GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Fakultät für
Physik

Master's Thesis

Messung des Wirkungsquerschnitts von $t\bar{t}Z$ mit zwei Neutrinos im Endzustand bei 13 TeV am ATLAS-Detektor

A measurement of the cross section of $t\bar{t}Z$ with two neutrinos in the final state at 13 TeV with the ATLAS detector

prepared by

Marie Reinecke

from Alfeld (Leine)

at the II. Physikalisches Institut

Thesis number: II.Physik-UniGö-MSc-2018/02

Thesis period: 8th January 2018 until 20th July 2018

First referee: Prof. Dr. Arnulf Quadt

Second referee: Prof. Dr. Stan Lai

Abstract

Die assoziierte Produktion eines Top-Antitop-Paares und eines Z-Bosons konnte erst in jüngster Zeit beobachtet werden und es existieren keine Wirkungsquerschnitte für den Zerfallskanal des Z-Bosons in zwei Neutrinos. Es wird gefordert, dass das Top-Antitop-Paar in dieser Analyse vollhadronisch zerfällt, um so das größtmögliche Verzweigungsverhältnis und außerdem keine zusätzlichen Neutrinos zu erhalten. Damit ergibt sich eine Signatur der Endzustände von vier leichten Jets, zwei b-Jets und zwei Neutrinos.

Diese Masterarbeit bieten einen ersten Einblick in eine Messung des Wirkungsquerschnitts in Daten, die bei einer Schwerpunktenergie von $\sqrt{s} = 13\,\text{TeV}$ am ATLAS-Detektor mit einer integrierten Luminosität von $14.61\,\text{fb}^{-1}$ aufgenommen wurden. Dabei wurde ein verstärkter Entscheidungsbaum trainiert um Signal von Untergrund zu trennen. Der resultierende Klassifikator wurde dann für einen Wahrscheinlichkeitsfit verwendet, in dem sowohl die Stärke des Untergrunds, als auch die Signalstärke für das Fitten an Daten verwendet wurden. Auf diese Weise wurde ein Wert für den Wirkungsquerschnitt von $\sigma_{t\bar{t}(Z\to\nu\bar{\nu},qq,\tau^-\tau^+)} = 4158^{+490}_{-475}(\text{stat})\,\text{fb}$ erreicht. Dieser Wert enthält keine systematischen Unsicherheiten, welche als sehr groß zu erwarten sind. Daher ist es nicht möglich, ohne ausführliche weitere strenge Tests des Standardmodells durchzufÃ¼hren, davon neue Physik abzuleiten.

Abstract

The associated production of a top-antitop pair and a Z boson has only recently been observed, and there are no previous cross section measurements for the channel with the Z boson decaying to two neutrinos. The top-antitop pair for this analysis is required to decay fully hadronically, to select for the largest possible branching fraction and no additional neutrinos from this decay. This gives a final state signature of four light jets, two b jets and two neutrinos.

This thesis constitutes a first look into a cross section measurement in data recorded at a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ at the ATLAS detector with an integrated luminosity of $14.61\,\text{fb}^{-1}$. For this, a boosted decision tree was trained to separate signal from background. This classifier output was then used for a likelihood fit where the signal strength, as well as the background strengths, are fitted to data. With this, a cross section value of $\sigma_{t\bar{t}(Z\to\nu\bar{\nu},qq,\tau^-\tau^+)} = 4158^{+490}_{-475}(\text{stat})\,\text{fb}$ was derived. This value does not include systematic uncertainties, which are expected to be very large, therefore it is not possible to claim new physics without extensive further studies and strict tests of the Standard Model.

Contents

1. Introduction

Elementary particle physics in this epoch revolves around the Standard Model of Particle Physics (SM) [6–9]. It accurately explains many phenomena which are observable on subatomic scale.

Large physics experiments, such as the ATLAS detector at the Large Hadron Collider (LHC) in Geneva, continuously test the hypotheses of the SM and are refining the precision of measurements. This includes searching for new particles such as the recently-discovered Higgs boson [1, 2] or trying to find physics that goes beyond the SM.

Even though the SM is a well-established theory, there are also observations that are not explained by it. One of these is the neutrino mass, which should be zero according to the SM. In experiments, (see for example [3]), however, neutrinos are found to oscillate between flavours. This effect is only possible if at least one of the neutrinos has a mass larger than zero, contradicting the Standard Model without any theoretical extensions.

An exciting research object in the Standard Model is the top quark. It was only discovered in 1995 [19, 20] and there are still certain properties which have, so far, only been possible to measure indirectly. It is also the only quark which decays before it can hadronise, making it an ideal quark to study due to direct access to its decay products. The top quark decays via the electroweak force, and the coupling between the weak gauge bosons W^+, W^-, and Z^0 (in the following mostly referred to as Z) and the top quark is a promising subject of study. By studying the coupling to the Z boson, it is possible to measure the third component of the weak isospin of the top quark. Because the process $Z \to t\bar{t}$ is kinematically suppressed at current particle accelerators and energies, this vertex is accessible only by studying the associated production of a Z boson and $t\bar{t}$. The first searches for this process were successful in the ATLAS and CMS experiments [4, 5].

This Master's thesis looks into the process $t\bar{t}Z$ with a $Z \to \nu\bar{\nu}$ decay, and carries out a first cross section measurement. The $t\bar{t}$ pair, here, decays fully hadronically, which means both W bosons originating from the top quark decay into two additional quarks. The $t\bar{t}Z$ process with a two-neutrino final state has not been studied before, which makes this analysis very exciting. However, it is also a challenging topology to search for, since its signature in the fully hadronic $t\bar{t}$ decay channel only consists of six jets and large

missing transverse momentum, where the missing transverse momentum originates from the neutrinos as final decay states of the Z boson. Additional $t\bar{t}Z$ signal channels which can pass the two-neutrino final state event selection are $t\bar{t}$ with $Z \to qq$ and $Z \to \tau^-\tau^+$.

First, an introduction into the theoretical background is given in Chapter 2. The Standard Model including its elementary particles and interactions, and the electroweak unification with emphasis on the Z boson are explained. Next, the top quark, its properties, as well as production and decay modes are investigated. The process $t\bar{t}Z$ is analysed with focus on previous cross section measurements at ATLAS and CMS and the signatures of the relevant decay channels. The experimental set-up, namely the LHC and the ATLAS detector are presented in the following chapter, Chapter 3. In Chapter 4, the physical objects are defined and events of the data formats are selected based on the desired signature in the detector, before they are used as input for the Analysis Method, as described in Chapter 5. In this chapter, boosted decision trees and template fitting are explained.

Chapter 6, then, presents the results of the complete analysis and a cross section value is extracted, followed by a summary and conclusion in Chapter 7.

2. Theoretical Background

This chapter gives an introduction into the theoretical background of the Standard Model of Particle Physics. Afterwards, the focus lies on the top quark, the Z boson and the important process for this thesis, the $t\bar{t}Z$ production and decay. While $t\bar{t}(Z \to \nu\bar{\nu})$ is the main signature in this thesis, the processes $t\bar{t}(Z \to qq)$ and $t\bar{t}(Z \to \tau^-\tau^+)$ also contribute to the signal here, and, with this, to the desired $t\bar{t}Z$ cross section.

2.1. The Standard Model of Particle Physics

The Standard Model of Elementary Particle Physics (SM) is currently the closest theory to the physicists' final goal of a theory which encompasses all known particles and interactions. Based on the theory of the Electroweak Unification by Glashow, Salam and Weinberg [6–9], the SM was developed in the 1960s and 1970s. Its second component is Quantum Chromodynamics (QCD) which enables the Standard Model to describe the behaviour of the particles around us and their interactions very precisely.

2.1.1. Elementary Particles

The known particles up to now can be sorted into two categories; fermions (half-integer spin values) and bosons (integer spin values). Bosons, as mediators of the interactions between elementary particles, are the subject of Section 2.1.2.

Both, quarks and leptons, are fermions, and come in three generations. The quarks are separated into up-type quarks (up u, charm c, and top t) and down-type quarks (down d, strange s, and bottom b), with every combination of generation and up-/down-type identifying one specific particle. Similarly, for the leptons, one can distinguish between the charged leptons electron e, muon μ, and tau τ and their respective neutrinos ν_e, ν_μ, and ν_τ. With increasing generation number, the masses of the particles also increase, even though the masses of the neutrinos are negligibly small and only upper limits on their masses could be determined so far. The masses of the particles are listed in Table 2.1. Other important properties can be found in Table 2.2. Since no right-handed neutrinos

or left-handed anti-neutrinos have been detected yet – if they even exist – the respective values could not be entered in the columns of T^3 and Y.

Quarks	**Mass m [MeV/c²]**	**Leptons**	**Mass m [MeV/c²]**
u	$2.2^{+0.5}_{-0.4}$	e	$0.5109989461 \pm 0.0000000031$
d	$4.7^{+0.5}_{-0.3}$	μ	$105.6583745 \pm 0.0000024$
c	1275^{+25}_{35}	τ	1776.86 ± 0.12
s	95^{+9}_{-3}	ν_e	$< 2 \cdot 10^{-6}$
t	$173210 \pm 510 \pm 710$	ν_μ	$< 2 \cdot 10^{-6}$
b	4180^{+40}_{-30}	ν_τ	$< 2 \cdot 10^{-6}$

Table 2.1.: Fermion masses according to most recent measurements [10].

Particle Type	**Generation**			**Q[e]**	**C**	**T³**		**Y**	
	1	**2**	**3**			**L**	**R**	**L**	**R**
Leptons	ν_e	ν_μ	ν_τ	0	none	$+\frac{1}{2}$	-	-1	-
	e	μ	τ	-1	none	$-\frac{1}{2}$	0	-1	-2
Quarks	u	c	t	$+\frac{2}{3}$	r,g,b	$+\frac{1}{2}$	0	$+\frac{1}{3}$	$+\frac{4}{3}$
	d	s	b	$-\frac{1}{3}$	r,g,b	$-\frac{1}{2}$	0	$+\frac{1}{3}$	$-\frac{2}{3}$

Table 2.2.: Properties of the fermions according to the Standard Model. Given properties are the electric charge Q, the colour charge C, the third component of the weak isospin T^3 and the weak hypercharge calculated by $Y = 2(Q - T^3)$. For the last two values, the particles need to be separated into left-handed (L) and right-handed (R). The listed down-type quarks d, s and b are the electroweak eigenstates and not the mass eigenstates, for details see Section 2.1.2 [10].

2.1.2. Interactions

According to the Standard Model, there are three different particle interactions:

- the electromagnetic force carried by the photon
- the strong force carried by one of the gluons
- the weak force carried either by the Z or a charged $W^\pm$ boson.

Gravitation is another elementary interaction which is left out here since it cannot yet be described by the Standard Model. For that, a carrier called the graviton is postulated,

and evidence for its existence has not yet been found. In theory, any interaction between two particles from Section 2.1.1 is mediated by a gauge boson. All bosons in the SM with their properties are listed in Table 2.3.

Boson	**Force**	**Mass m [GeV/c²]**	**Q[e]**	**T³**
Photon γ	electromagnetic	$0\ (< 1 \cdot 10^{-27})$	0	0
Gluon g	strong	0	0	0
Z	weak	91.1876 ± 0.0021	0	0
$W^{\pm}$	weak	80.379 ± 0.012	± 1	± 1
H^0	-	125.18 ± 0.16	0	0

Table 2.3.: Bosons with their properties. Given properties are the mass m, the electric charge Q, and the third component of the weak isospin T^3 [10].

The Standard Model is a gauge theory, thus the interactions can be described by local gauge symmetries and their gauge groups within separate field theories. While the theory of *Quantum Electrodynamics (QED)* describes the electromagnetic force with the group $\mathrm{U(1)_{EM}}$ and the invariance under the transformation $\phi \rightarrow \phi' = \exp(i\theta)\phi$, where ϕ is a spinor field and θ is a real number, the *Quantum Chromodynamics (QCD)* describes the strong force with the symmetry group $\mathrm{SU(3)_C}$. Here, the C in the index stands for the QCD colour charge. The QCD gauge group comprises the Gell-Mann matrices λ_k with $k = 1, \ldots, 8$. Each matrix λ_k stands for one of the eight gluons which all carry both a colour and an anti-colour of (anti)red, (anti)green or (anti)blue. The gluons cannot exist as single particles; this is called gluon confinement. Since they carry exactly the same type of charge that they are mediating, they can couple to each other the same way they can couple to quarks. The quarks also cannot exist separated from each other. This is called quark confinement and can be explained by the appearance of a quark-anti-quark pair when two initial quarks are separated. Therefore, they are bound to always form colour-neutral bound states which also have an electromagnetic charge of an integer value, called hadrons. The combination of two quarks is called a meson while three quarks form baryons in their confined state. The quark confinement implies that the coupling strength of the strong force, in contrast to the other forces, becomes stronger with larger distance [11, 12].

Lastly, the *Glashow-Weinberg-Salam (GWS) theory* does not only describe the weak force but also unifies the weak and the electromagnetic force. This way, the so-called electroweak force is subject to the GWS theory and the electromagnetic as well as the weak force are just two aspects of the same unified force. The gauge group for this unification is then $\mathrm{SU(2)_L} \times \mathrm{U(1)_Y}$. The index L stands for left-handed, the index Y

for the weak hypercharge $Y = 2(Q - T^3)$ which is listed in Table 2.2. Details for the electroweak force are given below [6–8].

The combination of these theories gives the Standard Model symmetry group

$$SU(3)_{\mathrm{C}} \times SU(2)_{\mathrm{L}} \times U(1)_{\mathrm{Y}} \tag{2.1}$$

to describe the Standard model as a local gauge and also Lorentz invariant Lagrange density function which can be shown to be fully renormalisable [13].

Electroweak Unification

As mentioned before, the electroweak force as a unification of the electromagnetic and weak forces is described by $\mathrm{SU}(2)_{\mathrm{L}} \times \mathrm{U}(1)_{\mathrm{Y}}$. The weak force in this is described by the symmetry group $\mathrm{SU}(2)_{\mathrm{L}}$ which comprises the three Pauli matrices σ_i, with $i = 1, 2, 3$. An L in the index stands for left-handed, as the weak isospin current T only couples to left-handed fermions (and right-handed antifermions) because the weak charged current maximally violates parity [14]. This means that the vertex structure is not purely vectorial, but instead, as indicated by a measurement of the neutrino helicity [15], has a vector-axial vector (V-A) structure in the form of,

$$-\frac{ig_W}{2\sqrt{2}}\left(\gamma^\mu - \gamma^\mu\gamma^5\right) \tag{2.2}$$

with γ^μ being the Dirac matrices, and $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$, and g_W being the electroweak coupling constant.

The V-A structure can be absorbed in the fermion spinor and so the projection operators to left- and right-handed fermion states become $P_L = \frac{1}{2}(1 - \gamma^5)$ and $P_R = \frac{1}{2}(1 + \gamma^5)$. With this, the electromagnetic force still preserves parity while also coupling to both left- and right-handed fermion states. At the same time, the weak force distinguishes between left- and right-handed states and acts as vectorial coupling. The electroweak force couples to the electroweak hypercharge Y instead of only the electromagnetic or a weak charge [7, 16].

Furthermore, the physical gauge bosons γ, Z, W^+ and W^- do not correspond directly to the generators of the electroweak symmetry group $\mathrm{SU}(2)_{\mathrm{L}} \times \mathrm{U}(1)_{\mathrm{Y}}$ which are referred to as W^1, W^2, W^3 and B. Mixing these states by using a mixing angle called the Weinberg

angle θ_W,

$$\begin{pmatrix} \gamma \\ Z \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B \\ W^3 \end{pmatrix} \tag{2.3}$$

$$\begin{pmatrix} W^+ \\ W^- \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & i \\ -1 & -i \end{pmatrix} \begin{pmatrix} W^1 \\ W^2 \end{pmatrix} \tag{2.4}$$

gives the physical gauge bosons. This mixing also accounts for the bosons having different masses. The Weinberg angle has been measured to be $\sin^2\theta_W = 1 - (m_W/m_Z)^2 = 0.2223(21)$ [6–9, 17].

After this unification, the Lagrangian of the Z boson interaction is,

$$\mathcal{L}_Z = \frac{g_e}{2\sin\theta_W\cos\theta_W} \sum_{f=L_L,e_R,Q_L,u_R,d_R} Z_\mu \bar{f} \left(C_V - C_A\gamma^5\right)\gamma^\mu f \tag{2.5}$$

with g_e the electromagnetic coupling constant, $L_L = \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}$ and $Q_L = \begin{pmatrix} u \\ d \end{pmatrix}$ the weak isospin doublets of the left-handed leptons and quarks and the denotations for the respective fermion f,

$$C_V = T^3 - 2Q\sin^2\theta_W \tag{2.6}$$

$$C_A = T^3\,. \tag{2.7}$$

The Z and W bosons as mediators of the weak force can couple to all fermions and even to each other and certain other bosons, in contrast to photons. Because the Z boson does not have an electric charge, the weak force mediated by it is also called the *neutral current*. In contrast, the *charged current* is mediated either by the W^+ or W^-. The weak force is the only interaction that is able to change the flavour of a particle and it is only possible under the effect of a charged current. This way, for example a muon μ^- is able to convert to a muon neutrino ν_μ while emitting a W^- or absorbing a W^+.

For quarks, a small percentage of interactions are also able to happen between different generations. This is possible because the stationary states of the quarks produced in QED or QCD, the physical quarks as described by the strong isospin doublets, are not the same as the weak eigenstates of the down-type quarks: d', s', b' to which the boson couples to. The mixing is similar to the electroweak generators and physical gauge bosons: the weak eigenstates are linear combinations of the physical quarks as described by the

Cabibbo-Kobayashi-Maskawa (CKM) matrix [18]

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{2.8}$$

The entries of the CKM matrix are not independent and can be reduced to three real valued parameters and a complex phase factor. The latter is the reason the weak force violates the conservation of charge and parity (CP violation).

The numerical values of the CKM parameters cannot be predicted by theory but have to be determined from experiments. Their magnitudes are [10]:

$$\begin{pmatrix} 0.97420 \pm 0.00010 & 0.22452 \pm 0.00044 & 0.00365 \pm 0.00012 \\ 0.22439 \pm 0.00044 & 0.97359^{+0.00010}_{-0.00011} & 0.04214 \pm 0.00076 \\ 0.00896^{+0.00024}_{-0.00023} & 0.04133 \pm 0.00074 & 0.999105 \pm 0.000032 \end{pmatrix}. \tag{2.9}$$

2.2. The Top Quark

The top quark was first proposed by Kobayashi and Maskawa in 1973 as the weak isospin partner of the bottom quark in the third generation of quarks. This was needed to explain the occurrence of CP violation in the weak interactions [18]. The first detection of the top quark happened 22 years later in $p\bar{p}$ collisions at the TEVATRON Run I at a centre-of-mass energy of $\sqrt{s} = 1.8\,\text{TeV}$ [19, 20]. Subsequently, the single top quark production was also first discovered at Fermilab in 2009 during the Run II at the TEVATRON at $\sqrt{s} = 1.96\,\text{TeV}$ [21, 22].

Top quark research is important since it is the only quark which decays before it can hadronise due to its large mass and short lifetime. This means that the top quark is the only quark that is not a main component in any existing hadron, and is therefore the only possibility to study an isolated quark.

In the following, the important particle properties as well as the production and decay of the heaviest of quarks are presented, before the topic of this analysis, the process $t\bar{t}Z$, is addressed.

2.2.1. Top Quark Properties

The latest and most precise value for the mass of the top quark is [23]

$$m_t = 173.34 \pm 0.27\,(\text{stat}) \pm 0.71\,(\text{syst})\,\text{GeV} \tag{2.10}$$

with a total uncertainty of 0.76 GeV. The relative uncertainty of 0.44% is small . This value combines measurements from the TEVATRON experiments CDF, and DØ, and the LHC Run I measurements from the ATLAS and CMS experiments.

In addition to the properties listed in Table 2.2, the top quark, as a fermion, carries a spin of $s = \frac{1}{2}$. The electromagnetic charge of the top quark has not been measured directly yet. Although ATLAS indirectly measured the top quark charge to be $0.64 \pm 0.02(\text{stat}) \pm 0.08(\text{syst})\, e$ from the charges of top quark decay products, and, with this, excluded the exotic scenario of a charge of $-\frac{4}{3}\, e$ at more than $8\,\sigma$[24]. In addition, CMS was able to exclude the exotic scenario at 99. % C.L. in the muon+jets final states in $t\bar{t}$ events [25].

The reason why no direct measurements were successful so far is the extremely short predicted lifetime of

$$\tau_t = \frac{\hbar}{\Gamma_t} = 3.29 \cdot 10^{-25}\,\text{s} \tag{2.11}$$

which was calculated with the value of $\hbar = 6.582119514(40) \cdot 10^{-16}\,\text{eV s}$ [17] and the decay width of the top quark [10]

$$\Gamma_t = 1.41^{+0.19}_{-0.15}\,\text{GeV}. \tag{2.12}$$

The lifetime in Equation 2.11 is much smaller than the time required for a top quark to form hadrons [26]

$$\tau_{\text{had}} = \frac{\hbar}{\Lambda_{QCD}} = \frac{\hbar}{213\,\text{MeV}} \approx 3 \cdot 10^{-24}\,\text{s}. \tag{2.13}$$

Thus, no top or anti-top quark can be found as a valence quark in any existing hadron. Here, Λ_{QCD} is the QCD time scale which is the only dimensionful parameter of the pure QCD interaction.

2.2.2. Production and Decay

At hadron colliders like the TEVATRON and LHC, top quarks can be produced in two different ways – either as top anti-top pairs via the strong interaction, or as single top quarks via the weak interaction mediated by the exchange of a W boson.

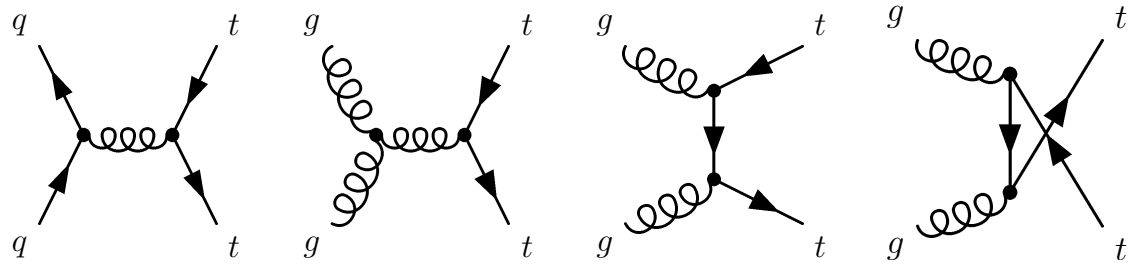


Figure 2.1.: Feynman diagrams of $q\bar{q}$ annihilation and gg fusion that produces top quark pairs in hadron collisions.

Top Quark Pair Production

As shown in Figure 2.1, $t\bar{t}$ pairs can be produced from two particle initial states. Either a quark and an anti-quark of the same type annihilate into a gluon, or two gluons fuse into one which then decays into the desired $t\bar{t}$ pair, or these two gluons exchange a quark under emission of the $t\bar{t}$ pair. The dominant process is determined by the type of colliding particles ($p\bar{p}$ or pp) and the provided centre-of-mass energy, $\sqrt{s}$. The first defines the valence quarks. If only protons collide with each other, processes that need anti-particles are suppressed at lower energies. Conversely, the centre-of-mass energy defines the parton distribution function that determines how the sea quarks and gluons in the hadrons are distributed, and therefore defines the fraction of the total momentum of the hadron which is carried by the collided parton. That momentum has impact on the favoured process of $t\bar{t}$ production by influencing the collision probabilities of all partons [27].

In accordance with the above, only 15 % of the production of $t\bar{t}$ pairs were initiated by gluon fusion during the Tevatron Run II ($\sqrt{s} = 1.96\,\text{TeV}$, $p\bar{p}$). In contrast, at the Lhc, the gluons provide the dominant process with a contribution of 80 % at Run I with $\sqrt{s} = 7\,\text{TeV}$ and around 90 % at Run II with $\sqrt{s} = 13\,\text{TeV}$ [10].

Single Top Quark Production

The second possibility to produce top quarks in hadron collisions is as single top quarks, produced by weak processes such as the diagrams in Figure 2.2.

Single top quark events are not as easy to identify as $t\bar{t}$ events since they have higher levels of background. In the production of a single top quark, another particle, a lighter quark or a W boson is also produced. Therefore, depending on the decay of the top quark, only a few jets (up to five but at least one) and a maximum of two charged leptons can be observed. This production mechanism of top quarks was first observed at the Tevatron [21, 22].

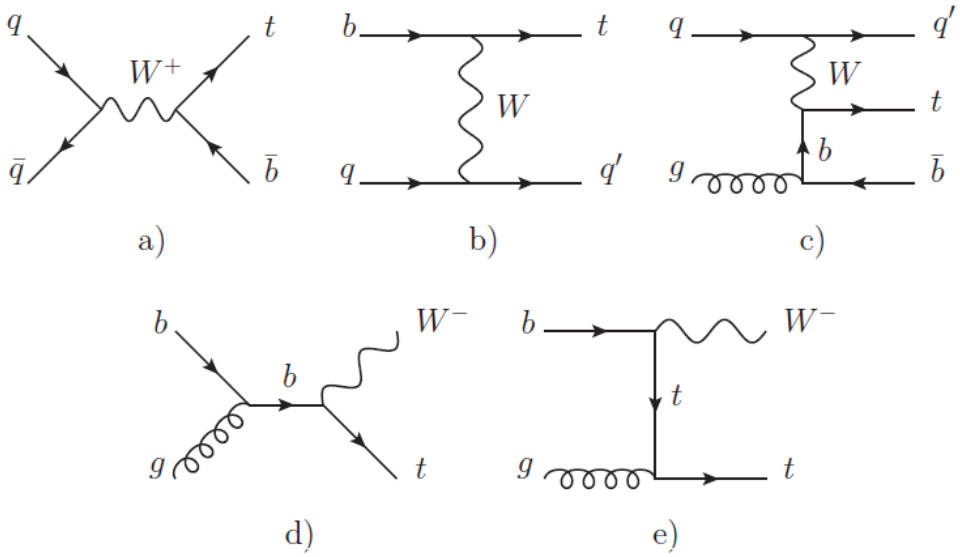


Figure 2.2.: Leading order Feynman diagrams for the production of single top quarks in the a) s-channel, b) t-channel $2 \to 2$, c) t-channel $2 \to 3$, d), e) and W associated production.

The Top Quark Decay

The decay of the top quark is a weak process, and thus determined by the CKM matrix element V_{tb}. The probability of the decay into a b quark and a W^+ boson is dependent on the matrix element, which is $|V_{tb}| = 0.999105 \pm 0.000032$ [10]. This means that cases where the top quark does not decay into a b quark can be neglected for this analysis, even though decays into other down-type quarks are possible with a very small fraction ($|V_{ts}| = 0.04$ and $|V_{td}| = 0.009$, see Section 2.1.2).

The produced b quark hadronises and forms jets which contain B mesons. The B mesons are able to travel a certain distance from the primary vertex. A b tagger, then, is able to identify these secondary vertices and corresponding decaying B mesons during the event reconstruction process.

The W boson can either decay hadronically into a light quark and a light anti-quark (u, d, s, c) with a probability of $(67.41 \pm 0.27)\,\% \approx \frac{2}{3}$ [10], or leptonically into a charged lepton (e, μ, τ) and the corresponding neutrino at a probability of $33\,\% \approx \frac{1}{3}$ [10], each charged lepton equally likely. The leptonic decay into a τ and its neutrino is not considered specifically in this analysis because the occurrence of τ leptons makes distinguishing between leptonic and hadronic decays difficult, but hadronically decaying τ leptons also contribute indirectly to hadronic decay channels.

This results in a branching fraction of $\mathcal{B}_{\text{lep}} = \frac{2}{9}$ for the decay into e or μ and a branching fraction of $\mathcal{B}_{\text{had}} = \frac{2}{3}$ for a hadronic decay including a single charged lepton. For a $t\bar{t}$ pair

decay, this gives three different possibilities for a decay, as shown in Figure 2.3. The separate processes are drawn in Figure 2.4.

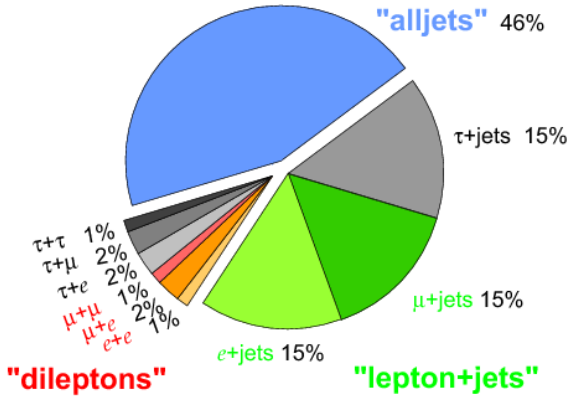
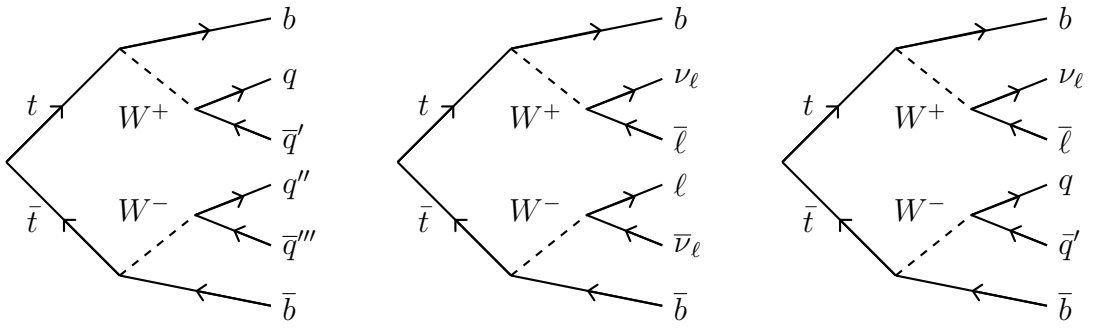


Figure 2.3.: Distribution of branching fractions in top pair decays.



(a) All-jets channel. ***(b)*** Dileptonic Channel. ***(c)*** Semileptonic channel.

Figure 2.4.: Feynman diagrams for the decay of $t\bar{t}$ events. In the semileptonic channel, in addition to the shown process, it is also possible that the W^+ decays hadronically while the W^- decays leptonically.

If both W bosons decay hadronically it is called the *all-jets* or *allhadronic decay channel*. The advantage of this decay channel is that first the branching fraction $\mathcal{B}_{\text{alljet}} = \frac{4}{9}$ is larger than for both other channels. Second, there are no neutrinos present which cannot be observed directly. The disadvantage is that six jets occur for which the experimental resolution is lower than, for example, for an electron or a muon. The largest difficulty with a signal which only consists of jets is that the background has very large contributions from QCD multijet events that look similar to the signal.

The second case shown is the *dileptonic decay channel*. Here, both W bosons decay leptonically so that the only two observed jets are clearly visible b jets. The remaining tracks are signals with even higher resolutions due to the high transverse momentum of

the electrons or muons. The main difficulty with this decay channel is the high missing transverse momentum. This is produced by two invisible neutrinos that cannot be distinguished by the detector. Another smaller disadvantage of analysing this $t\bar{t}$ decay channel is that this channel has the lowest branching fraction of $\mathcal{B}_{\text{dilep}} = \frac{4}{81}$, taking only decays to electrons and muons into account.

In the *semileptonic decay channel*, or also called $\ell + jets$ *channel*, one of the two W bosons decays hadronically while the other decays leptonically. This has a branching fraction of $\mathcal{B}_{\text{semilep}} = \frac{8}{27}$. Also the quality of the final state signatures combines both high and low features of the other two channels. The signal of two jets is less clear than the signal of a final state with a jet and a lepton, while here only one neutrino contributes to the missing energy so that theoretically all particles can be reconstructed accurately (minus the QCD uncertainties from the jets).

