First evidence of standard model $pp \to t\bar{t}t\bar{t}$ production and performance studies of the ATLAS tile calorimeter for HL-LHC

Dissertation

to achieve the degree of
Docteur d'université en physique / Doktor der Naturwissenschaften

presented by

M.Sc. Lennart Rustige

Université Clermont Auvergne
École Doctorale des Sciences Fondamentales

and

Technische Universität Dortmund
Lehrstuhl für Experimentelle Physik IV

2020

1. Reviewer : Prof. Dr. Kevin Kröninger

2. Reviewer : Dr. Frédéric Deliot

3. Reviewer : PD. Dr. Andreas Meyer

President of the defence commitee: Dr. Frédéric Deliot

Members of the defence commitee: Prof. Dr. Kevin Kröninger

Dr. Romain Madar

Prof. Dr. Lucia Masetti

PD. Dr. Andreas Meyer

Dr. Sarah Porteboeuf-Houssais

Prof. Dr. Wolfgang Wagner

Advisors: Dr. Romain Madar

Prof. Dr. Kevin Kröninger

Date of submission: 15.07.2020

Date of defence: 02.10.2020

Grade: Magna cum laude

Diese Arbeit wurde von der Deutsch-Französischen Hochschule (DFH) finanziell gefördert.

Université franco-allemande
Deutsch-Französische Hochschule

Ce travail a été soutenu financièrement par l'Université franco-allemande (UFA).

This work was financially supported by the German-French University.

Langsam blühte, langsam reifte in Siddhartha die Erkenntnis, das Wissen darum, was eigentlich Weisheit sei, was seines langen Suchens Ziel sei. Es war nichts als eine Bereitschaft der Seele, eine Fähigkeit, eine geheime Kunst, jeden Augenblick, mitten im Leben, den Gedanken der Einheit denken, die Einheit fühlen und einatmen zu können. Langsam blühte dies in ihm auf, strahlte ihm aus Vasudevas altem Kindergesicht wider: Harmonie, Wissen um die ewige Vollkommenheit der Welt, Lächeln, Einheit.

(Siddharta - Herman Hesse)

For the paths you have shown me,
this is for you.

Karl Rustige

(1958 - 2018)

Abstract

Two analyses in the field of particle physics are presented in this document. First, studies on the performance of the reconstruction of muons using calorimeter information under the conditions of the High-Luminosity Large-Hadron-Collider (HL-LHC) phase of the ATLAS detector. Second, the search for the Standard Model (SM) simultaneous production of four top quarks using the full Run-II data set recorded by ATLAS. This data set corresponds to an integrated luminosity of $L = 139\,\text{fb}^{-1}$ of proton–proton collisions at a centre of mass energy of $\sqrt{s} = 13\,\text{TeV}$.

Here, the performance of the reconstruction of muons is probed for different pile-up scenarios, as those expected for the HL-LHC phase, and in light of different noise scenarios that emulate the loss of energy resolution and deterioration of detector acceptance due to ageing and irradiation of detector components. This study is conducted to test proposed detector upgrade scenarios before their implementation.

The search for SM like four top quark production presented here, focuses on the decay modes with two same sign or more leptons in the final state. The search for this process is, among other factors, motivated by the very large energies involved and by the fact that it is likely on the verge of being discovered with currently available data sets. The final results are obtained in a profile likelihood fit involving the outcome of a boosted decision tree trained to discriminate between signal and background. The fit results in a production cross section of $\sigma(pp \to t\bar{t}t\bar{t}) = 24^{+7}_{-6}\,\text{fb}$ [1], which corresponds to an observed (expected) significance of $Z = 4.3$ ($Z = 2.4$). This represents the first evidence for this process. The obtained result is compatible with the SM prediction [2] within 1.7 standard deviations.

Following first evidence, the possibility of reconstructing the four top quark system using a kinematic likelihood approach is developed and tested. These developments are performed with the KLFitter [3] tool set and yield an efficiency of correctly matching all four top quarks of $\epsilon = 33 \pm 4\,\%$ under optimal conditions in the single lepton final state.

Zusammenfassung

In diesem Dokument werden zwei Analysen auf dem Gebiet der Teilchenphysik vorgestellt. Erstens, Studien über die Performanz der Rekonstruktion von Myonen unter Verwendung von Kalorimeterinformationen während der HL-LHC Phase des ATLAS Detektors. Zweitens, die Suche nach der simultanen Produktion von vier Top-Quarks, wie vom Standard Modell (SM) prognostiziert, unter Verwendung des vollständigen Run-II-Datensatzes, der von ATLAS aufgenommen wurde. Dieser Datensatz entspricht einer integrierten Luminosität von $L = 139\,\text{fb}^{-1}$ von Proton–Proton Kollisionen bei einer Schwerpunktsenergie von $\sqrt{s} = 13\,\text{TeV}$.

Für diese Arbeit, wird die Performanz der Rekonstruktion von Myonen für verschiedene pile-up-Szenarien untersucht, wie sie für die HL-LHC-Phase erwartet werden, und im Hinblick auf verschiedene Rauschszenarien, die den Verlust der Energieauflösung und die Verschlechterung der Detektorakzeptanz aufgrund von Alterung und Bestrahlung der Detektorkomponenten nachbilden. Diese Studie wird durchgeführt, um vorgeschlagene Szenarien für die Aufrüstung des Detektors vor ihrer Implementierung zu testen.

Die Suche nach der vom SM prognostizierten Produktion von vier Top-Quarks, konzentriert sich auf die Zerfallsmoden mit zwei Leptonen gleicher Ladung oder mehr Leptonen im Endzustand. Die Suche nach diesem Prozess ist unter anderem durch die sehr hohen involvierten Energien motiviert und durch die Tatsache, dass derzeit verfügbare Datensätze möglicherweise für eine Entdeckung ausreichen. Die finalen Messergebnisse werden mit einem Profile-Likelihood-Fit erzielt der das Ergebnis eines Boosted-Decision-Trees beinhaltet, welcher darauf trainiert ist zwischen Signal und Untergrund zu unterscheiden.
Der Fit führt zu einem Wirkungsquerschnitt von $\sigma(pp \to t\bar{t}t\bar{t}) = 24^{+7}_{-6}\,\text{fb}$ [1], was einer beobachteten (erwarteten) Signifikanz von $Z = 4{,}3$ ($Z = 2{,}4$) entspricht. Dies repräsentiert den ersten Nachweis (Evidence) für diesen Prozess. Das erhaltene Ergebnis ist mit der SM Vorhersage [2] innerhalb von 1,7 Standardabweichungen kompatibel.

Auf diesen ersten Nachweis aufbauend, wird die Möglichkeit der Rekonstruktion des vier Top-Quark Systems unter Verwendung eines kinematischen Likelihood Fits entwickelt und getestet. Die Entwicklungen und Tests werden mit dem KLFitter Software Paket [3] durchgeführt und erzielen eine Effizienz der korrekten Zuordnung aller vier Top-Quarks von $\epsilon = 33 \pm 4\,\%$ unter optimalen Bedingungen für den Endzustand mit einem Lepton.

Résumé

Deux analyses dans le domaine de la physique des particules sont présentées dans ce document. Premièrement, des études sur les performances de la reconstruction des muons à l'aide des informations calorimétriques dans les conditions de la phase HL-LHC du détecteur ATLAS. Deuxièmement, la recherche de la production simultanée de quatre quarks top en utilisant l'ensemble complet des données Run-II enregistrées par ATLAS. Cet ensemble des données correspond à une luminosité intégrée de $L = 139\,\text{fb}^{-1}$ des collisions protons-protons à un énergie dans le centre de masse de $\sqrt{s} = 13\,\text{TeV}$.

La performance de la reconstruction des muons est sondée pour differentes conditions de prise de données, en particulier avec des nombres de collisions parasites importants tels qu'attendu pour la phase du HL-LHC. La performance est également sondé et en vue de différents scénarios des bruits qui émulent la perte de résolution énergétique et la détérioration de l'acceptation du détecteur due au vieillissement et à l'irradiation des composants du détecteur. Cette étude est menée pour tester les scénarios proposés de mise à jour du détecteur avant leur mise en œuvre.

La recherche de la production de quatre quark top, prédite par le modèle standard (SM), présentée ici, se concentre sur les modes de désintégration avec deux leptons de même signe ou plusieurs leptons dans l'état final. La recherche de ce processus est, entre autres facteurs, motivée par les très grandes énergies impliquées et par le fait qu'il est potentiellement sur le point d'être découvert avec l'ensemble des données actuellement disponibles. Les résultats finaux sont obtenus dans l'ajustement d'une fonction de vraisemblance profilée impliquant le résultat d'un *boosted decision tree*, entraîné à discriminer entre le signal et les bruits de fond. L'ajustement donne une section efficace de $\sigma(pp \to t\bar{t}t\bar{t}) = 24^{+7}_{-6}\,\text{fb}$ [1], ce qui correspond à une significance observée (attendue) de $Z = 4{,}3$ ($Z = 2{,}4$). Cela correspond à la première evidence de ce processus. Le résultat obtenu est compatible avec la prédiction du SM [2] à 1,7 écart-type près.

Après la première evidence, la possibilité de reconstruire le système des quatre quark top en utilisant une approche de vraisemblance cinématique est explorée. Les études sont effectuées dans l'état final avec un seul lepton avec l'outil KLFitter [3] donnant une efficacité de correspondance correcte des quatre quarks top de $\epsilon = 33 \pm 4\,\%$ dans des conditions optimales.

Contents

1 Introduction

The search for the standard model like production of four top quarks belongs to the vast field of particle physics. This field revolves around the question what the most fundamental building blocks of nature are and how they interact to form the world around us. Our current understanding of this field is encoded in a comprehensive theoretical framework called the Standard Model (SM) of particle physics. This model has been devised over the course of the last century and has since been tested and confirmed in countless experiments over many decades.
Up to now, three of the four known fundamental forces in nature are described in the SM, where efforts to include gravity into the framework have so far not been proven experimentally. Furthermore, the building blocks of the SM, the elementary particles, only make up an estimated 5 % of the mass of the universe, where the remaining parts come from dark matter and energy. Finally, the baryon asymmetry that allows for any matter based life in the universe cannot be sufficiently explained by the current model, leaving this existential question unanswered. Therefore, and because humankind has always striven to find a more unified view of the world, Beyond Standard Model (BSM) unifications are devised that try to address the open questions of the field while fully incorporating the well tested SM.

Following the ideals of the scientific method, it is thus the work of experimentalists to determine the validity of the proposed scenarios and to measure, through controlled experimentation, the free parameters of the SM with high precision. In order to study the properties of nature at such minuscule scales, very large energies are needed. For this reason, the field is sometimes referred to as High Energy Physics (HEP) and these energies are typically released in particle collisions with very large particle momenta. The currently largest accelerator complex producing such collisions is the Large Hadron Collider (LHC) at the European Organization for Nuclear Research (CERN). As part of the experiments, these collisions need to not only be produced but rather also be measured, which is done with gargantuan machines called detectors. There are four major detectors around the LHC, the largest of which is called ATLAS. Detectors cannot observe short-lived particles, such as the top quark, directly, but rather rely on measuring their decay products that are stable at the scale of the detector and tracing those back to the original short-lived particle.

In order to probe the validity of the SM or of BSM scenarios in increasingly extreme areas of the phase space, experimentalists look for ever more elusive processes and

especially for those that behave differently for different scenarios. One such process is the very rare simultaneous production of four top quarks, predicted in both the SM as well as in BSM extensions.

If found, this process would represent the most energetic topology discovered at the LHC and, as part of the top quark sector, it plays a decisive role in many BSM scenarios. Searches for this process at ATLAS have therefore already begun in 2012 and have so far been focusing on probing those BSM scenarios where the probability for four top quark production is significantly enhanced. With the full data set of collisions recorded with the ATLAS experiment between 2015 and 2018, however, the observation of this process may come within reach, making this research one of the top priorities for the collaboration.
The main analysis presented in this document consequently represents precisely the search for four top quark production with the ATLAS detector in multilepton final states. As a result of these efforts, first evidence for four top production has finally been found.

This represents a very important result, but naturally the field does not end here. In order to continue the research in HEP for years to come, the LHC will be upgraded at the beginning of 2025. This upgrade will enable the collider to considerably enhance the frequency of particle interactions and thus provide a markedly larger integrated luminosity than obtained thus far. This gives the upgraded collider the name High-Luminosity Large-Hadron-Collider (HL-LHC). Evidently, this higher frequency of interactions add a string of new challenges to the detectors, most noticeably in the sheer amount of data that needs processing. In addition, and after years of detector operation, certain components of ATLAS are either out-dated or at the end of their lifetime due to irradiation effects. The detector is therefore upgraded simultaneously to the LHC, so as to continue its very high quality measurements during the HL-LHC phase.

To determine which components of the detector to replace and to assess the expected performance of the renewed machine, is a crucial task in the preparation for this challenging phase.
Consequently, the performance of muon reconstruction, a difficult task to convert raw detector information to physics objects usable in analyses, is studied in detail for procedures that rely on measurements from the calorimetry system. These type of muon objects are currently not used in most analyses at ATLAS, so that an improved performance could undoubtedly improve the reconstruction efficiency of processes with

multiple muons in their final states. This is, for example, precisely the case for measurements of the four top quark system in multilepton final states.

However, before recording first collisions of the HL-LHC in 2027, there are many additional studies involving four top quarks to be performed. This is particularly true following the upcoming data taking period that is likely to record twice as much collisions as have been recorded thus far. Many of these studies rely on the kinematic properties of the four top quark system, such as the measurements of spin correlations or energy spectra of the top quarks. It is therefore paramount to explore ways of accessing these kinematic properties, one of which is called event reconstruction.

Studies are therefore presented here that represent the first attempt of reconstructing the four top quark system using a kinematic likelihood approach. This event reconstruction technique is initially tested for the single lepton and not the multilepton final state of the four top quark system, due to experimental obstacles in disentangling the momenta of multiple neutrinos.

As mentioned above, this work represents experimental studies that try to measure parameters of the SM, so that a brief introduction to the components of this model and to specific properties of the top quark sector is given in chapter 2.

Following that the search for four top quark production is the main analysis of this doctorate, more detailed deliberations on the nature of this process are presented in chapter 3.

The introduced models and their properties are put to the test using collisions from the LHC measured by the ATLAS detector, so that a brief introduction to these machines is provided in chapter 4.

The measurement process is very complex and indeed relies on combining timed spatial information of energy deposits to reconstruct the properties of particles or showers of particles originating from the collisions that interact with the detector. A brief overview of the routines and tools used to do so is given in chapter 5.

The studies of the reconstruction performance for muons during the HL-LHC phase is presented in chapter 6. This chapter is followed by a comprehensive report on the first evidence of four top quark production in chapter 7.

As part of this search, extensive tests have been performed on the final fit model. These tests represent a main contribution of the author and a non-standard approach in ATLAS which can be applied to other analyses than the search for four top quarks. These tests are presented in chapter 8. Following the obtained first evidence for this process, the studies on the possibility of reconstructing the four top quark system

are described in chapter 9. Finally, a summary of the presented studies is given in chapter 10.

Contributions of the Author

The field of particle physics is a very collaborative discipline, with immense collaborations, groups, sub groups and teams. As a consequence, the work presented in this dissertation is not the exclusive work of the author, so that the main contributions to the chapters containing the results of analyses are listed here.

Muon Reconstruction under HL-LHC Conditions The entirety of the studies presented here represent the work of the author. However, the studies presented in that chapter rely on simulated processes and their detector response, which is the product of hundreds, if not thousands, of researchers and engineers. This is true for all studies conducted throughout this doctorate.

Evidence of Four Top Quark Production This analysis is the work of a team of $\mathcal{O}(30)$ researchers and students. The author was one of three team members responsible for all analysis related software tasks, contributed with early lepton isolation optimisations and background and event classification based on simulation. Furthermore, the author contributed significantly to the development of the fit model and in building confidence in its outcome, which is summarised in the succeeding chapter.

Design and Validation of the Fit Model This chapter represents the summary of the above mentioned studies on the fit model and consists solely of the work of the author.

Reconstruction of the Four Top Quark System The analysis presented here represents the work of the author, mostly accomplished during a research stay at TU Dortmund University as part of the cotutorship programme of this doctorate.

The work and contributions of this doctorate lead to one publication and one internally reviewed report:

- Lennart Rustige and Romain Madar. 'Study of noise and efficiency loss effects on calorimeter-tagged muons for HL-LHC.' ATL-TILECAL-INT-2019-001. 2019.[1]
- ATLAS Collaboration. 'Evidence for $t\bar{t}t\bar{t}$ production in the multilepton final state in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector.' Eur. Phys. J. C 80 (2020).

The results showing the first evidence of $t\bar{t}t\bar{t}$ were first presented at LHCP2020 as documented in [4].

[1] The titles here and on most bibliography entries are hyperlinks to the corresponding documents in the digital version of this dissertation.

2 A brief Introduction to Particle Physics

The field of particle physics encompasses research into the most fundamental building blocks of nature. These building blocks are represented by elementary particles and their interactions. In the last century, this has lead to a comprehensive theoretical framework that is able to describe three out of the four currently known fundamental forces and which has been constantly improved upon since its inception. This framework is commonly referred to as the Standard Model (SM) of particle physics [5–7] and it has been shown to accurately predict the outcome of countless experiments in many different approaches.

So far, it has not been possible to include the fourth fundamental force, gravity, into this model. Additionally, experimental evidence indicates that large parts of the universe are not described by the SM since it does not include neither dark energy nor dark matter. Furthermore, not only experimental challenges still persist, but also conceptual ones, such as the so called hierarchy problem of the observed Higgs boson mass [8]. In order to describe these phenomenae or to resolve the conceptual difficulties, a series of Beyond Standard Model (BSM) scenarios are devised that extend the SM by additional symmetries, particles or dimensions. It is then the challenge of experimentalists to determine which of these scenarios accurately model nature.

To this end, experimentalists intent to measure the free parameters of the SM with high precision and they investigate processes that have a different predicted outcome for different models.
Such is the case for the search for four top quark production, the cross section of which is greatly enhanced in BSM scenarios and which is on the verge of observation following the SM prediction. Especially that last point can be used to probe the validity of the SM in an extreme region of the phase space and represents the main motivation for the search presented in this dissertation.

Since the SM represents the cornerstone of the field and because its prediction of the four top quark production is investigated here, a very brief introduction is given below that closely follows deliberations in [9].

Table 2.1: Elementary particles and interactions in the Standard Model after electroweak symmetry breaking

Fermions			Gauge Bosons	
Generation	Leptons	Quarks	Interaction	Force Carrier
1^{st}	ν_e, e^-	u, d	strong force	8 massless gluons g
2^{nd}	ν_μ, μ^-	c, s	weak force	massive Z and $W^\pm$ bosons
3^{rd}	ν_τ, τ^-	t, b	electromagnetic force	massless photon γ
massive scalar Higgs boson H				

2.1 Symmetry Groups and Particles

The representation of the SM is given in a quantum field and gauge theory approach. This means that the underlying principles of the model originate from symmetry groups and assumptions on gauge invariance. The combined symmetry group of the SM is $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$, where each subgroup results in a fundamental interaction and the subscripts refer to the charges or the type of said interaction. Here, C stands for the colour charge, L for the left-handedness of the interaction and Y for the hypercharge. In the quantum field theory approach, the interactions, or forces, are mediated by spin-1 force carriers called gauge bosons, which are the result of the generators of the symmetry groups.

The group that give rise to the strong interaction is the $SU(3)_C$, where the charge of the interaction is the colour charge (red, green or blue). There are eight corresponding massless gauge bosons called gluons g, which themselves carry a colour and anti-colour charge. They interact with all colour-charged particles and thus also with each other.

The group that gives rise to the electroweak (EW) interaction is $SU(2)_L \otimes U(1)_Y$, generating the $W_{1,2,3}$ and B gauge bosons. These bosons are, however, massless and non-observable at currently available energies. Instead, the massive $W^\pm$ and Z bosons[1] are observed for the weak interaction and the massless photon γ for the electromagnetic interaction.

Following the gauge invariance principle massive particles are not allowed, so that an additional mechanism needs to give mass to observable elementary particles. For the SM, this is achieved by introducing a complex scalar field that gives rise to the massive spin-0 Higgs boson H. The vacuum expectation value of this scalar field is non zero and it is thus not invariant under gauge rotations leading to a spontaneous breaking of the EW symmetry that in turn results in the observed masses of the H,Z and $W^\pm$ bosons.

[1] The common superscript 0 is omitted for brevity here and in the following.

The symmetry-broken representation of the SM is given by the $SU(3)_C \otimes U(1)_{\text{QED}}$.

The SM does not only consist of interactions and their force carriers as described so far, but also of other elementary particles that interact through these forces. Among these are half-integer spin particles called fermions. In the current model, there are three generations of fermions that can be further categorised into colour charged quarks and colourless leptons.
Among these leptons, there are three charged leptons e, μ and τ and their corresponding anti-particles. A significant increase in mass is observed for higher generations, so that $m_e < m_\mu < m_\tau$.
Leptons are grouped in isoweak doublets, where the corresponding partners to the charged leptons are the neutrinos ν_e, ν_μ and ν_τ. The masses of neutrinos have not yet been determined and it remains an open question if they represent Maiorana or Dirac particles[2].
There are nine types of colour charged quarks and antiquarks in the SM. These can be grouped into up type (up, charm, top) or down type (down, strange, bottom) quarks according to their electric charge. As for the charged leptons, also the masses of the quarks increase for higher generations, so that the top quark is the most massive quark in the SM.

Quarks and charged leptons gain their mass through the Yukawa interaction which effectively couples the fermion field with the Higgs field according to a coupling strength y_f that is proportional to the mass of the fermion. For quarks, the Lagrangian density $\mathcal{L}_Y$ of this interaction is given by

$$\mathcal{L}_Y = -Y^d_{ij}\overline{Q^I_{Li}}\Phi d^I_{Rj} - Y^u_{ij}\overline{Q^I_{Li}}\epsilon_{ij}\Phi^* u^I_{Rj} + h.c., \tag{2.1}$$

where the $Y^{u,d}_{ij}$ are complex matrices for the quark generations i and j, the Q^I_{Li} are left-handed quark doublets of generation i, ϕ represents the scalar Higgs field and the $(u/d)^I_{Rj}$ represent right-handed up- and down-type quark singlets of generation j. The *h.c.* stands for hermitian conjugate. The basis I represents the basis of the weak eigenstates of the quark doublets and singlets and ϵ_{ij} is the anti symmetric tensor.

To derive the physical (propagating) states of the quarks after the symmetry breaking, the complex matrices $Y^{u,d}$ need to be diagonalised. This reflects a change from the basis of the weak eigenstates to the basis of mass eigenstates. This diagonalisation gives rise to the Cabibbo-Kobayashi-Maskaw (CKM) matrix $\mathbf{V}^{\text{CKM}}$ [11, 12]. With

[2] In fact it was only shown through neutrino oscillation in 1998 that neutrinos are not massless [10].

d,s,b describing the down, strange and bottom quark in their mass eigenstate, the rotation between the eigenstates is described by

$$\begin{pmatrix} d^I \\ s^I \\ b^I \end{pmatrix} = \begin{pmatrix} V_{\mathrm{ud}} & V_{\mathrm{us}} & V_{\mathrm{ub}} \\ V_{\mathrm{cd}} & V_{\mathrm{cs}} & V_{\mathrm{cb}} \\ V_{\mathrm{td}} & V_{\mathrm{ts}} & V_{\mathrm{tb}} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}, \tag{2.2}$$

where the elements $|\mathbf{V}_{ij}|^2$ can be associated to transition probabilities from quark flavour i to j.

The observable carriers of the weak and electromagnetic forces after the symmetry breaking represent superpositions of the generators from the symmetry group $SU(2)_L \otimes U(1)_Y$ of the electroweak unification before breaking the symmetry. The superpositions are given by

$$W^{\pm} = \frac{1}{\sqrt{2}} (W_1 \mp i W_2) \qquad \text{and} \tag{2.3}$$

$$\begin{pmatrix} \gamma \\ Z \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B \\ W_3 \end{pmatrix}, \tag{2.4}$$

where θ_W represents the Weinberg angle given by the ratio of the mass of the $W^{\pm}$ over the mass of the Z boson

$$\cos\theta_W = \frac{m_W}{m_Z}. \tag{2.5}$$

Given that the $W_{1,2,3}$ bosons interact with other particles with a coupling strength gT_3 (T_3 being the weak isospin) and that the B boson couples with a strength $g'Y_W$ (Y_W being the weak hypercharge), the couplings of the superpositions of these bosons to fermions after the symmetry breaking depends on the Weinberg angle.
This means that the coupling constant of $W^{\pm}$ is simply gT_3 as there is no mixing of the $W_{1,2,3}$ and B bosons in this case. For the Z boson and the photon γ, however, the couplings are given by

$$\frac{g}{\cos\theta_W}(T_3 - Q\sin^2\theta_W) \qquad \text{and} \qquad g\sin\theta_W, \tag{2.6}$$

respectively. Here, the electric charge Q plays a role in combination with the weak hypercharge.

A summary of the particles and forces of the SM after the spontaneous symmetry breaking is given in Table 2.1.

As mentioned above, the main topic of this thesis is the search for the simultaneous production of four top quarks with the ATLAS experiment, so that the top quark and intricacies of hadron collisions are discussed in more detail in the following.

2.2 The Top Quark

The most massive observed elementary particle of the SM is the top quark, which therefore opens the door to studying many interesting parameters of the SM or BSM scenarios. A small introduction to the properties of the top quark is therefore given here, where a more in depth discussion on the top quark in the context of the Large Hadron Collider (LHC) can be found in [13]. A summary of the state-of-the art research results on the top quark is given in [14].

The top quark was first discovered in $\bar{p}$–p collisions with the CDF and DØ experiments at Fermilab in 1995 [15–17]. Since then, the top quark has been an ardent topic in the field of High Energy Physics (HEP), an often used synonym for particle physics. This is especially the case since the beginning of data taking at the LHC, which is generally referred to as a *top-factory* due to the large number of produced top quarks following the large collision energies at this collider. The properties of the top quark have since been (and still are) measured in numerous analyses, where a combination by the particle data group of direct measurements of the mass of the top quark [14] yields $m_t = 172.76 \pm 0.30\,\text{GeV}$, with a corresponding total decay width of $\Gamma_t = 1.42^{+0.19}_{-0.15}\,\text{GeV}$.

Another important measure of any physics process is the probability with which this process occurs in an interaction. For the production of a particle or a set of particles in collisions at a hadron collider, this is the cross section σ, where a short discussion on the principles of this measure for top quarks follows below.

2.2.1 Production

When discussing interactions in the context of hadron collisions, i.e. of compound objects, different levels need to be taken into account. In hard-scattering processes, not the initially collided objects (such as protons in p–p collisions), but rather the internal elementary particles such as quarks and gluons, called partons, interact.
This interaction is then governed by the type of partons and by the centre-of-mass energy of the parton–parton system $\sqrt{\hat{s}}$. In order to properly predict the outcome of such interactions, it is therefore necessary to know how likely it is to find a specific parton i with a partial momentum x_i of the total momentum of the hadron. Precisely that is the information given by Parton Distribution Functions (PDFs), which are heavily

studied and constantly improved upon [18–21].

The predominant production mode of top quarks in p–p collisions is in $t\bar{t}$ pairs via the strong interaction (mediated by gluons), where representative Leading Order (LO), or tree-level, Feynman diagrams of $t\bar{t}$ production from gluon fusion and quark–antiquark annihilation are shown in Figure 2.1.

Following the deliberations above, the production cross section of $t\bar{t}$ production in p–p collisions are a convolution of the parton–parton cross section $\sigma^{ij\to t\bar{t}}$ and the PDF of protons. This PDF further depends on a factorisation scale μ_F, which determines the energy realms of pertubative Quantum Chromo Dynamics (QCD) and non pertubative effects. Additionally, a renormalisation scale μ_R is introduced for the pertubative calculation, so that the cross section can be calculated in finite order. The cross section is then given by

$$\sigma^{pp\to t\bar{t}}\left(\sqrt{s},m_t\right) = \sum_{i,j=q,\bar{q},g} \int dx_i dx_j f_i(x_i,\mu_F^2) f_i(x_j,\mu_F^2) \times \sigma^{ij\to t\bar{t}}\left(m_t,\sqrt{\hat{s}},x_i,x_j,\alpha_S(\mu_R^2),\mu_R^2\right), \tag{2.7}$$

where the cross section depends on the centre-of-mass energy of the collision $\sqrt{s}$ and the mass of the top quark m_t. The $f_{i,j}$ are the PDFs for the partons i and j and α_S represents the strong coupling constant.

The result of the computation of the inclusive production cross section of $t\bar{t}$ in p–p collisions at $\sqrt{s} = 13\,\text{TeV}$ at Next-to-next-to-leading Order (NNLO) including Next-to-next-to-leading logarithmic (NNLL) soft-gluon terms yields $\sigma(pp \to t\bar{t})_{\text{NNLO}} = 832^{+20}_{-29}\,(\text{scale}) \pm 35\,(\text{PDF} + \alpha_S)\,\text{pb}$ [22–24]. This result is very compatible with the most recent measurement in the lepton+jets final state using the full Run-II data set recorded by the ATLAS experiment. There, the cross section is measured to $\sigma(pp \to t\bar{t}) = 830 \pm 38\,\text{pb}$ [24].

The production of single top quarks occurs in such collisions via the EW interaction, where the predominant modes are the t-channel, the tW-channel and the s-channel. The corresponding LO Feynman diagrams are depicted in Figure 2.2 and the corresponding cross sections at Next-to-leading Order (NLO), or NNLO, computed with the HATHOR tool set [25, 26] following the prescriptions in [27] yield $\sigma(pp \to tq + pp \to \bar{t}q)^{\text{NLO}}_{t\text{-channel}} = 217.0^{+9.0}_{-7.7}\,\text{pb}$, $\sigma(pp \to tW + pp \to \bar{t}W)^{\text{NNLO}} = 71.1 \pm 3.8\,\text{pb}$ [28] and $\sigma(pp \to tq + pp \to \bar{t}q)^{\text{NLO}}_{s\text{-channel}} = 10.3 \pm 0.4\,\text{pb}$, respectively. Corresponding measure-

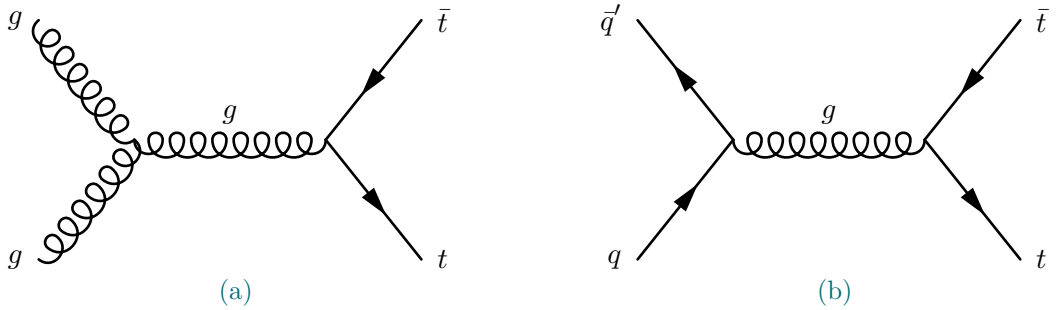


Figure 2.1: Representative Feynman diagrams of the LO production of $t\bar{t}$ pairs.

ments by the ATLAS collaboration can be found in [29, 30] and [31].

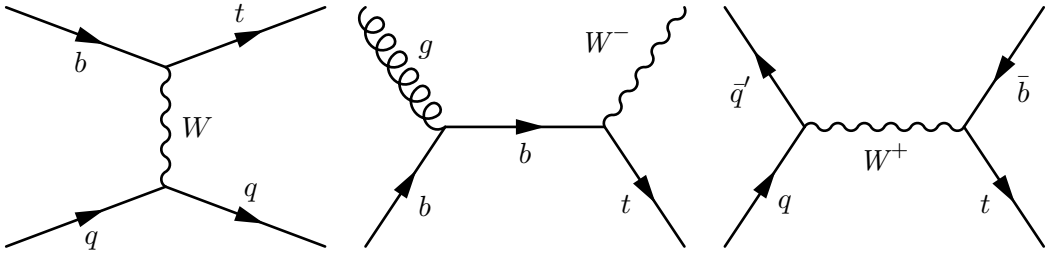
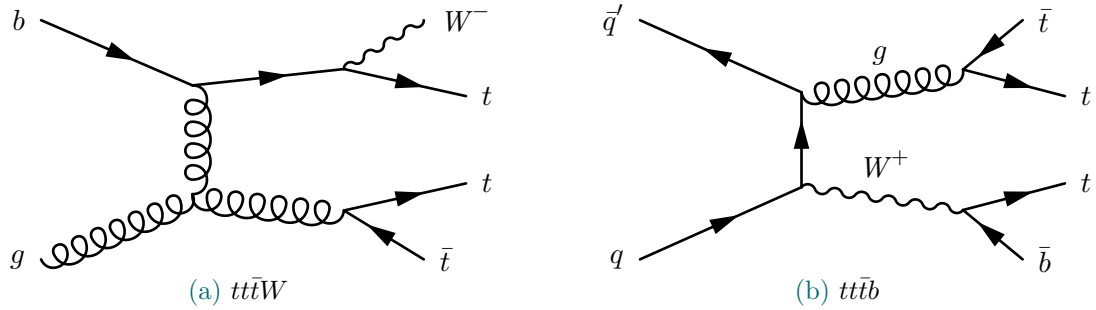


Figure 2.2: Representative Feynman diagrams of LO single top quark production

Finally, top quarks can also be produced in triplets or quartets, although with a production cross section that is several orders of magnitude lower than the pair or single top quark production. For the three top quark production, the cross section is estimated to $\sigma(pp \to t\bar{t}tW + pp \to t\bar{t}tq + pp \to t\bar{t}\bar{t}W + pp \to tb\bar{t}\bar{t}q) \approx 1.6\,\text{fb}$ using MadGraph5_aMC@NLO [32] and the NNPDF23 PDF set [33]. Representative LO Feynman diagrams of the three top quark production are shown for $tt\bar{t}W$ and $tt\bar{t}\bar{b}$ in Figure 2.3. The properties of the four top quark case, however, are discussed in more detail in chapter 3.

2.2.2 Decay

As a consequence to the very large mass of the top quark, it is the only quark that decays into a real W^+ boson which it does with the addition of a bottom quark. In fact, given the measured value for the V_{tb} component of the CKM matrix of $|V_{tb}| = 1.02 \pm 0.03$ [34], the top quark decays almost exclusively into a W^+ boson and b quark pair, where

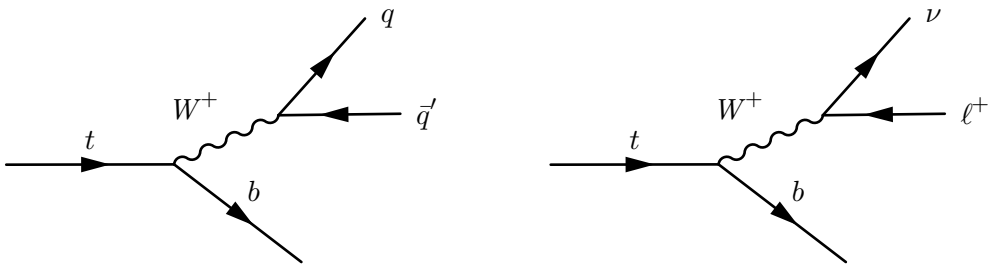


(a) $tt\bar{t}W$ (b) $tt\bar{t}b$

Figure 2.3: Representative Feynman diagrams of LO triple top quark production

the charge conjugate $\bar{t}$ quark decays into a W^- plus $\bar{b}$ pair.

Again due to its large mass, the lifetime of the top quark is very short at about $0.5 \cdot 10^{-24}$ s [14] for the above mentioned mass and width. This is particularly interesting considering the time scale of QCD interactions given by $\tau \simeq 1/\Lambda_{\mathrm{QCD}} = 3 \cdot 10^{-24}$ s, so that top quarks are likely to decay before hadronisation. The decay of the top quark then follows the decay modes of the W boson. For simplicity, a *leptonic* top quark decay refers to the case where the W boson of the decay of the top quark decays into a charged lepton and neutrino pair and a *hadronic* top quark decay refers to the case where that W boson decays into a quark–antiquark pair. The two modes are shown in Figure 2.4.



(a) *hadronic* (b) *leptonic*

Figure 2.4: Illustration of top quark decay modes depending on whether the W^+ boson decays leptonically or hadronically.

3 Four Top Quarks

3.1 Production

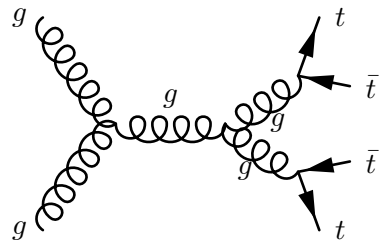
The most precise currently available prediction of the SM production cross section is given at NLO in QCD with NLO EW contributions in [2]. There, it was shown that especially the EW contributions that are usually subleading (w.r.t QCD contributions) indeed have a major impact, significantly enhancing previous predictions of the cross section purely at NLO in QCD [35].

Furthermore, the studies in [2] show a strong dependence of the predicted value on the choice of the renormalisation and factorisation scales as well as on the kinematics of the process. This leads to contributions of upward of 10 % for the $\mathcal{O}(\alpha_s^4\alpha)$, $\mathcal{O}(\alpha_s^3\alpha^2)$ and $\mathcal{O}(\alpha_s^2\alpha^3)$ contributions, where α is the coupling of the EW interaction.

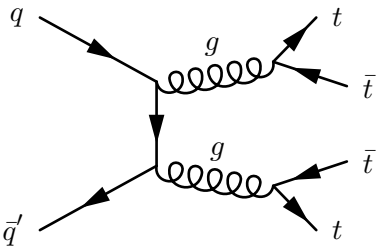
The production cross section for four top quarks in p–p collisions at 13 TeV that includes these contributions then amounts to $\sigma(pp \to t\bar{t}t\bar{t})_{\text{NLO}} \equiv \sigma_{4t\text{NLO}} = 12.0 \pm 2.4\,\text{fb}$ [2], where the uncertainty represents the uncertainty introduced by a variation of the above mentioned scales.

Representative Feynman diagrams of the production of four top quarks are given for QCD and EW processes in Figure 3.1, where particularly the production by an exchange of the H boson (Figure 3.1e) is interesting as it indicates a strong dependence between the production of four top quarks and the top-yukawa coupling y_t. Indeed, the cross section of this contribution scales with the fourth power of y_t as discussed in [36]. Moreover, due to the fact that $m_t > m_H$, this diagram does not depend on the width of the Higgs boson, opening potential direct measurements of y_t by measuring σ_{4t} at high precision.

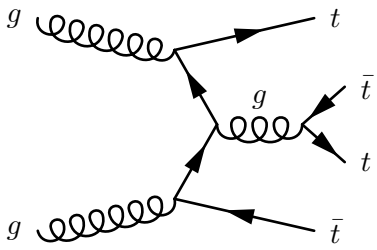
There are several BSM scenarios that predict values for the σ_{4t} that differ noticeably from the SM prediction. Among them are models using supersymmetry [37, 38], additional Higgs bosons [39, 40], scalar gluons [41] or additional dimensions [42].
In order to study possible BSM scenarios in a generic and model agnostic way, their influence on the four top quark production can be introduced as a $t\bar{t}t\bar{t}$ contact interaction model in an Effective Field Theory (EFT) approach. This has been done in a previous iteration of the search for four top quark production [43] using a data set that corresponds to an integrated luminosity of $L = 36.1\,\text{fb}^{-1}$ recorded by the ATLAS experiment.



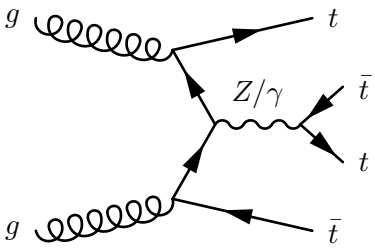
(a) QCD gluon fusion



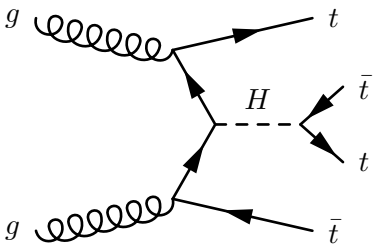
(b) QCD quark–antiquark annihilation



(c) QCD gluon splitting



(d) EW with Z boson or photon



(e) EW with H boson

Figure 3.1: Representative Feynman diagrams of LO QCD and EW production of four top quarks.

This search, and all iterations previous to that search, strongly focused on such BSM interpretations, as the available data set was unlikely to be sensitive to the very low cross section predicted by the SM. This, however, changes dramatically with the full Run-II data set recorded by ATLAS, so that the analysis strategy of the search presented in this thesis is exclusively tailored towards the SM prediction for the first time.

3.2 Decay Modes

Following the description of the decay of a single top quark in subsection 2.2.2, the deduction of the decay modes of the four top quark system is simply a matter of combinatorics. In this way, the four top quark system can either decay all-hadronically ($4h$), single- ($3h1l$), double- ($2h2l$), triple- ($1h3l$) or all-leptonically ($4l$) with the remaining top quarks decaying hadronically.

For studies at collider experiments, it is often advisable to combine decay modes according to the most prominent background sources. For this reason, the double-leptonical decay mode is first split into a decay mode that yields two leptons with opposite electric charge (opposite sign, OS) or with the same electric charge (same sign, SS). For the

search for four top quarks with the full Run-II data set of the ATLAS detector, the set of possible decay modes is split into two parts which are investigated independently[1].

The first group combines the $3h1l$ and the $2h2l$ OS decay modes so that the combined channel is called the 1LOS in the following. This channel represents approximately 57 % of the total branching ration $\mathcal{BR}$ of the decay of the four top quark system. The, by far most important, background process for this channel are $t\bar{t}$ events with additional b and light quarks in the final state. In comparison to the sheer size of this background in this channel, instrumental backgrounds are entirely negligible.

The second group combines the $2h2l$ SS, the $1h3l$ and the $4l$ decay modes into an analysis channel called SSML in the following. This channel represents approximately 13 % of the total decay width of the four top quark system, again not considering tau leptons[2].

Although the $\mathcal{BR}$ of this channel is considerably smaller than the $\mathcal{BR}$ of the 1LOS channel, it has been shown to be the more sensitive channel due to a much lower background contamination for the associated final state topologies [43].

The main analysis presented in this dissertation is the search for four top quarks in the SSML channel. The background processes to this channel are thus discussed in slightly more detail in the following section. The final state topologies of the different decay modes result in

$$\text{Final State Topology} = 4b + \begin{cases} \left.\begin{array}{ll} 8q & 4h \\ 6q + 1\ell + 1\nu & 3h1l \\ 4q + \ell^{\pm}\ell^{\mp} + 2\nu & 2h2l \text{ OS} \end{array}\right] \text{1LOS} \\ \left.\begin{array}{ll} 4q + \ell^{\pm}\ell^{\pm} + 2\nu & 2h2l \text{ SS} \\ 2q + 3\ell + 3\nu & 1h3l \\ 4\ell + 4\nu & 4l \end{array}\right] \text{SSML} \end{cases} , \quad (3.1)$$

where particles and antiparticles are grouped together unless specifically separated.

[1] However, the two analyses were devised and started at the same time and the same infrastructure and tools are applied as a baseline. Additionally, also basic optimisations at the object level are done in a combined effort so as to make sure that the areas of the phase space of the analyses are orthogonal and systematic uncertainties comparable. This significantly improves the prospects of a future combination of the two analyses.

[2] For this analysis, tau leptons are neither explicitly selected nor vetoed. Instead the leptons of a leptonically decaying tau lepton and the quarks of a hadronically decaying tau lepton are considered for the final state topology.

The all hadronic decay mode of the four top quark is not considered in any current analysis by ATLAS, even though it represents up to 30 %. This is due to the fact that the amount of $t\bar{t}$ or multijet events that result in a similar final state topology is simply overwhelming.

The corresponding $\mathcal{BR}$ for each of the decay mode is shown in Figure 3.2, where the magenta box represents the 1LOS and the teal box represents the SSML analysis channel.

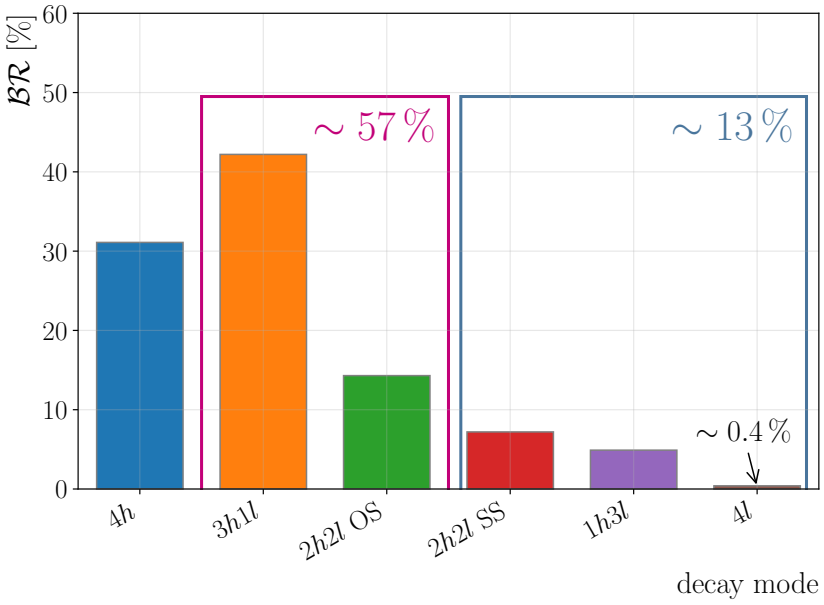


Figure 3.2: Branching ratios for the different decay modes of the four top quark system. The magenta box refers to the 1LOS and the teal box to the SSML analysis channel.

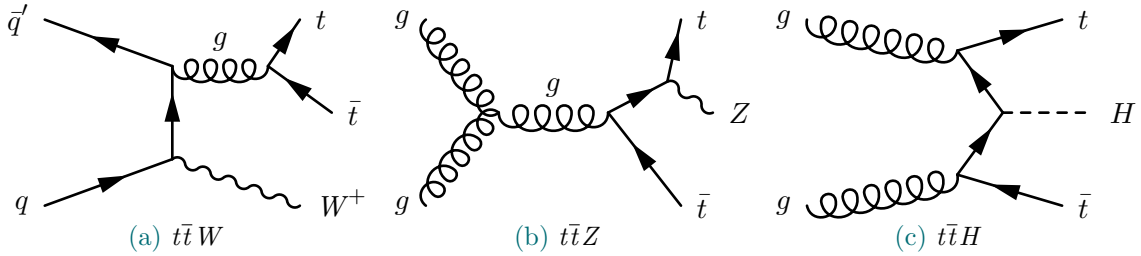
3.3 Background Processes

The main background processes that yield similar final state topologies as the decay of the four top quark system in the SSML channel are the $t\bar{t}$ plus vector boson production ($t\bar{t}V$), that is $t\bar{t}W$, $t\bar{t}Z$ and $t\bar{t}H$. Representative LO Feynman diagrams are shown for each of these processes in Figure 3.3.

The most relevant final state topologies of the three processes are given by

$$
\begin{aligned}
t\bar{t}\,W \to 2b + \begin{cases} 6q \\ 4q + 1\ell + 1\nu \\ 2q + \ell^{\pm}\ell^{\mp} + 2\nu \\ \ell^{\pm} + \ell^{\mp}\ell^{\mp} + 3\nu \end{cases}, \quad t\bar{t}\,Z \to \begin{cases} 2b + 4q \\ 2b + 2q + 1\ell + 1\nu \\ 2b + \ell^{\mp}\ell^{\pm} + 2\nu \end{cases} + \begin{cases} \ell^{\pm}\ell^{\mp} \\ q\bar{q} \\ 2\nu \end{cases}, \\
t\bar{t}\,H \to \begin{cases} 2b + 4q \\ 2b + 2q + 1\ell + 1\nu \\ 2b + \ell^{\mp}\ell^{\pm} + 2\nu \end{cases} + \begin{cases} ZZ \to 2\ell^{\pm} + 2\ell^{\mp} \\ WW \to \ell^{\pm}\ell^{\mp} + 2\nu \\ 2b \end{cases}
\end{aligned}
\tag{3.2}
$$

where it shows, that either additional b quarks, additional charged leptons ℓ or additional light quarks q are necessary in order to match the final state topology of the SSML channel.



(a) $t\bar{t}\,W$ (b) $t\bar{t}\,Z$ (c) $t\bar{t}\,H$

Figure 3.3: Representative LO Feynman diagrams of the most important background processes in the SSML analysis channel.

The inclusive production cross sections of these processes are considerably larger than σ_{4t} and amount to [44]

$$
\begin{aligned}
\sigma(pp \to t\bar{t}\,W)^{\text{NLO}}_{\text{QCD+EW}} &= 601 \pm 78\,\text{fb} \\
\sigma(pp \to t\bar{t}\,Z)^{\text{NLO}}_{\text{QCD+EW}} &= 839 \pm 94\,\text{fb} \\
\sigma(pp \to t\bar{t}H)^{\text{NLO}}_{\text{QCD+EW}} &= 507 \pm 42\,\text{fb}
\end{aligned}
\tag{3.3}
$$

for $\sqrt{s} = 13\,\text{TeV}$, where a mass of $m_H = 125\,\text{GeV}$ is assumed for the Higgs boson in the case of $t\bar{t}H$.

As will be shown in section 7.2, the event selection for a Signal Region (SR) does not exactly match the topology of the signal outlined in Equation 3.2 due to reconstruction and hadronisation effects. This means that also a larger part of events of the decay of $t\bar{t}V$ systems will enter this region. In addition to their larger production cross section,

this results in significant contributions of these processes to the optimised SR.

Besides the processes mentioned here, there are many additional processes that enter the signal region selection, such as the single and triple top production introduced in the previous chapter. More importantly, however, are events from $t\bar{t}$ +jets and single top production that enter the SR due to challenges in the measurement and reconstruction procedure (instrumental backgrounds), which are introduced at length in section 7.4.
In order to discuss these background types and in order to be able to perform any analysis (simulated or measured) at all, it is first necessary to describe the experiment, the simulation and the reconstruction process, which is therefore done in the next chapter and in chapter 5, respectively.

3.4 Previous Searches

The search for four top quark production presented in this thesis is not the first of its kind, as has been mentioned throughout this chapter, but rather follows roughly eight years of consecutive studies on this topic in ATLAS. This study, however, represents the first analysis in this collaboration that is exclusively dedicated to the four top quark production under the hypothesis of the SM. In addition to ATLAS, there have also been several searches for this process conducted by the CMS collaboration, where the last one has been published in January 2020.

This analysis by CMS has yielded the most stringent results on the production cross section and corresponding observed significance before the publication of the analysis presented here, while focusing on the same decay modes. The data set that has been used also corresponds to the full Run-II data set, albeit recorded by the CMS experiment, which corresponds to $L = 137\,\mathrm{fb}^{-1}$ of proton–proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$. The measured production cross section for this analysis reads [45]

$$\sigma(pp \to t\bar{t}t\bar{t}) = 12.6^{+5.8}_{-5.2}\,\mathrm{fb}, \tag{3.4}$$

which corresponds to an observed and expected significance of $Z_{\mathrm{obs}} = 2.6$ and $Z_{\mathrm{exp}} = 2.7$, respectively. This result is subsequently used to derive an upper limit on the measured Yukawa coupling between the top quark and the Higgs boson y_t over the coupling as predicted in the SM, y_t^{SM}. The CMS collaboration was thus able obtain a limit of $\left|y_t/y_t^{\mathrm{SM}}\right| < 1.7$ at the 95 % confidence level.

The previously strongest results by ATLAS are the combination of analyses of the 1LOS and the SSML analysis channels using $L = 36.1\,\text{fb}^{-1}$ worth of collision data. Using this data set, the amount of expected signal events is too small to directly measure the production cross section, so that upper limits are derived instead. These are measured to $\sigma(pp \to t\bar{t}t\bar{t}) < 49\,\text{fb}$ and $\sigma(pp \to t\bar{t}t\bar{t}) < 19\,\text{fb}$ for the observed and expected limit respectively [46], which corresponds to 5.3 and 2.1 times the standard model prediction at NLO in QCD, but without EW contributions.

4 Measuring Particles

In order to test the theoretical frameworks introduced in the previous chapter and in order to measure the free parameters of those models, very high energy interactions are studied. These can either come from astronomical processes, covered by the field of astroparticle physics, or they can be generated in particle colliders. In both cases, the outcome of these interactions in terms of particles that are stable at the scale of the experiment can then be measured by typically very large machines, called detectors. Since the nature of the models is probabilistic, and since it is usually necessary to back trace a series of interactions with the detector to draw a conclusion on the original hard scattering interaction, a very large amount of such collisions is required.
One such particle accelerator, the LHC, and one such detector, ATLAS, are introduced in the following.

4.1 CERN and the LHC

The largest laboratory dedicated to the field of particle physics is the European Organization for Nuclear Research (CERN)[1], an international organisation with 23 member states that was founded in 1954. With its main campus located right on the French-Swiss border near Geneva, CERN is active in the area of particle accelerators since the launch of the Synchro-Cyclotron in 1957 and has been at the forefront of accelerator based particle physics with increasingly more powerful and sophisticated accelerators. The latest of these is the LHC [47], a circular collider that achieved first collisions in 2010.

The LHC is currently the largest and most powerful particle accelerator at a circumference of approximately 26.7 km and it is designed for an operational period of 12 years from 2011 to 2023, after which it will be upgraded according to the proposed design of the High-Luminosity Large-Hadron-Collider (HL-LHC) [48]. A diagram of the accelerator complex at CERN is given in Figure 4.1, which shows the pre accelerators LINAC2 (LINAC3), BOOSTER, PS and SPS used to accelerate protons (Pb ions) sufficiently to enter the LHC ring. The current infrastructure therefore builds on the infrastructure from accelerators starting from the 1950s. Additionally, the four major experiments that use collisions from the LHC, namely ALICE [49], ATLAS [50], CMS [51] and LHCb [52] are depicted along ring.

[1] The acronym is derived from the name of the council that was tasked with founding the laboratory: *Conseil européen pour la recherche nucléaire*.

The LHC operates at a centre of mass energy of $\sqrt{s} = 7\ (8)\,\text{TeV}$ for 2011 and 2012, respectively, providing the ATLAS experiment with collision data corresponding to an integrated luminosity of $L = 5.46\ (22.8)\,\text{fb}^{-1}$ [53]. This data set was used to achieve one of the most important goals of the LHC: The discovery of the Higgs boson in 2012 [54, 55]. For the full Run-II period, that is from the beginning of 2015 to the end of 2018, the centre of mass energy was $\sqrt{s} = 13\,\text{TeV}$ and a data set corresponding to an integrated luminosity of $L = 147\,\text{fb}^{-1}$ was recorded with the ATLAS detector [53]. After evaluating the quality of the measured data, the full Run-II data set available to the ATLAS collaboration amounts to $L = 139\,\text{fb}^{-1}$.
An overview of the run conditions and time scales of the LHC and the upgraded HL-LHC is given in Figure 4.2.

For most of the time, the LHC is used to accelerate bunches of approximately 10^{11} protons in opposite directions which are brought to cross each other at multiple so called Interaction Points (IPs). Each proton beam consists of 2808 such bunches, which are spaced out at a distance corresponding to 25 ns at the nominal energy, with additional, larger, gaps between groups of bunches. During collision mode, the average rate of bunch crossings amount to 31.6 MHz.
Typically for a few weeks per year, however, heavy led (Pb) ions are accelerated instead of protons, so as to create a quark-gluon plasma with extremely high density and temperature during collisions. Studying this plasma is exceedingly interesting, especially because it resembles the conditions likely prevalent in the universe only micro seconds after the Big Bang.

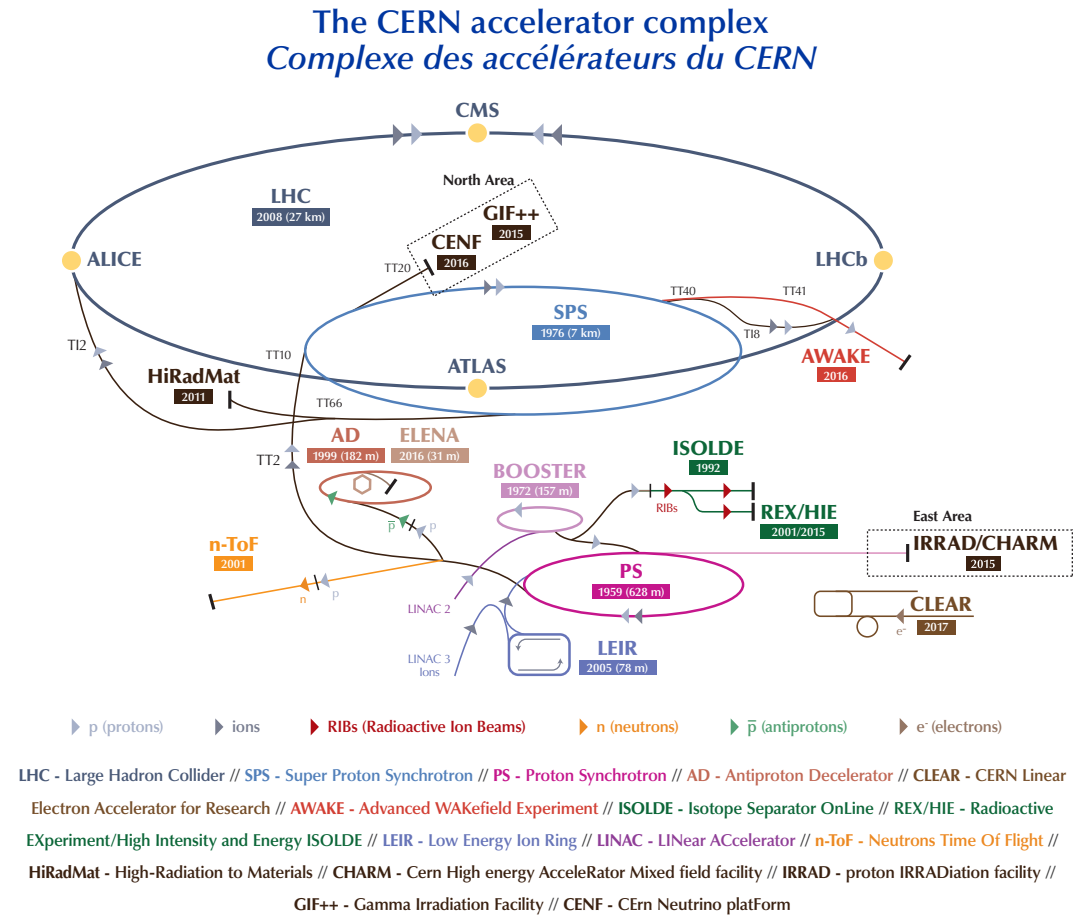


Figure 4.1: Schematic of the CERN accelerator complex [56].

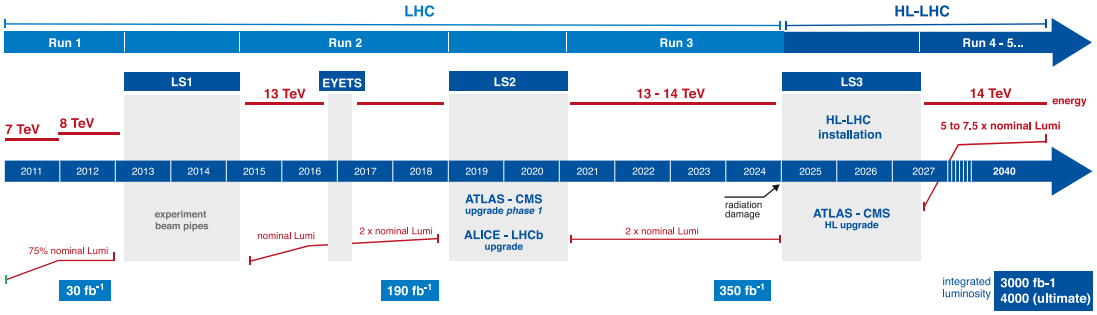


Figure 4.2: Schematic of the timescale of LHC and HL-LHC operations as well as expected integrated luminosities, centre of mass energies and upgrade and maintenance periods. With small changes taken from [57].

4.2 The ATLAS Detector

ATLAS is designed to be a multi purpose detector and it is thus constructed to cover a large variety of different physics programmes. Among these programmes are precision measurements of the SM as well as searches for BSM, with a strong focus on the production and decay of Higgs bosons and top quarks. Especially the discovery of the former has been the driving motivation for the design of the detector, which has been achieved in the summer of 2012 [54, 55].

4.2.1 Coordinate System and Variables

In order to consistently describe particle interactions with the detector or even just the geometry of the detector, a consistent coordinate system needs to be introduced. For ATLAS, the origin of such a coordinate system is set at the nominal IP at the centre of the detector. In Cartesian coordinates, the x-axis then points to the centre of the LHC, the y-axis points upwards and the z-axis is parallel to the beam axis and points in the counter-clockwise direction of the LHC as seen from above.
Seeing that the detector is mostly rotationally symmetric around the beam axis, it is often useful to use cylindrical coordinates. This system is then described by two angles, the polar angle θ in the xz plane, where $\theta = 0$ points into the positive direction of the z axis, and the azimuthal angle ϕ in the xy plane, where $\phi = 0$ coincides with the positive y direction.

The kinematic properties of reconstructed objects is given in terms of four-momenta p^μ, which represent the vector of energy E and the Cartesian spatial momentum $\vec{p}$ of the object in Minkowski space. With the rest mass m_0, the four-momentum is defined as

$$p^\mu = \begin{pmatrix} E \\ \vec{p} \end{pmatrix} = \begin{pmatrix} \sqrt{m_0^2 + |\vec{p}|^2} \\ \vec{p} \end{pmatrix}. \tag{4.1}$$

Again here, the spatial components of p^μ can be converted to a different coordinate system, where important properties are the transverse momentum p_T, i.e. the component of the momentum of the object that is perpendicular to the beam pipe (Equation 4.3), the azimuthal angle of the object (Equation 4.2) and the rapidity y or more often the pseudorapidity η (Equations (4.5) and (4.6)). Rapidity is used for hadron colliders instead of the polar angle as the colliding partons will very likely not have the same momentum in z, resulting in a boost in one of the z directions. Differences in rapidity,

however, are invariant under such a Lorentz boost so that it is unambiguous for any combination of parton momenta. Nevertheless, the rapidity also depends on the energy of the object and is therefore not a purely geometric measure. Consequently, it is often preferred to use the pseudorapidity η, differences of which are invariant under a Lorentz boost in case the object is massless or in case $p_{\mathrm{T}} >> m_0$. The pseudorapidity is approximatively related following Equation 4.7.

For many purposes it is interesting to determine a distance between the vectors given by the four-momenta of two objects (subscript 1 and 2). This can easily be expressed in terms of the difference in ϕ and η, where always the smallest angular difference is chosen for $\Delta\phi$ (Equation 4.9). The distance ΔR is then given by Equation 4.10.

Throughout this document, natural units are used, so that energies, momenta and masses are all given in units of energy (eV). Angles are expressed in radian, unless specified otherwise. The above described variables are given by

$$\phi = \arctan\left(\frac{p_x}{p_y}\right) \in (-\pi,\pi], \tag{4.2}$$

$$p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2}, \tag{4.3}$$

$$\theta = \arctan\left(\frac{p_{\mathrm{T}}}{p_z}\right) \in [0,\pi], \tag{4.4}$$

$$y = \frac{1}{2}\ln\left(\frac{E + p_z}{E - p_z}\right) \in (-\infty,\infty), \tag{4.5}$$

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \in (-\infty,\infty), \tag{4.6}$$

$$y \approx \eta - \frac{\cos\theta}{2}\left(\frac{m_0}{p_{\mathrm{T}}}\right)^2, \tag{4.7}$$

$$\Delta\eta = \eta_1 - \eta_2 \in (-\infty,\infty), \tag{4.8}$$

$$\Delta\phi = \arccos\left[\cos\left(\phi_1 - \phi_2\right)\right] \in [0,\pi), \tag{4.9}$$

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}. \tag{4.10}$$

Finally, two very important parameters of collider experiments are the centre of mass energy $\sqrt{s}$ and the integrated luminosity $\int \mathcal{L}dt$, which is generally shortened to L in

this document. The centre of mass energy is given by the sum of four momenta of the colliding particles

$$\sqrt{s} = \sqrt{\left(\sum_{i=1}^{n} p_i^{\mu}\right)^2}, \tag{4.11}$$

for n particles. This parameter is invariant under Lorentz transformation and indeed conserved, i.e. it remains the same before and after the collision. The integrated luminosity is, of course, the integral of the (instantaneous) luminosity $\mathcal{L}$. For two collinear intersecting particle beams with $N_{1,2}$ protons per bunch, the number of colliding bunches k, the frequency of bunch crossings f and the width of the beam in x and y, $\sigma_{x,y}$, the luminosity is given by

$$\mathcal{L} = \frac{fkN_1N_2}{4\pi\sigma_x\sigma_y}. \tag{4.12}$$

A precise knowledge of the integrated luminosity is crucial to compute the number of expected events N_{exp} of a given process with a cross section σ, which is given by $N_{\text{exp}} = \sigma L$.

4.2.2 The Detector in Run-II

The detector is built near rotationally symmetric around the beam pipe of the LHC and near point symmetric around the IP. This almost complete coverage in ϕ and the very large coverage in the η are crucial to minimise the loss of particles of interest due to detector acceptance effects. With a length of approximately 44 m and a height of about 25 m, the ATLAS detector is the largest of the detectors at the LHC in physical size and among the largest in the size of the collaboration of the experiment. This collaboration comprises about 3000 official authors from 181 institutions [58].

In order to measure the properties of a variety of different particles, different detector technologies are needed, that are stacked in shells around the beam pipe. These different shells, are the Inner Detector (ID), the Electromagnetic Calorimeter (EM), the hadronic calorimeter and the Muon System (MS), where each segment further consist of multiple sub detectors. A schematic cut away of the detector used in Run-I is given in Figure 4.3. The general concepts for each of the shells for Run-II conditions are outlined below, where in depth descriptions can be found in [50].

Inner Detector The main purpose of the ID is to measure the traces (tracks) of electrically charged particles and thus determine their momenta. These tracks can then be used to determine interaction and decay vertices which in turn are very important

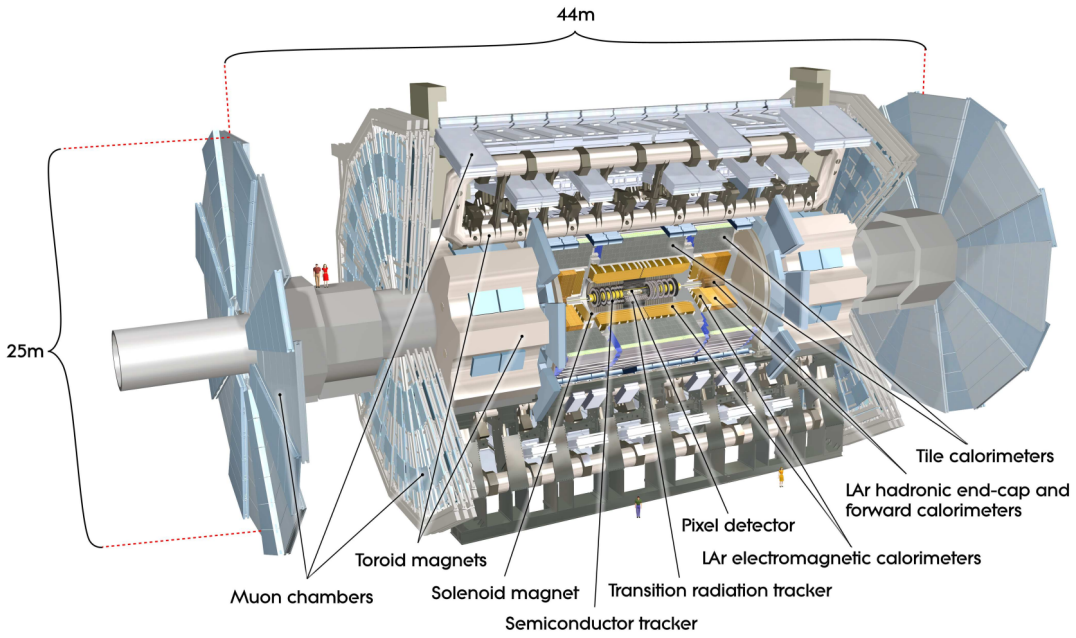


Figure 4.3: Schematic of the ATLAS detector under Run-I conditions. The overall layout remains essentially the same for Run-II, with the notable exception of the addition of the Insertable B-Layer (IBL). The graphic is taken from [50].

to differentiate the interaction of interest from other interactions in the same crossing of p–p bunches or to identify tracks that originate from B hadron decays.
The ID consists of the IBL, the Pixel Detector, the Silicon Microstrip Detector (SCT) and the Transition Radiation Tracker (TRT) sub detector and it is immersed in a magnetic field with a field strength of $B = 2\,\mathrm{T}$, generated by a solenoidal magnet. This magnetic field bends the path of charged particle, so that the curvature of the path of a particle through the ID can be used to estimate the momentum of said particle. The IBL [59] is the newest addition to the ID and it was installed during the Long-Shutdown (LS) 1 to cope with the significantly increased number of particle interactions encountered in Run-II with respect to the number of interactions in Run-I and to replace components that suffered radiation damage. In its current form, the sub detectors of this shell cover a range of $|\eta| < 2.5$.

