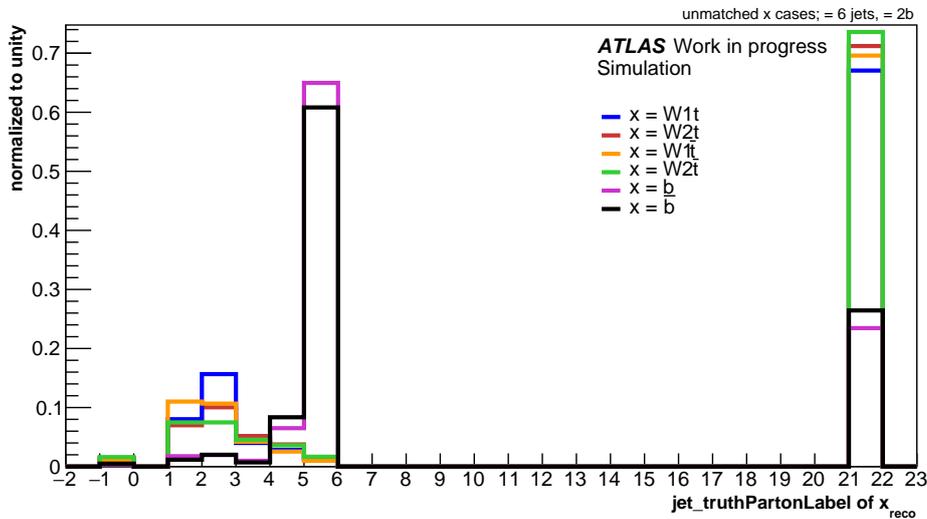
Figure 6.15: Normalized distributions of the jet-truthPartonLabel of a reco-jet x for the respective unmatched cases for events of the non-matched category (subsection 6.1.3) with x being either the W1t jet, the W2t jet, the W1$\bar{t}$ jet, the W2$\bar{t}$ jet, the b jet or the $\bar{b}$ jet.

Most of the b/$\bar{b}$ jets have been initiated by b partons, according to ghost association of the hardest parton. Again, a significant contribution of charm quarks as initiating partons is found due to misidentification. Light partons are only ghost-associated in a small number of cases. The striking observation in this plot is the large contribution of b jets which are ghost-associated to gluons. This might be a result of gluon emissions as visualized in 6.16.
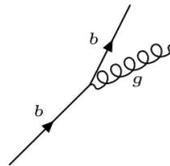
Figure 6.16: Sketch of a gluon emission from a b quark.

For the light jets, this contribution is even higher and dominates the one of light partons ghost-associated to them. In a negligible number of cases the light jets are ghost-associated to b quarks. This can be explained by misidentification.

The contribution from pile-up jets is negligible in all cases.

One reason for the fact that this large gluon contribution cannot be detected explicitly by studying the jet-truthflavor is the missing gluon label. For the jet-truthflavor, there is only one bin for all types except b/c quarks and taus. Another reason is that if the gluons emerge in final-state radiation, the gluon is still close to the b quark and hence should still be close to the B hadron as well. Due to the better 'resolution' of the flavors, the jet-truthPartonLabel can clearly show this.

Concluding, it can be retained that based on the ΔR studies two challenges in reconstruction have been revealed: on the one hand, there is the switching among b and $\bar{\text{b}}$ jets as well as among the light jets and, on the other hand, b-light merging poses a challenge. Both of them can hardly be avoided. In particular for the switching among the jets the masses of the reconstructed $t\bar{t}$ system as well as the reconstructed top-quark/W-boson masses only exhibit minor changes. A possible solution therefore could be to take into account jet charges or conditions on other sensitive observables during reconstruction. In order to develop a possible solution for the b-light merging, more detailed studies are conducted in subsection 6.2.5.
Within the scope of the flavor studies an additional challenge in the form of gluon-initiated jets has been identified. These can only be avoided by developing and applying a quark-gluon tagger. As this is a separate subject not targeted in this thesis, this has not been pursued further.

6.2.3 Validation of the matching condition ΔR(jet, parton) < 0.3

In the context of the ΔR studies in the previous subsection it has been found that a lot of unmatched b/$\bar{\text{b}}$ jets have ΔR values slightly greater than 0.3 to their corresponding truth partons. For that reason, it is checked within this section if the standard matching value of ΔR < 0.3 is a justified choice.
Therefore, the shares of events which can be assigned to the matched, combinatorial and non-matched category for different choices of the ΔR matching value are analyzed. In addition also the shares of matchable events (matched + combinatorial shares) as well as the percentage of matched to matchable events is given.
The results corresponding to values between 0.2 and 0.8 can be found in the table 6.1.

The percentage of events assigned to the matched category increases from 1.1% for a ΔR of 0.2 to a value of 2.0% for a ΔR of 0.5, before it decreases again at higher ΔR values. The share of events in the combinatorial category as well as the matchable share rise steadily while the share of events in the non-matched category and the percentage of matched to

matchable events decreases continuously with increasing ΔR.

This can be explained by envisioning the following situation: the more the matching distance is increased, the more jets can be matched to some parton. As a consequence, the contribution of non-matched events decreases with increasing ΔR while the percentage of matchable events increases. However, only the fact that some jet can be matched to a parton does not imply that the jet and the parton are the corresponding ones. This is visualized by the decreasing percentage of the matched to matchable events as well as the decreasing contribution of matched events for ΔR values above 0.5.

ΔR	matched [%]	combi-natorial [%]	non-matched [%]	matchable [%]	$\frac{\text{matched}}{\text{matchable}}$ [%]
0.2	1.1	2.6	96.2	3.7	29
0.3	1.6	4.3	94.1	5.9	27
0.4	1.9	5.5	92.6	7.4	26
0.5	2.0	6.8	91.2	8.8	23
0.6	1.9	8.2	89.9	10.1	19
0.7	1.8	10.0	88.3	11.8	15
0.8	1.5	12.3	86.2	13.8	11

Table 6.1: Table yielding the shares of events which can be assigned to the matched, combinatorial and non-matched category (subsection 6.1.3) for different choices of the ΔR matching value (0.2 to 0.8). In addition also the shares of matchable events (matched + combinatorial shares) as well as the percentage of matched to matchable events is given.

Subsequently, ΔR values above 0.5 can be excluded as one does not want the matched or rather the matched to matchable contribution to be unnecessarily low. Also, the value of 0.5 itself should not be chosen due to the high combinatorial share resulting in a small percentage of matched to matchable events. Therefore, a trade-off between a higher percentage of matched or rather matched to matchable events and a higher combinatorial share has to be conducted for the remaining ΔR values of 0.2, 0.3 and 0.4.

To determine the optimal matching value out of the aforementioned values, further investigation studies of the physical properties are conducted. For this reason, the ΔR and flavor studies are repeated and in order to cover both jet-types exemplified for the W1t and the b reco jets.

First, the minimum ΔR between the W1t/b reco jet and all truth partons originating from the $t\bar{t}$ system as well as the normalized distribution of the pdgId of the truth quark resulting in this minimal value are considered. This is shown in figure 6.17 for events of

the non-matched category in which the W1t/b reco jet could not be matched to a truth parton within the chosen ΔR radius of 0.2, 0.3 and 0.4.

The normalized distributions of the W1t reco jet in figure 6.17a and the b reco jet in figure 6.17c indicate that the shapes for the different ΔR values stay the same especially towards higher values. The only observable differences arise due to the moving matching threshold at lower values. In the case of the unmatched b jets, also the tail gets a bit broader with increasing ΔR values as in this cases also more jets are accepted which do not originate from the $t\bar{t}$ system. While in figure 6.17a the corresponding bins situated just above the particular threshold have identical heights, a decreasing height can be found for the distributions in figure 6.17c. In the aforementioned figure, the difference between the shapes corresponding to the 0.3 and the 0.2 values is a lot larger than the one between those corresponding to the 0.3 and 0.4 values.

The pdgId of the truth partons resulting in these minimal values is shown for the unmatched W1t cases in figure 6.17b and for the unmatched b cases in figure 6.17d.
In general, the underlying shapes stay the same. While in figure 6.17b there are only slight differences for the different ΔR values, in figure 6.17d significant differences especially in the b-quark contribution (pdgId of 5) can be observed. For a ΔR value of 0.2 there are clearly more b quarks leading to a minimal ΔR value. For the other two values, this share is reduced and instead the light contributions are larger. This is expected as a reduced ΔR value implies a stricter selection.
It is striking that the differences for the unmatched b-jet cases are larger than for the unmatched W1t-jet cases. This can be explained by considering the larger number of light jets compared to the number of b jets. Taking into account the phenomenon of switching among the light jets, there are subsequently more possible combinations which result in canceling out the differences for the light jets.

All in all, from the aspects covered in figure 6.17, it can be deduced that the choice of the ΔR value to be 0.3 represents a compromise between an overall higher matching percentage and a lower b-quark contribution in case of unmatched b reco jets.
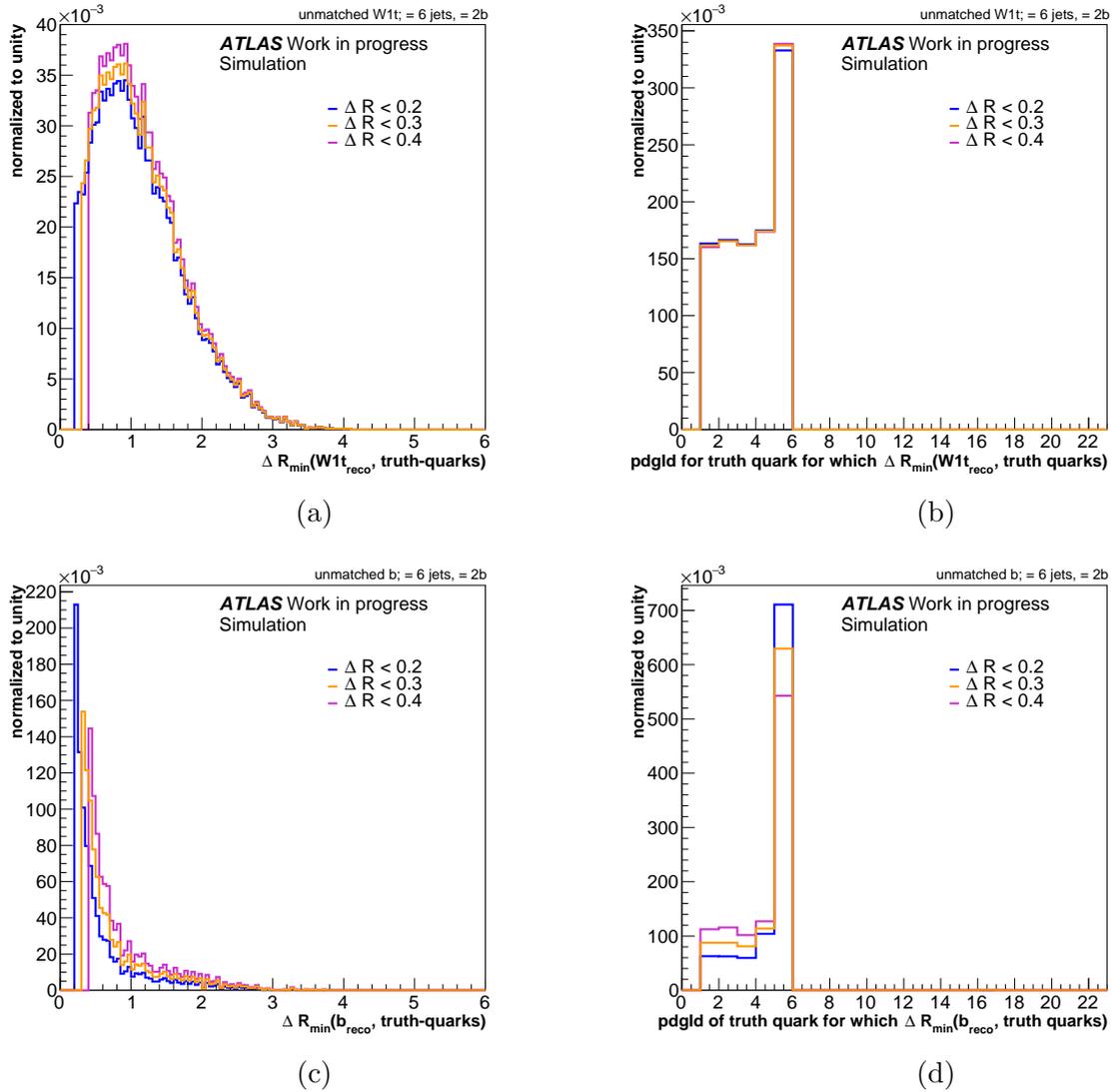
Figure 6.17: Normalized distributions of the minimum ΔR between the W1t reco jet (b reco jet) and all the truth quarks in a)/c) and the pdgId of the corresponding truth quark resulting in this minimum value in subfigure b)/d) respectively. The events considered are the unmatched W1t reco jet cases (b reco cases) for events of the non-matched category (subsection 6.1.3). The different colors indicated in the legend refer to different matching ΔR values.

In a next step, the normalized minimum ΔR distribution between, on the one hand, the W1t reco jet and the light truth partons as well as, on the other hand, the minimum ΔR distribution between the b reco jet and the $b/\bar{b}$ truth partons are analyzed. The results can be found in figure 6.18 in the same color-coding as in figure 6.17. This representation has been chosen in order to take into account the switching between the light jets which can hardly be avoided.
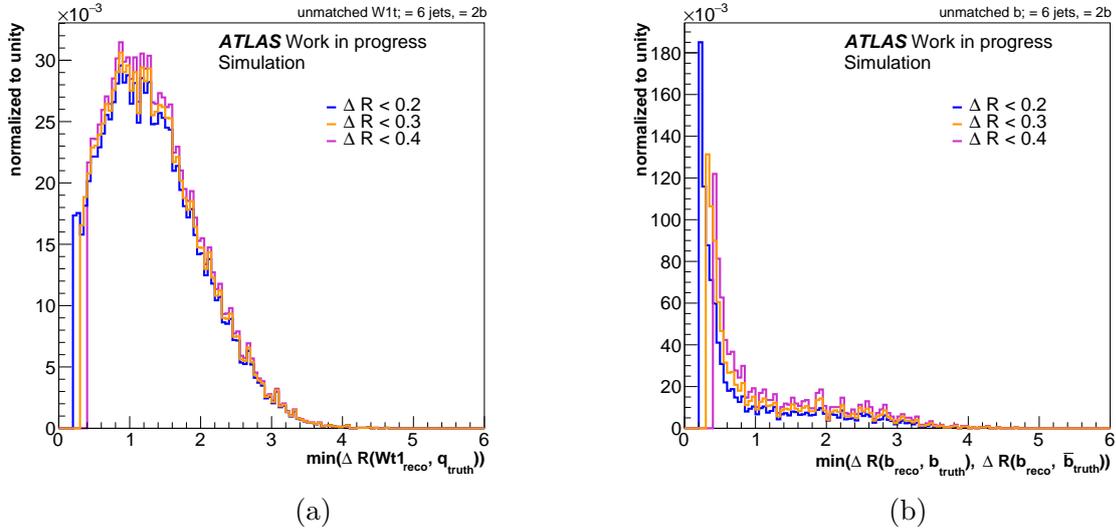
Figure 6.18: Normalized distribution of the minimum ΔR between the W1t reco jet and the light truth partons (b reco-jet and b/$\bar{\text{b}}$ truth partons) originating from the $t\bar{t}$ system for the unmatched W1t cases (unmatched b cases) in a)/b) for events of the non-matched category (subsection 6.1.3) for different values of the matching ΔR as indicated in the legend.

In both cases, changing the matching ΔR value only slightly changes the shapes of the distributions, but rather results in a moving threshold depending on the chosen value.

In figure 6.18a, a comparison of the threshold-bin heights manifests a dip. The minimum is traversed by the distribution corresponding to the ΔR value of 0.3. This means that not too many jets are accepted as with a 0.4 matching value while at the same time not too many of them are rejected as for a choice of a 0.2 matching value.

In figure 6.18b, decreasing heights with larger differences between the values of 0.2 and 0.3 than between 0.3 and 0.4 are found instead of a dip. Thus choosing a matching of 0.2 instead of 0.3 would exclude a lot more non-appropriate jets.

Concluding, figure 6.18a proves that the standard matching choice of $\Delta R < 0.3$ is a justified choice and figure 6.18b supports this too by showing that a choice of 0.3 as matching value yields a compromise between a higher matched percentage and a stricter jet selection.

Last but not least, also the flavor of the selected W1t/b reco jet is investigated for the different ΔR matching values. The normalized distributions of the jet-truthflavor and the jet-truthPartonLabel observable are shown in the same color-coding in figure 6.19 for the respective unmatched cases of events of the non-matched category.

Figures 6.19a and 6.19c illustrate that the different jet-truthflavor contributions of the selected W1t/b reco jets stay basically the same. Small differences can only be observed in

figure 6.19c for the bin related to the jets which can be assigned to a B hadron (value of 5). Here, the larger the ΔR matching value, the smaller the contribution of assigned b jets.
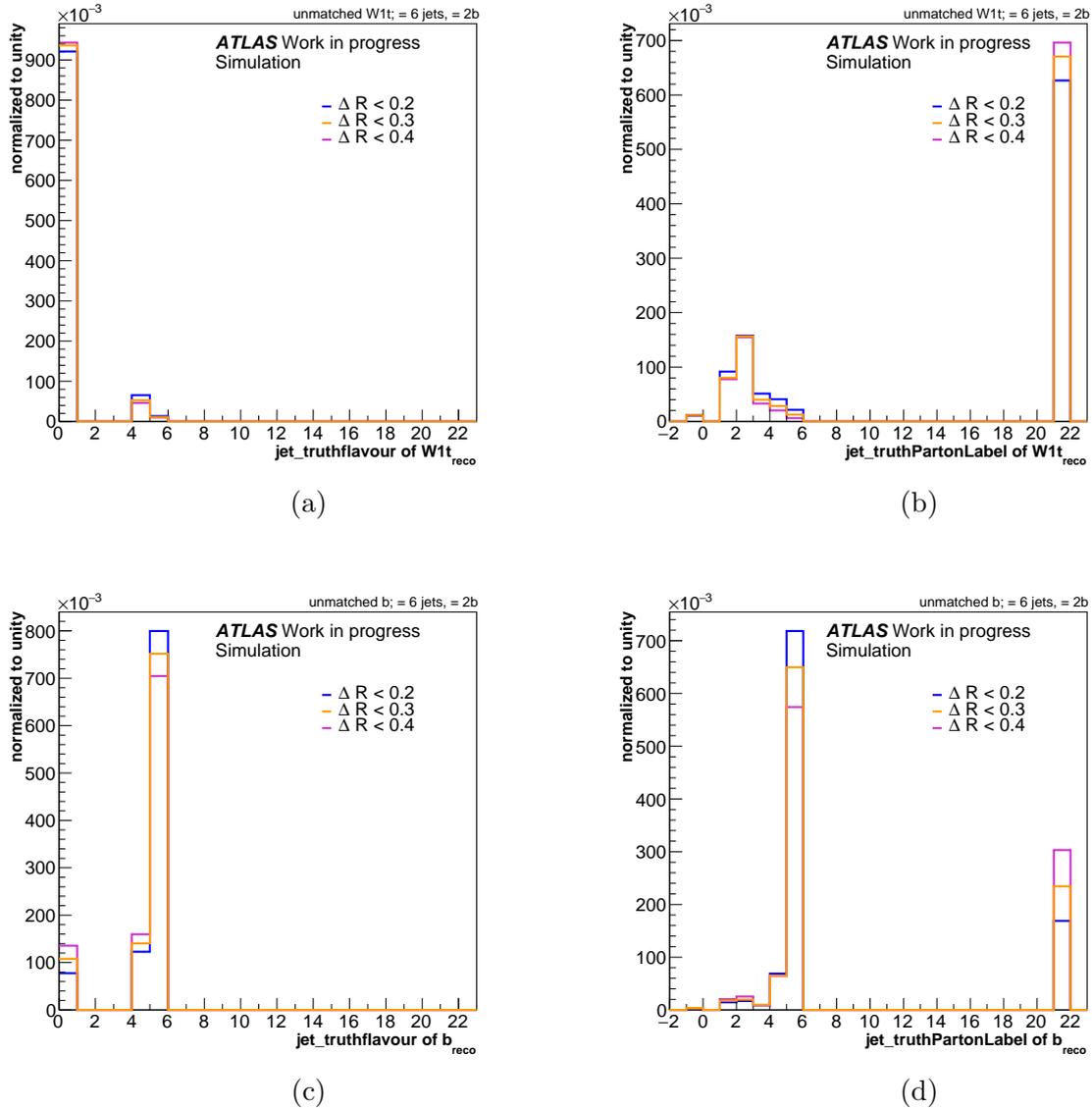
Figure 6.19: Normalized distributions of the jet-truthPartonLabel and jet-truthflavor of the W1t jet (a and b) and the b jet (c and d) for the respective unmatched cases for events of the non-matched category (subsection 6.1.3) for different values of the matching ΔR as indicated in the legend of the plots.

Regarding the jet-truthPartonLabel in figure 6.19b and 6.19d of the selected W1t/b reco jets, the basic trends stay also the same. In the case of the unmatched W1t reco-jet small differences can be found especially for the gluon contribution (value of 21). Here, the larger

the ΔR matching value, the more gluons contribute. For the unmatched b jets, these differences are even more significant. Here, not only the gluon, but also the b contributions vary significantly with larger ΔR matching values. The larger the ΔR matching value the smaller the b and the higher the gluon contribution.

Thus, as the gluon contribution should be reduced as much as possible and also a high b contribution should be sustained, choosing the standard value of the matching condition of 0.3 also represents a compromise regarding the reco-jet flavors.

Considering all the analyses above, it can be noted that the standard ΔR matching value of 0.3 represents in general a compromise between a higher percentage of matched events and a lower (larger) gluon (b) contribution. In figure 6.18a, the standard choice seems justified as the corresponding distribution traverses the minimum and thus assures that the right amount of 'good' jets are taken into account.

6.2.4   Phase-space adaption between reco and truth level: η and p_T studies

To increase the probability of successful matching, adapted phase-spaces between reco and truth level are essential: Truth partons of the $t\bar{t}$ system lying outside the allowed phase-space on reco level inevitably lead to a non-matched scenario.

For that purpose, different cut values are studied which should be understood as exploratory choices determining the direction for further studies rather than final, optimized choices. Table 6.2 displays the resulting shares of events assigned to the matched, combinatorial and non-matched category as well as the remaining number of events after the application of the respective cuts. In addition also the matched to matchable ratio is given. The matchable share represents in this context the sum of the matched and combinatorial shares. To ease comparison, the result of the initial situation, which considers exactly 6 jets as well as exactly 2 b jets and applies the signal-region and analysis cuts defined in chapter 5, is included in the table.

First, the adaption of the truth-level p_T-phase-space to the one on reco level is targeted. In this context, different p_T cuts are applied to all truth partons originating from the $t\bar{t}$ system. These are given in the form x/y with x corresponding to the cut value for the b and the $\bar{b}$ parton while y corresponds to the cut value for the light partons.

Imitating the p_T cut of 25 GeV on reco level, one cut-combination applied on truth level is 25/25 GeV. Taking into account hadronization and detector effects, a 30/30 GeV cut is applied as well. Due to the fact that b quarks originating from a $t\bar{t}$ system have in general a higher p_T than the light quarks, a 30/25 GeV cut is used in addition.

The cut-value pair leading to the highest matched percentage of 4.3% is the 30/30 GeV cut. It more than doubles the initial matching percentage of 1.6%. Unfortunately, this

combination also leads to the highest combinatorial contribution. In consequence, also the matched to matchable ratio is only slightly improved compared to the initial situation. In addition also the number of events is reduced by additional 6% compared to the 25/25 GeV cut. However, the increase in the percentage of matched events of the 30/30 GeV cut is with about 1% compared to the other two cuts quite significant and thus the 30/30 GeV cut pair is chosen despite the downsides.

cuts	number of events	matched	combi-natorial	non-matched	$\frac{\text{matched}}{\text{matchable}}$
initial	34083	1.6 %	4.3 %	94.1 %	27.1 %
$\mathbf{p_T}$ **cuts**					
30/30 GeV	8132	4.3 %	11.4 %	84.4 %	27.4 %
25/25 GeV	13100	3.5 %	8.9 %	87.6 %	28.2%
30/25 GeV	12538	3.5 %	9.0 %	87.5 %	28.0%
η **cuts**					
1.4	6827	2.4 %	7.7 %	89.9 %	23.7 %
1.7	12229	2.4 %	6.8 %	90.7 %	26.0 %
2.5	26667	2.0 %	5.5 %	92.5 %	26.7 %
combined					
p_t & $|\eta|$	7134	4.8 %	12.9 %	82.3 %	27.1%
p_t & $|\eta|$ & non-merged	2104	5.6 %	10.5 %	84.0 %	34.8 %

Table 6.2: Table yielding the results obtained by applying different p_T/η cuts to the truth partons originating from the $t\bar{t}$ system in order to adapt the p_T and η phase spaces on reco and truth level. The p_T cuts are given in the form x/y with x referring to the $p_T > x$ cut for the $b/\bar{b}$ truth partons and y referring to the $p_T > y$ cut applied to the light truth partons. The η cut values are given in the form x referring to $|\eta| < x$. Furthermore, the results obtained by combining the selected cuts on $p_T > 30$ GeV and $|\eta| < 2.5$ are included. In addition also another cut excluding merged partons (i.e. partons originating from the $t\bar{t}$ system having a distance of $\Delta R \leq 0.8$ between each other) has been added. Due to rounding, the numbers in the table do not always add up to 100%.

In a next step, different $|\eta|$ cuts on truth level are applied to all truth partons of the $t\bar{t}$ system in order to further adapt the phase space. The cut on $|\eta| < 2.5$ is motivated by the reco-level cut value. The $|\eta|$ cut on 1.7 is chosen as this corresponds to the acceptance of

the hadronic barrel calorimeter and the Pixel detector used for b-tagging. As the acceptance of the SCT barrel part also used for b-tagging is limited to $|\eta| < 1.4$, this cut value is used as well.

The cut values on $|\eta| < 1.7$ and $|\eta| < 1.4$ yield the highest matching percentages. At the same time also the number of events is reduced by more than half compared to the initial situation. Furthermore, they also yield the largest combinatorial shares which leads to a lower matched to matchable ratio compared to a cut on $|\eta| < 2.5$.
In order to keep as many events as possible and by taking into account that there is only a minor difference in the matched shares, the cut value of 2.5 is the preferred option.

Finally, the two selected cuts are combined. Furthermore an additional cut is applied which excludes events with merged partons, i.e. partons originating of the $t\bar{t}$ system with a distance $\Delta R \leq 0.8$ to each other, have been added. This addresses the challenge found in subsection 6.2.2.

From the results, the large impact of applying the chosen cuts is clearly visible: Adapting the phase space on reco and truth level by setting a 30/30 GeV p_T cut and an $|\eta| < 2.5$ cut improves the percentage of matched events considerably from 1.6% to 4.8%. At the same time the combinatorial contribution is increased while the non-matched contribution decreases. This indicates that there are indeed events for which the truth partons are situated outside of the allowed reco phase space and thus cannot be matched. The percentage of the matched events with respect to the matchable events is in both cases 27%, thus the assignment performance of the χ^2 reconstruction algorithm stays the same.
Taking into account the challenge of merged partons identified in subsection 6.2.2 and excluding these using an additional cut improves the percentage of matched events further to 5.6%. Compared to the situation above, the combinatorial contribution decreases while the unmatched contribution increases and thus also the share of matched to matchable events increases to 34 %. This means that by excluding one of the identified challenges in reconstruction, ambiguities in the assignment of the reco jets to the different decay-product positions are removed and thus the performance of the reconstruction is improved.
The chosen p_T and η cut as well as the cut excluding merged partons will be referred to as 'improvement cuts' in the rest of this thesis.

6.2.5   Looking closer: merged truth-parton studies

In subsection 6.2.2, merged partons have been identified as a challenge in the reconstruction of hadronically decaying $t\bar{t}$ systems in $t\bar{t}Z$ events. To develop a more detailed understanding of these cases, further investigation studies are conducted on truth level with applied improvement cuts on top of the initial event selection. In this context, the number of overall merged partons in an event as well as the number of merged cases to a particular parton are studied. Apart from that, it is also investigated which partons merge. In addition, the topology depending on the transverse momentum of the top quark $p_T(t)$/anti-top quark

$p_T(\bar{t})$ is studied. These studies are conducted for events of the non-matched category for which at least one ΔR distance between the partons of the $t\bar{t}$ system is ≤ 0.8 or in other words for the merged events of the non-matched category.

Number of partons with a distance of $\Delta R \leq 0.8$ between each other

As a starting point, the number of 'merged' partons in one event is plotted. In this context, 'merged partons' refer to partons originating from the $t\bar{t}$ system with a distance $\Delta R \leq 0.8$ to each other. The resulting distribution for the merged events of the non-matched category is depicted in figure 6.20.
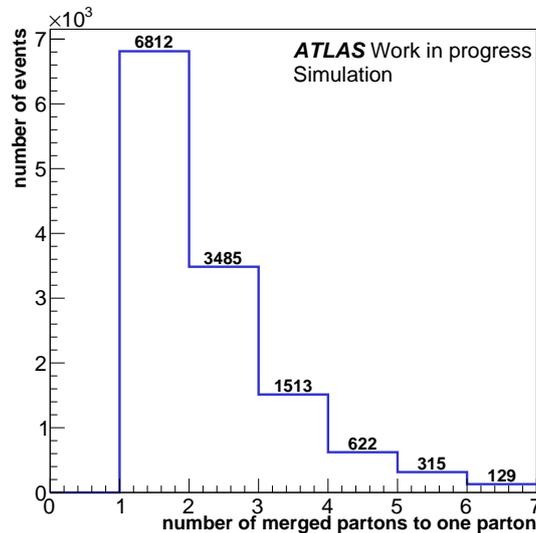
Figure 6.20: Number of merged partons, i.e. partons with a ΔR distance of ≤ 0.8 between each other, in an event considering merged events of the non-matched category (subsection 6.1.3).

In general, up to 6 merged parton-pairs can be found in an event. The largest share is represented by the 1-merged-parton-case, i.e. events in which exactly one merged parton-pair can be found. Looking at cases involving more merged partons in an event, a steadily decreasing number of events can be found.

Based on these observations, three categories are defined containing events with exactly 1-, 2- and 3-merged parton-pairs. These are used later on to enable more detailed investigation. The 4-, 5- and 6-merged-cases are neglected in this context due to a low number of events which leads to a large statistical uncertainty.

Number of partons merged to a particular parton originating from the $t\bar{t}$ system

In a next study, the focus is shifted from the whole event to a particular parton originating from the $t\bar{t}$ system. In this context, the number of partons merging to a particular parton is analyzed. This refers to the number of ΔR distances between the analyzed parton and all the other partons originating from the $t\bar{t}$ system which are ≤ 0.8. In figure 6.21, the resulting normalized distribution is shown examplarily for the b parton (blue)/the W1t parton (red)/the W2t parton (yellow) as the particular parton using the merged events of the non-matched category.
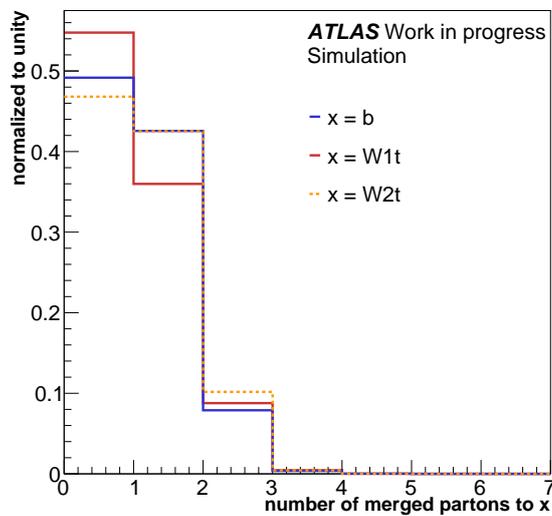
Figure 6.21: Normalized number of partons merged to a particular truth-parton x considering merged events of the non-matched category (subsection 6.1.3); x refers to either the b truth parton, W1t truth parton or the W2t truth parton.

For each of the b/W1t/W2t truth partons, there are up to 4 merged partons. The cases in which no partons merge to the analyzed truth partons represent the largest share. This case is not excluded in this study as the cut on merged partons requires only at least one merged parton-pair in the whole event. This pair does not necessarily involve the analyzed parton. For the cases with 2-, 3- and 4-merged-partons to the analyzed truth partons, the number of events decreases steadily.

The striking observation in this plot is the significant difference in the number of 1-merged-parton-cases for the particular partons W1t and W2t. This can be explained by considering the V-A structure of the electroweak interaction which can be visualized using the cosine of the helicity angle $\cos(\Theta^*)$ (subsection 2.1.2).

In subsection 2.1.2, it has been found that in top-quark decays within the SM only left-handed and longitudinally polarized W bosons are involved in the SM.

(a)

(b)
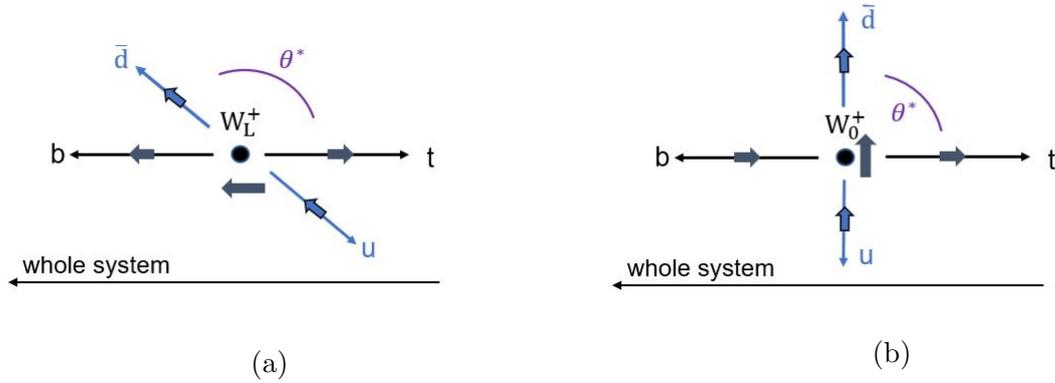
Figure 6.22: Visualization of the definition of the helicity angle Θ^* for left-handed (a) and longitudinally (b) polarized W bosons (subsection 2.1.2).

The situation for left-handed W bosons is depicted in figure 6.22a. Here, it can be observed that the quark situated closest to the b quark is the down-type quark (W2t). As the up-type quark (W1t) is emitted into the opposite direction than the down-type quark (W2t) in the W-boson rest frame considered in the sketch, the b and the up-type quark are situated farer apart. Consequently, the b quark merges more frequently with the down-type quark (W2t) than with the up-type quark (W1t). This would explain the mentioned difference in the 1-merged-parton shares found in figure 6.21.

The situation for a longitudinal W boson is depicted in figure 6.22b. It can be deduced that the up-/down-type quarks are emitted at helicity angles of about $90°$, i.e. perpendicular to the b-quark direction. Hence, this cannot explain the observation in figure 6.21.

For this assumption to be true, larger angles Θ^* and consequently more entries for the cosine of the helicity angle $\cos(\Theta^*)$ at a value of -1 are expected for the merged cases compared to non-merged cases.

To confirm this, figure 6.23 depicts the normalized distributions of the cosine of the helicity angle $\cos(\Theta^*)$ for merged (red) and non-merged (blue) events of the non-matched category, separately for the top quark (figure 6.23a) and the anti-top quark (figure 6.23b) on truth level. In these plots the outlined behavior is indeed observed.

Merged-parton configurations

Having analyzed the *number* of partons merged to a particular parton, it is now investigated *which* partons merge to the particular parton. For visualization, a matrix is used showing on the x-axis the particular parton of the $t\bar{t}$ system which is analyzed. On the y-axis, the parton of the $t\bar{t}$ system is shown which merges to the analyzed parton. The resulting shares in % are given in figure 6.24 together with the statistical uncertainty which is calculated as the square root of the number of entries.

(a)   (b)
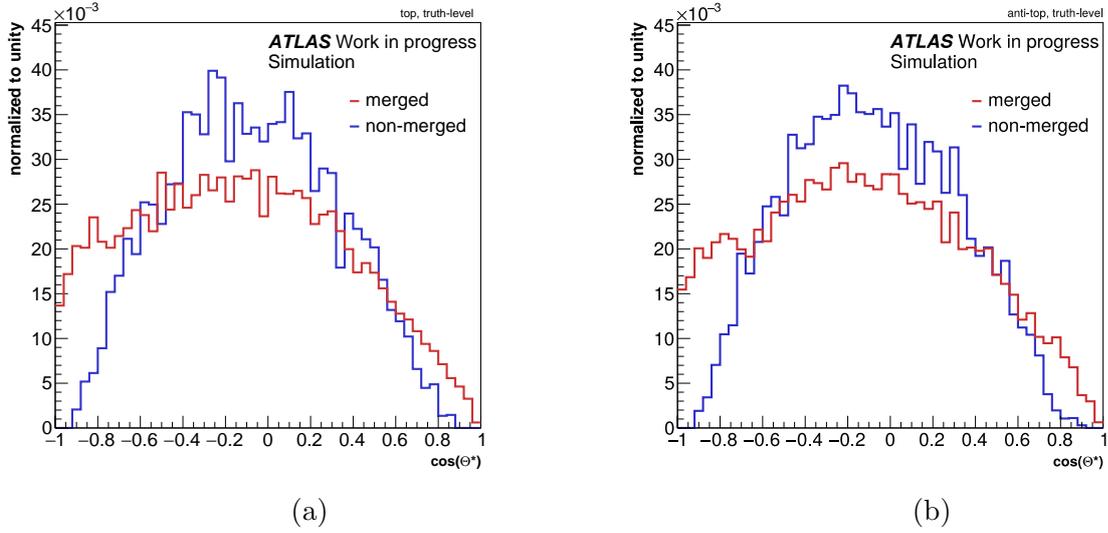
Figure 6.23: Normalized distributions of the cosine of the helicity angle $\cos(\Theta^*)$ (subsection 2.1.2) in the case of merged events and non-merged events for top (a) and anti-top (b) on truth level using events of the non-matched category (subsection 6.1.3).
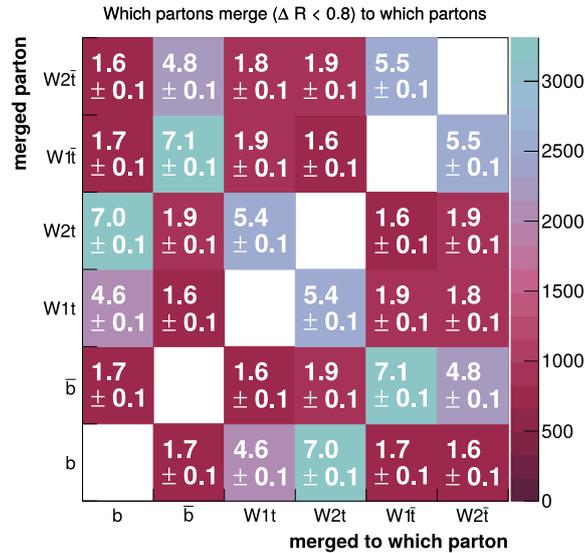
Figure 6.24: Visualization of the merged-parton configurations (in %) in matrix form for all merged events using the merged events of the non-matched category (subsection 6.1.3).