Which $t\bar{t}$ decay channel is the ideal choice as a subject of research depends highly on the general analysis and, with this, the other particles associated with the decay of the $t\bar{t}$. Because this thesis is about the process $t\bar{t}(Z \to \nu\bar{\nu})$, the allhadronic channel is selected for the high branching ratio and lack of additional neutrinos.

2.3. The Process $t\bar{t}Z$

As previously mentioned, the top quark is the only quark that decays before it hadronises and which can therefore be studied directly. The top quark properties are important to determine and to improve the current limits of the Standard Model and possibly access new physics beyond the SM. The third component of the weak isospin as a direct particle property can be measured through the measurement of the electroweak coupling strength of the top quark to the Z boson. From there, a comparison with coupling strengths of the W boson could give details about electroweak symmetry breaking. A collider does not yet exist that can produce a clean signature of a virtual Z boson decaying via $Z \to t\bar{t}$ with a well-known centre-of-mass energy to easily access the tZ coupling. Therefore, the $t\bar{t}$ pair production in association with a Z boson is the best channel to measure this coupling and a way to be sensitive to anomalous tZ couplings.

The SM Lagrangian of a $t\bar{t}$ pair coupling to a Z boson is given by

$$\mathcal{L}_{t\bar{t}Z} = \frac{e\bar{u}(p_t)\gamma^\mu}{2\sin\theta_W\cos\theta_W}\left(C_V - \gamma_5 C_A\right) v(p_{\bar{t}}) Z_\mu \tag{2.14}$$

where, like before, C_V and C_A are the vector and axial vector couplings which contain the third component of the weak isospin of the top quark, and $u(p_t)$ and $v(p_{\bar{t}})$ are the spinors

for the top quark and antitop quark, respectively. If the coupling shows deviations in the experimental result from the theory, this would indicate physics beyond the SM prediction of the third component of the weak isospin T_t^3, since the top quark electromagnetic charge Q_t and the Weinberg angle θ_W have been measured to a high precision in previous experiments. Measurements of the $t\bar{t}Z$ cross section are therefore sensitive to T_t^3 (and new particles) and are able to support or disprove the assumption of $T_t^3 = +\frac{1}{2}$.

2.3.1. $t\bar{t}$Z Production

The leading-order Feynman diagrams of the production of such a $t\bar{t}Z$ process are shown in Figure 2.5. The Z boson can be radiated both as *initial state* (ISR) or *final state radiation* (FSR). Even though the event signature of both processes, ISR and FSR, is the same, only the FSR process is of interest for the tZ coupling strength.

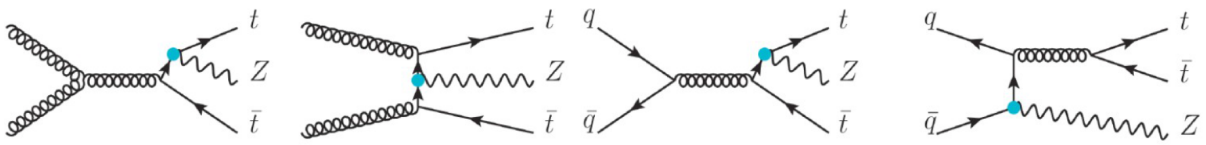


Figure 2.5.: Leading order Feynman diagrams for the $t\bar{t}$ production in association with a Z boson. The diagram on the right shows the initial state radiation which does not have a tZ vertex. The other diagrams have the top quark radiating the Z boson and thus making the coupling strength measurement possible.

The proposed production cross sections of $t\bar{t}Z$ processes as a function of $\sqrt{s}$ are shown in Figure 2.6 in comparison to processes $t\bar{t}W$ and $t\bar{t}H$. The $t\bar{t}Z$ cross section is higher than for the other processes with a value of [29]

$$\sigma_{pp\to t\bar{t}Z} = 759.8 \pm 2.6^{+9.7\%}_{-11.1\%}(\text{scale})^{+1.9\%}_{-2.2\%}(\text{PDF})\,\text{fb} \tag{2.15}$$

at next-to-leading-order calculations for a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ (which is relevant for this thesis).

2.3.2. Previous $t\bar{t}Z$ Cross Section Analyses

The first measurement of the $t\bar{t}Z$ process was achieved by ATLAS and CMS at the LHC at $\sqrt{s} = 7\,\text{TeV}$ [4, 5]. ATLAS used $4.7\,\text{fb}^{-1}$ of data to set an upper limit to the $t\bar{t}Z$ cross section of $\sigma_{t\bar{t}Z} < 0.71\,\text{pb}$ at a 95% confidence level in three charged lepton final states (where the Z boson decays to a pair of charged leptons and the $t\bar{t}$ pair decays semileptonically)

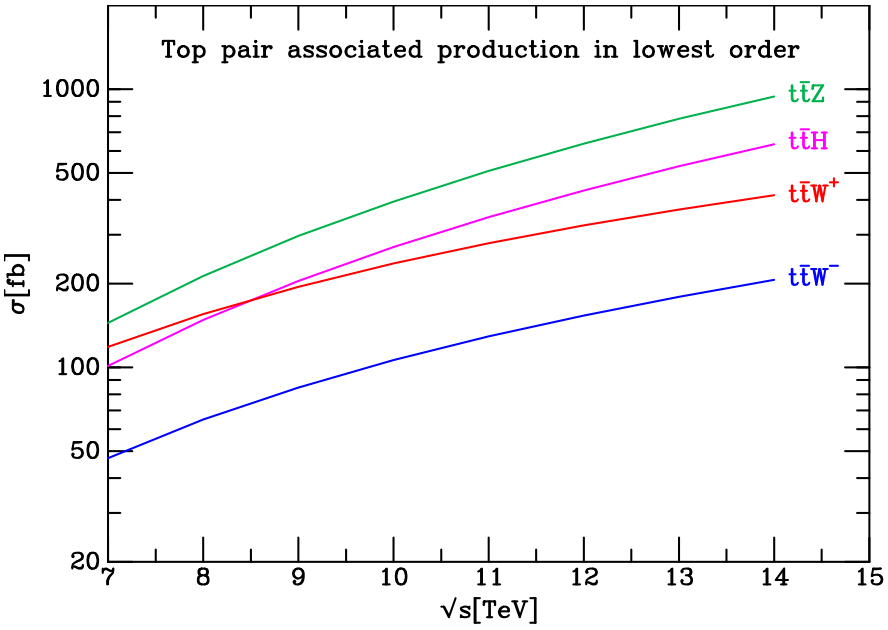


Figure 2.6.: Lowest order cross sections for $t\bar{t}$ pair production in association with a boson at a pp collider, as a function of the centre-of-mass energy $\sqrt{s}$. Renormalization and factorization scales have been set to $m_t = 172.5\,\text{GeV}$. The $t\bar{t}H$ process is computed for a Higgs boson mass of 125 GeV [28].

[4]. CMS, on the other hand, analysed data corresponding to an integrated luminosity of $5.0\,\text{fb}^{-1}$ in two different channels, the same three charged lepton final state ATLAS used for this result, as well as a combined search in $t\bar{t}Z$ and $t\bar{t}W$. The direct measurement lead to a cross section of $\sigma_{t\bar{t}Z} = 0.28^{+0.14}_{-0.11}(\text{stat})^{+0.09}_{-0.07}(\text{syst})\,\text{pb}$ with a significance of 3.3 standard deviations from the background hypothesis [5]. Within uncertainties, this value is compatible with the NLO prediction of $0.137^{+0.012}_{-0.016}\,\text{pb}$ at the considered centre-of-mass energy [30].

Measurements at $\sqrt{s} = 8\,\text{TeV}$ by ATLAS and CMS also showed agreement with the expectation with similar significance [31, 32].

Also for $\sqrt{s} = 13\,\text{TeV}$ data, the $t\bar{t}Z$ (and $t\bar{t}W$) cross sections have already been measured. ATLAS found the cross section to be,

$$\sigma_{t\bar{t}Z} = 0.9 \pm 0.3\,\text{pb} \tag{2.16}$$

from data of same-charge muons or three or four charged leptons (electrons or muons) corresponding to a total integrated luminosity of $3.2\,\text{fb}^{-1}$ [33].

Similarly, CMS measured a cross section of,

$$\sigma_{t\bar{t}Z} = 0.99^{+0.09}_{-0.08}(\text{stat})^{+0.12}_{-0.10}(\text{syst})\,\text{pb} \tag{2.17}$$

with the significance in excess of 5 standard deviations from the background-only hypothesis, using a total data sample corresponding to an integrated luminosity of $35.9\,\text{fb}^{-1}$ in three- and four-lepton channels [34]. All these values agree with the expected value from Equation 2.15.

2.3.3. The process $t\bar{t}Z$ with $Z \to \nu\bar{\nu}$

For the study presented here, the channel of $Z \to \nu\bar{\nu}$ was chosen, in contrast to the already well-established $Z \to \ell\ell$ analyses for which the previously mentioned results were obtained. The $Z \to \nu\bar{\nu}$ channel has not been studied before and is important to look at because it complements the leptonic $t\bar{t}Z$ cross section measurement with an even higher branching fraction for the Z. While the branching fraction for Z decaying to a lepton is 3.4%, respectively for each lepton, the branching fraction to invisible final states is 20.0% [10].

This means, the branching fraction of $Z \to \nu\bar{\nu}$ is 5-6 times higher than the branching fraction of the Z boson decaying into leptons. The formula for the partial decay width of the Z boson into any fermion is

$$\Gamma(Z \to f\bar{f}) = N_C \frac{\sqrt{2} G_F m_Z^3}{6\pi} \times \left[(T_3 - Q_f \sin^2\theta_W)^2 + (Q_f \sin^2\theta_W)^2\right] \tag{2.18}$$

where θ_W is the weak mixing angle, $G_F = 1.166 \cdot 10^{-5}\,\text{GeV}^{-2}$ the Fermi constant [17], m_Z the mass of the Z boson according to Table 2.3, $N_C = 1$ for leptonic and $N_C = 3$ for $q\bar{q}$ final states and all other values defined according to Table 2.2. With this, the partial decay width with neutrinos (where the separate flavours $i = e, \mu, \tau$ are not distinguished) is $\Gamma(Z \to \nu\bar{\nu}) = 3 \times \Gamma(Z \to \nu_i\bar{\nu}_i) = 0.4764\,\text{GeV}$ which is 5.9 times greater than the value $\Gamma(Z \to \ell^-\ell^+) = 0.0804\,\text{GeV}$ for electrons or muons, respectively.

In addition to the branching fraction, the $t\bar{t}Z(Z \to \nu\bar{\nu})$ process, with a signature of only 6 jets and high missing transverse energy, is a relevant background for many supersymmetry (SUSY) searches. Understanding this $t\bar{t}Z$ process is crucial for the best possible understanding of those background events and thus successfully conducting these further studies.

As mentioned in Section 2.2.2, the all-jets $t\bar{t}$ decay is the most promising choice of $t\bar{t}$ decay for this analysis, and therefore a signature of six jets and missing transverse

momentum from neutrinos is the desired event signal.

2.3.4. The process $t\bar{t}Z$ with $Z \to q\bar{q}$

Apart from the channel $t\bar{t}(Z \to \nu\bar{\nu})$, $t\bar{t}(Z \to q\bar{q})$ also contributes largely to the selected signature of high missing transverse energy and six jets. The missing transverse energy, here, can originate from the semileptonically or dileptonically decaying $t\bar{t}$ pair, where the charged lepton in the decay products was a hadronically decaying tau lepton or an electron or muon which did not pass the requirements in the event selection. This is in contrast to the $t\bar{t}(Z \to \nu\bar{\nu})$ channel, where the $t\bar{t}$ pair is required to decay fully hadronically in order to guarantee a minimum number of jets in the event selection. Another contribution to missing transverse energy can come in the form of contributions from jets, such as b jets containing leptonic decays with neutrinos.

For this channel, with quark final states from the Z boson, Equation 2.18 gives $\Gamma(Z \to q\bar{q}) = 0.2778\,\text{GeV}$ for up-type quarks and $\Gamma(Z \to q\bar{q}) = 0.3561\,\text{GeV}$ for down-type quarks. Not all of these events with a fully hadronic $t\bar{t}$ decay may pass the event selection, e.g. when jets are not detected or overlapping, however events with semileptonic or dileptonic $t\bar{t}$ decays can contribute instead. This shows that the $t\bar{t}(Z \to q\bar{q})$ is expected to be a large fraction of the final signal.

2.3.5. The process $t\bar{t}Z$ with $Z \to \tau^-\tau^+$

Additionally, the process $t\bar{t}(Z \to \tau^-\tau^+)$ is a part of the signature when the tau leptons decay hadronically, resulting in a process similar to the previous $t\bar{t}(Z \to q\bar{q})$. The branching ratio of 3.6% is only small in comparison to a direct decay of the Z boson into quarks. The significance of this channel for the analysis is even further decreased by the fact that the tau leptons also decay leptonically into electrons and muons in 35% of the cases [10].

3. Experimental Setup

This master thesis analysis is performed and written in the ATLAS collaboration with data from the LHC, which is the topic of the first section in this chapter. Afterwards, the ATLAS detector systems are explained in detail, followed by an introduction into the simulation of Monte Carlo samples, which are necessary to make predictions that are compared to the data taken at the LHC.

3.1. The Large Hadron Collider

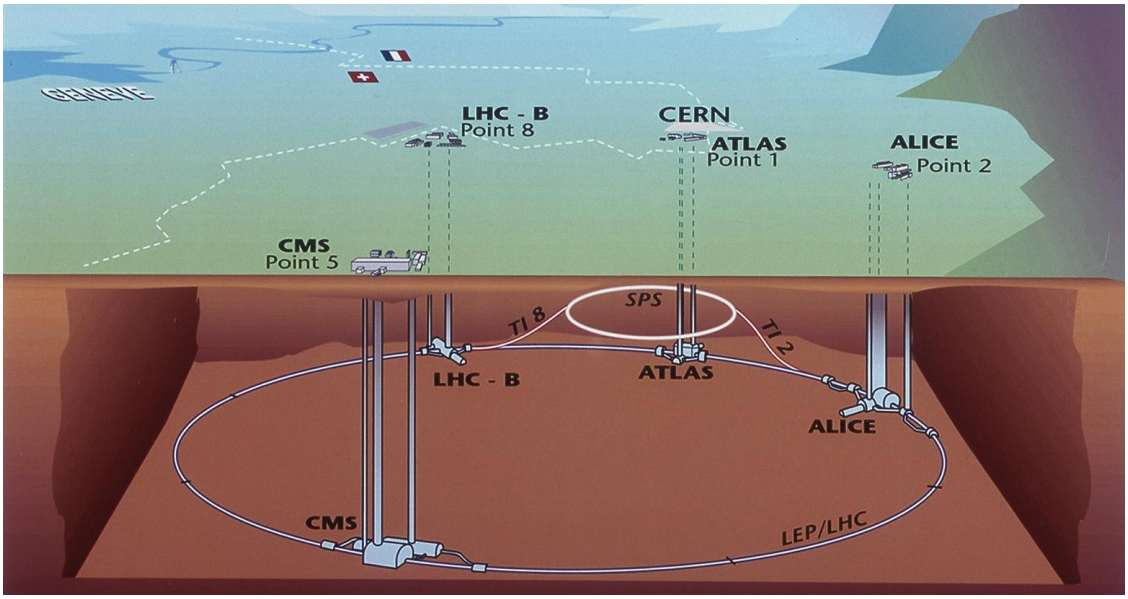


Figure 3.1.: The four main experiments at the LHC collider at CERN in Geneva. ©CERN

The Large Hadron Collider (LHC) is located at CERN[1] in Geneva on the Swiss-French border. The LHC is the largest particle collider in the world with a circumference of 26.7 km [35]. The LHC is a two-ring-superconducting-hadron accelerator and collider which is located in the already-existing tunnel of the Large Electron Positron (LEP)

[1]European Organization for Nuclear Research, abbreviation originating from: Centre Européen pour la Recherche Nucléaire.

Collider at a depth of approx. 100 m below ground. In Run I, the LHC produced proton-proton collisions at centre-of-mass energies of $\sqrt{s} = 7\,\text{TeV}$ (2011 dataset) and $\sqrt{s} = 8\,\text{TeV}$ (2012 dataset).

For the currently ongoing Run II, the LHC is colliding protons at a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ with the possibility of an increase to $\sqrt{s} = 14\,\text{TeV}$ in the future. Because the LHC collides same-charge particles, it includes two rings with counter-rotating beams, instead of one beam pipe where particles and antiparticles circle in opposite directions and share the same phase space [35].

The particle bunches are focused and bent by superconducting magnets which are cooled down with superfluid Helium to a temperature below 2 K, so that they produce fields above 8 T. The beam is focussed with the help of quadrupole magnets, and higher order multipole magnets, while bending of the trajectory of the particles to the required curvature is done by dipole magnets.

Before the particle bunches are injected into the LHC tunnel, they are accelerated to energies of 450 GeV by the CERN accelerator complex. Then, the particle bunches are injected, accelerated, and stored by a total of 16 radio frequency cavities, each operating at a maximum voltage of 2 MV, inside the colliders tunnel. This way, a total number of up to 2808 bunches can be stored inside the LHC, where each contains a total of about $1.15 \cdot 10^{11}$ protons [35].

Along the LHC ring, there are four main interaction points with the four main detector experiments located at each one of them, as shown in Figure 3.1. While the other three experiments focus on high-energy proton-proton collisions, the ALICE experiment investigates collisions of heavy ions, and therefore the theory of QCD. The ALICE collaboration searches for hints of the possible existence of a quark-gluon plasma, as well as looking at the behaviour of hadronic matter at high densities and temperatures [36].

LHCb uses an asymmetric detector which covers only part of the solid angle about the collision point. Here, physics at low scattering angles and measurements of B mesons happen to observe CP violation and look for physics beyond the Standard Model [37].

ATLAS and CMS are the biggest detectors at the LHC and cover almost the complete solid angle around their collision point. At these experiments, the Higgs boson which was discovered in 2012 [1, 2], is studied together with other rare Standard Model processes such as electroweak interactions involving the top quark. At the same time, the search for Dark Matter candidates and hints for the theory of Supersymmetry is ongoing [38, 39].

Besides the two main experiments, a few smaller ones exist near the interaction points. The aims of their research are, for example, the searches for magnetic monopoles (MoEDAL) [40] or the measurement of elastic and diffractive cross sections (TOTEM) [41].

3.2. The ATLAS Detector

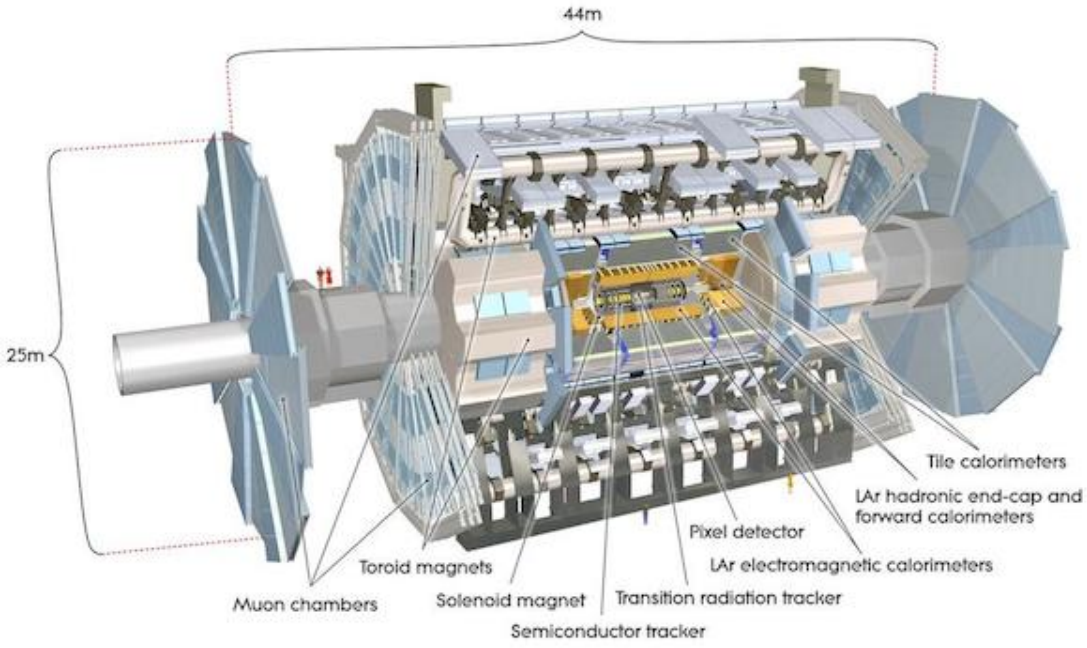


Figure 3.2.: Cut-away view of the ATLAS detector and its subsystems [38].

ATLAS, which stands for A Toroidal LHC ApparatuS, is a detector which covers, just like CMS, the complete solid angle and is able to detect particles up to a pseudorapidity of $|\eta| = 4.9$. The whole detector system has a weight of approximately 7000 t, is about 25 m in diameter and 44 m long. Its various layers are tracking detectors, calorimeters, and muon chambers. Unlike other detectors which are built for a specific purpose, ATLAS can perform measurements for many different studies. Its structure is shown in Figure 3.2 and explained in the following. The innermost layer contains the tracking detectors. These trace the trajectories of charged particles, and the recorded curvature provides information about the momentum of the particle passing through the detector. Because of the high lumiosity of the LHC, multiple different proton-proton collisions from the same bunch crossing overlap with each other in the detector, which is known as pile-up [42]. An excellent tracking system as used in the ATLAS detector is needed to separate these and measure a single event in a number of other events happening at the same time very close-by.

This tracking system is realised in four different sub-systems here. This inner detector measures the tracks of charged particles, which are bent by a magnetic field produced by solenoid magnets located between the tracking system and the electromagnetic calorime-

ter. This allows measurements of the charge and momentum of the charged particles, as well as vertex identification. The best spatial resolution in the inner detector is provided by the *Silicon Pixel Detector* having 80 million read-out channels with a minimum distance of $R = 4.55\,\text{cm}$ from the interaction point for the initial pixel detector used in Run I. This distance has been lowered to $R = 3.325\,\text{cm}$ with the addition of the Insertable B-Layer (IBL) in Run II. The IBL compensates radiation damage cased in Run I and improves the tracking and vertex identification [38, 43, 44].

Around the pixel detector, there is the *Silicon Microstrip Tracker* consisting of small silicon strips instead of pixels. The outermost tracking layer is formed by the *Transition Radiation Tracker*. It gives additional tracking information by measuring transition radiation of particles passing through. This allows the separation of lighter particles (like electrons) which transmit more transition radiation from heavier particles (like pions) which do not emit as much transition radiation. It has approximately 351,000 read-out channels. The whole tracking system has a momentum resolution of $\sigma_{p_T}/p_T = 0.05\,\% \cdot p_T[\text{GeV}] \oplus 1\,\%$ [38].

The next layers outside of the tracking and magnet systems are the electromagnetic and hadronic calorimeters. Those are needed to determine the energies of electrons, photons and jets precisely and thus need a high spatial resolution. The measurement is done by producing particle showers inside the detector material. The electromagnetic part is on the inside, surrounded by the hadronic calorimeter, so that in the electromagnetic part only the lightest particles, electrons and photons, are stopped completely with an energy resolution of $\sigma_E/E = 10\,\%/\sqrt{E} \oplus 0.7\,\%$. Except for muons and neutrinos all other known particles are then stopped by the hadronic layer. The energy resolution there depends on the position in the calorimeter but can be a maximum of $\sigma_E/E = 50\,\%/\sqrt{E} \oplus 3\,\%$. The electromagnetic calorimeter consists of liquid Argon (LAr) as active material while the hadronic one mostly uses stainless steel and scintillation plates [38].

Because muons do not emit enough Bremsstrahlung to be stopped by the calorimeters, special muon chambers are needed for their detection. These form the outermost layer of the whole detector system at the ATLAS experiment. The chambers have about a million read-out channels and deliver an absolute momentum resolution of $\sigma_{p_T}/p_T = 10\,\%$ at an energy of 1 TeV [38].

Figure 3.3 shows the different tracks of particles as they appear inside the ATLAS detector. By determining the specific track, the originating particle can be determined.

Because not all data recorded by the ATLAS experiment can be stored in general, and especially at the rate they are produced, triggers and event selection become important. There are three levels of triggering used. At first level, the hardware-based trigger is

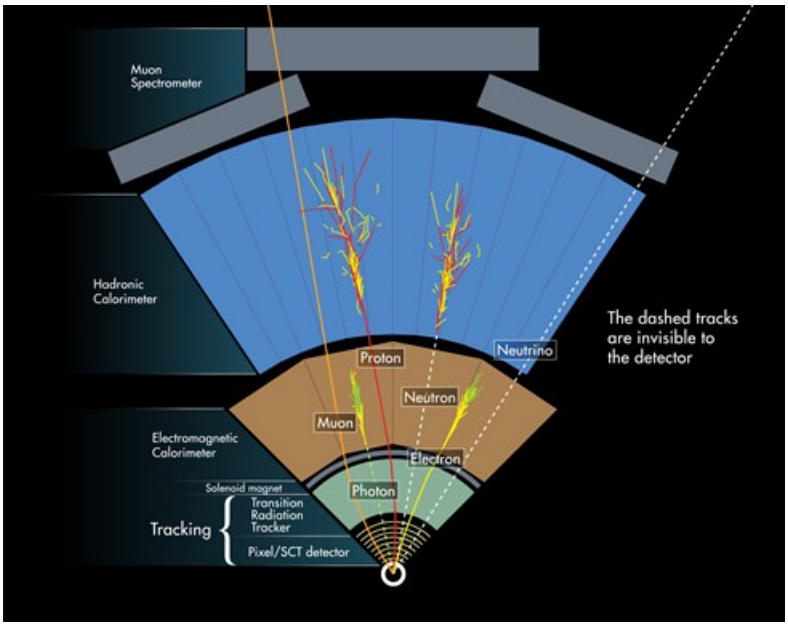


Figure 3.3.: Tracks, as they are measured in the respective sub-system in the ATLAS detector. ©CERN

sensitive to basic signatures such as electrons and muons with high transverse momentum, large energy deposits in the calorimeters or hadronic tau lepton decay signatures. The second level trigger accesses more information than the first level triggered on. Based on this, an event can be rejected or passed on to the high-level trigger. The high-level trigger uses the full information of the data event, and based on this selects the events to be stored permanently. Through this chain of triggers, in four seconds the event rate is reduced by a factor of $4 \cdot 10^4$ to about 1 kHz [45].

3.3. Data Samples and Monte Carlo Simulation

The data samples used in this analysis are the full datasets passing the data quality requirements[2] of the years 2015 and 2016 which were produced in pp collisions in the LHC and taken at the ATLAS detector, and correspond to an integrated luminosity of $\int \mathcal{L}\,\mathrm{d}t = 36.1\,\mathrm{fb}^{-1}$. However, only events with a total luminosity of $\int \mathcal{L}\,\mathrm{d}t = 14.61\,\mathrm{fb}^{-1}$ were used in the final results due to the implemented trigger. For details, see Section 4.2. In contrast to data, the sample files for all expected processes in signal and background

[2]The Good Run List (GRL) used is `https://atlasdqm.web.cern.ch/atlasdqm/grlgen/All_Good/data15_13TeV.periodAllYear_DetStatus-v79-repro20-02_DQDefects-00-02-02_PHYS_StandardGRL_All_Good_25ns.xml` for the 2015 dataset and `https://atlasdqm.web.cern.ch/atlasdqm/grlgen/All_Good/data16_13TeV.periodAllYear_DetStatus-v88-pro20-21_DQDefects-00-02-04_PHYS_StandardGRL_All_Good_25ns.xml` for the 2016 dataset.

are modelled with Monte Carlo (MC) simulations. The list of all files used in this analysis can be found in Appendix B.

This simulation is done in multiple steps. First, events are generated on parton level, where the bare quarks are present before building jets and hadronising to their hadron final states. In the next step, the parton shower and hadronisation process are simulated. As a final step, the resulting physics events are passed through a detector simulation.

In all generated samples, the top quark mass is set to 172.5 GeV. The detector response to particles is modelled by a dedicated simulation [46], either completely with GEANT [47] or based on a faster parametrisation for the calorimeter response and then with GEANT. To simulate pile-up in the form of additional pp interactions from the same or close-by bunch crossings, a set of minimum-bias interactions are generated using PYTHIA 8 [48]. To ensure the same pile-up levels in the simulated events that are present in the data, the distribution of the number of additional pp interactions in the MC samples is reweighted to match the one in the data. The samples are processed by the same reconstruction software as the data and the simulated events are corrected so that the object identification, reconstruction and trigger efficiencies, energy scales, and energy resolutions match those which are determined from data control samples.

The associated production of a top quark pair with a Z boson is generated at next-to leading-order (NLO) with `MG5_aMCNLO` interfaced to PYTHIA 8. The $t\bar{t}Z$ signal samples are normalised to the NLO QCD+EWK cross section calculation using fixed scale ($m_t + m_Z/2$) [49]. For the matrix elements, a dynamic scale is used ($H_T/2$) and the PDF set is `NNPDF3.0NLO` [50]. The `A14` tune [51] is applied together with the `NNPDF2.3LO` PDF set.

For the generation of the $t\bar{t}$ samples at NLO, the POWHEG-BOX v2 generator is used [52]. The parton shower, fragmentation, and the underlying event are simulated using the `NNPDF3.0` PDF set and PYTHIA 8 for the $t\bar{t}$ sample. The samples are normalised to their next-to-next-to-leading order (NNLO) cross sections including the resummation of soft gluon emission at next-to-next-to-leading-logarithmic (NNLL) accuracy using `Top++2.0` [53].

The t-channel production of a single top quark, as well as the W associated producation, is generated at leading order (LO) in QCD using `MG5_aMCNLO` interfaced with PYTHIA 6 [54] with the `CTEQ6L1` PDF [27] set and the Perugia2012 (P2012) tune [55]. For this, the four-flavour PDF scheme, treating the b quark as massive, is used. The t-channel single top samples are normalised to the NLO cross sections [56, 57], while the W associated production samples are normalised to the NNLO cross sections [58, 59].

Diboson processes with two charged leptons and two neutrinos ($\ell\ell\nu\nu$), as well as diboson processes having additional hadronic contributions ($\ell\ell\nu\nu jj$, $gg\ell\ell\nu\nu$) are simulated

using the SHERPA 2.1 generator [60]. Matrix elements contain all diagrams with four electroweak vertices. They are calculated using the COMIX [61] and OPENLOOPS [62] matrix element generators for up to one (4ℓ; $2\ell+2\nu$) or no additional partons ($3\ell+1\nu$) at NLO and up to three partons at LO and merged with the SHERPA parton shower using the `ME+PSNLO` prescription [63]. The `CT10` PDF set is used in conjunction with a dedicated parton shower tuning. The NLO cross sections calculated by the generator are used to normalise diboson processes.

The production of three massive vector bosons (triboson samples) with subsequent leptonic decays of all three bosons is modelled at LO with the SHERPA 2.1 generator. Up to two additional partons are included in the matrix element at LO.

Events simulating Z or W bosons with associated jets are modelled using the SHERPA 2.2.1 generator (for Z+jets and W+jets) or SHERPA 2.2 (for W+jets). The `NNPDF3.0` PDF set is used in conjunction with a dedicated parton shower tuning developed by the SHERPA authors. The samples are normalised to the NNLO cross sections [64–68].

An additional set of W+jets samples is modelled using SHERPA 2.1 as a generator and for showering. Then, the `CT10` PDF set is used in conjunction with a dedicated parton shower tuning. These samples are normalised to the LO cross sections.

The multijet Monte Carlo samples were generated using PYTHIA 8 for the matrix element and showering and are normalised to the LO cross sections.

