Observation of the Standard Model Higgs boson produced in association with a pair of top quarks at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS experiment at the LHC with emphasis on the decay of the Higgs boson into a $b\bar{b}$-pair in the single-lepton channel

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades
„Doctor rerum naturalium“
der Georg-August-Universität Göttingen

im Promotionsprogramm ProPhys
der Georg-August University School of Science (GAUSS)

vorgelegt von

Johannes Donatus Mellenthin

aus München

Göttingen, 2019

Betreuungsausschuss

Prof. Dr. Arnulf Quadt
Prof. Dr. Stan Lai

Mitglieder der Prüfungskommission:

Referent: Prof. Dr. Arnulf Quadt
II. Physikalisches Institut, Georg-August-Universität Göttingen

Koreferent: Prof. Dr. Stan Lai
II. Physikalisches Institut, Georg-August-Universität Göttingen

Weitere Mitglieder der Prüfungskommission:

PD Dr. Ralf Bernhard
II. Physikalisches Institut, Georg-August-Universität Göttingen

Prof. Dr. Ariane Frey
II. Physikalisches Institut, Georg-August-Universität Göttingen

Prof. Dr. Wolfram Kollatschny
Institut für Astrophysik, Georg-August-Universität Göttingen

Prof. Dr. Steffen Schumann
Institut für Theoretische Physik, Georg-August-Universität Göttingen

Tag der mündlichen Prüfung: 21. August 2019

Referenz: II.Physik-UniGö-Diss-2019/02

GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Observation of the Standard Model Higgs boson produced in association with a pair of top quarks at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS experiment at the LHC with emphasis on the decay of the Higgs boson into a $b\bar{b}$-pair in the single-lepton channel

Abstract

The top quark is the heaviest elementary particle in the Standard Model and has an expected Yukawa coupling to the Higgs boson of order unity. The value of this coupling is a key ingredient to unravel the nature of the observed Higgs boson. The most favourable production mode that has a direct sensitivity to this coupling is the production of a Higgs boson in association with a top-quark pair, $t\bar{t}H$. This process was observed based on the analysis of proton-proton collision data at $\sqrt{s} = 13\,\text{TeV}$ recorded with the ATLAS experiment at the LHC. Using data corresponding to integrated luminosities of up to $79.8\,\text{fb}^{-1}$, and considering the Higgs boson decays into $b\bar{b}$, WW^*, $\tau^+\tau^-$, $\gamma\gamma$, and ZZ^* yields a signal strength of

$$\mu = 1.32 \pm 0.18(\text{stat.})^{+0.21}_{-0.19}(\text{syst.}) = 1.32^{+0.28}_{-0.26},$$

corresponding to an observed (expected) signal significance of 5.8 (4.9) standard deviations. The analysis targeting the Higgs boson decay channel with the highest branching ratio, $t\bar{t}H(H \to b\bar{b})$, uses data corresponding to an integrated luminosity of $36.1\,\text{fb}^{-1}$ and will be presented in detail. A focus is placed on the single-lepton channel. The dominant background for this channel is $t\bar{t}b\bar{b}$. One of the small backgrounds originates from non-prompt leptons and fake leptons, which originate from jets misidentified as a reconstructed lepton. This background requires a special treatment in signal regions with many jets and b-jets. Despite its small contribution, an estimate of non-prompt leptons and fake leptons is important for a successful measurement in the $t\bar{t}H$ analysis as well as other analyses with leptonic final states. In this thesis, a fully data-driven technique – the matrix method – is presented. For the first time, efficiencies for the 2017 dataset are shown. In addition, a tag rate function could be employed to increase the performance of the matrix method for a fixed b-tagging working point. Finally, the performance of neural networks using low-level input variables is examined to discriminate the $t\bar{t}H$ signal from backgrounds.

GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Beobachtung der Produktion eines Standardmodell-Higgs-Bosons in Assoziation mit einem Top-Quark-Paar bei $\sqrt{s} = 13\,\text{TeV}$ mit dem ATLAS-Experiment am LHC mit Schwerpunkt auf dem Zerfall des Higgs-Bosons in ein $b\bar{b}$-Paar im Mono-Lepton-Kanal

Zusammenfassung

Das Top-Quark ist das schwerste Elementarteilchen des Standardmodells und hat eine erwartete Yukawa-Kopplung an das Higgs-Boson von etwa eins. Der Wert dieser Kopplung ist essentiell, um weitere Eigenschaften des beobachteten Higgs-Bosons präzise bestimmen zu können. Der bevorzugte Produktionsmechanismums mit einer direkten Sensitivität auf diese Kopplung ist die Produktion eines Higgs-Bosons in Assoziation mit einem Top-Quark-Paar, $t\bar{t}H$. Dieser Prozess wurde basierend auf der Analyse von Proton-Proton-Kollisionsdaten bei $\sqrt{s} = 13\,\text{TeV}$ mit dem ATLAS-Experiment am LHC beobachtet. Hierfür wurden Daten, die integrierten Luminositäten von bis zu $79.8\,\text{fb}^{-1}$ entsprechen, verwendet und Higgs-Boson Zerfälle in $b\bar{b}$, WW^*, $\tau^+\tau^-$, $\gamma\gamma$ und ZZ^* berücksichtigt. Daraus ergibt sich eine Signalstärke von

$$\mu = 1.32 \pm 0.18(\text{stat.})^{+0.21}_{-0.19}(\text{syst.}) = 1.32^{+0.28}_{-0.26},$$

was einer beobachteten (erwarteten) Signal-Signifikanz von 5.8 (4.0) Standardabweichungen entspricht. Die auf den Higgs-Boson-Kanal mit der höchsten Zerfallswahrscheinlichkeit, $t\bar{t}H(H \to b\bar{b})$, spezialisierte Analyse wird im Detail vorgestellt und verwendet Daten, die einer integrierten Luminosität von $36.1\,\text{fb}^{-1}$ entsprechen. Der Schwerpunkt liegt hierbei auf dem Mono-Lepton-Kanal. Der dominierende Untergrund in diesem Zerfallskanal ist $t\bar{t}b\bar{b}$. Ein weiterer Untergrund entsteht durch sekundäre Leptonen und durch als Leptonen fehlidentifizierte Jets. Dieser Untergrund benötigt eine gesonderte Behandlung in Signalregionen mit vielen Jets und b-Jets. Trotz eines nur geringen Anteils ist eine Abschätzung dieses Untergrundes wichtig für eine erfolgreiche Messung in der $t\bar{t}H$-Analyse sowie in Analysen mit leptonischen Endzuständen. In dieser Doktorarbeit wird die Modellierung dieses Untergrundes mittels einer datengestützten Methode, der Matrixmethode, vorgestellt. Zum ersten Mal werden Effizienzen für die 2017 aufgezeichneten Daten gezeigt. Zusätzlich wird eine Tag-Rate-Funktion verwendet, um die Performance der Matrixmethode bei einem festen b-Tagger-Arbeitspunkt zu verbessern. Abschließend wird die Performance von neuronalen Netzen, trainiert an Low-Level-Variablen, untersucht, um das $t\bar{t}H$-Signal vom Untergrund zu trennen.

Acknowledgements

First, I would like to thank Arnulf Quadt, my PhD supervisor, for giving me the opportunity to join his working group at the II. Physikalisches Institut, Georg-August-Universität Göttingen. His continuous support was essential for the completion of this thesis. I acquired a broad range of skills during my time in Göttingen and during a one year stay at CERN. I am thankful for the possibility to participate in many conferences and workshops.
Further, I am very grateful to Elizaveta Shabalina for all her help and support.
Thank you to Thomas Peiffer for proofreading my thesis and his helpful comments.
Thank you to Clara Nellist for final comments.
A special thank you to Nedaa Alexandra Asbah, Jeff Dandoy, Nello Bruscino, María Moreno Llácer, Claire David, and Roger Caminal Armadans for their time and patience to answer all my questions.
Thank you to my former office-mate Tomas Dado for his explanations and for the Slovak Tea.
Thank you to my friends and colleagues, especially Ishan Pokharel, for a very warm welcome to Göttingen.
Amanda, thank you for all your support. Thank you for your patience and confidence in me. And probably, I can't thank you enough for reading the thesis over and over.
Finally, thank you to my parents for their continuous support throughout all the years. It would not have been possible without you.

Contents

CHAPTER 1

Introduction

The discovery of the Higgs boson in 2012 by the ATLAS and CMS Collaborations [1,2] is a milestone in the history of particle physics and ended the almost 50 year long search for this boson. The Higgs boson was the last predicted, missing particle from the Standard Model (SM). After its discovery, measurements to examine the nature of the Higgs boson became the focus in order to further probe the SM. One property of fundamental interest is the coupling strength of the Higgs boson to the top quark, the heaviest elementary particle in the SM. A direct test of this coupling can be performed through the production of a Higgs boson in association with a top-quark pair, $t\bar{t}H$. However, this production process only contributes to about 1 % of the total Higgs boson production cross-section. The $t\bar{t}H$ production mechanism was observed by combining analyses targeting the Higgs boson decays into $b\bar{b}$, WW^*, $\tau^+\tau^-$, $\gamma\gamma$, and ZZ^* [3].

In this thesis, the measurement of the $t\bar{t}H$ production is presented. An emphasis is given on the analysis of the Higgs boson decay channel with the highest branching ratio $H \to b\bar{b}$ in the single-lepton channel. In this context, the estimation of non-prompt and fake leptons is examined and the performance of neural networks is studied.

At first, an introduction to the theory of the SM is given (Chapter 2). Hereby, the focus is set on two prominent particles in this analysis – the top quark and the Higgs boson. This introduction is followed by a description of the ATLAS detector at the Large Hadron Collider (LHC) in Chapter 3. Afterwards, the reconstruction mechanisms of the ATLAS detector are described including how Monte Carlo simulations are used to compare theoretical predictions to data (Chapter 4).

The main Chapter 5 presents the $t\bar{t}H(H \to b\bar{b})$ analysis [4], which comprises Section 5.3 where a data driven method to estimate non-prompt and fake leptons is studied. Section 5.5 describes the multivariate analysis techniques used in this analysis and also includes an independent study on deep neural networks employing different input variables.

Subsequently, the analyses of the remaining Higgs boson decay channels are summarised in Chapter 6 and the combination of these channels resulting in the observation of the $t\bar{t}H$ process is presented. The thesis concludes by a comparison to the $t\bar{t}H$ analysis of the CMS experiment and an outlook is given for future prospects of this analysis (Chapter 7).

Personal contributions

The success of the ATLAS experiment, one of the largest particle detectors in the world, would not be possible without the coordinated collaboration of more than 3000 scientists. The following list highlights my personal contributions for the collaboration in the scope of this thesis in chronological order.

- Improvements to algorithms for several calibration scans for the pixel detector and IBL.
- Offline monitoring of the inner detector and pixel detector data quality as well as general data quality monitoring in the ATLAS control room during data taking periods.
- Derivation contact for the jet/$E_{\mathrm{T}}^{\mathrm{miss}}$ group including submission and monitoring of data samples and work on the derivation software framework.
- Studies on the performance of deep neural networks with different high- and low-level input variables for the $t\bar{t}H(H \to b\bar{b})$ analysis in the single-lepton channel.
- Fake and non-prompt lepton background estimation for the $t\bar{t}H(H \to b\bar{b})$ analysis in the single-lepton channel.
  - Supporting the validation of the matrix method.
  - Extension of the matrix method to employ a tag rate function in order to reduce statistical uncertainties in high b-jet multiplicity regions.
  - Migration of the matrix method software packages to an updated analysis software release version resulting in first efficiencies for the 2017 data, which can also be used by other analyses with leptonic final states.
- Code quality review.

CHAPTER 2

The top quark and Higgs boson within the Standard Model

First, a general overview of the underlying theory in this thesis will be presented, the Standard Model of particle physics. This is followed by a more detailed description of the two most important elementary particles in this analysis, the top quark and the Higgs boson, including the concept of the Brout-Englert-Higgs mechanism [5–8]. An emphasis will be set on the production and decay processes for proton-proton collisions.

2.1. The Standard Model of particle physics

The Standard Model of particle physics (SM) is the theory that describes, to our best knowledge, Nature on a quantum level and predicts the interaction of elementary particles [9–11]. It includes three of the four fundamental forces; the electromagnetic, strong, and weak force. Gravitation, the weakest force, is not part of the SM and is described by the theory of general relativity. So far, no consistent quantum field description exists for gravity. The SM represents the combination of three renormalisable gauge field theories based on the $\mathrm{SU(3)_C \times SU(2)_L \times U(1)_Y}$ gauge symmetry, where $\mathrm{SU(3)_C}$ describes the strong force with the theory of quantum chromodynamics (QCD) and $\mathrm{SU(2)_L \times U(1)_Y}$ describes the electroweak interaction, which is the unification of quantum electrodynamics (QED) with the weak force. In the framework of the SM, particles are described as excitations of scalar and vector fields. These elementary particles (see Figure 2.1) can be split into two groups; fermions with half-integer spin and bosons with integer spin. Fermions include all quarks (up, down, charm, strange, top, and bottom) and six leptons. Charged leptons are electron (e), muon (μ), and tau (τ), each of which has an associated neutrino (ν_l). Bosons can be grouped into spin one gauge bosons, which mediate fundamental forces, photons (γ) for the exchange with the electromagnetic force, gluons (g) for the transfer of the strong force, and charged $W^{\pm}$ and neutral Z bosons transmitting the weak force. Finally, the one spin zero scalar boson, the Higgs boson (H), is associated with the Brout-Englert-Higgs mechanism that gives elementary particles their masses [5–8]. In addition, all particles have an associated antiparticle with the same mass but opposite internal quantum numbers, like the electrical charge. Neutrinos, the only fermions with zero electric charge, either differ from their antiparticles as all other fermions (called Dirac fermion) or are there own antiparticles (called Majorana fermion). It is not known which category applies to neutrinos and experiments are conducted to identify this property, e.g. searching for a neutrinoless double beta decay. When a particle and its antiparticle interact with each other they annihilate and produce other particles.

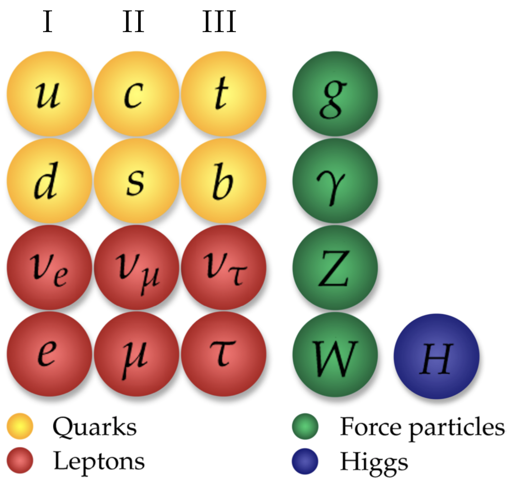


Figure 2.1.: Illustration of the Standard Model of particle physics. The quarks and leptons in each column represent one of the three generations.

The different fundamental forces can be associated with a conserved charge. For QED it is the *electric charge*, for QCD the *colour charge*, and for the weak interaction the *weak hypercharge* related to the *weak isospin*. Photons are the only gauge bosons that do not carry their respective charge, preventing them from self-interaction. This characteristic is also reflected in the abelian structure of $U(1)_{\mathrm{EM}}$. For the non-abelian groups $\mathrm{SU}(3)_{\mathrm{C}}$ and $\mathrm{SU}(2)_{\mathrm{L}}\times\mathrm{U}(1)_{\mathrm{Y}}$ self-interaction of their respective gauge bosons is possible. This theory is represented by the Lagrangian of the Standard Model

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{\Psi}\not{D}\Psi + \Psi_i Y_{ij}\Psi_j\Phi + h.c. + |D_\mu\Phi|^2 - V(\Phi), \tag{2.1}$$

where the first term describes the fundamental forces, the second term details how these forces act on fermions (leptons and quarks), and the remaining terms determine how these particles obtain their masses from the Higgs field Φ.

The Standard Model has been very successful in predicting new particles and is very effective in describing quantum mechanical phenomena. One example is the anomalous magnetic moment of the electron, where the prediction agrees with the experimentally measured value to twelve significant digits, making it the most precise prediction of physics [12]. Another example is the discovery of third generation particles. In 1973, Makoto Kobayashi and Toshihide Maskawa predicted the existence of a third generation of quarks to explain observed violations of CP-symmetry (charge conjugation parity symmetry) in kaon decays [13]. Two years later, the tau lepton was discovered at SLAC [14]. This particle improved the credibility of the prediction of third generation quarks, while implying the existence of a third-generation neutrino. The bottom quark was then discovered at Fermilab in 1977 [15]. This discovery strongly suggested that there must also be a sixth quark, the top quark, although it took almost 20 years until it was found in 1995 (see Section 2.2 for more details). The final third generation particle, the tau neutrino, was discovered at Fermilab in 2000 [16] making the prediction of the Higgs boson the last missing particle of the SM. This missing piece of the puzzle was discovered at CERN in 2012, see Section 2.3.

There are, however, phenomena that cannot be explained by the SM. One example is neutrino oscillations, where neutrinos change their flavour periodically [17–19]. This behaviour is only possible for neutrinos with different mass eigenstates, requiring at least two of the three neutrinos to have a mass greater than zero. Another example is the baryon asymmetry of the universe that is not explained by the SM [20]. A natural assumption is that at the time of the Big Bang the universe was neutral with respect to all conserved charges, including lepton and baryon numbers. A consequence of this is, an equal amount of matter and antimatter. However, in the observable universe, no significant amount of antimatter has been found, which suggests that there must be more matter than antimatter left after the Big Bang. Furthermore, the SM only describes 5 % of the overall mass-energy of the universe, the barionic matter, and neutrinos. The remaining part, accounting for dark matter (27 %) and dark energy (68 %), is not covered by this theory [21]. Dark matter is a form of matter that does not emit or absorb electromagnetic radiation and interacts only through gravity and, possibly, weak forces. So far, it has only been observed through its gravitational effects. The largest contribution, dark energy, permeates all space and accelerates the expansion of the universe, however, the mechanism behind this is still unknown. In addition to these phenomena, the SM is unable to explain the large discrepancy between the strength of the weak force and gravity, known as the hierarchy problem [22–25]. The Higgs boson mass of 125 GeV is close to the weak scale, while one would expect large loop corrections to the Higgs boson mass that drive it to the Planck scale of about 10^{19} GeV, the energy scale of quantum gravity. It is possible to achieve the observed mass by finely tuning the parameters of the SM, although it is likely more plausible that some mechanism suppresses the loop corrections.

Theories beyond the SM, such as supersymmetric theories, address many of these issues. Supersymmetry (SUSY) is a symmetry extending the Poincaré group of space-time symmetries that relates fermions and bosons [26–34]. Each SM fermion (boson) in a supersymmetric theory is associated with a SUSY boson (fermion) with the same quantum numbers except for the spin, which differs by one half. The loop corrections to the Higgs boson mass have opposite signs for fermions and bosons and, therefore, are cancelled with supersymmetry [35–40]. This naturally solves the hierarchy problem. Despite extensive searches for SUSY particles by experiments at the Sp$\bar{\text{p}}$S [41,42], HERA [43,44], LEP2 [45–49], Tevatron [50–53], and LHC [54,55] colliders, no such particles have been found and only exclusion limits on the masses of SUSY particles could be set.

Nevertheless, finding theories beyond what the SM cannot explain are crucial to gain a deeper understanding of quantum physics. Due to its large mass and the possibility to couple to unobserved particles beyond the SM, the top quark plays an important role.

2.2. The top quark

The top quark with a mass of 173.0 ± 0.4 GeV is the heaviest of all observed elementary particles [56]. After the discovery of the bottom quark, multiple searches for this last quark were performed. However, at that time no accelerator was powerful enough to produce this quark until the Tevatron was built at Fermilab. This proton-antiproton accelerator provided a high enough centre-of-mass energy of 1.8 TeV to produce top quarks leading to its discovery in 1995 by the CDF and DØ Collaborations [57,58].

Like all quarks, the top quark is subject to the strong force, however, it plays a key role within QCD. A unique behaviour of this theory is colour confinement. Colour confinement is a phenomenon, which describes that colour charged particles cannot be isolated. One consequence of gluons carrying a charge is that the strong force between two particles is constant regardless of their separation. Therefore, when a quark-antiquark pair is separated, the energy required to separate them increases up to a point, where creating another quark-antiquark pair becomes energetically favourable so that two quark-antiquark pairs exist rather than two isolated quarks. This process is also called hadronisation. Due to its high mass, the top quark has a mean lifetime of $5 \cdot 10^{-25}$ s (approximately a twentieth of the time scale for strong interactions) and, therefore, decays before it hadronises. This rapid decay process results in the top quark passing its spin information on to its decay products, which in turn provides an opportunity to study a “bare” quark.

2.2.1. Production of top quarks in a hadron collider

Until today, top quarks could only be produced in hadron colliders since the centre-of-mass energy of all operated e^+e^- colliders is too low. The primary top quark production in hadron colliders is the production of a top-antitop quark pair, see Figure 2.2. In proton-antiproton collisions at the Tevatron, the quark initiated production dominates, whereas the gluon initiated production dominates in proton-proton collisions of the LHC due to the lack of valence antiquarks in protons. Single top quarks can only be produced via the weak force corresponding to a lower cross-section. This is also the reason that the top quark was first discovered in $t\bar{t}$ production.

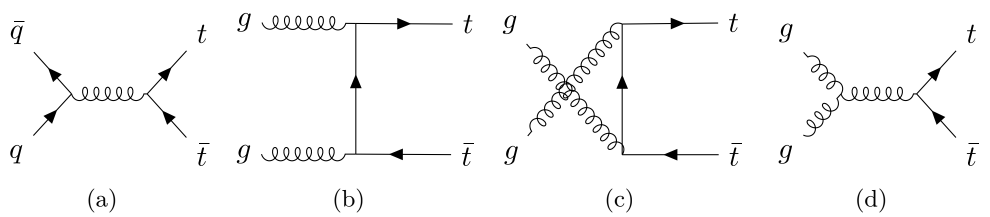


Figure 2.2.: The leading order Feynman diagrams for top-quark pair production at a hadron collider via (a) quark initiated production or (b)-(d) gluon initiated production.

2.2.2. Decay channels of the top quark

The top quark decays almost exclusively with a branching ratio (BR) of 99.8 % into a bottom quark and a W boson and, therefore, the top quark decay is categorised by the decay type of the W boson. The W can decay leptonically into $l\nu_l$ (BR 10.8 % for each generation) and hadronically into $q\bar{q}$ (BR 67.6 %). For top-quark pairs, the decay can be classified as *all hadronic*, where both W bosons decay hadronically, as *dilepton*, where both W bosons decay leptonically, and as *single-lepton*, where one of the W bosons decays hadronically and the other leptonically. Due to its highest BR, the all hadronic channel provides the highest event yield, but comes with the caveat of large background contributions. The dilepton channel gives a very clean signature for identification and reconstruction but has a much lower yield. The single-lepton channel combines both advantages of the other two channels, a high event yield with one hadronic decay and a distinct signature due to the leptonic decay. With increasing centre-of-mass energy, the $t\bar{t}$ cross-section increases faster than its corresponding backgrounds and, therefore, the LHC is also called a “top-quark factory”.

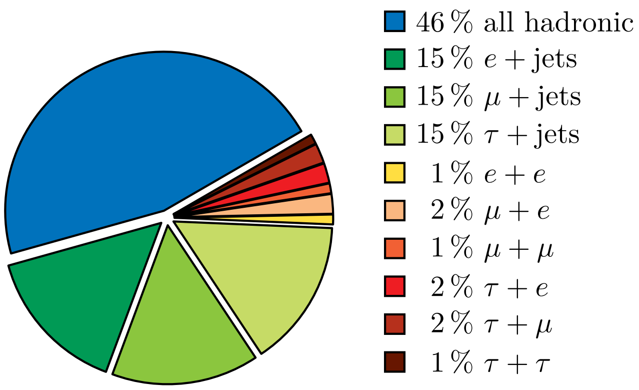


Figure 2.3.: The decay channels of a top-quark pair.

2.3. The Higgs boson

The Higgs boson plays a special role in the SM. Adding a mass term for a gauge boson breaks the gauge invariance of the Lagrangian. However, the gauge bosons of the weak force, the W and Z boson, have mass. To solve this issue the Brout-Englert-Higgs mechanism was proposed, including a new particle in the SM, the Higgs boson.

2.3.1. The Brout-Englert-Higgs mechanism

In 1964, three different groups independently proposed a theory to explain the masses of the W and Z gauge bosons, now termed the Brout-Englert-Higgs mechanism [5–8]. This mechanism is based on spontaneous symmetry breaking and introduces a complex scalar field, namely the Higgs field, with a potential in the shape of a "Mexican hat". This potential is symmetric with respect to a rotation around the centre axis. For a vacuum expectation value (VEV) of the Higgs boson $v = 0$ the symmetry is conserved. However, for a required non-zero VEV in one of the minima of the potential, the symmetry is spontaneously broken. In the SM, the VEV of the Higgs field depends on the mass of the W boson, the reduced Fermi constant, and the weak isospin coupling and has a value of 246 GeV [56].

The gauge bosons gain mass by interacting with this complex scalar field. The Higgs boson, an excitation of this field, has no electric or colour charge and, therefore, does not interact directly with photons or gluons leaving them massless. However, the Higgs boson can interact with W and Z bosons and gives them mass without the need of adding a mass term to these gauge bosons, conserving the gauge invariance of the Lagrangian.

Fermions gain mass via the so called Yukawa coupling (see Yukawa matrix Y_{ij} in Equation 2.1). This coupling describes the strength between the Higgs field and massless quark and lepton fields. It is assumed that the coupling strength is proportional to the corresponding fermion mass, m_f, leading to a top quark Yukawa coupling, y_t, of order unity [56]

$$y_t = \sqrt{2}\frac{m_f}{v} \approx \sqrt{2}\frac{173\,\mathrm{GeV}}{246\,\mathrm{GeV}} \approx 1. \tag{2.2}$$

Measuring this coupling strength is an important test for the SM and could give hints for possible new physics beyond the SM [59]. A process where this coupling can be measured directly is the production of a Higgs boson in association with a pair of top quarks, $t\bar{t}H$ [60–63].

Almost 50 years after its prediction [5–8], the Higgs boson was discovered with a mass of 125 GeV at CERN by the ATLAS and CMS collaborations in 2012 [1, 2].

An interesting consequence of a Higgs boson with a mass smaller than 130 GeV is that it can lead to an unstable vacuum in the universe. A loophole to avoid this vacuum decay is when the vacuum lifetime exceeds the age of the universe [64].

2.3.2. Production of a Higgs boson in proton-proton collisions

The different production mechanisms of a Higgs boson in proton-proton collisions as a function of the centre-of-mass energy are shown in Figure 2.4. For all centre-of-mass energies gluon-gluon fusion is the dominant production mechanism. It is important to note that for a higher centre-of-mass energy the $pp \rightarrow t\bar{t}H$ cross-section increases more than the other production mechanisms giving this process a higher signal over background ratio. The Feynman diagrams of these Higgs boson production mechanisms are presented in Figure 2.5.

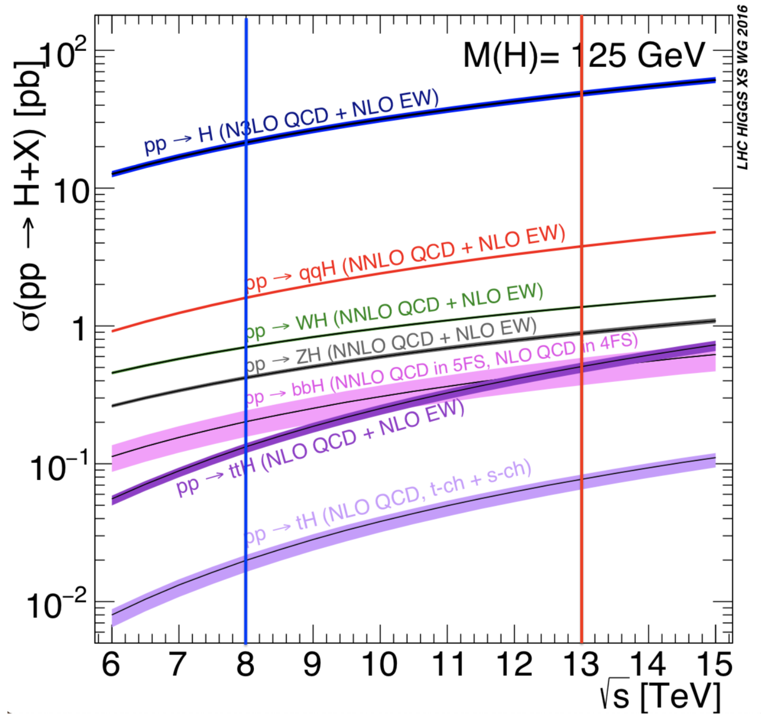


Figure 2.4.: The inclusive cross-section of the Higgs boson production mechanisms as a function of the centre-of-mass energy for proton-proton collisions. The process $t\bar{t}H$ has a cross-section of $\sigma_{\sqrt{s}=8\,\mathrm{TeV}}(t\bar{t}H) = 0.13\,\mathrm{pb}$ and $\sigma_{\sqrt{s}=13\,\mathrm{TeV}}(t\bar{t}H) = 0.50\,\mathrm{pb}$ [65].

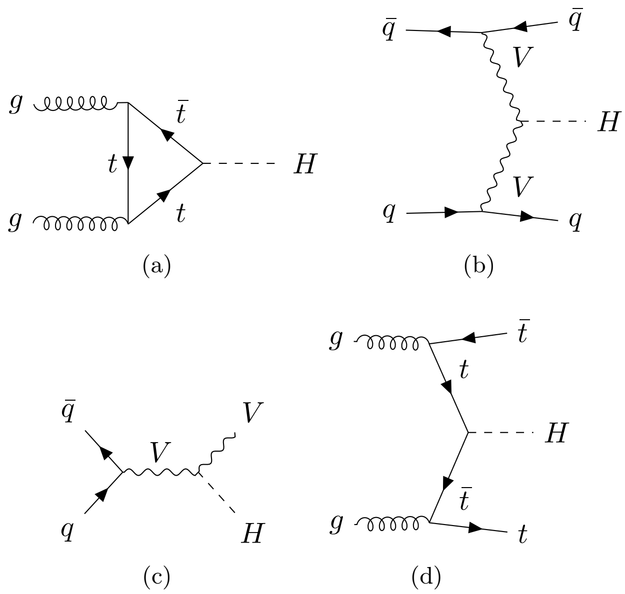


Figure 2.5.: Feynman diagrams of the main Higgs boson production mechanisms at the LHC. (a) gluon-gluon fusion, (b) vector boson fusion ($V = Z, W$), (c) associated production with a vector boson, (d) associated production with a top-quark pair.

2.3.3. Decay channels of the Higgs boson

The decay channels of a Higgs boson are shown in Figure 2.6. The dominant channel $H \to b\bar{b}$ with a BR of 57 % gives a final state that is difficult to distinguish from the large $t\bar{t}$ background of a hadron collider. Therefore, the Higgs boson was first discovered in the decay channels $H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ell$. The final states $b\bar{b}$, W^+W^-, $\tau^+\tau^-$, $c\bar{c}$, and ZZ^* originate from a direct decay of the Higgs boson, while the other final states are reached via loop contributions. The Feynman diagrams of these decays are depicted in Figure 2.7.

2.3.4. Extraction of the top quark Yukawa coupling

Higgs boson decays into massless particles (Figure 2.7) involve virtual loop contributions. The contribution from a top quark is favoured over a W boson and measuring these decay rates allows the top quark Yukawa coupling to be extracted. However, this is only an indirect measurement under the assumption that no additional particles outside of the SM contribute to this loop. In addition, the gluon-gluon fusion process (Figure 2.5a) is another indirect measurement of the top quark Yukawa coupling under the same assumption.

A direct measurement has the advantage of being model independent, thus no assump-

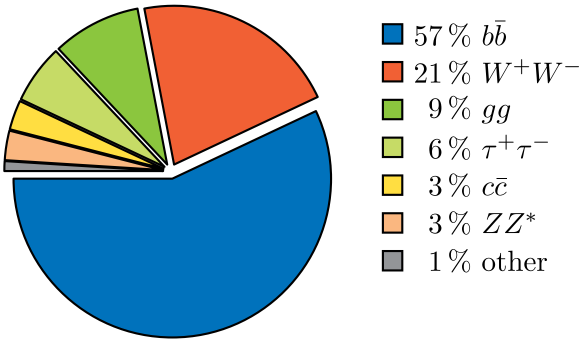


Figure 2.6.: The decay channels of a Higgs boson.

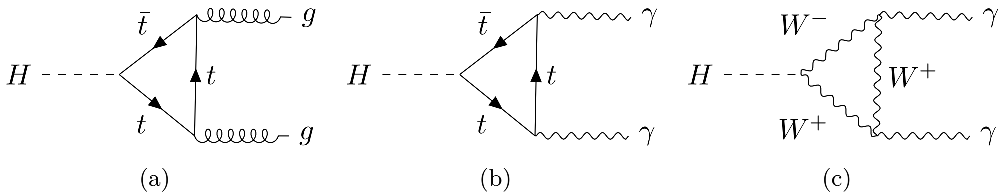


Figure 2.7.: Feynman diagrams of decay channels of the Higgs boson involving virtual loop contributions.

tions on additional particles outside of the SM need to be made. Any deviation from the SM prediction would give a clear sign for physics beyond the SM. The most favourable production mode, which has a direct sensitivity to this coupling, is the production of a Higgs boson in association with a top-quark pair, $t\bar{t}H$ (Figure 2.5d), and will be discussed in Chapters 5 and 6 .

CHAPTER 3

The ATLAS experiment at CERN

3.1. CERN

The European Organization for Nuclear Research with the acronym CERN, derived from the French name *Conseil Européen pour la Recherche Nucléaire*, is the largest particle physics laboratory in the world [66]. After the Second World War, European science lost its leading role, which led to the idea to create a European atomic physics laboratory. Louis de Broglie was one of the first supporters of this vision and with the help of Nobel laureate Isodor Rabi a UNESCO resolution concerning this idea was formed. On the 29$^{\text{th}}$ September 1954 the convention establishing CERN was ratified by twelve countries in Western Europe.

Over the course of the next decades, CERN played a significant role in particle physics leading to the award of several Nobel Prizes. New scientific discoveries go hand in hand with technological advances such as the invention of the World Wide Web. Today, CERN has 23 member states and over 12 000 scientists from all over the world. To this end, larger particle accelerators were built over the last 60 years, beginning with the Synchrocyclotron and the Proton Synchrotron, which consequently lead to the Large Hadron Collider.

3.2. The Large Hadron Collider

The Large Hadron Collider (LHC) is the largest and most powerful particle collider in the world and was constructed at CERN between 1998 and 2008 [67]. The 27 km long circular collider is located approx. 100 m beneath the France–Switzerland border near Geneva.

For proton-proton collisions the designed centre-of-mass energy is $\sqrt{s} = 14\,\text{TeV}$, while it is currently operated with 13 TeV. To reach these high energies, the protons are accelerated with electrical fields by several pre-accelerators (see Figure 3.1).

All protons originate from a hydrogen gas cylinder where hydrogen atoms are stripped of their electrons leaving protons. First, the protons gain energy by a linear accelerator, the Linac 2, to an energy of 50 MeV. After this, the protons are injected into circular accelerators, starting with the Booster. A circular accelerator uses magnets to bend particles with the Lorentz force on a circle so that particles can be accelerated multiple times from the same electrical field. The Booster consists of four superimposed synchrotron rings and boosts protons to 1.4 GeV. This machine splits the beam into bunches and increases the beam intensity. Next in the chain is the PS, one of the oldest accelerators at CERN, which is still in use today. The accelerator reaches proton energies of 25 GeV and injects the beam into the SPS. This synchrotron is the largest pre-accelerator reaching an energy of 450 GeV and prepares the proton bunches for the LHC. The beam is injected

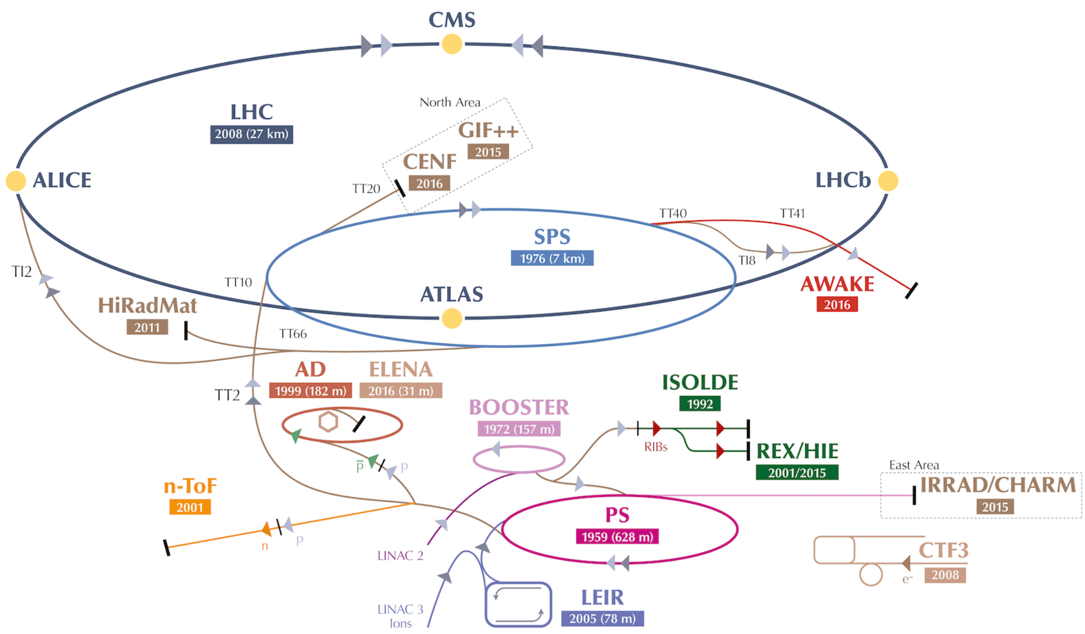


Figure 3.1.: The CERN accelerator complex. Protons are accelerated by traversing the Linear accelerator 2 (Linac 2), the Proton Synchrotron Booster (Booster), the Proton Synchrotron (PS), the Super Proton Synchrotron (SPS), and finally are injected into the Large Hadron Collider (LHC). The accelerating process for ions starts with the Linear accelerator 3 (Linac 3) and continues with the Low Energy Ion Ring (LEIR), PS, SPS, and finally LHC. In addition to the LHC accelerator chain, there are further experimental sites, e.g. the Antiproton Decelerator (AD) that slows down antiprotons so that they can be combined with positrons to form neutral antihydrogen. © CERN

clockwise and counterclockwise in two separated beam pipes of the LHC. The protons are further accelerated to 6.5 TeV and can be brought to collision at four interaction points. The strong magnetic field needed to contain particles with such high energies on their design orbit can only be generated by superconducting magnets. This is a phenomenon where the electrical resistance drops to exactly zero. Therefore, electrical currents can flow without emitting heat and very high currents can be reached, both of which are crucial properties for generating strong magnetic fields. Indeed, superconductivity only appears below a certain temperature, thus, the magnets of the LHC need to be cooled to a temperature of 1.9 K using liquid helium.

In addition to proton-proton collisions, the LHC can also collide (heavy-)ions, such as lead nuclei. The starting point for these is Linac 3. Prior to acceleration, atoms are evaporated in an oven, which removes some of their electrons, the rest of which are removed during acceleration. In a second step, atoms are transformed into “bare” nuclei, which are easier to accelerate than whole ions. After reaching an energy of 4.2 MeV, particles are injected into the LEIR accelerator where the beam is split into shorter bunches and further accelerated to 72 MeV. After this, the beam is passed to the PS, where the acceleration chain is the same as for protons.

When particles collide at one of the four interaction points, they transfer their energy into mass, $E = mc^2$, and create new particles. These particles can be measured by large particle detectors located at each of the four interaction points of the LHC, where the ATLAS [68], CMS [69], LHCb [70], and ALICE [71] detectors operate. In addition, three smaller detectors TOTEM, MoEDAL, and LHCf share the interaction point with one of the larger experiments (or are located close to it) and perform specialised research such as cross-section measurements or the search of magnetic monopoles.

The ATLAS and CMS experiments are multi-purpose particle detectors with a near 4π coverage in solid angle. Both experiments discovered the Higgs boson in 2012 leading to the 2013 Nobel Price for Physics, which was jointly awarded to Peter Higgs and François Englert for their theoretical work [1,2]. The main difference between the two detectors is that CMS uses all-silicon detectors for its inner tracker and is more compact than ATLAS. To compensate for the smaller dimensions, the CMS magnetic system can create a higher magnetic field and, therefore, increase the curvature of the charged particles to achieve a similar momentum resolution. The ATLAS experiment will be described in detail in Section 3.3. Of the remaining experiments, LHCb specialises in the measurement of b quarks to understand the matter-antimatter asymmetry in the universe and ALICE specialises in the study of heavy-ion collisions. When heavy-ions collide in the LHC, a quark gluon plasma is formed. This is a state of matter where quarks can move freely around what is hypothesised to have existed a few milliseconds after the Big Bang, and can also be found in neutron stars. At lower energies, quarks are confined in groups of at least two due to their interaction via the strong force.

After an extensive test period, first collisions (with low energies) were recorded in 2009. The first data taking period (Run 1) started in 2011 with a centre-of-mass energy of $\sqrt{s} = 7$ TeV, which increased to 8 TeV in 2012. After this initial run, the LHC entered Long Shutdown 1 (LS1) where further tests and upgrades where performed. Due to a better understanding of the LHC, collisions at 13 TeV were reached for Run 2, which lasted from 2015 until 2018. Run 3 is planned to start in 2021, which will explore the full potential of the LHC with $\sqrt{s} = 14\,\text{TeV}$.

3.3. The ATLAS experiment

3.3.1. Overview of the ATLAS detector

The ATLAS detector (A Toroidal LHC ApparatuS) [68] is one of the four main experiments at the LHC. It is a multi-purpose detector with a forward-backward-symmetric geometry. For proton-proton collisions an instantaneous luminosity of up to $\mathcal{L}_{\text{max}} = 2.1 \times 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$ was achieved during Run 2 [72]. This leads to a high particle density where a single bunch crossing produces several separate collisions (on average $\langle\mu\rangle = 34$ for Run 2). These so-called pile-up events interact with the same detector components at roughly the same time and can spoil the event identification, and, therefore, need to be taken into account during data analysis. Integrating the instantaneous luminosity with respect to time gives the integrated luminosity L [73–75]. This is an important quantity for particle colliders, as the higher the integrated luminosity, the more data is available to analyse. An integrated luminosity of $L = 139\,\text{fb}^{-1}$ was recorded during Run 2 that is available for analysis.

To maximise the detector area around the interaction point, ATLAS uses components that are orientated cylindrically around the beam axis in the central part and components that are positioned perpendicular to the beam pipe in the two end-caps. The detector consists of three main components and a magnetic system, see Figure 3.2. The innermost part is called the inner detector and uses pixels, strips, and TRT straws to measure tracks of charged particles. This inner detector is surrounded by the calorimeter system, an electromagnetic and hadronic calorimeter used to stop particles and measure their energy deposition. The outermost part of the detector incorporates a system of muon spectrometers used to detect muons that can pass through the calorimeter system. The interaction of different particles with these detector components is also visualised in Figure 3.3. Before describing the detector in more detail, it is important to introduce the nomenclature used to define components of the ATLAS detector. The nominal interaction point is defined as the origin of a right-handed coordinate system with the z-axis along the proton beam. The positive x-axis is directed from the interaction point towards the centre of the LHC ring, while the positive y-axis is directed upwards. Cylindrical coordinates (R, ϕ, z) are used in the transverse plane, ϕ being the azimuthal angle around the beam pipe. The pseudorapidity of a particle, η, is defined in terms of the polar angle, θ, as $\eta = -\ln(\tan(\theta/2))$, where the difference in pseudorapidity is invariant under boosts in the z-direction. The transverse momentum, $p_{\text{T}} = \sqrt{p_x^2 + p_y^2}$, and the transverse energy, $E_T = E/(\cosh(\eta))$, of a particle are defined in the x–y plane. Distance in the pseudorapidity-azimuthal angle space is defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

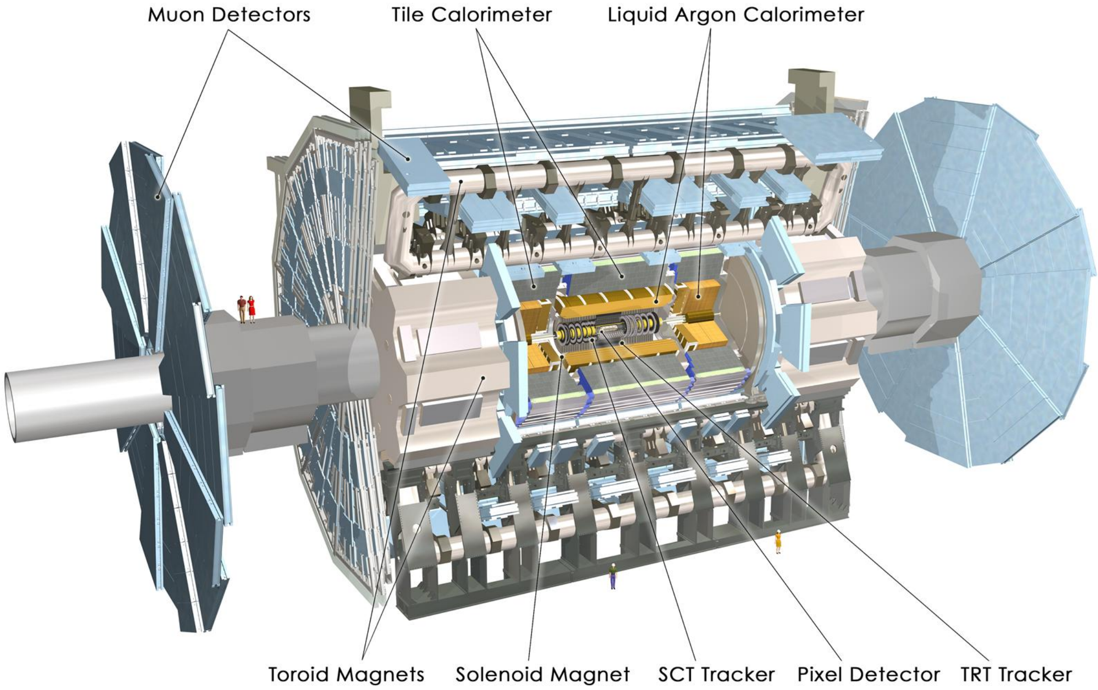


Figure 3.2.: Cutaway rendering of the ATLAS detector. © CERN

3.3.2. The magnet system

The ATLAS detector is built around a large superconducting magnetic system. It consists of one central solenoid close to the beam pipe, a large barrel toroid, and two toroids in its end-caps, see Figure 3.2. The purpose of this magnetic system is to create a strong magnetic field that deflects charged particles when passing through the detector, which vary in orientation depending on the sign of the particles' electric charge. By measuring the curvature of a particle track the momentum of the particle can be determined, whereby particles with a greater momentum experience a smaller deflection.