Electromagnetic Calorimeter Surrounding the magnet system of the ID, lie the calorimetry shells which are separated into an electromagnetic and a hadronic calorimeter. The general idea of most calorimeter systems is to determine the energy of a given particle through complete absorption of that energy into the detector material.
The EM is mostly designed to measure the energies of any electron or photon, thus also including those induced by particle showers originating from quarks or gluons. This

sub detector uses the possibility to ionise liquid argon by particle showers induced from lead-plate absorbers. It further consists of different modules that cover different areas in η. The most central region of $|\eta| < 1.475$ is instrumented by the liquid Argon electromagnetic barrel, where the outer regions ranging from $1.375 < |\eta| < 3.2$ are instrumented by end-cap components called EMEC. The layers of active and absorber material are arranged in an accordion-like structure, which allows for a full coverage in ϕ with a uniform response and to place the readout electronics outside of the sensitive area of the EM. This is an important fact as any additional material in front of or within the sensitive area would lead to a bias in the energy measurement. For this reason, an additional layer of liquid Argon, called the Presampler, is introduced in front of the main EM, to estimate the energy loss that particles incurred by the beam pipe and inner detector material.

The actual EM is then further split in three layers with a different granularity in η for $|\eta| \leq 2.5$ and two layers otherwise. The first layer is called the strip layer and for the most central region each strip covers a region of $\Delta\eta = 0.031$. The resolution of the second layer amounts to $\Delta\eta = 0.025$ and for the third layer it amounts to $\Delta\eta = 0.05$. Most of the energy of electrons and photons are deposited in the second layer, while the higher resolution of the first layer serves to better distinguish energy deposits of multiple particles from those of a single particle. A schematic cut away of the calorimeter system including the electromagnetic barrel and end-caps is given in Figure 4.4.

Hadronic Calorimeter Following the depiction in Figure 4.4, the EM is encapsulated by the hadronic calorimeter consisting of the tile barrel, the tile extended barrel, the Hadronic End-Cap (HEC) and the Forward Calorimeter (FCal). These sub detectors cover a range of $|\eta| < 4.8$, where the two barrels cover an area of $|\eta| < 1$ and $0.8 < |\eta| < 1.7$, respectively. The HEC covers the range of $1.5 < |\eta| < 3.2$ and the FCal spans over the range of $3.1 < |\eta| < 4.9$. Small redundancies therefore exist between the coverage of the different sub detectors so as to mitigate any irregularities caused by fringe effects. While the HEC and FCal use liquid argon as the active medium and copper plates as the absorber, similar to the EM, the two barrels are tile calorimeters that consist of 64 modules of alternating steel and scintillator tiles. The schematic of one such module is shown in Figure 4.5. The scintillating tiles consist of transparent polystyrene with a mixture of wavelength shifting dopants. The light of each scintillating tile is fed into two independent Photo-multiplier-tubes (PMTs) via wavelength shifting optical fibers that are connected to both sides of the tile. This redundancy ensures continuity of the measurement in case one of the PMTs fails and additionally allows for a statistical

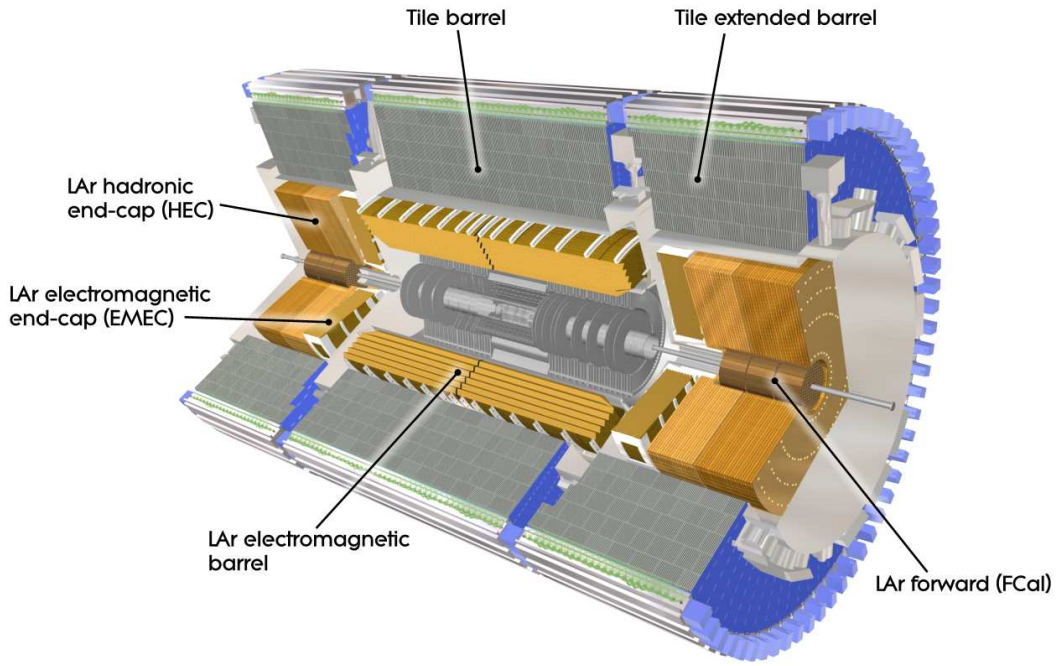


Figure 4.4: Schematic of the calorimetry system of the ATLAS detector [50].

treatment of the obtained signal as described in [60]. The tile calorimeter is calibrated using a Caesium source which traverses this sub detector through a system of tubes as shown in Figure 4.5.

Muon Spectrometer The outer-most shell of the detector represents the MS, instrumenting an area of $|\eta| < 2.7$. Muons can be considered Minimum Ionising Particles (MIPs) at the energies encountered at the LHC[2]. This means that muons are not generally stopped in the calorimeter system, and thus that their energy cannot be precisely determined there. Seeing that muons are charged particles, it is therefore beneficial to rather determine the momentum of the muon, analogously to the ID. This momentum measurement is the goal of the MS, where another important focus lies on the possibility to trigger the recording of a collision event based on the properties of muons. These two goals typically necessitate different detector technologies as for the former the highest priority is on the resolution of the measurement, while for the latter the speed of the measurement is of vital importance.
Consequently, the MS is divided into four sub detectors, where each is again divided into differently instrumented regions. For the measurement of the momentum and origin of the muons (tracking), the Monitored Drift Tube (MDT) and Cathode-Strip

[2]This will be introduced and discussed in detail in chapter 6.

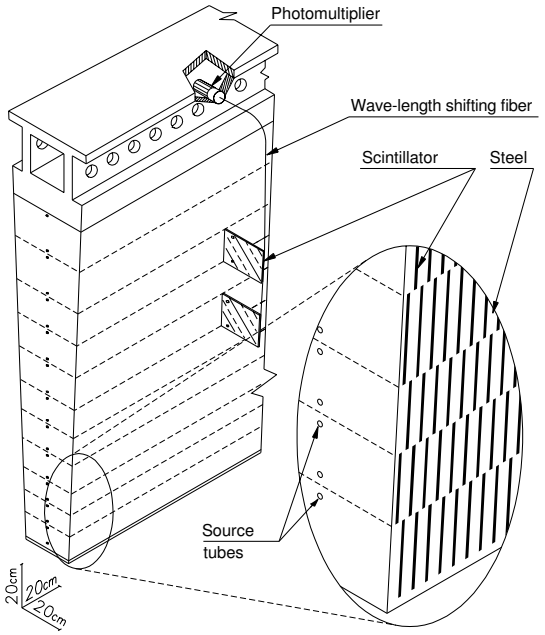


Figure 4.5: Schematic of a single module of the hadronic tile calorimeter [50].

Chamber (CSC) sub detectors are used. The MDT uses, as the name suggests, between three to eight layers of pressurised drift tubes (depending on the η region), while the CSC are multi wire proportional chambers at the innermost layer of the MS at $2 < |\eta| < 2.7$.
The subdetectors designed for triggering the recording of an event, are the Resistive Plate Chamber (RPC) and the Thin Gap Chamber (TGC) for $|\eta| < 1.05$ and $1.05 < |\eta| < 2.4$, respectively. The former are chosen because they provide a good spatial and time resolution and more importantly because they do not rely on wires, which are harder to produce and install. The latter technology is chosen for the more forward region as it allows for measurements at higher rates.

In order to measure the momentum of the muons, a strong magnetic field is needed to bend the trajectories of these particles. This is achieved with a toroidal magnet system that originally gave the ATLAS detector its name "**a** **t**oroidal **L**HC **a**pparatu**s**"[3]. The magnetic field generated by this system achieves a field strength of about 3.9 T and 4.1 T in the barrel and end-cap regions, respectively. This corresponds to a bending power of $\int Bdl = 7.5\,\mathrm{T\,m}$ [50] for an infinite-momentum muon traversing the MS. The field of the MS is perpendicular to the field of the solenoid used in the ID.

[3]This name has since been discarded and ATLAS represents the full name of the detector and not an acronym anymore.

As already indicated in the description of the MS, there is a need to specifically select collision events that should be recorded. This is due to the fact that there are in the order of 1 billion p–p collisions per second at the LHC during Run-II conditions, caused by approximately 40 million bunch crossings. This amount is considerably too large for conventional electronics and computer hardware to process or write out and, more importantly, mostly does not contain interactions of interest, so that a complex triggering system is put in place that selects collision events of interest to the different physics programmes. The complete system of triggering and data write out is called the Trigger and Data Acquisition (TDAQ) and it consists, among other things, of two levels of trigger systems and data merging and buffering facilities. The first trigger level, called L1 is a hardware trigger that uses very limited detector information to define regions of interest that are fed into the second trigger level within a timescale of less than 2.5 µs. This second level, called the High-Level Trigger (HLT), is a software based trigger that uses object reconstruction routines that are as close as possible to the routines used in the final physics analyses. The decision time for this trigger is in the order of 200 ms, which is therefore the order of buffering time needed for every collision event. The initial rate of bunch crossing at the LHC is in the order of $\mathcal{O}(40\,\text{MHz})$, where the rate of potentially interesting crossings selected by the L1 trigger system amounts to $\mathcal{O}(100\,\text{kHz})$. The final rate that is written to a data centre next to the detector and subsequently to the computing grid, given by the selection rate of the HLT, amounts to a manageable $\mathcal{O}(1\,\text{kHz})$ for Run-II conditions.

In order to assess the amount and quality of the recorded data, additional sub detectors are needed. For the measurement of the measurable luminosity, three such sub detectors are installed. The Beam Conditions Monitor (BCM) [61], located in the ID, the Luminosity Measurement Using Cerenkov Integrating Detector-2 (LUCID-2) [62] situated around the beam pipe to both sides of the main detector at an approximated distance of 17 m from the designed IP, and finally the Absolute Luminosity For ATLAS (ALFA) [63] sub detector situated at about 240 m from the IP.

The BCM is a diamond based solid state detector, that is used not only to measure bunch-by-bunch luminosities, but also the beam conditions in the form of lost beam particles. The LUCID-2 sub detector uses, as the name suggests, Cerenkov radiation as a proxy to the number of interactions per bunch crossing, where the active medium are thin quartz windows read out by PMTs. The active material of the ALFA sub detector are scintillating fibers and the detectors are mounted in movable *Roman pots* [63], that allow the detector to be placed at approximately 1 mm from the LHC beam. In this way, the detectors are used to measure the rate of proton–proton elastic scattering at

very low angles, which allows for a computation of the luminosity in several different ways, as described at length in Chapter 2 of [63].

4.2.3 The Detector for HL-LHC

Following the time line for the LHC, or rather the HL-LHC (see Figure 4.2), the accelerator complex will be upgraded to generate a significantly increased integrated luminosity of $L = 4000\,\mathrm{fb}^{-1}$. In order to achieve this within a reasonable time scale, the instantaneous luminosity represented by the average number of interactions per bunch crossing $\langle\mu\rangle$ as well as the frequency of bunch crossings need to be increased. Several scenarios for these parameter exist, where a commonly quoted scenario projects an average of $\langle\mu\rangle = 200$ for the HL-LHC phase, compared to $\langle\mu\rangle = 33.7$ [53] for the Run-II data taking period.
This poses an incredible challenge for detector and TDAQ technologies in terms of radiation, saturation and disentanglement effects. Generally, there is one interaction of interest (the hard scattering interaction) for each bunch crossing that successfully activated a trigger. With, on average, 200 interactions per bunch crossing, this leaves 199 interactions that represent a background to the interaction of interest. These interactions are called *in time* pile-up and the particles originating from these interactions overlay the signature of the sought-after particles in the different components of the detector. Furthermore, there is also *out-of-time* pile-up that originates from preceding and succeeding bunch crossings as and if the detector components integrate their measurement over more than 25 ns. This type of pile-up is particularly difficult for energy reconstruction techniques as it smears the expected shape of, for example, the pulse recorded by the PMTs in the case of the tile calorimeter.
As a consequence, also ATLAS is updated during the LS 3 starting in 2025. A plethora of studies is done in order to understand the impact of the increased pile-up contributions on the reconstruction process [64–66]. One such study, for a particular kind of muon reconstruction using ID and tile calorimeter information, is presented in chapter 6, so that possible scenarios of an upgraded tracking system for HL-LHC is briefly introduced here.

During this upgrade, the entire ID of ATLAS will be replaced by a silicon strip and a pixel detector, since these technologies allow for a very fast measurement and readout and are very robust with respect to the increased radiation dose. Additionally, the upgraded tracking system will be able to provide hardware triggers based on tracking information. This is crucial seeing that the amount of data generated by the very high instantaneous luminosities is considerably larger than for the Run-II phase, where the

rate after the first trigger level increases from $\mathcal{O}(100\,\text{kHz})$ to $\mathcal{O}(1\,\text{MHz})$. Additionally, the large menu of different physics programmes for the HL-LHC phase [67], increasingly include analyses with particles in the very forward region of the detector, so that the current range of the tracking system of up to $|\eta| = 2.7$ is not sufficient anymore. The designs for the upgrade Inner Tracker (ITK) therefore extend the coverage up to $|\eta| = 4$, further increasing the challenges of occupancy and radiation hardness of the sub detector components. One scenario of the geometry of ITK is shown in Figure 4.6, where the pixel modules are shown in red and the strip layers in blue. The horizontal axis is the distance from the IP parallel to the beam line and the vertical axis depicts the radius around the IP as the centre. One specificity of this geometry is that some of the pixel modules are inclined so that the incident angle between a particle from the IP and the module is closer to orthogonality. Additionally, this inclination creates an overlap between pixel modules that can be used to mitigate the impact of fringe effects on the measurement.

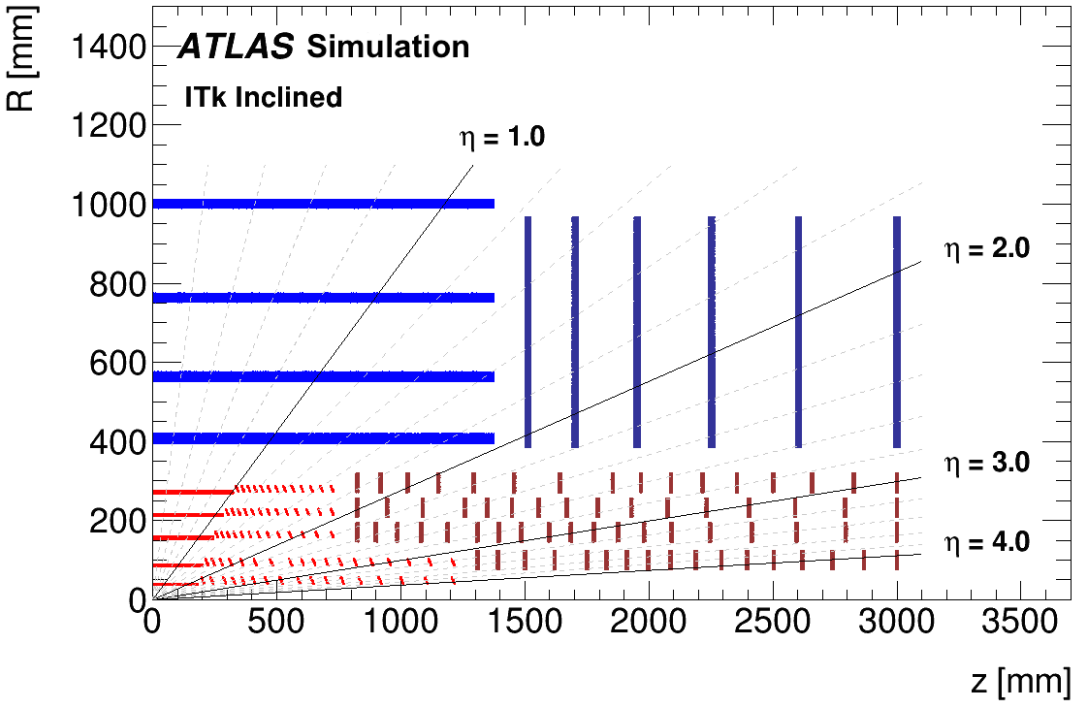


Figure 4.6: Schematic of the geometry of ITK in one quadrant for the HL-LHC phase of ATLAS. The horizontal shows the distance from the IP parallel to the beam pipe and the vertical axis represents the radius measured from the IP [64].

The measurements obtained in the ensemble of sub detectors generally yield timed space coordinates of energy deposits. However, the goal of physics analyses is to de-

termine the original interaction at the IP by back tracing and identifying the particles that interact with the detector as part of the measurement. For this reason, a very complex and comprehensive set of algorithms need to be applied to first correctly reconstruct the energy deposits and potential traces of particles and then actual particles or rather physics objects that can be associated to particles. A brief introduction to these procedures is given in the following chapter.

5 Reconstruction of Physics Objects

In most cases, the information given by the different sub detectors during data taking is either an electrical signal converted to a measure of an energy deposit or it is a hit, i.e. the activation of a sensitive detector piece by a passing particle, with its spatial coordinates. This low level information is often not directly useable for physics analyses, so that reconstruction procedures are necessary that aggregate the information from all subdetectors in order to associate different types of physical particles to these measurements.

5.1 Tracks and Vertices

Tracks are the reconstructed objects from the ID and the MS. This section heavily relies on [68], where the track reconstruction is explained in more detail.

In ATLAS, tracks (in the ID) are reconstructed following a fourfold approach. First, clusters from individual hits are formed. Second, multiple clusters are used to form track seeds. Third, track seeds result in track candidates by extrapolating the initial seed and fourth, the final track collection is selected from the list of track candidates. The different steps are described in more detail in the following.

Clusters are formed by grouping together pixels and strips, in the Pixel and SCT detectors respectively, that pass a certain energy threshold and which are adjacent to each other.
Especially in the very high activity conditions of Run-II, it is not uncommon that only one cluster is built for multiple charged particles traversing the detector, so that additional steps are needed to distinguish the clusters that only incorporate the energy deposits of a single particle from those that accumulate the deposits of multiple particles. For ATLAS, this is done using an Artifical Neural Network (ANN).

Consequently, clusters are reduced to three-dimensional coordinates called space-points. Three of these space-points can serve as a first seed for tracks, which are extrapolated towards the IP using the curvature, a momentum estimate, of the interpolation between the three space-points and assuming a uniform magnetic field. Track seeds need to have an additional (fourth) space-point that is compatible with the extrapolation from the original seed in order to be accepted. The seeds passing this criterion are used to build track candidates by using a Kalmann filter [69] to add additional viable space points. For those track candidates that pass another set of quality criteria, a final,

higher-precision fit on the space-points pertaining to the track candidate is performed (see [70]). The procedure may reconstruct multiple track candidates from a single track seed.

Given that a single cluster may correspond to multiple charged particles and that a single track seed might result in multiple track candidates, an ambiguity solving step is needed. This is achieved by appointing a *track score* to each of the track candidates, based on which, track candidates might be removed either by failing individual quality criteria or if they share a cluster with at least two track candidates that have a higher *track score*. The score itself is computed by taking into account, for example, the number of clusters associated to the candidate, the goodness-of-fit in terms of χ^2, and the p_{T}, where a higher p_{T} leads to a higher score.

In order to determine if a set of tracks originate from the same source, vertices are reconstructed that represent an area of intersection of multiple reconstructed tracks. The tracks used in this reconstruction need to satisfy a menu of quality criteria, e.g. a $p_{\mathrm{T}} > 400\,\mathrm{MeV}$ and $|\eta| < 2.5$, which can be found in [71]. With this collection of tracks, vertices can be found (built) and refitted as described in [72, 73]. The former is achieved by first determining vertex seeds around maxima in the z impact parameter distribution of the remaining tracks, followed by an iterative procedure that assigns a weight to every track and adds tracks to the initial seed. After each addition, the position of the vertex is recomputed by a χ^2 minimisation taking into account the weights of the tracks. The vertex build-up is stopped once the minimal χ^2 reaches a predefined cut-off value. The full procedure is repeated for all tracks that have not been associated to a vertex until either all tracks are associated to a vertex or no additional vertices can be found. In addition to this technique, also image reconstruction techniques are used [74].

In the complex realm of hadron collisions in bunches, multiple types of vertices can be defined. The most important one is called Primary Vertex (PV) and is supposed to represent the vertex of a hard-scattering interaction. It is determined to be the vertex with the largest sum of squared track momenta $\max_{\mathrm{vertices}}\left(\sum p_{\mathrm{T}}^2\right)$. Another important category is the one of secondary or displaced vertices which appears for example in the case of photon conversion (conversion vertex) or decays of heavy particles like b hadrons. Finally, there are the vertices that correspond to minimum-bias collisions in the same bunch crossing, called pile-up.

5.2 Electrons and Photons

Electrons and photons are the main particles to interact with the EM, where their interaction results in a cascade of Bremsstrahlung and pair production effects called an electromagnetic shower. The fact that the parameters of such a shower are very similar whether the shower has been initiated by a photon or by an electron makes it reasonable to treat these two particles in a coherent way instead of independently. This is also the case for the reconstruction procedure done in ATLAS.

In the following, the main ideas of electron and photon reconstruction and energy calibration are pointed out and the reader is referred to [75] for an overview and to [76] for an in-depth description of the procedure for electrons.

5.2.1 Reconstruction

There are three type of objects that are reconstructed in this category: Electrons that have energy deposits in the EM and an associated track in the ID, photons that only have energy deposits in the EM and converted photons, which have energy deposits in the EM and are associated to a conversion vertex, which is a point of origin of one or more tracks outside of the Primary Vertex (PV) compatible with a massless particle decay.

All three objects are reconstructed using topologically connected clusters of calorimeter cells (topo-clusters) which, in turn, are used to build superclusters. The object candidates reconstructed from these superclusters have to undergo an ambiguity solving step between electrons and photons to determine a final set of objects that, after energy calibration, is used in physics analysis.

The EM cell-level figure-of-merit used to build topo-clusters $\varsigma^{\mathrm{EM}}_{\mathrm{cell}}$ is defined as

$$\varsigma^{\mathrm{EM}}_{\mathrm{cell}} = \frac{E^{\mathrm{EM}}_{\mathrm{cell}}}{\sigma^{\mathrm{EM}}_{\mathrm{noise,cell}}}, \tag{5.1}$$

where $E^{\mathrm{EM}}_{\mathrm{cell}}$ is the energy of the cell and $\sigma^{\mathrm{EM}}_{\mathrm{noise,cell}}$ is the expected noise (electronic and activity from additional pp collisions called pile-up) of the same cell.
Topo-clusters are formed around initial seed cells that satisfy $\left|\varsigma^{\mathrm{EM}}_{\mathrm{cell}}\right| \geq 4$ by adding all cells adjacent to the seed cell as well as all cells connected to these cells that fulfil $\left|\varsigma^{\mathrm{EM}}_{\mathrm{cell}}\right| \geq 2$. Clusters are merged if they connect and may be split if two local maxima

above a given threshold are found. The clusters are surrounded by a ring of cells not satisfying the figure-of-merit requirement. Details of the formation and treatment of topo-clusters are best described in [77].

For electrons, tracks are reconstructed following the above procedure and loosely matched to clusters in the EM. This is followed by a re-fit of the tracks that have a match, using a generalisation of the Kalman Filter used in the initial fit of the track. The specificities of this generalisation can be found in [78]. The additional information from the re-fit allows for a more precise matching, not only spatially but also in terms of the matching between the momentum of the track and the energy of the cluster. Seeing that multiple tracks might be associated to the same cluster, a procedure is put in place that ranks the tracks depending on the quality of the track and its proximity to the cluster in terms of ΔR. Only the highest-ranked track is used.
For converted photons, the same loose matching between tracks and clusters in the EM as above is performed. If there are at least two opposite charge tracks matched to a cluster, a conversion vertex is built if the invariant mass of the two tracks are compatible with a massless particle. Following that, especially in the case of conversions, the environment for tracks is very dense, it might be the case that not both tracks corresponding to the electron-positron pair are reconstructed. As a consequence, also single-track conversion vertices are considered if the track does not have hits in the innermost layers of the ID, which could indicate a displaced vertex initiated by a chargeless particle.
Unconverted photons are reconstructed from clusters in the EM that cannot be matched to any track that passes certain quality criteria.

After obtaining a first collection of candidates using topo-clusters and tracks, superclusters are formed for each of the candidates in order to achieve a significantly improved energy reconstruction. Superclusters consist of a seed cluster and satellite clusters that are found within a fixed window around the seed. These superclusters are then used to build the final reconstructed objects by applying a first energy calibration and position adjustment and by redoing the track matching procedure. The step of building superclusters is done independently for electrons and photons, so that a single supercluster might be used to reconstruct both objects, making another ambiguity solving step necessary.
The ambiguity solver removes the photon entirely if a good track, but no good conversion vertex is matched to the supercluster (for the definition of *good*, see [75]). Otherwise, it removes the electron if no good track is matched and finally, both are

kept if none of these conditions match, setting a flag to mark them as still ambiguous.

Superclusters might not contain all energy deposits originating from one of the primary particles (electron or photon from the hard scattering process in this case). Additionally, there are other subdetectors that come before the main layers of the EM (from the beam pipe going outwards), like the presampler and the entire ID, where energy losses have likely already occurred. This, among other factors, means that the reconstructed energy does not correspond to the energy of the primary particle, making the case for an energy calibration. The effects mentioned here are treated with a multivariate algorithm trained exclusively on simulation. The full description of the energy calibration can be found in [79].

5.2.2 Identification

One major challenge for all object reconstruction techniques is to balance object reconstruction efficiency against object purity (or rather the purity of the collection of reconstructed objects). The former is the fraction of reconstructed objects from all real (prompt) objects from the collision $\epsilon = \frac{N_{\text{reco}}}{N_{\text{true,prompt}}}$, while the latter represents the fraction of correctly reconstructed prompt objects in the collection of all reconstructed objects (including erroneously reconstructed objects), $\pi = \frac{N_{\text{reco|true,prompt}}}{N_{\text{reco}}}$. Prompt means originating from the hard scattering process or heavy resonance decays in this context, and correctly reconstructed means that the correct object type is reconstructed, e.g. a real photon is reconstructed as a photon and not as an electron.

In the case of ATLAS, the baseline reconstruction yields a certain purity of the collection, which is improved upon by applying additional criteria in terms of identification and isolation for electrons, photons and muons, which are described later on. Different Working Points (WPs) for these identification and isolation criteria are given for these three object types in order to strike the optimal balance between efficiency and purity in a wide range of different physics analyses.

Photon identification is done in a *cut-based* approach using a series of variables designed to describe the shape of the electromagnetic shower. The full list of variables used in this approach are listed in Table 1 in [80].
Electron identification is done using a likelihood based approach, taking into account the Probabiliy Density Functions (PDFs) of the signal (prompt electrons) and multiple backgrounds such as jets (see next section) that are erroneously reconstructed

as electrons, electrons from photon conversions and electrons from decays of hadrons that contain a heavy flavour quark. The final discriminant for identification is the transformed ratio of the signal likelihood and the background likelihood. Some of the variables used for photon identification are also used for electron identification, where the latter is extended to also incorporate information from the track. The full list of variables used can be found in Table 1 in [76].
For electrons, there are four identification WPs: *VeryLoose, Loose, Medium,* and *Tight*, the efficiency of which is determined depending on E_{T} and η. For example, the latter three WPs correspond to an identification efficiency of 93 %, 88 % and 80 % [76] for an electron with a transverse component of the energy of the cluster $E_{\mathrm{T}} = 40\,\mathrm{GeV}$, respectively.

5.2.3 Isolation

As described above, an additional approach to purify the collection of reconstructed objects is by using isolation criteria. The idea behind isolation is that prompt electrons and photons tend to have a fairly collimated signature with the detector in comparison with the background interactions mentioned above. This means that for the signal process, there should be next to no activity around the reconstructed object, while this is not the case for erroneously reconstructed objects coming from high activity processes.

Both calorimeter and track based isolation is used in ATLAS. For electrons, calorimeter isolation is determined by a discriminant $E_{\mathrm{T}}^{\mathrm{isol}}$, which is the sum of E_{T} of clusters in a fixed size cone in the $\eta\phi$ plane of radius R around the reconstructed object (including the cluster used to reconstruct the object in question) subtracted by the E_{T} of the EM cells in a window of $\Delta\eta \times \Delta\phi = 0.125 \times 0.175$ around the object's epicentre. Since it can happen that a prompt electron has deposits also outside of this core window, an additional fit is used to determine the contamination of $E_{\mathrm{T}}^{\mathrm{isol}}$ by deposits from the prompt electron. Pile-up can additionally enhance $E_{\mathrm{T}}^{\mathrm{isol}}$, which is not the desired effect since this variable is used to separate prompt electrons from other hard background sources and not from pile-up. Consequently, also an estimate of the pile-up activity is subtracted.
The discriminant used in the track based isolation is called $p_{\mathrm{T,var}}^{\mathrm{isol}}$. It is computed by summing the p_{T} of tracks within a variably-sized cone around the reconstructed trajectory of the electron, subtracted by the track p_{T} of the electron. The size of the cone depends on the p_{T} of the electron and is narrower for higher p_{T} and which has a maximum radius R_{max}, i.e. $R = \min\left(\frac{10\,\mathrm{GeV}}{p_{\mathrm{T}}\,[\mathrm{GeV}]}, R_{\mathrm{max}}\right)$. For more details see [76].

Photon isolation follows the same general ideas, with the most notable difference that no track should be associated to the photon. For the full procedure, the reader is referred to Section 3.3. of [80].
In the following, the discriminant variables will be called $E_{\mathrm{T}}^{\mathrm{topoconeX}}$ and $p_{\mathrm{T}}^{\mathrm{varconeX}}$ for calorimeter based and track based isolation, respectively. The X in the superscript will be replaced to denote the size of the cone, where $X = 100R_{(\mathrm{max})}$.
As for the identification, different WPs are also defined for isolation, where the WPs for electrons are listed in Table 5.1 .

Table 5.1: Isolation Working points for electrons. The *Gradient* WP uses multiple cut values on the calorimeter and track isolation variables to obtain a specific efficiency ϵ, while the other WPs make use of a single, fixed, cut. Slightly adjusted and taken from [76].

Working Point	Calorimeter isolation	Track isolation
Gradient	$\epsilon = 0.1143 \times p_{\mathrm{T}}\,[\mathrm{GeV}] + 92.1\,\%$	*idem*
HighPtCaloOnly	$E_{\mathrm{T}}^{\mathrm{topocone20}} < \max\left(0.015 \times p_{\mathrm{T}}\,[\mathrm{GeV}], 3.5\,\mathrm{GeV}\right)$	—
FixedCutLoose	$E_{\mathrm{T}}^{\mathrm{topocone20}}/p_{\mathrm{T}} < 0.20$	$p_{\mathrm{T}}^{\mathrm{varcone20}}/p_{\mathrm{T}} < 0.15$
FixedCutTight	$E_{\mathrm{T}}^{\mathrm{topocone20}}/p_{\mathrm{T}} < 0.06$	$p_{\mathrm{T}}^{\mathrm{varcone20}}/p_{\mathrm{T}} < 0.06$

5.3 Muons

Muons can be considered MIPs for the energy range at the LHC and they provide a very distinguishable signature while passing through the entire detector. For this reason, information from almost all subdetectors is used to reconstruct muons.

5.3.1 Reconstruction

In a first step, muon candidates are reconstructed from each of the involved subdetectors individually [81]:

- Combined (CB) muons: Independent track reconstruction in the ID and the MS followed by a combined refit in an outside-in approach (default), i.e. the MS track is extrapolated towards the ID, or in an inside-out approach (complementary).
- Segment tagged (ST) muons: Track reconstructed in the ID which matches to a track segment in the MS. This collection is used if the original muon likely only crossed one segment of the MS

- Calorimeter tagged (CT) muons: Track reconstructed in the ID which matches to calorimeter deposits compatible with a MIP. Exclusively used in the MS gap region, i.e. $|\eta| < 0.1$. The specificities of this reconstruction type are discussed in detail in the next chapter.

- Extapolated (ME) muons: Candidate reconstructed only in the MS, given that its track can be extrapolated to the IP. This collection is designed to extend the muon acceptance beyond the pseudorapidity reach of the ID, i.e. $|\eta| > 2.5$.

One of the reasons why tracks are so important in muon reconstruction is that muons generally do not deposit their entire energy in the calorimeters, so that a momentum measurement has to come from the curvature of the track in the magnetic field, which could then be used to deduce the energy of the muon.
Since the above collections of muon candidates are reconstructed individually, they might contain overlaps where the same tracks are used to reconstruct multiple muons. Therefore, an Overlap Removal (OR) is applied that removes a muon candidate if it uses the same tracks as a muon candidate from a higher priority collection. The order of priority sets first CB muons, second ST muons and third CT muons, which reflects the hierarchy in terms of quality of the muon candidates. Since ME muons cannot share an ID track with other collections (the ID is not used for this muon type), its overlap is determined by looking at the track hit content in the MS and if the same has been used by another muon candidate, the one with a higher fit quality is kept.

5.3.2 Identification

As for electrons and photons, see section 5.2, identification and isolation WPs are defined for muons to give analysts a way of choosing the optimal level of purity and efficiency of the final muon collection for their specific task.
One main source for background muons are the decays of charged hadrons, which, as noted in subsection 5.4.2, may lead to a displaced vertex or at least a change of the direction of the track. This change consequently leads to a discrepancy between the reconstructed tracks of the muon in the ID and the MS, which in turn results in a low fit quality providing a powerful variable for identification.
Five WPs are defined: *Loose*, *Medium*, *Tight*, *LowPt* and *HightPt*, which correspond to the lowest, intermediate and highest purity as well as to a collection optimised for low p_{T} and high p_{T} muons, respectively. The results of efficiency measurements for these WPs are, with the exception of *LowPt*, found in [81].

5.3.3 Isolation

Background muons coming from hadron decays are generally not well isolated as the original hadron is part of a jet (see next section). Thus, isolation parameters can play a major role in suppressing this kind of background. The performance of the isolation WPs *LooseTrackOnly*, *Loose*, *Tight*, *GradientLoose*, *Gradient*, *FixedCutLoose* and *FixedCutTightTrackOnly* are also shown in [81].

Because there are some residual differences between, for example, the description of the magnetic field or the energy loss of the muon in the calorimeters between data and simulation, calibration techniques need to be applied to correct the momentum scale of the muon. This is commonly done using well known processes and well measured particles that decay into muons, such as $J/\psi \to \mu^+\mu^-$ and $Z \to \mu^+\mu^-$. For the full procedure, the reader is referred to Section 8 of [81].

5.4 Jets

Jets do not represent a singular particle as the previously introduced objects, but rather an entire hadronic shower, i.e. a large shower originating from a singular quark or gluon from the hard scattering process. The reason for this is that these two particle types are colour-charged and, thus, cannot be observed in an unbound state in nature. Instead, quarks form colourless bound states (hadrons) at the very short time scale of QCD which is approximately $\tau_{\mathrm{QCD}} \approx 1/\Lambda_{\mathrm{QCD}} = 3 \cdot 10^{-24}\,\mathrm{s}$ [14]. Because the initial quarks and gluons are very energetic during collisions at the LHC, scores of colour-neutral hadrons are produced that interact with the detector. This hadronic showering makes it very hard to reconstruct individual particles, so that a compound object is reconstructed instead. Fortunately, the produced particles are scattered conically around the original particle's trajectory, so that their ensemble can be reconstructed as a conically shaped jet that carries the kinematic information of the source particle.

5.4.1 Reconstruction

For this analysis, jets are reconstructed using topo-clusters from the electromagnetic and hadronic calorimeters as introduced in the previous section. Another option recently introduced to ATLAS is to reconstruct jets on the basis of particle flow objects, for which the reader is referred to [82].

The algorithm of choice to reconstruct jets in ATLAS is the anti-k_t algorithm introduced in [83]. It relies, such as its cousins the k_t and Cambridge/Aachen algorithms, on a distance parameter d_{ij} between the objects i and j to associate topo-clusters within a specific cone to a jet. A second parameter d_{iB} refers to the distance between object i and the beam axis and both distances are defined as

$$d_{ij} = \min\left(k_{ti},^{-2},k_{tj},^{-2}\right)\frac{\Delta_{ij}^2}{R^2} \tag{5.2}$$

$$d_{iB} = k_{ti},^{-2}, \tag{5.3}$$

where $k_t i$ refers to the transverse momentum of object i, which would be denoted $p_{\mathrm{T},i}$ in this document but is kept to showcase the origin of the algorithm's name. The distance Δ_{ij} is similar to the distance mostly used in this thesis, ΔR, except that it uses the rapidity y of the objects instead of the pseudo-rapidity η, so that $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$. The tunable parameter of the algorithm is R which determines the size of the cone and is fixed before the execution of the algorithm, where a value of $R = 0.4$ is commonly used in ATLAS analyses.
The anti-k_t clustering merges objects i and j if the condition $d_{ij} < d_{iB}$ is met, leading to perfectly cone shaped jets if all hard (high-p_{T}) objects are at a minimal distance of $2R$ from each other. This condition is not necessarily fulfilled in the busy environments at the LHC and especially not in the boosted regime (very high-p_{T} objects that are very collimated), leading to a difference in shape or even a merger of multiple hard objects into a single jet. The implementation of the anti-k_t algorithm used in ATLAS is done in the FastJet [84] package.

5.4.2 Flavour Tagging

For many analyses carried out within the ATLAS collaboration, it is crucial to determine more precisely the particle of origin for jets. One important distinction is between quarks and gluons as a whole (see [85, 86]), while another one is between the different flavours of the quarks. The algorithms used to determine the flavour of the source object of a jet are called jet-taggers (or taggers for short) and a jet classed as, for example, originating from a bottom quark, is called a b-tagged jet.
Recently, multi-class taggers are on the rise that are able to determine multiple flavours using the same algorithm such as DeepJet and DeepCSV used by the CMS collaboration [87, 88] or DL1 [89] which is used by ATLAS. There are top quark taggers making use of W boson taggers, since the decay time of a top quark is shorter than the hadronisation time τ_{QCD} mentioned above, so that the intermediary decay of the top quark into a W boson plays a role.

The most widely used type of taggers, however, are bottom quark taggers, because bottom quarks leave a very specific signature in the detector in comparison to light quarks (u,d,s,c) and because bottom quarks play a defining role in separating many processes that are currently under investigation in ATLAS from their background processes.

Many b tagging algorithms rely on the fact that the b hadrons formed by the initial bottom quark have a life time in the order of 1 ps to 2 ps, so that the mean flight path length $\langle l \rangle$ of a b hadron with $p_{\mathrm{T}} = 50\,\mathrm{GeV}$ is $\langle l \rangle \approx 3\,\mathrm{mm}$ [90]. This leads to a decay vertex of the b hadron that is spatially separated from the PV.
Initially, individual algorithms were used that either exploited the impact parameter d_0 (i.e. the shortest perpendicular distance between the extrapolated track and the PV), the distance of a secondary vertex to the PV or information from muons that emerge from the decay of the b hadron. Currently, variables of all three approaches are combined and passed to a Boosted Decision Tree (BDT) as implemented in the Toolkit for Multivariate Analysis (TMVA) [91], resulting in a group of classifiers called MV2 [92, 89], or they are passed to the multi-class deep feed-forward ANN called DL1.

The MV2 classifiers are trained on data sets combining events from $t\bar{t}$ and Z' processes where the $|\eta|$ and p_{T} distributions of b and c jets are reweighted to match the distributions of light jets. This is important so that the classifier is not able to use this information directly (and correlated variables thereof) to distinguish between the different origins of the jets, which would be dependant of the process used in the training.
Background compositions of the training data sets are chosen to be 0 % c jets (100 % light jets) for MV2c00, 7 % c jets (93 % light jets) for MV2c10 and 15 % c jets (85 % light jets) for MV2c20.

The main figures of merit for most classifiers are the signal efficiency ε and the background rejection $\varepsilon_{\mathrm{bkg}}^{-1}$. This is also the case for the b tagging algorithms presented here, with the b tagging efficiency ε_b, the c tagging efficiency ε_c and the light jet rejection ε_ℓ^{-1}. The discriminant of MV2, and of DL1 for that matter, is continuous so that several WPs (sometimes called operating points) are set for better comparability between analyses and for calibration purposes.
These WPs represent a specific b tagging efficiency and corresponding background rejection (choosing a WP is a trade-off between the two), giving analysts the possibility to optimise the b-tagged jets collection for their study. Currently, the WPs of $\varepsilon_b = 60\,\%$, 70 %, 77 % and 85 % are provided in ATLAS and the corresponding cuts on the BDT

discriminant as well as the rejection of c, τ and light jets in a simulated $t\bar{t}$ sample using the MV2 tagger trained with a 7 % c jet background fraction are listed in Table 5.2.

One elegant method to determine the b tagging efficiency is called $p_{\mathrm{T}}^{\mathrm{rel}}$ [93], which is applicable in the case where a muon can be matched to a jet using a cone algorithm (ΔR matching). The main discriminating variable, and the namesake to the method, is the component of the momentum of the muon that is transversal to the the combined momentum of the muon and the jet, called $p_{\mathrm{T}}^{\mathrm{rel}}$. This value is correlated to the mass of the mother particle and thus to the flavour composition of the original hadron if the muon originates from a hadron decay within the jet. Another important point about this method is that it is a perfect candidate for an efficiency measurement since $p_{\mathrm{T}}^{\mathrm{rel}}$ is only weakly correlated to the information used by MV2. Additionally, this method can be equally deployed in data and simulation and can thus be used to calibrate the b tagging procedure by deriving Scale Factors (SFs) between data and Monte Carlo (MC) or between different MC setups.

Table 5.2: Selection criterion, c, τ and light jet rejection in a simulated $t\bar{t}$ sample for different b tagging efficiency ε_b Working Points for the MV2 algorithm trained with a 7 % c jet fraction. Slightly adapted from [94].

	Selection on BDT discriminant	Rejection ($1/\varepsilon$)		
ε_b		c jet	τ jet	light jet
60 %	> 0.94	23	140	1200
70 %	> 0.83	8.9	36	300
77 %	> 0.64	4.9	15	110
85 %	> 0.11	2.7	6.1	25

5.5 Neutrinos

Neutrinos are among the most challenging particles to be measured at collider-experiment detectors due to the very low interaction cross-sections with any detector material. As a consequence, neutrinos are not reconstructed or even measured directly, but rather estimated from other measurements.

Given that the colliding particle beams at the IP are perfectly aligned and that the collision happens fully *head on*, the transverse component of the momentum of the

colliding particles in the lab frame is zero before collision. Following momentum conservation, this has to also be the case after collision, so that the sum of all transverse momenta of all particles produced in the collision should equally be zero. This way, the sum of momenta in the x-y plane of all particles that are invisible to the detector (expressed as energy deposits in the calorimeters) $E_{x(y)}^{\text{miss}}$ is determined by

$$E_{x(y)}^{\text{miss}} = E_{x(y)}^{\text{miss},\mu} + E_{x(y)}^{\text{miss},e} + E_{x(y)}^{\text{miss},\gamma} + E_{x(y)}^{\text{miss},\tau} + E_{x(y)}^{\text{miss,jets}} + E_{x(y)}^{\text{miss,soft}}\,, \tag{5.4}$$

where the terms for e,γ,τ and jets represent the negative vectorial sum of calibrated cell energies in the calorimeters, the muon term represents the the negative vectorial sum of calibrated muon momenta (from tracks) and $E_{x(y)}^{\text{miss,soft}}$ sums up all *soft* energy deposits and track momenta not associated to any of the particles that pass certain quality criteria.
The most important quantities used in physics analyses are the magnitude of the vector $\vec{E}_{\text{T}}^{\text{miss}}$, denoted $E_{\text{T}}^{\text{miss}}$ and its azimuthal angle ϕ^{miss} defined as

$$E_{\text{T}}^{\text{miss}} = \sqrt{\left(E_x^{\text{miss}}\right)^2 + \left(E_y^{\text{miss}}\right)^2} \quad \text{and} \tag{5.5}$$

$$\phi^{\text{miss}} = \arctan\left(\frac{E_y^{\text{miss}}}{E_x^{\text{miss}}}\right)\,, \tag{5.6}$$

as well as the total transverse energy in the detector $\sum E_{\text{T}}$ representing the scalar sum of the transverse momenta of the objects used in Equation 5.4 and defined as

$$\sum E_{\text{T}} = \sum p_{\text{T}}^{\mu} + \sum p_{\text{T}}^{e} + \sum p_{\text{T}}^{\gamma} + \sum p_{\text{T}}^{\tau} + \sum p_{\text{T}}^{\text{jets}} + \sum p_{\text{T}}^{\text{soft}}. \tag{5.7}$$

Seeing that each of the components used in Equation 5.4 is computed on individually reconstructed object collections, individual deposits or tracks might contribute multiple times. Therefore, an OR procedure is applied that uses a predefined hierarchy to remove objects that are already taken into account by objects that are higher up in the hierarchy. This hierarchy can be fine tuned by individual analysts but the most commonly used approach is to first consider muons then electrons, photons, hadronically decaying τ leptons and jets. The *soft* term only contains energy deposits and tracks not associated to any of the reconstructed objects by construction.

The computation of $E_{\text{T}}^{\text{miss}}$ uses its own set of reconstruction criteria for all of the objects mentioned above, which can be different to the criteria used for these objects used alongside $E_{\text{T}}^{\text{miss}}$ in an analysis. The full list of criteria and the procedure in detail is presented in [95].

The major caveats of the approach used here are first that only the transverse component of the neutrino can be determined, and in the case for multiple neutrinos only the vectorial sum of their p_{T}, and second that it is not possible to disentangle neutrinos from other potential particles that are invisible to the detector such as dark matter candidates.

5.6 Calibration of Objects

At this stage, the energy of the jet candidates is only estimated at the EM scale, which is incorrect for the hadronic content of the jet. Thus, a set of corrections and calibrations need to be applied, two of which are the absolute Jet Energy Scale (JES) calibration and the *in situ* calibration. The former uses the particle (truth) level information of simulated jets and matches them to reconstructed simulated jets for calibration, while the latter corrects for possible mis-modellings in the simulation using well known processes like the recoil of the Zboson in a $Z + 1$jet process. The full procedure is summed up in [96].
After calibration, the jet candidates are ready to be used in physics analyses.

5.7 Overlap Removal

As already mentioned for the computation of $E_{\mathrm{T}}^{\mathrm{miss}}$, an OR procedure needs to be in place in case multiple reconstructed objects are using the same track or energy deposit in the calorimeter. The most commonly used discriminant variable to check if two objects overlap is ΔR and a list of detailed discussions on different OR strategies may be found in [97]. One important recommendation following this document is that the OR algorithm should be applied on final reconstructed objects used in the analysis, specifically with the same ID and isolation criteria. One of the algorithms used in ATLAS is called BoostedSlidingDRMu, which solves the overlap by executing the following steps in the order they are presented:

- remove the electron track candidate if it overlaps with another electron
- remove CT muon if it uses the same track as an electron
- remove electrons that use the same track as a muon
- remove jets within a cone around an electron, where $\Delta R_{(e,j)} < 0.2$. If there are multiple jets inside this cone, only remove the closest one.

- remove jets within a cone of $\Delta R_{(\mu,j)} < 0.2$ around a muon if the jet contains fewer than three tracks.
- remove jets that contain fewer than three tracks if a muon track in the ID is matched to it.
- remove muon if the distance between any jet and the muon is $\Delta R < 0.4$.
- remove muon if the distance between any jet and the muon is $\Delta R < 0.4 + \frac{10\,\mathrm{GeV}}{p_{\mathrm{T}}^{\mu}}$.

The algorithm owes its name to the last condition, which varies the size of the cone around the muon where no jet should be found according to the muons p_{T} (sliding window).

The impact of different OR techniques on different analyses can be quite substantial and depend in most part on what levels of purities and efficiencies for each object are optimal for the specific analysis in question, so that extensive studies on different OR techniques are the starting point of many analyses.

6 Muon Reconstruction under HL-LHC Conditions

The previous chapter introduced the importance of a high quality reconstruction and identification procedure to balance the purity of objects against a loss in reconstruction efficiency. In order to ensure a good balance, or performance, in light of the proposed upgrade of the collider and detector for the HL-LHC phase, it is paramount to study the impact of different upgrade scenarios on this performance and consequently choose the optimal case.

For this reason, studies on the reconstruction performance of CT muons under HL-LHC conditions are presented and compared to the results of current conditions. These studies focus mainly on two aspects. First, evaluate the impact of increased pile-up and second, emulate the impact of ageing or damaged detector components on the reconstruction performance.

6.1 Overview

The study presented here has been conducted exclusively by the author of this dissertation as part of a *qualification task*, i.e. a task or service that has to be completed in order to become a verified author of the ATLAS collaboration, and it has a clear motivation in terms of the search for four top quarks presented below. The results of this analysis have been documented and internally reviewed in [98]. This chapter is in large part an adaptation from this document.

Muons and their reconstruction in the calorimeter are the focus of this study because they leave a very narrow and uniform energy deposit throughout the calorimeter (see subsection 6.1.1) and because they are not as strongly affected by clustering algorithms as jets or electrons, which allows for a better focus on effects purely at the calorimeter cell level.

The approach of this study is to investigate the effect of different types of emulated noise and efficiency loss effects on the CT muon reconstruction performance under current conditions and in light of the phase II upgrade for HL-LHC. The motivation for the noise scenarios are the harsh conditions expected for the HL-LHC phase, where it is crucial to identify which components of the detector need to be replaced due to ageing and irradiation effects or simply because they are not able to cope with the increased occupancy. Particularly the last point is not only a challenge for the

hardware of the detector, but also for the software, such as the choice of algorithm for energy reconstruction.
These studies should, in principle, be performed for all potential scenarios so as to choose the optimal one. Unfortunately, this is very time consuming and not feasible due to the complexity of the optimisation problem, so that the studies performed here only assume the most likely upgrade scenario. The results of this study can then be used to either flag significant problems of the scenario if the obtained performance is significantly reduced or to corroborate trust in the proposed model if the performance is the same or even improved.

As mentioned above, investigating the resilience of CT muons in an upgraded detector setup is also motivated by the current and future search for four top quarks, where muons are one of the possible final state particles of the four top quark decay.
This process has possible decay modes containing up to four muons, so that a high muon reconstruction efficiency can directly impact (up to the 4th order) the efficiency of the four top quark signal efficiency. This is additionally interesting seeing that the background contribution for the four muon final state is very low, strongly suggesting that already a migration from a four muon signal region to a three muon signal region, caused by not reconstructing one of the muons, would result in a lower overall significance.
An illustration of this argument can be found in Figure 6.1, which shows the migration of events being classified as either None (no signal region), same sign $\mu^{\mp}\mu^{\mp}$, 3μ or $\geq 4\mu$, if the nominal muon selection is applied (x-axis) to the same signal categories when muons that are reconstructed using calorimeter information are also used (y-axis). The colour code and therefore also the bin entries are normalised to unity. The figure shows that the overall number of events entering any signal region increases by approximately 3 % (the sum of the first column), and that there is indeed a migration towards signal regions with less background contributions (regions with a higher muon multiplicity). The final impact of these improvements on the signal significance is of course non-trivial and especially the erroneous reconstruction of other objects as muons (fakes) will play a role. But it is clear that studying muon reconstruction in the calorimeter remains an interesting subject for many physics analyses including the search for four top quarks.

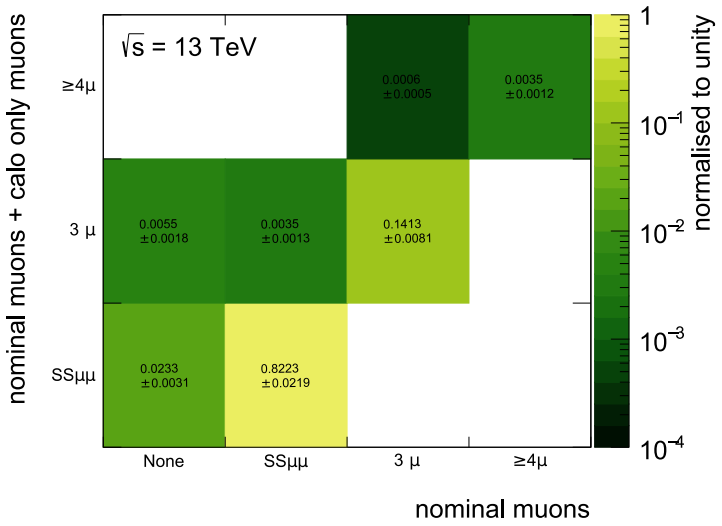


Figure 6.1: Event categorisation of muon events for the search for four top quark production, only using nominal muons (x-axis) or also including muons that are exclusively reconstructed in the calorimeters (y-axis). The colour-code and values are normalised to unity.

6.1.1 Muon Interaction with Matter

The starting point for any reconstruction of muons is a clear understanding of the measurement process, i.e. the interaction of the particle with the detector material, which is therefore highlighted in the following. This discussion relies heavily on the descriptions found in [99].

While passing through matter, muons experience an energy loss mainly originating from nuclear losses, ionisation losses and radiative losses, whose strength depend on the muon's momentum (see Figure 6.2). The mean rate of energy loss $\langle -\frac{dE}{dx} \rangle$ is called the stopping power and for muon momenta common at the LHC, i.e. $0.1 \lesssim \beta\gamma \lesssim 1000$. It is described by the Bethe equation with an accuracy of a few percent, which is described in detail in [99].

For lower muon momenta, it is necessary to add additional corrections that account for atomic binding, but which are neglected at this point as those momenta are usually well below the minimum requirements of analyses conducted at the LHC.

Figure 6.2 shows the mass stopping power of an anti-muon in copper in $\mathrm{MeVcm^2g^{-1}}$ in dependence of the muon's momentum in $\mathrm{MeVc^{-1}}$. It is divided into four regions: The Lindhard-Scharff, the Anderson-Ziegler, the Bethe and the radiative region. The first two are, as explained above, omitted. The Bethe region is given in [99] and the radiative region will be described more thoroughly in the next paragraph. The solid line represents the total mass stopping power, whereas the dashed and the dotted lines

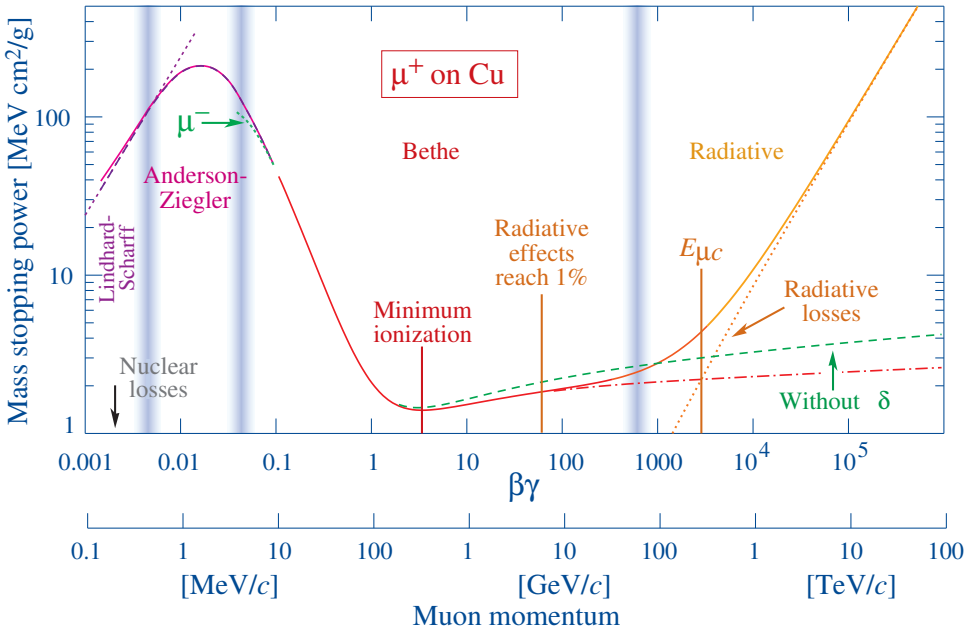


Figure 6.2: Mass stopping power $\langle -\frac{dE}{dx} \rangle$ per unit length in $\mathrm{g\,cm^{-2}}$ for anti-muons in copper in dependence on the momentum of the muon [99].

show the partial contributions. Specifically, the green dashed line represents the results of the Bethe equation without the density correction explained in section 33.2.5 in [99], while the red dashed-dotted line includes this correction. Furthermore, two interesting values for the muon momenta are detailed: The point of minimum ionisation and the muon critical energy $E_{\mu c}$. The former represents the minimum of the Bethe equation for common energies at the LHC, while the latter describes the point in phase-space, where the energy loss incurred by ionisation and radiation have the same value. For muon momenta below the critical energy, ionisation losses are the most important cause for the muons' energy loss, while it is radiative losses for higher momenta.

The radiative losses can be further separated into photonuclear interactions, Bremsstrahlung and e^+e^- pair production, which again strongly depend on the muon's momentum. Figure 6.3 shows the average energy loss of different radiative processes (dashed lines), ionisation losses (solid brown line) and the total average energy loss (red line) in iron depending on the momentum of the muon. Additionally, the total average energy loss of a muon is shown for hydrogen (solid yellow line) and for uranium (dotted blue line), to indicate the different behaviour for different values of the atomic number Z and for different states. The muon critical energy would correspond to

the crossing of the solid violet line (Fe total radiative) and the solid brown line (Fe ionisation) in this figure.

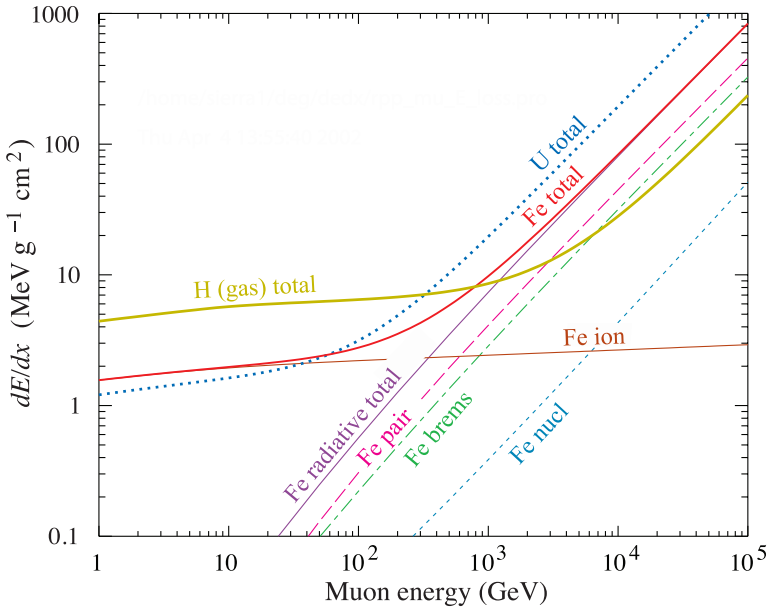


Figure 6.3: Energy loss of a muon in hydrogen, uranium and iron. The different radiative contributions are also shown independently in the case of iron [99].

Following the shape of the function for higher momenta, an extension of Bethe's equation can be parameterised as

$$\left\langle -\frac{dE}{dx} \right\rangle = a(E) + b(E)E \tag{6.1}$$

where $a(E)$ represents the ionisation losses covered by the equation and $b(E)$ represents the total radiative losses. With this parameterisation, the muon critical energy is given by

$$a(E_{\mu c}) = b(E_{\mu c})E_{\mu c} \quad \Leftrightarrow \quad E_{\mu c} = \frac{a(E_{\mu c})}{b(E_{\mu c})}. \tag{6.2}$$

The slopes of the the solid violet and brown lines in Figure 6.3 indicate that $a(E)$ and $b(E)$ are roughly constant in the momentum range displayed, so that the mean passing distance x_0 can be approximated by

$$x_0 \approx \frac{1}{b} \ln\left(1 + \frac{E_0}{E_{\mu c}}\right). \tag{6.3}$$

As a result, even if the entire ATLAS detector was one solid block of iron, a muon with an energy of $E = 200\,\text{GeV}$ would only loose an estimated 32 GeV or 16 % of its original

energy when passing from the IP through the detector. Inversely, a calorimeter layer of approximately 80 m of solid iron would be needed to only cover the mean passing distance of such a 200 GeV muon, so that a different technique is needed to measure the energy of muons in particle detectors. This explains why muon spectometers instead focus on determining the momenta of muons through measuring the curvature of muon tracks induced by very strong magnetic fields.

Finally, the fact that the ionisation losses are fairly constant over a wide range of muon momenta and that they are small compared to other particles passing the detector, leads to the name of Minimum Ionising Particle (MIP) often used to describe muons in ATLAS.

6.1.2 Calorimeter-Tagged Muons

The starting point for a CT muon is a reconstructed track in the ID as explained in section 5.1. These ID tracks are then extrapolated to the calorimeters using a combination of the Kalman filter technique and sequential fitting techniques. The procedure and its implementation for ATLAS can be found in [100]. The extrapolated tracks can then be associated to energy deposits in the calorimeter. This association sums up the energy deposits of the traversed cells per calorimeter sample since the energy deposits of muons are extremely narrow and almost entirely contained within the directly traversed cells (see Figure 4.8 in [101]). The way these energy deposits are reconstructed in the tile calorimeter is important for this study as they influence, already on a calorimeter cell level, whether or not the object reconstruction is affected by different pile-up and noise conditions. For this reason, the energy reconstruction is explained in more detail in the next section.
For the final CT muon collection, two algorithms are applied. The CaloMuonLikelihood and the CaloMuonTagger. The former is calculating the likelihood ratio of $\mathcal{L}_\mu / \left(\mathcal{L}_\mu + \mathcal{L}_{\text{bkg}}\right)$, where the likelihoods are associated to the probability that the energy deposit originates from a muon or from any other object respectively. The latter represents a tool that draws its conclusion on whether or not a track corresponds to a muon by applying cuts directly on the energy deposits. This is the tool probed here, so that a much more detailed explanation can be found in subsection 6.2.3.

6.1.3 Cell Energy Reconstruction in the Tile Calorimeter

As briefly outlined in chapter 4, the energy deposit in the tile calorimeter is measured through PMTs, which results in a well-defined shaped signal whose amplitude is proportional to the deposited energy after signal conditioning. Therefore, the signal

energy can be recovered by estimating the amplitude from the received time samples after digitisation. In the following, the currently used reconstruction technique is explained in more detail.

Optimal Filtering (OF): The main strategy of OF is to reconstruct the energy, timing and pedestal for a signal using weights that minimise variance introduced by noise. For this algorithm, the signal S_i is given by

$$S_i = p + Ag(t_i - \tau) + n_i, \tag{6.4}$$

where p represents the electronic pedestal, A the true amplitude of the signal, $g(t_i - \tau)$ is the shape form function of the signal including noise effects, τ denotes the phase between the output of the digitiser and the peak of the shape form function and n_i is an additional noise term. A first order Taylor expansion yields

$$S_i \approx p + Ag_i - A\tau g_i' + n_i. \tag{6.5}$$

Using a_i,b_i and c_i as parameters of the algorithm denoted as OF2 weights, u,v and w can be defined as

$$u = \sum_{i=1}^{N} a_i S_i, \quad v = \sum_{i=1}^{N} b_i S_i, \quad w = \sum_{i=1}^{N} c_i S_i \tag{6.6}$$

where N represents the number of available samples. In order to relate these quantities to Equation 6.5, the conditions

$$A = \langle u \rangle, \quad A\tau = \langle v \rangle, \quad p = \langle w \rangle, \tag{6.7}$$

are required, leading to a series of constraints for the OF weights. The noise contribution n_i has the effect that the distribution of the calculated u,v and w values has a non-zero variance. Therefore, OF tries to minimise the effect of noise on the signal reconstruction by choosing the weights that minimise $\mathrm{Var}(u)$, $\mathrm{Var}(v)$ and $\mathrm{Var}(w)$. One assumption of the optimisation procedure is that the noise contributions follow a normal distribution, which is the case for electronic noise, but not, for example, for pile-up effects. These effects can be approached by computing specific pile-up related weights or by an adapted algorithm as found in [102]. A more detailed documentation on OF in general is given found in [103].

6.2 Analysis Outline

6.2.1 Muon Reconstruction Efficiency

In order to study the noise scenarios mentioned above and to benchmark their impact, a figure-of-merit is defined which is the reconstruction efficiency of muons using the calorimeter tagging technique $\varepsilon_{\mathrm{CT}}$ given by

$$\varepsilon_{\mathrm{CT}} = \frac{\text{No. CT reconstructed and true-matched muons}}{\text{No. of true muons}}, \tag{6.8}$$

where the number of CT reconstructed muons contains only those muons which have successfully been matched to a truth muon to avoid counting other particles that have erroneously been reconstructed as muons (fakes). The efficiency $\varepsilon_{\mathrm{CT}}$ therefore also includes a truth-match efficiency[1]. This efficiency can then be studied in dependence of several variables, e.g. p_{T} and η, and in dependence of different noise or pile-up scenarios.

For this study, the noise scenarios (further described in subsection 6.2.4) are introduced simply by drawing random numbers from a normal distribution in an attempt to roughly emulate real noise or efficiency loss effects in a very generalist approach.

An unbiased and complete study would need to introduce certain noise scenarios, depending on the effects under investigation, through additional noise terms into the CaloMuonTagger tool, which is partly responsible for the CT muon collection and which is described in more detail later on. These additional noise terms need to be implemented at the time of building the muon collection during the reconstruction step, as all the effects would then directly be propagated to the final muon object. A simplified and incomplete chain of the data set generation is depicted in Figure A.1 and described in slightly more detail thereafter. Rebuilding the muon collection would need to be done for every choice of parameters and noise type and would be very involved in computing time, development time and overall person power, making it very difficult to proceed. In order to nevertheless obtain an estimate of the reconstruction efficiency within a reasonable time-frame, a different approach is chosen here.

Instead of introducing the noise terms during the production step, they are applied on the already built muon collection, or rather on the associated energy deposits in calorimeter samples, as this is the information used by the CaloMuonTagger. The information needed is therefore the energy deposits in the calorimeter that are asso-

[1] The truth matching procedure is done using the ATLAS internal MCTruthClassifier tool [104].

ciated to muons, the full (kinematic) information of the reconstructed muon collection built during the production step and the complete collection of generated (truth) muons.

Using data sets containing this information, it is possible to create a new *ad-hoc* reconstructed muon collection based on the outcome of the renewed CaloMuonTagger decision using the energy deposits that have been altered according to the noise scenario in question. This ad-hoc muon collection will then represent the numerator in Equation 6.8, while the truth muon collection represents the denominator. This way, $\varepsilon_{\mathrm{CT}}$ can easily be calculated inclusively or in dependence on any variable accessible for both reconstructed muons and true muons.

The main caveat of this approach is that it can only ever result in a drop in reconstruction efficiency or in leaving it constant, as its starting point is always the already existing muon collection. More precisely, the introduction of noise cannot lead to reconstructing additional muons (from true muons or otherwise) because it has no access to the energy deposits of these objects (so it cannot alter them). This also means that the obtained efficiency is only an estimator for the efficiency that would be obtained with the full approach as the effects of identifying new objects as muons and the effect of losing objects that have been reconstructed as muons before the introduction of noise will counteract each other. Additionally, this approach does not allow to study the effect of noise on the purity of the muon collection, which is a major limitation.

6.2.2 MC Samples and Object Definitions

Due to the large amount of prompt muons in the final state, i.e. muons originating from Z or W boson decays, simulated $Z \to \mu^+\mu^-$ events are a natural choice for this analysis. In order to study the effect of pile-up levels expected for the HL-LHC phase, three otherwise identical samples are used for values of $\langle\mu_{\mathrm{pu}}\rangle \approx$ 80,140,200, where $\langle\mu_{\mathrm{pu}}\rangle$ denotes the average number of interactions per bunch crossing. For the generation and showering of the physical process, Powheg+Pythia *8* [105, 106] is used with the *AU2CT10* tune. The centre-of-mass energy is set to $\sqrt{s} = 14\,\mathrm{TeV}$ and an upgraded detector geometry for the ITK as described in subsection 4.2.3 is used.

The full list of MC samples that are used can be found in Table A.1 in the appendix. For these samples, events are selected that have at least one muon, where muons are required to fulfil the *Medium* identification. There is no requirement on the isolation of the muons nor on their p_{T}.