Decays into heavy flavours for all signal samples and the background single top, $t\bar{t}$ and multijet samples, are modelled using the EvtGen [69] program. All other samples are produced using the SHERPA generator [60].

Table 3.1 gives a summary of the basic parameters of the MC samples used in the analysis. All MC samples used are from the MC15C production campaign. The complete list of all data and MC samples can be found in Appendix B.

Sample	**Generator**	**Shower**	**PDF**	**Normalisation**	σ [pb]
$t\bar{t}(Z \to \nu\bar{\nu})$	aMC@NLO	Pythia 8	NNPDF3.0	NLO	0.1549
$t\bar{t}(Z \to qq)$	aMC@NLO	Pythia 8	NNPDF3.0	NLO	0.52809
$t\bar{t}(Z \to \tau^-\tau^+)$	aMC@NLO	Pythia 8	NNPDF3.0	NLO	0.0036515
$t\bar{t}$	Powheg	Pythia 8	NNPDF3.0	NLO	646.86
Z+jets	Sherpa 2.2.1	Sherpa 2.2.1	NNPDF3.0	NNLO	41246.7
W+jets	Sherpa 2.2.1	Sherpa 2.2.1	NNPDF3.0	NNLO	61201.9
	Sherpa 2.1	Sherpa 2.1	CT10	LO	31.7
Single top	Powheg	Pythia 6	CT10	NLO/NNLO	137.52
Multiboson	Sherpa 2.1	Sherpa 2.1	CT10	NLO	119.23
multijet	Pythia 8	Pythia 8 (+EvtGen)	NNPDF3.0	LO	$80 \cdot 10^9$

Table 3.1.: Monte Carlo samples and generators used for the event production, showering and hadronization in the detector including the used PDF, level of normalisation, and corresponding cross section.

4. Sample Preparation

This chapter defines physical objects present in the collision data, before defining which high level trigger to use and event selection criteria. At the end, this event selection is defines as signal region, so that control regions are derived for use in the boosted decision tree which is evaluated later in this thesis.

4.1. Object Definition

The relevant signature of $t\bar{t}Z$ for this analysis, consists of four light jets, two b jets, and missing transverse energy. It does not contain any electrons or muons, however, it is important to define them for exclusion. Tau leptons are not reconstructed because they are not identified as such. This means that their decay final states are included in the samples as electrons, muons, or hadrons. In the following section, these objects are defined as they are identified in the ATLAS detector. When speaking of an electron or muon here, the definition is regardless of charge and also valid for the respective antiparticles.

An electron is identified as such if its signature corresponds to the image in Figure 3.3. It leaves a bent track in the inner detector and is stopped in the electromagnetic calorimeter where it deposits the largest fraction of its energy. In addition, it has to satisfy the `MediumLH` identification criteria for all regions [70]. Electrons are required to satisfy $p_T > 15\,\text{GeV}$ and be reconstructed in the central region of the detector with $|\eta| \leq 2.47$. Furthermore, they are required to fulfil the recommended impact parameter cuts $|d_0|/\sigma(d_0) < 5$, $z_0 \sin(\theta) < 0.5\,\text{mm}$. The electron isolation is required to follow the `FixedCutTight` working point for all regions.

Muons, as shown in Figure 3.3, deposit little energy in the calorimeters, so they are mostly identified by their reactions in the muon spectrometer. They are reconstructed using the unified muon identification chain which uses both, information from the inner detector and the muon spectrometer. Similar to the electrons, they are required to have $p_T > 15\,\text{GeV}$, lie in a region in the detector with $|\eta| \leq 2.5$, require an isolation following the `FixedCutTight` working point and fulfil the recommended impact parameter cuts $|d_0|/\sigma(d_0) < 3$, $z_0 \sin(\theta) < 0.5\,\text{mm}$. They are also required to pass the medium track

quality requirements. For isolation, the `FixedCutTightTrackOnly` criteria are used, which requires the total sum of track transverse momentum around the muon candidate within a cone of $\min(10\,\text{GeV}/p_T^{\text{candidate}}, 0.2)$ to be less than 6% of the respective $p_T^{\text{candidate}}$.

Except for the top quark, quarks hadronise after being produced. They form bound states and showers of mostly soft, charged components which leave signatures in the tracking detectors, but also deposit their energy partly in the electromagnetic and mostly in the hadronic calorimeter. Jets are identified as such with the anti-k_t algorithm [71] with a distance parameter of $R = 0.4$ (in η-ϕ space) and are reconstructed from topological calorimeter clusters [72]. The jet calibration is done with the so-called `EMTopo` scheme. To be identified as a jet, the candidate is required to have $p_T > 25\,\text{GeV}$ and $|\eta| < 2.5$. The Jet Vertex Tagger (JVT) [73] discriminant is required to be larger than 0.59 for jets with $p_T < 60\,\text{GeV}$ and $|\eta| < 2.5$ in order to reject jets arising from pileup collisions, as recommended by the ATLAS JetEtMiss group.

Up to a certain point, it is possible to distinguish b jets from lighter jets. Here, they are identified using the MV2c10 algorithm [74], which is based on a boosted decision tree (BDT) using the output weights of the JetFitter, IP3D and SV1 algorithms as input. The 77% efficiency working point corresponds to the fixed cut value on MV2c10 output > 0.645925. With this, an average of 77% efficiency in tagging b jets from top quark pair production is achieved and a factor of 6.21 and 134.34 rejection for c-jets and light-jets, respectively.

Missing transverse energy, $\not{E}_T$, is the magnitude of the missing transverse momentum vector p_T^{miss} which is calculated with an algorithm based on the "MetRefFinal" definition:

$$\left(p_T^{\text{miss}}\right) = \left(p_T^{\text{miss}}\right)^{\text{RefEle}} + \left(p_T^{\text{miss}}\right)^{\text{RefMuon}} + \left(p_T^{\text{miss}}\right)^{\text{RefJet}} + \left(p_T^{\text{miss}}\right)^{\text{SoftTerm}}, \tag{4.1}$$

which is a summation of terms obtained from the respective vectorial sum of the transverse momenta of preselected electrons, muons and jets as defined above. The soft term $(p_T^{\text{miss}})^{\text{SoftTerm}}$ is derived from tracks not assigned to any reconstructed object.

To avoid overlap of objects with the above definitions, overlap removal in the final state objects is necessary. This is done by the `harmonized` option [75] in the `AssociationUtils` package in the following sequence:

- when electron and muon candidates share a track, the electron is removed.
- At a distance $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.2$ between a jet and a baseline electron, the jet is dropped. If this is true for more than one jet, only the closest jet is dropped.
- At a distance of $0.2 < \Delta R < 0.4$ between a jet and a baseline electron, the electron is dropped.

2015 Trigger	2016 Trigger	Cut Details	MC Events	Norm. Events
-	-	basic event selection	1267866	6102
HLT_ht700_L1J75	HLT_ht1000	$H_T > 1000\,\text{GeV}$	111526	374
HLT_5j60	HLT_5j70_L14J15	≥ 5 jets with $p_T > 70\,\text{GeV}$	214785	860
HLT_xe70_mht	HLT_xe110_mht_L1XE50	$\not{E}_T > 110\,\text{GeV}$	580691	2481
HLT_xe70_mht	HLT_xe90_mht_L1XE50	$\not{E}_T > 90\,\text{GeV}$	739059	1316

Table 4.1.: Evaluation of the optimal available trigger, shown on the signal samples. The numbers represent the surviving events of the combination of 2015 and 2016 samples with no trigger applied, when the respective cut of the trigger is applied to the samples in addition to the basic event selection. The basic event selection is determined in Section 4.3, before applying cuts in $\Delta\phi_1$ and $\Delta\phi_2$. For the samples with only the basic event selection applied, as well as for the samples with the H_T and jet trigger applied, the $\not{E}_T$ cut of the event selection is lowered to 50 GeV. The normalised event numbers of the upper rows are determined using an integrated luminosity of $36.1\,\text{fb}^{-1}$, while the normalised event numbers in the last row is determined using a luminosity of $14.61\,\text{fb}^{-1}$.

- At a distance $\Delta R < 0.4$ between a jet and a baseline muon, the muon is dropped if the jet has more than 2 associated tracks, otherwise the jet is removed.

4.2. High Level Trigger

While the level 1 and level 2 triggers from Section 3.2 are already implemented in every event recorded in the detector, the high level trigger is applied to the buffered data afterwards.

Which trigger is the most appropriate depends on the signature of the signal. The channel in this thesis has a significant missing transverse energy which is not present in the dominant backgrounds of $t\bar{t}$ and multijet processes. In addition to that it includes four light jets and two b jets as a signature. With this, options for the trigger are $\not{E}_T$, the number of jets in combination with a minimum transverse momentum cut, as well as H_T, which is the sum of the transverse momenta of all final decay products in the event.

In Table 4.1, the respective unprescaled triggers best matching the signature are listed. Using the combination of 2015 and 2016 Monte Carlo samples without any trigger applied, the highest cut of the two triggers for 2015 and 2016 is applied to the set in addition to the basic event selection explained in Section 4.3. The results are then listed in Table 4.1.

It is obvious that none of the triggers is a perfect choice, all of them cut away a large fraction of the signal. Noting this, the $\not{E}_T$ triggers cut away the least amount of events, making them the trigger of choice for this analysis. So, for the 2015 samples, the trigger `HLT_xe70_mht` is chosen.

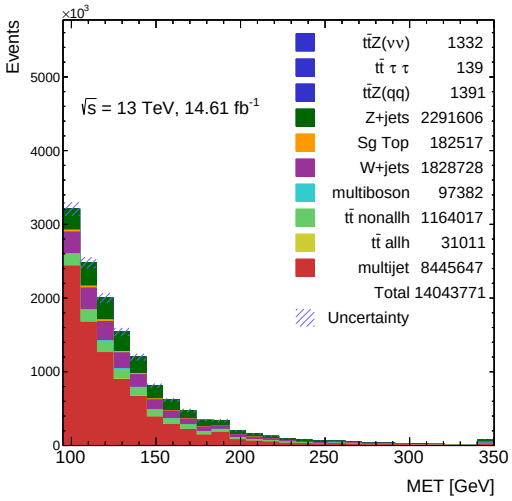
The respective trigger for the 2016 samples `HLT_xe110_mht_L1XE50`, however, is found to not be simulated in the MC samples, making it impossible to use on a short time scale. After considering the low rates let through by the other triggers, the next best $\not{E}_T$ trigger `HLT_xe90_mht_L1XE50` (also listed in Table 4.1) is chosen, which as an additional benefit does not cut away half of the peak of the $\not{E}_T$ distribution in the signal (see Figure 4.2a) since it cuts at a lower value. The disadvantage of this trigger, and reason why it was not the recommended trigger for analyses, is that for run numbers 303892 and above in the 2016 dataset, the trigger is not implemented anymore, lowering the analysed integrated luminosity from possible 36.1 fb^{-1} to about half the value, which is 14.61 fb^{-1}.

4.3. Event Selection

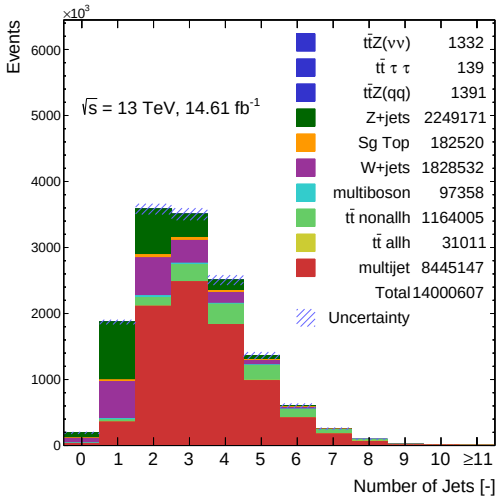
In this analysis, the `AnalysisTop` software (tag `2.4.32`) is used for pre-selection. The samples are split according to their correspondence to 2015 or 2016 data and the triggers evaluated in Section 4.2 are used which is `HLT_xe70_mht` for 2015 data and MC corresponding to 2015 data, and `HLT_xe90_mht_L1XE50` for the respective 2016 data and MC samples.

Besides that, the events have to be included in the Good Run List mentioned in Section 3.3, the `GOODCALO` selection is applied, which requires that the LAr and Tile are working properly for this event, ensuring there are no noise bursts. Every selected event is required to have a primary vertex, which is defined as the position of the hard scattering process, containing at least two tracks, which is the same as being labelled as `xAOD::VxType::PriVtx` defined by the Tracking CP group. To minimise contributions of fake jets originating from non-collision background or cosmic events, or from fake signals in the calorimeter originating from noise bursts or coherently present noise, jet cleaning is applied with the `LooseBad` selection of the `JetCleaningTool` after the overlap removal is included in the event. Here, all events where at least one jet fails the jet cleaning cut are rejected.

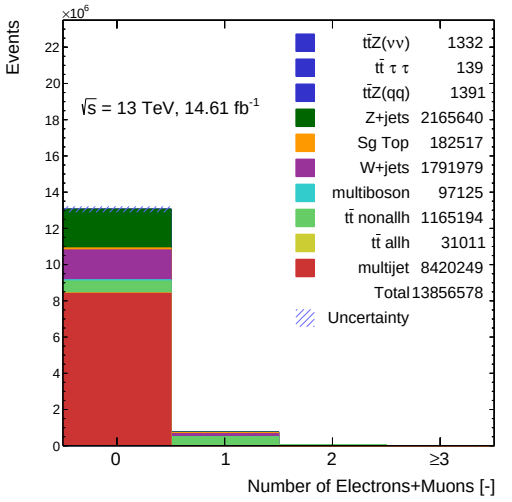
To normalise the Monte Carlo samples to data, event weights are applied, specifically a Monte Carlo weight, a pile-up weight and the b tag weight corresponding to the 77% working point.



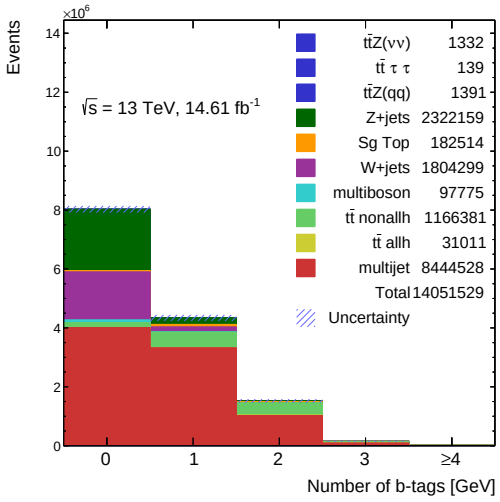
(a) $\not{E}_T$.



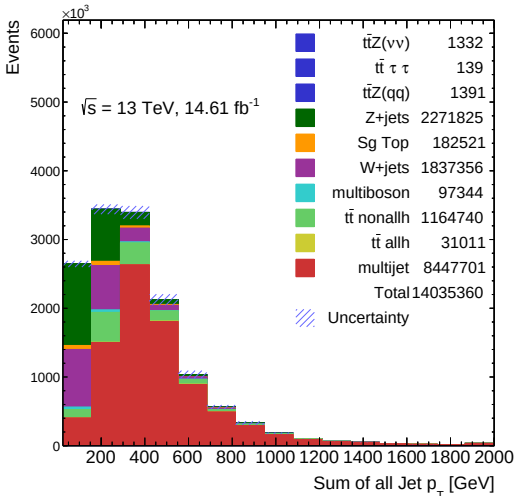
(b) Number of jets.



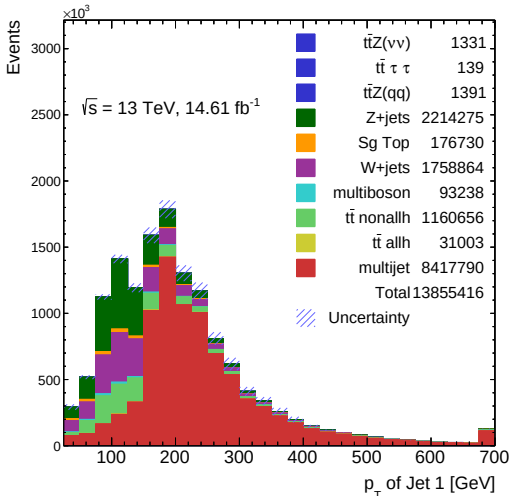
(c) Number of charged leptons.



(d) Number of b tags.

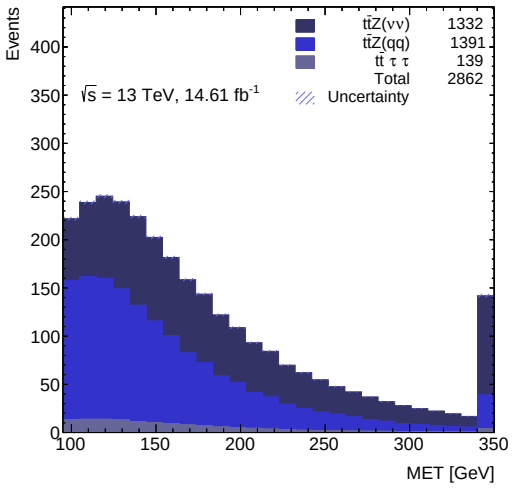


(e) H_T jets.

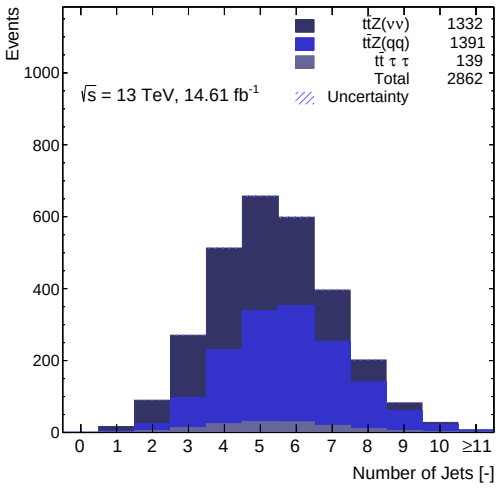


(f) p_T of leading jet.

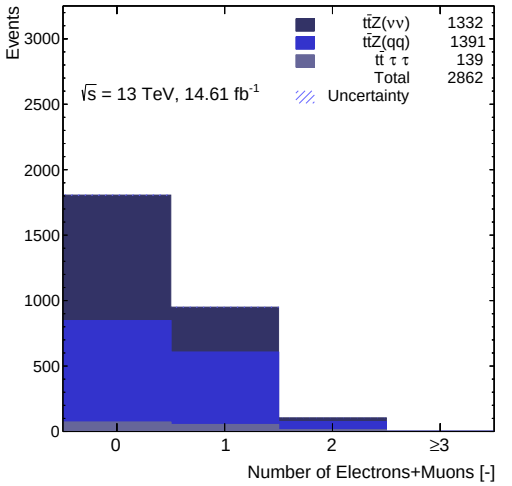
Figure 4.1.: Signal and background distribution of various observables before applying event selection cuts, but with the requirement $\not{E}_T \geq 95\,\text{GeV}$. Charged leptons, here, refers to either electrons or muons. Distributions of the number of electrons, muons and the p_T distributions of the second to fifth jet can be found in the appendix in Figure A.1.



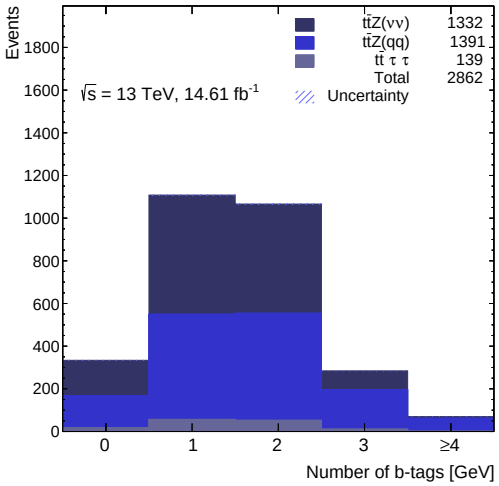
(a) $\not{E}_T$.



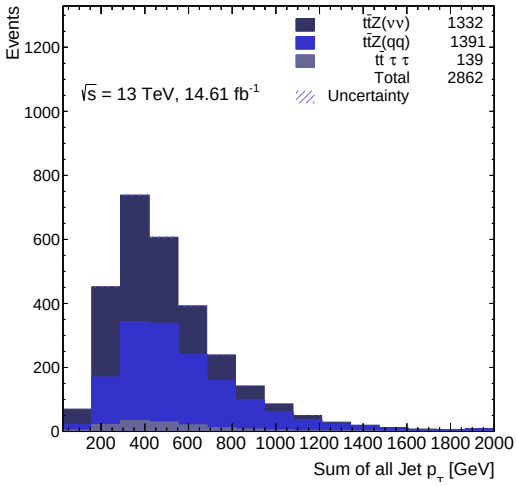
(b) Number of jets.



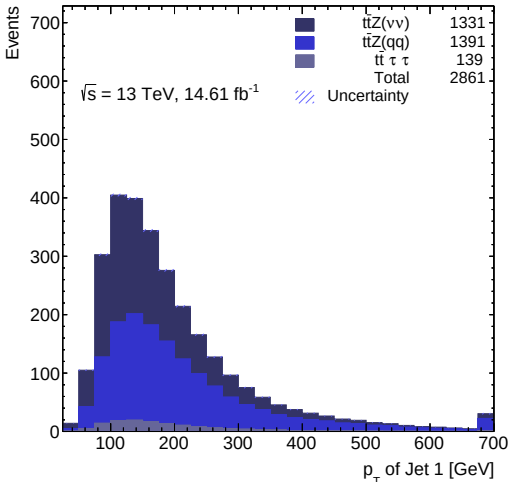
(c) Number of charged leptons.



(d) Number of b tags.



(e) H_T jets.



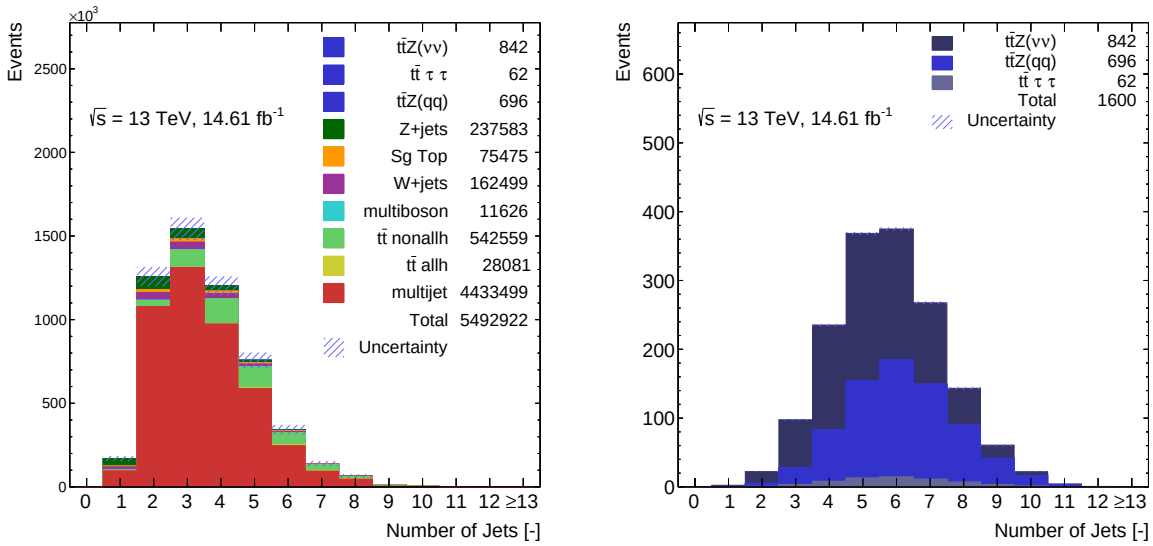
(f) p_T of leading jet.

Figure 4.2.: Signal distribution of various observables before applying event selection cuts, but with the requirement $\not{E}_T \geq 95\,\text{GeV}$. Charged leptons, here, refers to either electrons or muons. Distributions of the number of electrons, muons and the p_T distributions of the second to fifth jet can be found in the appendix in Figure A.2.

The ideal detected signal event contains six jets, of which two originate from b quarks, and large missing transverse momentum. Due to the trigger cut at an $\not{E}_T$ value of 90 GeV, an event selected for this analysis is required to have $\not{E}_T \geq 95$ GeV. Further cuts to increase the signal efficiency are determined by looking at relevant observables, which are the missing transverse energy, number of jets, number of b tags, H_T jets, as well as p_T of the leading jet, shown in Figure 4.1 for signal and background and in Figure 4.2 for signal only.

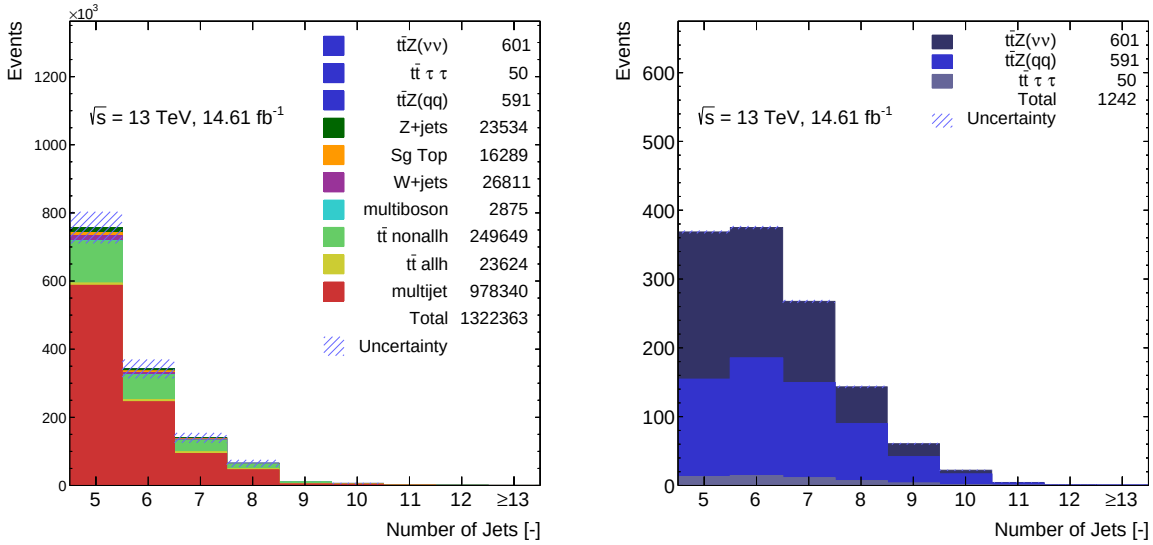
Since the physical process contains no charged leptons except for possible leptonic decays in jets, this is the first variable to cut on in compliance with Figure 4.2c, an event passing the event selection is required to have no muons and no electrons. With two b jets in the signal event coming from the $t\bar{t}$ decay, and considering the distributions of the number of b tags in Figure 4.2d and Figure 4.1d, requiring at least one b tag at 77% working point increases the signal to background ratio significantly.

When looking at Figure 4.3, it is apparent that the background samples have lower numbers of jets per event, while the signal distribution has its peak at the expected value of six jets. Therefore, the requirement of having at least five jets to pass the selection is implemented. With the combination of these cuts, the events contained in Figure 4.4 remain.



(a) Signal and Background. ***(b)*** Signal only.

Figure 4.3.: Signal and background, and signal only distributions of the number of jets after applying the cuts of $\not{E}_T > 95$ GeV, no electrons or muons, and at least one b tag.



(a) Signal and Background. ***(b)*** Signal only.

Figure 4.4.: Signal and background, and signal only distributions of the number of jets after applying the cuts of $\not{E}_T > 95\,\text{GeV}$, no electrons or muons, at least one b tag, and at least five jets.

Sample	**total Events**	**weights > 100**	**average pos. weight**
Multijet	49,820	114	6.9
Single Top	113,404	0	0.11
Multiboson	57,847	0	0.040
W+jets	326,559	1	0.067
Z+jets	856,693	0	0.038
$t\bar{t}$ allhad	21,347	0	0.59
$t\bar{t}$ nonallhad	1,734,129	0	0.12
$t\bar{t}(Z \to \nu\bar{\nu})$	354,507	0	0.0036
$t\bar{t}(Z \to qq)$	165,178	0	0.0062
$t\bar{t}(Z \to \tau^-\tau^+$	65,432	0	0.0014

Table 4.2.: Number of events which pass the final event selection from the end of this section in the respective samples, number of these events with positive event weights greater than a value of 100, and the average value of the positive weight (which corresponds to the contribution of the respective event to the final number of normalised events) in the different samples.

There are still large numbers of background, especially the possibly problematic multijet background, remaining. The QCD multijet Monte Carlo has significantly larger event weights, and number of events with a weight greater than 100 as shown in Table 4.2. These events cannot be removed from the analysis, as internally recommended by the ATLAS Physics Modelling Group, because that would also remove all multijet events in lower p_T regions, which are the crucial background events here.

To further lower the background, and especially the multijet background, the variables

$$\Delta\phi_1 = |\phi^{p_t^{\text{miss}}} - \phi^{\text{leading jet}}| \qquad \text{and} \qquad \Delta\phi_2 = |\phi^{p_t^{\text{miss}}} - \phi^{\text{subleading jet}}| \tag{4.2}$$

are evaluated for a possible cut. The distributions of $\Delta\phi_1$ for both all samples and signal samples only are shown in Figure 4.5, and of $\Delta\phi_2$ in Figure 4.6. Table 4.3 and Table 4.4 show the signal over background significance, calculated by the number of signal events divided by the square root of the total number of background events, and the signal over QCD multijet background significance, respectively, for a preliminary dataset. To determine the ideal cut, both values should be as high as possible to increase the signal purity by reducing the dominant background while keeping enough statistics of the small number of signal events. With this, the cuts

$$\Delta\phi_1 > 0.5 \qquad \text{and} \qquad \Delta\phi_2 > 0.6 \tag{4.3}$$

are determined. Another observable of interest in later parts of the analysis will be $\Delta\phi_{\text{closest}}$ which is defined in analogy to $\Delta\phi_1$ and $\Delta\phi_2$ but gives the minimum value of $\Delta\phi$ with any jet in the event.

	$\Delta\phi_2 > 1.2$	$\Delta\phi_2 > 1.0$	$\Delta\phi_2 > 0.8$	$\Delta\phi_2 > 0.6$	$\Delta\phi_2 > 0.5$	$\Delta\phi_2 > 0.4$
$\Delta\phi_1 > 1.2$	0.9854	1.0002	1.0200	1.0265	1.0303	1.0325
$\Delta\phi_1 > 1.0$	0.9941	1.0087	1.0283	1.0357	1.0384	1.0405
$\Delta\phi_1 > 0.8$	1.0056	1.0199	1.0392	1.0452	1.0488	1.0517
$\Delta\phi_1 > 0.6$	1.0089	1.0233	1.0425	1.0498	1.0524	1.0544
$\Delta\phi_1 > 0.5$	1.0104	1.0248	1.0441	1.0515	1.0542	1.0572
$\Delta\phi_1 > 0.4$	1.0088	1.0257	1.0439	1.0526	1.0548	1.0686

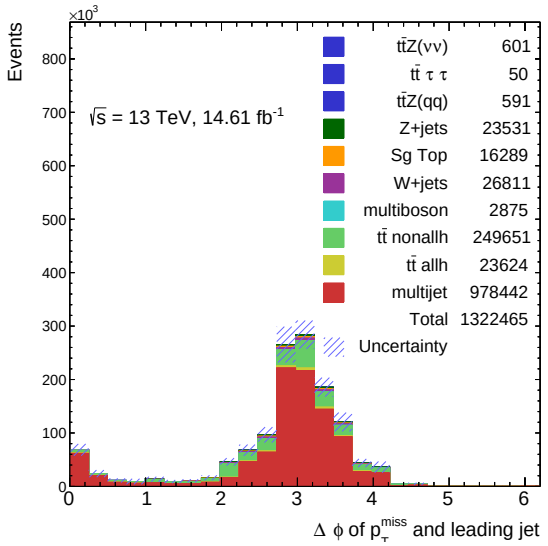
Table 4.3.: Signal over background significance for various cuts on $\Delta\phi_1$ and $\Delta\phi_2$ calculated by the number of signal events divided by the square root of the total number of background events. Background, here, means all background samples including QCD multijets background. These numbers have been obtained with a preliminary dataset which did not include all events used for the final analysis.