The central solenoid

The central solenoid encloses the inner detector and generates a magnetic field with a range of 0.9–2.0 T over pseudorapidities of $|\eta| < 2.5$ to bend charged particles in the R–ϕ plane and stores a magnetic energy of 40 MJ. The solenoid layout is optimised to keep the material thickness as low as possible to minimise its impact on the calorimeter performance. A total contribution of approximately 0.66 radiation lengths (X_0) at normal incidence could be achieved.

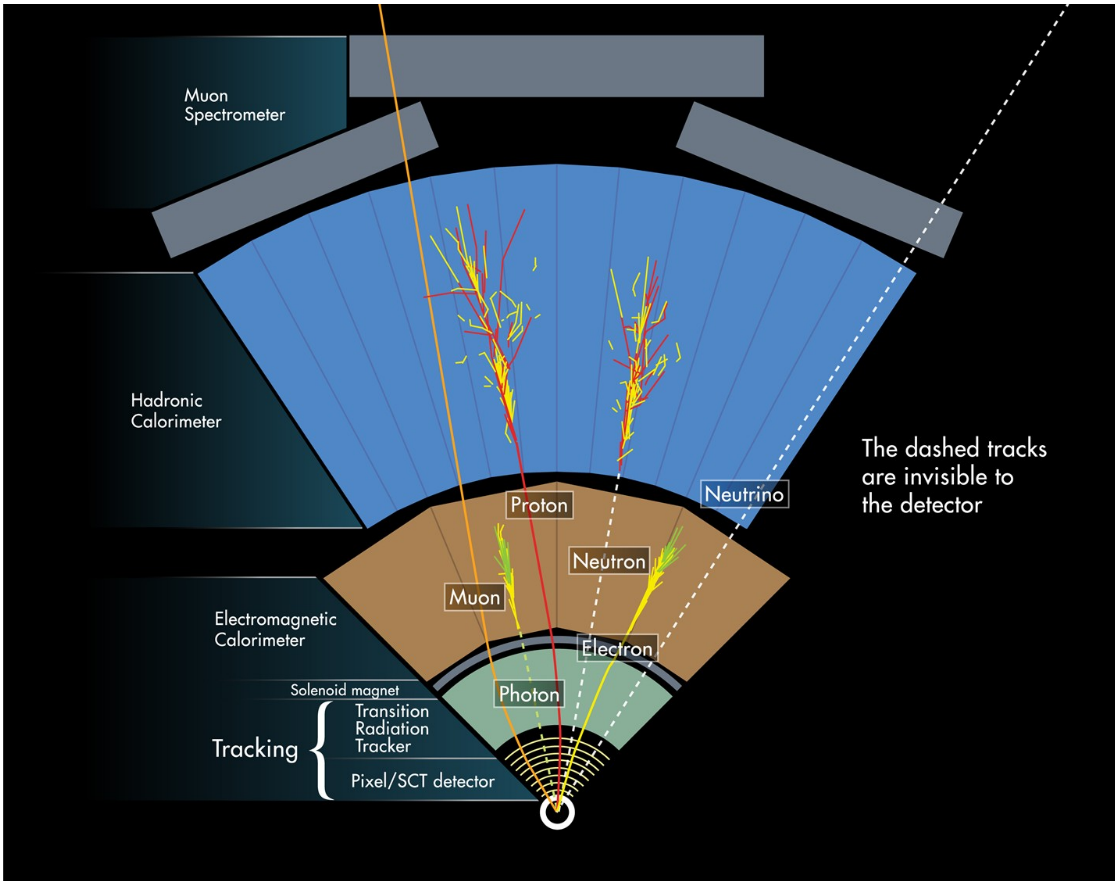


Figure 3.3.: Illustration of particle detection with the ATLAS detector. Particles emerge from the interaction point inside the beam pipe (white circle). Electrons are measured by the inner tracking detectors and are stopped by the electromagnetic calorimeter, where their energy deposition is determined. Photons deposit their energy only in the electromagnetic calorimeter. The energy of protons and neutrons is determined by the hadronic calorimeter. In addition, protons can also be detected by the tracking detector. Muons are visible to the tracking detector, and are identified by the muon spectrometers. Neutrinos pass through the detector without interaction, and can only be measured indirectly from the transverse momenta of visible particles. The curvature of different particles originates from applied magnetic fields in the inner detector and the muon spectrometer. © CERN

The barrel toroid

The barrel toroid consists of eight coils each individually encased in stainless-steel vacuum vessels. This is the largest structure of the ATLAS detector and can store an energy of 1.1 GJ. It generates a magnetic field with a range of 0.2–2.5 T over an axial length of

25 m and covers pseudorapidities of $|\eta| < 1.4$.

The end-cap toroids

The two end-cap toroids improve the bending power of the magnetic field in the end-cap regions of the muon spectrometer. Each toroid consists of eight coils and can store an energy up to 250 MJ. Both toroids generate a magnetic field with a range of 0.2–3.5 T over pseudorapidities of $1.6 < |\eta| < 2.7$.

In the transition region $1.4 < |\eta| < 1.6$ between the barrel and end-cap toroids, magnetic deflection is provided by a combination of the barrel and end-cap fields. With this, a field configuration mostly perpendicular to the particle trajectories can be achieved.

3.3.3. The inner detector

The inner detector tracks charged particles to measure their momenta and determines the primary vertex by extrapolating reconstructed tracks to the position of the main interaction point. The inner detector is located closest to the beam pipe and must cope with a very high particle density and harsh background. Because of this challenging environment, different detector technologies with fine granularity are used. These technologies include a silicon pixel detector, a silicon microstrip detector (SCT), and a transition radiation tracker (TRT), all displayed in Figure 3.4. Together the components cover pseudorapidities of $|\eta| < 2.5$. The combined effect of the strength of the central solenoids' magnetic field and the tracking precision of the inner detector achieves a momentum resolution of $0.05\,\%\ p_{\mathrm{T}}[\mathrm{GeV}] \oplus 1\,\%$.

The silicon pixel detector and IBL

To achieve the highest granularity closest to the collision region, silicon pixel detectors are used. These pixel detectors are arranged in three barrel layers made up of 1456 pixel modules and three end-cap disk layers with 288 modules. Most pixel modules have approximately 46 000 pixels with a size of $50\,\mu\mathrm{m} \times 400\,\mu\mathrm{m}$. During LS1 an additional layer, the Insertable B-Layer (IBL), was installed closest to the beam pipe [76,77]. This fourth layer consists of 448/224 single/double chip modules consisting of approximately 27 000 smaller pixels ($50\,\mu\mathrm{m} \times 250\,\mu\mathrm{m}$) and compensates radiation damage and degradation of installed modules of the other layers. In addition, it improves vertex detection and b-tagging [78]. With approximately 92 million readout channels, a precise measurement of 3D space-points is possible.

The silicon microstrip detector

The SCT is built of eight strip layers to measure four space-points for each track. This detector uses small-angle stereo strips to measure both coordinates. In the barrel, one set of strips in each layer is aligned parallel to the beam direction, measuring R–ϕ. In the end-caps, the SCT uses one set of radially arranged strips and one set consisting of

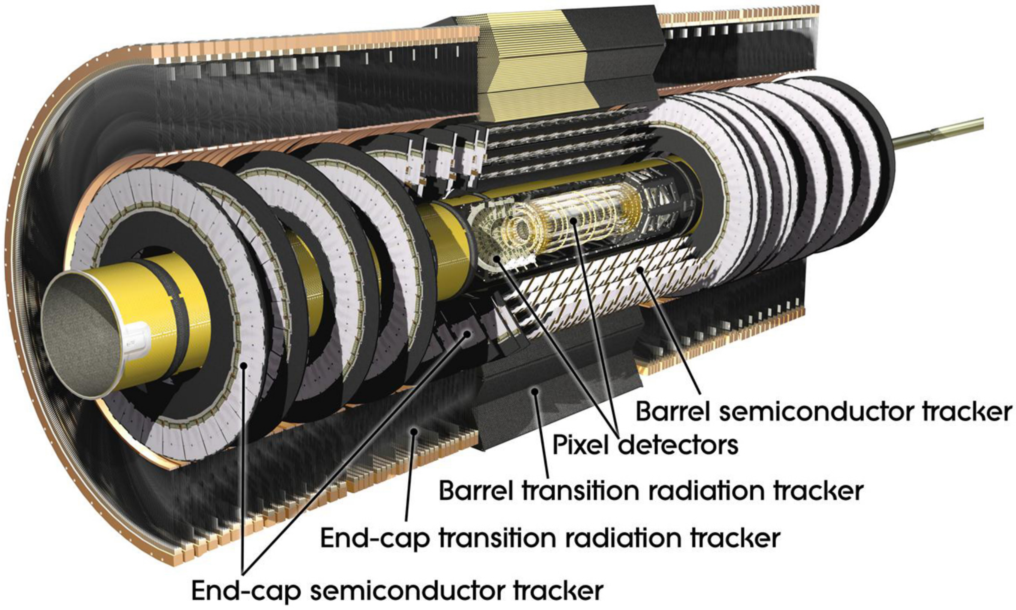


Figure 3.4.: Overview of the inner detector (without the IBL). © CERN

two stereo strips glued together at an angle of 40 mrad. With approximately 6.3 million total readout channels the SCT further improves the momentum resolution and vertex position.

The transition radiation tracker

The TRT has approximately 351 000 readout channels and uses single wire drift tubes called straw tubes, which provide a large number of hits per track in a range of $|\eta| < 2.0$. These straws can only provide R–ϕ information and are filled with a gas (Xe + CO_2 + O_2 mixture), such that any charged particle that traverses the tube ionises the surrounding gas. Resulting ions and electrons are accelerated by the electric field inside the tube and cause additional charged particles. This effect is repeated until the cascade of particles is collected on the wire, at which the number of particles is proportional to the energy of the primary particle. In addition, charged particles passing through different materials between the straws emit transition radiation photons. This effect is strongest for electrons, which means it can be used for particle identification.

A combination of precision trackers close to the beam pipe with the surrounding TRT gives very robust pattern recognition and a high precision momentum measurement. The precision of the ϕ-coordinate measurement is especially important to precisely measure

p_T, since the Lorentz force of the magnetic field acts in the direction of this coordinate.

3.3.4. The calorimeter system

The ATLAS calorimeter system is split into electromagnetic and hadronic calorimeters, see Figure 3.5. Each of these contains barrel, end-cap, and forward components, and together the system covers the range $|\eta| < 4.9$. The calorimeters are all sampling calorimeters, in which the material that produces the particle shower is separate from the material that measures the shower energy. With a fine granularity for the electromagnetic calorimeter and a coarser granularity for the hadronic calorimeter, the calorimeters can measure electrons and photons very precisely and have a good resolution for jet reconstruction. A jet consists of particles that form a narrow cone and are produced by the hadronisation of a quark or gluon.

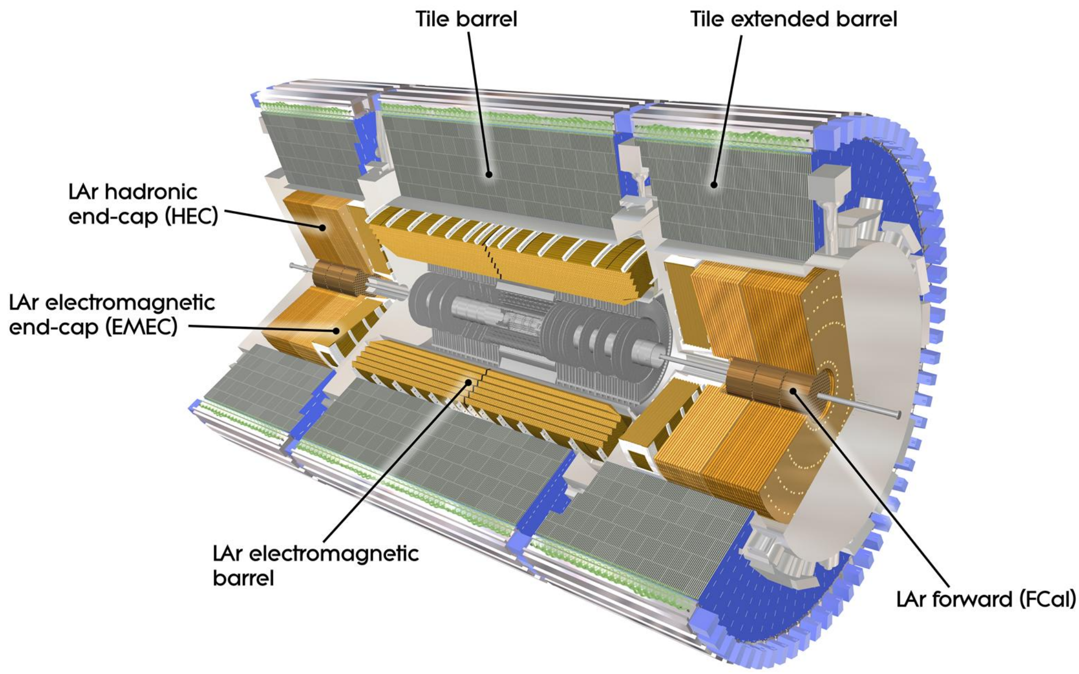


Figure 3.5.: Overview of the calorimeter system. © CERN

When electromagnetically- and strongly-interacting particles pass through the calorimeters they produce cascades of secondary particles called showers. Each shower produces many (up to approximately 10^9) low energetic particles, which are ultimately absorbed in the calorimeter. In order to absorb all these particles, calorimeters must be deep, otherwise charged particles, mostly hadrons, will escape the calorimeter and can be wrongly identified as muons from the muon spectrometer. This process is called punch-through and it is one effect that degrades jet energy scale and resolution. The total thickness of the electromagnetic calorimeter is greater than 22 (24) radiation lengths in the barrel (end-caps), where the radiation length is related to the energy loss of high energetic ($\gtrsim 10\,\mathrm{MeV}$) electrons by bremsstrahlung and photons by e^+e^- pair production within a material. The depth of the hadronic calorimeter is measured in interaction lengths λ, defined as the mean distance travelled by a hadronic particle before undergoing an inelastic nuclear interaction. Both barrel and end-cap calorimeters are approximately ten interaction lengths deep. Together with the large η-coverage, this thickness also ensures a good missing transverse momentum measurement.

The electromagnetic calorimeter

The electromagnetic calorimeter is divided into a barrel part ($|\eta| < 1.475$) and two end-caps ($1.375 < |\eta| < 3.2$). The end-cap calorimeters are further subdivided into outer wheels covering the range of $1.375 < |\eta| < 2.5$, and inner wheels covering the range of $2.5 < |\eta| < 3.2$. The electromagnetic calorimeter is a liquid-argon (LAr)-based detector with accordion-shaped kapton electrodes and lead absorber plates. This geometry provides complete ϕ symmetry without azimuthal cracks. Over the region $|\eta| < 2.5$ devoted to precision physics, the electromagnetic calorimeter is segmented into three sections in depth. Additionally, in the region $|\eta| < 1.8$ the energy lost by electrons and photons interacting with the inner detector is corrected by a presampler detector. For the end-cap inner wheel, the calorimeter is segmented into two sections in depth and has a coarser lateral granularity. The electromagnetic calorimeter has an energy resolution of $\sigma_E/E = 10\,\%/\sqrt{\mathrm{E}[\mathrm{GeV}]} \oplus 0.7\,\%$.

The hadronic calorimeters

The hadronic tile calorimeter is mounted directly outside the electromagnetic calorimeter. Its barrel covers the region $|\eta| < 1.0$, and it has two extended barrels to cover the range $0.8 < |\eta| < 1.7$. It uses steel as the absorber and scintillating tiles as the active material. The calorimeter is segmented in three layers.

The hadronic end-cap calorimeter (HEC) consists of two independent wheels per end-cap, located directly behind the end-cap electromagnetic calorimeter and shares the same LAr cryostats. The angular coverage of the calorimeter is $1.5 < |\eta| < 3.2$. Each wheel is divided into two segments in depth, for a total of four layers per end-cap. This construction is built from copper plates, which are interleaved with LAr gaps, where LAr is the active medium. The HEC has a resolution of $\sigma_E/E = 50\,\%/\sqrt{\mathrm{E}[\mathrm{GeV}]} \oplus 3\,\%$.

The forward calorimeter (FCal) is integrated into the end-cap cryostats and covers a range of $3.1 < |\eta| < 4.9$. It consists of three modules in each end-cap: the first is made of copper and optimised for electromagnetic measurements, while the other two are made of tungsten and predominantly measure the energy of hadronic interactions. LAr is used as the active medium in all three modules. The FCal has a resolution of $\sigma_E/E = 100\,\%/\sqrt{\mathrm{E}[\mathrm{GeV}]} \oplus 10\,\%$.

3.3.5. The muon spectrometer

The muon spectrometer is the outermost part of the ATLAS detector and is displayed in Figure 3.6. It is designed to detect charged particles exiting outside of the calorimeter system and to measure their momenta in the pseudorapidity range $|\eta| < 2.7$. This applies nearly exclusively to muons, which deposit little energy (approximately 3 GeV) in the ATLAS calorimeters.

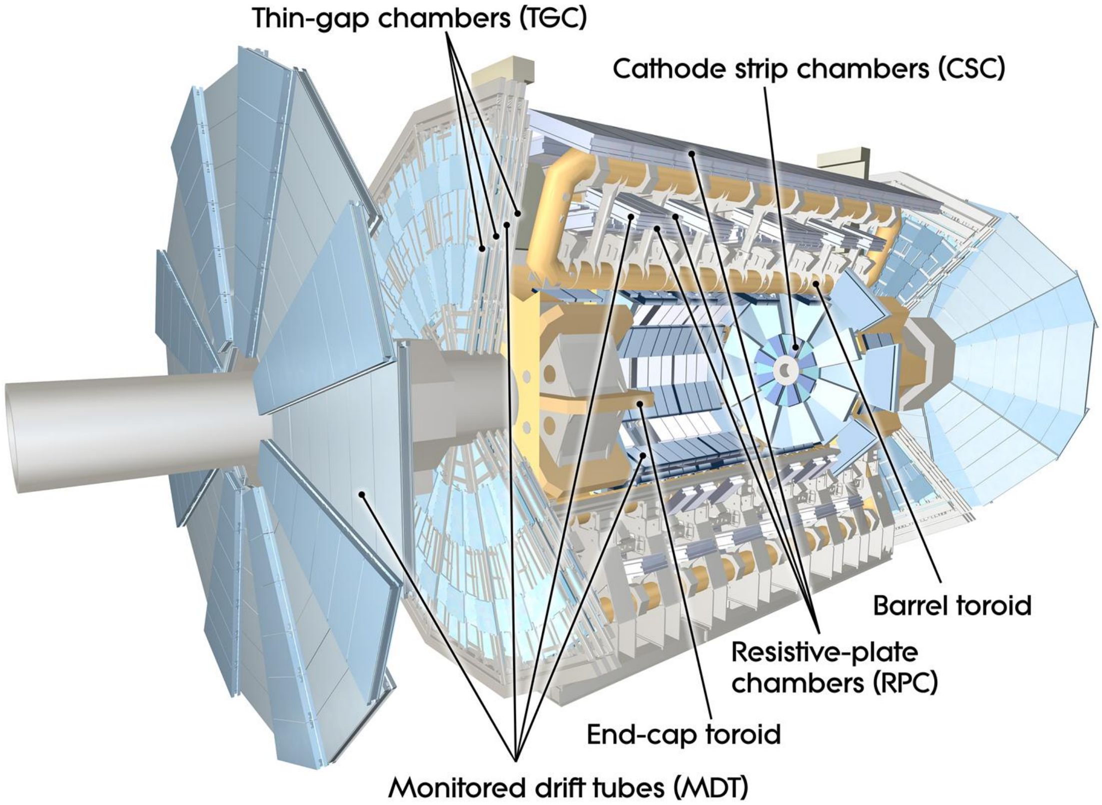


Figure 3.6.: Overview of the muon spectrometer. © CERN

The spectrometer measurements are based on the magnetic deflection of muons by the barrel and end-cap toroids, and are instrumented with separate trigger and high-precision tracking chambers. With this system the transverse momentum of muons with $p_{\mathrm{T}} \approx 1\,\mathrm{TeV}$ can be measured with a precision of about 10 %. The precision measurement of the muon tracks is made in the R–z projection, the direction parallel to the bending direction of the magnetic field. In the barrel region, muons are detected in chambers stacked in three cylindrical layers around the beam axis. In the transition and end-cap regions, the chambers are arranged in planes perpendicular to the beam, also in three layers.

The monitored drift tubes

Over most of the η-range, the precision measurement of muon track coordinates is conducted by monitored drift tubes (MDTs). These are proportional chambers with six layers of single wire drift tubes per chamber and operate with an Ar + CO_2 gas mixture. Each tube measures the distance to the wire at which the particle crosses the tube. This radius describes a drift circle, which is extracted from a precise time measurement of the signal. Multiple drift circles are required to reconstruct a track.

The cathode strip chambers

To withstand the high particle density and harsh background conditions at large pseudorapidities of $2 < |\eta| < 2.7$, cathode strip chambers (CSCs) with higher granularity are used. CSCs are multi-wire proportional chambers where the wires are oriented in the radial direction. The cathodes are segmented into strips perpendicular to the wires. The position of the track is obtained by interpolation between the charges induced on neighboring cathode strips.

The resistive plate chambers

The MDT readout is too slow to resolve 25 ns bunch crossings, therefore another system is required to trigger muons with $|\eta| < 2.4$. For this purpose, resistive plate chambers (RPCs) are used in the barrel and thin gap chambers (TGCs) in the end-caps. An RPC is a gaseous detector with two resistive plates with a high volume resistivity. The electric field between the plates allows avalanches to form along the ionising tracks towards the anode. This provides a very good time resolution of 1.5 ns. TGCs are multi-wire proportional chambers designed to achieve a very good time resolution. The trigger chambers for the muon spectrometer provide bunch-crossing identification, transmit input to the trigger system, and measure the muon coordinate in the (non-bending) ϕ direction.

3.3.6. Forward detectors

To provide good coverage in the very forward region, four additional smaller sets of detectors are used. They are all located close to the beam pipe outside of the ATLAS cavern. These detector systems are briefly described in the following and are ordered according to their distance from the interaction point.

LUCID (LUminosity measurement using Cherenkov Integrating Detector) [79]. This detector is the main online relative luminosity monitor in ATLAS and measures inelastic proton-proton scattering in the forward region ($|\eta| \approx 5.6$) by detecting Cherenkov light of charged particles with photomultiplier tubes. LUCID is located ± 17 m from the interaction point.

Zero-Degree Calorimeter (ZDC) [80]. The detector's main task is to detect forward neutrons and photons with $|\eta| > 8.3$ in both proton-proton and heavy-ion collisions. For this purpose, layers of alternating quartz rods and tungsten plates are used. The ZDC plays a crucial role in determining the centrality of heavy-ion collisions, which is strongly correlated with the number of very forward neutrons. The detector is located at ± 140 m from the interaction point where the straight-section of the beam-pipe divides back into two independent beam-pipes.

ATLAS Forward Proton (AFP) [81]. This project consists of four detectors located at ± 205 m and ± 217 m from the interaction point and measures the momentum of protons originating from elastic and diffractive scattering. This is achieved by installing the detectors inside Roman Pots, which are special devices that allow the detectors to be moved close to the beam.

ALFA (Absolute Luminosity For ATLAS) [82]. This module is located at ± 240 m from the interaction point and covers the region $10.6 < |\eta| < 13.5$. To achieve this, scintillating fibre tracks inside Roman Pots can be moved to the beam as close as 1 mm and measure the absolute luminosity from elastic proton-proton scattering down to very small angle ($\theta_{\min} \sim$ µrad).

3.3.7. The trigger system

In Run 2, proton-proton collisions occur every 25 ns corresponding to a bunch crossing rate of 40 MHz. These collisions have an average pile-up of $\langle\mu\rangle = 34$, which leads to an interaction rate of over 1 GHz. Multiplying the rate by 1.5 MB, the average disc storage size of a single event, gives a data rate of about $2\,\mathrm{PB\,s^{-1}}$. With today's technology, it is impossible to save and store such a high data rate. The ATLAS trigger system is designed to reduce the bandwidth by selecting and storing events with interesting physics scenarios and discarding the remaining ones. It consists of a hardware Level-1 trigger and a software-based high level trigger [83].

The Level-1 trigger. This system uses custom electronics to determine Regions-of-Interest (RoIs) in the detector. RoIs only consider a subset of detectors with reduced granularity and precision. The trigger algorithm searches in these regions for high-transverse-momentum muons, electrons, photons, jets, and τ-leptons decaying into hadrons, as well as large missing transverse energy. The selection is made in less than 2.5 µs and reduces the event rate to 100 kHz.

The high level trigger (HLT). It uses the RoI information from the Level-1 trigger and performs complex selection algorithms using the full granularity detector information in either the RoI or the whole event. The HLT reduces the event rate to 1 kHz, which reduces the data rate to approximately $2\,\mathrm{GB\,s^{-1}}$, which is sufficiently low for storage. In a next step, this data can be analysed by members of the ATLAS collaboration.

3.4. The Worldwide LHC Computing Grid

The data recorded by ATLAS is about 15 PB per year. To analyse this data and search for potential new physics it needs to be further reconstructed and processed. In addition, computational intensive simulations describing different theories need to be performed. For this purpose, the Worldwide LHC Computing Grid (WLCG) was established. It consists of over 170 computing centres in 42 countries and splits the workload (and financial cost) of building, maintaining, and upgrading the infrastructure. It is structured in one Tier 0 computing centre located at CERN (with a mirrored back-up at the Wigner Research Centre for Physics in Hungary) performing the imminent event reconstruction, thirteen large Tier 1, and several smaller Tier 2 centres across the world. A dedicated software keeps track on where the data is stored and a batch system distributes the workload equally between the different computing centres to ensure a smooth computing environment for the different physics analyses.

3.5. Planned upgrades

Since the successful Run 2 ended in 2018, the LHC is currently in Long Shutdown 2 (LS2) to prepare for Run 3 with a centre-of-mass energy of $\sqrt{s} = 14\,\text{TeV}$ and an increased instantaneous luminosity. After this Long Shutdown 3 (LS3), a thorough upgrade will be performed to boost the instantaneous luminosity even further.

3.5.1. The High Luminosity Large Hadron Collider

The High Luminosity Large Hadron Collider (HL-LHC) is an upgrade of the current LHC to increase its instantaneous luminosity by a factor of approximately five with respect to the design value of LHC [84]. To achieve this, the proton bunches will be packed with even more protons and the beam will be focused even stronger at the collision points. For this, more powerful focusing magnets will be installed. In addition, crab cavities in front of the interaction points will provide a transverse deflection of the bunches to enlarge the overlap area of the two colliding bunches and, therefore, increasing the probability of interactions. Over the course of the HL-LHC programme the goal is to collect up to $4\,\text{ab}^{-1}$ of data. With this amount, it will be possible to explore extremely rare interactions such as the Higgs self-coupling and the SM can be probed with even higher precision, which might lead to signs of physics beyond the SM.

3.5.2. Upgrades for the ATLAS experiment

Run 3 followed by the HL-LHC upgrade will both increase the instantaneous luminosity. This increase leads to even more demanding requirements on the ATLAS experiment such as greater radiation hardness of electronical components. With an expected mean pile-up of $\langle\mu\rangle \sim 200$ and an increased background radiation the reconstruction of events becomes much more challenging and requires new components.

For this purpose, the inner detector will be fully replaced by the all-silicon Inner Tracker (ITk) in LS3 [85, 86]. This detector increases the number of readout-channels by a factor of approximately 7, to over 600 million pixels and over 60 million strips, and will extend the acceptance from $|\eta| < 2.5$ to $|\eta| < 4.0$. It needs to be radiation hard to withstand the high particle flux closest to the interaction point.

In LS2 the end-cap system of the muon spectrometer will be replaced with so-called New Small Wheels [87]. These components can be operated in a higher background radiation and use novel detector technologies to improve the muon tracking and the trigger system. In LS3 the trigger and readout system of the muon spectrometer will be completely replaced to cope with the increasing performance of the HL-LHC.

In addition to these detector parts, the ATLAS trigger and data acquisition system will be upgraded and new readout electronics will be installed to cope with the increased pile-up and data rate [88].

With these upgrades, ATLAS will be able to increase its performance under more challenging conditions and provide precise collision data for the next decades.

CHAPTER 4

Monte Carlo simulation and object reconstruction

To be able to test theoretical models against data from the ATLAS detector, a Monte Carlo (MC) simulation needs to be performed. In an MC simulation, a random number generator is used to describe the interaction of particles and this simulated dataset can then be compared to the reconstructed collision data. The first part of this chapter will briefly summarise the necessary steps to create such an MC dataset and the second part will elaborate on the techniques used to reconstruct collision data.

4.1. The Monte Carlo method

4.1.1. Event generation

The first step is the event generation where proton-proton collisions are simulated. Here, the underlying structure of the protons, the quarks and gluons need to be considered. These constituents can be described as partons. During a collision, the partons transfer a fraction of the proton's momentum. This fraction can be determined using parton distribution functions (PDFs), which are extracted from experimental results and are scale dependant. These PDF sets serve as input for the actual interaction, called the hard scattering process, which is calculated with the matrix element (ME) method. This method is the underlying mathematical description of Feynman diagrams and perturbatively models the hard scattering process with different orders of precision. Higher orders include additional corrections leading to more complex calculations. It is not guaranteed that a higher order calculation gives a correction closer to the true value of the full calculation. The ME result gives the probability for a certain process and the cross-section can be determined. So far, not considered is the photon and gluon emission of the partons due to bremsstrahlung. This effect is incorporated by adjusting the kinematic distributions of the final ME particles, called showering. Finally, QCD confinement requires partons to form hadrons, which subsequently decay until stable particles are reached. This is the final state of the proton-proton collision.

4.1.2. Detector simulation

It is important to account for the detector acceptance defined by the geometry and the resolution of the different sub-detectors, which affects all data collected by ATLAS. This computational intensive task is done by simulating the interaction of particles and their decay products with the ATLAS detector and all its subsystems using GEANT4 [89]. A faster simulation algorithm can be used to replace the calorimeter response with a parameterisation of the shower shapes [90]. In addition, pile-up effects need to be considered and the MC samples are re-weighted accordingly.

4.1.3. Object reconstruction

All simulated events are processed through the same reconstruction algorithms as the data creating an MC dataset in the same format [91]. This procedure ensures a consistent comparison of data and MC simulation.

4.2. Object reconstruction and identification of particles

In the following, an outline of the main reconstruction algorithms will be presented. Where details vary from analysis to analysis, the techniques used in the $t\bar{t}H(H \to b\bar{b})$ analysis [4] are described.

4.2.1. Vertex reconstruction

Multiple vertex candidates from the proton-proton interactions are reconstructed using inner detector tracks. The vertex with the highest scalar sum of the transverse momentum squared, $\sum p_{\mathrm{T}}^2$, of the corresponding tracks is defined as the primary vertex. Only events with at least one vertex with two or more tracks with a transverse momentum $p_{\mathrm{T}} > 0.4\,\mathrm{GeV}$ are considered for analysis.

4.2.2. Leptons

Electrons are reconstructed by associating tracks of the inner detector with an energy deposition (cluster) in the electromagnetic calorimeter [92–94]. They must satisfy additional requirements on $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ and $|\eta| < 2.47$, with the exclusion of the calorimeter barrel end-cap transition region ($1.37 < |\eta| < 1.52$). Further, electrons must fulfil *Loose* identification criteria, as described in Reference [94].

Muons are reconstructed by combining tracks in the inner detector and track segments of the muon spectrometer [95, 96] and are required to have $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ and $|\eta| < 2.5$.

With an average lifetime of only $2.9 \cdot 10^{-13}\,\mathrm{s}$, tau leptons decay before they reach the inner detector and are not classified as (light) leptons in this analysis.

A contribution of non-prompt leptons originating from hadronic decays can be reduced by choosing a *Loose* lepton isolation working point [94, 95] for both electrons and muons. This isolation criterion is based on information from the inner tracking detector and calorimeter. Finally, the vertex matched to the lepton tracks is required to be the primary vertex with constraints on the transverse and longitudinal impact parameters $|z_0 \sin\theta| < 0.5\,\text{mm}$ and $|\frac{d_0}{\sigma_{d_0}}| < 5\,(3)$ for electrons (muons).

4.2.3. Jets

Jets are reconstructed from three-dimensional topological clusters in the calorimeters [97] using the anti-k_t jet algorithm [98] implemented in the FastJet package [99] with a radius parameter $R = 0.4$. The cluster energy is corrected using local cluster calibration, consisting of weighting the energy deposits arising from electromagnetic showers and those from hadronic showers. The final jet energy calibration factors are obtained from simulation and in situ corrections based on 13 TeV data [100]. After applying these factors, jets are required to satisfy $p_{\mathrm{T}} > 25\,\text{GeV}$ and $|\eta| < 2.5$ [101]. To suppress pile-up effects a jet vertex tagger (JVT) algorithm [102] is used that matches jets with $p_{\mathrm{T}} < 60\,\text{GeV}$ and $|\eta| < 2.4$ consistently to tracks originating from the primary vertex.

Hadronically decaying tau leptons can be distinguished from jets using a multivariate discriminant based on calorimeter and tracking information [103]. These τ_{had} candidates are required to satisfy $p_{\mathrm{T}} > 25\,\text{GeV}$, $|\eta| < 2.5$, and the *Medium* tau identification working point.

4.2.4. *b*-tagging

For signal and background selection it is crucial to identify jets containing b-hadrons and separate them from jets containing c-hadrons and light jets. For this, the multivariate b-tagging algorithm *MV2c10* is used, which combines the output of an impact-parameter-based algorithm with the reconstruction of an inclusive secondary vertex and the information of a multi-vertex fitter that reconstructs the b- to c-hadron decay chain [104, 105].

The algorithm uses four working points referred to as *loose*, *medium*, *tight*, and *very tight* corresponding to a b-jet efficiency of 85 %, 77 %, 70 %, and 60 %, respectively.

4.2.5. Tag rate function

Modelling regions with a high b-tagging multiplicity is a challenging task. In regions with three or four b-tags, the number of simulated Monte Carlo events is drastically reduced. An approach to solve this problem is the use of the tag rate function (TRF) method [106]. Here, no cut is applied based on the b-tagging requirements, all events are considered and multiplied by a TRF weight. This weight reflects the probability of the given event to contain the desired number of b-jets and is obtained through the jet tagging efficiency $\epsilon(f, \eta, p_{\mathrm{T}})$, which depends on the jet flavour f, the pseudorapidity η, and the transverse momentum p_{T}. With this efficiency the probability for an event with N jets to contain exactly N_b b-jets is given by

$$P(N_{\mathrm{tag}} = N_b | N_{\mathrm{jets}}) = \sum_{m+n=N_{\mathrm{jets}}} \left(\prod_{i \in T_m} \epsilon_i \prod_{j \in U_n} (1-\epsilon_j) \right), \tag{4.1}$$

where the sum is computed for all permutations in which T_m (U_n) designates the subset of m (n) jets considered (un)tagged.

Therefore, the probability for inclusive b-tagging regions can be computed with

$$P(N_{\mathrm{tag}} \geq N_b | N_{\mathrm{jets}}) = 1 - \sum_{N_{b'} < N_b} P(N_{\mathrm{tag}} = N_{b'} | N_{\mathrm{jets}}). \tag{4.2}$$

For this procedure, a permutation is selected among all the possible combinations of N jets and a given number of b-tags. In a first step, the sum of the TRF weights, S, of all permutations corresponding to the number of b-jets is calculated, and each partial sum, Si, is recorded. Next, a random number uniformly distributed between 0 and S is chosen and finally, the permutation i corresponding to the partial sum up to i, which is greater or equal to the random number, is selected. This method is illustrated in Figure 4.1.

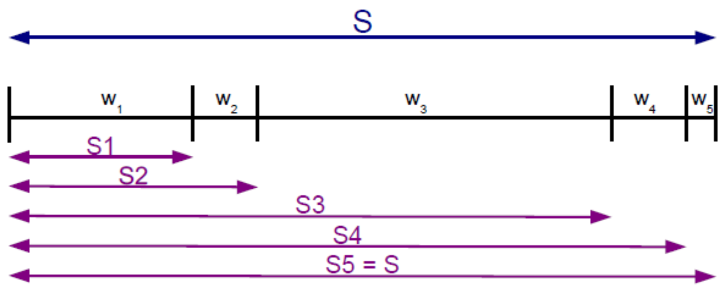


Figure 4.1.: Illustration of the choice of permutation in a case with five possible permutations (e.g. one b-tag among five jets). The total sum S is divided in partial sums Si, with $S5 = S$. A random number is then generated between 0 and S. Depending on the interval $Si - S(i-1)$ (with $S0 = 0$) that includes this number, a corresponding TRF weight w_i is selected, e.g. for a random number between $S2$ and $S3$ the permutation three with TRF weight w_3 is chosen.

4.2.6. Missing transverse energy

Due to conservation of momentum the total momentum in the beam direction is expected to be zero for a collision. Particles that are not reconstructed in the detector, e.g. neutrinos, can lead to non-zero values. This missing transverse momentum $\boldsymbol{p}_{\mathrm{T}}^{\mathrm{miss}}$, and its magnitude $E_{\mathrm{T}}^{\mathrm{miss}}$, can be used to identify escaping particles and includes contributions from energy deposits in the calorimeters and muon momenta measured in the muon spectrometer. The estimation of $E_{\mathrm{T}}^{\mathrm{miss}}$ is given by the negative vector sum of the transverse momenta of all identified and calibrated objects (leptons and jets) and remaining unclustered energy that is not associated with any of these. The latter is calculated from low-p_{T} tracks from the inner detector matched to the primary vertex to make it more robust against pile-up contamination [107, 108]. The contributions to $\boldsymbol{p}_{\mathrm{T}}^{\mathrm{miss}}$ in the transverse (x, y) plane are therefore:

$$E_{\mathrm{x,y}}^{\mathrm{miss}} = E_{\mathrm{x,y}}^{\mathrm{miss},\,e} + E_{\mathrm{x,y}}^{\mathrm{miss},\,\gamma} + E_{\mathrm{x,y}}^{\mathrm{miss,\,jets}} + E_{\mathrm{x,y}}^{\mathrm{miss,\,clus}} + E_{\mathrm{x,y}}^{\mathrm{miss},\,\mu}, \tag{4.3}$$

and the value of $E_{\mathrm{T}}^{\mathrm{miss}}$ is calculated as:

$$E_{\mathrm{T}}^{\mathrm{miss}} = \sqrt{(E_{\mathrm{x}}^{\mathrm{miss}})^2 + (E_{\mathrm{y}}^{\mathrm{miss}})^2}. \tag{4.4}$$

This can be a very useful quantity for searches beyond the SM since many theories predict heavy minimally interacting particles that give large contributions to $E_{\mathrm{T}}^{\mathrm{miss}}$.

4.2.7. Overlap removal

To remove overlaps and resolve ambiguities between reconstructed particles from double-counting, the distance $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ in the pseudorapidity-azimuthal angle space between them is used. Jet reconstruction does not take into account the energy deposition of electrons and, to avoid double-counting, the closest jet with $\Delta R < 0.2$ of an electron is rejected. If the nearest jet passing this selection is within $\Delta R = 0.4$ of the electron, the electron is removed. To suppress background contributions of muons originating from semi-leptonic decays of c- and b-quarks, muons are removed if they are with $\Delta R < 0.4$ of a jet. However, the closest jet is removed instead if it is reconstructed with fewer than three inner detector tracks. This procedure avoids an inefficiency caused by the significant energy loss of high-energetic muons traversing the calorimeter. τ_{had} candidates are required to be separated by more than $\Delta R = 0.2$ from the closest electron or muon; otherwise they are discarded.

CHAPTER 5

The search for $t\bar{t}H(H \to b\bar{b})$

The decay of a Higgs boson into a pair of bottom quarks has the largest branching fraction of about 57 %. The analysis targets events in which one or both top quarks decay semi-leptonically, producing an electron or a muon, and uses a dataset of 36.1 fb^{-1} of proton-proton collisions collected by the ATLAS detector in 2015 and 2016 [4]. In addition to a direct sensitivity to the top quark Yukawa coupling, this decay channel is also sensitive to the Yukawa coupling of the b-quark. The main experimental challenge is to reconstruct and separate the signal from large backgrounds, mainly from $t\bar{t}$ + jets. In this chapter, the $t\bar{t}H(H \to b\bar{b})$ analysis with emphasis on the single-lepton channel is presented. Further, studies for the next round of the analysis using an updated reconstructed software, which is required for the full Run 2 dataset are shown.

5.1. Event selection

All recorded events were required to pass single-lepton triggers. The reason for this is that at least one electron or one muon is expected from a single-lepton or dilepton top quark decay. The conditions to fulfil these triggers were either a low lepton p_{T} threshold and a lepton isolation requirement, or with a higher threshold a looser identification criterion and no isolation requirement. A detailed list can be found in Appendix A.

To ensure good separation from other objects and prevent overlap with large energy deposits in the calorimeters or high p_{T} tracks, both electrons and muons must fulfil the *Gradient* isolation [94, 95]. Electrons are required to pass the *TightLH* working point [94], while muons must have the *Medium* quality criterion [95]. These criteria favour the selection of prompt leptons originating from W and Z boson decays.

In the single-lepton channel, the transverse momentum of the lepton must be greater than 27 GeV.

In the dilepton channel, exactly two leptons with opposite charge are selected. The invariant mass of an electron or muon pair must be above 15 GeV (to reject low-mass resonances) and outside of the Z boson mass window (83–99 GeV, to reject Z + jets events). Only one of the leptons is required to fulfil the $p_{\mathrm{T}} > 27$ GeV requirement of the single-lepton channel, for the other this threshold is reduced to 15 GeV in case the of two electrons and to 10 GeV for the presence of a muon.

There is a small chance that the decay products of at least one of the top quarks and Higgs boson are Lorentz boosted in such a way that they cannot be resolved by the detector and are, therefore, reconstructed as a single large radius jet [109]. This scenario is studied in a dedicated boosted channel, included as a sub-channel in the single-lepton channel. The methods employed to analyse the boosted topology will not be discussed in this thesis.

The high b-tagging multiplicity of the $t\bar{t}H(H \to b\bar{b})$ signal can be utilised to separate signal from background, therefore, a b-tagging discriminant value is assigned to each jet according to the tightest b-tagging WP it satisfies. The numeric value ranges from 1 for a jet that does not satisfy any of the b-tagging criteria defined by the loosest WP up to 5 for jets satisfying the very tight WP.

5.2. Signal and background modelling

The $t\bar{t}H$ signal and the background processes need to be accurately modelled using MC simulations following the description of Section 4.1. Pile-up effects were simulated with additional interactions generated with Pythia 8.186 [110]. In samples not simulated by the Sherpa event generator, the decays of b- and c-hadrons were computed with EvtGen v1.2.0 [111].

5.2.1. Modelling of the $t\bar{t}H$ signal

The $t\bar{t}H$ signal was generated with MadGraph5_aMC@NLO [112] for the ME interfaced to the Pythia 8.210 parton shower using the A14 tune [113]. As a PDF set, NNPDF3.0NLO [114] was used. The top quark decays including spin information were simulated with MadSpin [115], where the Higgs boson mass was fixed at 125 GeV and all decay modes were considered. The $t\bar{t}H$ cross-section of 507^{+35}_{-50} fb was calculated at NLO accuracy in QCD including NLO electroweak corrections [65, 116–120]. The different branching ratios were calculated with HDECAY [65, 121].

5.2.2. The main background: $t\bar{t}$ + jets

The largest background contribution originates from $t\bar{t}$ + jets. This background was normalised to the predicted cross-section of 832^{+46}_{-51} pb calculated by Top++2 [122] at NNLO+NNLL accuracy in QCD [122–125]. The inclusive $t\bar{t}$ background was modelled using Powheg-Box v2 NLO [126–129] with $h_{\mathrm{damp}} = 1.5\ m_{\mathrm{top}}$ [130]. Parton shower and hadronisation were computed with the same parton shower model as the signal. This method only simulated the ME of the $t\bar{t}$ decay at NLO accuracy, whereas additional jets beyond the first parton were included with parton shower radiation.

The $t\bar{t}$ + jets background is categorised by the flavour of additional jets in the event by counting the number of b- or c-hadrons within $\Delta R < 0.4$ of the jet [131].