In order to estimate the increase of sensitivity for the search for four top quarks achieved by including CT muons, an MC sample of the standard model four top quark production is used. A more detailed description of this sample can be found in subsection A.II.

6.2.3 CaloMuonTagger

The CaloMuonTagger is a cut based decision tool that flags an ID track as compatible with originating from a muon if the energy deposits in the calorimeter that are associated to this track fulfil certain cut conditions. A full description of the tool can be found in [101].

There are three types of cut conditions: *veto*, *diffLow* and *sig* which veto a track as a muon if the energy deposited in the corresponding calorimeter layer is greater than a cut value for the first condition or if it is lower than a cut value for the last two conditions. The cut values themselves can be set by the user and two WPs (i.e. predefined cut-value menus) are defined: *Loose* and *Default*, where, contrary to the name, Loose is used in most analyses in ATLAS.

The actual cut value that is applied depends on the transverse momentum p_{T} of the muon candidate and the user has to define two cut values per cut condition and calorimeter layer. One that will be applied if $p_{\mathrm{T,muon}} < p_{\mathrm{T,\,min}}$ and the second which is applied if $p_{\mathrm{T,muon}} > p_{\mathrm{T,\,max}}$, where $p_{\mathrm{T,\,min}}$ and $p_{\mathrm{T,\,max}}$ are also user defined parameters of the tagger and by default set to $p_{\mathrm{T,\,min}} = 15\,\mathrm{GeV}$ and $p_{\mathrm{T,\,max}} = 35\,\mathrm{GeV}$. For transverse momenta in between those two values, the corresponding cut value is a linear interpolation between the two values given by the user, where the interpolation is given by

$$\text{cut value} = \text{cut1} + (\text{cut2} - \text{cut1}) \times \frac{p_{\mathrm{T,muon}} - p_{\mathrm{T,\,min}}}{p_{\mathrm{T,\,max}} - p_{\mathrm{T,\,min}}}. \tag{6.9}$$

Figure 6.4 shows the energy deposited in the first layer of the tile calorimeter barrel which is associated to the ID track in MeV in bins of the muon's (reconstructed from the ID track) p_{T} in GeV. The two coloured lines represent the p_{T} dependent cut values, where the arrows indicate the region that is accepted by the cut. The tagger accepts the track as a muon if all of the cut conditions are met.

In order to quickly reevaluate the decision of the CaloMuonTagger after changing the value of the deposited energy to emulate noise or efficiency losses, the tagger itself has been fully reimplemented in a standalone C++-tool by the author. The validation of this implementation against the standard implementation in the main ATLAS framework athena can be found in subsection A.III.

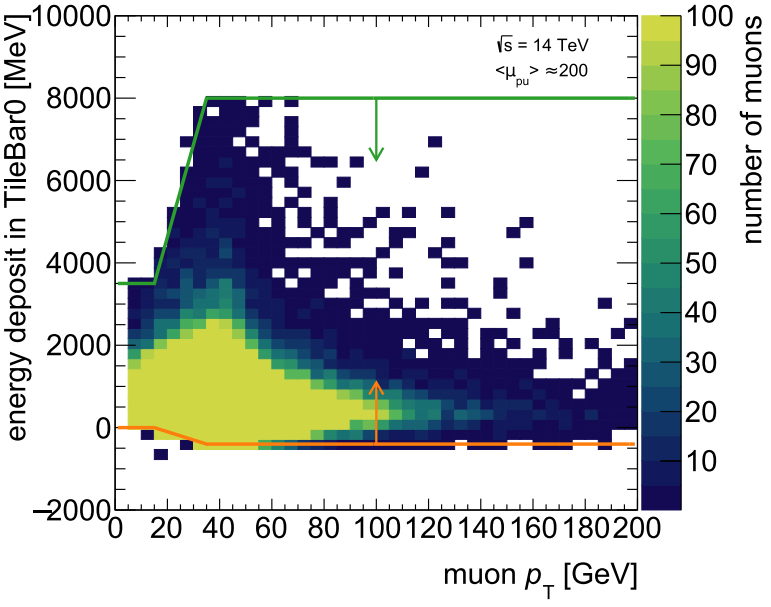


Figure 6.4: Nominal energy deposits associated to muons in the first layer of the tile barrel in MeV versus the corresponding p_{T} of the muon in GeV. The colour code represents the number of simulated muons in each bin. The orange and green lines correspond to cut values set for the CaloMuonTagger, the arrows indicate the acceptance region of those cuts.

Figure 6.5 shows $\varepsilon_{\mathrm{CT}}$ in dependence on η and p_{T} of the truth muon for the nominal simulated energy deposits, i.e. without noise. The figure shows a clear structure over the entire p_{T} spectrum in the η distribution depending on the corresponding sub-detectors. The highest efficiency is obtained for the most central part $|\eta| < 0.8$, where the calorimeter is the most instrumented. A drop in efficiency can then be seen between the tile barrel (TileBar) and the tile extended barrels (TileExt) at $|\eta| < 0.8$, between TileExt and the hadronic end caps (HEC) at $|\eta| < 1.7$ and where the electromagnetic end caps (EMEC) are reduced from three to two layers ($|\eta| < 2.5$). Seeing that CT muons require a track in the ID, it is important to keep in mind that the data sets used for this study assume a geometry of the ID that has been proposed for the HL-LHC phase and which covers a region of $|\eta| < 4$.

Since this study focuses on the effect on the tile calorimeter, most distributions are shown for $|\eta| < 1.7$. Additionally, distributions for $|\eta| < 0.1$ are shown as that region represents a blind spot of the muon system, so that the performance of the calorimeter in that region is especially interesting.

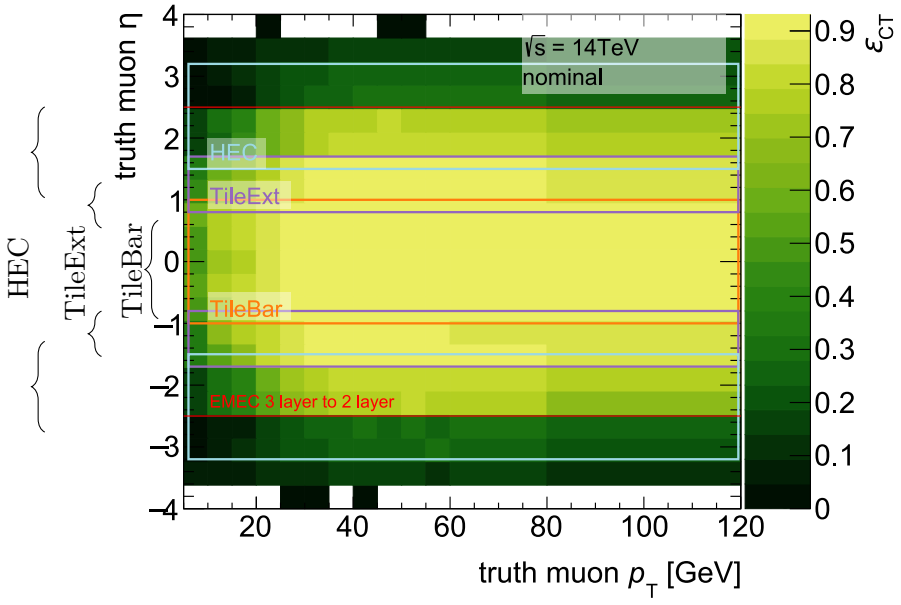


Figure 6.5: Calorimeter tagged muon reconstruction efficiency $\varepsilon_{\mathrm{CT}}$ in dependence of the η and p_{T} of the truth muon. The coloured boxes represent the η-coverage of different sub-detectors of the calorimeter.

6.2.4 Noise

Two different types of noise are introduced in three different scenarios. The two different types will be denoted *additive noise* and *multiplicative noise* in the following and are defined as

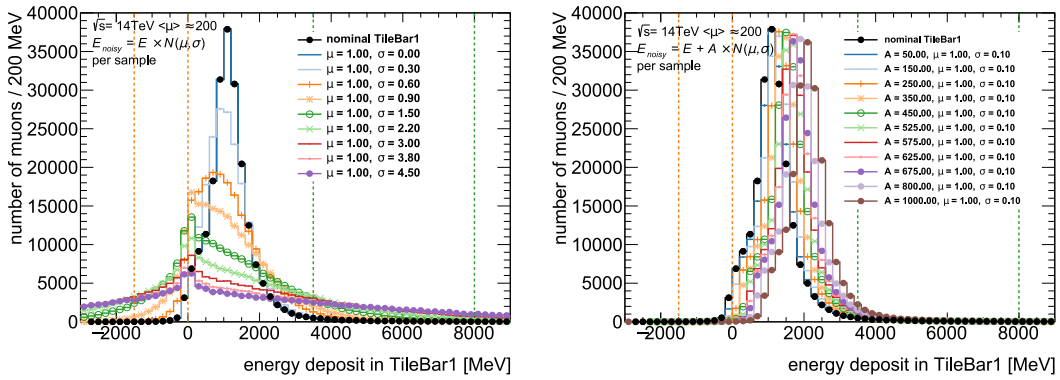
$$E_{\mathrm{dep,\ noisy}} = E_{\mathrm{dep}} + A \times \mathcal{N}(\mu,\sigma) \tag{6.10}$$

$$E_{\mathrm{dep,\ noisy}} = E_{\mathrm{dep}} \times \mathcal{N}(\mu,\sigma), \tag{6.11}$$

respectively. E_{dep} represents the energy deposited in a given calorimeter layer, A is a user defined scaling parameter called amplitude and $\mathcal{N}(\mu,\sigma)$ represents a random variable drawn from a normal distribution with mean μ and variance σ.
Here, the random variable is either drawn once for each muon individually and then the same random variable will be applied to all energy deposits in the tile calorimeter associated to this muon (uniform) or it is drawn for each muon and each associated deposit independently (per sample).

The additive noise is introduced in order to emulate noise that comes from simultaneous energy deposits originating, for example, from additional particles. Those types of effects add to the energy originally deposited by the muon and are therefore reasonably

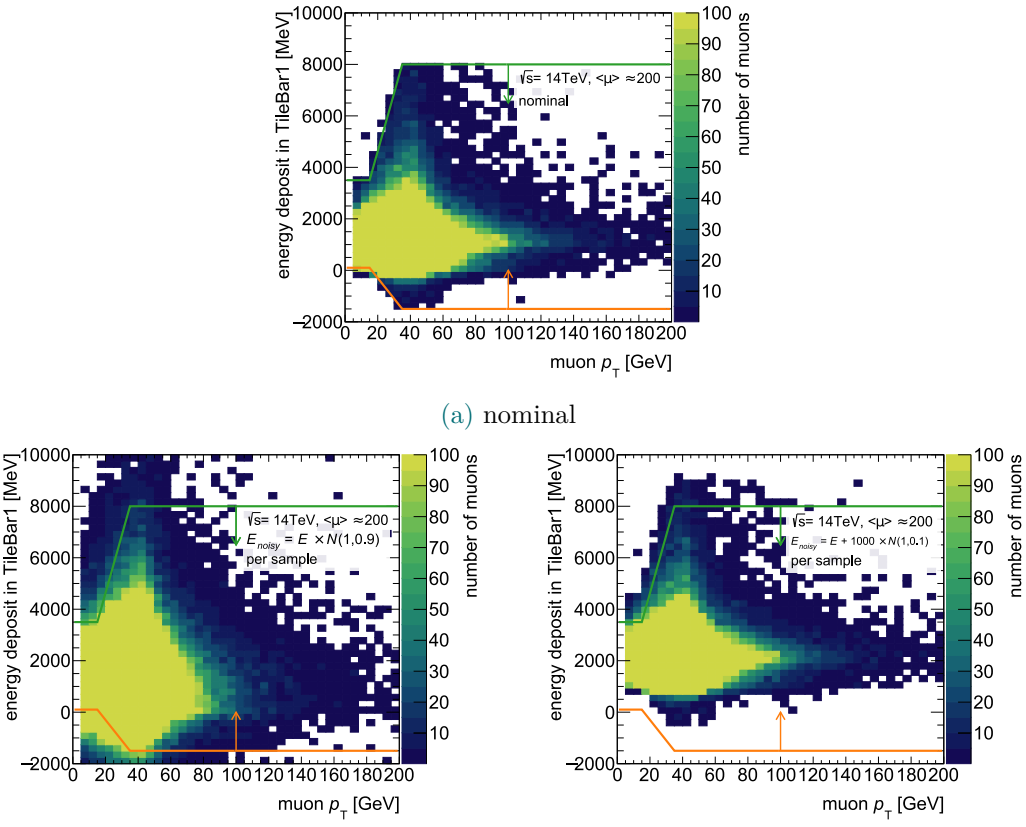
well emulated if $A \times \mathcal{N}(\mu,\sigma) > 0$. This is not strictly required in the analysis, but rather assured by the choice of parameters, where the mean and variance are fixed to $\mu = 1$, $\sigma = 0.1$ and A is scanned in a range of $A \in [1\,\text{MeV},1000\,\text{MeV}]$.
The multiplicative noise is the more interesting case and it is introduced in order to emulate two types of effects: A smearing of the deposited energy to account for a loss in resolution and the loss of energy reconstruction efficiency of the ageing optical system. This effect has not been estimated before, and has been requested during the edition of [65].
For the former the mean will be set to $\mu = 1$ and the variance is scanned, while for the latter the variance is set to $\sigma = 0.1$ and the mean is scanned in the range of $\mu \in [0,1)$. The effect of different noise scenarios on a specific tile calorimeter layer can be seen in Figure 6.6, where the deposit is shown for the second layer of the tile calorimeter barrel. Different colours represent different parameter choices for the normal distribution or its amplitude in the noise term. The vertical dashed lines indicate the cut values used by the CaloMuonTagger, where the tighter cut value is applied on muons with $p_\text{T} < 15\,\text{GeV}$ and the looser cut value is applied on muons with $p_\text{T} > 35\,\text{GeV}$. For muons with $15\,\text{GeV} < p_\text{T} < 35\,\text{GeV}$, a linear interpolation between the cut values is used, as explained above. The cuts represented by the orange dashed lines accept the muon if the deposit is larger than the cut value, while the cuts represented by the green dashed lines require the deposit to be lower than the cut value. These distributions only show the case for $\langle\mu_\text{pu}\rangle \approx 200$ and no specific cut on the η distribution of the muon is required as this is already limited by the dimensions of the tile calorimeter barrel. The displayed uncertainty in this, and in any of the following figures in this chapter originates from the statistical uncertainty due to the size of the MC simulation.
For the multiplicative noise, the smearing caused by the normal distribution is clearly visible and a shift of the mean towards lower values can be observed, which is likely caused by the fact that the random variable can be negative, and thus can flip the sign of the energy deposit. Negative energy deposits are of course nonphysical and a result of calibration and reconstruction effects. Since the mean of the energy deposit distribution is positive, more positive deposits become negative through this flip than vice versa, resulting in the observed behaviour.
The energy deposit distribution for the additive noise shows the expected behaviour, where a slight smearing ($\sigma = 0.1$) of the nominal distribution is visible and otherwise only the mean of the distribution is shifted according to the amplitude parameter A.
Figure 6.7 again shows the energy deposit in the second layer of the tile calorimeter barrel, but this time only for one specific choice of parameters (or nominal) and against the p_T of the muon. The shown distribution is also representative of the other layers in the tile barrel. The parameters of choice are $\mu = 1$ and $\sigma = 0.9$ for the multiplicative



(a) multiplicative noise (b) additive noise

Figure 6.6: Energy deposits in the second layer of the tile barrel associated to muons for different noise scenarios. The vertical dashed lines represent the *Loose* cut of the CaloMuonTagger. The orange lines represent cuts requiring values to be larger than the cut and the green lines represent cuts that accept deposits that are lower than the cut value. There are two lines per colour as the cut is p_{T} dependent, so that the two lines represent the lower and the upper bound of the cut value.

noise and $A = 1\,\mathrm{GeV}$ for the additive noise. The green and orange lines again represent the cut values of the CaloMuonTagger. The comparison of the nominal to the noise distributions shows nicely that the introduced noise terms can have a significant effect on the decision of the CaloMuonTagger.



(a) nominal

(b) multiplicative noise

(c) additive noise

Figure 6.7: Energy deposits associated to muons in the second layer of the tile barrel for nominal and for different noise types. The colour code represents the number of simulated muons in each bin. The orange and green lines correspond to cut values of the *Loose* working point of the CaloMuonTagger, the arrows indicate the acceptance region of those cuts.

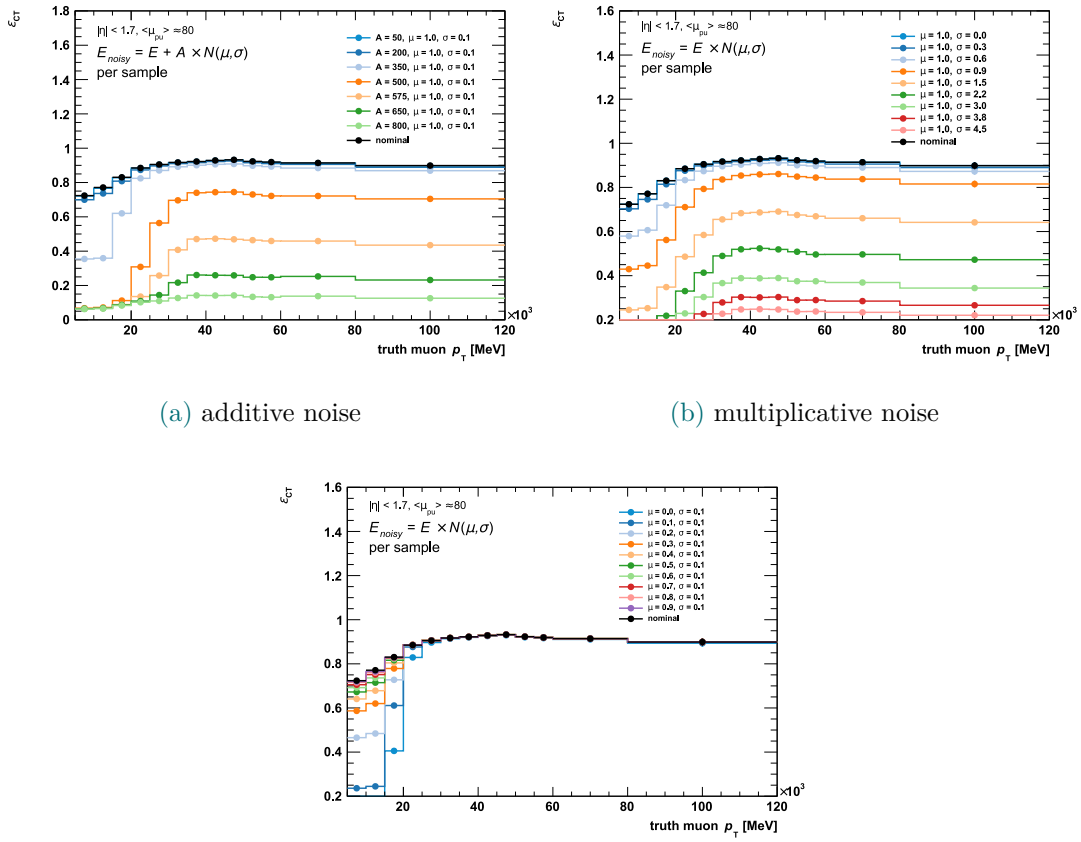
6.3 Analysis Results

6.3.1 Efficiency Dependence on Noise

The distributions of $\varepsilon_{\mathrm{CT}}$ are shown for the additive noise in Figure 6.8a, for the multiplicative noise with a fixed mean and varying standard deviation in Figure 6.8b and for the multiplicative noise with a fixed standard deviation and varying mean in Figure 6.8c, where it is plotted in dependence on the p_{T} of the truth muon for the tile calorimeter region of the calorimeter $|\eta| < 1.7$ and for a pile-up scenario of $\langle\mu_{\mathrm{pu}}\rangle \in [70,90]$. The different colours represent different parameter sets of the normal distribution or its amplitude in the noise term.

Whereas the effect of the noise can be clearly seen in the case of Figure 6.8a and Figure 6.8b, there is next-to-no effect visible for muons with a $p_{\mathrm{T}} > 35\,\mathrm{GeV}$ for the

multiplicative noise if the mean is chosen to be $\mu \in [0,1)$, i.e. the emulation of efficiency loss effects, see Figure 6.8c. This is true even for the very extreme case of $\mu = 0$, which leads to an expected noisy energy deposit of $\langle E_{\text{dep,noisy}} \rangle = 0$. This can be expected seeing that the lower energy thresholds implemented in the CaloMuonTagger are mostly 0 or even negative (see Figure 6.4 or Table A.2). This noise scenario only has an effect once the more stringent cut value plays a role, i.e. for muons with $p_{\text{T}} < 35\,\text{GeV}$. Even then, a conservative emulated efficiency drop of 50 % leads to no more than a one percentage point drop in the reconstruction efficiency.
This shows that the cut values of the loose working point are indeed quite loose regarding a minimum signal requirement, directly leading to the question of the reconstruction efficiency for other particles / energy deposits as muons, called fake efficiency, which is desired to be low for physics analyses. However, this cannot be studied with the current set up as has already been demonstrated above. As a result, only the cases of the additive noise and the multiplicative noise with fixed mean are considered in the following.



(a) additive noise

(b) multiplicative noise

(c) multiplicative noise low mean

Figure 6.8: CT muon reconstruction efficiency $\varepsilon_{\mathrm{CT}}$ against the transverse momentum p_{T} of the truth muon for the nominal and for different noise scenarios for $|\eta| < 1.7$ and $\langle\mu_{\mathrm{pu}}\rangle \in [70,90]$. The colours indicate different sets of parameters for the random noise distribution. The displayed uncertainty is MC statistical, but so small that it is not visible.

All distributions in Figure 6.8 show an expected p_{T} dependent loss of $\varepsilon_{\mathrm{CT}}$ for increasingly noisy scenarios. In order to compare the evolution of $\varepsilon_{\mathrm{CT}}$ with the choice of parameters for the noise distribution in a more visible way, a constant function $c = p0$ is fitted to the $\varepsilon_{\mathrm{CT}}$ distribution for every parameter set, noise type and pile-up scenario generated. In order to render the distributions more easily comparable and because for many physics analyses the reconstruction efficiency for higher values of p_{T} are of interest, c is only fitted in a window of $p_{\mathrm{T}} \in [40\,\mathrm{GeV}, 120\,\mathrm{GeV}]$. This region is denoted plateau in the following, so that the result of the fit for $p0$ is called plateau efficiency. The procedure is illustrated in Figure 6.9, where the distribution of $\varepsilon_{\mathrm{CT}}$ in bins of the truth muon's p_{T} is shown for nominal energy deposits, i.e. without added noise. The

fitted constant function c is depicted as a red dashed line and the green area represents the value of the plateau efficiency $p0$ plus or minus the uncertainty on the plateau efficiency as obtained from the fit. The goodness-of-fit test of $\chi^2/ndf = 1.1$ suggests that the fit has converged well, which is expected given the simple shape of both the distribution and of the fit function. The fit is performed using the minuit package as implemented in ROOT [107, 108].

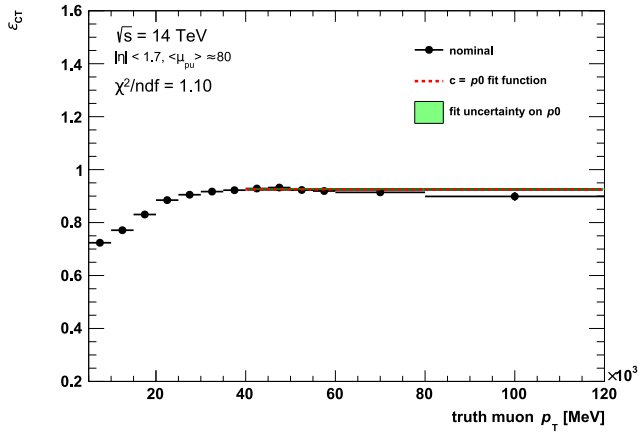
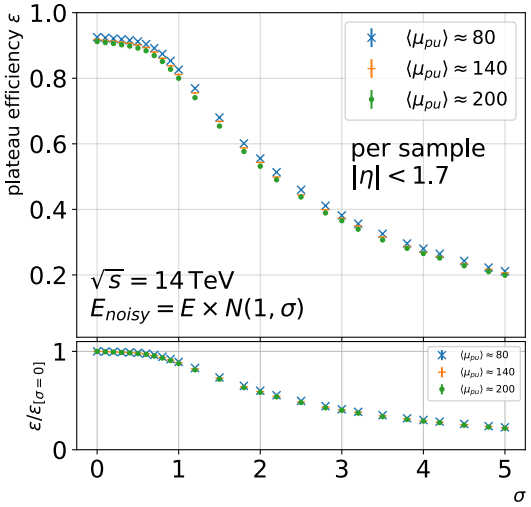


Figure 6.9: The distribution of $\varepsilon_{\mathrm{CT}}$ in bins of the p_{T} of the truth muon for nominal energy deposits, i.e. without added noise. The fitted constant function $c = p0$ is depicted as a red dashed line and the green area represents the value of the plateau efficiency $p0$ plus or minus the uncertainty on $p0$ as obtained from the fit.

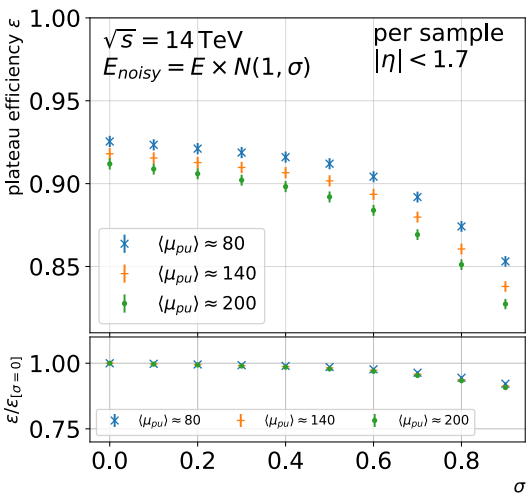
The result of this procedure is shown in Figure 6.10 for the multiplicative noise and in Figure 6.11 for the additive noise. The former shows the dependence of the plateau efficiency on the standard deviation of the normal distribution of the noise term in the top two plots and in dependence of the mean in the bottom plot. The different colours represent the behaviour for different pile-up scenarios and show the expected behaviour of a deterioration of the plateau efficiency for busier environments. It is important to note that the underlying simulated events for different pile-up scenarios are the same and thus the opposite of statistically independent. This should be taken into account when interpreting the statistical uncertainty given on each of the distributions individually.

Figure 6.10a shows the plateau efficiency for the multiplicative noise for a very large range of possible choices for σ of up to $\sigma = 5$. Considering that this noise scenario tries to emulate a loss of energy measurement resolution, these very high values are not very likely scenarios. Therefore, the same distribution is shown in Figure 6.10b for values of up to $\sigma = 1$. The deterioration is not very large for $\sigma < 0.6$, where the plateau efficiency drops about two percentage points, but considerably more noticeable for

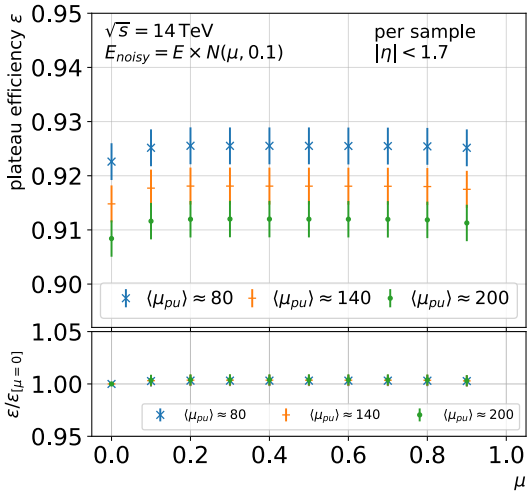
higher values of σ. Again here, higher values of σ are much less expected, so that the muon reconstruction efficiency can be considered fairly stable against moderate losses of resolution, even without re-optimising the CaloMuonTagger for these scenarios. The distributions are shown normalised by their value for the lowest scanned parameter in the bottom part of these figures so as to allow for a comparison of the shape of the distribution for different pile-up scenarios. This shows that the effect of pile-up on the impact of noise on the reconstruction efficiency is negligible. As a comparison to the overall deterioration caused by increasing pile-up, Figure 3.7 in [65] states an effect of up to four percentage points for the plateau region, when comparing the pile-up scenarios of ATLAS run 2 to $\langle \mu_{\text{pu}} \rangle \approx 200$ for the muon gap region $|\eta| < 0.1$.
As already seen and discussed above, Figure 6.10c shows next-to-no effect of the noise term on the obtained plateau efficiency.



(a) All variations of σ.



(b) Only small variations of σ.



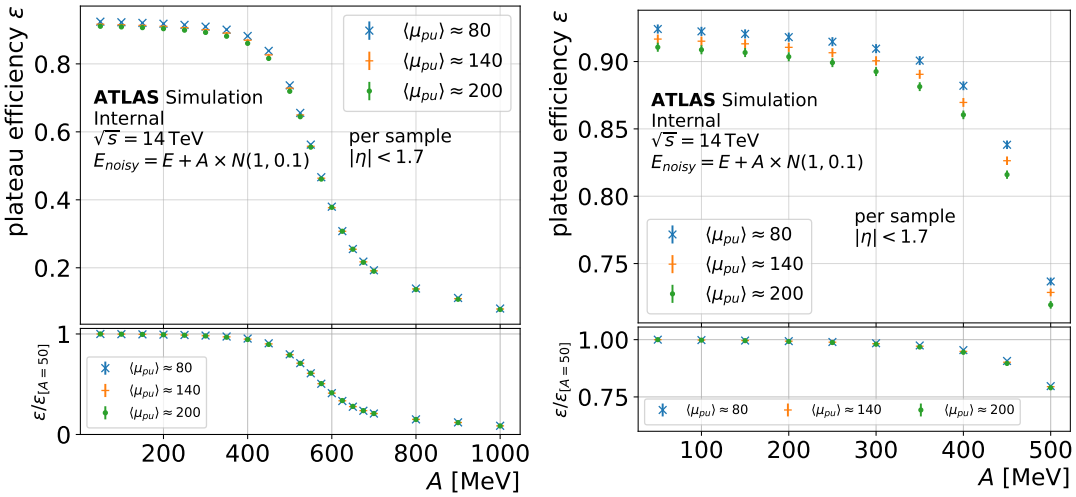
(c) Variations of μ.

Figure 6.10: Plateau efficiencies resulting from the fit of ε_{CT} in the truth muon p_{T} distribution for different noise parameters of the multiplicative noise. The two figures at the top vary the standard deviation of the noise term, while the bottom figure varies the mean of the noise term. The colours indicate different pile-up scenarios

Figure 6.11a shows the plateau distribution in the additive noise case for amplitudes of up to $A = 1\,\text{GeV}$ and Figure 6.11b shows the same distribution zoomed in to values of up to $A = 500\,\text{MeV}$. The distributions show an almost heavy-side-like behaviour, where the efficiency drop for amplitudes $A < 400\,\text{MeV}$ is very low (two to three percentage points) and very large for higher values of A (drop from 90 % to 10 %). It has already been discussed in subsection 6.2.4 that the effect of the additive noise on the energy deposits is mainly a shift in the mean of an almost Gaussian shaped distribution. The plateau efficiency therefore has a similar shape to the inverse Cumulative Density Function (CDF) of the normal distribution (i.e. the CDF with the sign of the random variable flipped), since the shift of the mean caused by the noise term "pushes" the Gaussian shaped energy deposit distribution over the cut value of the CaloMuonTagger.

This would suggest that the shape of the plateau efficiency distribution also depends very strongly on the choice of the σ parameter of the normal distribution in the noise term, which is therefore shown in Figure 6.12a. This figure shows the plateau efficiency against different choices of A but only for one pile-up scenario ($\langle\mu_{\text{pu}}\rangle \approx 200$). The different colours represent different choices of σ and the lines are linear interpolation between the efficiencies corresponding to the simulated values of A for better visualisation. As a reference, the cumulative density function Φ of the normal distribution is shown for the same values of σ in Figure 6.12b. The distribution in Figure 6.12a does not correspond exactly to the functions inverting the argument in Figure 6.12b, which is to be expected seeing that Figure 6.12a shows the effect of adding an amplitude scaled random variable drawn from a normal distribution with $\mathcal{N}(1.0,\sigma_{\text{noise}})$ to a series of energy deposits each "drawn" from more or less Gaussian shaped distributions (the nominal energy deposits) which would correspond to varying values of μ_i and σ_i.

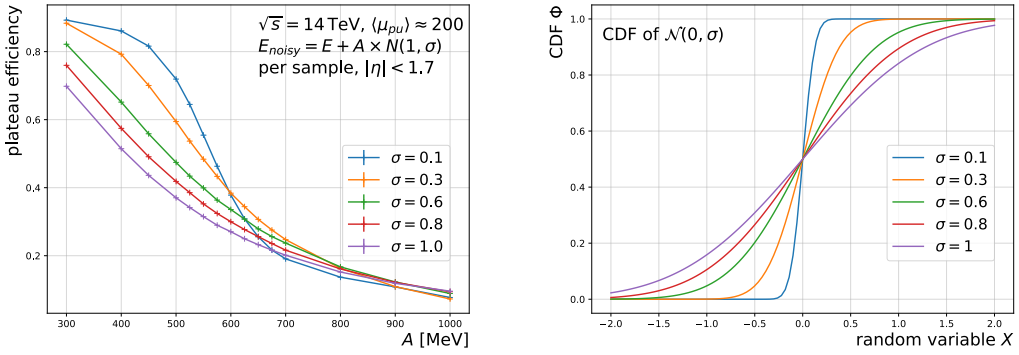
As for the multiplicative case, the effect of different pile-up scenarios on the impact of noise on the reconstruction efficiency is negligible.



(a) the full simulated spectrum

(b) for low simulated values of A

Figure 6.11: Plateau efficiencies resulting from the fit of $\varepsilon_{\mathrm{CT}}$ in the truth muon p_{T} distribution for different noise parameters of the additive noise term. The colours indicate different pile-up scenarios



(a) Plateau efficiencies

(b) Cumulative density functions

Figure 6.12: On the left, plateau efficiencies resulting from the fit of $\varepsilon_{\mathrm{CT}}$ in the truth muon p_{T} distribution against different choices of the amplitude parameter A of the additive noise term. The different colours represent different choices of the standard deviation σ of the normal distribution used for the noise term. The lines are linear interpolations between the simulated parameters, for visualisation purposes. On the right, different shape examples of the cumulative density function of normal distributions with $\mu = 0$ in dependence on the random variable X are shown, where the different colours represent different values for σ.

6.3.2 Conclusion

The studies above investigate the effect of different pile-up scenarios and the effect of noise scenarios in energy deposits in the different layers of the tile calorimeter on the CT muon reconstruction efficiency. This is done in the context of the phase II upgrade of ATLAS for the HL-LHC period and thus under extreme pile-up conditions and using the proposed geometry of the different sub detectors.
The calorimeters do not currently play a major role in the muon reconstruction, but already in the introduction to this chapter it has been argued that this might be an interesting improvement to look at, especially for analyses with high muon multiplicities, among which the search four four top quarks can be found. In order to keep this possibility open, it is important to ensure a high quality of the muon reconstruction in the calorimeter which warrants for a study on the impact of the upgrade on this particular sub module. A first step of this has been achieved with the studies presented here, where it shows that the impact of strongly increased pile-up from $\langle \mu_{\mathrm{pu}} \rangle = 80$ to $\langle \mu_{\mathrm{pu}} \rangle = 200$, otherwise using the same setup, is within three percentage points for the reconstruction efficiency $\varepsilon_{\mathrm{CT}}$.
The introduction of a multiplicative noise to the energy deposits, essentially emulating energy resolution losses due to ageing detector components, shows that $\varepsilon_{\mathrm{CT}}$ may drop around two percentage points for $\sigma < 0.6$, where σ represents the standard deviation of the normal distribution introducing the noise. This translates very roughly to a loss in resolution of up to 60 % which can be considered conservative in the context of the proposed upgrade. The efficiency drops significantly only for higher values of σ which would correspond to extreme losses of signal resolution, that are unlikely to occur.
The behaviour of the muon reconstruction efficiency in dependence of a loss in energy reconstruction efficiency, showed that it is fairly independent from this effect due to very loose requirements on the minimum energy deposited in the tile calorimeter. Indeed, it becomes clear that for this study not only the reconstruction efficiency, but more importantly the purity of the reconstructed muons should be considered. Unfortunately, this cannot be achieved with the analysis setup outlined above, but instead two alternative approaches have been discussed.
Additionally to the main conclusion regarding the energy resolution, an effect has been studied that would randomly increase the deposited energy in a calorimeter sample, which could, for example, be caused by a different particle crossing the same cells as the muon at the same time. The effect of a surge in the reconstructed / deposited energy is within a few percentage points for energies $E < 400\,\mathrm{GeV}$ and very large (70 percentage points) for $E > 600\,\mathrm{GeV}$. The shape of the effect strongly depends on

the arbitrarily chosen values of the Gaussian noise so that a detailed study and good simulation of the noise effects remains as an outlook.
In all noise scenarios differences in the amount of pile-up does not change the impact of the noise on the reconstruction efficiency in any significant way.

The studied effects only exhibit a small impact on the performance of muon reconstruction using calorimetry information, even during HL-LHC conditions. This is true even though none of the routines, WPs or weights used in the (energy) reconstruction steps were optimised and instead remained the same as for Run-II conditions.
The proposed upgrade scenario studied here can therefore be considered highly effective in ensuring a high quality of CT muon reconstruction, potentially even allowing to spread their use to more analyses in the future.

6.4 Optimisation of Energy Reconstruction beyond Muons

It has been pointed out at multiple occasions throughout this chapter, that not only the hardware components are upgraded to cope with the significantly increased activity of the HL-LHC phase, but rather that new or improved algorithms for the reconstruction procedure are needed as well. In order to design these algorithms, such as the routine for tile calorimeter energy reconstruction presented in subsection 6.1.3, sets of Figures of Merit (FOMs) need to be defined. These FOMs provide the possibility to tune the free parameters of any algorithm to obtain optimal results and subsequently to render the performance of different algorithms comparable.

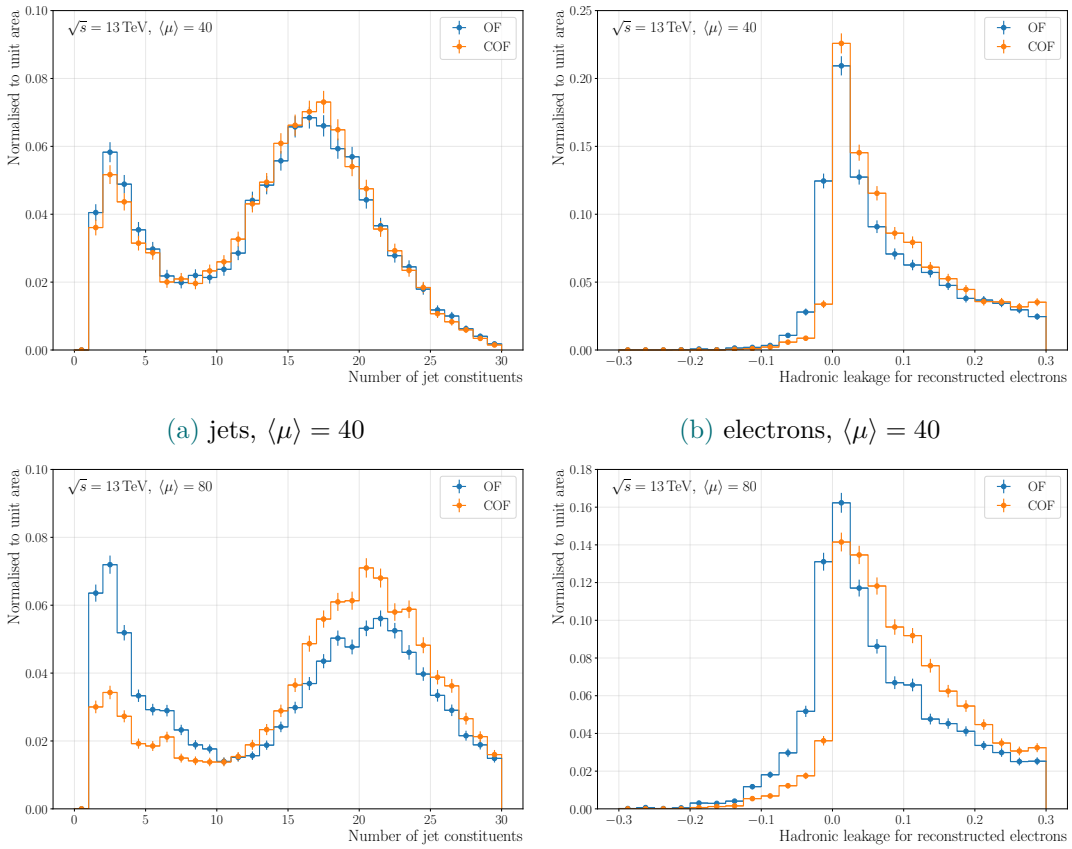
The most important global FOM in this context is the quality of the reconstructed objects used in physics analysis for each of the optimisation steps. However, it is often not practical to re-execute the entire reconstruction procedure in order to scan for optimal parameters of a subroutine of this procedure. Therefore, local FOMs are often defined for the subroutines that can be quickly computed without a full run of the entire software suite.
This is, for example, the case for the design of an alternative cell energy reconstruction technique, called Constrained Optimal Filter (COF), as described in [102]. Here, the algorithm is tuned towards optimising the resilience of the algorithm against pile-up in the form of energy bias and resolution. These FOMs are likely to be a good choice and propagate to the final reconstructed objects, as they are unlikely to be detrimental to any of the subsequent steps. However, this is a hypothesis that needs to be tested and, more importantly, the overall impact of this choice of algorithms on the final objects should be studied.

Unfortunately, these studies are often missing, in part because of technical difficulties, but also in part because of different expertise of the groups that optimise detector components and energy reconstruction techniques and the groups that conduct physics analyses using the final reconstructed objects.

For this reason, a small framework is developed by the author and in collaboration with the Federal University of Rio de Janeiro that tries to bridge this gap. The first prototype of this tool is developed in light of comparing the performance of OF and COF used to reconstruct the energy in tile calorimeter cells.

Here, the main idea is that the group developing a new algorithm for the reconstruction procedure can simply feed a data set generated with the changed procedure to the tool, which will then automatically generate a large variety of control figures that compare the properties of final reconstructed objects that are important for a large variety of physics programmes or that are sensitive to changes in energy reconstruction. Moreover, this tool has a modular structure and it is highly configurable so that it is, in principle, applicable for all kinds of changes in the reconstruction procedure.
One variable that is particularly sensitive to minor differences in the reconstructed cell energies is the number of constituents of a reconstructed jet. This number is heavily affected by the energy thresholds of a reconstruction algorithm and it represents an important information for identifying merged jets. The distributions of this variable for the same simulated $t\bar{t}$ data set, where the tile calorimeter cell energy is reconstructed with either the OF or the COF algorithm is shown in Figure 6.13. This simulation assumes Run-II (Run-III) conditions, and thus a number of average interactions per bunch crossing of $\langle\mu\rangle = 40$ (80). While the difference between the two algorithms is small for Run-II conditions, the number of jet constituents is significantly larger for COF than for OF when dealing with double the amount of pile-up events. This is due to the fact that the outcome of COF for this variable is more or less stable for different pile-up conditions, while OF is heavily affected.

Another such variable is the hadronic leakage for reconstructed electrons, which is the ratio of cell energies associated to the electron that belong to the hadronic calorimeter over the energies of the EM calorimeter. This variable plays an important role for the identification of electrons and this leakage usually is in the order of a few GeV, and thus highly susceptible to changes in the reconstruction of cell energies. Here, the distributions widen for both algorithms and more studies are needed to properly understand the behaviour. The point of the figures in the context of this dissertation, however, is simply to illustrate a typical output of the developed tool.



(a) jets, $\langle\mu\rangle = 40$

(b) electrons, $\langle\mu\rangle = 40$

(c) jets, $\langle\mu\rangle = 80$

(d) electrons, $\langle\mu\rangle = 80$

Figure 6.13: Number of jet constituents and hadronic leakage for electrons obtained in data sets of simulated $t\bar{t}$ production, where the tile calorimeter cell energy in is reconstructed with either the COF or the OF algorithm. The number of average interactions of bunch crossings is set to $\langle\mu\rangle = 40$ (80), which corresponds to the roughly average or expected number for Run-II and Run-III, respectively.

This tool is currently still in an early prototype stage and will likely be documented in a technical report following the publication of this dissertation.

7 Evidence of Four Top Quark Production

Following the motivations laid out in chapter 3, the search for the SM simultaneous production of four top quarks is a very interesting and promising topic at the verge of observation with currently available data sets. Thus, the search for this process using the full Run-II data set taken by the ATLAS collaboration amounting to $\mathcal{L} = 139\,\text{fb}^{-1}$ at $\sqrt{s} = 13\,\text{TeV}$, is presented in this chapter.
This analysis represents the fruits of the labour of an entire analysis team so that results presented here are not the exclusive work of the author. The main contributions by the author are:

- Software development and data set production (one of three people responsible)
- Early lepton isolation optimisation
- Background and event classification based on MC truth information on which the template fit method relies
- Building, improving and testing the fit model as described in chapter 8

7.1 Analysis Overview

The final goal of this analysis is to determine whether or not the simultaneous production of four top quarks has been observed in the measured data set and if this observation is compatible with the SM prediction. This is effectively achieved by searching for an excess over the background only model in data and by evaluating this excess in terms of the kinematic properties of the expected signal.

Methodology

In the following, the overall fit procedure is described in detail as it represents the final step of this analysis. In this way, the individual parts of the fit setup are used to describe and give an overview of the different steps of the analysis.

The fit model is built using a global binned likelihood approach which allows for the combined background plus signal model (S+B) or the background only model (B) to be simultaneously adjusted to data. The Parameter of Interest (PoI) of the fit model is the signal strength μ_{4t}, defined as the ratio between the observed and the predicted signal cross section.

The fact that the background contributions are adjusted to data simultaneously to the PoI is particularly useful when a precise prediction of all backgrounds in both shape of the distribution and normalisation is not available.

This is the case for this analysis, so that a fit model is devised that allows for a constraint and a more precise determination of the normalisation of certain backgrounds by adding auxiliary measurements in the form of Control Regions (CRs). Here, CRs are regions of the phase-space enriched with one or more of the background processes, while only having a negligible contribution from the signal. Additionally, several degrees of freedom in the form of Norm Factors (NFs) are added to the fit model that represent a scale to the expected normalisation of certain background processes. These NFs are unconstrained and treated in exactly the same way as the PoI by the fit.

The likelihood of this global fit is then given by

$$\mathcal{L}(\boldsymbol{d}\,|\,\mu_{4t},\boldsymbol{\theta},\boldsymbol{\kappa}) \;=\; \left(\prod_{i\,\in\,\text{bin, regions}} \mathcal{P}_{\text{Poisson}}\left(d_i \,\middle|\, \mu_{4t} s_i(\boldsymbol{\theta}) \,+\, \sum_j \kappa_j b_{ij}(\boldsymbol{\theta}) \right) \right) \,\times\, \mathcal{G}(\boldsymbol{\theta}), \tag{7.1}$$

where $\mathcal{P}_{\text{Poisson}}(n\,|\,\lambda)$ is the Poisson probability to observe n events given an expectation of λ events, $\boldsymbol{d} = (d_0,...,d_i,...,d_{N_{\text{bin}}\times N_{\text{regions}}})$ represents the vector of yields of (pseudo-) data in each bin, where the bins of different regions are simply concatenated and μ_{4t} represents the signal strength as introduced above, so that $\mu_{4t} = \sigma_{4t}^{\text{obs}}/\sigma_{4t}^{\text{theo}}$. The vector $\boldsymbol{\theta} = (\theta_0,...,\theta_{N_{\text{syst}}})$ represents the vector of constrained Nuisance Parameters (NPs), accounting for systematic uncertainties, where the constraint is given by the Gaussian constraint term $\mathcal{G}(\boldsymbol{\theta})$. Systematic uncertainties are given by the distance of variations caused by the uncertainty to the nominal in units of standard deviations, e.g. $\theta \equiv (\xi_{\text{var}} - \xi_{\text{nominal}})/\sigma_\xi$, and can either be measured in auxiliary measurements or defined *ad-hoc*. The symbol $s_i(\boldsymbol{\theta})$ represents the yield of the simulated signal model in bin i in dependence of $\boldsymbol{\theta}$ when scaled by the signal strength μ_{4t}. Each $b_{ij}(\boldsymbol{\theta})$ represents the yield for the background process j in bin i in dependence of $\boldsymbol{\theta}$ when scaled by $\boldsymbol{\kappa}$, where in turn $\boldsymbol{\kappa} = (\kappa_0,...,\kappa_{N_{\text{bkg}}})$ denominates the vector of NFs to each background. The value of κ_i is fixed to 1 for backgrounds that are well estimated either in simulation or by other means and it is a free parameter of the fit for backgrounds where this is not the case.

The fit is performed using the RooFit framework [109] interfaced by the TRexFitter tool set [110], where the minimisation is achieved using MINUIT [107], so that the uncertainties on the estimated parameters of the fit are given by MINOS [107].

For this fit, the functions $s_i(\boldsymbol{\theta})$ and $b_{ij}(\boldsymbol{\theta})$ are obtained by interpolating between their discrete values obtained for the nominal, a variation of $+1\sigma$ and a variation of -1σ for each θ.

An unconditional fit refers to the fit where all parameters (μ_{4t}, κ_j and $\boldsymbol{\theta}$) are freely varied, whereas μ_{4t} is fixed in a conditional fit.

Following the different terms of the likelihood, three steps are needed for this analysis.

First, a SR has to be designed that yields a good separation between the signal and background processes so that the fit is correctly able to estimate μ_{4t}. For this analysis, this is achieved by using a BDT resulting in a single discriminant variable (the output weight of the BDT), and a single set of event selections that further optimise the ratio of signal over backgrounds in that region. More details of the event selection are given in section 7.2 and the signal extraction in form of a BDT is described in section 7.5.

Second, the background processes have to be estimated. To this end, additional information available through the MC, called *truth*, can be used to study the impact of different backgrounds and their origins. This, in turn, is used in the template fit method, which is applied to estimate the normalisation for those backgrounds of which the normalisation shows to be disparate between data and simulation. Additionally, a data-driven technique is used for one of the background types. A detailed description of the estimation of backgrounds is given in section 7.4.

Third, the set of systematic uncertainties, e.g. uncertainties from theory predictions, detector properties and so on, needs to be devised and tested. The impact of chosen systematic uncertainties needs to be evaluated, which is particularly the case for those uncertainties that are difficult to determine in auxiliary (self-contained) measurements, denoted *ad-hoc* in the following. More information on the systematic model and its impact on the fit and final significance is given in section 7.6 and section 8.2.

A very important tool for most complex statistical analyses and especially for those that optimise the analysis setup with regards to a high statistical significance, is the idea of blinding. A blinded analysis is an analysis that does not use the full available data set for the optimisation steps. Unblinding, i.e. revealing the full data set, is only done once the optimisations are finalised. The main goal of this procedure is to reduce the probability of over-engineering the optimisation in a way that the conclusion of the statistical test becomes biased.

Blinding is used in this analysis, resulting in two different stages of the analysis: pre and post unblinding. The blinding procedure is outlined in section 7.2 and the impact on the design of the fit model is pointed out in section 8.3.

Outcome

The final measures of interest for this analysis are the post fit signal strength $\hat{\mu}_{4t}$, the corresponding measured production cross-section $\sigma(pp \to t\bar{t}t\bar{t}) \equiv \sigma_{4t} = \hat{\mu}_{4t} \times \sigma_{4t}^{\text{theo}}$ and the expected and observed significance introduced in the following.

In order to estimate the uncertainty on the estimator $\hat{\mu}_{4t}$ and in order to define p values for inference, a test statistic q_μ[1] is defined as

$$q_\mu = -2 \ln \left(\frac{\mathcal{L}\left(\mu, \hat{\hat{\boldsymbol{\theta}}}_\mu, \hat{\hat{\boldsymbol{\kappa}}}_\mu\right)}{\mathcal{L}\left(\hat{\mu}_{4t}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\kappa}}\right)} \right) , \tag{7.2}$$

where the denominator represents the maximised, or post fit, likelihood of the unconditional fit (μ_{4t} free) and the numerator represents the maximised likelihood of the conditional fit (μ fixed) for a given μ.
In this way, the test statistic represents a measure of compatibility between the estimator $\hat{\mu}_{4t}$ and the probed value μ, where larger values of the fraction (lower values of q_μ) correspond to a better compatibility[2].

In order to use this test statistic for statistical inference, the PDF of the test statistic needs to be determined. This could be achieved in a straight forward way by generating a series of pseudo-experiments. However, this approach is often difficult to realise due to the amount of computational power needed.
For this analysis, the PDF is therefore determined by using an asymptotic approximation [111] of the test statistic under the assumption that $\hat{\mu}$ follows a normal distribution with mean $\bar{\mu}$ and width σ_μ, where $\bar{\mu}$ represents the real signal strength of the probed data set. The simplified test statistic is then given by

$$q_\mu = \frac{(\mu - \hat{\mu})^2}{\sigma_\mu^2} + \mathcal{O}\left(\sqrt{N}^{-1}\right) . \tag{7.3}$$

[1] The signal strength μ is equivalent to μ_{4t}, it is however written without the subscript here in order to stress that it represents the independent (probed) variable of q_μ

[2] By construction, the unconditional likelihood is strictly larger than the conditional one, $\mathcal{L}(\mu, \hat{\hat{\boldsymbol{\theta}}}_\mu, \hat{\hat{\boldsymbol{\kappa}}}_\mu) < \mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\kappa}})$ [111], so that $q_\mu > 0$.

This simplified test statistic then follows the PDF of a central (non-central) χ^2 distribution if the B (S+B) hypothesis is true [112].

The properties of the PDF for the background only hypothesis $f(q_\mu|\mu = 0)$ are obtained from the Asimov pseudo data set[3]. This pseudo data set is built by taking the sum of expectation values of all components of the pre fit model (the expected event yield in each bin), suppressing the information of the uncertainty on the expectation values.

As stated above, the goal of the statistical inference is to determine if it can be claimed that the signal has been observed in data. In terms of a hypothesis test, the question to be answered is thus if the null hypothesis can be rejected, which is the hypothesis that the observed data can be obtained assuming the background only model.
For this reason, a p_{obs} value is defined that provides a measure of probability to obtain a data set at least as extreme in terms of the excess as the observed data, while assuming the background only hypothesis.
In order to have an estimate of the p value, and thus the significance, during the blinded stage of the analysis, an additional p value is defined that corresponds to an expected significance. This value is a measure of the probability of obtaining a data set that is at least as extreme as the Asimov pseudo data set of the S+B model, again given the background only hypothesis. This expected p value is also useful to illustrate potential differences between the observed data and the S+B model. The two p values are given by

$$p_{\text{exp}} = \int_{q_{\mu,\text{exp}}}^{\infty} f(q_\mu|\mu = 0)\, dq_\mu, \qquad p_{\text{obs}} = \int_{q_{\mu,\text{obs}}}^{\infty} f(q_\mu|\mu = 0)\, dq_\mu, \tag{7.4}$$

where $q_{\mu,\text{exp}}$ and $q_{\mu,\text{obs}}$ refer to $q_{\mu=0}$ obtained with the S+B pseudo data set and the measured data set, respectively.

The expected and observed significances Z that correspond to these two p values are given by a transformation with the quantile function. The quantile function is the inverse of the CDF of the normal distribution Φ, so that the significances are given by $Z = \Phi^{-1}(1 - p)$ and are thus quoted in units of standard deviations.

[3] This refers to the author Isaac Asimov, whose name is commonly used by members of the ATLAS collaboration to describe this particular type of pseudo data sets. The term was coined in [111] where it is derived from the electoral system pictured in the short story *Franchise*, written by Mr. Asimov, where the single most representative voter (expected voter) is used to replace all other votes (full distribution).

The results of the fit and the statistical inference are given in section 7.7. Following these results, a series of tests are performed to investigate potential origins of a moderate excess observed data, which are described in chapter 8.

7.2 Object and Event Selection

Following the discussion of the reconstruction procedure in chapter 5, the selected WPs, cuts and overall objects definitions for the search for four top quarks is given in the following.

There are two definitions for leptons used in this analysis called *loose* and *tight*. The *tight* definition is used in most cases, while the *loose* definition is only used in cases where it is useful to look at events enriched with incorrectly reconstructed leptons.

Electrons are required to have a $p_{\mathrm{T}} > 28\,\mathrm{GeV}$ and the lowest trigger thresholds are $p_{\mathrm{T}} = 24, 26, 17$ and $7\,\mathrm{GeV}$ for a single electron trigger in data taking years 2015-2017 and 2018, for a dielectron trigger and for an electron-muon trigger, respectively. This set of triggers is chosen (in combination with the p_{T} requirement) so that all selected electrons are on the trigger efficiency plateau while still being *unprescaled*, i.e. all triggered events are recorded.
The requirement on the pseudorapidity of the electron is set to be $|\eta| < 2.47$ & $|\eta| \notin [1.37, 1.52]$. This corresponds to the dimensions of the EM, where the region of $1.37 \leq |\eta| \leq 1.52$ corresponds to an area not instrumented by this sub detector.
The chosen identification WP is *tightLH* (*mediumLH*) for the tight (loose) lepton definitions and the isolation used is *FCTight* for tight electrons and no isolation requirement is set for loose electrons. Additionally, a cut on the output of the ECIDS discriminant is applied for electrons in the ee and $e\mu$ channels. This discriminant is the output of a BDT trained to further suppress electrons that have the wrong charge associated to them by the reconstruction process. Details about its implementation can be found in [76].

Muons are required to have a $p_{\mathrm{T}} > 28\,\mathrm{GeV}$ and the lowest trigger thresholds are $p_{\mathrm{T}} = 50, 8,$ and $24\,\mathrm{GeV}$ for a single muon trigger, for a dimuon trigger and for an electron-muon trigger, respectively. The choice of triggers follows the same principle as for electrons.
The requirement on the pseudorapidity of the muon is set to be $|\eta| < 2.5$, which is limited by the dimensions of the ID.

The muon identification WP is *medium* for both the loose and the *tight* lepton definitions and muons are not required to be isolated for the former, while the *FixedCutTightTrackOnly* isolation WP is used for the latter.

Jets are required to have a $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and to have a pseudorapidity of $|\eta| < 2.5$. As mentioned in section 5.4, the anti-k_t algorithm with a size $R = 0.4$ is used for the jet reconstruction based on topological clusters.
Two tools are applied to further purify the jet collection: JetCleaning and the Jet-Vertex-Tagger (JVT) [113]. The latter is only applied to jets with a $p_{\mathrm{T}} < 60\,\mathrm{GeV}$ and a pseudorapidity of $|\eta| < 2.4$ and it is targeting to reduce the amount of jets caused by additional (pile-up) collisions.
In order to identify jets that originate from a b-hadron, the MV2c10 algorithm is used at the WP which corresponds to a 77 % efficiency.

$E_{\mathrm{T}}^{\mathrm{miss}}$ is computed using the *loose* lepton definitions and the same jet collection as stated above.

Overlap Removal The overlap removal procedure applied in this analysis is the BoostedSlidingDRMu setup as outlined in section 5.7.

The different object selection criteria are summarised in Table 7.1.

Table 7.1: Overview of object definitions and selections used in this analysis.

Object	WP	p_{T} [GeV]	$\lvert\eta\rvert$	Identification	Isolation
Electrons	*loose*	$> 28\,\mathrm{GeV}$	< 1.37 or $\in (1.52,2.47)$	*mediumLH* ECIDS(ee,$e\mu$)	None
	tight	$> 28\,\mathrm{GeV}$	< 1.37 or $\in (1.52,2.47)$	*tightLH* ECIDS(ee,$e\mu$)	*FCTight*
Muons	*loose*	$> 28\,\mathrm{GeV}$	< 2.5	*medium*	None
	tight	$> 28\,\mathrm{GeV}$	< 2.5	*medium*	*FixedCutTightTrackOnly*
Jets		$> 25\,\mathrm{GeV}$	< 2.5	cleaning & JVT [113]	
b-tagged Jets		$> 25\,\mathrm{GeV}$	< 2.5		MV2c10 at 77 % eff.
$E_{\mathrm{T}}^{\mathrm{miss}}$	*loose* lepton definitions and anti-k_t jets with $R = 0.4$ from topological clusters				
Overlap Removal	BoostedSlidingDRMu (see section 5.7)				

7.2.1 Event Selection

As mentioned in the introduction to this chapter, a set of event selections have to be devised to define the SR and the CRs used in the global fit. As a first step, however, a preselection is defined that is the basis of all other selections and which is used in order to reduce the otherwise very large simulated data sets to a manageable level. Data sets following the preselection can then be used to study the impact of different additional selections and thus arrive at the definitions for the final selections. The preselection and signal selection are described in the following, while the selections used to define the CRs are given in subsection 7.4.2.

7.2.1.1 Preselection

The general strategy of the preselection follows the strategy used in a previous search for four top quarks using $\mathcal{L} = 36.1\,\text{fb}^{-1}$ of ATLAS data [43] and it is somewhat convoluted at first glance. This convoluted structure is kept, because it is mandated by one of the background estimation methods which was initially considered.
The preselection separates events into seven different categories depending on lepton multiplicity and flavour: $e^{\pm}e^{\pm}$, $e^{\pm}\mu^{\pm}$, $\mu^{\pm}\mu^{\pm}$, eee, $ee\mu$, $e\mu\mu$ and $\mu\mu\mu$. The first three categories are often denoted as (same sign) dilepton categories while the last four categories will be denoted as multilepton categories in the following.

For the event classification, only the first three p_{T}-leading *loose* leptons are considered. If out of these three *loose* leptons, there are exactly two leptons that also pass the *tight* criteria and if these two *tight* leptons have the same electric charge, they are put in the dilepton category that corresponds to the flavours of the leptons[4].
Electron pairs originating from J/ψ and Z boson decays can enter the $e^{\pm}e^{\pm}$ category if the charge of one of the electrons is mis-identified. As a means to reduce their contribution, a requirement is set on the invariant mass of the pair: $m_{ee} > 15\,\text{GeV}$ and $m_{ee} \notin [81\,\text{GeV}, 101\,\text{GeV}]$ for this category.

If out of the three p_{T}-leading *loose* leptons, all three also pass the *tight* requirements, the event will be selected into one of the multilepton categories according to the flavours of the leptons and no additional requirement on the charge of these leptons is set.
Similar to the dilepton case, also in the multilepton channels there are significant contributions by leptons originating from Z boson decays. Here, events are rejected if any

[4] The order (in p_{T}) of the flavour of the leptons does not play a role for this classification so that, for example, the category $e^{\pm}\mu^{\pm}$ would also contain events where the muon has a larger p_{T} than the electron. Simply put, the category names of $e^{\pm}\mu^{\pm}$ and $\mu^{\pm}e^{\pm}$ are interchangeable unless stated otherwise.

opposite sign same flavour lepton pair has an invariant mass close to $m_Z \approx 91\,\text{GeV}$, i.e. if any $m_{\ell^\pm\ell^\mp,\text{SF}} \in [81\,\text{GeV}, 101\,\text{GeV}]$.

If none of the two cases above apply, the event is discarded.

The fact that only the first three p_T-leading *loose* leptons are considered has some intricate, although marginal, consequences. First, it means that events with four (or more) *tight* leptons can be discarded or classed as either dileptonic or multileptonic, where the latter is by far the most common case. Thus, the fully leptonic decay mode of a four top quark event is implicitly covered. Second, an additional *tight* lepton at the 4th or lower position in p_T order will be missed, potentially migrating a trilepton event to a dilepton category or even discard it, or discard an otherwise valid dilepton event.
However, these migratory effects are very small and do not have a measurable impact on the outcome of the analysis. This is especially the case since the same selection criteria are applied on data and simulation.

In addition to the lepton selection above, events are also required to contain at least one b-tagged jet.

7.2.1.2 Signal Selection

The definition of the SR has been iteratively improved upon, where first iterations optimise the selection according to the FOM defined by $S/\sqrt{B}$, where S represents the signal yield and B the sum of background yields. The FOM used in later iterations is the expected significance as described in the introduction to this chapter.

In order to schematically illustrate the signal efficiency and background rejection of the selection, the distributions used for the SR definition are shown in Figure 7.1 for the signal, for two major physics backgrounds ($t\bar{t}\,W$ and $t\bar{t}Z$) and for one instance of an instrumental background (HF_μ)[5]. The figure shows the distributions of the number of jets N_j (Figure 7.1a), the number of b-tagged jets N_b (Figure 7.1b) and of $H_\text{T} = \sum_\ell(p_{\text{T},\ell}) + \sum_j(p_{\text{T},j})$ (Figure 7.1c) for each of the mentioned processes and individually normalised to unit area. The normalisation to unit area allows for a comparison of the shape of the distributions without being overburdened by the difference in overall yields between the processes. Nevertheless, this information is given in terms

[5] A more thorough description of these backgrounds and the used data sets will be given in section 7.4 and section 7.3, respectively.

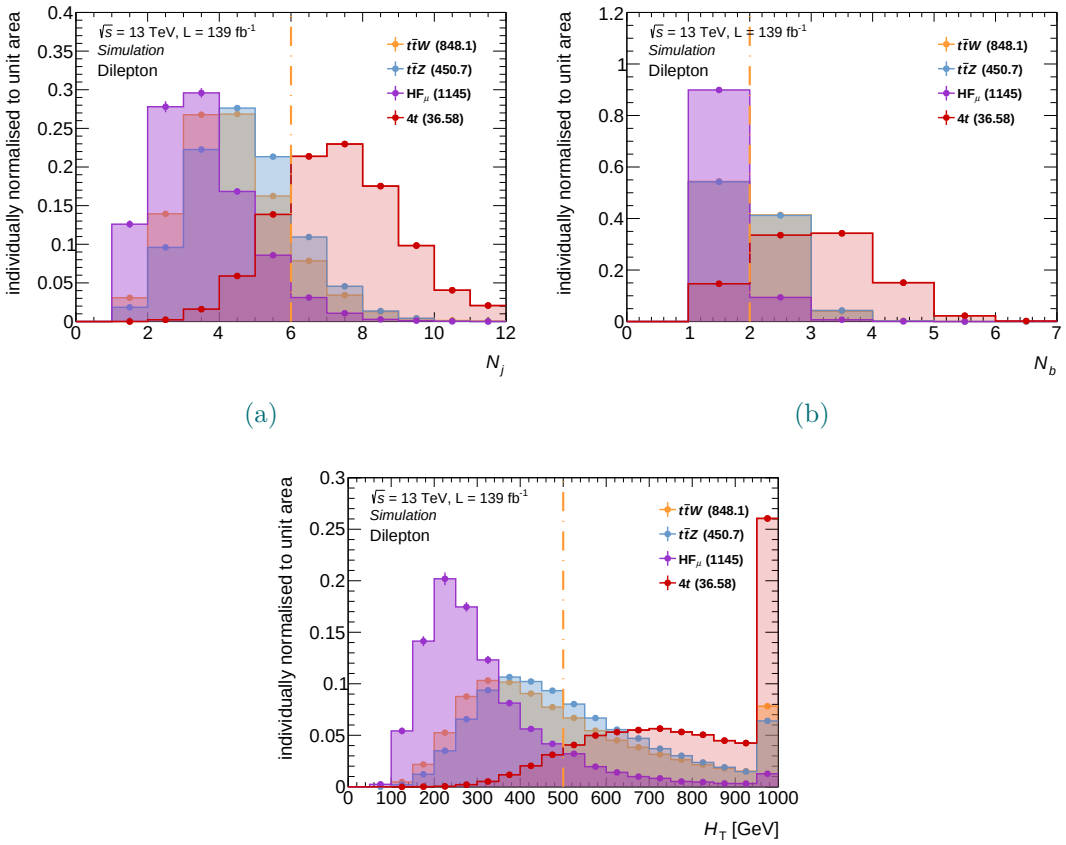
of numbers of expected events in the legend.

The distributions are shown only for the dilepton categories for illustration purposes. The uncertainty corresponding to the size of the simulated data set is displayed as vertical error bars, which is the case for histograms of simulated processes in the following unless specified otherwise.

The lowest and the highest bin do not have a lower and upper bound, respectively and in general, the binning scheme is to include the lower bin edge and exclude the upper one. As an example in Figure 7.1, the first (empty) bin covers the range of $[-\infty,1)$, the second bin covers the range of [1,2) and the last bin includes all entries in $[6,\infty)$. This scheme is used for all histograms unless explicitly specified otherwise.

The orange dashed lines represent the cut values used to define the SR. Therefore, the SR is defined as the part of the phase-space with at least six jets ($N_j \geq 6$), at least two b-tagged jets ($N_b \geq 2$) and $H_{\mathrm{T}} > 500\,\mathrm{GeV}$.

