Universität Bonn

Physikalisches Institut

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 at the LHC

Ö. Oğul Öncel

The first evidence for simultaneous production of four top quarks is presented. The measurement is conducted by selecting events having two same-charge leptons or at least three leptons and by analysing 139 fb^{-1} of proton–proton collision data at a centre-of-mass energy of $\sqrt{s}$ = 13 TeV provided by the LHC and collected by the ATLAS detector. In addition to simulated samples, data-driven methods are utilised for the estimation of background contributions. A multivariate discriminant is developed to separate $t\bar{t}t\bar{t}$ signal from the background processes using the Boosted Decision Tree (BDT) method. The production cross-section is extracted using a binned profile-likelihood fit to all control regions and the BDT distribution in the signal region simultaneously and is measured to be

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

corresponding to an observed (expected) statistical significance of 4.3 (2.9) standard deviations from the background-only hypothesis. The analysis model is used to set limits on the top-quark Yukawa coupling parameter y_t. A pure CP-even (CP-odd) coupling is excluded for $|y_t| > 2.2$ ($|y_t| > 1.6$) at the 95% confidence level. No deviations from the Standard Model of particle physics prediction has been found.

Physikalisches Institut der
Universität Bonn
Nussallee 12
53115 Bonn
Germany

January 2025

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 at the LHC

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

von

Ö. Oğul Öncel

aus

Çankaya, Ankara — Türkei

Bonn, 2024

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

Gutachter/Betreuer:	Prof. Dr. Markus Cristinziani
Gutachter:	Prof. Dr. Jochen Dingfelder
Tag der Promotion:	27.8.2024
Erscheinungsjahr:	2025

“A hundred times every day I remind myself that my inner and outer life depend on the labours of other men, living and dead, and that I must exert myself in order to give in the same measure as I have received and am still receiving.”

“One thing I have learned in a long life: that all our science, measured against reality, is primitive and childlike – and yet it is the most precious thing we have.”

Albert Einstein

Yaşamak şakaya gelmez,
büyük bir ciddiyetle yaşayacaksın
bir sincap gibi mesela,
yani, yaşamanın dışında ve ötesinde hiçbir şey beklemeden,
yani bütün işin gücün yaşamak olacak.

Yaşamayı ciddiye alacaksın,
yani o derecede, öylesine ki,
mesela, kolların bağlı arkadan, sırtın duvarda,
yahut kocaman gözlüklerin,
beyaz gömleğinle bir laboratuvarda
insanlar için ölebileceksin,
hem de yüzünü bile görmediğin insanlar için,
hem de hiç kimse seni buna zorlamamışken,
hem de en güzel en gerçek şeyin
yaşamak olduğunu bildiğin halde.

Yani, öylesine ciddiye alacaksın ki yaşamayı,
yetmişinde bile, mesela, zeytin dikeceksin,
hem de öyle çocuklara falan kalır diye değil,
ölmekten korktuğun halde ölüme inanmadığın için,
yaşamak yani ağır bastığından...

...Bu dünya soğuyacak,
yıldızların arasında bir yıldız,
hem de en ufacıklarından,
mavi kadifede bir yaldız zerresi yani,
yani bu koskocaman dünyamız.

Bu dünya soğuyacak günün birinde,
hatta bir buz yığını
yahut ölü bir bulut gibi de değil,
boş bir ceviz gibi yuvarlanacak
zifiri karanlıkta uçsuz bucaksız.

Şimdiden çekilecek acısı bunun,
duyulacak mahzunluğu şimdiden.
Böylesine sevilecek bu dünya
‘Yaşadım’ diyebilmen için...

Nâzım Hikmet Ran

Acknowledgements

Aside from the training in scientific discipline that was the goal of the doctorate, it also taught me how dependent achieving any result is on a large network of collaboration in this field. Any attempt to capture the importance of collaboration and how much a person's work and success depend on the efforts and generosity of many others will inevitably be an understatement. My time as a doctoral student at the Physics Institute in Bonn coincided with various events of specific and general relevance such as the 200^{th} anniversary of the university, the construction of the Forschungs- und Technologiezentrum Detektorphysik (FTD), the promotion of my supervisor to full professorship and his subsequent departure to Siegen, the retirement of our decades-long institute director, Prof. Dr. Norbert Wermes, and the COVID-19 pandemic. A large number of people have supported me in various ways and at different stages, allowing this work to continue and be completed. Therefore, I feel not only entitled but also compelled to generously acknowledge many of them here.

First and foremost, I would like to express my profound gratitude to my *Doktorvater*, Prof. Dr. Markus Cristinziani. His guidance was indispensable to the successful completion of my research project. He played a major role in shaping my scientific thinking, as well as my ability to navigate the sociological structures of academia in Germany and internationally. He encouraged and enabled me to participate in multiple graduate schools, workshops and conferences during these years, enriching my research experience and education. At times of personal difficulties, he approached me with compassion and understanding. He was always available for discussions, giving advice and assistance, treating me with respect and fairness. His careful reading and detailed feedback has significantly increased the quality of this dissertation, as it did with all other important presentations and publications I have worked on during my studies.

My heartfelt thanks go to my second supervisor, Prof. Dr. Jochen Dingfelder. He supported me financially in the final months of my PhD, provided valuable insights and advice whenever I sought them, and was always kind and approachable. His cheerful and optimistic personality uplifted not only me but also others, making the institute a more comfortable and enjoyable workplace. I would also like to acknowledge his calm and composed leadership during the COVID-19 pandemic, as well as his support for institute members during that challenging time.

My sincere thanks to Prof. Dr. Christiane Dahl and Prof. Dr. Carsten Urbach for agreeing to serve on my doctoral examination committee despite the tight timeline during the lecture-free period. Their availability and accommodating attitude ensured that my doctoral colloquium could proceed without significant delays.

I had a truly pleasant and productive time in the Cristinziani working group, surrounded by friendly, fun, and competent colleagues. Like many others, I started by running away but ended up running after Vadim Kostyukhin, whose deep and well-structured thinking saved the day on multiple occasions and inspired many ideas. It was a great pleasure to work alongside and befriend the indispensable Italian core of the group, Andrea Sciandra and Carlo Gottardo, who taught me a lot about particle physics and computing while generously sharing their experiences and advice. Sebastian Heer, Niklas Schwan, and Matei Climescu have been wonderful companions both in and out of the office, becoming good friends providing me with a steady supply of jokes and encouragement on top of lot of physics talk.

I spent several beautiful years at the Physics Institute in Bonn until the onset of the pandemic. Christopher Deutsch, Christian Nass, Shubham Bansal, Roman Küsters, Ruth Jacobs, Emily Thompson, Christos Vergis, and David Hohn provided valuable help with various technical topics, as well as occasional banter in the kitchen and offices on the third floor, making the workplace more enjoyable. Tatjana Lenz and Eckhardt von Toerne offered insightful comments on my presentations and studies, with their careful attention and expertise. I am grateful to Prof. Dr. Norbert Wermes for including me in his weekly group meetings, which allowed me to benefit from his and his group's invaluable feedback. Peter Falke was always ready to listen, assist, and advise on any topic I brought up, for which I am very thankful. Oliver Freyermuth provided unparalleled IT support, enabling me to parallelize thousands of BDT jobs and saving me countless hours of work. Andrea Fürstenberg, who signed my initial contract on my behalf, was always friendly and helpful in navigating the intricacies of paperwork and German bureaucracy. Barbara Valeriani-Kaminski was an exceptional organizer and communicator for the laboratory courses I tutored, ensuring everything ran smoothly.

During my short time at the Physics Institute in Siegen, I was warmly welcomed and had the pleasure of exchanging meaningful conversations with members of the Cristinziani working group. Thank you, Wolfgang Walkowiak, Katharina Voss, Florian Fischer, Tongbin Zhao, Beatrice Cervato, and Gabriel Gomes, for your kindness and the enjoyable moments we shared.

In my day-to-day work, I collaborated closely with the four top quarks analysis team from various ATLAS member institutions. By the time I joined, the team had already been working on the analysis for over a year. Romain Madar helped me integrate smoothly and quickly contribute to the project by assigning me responsibilities and entrusting me with important tasks. Frédéric Deliot steered and organized the team towards the publication in his role as analysis contact and personally supported me on many occasions. I worked closely with and learned a great deal from Lennart Rustige and Zhi Zheng, who were responsive, friendly, and helpful throughout. I would also like to thank the TopWG convener at the time, Prof. Dr. Wolfgang Wagner, for his support in general and also personally.

The writing process for this dissertation extended several years beyond the principal research phase, making my three-year-long research associate position in Freiburg an integral part of the story. I am deeply grateful to our group leader, Prof. Dr. Markus Schumacher, for welcoming me into his group and tolerating my unexpectedly lengthy writing process, which was not initially planned to take so long. Christina Skorek, with her ever-kind and sympathetic demeanor, helped me from the very beginning to the very end with all the necessary paperwork and administrative matters. My gratitude and appreciation run deep for Valerie Lang, my office mate and the person I interacted with the most during those three years. She was an exceptional mentor whose moral

compass, work ethic, and logical reasoning were constant sources of inspiration. Lorenzo Rossini brought fresh energy to our analysis team with his cheerfulness and optimism, effectively ending my chronic "great-to-be-around-Italian-colleague withdrawal syndrome" after Bonn. I truly appreciated the synergy we shared and will fondly remember how every disagreement turned into a fun memory. I thank my doctoral student collaborators, Daniel Bahner and Naman Kumar Bhalla, for showing understanding when my creative-engine ran out of steam at times due to preoccupation with finishing my thesis, for the great times and kindness they brought to our almost daily interactions. I learned a great deal from you both and thoroughly enjoyed working alongside you. It has also been a pleasure discussing and learning about the theorist's perspective of our work from Yann Stoll. Engaging in enriching discussions about physics and other intellectual topics with Kartik Bhide was always a highlight. Furthermore, I extend my thanks to Anton J. Gamel for *Märchenstunden*, Michael Böhler, Dirk Sammel, and Benjamin Rottler for their technical support and engaging conversations, and Katharina Schleicher for her compassion.

One of my closest ATLAS colleagues during my Freiburg years was Erlend Aakvaag, a responsive and kind collaborator. I also greatly appreciated the many engaging conversations we shared on non-academic topics. I am also very happy to have gotten to know Alessio Pizzini and Alice Alfonsi, with whom I had some of the most insightful and focused discussions, which I thoroughly enjoyed. Thank you to Antonio de Maria for being an inspirational role model, demonstrating that kindness, modesty, and competence can coexist even at the highest levels and during the busiest of days. I deeply appreciate the support and guidance of ATLAS giants Zach Marshall and Tomas Dado, who provided major assistance and valuable exchanges during my learning of event generator tools and TRExFitter, respectively. I benefited greatly from working with Jonathan Shlomi, who introduced me to important aspects of the Athena software and b-tagging tools, enabling me to quickly make progress on my Qualification Task. I would also like to thank my ATLAS colleagues for the fruitful, friendly, and productive times we shared, particularly Nedaa-Alexandra Asbah, Alessandra Betti, Kathrin Becker, Quentin Buat, Andrea Knue, Elin Bergeås Kuutmann, Ki Lie, Michaela Mlynarikova, Gianantonio Pezzullo, and Duc Bao Ta.

My special thanks go to the people who helped me review the drafts, provided feedback on my colloquium presentation, and organized various aspects of the day of my defense. I would like to mention them and reiterate my gratitude: Peter and Wolfgang reviewed several chapters of my thesis. Nikolaus Owtscharenko reviewed additional chapters and provided valuable feedback on my rehearsal. Sebastian offered constructive feedback on my rehearsal and helped organize logistics on the day of the colloquium. Christian assisted me with arranging, setting up, and booking rooms as well as organizing anything related to the event. Naman, Kartik, Yann and Lorenzo suggested important changes that improved my presentation. Finally, I extend my heartfelt thanks to Matei for reviewing most of my thesis, providing detailed feedback on both the draft and the presentation, and supporting me in person with rehearsals and preparations the day before the defense.

I had the good fortune to come across Anıl, Mehmet, and Melek in my early days in Cologne. I am deeply grateful for their friendship and for always extending a helping hand whenever a need arose. Their presence provided me with a sense of family and security as I adjusted to living far from home. I count myself incredibly lucky to have my good friends Beyza, Duygu, and Selcan, who were always there to listen and support me—no matter the topic, time, or place. As fellow travelers through the slippery labyrinths of academia, Özgür and Ceylan patiently endured my vents and generously offered advice based on their experiences. I am truly thankful for their compassion and genuine interest in my progress. My short-term flatmate and friend Chiara supported me with excellent food, endless encouragement, and intellectually stimulating philosophical discussions that brought me immense joy. Thank you Melanie and Julia for helping me navigate life in Germany, for not giving up on making me a part of your lives despite my occasional unavailability, and for the many wonderful and memorable moments we shared during these years.

I thank all members of the Demirtaş family for their endless support in every way possible, always given with kindness and without hesitation. Their presence has been a constant source of empowerment, facilitating my progress and bringing happiness since my childhood. My gratitude goes to the Damschen family for their love and care. Much of this thesis was written while I was their guest, enjoying delicious food, engaging in pleasant conversations, going to *spazieren* or short trips, and being reminded that everything will turn out well. In addition to being a great listener, my brother Çağlar patiently reassured me and made sure that I always have the big picture in my mind and put things into perspective. Thank you to my cousins, Kemal and Begam, for always checking in on me, and boosting my morale when I needed it. I also thank Çağla for attending my defense and supporting me on this important day. Lastly, I would like to acknowledge my teachers from earlier stages of my education, who had a significant impact on my academic development; Vicdan İnceoğlu, Gökhan Aldemir, Hakan Uzun, Prof. Dr. Ceyhun Bulutay and Prof. Dr. Tuğrul Hakioğlu.

The person who suffered the most from the day-to-day progress of my research is undoubtedly Amelie. She not only endured my physical and mental unavailability due to work but also selflessly gave of herself to keep me afloat—ensuring I remained functional, optimistic, cheerful, and healthy. She spent significant energy organizing the little time and strength I had left after work so that we could enjoy meaningful moments together. Without her efforts, my memories from years of doctoral research would not have the beautiful, warm, and colorful texture they have now. I am grateful for her love, compassion, and kindness.

Finally, I would like to express my profound love and gratitude to my mother and grandmother, who have nurtured and supported me every step of the way. None of my achievements would have been possible without their sacrifices and devotion. They gave up precious time that we could have spent together so that I could dedicate myself to my scientific studies in a foreign country for many years. I hope that, throughout these years of intense work, I have lived up to the high standards of humanity and ethical conduct that they instilled in me and have always expected of me. This work is dedicated to them.

Ankara, December 2024

Contents

CHAPTER 1

Introduction

Scientific progress in physics requires a conjunction of advancement in theoretical and experimental fronts. Historically, the distinction between theory and experiment, and theoretician and experimentalist was not as pronounced as it is today. Although there had been examples such as Ernest Rutherford and Enrico Fermi, the rapid expansion of scientific knowledge and the creation of numerous specialized sub-fields in the course of the 20^{th} century has made it gradually more and more difficult for scientists to develop sufficient expertise in both areas within a reasonable amount of time.

Whether one or the other guides the progress in physics more is an often debated topic. Independent from a purely philosophical argumentation, this question might also be answered differently at different periods of time for more practical reasons. For example, lack of funding for one type of research due to political or economical reasons, or limitations in technology could prevent an otherwise scientifically correct strategy from being pursued. Ultimately, both a theoretical explanation and an experimental observation are required for establishing scientific consensus about a natural phenomenon which is being studied. It is the topic of research when either of the two miss, or disagree with each other. When a research enterprise succeeds in filling the gap in understanding, new scientific knowledge is produced.

A famous example from the history of physics is the *ultraviolet catastrophe*, which refers to the early 20^{th} century problem regarding blackbody radiation. The accepted theory of classical statistical mechanics at the time predicted a divergent energy density spectrum in the ultraviolet regime for a blackbody in thermal equilibrium. In contrast, the corresponding observed spectrum was measured to be finite. Max Planck, guided and compelled by the observation, had been able to resolve this disagreement between the two spectra by modifying the theoretical prediction through making the revolutionary assumption that energy varies in discrete amounts and is not a continuous parameter [1]. This paradigm shift has led to the birth of Quantum Mechanics.

In a similar logic, this thesis presents an experimental particle physics measurement which tests the corresponding prediction of the established theoretical model of particle physics, the Standard Model (SM). With its accurate predictions over an energy scale range of many orders of magnitude, the SM is one of the strongest theoretical frameworks ever conceived. Yet, it does not address all the observed phenomena known to us. In order to introduce the theory being tested in this work, the conceptual and mathematical properties as well as limitations of the SM are summarised in Chapter 2.

Directly observing a process forbidden in the SM, or finding a disagreement between the measurement of a process and its corresponding SM prediction, are the two ways new physics might be discovered. Considering the latter, measurements of top-quark production made by the ATLAS collaboration at CERN are shown in Figure 1.1, together with their theoretical SM predictions.

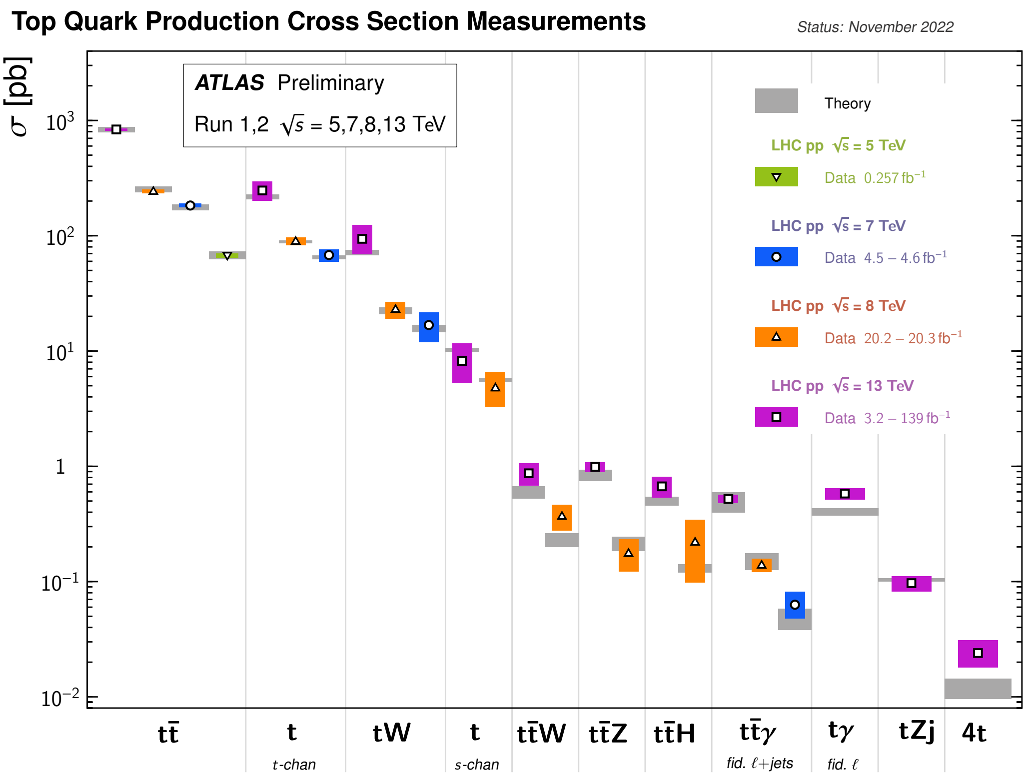


Figure 1.1: Top-quark production measurements made by the ATLAS collaboration, compared to SM predictions. The grey (coloured) rectangles are the predicted (observed) results. The height of each box represents the uncertainty associated with the result. The plot is taken from Reference [2].

The result of the analysis described in this thesis can readily be seen on the right-most column titled "4t". This process is the simultaneous production of four top-quark particles and is referred to as "$t\bar{t}t\bar{t}$" in this thesis. As it is the subject of the measurement, properties of the $t\bar{t}t\bar{t}$ process as predicted by the SM and probed by the previous experimental efforts are detailed in Chapter 3.

Due to their subatomic nature, producing, accessing and measuring the properties of elementary particles and their interactions requires advanced instruments and computational methods. Observational data used in this work are recorded by the ATLAS detector, from the proton–proton collisions provided to it by the Large Hadron Collider (LHC). Large numbers of collision events are stored and processed for use in experimental analyses. Understanding how and under what conditions data is collected and processed is crucial in any measurement. Therefore, the working principles of the accelerator and the detector as well as the accompanying hardware and software components enabling their functions are explained in Chapter 4.

As discussed above, deviations in measurements from the prediction would be an indication of new physics, similar to how the ultraviolet catastrophe allowed Planck and his contemporaries to uncover the quantum nature of energy. However, for example when looking at Figure 1.1, deciding on what constitutes a ‘deviation’, or more generally a significant observation, requires a metric for decision. How much of a disagreement would imply that the sought-after phenomenon exists or not? Analysis and inference of data thus necessitates the use of statistical methods. Taking into consideration the mathematical challenges and instrumental complexities, such a measurement involves various types and amounts of uncertainties. This challenge is represented quantitatively in Figure 1.1 by the error rectangles. A statistical optimisation is often pursued in order to reduce the uncertainties and increase the performance of the statistical analysis being conducted. Such a statistical optimisation has been made in the $t\bar{t}t\bar{t}$ measurement using machine-learning (ML) methods and is described in this thesis. In Chapter 5 the statistical methods and ML tools employed in this work are explained.

The analysis of data is performed in several steps in order to probe the $t\bar{t}t\bar{t}$ process as precisely and accurately as possible. All steps are elaborated upon in Chapter 6 which concludes with the results of the measurement. It must be made clear that this analysis work has been conducted by a team of circa 30 members of the ATLAS collaboration. Computational resources and tools of the ATLAS collaboration in large and of CERN were actively used by the team at every stage and for each scientific result. The author was one of the members of the analysis team and has made personal contributions to the definition and improvement of the signal region by developing, optimising and validating a ML algorithm. He also participated in the fit and post-unblinding studies that have been done to probe the behaviour of the fit setup under certain variations. The results presented in this thesis are published by the ATLAS collaboration in the form of Reference [3]. The author has also presented the work in an international conference, whose proceedings are published in the form of Reference [4]. The top-quark Yukawa coupling interpretation studies presented in Section 6.7 are not part of the publications and is the author’s own work, conducted using resources and tools of the ATLAS collaboration, but not approved by or representing it.

In Chapter 7 the research and its findings are summarised, and limitations of the current results as well as the outlook for future directions are discussed, which concludes the document.

Note on the observation of the $t\bar{t}t\bar{t}$ process by the ATLAS and CMS collaborations

In May 2023, the ATLAS and CMS collaborations both reported the observation of the $t\bar{t}t\bar{t}$ production process [5, 6]. As this document was almost finalised at the time of these announcements, the discussions in this thesis does not include these new results and their consequences. The reader is referred to the respective publications for the details of these analyses.

CHAPTER 2

Overview of theoretical foundations

This Chapter gives a brief summary of the theoretical foundations of modern particle physics with a focus on parts related to top-quark physics and LHC phenomenology as they are essential to the work presented in this dissertation. The theoretical framework considered is the SM. The SM classifies the fundamental constituents of matter and describes how the forces of nature interact with them. These categories, particles and forces are introduced in Section 2.1. The mathematical foundation of the SM is grounded in the Quantum Field Theory (QFT). Basic notions of QFT and how they are used to build the structure of the SM are discussed in Section 2.2.

Although the SM is a successful theory in terms of its predictive power, it does not encompass all natural phenomena known to humankind. A prominent example is the gravitational force, which does not exist in the SM. Few examples of the shortcomings and possible ways to extend the SM are discussed in Section 2.3.

Section 2.4 describes physics and properties of the top-quark, the particle of interest for this dissertation. A more detailed description of one of the SM parameters that could point at new physics, namely the top-quark Yukawa coupling, is discussed at the end of this Section.

As the data used to test the theoretical models is collected from proton–proton collisions at the LHC, all subsequent outputs originate from the QCD-dominated interactions between them. Starting from that collision moment, each step for each event; up to the final, human-readable data formats used in the analyses are simulated using computers. These simulations are essential to the understanding and quantification of the physics processes as well as the detectors' response to them. In Section 2.5 the phenomenological basis and the simulation steps of the proton–proton collisions are summarised.

2.1 Matter and forces

In the SM, matter and forces are classified into two categories called fermions and bosons according to the spin-statistics they follow.

Bosons are integer-spin particles subject to Bose-Einstein statistics. In the SM there are five types of bosons. Four of them are spin-1 particles and are associated with the three fundamental forces of nature: The mediator of the electromagnetic force is the massless spin-1 boson *photon* (γ). All particles possessing an electric charge are subject to the electromagnetic force. The photon is neutral, therefore it cannot interact with itself, i.e. self-couple. Another neutral spin-1 particle is the massive Z-boson with a mass of ≈91 GeV. The Z-boson, together with the lighter $W^{\pm}$ bosons that have a mass of ≈80 GeV, constitute the mediators of the *weak force*. In the SM, electromagnetic and weak forces are interpreted as different low-energy manifestations of a single, *electroweak force*. This is called *electroweak unification* and discussed in Section 2.2.3. The *strong force* is mediated by *gluon*s, which are electrically neutral massless spin-1 bosons, and is only felt by particles that have *colour charge* which takes three values: red, blue and green.

Fermions are half-integer-spin particles and follow the Fermi-Dirac statistics. They describe matter, and further split into two categories according to their interaction with the strong force: The group of particles that interact with the strong force is known as *quarks*. There exist six types of quarks, which are grouped into three generations of doublets. In order of increasing generation, the quarks are: up (u) and down (d), charm (c) and strange (s), top (t) and bottom (b). The d, s and b (u, c and t) quarks are together referred to as down (up)-type quarks. Each quark has one of three colour charges, and a fractional electric charge that can be 2/3 or -1/3. The second category is *leptons*. They lack colour charge and thus do not interact with the strong force. Leptons are also grouped into three generations of doublets. In each generation there exists a charged particle and its corresponding uncharged neutrino. In order of increasing generation, leptons are: electron (e) and electron neutrino (ν_e), muon (μ) and muon neutrino (ν_μ), and tau (τ) and tau neutrino (ν_τ). Each particle also has a corresponding *anti-particle*. A particle and its anti-particle have the same mass but differ by having opposite quantum numbers. In Table 2.1 the particles in the SM and their basic properties are listed.

Properties of forces and particles, as well as the relationships between them, conceptually described above, appears arbitrary at first. Using mathematics, this scaffolding of SM can be explained, revealing a logical structure, patterns of which guide physicists in the search for new theoretical extensions and enable them to make predictions.

Category	Type	Symbol	Name	Electric Charge [e]	≈Mass [GeV]
Fermion	Lepton	e^-	Electron	-1	511 x 10^{-6}
		ν_e	Electron Neutrino	0	< 10^{-10}
		μ^-	Muon	-1	105 x 10^{-3}
		ν_μ	Muon Neutrino	0	< 10^{-10}
		τ^-	Tau	-1	1.78
		ν_τ	Tau Neutrino	0	< 10^{-10}
	Quark	u	Up	2/3	2.16 x 10^{-3}
		d	Down	-1/3	4.67 x 10^{-3}
		c	Charm	2/3	1.27
		s	Strange	-1/3	93 x 10^{-3}
		t	Top	2/3	172.5
		b	Bottom	-1/3	4.2
Boson	Vector	γ	Photon	0	0
		$W^\pm$	W-boson	±1	80.4
		Z	Z-boson	0	91.2
		g	Gluon	0	0
	Scalar	H	Higgs	0	125

Table 2.1: List and selected properties of the particles in the Standard Model. In the SM neutrinos are considered as massless particles, however, experimental observations established that they are also massive particles. Some mass values are rounded for ease of comparison and taken from Reference [7].

2.2 General formalism of quantum field theory

The mathematical foundation of modern particle physics is Quantum Field Theory (QFT). In QFT, fields are the most fundamental elements and particles are interpreted as their local excitations or *quanta*. This is valid for both matter and mediator particles. In classical mechanics, the idea of using fields was motivated by solving the problem of spontaneous action-at-a-distance. A prominent classical example of spontaneous action-at-a-distance is the theory of gravitation developed by Isaac Newton.

Newtonian mechanics lacks locality; for instance, the effect of the change in the motion of a far-away galaxy is instantly felt on Earth, without any delay due to distance. Having fields as mediators of any type of influence allows for a description in which the interactions propagate gradually and locally. This idea is confirmed by the modern understanding of the speed of light—which is finite—as the ultimate transmission speed in the universe.

As locality remains a fundamental requirement, the motivation for the development of QFT came from the necessity and desire to combine Quantum Mechanics (QM) and Special Relativity (SR) into one relativistic theory of quantum mechanics. With this achievement, unlike QM, QFT is able to describe and explain quantum mechanical systems at relativistic regime, where the number of particles are not conserved. Excitation of fields can increase the number of particles and their de-excitation can reduce them. The former is known as particle creation whereas the latter is known as particle decay. This also holds true for vacuum. Since the Heisenberg uncertainty principle prevents the ground state of a system to be absolutely at zero, vacuum is also made of fluctuating quantum fields which can create particle–antiparticle pairs.

The mathematical basis of QFT follows from classical mechanics, most commonly expressed using the Lagrangian formalism. A Lagrangian (L) of a system is defined as the difference between the kinetic (T) and potential energy (V) it has:

$$L = T - V. \tag{2.1}$$

In the case of a field $\phi(x)$ which is a function of a continuous variable x, the Lagrangian is

$$L(x) = \int \mathscr{L}(\phi, \partial_\mu \phi) d^3x, \tag{2.2}$$

with $\mathscr{L}$ being the Lagrangian density and $\partial_\mu \phi$ the four-gradient of the field $\phi(x)$. In order to arrive at the equations of motion from a Lagrangian, the *principle of stationary action* is invoked. According to this principle, the full information of a physical system can be described by its action (S)

$$S(x) = \int \mathscr{L}(\phi, \partial_\mu \phi) d^4x, \tag{2.3}$$

and the system moves along the path on which S remains stationary, i.e.

$$\delta S(x) = 0. \tag{2.4}$$

Imposition of this condition results in what is known as the Euler–Lagrange equations of motion, which for a given field $\phi(x)$ are

$$\partial_\mu \left[\frac{\partial \mathscr{L}}{\partial(\partial_\mu \phi)} \right] - \frac{\partial \mathscr{L}}{\partial \phi} = 0. \tag{2.5}$$

Different field functions can lead to the same Euler–Lagrange equations of motion as long as the $\delta S(x) = 0$ condition is satisfied. If such different fields are related via a continuous transformation, they are said to possess a continuous symmetry and belong to the same symmetry group. If the transformation is independent of coordinates, and operates exactly the same at all points in space-time, it is a global symmetry. On the other hand, if the transformation operator is dependent on the coordinates, it is a local symmetry, also known as a gauge symmetry. For a field $\phi(x)$, both type of transformations can be exemplified by a simple complex phase transformation:

$$\begin{aligned} \text{Global:} \phi(x) &\rightarrow \phi(x) e^{i\theta}, \\ \text{Gauge:} \phi(x) &\rightarrow \phi(x) e^{i\theta(x)}. \end{aligned} \tag{2.6}$$

According to Noether's first theorem [8], each global symmetry is associated with a conservation law. For example rotational symmetry is associated with the angular momentum conservation.

Gauge symmetries are used to work with redundancy in the description of physical systems. When a Lagrangian has redundant degrees of freedom, it can be expressed in multiple but equivalent ways. That is, systems with different gauges describe the same physical situation with different formulations. However, in order to solve the system, one of those needs to be chosen (known as *gauge fixing*). All formulations in the SM are gauge invariant.

Once the Lagrangian of an elementary particle process is found, it can be used to calculate the production cross-section of this process, which is the rate of occurrence of that process, a value that can be measured experimentally. The theoretical calculation of cross-sections can be made by applying the *Feynman rules* to the Lagrangian of the process. Feynman rules are set of associations between the terms of a Lagrangian and mathematical expressions that could be used to compute the *matrix-element* (ME). The absolute square of the matrix-element gives the transition probability between initial and final states of the process.

Feynman rules distinguish between particles and interactions, and can be represented visually through Feynman diagrams. In a Feynman diagram, particles and interactions are represented with lines, form of which can change depending on the type of the particle. Solid lines are used for fermions, wavy lines for vector-bosons and photons, spring-like lines for gluons and dashed lines for the Higgs-boson. The lines that represent the particles mediating the forces are called *propagators*. The vertices are defined as the points where particle lines interact with each other. In this thesis, Feynman diagrams are drawn using the convention that the time flows from left to right, or in other words, the initial state is on the left and the final state is on the right. Three example Feynman diagrams are shown in Figure 2.2. Through Feynman rules, each vertex and propagator in a diagram is assigned a factor that contributes to the matrix-element calculation, with the vertex (propagator) factor accounting for the interaction (free) term of the Lagrangian. The strength of an interaction at a vertex is proportional to the *coupling constant* of that interaction. In this visual representation, summing over the transition probabilities calculated for each possible diagram leads to the differential cross-section. Integrating this over the total available phase space results in the total cross-section of the process. It is possible to draw infinitely many Feynman diagrams for any given process by adding internal closed loops. The type of diagram with the lowest number of possible vertices to depict a given process is called its *tree-level* or *leading-order* (LO) diagram. The depiction with the second least possible vertices is then called next-to-leading (NLO) order. This logic continues for higher order terms, leading to a perturbation series expansion that needs to be integrated over. However, it gets computationally very complex to calculate higher order contributions and therefore in most cases only the first few orders are computed.

The phenomena happening in the closed loops are not directly observable experimentally, and particles appearing in them are called *virtual particles*. A virtual particle can have mass values other than its rest mass, to the extent allowed by the Heisenberg uncertainty principle, called *off-shell* mass. The contributions from internal loops could lead to divergent terms in the integration when the momenta of the virtual particles are high, implying infinite cross-section values.

This problem of divergences had been a major obstacle in the historical development and acceptance of QFT as a valid mathematical formalism. The divergences can be remedied by the mathematical tools of *regularisation* and *renormalisation*. Regularisation is the name of the method given to splitting the divergent part of an integral into two parts: one being finite and the other being divergent. The choice of the point where this split happens is the *cutoff* value, and it defines (in terms of momenta or energy) the regime of validity of the calculation, which is no longer a general calculation valid at all energies. This cutoff is known as the *renormalisation scale*, and represented by the symbol μ_R.

If it is possible to remove the divergent part from the integration, by absorbing it into the coupling constant of the theory under consideration, that theory is said to be renormalisable and its coupling constant becomes an effective coupling parameter which varies with the renormalisation scale. However, since the strength of interaction i.e. the coupling is an observable quantity, it should not be depending on an arbitrary choice of a cutoff value. As such, the change in the coupling with the choice of a different scale can be computed by using the renormalisation group equations (RGE). The evolution of the coupling constant (which is now only constant in name) with the energy scale is known as the *running* of the coupling constant, and described by the beta functions (β)

$$\beta(g) = \frac{\partial g}{\partial \ln \mu_R}, \tag{2.7}$$

where g is the coupling constant.

Through their ability to extrapolate the calculations to any energy scale, the running of coupling constants provide a way to theoretically probe the behaviour of interactions at the high energy regimes that are not yet experimentally accessible. This is particularly important because of the prospect of a Grand Unified Theory (GUT), in which forces that look distinct to us at the lower energy scales we can perceive or study, become one and unified at a certain higher energy scale. When the renormalisation calculations are done for a fixed order (such as at NLO), the scale dependence cannot be removed. In this case the scale used can be varied in different ways to estimate an uncertainty that accounts for the incompleteness of the theoretical calculation.

These fundamental properties leads to the definition of the SM as a renormalisable, gauge-invariant quantum field theory that is compatible with the special theory of relativity. The various sectors of the SM are shortly discussed in the following. For a detailed account the reader is referred to the vast literature of QFT, e.g. References [8, 9, 10].

2.2.1 Quantum Electrodynamics

Quantum Electrodynamics (QED) is the quantum mechanical extension of classical electrodynamics described by Maxwell's Equations. QED explains the behaviour and interaction of particles that possess electric charge. QED has the gauge symmetry of the Unitary group U(1) which is associated to its gauge boson, the photon, and to the conservation of electric charge. The Lagrangian for a free spinor field (ψ) that is not subject to any external influence is

$$\mathscr{L}_{\text{Free}} = \bar{\psi}(i\gamma^{\mu}\partial_{\mu} - m)\psi. \tag{2.8}$$

In Equation 2.8, m is the mass of the electron, $\bar{\psi}$ is the adjoint spinor[1] of ψ and γ^μ are the Dirac matrices.

In order to apply the U(1) gauge symmetry, the field is subjected to a transformation by a 1×1 unitary matrix which can be equivalently represented by a complex phase

$$\psi(x) \to \psi(x)e^{iq\theta(x)}, \tag{2.9}$$

for which the $\mathscr{L}_{\text{Free}}$ is not invariant due to an additional term from the derivative, leading to the modified Lagrangian density

$$\mathscr{L} \to \mathscr{L} - q\partial_\mu\theta(x)\bar{\psi}\gamma^\mu\psi. \tag{2.10}$$

It is a general property that free field Lagrangians do not satisfy gauge invariance conditions. In order to render the Lagrangian invariant, a modified, covariant derivative D_μ, defined as

$$D_\mu = \partial_\mu - iqA_\mu, \tag{2.11}$$

is introduced, which replaces the partial derivative ∂_μ. This modification is known as the *minimal coupling rule* [11]. A_μ is a new vector field that transform as:

$$A_\mu \to A_\mu + \frac{1}{q}\partial_\mu\theta(x). \tag{2.12}$$

After these substitutions, $\mathscr{L}$ is now invariant under the local gauge transformation. The new field A_μ has to have a corresponding free Lagrangian as well. Since A_μ is a vector field, the Proca Lagrangian describing a particle with a mass of m and a spin of 1 can be used [11]:

$$\mathscr{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + m^2A_\mu A^\nu. \tag{2.13}$$

Here, $F^{\mu\nu}$ is the *field strength tensor* with the definition $F^{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. While $F^{\mu\nu}F_{\mu\nu}$ is gauge invariant, the term $A_\mu A^\nu$ isn't. However, the gauge invariance can be kept if we require the particles to be massless, i.e. $m = 0$. This new gauge field describes the gauge boson photon, which is a massless, spin-1 particle. The resulting Lagrangian is

$$\mathscr{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \bar{\psi}(i\gamma^\mu D_\mu)\psi = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \bar{\psi}(i\gamma^\mu\partial_\mu)\psi + q\bar{\psi}\gamma^\mu A_\mu\psi. \tag{2.14}$$

The introduction of the new vector field creates an interaction term $q\bar{\psi}\gamma^\mu A_\mu\psi$, whose strength is governed by the coupling constant q, which, in this case is interpreted as the charge of the electron: $q = e$. In the case of QED, running of this coupling constant, α_e, is understood with the concept of charge screening due to *vacuum polarisation*. In the vicinity of an electron, virtual electron-positron pairs appear due to the excitations of the vacuum. Since opposite charges attract each other, virtual positrons will cluster around the electron and thus lead to a screened, reduced effective charge $e(r)$ when probed from a distance larger than the radius (r) of this cluster. The coupling constant decreases with increasing distance and converges to the well-known asymptotic value of *fine structure constant*, $\alpha_e(r \to \infty) \approx 1/137$.

[1] $\bar{\psi} = \psi^\dagger\gamma^0$

2.2.2 Quantum Chromodynamics

The strong force is described by Quantum Chromodynamics[2] (QCD). The defining symmetry group in this theory is the Special Unitary Group SU(3). The corresponding gauge bosons are named gluons which are particles without mass or electric charge. Instead, they carry colour charge which can have three values: red (r), green (g) and blue (b). There exist eight gluons associated with the group generators. Similar to QED, one can start from the free Lagrangian density for a given quark flavour q:

$$\mathscr{L}_{\text{Free,q}} = \bar{\psi}_q (i\gamma^\mu \partial_\mu - m_q)\psi_q. \tag{2.15}$$

However, since QCD has three colour charges, the field ψ_q now is a vector of Dirac spinors, with one spinor for each colour:

$$\psi_q = \begin{pmatrix} \psi_r \\ \psi_b \\ \psi_g \end{pmatrix}, \ \bar{\psi}_q = (\bar{\psi}_r \ \bar{\psi}_b \ \bar{\psi}_g). \tag{2.16}$$

The SU(3) symmetry group can be represented by a set of eight Gell-Mann matrices ($\lambda_{1,2...8}$), each being a 3×3 matrix with a determinant of one. In order to impose SU(3) gauge symmetry, one then requires that fields transform under

$$\psi_q(x) \to \psi_q(x) e^{i g_s \theta_a(x) \frac{\lambda_a}{2}}, \tag{2.17}$$

where g_s is the coupling constant of the strong interaction and $a = 1, 2...8$. The covariant derivative to compensate for new terms is $D_\mu = \partial_\mu - i g_s A_\mu(x)$. There are eight new gauge fields A_μ^a introduced:

$$A_\mu = \sum_{a=1}^{8} A_\mu^a(x) \frac{\lambda_a}{2}. \tag{2.18}$$

Accounting for the terms linear in $\theta_a(x)$, A_μ^a transform as

$$A_\mu^a \to A_\mu^a - \frac{1}{g_s}\partial_\mu \theta_a(x) + f_{abc}\theta_b(x) A_\mu^c, \tag{2.19}$$

where f_{abc} give the structure constants of the SU(3) group. Adding the free Lagrangian terms for the new gluon fields, the gauge invariant QCD Lagrangian density is finally given by a sum over six quark flavours q

$$\mathscr{L}_{\text{QCD}} = \sum_{q=1}^{6} -\frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu} + \bar{\psi}_q (i\gamma^\mu D_\mu - m_q)\psi_q, \tag{2.20}$$

where $G_{\mu\nu}^a$ is the field strength tensor defined as $G_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + f_{abc} A_\mu^b A_\nu^c$. Unlike the QED field tensor in Equation 2.13, self-coupling of field mediators are not proportional to the mass, but to the structure constants f_{abc}. As such, gluons are massless, but they can self-interact through triple and quadruple vertices.

[2] From *khrôma*, ancient greek word for “colour”.

Quarks have a similar screening effect through vacuum polarisation, as explained previously with electrons for the QED case. However, unlike QED, the mediator bosons in QCD also contributes to the vacuum polarisation since they can self-couple. Because gluons enhance the field strength they lead to anti-screening effects at distance.

The interplay between QCD screening and anti-screening can be quantified by the beta-function at lowest order [11]:

$$\beta(\alpha_s) \propto -\left(11 - \frac{n_s}{6} - \frac{2n_f}{3}\right)\alpha_s. \tag{2.21}$$

In the SM, the number of flavours (n_f) is 6, and the number of scalar coloured bosons (n_s) is 8. Therefore, in this case the interplay is dominated by the anti-screening effect due to gluon self-couplings, leading to the uniquely-QCD phenomenon of interaction strength of the strong force being lower at smaller distances, and increasing at higher energies. This behaviour is called *asymptotic freedom*, as at the smallest distance scales the strong force asymptotically approaches zero, implying quarks behaving as free particles in those regimes. On the other hand, in the case of larger distances and correspondingly lower energies, the increase in the strong force leads to accumulation of energy between two quarks, eventually leading to the creation of new quark-antiquark pairs when the threshold energy for particle creation is reached. When the system falls below an energy scale of $\Lambda_{\text{QCD}} \simeq 300$ MeV [11] the perturbative expansion regime breaks down. These newly formed quarks combine and form bound states, in a process called *hadronisation*. Quarks cannot be observed as single entities, and their combinations always lead to colour-neutral bound states. This is called *colour confinement*. These bound states are called hadrons and they come in two types depending on the number of quarks in the bound state. Quark-antiquark particles are called *meson*s, and three-quark particles are called *baryon*s. Due to difficulties in studying the non-perturbative regime, the exact mechanism of how colour-confinement is realised in nature is still a topic of research. Two commonly used models with predictive power and an ability to describe data well are the Lund string [12] and the Cluster model [13].

2.2.3 Electroweak theory

Weak Force

In the SM, the weak force interacts with all fermions and the Higgs-boson. It has two charged (W^+ and W^-) and one neutral (Z) mediators. Crucially, in contrast to QED and QCD, all three bosons are massive, leading to a finite range relative to other forces. Massive force carriers are one of the many aspects where the weak force differs from other forces: it is also the only force that can change the flavour of fermions. This happens when $W^\pm$ bosons interact with quarks. In the case of interactions with leptons, $W^\pm$ bosons affect the charged lepton and its corresponding anti-neutrino of the same generation only. Neutral Z-boson can couple to any fermion pair but does not change their types.

Weak interaction is described by a chiral[3] theory, that is, it does not conserve the *parity* (defined at the end of this section). For a fermion field ψ, chirality is defined using the γ^5 operator, which is related to other gamma matrices through the relation $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$. With the corresponding projection

[3] From *kheir*, ancient greek word for "hand".

operators $\frac{1}{2}(1 \pm \gamma^5)$ one distinguishes between the left-handed (ψ_L) and right-handed (ψ_R) chiral components of fermionic fields:

$$\psi_L = \frac{1}{2}(1-\gamma^5)\psi,\ \psi_R = \frac{1}{2}(1+\gamma^5)\psi. \tag{2.22}$$

The chirality of fermions is quantified using the weak isospin (I) number and its third component (I_3), related to 2×2 Pauli matrices of the SU(2) symmetry group. Left-handed fermions have $I = 1/2$ and they are grouped into doublets having $I_3 = \pm 1/2$:

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L, \begin{pmatrix} u \\ d \end{pmatrix}_L, \begin{pmatrix} c \\ s \end{pmatrix}_L, \begin{pmatrix} t \\ b \end{pmatrix}_L. \tag{2.23}$$

On the other hand, right-handed fermions have $I = 0$ and form singlets having $I_3 = 0$:

$$e_R,\ \mu_R,\ \tau_R,\ u_R,\ d_R,\ c_R,\ s_R,\ t_R,\ b_R. \tag{2.24}$$

Left-handed leptons are grouped into doublets made up of a charged lepton and its corresponding neutrino. In the SM, neutrinos are massless, therefore they do not have right-handed singlet states. In the case of quarks, left-handed doublets are formed by up- and down-type quarks belonging to the same generation.

Electroweak interaction

Through work of Glashow, Weinberg and Salam, weak and electromagnetic interactions are combined under a common symmetry group $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$, and known as the electroweak (EW) interaction. Here, Y is the weak hypercharge and it is related to I_3 and electric charge Q via the Gell-Mann–Nishijima relation [11]:

$$Q = I_3 + \frac{Y}{2}. \tag{2.25}$$

The local gauge transformations of the respective symmetry groups lead to the transformation of the fermion fields as,

$$\begin{aligned} SU(2)_L : \phi_L &\rightarrow e^{i\frac{g_w}{2}\sum_{a=1}^{3}\tau_a\alpha_a(x)}\phi_L, \\ U(1)_Y : \phi_L &\rightarrow e^{i\frac{g_w'}{2}Y\beta(x)}\phi_L,\ \phi_R \rightarrow e^{i\frac{g_w'}{2}Y\beta(x)}\phi_R. \end{aligned} \tag{2.26}$$

Here $\alpha_a(x)$ and $\beta(x)$ are local phases, τ_a Pauli matrices, g_w and g_w' are two coupling constants. In order to account for extra terms brought in by the gauge transformation, two covariant derivatives are introduced:

$$\begin{aligned} D_L^\mu &= \partial^\mu + i\frac{g_w}{2}\sum_{a=1}^{3}\tau_a W_a^\mu + i\frac{g_w'}{2}YB^\mu, \\ D_R^\mu &= \partial^\mu + i\frac{g_w'}{2}YB^\mu. \end{aligned} \tag{2.27}$$

Through these definitions, four new gauge fields appear: W_1^μ, W_2^μ, W_3^μ related to the $SU(2)_L$, and B^μ related to the $U(1)_Y$ symmetry group. These fields have the transformation properties of

$$\begin{aligned} W_a^\mu &\rightarrow W_a^\mu - \frac{1}{g_w}\partial_\mu \alpha_a(x) - \alpha_a(x) W_a^\mu, \\ B^\mu &\rightarrow B^\mu - \frac{1}{g_w'}\partial_\mu \beta(x). \end{aligned} \tag{2.28}$$

Adding the free field Lagrangian densities, the final EW Lagrangian density can be written as:

$$\mathscr{L}_{\mathrm{EW}} = \bar{\psi}_L(i\gamma_\mu D_L^\mu)\psi_L + \bar{\psi}_R(i\gamma_\mu D_R^\mu)\psi_R - \frac{1}{4}\sum_{a=1}^{3} W_{\mu\nu}^a W_a^{\mu\nu} - \frac{1}{4} B_{\mu\nu}B^{\mu\nu}, \tag{2.29}$$

where the two field strength tensors $W_{\mu\nu}^a$ and $B^{\mu\nu}$ are defined as

$$\begin{aligned} W_a^{\mu\nu} &= \partial_\mu W_a^\nu - \partial_\nu W_a^\mu - g_w \epsilon_{abc} W_b^\mu W_c^\nu, \\ B^{\mu\nu} &= \partial_\mu B^\nu - \partial_\nu B^\mu, \end{aligned} \tag{2.30}$$

with ϵ_{abc} providing the structure constants for the $SU(2)_L$ symmetry group. Linear combinations of these four fields lead to four physical fields present in the QED and weak sectors. The charged gauge bosons $W^\pm$ are expressed by a combination of W^1 and W^2:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp W_\mu^2). \tag{2.31}$$

Due to their charge neutrality, a relationship between the Z-boson (Z) and the photon (A) fields is anticipated and expressed through a two-dimensional rotation of the W^3 and B^μ fields

$$A_a^\mu = W_\mu^3 \sin(\theta_w) + B^\mu \cos(\theta_w), Z_a^\mu = W_\mu^3 \cos(\theta_w) - B^\mu \sin(\theta_w). \tag{2.32}$$

by an angle θ_w known as the *weak mixing angle* or Weinberg angle [11]

$$\tan(\theta_w) = \frac{g_w'}{g_w}. \tag{2.33}$$

As in the case of QCD, there are self-interaction terms proportional to the field strength tensor ϵ_{abc} for the W fields. The EW theory at this point still does not provide a way for the three massive bosons to acquire mass. This is achieved by breaking the symmetry of the EW theory by introducing a new scalar field.