	$\Delta\phi_2 > 1.2$	$\Delta\phi_2 > 1.0$	$\Delta\phi_2 > 0.8$	$\Delta\phi_2 > 0.6$	$\Delta\phi_2 > 0.5$	$\Delta\phi_2 > 0.4$
$\Delta\phi_1 > 1.2$	1.7064	1.7029	1.7269	1.6953	1.6762	1.6547
$\Delta\phi_1 > 1.0$	1.7175	1.7153	1.7391	1.7096	1.6889	1.6674
$\Delta\phi_1 > 0.8$	1.7448	1.7412	1.7641	1.7312	1.7113	1.6906
$\Delta\phi_1 > 0.6$	1.7383	1.7366	1.7604	1.7315	1.7110	1.6877
$\Delta\phi_1 > 0.5$	1.7352	1.7343	1.7585	1.7307	1.7108	1.6916
$\Delta\phi_1 > 0.4$	1.7205	1.7253	1.7485	1.7246	1.7058	1.6859

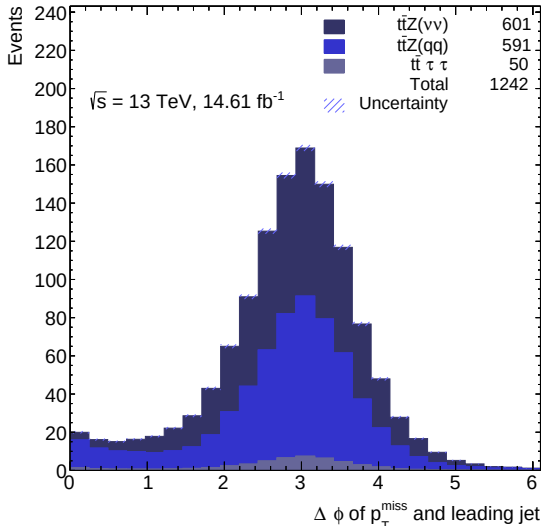
Table 4.4.: Signal over QCD multijet background significance for various cuts on $\Delta\phi_1$ and $\Delta\phi_2$ calculated by the number of signal events divided by the square root of the number of QCD multijet background events. These numbers have been obtained with a preliminary dataset which did not include all events used for the final analysis.

Sample	**Total Events**	**Passing Selection**	**Normalised**
Multijet	5,230,333	49,820	341,227
Single Top	2,806,374	113,404	12,051
Multiboson	5,201,067	57,847	2,276
W+jets	16,206,895	326,559	20,201
Z+jets	44,265,287	856,693	19,180
$t\bar{t}$ allhad	171,844	21,347	12,352
$t\bar{t}$ nonallhad	15,898,396	1,734,129	197,703
Signal	1,944,015	1,174,017	1,011

Table 4.5.: Number of events in each sample after applying the event selection cuts.

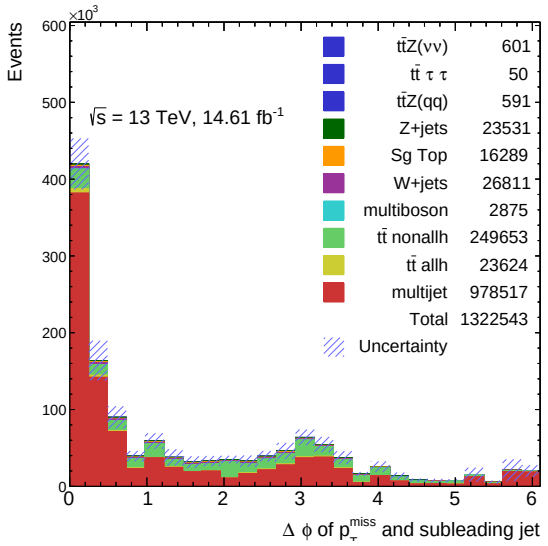


(a) Signal and Background.

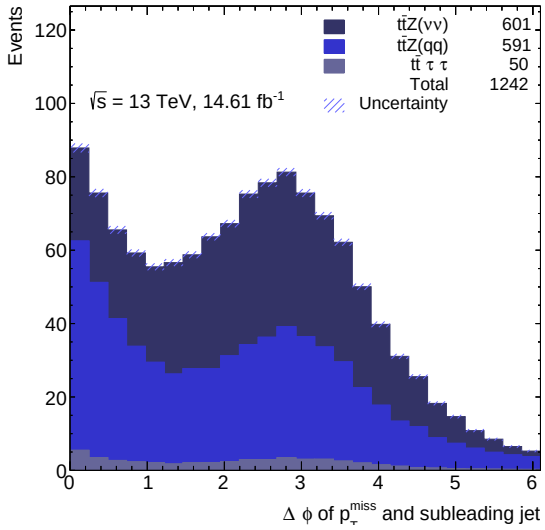


(b) Signal only.

Figure 4.5.: Signal and background and signal only distributions of $\Delta\phi_1$ after applying first event selection cuts of $\not{E}_T \geq 95\,\text{GeV}$, at least one b tag, no electrons or muons and at least five jets.



(a) Signal and Background.



(b) Signal only.

Figure 4.6.: Signal and background and signal only distributions of $\Delta\phi_2$ after applying first event selection cuts of $\not{E}_T \geq 95\,\text{GeV}$, at least one b tag, no electrons or muons and at least five jets.

4.4. Signal, Control and Validation Regions

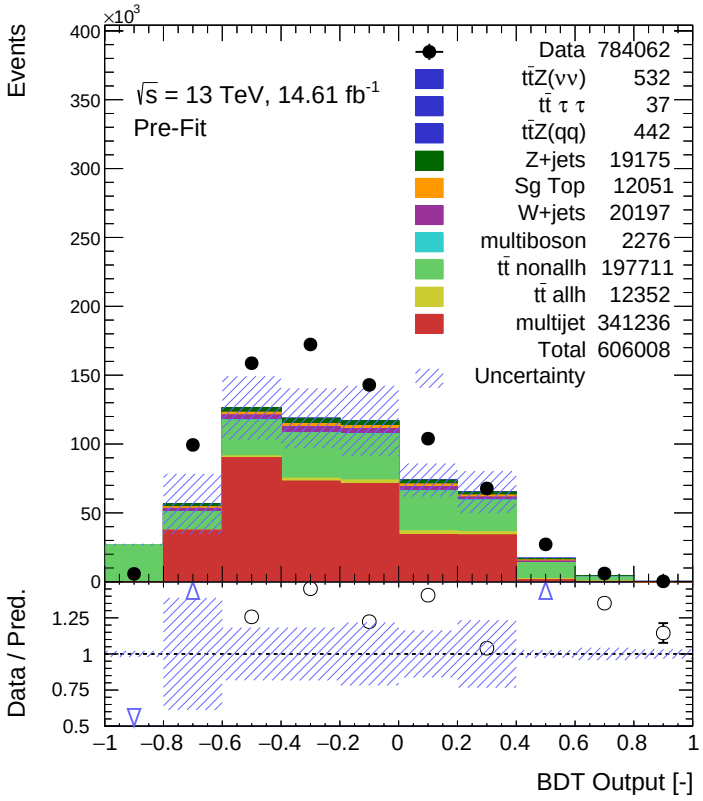


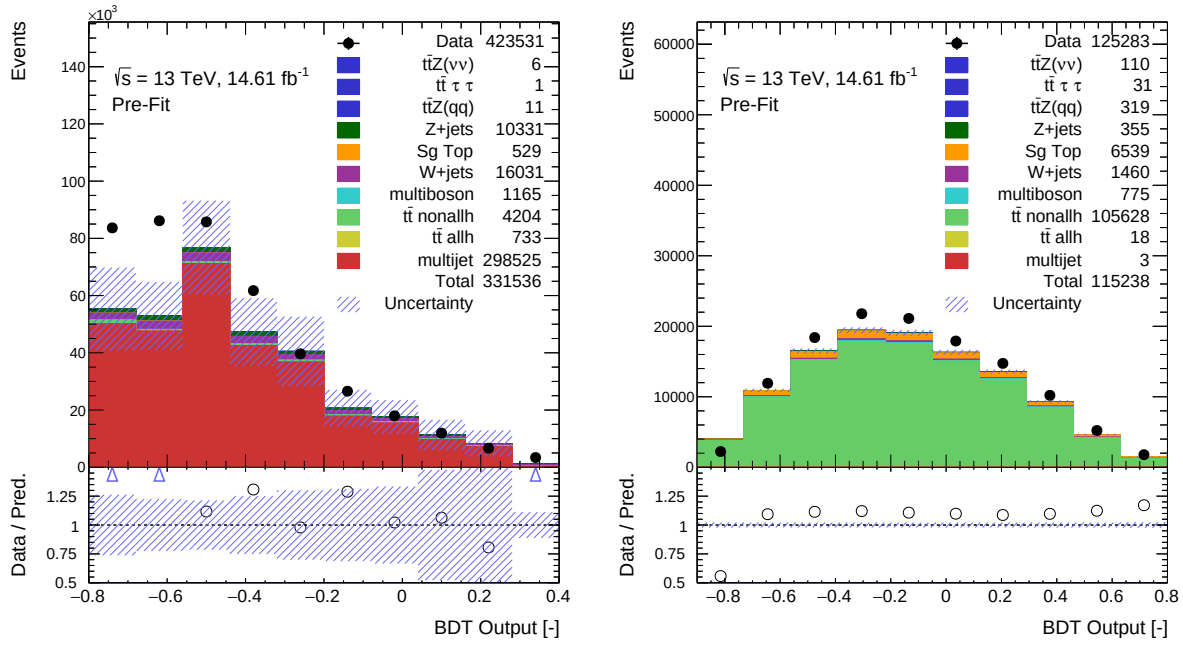
Figure 4.7.: The preliminary signal region defined by the BDT output and the cuts from Chapter 4.3. To reduce the influence of QCD multijet uncertainties, a cut of the region at 0.6 is performed in Section 6.3.

The samples passing the event selection explained in Section 4.3 are processed by a boosted decision tree (BDT). The concept of the BDT is explained in Section 5.1. The output of the BDT as explained in Chapter 5.1 and evaluated in Chapter 6.1 is the preliminary signal region for the statistical fit as explained in Section 5.2 and evaluated in Section 6.3. As shown in Figure 4.7, the contribution of the QCD multijet Monte Carlo is dominant in the lower regions of the histogram. Because of the large uncertainties of this background, it is desired to have as little contribution to the region used for the fit as possible. To achieve this, the final signal region is defined as the BDT output in the range from 0.6 to 1.0.

The largest background contributions originate from QCD multijets and the non-fully-hadronic $t\bar{t}$ samples. For these, separate control regions are defined to be able to evaluate the quality of a fit on specific samples. To get a more accurate multijet background estimate, the initial event selection cut on b tags is set to 0, as taken from Figure 4.1d.

In addition, the cut on $\Delta\phi_2$ is reversed to $\Delta\phi_2 < 0.6$, which is suggested by Figure 4.6a. The final multijet control region is shown in Figure 4.8a, where the first and last bin are the under- and overflow bins, respectively, so that the plot shows all events of this region.

For the fully hadronic $t\bar{t}$ control region, it is sufficient to invert the restriction on electrons. With the information from Figure 4.1c, the requirement of at least one charged lepton (electron or muon) has to be present in the event is deduced. The other event selection cuts defined in Section 4.3 remain the same. With this, the final BDT distribution in the fully hadronic $t\bar{t}$ control region is shown in Figure 4.8b. Here, the last bin is the overflow bin, while the first bin has been set to start at a value of -0.9. This is to exclude the lowest BDT region due to mismodelling of the MC compared to data in that region. It is recommended for future analyses that the BDT is also trained for the control regions to help to resolve this. In addition, there is a large contribution of the signal remaining in this region. This is because there is also a significant fraction of the signal sample containing one or more charged leptons, as visible in Figure 4.2c.



(a) QCD multijet. ***(b)*** allhadronic $t\bar{t}$.

Figure 4.8.: Control regions of the in Section 6.1 evaluated final BDT output for the QCD multijet and allhadronic $t\bar{t}$ background.

5. Analysis Method

This chapter explains the analysis strategies of a boosted decision tree which is going to be evaluated on the analysis data, and the fitting process which is going to be used on the output of the tree in order to extract the signal.

5.1. Boosted Decision Trees

A first approach to separate signal from background in this analysis has been conducted by implementing a boosted decision tree (BDT), which is the topic of this section.

Due to the complex nature of the signal, a more sophisticated method than a simple cut analysis is needed for the separation. For this, a multivariate discriminant is ideal, to take various observables and their correlations into account by expanding the one-dimensional diagram of a cut analysis into a multi-dimensional one, with geometrically complex cuts being applied. Due to the complexity of the patterns, a machine learning algorithm such as a boosted decision tree is implemented for that.

5.1.1. Decision Trees

A Decision Tree, as shown in Figure 5.1, consists of different so-called *nodes*, where a question about the event is asked and the answer determines the next node or the final classification. In this example, the initial node, the root node, asks for the value of xi. If for the event in question the value of xi is greater than a constant $c1$, the next node is the lower left daughter node. If the value of xi is smaller than $c1$, the consecutive node would be the lower right daughter node. These kind of questions are characteristics of binary trees like the ones used for this thesis.

The questions are not given to the algorithm beforehand, but instead, the decision tree is trained. In this case, it is presented with Monte Carlo samples split into signal and background. The tree then learns the pattern of what the signal and background looks like, and optimises the cut parameters in the individual nodes. After going through all the decision nodes, the event ends on a leaf node which is classified as either a signal or background leaf node, which is why the algorithm is also called a classifier.

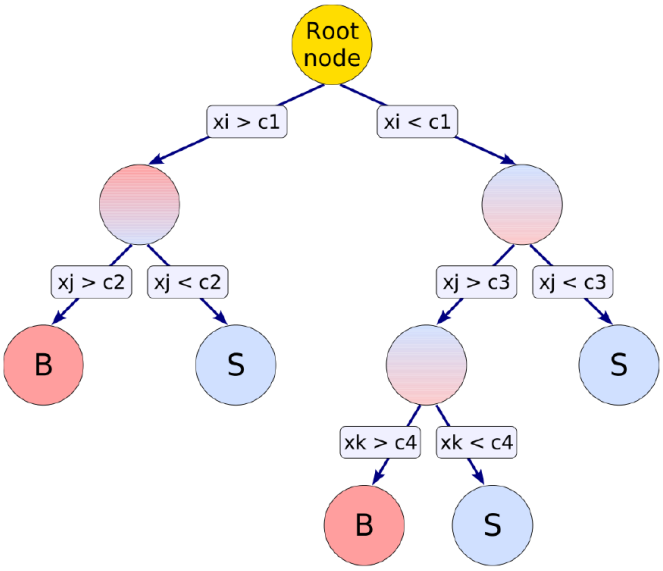


Figure 5.1.: Exemplary structure of a decision tree.

The path from the root node to one of the leaf nodes, corresponds to a single cut analysis. Since the whole tree is a combination of many cut analyses, the criteria of the final tree are simple and clear to see. This multitude of branches naturally makes the tree superior to a standard cut analysis.

This, however, is also one of the weaknesses of the decision tree. Maximising the effectiveness of the cuts for the event classification on all variables at the same time is computationally difficult to achieve. Usually, a decision tree is built starting from the root node. At every new node, the variable with the greatest separation gain at this point of the tree is used for the cut. The default to determine the greatest separation gain is the Gini Index defined as $p \cdot (1 - p)$ where p is the signal purity of the node which is defined as $S/(S + B)$ with S and B being the number of signal and background events, respectively. This way it is possible that several variables are used at multiple nodes, while other variables are not used at all in this tree.

This way, the tree is very unstable to statistical fluctuations. If a fluctuation in the training sample is caused for selecting a cut variable in an individual node, all the following nodes propagate that error. A technique to lower this instability is called boosting.

5.1.2. Boosting

Here, a so-called forest of decision trees is built which then vote on if an event is classified as signal or background. The separation performance is increased and statistical errors are lowered significantly when the classification is not only dependent on one tree but

averaged over many. This comes at the cost of losing the clear interpretation of cuts in the decision tree with increasing number of trees.

The individual trees in a forest are trained with the same training sample, although the event weights of the input sample are modified based on whether the event was correctly or incorrectly classified in the previous tree. With this boost weight, α, the next tree serves the purpose of compensating for and improving on the mistakes of the previous tree.

The boosting usually performs best when applied to trees with small individual classification power. These weak classifiers are typically smaller trees limited in growth, with depth ranging from two to six nodes. When limiting the tree depth during the training, the overtraining of trees (see Sec. 5.1.3) with large depth and pruning applied is close to eliminated. Pruning is when statistically insignificant nodes are removed after the tree has been fully grown. This, however, is usually not necessary for boosted decision trees, since it is more effective to train trees with small depths right away.

In this thesis, two different ways of boosting the decision trees are tested, the AdaBoost (adaptive boost) [76] and the Gradient Boost [77]. Both boosting methods use the boost weight α for the subsequent tree, which is defined as

$$\alpha = \frac{1 - \text{err}}{\text{err}} \tag{5.1}$$

where err is the misclassification rate of the previous tree. The misclassification rate is determined by the same events used for the training of the tree itself, so by definition, it is err ≤ 0.5. After applying this adjusted weight, all event weights are normalised so that the total sum of weights is the same as in the previous tree.

With x being the tuple of the input variables in the boosted decision tree, the result of an individual decision tree in the forest becomes $h(x)$, which takes values of ± 1 for signal or background, respectively. Then, the total boosted decision tree classifier becomes

$$y_{\text{Boost}}(x) = \frac{1}{N_{\text{collection}}} \sum_{i}^{N_{\text{collection}}} \ln(\alpha_i) \cdot h_i(x) \tag{5.2}$$

in the case of AdaBoosting. The sum here is over the complete collection, forest, of the classifiers. The value of $y_{\text{Boost}}(x)$ is smaller for background-like events, and larger for signal-like events.

For gradient boosting, the decision trees and the whole forest need to be represented mathematically. This is universally applicable also to AdaBoost BDTs, although due to the simplicity of the boosting there, it is not necessary. With the mathematical represen-

tation, the gradient boosting process, then, corresponds to a simple additive expansion approach.

The individual decision trees are described by functions $f(x, a_i)$, so called weak learners. The boosted decision tree then becomes the weighted sum of such base functions

$$F(x, P) = \sum_{i=0}^{N_{\text{collection}}} \beta_i f(x, a_i), \qquad P \in \{\beta_i; a_i\}_0^{N_{\text{collection}}} \tag{5.3}$$

with β_i being the weight of the individual tree. The boosting process now optimises P so that the difference between the mathematical model and the true value y from the training sample is minimised. This difference is described by the loss function $L(F, y)$ which fully determines the boosting procedure.

The loss function corresponding to AdaBoosting is

$$L(F, y) = e^{-F(x)y} \tag{5.4}$$

which then gives the boosting weights mentioned above. This exponential loss function is less stable in regards to outlying and mislabelled data points and thus does not perform well for noisy samples.

Without giving up the good performance of an exponential loss function, the gradient boosting algorithm implements more robust functions. In TMVA [78], a Toolkit for Multivariate Data Analysis implemented in ROOT [79], the gradient boosting is realised by the binomial log-likelihood loss

$$L(F, y) = \ln\left(1 + e^{-2F(x)y}\right) . \tag{5.5}$$

This corresponds to a complicated and difficult to obtain boosting algorithm, so the minimisation is done by a steepest-descent approach. For this, the current gradient of the loss function is calculated and stored. The decision tree is then grown in a way so that the leaf values match the stored mean value of the current gradient of the loss function. in each region of the tree structure. Because this value changes, the different decision trees of the collection are built and the loss function is minimised. While TMVA uses the loss function in Eq. 5.5, gradient boost can be applied to any loss function with a feasible gradient.

5.1.3. Overtraining

The goal of a boosted decision tree is to identify patterns in the analysed data as accurately as possible. With increasing complexity of the boosted tree, the contribution of even smaller details in the pattern becomes larger. Samples in real life, however, are limited to finite sample sizes and are subject to statistical fluctuations, which can mislead the BDT. This is called overtraining, when the algorithm follows the fluctuations of the sample points too closely, getting every sample point right. If such an overtrained algorithm is then applied to a new, similar dataset, its performance decreases drastically, making the shortcomings of the boosted decision tree obvious.

To prevent overtraining from happening, the sample used to train the BDT is split into two individual samples, the training and the test sample. In this thesis, events are randomly assigned to one of these separate samples, such that in the end both contain the same number of events. The tree is then trained on only the training set, while the test set is used to evaluate the performance on a new, statistically independent sample which was not used for training.

When overlaying the histograms of training and testing sample for the BDT classifier output, they should be approximately the same. A more quantifiable criterion for overtraining than observing their overlap is the Kolmogorov-Smirnov test.

For this, the two empirical distributions $F_{1,n}(x)$ and $F_{2,m}(x)$, which are here evaluated for the training and testing, respectively, are taken to calculate the value

$$D_{n,m} = \sup |F_{1,n}(x) - F_{2,m}(x)| \tag{5.6}$$

where sup is the supremum function. For a good agreement between training and testing sample, the value of $D_{n,m}$ should be close to 1 [80, 81].

5.1.4. Optimisation

Because the boosted decision tree is no longer as straightforward as a single decision tree or a cut analysis, there are several parameters which can be adjusted to optimise the classification power of the BDT. These are listed in Tab. 5.1. For the separation criterion in the node splitting, the options are GiniIndex, MisClassificationError and SdivSqrtSPlusB. When p is the signal purity in a node defined as $S/(S+B)$, where S and B are the number of signal and background events, respectively, the GiniIndex is defined as $p \cdot (1-p)$, and the misclassification error corresponds to $1 - \max(p, 1-p)$. The third option is the definition of the significance, which is $S/\sqrt{S+B} = p^2/S$.

As mentioned in Sec. 5.1.2, AdaBoost performs best on weak classifiers. This is achieved

Parameter	default	Explanation
SeparationType	GiniIndex	Separation criterion for node splitting
NodePurityLimit	0.5	In the boosting process, nodes with purity greater than this value are considered signal, background otherwise
MinNodeSize	5	Minimum percentage of events in the training sample required in a leaf node before the growing process is stopped
NTrees	800	Number of decision trees grown in the boosting process
MaxDepth	3	Maximum depth value allowed for a single decision tree
nCuts	20	Number of points considered when finding the optimal cut in a variable range in node splitting
Shrinkage	1	Learning rate of the gradient boost algorithm
AdaBoostBeta	0.5	Learning rate of the AdaBoost algorithm

Table 5.1.: The different parameters used for optimising the BDT performance.

with small depth of the individual trees which have little discrimination power. Small trees also reduce the risk of overtraining. Further improvements are usually achieved by slowing down the learning rate, which is described with the parameter β given as an exponent to the boost weight.

The same is valid for gradient boost. While the performance of gradient boosting is more stable even for non-ideal setups, it still works best on weak classifiers with small depths. This makes the gradient boosted decision tree less affected by overtraining which can be even further improved when slowing down the learning rate by decreasing the Shrinkage parameter, which is an additional factor given to the boost weight of the individual tree. With this, the number of grown trees has to increase. This way, the performance even in difficult settings is improved.

The optimisation of many parameters of a BDT to achieve the best possible performance on the test set is prone to overtraining in itself since it is affected by fluctuations of the influence the different parameters have on the BDT. To minimise this effect, another independent test set is useful. In this approach, the previous test set is re-named as validation set, for it validates a single BDT. The new test set then serves to assure that the best BDT does not only perform exceptionally well on the validation set in the BDT by chance but that it is actually performing well in general, also on a new, statistically independent sample. The test set is separated from the training and validation set even

before any BDT is trained or evaluated. This way it is ensured that the test set does not influence the training and validation sets. In the presented analysis, it contains 10% of all Monte Carlo event samples, so that both training and validation set are then both equal to 45% each.

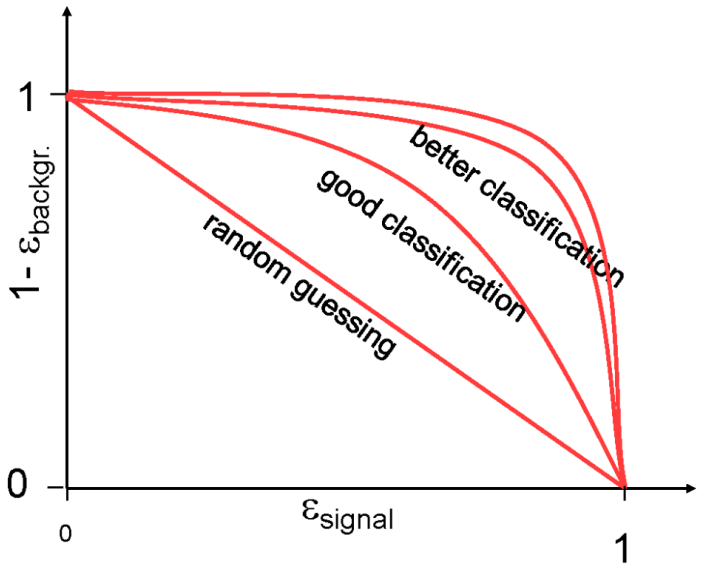


Figure 5.2.: Features of a ROC curve.

The performance of a boosted decision tree can be evaluated with the help of a Receiver Operating Characteristic (ROC) curve. As shown in Figure 5.2, the ROC curve gives the relationship between the signal efficiency $\varepsilon_{\text{signal}}$ and the background rejection, defined as $1 - \varepsilon_{\text{background}}$. By random guessing of the signal or background classification, a straight line is achieved. The better the evaluated classifier, such as the BDT, is, the more the ROC curve tends towards the top right. This can be quantified by taking the integral of the curve, which is always smaller than 1, a perfect classification, and greater than 0.5, the ROC integral for the straight line corresponding to random guessing.

5.2. Signal Extraction

To further use the output of the BDT, which assigns every event a value between -1 and 1 and combines the information of all input variables, a likelihood fit is performed in order to extract the signal. For this, the software framework `TRExFitter` is used, which includes the `HistFactory` tool [82] designed to include signal, validation and control regions as defined in Section 4.4.

To test the existence of the signal in data, a profile likelihood fit is done using Monte Carlo and data samples. For this, a signal strength $\mu = \frac{\sigma}{\sigma_{\text{SM}}}$ is introduced such that $\mu = 0$

equals the background-only hypothesis and $\mu = 1$ is the expectation derived from the Standard Model. This is the parameter of interest and final result in Chapter 6. Similarly, a normalisation factor is introduced on the background strength. For this, the background is split into two groups to separate the *multijet* background and the other backgrounds due to the large uncertainties on the former. All three of these normalisation factors, μ, the background strength of the *multijet* background, and the background strength of the combination of all other background samples, are able to float within the fit.

For this analysis, the systematic uncertainty on the luminosity is the only nuisance parameter implemented. Determining more complex systematic uncertainties requires time and deep understanding through multiple studies of the processes to achieve a sensible result and is therefore a subject for future studies.

For the fitting process, the implemented parametrized probability density function in `TRExFitter` is

$$\mathcal{P} = \prod_{b\in\text{bins}} \text{Pois}(n_{\text{SR},b}|\nu_{\text{SR},b}) \cdot \prod_{\text{CR}_i\in\text{CR}} \prod_{b\in\text{bins}} \text{Pois}(n_{\text{CR}_i,b}|\nu_{\text{CR}_i,b}) \cdot C(\theta_{\text{nom}}|\theta) \tag{5.7}$$

where Pois is the Poisson distribution of events in the one signal (denoted by SR) and two control regions (denoted by CR), n_b is the content of the data histogram and $C(\theta_{\text{nom}}|\theta)$ gives the influence of the floating nuisance parameter θ with nominal value θ_{nom}. The expected number of events in bin b, ν_b, is dependant on all of the fit parameters: the signal strength, the background normalisation factors, and the nuisance parameter θ. The fit minimises $\mathcal{P}$ to find the optimal values for the fit parameters.

For the signal strength of $\mu = 1$, a nominal cross section value of $\sigma_{t\bar{t}(Z\to\nu\bar{\nu},qq)} = 683.1$ fb is used [83]. With the addition of the process $t\bar{t}(Z \to \tau\tau)$, this becomes $\sigma_{t\bar{t}(Z\to\nu\bar{\nu},qq,\tau^-\tau^+)} = 719.5$ fb.

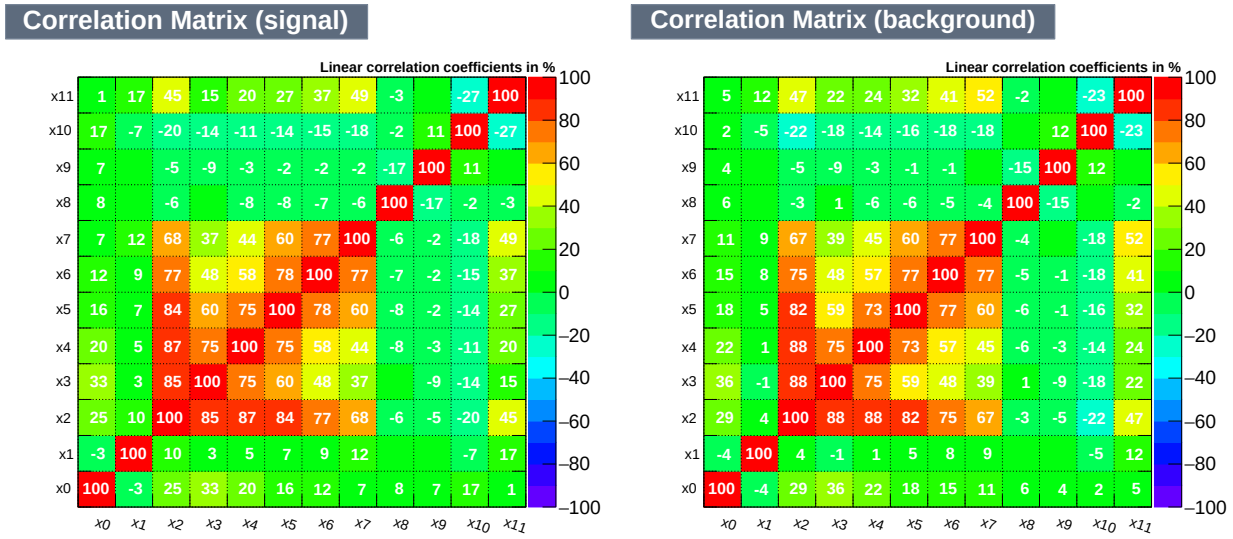
6. Results and Discussion

This chapter gives the results of this master thesis and places them into perspective. First, the boosted decision tree (BDT) optimisation is presented. For this, a preliminary set of samples has been used to understand the behaviour of the different parameters. The best performing BDT then has been tested for its performance on the final samples employing all criteria mentioned in Chapter 4, before the output of that BDT serves as input for the subsequent statistical fitting procedure. In the end, a value for the signal strength is derived and compared to the expectation.