In the case of a search for a new particle or process, such as the search for four top quark, it is often desirable to blind the regions of the phase-space in data that are likely to contain the sought-after signal, seeing that the (non)existence of this signal in data is the main outcome of the statistical test. For this analysis, three selections are proposed, at least one of which has to be applied in order to blind the data. The selections are either ($N_j \leq 5 \& N_b \leq 2$), $H_{\mathrm{T}} \leq 500\,\mathrm{GeV}$ or $BDT < 0$, where BDT represents the discriminant output of the BDT used for signal extraction as described in section 7.5.



(a) (b)

(c)

Figure 7.1: Distributions of the number of jets N_j, the number of b-tagged jets N_b and of H_{T} for the signal and three different background sources in the dilepton channel after preselection. The area of each distribution is normalised to unit area and the numbers in the legend correspond to the total number of expected events of the corresponding process. The orange dashed lines represent the lower bounds on these distributions used to define the SR.

7.3 Data Sets

The basis of any quantitative study is the quality, quantity and type of data sets used. For this reason, the measured data and the methods used for simulating data are outlined below. The description closely follows the summary in [4].

7.3.1 Measured Data

The data set of measured data comprises measurements by the ATLAS detector between 2015 and 2018, i.e. the entire LHC Run-II period, that pass *good run* quality criteria, i.e. all subdetectors were fully operational and the particle beam was stable. The collision energy in this period was set to $\sqrt{s} = 13\,\text{TeV}$ and the final data set corresponds to an integrated luminosity of $\mathcal{L} = 139.0 \pm 2.4\,\text{fb}^{-1}$ measured using the LUCID-2 detector [62]. As indicated in chapter 6, the mean number of interactions per bunch crossing $\langle\mu\rangle$ is an important parameter for the reconstruction procedure. Therefore, the distribution of $\langle\mu\rangle$ is shown in terms of recorded luminosity in Figure 7.2, which is taken from [114] and where the different colours represent the distribution for different years. The parameter $\langle\mu\rangle$ written in the legend corresponds to the luminosity weighted mean of the distribution of $\langle\mu\rangle$ per year.

For the selection of measured data for this analysis, dilepton triggers are used as well as single lepton triggers and the leptons used for the event selection as described in the previous section are required to match at least one of the leptons that activated at least one of these triggers.

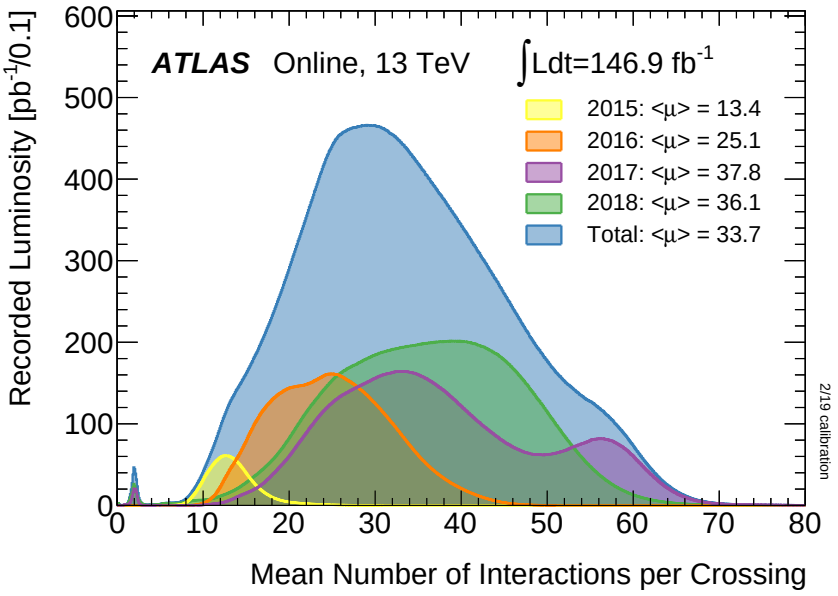


Figure 7.2: Distribution of $\langle\mu\rangle$ in terms of recorded luminosity, where the different colours represent the distribution for different years. The parameter $\langle\mu\rangle$ written in the legend corresponds to the luminosity weighted mean of the distribution of $\langle\mu\rangle$ per year. Taken from [114].

7.3.2 Simulation

The data sets (samples) comprising simulated collisions are produced in three different campaigns that correspond to different data taking conditions (most notably the difference in the $\langle\mu\rangle$ distributions above) in the years 2015-2016, 2017 and 2018 respectively. The campaigns are called MC16a, MC16d and MC16e, where the first two account for the measured distribution of $\langle\mu\rangle$ in the corresponding years and the latter is reweighted to match the expected $\langle\mu\rangle$ distribution of 2018 as they were produced before the data taking period in 2018 was finalised. The libraries, tools and settings used for the simulation of each of the considered processes is outlined in the following.

$t\bar{t}t\bar{t}$: Three different signal data sets (samples) are produced for this analysis. For the main sample, the production of $t\bar{t}t\bar{t}$ is modelled at NLO for the strong coupling constant α_s using MadGraph5_aMC@NLO v2.6.2 [32] and following the NNPDF3.1NLO [33] PDF.

The consequent decay of the top quarks is simulated at LO by using the MadSpin [115, 116] framework. Hadronisation and showering effects are taken into account using Pythia 8.230 [117] with the ATLAS specific set of tuned parameters called A14 [118] and the NNPDF23LO [33] PDF. In contrast, the decays of b and c hadrons are

simulated using EvtGen in version 1.6.0 [119]. The detector response is simulated using the AtlFast II (AFII) [120] framework.

Additionally, a signal sample is produced with the generation of $t\bar{t}t\bar{t}$ events at LO, while all other steps are the same as for the main sample. This data set is solely used for the training of the signal extraction BDT and is preferred over the sample at NLO, because it does not yield negative event weights that would interfere with the training of the BDT. This is possible since the difference in terms of the shape of the distributions between the LO and NLO expectation is small for all variables used in the training.

In order to test the impact of changes to the modelling of the hadronisation and showering of $t\bar{t}t\bar{t}$ events, the same set up as for the main sample is used, albeit exchanging Pythia 8.230 with Herwig 7.04 [121, 122] using the H7UE [122] set of parameters and the MMHT2014LO PDF [123] for this simulation step.

$t\bar{t}$: Events of the production of $t\bar{t}$ are modelled at NLO using Powheg-Box version 2 [105, 124–126] and using the NNPDF3.0NLO [33] PDF. One major parameter in $t\bar{t}$ production modelling is h_{damp} [127] which is a measure for the p_{T} of the first additional emission in NLO (w.r.t. LO) during the showering. This value is set to $h_{\text{damp}} = 1.5\,m_t$, where m_t represents the mass of the top quark. The showering and hadronisation is approached by using Pythia 8.230 with the A14 set of parameters and the NNPDF23LO PDF, like the main signal sample. The detector response is estimated using a full detector simulation implemented in the Geant 4 [128] framework.

The distribution of H_{T} for $t\bar{t}$ declines sharply for larger values of H_{T}, which is an important region for this analysis. Therefore and in order to avoid producing unrealistically high number of overall events, a set of nominal $t\bar{t}$ samples is produced using H_{T}-slicing. This slicing yields a similar number of simulated events for different ranges in H_{T}, resulting in a much better MC precision in higher H_{T} regions (w.r.t. the nominal sample) at the very low cost of having to reweight the simulated events to restore the correct H_{T} distribution.

t and tW: These processes are generated using Powheg-Box v2 at NLO in QCD using NNPDF3.0NLO for the PDF model, with the exception of the single top s-channel production, which utilises the NNPDF3.0NLOnf4 PDF. Showering and hadronisation are achieved using Pythia 8.230 with the A14 set of parameters and NNPDF23LO PDF.

$t\bar{t}Z$: The production of $t\bar{t}Z$ is modelled using MadGraph5_aMC@NLO 2.3.3 at NLO with the NNPDF3.0NLO PDF model followed by Pythia 8.210 with A14 parameter and NNPDF2.3LO PDF sets.

$t\bar{t}W$: The production of $t\bar{t}W$ is achieved using SHERPA 2.2.1 with the NNPDF3.0nlo PDF set. Additionally and because this process can become more signal-like if there are additional partons in the final state, a special interest is put on generating these events as well as possible. The calculation of the matrix element of the diagram with one additional particle is achieved at NLO, while the calculation for up to two additional partons is done at LO. The tools used are comix [129] and OpenLoops [130, 131], where the additionally generated events are merged with the nominal ones in the parton showering step using Sherpa [129].

$t\bar{t}H$: The production of $t\bar{t}H$ is modelled using PowhegBox at NLO and the NNPDF-3.0NLO PDF set, interfaced with Pythia 8.230 using the A14 set of parameters and NNPDF2.3LO PDF.

tWZ: The production of tWZ is modelled using MadGraph5_aMC@NLO 2.3.3 with the NNPDF3.0NLO PDF model and followed by Pythia 8.212 set to use the A14 set of parameters and NNPDF2.3LO PDF.

tZ, $t\bar{t}VV$, ttt: These processes are modelled using MadGraph5_aMC@NLO 2.3.3 but contrary to the name of the generator only at LO. The showering and hadronisation is achieved by using Pythia 8, the A14 set of parameters and the NNPDF2.3lo PDF. The V in $t\bar{t}VV$ is used to represent vector bosons and the possible diboson combinations used here are $VV = WW, ZZ, WZ, HH$, and WH.

V+jets: Samples containing the process of a vector boson (V = W,Z) with additional jets are generated using Sherpa 2.2.1 at NLO for up to two additional jets and Comix and OpenLoops at LO for up to four additional jets. As in the case of $t\bar{t}W$, the additional processes are joined with the nominal ones during the parton showering using Sherpa. The NNLO PDF set NNPDF3.0nnlo [33] is used so that the generated samples need to be normalised (using a k-factor) to the NNLO prediction of the cross-sections of the corresponding processes.

VV and VVV: The production of diboson and triboson ($VV(V)$, where V = W,Z) events is modelled using Sherpa 2.2.1-2 using the NNPDF3.0NNLO PDF model. Fully leptonic and semi leptonic final states, i.e. both vector bosons decay leptonically

or one of the bosons decays hadronically are simulated at NLO in QCD for diagrams with one additional parton and at LO for those including up to three additional partons.

The author of this document was one of the three people in the analysis team responsible for the creation and maintenance of the entire software framework and organisation used to produce samples that are directly usable by the analysis team. This uses data sets that are centrally produced in ATLAS and applies pre- and object selections to those data sets and enhances them with a series of additional variables from many different components of the ATLAS wide software project.

7.4 Background Estimation

As discussed in the introduction to this chapter, the different background sources, or rather the combined background to the search for four top quarks, need to be precisely estimated. Without a good understanding of the background model, the fit, and consequently the hypothesis test, becomes un-interpretable.

Approximately 60 % of background contributions to the SR enter that region because they include final states topologies that match the signal selection. This type of backgrounds is denominated *physics backgrounds* and the processes falling into this category are, $t\bar{t}V$ ($V = W,Z,H$) production, di- and triboson production, as well as rare $t\bar{t}VV$, ttt and single top production.
With the exception of $t\bar{t}\,W$, these backgrounds are estimated relying entirely on the MC prediction. For $t\bar{t}\,W$ however, the predicted normalisation has been shown to underestimate the observed one in the region of the phase space that is important for this analysis. This discrepancy has recently been shown in the search for $t\bar{t}\,H$ production in multilepton final states [132] triggering the investigation of spin correlation effects as well as EW corrections on the predicted normalisation of $t\bar{t}\,W$ [133], which is discussed in section 7.7. As a consequence, the normalisation of this process is determined in the fit to data where it is represented by one of the NFs.

Roughly 20 % of background contributions are coming from a series of different processes by passing the signal selection due to difficulties in the reconstruction (instrumental backgrounds)[6]. These can be further grouped into events with an opposite sign lepton pair where the charge of one of the leptons is mis reconstructed (QMR), and into fake/non prompt leptons. Fake leptons are leptons that are erroneously recon-

[6] The remaining contributions come from minor background processes of both physics and instrumental background type that are grouped together *Others*

structed from other objects, such as from quarks or gluons (LF). Non prompt leptons are leptons that do not directly emanate from the hard scattering process but rather from the decay of c and b hadrons (HF_e and HF_μ) or from photon conversions (γ^{*}[7] and Conv).

In the case of ATLAS, the mis reconstruction of the charge of a lepton is negligible for muons and it is mostly due to two processes for electrons[8]. First, the charge can be mis reconstructed for very p_T-hard electrons, because in this case the curvature of the track of the electron can become very small, so that the sign of the curvature can be mis reconstructed within the precision of the measurement. Second, a so called trident electron is reconstructed instead of the correct primary electron (see Figure 7.3).

A trident electron is created from the primary electron by emitting a Bremsstrahlung photon which in turn produces an electron–positron pair. The event is then considered to have an electron with a mis reconstructed charge if the charge of the trident electron is the opposite of the charge of the primary electron. In the trident case not only the charge is mis reconstructed, but likely also other properties. The likelihood for trident leptons increases for larger values of $|\eta|$ because the primary electron traverses more detector material than for lower values of $|\eta|$.

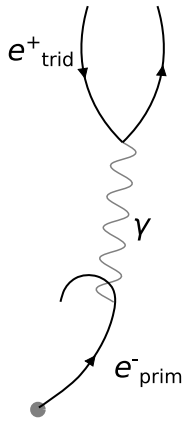


Figure 7.3: Schematic of QMR due to a trident electron.

The QMR background is estimated using a data driven technique that first estimates the rate of charge mis identification using regions enriched in Z boson decays by requiring two electrons with a combined invariant mass of $81\,\mathrm{GeV} < m_{ee} < 101\,\mathrm{GeV}$. The rate is devised in terms of η and p_T since it is highly correlated with these properties as described above. Subsequently, weights following the obtained rates are applied to electrons in a region identical to the SR, albeit with the requirement of an opposite sign lepton pair instead of the same sign one. QMR only plays a role in the $e^{\pm}e^{\pm}$ and $e^{\pm}\mu^{\pm}$ channels. The method is described in more detail in [1]. In order to cross-check this estimation, the weights are also applied to simulation and the results are compared

[7] This background source does not strictly represent a conversion. The difference and the reason why it is mentioned here nevertheless is described in the next section.

[8] As mentioned earlier, the term *electron* refers to positrons and electrons alike, unless a positron is specifically mentioned.

to an estimation obtained purely from the MC simulation.

The normalisation of non prompt and fake lepton contributions are mostly estimated during the fit by using dedicated CRs and NFs. This is not the case, however, for minor contributions which are estimated from simulation, which is listed more clearly in subsection 7.4.2. As a cross-check for the estimation using this approach, the matrix method is used, which is described in [43].

A detailed list of event level background classifications and their definitions is given in Table 7.2.

Table 7.2: Event-level background types and their properties.

Physics Backgrounds	
Process	Description / Origin
$t\bar{t}V$ (V = W,Z,H), ttt	events where all leptons are classed as *physics*
Instrumental Backgrounds originating from $t\bar{t}$+jets, V+jets, s/t-channel and tW single top processes	
Type	Description / Origin
QMR	one lepton charge mis identified
Conv	one lepton coming from photon conversion in the presence of detector material
$\gamma*$	one lepton originating from a virtual photon decay in the hard-scattering process
HF_e	one electron originating from a heavy-flavour hadron decay
HF_μ	one muon originating from a heavy-flavour hadron decay
LF	lepton originating from a light hadron decay or from the mis reconstruction of a quark or gluon
Mixed Backgrounds (Others)	
Process / Type	Description / Origin
other fakes	instrumental background events that do not fall into any of the above categories or where both leptons are of non-physics type
$t\bar{t}VV$, tX, VV, VVV, VH	events where all leptons are classed as *physics*

7.4.1 Instrumental Backgrounds

The studies conducted in order to arrive at a consistent background classification were strongly driven by work done within the frame of this thesis. In principle, this represents a non-trivial challenge with a fair amount of arbitrariness, which therefore needs to be studied to some detail.

In order to arrive at an event level classification as presented in Table 7.2, a lepton level classification needs to be derived first. Following this classification, studies can then be performed to group different lepton level classifications into event level classifications, which is presented in the following.

The classification of individual leptons largely follows the recommendation by the Isolation and Fakes Forum (IFF) of the ATLAS collaboration, making use of the information provided by the MCTruthClassifier tool [104]. This tool provides summarised information about the provenance of a reconstructed object by matching it to a *truth* object coming from the MC simulation. This is achieved by computing the angular distance ΔR between all reconstructed and truth objects and matching the pair with the minimal distance. The types defined for individual leptons very closely follow the event level classification mentioned above.

Physics prompt leptons, as well as trident electrons, with the correct reconstructed charge. In the case of the trident electrons, this means that the charge of the reconstructed trident electron has to be the same as the charge of the primary electron.

***Q*MR** prompt electrons with the wrong associated charge due to a mis interpretation of the curvature of the track, which is mostly the case for high-p_{T} electrons, or trident electrons with the opposite charge with respect to the primary electron.

Conv electrons originating from electron pair production in the presence of detector material.

$\gamma*$ leptons originating from a virtual photon during the hard scattering process. Studies on this background type are relatively recent in ATLAS and a more thorough description can be found in [132]. The difficulty of this type for the MCTruthClassifier lies in the fact that the simulation step of this background differs according to the p_{T} of the virtual photon. Its decay is included in the calculation of the matrix element if the photon is p_{T}-hard, while it is simulated during the parton showering step if the photon is p_{T}-soft. In the former case, the leptons are classed as prompt,

while in the latter they are denominated non-prompt by the MCTruthClassifier. From a reconstruction point of view, this background source resembles the material conversion, which is why it was initially grouped together.

HF$_{e,\mu}$ leptons originating from the decays of c and b hadrons.

Seeing that all events considered for this analysis have at least two leptons with potentially different types, an event-level categorisation using the lepton level information needs to be defined. This is generally non-trivial as in principle 28 different lepton category combinations need to be considered.

Figure 7.4 shows the expected number of events for all possible combinations of lepton classifications in the $e^{\pm}e^{\pm}$, $e^{\pm}\mu^{\pm}$ and $\mu^{\pm}\mu^{\pm}$ channels after pre selection and given a luminosity of $\mathcal{L} = 125\,\text{fb}^{-1}$. These results were obtained before finalising the data taking period so that a reduced luminosity is considered. Additionally, the simulation of some of the, mostly minor, background processes were not yet available at this point. This is due to the fact that these studies were an intermediate step used to define the event level background classifications and they were not repeated after finalising the classification. The different colours correspond to different physics processes from which the events originate.
It is important to note that the colour code is deliberately different from the rest of this chapter to stress that these processes do not directly correspond to the processes used in the final analysis. They do not correspond as they are presented inclusively here, while they are separated according to the event level classification otherwise. Additionally, the contributions from QMR are fully taken from simulation for this study.

The *None* category represents the case where the MCTruthClassifier is not able to determine a lepton category, which is mostly due to the inability to match the reconstructed object to a truth one.

The figure indicates that events where both leptons pertain to any lepton category other than *physics* are very rare, i.e. less than 5 events for the combination with the largest number of events not counting *None*, namely (QMR, Conv) in $e^{\pm}e^{\pm}$. This conclusion holds up also for more refined selections, like the signal selection, and for the trilepton case which are not shown here.

With this result in mind, an event level categorisation becomes straightforward, defining an event as *physics* if both leptons are of *physics* type or otherwise according to

the lepton type of the lepton that is not of *physics* type. If both leptons have a type other than *physics*, the event is classed as *other fakes*[9]. This categorisation is the one already given in Table 7.2.

Finally, Figure 7.4 shows that the contributions to the instrumental background categories come almost exclusively from $t\bar{t}$+jets, V+jets and single top processes, so that in the following only contributions from these processes are considered for the instrumental background.

This background categorisation is the basis for the background estimation of this analysis, which relies heavily on the simulation. This is particularly the case for the template fit method as will be described in the next section.

[9] This term is not very precise at it is not clear that the events in that category are indeed fakes, but it is kept for consistency with [4].

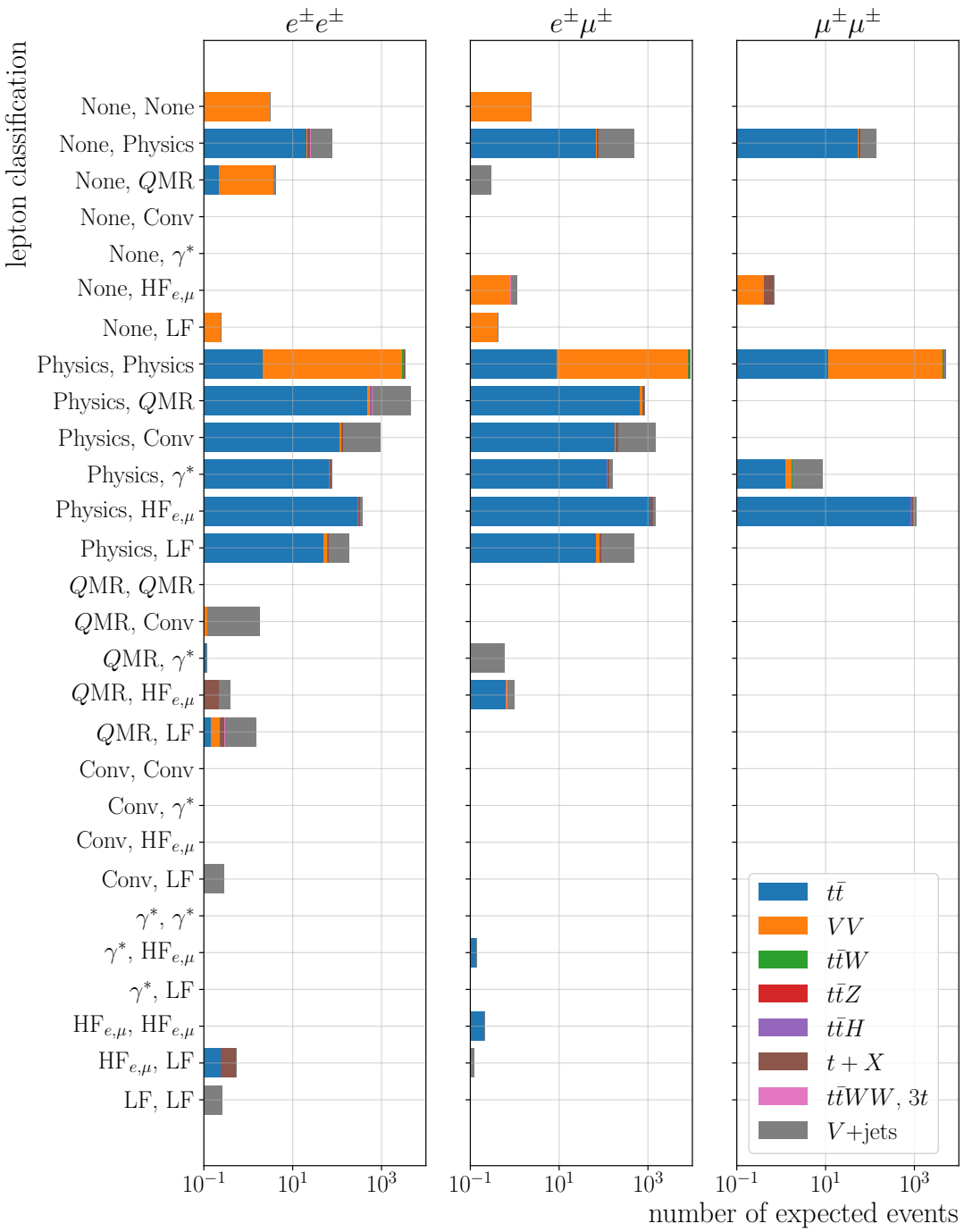


Figure 7.4: Number of expected events for all combinations of lepton classifications for the dilepton channels after preselection at $\mathcal{L} = 125\,\mathrm{fb}^{-1}$. These results were obtained with an older setup and not with the complete list of processes that are used in the final analysis as they were an intermediate step used to define event level background classifications. The different colours correspond to different physics processes from which the events originate and the colour code is deliberately different from the rest of this chapter.

With the event level background categorisation in place, the contributions of each background process to the overall background in the SR can be investigated. This is shown in Figure 7.5 for the SR at pre fit level. The size of each coloured segment is proportional to the contribution of the corresponding background and the inner circle corresponds to the outer circle by grouping the different background types by their major category.

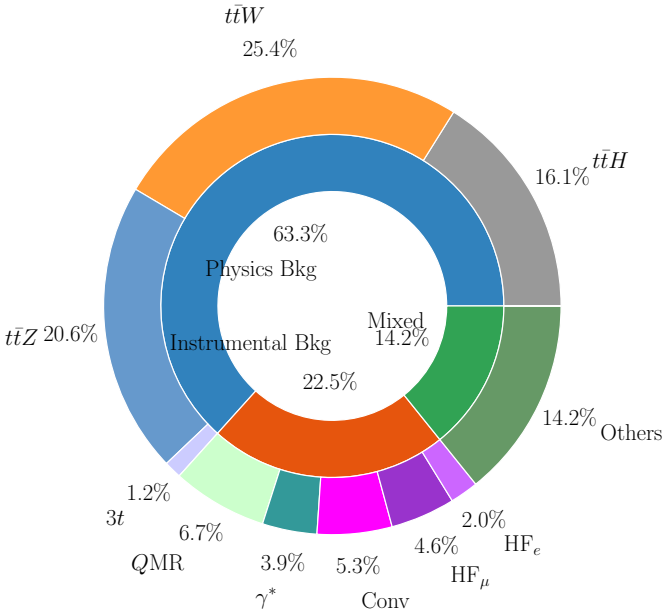


Figure 7.5: Pre fit background composition of the SR. The outer circle represents the proportions of the individual background types, while the inner circle represents the proportions for the major background categories that correspond to the background types on the outer ring.

7.4.2 Template Fit Method

For this analysis, the data-assisted template fit method is used to estimate the normalisation of the background contributions of $t\bar{t}\,W$, HF_e, HF_μ, Conv and γ^*. The instrumental background types are estimated in this way as the correct normalisation of these background types is difficult to obtain in simulation. The normalisation of $t\bar{t}\,W$ is estimated using this technique because of the discrepancy between the observation and prediction in [132] stated above.

The version of the template fit method applied here consists of designing CRs and discriminant variables for each of the backgrounds individually (ideal case) or regions

where the distributions of different background processes are mostly uncorrelated. In this case, distribution refers to the frequentist distribution as obtained from the simulation, i.e. the histogram[10] normalised to unity of background i in terms of variable j. These distributions are often referred to as templates, which gives the method its name. As a consequence, this method is only data assisted (and not data driven) as the normalisation is obtained from data, while fully relying on the simulation for the determination of the templates.

The normalisation to these templates is determined during the likelihood fit to data, as described in section 7.1, where the full background (data-driven for QMR, simulation for all other sources) plus signal model is used as a starting point. The normalisations of the above mentioned backgrounds are free parameters of the fit, called NFs, while the uncertainties on the shape of the templates is treated by additional NPs to the fit.

The definitions of the different CRs are given in the following.

CR Conv./γ^* is designed for the estimation of the Conversion ($\mathrm{NF_{Conv}}$) and γ^* (NF_{γ^*}) backgrounds. Only $e^\pm e^\pm$ and $e^\pm\mu^\pm$ events are considered, where at least one of the electrons has an additional track t_c close by that can be associated to a Conversion Vertex (CV) and where the invariant mass of this track and the track associated to the electron t_e is $m^{\mathrm{CV}}_{(t_c,t_e)} \in (0\,\mathrm{GeV}, 0.1\,\mathrm{GeV})$. Additionally, $4 \leq N_j < 6$, at least one b-tagged jet and $H_\mathrm{T} \in (200\,\mathrm{GeV}, 500\,\mathrm{GeV})$ are required. The discriminant variable is the invariant mass of the two tracks evaluated at the PV $m^{\mathrm{PV}}_{(t_c,t_e)}$ and they are described in more detail in [132].

CR HF$_\mu$ is designed for the estimation of the HF_μ background normalisation ($\mathrm{NF_{HF\mu}}$). Only $\mu\mu\mu$ and $e\mu\mu$ events are considered that have exactly one b-tagged jet and an $H_\mathrm{T} \in (100\,\mathrm{GeV}, 250\,\mathrm{GeV})$. No template is used in this region and instead the (expected) overall event yield is used to determine $\mathrm{NF_{HF\mu}}$.

CR HF$_e$ is designed for the estimation of the HF_e background normalisation ($\mathrm{NF_{HF}}e$). Only eee and $ee\mu$ events are considered that have exactly one b-tagged jet and an $H_\mathrm{T} \in (100\,\mathrm{GeV}, 250\,\mathrm{GeV})$. As for CR HF_μ , only the (expected) event yield is used.

CR $t\bar{t}W$ is designed for the estimation of the $t\bar{t}W$ background normalisation ($\mathrm{NF}_{t\bar{t}W}$). Only $e^\pm\mu^\pm$ and $\mu^\pm\mu^\pm$ events are considered that have at least four jets

[10] The term 'histogram' slightly differs from its use in the field of statistics where it is normalised to unity by definition. In particle physics, however, it most commonly shows the number of (expected) events per bin.

among which at least two are b-tagged and where the pseudo-rapidity of all electrons is required to be $|\eta| < 1.5$. Additionally, the inverse of the selection of the CR Conv./γ^* region is required to ensure orthogonality. A separation from the SR is achieved by requiring either $H_{\mathrm{T}} < 500\,\mathrm{GeV}$ or $N_j < 6$ in the case where $N_b = 2$ and by requiring $H_{\mathrm{T}} < 500\,\mathrm{GeV}$ if $N_b \geq 3$. The discriminant variable is the sum of transverse momenta of all leptons $\sum p_{\mathrm{T},\ell}$.

The different regions, their definitions and discriminant variables are listed in a more concise form in Table 7.3.

Table 7.3: Summary of the definitions of the regions used in the fit. In the case of CRs, the NF they are designed to constrain and estimate is also given.

Region	NF	Discr.	Channel	N_j	N_b	add. selection
SR		BDT	SSML	≥ 6	≥ 2	$H_{\mathrm{T}} > 500\,\mathrm{GeV}$
CR Conv./γ^*	$\mathrm{NF}_{\mathrm{Conv}}$, NF_{γ^*}	$m^{\mathrm{PV}}_{(t_c,t_e)}$	$e^{\pm}e^{\pm}$, $e^{\pm}\mu^{\pm}$	$\in [4,6)$	≥ 1	$H_{\mathrm{T}} \in (200\,\mathrm{GeV}, 500\,\mathrm{GeV})$, $m^{\mathrm{CV}}_{(t_c,t_e)} \in (0\,\mathrm{GeV}, 0.1\,\mathrm{GeV})$
CR HF_e	$\mathrm{NF}_{\mathrm{HF}e}$	yield	eee, $ee\mu$		$= 1$	$H_{\mathrm{T}} \in (100\,\mathrm{GeV}, 250\,\mathrm{GeV})$
CR HF_μ	$\mathrm{NF}_{\mathrm{HF}\mu}$	yield	$\mu\mu\mu$, $e\mu\mu$		$= 1$	$H_{\mathrm{T}} \in (100\,\mathrm{GeV}, 250\,\mathrm{GeV})$
CR $t\bar{t}\,W$	$\mathrm{NF}_{t\bar{t}W}$	$\sum p_{\mathrm{T},\ell}$	$e^{\pm}\mu^{\pm}$, $\mu^{\pm}\mu^{\pm}$	≥ 4	≥ 2	$\notin$ CR Conv./γ^*, $\lvert\eta_e\rvert < 1.5$, $H_{\mathrm{T}} < 500\,\mathrm{GeV}$ if $N_b \geq 3$, ($H_{\mathrm{T}} < 500\,\mathrm{GeV}$ || $N_j < 6$) if $N_b = 2$

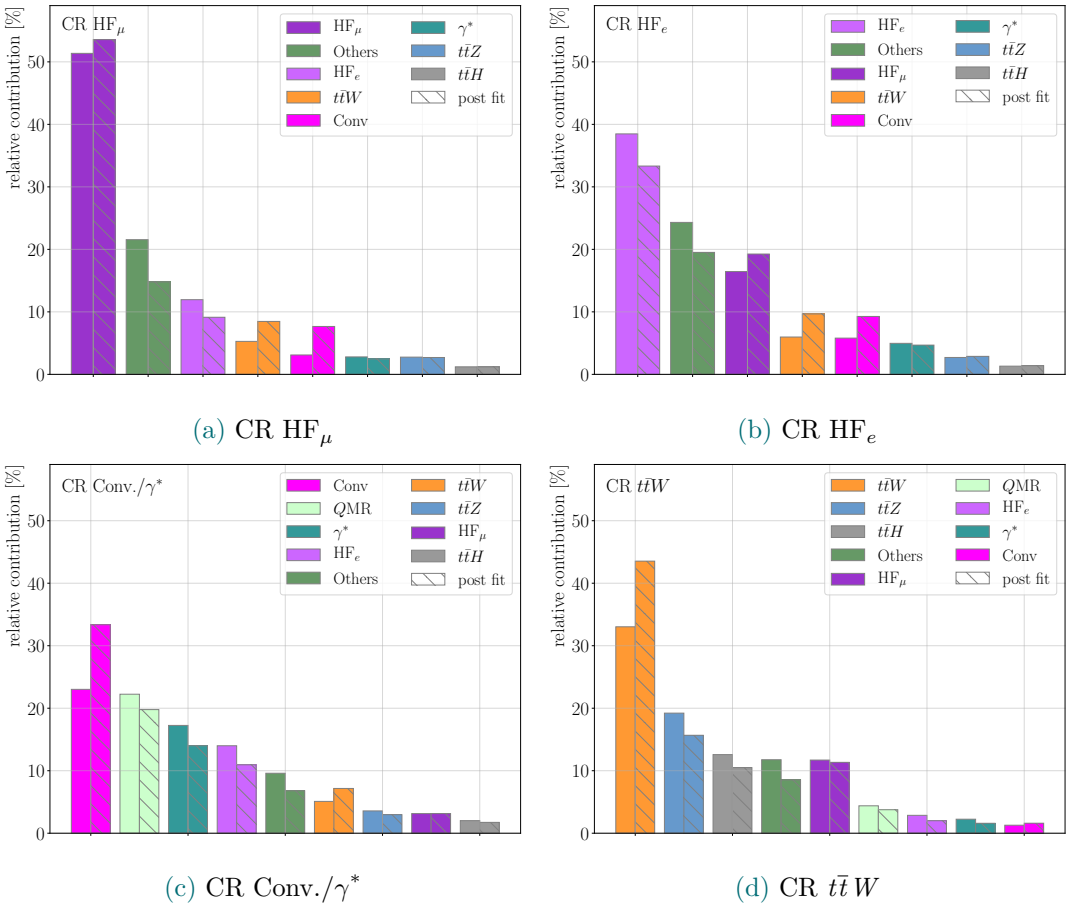
The success of the template fit method depends on two properties of the designed CRs. First, the purity of each CR in terms of the background types the region is designed for. This purity is simply given by the fraction of the corresponding background to the overall yield in that region. Second, the template of the background in question should be as uncorrelated as possible to the templates of all other processes in the CR.

For this reason, the compositions, and thus the purity, of each region is given in Figure 7.6 at pre and post fit level[11] (clear and hatched bars, respectively) for all contributions that are larger than 1 %. The largest contributor to each region is indeed the background type, the region is designed to estimate, with the small exception of γ^* in the CR Conv./γ^* region. However, this is not a problem for the template fit method for this process, as indeed the shape of the template differs strongly between γ^* and all other processes.

[11] The post fit level is shown here for the observant reader, but is introduced fully only in section 7.7.

This is shown in Figure 7.7a, where the templates of background processes contributing more than 5 % to the CR Conv./γ^* region are shown. While γ^* is mostly contained in the first bin, the other background estimated in that region, Conversions, is mostly contained in the second and third bin. Moreover all other processes tend towards the last bin of the distribution, so that the condition that the templates should be uncorrelated is given.
Figure 7.7b shows the templates of the CR $t\bar{t}\,W$ region. In this case, the template does not significantly differ from the templates of other processes such as $t\bar{t}\,Z$ and $t\bar{t}\,H$, which slightly minimises the ability of the template fit method to determine the correct normalisation for $t\bar{t}\,W$. However, this region is entirely dominated by contributions from $t\bar{t}\,W$, so that the method is still valid in this region.

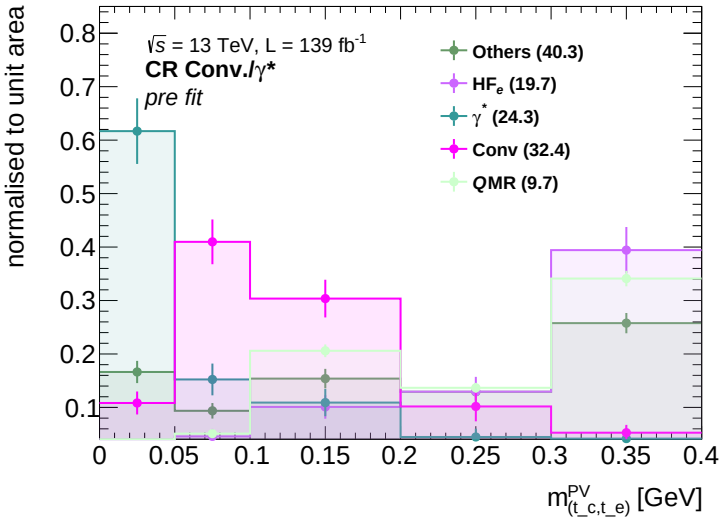
Seeing that the NFs that are part of the template fit method are determined simultaneously to the PoI, the results of the template fit are presented and discussed alongside the result on the PoI in section 7.7.



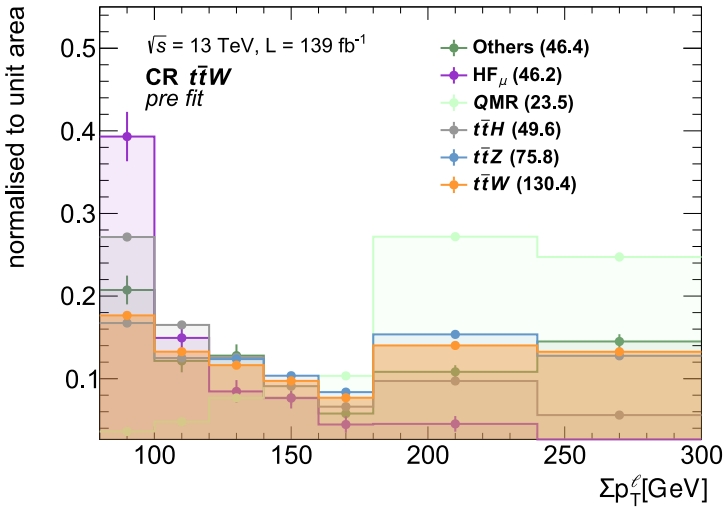
(a) CR HF$_\mu$ (b) CR HF$_e$

(c) CR Conv./γ^* (d) CR $t\bar{t}\,W$

Figure 7.6: Relative pre fit (clear) and post fit (hatched) contributions to the CRs used in the template fit method for all processes with a relative pre fit contribution larger than 1 %.



(a) CR Conv./γ^*



(b) CR $t\bar{t}\,W$

Figure 7.7: Templates obtained from MC simulation in the CR Conv./γ^* and CR $t\bar{t}\,W$ regions. Only templates for background processes that constitute more than 5 % of the overall background yield of the corresponding region are shown to simplify the visualisation. The CR Conv./γ^* region is designed to cover both, the γ^* as well as the Conv backgrounds, while the CR $t\bar{t}\,W$ region is exclusively designed for $t\bar{t}\,W$.

7.4.3 Validation of the Estimation

Because the fit has a large number of degrees of freedom to find a post fit model, it is important to study the possibility of extrapolating the results of the fit to regions and distributions of the phase space that were not (fully) available to the fit during the minimisation.
This is particularly important as the ability for extrapolation is the most important assumption for the use of the template fit method. Studying these properties is therefore a direct probe of the validity of the background estimation. As a consequence, the regions of the phase space that are used for this study, and that are (next-to) orthogonal to all regions used in the fit, are called Validation Regions (VRs). For this analysis, two such VRs are defined to study the quality of the estimation of the two largest background processes of the SR, namely $t\bar{t}Z$ and $t\bar{t}W$[12].

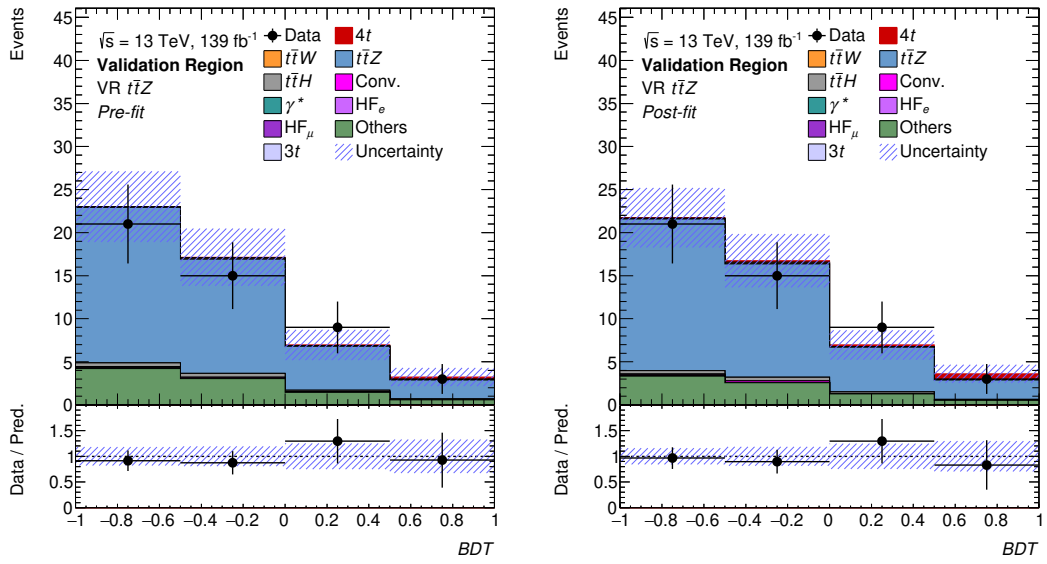
VR $t\bar{t}Z$ requires three leptons where at least one same flavour opposite sign lepton pair has an invariant mass of $m_{\ell^{\pm}\ell^{\mp},\mathrm{SF}} \in (81\,\mathrm{GeV}, 101\,\mathrm{GeV})$. Additionally, at least six jets are required among which at least two are b-tagged. Finally H_{T} is set to be larger than $H_{\mathrm{T}} > 500\,\mathrm{GeV}$. This region is similar to the trilepton contribution to the SR with the notable difference that the invariant mass requirement is inverted. It is thus fully orthogonal to the signal and control regions. The pre and post fit distributions of the BDT output in terms of (expected) events in this region is shown in Figure 7.8. These figures show a reasonable agreement between data and simulation resulting in only a very small difference between the pre and the post fit distribution. This is to be expected since $t\bar{t}Z$ enters the fit only through normalisation and shape NPs, which in turn are not strongly pulled (i.e changed, cf. Figure B.3) by the fit due to the agreement already present before the fit.

VR $t\bar{t}W$ exploits the fact that $t\bar{t}W$ exhibits a notable charge asymmetry for events that pass the preselection of this analysis. For pp collisions, $t\bar{t}W$ events with two positively charged leptons are more likely to occur than those with two negatively charged leptons. The region requires at least four jets among which at least two have to be b-tagged and the output of the BDT is used as discriminant variable. The charge asymmetry is given by the difference $N_p - N_m$ for each bin of the distribution, where N_p is the (expected) event yield for all events with $\ell^+\ell^+$, $\ell^+\ell^+\ell^+$ and $\ell^+\ell^+\ell^-$ final

[12] Extensive tests of the extrapolation of the post fit model obtained solely in the CRs to the prediction in the SR are described in detail in section 8.3.

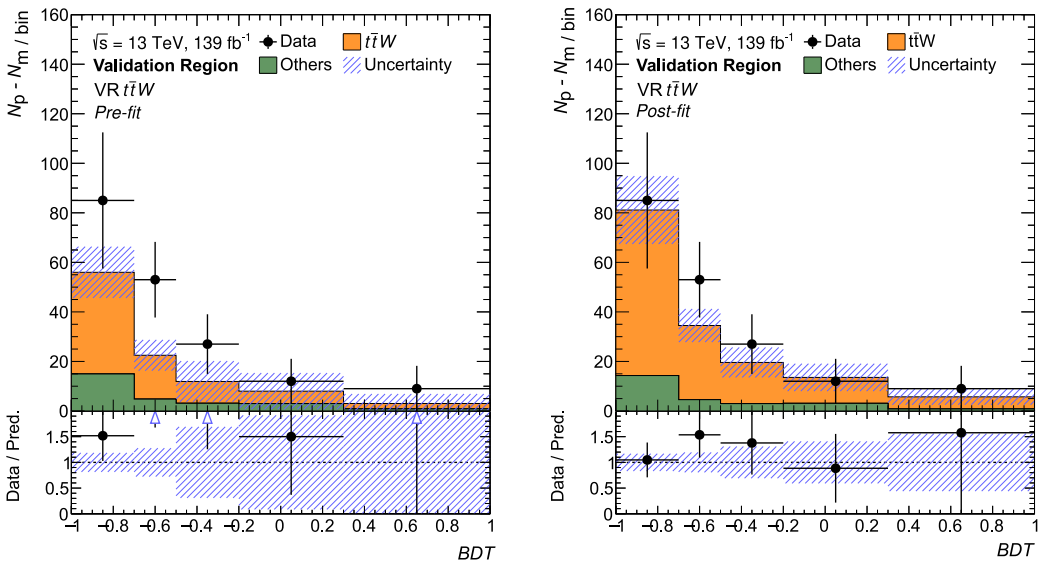
states and N_m represents the (expected) event yield for $\ell^-\ell^-$, $\ell^-\ell^-\ell^-$ and $\ell^-\ell^-\ell^+$ final states.

This region is not strictly orthogonal to neither the SR nor the CRs, but signal and non $t\bar{t}W$ contributions are expected to be negligible due to the charge symmetry of these processes. The pre and post fit distributions of the $t\bar{t}W$ VR are shown in Figure 7.9, where an important improvement in the agreement between the prediction and data can be seen as a result of the fit.



(a) pre fit (b) post fit

Figure 7.8: Pre and post fit expected and observed event yields for the output of the signal extraction BDT in the VR $t\bar{t}Z$ validation region. The uncertainties include statistical as well as systematic uncertainties.



(a) pre fit (b) post fit

Figure 7.9: Difference between the expected and observed yields for events with a positive sum of lepton charges N_p to those with a negative sum N_m in terms of the the output of the signal extraction BDT in the VR $t\bar{t}\,W$ validation region. The results are given at pre and post fit level and the uncertainties include statistical as well as systematic uncertainties.

7.5 Signal Extraction

With the background thoroughly studied, a reliable method to single out the signal (signal extraction) can be devised. The approach that is used as well as its performance is described in the following.

The main algorithm used to separate signal events from background events in this analysis is a Boosted Decision Tree (BDT) as implemented in the TMVA framework, which in turn relies on ROOT.
A BDT is an ensemble classifier consisting of N_{T} decision trees iteratively built starting from a single tree. In each iteration, a weight is assigned to each data point which corresponds to the frequency and quality of mis classification in the previous iteration, where the worse the mis classification the larger the weight. Therefore, the algorithm *boosts* the importance of mis classified events. For this analysis, the bagged gradient boost algorithm [134] is used.
Both individual decision trees and the boosting procedure have a set of tunable *hyper parameters* which can be optimised according to one or multiple figures of merit. Here, this is the case for the maximum depth of each individual tree $\mathrm{max_D}$, the number of

grid points used to determine the optimal split in each node N_{C} and the minimum fraction of events required for training a node $\mathrm{min_f}$ for the decision tree parameters. The boosting parameters include the learning rate of the gradient descent *shrinkage*, the fraction of events used in the boosting *bagging* and the final number of trees of the ensemble (forest) N_{T}.

7.5.1 Multivariate Analysis Setup

As customary for most supervised Multivariate Analysis (MVA) techniques, the full building and application chain of the BDT can be divided into three parts: Training, validation and application.

Training: The BDT in this case is a *yes-no* or rather a *signal-background* classifier. As a consequence, the background model used for the training is simply the combination of all background predictions directly from the MC simulation. More specifically, no data-driven or data-assisted correction to the MC prediction is applied. The signal model used in the training is the MC prediction at LO, to avoid negative weights which cannot be handled by this implementation of a BDT, while the nominal (NLO) signal sample is used for testing and validation.
A potential over specialisation of a classifier on the given training data set leads to a significant loss of ability for the classifier to generalise on other data sets. This fact especially poses a problem when this over specialisation (over training) confuses statistical fluctuations that are particular to the training data set with real correlations. The level of over training is probed using a twofold cross validation technique in this analysis. This technique requires that the training samples be split in two parts, where the classifier trained on one part is tested on the other part and vice versa. Often, the final classifier is then either a smart combination of the two trained instances of the classifier or the final classifier is trained on the full training data set. The latter poses the threat of loosing a potential minimisation of over training effects achieved by splitting the data set. For this analysis, the two classifiers, each trained on one half of the training data set, are kept and applied individually, which is described later on. The split is achieved by separating events with an even event number (event index) from those with an odd event number.

The training is done on the full LO signal sample and 80 % of the background samples. The number of signal and background events that enter the training are reweighted so that there are as many background events as there are signal events, while retaining

the proportions of the backgrounds taken from the simulation. The cross validation (testing) step is done on the odd 40 % of the background sample and on the odd 40 % of the NLO signal sample for the classifier trained on the same fraction of even events and vice versa. This is a slightly out-of-the-norm setup, where the cross validation is indeed one for the background, but is rather an out-of-sample validation for the signal, i.e. the training and testing samples are entirely distinct. This strategy allows not only to test for an over training of the classifier *per se*, but also for an over training of the classifier on peculiarities of the LO signal sample with respect to the NLO one.

Validation: The validation of both classifiers (odd and even) is done in data sets coming from the remaining 20 % of background and signal at NLO events. The bootstrap method is used to derive an uncertainty on the FOM, where 50 test samples are randomly selected from the original validation sample, so that the FOM can be calculated 50 times of which the mean and standard deviation can be computed.

Application: For the final analysis, the two classifiers are applied to all of the samples and to data. This is done in a simple way, where the classifier trained on odd events is applied to even ones and vice versa. One caveat of a *yes-or-no* BDT becomes evident here, where classifiers that are trained on the sum of background events are now applied on each background process individually. It is clear that this should be sub optimal as it cannot take advantage of peculiarities of any specific background.

7.5.2 Optimisations and Results

The final BDT setup used for this analysis has been thoroughly optimised in terms of input variables and hyper parameters. The optimisation uses the Area under the Curve (AUC) of the Receiver operating Characteristic (ROC), i.e. the integral of the curve given by the true positive rate over the false positive rate of the classifier. In order to have just a single FOM, the arithmetic mean of the AUCs obtained for the odd and even classifiers in the test sample is used.
The optimisation of hyper parameters is then achieved by scanning a grid of five variations for each of the six hyper parameters, so that $6^5 = 15625$ setups are probed. The optimal set of hyper parameters leading to a combined ROC estimated using the validation data set of $ROC = 0.853 \pm 0.006$ is given in Table 7.4.

The list of input variables to the BDT is optimised from an initial set of 21 variables. The optimisation procedure used is called Iterative Removal (IR). For this procedure,

Table 7.4: Optimal set of hyperparameters for the signal extraction BDT.

N_{T}	$\max_{\mathrm{D}}$	$\min_f$	*shrinkage*	N_{C}	*bagging*
800	6	3	0.01	15	0.7

a BDT classifier is trained using the previously optimised hyper parameters and the full list of variables bar one. This is iteratively repeated until every variable has been left out once and the impact of removing the variable in question on the FOM is evaluated. For this analysis, only variables are removed that leave no measurable impact on the FOM. Additionally, two variables are removed that show a disagreement between data and the simulation. This is done as the training relies on the information from the simulation so that a variable that is not well modelled would need a dedicated calibration procedure. The final set of input variables that are used amounts to 12 variables. The full list of initial input variables, an attempt to categorise them and the outcome of the IR is given in Table 7.5. The variables are listed in order of their rank as determined by the ranking score r defined as $r = \sum_{i=0} kS_i^2 N_i$, where k is the number of nodes where the algorithm chooses the variable in question for the split, S_i represents the separation gain and N_i the number of events in node i.

The final AUC achieved after input variable and hyper parameter optimisation amounts to AUC $= 0.8466 \pm 0.0021$ on the test data set, and to AUC $= 0.8526 \pm 0.0063$ on the validation data set, indicating a reasonable stability of the performance of the BDT with respect to differing data sets. The full distribution of the output of the BDT in the SR is shown in Figure 7.10. There, the background is given by the estimation using plain MC, except for the QMR background, which is derived by the data-driven method. The figure shows the difference in the shape of the distributions between the signal (larger for higher values of the BDT output) and different backgrounds (larger for lower values of the BDT output, with the notable exception of *ttt*). The uncertainties include the statistical as well as the systematic uncertainties which are introduced in section 7.6.

Table 7.5: List of initial input variables used for the signal extraction BDT ranked by importance determined by the algorithm during its training. The outcome of the IR procedure is given, where a variable is either (K)ept or (R)emoved. The variables marked with an (M) are removed due to an observed mis modelling.

Rank	Variable	Category	Description	IR
1	$\sum w_{\mathrm{MV2c10}}$	b-tagging	Sum over all jets of the index of the MV2c10 pseudo-continuous b-tagging score interval	K
2	$p_{\mathrm{T}}^{\ell_0}$	Lepton	Transverse momentum of p_{T} leading lepton	K
3	$E_{\mathrm{T}}^{\mathrm{miss}}$	Energy	Missing transverse momentum	K
4	$\Delta R(\ell,\ell)_{\mathrm{min}}$	Distance	Minimum distance between any lepton pair	K
5	$p_{\mathrm{T}}^{j_5}$	Jet	Transverse momentum of 6th p_{T} leading jet	K
6	$\Delta R(\ell,b)_{\mathrm{max}}$	Distance	Maximum distance between leptons and b-tagged jets	K
7	$H_{\mathrm{T}}^{\text{no lead jet}}$	Energy	Scalar sum of the p_{T} of all leptons and jets except leading jet	K
8	$\sum \Delta R(\ell,\ell)_{\mathrm{min}}$	Distance	Sum of the distance between all *tight* leptons	K
9	m_j/p_{T}^{j}	Event	Mass of the jet divided by its p_{T}	RM
10	$\Delta\phi(\ell_0,j_0)$	Distance	Difference in azimuthal angle between p_{T} leading lepton and jet	R
11	$p_{\mathrm{T}}^{j_0}$	Jet	Transverse momentum of p_{T} leading jet	K
12	$\Delta R(j,b)_{\mathrm{min}}$	Distance	Minimum distance between any b-tagged jet and to any non-tagged jet	K
13	$\Delta R(\ell,j)_{\mathrm{min}}$	Distance	Minimum distance between any lepton and any jet	R
14	$p_{\mathrm{T}}^{b_0}$	Jet	Transverse momentum of leading b-tagged jet	K
15	$\Delta R(\ell,b)_{\mathrm{min}}$	Distance	Minimum distance between any lepton and any b-tagged jet	R
16	$p_{\mathrm{T}}^{\ell_1}$	Lepton	Transverse momentum of the p_{T} subleading lepton	R
17	$p_{\mathrm{T}}^{j_2}$	Jet	Transverse momentum of the p_{T} third-leading jet	R
18	$p_{\mathrm{T}}^{j_1}$	Jet	Transverse momentum of the p_{T} subleading jet	K
19	N_j	Jet	Number of jets	RM
20	N_ℓ	Lepton	Number of leptons	R
21	$p_{\mathrm{T}}^{\ell_2}$	Lepton	Transverse momentum of p_{T} third-leading lepton	R

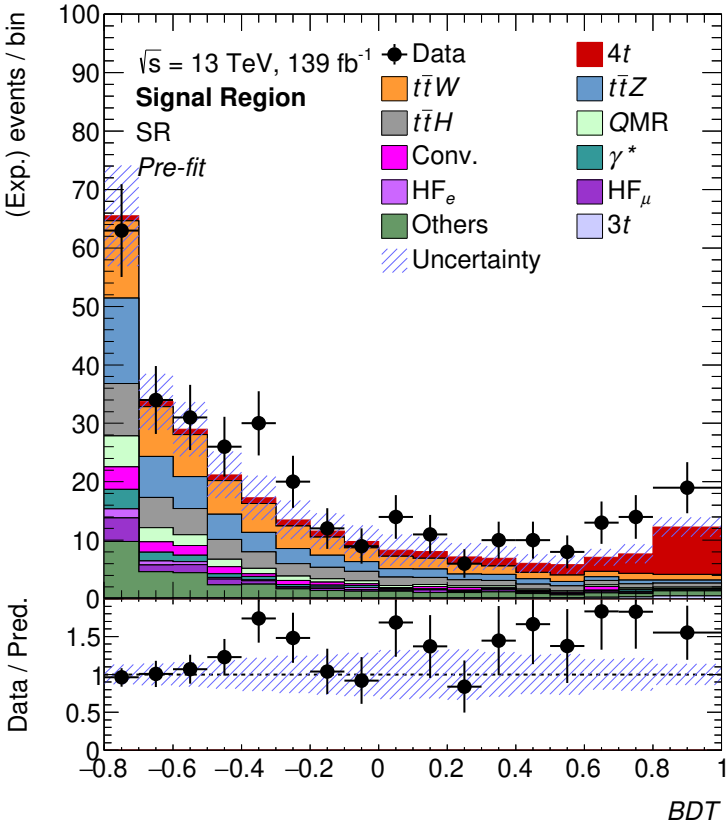


Figure 7.10: Distribution of the output of the BDT used for signal extraction in the SR. The background contributions are estimated from MC, with the exception of QMR, which is estimated using the data-driven method. The uncertainty includes the statistical as well as systematic uncertainties, where the systematic model is introduced in section 7.6.

7.6 Systematic Uncertainties

Following the argument in section 7.1, there are many source of systematic uncertainties associated to complex analyses as those that take place in HEP. For the global fit, these uncertainties are considered in form of NPs θ with a Gaussian constraint. The different sources, types and magnitudes of the systematic uncertainties taken into account in this analysis are described in this section.

The systematic uncertainties are translated into up and down variations of the nominal distributions that enter the fit. These variations can, for example, be global variations of the normalisation of a certain process (and in a certain region) or variations of the shape of the distribution of a certain process (in a certain region). Before being con-

sidered by the fit, uncertainties have to undergo three pre processing steps. These steps are *smoothing*, *symmetrisation* and *pruning* and they are explained in the following.

Smoothing is used to suppress statistical fluctuations in the systematic variation, as these would destabilise the fit and, more importantly, lead to double counting of statistical uncertainties as those are treated separately in the fit. Here, an iterative approach is used to smooth the distributions. First, the distributions are rebinned in a way that the relative statistical uncertainty in each bin is below 8 % and so that there are less than five inflexions in the distribution. If these conditions cannot be met, the procedure is repeated with an upper bound on the relative statistical uncertainty that is half that of the previous iteration until both conditions are fulfilled. The rebinned distributions are then used as input for the 353QH algorithm (broadly discussed in section 4.2 of [135]) which uses quadratic interpolation and running averages as implemented in the ROOT framework.

Symmetrisation follows the assumption that variations introduced by the systematic uncertainty follow a normal distribution $\mathcal{N}(\mu,\sigma)$ around the nominal μ in each bin. The variance of the normal distribution σ is then given by the difference between the yield of the varied distributions and the nominal one for each bin. Since the normal distribution is symmetric around μ, also the variations are symmetrised.

If the uncertainty is given in a one-sided way, i.e. only one variation (up or down) is known, the symmetrised variation is obtained by simply mirroring the known variation around the nominal. This is, for example, the case for uncertainties related to resolution effects, as a degradation of the resolution only yields one variation of the nominal.
If both variations are known, the symmetrisation is called two sided and the symmetrised variations are given by $x_i^{\text{nominal}} \pm \sigma_i^{sym}$, where σ_i^{sym} is defined to be the arithmetic mean of the difference between the yield of the up and down variation and the yield of the nominal distribution in each bin

$$\sigma_i^{\text{sym}} = \frac{\left|x_i^{\text{up}} - x_i^{\text{nominal}}\right| + \left|x_i^{\text{down}} - x_i^{\text{nominal}}\right|}{2}. \tag{7.5}$$

Some systematic uncertainties are not symmetrised, which is mostly the case when there is a physical argument to be made that the variation should not have a symmetric impact. In this case, the smoothed but not symmetrised variations are used in the fit.

The effects of smoothing on the uncertainty on the signal prediction obtained by varying the renormalisation and the factorisation scales, denominated `tttt varRF`, are shown in Figure 7.11. This uncertainty is described in more detail below and serves at this point for a visualisation of the impact of smoothing on a systematic uncertainty that is not symmetrised. In this figure, the distribution of the nominal (black, solid), the original up variation (red, dashed), the original down variation (blue, dashed) and their smoothed counterparts (red and blue, solid, resp.) of the signal extraction BDT for $4t$ events in the SR are shown. The shaded blue band represents the statistical uncertainty on the nominal distribution. The bottom panel of Figure 7.11 depicts the relative deviation of the variation from the nominal in percent. Two things are noticeable here. First, the smoothed variations appear considerably smoother than their non-smoothed counterparts, while still conserving trends (the red and blue solid lines in the bottom panel are not flat). This is important as otherwise any shape information from the systematic uncertainty would be lost in the smoothing. Second, the smoothed variations are not symmetric, which is of course desired in this case.

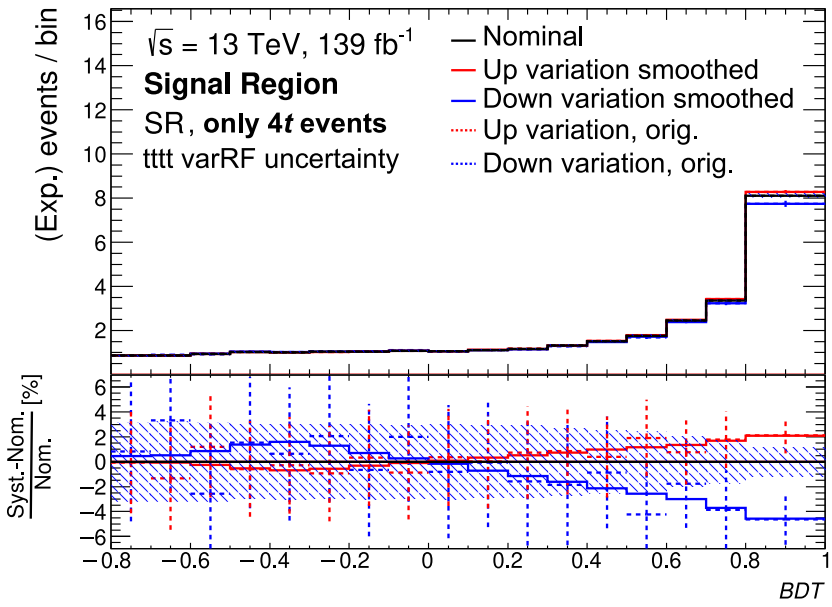


Figure 7.11: Distributions of the nominal (black, solid), the original up variation (red, dashed), the original down variation (blue, dashed) and their smoothed counterparts (red and blue, solid, resp.) of the signal extraction BDT in the SR. The variations are shown for the `tttt varRF` uncertainty and only for $4t$ events. The shaded blue band represents the statistical uncertainty on the nominal distribution. The bottom panel depicts the relative deviation of the variation from the nominal in percent.

Pruning refers to the removal of uncertainties and is mostly applied in order to stabilise and significantly speed up the global fit. Evidently, pruning too many uncertainties artificially reduces the obtained uncertainty on the fit results, while pruning too few can impede the fit to converge within a reasonable time scale. Different pruning methods are applied on uncertainties that affect the normalisation of a process and on those that affect the shape of the distribution of a given process. For this analysis, normalisation uncertainties are pruned if the relative difference between the integral of the varied distribution and the one of the nominal distribution is lower than 0.5 %. For shape uncertainties, the difference between the variation and the nominal is evaluated for each bin and the uncertainty is removed if the difference is smaller than 0.5 % in all bins.

7.6.1 Reconstruction Uncertainties

Data taking conditions include the uncertainty on the integrated luminosity (`luminosity`) of the full Run-II data set, measured to be 1.7 % [62, 136] by dedicated groups of the ATLAS collaboration. This uncertainty is applied to all contributions from MC simulation as the nominal luminosity is used to derive the corresponding expected yields. Additionally, a `Pileup reweighting` uncertainty is applied, which relates to the $\pm 1\sigma$ variation of the *pileup weight* used to reweight the original simulated pileup distribution to the one measured in data (see Figure 7.2).

Lepton uncertainties encompass systematic uncertainties coming from the estimation of the electromagnetic energy scale `ATLAS_EM_SCALE(_AFII)`, where AFII refers to a simplified detector simulation, from energy resolution `ATLAS_EM_RES`, which is estimated following [75, 81], and from disagreement between data and simulation represented by SFs. Different SFs are derived for trigger, reconstruction, identification and isolation efficiencies and the corresponding uncertainties are called `ATLAS_[EL/MU]_SF_[TRIGGER/RECO/ID/ISO]`. Additionally, one uncertainty is assigned for the SF applied to the outcome of the ECIDS BDT, used to suppress QMR electrons, denoted as `ATLAS_EL_SF_ChargeID_Stat`. For muons, uncertainties are assigned for properties of the track in the ID, `ATLAS_MU_ID`, properties of the track in the muon system, `ATLAS_MUF_MS`, the momentum scale `ATLAS_MU_SCALE`, `ATLAS_MU_SAGITTA_[RHO,RESBIAS]`, where are more detailed description is found in [81]. Finally, an uncertainty is given for the SF of the Track-To-Vertex-Association (TTVA) of muons separated into a statistical and a systematic component `ATLAS_MU_SF_TTVA_[Stat,Syst]`. This separation is also used for the trigger, reconstruction, identification and isolation systematics of the muon, while they are not split for electrons.

Jet uncertainties are numerous seeing that jets are complex composite objects. As described in chapter 5, the JVT is used to suppress pile-up jets and its efficiency is determined on a data set containing events from $Z+$jets and $t\bar{t}$ decays. The uncertainty obtained in this measurement is called `Jet vertex tagger efficiency`. Following the procedure given in [113], a list of 29 systematic uncertainties is assigned to the JES determination, grouped into `JES EffectiveNP`, `JES EtaIntercalibration`, `JES Flavor`, `JES Pileup`, `JES PunchThrough MC16 [AFII]`, `JES RelativeNonClosure-MC16AFII` and `JES SingleParticle HighPT`. The set of uncertainties corresponding to `JES EffectiveNP` is the result of diagonalisation of the matrix of uncertainties for all kinematic bin pairs used for the measurement of JES.

Flavour tagging uncertainties represent a set of 85 systematic variations coming from the diagonalisation of the matrix of uncertainties for all kinematic bin pairs used to determine the b-tagging efficiencies and SFs. The nuisance parameters related to the b-tagging efficiency are then denoted `b-tagging MV2c10 B[0-44]`, while the nuisance parameters relating to the erroneous tagging of c jets and light jets as b-tagged jets are grouped in `b-tagging MV2c10 C/Light[0-19]`. Seeing that these uncertainties are the results of the diagonalisation of a matrix, they each represent linear combinations of a series of physical uncertainties.

$E_{\mathrm{T}}^{\mathrm{miss}}$ uncertainties are all related to the determination of the *soft* term of the reconstruction of $E_{\mathrm{T}}^{\mathrm{miss}}$ as described in chapter 5. These uncertainties are estimated in events originating from processes without real $E_{\mathrm{T}}^{\mathrm{miss}}$ contributions, such as $Z \to ee$. In events of those processes, the soft term should correspond exactly to the negative sum of all other terms used to compute $E_{\mathrm{T}}^{\mathrm{miss}}$, so that the resulting $E_{\mathrm{T}}^{\mathrm{miss}}$ would be zero. Due to resolution and other effects, the obtained $E_{\mathrm{T}}^{\mathrm{miss}}$ value is generally not exactly zero but rather some residual, which is determined in data and simulation. The difference between the two residuals is taken as uncertainty on the estimation of the soft term in MC. This procedure provides three nuisance parameters, namely `ATLAS_MET_SoftScale` and `ATLAS_MET_SoftRes[Perp,Para]`, where the former relates to the parallel component (parallel to the beam pipe) of the energy scale and the latter to the resolution of the parallel and perpendicular components, respectively. More detail can be found in [137].

7.6.2 Physics Background Uncertainties

In general, five uncertainty types are estimated for physics processes followed by additional uncertainties that are required by the setup of the analysis.

First, the impact of different ways to compute the matrix element, represented as the choice of MC generator, called `[process] modelling (generator)`. This uncertainty is estimated by exchanging the nominally used generator with an alternative one.

Second, different approaches to the modelling of the parton showering are studied by replacing the nominally used showering algorithm with an alternative one, denominated `[process] modelling (shower)`.

Third, the effect of currently missing higher order QCD corrections to the prediction is studied by a variation of the renormalisation scale μ_{R}, while the impact of different scale thresholds between the PDF and the hard-scattering regime is estimated by a variation of the factorisation scale μ_{F}. For this analysis, both scales are varied simultaneously by a factor of 2 and 0.5, providing a two-sided uncertainty called `[process] varRF`.