A single b-jet contains exactly one b-hadron with p_{T} above 5 GeV. Jets containing more than one b-hadron are labelled as a B-jet (without p_{T} requirement on the second hadron). c- and C-jets are defined analogously, excluding b- and B-jets. With this labelling the following exclusive background categories are defined for the fit:

- $t\bar{t} + \geq 1b$: events with at least one b- or B-jet, not counting heavy-flavour jets from top quark or W boson decays
- $t\bar{t} + \geq 1c$: events with no b- or B-jet but at least one c- or C-jet
- $t\bar{t}$ + light jets: events not containing any heavy-flavour jets (besides from top quark or W boson decays)

Additionally, subcategories of $t\bar{t} + \geq 1b$ and $t\bar{t} + \geq 1c$ are defined to assess uncertainties:

- $t\bar{t} + b\bar{b}$: events with exactly two b-jets
- $t\bar{t} + b$: events with only one b-jet
- $t\bar{t} + B$: events with only one B-jet
- $t\bar{t} + \geq 3b$: remaining $t\bar{t} + \geq 1b$ events
- $t\bar{t} + b$ (MPI/FSR): events with additional b-jets entirely originating from multi-parton interactions (MPI) or b-jets from final-state radiation (FSR)

Events containing c-jets are categorised analogously.

The dominant $t\bar{t} + \geq 1b$ background is modelled with the highest available theoretical precision. This is achieved by scaling the relative contributions of the different subcategories $t\bar{t} + \geq 3b$, $t\bar{t} + b\bar{b}$, $t\bar{t} + B$, and $t\bar{t} + b$ in the Powheg+Pythia 8 sample to those predicted by an NLO $t\bar{t}b\bar{b}$ sample, which was modelled with Sherpa+OpenLoops [132, 133] including parton showering and hadronisation [134]. The simulation was computed with Sherpa 2.1.1 and the CT10 four-flavour (4F) scheme PDF set [135, 136]. This $t\bar{t}b\bar{b}$ sample was renormalised to the CMMPS value [137] and hereafter will be referred to as Sherpa4F. Figure 5.1 shows the subcategories of the $t\bar{t} + \geq 1b$ background, modelled with the Powheg+Pythia inclusive $t\bar{t}$ sample compared to the $t\bar{t}b\bar{b}$ Sherpa4F sample.

5.2.3. Other backgrounds

Other background contributions can arise from $t\bar{t}W$ and $t\bar{t}Z$ events that were generated using MadGraph5_aMC@NLO for the ME interfaced to the Pythia 8.210 parton shower with the A14 parameter set and the NNPDF3.0NLO PDF. Wt and s-channel single top quarks were simulated with Powheg-Box v1 at NLO accuracy using the CT10 PDF tune. Overlap removal between $t\bar{t}$ and Wt is performed employing the *diagram removal* scheme [138]. Single top quarks in the t-channel were computed with the

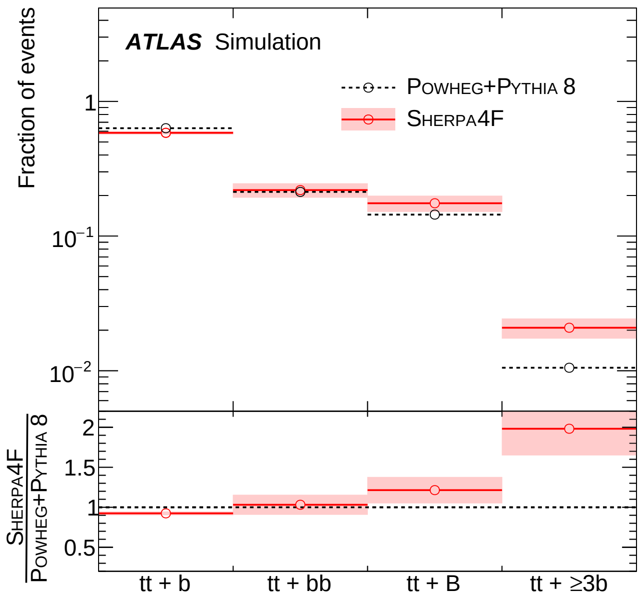


Figure 5.1.: Relative event fractions of the $t\bar{t} + b$, $t\bar{t} + b\bar{b}$, $t\bar{t} + B$, and $t\bar{t} + \geq 3b$ subcategories before event selection. The inclusive Powheg+Pythia 8 sample is compared to the $t\bar{t}b\bar{b}$ Sherpa4F sample. The fractions are normalised to the sum of these four contributions, without considering the $t\bar{t} + b$ (MPI/FSR) subcategory.

Powheg-Box v1 event generator at NLO accuracy with the four-flavour PDF set CT10 4F. All single top quark samples were interfaced to Pythia 6.428 [139] using the Perugia 2012 set of tuned parameters for the parton shower and hadronisation [140]. All samples are normalised according to theoretical NNLO cross-section calculations [141–143].

W/Z + jet events were generated using Sherpa 2.2.1. The ME was calculated using Comix [144] and OpenLoops, and merged to the Sherpa parton shower [145] using the ME+PS@NLO prescription [146] and the NNPDF3.0NNLO PDF tune. The normalisation is performed with NNLO cross-section calculations [147]. Diboson + jet samples were generated with Sherpa 2.1.1 [148].

Higgs boson production associated with a single top quark is included as a minor background contribution. tWH was generated with MadGraph5_aMC@NLO interfaced to Herwig++ [149] with the CTEQ6L1 PDF set. $tHqb$ was produced at LO with MadGraph5_aMC@NLO interfaced to Pythia 8 using the CT10 4F PDF set.

$t\bar{t}t\bar{t}$ as well as $t\bar{t}WW$ were generated at LO accuracy with MadGraph5_aMC@NLO interfaced to Pythia 8. tZ events use the same generator but were interfaced to Pythia 6. tZW also uses the same generator interfaced with Pythia 8 but at NLO.

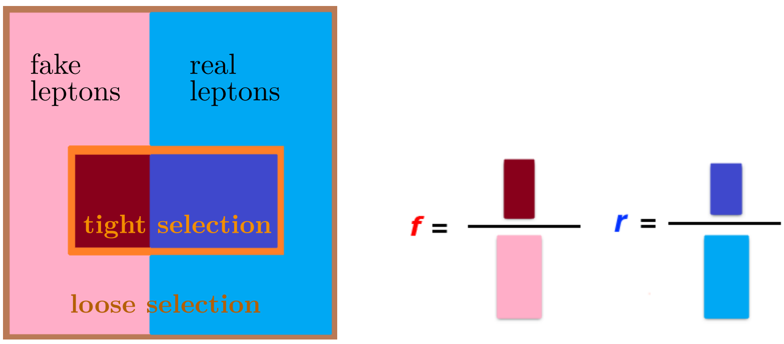
Another source of backgrounds is non-prompt leptons and fake leptons, which will be discussed in detail in the following section.

5.3. Non-prompt leptons and fake leptons

Background from non-prompt leptons and fake leptons requires a special treatment in the events with many jets and b-jets. Non-prompt leptons originate mostly from semi-leptonic decays of c- and b-quarks, photon conversion, and kaon decays. Fake lepton signatures can occur when jets or photons are misidentified as a reconstructed lepton. This scenario happens mostly for electrons. The non-prompt leptons and fake lepton background will be referred to collectively as fakes. To account for the different lepton identification and reconstruction, electrons and muons will be considered separately. This section will focus on the fake estimation using data-driven techniques in the single-lepton channel. In the dilepton channel, the background is extracted from simulation and normalised to data in a control region with two same-sign leptons.

5.3.1. Fake estimation with the matrix method

A robust fake estimation cannot be accurately modelled in MC. Therefore, a data-driven approach, the matrix method, is used [150]. This method requires two event regions with different lepton selection criteria. One region, referred to as *tight*, has the same lepton selection criteria as the analysis. By loosening the lepton selection requirements, the *loose* region can be obtained (see Figure 5.2). Both regions contain fake and real leptons.



(a) Tight and loose selection regions for the matrix method, where the tight region is included in the loose.

(b) The fake (real) efficiency is defined as the number of fake (real) leptons in the tight selection divided by the number of fake (real) leptons in the loose selection.

Figure 5.2.: Illustration of the matrix method.

Therefore, the number of leptons in the loose region (N^{loose}) and the number of leptons in the tight region (N^{tight}) region can be written as:

$$N^{\text{loose}} = N^{\text{loose}}_{\text{real}} + N^{\text{loose}}_{\text{fake}}, \tag{5.1}$$

$$N^{\text{tight}} = N^{\text{tight}}_{\text{real}} + N^{\text{tight}}_{\text{fake}}. \tag{5.2}$$

To calculate the number of fake leptons in the tight selection, real and fake efficiencies are introduced:

$$r = \frac{N_{\text{real}}^{\text{tight}}}{N_{\text{real}}^{\text{loose}}}, \tag{5.3}$$

$$f = \frac{N_{\text{fake}}^{\text{tight}}}{N_{\text{fake}}^{\text{loose}}}. \tag{5.4}$$

With these efficiencies Equations 5.1 and 5.2 can be rewritten to extract the parameter of interest:

$$N_{\text{fake}}^{\text{tight}} = \frac{f}{r-f} \cdot (r \cdot N^{\text{loose}} - N^{\text{tight}}). \tag{5.5}$$

The number of fake leptons in the analysis region can be obtained if the fake and real efficiencies are estimated and the number of tight and loose events are extracted from the data. To apply Equation 5.5 to a binned distribution, it is rewritten as a per-event weight that is applied on the loose selection to determine the fake background in the analysis:

$$w_i = \frac{f}{r-f}(r - P_i), \tag{5.6}$$

where i stands for the event and $P_i = 1$ if the loose event passes also the tight selection and $P_i = 0$ otherwise. Equation 5.5 can then be rewritten as

$$N_{\text{fake}}^{\text{tight}} = \sum_i w_i N^{\text{loose}}. \tag{5.7}$$

If a loose event passes the tight selection ($P_i = 1$), the weight will be negative. Tight leptons have a high possibility to include real leptons, consequently, they should be removed from the fake estimation. In contrast, lepton events in the loose and not-tight selection are likely fake leptons and contribute to positive weights. Therefore, a large difference between the loose and tight selection is desired to reduce the possibility of a large number of negative weights. This argument highlights the importance of the loose selection choice for the matrix method to be successful. Two important requirements for the loose selection are to include both the tight selection and all possible sources of fakes, which are expected in the analysis region. Ideally, both the real and fake efficiencies should be determined in the analysis regions. However, fake contributions are already highly suppressed in these regions because of the efficient background rejection in the analysis region. Therefore, real and fake efficiencies are extracted in custom-built regions, which are enriched in either real or fake leptons. Fake enriched regions differ for electrons and muons. A high electron fake contribution is expected in the low $E_{\text{T}}^{\text{miss}}$ region of $W \to \ell\nu_\ell / t\bar{t}$ events, therefore $E_{\text{T}}^{\text{miss}} < 20\,\text{GeV}$ is required. Muon fakes can originate from semi-leptonic b-decays. These events can be selected with $|d_0^{\text{sig}}| > 5$, where the muon impact parameter significance is defined as $d_0^{\text{sig}} = d_0/\sigma_{d_0}$.

Fake efficiencies can then be estimated in the following:

$$f = \frac{N_{\text{data}}^{\text{tight}} - N_{\text{MC}_{\text{real}}}^{\text{tight}}}{N_{\text{data}}^{\text{loose}} - N_{\text{MC}_{\text{real}}}^{\text{loose}}}, \tag{5.8}$$

where MC events are subtracted in order to reduce a real lepton contribution. These events are estimated from MC simulations of all relevant processes for the analysis, such as $t\bar{t}$, single top, $t\bar{t}$ + W/Z, W/Z + jet, and diboson.

Real efficiencies are determined using a tag-and-probe method on $Z \to ee$ and $Z \to \mu\mu$ events. To ensure a clean signature, events with a pair of same-flavour opposite-sign loose or tight leptons containing at least one jet are selected. In addition, the invariant mass of the dilepton is required to be between 60–120 GeV. Leptons passing the tight selection are labelled as tag, while leptons passing the loose selection are considered as probe. The efficiency is then computed as the number of probes that pass the tight criteria divided by the number of all probes:

$$r = \frac{N_{\text{probe}}^{\text{tight}}}{N_{\text{probe}}^{\text{tight+loose}}}. \tag{5.9}$$

The real and tight efficiencies depend on kinematic properties of the event such as lepton p_{T}, leading jet p_{T}, and lepton η and are parameterised accordingly. In a next step, these efficiencies can be used to calculate the fake contribution of the analysis region.

Migration to an updated analysis software

For the 2017 data recording year the ATLAS analysis software received a major update. This update is not included in the current $t\bar{t}H(H \to b\bar{b})$ analysis targeting the 2015 and 2016 dataset. However, for an updated analysis including the full Run 2 dataset a migration to this updated analysis software is necessary. Therefore, the matrix method software package was ported to this updated analysis software. Besides major improvements of the reconstruction algorithms, one of the structural code changes was switching from the code management tool CMT to CMake. This change significantly increased the compiler speed and robustness but required a major revision of the code packages used to employ the matrix method. The software version and revision control system was also changed from SVN to GitLab.

After successfully porting all packages, the performance of the matrix method is evaluated in both analysis release versions. For this purpose, the majority of the parameters needed for the reconstruction and the definition of the regions are kept the same between the two analysis versions. It is not possible to use exactly the same parameters, for example, the jet energy scale and resolution and the modelling of pile-up effects could be significantly improved in the updated analysis version due to a better understanding of the detector and no previous version of the modelling can be selected. Another example is updated data quality criteria that define which collision event is sufficiently precisely and accurately recorded and can be used for the analysis or needs to be excluded, which leads to a change of the integrated luminosity. Therefore, even for the same dataset small deviations between the two analysis versions are expected.

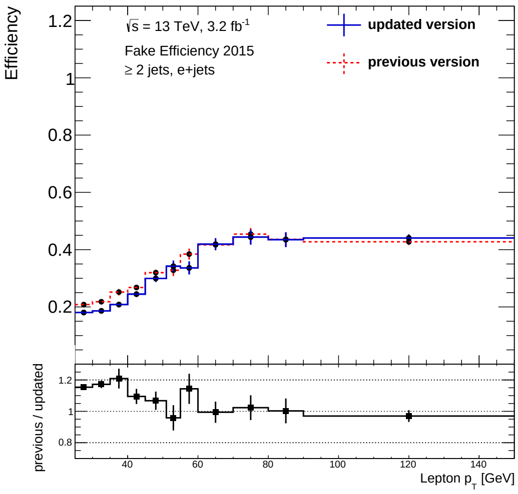
Both versions use the same lepton triggers. As stated in Section 5.1, the 2016 (and 2017) low lepton p_{T} threshold triggers apply isolation requirements to limit the bandwidth and to reduce the contribution from leptonically decaying hadrons, whereas at high p_{T} this background is insignificant and no isolation requirements are applied in order to increase the trigger efficiency. In addition, these isolation requirements reduce the number of fakes, which is usually desirable. However, for the loose region a fake enriched environment is desired. Therefore, pre-scale (PS) triggers without isolation requirements for the low lepton p_{T} regions of the 2016 and 2017 data are studied. A PS trigger can be used for a region where it is not possible to record all events due to the bandwidth limitation of the detector. These PS triggers reduce the data rate by only selecting every n^{th} event. A detailed list of the lepton triggers can be found in Appendix A.

When available, the same MC samples for the single top, $t\bar{t} + W/Z$, W/Z + jet and diboson events are used. For $t\bar{t}$ events, the MC sample of the updated version uses a different Powheg+Phythia 8 setup. For the 2015 and 2016 dataset pile-up reweighting of MC events is done with a distribution of the average pile-up $\langle\mu\rangle$ (that differs between the analysis versions) and for the 2017 dataset the MC events are reweighted according to the actual pile-up μ. New muon isolation working points with a higher pile-up robustness were recently added to the updated analysis software. However, to have a consistent loose and tight lepton definition for both release versions, identical, but soon obsolete lepton identification and isolation requirements are used. To pass the loose (tight) selection, electrons are required to satisfy the *LooseAndBLayerLH* (*TightLH*) working point [94], while muons must have *Loose* (*Medium*) quality criteria [95]. Each lepton is required to fulfil *None* (the *Gradient*) isolation [94, 95]. Overlap removal is performed by considering the loose lepton definition to correctly account for this region. In addition, the standard procedure where the overlap removal is based on the tight lepton definition was also tested and showed a similar result.

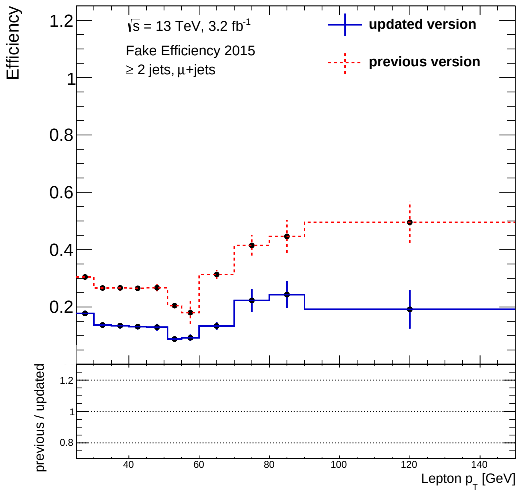
Following these conditions, real and fake efficiencies are computed for different jet and b-jet multiplicities as a function of kinematic variables. Higher jet and especially b-jet multiplicities lead to drastically reduced statistics and, therefore, a significant increase in statistical uncertainty. Sufficient statistics are assured when using regions with exactly one jet or at least two jets and no b-jet requirements. Real and fake efficiencies as a function of lepton p_{T} are compared between the two release versions and can be seen in Figure 5.3 for the 2015 dataset, in Figure 5.4 for the 2016 dataset, and in Figure 5.5 for the 2016 dataset using PS triggers for the low lepton p_{T} region. Figure 5.6 shows the efficiencies for the 2017 dataset using the nominal and PS triggers, which were only recorded with the updated analysis release.

Throughout all years, the electron and muon real efficiencies as well as the electron fake efficiencies estimated from data show a good agreement. Small deviations between the release versions are expected because of different data quality criteria and changes in the MC modelling used to estimate the real lepton contributions, see Equation 5.8. On the contrary, muon fake efficiencies show a large discrepancy between the two release versions. The shape of the distributions remain similar, whereas the absolute value decreases by a factor of about 2/3 for the updated analysis version. Different parameterisations show the same characteristics between the two release versions.

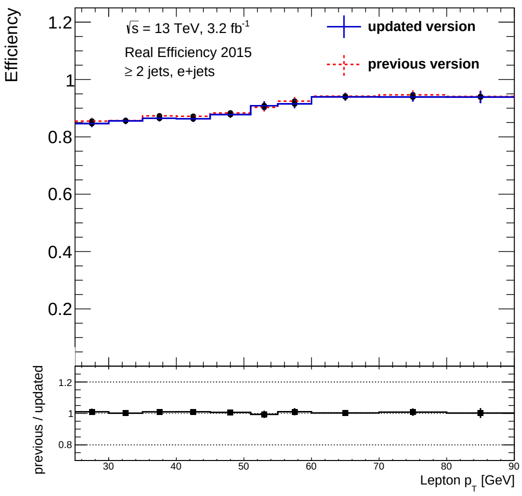
The sudden fake efficiency drop for 2016 and 2017 at $p_{\mathrm{T}} = 61\,\mathrm{GeV}$ ($51\,\mathrm{GeV}$) for electrons (muons) is caused by the isolation requirements of the low lepton p_{T} triggers. Applying PS triggers smoothed the distributions in this transition region. However, these



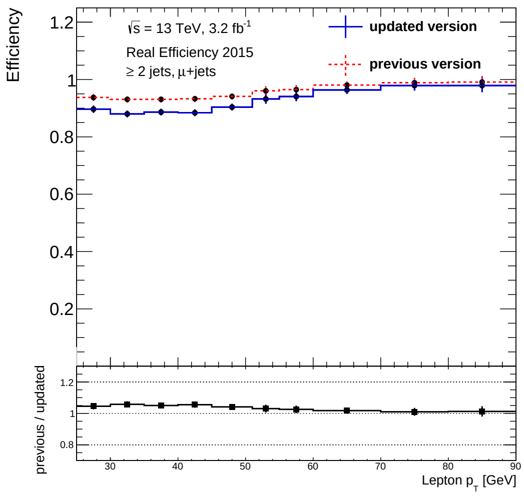
(a) Fake efficiency for the e + jets channel.



(b) Fake efficiency for the μ + jets channel.

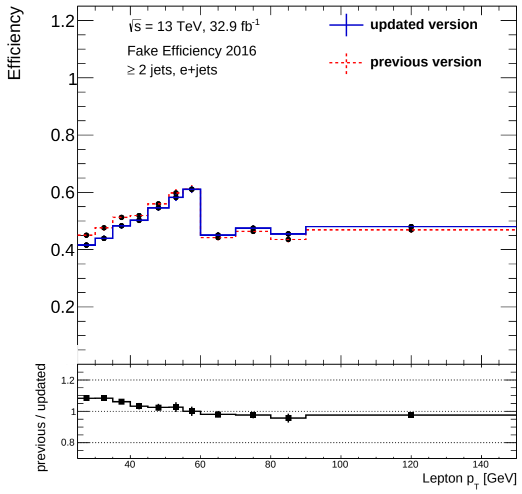


(c) Real efficiency for the e + jets channel.

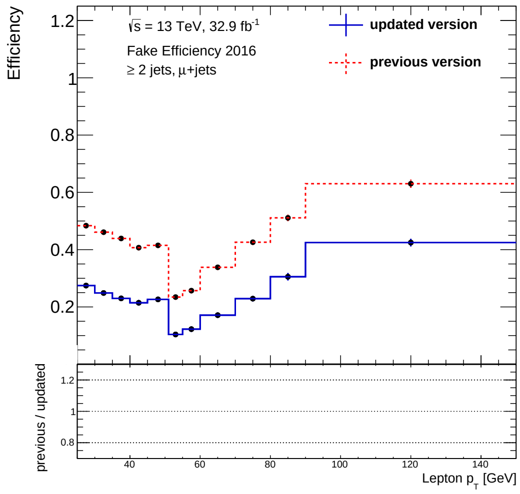


(d) Real efficiency for the μ + jets channel.

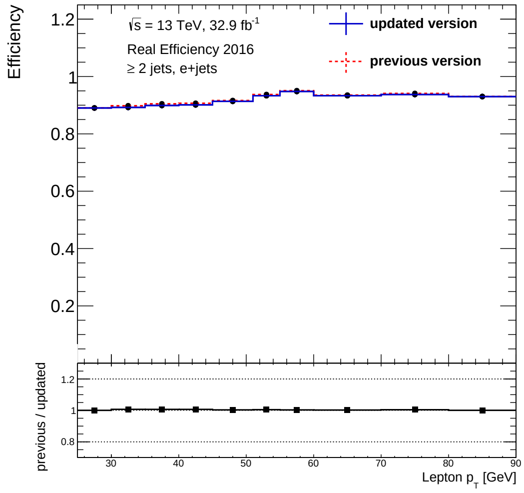
Figure 5.3.: Comparison of fake and real efficiencies between two release versions for the 2015 dataset as a function of lepton p_{T}.



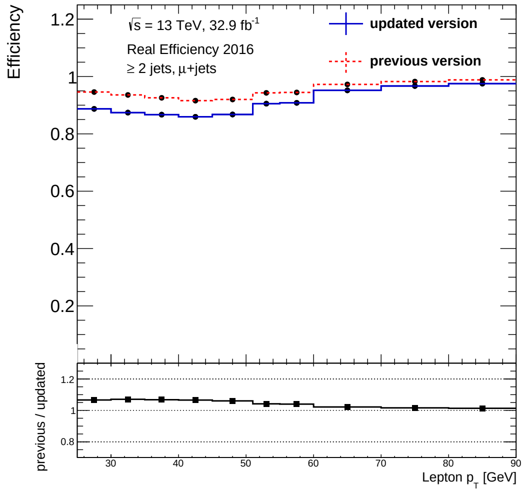
(a) Fake efficiency for the e + jets channel.



(b) Fake efficiency for the μ + jets channel.

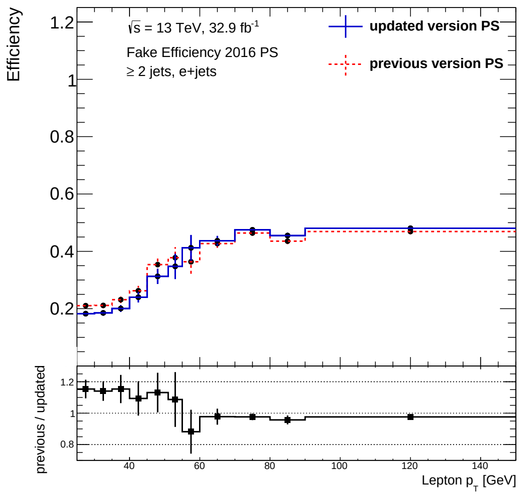


(c) Real efficiency for the e + jets channel.

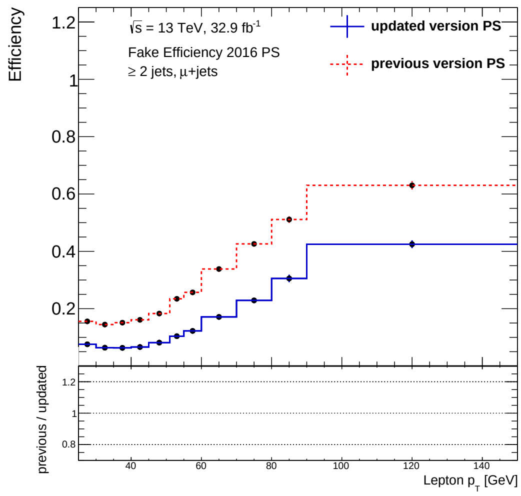


(d) Real efficiency for the μ + jets channel.

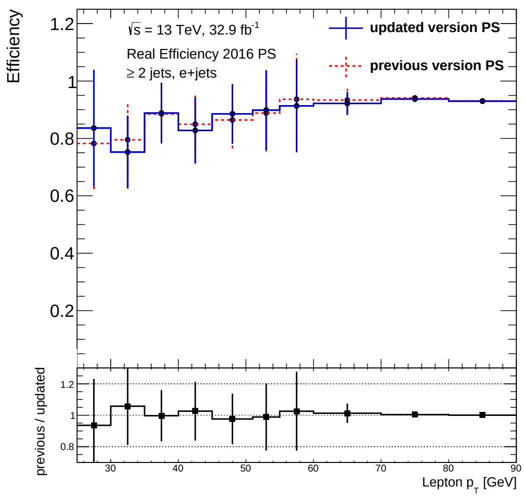
Figure 5.4.: Comparison of fake and real efficiencies between two release versions for the 2016 dataset as a function of lepton p_{T}.



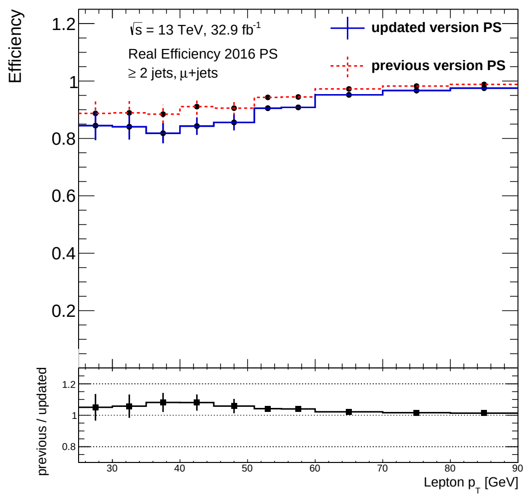
(a) Fake efficiency for the e + jets channel.



(b) Fake efficiency for the μ + jets channel.

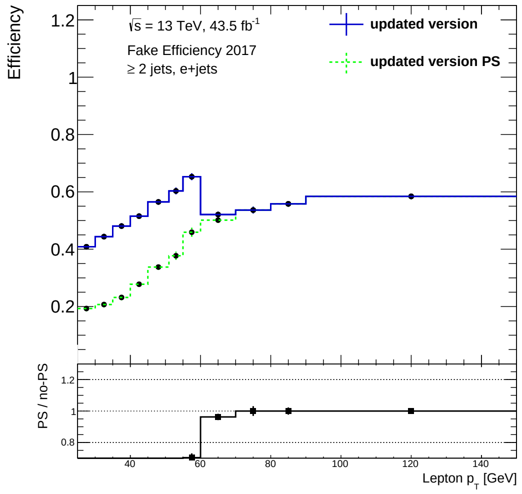


(c) Real efficiency for the e + jets channel.

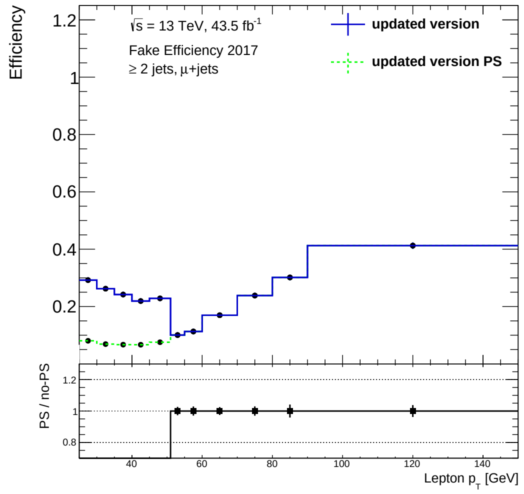


(d) Real efficiency for the μ + jets channel.

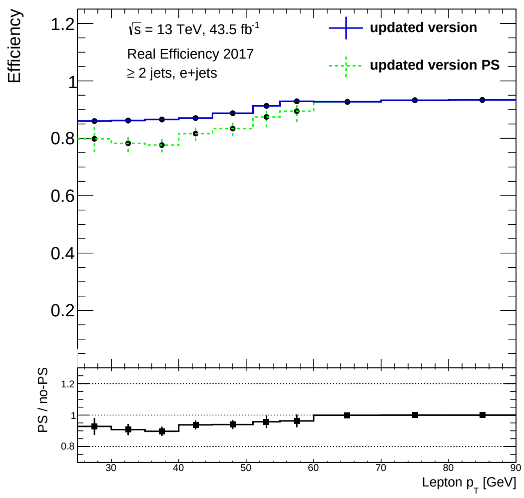
Figure 5.5.: Comparison of fake and real efficiencies between two release versions for the 2016 dataset using PS triggers for the low lepton p_{T} regions (below 61 GeV for electrons and below 51 GeV for muons) as a function of lepton p_{T}.



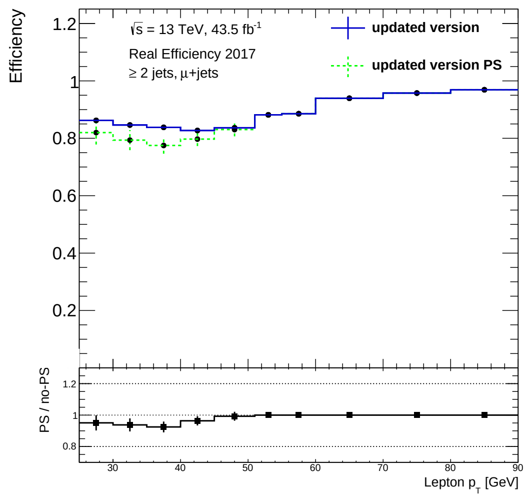
(a) Fake efficiency for the e + jets channel.



(b) Fake efficiency for the μ + jets channel.



(c) Real efficiency for the μ + jets channel.



(d) Real efficiency for the μ + jets channel.

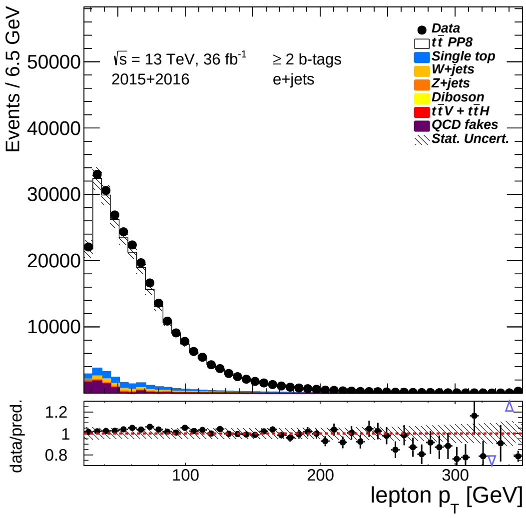
Figure 5.6.: Comparison of fake and real efficiencies for the 2017 dataset using the nominal and PS triggers for the low lepton p_{T} regions (below 61 GeV for electrons and below 51 GeV for muons) as a function of lepton p_{T}. Contrary to the previous figures, a different efficiency distribution is expected when applying PS triggers due to different lepton isolation requirements.

PS triggers can cause events with large trigger PS weights leading to spikes in the fake estimates, and thus significant larger statistical uncertainties. In all tested regions (up to six jets inclusive with at least four b-jets) the matrix method could achieve a better performance in absence of PS triggers.

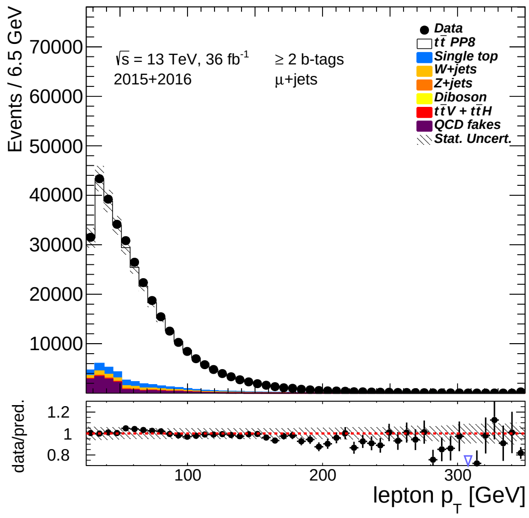
A comparison between data and prediction in a region requiring at least four jets and at least two b-jets for both analysis software versions can be seen in Figure 5.7. The fake and real efficiencies use a combination of the regions requiring exactly one jet and at least two jets, where no cut on the b-jet multiplicity is applied and no PS triggers are used. Electrons use a parameterisation of leading jet p_{T} and $\Delta R(l_{\mathrm{probe}}, \text{closest jet})$, whereas muons use lepton p_{T} and $\Delta\phi(l_{\mathrm{probe}}, E_{\mathrm{T}}^{\mathrm{miss}})$. The overall agreement between data and prediction is within statistical uncertainties; changes in the updated analysis reconstruction software described at the beginning of this paragraph do not yield to a significantly improved agreement. The electron selection shows a similar agreement between both analysis release versions, whereas the muon selection shows a better agreement for the previous release version. A larger fake contribution from muons compared to electrons as seen in the previous analysis version is unexpected, but leads to a better agreement between data and prediction. The significantly lower muon fake rate in the updated release version is caused by the reduced muon fake efficiency. In addition, modelling discrepancies could be caused by extrapolating from a combined exactly one jet and at least two jets region to a four jet inclusive region with at least two b-jets.

Performance of the matrix method

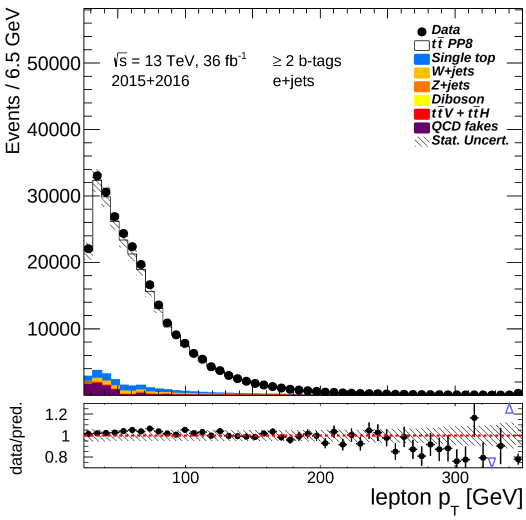
The matrix method was successfully ported to the updated release version. The efficiency distributions between the previous and updated analysis versions show a good agreement for multiple parameterisations except for the muon fake efficiencies. However, the significantly reduced muon fake efficiency is caused by a larger number of events passing the loose muon selection criteria in the updated version relative to the previous version. These additional events fail the isolation criteria of the tighter selection, which leads to a lower muon fake efficiency, see Equation 5.8. One explanation could be related to a change in the overlap removal for muons between the two analysis versions. The overlap of muons and light jets increases and the overlap between muons and b-jets decreases. The muons are not removed, because they are favoured over light jets, whereas the muons would be removed if overlapping with a b-jet. These additional muon events accumulate in the loose selection and, therefore, reduce the fake efficiency. However, this cannot explain why the scale factors are correctly applied to the analysis region to which the efficiency is extrapolated, resulting in a worse agreement between data and prediction (see Figure 5.7). This difference could also be related to the impact parameter significance used to define the muon fake enriched loose region. The modelling of this parameter changed substantially between the two release versions, however its distribution remains similar. This effect is also seen by an analysis targeting $H \to WW^*$ decays using a different framework for the fake estimation, supporting the hypothesis that the source of this reduced efficiency lies outside of the matrix method framework. A temporary solution could be to scale the muon fake efficiencies for the 2015 and 2016 dataset of the



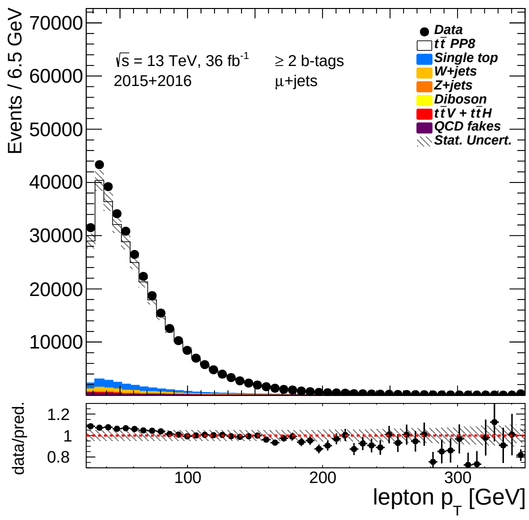
(a) e + jets channel & previous analysis release.



(b) μ + jets channel & previous analysis release.



(c) e + jets channel & updated analysis release.



(d) μ + jets channel & updated analysis release.

Figure 5.7.: Comparison between data and prediction for both analysis software versions for a selection of at least four jets with at least two b-jets as a function of lepton p_{T}. Figures (a) and (b) use the previous release, whereas Figures (c) and (d) use the updated release.

updated analysis version to the previous version and use this scale factor to reweight the 2017 muon fake efficiency accordingly.

For the first time, fake and real efficiencies were estimated for the 2017 dataset. With similar pile-up conditions for the 2017 and 2018 data taking periods, the 2017 efficiencies can be used for both datasets resulting in efficiencies for the full Run 2 dataset, which can not only be used in the $t\bar{t}H$ analysis but also in other analyses with leptonic final states.

The performance of the matrix method highly depends on the definition of the loose region and sufficient statistics. A large amount of negative weights can occur if a loose lepton passes the tight selection, see Equation 5.6 ($P_i = 1$). On the other hand, for similar real and fake efficiencies, the denominator in Equation 5.6 ($r - f$) can become very small compared to the numerator (f) and even converge towards zero resulting in very large weights. Thus, a single event could cause a spike in the distribution. In the next section, a method will be presented that has the potential to remove some of the limitations of the matrix method.

5.3.2. Fake estimation with a tag rate function

Initially, the TRF method was developed to avoid fluctuations caused by low MC statistics in high b-tagging multiplicity regions by extrapolating distributions from low to high b-tag regions, see Section 4.2.5. One premise of this method is the knowledge of the true jet flavour, which is needed for the jet tagging efficiency ϵ. Since this information is not accessible for a fully data-driven approach like the matrix method, the TRF method cannot be applied directly. This issue can be bypassed by using b-tagging information to label jets as b-tagged or not-b-tagged:

$$\epsilon_b(x|N_{\text{jets}}) = \frac{x_{b\text{-tagged}}}{x_{\text{all}}}. \tag{5.10}$$

This redefined jet efficiency can be inserted in Equation 4.1 to obtain a TRF weight, which can be used as in the standard TRF approach.

This section will examine if it is possible to adapt the TRF technique for a fake estimation using the matrix method. Additionally, the result is compared to a fake estimation using the matrix method without TRF. In a first step, the jet tagging efficiency is calculated. After that, the matrix method can be applied to an event selection with sufficient statistics and finally, the TRF method is employed to extrapolate from this region to the desired high b-tagging multiplicity regions.

The following study is performed in the single-lepton channel with the 2015 and 2016 dataset corresponding to $36.1\,\text{fb}^{-1}$ and requires a fixed b-tagging WP. The study is based on the previous version of the analysis software. Applying PS triggers for the low lepton p_{T} region in the 2016 dataset was studied, but it did not show an improvement, therefore, PS triggers are not considered for the final result. The dataset is again split in electron + jets and muon + jets samples, where both of the sets are divided further, each into the following eight regions:

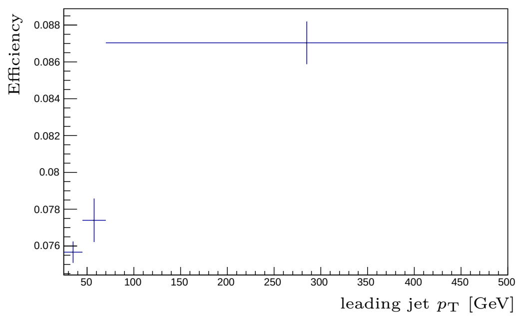
- 4 jets and 2 b-jets, 4 jets and ≥ 3 b-jets

- 5 jets and 2 b-jets, 5 jets and 3 b-jets, 5 jets and ≥ 4 b-jets
- ≥ 6 jets and 2 b-jets, ≥ 6 jets and 3 b-jets, ≥ 6 jets and ≥ 4 b-jets

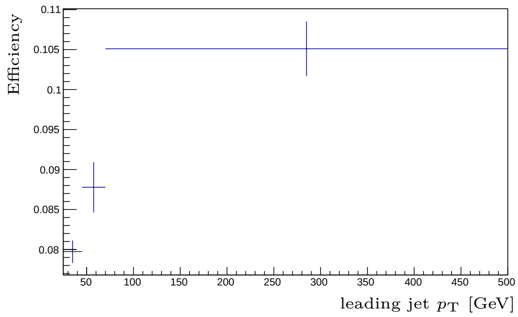
The best result is achieved with a hybrid TRF approach. In this approach, the TRF method is applied on a dataset that has an inclusive (low) b-tagging cut. This cut prevents modelling discrepancies, which can occur when the TRF method is used to extrapolate from a region without a b-jet multiplicity cut to a high b-tag multiplicity region, such as ≥ 4 b-jets. Indeed, this approach comes with the caveat of reduced statistics and, therefore, a balance between reducing modelling discrepancies and losing statistical power needs to be found. Different configurations depending on the number of inclusive jets and b-jets (corresponding to the four different WPs) have been tested and the best result is obtained for a region requiring at least four jets and at least one b-jet corresponding to a WP of 85 %.

Computing different efficiencies depending on the number of jets was also tested, but regions with a jet multiplicity of five or greater lead to large statistical fluctuations and, therefore, a four jet inclusive region is used to calculate the TRF weights.

The jet tagging efficiency depends on different parametrisations. The leading jet p_{T} dependence can be seen in Figure 5.8. In Appendix B.1 additional plots for the remaining WPs can be found.



(a) Efficiency in the case of an electron.



(b) Efficiency in the case of a muon.

Figure 5.8.: TRF b-tagging efficiency calculated in a 4 jet inclusive region with at least 1 b-jet using a WP of 85 % in dependence of the leading jet p_{T}; electrons (a) and muons (b) are considered separately.

The jet tagging efficiency is not only dependent on the jet flavour but also dependent on jet kinematics, therefore, four different parametrisations are examined:

- leading jet p_{T}
- leading jet p_{T} and leading jet η
- leading jet p_{T} and ΔR(leading jet, lepton)

- leading jet p_{T} and $\Delta\phi(\text{leading jet}, E_{\mathrm{T}}^{\mathrm{miss}})$

Figure 5.9 and Figure 5.10 show a comparison of the event yields with and without the TRF method for a WP of 85 % as a function of $H_{\mathrm{T}}^{\mathrm{had}}$, the scalar sum of p_{T} of all jets. Additional plots for the remaining WPs can be found in Appendix B.2.

Performance of the TRF method

All four TRF parametrisations show very similar results and the simplest of them, leading jet p_{T}, can be favoured. The matrix method with and without TRF has comparable results. The TRF method is able to reduce the statistical uncertainty in high b-jet multiplicity regions greatly and can also reduce spikes in the distribution, which are caused by single large weights of the matrix method and are not compensated by other events due to a lack in statistics, see Equation 5.6. A single event completely dominating the bin content is unfavourable and this effect can be reduced with TRF and the distributions are smoothed. However, this method is only tested with a fixed b-tagging WP and is not yet applicable for the complex definition of signal and control regions used in this analysis.

5.4. Signal and control regions

The main experimental challenge in this analysis is to reconstruct and separate the signal from large backgrounds, mainly from $t\bar{t}$ + jets. To cope with this challenging background, events are categorised into multiple, non-overlapping analysis regions based on the number of jets and the number of b-jets. With this procedure the high jet and b-jet multiplicity of the $t\bar{t}H(H \to b\bar{b})$ signal can be utilised, which provides a strong discriminating power. Events in the single-lepton (dilepton) channel are first split into two categories depending on whether the number of jets in the final state is exactly five (three) or at least six (four). These events are then further divided into analysis regions depending on the values of the b-tagging discriminants for the jets. This procedure ensures to create regions either enriched in $t\bar{t}H$ plus $t\bar{t} + b\bar{b}$ or $t\bar{t} + b$, $t\bar{t} + \geq 1c$, $t\bar{t}$ + light jets. The analysis regions where the $t\bar{t}H$ and $t\bar{t} + b\bar{b}$ contribution is higher relative to the other backgrounds are referred to as *signal regions* (SRs). Here, multivariate analysis techniques are applied to further separate the $t\bar{t}H$ signal from the backgrounds. The remaining analysis regions, referred to as *control regions* (CRs), provide constraints on backgrounds and systematic uncertainties; signal and background are not separated in these regions.