The highest shares are found for the merging between the b/$\bar{\text{b}}$ truth partons and the light quarks of the same top quark, on the one hand, and for the merging among the light quarks of the same W boson, on the other hand. Furthermore, it is interesting to see that there is also a small share of b-$\bar{\text{b}}$ merging.

Attention should be drawn to the difference in the merging shares of W1t-b (W2t̄-b̄) and W2t-b (W1t̄-b̄). This is again a consequence of the V-A structure of the electroweak interaction and can be explained in the same way as in figure 6.21 considering that W1t/W2t̄ is the up-type and W2t/W1t̄ is the down-type quark.

Comparing all shares to their respective statistical errors, it can be found that all percentages are significant.

To gain further insights, this matrix is redrawn for the cases with exactly 1-, 2- and 3-merged-partons in the event. The resulting shares in % are shown in figure 6.25 for the merged events of the non-matched category. Figures 6.25a, 6.25b and 6.25c represent the results of the 1-, 2- and 3-merged-parton-cases respectively.

(a) 1-merged parton case   (b) 2-merged parton case   (c) 3-merged parton case
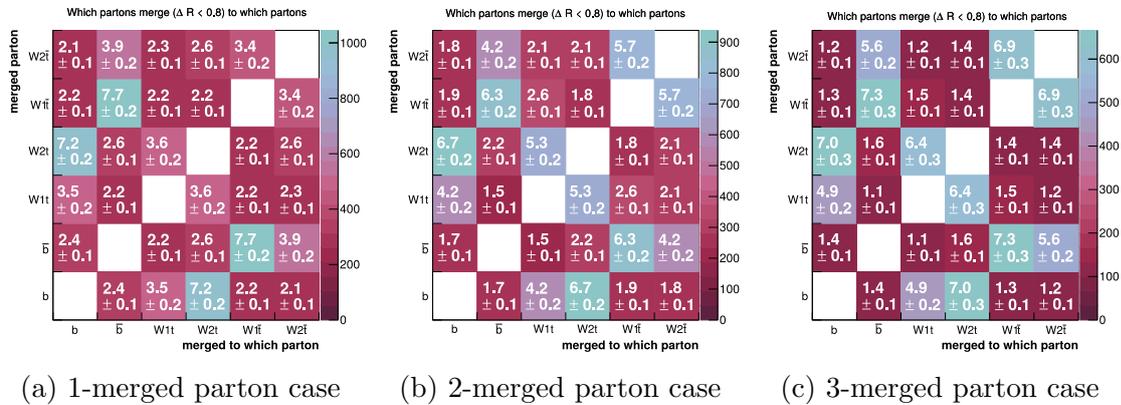
Figure 6.25: Visualization of the merged-parton configurations (in %) in matrix form for events with exactly 1- (a), 2- (b) and 3- (c) merged-partons in an event using merged events of the non-matched category (subsection 6.1.3).

At first sight, one notices that the colors change from the 1-merged-parton-case (figure 6.25a) to the 3-merged-parton-case (figure 6.25c). This is due to the increase of the merging between the decay products of the same top quark. At the same time the b-b̄-merging share as well as the merging between the decay products originating from different top quarks decrease. Also the difference between the b-W1t (b̄-W2t̄) and the b-W2t (b̄-W1t̄) decreases.

Based on these observations, it seems likely that by going from the 1-merged to the 3-merged-parton-case a transition from a non-boosted to a boosted topology is conducted. In a boosted topology, the top quark has a high transverse momentum which, according to momentum conservation, leads to a high p_T of the decay products. In consequence, these are emitted in smaller angles around the top/anti-top quark and thus are situated closer together. Hence, the merging within the decay products of the top quark is preferred and also the b-b̄ merging share decreases.

Looking at the denoted error, it can be deduced that all given share values are significant.
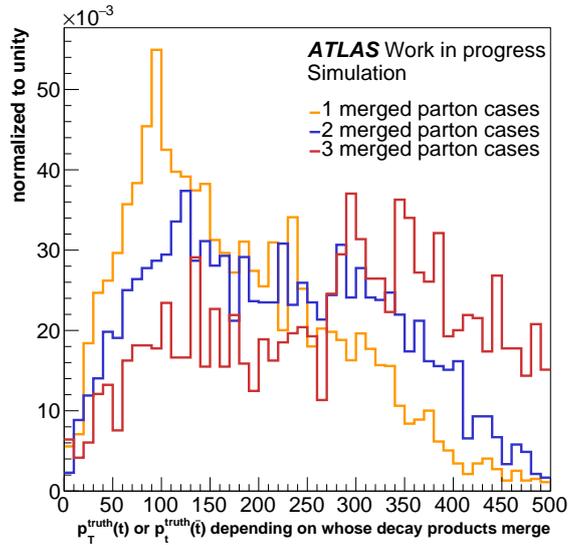
Figure 6.26: Normalized distributions of the p_T of top-/anti-top quark if their decay products merge for the 1-, 2- and 3-merged-parton-cases for the merged events of the non-matched category (subsection 6.1.3).

To verify the assumption of a boosted topology, the p_T of the top/anti-top quark is plotted if their decay products merge. This is conducted for the three distinct cases of 1- (yellow), 2- (blue) and 3-merged (red) -partons in the event for merged events of the non-matched category. The resulting normalized distributions are shown in figure 6.26.

For the 1-merged-parton-cases, a peak at about a transverse momentum of 100 GeV with a broad decreasing shoulder towards higher p_T values can be found. In contrast, the distributions representing the 2- and 3-merged-parton-cases are a lot broader. This leads to a mean p_T of the top/anti-top quark of 225 GeV for the 2-merged-parton-case. For the 3-merged-parton-case even a peak in the range between 300 and 350 GeV and a broad shoulder towards lower values can be identified.

Concluding, it can therefore be retained that the mean transverse momentum of the top quark increases indeed with a larger number of merged partons in the event. Hence, the assumption of a transition to a boosted topology is supported.

$p_T(t)/p_T(\bar{t})$-dependent topology studies

Based on the previous observation of an increased mean of the top/anti-top quark p_T for a larger number of merged partons in the event, the topology is now further studied for 4 different $p_T(t)/p_T(\bar{t})$ bins: the first bin includes events in which the p_T values of the top/anti-top quark lie between 0 GeV and 100 GeV (blue). The second one includes $p_T(t)/p_T(\bar{t})$ values between 100 GeV and 200 GeV (red). In the third bin (orange), events

with $p_T(t)/p_T(\bar{t})$ values between 200 GeV and 300 GeV can be found while the events with $p_T(t)/p_T(\bar{t})$ values greater than 300 GeV are assigned to the fourth bin (pink).

For each of the 4 p_T bins, the angular difference $\Delta\phi$ between the top quark and the b quark is plotted in a first step. The resulting normalized distributions using the merged events of the non-matched category can be found in figure 6.27.
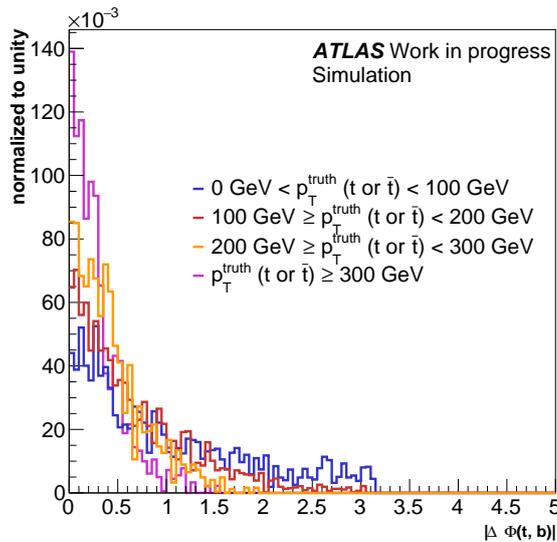
Figure 6.27: Normalized distribution of the absolute value of $\Delta\phi$ between the top quark and the b quark on truth level for 4 different ranges of the p_T of the top/anti-top quark as defined in the text using the merged events of the non-matched category (subsection 6.1.3).

For a higher mean p_T of the top/anti-top quark, corresponding to higher merged-parton cases (subsection 6.2.5), the peak towards lower $\Delta\phi$-values becomes sharper. This means that the higher the transverse momentum of the top quark, the smaller the angle in which its decay products, in this case the b quark, are emitted. In consequence, the decay products are situated closer together and the merging between the decay products of the same top quark is preferred. This supports the result found in the context of the merged-parton configuration studies above.

Finally, the topology of the $t\bar{t}Z$ system is investigated. Therefore, the minimum of the absolute $\Delta\phi$ value between the top/anti-top quark and the Z boson is plotted against the absolute value of the $\Delta\phi$ between the top and anti-top quark. Figure 6.28 displays the result for the merged events of the non-matched category.
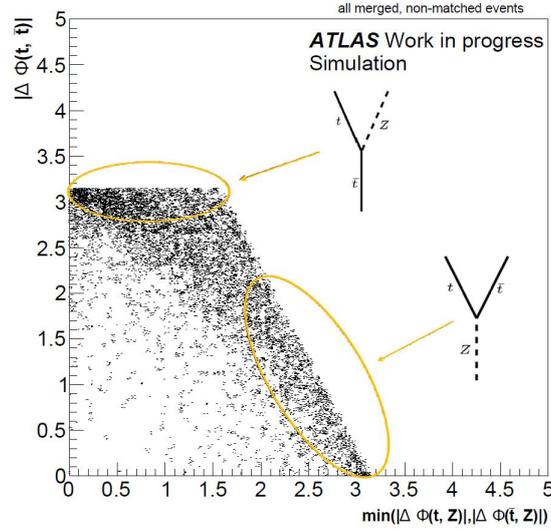
Figure 6.28: Scatter plot showing the minimum of the absolute values of $\Delta\phi$ between the top/anti-top quark and the Z boson against the absolute value of $\Delta\phi$ between the top and the anti-top quark using the merged events of the non-matched category (subsection 6.1.3). In addition, sketches of the two preferred topologies have been added.

Most of the events gather at two positions with opposed characteristics: one preference is the topology having a large absolute value of $\Delta\phi$ between the top/anti-top quark and the Z boson in combination with a small $\Delta\phi$ value between the top and the anti-top quark. This corresponds to the case in which the Z boson recoils against the t$\bar{\text{t}}$ system. The other preferred topology is the one having a small absolute value of $\min(|\Delta\phi(\text{t},\text{Z})|,|\Delta\phi(\bar{\text{t}},\text{Z})|)$ combined with a large $|\Delta\phi(\text{t},\bar{\text{t}})|$ value. This represents a topology in which the top quark and the anti-top quark are emitted almost back-to-back.

Redrawing this plot for the first 3 different p_T bins[4] yields the results shown in figure 6.29.

With a larger mean p_T value of the top/anti-top quark, more and more points in figure 6.29c gather at relatively small $\min(|\Delta\phi(\text{t},\text{Z})|,|\Delta\phi(\bar{\text{t}},\text{Z})|)$ values and large $|\Delta\phi(\text{t},\bar{\text{t}})|$ values. Thus, it can be deduced that the topology in which the top and the anti-top quark are emitted almost back-to-back is preferred. This can also be quantified by looking at the mean values and standard deviations in x- and y-direction of figure 6.29:

This is also in line with the observation of a decreasing b-$\bar{\text{b}}$-merging share for events with a higher number of merged partons per event as the b and the $\bar{\text{b}}$ quark are emitted in opposite directions in the corresponding topology.

---

[4]There is no plot for the fourth p_T bin due to a very low number of events which would lead to large statistical uncertainties.
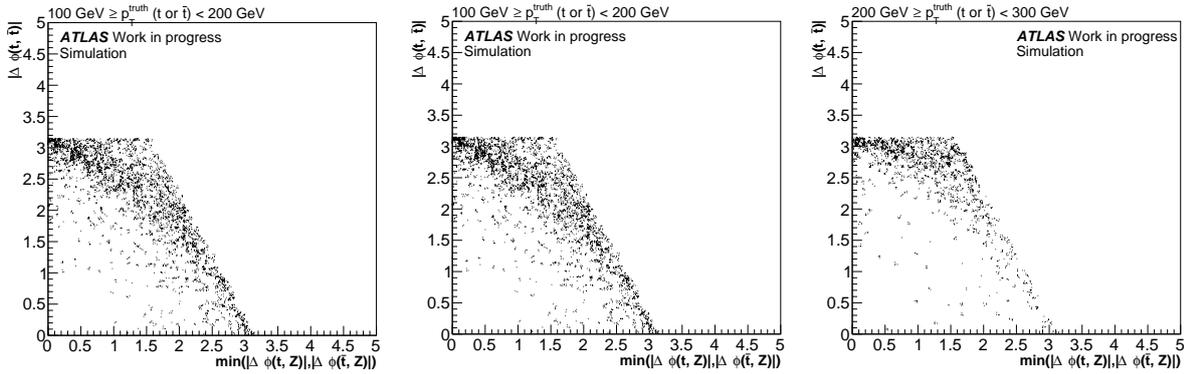
Figure 6.29: Scatter plot showing the minimum of the absolute values of $\Delta\phi$ between the top/anti-top quark and the Z boson against the absolute value of $\Delta\phi$ between the top and the anti-top quark for 3 different $p_T(t)$ bins defined in the text using the merged events of the non-matched category (subsection 6.1.3).

p_T bin	mean x	mean y	std. dev. x	std. dev. y
0 GeV $<$ p_T $<$ 100 GeV	1.514	1.762	0.8322	0.9196
100 GeV $<$ p_T $<$ 200 GeV	1.365	2.108	0.7878	0.833
200 GeV $<$ p_T $<$ 300 GeV	1.158	2.517	0.6861	0.6869

Table 6.3: Table yielding the mean and standard-deviation values in x- and y-direction of figure 6.29.

6.3 Conclusion and cut-flow

In this chapter, the χ^2 reconstruction as well as a procedure called truth matching were outlined and applied. It has been shown that the χ^2 value calculated in the reconstruction process can be used to separate the signal $t\bar{t}Z$ process from background processes like the Z+jets and the dileptonic $t\bar{t}$ decay. The performance of the minimum-χ^2 algorithm has been analyzed by comparing the selected best χ^2 hypotheses to all hypotheses considered during reconstruction. This has shown a better performance for the selected hypotheses. However, the resulting share of matched events in the context of truth matching was found to be with 1.1%[5] extremely low which indicates that the reconstruction of hadronic $t\bar{t}$ systems is not a trivial task. In addition, also the matchable share, i.e. $\frac{\text{matched}}{\text{matched}+\text{combinatorial}}$ was found to be with about 27% quite low.

Based on these findings, detailed investigation studies were conducted in order to identify and understand the challenges faced. In this context, some challenges in the form of gluon jets, switched jets, merging of jets as well as a mix of boosted and non-boosted topologies have been identified. Furthermore, studies have been conducted to successfully validate

---

[5]for an event selection with $\geq$ 6 jets and $\geq$ 2 b jets

the standard choice of the matching condition $\Delta R < 0.3$.

On top of that, a phase-space adaption between reco and truth level has been performed in order to improve the attainable matching performance. Therefore, a p_T cut as well as an η cut on all truth partons originating from the $t\bar{t}$ system were selected. In addition, the challenge of merged partons was addressed with an additional cut to exclude merged partons on truth level. All these cuts were grouped under the name 'improvement cuts' as their application resulted in an increase of the matched share from 1.6% to 5.3%[6] as well as the matchable share ($\frac{\text{matched}}{\text{matched+combinatorial}}$) of 27.1 % to 34.8%.

The overall reduction of events by the applied cuts is visualized in the cut-flow shown in figure 6.30.
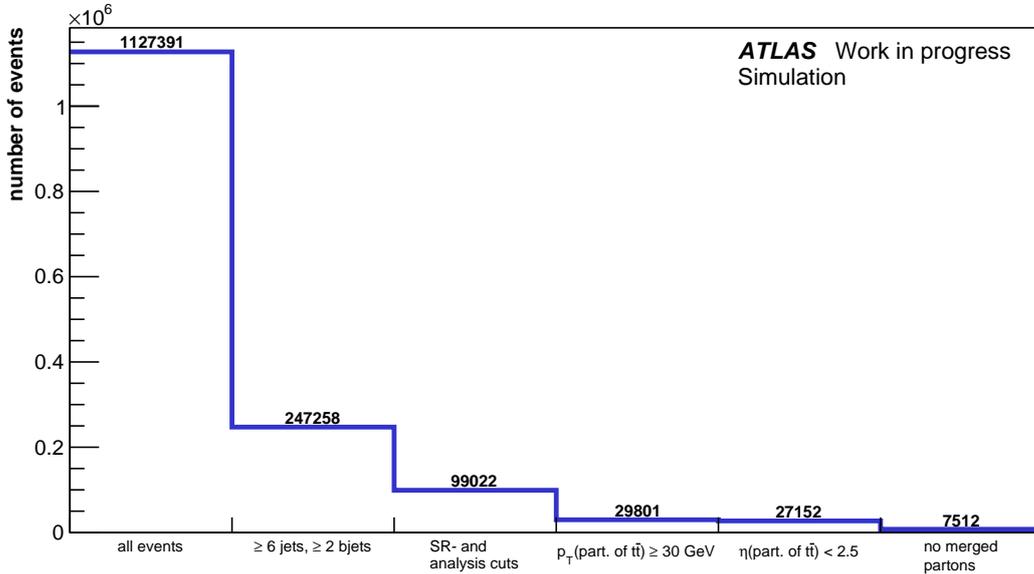
Figure 6.30: Visualization of the reduction of events by applying the event selection presented in chapter 5 and the improvement cuts found in subsection 6.2.4.

After applying the analysis cuts and the cuts focusing on the signal process (hadronically decaying $t\bar{t}$ system and a Z decaying to ee/$\mu\mu$), 8% of the initial number of events remain. Setting the improvement cuts, i.e. a p_T cut of 30 GeV and an $|\eta| < 2.5$ cut on all truth partons originating from the $t\bar{t}$ system as well as a cut to exclude merged partons, reduces the number of events further by 24%. Thus, 0.6% of the initial number of events remain. For these remaining events, the matched share is then found to be 2.3%[7]. This is already better than the 1.1% obtained without the application of the improvement cuts, but is still extremely bad for a reconstruction algorithm.

---

[6]for an event selection of exactly 6 jets and exactly 2 b jets

[7]for an event selection with $\geq$ 6 jets and $\geq$ 2 b jets

Chapter 7

Multivariate reconstruction

In chapter 6 it has been outlined that the reconstruction of hadronically decaying $t\bar{t}$ systems in $t\bar{t}Z$ events is an essential, but not at all trivial task. In this context challenges have been identified which clearly limit the performance of the standard minimum-χ^2 reconstruction technique. Improvement cuts are only able to address these up to a certain point.

For that reason, an alternative reconstruction approach for hadronically decaying $t\bar{t}$ systems in $t\bar{t}Z$ events is developed and studied in this chapter. The basis for this approach is a multivariate (MVA) method. This concept has already been applied in the context of the associated production of a Higgs boson and a leptonically decaying single top quark [106, 107].

In this thesis, a Neural Network (NN) is chosen as a multivariate method. Its application consists of two steps as described in chapter 4.

First, the NN has to be trained with example data to learn the underlying features. This is done within the context of supervised learning. Therefore, the examples are required to be in the form of labeled hypotheses, i.e. the hypotheses need to be classified into a signal category (well-reconstructed $t\bar{t}$ systems) and a background category (badly-reconstructed $t\bar{t}$ systems). To obtain input data in this form, a suitable method to produce these has to be selected. This is conducted in section 7.1. Afterwards, also the hypotheses assigned to the background category have to be studied further as they combine a variety of physical scenarios. Therefore, different background categories are defined and investigated in section 7.2. Finally, the NN is trained on the selected data in section 7.3.

The second step of using an MVA method consists of evaluating the trained model. That is why the performance of the Neural Network is evaluated on un-labeled hypotheses in section 7.4.

To remain compatible with the 6j2b signal-region of the 2LOS $t\bar{t}Z$ analysis, a cut on ≥ 6 jets and ≥ 2 b jets is set together with all cuts, in particular the analysis and signal-region cuts, defined in chapter 5. If not stated otherwise, this event selection is used for the studies conducted in this chapter and is referred to as 'plain analysis cuts'.

7.1 Obtaining labeled input data for training

In order to perform an NN-based reconstruction of hadronically decaying $t\bar{t}$ systems in $t\bar{t}Z$ events, input data in the form of labeled hypotheses are required for the training process. Therefore, three different methods to produce these are presented in this section. Subsequently, their performance is compared using two event selections.

7.1.1 Description of three methods to obtain labeled input data for training

To obtain labeled input data for the training process, three different methods are considered. These are described below:

- **best possible hypothesis on quarklevel ('quarklevel method')**
  For this approach based on [108], the jets in the event are grouped into groups of 6 jets. Then, all combinations of placing these onto the positions of the 6 decay products of the $t\bar{t}$ system (b, $\bar{b}$, W1t, W2t, W1$\bar{t}$, W2$\bar{t}$) are formed[1]. For each of these combinations/hypotheses a ΔR_{total} between the selected jets and the truth partons originating from the $t\bar{t}$ system is calculated on quarklevel using the following definition:

$$\Delta R^{q}_{total} = \Delta R_{W1t} + \Delta R_{W2t} + \Delta R_{W1\bar{t}} + \Delta R_{W2\bar{t}} + \Delta R_{b} + \Delta R_{\bar{b}}$$

  Here, the individual ΔR values indicate the distances between the particular jets and their assigned truth partons.
  In the end, the combination/hypothesis with the minimum overall ΔR_{total} is labeled as the best possible reconstructed $t\bar{t}$ system and thus the 'signal'. The 'background' is formed by all the other hypotheses in the event. This procedure is repeated for every event.

- **best possible hypothesis on Wtlevel ('Wtlevel method')**
  This method follows the same procedure as described for the quarklevel method. However, here the 'physical' particles involved instead of the partons/jets resulting from their decay are used. In consequence, ΔR_{total} is now calculated on Wtlevel:

$$\Delta R^{Wt}_{total} = \Delta R_{W \text{ from } t} + \Delta R_{W \text{ from } \bar{t}} + \Delta R_{t} + \Delta R_{\bar{t}}$$

  In order to determine the individual ΔR values in this definition, the Lorentz vectors of the decay partons/jets have to be added up first in order to obtain the Lorentz vectors for the W bosons/top quarks. Using these, the ΔR values between the W bosons/top quarks on reco and truth level can be determined.

---

[1]In contrast to the minimum-χ^2 reconstruction, no b-tagging conditions are set on the b/$\bar{b}$ positions.

- **'assign all combinations to categories' (ctc method)**
  For this method again all possible groups of 6 jets are formed in an event and all possibilities to assign these to the 6 decay positions (b, $\bar{\text{b}}$, W1t, W2t, W1$\bar{\text{t}}$, W2$\bar{\text{t}}$) are constructed[2]. In the following, every single combination/hypothesis is assigned to one of the 3 'truth-matching' categories (matched, combinatorial and non-matched) defined in subsection 6.1.3.
  Finally, all hypotheses of the 'matched' category are considered as the signal. The background is in consequence formed by events of the 'combinatorial' and 'non-matched' categories.

7.1.2 Performance comparison of the three methods

In order to determine the best performing method to produce input data for the training of the MVA method, the results of the described methods are now compared. Therefore, selected observables, which represent the fundamental trends best, are investigated.
First, the two ΔR_{total} distributions, i.e. $\Delta R_{\text{total}}^{\text{q}}$ and $\Delta R_{\text{total}}^{\text{Wt}}$, are studied. This targets the alignment of the truth partons originating from the $t\bar{t}$ system and the selected jets or rather the alignment of the truth particles with the reconstructed particles within the signal hypotheses of the particular method. In a next step, further observables are used to analyze the extent to which the signal hypotheses are able to reproduce the physical properties of the $t\bar{t}$ system. For that purpose, some Wtlevel observables, i.e. observables related to the reconstructed top quarks/W bosons ($m_{t\bar{t}}$, $\cos(\Theta^*)$), and some quarklevel observables, i.e. observables related to the separate partons ($p_T(\text{W1t})$, is-btagged(b)), are examplarily presented, covering different physical aspects.

In addition, the methods are not only compared among each other, but also their performance using two event selections is analyzed: On the one hand, the plain analysis cuts are used which are referred to as 'without improvement cuts'. On the other hand, the improvement cuts found during the investigation of the χ^2-reconstruction method in chapter 6 are applied on top of the plain analysis cuts which is referred to as 'with improvement cuts'.
Towards the end, also the hypotheses yields of the signal and background categories of the three methods are compared, each using the two event selections. Finally, the best performing method to obtain labeled input data for the training process is chosen together with an event selection.

Performance comparison using ΔR_{total}

At first, the normalized ΔR_{total} distributions of the three methods are compared using the particular signal hypotheses. ΔR_{total} is calculated on quarklevel ($\Delta R_{\text{total}}^{\text{q}}$) and on Wtlevel

---

[2]In contrast to the minimum-χ^2 reconstruction, no b-tagging conditions are set on the b/$\bar{\text{b}}$ positions.

($\Delta R_{\text{total}}^{\text{Wt}}$) for all methods in order to allow for a complete comparison. All resulting normalized distributions are depicted for both event selections in figure 7.1.
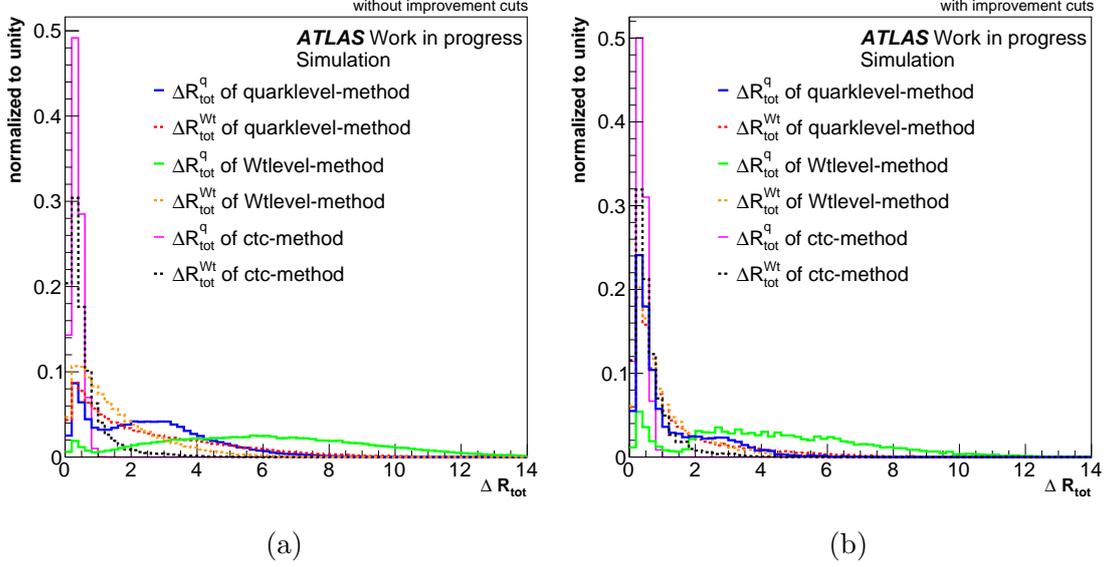
Figure 7.1: Normalized ΔR_{total} distributions for the particular signal hypotheses of the three presented methods (subsection 7.1.1) for the two event selections, without improvement cuts (a) and with improvement cuts (b). ΔR_{total} is calculated on quarklevel ($\Delta R_{\text{total}}^{\text{q}}$) and Wtlevel ($\Delta R_{\text{total}}^{\text{Wt}}$) for every method.

In figure 7.1a the normalized results for the event selection without improvement cuts are shown. The ΔR_{total} distributions calculated on Wtlevel and quarklevel for the three methods are depicted in different colors: The distributions produced by the quarklevel method are indicated in blue and red for $\Delta R_{\text{total}}^{\text{q}}$ and $\Delta R_{\text{total}}^{\text{Wt}}$ respectively. The yellow and green distributions are obtained using the Wtlevel method for a $\Delta R_{\text{total}}^{\text{q}}$ and $\Delta R_{\text{total}}^{\text{Wt}}$ respectively. The two distributions resulting from the use of the ctc method are shown in purple and black.

Focusing on the distributions obtained by using the ctc method, quite narrow distributions with a sharp peak at low ΔR_{total} values can be identified. In contrast, the remaining distributions are broader and some even show long tails towards larger values.
This observation is not surprising as an implicit cut on $\Delta R < 0.3$ between the jet and the parton is included in the definition of the 'matched category' (subsection 6.1.3) which defines the signal of the ctc method. The other two methods only use the *best possible* reconstructed $t\bar{t}$ systems as their signal, i.e. the hypotheses with the lowest $\Delta R_{\text{total}}^{\text{q}}$ or rather $\Delta R_{\text{total}}^{\text{Wt}}$ value in the events.

A striking fact to mention is that the distributions related to a ΔR_{total}^{Wt} are always broader than those related to a ΔR_{total}^{q}. This can be explained by considering that multiple configurations of jets are possible for one particular ΔR_{total}^{Wt} value. This is a consequence of the fact that the Lorentz vectors of the single jets and partons have to be added up first before the ΔR and the ΔR_{total}^{Wt} value can be calculated on Wtlevel. Thus, in an extreme case, two jets with quite large ΔR distances to their assigned truth partons can result in an acceptable ΔR value between the W bosons/top quarks on reco and truth level as the deviations might cancel out. The result is then a relatively small ΔR_{total}^{Wt} value.

Another fact worth emphasizing is related to the distributions produced by the quarklevel method and by the Wtlevel method for ΔR_{total}^{q}: both ΔR_{total}^{q} distributions have a dip situated at about a value of 1 followed by a shoulder towards higher values. The value of the dip also corresponds to the largest value of the distribution obtained by using the ctc method with ΔR_{total}^{q}. Hence, it can be concluded that, if jets originating from the $t\bar{t}$ system are considered and assigned to the right positions during reconstruction, they also match quite closely the truth-parton positions. This results in small ΔR(jet, parton) distances and leads to small ΔR_{total}^{q} values. If jets not belonging to the $t\bar{t}$ system are considered, the ΔR_{total}^{q} values increase considerably. This leads to the formation of the observed shoulder towards larger values.

In figure 7.1b the normalized results of the event selection with improvement cuts are shown in the same colors as in figure 7.1a.
All in all, the same basic trends can be observed: Regarding the distributions produced by the ctc method, a narrow and sharp peak at low values with no tail towards higher values can be observed for both ΔR_{total} calculations.
The distributions resulting from the use of the quarklevel method and of the Wtlevel method for ΔR_{total}^{q} still exhibit a dip at about a value of 1 and a shoulder towards higher values. However, this shoulder is reduced considerably by applying the improvement cuts and also the distributions end at lower values. Moreover, the peak gets sharper and rises higher at the same time. Especially for the quarklevel method the observed difference is significant.
From these observations, it can be deduced that applying the improvement cuts clearly improves the alignment of jets and partons and consequently leads to overall lower ΔR_{total} values. This is expected as with the application of the improvement cuts an adaption of reco-level and truth-level phase spaces is implied. It results in an increased matching probability of the truth partons and the jets. Furthermore also the implied exclusion of merged-parton scenarios reduces matching ambiguities and supports the aforementioned trend.

All in all, it can be retained that the distributions resulting from an event selection without improvement cuts are in general broader and reach to higher ΔR_{total} values than those resulting from an event selection with improvement cuts. Only the distributions produced by the ctc method stay sharp and also end at the same values in both cases. Thus, the

ctc method is a good option in combination with both event selections.

Regarding the distributions produced by the quarklevel method and the Wtlevel method, broader distributions in comparison to the ctc method have been found for both event selections. Comparing the distributions obtained by using these two methods among each other, the best result is achieved by the quarklevel distribution in combination with an event selection including improvement cuts.

Performance comparison using the invariant mass of the $t\bar{t}$ system $m_{t\bar{t}}$

In a next step, the Wtlevel observable $m_{t\bar{t}}$ is used to investigate the ability of the three methods to reproduce physical information on Wtlevel.

Figure 7.2 shows the normalized distributions of the signal (blue) and the background (red) hypotheses selected by the three methods. In order to provide a reference, also the truth information (yellow) for the signal hypotheses is added. This is done for an event selection with (b, d, f) and without (a, c, e) improvement cuts.

In figure 7.2a and 7.2b the normalized results of the quarklevel method are depicted for the two event selections. In the case without improvement cuts (figure 7.2a), a significant discrepancy between the shapes of signal and background can be observed. This indicates a good separation power of the $m_{t\bar{t}}$ observable. In addition, also a significant difference between the shapes of the signal and the truth information can be found. So, it can be concluded that the quarklevel method used in combination with the event selection without improvement cuts is not very reliable in selecting well-reconstructed $t\bar{t}$ systems for the signal category.

With improvement cuts applied (figure 7.2b), the discrepancy between the truth and the signal distribution vanishes within statistical fluctuations, except at low $m_{t\bar{t}}$ values. The separation between signal and background is kept at the same time.

This means that by setting the improvement cuts, the truth information fits the reco information better and thus, hypotheses containing well-reconstructed $t\bar{t}$ systems are selected more reliably. Furthermore, it can be seen that $m_{t\bar{t}}$ is a quite well-separating observable. Referring to the difference between signal and truth information at low $m_{t\bar{t}}$, it is striking that it can be found in all 6 cases. The truth distributions always start at a value of about 350 GeV which is about twice the top-quark mass used by the MC generator. Hence, the deviation can be assumed to result from hadronization.

The normalized results of the Wtlevel method are shown in figure 7.2c and 7.2d. The situation is similar to the one of the quarklevel method. However, compared to the signal hypotheses produced by the quarklevel method, those produced by the Wtlevel method with applied improvement cuts reproduces the truth information slightly worse.

In figure 7.2e and 7.2f the normalized results of the ctc method are displayed. Here, the truth information matches the signal shape quite perfectly for both event selections. At the

same time there is a significant separation between the signal and background hypotheses.
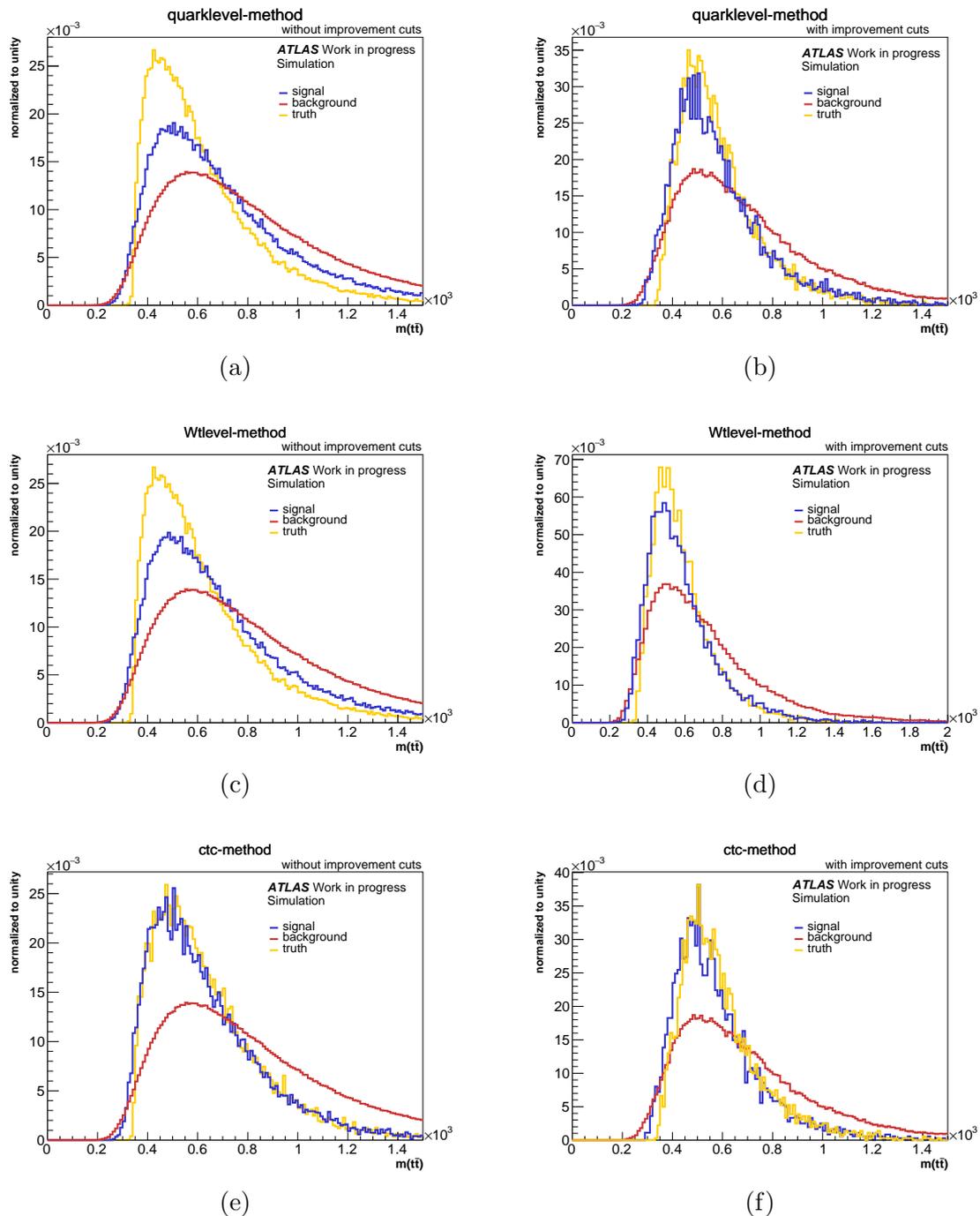
Figure 7.2: Normalized $m_{t\bar{t}}$ distributions for the particular signal and background hypotheses of the three methods (subsection 7.1.1) in combination with the truth information for the signal hypotheses for the two event selections without (a, c, e) and with improvement cuts (b, d, f).

Concluding, the ctc method is the only method which is able to reproduce the physical observables on Wtlevel for the event selection without improvement cuts well. When choosing the event selection with improvement cuts, either the quarklevel or the ctc method should be used as they reproduce the physical information slightly better than the Wtlevel method.

Performance comparison using the transverse momentum of the W1t jet/parton p_T(W1t)

In a next step, it is checked how well physical information on quarklevel can be reproduced. Therefore, another kinematic observable, the transverse momentum of the W1t jet/parton, p_T(W1t), is used to compare the three methods and the two event selections. Figure 7.3 displays the normalized distributions of the signal (blue) and background (red) hypotheses for each of the three methods. In addition, also the truth information for the signal hypotheses (yellow) is included. The results are shown for the event selections without improvement cuts (a, c, e) and with improvement cuts (b, d, f).

In figure 7.3a the normalized results of the quarklevel method for the event selection without improvement cuts are depicted. Here, the shapes of signal and background are quite similar which indicates a low separation power of the p_T(W1t) observable. The shape of the truth information is however quite different to the signal one. Especially the discrepancy between the starting point of the truth distribution at 0 GeV and the start of the signal/background distributions at 25 GeV is eye-catching. This is a result of the different p_T cuts on truth and reco level: While on reco level a cut on $p_T > 25$ GeV is applied, no p_T cut is applied on truth level.
Setting the improvement cuts (figure 7.3b), the observed discrepancy between the starting points of the three distributions is visibly reduced. This is due to the implied p_T cut of 30 GeV on truth level. As a consequence of this adaption, the truth and the signal distributions are a lot closer together. This indicates a good reproduction of the truth information by the signal hypotheses which is not the case for the event selection without improvement cuts. The small remaining differences could be eliminated by optimizing the improvement cuts for the multivariate reconstruction.
At the same time, the small difference between the signal and the background shapes is still present. This indicates a low separation power of the p_T(W1t) observable also for the event selection with improvement cuts.