Spontaneous symmetry breaking and the Higgs mechanism

In order to resolve the conflict between experimental observation of massive electroweak gauge bosons and EW theory's prediction of massless particles due to gauge-invariance symmetry, the concept of spontaneous symmetry breaking (SSB) was introduced. Here, "spontaneous" implies that symmetry is broken by the system's own internal dynamics, rather than being the result of an external effect (such as breaking of the up/down symmetry of a system that is under the influence of gravitational force).

Since the electroweak Lagrangian discussed in the previous Section cannot possess mass terms, a new, complex scalar field (ϕ) is introduced in the form of an SU(2) isospin doublet:

$$\begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \tag{2.34}$$

After the realisation of SSB by imposing the gauge-invariance condition on this new scalar field, it results in a new particle, identified as the Higgs-boson. This process is known as the Higgs Mechanism.

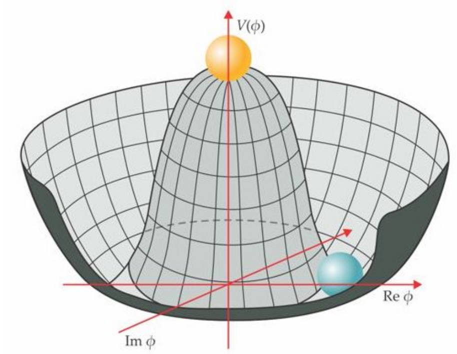
The Lagrangian density associated with this field is given by

$$\mathscr{L}_{\mathrm{H}} = (D_\mu \phi^\dagger)(D^\mu \phi^\dagger) - V(\phi), \tag{2.35}$$

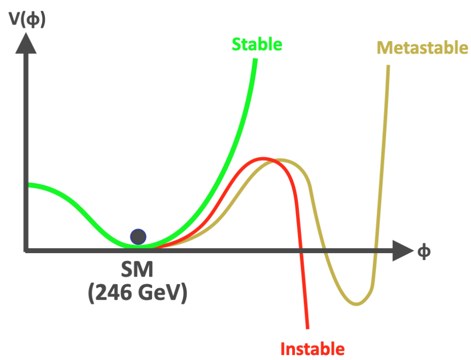
where covariant derivatives are those defined in Equation 2.27. A general potential function fulfilling the conditions of SU(2)-invariance and renormalisability needs to be a quadratic function, and can be expressed as

$$V(\phi) = \mu^2(\phi^\dagger\phi) + \lambda(\phi^\dagger\phi)^2 = \mu^2|\phi|^2 + \lambda|\phi|^4. \tag{2.36}$$

Here, μ and λ are parameters of the potential and $\lambda > 0$ is assumed. The sign of the parameter μ^2 then defines where the ground state (ϕ_0) of the system is located. For positive values of μ^2 the ground state is unique and occurs at $V = 0$ for $\phi_0 = 0$. This ground state is symmetric with respect to SU(2) transformations. When μ^2 is negative, the ground state is no longer unique but becomes a continuous range of degenerate ground states, located at a radius of $\phi_0^2 = -\mu^2/2\lambda$ from the origin on the complex plane as depicted in Figure 2.1(a). Since a perturbative expansion needs to be done in the vicinity of



(a) Drawing of the potential function given by $V(\phi) = \mu^2|\phi|^2 + \lambda|\phi|^4$ for $\lambda > 0$ and $\mu^2 < 0$. Graphical representation is taken from Reference [14].



(b) Various effective potential shapes representing the stable, metastable and instable SM electroweak vacuum.

Figure 2.1: Two figures illustrating the shape of the Higgs field (ϕ) potential under various conditions.

the ground state, $\phi_0 = 0$ cannot be used. In order to proceed with the calculation, a particular ground state among the infinite possibilities needs to be chosen. This specific choice among degenerate ground states creates an asymmetry and results in the spontaneous breaking of the SU(2) symmetry.

A ground state is conventionally chosen to be

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \sqrt{-\mu^2/2\lambda} \end{pmatrix} = \begin{pmatrix} 0 \\ v \end{pmatrix}, \tag{2.37}$$

where v is the vacuum expectation value (vev) of the Higgs field defined as $v = \sqrt{-\mu^2/\lambda}$. Around the now-chosen ground state, the field ϕ_0 can be expanded and expressed as

$$\phi(x) = \frac{1}{\sqrt{2}} e^{i \sum_{a=1}^{3} \frac{\tau_a \eta_a}{v}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}, \tag{2.38}$$

where four new real fields are introduced. Imposing the local gauge invariance condition

$$\phi_0(x) \to e^{-i \sum_{a=1}^{3} \frac{\tau_a \eta_a}{v}} \phi_0(x), \tag{2.39}$$

on the field $\phi_0(x)$, the three η fields are removed and the remaining field H is associated with the Higgs-boson.

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}. \tag{2.40}$$

The kinetic and potential terms of the free Lagrangian density are separately expressed as:

$$\begin{aligned} (D_\mu \phi^\dagger)(D^\mu \phi^\dagger) &\to \frac{1}{2}\partial^\mu H \partial_\mu H + \frac{1}{8} g_w^2 (v+H)^2 |W_\mu^1 + W_\mu^2|^2 + \frac{1}{8}(v+H)^2 |g'_w W_\mu^3 - B_\mu|^2, \\ V(\phi) &\to -\lambda v^2 H^2 - \lambda v H^3 - \frac{\lambda}{4} H^4. \end{aligned} \tag{2.41}$$

Using Equation 2.31 in the W_μ terms, the W-boson mass can be read from the second term in the kinetic expansion using the relation

$$\frac{1}{2} m^2 \phi^2 \propto \frac{1}{2} \frac{g_w^2 v^2}{4} W_\mu^\pm \to m_W = \frac{g_w v}{2}. \tag{2.42}$$

The Z-boson is in a mixed term, but inserting the relations from Equation 2.32 in the combined term of W_3 and B fields, the Z-boson mass can be read from the last term in the kinetic expansion using the relation

$$\frac{1}{2} m^2 \phi^2 \propto \frac{1}{2} \frac{g_w^2 v^2}{2\cos(\theta_w)} Z^\mu Z_\mu \to m_Z = \frac{m_W}{\cos(\theta_w)}. \tag{2.43}$$

There is no mass term for the resulting photon field, a result that is in agreement with the experimental observation of a massless neutral boson. Looking into the potential expansion, the mass of the Higgs-boson can be found from the relation

$$\frac{1}{2} m^2 \phi^2 \propto \lambda v^2 H^2 \to m_H = \sqrt{2\lambda v^2}. \tag{2.44}$$

At this point, the mass values of all bosons in the SM are defined. Experiments dictate that fermions are also massive. The SM accommodates this fact for fermions other than neutrinos, via introducing a new coupling between the Higgs-boson and fermions. This coupling is called the Yukawa coupling. As the addition of a mass term would violate the gauge invariance, a similar SSB approach presented above is used to incorporate the mass term whilst preserving gauge invariance. For leptons, one can start from the same complex scalar doublet field ϕ described in Equation 2.40. For the quark sector, the charge conjugate of that field is used:

$$\phi_c = i\sigma_2\phi^* = \frac{1}{\sqrt{2}}\begin{pmatrix} v + H(x) \\ 0 \end{pmatrix}. \tag{2.45}$$

The Lagrangian density for the Yukawa couplings is then given by

$$\mathscr{L}_{\mathrm{Y}} = y_f \bar{\psi_L}\phi\psi_R + \bar{\psi_R}\phi_c\psi_L. \tag{2.46}$$

Expanding the field ϕ around the vev leads to the Yukawa Lagrangian density

$$\mathscr{L}_{\mathrm{Y}} = \frac{y_f}{\sqrt{2}}H\bar{\psi}^f\psi^f - \frac{v}{\sqrt{2}}y_f\bar{\psi}^f\psi^f, \tag{2.47}$$

where the first term is the coupling of the Higgs-boson to fermion field and the second term is the corresponding fermion mass given by

$$m_f = \frac{v}{\sqrt{2}}y_f. \tag{2.48}$$

The full Lagrangian density of the SM is then defined as the sum of all these sectors discussed:

$$\mathscr{L}_{\mathrm{SM}} = \mathscr{L}_{\mathrm{EW}} + \mathscr{L}_{\mathrm{QCD}} + \mathscr{L}_{\mathrm{H}} + \mathscr{L}_{\mathrm{Y}}. \tag{2.49}$$

The SM Lagrangian has been developed by exploiting continuous symmetries. There exists also three discrete symmetries that are of central importance to particle physics.

- Charge conjugation (C-Symmetry): Application of charge conjugation to a field changes it to its complex conjugate field and also inverts its quantum numbers. As such, charge conjugation also relates particles to their anti-particles.

- Parity transformation (P-Symmetry): If a system has parity symmetry then it behaves exactly the same way when all spatial coordinates are inverted. This is commonly exemplified by the analogy of a movement and its reflection in a mirror. If one cannot tell whether one is observing the movement itself or its reflection on the mirror, this action is said to have a mirror symmetry or an even parity, conventionally represented by +1. If parity is not conserved, it is assigned a value of -1 and the system is said to have odd parity.

- Time reversal (T-Symmetry): Time reversal amounts to exchange of initial and final states of a process as well as inverting the sign of particle's spin and momentum.

According to current understanding, the combined application of these three transformations to a physical system is a fundamental symmetry of nature, a statement known as the CPT theorem.

It is experimentally confirmed that none of these symmetries alone, or any combination of two of them are absolutely conserved. For example, the combination of charge and parity symmetries (CP) are shown to be not conserved in Kaon [15], B meson [16] and charm hadron [17] decays. This phenomenon, known as CP-violation, is also theoretically predicted by the SM, manifesting itself in the mixing of quark-generations. In the SM, this mixing is quantified by the 3×3 unitary Cabibbo–Kobayashi–Maskawa (CKM) matrix that relates mass eigenstates (d, s, b) of down-type quarks[4] to their quark eigenstates (d', s', b'):

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{2.50}$$

with [7]

$$\begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} = \begin{pmatrix} 0.97435 \pm 0.00016 & 0.22500 \pm 0.00067 & 0.00369 \pm 0.00011 \\ 0.22486 \pm 0.00067 & 0.97349 \pm 0.00016 & 0.04182^{+0.00085}_{-0.00074} \\ 0.00857^{+0.00020}_{-0.00018} & 0.04110^{+0.00083}_{-0.00072} & 0.999118^{+0.000031}_{-0.000036} \end{pmatrix}. \tag{2.51}$$

Each of the nine matrix elements V_{xy} represent the inter-flavour transition strength $x \to y$ that is proportional to $|V_{xy}|^2$. Diagonal elements dominate the CKM matrix, indicating that flavour transitions within the same generation are favoured. A non-unitarity of CKM matrix would indicate the existence of additional quark generations. The study of CP-violation is one of the ways where BSM physics could reveal itself. A brief summary of various shortcomings of the SM are discussed in the next Section.

2.3 The Standard Model as an incomplete description of Nature

Examples of phenomena observed in nature but not addressed by the SM are listed in the following:

- Gravity: Gravity is not part of the SM. A quantum theory of gravitational force consistent with General Relativity is yet to be developed. Up to now, all forces in the SM had associated mediator particles and the hypothetical candidate for the mediator of gravitational interaction is called the *graviton*. The addition of the graviton to the SM is a significant theoretical challenge because its interactions are not renormalisable and therefore lead to divergences. On the experimental side, direct observations of gravitons are out of reach given the current technological level, however, the study of gravitational waves [18] could lead to further understanding of the possible properties of gravitons indirectly.

- Neutrino mass: In the SM, neutrinos are massless particles. However, the discovery of neutrino oscillations have shown that neutrinos can change flavour [19, 20, 21], requiring them to possess

[4] The choice of down-type quarks assumes up-type quarks are same in both bases. This is only a conventional choice, and the usage of up-type quarks would be equally valid.

a non-zero mass. The addition of neutrinos with right-chiral states would result in a SM extension with massive neutrinos. However, using Dirac fermions warrants the question of what the origin of extremely small neutrino masses are in comparison to the rest of the fermions. Thus, further theoretical methods are being explored [7, 22]. One leading candidate for such an extension is the *seesaw mechanism* [22]. Current cosmological limits for the total mass of all three flavours of neutrinos are around 0.12 eV (95% confidence level) [7, 23].

- Matter-antimatter symmetry: Also known as the baryon asymmetry or *baryogenesis*, this topic refers to the problem of the observed abundance of matter over antimatter which is not necessitated by the SM. This asymmetry is commonly quantified using the density parameter η:

$$\eta = \frac{n_B - n_{\bar{B}}}{n_\gamma}$$

where n_B $(n_{\bar{B}})$ is the baryon (anti-baryon) density. This density difference is divided by the photon density parameter, n_γ. Currently, $\eta \approx 6 \times 10^{-10}$ [24]. Any theory explaining the baryogenesis has to satisfy the three conditions known as the Sakharov conditions [25]. These are the baryon number violation, C and CP symmetry violation, and interactions out of thermal equilibrium. In the SM all three could occur: via sphalerons [26], EW interactions and the CKM matrix, and through electroweak symmetry breaking, respectively. However, the amount of baryon asymmetry created by the SM is at the order of $\eta \approx 10^{-19}$ [27] and thus is insufficient to explain the observation.

- Non-SM matter: Matter composed of particles whose nature is explained by the SM and observed by humans makes up less than 5% of the composition of the universe according to cosmological data [28]. The rest are attributed to two sources: Dark Matter and Dark Energy. Leading evidence for dark matter is the observed flatness of the rotation curves in the disc galaxies. If existing, dark matter is estimated to constitute about 27% of the matter-energy content of the universe. The remaining 68% of the composition of the universe is then attributed to hypothetical dark energy which is understood as a property of the vacuum in space. As such, it is thought of as non-local, homogeneous, and omnipresent at the scale of the universe. Evidence for dark energy comes from the crucially important observation that the expansion of the universe is accelerating [29]. There exist alternative hypotheses to both, dark matter and dark energy [30].

Among many avenues and approaches, studying the properties of the top-quark, the particle of interest for the analysis presented in this thesis, can provide insights into how the shortcomings of the SM could be scrutinised. Details on the top-quarks and its possible potential to probe BSM physics via its Yukawa coupling to the Higgs-boson are described in the next Section.

2.4 The Standard Model top-quark

Predicted by Kobayashi and Maskawa in 1973 [31] based on the CKM matrix, and named by Harari in 1975 [32], the possibility of the existence of the top-quark became widely accepted after the discovery of the bottom-quark in 1977 by the E288 experiment at Fermilab [33]. It was discovered 18 years afterwards, in 1995, jointly by the CDF [34] and DØ collaborations [35].

The top-quark has an electric charge of 2/3 and a spin value of 1/2. The most outstanding property of the top-quark is its mass of about 172.5±0.3 GeV [7], which makes it by far the heaviest particle in the SM. The top-quark is an unstable particle with a lifetime of about $\approx 5.0 \times 10^{-25}$ seconds [7] corresponding to a width of $1.42^{+0.19}_{-0.15}$ GeV at NLO accuracy [7]. This lifetime is remarkably and uniquely shorter than the hadronisation scale, making the top-quark the only quark that decays before hadronisation. This implies that decay products of the top-quark carry information (such as spin value) directly descending from the original particle. The anti-particle of top-quark is called anti top-quark.

Production

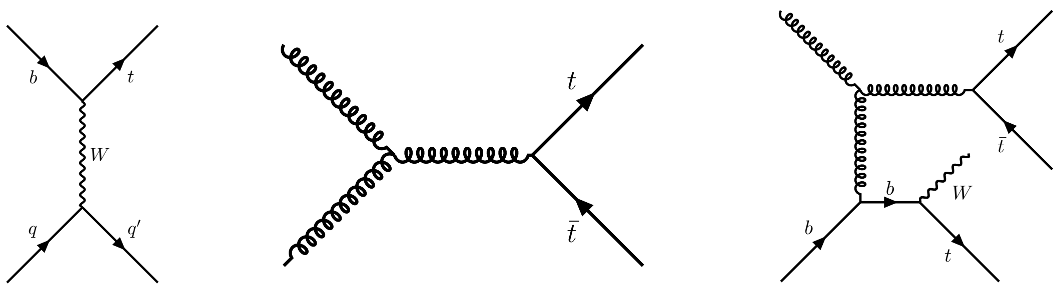
Top-quark production can occur at different multiplicities, with pair-production ($t\bar{t}$) and single-top productions being the leading and sub-leading modes at the LHC, respectively. Single-top production in the SM happens through the electroweak interaction. Production is only possible in association with bosons or quarks and occur via three decay modes:

- t-channel: In this channel a single-top quark is produced together with a quark. It is the single top production channel with the highest cross-section at the LHC.

- tW-channel: The production of a single top-quark happens in association with a W-boson in this channel which contributes to roughly a quarter of the total single-top production at 13 TeV. A known complication in the study of this channel is the quantum interference with the $t\bar{t}$ process when considering higher-order QCD diagrams [36]. The event generators employ methods such as Diagram Removal and Diagram Subtraction to account for this effect [36].

- s-channel: This channel comprises only ≈3% of the single-top production at 13 TeV collision energy. At tree-level, s-channel production is initiated via $q\bar{q}'$ annihilation and as such its cross-section in $p\bar{p}$ collisions are higher than that of pp collisions.

One example Feynman diagram for t-channel single top-quark production process described above is shown in Figure 2.2(a). Studying of single-top production is important among other reasons [37] because it allows for a direct measurement of the CKM matrix element $|V_{tb}|$, a fundamental parameter of the SM.



(a) t-channel single top-quark production

(b) Gluon-gluon fusion initiated $t\bar{t}$ production

(c) Triple top-quark production

Figure 2.2: Example Feynman diagrams for three different top-quark-multiplicity production modes.

$t\bar{t}$ production is the dominant top-quark production mode at the LHC. Close to 90% of $t\bar{t}$ production at the LHC is initiated by gluon-gluon fusion at 13 TeV. An example Feynman diagram for this process is shown in Figure 2.2(b). Other $t\bar{t}$ production processes originate from quark–anti-quark annihilations or quark–gluon interactions. In addition to standalone production, associated production modes of $t\bar{t}$ with bosons are of interest, both as signal processes themselves and as important backgrounds to a large range of analyses. $t\bar{t}Z$, $t\bar{t}W$ and $t\bar{t}H$ are examples of such processes and they are leading backgrounds in the analysis discussed in this dissertation.

Similarly to the single-top quark production, another odd-top-multiplicity process, triple top-quark production in the SM require a Wtb vertex and can only happen in association with other particles. Most commonly, triple top production occurs via a b-quark in the initial state. Combined effects of electroweak production with the requirement of a heavy quark in the initial state leads to $3t$ cross-section at the LHC to be strongly suppressed. At the LO, a production cross-section of 1.9 fb [38] is predicted for the $3t$ process at the LHC with 14 TeV centre-of-mass energy. An example Feynman diagram of this process is shown in Fig. 2.2(c). Triple top-quark production is currently experimentally unexplored.

The simultaneous production of four top-quarks is the subject of the measurement discussed in this dissertation. It is the third most frequent production mode at the LHC at 13 TeV. This process is discussed in detail in the next Chapter. The simultaneous production of more than four top-quarks in the SM is possible, however the predicted cross-sections are extremely small. For example, at 13 TeV, six top-quark production has a cross-section that is five orders of magnitude smaller than that of the four top-quark production [39]. Thus, even at the maximum design energy of the LHC, observing a single six top-quark event requires an integrated luminosity of about 10^4 fb^{-1}, a value larger than what the LHC will collect.

Decay

Top-quarks almost always decays into a b-quark and a W-boson. The identification of b-quarks is therefore central to the experimental investigation of top-quarks, and experimental methods are generically called *b-tagging*. The W-boson can further decay leptonically via $W \to \ell$ or hadronically via $W \to q\bar{q}'$ with fractions of approximately 1/3 and 2/3, respectively. The decay mode of the W-boson thus determines whether the top-quark has decayed semi-leptonically or hadronically. In multi-top-quark processes, various combinations of the decay of W-bosons emerging from each top-quark collectively determines the lepton and jet multiplicities[5] of the final state.

Top-quark Yukawa coupling and its relevance in the search for new physics

Its large mass implies that the top-quark is strongly coupled to the Higgs-boson. The top-quark Yukawa coupling (y_t) is the largest of all Yukawa coupling parameters, with a value of approximately 1. It is a CP-conserving, i.e. CP-even parameter in the SM. As such, a deviation from the predicted magnitude or the CP-characterisation of top-quark Yukawa coupling could indicate new physics.

[5] Here, only jets emerging from particle decays are implied, excluding QCD radiation.

These BSM effects can be described with a modified Yukawa Lagrangian density of the form [40]

$$\mathscr{L}_{\mathrm{Y}}^{\mathrm{BSM}} = -\frac{y_t}{\sqrt{2}} H\bar{t}(a_t + i\gamma_5 b_t)t, \tag{2.52}$$

where a_t (b_t) parametrise the CP-even (CP-odd) couplings. Equation 2.52 reduces to the SM Yukawa Lagrangian 2.47 when the SM values of the parameters are substituted:

$$\begin{aligned} y_t &\to y_t^{\mathrm{SM}} \approx 1, \\ a_t &\to 1, \\ b_t &\to 0. \end{aligned} \tag{2.53}$$

As discussed in Section 2.3, additional sources of CP violation can help explaining the baryon asymmetry. There are also searches for CP violation in other sectors of the SM and in various couplings. The magnitude of the top-quark Yukawa coupling is a rather unique parameter. Its value being close to unity possibly implies a special relationship between the Higgs-boson and the top-quark.

In electroweak theory, the shape of the Higgs potential has a local minimum at the vev ≈246 GeV, which defines the electroweak scale. It is of interest how the shape of the Higgs potential evolves for larger values of the Higgs field. This shape determines whether the EW vacuum is in a stable, metastable or unstable state. This is demonstrated in Figure 2.1(b) with a one dimensional plot of different effective potential shapes each representing one of the possible outcomes. If there is a global minimum other than our local minima in the EW scale, we are currently in a false, metastable vacuum state and there is a possibility of the vacuum decaying into the global minimum. If the potential is unstable, then there is a probability that the vacuum decays within a timescale shorter than the lifetime of the universe. As such, the electroweak vacuum has also important implications on cosmology and the evolution of the Universe.

The shape of the potential at higher energies can be expressed by using the RGE of the Higgs self-coupling parameter, by expressing it as a running coupling incorporating quantum corrections. At the electroweak scale, the self-coupling parameter λ (as introduced in 2.2.3) and y_t can be calculated using experimentally determined parameters. The Higgs self-coupling is then

$$\lambda = \frac{m_H^2}{2\mathrm{vev}^2} \approx \frac{(125)^2}{2(246)^2} \approx 0.13, \tag{2.54}$$

whereas the top-quark Yukawa coupling is[6]

$$y_t = \sqrt{2}\frac{m_t}{\mathrm{vev}} \approx \sqrt{2}\frac{173}{246} \approx 0.99. \tag{2.55}$$

[6] It should be noted that this value is computed at the LO accuracy. Corrections to y_t when including higher orders are large enough to become relevant in the context of vacuum stability argument discussed here. At the NLO (NNLO) accuracy $y_t \approx 0.95$ (0.94) [41].

Taking only the two leading corrections, the RG evolution equation is approximately

$$\frac{\mathrm{d}\lambda}{\mathrm{d}\ln(\mu)} \approx \lambda^2 - y_t^4. \tag{2.56}$$

where λ is the self coupling constant of the Higgs-boson. Thus, the renormalised running coupling (λ_r) for the Higgs field is approximated as

$$\lambda_\mathrm{r} \approx \lambda + \lambda^2 - y_t^4. \tag{2.57}$$

These quantum corrections are also depicted in terms of Feynman diagrams in Figure 2.3. The crucial role of y_t in connection with the shape of the potential is now clear: It has a large but negative contribution, and thus the stability of the vacuum depends on the interplay between the two couplings, where y_t can push the potential to lower values which could lead to meta- or instability in the vacuum.

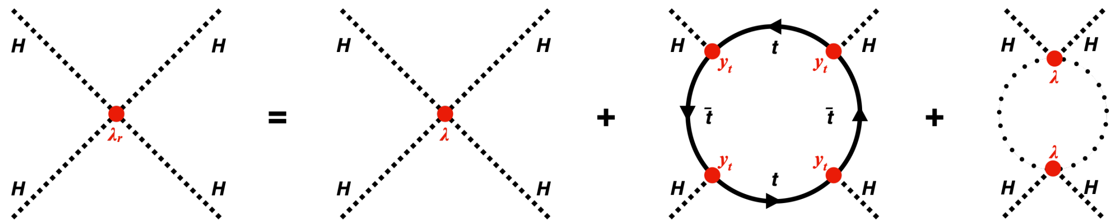


Figure 2.3: Graphical representation of the perturbative expansion of the first terms of the Higgs self coupling parameter λ_r.

In Figure 2.4(a) the running of λ is plotted for a few values of y_t at the various scales. Here, it can be seen that in the SM-like y_t regime of about $y_t \approx 0.93$, the self-coupling of the Higgs-boson becomes negative around 10^{10} GeV. It is important to note that this critical value of the coupling can move orders of magnitude by minuscule changes in y_t. The self-coupling value around $y_t \approx 0.92$ already crosses the abscissa at about 10^{15} GeV, five orders of magnitude higher. In Figure 2.4(b), even smaller variations on this latter y_t value are shown in the context of the Higgs field and effective potential values. Again, the shape of the potential is demonstrated to be very sensitive to small changes in y_t.

It appears that the SM sits at the border between the stable and metastable vacuum with a slight tendency for metastability. The phase transition diagram, expressed in terms of Higgs-boson and top-quark masses, are shown in Figure 2.5 on the left. The zoomed-in version on the right indicates, similarly to the y_t plots above, how sensitive the behaviour of the EW vacuum is to the masses of these two particles.

It should be noted and emphasized that above arguments rely on the assumption that no new physics appear before the Planck scale, and that the SM holds up to that point. Although this is probably not the case, it is interesting to look at the phase diagrams evaluated at the Planck scale (M_Pl) because couplings could carry information related to new physics appearing at a larger scale than they are evaluated at, that is not accessible to us. Top-quark Yukawa and Higgs self coupling phase diagrams evaluated at the Planck scale shown in Figure 2.6 demonstrate that, also at the Planck

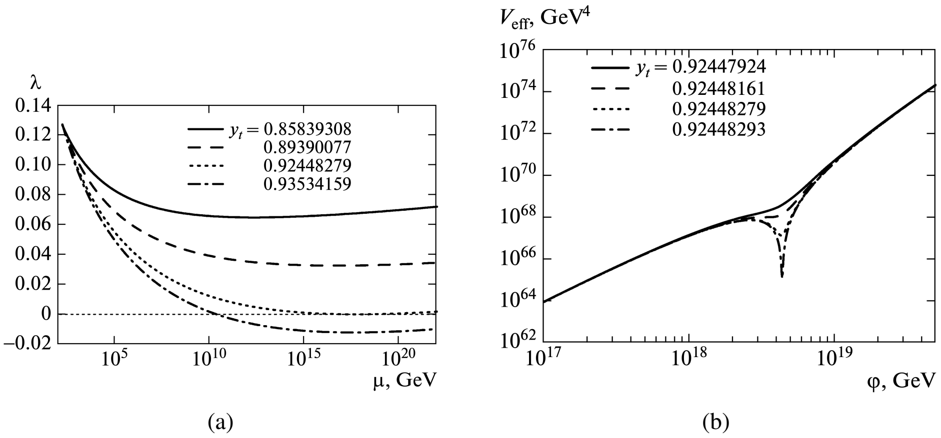


Figure 2.4: Behaviour of (a) the Higgs self coupling parameter λ and of the (b) Higgs field (φ) potential V_{eff}, for few values of top-quark Yukawa coupling y_t in the vicinity of its critical value. The renormalisation scale μ is set to 173.2 GeV. Both figures are taken from Reference [42].

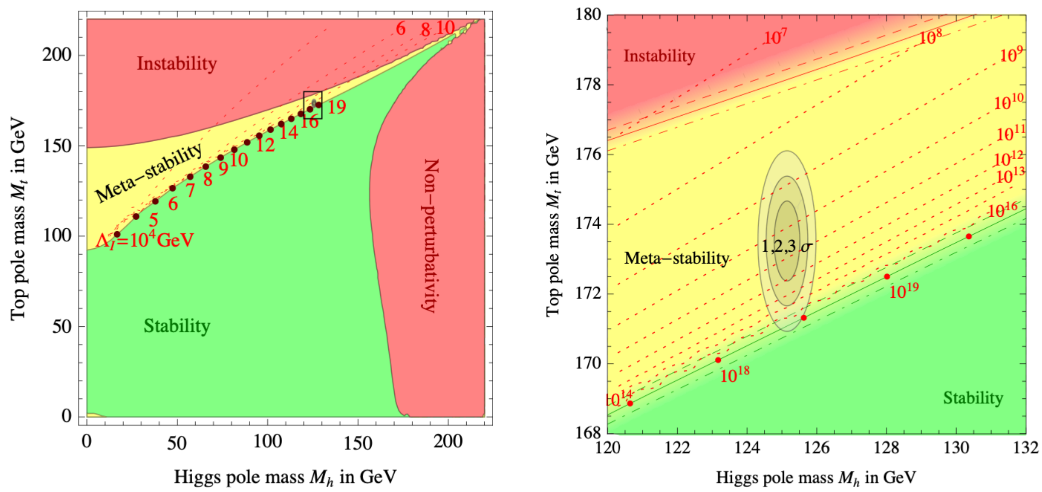


Figure 2.5: Phase diagram of the SM EW vacuum at the EW scale, parametrised in terms of top-quark (M_t) and Higgs-boson (M_h) masses. Both figures are taken from Reference [41].

scale, the Universe described by the SM seems to sit at the edge of a tiny boundary of metastability. Particularly noteworthy is that the calculated phase space point corresponds to the minimum of the metastable region in both the Higgs self coupling and the top-quark Yukawa couplings, and not just one of them. It could be the case that at some intermediate energy between the EW and Planck scales, new physics interfere and modify the shape of the Higgs potential already, effects of which could propagate to the Planck scale measurement. Possible new physics candidates that are discussed in Section 2.3 such as matter–antimatter asymmetry and neutrino masses can be given as examples. On the other hand, there exist also explanations to intermediate-energy BSM physics that do not significantly impact the Higgs potential shape and phase space diagrams discussed here. For example, when explaining neutrino masses with the seesaw mechanism, right-handed neutrino masses are above the instability scale, whereas Yukawa couplings of neutrinos are too small to significantly alter the running of λ, providing a solution that does not impact the Planck scale phase space diagram picture [41]. As such, it is possible that Planck scale couplings and their near-critical values could contain crucial information by themselves. In Reference [41] it is argued that this multiple near-criticality provides an argument for the existence of multiverses assuming it is not a coincidence.

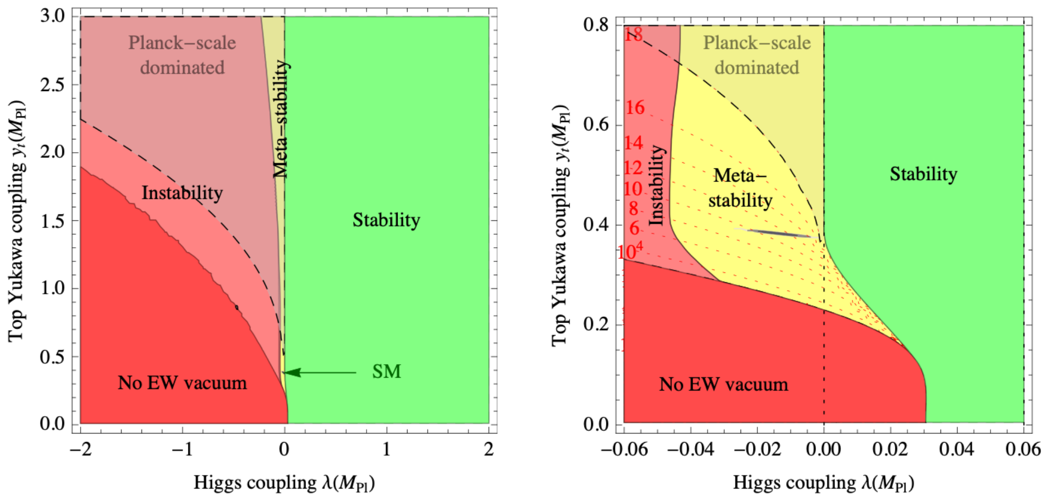


Figure 2.6: Phase diagram of the SM EW vacuum evaluated at the Planck scale ($M_{\rm Pl}$) and parametrised in terms of top-quark Yukawa coupling (y_t) and Higgs self coupling (λ) strengths. Both figures are taken from Reference [41].

2.5 Proton–Proton collisions at the LHC

At a first glance, protons are composite particles made of two up- and one down-type valence quarks, and thus belong to the class of baryons. These valence quarks are held together by gluons. Gluon connecting the valence quarks can transition to virtual quark–antiquark pairs that are called sea quarks. The complex sub-structure of proton emerging from the interplay between quarks and gluons can be approximated using the *parton* model.

A parton is a point-like object that carries a fraction of the momentum of the proton. Quarks and gluons are partons. The interactions between two colliding protons at the LHC can effectively be described and modelled as interactions between their partons. When two protons collide, the interaction between one parton from each proton results in a large momentum transfer which defines the nominal collision event, known as *hard scattering* process. Any type of interaction happening other than the hard scattering process is called the *underlying event* (UE). Examples of underlying events are interaction of beam remnants and interaction of other partons (multiple parton interaction).

In order to calculate the production cross-section of a two-proton hard-scattering process, the associated momentum fractions of involved the partons x_a and x_b need to be determined. These fractions are expressed as probability density functions $f(x_i, Q^2)$ for each parton i to be in possession of momentum fraction x_i of the proton at a given energy scale Q^2. These functions are extracted from available experimental datasets and known as Parton Distribution Functions (PDFs). The most effective way of probing the partons are the Deep Inelastic Scattering (DIS) experiments where leptons are scattered off nuclei, such as the $e^{\pm}$-p collisions that were produced at the HERA collider. Hadron collision data provided by the Tevatron and the LHC are also used.

The cross-section can then be calculated by using the QCD factorisation formula [43]

$$\sigma_{pp\to X}(\hat{s}, \mu_F, \mu_R) = \sum_{ij} \int_0^1 dx_a \int_0^1 dx_b f_i(x_a, \mu_F) f_j(x_b, \mu_F) \hat{\sigma}_{ij}(\hat{s}, \mu_F, \mu_R)_{ab\to X}, \tag{2.58}$$

where $\hat{\sigma}$ is the partonic cross-section, $\hat{s}$ is the centre-of-mass energy of the partonic system related to the centre-of-mass energy of the pp collisions via, $\hat{s} = sx_a x_b$ and μ_F and μ_R are the factorisation and renormalisation scales, respectively. Equation 2.58 combines perturbative (short distance, given by partonic cross-section) and non-perturbative (long-distance, given by PDFs) regimes. The factorisation scale is the cutoff value that defines these two regimes. While the renormalisation scale is used against the loop divergences, the factorisation scale is used to handle infinities arising from radiation emissions from massless particles (collinear divergences). Similarly to beta-functions used to study the evolution of coupling constants with the renormalisation scale, the Dokshitzer–Gribov–Lipatov–Altarelli–Parisi (DGLAP) equation [44, 45, 46] is used to re-evaluate PDFs at different factorisation scale values. In principle, summing over all orders in a calculation leads to an exact result, independently of scales. In practice, calculations are done for finite orders and the effect of higher orders are accounted for by introducing theoretical uncertainties in the scales. In ATLAS, analyses of these uncertainties are typically performed through varying the scales such as by doubling and halving them in a calculation and observing the difference in the result relative to the one obtained with the original scales.

Monte Carlo simulations of collision events

Simulations are an important part of the experimental particle physics research. Two main applications of simulations are detector design and particle collisions. Simulations help designing detectors such that they can have required shape, volume, material, depth and other parameters which would enable them to perform targeted physics measurement. The second application is the simulation of particle physics processes starting from the initial collision and all the way up to the final states. This way different physics processes can be studied with the information provided such as their kinematic distributions and predicted cross-sections. A simulated instance of a given process is called an *event*. Depending on the nature and requirements of the study at hand, the numbers of simulated events can vary, with typical ranges being of the order of $10^5 - 10^7$ events per process. The event simulation involves calculating multi-dimensional phase-space integrals. This is achieved by Monte Carlo random sampling methods, and therefore resulting samples are colloquially referred to as *Monte Carlo samples* (shortened as "MC"). Simulations are performed using programs known as event generators. All event generators split the simulation of the event into several steps, a general scheme of which is given in the following flowchart: Hard scattering → Parton shower → Hadronisation → Particle decay → Detection

As introduced in the previous paragraph, the hard scattering (or hard process) defines the main interaction from which the final state emerges. The computation at this step uses ME, PDF, factorisation and renormalisation scale information as input. Typically, more interactions take place in a collision other than the hard process. These underlying events are also simulated along with the hard scattering.

Partons formed in the hard process and the UE split further by radiating new partons, resulting in a parton shower (PS). The creation of this cascade of new particles costs energy at each vertex split. The termination of the PS depends on the cutoff threshold on energy and some other parameters of the PS algorithm. These internal parameters are tunable and determined based on data provided by the experiments. Several such sets of parameters are grouped into what is called a *tune*. Different tunes can be set for a PS algorithm.

The Lund string model or the cluster model are used to simulate the hadronisation of particles. As generators are based on either model, the choice of the generator also fixes the hadronisation model being used. Since the hadronisation follows directly after the PS, the choice of tunes also has an effect on how the hadronisation step is realised in an event generator.

After hadronisation, some outgoing particles become stable particles with long lifetimes, such that they reach the detector volume. Particles that are unstable will decay into stable products, and these decay products will be interacting with the detector. Although most general-purpose event generators are capable of decaying particles, there exist also tools developed for particular decay processes, which gives more precise results in their particular domain. In top-quark physics, because top-quark decay mostly into heavy hadrons, decays of b- and c-quarks are often simulated by one such specific program called EvtGen [47].

The events generated up to the level of stable particles are then embedded into a virtually constructed detector geometry which simulates the interaction of the particles with matter. This step is typically provided by a different, specialized software. Due to the complicated nature of modelling detector response accurately, they are computationally expensive. For this reason, in ATLAS, fast simulation (FastSim) is occasionally preferred to full simulation (FullSim). The difference between the two is that in FastSim, the calorimeter response is simplified, by approximating the geometry of calorimeter cells through modelling them as cuboids, as well as parameterising the longitudinal and lateral shower profiles. FastSim can shorten the simulation time by several orders of magnitude [48].

There exists a large set of tools for event generation and particle simulation. Amongst them, the simulation tools used in this thesis work are listed and shortly described below.

- PYTHIA[7] [49]: PYTHIA has the functionalities for all steps involved in the generation of an event. It can however also be used in combination with other tools, where certain steps are achieved through different software. PYTHIA uses the Lund string model and only has LO level accuracy in QCD.

- Hadron Emission Reactions With Interfering Gluons (HERWIG) [50]: HERWIG is an event generator capable of producing events at NLO level. Cluster model is used for simulating the hadronisation process.

- MadGraph5_aMCatNLO (MG5) [51]: MG5 can generate MEs at LO or NLO level but cannot simulate parton showers. Events generated with MG5 can be provided to another tool which possesses parton shower capabilities. This process is called interfacing the generators. In the analysis described in this thesis MG5 generated events are interfaced with PYTHIA, meaning PYTHIA is only used for the parton showering step of the event generation.

- Simulation of High-Energy Reactions of PArticles (SHERPA) [52]: Similar to PYTHIA, SHERPA also has a wide range of functionalities. However, unlike PYTHIA, it can also generate QCD events at NLO level. Another difference is that SHERPA employs the Cluster model instead of the Lund string model.

- POsitive Weight Hardest Emission Generator (POWHEG BOX)[53]: POWHEG BOX is an ME generator. It is capable of generating events at NLO level accuracy. Remarkably, events generated with POWHEG BOX have only positive weights.

- GEometry ANd Tracking 4 (GEANT4) [54]: GEANT4 is the software tool that can simulate the interaction of particles with matter. In the case of ATLAS, it used to simulate the detector response to the particles produced by the previously mentioned event generators. GEANT4 has the complete geometry and material budget information of the ATLAS detector.

[7] From the Oracle of Delphi in Ancient Greece.

CHAPTER 3

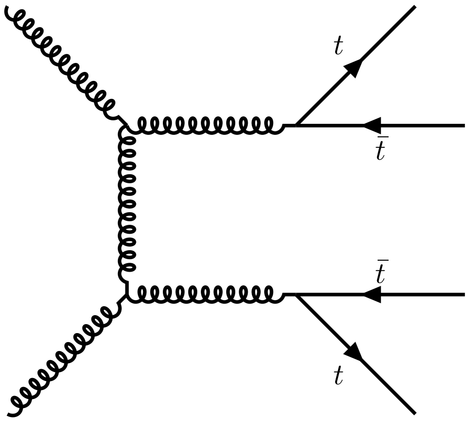
Four top-quarks

As focus of interest for this dissertation, theoretical properties and experimental relevance of the $t\bar{t}t\bar{t}$ process are detailed in this Chapter. The Chapter is divided into five sections. In Section 3.1 the production of the $t\bar{t}t\bar{t}$ process and its theoretical predictions are described. In Section 3.2 various decay channels of the $t\bar{t}t\bar{t}$ process and their properties such as final state composition and branching ratios are given. In Section 3.3 the most important background processes expected in the $t\bar{t}t\bar{t}$ measurement are introduced. Experimental estimation and methods of suppression for these leading, as well as other minor background contributions are detailed in Chapter 6. Previous experimental efforts by the ATLAS and CMS collaborations at centre-of-mass energies of 8 and 13 TeV are summarized in Section 3.4. Finally, the last Section discusses some examples of BSM models that could be probed via the $t\bar{t}t\bar{t}$ process as well as the capability of this process to measure the top-quark Yukawa coupling without using assumptions on the width of the Higgs-boson.

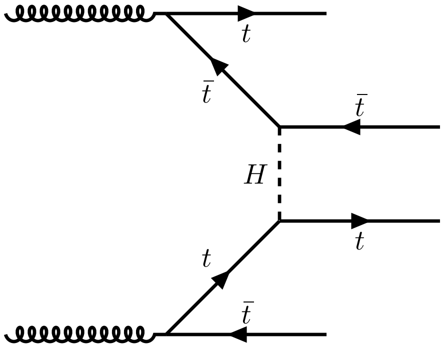
3.1 Production

According to the SM the production of $t\bar{t}t\bar{t}$ is a rare process. For the pp collisions at $\sqrt{s}$ = 13 TeV, the production cross-section calculation of $\sigma_{t\bar{t}t\bar{t}} = 12.0 \pm 2.4$ fb at next-to-leading order (NLO) in QCD, including NLO electroweak corrections, was the prediction with the highest precision available in the literature at the time of the measurement [55]. As such, this prediction is used in the analysis presented in this document. However, it should be noted that current state-of-the-art prediction is different and accounts for the corrections from soft-gluon emissions at the next-to-leading logarithmic accuracy [56]. This latest calculation report the $t\bar{t}t\bar{t}$ production cross-section to be $\sigma_{t\bar{t}t\bar{t}} = 13.4^{+1.0}_{-1.8}$ fb, which corresponds to an enhancement of 15% compared to the NLO prediction. The NLO prediction itself has also led to an approximately 30% increase in the cross-section prediction in comparison to a previous, highest-perturbative-order $\mathcal{O}(\alpha_s^4)$ and QCD-only NLO computation with a calculated cross-section of $\sigma_{t\bar{t}t\bar{t}} = 9.2 \pm 2.6$ fb [57]. Perturbative orders are calculated by counting the number of QCD (EW) coupling vertices, whose strength is proportional to $\sqrt{\alpha_s}$ ($\sqrt{\alpha}$). The square-root convention helps with the counting orders of squared matrix elements $|\mathcal{M}|^2$. The inclusion of lower perturbative orders at LO and NLO in QCD accounts for ≈20% of the enhancement. It is due to the uncertainties in the renormalisation and factorization scales. A dynamical scale choice of $\mu_{\mathrm{R,F}} = H_{\mathrm{T}}/4$ is used in this calculation, where H_{T} is the scalar sum of transverse momenta of all charged leptons and jets in the event. The remaining ≈10% are due to the EW corrections.

This Section follows closely the results and arguments of Reference [55]. Using the notation therein, contributions from each perturbative order are labeled according to the (N)LO order they belong and numbered starting from the highest to the lowest possible QCD orders. LO_1 and NLO_1, in this case, corresponds to $\mathcal{O}(\alpha_s^4)$ and $\mathcal{O}(\alpha_s^5)$, respectively. An example of LO_1 contribution comes from the Feynman diagram shown in Figure 3.1(a). In Figure 3.1(b) an example diagram for LO_3 contribution is shown.



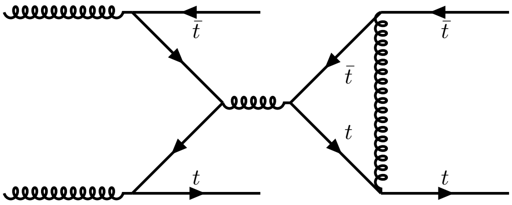
(a) $t\bar{t}t\bar{t}$ at $\mathcal{O}(\alpha_s^2)$



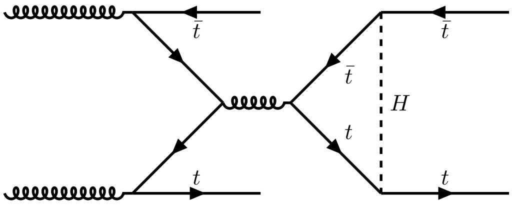
(b) $t\bar{t}t\bar{t}$ at $\mathcal{O}(\alpha \cdot \alpha_s)$

Figure 3.1: Example Feynman diagrams for $t\bar{t}t\bar{t}$ (a) at $\mathcal{O}(\alpha_s^2)$ and (b) at $\mathcal{O}(\alpha \cdot \alpha_s)$.

Interference of these two diagrams leads to an example of the LO_2-level contribution. In Figure 3.2 two examples of 1-loop $t\bar{t}t\bar{t}$ Feynman diagrams are shown. Interference of the diagram shown in Figure 3.1(a) with the diagram in Figure 3.2(a) leads to NLO_1, and with the diagram in Figure 3.2(b) it leads to NLO_2 contributions. Finally, the interference of the diagram in Figure 3.1(b) and diagram in Figure 3.2(b) leads to NLO_3 contributions. The complete list of perturbative orders and their relative contributions with respect to LO_1 are given in Table 3.1. Across all orders, gluon-gluon initiated contributions from LO(1$\rightarrow$ 3) and NLO(1$\rightarrow$ 4) are dominant.



(a) $t\bar{t}t\bar{t}$ at $\mathcal{O}(\alpha_s^3)$



(b) $t\bar{t}t\bar{t}$ at $\mathcal{O}(\alpha \cdot \alpha_s^2)$

Figure 3.2: Example Feynman diagrams for 1-loop $t\bar{t}t\bar{t}$ at orders (a) $\mathcal{O}(\alpha_s^3)$ and (b) $\mathcal{O}(\alpha \cdot \alpha_s^2)$.