The single-lepton channel forms five SRs with different levels of purity for the $t\bar{t}H$ and $t\bar{t} + b\bar{b}$ components and six CRs enriched in $t\bar{t} + b$, $t\bar{t} + \geq 1c$, and $t\bar{t}$ + light jets. Subsequently, the iterative process is listed how these regions are defined.

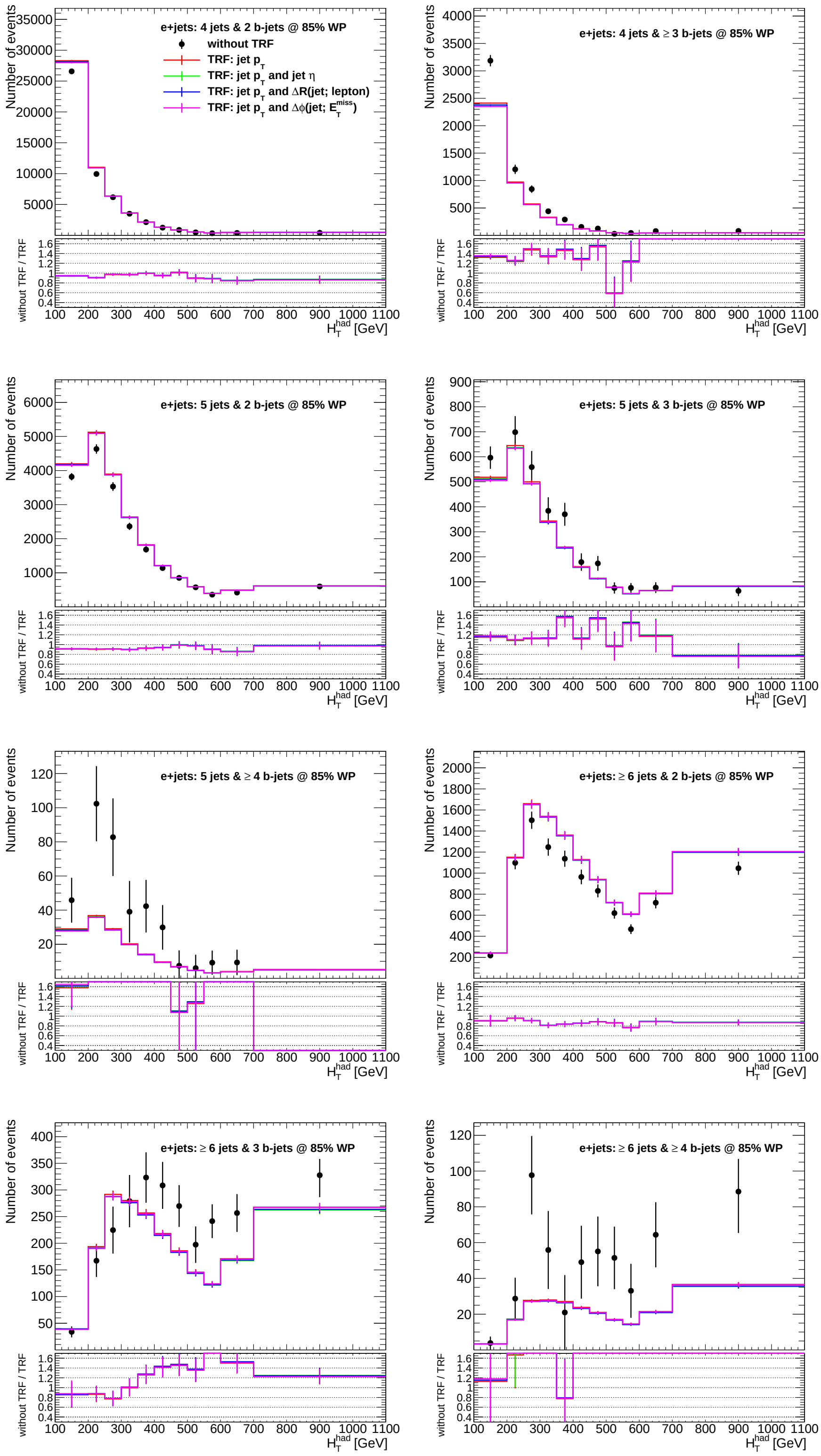


Figure 5.9.: Comparison of the event yields of the fake estimation with and without the TRF method. Selection: electron + jets with a b-tagging WP of 85 %.

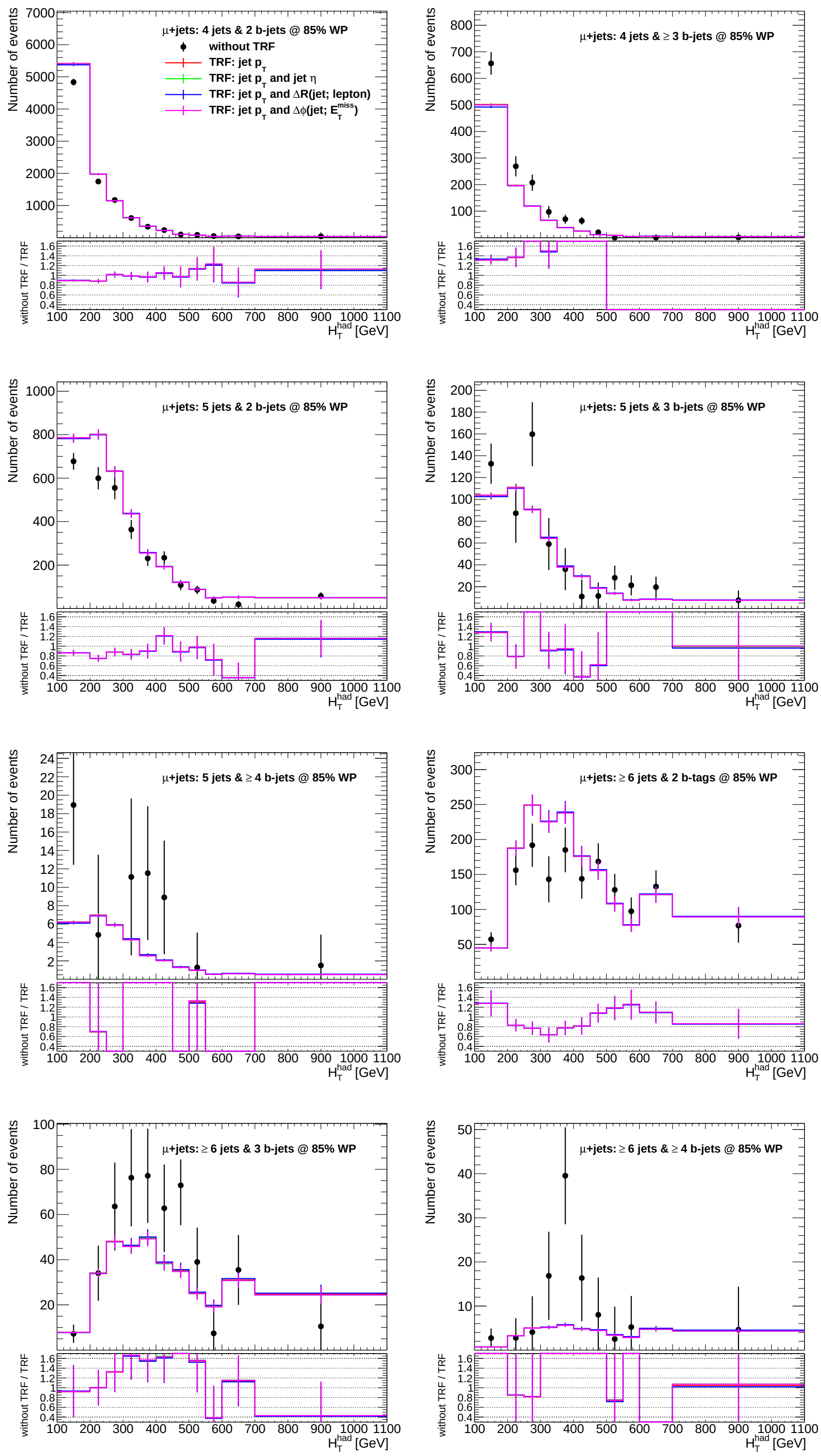


Figure 5.10.: Comparison of the event yields of the fake estimation with and without the TRF method. Selection: muon + jets with a b-tagging WP of 85 %.

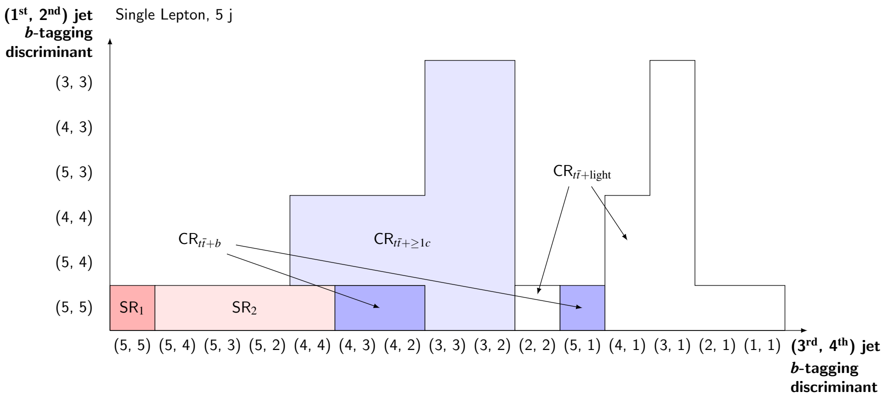
- Events with exactly 5 jets are selected.
  - SR_1^{5j}: ultra-pure SR that requires four b-jets at the very tight WP labelled $(5, 5, 5, 5)$ with a $t\bar{t} + \geq 2b$ content of at least 60 %.
  - $\mathrm{CR}_{t\bar{t}+b}^{5j}$: contains remaining events with a $t\bar{t} + b$ content of at least 20 %.
  - SR_2^{5j}: contains remaining events with a $t\bar{t} + \geq 2b$ content of at least 20 %.
  - $\mathrm{CR}_{t\bar{t}+\geq 1c}^{5j}$: contains remaining events with a $t\bar{t} + \geq 1c$ content of at least 20 %.
  - $\mathrm{CR}_{t\bar{t}+\mathrm{light}}^{5j}$: contains remaining events.
- Events with at least 6 jets are selected.
  - $\mathrm{SR}_1^{\geq 6j}$: same requirements as for the 5 jets scenario. This is the most signal-enriched region.
  - $\mathrm{SR}_2^{\geq 6j}$: contains remaining events with a $t\bar{t} + \geq 2b$ content of at least 45 %.
  - $\mathrm{SR}_3^{\geq 6j}$: contains remaining events with a $t\bar{t} + \geq 2b$ content of at least 30 %.
  - $\mathrm{CR}_{t\bar{t}+b}^{\geq 6j}$: contains remaining events with a $t\bar{t} + b$ content of at least 30 %.
  - $\mathrm{CR}_{t\bar{t}+\geq 1c}^{\geq 6j}$: contains remaining events with a $t\bar{t} + \geq 1c$ content of at least 20 %.
  - $\mathrm{CR}_{t\bar{t}+\mathrm{light}}^{\geq 6j}$: contains remaining events.

The procedure in the dilepton channel is similar and leads to three SRs and four CRs. In addition, categories with less than eight background events or a $S/\sqrt{B}$ ratio smaller than 0.08 are not considered.

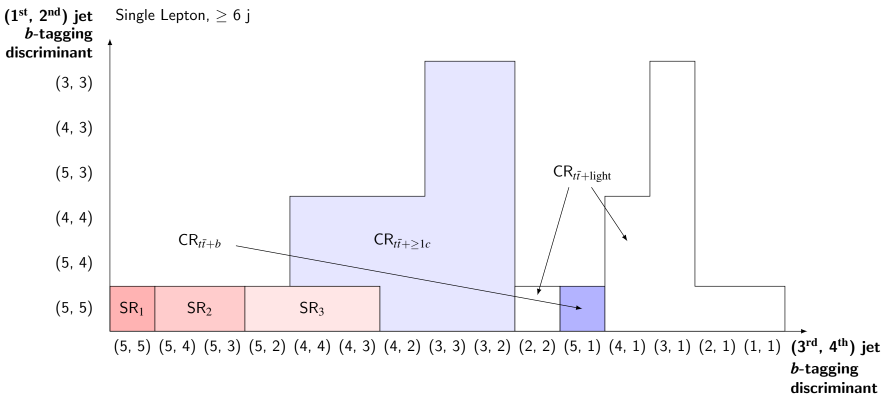
- Events with exactly 3 jets are selected.
  - $\mathrm{CR}_{t\bar{t}+\geq 1b}^{3j}$: contains events with a $t\bar{t} + \geq 1b$ content of at least 30 %.
  - $\mathrm{CR}_{t\bar{t}+\mathrm{light}}^{3j}$: contains remaining events.
- Events with at least 4 jets are selected.
  - $\mathrm{SR}_1^{\geq 4j}$: contains events with a $t\bar{t} + \geq 2b$ content of at least 70 %.
  - $\mathrm{SR}_3^{\geq 4j}$: contains remaining events with a $t\bar{t} + \geq 1b$ content of at least 30 %.
  - $\mathrm{SR}_2^{\geq 4j}$: contains remaining events with a signal content of at least 1.5 %.
  - $\mathrm{CR}_{t\bar{t}+\geq 1c}^{\geq 4j}$: contains remaining events with a $t\bar{t} + \geq 1c$ content of at least 25 %.
  - $\mathrm{CR}_{t\bar{t}+\mathrm{light}}^{\geq 4j}$: contains remaining events.

Figure 5.11 illustrates the regions for the single-lepton channel and Figure 5.12 for the dilepton channel. To avoid disconnected areas between the regions a few b-tagging categories were manually moved between regions.

The different background components and the $t\bar{t}H$ signal purity for all SRs and CRs are shown in Figures 5.13 and 5.14. 96 % of the $t\bar{t}H$ signal events in the signal regions of the single-lepton channel have a $H \to b\bar{b}$ decay, while in the dilepton channel this is the case for 89 %.

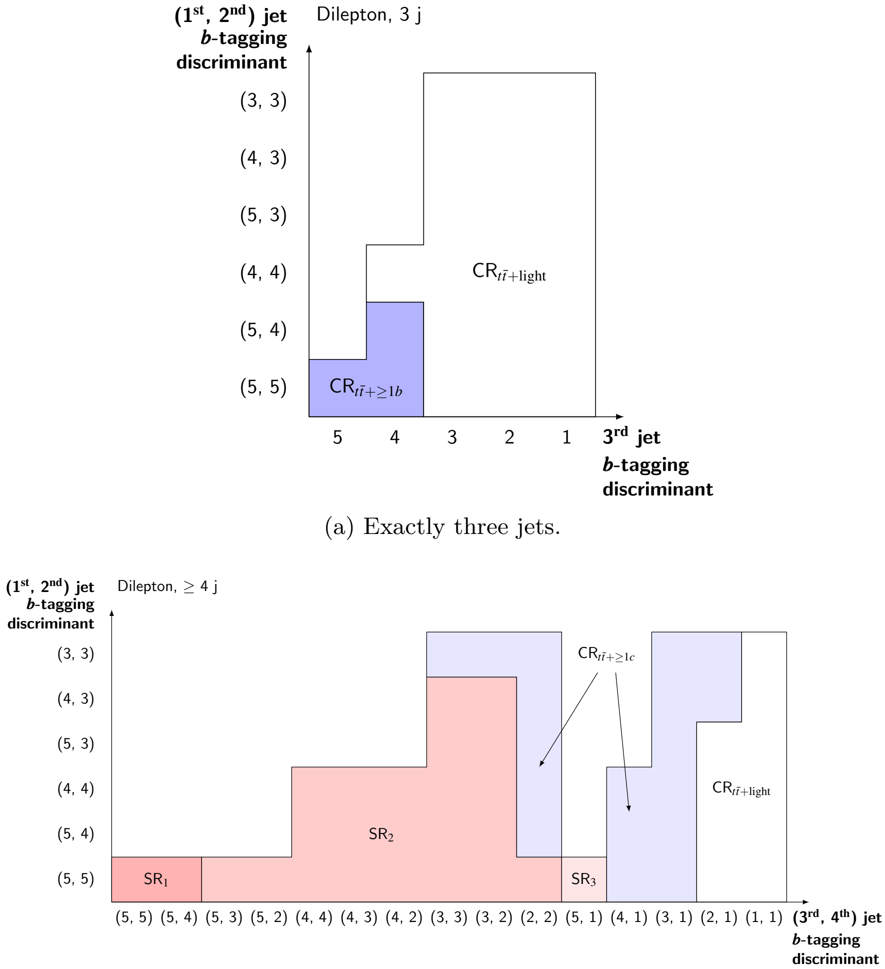


(a) Exactly five jets.



(b) At least six jets.

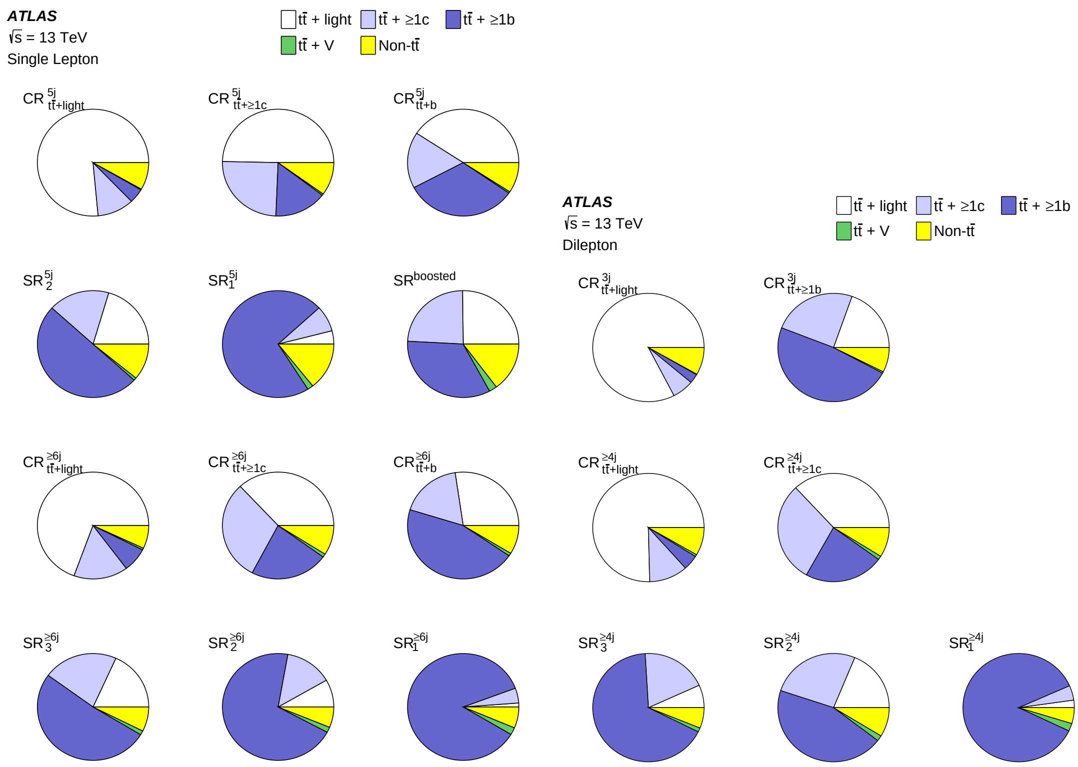
Figure 5.11.: Definition of the SRs and CRs in the single-lepton channel. The categorisation is made in dependence of the b-tagging discriminant. The vertical axis displays the values of the b-tagging discriminant assigned to the first two jets, while the horizontal axis displays these values for the third and fourth jets.



(a) Exactly three jets.

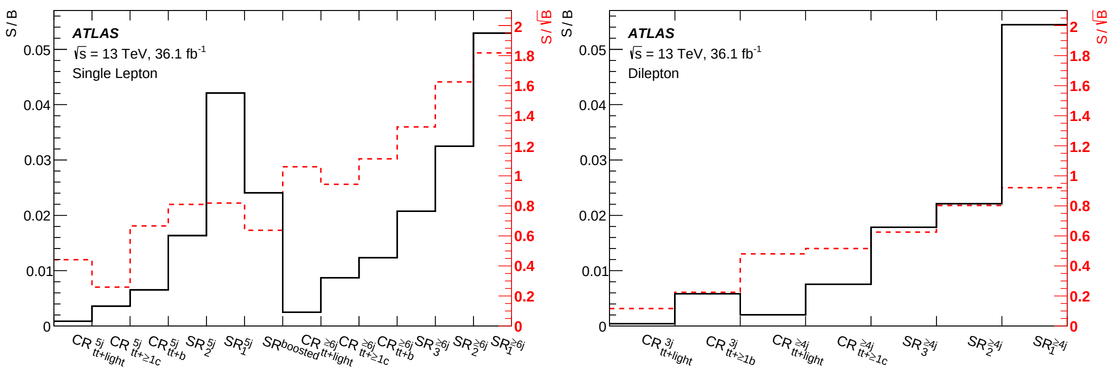
(b) At least four jets.

Figure 5.12.: Definition of the SRs and CRs in the single-lepton channel. The categorisation is made in dependence of the b-tagging discriminant. The vertical axis displays the values of the b-tagging discriminant assigned to the first two jets, while the horizontal axis displays these values for (a) the third jet or (b) the third and fourth jets.



(a) Single-lepton channel. (b) Dilepton channel.

Figure 5.13.: Background contribution in the SRs and CRs displayed in pie charts for the (a) single-lepton and (b) dilepton channel.



(a) Single-lepton channel. (b) Dilepton channel.

Figure 5.14.: S/B ratio (black solid line) and $S/\sqrt{B}$ ratio (red dashed line) displayed for each of the analysis regions (a) in the single-lepton channel and (b) in the dilepton channel.

5.5. Multivariate analysis techniques

After assigning events to analysis regions, multivariate analysis techniques are applied to the SRs of the single-lepton and dileptonic topologies in order to discriminate signal from the background. First, a *reconstruction boosted decision tree* (BDT) is used to match jets to Higgs boson or top quark decays considering the large b-jet combinatorics. In addition, a likelihood discriminant and a matrix element method are employed for the reconstruction in some of the regions. The outputs of these three methods are given as input to a *classification BDT*. Here, the events are classified as more signal- or background-like. The background in the fit is constrained by the total event yield of all CRs, with the exception of $\mathrm{CR}^{5j}_{t\bar{t}+\geq 1c}$ and $\mathrm{CR}^{\geq 6j}_{t\bar{t}+\geq 1c}$, where the $H_{\mathrm{T}}^{\mathrm{had}}$ distribution is used as input to the fit.

5.5.1. Reconstruction of the signal

Reconstruction BDT

The reconstruction BDT is built with the toolkit for multivariate analysis (TMVA) [151] and employed in all signal regions. To avoid bias, the training and evaluation are performed on statistically independent samples. This tool is trained to find the best combination of jet-parton matches to construct the Higgs boson and top quark candidates by assigning reconstructed jets to partons originating from top quarks and Higgs boson decays. For this task, W boson, top quark, and Higgs boson candidates are formed from combinations of jets and leptons. To reduce the number of combinations, b-quarks can only be paired with the four leading jets ranked by their b-tagging discriminant.

In the single-lepton channel, the W boson is, in the case of a leptonic decay, built from the lepton's and neutrino's four momenta (p_ℓ and p_ν). The neutrino four momentum is obtained from the missing transverse momentum, while the z component is derived by solving the equation $m_W^2 = (p_\ell + p_\nu)^2$, where m_W represents the W boson mass. Both solutions of this quadratic equation are used and treated as individual configurations. However, if no real solution exists, the discriminant of the quadratic equation is set to zero, giving a unique solution. In the case of a hadronic decay, the W boson is formed from a pair of jets. The latter procedure is also employed for Higgs boson candidates, whereas top quark candidates are built from one W boson candidate and one jet. In signal regions requiring exactly five jets, top quark candidates with a hadronic W boson decay are formed from one jet and one b-jet, since less than 30 % of the events contain both jets form the hadronic W boson decay. In the dilepton channel, top quark candidates are formed by one lepton and one jet and no attempt is made to build leptonic W boson decay candidates.

The training is performed with simulated $t\bar{t}H$ events by iterating over all allowed combinations to distinguish correct from incorrect matches. If additional information related to the Higgs boson is added to the kinematic input variables, the performance increases, however, this also biases the background distributions. Therefore, two versions of the reconstruction BDT are employed, one containing this additional information and one without. Both BDTs or just one are then used for the classification BDT. A full list of the input variables is given in Appendix C.1.

In the $\mathrm{SR}_1^{\geq 6j}$ region of the single-lepton channel, the Higgs boson can be correctly reconstructed in 48 % (32 %) of the selected $t\bar{t}H$ events using the reconstruction BDT with (without) the additional Higgs boson information. For the $\mathrm{SR}_1^{\geq 4j}$ region of the dilepton channel, a reconstruction efficiency of 49 % (32 %) is achieved.

Likelihood discriminant

Because of the high sensitivity in the single-lepton channel a likelihood discriminant (LHD) method is also used. This probability is computed in the same way as in Reference [152] and combines the signal and background probabilities of all possible combinations in each event, defined as

$$D = \frac{p^{\mathrm{sig}}}{p^{\mathrm{sig}} + p^{\mathrm{bkg}}}, \tag{5.11}$$

where p^{sig} gives the probability for the signal hypothesis that the event originates from the $t\bar{t}H$ signal and p^{bkg} gives the probability for the background hypothesis stated below. Hence, a probability close to one suggests a strong compatibility with the signal hypothesis, whereas a value close to zero favours the background hypothesis. The probabilities are obtained by multiplying one-dimensional probability density functions (pdfs) for the different kinematic distributions, averaged over all jet-parton assignments. These combinations are weighted according to b-tagging information to suppress incorrectly matched flavour candidates. Two likelihoods are considered with two different background hypotheses; 1) the event originates from the $t\bar{t}$ + $\geq 2b$ background, or 2) the event originates from the $t\bar{t}$ + b background. Both likelihoods are averaged and weighted according to their relative fraction of the $t\bar{t}$ + jets background. An additional signal and background hypothesis is considered to account for topologies where only one jet from the hadronic W decay is selected, which is a significant fraction for the regions requiring at least six jets.

Contrary to the reconstruction BDT, this method fully utilises all possible combinations in the event, but does not entirely account for correlations between variables within one combination, because a product of one-dimensional pdfs is used.

Matrix element method

The matrix element method is used to construct a discriminant ($\mathrm{MEM}_{\mathrm{D1}}$) similar as in Reference [131]. For this purpose, two likelihoods L_S and L_B are introduced that express how compatible an event is with the signal ($t\bar{t}H(H \to b\bar{b})$) and background ($t\bar{t}$ + $b\bar{b}$) hypotheses, respectively. Instead of using simulated MC samples as for the LHD method, each likelihood is calculated using ME calculations at the parton level. Due to its high computational intensity, this method is only applied in the signal region with the highest sensitivity $\mathrm{SR}_1^{\geq 6j}$. The likelihoods are defined as:

$$L_i = \sum \int \frac{f_1(x_1, Q^2) f_2(x_2, Q^2)}{|\vec{q}_1||\vec{q}_2|} |\mathcal{M}_i(\boldsymbol{Y})|^2 T(\boldsymbol{X}; \boldsymbol{Y}) d\boldsymbol{\Phi}_n(\boldsymbol{Y}). \tag{5.12}$$

Each likelihood contains a product of PDFs f_1 and f_2, each of them relating to a parton with momentum $\vec{q}_j$ to carry the energy fraction x_j of the proton in a collision at energy scale Q^2 (for the initial $j = 1, 2$). $\mathcal{M}_i$ denotes the LO ME calculation of either the signal or background Feynman diagrams for a phase space configuration $\boldsymbol{Y}$ at parton level. The transfer function T gives the probability that a jet measurement on reconstruction level $\boldsymbol{X}$ originates from a parton level configuration $\boldsymbol{Y}$. Only the reconstruction level information $\boldsymbol{X}$ is available and, therefore, all unknown parameters need to be integrated out over the phase space factor $d\boldsymbol{\Phi}_n$, including undetected neutrinos. Finally, a sum is performed over all different possible initial states.

The ME calculations are computed with MadGraph5_aMC@NLO at LO accuracy using the CT10 PDF set, interfaced via the LHAPDF package [153]. The transfer functions are extracted from a $t\bar{t}$ sample simulated with Powheg+Pythia 6 and validated with the nominal Powheg+Pythia 8 $t\bar{t}$ sample.

CPU time can be reduced by applying the following methods: only gluon-induced Feynman diagrams are considered; the transfer function significantly constrains the phase space by assuming δ-functions for well measured directions η and ϕ; imposing transverse momentum conservation restricts the neutrino's momentum by integrating over its z component using VEGAS [154], following Reference [155]; employing b-tagging information to reduce the number of jet-parton matches.

Finally, combining both likelihoods for signal and background leads to the powerful discriminating variable, $\mathrm{MEM}_{\mathrm{D1}}$, defined as:

$$\mathrm{MEM}_{\mathrm{D1}} = \log_{10}(L_S) - \log_{10}(L_B). \tag{5.13}$$

5.5.2. Signal and background classification

Finally, the classification BDT is employed to classify events as signal- and background-like. This BDT is also trained with TMVA and uses as input the outputs of three intermediate multivariate methods; the reconstruction BDT outcome, the likelihood discriminant's value, and the full matrix element result. In addition, information provided by general kinematic variables as well as b-tagging discriminants of the selected jets are also exploited. A full list of the input can be found in Appendix C.2. To ensure a good fit result, only variables with good modelling of data are considered.

5.5.3. Artificial neural networks

A different multivariate analysis approach referring to the use of an artificial neural network (NN), which is not included in the analysis described in this chapter, will be examined in the following section. A neural network consists of several neurones or *nodes* organised in computational layers, see Figure 5.15. The first layer (input layer) consists of multiple input variables that pass the information onto the nodes of the next layer (hidden layer). Here, the NN "learns" patterns of the given input. The final layer (output layer) can consist of multiple nodes that show the response of the NN. In the following, a binary NN is used where the output layer consists of only a single node that classifies an

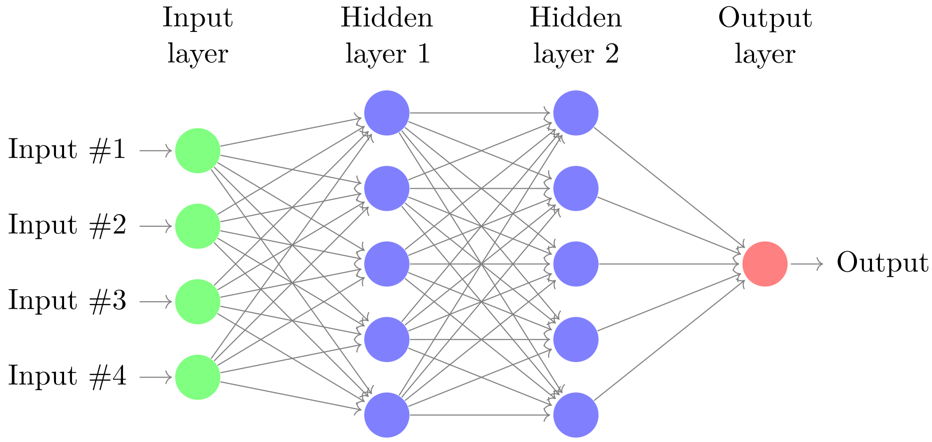


Figure 5.15.: Schematic diagram of a neural network, where the number of hidden layers depends on the architecture of the network.

event as signal- or background-like. A neural network can be designed to have multiple hidden layers. In this case, it is called a deep neural network (DNN). For less than three hidden layers the term shallow neural network or simply NN is used. The performance is quantified by the separation:

$$S = \frac{1}{2}\sum_{i}^{\text{bin}} \frac{(N_i^{\text{sig}} - N_i^{\text{bkg}})^2}{N_i^{\text{sig}} + N_i^{\text{bkg}}}, \tag{5.14}$$

where N_i^{sig} (N_i^{bkg}) is the number of signal (background) entries in each bin after histograms have been normalised to unity. Reference [156] examines the idea to employ a shallow NN instead of a boosted decision tree as a multivariate analysis tool in the single-lepton channel and a separation power of 16.9 % was achieved. In this section, the performance of DNNs under the same conditions will be studied to determine if a similar separation power can be achieved with a simpler set of input variables. The study[1] uses the single-lepton channel with the full 2015 and part of the 2016 dataset corresponding to $13.2\,\text{fb}^{-1}$.

Each link between the nodes of an NN is associated with a weight corresponding to the strength of this connection. A node takes the sum of the weighted contributions from the previous layer as input and the output of each node can be described by an activation function $f(x \cdot w + \mu)$, where x is the vector containing the node outputs of the previous layer (= the input variables for the first layer), w is the vector of the individual weights w_i, and μ denotes the bias value of a node. In the following, a sigmoid function is employed as an activation function:

$$f(z) = \frac{1}{1 + e^{-z}}. \tag{5.15}$$

[1] Raymond Han, a summer student under my supervision, was essential for the technical implementation of this idea and he provided the figures that show the performance of different NN structures.

The previously described case, where the output of a node is always passed on to the next layer is known as a feed-forward NN.

Before a neural network can be used to classify events as signal- or background-like, the architecture needs to be trained. For this, a training dataset is given to the NN and the classification error is characterised by a loss function such as cross-entropy loss, which is a logarithmic loss function. The goal is to minimise this loss function by successively adjusting the weight and bias parameters of the NN. This minimisation is achieved by propagating the error backwards through the network with a learning algorithm. One of these iterations is called *epoch* and the magnitude of this adjustment is given by the *learning rate*. In this study, the *Adam* algorithm is used for the minimisation. Adam is an adaptive learning rate optimiser with good performance for large datasets [157]. After this procedure, the NN can be employed to recognise patterns in an unknown dataset and separate the signal from backgrounds.

The neural networks are built using Keras, a deep learning library for Python based on TensorFlow, an open source machine learning platform [158, 159].

Shallow neural networks

At first, the separation power of a shallow neural network with high-level input variables will be examined. Here, the high-level input classifies variables that contain complex information about the underlying event are built from combinations of low-level object kinematic variables, such as jet p_{T}, jet angle, and jet energy. Variables with a high signal and background separation are selected to ensure a good performance of the NN. A list of 15 variables with highest separation power can be seen in Table 5.1.

Separation	Variable	Definition
7.07 %	$\Delta R_{bb}^{\mathrm{avg}}$	average ΔR for all b-tagged jet pairs
5.64 %	N_{Higgs}^{30}	number of b-jet pairs with a mass within 30 GeV of m_{Higgs}
4.75 %	$\Delta\eta_{jj}^{\max\Delta\eta}$	maximum $\Delta\eta$ between a pair of jets
4.14 %	$\Delta R_{bb}^{\max p_{\mathrm{T}}}$	ΔR between two b-jets with the largest p_{T}
3.14 %	$M_{bb}^{\min\Delta R}$	mass of the combination of two b-jets with the smallest ΔR
3.06 %	$\mathrm{Aplanarity}_{b\text{-jets}}$	1.5 times the 2$^{\mathrm{nd}}$ eigenvalue of the momentum tensor [160] built with all b-jets
3.04 %	$\mathrm{Centrality}_{\mathrm{all}}$	p_{T} sum divided by energy sum of all jets and the lepton
2.46 %	$p_{\mathrm{T}}^{\mathrm{jet5}}$	5$^{\mathrm{th}}$ leading jet p_{T}
1.77 %	$H_{\mathrm{T}}^{\mathrm{jets}}$	scalar sum of jet p_{T} in the final state
1.73 %	$p_{\mathrm{T}}^{\mathrm{jet3}}$	3$^{\mathrm{rd}}$ leading jet p_{T}
1.63 %	H_{all}^{4}	4$^{\mathrm{th}}$ Fox-Wolfram moment [161, 162] computed from all jets and the lepton
1.56 %	$\Delta R_{\mathrm{lep}\text{-}bb}^{\min\Delta R}$	ΔR between the lepton and the combination of two b-jets with smallest ΔR
1.10 %	$\Delta R_{lj}^{\min\Delta R}$	smallest ΔR between the lepton and a jet
0.90 %	M_{jj}^{Higgs}	mass of the combination of two jets closest to the Higgs mass
0.87 %	$\Delta R_{Hl}^{\min\Delta R}$	smallest ΔR between Higgs boson decay products and the lepton

Table 5.1.: Candidate variables for an NN input with their separation power among events with at least 6 jets and at least 4 b-jets.

Multivariate analysis techniques can give a greater separation than any individual variable. This might lead to the assumption that a larger set of input variables leads to an even better performance. A large set of variables comes with the caveat of an increased computation time as well as the risk of overtraining. This can happen when the NN incorrectly identifies statistical fluctuations as patterns in the dataset and bias parameters are adjusted accordingly. The network memorises details of the training dataset without gaining any predictive power. Therefore, a set of ten variables with small correlations is chosen to gain the maximum benefit from each individual variable. The correlation matrix for the 15 input variables can be seen in Figure 5.16. The redundant variables $p_{\mathrm{T}}^{\mathrm{jet3}}$ and $H_{\mathrm{T}}^{\mathrm{jet}}$ are discarded in favour of $p_{\mathrm{T}}^{\mathrm{jet5}}$ and the three variables with the lowest separation power ($\Delta R_{lj}^{\min\Delta R}$, M_{jj}^{Higgs}, $\Delta R_{Hl}^{\min\Delta R}$) are also removed. The signal and background distributions of the remaining ten variables are then shifted to have a mean of zero and rescaled to have a standard deviation of one in order to restrict inputs to the sensitive range ($[-4, 4]$) of the sigmoid function.

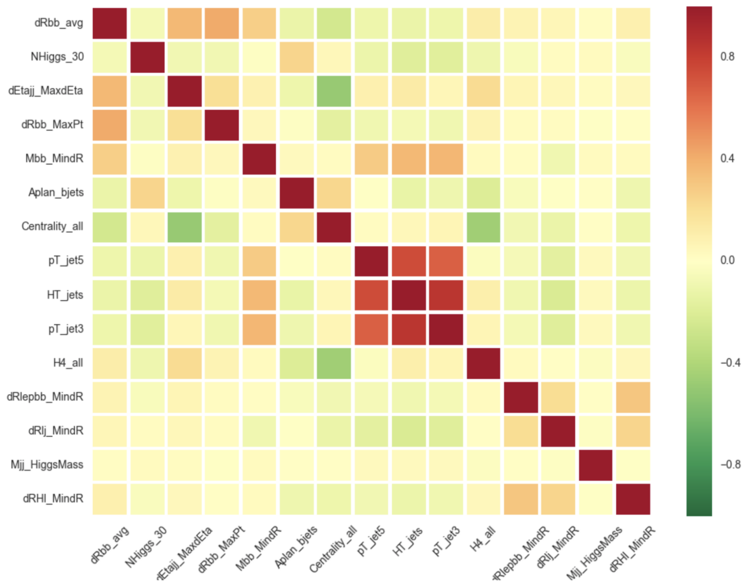
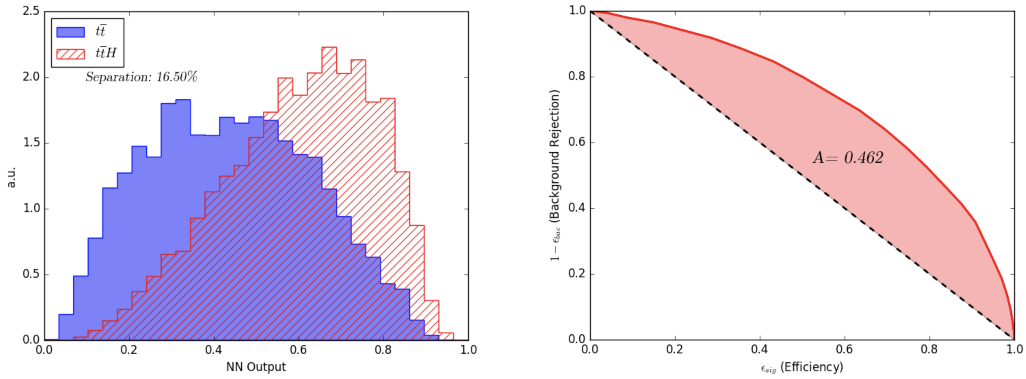


Figure 5.16.: Correlation matrix for the input variables. The scale on the right gives the strength of the correlation, where dark red corresponds to high correlation and dark green to high anti-correlation.

Finding the optimal architecture for a neural network does not follow a strict guideline. However, the neural network package NeuroBayes for ROOT suggests that $N+2$ nodes for the hidden layer of a shallow NN are sufficient when high-level variables are used [163,164]. Therefore, the network is designed to have 10 input nodes, 12 hidden nodes, and 1 output node, where the output node returns a value between 0 and 1 with the specification that values near 0 correspond to predicted background events and values near 1 represent predicted signal events.

The input dataset consists of 167 700 MC events with at least six jets and at least four b-jets and is composed of 86 % signal ($t\bar{t}H$) and 14 % background ($t\bar{t}$) events. To ensure statistical independence, the dataset is divided into two equal sets, one used for the training and the other for evaluating the performance of the neural network. First, the training set is applied to the NN and the architecture learns to classify the input in signal- and background-like events. Several tests showed that a training period over 15 epochs is sufficient and additional epochs only lead to a negligible learning rate. After that, the NN is applied to the testing sample and a separation power of 16.5 % is achieved, see Figure 5.17a. This is a significant improvement to the highest separation power of a single variable (7.07 % for $\Delta R_{bb}^{\text{avg}}$). Another characterisation of the performance of a binary neural network is given by the receiver operating characteristic (ROC) curve. This distribution relates the probability that a background event is correctly classified (background rejection) to the probability that the signal is detected (signal efficiency). The efficiency can then be expressed by twice the fraction of the area under the ROC curve (AUC) above the line of non-discrimination, shown as a dashed line in Figure 5.17b. Perfect separation gives a ROC curve extending up to the point $[1, 1]$ with an AUC of exactly one. The ROC curve of the tested NN corresponds to an AUC of 46.2 %.



(a) Normalised NN response for the test dataset including signal (red) and background (blue) events. The general shape of the distribution is in good agreement with the separation power.

(b) ROC curve corresponding to the separation power of the NN on the test dataset.

Figure 5.17.: Separation power of the neural network applied on the test dataset.

In the next step, the NN is tested to ensure that overtraining is avoided. For this purpose, the network is employed to the same training dataset it was trained on and the training separation power is compared to the test separation power obtained with the test dataset. If the separation is similar, overtraining is avoided. A significant difference is a strong indication for overtraining as the NN recognised statistical fluctuations in the training sample. The results obtained for the training sample give an AUC of 46.78 %, which is in good agreement with an AUC of 46.16 % for the testing sample, see Figure 5.18.

A difference of less than 1 % suggests the overtraining effects of the neural network are insignificant.

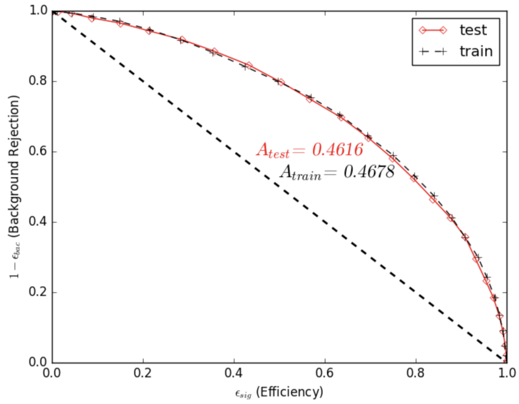


Figure 5.18.: Comparison of ROC curves for an overtraining test. The solid red (dashed black) curve corresponds to the separation power of the NN applied on the test (training) sample. Both curves are in good agreement suggesting the overtraining is insignificant.

Another important check is a two-fold validation test to ensure the NN performance is independent of the training set. For this cross-validation, the MC dataset is split again into two subsets of equal size labelled *even* and *odd*. The NN is now trained in the even sample and tested in the odd sample. Afterwards, the performance of this network is compared to the performance of a network, which is trained in the odd sample and evaluated in the even sample. Figure 5.19 displays both of the ROC curves. Again, the difference between the two AUCs is under 1 % ensuring the learning procedure of the neural network is sufficiently general.

Deep neural networks

After this initial study, the potential of deep learning will be explored. This machine learning technique is made possible by recent advances in computing power and is implemented by increasing the number of hidden layers of a neural network, see Figure 5.15.

The design of deep neural network architectures relies even more on experience and trial and error as in the case of a shallow NN. Pre-defined parameters that need to be specified before the training procedure are called *hyper-parameters*. These parameters include the activation function, weight initialisation, regularisation, number of hidden layers, and number of nodes per layer. Scenarios exist where a dedicated NN is used to compute the hyper-parameters of the desired NN. For the scope of this study, the choice of hyper-parameters follows the following arguments.