In figures 7.3c and 7.3d the normalized results of the Wtlevel method are displayed for the two event selections.
In the case without improvement cuts (figure 7.3c), the same discrepancy between the truth and the signal/background distributions can be found as in figure 7.3a. This is again due to the missing p_T cut on truth level. In contrast to the distributions resulting from the use of the quarklevel method, also a significant difference between the signal shape and background shape is found. The truth distribution is in this case better reproduced by the

background distribution though.
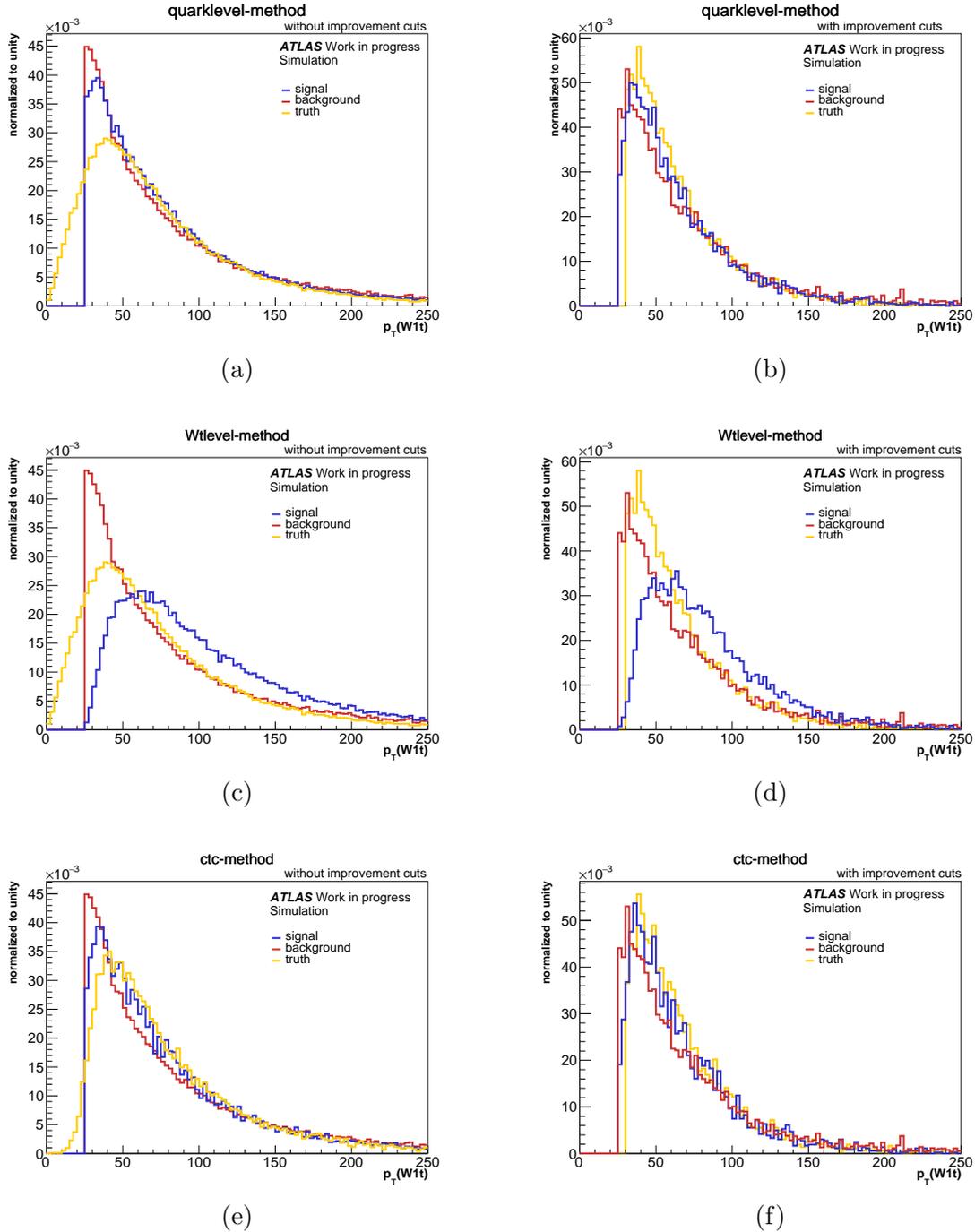
Figure 7.3: Normalized $p_T(W1t)$ distributions for the signal and background hypotheses of the three methods (section 7.1) combined with the truth information for the signal hypotheses for an event selection without (a, c, e) and with improvement cuts (b, d, f).

Applying the improvement cuts (figure 7.3d) reduces the discrepancy between the truth and the signal/background shapes. However, now the shape of the truth information resembles the background shape even more. Thus, for both event selections, the Wtlevel method is obviously not able to select signal hypotheses which reproduce the physical information on quarklevel properly. This fact is not surprising as the signal hypotheses are selected based on their low $\Delta \text{R}^{\text{Wt}}_{\text{total}}$ value. Thus, information on quarklevel is not used explicitly and obviously cannot be resolved in consequence. As for the reconstruction process the best-fitting jets have to be selected, quarklevel information is important. Therefore, the Wtlevel method should be used cautiously.

Finally, the normalized results of the ctc method are depicted for the two event selections in figure 7.3e and 7.3f.
In the case without improvement cuts (figure 7.3e) all three distributions start at almost the same value even though there is no p_T cut on truth level. This might be due to the fact, that here only matched hypotheses are considered as signal for both reco and truth level. Moreover, in particular the signal and background distributions resemble each other a lot which indicates a low separation power of the $p_\text{T}(\text{W1t})$ observable. Overall, the truth information is reproduced slightly better by the signal information. Thus, signal hypotheses are selected which are also able to reproduce the information on quarklevel.
These trends become even more obvious when applying the improvement cuts (figure 7.3f). Here, the signal is even closer to the truth information due to the additional p_T cut on truth level and also the signal and background distributions are more similar. Thus, the ctc method is a suitable option to produce labeled input data for both event selections.

All in all, it can be noted, that the Wtlevel method should not be used for obtaining labeled input data. The reason is that the signal hypotheses cannot reproduce the physical properties on quarklevel[3] which is essential for the reconstruction process.
The quarklevel method should only be used in combination with the event selection including improvement cuts as in this case the truth information is reproduced well by the signal hypotheses.
The ctc method shows good results for both event selections and thus is suitable in all cases.

Performance comparison using is-btagged(b)

In a next step, another (non-kinematic) observable on quarklevel is used: is-btagged(b). It is a binary observable indicating whether the jet placed on the b reconstruction position has a b-tag (value of 1) or does not have a b-tag (value of 0). As the b reconstruction position is considered here, a high share of signal hypotheses for a value of 1 and a high share of background hypotheses for a value of 0 are expected.
On truth level, the is-btagged observable does not exist as b-tagging is not necessary for

---

[3]Further observables were studied showing the same trend.

truth partons. Therefore, the truth information is not considered in the following. In consequence, figure 7.4 depicts only the normalized distributions of the signal (blue) and background (red) hypotheses of the three methods for the two event selections.

In figures 7.4a and 7.4b the normalized results of the quarklevel method are shown for event selections without and with improvement cuts.
In the case without improvement cuts (figure 7.4a), the share of b-tagged jets on the b reconstruction position is 85%. This is quite high as expected. Only about 15% of the jets placed on the b reconstruction position are not b-tagged. This might be due to the fact that the b-tagging working point chosen has an efficiency of 77%, i.e. only 77% of b-quark initiated jets are b-tagged. Another possibility would be that the selected signal hypothesis contains a jet which does not originate from the $t\bar{t}$ system or contains a light jet on the b reconstruction position due to merging.
Regarding the background, there is a 50% chance of finding a b-tagged jet on the b reconstruction position. The large share of b-tagged jets on the b reconstruction position can have multiple reasons. One of them is misidentification of charm/light jets as b jets which happens in about 16% of the cases (subsection 3.2.5). Furthermore, the jet placed on the b reconstruction position might indeed be the correct one, but some other jets have large ΔR(jet, parton) values. Therefore, this hypothesis is not assigned to the signal category due to a large ΔR_{total} value. Moreover, the selected signal hypothesis might contain a b jet not originating from the $t\bar{t}$ system.

In the case with improvement cuts (figure 7.4b), the signal share of b-tagged jets on the b reconstruction position rises to about 90%. In consequence, the signal share of jets on the b reconstruction position without a b-tag decreases to about 10%. Also the share of background hypotheses with a b-tagged jet on the b reconstruction position is reduced to about 25%, which corresponds roughly to the misidentification rates mentioned above. The share of background hypotheses containing a non-b-tagged jet on the b reconstruction position rises to about 75% correspondingly. This is a result of excluding events containing merged truth partons which also has an impact on reco level. Furthermore, adapting the phase spaces on reco and truth level allows for better alignment of the partons and the jets originating from the $t\bar{t}$ system.
All in all, the separation power between signal and background distributions is considerably increased by applying the improvement cuts. Furthermore, also the result for this event selection is closer to the expected situation.

In figures 7.4c and 7.4d the normalized results of the Wtlevel method are shown for the two event selections. Here, the same basic trends can be found as mentioned in the discussion of figure 7.4a and 7.4b.
In the case without improvement cuts (figure 7.4c), a lower share of signal hypotheses having a b-tagged jet on the b reconstruction position (circa 65%) can be found compared to figure 7.4a (85%). Consequently the share of signal hypotheses yielding a non-b-tagged jet on the b reconstruction position is increased to about 35% (15%).
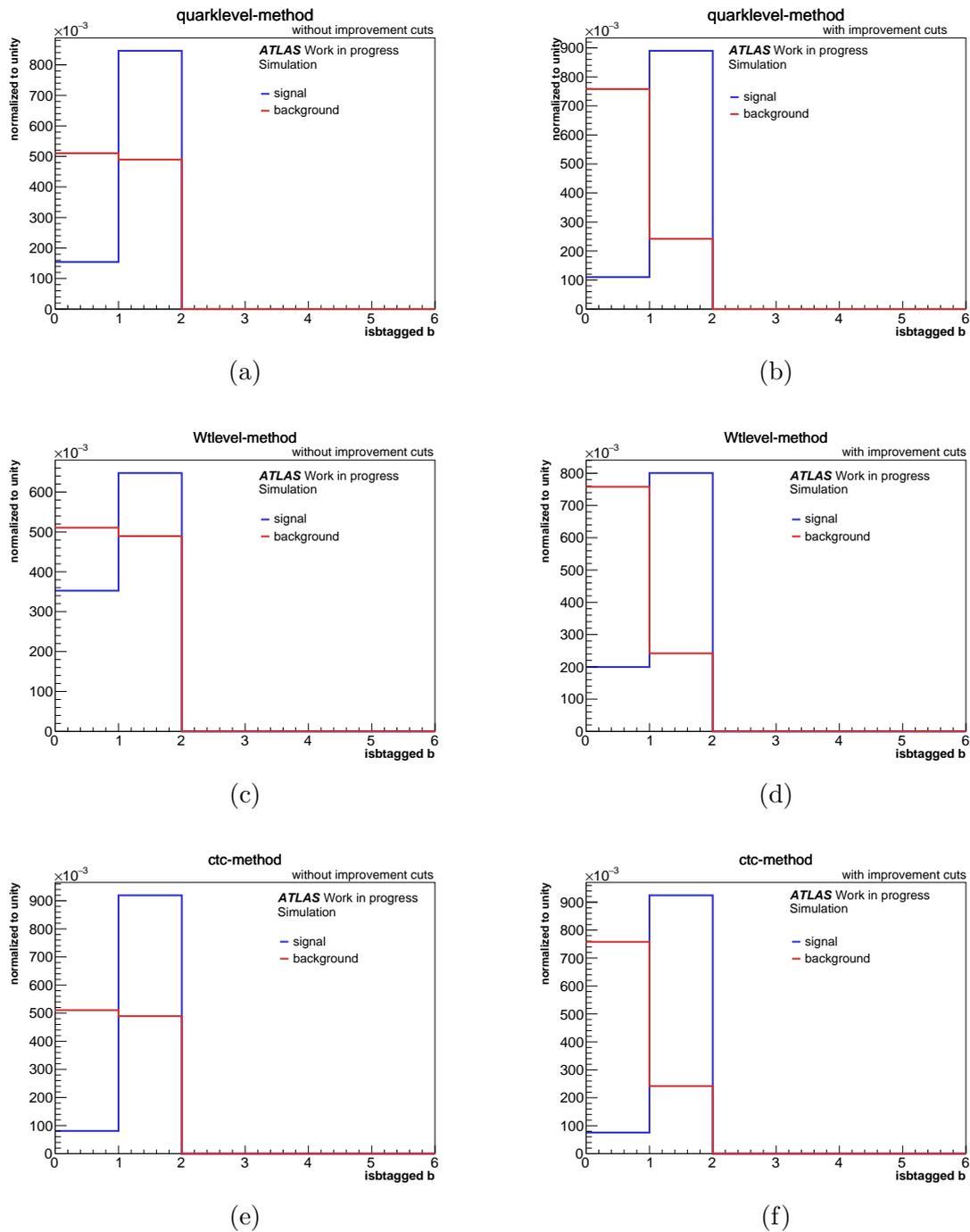
Figure 7.4: Normalized is-btagged(b) distributions for the particular signal and background hypotheses of the three methods (section 7.1.1) for event selections without (a, c, e) and with improvement cuts (b, d, f).

Here, an additional reason for this high signal share of non-b-tagged jets on the b recon-

struction position is the possibility to have multiple jet configurations for one particular ΔR_{total} value on Wtlevel. This is a consequence of the fact that the Lorentz vectors of the single jets and partons have to be added up first.

For the background hypotheses, there is still a 50% chance to have a b-tagged or non-b-tagged jet on the b reconstruction position.

For the event selection with improvement cuts (figure 7.4d), the signal share of b-tagged jets on the b reconstruction position is 80% which is lower than the one for the quarklevel method (90%). In consequence also the signal share of non-b-tagged jets on the b reconstruction position rises to about 20%. This roughly doubles the share of the quarklevel method indicating once again that the Wtlevel method cannot resolve physical information on quarklevel well. The background shares are approximately the same as for the quarklevel method. Based on this, an increased separation power can again be noticed for an event selection with improvement cuts. In addition, the result for this event selection is closer to the expected one.

In the last figures 7.4e and 7.4f, the normalized results of the ctc method are shown for the two event selections. They exhibit again the same basic trends as found in figure 7.4a and 7.4b.

In figure 7.4e, representing the case without improvement cuts, a larger signal share of b-tagged jets on the b reconstruction position (92%) is found. Consequently, a lower signal share of non-b-tagged jets on the b reconstruction position (8%) is found compared to figure 7.4a. For the background, still half of the hypotheses contain a b-tagged jet on the b position and the other half does not.

Applying the improvement cuts (figure 7.4f) yields the same signal distribution as for an event selection without improvement cuts. In contrast, the background share of b-tagged jets on the b position decreases to about 25%. Analogously to the other two methods, a huge increase in the separation power between signal and background distributions is found when applying improvement cuts. Furthermore a constantly good reproduction of the truth information by the signal hypotheses of the ctc method can be deduced.

Concluding, it can be retained that the expected observation, i.e. a high signal and a low background share of b-tagged jets on the b reconstruction position, is reproduced best by the ctc method and the quarklevel method combined with an event selection with improvement cuts. As applying these cuts increases the separation power of the is-btagged(b) observable considerably, the event selection without improvement cuts should not be used. A significant separation power is important as this enables the NN later on to better distinguish between signal and background.

Performance comparison using the cosine of the helicity angle $\cos\Theta^*$

As a last observable, the cosine of the helicity angle $\cos(\Theta^*)$ (subsection 2.1.2) is used for comparison of the three methods and the two event selections.

In figure 7.5 the normalized distributions of $\cos(\Theta^*)$ of the top quark[4] are displayed for the signal and background hypotheses of the three methods (in red/blue respectively). In addition the truth information (yellow) for the signal hypotheses is included. The normalized results of the three methods are depicted for event selections without improvement cuts (a, c, e) and with improvement cuts (b, d, f).

In figures 7.5a and 7.5b, the normalized results of the quarklevel method are illustrated for the two event selections.

In the case without improvement cuts (figure 7.5a), a significant difference in truth and signal shapes can be observed. While the signal distribution is symmetric and centered around 0, the shape of the truth information is shifted to the left. This difference can be explained by considering the selection of the down-type quark within the definition of $\cos\Theta^*$ (subsection 2.1.2): On truth level the down-type quark is selected using the pdgId which is a reliable and precise observable. On reco level, this truth-level observable cannot be used. Instead, the so-called jet charge (subsection 3.2.6) has to be used to distinguish between up- and down-type jets. Unfortunately, this observable is not very precise and thus the left-shift observed for the truth information cannot be resolved on reco level.

Physically, the left-shifted shape of the truth information indicates that the W boson originating from the top quark is more often left-handed. Thus the down-type quark is emitted at large angles Θ^* measured in the rest frame of the W boson. This corresponds to merged-parton scenarios.

Another striking aspect is the high separation power of the cosine of the helicity angle regarding signal and background distributions. The values of the background hypotheses are almost equally distributed and show a symmetric dip around zero. The slight peaks at values of -1 and 1 indicate, that there are some well-reconstructed top quarks in the background category. Considering them as signal would help to reduce the significant difference between the signal and background shape.

Applying improvement cuts (figure 7.5b), it can be found that the shape of the truth information matches the signal shape well within statistical fluctuations. It is also eye-catching that the shape of the truth information is now centered around 0 and shows no longer the left-shift observed in the case without improvement cuts. This is a result of the implied exclusion of merged partons within the improvement cuts. The partons belonging to the $t\bar{t}$ system now originate more often from a longitudinal W boson. Therefore, they are emitted preferentially perpendicular to the W boson in the rest frame of the W boson. Thus, the quarklevel method should only be used in combination with improvement cuts to ensure that the same information is contained in the truth and in the signal hypotheses.

---

[4]The same distributions can also be found for the case of the anti-top quark, thus only the results of the top quark are discussed here.
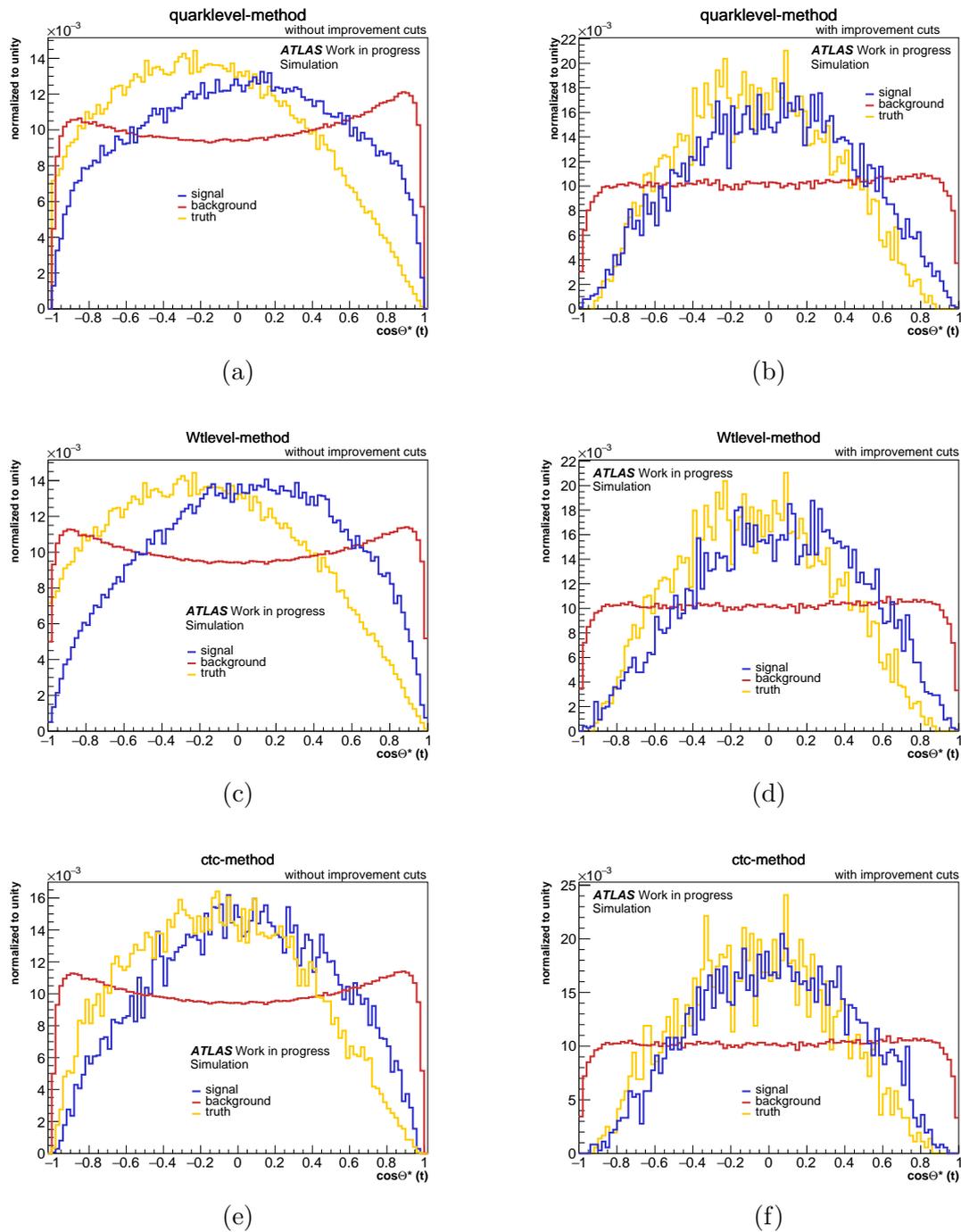
Figure 7.5: Normalized $\cos(\Theta^*)$ distributions for the particular signal and background hypotheses of the three methods (subsection 7.1.1) in combination with the truth information for the signal hypotheses for event selections without (a, c, e) and with improvement cuts (b, d, f).

Furthermore, the shapes of signal and truth information are both a lot narrower than in the case without improvement cuts. This change in shape and the corresponding change of the physical situation is investigated step by step in figure 7.6 later on. Another observation to mention in figure 7.5b is that the dip in the background distribution vanishes. This is assumed to be due to the cut on merged partons. As in these cases the whole event is rejected, they are no longer present in the background category.

The normalized results of the Wtlevel method are displayed in figure 7.5c and 7.5d. Here, the same basic trends as for the quarklevel method can be found and therefore are not discussed again. The only difference is that in the case with improvement cuts, the discrepancy between the shapes of the truth information and the signal is slightly larger than for the quarklevel method. Thus, the use of the Wtlevel method should be restricted to the use in combination with improvement cuts.

Figures 7.5e and 7.5f display the normalized results of the ctc method for the two event selections. This corresponds well to the trends found in the discussion of the quarklevel-method results (figure 7.4a and figure 7.4b). However, for the event selection without improvement cuts (figure 7.5e), the truth information matches the signal information a lot better than for the other two methods.
In the case with improvement cuts the two distributions match perfectly within statistical fluctuations. The background shape stays the same compared to the other methods.

Concluding, it can be retained that the signal hypotheses of the ctc method reproduce the truth information well for both event selections. Regarding the other two methods, this is only the case for an event selection with improvement cuts. Thus, in order to assure that the signal hypotheses selected by the three methods contain the same information as on truth level, the event selection with improvement cuts is preferred.

In a next step, the observed change of the shape in $\cos(\Theta^*)$ of the truth information between the event selections without and with improvement cuts is investigated in more detail. For this reason, the improvement cuts are set successively and also different combinations of them are used. The normalized distributions of the signal and background hypotheses (in blue/red respectively) as well as the truth information (yellow) of the signal hypotheses are examplarily displayed for the quarklevel method for the different cuts.

As a reference, figure 7.6a depicts the normalized results for an event selection without improvement cuts. Here a left-shifted shape of the truth information compared to a symmetric signal shape centered at 0 can be observed.

Figure 7.6b presents the normalized results if only the η^{truth} cut is applied. Compared to figure 7.6a, no significant differences can be found. This indicates that events with truth partons originating from the $t\bar{t}$ system situated outside of $|\eta| < 2.5$ are not responsible for

the change in shape.

(a) without improvement cuts

(b) η^{truth}-cut

(c) $p_{\text{T}}^{\text{truth}}$-cut

(d) exclusion of merged partons

(e) $p_{\text{T}}^{\text{truth}}$ and merged cut

(f) with improvement cuts
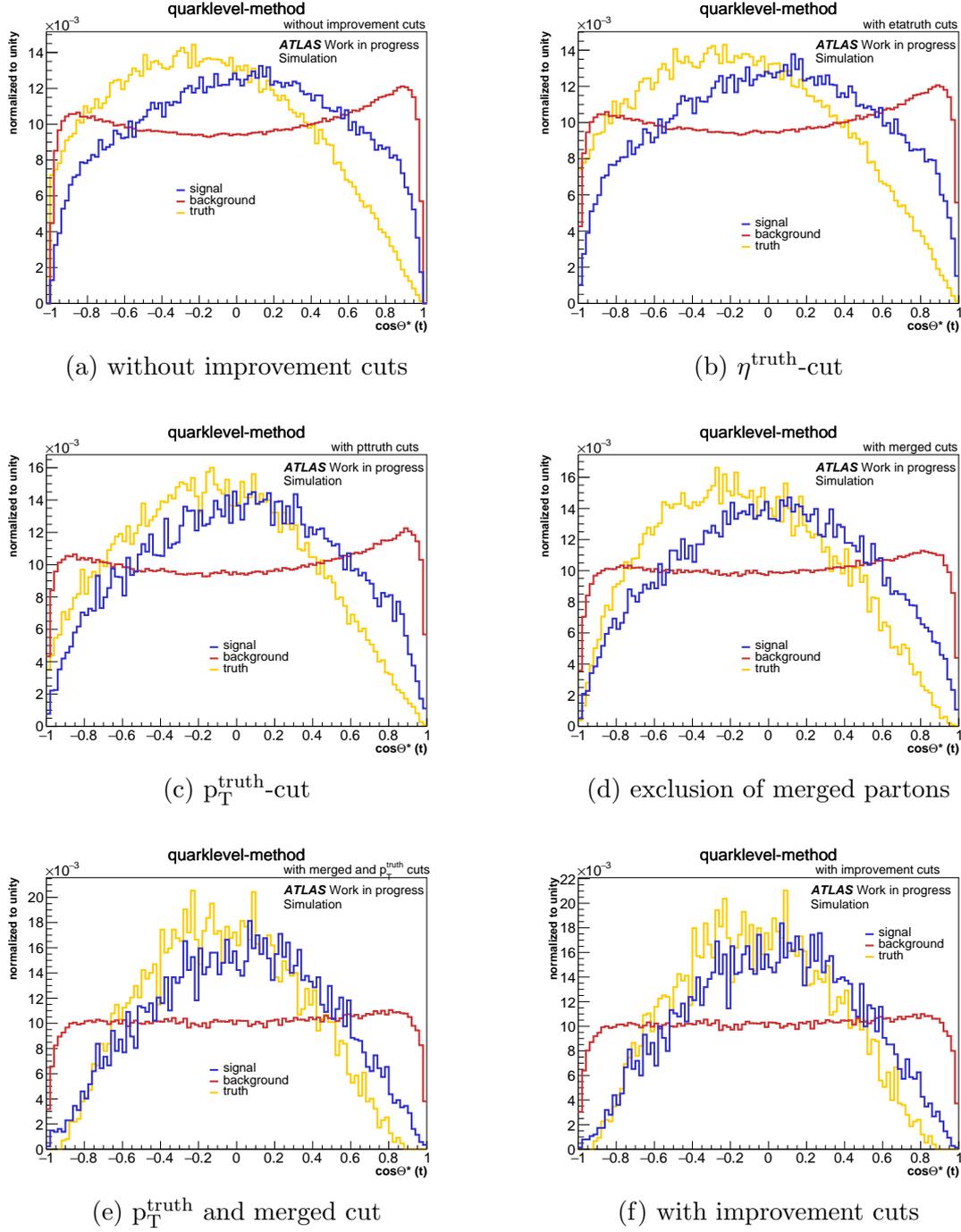
Figure 7.6: Normalized $\cos(\Theta^*)$ distributions for the signal and background class of the quarklevel method (subsection 7.1.1) in combination with the truth information for the signal hypotheses for various additional cuts on top of the plain analysis cuts.

The normalized results of an event selection applying only the p_T^{truth} cut on top of the plain analysis cuts are shown in figure 7.6c. In comparison to figure 7.6a, the shoulder in the signal and truth shapes towards low values is reduced. In consequence, the shapes appear to be more central. This can be explained by the fact that in the case of a left-handed W boson the down-type quark is emitted at large angles of Θ^* in the rest frame of the W boson and thus in the opposite direction than the direction of motion of the W boson. Hence, the p_T of the down-type quark is lower and can no longer fulfill the additional cut on p_T^{truth}. In consequence, entries at low values of $\cos(\Theta^*)$ are cut off and the shape becomes more central. This represents the fact that more decay products from longitudinal W bosons are used for the reconstruction.

Furthermore, for the signal shape also the small shoulder at high $\cos(\Theta^*)$ values vanishes. This can be explained by considering that the p_T^{truth} cut is applied to all partons originating from the $t\bar{t}$ system, thus also to the up-type quark. As events which do not fulfill this cut on truth level are removed completely, this has also an impact on the reco-level shapes.

All in all, the signal and truth shapes become narrower and are situated more central. The background shapes are unchanged.

Figure 7.6d displays the normalized results of an event selection excluding merged partons on truth level on top of the plain analysis cuts.

In comparison to figure 7.6a, a lot of entries are lost at low $\cos(\Theta^*)$ values, i.e. at large angles Θ^*. This can be explained by considering that for left-handed W bosons, the down-type quark and the b quark are situated close together, i.e. in a ΔR distance ≤ 0.8, and thus merging occurs quite often as found in chapter 6. Thus, the exclusion of merged partons results in a more centrally situated shape of the truth information.

Regarding the shape of the signal hypotheses it can be observed that the small shoulder at large $\cos(\Theta^*)$ values vanishes. Also the depth of the dip in the background shape is reduced. As before, this can be explained by the fact, that events which do not pass this cut on truth level are removed completely and thus an impact can also be observed on reco level.

Combining the merged and the p_T^{truth} cut yields the normalized results in figure 7.6e. Here, the separate shapes as well as the deviations between them are almost the same as for an event selection using all the improvement cuts (7.6f). The small remaining differences are then removed by the η^{truth} cut.

All in all, it has been deduced, that the left-shifted shape of the truth information is mostly due to events containing merged partons and low-p_T partons.

Discussion of hypotheses yields

Finally, the hypotheses yields for both event selections and all three methods are compared. The resulting numbers are presented in the following tables:

method	signal	background
quarklevel	99022	1.95×10^9
Wtlevel	99022	1.95×10^9
ctc	17428	1.95×10^9

(a) without improvement cuts

method	signal	background
quarklevel	7512	0.18×10^9
Wtlevel	7512	0.18×10^9
ctc	3614	0.18×10^9

(b) with improvement cuts

Table 7.1: Hypotheses yields of the three methods (subsection 7.1.1) and the two event selections, without (a) and with (b) improvement cuts.

For the event selection 'without improvement cuts', the number of signal hypotheses is the same for the quarklevel and the Wtlevel method as expected. Both of them have one signal hypothesis per event. The number of signal hypotheses for the ctc method is smaller compared to the other two, as here also the matching probability needs to be taken into account. In consequence, it is not possible to assign a hypothesis to the matched category for every event.
The same trends can also be observed for the event selection with improvement cuts.

Focusing on the exact number of signal hypotheses of the different methods, it can be found that the number of 3614 signal hypotheses obtained by the ctc method in combination with improvement cuts is far too small to train an MVA method.[5] For the training process, this sample has to be split up into a training and testing subset. Thus, the actual number of signal events from which the method can learn is only half and thus even smaller. The number of 7512 hypotheses provided by the Wtlevel and quarklevel method in combination with applied improvement cuts is already a bit better, but is still at the lower limit.
For the event selection 'without improvement cuts', the yields are in general larger. The number of signal hypotheses for the ctc method increases to 17428 and for the Wtlevel and quarklevel methods increases to 99022. Both yields are one order of magnitude greater compared to the ones for the event selection 'with improvement cuts' and thus yield a more suitable number for training.

Another fact to mention is that for all methods and both event selections the number of background hypotheses is orders of magnitudes larger than the number of signal hypotheses. Thus, an additional requirement for the MVA method used for the reconstruction is the ability to cope with the imbalance between the numbers of signal and background.

---

[5]As a rule of thumb about 10000 examples are necessary in order to be able to train an MVA method which also performs well on new data. [74]

Conclusion of the performance comparison and selection of a method to obtain labeled input data

In this section, the three methods investigated - the quarklevel , the Wtlevel and ctc method - have been compared for two event selections, with and without improvement cuts.

It has been found that the methods capable of truthfully reproducing the physical properties are the ctc method for both event selections and the quarklevel method in combination with applied improvement cuts. The Wtlevel method is excluded as the selected signal hypotheses are not able to reproduce physical information on quarklevel well as found for the transverse momentum $p_T(W1t)$. This is essential as within the reconstruction process the single jets have to be selected on quarklevel. Thus, out of the 6 options 3 remain: the ctc method in combination without improvement cuts, the ctc method with applied improvement cuts and the quarklevel method with applied improvement cuts.

In addition, it has been observed in the performed studies that applying the improvement cuts helps selecting good signal hypotheses and leads to a truthful reproduction of the physical properties as observed for e.g. the helicity angle and the mass of the $t\bar{t}$ system. Furthermore, it improves the separation between the signal and background shapes. Therefore, preferably a method in combination with applied improvement cuts should be used. Thus, the final decision has to be made between the ctc method and the quarklevel method, both used in combination with improvement cuts.

Comparing the hypotheses yields, it has been found furthermore that the ctc method for an event selection with applied improvement cuts has to be excluded. The reason for this is the small number of signal hypotheses selected by this method. Even though this method reproduces the physical information very well, the MVA method needs a sufficiently high number of signal hypotheses to learn from which this method cannot provide.

Based on this finding, the quarklevel method with applied improvement cuts is finally chosen to produce the labeled input data for the training process in this thesis[6].

7.2   Investigation of the background hypotheses

In the previous section, the quarklevel method in combination with the event selection 'with improvement cuts' has been chosen to produce the input data for the multivariate method. In this context, a significant difference in the number of signal and background hypotheses has been found. This is not surprising as all possible combinations are con-

---

[6]If instead of a high quality of the input data a larger number of signal hypotheses is preferred, the ctc method in combination with the event selection 'without improvement cuts' would also be an option.

sidered in the process of selecting the signal hypotheses and all non-selected hypotheses are labeled as background. In consequence, the background category is a collection of hypotheses containing various physical scenarios.

However, this large discrepancy between the number of signal and background hypotheses might limit the performance of the multivariate method. As it has been found in the previous section, that the number of signal hypotheses is already quite low, the background hypotheses are studied in more detail in this section. In a first step, different background categories for the specific physical scenarios are defined and the corresponding hypotheses are assigned to them. Subsequently, the distributions resulting from the hypotheses of the different categories are compared to the signal distribution in order to assess the extent of similarity. Therefore, some well-discriminating observables are examplarily investigated. The plots of further observables can be found in appendix A. The goal is to identify the categories whose distributions are most similar to the signal as these are expected to be most difficult for the MVA method to distinguish from the signal.

7.2.1 Definition of categories for the background hypotheses

In order to identify the background hypotheses most similar to the signal, distinct background categories are defined in the following.

In a first step, the so-called 'flip categories' are defined. For that purpose, each jet needs to be assigned to the best-fitting truth parton with regard to the ΔR distance. Then, it is counted how many jets are not assigned to the corresponding parton. Based on this number, a number of 'flips' is determined which indicates how many switches would be needed in order to obtain the correct $t\bar{t}$ configuration. As one switch always involves two jets, the number of incorrectly assigned jets needs to be divided by two to obtain the number of flips. In case the number of not correctly assigned jets is odd, the number of flips needs to be rounded up as there are no 'half flips'. Thus, considering the case in which 5 jets cannot be assigned to the correct partons, there are $\frac{5}{2} = 2.5 \rightarrow 3$ flips.

With this in mind, the hypotheses can now be assigned to the different flip categories: If all jets of the analyzed hypothesis can be assigned to their corresponding truth partons, the hypothesis is assigned to the *0-flip category*. If two jets cannot been assigned to their corresponding truth partons, the hypothesis is assigned to the *1-flip category* as one flip would be needed to produce the correct configuration. The case in which two flips are needed to 'restore' the correct configuration, the hypothesis is assigned to the *2-flip category* and so on.

If at least two jets fit best with the same truth parton, the hypothesis is assigned to the *merged category* (also referred to as *category 7*). This category is orthogonal to the flip categories.

Figure 7.7 depicts the number of background hypotheses which are assigned to the different

flip categories and the merged category.
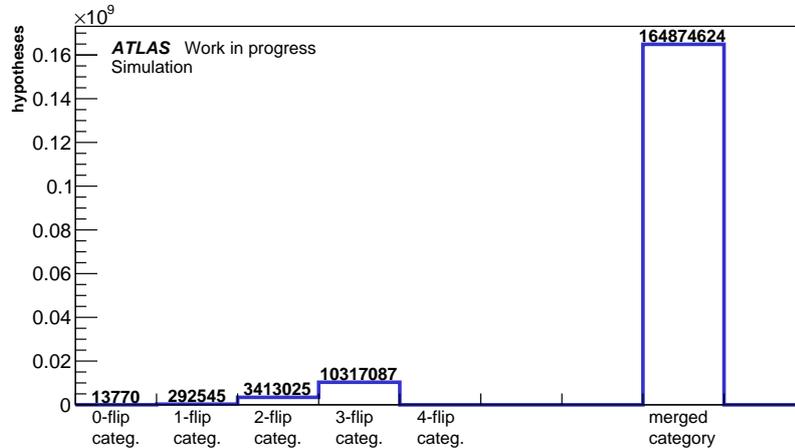
Figure 7.7: Visualization of the number of hypotheses assigned to the different flip categories and the merged category defined in the text.

On the x-axis of this plot, the different flip categories and the merged category are shown. On the y-axis the number of hypotheses assigned to those is plotted.
The first eye-catching aspect is that most hypotheses are assigned to the merged category. This proves once again that merged partons/jets are a challenge faced in reconstruction.
Another fact worth emphasizing is, that in the worst case up to 3 flips are needed to assign the selected jets to the corresponding truth partons and thus a well-reconstructed $t\bar{t}$ system. This is explainable considering the fact that the $t\bar{t}$ system has 6 decay products. Thus, a maximum number of 6 jets cannot be assigned to the correct truth parton which implies a maximum number of 3 flips.
Furthermore, the numbers related to the different categories indicate that it is quite likely that multiple flips are required to obtain the correct configuration. Thus, only considering the plain numbers, the main challenge for reconstruction seems to be flipped scenarios and merged cases. The signal-like background hypotheses (0-flip category) have in comparison a negligible contribution.