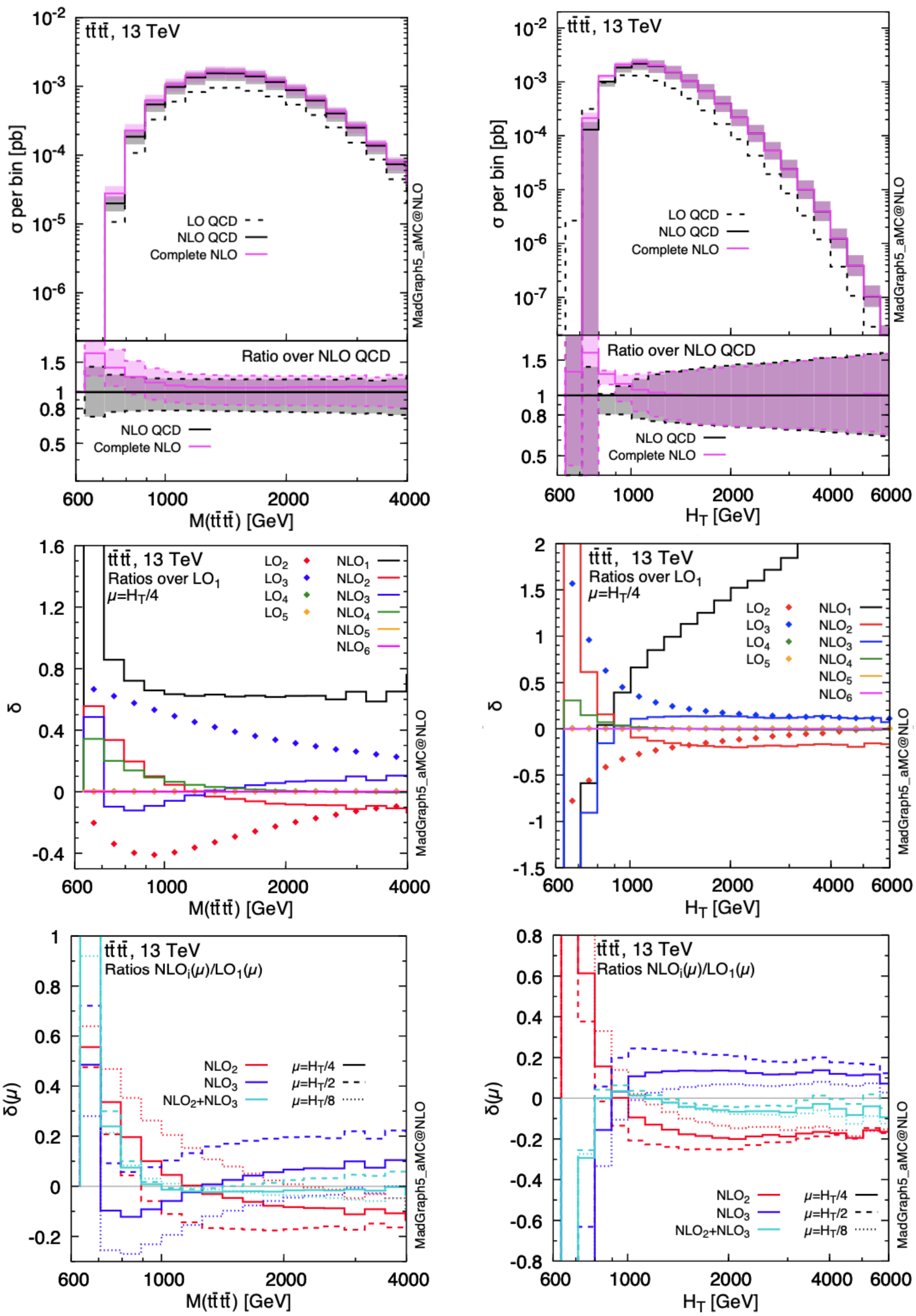
Presence of EW corrections from the $t\bar{t} \rightarrow t\bar{t}$ scattering already at LO (see Figure 3.1(b) for an example Feynman diagram) causes LO_2 and LO_3 corrections to be relatively large compared to QCD-only LO_1 [55]. NLO_2 and NLO_3 contributions include sizeable QCD corrections, inferred from the strong scale dependence of these contributions as can be seen in Table 3.1. At the nominal scale of $\mu_{\mathrm{R,F}} = H_{\mathrm{T}}/4$ used by the analysis, $\Delta\mathrm{NLO}_2$ and $\Delta\mathrm{NLO}_3$ are smallest among the three scales studied, but their values are larger than what would be predicted from their coupling strengths. For example, $\Delta\mathrm{NLO}_3$ is 1.8% which is an order of magnitude larger than the expected value of $\alpha_s^3\alpha^2/\alpha_s^4 = 0.1\%$ [55]. The NLO_2 and NLO_3 contributions partially cancel each other and this cancelation is found to be not affected much by the scale choice [55].

Label	$\lvert\mathcal{M}\rvert^2 \propto$	$\Delta\mathrm{LO}_1$ [%]		
		$H_{\mathrm{T}}/8$	$H_{\mathrm{T}}/4$	$H_{\mathrm{T}}/2$
LO_2	$\alpha.\alpha_s^3$	-26.0	-28.3	-30.5
LO_3	$\alpha^2.\alpha_s^2$	32.6	39.0	45.9
LO_4	$\alpha^3.\alpha_s$	0.2	0.3	0.4
LO_5	α^4	0.02	0.03	0.05
NLO_1	α_s^5	14.0	62.7	103.5
NLO_2	$\alpha.\alpha_s^4$	8.6	-3.3	-15.1
NLO_3	$\alpha^2.\alpha_s^3$	-10.3	1.8	16.1
NLO_4	$\alpha^3.\alpha_s^2$	2.3	2.8	3.6
NLO_5	$\alpha^4.\alpha_s$	0.12	0.16	0.19
NLO_6	α^5	<0.01	<0.01	<0.01

Table 3.1: Relative contribution of different perturbative orders as a percentage of LO_1 for three different $\mu = \mu_{\mathrm{R}} = \mu_{\mathrm{F}}$. Numerical values are taken from Reference [55].

So far, the top-quark Yukawa coupling of the $t\bar{t}t\bar{t}$ process was only studied at LO-level. Details of this study are discussed in Chapter 6. The extent of these seemingly accidental cancelations among sub-leading perturbative orders may be significantly modified by BSM contributions. Thus, LO-level studies may not be sufficient for probing BSM effects in the $t\bar{t}t\bar{t}$ process. One example is the measurement of the top-quark Yukawa coupling y_t. (N)LO_i with $i \geq 2$ and (N)LO_i with $i \geq 3$ contain terms proportional to y_t^2 and y_t^4, respectively.

At the differential level, similar patterns to the inclusive case are observed. Differential distributions for the invariant mass of the $t\bar{t}t\bar{t}$ system ($m_{t\bar{t}t\bar{t}}$) and H_{T} are shown on Figure 3.3. At the high-end tail of the invariant mass distribution, the complete NLO calculation is close to the $\mathrm{NLO}_{\mathrm{QCD}}$ with an approximately constant difference of 10%. This trend is observed despite of the shape dependence of different contributions, because these differences mostly cancel out in the sum. In the vicinity of the on-shell $m_{t\bar{t}t\bar{t}}$, the cross-section from the complete-NLO calculation is almost



(a) (b)

Figure 3.3: (a) Mass and (b) H_{T} distributions for $t\bar{t}t\bar{t}$ production at various perturbative orders [55].

exclusively dominated by the LO_{QCD} contribution. Around the threshold region ($m_{t\bar{t}t\bar{t}} < 900$ GeV) the NLO_2 and NLO_3 contributions are larger than their inclusive counterparts, and display a strong scale and kinematic dependence. The (N)LO_2 contribution reverses sign around 900 GeV while (N)LO_3 increases closer to the invariant mass threshold. This opposite trend removes the cancelation effect. NLO_4 contributions are also present in sizeable amounts. NLO enhancements in the threshold region are attributed to two reasons:

- LO_2 and LO_3 contributions are large in the threshold region and so are their QCD-corrections, contributing to the NLO terms. Kinematic display of scale dependence confirm the QCD origins of $NLO_{2,3}$ corrections and that their sum is less sensitive to scale than the individual contributions of each separately.

- Under non-relativistic conditions, the exchange of heavy particles such as Z and Higgs-bosons among top-quarks prompt *Sommerfeld enhancements*. This effect has been previously observed in $t\bar{t}$ production [58]. Thus, particularly for the $m_{t\bar{t}t\bar{t}}$ distribution, $t\bar{t}t\bar{t}$ processes in which two top-quarks exchange a heavy particle can facilitate such effects. An example Feynman diagram is shown in Figure 3.1(b), where two top-quarks exchange a Higgs-boson.

The H_T distributions show similar patterns as the invariant mass distribution: the threshold region and high-end tail characterise two different regimes. Around the threshold region of $4{\times}m_t$, the NLO_{QCD} distribution differs from the complete-NLO calculation. In this regime, NLO contributions have opposite signs similar to the LO contributions. The H_T distribution includes the p_T of all jets in its definition, therefore it acts as a veto on any possible extra soft jets at the invariant mass limit. In fixed-order perturbation theory such veto on soft jets leads to negative and large QCD Sudakov logarithms [59]. These negative NLO contributions are enhanced in magnitude through Sudakov logarithms, becoming large for $H_T \to 4{\times}m_t$.

Another feature to note in Figure 3.3 is that ΔNLO_2 has a smaller scale dependence at larger H_T values. As LO_2 in those ranges is small, ΔNLO_2 is mostly due to EW effects, explaining the approximate independence of this contribution from the scale. Towards higher values beyond the peak around 1500 GeV, the two predictions become similar due to the cancelation among contributions, and scale uncertainties diminish.

3.2 Decay

The final states of the $t\bar{t}t\bar{t}$ system are determined by the decay of the four W-bosons, each being a decay product of one of the four top-quarks. In this analysis final states with tau leptons are not directly considered. Leptonically decaying tau leptons are classified according to the lighter leptons they decay into. As such, leptons (ℓ) in this analysis refer to either electrons or muons, including those that originate from tau lepton decays. Similarly, hadronically decaying tau leptons are classified as quarks they decayed into. The decay mode where all W-bosons decay hadronically into quarks is called the all-hadronic final state. This final state has a branching ratio of 31%. The all-hadronic final state of the $t\bar{t}t\bar{t}$ process has not been analysed by ATLAS or CMS collaborations so far, due to the overwhelming amount of background contributions expected in this final state that renders it quite insensitive to the signal. As a proton–proton collider, hadronic final states have very high production cross-sections at the LHC. In the all-hadronic $t\bar{t}t\bar{t}$ final state, contributions from all-hadronic $t\bar{t}$ and multi-jet final states are therefore expected to be several orders of magnitude higher than the signal itself.

In case one or more W-bosons decay leptonically, the decay modes are named according to the number of leptons in the final state: one lepton (1L), two lepton (2L), three lepton (3L) and four lepton (4L) final states. The 1L final state has the largest branching fraction with 42%. The 2L and 3L final states comprises 22% and 5%, respectively. The 4L final state has the smallest branching fraction corresponding to 0.5% of expected decays.

Although the above categorization follows a numerical logic, it does not necessarily point to an optimal classification from the physics analysis point of view. An analysis is better divided into final states according to similar background processes they share in common. Thus, different strategies can be developed for each of these final states, that target single or few dominant backgrounds, optimally. Applying such a background-composition-based classification in $t\bar{t}t\bar{t}$ final states, as it is used in the analyses by the ATLAS and CMS collaborations, the 2L final state is split according to the charge of the lepton pair. The 2L final states where two leptons have opposite-sign (OS) charges is called 2LOS. For the 2LOS final state the dominant background is the $t\bar{t}$ process, as dileptonic $t\bar{t}$ decays also have opposite-sign final states. The 2LOS final state is therefore merged into the same category with the 1L final state, where $t\bar{t}$ is also the leading background contribution via its lepton+jets decay channel. This combined 1L/2LOS channel has a branching fraction of 57%. In the 139 fb^{-1} ATLAS analysis of this channel, after the event selection, more than 90% of the background contributions are found to be coming from $t\bar{t}$ events with additional jets.

The final states where two leptons have same-sign charges is called 2LSS. In the 2LSS final state, the $t\bar{t}$ background is suppressed as its decay topology does not have same-sign final states. In this way, the 2LSS final state is similar to the 3L and 4L final states where at least one SS lepton pair is expected in the final state. The combined analysis channel of these three final states with two same-sign leptons or at least three leptons are called 2LSS/3L. Various final states, their classification and branching fractions are visualized in Figure 3.4. The ATLAS measurement performed in the 2LSS/3L channel is the main topic of this thesis. The 2LSS/3L channel has a branching fraction of 13% and a more diverse spectrum of background processes some of which are discussed in the next Section.

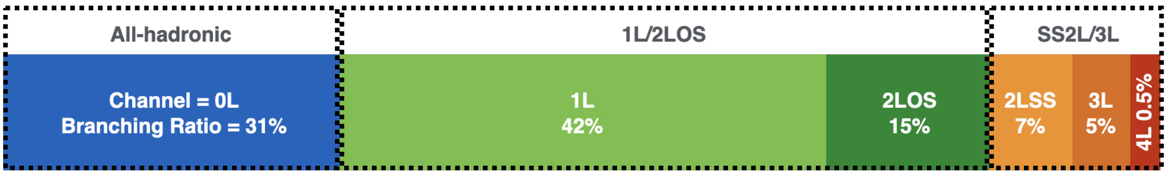


Figure 3.4: Final state categories and branching fractions for the $t\bar{t}t\bar{t}$ process.

3.3 Background processes

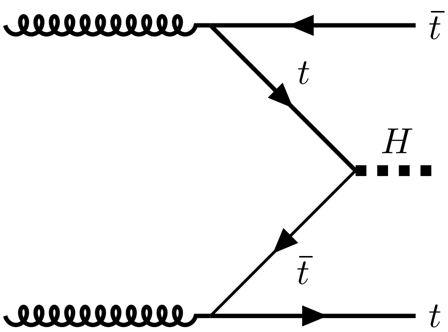
In this Section the properties of the three major background processes in the 2LSS/3L channel are discussed. The complete list of background processes and their properties are provided in Chapter 6.

The final state topology of the $t\bar{t}t\bar{t}$ process is characterised by the multiple number of the same type of particle: four top-quarks. As such, background processes that are most relevant for the $t\bar{t}t\bar{t}$ process are expected to come from similar processes with multiple top-quark final states. $3t$ production is the process with the highest top-quark multiplicity after the $t\bar{t}t\bar{t}$ process. The $3t$ process is indeed found to be the hardest to separate background (see Chapter 6). However, since its production cross-section is an order of magnitude smaller than that of $t\bar{t}t\bar{t}$, its impact on the analysis sensitivity is not the largest one in absolute terms.

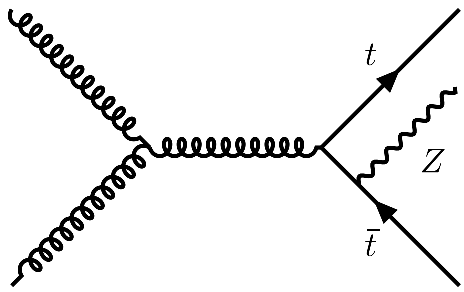
The process with the highest top-quark final state multiplicity after the $3t$ process is the $t\bar{t}$ production. As discussed in the previous Section, the $t\bar{t}$ final state topology by itself is suppressed in the 2LSS/3L channel. Associated production of $t\bar{t}$ with a massive boson can provide final state topologies closer to the $t\bar{t}t\bar{t}$ process through additional decay products provided by these bosons. In contrast to the $3t$ process, the $t\bar{t}H$, $t\bar{t}W$ and $t\bar{t}Z$ processes have approximately two orders of magnitude larger production cross-sections than the $t\bar{t}t\bar{t}$ process. These three background processes account for approximately 75% of the total background contamination in the signal-like phase-space selection of the analysis. Their properties and relevance for the $t\bar{t}t\bar{t}$ analysis are summarized below.

$t\bar{t}H$

At $\sqrt{s}$ = 13 TeV, the $t\bar{t}H$ process has a calculated production cross-section of $\sigma_{t\bar{t}H} = 507 \pm 42$ fb at NLO precision including QCD and EW corrections [60]. A representative LO Feynman diagram of this process is shown in Figure 3.5(a). The $t\bar{t}H$ process is sensitive to the top-quark Yukawa coupling at the first order as it contains diagrams with a vertex that couples $t\bar{t}$ to Higgs-boson. The $t\bar{t}H$ background process contributes ≈15% of the backgrounds.



(a) $t\bar{t}H$



(b) $t\bar{t}Z$

Figure 3.5: Example LO Feynman diagrams for (a) $t\bar{t}H$ and (b) $t\bar{t}Z$ production.

$t\bar{t}Z$

The $t\bar{t}Z$ production process has a cross-section of $\sigma_{t\bar{t}Z} = 839 \pm 94$ fb at NLO precision including QCD and EW corrections at $\sqrt{s} = 13$ TeV at the LHC [60]. The $t\bar{t}Z$ process particularly contributes to 3L final state as it can produce two extra leptons via $Z \to \ell\ell$ decays. In total, this process comprises 20% of the background contribution. A representative LO Feynman diagram for $t\bar{t}Z$ is shown in Figure 3.5(b).

$t\bar{t}W$

The associated production of $t\bar{t}$ with a W-boson at the LHC has a production cross-section of $\sigma_{t\bar{t}W} = 601 \pm 78$ fb at NLO precision including QCD and EW corrections at $\sqrt{s} = 13$ TeV [60]. It is the most dominant background in this analysis with a contribution of about 40%. The $t\bar{t}W$ process differs from the $t\bar{t}Z$ and the $t\bar{t}H$ processes as its dominant production mode is not initiated by gluon-gluon scattering or fusion. Furthermore, in the 2LSS final state, $t\bar{t}W$ production has a 2:1 charge asymmetry favouring positively charged final states. This is an expected outcome at the LHC as protons consist of two valence u-quarks and one valence d-quark. A representative LO Feynman diagram for quark-quark initiated production of the $t\bar{t}W$ process is shown in Figure 3.6(a).

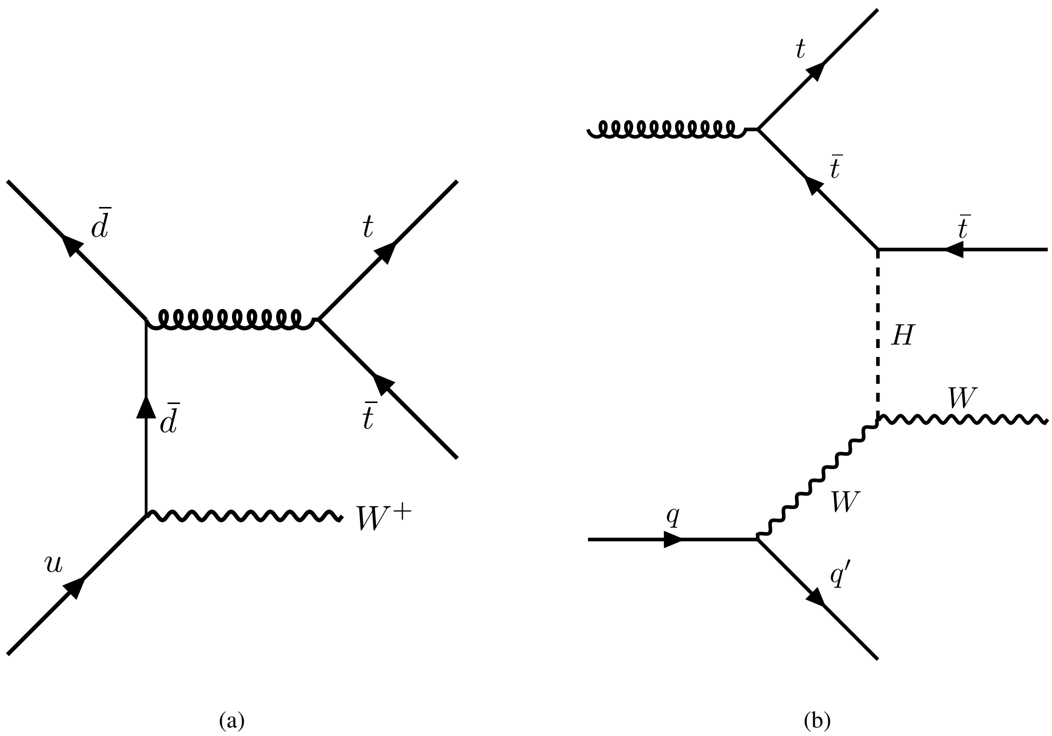


Figure 3.6: Example tree level Feynman diagrams for $t\bar{t}W$ at order (a) $\mathcal{O}(\alpha^{1/2} \cdot \alpha_s)$ and (b) $\mathcal{O}(\alpha^{3/2} \cdot \alpha_s^{1/2})$.

Measurements of the $t\bar{t}W$ production by the ATLAS and CMS collaborations in final states with several leptons have consistently yielded larger cross-sections compared to the ones expected from the simulation [61]. This is particularly the case when the $t\bar{t}W$ cross-section is estimated by an auxiliary measurement in $t\bar{t}H$ analyses. Depending on the final states, the observed excesses are approximately 30–60% larger than the corresponding SM predictions.

Difficulties in the modelling of the $t\bar{t}W$ process in simulation have been attributed to various reasons, some of which are summarised in the following:

- Off-shell effects: Enhancement of the off-shell NLO QCD cross-section at the tails of the differential distributions compared to the Narrow Width Approximation (NWA) are observed [62]. These enhancements are significant because they are not covered by scale uncertainties. An example is shown in Figure 3.7 for the H_T distribution, where an enhancement of about 30% is seen at the tails compared to approximately 10% scale uncertainty in this range.

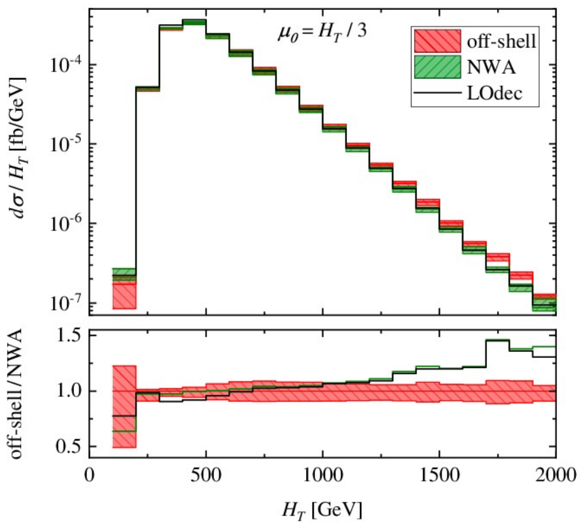


Figure 3.7: Expected H_T distribution of the $t\bar{t}W$ process as predicted by different calculations [62].

- Spin-correlation effects: Spin correlations of the two top-quarks have a sizeable impact on the shapes of lepton rapidities and jet multiplicities of the $t\bar{t}W$ process [63]. In Figure 3.8 it can be seen that in the central region more positively charged leptons are favoured over negatively charged leptons. This, in combination with the 2:1 charge asymmetry of the $t\bar{t}W$ process at the LHC, causes a significant increase in the fiducial cross-section at central rapidities. As such, the selection cuts for the event signature will have an impact on the $t\bar{t}W$ measurement. The spin-correlation effect also enhances the jet multiplicities especially at small multiplicities.

- Sub-leading EW corrections: Studies show that there are large sub-leading EW corrections with differential characteristics [63]. The leading cause for this behaviour is the involvement of scattering between heavy particles (tW-tW), similar in logic to the $t\bar{t}t\bar{t}$ case explained above in Section 3.1. These EW contributions come mostly from $\alpha^3 \cdot \alpha_s$ orders, one Feynman diagram of which is exemplified in Figure 3.6(b). On average, these effects increase the cross-section by about 10%, and at the differential level they are correlated with the jet multiplicity, as shown

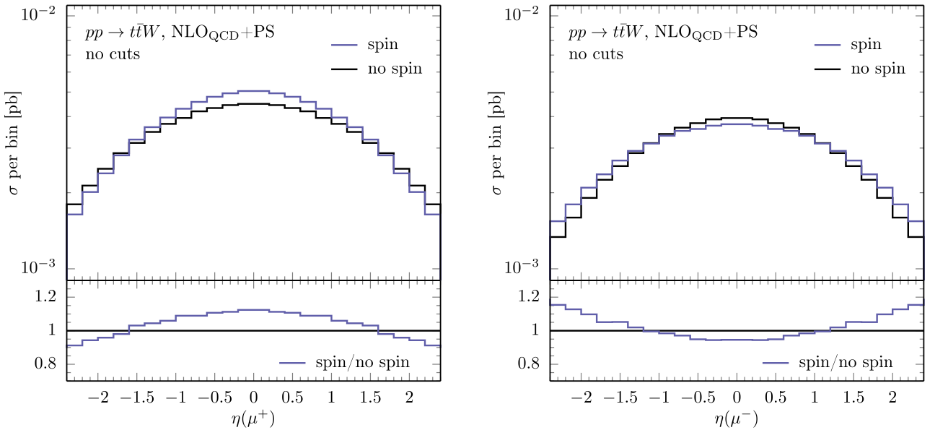


Figure 3.8: Rapidity distribution of positive (left) and negative (right) muons in the $t\bar{t}W$ process with and without spin-correlation effects [63].

in Figure 3.9(b). Although spin-correlation effects (see Figure 3.9(a)) and EW corrections affect different ranges of jet multiplicity, the difference in the shapes of the two distributions prevents their combined effect to be represented through a flat form factor. Thus, the differential distributions need to be accounted for.

- BSM effects: The observed discrepancy between theoretical predictions and experimental measurements of $t\bar{t}W$ processes could be due to some BSM processes that mimic the $t\bar{t}W$ final states, and therefore contribute to the measured cross-section under the SM assumption. One example of a BSM process that could enhance the SM $t\bar{t}W$ cross-section is described in Reference [64] where a new, top-philic neutral spin-1 Z'-boson with a mass above m_t is postulated.

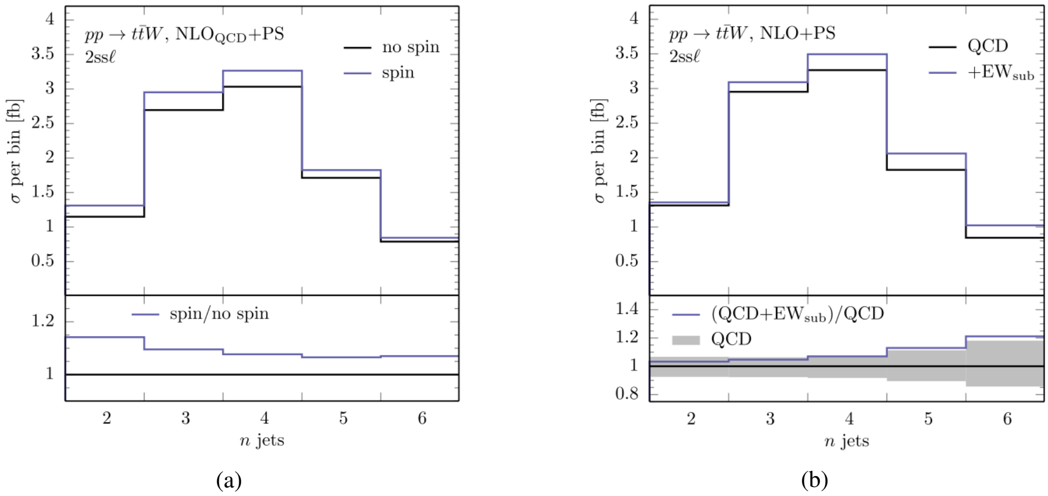


Figure 3.9: Impact of (a) spin-correlation effects and (b) EW effects on jet multiplicity distributions of the $t\bar{t}W$ process [63].

3.4 Previous searches

The ATLAS and CMS collaborations have both performed multiple measurements of the $t\bar{t}t\bar{t}$ production cross-section using LHC data at $\sqrt{s}$ = 8 and 13 TeV centre-of-mass energies targeting different final states. This Section summarizes relevant information of these previous measurements and their findings.

Searches at 8 TeV

In 2014 the CMS collaboration published its first result for the $t\bar{t}t\bar{t}$ process, studying the 1L channel using 19.6 fb^{-1} of pp collisions [65]. The theoretical SM cross-section used in the analysis was about 1 fb, based on a LO prediction at 8 TeV. A kinematic top-quark reconstruction and an event-level Boosted Decision Tree (BDT) are used to improve the signal discrimination against the backgrounds. The observed result was reported as an 95% confidence level (CL) upper limit on the $t\bar{t}t\bar{t}$ production and found to be 32 fb, with the expected limit being 32 fb. The leading systematic uncertainty in this analysis was the cross-section uncertainty on the leading $t\bar{t}$ background process.

The ATLAS collaboration has published two results at 8 TeV, one for the 1L [66] and one for the 2LSS/3L [67] channels. However, unlike the CMS search, these studies are not dedicated SM $t\bar{t}t\bar{t}$ searches. Instead, they were done as parts of studies that were mainly targeting the measurement of various BSM models with similar final state compositions as the $t\bar{t}t\bar{t}$ production. For this reason, it is important to note that these studies were not fully optimized for the $t\bar{t}t\bar{t}$ process. Both analyses use 20.3 fb^{-1} of data collected by the ATLAS detector at the LHC, and use the same SM prediction as the measurement performed by the CMS collaboration. In the 2LSS/3L channel an observed (expected) upper limit of 70 fb (27 fb) at 95% CL is measured, and an excess of 2.5σ is reported for the final states with two b-tagged jets. The 1L channel yields an observed (expected) upper limit of 23 fb (32 fb) at 95% CL and therefore achieves the most stringent limit of all the 8 TeV $t\bar{t}t\bar{t}$ measurements.

Searches at 13 TeV

There exist multiple measurements performed by both collaborations at 13 TeV. Only the most recent results for each of the channels from both experiments are described in this Section.

The most recent measurement by the CMS collaboration at 13 TeV in the 1L/2LOS channel has been performed using 35.8 fb^{-1} of pp collisions [68]. Similarly to the 8 TeV analysis, two BDTs are used in order to increase the signal sensitivity, one for the reconstruction of hadronic top-quarks and another as a general event-level discriminator between signal and the background processes. Furthermore, multiple regions are used according to the jet multiplicity and the multiplicity of b-tagged jets. An observed (expected) upper limit of 48 fb (52 fb) at 95% CL is obtained corresponding to an observed (expected) significance of 0.0σ (0.4σ). The QCD prediction at NLO of 9.2 fb is used as the reference SM value, which was described in the Section 3.1 as the calculation with the highest-order QCD coupling, only. The leading systematic uncertainty in this analysis are the reweighting uncertainties of $t\bar{t}$ +HF processes. The statistical and systematic uncertainties have similar impact.

The analysis in the 2LSS/3L channel is performed by the CMS collaboration using the full Run II dataset of 137 fb^{-1} [69]. As such, this analysis is the closest CMS equivalent to the ATLAS analysis described in this thesis. The analysis is performed using two different approaches defined by the use of cuts or an MVA in the signal region selection. These are referred to as "cut-based" and "BDT" analysis, respectively. The normalizations of the $t\bar{t}Z$ and $t\bar{t}W$ backgrounds together with the $t\bar{t}H$ process are corrected for discrepancies between data and simulation through reweighting methods leading to post-fit normalization factors of 1.3 ± 0.2 for the former two and 1.1 ± 0.3 for the $t\bar{t}H$ background. The analysis is dominated by the statistical uncertainty whose magnitude is almost twice that of the systematic uncertainties. The leading systematic uncertainties are found to be the modelling of additional b-tagged jets and experimental jet-related uncertainties.

In the cut-based analysis, 14 Signal regions[1] (SR) are defined using jet, b-tagged jet and lepton multiplicities. Additionally, two Control regions (CR) are defined for the $t\bar{t}Z$ and $t\bar{t}W$ backgrounds. The measurement yields an upper limit of 20 fb at 95% CL, a measured cross-section of 9.4 fb and an observed (expected) significance of 1.7σ (2.5σ).

The BDT analysis define, 17 SRs through bins of an event-level BDT distribution. In this analysis only a CR for the $t\bar{t}Z$ process is used. The measurement yields an upper limit of 22.5 fb at 95% CL, measured cross-section of 12.6 fb and an observed (expected) significance of 2.6σ (2.7σ), therefore exceeding the sensitivity of the cut-based analysis. An upper limit on the top-quark Yukawa coupling is also set at 1.7 times the SM value.

The latest ATLAS analysis in the 1L/2LOS channel has used the same 139 fb^{-1} dataset, corresponding to the full Run II data, with many common settings with the 2LSS/3L analysis described in this thesis [70]. The main challenge of the analysis has been the careful estimation of the $t\bar{t}$ process with additional jets, which constitutes more than 90% of the background in this final state. The MC simulation of the $t\bar{t}$ process in the presence of additional jets was found to be unreliable and displayed discrepancies between data and simulation. These discrepancies have been corrected for, through multiple steps of reweighting schemes developed in the analysis. The signal extraction has been achieved using event-level BDTs in four different SRs that are exclusive in jet multiplicities and inclusive in b-tagged jet multiplicities. Twelve CRs have been included in the final fit, using the H_T distributions in all but one. The measured cross-section is found to be 26 fb with an observed (expected) significance of 1.9σ (1.0σ), reaching a sensitivity considerably larger than that of the corresponding CMS analysis. The CMS collaboration currently does not have an analysis in this channel using the full Run II dataset. A more complete list of measurements, including those that were published after 2021 and not discussed above, are shown in Figure 3.10.

[1] Concept of Signal and Control regions are introduced in Chapter 5.

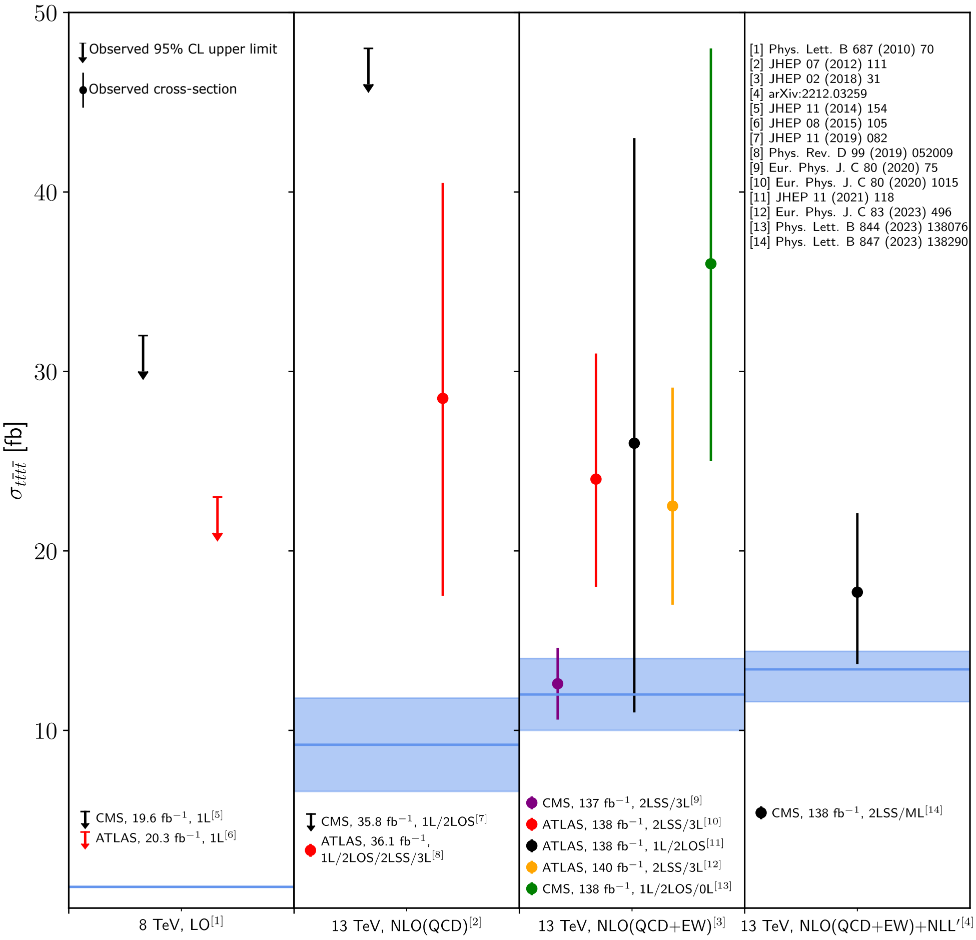


Figure 3.10: Selected $t\bar{t}t\bar{t}$ cross-section measurements performed by the ATLAS and CMS collaborations at different centre-of-mass energies. SM predictions are shown in blue colour, where the line and the rectangle represents the central value and its uncertainty, respectively.

3.5 $t\bar{t}t\bar{t}$ as a probe of new physics

As a rare process with a small cross-section, a sensitivity to Higgs-induced processes, a large mass and a rich final state composition, the $t\bar{t}t\bar{t}$ process can be used to probe various BSM scenarios through testing proposed models or in a model-independent way via EFT formalism.

In one example of a BSM $t\bar{t}t\bar{t}$ model, a new intermediate scalar is postulated with the property that it can decay into a $t\bar{t}$ pair along with two SM top-quarks. A representative Feynman diagram of such a process is shown in Figure 3.11(a) with the resonant scalar represented using the letter X. Various models have been developed that characterise and interpret the nature of such a scalar. In some simplified top-philic dark matter models [71], this intermediate particle is a dark matter candidate. In Type II two-Higgs-Doublet Models, it can be a heavy Higgs-boson or a pseudo-scalar [72]. A non-exhaustive list of further BSM models proposed to be relevant for the $t\bar{t}t\bar{t}$ process are listed in References 1-10 of Reference [73].

An example of an EFT analysis is the Top Contact Interaction where right-handed top-quarks are assumed to be composite particles. In EFT terms this implies four-fermion operators to become significant and as a result the $t\bar{t}t\bar{t}$ cross-section could become approximately 10^3 times larger than its SM prediction. An example Feynman diagram for the contact interaction is shown in Figure 3.11(b).

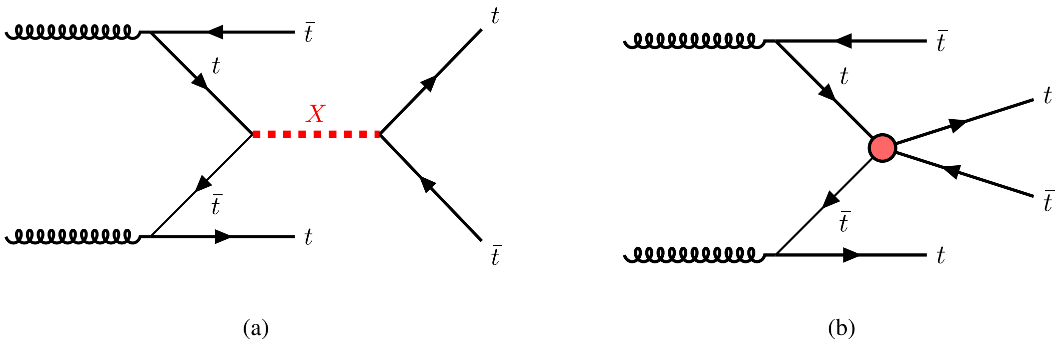


Figure 3.11: Example Feynman diagrams for (a) intermediate resonant scalar and (b) contact interaction BSM scenarios. The BSM parts are coloured in red.

3.5.1 Extraction of top-quark Yukawa coupling

Although the top-quark Yukawa coupling is not a BSM parameter, an observation of deviations from its SM prediction could reveal important hints about the scale and characteristic of new physics, as detailed in Section 3.5. This Section follows the work of Reference [74], where it is argued that the $t\bar{t}t\bar{t}$ process provides a unique opportunity to estimate the top-quark Yukawa coupling without making assumptions on the width of the Higgs-boson. Using the kappa-framework notation [75], deviations from SM Yukawa couplings shown in Figure 3.12(a) are parametrised as

$$\kappa_t = \frac{y_t}{y_t^{\mathrm{SM}}}, \kappa_x = \frac{y_x}{y_x^{\mathrm{SM}}} \tag{3.1}$$

where x can be any final state particle allowed in the SM.

The final state production cross-section can be approximated using the narrow width approximation,

$$\sigma(pp \to t\bar{t}H \to t\bar{t}xx) = \sigma^{\mathrm{SM}}(pp \to t\bar{t}H \to t\bar{t}xx)\kappa_t^2\kappa_x^2\frac{\Gamma_H^{\mathrm{SM}}}{\Gamma_H} \tag{3.2}$$

$$= \sigma^{\mathrm{SM}}(pp \to t\bar{t}H \to t\bar{t}xx)\mu_{t\bar{t}H}^{xx} \tag{3.3}$$

with

$$\mu_{t\bar{t}H}^{xx} = \frac{\sigma}{\sigma^{\mathrm{SM}}} = \kappa_t^2\kappa_x^2\frac{\Gamma_H^{\mathrm{SM}}}{\Gamma_H}. \tag{3.4}$$

A cross-section measurement can be used to estimate y_t only after assuming values for κ_x and Γ_H. For the former, $\kappa_x \approx 1$ is assumed if the coupling to BSM x particles are very weak, such as when x represents invisible particles. For the latter, $\Gamma_H \approx \Gamma_H^{\mathrm{SM}}$ is used if the Higgs-boson decay modes are rare, such as $H \to \mu\mu$ or $H \to \gamma\gamma$. Due to their low branching fraction, rare decay modes are not expected to modify the decay width significantly. In the case of large branching fractions, y_t is estimated by assuming that the decay width is equal to its SM value. An example is the y_t measurement performed by the ATLAS collaboration using the $H \to b\bar{b}$ decay process [76], which is the Higgs-boson decay mode with the largest branching fraction.

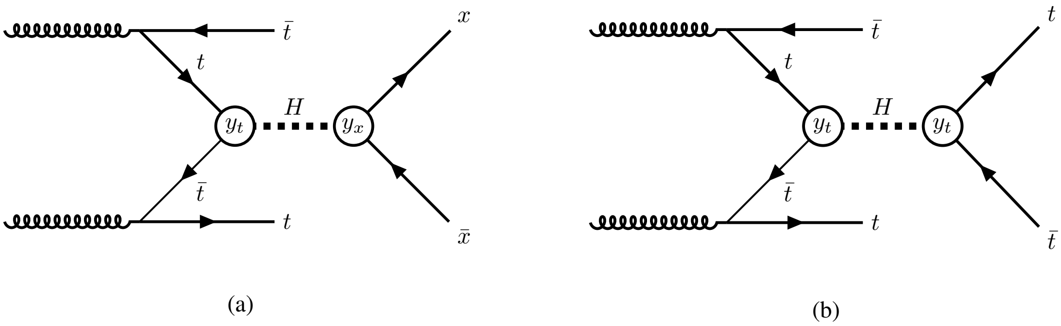


Figure 3.12: Example Feynman diagrams for $t\bar{t}t\bar{t}$ final states with an intermediate Higgs-boson decaying into (a) $H \to x\bar{x}$ and (b) $H \to t\bar{t}$.

The $t\bar{t}t\bar{t}$ final state for this Higgs-mediated process is realized if $x = t$ as shown in Figure 3.12(b). Having a top-quark Higgs-boson coupling vertex means that κ_x is equal to y_t and no assumptions are needed for the coupling. The cross-section is then dependent on the fourth order power of y_t. Since the Higgs-boson has a lower mass than the top-quark, the propagator is dominated by off-shell contribution,

$$\frac{1}{(s - M_H^2) + M_H\Gamma_H} \approx \frac{1}{(s - M_H^2)} \tag{3.5}$$

and therefore the Higgs-boson width can be ignored, rendering the measurement independent of the assumptions on its value. This is an important feature for the experimental efforts because the width of the Higgs-boson is $\approx \mathcal{O}(2)$ smaller than the typical detector resolution.

CHAPTER **4**

ATLAS detector at the LHC

The data used in this dissertation is collected by the ATLAS detector and originates from the proton–proton (pp) collisions provided by the Large Hadron Collider (LHC). Both structures are located at and hosted by the European Organisation for Nuclear Research, known as CERN, the french abbreviation for the name of the provisional council which laid the groundwork for the establishment of the institution: *Conseil Européen pour la Recherche Nucléaire*. CERN was officially founded on 1954 and is located in the vicinity of the Franco-Swiss border near the city of Geneva. This Chapter is divided into three sections describing in the following order: the LHC and its operation, the structure of the ATLAS detector, and the particle detection methods in the ATLAS detector.

4.1 The Large Hadron Collider

The LHC is an underground, approximately circular particle accelerator with a circumference of 27 km [77]. It is housed in the same, 50–175 m deep tunnel that was previously used by the Large Electron Positron (LEP) collider [78]. The LHC has two separate, concentric, high vacuum, and magnetically regulated beam pipes that can accelerate protons or ions in opposite directions. 1232 superconducting dipole magnets create a 8.3 Tesla (T) strong magnetic field to maintain the circular trajectory of the beams. Additional 392 quadrupole magnets are used to counteract against the spread of beams and hold them focused. In this dissertation only data collected from proton collisions are used and therefore ion collisions are not discussed. In order to finally collide the proton beams, the acceleration of particles occur at several steps called the injection chain, involving other pre-accelerators. Protons are initially produced by ionising gaseous hydrogen with the help of an electric field, that are then fed into the Linac2 linear accelerator. At the end of the 33-meters linear acceleration track, protons reach 50 MeV of energy and are passed on to the circular Proton Synchrotron Booster (PSB) accelerator. Protons leave the PSB with a 1.4 GeV of energy and are further accelerated at the Proton Synchrotron (PS) up to 25 GeV. Finally, they reach an energy of 450 GeV at the Super Proton Synchrotron (SPS) and are injected into the LHC ring. This injection chain is summarized below and shown in Figure 4.1.

$$0 \rightarrow \text{Linac2} \rightarrow 50\text{ MeV} \rightarrow \text{PBS} \rightarrow 1.4\text{ GeV} \rightarrow \text{PS} \rightarrow 25\text{ GeV} \rightarrow \text{SPS} \rightarrow 450\text{ GeV} \rightarrow \text{LHC}$$

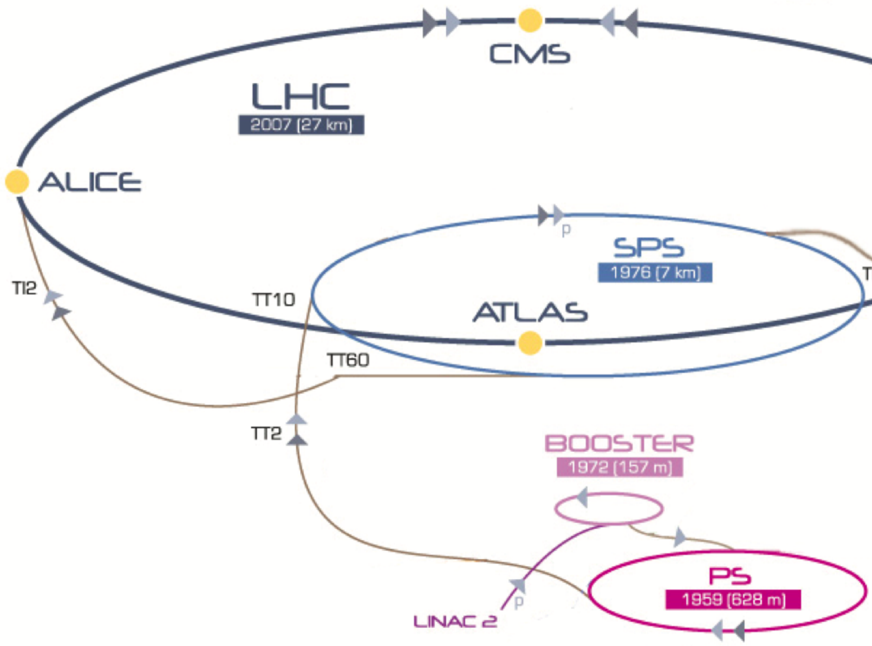


Figure 4.1: Schematic depiction of the LHC injection chain. The largest ring is the main LHC ring with each of the four main experiments (ALICE, ATLAS, CMS and LHCb) built around an interaction point, highlighted by a yellow filled dot. The graphic is taken from Reference [79].

The injected beams are bundled and grouped in a specific way. This is necessitated by the fact that oscillating electric fields are used to accelerate protons. Thus, the timing of arrival of particles into the region with electric field has to be synchronised with the oscillation frequency such that particles are always accelerated. This is not possible in the case of continuous particle injection into the ring. First, about 10^{11} protons are grouped together in what is called a *bunch*. Then every 72 bunches are arranged together by having 25 ns spacing between each other, creating *trains*. Each train is separated from one another by a space equivalent to 12 bunches. At LHC, every beam has a design value of 2808 bunches. For the main ring, eight radio-frequency cavities are designed to accelerate proton bunches up to 7 GeV. The data used in this dissertation is from the data taking period 2015–2018, known as the Run II, where 6.5 GeV of proton beam energy was achieved, leading to a centre-of-mass energy ($\sqrt{s}$) of 13 TeV. Selected operational parameters of the LHC and ATLAS during different data-taking periods of Run II are listed in Table 6.2. The amount of data a collider can provide is proportional to the number of collisions it can deliver. At LHC, for two beams of proton bunches this can be quantified by the instantaneous luminosity

$$\mathcal{L} = \frac{N_p^2 N_b f}{4\pi\sigma_x\sigma_y} \tag{4.1}$$

where N_p is the number of protons per bunch, N_b the number of proton bunches, f the revolution frequency, σ_x and σ_y the spread of the beam in the corresponding axis. The peak instantaneous design luminosity of LHC is $\mathcal{L}_{\text{peak}} = 10^{34}\text{cm}^{-2}\text{s}^{-1}$.