In each training cycle, the weights receive an update proportional to the partial derivative used to calculate the gradient of the activation function. In the case of a sigmoid function, which is used in the shallow NN study, this can lead to small values

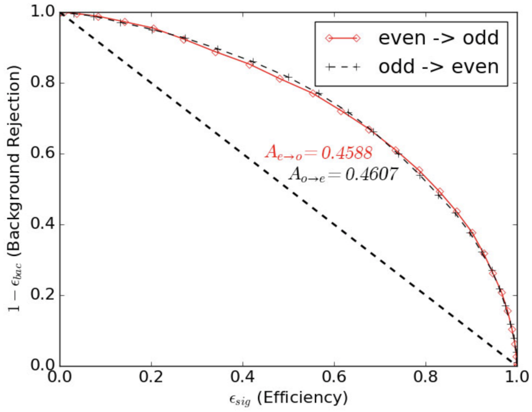


Figure 5.19.: Comparison of ROC curves for a two-fold validation test. The solid red (dashed black) line corresponds to the ROC curve obtained by training in the even (odd) and testing in the odd (even) dataset. Both curves are in good agreement suggesting the training procedure is independent of the training set.

as the gradient decreases for large positive and negative inputs. This effect increases with multiple hidden layers until the final product approaches zero, which will ultimately prevent the weights to change and leads to unresponsive layers and a poor predictive performance of the DNN. It is also known as the *vanishing gradient* problem [165] and can be avoided by using a rectified linear unit (ReLU) function, defined as

$$f(z) = max(0, z). \tag{5.16}$$

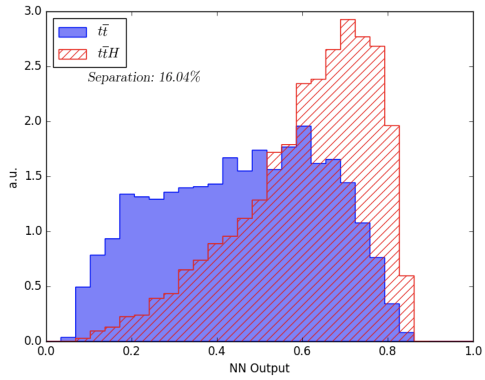
Weights are initialised as Gaussian distributions with a standard deviation of 0.05. Different non-standard initialisation procedures were studied but none improved the separation power or convergence speed of the tested DNNs.

Regularisation describes techniques used to avoid overtraining. With an increased number of nodes and hidden layers the chance of overtraining increases. Therefore, additional techniques need to be applied for a DNN. One common procedure is to implement a weight decay that penalises large weights that are not constantly reinforced by the network so that the network is encouraged to generalise more efficiently. Another method is the implementation of *dropout* nodes. For each training cycle, a random selection of nodes is temporarily deactivated and with each cycle, a slightly different neural network is used. This allows the training of a large number of different NN and the combined response corresponds to the average response of all trained variations.

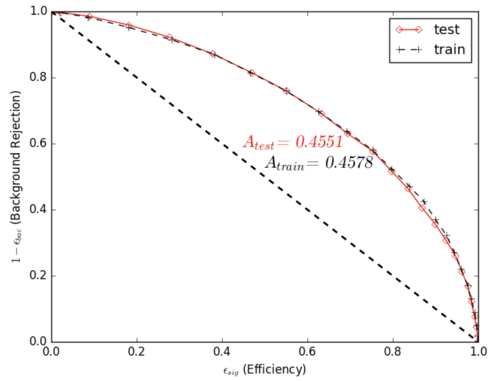
Finally, the number of hidden layers and nodes per layer needs to be determined. A general guideline is that the training of deeper neural networks requires more data and is more computationally intensive. For the following study, networks with three, four, and five hidden layers are considered. The number of nodes per layer is strongly dependent on the number of input variables. Using too few nodes significantly decreases the performance of the network while too many can lead to overtraining. Experience shows that the first hidden layer should have at least the same number of nodes as input variables are used. Further, networks with the same number of nodes for each hidden layer generally perform the same or better as DNN with a varying number of nodes per layer [166]. The effect of overtraining can be reduced by applying regularisation techniques and *early stopping* the training when performance stops improving.

Deep neural networks require a large dataset for training. Therefore, to increase the number of MC events in a high b-tagging multiplicity region, TRF is applied, see Section 4.1 for further details. With this procedure a dataset of 3.7 million MC events with at least six jets and at least four b-jets can be used. Again, training and testing samples with equal size are defined. In addition, 20 % of the training sample is used as a validation sample to enable early stopping. The maximum number of epochs is set to 50 while at least 10 epochs are required before early stopping can be applied.

High-level input variables. For a performance comparison the same high-level input variables used for the shallow NN are utilised in a DNN. The structure of this network has 10 input nodes, 3 hidden layers, each consisting of 12 nodes, and 1 output node. This leads to a separation power of 16.04 % corresponding to an AUC of 45.51 % (Figure 5.20) and no signs of overtraining are seen in Figure 5.20b.



(a) Normalised DNN response for the test dataset including signal (red) and background (blue) events. The general shape of the distribution is in good agreement with the separation power.



(b) Comparison of ROC curves for an overtraining test. The solid red (dashed black) curve corresponds to the separation power of the DNN applied on the test (training) sample. Both curves are in good agreement suggesting the overtraining is insignificant.

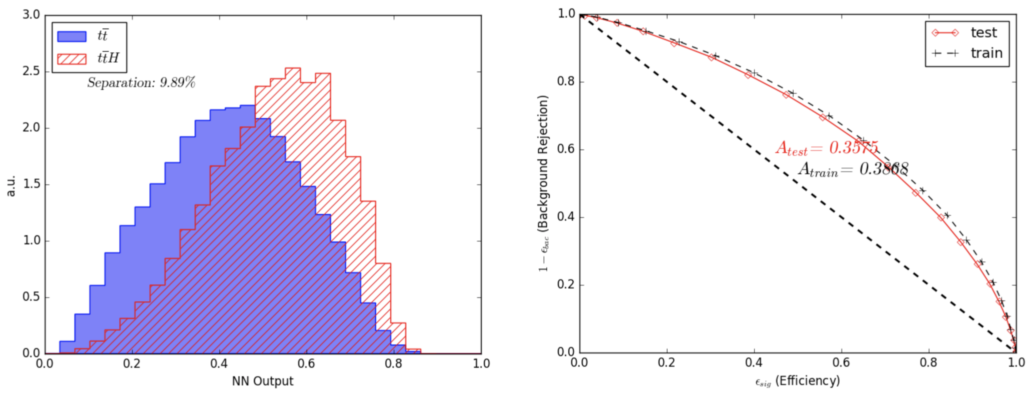
Figure 5.20.: Separation power of the deep neural network using high-level variables.

The almost identical separation power of this DNN compared to the shallow NN can be explained by the choice of input variables. Each of the high-level variables contains complex information on the underlying event and behaves differently for signal and background and the full separation potential can already be exploited by a shallow neural network. This raises the question if a DNN can be used to gain the same understanding of the underlying events without the need of high-level input variables.

Low-level input variables. In this scenario, a deep neural network is only provided with low-level input variables. The network needs to learn additional information about the underlying event, which is needed to separate the signal from background. Ideally, with the help of deep learning, the task to manually construct a set of high-level variables with large separation power can be eliminated or for an existing set of variables, missing information can be discovered.

For this study, the four-momentum vectors of each kinematic object are used as input. However, the highest jet multiplicity in the dataset has 14 jets. While including all jets for each event gives more data, this also drastically increases the dimension of the NN parameter space. Therefore, to reduce the computation time, only the six leading jets and the associated lepton are included. In addition to these 28 variables, the missing transverse momentum is also added.

To set a baseline performance, a shallow NN with one hidden layer consisting of 30 nodes is trained with this set of 29 input variables. The separation power can be seen in Figure 5.21. As expected, the network when trained on low-level variables performs



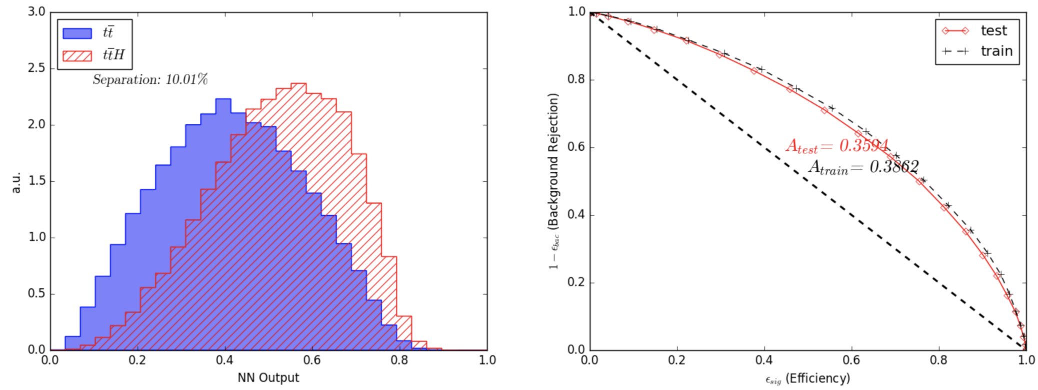
(a) Normalised shallow NN response for the test dataset including signal (red) and background (blue) events. The general shape of the distribution is in good agreement with the separation power.

(b) Comparison of ROC curves for an overtraining test. The solid red (dashed black) curve corresponds to the separation power of the shallow NN applied on the test (training) sample. The small deviation for training and test sample shows signs of overtraining.

Figure 5.21.: Separation power of a shallow neural network using low-level variables.

significantly worse compared to a network with the same structure but trained on high-level variables, this results in a 10 % lower AUC. In addition, with a greater number of input variables, effects of overtraining lead to a 3 % larger AUC when the network is evaluated on the training sample. Because the purpose of this test is to simply establish a baseline performance, no further tuning is performed. With this result, the performance of a DNN using only the set of low-level input variables will be classified.

The following DNN is designed to have 3 hidden layers with 10 nodes each and no regularisation techniques are applied to the network. The result can be seen in Figure 5.22. Unfortunately, the performance is again significantly worse compared to training on high-level variables. Only a separation of 10 % is achieved. Additionally, the deviation of the two ROC curves for the overtraining test shows signs of overtraining. Fortunately, this effect does not increase compared to the shallow NN using the same set of low-level input variables. A comparison between this DNN and the previously used shallow NN trained on high-level or low-level variables can be seen in Figure 5.23. The performance of the network strongly depends on the choice of input variables. Simply adding hidden layers to a neural network does not lead to a better understanding of the underlying events.



(a) Normalised DNN response for the test dataset including signal (red) and background (blue) events. The general shape of the distribution is in good agreement with the separation power.

(b) Comparison of ROC curves for an overtraining test. The solid red (dashed black) curve corresponds to the separation power of the DNN applied on the test (training) sample. The small deviation for training and test sample shows signs of overtraining.

Figure 5.22.: Separation power of a deep neural network using low-level variables.

To gain more insight into the potential of deep learning, a structured hyper-parameter grid search is conducted. Hereby, the following parameters are used:

- Number of hidden layers: 3, 4, 5
- Nodes per layer: 30, 60
- Regularisation: none, 20 % dropout in each layer

Each of these twelve neural network architectures is trained and their performance is evaluated. Figure 5.24 displays the results for a subset of these. A table containing the separation power and AUC for all tested networks can be found in Appendix D.1. All network configurations show a similar performance. The best separation is achieved

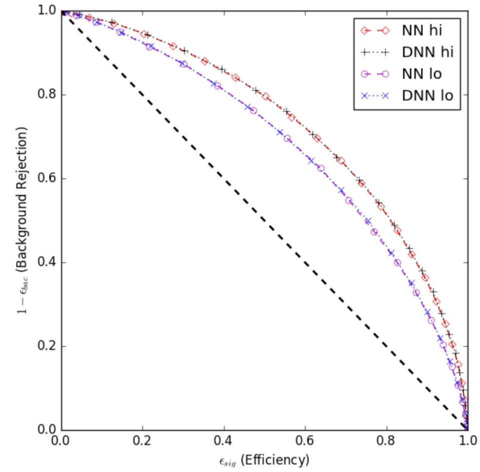


Figure 5.23.: Comparison of ROC curves for a single (shallow) and three layered (deep) NN using low- (lo) and high-level (hi) variables.

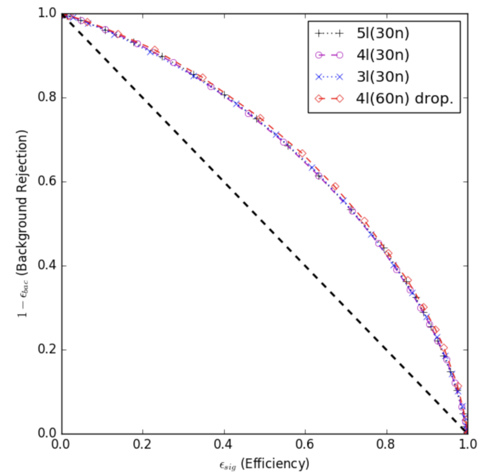


Figure 5.24.: ROC curves of several DNN configurations. ‘l’ referrs to the number of hidden layers, ‘n’ to the number of nodes, and ‘drop.’ indicates that a 20 % dropout in each layer is used as regularisation technique. The best performance with an AUC of 36.0 % is achieved with four hidden layers each containing 60 nodes and a 20 % dropout.

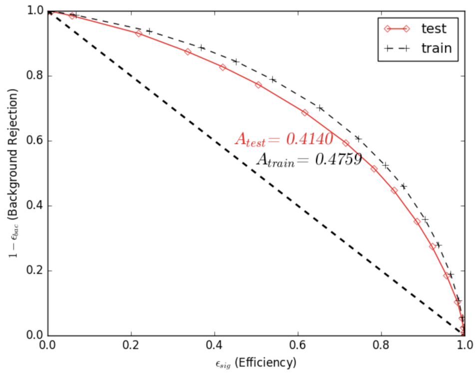
with 4 hidden layers each using 60 nodes and a 20 % dropout. Overtraining effects increase significantly, however, employing regularisation techniques such as dropout helps to decrease this effect. Nevertheless, none of the networks could achieve a similar performance as an NN trained on high-level variables. One explanation for this poor performance may be that the tested DNNs cannot learn complex information provided by the set of high-level variables. Another factor is that the high-level variables include additional information such as reconstructed secondary vertex information, which is not contained in the low-level variables. This leads to the question if a DNN can perform significantly better when additional key information is added as input.

Low-level input variables and additional information. One of the central points of this analysis is the identification of b-jets and using this information to separate signal from background. Two different algorithms exist for b-tagging; *MV2c10* where a boosted decision tree is employed, and *DL1* where a deep neural network is utilised. Both methods use dedicated multivariate analysis techniques. In the following, b-tagging information is added for each jet in the form of a TRF b-tagging weight. This increases the number of input variables to 35 and the performance of the DNNs is examined.

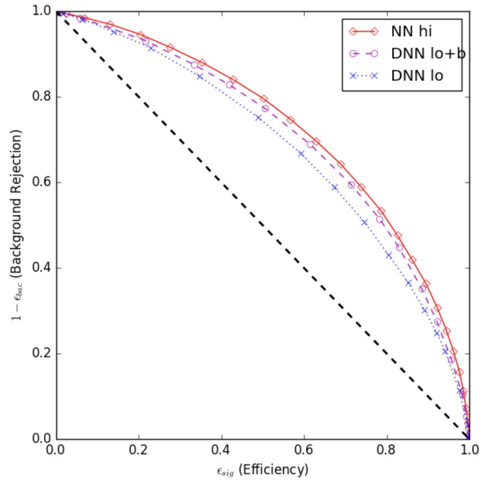
The network configurations use either 4 or 5 hidden layers with 60 or 90 nodes per layer. As regularisation a 20 %, 40 %, or 60 % dropout is applied. These additional dropout levels were added because of the good dropout performance seen in the previous study. However, configurations such as 60 nodes with a 60 % dropout resulting in only 24 active nodes per training epoch, which is lower than the number of input variables and, therefore, were not tested. A table containing the performance of all tested configurations can be found in Appendix D.2. The optimal architecture is found to have 4 layers of 60 nodes each with a 40 % dropout resulting in a separation power of 13.24 %, see Figure 5.25a for the ROC curve. Although the performance increases, it still remains poor compared to networks using a high-level input. A comparison of ROC curves for different sets of variables can be seen in Figure 5.25b. Overtraining could be reduced while performance is enhanced with a 40 % dropout. A 60 % dropout further lowered overtraining but negatively impacted the performance.

Finally, DNNs using a combination of low-level variables, TRF b-tagging weights, and the set of ten high-level variables are also studied. For this complete set consisting of 45 variables the provided amount of training data is not sufficient and, therefore, the results need to be interpreted with caution.

The DNNs have again 4 or 5 hidden layers with 60 or 90 nodes per layer. Despite increasing the dropout, overtraining remains a significant problem. Further, the large number of deactivated nodes in each training epoch can lead to an increased number of nodes becoming stuck in the negative region of the ReLu function where the gradient is zero. A proposed solution for this is the introduction of Leaky ReLu functions. Here, the negative flat region is replaced with a small negative gradient. With this change, the performance could be increased while overtraining effects could be reduced in networks with a 40 % dropout. The performance of all tested networks can be seen in Appendix D.3. With a separation power of 19.09 %, the optimal tested configuration



(a) ROC curves for a DNN in the best configuration using jet b-tag weights in addition to low-level variables.

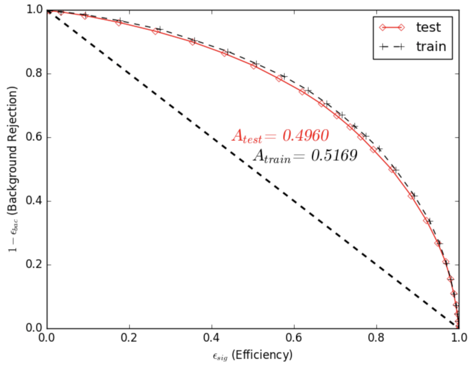


(b) Comparison of variable set performance. The addition of TRF b-tagging weight information (DNN lo+b) recovers some of the performance lost when changing from high-level variables (hi) to low-level variables (lo).

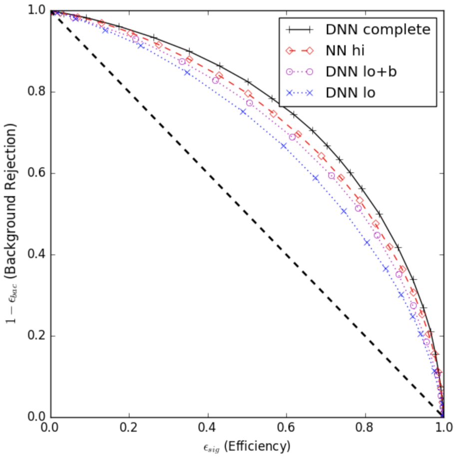
Figure 5.25.: Comparison of performance for different sets of variables including TRF b-tagging weights.

utilising the complete variable set has 5 hidden layers, 90 nodes per layer, and uses a 40 % dropout with Leaky ReLu activation functions, see Figure 5.26a for the ROC curve and Figure 5.26b for a comparison of all tested input variable sets. This network could improve the separation by an additional 2.6 % compared to a shallow NN using only high-level variables. Switching to Leaky ReLu activation functions resulted in significant better overtraining margins.

Conclusions. It is found that the choice of input variables influences the performance of the network the most, see Figure 5.26b. DNNs trained only with low-level variables performed significantly worse than any network trained with the set of high-level variables. Indeed, finding such sets needs a lot of studies on its own. An important result is that a DNN with low-level input variables and b-tagging information without any additional information could provide a fair separation power. Ultimately, a DNN with the complete set of variables showed the best performance but such a large set of variables increases the complexity of the analysis. Machine learning remains a powerful tool and should be exploited further in the future.



(a) ROC curves for a DNN in the best configuration using the complete variable set.



(b) Comparison of performance using a variety of variable sets. A DNN using the complete variable set outperforms all other networks.

Figure 5.26.: Comparison of performance for different sets of variables including a complete set.

5.6. Systematic uncertainties

Various sources of systematic uncertainty need to be considered for the $t\bar{t}H(H \to b\bar{b})$ measurement. These sources can be grouped in two main categories; experimental uncertainties originating from imperfect ATLAS and LHC measurements, and modelling uncertainties affecting the normalisation of the samples and the shapes of the distributions. In the following, a description of the sources of systematic uncertainty will be given together with their impact on the signal strength $\mu = \sigma/\sigma_{\mathrm{SM}}$, defined as the ratio of the measured cross-section with respect to the predicted cross-section of the SM. Hereby, a single independent nuisance parameter is assigned to each source of systematic uncertainty. Pruning and smoothing is applied to certain uncertainties to ensure convergence and a stable result of the profile likelihood fit, which receives the nuisance parameters as input.

5.6.1. Experimental uncertainties

Integrated luminosity. The analysed data in this search corresponds to the integrated luminosity collected by the ATLAS detector in 2015 and 2016. This quantity is derived from the instantaneous luminosity, which is determined by various detector components within ATLAS and at the LHC [73–75, 167]. For the combined 2015 and 2016 integrated luminosity measurement an uncertainty of 2.1 % was achieved.

Charged leptons. Systematic uncertainties arise from the trigger system, as well as the reconstruction, identification, and isolation efficiencies for electrons and muons. In addition, the lepton momentum scale and resolution needs to be accounted for. All these sources are measured in data using leptons originating from $Z \to \ell\ell$, $J/\Psi \to \ell\ell$, and $W \to e\nu_e$ decays [92–94]. In total, 24 independent sources are considered, but have only a minor impact on the analysis.

Jets. A major contribution comes from the jet energy scale and resolution, where the latter consists of two independent components. The jet energy scale is determined from test-beam data, LHC collision data, as well as simulation data [100], which leads to eight independent parameters. With additional uncertainties related to jet flavour, pile-up, η- and p_{T}-dependence, a total of 20 parameters is reached. The uncertainties per jet are rather small, varying from 1 % up to 6 %, however, due to a large number of jets in the final state the total contribution strongly increases. The efficiency of the JVT requirement to remove jets from pile-up is also considered.

Flavour tagging. Three different flavour categories are considered. The b-jet efficiency is measured in dileptonic $t\bar{t}$ events, while the c-jet efficiency is determined from semi-leptonic $t\bar{t}$ events where one of the W boson decays into a c-jet [168]. For light jets, the efficiency is extracted from QCD multijet events originating from secondary vertices and tracks that have an impact parameter implying a negative lifetime [104]. These measurements depend on the jet p_{T} and the different WPs, which leads to uncertainties of 2 % to 10 % for b-jets, 5 % to 20 % for b-jets, and 10 % to 50 % for light jets. These uncertainties are expressed in 30, 15, and 80 independent parameters, respectively. Jets arising from τ_{had} are treated as c-jets for determining the systematic uncertainties.

Missing transverse energy. All energy scale and resolution related uncertainties are considered for the calculation of the missing transverse momentum. Also included are three sources of systematic uncertainty related to the soft term, which is needed for the reconstruction of this quantity.

5.6.2. Modelling uncertainties

Modelling of the $t\bar{t}H$ signal. The theoretical cross-section of this process employs a $^{+5.8\,\%}_{-9.2\,\%}$ QCD scale and ±3.6 % PDF+α_S uncertainty [65,116–120]. Also included is a 2.2 % theory uncertainty related to the $b\bar{b}$ BR of the Higgs boson [65]. Modelling uncertainties related to the choice of parton shower and hadronisation model are estimated by comparing the nominal sample to MadGraph5_aMC@NLO interfaced to Herwig++.

Modelling of the $t\bar{t}$ + jets background. This background has the largest contribution of all backgrounds and a special treatment is required. All sources of systematic uncertainty are listed in Table 5.2. The $t\bar{t}$ + jets background is normalised to the theoretical NNLO+NNL cross-section and a 6 % uncertainty is assumed to account for variations of scales, the PDF, and the top quark mass [122]. Since the different flavour components are affected by different types of uncertainties (e.g. the flavour scheme used for the PDF), each of them are assigned independent nuisance parameters (except for the inclusive cross-section). For a shape comparison between the nominal sample and alternative samples, all alternative samples are reweighted to include the same fractions of $t\bar{t}$ + $\geq$1c and $t\bar{t}$ + $\geq$1b as the nominal sample. If not stated otherwise, all $t\bar{t}$ + $\geq$1b subcategories are scaled to agree with the predictions of SHERPA4F. The normalisations of $t\bar{t}$ + $\geq$1c and $t\bar{t}$ + $\geq$1b can float freely in the fit.

Modelling uncertainties associated with the choice of the event generator and the parton shower and hadronisation model are estimated by comparing simulations from POWHEG+PYTHIA 8 with SHERPA and with POWHEG interfaced with HERWIG 7 [169]. This procedure ensures a simultaneous variation of the event generator and the parton shower and hadronisation model or varying just the parton shower and hadronisation model. For this test, SHERPA version 2.2.1 with the ME+PS@NLO configuration, interfaced with OPENLOOPS and the NNPDF3.0NNLO PDF set is used. This setup is able to simulate one additional parton at NLO and four additional partons at LO accuracy by employing a five flavour (5F) scheme in the PDF. In contrast to the four flavour scheme, which considers the b-quark mass, the five flavour scheme treats the b-quark as massless, therefore, the sample will be referred to as SHERPA5F. Additionally, initial- and final-state radiation (ISR/FSR) modelling discrepancies are simulated with alternative POWHEG+PYTHIA 8 samples [170]. For these samples, no scaling is applied to the nominal value. All modelling uncertainties relate to three independent sources for each of the $t\bar{t}$ + jets flavour components.

For the $t\bar{t}$ + $\geq$1c background modelling, a $t\bar{t}c\bar{c}$ sample is generated with MADGRAPH5_aMC@NLO interfaced to HERWIG++ using a three flavour (3F) scheme (including massive c-quarks) for the PDFs, analogously to Reference [171]. The difference between this 3F sample and an inclusive $t\bar{t}$ 5F sample is added as an additional uncertainty.

The $t\bar{t}$ + $\geq$1b process includes modelling differences between 5F predictions (POWHEG+PYTHIA 8) and 4F predictions (SHERPA4F).

The subcategories $t\bar{t}$ + b, $t\bar{t}$ + $b\bar{b}$, $t\bar{t}$ + B and $t\bar{t}$ + $\geq$3b all depend on SHERPA4F predictions and are therefore not affected by the uncertainties described above. To evaluate sources for these subcategories, multiple variations of the renormalisation scale are examined; scaling by a factor, changing the functional form, and adopting a global scale. Additionally, two alternative PDF sets are considered [172] as well as an alternative set of tuned parameters for the underlying event. An extra 50 % normalisation uncertainty is added for the $t\bar{t}$ + $\geq$3b process in order to reflect the large discrepancy between the 4F prediction and 5F predictions, see Figure 5.1.

Another 50 % normalisation uncertainty is incorporated into the $t\bar{t}$ + b (MPI/FSR) sample to account for MPI contributions. The shape uncertainty of this subcategory is already considered by the comparison of the nominal sample to alternative ones as described above.

This leads to additional 20 independent sources of modelling uncertainties, of which 13 are related to $t\bar{t}$ + $\geq$1b, 4 to $t\bar{t}$ + $\geq$1c, and 3 to $t\bar{t}$ + light jets.

Systematic source	Description	$t\bar{t}$ categories
$t\bar{t}$ cross-section	up or down by 6 %	all, correlated
$k(t\bar{t} + \geq 1c)$	free-floating $t\bar{t} + \geq 1c$ normalisation	$t\bar{t} + \geq 1c$
$k(t\bar{t} + \geq 1b)$	free-floating $t\bar{t} + \geq 1b$ normalisation	$t\bar{t} + \geq 1b$
Sherpa5F vs. nominal	related to the choice of the event generator	all, uncorrelated
parton shower & hadronisation	Powheg+Herwig 7 vs. Powheg+Pythia 8	all, uncorrelated
ISR/FSR	variations of μ_R, μ_F, h_{damp}, and A14 Var3c parameters	all, uncorrelated
$t\bar{t} + \geq 1c$ ME vs. inclusive	MadGraph5_aMC@NLO+Herwig++: ME prediction (3F) vs. incl. (5F)	$t\bar{t} + \geq 1c$
$t\bar{t} + \geq 1b$ Sherpa4F vs. nominal	comparison of $t\bar{t}b\bar{b}$ NLO (4F) vs. Powheg+Pythia 8 (5F)	$t\bar{t} + \geq 1b$
$t\bar{t} + \geq 1b$ renorm. scale	up or down by a factor of two	$t\bar{t} + \geq 1b$
$t\bar{t} + \geq 1b$ resumm. scale	change μ_Q from $H_T/2$ to μ_{CMMP}	$t\bar{t} + \geq 1b$
$t\bar{t} + \geq 1b$ global scale	set $\mu_Q = \mu_R = \mu_F \equiv \mu_{\text{CMMP}}$	$t\bar{t} + \geq 1b$
$t\bar{t} + \geq 1b$ shower recoil scheme	alternative model scheme	$t\bar{t} + \geq 1b$
$t\bar{t} + \geq 1b$ PDF (MSTW)	MSTW2008NLO vs. CT10	$t\bar{t} + \geq 1b$
$t\bar{t} + \geq 1b$ PDF (NNPDF)	NNPDF2.3NLO vs. CT10	$t\bar{t} + \geq 1b$
$t\bar{t} + \geq 1b$ underlying event	alternative tune for the underlying event	$t\bar{t} + \geq 1b$
$t\bar{t} + \geq 1b$ multiple parton int.	up or down by 50 %	$t\bar{t} + \geq 1b$
$t\bar{t} + \geq 3b$ normalisation	up or down by 50 %	$t\bar{t} + \geq 1b$

Table 5.2.: Summary of the sources of systematic uncertainty for $t\bar{t}$ + jets modelling. If a systematic source effects more than one $t\bar{t}$ category, the last column indicates whether a correlation is considered or not.

Modelling of the W/Z + jets background. Varying factorisation and renormalisation scales and matching parameters in the Sherpa computation leads to an uncertainty of 35 % for Z + jets and 40 % for W + jets, in which another 30 % is added to the heavy flavour component.

Single top modelling. Each of the Wt, t-channel, and s-channel cross-sections are considered with an uncertainty of $^{+5\,\%}_{-4\,\%}$ [141–143]. For the Wt and t-channel production modes, modelling uncertainties related to the choice of the parton shower and hadronisation model are assessed by comparing the modes to alternative MC simulations. In addition, the interference between Wt and $t\bar{t}$ is accounted [120].

Diboson modelling. To include uncertainties related to the inclusive cross-section and additional jet production, a 50 % normalisation uncertainty is assumed [148].

$t\bar{t} + W/Z$ modelling. The theoretical NLO cross-section prediction has a 15 % uncertainty [173]. To assess modelling uncertainties caused by the choice of MC generator, the nominal sample is compared to alternative ones simulated by Sherpa. No correlations between $t\bar{t} + W$ and $t\bar{t} + Z$ are considered.

Fake estimation. In the single-lepton channel, a 50 % uncertainty is applied to the estimated number of fake leptons. The tag-and-probe method used in the matrix method to estimate the real efficiency introduces a bias related to the detector geometry. If one of the two leptons from the Z decay passes through an acceptance gap of the detector, the other successfully reconstructed lepton is not considered in the tag-and-probe method and the whole event is rejected for the real efficiency calculation. The bias related to the tag-and-probe method was estimated in Reference [174] in MC simulations by randomly removing leptons from the events according to a Gaussian probability distribution corresponding to the lepton reconstruction efficiency uncertainty and was found to be small. This effect is accounted for in the large uncertainty applied to the single-lepton channel. No correlation between events containing an electron or a muon as well as between events in analysis regions containing exactly five jets or regions requiring at least six jets is assumed. Conversely, in the dilepton channel correlation across lepton flavours and all analysis regions is considered and a 25 % uncertainty is assumed.

Modelling of rare processes. For the $t\bar{t}t\bar{t}$ process, a normalisation uncertainty of 50 % is assumed. In addition, PDF and scale uncertainties for tZ, $t\bar{t}WW$, $tHjb$, WtH and tWZ are also considered.

5.7. Results

The signal strength μ can be determined by constructing a binned likelihood function, $\mathcal{L}(\mu, \boldsymbol{\theta})$, which is the product of Poisson probability terms over all bins in each distribution of the analysis regions. The uncertainties are characterised by $\boldsymbol{\theta}$, which contains the set of systematic nuisance parameters as Gaussian, log-normal, or Poisson prior, and the two free floating normalisation factors for the $t\bar{t} + \geq 1c$ and $t\bar{t} + \geq 1b$ background contributions. The statistical uncertainty of the prediction comprises the statistical uncertainty of the MC events and the data-driven fake estimation. This uncertainty is added in the form of one additional nuisance parameter for each bin of the distributions. Finally, the likelihood function is maximised in a combined profile likelihood fit simultaneously performed to data in all 19 analysis regions. In addition, the probability that the measurement is compatible with the background-only hypothesis is evaluated and upper limits are obtained using the $\mathrm{CL_s}$ method [175–177].

The number of observed events compared to the prediction in each of the analysis regions can be seen in Figure 5.27 before the fit to data (pre-fit) and after the fit to data (post-fit) for the signal-plus-background hypothesis.

As described in Section 5.5, CRs are used to constrain backgrounds in the fit. H_T distributions for the $t\bar{t} + \geq 1c$ enriched CRs in the single-lepton channel are displayed in Figure 5.28, both before and after the fit. The outputs of the classification BDTs are depicted in Figure 5.29 for the five-jet SRs, and in Figure 5.30 for the six-jet SRs of the single-lepton channel, and in Figure 5.31 for the SRs of the dilepton channel, also before and after the fit. All distributions are well modelled pre-fit within the uncertainties. Subsequently, the profile likelihood fit adjusts the nuisance parameters accordingly and, therefore, the agreement between data and prediction is improved post-fit. In addition, the post-fit uncertainty is significantly reduced by these nuisance-parameter constraints and the correlations generated by the fit.

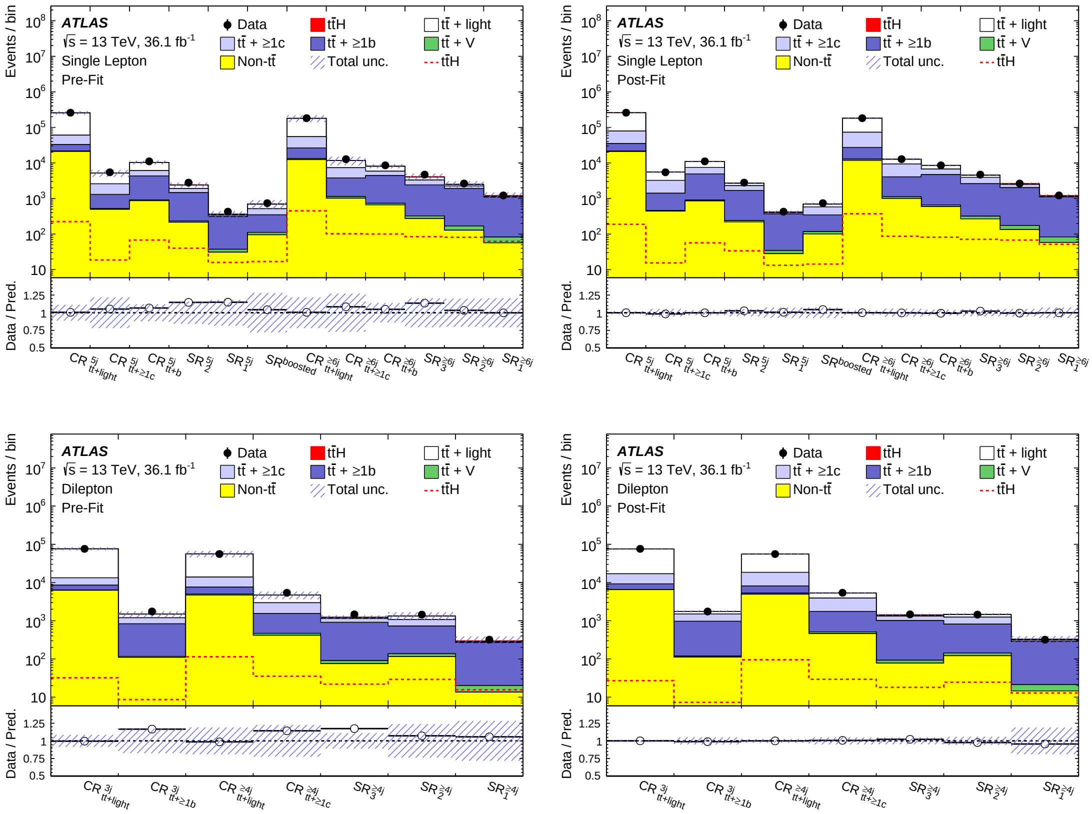


Figure 5.27.: Comparison of predicted and observed event yields in all 19 regions before (left) and after (right) the fit to the data for the single-lepton (top) and dilepton (bottom) channels.

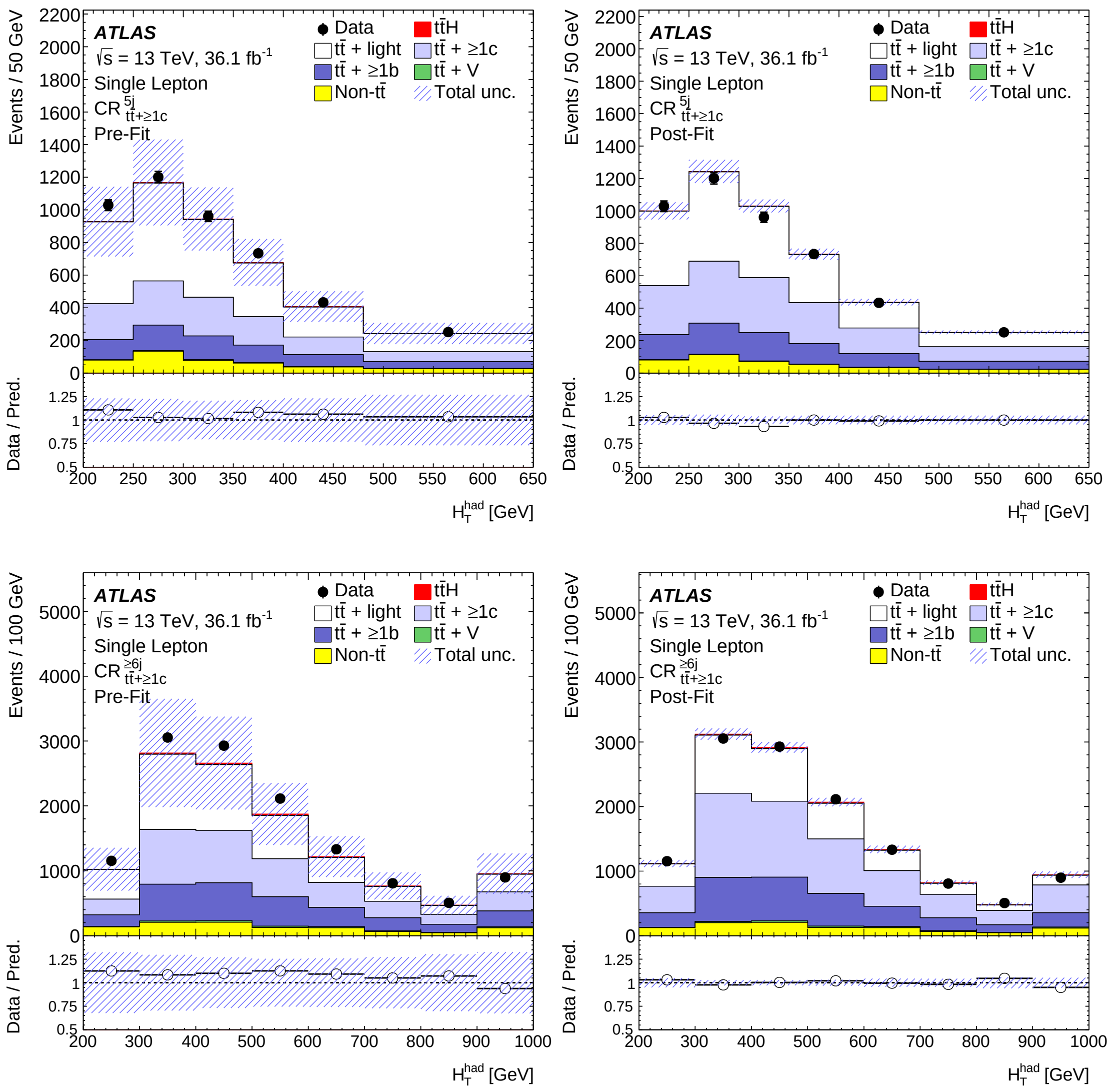


Figure 5.28.: Comparison between data and prediction for the H_T^{had} distributions in $t\bar{t} + {\geq}1c$ enriched CRs in the single-lepton channel before and after the fit.

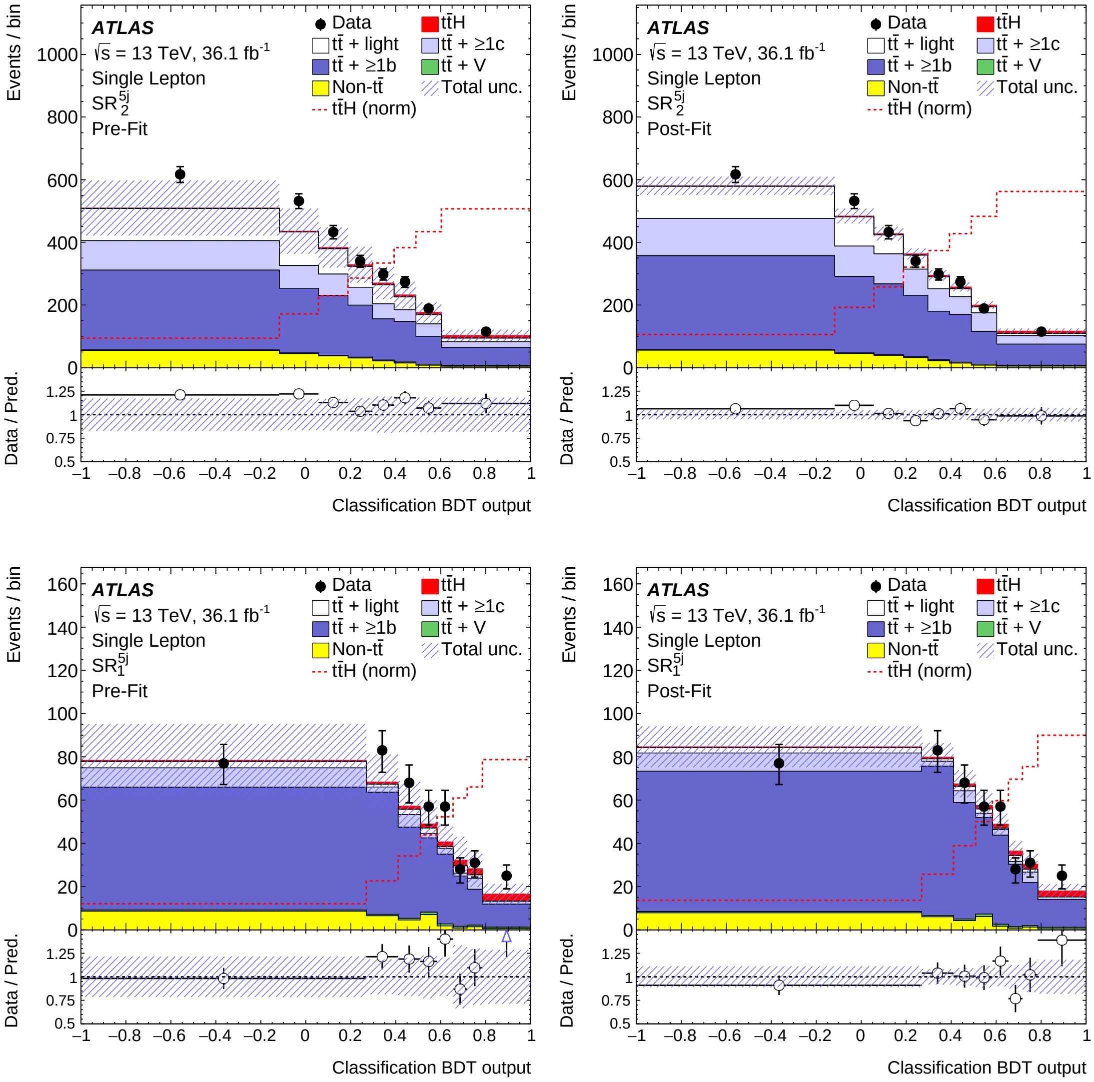


Figure 5.29.: Comparison between data and prediction for the BDT discriminant in the single-lepton channel five-jet SRs before and after the fit.