6.1. Boosted Decision Tree

Variable	x_i	Separation Power	Importance
H_T jets	x_2	0.06643	0.08789
p_T of jet 5	x_7	0.06634	0.08565
p_T of jet 4	x_6	0.06465	0.07302
number of jets	x_{11}	0.05405	0.07715
p_T of jet 3	x_5	0.05397	0.06807
$\not{E}_T$	x_0	0.04530	0.08389
p_T of jet 2	x_4	0.04290	0.06610
p_T of jet 1	x_3	0.02196	0.09929
$\Delta\phi_1$	x_8	0.01914	0.08482
Number of b tags	x_1	0.01705	0.08503
$\Delta\phi_2$	x_9	0.01638	0.08247
$\Delta\phi_{\text{closest}}$	x_{10}	0.01020	0.1066

Table 6.1.: A list of the input variables used for training and evaluation of the boosted decision tree optimisation, sorted by their separation power and therefore significance for the analysis in the final BDT evaluated in Section 6.2. The variables $\Delta\phi_1$ and $\Delta\phi_2$ are defined according to Equation 4.3, $\Delta\phi_{\text{closest}}$ uses the minimum value of the ϕ difference between the direction of p_T^{miss} and any jet. Also the corresponding x_i name for Figure 6.1 is listed.



(a) Signal. ***(b)*** Background.

Figure 6.1.: The correlation between individual input variables used in the boosted decision tree. Which variables corresponds to which x_i is noted in Table 6.1.

Several boosted decision trees are trained on a preliminary sample set to evaluate the best one to use for further analysis on the final event selection. The criterion used to determine the effectiveness of a single BDT is the ROC integral as explained in Section 5.1.4. If this value is the same for several BDTs, the signal efficiency at different background levels, as well as the level of overtraining, is taken into account.

The list of input variables for the BDT and their respective separation power, which is a measure for how values of an observable differ between signal and background samples, in the final BDT can be found in Table 6.1. The correlation between the separate input variables is shown in Figure 6.1. The distribution of the considered variables in MC and data samples after event selection but before fitting is also shown in Figure 6.14 and Figure 6.15 in Section 6.3.

6.1.1. Treatment of Negative Event Weights

The NLO and NNLO Monte Carlo samples used for the BDT can include negative event weights which might cause problems. This makes it necessary to define how such events are treated during training and testing, which is the first step for optimising the output of any BDT.

A common choice for negative event treatment is to ignore events with negative weights during the training of the BDT and only include them in testing. Whether this option is

Sample	Events	Events with neg weights	average weight
Multijet	49,820	0	-
Single Top	113,404	2,072	0.12
Multiboson	57,847	349	0.024
W+jets	326,559	4,533	0.28
Z+jets	856,693	13,479	0.093
$t\bar{t}$ allhad	21,347	147	0.59
$t\bar{t}$ nonallhad	1,734,129	9,111	0.12
$t\bar{t}(Z \to \nu\bar{\nu})$	354,507	103,782	0.0036
$t\bar{t}(Z \to qq)$	165,178	46,681	0.0062
$t\bar{t}(Z \to \tau^{-}\tau^{+})$	65,432	19,152	0.0014

Table 6.2.: Number of events (which pass the event selection) in the respective samples, number of these events with negative event weights, and the average value of the negative weight (which corresponds to the contribution of the respective event contribution on the final number of normalised events) in the different samples.

sensible highly depends on the evaluated input samples and fraction of negative weights. Table 6.2 lists the respective input samples with the number of events which have a negative event weight and the average order of magnitude of this weight.

An alternative option for handling negative events is to use these events in the training, but boosting them inversely with the absolute value of the event weight. This way, the event can still be used in training, which is a reasonable choice in the case of low statistics.

To test both options, a BDT is run for comparison, both for AdaBoost and gradient boost. The integral under the ROC curve shown in Figure 6.2 is 0.707 for all variations except for the AdaBoost when ignoring negative events in training, which gives a ROC integral of 0.708. The detailed signal efficiencies can be found in Table A.1. BDT output plots with overtraining tests can be found in Figure A.3.

Considering the similar results from both options and the details about the weights in Table 6.2, the further analysis is conducted using negative event weights and boosting them inversely during the training process, to keep as much statistics in the BDT as possible.

6.1.2. General Parameters

After the treatment of negative weights in the input events, the next basic optimisation parameter is the process by which the different nodes in a tree define the quality of the

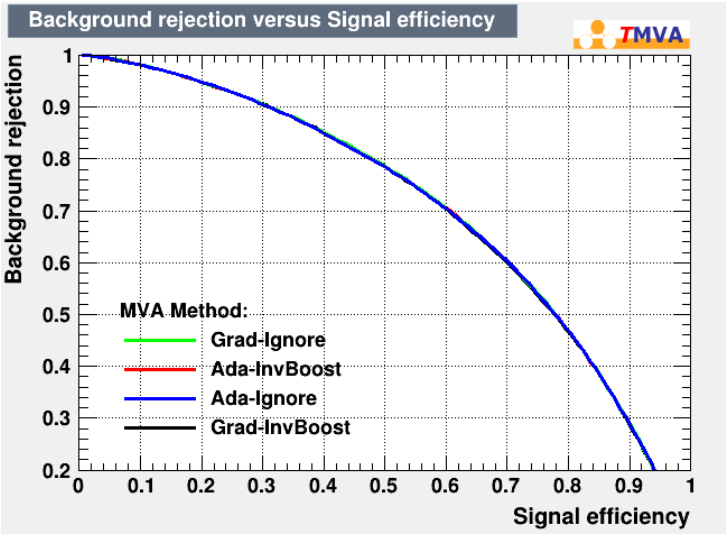


Figure 6.2.: ROC curve for BDTs employing either gradient boosting or AdaBoost, and varied ways of treating negative events, by either ignoring them in training and using them in testing, or boosting events with negative weights inversely.

evaluated cut. Options here are the GiniIndex, the misclassification error, and the number of signal events divided by the square root of the total number of events, as defined in Section 5.1.4. The best result for the ROC integral is achieved by the BDT employing the GiniIndex, making the default setting also the most effective separation type here, as shown in Table 6.3.

In parallel, the two available boosting types, AdaBoost and gradient boost are compared to each other. The results of the different separation types evaluated on a BDT with inversely boosted events with negative weights and otherwise default settings is shown in Table 6.3.

Similarly, a default BDT (with inversely boosted events from negative event weights) with varying node purity limit was run, and the results are found in Table 6.4. The value giving the best result here is a node purity limit of 0.5, which is also the default value.

Both these tables show, that BDTs with gradient boosting instead of AdaBoost are superior in their stability, and therefore gradient boost is the desired boosting type for further BDTs from this point.

The results of the default BDT (with inversely boosted events from negative event weights) with varied minimum node size can be found in Table 6.5a. The best result here is not achieved by the default value of 5%, but 0.5% instead, so this is the value of choice for the final BDT. Just like the minimum node size, the number of considered cuts for every node is also only evaluated for BDTs implementing gradient boost. The results for

Separation Type	AdaBoost	Grad Boost
GiniIndex	0.707	0.707
MisClassificationError	0.706	0.707
SdivSqrtSPlusB	0.572	0.707

Table 6.3.: ROC integrals of the default BDT with varied separation type for AdaBoost and gradient boost BDTs.

NodePurityLimit	AdaBoost	Grad Boost
0.3	0.500	0.707
0.4	0.654	0.707
0.5	0.707	0.707
0.6	0.670	0.707
0.7	0.505	0.707

Table 6.4.: ROC integral of the default BDT with varied node purity limits for AdaBoost and gradient boost BDTs.

a varied number of cuts can be found in Table 6.5b. Here, the preferred value is chosen to be 200 cuts per node as a compromise between the time the training takes, and to ensure at the same time that the best result is still achieved. However, it is worth noting that the difference between 100, 200, and 2000 cuts per node are minimal, as visible in Table A.2 in the appendix.

6.1.3. Shrinkage, Number of Trees, Maximum Depth

Finally, the remaining parameters of shrinkage, which is the learning rate of the boosting algorithm with a default value of 1, number of trees and maximum depth were optimised. The influence of these three parameters on the BDT are correlated, so they were varied simultaneously. For this, an individual BDT has been trained for each combination of maximum depths in the range of 2-6, number of trees from 200 to 2000 and shrinkage values from 0.2 to 5. The results have been split according to the integer values of the maximum depth. The resulting ROC integrals corresponding to a maximum depth of the tree of 2 can be found in Figure 6.3, with the values listed in Table 6.6. The same for a maximum depth of 3 can be found in Figure 6.4 and Table 6.7, for a maximum depth of 4 in Figure 6.5 and Table 6.8, for a maximum depth of 5 in Figure 6.6 and Table 6.9, and for a maximum depth of 6 in Figure 6.7 and Table 6.10.

The highest achieved ROC integral has a value of 0.711 for the combinations of 2000

MinNodeSize	ROC Integral
0.1	0.708
0.2	0.709
0.3	0.709
0.4	0.709
0.5	0.710
1	0.709
2	0.709
2.5	0.709
3	0.709
4	0.708
5	0.707
6	0.707
8	0.703
10	0.697

(a) Minimum Node Size

nCuts	ROC Integral
5	0.642
10	0.642
20	0.707
30	0.708
40	0.708
50	0.709
60	0.709
70	0.709
80	0.709
90	0.709
100	0.710
200	0.710
300	0.710
500	0.709
1000	0.710
2000	0.710

(b) Number of Cuts

Table 6.5.: ROC integral of the default BDT with varied number of cuts for gradient boost BDTs.

trees and a shrinkage of 0.2 at a maximum depth of 4, for both 2000 and 1500 trees and a shrinkage of 0.2 at a maximum depth of 5, and several combinations for a maximum depth of 6, as shown in Table 6.10.

Observing the trend of achieving better results at lower shrinkage values and higher number of trees and maximum depths, another set of BDTs is trained for a shrinkage of 0.1, and a number of trees of 5000, respectively. The results for this can be found in Table 6.11.

Taking the best performing variations from these analyses gives a final set of useful BDTs for the further analysis, as listed in Table 6.12.

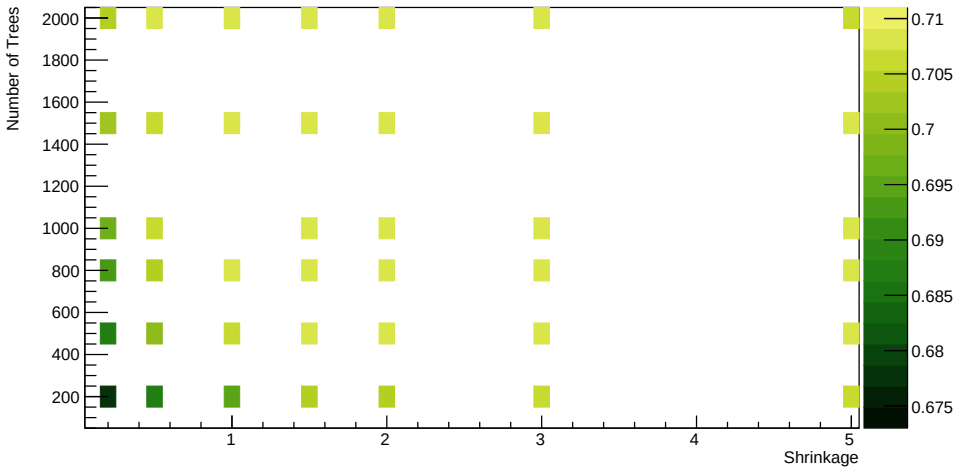


Figure 6.3.: 2D plot of the ROC integral of BDTs with various values of shrinkage and number of trees at fixed maximum depth of 2.

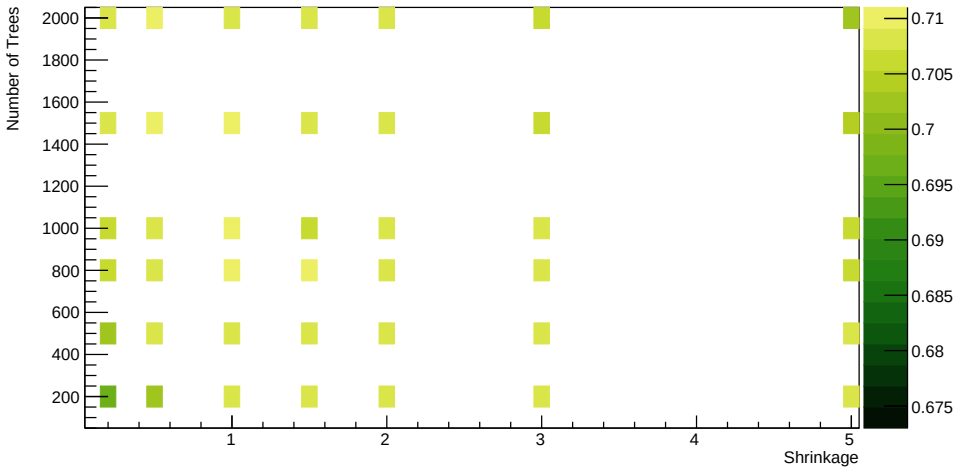


Figure 6.4.: 2D plot of the ROC integral of BDTs with various values of shrinkage and number of trees at fixed maximum depth of 3.

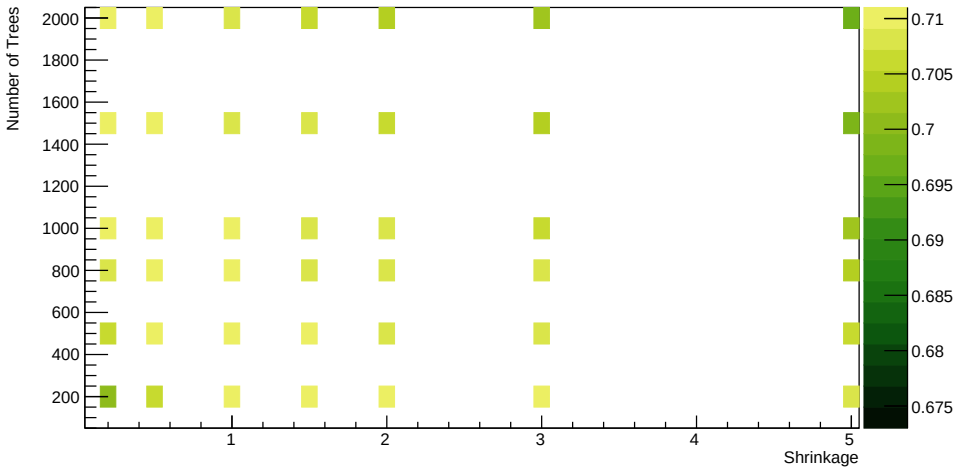


Figure 6.5.: 2D plot of the ROC integral of BDTs with various values of shrinkage and number of trees at fixed maximum depth of 4.

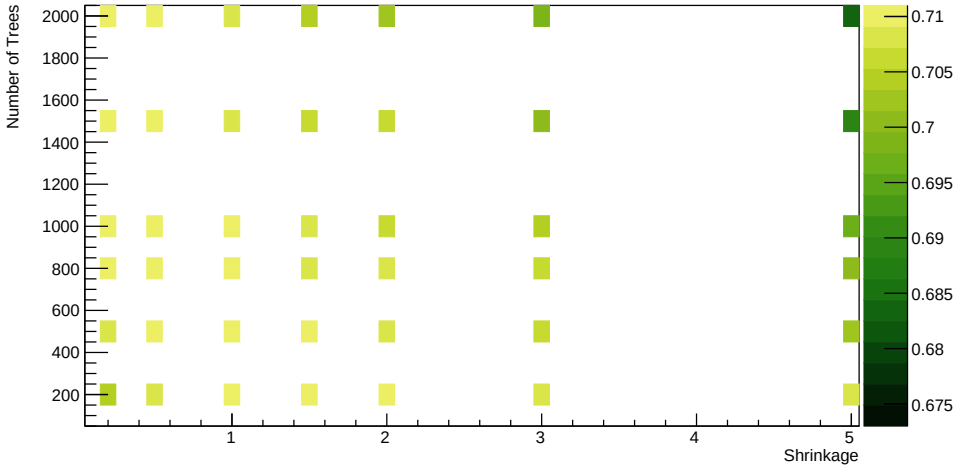


Figure 6.6.: 2D plot of the ROC integral of BDTs with various values of shrinkage and number of trees at fixed maximum depth of 5.

Shrinkage	nTrees	ROC Integral	Shrinkage	nTrees	ROC Integral
0.2	200	0.677	1.5	1000	0.709
0.2	500	0.687	1.5	1500	0.709
0.2	800	0.693	1.5	2000	0.709
0.2	1000	0.697	2	200	0.704
0.2	1500	0.703	2	500	0.708
0.2	2000	0.705	2	800	0.708
0.5	200	0.687	2	1000	0.708
0.5	500	0.701	2	1500	0.708
0.5	800	0.705	2	2000	0.708
0.5	1000	0.706	3	200	0.706
0.5	1500	0.707	3	500	0.708
0.5	2000	0.708	3	800	0.708
1	200	0.695	3	1000	0.708
1	500	0.706	3	1500	0.708
1	800	0.708	3	2000	0.708
1	1000	0.708	5	200	0.707
1	1500	0.708	5	500	0.709
1	2000	0.709	5	800	0.708
1.5	200	0.704	5	1000	0.708
1.5	500	0.708	5	1500	0.708
1.5	800	0.708	5	2000	0.707

Table 6.6.: ROC integral of BDTs with optimised settings from Section 6.1.2, a maximum depth of the single tree of 2, and varied learning rate (shrinkage) and number of trees. The graphical representation can be found in Figure 6.3.

Shrinkage	nTrees	ROC Integral	Shrinkage	nTrees	ROC Integral
0.2	200	0.697	1.5	1000	0.710
0.2	500	0.703	1.5	1500	0.709
0.2	800	0.706	1.5	2000	0.708
0.2	1000	0.707	2	200	0.709
0.2	1500	0.709	2	500	0.709
0.2	2000	0.709	2	800	0.709
0.5	200	0.703	2	1000	0.709
0.5	500	0.708	2	1500	0.709
0.5	800	0.709	2	2000	0.708
0.5	1000	0.709	3	200	0.709
0.5	1500	0.710	3	500	0.709
0.5	2000	0.710	3	800	0.708
1	200	0.708	3	1000	0.708
1	500	0.709	3	1500	0.707
1	800	0.710	3	2000	0.706
1	1000	0.710	5	200	0.709
1	1500	0.710	5	500	0.708
1	2000	0.709	5	800	0.707
1.5	200	0.709	5	1000	0.706
1.5	500	0.709	5	1500	0.704
1.5	800	0.710	5	2000	0.703

Table 6.7.: ROC integral of BDTs with optimised settings from Section 6.1.2, a maximum depth of the single tree of 3, and varied learning rate (shrinkage) and number of trees. The graphical representation can be found in Figure 6.4.

Shrinkage	nTrees	ROC Integral	Shrinkage	nTrees	ROC Integral
0.2	200	0.701	1.5	1000	0.709
0.2	500	0.707	1.5	1500	0.708
0.2	800	0.709	1.5	2000	0.707
0.2	1000	0.710	2	200	0.710
0.2	1500	0.710	2	500	0.709
0.2	2000	0.711	2	800	0.709
0.5	200	0.707	2	1000	0.708
0.5	500	0.710	2	1500	0.707
0.5	800	0.710	2	2000	0.705
0.5	1000	0.710	3	200	0.710
0.5	1500	0.710	3	500	0.709
0.5	2000	0.710	3	800	0.708
1	200	0.710	3	1000	0.707
1	500	0.710	3	1500	0.705
1	800	0.710	3	2000	0.703
1	1000	0.710	5	200	0.708
1	1500	0.709	5	500	0.706
1	2000	0.709	5	800	0.704
1.5	200	0.710	5	1000	0.702
1.5	500	0.710	5	1500	0.699
1.5	800	0.709	5	2000	0.696

Table 6.8.: ROC integral of BDTs with optimised settings from Section 6.1.2, a maximum depth of the single tree of 4, and varied learning rate (shrinkage) and number of trees. The graphical representation can be found in Figure 6.5.

Shrinkage	nTrees	ROC Integral	Shrinkage	nTrees	ROC Integral
0.2	200	0.704	1.5	1000	0.709
0.2	500	0.709	1.5	1500	0.707
0.2	800	0.710	1.5	2000	0.705
0.2	1000	0.710	2	200	0.710
0.2	1500	0.711	2	500	0.709
0.2	2000	0.711	2	800	0.708
0.5	200	0.708	2	1000	0.707
0.5	500	0.710	2	1500	0.706
0.5	800	0.710	2	2000	0.703
0.5	1000	0.710	3	200	0.709
0.5	1500	0.710	3	500	0.707
0.5	2000	0.710	3	800	0.706
1	200	0.710	3	1000	0.705
1	500	0.710	3	1500	0.701
1	800	0.710	3	2000	0.698
1	1000	0.710	5	200	0.708
1	1500	0.709	5	500	0.703
1	2000	0.708	5	800	0.700
1.5	200	0.710	5	1000	0.697
1.5	500	0.710	5	1500	0.690
1.5	800	0.709	5	2000	0.684

Table 6.9.: ROC integral of BDTs with optimised settings from Section 6.1.2, a maximum depth of the single tree of 5, and varied learning rate (shrinkage) and number of trees. The graphical representation can be found in Figure 6.6.

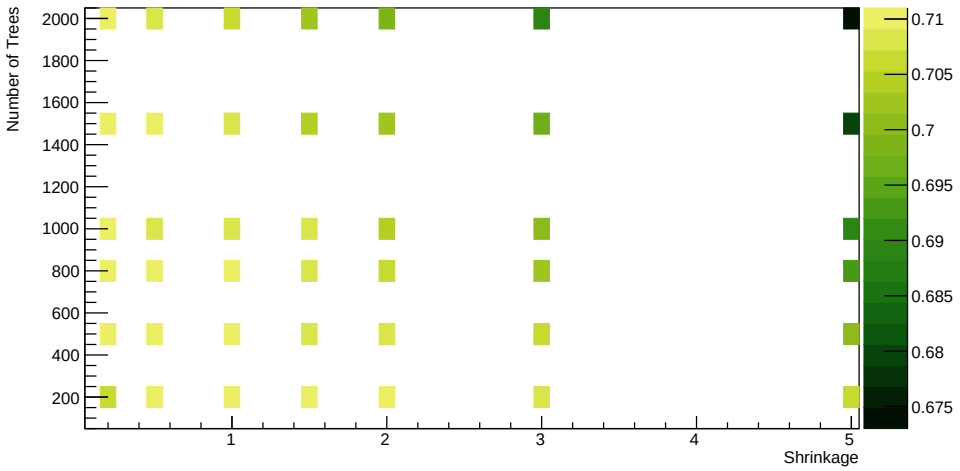


Figure 6.7.: 2D plot of the ROC integral of BDTs with various values of shrinkage and number of trees at fixed maximum depth of 6.

Shrinkage	nTrees	ROC Integral	Shrinkage	nTrees	ROC Integral
0.2	200	0.706	1.5	1000	0.708
0.2	500	0.710	1.5	1500	0.705
0.2	800	0.711	1.5	2000	0.703
0.2	1000	0.711	2	200	0.710
0.2	1500	0.711	2	500	0.708
0.2	2000	0.711	2	800	0.706
0.5	200	0.710	2	1000	0.705
0.5	500	0.711	2	1500	0.702
0.5	800	0.711	2	2000	0.699
0.5	1000	0.710	3	200	0.709
0.5	1500	0.710	3	500	0.706
0.5	2000	0.709	3	800	0.703
1	200	0.711	3	1000	0.701
1	500	0.710	3	1500	0.696
1	800	0.710	3	2000	0.690
1	1000	0.709	5	200	0.706
1	1500	0.708	5	500	0.700
1	2000	0.706	5	800	0.693
1.5	200	0.710	5	1000	0.689
1.5	500	0.709	5	1500	0.680
1.5	800	0.708	5	2000	0.673

Table 6.10.: ROC integral of BDTs with optimised settings from Section 6.1.2, a maximum depth of the single tree of 6, and varied learning rate (shrinkage) and number of trees. The graphical representation can be found in Figure 6.7.

mDepth	Shrink.	nTrees	ROC Int	Signal Eff. from Test (Training) Sample at B=0.01	at B=0.10	at B=0.30
4	0.1	1500	0.709	0.071 (0.072)	0.316 (0.317)	0.606 (0.605)
4	0.1	2000	0.710	0.073 (0.073)	0.318 (0.320)	0.608 (0.608)
4	0.1	5000	0.710	0.074 (0.075)	0.320 (0.321)	0.610 (0.610)
4	0.2	5000	0.710	0.074 (0.075)	0.319 (0.320)	0.608 (0.609)
4	0.5	5000	0.708	0.074 (0.076)	0.316 (0.319)	0.607 (0.608)
5	0.1	1500	0.710	0.075 (0.074)	0.319 (0.320)	0.607 (0.607)
5	0.1	2000	0.710	0.076 (0.076)	0.320 (0.321)	0.608 (0.608)
5	0.1	5000	0.710	0.075 (0.076)	0.321 (0.323)	0.609 (0.610)
5	0.2	5000	0.710	0.076 (0.077)	0.321 (0.323)	0.608 (0.609)
5	0.5	5000	0.707	0.075 (0.077)	0.316 (0.320)	0.603 (0.605)
6	0.1	1500	0.710	0.074 (0.074)	0.320 (0.321)	0.608 (0.608)
6	0.1	2000	0.711	0.074 (0.075)	0.321 (0.321)	0.608 (0.609)
6	0.1	5000	0.710	0.075 (0.076)	0.321 (0.322)	0.608 (0.610)
6	0.2	5000	0.709	0.074 (0.076)	0.320 (0.322)	0.607 (0.608)
6	0.5	5000	0.705	0.073 (0.075)	0.313 (0.318)	0.599 (0.602)

Table 6.11.: The ROC integral value and signal efficiency at different background efficiencies for BDTs with optimised settings and varied shrinkage, number of trees and maximal tree depth, values around the optimal areas.

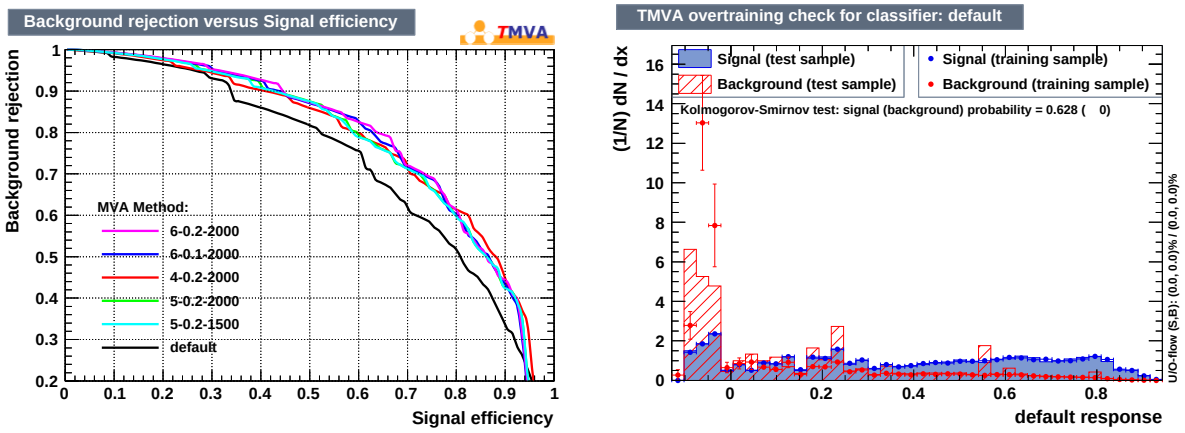
mDepth	Shrink.	nTrees	ROC Int	Signal Eff. from Test (Training) Sample at B=0.01	at B=0.10	at B=0.30
4	0.2	2000	0.711	0.074 (0.074)	0.319 (0.320)	0.610 (0.610)
5	0.2	1500	0.711	0.075 (0.075)	0.320 (0.322)	0.609 (0.610)
5	0.2	2000	0.711	0.075 (0.076)	0.320 (0.322)	0.610 (0.611)
6	0.1	2000	0.711	0.074 (0.075)	0.321 (0.321)	0.608 (0.609)
6	0.2	2000	0.711	0.075 (0.075)	0.320 (0.322)	0.608 (0.609)

Table 6.12.: Final BDT candidates for the further analysis, containing optimised parameters and varied shrinkage, number of trees, and maximal tree depth.

6.2. Applying the BDT to the Final Samples

To determine the best BDT settings from the ones in Table 6.12 to be used in the classifier, the BDTs are separately trained on the final event selection instead of the preliminary samples, together with the default BDT for comparison.

The ROC curves, together with the overtraining plot of the default BDT, are shown in Figure 6.8. The overtraining, here, is extraordinarily large, and because this is already the case for default parameters before optimisation, the problem is likely caused by the input samples. The most likely candidate causing this is the multijet sample which has several events with event weights greater than 100. The contribution of these events in the training is therefore disproportionate to other events of the same sample, as well as the other samples.



(a) ROC curve. *(b)* Overtraining plot.

Figure 6.8.: ROC curve of the BDTs from Table 6.12 trained and evaluated on the final event selection including the multijet sample, as well as the overtraining plot for the default BDT on these samples.

For this reason, the multijet sample is excluded from the training and validation of the BDT, but again included into the analysis for the classification of the samples and the fitting procedure in Section 6.3. This is accepted because, as shown in Figure 4.7, the BDT still performs well on the multijet sample and classifies these events as more background-like.

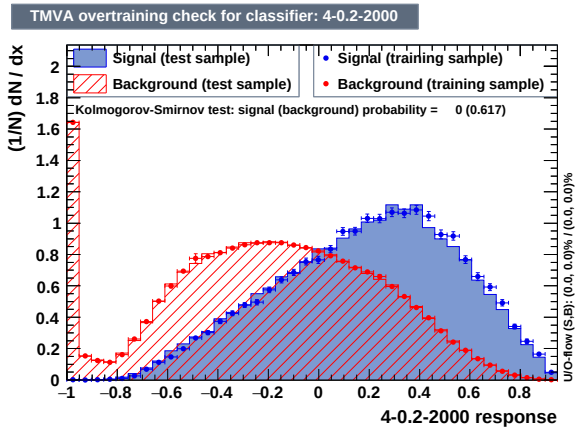
Implementing this, the results of the BDT trained and evaluated on the final event selection, except the multijet sample, can be found in Table 6.13. The overtraining test plots are shown in Figure 6.9. While the background distribution has hardly any problem with overtraining, the BDT does not perform exceptionally well on the signal sample, as visible especially from the value given for the Kolmogorov-Smirnov test. This is 0.007 for

the default BDT and smaller than 0.001 for the others. Because this is the case for all of the BDTs, the analysis is continued with overtraining on the signal sample. However, it is taken into account as a source for uncertainties. The best test in the Kolmogorov-Smirnov test of 0.85 in the background distribution is achieved with the BDT using a maximum depth of 6, a shrinkage of 0.1, and training on 2000 trees.

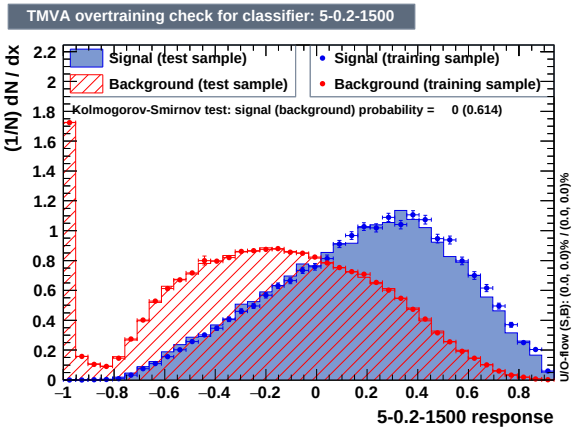
mDepth	**Shrink.**	**nTrees**	**ROC Int**	**Signal Eff. from Test (Training) Sample**		
				at B=0.01	**at B=0.10**	**at B=0.30**
4	0.2	2000	0.746	0.084 (0.086)	0.349 (0.351)	0.651 (0.651)
5	0.2	1500	0.746	0.084 (0.085)	0.350 (0.352)	0.651 (0.651)
6	0.1	2000	0.746	0.084 (0.085)	0.350 (0.352)	0.653 (0.653)
5	0.2	2000	0.746	0.084 (0.086)	0.349 (0.351)	0.651 (0.651)
6	0.2	2000	0.746	0.084 (0.085)	0.350 (0.353)	0.652 (0.653)
	default		0.739	0.076 (0.076)	0.345 (0.345)	0.645 (0.644)

Table 6.13.: Final BDT candidates for the further analysis, containing optimised parameters and varied shrinkage, number of trees, and maximal tree depth, trained on the final event selection samples.