In order to gain insight into the specific physical scenarios implied in these flips, further detailed (sub-)categories are defined based on the 0-flip and 1-flip categories as for these the physical causes can be well identified:

- *category 1*: signal-like background, i.e. all jets are assigned to the corresponding truth partons (= 0-flip category)

- *category 2*: the light jets of the same top quark are switched among each other

- *category 3*: the light jets of different top quarks are switched among each other, thus a light jet of the top quark is assigned to a light parton of the anti-top quark and

vice versa

- *category 4*: one light jet and the b jet of the same top quark are switched

- *category 5*: one light jet is switched with the b jet of the other top quark

- *category 6*: the b jet and the $\bar{\text{b}}$ jet are switched

The distribution of hypotheses assigned to these (sub-)categories is presented in figure 7.8.
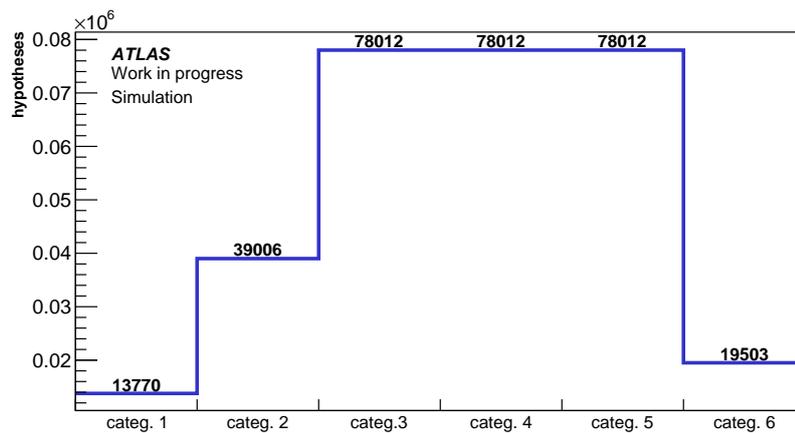
Figure 7.8: Visualization of the number of hypotheses assigned to the different subcategories of the 0-flip and 1-flip categories named categories 1-6 as defined in the text.

The categories 3, 4 and 5 (switching between the light and the b jets of the same/different top quark and switching among the light jets of one top quark) have with 78012 entries the largest number of hypotheses. In comparison, the number of 39006 assigned hypotheses for category 2 (switched light jets of the same top quark), is exactly half of the number assigned to categories 3, 4 and 5. The number of hypotheses assigned to category 6 (switched b-$\bar{\text{b}}$), is exactly a quarter of the number of categories 3, 4 and 5. The lowest number of hypotheses is assigned to category 1 (signal-like background).

This observation can be explained by using combinatorics and by considering the fact that the $t\bar{t}$ system consists of 2 top quarks: For category 2, there are 2 light-quarks involved which both have one possibility to be switched. Thus, for the 2 top quarks there are overall $2 \cdot 1 \cdot 2 = 4$ possibilities. For category 3, there are 2 light quarks for each top quark which have 2 possibilities for a switch, resulting in $2 \cdot 2 \cdot 2 = 8$ possibilities. For categories 4 and 5 two perspectives have to be considered: From the point of view of the b quark there are 2 light quarks per top quark with which it can switch (= 2 exchange possibilities per top quark). In addition, also the perspective of the light partons has to be considered: there are 2 light quarks which can both switch with the b quark (1 possibility each). All in all, there are thus $(1 \cdot 2 + 2 \cdot 1) \cdot 2 = 8$ possibilities. Last but not least, for category 6 there

are 2 b quarks which can only be switched among each other, thus resulting in $1 \cdot 1 \cdot 2 = 2$ possibilities.

From these calculations, it can be deduced that category 6 is expected to have the lowest number of assigned hypotheses compared to categories 2 to 5. Category 2 should have twice as many assigned hypotheses as category 6 and half as many as categories 3 to 5. This matches exactly the observation found in figure 7.8. As in category 1 no switching happens, no possibilities can be calculated and the number is identical to the one of the 0-flip category in figure 7.7.

The combinatorics can also be visualized in form of a matrix which is shown in figure 7.9. Here, in the columns the 6 reco jets of the reconstructed $t\bar{t}$ system are shown. In the rows, the 6 truth partons originating from the $t\bar{t}$ system are given. The cells are then colored according to the definition of the 6 (sub-)categories. Counting the same-colored squares then yields the identical numbers as the calculated ones.

jets → partons ↓	W1t	W2t	W1t̄	W2t̄	b̄	b
b						
b̄						
W2t̄						
W1t̄						
W2t						
W1t						

Figure 7.9: Visualization of the number of occurrence of the sub-categories based on the 1-flip category: category 2 in yellow (4x), category 3 in red (8x), category 4 in orange (8x), category 5 in blue (8x) and category 6 in gray (2x).

7.2.2   Observable distributions for different background categories

Having defined different background categories for specific physical scenarios, the extent of similarity between the distributions belonging to those categories and the signal distribution is now investigated. For this reason, some observables with a clear shape difference between signal and background have been selected examplarily. Some more observables can be found in appendix A.

At first, the mass of the reconstructed top quark m_t is investigated. In figure 7.10, the resulting normalized distributions are depicted for the different flip categories (dotted lines) and the (sub-)categories 1 to 6 (solid lines) in different colors. As a reference, the signal distributions (green) and the full-background distribution (dark blue) have been included.
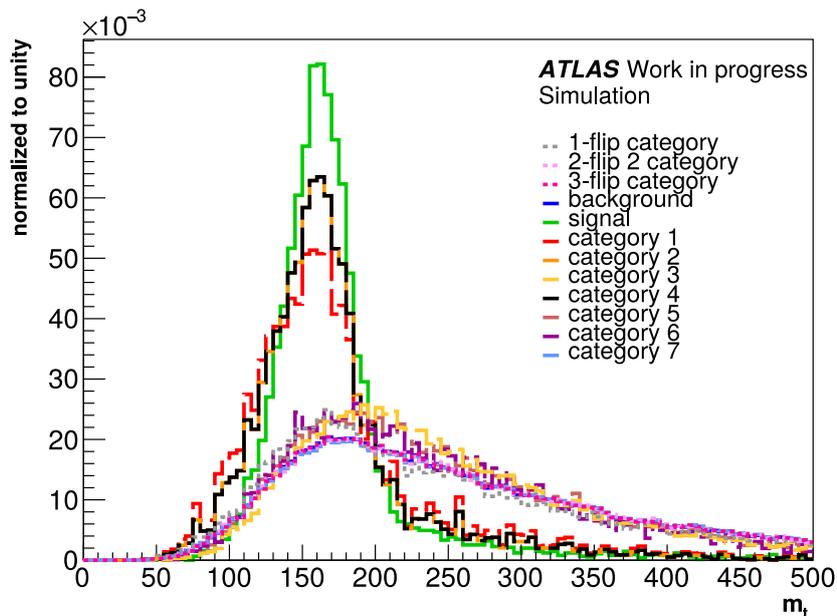
Figure 7.10: Normalized distribution of the top-quark mass m_t related to the different background categories defined in section 7.2.

It can be deduced that the most signal-like categories are category 1 (red), category 2 (orange) and category 4 (black).
For category 1 (signal-like background) this is expected. For category 2 and 4 (switching among the light jets within one top quark/switching among the light jets and the b jet of the same top quark), a signal-like distribution is expected as well. If two decay products of the same top quark are switched, the mass remains unchanged.
Furthermore, it can be observed that the distributions of the remaining categories namely category 3 (yellow), category 5 (red-brown), category 6 (purple) and category 7 (light blue) are clearly more background-like, i.e. their shapes resemble the one of the overall background (dark blue). Interestingly, also the 1-flip (grey), 2-flip (light pink) and 3-flip (dark pink) categories are significantly more background-like. Thus, the combination of all these separate scenarios should be well-distinguishable from the signal by the MVA method.

Another (non-kinematic) variable considered is the is-btagged variable for the W2t jet. In figure 7.11 the normalized results are displayed for the different flip categories (dotted lines) and the (sub-)categories 1 to 6 (solid lines) in different colors. As a reference, the signal

distributions (green) and the full-background distribution (dark blue) have been included.
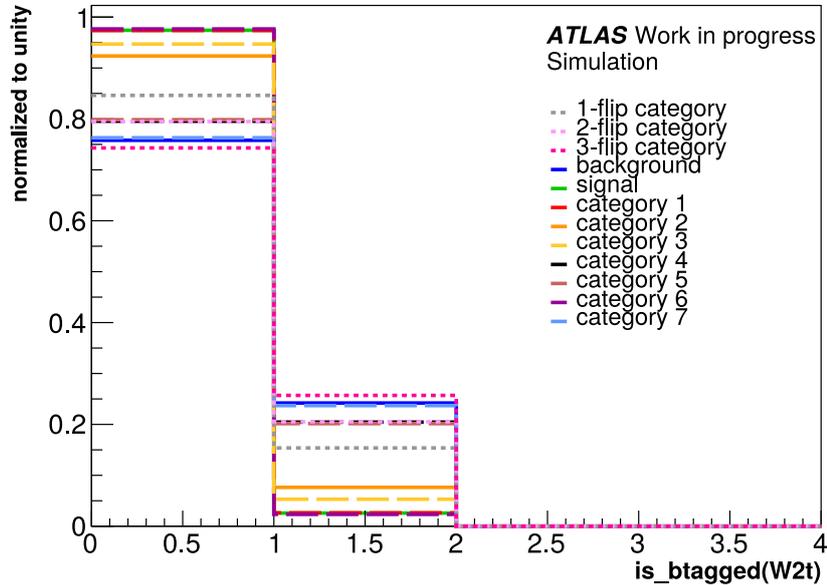
Figure 7.11: Normalized distribution of the is-btagged(W2t) observable related to the different background categories defined in section 7.2.

Here, the most signal-like shapes are those of category 1 (red), category 2 (orange), category 3 (yellow) and category 6 (purple). For category 1 (signal-like background) this behavior is expected. For categories 2 and 3 (switching between light jets) the observation can be explained by the fact, that even if light jets are switched, there is still a light jet on the W2t reconstruction position. Thus, the jet is non-btagged as in the signal scenario. The small share ($< 1\%$) of b-tagged jets on the W2t position is due to misidentification (subsection 3.2.5). For hypotheses of category 6 (switched b and $\bar{\mathrm{b}}$), the W2t jet is not involved and thus, the situation for this jet is signal-like.

The most background-like shapes belong to the 3-flip category (pink) and category 7 (light blue). The shapes of category 4 and category 5 as well as the 1-flip and 2-flip categories are very background-like as well. In between signal and background lie the 2-flip and 1-flip category as well as category 4 (black) and category 5 (red-brown). Category 4 and category 5 both involve b-light switching and thus more b-tagged jets can be found on the W2t position compared to the signal. As the switching can in principle involve all jets considered in the hypotheses, the W2t jet is not necessarily involved. In consequence, the share of non-b-tagged jets is still quite high.

Finally, an angular variable is considered: the rapidity of the W boson originating from the anti-top quark. Figure 7.12 displays the normalized distributions for the different background categories using the same color assignments and linestyles as before. In addition

also the signal (green) and the full background (dark blue) are included.
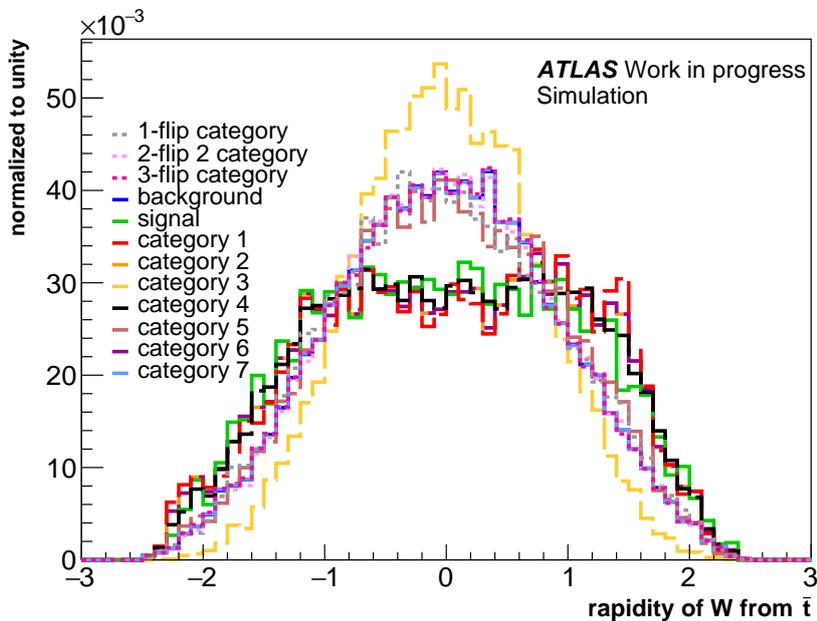
Figure 7.12: Normalized distribution of the rapidity of the W boson originating from the anti-top quark related to the different background categories defined in section 7.2.

The shapes closest to the signal are those of category 1 (red), category 2 (orange), category 4 (black) and category 6 (purple).

For category 1 (the signal-like background), this is again the expected behavior. For category 6 (switched b-$\bar{\text{b}}$) this is also expected as b-$\bar{\text{b}}$ switching does not have an impact on the W boson properties. For category 2 (switching between the light jets of the same top quark) the correct decay products have been chosen. Thus, switching implies no change in the properties of the W boson. Furthermore, also the other W boson can be addressed by the flipping and thus the analyzed W boson - in particular its rapidity distribution - is unchanged. For category 4 (switching between the b jet and the light jets of the same top quark) the signal-like shape is less obvious. It is assumed that the b jets and the light jets involved in the switching are situated close together and thus the switching hardly affects the angular position of the jet used for the reconstruction of the W boson.

The shapes of the remaining categories are closer to the background distribution. Interesting here is that the shape of category 3 (yellow) has an even sharper peak centered around 0 than the full background. This means that the hypotheses belonging to category 3 describe even more centrally produced W bosons.

In summary, it has been found that the shapes of the 1-flip, 2-flip and 3-flip categories are in all cases well-distinguishable from the signal shape.

Regarding the (sub-)categories of the 1-flip and 0-flip categories, i.e. categories 1-6, it can be retained that especially categories 1 and 2 are very signal-like in all cases. For category 1 this is obvious as it represents hypotheses of signal-like background thus all jets have been assigned to the corresponding partons. Category 2 is a challenging category to separate from the signal, as the light jets can hardly be distinguished between each other besides from using the relatively unprecise jet charge. Consequently an option could be to accept this category as signal. This is just mentioned as an outlook here, but is not pursued further in this thesis.

In addition, different categories have been found to be most signal-like for each of the different observables: category 4 for m_t, categories 3 and 6 for is-btagged(W2t) and categories 4 and 6 for the rapidity of the W boson originating from the anti-top quark. Thus, to assure that the MVA method is able to well-distinguish especially between the signal and the signal-like categories, all of the categories 1-6 are expected to be extremely important in the training process.

7.3   Training of the Neural Network

Having selected a method to produce labeled input data and having developed a good understanding of the physical scenarios contained in the background category, it is now possible to train the MVA method which is the focus of this section.

Basically, there are two 'classic' MVA methods which can be used for the reconstruction: the Neural Network (NN) and the Boosted Decision Tree (BDT). However, the NN should be able to cope better with the imbalanced numbers of signal and background hypotheses as special weights addressing these can be used [74]. Therefore, a Keras NN [109] with Tensorflow back-end [110] has been chosen as the MVA method in this thesis. It is implemented in [111].

To train the Neural Network, a decision regarding the input variables and the NN architecture needs to be taken. Due to memory limitations also the dataset used for training needs to be selected. The chosen dataset, input variables and NN architecture are presented in subsection 7.3.1. Having taken these decisions, the training can be conducted. Its performance is subsequently investigated in subsection 7.3.2.

7.3.1   Input variables, training dataset and Neural-Network architecture

To prepare the training process, the dataset, the input variables and the architecture of the Neural Network have to be determined.

Training dataset

As described in section 7.1, the quarklevel method combined with the event selection 'with improvement cuts' is chosen to produce the labeled input data for the training process.
Unfortunately, using all hypotheses labeled by this method exceeded the memory capacity of the work-station used which is 32 GB. As this is the work-station with the highest memory capacity at the chair, a possible solution would have been to use the resources of the computing center LRZ. The preparation for this is very costly in terms of time though. It was therefore not possible to conduct this within the scope of this thesis.
That is why an alternative approach has been developed: Instead of using the full input dataset for the training, only a subset is selected. As the number of signal hypotheses is already low, it seems reasonable to reduce the number of background hypotheses. The idea is then to train the NN on the whole signal and only on the background categories most similar to the signal. The final performance evaluation of the NN is subsequently conducted using the whole dataset as the evaluation is a lot less memory intensive than the training[7].

In the studies conducted in section 7.2, it was discovered that the sub-categories of the 0-flip and 1-flip categories, i.e. categories 1 to 6, are quite signal-like while the multi-flip categories as well as the merged category are very background-like. In consequence, the latter are excluded for the training. According to figure 7.7 this also reduced the number of background hypotheses considerably as desired.

Input variables for the NN training

In a next step, also the observables used as input variables for training have to be selected. Here, only observables on reco level are chosen as - in case of a good performance - the NN-based reconstruction is applied to experimental data. For these, no truth information is available.

The observables used as input variables for the training are summarized in the tables 7.2 and 7.3.

In these tables, the input variables are grouped in different categories: the first two categories are formed by the kinetic observables, the masses and the transverse momenta. Furthermore, also angular information is included in the form of the rapidities of the reconstructed objects, i.e. the rapidities of the W bosons and the top/anti-top quark[8]. The angular information provided by η for the single jets[9] as well as by ϕ for the single jets

---

[7]For the training, all input data need to be imported as they are reshuffled after each epoch. In this way, it is assured, that not always the same hypotheses are used for training and testing.

[8]In this context, a Lorentz-invariant property like the rapidity is required as the reconstructed objects are quite massive.

[9]The non-Lorentz-invariant property η can be used for the single jets as they can be approximated to be massless compared to the reconstructed objects.

and the reconstructed objects cannot be used here due to the lack of separation power between signal and background. Moreover, also the is-btagged observable as well as helicity properties are used.

In general, no criteria to differentiate between the top quark and the anti-top quark have been used. It would thus be legitimate to think of them as top 1 and top 2 and analogous for their decay products. Only to determine the down-type quark, the jet charge introduced in subsection 3.2.6 has been used.

To address the identified reconstruction challenge of a changing topology, observables targeting the angular distance between the decay products have been included. As they have been developed in this thesis, their shapes are investigated in more detail in the following.

observables	description
masses	
m_t/m_{t1}	mass of the top quark
$m_{\bar{t}}/m_{t2}$	mass of the anti-top quark
$m_{W \text{ from } t}/m_{W1}$	mass of the W boson originating from the top quark
$m_{W \text{ from } \bar{t}}/m_{W2}$	mass of the W boson originating from the anti-top quark
$m_{t\bar{t}}$	mass of the top-antitop-quark system
m_{W1t}/m_{j1} and m_{W2t}/m_{j2}	masses of the jets originating from the W boson of the top quark
$m_{W1\bar{t}}/m_{j3}$ and $m_{W2\bar{t}}/m_{j4}$	masses of the jets originating from the W boson of the anti-top quark
m_b/m_{j5b}	mass of the b jet originating from the top quark
$m_{\bar{b}}/m_{j6b}$	mass of the $\bar{b}$ jet originating from the anti-top quark
transverse momenta	
$p_T(t)/p_T(t1)$	transverse momentum of the top quark
$p_T(\bar{t})/p_T(t2)$	transverse momentum of the anti-top quark
$p_T(W \text{ from } t)/p_T(W1)$	transverse momentum of the W boson originating from the top quark
$p_T(W \text{ from } \bar{t})/p_T(W2)$	transverse momentum of the W boson originating from the anti-top quark
$p_T(W1t)/p_T(j1)$ and $p_T(W2t)/p_T(j2)$	transverse momenta of the jets originating from the W boson of the top quark
$p_T(W1\bar{t})/p_T(j3)$ and $p_T(W2\bar{t})/p_T(j4)$	transverse momenta of the jets originating from the W boson of the anti-top quark
$p_T(b)/p_T(j5b)$	transverse momentum of the b jet originating from the top quark
$p_T(\bar{b})/p_T(j6b)$	transverse momentum of the $\bar{b}$ jet originating from the anti-top quark
rapidities	
rapidity(t)/rapidity(t1)	rapidity of the top quark
rapidity($\bar{t}$)/rapidity(t2)	rapidity of the anti-top quark
rapidity(W from t)/rapidity(W1)	rapidity of the W boson originating from the top quark
rapidity(W from $\bar{t}$)/rapidity(W2)	rapidity of the W boson originating from the anti-top quark

Table 7.2: Input variables for the training of the Neural Network (part 1).

observables	description
is-btagged	
is-btagged(W1t)/is-btagged(j1)	is-btagged of the jet of the first decay product of the W boson originating from the top quark
is-btagged(W2t)/is-btagged(j2)	is-btagged of the jet of the second decay product of the W boson originating from the top quark
is-btagged(W1$\bar{\text{t}}$)/is-btagged(j3)	is-btagged of the jet of the first decay product of the W boson originating from the anti-top quark
is-btagged(W2$\bar{\text{t}}$)/is-btagged(j4)	is-btagged of the jet of the second decay product of the W boson originating from the anti-top quark
is-btagged(b)/is-btagged(j5b)	is-btagged of the b jet originating from the top quark
is-btagged($\bar{\text{b}}$)/is-btagged(j6b)	is-btagged of the $\bar{\text{b}}$ jet originating from the anti-top quark
helicity properties	
$\cos(\Theta^*)(\text{t})/\cos(\Theta^*)(\text{t1})$	helicity angle for the top quark
$\cos(\Theta^*)(\bar{\text{t}})\cos(\Theta^*)(\text{t2})$	helicity angle for the anti-top quark
p_T(down-type quark of t)/ p_T(down-type quark of t1)	transverse momentum of the down-type quark of the top quark
p_T(down-type quark of $\bar{\text{t}}$)/ p_T(down-type quark of t2)	transverse momentum of the down-type quark of the anti-top quark
sum p_T (down-type quark)	p_T(down-type quark of t) + p_T(down-type quark of $\bar{\text{t}}$)
angular distance between the decay products	
$\Delta\phi$(decay products t)/ $\Delta\phi$(decay products t1)	$\|\Delta\phi_{(\text{b,W1t})}\| + \|\Delta\phi_{(\text{b,W2t})}\| + \|\Delta\phi_{(\text{W1t,W2t})}\|$
$\Delta\phi$(decay products $\bar{\text{t}}$)/ $\Delta\phi$(decay products t2)	$\|\Delta\phi_{(\bar{\text{b}},\text{W1}\bar{\text{t}})}\| + \|\Delta\phi_{(\bar{\text{b}},\text{W2}\bar{\text{t}})}\| + \|\Delta\phi_{(\text{W1}\bar{\text{t}},\text{W2}\bar{\text{t}})}\|$
$\Delta\phi$(decay products of t and $\bar{\text{t}}$)	$\Delta\phi$(decay products $\bar{\text{t}}$) + $\Delta\phi$(decay products t)
$\Delta\eta$(decay products t)/ $\Delta\eta$(decay products t1)	$\|\Delta\eta_{(\text{b,W1t})}\| + \|\Delta\eta_{(\text{b,W2t})}\| + \|\Delta\eta_{(\text{W1t,W2t})}\|$
$\Delta\eta$(decay products $\bar{\text{t}}$)/ $\Delta\eta$(decay products t2)	$\|\Delta\eta_{(\bar{\text{b}},\text{W1}\bar{\text{t}})}\| + \|\Delta\eta_{(\bar{\text{b}},\text{W2}\bar{\text{t}})}\| + \|\Delta\eta_{(\text{W1}\bar{\text{t}},\text{W2}\bar{\text{t}})}\|$
$\Delta\eta$(decay products of t and $\bar{\text{t}}$)	$\Delta\eta$(decay products $\bar{\text{t}}$) + $\Delta\eta$(decay products t)
ΔR(decay products t)/ ΔR(decay products t1)	$\|\Delta\text{R}_{(\text{b,W1t})}\| + \|\Delta\text{R}_{(\text{b,W2t})}\| + \|\Delta\text{R}_{(\text{W1t,W2t})}\|$
ΔR(decay products $\bar{\text{t}}$)/ ΔR(decay products t2)	$\|\Delta\text{R}_{(\bar{\text{b}},\text{W1}\bar{\text{t}})}\| + \|\Delta\text{R}_{(\bar{\text{b}},\text{W2}\bar{\text{t}})}\| + \|\Delta\text{R}_{(\text{W1}\bar{\text{t}},\text{W2}\bar{\text{t}})}\|$
ΔR(decay products of t and $\bar{\text{t}}$)	ΔR(decay products $\bar{\text{t}}$) + ΔR(decay products t)

Table 7.3: Input variables for the training of the Neural Network (part 2).

In figure 7.13, the normalized distributions of the $\Delta\eta$ between the decay products of the top quark (figure 7.13a) and the anti-top quark (figure 7.13b) as well as their event-wise sum (figure 7.13c) are displayed. Their definition is given in table 7.3. The signal and background distributions are depicted in blue and red respectively. As a reference, also the truth information (yellow) is included.

(a)

(b)

(c)
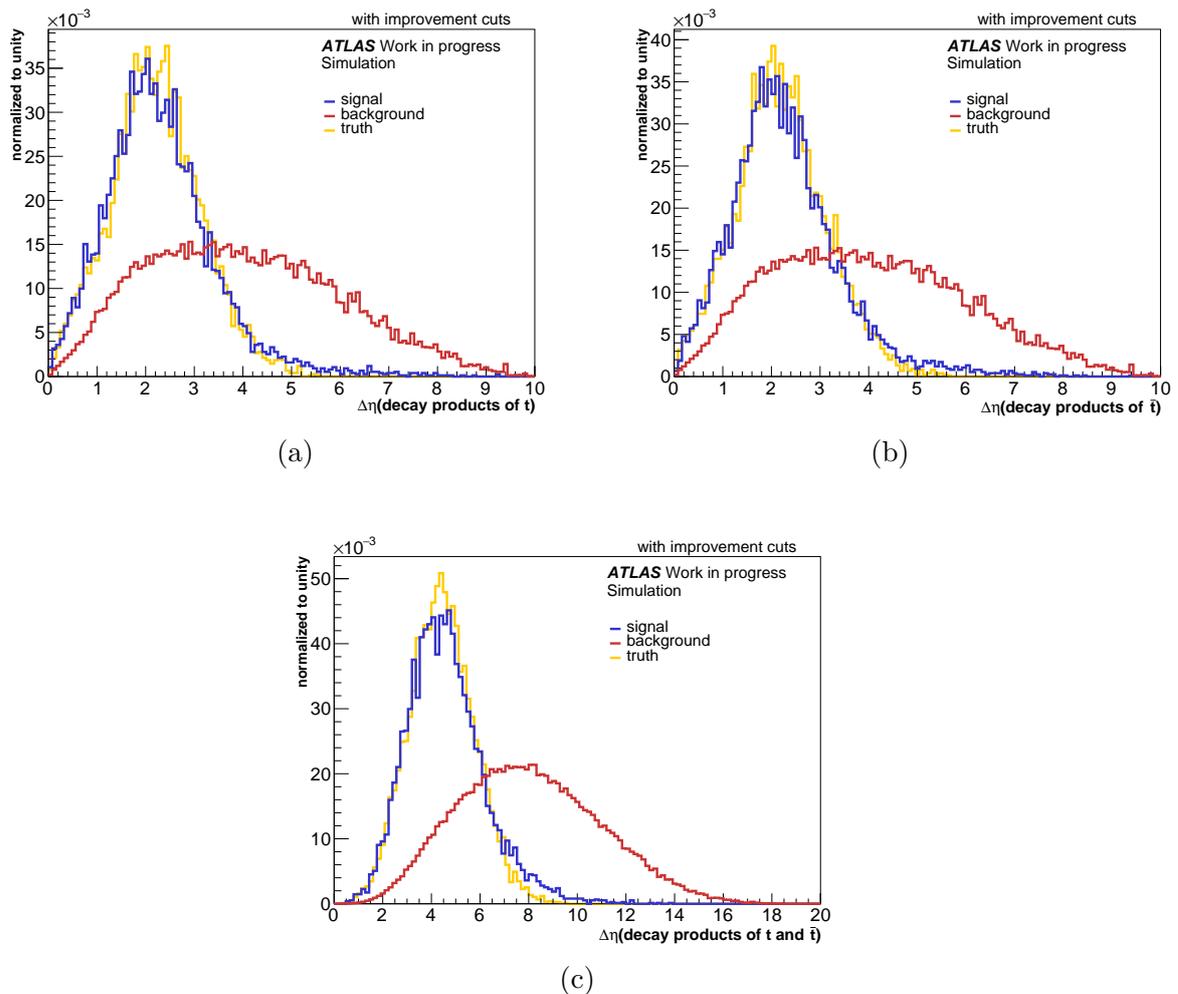
Figure 7.13: Normalized signal, background and truth distributions of the angular $\Delta\eta$ distance between the decay products of the top quark (a), the anti-top quark (b) and their event-wise sum (c).

In both, figure 7.13a and figure 7.13b, the signal distributions peak at a value of about 2. Consequently, the signal peak is situated at a value of about 5 in figure 7.13c as here the sum of $\Delta\eta$ between the decay products of the top and the anti-top quark is shown. The background clearly combines scenarios exhibiting higher distances between the decay particles.

Due to the different shapes of the signal and background distributions, a high separation power can be identified which makes this observable a suitable input variable.

Furthermore, the distribution of the signal is in line with the distribution of the truth information in all cases. This confirms the choice of the method as the physical information is reproduced truthfully.

In figure 7.14, the normalized signal, background and truth distributions of the $\Delta\phi$ between the decay products of the top quark (figure 7.14a) and the anti-top quark (figure 7.14b) as well as their event-wise sum (figure 7.14c) are depicted in the same colors as in figure 7.13. Their definition is given in table 7.3.

(a)   (b)

(c)
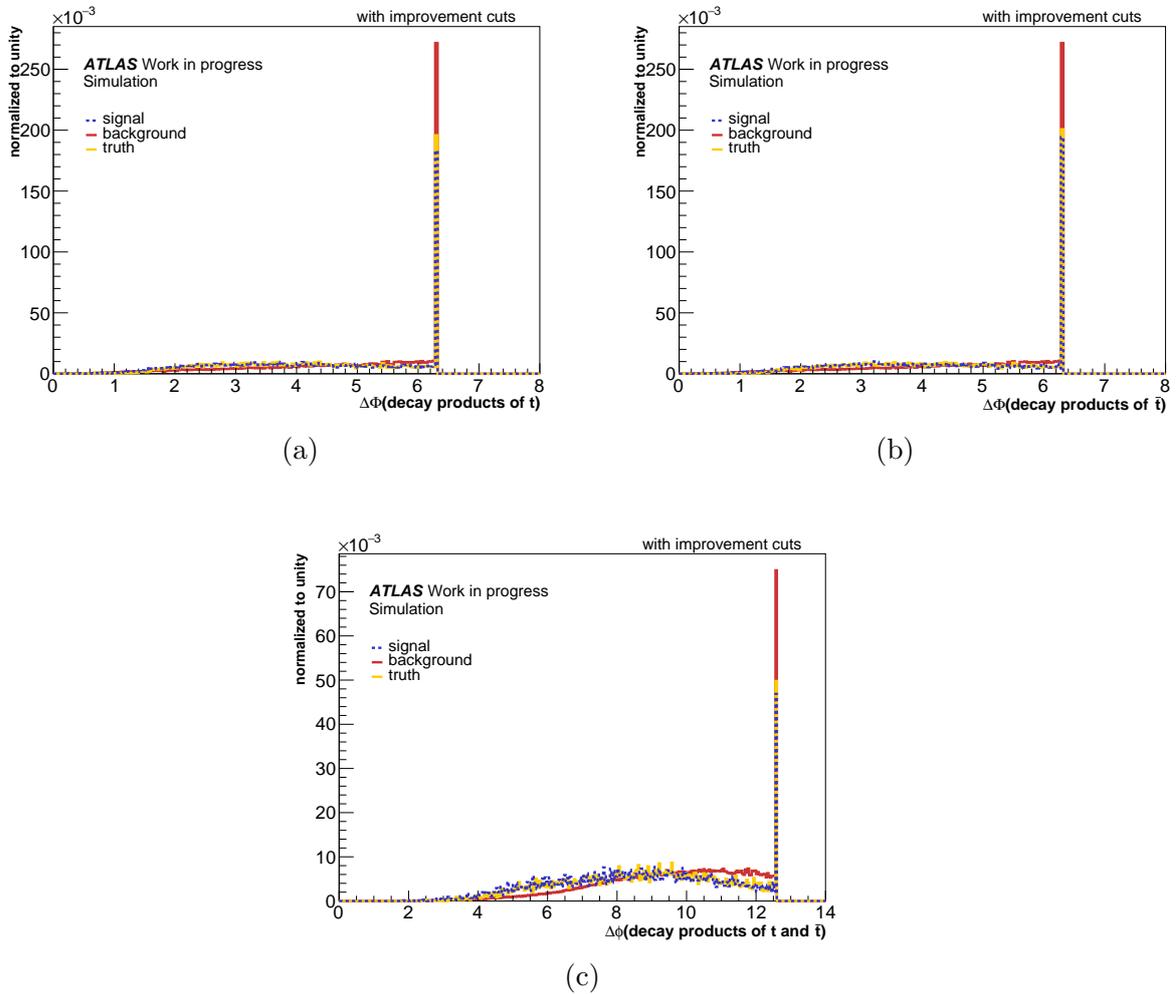
Figure 7.14: Normalized signal, background and truth distributions of the angular $\Delta\phi$ distance between the decay products of the top quark (a), the anti-top quark (b) and their event-wise sum (c).

Looking at the resulting distributions, a quite exotic shape exhibiting a high and very sharp peak at a value of exactly 2π (≈ 6.2) in figures 7.14a and 7.14b can be observed. In figure 7.14c, the peak is situated at exactly 4π (≈ 12.3) as here the sum of the $\Delta\phi$ between the decay products of the top and anti-top quark is displayed. Towards lower values of $\Delta\phi$, a long and narrow tail ending at values greater 0 can be found in all subfigures.

Apart from that, there is a perfect fit of the truth information and the signal. This confirms again the choice of the method.

Moreover, also a small signal-background separation is observed which can hardly be spotted due to the high peak in figures 7.14a and 7.14b. The separation becomes more obvious when looking at figure 7.14c as here the sum is displayed. This implies that not only the distributions, but also their separation powers are added.

To explain the observed exotic shapes, further investigation is necessary. For this reason, the truth distribution of the transverse momentum of the anti-top quark[10] is plotted for $\Delta\phi$(decay products $\bar{\mathrm{t}}$) $> 2\pi \approx 6.2$ (peak) and $\leq 2\pi \approx 6.2$ (tail). The normalized results are depicted in figure 7.15.
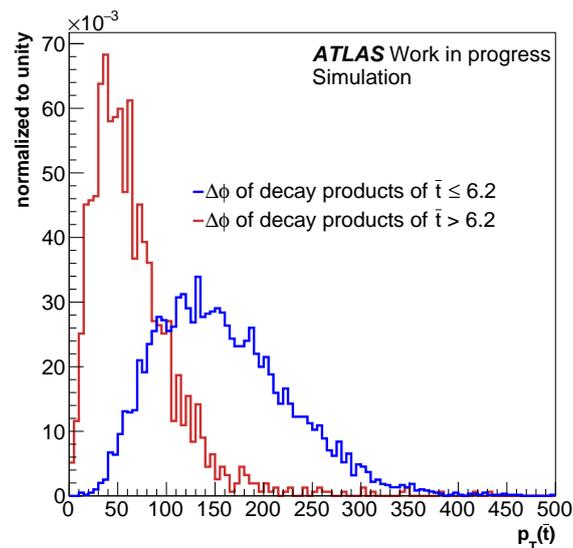
Figure 7.15: Normalized truth distributions of the transverse momentum of the anti-top quark $\mathrm{p_T}(\bar{\mathrm{t}})$ separately for $\Delta\phi$(decay products $\bar{\mathrm{t}}$)$> 2\pi$ and $\leq 2\pi$.

Here, a clear separation between the $\mathrm{p_T}(\bar{\mathrm{t}})$ distributions of the 2 $\Delta\phi$ cases can be observed. From this, it can be deduced that the peak in figure 7.14 results from (anti-)top quarks with low $\mathrm{p_T}$ and that the tail is a result of (anti-)top quarks with high $\mathrm{p_T}$.

This is plausible as in general for a (anti-)top quark with low $\mathrm{p_T}$ the decay products are emitted in larger angles and thus are spread further. Looking at the positions of the decay

---

[10]The distributions are the same for the top quark and the anti-top quark. Therefore, the discussion is examplarily done based on the $\mathrm{p_T}(\bar{\mathrm{t}})$.

products in the η-ϕ plane, the three jets j1, j2, j3 form a triangle. Figure 7.16 depicts examplarily the configuration for a vanishing transverse momentum of the top/anti-top quark.
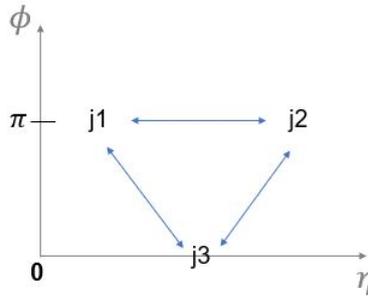
Figure 7.16: Sketch of the triangle formed by the three decay products of the (anti-)top quark, referred to as j1, j2 and j3, in the ϕ-η plane for the case of a vanishing transverse momentum of the top/anti-top quark.

Considering the projection onto the ϕ-axis, $\Delta\phi$(j1, j3) $= \pi$ can be determined as ϕ(j1) $= \pi$ and ϕ(j3) $= 0$. Analogously, $\Delta\phi$(j1, j2) $= 0$ and $\Delta\phi$(j2, j3) $= \pi$. Thus, overall $\Delta\phi$(decay products $t/\bar{t}$) $= \pi + 0 + \pi = 2\pi$. This is exactly the position of the sharp peak in figures 7.14a and 7.14b.
Furthermore, it is also the maximum distance possible between the decay products of the top/anti-top quark as ϕ ranges from $-\pi$ to π and only the absolute values are considered in the calculation.
In case the (anti-)top quark has a higher p_T, the decay products are emitted in smaller angles and are in consequence closer together. Thus, a smaller triangle can be found which results in lower values of $\Delta\phi$(between decay products $t/\bar{t}$). As there are multiple configurations leading to a $\Delta\phi$(decay products of $t/\bar{t}$)$< 2\pi$, a continuous and narrow tail can be found.
Moreover, this visualization also explains why the tail ends at values greater 0. This would corresponds to a configuration in which the decay products are 'glued' to each other which cannot be achieved with a $p_T(\bar{t})/p_T(t)$ of maximum 400 GeV using the setup of this thesis.

In figure 7.17, the normalized distributions of the ΔR between the decay products of the top quark (figure 7.17a) and the anti-top quark (figure 7.17b) as well as their event-wise sum (figure 7.17c) are displayed in the same color-scheme as in figure 7.13. Their definition is given in table 7.3.