The total amount of collisions in a given time interval is then given by the integrated luminosity

$$L = \int \mathcal{L}\, dt.$$

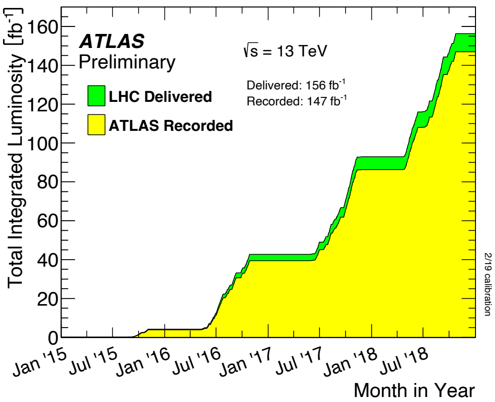
For a process x with a cross-section $\sigma(pp \to x)$, the expected number of events $N(pp \to x)$ is given by

$$N(pp \to x) = L \times \sigma(pp \to x). \tag{4.3}$$

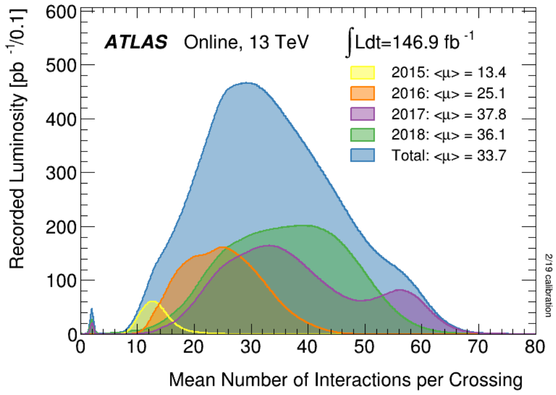
Year	$\sqrt{s}$ [TeV]	L_{LHC} [fb^{-1}]	L_{ATLAS} as % of L_{LHC}	$\langle\mu\rangle$	N_p [$\times 10^{11}$]	max(N_b)	% of $\mathcal{L}_{\text{peak}}$
Design	14	-	100	25.0	1.15	2808	100
2015	13	4.0	88.8	13.4	1.10	2244	50
2016	13	38.5	93.1	25.1	1.10	2076	138
2017	13	50.2	95.7	37.8	1.10	2556	209
2018	13	63.4	97.5	36.1	1.10	2556	214

Table 4.1: Selected operational parameters of the LHC and ATLAS during Run 2. $\sqrt{s}$ is the centre-of-mass energy, L_{LHC} and L_{ATLAS} are the total integrated luminosity of the LHC and ATLAS respectively, $\langle\mu\rangle$ is the number of average pile-up interactions, N_p is the number of protons per bunch, N_b the number of proton bunches, $\mathcal{L}_{\text{peak}}$ is the peak instantaneous luminosity. Values are taken from References [80] and [81].

During Run II, LHC had delivered 156 fb^{-1} of data, 147 fb^{-1} of which was recorded by the ATLAS detector. The accumulation of the delivered and recorded data over the years are shown in Figure 4.2(a). Another notable parameter is the average *pile-up* ($\langle\mu\rangle$), defined as the number of average pp collisions in a bunch crossing, predominantly stemming from QCD interactions. Shape of pile-up distributions, known as *profiles* are shown in Figure 4.2(b). An event display of a collision event resulting in pile-up is shown in Figure 4.8.



(a)



(b)

Figure 4.2: Luminosity accumulation and pile-up profiles during Run II. In (a) the total integrated luminosity delivered by the LHC (green) and recorded by the ATLAS (yellow) over the duration of data taking is shown. In (b) average number of pile-up interactions in 2015 (yellow) , 2016 (orange) , 2017 (purple) , 2018 (green) and their average (blue) are shown. Both figures are taken from Reference [82].

4.2 The ATLAS detector

ATLAS[1] [83] is a cylindrical particle detector weighing 7000 tonnes with a length of 46 meters and a diameter of 25 meters [83]. A layered, computer-generated drawing of the ATLAS detector with its main components labelled are shown in Figure 4.3.

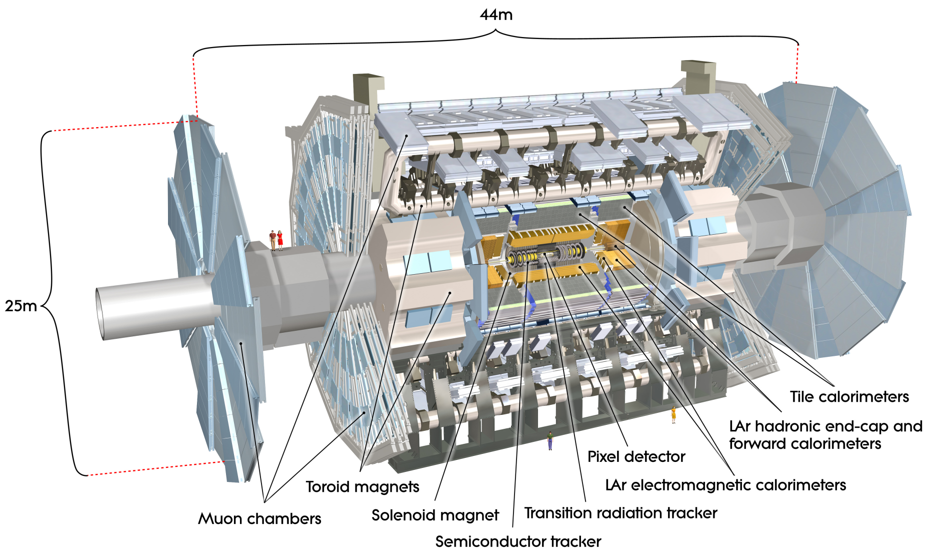


Figure 4.3: Computer-generated view of the ATLAS detector. The drawing is copyrighted by CERN [84].

[1] The name ATLAS is both an acronym (A Toroidal LHC Apparatus) and a backronym (The titan Atlas in the Greek mythology as represented on the logo of the ATLAS collaboration).

It is placed so that the central axis of the cylinder overlaps with the LHC beam pipe and its centroid coincides with one of the LHC interaction points (IP) where the collisions happen. A Cartesian coordinate system is commonly used in ATLAS with the IP as the origin, the z-axis along the beam pipe, the x-axis pointing towards the centre of the LHC ring and the y-axis pointing vertically upwards. In the x-y plane, an important quantity is the transverse momentum, defined as

$$p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2}, \tag{4.4}$$

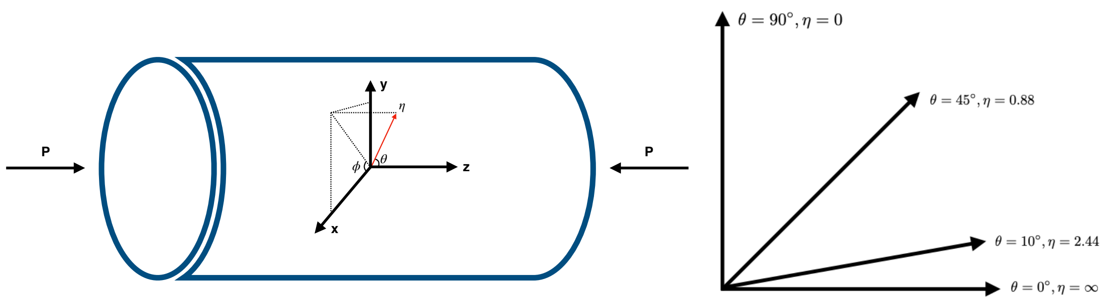
which is the transverse component of the momentum of a particle. Since the initial momentum on the transverse plane is zero, the law of momentum conservation can be invoked when working with p_{T}. There are also quantities expressed in cylindrical coordinates. ϕ is the azimuthal angle with a range of $[-\pi, +\pi]$. θ measures the angle starting from the positive z-axis. This angle is important in measuring the angular difference between the incoming beams and outgoing particles produced in the collisions. The difference between the polar angles θ_1 and θ_2 of two particles is not a Lorentz invariant quantity. Instead, the rapidity variable, defined as

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

can be used. The rapidity difference is a Lorentz invariant quantity. For practical purposes, a transformed quantity called pseudo-rapidity,

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) \tag{4.6}$$

is used, which is a valid approximation to the rapidity for particles with $E \gg m$. Cartesian and cylindrical coordinates are shown in the context of a simplified ATLAS geometry in Figure 4.4(a). The relationship between η and θ is shown for few example values in Figure 4.4(b). The pseudo-rapidity is zero at the right angle ($\theta = 90^\circ$) to the z-axis and increases as the angle θ decreases, reaching infinity at the axis corresponding to $\theta = 0^\circ$.



(a) (b)

Figure 4.4: (a) Schematic depiction of Cartesian and cylindrical coordinates in the ATLAS detector. The blue cylinder represents the ATLAS detector volume, incoming protons are labelled with the letter “P”. (b) Corresponding values of η and θ for given values.

The distance between two objects i and j is defined as

$$\Delta R = \sqrt{(\Delta\eta_{ij})^2 + (\Delta\phi_{ij})^2}. \tag{4.7}$$

ATLAS is designed as a hermetic detector composed of three sub-detector units (Inner Detector, Calorimeter and Muon Spectrometer) each of which has its own complex sub-components. The detector units have two main geometrical structures that complement each other in achieving good coverage. The cylindrical component parallel to the beam pipe covers the central region and is called the *barrel*. Forward regions that go beyond the edges of the barrel are covered by *endcaps* that are generally wheel or disc shaped, and lie perpendicularly to the beam pipe.

4.2.1 Inner detector

The inner detector (ID) is the first sub-component enclosing the beam pipe. It has a length of 6.2 m and a radius of 1.1 m, corresponding to a coverage of $|\eta| < 2.5$. The ID mainly provides tracking and momentum information for charged particles through information provided by three different units that are immersed in a 2 T magnetic field. The proximity to the collision point means ID is the sub-detector unit that is most impacted from radiation damage. In order of increasing radius these three units are described below and a schematic cross-section drawing is given in Figure 4.5 on the left.

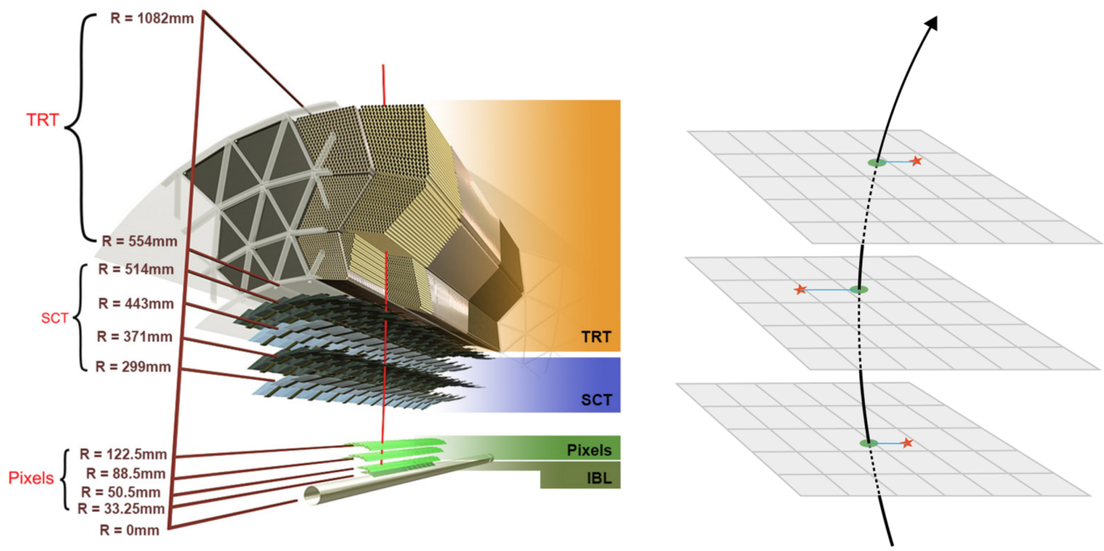


Figure 4.5: (Left) Computer-generated view of the ID with its sub-detectors. (Right) Trajectory of a particle established from hits recorded by three layers of the detector [85].

Silicon pixel detector

The silicon pixel detector [86] occupies the space from the radius of 33.25 mm to 122.5 mm and comprises four layers. Particles enter the ATLAS detector at the first pixel detector layer called *Insertable b-layer* (IBL) [87]. The detector design strategy of ATLAS is such that, granularity of units increase closer to the beampipe. The benefit of this approach is to have high spatial resolution resulting in improved tracking and vertexing capabilities as well as to be able to handle higher particle rates. Therefore, in the innermost layers, large number of pixel detectors are used. A *pixel* is defined by a unit called *module*, which consists of an active semiconductor detector material and associated electronics that can read out the response. The response is an electric signal created due to the voltage difference induced by ionisation of the active semiconductor material when a charged particle traverses it. The IBL has 12 million pixels each having a size of 50 x 250 μm^2. In the $r - \phi$ plane the IBL has a spatial resolution of 8 μm. In z-direction the resolution is five times lower and about 40 μm. The IBL is particularly useful-as the name suggests-for the b-tagging performance.

The three remaining layers are accompanied by three endcap discs and they share similar technical capabilities. The pixel size for these layers are 50 x 400 μm^2 which is 60% larger than that of IBL. Spatial resolution in the $r - \phi$ plane is 10 μm and 115 μm in the z-direction. The pixel detector has in total 92 million pixels.

Semiconductor tracker

The second unit is the Semiconductor Tracker (SCT) [88] occupying the radius range of 299-514 mm. The SCT has four barrels and two endcaps. Both structures are equipped with silicon chip modules know as *microstripes*, using semiconductors similar to pixels, but with module units in the shape of thin stripes. Each barrel layer has 2112 modules. Each endcap has 988 modules and is composed of nine discs. Only the spatial information in the plane perpendicular to it can be obtained from a single module. Thus, strip layers are arranged in a crossed, twisted pattern in order to have spatial information in two dimensions. The crossing angle between layers is 40 mrad. Choosing a non-orthogonal crossing angle has certain benefits. For example, it allows for a geometrical structure where electronic parts of each module face to the same side and do not obscure the active detection material of another module. The SCT has a spatial resolution of 17 μm in $r - \phi$ plane and 580 μm on the z-axis.

Transition radiation tracker

The outermost ID unit is the Transition Radiation Tracker (TRT) which comprises various sizes and arrangements of straw tubes. A straw is a drift tube filled with a gaseous mixture that functions as a proportional counter. In the barrel region, 1.5 m long straw tubes are placed parallel to the beam pipe. 52544 parallel straws covers a range of $|\eta| < 1$ in a arrangement of 73 layers on three cylindrical rings. At the each endcap region 12288 shorter straws of 0.4 m are placed in an arrangement of eight layers each having 768 straws that are perpendicular to the beam pipe in a radially outward manner, allowing detection in the $1 < |\eta| < 2$ range. The large length of the detector system enables the TRT to contribute to tracking and momentum measurements. The energy deposited into the straws are enhanced by filling the spaces between the straws with polypropylene or polyethylene fibres in the barrel region and polypropylene foils in the endcap regions. Transition radiation emitted from these materials are absorbed by the gas in the straw tubes. The TRT is particularly useful in electron identification by helping to distinguish them from pions. This is because the probability of emitting transition radiation is dependent on the Lorentz factor γ. Thus, it is more likely for lighter particles, like electrons, to radiate compared to relatively heavier pions.

4.2.2 Calorimeters

The middle sector of the detector is instrumented with calorimeters that are used to measure the energy of the particles by absorbing them. The ATLAS calorimeters comprise three main units: electromagnetic and hadronic calorimeters accompanied by a forward calorimeter capturing the particles with trajectories closer to the beam pipe. These components are shown in a computer generated image in Figure 4.6. The ATLAS calorimeters do not solely consist of continuous passive absorbing material, but are endowed with a layered internal structure which allows them to probe depth and therefore register the profile of showers developed by the particles traversing them. This type of detectors are called sampling detectors.

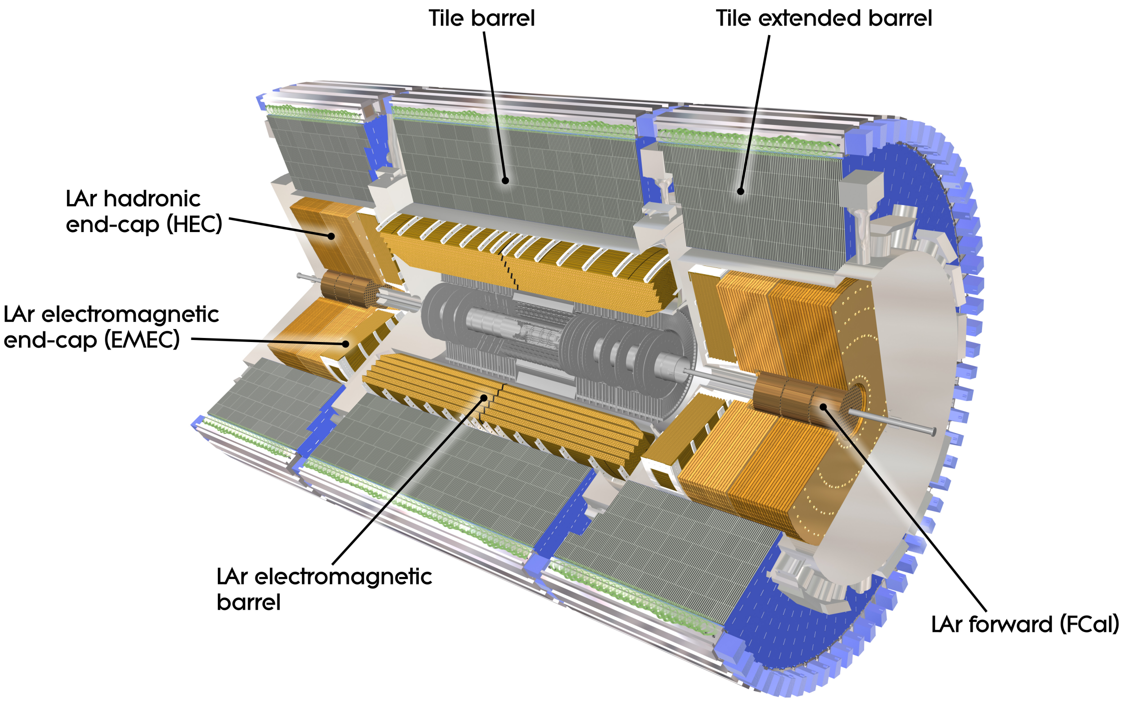


Figure 4.6: Computer-generated cutaway view of the different ATLAS calorimeter sub-systems [89].

Electromagnetic calorimeter

When moving radially outwards from the ID; the first calorimeter unit is the electromagnetic calorimeter (ECAL) which is designed to measure the energies of electrons and photons. A small pre-sampler is placed between the TRT and the ECAL that corrects for the energy loss due to interactions with the detector components up to that point such as with the ID, the cryostat and the solenoid coil. The pre-sampler as well as the ECAL relies on liquid Argon (LAr) as active material that is ionised by the incident particles. With the help of an electric field, ionised charges are collected by copper electrodes.

The barrel region is 22 radiation lengths long, covers a range of $|\eta| < 1.475$. It comprises an alternating series of active LAr and passive lead absorber volumes formed in an accordion-like geometry. The electromagnetic endcap detector (EMEC) is 24 radiation lengths long and covers a total range of $1.375 < |\eta| < 3.2$ with two coaxial wheels split at $|\eta| = 2.5$. The region $1.375 < |\eta| < 1.52$ has additional support and structural material (cables, cooling utilities etc.) used for the operation of the detector and therefore has reduced detection capabilities. This gap is known as the *LAr crack region* and data collected from this sector is often not used in ATLAS physics analyses.

Hadronic calorimeter

The hadronic calorimeter (HCAL) begins where the ECAL ends. The HCAL is designed to effectively sample hadronic showers coming from mesons and baryons. The hadronic endcap calorimeter (HEC) is similar in structure and shape to its electromagnetic counterpart: A LAr calorimeter with two coaxial wheels on either side of the barrel region, covering a range of $1.5 < |\eta| < 3.2$. The barrel region comprises 64 tile modules each with an alternating active-passive volume structure. Plastic scintillators of 3 mm thickness are used as active material and interleaved with volumes of steel that act as passive material. Signals from each tile module are amplified by Photo-multiplier tubes (PMTs) attached to its edge.

Forward calorimeter

The forward calorimeter (FCAL) extends the pseudo-rapidity coverage of the two calorimeters from $|\eta| < 3.1$ to $|\eta| < 4.9$. It has three layers of equal depth with a total depth of 10 interaction lengths. The innermost layer is similar to the ECAL with LAr as active material, but has copper instead of lead as passive material. The remaining two layers are designed to capture hadronic showers and employ tungsten as passive material.

4.2.3 The Muon spectrometer

The third and outermost component of the ATLAS detector is the muon spectrometer (MS) which has a coverage of $|\eta| < 2.7$. The MS is designed to optimally provide tracking and momentum measurement information for muons. This is achieved by a combination of four distinct sub-detectors immersed in a strong magnetic field ranging from 3.9 T in the barrel to 4.1 T at the endcap region. A cross-sectional view featuring sub-detectors are shown in Figure 4.7. Monitored Drift Tubes (MDT) are the most abundant detectors in the MS and used for high accuracy tracking. MDTs are tubes with a width of 3 cm, made of aluminum and filled with gas. More than 380000 MDT tubes are arranged in layers of 3 to 8, and cover a range of $|\eta| < 2.7$. Cathode-Strip Chambers (CSC) are multi-wire proportional chambers that complement the function of MDTs in the endcap regions with a coverage of $2.0 < |\eta| < 2.7$. There exist 70000 CSC channels. Most of the central region is instrumented with over 1100 Resistive Plate Chambers (RPCs) at $|\eta| < 1.05$. RPCs are gaseous detectors working in avalanche mode. In the middle pseudo-rapidity range, $1.05 < |\eta| < 2.4$, another type of multi-wire proportional chambers called Thin-Gap Chambers (TGCs), are used. TGCs provide 2-7 mm resolution depending on the angle. Thanks to their fast response times, RPCs and TGCs play an important role in trigger decisions, which are described in the next Section.

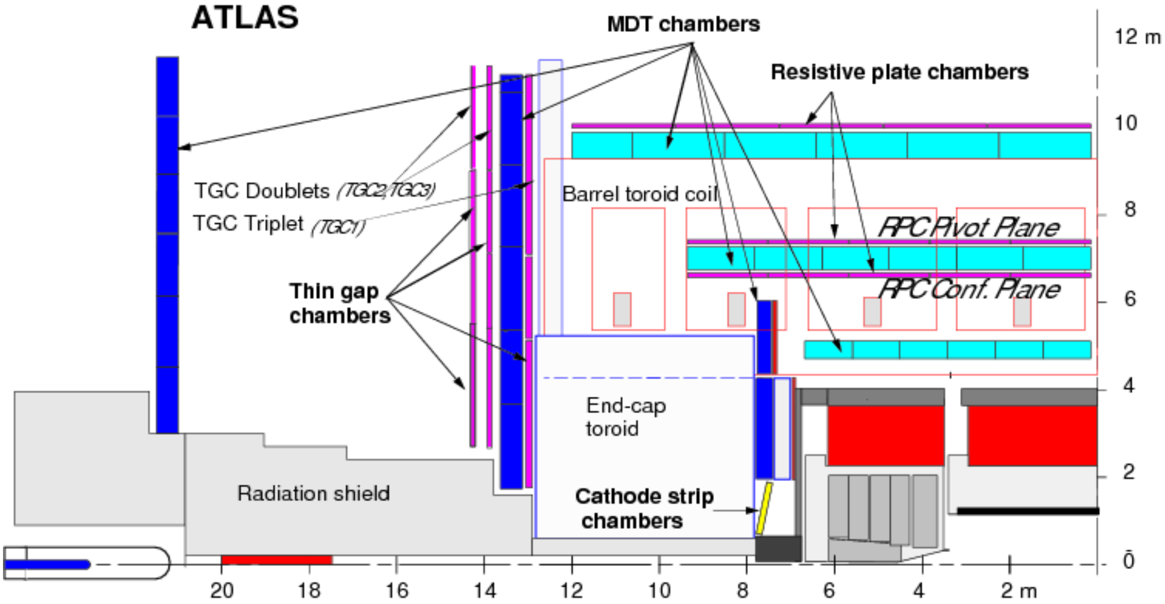


Figure 4.7: Schematic cross-sectional view of the sub-components of MS [90].

4.2.4 Trigger and data acquisition

The computational size of an average event recorded by the ATLAS detector is ≈ 1 MB. ATLAS has limited computational storage that allows for recording ≈ 1500 events per second. On the other hand, during Run II the LHC delivered approximately 40 million bunch crossings per second with an average 25 interactions per bunch crossing. As such, the ATLAS detector is only able to record approximately 1 event out of every 1 million events. A two step trigger system is therefore developed in order to decide which events are to be picked and registered by the detector:

- Level-1 (L1) [91]: This first trigger reduces the event rate to 100 kHz. The L1 trigger uses the input from the calorimeters and the MS. Information is provided in a coarse, low-precision format that can be quickly processed. This is used to look for interesting event topologies.

  Although what constitutes interesting depends on the target process, some general examples are: particles above certain threshold energies, events with high amounts of missing transverse energy, or leptons and jets with high p_{T} as well as geometrically isolated objects. Based on such criteria, L1 decides whether to accept or reject an event in about 2.5 μs; a speed enabled by the hardware-based architecture of the L1 trigger. Events selected by L1 are provided to the second trigger step as input.

- High Level Trigger (HLT) [91]: The software-based HLT is less burdened compared to L1 due to several orders of magnitude less number of events it needs to evaluate per second. Therefore it exploits more precise information from all detector units for events identified as relevant by L1. HLT accepts only 1% of the events selected by the L1 and has an output rate of approximately 1000 events per second. Events accepted by HLT are recorded by the ATLAS detector permanently in a file format called Event Summary Data (ESD).

ESD files are later reduced in size by keeping only physics information needed in analyses and transformed into the Analysis Object Data (xAOD) [92] file format. xAOD files are compatible with the ROOT Data Analysis Framework [93], the main computational framework used in ATLAS analyses. Data and MC simulations are both available in xAOD format and can be further reduced into smaller files by different analysis or study groups according to their interests, in a process known as derivation, resulting in Derived Analysis Object Data (DxAOD or DAOD) [94] files. Various filtering, customisation and object redefinitions are applied on DAODs by analysis teams to create a stable and practical to use dataset, known as *ntuples* [95].

4.3 Object reconstruction and identification in the ATLAS detector

4.3.1 Tracks

A pp collision in Run II produces around 600 tracks on average. ATLAS relies on precise localization capabilities of the ID and MS detectors for tracking. In this Section only track reconstruction in the ID is discussed. Tracks associated with MS are described in Section 4.3.4 together with the muons as they are used in the definition of different muon types in ATLAS. Tracks in the ID are reconstructed in four steps [96], where the output of each step is provided as input for the next one. These four steps are listed and summarised below:

1. Space-point formation: In the first step, hits close in space recorded by the pixel and SCT detectors are grouped into clusters of energy. Each of these clusters are transformed into a three dimensional coordinate point and called space-point.

2. Seeding and track candidates: Three space-points coming from three different layers define a seed. The curve defined by the bending of the seed is used to extend the seed by extrapolating it towards the interaction point, assuming a perfect helical bending dictated by the magnetic field. Some of the set of remaining space-points not associated with a seed might be consistent with its extrapolated trajectory. Such points are added to the seed by using a Kalman filter (KF), resulting in track candidates.

3. Ambiguity solving: Due to the high number of hits and the dense environment, one-to-one mapping between a seed and a track candidate is often not possible. When a seed is shared by multiple track candidates an ambiguity-solving algorithm is used to decide on the track assignment. A weight called the track score is used to aid the decision making process. It is composed of parameters representative of the quality of the track, such as the track momentum, resolution and multiplicity of the cluster hits.

4. TRT extension: The uniquely mapped set of tracks are fitted again using information provided by all three ID sub-detectors, including the TRT. A comparison between the old and the new tracks are made based on the quality of the fit. The higher quality track is retained.

Various tracks reconstructed by the ID are shown in Figure 4.8. Tracks are used in the reconstruction of all other objects discussed in this Chapter.

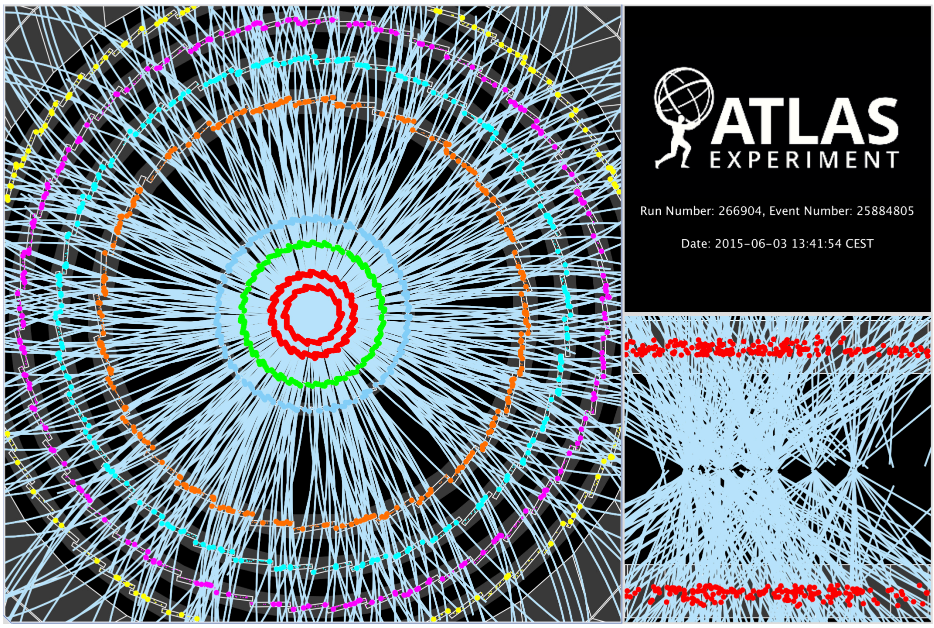


Figure 4.8: Event display of a proton–proton collision event as seen by the ATLAS detector. The large image on the left is the cross-sectional view of the detector with the origin corresponding to the beam axis. There are eight concentric circles each representing various detector components and hits recorded by them. The innermost circle is the IBL, the next three layers outwards are pixel detector layers, followed by four layers of SCT. Light blue lines represent tracks reconstructed by the ID. Circles/rectangles filled with colour are hits detected in a given region. Filled proper circles representing hits that are associated with a reconstructed track are larger in size compared to circles/rectangles that are not associated to tracks. The plot on the lower right shows the longitudinal section with the middle of the image corresponding to the beam line. Multiple nodes lying on the beam line that are created by set of converging tracks are representing the reconstructed vertices. The red dots represent the hits in the IBL similar to the left plot [97].

4.3.2 Vertices

A vertex is a point where two or more tracks intersect. Due to the limited resolution of the ATLAS detector, tracks originating from the same vertex might not exactly intersect. Vertex-finding algorithms [98] are used in ATLAS to decide whether a given set of tracks in the close vicinity of each other could be assigned to a common vertex or not. Nearby tracks are initially used to define a vertex seed, which in return creates a point of reference to assign a compatibility score to each track. The vertex is then re-defined using a χ^2 fit, taking the compatibility scores into account as track weights. This iterative process continues until the fit converges and certain stop conditions are met. A set of vertices reconstructed using ID tracks are shown in Figure 4.8. The most important type of vertex is the Primary Vertex (PV), which is the reconstruction of the interaction point of the hard-scattering process. Among the collection of reconstructed vertices, the one with the maximum $\Sigma_{\text{track}} p_{\text{T}}^2$, where the sum runs over tracks having a p_{T} above 400 MeV, is chosen as the PV [99]. There are other vertex types such as displaced vertices, pile-up vertices and conversion vertices. Two types of displaced vertices, namely Secondary Vertex and Tertiary Vertex are discussed in Section 4.3.6 within the context of jet flavour tagging. An example of conversion vertex in connection to photons are discussed in Chapter 6.

4.3.3 Electrons

Electron reconstruction [100] is achieved by combining tracking information from the ID with energy measurement from the electromagnetic calorimeter which absorbs electrons efficiently. Algorithms are used to determine which energy deposits in the calorimeters should be clustered together and associated with a specific electron. Up to recent past, including many analyses using a partial Run II dataset, ATLAS has used fixed-size clusters determined via a sliding-window algorithm [101]. In order to improve the electron reconstruction performance in full Run II analyses, a change has been introduced to switch to variable-sized clusters, known as *superclusters* [100]. A supercluster is constructed in several steps, starting with the topological cluster (topo-cluster) construction, which is a group of geometrically connected calorimeter cells. In simplified terms, whether two cells are considered to be connected or not is decided by an algorithm based on the EM cell significance $\zeta_{\mathrm{cell}}^{\mathrm{EM}}$ [100],

$$\zeta_{\mathrm{cell}}^{\mathrm{EM}} = \frac{E_{\mathrm{cell}}^{\mathrm{EM}}}{\sigma_{\mathrm{noise,cell}}^{\mathrm{EM}}}, \tag{4.8}$$

with $E_{\mathrm{cell}}^{\mathrm{EM}}$ being the energy of the EM cell and $\sigma_{\mathrm{noise,cell}}^{\mathrm{EM}}$ its expected noise. A topo-cluster candidate seed is formed by a cell if it fulfils $|\eta_{\mathrm{cell}}^{\mathrm{EM}}| \geq 4$. Neighbouring cells of the seed are then checked and merged if they satisfy $\zeta_{\mathrm{cell}}^{\mathrm{EM}} \geq 2$. When no bordering cell is left to satisfy the merging condition, the formed cluster is defined as a proto-cluster. For the second iteration, each cell with $\zeta_{\mathrm{cell}}^{\mathrm{EM}} \geq 2$ is taken as a seed and the process is repeated for each seed. Two proto-clusters are merged if they share a common cell satisfying the $\zeta_{\mathrm{cell}}^{\mathrm{EM}} \geq 2$ condition. Further conditions and details regarding when proto-clusters can be merged or split are described in Reference [100]. A topo-cluster is defined when no bordering cell is left that satisfies the merging condition.

The association of ID tracks to the topo-clusters is initially done by a loose matching process. Tracks that are matched to a topo-cluster are then subjected to another fitting process necessitated by the observed inability of the KF method used in the standard track reconstruction to account for electron bremsstrahlung properly.

The probability density of the electron bremsstrahlung process follows a Bethe-Heitler (BH) distribution. Although BH is an asymmetric distribution with a long tail, the standard KF method approximates it as a symmetric Gaussian distribution. Therefore, a generalisation of the Kalman filter called the Gaussian Sum Filter (GSF) [102] is used instead, that recovers the bremsstrahlung effect. The GSF captures the features of BH better by combining several Gaussian distributions.

If multiple tracks are matched to an energy cluster, the track with the minimum ΔR distance to the cluster is selected. The set of matched clusters and re-fitted tracks are used as input to further define superclusters. They are built from the combination of seed clusters and their associated satellite clusters. Among the topo-clusters with matched tracks having at least 4 SCT hits, therefore satisfying $E_{\mathrm{T}} > 1$ GeV are declared cluster seeds. For each seed cluster, neighbouring clusters that are found within a fixed window from its barycentre are defined as its associated satellite clusters. The combination of a seed cluster with its satellite clusters complete the construction of superclusters. An ambiguity resolver is applied to create disjoint collections of electrons and photons. Details of the supercluster building process are described in Reference [100].

Identification

Oftentimes, detector signatures from photon conversions or jets are falsely reconstructed as electron candidates. A likelihood-based multivariate discriminant is used to suppress these mis-reconstructed objects and increase the correct identification of the prompt electrons [103]. The discriminator utilizes 20 input variables from ID and EM clusters such as shower shape, track quality and track matching information. Four benchmark working points (WP) are defined by applying cuts on the LH discriminant output representing different efficiency and background rejection fractions dependent on the E_{T} and η. These are named VeryLoose, LooseLH, MediumLH and TightLH. In the $t\bar{t}t\bar{t}$ analysis only the latter two WPs are used. At a given benchmark transverse energy of 40 GeV, MediumLH (TightLH) has an identification efficiency of 88% (80%). The corresponding background efficiencies are reported to be similar to the ones estimated during Run I [104]: 0.51% (0.29%) at a transverse energy range of 20-50 GeV.

Charge identification

In the ATLAS detector, the charge of an electron can be identified wrongly for reasons such as a mis-identified track curvature direction. The Electron Charge Identification Selector (ECIDS) is a tool which is a multivariate discriminant trained on data with a selection of $Z \to ee$ events using the Boosted Decision Tree (BDT) method. The tool is able to determine the correct charge of the selected electron in about 98% of the cases and to reject the electrons with mis-identified charges in around 90% of the cases. Unlike standard identification, ECIDS is not used by default in many ATLAS analyses. However, it is used in the $t\bar{t}t\bar{t}$ analysis because the final state selection is sensitive to the lepton charge information as two same-charged leptons are required.

4.3.4 Muons

In detecting muons, tracks become more important as low energy deposition renders calorimetry less useful. Tracking information provided by the ID and MS sectors of the ATLAS detector is used to reconstruct and identify the muons. For ID information, similar to the case with electrons, no new tracking methods or criteria are applied and tracks described in Section 4.3.1 are used. The track reconstruction in MS begins with finding MDT segments, defined as the straight line fit between the hits in the four different layers of the MS. Tracks are then reconstructed by performing a fit to segments from different layers, taking into account the bending due to the magnetic field and the different material composition of the layers. Different ways of combining ID and MS track informations lead to four type of muon definitions [105]:

- Combined (CB) Muon: These muons are reconstructed from a new fit made using the combined ID and MS tracks, where input ID and MS tracks are independently reconstructed in a previous step.
- Extrapolated (ME) Muon: Extrapolated muons are reconstructed from MS tracks that can be extrapolated to the IP. No ID information is used for these type of muons therefore the pseudo-rapidity acceptance is not limited to $|\eta| < 2.5$, which is dictated by the ID geometry.

- Segment Tagged (ST) Muon: This type of muons use segment information from the MS and fits it together with ID tracks. ST muon candidates make reconstruction possible for the cases when only one segment is found.
- Calorimeter Tagged (CT) Muon: If the extrapolation from the ID track of the muon is matched with a calorimeter deposit consistent with that of a minimum-ionizing particle, it is defined as a CT muon. CT muons are used in the gap region of the MS ($|\eta| < 0.1$), where this detector unit can not provide information.

When the same ID tracks are used to reconstruct multiple types of muons, a quality-based hierarchical selection is used for overlap removal: CB first, ST second and CT last. ME muons do not have ID tracks but can overlap with other muons' MS tracks. In that case, the muon candidate with the better fit performance is retained and the other one is discarded.

Identification

Muon identification [105] mainly aims at increasing the prompt muon rate by suppressing the background muon contaminations originating from decays of hadrons. This is achieved by applying selection cuts on discriminating variables such as the relative difference of transverse momenta, as well as the charge-to-momentum ratio significance measured in the ID and MS, or the χ^2 value of the combined ID and MS track fit. Further quality criteria are applied to the number of hits in the ID and MS, sectors to increase the accuracy of the momentum measurement. Various combinations of discrimination and quality criteria are grouped into five WPs: Loose (all), Medium (CB and ME only), Tight (CB and ME only), low-p_{T} ($p_{\mathrm{T}} < 5$ GeV) and medium-p_{T} ($p_{\mathrm{T}} > 100$ GeV) [105]. The medium identification WP is used in the $t\bar{t}t\bar{t}$ analysis. Based on the measurements made in data by selecting muons decaying from Z-boson or J/ψ events, an efficiency of 98% is estimated in the range $0.1 < |\eta| < 2.5$. The charge mis-identification rate is very small for muons and therefore a separate tool is not considered in this analysis.

4.3.5 Lepton isolation

The collection of reconstructed and identified leptons can still be further separated from the non-prompt background contamination. The usefulness of isolation relies on the observation that prompt leptons produce more collimated and dense detector signals whereas their non-prompt backgrounds have scattered and sparse signals.

- Electrons: In the case of electrons, isolated prompt candidates are defined as the electrons decaying from heavy resonances (typically W- and Z-bosons). The isolation variables are designed such that they distinguish these electrons from other poorly isolated candidates that could arise from mis-identification of light hadrons, decays from heavy-flavour hadrons or conversion of photons emerging from hadron decays, into electron-positron pairs [103].
- Muons: Isolated prompt muon candidates decay from heavy particles, similarly to electrons. The main target of muon isolation is to reject the muons that originate from jets and occur via semileptonic decays. Being in the vicinity of jets implies that these muons will have hadronic activity around them and thus will have less isolation compared to their prompt counterparts.

As such, selecting isolated leptons amounts to rejecting background events. Two variables are used to determine whether leptons are considered isolated or not, with the first one being based on information from the calorimeters and the second one using the input from trackers:

- Calorimeter-based isolation variable $E_{\mathrm{T}}^{\mathrm{topocone}R}$: This variable gives the total transverse energy deposited within a cone of radius $r = R$ of a cluster assigned to a lepton candidate. Typical cone radius values are 0.2 or 0.4.

- Track-based isolation variable $p_{\mathrm{T}}^{\mathrm{varcone}R}$: Here, instead of the total transverse energy, the total transverse momentum is used. The cone radius r is dynamically parametrised as a function of p_{T}:
$$r = \min\left(R, \frac{10\ \mathrm{GeV}}{p_{\mathrm{T}}}\right) \tag{4.9}$$

Using selections on isolation variables, various isolation working points with different efficiencies are defined for each type of lepton in ATLAS. Among many possible WPs, only the ones relevant for the $t\bar{t}t\bar{t}$ analysis are described. In the $t\bar{t}t\bar{t}$ analysis, the *FixedCutTight* WP is used for electrons which is defined by requiring the calorimeter-based isolation variable with $R = 0.2$ to satisfy the condition $E_{\mathrm{T}}^{\mathrm{topocone20}} < 0.06\ p_{\mathrm{T}}$ and the track-based isolation variable with R = 0.2 to fulfil the condition $E_{\mathrm{T}}^{\mathrm{topocone20}} < 0.06\ p_{\mathrm{T}}$. This WP has an approximate efficiency of 95% at a given transverse energy of 40 GeV [103]. In the case of muons, the *FixedCutTightTrackOnly* isolation WP is used. This WP is defined by only requiring the same track-based isolation condition used for the electrons, and has an efficiency of 94% measured with simulated $t\bar{t}$ events using tracks with a p_{T} range of 20–100 GeV [105]. The properties of the two lepton isolation WPs are summarised in Table 4.2.

Lepton	Isolation WP	Efficiency	Calorimeter Isolation	Track Isolation
Electron	FixedCutTight	≈ 95%	$\frac{E_{\mathrm{T}}^{\mathrm{topocone20}}}{p_{\mathrm{T}}} < 0.06$	$\frac{p_{\mathrm{T}}^{\mathrm{varcone20}}}{p_{\mathrm{T}}} < 0.06$
Muon	FixedCutTightTrackOnly	≈ 94%	-	$\frac{p_{\mathrm{T}}^{\mathrm{varcone30}}}{p_{\mathrm{T}}} < 0.06$

Table 4.2: Isolation parameters for lepton WPs used in the $t\bar{t}t\bar{t}$ analysis. Values are taken from Reference [103] for electrons and from Reference [105] for muons.

4.3.6 Jets

Jets are hadronic objects that result from the fragmentation and hadronisation of the quarks or gluons and are color-neutral. A jet is an ensemble of collimated particles around the flight direction of the original parton. These ensembles will form various topological clusters that manifest itself as deposits in the nearby calorimeter cells.

Jet reconstruction algorithms are used to define the boundaries of these calorimeter clusters, therefore deciding which particles originate from the same parton. These algorithms are also called cone algorithms as the process can be visualized as cones around a bunch of particles with the cone radius determining the boundary.

An important feature of a jet cone is its *safety*. Connected to theoretical calculations of perturbative QCD, safety in this context refers to the clustering algorithms' outcome being not altered when a narrow or soft gluon emission is introduced. These properties are known as *collinear safety* and *infrared safety*, respectively [106].

The common mechanism behind the jet finding algorithms discussed here is the iterative recombination of pair of particles until none are left or some stopping criteria are reached.
The merging of a pair of particles i and j takes place according to a criterion based on their proximity, which is defined by two distances

$$d_{i,B} = p_{\mathrm{T,i}}^{2\kappa} \tag{4.10}$$

and

$$d_{ij} = \min(p_{\mathrm{T,i}}^{2\kappa}, p_{\mathrm{T,j}}^{2\kappa})\frac{\Delta R_{ij}^2}{R^2}, \tag{4.11}$$

where ΔR_{ij} is

$$\Delta R_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}, \tag{4.12}$$

with y being the rapidity. The variable κ is an exponent that regulates the priority given to clustering in relation to p_{T} of the particle. Three different κ values lead to three different clustering algorithms [107]: The k_t algorithm for $\kappa = 1$, the anti-k_t algorithm for $\kappa = -1$ and the Cambridge/Aachen algorithm for $\kappa = 0$. The anti-k_t algorithm is the main algorithm used by the ATLAS collaboration, fulling the QCD safety conditions, and prioritizing the clustering of higher p_{T} jets into approximately conical shaped jets. The algorithm merges two objects i and j if $d_{ij} < d_{i,B}$, adds the merged object as a new particle and removes i and j from the list of particles. The process is repeated until no further merging is possible. An example calorimeter clustering output given by the anti-k_t algorithm with $R = 1$ is shown in Figure 4.9.

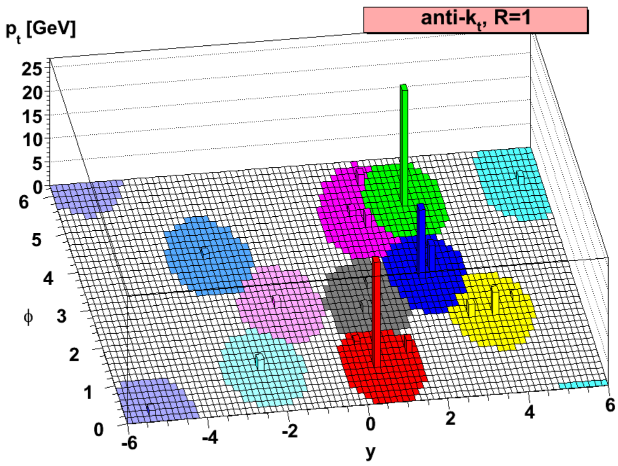


Figure 4.9: An example calorimeter clustering output given by the anti-k_t algorithm with $R = 1$. The two dimensional grid defines the detector plane (ϕ, y). The vertical axis gives the p_{T} measured through the energy deposited in each square. The near-circular coloured shapes show the clusters defined by the algorithm. Image taken from Reference [107].

Flavour tagging

For jets emerging from quarks, the flavour information of the original quark could be retrieved with some efficiency, through a method called *tagging*. In top-quark physics, b-tagging is very important since top quarks decay almost always into a b-quark and a W-boson. b-quarks are the heaviest quarks that can hadronise, forming B-hadrons, therefore their decay time is relatively short compared to lighter hadrons formed by other quarks. This results in B-hadrons decaying at a secondary vertex (SV), on average few millimetres away from the PV in the laboratory frame, before entering into the detector volume [108]. b-taggers using SV information are called SV-based taggers. A related b-tagging algorithm is JetFitter [108], which relies on the assumption that a b-hadron formed at the PV will decay to a c-hadron at the SV, which later decays further at a tertiary vertex, and that the three vertices will lie on a common flight path.

Another important variable is the Impact Parameter (IP) which is defined as the closest distance (d_0) of a track to the PV. Tracks from b-tagged jets are more likely to have larger IPs as they are more likely to emerge from SVs, whereas tracks originating from the PV will have a d_0 compatible with 0. The relationship between PV, SV and d_0 with the tracks within a b-jet is depicted in Figure 4.10. In

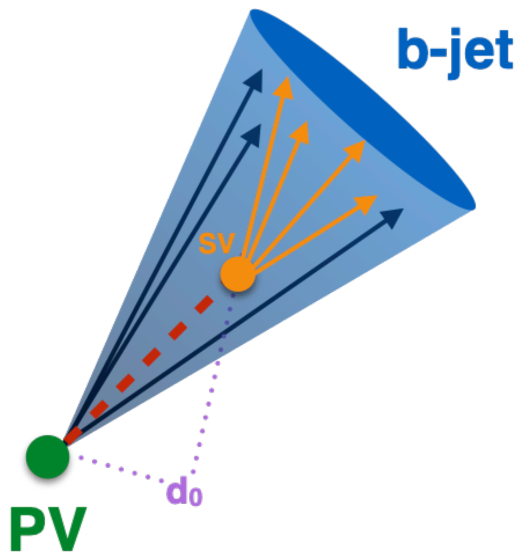


Figure 4.10: Schematic drawing of a b-jet defined by the blue cone. Black (orange) arrows represent the tracks originating from the PV (SV). The dashed red line is the flight length of the B-hadron and the dotted purple lines show the impact parameter (d_0).