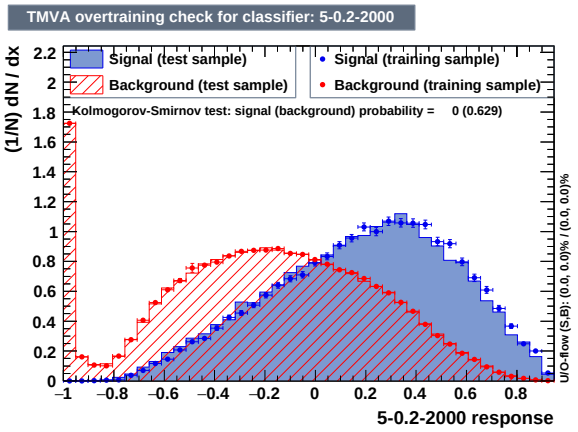
Because all of the selected BDTs have the same ROC integral value of 0.746, the BDT with the least overtraining, measured by the Kolmogorov-Smirnov-Test in Figure 6.9 is selected, which is the BDT with a maximum depth of 6, a shrinkage of 0.1, and a number of trees of 2000. This classifier will serve as input for the statistical fit analysis following in Section 6.3. Before that is possible, the BDT has to be tested on the test set as explained in Section 5.1.4. For this, the remaining 10% of events is plotted in Figure 6.11. The shape of the distributions in this plots looks similar to the distributions in Figure 6.9d. Minor deviations, such as the height of the peak in the first bin, can be explained with differences in the binning in both graphs. This makes the considered BDT fit as input for the statistical fitting.



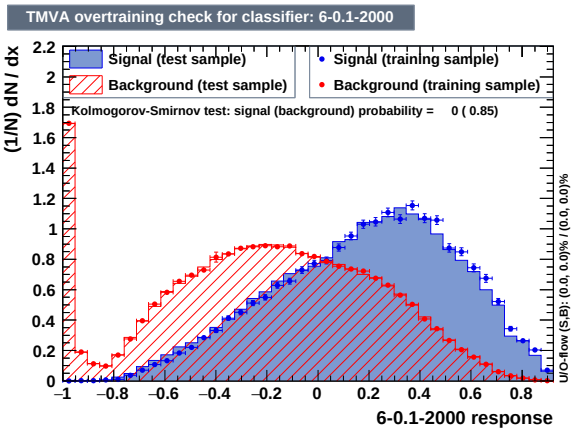
(a) maxDepth = 4, Shrinkage = 0.2, nTrees = 2000.



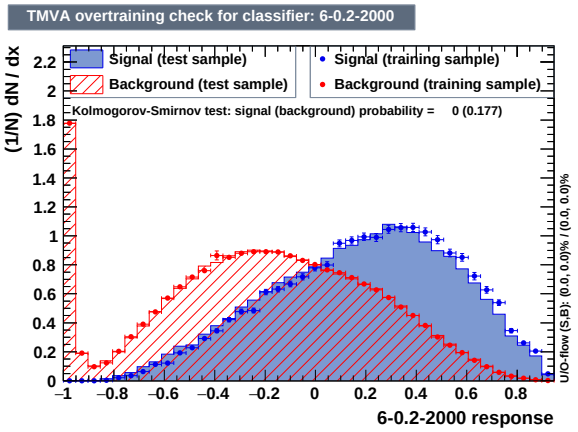
(b) maxDepth = 5, Shrinkage = 0.2, nTrees = 1500.



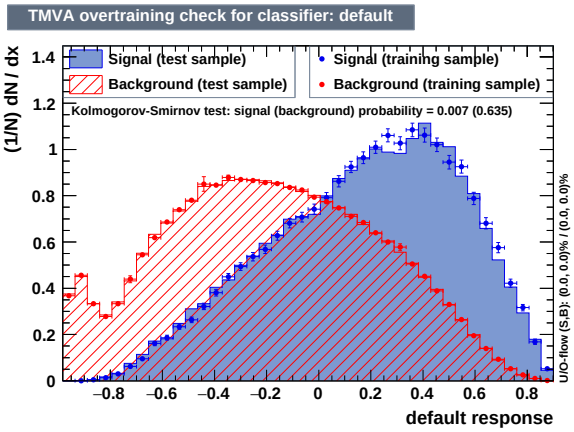
(c) maxDepth = 5, Shrinkage = 0.2, nTrees = 2000.



(d) maxDepth = 6, Shrinkage = 0.1, nTrees = 2000.



(e) maxDepth = 6, Shrinkage = 0.2, nTrees = 2000.



(f) default Parameters

Figure 6.9.: Overtraining tests for the selected BDTs from Table 6.13 and the default BDT for comparison. A value of 0 for the Kolmogorov-Smirnov test does not mean that the actual value is 0, but only indicates that it is smaller than 0.001, as implemented in TMVA.

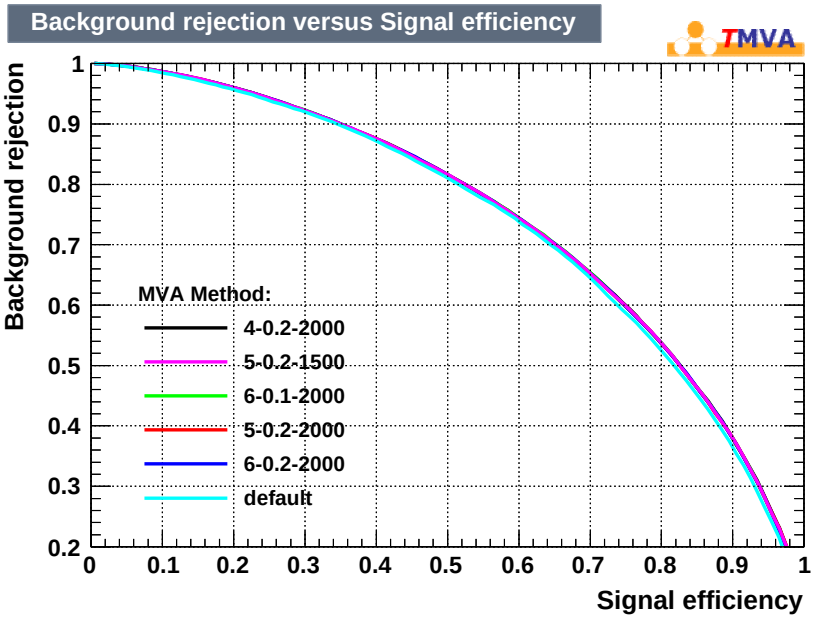


Figure 6.10.: ROC curves for the selected BDTs from Table 6.13 and the default BDT for comparison.

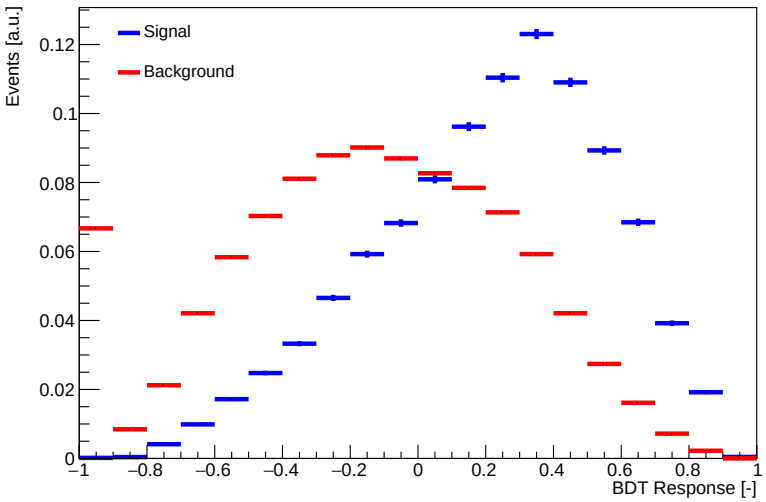


Figure 6.11.: Distribution of the BDT classifier with a maximum depth of 6, a shrinkage of 0.1, and a number of trees it was trained on of 2000, applied to the 10% of the input samples on which the BDT was not trained. This serves as validation that the optimisation process of the BDT parameters is not overtrained.

6.3. Cross Section Extraction

With the regions defined in Section 4.4, a likelihood fit is performed to fit the Monte Carlo samples to data. Figure 6.14 and Figure 6.15 show the distribution of all variables of interest for MC and data samples for all events passing the event selection in Section 4.3. The final signal region of the BDT output with the cut at greater than 0.6. This is shown in Figure 6.12, where the last bin with the upper limit of 0.9 is the overflow bin. In this region, the fit is performed on five bins, whereas a total of ten bins have been set for the control regions, respectively.

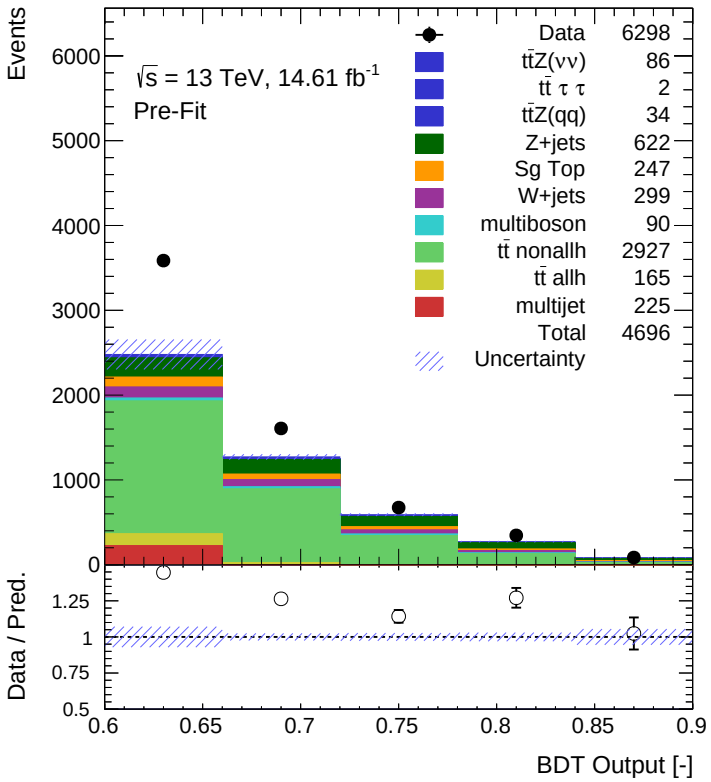


Figure 6.12.: Signal region of the BDT output as defined in Section 4.4 with a cut at 0.6 applied to reduce the contribution from the QCD multijet. The right-most bin is an overflow bin here.

The fit is performed including the luminosity as the only nuisance parameter, which is allowed to vary by 2.1% in both positive and negative direction as measured with the *LUCID-2* system in Run II [84]. The normalisation for the combination of all background samples except the QCD multijet are set to be in the range of 0 and 2, with a nominal value of 1, the signal strength as well as the norm factor for QCD multijets are set to be in the range of 0 to 10 with a nominal value of 1, since they are required to float more

freely due to their role as point of interest in the fit or their very high uncertainty.

The fit configuration has been tested for consistency by fitting the MC samples to MC expectations while blinding data, which gave the nominal values as expected from a working setup. Also several variations in binning and extensions of the signal region have been tested, which gave consistent results. The results given by the final fit, when using this configuration to fit the MC samples to data, are

$$\mu_{\text{multijet}} = 3.03^{+0.10}_{-0.10} \tag{6.1}$$

$$\mu_{\text{other bkg}} = 1.067^{+0.023}_{-0.022} \tag{6.2}$$

$$\mu_{t\bar{t}Z} = 5.78^{+0.68}_{-0.66} \tag{6.3}$$

$$\alpha_{\text{Lumi}} = 0.07^{+1.0}_{-1.0} \tag{6.4}$$

With the cross section value calculated in Section 5.2 and the resulting signal strength here, this gives a final result for the cross section of $t\bar{t}Z$ of

$$\sigma_{t\bar{t}(Z\to\nu\bar{\nu},qq,\tau^-\tau^+)} = 4158^{+490}_{-475}(\text{stat})\,\text{fb} \tag{6.5}$$

in contrast to the theoretical value given by the Standard Model of

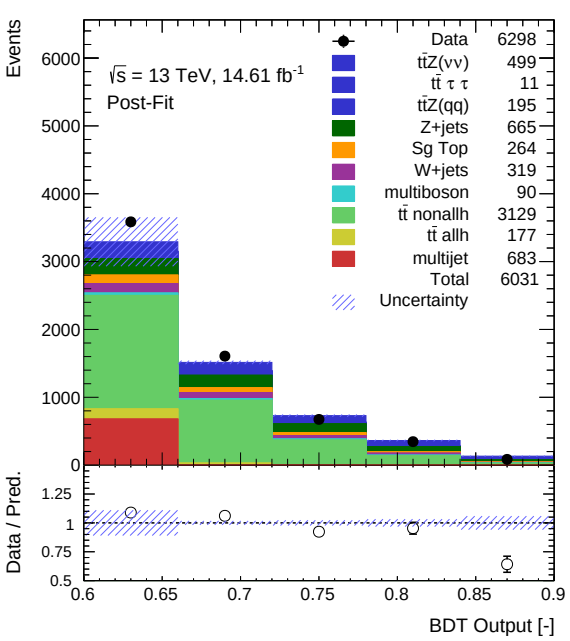
$$\sigma_{t\bar{t}(Z\to\nu\bar{\nu},qq,\tau^-\tau^+)} = 759.8 \pm 2.6^{+9.7\%}_{-2.2\%}\,(\text{scale})^{+1.9\%}_{-2.2\%}(\text{PDF})\,\text{fb}. \tag{6.6}$$

The signal region and control regions after the fit, as well as the distribution of the complete BDT output are shown in Figure 6.13. It is visible that the non-fully-hadronic $t\bar{t}$ control region responded well to the fit, and the MC sample points agree well with data. In contrast to that, all regions where there is contribution from the QCD multijet sample, there is very little agreement. In both distributions of the BDT output, before the fit in Figure 4.7 and after the fit in Figure 6.13d the high peak in the first bin is not visible, although it is visible in Figure 6.9d and Figure 6.11. This is due to the selected binning. If the number of bins in Figure 4.7 and Figure 6.13d are set to 100 bins instead of 10, this peak becomes visible again.

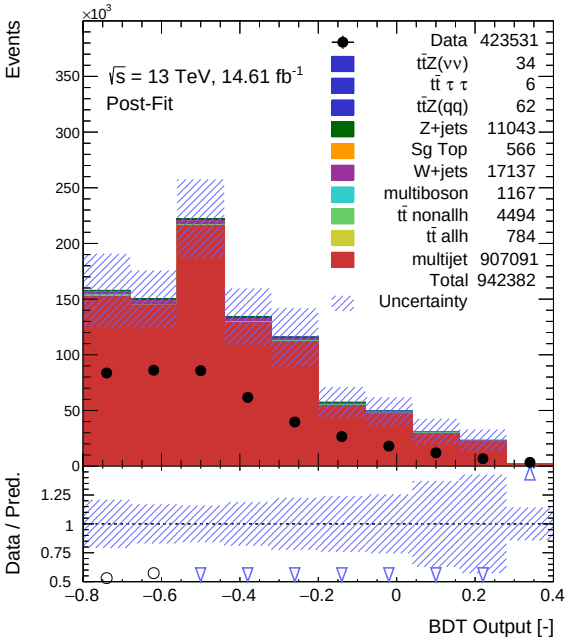
The estimation of a complete systematic error has not been performed in this thesis, due the complexity of such studies and the time required for them. However, large sources of the systematic uncertainties, here, is the simulation of the MC samples, especially the QCD multijet sample, which in further analyses should be replaced by a data-driven approach. Additional systematics for fitting can be derived from theoretical cross section uncertainties, as well as jet property uncertainties.

The distribution of all variables of interest for MC and data samples after the fit to data

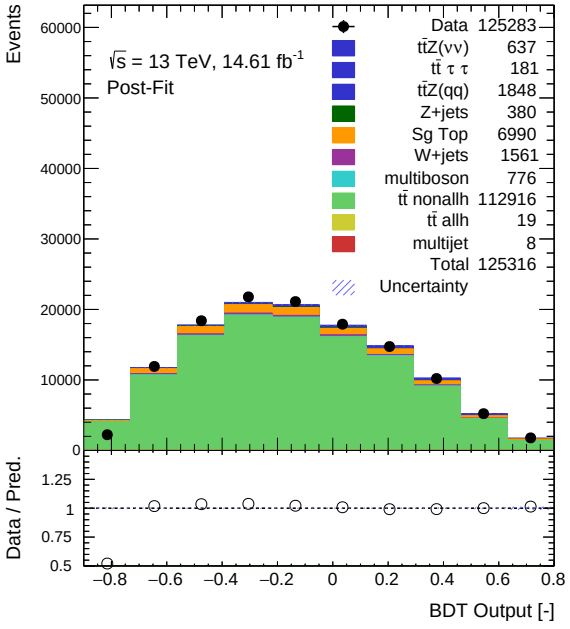
are shown in Figure 6.16 and Figure 6.17. It is visible that the agreement between fitted MC and data events is not optimal, which reflects the quality of the fit output as well. By making optimisations in the input samples, e.g. with data-driven QCD multijet samples, in the training of the BDT, e.g. by training on the final event selection and including control regions, as well as including and optimising more systematic uncertainties, the results here may improve and have larger significance.



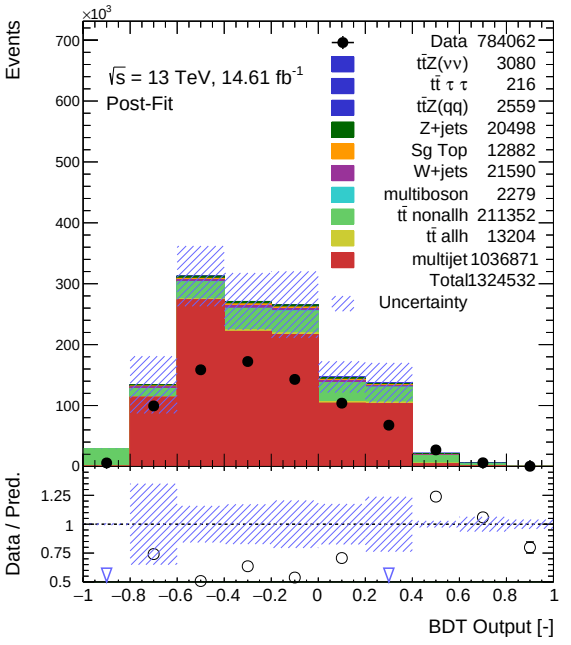
(a) Signal region.



(b) QCD multijet control region.

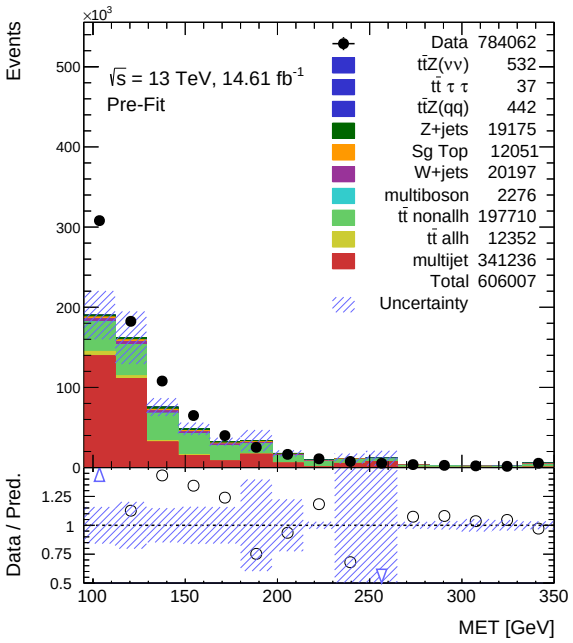


(c) Allhadronic $t\bar{t}$ control region.

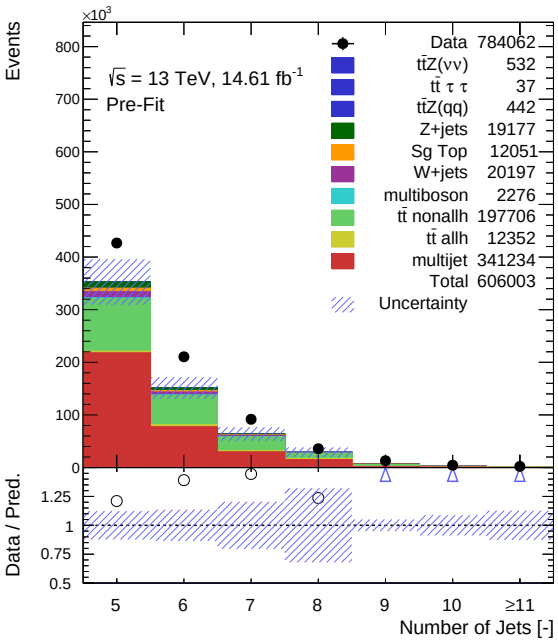


(d) Extended signal region over the complete BDT output.

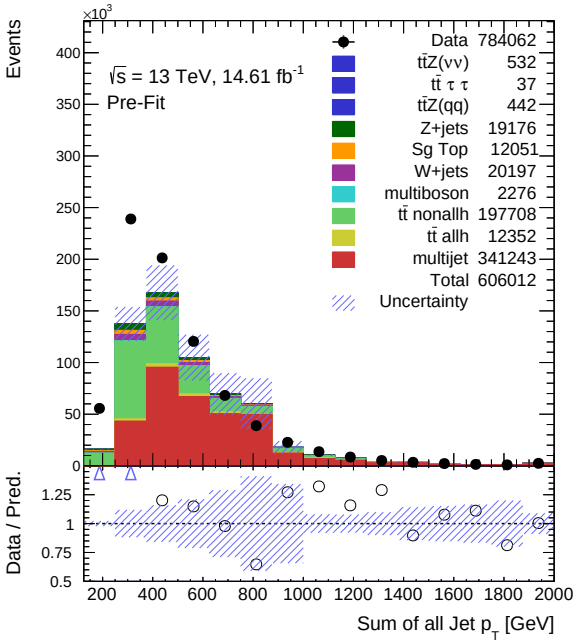
Figure 6.13.: Monte Carlo and data distribution of the signal region, control regions, and the complete BDT output distribution after the fitting.



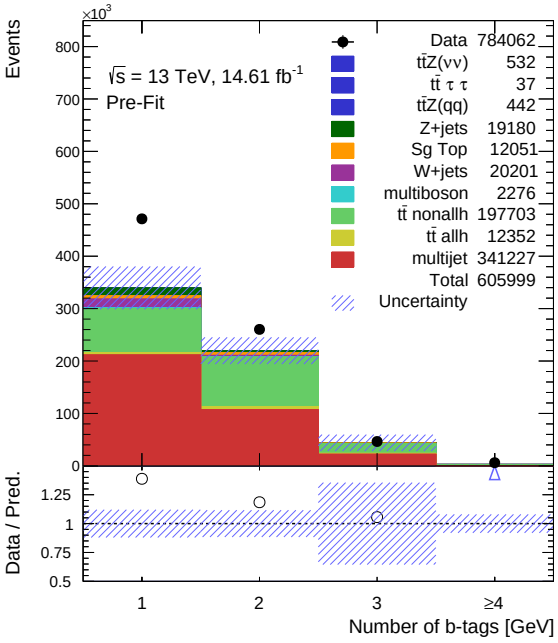
(a) $\not{E}_T$.



(b) Number of jets.

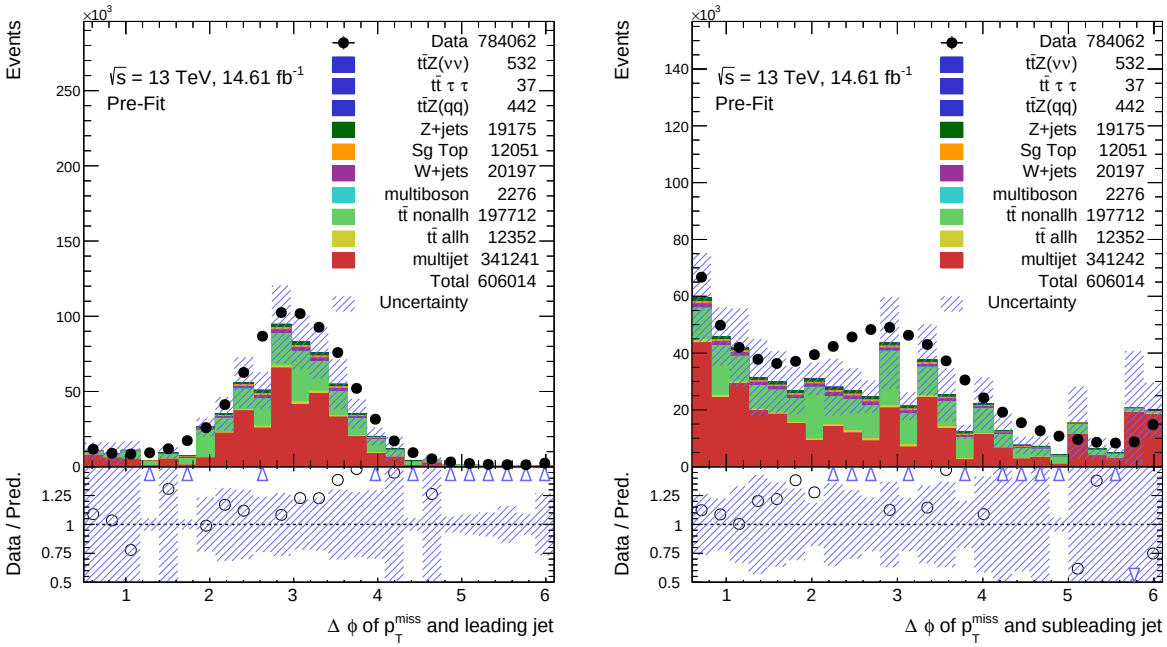


(c) H_T jets.



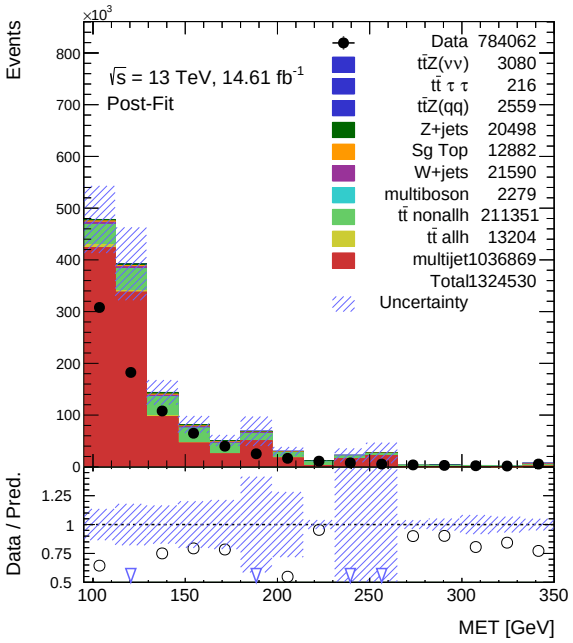
(d) Number of b tags.

Figure 6.14.: Monte Carlo and data distribution of various observables after applying event selection cuts, and before performing the fit. The p_T distributions of the second to fifth jet and the distribution of $\Delta\phi_{\text{closest}}$ can be found in the appendix in Figure A.4 and Figure A.5.

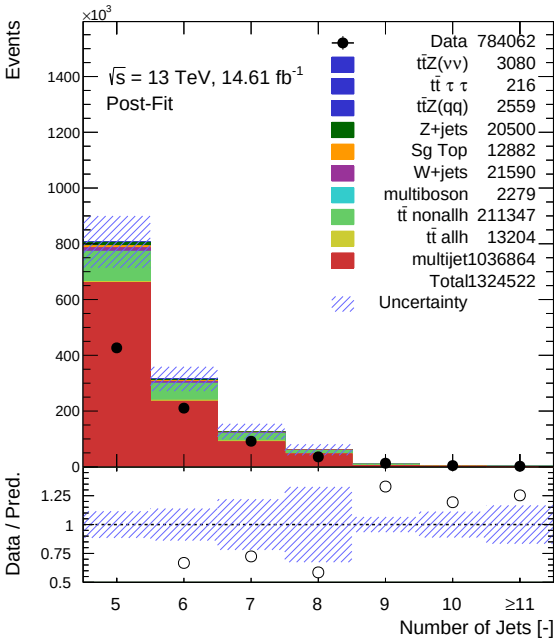


(a) $\Delta\phi_1$. *(b)* $\Delta\phi_2$.

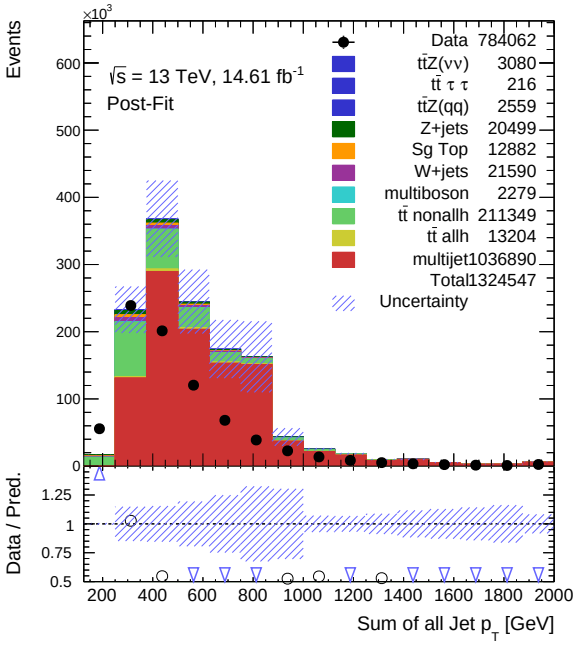
Figure 6.15.: Monte Carlo and data distribution of various observables after applying event selection cuts, and before performing the fit. The p_T distributions of the second to fifth jet and the distribution of $\Delta\phi_{\text{closest}}$ can be found in the appendix in Figure A.4 and Figure A.5.



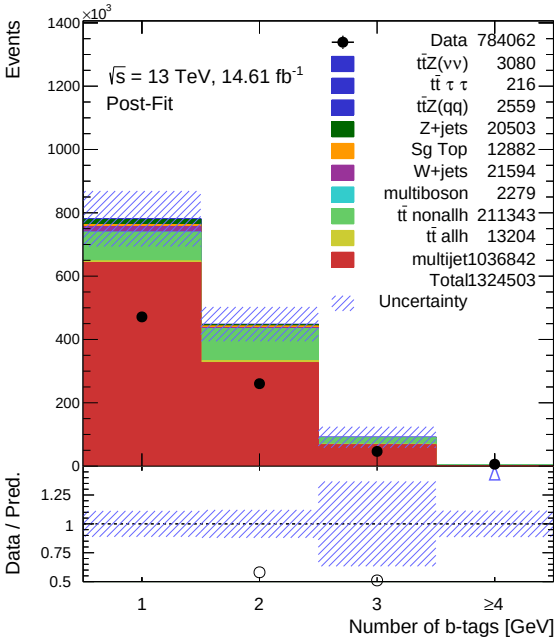
(a) $\not{E}_T$.



(b) Number of jets.