(a)

(b)

(c)
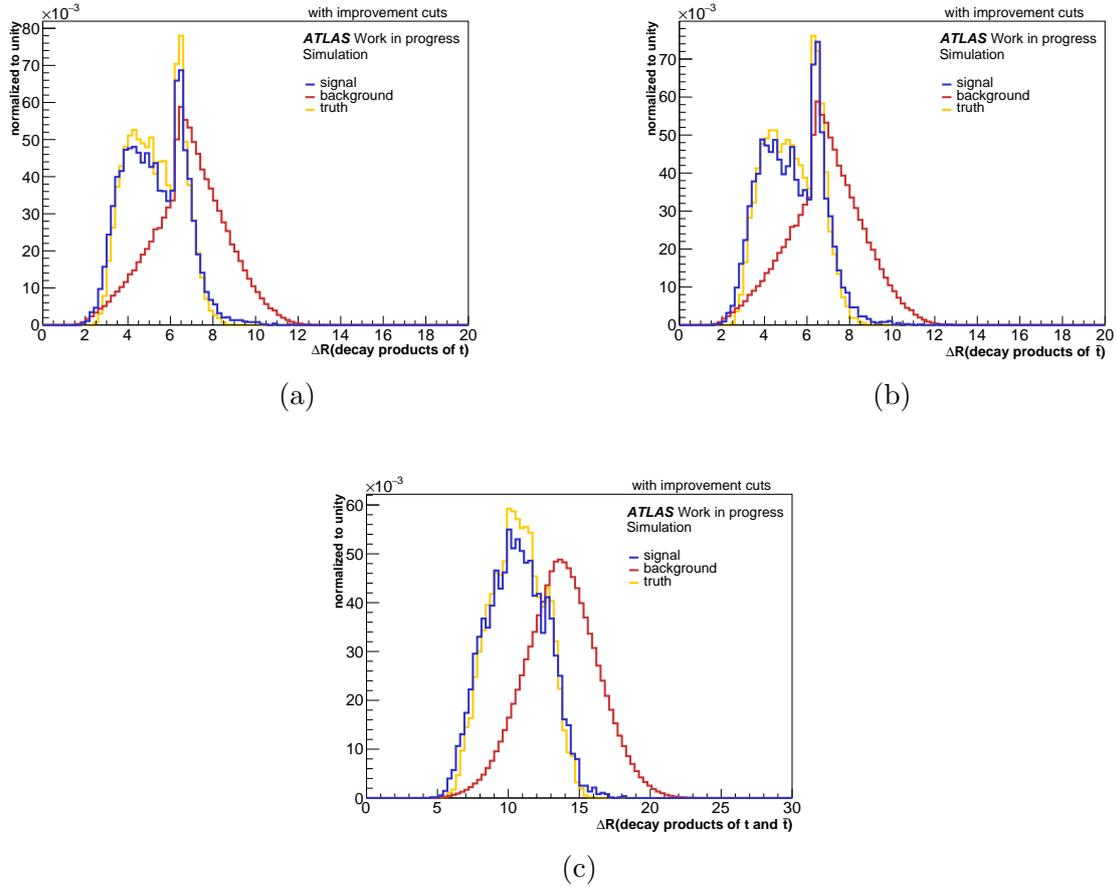
Figure 7.17: Normalized signal, background and truth distributions of the angular ΔR distance between the decay products of the top quark (a), the anti-top quark (b) and their event-wise sum (c).

In figures 7.17a and 7.17b again some exotic shapes are found. They can be explained by considering that ΔR is defined as $\sqrt{\Delta\phi^2 + \Delta\eta^2}$. Thus, figure 7.17a (figure 7.17b) is roughly the sum of figure 7.13a (figure 7.13b) and figure 7.14a (figure 7.14b).

The exotic shape vanishes in figure 7.17c. This is a result of an event-wise addition of the ΔR values between the decay products of the top quark to those of the anti-top quark. Obviously in one event, there is always one top quark with a large ΔR value and an anti-top quark with a low ΔR value or vice versa and thus the sum exhibits only one peak. This corresponds to one of the two preferred topologies found in chapter 6: if the one of the top quarks recoils against the other top quark and the Z boson, the 'single' top quark will have a higher transverse momentum than the top quark on the Z boson side. As a higher (lower) transverse momentum leads to decay products which are situated closer together (farther apart), the top quark with the higher (lower) p_T value has also a lower (higher) ΔR value for its decay products.

Overall, the investigations conducted indicate that all observables targeting the angular difference between the decay products exhibit a significant separation power and are thus used for training.

In a next step, also the normalized signal and background distributions are depicted ranked according to their separation power in figures 7.18 to 7.25.

In this context, the separation power is calculated as follows: first, for all signal bins, the bin content is divided by the number of bins and referred to as s. The same is done for the background and referred to as b. Then, the separation power is calculated according to $\sum_{i}^{N_{bins}} 0.5 \frac{(s_i - b_i)^2}{(s_i + b_i)}$ using all bins with $s_i + b_i > 0$ [98]. The error of the separation power is calculated according to linear error propagation theory.

(a) separation power: 0.4275
(0.427 ± 0.006)

(b) separation power: 0.4271
(0.427 ± 0.006)

(c) separation power: 0.3982
(0.39815 ± 0.00020)

(d) separation power: 0.3374
(0.33741 ± 0.00014)
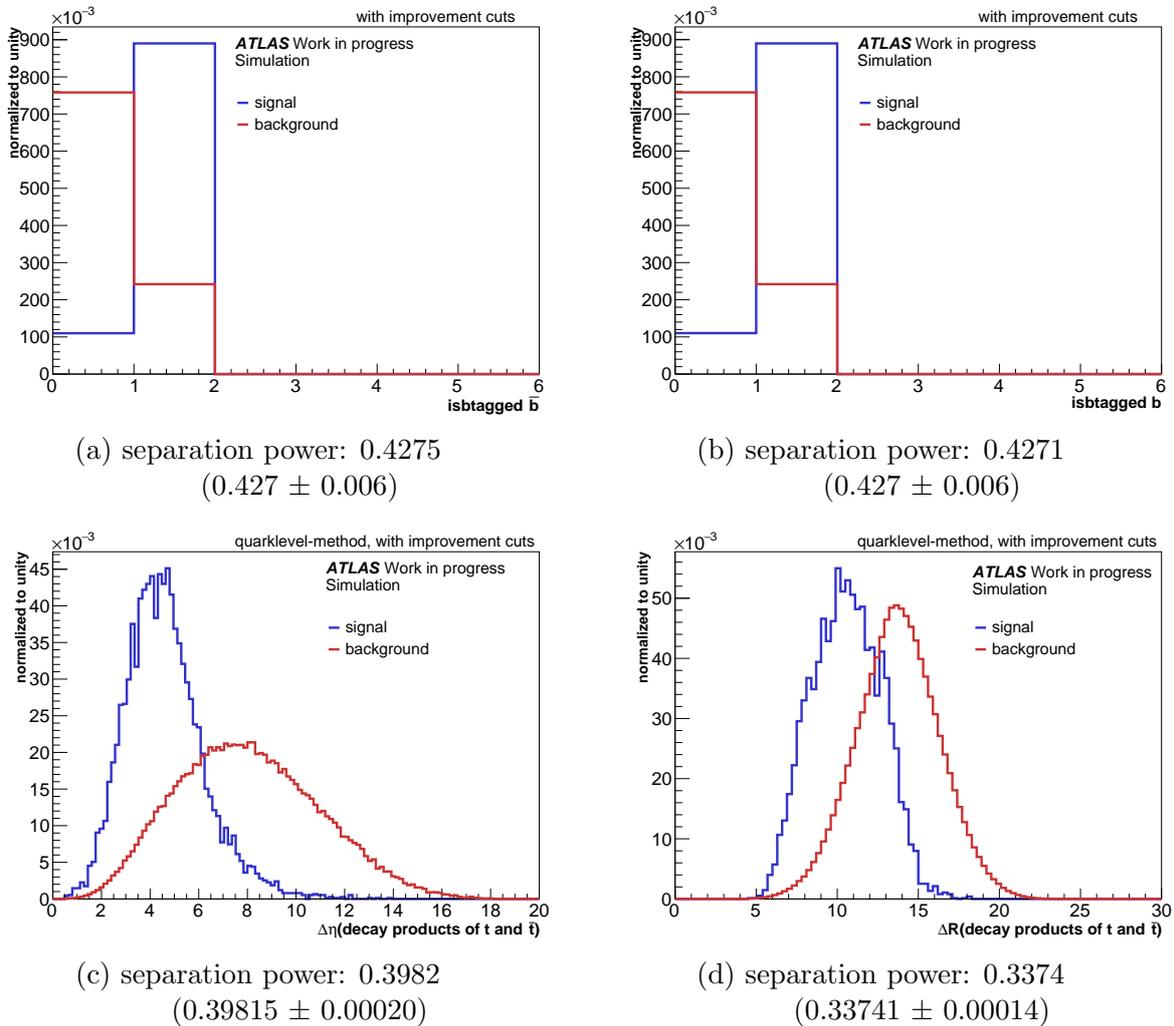
Figure 7.18: Normalized signal and background distributions of the most separating input variables (listed in table 7.2 and 7.3) for the NN training ordered by decreasing separation power. (part 1)
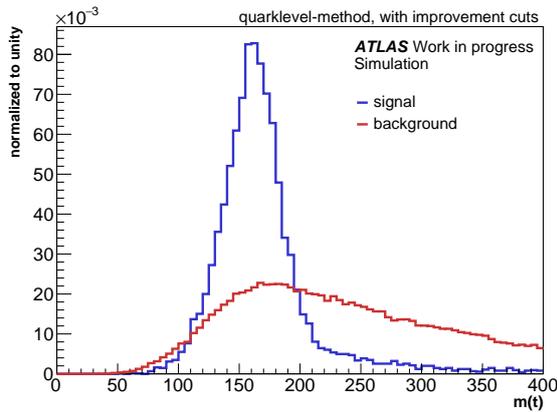
(a) separation power: 0.3076
(0.308 ± 0.005)
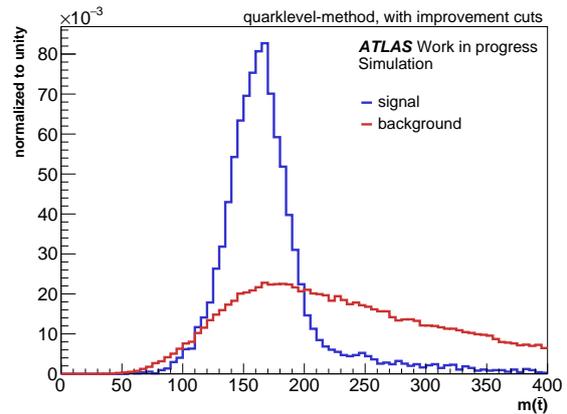
(b) separation power: 0.3071
(0.307 ± 0.005)
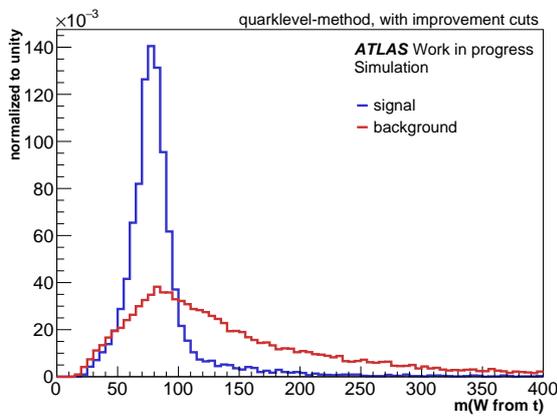
(c) separation power: 0.3059
(0.306 ± 0.005)
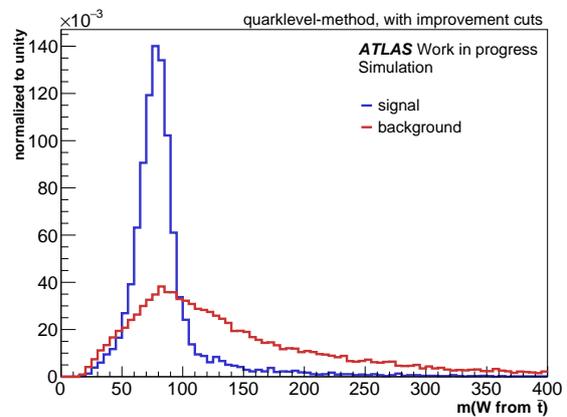
(d) separation power: 0.3029
(0.303 ± 0.005)
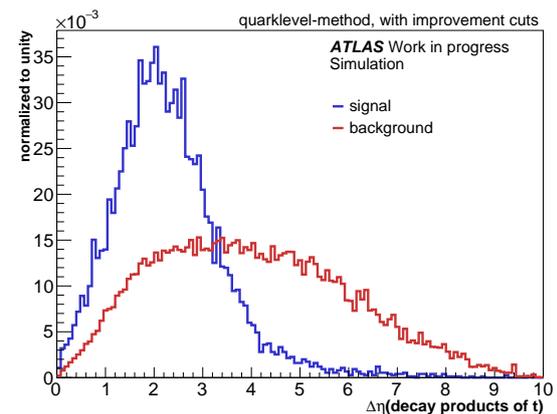
(e) separation power: 0.2610
(0.26096 ± 0.00017)
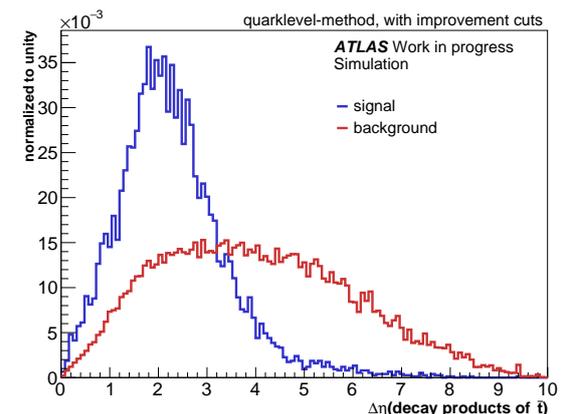
(f) separation power: 0.2595
(0.25948 ± 0.00017)

Figure 7.19: Normalized signal and background distributions of the most separating input variables (listed in table 7.2 and 7.3) for the NN training ordered by decreasing separation power. (part 2)
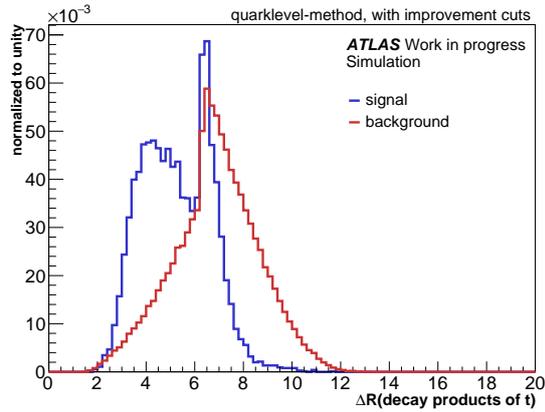
(a) separation power: 0.2366
(0.23663 ± 0.00008)
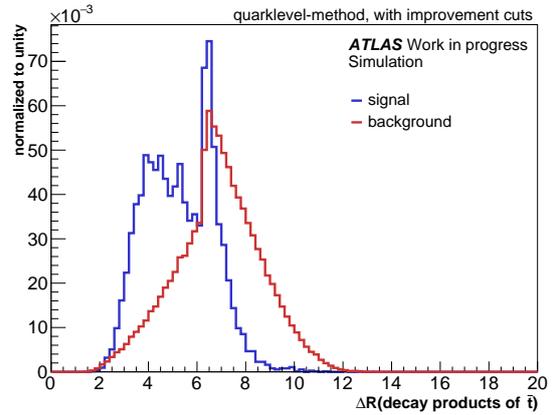
(b) separation power: 0.2311
(0.23115 ± 0.00008)
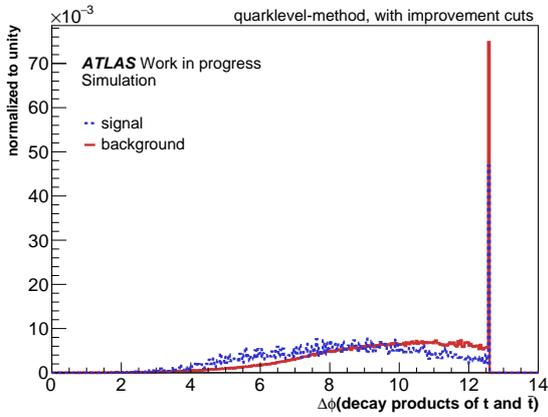
(c) separation power: 0.1295
(0.129473 ± 0.000014)
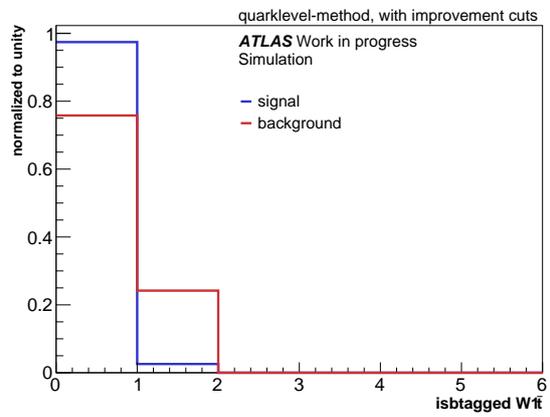
(d) separation power: 0.1009
(0.1009 ± 0.0025)
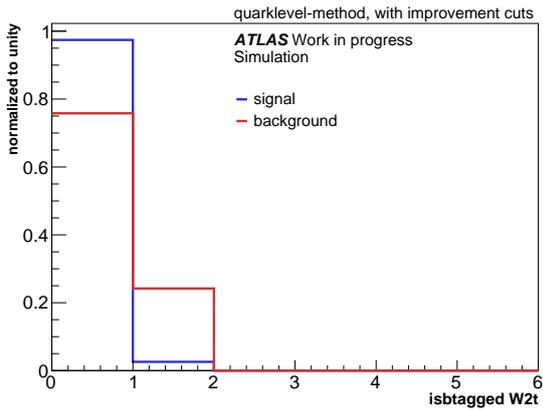
(e) separation power: 0.1007
(0.1007 ± 0.0025)
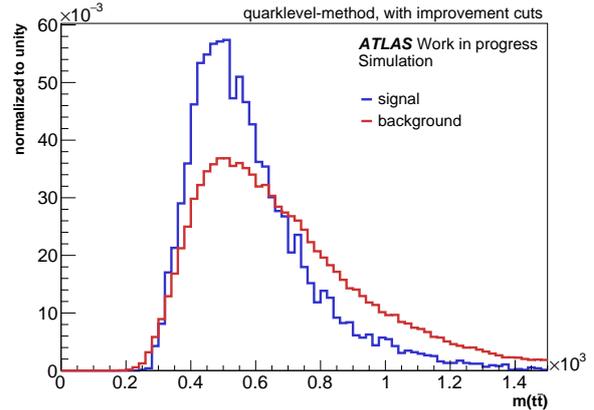
(f) separation power: 0.0997
(0.0997 ± 0.0025)

Figure 7.20: Normalized signal and background distributions of the most separating input variables (listed in table 7.2 and 7.3) for the NN training ordered by decreasing separation power. (part 3)
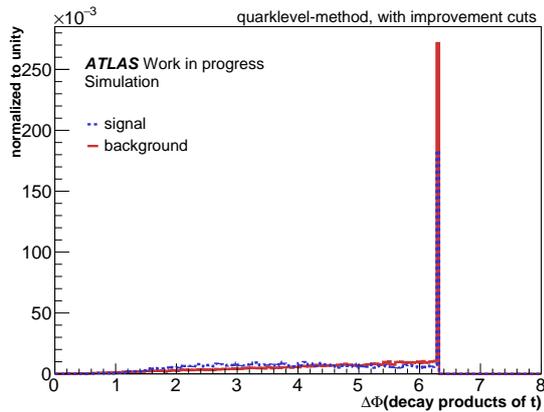
(a) separation power: 0.0906
(0.090648 ± 0.000012)
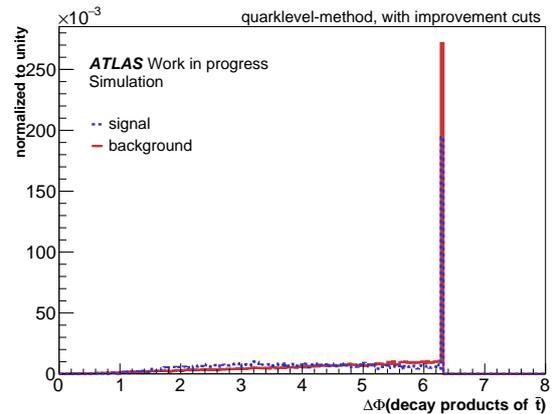
(b) separation power: 0.0893
(0.089297 ± 0.000012)
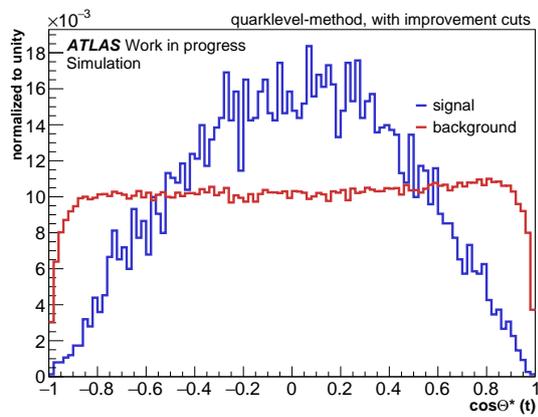
(c) separation power: 0.0754
(0.07544 ± 0.00005)
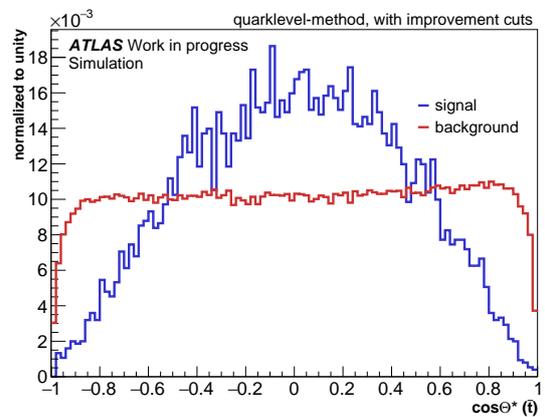
(d) separation power: 0.0753
(0.07530 ± 0.00005)
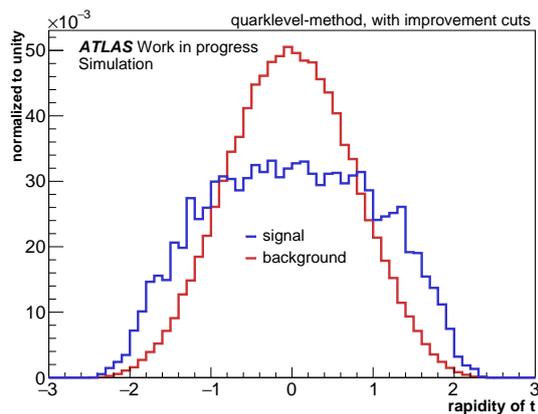
(e) separation power: 0.0580
(0.05804 ± 0.00026)
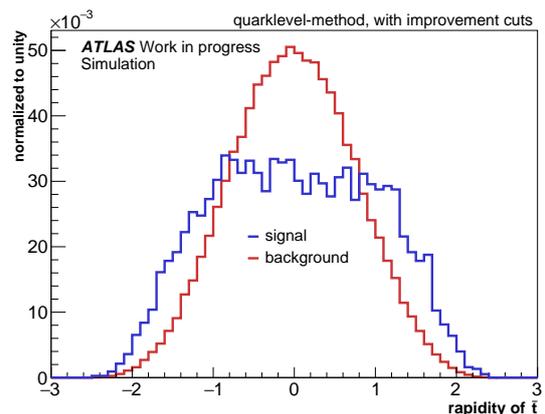
(f) separation power: 0.0572
(0.05723 ± 0.00026)

Figure 7.21: Normalized signal and background distributions of the most separating input variables (listed in table 7.2 and 7.3) for the NN training ordered by decreasing separation power. (part 4)
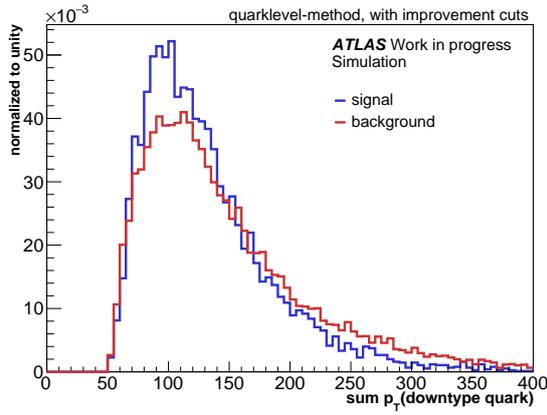
(a) separation power: 0.0539
(0.0539 ± 0.0010)
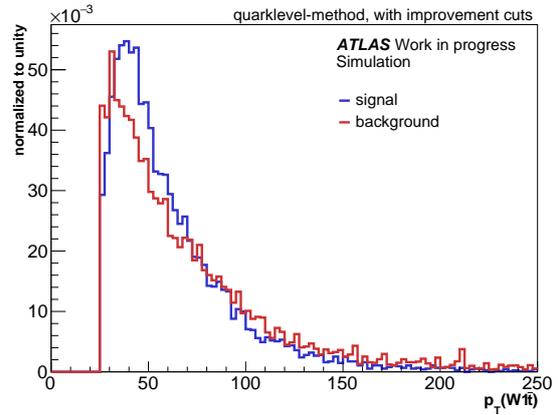
(b) separation power: 0.0511
(0.0511 ± 0.0010)
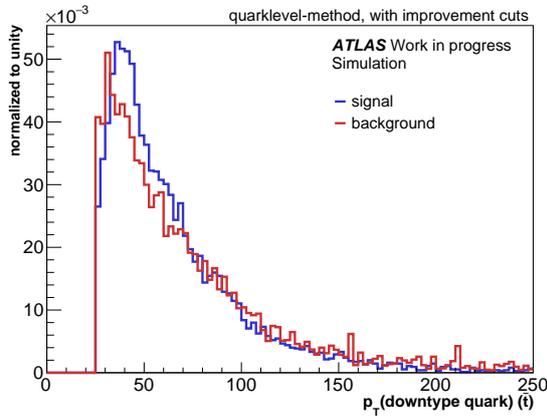
(c) separation power: 0.0496
(0.0496 ± 0.0010)
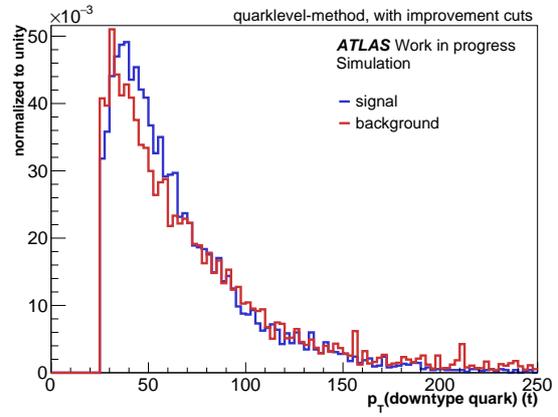
(d) separation power: 0.0460
(0.0460 ± 0.0009)
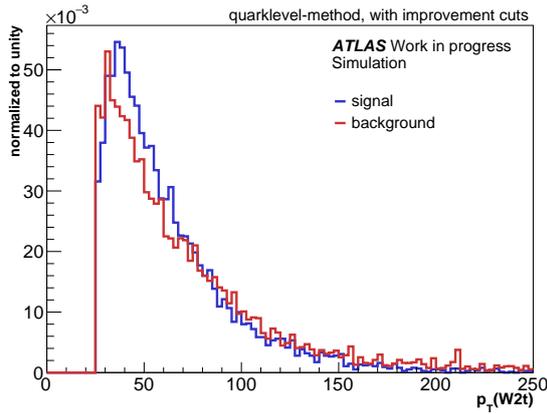
(e) separation power: 0.0455
(0.0455 ± 0.0009)
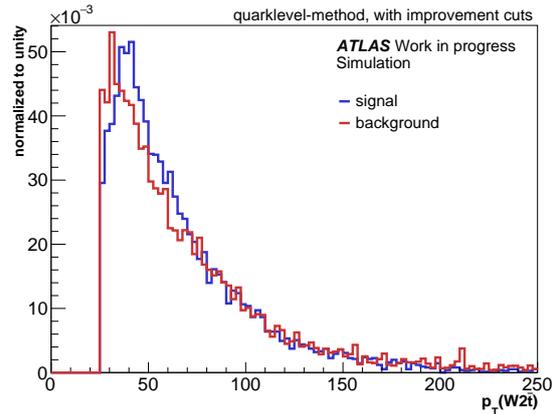
(f) separation power: 0.0447
(0.0447 ± 0.0009)

Figure 7.22: Normalized signal and background distributions of the most separating input variables (listed in table 7.2 and 7.3) for the NN training ordered by decreasing separation power. (part 5)
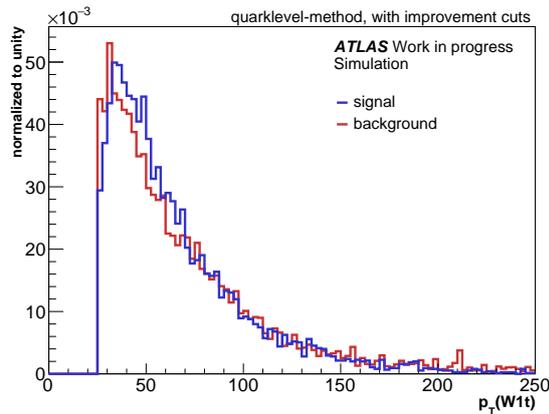
(a) separation power: 0.0436
(0.0436 ± 0.0009)
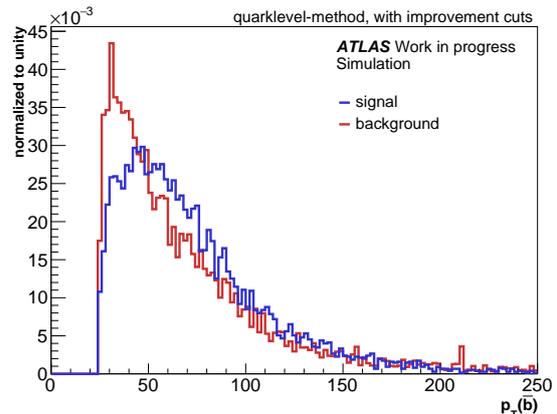
(b) separation power: 0.0415
(0.0415 ± 0.0019)
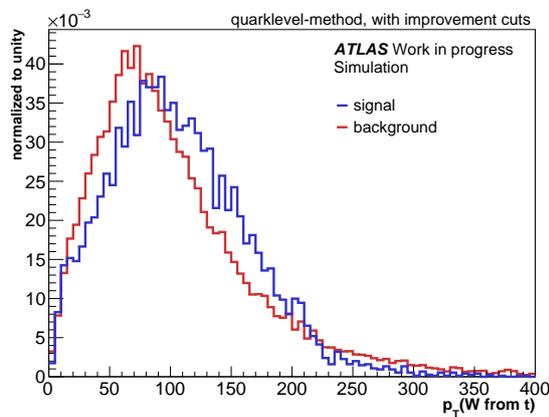
(c) separation power: 0.0378
(0.0378 ± 0.0019)
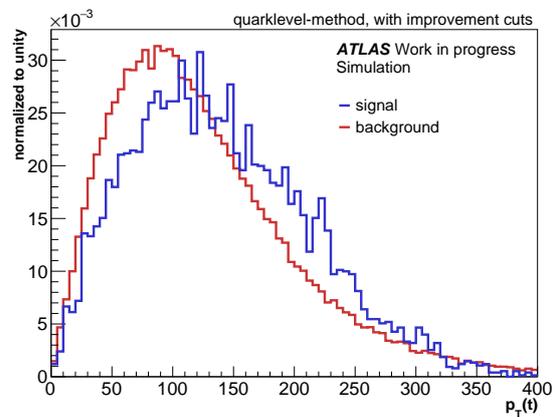
(d) separation power: 0.0368
(0.0368 ± 0.0019)
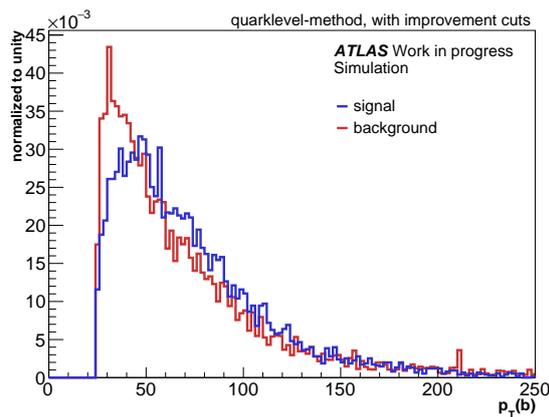
(e) separation power: 0.0355
(0.0355 ± 0.0018)
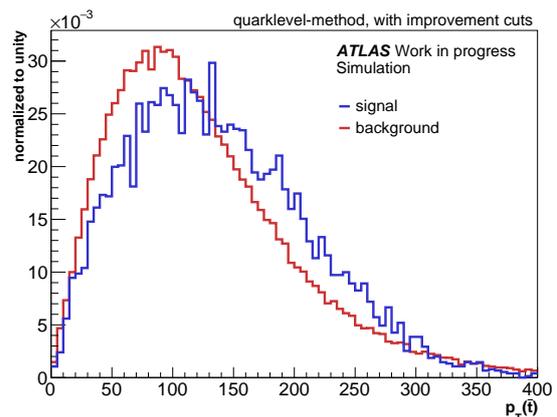
(f) separation power: 0.0338
(0.0338 ± 0.0019)

Figure 7.23: Normalized signal and background distributions of the most separating input variables (listed in table 7.2 and 7.3) for the NN training ordered by decreasing separation power. (part 6)
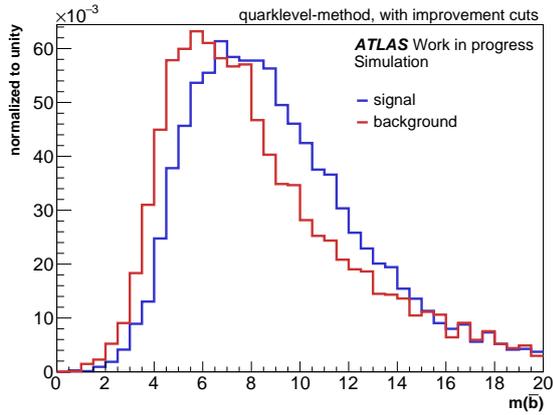
(a) separation power: 0.0332
(0.0332 ± 0.0009)
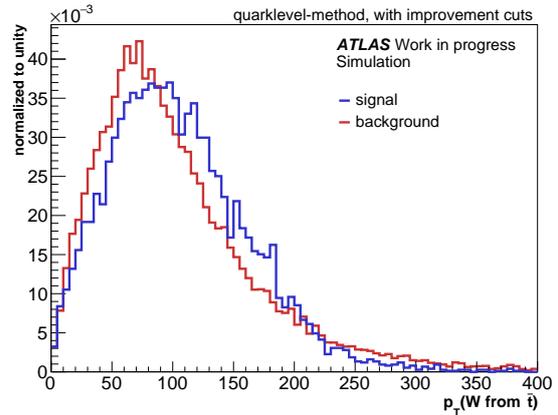
(b) separation power: 0.0313
(0.0313 ± 0.0016)
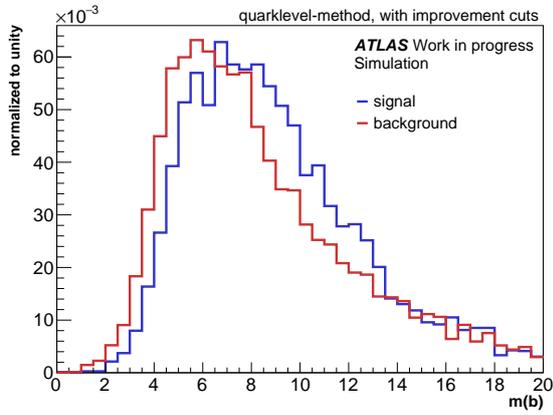
(c) separation power: 0.0311
(0.0311 ± 0.0009)
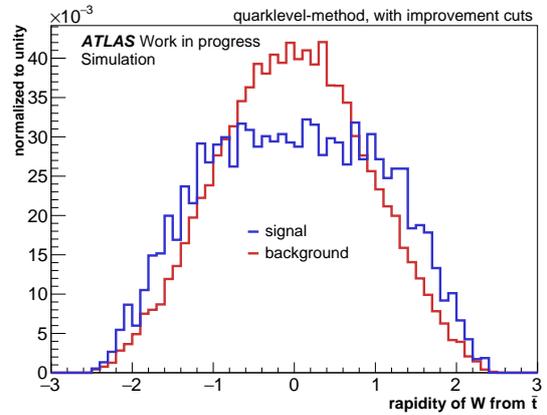
(d) separation power: 0.0232
(0.02317 ± 0.00017)
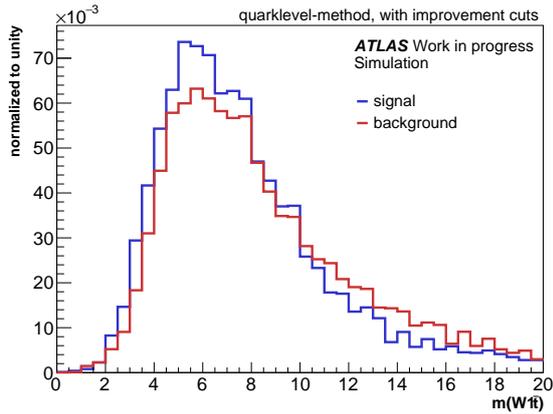
(e) separation power: 0.0228
(0.0228 ± 0.0008)
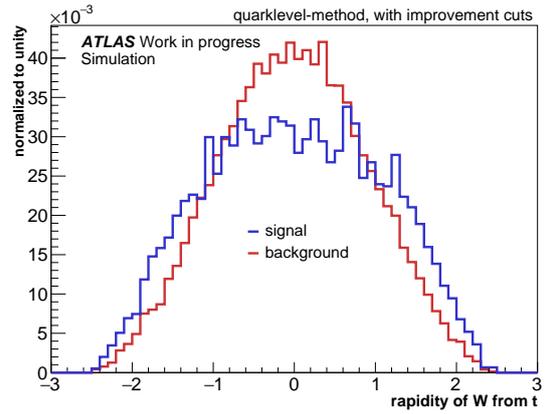
(f) separation power: 0.0220
(0.02197 ± 0.00017)

Figure 7.24: Normalized signal and background distributions of the most separating input variables (listed in table 7.2 and 7.3) for the NN training ordered by decreasing separation power. (part 7)
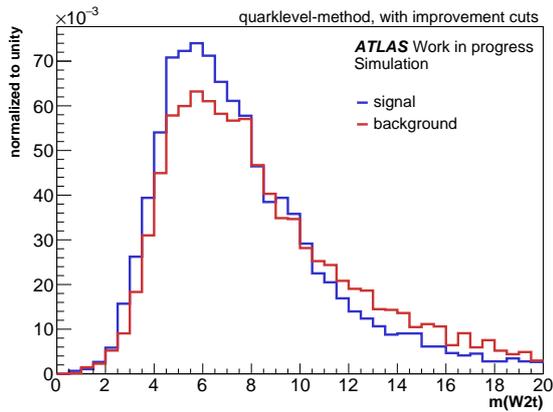
(a) separation power: 0.0217
(0.0217 ± 0.0007)
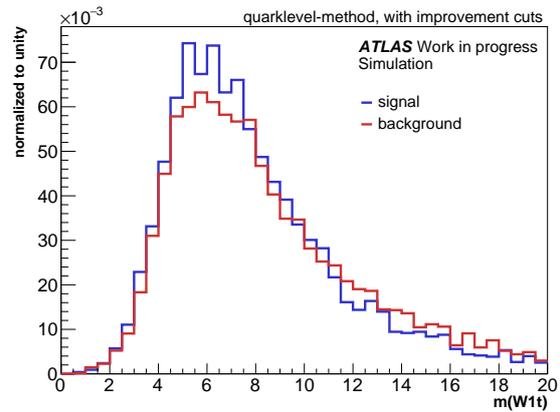
(b) separation power: 0.0153
(0.0153 ± 0.0006)
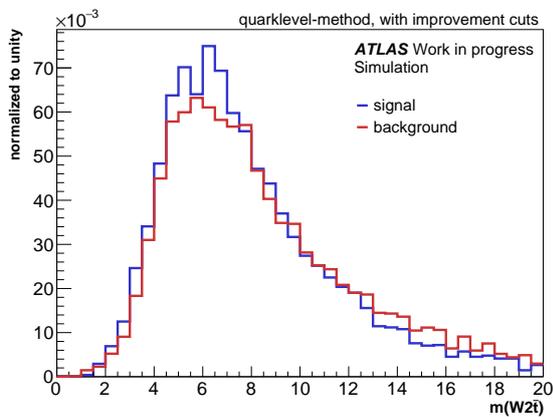
(c) separation power: 0.0143
(0.0143 ± 0.0006)
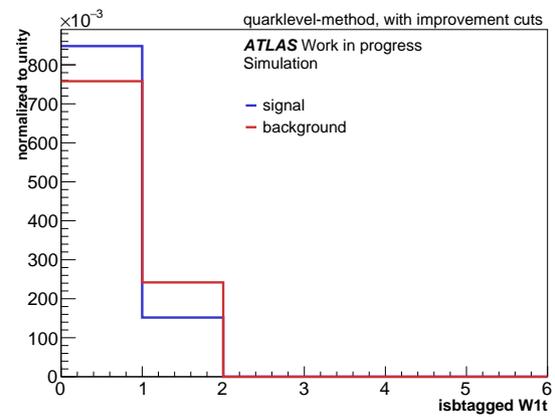
(d) separation power: 0.0128
(0.0128 ± 0.0013)
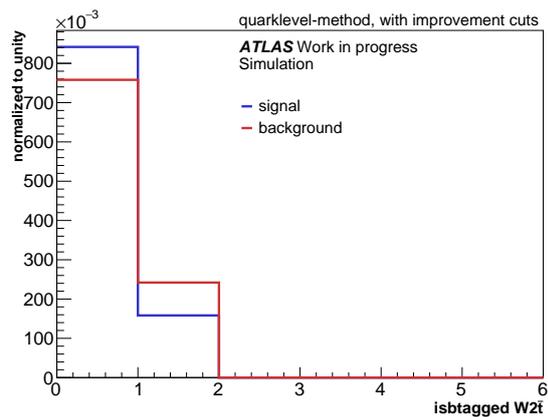
(e) separation power: 0.0109
(0.0109 ± 0.0012)
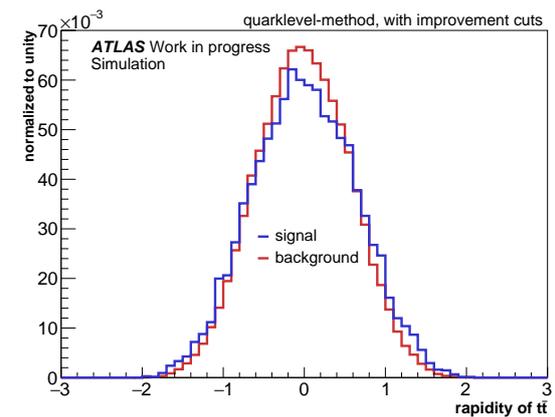
(f) separation power: 0.0067
(0.00667 ± 0.00009)

Figure 7.25: Normalized signal and background distributions of the most separating input variables (listed in table 7.2 and 7.3) for the NN training ordered by decreasing separation power. (part 8)

In order to assure that the Neural Network is able to learn from these diverse observables, they need to be standardized, i.e. their ranges need to be made comparable. To achieve this, a standard approach is to transform the input observables in a way that they exhibit a mean of zero and a standard deviation of 1.