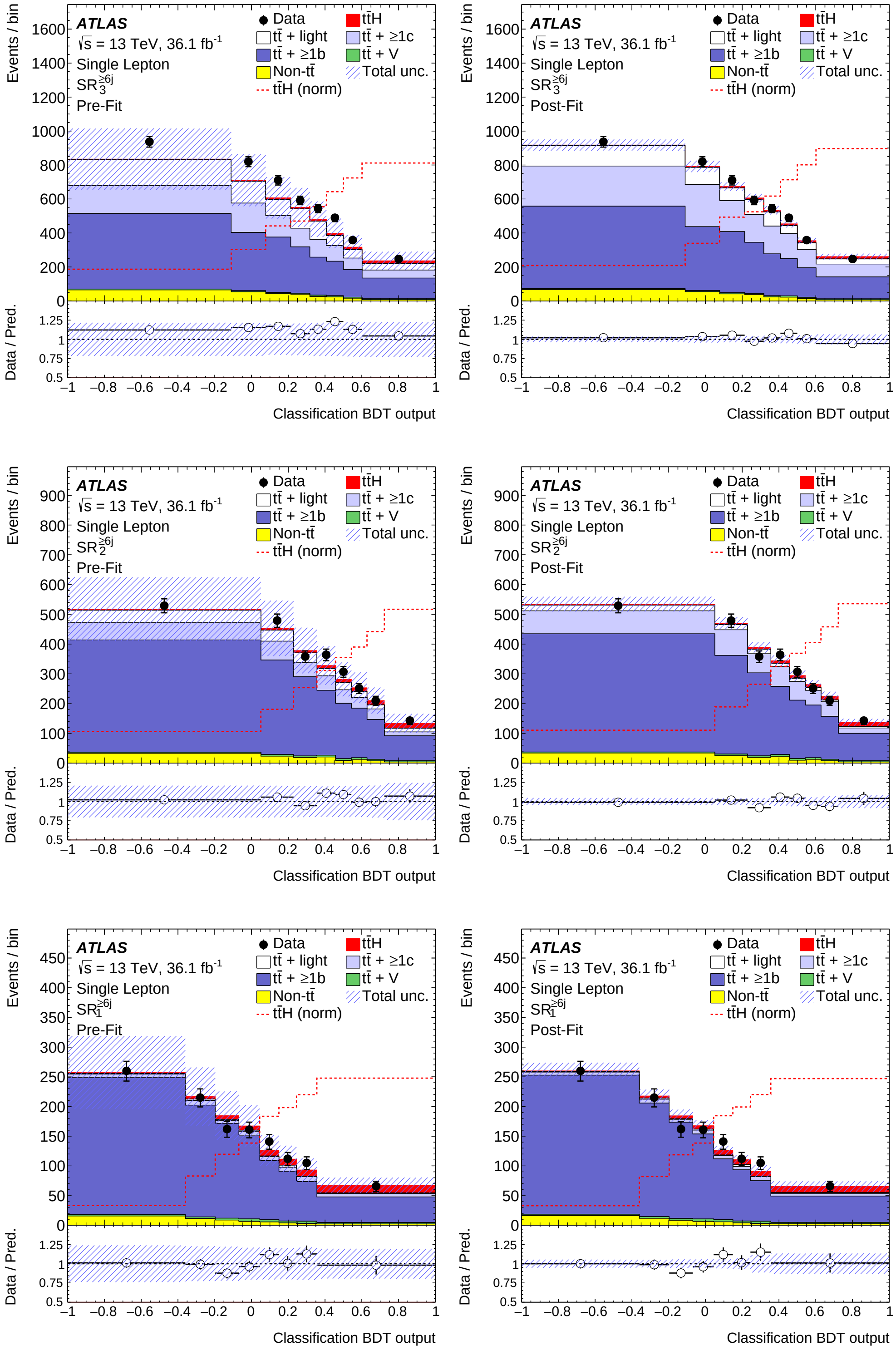


Figure 5.30.: Comparison between data and prediction for the BDT discriminant in the single-lepton channel six-jet SRs before and after the fit.

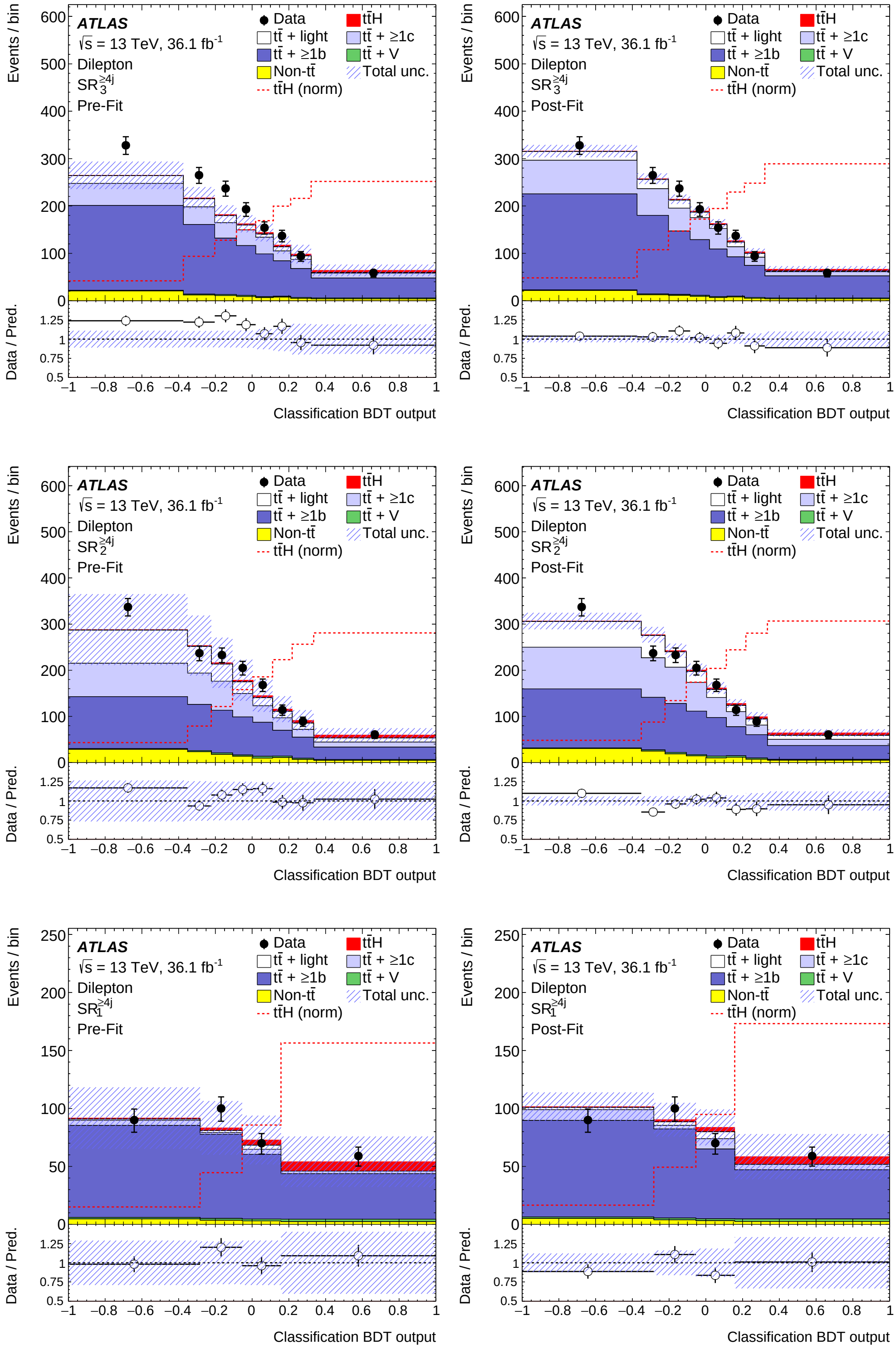


Figure 5.31.: Comparison between data and prediction for the BDT discriminant in the dilepton channel SRs before and after the fit.

As a sanity check, all input variable distributions of the classification BDTs are evaluated post-fit and no significant deviations between predictions and data are found. Distributions of the Higgs boson candidate mass for $\mathrm{SR}_1^{\geq 6j}$ and $\mathrm{SR}_1^{\geq 4j}$ can be seen in Figure 5.32.

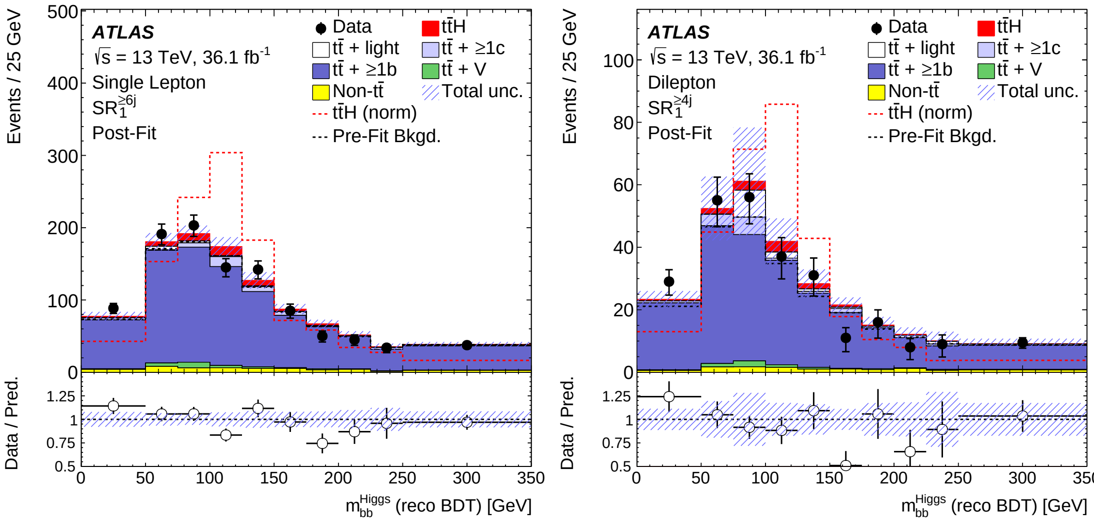


Figure 5.32.: Comparison between data and prediction for the Higgs boson candidate mass from the reconstruction BDT.

The combined fit in all signal and control regions of the single-lepton and dilepton channel corresponds to the best-fit μ value:

$$\mu = 0.84 \pm 0.29(\text{stat.})^{+0.57}_{-0.54}(\text{syst.}) = 0.84^{+0.64}_{-0.61}.$$

Additionally, an alternative combined fit is performed where an independent signal strength is assigned to each of the two channels. A μ value of $0.95^{+0.65}_{-0.62}$ is obtained for the single-lepton channel, whereas a value of $-0.24^{+1.02}_{-1.05}$ is received for the dilepton channel. A comparison between the combined signal strength and the two different channels can be seen in Figure 5.33. To evaluate the statistical uncertainty of the signal strength the fit to data is repeated with post-fit nuisance parameter values with the exception of μ and the normalisation factors for the $t\bar{t} + \geq 1c$ and $t\bar{t} + \geq 1b$ background contributions. The total systematic uncertainty is then calculated by subtracting in quadrature the statistical uncertainty from the total uncertainty. Overall, the analysis is dominated by systematic uncertainties, for which the largest contribution originates from the $t\bar{t} + \geq 1b$ modelling and the second largest relates to the limited number of events in the MC samples. Table 5.3 lists the contributions from the different sources of uncertainty in the combined fit to μ.

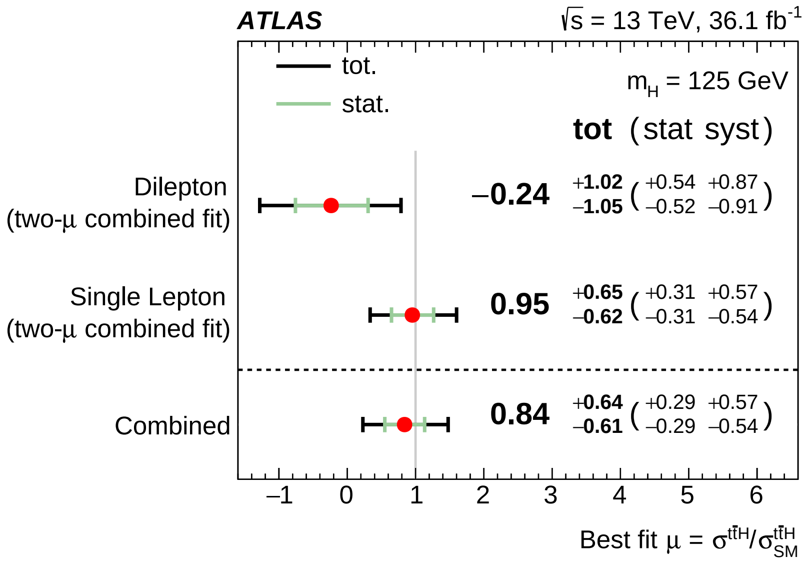


Figure 5.33.: Signal strength measurements in the single-lepton and dilepton channel (top) and for the combination (bottom) obtained from a simultaneous profile likelihood fit to data. The results for the individual channels are calculated without correlation between the signal strengths, while including correlations between nuisance parameters across channels.

Uncertainty source	$\Delta\mu$	
Topological information from $t\bar{t}$		
$t\bar{t}$ + $\geq$1b modelling	+0.46	-0.46
Background-model statistical uncertainty	+0.29	-0.31
b-tagging efficiency and mistag rates	+0.16	-0.16
Jet energy scale and resolution	+0.14	-0.14
$t\bar{t}H$ modelling	+0.29	-0.05
$t\bar{t}$ + $\geq$1c modelling	+0.09	-0.11
JVT, pileup modelling	+0.03	-0.05
Other background modelling	+0.08	-0.08
$t\bar{t}$ + light jets modelling	+0.06	-0.03
Luminosity	+0.03	-0.02
Lepton identification, isolation, trigger	+0.03	-0.04
Total systematic uncertainty	+0.57	-0.54
$t\bar{t}$ + $\geq$1b normalisation	+0.09	-0.10
$t\bar{t}$ + $\geq$1c normalisation	+0.02	-0.03
Intrinsic statistical uncertainty	+0.21	-0.20
Total statistical uncertainty	+0.29	-0.29
Total uncertainty	+0.64	-0.61

Table 5.3.: Breakdown of the contributions of the uncertainties in μ.

A ranking of the 20 nuisance parameters with the largest contribution to the total uncertainty is shown in Figure 5.34. The largest impact on the signal strength is driven by the deviation between the SHERPA5F and the nominal prediction for the $t\bar{t}$ + $\geq$1b background, followed by three $t\bar{t}$ + $\geq$1b background modelling uncertainties. These top four uncertainties suffer all from large theoretical uncertainties of the simulation of the $t\bar{t}$ + $\geq$1b process and are the limiting factor for this search.

Further tests were conducted to evaluate the impact of the fit on the nuisance parameters. It was seen that shifts of the nuisance parameters from their nominal values correct mainly the predictions of the $t\bar{t}$ background to the observed data. Additionally, the capability of the fit to constrain systematic uncertainties was verified on a pseudo-dataset based on the nominal samples, the Asimov dataset [175].

The measured signal strength corresponds to an observed (expected) significance of 1.4 (1.6) standard deviations. A signal strength larger than 2.0 can be excluded at the 95 % confidence level (see Figure 5.35). Figure 5.36 displays the event yields in data compared to the combined post-fit prediction for all analysis regions for the background-only and the signal-plus-background hypotheses in bins ordered in S/B ratio.

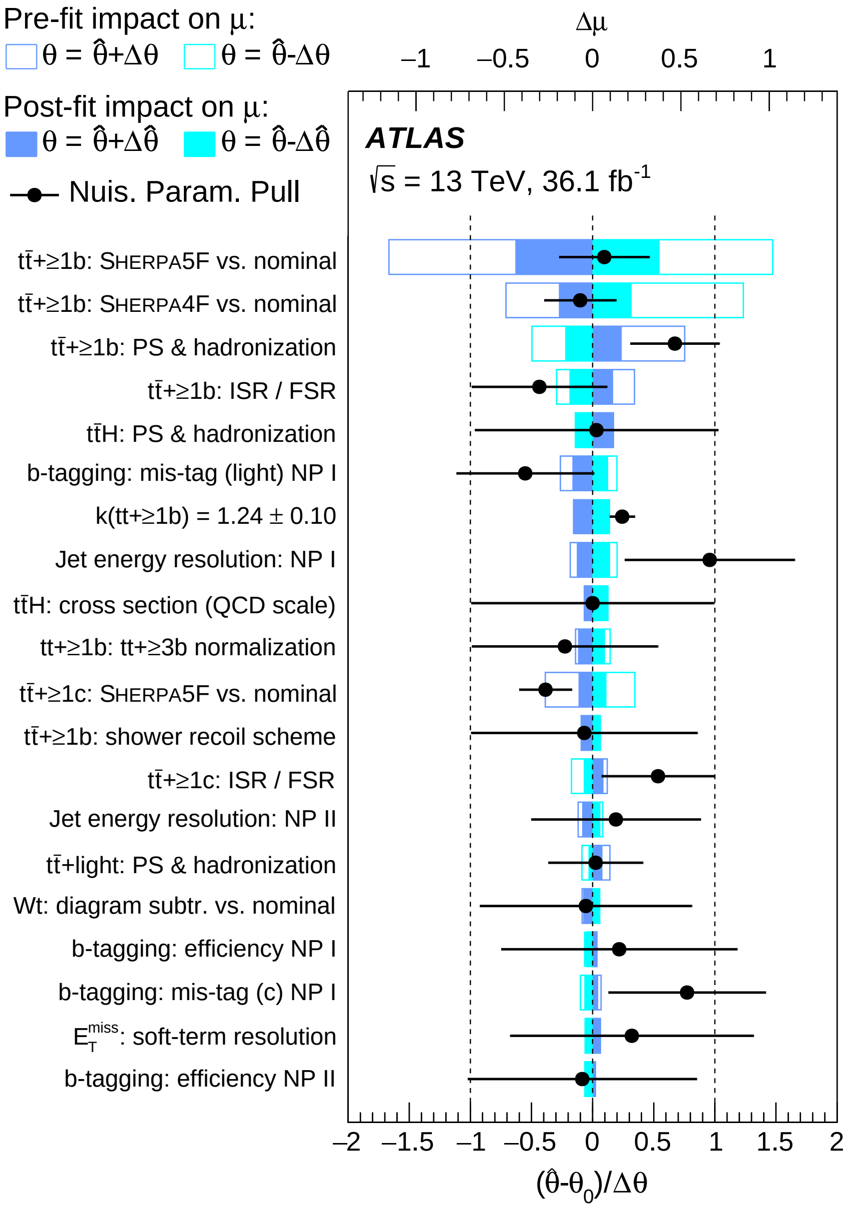


Figure 5.34.: Ranking of the 20 nuisance parameters with the largest contribution to the total uncertainty (ranked by decreasing contribution). Comparing the best-fit value of μ to the fit result obtained with post-fit nuisance parameter values, $\hat{\theta}$, shifted by their pre-fit (post-fit) uncertainties $\pm\Delta\theta$ ($\pm\Delta\hat{\theta}$), gives the impact of each nuisance parameter, $\Delta\mu$. The black points indicate the pulls of the nuisance parameters relative to their nominal values, θ_0. $k(t\bar{t} + \geq 1b)$ denotes to the floating normalisation of the $t\bar{t} + \geq 1b$ background.

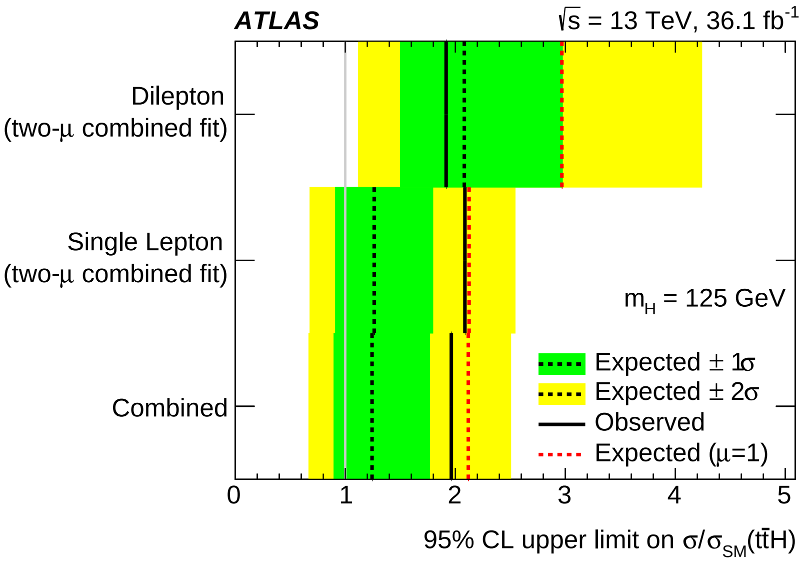


Figure 5.35.: Upper limits on $\sigma(t\bar{t}H)$ at the 95 % confidence level relative to the SM prediction in the single-lepton and dilepton channel and for the combination. Observed limits (solid black lines) as well as expected limits for the background-only hypothesis (dotted black lines) and for the SM hypothesis (dotted red lines) are shown.

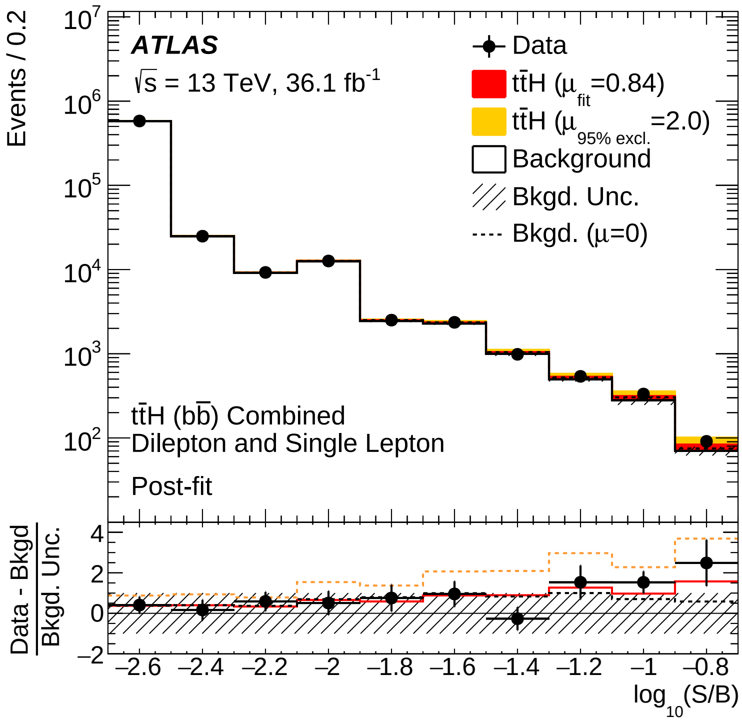


Figure 5.36.: Signal and background yields as a function of $\log_{10}(S/B)$, compared to data for all bins used in the combined fit of the single-lepton and dilepton channels. The signal is shown normalised to the best-fit value and to the excluded value at 95 % confidence level, including the background prediction from the fit. In addition, the pulls of the data relative to the background prediction are compared to the pulls of the signal-plus-background prediction from the fit for $\mu = 0.84$ ($\mu = 2$) and represented as a solid red line (dashed orange line). For the background-only hypothesis the pulls are drawn as a dashed black line. Underflow is included in the first bin.

CHAPTER 6

The observation of $t\bar{t}H$

In this chapter, $t\bar{t}H$ analyses targeting the remaining Higgs boson decay channels will be summarised and a combination of all results leading to the observation of this process will be presented. In addition, a comparison to the result of the $t\bar{t}H$ measurement by the CMS Collaboration is given.

6.1. Further Higgs boson decay channels

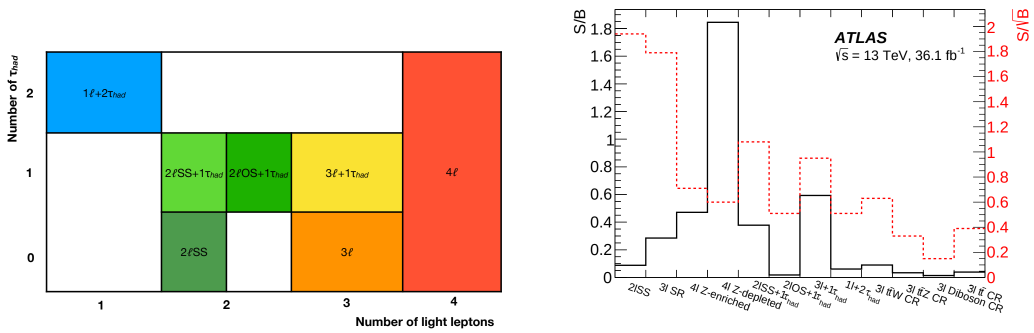
6.1.1. $H \rightarrow$ multilepton

The multilepton analysis considers the Higgs boson decays into WW^*, ZZ^*, $\tau^+\tau^-$, and uses a dataset of $36.1\,\mathrm{fb}^{-1}$ corresponding to the 2015 and 2016 ATLAS dataset [178]. Top-quark pairs decaying into a single-lepton or dilepton final state are considered. For this purpose, seven different channels depending on the number of leptons and hadronic taus are examined and are split into eight SRs and four CRs, see Figure 6.1.

Background contributions with prompt leptons originate mainly from top production in association with a vector boson, $t\bar{t}W$, $t\bar{t}(Z/\gamma^*)$, and diboson production. Data-driven methods are used to estimate non-prompt leptons and hadronic tau fakes. The modelling of this type of background is the greatest challenge of the analysis. To separate signal from background and to suppress backgrounds, a set of different BDTs is used. One dedicated BDT is used to reduce the misidentification of the electron charge, another BDT reduces the non-prompt electrons or muons. In the channels including tau leptons, the fake contribution from hadronic tau fakes is a significant background and needs to be well modelled. Prompt electrons or muons are estimated from MC simulations.

A maximum likelihood fit of the twelve categories is performed simultaneously to extract the $t\bar{t}H$ signal cross-section normalised to the prediction from the SM, see Figure 6.2. For five SRs the shape of the BDT distributions is used as the final discriminant, whereas the total yield is used in regions with low statistics.

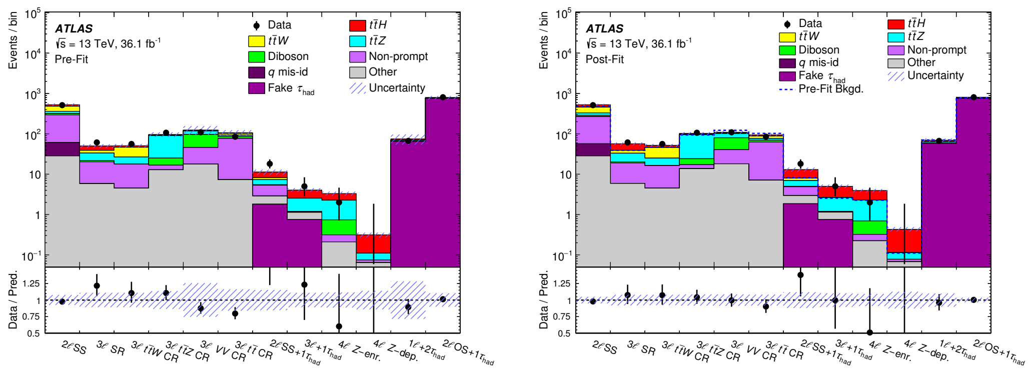
The main sources of systematic uncertainty are related to the signal modelling (cross-section prediction), the jet energy scale and resolution, and the fake estimation. Overall, the systematic and statistical uncertainties both account for about 30 %. A combined signal strength of $\mu = 1.6^{+0.5}_{-0.4}$ is measured, see Figure 6.3. This corresponds to an observed (expected) significance of 4.1 (2.8) standard deviations and gives evidence for the $t\bar{t}H$ production in the multilepton channel.



(a) Illustration of the channels used in the analysis.

(b) Pre-fit S/B (black line) and $S/\sqrt{B}$ (red dashed line) ratios for each of the twelve analysis regions.

Figure 6.1.: Channels and regions of the $H \to$ multilepton analysis.



(a) Pre-fit.

(b) Post-fit.

Figure 6.2.: Observed number of events in data compared to the background and signal yields before and after the fit in the twelve fit regions for the $H \to$ multilepton decay channel.

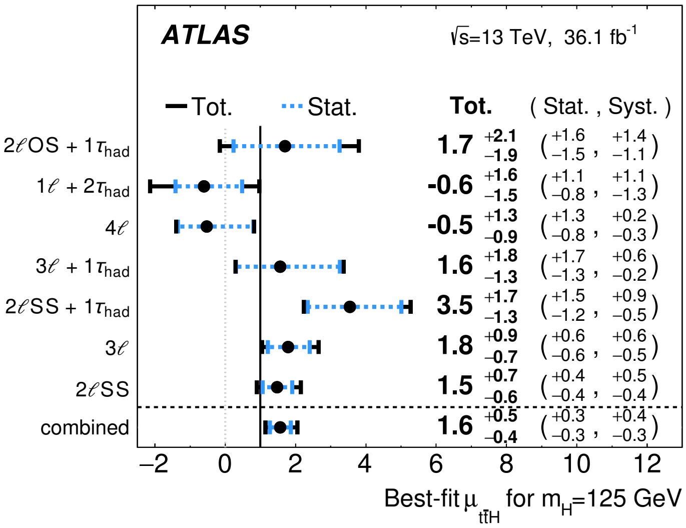


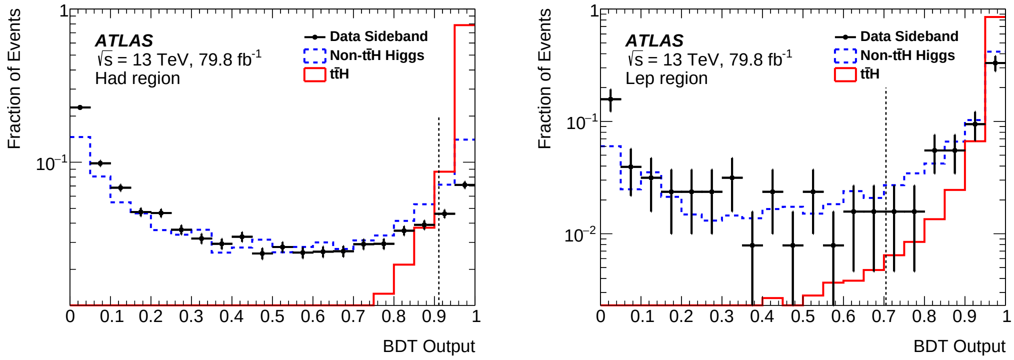
Figure 6.3.: The observed best-fit values of the $t\bar{t}H$ signal strength μ and their uncertainties of the $H \to$ multilepton analysis for each channel (top) and combined (bottom).

6.1.2. $H \to \gamma\gamma$

In the $H \to \gamma\gamma$ analysis events with two isolated photon candidates are selected [3, 179]. A dataset of 79.8 fb^{-1} corresponding to the 2015 – 2017 ATLAS dataset is analysed. When including the 2017 dataset, the sensitivity could be improved by about 50 % for the same integrated luminosity compared to the previous version of the analysis due to a refined analysis strategy and an updated reconstruction software.

The diphoton invariant mass, $m_{\gamma\gamma}$, is chosen to be in the range 105 GeV $\leq m_{\gamma\gamma} \leq$ 160 GeV and at least one b-tagged jet is required. Two signal regions are defined, a hadronic SR with at least two jets and zero isolated leptons, and a leptonic SR with at least one isolated lepton. A BDT is trained in each region with object-level variables, see Figure 6.4. The events are classified depending on the value of the BDT response in four (three) categories for the hadronic (leptonic) channel. This is done to optimise the sensitivity to the $t\bar{t}H$ signal. Figure 6.5 shows the weighted global fit of the diphoton mass.

The main systematic uncertainties are signal modelling, photon isolation and energy scale and resolution, and jet energy scale and resolution. The event yields are presented in Figure 6.6, here a signal strength of 1.4 is assumed. With an observed (expected) significance of 4.1 (3.7) standard deviations this Higgs boson decay channel also shows evidence for the $t\bar{t}H$ production.



(a) Hadronic signal region.

(b) Leptonic signal region.

Figure 6.4.: BDT output for the hadronic and leptonic signal regions for the $H \to \gamma\gamma$ decay channel. Events to the left of the vertical dashed line are rejected. The distributions are normalised to unity.

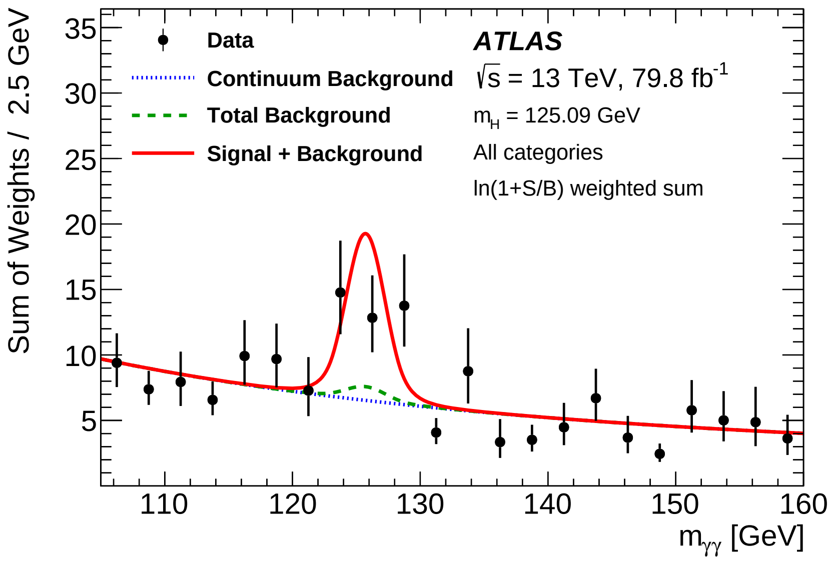


Figure 6.5.: Weighted diphoton invariant mass distribution in the $t\bar{t}H$-sensitive BDT bin of the $H \to \gamma\gamma$ analysis.

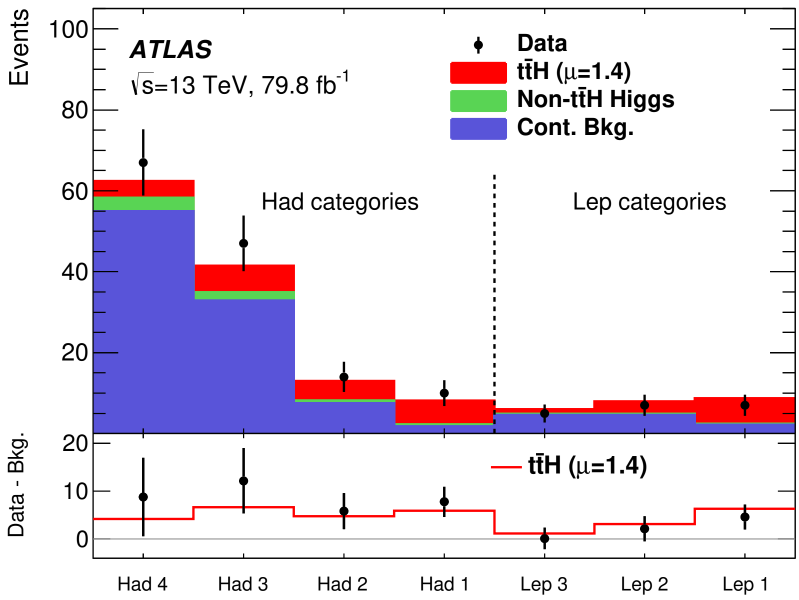


Figure 6.6.: Number of events in the different analysis regions of the $H \to \gamma\gamma$ analysis, in a diphoton mass windows that contains 90 % of the $t\bar{t}H$ signal.

6.1.3. $H \to ZZ^* \to 4\ell$

This channel was first considered for a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ and uses a dataset of $79.8\,\text{fb}^{-1}$ corresponding to the 2015 – 2017 ATLAS data [3, 180]. Here, the Higgs boson decays into same-flavour opposite-sign pairs of four electrons, four muons, or two electrons and two muons. A Higgs candidate is considered within the range $115\,\text{GeV} \leq m_{4\ell} \leq 130\,\text{GeV}$, a region that is excluded from the $H \to$ multilepton analysis.

Two signal regions enriched in $t\bar{t}H$ are selected by requiring at least one b-jet. The hadronic SR requires in addition at least four jets, whereas the leptonic SR requires in addition at least two jets and at least one lepton. In the hadronic region, a BDT is used to separate signal from backgrounds and the output discriminant is divided into two bins to maximise the expected $t\bar{t}H$ significance, see Figure 6.7. With this procedure, a signal purity of more than 80 % is anticipated for the leptonic region and the bin with the higher value of the BDT discriminant in the hadronic region. The other BDT bin is estimated to have a signal purity of about 35 %.

No event is observed and as an upper limit $\mu < 1.8$ can be excluded at 68 % confidence level. The expected significance is 1.2 standard deviations.

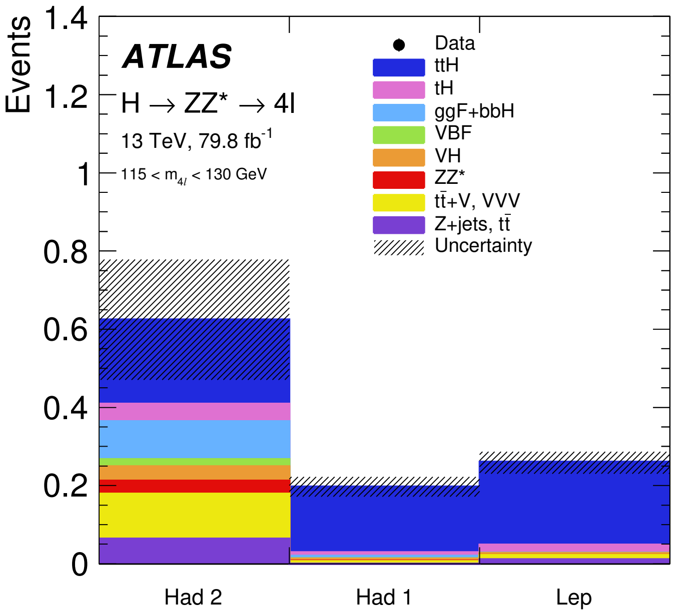


Figure 6.7.: Expected number of events in the three bins of the $H \to ZZ^* \to 4\ell$ analysis, including systematic uncertainties. No events are observed in data.

6.2. Combination of the Higgs boson decay channels

Combining the $H \to b\bar{b}$, $H \to$ multilepton, $H \to \gamma\gamma$, and $H \to ZZ^* \to 4\ell$ analyses, the $t\bar{t}H$ production was observed [3]. This was achieved with ATLAS Run 2 data up to $79.8\,\text{fb}^{-1}$. When combining the 8 TeV and 13 TeV data, the expected significance is also larger than five standard deviations, resulting in an observed (expected) significance of 6.3 (5.1) standard deviations, see Table 6.1.

Analysis	Integrated luminosity $[\text{fb}^{-1}]$	$t\bar{t}H$ cross-section $[\text{fb}^{-1}]$	Obs. sign.	Exp. sign.
$H \to \gamma\gamma$	79.8	710^{+210}_{-190} (stat.) $^{+120}_{-90}$ (syst.)	$4.1\ \sigma$	$3.7\ \sigma$
$H \to$ multilepton	36.1	790 ± 150 (stat.) $^{+150}_{-140}$ (syst.)	$4.1\ \sigma$	$2.8\ \sigma$
$H \to b\bar{b}$	36.1	400^{+150}_{-140} (stat.) ± 270 (syst.)	$1.4\ \sigma$	$1.6\ \sigma$
$H \to ZZ^* \to 4\ell$	79.8	< 900 (68 % CL)	$0\ \sigma$	$1.2\ \sigma$
Combined (13 TeV)	36.1–79.8	670 ± 90 (stat.) $^{+110}_{-100}$ (syst.)	$5.8\ \sigma$	$4.9\ \sigma$
Combined (7, 8, 13 TeV)	4.5, 20.3, 36.1–79.8	–	$6.3\ \sigma$	$5.1\ \sigma$

Table 6.1.: Measured total $t\bar{t}H$ production cross-sections as well as observed and expected significances relative to the background-only hypothesis.

The main systematic uncertainties in the combination are also related to modelling uncertainties, with the highest contribution from $t\bar{t}$ + heavy flavour modelling. Other sources arise from uncertainties in the jet energy scale and resolution, as well as fake leptons, which are estimated from leptons originating from heavy-flavour decays, conversions, or misidentified hadronic jets.

Figure 6.8 shows the ratios of the combined $t\bar{t}H$ production cross-section, and cross-sections measured in the individual analyses, to the SM prediction. A combined signal strength of

$$\mu = 1.32 \pm 0.18(\text{stat.})^{+0.21}_{-0.19}(\text{syst.}) = 1.32^{+0.28}_{-0.26}$$

is obtained, which corresponds to a measured total-cross section of $670 \pm 0.90(\text{stat.})^{+110}_{-100}(\text{syst.})$ fb at $\sqrt{s} = 13$ TeV. This result is in agreement with the SM prediction of 507^{+35}_{-50} fb calculated at QCD and electroweak NLO accuracy [65, 116–120]. A comparison of the signal and background yields to data in bins ordered in S/B ratio is given in Figure 6.9.

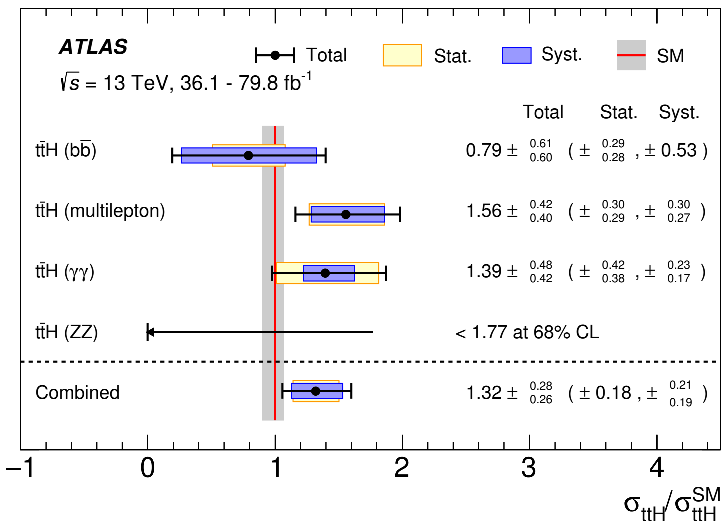


Figure 6.8.: Combined $t\bar{t}H$ production cross-section and cross-sections measured in the individual analyses, divided by the SM predictions for each process.

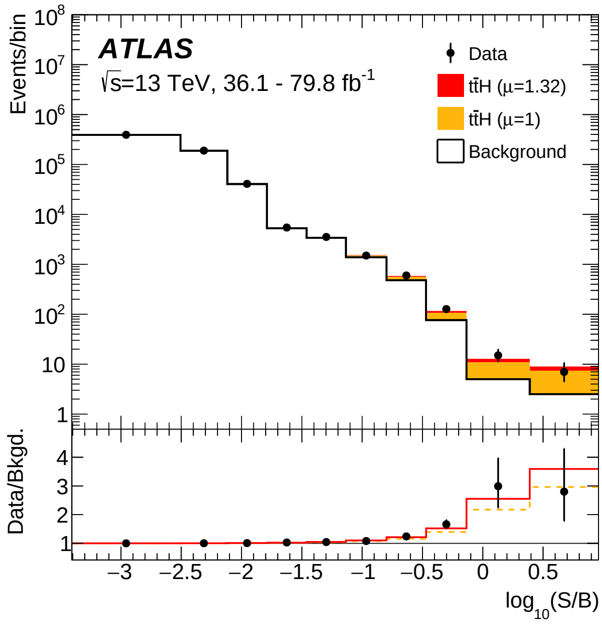


Figure 6.9.: Signal and background yields as a function of $\log_{10}(S/B)$, compared to data for all analysis categories. The signal is shown normalised to the best-fit value and to the SM prediction, including the background prediction from the fit. In addition, the ratio of the data to the background prediction is compared to the expected distribution including the signal for $\mu = 1.32$ ($\mu = 1$) and represented as a solid red line (dashed orange line).

6.3. Comparison to the results of the $t\bar{t}H$ analysis from CMS

The $t\bar{t}H$ process is also observed by the CMS experiment in a dataset corresponding to $5.1\,\text{fb}^{-1}$ (7 TeV) + $19.7\,\text{fb}^{-1}$ (8 TeV) + $35.9\,\text{fb}^{-1}$ (13 TeV) [181]. This is achieved by combining different analyses targeting the individual Higgs boson decays into WW^*, ZZ^*, $\gamma\gamma$, $\tau^+\tau^-$, and $b\bar{b}$, and results in a signal strength of $\mu = 1.26^{+0.31}_{-0.26}$, see Figure 6.10. Therefore, the background-only hypothesis can be excluded with an observed (expected) significance of 5.2 (4.2) standard deviations. The result is in good agreement with the SM prediction as well as the measurement from ATLAS. Figure 6.11 shows the signal and background yields compared to data in bins ordered in S/B ratio.

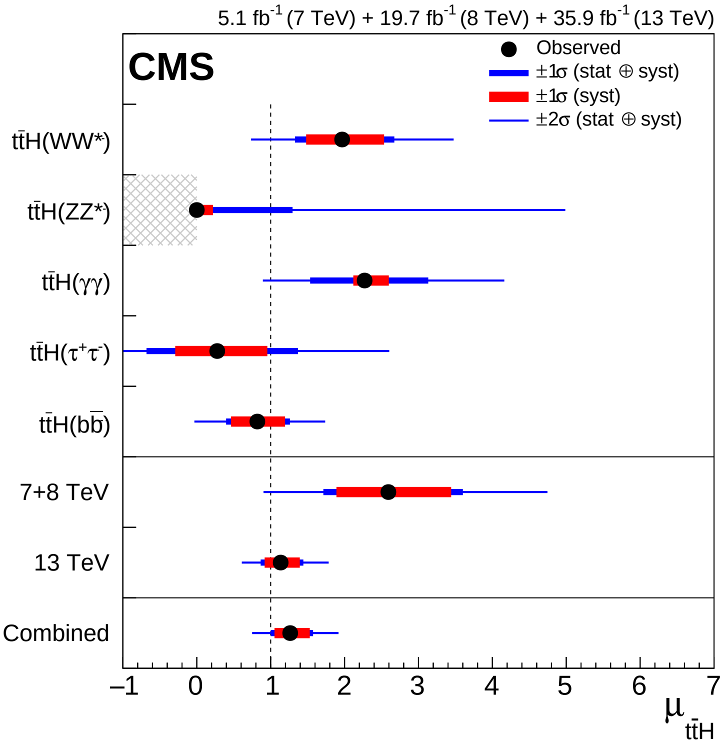


Figure 6.10.: Best fit value of the $t\bar{t}H$ signal strength for the five individual decay channels considered (top), the combined result for 7 and 8 TeV alone and for 13 TeV alone (middle), and the overall combined result (bottom) from CMS.

The uncertainty of the measurement is given with ± 0.16 (stat.)$^{+0.26}_{-0.21}$ (syst.), where the systematic component comprises an experimental uncertainty of $^{+0.17}_{-0.15}$, a theoretical uncertainty on the signal of $^{+0.15}_{-0.07}$, and an uncertainty related to the background modelling of $^{+0.13}_{-0.12}$. The ATLAS result has a larger statistical uncertainty of ± 0.18, which is hypothesised to be related to the statistical uncertainty on the freely floating fit parameters applied to a larger dataset. However, the systematic uncertainty of $^{+0.21}_{-0.19}$ is significantly smaller. On the one hand, this is related to the different experimental setup and on the other, to the estimation of modelling uncertainties, especially for the $t\bar{t}b\bar{b}$ background, which is one of the leading uncertainties in the combined analysis of ATLAS.