ATLAS, the family of b-taggers using IP information are called IP-based taggers. SV- and IP- based taggers both use a likelihood discriminator. A better b-tagging performance is achieved by using the boosted-decision-tree (BDT)-based MV2c10 tagger [108]. It has 24 input variables, mostly similar to the ones used by the SV- and IP-based taggers, but also includes outputs of these low-level taggers. The MV2c10 tagger provides a BDT discriminator score as output which results from a training on $t\bar{t}$ samples using b-jets as the signal and light-flavour jets as the background. The fraction of c-jets in the background composition of the training is separated since c-jets have characteristic properties closer to b-jets compared to the rest of the light jets. Thus, c-jets are more likely to be confused by the algorithm with the b-jets, impacting the performance. For the MV2c10 algorithm, the training background comprises 7% c-jets and 93% light-jets.

The continuous BDT score output of MV2c10 is calibrated at 4 working points (WP) at the b-tagging efficiencies of 60%, 70%, 77% and 85%. The characteristic efficiency and selection values for the WPs are listed in Table 4.3. The arguments given above for b-hadrons are to a lesser extend valid for c-quarks as their characteristic lie between the light jets and the b-quarks.

WP	BDT Score	Jet Rejection Rate		
ε_b	Selection	ε_c^{-1}	ε_τ^{-1}	$\varepsilon_{\text{light}}^{-1}$
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

Table 4.3: Benchmark values for the four working points (WP) available for the MV2c10 b-tagging algorithm. The b-tagging efficiency (ε_b) with their corresponding BDT discriminant selection cut and the rejection fractions (ε^{-1}) for charm (c), tau (τ) and light flavour jets are reported. Table adapted from Reference [109].

4.3.7 Neutrinos and missing transverse energy

Unlike the rest of the particles described above, neutrinos are not reconstructed in the ATLAS detector. As discussed in Section 2.1 neutrinos are uncharged and only interact with matter weakly. Therefore they neither create tracks nor deposit energy into the calorimeters that can be registered by the ATLAS detector. Neutrinos leaving the detector volume undetected carry away some momentum. Conservation of momentum dictates that total momentum before and after the collision needs to be conserved. In the laboratory frame, the total transverse momentum in the detector volume before the collision is zero. Assuming no BSM contributions and that the ATLAS detector registers all the particles except the neutrinos, the total missing transverse momentum after the collision is attributed to escaping neutrinos. The missing transverse energy $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ is defined as [110]

$$\vec{E}_{\mathrm{T}}^{\mathrm{miss}} = \Sigma \vec{p}_{\mathrm{T}}^{\,\mathrm{i}} + \Sigma \vec{p}_{\mathrm{T}}^{\,\mathrm{soft}}, \tag{4.13}$$

where the first term is called the hard term and is the vectorial sum of the calibrated p_{T} of all i reconstructed and selected objects in an analysis. The second term is called the soft term and is the sum of the transverse momentum of all tracks that are not used in the reconstruction of any objects but associated with the PV. Two derived quantities commonly used are the scalar

$$E_{\mathrm{T}}^{\mathrm{miss}} = \sqrt{(E_{\mathrm{x}}^{\mathrm{miss}})^2 + (E_{\mathrm{y}}^{\mathrm{miss}})^2}, \tag{4.14}$$

corresponding to the magnitude of $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ and the azimuthal angle

$$\phi^{\mathrm{miss}} = \arctan\left(\frac{E_{\mathrm{y}}^{\mathrm{miss}}}{E_{\mathrm{x}}^{\mathrm{miss}}}\right), \tag{4.15}$$

on the transverse plane. The $E_{\mathrm{x,y}}^{\mathrm{miss}}$ terms are the sums of momenta on these specified axes.

4.3.8 Energy and efficiency corrections

The accuracy of the energy determination of physics objects often deteriorate for various reasons such as large shower profile sizes not contained by calorimeter volumes, limitations on detector geometry simulation, energy losses in non-instrumented regions of the detector, or algorithm designs not using the full energy information. In electron reconstruction, multiple energy calibrations are applied during intermediate steps and the details are given in Reference [100]. For muons, an important process is the correction of the momentum measurement because the performance of the detector simulation in modelling muon energy losses are below the observed level.

Jets are subjected to a complex correction scheme in ATLAS. The purpose of the jet energy scale (JES) calibration is to recover the original energy of the first stable reconstructed jet object defined by the MC simulation [111]. Before energy scale corrections, the direction of a jet is recalculated so that it is compatible with the PV. After the recalculation of the jet direction, additional energy attributed to jets due to pile-up events are removed as the first correction step. MC-simulation-based energy corrections are applied next, in order to determine the jet energy at particle level and address the energy loss through leakage in the calorimeter.

A separate correction known as *in situ* is applied only to data events with reconstructed jets. It accounts for the differences in the jet reconstruction between the MC simulation and data. This difference is quantified by comparing the discrepancy between the p_{T} of jets, using processes that are known to be measured with a high precision in data, such as the ones involving photons and Z-bosons [111].

Efficiencies also differ when applied to simulation (ϵ_{MC}) and data (ϵ_{Data}). This leads to a discrepancy, manifesting itself as an apparent mis-modelling between data and simulation, where a good closure is expected otherwise. In order to correct for this effect, scale-factors SF

$$\mathrm{SF} = \frac{\epsilon_{\mathrm{Data}}}{\epsilon_{\mathrm{MC}}} \tag{4.16}$$

are defined as the ratio of two efficiencies. Various SFs for efficiencies are calculated for selecting, reconstructing, identifying and isolating physics objects. All relevant SFs are multiplied with each other and applied as a combined weight factor to each simulated event in an analysis.

CHAPTER **5**

Statistical methods and multivariate techniques

In particle physics analyses at the LHC, quantitative results are extracted by constructing a statistical model, which is then used to estimate certain parameters from data, a process known as parameter estimation or “fitting”. The interpretation of the fit results are based on the statistical methods as well as commonly agreed standards of the HEP community. The development of an analysis can be seen as an attempt at a statistical optimization of the constructed model, where a sought-after signal (S) process is to be separated from established background (B) processes to the best extent possible, taking systematic uncertainties into account. Through optimisation, the precision of a measurement on a Parameter of Interest (PoI) can be improved. What actually constitutes to PoI, signal, background and uncertainties is determined by the physics goals and experimental setup. In the $t\bar{t}t\bar{t}$ analysis, the goal is to measure the inclusive cross-section of the $t\bar{t}t\bar{t}$ process. In this setting the PoI is the signal strength μ, defined as the ratio of the measured cross-section divided by the cross-section predicted by the SM

$$\mu = \frac{\sigma}{\sigma_{\mathrm{SM}}} \,. \tag{5.1}$$

While $t\bar{t}t\bar{t}$ is the signal process being searched, the leading background processes are $t\bar{t}H$, $t\bar{t}Z$ and $t\bar{t}W$ as described in 3, and the leading uncertainty is given by the modelling of the $t\bar{t}W$ background process. Further background processes and details of uncertainties are discussed in 6. Some of the parameters of these background processes and uncertainties can be estimated by the fit along with the PoI. For example the amount of a poorly known or modelled background process may be scaled by the fit, resulting in a Normalization Factor (NF). Parameters of the model other than PoI are named Nuisance Parameters (NPs), and represented by $\boldsymbol{\theta} = \theta_1, \theta_2, ..\theta_n$. Each systematic uncertainty is a NP, that is required in the model, however they are not the actual measurements targeted by the analysis. Depending on the designation of the model, NPs can be estimated either in the same fit, or in an auxiliary measurement.

An important step in the statistical optimization is the definition of *region*s. A region is a phase-space defined by certain selection criteria applied to events that are decided based on a purpose, and can be classified into two:

- Signal Region (SR): This region is designed to maximize the signal events and minimize the background events passing the selection criteria. SRs are where the analysis sensitivity mostly come from as they try to enhance the signal. Multiple SRs can be defined in a given analysis if e.g. different channels or phase-space regions have to be separated.

- Control Regions (CR): A Control Region is where ideally the fraction of one of the background processes is maximized, with signal and all other background processes minimized. A CR helps to constrain the uncertainty of the estimation of the process or of the auxiliary measurement under question by comparing the simulation to the data.

The definition of regions can be "cut-based" or defined by multivariate analysis (MVA) techniques. Both methods rely on the use of discriminating variables that may be selected by physics reasoning or experimental tests such as trial-and-error. In a cut-based method, the range of various variables may be set to define the phase-space of a region. In the MVA method, typically a machine-learning algorithm is used to combine multiple variables to provide a single output discriminant, which is then used to define the region. An MVA discriminant can also be combined with other cut-based selections to define regions.

The first Section of this Chapter is largely based on Reference [112] and describes the construction of the statistical fit model, by incorporating above-mentioned concepts into the profile-likelihood method used in the $t\bar{t}t\bar{t}$ analysis. The second Section is dedicated to a discussion of MVA methods. The scope of the discussion presented in this Chapter is limited to methods that have been used in the $t\bar{t}t\bar{t}$ analysis. Similarly, in Section 5.2, only the supervised learning method of Boosted Decision Trees is described and many other machine-learning methods (such as Neural Networks) or unsupervised methods are not discussed.

5.1 Statistical methods

In a measurement, where the distribution of the measured variable x is expressed by a binned histogram, each bin is populated by various fractions of the signal and background events both of which depend on systematic uncertainties that are parameterized by NPs $\boldsymbol{\theta}$. The expectation value $[n_i]$ for the i^{th} bin is computed by

$$[n_i] = \mu s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta}), \tag{5.2}$$

where μ is the signal strength, s_i and b_i are the expectation values of the number of signal and background events in the bin i, respectively. Assuming another histogram designed as a CR and therefore dominated by background events, the signal contribution can be ignored and the expectation value $[m_j]$ for the j^{th} bin is computed by

$$[m_j] = u_j(\boldsymbol{\theta}), \tag{5.3}$$

where u_j are the expectation values of background events in the bin j.

For an event, the probability distribution function (pdf) $f(x|\mu, \boldsymbol{\theta})$ quantifies the probability density of measuring observable x, given the condition defined by the set of parameters $(\mu, \boldsymbol{\theta})$. The same mathematical expression may also be used to define the likelihood function $L(\boldsymbol{x}|\mu; \boldsymbol{\theta})$, the joint probability distribution for a set of i events $\boldsymbol{x} = x_i$. The likelihood is a tool for *inference*, in contrast to $f(x|\mu, \boldsymbol{\theta})$, where one computes the probability density of getting a measurement given a condition

$(\mu, \boldsymbol{\theta})$. The likelihood function gives the probability density that $(\mu, \boldsymbol{\theta})$ could lead to a measurement $\boldsymbol{x}$. The likelihood construction from measured data is used to check how likely it is that physics model tested under the given experimental setup in consideration could have led to the measured values. According to the *likelihood principle*, all the necessary information of a measurement is contained in the likelihood function.

Given a simplified model of a typical measurement with one SR, one CR and a set of NPs, the likelihood function is constructed as

$$L(\mu, \boldsymbol{\theta}) = \prod_{i=0}^{N} \frac{(\mu s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta}))^{n_i}}{n_i!} e^{-(\mu s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta}))} \times \prod_{j=0}^{M} \frac{u_j(\boldsymbol{\theta})^{m_j}}{m_j!} e^{-u_j(\boldsymbol{\theta})} \times \prod_{k=1}^{P} \rho(\theta_k), \tag{5.4}$$

where in the first two terms corresponding to SR and CR, a Poisson pdf

$$f(n; \lambda) = \frac{\lambda^n}{n!} e^{-\lambda} \tag{5.5}$$

is used for each bin and the contributions from all bins are accounted for by multiplication. The third multiplicative term in Equation 5.4 incorporates NPs into the likelihood and is described in the next Section. N, M and P represent the number of bins.

As it contains the full information of the given model and data, the likelihood function is used to estimate the desired parameters in the model. Under the assumption that the parameter distributions used to construct the likelihood function are correct, the maximum likelihood gives the best achievable estimate of parameters. Therefore, the maximum likelihood estimate (MLE) $(\hat{\mu}, \hat{\boldsymbol{\theta}})$ of model parameters will result in the global maximum of the likelihood function.

5.1.1 Incorporation of systematic uncertainties

The last term in Equation 5.4 accounts for the pdfs of the NPs, assuming there are P of them and that they are uncorrelated. Represented by the NPs, the inclusion of systematic uncertainties enables the model to account for the impact of these parameters reflecting the conditions under which the collected data is obtained and the MC simulations are produced. Pdfs for NPs are given by the application of Bayes' theorem,

$$\rho(\theta|\hat{\theta}) \simeq \rho(\hat{\theta}|\theta)\pi(\theta) \tag{5.6}$$

where $\rho(\theta|\hat{\theta})$ is the posterior pdf, $\pi(\theta)$ is the prior probability density of θ, and $\hat{\theta}$ an auxiliary measurement. The prior $\rho(\hat{\theta}|\theta)$ typically takes three functional forms corresponding to three use cases:

- Log-normal distribution for normalisation uncertainties: NPs related to normalisation uncertainties cannot be negative and thus a log-normal distribution is a suitable choice as it is positive-definite only. A pdf distribution of the form

$$\rho(\theta|\hat{\theta}) = \frac{1}{\sqrt{2\pi}\theta \ln(\sigma)} e^{-\frac{\ln^2(\theta/\hat{\theta})}{2\ln^2(\sigma)}} \tag{5.7}$$

is used in this case, where σ is the spread around the central value $\hat{\theta}$, derived independently from an auxiliary measurement. The log-normal distribution describes the variables that result from the multiplication of random factors in the limit where the number of these factors go to infinity. This is similar to the Central Limit Theorem, where, instead of multiplication, the sum of many random factors can be described by a Gaussian in the limit of large numbers.

- Gaussian distribution for shape uncertainties: When the NPs affecting the shape of distributions are considered, pdfs for systematic uncertainties are built from the Gaussian distribution

$$\rho(\theta|\hat{\theta}) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\theta-\hat{\theta})^2}{2\sigma^2}} . \tag{5.8}$$

  This allows for both, negative and positive variations, and per property of the Gaussian function, variations can be interpreted in terms of standard deviations (σ) around the central value, conventionally quoting $\pm 1\sigma$. Positive (negative) variations are often called *up* (*down*) variations.

- Gamma distribution for statistical uncertainties: Statistical and systematic uncertainties are normally two disjoint sets of parameters. Statistical uncertainties are accounted for in the likelihood through Poisson distributions. This Poissonian statistical uncertainty may not be very accurate if the event yield used in its computation is a reweighted outcome of an actually small number of raw events. To account for the large uncertainty in this reweighted yield due to the low number of original unweighted events, one NP for each histogram bin is introduced. These NPs have pdfs built from the Gamma distribution.

Often there are hundreds of NPs considered in a fit model, requiring significant computation time and resources as well as causing concern for the stability of the output with respect to statistical fluctuations. In order to address these issues, various preprocessing methods have been developed. Three of those have been used in the $t\bar{t}t\bar{t}$ analysis are described below:

- Pruning: A term coined in analogy to the pruning of trees, this method removes the uncertainties that fail criteria reducing the processing time of the full fit. The removal criteria differ for normalization and shape uncertainties. For the normalization uncertainties, if the difference between the total yield of a nominal distribution and the total yield of its systematic variation is below 0.5%, that NP can usually be dropped. For the shape systematics the same 0.5% difference is checked per bin, and if for all the bins the difference is below 0.5%, that NP can be removed. The removal of too many uncertainties with a large cumulative impact could lead to underestimated uncertainties, as the statistical model at hand will no longer correspond to the actual conditions. On the other hand, if too few NPs are pruned, it will be harder to evaluate the model and thus convergence of the fit will take considerably longer.

- Smoothing: This method aims at reducing the statistical fluctuations due to the low number of events in the alternative MC samples that are used to define the systematic variations in order to achieve a more stable fit model. In the case of the $t\bar{t}t\bar{t}$ analysis presented in this thesis, an iterative rebinning method is used. Neighbouring bins in a distribution are merged until each bin has a statistical uncertainty of less than 8%. If the number of local extrema in the outcome is less than 5, the rebinning is repeated from the beginning with half of the uncertainty used in the previous iteration. The obtained rebinned distribution is then processed with the *running* *medians* algorithm (implemented in ROOT [93] with the 353QH algorithm [113]) against the artificial flattening of the distributions. Similarly to the pruning, the amount of smoothing needs to be set to a moderate value. Too little smoothing might not create the intended effect whereas too much smoothing might lead to an alteration of the shape of the distribution causing loss of information.

- Symmetrization: Symmetry in this context refers to the relative size of the up and down systematic variations around the nominal value. In some cases, certain systematic uncertainties can have large asymmetries between the up and down variations. Symmetrization is the process of making the magnitude of two variations similar, except for the cases where there is a physics reason for an asymmetry to exist. This is considered important because NPs with a high asymmetry might get under-constrained in a fit. Systematic variations can be one-sided (for example some resolution effects that can only get worse) or two-sided depending on whether only one or both variations are given for the NP under consideration. In the case of one-sided variation, the opposite variation is assumed by taking the magnitude of the given variation and replacing its direction i.e. sign. Therefore a mirroring is applied. If for a positive definite NP (such as cross-section uncertainties) the one-sided variation is an up variation with a value above 100%, its mirrored, downward variation will be negative. In that case the downward variation will be set to 100% of the central value as for a positive definite NP, values cannot be below zero. For the cases where both up, (σ^{up}) and down (σ^{down}) variations are known, the common, symmetrised variation is given by

$$\sigma_{\text{sym}} = \frac{|\Delta^{\text{up}}| + |\Delta^{\text{down}}|}{2}, \tag{5.9}$$

where $\Delta^{\text{up/down}} = N_{\text{up/down}} - N_{\text{nominal}}$ and N_{x} is the yield for the given case.

5.1.2 Hypothesis testing and test statistics

A search for a new phenomenon is in essence the testing of an hypothesis, which is a statement of a prediction for the given natural phenomenon under certain conditions, that can experimentally be tested. If a hypothesis fails to explain the experimental results, it is rejected. Hypotheses can not be proven, they can only be rejected. Depending on the quantitative strength of the rejection, the result can be translated into the scientific statements like evidence, discovery or exclusion. Such qualitative statements about the outcome of an experiment can be reached through statistical inference. This Section describes the formulation of hypothesis testing in statistical terms and the resulting metric, the test statistic, that incorporates the full information of the hypothesis test. The following Section describes how this metric may be transformed into qualitative statements about the outcome of an experiment.

The formulation begins by defining two hypotheses:

- Null Hypothesis (H_0): This hypothesis defines the state with the established conditions and existing knowledge where the hypothesised effect under investigation is not present. What corresponds to the null hypothesis depends on the goal of the experiment. In the common case of a search for a new signal process (such as a new particle or process) H_0 is called "background-only hypothesis" as it represents all the established processes and particles in the model, excluding the sought-after signal.

- Alternative Hypothesis (H_1): In contrast to the null hypothesis, the alternative hypothesis includes the new effect being investigated and therefore corresponds to an expanded version of the currently established knowledge. In the example case, where a new signal is being searched, H_1 is called "Signal+Background (S+B) hypothesis".

According to the Neyman-Pearson lemma [114], the most powerful way to compare two competing hypotheses for a value of μ can be achieved by a test using the likelihood ratio (LR) $\lambda(\mu)$,

$$\lambda(\mu) = \frac{L(\mu, \theta_0)}{L(\mu, \theta_1)}, \tag{5.10}$$

where θ_0 and θ_1 are the associated sets of NPs for a given hypotheses H_0 and H_1 respectively, μ is the PoI and the numerator (denominator) is the likelihood function for the null (alternative) hypothesis [112].The parameters $\lambda(\mu)$ has values between 0 and 1, with smaller values pointing to an incompatibility with the null hypothesis. If a model consists of many NPs but only one PoI, the LR can be turned into a profile-likelihood ratio,

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\theta}})}{L(\hat{\mu}, \hat{\theta})}, \tag{5.11}$$

where $\hat{\hat{\theta}}$ in the numerator are the MLE values of θ for a fixed μ, and both μ and θ in the denominator are MLE values giving the global maximum of the likelihood function. If $\hat{\mu} < 0$, it will be set to 0.

The test statistic t_μ is defined as,

$$t_\mu = -2\ln\lambda(\mu) \tag{5.12}$$

and is used to quantify the outcome of the hypothesis tests.

5.1.3 Statistical interpretation

The level of discrepancy between the data and the hypothesis under consideration for a parameter μ is quantified by computing the p-value

$$p_\mu = \int_{t_\mu^{\text{obs}}}^{\infty} f(t_\mu|\mu)\, dt_\mu,$$

where t_μ^{obs} is the observed test statistic value calculated from data and $f(t_\mu|\mu)$ is the pdf of t_μ for a given value of μ. This pdf distribution may be constructed by repeating a pseudo-experiment of hypothesis testing many times and calculating the test statistic. However, achieving reasonable accuracies requires running large number of such pseudo-experiments, rendering this method a computational challenge. Instead, an approximation is used, namely the asymptotic approximation, based on Wald's approximation [115] that is valid in the limit of large numbers.

The p-value for a given hypothesis provides a measure how unlikely it is for the measured data to result from that hypothesis. As such, the lower the p-value the lower is the support for the hypothesis under consideration. For ease of interpretation, it is customary to transform and re-express the p-value in terms of the standard deviations from the mean of a variable following a Gaussian distribution. In this case, the point on the Gaussian distribution whose upper-tail probability is equal to the p-value is determined, and the significance Z is the distance in units of standard deviations, between this point and the mean of the distribution. Z is mathematically determined from the p-value through the inverse of the cumulative distribution of the Gaussian function Φ^{-1}, using the equation

$$Z = \Phi^{-1}(1 - p). \tag{5.14}$$

According to commonly agreed criteria of the scientific community an evidence (discovery) is claimed with a $Z \geq 3$ (5) σ measurement sensitivity corresponding to a p-value of 1.349×10^{-3} (2.87×10^{-7}). If neither of the thresholds is reached, an upper limit can be set to exclude the alternative hypothesis typically using $p = 0.05$ corresponding to a 95% confidence level and a Z of 1.64σ. The calculation of the p-value and its corresponding Z value from the pdf function and the test statistic is depicted in Figure 5.1.

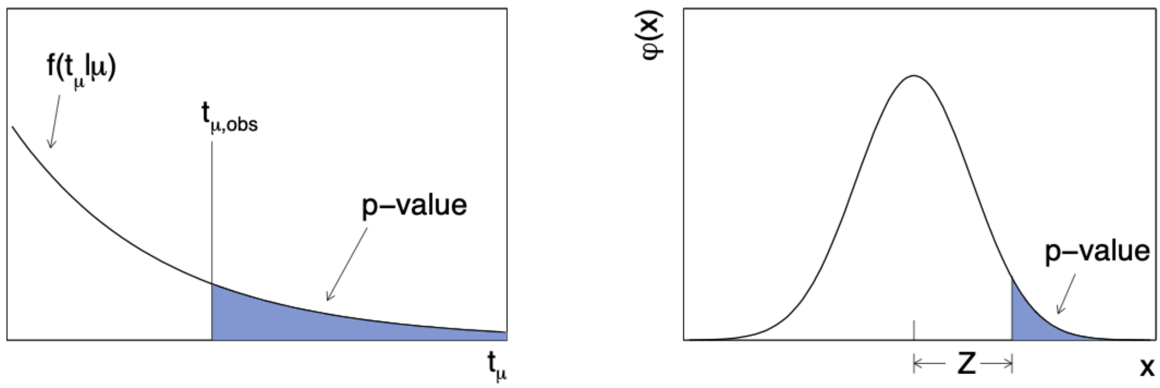


Figure 5.1: Plots depicting the relationship of p-value to (left) the test statistic and (right) to the Z value. Figures are taken from Reference [112].

The discovery case

The rejection of the background-only hypothesis H_0 is needed to claim the presence of a signal that is being searched for. The background-only hypothesis means that signal is absent i.e. $\mu = 0$ and the corresponding test statistic is,

$$q_0 = \begin{cases} -2\ln\lambda(0) & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0. \end{cases} \tag{5.15}$$

The observation of more events than expected by the background-only prediction increases μ as the excess favours the hypothesis with the presence of signal. The p-value

$$p_0 = \int_{q_0^{\text{obs}}}^{\infty} f(q_0|0)\, dq_0$$

quantifies the incompatibility of the background-only hypothesis with the observation made, where the pdf $f(q_0|0)$ can be defined as [112]

$$f(q_0|0) = \frac{1}{2}\delta(q_0) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_0}}e^{-q_0/2} \tag{5.17}$$

under the assumption of the asymptotic approximation.

Setting of an upper limit

If the goal of the statistical inference is to set a limit on the expected signal process, the null and alternative hypotheses are exchanged relative to the discovery case i.e. signal plus background hypothesis becomes the null hypothesis subject to testing.

$$t_\mu = \begin{cases} -2\ln\lambda(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu. \end{cases} \tag{5.18}$$

The case $\hat{\mu} > \mu$ occurs if the observed data exceeds the signal plus background prediction. This would still be more compatible with the signal plus background hypothesis compared to the background-only hypothesis. Therefore, the MLE is set to zero for this case that does not favour the rejection of the hypothesis. The p-value is

$$p_\mu = \int_{t_\mu^{\mathrm{obs}}}^{\infty} f(t_\mu|\mu)dt_\mu,$$

where the pdf $f(t_\mu|\mu)$ can be estimated in the asymptotic approximation [112], and using $p = 0.05$ corresponding to 95% confidence level, it leads to the upper limit

$$\mu_{\mathrm{up}} = \hat{\mu} + \sigma\Phi^{-1}(1 - 0.05) = \hat{\mu} + 1.64\sigma. \tag{5.20}$$

If the observed data is less than the prediction, while not excluding $\mu = 0$, the estimate $\hat{\mu}$ can get very small yielding tighter upper limits. These limits are actually tight beyond the sensitivity of the experimental measurement. This problem is addressed by the CLs [112] method, which modifies the p-value p_μ to

$$p'_\mu = \frac{p_\mu}{1 - p_b}, \tag{5.21}$$

where p' is the modified p-value given by the CLs method and p_b is the p-value of the background-only hypothesis

$$p_b = 1 - \int_{t_\mu^{\mathrm{obs}}}^{\infty} f(t_\mu|0)dt_\mu.$$

Expected results and the Asimov dataset

In HEP analyses, a measurement is often quoted together with the expected result, which is the median result of an hypothesis testing obtained using distributions of test statistics generated by repeated MC pseudo-experiments. The expected results provide crucial information about an experiment's designated sensitivity as they do not depend on the fluctuations in data. It is also possible to compute the expected results using a single and artificial MC sample called the Asimov dataset. This way the Asimov dataset constitutes a reference point from which the deviations in data could be interpreted. The special feature of the Asimov dataset is that, in it, all observed quantities are equal to their expected values. Thus, instead of calculating the median (med) of many pseudo-experiments, the Asimov dataset provides the median values immediately by construction. For a given μ' , the test statistic $t_{\mu,A}$ provided by the Asimov (A) dataset can be used to calculate the variance (σ^2) using the expression [112]

$$\sigma^2 = \frac{(\mu - \mu')^2}{t_{\mu,A}}. \tag{5.23}$$

The expected discovery significance is then given by setting $\mu = 0$:

$$\mathrm{med}[Z_0|\mu'] = \sqrt{t_{0,A}}. \tag{5.24}$$

5.2 Multivariate techniques

A multivariate discriminant is a combination of multiple variables into a single discriminator in a way that separates a target better than any of the variables alone. Such a discriminant resulting from combining N input variables reduces an N-dimensional space into a single variable, enabling the exploitation of correlations between variables. In HEP analysis, it is common to develop multivariate discriminants by using machine-learning methods. ML is an automated way of statistical optimization that develops a model by processing the input data with an algorithm that provides feedback for iterative improvement. This process of iteratively adjusting a model is called *learning*. There are three general approaches to learning:

- Supervised Learning: Supervision in this context indicated that the information of the correct outcome is available to the method. This information is used to improve the discrimination performance by comparing the mapping between the calculated outcome and the true value. The learning happens through iteratively modifying the model so as to increase the rate of matching between the calculated and the true outcomes.

- Unsupervised Learning: In unsupervised learning, the true outcome is not used to correct the calculation made by the model. Instead, the algorithm tries to extract features by seeking and recognizing patterns (such as clustering, segmentation etc.) in the unlabelled dataset in a purely data-driven way.

- Reinforced Learning: In reinforced learning, the input is not data but an *action*, which is defined as the mechanism of changing from the present *state* into a new one. This is particularly useful for handling problems with an insufficient amount of data or no clear final outcome. The learning algorithm defines a *reward* to be maximised, and actions that maximise the reward are chosen, leading to a series of actions resulting in a path to the optimal outcome.

In this thesis only supervised learning is used. True outcomes used for learning are provided by MC simulation and their format depends on the type of problem. In supervised learning two types of prediction problems are distinguished:

- Regression: In regression tasks, the target is to estimate the continuous numerical value of a variable, for example the energy deposited in the detector. The true outcome used for learning is a numerical value and is called the *target*.
- Classification: In classification problems, the aim is to make a discrete categorisation between types of events, such as signal and background. Classification problems can be binary or multi-dimensional. In binary classifiers only one type of signal and one type of (combined) background is defined. In multi-classification problems there are more than two categories. The true outcomes are the labels of the correct categories events belong to.

5.2.1 Training, testing and validation of machine-learning models

In supervised learning methods, the model building process where learning happens is called *training*. Given its parameters, the data sample used during the training determines the output of the model. The training sample comes from MC simulated events that are used for training as the MC truth information provides the true values needed. However, the trained model later has to be evaluated on data as well. Evaluation of the model response on samples is called *application*. The true discrimination performance of the model on data cannot be determined as data events do not have true values that could be compared to predictions from the training. Thus, the generalizability of the discrimination performance achieved by training on a certain dataset into another dataset that has not been used in the training, needs to be studied. In principle, a discriminant performs best on the training dataset as exactly the same events that have been used to develop the model are the ones the model is applied to. In case the training becomes sensitive to statistical fluctuations in a dataset, it might learn to separate based on the characteristics of individual events. However, this fine-tuning is often not generalizable as statistical fluctuations in another dataset might cause individual events to flip to the other side of this highly specialized discrimination criterion. This is visualised in Figure 5.2 (right) where the discrimination boundary becomes specific to training events and loses its generalizability. Such loss of performance is called *overtraining*.

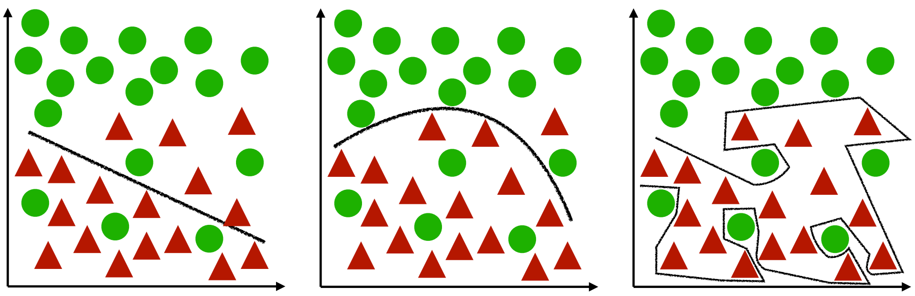


Figure 5.2: Depiction of a (left) suboptimal, (middle) optimal and (right) overtrained discriminator. The black lines represent the discrimination boundary between two different classes of events represented by green circles and red triangles.

To check the overtraining, a process called *testing* is used. During testing, the resulting discriminant from the training is applied to another sample which is not used for training. For example if the training is done using even numbered events in a sample, odd numbered events are used in the testing and vice versa. This ensures that the performance of the trained discriminant is evaluated on a set of events that has not been used for the training. Therefore it represents a test of overtraining. Although the impact of overtraining is mostly spotted in the training–testing process, it can induce a bias. When both, training and testing sets are used during optimization of the final model, errors on these models will be lower than the actual error. This effect becomes important when setups are highly optimized as the optimization performance is given and guided by the testing result, and in highly optimized setups this feedback is repeated many times.

A third sample, called *validation* set, is introduced to resolve this bias. A validation set has the same properties as the testing set, except that it is not used in the optimization and tuning of the model. Once the model is finally optimized to the best possible testing performance, to validate that it actually performs well with never-seen-before data, the training is applied to the validation set and the outcome is compared to the result of applying the model only to the testing set. In ML theory, proper optimization requires this three step approach of training–testing–validation. However, in HEP analyses the validation step is often ignored as the expected bias is often minuscule compared to the one that could be addressed using only the training–testing scheme.

Some of the commonly used ML techniques are Neural Networks, Support-Vector-Machines, Boosted Decision Trees (BDTs) and k-Nearest-Neighbour algorithms. In this thesis, BDTs are used and their properties are introduced in the next Section.

5.2.2 Boosted Decision Trees

The MVA studies in this thesis are conducted using the Toolkit for Multivariate Analysis (TMVA) [116], a software package for multi-variate analysis that is integrated in the ROOT data analysis framework [93]. TMVA provides many methods to tune, train, test and investigate various ML methods, including a graphical interface that produces plots. Relevant settings and outputs used in this thesis are described here; for more technical details the reader is referred to the TMVA manual [116]. It has been observed that the ROOT version used in the BDT studies can affect the output. Re-running the same setup while changing the ROOT version can shift the distribution shapes and outcomes slightly. This is understood in relation to changes in random number generating methods between different ROOT versions. Different ways of generating random numbers affect the TMVA calculation methods, hence the results. In this thesis ROOT version 6.14.08 is used.

Decision tree construction

A decision tree is a way of splitting input data using a given set of input variables. The choice of variables happens according to their expected discrimination power which is often guided by a mix of physics insight and trial and error.

Given two normalised distributions of signal $s(x)$ and background $b(x)$ of a variable x, TMVA uses the separation formula

$$D^2 = \frac{1}{2}\int \frac{(s(x) - b(x))^2}{s(x) + b(x)}\,dx$$

to quantify and rank the variables with $D = 0$ and $D = 1$ corresponding to identical and disjoint distributions, respectively. In the typical case of classification task in HEP analyses, the decision tree is a binary classifier that decides whether an event belongs to the signal or a background category.

The basic structure of a decision tree is schematically shown in Figure 5.3. Growing a tree happens through a series of bifurcations at *nodes*, a process called splitting. Each split is a decision that attempts at separating signal and background events, therefore creating two (left and right) child nodes, each enriched in one class of events.

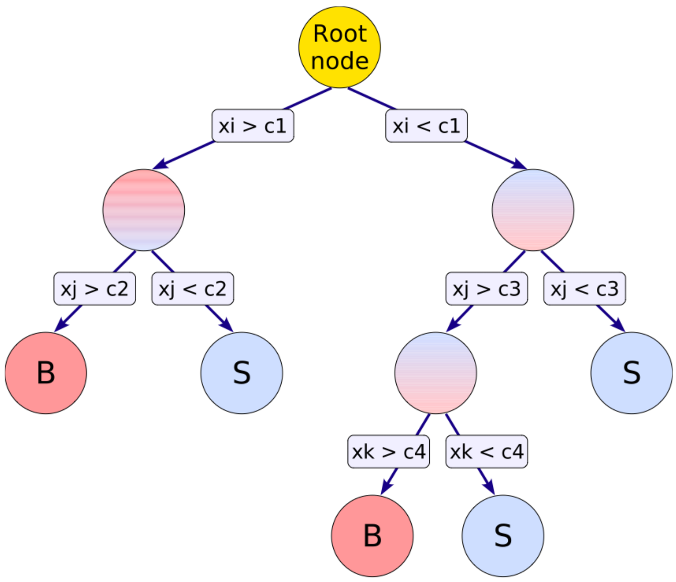


Figure 5.3: Schematic depiction of a decision tree. Circles represent the nodes, labels in the middle of arrows are the optimal cuts on a variable x creating the split. The labels B and S at the leaves represent the background and signal classification based on the majority of the events in that leaf. The Figure is taken from Reference [116].

The purity p is defined as the fraction of signal events to the all events in a given node. If each child node only comprises one type of events, then the node with signal events will have $p = 1$ and the node with background events will have $p = 0$. In this case no further splits are needed as a complete separation has been achieved. However, this is rarely possible since signal and backgrounds often have some overlap in the variable distributions. Those mis-classified events introduce an impurity $(1 - p)$ in the nodes and this is called the error of the training. The split with maximum separation is the choice that leads to the maximum of the purity or minimum of the impurity as they are complementary. The separation is quantified using the Gini Index [116]

$$G = p(1 - p). \tag{5.26}$$

The Gini distribution has a maximum at $G = 1/4$ corresponding to $p = 0.5$ where no separation exists as equal number of signal and background events populate a node. The minimum (maximum) of p is reached for $G = 0$ ($G = 1/4$). The separation gain at a node is given by

$$\Delta G = N_{\text{parent}}.G_{\text{parent}} - N_{\text{left node}}.G_{\text{left}} - N_{\text{right}}.G_{\text{right}} \tag{5.27}$$

where the multiplication by the number of events N of the specified nodes weighs the purity by the node size.

The splits are created in a sequential manner and terminate at the *leave*s, that is the final classification attribute of the event, once a stop condition is met. The stop conditions vary depending on tunable parameters of the tree-growing algorithm.

Hyperparameter optimization and performance metric

The properties of decision trees and how the ML algorithm processes information provided by them are determined by a set of tunable *hyperparameters*. In this thesis six common hyperparameters are studied. Three of them regulate the learning process and are described later in the Section. The three other parameters determine the structure of a decision tree and are described by the following list:

- Maximum Depth (Max.D.): This parameter determines the number of maximum split nodes allowed, thus the depth of the decision tree. Deeper trees exploit the correlations better and generally lead to an increased performance. However, deeper trees also increase the risk of overtraining as they could lead to leafs populated with small number of events.
- Number of Cuts (nCuts): TMVA accepts variables in the form of binned histograms. This parameter sets the granularity of the variable distributions i.e. number of bins covering the range between the minimum and the maximum values of a variable. An increased number of cuts enable the algorithm to utilize narrower ranges and is expected to improve the performance.
- Minimum event fraction (nMin): The minimum number of allowed fraction of events in a leaf is set by this parameter. When nMin is reached, the growth of the tree is terminated. Smaller values of nMin enable the tree to process more events but could result in an overtraining if leaves do not have enough statistics.

The set of hyperparameters that yields the best performance for a ML algorithm is not known *a priori*. The process of finding the best set of hyperparameters is called *optimization* and often a grid-scan over a range of hyperparameter values is performed to find the optimal set. The ML setup is re-run by changing a hyperparameter value each time and the resulting performance is compared to the outcome of the previous setup with different parameter values. If the performance is better, then the hyperparameter value is updated.
In a classification problem, a commonly used metric to visualise and quantify the discrimination performance is the Receiver-Operator-Characteristic (ROC) curve and its integral. The ROC curve is created by plotting the background rejection rate ($1 - \epsilon_{bkg}$) as a function of the signal efficiency (ϵ_{sig}) with ranges of both parameters between 0 and 1. The area under the curve (AUC) is defined as the integral of the area under the ROC curve, with larger values corresponding to a better separation. A maximum AUC of 1 corresponds to a perfect discrimination whereas an AUC of 0.5 quantifies the

separation power of a random classification of events. In Figure 5.4, two example ROC curves are depicted on the same plot.

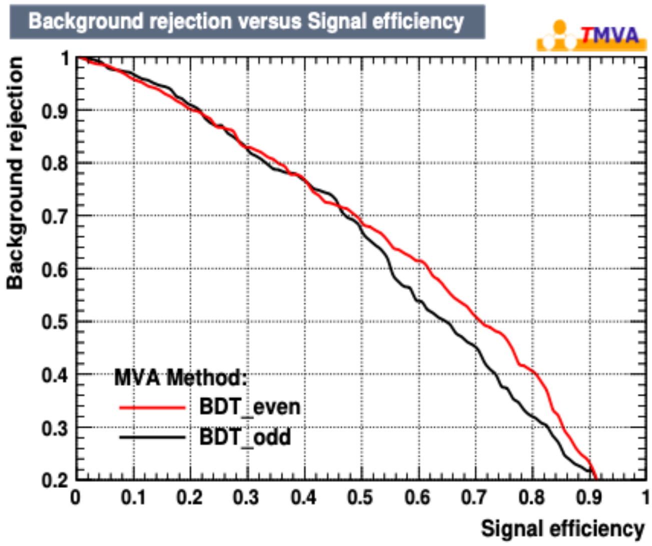


Figure 5.4: Visualisation of two ROC curves with the y-axis showing the background rejection rate ($1 - \epsilon_{bkg}$) and the x-axis showing the signal efficiency (ϵ_{sig}).

Since no estimation of uncertainty can be done using a single AUC value, the stability of the BDT performance in relation to statistical fluctuations due to the choice of the specific sample used, needs to be studied. In order to arrive at an estimate on the testing response of the BDT training, the Poisson bootstrap resampling method [117, 118] with replacement is used. In this method, the test sample is resampled by randomly drawing events until the original sample size is reached. Replacement in this context refers to allowing the same event to be drawn multiple times. By resampling, a set of test samples is generated and these are used to estimate the variations in the training. Randomly drawing events follows binomial distribution and is a computationally expensive process. An approximate resampling method with a lower computational cost can be achieved by assigning a random weight to each original event, which is distributed according to a Poisson distribution (Eqn. 5.5) with $\lambda = 1$. This method is justified in the limit of large numbers, where the Poisson distribution is a good approximation of the binomial distribution as shown in Figure 5.5.

Boosting

A single decision tree by itself is a *weak learner*. It cannot utilize a feedback response and improve its mistakes, that is the categorization of mis-classified events in the final leaves. This problem is solved via *boosting*. The term boosting refers to the iterative learning process by training multiple decision trees, where results from the previous decision tree training are used to guide the training of the next one. This is mathematically expressed as the weighted sum $F(\vec{x}, \mathrm{P})$ resulting from adding functions $f(\vec{x}, a_m)$ representing individual decision trees m with a set of inputs $\vec{x}$ and parameters a_m, multiplied with an associated weight of w_m:

$$F(\vec{x}; \mathrm{P}) = \Sigma_{m=0}^{M} w_m f(\vec{x}; a_m) \quad , \quad \mathrm{P} \in \{w_m, a_m\}.$$

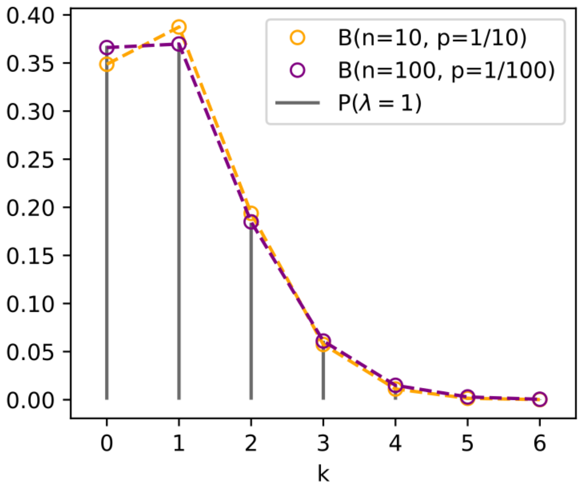


Figure 5.5: A normalised Poisson distribution (P) is compared to two binomial distributions (B), having $n = 10$ and $n = 100$ events. The Figure is taken from Reference [119].

Boosting is the attempt to adjust the parameters P such that $F(\vec{x}) - y$, corresponding to the deviation of constructed model $F(\vec{x})$ from the true value y, is minimized. The deviation between the two values is quantified by the *loss-function* $L(F, y)$. The minimization of the loss-function is done by taking the gradient of the loss-function with respect to $F(\vec{x})$. The negative of this gradient is then used to update the model.

The three hyperparameters related to the boosting process are:

- Number of Trees (nTrees): The number of trees to be trained is set by this variable.
- Bagging fraction (nBag): In order to stabilize the training process, a randomly sampled subset of training events is used in each tree. The fraction of training events that are sampled are set using the bagging fraction.
- Shrinkage (LR%): This parameter, also known as learning rate, sets the weight of the individual trees and regulates their contribution to the final sum $F(\vec{x})$.

There are two main boosting algorithms: Adaptive Boosting (AdaBoost) and Gradient Boosting (GradBoost). Gradient Boosting allows for different types of loss functions. The TMVA default for the loss function in classification problems is the binomial log-likelihood loss:

$$L_{\text{Grad}}(F, y) = \ln(1 + e^{-2F(\vec{x})y}). \tag{5.29}$$

AdaBoost is a specific case of the more general method of Gradient Boosting, with the loss-function being equal to an exponential function:

$$L_{\text{Ada}}(F, y) = \mathrm{e}^{-F(\vec{x})y}. \tag{5.30}$$

CHAPTER 6

Evidence for four top-quark production in the same-sign dilepton and multilepton channel

This chapter details the production cross-section measurement of the four-top-quark process in the same-sign dilepton and multilepton (2LSS/3L) channels leading to the first evidence of this process. The measurement has been performed using full Run II proton–proton collision data collected by the ATLAS detector between 2015 and 2018, corresponding to an integrated luminosity of 139.0 $\pm$ 2.4 fb^{-1} at a centre-of-mass energy of 13 TeV. The $t\bar{t}t\bar{t}\rightarrow$ 2LSS/3L process has been discussed in Chapter 3. At lowest order, the 2LSS (3L) processes presents itself with a final state composition of 2 same-sign (3) leptons, 2 (3) neutrinos, 4 b-tagged jets, 6 (4) light jets and a large amount of H_{T} .

In Section 6.1, the simulated samples of both, signal and background processes are described. These datasets are then subjected to certain selection criteria filtering the type of analysis objects and events available for further analysis. These selections are motivated and listed in Section 6.2.

The accurate estimation of background processes is a crucial part of this analysis as it affects the uncertainty on the measurement of the signal. Section 6.3 describes various classes of backgrounds and control regions considered in this analysis, as well as different methods used for their estimation and validation. In Section 6.4 an MVA discriminant built and optimised for increasing the sensitivity to the signal process is described.

Various sources of systematic uncertainties in analysis methods and tools, or instrumentation are defined and listed in Section 6.5. The results are presented in Section 6.6, estimated using a simultaneous maximum likelihood fit to the CRs and the SR, including the systematic uncertainties. In Section 6.6.3, various post-unblinding checks are summarised, that have been conducted to validate and study the stability of the results provided by the fit setup.

In the last section the cross-section results are employed to set an upper limit on the SM top-quark Yukawa coupling parameter y_t.

6.1 Monte-Carlo simulation samples

The simulated Monte-Carlo samples are produced in three campaigns that reflect three different periods of data-taking conditions corresponding to the years 2015–2016, 2017 and 2018. In most cases, in order to estimate the uncertainties emerging from choosing a certain setting for the simulation, more than one sample for the same process is produced. The main sample used in the estimation of yields will be called the *nominal*, and the other sample or samples used only for comparison and therefore extraction of the relevant systematic uncertainties, will be called *alternative* samples. Nominal samples are processed with the full detector simulation using the GEANT4 software. Alternative samples are in most cases processed with Fast Simulation, only. The samples and their main simulation settings are summarised in Table 6.1.

In this analysis, the masses of the Higgs-boson and the top-quark are set to 125.0 GeV and 172.5 GeV, respectively. Apart from the ones where the SHERPA event generator is used, decays of b- and c-hadrons are simulated with the `EVTGEN v1.2.0` software. Pile-up events and their profiles are simulated by adding minimum-bias events generated with Pythia8.186 to the hard-process, using the A3 tune. In order to account for the modelling differences between measured data and the simulation samples (such as selection efficiency of physics objects), various correction factors are applied in the form of event weights.