This procedure is essential, as a change of 1 for 2 observables, say first observable ranging from 0 to 1×10^6 and second observable ranging from 0 to 2, has a quite different impact. The standardization is performed during a step called pre-processing.

NN architecture

Up to now a dataset as well as input variables have already been chosen for the training process. So, the definition of the final NN architecture is the only open point.

In order to select the most suitable architecture, various hyperparameter combinations have been tested. The table below shows the hyperparameters (introduced in chapter 4) leading to the best-performing NN together with the results (area-under-curve (AUC) value of the receiver-operating-curve (ROC) curve, accuracy, loss):

Hyperparameters	
neurons per layer	500
ncycles	100
batchSize	32
dropout	0.5
optimizer	adam
activation	leaky ReLu
initializer	normal
regularizer	0.01
classWeight	balanced
learning rate	0.001
decay	0.0001
Results	
AUC value	train: 0.97, test: 0.97
loss	20.52
accuracy	91.94 %

Table 7.4: Hyperparameters leading to the best-performing NN together with its results.

Investigation of deeper as well as narrower NNs exhibited more overtraining. Varying the

activation functions has not had a significant impact. So the state-of-the-art activation function, the leaky ReLu, was chosen. A table summarizing the results (ROC value, loss and accuracy) of all tested hyperparameter combinations can be found in appendix C.

Overall, only a limited number of combinations of hyperparameters could be tested within the scope of this thesis. To optimize the hyperparameter set further e.g. a grid search could be used. As the focus of this thesis is not the optimization of the multivariate reconstruction, this has not been conducted.

7.3.2   Training-performance investigation

Having selected the dataset, the input variables and the architecture of the Neural Network, it is all set to finally perform the training.
To evaluate its performance, multiple visualizations showing the separation power as well as the extent of overtraining are analyzed. Furthermore, it is investigated which input variables are most important for the learning process.

Training output plots

The training is evaluated using several different visualizations of the separation power and the extent of overtraining. The plots discussed below, have already been introduced in subsection 4.2.

First of all, the train-test plot is considered. Figure 7.26 displays the normalized output-score EPD (= Event Probability Discriminant) for the signal and the background hypotheses (in red and blue respectively). The filled distributions are the results of the training process while the points result from the testing process.

Looking at the training results, a clear separation between signal and background can be observed: Most of the signal hypotheses can be found at a value close to 1 and most of the background hypotheses have an output-score close to 0. The overlap between the two distributions is very small. The same trends are also found by looking at the testing results. Regarding the Signal-to-Background ratio (S/B) calculated from the testing distributions a steadily rising trend is noticeable which indicates that moving towards an output-score of 1 more signal hypotheses can be found. The bins indicating an output-score EPD greater than 0.8 have a signal-to-background ratio of greater than 2.

Moreover, the training and testing distributions match each other perfectly. Thus it can be deduced from this plot that no significant overtraining is happening.
This is also confirmed by the results of the Kolmogorov-Smirnoff test (subsection 4.2.1). The values of 0.16 for the signal and 0.04 for the background are good results.
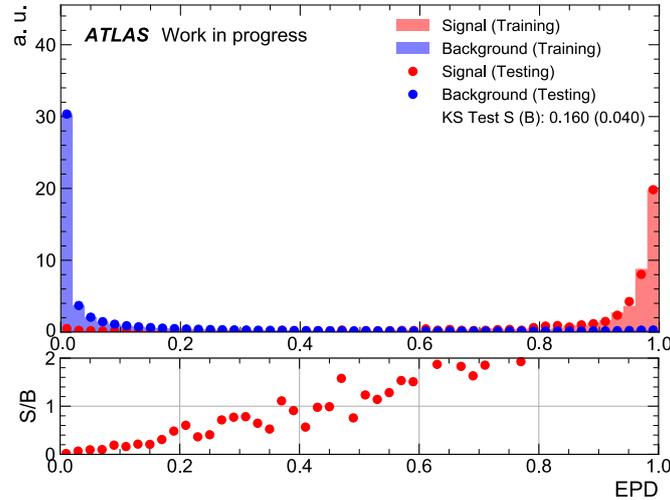
Figure 7.26: Normalized train-test plot (introduced in subsection 4.2.1). The bins indicating an output-score EPD greater than 0.8 have a signal-to-background ratio of greater than 2.

In a next step, the non-normalized, logarithmic distributions are depicted in figure 7.27 using training and testing data. The latter are weighted and scaled to a luminosity of $140\ \mathrm{fb}^{-1}$. The blue distribution indicates the background while the signal distribution is shown in red. The label events on the y-axis of this plot refers to the considered hypotheses.

Here, it can be found that indeed most of the signal (background) hypotheses have an output-score EPD close to 1 (0). In addition, also the imbalanced numbers of signal and background hypotheses are visualized. This shows quite impressively to which extent the number of signal hypotheses is swamped by the number of background hypotheses[11].
Due to this large number of background hypotheses the background distribution is smooth. In contrast, the signal distribution is quite 'bumpy' which indicates that the number of signal hypotheses is at the lower limit as mentioned in section 7.1. Thus, for optimization of the Neural Network a higher number of signal hypotheses would be desirable as mentioned earlier.

To gain further information about the training performance of the NN, the ROC curves together with the AUC value are analyzed (figure 7.28).

---

[11]The small rise of the background hypotheses in the signal bin is assumed to be a result of the chosen binning.
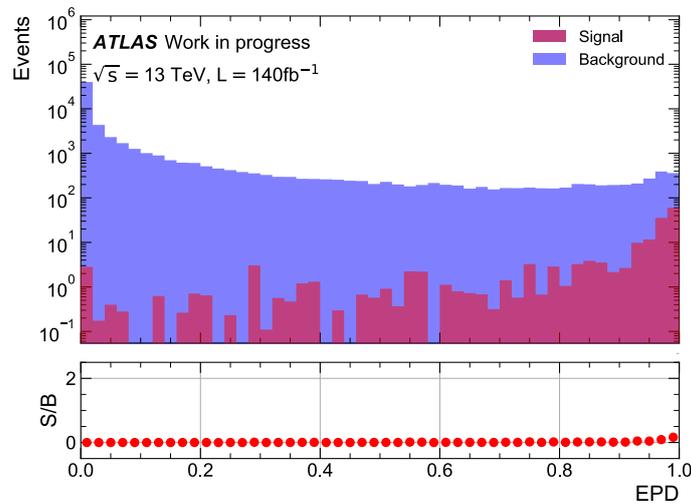
Figure 7.27: Logarithmic and weighted training output-score scaled to a luminosity of 140 fb^{-1} using training and testing data. The label events on the y-axis of this plot refers to the considered hypotheses.
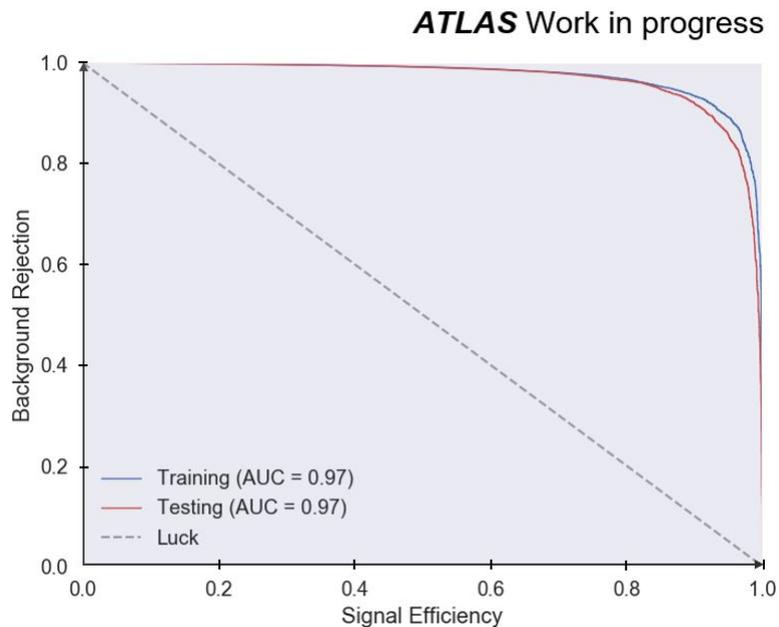
Figure 7.28: ROC curves of training and testing together with their AUC values as introduced in subsection 4.2.2.

Here, the training and testing curves have both almost a rectangular shape. Thus, the NN is obviously able to extract useful information during the learning process as the shapes

are significantly different to the diagonal which would indicate a decision based on luck. This can also be confirmed by looking at the AUC scores of 0.97 which are very high.

Moreover, the two distributions lie very close to each other which indicates only a small amount of overtraining.

The fact that there is a small difference in the ROC curves for the same AUC values is a result of the approximation which has to be chosen in order to numerically calculate the integral to determine the AUC value.

Another way of visualizing the training performance are the confusion matrices shown in figure 7.29.

In figure 7.29a, the results of the training process are summarized. Looking at the deviation of the matrix from the unity matrix, it can be deduced that there is a good, but not perfect, signal-background separation. The same can be observed for the testing results depicted in figure 7.29b.

Furthermore, small differences between the shares of the training and testing matrices can be found. This indicates that overtraining is happening to a small extent.

(a) Training   (b) Testing
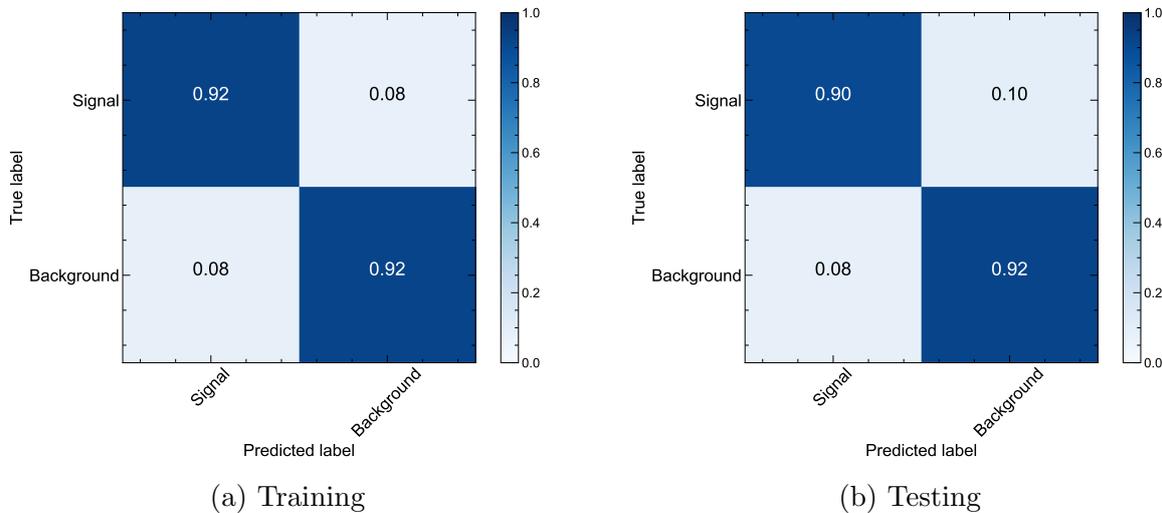
Figure 7.29: Confusion matrices for training (a) and testing (b) as introduced in subsection 4.2.3.

All in all, evaluating the training performance, good separation between signal and background and a small amount of overtraining has been found. Furthermore, the high AUC value of 0.97 indicates a very good performance of the Neural Network.

Variable ranking and correlations

To investigate which observables are most helpful for the NN to perform the reconstruction, the linear correlations among them as well as their linear correlation to the output-score EPD is investigated. The results are displayed in the correlation matrices in figures 7.30 (signal) and 7.31 (background). The values of the correlation coefficients are not included in the matrices as this would overload the plot due to the high number of input variables.

The correlation matrix of the signal (figure 7.30) indicates that quite a few observables are correlated to each other. This is obvious for the different masses and the different transverse momenta because the masses and the transverse momenta are related by $E = mc^2 = \sqrt{\vec{p}^2 c^2 + m_0^2 c^4} = \sqrt{\left(\vec{p_T} + p_z\right)^2 c^2 + m_0^2 c^4}$ with $E/m/\vec{p}$ the energy/mass/momentum of the particle, m_0 the rest-mass of the particle and the velocity of light c. The masses of the top/anti-top quark and the masses of the W bosons are also correlated to the angular differences between the decay products. This can be explained by the formula above and considering that $\vec{p_1} \cdot \vec{p_2} = |\vec{p_1}| \cdot |\vec{p_2}| \cdot \cos\alpha$ with α related to ϕ and η. Interestingly, the masses of the jets are slightly anti-correlated to the observables targeting the angular difference between the decay products. For the same reason, also the p_T values are (anti-)correlated to the angular variables, especially $\Delta\phi$ and ΔR. They are correlated by the relation $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$. Moreover, also a correlation between the helicity properties and the observables targeting the angular distance between the decay products is found. This can be explained by the fact that both observables rely on angular information related to the topology of the system.

Regarding the correlation of the input variables to the output-score EPD, it can be found that the most contributing observables are the is-btagged(b) and the is-btagged($\bar{\text{b}}$). Furthermore, also the masses of the top/anti-top quark, the masses of the W bosons as well as the observables targeting the ΔR and the $\Delta\eta$ between the decay products are quite strongly (anti-)correlated to the output-score. This agrees well with the high separation power of these observables found in subsection 7.3.1.

Thus, on top of the information used in the χ^2 reconstruction (masses and is-btagged), also the developed observables targeting the angular distance between the decay products are important for the NN.

Figure 7.31 shows the correlation matrix of the background. Here, the same basic trends can be found as for the signal. In contrast to the signal correlation matrix, it can be observed that the correlation between the masses/transverse momenta and the variables targeting the angular differences between the decay products get smaller. Moreover, the mass of the $t\bar{t}$ system $m_{t\bar{t}}$ is found to be stronger correlated to the angular observables, while the transverse momenta of the down-type quarks are less correlated to the observables targeting the angular distance between the decay products. The correlation between $\Delta\phi$ and $\Delta\eta$ has vanished as well.

Regarding the correlation between the output-score EPD and the input variables, also the same correlations can be found. Apart from that, there is a stronger correlation to the mass of the $t\bar{t}$ system $m_{t\bar{t}}$.
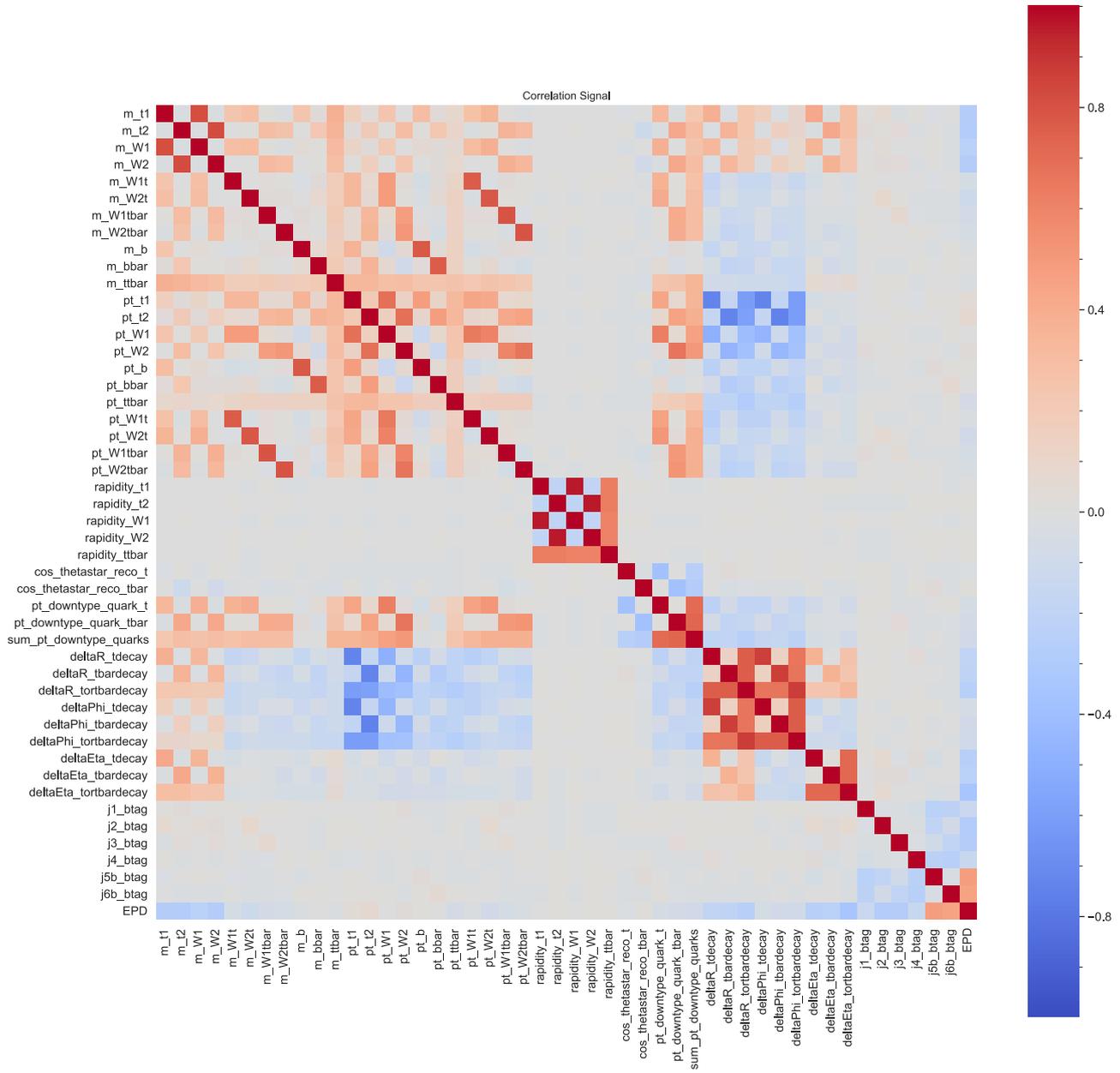
Figure 7.30: Correlation matrix of the signal.
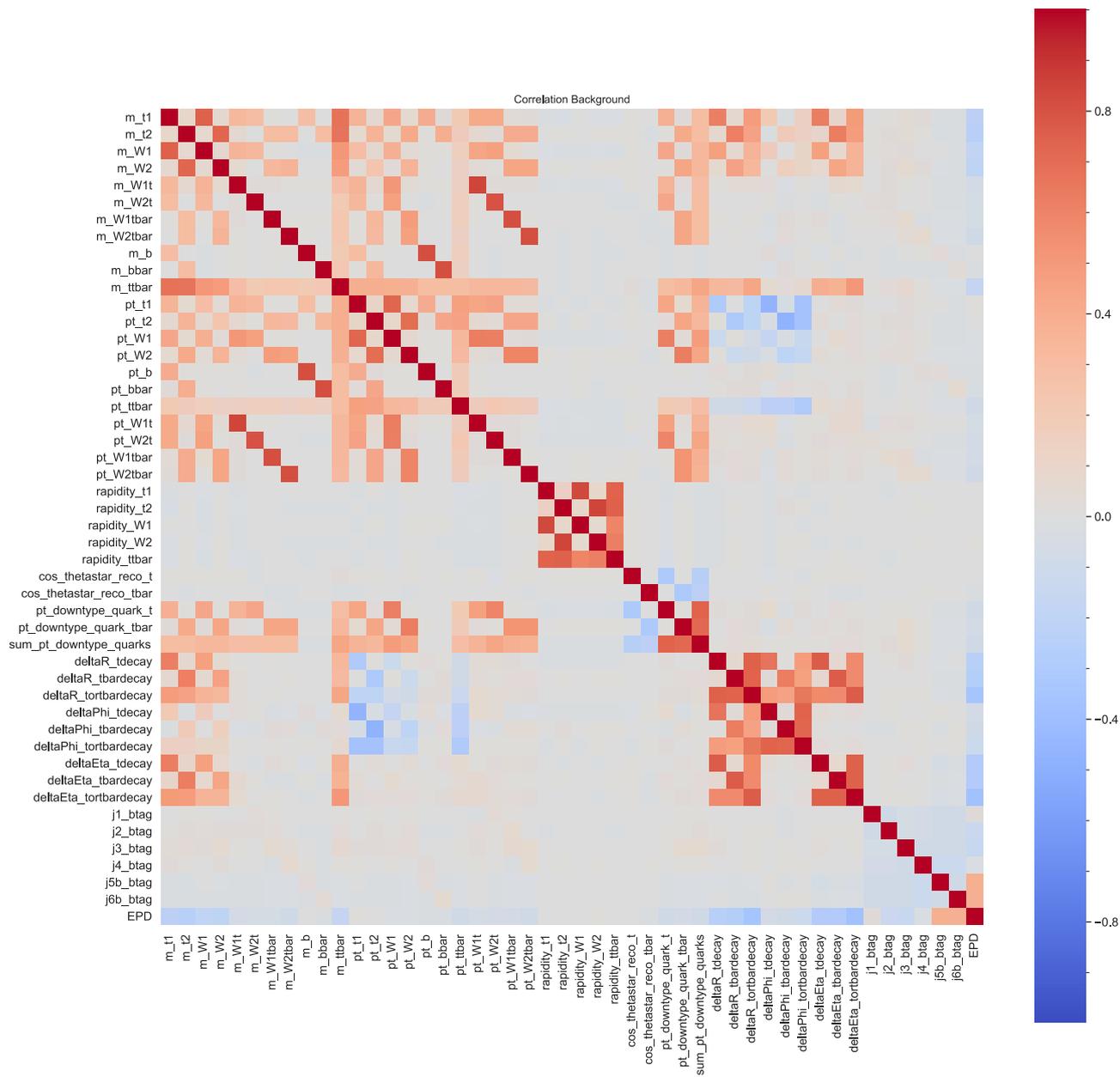
Figure 7.31: Correlation matrix of the background.

Another complementary way to discover which observables contribute most to the decision is to rank the variables. For that purpose, multiple metrics exist. Three of them are shown in the following, in particular the sum of absolute weights, the sum of squared weights and the sum of gradients.

For the 'sum of the absolute weights' ranking, the absolute values of the weights of the connection between the first hidden layer and the neuron related to the particular variable in the input layer are added. A similar approach is performed for the 'sum of squared weights' ranking with the only difference that here the squared weights are summed up. For the 'sum of gradients' ranking, the variation of the output ∂y in the first hidden layer depending on the particular variable x_i is calculated and multiplied with the standard deviation of the particular variable σ_{x_i}. As within the pre-processing the input variables x_i underwent a transformation which resulted in a mean of $\mu_{x_i} = 0$ and a standard deviation $\sigma_{x_i} = 1$, only $\frac{\partial y}{\partial x_i}$ is used here.

Highly discriminating variables should be ranked quite high for each of the three metrics.

In table 7.5, the 15 most important input variables according to the three metrics are listed. The results are highlighted in different colors depending on the physical categories: kinematic variables (blue and orange), is-btagged (green), helicity properties (magenta/pale red), angular differences between the decay products (yellow) and the rapidities (gray). The complete list containing all input variables is contained in appendix B.

Obviously, the most contributing observable is the mass of the $t\bar{t}$ system $m_{t\bar{t}}$ being placed on the top position for all three metrics. According to the two weight-based metrics, the is-btagged observables as well as the transverse momenta contribute significantly to the training. Furthermore, also the $\Delta\eta$ differences between the decay products are ranked quite high. This is reassuring to see as they have been developed especially for the multivariate reconstruction addressed in this thesis.

From the ranking based on the gradient metric, it can be seen that here the transverse momenta play an important role in the training as well. In contrast to the other two metrics, the is-btagged observables as well as the $\Delta\eta$ distances between the decay products are not ranked within the top 15. Rather the masses of the W bosons and the masses of the top quarks as well as the transverse momenta of the down-type quarks can be found.

Furthermore, it is eye-catching that no rapidities of the top quarks/W bosons are within the top 15 most contributing observables.

Concluding, it can be retained that besides the masses and the is-btagged observable which is exactly the information used in the χ^2 reconstruction, also the observables targeting the angular difference between the decay products, transverse momenta as well as the helicity properties contribute significantly to the training process and hence to the reconstruction of the hadronic $t\bar{t}$ systems in $t\bar{t}Z$ events.

	sum of squared weights	sum of absolute weights	gradients
1.	$m_{t\bar{t}}$	$m_{t\bar{t}}$	$m_{t\bar{t}}$
2.	is-btagged($\bar{b}$)	is-btagged($\bar{b}$)	$m_{W\text{ from t}}$
3.	is-btagged(b)	is-btagged(b)	m_t
4.	$p_T(W1\bar{t})$	$p_T(W1\bar{t})$	$p_T(\text{down-type quark t})$
5.	is-btagged($W1\bar{t}$)	is-btagged($W1\bar{t}$)	$p_T(t)$
6.	is-btagged(W2t)	$p_T(t)$	$p_T(W2\bar{t})$
7.	$p_T(W1t)$	$p_T(W1t)$	$p_T(\bar{t})$
8.	$p_T(b)$	$p_t(\bar{b})$	$p_T(W1\bar{t})$
9.	$p_T(\bar{b})$	is-btagged(W2t)	$p_T(W1t)$
10.	$\Delta\eta(\text{decay products of t})$	$p_T(b)$	$m_{W\text{ from }\bar{t}}$
11.	$p_T(t)$	is-btagged($W2\bar{t}$)	$p_T(\text{W from t})$
12.	$\Delta\eta(\text{decay products of }\bar{t})$	$p_T(W2t)$	$p_T(\text{down-type quark }\bar{t})$
13.	is-btagged($W2\bar{t}$)	$p_T(W2\bar{t})$	$p_T(W2t)$
14.	$p_T(W2t)$	m_t	$p_T(\bar{b})$
15.	$m_{\bar{b}}$	$\Delta\eta(\text{t or }\bar{t}\text{ decay})$	$m_{\bar{t}}$

Table 7.5: Table displaying the 15 most important input variables ranked according to three metrics, 'sum of squared weights', 'sum of absolute weights' and 'gradients', highlighted in different colors depending on the physical categories: kinematic variables (blue and light orange), is-btagged (green), helicity properties (magenta/pale red) and the angular differences between the decay products (yellow).

7.4   Performance investigation of the multivariate reconstruction

In the previous section, the Neural Network has been trained and its training performance has been studied. As the investigations indicated a good training performance, the trained Neural Network is assumed to be able to truthfully classify hypotheses as signal (well-reconstructed $t\bar{t}$ systems) and background (badly-reconstructed $t\bar{t}$ systems). As this represents the multivariate reconstruction of hadronic $t\bar{t}$ systems in $t\bar{t}Z$ events, also a good reconstruction performance is expected.

To verify this, un-labeled data are used which consist basically of the same hypotheses as the ones obtained by applying the quarklevel method in combination with the event selection 'with improvement cuts'. The only difference is that no labels have been assigned to the hypotheses.
To find out how well the classification/reconstruction works, the trained NN is now applied to these un-labeled hypotheses and is asked to classify them, i.e. to conduct the reconstruction. This step is referred to as 'evaluation'. Based on the obtained results, detailed investigations regarding the classification quality/reconstruction performance are conducted in the following.

In a first study, the hypothesis yielding the highest output-score in the event is selected for each event. These hypotheses are in the following referred to as the 'best NN hypotheses' and also represent the best reconstructed $t\bar{t}$ systems of the multivariate reconstruction. The output-scores of the best NN hypotheses are plotted in figure 7.32.

It can be observed that the output-score is extremely high for all best NN hypotheses and thus, according to the NN, they are all extremely signal-like. Based on this, it seems reasonable to expect that the NN can apply its 'knowledge' successfully to un-labeled data and that the multivariate reconstruction works well.

To confirm this, the best NN hypotheses are further investigated. In this context, it is identified to which categories (signal, flip categories, categories 1 to 6) these hypotheses were initially assigned, i.e. before training. Thus their 'true' label is queried. Table 7.6 displays the number of best NN hypotheses carrying a particular 'true' label.

In contrast to the expectation, the share of the signal is vanishingly small while the share of the background is extremely large. This indicates that the NN was not able to truthfully differentiate between signal (well-reconstructed $t\bar{t}$ systems) and background (badly-reconstructed $t\bar{t}$ systems).
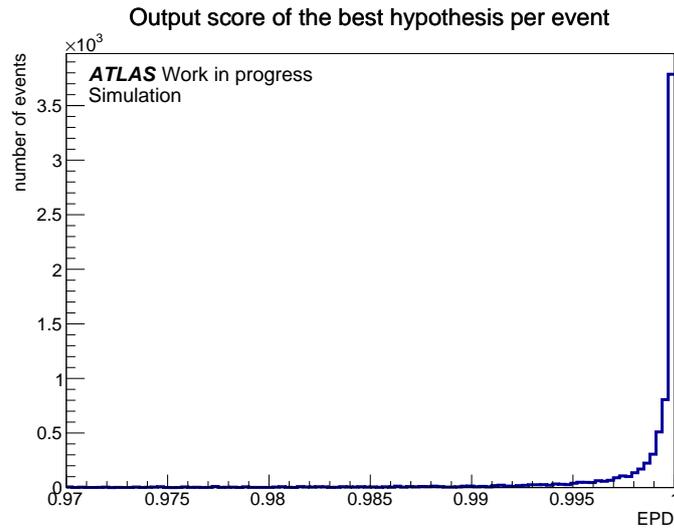
Figure 7.32: Distribution of the output-score using the hypotheses exhibiting the highest output-score in the event.

category	number of hypotheses within best NN hypotheses (7512)
signal	14 (0.2%)
background	7498 (99.8%)
background categories	
merged category	4535 (60.4%)
3-flip category	1277 (17.0%)
2-flip category	1361 (18.1%)
1-flip category	323 (4.3%)
0-flip category	2 (0%)
(sub-)categories	**based on 0-flip and 1-flip categories (325)**
category 1	2 (0.7%)
category 2	27 (8.3%)
category 3	279 (85.8%)
category 4	5 (1.5%)
category 5	4 (1.2%)
category 6	8 (2.5%)

Table 7.6: Table indicating the number of best NN hypotheses exhibiting a particular 'true' label as defined in the text. The different background categories are defined in section 7.2.

The most represented background categories among the hypotheses with the best output-scores are the merged category and the multi-flip categories, thus exactly the categories which were not included in the training. This indicates that the NN obviously does not know how to handle hypotheses of these categories as it has never 'seen' them before. The NN has not recognized that these categories are only combinations of the physical scenarios used in the training. The fact that their distributions are very background-like, apparently does not make a difference.

Looking at the categories used for training (0-flip and 1-flip categories), the 1-flip category holds the largest share. Especially (sub-)category 3 (switching among light jets between top quarks) and (sub-)category 2 (switching among light jets of the same top quark) are obviously challenging for the NN. This is indicated by their high misclassification shares. Hypotheses of these categories are indeed hardly distinguishable from the signal. The only possibility would be to use the relatively unprecise jet charge (which was not included in the list of input variables). As already outlined in section 7.2, it might therefore be reasonable to accept category 2 as signal. This is not discussed further and is just mentioned as an outlook here.

Focusing on the signal-like background (category 1/0-flip category) only a negligible contribution can be found.

These findings can also be visualized using a stack-plot showing the mass of the top quark and the mass of the anti-top quark. This is shown in figure 7.33.

(a)   (b)
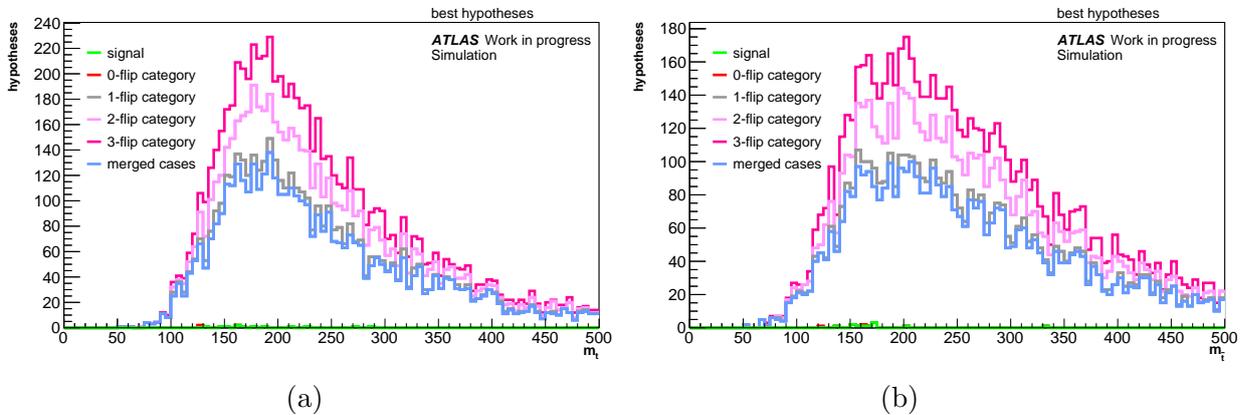
Figure 7.33: Stack-plots of the masses of the top quark (a) and the anti-top quark (b) for the best NN hypotheses initially assigned to a particular category (signal, different background categories).

Here, the huge impact of both, the multi-flip categories as well as the merged category, on the overall shape is striking. The impact of the shapes of the 0-flip and 1-flip categories is negligible.

Furthermore, it is eye-catching that the distribution of the anti-top quark (figure 7.33b) is significantly broader than the one of the top quark (figure 7.33a). This indicates that

the NN used is either unable to deal with the huge number of possible combinations or it erroneously learns to differentiate between the mass of the two top quarks. Thus, it might be reasonable to use a more sophisticated model for this complex reconstruction problem. One possible solution could be to use several Neural Networks for the different categories and then use the combined output for classification. Furthermore, increasing the number of available signal hypotheses or enable training on all hypotheses might be helpful as well.

To gain further insights about the classification quality or rather reconstruction performance, the focus is shifted. In the following only hypotheses with an output-score EPD $\geq$ 0.95 are considered[12] and their distribution (representing the situation after evaluation) is compared to the distribution of the initial situation, i.e. the distributions before training (based on all hypotheses of all events).
For this purpose, the distributions of two observables found to be strongly correlated to the output-score (figures 7.30 and 7.31) are plotted: the mass of the top quark and the sum of the ΔR values between the decay products of the top quark and the anti-top quark. In addition, also the distribution of an observable with negligible correlation to the output-score, i.e. is-btagged(W2t), is drawn.
Querying the 'true' label of the hypotheses, the distributions are further differentiated. This is indicated by depicting the distributions of hypotheses with different 'true' labels in different colors.

Figure 7.34 shows the normalized distributions of the mass of the top quark before the training (for all hypotheses of all events, figure 7.34a) and after evaluation separately for the (sub-)categories and the flip categories (for hypotheses with an output-score $\geq$ 0.95, figures 7.34b and 7.34c). In all subfigures, the signal is included as a reference.

Comparing figures 7.34a to figures 7.34b and 7.34c, no significant changes can be observed apart from statistics which is a result of only displaying the distributions for hypotheses with an output-score $\geq$ 0.95. The distributions after training are thus the same as before the training which contrasts the expectation: As the mass of the top quark has been found to be a strongly correlated observable, it is assumed that it contributes significantly in the learning process of the NN. Considering only the most signal-like hypotheses, extremely similar and signal-like shapes are expected in consequence. As still significant differences between the shapes of the diverse categories are visible, it can be concluded that the Neural Network was not able to extract useful information based on this observable to truthfully differentiate between well-reconstructed from badly-reconstructed $t\bar{t}$ systems.