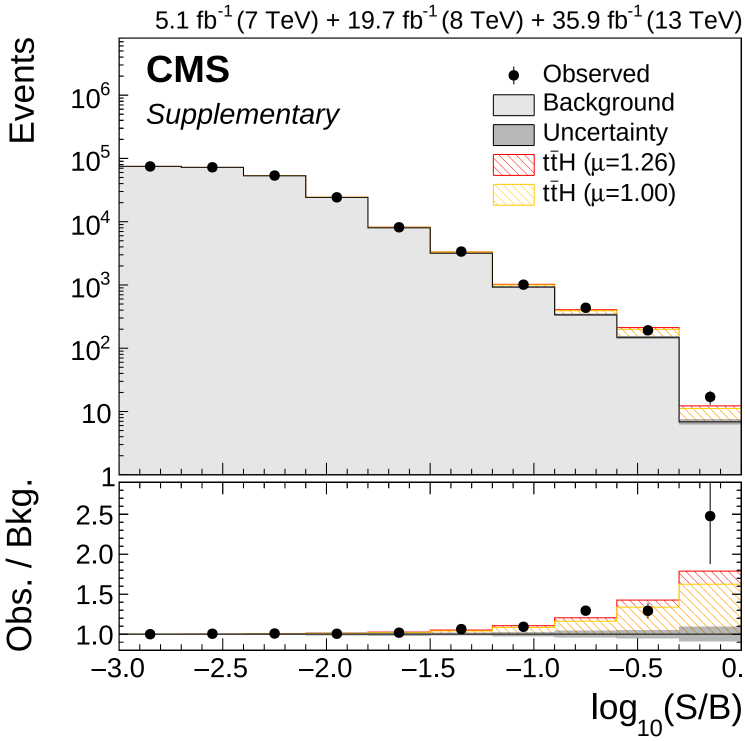


Figure 6.11.: Signal and background yields as a function of $\log_{10}(S/B)$, compared to data for all analysis categories from CMS. The signal is shown normalised to the best-fit value and to the SM prediction, including the background prediction from the fit. In addition, the ratios of the expected signal and observed results relative to the expected background are shown for $\mu = 1.26$ ($\mu = 1$) and represented in red (orange).

The $t\bar{t}H(H \to b\bar{b})$ analysis from CMS. Not included in the combination of CMS is an updated $H \to b\bar{b}$ analysis, which also considers all hadronic top-quark pair decays and includes the 2017 dataset [182–184]. Due to one additional layer in the pixel detector and an improved b-tagging algorithm, the background modelling could be significantly improved for the 2017 dataset. This updated analysis measures a signal strength of $\mu = 1.15 \pm 0.15(\text{stat.})^{+0.28}_{-0.25}(\text{syst.}) = 1.15^{+0.32}_{-0.29}$, see Figure 6.12. This corresponds to an observed (expected) significance of 3.9 (3.5) standard deviations and allows evidence for the $t\bar{t}H$ production in the $H \to b\bar{b}$ channel to be claimed.
Compared to the signal strength of the ATLAS analysis, $\mu = 0.84 \pm 0.29(\text{stat.})^{+0.57}_{-0.54}(\text{syst.}) = 0.84^{+0.64}_{-0.61}$, the measurement performed by CMS exhibits significantly lower uncertainties. The analysis strategy from CMS is similar to the one of ATLAS; both perform a combined

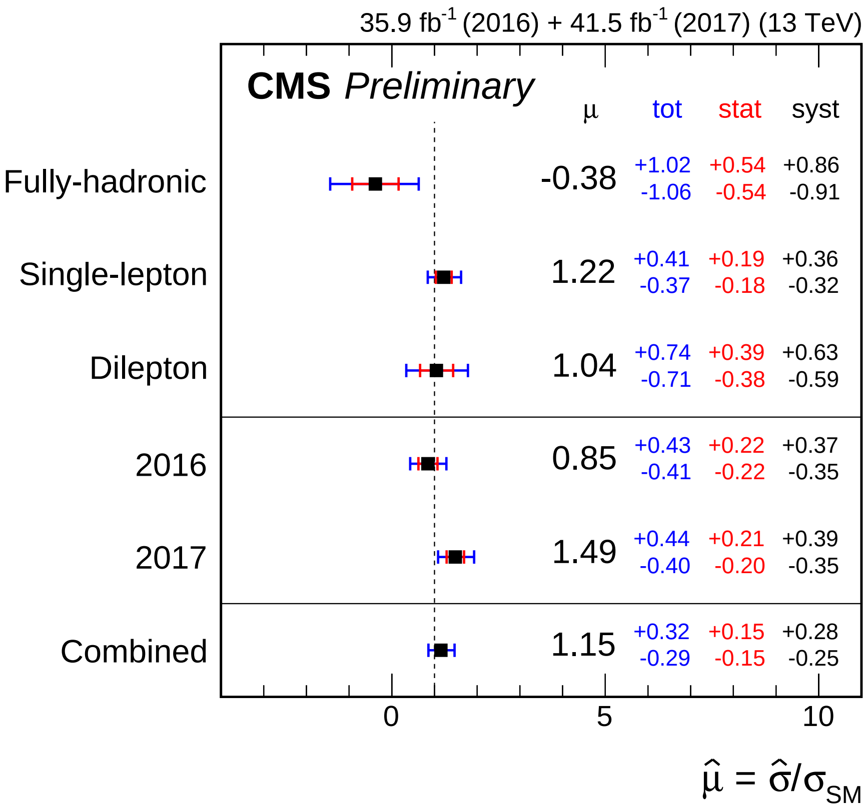


Figure 6.12.: Best fit value of the $t\bar{t}H(H \to b\bar{b})$ signal strength for the three individual decay channels considered (top), the combined result for the 2016 dataset and 2017 dataset (middle), and the overall combined result (bottom) from CMS. The error bars indicate the 1 σ expected confidence intervals, also split into their statistical (red inner error bar) and systematic component (blue outer error bar).

maximum likelihood fit to data. In the single-lepton channel, an NN is employed instead of a BDT to classify events. The all hadronic channel requires the presence of a W boson in $t\bar{t}H$ events and cuts on the invariant dijet mass within the W mass window. In addition, differences between QCD processes and the signal such as a smaller different angular distance between any two jets and a lower number of gluon-initiated jets in $t\bar{t}H$ events are leveraged by a cut on η and the quark-gluon likelihood ratio, respectively. The events are then categorised according to a matrix element method calculation. Table 6.2 shows the contributions of the uncertainty sources (the uncertainty sources of the ATLAS analysis can be found in Table 5.3).

Uncertainty source	$\Delta\mu$	
Experimental uncertainties		
b-tagging	+0.08	-0.07
Jet energy scale and resolution	+0.05	-0.04
Total experimental uncertainty	+0.15	-0.13
Theory uncertainties		
Signal	+0.15	-0.06
$t\bar{t}$ + heavy flavour modelling	+0.14	-0.15
Total theory uncertainty	+0.23	-0.10
QCD background prediction	+0.10	-0.08
Size of simulated samples	+0.10	-0.10
Total systematic uncertainty	+0.28	-0.25
Total statistical uncertainty	+0.15	-0.15
Total uncertainty	+0.32	-0.29

Table 6.2.: Breakdown of the contributions of the uncertainties in μ from CMS.

The CMS analysis uses a different choice of generators to simulate the $t\bar{t}$ + jets background, however the background is also categorised by heavy flavour components, $t\bar{t} + b\bar{b}$, $t\bar{t} + B$, $t\bar{t} + b$, and $t\bar{t} + c\bar{c}$. Each of these processes has a 50 % normalisation uncertainty assigned, whereas the fit constrains the nuisance parameters related to the heavy flavour cross-section to 30 % of the prior value. Not considered by CMS are effects on the shape of the discriminating distributions as well as uncertainties related to the ME and parton shower generators. In the ATLAS analysis this is accounted for by evaluating differences between the nominal $t\bar{t}$ + jets sample to an alternative five flavour simulation, depicted in a nuisance parameter with the impact on the signal strength. This is also reflected in the leading contribution to the systematic uncertainty in the ATLAS analysis, the $t\bar{b} + \geq 1b$ modelling.

In conclusion, both ATLAS and CMS have observed the $t\bar{t}H$ production, and the results show a good agreement between the experimental measurement and the prediction of the SM.

CHAPTER 7

Conclusions

In this thesis, the observation of the $t\bar{t}H$ process at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS experiment was presented. The signal strength $\mu = 1.32^{+0.28}_{-0.26}$ was measured, corresponding to an observed (expected) significance of 5.8 (4.9) standard deviations. Observing this process is one of the milestones of Run 2 and was achieved by combining several analyses targeting the different Higgs boson decay channels, namely the $H \to$ multilepton analysis considering Higgs boson decays into WW^*, ZZ^*, and $\tau^+\tau^-$, the $H \to \gamma\gamma$ analysis, the $H \to ZZ^* \to 4\ell$ analysis, and the $H \to b\bar{b}$ analysis. The latter targets the Higgs boson decay with the highest branching fraction of about 57 % and is discussed in detail in Chapter 5, while the remaining analyses are summarised in Chapter 6.

Chapter 5 includes a detailed description of the fake and non-prompt lepton estimation in the single-lepton channel of the $t\bar{t}H(H \to b\bar{b})$ analysis. For this purpose, a data-driven technique, the matrix method was employed. In the scope of this thesis, the matrix method was ported to an updated analysis reconstruction software release and performance studies between these release versions are shown. In general, a good agreement between both release versions was seen with the exception of the muon fake efficiency, where an approximately 2/3 lower efficiency for the updated release version was seen. The source of this discrepancy was expected to be related to a change in the overlap removal for muons between the two analysis versions. This effect was also seen in another ATLAS analysis targeting $H \to WW^*$ decays using a different framework for the fake estimation. This also supports the assumption that the source of this discrepancy lies outside of the matrix method's framework. For the first time, fake and real efficiencies were estimated for the 2017 data. With similar pile-up conditions for the 2017 and 2018 data taking periods these efficiencies can be used for the full Run 2 dataset not only in the $t\bar{t}H$ analysis but also in other analyses with leptonic final states.

It was also examined if a tag rate function could be used to increase the performance of the matrix method and this was successfully proven for a fixed b-tagging working point.

In addition, prospects of using deep neural networks for the separation of the signal from backgrounds were studied. A deep neural network only provided with low-level variables did not show a high separation power. However, including TRF b-tagging weights could improve the performance significantly, which shows machine learning is a powerful technique that should be exploited further in the future.

Only using the 2015 and 2016 dataset without the refined analysis reconstruction software, the $t\bar{t}H(H \to b\bar{b})$ analysis could measure a signal strength of $\mu = 0.84^{+0.64}_{-0.61}$ corresponding to an observed (expected) significance of 1.4 (1.6) standard deviations. The last section will give an outlook on the next steps for the $t\bar{t}H$ measurement.

7.1. Outlook

After the discovery of the $t\bar{t}H$ process, the next step is to reduce the uncertainties of this measurement. With a higher precision it will be possible to extract the top quark Yukawa coupling and compare it to its SM prediction. However, this can only be achieved with a larger dataset and a refined analysis strategy.

Consequently, the full Run 2 dataset corresponding to $139\,\mathrm{fb}^{-1}$ will be included in the different analyses. In this context, the analysis reconstruction software currently used will be updated, which includes various improvements for object reconstruction such as enhanced flavour tagging performance and a refined jet energy scale and resolution. The $t\bar{t}H(H \to b\bar{b})$ analysis is dominated by systematic uncertainties, where the largest contribution is related to the $t\bar{t}$ + jets background, especially the $t\bar{t} + \geq 1b$ modelling. Simply including more data will not lead to an improved measurement and, therefore, the modelling uncertainties need to be reduced with a different MC generation strategy. A large systematic uncertainty is introduced by reweighting MC samples simulated by different generators to the nominal one. This uncertainty could be drastically reduced with an updated generator setup, where the four flavour $t\bar{t}b\bar{b}$ simulation can be merged with the $t\bar{t}$ + jets sample as well as identifying and removing overlapping events. Currently in development is a setup, where Powheg interfaced to OpenLoops is employed for the ME calculation of the $t\bar{t}b\bar{b}$ process using a four flavour scheme at NLO precision [185]. The parton showering and hadronisation can then be performed with Pythia 8. This method has the advantage of using the same MC setup as for the inclusive $t\bar{t}$ + jets generation (Powheg+Pythia 8) and a comparison can be done in a more comprehensive way, which will improve modelling uncertainties. In addition, with a larger number of MC events it will be possible to cover a larger phase space resulting in a decreased statistical uncertainty.

The top quark Yukawa coupling can be extracted by measuring the ratio of the $t\bar{t}H/t\bar{t}Z$ production. Here, the $t\bar{t}Z$ process is the ideal production mechanism due to a similar mass of the Z boson compared to the Higgs boson. A combination of different Higgs boson decay channels (e.g. $t\bar{t}H(H \to b\bar{b}) \sim \kappa_t\kappa_b$ and $t\bar{t}H(H \to \tau\tau) \sim \kappa_t\kappa_\tau$) allows by comparing the coupling modifiers κ_i to determine the top quark Yukawa coupling more precisely.

Another prospect lies in the extraction of the fiducial $t\bar{t}H(H \to b\bar{b})$ cross-section, which can be combined with the $W/ZH(H \to b\bar{b})$ analysis to obtain a measurement for the $H \to b\bar{b}$ decay branching ratio.

For Run 3 the centre-of-mass energy of the LHC will be increased to $\sqrt{s} = 14\,\text{TeV}$, which will result in a higher $t\bar{t}H$ signal-over-background ratio (see Figure 2.4). With more data it will also be possible to search for a single top associated Higgs boson production, tH, which will allow to determine the sign of the top quark Yukawa coupling. In the SM, this process is suppressed by destructive interference. However, models beyond the SM predict constructive interference, which would increase the tH cross-section and give signs for new physics. The $t\bar{t}H$ and tH process are expected to have different kinematics, where the single top Higgs production is expected to be more in the forward region.

This outlook shows that while the observation of the $t\bar{t}H$ production was an important first step, further studies of this process will be able to probe the SM with even greater precision and might lead to possible signs of new physics.

Bibliography

[1] ATLAS Collaboration, G. Aad et al., *Observation of a New Particle in the Search for the Standard Model Higgs Boson with the ATLAS Detector at the LHC*, Phys. Lett. **B716** (2012) 1–29.

[2] CMS Collaboration, S. Chatrchyan et al., *Observation of a New Boson at a Mass of* 125 *GeV with the CMS Experiment at the LHC*, Phys. Lett. **B716** (2012) 30–61.

[3] ATLAS Collaboration, M. Aaboud et al., *Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector*, Phys. Lett. B **784** (2018) 173.

[4] ATLAS Collaboration, M. Aaboud et al., *Search for the standard model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, Phys. Rev. D **97** (2018) 072016.

[5] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321–323.

[6] P. W. Higgs, *Broken Symmetries, Massless Particles and Gauge Fields*, Phys. Lett. **12** (1964) 132–133.

[7] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508–509.

[8] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (1964) 585–587.

[9] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579.

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

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

[12] D. Hanneke, S. Fogwell, and G. Gabrielse, *New Measurement of the Electron Magnetic Moment and the Fine Structure Constant*, Phys. Rev. Lett. **100** (2008) 120801.

[13] M. Kobayashi and T. Maskawa, *CP-Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652–657.

[14] M. L. Perl et al., *Evidence for Anomalous Lepton Production in $e^+ - e^-$ Annihilation*, Phys. Rev. Lett. **35** (1975) 1489–1492.

[15] S. W. Herb et al., *Observation of a Dimuon Resonance at 9.5 GeV in 400 GeV Proton-Nucleus Collisions*, Phys. Rev. Lett. **39** (1977) 252–255.

[16] DONUT Collaboration, K. Kodama et al., *Observation of Tau Neutrino Interactions*, Phys. Lett. **B504** (2001) 218–224.

[17] Super-Kamiokande Collaboration, Y. Fukuda et al., *Evidence for oscillation of atmospheric neutrinos*, Phys. Rev. Lett. **81** (1998) 1562–1567.

[18] SNO Collaboration, Q. R. Ahmad et al., *Measurement of the rate of $\nu_e + d \to p + p + e^-$ interactions produced by ^8B solar neutrinos at the Sudbury Neutrino Observatory*, Phys. Rev. Lett. **87** (2001) 071301.

[19] SNO Collaboration, Q. R. Ahmad et al., *Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory*, Phys. Rev. Lett. **89** (2002) 011301.

[20] U. Sarkar, *Particle and Astroparticle Physics*. Taylor & Francis, ISBN: 978-1584889311, 2007.

[21] Planck Collaboration, P. A. R. Ade et al., *Planck 2013 Results. XVI. Cosmological Parameters*, Astron. Astrophys. **571** (2014) A16.

[22] S. Weinberg, *Implications of Dynamical Symmetry Breaking*, Phys. Rev. D **13** (1976) 974–996.

[23] E. Gildener, *Gauge Symmetry Hierarchies*, Phys. Rev. D **14** (1976) 1667–1672.

[24] S. Weinberg, *Implications of Dynamical Symmetry Breaking: An Addendum*, Phys. Rev. D **19** (1979) 1277–1280.

[25] L. Susskind, *Dynamics of Spontaneous Symmetry Breaking in the Weinberg-Salam Theory*, Phys. Rev. D **20** (1979) 2619–2625.

[26] H. Miyazawa, *Baryon Number Changing Currents*, Prog. Theor. Phys. **36 (6)** (1966) 1266–1276.

[27] P. Ramond, *Dual Theory for Free Fermions*, Phys. Rev. D **3** (1971) 2415–2418.

[28] Yu. A. Golfand and E. P. Likhtman, *Extension of the Algebra of Poincare Group Generators and Violation of p Invariance*, JETP Lett. **13** (1971) 323, [Pisma Zh. Eksp. Teor. Fiz. **13** (1971) 452].

[29] A. Neveu and J. H. Schwarz, *Factorizable Dual Model of Pions*, Nucl. Phys. **B31** (1971) 86–112.

[30] A. Neveu and J. H. Schwarz, *Quark Model of Dual Pions*, Phys. Rev. D **4** (1971) 1109–1111.

[31] J. L. Gervais and B. Sakita, *Field Theory Interpretation of Supergauges in Dual Models*, Nucl. Phys. **B34** (1971) 632–639.

[32] D. V. Volkov and V. P. Akulov, *Is the Neutrino a Goldstone Particle?*, Phys. Lett. B **46** (1973) 109.

[33] J. Wess and B. Zumino, *A Lagrangian Model Invariant under Supergauge Transformations*, Phys. Lett. **B49** (1974) 52–54.

[34] J. Wess and B. Zumino, *Supergauge Transformations in Four-Dimensions*, Nucl. Phys. B **70** (1974) 39.

[35] S. Dimopoulos and H. Georgi, *Softly Broken Supersymmetry and SU(5)*, Nucl. Phys. **B193** (1981) 150–162.

[36] E. Witten, *Dynamical Breaking of Supersymmetry*, Nucl. Phys. **B188** (1981) 513–554.

[37] M. Dine, W. Fischler, and S. M., *Supersymmetric Technicolor*, Nucl. Phys. **B189** (1981) 575–593.

[38] S. Dimopoulos and S. Raby, *Supercolor*, Nucl. Phys. **B192** (1981) 353–368.

[39] N. Sakai, *Naturalness in Supersymmetric GUTs*, Zeit. Phys. **C11** (1981) 153–157.

[40] R. K. Kaul and P. Majumdar, *Cancellation of Quadratically Divergent Mass Corrections in Globally Supersymmetric Spontaneously Broken Gauge Theories*, Nucl. Phys. **B199** (1982) 36–58.

[41] UA1 Collaboration, C. Albajar et al., *Events with Large Missing Transverse Energy at the CERN Collider: III. Mass Limits on Supersymmetric Particles*, Phys. Lett. **B198** (1987) 261–270.

[42] UA2 Collaboration, R. Ansari et al., *Search for Exotic Processes at the CERN $p\bar{p}$ Collider*, Phys. Lett. **B195** (1987) 613–622.

[43] ZEUS Collaboration, S. Chekanov et al., *Search for Stop Production in R-Parity-Violating Supersymmetry at HERA*, Eur. Phys. J. **C50** (2007) 269–281.

[44] H1 Collaboration, S. Aid et al., *A Search for Selectrons and Squarks at HERA*, Phys. Lett. **B380** (1996) 461–470.

[45] LEPSUSYWG, ALEPH, DELPHI, L3 and OPAL Experiments, *Charginos, Large m_0*, http://lepsusy.web.cern.ch/lepsusy/Welcome.html (visited 15.05.2019).

[46] ALEPH Collaboration, A. Heister et al., *Absolute Mass Lower Limit for the Lightest Neutralino of the MSSM from e^+e^- Data at $\sqrt{s}$ up to 209 GeV*, Phys. Lett. **B583** (2004) 247–263.

[47] DELPHI Collaboration, J. Abdallah et al., *Searches for Supersymmetric Particles in e^+e^- Collisions up to 208 GeV and Interpretation of the Results within the MSSM*, Eur. Phys. J. **C31** (2003) 421–479.

[48] L3 Collaboration, M. Acciarri et al., *Search for Charginos and Neutralinos in e^+e^- Collisions at $\sqrt{s} = 189$ GeV* , Phys. Lett. **B472** (2000) 420–433.

[49] OPAL Collaboration, G. Abbiendi et al., *Search for Chargino and Neutralino Production at $\sqrt{s} = 192$ GeV to 209 GeV at LEP*, Eur. Phys. J. **C35** (2004) 1–20.

[50] CDF Collaboration, T. Aaltonen et al., *Inclusive Search for Squark and Gluino Production in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV*, Phys. Rev. Lett. **102** (2009) 121801.

[51] CDF Collaboration, T. Aaltonen et al., *Search for Supersymmetry in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV Using the Trilepton Signature for Chargino-Neutralino Production*, Phys. Rev. Lett. **101** (2008) 251801.

[52] DØ Collaboration, V. M. Abazov et al., *Search for Squarks and Gluinos in Events with Jets and Missing Transverse Energy Using 2.1 fb^{-1} of $p\bar{p}$ Collision Data at $\sqrt{s} = 1.96$ TeV*, Phys. Lett. **B660** (2008) 449–457.

[53] DØ Collaboration, V. M. Abazov et al., *Search for Associated Production of Charginos and Neutralinos in the Trilepton Final State using 2.3 fb^{-1} of Data*, Phys. Lett. **B680** (2009) 34–43.

[54] ATLAS Collaboration, G. Aad et al., *ATLAS Supersymmetry (SUSY) searches*, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/SupersymmetryPublicResults (visited 15.05.2019).

[55] CMS Collaboration, S. Chatrchyan et al., *CMS Supersymmetry Physics Results*, https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS (visited 15.05.2019).

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

[57] CDF Collaboration, F. Abe et al., *Observation of Top Quark Production in $p\bar{p}$ collisions*, Phys. Rev. Lett. **74** (1995) 2626.

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

[59] C. Englert et al., *Precision measurements of Higgs couplings: implications for new physics scales*, J. Phys. G **41** (2014) 113001.

[60] J. N. Ng and P. Zakarauskas, *QCD-parton calculation of conjoined production of Higgs bosons and heavy flavors in p anti-p collisions*, Phys. Rev. D **29** (1984) 876.

[61] Z. Kunszt, *Associated production of heavy Higgs boson with top quarks*, Nucl. Phys. B **29** (1984) 876.

[62] S. Dawson et al., *Associated top quark Higgs boson production the LHC*, Phys. Rev. D **67** (2003) 071503.

[63] W. Beenakker et al., *Higgs radiation off top quarks at the Tevatron and the LHC*, Phys. Rev. Lett. **87** (2001) 201805.

[64] M. Sher, *Electroweak Higgs potentials and vacuum stability*, Phys. Rept. **179** (1989) 273–418.

[65] de Florian, D. and others, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*. CERN, ISSN 2519-8076, 2017.

[66] *CERN homepage*, `https://home.cern` (visited 02.05.2019).

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

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

[69] CMS Collaboration, S. Chatrchyan et al., *The CMS experiment at the CERN LHC*, JINST **3** (2008) S08004.

[70] LHCb Collaboration, J. Alves et al., *The LHCb Detector at the LHC*, JINST **3** (2008) S08005.

[71] ALICE Collaboration, K. Aamodt et al., *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.

[72] ATLAS Collaboration, G. Aad et al., *ATLAS Luminosity Run-2*, `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2` (visited 02.05.2019).

[73] ATLAS Collaboration, G. Aad et al., *Luminosity determination in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector at the LHC*, Eur. Phys. J. C **71** (2011) 1630.

[74] ATLAS Collaboration, G. Aad et al., *Improved luminosity determination in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector at the LHC*, Eur. Phys. J. C **73** (2013) 2518.

[75] ATLAS Collaboration, G. Aad et al., *Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC*, Eur. Phys. J. C **76** (2016) 653.

[76] ATLAS Collaboration, M. Capeans et al., *ATLAS Insertable B-Layer Technical Design Report*, ATLAS-TDR-19, 2010, https://cds.cern.ch/record/1291633, *ATLAS Insertable B-Layer Technical Design Report Addendum*, ATLAS-TDR-19-ADD-1, 2012, https://cds.cern.ch/record/1451888.

[77] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*, JINST **13** (2018) T05008.

[78] B. Abbott et al., *Expected performance of the ATLAS b-tagging algorithms in Run-2*, ATL-PHYS-PUB-2015-022, 2015, https://cds.cern.ch/record/2037697.

[79] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, JINST **13** (2018) P07017.

[80] ATLAS Collaboration, P. Jenni, M. Nessi, and M. Nordberg, *Zero degree calorimeters for ATLAS*, CERN-LHCC-2007-001, 2007, http://cds.cern.ch/record/1009649.

[81] L. Adamczyk et al., *Technical Design Report for the ATLAS Forward Proton Detector*, CERN-LHCC-2015-009, 2015, https://cds.cern.ch/record/2017378.

[82] S. Abdel Khalek et al., *The ALFA Roman Pot detectors of ATLAS*, JINST **11** (2016) P11013.

[83] ATLAS Collaboration, G. Aad et al., *Performance of the ATLAS Trigger System in 2015*, Eur. Phys. J. C **77** (2016) 317.

[84] Apollinari, G. and others, *High-Luminosity Large Hadron Collider (HL-LHC) : Technical Design Report V. 0.1*, CERN-2017-007-M, 2017, https://cds.cern.ch/record/2284929.

[85] ATLAS Collaboration, G. Aad et al., *Technical Design Report for the ATLAS Inner Tracker Pixel Detector*, CERN-LHCC-2017-021. ATLAS-TDR-030, 2017, https://cds.cern.ch/record/2285585.

[86] ATLAS Collaboration, G. Aad et al., *Technical Design Report for the ATLAS Inner Tracker Strip Detector*, CERN-LHCC-2017-005. ATLAS-TDR-025, 2017, https://cds.cern.ch/record/2257755.

[87] T. Kawamoto et al., *New Small Wheel Technical Design Report*, CERN-LHCC-2013-006. ATLAS-TDR-020, 2013, https://cds.cern.ch/record/1552862.

[88] ATLAS Collaboration, G. Aad et al., *Technical Design Report for the Phase-II Upgrade of the ATLAS TDAQ System*, CERN-LHCC-2017-020. ATLAS-TDR-029, 2017, https://cds.cern.ch/record/2285584.

[89] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4 - A Simulation Toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250–303.

[90] ATLAS Collaboration, G. Aad et al., *The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim*, ATL-PHYS-PUB-2010-013, 2010, https://cds.cern.ch/record/1300517.

[91] ATLAS Collaboration, G. Aad et al., *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823.

[92] ATLAS Collaboration, G. Aad et al., *Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data*, Eur. Phys. J. C **72** (2012) 1909.

[93] ATLAS Collaboration, G. Aad et al., *Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton-proton collision data*, Eur. Phys. J. C **74** (2014) 2941.

[94] ATLAS Collaboration, G. Aad et al., *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data*, ATLAS-CONF-2016-024, 2016, https://cds.cern.ch/record/2157687.

[95] ATLAS Collaboration, G. Aad et al., *Muon reconstruction performance of the ATLAS detector in proton-proton collision data at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **76** (2016) 292.

[96] ATLAS Collaboration, G. Aad et al., *Muon Combined Performance in Run 2 (25 ns runs)*, ATL-COM-MUON-2015-093, 2015, https://cds.cern.ch/record/2105495.

[97] ATLAS Collaboration, G. Aad et al., *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. C **77** (2017) 490.

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

[99] M. Cacciari, P. Salam, and G. Soyez, *FastJet User Manual*, Eur. Phys. J. C **72** (2012) 1896.

[100] ATLAS Collaboration, M. Aaboud et al., *Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. D **96** (2017) 072002.

[101] ATLAS Collaboration, G. Aad et al., *Selection of jets produced in 13 TeV proton-proton collisions with the ATLAS detector*, ATLAS-CONF-2015-029, 2015, `https://cds.cern.ch/record/2037702`.

[102] ATLAS Collaboration, G. Aad et al., *Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, Eur. Phys. J. C **76** (2016) 581.

[103] ATLAS Collaboration, G. Aad et al., *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC*, ATL-PHYS-PUB-2015-045, 2015, `https://cds.cern.ch/record/2064383`.

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

[105] ATLAS Collaboration, G. Aad et al., *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*, ATL-PHYS-PUB-2016-012, 2016, `https://cds.cern.ch/record/2160731`.

[106] DØ Collaboration, V. M. Abazov et al., *Measurement of the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV using secondary vertex b tagging*, Phys. Rev. D **74** (2016) 112004.

[107] ATLAS Collaboration, G. Aad et al., *Performance of missing transverse momentum reconstruction with the ATLAS detector in the first proton-proton collisions at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-027, 2015, `https://cds.cern.ch/record/2037904`.

[108] ATLAS Collaboration, G. Aad et al., *Performance of algorithms that reconstruct missing transverse momentum in $\sqrt{s} = 8$ TeV proton-proton collisions in the ATLAS detector*, Eur. Phys. J. C **77** (2017) 241.

[109] ATLAS Collaboration, B. Nachman et al., *Jets from jets: re-clustering as a tool for large radius jet reconstruction and grooming at the LHC*, JHEP **02** (2015) 075.

[110] T. Sjöstrand, S. Mrenna, and P. Z. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852.

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

[112] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) 079.

[113] ATLAS Collaboration, G. Aad et al., *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014, https://cds.cern.ch/record/1966419.

[114] R. D. Ball et al., *Parton distributions for the LHC Run II*, JHEP **04** (2015) 040.

[115] P. Artoisenet et al., *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations*, JHEP **03** (2013) 015.

[116] R. Raitio and W. W. Wada, *Higgs-boson production at large transverse momentum in quantum chromodynamics*, Phys. Rev. D **19** (1979) 941.

[117] W. Beenakker et al., *NLO QCD corrections to $t\bar{t}H$ production in hadron collisions*, Nucl. Phys. B **653** (2003) 151.

[118] S. Dawson et al., *Associated Higgs boson production with top quarks at the CERN Large Hadron Collider: NLO QCD corrections*, Phys. Rev. D **68** (2003) 034022.

[119] Y. Zhang et al., *QCD NLO and EW NLO corrections to $t\bar{t}H$ production with top quark decays at hadron collider*, Phys. Lett. B **738** (2014) 1.

[120] S. Frixione et al., *Electroweak and QCD corrections to top-pair hadroproduction in association with heavy bosons*, JHEP **06** (2015) 184.

[121] A. Djouadi, J. Kalinowski, and M. Spira, *HDECAY: a program for Higgs boson decays in the Standard Model and its supersymmetric extension*, Comput. Phys. Commun. **108** (1998) 56.

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

[123] M. Cacciari et al., *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*, Phys. Lett. B **710** (2012) 612.

[124] M. Czakon and A. Mitov, *NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels*, JHEP **12** (2012) 054.

[125] M. Czakon and A. Mitov, *NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction*, JHEP **01** (2013) 080.

[126] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040.

[127] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, JHEP **11** (2007) 070.

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

[129] J. M. Campbell et al., *Top-pair production and decay at NLO matched with parton showers*, JHEP **04** (2015) 114.

[130] ATLAS Collaboration, G. Aad et al., *Studies on top-quark Monte Carlo modelling for Top2016*, ATL-PHYS-PUB-2016-020, 2016, `https://cds.cern.ch/record/2216168`.

[131] ATLAS Collaboration, G. Aad et al., *Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, Eur. Phys. J. C **75** (2015) 349.

[132] F. Cascioli et al., *NLO matching for $t\bar{t}b\bar{b}$ production with massive b-quarks*, Phys. Rev. Lett. B **734** (2014) 210.

[133] T. Gleisberg et al., *Event generation with SHERPA 1.1*, JHEP **02** (2009) 007.

[134] M. Czakon, P. Fiedler, and A. Mitov, *Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $\mathcal{O}(\alpha_S^4)$*, Phys. Rev. Lett. **110** (2013) 252004.

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

[136] M. Guzzi et al., *CT10 parton distributions and other developments in the global QCD analysis*, 2011, `arXiv:1101.0561 [hep-ph]`.

[137] J. Gao et al., *CT10 next-to-next-to-leading order global analysis of QCD*, Phys. Rev. D **89** (2014) 033009.

[138] S. Frixione et al., *Single-top hadroproduction in association with a W boson*, JHEP **07** (2008) 029.

[139] T. Sjöstrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 physics and manual*, JHEP **05** (2006) 026.

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

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

[142] N. Kidonakis, *NNLL resummation for s-channel single top quark production*, Phys. Rev. D **81** (2010) 054028.

[143] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production*, Phys. Rev. D **83** (2011) 091503.

[144] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, JHEP **12** (2008) 039.

[145] S. Schumann and F. Krauss, *A Parton shower algorithm based on Catani-Seymour dipole factorisation*, JHEP **03** (2008) 038.

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

[147] ATLAS Collaboration, G. Aad et al., *Measurement of $W^{\pm}$ and Z Boson Production Cross Sections in pp Collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector*, ATLAS-CONF-2015-039, 2015, https://cds.cern.ch/record/2045487.

[148] ATLAS Collaboration, G. Aad et al., *Multi-boson simulation for 13 TeV ATLAS analyses*, ATL-PHYS-PUB-2016-002, 2016, https://cds.cern.ch/record/2119986.

[149] M. Bahr et al., *Herwig++ physics and manual*, Eur. Phys. J. C **58** (2008) 639.

[150] DØ Collaboration, V. M. Abazov et al., *Measurement of the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV using kinematic characteristics of lepton + jets events*, Phys.Rev. D **76** (2017) 092007.

[151] A. Hoecker et al., *TMVA - Toolkit for Multivariate Data Analysis*, 2007, arXiv:physics/0703039 [physics.data-an].

[152] ATLAS Collaboration, G. Aad et al., *Search for flavour-changing neutral current top quark decays $t \to Hq$ in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, JHEP **12** (2015) 061.

[153] M. R. Whalley, D. Bourilkov, and R. C. Group, *'The Les Houches accord PDFs (LHAPDF) and LHAGLUE', HERA and the LHC: A Workshop on the implications of HERA for LHC physics. Proceedings, Part B*, 2005, arXiv:hep-ph/0508110 [hep-ph].

[154] G. P. Lepage, *A New Algorithm for Adaptive Multidimensional Integration*, J. Comput. Phys. **27** (1978) 192.

[155] D. Schouten, A. DeAbreu, and B. Stelzer, *Accelerated matrix element method with parallel computing*, Comput. Phys. Commun. **192** (2015) 54.

[156] M. Mantoani, *Search for the Standard Model Higgs boson produced in association with a pair of top quarks and decaying into a $b\bar{b}$-pair in the single lepton channel at $\sqrt{s} = 13$ TeV with the ATLAS experiment at the LHC*. PhD thesis, II.Physik-UniGö-Diss-2017/02, 2017.

[157] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, Conf. Proc. ICLR, 2015.

[158] F. Chollet et al., *Keras*, 2015, https://keras.io.

[159] M. Abadi et al., *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, 2015, http://tensorflow.org/.

[160] V. Barger, J. Ohnemus, and R. Phillips, *Event shape criteria for single lepton top signals*, Phys. Rev. D **48** (1993) 3953.

[161] G. Fox and S. Wolfram, *Observables for the Analysis of Event Shapes in $e^+ - e^-$ Annihilation and other Processes*, Phys. Rev. Lett. **41** (1978) 1581.

[162] C. Bernaciak et al., *Fox-Wolfram Moments in Higgs Physics*, Phys. Rev. D **87** (2013) 073014.

[163] M. Feindt, *A Neural Bayesian Estimator for Conditional Probability Densities*, 2004, arXiv:physics/0402093 [physics.data-an].

[164] M. Feindt and U. Kerzel, *The NeuroBayes neural network package*, Nucl. Instrum. Meth. A **559** (2006) 190.

[165] S. Hochreiter et al., *Gradient flow in recurrent nets: the difficulty of learning long-term dependencies.* IEEE Press, ISBN: 0780353692, 2001.

[166] H. Larochelle et al., *Exploring strategies for training deep neural networks*, JMLR **10** (2009) 1-40.

[167] ATLAS Collaboration, M. Aaboud et al., *Measurement of the Inelastic Proton-Proton Cross Section at $\sqrt{s} = 13$ TeV with the ATLAS Detector at the LHC*, Phys. Rev. Lett. **117** (2016) 182002.

[168] ATLAS Collaboration, M. Aaboud et al., *Measurement of b-tagging efficiency of c-jets in $t\bar{t}$ events using a likelihood approach with the ATLAS detector*, ATLAS-CONF-2018-001, 2018, https://cds.cern.ch/record/2306649.

[169] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, Eur. Phys. J. C **76** (2016) 196.

[170] ATLAS Collaboration, M. Aaboud et al., *Studies on top-quark Monte Carlo modelling with Sherpa and MG5_aMC@NLO*, ATL-PHYS-PUB-2017-007, 2017, https://cds.cern.ch/record/2261938.

[171] ATLAS Collaboration, M. Aaboud et al., *Studies of $t\bar{t} + c\bar{c}$ production with MadGraph5_aMC@NLO and Herwig++ for the ATLAS experiment*, ATL-PHYS-PUB-2016-011, 2016, https://cds.cern.ch/record/2153876.

[172] A. D. Martin et al., *Parton distributions for the LHC*, Eur. Phys. J. C **63** (2009) 189.

[173] J. M. Campbell and R. K. Ellis, *$t\bar{t}W^{\pm}$ production and decay at NLO*, JHEP **07** (2012) 052.

[174] F. Kohn, *Measurement of the charge asymmetry in top quark pair production in in pp collision data at* $\sqrt{s} = 7$ *TeV using the ATLAS detector*. PhD thesis, II.Physik-UniGö-Diss-2012/02, 2012.

[175] G. Cowan et al., *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C **71** (2011) 1554, [Erratum: Eur. Phys. J. C **73** (2013) 2501].

[176] A. L. Read, *Presentation of search results: The CLS technique*, J. Phys. G **28** (2002) 2693.

[177] T. Junk, *Confidence level computation for combining searches with small statistics*, Nucl. Instrum. Meth. A **434** (1999) 435.

[178] ATLAS Collaboration, M. Aaboud et al., *Evidence for the associated production of the Higgs boson and a top quark pair with the ATLAS detector*, Phys. Rev. D **97** (2018) 072003.

[179] ATLAS Collaboration, M. Aaboud et al., *Measurements of Higgs boson properties in the diphoton decay channel with 36* fb^{-1} *of pp collision data at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Phys. Rev. D **98** (2018) 052005.

[180] ATLAS Collaboration, M. Aaboud et al., *Measurement of the Higgs boson coupling properties in the* $H \to ZZ \to 4l$ *decay channel at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, JHEP **03** (2018) 095.

[181] CMS Collaboration, A. M. Sirunyan et al., *Observation of* $t\bar{t}H$ *Production*, Phys. Rev. Lett. **120** (2018) 231801.

[182] CMS Collaboration, A. M. Sirunyan et al., *Search for* $t\bar{t}H$ *production in the* $H \to b\bar{b}$ *decay channel with leptonic* $t\bar{t}$ *decays in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, JHEP **2019** (2019) 26.

[183] CMS Collaboration, A. M. Sirunyan et al., *Search for* $t\bar{t}H$ *production in the all-jet final state in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, JHEP **06** (2018) 101.

[184] CMS Collaboration, A. M. Sirunyan et al., *Measurement of* $t\bar{t}H$ *production in the* $H \to b\bar{b}$ *decay channel in 41.5* fb^{-1} *of proton-proton collision data at* $\sqrt{s} = 13$ *TeV*, CMS-PAS-HIG-18-030, 2019, `https://cds.cern.ch/record/2675023`.

[185] T. Ježo et al., *New NLOPS predictions for* $t\bar{t} + b$*-jet production at the LHC*, Eur. Phys. J. C **78** (2018) 502.

List of Figures

List of Tables

Appendices

APPENDIX A

Single-lepton trigger list

In the following, the lepton p_T, identification, and isolation requirements for the single-lepton triggers are listed [94, 95]. The 2017 triggers and pre-scale (PS) triggers are only used for the fake efficiency studies in Section 5.3.

- 2015
  - Electron
    - $p_T < 61\,\text{GeV}$
      - $p_T > 24\,\text{GeV}$ and medium identification
    - $p_T \geq 61\,\text{GeV}$
      - medium identification; for $p_T > 120\,\text{GeV}$ loose identification
  - Muon
    - $p_T < 51\,\text{GeV}$
      - $p_T > 20\,\text{GeV}$ and loose identification
    - $p_T \geq 51\,\text{GeV}$
      - no identification and isolation requirements

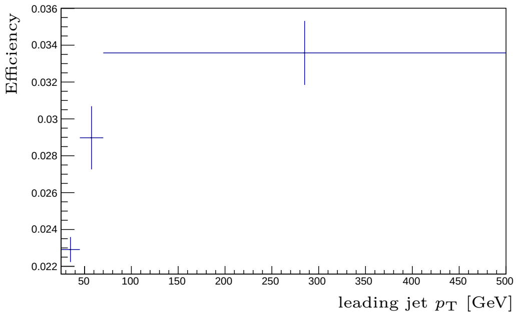
- 2016
  - Electron
    - $p_T < 61$ GeV
      - $p_T > 26$ GeV and tight identification and loose isolation
      - PS: $p_T > 24$ GeV and medium identification
    - $p_T \geq 61$ GeV
      - medium identification; for $p_T > 140$ GeV loose identification
  - Muon
    - $p_T < 51$ GeV
      - $p_T > 26$ GeV and medium isolation
      - PS: $p_T > 24$ GeV
    - $p_T \geq 51$ GeV
      - no identification and isolation requirements
- 2017
  - Electron
    - $p_T < 61$ GeV
      - $p_T > 26$ GeV and tight identification and loose isolation
      - PS: $p_T > 26$ GeV and medium identification
    - $p_T \geq 61$ GeV
      - medium identification; for $p_T > 140$ GeV loose identification
  - Muon
    - $p_T < 51$ GeV
      - $p_T > 26$ GeV and medium isolation
      - PS: $p_T > 24$ GeV
    - $p_T \geq 51$ GeV
      - no identification and isolation requirements

APPENDIX B

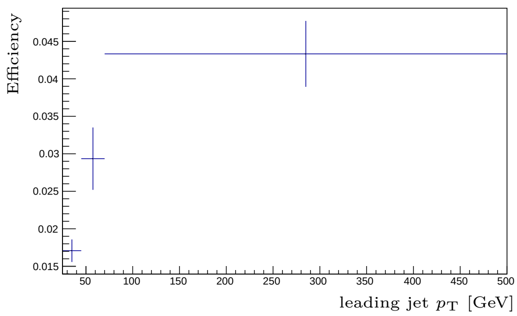
Fake estimation with a tag rate function

B.1. TRF b-tagging efficiency

Figure B.1 shows the TRF b-tagging efficiency in dependence of the leading jet p_{T} calculated in a four jet inclusive region with at least one b-jet using a working point of 60 %, 70 %, and 77 %. Electrons and muons are considered separately. The efficiency for a WP of 85 % is shown in Figure 5.8. The efficiency increases with a looser working point.

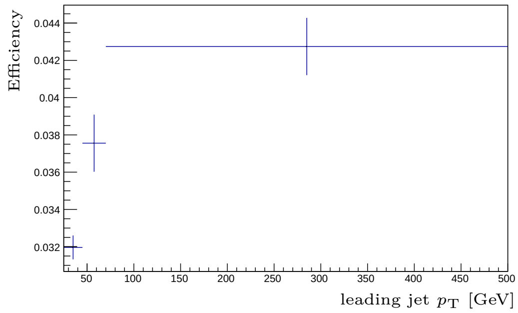


(a) 60 % WP: Efficiency in the case of an electron.

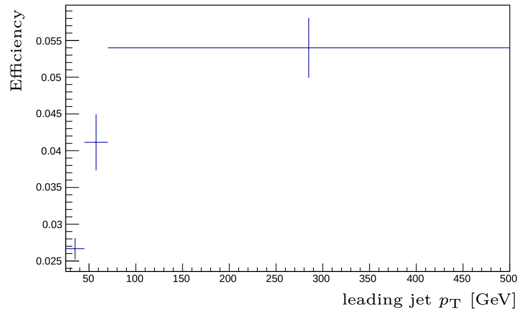


(b) 60 % WP: Efficiency in the case of a muon.