Fourth, the impact of the choice of the PDF set is estimated by exchanging the nominal set with a alternative ones. Here, this is done in separate from the main analysis and the envelop on the variations are derived. The uncertainty used in this analysis is then a global variation of event yields according to this envelop and it is referred to as `[process] PDF` in the following. This corresponds to the procedure described in [18].

Fifth, an uncertainty on the predicted cross section of a given process is applied, which is either the result of a previous measurement, which is evidently only possible if this process has already been observed, or the uncertainty given by the theoretical computation of the cross section. This set of uncertainties is labelled `[process] Cross-Section`.

In addition to the above uncertainties, additional shape uncertainties are added for the processes of which the normalisation is estimated with the template fit method. This is because for these processes, the shape of the distributions used in the fit play a major role and rely entirely on the information provided by the simulation. A reasonable uncertainty on this shape therefore directly determines the quality of the background estimation and allows the use of data to slightly correct the used templates. These uncertainties need to be derived for every region but are given as a single, combined uncertainty called `[process] ShapeSyst` in this chapter.

Finally, the simulation of the correct number of b jets is fairly difficult, while at the same time N_b is a crucial variable for the analysis. Therefore, additional uncertainties are assigned to physics processes based on the true number of b jets (truth from MC). An uncertainty of 50 % is set for events with exactly three true b jets and for events with at least four true b jets and the uncertainties are called `[process] truth [3b,4b]`, respectively.

The exact choices and estimations as well as potential additional uncertainties for each of the physics backgrounds is given in the following.

$t\bar{t}W$, $t\bar{t}Z$ and $t\bar{t}H$ uncertainties For $t\bar{t}W$, the `ttW modelling (generator)` uncertainty is estimated by comparing an alternative MadGraph5_aMC@NLO to the nominal Sherpa with one additional jet. For $t\bar{t}Z$ the comparison is vice versa, albeit with no additional jet and for $t\bar{t}H$, the nominal PowhegBox is compared to MadGraph5_aMC@NLO.
The assigned PDF uncertainties for $t\bar{t}Z$ and $t\bar{t}H$ amount to 1 %, while the cross section uncertainties are set to 15 % and 20 %, respectively. No PDF nor cross section uncertainties are assigned to $t\bar{t}W$ as the normalisation of this process is freely determined by the fit to data and are thus included in the nominal value of $\mathrm{NF}_{t\bar{t}W}$.
The scale uncertainties (`[ttW,ttZ,ttH] varRF`) as well as the truth heavy flavour uncertainties (`[ttW,ttZ,ttH] truth [3b,4b]`), are determined as described above.

Additional uncertainties are assigned for $t\bar{t}W$ events with $N_j = 7$ and $N_j \geq 8$ as a result of the studies in section 8.3, showing that the result of this analysis is sensitive to these events in particular. Furthermore, a slight tension between data and prediction for these jet multiplicities is seen in the VR $t\bar{t}W$ region (see Figure 8.12). Most importantly, the studies in [133] show a strong dependence of additional, not yet simulated, contributions to the $t\bar{t}W$ cross section on the number of jets, so that additional degrees of freedom associated to $\mathrm{NF}_{t\bar{t}W}$ seem reasonable. The uncertainties amount to 125 % and 300 %, respectively, which is loosely based on the size of the tension observed in Figure 8.12.

$3t$ uncertainties include an uncertainty of 100 % on the predicted cross section and an uncertainty of 50 % is set on $3t + b$ events with at least four true b jets. The uncertainty on the cross section of this process is considerably large because this process has not yet been observed and because its normalisation has a noticeable impact on the outcome of the analysis. This is despite the fact that the nominal cross section of this process is an order of magnitude smaller than the signal cross section and it is due to the fact that the BDT is not able to easily separate $3t$ events from the signal (see Figure 8.3). The impact of the choice of the size of this uncertainty is thoroughly studied in subsection 8.2.2.

Others and V+jets uncertainties include uncertainties on processes from $t\bar{t}VV$, tX ($X = Z$,WZ), VV, VVV and VH. A 30 % uncertainty is set for the cross section of tX and V+jets, while it is set to 40 % for VV and to 50 % for $t\bar{t}VV$. The truth

heavy flavour uncertainties are estimated as described above and denoted `other truth [3b,4b]`[13].

7.6.3 Instrumental Background Uncertainties

QMR uncertainties encompass uncertainties originating from the data driven method used to determine this background contribution. This includes the statistical uncertainty as given by the likelihood fit, a variation of the window in the dielectron invariant mass around $m_{ee} = 91\,\text{GeV} \approx m_Z$, and a variation of the side band regions used to subtract background events from the region defined by this invariant mass window. Moreover it also includes differences between the estimation using this method (applied to MC) and the estimation using purely the information given by the MC (closure test). The uncertainties are derived for the SR, CR $t\bar{t}\,W$ and CR Conv./γ^* regions separately and consequently combined in a single NP called Q`MR Sys`.

Conversion and γ^* uncertainties include uncertainties on the shape of the distributions used in the fit as these background types are estimated using the template fit method. The uncertainties are estimated in a $Z(\rightarrow \mu\mu) + \gamma$ region, where data is compared to the prediction of Powheg + Pythia 8 of $Z(\rightarrow \mu\mu)$+jets events. The normalisation of the prediction is adjusted to match the normalisation of data, so that any discrepancy between the distributions is related exclusively to a difference in shape. The uncertainties are denoted `[CO,Gstr] ShapeSyst`.

HF$_{e,\mu}$ uncertainties include, similar to above, uncertainties on the shape of the distributions and not on the normalisation, as both contributions are estimated using the template fit method. The shape uncertainty is estimated by using the *loose* object definitions of leptons (see section 7.2), so as to further enrich the sample with non-prompt leptons. The uncertainties are then given by the difference between the normalised data and simulation distributions in each bin of and for each region of the fit. The uncertainties are derived separately for electrons and muons and the uncertainties for different regions are condensed (in a fully correlated way) into global NPs, called `[HFe,HFm] ShapeSyst`, respectively.

Light Flavour uncertainties include an uncertainty on the normalisation of events with light flavour non prompt leptons of 100 %. This follows the results of [132], and the corresponding NP is called `ttbar_light Cross-Section`. The light flavour non prompt background is merged with other backgrounds in the Others category as its

[13] Here, *other* does not fully correspond to the Other background category, but the naming is kept here for comparability to [1].

contribution to the regions in the fit is low.

Additionally to the above mentioned uncertainties, all events coming from $t\bar{t}(+b)$ events with exactly three (at least four) b-tagged jets have an uncertainty of 30 % assigned to their normalisation. The corresponding NPs are called `ttbar Cross-Section [3b,4b]`. These uncertainties try to account for difficulties in correctly estimating the heavy flavour contributions to $t\bar{t}$ processes as pointed out in the introduction to the physics backgrounds as well as in [138].

7.6.4 Signal Uncertainties

Seeing that the signal is a physics process, the estimation of uncertainties largely follows the recipe quoted above.
The uncertainty on the modelling of the parton showering and hadronisation of the signal process is estimated by a two-point variation, where the distributions of the signal process given by MadGraph5_aMC@NLO interfaced with Pythia8 is compared to the distributions using MadGraph5_aMC@NLO interfaced with Herwig7. The difference between the two cases is used as a one-sided variation, so that symmetrisation is used to obtain an up and down variation.
The impact of the scale variation on the obtained signal yields is estimated, as described above, by the simultaneous variation of μ_{R} and μ_{F} by a factor of 2 and 0.5.
The impact of the choice of PDF set is determined following the procedure given in [18] using the NNPDF30_nlo_as_0118 PDF set. The RMS is found to be lower than 1 % so that this value is set as the PDF uncertainty for the signal.

Finally, an additional uncertainty of 20 % on the predicted signal cross section is added, following the theoretical uncertainty of the prediction in [2]. This uncertainty is denominated `tttt cross-section` and it is only used during the estimation of μ_{4t} and the corresponding statistical significance. More specifically, it is not applied for the measurement of the signal cross section as is described in more detail in section 7.7.

7.7 Results

Following the analysis steps outlined above, the global fit to observed data in all regions results in a post fit value for the signal strength of [1]

$$\hat{\mu}_{4t} = 2.0^{+0.4}_{-0.4}\,(\text{stat.})\,^{+0.7}_{-0.5}\,(\text{syst.}) = 2.0^{+0.8}_{-0.6}\,.$$

This signal strength represents the ratio between the observed cross section of the four top quark production σ_{4t} and the SM prediction at NLO in QCD and EW[14], $\sigma_{4t}^{\text{theo}} = 12.0 \pm 2.4\,\text{fb}$ [2], so that a result of $\hat{\mu}_{4t} = 1$ would indicate a perfect agreement between observation and prediction. Since the obtained value of $\hat{\mu}_{4t}$ directly depends on $\sigma_{4t}^{\text{theo}}$, it also directly depends on the associated theoretical uncertainty of 20 %, which is therefore propagated as part of the systematic uncertainty on $\hat{\mu}_{4t}$. A detailed list of the different sources of uncertainties associated to $\hat{\mu}_{4t}$ is given in Table 7.7. There, the contribution of each uncertainty source to the global uncertainty on $\hat{\mu}_{4t}$ is estimated by doubling or removing the uncertainty in question at its post fit value and subsequently recomputing $\hat{\mu}_{4t}$.

The observed $\hat{\mu}_{4t}$ is compatible with the SM prediction at NLO within 1.7 standard deviations.

As described in subsection 7.4.2, the normalisations of five major background contributions are estimated simultaneously with μ_{4t} during the fit. This is mostly achieved by extending the fit with dedicated CRs. The post fit results for the NFs, that is the ratio between the observed and the predicted normalisation for each of these background contributions are given in Table 7.6, where the quoted uncertainties contain

Table 7.6: Post fit NFs obtained in the global fit to data. The results are listed for the nominal setup as presented in [1], as well as for a setup with an uncertainty of 20 % assigned to the cross section of each of the processes corresponding to the NFs.

Setup	$\hat{\text{NF}}_{t\bar{t}W}$	$\hat{\text{NF}}_{\text{Conv}}$	$\hat{\text{NF}}_{\gamma^*}$	$\hat{\text{NF}}_{\text{HF}e}$	$\hat{\text{NF}}_{\text{HF}\mu}$
nominal	$1.6^{+0.3}_{-0.3}$	$1.6^{+0.5}_{-0.4}$	$0.9^{+0.4}_{-0.4}$	$0.8^{+0.4}_{-0.4}$	$1.0^{+0.4}_{-0.4}$
w/ cross section uncertainties	$1.6^{+0.5}_{-0.4}$	$1.6^{+0.7}_{-0.5}$	$0.9^{+0.5}_{-0.4}$	$0.8^{+0.5}_{-0.4}$	$1.0^{+0.5}_{-0.4}$

[14] For a more detailed description of the predicted cross section, the reader is referred back to chapter 2.

both statistical as well as systematic uncertainties.
As described in section 7.6, no uncertainty is set for the cross sections of the processes that correspond to the different NFs, arguing that this is not needed seeing that these normalisations are determined in data.
This is true for the final FOMs of this analysis as there, only the nominal value of the NFs is propagated. In order to properly discuss the compatibility of the observed and the predicted normalisation, however, these uncertainties play an important role. This closely follows the argument made for adding the uncertainty on the signal cross section when discussing $\hat{\mu}_{4t}$.
To illustrate the impact of these uncertainties on the uncertainties of the obtained post fit NFs, the fit is repeated with the addition of arbitrary uncertainties of 20 % on the cross section of each of the processes to illustrate the impact. The results of this fit are also given in Table 7.6.

The post fit NFs of $\hat{\mathrm{NF}}_{\gamma^*}$, $\hat{\mathrm{NF}}_{\mathrm{HF}e}$ and $\hat{\mathrm{NF}}_{\mathrm{HF}\mu}$ obtained in the nominal setup show a good agreement between the prediction and observation of the corresponding background categories. This is represented by the fact that the NFs are compatible with 1 within the post fit uncertainties.
These background contributions come from $t\bar{t}$ + jets events, where no systematic uncertainty is assigned for the $t\bar{t}$ cross section for $N_b < 3$ and a 30 % uncertainty is assigned for $N_b > 3$.

$\hat{\mathrm{NF}}_{\mathrm{Conv}}$ is compatible with the prediction within 1.5 and 1.2 standard deviations considering the same systematic uncertainties on $t\bar{t}$ + jets events or additionally considering the uncertainty on the cross section, respectively.

The obtained value for $\hat{\mathrm{NF}}_{t\bar{t}W}$ (with additional uncertainty) is compatible with the prediction within 2 (1.5) standard deviations and it is compatible with previous results from the search for $t\bar{t}H$ production in multilepton final states with the ATLAS detector [132][15].
The selections used in this analysis in terms of jet and b jet multiplicities represent a relatively little studied area of the phase space of $t\bar{t}W$ production so that tensions between observation and prediction within two standard deviations is not necessarily unexpected. Additionally, a series of studies have been performed in [133], investigating the impact of sub leading EW effects ($tW \to tW$ scattering) and spin correlations

[15] For a correct comparison between the results in [132] and the results presented here, it is important to note that an additional correction to the predicted $t\bar{t}W$ cross section of 20 % is added for the search for $t\bar{t}H$ to account for additional jets and electroweak effects.

on the differential cross section of $t\bar{t}W$. These studies found an impact of these contributions on the $t\bar{t}W$ event yield for the phase space used in the $t\bar{t}H$ analysis of 10 % and corroborated the notion that their impact is dependent on jet multiplicity.

In order to study the impact of different NF, particularly in light of the slight tension of $\mathrm{NF}_{t\bar{t}W}$, extensive tests are performed and presented in subsection 8.2.1.

Table 7.7: Breakdown of the uncertainties on the post fit signal strength $\hat{\mu}_{4t}$ grouped in categories. The contribution of each uncertainty source to the global uncertainty on $\hat{\mu}_{4t}$ is estimated by doubling or removing the uncertainty in question at its post fit value, subsequently recomputing $\hat{\mu}_{4t}$.

Uncertainty source	$\Delta\hat{\mu}_{4t}$	
Signal modelling		
$4t$ cross section	+0.56	−0.31
$4t$ modelling	+0.15	−0.09
Background modelling		
$t\bar{t}W$ modelling	+0.26	−0.27
$3t$ modeling	+0.10	−0.07
Non-prompt leptons modeling	+0.05	−0.04
$t\bar{t}H$ modelling	+0.04	−0.01
$t\bar{t}Z$ modelling	+0.02	−0.04
Others modelling	+0.03	−0.02
QMD	+0.01	−0.02
Instrumental		
Jet uncertainties	+0.12	−0.08
Jet flavour tagging (light-jets)	+0.11	−0.06
Simulation sample size	+0.06	−0.06
Luminosity	+0.05	−0.03
Jet flavour tagging (b-jets)	+0.04	−0.02
Other experimental uncertainties	+0.03	−0.01
Jet flavour tagging (c-jets)	+0.03	−0.01
Total systematic uncertainty	+0.70	−0.44
Others		
Statistical	+0.42	−0.39
$\mathrm{NF}_{\mathrm{Conv}}$, NF_{γ^*}, $\mathrm{NF}_{\mathrm{HF}e}$ and $\mathrm{NF}_{\mathrm{HF}\mu}$	+0.05	−0.04
$\mathrm{NF}_{t\bar{t}W}$	+0.04	−0.04
Total uncertainty	+0.83	−0.60

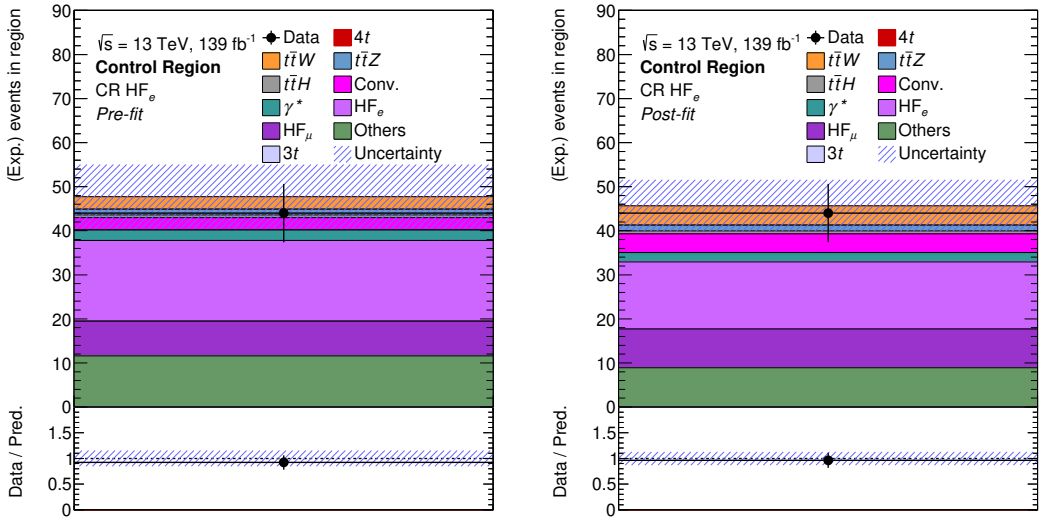
As a means to visually validate if the template fit method is able to properly describe the data in the CRs, the pre and post fit distributions of the discriminant variable (or overall event yield) in each of the CRs is shown in Figures 7.12 and 7.13. While the upper panel shows the (expected) yield, the lower panel depicts precisely the agreement between data and prediction. The uncertainties shown in these figures are comprised of the statistical as well as of the systematic uncertainties.

Already the pre fit distributions show a reasonable agreement between data and prediction for the $\gamma*$, HF_{μ} and HF_{e} backgrounds, which can be seen in Figure 7.12 and the first bin of Figure 7.13a and which explains the good compatibility of the corresponding NFs with 1.
On the other hand, the Conversion and $t\bar{t}\,W$ backgrounds are consistently under estimated, which can be seen in Figure 7.13a and Figure 7.13b, and which is reflected in the obtained post fit NFs as well.

In contrast to the pre fit agreement, the post fit prediction agrees reasonably well with data in all CRs. This indicates that the fit is indeed able to converge and find a set of NFs values to describe the observed data in the CRs.

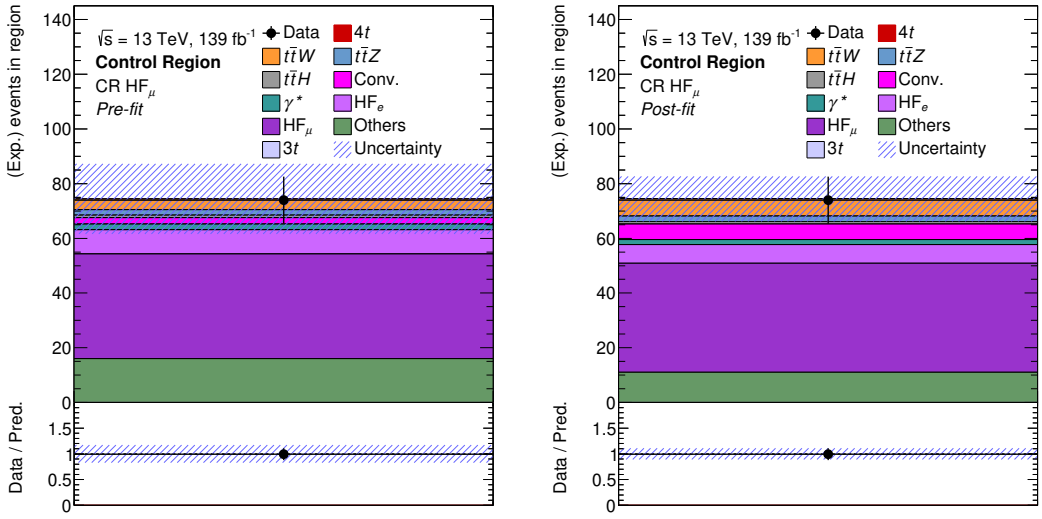
Only very few of the nuisance parameters that enter the fit are pulled or constrained by it. Most notable are the systematic uncertainties `ttW Syst [7j,8j]`, resulting in values of $0.18^{+0.73}_{-0.61}$ and $0.22^{+0.56}_{-0.42}$, respectively, which in turn result in an increase of 22 % (65 %) of $t\bar{t}\,W$ events with $N_j = 7$ ($N_j \geq 8$). All other nuisance parameters are only marginally or not at all affected by the fit. The full list of post fit nuisance parameters is given in Figure B.3.

The correlations between the parameters entering the fit are shown in Figure 7.14, for all parameters that have at least one $|\rho| > 0.1$, where ρ is the correlation coefficient for one pair of parameters. The parameter most correlated with the signal strength is the uncertainty on the SM signal cross section with $\rho = -0.59$, which is expected seeing that the SM signal cross section is the denominator of μ_{4t}. The second to most correlated parameter is the uncertainty for $t\bar{t}\,W$ events with at least eight jets, yielding $\rho = -0.32$, which is discussed in more detail in subsection 8.3.3. Another notable correlation exists between NF_{γ^*} and $\mathrm{NF}_{\mathrm{Conv}}$, $\rho = -0.42$, which boils down to the fact that both normalisation factors are essentially estimated in the same CR as they leave a similar signature. No unexpected correlations arise from the fit.



(a) CR HF_e pre fit

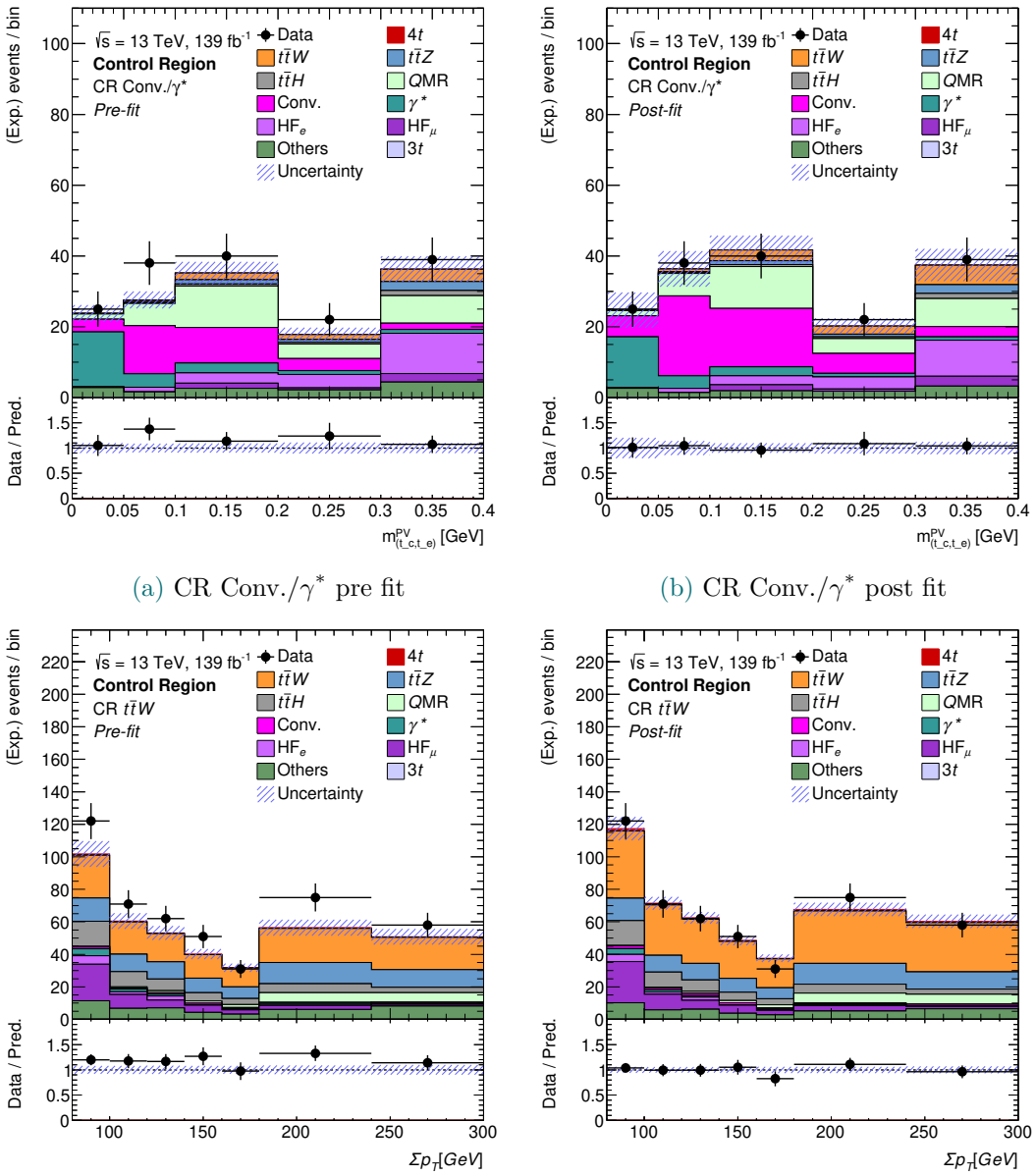
(b) CR HF_e post fit



(c) CR HF_μ pre fit

(d) CR HF_μ post fit

Figure 7.12: Pre and post fit expected and observed event yields for the CR HF_e and CR HF_μ regions. The uncertainties include statistical as well as systematic uncertainties.



(a) CR Conv./γ^* pre fit

(b) CR Conv./γ^* post fit

(c) CR $t\bar{t}\,W$ pre fit

(d) CR $t\bar{t}\,W$ post fit

Figure 7.13: Pre and post fit expected and observed event yields for the discriminant variables used in the template fit for the CR Conv./γ^* and CR $t\bar{t}\,W$ regions. The uncertainties include statistical as well as systematic uncertainties.

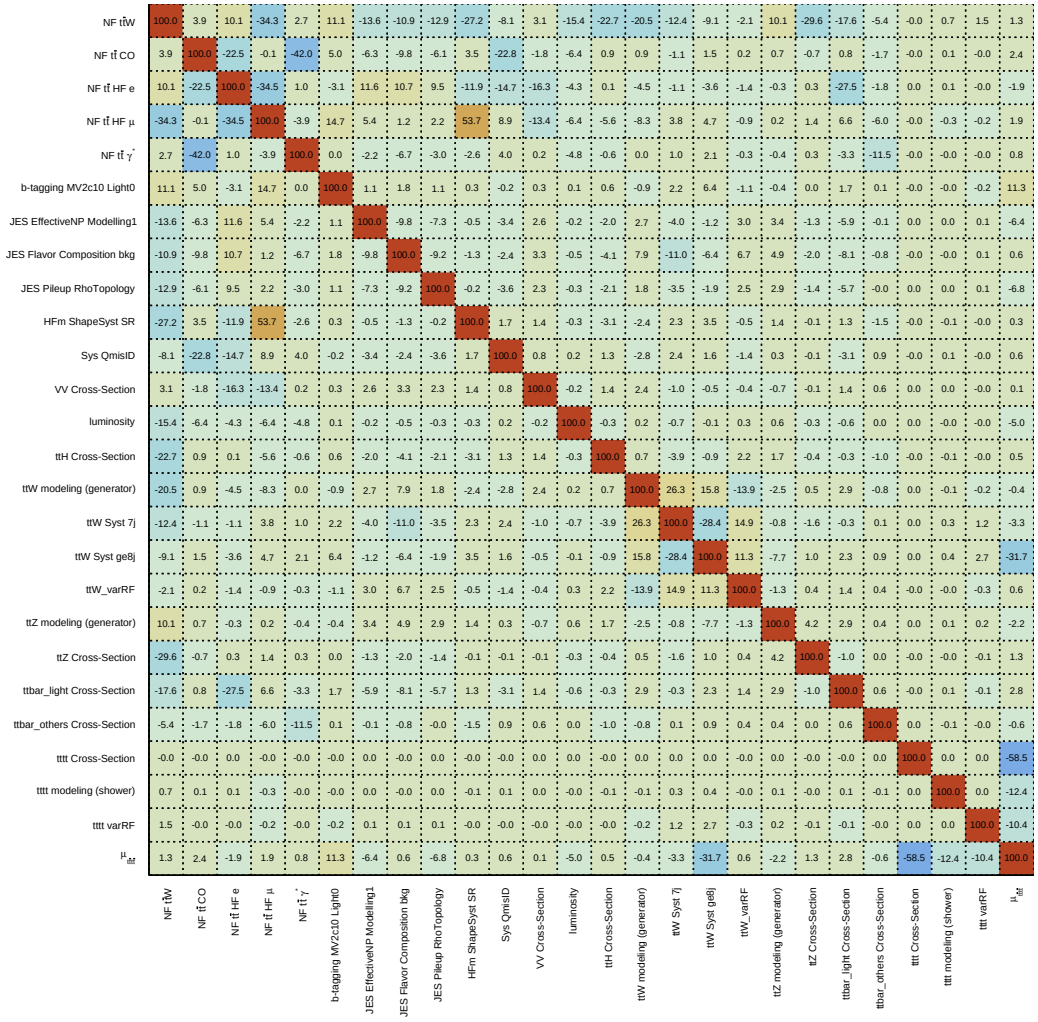


Figure 7.14: Correlation coefficients for pairs of parameters entering the fit ρ only considering parameters that have at least one $|\rho| > 0.1$, where ρ is computed for all possible parameter pairs. The numbers shown here correspond to $100 \times \rho$.

The post fit model is used to determine the significance of rejecting the background only hypothesis, resulting in an observed significance of [1]

$$Z_{\text{obs}} = 4.3$$

with an expected significance of

$$Z_{\text{exp}} = 2.4\,,$$

so that first evidence for four top quark production assuming SM kinematics is achieved.

Using the kinematic information from the simulation of the signal process, the observed production cross section σ_{4t} is determined using $\hat{\mu}_{4t}$ by computing

$$\sigma_{4t} \equiv \sigma_{t\bar{t}t\bar{t}}^{\text{SM, nominal}} \cdot \hat{\mu}_{t\bar{t}t\bar{t}}^{\text{no } \sigma \text{ uncert.}}, \tag{7.6}$$

where $\sigma_{t\bar{t}t\bar{t}}^{\text{SM, nominal}}$ represents the nominal value of the predicted $t\bar{t}t\bar{t}$ cross section, i.e. without theoretical uncertainties and $\hat{\mu}_{t\bar{t}t\bar{t}}^{\text{no } \sigma \text{ uncert.}}$ represents the post fit measured signal strength also removing the uncertainty on the predicted production cross section[16]. This results in

$$\sigma(pp \to t\bar{t}t\bar{t}) = \sigma_{4t} = 24^{+5}_{-5}\ (\text{stat.})\ ^{+5}_{-4}\ (\text{syst.})\ \text{fb} = 24^{+7}_{-6}\ \text{fb}.$$

Figure 7.15 shows the distributions of the signal extraction BDT output in the SR for the post fit model and data. The post fit and data yields are given in Table 7.8 for the entire SR and for the SR high BDT region. The latter represents a sub region of the the SR with the additional requirement of $BDT \geq 0$. This region is interesting since it represents the area of the phase space with a large signal to background ratio and therefore the area that has the largest impact on the determination of $\hat{\mu}_{4t}$. The region also serves to better illustrate the agreement between data and prediction, where the prediction is mostly following the kinematics of the signal.

The post fit model is propagated to other distributions and regions not used in the fit by applying the ensemble of parameters of the model that are not specific to an exact region to the expected yields in those new regions given by the simulation. Parameters that are specific to a region are for example shape uncertainties, while NFs represent

[16] This uncertainty is removed here, seeing that $\sigma_{t\bar{t}t\bar{t}}$ is the measured, free, parameter and does not depend on the normalisation of the prediction, but rather only on its shape.

one group of parameters that are assumed to be independent of the region. This is evidently not entirely true, so that this assumption in particular is probed in subsection 8.3.3.

The distributions of selected variables not explicitly used in the fit are shown for the expectation, propagated from the post fit model, and data in Figure 7.16.
Seeing that these distributions are not explicitly used in the fit, they can give an indication of the quality of the fit and the extrapolation of the post fit model (similar to the VRs introduced in subsection 7.4.2).
Figure 7.16a shows the distribution of the jet multiplicity, which had been removed from the list of input variables to the signal extraction BDT due to a potential mis modelling seen at the pre fit level for events with $N_j = 7$ (see Figure B.1). After the fit, and thus after the background estimation with the template fit method, this discrepancy is noticeably reduced and overall agreement is very good even for very high multiplicities of $N_j \geq 10$.
Figure 7.16b shows the number of b-tagged jets, which is not explicitly used in the BDT, but rather implicitly through the sum of b-tagging weights shown in Figure 7.16c. Both distributions exhibit a very good agreement between the expectation and data. This is particularly interesting seeing that b-tagging information is the most powerful information for discriminating the signal from background processes, so that the corresponding distributions provide a good handle to evaluate the agreement for signal and background separately.
Figure 7.16d shows the distribution of the minimum distance between two leptons, which is the fourth most important variable for signal extraction according to the BDT. This variable provides angular information which is notoriously difficult to model. This is not least the case because differential cross section measurements are not available for all involved processes. However, the expectation and data are found to agree well for this distribution.

In conclusion, all distributions that have been studied show a good agreement between the post fit expectation and data, indicating that the observed excess follows the properties of the signal. Nevertheless, a series of tests are performed to mitigate the possibility of estimating the signal content in data incorrectly, which are described in the following.

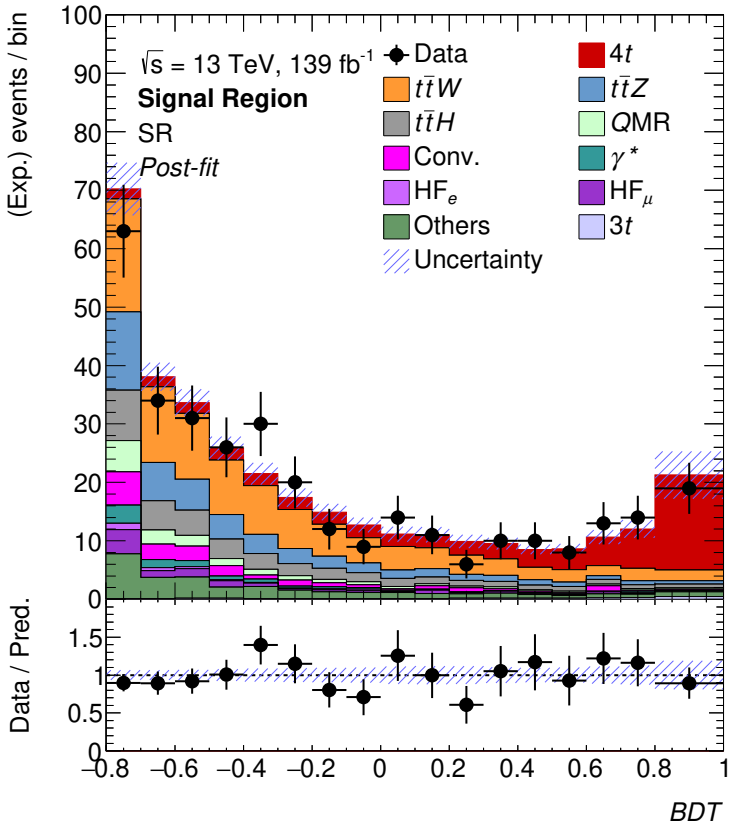
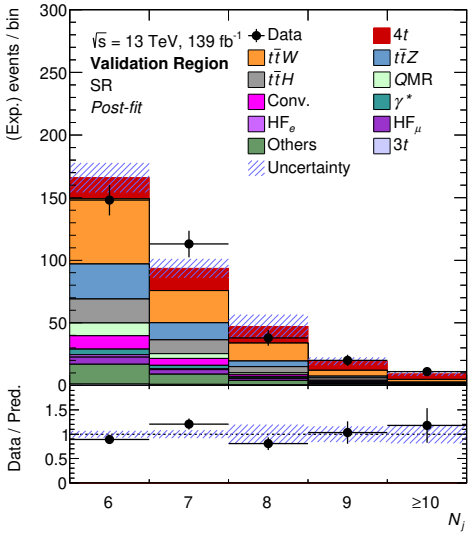


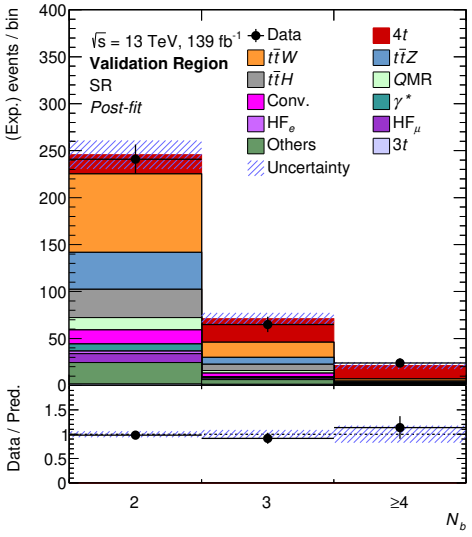
Figure 7.15: Distribution of the signal extraction BDT output for the post fit model (coloured areas) and data (black points) in the SR. The uncertainty band includes systematic as well as statistical uncertainties. This distribution is the only distribution used in the fit with a non-negligible impact on the signal strength μ_{4t}.

Table 7.8: Background and signal yields for the post fit model as well as data yields for the SR and SR high BDT regions. The latter represents the sub-region most enriched with the signal and is defined by the signal selection with the additional condition of $BDT \geq 0$.

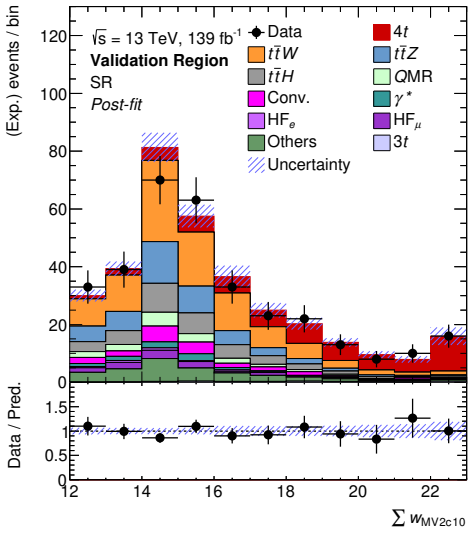
Process	SR	SR high BDT
$t\bar{t}W$	102 ± 26	23 ± 10
$t\bar{t}Z$	48 ± 9	9 ± 2
$t\bar{t}H$	38 ± 9	8 ± 2
Others	29 ± 7	7 ± 2
Conv	19 ± 6	3 ± 1
QMR	16.3 ± 0.4	2.7 ± 0.2
HF_μ	12 ± 6	3 ± 2
γ^*	9 ± 4	0.9 ± 0.5
HF_e	3 ± 3	1 ± 1
$3t$	3 ± 3	2 ± 2
Total bkg	278 ± 22	59 ± 10
$4t$	60 ± 17	44 ± 12
Total pred	337 ± 18	103 ± 10
Data	330	105



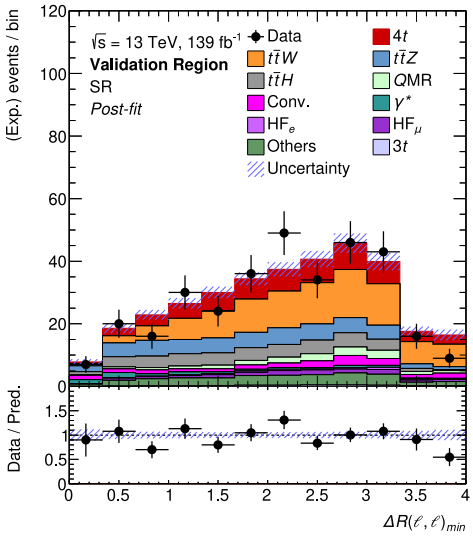
(a) Jet multiplicity



(b) b-tagged jet multiplicity



(c) Sum of b-tagging weights



(d) Min distance between two leptons

Figure 7.16: Selected distributions of the post fit model (coloured areas) and data (black points) in the SR. The uncertainty band includes statistical as well as systematic uncertainties. The denomination "Validation" indicates that these distributions are not used explicitly in the fit.

7.8 Post-Unblinding Tests

The obtained result for μ_{4t}, is compatible with the expectation of the SM within 1.7 standard deviations. Nonetheless, a series of tests is warranted to better understand the nature of the moderate excess observed in data.
These tests are conducted at the post unblinding stage and using real data in all regions. The following tests are described in this section:

1. Different scenarios for the signal strength μ_{4t}
2. Assuming different cross sections of $3t$ production
3. Testing for a bias introduced by the MVA approach by replacing the fit on the BDT distribution with a fit on H_{T}.
4. Comparing $\hat{\mu}_{4t}$ obtained in events with a positive sum of lepton charges with respect to events with a negative sum
5. Comparing $\hat{\mu}_{4t}$ obtained in different analysis channels
6. Comparing $\hat{\mu}_{4t}$ obtained in different data taking periods

7.8.1 μ_{4t} Scenarios

In order to test if the data could be described well without (or with less) signal, three different settings for μ_{4t} in the fit are tested. First, the background only fit with $\mu_{4t} = 0$, second, the fit assuming the SM prediction by fixing $\mu_{4t} = 1$ and third, the nominal fit with μ_{4t} as a free parameter of the fit. The uncertainty on the $4t$ cross section is not taken into account for the first two scenarios.

Figure 7.17 shows the comparison of the obtained post fit NFs and $\hat{\mu}_{4t}$ for the three different scenarios. The NFs differ only marginally between the three scenarios. Since the NFs are mostly constrained by the CRs, the fit uses systematic uncertainties that can be associated to the higher part of the BDT distribution in the SR, and thus where most of the excess is, especially if no signal is assumed.

Figure 7.18 shows the post fit distributions of the BDT in the SR for the three different setups for μ_{4t}. There, the background only, $\mu_{4t} = 0$, and the $\mu_{4t} = 1$ scenario are not able to convincingly describe the data, whereas the agreement between the post fit model and data is very good for the nominal fit scenario. This is a strong suggestion

that the excess in data in the SR is signal-like in terms of the information learnt by the BDT.

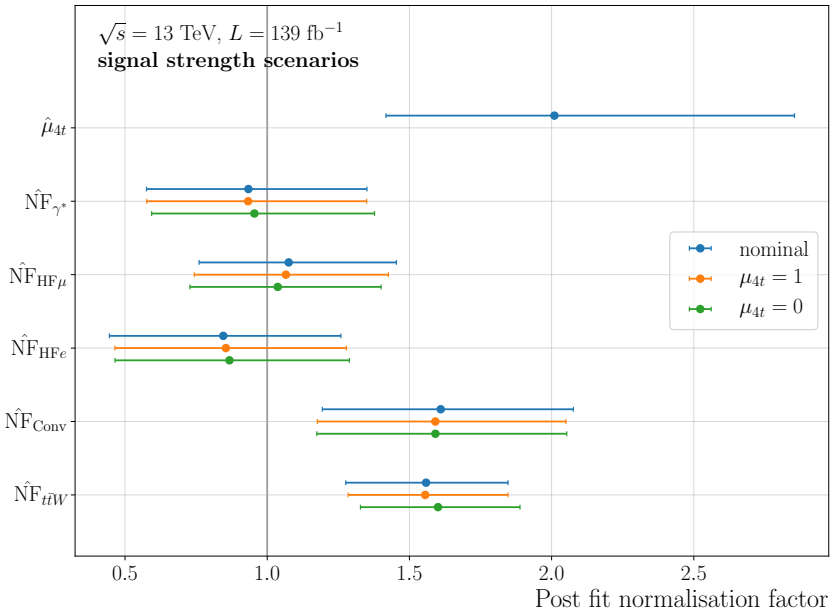
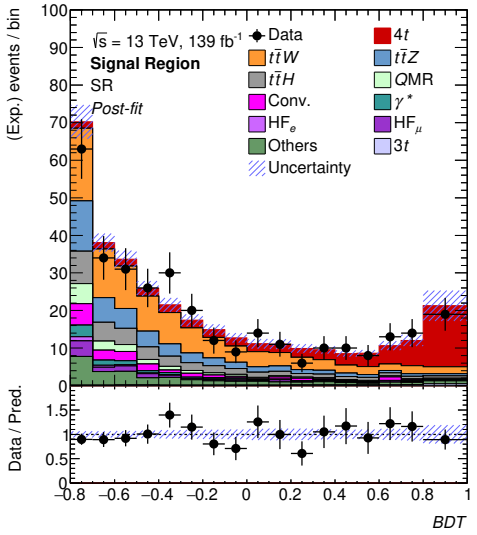
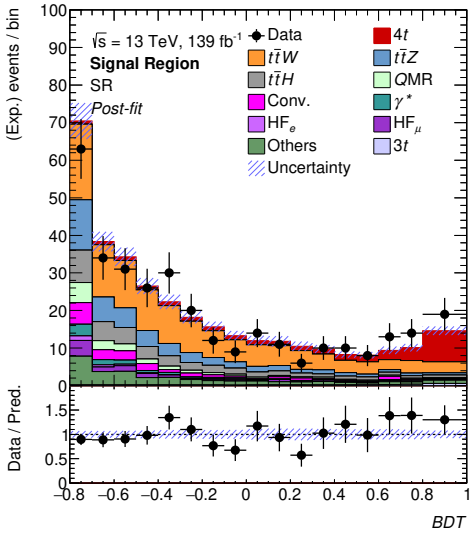


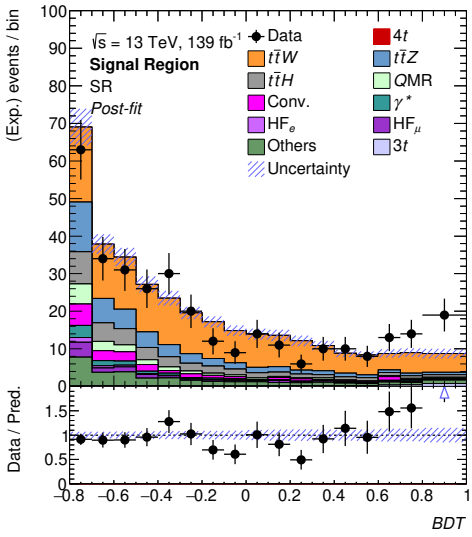
Figure 7.17: Post fit NFs and signal strength for the nominal fit, a fit scenario with $\mu_{4t} = 1$ fixed and with $\mu_{4t} = 0$ fixed, corresponding to the SM signal prediction and the background only fit, respectively. The post fit signal strength is only given for the nominal fit as it is constant for the other two scenarios.



(a) nominal



(b) $\mu_{4t} = 1$



(c) $\mu_{4t} = 0$

Figure 7.18: Distributions of the signal extraction BDT of the post fit model and data in the SR for three different setups of μ_{4t}. The setups correspond to the nominal fit, the SM prediction and the background only model respectively

7.8.2 $3t$ Cross Section Scenarios

As mentioned in section 7.6, the value of the predicted cross section of $3t$ production can have an impact on the obtained $\hat{\mu}_{4t}$, because the BDT is not easily able to differentiate between events from $3t$ production and the signal. The exact impact is studied here, where the fit is performed with the cross section scaled by $S_{3t} = 1.0, 1.5, 2.0$ and 5. The uncertainty on the cross section remains 100 %, as in the nominal case. The results of these fits are shown in Figure 7.19.

While the NFs show no change in both the nominal value as well as in the uncertainty, a change of up to 14 % can be seen for the nominal value of $\hat{\mu}_{4t}$. More importantly for the significance, the uncertainty on $\hat{\mu}_{4t}$ increases by approximately the same amount for $S_{3t} = 5$. However, even in a very extreme scenario of five times the predicted cross section, evidence can still be claimed for the four top quark production.

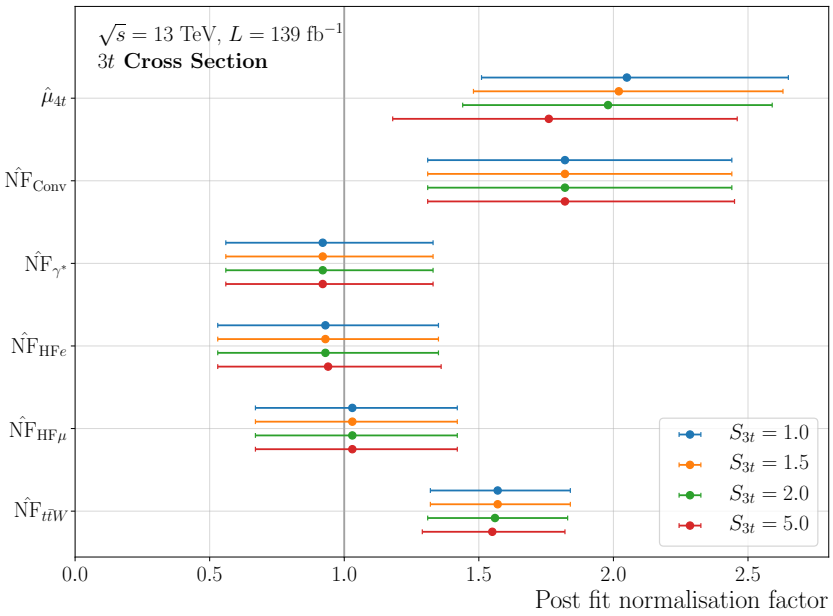


Figure 7.19: Post fit NFs and signal strength for different scales S_{3t} applied to the cross section of $3t$ production.

7.8.3 Fit using H_{T}

In order to exclude the possibility that the observed excess in the SR is an artefact from the MVA approach, a fit is performed using the distributions of H_{T} in SRs split between same-sign dilepton and trilepton events and N_b.
This split results in five SRs, i.e. $2b2l$, $3b2l$, $2b3l$, $3b3l$ and $4b$, which is inclusive over the lepton multiplicity. This split is loosely comparable to the SR definition of the previous search for four top quark production in ATLAS [43]. The obtained NFs and $\hat{\mu}_{4t}$ are shown in Figure 7.20, where they are compared to the nominal results. The NFs are essentially unchanged by the different fit set up, which is expected mostly because the CRs are unchanged. The obtained $\hat{\mu}_{4t}$ is compatible with the results from the nominal setup. The significances for the fit on H_{T} are $Z_{\mathrm{obs}} = 4.26$ (compared to 4.4) and $Z_{\mathrm{exp}} = 2.07$ (compared to 2.4), where the latter is an indicator on the gain of using an MVA approach over a fit on H_{T}.

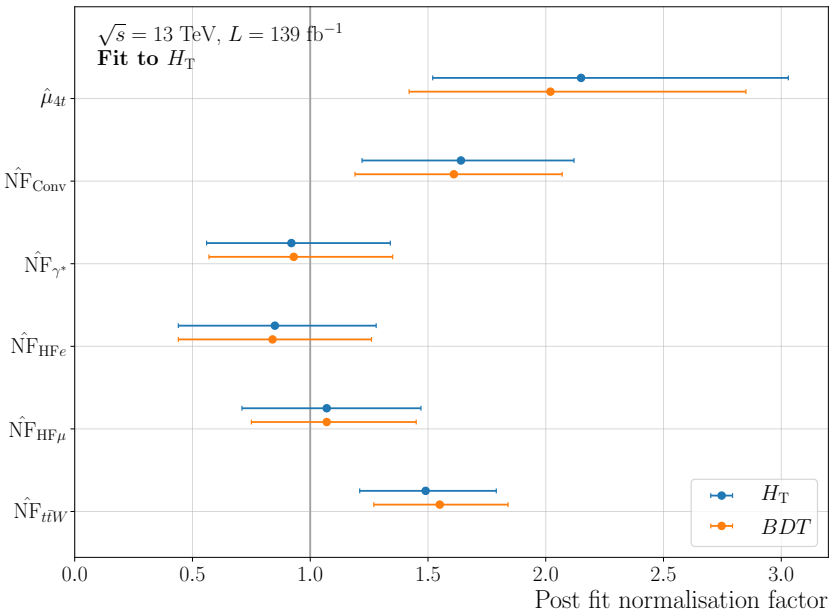


Figure 7.20: Post fit NFs and signal strength for the nominal fit setup using the BDT in the SR and a setup using H_{T} in a group of SRs following [43].

7.8.4 Sum of Lepton Charge

Seeing that $t\bar{t}W$ is (one of) the most influential backgrounds and that a slight tension between the observed and predicted normalisation has been found, additional studies of the impact of $\mathrm{NF}_{t\bar{t}W}$ on $\hat{\mu}_{4t}$ are performed.

Following [133], one reason for the observed tension could lie in an incorrect prediction of the charge asymmetry of $t\bar{t}\,W$, which is therefore studied here.

In order to do so, the SR is split into a positive SR, where the sum of lepton charges is greater than zero, and a negative one. The same split is applied to the $t\bar{t}\,W$ CR and μ_{4t} is replaced by a positive μ_{4t}^{+} and a negative μ_{4t}^{-} and fixed to one in all CRs but CR $t\bar{t}\,W$. Likewise, $\mathrm{NF}_{t\bar{t}W}$ is split into a positive and a negative part and all systematic uncertainties pertaining to $t\bar{t}\,W$ are fully decorrelated between the positive and negative selections. The obtained NFs and signal strengths are shown in Figure 7.21.

The obtained μ_{4t}^{+} and $\hat{\mathrm{NF}}_{t\bar{t}\,W}^{+}$ are consistently larger than their negative counterparts indicating that the excess does show a slight charge asymmetry. However, the NFs and signal strengths obtained in the different scenarios agree within their uncertainties so that the trend towards asymmetry has no significant impact.

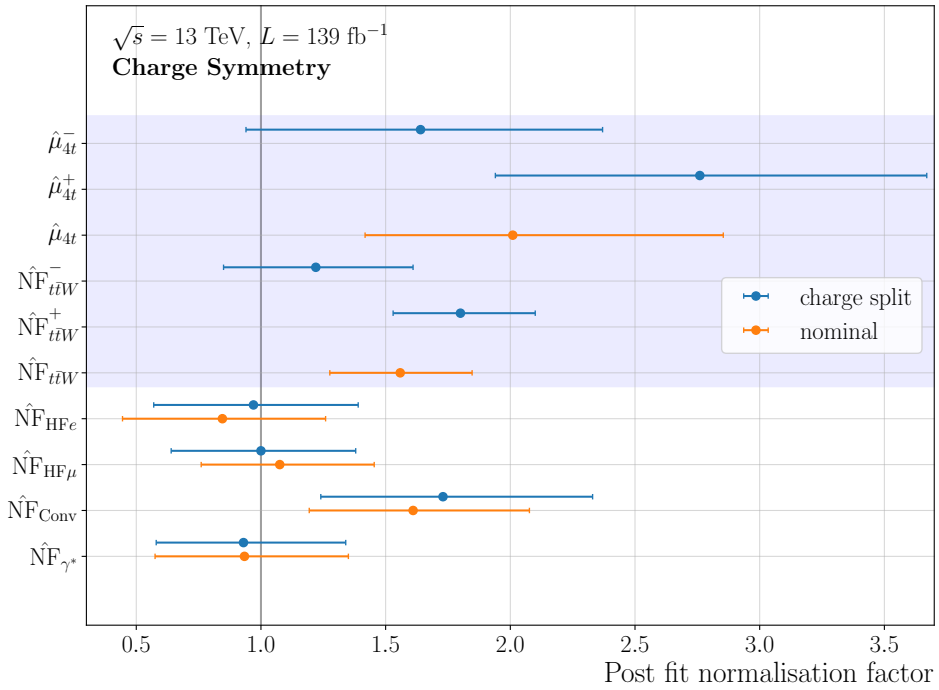


Figure 7.21: Post fit NFs for the nominal fit and a fit where the SR and the $t\bar{t}\,W$ CR are each split into two regions, where one only contains events with a positive sum of lepton charges and the other only those with a negative sum of lepton charges. The blue area highlights the NFs and signal strengths that do not have a direct correspondence to their counterpart in the other scenario. For example, no combined $\hat{\mu}_{4t}$ is obtained in the charge split scenario and no $\hat{\mu}_{4t}^{\pm}$ is used in the nominal fit setup.

7.8.5 Analysis Channels

Following that each analysis channel, i.e. $e^{\pm}e^{\pm}$,$e^{\pm}\mu^{\pm}$,$\mu^{\pm}\mu^{\pm}$,eee,$ee\mu$,$e\mu\mu$ and $\mu\mu\mu$, face different background compositions and difficulties, splitting the analysis between them can be particularly helpful to investigate the properties of the observed moderate excess.

Due to limited statistics, the test is not performed by splitting the SR into each of the channels, but rather it is split into a same sign dilepton and a multilepton SR. Two fit scenarios are performed using the split SRs. First, a single signal strength parameter $\mu_{4t}^{\text{split SR}}$ is used and second, two signal strength parameters are estimated, called $\mu_{4t}^{ss\ell\ell}$ and $\mu_{4t}^{\ell\ell\ell}$ and which correspond to the same sign dilepton and the multilepton SRs, respectively.

The NFs are mostly unchanged between the scenarios, so that only the obtained post fit signal strengths are given in Table 7.9. The values there are obtained with a next-to-final fit set up with the most important difference being that no uncertainty on the signal cross section is applied. As a result, the nominal $\hat{\mu}_{4t}$ given in that table does not correspond to the results in [1].
All obtained signal strengths agree very well with each other, so that the excess is found for all probed lepton multiplicities.

Table 7.9: Post fit signal strengths for the nominal fit setup $\hat{\mu}_{4t}$ and two scenarios where the SR is split into a same sign dilepton and a multilepton SR. This test has been performed with a fit set up that is very close to the final fit set up, where the major difference is that no uncertainty on the signal cross section is applied. Therefore, $\hat{\mu}_{4t}$ does not correspond to the value cited in the main results section.

$\hat{\mu}_{4t}$	$\hat{\mu}_{4t}^{\text{split SR}}$	$\hat{\mu}_{4t}^{ss\ell\ell}$	$\hat{\mu}_{4t}^{\ell\ell\ell}$
$2.1^{+0.6}_{-0.5}$	$1.9^{+0.6}_{-0.6}$	$1.9^{+0.7}_{-0.7}$	$1.9^{+0.7}_{-0.7}$

7.8.6 Data Taking Periods

Another test is performed where the SR is split into sub regions. In this case it is split into two data taking periods, namely the years 2015-2017 and 2018. The main motivation for a split in data taking periods is the difference of the average number of interactions per bunch crossing for different periods. This is particularly interesting as

these pile-up events can have a significant impact on particle reconstruction and thus especially on instrumental backgrounds, as broadly discussed in chapter 6.

Additionally, other, potentially unnoticed, effects of data taking conditions could play a role in producing an artificial excess in data.

The post fit NFs and $\hat{\mu}_{4t}$ are shown for the nominal full Run-II period, only for 2015-2017 and only for 2018 in Figure 7.22.

While the signal strength remains relatively stable, the NFs vary noticeably, although remaining compatible within their relatively large uncertainties. The nominal fit setup quoted here is the same as for the analysis channel split study and thus differs slightly from the final setup.

As a result, the fit remains stable for different data taking conditions, further suggesting that the observed excess is not an artefact from data taking and reconstruction.

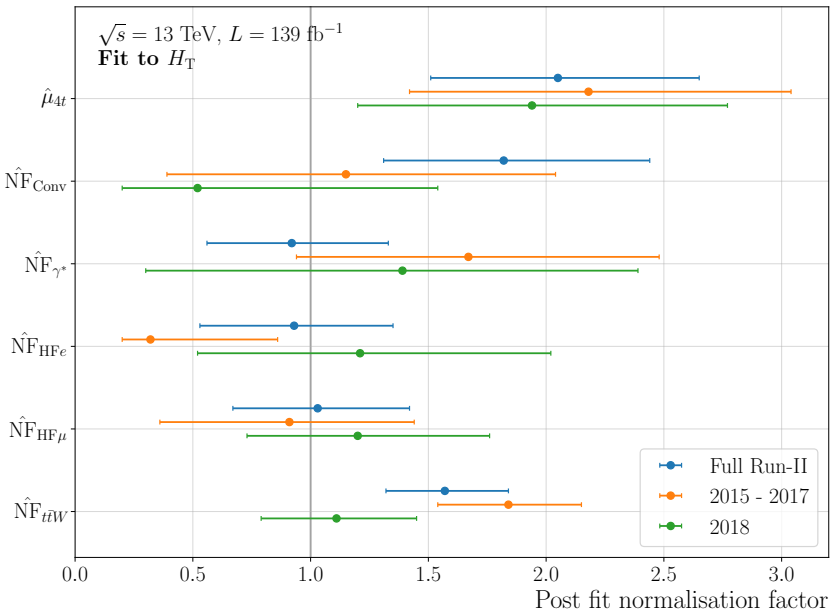


Figure 7.22: Post fit NFs and signal strength for the nominal fit setup using the full Run-II data set (blue), only using data recorded between 2015-2017 (orange) or only using data recorded in 2018 (green).

In summary, none of the studies conducted suggest that the observed excess is artificial. Instead, the opposite is the case, where most studies strongly confirm the likeness between the excess in data and the properties of the predicted signal[17].

7.9 Conclusion

The analysis presented in this chapter has been able to achieve the first evidence for the simultaneous production of four top quarks with an observed significance of $Z = 4.3$, while a significance of $Z = 2.4$ had been expected. As a consequence, the production cross section of this very rare process could be measured, which represents the culmination of many years of efforts within the ATLAS collaboration and is therefore a great success.

The obtained results and many parts of the analysis as it is presented in this dissertation are achieved in a group effort. The author of this document has been able to provide significant contributions in the areas of background estimations, software needs and most importantly in the thorough evaluation of the fit model for many of the ever evolving setups devised on the way to the final model.

Especially this last set of contributions was used in part to design the analysis strategy and in part to give confidence and insights into the behaviour of the fit. They represent the most important contributions exclusively provided by the author and are thus described in detail in the next chapter.

[17] With the obvious exception of the normalisation of the signal.

8 Design and Validation of the Fit Model

The final step of the analysis presented in the previous chapter is the global fit of a signal plus background model to data in a set of CRs and the SR.
Because of the importance of this fit on the outcome of the analysis, extensive tests are performed to better understand the behaviour and limitations of the fit, some of which are presented here in their final version. They are said to be in their final version as most of the studies have been performed multiple times (each for one iteration of the design of the fit model) in order to help build the final model. The final version of these studies are shown instead of the intermediate ones so that the results can be directly compared to the results in chapter 7.

For the purpose of the studies shown here, the results of the global fit basically depend on two components. First, the fit model in terms of the expected yields given by background estimations and smeared by systematic uncertainties and second, the data set used to adjust this model.

To examine the impact of any of the parameters of either of the two components, slight variations on one of the parameters are introduced and the fit reevaluated. The newly obtained results from the varied model or data set can then be compared to the nominal scenario, thus providing a measure of stability of the fit with respect to the varied parameter. This can then ultimately be used to design the strategy of the analysis prioritising on the most influential parameters or to provide an additional measure of confidence on the outcome of the fit.

This kind of fit stability tests is presented in this chapter, separated into whether the model or the data set is varied.

8.1 Fit Setups

Following the description of the global fit and the blinding procedure given in section 7.1 as well as the description of the template fit method used for the estimation of selected backgrounds (see subsection 7.4.2), multiple fit scenarios or configurations can be defined. Each fit scenario serves a specific purpose, albeit only the fit scenario called *Data fit* is used to obtain the final results of the analysis after unblinding, as presented in the previous chapter. The different scenarios are described in the following and an overview is given in Table 8.2. The main parameters that are changed in the different

scenarios are whether real data or pseudo data is used in the CRs and SRs, whether μ_{4t} is a free parameter in the fit, whether the SR is split into two sub regions and how the pseudo data is built.

At the blinded stage of the analysis, the SR can be split into two sub regions called CR *SR* low BDT and SR high BDT that follow the same selections as the original SR and additionally require $BDT < 0$ and $BDT \geq 0$, respectively. This split is done so that as much of the phase space used in the final unblinded fit as possible can be studied using real data while keeping the signal blinded.

The distinction between a CR and an SR is, in principle, irrelevant for the fit and it is only used for an easier communication. This means that this split should have no impact on the outcome of the fit if the same data set as in the inclusive (non-split) region is used in the split regions. This is however not the case here because of the smoothing procedure applied to the set of systematic uncertainties, which is applied once over the inclusive region for the non-split case and independently for the sub regions in the split case.

Two types of pseudo data sets are used in the following, the Plain Asimov pseudo data set (PA) and the Realistic Asimov pseudo data set (RA). The general concept of these data sets has already been introduced in section 7.1, so that a more precise definition is given here.

The PA is built by taking the sum of expectation values of all components of the pre fit S+B model (the expected event yield in each bin), suppressing the information of the uncertainty on the expectation values. In this case the pre fit model is taken directly from the MC simulation for all processes bar QMD, which is estimated with the data-driven technique. The uncertainty on the pseudo data event yield per bin is given for visualisation purposes by $\sqrt{N_i}$, where N_i is the yield in bin i. This follows the idea of real data, where each observation is assumed to follow the Poisson distribution.

The RA is built analogously, albeit with a *correction* to the background expectation by a fit to data in all CRs including CR *SR* low BDT. In this way, the background expectation is more realistic as the normalisation of the backgrounds determined by the template fit method are already adjusted to as much data as possible without unblinding the signal. The fit setup used to derive the model to generate the RA as well as all other setups are described in more detail in the following.

Proto Fit is the fit setup used to derive the RA. Thus, only the CRs including the CR *SR* low BDT are used to adjust the S+B model to real data. In this case the signal strength is fixed to one ($\mu_{4t} = 1$), as the CRs do not provide any information on the signal by design.
The set of parameters obtained in this fit are called $\hat{\alpha}_{\text{proto}}$. The name of the fit setup refers to the fact that this fit needs to be done *before* any fit setup that uses the RA. The yield for each bin of the RA is therefore the cumulative sum of all expected yields for each bin of the post Proto fit model given by the NFs in Table 8.1 and the NPs shown in Figure C.4. A visual representation of the BDT distribution in the SR of this data set is given in Figure 8.1b.

Table 8.1: Post fit NFs of $\hat{\alpha}_{\text{proto}}$, used to generate the Realistic Asimov pseudo data set, as obtained with the Proto fit setup. Additionally, the NFs obtained in the nominal fit as quoted in section 7.7 are shown for comparison.

Fit Setup	$\hat{\text{NF}}_{t\bar{t}W}$	$\hat{\text{NF}}_{\text{Conv}}$	$\hat{\text{NF}}_{\gamma^*}$	$\hat{\text{NF}}_{\text{HF}e}$	$\hat{\text{NF}}_{\text{HF}\mu}$
Proto	$1.60^{+0.29}_{-0.27}$	$1.63^{+0.47}_{-0.42}$	$0.92^{+0.42}_{-0.36}$	$0.83^{+0.41}_{-0.38}$	$1.06^{+0.33}_{-0.31}$
Nominal	$1.56^{+0.29}_{-0.28}$	$1.61^{+0.47}_{-0.42}$	$0.93^{+0.42}_{-0.36}$	$0.85^{+0.41}_{-0.40}$	$1.08^{+0.38}_{-0.32}$

The RA has been devised before unblinding, as that is its motivation, and the unblinding has shown that the name *realistic* is well deserved as the NFs and NPs do not differ significantly between the Proto and Nominal fit setup, with the natural exception of signal related parameters on which the Proto fit has no handle by construction. This can be seen in Table 8.1 for the NFs and by comparing Figure C.4 to Figure B.3 for the NPs.

Plain Asimov Fit uses the PA set as introduced in section 7.1 in all CRs and the SR to adjust the S+B model. The PA and pre fit model in the SR are shown in Figure 8.1a.

Using the PA allows to probe expected constraints on all parameters involved in the fit and enables an estimation of the expected significance. The signal strength μ_{4t} is a free parameter in this fit setup, albeit the post fit result yields the expected value of the signal model $\hat{\mu}_{4t} = \mu_{4t} = 1$, due to the nature of the pseudo data set.
Both a split and non-split version of this fit setup exists (where the split is described above), even though blinding is in principle not needed for this setup as only pseudo data is used. The split is nevertheless probed in order to be consistent with other fit setups

Realistic Asimov Fit uses the same setup as the Plain Asimov (Split) case, although the RA is used instead of the PA. The signal strength is in principle a free parameter in this fit setup, but just as in the Plain Asimov case, it is highly unlikely to obtain a post fit signal strength other than one as the signal strength is assumed to be one for the generation of the RA.

The Realistic Asimov (non-split) fit setup is used to determine the expected sensitivity of the analysis as it only uses real data implicitly as part of the background estimation in form of the RA, while otherwise representing the same structure as the final unblinded Data fit.
Both a split and non-split version of this fit setup exist, where, as in the Plain Asimov case, the split is not strictly needed and only kept for comparisons with other fit setups.

Realistic SR Fit uses the same RA as the Realistic Asimov fit in the SR, while real data is used in all CRs. As a consequence, the split and non split versions of this setup differ significantly. In the former case, the SR is split so that the RA is used in SR high BDT and data in CR *SR* low BDT, while the RA is used for both in the latter scenario.

The signal strength is a free parameter of the fit and this setup is used to probe the impact on $\hat{\alpha}$ (which includes $\hat{\mu}_{4t}$) of statistical fluctuations of real data in the CRs.
Additionally, the split version of this set up is used to probe possible outcomes of the fit after unblinding the analysis, before actually doing so. This is described in detail in subsection 8.3.3.
Finally, the setup is also used to cross-check the expected significance given by the Realistic Asimov fit.