---

[12]It was not possible to use the best hypotheses for this studies due to their low number and thus large statistical uncertainties.
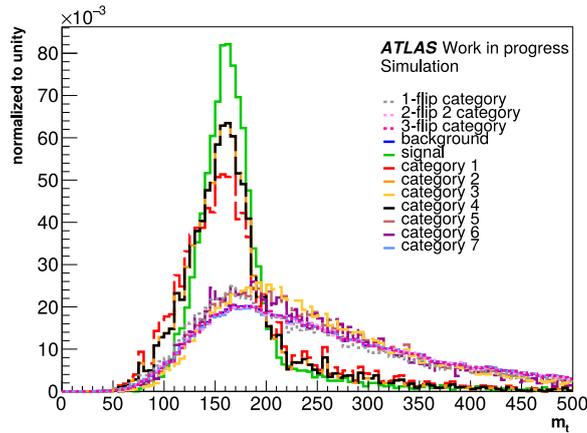
(a) initial (before training)
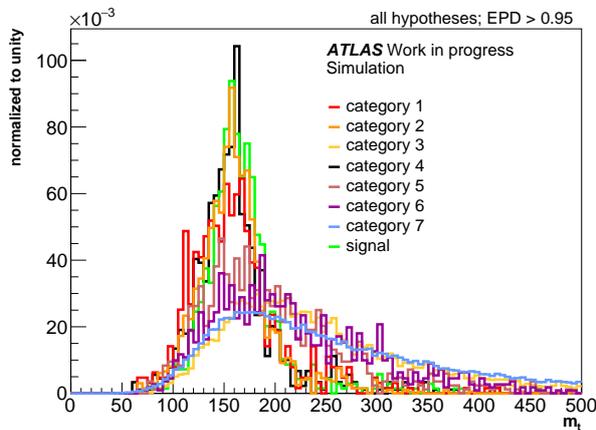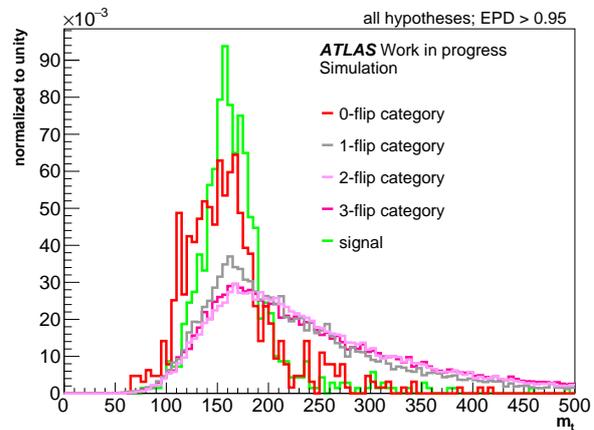
(b) after evaluation

(c) after evaluation

Figure 7.34: Normalized distributions of the top-quark mass m_t for the signal and background category as well as for the different background categories (section 7.2) for the initial situation (before training) (a) and after the evaluation (b and c). For the distributions after the evaluation only hypotheses with an output-score ≥ 0.95 have been considered.

This contradiction might seem surprising at first sight. However, a closer look reveals that only the *linear* correlation coefficients have been considered in figures 7.30 and 7.31. The Neural Network is a *nonlinear* model though as explained in chapter 4. Thus the most contributing observables exhibiting a large linear correlation coefficient are not necessarily the observables from which the NN extracts most information.

Furthermore, only hypotheses with an output-score ≥ 0.95 are considered while the correlation matrices have been determined using all hypotheses. Recalculating the correlation matrices only for this subset of hypotheses might yield some changes in the values of the linear correlation coefficient. Due to statistical reasons, this has not been conducted here.

Nevertheless, the NN was obviously not able to extract useful information for truthful classification during reconstruction. This indicates that a more sophisticated MVA method would be required in order to improve the reconstruction performance. Another way would be to increase the number of available signal hypotheses in training or to enable training using all hypotheses.
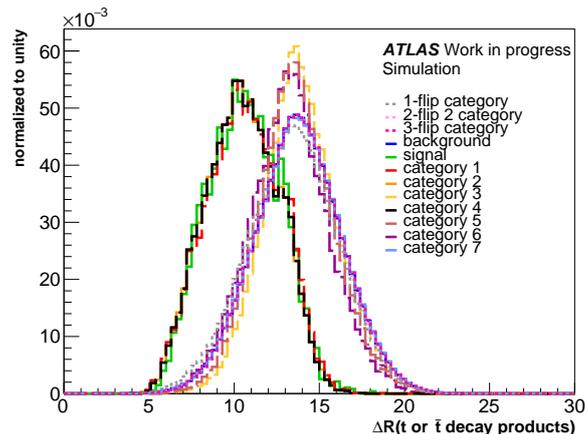
(a) initial (before training)
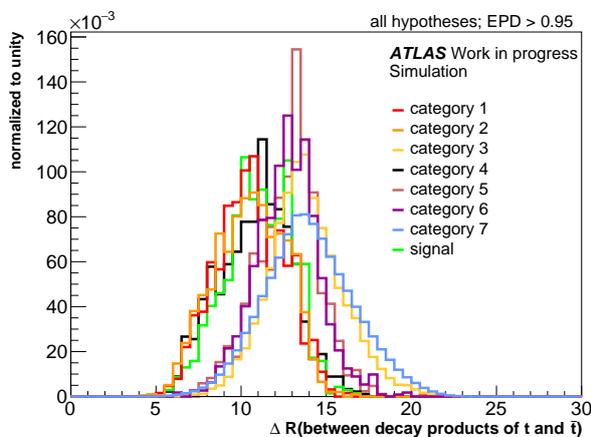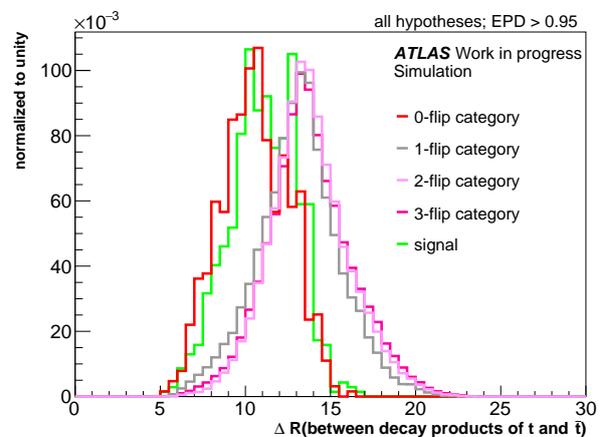
(b) after evaluation

(c) after evaluation

Figure 7.35: Normalized distributions of the sum of the ΔR between the decay products of the top quark and the anti-top quark the signal and background category as well as for the different background categories (section 7.2) for the initial situation before training (a) and after the evaluation (b and c). For the distributions after the evaluation only hypotheses with an output-score ≥ 0.95 have been considered.

In a next step, the sum of the ΔR values between the decay products of the top quark and the anti-top quark is studied. Figure 7.35 displays the normalized distributions before the

training (for all hypotheses of all events, figure 7.35a) and after the evaluation separately for the (sub-)categories and the flip categories (for hypotheses with an output-score ≥ 0.95, figures 7.35b and 7.35c). In all subfigures, the signal is included as a reference.

Here, the same trends can be found as for the mass of the top quark in figure 7.34.
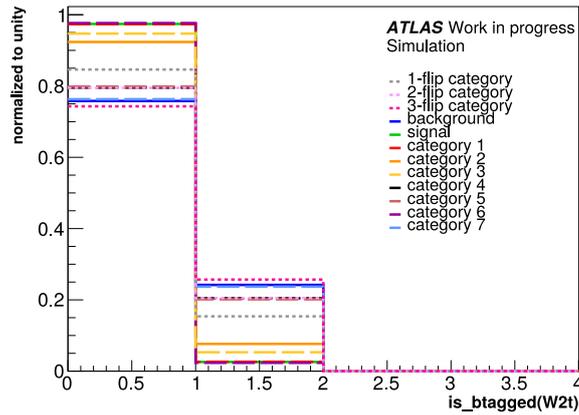
(a) before training
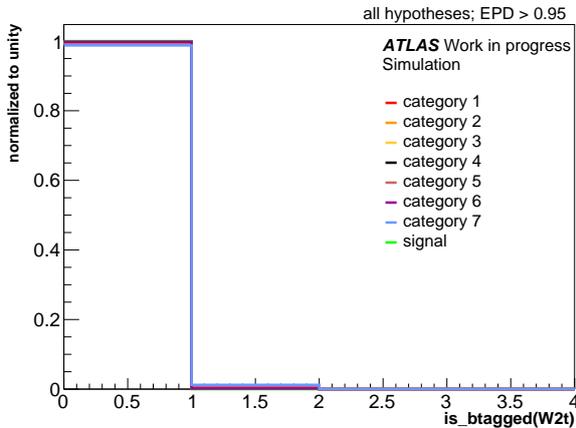
(b) after evaluation
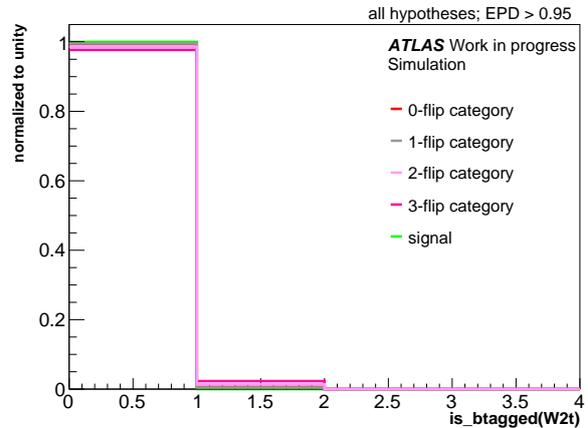
(c) after evaluation

Figure 7.36: Normalized distributions of the is-btagged(W2t) observable for the signal and background category as well as for the different background categories (section 7.2) before (a) and after the training (b and c). For the distributions after the training only hypotheses with an output-score ≥ 0.95 have been considered.

Looking at the results of the is-btagged(W2t) observable in figure 7.36, a different trend can be found. Here, the distributions after the evaluation (figures 7.36b and 7.36c) exhibit all the same shape, i.e. the shape of the signal, which is not the case for the distributions before the training (figure 7.36a). As only very signal-like hypotheses (EPD ≥ 0.95) are

considered, this is in principle the expected result. Thus, apparently, the NN can extract some information from this observable during the training process.

However, the is-btagged(W2t) exhibits only a small correlation to the output-score in figures 7.30 and 7.31 and was consequently expected to contribute negligibly to the learning process. However, taking into account that the small correlation found refers to the *linear* correlation coefficient and the Neural Network is a *nonlinear* model, the observations are plausible.

Concluding, it can be retained that even though the output-scores of the best hypotheses in the event, i.e. the reconstructed $t\bar{t}$ systems, are extremely high, the chosen Neural Network is obviously not able to truthfully differentiate between well- and badly-reconstructed $t\bar{t}$ systems for most hypotheses and thus does not exhibit a good reconstruction performance. One reason might be that not all hypotheses of the different background categories could be included in the training due to memory limitations. Those are exactly the ones which could not be classified well. Furthermore, also the huge number of possible configurations present in the reconstruction of hadronic $t\bar{t}$ systems is quite challenging for the chosen NN. It also seems to learn to erroneously differentiate between the properties of the two top quarks. Moreover, the chosen NN apparently extracts too little information for classification as some distributions after the evaluation (based on hypotheses exhibiting an output-score ≥ 0.95) are the same as the initial ones (before the training). This indicates that a more sophisticated model is needed for truthful classification and for good reconstruction performance. Another way to improve the reconstruction performance could be to increase the number of available signal hypotheses for training.

7.5 Conclusion of the multivariate reconstruction

In this chapter, an multivariate reconstruction of hadronic $t\bar{t}$ systems in $t\bar{t}Z$ events has been developed and applied using a Neural Network which was trained and evaluated.

As the training was conducted in the context of supervised learning, labeled input data were required. That is why a method to produce labeled input data was selected in a first step. In this context, three different methods have been introduced: the Wtlevel method, the quarklevel method and the ctc method. Their performance was then compared among each other as well as for two different event selections, i.e. 'without improvement cuts' and 'with improvement cuts'. Based on this comparison, the quarklevel method in combination with the event selection 'with improvement cuts' was found to be the most suitable method to produce labeled input data for the training process in this thesis.

In a next step, the hypotheses assigned to the background category by the selected method have been studied in more detail as they combine a variety of physical scenarios. Therefore, different background categories, the flip categories and the (sub-)categories 1 to 6 have been defined and investigated. In this context, it was found that the 0-flip and 1-flip

categories are very signal-like and the merged category as well as the multi-flip categories are background-like.

As only a subset of the labeled input data could be used for training due to memory limitations, the training was conducted on the full signal and only on the most signal-like background categories, i.e. the (sub-)categories of the 0-flip and 1-flip categories.

Furthermore, for the training of the NN also decisions regarding suitable input variables and the NN architecture were required. For the input variables, observables covering kinematic information (masses and transverse momenta), angular information (rapidities of the W bosons and the top/anti-top quark), b-tag and helicity information have been chosen. In this context, also input variables especially targeting the identified reconstruction challenge of a changing topology in chapter 6 were developed and included in the list of input variables. Regarding the NN architecture, a Neural Network with one hidden layer containing 500 neurons has been selected based on its good performance regarding signal-background separation and the exhibited amount of overtraining. Subsequently the training was conducted and its performance was studied. In this context, a good training performance of the NN has been found. Furthermore, variables most important for the training process have been identified. It has been found that, apart from the masses and the is-btagged observables which are also used in the χ^2 reconstruction, the transverse momenta, the helicity properties and the angular distances between the decay products of the top/anti-top quark contributed significantly to the learning process.

Finally, the trained NN was then evaluated on un-labeled data and the classification quality, i.e. the reconstruction performance, was investigated. It was found that the chosen NN is unfortunately not able to truthfully classify un-labled hypotheses even though the output-scores of the best hypotheses in the event, i.e. the reconstructed $t\bar{t}$ systems, were found to be extremely high.

One possible reason might be the fact that not all hypotheses of the different background categories could be included in the training due to memory limitations. It was striking that hypotheses belonging to the excluded categories were exactly the ones which could not be classified truthfully. Furthermore, also the huge number of possible configurations present in the reconstruction of hadronic $t\bar{t}$ systems was found to be challenging for the chosen NN. While 3 decay products could be handled acceptably well, 6 decay products overchallenged the chosen NN. In addition, the chosen NN seems to erroneously learn a non-existent difference in the properties of the two top quarks. Moreover, it was found that the chosen NN extracts too little information for classification. This might be due to the low number of signal hypotheses available for training. It might also be an indication for the fact that a more sophisticated model is needed to perform a good classification/reconstruction.

A possible approach could be to use several NNs for the distinct physical scenarios and consequently to use their combined output-scores for classification/reconstruction.

Chapter 8

Performance comparison of the minimum-χ^2 and the multivariate reconstruction

In the previous chapters, two approaches to reconstruct hadronic $t\bar{t}$ systems in $t\bar{t}Z$ events have been presented: the minimum-χ^2 reconstruction and the multivariate reconstruction using a Neural Network. Both of them have been investigated in detail and their individual reconstruction performance has been studied. In this chapter, they are now compared.

In this context, the 'best hypothesis', i.e. the hypothesis containing the best-reconstructed $t\bar{t}$ system according to the particular reconstruction method (minimal χ^2 value for the χ^2 method and largest output score for the NN-based reconstruction), is selected for each event. To gain more detailed insights regarding the quality of the input data, also the quarklevel method is considered which selects its best hypotheses based on the largest $\Delta R_{\text{total}}^{\text{q}}$ value (subsection 7.1.1). It is then investigated how often the respective methods have selected the same best hypotheses.
Moreover, also the matched (truth matching) shares of the different reconstruction methods are determined by comparing the best hypotheses to the matched hypotheses. These represent hypotheses which contain $t\bar{t}$ systems assigned to the matched category (subsection 6.1.3).

The percentage of cases in which the same best hypotheses are selected by two methods are visualized in the diagram in figure 8.1.
Focusing on the identical best hypotheses of the NN-based reconstruction and the minimum-χ^2 method, a vanishingly small share of 0.2% can be identified. The same share can be found comparing the best hypotheses of the NN-based reconstruction to the quarklevel method, i.e. the method which provides the labeled input data for the NN. Moreover, also a vanishingly low share (0.1%) of best hypotheses of the NN-based reconstruction coincides with the matched hypotheses. In contrast to this, the share of best hypotheses of the quarklevel method which are in line with the matched hypotheses is found to be almost

145

50% which is in comparison very high.
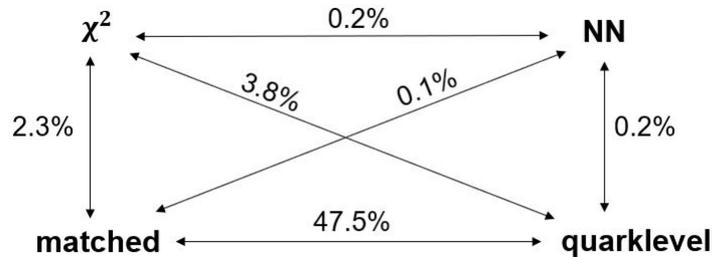
Figure 8.1: Diagram visualizing the percentage of cases in which the same best hypotheses were selected by two reconstruction methods. In addition also the matched shares (subsection 6.1.3) of the best hypotheses of the different methods are determined by comparing the best with the matched hypotheses.

Looking at the minimum-χ^2 reconstruction method and the quarklevel method a small share (3.8%) of identical best hypotheses has been selected by both methods. Thus, the different information used in the two approaches apparently results in a selection of different hypotheses as best hypotheses in a majority of cases.
Comparing the best hypotheses of the χ^2 reconstruction to the matched hypotheses, a share of 2.3% is found. This coincides with the matched share found in the context of the truth-matching procedure in chapter 6.

Based on these findings, it can be deduced that using the quarklevel method to produce the input data for the training is in principle a good approach. However, the reconstruction performance of the chosen Neural Network is quite bad. Thus, even though its input data yield great potential for a good truth-matching share, in the end, this reconstruction approach does not work well.
Unfortunately, the quarklevel method cannot be used as a stand-alone method for reconstruction as truth information is used. Hence, it is not applicable to experimental data as for those no truth information is available.

To gain further insights regarding the reconstruction performance of the NN-based approach in comparison to the minimum-χ^2 approach, the distributions of some key observables are plotted. In this context, only the best hypotheses of the χ^2 reconstruction (red) and of the NN-based reconstruction (blue) are considered.
As a first observable, the mass of the $t\bar{t}$ system is investigated. The normalized distributions of the two reconstruction approaches can be found together with the truth information (yellow) in figure 8.2.

Here, both distributions based on the best hypotheses of the χ^2 reconstruction and of the NN-based reconstruction have a peak at a value of roughly twice the top-quark mass. Moreover, they fit each other very well. In comparison, the distribution of the truth information exhibits a slight right-shift. This might be explicable considering hadronization which is not taken into account by the distribution of the truth information. It could also indicate that both reconstruction methods have chosen some hypotheses as best hypotheses which contain jets not originating from the $t\bar{t}$ system. Apart from this difference, the mass of the $t\bar{t}$ system is very well reproduced.
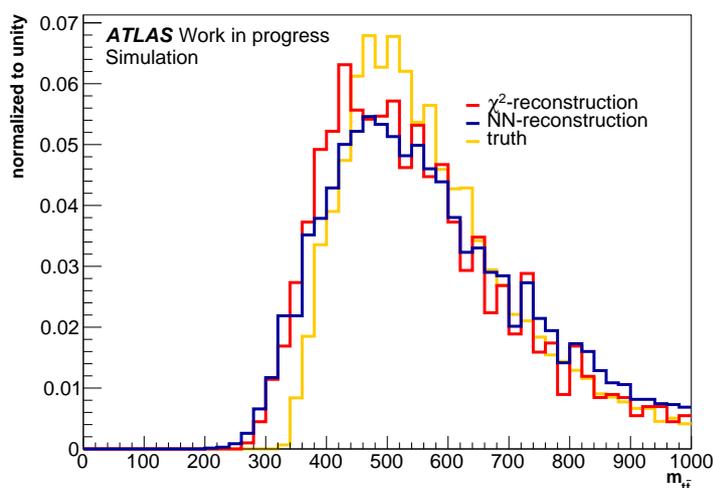
Figure 8.2: Normalized distributions of the mass of the $t\bar{t}$ system for the best hypotheses of the χ^2 reconstruction and the NN-based reconstruction. As a reference, also the distribution of the truth information is included.

In a next step, the mass distributions of the top and the anti-top quark are studied. They are depicted in figure 8.3 in the same color-scheme as above. The truth information is not included here in order to assure good visibility of the shape differences.

In contrast to the findings of figure 8.2, significant differences can be identified here: While the distributions based on the best hypotheses of the χ^2 method exhibit nice sharp peaks with small tails situated around a value of about 173 GeV, thus exactly at the top-quark mass, the distributions based on the best hypotheses of the NN-reconstructed top quarks are significantly broader. In particular, they exhibit significantly broader tails towards larger mass values. This indicates that the Neural Network often selects hypotheses containing wrong configurations of the selected jets as found in figure 7.10.
All in all, this indicates that the best hypotheses of the NN-based reconstruction cannot reproduce some of the key properties of the $t\bar{t}$ system and hence that the reconstruction performance and quality of the NN-based approach is far lower than the one of the χ^2

approach.

(a)   (b)
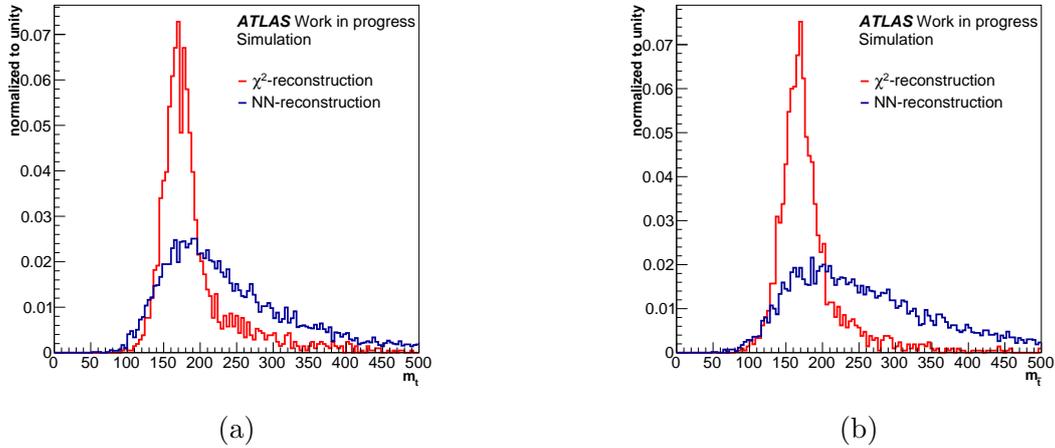
Figure 8.3: Normalized distributions of the mass of the top quark (a) and anti-top quark (b) for the best hypotheses of the χ^2 reconstruction and the NN-based reconstruction.

Comparing the mass distributions of top and anti-top quark among each other, it can be found that in the case of the NN-based reconstruction, the mass of the top quark is slightly better reproduced than the mass of the anti-top quark as the latter is found to be a lot broader. While the maximum of the top-quark mass distribution is situated roughly at the same value, the maximum of the anti-top-quark mass distribution is situated more at about 200 GeV. This indicates that a more sophisticated model is required for the multivariate approach in order to take into account all the necessary information for a good reconstruction of the complex hadronic $t\bar{t}$ system. It might also hint at the fact that the NN erroneously learned to distinguish between the masses of the two top quarks.

This trend becomes even more obvious when looking at the mass distributions of the reconstructed W bosons in figure 8.4. Here, for the W boson originating from the top quark a quite sharp peak can be found in comparison to the W boson originating from the anti-top quark. This supports the finding that the NN can handle two and three decay products acceptably well, but is overchallenged when having to deal with 6 decay products. Furthermore, false information are extracted allowing the NN to assign different masses to the top quarks.

In contrast to figure 8.3, the mass distribution of the W boson originating from the top quark peaks at a value of about 80 GeV, thus the W-boson mass. However, the shapes related to the χ^2 reconstruction still exhibit high and sharp peaks with small tails, while the shapes related to the NN-based reconstruction are significantly broader. This supports the result found above.
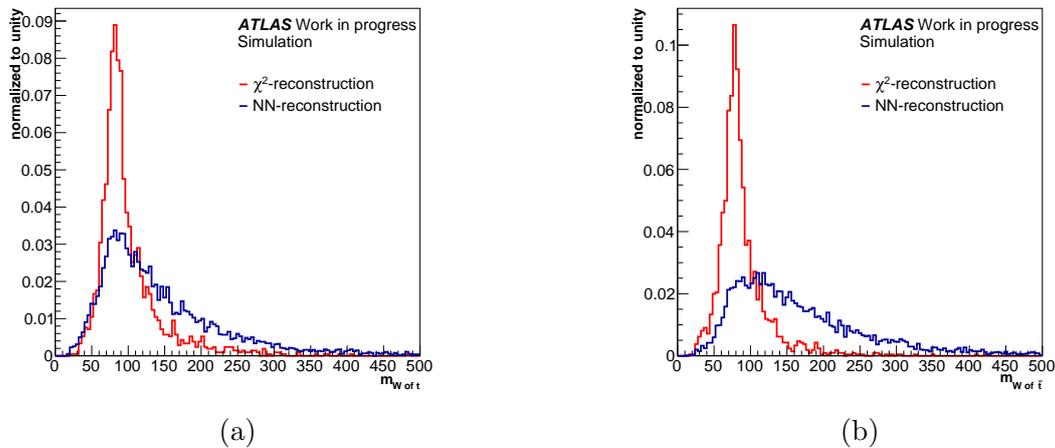
Figure 8.4: Normalized distributions of the mass of the W boson originating from the top quark (a) and originating from the anti-top quark (b) for the best hypotheses of the χ^2 reconstruction and the NN-based reconstruction.

Concluding, it can be retained that the minimum-χ^2 reconstruction performs significantly better than the NN-based reconstruction. It was found on the one hand that the best hypotheses selected by the χ^2 reconstruction exhibited a higher matched share than the hypotheses selected by the NN-based reconstruction. On the other hand, the hypotheses selected by the χ^2 method are able to better reproduce the information of the reconstructed $t\bar{t}$ system.

However, based on these findings, the NN-based reconstruction should not be rejected in general. In principle, the approach of using a multivariate model for reconstruction is very promising as the matched share for the quarklevel method, i.e. the input data for the MVA training, was found to be almost 50%, which is very high. It has also been found that one reconstructed object (top quark or W boson) is better reconstructed than the other one. Thus, the MVA model used for the multivariate reconstruction is simply not sophisticated enough. Applying a more sophisticated, well-performing MVA model on the 'high-quality' input data, the matched share of almost 50% of the input data provided by the quarklevel method could be increased. In consequence, a significantly better reconstruction performance than for the classic χ^2 method could be reached.

Chapter 9

Conclusion and outlook

In order to conduct a precise cross-section measurement of the associated production of a top-antitop-quark pair and a Z boson ($t\bar{t}Z$), a good reconstruction of the $t\bar{t}(Z)$ systems is of utmost importance. This is however quite challenging as it implies selecting the correct jets among the multiple physical objects produced in a high number of interactions.

That is why in this thesis two methods for the reconstruction of $t\bar{t}(Z)$ systems have been presented, investigated and compared: the standard minimum-χ^2 method and a self-developed multivariate (MVA) method using a Neural Network. The studies presented are based on Monte Carlo simulated data for the ATLAS experiment and are conducted in the context of the full Run-2 (140 fb^{-1}) analysis. They consider $t\bar{t}Z$ systems in which the top and the anti-top quark decay hadronically and the Z boson decays either in an electron-positron or muon-antimuon pair.

In the first part, the standard minimum-χ^2 reconstruction was the focus of investigation. This method calculates a χ^2 value for every possible combination of jets to form a $t\bar{t}$ system in each event. The combination/hypothesis with the minimal χ^2 value is then considered as the reconstructed $t\bar{t}$ system. In addition, this χ^2 value can also be used to separate the signal $t\bar{t}Z$ process from dominant background processes like the Z+jets and the dileptonic $t\bar{t}$ decay as confirmed in a first study.

The performance of the minimum-χ^2 reconstruction was analyzed by comparing the selected best χ^2 hypotheses to all hypotheses considered during reconstruction. This exhibited a better performance for the selected hypotheses as expected. The performance was then further quantified by applying a procedure called truth matching which tests the alignment of the χ^2-selected jets and the truth partons. In this context, three categories - 'matched', 'combinatorial' and 'non-matched'- have been defined based on two criteria: it is checked if the distance ΔR of the jet to the assigned truth parton is smaller than 0.3 and if the assigned truth parton is also the corresponding one. For $t\bar{t}$ systems of the matched category both criteria are fulfilled for all 6 selected jets. For the combinatorial category all 6 selected jets have a distance ΔR smaller than 0.3 to their partons, but at least one truth parton is not the corresponding one. In case at least one jet has a distance $\Delta R \geq 0.3$ to the assigned truth parton, the $t\bar{t}$ system and thus also the event is assigned

to the non-matched category. Applying this procedure to t$\bar{\text{t}}$Z events resulted in a share of 1.1%[1] of matched events. This is extremely low and indicates that the reconstruction of hadronic t$\bar{\text{t}}$ systems is not a trivial task.

Based on these findings detailed investigation studies were conducted which identified some challenges faced in reconstruction: the erroneous selection of gluon jets, the presence of switched jets and merged partons/jets as well as the mix of boosted and non-boosted topologies. As in this context the question arose whether the choice of the matching condition $\Delta R < 0.3$ was justified, further studies were conducted which validated it successfully. To improve the attainable matching performance, a phase-space adaption between reco and truth level was performed. For this purpose, a p_T cut as well as an η cut on all truth partons originating from the t$\bar{\text{t}}$ system were selected. In addition, the challenge of merged partons was addressed with an additional cut to exclude merged partons, i.e. partons with a distance $\Delta R < 0.8$ to each other, on truth level. All these cuts were grouped under the name 'improvement cuts' as their application resulted in an increase of the matched share from 1.6% to 5.3%[2]. Regarding the matchable share, i.e. the ratio of matched to the sum of matched and combinatorial events, an increase from 27.1% to 34.0% was achieved. As these cuts represent only exploratory choices, further optimization of the cuts is expected to lead to an additional increase in the matched and matchable shares.

To further improve the (χ^2-)reconstruction, the remaining challenges identified above need to be addressed: in order to reduce the contribution of gluon jets a quark-gluon tagger could be developed and applied. Regarding the switching between the light jets as well as the mix of boosted and non-boosted topologies, incorporating observables sensitive to these challenges in the reconstruction process is expected to reduce their impact. To address the switching between the light jets, the jet charge could be a suitable choice. To target the changing topology, the angular distances between the decay products of the top quarks could be included.

As improvement cuts are only able to address the identified reconstruction challenges up to a certain point, an alternative reconstruction approach based on an MVA method was developed and studied. The latter was chosen to be a Neural Network in this thesis. Its application consists of two steps: training and evaluation.

To conduct the training in the context of supervised learning, labeled input data were required. Therefore, a method to produce labeled input data had to be selected. In this context, three methods were developed: the Wtlevel method, the quarklevel method and the ctc method. Both the Wtlevel and the quarklevel method calculate a ΔR_{total} value for every possible combination to form a t$\bar{\text{t}}$ system in each event. This ΔR_{total} value is determined by considering the distances ΔR either between the 6 jets and their assigned partons or between the 4 reconstructed objects (two top quarks and two W bosons) on truth and

---

[1]for an event selection with ≥ 6 jets and ≥ 2 b-jets
[2]for an event selection of exactly 6 jets and exactly 2 b jets

reco level. The combination/hypothesis with the minimal ΔR_{total} value is then labeled as signal while all other hypotheses are labeled as background. In contrast, the ctc method assigns every possible combination to form a $t\bar{t}$ system to the three categories defined for the truth-matching process. Here, the signal is represented by hypotheses of the matched category while the combinatorial and non-matched categories form the background.

The performance of these three methods was compared among each other as well as for two different event selections, i.e. 'without improvement cuts' and 'with improvement cuts'. Based on this comparison, the quarklevel method in combination with the event selection 'with improvement cuts' was found to be the most suitable method for the goal of this thesis.

Driven by the huge number of background hypotheses provided by this method which exceeded the number of signal hypotheses by orders of magnitude, the former were studied in more detail. For this reason, the background category was split up into more detailed categories representing diverse physical scenarios: the flip categories and the (sub-)categories 1 to 6. Investigation of them laid out that the 0-flip and 1-flip categories were very signal-like and the merged category as well as the multi-flip categories were background-like.

As only a subset of the labeled input data could be used for training due to memory limitations, the training was conducted on the full signal and only on the most signal-like background categories, i.e. the (sub-)categories of the 0-flip and 1-flip categories. They were assumed to be most challenging for the Neural Network to distinguish from the signal.

Before finally conducting the training of the Neural Network also its architecture and suitable input variables had to be specified. For the input variables, observables covering different kinds of physical information (kinetic, angular, b-tag and helicity) were chosen. In this context, also 6 input variables especially targeting the identified reconstruction challenge of a changing topology were developed and included in the list of input variables. In particular, they are $\Delta R/\Delta\phi/\Delta\eta$ between the decay products of the top/anti-top quark and their event-wise sum. Regarding the NN architecture, a Neural Network with one hidden layer containing 500 neurons with a leaky ReLu activation function was selected based on its good training performance in terms of signal-background separation and the exhibited small amount of overtraining. Furthermore, the most important observables for the training process were identified: apart from the masses and the is-btagged observables, which were also used in the χ^2 reconstruction, the transverse momenta, the helicity properties and the angular distances between the decay products of the top/anti-top quark contributed significantly to the learning process.

In the second step of the application of the MVA method, the performance of this reconstruction approach was investigated by evaluating the trained Neural Network on unlabeled data. Investigation of the classification quality or rather the reconstruction performance, showed that the chosen Neural Network was not able to truthfully classify unlabeled hypotheses, even though the output-scores of the best hypotheses in the event, i.e. the reconstructed $t\bar{t}$ systems, were found to be extremely high.

One possible reason could have been the fact that not all hypotheses of the different back-

ground categories were included in the training due to memory limitations. It was striking that hypotheses belonging to the excluded categories were exactly the ones which could not be classified truthfully. Furthermore, also the huge number of possible configurations present in the reconstruction of hadronic $t\bar{t}$ systems was found to be challenging for the chosen Neural Network. While 3 decay products could be handled acceptably well, 6 decay products overchallenged the selected Neural Network. Moreover, it was found that the chosen Neural Network extracted too little and also erroneous information for classification. A reason for this could have been the low number of signal hypotheses available for training. This could also be regarded as an indication that a more sophisticated model is needed to achieve a good classification/reconstruction performance.

The final performance comparison of the two approaches exhibited a significantly better performance for the minimum-χ^2 reconstruction: the best hypotheses selected by the χ^2 reconstruction yielded a higher matched share and were also able to better reproduce the information of the reconstructed $t\bar{t}$ system.
Nevertheless, the NN-based reconstruction should not be rejected in general: It has also been found that one reconstructed object (top quark or W boson) of the $t\bar{t}$ system was better reconstructed than the other one. Thus, the model used for the multivariate reconstruction is obviously not sophisticated enough. Furthermore, by comparing the best NN hypotheses to all hypotheses before training, it was found that a lot of observable shapes stayed unchanged. This indicates that the training needs to be improved, so that the NN can extract enough information for classification. Thus, with application of a more sophisticated and well-performing Neural Network on the 'high-quality' input data, their matched share of almost 50% could be increased. In consequence, a significantly better reconstruction performance than for the classic χ^2 method could be realized.

To achieve this, there are various options. One is to provide sufficient memory capacity in order to be able to train on the whole dataset. This would assure that the Neural Network is exposed to all possible configurations and thus should know how to handle all of them. A challenge which could arise in this context is the highly imbalanced number of signal and background hypotheses. A way to circumvent this could be to still train on the whole signal, but randomly choose the background hypotheses from the different categories.

Another option could be to increase the number of signal hypotheses available for training. In this context, it would be reasonable to change the method which provides the labeled input data for the Neural Network training: Instead of using the quarklevel method in combination with the event selection 'with improvement cuts', it could be worth to consider also the ctc method in combination with the event selection 'without improvement cuts'. This would slightly decrease the reproduction quality of the input data, but would at the same time increase the number of available signal hypotheses for training by one order of magnitude. With the higher number of signal hypotheses available for training, also the full power of Neural Networks could be exploited: In general, deeper Neural Networks are known to be more powerful assuming that a large number of (signal) hypotheses

is available. Otherwise, significant overtraining will be observed as it was the case in this thesis for the discarded Neural Networks.

To support this, also the signal definition of the ctc method could be investigated further: it might also be reasonable to accept the combinatorial category or some parts of the latter as signal. This would also reduce the imbalance in the number of signal and background hypotheses.

A disadvantage of this approach could be that even more memory capacity would be required. A randomized selection of background hypotheses of the different background categories could be a possible solution.

If one wants to stay with the quarklevel method in combination with the event selection 'with improvement cuts', it could be worth thinking about possibilities to obtain a larger dataset in order to be able to increase the number of signal hypotheses. One option could be to use smeared truth-MC data for training. If the number of signal hypotheses is then large enough, an approach worth considering could be to accept only signal hypotheses with a ΔR_{total} value ≤ 1 as signal. The reason is that for this value a dip was found which clearly separates well- from badly-reconstructed $t\bar{t}$ systems. This would improve the quality of the input data even further and might also yield promising results in combination with the event selection 'without improvement cuts'.

In addition also the improvement cuts could be optimized further to better suit the needs arising in the context of producing labeled input data.

Moreover, it could also be worth to consider 'multiclassification', i.e. simultaneous classification into more than two classes. Here, in contrast to the 'classic' Neural Network, there is no longer only one output node stating if the hypothesis is signal (output-score ≥ 0.5) or background (output-score < 0.5). Instead, there is one output node per defined class which indicates for each input value with which probability it belongs to the particular class. A challenge which could arise in this context is again related to the low number of signal hypotheses. As extracting enough information was already challenging for the classic Neural Network, it would now be tasked with performing multiple classifications related to the different output classes.

To overcome this, an alternative idea is to train different Neural Networks for the separate background categories. This would have the advantage that all categories could be included in the training while at the same time, the Neural Network would not be overchallenged by having to deal with all possible configurations at the same time. The distinct classes could be trained separately and hence could be better learned in the different Neural Networks. Furthermore also the imbalanced number of signal and background hypotheses would be addressed. In the end, the output of the different Neural Networks could be used in combination for classification/reconstruction: either another Neural Network could be used which obtains the outputs of the different Neural Networks as an input or one could also combine the different output-scores using a likelihood method.

A challenge which could arise in this context is that the training will be very costly in terms of time. In this case it could be a solution to use GPU computers, which accelerate the processing.

Concluding, it can be retained that using a multivariate approach for the reconstruction of hadronically decaying $t\bar{t}$ systems in $t\bar{t}$Z events has been found promising. Even though the χ^2 reconstruction performed better in direct comparison with the reconstruction using the chosen Neural Network, reconstruction based on a more sophisticated and well-performing multivariate method has high potential to achieve a better performance.

Bibliography

[1] CDF Collaboration, "Observation of Top Quark Production in p$\bar{\text{p}}$ Collisions with the Collider Detector at Fermilab," *Phys. Rev. Lett. 74 (1995) 2626, arXiv: hep-ex/9503002*, 1995.

[2] D0 Collaboration, "Observation of the Top Quark," *Phys. Rev. Lett. 74 (1995) 2632, arXiv: hep-ex/9503003*, 1995.

[3] Weinberg, S., "The making of the Standard Model," *Eur. Phys. J. C34 (2004) 5*, 2004.

[4] Weinberg, S., "Conceptual foundations of the unified theory of weak and electromagnetic interactions," *Nobel Lecture, https://www.nobelprize.org/uploads/2018/06/weinberg-lecture.pdf*, 1979.

[5] Gross, D. J. and Wilczek, F., "Asymptotically Free Gauge Theories," *Phys. Rev. D8 (1973) 3633*, 1973.