Process	Use Case	Order	ME	PDF	PS	PS PDF	Tune
$t\bar{t}t\bar{t}$	Nominal	QCD NLO	MG5 2.6.2	NNPDF3.1NLO	Pythia8.230	NNPDF2.3LO	A14
	BDT	QCD LO	MG5 2.6.2	NNPDF3.1NLO	Pythia8.230	NNPDF2.3LO	A14
	Alt.	QCD NLO	MG5 2.6.2	NNPDF3.1NLO	HERWIG 7.04	MMHT2014LO	H7UE
$t\bar{t}$	Nominal	QCD NLO	POWHEGBOX v2 (h_{damp}=1.5 m_t)	NNPDF3.1NLO	Pythia8.230	NNPDF2.3LO	A14
	Alt. ME	QCD NLO	MG5 2.6.0	NNPDF3.1NLO	Pythia8.230	NNPDF2.3LO	A14
	Alt. h_{damp}	QCD NLO	POWHEGBOX v2 (h_{damp}=3.0 m_t)	NNPDF3.1NLO	Pythia8.230	NNPDF2.3LO	A14
	Alt. PS	QCD NLO	POWHEGBOX v2 (h_{damp}=1.5 m_t)	NNPDF3.1NLO	HERWIG 7.04	MMHT2014LO	H7UE
S-top tW-	Nominal	QCD NLO	POWHEGBOX v2 (DR)	NNPDF3.1NLO	Pythia8.230	NNPDF2.3LO	A14
s-	Nominal	QCD NLO	POWHEGBOX v2	NNPDF3.1NLO	Pythia8.230	NNPDF2.3LO	A14
t-	Nominal	QCD NLO	POWHEGBOX v2 (4FS)	NNPDF3.1NLOnf4	Pythia8.230	NNPDF2.3LO	A14
$t\bar{t}Z$	Nominal	QCD NLO	MG5 2.3.3	NNPDF3.1NLO	Pythia8.210	NNPDF2.3LO	A14
$t\bar{t}W$	Nominal	QCD NLO	MEPS@NLO (add. 1/2 partons)	NNPDF3.1NLO	SHERPA	NNPDF2.3LO	SHERPA
	Alt.	QCD NLO	MG5 2.3.3	NNPDF3.1NLO	Pythia8.210	NNPDF2.3LO	A14
$t\bar{t}H$	Nominal	QCD NLO	POWHEGBOX	NNPDF3.1NLO	Pythia8.230	NNPDF2.3LO	A14
	Alt. ME	QCD NLO	MG5 2.3.3	NNPDF3.1NLO	Pythia8.230	NNPDF2.3LO	A14
tWZ	Nominal	QCD NLO	MG5 2.3.3	NNPDF3.1NLO	Pythia8.212	NNPDF2.3LO	A14
tZ and $t\bar{t}VV$	Nominal	QCD LO	MG5 2.3.3	NNPDF3.1NLO	Pythia8.230	NNPDF2.3LO	A14
$3t$	Nominal	QCD LO	MG5 2.3.3	NNPDF3.1NLO	Pythia8.230	NNPDF2.3LO	A14
V+jets	Nominal	QCD LO(+4j), NLO(+2j)	MEPS@NLO	NNPDF3.1NLO	SHERPA	NNPDF2.3LO	SHERPA
VH	Nominal	QCD LO	Pythia8.230	NNPDF2.3LOlo	Pythia8.230	NNPDF2.3LO	A14

Table 6.1: Summary of the main simulation parameters of the Monte-Carlo samples. Abbreviations are: MadGraph5_aMC@NLO(MG5), Matrix Element (ME), Parton Shower (PS), Alternative (Alt.), Boosted Decision Tree (BDT), top-quark mass (m_t), Diagram Removal (DR), 4-flavour scheme (4FS), Single-top (S-top).

$t\bar{t}t\bar{t}$

Signal samples are produced at both, LO and NLO accuracies. The LO-level sample is exclusively employed for the MVA training, and not used in the final fit. This choice stems from the fact that the MVA training algorithm cannot use negative-weight events, which are only occurring in the NLO sample. The shapes of LO and NLO samples are shown to be in good agreement in Section 6.4.1 and in Figure 6.5, validating this approach. The LO sample is generated with QCD diagrams using MadGraph5_aMC@NLOwith the NNPDF3.1NLO PDF set. The renormalisation and factorisation scales are set to the functional form of $H_T/4$. Top-quarks are decayed via the MADSPIN module of the MG5 generator. Generated and decayed parton-level events then undergo hadronisation and showering using PYTHIA 8.230 with the A14 tune and the NNPDF2.3LOPDF set. The `EVTGEN v1.6.0` program is used to simulate the decays of bottom and charm hadrons.

At NLO accuracy, two samples are produced: The nominal sample has the same settings as the LO sample, except that it is generated at NLO accuracy. The alternative sample also has NLO accuracy and differs from the nominal sample at the hadronisation and showering stage, using HERWIG 7.04 with the H7UE tune with the MMHT2014LOPDF set.

$t\bar{t}$

$t\bar{t}$ events are produced at NLO precision in QCD using the POWHEGBOX v2 generator with NNPDF3.1NLO. Parton shower and hadronisation are simulated with Pythia8.230 with the A14 set of tunes and the NNPDF2.3LOPDF set. The h_{damp} parameter regulating the downscaling of the cross-section of real emissions is set to 1.5 times the mass of the top-quark [120]. When considering high-p_T radiation, lower h_{damp} values lead to larger suppression of the higher-order contributions [121]. LO processes are not impacted from the tuning of the h_{damp} parameter. A correction to the H_T distribution is performed to account for the relatively low number of events at the high-end tail. This correction is applied by reweighting the nominal sample events, relative to another sample generated with H_T -slicing, which is a method that produces a more flat distribution over the whole H_T spectrum. Three alternative samples are produced to estimate the impact of various generator parameters where only parameters of interest are changed and all other settings are kept as for the nominal sample. To probe the effect of the choice of the generator, another sample with MadGraph5_aMC@NLO2.6.0 with the NNPDF3.1NLO PDF set and A14 tune is used. Another alternative sample is produced using the same generators but HERWIG 7.04 with the H7UE tune and the MMHT2014LO set for parton shower and hadronisation. Lastly, an alternative sample is generated while only changing the h_{damp} parameter from 1.5 to 3.0 times the top-quark mass to assess the influence of this variable. The production cross-section is normalised to its corresponding next-to-next-to-leading order (NNLO) predictions where soft-gluon terms at next-to-next-to-leading logarithmic (NNLL) order are also accounted for via resummation.

Single-top

Single-top quark events comprises s-, t- and tW- channels. Single-top quark production modes associated with Z-bosons are discussed below among the rare processes. Sharing the same nominal settings with $t\bar{t}$ production, single-top production is simulated at NLO precision in QCD using 0OWHEG generator with NNPDF3.1NLO for the tW- and s- channels, and with the NNPDF3.1NLONF4 PDF set for the t-channel. The interference between the $t\bar{t}$ and the tW-channel is removed with the Diagram Removal (DR) method [36]. Generated events in all channels are interfaced with PYTHIA8.230 with the A14 set of tunes and the NNPDF2.3LOPDF set for parton shower and hadronisation.

$t\bar{t}W$

The $t\bar{t}W$ process is generated using SHERPA v2.2.1 with the NNPDF3.1NLO PDF set. The $t\bar{t}W$ background is the most significant background for this analysis and special emphasis is put on the modelling of the background with additional partons in the production, since $t\bar{t}W$ events with extra partons appear similar to the signal process. Using the COMIX [122] and OpenLoops [123, 124] libraries, the matrix elements are calculated at LO QCD with up to two additional partons, and at NLO QCD with one additional parton. The merging of generated events with different additional parton multiplicities into the SHERPA parton showering algorithm is achieved with the MEPS@NLO [125, 126, 127, 128] prescription at a merging scale of 30 GeV. Events where a W-boson decays into a c- and a b-quark are not simulated.[1] The total production cross-section is normalised to $601 \pm 76\ \text{fb}^{-1}$ which is the calculated cross-section at NLO QCD including the NLO EW corrections [129, 130, 131]. An alternative sample is produced at QCD NLO precision using the MADGRAPH5_AMC@NLO2.3.3 generator with the NNPDF3.1NLO PDF set and the SHERPA tune. Generated events are interfaced with PYTHIA8.210 with the NNPDF2.3LO PDF set using the A14 tune.

$t\bar{t}H$

The production of the $t\bar{t}H$ process is done at NLO precision in QCD using the POWHEGBOX v2 generator with the NNPDF3.1NLO PDF set. The h_{damp} parameter has the functional form of $1.5 \times (2m_t - m_H)/2$. The generated events are interfaced with PYTHIA8.230 with the A14 set of tunes and the NNPDF2.3LO PDF set for parton shower and hadronisation. Similarly to the $t\bar{t}W$ process, the total production cross-section is normalised to the calculated cross-section at NLO QCD including the NLO EW corrections. The impact of different generators is evaluated with an alternative sample with the same settings except using the MADGRAPH5_AMC@NLO2.3.3 ME generator instead of POWHEGBOX .

$t\bar{t}Z$

The $t\bar{t}Z$ process has its events produced at QCD NLO precision using the MADGRAPH5_AMC@NLO2.3.3 generator with the NNPDF3.1NLO PDF set. Parton showering and hadronisation are simulated using PYTHIA8.210 with the NNPDF2.3LO PDF set.

tZ and $t\bar{t}XX$

These rare top-quark processes (where $X \in \{H, W, Z\}$) are generated at LO precision in QCD and normalised to NLO QCD cross-section values based on theoretical calculations. No alternative samples are produced for comparison due to the relatively minor impact of these samples on the analysis. The production is done with the MADGRAPH5_AMC@NLO2.3.3 generator with the NNPDF3.1NLO PDF set. Parton showering and hadronisation are simulated using PYTHIA with the NNPDF2.3LO PDF set.

[1] Due to $|V_{cb}|^2/|V_{cs}|^2 \sim 10^{-3}$, $W \to cb$ events are suppressed relative to $W \to cs$.

tWZ

Similarly to tZ and $t\bar{t}XX$ above, only a nominal set is produced for the tWZ process, albeit at NLO precision. Events are produced using the MadGraph5_aMC@NLO2.3.3 generator with the NNPDF3.1NLO PDF set. Parton showering and hadronisation are simulated using Pythia8.212 with the NNPDF2.3LO PDF set.

$3t$

The $3t$ background process is modelled with settings similar to the rare processes above and a cross-section normalised to 1.64 fb^{-1}, the LO QCD prediction, as no NLO level theory calculation was available at the time of this analysis. The simulation is performed based on the description in Reference [132]. The production is done with the MadGraph5_aMC@NLO2.3.3 generator with the NNPDF3.1NLO PDF set. Parton showering and hadronisation are simulated using Pythia with the NNPDF2.3LO PDF set.

Z+jets and W+jets

Samples for the Z+jets and W+jets processes are generated using SHERPA v2.2.2 with the NNPDF3.0NNLOPDF set. The samples have two different accuracies at the ME generation step that correspond to two different setups for generating additional jets: at LO, up to four jets, at NLO, up to two jets are included using the COMIX and OpenLoops libraries. The merging of generated events with different jet multiplicities is achieved according to the MEPS@NLO rescription with the SHERPA parton showering algorithm. The production cross-sections are normalised to their corresponding NNLO order predictions.

VH

The associated production of a Higgs boson with a vector boson ($V \in \{W, Z\}$) is simulated using the Pythia8.230 generator with the NNPDF2.3LO PDF set and A14 tune. These LO precision samples are then normalised to the theoretical predictions at QCD NNLO and EW NLO accuracies.

VV and VVV

Diboson (VV) and triboson (VVV) samples ($V \in \{W, Z\}$) are generated with SHERPA v2.2.1 using the NNPDF3.0NNLOPDF set. Both processes are normalised to the NLO level cross-section value, using QCD calculations. Diboson events with non-all-hadronic final states have two different accuracies at the ME generation step that corresponds to two different setups for generating additional partons: at LO, up to three additional partons, at NLO, up to one additional parton. In the case of triboson events, the ME has NLO accuracy for the inclusive process and accounts for up to two additional parton emissions at LO-level. The merging of generated events with different jet multiplicities is achieved according to the MEPS@NLO prescription with the SHERPA parton showering algorithm based on the Catani–Seymour dipole [133]. The OpenLoops library is used for calculating virtual QCD corrections.

6.2 Object and event selection

6.2.1 Object selection

This section describes the requirements on the objects and the selection of events that are used in the analysis. The definition and reconstruction of objects and events in ATLAS is discussed in Chapter 4. The properties of the main objects are summarized in Table 6.2, followed by a detailed discussion below.

Object	WP	Identification	Isolation	Minimum p_{T} [GeV]	$\lvert\eta\rvert$
Electron	Tight	TightLH, ECIDS ($e^{\pm}e^{\pm}, e^{\pm}\mu^{\pm}$)	FCTight	28	< 2.47 and $\notin [1.37, 1.52]$
	Loose	MediumLH, ECIDS ($e^{\pm}e^{\pm}, e^{\pm}\mu^{\pm}$)	-	28	< 2.47 and $\notin [1.37, 1.52]$
Muon	Tight	MediumLH	FixedCutTightTrackOnly	28	< 2.5
	Loose	MediumLH	-	28	< 2.5
Jets	-	JVT (for p_{T}< 60 GeV and $\lvert\eta\rvert$ < 2.4) and Jet Cleaning	-	25	< 2.5
b-tagged jets	77%	MV2c10	-	25	< 2.5

Table 6.2: Summary of definition and identification criteria of the main objects used in the analysis.

Electrons

Kinematic requirements for the selection of electrons are twofold: Electrons with a minimum p_{T} of 28 GeV and a pseudo-rapidity of $|\eta| < 2.47$ are selected excluding the transition region between the barrel and the endcaps, corresponding to $1.37 < |\eta| < 1.52$, where the detector performance degrades significantly. The transverse impact parameter significance ($d_0/\sigma(d_0)$) and the longitudinal impact parameter ($z_0 \sin(\theta)$) are required to be below 5 and 0.5 mm, respectively. Electrons used in the analysis have the Tight WP with TightLH identification and FCTight isolation criteria. The Loose WP is used in cases where an enriched selection of non-prompt leptons is needed to estimate this type of backgrounds with a large number of events which are otherwise suppressed by tighter identification criteria. Loose electrons have the LooseLH identification WP and they require no isolation criteria.

For the electrons in the $e^{\pm}e^{\pm}$ and $e^{\pm}\mu^{\pm}$ channels, an extra condition is required in order to reduce the amount of charge mis-assignment in the reconstruction. This condition is based on the Electron Charge Identification Selector (ECIDS) [103] tool, which is a multivariate discriminant trained on data with a $Z \to ee$ selection using the Boosted Decision Tree (BDT) method. The tool is able to select electrons with the correct charge in about 98% of cases and reject the electrons with mis-identified charges around 90% of the time.

Muons

The p_{T} condition of 28 GeV or above is applied to muon candidates as well. A pseudo-rapidity coverage of $|\eta| < 2.5$ is selected without excluding the calorimeter transition region as muons generally do not deposit much energy in the calorimeter. The transverse impact parameter significance and longitudinal impact parameters are required to be below 3 and 0.5 mm, respectively. Similarly to electrons, there are Tight and Loose variants of muons, with Tight being the nominal use case. Loose muons only have an identification requirement of Medium WP and they lack isolation criteria. A Medium identification requirement and, in addition, the FixedCutTightTrackOnly isolation is imposed on Tight muon tracks.

Jets

Jets are reconstructed by applying the anti-k_t algorithm with a fixed jet cone radius of $R = 0.4$ to the energy deposits from electromagnetic topological clusters. To veto jets that are likely to originate from non-collision sources, jet cleaning quality criteria are applied [134]. In order to suppress contributions from pile-up collisions, jets with a $p_{\mathrm{T}} < 60$ GeV and $|\eta| < 2.4$ have to satisfy a Jet Vertex Tagger (JVT) [135] discriminant value of 0.59 or above, corresponding to the WP at which the prompt jet efficiency is 92%. Jets with an MV2c10 discriminant value of at least 0.64 are identified as b-tagged jets. This is the default b-tagging WP used in the analysis, and it corresponds to 77% b-tagged jet selection efficiency. All jets are required to have $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.5$.

$E_{\mathrm{T}}^{\mathrm{miss}}$

The missing transverse energy used in this analysis is defined as described in Section 4.3.7 without any specific selection cuts. Reconstructed and selected objects of the analysis are used in the computation of $E_{\mathrm{T}}^{\mathrm{miss}}$ as well as the soft terms.

Overlap Removal

Overlap removal is a way of resolving the selection ambiguities between the above-defined objects except $E_{\mathrm{T}}^{\mathrm{miss}}$. An overlap removal method inspired from the BOOSTEDSLIDINGDRMU [136] algorithm is used in this analysis. This algorithm sequentially applies the criteria listed below, therefore creating an event selection where no overlap between electrons, muons and jets occurs. In each step, only the set of objects that have not been filtered out in any of previous steps is considered. The algorithm proceeds as follows:

1. If an electron candidate shares a track with a muon candidate, the electron candidate is removed.

2. If there is a jet within a distance $\Delta R < 0.2$ of an electron, the jet is removed. If this is the case for multiple jets, only the jet with the smallest ΔR to the electron is removed.

3. If there is an electron candidate within $\Delta R < 0.4$ of a jet, the electron is removed.

4. If a jet is within $\Delta R < 0.2$ of a muon, and has less than three associated tracks, it is removed.

5. If a muon track in the Inner Detector is matched with a jet that has less than three associated tracks, the jet is removed.

6. If the distance between a muon and a jet is $\Delta R < 0.4 + \frac{10\ \mathrm{GeV}}{p_{\mathrm{T}}^{\mu}}$, the muon is removed.

6.2.2 Event selection

As discussed in Section 4.1, the data used in this analysis is collected by the ATLAS detector between 2015 and 2018, which corresponds to the full Run-II period of LHC operations. During this period, the LHC provided proton–proton collisions at a center-of-mass energy of $\sqrt{s}$ = 13 TeV that were registered by the ATLAS detector with varying pile-up and luminosity values. The total amount of data categorised by ATLAS as good-enough to be used in analyses is 139.0 ± 2.4 fb^{-1}. Only events with at least one reconstructed primary vertex are used in this analysis. A vertex can be reconstructed from a minimum of two tracks in the inner detector, with a p_{T} of at least 0.4 GeV and a certain proximity to each other [137]. When multiple vertices can be reconstructed in an event, the vertex having the maximum $\Sigma p^2_{\mathrm{Ttrack}}$ value is selected as the Primary Vertex. Single lepton and dilepton triggers are used to select the events. Five types of lepton triggers are used in the selection of events corresponding to two lepton types and their possible combinations. Single lepton triggers are used to select the events with either isolated leptons with low p_{T}, or high-p_{T} leptons with loose definitions. There is a p_{T} threshold range of 20–26 GeV over different data-taking periods and lepton flavours. Dilepton triggers are used to select events where no isolation requirements are applied to leptons, thus selecting loose leptons. Since it is less likely to trigger two leptons simultaneously, dilepton triggers have a lower threshold compared to single lepton triggers, with a range of 8–24 GeV. Details can be found in References [138, 139].

6.2.3 Event categorisation and analysis preselection

Before the actual analysis regions are defined, a loose preselection is applied in order to reduce the size of the simulated samples. It is loose because the purpose of the preselection is not to tightly fix any region for a specific analysis purpose, as it would be for a CR or SR. Instead, the aim is to generally remove events that are neither signal-like nor relevant for the background estimation studies. First, at least one b-tagged jet is required. This loose requirement removes events with no b-tagged jets such as QCD multi-jet production. Second, events are grouped into two main categories and seven final states, based on their lepton flavour and multiplicities. Events that do not fall into one of these categories are discarded. The definition of categories is based on the three p_{T}-leading loose leptons only. This approach is a legacy of the background estimation method requirements of the previous ATLAS analysis, conducted using partial Run II data. The method has not been used in the current analysis eventually, however, the structure has been kept due to practical and timing reasons.

An event is classified as same-sign dilepton (2LSS), if, out of three loose leptons, two same-charged leptons pass the tight requirement. The 2LSS channel comprises three final states: $e^{\pm}e^{\pm}$, $e^{\pm}\mu^{\pm}$, $\mu^{\pm}\mu^{\pm}$. The order of particle symbols does not contain any additional information such as p_{T} ordering of the particle types, but merely follows alphabetical order. For the $e^{\pm}e^{\pm}$ channel, two additional selections employing the invariant mass of the di-electron system (M_{ee}) are applied to suppress events that could enter this channel due to mis-identification of one of the electrons:

- $M_{ee} > 15$ GeV: This selection aims at removing $e^{\pm}e^{\mp}$ events produced by decays of J/ψ resonances.
- $|M_{ee} - 91| > 10$ GeV: This requirement removes events whose invariant mass is in the vicinity of the Z-boson mass, therefore likely to originate from the $e^{\pm}e^{\mp}$ decay mode of this particle.

The event is classified as multilepton (ML), if all of the three p_{T}-leading loose leptons pass the tight selection without any requirement on their charge. Four final states make up the ML category: eee, $ee\mu$, $e\mu\mu$, $\mu\mu\mu$. As in the 2LSS case, the order of particle flavours does not contain any information. For the ML events having opposite-sign same-flavour lepton pairs, it is important to suppress $Z(\rightarrow \ell^{\pm}\ell^{\mp})$ contributions. To achieve this, the invariant mass of the dileptonic system is required to be out of the 10 GeV mass window defined around the Z-boson mass i.e. $|M_{ee} - 91| > 10$ GeV.

It is important to note, that, considering a p_{T}-sorted lists of only three loose leptons for the event classification process, the type of a fourth lepton does not influence the categorisation decision. For example, an event with first, second and fourth p_{T}-leading loose leptons passing also the tight requirement will be classified as a dilepton event although three tight leptons are present. The rate of migration between categories due to mis-classification from the above-described effect is small and practically irrelevant for this analysis. Another noteworthy consequence is the implicit inclusion of the tetralepton channel. Events with four loose leptons which also pass the tight requirement will be categorised under ML events. For this analysis, only a few weighted events of the tetralepton channel are available. Thus the channel is not considered as a separate final state.

6.3 Background categorisation and estimation

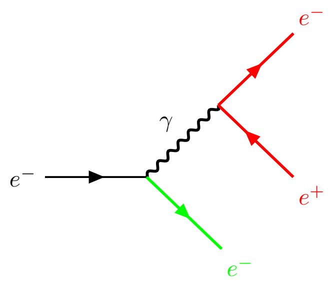
Background processes used in this analysis are divided into two main categories, based on the origin of final state leptons in the events. The origin is defined using the truth particle record in the simulated samples, and the classification scheme based on the `MCTruthClassifier` [140] tool. The association of a reconstructed lepton to its corresponding originating particle in the truth record is done via pairing it to the object that minimises the ΔR distance. This is also known as *truth matching* in the particle physics jargon.

6.3.1 Irreducible backgrounds

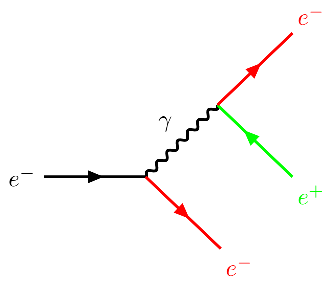
Background processes with prompt leptons having the same final state composition (charge and multiplicity) as the signal process are called *irreducible* backgrounds. This is because the final state composition is the same as the signal process not due to a measurement or detection error, but because of the nature of the physics process. Reconstructed prompt leptons with their charges flipped relative to their truth-particle counterparts are not included in this background category. In the case of the $t\bar{t}t\bar{t}$ analysis, there is another background process considered, namely the *trident process*, in which electrons are not prompt but have the same final state composition as the initial process. An electron leads to a trident process when it undergoes bremsstrahlung and the bremsstrahlung photon further creates an electron–positron pair.[2] If the reconstructed particle is the same as the incoming one, as depicted in Figure 6.1(a), the correct electron is reconstructed and the process is not a background process. If the reconstructed particle decays from the photon and has the opposite-charge of the incoming particle, as depicted in Figure 6.1(b), the process results in a charge mis-identification and is categorised as a "reducible" background process. This category is introduced and discussed in the next Section. If, among the three outgoing electrons, only the one decaying

[2] The three outgoing parts originating from one common origin is reminiscent of the trident of Poseidon, ancient Greek God of the sea.

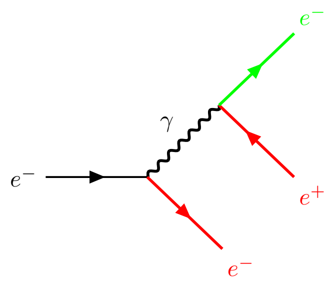
from the photon with the same charge as the incoming electron is reconstructed, the final state lepton composition will be same as the initial one. An example Feynman diagram of this case is shown in Figure 6.1(c). These events are also considered as irreducible background processes along the events with prompt leptons. It could be that such trident electrons will display kinematically different characteristics compared to prompt electrons. For instance, if a large fraction of the initial electron's momentum is transferred to the opposite-charge outgoing electron, the reconstructed same-charge electron will be softer. But since this background process is rare, such effects are found to be negligible.



(a) Trident process where the electron undergoing bremsstrahlung is reconstructed.



(b) Trident process where the positron decaying from the photon is reconstructed.



(c) Trident process where the electron decaying from the photon is reconstructed.

Figure 6.1: Example Feynman diagrams for trident decay process of an electron. The outgoing particle reconstructed by the detector is coloured in green. Outgoing particles that are not reconstructed by the detector are coloured red.

Irreducible backgrounds are the leading source of background in this analysis and dominated by the $t\bar{t}W$, $t\bar{t}Z$ and $t\bar{t}H$ processes. Minor irreducible backgrounds from $t\bar{t}XX$, VV and VVV productions, tWZ and tZq are collected into the *Others* category. Although small in amount, due to its similarity with the signal process, the triple top-quark production ($3t$) process is not included in the *Others* but kept separate. Among the irreducible backgrounds, only the $t\bar{t}W$ process is not directly estimated from the MC prediction. The special treatment of $t\bar{t}W$ is due to the observed discrepancies in the modelling of this process as motivated in Section 3.3. The normalisation of the $t\bar{t}W$ background process is scaled with a NF that is estimated from the data-assisted template-fit method using a CR enriched in $t\bar{t}W$ events and described in Section 6.3.3.

6.3.2 Reducible backgrounds

The second category, in contrast to the first one, is called *reducible* backgrounds. Reducible backgrounds are processes with a different reconstructed final state compared to the original initial physics process. Some examples are mis-identified charges, *non-prompt* leptons originating from jets or photons, or non-leptonic particles reconstructed as leptons (also known as *fake* leptons). There are various instrumental effects or limitations that could lead to such a difference. These backgrounds are called "reducible" because using better, more efficient detection tools could in principle improve particle identification and reduce these background contributions. Wrong reconstruction of more than one lepton in an event can occur, but it is rare (less than 5 raw events after the final selection). These events are included in the *Others* event category. There are also instances where the truth classifier fails to match true and reconstructed leptons, and thus returns no lepton information. Such events are discarded. Various types of reducible background processes with one wrongly reconstructed lepton that are considered in this analysis are described below. In Table 6.3, processes belonging to each background category are listed together with the background estimation methods used in the calculation of their contributions. In the last column, relative amount of each process in the SR selection before performing any fit are reported in terms of percentages. The fit setup is introduced and described in Section 6.6. A comparison between pre- and post-fit yields are given in Table 6.15.

Label	Background Type	Constituent Processes	Lepton (ℓ) Classification	Estimation Method	≈Pre-Fit % in SR
$t\bar{t}W$	Irreducible	$t\bar{t}W$	All leptons prompt or same-charge trident	Template Fit	22.6
$t\bar{t}Z$		$t\bar{t}Z$		MC Simulation	18.4
$t\bar{t}H$		$t\bar{t}H$		MC Simulation	14.3
$3t$		$3t$		MC Simulation	1.1
Others	Irreducible	VV, VVV, $t\bar{t}XX$, VH, tZ , tWZ , V+Jets	All leptons prompt or same-charge trident	MC Simulation	12.5
	Reducible	$t\bar{t}$+jets, V+jets, Single-top	1ℓ from q/g jets or light-flavour (LF) hadron decays		
	Reducible	All	$> 1\ell$ not prompt or same-charge trident		
HF_e	Reducible	$tbart$+jets, V+jets, Single-top	$1e$ originate from b- or c-hadron decays	Template Fit	1.8
HF_μ			1μ from b- or c-hadron decays		4
Mat. Cv.			1ℓ from γ conversion in detector material		4.7
$\gamma*$			1ℓ from virtual γ conversion in hard-scattering process		3.4
QMID	Reducible	$t\bar{t}$+jets, V+jets, Single-top		Data-Driven	6

Table 6.3: Classification of processes used in the analysis and information associated to their definition and estimation. q/g stands for quark/gluon.

Electron charge mis-identification (QMID) background

In the $t\bar{t}t\bar{t}$ analysis only the charge mis-identification of electrons is considered. Due to their higher mass, muons produce significantly less bremsstrahlung. In addition, the ATLAS detector has the MS for precise muon momentum measurements. In combination, these two factors render charge mis-identification for muons negligible. The QMID background is only considered for the $e^\pm e^\pm$ and $e^\pm\mu^\pm$ final states in the 2LSS channel and mostly originates from opposite-charge $t\bar{t}$+jets events, which fall into this category due to mis-identification of one of the electrons. The two most important scenarios for charge mis-identification of electrons are:

- High-p_T electrons: The electron charge can be determined from the direction of the curvature of the flight path of electrons under the influence of the detector's magnetic field. Higher-p_T electrons have less curvature in the detector volume, creating difficulties in correctly determining the direction of curvature of the track. When the direction of curvature is incorrectly determined, the charge of the electron is mis-identified.
- Trident process: As defined in the previous section, this process results in an increase in the electron multiplicity. Since the rate of bremsstrahlung is related to the amount of material along the electron's flight path, the trident process takes place more frequently in the forward regions (larger $|\eta|$) of the detector. Among the three final state electrons, a charge mis-identification

occurs, if only the one with the opposite-charge to the original incident electron is reconstructed by the detector (see Figure 6.1).

The QMID background is estimated using a data-driven method, described in Section 6.3.3.

Photon conversion backgrounds

When photons decay and produce an electron–positron pair, leptons are produced without an actual incident electron. Two types of photon conversion backgrounds are distinguished in the $t\bar{t}t\bar{t}$ analysis:

- Material Conversion (Mat. Cv.): Events with photons that decay into an electron–positron pair within the detector volume, induced by material interaction, are called material conversion events. Electrons decaying from material conversion originate from a decay vertex located inside the detector, called the conversion vertex (CV). The CV can then be identified as a displaced vertex and defined as the point where the track associated to the electron and the track closest to it in the ΔR plane have the same ϕ value. CV and PV are sketched in Figure 6.2 together with a electron–positron pair emerging from each vertex.

- Virtual Conversion (γ^*) [141]: Another case of a photon producing an electron–positron pair happens inside the beam pipe before entering into the detector volume. Electrons emerging from this conversion process are features of an actual physics process rather than an instrumental effect as in the case of material conversions. Virtual photon conversions originate from the PV.

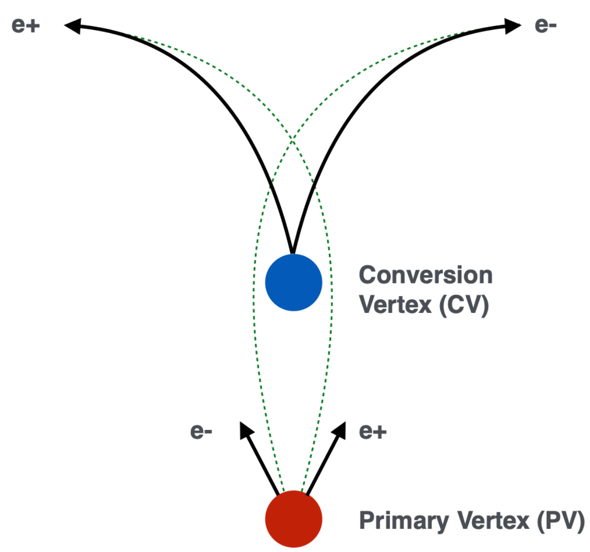


Figure 6.2: Sketch of the primary vertex (PV) and the conversion vertex (CV) used in the context of photon conversions. The solid black arrows represent the reconstructed electrons. The dashed green lines represent the tracks extrapolated to the PV that originate from electrons associated to the CV.

Photon conversion backgrounds are estimated using a data-assisted template-fit method described in Section 6.3.3.

Heavy-flavour non-prompt background

In the case of a non-prompt lepton from heavy-flavour (bottom or charm) hadron decays, outgoing electrons and muons are distinguished and labelled as HF_e and HF_μ, respectively. They are estimated using a data-assisted template-fit method described in Section 6.3.3.

Light-flavour fake background

One category is defined for events where a light-jet is wrongly identified as a lepton. Such leptons were already introduced in Section 6.3.2, since in comparison to non-prompt leptons, these events do not contain real leptons. Due to their low rate, leptons originating from light hadrons or quark/gluon jets are not distinguished by flavour, and their contribution is estimated directly from MC simulations. These light-flavour fake contributions (LF) are included in the *Others* category.

6.3.3 Estimation of backgrounds

Background processes that are not estimated directly from the predictions of MC simulation are determined using two different methods. QMID is estimated using a fully data-driven approach. The remaining processes are calculated using a data-assisted method called the *template fit*. In this method, the template refers to the shape of the distribution of the background process of interest, which is taken directly from its MC simulation prediction and thus not data-driven. On the other hand, the normalisation of the process is left as a free NF, to be determined by a fit to data. A CR enriched in the target background process is included in the fit in order to help constrain the varying NF. There are multiple systematic uncertainties associated with both background estimation methods. The different origins of systematic uncertainties are discussed in the Section 6.5. The results of the estimation, encapsulated in the numerical values of the NFs, are discussed along with the fit setup in Section 6.6.

CR Conv.

The CR Conv. simultaneously targets virtual photon conversion and material photon conversion backgrounds together, which makes up about 40% of the event yield in this region. As such, constraining the two backgrounds in one region requires their template shapes to be different. This is achieved by using the variable m_{ee}^{PV} which is the invariant mass of the converted electron-positron pair at the PV. It is calculated by using the track associated to the electron[3] and the track closest to it at the PV. This is sketched in Figure 6.2.

The virtual photons originating from the PV, will have an invariant mass m_{ee}^{PV} close to zero. Additionally, for material conversion, the CV lies within the detector volume, separated from the PV. The invariant mass defined at the CV using these two tracks is called m_{ee}^{CV}. In the case of material conversions, the extrapolation of the associated tracks to the PV, whilst the correct vertex being the CV, will induce a larger invariant mass of m_{ee}^{PV}. Thus, m_{ee}^{PV} will have lower values for virtual conversions and larger values for material conversions. As such, this variable has the desired property of having a shape difference that can distinguish between both processes.

[3] In this context, the term “electron” is used for both electrons and positrons.

The CR selection requires an m_{ee}^{CV} below 0.1 GeV in order to increase the fraction of electron-positron pairs originating from a massless particle, i.e. photons. Only $e^{\pm}e^{\pm}$ or $e^{\pm}\mu^{\pm}$ dilepton events satisfying $200 < H_{\mathrm{T}} < 500$ GeV are considered with four or five jets, where at least one of them is b-tagged. This CR is used to constraint the $\mathrm{NF}_{\mathrm{Conv.}}$ parameter of the fit.

Charge misidentification background

A data-driven estimation of the QMID contribution aims at estimating the probability of mis-identifying the electron charge, parametrised in p_{T} and $|\eta|$. For the CR Conv. the m_{ee}^{PV} variable is also used. The charge-flip rates are derived using a selection of events consistent with the $Z \to ee$ process in a data sample. This is achieved by defining an invariant di-electron mass distribution window within ±10 GeV of the Z-boson mass, where no condition on particle charges is required. A sideband subtraction method [142] is used in the region outside of the Z-boson mass window ($|m_{ee} - m_Z| > 10$ GeV) to remove events in the Z-boson resonance peak region. Indexing the first (second) electron as i (j) according to the two dimensional (p_{T}, $|\eta|$) bin it falls into, the charge mis-identification rates $\epsilon_{i,j}$ relate the total number of measured events ($N_{i,j}$) to the total number of same-charged events ($N_{i,j}^{\mathrm{SS}}$) through

$$N_{i,j}^{\mathrm{SS}} = N_{i,j}[\epsilon_i(1-\epsilon_j) + \epsilon_j(1-\epsilon_i)]. \tag{6.1}$$

The first summand is the multiplication of the rates for the first electron being charge-flipped and the second being correctly measured and vice-versa for the second summand. This is because a final state is wrongly identified from opposite-sign to same-sign only if one electron's charge is flipped. The rates $\epsilon_{i,j}$ are calculated by using $N_{i,j}^{\mathrm{SS}}$ and maximising the Poisson likelihood $\mathcal{L}(\epsilon_{i,j}|N_{i,j}^{\mathrm{SS}}, N_{i,j})$ according to the total number of events $N_{i,j}$. Charge mis-assignment rates increase with growing p_{T} and $|\eta|$ values, spanning several orders of magnitude between 0.002% to 4%.

After having determined the charge mis-assignment rates, the expected contribution of the QMID background in a given 2LSS channel analysis region is estimated in two steps. First, an orthogonal region is defined by only inverting the same-sign requirement of the original region. In this opposite-sign region, each data event is weighted by an event weight w, defined as

$$w = \frac{\epsilon_1 + \epsilon_2 - 2\epsilon_1\epsilon_2}{1 - \epsilon_1 - \epsilon_2 + 2\epsilon_1\epsilon_2}. \tag{6.2}$$

The event weight gives the reconstruction probability of an opposite-sign dileptonic final state as a same-sign dileptonic final state. The sum of weighted events provides the QMID background yield in the region.

CR HFe

This CR is enriched with non-prompt electrons from decays of heavy-flavour jets, constraining the $\mathrm{NF}_{\mathrm{HF}_e}$ normalisation factor for the HF_e background contribution in the fit. The region is defined in trilepton channels with at least two electrons: eee or $ee\mu$. Exactly one b-tagged jet is required without any restriction on the total number of jets. A medium range of 100–250 GeV for the H_{T} selection completes the definition of the region. This selection attempts at capturing the dileptonic $t\bar{t}$ decay events with an additional non-prompt electron. Unlike the conversion region discussed previously, CR HFe does not employ a distribution, but consists of a single bin counting the events passing the selection. The obtained HFe fraction in this region is about 40% with the second highest contribution coming from "Others" with a fraction of about 25%.

CR HFm

This CR is similar to the previously defined CR HFe in structure and design. However, instead of electrons, it targets the non-prompt muons coming from the decays of the heavy-flavour jets, and therefore is constraining the $\mathrm{NF}_{\mathrm{HF}_\mu}$ normalisation factor for the HF_m background. The region is defined in trilepton channels with at least two muons: $e\mu\mu$ or $\mu\mu\mu$. Exactly one b-tagged jet is required without any restriction on the total number of jets. The requirement $100 < H_\mathrm{T} < 250$ GeV completes the definition of the region. The CR HFm consists also of a single bin counting the events passing the selection. The achieved HFm fraction in this region is about 50% with the second highest contribution coming from "Others" with a fraction of about 20%, thus having a purity slightly higher than CR HFe.

CR $t\bar{t}W$

This CR is different from the others as it targets the $t\bar{t}W$ process, which is an irreducible background. The normalisation of the $t\bar{t}W$ background process is left as a free parameter in the fit due to previous knowledge as to the underestimation of this process in the simulations (see Section 3.3). The CR constrains the $\mathrm{NF}_{t\bar{t}W}$ normalisation factor in the fit and is defined in the dileptonic channels excluding the ee channel in order to suppress QMID contributions. Two additional selections are applied to further reduce the virtual and material photon conversion backgrounds:

- $m_{ee}^{\mathrm{CV}} > 0.1$ GeV: As explained in the description of the conversion CR above, leptons from photon decays are expected to have lower invariant masses at the Conversion Vertex.

- $|\eta(e)| < 1.5$: The larger the η region, the larger the material which the electrons traverse, increasing the likelihood of conversion due to material interaction. To reduce this effect a central $|\eta|$ region is selected.

In addition, at least four jets and at least two b-tagged jets are required. As this selection overlaps with the SR selection for events with at least six jets, orthogonality is achieved by requiring an H_T below 500 GeV for events with exactly two b-tagged jets. In the case of events with at least three b-tagged jets, those that fail to pass this H_T selection are still accepted if they have less than six jets. The sum of the lepton p_T ($\Sigma p_\mathrm{T}^{\ell}$) variable is used as the discriminating distribution for this CR. The achieved $t\bar{t}W$ fraction in this region is about 33% with second and third highest contributions coming from $t\bar{t}Z$ and $t\bar{t}H$ backgrounds with fractions of about 20% and 15%, respectively. It is in general difficult to disentangle $t\bar{t}W$ from these two other irreducible background contributions. Nevertheless, unlike the disjoint shapes of two photon conversion processes in CR Conv., here, $t\bar{t}W$ dominates the full range of the CR distribution and thus could still be used. The properties of the regions used in the fit setup are summarised in Table 6.4. In this section only the CRs are described. The SR is detailed in the next section.

Region	NF	Channel	N_j	N_b	H_T [GeV]	Other Requirements	Distribution
SR	-	2LSS/ML	≥ 6	≥ 2	> 500	-	BDT score
CR Conv.	$\mathrm{NF_{Conv.}}$	$e^{\pm}e^{\pm} \| e^{\pm}\mu^{\pm}$	$4 \leq \mathrm{N}_j < 6$	≥ 1	$200 < H_T < 500$	$m_{ee}^{\mathrm{CV}} \in [0{,}0.1\ \mathrm{GeV}]$	m_{ee}^{PV}
CR HF$_e$	$\mathrm{NF_{HF_e}}$	$eee \| ee\mu$	-	$= 1$	$100 < H_T < 250$	-	-
CR HF$_\mu$	$\mathrm{NF_{HF_\mu}}$	$e\mu\mu \| \mu\mu\mu$	-	$= 1$	$100 < H_T < 250$	-	-
CR $t\bar{t}W$	$\mathrm{NF}_{t\bar{t}W}$	$e^{\pm}\mu^{\pm} \| \mu^{\pm}\mu^{\pm}$	≥ 4	≥ 2	-	$m_{ee}^{\mathrm{CV}} \notin [0{,}0.1\ \mathrm{GeV}]$, $\lvert\eta(e)\rvert < 1.5$	Σp_T^{ℓ}
						For $N_b = 2$, $H_T < 500$ GeV || $N_j < 6$	
						For $N_b \geq 3$, $H_T < 500$ GeV	

Table 6.4: Names and properties of regions used as signal and control regions in the profile-likelihood fit. Abbreviations are: normalisation factor (NF), jet multiplicity (N_j), b-tagged jet multiplicity (N_b), conversion vertex (CV), primary vertex (CV), Boosted Decision Tree (BDT).

6.3.4 Validation of background estimations

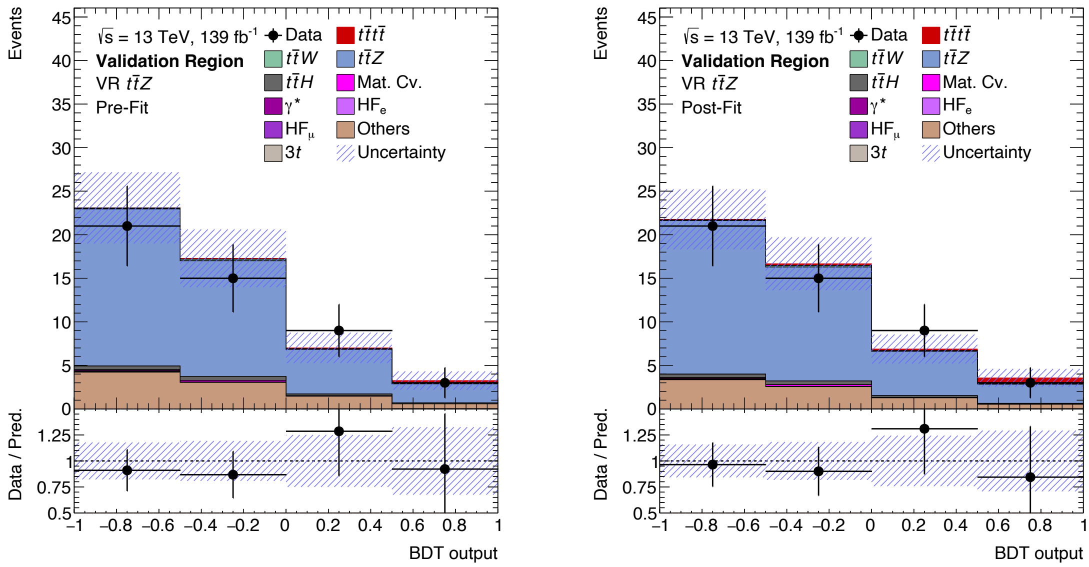
Two validation regions are defined in order to verify the modelling of the MC prediction for the two dominant background sources of the analysis, namely the $t\bar{t}Z$ and $t\bar{t}W$ processes. A good agreement in the validation regions is important to have confidence in the assumption that estimates done in CRs can be extrapolated to other regions. In this section, pre- and post-fit data–MC comparisons are shown for two validation regions for the sake of clarity, while, the fit setup is introduced in Section 6.6.

$t\bar{t}Z$ VR

The $t\bar{t}Z$ validation region is defined using a similar selection to the SR in the trilepton channel by requiring $H_T > 500$ GeV, at least six jets with at least two of them being b-tagged. Orthogonality to the SR and CRs is established by requiring at least one opposite-sign same-flavour lepton pair with an invariant mass within 10 GeV of the Z-boson. This reverts the Z-veto applied in the SR and all CRs. The invariant mass condition enriches the region in $t\bar{t}Z$ events by selecting leptons that are likely to come from the decay of a Z-boson. In Figure 6.3(a), the $t\bar{t}Z$ validation region shows good agreement between data and simulation already prior to the fit. The post-fit agreement in Figure 6.3(b) shows only minor changes in the distribution. Therefore, the NPs related to the $t\bar{t}Z$ process are not expected to be disrupted by the fit nor pulled.

$t\bar{t}W$ VR

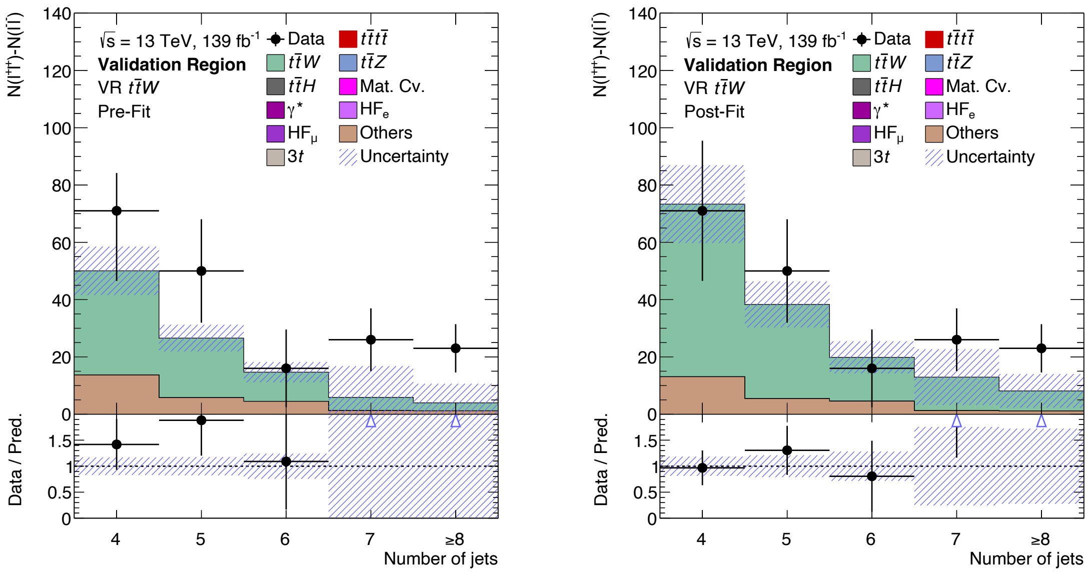
A validation region dominated by $t\bar{t}W$ events is achieved by taking advantage of the charge-asymmetric nature of this process in proton–proton collisions at the LHC. The charge asymmetry is defined by the difference between events with at least two positively charged same-sign leptons ($N_+ \in \ell^+\ell^+$, $\ell^+\ell^+\ell^+$, $\ell^+\ell^+\ell^-$) and events with at least two negatively charged same-sign leptons ($N_- \in \ell^-\ell^-$, $\ell^-\ell^-\ell^-$, $\ell^-\ell^-\ell^+$). A selection of at least four jets, with at least two of them being b-tagged, is applied. This VR is not orthogonal to the other regions, however largely deprived of signal. In regions with high jet multiplicities, it is observed that the MC simulation predicts a lower estimation compared to data. Based on the difference between data and prediction in the last two bins of Figure 6.4(a), an additional uncertainty of 125% for events with seven jets, and additional uncertainties of 300% for events with at



(a) Pre-fit (b) Post-fit

Figure 6.3: Pre- and post-fit BDT output score distributions for the $t\bar{t}Z$ VR using the data fit. Shaded bands represent both, statistical and systematic uncertainties.

least eight jets are introduced. Pre-fit and post-fit comparisons are shown in Figure 6.4, where these two additional uncertainties are included among all other statistical and systematic uncertainties.



(a) Pre-fit (b) Post-fit

Figure 6.4: Pre- and post-fit number of jets distributions for $t\bar{t}W$ VR using the data fit. Shaded bands represent both, statistical and systematic uncertainties.