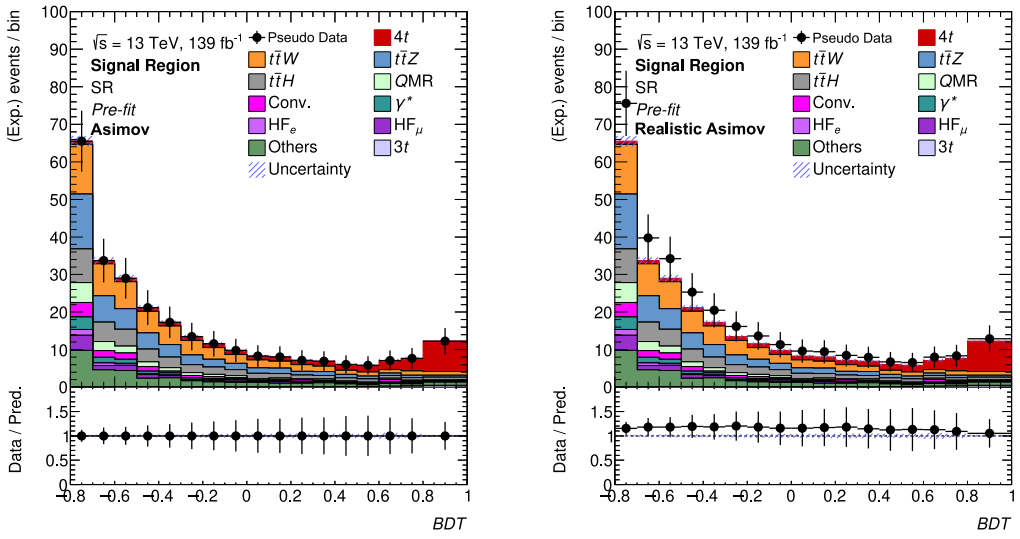
Data Fit uses real data in all regions and is thus fully unblinded. The regions CR *SR* low BDT and SR high BDT are removed as the nominal SR is used. The signal strength is a free parameter and this fit represents the final fit to data after unblinding.

This fit setup is used for obtaining the results on the signal strength μ_{4t}, the signal cross section $\sigma(pp \to t\bar{t}t\bar{t})$ and the observed significance as presented in chapter 7.

Both a split and non-split version of this fit setup exists, where this is not strictly needed as this setup can only be used after unblinding. The split case is used in order to study the impact of splitting the SR on the outcome of the fit.

Table 8.2: Overview of different fit setups and their purpose. P(R)A refers to the Plain (Realistic) Asimov pseudo data set and the SR split refers to the split of the nominal SR into an CR *SR* low BDT and an SR high BDT region.

Setup	(Pseudo) Data		SR Split	μ_{4t}	Purpose
	CRs	SR			
Plain Asimov	PA	PA	both	free	expected constraints on fit parameters
Proto	Data	-	yes	$= 1$	background estimation before unblinding
Realistic Asimov	RA	RA	both	free	expected significance
Realistic SR	Data	RA	both	free	evaluate impact of residuals of proto fit on μ_{4t} and exp. significance
Data	Data	Data	both	free	final results



(a) Plain Asimov (b) Realistic Asimov

Figure 8.1: Distributions of the signal extraction BDT in the SR for the pre fit model and the Plain (Realistic) Asimov pseudo data set.

8.2 Variations of the Model

In order to define an efficient analysis strategy it is paramount to understand and evaluate the impact of each of the parameters of the analysis (and fit) model on the outcome of the study. For this reason, a series of studies are performed where parameters of the fit, such as background normalisations or the size of the variation induced by uncertainties, are varied individually while all other parameters, the general fit setup and the data set remain unchanged. This *profiling* is then used to estimate the impact of the varied parameter on the final outcome.

8.2.1 Variation of the Background Model

Following the introduction above, the variation of the background model is first probed by a variation of the normalisation of the different background processes as well as of the uncertainty on said normalisation.

8.2.1.1 Background Normalisation

The variation of the nominal value of the normalisation of each background process is achieved by scaling the expected event yield N of a given background with a factor S_N.

The Plain Asimov fit setup is used for this study and no systematic uncertainties are considered to speed up the convergence of the fit. The impact on the outcome of the fit is evaluated using the expected significance $Z_{\text{stat.}}$.

The impact of the variations in form of different values of S_N on the significance $Z_{\text{stat.}}$ is shown in Figure 8.2. There, the impact is displayed as the relative change of $Z_{\text{stat.}}$ with respect to the nominal (unvaried) background model. Each colour represents the variation of a different background, where all other contributions remain unchanged. This artificial scaling of the normalisation of a component of the S+B model visualises the impact a certain component has on the final outcome. It therefore emulates the effect of, for example, an over or under estimation of the normalisation of a certain background, or the potential gain of a successful reduction of one of the background categories through an improved analysis.

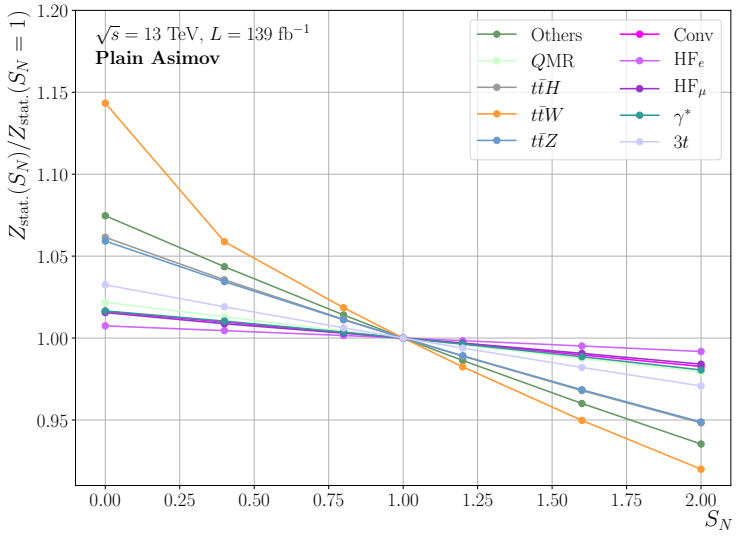


Figure 8.2: Relative change of the expected significance without systematic uncertainties in dependence of different scales applied to the normalisation of a certain background used to build the pre fit S+B model. Each colour represents a different background that is scaled. The Plain Asimov fit setup is used, so that the background model is not updated with the results of the template fit method.

Figure 8.4 shows the most influential backgrounds to be $t\bar{t}\,W$, Others, $t\bar{t}\,H$, $t\bar{t}\,Z$ and $3t$, followed by background contributions with a maximum impact of less than 3 % for the scanned range.

While interpreting the results it is important to note, that a relative change of one of the background processes with a large number of events in the SR corresponds to a larger absolute change in expected yields of the S+B model as would be the case for the variation of a smaller contribution.

It is therefore not surprising to see the largest variability of $Z_{\text{stat.}}$ to be associated to $t\bar{t}\,W$, Others, $t\bar{t}\,H$ and $t\bar{t}\,Z$, which are the largest, fourth, third and second largest contributors to the SR, respectively.

However, it is interesting to note that the Others category shows the second largest variability while it only provides the fourth largest contribution to the SR.

This is due to the fact that, in contrast to the $t\bar{t}V$ processes, this group of processes has a non-negligible event yield in the highest bin of the signal extraction BDT, which

is shown most clearly in Figure 8.3[1].

The expected yield for $3t$ events in the SR is the smallest of all yields and approximately an order of magnitude lower than the expected yield of the signal with 3 ± 3 with respect to 30 ± 12 at the pre fit level, respectively. Nevertheless, it ranks as the fifth most impactful background source. This has already been mentioned in section 7.7 and is explained by the high similarity between the shapes of the distribution of the BDT for the two processes shown in Figure 8.3. For this reason and because $3t$ is a process that has not yet been observed, which means complete reliance on theoretical predictions, the $3t$ production is removed from the "Others" category of which it has been a part for all previous fit setups. This represents a noticeable change of the fit strategy as a result of the study described here.

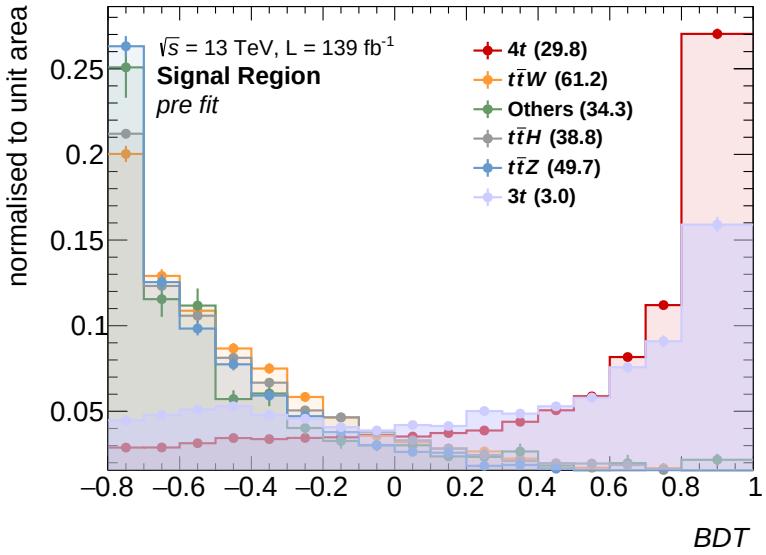


Figure 8.3: Comparison of the shape of the signal extraction BDT distribution in the SR for the signal and major backgrounds from simulation. The uncertainties shown in form of error bars only include the uncertainty corresponding to the sample size of the simulation. The absolute number of events in the SR is shown for each background in parentheses in the legend.

[1] This would indeed warrant for additional studies to better understand this category, but since it groups together a list of processes that were added during the finalising steps of the analysis this has not been possible within the time frame of this thesis.

8.2.1.2 Uncertainty on the Background Normalisation

In order to further study the impact of each of the backgrounds on the analysis, not only the nominal value of the normalisation of each component is of interest, but rather its uncertainty as well. This uncertainty is of interest as it is directly related to the precision of the background estimation techniques used in this analysis.

To emulate a variation of the normalisation uncertainty, a single flat systematic uncertainty on the normalisation of the corresponding background ΔN is introduced to the otherwise nominal Plain Asimov fit setup without systematic uncertainties. A range of different values for the size of the uncertainty, that is the up and down variation, is probed. The sizes range between 0 % and 200 %, where the variations are symmetric for values up to 100 % and become asymmetric for larger variations. This is due to the fact that the downward variations are not allowed to go below $N - 100\,\% = 0$ from the nominal. This rather large range of probed values is chosen seeing that uncertainties of up to 300 % are used in the main analysis, albeit only for a very specific region of the phase space.

The results of this scan are shown in Figure 8.4a, where again each colour represents the variation of the uncertainty on the normalisation of a specific contribution to the S+B model and the relative change of Z is with respect to the case where no additional systematic is introduced $\Delta N = 0\,\%$. No other uncertainties besides the scanned normalisation uncertainty are used for the determination of Z. The blued-out area represents the range where the variation due to the uncertainty becomes asymmetric.

The strongest impact is given by $3t$, which is particularly problematic since, as has been pointed out above, this process has not yet been measured. Thus, any uncertainty on the normalisation can only come from the theoretical prediction and not from measurements. The other contributions that have a measurable effect come from $t\bar{t}H$, Others and QMD processes.

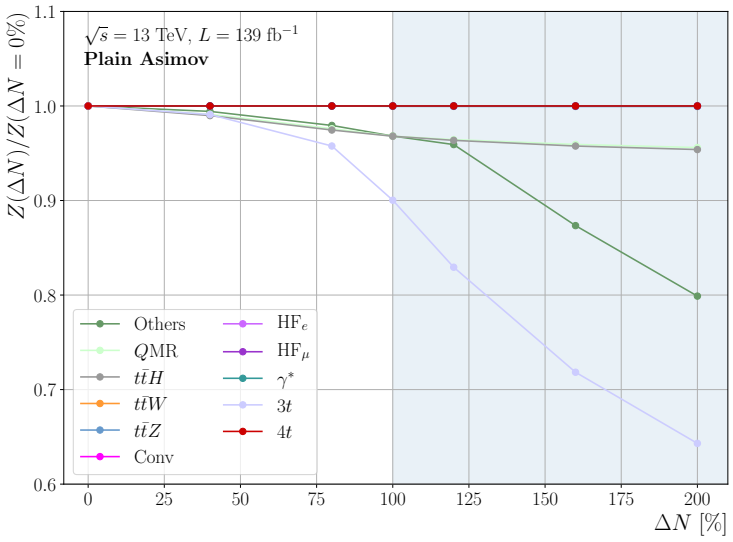
In order to correctly interpret this result it is again important to note that the absolute variation is much larger for contributions with a large normalisation, as is for example the case for $t\bar{t}W$. On the other hand, this is by far not the only factor seeing that the process with the smallest event yield ($3t$) results in the largest impact.

Furthermore, the introduced uncertainty is likely to play a smaller role for the background contributions estimated with the template fit method seeing that in their case

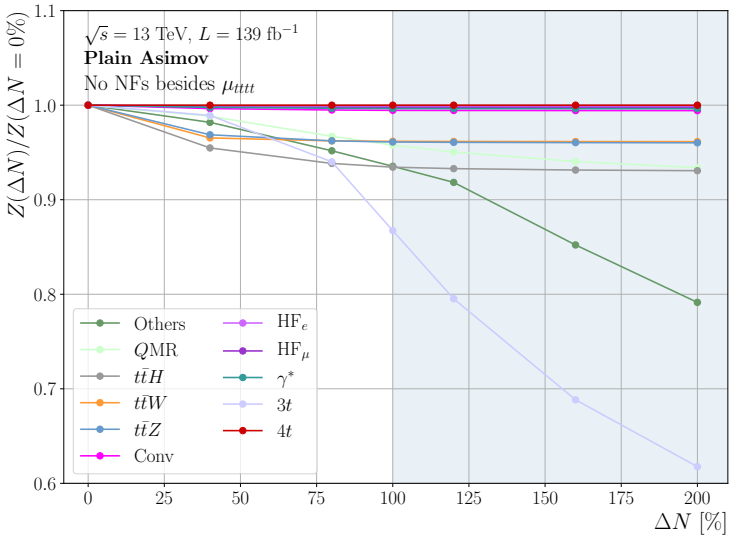
the normalisation and the corresponding uncertainty is given by the free floating NFs, so that the additional uncertainty will be mostly ignored by the fit. This effect is showcased in Figure 8.4b, where the same fit setup as for Figure 8.4a is used albeit without any of the NFs. Doing so severely alters the fit, resulting in a different ranking, where the uncertainty on $t\bar{t}H$, $t\bar{t}Z$ and $t\bar{t}W$ play a major role.

In conclusion, this result lead to a focus on $t\bar{t}W$ and demonstrated the importance of estimating the normalisation of $t\bar{t}W$ with the data assisted template fit method. The argument is strongly enhanced by the fact that tensions between data and prediction have been found for this process in chapter 7.

Finally, the normalisation and corresponding uncertainty on the different components of the background model play an important role in the definition of the analysis strategy. In order to better understand the full fit model, however, it is also fundamental to study the impact of all other NPs that make up the fit model. This is described in the following.



(a) nominal



(b) no NFs

Figure 8.4: Relative change of the expected sensitivity Z in dependence of the size of an uncertainty on the normalisation ΔN for different background expectations. Different colours represent the uncertainties pertaining to different background sources. The Plain Asimov fit setup is used and no systematic uncertainties except for the additional normalisation uncertainty are applied to determine Z. The results are shown for the nominal fit setup up (top) and a fit setup where all NFs are removed so that μ_{4t} is the only unconstrained parameter of the fit (bottom).

8.2.2 Variation of Systematic Uncertainties

The full systematic uncertainty model of this analysis has been described in section 7.6, where the sheer number of uncertainties makes it clear that not all of these uncertainties can be studied in detail for the peculiarities of this analysis. Moreover, a significant number of these uncertainties are assigned *ad hoc*, following reasonable assumptions, but without a dedicated measurement. For these reasons, it is important to establish the overall impact of each systematic uncertainty on the results of this analysis and to probe the impact of varying the magnitude of an uncertainty on said results. Studying this allows to rank the different systematic uncertainties, so as to focus more in depth studies on those uncertainties that have a larger impact or on those with a high variability with regards to the assigned *ad hoc* size of the uncertainty.

Figure 8.5 and Figure 8.6 show the relative change of the expected significance Z with respect to the nominal for the systematic uncertainty model where one of the systematics uncertainties is scaled by a factor S_{syst}. This scale is applied to the difference of the up (down) variation of the systematic uncertainty to the nominal so that the scaled up (down) variation follows $\text{var}_{\text{scaled}} = \text{nominal} + S_{\text{syst}} \times (\text{var} - \text{nominal})$. The change in significance is shown in dependence on S_{syst}, where a range of $S_{\text{syst}} \in [0, 2]$ is probed.
For better visibility, only the 18 most influential systematic uncertainties[2] are shown, where each is represented in a different colour. The study is performed with a previous fit configuration, where most notably the `tttt cross-section` uncertainty is not applied, so that it does not appear in the list. The connections (lines) between the data points are piecewise linear interpolations. Figure 8.5 shows the results for the Plain Asimov fit set up, while for Figure 8.6 the Data fit setup is used[3].

Following Figure 8.6, the two most volatile systematic uncertainties are `ttW Syst` $\geq 8j$ and `ttt Cross-Section`, where a removal ($S_{\text{syst}} = 0$) of the former increases the expected significance by almost 20 %, while enlarging it has no measurable effect. For the latter, the exact opposite is the case, where a doubling of the assigned uncertainty would result in a loss of significance of approximately 12 %. It is important to note that both uncertainties are already considerably large in the nominal scenario, 300 % and 100 %, respectively, so that the application of a multiplicative scale represents a larger

[2] The technique used to determine this set of uncertainties slightly differs from the technique used in this section and it is shown in Figure B.2.
[3] The studies were first performed during the blinded stage of the analysis, where the Realistic Asimov fit setup was used. The differences in the outcome between the two setups is negligible.

variation of the uncertainty than it does for systematic uncertainties that are initially smaller. As for the normalisation uncertainties introduced above, uncertainties above 100 % become asymmetric as the down variation is truncated at zero.

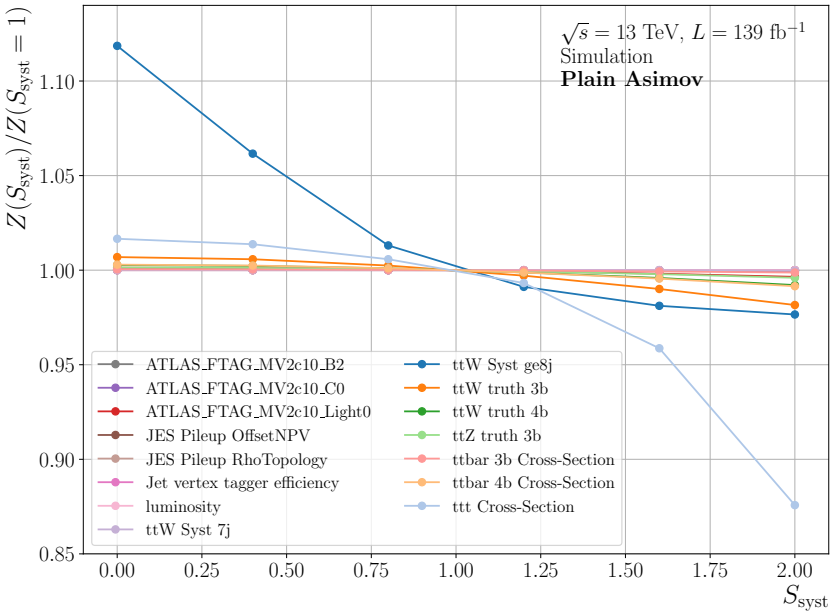


Figure 8.5: Relative change of the expected significance Z in terms of the scale applied to a systematic uncertainty S_{syst}. The change is relative with respect to the nominal, i.e. $S_{\text{syst}} = 1$, case. Different colours represent the uncertainty that is varied, where only one uncertainty is varied at a time. The Plain Asimov fit setup is used, albeit without the `tttt cross-section` uncertainty.

In order to condense the results of Figure 8.6 into a measure of the dependence of Z on S_{syst} for any given systematic uncertainty, two FOMs are defined. First, the average of the distance between the obtained ratio of $\frac{Z_S}{Z_{(S=1)}}$ to one, meaning $\left\langle \left| 1 - \frac{Z_S}{Z_{(S=1)}} \right| \right\rangle$, which condenses the impact over the whole range of S_{syst}. Second, the maximum of the same argument, generally returning the result of $S_{\text{syst}} = 0$ or $S_{\text{syst}} = 2$. The aggregated values are listed in Table 8.3 for the 18 most influential uncertainties. As mentioned above, the initial size of the variation plays an important role for the obtained measure of impact, as the scanned variations are obtained in a multiplication to S_{syst}. Therefore, the initial variations are also listed in Table 8.3. All uncertainties listed in this table are two-sided.

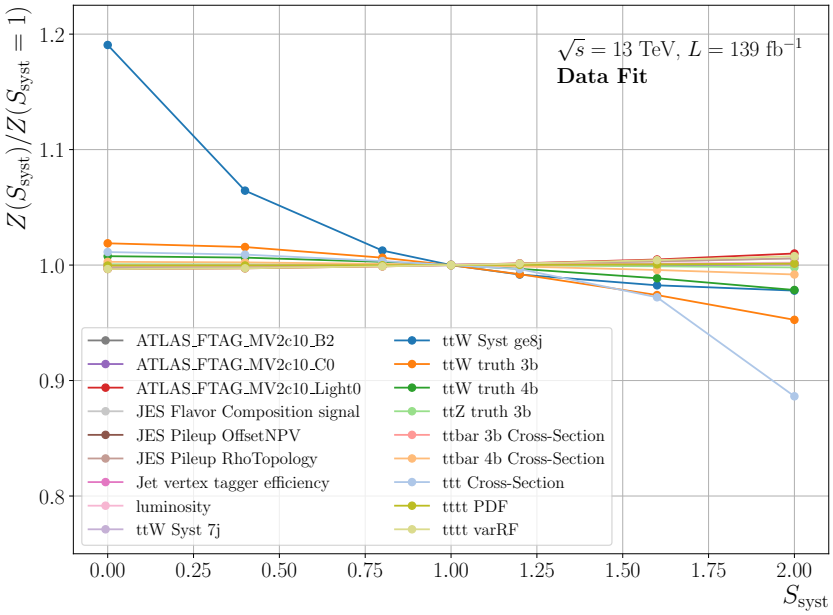
Studying the impact of the different components of the systematic uncertainty model in this way, allowed for the identification of critical uncertainties for a prioritisation

and to corroborate confidence in the obtained results. Moreover, these studies were used to obtain a better understanding of the impact of the rather arbitrary size of variation set for the *ad hoc* uncertainties.

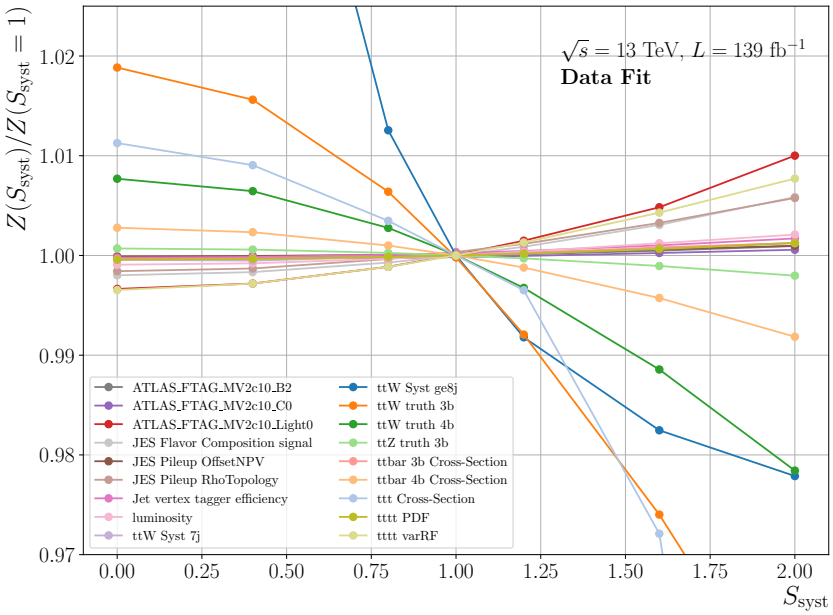
Table 8.3: List of the 18 systematic uncertainty with the largest impact on the observed significance Z. The fit setup used for this study does not include the uncertainty on the normalisation of the signal prediction.

	Uncertainty	Initial Variation	$\max\left\lvert 1-\frac{Z_S}{Z_{(S=1)}}\right\rvert$	$\left\langle\left\lvert 1-\frac{Z_S}{Z_{(S=1)}}\right\rvert\right\rangle$
0	ttW Syst $\geq 8j$	300 % if $N_j \geq 8$	0.191	0.045
1	ttt Cross-Section	90 %	0.114	0.024
2	ttW truth 3b	50 % if $N_{b,\text{truth}} = 3$	0.047	0.017
3	ttW truth 4b	50 % if $N_{b,\text{truth}} \geq 4$	0.022	0.008
4	ATLAS_FTAG_MV2c10_Light0	Weight[1]	0.010	0.003
5	ttbar 4b Cross-Section	30 % if $N_b \geq 4$	0.008	0.003
6	tttt varRF	Weight[1]	0.008	0.003
7	JES Flavor Composition signal	Weight[1]	0.006	0.002
8	JES Pileup RhoTopology	Weight[1]	0.006	0.002
9	luminosity	1.7 %[2]	0.002	< 0.001
10	ttZ truth 3b	50 % if $N_{b,\text{truth}} = 3$	0.002	< 0.001
11	Jet vertex tagger efficiency	Weight[1]	0.002	< 0.001
12	ttbar 3b Cross-Section	30 % if $N_b = 3$	0.001	< 0.001
13	ttW Syst $7j$	125 % if $N_j = 7$	0.001	< 0.001
14	tttt PDF	1 %	0.001	< 0.001
15	JES Pileup OffsetNPV	Weight[1]	< 0.001	< 0.001
16	ATLAS_FTAG_MV2c10_B2	Weight[1]	< 0.001	< 0.001
17	ATLAS_FTAG_MV2c10_C0	Weight[1]	< 0.001	< 0.001

[a] The size of the variation is determined in an auxiliary measurement and provided here as a per event weight so that an overall size of the variation cannot be given.
[b] And none otherwise.



(a) full y-axis range



(b) reduced y-axis range

Figure 8.6: Relative change of the expected significance Z in terms of the scale applied to a systematic uncertainty S_{syst}. The change is relative with respect to the nominal, i.e. $S_{\text{syst}} = 1$, case. Different colours represent the uncertainty that is varied, where only one uncertainty is varied at a time. The Data fit setup is used, albeit without the `tttt cross-section` uncertainty.

8.3 Variations of Data

The studies in the previous section concentrate on the impact of changes to the fit model, be it a change in the estimated background contributions or changes to the systematic uncertainty model. In contrast, the studies presented in the following use a fixed model and variations are done on the composition of the pseudo data set used in the fit.

As it is the goal of the fit to determine background and signal contributions in data, multiple studies have to be performed to probe its (hypothetical) ability to do so. This is particularly true as the fit has a very large number of degrees of freedom in the form of the signal strength, NFs and NPs.

First, it needs to be shown that the fit is able to correctly recover the composition of the data set. For this, artificial pseudo data sets are used, where the composition can be arbitrarily varied. This change of the contribution of a certain process to the data set is referred to as injection.
The fit is then realised for a set of injected pseudo data sets (one at a time) and the post fit model of these fits is compared to the correct (injected) model used to build the pseudo data set.

Second, an important assumption of estimating certain background contributions in dedicated regions is that the same estimation should also be valid in the SR. This is not necessarily correct, so that studies are performed where different data sets with a controlled discrepancy in their composition are used for the CRs and the SR[4].

The fit itself will adjust in part to these discrepancies as the NFs used for the estimation of backgrounds are not limited to the CRs but instead are fitted simultaneously in the SR as well. This means that the obtained post fit NFs represent a *compromise* between the NFs that would be obtained in each region individually. Studying the effect of different discrepancies between the regions is therefore also a handle to visualise the importance of each region on a certain NF or the signal strength.

Third, a discrepancy as above can also be introduced between the data sets used in the unblinded region, i.e. all CRs and the lower part of the BDT distribution in the SR,

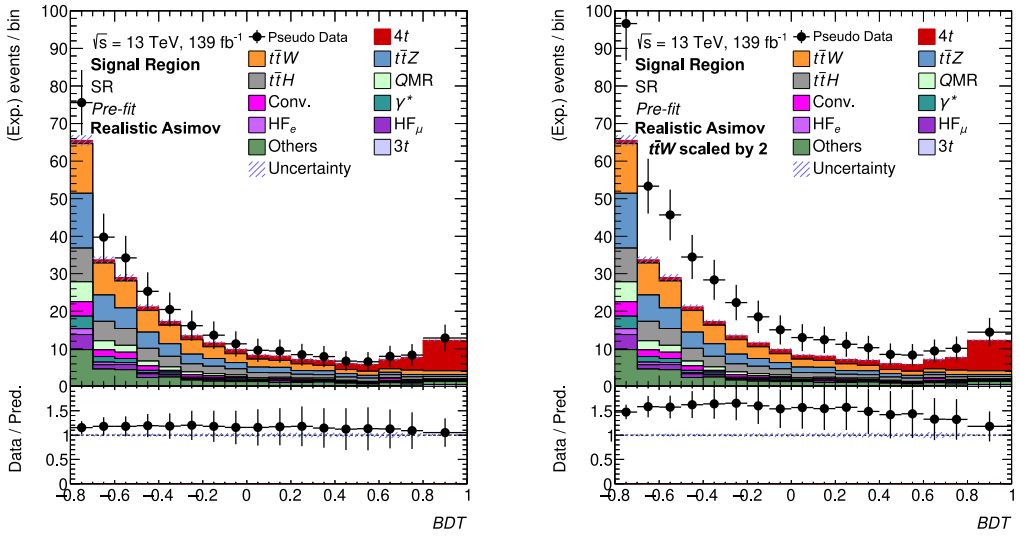
[4] One approach to mitigate the impact of this assumption is by devising extrapolation uncertainties in auxiliary measurements that can be included in the uncertainty model of the fit.

and the blinded region, i.e. the upper part of the distribution in the SR. This setup is used to systematically probe possible scenarios for the unblinding of the analysis based on the data that is already available before unblinding and represents an important step for the decision to start the unblinding procedure.

8.3.1 Implementation

As introduced above, the studies in this section rely on differences in pseudo data sets. The nominal pseudo data set used here is the RA and differences are introduced by individually scaling one of the parameters of $\hat{\alpha}_{\text{proto}}$ used to build the RA with a scaling factor S_{PF}. In order to disentangle the post *Proto fit* NFs used to build the pseudo data set from the post *fit* NFs obtained in the fit to this pseudo data set, the former are renamed Proto Factors (PFs) in the following.

An example of the change in the pseudo data set is given in Figure 8.7, where the nominal case of the realistic Asimov pseudo data set is shown in Figure 8.7a and the case where the contribution of $t\bar{t}\,W$, $\text{PF}_{t\bar{t}\,W}$, is scaled by 2 is given in Figure 8.7b.



(a) Realistic Asimov

(b) $t\bar{t}\,W$ scaled by 2

Figure 8.7: Distributions of the signal extraction BDT in the SR for the pre fit model, the realistic Asimov pseudo data set and the realistic Asimov pseudo data set where the contribution of $t\bar{t}\,W$ is scaled by two. This figure serves as an illustration on how a change in the construction of the pseudo data set is achieved.

8.3.2 Linearity of the response

In order to test if the fit is able to correctly recover the composition of the (pseudo) data set, different fits with the Realistic Asimov fit setup are performed where the RA is altered by scaling one of the PFs of $\hat{\alpha}_{\text{proto}}(\mu_{4t} = 1)$ with S_{PF}. The scaled PF used to build the altered data set is therefore given by $\text{PF}_{\text{scaled}} = \text{PF}_{\text{proto}} \times S_{\text{PF}}$.
Subsequently, the outcome of each of these fits in terms of post fit NFs is compared to the (scaled) PFs used to build the pseudo data set. The results of this are shown in Figure 8.8, where the relative change of the post fit NF associated to the same process as the scaled PF is shown in dependence of the value of the applied scale. Different colours represent different processes and the uncertainties represent the propagated uncertainties of the post fit NFs.
The fit can be considered to correctly assess the contribution of a certain process in the pseudo data set, when the response (in terms of post fit NFs) to changes in the scaled PFs is exactly proportional. Therefore, also the bisecting line (grey, dashed) is shown in Figure 8.8.

As a result, the fit exhibits a very linear response with only very small deviations from linearity for the extreme cases of $S_{\text{PF}} = 0$ and is thus able to correctly estimate the composition of the pseudo data set.

8.3.3 Signal Region Injections

In order to study the impact of potential discrepancies between the data composition in the CRs and the SR, the Realistic SR fit setup is used. In this way, the real data composition is present in the CRs and discrepancies are introduced by using pseudo data sets with different compositions (injections) in the SR. For the sake of this study it is assumed that the post fit model of the Proto fit setup, that is $\hat{\alpha}_{\text{proto}}$, correctly estimates the composition of data in the CRs. Consequently, and analogously to the studies above, the pseudo data used in the signal region is the RA where one of the PFs used to build the pseudo data set is scaled.

The signal region injections are separated into background injections and signal injections as their behaviour and interpretation is considerably different, seeing that the former contributions are constrained by the CRs while the latter is not.

8.3.3.1 Background Injections without $t\bar{t}W$ uncertainties

The relative evolution of each of the post fit NFs with respect to the scale applied to one of the PFs used to build the pseudo data set is shown in Figure 8.9. There,

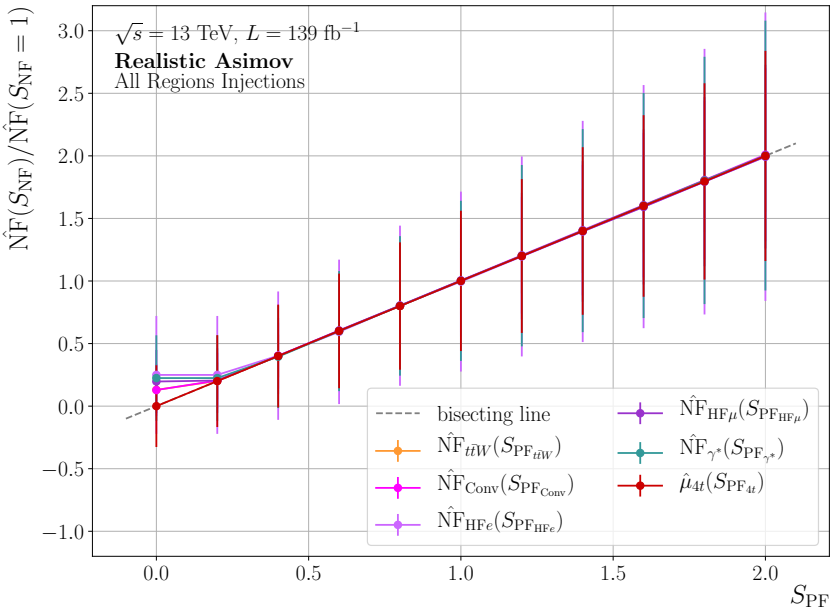


Figure 8.8: Linearity of the response of the fit, i.e. the relative change of the post fit NFs associated to the scaled PFs used to build the pseudo data set in dependence of the applied scale S_{PF}. Different colours represent different processes and the uncertainties represent the propagated uncertainty from the post fit NF as obtained in the fit. The Realistic Asimov fit set up is used. In order to visualise any deviation from linearity, the bisecting line is shown (grey, dashed).

each figure represents a different PF that is scaled, while each colour represents the evolution of a specific post fit NF. The uncertainties on the post fit NFs are not shown as they are large in comparison to any variation of the nominal value induced by the scaling and would thus overburden the figure. For the fit setup shown here, the `ttW Syst` $7j$ and the `ttW Syst` $\geq 8j$ uncertainties are not applied, as their introduction is precisely the result of this study.

Scaling the PFs has only a very limited, i.e. less than 3 %, impact on the post fit NFs with the notable exception of scaling $\mathrm{PF}_{t\bar{t}W}$. There, the larger impact is to some extent caused by the fact that it is a large PF to begin with, $\mathrm{PF}_{t\bar{t}W} \approx 1.6$, so that scaling it has a larger effect by construction. On the other hand, this is also the case for $\mathrm{PF}_{\mathrm{Conv}}$, where the observed effect is not as strong.

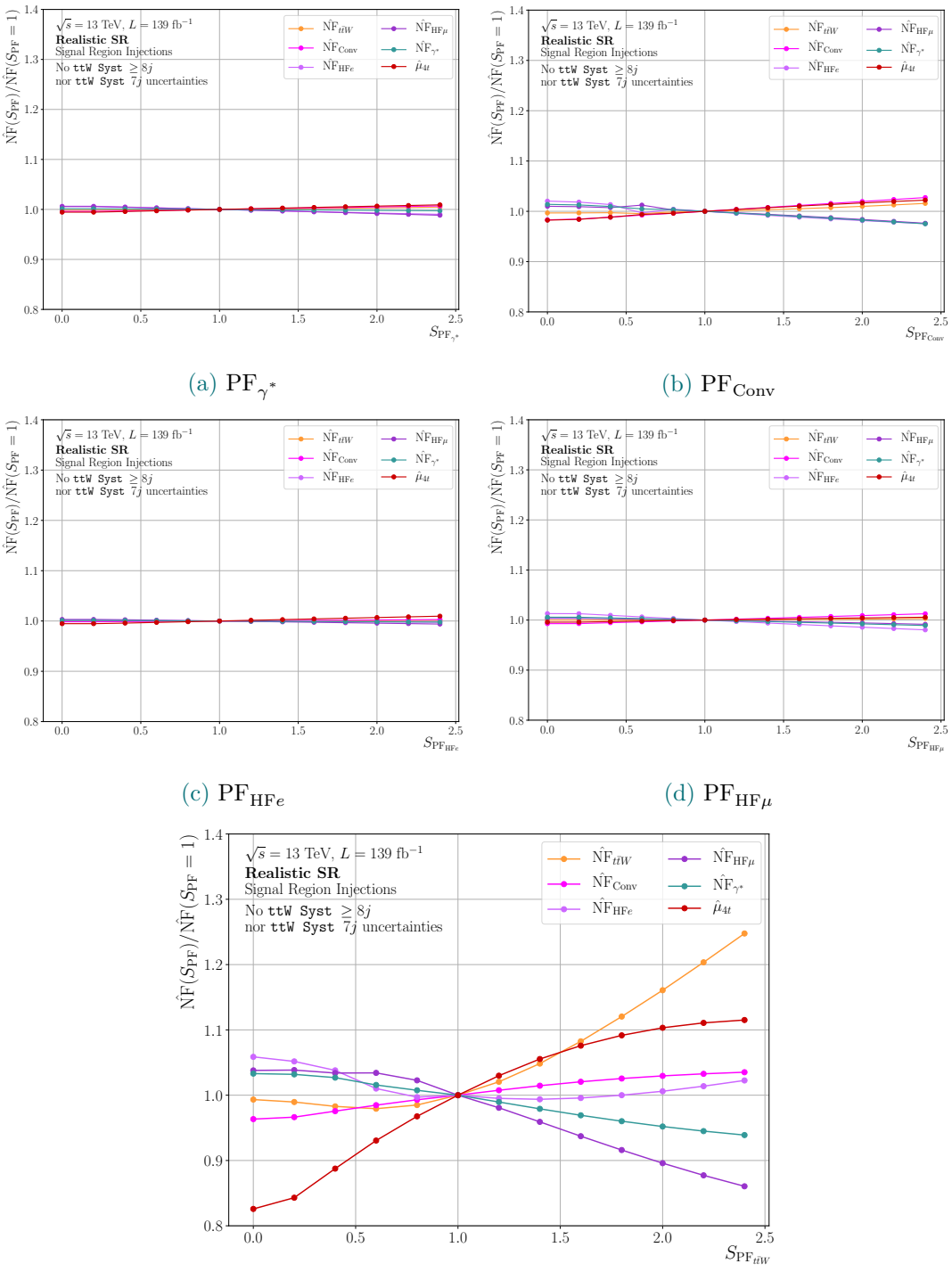
It is therefore likely that the large impact on $\hat{\mathrm{NF}}_{t\bar{t}W}$ is also due to the fact that $\mathrm{NF}_{t\bar{t}W}$ is not only constrained by the $t\bar{t}\,W$ CR, but rather also by the SR. This is particularly true for the lower part of the BDT distribution in the SR, where $t\bar{t}\,W$ provides the largest contribution.

In addition to $\hat{\mathrm{NF}}_{t\bar{t}W}$, also $\hat{\mu}_{4t}$ is strongly affected by a change in $\mathrm{PF}_{t\bar{t}W}$, where the obtained value for $\hat{\mu}_{4t}$ can differ up to 20 % from nominal for the scanned range of $S_{\mathrm{PF}_{t\bar{t}W}}$. This can be partly explained by the fact that μ_{4t} is the parameter with the weakest constraint by other regions, so that it will be generally preferred by the fit to accommodate the excess in the pseudo data caused by the injected contribution of $t\bar{t}\,W$.

This behaviour of the fit, however, is highly undesirable in a real case scenario as it shows that the fit could potentially associate a discrepancy of $t\bar{t}\,W$ contributions between the CRs and the SR to the signal.

Figure 8.10 shows the impact of scaling any of the processes on the obtained *hybrid* significance. The significance is denoted a hybrid as it is computed using real data in the CRs and the injected pseudo data set in the SR. The figure shows that even though $\hat{\mu}_{4t}$ could be up to 15 % larger for a $t\bar{t}\,W$ contribution that is 2.5 times larger in the SR than in the CRs (see Figure 8.9), the obtained significance is not. Instead, the significance is lower for most values of $S_{\mathrm{PF}_{t\bar{t}W}}$ than for the nominal case. This is even more striking as it means that the significance diminishes with a lower background contribution (as given by $S_{\mathrm{PF}_{t\bar{t}W}}$). This behaviour can be explained again following Figure 8.9, where for lower contributions of $t\bar{t}\,W$ to the pseudo data set, the obtained $\hat{\mathrm{NF}}_{t\bar{t}W}$ stays the same, while $\hat{\mu}_{4t}$ decreases by almost 20 %. However, the observed behaviour of the significance therefore represents a conservative result regardless of the discrepancy in terms of $t\bar{t}\,W$.

Nevertheless, seeing that the value of $\hat{\mu}_{4t}$ is used to determine the observed signal cross section, any dependence on an incorrect extrapolation of $\mathrm{NF}_{t\bar{t}W}$ to the SR should be mitigated. This becomes even more clear when investigating possible discrepancies between the data composition of the blinded and unblinded areas of the phase space which is described in the following. Additionally, the approach to mitigate the above mentioned dependence, used in the nominal fit setup, is introduced and its effect is presented in subsubsection 8.3.3.2 and subsubsection 8.3.3.3.



(a) PF_{γ^*}

(b) $\mathrm{PF}_{\mathrm{Conv}}$

(c) $\mathrm{PF}_{\mathrm{HF}e}$

(d) $\mathrm{PF}_{\mathrm{HF}\mu}$

(e) $\mathrm{PF}_{t\bar{t}W}$

Figure 8.9: Relative change of the post fit NFs associated to the scaled PFs used to build the pseudo data set in dependence on the applied scale S_{PF}. Different colours represent the different post fit NFs and each figure corresponds to a single PF that is scaled. The Realistic SR fit setup is used, so that real data is used in the CRs and the scaled realistic Asimov pseudo data is used in the SR. The `ttW Syst (7j) ≥8j` uncertainties are not used in this fit in order to illustrate their impact when compared to Figure 8.14. The uncertainties on the post fit NFs are not shown for visualisation purposes as they are considerably larger than any variation seen in the figures.

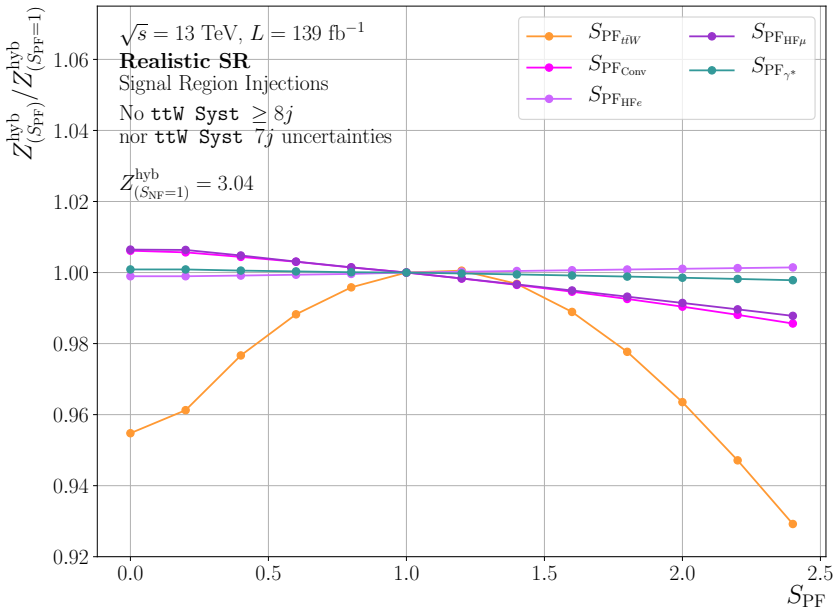


Figure 8.10: Relative change of the obtained hybrid significance in dependence of the value of the applied scale S_{PF}. Different colours represent different processes to be scaled while generating the pseudo data set used in the fit. The Realistic SR fit setup is used, so that real data is used in the CRs and the scaled realistic Asimov pseudo data is used in the SR. The `ttW Syst (7j)` $\geq$`8j` uncertainties are not used in this fit.

8.3.3.2 Intermezzo: Unblinding

As mentioned above, the studies shown here can also be used to scan the stability of the fit at the pre unblinding stage with respect to the unblinding. These stability tests are key for the decision to unblind the analysis as this step is not easily retractable and the fit model is generally frozen.

For testing different possible outcomes of the fit after unblinding using only information available before unblinding, a similar setup to the above study is used. The main difference to the above setup is that the SR is split into the CR *SR* low BDT and SR high BDT regions, which was not the case before. Real data is then used in the former, while the scaled pseudo data is only used in the latter. In this way, the largest possible amount of data is unblinded without unblinding the signal, so that any injection to the pseudo data set emulates the difference of the data composition between the unblinded and blinded regions of the phase space.

The impact on the post fit NFs and $\hat{\mu}_{4t}$ of scaling the $t\bar{t}\,W$ contribution in the pseudo data set in the SR high BDT is displayed in Figure 8.11. Since in this case also the low part of the BDT distribution in the SR serves as a CR with real data, the constraints on the NFs are even stronger than in the studies presented previously. For this reason, and because the signal represents a much larger fraction of the expected yield in the SR high BDT than in the nominal SR, any excess in pseudo data is almost exclusively taken up by $\hat{\mu}_{4t}$. A doubling of the $t\bar{t}\,W$ contribution in data in the SR high BDT would result in an increase of $\hat{\mu}_{4t}$ of almost 50 %.

It is important to note that the chosen range of $S_{PF_{t\bar{t}\,W}}$ is arbitrary and it is a priori not clear how large possible differences between the compositions of the data in the blinded and unblinded regions are likely to be. The studies shown here therefore present scenarios and an actual estimation of the difference would be needed to draw a definite conclusion on the impact[5].

In order to nevertheless mitigate the impact of any such discrepancy in $t\bar{t}\,W$, an additional set of systematic uncertainties is introduced. The uncertainties are the aforementioned `ttW Syst` $7j$ and the `ttW Syst` $\geq 8j$ uncertainties that are used in the nominal final fit setup and were left out in the above studies for illustration purposes.

[5] In this case, however, this can only be achieved after unblinding and it would thus be easier to simply define individual NFs for the unblinded and previously blinded regions and test their compatibility.

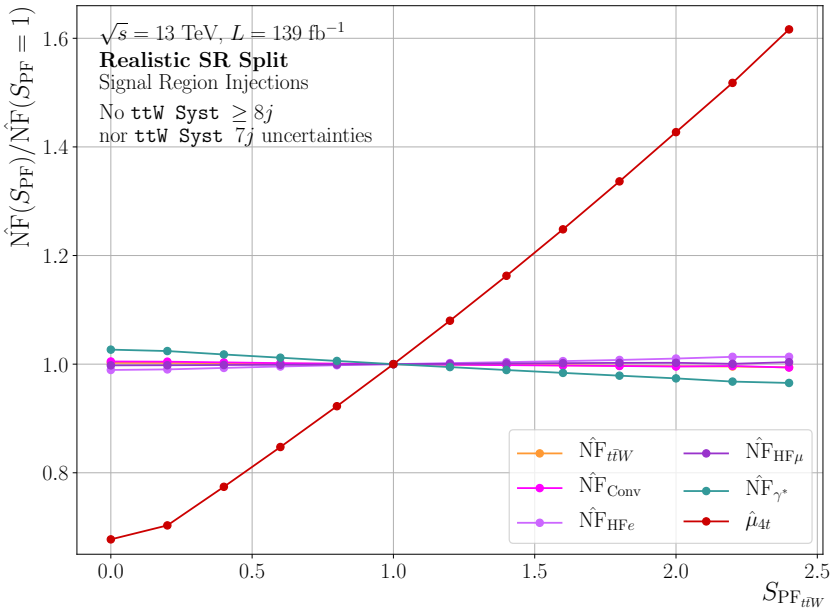


Figure 8.11: Relative change of the post fit NFs obtained in a fit where the contribution of $t\bar{t}\,W$ to the pseudo data set is scaled by $S_{\mathrm{PF}_{t\bar{t}W}}$. Different colours represent the different post fit NFs. The Realistic SR Split fit setup is used in order to emulate the response of the fit during the unblinding step of the analysis. The `ttW Syst (7j) ≥8j` uncertainties are not used in this fit and the uncertainties on the post fit NFs are not shown for a clearer visualisation. The orange curve is hidden behind other curves and is essentially constant at 1.

The uncertainties represent a 125 % (300 %) uncertainty on the normalisation of the event yield from $t\bar{t}\,W$ events with exactly seven jets ($\geq 8j$) and are based on a disagreement between data and expectation in the VR $t\bar{t}\,W$ region. The jet multiplicity in that region is shown in Figure 8.12, where the discrepancy can be seen in the two highest bins. The y-axis in this figure shows the difference of the yields for events where the sum of lepton charges is positive with respect to the yield of events where the sum is negative. This difference is expected to be zero for charge symmetric processes while any asymmetry towards positively charged leptons, as is the case for $t\bar{t}\,W$, would yield a positive result.

Introducing this additional set of uncertainties significantly reduces the impact of any discrepancy of the $t\bar{t}\,W$ contributions in the blinded and unblinded areas of the phase space, which is shown in Figure 8.13. There, the hybrid significance Z^{hyb} is shown with respect to the scale applied to the $\mathrm{PF}_{t\bar{t}W}$ for the fit setup without the additional $t\bar{t}\,W$ uncertainties (orange) and the nominal setup (blue).

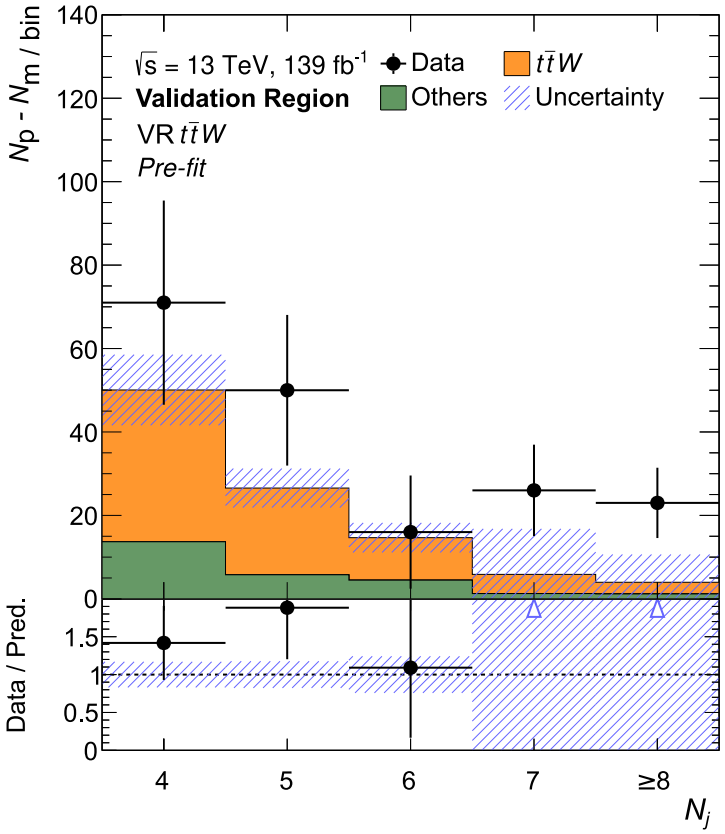


Figure 8.12: Jet multiplicity in the VR $t\bar{t}W$ region. The y-axis shows the difference of the per-bin yield for events with a positive sum of lepton charges to the yield with a negative sum. For charge symmetric processes, this difference should cancel out, while charge asymmetric processes such as $t\bar{t}W$ should result in a positive difference. The discrepancy between data and expectation for seven and eight or more jets is used as a basis for the `ttW Syst` $7j$ ($\geq 8j$) systematic uncertainties.

As expected, the case without the additional uncertainties is highly critical as a higher contribution of the $t\bar{t}W$ background in the data composition in the SR high BDT leads to an important increase in the significance. Adding the uncertainty however, entirely mitigates this effect and the obtained significance becomes relatively stable with respect to the $t\bar{t}W$ contribution. This is an important result, albeit it comes at a cost, which is seen in the difference of the obtained significances between the two setups for $S_{\mathrm{PF}_{t\bar{t}W}} = 1$ of approximately 30 %. This effectively means that if there is no noticeable difference between the $t\bar{t}W$ contributions to data in different regions, the introduction of this additional uncertainty lowers the obtained significance by 30 % for the setup where the signal region is split.

Fortunately, the impact is considerably lower (around 15 %) for the nominal non-split setup as can be seen in Figure 8.15. This large difference between the split and non-split setup is likely due to smoothing and pruning effects, where the additional systematic uncertainties are smoothed inclusively over the entire BDT spectrum for the non-split setup, while the smoothing is applied separately for the two sub regions in the split case.

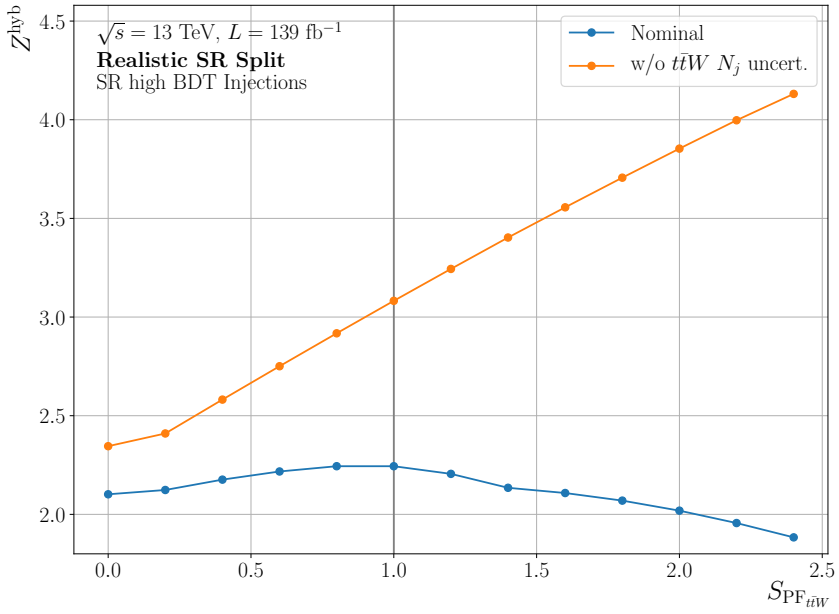


Figure 8.13: Change of the obtained hybrid significance in dependence of the value of the scale applied to the $t\bar{t}\,W$ contribution in the pseudo data $S_{\text{PF}_{t\bar{t}W}}$. The nominal fit setup is depicted in blue, while the setup without the additional $t\bar{t}\,W$ uncertainties are shown in orange. A grey vertical line is drawn for $S_{\text{PF}_{t\bar{t}W}} = 1$ to illustrate the impact of adding the additional uncertainties on the obtained significance for the unaltered realistic Asimov pseudo data set. The Realistic SR Split fit setup is used in order to emulate the response of the fit during the unblinding step of the analysis.

8.3.3.3 Background Injections with $t\bar{t}W$ uncertainties

However large the loss in significance by introducing the additional uncertainties is, they are also shown to considerably stabilise the fit and the obtained significance with respect to the studied differences. This is further demonstrated in Figure 8.14, where the same information is shown as in Figure 8.9, albeit for the nominal setup including the additional uncertainties on $t\bar{t}\,W$. The results presented in this figure are again obtained with the combined, nominal, SR where the pseudo data is applied, while real data is used in the nominal CRs.

In contrast to Figure 8.9, this figure shows that the obtained $\hat{\mu}_{4t}$ is very stable with respect to differences in the $t\bar{t}W$ contribution, which is the desired outcome of introducing the additional uncertainties, and that the overall impact on any post fit NF is lower than in the scenario without the additional uncertainties. While in the former the largest impact amounts to 12 %, it reaches up to 25 % in the latter.

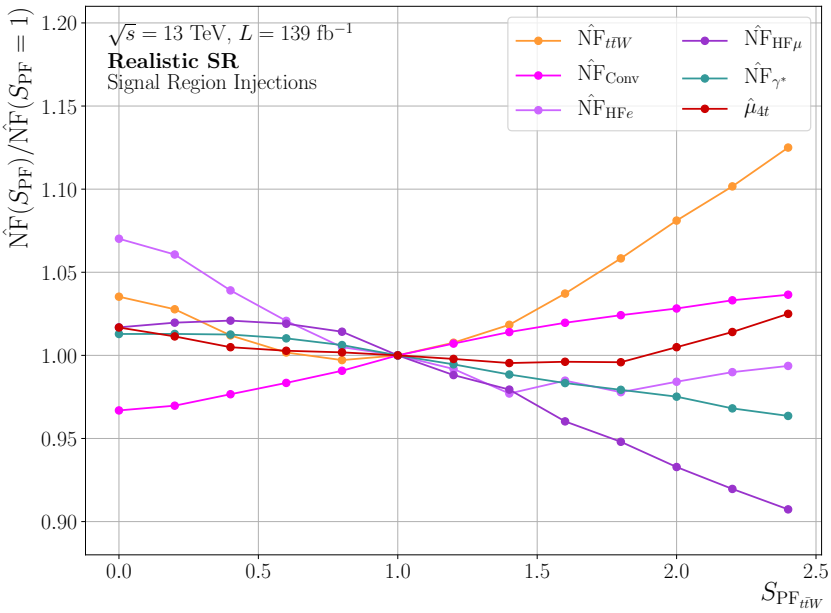


Figure 8.14: Relative change of the post fit NFs obtained in a fit where the contribution of $t\bar{t}W$ to the pseudo data set is scaled by $S_{\mathrm{PF}_{t\bar{t}W}}$. Different colours represent the different post fit NFs. The Realistic SR fit setup is used so that a change in the pseudo data set represents a discrepancy between the data set used in the CRs and the one used in the SR. The `ttW Syst (7j)` $\geq$`8j` uncertainties are used in this fit so that the fit represents the nominal fit setup and the uncertainties on the post fit NFs are not shown for a clearer visualisation.

The impact of introducing the additional uncertainties is also studied for the evolution of the hybrid significance in terms of the scale applied to the $\mathrm{PF}_{t\bar{t}W}$, which is shown in Figure 8.15. In this figure, the cost of introducing the additional systematic uncertainty is represented by the difference between the nominal and the case without the additional uncertainties for $S_{\mathrm{PF}_{t\bar{t}W}} = 1$ and amounts to approximately 15 % as mentioned in subsubsection 8.3.3.2. Apart from this loss, the obtained significance is less independent from the scale applied to the $t\bar{t}W$ contribution in the pseudo data when adding the uncertainties, although it evolves in the expected way, where the significance diminishes for larger contributions of $t\bar{t}W$ and where it increases for smaller

contributions.

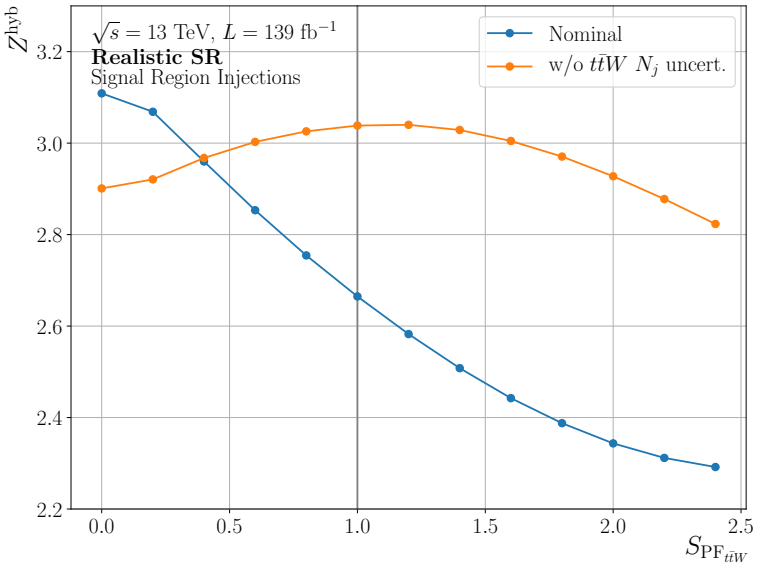


Figure 8.15: Change of the obtained hybrid significance in dependence of the value of the scale applied to the $t\bar{t}\,W$ contribution in the pseudo data $S_{\text{PF}_{t\bar{t}W}}$. The nominal fit setup is depicted in blue, while the setup without the additional $t\bar{t}\,W$ uncertainties is shown in orange. A grey vertical line is drawn for $S_{\text{PF}_{t\bar{t}W}} = 1$ to illustrate the impact of adding the additional uncertainties on the obtained significance for the unaltered realistic Asimov pseudo data set. The Realistic SR fit setup is used.

8.3.3.4 Signal Injections

So far only the background contributions to the pseudo data set in the SR have been scaled, while it is at least equally important to probe the response of the fit to a change in the amount of signal in the pseudo data. In retrospect, this is particularly intriguing seeing that the observed $\hat{\mu}_{4t}$ of the final analysis considers the proportion of the signal in real data to be about twice as large as expected.

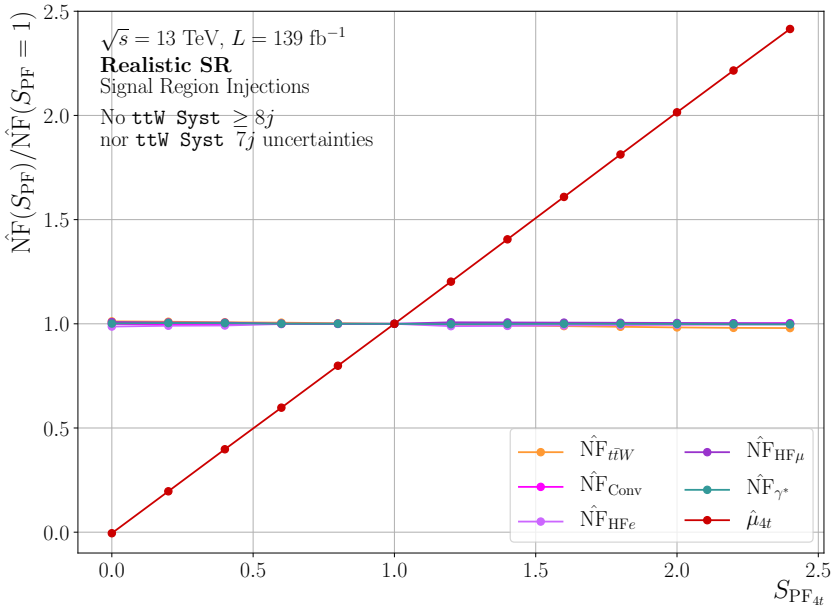
It is important to note that an injection of signal into the pseudo data in the SR is almost identical to injecting the signal in pseudo data in all regions as the CRs contain only a negligible amount of signal.

The sub figures in Figure 8.16 show the relative change of the post fit NFs for different scales $S_{\text{PF}_{4t}}$ applied to the amount of signal in the pseudo data used in the SR.

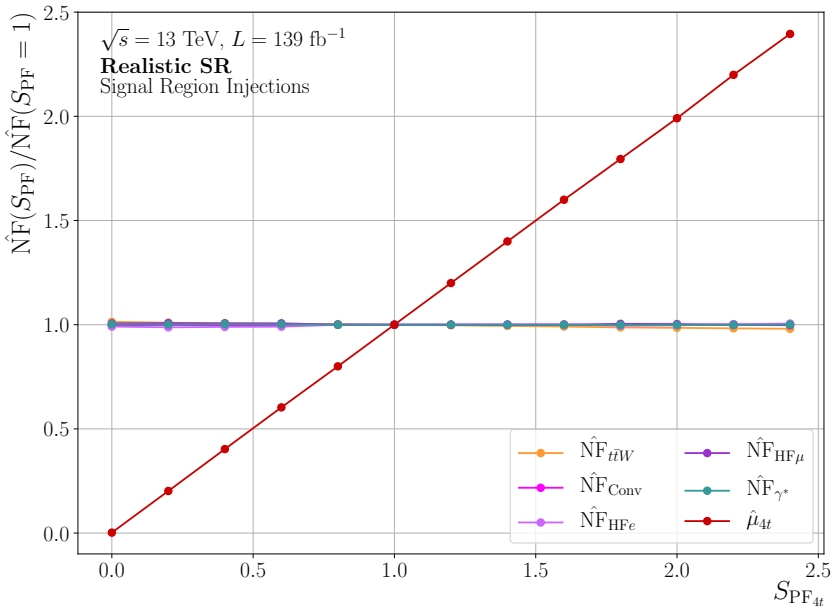
The same setup as above is used and the results are shown without the additional uncertainties on $t\bar{t}\,W$ in Figure 8.16a and for the nominal case in Figure 8.16b.
In both cases the response of the fit to changes in the signal contribution is almost exactly linear. This is the desired outcome and also expected seeing again that all other NFs are strongly constrained by the CRs and that the excess (deficiency) in the pseudo data set caused by additional (missing) signal contributions has the exact shape of the signal prediction. It is important to see that also the introduction of the additional uncertainties on $t\bar{t}\,W$ do not reduce the signal efficiency, i.e. the linearity of the response of $\hat{\mu}_{4t}$ to $S_{\mathrm{PF}_{4t}}$.

The change of the obtained hybrid significance for different $S_{\mathrm{PF}_{4t}}$ is shown in Figure 8.17, where it can be seen that the difference in obtained significance between the setup without the additional uncertainties and the nominal diminishes for a lower amount of signal in the pseudo data than the nominal $S_{\mathrm{PF}_{4t}} < 1$ and for a considerably larger amount of signal in the pseudo data starting from $S_{\mathrm{PF}_{4t}} \geq 2$. While the loss in significance for $S_{\mathrm{PF}_{4t}} = 1$ is approximately 15 % (0.37 in absolute terms), it is reduced to approximately 5 % (0.25 in absolute terms) for $S_{\mathrm{PF}_{4t}} = 2$, which represents the actual result of the final analysis. Thus, the cost of introducing the additional uncertainties on $t\bar{t}\,W$ in the Data fit is potentially lower than expected following the $\mu_{4t} = 1$ expectation.

In conclusion, the additional uncertainties on $t\bar{t}\,W$ are considered crucial for obtaining stable and expected results after unblinding as well as in the case of possible discrepancies between the behaviour of $t\bar{t}\,W$ in the SR and the CRs. The `ttW Syst` $\geq 8j$ is the second most impactful uncertainty on the final result of $\hat{\mu}_{4t}$ and the most impactful on the measurement of σ_{4t}. It has been introduced as a consequence to the studies shown here.



(a) without additional $t\bar{t}W$ uncertainties



(b) nominal

Figure 8.16: Relative change of the post fit NFs obtained in a fit where the signal contribution to the pseudo data set is scaled by $S_{\mathrm{PF}_{4t}}$. Different colours represent the different post fit NFs. The Realistic SR fit setup is used so that a change in the pseudo data set represents a discrepancy between the data set used in the CRs and the one used in the SR. However, since the amount of signal in the CR is negligible, the results are identical with a signal injection in the pseudo data set in all regions. The results are shown for the case where he `ttW Syst (7j) ≥8j` uncertainties are not used (top) and where they are used (bottom). Uncertainties on the post fit NFs are not shown for a clearer visualisation.