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

[7] Plitzer, H. D., "Asymptotic freedom: An approach to strong interactions," *Phys. Rept. 14/4 (1974) 129*, 1974,.

[8] Thomson, M., *Modern particle physics.* Cambridge University Press, 2013. ISBN: 978-1-107-03426-6.

[9] Perkins, D. H., *Introduction to High Energy Physics, 4th edition.* Cambridge University Press, 2014. ISBN: 978-0-521-62196-0.

[10] Griffiths, D., *Introduction to Elementary Particles, Second Revised Edition.* WILEY-VCH Verlag GmvH & Co. KGaA, 2010. ISBN: 978-3-527-40601-2.

[11] Particle Data Group, "The Review of Particle Physics." http://pdg.lbl.gov/, accessed 20th April 2019.

[12] Kröninger, K., Meyer, A. B. and Uwer, P., "Top-Quark Physics at the LHC," *arXiv:1506.02800 [hep-ex]* , 2015.

[13] Moch, S. and Uwer, P., "Theoretical status and prospects for top-quark pair production at hadron colliders," *Phys.Rev. D78 (2008) 034003, arXiv:0804.1476 [hep-ph]*.

[14] Particle Data Group, "The Review of Particle Physics, Chapter 18: Structure functions, archived 2013 Edition." `http://pdg.lbl.gov/2013/reviews/rpp2013-rev-structure-functions.pdf`.

[15] Demina, R. and Thomson, E. J., "Top Quark Properties and Interactions." `https://hep.physics.utoronto.ca/WilliamTrischuk/courses/phy189/annurev.nucl.58.110707.171224.pdf`, accessed 18th April 2019.

[16] Wagner, W., "Top quark physics in hadron collisions," *Rep. Prog. Phys. 68 (2005) 2409, arXiv:hep-ph/0507207*, 2005.

[17] Povh, B. et al., *Teilchen und Kerne - Eine Einführung in die physikalischen Konzepte, 9. Auflage*. Springer Spektrum, 2014. ISBN: 978-3-642-37821-8.

[18] Czarneck, A., "Helicity fractions of W bosons from top quark decays at NNLO in QCD," *Phys.Rev. D81 (2010) 111503, arXiv:1005.2625 [hep-ph]*.

[19] ATLAS Collaboration, "Measurement of the W boson polarisation of $t\bar{t}$ events from pp collisions at $\sqrt{s} = 8$ TeV in the lepton + jets channel with ATLAS," *Eur. J. C 77 (2017) 264, arXiv:1612.02577v3 [hep-ex]*.

[20] CMS Collaboration, "Measurement of the W-boson helicity in top-quark decays from $t\bar{t}$ production in lepton+jets events in pp collisions at $\sqrt{s} = 7$ TeV," *JHEP 10 (2013) 167, arXiv:1308.3879v2 [hep-ex], CMS-TOP-11-020, CERN-PH-EP-2013-133*, 2013.

[21] CMS Collaboration, "Measurement of the W boson helicity fractions in the decays of top quark pairs to lepton + jets final states produced in pp collisions at $\sqrt{s} = 8$ TeV," *Phys. Lett. B 762 (2016) 512, arXiv:1605.09047v2 [hep-ex]*.

[22] ATLAS Collaboration, "Standard Model Total Production Cross Section Measurements." `https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SM/ATLAS_a_SMSummary_TotalXsect/ATLAS_a_SMSummary_TotalXsect.png`, accessed 20th April 2019.

[23] CMS Collaboration, "Observation of top quark pairs produced in association with a vector boson in pp collisions at $\sqrt{s} = 8$ TeV," *JHEP 01 (2016) 096, arXiv:1510.01131 [hep-ex], CMS-TOP-14-021, CERN-PH-EP-2015-248*, 2016.

[24] ATLAS Collaboration, "Measurement of the $t\bar{t}$W- and $t\bar{t}$Z-production cross sections in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector," *JHEP 11 (2015) 172, arXiv:1509.05276 [hep-ex], CERN-PH-EP-2015-208*, 2015.

[25] ATLAS Collaboration, "Measurement of the t$\bar{\text{t}}$Z and t$\bar{\text{t}}$W cross sections in proton-proton collisions at $\sqrt{\text{s}} = 13$ TeV with the ATLAS detector," *arXiv:1901.03584v1 [hep-ex]*, 2019.

[26] CMS Collaboration, "Measurement of top quark pair production in association with a Z boson in proton-proton collisions at $\sqrt{\text{s}} = 13$ TeV,"

[27] CMS Collaboration, "Observation of t$\bar{\text{t}}$H Production," *Phys. Rev. Lett. 120 (2018) 231801, arXiv:1804.02610v2 [hep-ex], CMS-HIG-17-035, CERN-EP-2018-064*, 2018.

[28] ATLAS Collaboration, "Measurement of the t$\bar{\text{t}}$Z and t$\bar{\text{t}}$W production cross sections in multilepton final states using 3.2 fb^{-1} of pp collisions at $\sqrt{\text{s}} = 13$ TeV with the ATLAS detector," *Eur. Phys. J. C77 (2017) 49, arXiv:1609.01599v2 [hep-ex]*, 2017.

[29] ATLAS Collaboration, "Measurement of the t$\bar{\text{t}}$Z and t$\bar{\text{t}}$W cross sections in proton-proton collisions at $\sqrt{\text{s}} = 13$TeV with the ATLAS detector," *arXiv:1901.03584v1 [hep-ex]*, 2019.

[30] Brüning, O. et al., "LHC Design Report," *CERN-003-V-1*, 2004.

[31] ATLAS Inner Detector Community, "ATLAS Inner Detector Technical Design Report – Volume I," *CERN-LHCC-97-016, ATLAS-TDR-4*, 1997.

[32] ATLAS Collaboration, "ATLAS muon spectrometer: Technical Design Report," *CERN-LHCC-97-022, ATLAS-TDR-10*, 1997.

[33] ATLAS Collaboration, "ALTAS calorimeter performance: Technical Design Report," *CERN-LHCC-96-040, ATLAS-TDR-1*, 1996.

[34] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider," *JINST3 (2008) S08003,*, 2008.

[35] CERN Document Server, "About CERN," 2012. `http://cds.cern.ch/record/1997225`, accessed 20th April 2019.

[36] Lopienska, E., "LHC Guide," 2017. `https://cds.cern.ch/record/2255762?ln=de;CERN-Brochure-2017-002-Eng`, accessed 10th April 2019.

[37] Evans, L. and Bryant, P., "LHC Machine," *JINST 3 (2008) S08001*, 2008.

[38] Ruiz-Martinez, A. on behalf of the ALTAS Collaboration, "The RUN-2 ATLAS Trigger System," 2016. *J.Phys. Conf. Ser.762no. 1, (2016) 012003 , ATL-DAQ-PROC-2016-003*.

[39] CERN Document Server, "Overall view of the LHC." `http://cds.cern.ch/record/1708849?ln=de`, accessed 14th April 2019.

[40] Lamprecht, M., *Studien zu Effizienz und Akzeptanz des ATLAS-Myontriggers mit simulierten Messdaten*. Ludwig-Maximilians-Universität München. Diplomarbeit, https://www.etp.physik.uni-muenchen.de/publications/theses/download/dipl_mlamprecht.pdf.

[41] CERN Document Server, "Computer generated image of the whole ATLAS detector." https://cds.cern.ch/record/1095924?ln=de, accessed 14th April 2019.

[42] Schleper, P., "Teilchenphysik für Fortgeschrittene - Notizen zur Vorlesung im Wintersemester 2010-11." http://www.desy.de/~schleper/lehre/TeilchenFortgeschrittene/WS_2010_11/TeilchenFortgeschrittene.pdf, accessed 17th April 2019.

[43] Andersson, B., "Parton fragmentation and string dynamics," *Phys. Rep. 97 (1983) 31*, 1983.

[44] Webber, B. R., "A QCD model for jet fragmentation including soft gluon interference," *Nucl. Phys. B 238/3 (1984) 492*, 1984.

[45] Agostinelli, S. et al., "Geant4 - a simulation toolkit," *Nucl. Instrum. Meth. A 506 (2003) 250, CERN-IT-2002-003, SLAC-PUB-9350*, 2002.

[46] ATLAS Collaboration, "The ATLAS Simulation Infrastructure," *Eur. Phys. J. C70 (2010) 823, arxiv: 1005.4568 [physics.ins-det]*, 2010.

[47] Seymour, M. H. and Marx, M., "Monte Carlo Event Generators," *arXiv:1304.667v1 [hep-ph]*, 2013.

[48] Sjöstrand, T., "Monte Carlo Tools," *arXiv:0911.5286 [hep-ph]*, 2009.

[49] Alwall, J. et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations," *JHEP 07 (2014) 079, arXiv:1405.0301v2 [hep-ph]*, 2014.

[50] Hamilton, K., "A positive-weight Next-to-Leading Order Monte Carlo simulation for Higgs boson production," *JHEP 0904 (2009) 116, arXiv:0903.4345 [hep-ph]*, 2009.

[51] Sjöstrand, T. et al., "An Introduction to PYTHIA 8.2," *Computer Phys. Comm. 191 (2015) 159, arXiv:1410.3012 [hep-ph]*, 2014.

[52] Gleisberg, T. et al., "Event generation with SHERPA 1.1," *JHEP 0902 (2009) 007; arXiv:0811.4622 [hep-ph]*, 2009.

[53] Ball, R. D. et al., "Parton distributions for the LHC run II," *JHEP 04 (2015) 040, arXiv:1410.8849 [hep-ph]*, 2015.

[54] ATLAS Collaboration, "ATLAS Run 1 Pythia8 tunes," *ATL-PHYS-PUB-2014-021*, 2014.

[55] Ball, R. D. et al., "Parton distributions with LHC data," *Nucl. Phys. B 867 (2013) 244, arxiv: 1207.1303 [hep-ph]*, 2013.

[56] Frixione, S., Ridolfi, G. and Nason, P., "A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction," *JHEP 09 (2007) 126, arxiv: 0707.3088 [hep-ph]*, 2007.

[57] Nason, P., "A new method for combining NLO QCD with shower Monte Carlo algorithms," *JHEP 11 (2004) 040, arxiv: hep-ph/0409146*, 2004.

[58] Frixione, S., Nason, P. and Oleari, C., "Matching NLO QCD computations with parton shower simulations: the POWHEG method," *JHEP 11 (2007) 070, arxiv: 0709.2092 [hep-ph]*, 2007.

[59] Alioli, S., Nason, P., Oleari, C. and Re, E., "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX," *JHEP 043 (2010) 1006, arXiv:1002.2581 [hep-ph]*, 2010.

[60] Lange, D. J., "The EvtGen particle decay simulation package," *Nucl.Instrum.Meth. A 462 (2001) 152*, 2001.

[61] Sherpa Team, "Sherpa Manual Version 2.2.1." `https://sherpa.hepforge.org/doc/SHERPA-MC-2.2.1.html`, accessed 18th April 2019.

[62] ATLAS Collaboration, "Performance of primary vertex reconstruction in proton-proton collisions at $\sqrt{s} = 7$ TeV in the ATLAS experiment," *ATLAS-CONF-2010-069*, 2010.

[63] ATLAS Collaboration, "Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data," *ATLAS-CONF-2016-024*, 2016.

[64] ATLAS Collaboration, "Muon reconstruction performance of the ATLAS detector in proton-proton collision data at $\sqrt{s} = 13$ TeV," *Eur. Phys. J. C76 (2016) 292arXiv:1603.05598 [hep-ex]*, 2016.

[65] ATLAS Collaboration, "Measurement of jet charge in dijet events from $\sqrt{s} = 8$ TeV pp collisions with the ATLAS detector," *Phys. Rev. D93 (2016) 052003, arXiv:1509.05190v2 [hep-ex]*, 2016.

[66] Cacciari, M., Salam, G. P. and Soyez, G., "The anti-k_t jet clustering algorithm," *JHEP 04 (2008) 063, arXiv:0802.1189 [hep-ph]*, 2008.

[67] Cacciari, M., Salam, G. P. and Soyez, G., "The catchment area of jets," *JHEP 04 (2008) 005, arXiv:0802.1188 [hep-ph]*, 2008.

[68] ATLAS Collaboration, "Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run1," *Eur. Phys. J. C77 (2017) 490, arXiv:1603.02934 [hep-ex]*, 2017.

[69] Salam, G. P., "Towards Jetography," *arXiv:0906.1833 [hep-ph], Eur. Phys. J. C67 (2010) 637-686*, 2010.

[70] ATLAS Collaboration, "Tagging and suppression of pileup jets with the ATLAS detector," *ATLAS-CONF-2014-018*, 2014.

[71] ATLAS Collaboration, "Performance of jet substructure techniques for large-R jets in proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector," *JHEP 09 (2013) 076, arXiv:1306.4945v1 [hep-ex]*, 2013.

[72] ATLAS Collaboration, "Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS detector at $\sqrt{s} = 13$ TeV," *ATL-PHYS-PUB-2015-015*, 2015.

[73] ATLAS Collaboration, "Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run," *ATL-PHYS-PUB-2016-012*, 2016.

[74] private conversation.

[75] Adams, D. et al., "Recommendations of the Physics Objects and Analysis Harmonisation Study Groups 2014," *ATL-PHYS-INT-2014-018*, 2016.

[76] ATLAS Collaboration, "Measurement of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV," *JHEP 08 (2018) 89, arXiv:1805.01845v2 [hep-ex]*, 2018.

[77] Goodfellow, I., Bengio, Y. and Courville, A., *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[78] Buduma, N., *Fundamentals of Deep Learning - Designing Next-Generation Machine Intelligence Algorithms*. O'Reilly Media, Inc., 2017. ISBN:978-1-491-92561-4.

[79] Alpaydin, E., *Introduction to Machine Learning, 3rd edition*. The MIT Press, 2014. ISBN: 978-0-262-02818-9.

[80] Wartala, R., *Praxiseinstieg Deep Learning - mit Python, Caffe, TensorFlow und Spark eigene Deep-Learning-Anwendungen erstellen, 1. Auflage*. O'Reilly, 2018.

[81] Behnke, O. et al., *Data Analysis in High Energy Physics - A Practical Guide to Statistical Methods*. WILEY-VCH GmbH & Co. KGaA, 2013. ISBN:978-3-527-41058-3.

[82] Skansi, S., *Introduction to Deep Learning - From Logical Calculus to Artificial Intelligence*. Springer International Publishing AG, 2018. ISBN: 978-3-319-73003-5.

[83] Rosenblatt, F., "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychol. Rev. 65 (1958) 386*, 1958.

[84] Lippmann, R., "An introduction to computing with neural nets," *IEEE ASSP Magazine 4/2 (1987) 4*, 1987.

[85] Rumelhart, D. E. and McClelland, J. L., *Eds. Parallel Distributed Processing, Vol. 1, Cambridge.* MIT Press, 1986. ISBN: 9780262680530.

[86] Kohonen, T., "An introduction to neural computing," *Neural Networks 1/1 (1988) 3*, 1988.

[87] Kohonen, T., *Self-Organization and Associative Memory.* Springer-Verlag, 1989. ISBN: 978-3-642-88163-3.

[88] Robbins, H. and Monro, S., "A Stochastic Approximation Method," *Ann. Math. Statist. 22/3 (1951) 400*, 1951.

[89] Kiefer, J. and Wolfowitz, J., "Stochastic Estimation of the Maximum of a Regression Function," *Ann. Math. Statist. 23/3 (1952) 462*, 1952.

[90] Bottou, L., Curtis, F. E. and Nocedal, J., "Optimization Methods for Large-Scale Machine Learning," *arXiv:1606.04838 [stat.ML]*.

[91] Kingma, D. P. and Ba, J., "Adam: A Method for Stochastic Optimization," *arXiv:1412.6980 [cs.LG]*, 2014.

[92] Linnainmaa, S., "Taylor expansion of the accumulated rounding error," *BIT Numerical Mathematics 16/2 (1976) 146*, 1976.

[93] Srivastava, N. et al., "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research 15 (2014) 1929*, 2014.

[94] Smith, V. et al., "L1-Regularized Distributed Optimization: A Communication-Efficient Primal-Dual Framework," *arXiv:1512.04011 [cs.LG]*, 2016.

[95] Kolmogorov, A. N., "Sulla determinazione empirica di una legge di distribuzione," *Inst. Ital. Attuari*, 1933.

[96] Smirnov, N. V., "On the estimation of the discrepancy between empirical curves of distribution for two independent samples," *Bull. Math. Univ. Moscou*, 1939.

[97] Lopez-Paz, D. and Oquab, M., "Revisiting Classifier Two-Sample Tests," *arXiv:1610.06545v4 [stat.ML]*.

[98] Hoecker, A. et al., "TMVA 4: Toolkit for Multivariate Data Analysis," *arXiv:physics/0703039 [physics.data-an]*.

[99] ATLAS Data Quality Monitoring, "Good Runs Lists data 2016." `https://atlasdqm.web.cern.ch/atlasdqm/grlgen/All_Good/data16_13TeV.periodAllYear_DetStatus-v89-pro21-01_DQDefects-00-02-04_PHYS_StandardGRL_All_Good_25ns.xml`, accessed 25th April 2019.

[100] ATLAS Data Quality Monitoring, "Good Runs List data 2015." `https://atlasdqm.web.cern.ch/atlasdqm/grlgen/All_Good/data15_13TeV.periodAllYear_DetStatus-v89-pro21-02_Unknown_PHYS_StandardGRL_All_Good_25ns.xml`, accessed 25th April 2019.

[101] Markovec, N., "Top xAOD Start-Guide." `https://twiki.cern.ch/twiki/bin/view/AtlasProtected/TopxAODStartGuide`, accessed 24th April 2019.

[102] Markovec, N., "How to clean jets." `https://twiki.cern.ch/twiki/bin/view/Sandbox/EgammaPhysCalib`, accessed 24th April 2019.

[103] Particle Data Group, "The Review of Particle Physics, chapter 18: Structure Functions, archived 2017 Edition." `http://pdg.lbl.gov/2017`, accessed 25th April 2019.

[104] Particle Data Group, "Monte Carlo particle numbering scheme." `http://pdg.lbl.gov/2007/reviews/montecarlorpp.pdf`, accessed 7th April 2019.

[105] Buckley, A. and Pollard, C., "QCD-aware partonic jet clustering for truth-jet flavour labelling," *arXiv:1507.00508 [hep-ph]*.

[106] CMS Collaboration, "Search for the associated production of a Higgs boson with a single top quark in proton-proton collisions at $\sqrt{s} = 8$ TeV," *JHEP 06 (2016) 177,arXiv:1509.08159 [hep-ex]*, 2016.

[107] CMS Collaboration, "Search for $t\bar{t}H$ production in the $H \rightarrow b\bar{b}$ decay channel with leptonic $t\bar{t}$ decays in proton-proton collisions at $\sqrt{s}= 13$ TeV," *JHEP 03 (2019) 026, arXiv:1804.03682v2 [hep-ex]*, 2019.

[108] Chwalek, T., "Messung der W-Boson-Helizitätsanteile in Top-Quark-Zerfällen mit dem CDF II Experiment und Studien zu einer frühen Messung des $t\bar{t}$-Wirkungsquerschnitts mit dem CMS Experiment, Doktorarbeit," 2012. `http://www-ekp.physik.uni-karlsruhe.de/%7Ethesis/data/iekp-ka2010-5.pdf`, accessed 6th April 2019.

[109] Chollet, F. and et al., "Keras," 2015. `https://keras.io`, accessed 30th March 2019.

[110] Abadi, M. and et al., "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems," 2015. software available from: `https://www.tensorflow.org/`.

[111] Handl, D., "stop1l-MLkit," 2019. `https://gitlab.physik.uni-muenchen.de/David.Handl/stop1l-MLkit`.

Appendix A

Additional plots for the investigation of the background categories
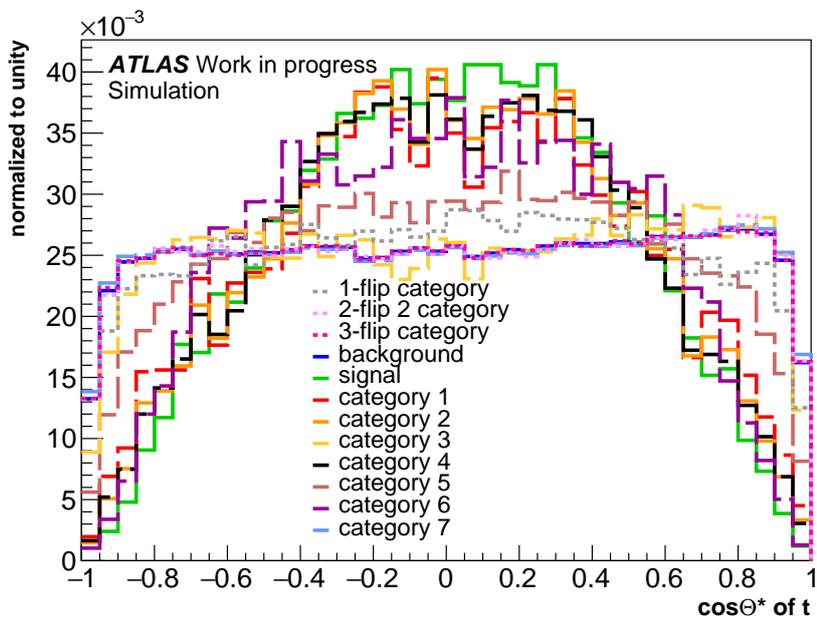
Figure A.1: Normalized distribution of the cosine of the helicity angle of the top quark $\cos(\Theta^*)$ related to the different backgroundmcategories defined in section 7.2.
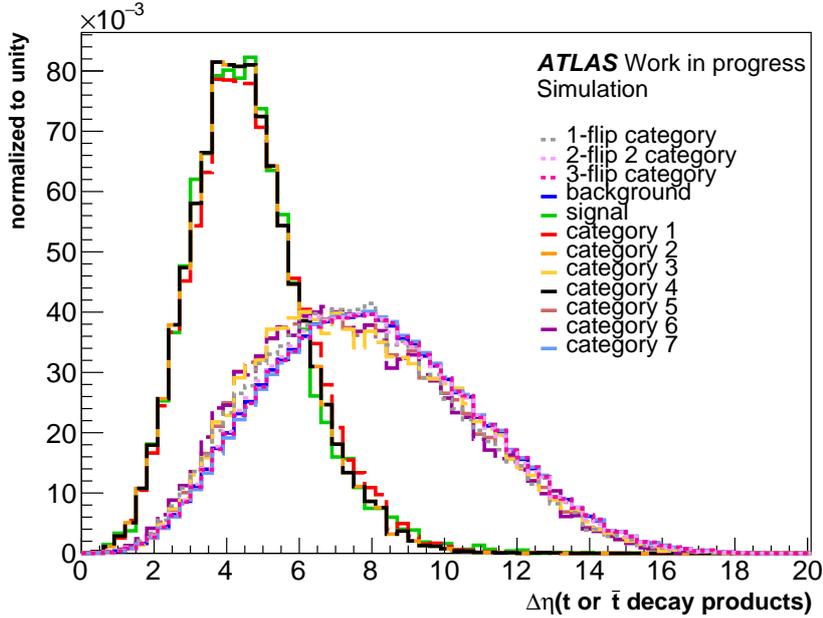
Figure A.2: Normalized distribution of the event-wise sum of the $\Delta\eta$ between the decay products of the top- and anti-top quark (defined in section 7.3) related to the different background categories defined in section 7.2.
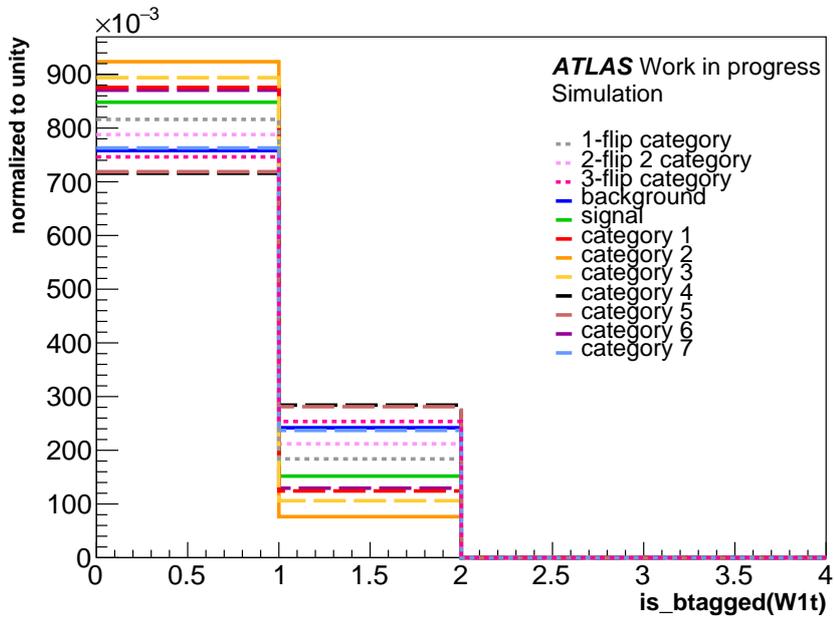
Figure A.3: Normalized distribution of is-btagged(W1t) related to the different background categories defined in section 7.2.
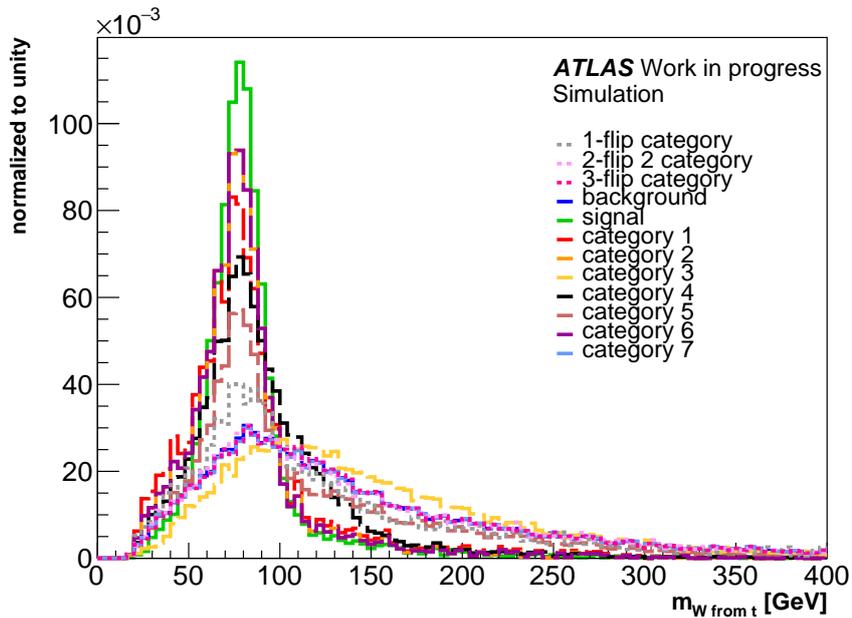
Figure A.4: Normalized distribution of the W-boson mass originating from the top quark $m_{W \text{ from } t}$ related to the different background categories defined in section 7.2.
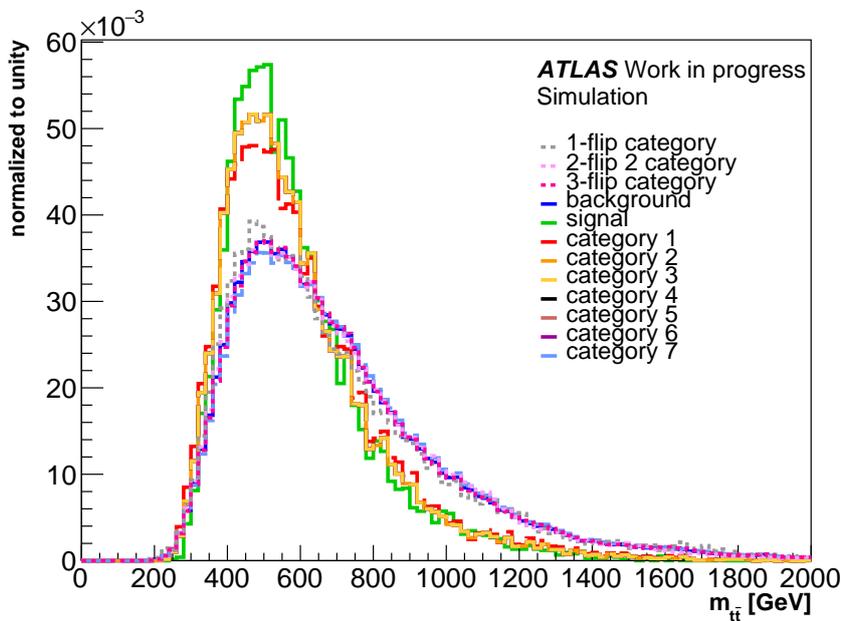
Figure A.5: Normalized distribution of the mass of the top-anti-top-quark system $m_{t\bar{t}}$ related to the different background categories defined in section 7.2.
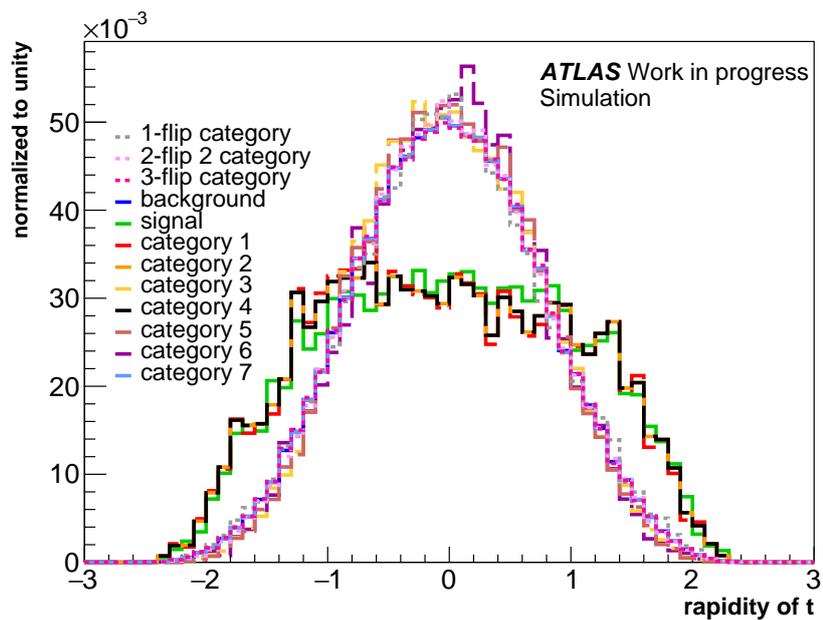
Figure A.6: Normalized distribution of the rapidity of the top quark related to the different background categories defined in section 7.2.

Appendix B

Complete variable ranking

	sum of squared weights	sum of absolute weights	gradients
1.	$m_{t\bar{t}}$	$m_{t\bar{t}}$	$m_{t\bar{t}}$
2.	is-btagged($\bar{b}$)	is-btagged($\bar{b}$)	$m_{W\text{ from }t}$
3.	is-btagged(b)	is-btagged(b)	m_t
4.	$p_T(W1\bar{t})$	$p_T(W1\bar{t})$	p_T(down-type quark t)
5.	is-btagged($W1\bar{t}$)	is-btagged($W1\bar{t}$)	$p_T(t)$
6.	is-btagged(W2t)	$p_T(t)$	$p_T(W2\bar{t})$
7.	$p_T(W1t)$	$p_T(W1t)$	$p_T(\bar{t})$
8.	$p_T(b)$	$p_t(\bar{b})$	$p_T(W1\bar{t})$
9.	$p_T(\bar{b})$	is-btagged(W2t)	$p_T(W1t)$
10.	$\Delta\eta$(decay products of t)	$p_T(b)$	$m_{W\text{ from }\bar{t}}$
11.	$p_T(t)$	is-btagged($W2\bar{t}$)	p_T(W from t)
12.	$\Delta\eta$(decay products of $\bar{t}$)	$p_T(W2t)$	p_T(down-type quark $\bar{t}$)
13.	is-btagged($W2\bar{t}$)	$p_T(W2\bar{t})$	$p_T(W2t)$
14.	$p_T(W2t)$	m_t	$p_T(\bar{b})$
15.	$m_{\bar{b}}$	$\Delta\eta$(t or $\bar{t}$ decay)	$m_{\bar{t}}$
16.	$\cos(\Theta^*)(\bar{t})$	$\Delta\eta$(decay products $\bar{t}$)	$p_T(t\bar{t})$
17.	$p_T(W2\bar{t})$	$p_T(\bar{t})$	p_T(W of $\bar{t}$)
18.	p_T(down-type quark t)	$\Delta\eta$(decay products t)	m_{W1t}
19.	$\Delta\eta$(t or $\bar{t}$ decay)	sum p_T(down-type quarks)	$p_T(b)$
20.	m_b	p_T(down-type quark t)	m_b
21.	$p_T(\bar{t})$	$m_{W\text{ of }\bar{t}}$	m_{W2t}
22.	m_t	$m_{W\text{ of }t}$	$m_{W1\bar{t}}$
23.	$m_{W\text{ of }\bar{t}}$	$m_{\bar{b}}$	$\Delta\eta$(t or $\bar{t}$ decay)
24.	ΔR(decay products t)	$m_{\bar{t}}$	$m_{\bar{b}}$
25.	p_T(W of t)	p_T(W of t)	$\Delta\eta$(decay products of $\bar{t}$)
26.	$m_{\bar{t}}$	$\cos(\Theta^*)(\bar{t})$	ΔR (t or $\bar{t}$ decay)

27.	$\cos(\Theta^*)(t)$	p_T(down-type quark $\bar{t}$)	sum p_T(down-type quarks)
28.	p_T(down-type quark $\bar{t}$)	ΔR(decay products $\bar{t}$)	$m_{W2\bar{t}}$
29.	$\Delta\phi$(t or $\bar{t}$ decay)	$\Delta\phi$(t or $\bar{t}$ decay)	$\Delta\phi$(decay products t)
30.	is-btagged(W1t)	is-btagged(W1t)	ΔR(decay products $\bar{t}$)
31.	rapidity(W of $\bar{t}$)	m_b	$\Delta\eta$(decay products t)
32.	sum p_T(down-type quarks)	$\Delta\phi$(decay products $\bar{t}$)	ΔR(decay products t)
33.	$m_{W \text{ of } t}$	rapidity(W of $\bar{t}$)	$\cos(\Theta^*)(t)$
34.	$\Delta\phi$(decay products $\bar{t}$)	$\cos(\Theta^*)(t)$	$\cos(\Theta^*)(\bar{t})$
35.	m_{W2t}	ΔR(t or $\bar{t}$ decay)	$\Delta\phi$(t or $\bar{t}$ decay)
36.	ΔR(decay products $\bar{t}$)	m_{W2t}	rapidity(W of $\bar{t}$)
37.	$m_{W1\bar{t}}$	ΔR(decay products $\bar{t}$)	rapidity(t)
38.	rapidity(t)	p_T(t$\bar{t}$)	rapidity(W of t)
39.	m_{W1t}	$m_{W1\bar{t}}$	is-btagged(W1$\bar{t}$)
40.	ΔR(t or $\bar{t}$ decay)	$m_{W2\bar{t}}$	is-btagged(W2$\bar{t}$)
41.	p_T(t$\bar{t}$)	p_T(W of $\bar{t}$)	rapidity(t$\bar{t}$)
42.	rapidity($\bar{t}$)	m_{W1t}	is-btagged(b)
43.	p_T(W of $\bar{t}$)	rapidity(t)	is-btagged($\bar{b}$)
44.	$m_{W2\bar{t}}$	$\Delta\phi$(decay products t)	is-btagged(W1t)
45.	$\Delta\phi$(decay products t)	rapidity($\bar{t}$)	rapidity($\bar{t}$)
46.	rapidity(t$\bar{t}$)	rapidity(W of t)	is-btagged(W2t)
47.	rapidity(W of t)	rapidity(t$\bar{t}$)	$\Delta\phi$(decay products $\bar{t}$)

Appendix C

List of discarded Neural Networks

neurons per layer	128	356	356,128	128, 128	500
ncycles	100	100	100	100	100
batchSize	32	32	32	32	32
dropout	0.5	0.5	0.5	0.5	0.5
optimizer	adam	adam	adam	adam	adam
activation	ReLu	ReLu	ReLu	ReLu	ReLU
initializer	normal	normal	normal	normal	normal
regularizer	0.01	0.01	0.01	0.01	0.01
classWeight	balanced	balanced	balanced	balanced	balanced
learningRate	0.001	0.001	0.001	0.001	0.001
decay	0.0	0.0	0.0	0.0	0.0
ROC	0.97 (0.96)	0.96(0.96)	0.96 (0.96)	0.96 (0.96)	0.99 (0.98)
accuracy	86.62%	90.77%	83.01%	87.82%	88.11%
loss	32.23%	26.7%	136.98%	57.23%	36.99%

neurons per layer	128	356, 128	356	500	800
ncycles	100	100	100	100	100
batchSize	32	32	32	32	32
dropout	0.5	0.5	0.5	0.5	0.5
optimizer	gradient descent	adam	adam	adam	adam
activation	ReLu	ReLu	ReLU	ReLu	ReLu
initializer	normal	normal	normal	normal	normal
regularizer	0.01	0.01	0.01	0.1	0.01
classWeight	balanced	balanced	balanced	balanced	balanced
learningRate	0.001	0.001	0.001	0.001	0.001
decay	0.0	0.0	0.0	0.0	0.0
ROC	0.98 (0.97)	0.98 (0.97)	0.97 (0.98)	0.97 (0.97)	0.98(0.97)
accuracy	93.26%	91.22%	91.79%	91.94%	91.87%
loss	16.59%	21.84%	20.62%	20.52%	20.48%

neurons per layer	500	500	128	128
ncycles	100	100	100	100
batchSize	32	32	32	32
dropout	0.5	0.5	0.5	0.5
optimizer	adam	adam	adam	adam
activation	ReLu	ReLu	sigmoid	tanh
initializer	normal	normal	normal	normal
regularizer	0.1	0.001	0.01	0.01
classWeight	balanced	balanced	balanced	balanced
learningRate	0.001	0.001	0.001	0.001
decay	0.0	0.0	0.0	0.0
ROC	0.5 (0.5)	0.98 (0.96)	0.98 (0.97)	0.98(0.97)
accuracy	99.81%	93.62%	93.53%	91.74%
loss	3.04%	16.07%	16.82%	20.94%

Acknowledgments

At this point I want to say a big thank you for all the support I experienced during my master thesis.

In this context, there are some people to which I want to contribute a special thanks:

- Prof. Dr. Dorothee Schaile for giving me the opportunity to write my master thesis at her chair

- PD Dr. Jeannine Wagner-Kuhr for the great supervision and a lot of insightful discussions

- Prof. Dr. Otmar Biebel for stepping in as supervisor, always taking the time for long discussions and for a lot of motivating comments

- Florian for dealing with loads of questions

- the 2LOS-analysis team for their support and a lot of great ideas

On top of that, I want to thank the complete group at the chair for welcoming me with open arms and always providing me with support.

Zu guter Letzt ein riesiges Dankeschön auch an meine Eltern und meine Schwester für all die Unterstützung!

Selbständigkeitserklärung

Ich versichere hiermit, die vorliegende Arbeit mit dem Titel

Development of a multivariate reconstruction algorithm for
hadronic $t\bar{t}$ systems in $t\bar{t}Z$ events with $Z \rightarrow ee/\mu\mu$ using the ATLAS detector

selbständig verfasst zu haben und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet zu haben.

Nina Wenke

München, den 08. Mai 2019