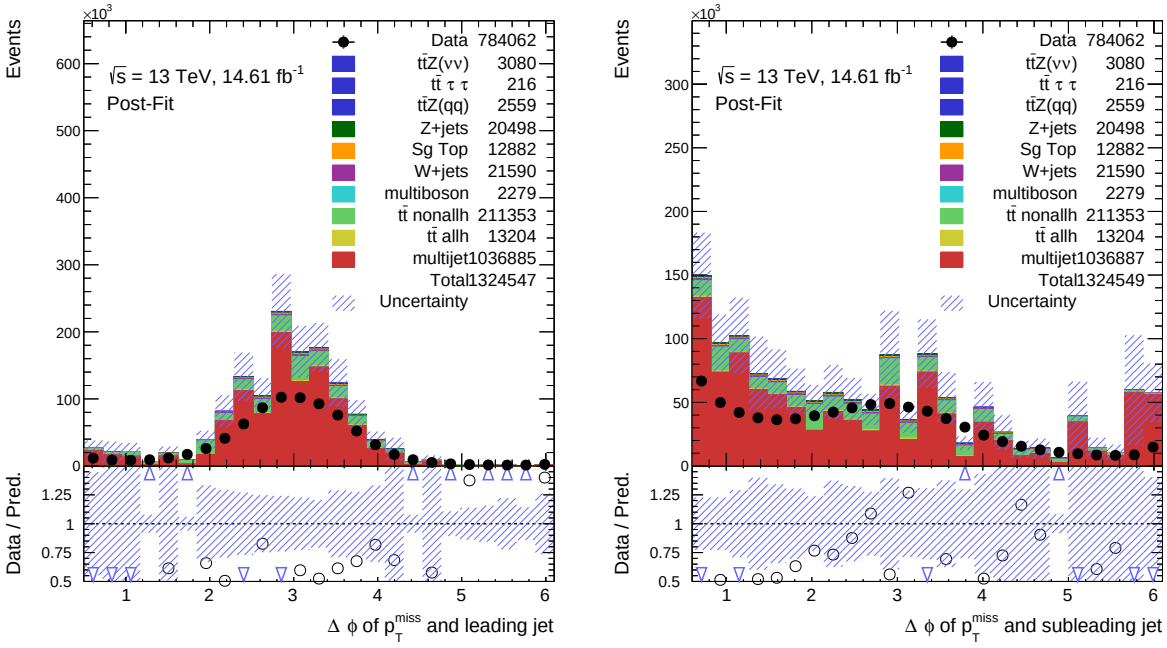


(c) H_T jets.



(d) Number of b tags.

Figure 6.16.: Monte Carlo and data distribution of various observables after the fitting. The p_T distributions of the second to fifth jet and the distribution of $\Delta\phi_{\text{closest}}$ can be found in the appendix in Figure A.6 and Figure A.7.



(a) $\Delta\phi_1$. *(b)* $\Delta\phi_2$.

Figure 6.17.: Monte Carlo and data distribution of various observables after the fitting. The p_T distributions of the second to fifth jet and the distribution of $\Delta\phi_{\text{closest}}$ can be found in the appendix in Figure A.6 and Figure A.7.

7. Summary and Outlook

This thesis presented a measurement of the cross section of $t\bar{t}Z$ in the two neutrino final state with contributions from two quarks, and two tau lepton final states using a dataset of $14.61\,\text{fb}^{-1}$ recorded at a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ at the ATLAS detector at the LHC.

The $t\bar{t}Z$ process is the combination of both initial and final state radiation processes in the associated production of a $t\bar{t}$ pair and a Z boson. The final state radiation contributions make it possible to measure the coupling between the Z boson and the top quark, and with this the third component of the weak isospin for the top. There, deviations from the expected value of $+1/2$ can hint at new physics beyond the Standard Model.

The $t\bar{t}(Z \to \nu\bar{\nu})$ channel specifically is a very important background for supersymmetry searches, and a deep understand of it is therefore necessary for such discoveries. The process $t\bar{t}Z$ has a relatively low production cross section of about $0.76\,\text{pb}$ and so its observation is still recent. While searches which include charged leptons in the final state are now established channels, the $t\bar{t}(Z \to \nu\bar{\nu})$ is also crucial since the $Z \to \nu\bar{\nu}$ branching fraction is 5-6 times higher than the branching fraction of the Z boson decaying into charged leptons.

For the $t\bar{t}$ decay in this process, the allhadronic channel is chosen due to its largest branching fraction and the advantage that there are no additional neutrinos in the final state signature.

The cross section of the $t\bar{t}Z$ process with neutrinos in the final state has not been measured before, and so the analysis included several complications. The trigger applied to MC and data samples was not ideal and cut away more than half of the possible signal events, while decreasing the statistics in the data sample to $14.61\,\text{fb}^{-1}$. On a longer time scale in the future of the $t\bar{t}(Z \to \nu\bar{\nu})$ analysis, improving the trigger provides a first opportunity to improve the resulting cross section by doubling the statistics. Possibly the greatest source of uncertainty, and another good working point for further analysis, is the simulation of the QCD multijet background. This is the background with the largest contribution and it is therefore necessary to model it well. For this, Monte Carlo samples where sometimes only few events with very large weights comprise an individual bin are

not ideal, and a data-driven approach would be more promising. This would decrease the statistical uncertainties in the bins with highest weights in the plots and likely reduce the systematic uncertainty to improve the overall accuracy.

Here, a boosted decision tree was trained and optimised to separate signal from background and thus improve signal purity in the relevant regions. The BDT was trained without including the QCD multijet sample because it did not provide sensible results, possibly due to the high event weights in this MC sample. In further analyses, after achieving a better simulation of the multijet background, this sample should also be included throughout the analysis, especially in the BDT. It has to be mentioned that, although the multijet sample was not used for the training, its distribution from the BDT classifier was still background-like and so it was acceptable to use a BDT only trained on the remaining background samples. When improving the BDT, it can prove beneficial to train and optimise the BDT on the final sample set instead of a preliminary one. This way, it might be possible to achieve less overtraining in the signal sample than was the case here.

The signal-like range of the BDT classifier then served as input to a likelihood fit. For this, the signal strength, as well as the QCD multijet background strength, and the strength of the combination of the other backgrounds have been evaluated. The signal region is the signal-like BDT output in a range from 0.6 to 1.0. In addition, separate control regions for the two greatest background contributions, from the QCD multijet and non-fully-hadronic $t\bar{t}$ samples have been included in the fit. The lower BDT output values in the non-fully-hadronic $t\bar{t}$ control region have been excluded from the fit due to mismodelling of the MC in comparison to data. In addition, there is a non-negligible signal contribution in this region. In the further analysis, these control regions can be chosen to contain less signal contribution and better modelling to data to improve the overall result.

For systematics, only the luminosity has been taken into account, since estimation of systematic uncertainties is very complex and if not done thoroughly might do more harm than good. Here, it is advised to begin with the main systematics such as the uncertainties derived from the generation of the background sample. In further studies it is possible to include even more systematics such as theoretical cross section uncertainties and uncertainties originating from the jets.

With this likelihood fit, a value of $\sigma_{t\bar{t}(Z\to\nu\bar{\nu},qq,\tau^-\tau^+)} = 4158^{+490}_{-475}(\text{stat})\,\text{fb}$ has been measured for the cross section, although systematic uncertainties are expected to be larger than the value itself. This means that although the value, here, does not agree with the standard model value of $\sigma_{t\bar{t}(Z\to\nu\bar{\nu},qq,\tau^-\tau^+)} = 759.8 \pm 2.6^{+9.7\%}_{-2.2\%}\,(\text{scale})^{+1.9\%}_{-2.2\%}(\text{PDF})\,\text{fb}$, the mag-

nitude of the possible systematic uncertainties makes it impossible to claim new physics beyond the standard model from this analysis.

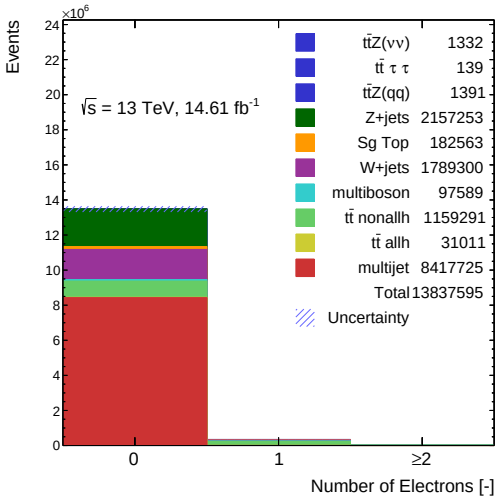
A. Additional Plots and Tables

Boosting	Neg. Weight Events	Signal Eff. from Test (Training) Sample at B=0.01	at B=0.10	at B=0.30
Gradient	Ignored	0.066 (0.067)	0.311 (0.310)	0.605 (0.604)
AdaBoost	Inverse boost	0.067 (0.067)	0.311 (0.311)	0.606 (0.606)
AdaBoost	Ignored	0.065 (0.065)	0.310 (0.310)	0.604 (0.604)
Gradient	Inverse boost	0.064 (0.064)	0.312 (0.312)	0.604 (0.604)

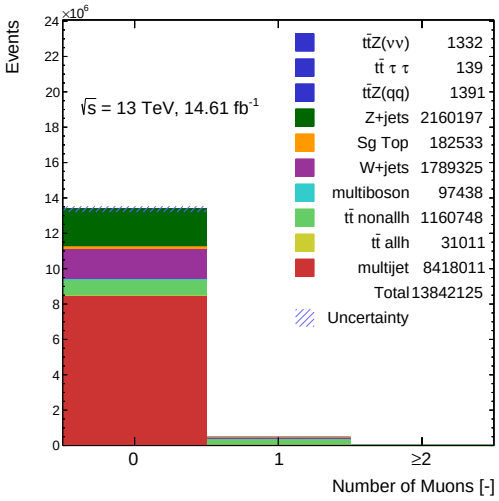
Table A.1.: The signal efficiency at different background efficiencies for BDTs with default settings and either inversely boosted events with negative weights or with ignored negative event weights in training but used in testing.

nCuts	Signal Eff. from Test (Training) Sample		
	at B=0.01	at B=0.10	at B=0.30
5	0.032 (0.032)	0.226 (0.226)	0.509 (0.508)
10	0.033 (0.033)	0.228 (0.228)	0.509 (0.508)
20	0.064 (0.064)	0.312 (0.312)	0.604 (0.604)
30	0.070 (0.070)	0.313 (0.313)	0.605 (0.605)
40	0.070 (0.070)	0.315 (0.313)	0.607 (0.607)
50	0.072 (0.071)	0.314 (0.313)	0.608 (0.608)
60	0.070 (0.070)	0.315 (0.315)	0.607 (0.607)
70	0.070 (0.070)	0.315 (0.316)	0.609 (0.609)
90	0.071 (0.071)	0.317 (0.316)	0.607 (0.607)
80	0.070 (0.070)	0.315 (0.315)	0.608 (0.608)
100	0.072 (0.072)	0.317 (0.318)	0.608 (0.608)
200	0.071 (0.071)	0.318 (0.317)	0.608 (0.608)
300	0.071 (0.071)	0.317 (0.316)	0.607 (0.607)
500	0.071 (0.071)	0.318 (0.318)	0.607 (0.608)
1000	0.072 (0.072)	0.317 (0.317)	0.607 (0.608)
2000	0.072 (0.072)	0.317 (0.317)	0.608 (0.608)

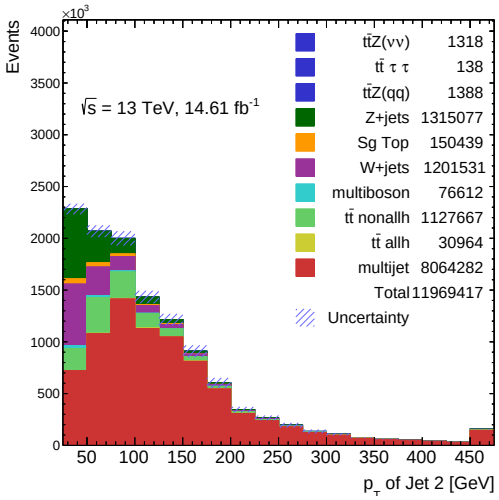
Table A.2.: The signal efficiency at different background efficiencies for BDTs with default settings, inversely boosted events with negative weights, and varying number of cuts considered for a single node.



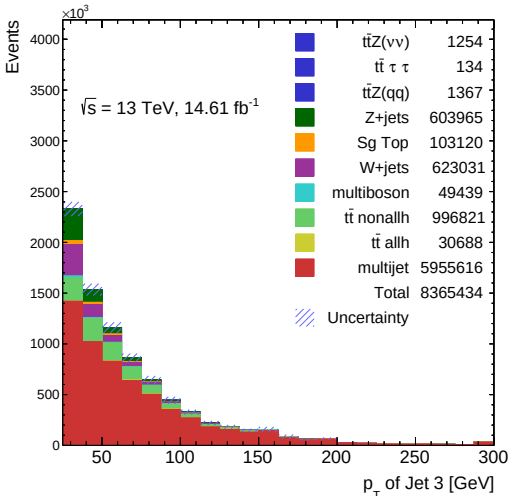
(a) Number of Electrons.



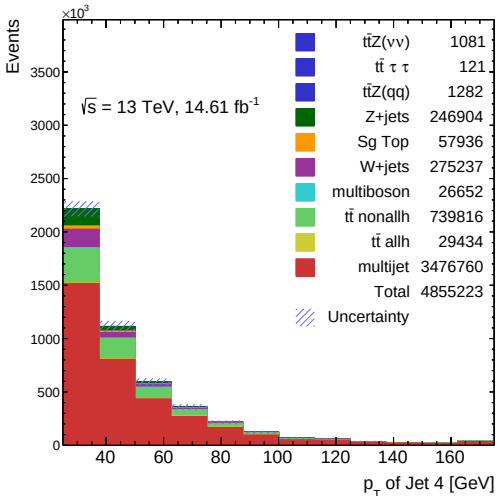
(b) Number of Muons.



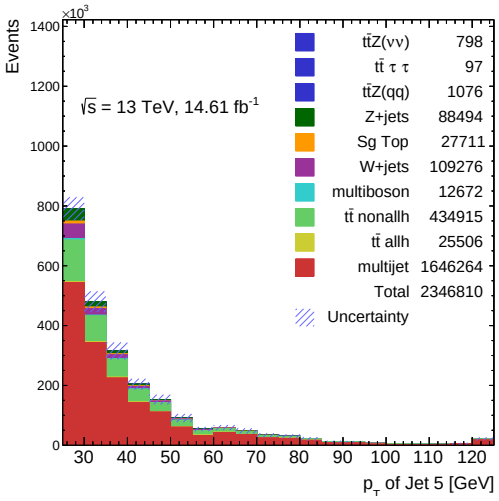
(c) p_T of second jet.



(d) p_T of third jet.

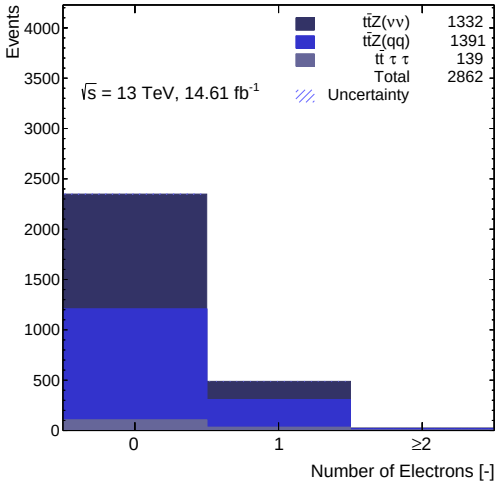


(e) p_T of fourth jet.

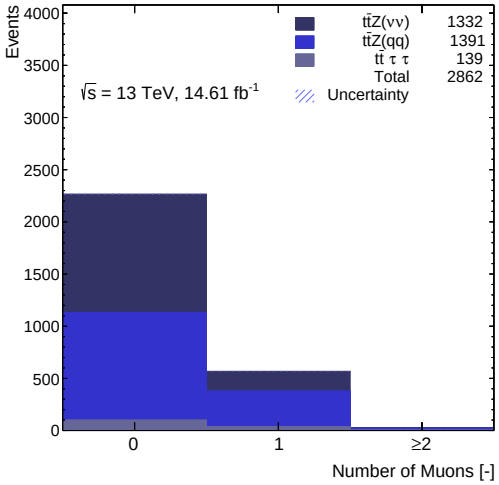


(f) p_T of fifth jet.

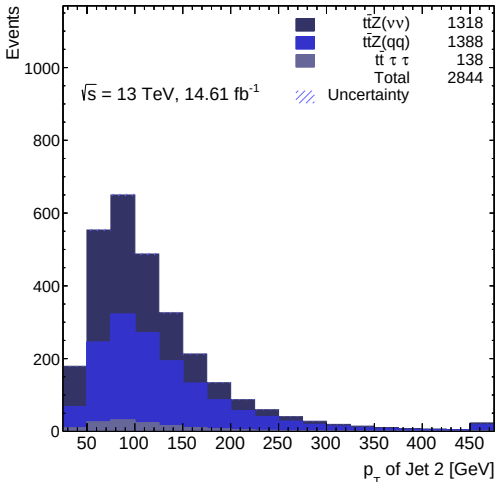
Figure A.1.: Signal and background distribution of various observables before applying event selection cuts, but with the requirement $\not{E}_T \geq 95\,\text{GeV}$.



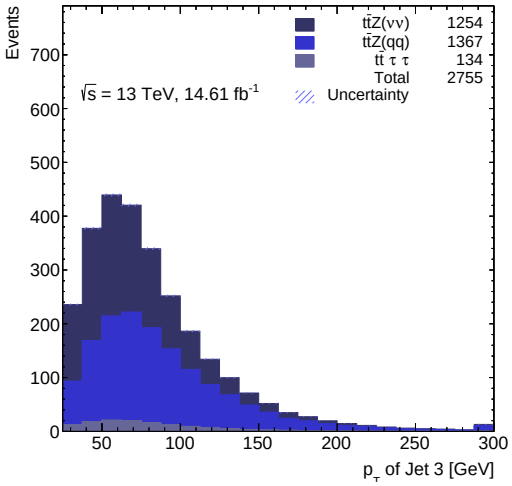
(a) Number of Electrons.



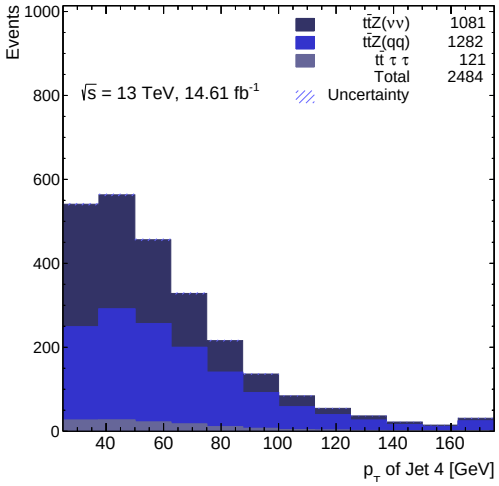
(b) Number of Muons.



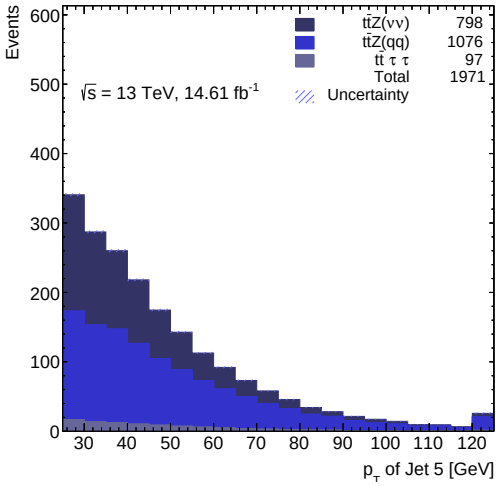
(c) p_T of second jet.



(d) p_T of third jet.

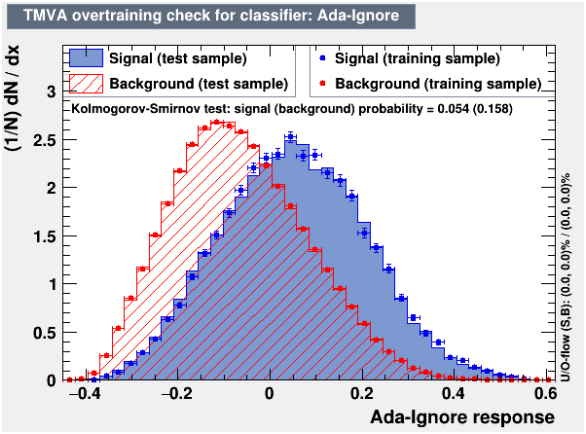


(e) p_T of fourth jet.

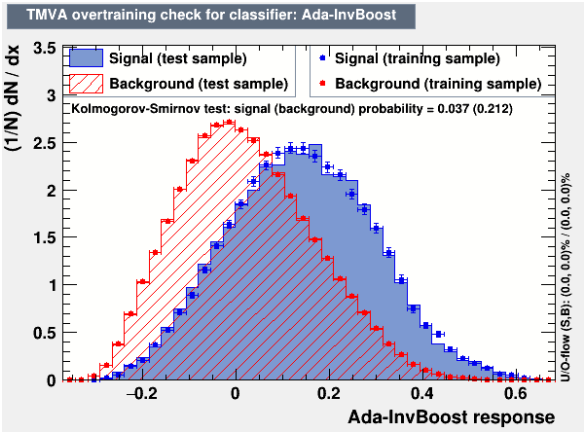


(f) p_T of fifth jet.

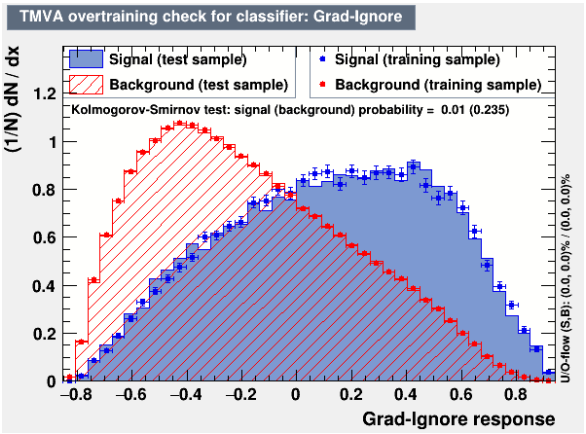
Figure A.2.: Signal distribution of various observables before applying event selection cuts, but with the requirement $\not{E}_T \geq 95\,\text{GeV}$.



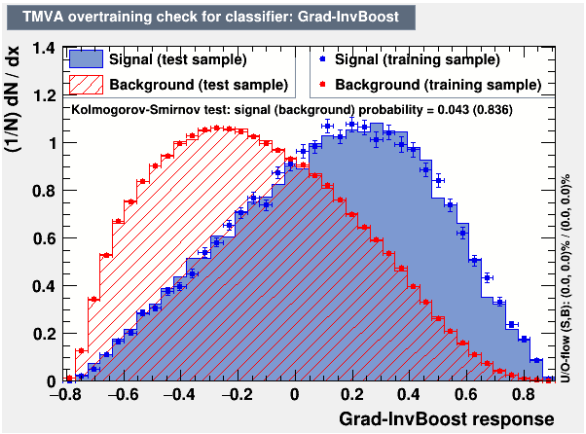
(a) AdaBoost, ignoring negative event weights.



(b) AdaBoost, inverse boosting negative event weights.

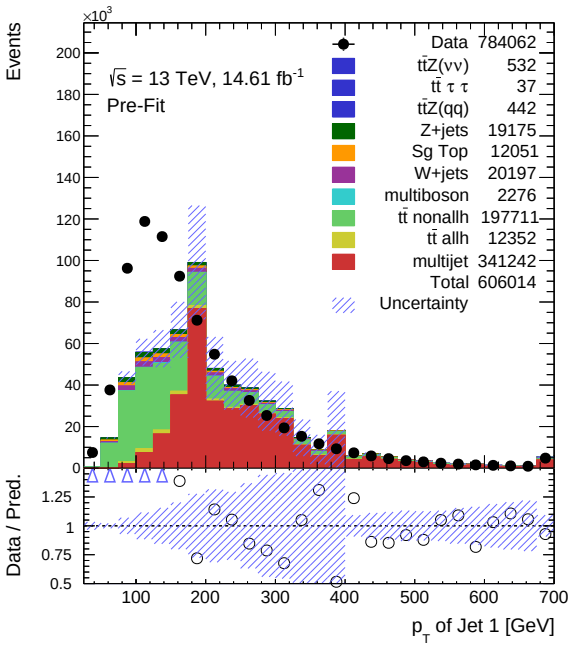


(c) Gradient boost, ignoring negative event weights.

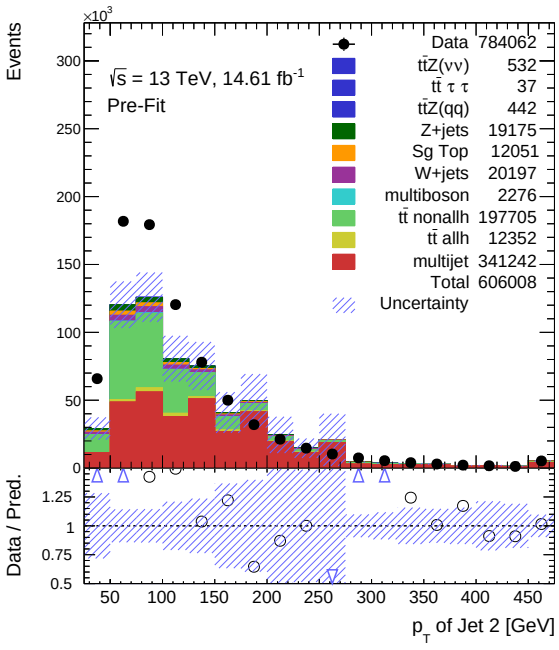


(d) Gradient boost, inverse boosting negative event weights.

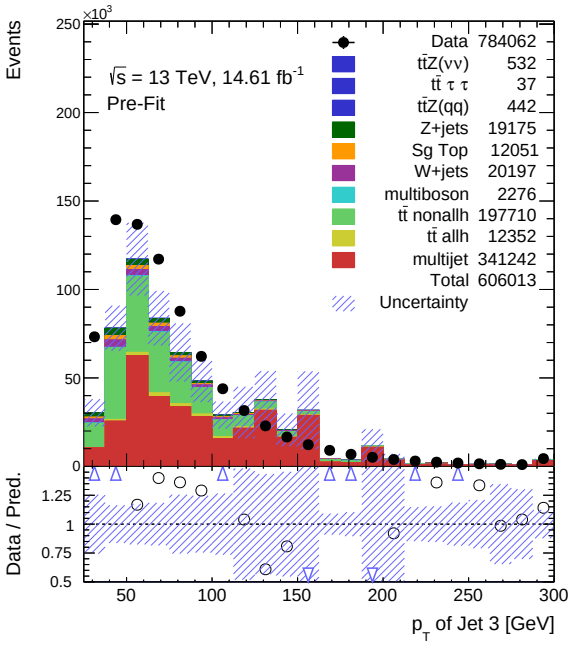
Figure A.3.: Overtraining tests for BDTs employing either gradient boosting or AdaBoost, and varied way of treating negative events, by either ignoring them in training and using them in testing, or boosting events with negative weights inversely.



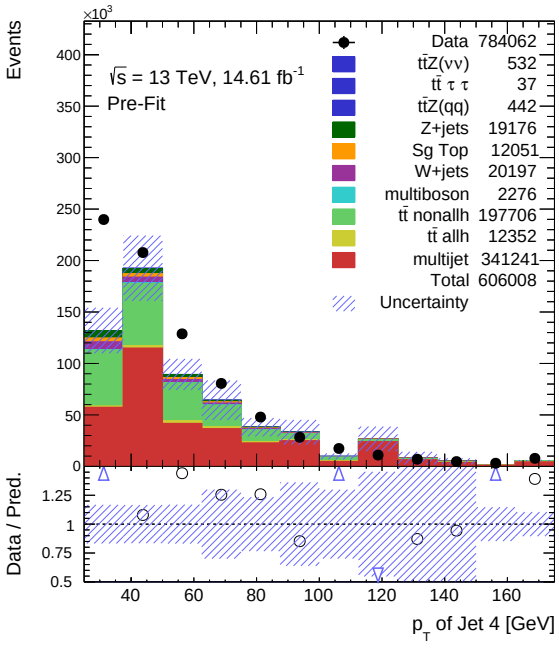
(a) p_T of first jet.



(b) p_T of second jet.

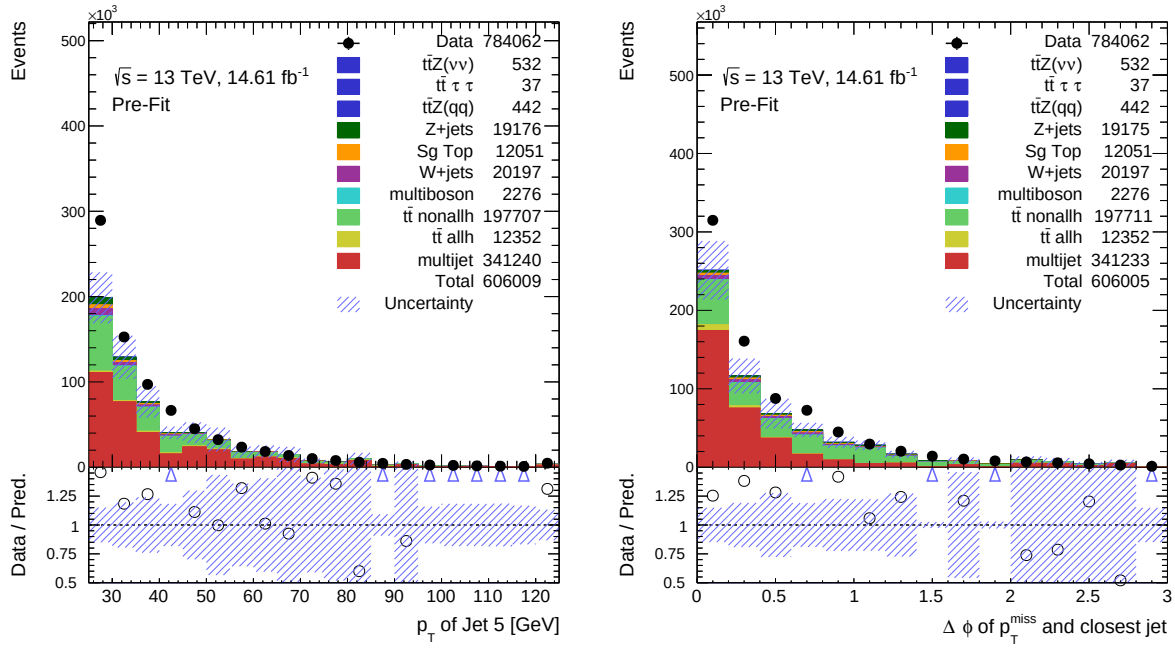


(c) p_T of third jet.



(d) p_T of fourth jet.

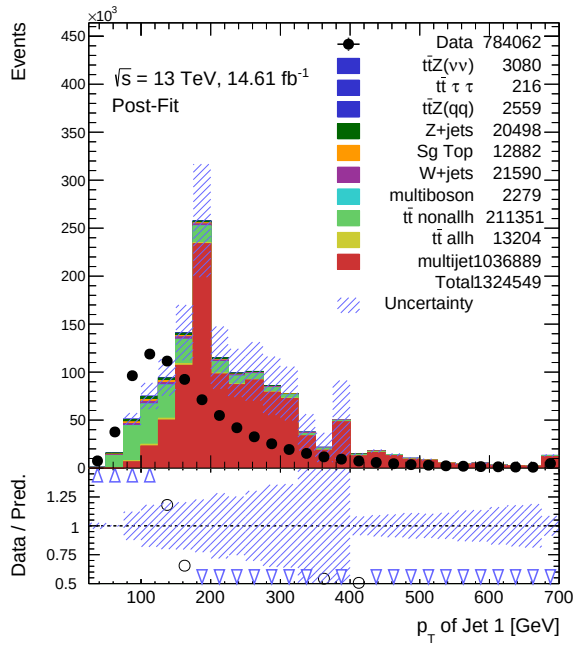
Figure A.4.: Monte Carlo and data distribution of various observables after applying event selection cuts, and before performing the fit.