6.4 Multivariate signal extraction

A BDT discriminant is used to define the SR of the analysis. In this section, the three steps in the development of the discriminant are described. The second step is the optimisation of the discriminant by studying the impact of various hyperparameters and variables on the performance of the discriminator. The first and third steps are pre- and post-optimisation steps, respectively. An MVA architecture as well as an initial set of selections and inputs should be chosen before an optimisation can be done. These are categorised under pre-optimisation studies. After the optimisation, the stability of the performance and its generalisability need to be validated, which are grouped under the post-optimisation checks. Comparisons with the corresponding analysis by the CMS collaboration [69] and the MVA discriminant used therein are provided where possible, in order to highlighted the differences and similarities between the two approaches. For a more in-depth description of machine-learning and statistics related concepts and methods mentioned in this Section, the reader is referred to Chapter 5.

6.4.1 Pre-optimisation studies

Signal region selection

Three variables are used in the definition of the signal region with their values selected such as to maximize the sensitivity defined by the ratio of signal yield to the square root of the total background yield. Two out of three variables are related to jet and b-tagged jet multiplicities. In the 2LSS channel, 8 jets are expected in the final state, whereas in the trilepton final state 6 jets are expected. Final states of at least six jets are therefore expected to make up a large fraction of the signal. In both categories, four b-tagged jets are expected. However, possible losses due to the b-tagging efficiency, a sizeable fraction of four-top-quark events will be reconstructed with less than four b-tagged jets. In order to include such events, a looser b-tagging requirement is preferred. The signal in events with one b-tagged jet is dominated by background. Based on these motivations, at least two b-tagged jets are required at the 77% WP. The third and final variable is H_{T}. Simultaneously producing four top-quarks results in a crowded final state with many jets, and at least two leptons in the final states considered in this analysis. This yields a large H_{T} value that can distinguish the signal process from backgrounds. As discussed in Section 3.4, due to this property, H_{T} distributions were used in the SRs in the previous ATLAS measurement. Based on optimisation studies, an H_{T} value of 500 GeV is selected to finalize the SR definition. It is important to note that the SR selection defined above is tighter in comparison to the most recent CMS analysis in this final state, where the H_{T} selection cut is 300 GeV and at least six jets are required. Preliminary studies with BDTs showed that, using the tighter selection above as the training region leads to a better performing MVA discriminant, compared to using the looser selection preferred by the CMS Collaboration's analysis.

In terms of lepton multiplicities, the analysis presented in this thesis has one inclusive signal region. The possibility of having more than one SR and an individual BDT discriminant for each SR was considered but found not to be useful. In the extreme case, each of the seven leptonic channels could be taken as an individual SR. Due to reduced sample sizes available in the per channel training, this option was discarded. A more inclusive but still physically motivated option was to have two SRs: 2LSS and ML. In this case, studies showed that the combined performance of two separate BDT discriminants is less powerful than the inclusive one. Including the lepton multiplicity information explicitly as a variable in the training did not result in increased performance either. This variable did not rank high in importance. This is in contrast with the findings of the CMS analysis, where the lepton multiplicity is found to be the 3rd highest ranking variable [69]. The inclusive region's robustness in the ATLAS analysis can be attributed to the implicit availability of the lepton multiplicity information through other input variables. The minimum distance $\Delta\mathrm{R}(\ell,\ell)_{\mathrm{min}}$ between any lepton pair, for example, is a very high ranking variable in importance and is larger (smaller) for dilepton (trilepton) final states. The SR selection used as the MVA training region is summarised in Table 6.5 for both analyses. It should be noted, that the CMS analysis also had a single inclusive SR and the object reconstruction methods and working points efficiencies differ between the two analyses. As an example, for the b-tagging, the ATLAS analysis employs the MV2c10 algorithm at 77% WP, whereas the CMS analysis used the DeepCSV algorithm with an identification efficiency of 55–70%.

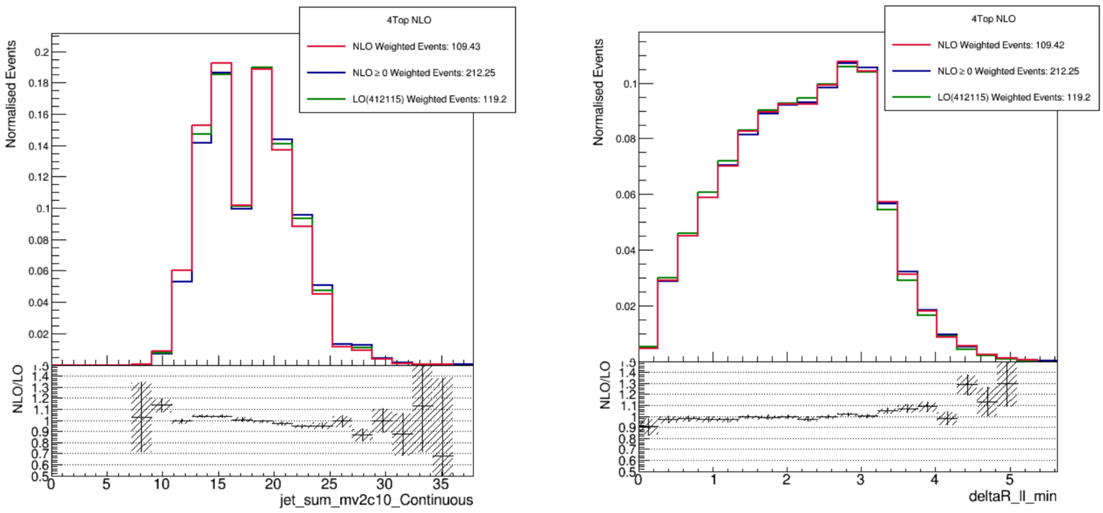
Region / Variable	H_T [GeV]	N_j	N_b	Leptons
ATLAS SR	>500	≥6	≥2	2LSS || 3L
CMS SR	>300	≥2	≥2	2LSS || 3L

Table 6.5: Main signal region selection criteria used in the ATLAS and CMS analyses. The object reconstruction methods and working point efficiencies are different between the two analyses.

Sample selection, classification and splitting

The BDT trainings are done using the LO signal sample against the sum of all background processes as predicted by the simulation, i.e. without a data-driven QMID estimation and using the pre-fit prediction. At the time of the BDT studies, the data-driven QMID estimation had not been finalised. Due to its small size, the impact of missing data-driven corrections on the training outcome is negligible. The decision to use the LO prediction instead of the more accurate NLO prediction was motivated by the presence of negative-weight events in the NLO sample. Such events could destabilise the training process as they could lead to unbounded loss function values which in return causes the training process to diverge. As both, the LO and the NLO samples are available, the LO sample is preferred to the NLO sample without any negative-weight events. A comparison between LO and NLO signal sample distributions, along with the NLO signal sample distribution with positive-weight events only, shows that the kinematics are not significantly different. Two example distributions are shown in Figure 6.5 using variables utilized for the BDT training, albeit without the SR selection applied. These variables are defined in the next sub-section. Further distributions are available in Reference [143]. The final fit uses the full NLO prediction of the signal sample.

As alternatives to a binary, signal-versus-total-background training architecture, three other models were tested. In the first one, the $t\bar{t}W$ process is separated and defined as an individual background output class and all remaining backgrounds are classified together. This signal versus $t\bar{t}W$ and remaining backgrounds training was motivated by the dominant role of the $t\bar{t}W$ background process in the analysis. The second model also comprises two background classes to improve the targeting of irreducible backgrounds: $t\bar{t}W$, $t\bar{t}Z$ and $t\bar{t}H$ in one class and remaining background processes in another



(a) Sum of the MV2c10 pseudo-continuous b-tagging score over all jets ($\Sigma w_{\mathrm{MV2c10}}$)

(b) The minimum distance between any lepton pair ($\Delta R(\ell,\ell)_{\min}$)

Figure 6.5: Distribution of two BDT training input variables for LO, NLO and positive-weight-only NLO $t\bar{t}t\bar{t}$ signal samples. No SR selection has been applied to the samples. Plots are taken from Reference [143].

class. The third setup splits the backgrounds in four classes with the three previously mentioned irreducible processes each becoming a class, and the remaining background processes again collected under a single class. These models all underperformed compared to the original binary classification setup. A training distinguishing the $3t$ process is also desirable, however, due to the small sample size it could not be properly tested.

To check against overtraining and to quantify the performance, three concepts and three corresponding subsets of samples are distinguished. First, 20% of the NLO signal sample and the background samples are spared for later use and not used in any of the optimisation steps. The remaining 80% of the background sample is then divided into two samples with odd and even event numbers, creating two orthogonal subsets each with 40% of the total events. The remaining 80% of the NLO signal sample are set aside without further splitting. In conclusion, 100% of the LO signal sample is used in training, 80% of the NLO signal sample is used in the testing and 20% of the NLO signal sample is used in the validation. A schematic summary of the strategy described here is shown in Figure 6.6.

The number of raw simulated events used in the BDT setups is listed in Table 6.6. Reweighted events corresponding to similar numbers of signal and background events are fed to TMVA in the training. As a result, the ratio of signal-to-background events used in the training is scaled to approximately 1:1. The reweighting uses the default TMVA method *EqualNumEvents*, which reweights events in two steps. First, events of the signal sample are reweighted such that the weight of each event on average becomes equal to one. Using this reweighted signal sample as a reference, in the second step, a scaling is applied to the background sample to make its total weight equal to the total weight of the signal sample. Each background event is then reweighted based on this constraint.

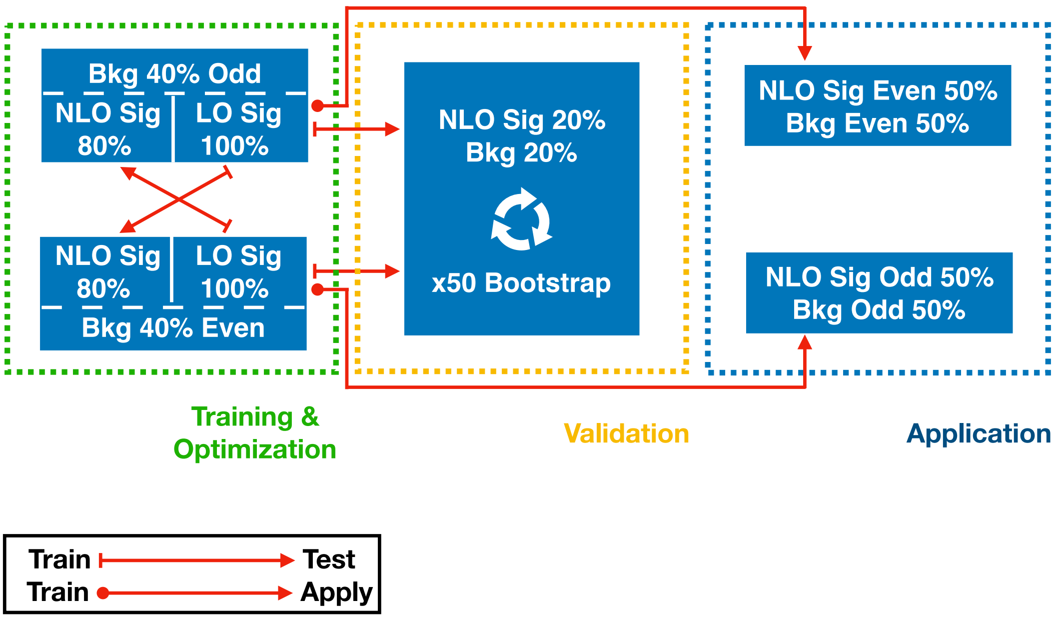


Figure 6.6: The schematic representation of sample fractions and splitting using the training–testing–validation method. Sig (Bkg) refers to signal (background) samples. Even (Odd) refers to even (odd) numbered events, respectively.

Sample	Type	Odd	Even
Signal	Training	282449	282449
Background	Training	45550	46414
Reweighted	Training	164000	164432

Table 6.6: Number of MC simulated events used in different BDT trainings. Odd (Even) refers to the event number being odd (even).

Weighted event yields used in the BDT trainings are reported in Table 6.7. It should be noted that negative-weight events in NLO samples are not used in the training, causing an increase in the yields of the background processes. The application of the BDT training response to the observed data events is discussed in the next Section.

Samples	Odd	Even
LO Signal 100%	100%	100%
$t\bar{t}t\bar{t}$	21.14	21.14
NLO Backgrounds 80% (w>0)	40%	40%
$t\bar{t}W$	25.45	26.98
$t\bar{t}Z$	40.82	41.33
$t\bar{t}H$	18.67	18.91
QMID	3.71	5.00
Mat. Cv.	5.19	5.32
$\gamma*$	1.97	2.02
HF_e	2.40	2.87
HF_μ	2.54	3.04
Others	12.42	10.14
$3t$	1.13	1.12
Total Background	114.30	119.46

Table 6.7: Weighted yields of samples used in the BDT training. Odd (Even) refers to event number being odd (even). “w>0” indicates that only positive-weight events are used.

Choice of initial algorithm parameters and figure of merit

The GradBoost and AdaBoost algorithms, as they are implemented in the TMVA and described in Section 5.2.2, have been tested with a preliminary setup using a few trainings. No significant differences in the discrimination performance have been found in these studies. Due to the slightly faster convergence rate and the possible prospect of testing other loss functions, the GradBoost algorithm is chosen. The TMVA default, binomial log-likelihood loss function provided a satisfactory performance. Node splitting decisions are made by maximising the Gini Index at each node.

The remaining parameters of the model are not fixed in the beginning, but left open to be determined in an optimisation process. These parameters are: Number of trees (nTrees), maximum tree depth (MaxD), minimum fraction of events in a leaf (nMin%), learning rate (Shrinkage), granularity (nCut) and bagging rate. The minimum and maximum values for each parameter are defined, using a preliminary training setup. Additional trainings are performed with some relatively large and small values for each parameter (compared to a baseline setup with average values for each parameter). The minimum and maximum ranges for each parameter, beyond which the performance is found not to be improving significantly are determined and shown in Table 6.8.

nTrees	MaxD	nMin%	Shrinkage	nCut	Bagging
200	2	1	0.01	15	0.4
400	3	3	0.02	20	0.5
600	4	4	0.05	25	0.6
800	5	5	0.10	30	0.7
1000	6	9	0.20	35	0.8

Table 6.8: Grid-scan points for different BDT training hyperparameters. For the definitions of the hyperparameters, the reader is referred to Section 5.2.2.

The performance has been checked using the arithmetic average of even and odd testing BDT ROC curve integrals i.e. the area under the curve (AUC):

$$\mathrm{ROC} = \frac{\mathrm{AUC}_{\mathrm{test}}^{\mathrm{even}} + \mathrm{AUC}_{\mathrm{test}}^{\mathrm{odd}}}{2}. \tag{6.3}$$

The performance of the model as quantified by the validation set ROC value is compared to the expected significance of a preliminary fit setup where the same model is applied. This is repeated for several trainings where it is ensured that performance of each vary noticeably. An approximately linear positive correlation is observed between ROC curve values and the expected significances, confirming the ROC curve value as a valid proxy for the performance of the fit. Details of the optimisation process and its outcomes are described in the next section.

Choice of initial input variables

More than 60 variables such as kinematics and multiplicities of leptons and jets are initially considered as input variables. Signal region MVA discriminant input variables used in the analysis by the CMS collaboration [69] are also considered. Several variables using Fox-Wolfram moments [144, 145] for jets are included for testing. A BDT is trained with average hyperparameter settings without any optimisation or further elaboration. Variables are ranked according to their impact on the performance in the training, where the ranking of the variable is reported by TMVA. It is calculated by counting how often a variable is used for the node split decisions (k) and weighing this number by the square of the separation gain (S) defined in Equation 5.27 and the number of events in the node (N):

$$\text{Ranking score} = \Sigma_{i=1}^{k} S_i^2 N_i. \tag{6.4}$$

The ranking score is not equal to the integer number of the rank itself. The rank is determined by ordering the variables using the ranking score.
After each training, the 5 lowest ranking variables are dropped and the training is repeated. If the

performance is not decreased significantly in the new training, dropped variables are no longer considered. The procedure is repeated until performance degradation is observed. For the set of 5 variables that leads to poorer performance, each member of the set is individually examined. This is done by training five individual BDTs with each of them lacking one of the five variables from the set. If a training in which a variable is missing, performs as good as the training with all 5 variables included; then this variable is deemed redundant and also removed. A final set of 21 variables is reached as a result. The ranking and properties of these variables are reported in Table 6.9.

Rank	Variable	Category	Description	IR
1	Σw_{MV2c10}	b-tagging	Sum of MV2c10 pseudo-continuous b-tagging score over all jets	✔
2	$\text{p}_{\text{T}}^{\ell_0}$	Lepton	p_{T} of leading lepton	✔
3	$\text{E}_{\text{T}}^{\text{miss}}$	Energy	Missing transverse energy	✔
4	$\Delta R(\ell,\ell)_{\text{min}}$	Distance	Minimum distance between any lepton pair	✔
5	$p_{\text{T}}^{\text{jet}_5}$	Jet	p_{T} of 6$^{\text{th}}$ leading jet	✔
6	$\Delta R(\ell,b)_{\text{max}}$	Distance	Maximum distance between leptons and b-tagged jets	✔
7	$H_{\text{T}}^{\text{no lead jet}}$	Energy	Scalar sum of all lepton and jet p_{T} except the leading jet	✔
8	$\Sigma\Delta R(\ell,\ell)_{\text{min}}$	Distance	Sum of the distance between leading and sub-leading leptons in SS channel or leading, sub-leading and third-leading leptons in 3ℓ channel	✔
9	$m_{\text{jet}}/p_{\text{T}}^{\text{jet}}$	Event	Jet mass divided by p_{T} for the highest ratio	✖M
10	$\Delta\phi(\ell_0, j_0)$	Distance	Transverse angle between leading lepton and jet	✖
11	$p_{\text{T}}^{\text{jet}_0}$	Jet	p_{T} of leading jet	✔
12	$\Delta R(j,b)_{\text{min}}$	Distance	Minimum distance between b-tagged jets and jets	✔
13	$\Delta R(\ell,j)_{\text{min}}$	Distance	Minimum distance between leptons and jets	✖
14	$p_{\text{T}}^{b\text{-jet}_0}$	Jet	p_{T} of leading b-tagged jet	✔
15	$\Delta R(\ell,b)_{\text{min}}$	Distance	Minimum distance between leptons and b-tagged jets	✖
16	$p_{\text{T}}^{\ell_1}$	Lepton	p_{T} of sub-leading lepton	✖
17	$p_{\text{T}}^{\text{jet}_2}$	Jet	p_{T} of third-leading jet	✖
18	$p_{\text{T}}^{\text{jet}_1}$	Jet	p_{T} of sub-leading jet	✔
19	nJets	Jet	Number of jets	✖M
20	nLeps	Lepton	Number of leptons	✖
21	$p_{\text{T}}^{\ell_2}$	Lepton	Transverse momentum of third-leading lepton	✖

Table 6.9: Initial set of variables used during the optimization studies. The rank is reported by TMVA and its calculation is described in Section 6.4.1. The last column refers to results of a variable optimization procedure. Two variables with (M) are removed due to mismodeling issues. The sum of the MV2c10 pseudo-continuous b-tagging score is built using the integer sum of the b-tagging score intervals (see Section 4.3.6).

6.4.2 Optimisation of BDT discriminant

The BDT discriminant is optimised in two steps. In a first step, a grid-scan of six selected hyperparameters is performed with five selected values for each of the parameters. Secondly, setups with the fixed hyperparameters from the previous step are used to probe the power of individual variables. In this way, redundant variables are removed from the setup.

Optimization of hyperparameters

For each of the hyperparameters, five values are scanned. These values are defined based on the previous studies, such that the central value corresponds to a roughly above-average performance. The two upper and lower variations are then understood as the scan of the region in the vicinity. Parameters and their chosen ranges are summarised in Table 6.8.

The list of variables used in this optimisation step is described in Table 6.9. There is no order or hierarchy in the way hyperparameters are combined. An individual setup of the grid-scan is composed of one value from each column. All combinations of values have been processed. In total, $6 \times 3125 = 15625$ different BDT setups have been processed. A summary plot of these trainings is shown in Figure 6.7, where each point corresponds to one unique BDT training setup. The y-axis shows the corresponding ROC value of the setups. Since many trainings yielded similar performances, the number of dots appearing is smaller than the total number of trainings. Furthermore, the trainings where a difference of more than 0.05 in the ROC curve integral values between the Odd and the Even testing responses are ignored, since those are considered to be overtrained.

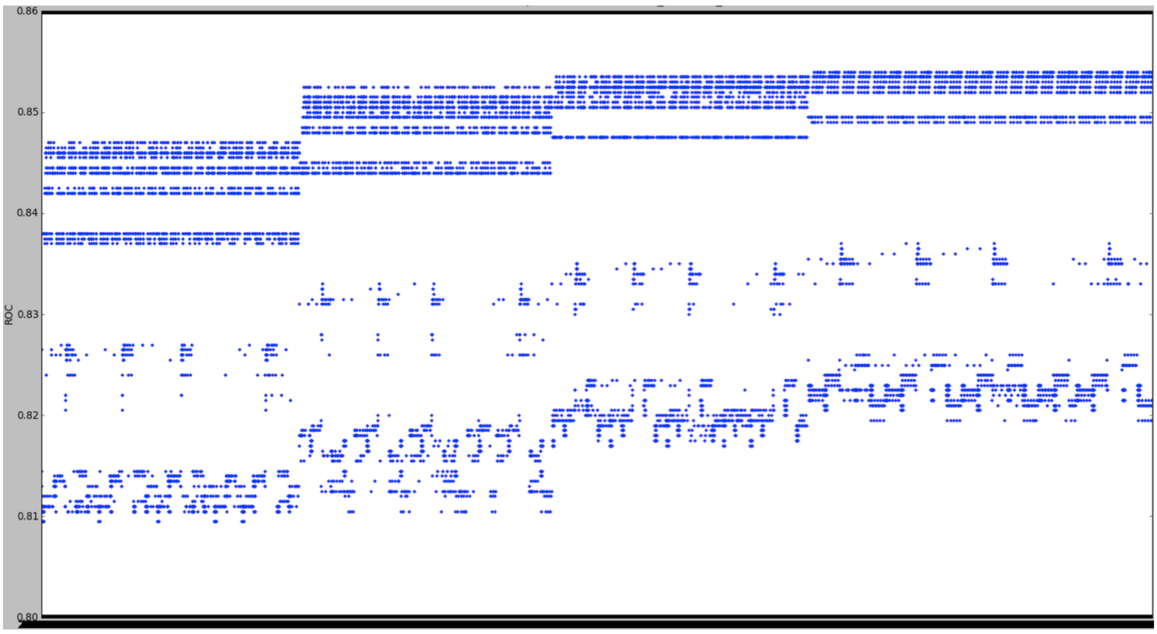


Figure 6.7: Summary plot of the BDT hyperparameter optimisation studies. Each point corresponds to one unique BDT training setup. The y-axis shows the corresponding ROC value of the setups. Trainings that yielded very close ROC values are overlayed and are not distinguishable.

The final discriminant returns an average ROC integral value of 0.854. The chosen technical parameters and the optimised hyperparameter settings are listed in Table 6.10. For three parameters (MaxD, Shrinkage, nCut) the optimal choice is either at the maximum or minimum value of their respective ranges. Thus, in these cases training is repeated for each of these parameter by using the value one-step-size beyond the determined value. No improvement over the original choices have been observed. Among the parameters reported by the CMS collaboration in their analysis, the choice of the Gradient Boosting method is common, whereas ATLAS analysis uses more (400 vs. 600) and deeper (4 vs. 6) trees.

Parameter	Value
ROOT Version	6.10.04
Separation Gain	Gini
Boosting	GradBoost
Loss-function	Binomial log-likelihood
Signal Region	$H_{\mathrm{T}}^{\mathrm{all}} > 500$ & $N_{\mathrm{jets}} \geq 6$ & $N_{b\text{-jets}} \geq 2$ & (2LSS || 3L)
nTrees	800
MaxD	6
nMin%	3
Shrinkage	0.01
nCuts	15
Bagging	0.7

Table 6.10: Final settings of the optimised BDT model.

Optimisation of variable selection

Using the determined set of BDT hyperparameters, an optimization of the variables is studied. It is often observed that some variables of the BDT can be removed without much loss of separation power. To test the effect of individually including or excluding variables on the overall BDT performance, the Iterative Removal (IR) method [146] is used.

For the IR method, a fixed BDT setup is trained repeatedly, where in each iteration one variable is removed from the setup. The resulting ROC value is then compared to the nominal setup with the said variable. The performance is then evaluated with a 1% threshold for the loss. In this study a stricter criterion is used by setting the threshold to 0, i.e. only variables, whose removal do not cause any degradation in the performance, are kept. By this approach 7 variables were removed and the total number of variables has been reduced from 21 to 14.

The most important variable is found to be the b-tagging related $\Sigma w_{\mathrm{MV2c10}}$. It is defined as the "sum of MV2c10 pseudo-continuous b-tagging score over all jets". The variable is built by defining five efficiency WPs (see Section 4.3.6) of the MV2c10 algorithm. An integer is assigned to each range. The associated ranges and scores are listed in Table 6.11. Every jet in the event is then

Minimum of WP	Maximum of WP	Score
0%	60%	1
60%	70%	2
70%	77%	3
77%	85%	4
85%	100%	5

Table 6.11: Working point (WP) ranges and corresponding scores used in the definition of the $\Sigma w_{\mathrm{MV2c10}}$ variable. The reader is referred to Section 4.3.6 for the description of WPs.

assigned a score. Then all numbers from all jets are summed and their total gives the variable's value. The *pseudo-continuous* in the definition is due to the fact that the variable uses ranges defined by the five calibrated WPs. In this sense each jet is assigned a b-tagging score in a *continuous* manner. However, the WPs are actually limited to five and are discrete, and therefore this continuity is a "*pseudo-*" continuity. An example distribution is shown in Figure 6.8(a).

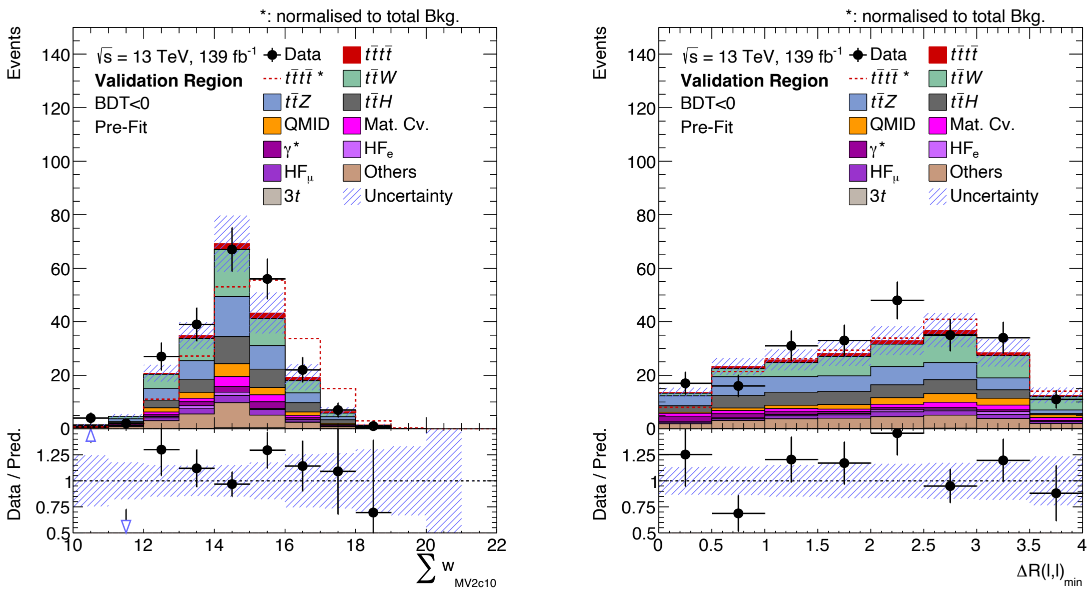
6.4.3 Post-optimisation checks

After the optimisation of the MVA model, several checks have been made in order to validate the performance against several possible causes of error.

Modelling of input variables and discriminant output

This analysis is conducted *blinded*, that is, data is not used during the optimisation of the whole analysis setup in the parts of phase-space where the expected fraction of the signal is relatively high. This is done to avoid conscious or unconscious attempts to tune parameters in favour of the expected analysis results. The blinded analysis thus is a most objective study. The signal extraction is achieved through a BDT discriminant. This is also used for delimiting the blinded and unblinded regions of the analysis. The discriminant assigns a score between -1 and 1 to all events, where 1 is most signal like and -1 the opposite. The region below a BDT discriminant score of 0 (BDT < 0) is defined as the unblinded region deprived (< 5%) of signal with a fraction of only 3.5%, compared to the blinded region (BDT > 0) with a fraction of 18.9%. The full SR has a signal fraction of 10.4%.

This unblinded region is used for checks on the modelling of input variables provided to the MVA training. It is an important test to check the discriminant performance on data which it was not trained and optimised on. Modelling checks in the unblinded region for the two highest ranking input variables are shown in Figure 6.8.
Based on these checks, two variables are removed due to mismodeling issues, ending up with a final



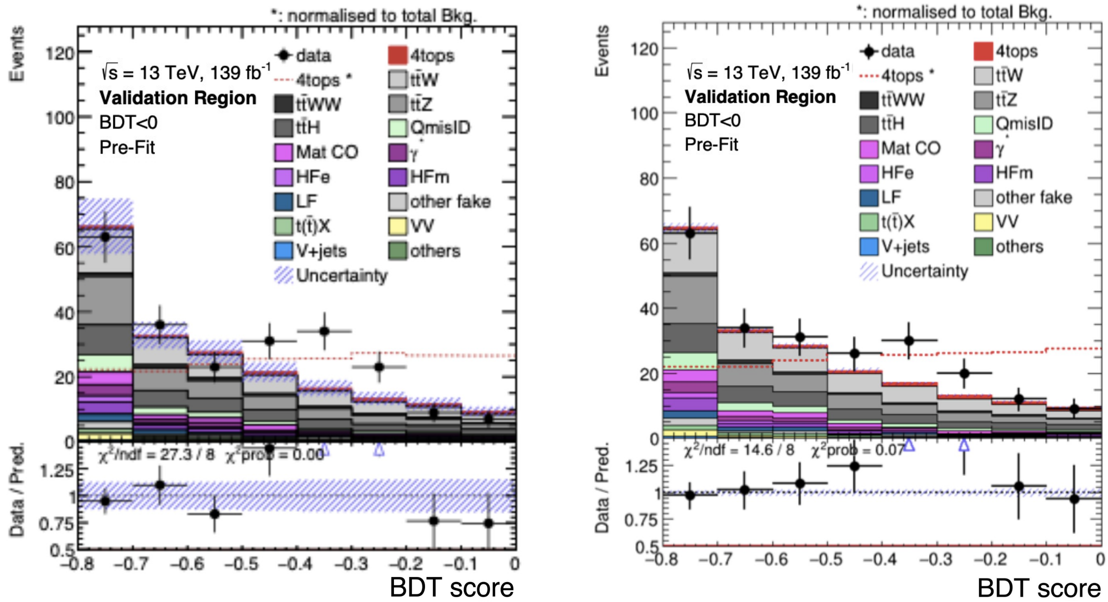
(a) Sum of the MV2c10 pseudo-continuous b-tagging score over all jets (Σw_{MV2c10}).

(b) The minimum distance between any lepton pair ($\Delta R(\ell,\ell)_{min}$).

Figure 6.8: Modelling checks in the BDT<0 validation region for the two highest ranking input variables provided to the BDT training. Both distributions are pre-fit and shaded bands represent both, statistical and systematic uncertainties.

set of 12 variables. The removed variables are marked with a ✖ in Table 6.9 and variables removed due to mismodeling are marked with an additional letter "M". The $m_{\text{jet}}/p^{\text{jet}}$ variable is removed upon the recommendation of the ATLAS Jet/$E_{\text{T}}^{\text{miss}}$ combined performance group because of the lack of proper associated systematic uncertainties for this variable. The nJets variable showed a tendency to contribute to a slight mismodeling around the central region (-0.4 < BDT Score < -0.2) of the BDT score as shown in Figure 6.9(a). The sample names, colour scheme and yields in Figure 6.9 are different from the rest of this document as these plots were made during an earlier phase of the analysis. Although the nJets distribution itself was not found to be mismodeled, tests show that, among all variables, its removal showed the largest improvement in data/MC ratio in this region for the BDT score distribution. The BDT score distribution after a training without the nJets variable is shown in Figure 6.9(b). Thus, this variable is also removed from the list of input variables.

The BDT discriminant is re-trained with the existing set of hyperparameters using the reduced set of 12 variables. With this training the ROC value was reduced from 0.854 to 0.847. The final set of chosen input variables is listed in Table 6.12. The modelling of the BDT score distribution in the unblinded BDT < 0 region is shown in Figure 6.10. Distributions of input variables in this region are provided in Appendix A. The small excess in data in the two bins between -0.4 and -0.2 is a remnant of the mismodeling despite the removal of the nJets variable from the training inputs as described above.



(a) BDT score distribution produced from the training including the nJets variable. Shaded bands represent both statistical and systematic uncertainties.

(b) BDT score distribution produced from the training without the nJets variable. Shaded bands represent statistical uncertainties, only.

Figure 6.9: Modelling checks in the BDT<0 validation region for two BDT score distributions where the only difference between left and right plots is that the right plot resulted from a training without the nJets variable. Both distributions are pre-fit. The sample names, colour scheme and yields are different from the rest of this document as these plots were made during an earlier phase of the analysis.

Rank	Variable	Category	Description
1	$\Sigma w_{\mathrm{MV2c10}}$	b-tagging	Sum of the MV2c10 pseudo-continuous b-tagging score over all jets
2	$\Delta R(\ell,\ell)_{\mathrm{min}}$	Distance	Minimum distance between any lepton pair
3	$p_{\mathrm{T}}^{\mathrm{jet}_0}$	Jet	Transverse momentum of leading jet
4	$p_{\mathrm{T}}^{b-\mathrm{jet}_0}$	Jet	Transverse momentum of leading b-tagged jet
5	$p_{\mathrm{T}}^{\ell_0}$	Lepton	Transverse momentum of leading lepton
6	$\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$	Energy	Missing transverse energy
7	$\Sigma\Delta R(\ell,\ell)_{\mathrm{min}}$	Distance	Sum of the distance between leading and sub-leading leptons in SS channel or leading, sub-leading and third-leading leptons in 3ℓ channel
8	$H_{\mathrm{T}}^{\mathrm{no\ lead\ jet}}$	Energy	Scalar sum of all lepton and jet p_{T} except leading jet
9	$\Delta R(\ell,b)_{\mathrm{max}}$	Distance	Maximum distance between leptons and b-tagged jets
10	$p_{\mathrm{T}}^{\mathrm{jet}_5}$	Jet	Transverse momentum of 6^{th} leading jet
11	$\Delta R(j,b)_{\mathrm{min}}$	Distance	Minimum distance between b-tagged jets and jets
12	$p_{\mathrm{T}}^{\mathrm{jet}_1}$	Jet	Transverse momentum of sub-leading jet

Table 6.12: Final set of variables used in the training. The rank is reported by TMVA. The computation procedure for the rank is described in Section 6.4.1. The sum of the MV2c10 pseudo-continuous b-tagging score is built with the integer of the pseudo-continuous b-tagging score intervals (see Section 4.3.6).

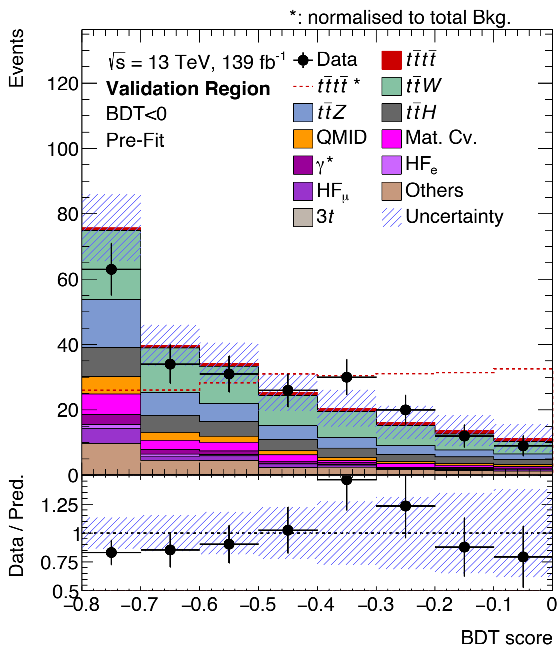
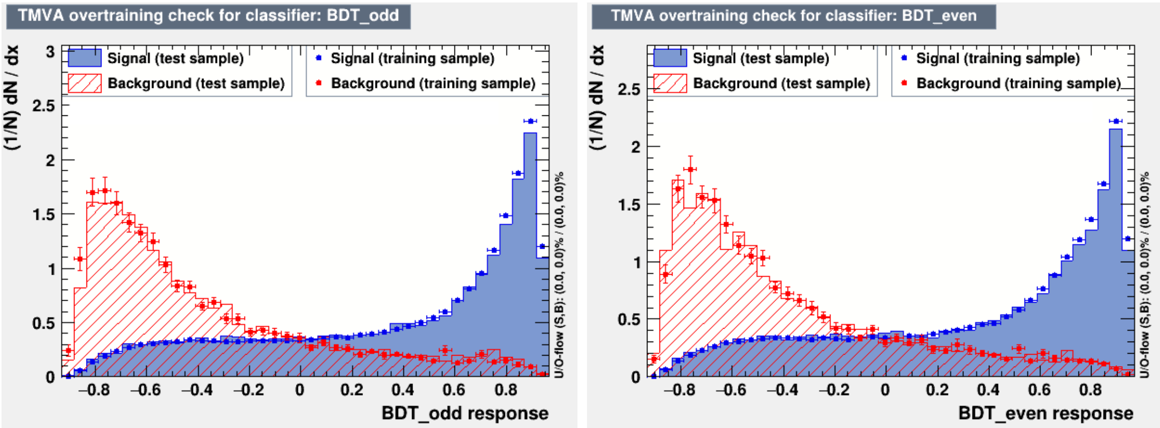


Figure 6.10: Pre-fit BDT score distribution in the BDT<0 validation region. Shaded bands represent both, statistical and systematic uncertainties.

Checks against overtraining

The results of the optimised training is checked against possible signs of overtraining by comparing training and testing sample distributions as a function of the BDT discriminant. In Figure 6.11(a) the normalised signal (background) samples used in the training of the BDT using odd-numbered events are represented with blue (red) dots. The corresponding testing response is overlayed and shown in solid blue (dashed red) pattern for the signal (background) contribution. Good agreement is observed between the training and testing responses of individual classes, indicating that the discriminant was not overtrained. The same information is shown in Figure 6.11(b) for the case of even-numbered BDT training, where also no overtraining is observed.



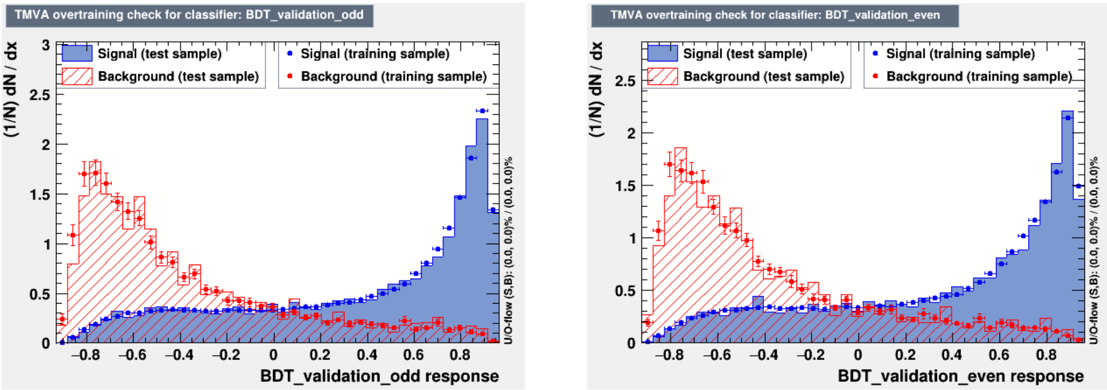
(a) BDT odd (b) BDT even

Figure 6.11: Comparison of training and testing sample responses of signal and background contributions for odd-numbered and even-numbered BDT trainings, as a functions of the discriminant score. Signal process is represented by blue dots (solid patterns) for the training (testing). The background process is represented by red dots (dashed patterns) for the training (testing). The distributions are normalised and the error bars represent the statistical uncertainties only.

Validation of performance

In Figure 6.12 the normalised signal (background) samples used in the training of the BDT are represented with blue (red) dots. The corresponding validation response is overlayed and shown in solid blue (dashed red) pattern for the signal (background) contribution, showing no sign of overtraining. Since no average and error estimation can be done on a single test or validation sample, the stability of the BDT performance in relation to statistical fluctuations in the choice of this specific sample needs to be studied. In order to have such an estimate on the testing response of the BDT training, the *bootstrap resampling method with replacement*, which is described in Section 5.2.2, is used. In this method, a sample is resampled by randomly picking events within the sample until the same number of events have been reached, allowing to choose the same event multiple times. In this way, a set of samples is generated and used to estimate the variations on the training.

The method is implemented with the Poisson Bootstrap approach. The events in the resampled samples are assigned a random weight, which is distributed according to a Poisson distribution with $\lambda = 1$



(a) BDT odd

(b) BDT even

Figure 6.12: Comparison of training and validation sample responses of signal and background contributions for odd-numbered and even-numbered BDT trainings, as a function of the discriminant score. The signal process is represented by blue dots (solid pattern) for the training (validation). Background processes are represented by red dots (dashed pattern) for the training (validation). Distributions are normalised and error bars represent the statistical uncertainties only.

using the TRandom3 function in the ROOT data analysis framework. Test and validation samples are each generated 50 times and the average and standard deviation of the testing and validation response are individually estimated using their corresponding reweighted samples. The original testing sample response returns a ROC-integral value of 0.8470. In Figure 6.13 the ROC curves for bootstrapped samples (in green) are shown together with default Odd and Even BDT response ROC curves (in red) for the testing case. The variations are observed to be small, and the average of the red curves agrees with the average performance among the many cases tested.

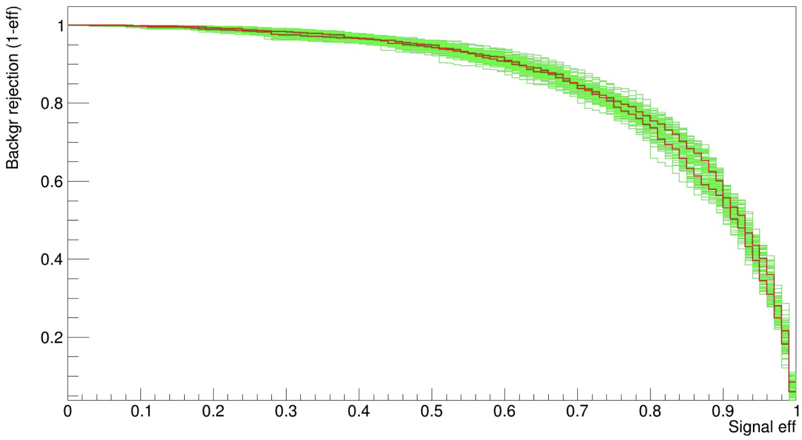


Figure 6.13: ROC curves for bootstrapped samples (green) and default Odd and Even BDT response ROC curves (red) for the testing case.

The bootstrap estimates for the testing and validation steps are 0.8466 ± 0.0021 and 0.8526 ± 0.0063 respectively, showing a stable performance. These results are plotted in Figure 6.14.

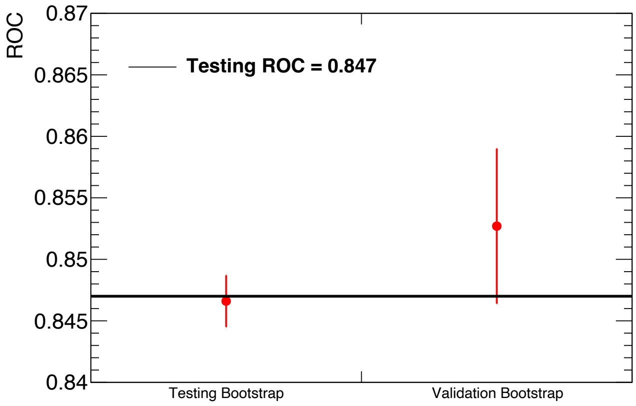


Figure 6.14: ROC values for the Poisson bootstrap estimation for the testing and validation sets compared to the testing score, represented by a black horizontal line.

Impact of sample splitting on application to data

Inheriting from the splitting strategy of the MC simulation samples, by default, the BDT score is applied to events based on the event number being odd or even. However, in the case of data events, no such requirement is necessary as none of the data events were used during the training process. This raises the question as to whether the data sample responds differently to the default and alternative ways of applying the BDT output score. This is checked in the unblinded validation region by comparing the default with three alternative applications: using only the odd BDT response, using only the even BDT response or using the average of the two; on all data events. The results are shown in Figure 6.15 and lead to the conclusion that various applications do not differ significantly. Thus, the default application is kept as the final choice.

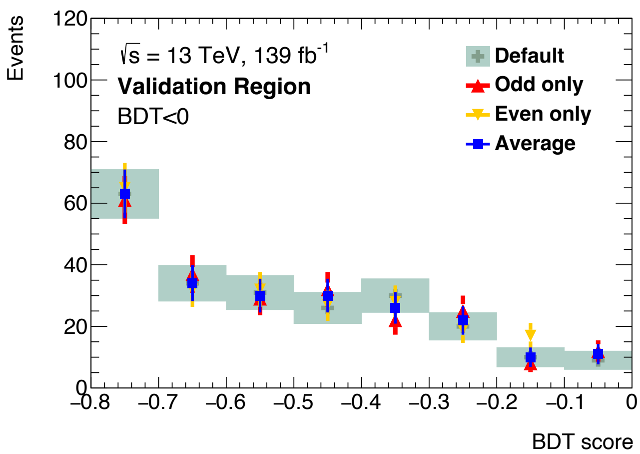


Figure 6.15: Various BDT response applications to data events in the low BDT score validation region. *Default* corresponds to cross-application, whereas *Odd* (*Even*) *only* refers to the application of only odd (even) BDT response to all events. In the case of *Average*, the average of both BDT responses is applied to all events.

6.5 Systematic uncertainties

6.5.1 Experimental uncertainties

Data-taking conditions

To reflect the uncertainties related to the interaction between the LHC and the ATLAS detector, the data-taking conditions are parametrised by two values. The uncertainty on the total integrated luminosity (`luminosity`) of the full Run-II dataset is a global uncertainty applied to all simulated samples. It is determined in ATLAS using the LUCID-2 detector [147], and is independent of this analysis. It is found to be 1.7% [148]. The second parameter that is considered is the effect of pile-up. A pile-up reweighting factor is applied to all simulated samples in order to correct them according to the pile-up profile of the data. The uncertainty (`Pileup reweighting`) is then calculated as the $\pm 1\sigma$ variation of this reweighting factor.

Leptons

Seven systematic uncertainties are considered for electrons. The labels `ATLAS_EM_SCALE` and `ATLAS_EM_RES` are used for the electromagnetic scale and resolution uncertainties, respectively [100, 149]. The same uncertainties in the case of Fast simulation are represented by the same labels with an additional suffix of `_AFII` at the end. A nuisance parameter, `ATLAS_EL_SF_ChargeID_Stat`, is used to account for the uncertainty on the efficiency of the ECIDS tool (see Section 6.2.1 for the description). The remaining four uncertainties are associated with scale factors applied as corrections to the efficiencies of trigger (`ATLAS_EL_SF_TRIGGER`), reconstruction (`ATLAS_EL_SF_RECO`), identification (`ATLAS_EL_SF_ID`) and isolation (`ATLAS_EL_SF_ISO`).

Twelve systematic uncertainties are used for muons. Two nuisance parameters account for the tracking resolution in the inner detector (`ATLAS_MU_ID`) and muon spectrometer (`ATLAS_MU_MS`). Similar to the case for electrons, four types of nuisance parameters are defined for efficiency of the correction scale factors: trigger (`ATLAS_MU_SF_TRIGGER`), reconstruction (`ATLAS_MU_SF_RECO`), identification (`ATLAS_MU_SF_ID`) and isolation (`ATLAS_MU_SF_ISO`). Furthermore, a muon-specific systematic uncertainty (`ATLAS_MU_SF_TTVA`) is introduced for the scale factor related to the track-to-vertex association (TTVA). These five uncertainties are each split into a statistical and a systematic component, distinguished by the addition of `_STAT` and `_SYST` suffixes, respectively.

The three remaining uncertainties address the momentum scale, one being charge-independent (`ATLAS_MU_SCALE`) and two charge-dependent (`ATLAS_MU_SAGITTA_RHO`, `ATLAS_MU_SAGITTA_RESBIAS`). The latter two parameters are related to the track curvature induced by the magnetic field and the bias in the track reconstruction due to the radial and rotational physical deformations in the detector geometry, respectively [150].