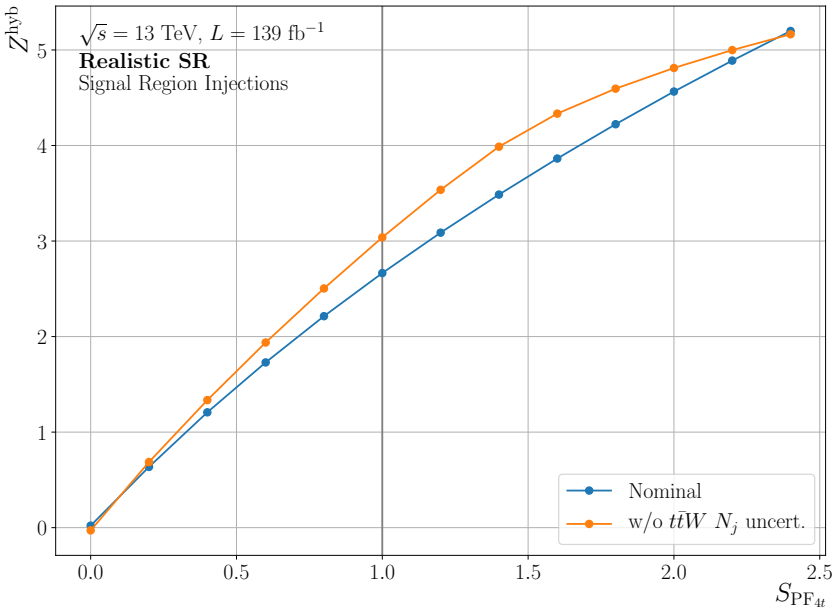


Figure 8.17: Change of the obtained hybrid significance in dependence of the value of the scale applied to the signal contribution in the pseudo data $S_{\text{PF}_{4t}}$. The nominal fit setup is depicted in blue, while the setup without the additional $t\bar{t}\,W$ uncertainties are shown in orange. A grey vertical line is drawn for $S_{\text{PF}_{4t}} = 1$ to illustrate the impact of adding the additional uncertainties on the obtained significance for the unaltered realistic Asimov pseudo data set. The Realistic SR fit setup is used.

9 Reconstruction of the Four Top Quark System

Finally achieving first evidence for the simultaneous production of four top quarks (see previous chapter) opens the possibility to study this process in detail. This is interesting for multiple reasons.

First, in depth studies on the properties of the signal can help significantly in determining if the evidenced process actually represents the SM prediction, or if it rather follows a BSM description that just yields a production cross section similar to the one predicted by the SM.
Second, these in depth studies could include differential cross section measurements, which would be particularly interesting because the process covers an area of the phase space not previously studied.
Third, the top-yukawa coupling y_t can be measured using $4t$ events, which has been one of the motivations to look for this process in the first place (see chapter 3).

Most of these newly opened opportunities rely on a precise knowledge of the kinematic properties of the four top quarks. One way to access this information is by using event reconstruction techniques. The application of one such technique on one of the decay modes of the four top quark system is therefore studied and presented in this chapter.

In addition to the possibilities of detailed studies, as mentioned above, event reconstruction could also be used in order to provide additional discriminant observables for future measurements of four top quark production aiming to obtain first observation.

To the knowledge of the author, the studies presented here represent the first attempt at reconstructing the four top quark system with a likelihood based method in ATLAS and they were conducted as part of the cotutorship programme (*cotutelle de thèse*) with TU Dortmund University.

The signal that is considered for reconstruction in this chapter is described in section 9.1, while a clearer picture of what is meant by event reconstruction and how it can be achieved is given in section 9.2. Section 9.3 introduces the implementation of the reconstruction of $4t$ events in a modular framework. The performance of this implementation is evaluated in section 9.6 and a rough outlook on possible improvements is given in section 9.7.

9.1 Reconstructed Signal

The studies performed in this chapter are used to probe possible pathways of future searches and measurements of the four top quark production, so that the sought-after signal in this chapter is precisely that process.

As pointed out in chapter 3, the four top quark system can decay in several different modes, where the mode used for the studies shown here is the single-lepton (1ℓ) one. This decay mode has not been covered in chapter 7 but it is equally taken into account for the previous and ongoing searches for the production of four top quarks [46].

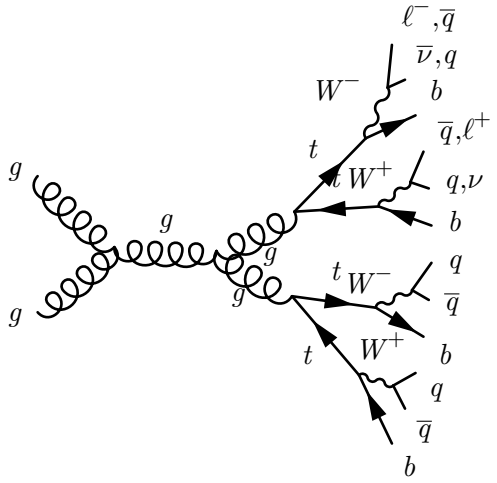


Figure 9.1: Representative tree-level Feynman diagram of four top quark production in the single-lepton decay mode.

As a reminder, the single leptonic decay mode refers to the case where, out of the four W bosons that originate from the decay of the four top quarks, one decays into a charged lepton and neutrino pair (leptonic decay), while the other three decay into quark–antiquark pairs (hadronic decay)[1]. This yields a large set of final state particles, even only considering the tree-level process and without considering any radiation. The set of particles includes 10 jets, where four jets are initiated by b quarks, one charged lepton and one neutrino. A representative Feynman diagram of the production of four top quarks by gluon fusion with the subsequent decay in the single-lepton mode is shown in Figure 9.1.

The single-lepton decay mode has been chosen for this first attempt to reconstruct the four top quark system, because it is the only mode with just one neutrino in the final state. This is an important property, as the disentanglement of multiple neutrinos from the information contained in $E_{\mathrm{T}}^{\mathrm{miss}}$ is far from trivial.

[1] The expression of the decay mode of the W boson is adapted to the corresponding top quark in the following. Thus, a *leptonically decaying* top quark refers to the case where the W boson from the decay of the top quark decays leptonically.

9.2 Event Reconstruction

In order to describe the principles and challenges of event reconstruction, it is necessary to introduce the notion of different *levels* at which objects or particles can exist. These are briefly described in the following.

Parton Level describes actual and virtual particles, usually without taking into account any hadronisation or radiation effects. In the context of this chapter, they can best be thought of as the *particles* drawn in Feynman diagrams.

Particle Level describes particles and objects after hadronisation and radiation effects. Most importantly, quarks and gluons are no longer individual particles, but rather compound objects called jets (see chapter 5). The particle level originates from the parton level (by the aforementioned processes), but without considering measurements of any kind. Generally, there can be a *one-to-one*, a *one-to-many* and a *many-to-one* correspondence between the parton level and the particle level. This refers to the fact, that one parton level particle can directly correspond to one particle level object, to multiple particle level objects due to radiation, or multiple parton level particles can correspond to a single particle level object in the case that objects are merged.

Reconstruction Level refers to the ensemble of objects reconstructed from detector information using the techniques described in chapter 5. These objects correspond to the particle objects through the interaction of those objects with the detector and the subsequent reconstruction of these interactions. The same types of relations as between the parton and particle levels exist between the particle and reconstruction levels. This is the case as there can first be an exact match between objects of the two levels, second, a single particle level object can be reconstructed as multiple separate objects and finally, multiple particle level objects can also be merged and reconstructed as a single object. The latter is especially the case for very dense environments and if the distance between two particle level objects is within the order of magnitude of detector resolution effects.

Additionally, these resolution effects, in combination with acceptance effects, can lead to loosing a particle level object during the reconstruction step, which would thus not be able to be matched. Furthermore, there can also be additional objects that are reconstructed which have no connection to the parton level process. This is mainly caused by pile-up events as described at length in chapter 6.

Model Level is much less commonly used in literature as the other levels and refers to the ensemble of objects that are the outcome of the event reconstruction technique. These objects can either correspond exactly to the objects from the reconstruction level in the case where the event reconstruction is only achieved by determining which of the reconstruction level objects correspond to the parton or particle level ones, or they can be altered as the result of a kinematic fit used for the event reconstruction as is described in subsection 9.2.2.

Following the remarks at the beginning of this chapter, the goal of event reconstruction is to obtain the (kinematic) properties of particles of interest at the parton or particle level using only the measured properties of objects at the reconstruction level. Given a series of simplifications listed below, this can be achieved by finding the correct combination of reconstructed (or model) objects that correspond to the final state objects or particles of the decays of the particles of interest.

To concretise this for the analysis presented here, the particles of interest are the four top quarks, while the final state objects are the decay products of the four top quark system in the single-lepton mode at particle level. The reconstructed objects are the objects built from detector information, such as jets, leptons and $E_{\mathrm{T}}^{\mathrm{miss}}$ that are associated to the sought-after topology by the algorithm. Model objects can either be exactly the same as the reconstruction level or they can be slightly altered versions of these objects as a result of a fit. One exception to this are model neutrinos, which have *pendant* at reconstruction level and therefore need to be determined by the event reconstruction algorithm.

9.2.1 Mappings and Permutations

One major simplification that is generally applied in reconstruction studies is that the relations between the parton and particle levels should be *one-to-one*, i.e. there is a direct correspondence between every particle of the former to exactly one object of the latter. In this way, the two levels become interchangeable and effects like radiation are entirely neglected. This is generally required because it is much more straightforward to reconstruct particle level objects than it is for parton level particles as the latter are not physical in the context of a measurement.

This simplification allows for a mapping of reconstructed to particle level objects, where another simplification, that all particle level objects should have exactly one corresponding reconstruction level object, is required for one selection of this ana-

lysis. Generally, additional reconstructed objects (that therefore do not correspond to particle level objects) are allowed as these are ubiquitous and because the reconstruction algorithms are precisely designed to discard these additional objects.

Nevertheless, only events without such additional reconstructed objects are used in the main event selection of this analysis, so that the number of particles and reconstructed objects is identical and there is an unambiguous, correct mapping between the two levels.

In this case, it is the task of the event reconstruction algorithm to determine this correct mapping out of the set of all possible mappings. And an event is defined to be correctly reconstructed if the algorithm has been able to select the correct mapping.

Since the mapping is assumed to be for a *one-to-one* relation, all possible mappings are given by the set of permutations of pairs of particle and reconstruction level objects.

This is insinuated in Figure 9.2, where the final state objects of the *hadronic* decay of a top quark are mapped in *one-to-one* relations to jets in the set of reconstructed objects. If the number of reconstructed and particle level jets are the same $N_j = 3$, there are $3! = 6$ possible permutations, out of which the correct one needs to be determined. If, however, additional reconstructed jets were allowed and $N_j^{\text{reco}} = 5$ (depicted by the two additional jets in teal), the number of possible permutations rises sharply to $N_j^{\text{reco}}!/(N_j^{\text{reco}} - N_j^{\text{particle}})! = 60$. The number of possible permutations is an important parameter, especially for computationally intensive event reconstruction algorithms and is discussed in detail in subsection 9.3.3.

There are a series of available algorithms that attempt to find the correct mapping and thus reconstruct the event. One group of such algorithms applies kinematic fitting to adjust the properties of the objects at reconstruction level with respect to the sought-after topology in order to improve the reconstruction efficiency. This is the case for the method used for the studies presented here, so that kinematic fitting is introduced in detail in the following.

9.2.2 Kinematic Fitting

Kinematic fitting refers to the technique of adjusting the kinematic properties of reconstructed objects to the hypothesised event topology to help find the correct mapping of the former to the latter and to correct for detector resolution and acceptance effects.

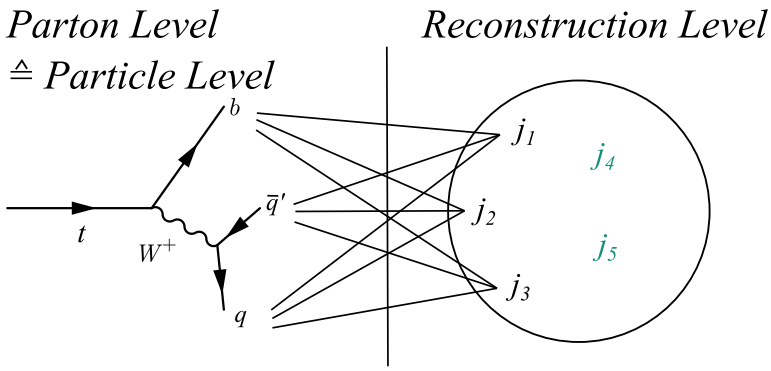


Figure 9.2: Schematic of all possible *one-to-one* mappings between the particle level final state jets of the *hadronic* decay of a top quark and reconstructed jets. The relations are only shown for the case that the number of reconstructed and particle jets are the same (black). However, two additional reconstructed jets are shown in teal to illustrate the possibility of reconstructing jets that are not associated to the process in question and rather originate from pile-up events.

The fit results in a new class of objects called *model* objects that represent the reconstructed objects after the fit (and thus relate to the fit model). The main idea behind the application of such kinematic fits is to mitigate differences between the *particle level* and the *reconstruction level* objects that are due to detector resolution effects. These effects are introduced to the fit in the form of constraint terms given by so called Transfer Functions (TFs) W, which are obtained in auxiliary measurements in simulation and which yield the probability of observing an energy E_o at *reconstruction level*, given the real energy E_o^{true} for any object o. The goal of the fit is then to find an estimator to the true energy, called $\tilde{E}_o$. The TFs are usually different for different objects and are described in more detail in subsection 9.3.2.

One instance of an event reconstruction algorithm using kinematic fitting is the χ^2-method, which is described in the following.

9.2.3 The χ^2 Method

The determination of the *most probable* mapping in case of the χ^2 method is achieved by maximising a variable χ^2 (hence the name), which, for the single-lepton decay of

$t\bar{t}$[2] can be defined as

$$\begin{aligned}\chi^2 = &8\ln 2\frac{(m_{q_1q_2q_3}-m_t)^2}{\Gamma_t^2} + 8\ln 2\frac{(m_{q_1q_2}-m_W)^2}{\Gamma_W^2}\\ &+ 8\ln 2\frac{(m_{q_4\ell\nu}-m_t)^2}{\Gamma_t^2} + 8\ln 2\frac{(m_{\ell\nu}-m_W)^2}{\Gamma_W^2}\\ &- 2\sum_{i=1}^{4}\ln W_j(E_{j,i}|\tilde{E}_{j,i}) - 2\ln W_\ell(E_\ell|\tilde{E}_\ell)\\ &- 2\ln W_{E^{\text{miss}}}(E_x^{\text{miss}}|\tilde{p}_\nu^x) - 2\ln W_{E^{\text{miss}}}(E_y^{\text{miss}}|\tilde{p}_\nu^y).\end{aligned} \tag{9.1}$$

There, $m_{q_iq_j}$ represents the invariant mass of the two-jet system of the model jets i and j, m_t (m_W) represents the mass of the top quark (W boson), Γ_t (Γ_W) represents the decay width of the top quark (W boson) and $m_{\ell\nu}$ is the invariant mass of the charged lepton and neutrino system at the model level. The $W_{j,\ell,E^{\text{miss}}}$ are the TFs for jets, charged leptons and $E_{\text{T}}^{\text{miss}}$, respectively. For this method, the TFs are described by normal distributions. $E_{x,y}^{\text{miss}}$ are the x and y components of the missing energy (see section 5.5), while the $\tilde{p}_\nu^{x,y}$ are the x and y components of the momentum of the model neutrino.

For the computation of $m_{q_4\ell\nu}$ from *reconstruction level* information, also the z component of p_ν needs to be determined. This can either be achieved by assuming the p_{T} of the top quark system to be zero, so that a constraint on the mass of the W boson of $m_W^2 = (p_v + p_\ell)^2$ can directly be used to extract the missing component from $E_{\text{T}}^{\text{miss}}$, or by leaving all spatial components of the momentum of the model neutrino as free parameters to the fit.

The value of χ^2 is then minimised for each possible mapping of reconstructed objects to model objects and the algorithm chooses the mapping with the lowest obtained χ^2 to be the correct one.

Another instance of an event reconstruction algorithm using kinematic fitting is based on the maximisation of a likelihood. This algorithm is used for the reconstruction of four top quark events studied in this chapter and is thus described in section 9.3.

[2] Analogous the the single-lepton decay of $4t$, this decay is defined by one leponically and one hadronically decaying top quark. This yields a set of final state particles comprised of one charged lepton, one neutrino and four jets among which two contain B hadrons.

9.3 The Kinematic Likelihood Fitter

The likelihood method introduced in the following is implemented in the Kinematic Likelihood Fitter (KLFitter), tool set first introduced in [3]. The tool set is based itself on the ROOT framework and the Bayesian Analysis Toolkit (BAT) introduced in [139]. The algorithms in the KLFitter tool set can easily be adapted to different topologies, which has been done for the reconstruction of four top quark events presented here.

9.3.1 The Likelihood Method

The likelihood method closely follows the approach of the χ^2 method, where a likelihood is maximised instead of the minimisation of the χ^2 variable. For the sought-after signal, this likelihood is given by

$$\begin{aligned}\mathcal{L} = &\mathcal{B}(m_{b_1q_1q_2}|m_t,\Gamma_t)\cdot\mathcal{B}(m_{q_1q_2}|m_W,\Gamma_W)\cdot\\ &\mathcal{B}(m_{b_2q_3q_4}|m_t,\Gamma_t)\cdot\mathcal{B}(m_{q_3q_4}|m_W,\Gamma_W)\cdot\\ &\mathcal{B}(m_{b_3q_5q_6}|m_t,\Gamma_t)\cdot\mathcal{B}(m_{q_5q_6}|m_W,\Gamma_W)\cdot\\ &\mathcal{B}(m_{b_4\ell\nu}|m_t,\Gamma_t)\cdot\mathcal{B}(m_{\ell\nu}|m_W,\Gamma_W)\cdot\\ &\prod_{i=1}^{10} W_{\mathrm{j}}\left(E_{j,i}|\tilde{E}_{j,i}\right)\cdot W_{\ell}\left(E_{\ell}|\tilde{E}_{\ell}\right)\cdot\\ &W_{E_{\mathrm{T}}^{\mathrm{miss}}}\left(E_x^{\mathrm{miss}}|\tilde{p}_x^{\nu}\right)\cdot W_{E_{\mathrm{T}}^{\mathrm{miss}}}\left(E_y|\tilde{p}_y^{\nu}\right),\end{aligned} \tag{9.2}$$

where the $\mathcal{B}\left(m_{b_iq_jq_k}|m_t,\Gamma_t\right)$ represent the probability of obtaining an invariant mass of the three jet system of the b tagged jet i and the jets j and k, given the relativistic Breit-Wigner distribution for the mass and decay width of the top quark. Analogously, the $\mathcal{B}\left(m_{q_jq_k}|m_W,\Gamma_W\right)$ represent the probability of the two jet system given the properties of the W boson and the $\mathcal{B}\left(m_{(b_i)\ell\nu}|m_{(t)W},\Gamma_{(t)W}\right)$ represent the probabilities for the (b-tagged jet and) charged lepton plus neutrino system given the properties of the (top quark) W boson.

In contrast to the χ^2 method, the transfer functions W are no longer constrained to be described by normal distributions and their parameterisation is given in subsection 9.3.2.

Here, it is already assumed that b tagging information is used to differentiate model b jets from light model jets. However, this is in general not necessary and the b_i can just as well be considered q_i for all intents and purposes.

Similar to the χ^2 method, the full four-vector of the model neutrino needs to be determined, which is achieved by leaving the three spatial components of the momentum of the model neutrino as free parameters of the fit.

The full list of free parameters of the fit are therefore the energies of all involved model objects (constrained by the TFs, except for the neutrino) and all three spatial components of the momentum of the neutrino. Additionally, m_t can also be left as a free parameter in case the reconstruction is used in a measurement of the mass of the top quark, which is not the case here.

9.3.2 Detector Constraints

It is clear from the likelihood given in Equation 9.2, that the TFs that incorporate detector resolution effects play a very important role in the kinematic fit.
These functions are PDFs and need to be derived for all involved objects and for each specific detector setup. For this first attempt, the TFs derived in [140] are used, which correspond to the conditions of the ATLAS detector at $\sqrt{s} = 8\,\text{TeV}$. The functions are derived for multiple segments of $|\eta|$, that correspond to areas of the detector that are likely to have a different resolution due to different instrumentation. The parameters of the PDFs additionally depend on the true properties such as E, η and p_T of the object, which the fit tries to estimate.

For the objects involved in this analysis, two classes of PDFs are used. The single and the double Gaussian distribution, where the latter is given in the case of energy by

$$\mathcal{N}_2(\tilde{E},E;\mu_1,\sigma_1,A,\mu_2,\sigma_2) = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sigma_1 + A\sigma_2} \cdot e^{-\frac{(\Delta E-\mu_1)^2}{2\sigma_1^2}} + A \cdot e^{-\frac{(\Delta E-\mu_2)^2}{2\sigma_2^2}}, \qquad (9.3)$$

where $\Delta E = \tilde{E} - E$ and the parameters themselves depend on $\tilde{E}$. The parameters $\mu_{1(2)}$ and $\sigma_{1(2)}$ represent the mean and standard deviation of the first (second) Gaussian and the amplitude A regulates the relative contribution of each of the Gaussian distributions to the combined distribution.

Representative TFs are shown for b jets (Figure 9.3a), light jets (Figure 9.3b), electrons (Figure 9.3c), muons (Figure 9.3d) and neutrinos (Figure 9.3e) for a single η range and for multiple values of $\tilde{E}$ to illustrate the dependence of the parameters on that value. The TF for the neutrino is given by the standard Gaussian distribution.
These figures show a symmetric smearing of the energy due to detector effect for charged leptons and neutrinos, which is relatively small for the former and relatively

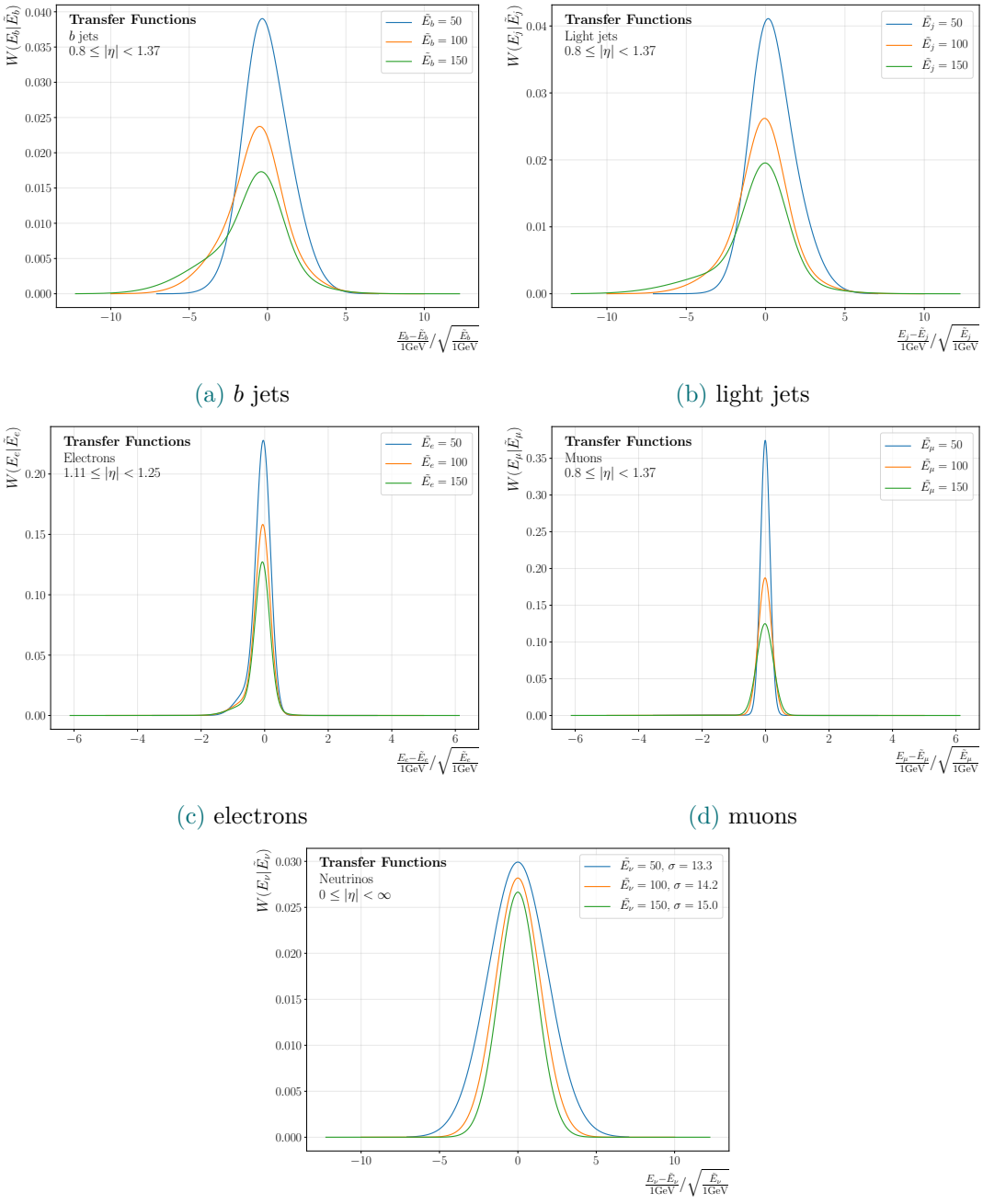
large for the latter. For b and light jets, however, the smearing is skewed towards measuring lower values than the true values. This asymmetry grows stronger for larger true jet energies.
The possibility to use more complex TFs in the likelihood method (such as the double Gaussian distribution) represents a significant advantage of this method over the χ^2 method, as the double Gaussian distribution is considerably better at describing detector resolution effects for the objects mentioned here, as can be seen in Figure 2 in [3].

9.3.3 Permutations

As briefly discussed above, the number of permutations per event is time-critical, especially for more evolved reconstruction algorithms, such as the likelihood method. Even with the very tight constraint that the number of reconstructed jets N_j should be the same as the number of particle level jets, the number of permutations scales with $N_j!$. For the sought after signal, this results in $10! = 3{,}628{,}800$ permutations that need to be evaluated for every event.
This is of course not a feasible number of computations, even for very small data sets, so that additional knowledge of the underlying process needs to be applied to reduce the number of possible permutations.

First, for the computation of the invariant mass of the two (three) jet system, the ordering of the jets has no effect, so that $m_{(q_i)q_jq_k} = m_{\text{permuted}[(q_i)q_jq_k]}$ for all possible permutations of $[(i),j,k]$. This is straightforward and does not need any additional information from the event, so that it is always applied.
Second, b-tagging information can be used to discard any three-jet system with more or less than 1 b-tagged jet and any two jet system with at least one b-tagged jet. This follows the topology of the tree-level signal process, where each top quark has exactly one b quark in the final state of its decay and where no b quark can originate from the decay of the W bosons.

The number of valid permutations that have to be evaluated for every event is then given by $N_{j,\text{light}}!/2 = 360$, where the $N_{j,\text{light}} = 6$ is the number of reconstructed jets that are not b-tagged and the division by two is due to the invariance of the invariant masses for different permutations of light model objects.



(a) b jets

(b) light jets

(c) electrons

(d) muons

(e) neutrinos

Figure 9.3: Representative Transfer Functions as obtained in [140] for the objects used in this analysis. Only one segment in η, but multiple values of $\tilde{E}$ are shown.

9.4 Data Sets

The nominal signal sample at NLO using the AFII detector simulation from the analysis presented in chapter 7 is also used as the nominal signal sample for the studies presented here. In principle, studies using a LO sample would be interesting as this is the assumed model by the reconstruction algorithm. For this first attempt however, the NLO sample is chosen to stay as close as possible to the analysis in chapter 7 and to the situation in data.

One goal of this first attempt at reconstructing the four top quark system is to find potential discriminating variables to be used in future efforts for a first observation of the four top quark production. For this reason, also a $t\bar{t}$+jets sample is used, which is the same as the nominal sample in section 7.3, which is equally at NLO, but uses the full GEANT4 detector simulation.

9.5 Object and Event Selection

The signal defined in section 9.1, is also investigated in parallel to the studies in chapter 7, so that the ongoing analysis uses the same object definitions as the ones stated in section 7.2. For compatibility, the same object definitions are therefore used for the studies presented here.

As a reminder, the sought-after signal yields one charged lepton, one neutrino and 10 jets (among which 4 contain a B hadron) in the final state at particle level. Following the simplifications outlined above, this leads to an 'SR'[3] of

- exactly one electron or muon with a transverse momentum of $p_{\mathrm{T}} > 28\,\mathrm{GeV}$. This lepton has to be the one that activated one of the lepton triggers.
- exactly ten jets with a transverse momentum of $p_{\mathrm{T}} > 25\,\mathrm{GeV}$
- exactly four b-tagged jets using the MV2c10 algorithm at the 77 % working point

For this study, no requirement on $E_{\mathrm{T}}^{\mathrm{miss}}$ is set for the signal region definition, following the prescription of the ongoing analysis.

[3] SR is put in quotes here, as the selection is not in any way optimised or tailored in detail to the signal.

This SR definition follows the simplification that there should be exactly the same number of reconstructed objects as there are particle level objects expected. Another simplification introduced in subsection 9.2.1, is to only select such events where there is a direct correspondence between the parton and particle levels and where there is an unambiguous mapping between the particle and the reconstruction level. The selection that requires this in addition to the SR selection is called the Direct Correspondance (DC) selection further on.

The selection efficiencies of different sub selections of the full DC selection are shown in Figure 9.4, where the preselection requires exactly one trigger-matched charged lepton with a $p_{\mathrm{T}} > 28\,\mathrm{GeV}$ and at least five jets with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$. There, the number of weighted events passing each selection is shown, where the applied selections are cumulative, meaning that each selection also includes all the selection that are to the left of it in the figure. The number of weighted events is given in arbitrary units as the yield is not normalised to any specific luminosity. This is the case for all studies conducted in this chapter and omitted as only relative measures such as efficiencies are studied.
In addition to the absolute number of events, also three types of selection efficiencies are given in Figure 9.4. First, the efficiency of the selection with respect to the previous selection (orange). Second, the selection efficiency with respect to the preselection (green) and third, the selection efficiency with respect to the full SR selection (which is insinuated by a red dashed line). This efficiency is only given for the final DC selection, where it represents, in principle, the four top reconstruction efficiency of a perfect algorithm and thus the upper bound for any reconstruction efficiency studied here. This is the case here, because for any event that passes the SR but not the DC selection, no correct mapping between reconstruction and particle level objects exists.

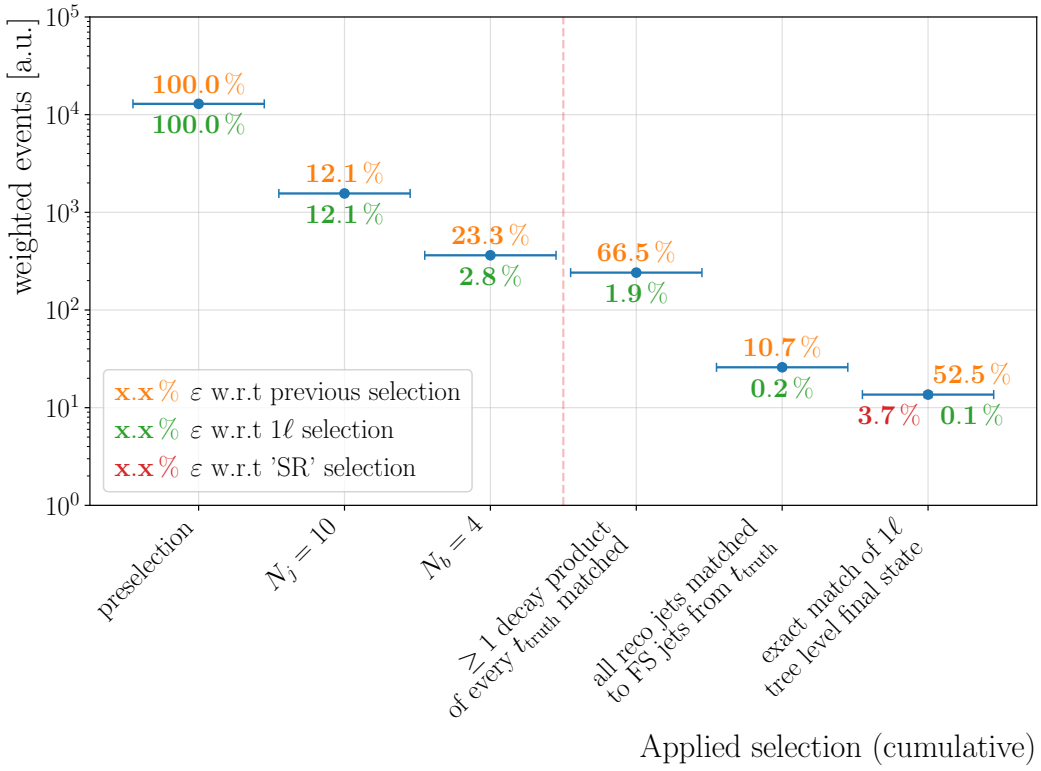


Figure 9.4: Selection efficiencies for the simulated four top quark data set.

9.6 Results

The physical properties of the top quark and the W boson are set to

$$\begin{aligned} m_t &= 172.5\,\text{GeV} \\ \Gamma_t &= 1.5\,\text{GeV} \end{aligned} \quad \text{and} \quad \begin{aligned} m_W &= 80.4\,\text{GeV} \\ \Gamma_W &= 2.1\,\text{GeV} \end{aligned} \tag{9.4}$$

9.6.1 Reconstruction Efficiencies

As part of the definition of a reconstruction efficiency, it is necessary to define a criterion to determine if a truth top quark at particle level has been correctly reconstructed by the algorithm. For this reason, two types of such truth top quarks and jets[4] are defined in addition to so called reconstruction and KLF top jets. Their definition is given in the following.

Truth top (t_{truth}) are given by the four-vector of the top quark at particle level.

[4] Top quarks do not hadronise and thus do not form jets. The term jet is however use to showcase the fact that kinematic properties of this top quark are obtained from the jets of the decay products.

Truth top FS (t_{truthFS}) are given by the vectorial sum of the four-vectors of the three decay products of the truth top quark at particle level. In a pure, radiation-less scenario, this would yield the same four-vector as the t_{truth}, which is not the case here, exactly due to radiation effects considered in the simulation.

Reconstruction top (t_{reco}) are given by the vectorial sum of the four-vectors of the three objects that the KLFitter module associates to one top quark at reconstruction level (so not using the outcome of the kinematic fit). This means, that the reconstructed top jet is not available for the leptonic decay due to the missing component of the neutrino momentum.

KLF top (t_{KLF}) are given by the vectorial sum of the four-vectors of the three objects that the KLFitter module associates to one top quark at model level, so that the energies and the neutrino momentum are adjusted in the kinematic fit.

For every event, there are four truth, truth FS and KLF top quarks and jets and three reconstruction top jets that are ordered by their association to the corresponding KLF top quark, i.e. $t^{\text{had1}}_{\text{KLF}}$, $t^{\text{had2}}_{\text{KLF}}$, $t^{\text{had3}}_{\text{KLF}}$ and $t^{\text{lep}}_{\text{KLF}}$ in case they were correctly mapped.

The mapping, or matching between the objects is achieved using the MCTruthClassifier as introduced in chapter 6 using the Ghost-matching technique [96] in the case of jets.

As an artefact of how the permutations are implemented, the b jet of $t^{\text{had1}}_{\text{KLF}}$ has generally a larger p_{T} than the b jet of $t^{\text{had2}}_{\text{KLF}}$, which in turn has generally a larger p_{T} than the b jet associated to $t^{\text{had3}}_{\text{KLF}}$. The p_{T} associated to the leptonic top $t^{\text{lep}}_{\text{KLF}}$, however, has no particular order. This is shown in Figure 9.5, which depicts the position of the b-tagged jet associated to each of the model top quarks in the p_{T} ranking of all reconstructed jets[5].

With the matching in place, the reconstruction efficiency ε can be defined as the fraction of events where all four top quarks are correctly matched by the algorithm. This means that the fit has been able to determine the correct permutation out of all the possible permutations given by the reconstructed objects. One important remark to make here is that the reconstruction efficiency, naturally, does not provide any information on the quality of the reconstruction which has thus to be determined separately.

[5] An even more stringent but less visually intuitive depiction of this is shown in Figure D.5.

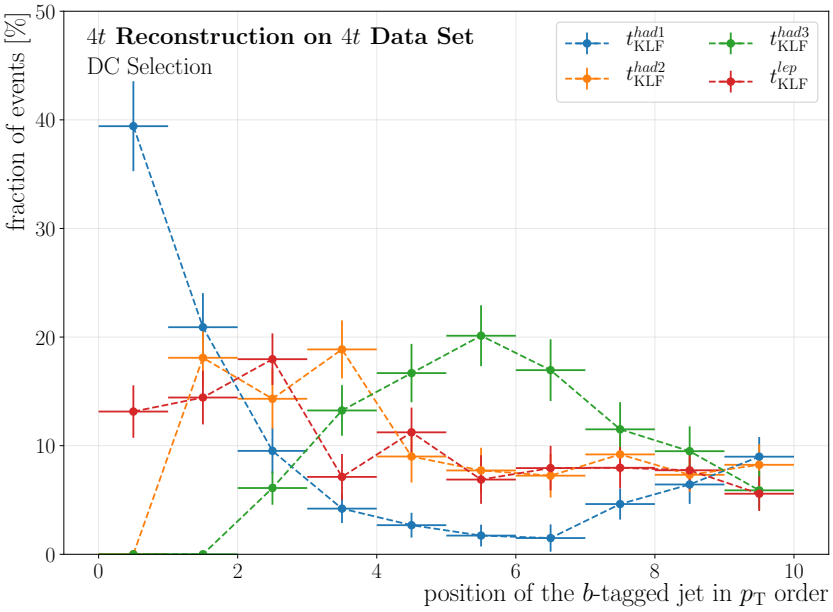


Figure 9.5: Distribution of the position of the b-tagged jet associated to each of the model top quarks in the p_{T} ranking of all reconstructed jets.

Figure 9.6 shows the number of correctly matched top quarks for the DC selection, where the matching is done between the t_{truth} and the t_{KLF}. Because of the very strict selection where the number of reconstructed objects and particle level FS objects is the same, there can never be three top quarks matched. This is the case since not matching one of the truth top quarks means that at least one of the associated final state objects is not correct and is thus originating from another top quark. This other top quark is therefore missing one of the correct final state objects so that at least one of the associated objects to this top quark is also incorrect, leading to the mismatch of at least two top quarks. Following the results of Figure 9.6, the efficiency to correctly reconstruct the four top quark system is

$$\varepsilon = 33 \pm 4\,\%, \tag{9.5}$$

which is comparable to typical values for $t\bar{t}$ reconstruction [3] and is much larger than a random guess which would tend to $\varepsilon_{\mathrm{rand}} = 360^{-1} = 0.3\,\%$ for a large number of events.

The reconstruction efficiency for each individual model top quark is shown in Figure 9.7. While the hadronic model jets present a very similar efficiency within the given uncertainties, the leptonic model jet is more likely to be correctly matched than

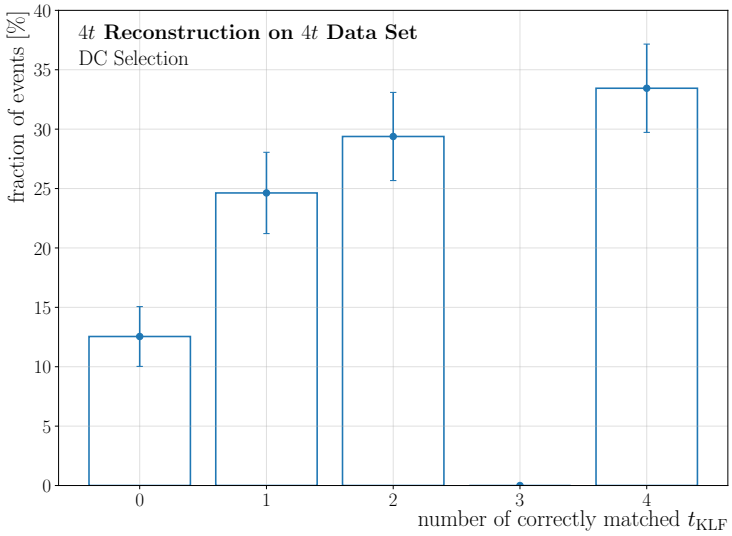


Figure 9.6: Reconstruction efficiencies in terms of number of correctly matched tops.

the hadronic ones. This is expected to some extent as in this case only one jet, the b-tagged jet, has to be associated to the charged lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$ components, which are unique.

The goal of event reconstruction is of course to eventually be applied to real data, where the truth information used to define the DC selection is not available. For this reason, the performance of the algorithm is also shown for the SR selection in Figure 9.8. In Figure 9.8a, it is now possible to match three of the four top quarks or rather incorrectly determine just one of the top quarks because reconstructed jets can now come from pile-up events or other sources, so that any (single) model top quark using one of these non-related reconstructed jets, would be incorrect.

Additionally, the reconstruction efficiency for the SR selection lowers drastically to $\varepsilon = 1.9 \pm 0.2\,\%$, which is expected seeing that an ideal algorithm would already only yield $\varepsilon = 3.7\,\%$[6]. This is because vital information to map the model to the reconstructed objects is missing, which can be due to the above mentioned effects of merging, or multiplication of parton level objects during the transition to the particle level, or because one or more of the particle level objects are lost due to detector acceptance effects.

Figure 9.8b shows that the reconstruction of the leptonically decaying top quark is much less affected by these issues, which is likely due to the fact that the transition

[6] Compare to the efficiency in red in Figure 9.4.

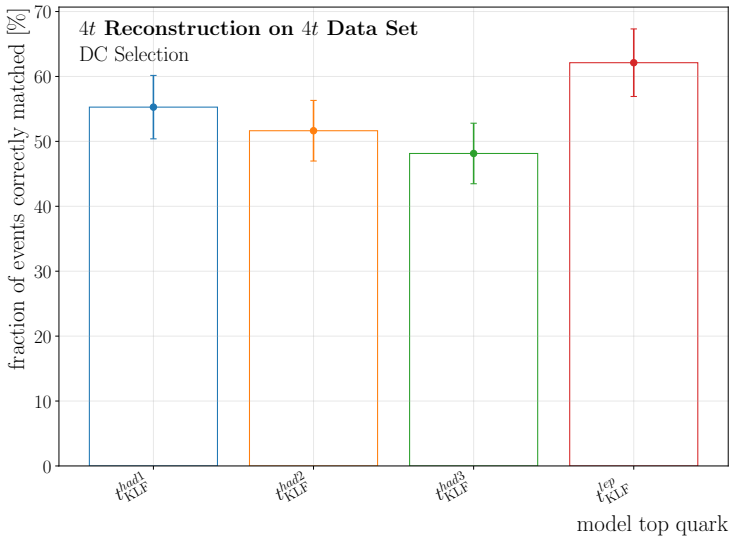


Figure 9.7: Reconstruction efficiencies for each of the model top quarks individually.

between parton level and particle level leptons is much more straightforward.

As mentioned above, the numbers presented so far represent the reconstruction efficiency as the fraction of events where the correct permutation is found by the algorithm. This however does not provide any information on the quality of the event reconstruction, which is therefore discussed in the next section.

9.6.2 Quality of the Reconstruction

As for the reconstruction efficiency, also for the quality of the reconstruction there is a multitude of possible definitions. For this study, the quality of the reconstruction is assessed by how close the model properties of p_{T}, η, ϕ and E are to the true values of the particle level top quarks. These properties are only evaluated for the model to truth pairs that are correctly matched, as there is otherwise no straightforward way to determine which pair to evaluate.

This means that the fraction of events after the DC selection, used for the distributions of the different t_{KLF}–t_{truth} pairs is given by Figure 9.7.

Figure 9.9 shows the distributions of the difference in pseudorapidity between the two objects in Figure 9.9a and the difference in the azimuthal angle in Figure 9.9b. For the latter it is important to note, that always the smallest angular distance between the objects is chosen so that $\Delta\phi \in [0,\pi]$. These two properties are not varied in the fit

for the hadronically decaying top quarks, while they are for the leptonically decaying one, albeit only as part of the neutrino momentum.
The figures show a good agreement between the model and the truth objects, with expected detector resolution effects, with the slight exception of the leptonically decaying top quark, which is likely influenced by the fact that three spatial components are free in the fit.

The difference between the two matched objects is shown in terms of p_{T} in Figure 9.10a and for the energy of the objects in Figure 9.10b. Since the energy is the principal free parameter of the fit, both it and consequently the p_{T} are varied by the reconstruction algorithm. The distributions of both parameters show indeed a peak close to zero, but especially the distributions of the energy also show a large overestimation of the true energy by the algorithm (see the first bin in Figure 9.10b).
This is not ideal and even slightly more puzzling considering that the algorithm uses the reconstructed jets and leptons as a starting point, which in turn should correspond to the final state objects of the particle level. Figure 9.11 shows the difference in p_{T} and energy between the t_{truth} and the t_{truthFS}, where, as a reminder, the latter is constructed from the three expected decay products of the truth top quark at particle level. This shows that the energy of the t_{truthFS} is generally lower than the energy of the t_{truth}, which is mostly due to radiative losses that are not captured here as only exactly three decay products are used to build the t_{truthFS}.

This would mean that, unless the object reconstruction is able to incorporate these effects, it is likely that the starting point for the energy reconstruction is even further off from the $E_{t_{\mathrm{truth}}}$ than should be expected given the TFs.
The potentially large discrepancies between the model top jets and the truth top quarks in Figure 9.10b could thus come frome radiative losses that are not covered by the TFs for the hadronically decaying top quarks. For the leptonically decaying top quark, the disagreement could arise from difficulties in estimating $E_{\mathrm{T}}^{\mathrm{miss}}$ and subsequently from too many degrees of freedom for the model neutrino. This is by no means conclusive, so that these discrepancies need to be studied in more detail, which is left for future analyses at this point.

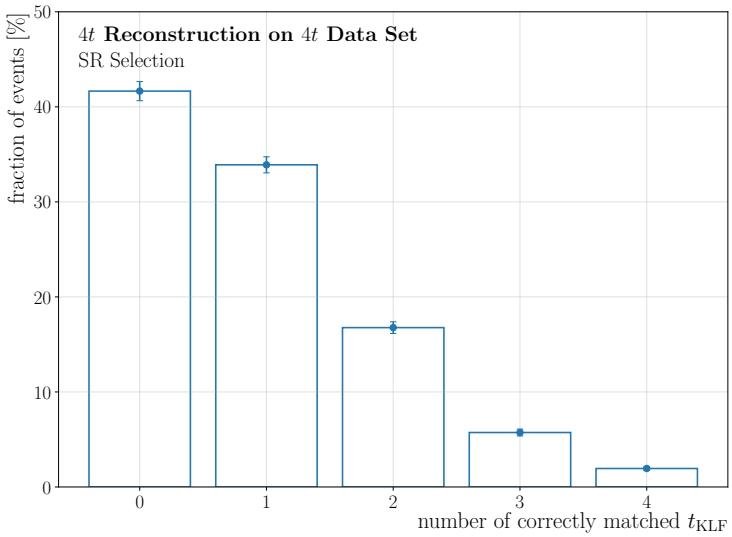
One interesting variable that showcases the advantage of using kinematic fitting for the event reconstruction is the invariant mass of the top quark[7]. This variable is interesting because it represents an important condition for the likelihood in the likelihood method or the χ^2 variable in the χ^2 method.

[7] The same is true for the invariant mass of the W boson, which is not shown here for brevity.

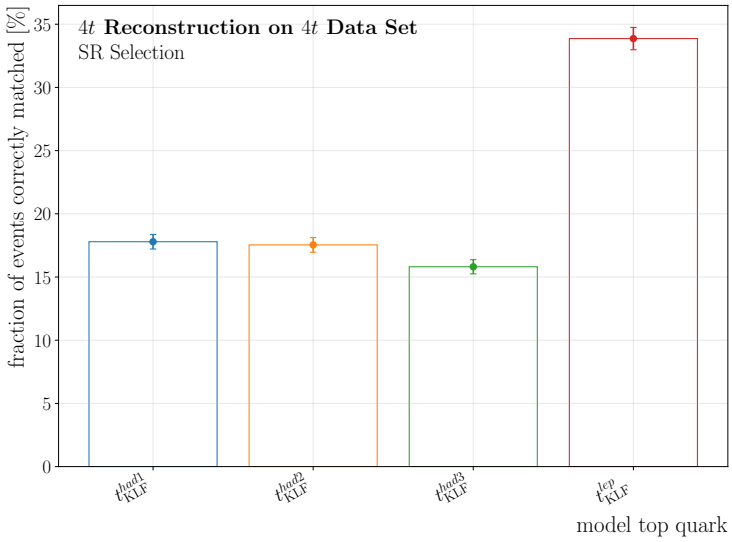
Thus, the distribution of m_t is shown for the truth top quarks and model top jets in Figure 9.12, for the final state truth top jet in Figure 9.13 and finally for the top jets built using reconstruction level objects in Figure 9.14.
The first two figures show the expected invariant mass following the simulated or defined m_t and Γ_t parameters, respectively. Figure 9.13 again shows that the invariant mass of the composite truth top jet (t_{truthFS}) is generally lower than the expected mass shown as a bordeaux dashed line. This is the likely the result of the radiative losses already discussed above.
Figure 9.14 shows the distribution of $m_{t,\text{reco}}$, which follows the expected shape around the correct mass (again indicated by a dashed line), albeit smeared to lower and higher values due to detector resolution effects.
The mitigation of these detector resolution effects for the computation of the score of any possible permutation of objects is what gives an advantage to event reconstruction algorithms that apply kinematic fitting over those that do not.

In summary, there is a decent agreement between the reconstructed and true properties of the top quarks, although additional studies are needed to better understand an overestimation of the energy by the kinematic fit.

Following the deliberations at the introduction to this chapter, event reconstruction is not only a promising topic for more precise measurements of kinematic and angular properties of the four top quark system, but rather it might also provide additional observables that can be used to further discriminate a four top quark signal from background sources. One such observable is introduced and studied in the following.

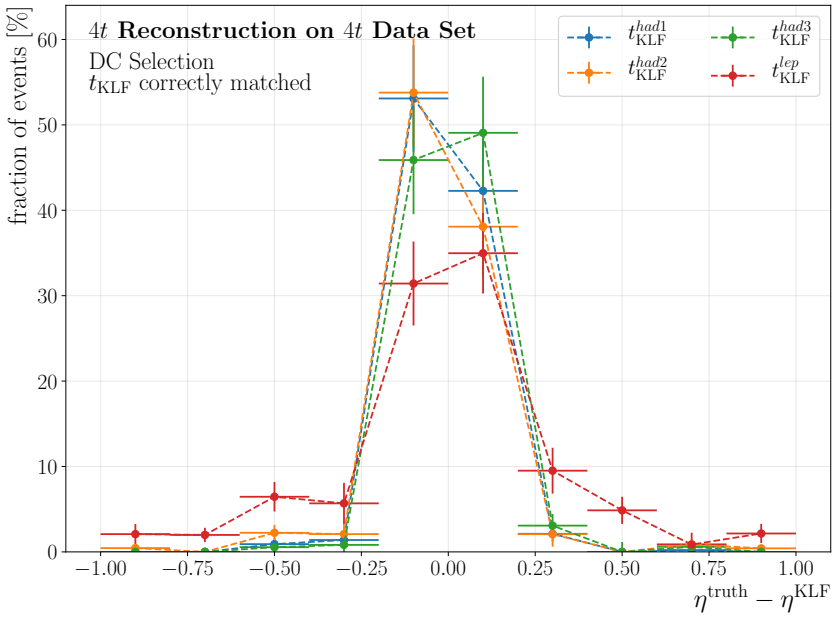


(a) Global

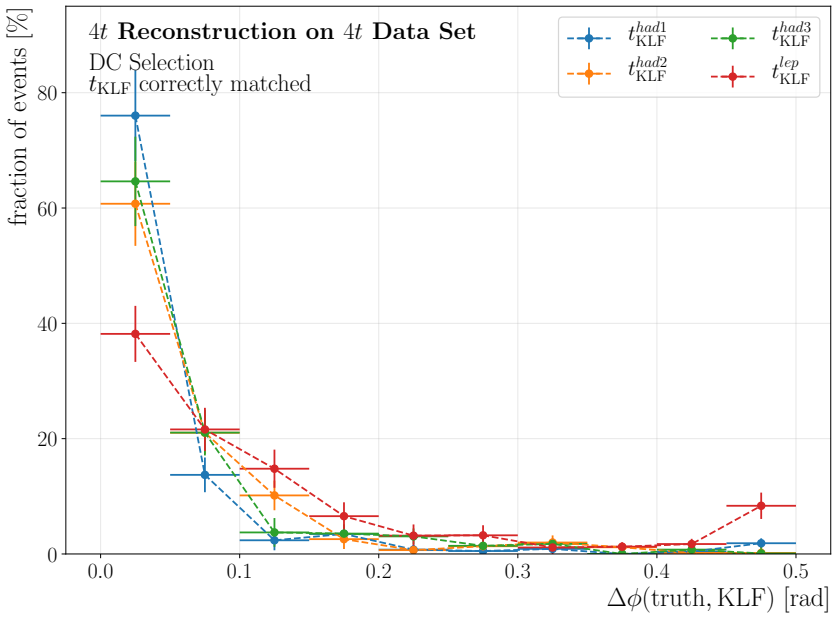


(b) Individual

Figure 9.8: Reconstruction efficiencies in terms of number of correctly matched tops and for each of the model top quarks individually.

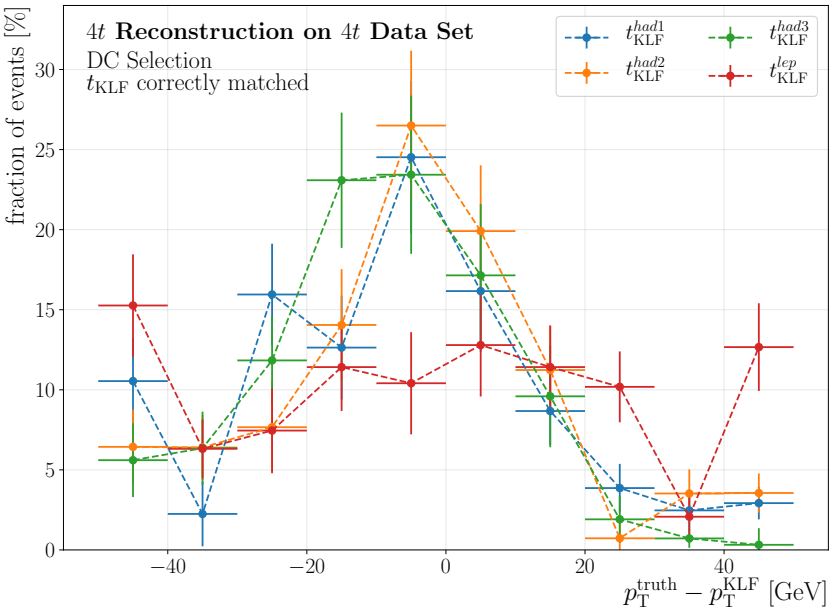


(a) Pseudorapidity

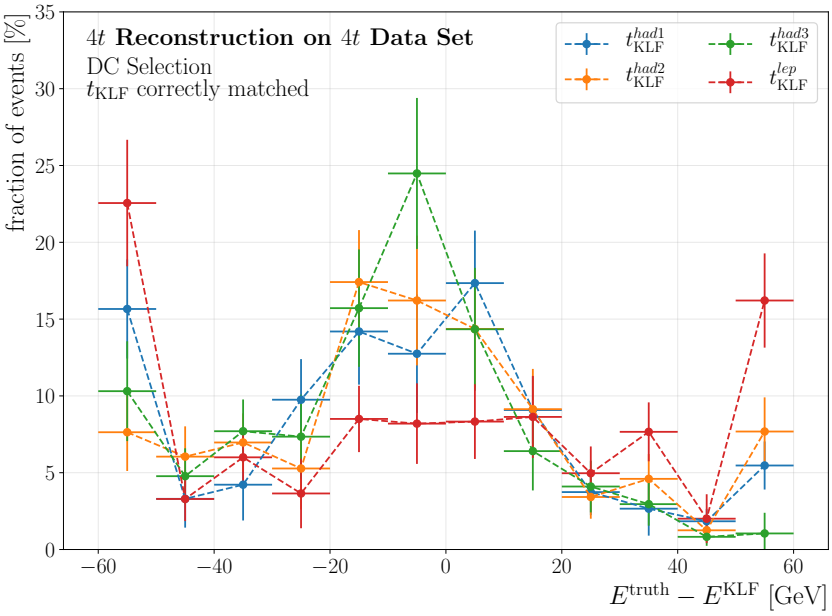


(b) Azimuthal angle

Figure 9.9: Distribution of the difference in pseudorapidity and azimuthal angle between truth and model top quarks, where the truth top quarks have been correctly matched by the model. For the difference in azimuthal angle, always the smaller possible angle is chosen, so that $\Delta\phi \in [0,\pi]$.

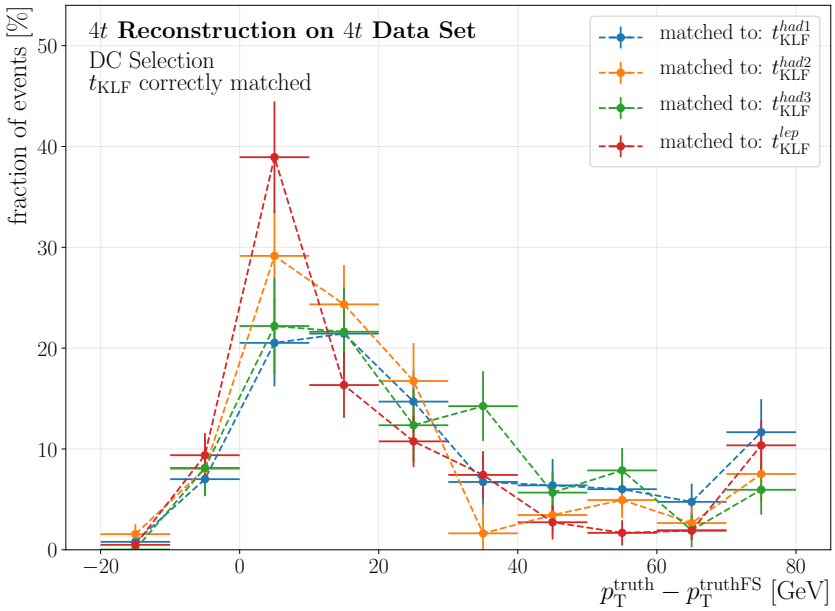


(a) p_{T}

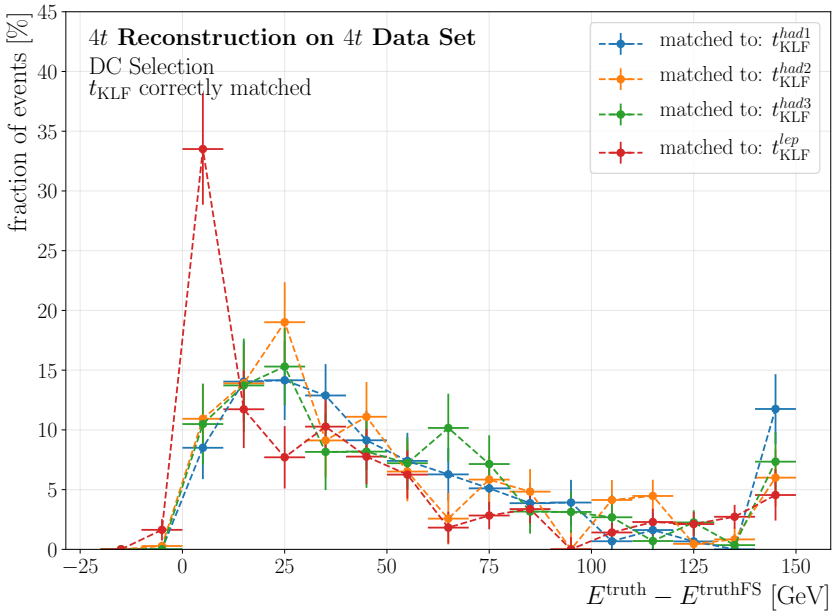


(b) Energy

Figure 9.10: Energy and p_{T} difference between truth and model top quarks, where the truth top quarks have been correctly matched by the model.

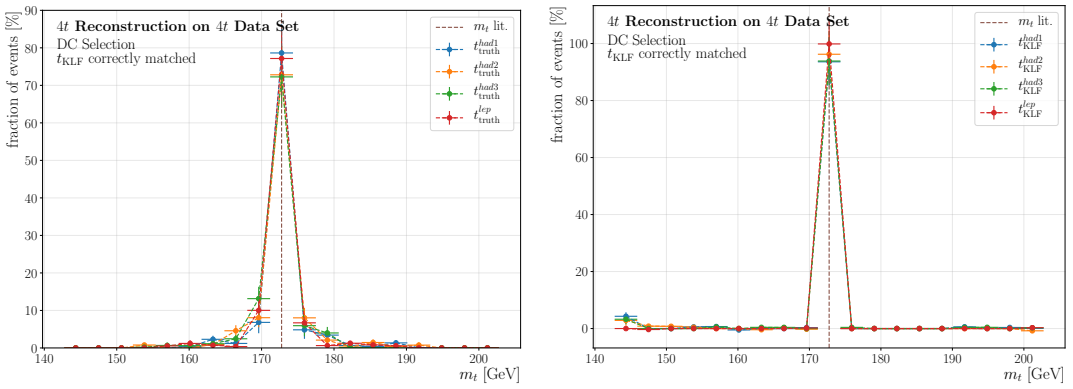


(a) p_{T}



(b) Energy

Figure 9.11: Energy and p_{T} difference between truth top quarks and composite truth top quarks, i.e. the truth top quarks built from the three expected final state objects at particle level. Only entries where the truth objects have been correctly matched by the model objects are shown.



(a) Truth (b) Model

Figure 9.12: Distribution of the invariant mass of the top quark built from truth and model level objects.

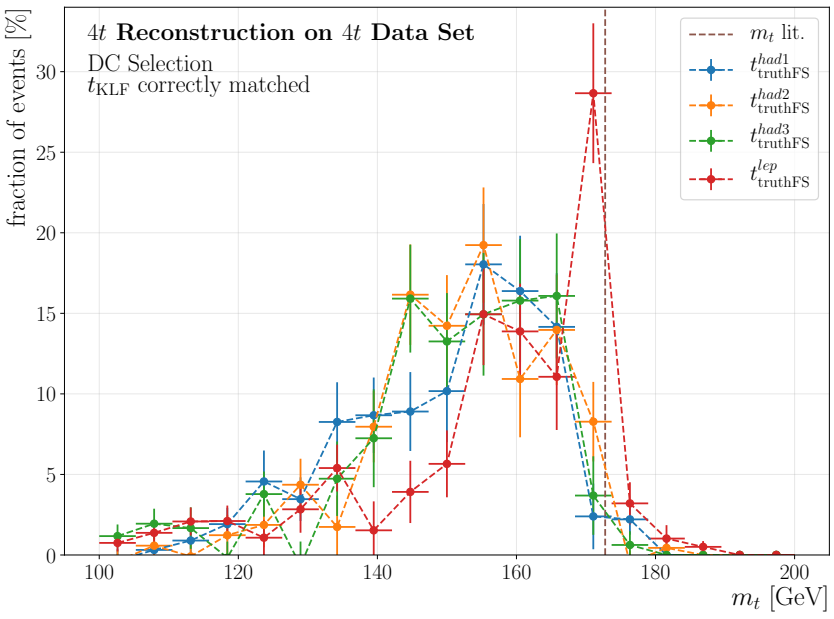


Figure 9.13: Distribution of the invariant mass of the top quark built from truth FS objects.

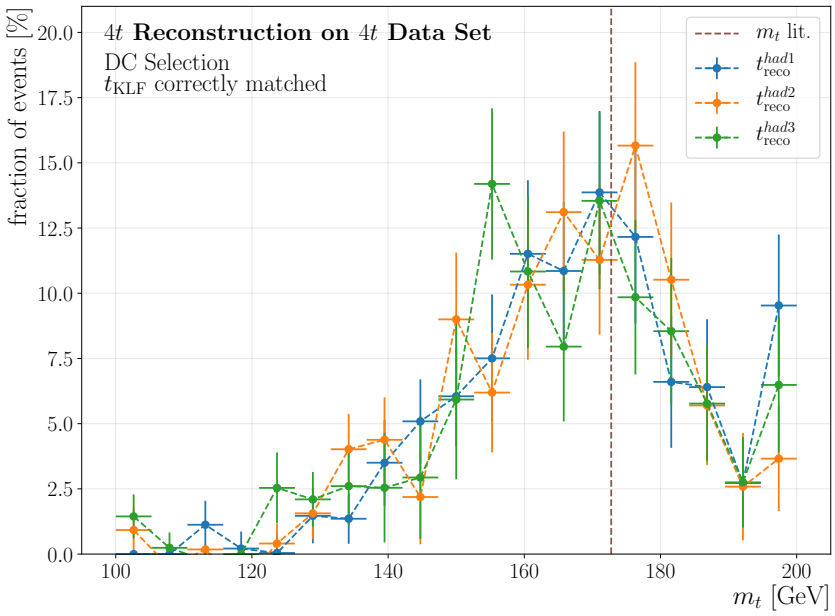


Figure 9.14: Distribution of the invariant mass of the top quark built from reconstruction level objects.

9.6.3 Discrimination Power

A first natural observable to study in terms of its discriminating power between the four top quark signal and background events is the value of the likelihood of chosen permutation. This measure directly codifies how likely it is that the most likely permutation of an event corresponds to the signal process, making it an ideal candidate for discrimination studies.

Figure 9.15 shows the distribution of the maximal log-likelihood per event, which is the maximum out of the set of log-likelihoods for all possible permutations. The distributions are shown for four different selections.
First, the SR selection which represents the widest distribution and which tends to lower values with respect to the other selections. This is expected seeing that not all information needed for the reconstruction might be available as has already been discussed for the low reconstruction efficiency of this selection. However, this selection represents the most widely applicable, as it can be required for any simulated and even for measured data sets.

Second, the distribution is shown for the DC selection which is specific to data sets containing four top quark events. Here, the distribution is much narrower and located at higher values, which is connected to the increased reconstruction efficiency for this selection.

Third, the distribution is again shown for the SR selection, albeit only for those events where all four top quarks have nevertheless been correctly reconstructed. The same is shown for the SR selection as a fourth distribution and both the third and fourth distributions are plotted to illustrate that correctly reconstructed events indeed yield larger likelihood values than events where at least one top quark is not correctly reconstructed.

In order to gauge the quality of this observable in terms of its discrimination power, the likelihood method assuming the four top quark topology is also applied to a data set of the most important background process in the single-leptonic decay mode, $t\bar{t}$+jets. The resulting distribution of the maximal log-likelihood values for this data set is shown in Figure 9.16, where it is compared to the distribution obtained from the $4t$ data set using the same SR event selection in both cases.

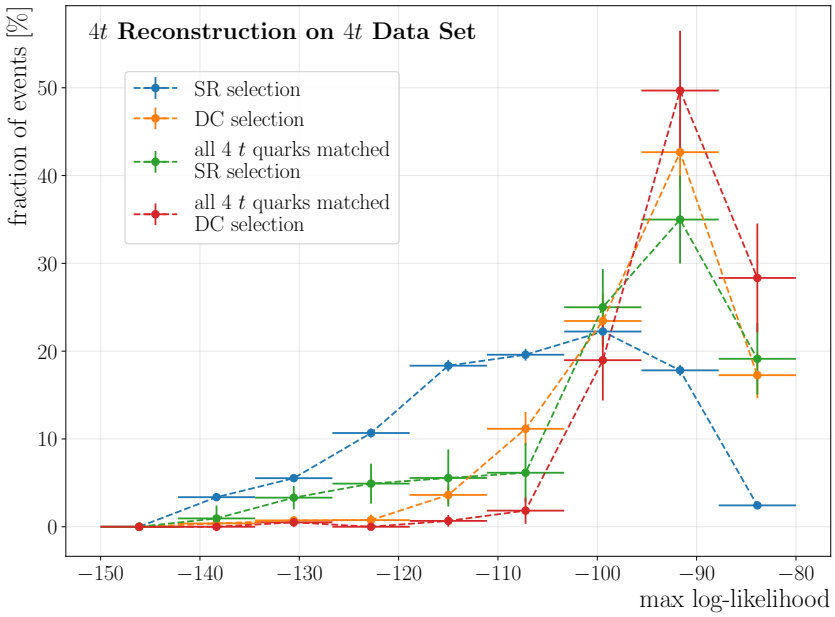


Figure 9.15: Distribution of the maximum of the log-likelihood values computed for all possible permutations of reconstructed objects for the event reconstruction using the four top quark topology on the signal events. Different colours represent different selections.

Unfortunately, the two distributions exhibit no clear difference in their shape, with a potential minuscule difference for values around a maximal log-likelihood of -90, which is the peak region of correctly reconstructed four top events (insinuated by the faded out distribution in green). This is especially the case for the last bin of the distribution.
However, due to the limited number of simulated $t\bar{t}$+jets events in this extreme region of the phase space that are available for this analysis, it is not clear that this difference does not simply originate from a statistical fluctuation.