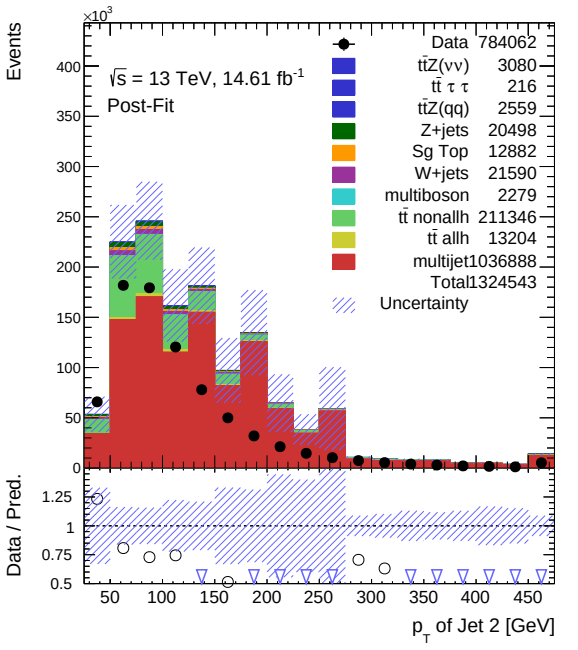
(a) p_T of fifth jet.

(b) $\Delta\phi_{\text{closest}}$.

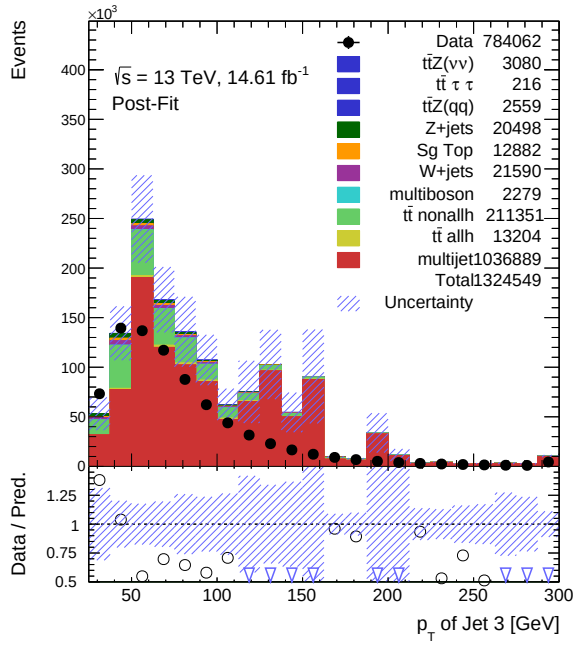
Figure A.5.: Monte Carlo and data distribution of various observables after applying event selection cuts, and before performing the fit.



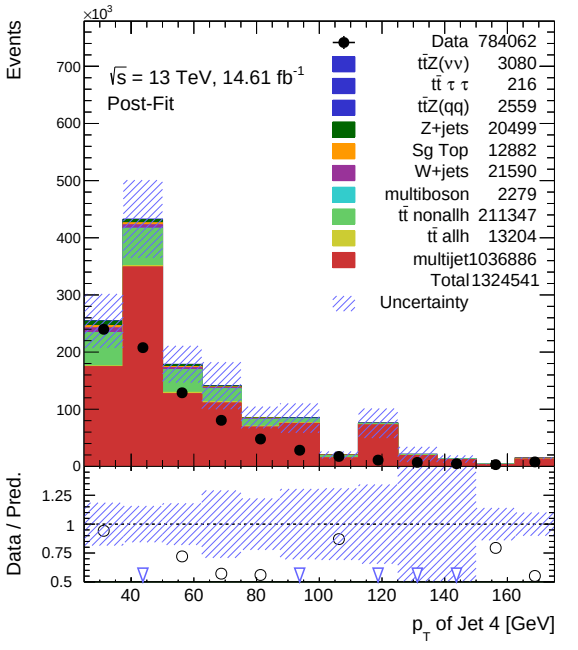
(a) p_T of first jet.



(b) p_T of second jet.

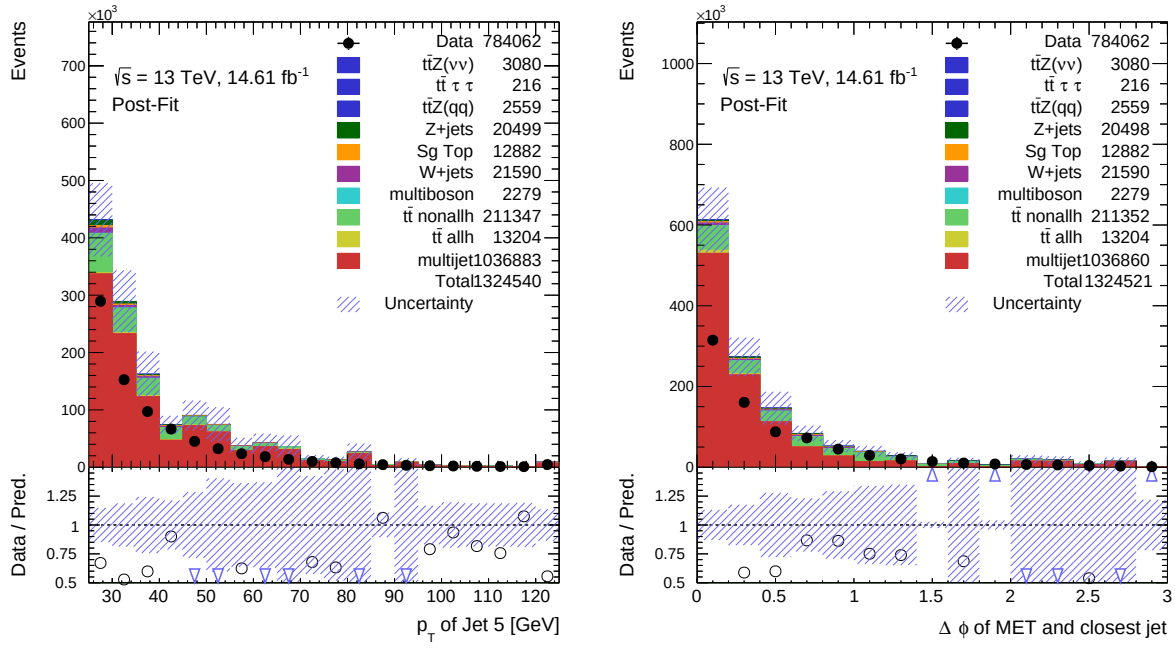


(c) p_T of third jet.



(d) p_T of fourth jet.

Figure A.6.: Monte Carlo and data distribution of various observables after performing the fit.



(a) p_T of fifth jet.

(b) $\Delta\phi_{\text{closest}}$.

Figure A.7.: Monte Carlo and data distribution of various observables after performing the fit.

B. Sample Files

B.1. Data Samples

Sample Name	integrated Lumi [pb^{-1}]
data15_13TeV.periodD.physics_Main.PhysCont	3212.96
data15_13TeV.periodE.physics_Main.PhysCont	
data15_13TeV.periodF.physics_Main.PhysCont	
data15_13TeV.periodG.physics_Main.PhysCont	
data15_13TeV.periodH.physics_Main.PhysCont	
data15_13TeV.periodJ.physics_Main.PhysCont	
data16_13TeV.periodA.physics_Main.PhysCont	11394.6
data16_13TeV.periodB.physics_Main.PhysCont	
data16_13TeV.periodC.physics_Main.PhysCont	
data16_13TeV.periodD.physics_Main.PhysCont	
data16_13TeV.periodE.physics_Main.PhysCont	
data16_13TeV.periodF.physics_Main.PhysCont	
data16_13TeV.periodG.physics_Main.PhysCont	
data16_13TeV.periodI.physics_Main.PhysCont	
data16_13TeV.periodK.physics_Main.PhysCont	
data16_13TeV.periodL.physics_Main.PhysCont	

Table B.1.: Data samples, corresponding to the full 2015 and 2016 datasets. For all samples, the `DAOD_TOPQ4` derivation is used with the tag `p2950`.

B.2. Monte Carlo Samples

For all samples, the `DAOD_TOPQ1` derivation is used, with the tag `p2952`, unless otherwise stated. Samples from the mc15c campaign were used.

Sample Name	Num Events	σ [pb]	k-factor	Lumi [fb^{-1}]
410156.aMcAtNloPythia8EvtGen_MEN30NLO_A14N23LO_ttZnunu	1495400	0.1549	1.11	9653.97030342
410157.aMcAtNloPythia8EvtGen_MEN30NLO_A14N23LO_ttZqq	2996400	0.52809	1.11	5674.03283531
410220.aMcAtNloPythia8EvtGen_MEN30NLO_A14N23LO_tttautau	934300	0.036515	1.12	25586.7451732

Table B.2.: NLO signal samples. The number of events, cross section and luminosity are given before the event selection.

Sample Name	Num Events	σ [pb]	k-factor	Lumi [fb^{-1}]
364142.Sherpa_221_NNPDF30NNLO_Znunu_MAXHTPTV0_70_CVetoBVeto	9923000	10706.0	0.9728	1.12812005971
364143.Sherpa_221_NNPDF30NNLO_Znunu_MAXHTPTV0_70_CFilterBVeto	7881000	10705.0	0.9728	6.61870033534
364144.Sherpa_221_NNPDF30NNLO_Znunu_MAXHTPTV0_70_BFilter	7926800	10705.0	0.9728	11.1896700097
364145.Sherpa_221_NNPDF30NNLO_Znunu_MAXHTPTV70_140_CVetoBVeto	14805000	607.65	0.9728	35.3495944524
364146.Sherpa_221_NNPDF30NNLO_Znunu_MAXHTPTV70_140_CFilterBVeto	14763800	607.59	0.9728	133.196029138
364147.Sherpa_221_NNPDF30NNLO_Znunu_MAXHTPTV70_140_BFilter	19728500	607.97	0.9728	271.432805776
364148.Sherpa_221_NNPDF30NNLO_Znunu_MAXHTPTV140_280_CVetoBVeto	14102800	222.33	0.9728	104.440331868
364149.Sherpa_221_NNPDF30NNLO_Znunu_MAXHTPTV140_280_CFilterBVeto	12364500	222.31	0.9728	246.896083718
364150.Sherpa_221_NNPDF30NNLO_Znunu_MAXHTPTV140_280_BFilter	9864900	222.38	0.9728	293.720148359
364151.Sherpa_221_NNPDF30NNLO_Znunu_MAXHTPTV280_500_CVetoBVeto	4937400	47.423	0.9728	186.293838446
364152.Sherpa_221_NNPDF30NNLO_Znunu_MAXHTPTV280_500_CFilterBVeto	3445500	47.401	0.9728	277.425823092
364153.Sherpa_221_NNPDF30NNLO_Znunu_MAXHTPTV280_500_BFilter	8887350	47.421	0.9728	1070.07995895
364154.Sherpa_221_NNPDF30NNLO_Znunu_MAXHTPTV500_1000	9869000	69.9101	0.9728	995.852715916
364155.Sherpa_221_NNPDF30NNLO_Znunu_MAXHTPTV1000_E_CMS	4951000	0.818	0.9728	6052.56723716
364128.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV0_70_CVetoBVeto	7907000	1981.8	0.9751	4.85720733113
364129.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV0_70_CFilterBVeto	4941000	1981.7	0.9751	22.0374210842
364130.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV0_70_BFilter	7890600	1982.0	0.9751	61.7679576054
364131.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV70_140_CVetoBVeto	5935500	110.61	0.9751	77.9024046943
364132.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV70_140_CFilterBVeto	1961200	110.56	0.9751	96.9862458747
364133.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV70_140_BFilter	5912550	110.62	0.9751	482.019330158
364134.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV140_280_CVetoBVeto	4956000	40.79	0.9751	199.76713263
364135.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV140_280_CFilterBVeto	2973000	40.716	0.9751	318.897576933
364136.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV140_280_BFilter	4932950	40.739	0.9751	900.808463795
364137.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV280_500_CVetoBVeto	1973000	8.6639	0.9751	406.393289088
364138.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV280_500_CFilterBVeto	986000	8.6756	0.9751	433.042795582
364139.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV280_500_BFilter	1974950	8.6794	0.9751	1314.2986815
364140.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV500_1000	2944800	1.8082	0.9751	1628.58090919
364141.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV1000_E_CMS	980000	0.14826	0.9751	6610.00944287

Table B.3.: NNLO Z+jets background samples. The number of events, cross section and luminosity are given before the event selection.

Sample Name	Num Events	σ [pb]	k-factor	Lumi [fb^{-1}]
410502.PowhegPythia8EvtGen_A14_ttbar_hdamp258p75_allhad	9789000	249.81	1.5216	39.1857811937
410501.PowhegPythia8EvtGen_A14_ttbar_hdamp258p75_nonallhad	58829000	397.05	1.1390	148.164530895

Table B.4.: NLO $t\bar{t}$ background samples. The number of events, cross section and luminosity are given before the event selection.

Sample Name	Num Events	σ [pb]	k-factor	Lumi [fb^{-1}]
361001.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ1	1997200	78420000000.0	1.0	1.02614903336e-06
361002.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ2	1998600	2433400000.0	1.0	8.18047772746e-05
361000.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ0	1999000	78420000000.0	1.0	2.61276776926e-08
361008.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ8	1989000	0.62505	1.0	308019.110286
361006.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ6	1796000	257.52	1.0	696.168456269
361004.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ4	1992400	254640.0	1.0	0.582171094954
361009.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ9	1179400	0.01964	1.0	4958787.48943
361007.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ7	1948800	16.214	1.0	10808.6714297
361005.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ5	1998800	4553.6	1.0	30.1124650251
361010.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ10	1995800	0.0011961	1.0	283190981.621
361011.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ11	1083600	4.226e-05	1.0	9584445983.2
361003.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ3	1832000	26454000.0	1.0	0.00586983276731

Table B.5.: LO multijet background samples. The number of events, cross section and luminosity are given before the event selection.

Sample Name	Num Events	σ [pb]	k-factor	Lumi [fb^{-1}]
304309.Sherpa_CT10_Wqq_Pt1000	15000	0.046461	1.0	322.851423775
304308.Sherpa_CT10_Wqq_Pt500_1000	29000	2.1559	1.0	13.4514587875
304307.Sherpa_CT10_Wqq_Pt280_500	148000	29.504	1.0	5.01626898048
364197.Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV1000_E_CMS	3946000	1.2341	0.9702	3197.47184183
364196.Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV500_1000	5945000	15.052	0.9702	394.964124369
364195.Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV280_500_BFilter	2954100	71.945	0.9702	305.828484665
364194.Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV280_500_CFilterBVeto	2956400	71.994	0.9702	129.753959667
364193.Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV280_500_CVetoBVeto	4931200	72.085	0.9702	125.360789627
364191.Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV140_280_CFilterBVeto	7265000	339.68	0.9702	73.6874305979
364190.Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV140_280_CVetoBVeto	9899000	339.78	0.9702	48.8637700974
364189.Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV70_140_BFilter	9857000	945.96	0.9702	100.280072466
364188.Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV70_140_CFilterBVeto	9860000	946.23	0.9702	46.8918166877
364187.Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV70_140_CVetoBVeto	14808500	944.97	0.9702	23.256756787
364186.Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV0_70_BFilter	17273200	19149.0	0.9702	20.2278755456
364185.Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV0_70_CFilterBVeto	9865600	19149.0	0.9702	3.98331404218
364184.Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV0_70_CVetoBVeto	24784000	19155.0	0.9702	1.56841727544

Table B.6.: LO and NNLO W+jets background samples. The number of events, cross section and luminosity are given before the event selection.

Sample Name	Num Events	σ [pb]	k-factor	Lumi [fb^{-1}]
361097.Sherpa_CT10_ZqqZvv_SHv21_improved	4483500	16.43	0.91	966.017888751
361095.Sherpa_CT10_WqqZvv_SHv21_improved	4962400	6.777	0.91	732.241404751

Table B.7.: NLO multiboson background samples with the tag `p3317`. The number of events, cross section and luminosity are given before the event selection.

Sample Name	Num Events	σ [pb]	k-factor	Lumi [fb^{-1}]
361627.Sherpa_CT10_ZZZ_2l4v	35000	0.0004453	0.000199561	175384.798633
361626.Sherpa_CT10_ZZZ_4l2v	34600	0.00044125	9.9467e-05	347855.548459
361625.Sherpa_CT10_ZZZ_6l0v	35000	1.7059e-05	1.7059e-05	2051702.91342
361624.Sherpa_CT10_WZZ_3l3v	49800	0.0019248	0.000855458	58214.4225432
361623.Sherpa_CT10_WZZ_5l1v	49800	0.00021783	0.00021783	228618.647569
361622.Sherpa_CT10_WWZ_2l4v	59800	0.0034299	0.0034299	17434.9106388
361621.Sherpa_CT10_WWZ_4l2v	59600	0.001734	0.001734	34371.3956171
361620.Sherpa_CT10_WWW_3l3v	59800	0.008343	0.008343	7167.68548484
361096.Sherpa_CT10_ZqqZll_SHv21_improved	3988900	16.445	0.91	1690.07837686
361094.Sherpa_CT10_WqqZll_SHv21_improved	3990500	3.4234	0.91	1165.65402816
361093.Sherpa_CT10_WlvZqq_SHv21_improved	3993600	11.494	0.91	347.450843919
361092.Sherpa_CT10_WpqqWmlv_SHv21_improved	3993700	24.857	0.91	160.667015328
361091.Sherpa_CT10_WplvWmqq_SHv21_improved	3993900	24.885	0.91	160.494273659
361077.Sherpa_CT10_ggllvv	6289000	0.85401	0.91	7364.08238779
361070.Sherpa_CT10_llvvjj_ss_EW6	10000	0.043004	0.91	232.536508232
361068.Sherpa_CT10_llvv	5929600	14.0	0.91	423.542857143

Table B.8.: NLO multiboson background samples with the tag `p2952`. The number of events, cross section and luminosity are given before the event selection.

Sample Name	Num Events	σ [pb]	k-factor	Lumi [fb^{-1}]
410013.PowhegPythiaEvtGen_P2012_Wt_inclusive_top.merge	4965000	34.009	1.054	145.99076715
410014.PowhegPythiaEvtGen_P2012_Wt_inclusive_antitop	4972000	33.989	1.054	146.282620848
410012.PowhegPythiaEvtGen_P2012_singletop_tchan_lept_antitop	4775000	25.778	1.0193	185.235472108
410011.PowhegPythiaEvtGen_P2012_singletop_tchan_lept_top	4941000	43.739	1.0094	112.965545623

Table B.9.: Single top background samples with the tag `p3317`. The W associated production samples are normalised to NNLO, the t channel samples are normalised to NLO. The number of events, cross section and luminosity are given before the event selection.

Bibliography

[1] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716(1)**, 1 (2012)

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

[3] Y. Ashie, et al. (The Super-Kamiokande Collaboration), *Evidence for an Oscillatory Signature in Atmospheric Neutrino Oscillations*, Phys. Rev. Lett. **93**, 101801 (2004)

[4] ATLAS Collaboration, *Search for $t\bar{t}Z$ production in the three lepton final state with 4.7 fb^{-1} of $\sqrt{s} = 7$ TeV pp collision data collected by the ATLAS detector*, Technical Report ATLAS-CONF-2012-126, CERN (2012)

[5] CMS Collaboration, *Measurement of associated production of vector bosons and top quark-antiquark pairs in pp collisions at $\sqrt{s} = 7$ TeV*, Phys. Rev. Lett. **110**, 172002 (2013)

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

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

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

[9] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. **C680519**, 367 (1968)

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

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

[12] H. D. Politzer, *Asymptotic freedom: An approach to strong interactions*, Physics Reports **14** (1974)

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

[14] H. Frauenfelder, et al., *Parity and the Polarization of Electrons from* Co^{60}, Phys. Rev. **106**, 386 (1957)

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

[16] R. P. Feynman, M. Gell-Mann, *Theory of the Fermi Interaction*, Phys. Rev. **109**, 193 (1958)

[17] P. J. Mohr, D. B. Newell, B. N. Taylor, *CODATA recommended values of the fundamental physical constants: 2014*, Rev. Mod. Phys. **88**, 035009 (2016)

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

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

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

[21] T. Aaltonen, et al. (CDF Collaboration), *Observation of Electroweak Single Top-Quark Production*, Phys. Rev. Lett. **103**, 092002 (2009)

[22] V. M. Abazov, et al. (DØ Collaboration), *Observation of Single Top-Quark Production*, Phys. Rev. Lett. **103**, 092001 (2009)

[23] ATLAS Collaboration, CDF Collaboration, CMS Collaboration, DØ Collaboration, *First combination of* TEVATRON *and* LHC *measurements of the top-quark mass* (2014), ATLAS-CONF-2014-008, CDF-NOTE-11071, CMS-PAS-TOP-13-014, D0-NOTE-6416

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

[25] CMS Collaboration, *Constraints on the Top-Quark Charge from Top-Pair Events*, Technical Report CMS-PAS-TOP-11-031, CERN (2012)

[26] I. Bigi, et al., *Production and decay properties of ultra-heavy quarks*, Phys. Lett. B **181(1,2)**, 157 (1986)

[27] J. Pumplin, et al., *New Generation of Parton Distributions with Uncertainties from Global QCD Analysis*, JHEP **2002(07)**, 012 (2002)

[28] J. M. Campbell, R. K. Ellis, $t\bar{t}W^{\pm}$ *production and decay at NLO*, JHEP **2012(7)**, 52 (2012)

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

[30] M. V. Garzelli, et al., $t\bar{t}W^{\pm} + t\bar{t}Z$ *hadroproduction at NLO accuracy in QCD with Parton Shower and Hadronization effects*, JHEP **2012(11)**, 56 (2012)

[31] 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 **2015(11)**, 172 (2015)

[32] Cms Collaboration, *Observation of top quark pairs produced in association with a vector boson in pp collisions at* $\sqrt{s} = 8$ *TeV*, JHEP **2016(1)**, 96 (2016)

[33] Atlas Collaboration, *Measurement of the* $t\bar{t}Z$ *and* $t\bar{t}W$ *production cross sections in multilepton final states using* $3.2\,fb^{-1}$ *of pp collisions at* $\sqrt{s} = 13$ *TeV with the Atlas detector*, Eur. Phys. J. C **77(1)**, 40 (2017)

[34] Cms Collaboration, *Measurement of the cross section for top quark pair production in association with a W or Z boson in proton-proton collisions at* $\sqrt{s} = 13$ *TeV* (2017), accepted for publication in JHEP, `1711.02547`

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

[36] K. Aamodt, et al. (Alice Collaboration), *The Alice experiment at the Cern Lhc*, JINST **3(08)**, S08002 (2008)

[37] A. Augusto Alves Jr., et al. (LHCb Collaboration), *The LHCb Detector at the Lhc*, JINST **3(08)**, S08005 (2008)

[38] Atlas Collaboration, *The Atlas Experiment at the Cern Large Hadron Collider*, JINST **3(08)**, S08003 (2008)

[39] Cms Collaboration, *The Cms experiment at the Cern Lhc*, JINST **3(08)**, S08004 (2008)

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

[41] V. Berardi, et al. (TOTEM Collaboration), *TOTEM: Technical design report. Total cross section, elastic scattering and diffraction dissociation at the Large Hadron Collider at CERN*, Technical Report CERN-LHCC-2004-002, TOTEM-TDR-001, CERN (2004)

[42] ATLAS Collaboration, *Performance of the ATLAS Inner Detector Track and Vertex Reconstruction in the High Pile-Up LHC Environment*, Technical Report ATLAS-CONF-2012-042, CERN (2012)

[43] ATLAS Collaboration (ATLAS Collaboration), *ATLAS Insertable B-Layer Technical Design Report*, Technical Report CERN-LHCC-2010-013. ATLAS-TDR-19, CERN (2010)

[44] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report Addendum*, Technical Report CERN-LHCC-2012-009. ATLAS-TDR-19-ADD-1, CERN (2012), addendum to CERN-LHCC-2010-013, ATLAS-TDR-019

[45] ATLAS Collaboration, *The ATLAS Data Acquisition System: from Run 1 to Run 2*, Nucl. Part. Phys. Proc. **273(Supplement C)**, 939 (2016)

[46] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70(3)**, 823 (2010)

[47] S. Agostinelli, et al. (GEANT4), *GEANT4: A Simulation toolkit*, Nucl. Instrum. Meth. **A506**, 250 (2003)

[48] T. Sjöstrand, et al., *An Introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191**, 159 (2015)

[49] D. de Florian, et al. (LHC Higgs Cross Section Working Group), *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, Technical report (2016)

[50] R. D. Ball, et al., *Parton distributions with LHC data*, Nucl. Phys. B **867(2)**, 244 (2013)

[51] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021 (2014)

[52] S. Alioli, et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06**, 043 (2010)

[53] M. Czakon, A. Mitov, *Top++: A program for the calculation of the top-pair cross-section at hadron colliders*, Comput. Phys. Commun. **185(11)**, 2930 (2014)

[54] T. Sjöstrand, S. Mrenna, P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **0605**, 026 (2006)

[55] P. Z. Skands, *Tuning Monte Carlo generators: The Perugia tunes*, Phys. Rev. D **82**, 074018 (2010)

[56] M. Aliev, et al., *HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR*, Comput. Phys. Commun. **182**, 1034 (2011)

[57] P. Kant, et al., *HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions*, Comput. Phys. Commun. **191**, 74 (2015)

[58] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^-*, Phys. Rev. D **82**, 054018 (2010)

[59] N. Kidonakis, *Top Quark Production*, in *Proceedings, Helmholtz International Summer School on Physics of Heavy Quarks and Hadrons (HQ 2013): JINR, Dubna, Russia, July 15-28, 2013* (2014)

[60] T. Gleisberg, et al., *Event generation with SHERPA 1.1*, JHEP **0902**, 007 (2009)

[61] T. Gleisberg, S. Höche, *Comix, a new matrix element generator*, JHEP **0812**, 039 (2008)

[62] F. Cascioli, P. Maierhofer, S. Pozzorini, *Scattering Amplitudes with Open Loops*, Phys. Rev. Lett. **108**, 111601 (2012)

[63] S. Höche, et al., *QCD matrix elements + parton showers: The NLO case*, JHEP **04**, 027 (2013)

[64] K. Melnikov, F. Petriello, *Electroweak gauge boson production at hadron colliders through $\mathcal{O}(\alpha_s^2)$*, Phys. Rev. D **74**, 114017 (2006)

[65] R. Gavin, et al., *FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order*, Comput. Phys. Commun. **182(11)**, 2388 (2011)

[66] S. Quackenbush, et al., *W physics at the LHC with FEWZ 2.1*, Comput. Phys. Commun. **184(1)**, 209 (2013)

[67] Y. Li, F. Petriello, *Combining QCD and electroweak corrections to dilepton production in the framework of the FEWZ simulation code*, Phys. Rev. D **86**, 094034 (2012)

[68] ATLAS Collaboration, *Single Boson and Diboson Production Cross Sections in pp Collisions at* $\sqrt{s} = 7$ *TeV*, Technical Report ATL-COM-PHYS-2010-695, CERN, Geneva (2010)

[69] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462**, 152 (2001)

[70] ATLAS Collaboration, *Expected electron performance in the ATLAS experiment*, ATL-PHYS-PUB-2011-006

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

[72] ATLAS Collaboration, *Calorimeter Clustering Algorithms: Description and Performance*, Technical Report ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, CERN, Geneva (2008)

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

[74] ATLAS Collaboration, *Expected performance of the ATLAS b-tagging algorithms in Run-2*, ATL-PHYS-PUB-2015-022

[75] ATLAS Collaboration, *Recommendations of the Physics Objects and Analysis Harmonisation Study Groups 2014*, Technical Report ATL-PHYS-INT-2014-018, CERN, Geneva (2014)

[76] Y. Freund, R. E. Schapire, *A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting*, J. Comput. Syst. Sci. **55(1)**, 119 (1997)

[77] L. Breiman, *Arcing the edge*, Technical report, Statistics Department, University of California at Berkeley (1997)

[78] A. Hoecker, et al., *TMVA: Toolkit for Multivariate Data Analysis*, PoS **ACAT**, 040 (2007)

[79] R. Brun, F. Rademakers, *ROOT - An object oriented data analysis framework*, Nucl. Instr. Meth. Phys. Res. A **389(1-2)**, 81 (1997)

[80] A. Kolmogoroff, *Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung*, Math. Ann. **104(1)**, 415 (1931)

[81] N. V. Smirnov, *On approximation of distribution laws of random variables by empirical data*, Uspekhi Mat. Nauk **10**, 179 (1944)

[82] K. Cranmer, et al. (ROOT Collaboration), *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, Technical Report CERN-OPEN-2012-016, New York U., New York (2012)

[83] Atlas Collaboration, *Modelling of the $t\bar{t}H$ and $t\bar{t}V$ ($V = W, Z$) processes for $\sqrt{s} = 13$ TeV ATLAS analyses*, Technical Report ATL-PHYS-PUB-2016-005, CERN, Geneva (2016)

[84] Atlas LUCID Collaboration, *Measurement of the luminosity with the new LUCID-2 detector in 2015*, Technical Report ATL-COM-FWD-2016-008, CERN, Geneva (2016)

Acknowledgements

Seeing that I have completed this master thesis feels like a miracle happening after coming across all the uncountable obstacles throughout the analysis and writing process. This would not have been possible without the support of several people.

First of all, I would like to thank my supervisor and first referee Prof. Dr. Arnulf Quadt, who sparked my interest in particle physics seven years ago, and has provided academic guidance all the way from my bachelor thesis to my master thesis, now. While expecting only the highest quality scientific output from his students, he always had our best interest in mind, and, with this, I would like to confirm that no matter how busy I was, yes, I was always eating regularly.

Second, I would like to thank Prof. Dr. Stan Lai for his unhesitant willingness to be my second referee. I am grateful to Dr. Clara Nellist, Dr. Nils-Arne Rosien, and Dr. Thomas Peiffer for all the advice, and support in questions about the analysis and the following documentation into this thesis.

I would also like to thank Dr. Elizaveta Shabalina for sharing her knowledge in the weekly top meetings and steering my results in the right direction. Special thanks go to Ishan Pokharel and Tomas Dado for their patience and efforts in explaining to me what I am (supposed to be) doing and providing troubleshooting ideas for technical issues. Also thanks to everyone at the II. Physikalisches Institut for the friendly, helpful atmosphere.

Despite being in physics, I did not swear off private life completely, and so I would like to thank my friends who took my mind off work every once in a while. In addition to that, I am deeply grateful for all the loving support I received from Darren throughout even the busiest of phases. You always knew how to help me through my frustration and stress, be it in Germany, Singapore, or even in Japan.

Thanks a lot also to my brother, who made me see everything from different perspectives that are not my figurative science goggles, keeping me grounded in some crucial moments. Most of my thanks I probably owe to my parents, who always encouraged me to follow my own path, and, although not understanding anything of what I did from the beginning of university onwards, listened to my problems and always supported me to the best of their abilities.

Erklärung nach §17(9) der Prüfungsordnung für den Bachelor-Studiengang Physik und den Master-Studiengang Physik an der Universität Göttingen:

Hiermit erkläre ich, dass ich diese Abschlussarbeit selbständig verfasst habe, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen wurden, als solche kenntlich gemacht habe.
Darüberhinaus erkläre ich, dass diese Abschlussarbeit nicht, auch nicht auszugsweise, im Rahmen einer nichtbestandenen Prüfung an dieser oder einer anderen Hochschule eingereicht wurde.

Göttingen, den 20. Juli 2018

(Marie Reinecke)