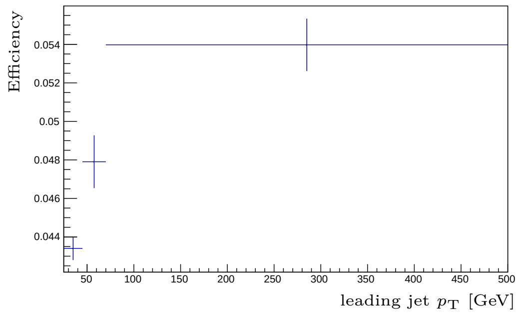
Figure B.1.: TRF b-tagging efficiency calculated in a four jet inclusive region with at least one b-jet using different b-tagging WPs in dependence of the leading jet p_{T}; electrons (left) and muons (right) are considered separately. Continued on the next page.



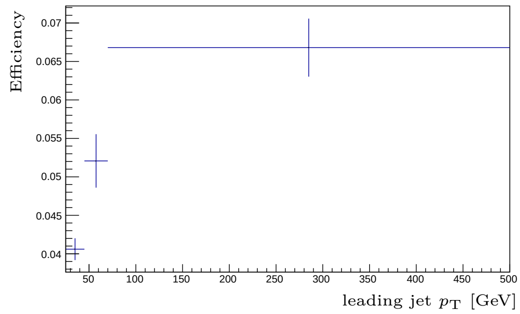
(c) 70 % WP: Efficiency in the case of an electron.



(d) 70 % WP: Efficiency in the case of a muon.



(e) 77 % WP: Efficiency in the case of an electron.



(f) 77 % WP: Efficiency in the case of a muon.

Figure B.1.: TRF b-tagging efficiency calculated in a four jet inclusive region with at least one b-jet using different b-tagging WPs in dependence of the leading jet p_T; electrons (left) and muons (right) are considered separately.

B.2. Comparison of the fake estimation with and without TRF

A full comparison of the event yields of the fake estimation with and without the TRF method is displayed on the following pages, see Table B.1.

b-tagging WP	Electron + jets	Muon + jets
60 %	Figure B.2	Figure B.3
70 %	Figure B.4	Figure B.5
77 %	Figure B.6	Figure B.7
85 %	Figure 5.9	Figure 5.10

Table B.1.: Reference to all figures for the different selections.

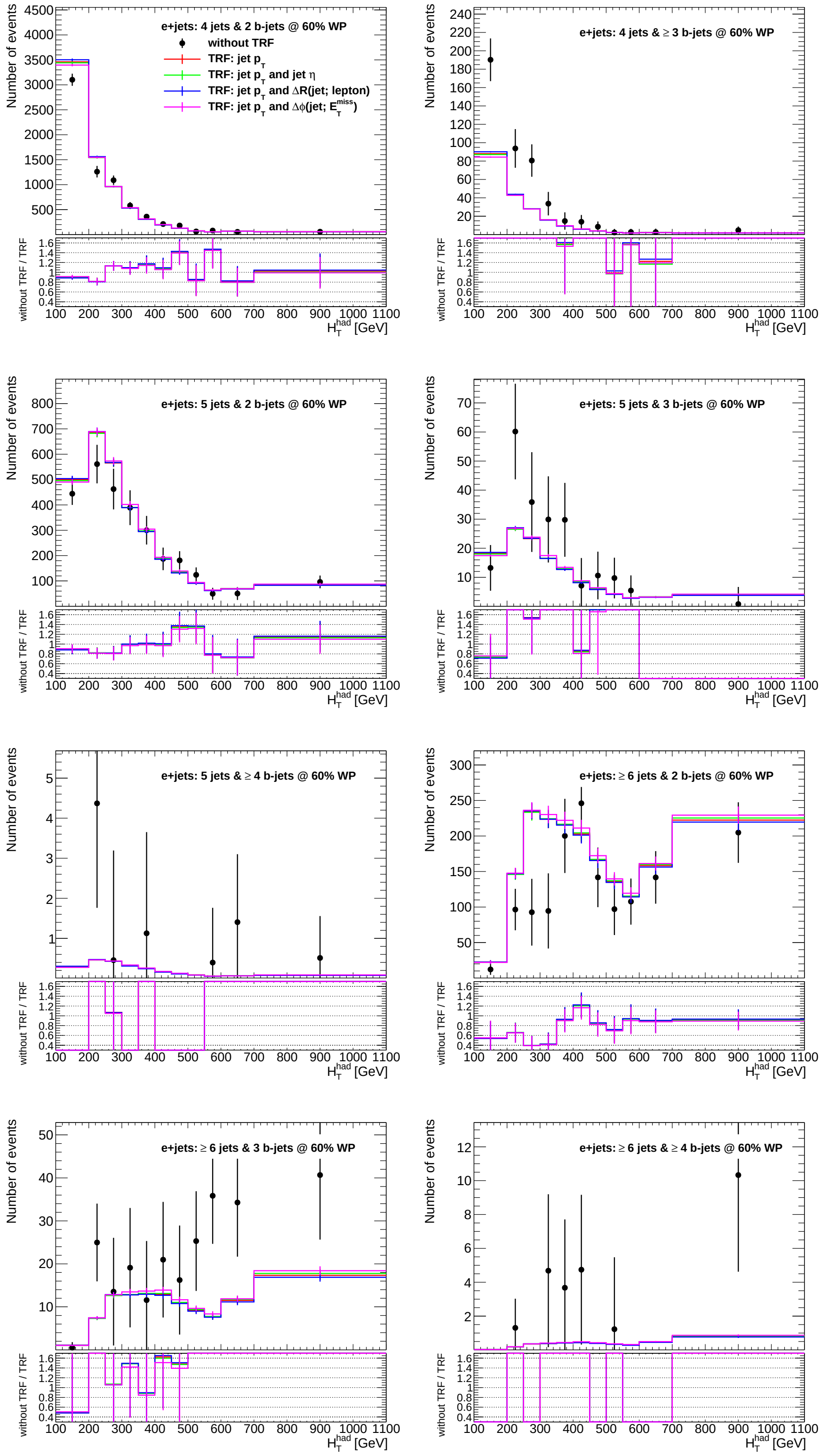


Figure B.2.: Comparison of the event yields of the fake estimation with and without the TRF method. Selection: electron + jets with a b-tagging WP of 60 %.

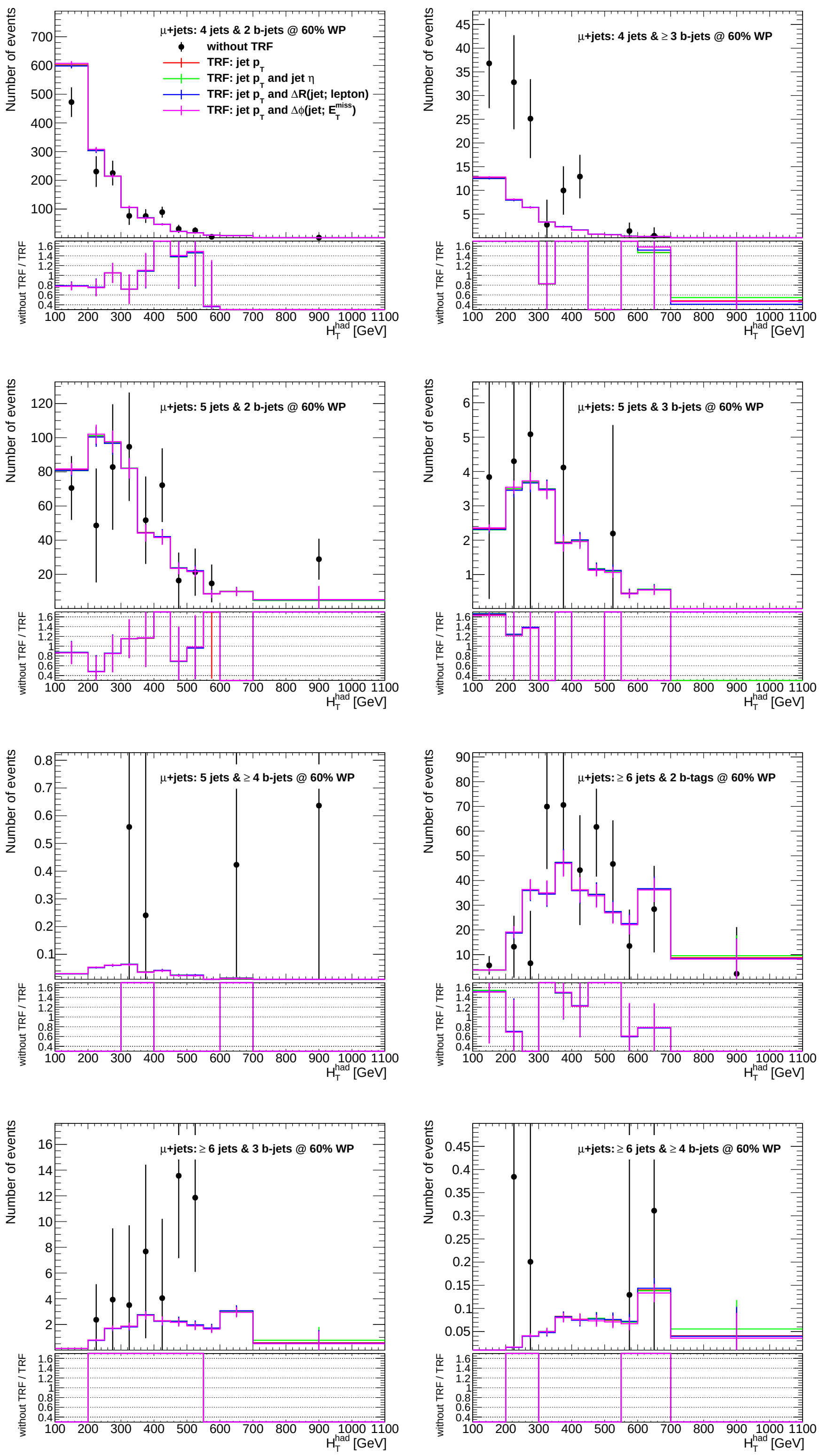


Figure B.3.: Comparison of the event yields of the fake estimation with and without the TRF method. Selection: muon + jets with a b-tagging WP of 60 %.

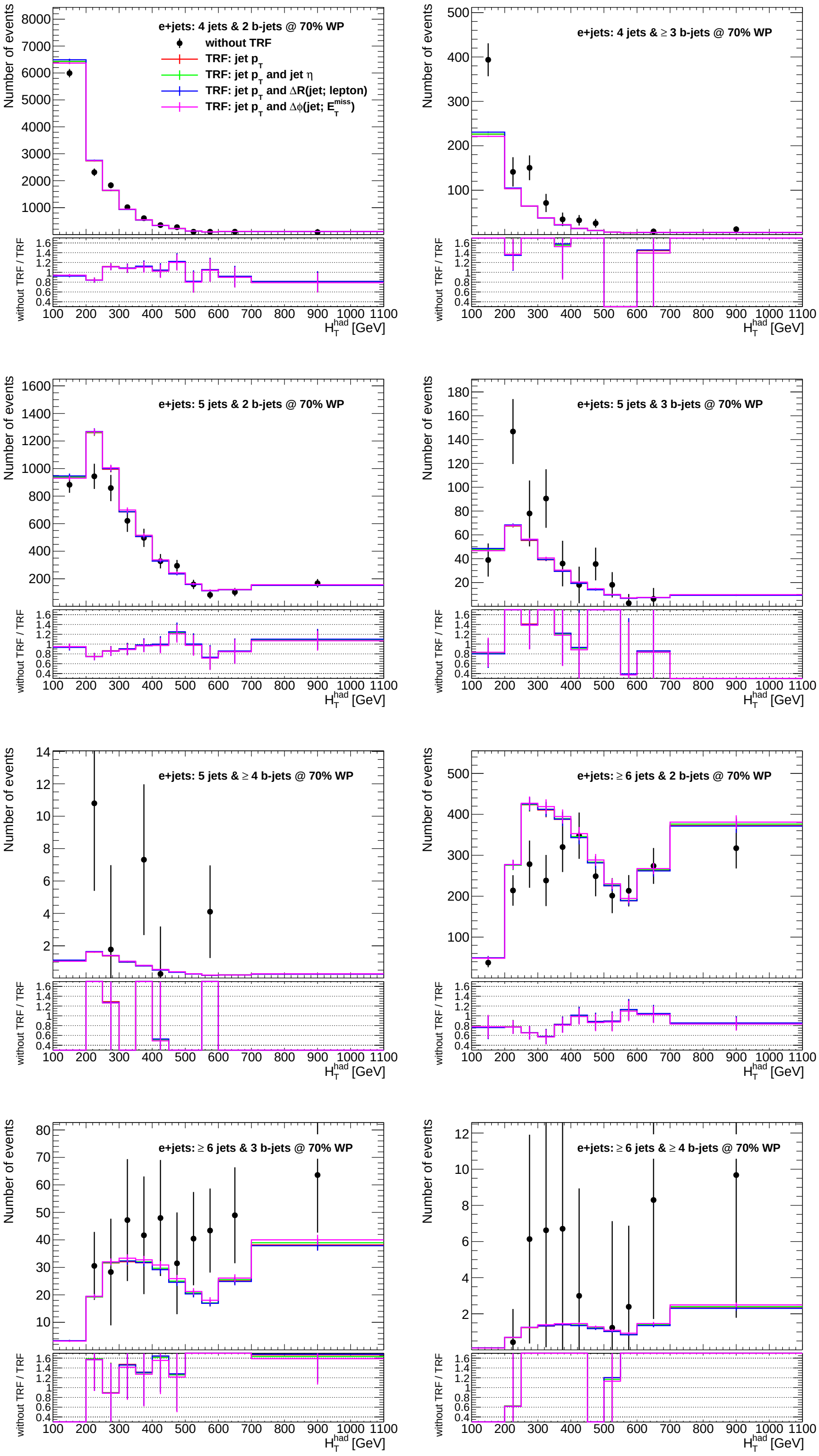


Figure B.4.: Comparison of the event yields of the fake estimation with and without the TRF method. Selection: electron + jets with a b-tagging WP of 70 %.

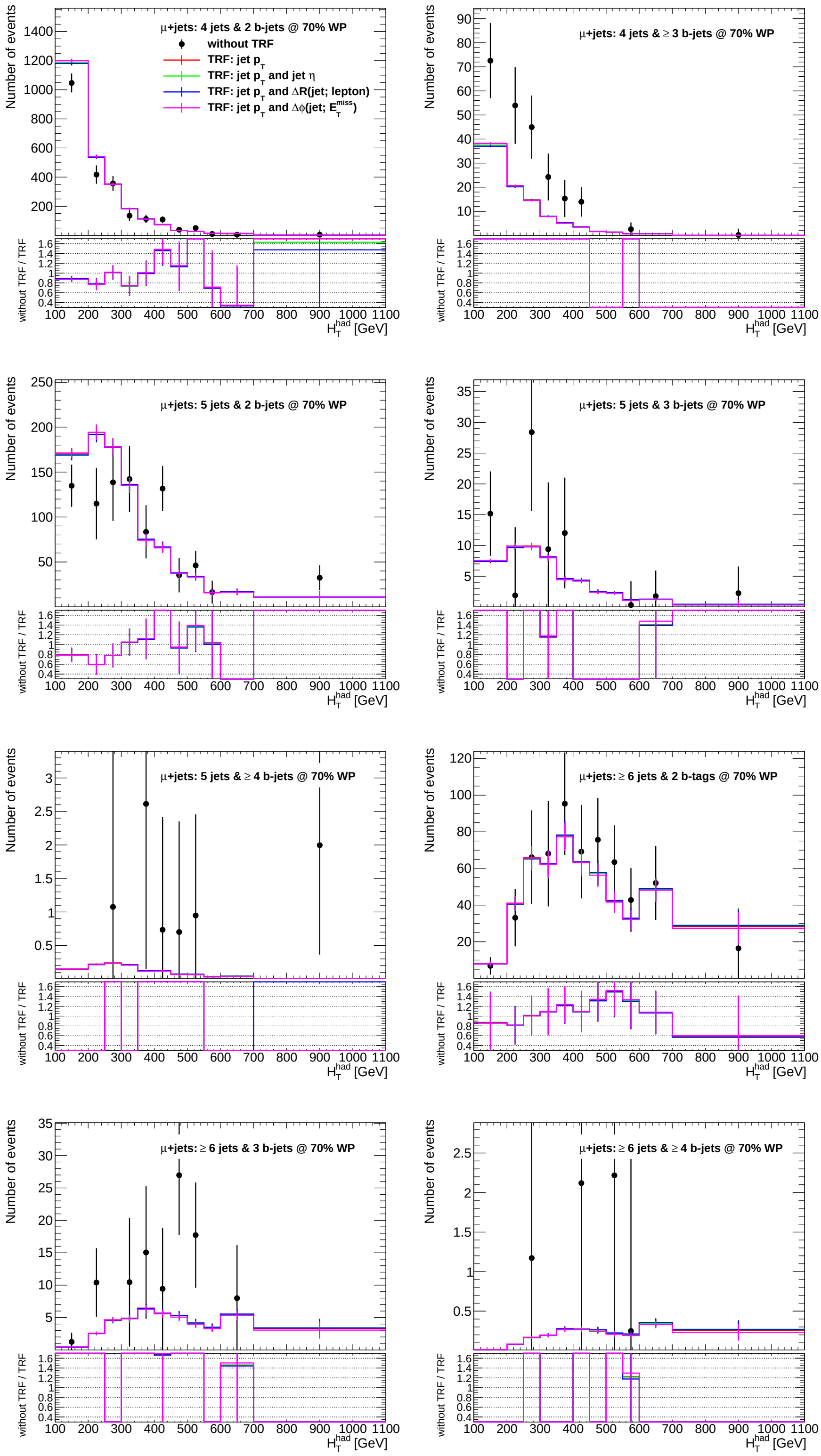


Figure B.5.: Comparison of the event yields of the fake estimation with and without the TRF method. Selection: muon + jets with a b-tagging WP of 70 %.

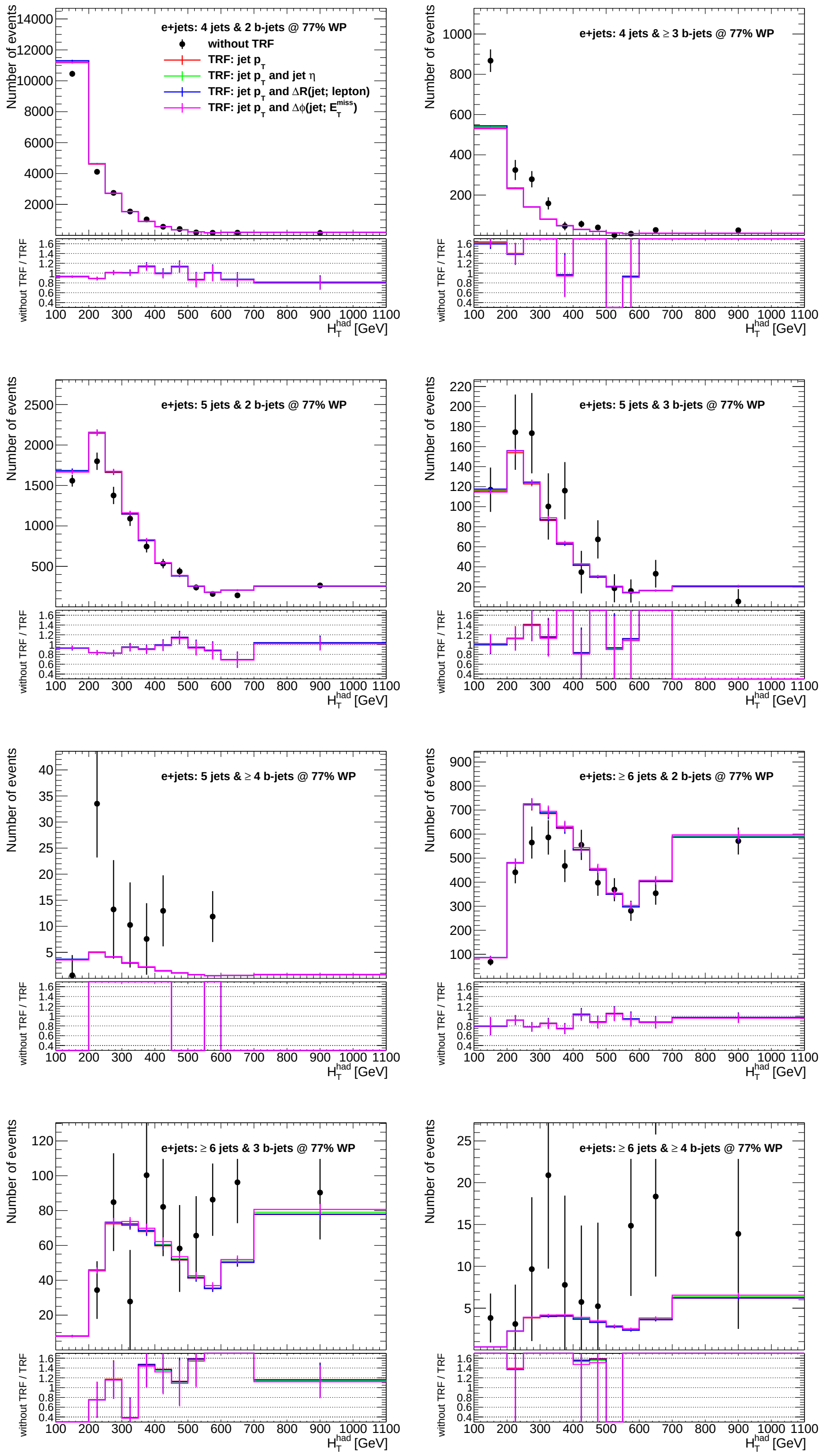


Figure B.6.: Comparison of the event yields of the fake estimation with and without the TRF method. Selection: electron + jets with a b-tagging WP of 77 %.

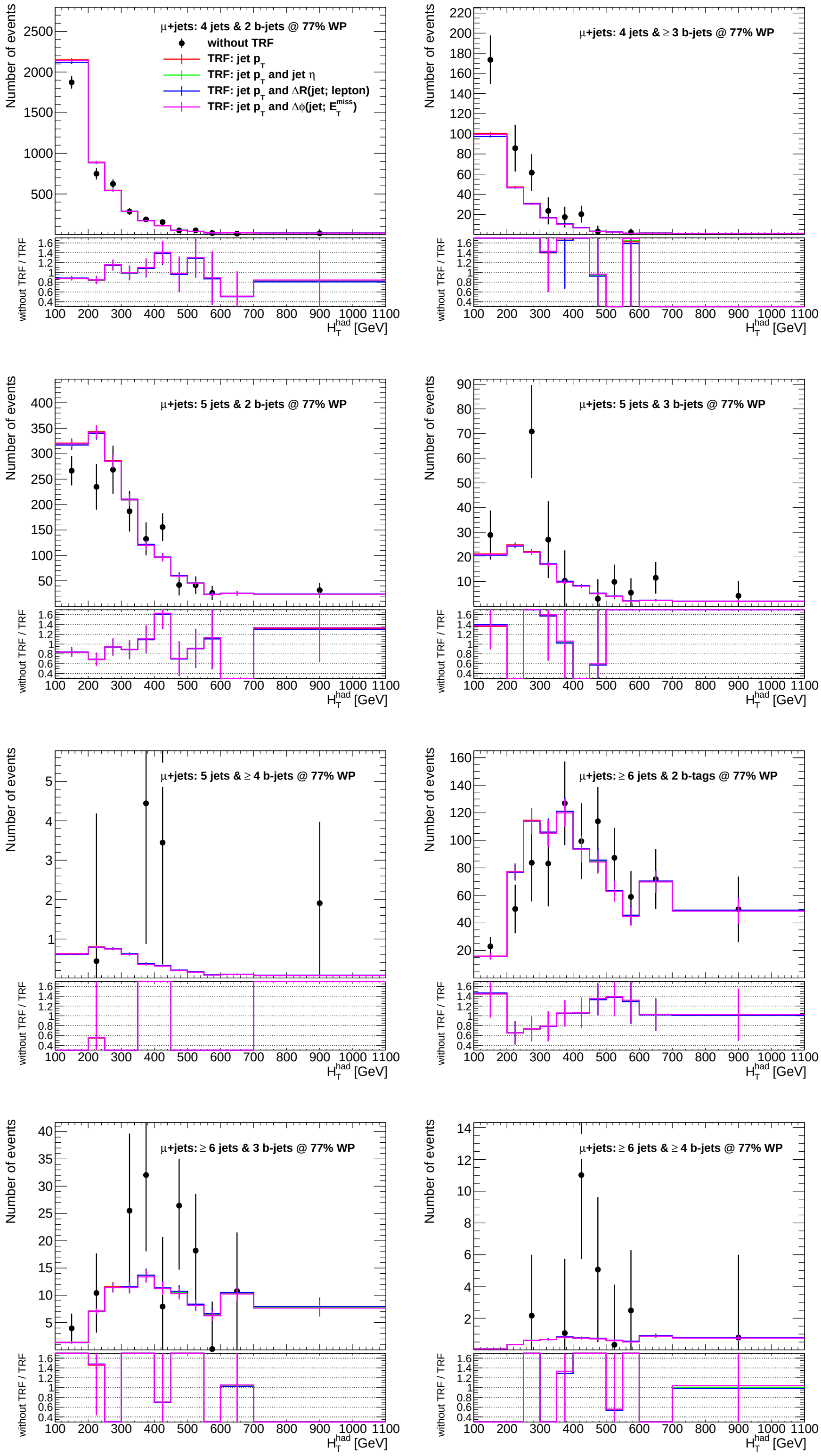


Figure B.7.: Comparison of the event yields of the fake estimation with and without the TRF method. Selection: muon + jets with a b-tagging WP of 77 %.

APPENDIX C

Input variables for the boosted decision trees

C.1. Input variables for the reconstruction BDTs

The input variables for the reconstruction BDTs are listed in Table C.1 for the single-lepton channel and in Table C.2 for the dilepton channel.

C.2. Input variables for the classification BDTs

The input variables for the classification BDTs are listed in Table C.3 for the single-lepton channel and in Table C.4 for the dilepton channel.

Variable	Definition	Region
Topological information from $t\bar{t}$		
$m_{t_{\mathrm{lep}}}$	mass of the leptonically decaying top quark	$\geq 6j$, $= 5j$
$m_{t_{\mathrm{had}}}$	mass of the hadronically decaying top quark	$\geq 6j$
$m_{t_{\mathrm{had}}^{\mathrm{incomp}}}$	mass of the hadronically decaying top quark using only one jet and one b-jet	$= 5j$
$m_{W_{\mathrm{had}}}$	mass of the hadronically decaying W boson	$\geq 6j$
$m_{W_{\mathrm{had}},b_{t_{\mathrm{lep}}}}$	mass of the hadronically decaying W boson and the b-jet from the leptonically decaying top quark	$\geq 6j$, $= 5j$
$m_{W_{\mathrm{lep}},b_{t_{\mathrm{had}}}}$	mass of the leptonically decaying W boson and the b-jet from the hadronically decaying top quark	$\geq 6j$, $= 5j$
$\Delta R(W_{\mathrm{had}}, b_{t_{\mathrm{had}}})$	ΔR between W_{had} and $b_{t_{\mathrm{had}}}$	$\geq 6j$, $= 5j$
$\Delta R(W_{\mathrm{had}}, b_{t_{\mathrm{lep}}})$	ΔR between W_{had} and $b_{t_{\mathrm{lep}}}$	$\geq 6j$, $= 5j$
$\Delta R(l, b_{t_{\mathrm{had}}})$	ΔR between the lepton and $b_{t_{\mathrm{had}}}$	$\geq 6j$, $= 5j$
$\Delta R(l, b_{t_{\mathrm{lep}}})$	ΔR between W_{had} and $b_{t_{\mathrm{lep}}}$	$\geq 6j$, $= 5j$
$\Delta R(b_{t_{\mathrm{had}}}, b_{t_{\mathrm{lep}}})$	ΔR between $b_{t_{\mathrm{had}}}$ and $b_{t_{\mathrm{lep}}}$	$\geq 6j$, $= 5j$
$\Delta R(q_{1,W_{\mathrm{had}}}, q_{2,W_{\mathrm{had}}})$	ΔR between the leading quark from the hadronic W decay and the subleading quark from the hadronic W decay	$\geq 6j$
$\Delta R(b_{t_{\mathrm{had}}}, q_{1,W_{\mathrm{had}}})$	ΔR between $b_{t_{\mathrm{had}}}$ $q_{1,W_{\mathrm{had}}}$	$\geq 6j$
$\Delta R(b_{t_{\mathrm{had}}}, q_{2,W_{\mathrm{had}}})$	ΔR between $b_{t_{\mathrm{had}}}$ $q_{2,W_{\mathrm{had}}}$	$\geq 6j$
$\min\Delta R(b_{t_{\mathrm{had}}}, q_{W_{\mathrm{had}}})$	minimum ΔR between $b_{t_{\mathrm{had}}}$ and a quark from the hadronic W decay	$\geq 6j$
$\min\Delta(Rb_{t_{\mathrm{had}}}, q_{W_{\mathrm{had}}})$ $-\Delta R(l, b_{t_{\mathrm{lep}}})$	minimum ΔR between $b_{t_{\mathrm{had}}}$ and a quark from the hadronic W decay minus ΔR between lepton and $b_{t_{\mathrm{lep}}}$	$\geq 6j$, $= 5j$
Topological information from the Higgs boson		
m_H	Higgs boson mass	$\geq 6j$, $= 5j$
$m_{H,q_{1,W_{\mathrm{had}}}}$	mass of the Higgs boson and the quark from the hadronically decaying W boson	$\geq 6j$, $= 5j$
$\Delta R(b_{1,H}, b_{2,H})$	ΔR between the b-jets from the Higgs boson decay	$\geq 6j$, $= 5j$
$\Delta R(b_{1,H}, l)$	ΔR between the leading b-jets from the Higgs boson decay and the lepton	$\geq 6j$, $= 5j$
$\Delta R(b_{1,H}, b_{t_{\mathrm{had}}})$	ΔR between the leading b-jets from the Higgs boson decay and the b-jet from the hadronically decaying top quark	$= 5j$
$\Delta R(b_{1,H}, b_{t_{\mathrm{lep}}})$	ΔR between the leading b-jets from the Higgs boson decay and the b-jet from the leptonically decaying top quark	$= 5j$

Table C.1.: Input variables used in the reconstruction BDTs in the single-lepton SRs.

Variable	Definition	Region
Topological information from $t\bar{t}$		
$\Delta m^{\text{tops}}_{l^{\pm},b}$	difference of $m^{\text{top}}_{l,b}$ and $m^{\text{antitop}}_{l,b}$	$\geq 4j$, $= 3j$
$m^{\text{top}}_{l,b}$	mass of the lepton and b-jet originating from the top quark decay	$\geq 4j$, $= 3j$
$m^{\text{antitop}}_{l,b}$	mass of the lepton and b-jet originating from the antitop quark decay	$\geq 4j$, $= 3j$
$\Delta R^{\text{top}}(l,b)$	ΔR between the lepton b-jet originating from the top quark decay	$\geq 4j$, $= 3j$
$\Delta R^{\text{antitop}}(l,b)$	ΔR between the lepton b-jet originating from the antitop quark decay	$\geq 4j$, $= 3j$
$\Delta\phi^{\text{tops}}(b,b)$	$\Delta\phi$ between b-jets originating from top quark decays	$\geq 4j^{*}$, $= 3j$
p^{b}_{T}	p_{T} of the b-jet originating from the top quark decay	$\geq 4j^{*}$, $= 3j^{*}$
$p^{\bar{b}}_{\text{T}}$	p_{T} of the b-jet originating from the antitop quark decay	$\geq 4j^{*}$, $= 3j^{*}$
$\Delta(\Delta R^{\text{tops}}(l,b))$	difference of $\Delta R^{\text{top}}(l,b)$ and $\Delta R^{\text{antitop}}(l,b)$	$\geq 4j^{*}$, $= 3j^{*}$
min $\Delta\eta(l,b)$	minimum $\Delta\eta$ between a lepton and b-jet	$\geq 4j^{*}$, $= 3j^{*}$
Topological information from the Higgs boson		
m_H	Higgs boson mass	$\geq 4j$, $= 3j$
$\Delta\phi(H,t\bar{t})$	$\Delta\phi$ between the Higgs boson and top quark pair	$= 3j$
$\Delta R(H,t\bar{t})$	ΔR between the Higgs boson and top quark pair	$\geq 4j$
$\Delta R(b_{1,H},b_{2,H})$	ΔR between the leading and subleading b-jet from the Higgs boson decay	$\geq 4j$, $= 3j$
$p^{b_{2,H}}_{\text{T}}$	p_{T} of the subleading b-jet from the Higgs boson decay	$= 3j$
min $\Delta R(b_H,l)$	ΔR between a b-jet from the Higgs boson and a lepton	$= 3j$
max $\Delta R(H,b)$	ΔR between the Higgs boson and a b-jet	$\geq 4j$
$\Delta R^{\text{tops}}(b,b)$	ΔR between b-jets originating from top quark decays	$\geq 4j$

Table C.2.: Input variables used in the reconstruction BDTs in the dilepton SRs. Variables indicated with a * are only from the reconstruction BDT without Higgs boson information.

Variable	Definition	Region
General kinematic variables		
$\Delta R_{bb}^{\text{avg}}$	average ΔR for all b-tagged jet pairs	$\text{SR}_{1,2,3}^{\geq 6j}$, $\text{SR}_{1,2}^{5j}$
$\Delta R_{bb}^{\max p_{\text{T}}}$	ΔR between the two b-jets with the largest vector sum p_{T}	$\text{SR}_{1,2,3}^{\geq 6j}$
$\Delta \eta_{jj}^{\max \Delta\eta}$	maximum $\Delta\eta$ between any two jets	$\text{SR}_{1,2,3}^{\geq 6j}$, $\text{SR}_{1,2}^{5j}$
$M_{bb}^{\min \Delta R}$	mass of the combination of two b-jets with the smallest ΔR	$\text{SR}_{1,2,3}^{\geq 6j}$
$M_{jj}^{\min \Delta R}$	mass of the combination of any two jets with the smallest ΔR	$\text{SR}_{1,2,3}^{\geq 6j}$, $\text{SR}_{1,2}^{5j}$
$N_{bb}^{\text{Higgs } 30}$	number of b-jet pairs with invariant mass within 30 GeV of m_{Higgs}	$\text{SR}_{1,2,3}^{\geq 6j}$, $\text{SR}_{1,2}^{5j}$
$H_{\text{T}}^{\text{had}}$	scalar sum of jet p_{T}	$\text{SR}_{1,2}^{5j}$
$\Delta R_{\text{lep},bb}^{\min}$	ΔR between the lepton and the combination of two b-jets with smallest ΔR	$\text{SR}_{1,2}^{5j}$
Aplanarity	1.5 times the 2nd eigenvalue of the momentum tensor built with all jets	$\text{SR}_{1,2,3}^{\geq 6j}$, $\text{SR}_{1,2}^{5j}$
H_1	2nd Fox-Wolfram moment computed from all jets and the lepton	$\text{SR}_{1,2,3}^{\geq 6j}$, $\text{SR}_{1,2}^{5j}$
Variables from reconstruction BDT		
BDT output	output of the reconstruction BDT	$^*\text{SR}_{1,2,3}^{\geq 6j}$, $^*\text{SR}_{1,2}^{5j}$
m_{bb}^{Higgs}	Higgs boson candidate mass	$\text{SR}_{1,2,3}^{\geq 6j}$, $\text{SR}_{1,2}^{5j}$
$m_{H,b_{\text{lep top}}}$	mass of Higgs boson candidate and b-jet from leptonic top candidate	$\text{SR}_{1,2,3}^{\geq 6j}$
$\Delta R_{bb}^{\text{Higgs}}$	ΔR between b-jets from the Higgs boson candidate	$\text{SR}_{1,2,3}^{\geq 6j}$
$\Delta R_{H,t\bar{t}}$	ΔR between Higgs boson candidate and $t\bar{t}$ candidate system	$^*\text{SR}_{1,2,3}^{\geq 6j}$, $^*\text{SR}_{1,2}^{5j}$
$\Delta R_{H,\text{lep top}}$	ΔR between Higgs boson candidate and leptonic top candidate	$\text{SR}_{1,2,3}^{\geq 6j}$
$\Delta R_{H,b_{\text{had top}}}$	ΔR between Higgs boson candidate and b-jet from adronic top candidate	$\text{SR}_{1,2,3}^{\geq 6j}$, $^*\text{SR}_{1,2}^{5j}$
Variables from LHD and ME calculations		
LHD	likelihood discriminant	$\text{SR}_{1,2,3}^{\geq 6j}$, $\text{SR}_{1,2}^{5j}$
MEM_{D1}	matrix element discriminant	$\text{SR}_{1}^{\geq 6j}$
Variables from b-tagging		
$w_{b\text{-tag}}^{\text{Higgs}}$	sum of b-tagging discriminants of jets from best Higgs boson candidate from the reconstruction BDT	$\text{SR}_{2,3}^{\geq 6j}$, $\text{SR}_{1,2}^{5j}$
B_{jet}^{3}	3rd largest b-tagging discriminant	$\text{SR}_{2,3}^{\geq 6j}$, $\text{SR}_{1,2}^{5j}$
B_{jet}^{4}	4th largest b-tagging discriminant	$\text{SR}_{2,3}^{\geq 6j}$, $\text{SR}_{1,2}^{5j}$
B_{jet}^{5}	5th largest b-tagging discriminant	$\text{SR}_{2,3}^{\geq 6j}$, $\text{SR}_{1,2}^{5j}$

Table C.3.: Input variables used in the classification BDTs in the single-lepton SRs. For variables from the reconstruction BDT, those indicated with a * are from the BDT using Higgs boson information, those with no * are from the BDT without Higgs boson information.

Variable	Definition	Region
General kinematic variables		
$M_{bb}^{\min}$	minimum invariant mass of a b-jet pair	$\mathrm{SR}_{1,2}^{\geq 4j}$
$M_{bb}^{\max}$	maximum invariant mass of a b-jet pair	$\mathrm{SR}_{3}^{\geq 4j}$
$M_{bb}^{\min\Delta R}$	invariant mass of the b-tagged jet pair with minimum ΔR	$\mathrm{SR}_{1,3}^{\geq 4j}$
$M_{jj}^{\max p_{\mathrm{T}}}$	invariant mass of the jet pair with maximum p_{T}	$\mathrm{SR}_{1}^{\geq 4j}$
$M_{bb}^{\max p_{\mathrm{T}}}$	invariant mass of the b-jet pair with maximum p_{T}	$\mathrm{SR}_{1,3}^{\geq 4j}$
$\Delta\eta_{bb}^{\mathrm{avg}}$	average $\Delta\eta$ between any two b-jets	$\mathrm{SR}_{1,2,3}^{\geq 4j}$
$\Delta\eta_{lj}^{\max}$	maximum $\Delta\eta$ between a jet and a lepton	$\mathrm{SR}_{2,3}^{\geq 4j}$
$\Delta R_{bb}^{\max p_{\mathrm{T}}}$	ΔR between the two b-jets with the largest p_{T}	$\mathrm{SR}_{2,3}^{\geq 4j}$
$N_{bb}^{\mathrm{Higgs}\ 30}$	number of b-jet pairs with invariant mass within 30 GeV of m_{Higgs}	$\mathrm{SR}_{1,2}^{\geq 4j}$
$N_{\mathrm{jets}}^{p_{\mathrm{T}}>40}$	number of jets with $p_{\mathrm{T}} > 40\,\mathrm{GeV}$	$\mathrm{SR}_{2,3}^{\geq 4j}$
$\mathrm{Aplanarity}_{b\text{-jet}}$	1.5 times the 2nd eigenvalue of the momentum tensor built with all b-jets	$\mathrm{SR}_{2}^{\geq 4j}$
$H_{\mathrm{T}}^{\mathrm{all}}$	scalar sum of p_{T} of all jets and leptons	$\mathrm{SR}_{3}^{\geq 4j}$
Variables from reconstruction BDT		
BDT output	output of the reconstruction BDT	$^{**}\mathrm{SR}_{1,2}^{\geq 4j}$, $\mathrm{SR}_{3}^{\geq 4j}$
m_{bb}^{Higgs}	Higgs boson candidate mass	$\mathrm{SR}_{1,3}^{\geq 4j}$
$\Delta R_{H,t\bar{t}}$	ΔR between Higgs boson candidate and $t\bar{t}$ candidate system	$^{*}\mathrm{SR}_{1}^{\geq 4j}$
$\Delta R_{H,l}^{\min}$	minimum ΔR between Higgs boson candidate and lepton	$\mathrm{SR}_{1,2,3}^{\geq 4j}$
$\Delta R_{H,b}^{\min}$	minimum ΔR between Higgs boson candidate and b-jet from top	$\mathrm{SR}_{1,2}^{\geq 4j}$
$\Delta R_{H,b}^{\max}$	maximum ΔR between Higgs boson candidate and b-jet from top	$\mathrm{SR}_{2}^{\geq 4j}$
$\Delta R_{bb}^{\mathrm{Higgs}}$	ΔR between b-jets from the Higgs boson candidate	$\mathrm{SR}_{2}^{\geq 4j}$
Variables from b-tagging		

Table C.4.: Input variables used in the classification BDTs in the dilepton SRs. For variables from the reconstruction BDT, those indicated with a * are from the BDT using Higgs boson information, those with no * are from the BDT without Higgs boson information while for those indicated with a ** both versions are used.

APPENDIX D

Performance of artificial neural networks

D.1. Neural network performance with low-level input variables

The performance of 14 NNs trained on low-level variables, where each of the networks was designed with a different choice of hyper-parameters, is presented in Table D.1.

Layers	Nodes/layer	Regularisation	Separation	AUC	Overtraining
1	30	none	9.89 %	35.8 %	+2.9 %
1	90	none	9.78 %	35.6 %	+2.5 %
3	30	none	9.04 %	34.2 %	+7.1 %
3	30	20 % dropout	9.75 %	35.6 %	+4.8 %
3	60	none	8.07 %	33.7 %	+7.3 %
3	60	20 % dropout	10.01 %	33.2 %	+4.5 %
4	30	none	9.13 %	34.1 %	+7.0 %
4	30	20 % dropout	10.02 %	35.7 %	+4.4 %
4	60	none	8.63 %	33.2 %	+9.2 %
4	60	20 % dropout	10.16 %	36.0 %	+5.5 %
5	30	none	9.27 %	34.4 %	+5.4 %
5	30	20 % dropout	10.10 %	35.9 %	+3.8 %
5	60	none	8.50 %	32.3 %	+11.1 %
5	60	20 % dropout	9.97 %	35.7 %	+5.2 %

Table D.1.: Performance of different NN configurations with low-level input variables. ‘AUC’ referrers to the area under the ROC curve of the test sample and ‘Overtraining’ denotes the absolute difference of the AUC between training and test sample.

All networks show a similar separation power, where four layered NN are slightly better. However, the performance of these networks is inferior to networks trained with high-level variables. A 20 % dropout in each hidden layer during training increased the separation power of larger networks and lead to a significant reduction of overtraining. The best NN architecture consists of 4 hidden layers with 60 nodes per layer and a 20 % dropout in each layer is applied.

D.2. Neural network performance with low-level input variables including *b*-tagging information

In this scenario, *b*-tagging information is added in the form of a TRF weight to six different NNs, see Table D.2. This additional information could increase the performance of all tested networks significantly. A larger dropout in each hidden layer during training reduced overtraining further but negatively affected performance. The best NN architecture is built of 4 hidden layers with 60 nodes per layer and a 40 % dropout in each layer is enforced.

Layers	Nodes/layer	Regularisation	Separation	AUC	Overtraining
4	60	20 % dropout	12.69 %	40.58 %	+8.2 %
4	60	40 % dropout	13.24 %	41.4 %	+6.2 %
4	90	40 % dropout	12.99 %	41.1 %	+6.5 %
4	90	60 % dropout	12.76 %	40.3 %	+4.5 %
5	60	20 % dropout	12.74 %	40.6 %	+7.2 %
5	60	40 % dropout	12.76 %	40.5 %	+4.6 %

Table D.2.: Performance of different NN configurations with low-level input variables including *b*-tagging information in the form of a TRF weight. ‘AUC’ referrers to the area under the ROC curve of the test sample and ‘Overtraining’ denotes the absolute difference of the AUC between training and test sample.

D.3. Neural network performance with a complete set of input variables

When extending the input to include low-level variables, *b*-tagging information, and the set of ten high-level variables, the best separation power could achieved, see Table D.3. Using Leaky ReLu functions as activation functions increased the performance of all networks. The best NN architecture is found to have 5 hidden layers with 90 nodes per layer and uses Leaky ReLu activation functions with a 40 % dropout in each layer.

Layers	Nodes/layer	Regularisation	Activation	Separation	AUC	Overtraining
1	60	none	sigmoid	16.76 %	42.1 %	+3.3 %
4	60	40 % dropout	ReLu	18.26 %	48.8 %	+4.1 %
4	60	40 % dropout	Leaky ReLu	18.26 %	48.9 %	+2.0 %
4	90	40 % dropout	ReLu	18.29 %	48.8 %	+4.5 %
4	90	40 % dropout	Leaky ReLu	18.97 %	49.2 %	+2.2 %
4	90	60 % dropout	ReLu	18.06 %	48.1 %	+3.2 %
5	60	40 % dropout	ReLu	18.36 %	48.7 %	+3.7 %
5	60	40 % dropout	Leaky ReLu	18.83 %	49.5 %	+2.3 %
5	90	40 % dropout	ReLu	18.25 %	48.5 %	+5.8 %
5	90	40 % dropout	Leaky ReLu	19.09 %	49.6 %	+2.1 %
5	90	60 % dropout	ReLu	17.82 %	47.9 %	+2.5 %

Table D.3.: Performance of different NN configurations with a complete set of input variables. 'Activation' shows the choice of the activation function, 'AUC' referrers to the area under the ROC curve of the test sample, and 'Overtraining' denotes the absolute difference of the AUC between training and test sample.