Therefore, it is likely that this observable alone will not provide an advantage in discriminating the signal from background events, leaving it open for future analyses to investigate the viability of using other variables accessible through the even reconstruction, such as the kinematic properties of the model objects, for this task.

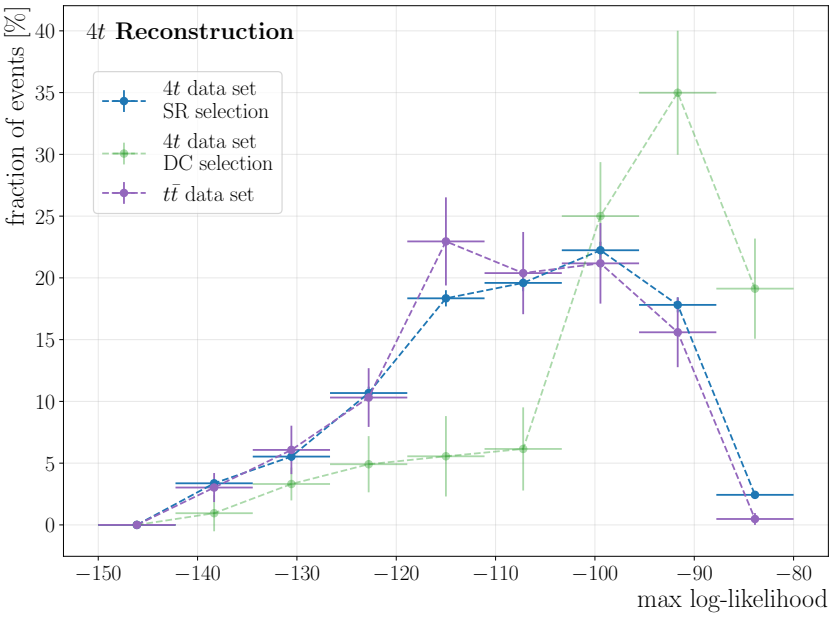


Figure 9.16: Distribution of the maximum of the log-likelihood values computed for all possible permutations of reconstructed objects for the event reconstruction using the four top quark topology on different data sets and selections.

9.7 Summary

The studies described in this chapter represent the first event reconstruction of the four top quark system in ATLAS using the likelihood method. In order to accomplish this, a tool set commonly used for the reconstruction of $t\bar{t}$ topologies, the KLFitter framework, has been used. In order to implement the four top quark topology into the tool, a series of readjustments were necessary. This included a thorough re-write of core modules of this framework as well as the implementation of the actual likelihood of the sought-after topology.

As a result, the method is indeed able to reconstruct four top quark events and correctly associated all four top quarks to the right permutation of reconstructed objects with an efficiency of $\varepsilon = 33 \pm 4\,\%$ for a best case event selection. This result is comparable to results previously achieved for $t\bar{t}$ events [3].

In addition, also the quality of the reconstruction in terms of kinematic properties has been evaluated, which is generally considered decent, with the exception of the energy of the reconstructed objects, warranting for additional studies.

Finally, the viability of using the outcome of the event reconstruction to discriminate signal from background events has been evaluated for the $t\bar{t}$+jets background and for the maximum of the likelihood values obtained for each of the possible probed permutations. These studies did not show a strong separation power for this particular variable, leaving room for additional studies.

Among these additional studies is certainly to use this variable in the training of a MVA technique to see its impact on a real life scenario. In this case, the studied variable can also be combined with all other observables obtained in the event reconstruction such as the kinematic properties. This has, for example, been done for the top mass measurements in $t\bar{t}$ events in [140].

Furthermore, it would be interesting to repeat the above studies with less draconian event selections, potentially even allowing for more or less reconstructed objects than expected from the event topology. This is not at all straightforward and quickly results in unmanageable amounts of possible permutations, so that new ways of pre selecting permutations would need to be devised.

However, the most interesting next step (for the author) would be to include additional decay modes of the four top quark system to eventually cover all the modes that are currently investigated in ATLAS. Here, an advantage of the decay mode resulting in two charged leptons, two neutrinos, four b and four light jets would be the considerably reduced amount of possible permutations. Unfortunately, this is achieved at the cost of having to disentangle the detector signature of two neutrinos, which is all but trivial.

The studies presented in this chapter conclude the investigations of four top quark processes conducted during this doctorate, so that concluding remarks of the combined studies presented in this document are given in the following.

10 Conclusion

The main analysis presented in this dissertation is the search for the Standard Model (SM) like simultaneous production of four top quarks. This analysis uses a data set recorded by the ATLAS experiment that corresponds to an integrated luminosity of $L = 139\,\text{fb}^{-1}$ of p–p collisions with a centre of mass energy of $\sqrt{s} = 13\,\text{TeV}$.
As the main result of this analysis, first evidence for the production of four top quarks is found. This represents the culmination of more than eight years of consecutive searches for this process at ATLAS, which were initiated in 2012 with a very strong contribution of the *Laboratoire de Physique de Clermont* (LPC).

The search presented here, targets events of the decay modes of the four top quark system with two same sign or more leptons in the final state. The analysis is published in [1] and it is soon to be complemented by studies exploring the one and two opposite sign lepton final states.
Following the evidence of four top quarks, the available data set is used to measure the production cross section $\sigma(pp \to t\bar{t}t\bar{t})$. The obtained result is [1]

$$\sigma(pp \to t\bar{t}t\bar{t}) = 24^{+7}_{-6}\,\text{fb},$$

which is compatible within 1.7 standard deviations with the latest SM prediction at Next-to-leading Order (NLO) in Quantum Chromo Dynamics (QCD) and electroweak (EW) contributions [2], given by $\sigma_{4t}^{\text{NLO}} = 12.0 \pm 2.4\,\text{fb}$. The statistical significance of rejecting the hypothesis that there is no four top quark production is

$$Z_{\text{obs}} = 4.3 \qquad \text{and} \qquad Z_{\text{exp}} = 2.4,$$

for the observed and expected data sets, respectively.

This result represents a considerably larger observed significance than the previously strongest results obtained by the CMS collaboration [45] focusing on the same final state topology. There, the cross section is measured to $\sigma_{4t} = 12.6^{+5.8}_{-5.2}\,\text{fb}$ with an observed significance of $Z_{\text{obs}} = 2.6$, while $Z_{\text{exp}} = 2.7$ were expected. The previous effort by the ATLAS collaboration using a data set of $L = 36.1\,\text{fb}^{-1}$, but combining the different final state topologies described above [46], set an observed (expected) upper limit on σ_{4t} of 5.3 (2.1) times the predicted cross section at NLO in QCD at the 95 % confidence level.

The search presented here, represents the fruit of the labour of a multitude of people and the author strongly contributed to the areas of analysis software, early lepton isolation optimisation, event classification in the simulation and extensive studies for building and benchmarking the global fit model and overall analysis strategy.

Following this first evidence, there are many ways to use measurements of the four top quark system to probe parameters of the SM or of Beyond Standard Model (BSM) scenarios. Among these will be more in depth studies of the properties of the four top quarks, such as energy spectra or spin correlations as well as significantly more precise measurements of the inclusive or even differential cross section. This last point is particularly interesting following a simple extrapolation of the analysis in chapter 7 to a simulated data set corresponding to $L = 4000\,\mathrm{fb}^{-1}$, which corresponds to the design luminosity of the High-Luminosity Large-Hadron-Collider (HL-LHC) phase. This extrapolation shows an expected statistical only uncertainty of less than 7 % on the measured cross section. Compared to the statistical uncertainty of approximately 30 % obtained with the simulation corresponding to the full Run-II data set, and especially compared to the uncertainty on the SM prediction of around 20 %, this will constitute a great advancement for the field.

Consequently, a series of studies were conducted in a second part of this doctorate that precisely look at the tasks and tools needed to continue searching for four top quarks. This includes studies on the performance of a proposed upgrade of ATLAS for HL-LHC conditions as well as on advanced techniques for the reconstruction of the four top quark system.

The studies were done in part in cooperation with the *Universidade do Estado do Rio de Janeiro*, financially supported by the CAPES–COFECUB [141] programme between the governments of Brazil and France and as part of a cotutorship programme with TU Dortmund university, which was supported by a grant from the German-French university [142].
Except for the development of a tool for visualising the impact of different reconstruction algorithms on final physics objects, these studies are the exclusive work of the author.

In more detail, studies were conducted on the expected peformance of the reconstruction of muons only using the information of the inner detector and the calorimeters of the ATLAS detector under the conditions of the HL-LHC phase. For this reason, the impact of arteficial noise and pile-up scenarios on the reconstruction process were tested. Here, the goal of the noise scenarios is to emulate, among other effects, a loss of acceptance due to ageing detector components as well as losses in resolution for the same reason and due to higher occupancy levels expected for this phase.
These studies represent a first estimation of these effects while more in depth studies, where the entire reconstruction procedure is repeated for different noise scenarios, would be in order. This, however, would require too much expensive CPU time, so that the presented studies represent a compromise between feasability and precision. As a result, these studies show that the calorimetry system can be used for the reconstruction of muons during the HL-LHC phase even before tuning the algorithms to these new conditions. This is particularly interesting for any measurement with large numbers of muons, such as future measurements of four top quark events.

Finally, and as mentioned above, a more immediate way of improving and progressing future measurements of four top quark production, namely the event reconstruction using the kinematic likelihood fit approach is studied. This approach is implemented in the KLFitter tool set and has previously been very successfully used for the reconstruction of $t\bar{t}$ pairs. In order to assess the possibility to reconstruct the four top quark system, the author implemented the likelihood that corresponds to the decay mode with exactly one lepton in the final state. This particular topology results in a very large number of possible permutations of physics objects that need to be probed for every suitable event. As a consequence, it has been neccessary to rewrite and optimise significant parts of the core software of the KLFitter tool set.
Applying the method on a simulated data set of four top quark production, an event reconstruction efficiency of $\epsilon = 33 \pm 4\,\%$ is achieved under optimal conditions. This is comparable to results obtained for the reconstruction of $t\bar{t}$ pairs, which is an important result given that the four top quark topology requires four instead of two top quarks to be correctly reconstructed. To the knowledge of the author, this study represents the first attempt to reconstruct the four top quark system with a likelihood based method and the fact that it has been successful in doing so opens the door to a host of new areas to explore.

Among these new opportunities is the possibility to use the additional information obtained through the event reconstruction in order to discriminate signal from background events. A study of one of these variables, the maximal likelihood, is therefore

tested for the discrimination between $4t$ and $t\bar{t}$ events, where no strong discrimination power was found. However, with growing data sets and using multiple variables and sophisticated machine-learning techniques, this possibility remains an interesting opportunity.

In summation, the author has been heavily involved in obtaining the first evidence of four top quark production, while also opening the way for a large variety of future studies involving four top quarks, be it in terms of an upgraded detector or using previously unexplored techniques of reconstructing the four top quark system. The author generously leaves these exciting future studies to the *geneigte Leser*[1].

[1] A difficult to translate concept used by university professors to burden their students with homework. The closest translation might be "the gentle reader".

Acknowledgements

These three years would not have been possible without the help of numerous people whom I thank dearly.

First and foremost I would like to thank the Université Clermont Auvergne, CAPES-COFECUB and DFH for providing the financial support for the various projects of this doctorate.
Following that, I wholeheartedly thank Romain Madar for being a highly motivated and incredibly knowledgeable advisor who spent nights and holidays helping me. Naturally, I would also like to thank Kevin Kröninger, for giving me some perspective, for getting me back home for a spell and for reading this document in its entirety. However, I'm also slightly disappointed that promises made in [9] to take me back to Corsica have so far not been fulfilled[2]. I also want to thank Frédéric Deliot and Andreas Meyer for combing through this, admittedly quite long, document and I want to thank all members of the Jury for taking so much of your time for me.

I am deeply grateful to have met Lina during this time and do not have the words to describe my gratitude for countless Duo calls, vegetarian care packages, beautiful canoe trips and her love. I am really looking forward to the next stage of our life together.

Further, I would like to thank my family for unlimited support, visits, distractions and for overall being the best family I could have hoped for. To name all would further significantly enlarge the number of pages of this document, so I just want to explicitly thank my parents, Karl and Ute, and Claudia and Ralf for getting me this far and for always being there for me. I sincerely wished Karl was still here to celebrate with us. I also want to explicitly thank my siblings. Janina, for the support and good advice and Pascal for the support, great discussions, amazing cooking sessions and of course the circular saw. Additionally, I would like to thank Hannah, Paul, Luna and Tino for showing me a bright future and especially for putting everything into perspective when things got out of control. You are amazing and I'm really looking forward to see you grow and flourish. This thanks definitely includes my Argentinian family, who have shown me a beautiful continent and significantly shaped my future. Special thanks go to Alfredo for unknowingly giving me the last big push to dedicate (parts of) my life to science. I'm particularly happy to find my sister Diana and my cousin Jorgelina in

[2] Aber heut ist ja noch nicht aller Tage Abend.

Germany when I return.

Besides family, there were also numerous good friends who helped me through the difficult periods of this doctorate. There are of course my dear "colleagues" from Dortmund: Gregor, Peter, Johannes, Philipp and Elena, as well as my creative, energetic and overall amazing cultural advisor Sophie. There is of course Wiebke, who provided me with good recipes, good times in the *puys* and with *die Blätter* on so many occasions. There are my beautiful *Peruvian* friends Stefan, Lisa and Farah, who are always there to support me and discuss politics and recent events (from) all over the world.
Infinite thanks are in order for Megan, the entire Carnal family, Nico and Matthieu, who are the main reason my stay in France has been a great success by fully accepting me into their lives. Thank you for the barbecues, the wine, the hikes, the enticing philosophical discussions and so much more.
Even though I have been absent a lot in the last period of writing this thesis, I would like to thank the infamous group of *non-permanents* of the LPC, who are a great bunch of interesting people. Thank you for yoga, beer and oh so much Greek and Turkish food. Equally infamous, I also want to cherish my time with the *expats* at Clermont, most notably with Treasa and Karolina. Thank you Sergej and Alisa, for our interesting discussions, for being such kind people and for the beginning of our shared kayak experiences.

Thanks go to the ATLAS teams at Clermont and Dortmund for simply everything as well as to some amazing people in the administration at both institutes, most notably Andrea, Marine and Florence. Among the team in Clermont, I sincerely want to thank my partner in confinement and thesis writing, Nazlim, without whom two months of isolation would have been much harder to bear.
Thank you Bernardo and Elisa for taking complete care of me for a month and for welcoming me into your country and families.

Finally, I would really like to mention so many more people here, but it is definitely much better to just thank you in person!

Bibliography

[1] ATLAS Collaboration. 'Evidence for $t\bar{t}t\bar{t}$ production in the multilepton final state in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector'. *Eur. Phys. J. C 80* (2020) (cit. on pp. I–III, 103, 129, 131, 138, 149, 217).

[2] Rikkert Frederix et al. 'Large NLO corrections in $t\bar{t}W^{\pm}$ and $t\bar{t}t\bar{t}$ hadroproduction from supposedly subleading EW contributions'. *JHEP* 02 (2018), 31 (cit. on pp. I–III, 19, 130 sq., 217).

[3] Johannes Erdmann et al. 'A likelihood-based reconstruction algorithm for top-quark pairs and the KLFitter framework'. *Nucl. Instrum. Meth. A* 748 (2014) (cit. on pp. I–III, 192, 194, 200, 213).

[4] ATLAS Collaboration. 'Evidence for $t\bar{t}t\bar{t}$ production in the multilepton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector'. ATLAS-CONF-2020-013. CERN, 2020 (cit. on pp. 7, 98, 107).

[5] Sheldon L. Glashow. 'Partial-symmetries of weak interactions'. *Nucl. Phys.* 22.4 (1961), 579 (cit. on p. 9).

[6] Steven Weinberg. 'A Model of Leptons'. *Phys. Rev. Lett.* 19 (1967), 1264 (cit. on p. 9).

[7] Abdus Salam. 'Weak and Electromagnetic Interactions'. *Conf.Proc.* C680519 (1968), 367 (cit. on p. 9).

[8] Steven Weinberg. 'Implications of dynamical symmetry breaking'. *Phys. Rev. D* 13 (1976) (cit. on p. 9).

[9] Lennart Rustige. 'Studies for the background estimation to the search for FCNCs in single top quark + photon production with the ATLAS experiment at $\sqrt{s} = 13\,\text{TeV}$'. Master Thesis. TU Dortmund, 2017 (cit. on pp. 9, 223).

[10] The Super-Kamiokande Collaboration. 'Evidence for Oscillation of Atmospheric Neutrinos'. *Phys. Rev. Lett.* 81 (1998) (cit. on p. 11).

[11] Nicola Cabibbo. 'Unitary Symmetry and Leptonic Decays'. *Phys. Rev. Lett.* 10 (1963), 531 (cit. on p. 11).

[12] Makoto Kobayashi and Toshihide Maskawa. 'CP-Violation in the Renormalizable Theory of Weak Interaction'. *Prog. Theor. Phys.* 49 (1973), 652 (cit. on p. 11).

[13] Kevin Kröninger et al. 'Top-Quark Physics at the LHC'. 2015, 35 (cit. on p. 13).

[14] P.A. Zyla et al. (Particle Data Group). '2020 Review of Particle Physics'. *Prog. Theor. Exp. Phys.* 0831C01 (2020) (cit. on pp. 13, 16, 51).

[15] CDF Collaboration. 'Evidence for top quark production in $\bar{p}p$ collisions at $\sqrt{s} =$ 1.8 TeV'. *Phys. Rev. Lett.* 73 (1994) (cit. on p. 13).

[16] D0 Collaboration. 'Observation of the top quark'. *Phys. Rev. Lett.* 74 (1995) (cit. on p. 13).

[17] CDF Collaboration. 'Observation of top quark production in pp collisions with the collider detector at fermilab'. *Phys. Rev. Lett.* 74 (1995) (cit. on p. 13).

[18] Jon Butterworth et al. 'PDF4LHC recommendations for LHC Run II'. *J. Phys. G* 43 (2016), 23001 (cit. on pp. 14, 127, 130).

[19] H L Lai et al. 'Global QCD analysis of parton structure of the nucleon: CTEQ5 parton distributions'. *Eur. Phys. J. C* 12 (2000), 375 (cit. on p. 14).

[20] J Pumplin et al. 'New generation of parton distributions with uncertainties from global QCD analysis'. *JHEP* 07 (2002), 12 (cit. on p. 14).

[21] A Sherstnev and R S Thorne. 'Parton Distributions for LO Generators'. *Eur. Phys. J. C* 55 (2008), 553 (cit. on p. 14).

[22] Matteo Cacciari et al. 'Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation'. *Phys. Lett. B* 710 (2011) (cit. on p. 14).

[23] Stefano Catani et al. 'Top-quark pair hadroproduction at next-to-next-to-leading order in QCD'. *Phys. Rev. D* 99 (2019) (cit. on p. 14).

[24] ATLAS Collaboration. 'Measurement of the $t\bar{t}$ production cross-section in the lepton+jets channel at $\sqrt{s} = 13$ TeV with the ATLAS experiment'. *Submitted to: Phys. Lett. B* (2020) (cit. on p. 14).

[25] M. Aliev et al. 'HATHOR, HAdronic Top and Heavy quarks crOss section calculatoR'. *Comp. Phys. Comm.* 182 (2010) (cit. on p. 14).

[26] P. Kant et al. 'HATHOR for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions'. *Comp. Phys. Comm.* 191 (2014) (cit. on p. 14).

[27] ATLAS and CMS Collaborations. NLO single-top channel cross sections. URL: `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopRefXsec` (visited on 03/07/2020) (cit. on p. 14).

[28] Nikolaos Kidonakis. 'Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^-'. *Phys. Rev. D* 82 (2010) (cit. on p. 14).

[29] ATLAS Collaboration. 'Measurement of the inclusive cross-sections of single top-quark and top-antiquark t-channel production in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector'. *JHEP* 4 (2017) (cit. on p. 15).

[30] ATLAS Collaboration. 'Measurement of the cross-section for producing a W boson in association with a single top quark in pp collisions at $\sqrt{s} = 13$ TeV with ATLAS'. *JHEP* 1 (2018) (cit. on p. 15).

[31] ATLAS Collaboration. 'Evidence for single top-quark production in the s-channel in proton–proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector using the Matrix Element Method'. *Phys. Lett. B* 756 (2016) (cit. on p. 15).

[32] 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* 1407 (2014), 79 (cit. on pp. 15, 99).

[33] NNPDF Collaboration, R.D. Ball et al. 'Parton distributions for the LHC Run II'. *JHEP* 04 (2015), 40 (cit. on pp. 15, 99–101).

[34] M. Aaboud et al. 'Combinations of single-top-quark production cross-section measurements and $|f_{\mathrm{LV}}V_{tb}|$ determinations at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS experiments'. *JHEP* 5 (2019) (cit. on p. 15).

[35] G. Bevilacqua and M. Worek. 'Constraining BSM physics at the LHC: four top final states with NLO accuracy in perturbative QCD'. *JHEP* 2012.7 (2012) (cit. on p. 19).

[36] Qing-Hong Cao et al. 'Probing Higgs Width and Top Quark Yukawa Coupling from $\bar{t}tH$ and $\bar{t}t\bar{t}t$ Productions' (2016) (cit. on p. 19).

[37] H. P. Nilles. 'Supersymmetry, Supergravity and Particle Physics'. *Phys. Rept.* 110 (1984) (cit. on p. 19).

[38] G. R. Farrar and P. Fayet. 'Phenomenology of the production, decay, and detection of new hadronic states associated with supersymmetry'. *Phys. Lett. B* 76 (1978) (cit. on p. 19).

[39] D. Dicus et al. 'Higgs decay to top quarks at hadron colliders'. *Phys. Lett. B* 333 (1994) (cit. on p. 19).

[40] Nathaniel Craig et al. 'The Hunt for the Rest of the Higgs Bosons'. *JHEP* 06 (2015) (cit. on p. 19).

[41] T. Plehn and T. M. P. Tait. 'Seeking sgluons'. *J. Phys. G* 36 (2009) (cit. on p. 19).

[42] Giacomo Cacciapaglia et al. 'Four tops on the real projective plane at LHC'. *JHEP* 10 (2011) (cit. on p. 19).

[43] ATLAS Collaboration. 'Search for new phenomena in events with same-charge leptons and b-jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector'. *JHEP* 12 (2018) (cit. on pp. 19, 21, 94, 104, 147).

[44] D. de Florian et al. 'Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector'. 2017 (cit. on p. 23).

[45] CMS Collaboration. 'Search for production of four top quarks in final states with same-sign or multiple leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV'. *Eur. Phys. J. C* 80 (2020) (cit. on pp. 24, 217).

[46] ATLAS Collaboration. 'Search for four-top-quark production in the single-lepton and opposite-sign dilepton final states in p–p collisions at $\sqrt{s} = 13TeV$ with the ATLAS detector'. *Phys. Rev. D* 99.052009 (2019) (cit. on pp. 25, 186, 217).

[47] Oliver Sim Brüning et al. 'LHC Design Report Vol. I: The LHC Main Ring'. 1.CERN-2004-003 (2004) (cit. on p. 27).

[48] A Barachetti et al. 'Final Project Report: High Luminosity Large Hadron Collider Design Study'. CERN-ACC-2016-0007 (2016) (cit. on p. 27).

[49] ALICE Collaboration. 'The ALICE experiment at the CERN LHC'. *JINST* 3 (2008), S08002 (cit. on p. 27).

[50] ATLAS Collaboration. 'The ATLAS Experiment at the CERN Large Hadron Collider'. *JINST* 3 (2008), S08003 (cit. on pp. 27, 32 sq., 35 sq.).

[51] CMS Collaboration. 'The CMS experiment at the CERN LHC'. *JINST* 3 (2008), S08004 (cit. on p. 27).

[52] LHCb Collaboration. 'The LHCb Detector at the LHC'. *JINST* 3 (2008), S08005 (cit. on p. 27).

[53] ATLAS Collaboration. Luminosity Run-II. URL: `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2` (visited on 10/07/2020) (cit. on pp. 28, 38).

[54] ATLAS Collaboration. 'Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC'. *Phys. Lett. B* 716 (2012) (cit. on pp. 28, 30).

[55] CMS Collaboration. 'Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC'. *Phys. Lett. B* 716 (2012) (cit. on pp. 28, 30).

[56] Esma Mobs and CERN. The CERN accelerator complex. URL: `https://cds.cern.ch/record/2636343` (visited on 09/07/2020) (cit. on p. 29).

[57] CERN. LHC / HL-LHC Plan. 2020. URL: `https://hilumilhc.web.cern.ch/content/hl-lhc-project` (visited on 05/07/2020) (cit. on p. 29).

[58] ATLAS Collaboration. The Collaboration | ATLAS Experiment at CERN. URL: `https://atlas.cern/discover/collaboration` (visited on 08/07/2020) (cit. on p. 32).

[59] ATLAS Collaboration. 'ATLAS Insertable B-Layer Technical Design Report'. CERN-LHCC-2010-013. 2010 (cit. on p. 33).

[60] ATLAS Tile Calorimeter Collaboration. 'The optical instrumentation of the ATLAS Tile Calorimeter'. *JINST* 8 (2013), P01005 (cit. on p. 35).

[61] V. Cindro et al. 'The ATLAS Beam Conditions Monitor'. *JINST* 3 (2008), P02004 (cit. on p. 37).

[62] G Avoni et al. 'The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS'. *JINST* 13 (2018), P07017 (cit. on pp. 37, 98, 125).

[63] Peter Jenni et al. 'ATLAS Forward Detectors for Measurement of Elastic Scattering and Luminosity ATLAS Collaboration Technical Design Report'. CERN-LHCC-2008-004. 2008 (cit. on pp. 37 sq.).

[64] ATLAS Collaboration. 'Technical Design Report for the ATLAS Inner Tracker Strip Detector'. CERN-LHCC-2017-005. 2017 (cit. on pp. 38 sq.).

[65] ATLAS Collaboration. 'Technical Design Report for the Phase-II Upgrade of the ATLAS Tile Calorimeter'. CERN-LHCC-2017-019. 2017 (cit. on pp. 38, 71, 77).

[66] ATLAS Collaboration. 'Technical Design Report for the ATLAS Inner Tracker Pixel Detector'. CERN-LHCC-2017-021. 2017 (cit. on p. 38).

[67] ATLAS Collaboration. 'Letter of Intent for the Phase-II Upgrade of the ATLAS Experiment'. CERN-LHCC-2012-022. 2012 (cit. on p. 39).

[68] ATLAS Collaboration. 'Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2'. *Eur. Phys. J. C* 77 (2017), 673 (cit. on p. 43).

[69] R. Frühwirth. 'Application of Kalman filtering to track and vertex fitting'. *Nucl. Instrum. Meth. A* 262 (1987), 444 (cit. on p. 43).

[70] T. G. Cornelissen et al. 'The global χ^2 track fitter in ATLAS'. *J. Phys. Conf. Ser.* 119 (2008), 032013 (cit. on p. 44).

[71] ATLAS Collaboration. 'Vertex Reconstruction Performance of the ATLAS Detector at $\sqrt{s} = 13\,\text{TeV}$'. ATL-PHYS-PUB-2015-026. 2015 (cit. on p. 44).

[72] ATLAS Collaboration. 'Reconstruction of primary vertices at the ATLAS experiment in Run$\sim$1 proton–proton collisions at the LHC'. *Eur. Phys. J. C* 77 (2017), 332 (cit. on p. 44).

[73] G. Piacquadio et al. 'Primary vertex reconstruction in the ATLAS experiment at LHC'. *Phys. Conf. Ser.* 119 (2008), 032033 (cit. on p. 44).

[74] S. Hageböck and E. Von Toerne. 'Medical imaging inspired vertex reconstruction at LHC'. *Phys. Conf. Ser.* 396 (2012), 022021 (cit. on p. 44).

[75] ATLAS Collaboration. 'Electron and photon performance measurements with the ATLAS detector using the 2015-2017 LHC proton–proton collision data'. *JINST* 14 (2019), P12006 (cit. on pp. 45 sq., 125).

[76] ATLAS Collaboration. 'Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at $\sqrt{s} = 13\,\text{TeV}$'. *Eur. Phys. J. C* 79 (2019), 639 (cit. on pp. 45, 48 sq., 92).

[77] ATLAS Collaboration. 'Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1'. *Eur. Phys. J. C* 77 (2017), 490 (cit. on p. 46).

[78] ATLAS Collaboration. 'Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung'. ATLAS-CONF-2012-047. 2012 (cit. on p. 46).

[79] ATLAS Collaboration. 'Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton–proton collision data'. *JINST* 14 (2019), P03017 (cit. on p. 47).

[80] ATLAS Collaboration. 'Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run 2 data collected in 2015 and 2016'. *Eur. Phys. J. C* 79 (2019), 205 (cit. on pp. 47, 49).

[81] ATLAS Collaboration. 'Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$=13 TeV'. *Eur. Phys. J. C* 76 (2016) (cit. on pp. 49–51, 125).

[82] ATLAS Collaboration. 'Jet reconstruction and performance using particle flow with the ATLAS Detector'. *Eur. Phys. J. C* 77 (2017), 466 (cit. on p. 51).

[83] Matteo Cacciari et al. 'The anti-k_t jet clustering algorithm'. *JHEP* 04 (Feb. 2008), 63 (cit. on p. 52).

[84] Matteo Cacciari et al. 'FastJet user manual' (2011) (cit. on p. 52).

[85] ATLAS Collaboration. 'Quark versus Gluon Jet Tagging Using Charged-Particle Constituent Multiplicity with the ATLAS Detector'. ATL-PHYS-PUB-2017-009. 2017 (cit. on p. 52).

[86] ATLAS Collaboration. 'Quark versus Gluon Jet Tagging Using Jet Images with the ATLAS Detector'. ATL-PHYS-PUB-2017-017. 2017 (cit. on p. 52).

[87] Markus Stoye and on behalf of the CMS collaboration. 'Deep learning in jet reconstruction at CMS'. *J. Phys. Conf. Ser.* 1085 (2018), 042029 (cit. on p. 52).

[88] CMS Collaboration. 'Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV'. *JINST* 13 (2018), P05011 (cit. on p. 52).

[89] ATLAS Collaboration. 'Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run'. ATL-PHYS-PUB-2017-013. 2017 (cit. on pp. 52 sq.).

[90] ATLAS Collaboration. 'Performance of b-jet identification in the ATLAS experiment'. *JINST* 11 (2016), P04008 (cit. on p. 53).

[91] Andreas Höcker et al. 'TMVA - Toolkit for Multivariate Data Analysis'. CERN-OPEN-2007-007. 2007 (cit. on p. 53).

[92] ATLAS Collaboration. 'Expected performance of the ATLAS b-tagging algorithms in Run-2'. ATL-PHYS-PUB-2015-022. 2015 (cit. on p. 53).

[93] Ingo Burmeister. 'Measuring the b-tagging efficiency in ATLAS data at $\sqrt{s} = 8$ TeV and $\sqrt{s} = 13$ TeV using the $p_{\mathrm{T}}^{\mathrm{rel}}$-method'. PhD Thesis. TU Dortmund, 2017 (cit. on p. 54).

[94] ATLAS Collaboration. 'ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV'. *Eur. Phys. J. C* 79 (2019) (cit. on p. 54).

[95] ATLAS Collaboration. '$E_{\mathrm{T}}^{\mathrm{miss}}$ performance in the ATLAS detector using 2015-2016 LHC pp collisions'. ATLAS-CONF-2018-023. 2018 (cit. on p. 55).

[96] ATLAS Collaboration. 'Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ with the ATLAS detector'. *Phys. Rev. D* 96 (2017), 72002 (cit. on pp. 56, 199).

[97] D Adams et al. 'Recommendations of the Physics Objects and Analysis Harmonisation Study Groups 2014'. Internal to ATLAS. ATL-PHYS-INT-2014-018. 2014 (cit. on p. 56).

[98] Lennart Rustige and Romain Madar. 'Study of noise and efficiency loss effects on calorimeter-tagged muons for HL-LHC'. Internal to ATLAS. 2019 (cit. on p. 59).

[99] M. Tanabashi et al. (Particle Data Group). 'Passage of particles through matter'. *Phys. Rev. D* 98 (2018), 030001 (cit. on pp. 61–63).

[100] A Salzburger. 'The ATLAS Track Extrapolation Package'. ATL-SOFT-PUB-2007-005 (2007) (cit. on p. 64).

[101] Gustavo Ordonez Sanz. 'Muon identification in the ATLAS calorimeters'. PhD thesis. Nijmegen U., 2009 (cit. on pp. 64, 68).

[102] Luciano M. de A. Filho et al. 'Calorimeter Response Deconvolution for Energy Estimation in High-Luminosity Conditions'. *IEEE Trans. on Nucl. S.* 62 (2015) (cit. on pp. 65, 82).

[103] E Fullana et al. 'Optimal Filtering in the ATLAS Hadronic Tile Calorimeter'. ATL-COM-TILECAL-2005-001. 2005 (cit. on p. 65).

[104] ATLAS Collaboration. The MCTruthClassifier Tool. URL: `https://twiki.cern.ch/twiki/bin/view/AtlasProtected/MCTruthClassifier` (visited on 02/06/2020) (cit. on pp. 66, 105).

[105] Paolo Nason. 'A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms'. *JHEP* 11 (2004) (cit. on pp. 67, 100).

[106] Torbjorn Sjöstrand et al. 'A Brief Introduction to PYTHIA 8.1'. *Comput. Phys. Commun.* 178 (2008) (cit. on pp. 67, 238).

[107] F. James and M. Roos. 'Minuit - a system for function minimization and analysis of the parameter errors and correlations'. *Comp. Phys. Comm.* 10.6 (1975) (cit. on pp. 76, 88).

[108] Rene Brun and Fons Rademakers. 'ROOT: An object oriented data analysis framework'. *Nucl. Instrum. Meth. A* 389 (1997) (cit. on p. 76).

[109] Wouter Verkerke and David Kirkby. 'The RooFit toolkit for data modeling' (2003) (cit. on p. 88).

[110] ATLAS Collaboration. The TRExFitter Framework. URL: `https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/TtHFitter` (visited on 10/07/2020) (cit. on p. 88).

[111] Glen Cowan et al. 'Asymptotic formulae for likelihood-based tests of new physics'. *Eur. Phys. J.* C71 (2010) (cit. on pp. 90 sq.).

[112] Frederick E James. 'Statistical Methods in Experimental Physics'. 2nd. Singapore : World Scientific, 2006, 345 (cit. on p. 91).

[113] ATLAS Collaboration. Tagging and suppression of pileup jets with the ATLAS detector. ATLAS-CONF-2014-018. 2014 (cit. on pp. 93, 126).

[114] ATLAS Collaboration. Number of Interactions per Crossing 2018. 2020. URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2 (visited on 13/04/2020) (cit. on pp. 98 sq.).

[115] Stefano Frixione et al. 'Angular correlations of lepton pairs from vector boson and top quark decays in Monte Carlo simulations'. *JHEP* 04 (2007) (cit. on p. 99).

[116] Pierre Artoisenet et al. 'Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations'. *JHEP* 03 (2013) (cit. on p. 99).

[117] Torbjörn Sjöstrand et al. 'An Introduction to PYTHIA 8.2'. *Comput. Phys. Commun.* 191 (2015) (cit. on p. 99).

[118] ATLAS Collaboration. 'ATLAS Pythia 8 tunes to 7 TeV datas'. ATL-PHYS-PUB-2014-021. 2014 (cit. on p. 99).

[119] D. J. Lange. 'The EvtGen particle decay simulation package'. *Nucl. Instrum. Meth. A* 462 (2001), 152 (cit. on pp. 100, 238).

[120] ATLAS Collaboration. 'The new ATLAS Fast Calorimeter Simulation'. *J. Phys. Conf. Ser.* 898 (2017), 042006 (cit. on p. 100).

[121] M. Bahr et al. 'Herwig++ Physics and Manual'. *Eur. Phys. J. C* 58 (2008) (cit. on p. 100).

[122] Johannes Bellm et al. 'Herwig 7.0/Herwig++ 3.0 release note'. *Eur. Phys. J. C* 76 (2016) (cit. on p. 100).

[123] L A Harland-Lang et al. 'Parton distributions in the LHC era: MMHT 2014 PDFs'. *Eur. Phys. J. C* 75 (2015) (cit. on p. 100).

[124] Stefano Frixione et al. 'A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction'. *JHEP* 09 (2007) (cit. on p. 100).

[125] Stefano Frixione et al. 'Matching NLO QCD computations with Parton Shower simulations: the POWHEG method'. *JHEP* 11 (2007) (cit. on p. 100).

[126] Simone Alioli et al. 'A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX'. *JHEP* 1006 (2010) (cit. on p. 100).

[127] ATLAS Collaboration. 'Studies on top-quark Monte Carlo modelling for Top2016'. ATL-PHYS-PUB-2016-020. 2016 (cit. on p. 100).

[128] GEANT4 Collaboration. 'Geant4—a simulation toolkit'. *Nucl. Instrum. Meth. A* 506 (2003) (cit. on pp. 100, 237).

[129] Enrico Bothmann et al. 'Event Generation with Sherpa 2.2'. *SciPost Phys.* 7 (2019) (cit. on p. 101).

[130] F. Cascioli et al. 'Scattering Amplitudes with Open Loops'. *Phys. Rev. Lett.* 108 (2012), 111601 (cit. on p. 101).

[131] Ansgar Denner et al. 'Collier: a fortran-based Complex One-Loop LIbrary in Extended Regularizations'. *Comput. Phys. Commun.* 212 (2017) (cit. on p. 101).

[132] ATLAS Collaboration. 'Analysis of $t\bar{t}H$ and $t\bar{t}W$ production in multilepton final states with the ATLAS detector'. *12th International Workshop on Top Quark Physics, Beijing, China* (2019) (cit. on pp. 102, 105, 109 sq., 129, 132).

[133] Rikkert Frederix and Ioannis Tsinikos. 'Subleading EW corrections and spin-correlation effects in $t\bar{t}W$ multi-lepton signatures' (2020) (cit. on pp. 102, 128, 132, 148).

[134] Leo Breiman. 'Bagging predictors'. *Machine Learning* 24 (1996) (cit. on p. 117).

[135] J. H. Friedman. 'Data analysis techniques for high energy particle physics'. *Proceedings of 3rd CERN School of Computing* (1974) (cit. on p. 123).

[136] ATLAS Collaboration. 'Luminosity determination in pp collisions at $\sqrt{s} = 8\,\text{TeV}$ using the ATLAS detector at the LHC'. *Eur. Phys. J. C* 76 (2016) (cit. on p. 125).

[137] ATLAS Collaboration. 'Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 13$ TeV'. *Eur. Phys. J. C* 78 (2018) (cit. on p. 126).

[138] ATLAS Collaboration. 'Measurements of inclusive and differential fiducial cross-sections of $t\bar{t}$ production with additional heavy-flavour jets in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector'. *JHEP* 04 (2019) (cit. on p. 130).

[139] Allen Caldwell et al. 'BAT: The Bayesian Analysis Toolkit'. *Comp. Phys. Comm.* 180 (2009) (cit. on p. 192).

[140] ATLAS Collaboration. 'Measurement of the top quark mass in the $t\bar{t} \to$ lepton+jets channel from $\sqrt{s} = 8$ TeV ATLAS data and combination with previous results'. *Eur. Phys. J. C* 79 (2019) (cit. on pp. 193, 195, 214).

[141] Campus France. Le Programme CAPES-COFECUB 2020. URL: `https://www.campusfrance.org/en/node/2338` (visited on 10/07/2020) (cit. on p. 218).

[142] DFH. German French University. URL: `https://www.dfh-ufa.org/` (visited on 10/07/2020) (cit. on p. 218).

[143] Johan Alwall et al. 'MadGraph 5 : Going Beyond'. *JHEP* 06 (2011) (cit. on p. 238).

Appendix

A Addendum to chapter 6

A.I Computing Setup and Data Sets

After the generation and showering of simulated particles using the dedicated tools, their interaction with the detector needs to be emulated. For ATLAS, this can be achieved using a full simulation called FullSim that employs GEANT4 [128].
All data sets used for this study use the FullSim approach. The result of the detector response simulation is saved in *HITS* files (see Figure A.1). The *HITS* files do not contain any information about pile-up conditions, so particular minimum bias data sets need to be overlaid onto the *HITS* files depending on the desired pile-up scenario before reconstruction. Using this combination of data sets, the full reconstruction can be done, leading to Event Summary Data (ESD) files that contain the final reconstructed physics objects as well as some more detailed detector information that was needed in the reconstruction process. In order to further reduce the size of the data sets for final storage, ESD files are filtered so that they only contain information used by most physics analysis groups, leading to Analysis Object Data (xAOD) files. To assure a quick turn around for analysis teams and to further reduce data set sizes to a level where advanced data analysis techniques can be applied, different ATLAS sub groups can define further filtering, resulting in the most commonly used data format, Derived Analysis Object Data (DxAOD).
The information needed for the track extrapolation and for retrieving the energy depositions in the calorimeter are stored in ESD files and removed in later steps, which is why this study is mostly based on this file format.

Table A.1: List of Monte Carlo (MC) samples used for the reconstruction efficiency study

Process	Gen. Events	$\langle \mu_{\mathrm{pu}} \rangle$
$Z \to \mu^- \mu^+$	400k	190-210
$Z \to \mu^- \mu^+$	400k	130-150
$Z \to \mu^- \mu^+$	400k	70-90

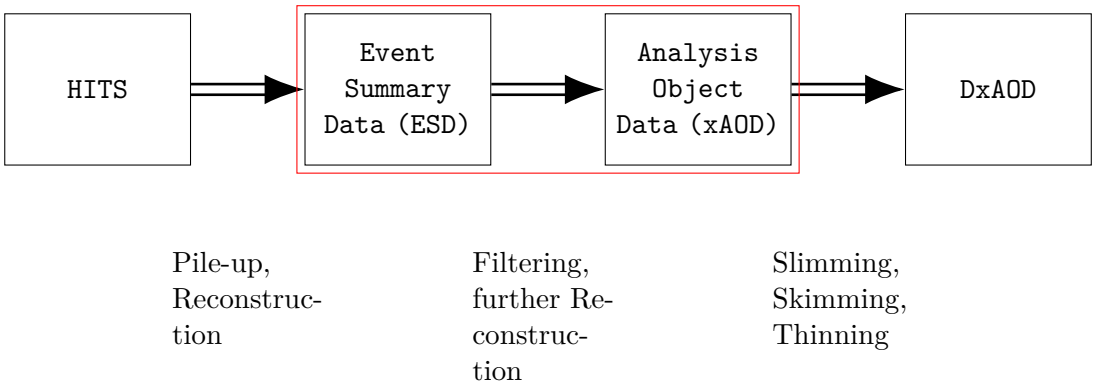


Figure A.1: Chain of different MC file formats and the corresponding steps applied in between. Only the two file formats in the red box are directly used for this study.

A.II Four Top Quark MC Sample

In order to estimate the impact of muon reconstruction on the search for four top quarks, a sample is produced that represents the LO process $pp \to t\bar{t}t\bar{t}$ at a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$, with a simulated cross section of $\sigma = 9.2\,\text{fb}$ and which is reconstructed with estimated pile-up conditions for the 2017 data taking period of ATLAS. The MC event generation is done using MadGraph [143], the showering is achieved using Pythia 8 [106], where the decay of b-hadrons is processed by EvtGen [119]. For these samples, muons are required to pass the *Loose* identification and have $p_{\text{T}} > 15\,\text{GeV}$, but there is no requirement on the isolation of the muons.

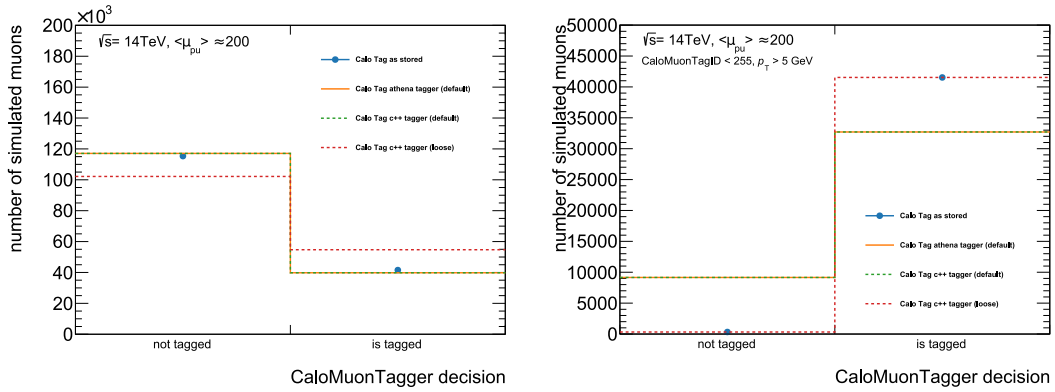
A.III Validation of Standalone CaloMuonTagger

As has been mentioned in subsection 6.2.3, the CaloMuonTagger has been reimplemented as a standalone C++ tool for the purpose of this analysis. The reimplemented tagger corresponds to the CaloMuonTagger of the release 20.7 series and has multiple defined Working Points (WPs). The cut values of the working point called *Loose* in the code base can be found in Table A.2, another working point, called *Default* in the code base, is also defined although it is not the default for physics analysis.
In order to validate the reimplementation, the decision of this reimplementation is compared to the CaloMuonTagger decision which is stored in the data set during the main reconstruction procedure. Additionally, it is also directly compared to the decision obtained by re-running the athena-based CaloMuonTagger, which should - in principle - have the same value as the one stored in the data set. The comparison of the different approaches and WPs are shown in Figure A.2, where the blue dot represents the decision as stored, the orange line represents the decision of re-running the standard (athena-based) implementation of the CaloMuonTagger with the *De-*

Table A.2: Cut values for the *Loose* working point of the CaloMuonTagger. The Sig cut values require the energy deposit to be higher than the cut value, the Veto cut values require the energy deposits to be lower. The $p_{\mathrm{T,min}}$ is chosen for muons with $p_{\mathrm{T}} < 15\,\mathrm{GeV}$ and the $p_{\mathrm{T,max}}$ for muons with $p_{\mathrm{T}} > 35\,\mathrm{GeV}$. For muon momenta in between, a linear interpolation between the cut values is applied.

Calorimeter Sample	Sig [MeV]		Veto [MeV]	
	$p_{\mathrm{T,min}}$ cut	$p_{\mathrm{T,max}}$ cut	$p_{\mathrm{T,min}}$ cut	$p_{\mathrm{T,max}}$ cut
EMB1	-	-	400	800
EMB2	-	-	700	1400
EMB3	-	-	500	600
EME1	-	-	1100	1300
EME2	-	-	2200	2100
EME3	-	-	1100	1100
HEC0	-50	-1100	2500	9000
HEC1	-50	-1100	3000	9000
HEC2	0	-1100	3000	9000
HEC3	150	-1100	3000	9000
TileBar0	0	-400	3500	8000
TileBar1	100	-1500	3500	8000
TileBar2	0	-400	3500	8000
TileExt0	-50	-500	3000	7000
TileExt1	200	-600	3500	9000
TileExt2	150	-1200	3500	9000

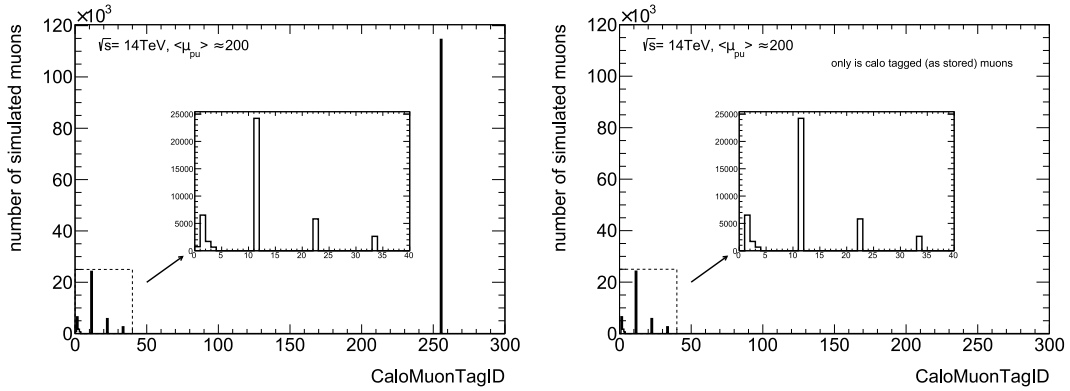
fault working point, and the dashed lines represent the decision of the re-implemented CaloMuonTagger for the *Default* (green) and the *Loose* (red) working point.
The athena-based implementation agrees fully with the re-implemented version, but neither of those agree with the value stored in the data set. Additionally, it is interesting to note that most muons are not classed as tagged for this configuration, which could potentially correspond to a very poor reconstruction efficiency of the tool. After applying a cut on the muon's p_{T} of $p_{\mathrm{T}} > 5\,\mathrm{GeV}$ and a cut on the *CaloMuonTagID* of *CaloMuonTagID* < 255, the conclusion changes significantly and the decision stored in the data set now corresponds mostly with the re-implemented tagger using the *Loose* working point and almost all of the muons passing the selection are tagged. The *CaloMuonTagID* is the direct output of the CaloMuonTagger and according to the code base it should only assume values in [0,3], where 0 corresponds to "not calo tagged" and 1-3 to "is calo tagged". However, looking at the *CaloMuonTagID* values stored in the data set (see Figure A.3), it often equals 255. The full value of *CaloMuonTagID* is a combination of the individual IDs for both WPs, where *CaloMuonTagID* = loose ID + 10× default ID.



(a) without additional cuts (b) with additional cuts

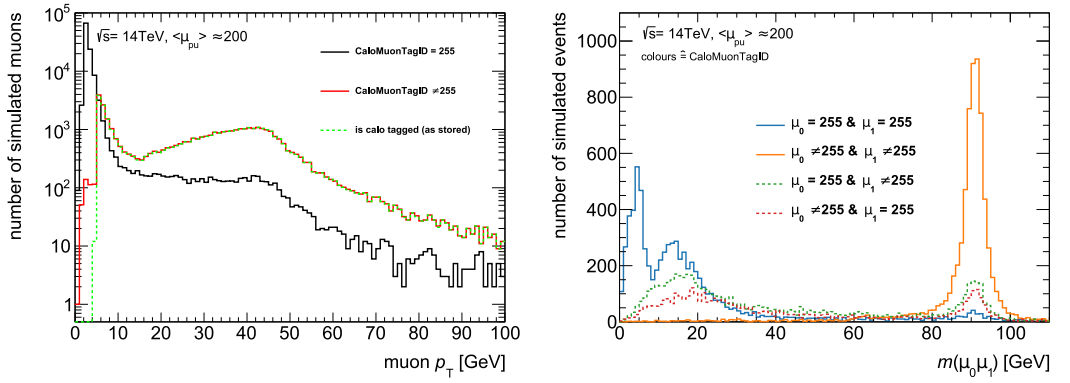
Figure A.2: Tagging decision of the CaloMuonTagger for different implementations and WPs of the tool. The blue markers represent the decision as obtained during the initial reconstruction and stored in the data set, the orange line represents a re-run of the official athena-based tagger using the *default* working point and the dashed lines show the decision for the re-implemented standalone C++ CaloMuonTagger for the *loose* working point (red) and the *default* working point (green).

Figure A.3b shows the *CaloMuonTagID* distribution for all muons that are tagged by the CaloMuonTagger according to the value stored in the data set. It shows that indeed none of these muons have $ID = 0$, so that this variable is indeed connected to the final decision, and none of these muons have a value of $ID = 255$. Investigating further, Figure A.4 shows the p_{T} distribution of the muons in Figure A.4a and the invariant mass of the p_{T}-leading and p_{T}-subleading muons in Figure A.4b for different *CaloMuonTagID* configurations. The p_{T} distribution shows, that the muons that passed the CaloMuonTagger (as stored) also require a p_{T} of $p_{\mathrm{T}} > 5\,\mathrm{GeV}$ and that it mostly follows the *CaloMuonTagID* $\neq 255$ for higher momenta. Muons classed as *CaloMuonTagID* $= 255$ mostly have momenta of $p_{\mathrm{T}} < 5\,\mathrm{GeV}$.
The invariant mass distribution shows that the p_{T}-leading and subleading muons are mostly prompt muons originating from a Z boson if their *CaloMuonTagID* $\neq 255$. If one of the muons has *CaloMuonTagID* $= 255$, the invariant mass goes to lower values peaking at $m_{\ell\ell} \approx 15\,\mathrm{GeV}$ and $m_{\ell\ell} \approx 5\,\mathrm{GeV}$. If those muons are not only non-prompt but also fake remains to be studied, but the goal of this validation was to find a working point proving that the reimplementation of the CaloMuonTagger was successful, which has been achieved. Further studies therefore remain as an outlook.



(a) (b)

Figure A.3: CaloMuonTagID as stored in the data set for all reconstructed muons (left) or only for those tagged as muons by the *CaloMuonTagger* - also as stored in the data set (right).



(a) muon p_{T}

(b) invariant mass of p_{T}-leading and subleading muon

Figure A.4: The p_{T} distribution and the distribution of the invariant mass of the p_{T}-leading and subleading muon for different values of the *CaloMuonTagID* (for calo tagged muons).

B Addendum to chapter 7

B.I Results

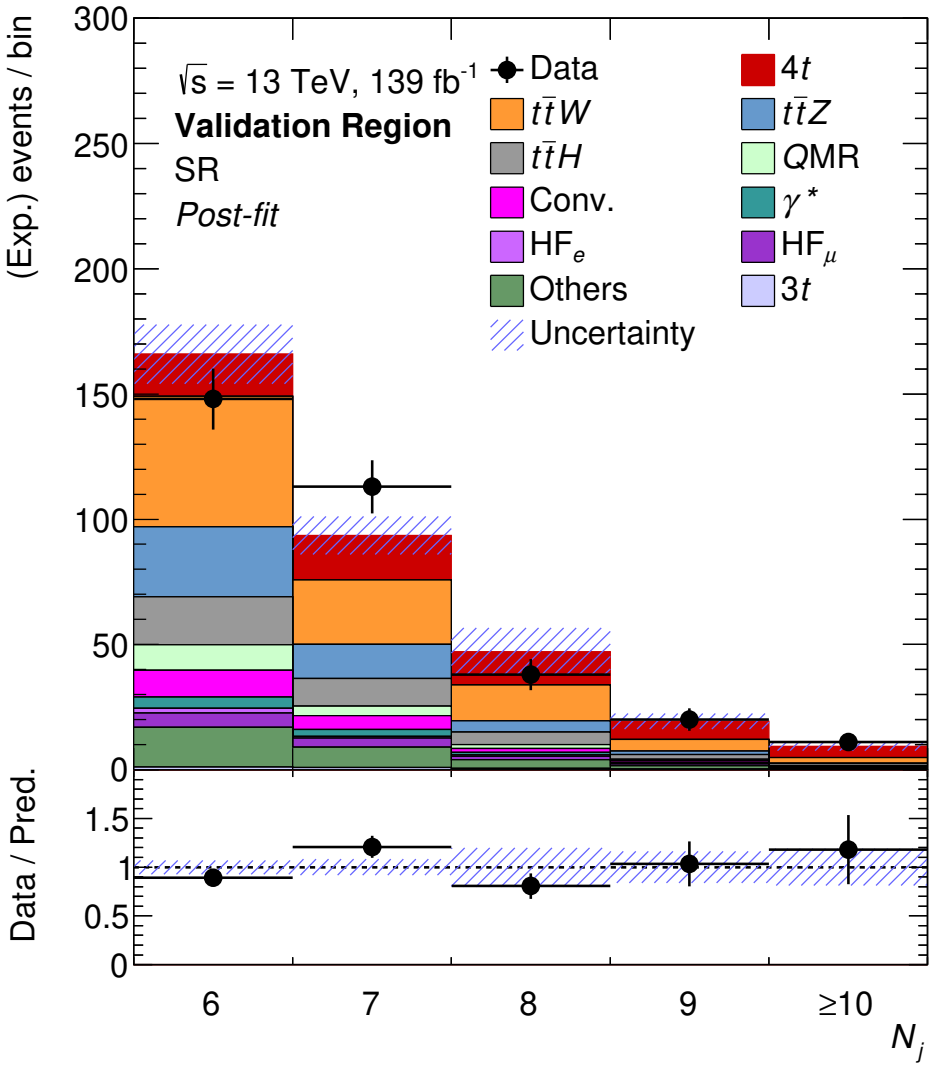


Figure B.1: Distributions of the jet multiplicity for the pre template fit estimation (coloured areas) and data (black points) in the Signal Region (SR). The uncertainty band includes statistical as well as systematic uncertainties. The denomination "Validation" indicates that this distribution is not explicitly used in the fit.

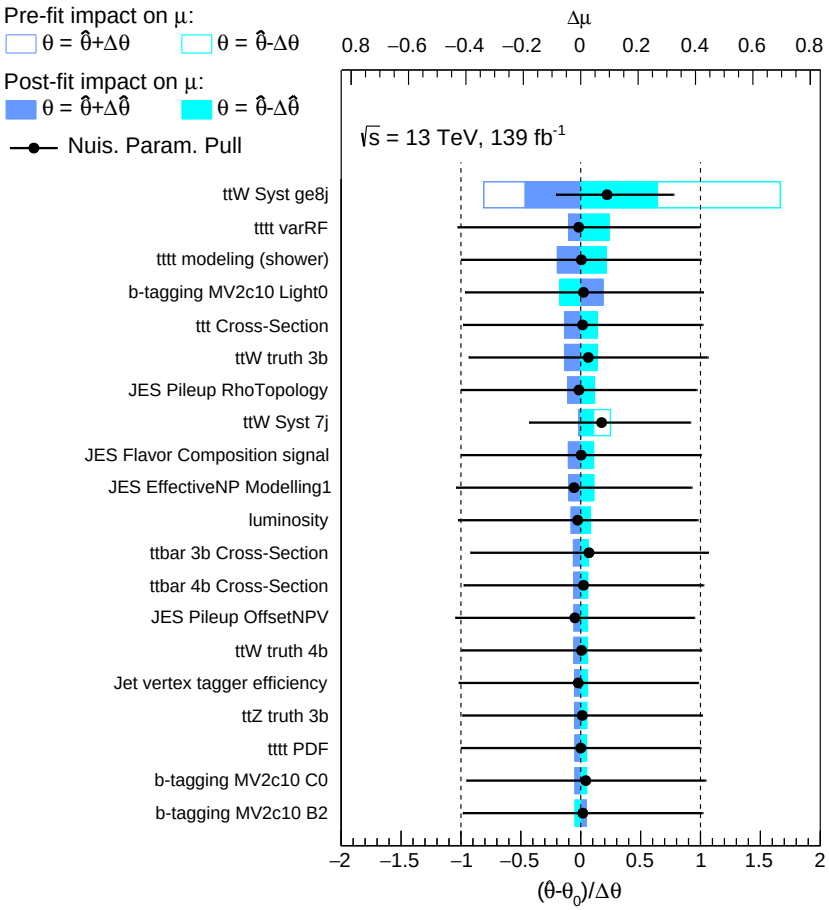


Figure B.2: Ranking of systematic uncertainties excluding the uncertainty on the prediction of the four top quark production cross section.

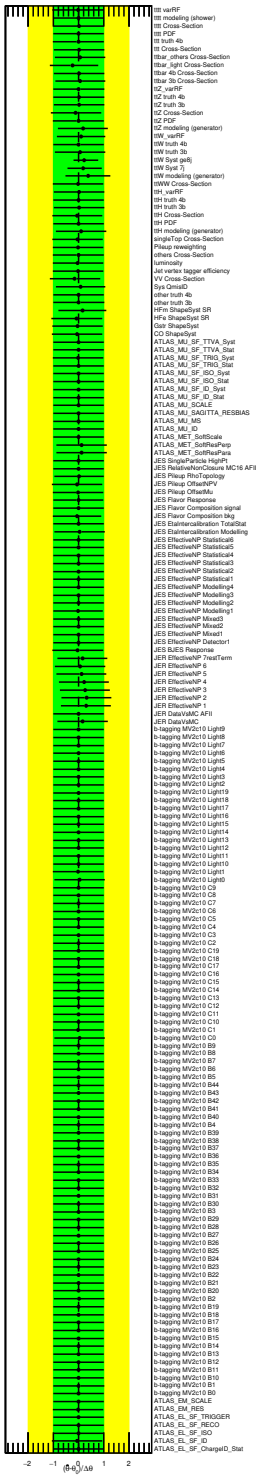


Figure B.3: Post fit Nuisance Parameters excluding the uncertainty on the prediction of the four top quark production cross section.

C Addendum to chapter 8

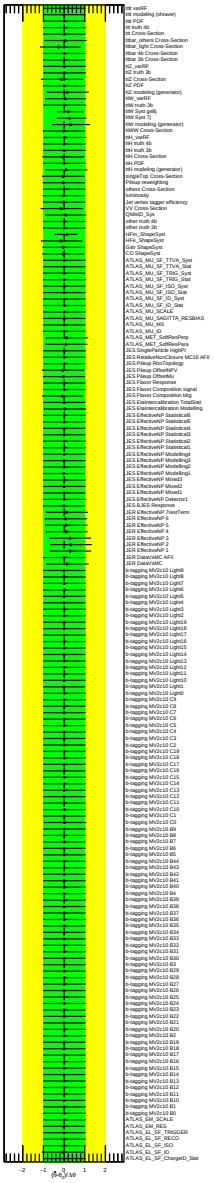


Figure C.4: Post proto fit Nuisance Parameters excluding the uncertainty on the prediction of the four top quark production cross section.

D Addendum to chapter 9

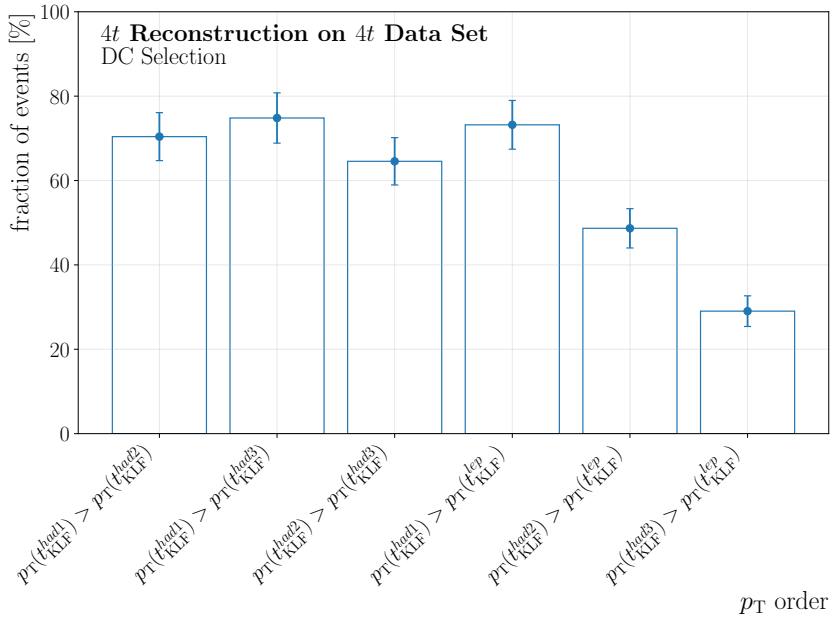


Figure D.5: Boolean conditions of p_T ordering of model top quarks.

Glossary

AFII AtlFast II. 100

ALFA Absolute Luminosity For ATLAS. 37

ANN Artifical Neural Network. 43, 53

AUC Area under the Curve. 119, 120

BCM Beam Conditions Monitor. 37

BDT Boosted Decision Tree. 53, 89, 92, 96, 100, 115, 116, 117, 118, 119, 120, 121, 122, 124, 125, 128, 138, 139, 140, 143, 144, 145, 146, 147, 155, 157, 159, 160, 168, 169, 171, 175, 178

BSM Beyond Standard Model. 3, 4, 9, 13, 19, 20, 30, 185, 218

CB Combined. 49, 50

CDF Cumulative Density Function. 79, 91

CERN European Organization for Nuclear Research. 3, 27

CKM Cabibbo-Kobayashi-Maskaw. 11, 15

COF Constrained Optimal Filter. 82, 83, 84

CR Control Region. 88, 94, 104, 109, 110, 111, 113, 115, 116, 131, 134, 143, 147, 148, 153, 154, 155, 156, 157, 168, 170, 171, 172, 173, 174, 175, 178, 179, 180, 181, 182

CSC Cathode-Strip Chamber. 35, 36

CT Calorimeter tagged. 50, 56, 59, 60, 64, 66, 68, 69, 75, 81, 82

CV Conversion Vertex. 110

DC Direct Correspondance. 197, 200, 201, 202, 211

DxAOD Derived Analysis Object Data. 237

EFT Effective Field Theory. 19

EM Electromagnetic Calorimeter. 32, 33, 34, 45, 46, 47, 48, 83, 92

ESD Event Summary Data. 237

EW electroweak. 10, 14, 19, 20, 25, 102, 131, 132, 217

FCal Forward Calorimeter. 34

FOM Figure Of Merit. 82, 95, 119, 120, 132, 165

HEC Hadronic End-Cap. 34

HEP High Energy Physics. 3, 4, 13, 122

HL-LHC High-Luminosity Large-Hadron-Collider. I, II, III, 4, 5, 27, 28, 29, 38, 39, 59, 67, 69, 81, 82, 218, 219

HLT High-Level Trigger. 37

IBL Insertable B-Layer. 33

ID Inner Detector. 32, 33, 35, 36, 37, 38, 43, 45, 46, 47, 49, 50, 56, 57, 64, 68, 69, 92, 125

IFF Isolation and Fakes Forum. 105

IP Interaction Point. 28, 30, 32, 37, 39, 40, 43, 50, 54, 64

IR Iterative Removal. 119, 120, 121

ITK Inner Tracker. 39, 67

JES Jet Energy Scale. 56, 126

JVT Jet-Vertex-Tagger. 93, 126

LHC Large Hadron Collider. 3, 4, 5, 13, 27, 28, 29, 30, 32, 35, 37, 38, 49, 51, 52, 61, 62, 98

LO Leading Order. 14, 15, 16, 20, 22, 23, 99, 100, 101, 102, 118, 119, 196

LPC *Laboratoire de Physique de Clermont*. 217

LS Long-Shutdown. 33, 38

LUCID-2 Luminosity Measurement Using Cerenkov Integrating Detector-2. 37

MC Monte Carlo. 54, 67, 68, 71, 87, 89, 100, 102, 104, 105, 114, 118, 120, 122, 125, 126, 127, 129, 154, 237, 238

MDT Monitored Drift Tube. 35, 36

ME Extapolated. 50

MIP Minimum Ionising Particle. 35, 49, 50

MS Muon System. 32, 35, 36, 37, 43, 49, 50

MVA Multivariate Analysis. 118, 143, 147, 214

NF Norm Factor. 88, 102, 104, 110, 111, 112, 131, 132, 133, 134, 138, 143, 144, 146, 147, 148, 149, 150, 155, 162, 163, 168, 169, 170, 171, 173, 175, 176, 179, 180, 181, 182

NLO Next-to-leading Order. 14, 19, 25, 99, 100, 101, 102, 118, 119, 131, 196, 217

NNLL Next-to-next-to-leading logarithmic. 14

NNLO Next-to-next-to-leading Order. 14, 101

NP Nuisance Parameter. 88, 110, 115, 122, 129, 130, 155, 162, 168

OF Optimal Filtering. 65, 83, 84

OR Overlap Removal. 50, 55, 56, 57

PA Plain Asimov pseudo data set. 154, 155, 156

PDF Parton Distribution Function. 13, 14, 15, 99, 100, 101

PDF Probabiliy Density Function. 47, 90, 91, 127, 128, 130, 193

PF Proto Factor. 169, 170, 171, 173

PMT Photo-multiplier-tube. 34, 37, 38, 64

PoI Parameter of Interest. 87, 88, 112

PV Primary Vertex. 45, 53, 110

QCD Quantum Chromo Dynamics. 14, 16, 19, 20, 25, 51, 100, 102, 131, 217

RA Realistic Asimov pseudo data set. 154, 155, 156, 169, 170

ROC Receiver operating Characteristic. 119

RPC Resistive Plate Chamber. 36

SCT Silicon Microstrip Detector. 33, 43

SF Scale Factor. 54, 125, 126

SM Standard Model. I, II, III, 3, 4, 5, 9, 10, 11, 12, 13, 19, 20, 24, 30, 87, 131, 134, 138, 143, 144, 145, 185, 217, 218

SR Signal Region. 23, 24, 89, 94, 95, 96, 97, 102, 103, 109, 111, 115, 116, 120, 122, 124, 129, 138, 140, 141, 142, 143, 144, 145, 147, 148, 149, 153, 154, 155, 156, 157, 159, 160, 168, 169, 170, 171, 172, 173, 174, 175, 178, 179, 180, 181, 182, 197, 201, 211, 242

ST Segment tagged. 49, 50

TDAQ Trigger and Data Acquisition. 37, 38

TF Transfer Function. 190, 191, 193, 194, 203

TGC Thin Gap Chamber. 36

TMVA Toolkit for Multivariate Analysis. 53, 117

TRT Transition Radiation Tracker. 33

TTVA Track-To-Vertex-Association. 125

VR Validation Region. 115, 116, 139

WP Working Point. 47, 48, 49, 50, 51, 53, 68, 82, 92, 93, 238, 239, 240

xAOD Analysis Object Data. 237