Jets

The systematic uncertainty in the jet energy resolution (JER) takes into account the differences between data and simulation in different regions of the p_{T}-η phase-space. Two uncertainties (`JER DataVsMC [1-2]`) are introduced to include the impact of the difference between full and fast simulation, and data. Seven further effective nuisance parameters (`Jet EffectiveNP [1-7]`) are defined through diagonalising the uncertainty matrix of each p_T-η region. Similarly, for the jet energy scale (JES) uncertainties [151], 15 such effective nuisance parameters (`JES EffectiveNP [1-15]`) are defined. Another four nuisance parameters are used for the η inter-calibration (`EtaInterCalibration [1-4]`). Three nuisance parameters are introduced for the jet flavour determination (`JES Flavour [1-3]`) and pile-up subtraction (`JES Pileup [1-3]`), each. Two nuisance parameters (`JES PunchThrough MC16 (AFII) [1-2]`) are used to account for the uncertainties on the treatment of jets that could not be contained in the calorimeters (punch-through effect). The labels `JES RelativeNonClosure MC16 (AFII)` and `JES SingleParticle HighPt` are used for the nuisance parameters related to non-closure in the fast simulation and treatment of high-p_T single hadrons. Finally, one nuisance parameter (`Jet Vertex tagger efficiency`) is defined in order to estimate the uncertainty in the scale factor used for the Jet Vertex Tagger (JVT) algorithm [152].

Jet flavour tagging

Uncertainties in the jet flavour tagging performance are calculated separately for b-, c- and light-jets, and represented with 85 independent systematic variations. 45 of these are related to the b-tagging efficiency (`b-tagging MV2c10 B[0-44]`), 20 are for the mis-tagging rate of c-jets (`b-tagging MV2c10 C[0-19]`) and the remaining 20 consider the mis-tagging rate for light-jets (`b-tagging MV2c10 B[0-19]`). Details can be found in References [153, 154, 155].

Missing transverse energy

Uncertainties in E_T^{miss} solely address the possible errors in the calibration of the soft term component as described in Section 4.3.7. For the $Z \to ee$ process there is no E_T^{miss}. Recalling the formula for the E_T^{miss} computation as provided in Equation 4.13, an $E_T^{miss} = 0$ can be achieved if the soft track term exactly cancels out the contributions from the reconstructed objects. Due to resolution effects, this does not happen and the amount of non-compensation can be measured for this decay channel in both, data and simulation. Discrepancies between the two measurements are attributed to the error on the estimation of the soft term. Relative to the beam axis, the parallel and perpendicular components are each assigned one nuisance parameter: `ATLAS_MET SoftResPerp` and `ATLAS_MET SoftResPara`. A third uncertainty, labelled `ATLAS_MET SoftScale` is used to account for the scale of the parallel component. Details can be found in Reference [110].

6.5.2 Signal modelling uncertainties

A theoretical uncertainty of 20% is assigned to the cross-section prediction for the signal process, based on the calculations at NLO precision as estimated in [55]. Labelled as `tttt Cross-Section`, this uncertainty is considered for the estimation on the error on μ_{4t}. For the case where the cross-section, as opposed to the signal-strength, was calculated, this uncertainty is not included. In order to account for differences stemming from the choice of the parton showering and hadronisation algorithm, an alternative set of distributions is created through replacing the nominal Pythia interface with HERWIG7. The alternative estimate is used as a one-sided variation, relative to the nominal, while the opposite variation is created through symmetrisation of the calculated difference. This uncertainty is labeled as `tttt (modeling) shower`. Under the name `tttt renorm./fact.scale`, the renormalisation and factorisation scale uncertainties are calculated via multiplying both values with a common factor and getting the difference to the nominal case. For the up- (down-) variation, the central values are multiplied by 2 (0.5).

The uncertainty introduced due to the choice of the PDF set is estimated by studying the effect of an alternative PDF set, NNPDF30_nlo_as_0118, on the acceptance of the signal sample in the signal region selection. First, 100 varied samples have been produced using the alternative PDF set. Then, the root-mean-square (RMS) for each bin is calculated. As differences among bins are found to be small, a flat 1% uncertainty is assigned to all bins, which is the uncertainty calculated for the last bin covering the BDT score between 0.8 and 1.0. This uncertainty is labelled as `tttt PDF`.

6.5.3 Physics background uncertainties

$t\bar{t}W$

The $t\bar{t}W$ background uncertainty related to the matrix element computation is obtained by comparing the nominal SHERPA sample with the alternative MadGraph5_aMC@NLO. Labelled as `ttW modelling (generator)`, it is a one-sided uncertainty. The scaling and renormalisation uncertainties follow the same method as described in Section 6.5.2 and are labelled `ttW varRF`. Unlike $t\bar{t}Z$ or $t\bar{t}H$, the $t\bar{t}W$ process does not have dedicated PDF or cross-section normalisation uncertainties, since its normalisation is left free to be determined by the fit and incorporated through $\mathrm{NF}_{t\bar{t}W}$.

The events that contain exactly three true b-jets or at least four true b-jets are both assigned a one-sided, 50% uncertainty, with the labels of `ttW truth 3b` and `ttW truth 4b`, respectively. Based on the rate of data/MC discrepancy observed in the ttW VR for events with at least seven jets at the detector level, two additional uncertainties are assigned. A one-sided uncertainty of 125% and 300% is assigned to events with exactly seven jets and at least eight jets, respectively. These uncertainties, labelled `ttW syst 7jets` and `ttW syst ≥8jets` introduce two new degrees of freedom to the normalisation of the $t\bar{t}W$ process governed by $\mathrm{NF}_{t\bar{t}W}$.

$t\bar{t}Z$ and $t\bar{t}H$

The $t\bar{t}Z$ and $t\bar{t}H$ backgrounds have exactly the same number of uncertainties in the same categories, and are thus discussed together. In the $t\bar{t}Z$ sample, the uncertainty related to the matrix-element computation is obtained by comparing the nominal POWHEGBOX sample to the alternative MadGraph5_aMC@NLO. For ttH, the nominal sample uses the POWHEGBOX generator, which is compared to the alternative MadGraph5_aMC@NLOgenerator. This one-sided uncertainty is labeled as `ttZ/ttH modelling (generator)`. The scaling and renormalisation uncertainties (`ttZ/ttH varRF`) as well as PDF uncertainties (`ttZ/ttH PDF`) both follow the same method as described in Section 6.5.2. The PDF uncertainties for both processes are estimated to be about 1%, the same as the signal process.

Events that contain exactly three true b-jets or at least four true b-jets are both assigned a one-sided 50% uncertainty, with the labels of `ttZ/ttH truth 3b/4b`. Finally, cross-section uncertainties of 15% and 20% are assigned to $t\bar{t}Z$ and ttH, respectively.

3t

Among all background processes, the 3t process is most similar to the signal, as it is the only process with more than two top-quarks. This fact is reflected in the BDT output where a good proportion of 3t events peaks at the high end of the distribution. The absolute impact of the 3t background is moderated by the low production cross-section predicted for 3t, about one order of magnitude smaller than the cross-section for $t\bar{t}t\bar{t}$. This prediction originates from the LO precision MC sample, as the 3t process is currently experimentally unexplored. To account for this lack of information, an ad-hoc 100% uncertainty is assumed for the cross-section prediction. An additional 50% uncertainty is assumed as one-sided variation for events with at least 4 true b-tagged jets. These uncertainties are labelled as `ttt Cross-Section` and `ttt+1 truth b`, respectively. Various variations of the 3t normalisation uncertainties and their effects on the final fit are studied in Section 6.6.3.

Single top-quark

Single-top quark production processes of the tZ, tWZ, t-channel, s-channel and tW-channels are assigned a combined 30% cross-section uncertainty (`singleTop cross section`) based on previous work [156, 157].

Minor processes

V+jets and diboson events are assigned cross-section uncertainties of 30% (`Vjets Cross-Section`) and 40% (`VV Cross-Section`) respectively. $ttXX$ processes are assigned a combined cross-section uncertainty (`Other Cross-Section`) of 50%. For the remaining minor processes (VH and VVV) the events that contain exactly three true b-jets or at least four true b-jets are both assigned a one-sided, 50% uncertainty, with the labels of `Other+3/4 truth b`.

6.5.4 Instrumental background uncertainties

Material and virtual photon conversions

These two background contributions are estimated in the fit and quantified with $\mathrm{NF}_{\mathrm{Conv.}}$, thus no normalisation uncertainties are considered in the analysis. The shape uncertainties still need to be included since in the template method these are purely derived from MC simulation. These uncertainties are computed by comparing data to simulated $Z(\rightarrow \mu\mu) + \gamma$ and $Z(\rightarrow \mu\mu)+$ jets samples in a region enriched in $Z(\rightarrow \mu\mu) + \gamma$ events. In this region, a $\mu^{\pm}\mu^{\mp}e$ trilepton selection is used to target $\mu \rightarrow \mu + \gamma$ and $\gamma \rightarrow e^{+}e^{-}$ events, where one electron is soft. The invariant mass of the trilepton system ($m_{\ell\ell\ell}$) is required to be $m_{\ell\ell\ell} \in [84, 96]$ GeV so as to select Z-boson candidates.

The data/MC comparisons show a maximum discrepancy of about 25% for both background sources in the corresponding regions where they are dominant. A flat 25% uncertainty is assumed for the material conversion (`CO ShapeSyst`) and virtual photon conversion (`Gstr ShapeSyst`) events that do not match with the selection of ConvCR, therefore accounting for an extrapolation of these background events into other regions.

Heavy-flavour non-prompt leptons

Similarly to the photon conversion case, heavy-flavour non-prompt lepton contributions are estimated in the fit and quantified with $\mathrm{NF}_{\mathrm{HF}_e}$ and $\mathrm{NF}_{\mathrm{HF}_\mu}$, thus no normalisation uncertainties are considered for these processes. The shape uncertainties are separately estimated for electrons and muons, and estimated in each bin of all regions used in the final data fit. As fit regions nominally use Tight lepton definitions that suppress the non-prompt contributions, here, Loose lepton definitions are used to avoid statistical fluctuations due to low number of non-prompt lepton events. The computation of uncertainties is performed in two steps. First, the difference between data and MC simulation contributions of all processes other than the non-prompt heavy-flavour processes ($\mathrm{HF}^{\mathrm{no}}_{e,m}$) is calculated. Second, the ratio ($\mathrm{R}_{e,m}$) of this number to the $\mathrm{HF}_{e,m}$ event yields is found:

$$\mathrm{R}_{e,m} = \frac{\mathrm{Data} - \mathrm{HF}^{\mathrm{no}}_{e,m}}{\mathrm{HF}_{e,m}}. \tag{6.5}$$

This ratio is taken as the uncertainty, and contributions from all regions are combined into a single, one-sided NP per lepton channel (`HFe_ShapeSyst` and `HFm_ShapeSyst`). These two nuisance parameters are correlated due to their common physics origin.

Light flavour non-prompt leptons

The non-prompt lepton events from light-flavour jets are assigned a 100% normalisation uncertainty, based on previous results [158], where it is shown to be sufficient to account for the discrepancy between data and the MC simulation in regions using loose lepton definitions. This uncertainty is labeled as `ttbar_light Cross-Section`.

QMID

The charge mis-identification background uncertainty (`QMisID_Sys`) is given in the form of two-dimensional maps binned in p_{T} and η. The maps are calculated separately for Conv. CR, $t\bar{t}W$ CR and SR. The uncertainties consist of three components:

- The statistical uncertainty of the fit: QMID rates are calculated by a likelihood fit to the data. The statistical uncertainty of this fit is the first source of uncertainty.
- Impact of the dielectron invariant mass requirements: Selections are applied to the invariant mass distribution of the dielectron system to define the Z-boson resonance peak and the sideband regions. The width of these regions can change the outcome. The impact of these choices is estimated by varying the range of regions and using the resulting differences between the calculated rates.
- Non-closure: Differences occur between predicted and observed mis-identification rates. This non-closure is included as an additional uncertainty.

These three uncertainties are combined into a single NP in the final fit setup under the label `QMR Sys.`.

Additional $t\bar{t}$+jets associated processes

Remaining minor processes from the $t\bar{t}$+jets production are assigned a 30% normalisation uncertainty (`ttbar_others Cross-Section`), whereas the shape-related uncertainties are not considered due to the negligible amount of contributions from these processes. For $t\bar{t}$+jets events with three true b-jets, an uncertainty (`ttbar Cross-Section [3b]`) of 30% is assigned. If the events have four or more true b-jets, an additional 30% uncertainty (`ttbar Cross-Section [4b]`) is assigned to these events. These HF-related uncertainties are motivated by a previous $t\bar{t}$+HF measurement performed by the ATLAS Collaboration [159].

6.6 Results

The binned profile-likelihood fit method is used to extract the cross-section and the significance of the $t\bar{t}t\bar{t}$ process using the analysis setup and statistical model described in this chapter. The fit is performed using the TRExFitter [160] software developed within the ATLAS Collaboration which is built upon the RooFit [161] framework and features available in the RooStats [162] tool to build statistical models. The output provided by TRExFitter based on these programs have the HistFactory [163] format. Uncertainties are estimated by means of the MINOS [164] algorithm after the minimisation of the likelihood which is performed with MINUIT [164].

In the SR, the BDT score is used as the discriminating distribution. The binning of this distribution is determined based on the *TransfoD* [165] algorithm available in the TRExFitter. The algorithm tries to maximise signal-to-background separation while trying to keep MC statistical uncertainties small. This is achieved by starting from an initial distribution with a relatively large number of bins and sequentially merging them according to given criteria. The scanning starts from the bin with the largest BDT score, merging neighbouring bins until the score Z, defined as

$$Z = z_b \frac{n_b}{N_b} + z_s \frac{n_s}{N_s}, \tag{6.6}$$

becomes larger than 1. The procedure is then repeated starting from the first neighbouring bin that is not merged. The parameters z_s and z_b can be adjusted, regulating the fraction of signal and background events, respectively. The sum of these two parameters is required to be equal to the total number of bins. Here, n_b/N_b (n_s/N_s) is the ratio of number of signal (background) events in the merged bin to the total number of signal (background) events.

The algorithm converges to parameters that maximise the expected sensitivity of the fit, ensuring a non-zero contribution of all background bins at the same time. The obtained binning is shown in Figure 6.16 with an equidistant bin width of 0.1 between -0.8 and 1.0 and with the last bin with a width of 0.2.

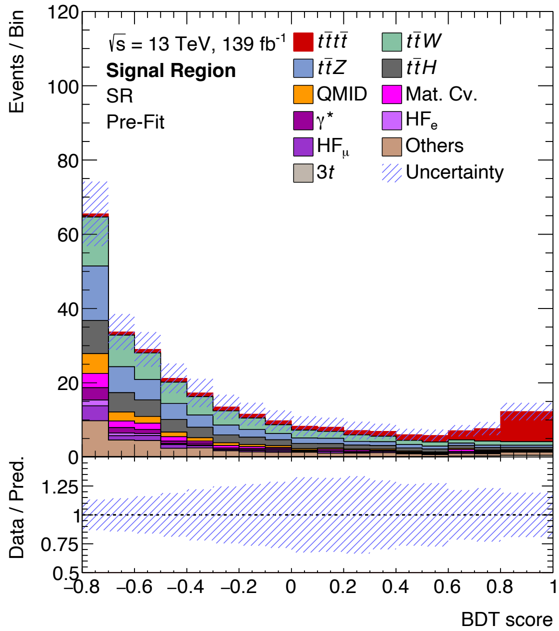


Figure 6.16: Pre-fit distributions for the SR using the Asimov fit. The shaded bands represent both statistical and systematic uncertainties. The leftmost bin is an underflow bin.

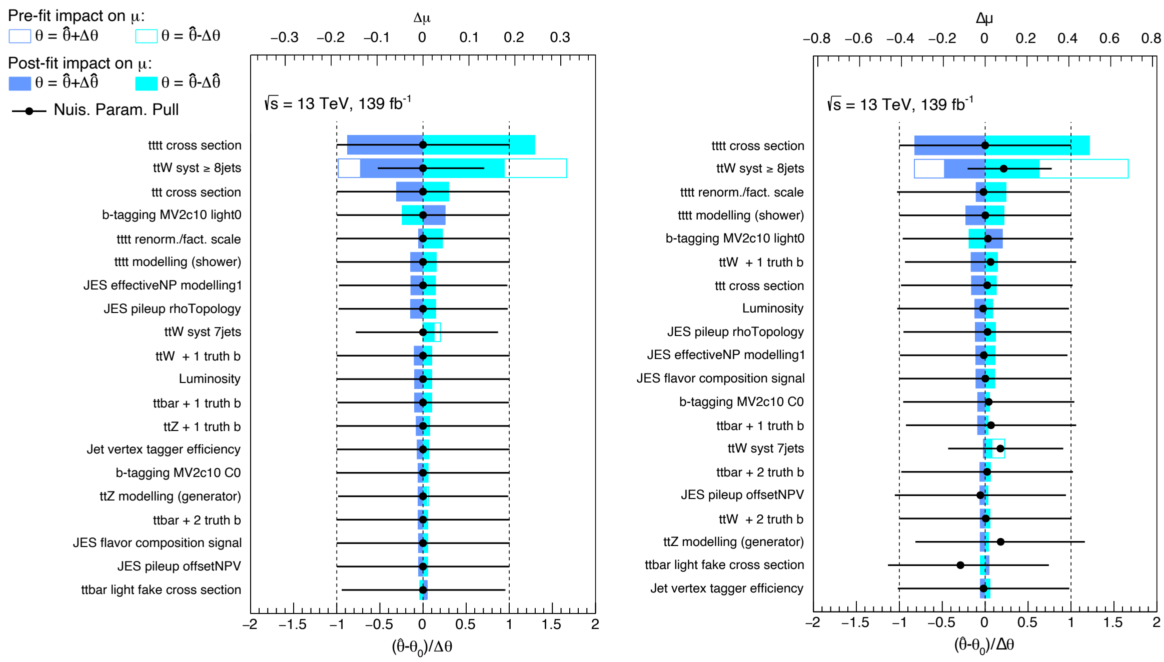
6.6.1 Expected performance of the fit

Before the unblinded fit to observed data, the expected performance and behaviour of the fit model are studied using the Asimov dataset, described in Section 5.1.3, where no observed data is used. An exception is the QMID process, which is estimated by the data-driven method, described in Section 6.3.2. A fit is performed including the four control regions introduced, as well as the full signal region.

It should be noted that the expected fit results reported in Reference [3] are different from the ones reported here. This is because the definition of "expected fit" is different between setups. The classical definition of the Asimov fit is used in this thesis in order to be comparable to the corresponding CMS publication [69]. In the case of Reference [3], first, a fit to data in CRs was performed. NFs and NPs estimated from this fit are then used as initial values in an Asimov fit. There exist also two different versions of this fit in the literature, depending on whether the BDT distribution in SR is split or not, are performed. The result of Reference [3] is estimated from a fit using the BDT < 0 unblinded distribution as an additional CR and the BDT > 0 blinded region as SR. Whereas results reported in Reference [166] are obtained with CRs described in this thesis, and the blinded unsplit SR.

A post-fit expected signal strength of $\mu = 1.00^{+0.55}_{-0.41}$ is measured. The expected significance of the fit is found to be 2.9σ. The normalisation factors are listed in Table 6.13. All NFs have their central values equal to 1 per definition, as the Asimov dataset has been used. Different NPs affect the fit at varying levels. Their impact is quantified based on the following definition: For a nuisance parameter θ, its impact $\Delta\mu$ on the signal-strength μ is given by the difference in μ between its nominal fit value and another fit in which the same nuisance parameter is set to a value of $\hat{\theta} \pm x$. Here, $\hat{\theta}$ is the maximum likelihood estimator of θ also known as the post-fit value. The value x depends on the impact estimation either from pre-fit or post-fit. For the former, x is equal to the pre-fit uncertainty $\Delta\theta$ following a Gaussian distribution i.e. $x = \Delta\theta$ corresponding to a 1σ variation. For the latter, x is equal to the uncertainty of $\hat{\theta}$ i.e. $x = \Delta\hat{\theta}$ not corresponding to a 1σ variation. It is possible that a NP could be constrained by the fit and thus, post-fit impact of an NP might be smaller than its pre-fit counterpart.

Sorted by their post-fit impact (solid blue and cyan bars) on the signal-strength, the 20 top ranking systematic uncertainties are shown in Figure 6.17(a) together with their pre-fit impact (hollow blue and cyan bars). The dot and corresponding error line represent the relative deviation of the fitted value $\hat{\theta}$ from its nominal value θ_0 and the post-fit uncertainty of the NP, respectively. Both values are expressed in units of pre-fit uncertainty $\Delta\theta$ shown in the bottom horizontal scale. The dashed vertical lines at ± 1 mark the case where pre- and post-fit uncertainties are equal. If an NP is constrained in the post-fit case, then the width of the error line will be smaller than one. For the signal strength



(a) Ranking plot for the Asimov Fit

(b) Ranking plot for the data Fit

Figure 6.17: Ranking plots for (a) Asimov and (b) data fits. The top ranking 20 NPs are shown, sorted by their post-fit impact (solid blue and cyan bars) on the signal strength. The reader is referred to amin text for a detailed description.

measurement, the most important systematic uncertainty is the cross-section uncertainty on the signal process, followed by the modelling uncertainty of the $t\bar{t}W$ process for events with more than or equal to eight jets. Similarly, the modelling uncertainty for $t\bar{t}W$ events with exactly seven jets ranks higher. Both uncertainties could be constrained by the fit. The third highest-ranking uncertainty is the cross-section uncertainty of the $3t$ process. This reflects the fact that this process is difficult to distinguish from the signal process and thus contributes to the bins at the high-end of the BDT distribution where the sensitivity to the signal process is the largest.

Fit Model	$\mathrm{NF}_{\mathrm{Conv.}}$	NF_{γ^*}	$\mathrm{NF}_{\mathrm{HF}_e}$	$\mathrm{NF}_{\mathrm{HF}_\mu}$	$\mathrm{NF}_{t\bar{t}W}$	μ	σ
Asimov Fit	$1.00^{+0.42}_{-0.38}$	$1.00^{+0.41}_{-0.35}$	$1.00^{+0.43}_{-0.40}$	$1.00^{+0.37}_{-0.34}$	$1.00^{+0.27}_{-0.25}$	$1.00^{+0.55}_{-0.41}$	2.9
Data Fit	$1.61^{+0.46}_{-0.42}$	$0.93^{+0.42}_{-0.36}$	$0.85^{+0.42}_{-0.40}$	$1.07^{+0.38}_{-0.32}$	$1.56^{+0.30}_{-0.28}$	$2.02^{+0.83}_{-0.61}$	4.3

Table 6.13: Comparison of results for the Asimov and data fits. μ and σ stands for signal strength and significance, respectively.

6.6.2 Fit to data

After the study of the expected behaviour with the Asimov data, the fit is repeated with the observed data. The four CRs and the unsplit SR distribution is used in the profile-likelihood fit. The fit simultaneously determines the signal strength as well as the five previously mentioned NFs. The observed signal strength is

$$\mu = 2.0 \pm 0.4(\text{stat.})^{+0.7}_{-0.4}(\text{syst.}) = 2.0^{+0.8}_{-0.6}, \tag{6.7}$$

corresponding to a significance of 4.34σ, presenting evidence for this process. The measured signal strength is found to be compatible with the SM prediction at NLO accuracy, within 1.7σ.

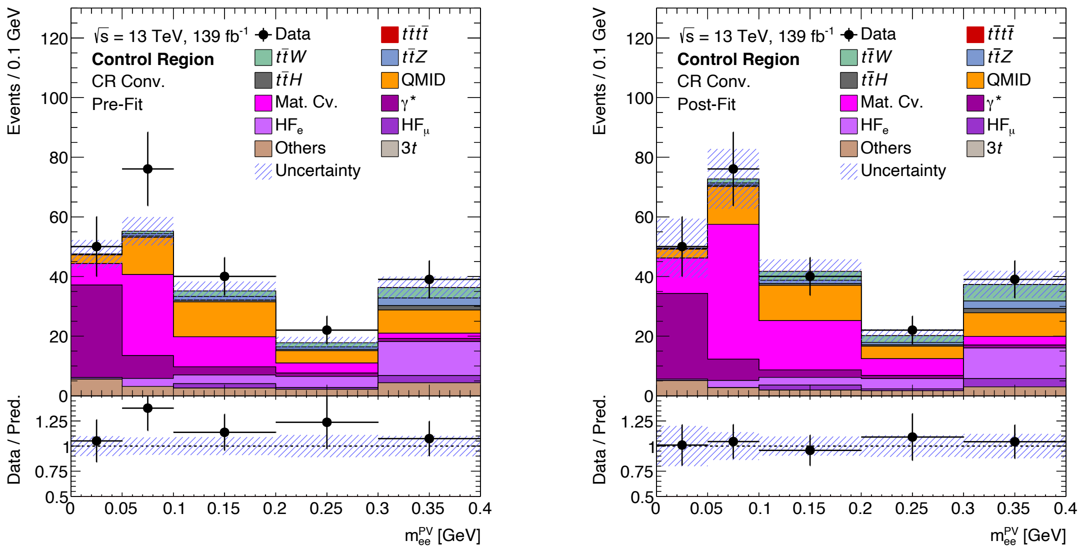
The inclusive cross-section of the $t\bar{t}t\bar{t}$ process is extracted by multiplying the QCD NLO prediction of $\sigma_{t\bar{t}t\bar{t}} = 12.0 \pm 2.4$ fb with the signal strength measured without the 20% theoretical cross-section uncertainty, as in this case the cross-section itself is the free parameter to be measured that is independent of the normalisation. The cross-section is measured to be

$$\sigma_{tttt} = 24 + 5(\text{stat.})^{+5}_{-4}(\text{syst.})\text{fb} = 24^{+7}_{-6}\text{fb}. \tag{6.8}$$

In comparison, the CMS Collaboration measured a $t\bar{t}t\bar{t}$ cross-section of $12.6^{+5.8}_{-5.2}$ fb [69], and thus a result much closer to the SM prediction. The expected (observed) sensitivity achieved by the CMS collaboration in that publication is 2.6σ (2.7σ).

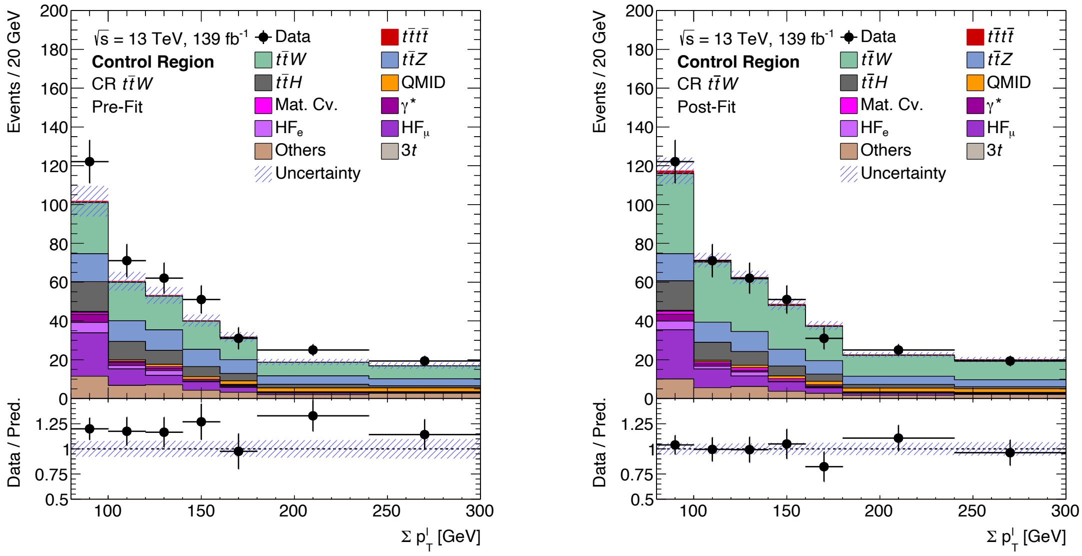
All normalisation factors are given in Table 6.13. $\mathrm{NF}_{\mathrm{conv}}$ is estimated to be $1.61^{+0.46}_{-0.42}$. The increase relative to the prediction is compatible with the observation that there are more data events in all the bins of the m_{ee}^{PV} distribution in the CR Conv., with a particularly strong discrepancy in the second bin where material conversion is the dominant contribution, as shown in Figure 6.18(a). NF_{γ^*} is estimated to be $0.93^{+0.42}_{-0.36}$, close to the pre-fit value. In CR Conv., the γ^* contribution is most dominant in the very first bin, where the discrepancy between data and prediction is small and the fit model prefers to bridge the gap by enhancing $\mathrm{NF}_{\mathrm{conv}}$ in general. The post-fit distribution is shown in Figure 6.18(b).

Besides $\mathrm{NF}_{\mathrm{conv}}$, also $\mathrm{NF}_{t\bar{t}W}$ resulted in an increased estimation with a value of $1.56^{+0.30}_{-0.28}$. As shown in Figure 6.19(a), the number of data events exceeds the prediction in all the bins of the distribution in the $t\bar{t}W$ CR. Considering a $t\bar{t}W$ production cross-section prediction increased by 20% in an earlier ATLAS analysis, the 56% larger estimate in $t\bar{t}t\bar{t}$analysis is qualitatively in agreement. For a discussion on the jet and electroweak effects considered for the increased prediction the reader is referred to Section 3.3. Good post-fit agreement is observed in Figure 6.19(b) between data and the simulation in the $t\bar{t}W$ CR, where the initial excess in data over the whole range is accompanied by an increase in the amount of $t\bar{t}W$ events.



(a) Pre-fit (b) Post-fit

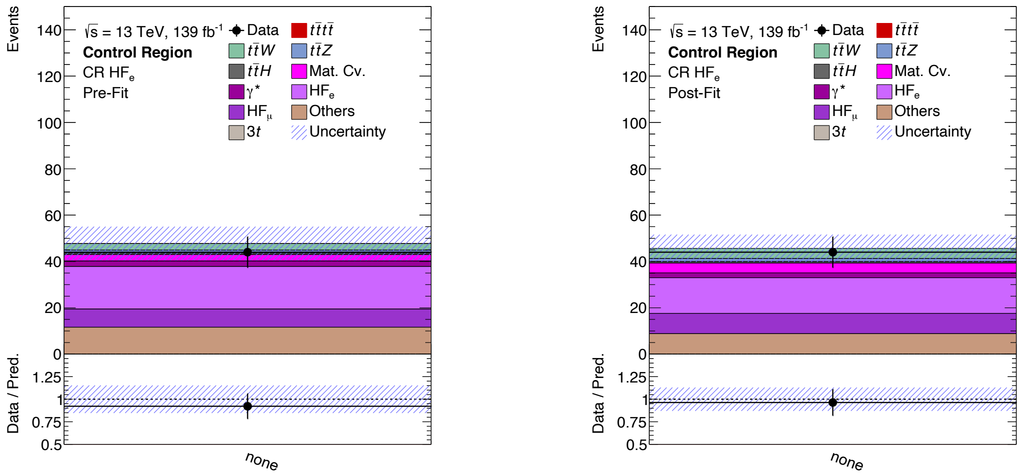
Figure 6.18: Pre- and post-fit distributions for CR Conv. using the data fit. Shaded bands represent both, statistical and systematic uncertainties.



(a) Pre-fit (b) Post-fit

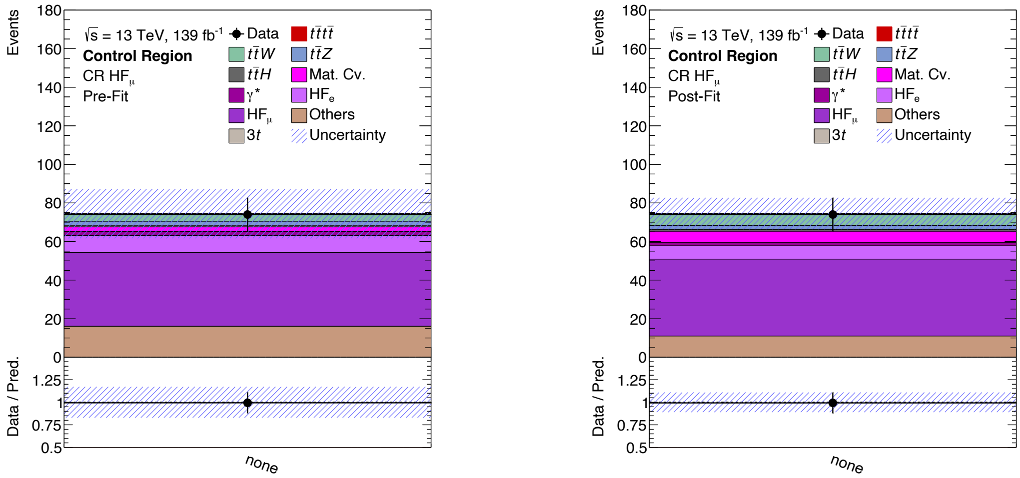
Figure 6.19: Pre- and post-fit distributions for the $t\bar{t}W$ CR obtained in the fit to data. The shaded bands represent both, statistical and systematic uncertainties.

$\mathrm{NF}_{\mathrm{HF}_e}$ is found to be lower than the prediction with a value of $0.85^{+0.42}_{-0.40}$, whereas $\mathrm{NF}_{\mathrm{HF}_\mu}$ is found to be close to its nominal prediction with an estimate of $1.07^{+0.38}_{-0.32}$. Both results are qualitatively in agreement with data-to-prediction ratios in CRs as shown in Figures 6.20 and 6.21. The agreement between data and the prediction improved in all CRs in the post-fit case with remaining discrepancies mostly lying within the uncertainties.



(a) Pre-fit (b) Post-fit

Figure 6.20: Pre- and post-fit distributions for the HF_e CR obtained from the data fit. The shaded bands represent both, statistical and systematic uncertainties.



(a) Pre-fit (b) Post-fit

Figure 6.21: Pre- and post-fit distributions for the HF_μ CR obtained from the data fit. The shaded bands represent both, statistical and systematic uncertainties.

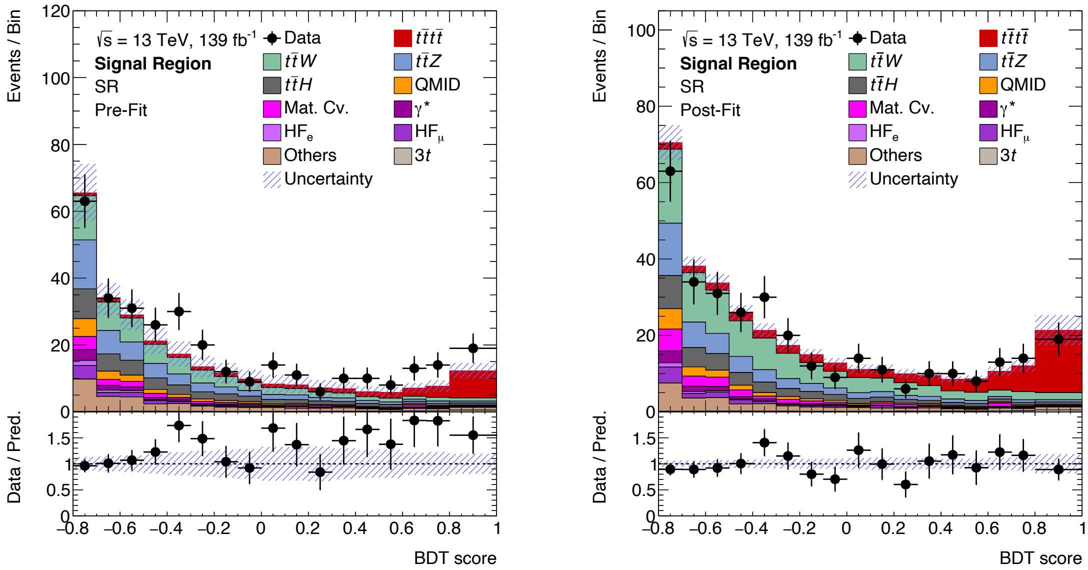
Most NPs are not significantly modified by the fit. The largest changes are seen in the `ttW syst`

`7jets` and `ttW syst ≥8jets` systematic uncertainties, both of which have been shifted to larger values in the fit: $0.18^{+0.73}_{-0.61}$ and $0.22^{+0.56}_{-0.42}$, respectively. This has led to an increase of about 22% (65%) in the $t\bar{t}W$ events with 7 jets (≥ 8 jets). Combined with the effect of $\mathrm{NF}_{t\bar{t}W}$, the total yield of $t\bar{t}W$ events in the signal-enriched (BDT > 0) region has increased to 23.2 ± 10.1 from an initial pre-fit value of 12.4 ± 8.8. The total list of post-fit NPs are provided in Appendix B. Main categories of systematic uncertainties and their impacts are listed in Table 6.14.

Category	Uncertainty	$\Delta\mu$	
Signal Modelling			
	$t\bar{t}t\bar{t}$ cross-section	+0.56	-0.31
	$t\bar{t}t\bar{t}$ modeling	+0.15	-0.09
Background Modelling			
	$t\bar{t}W$+jet modeling	+0.26	-0.27
	$t\bar{t}Z$+jet modeling	+0.02	-0.04
	$t\bar{t}H$+jet modeling	+0.04	-0.01
	QMID modeling	+0.01	-0.02
	Non-prompt leptons' modeling	+0.05	-0.04
	Modelling of other backgrounds	+0.03	-0.02
	$3t$ modeling	+0.10	-0.07
Instrumental			
	Luminosity	+0.05	-0.03
	Jet uncertainties	+0.12	-0.08
	Jet flavour tagging (LF)	+0.11	-0.06
	Jet flavour tagging (b-jets)	+0.04	-0.03
	Jet flavour tagging (c-jets)	+0.03	-0.01
	Simulated sample size	+0.06	-0.06
	Other experimental uncertainties	+0.03	-0.01
Total Systematic Uncertainty		+0.70	-0.44
Statistical		+0.42	-0.39
	Non-prompt lepton normalisation ($\mathrm{HF}_{e,\mu}$, Mat. Cv., γ^*)	+0.05	-0.04
	$t\bar{t}W$ normalisation	+0.04	-0.04
Total Uncertainty		0.83	-0.60

Table 6.14: Breakdown of uncertainties into main categories and their impact on the parameter of interest.

Pre- and post-fit SR distributions are shown in Figure 6.22. In the last six bins of the pre-fit SR distribution, the amount of data events are observed to be consistently larger than the prediction increasing bin-by-bin. These bins are the bins with the highest signal fractions. In the corresponding post-fit distribution, it can be seen that the gap between data and prediction is reduced by increasing the amount of signal events. As such, the fit determines additional data events to be signal-like. Distributions of input variables in this region are provided in Appendix C.

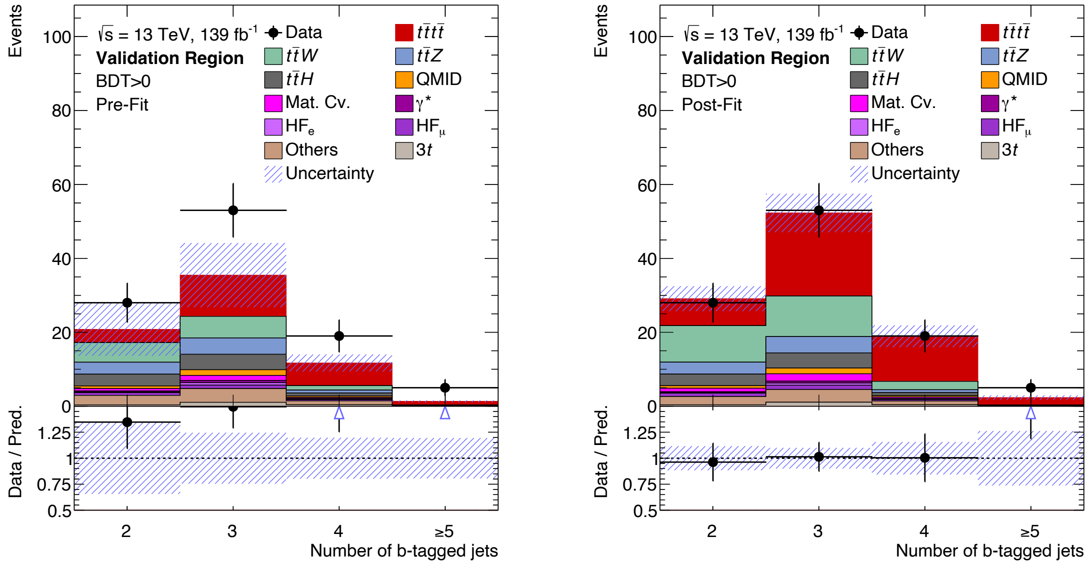


(a) Pre-fit

(b) Post-fit

Figure 6.22: Pre- and post-fit distributions for SR obtained from the data fit. The shaded bands represent both, statistical and systematic uncertainties. The leftmost bin contains the underflow.

This behaviour can be investigated when looking at distributions in the high-BDT score (BDT > 0) region. An example is shown in Figure 6.23 with the b-tagged jet multiplicity distributions in the pre- and post-fit cases where the fit increases the signal contribution significantly in the post-fit case. Further examples are provided in Appendix A. The post-fit yields for the whole SR and the high-BDT selection are given in Table 6.15. The corresponding SR BDT distribution used by the CMS collaboration [69] is also shown in Figure 6.24. Here, it can be seen that the MVA discriminant used by the CMS collaboration peaks at the center in contrast to a left-peaked BDT discriminant used in this analysis.



(a) Pre-fit (b) Post-fit

Figure 6.23: Pre- and post-fit distributions for the number of b-tagged jets in the BDT > 0 region, using the data fit. The shaded bands represent both, statistical and systematic uncertainties.

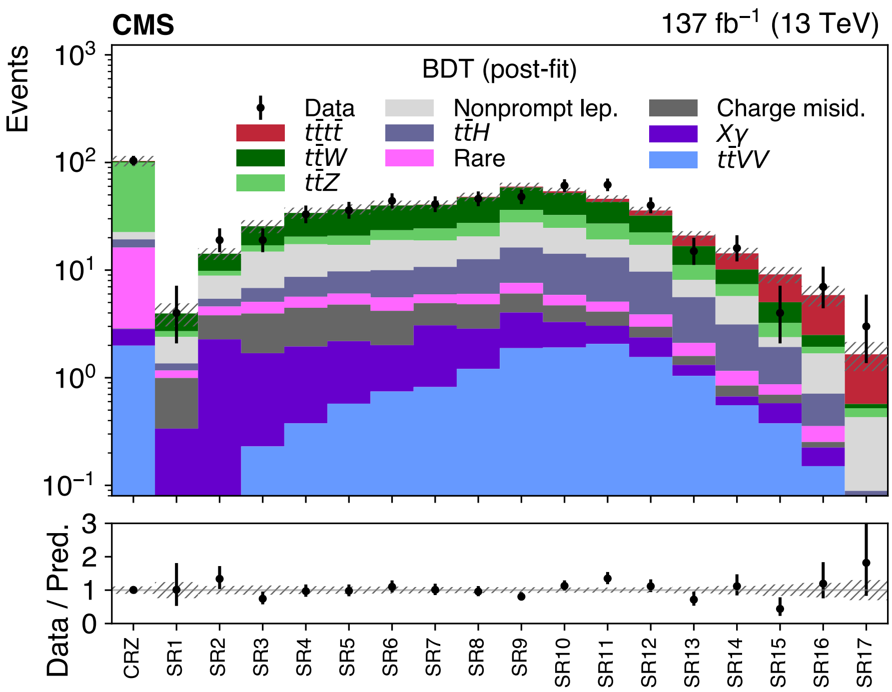


Figure 6.24: Post-fit SR BDT distribution used by the CMS collaboration in Reference [69]. The first bin from the left is not part of the BDT distribution and shows the yields in a control region named “CRZ”.

Process	Pre-Fit SR Yield	Post-Fit SR Yield	Pre-fit % → Post-fit %
$t\bar{t}W$	61.2 ± 37.9	99.6 ± 24.9	23% → 30%
$t\bar{t}Z$	49.7 ± 9.9	47.7 ± 8.9	18% → 14%
$t\bar{t}H$	38.8 ± 9.3	37.8 ± 8.7	14% → 11%
QMID	16.2 ± 1.4	16.3 ± 1.3	6.0% → 4.8%
Mat. Cv.	12.7 ± 2.6	18.9 ± 5.9	4.7% → 5.6%
γ^*	9.3 ± 1.5	8.8 ± 3.8	3.4% → 2.7%
HF_e	4.8 ± 1.3	3.6 ± 1.9	1.8% → 1.1%
HF_μ	11.0 ± 2.1	11.1 ± 3.8	4.0% → 3.3%
Others	34.1 ± 3.1	31.2 ± 3.5	12% → 9.3%
$3t$	2.9 ± 2.9	3.1 ± 3.0	1.1% → 0.9%
Total Background	240.9 ± 46.6	275.0 ± 52.2	89% → 82%
$t\bar{t}t\bar{t}$	29.7 ± 6.2	59.7 ± 16.7	11% → 18%
Total	270.7 ± 47.5	335.0 ± 19.6	
Data	330	330	

Table 6.15: Pre- and post-fit yields and their relative fractions in the SR for the default fit.

6.6.3 Post-unblinding checks

After the unblinding of the SR, the response of the fit model to data is studied by varying the model parameters. This is done to probe the impact of certain choices that are part of the default fit model.

Performance of the fit to H_T distributions

Within ATLAS collaboration, the MVA discriminant developed for this analysis is the first one employing a BDT discriminant for $t\bar{t}t\bar{t}$ analyses. In previous ATLAS $t\bar{t}t\bar{t}$ analyses H_T distributions were used as SRs and split according to lepton (N_ℓ) and b-tagged jet multiplicities (N_b). In order to compare the performance, the fit is repeated by using the SR splitting from the previous analysis instead of the single SR in the BDT distribution. All other selections are kept identical to those of the default analysis. These splitted signal regions have the following (N_ℓ, N_b) selections: (2, 2), (2, 3), (3, 2), (3, 3), (≥2, 4). The observed (expected) sensitivity of this fit is found to be 4.23σ (2.35σ). Post-fit distributions of the H_T variable in five SRs are shown in Figure 6.25. The SR with the MVA discriminant outperforms the legacy method, particularly showing a strong improvement in the expected significance. However, the fit based on the H_T distribution is also found to be strong enough to claim evidence. The NFs in this setup are very similar to the ones of the default fit model, as shown in Figure 6.27.

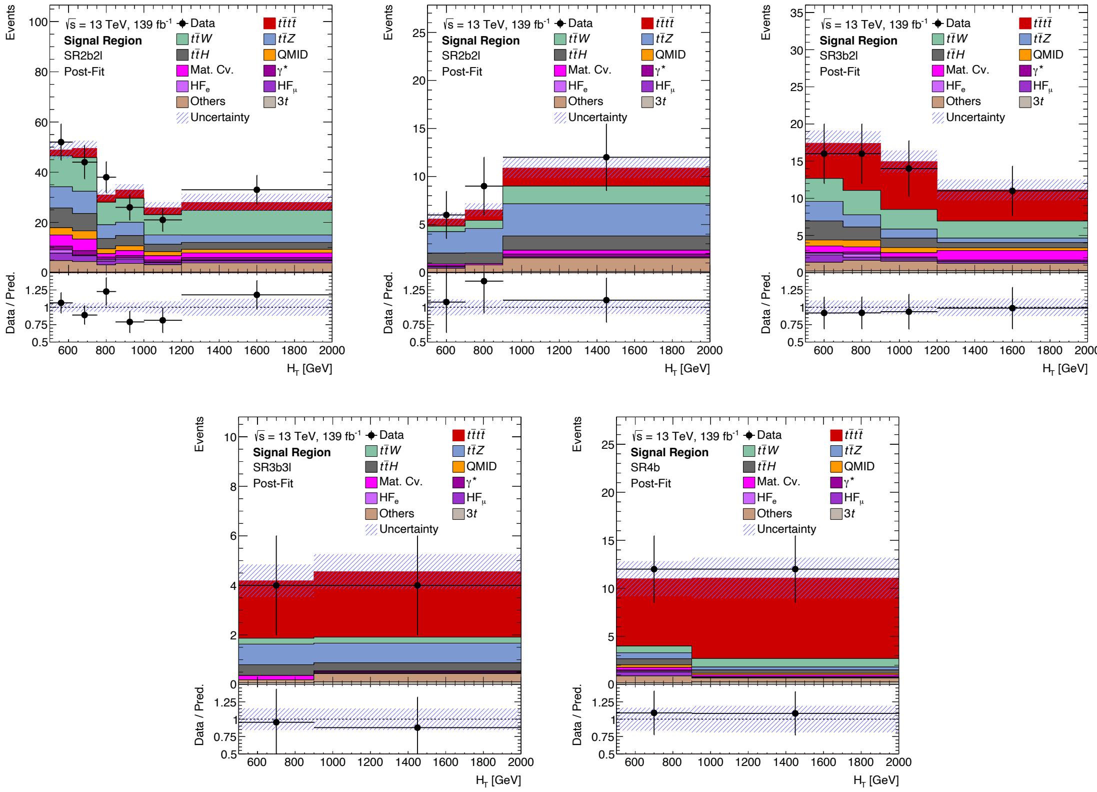


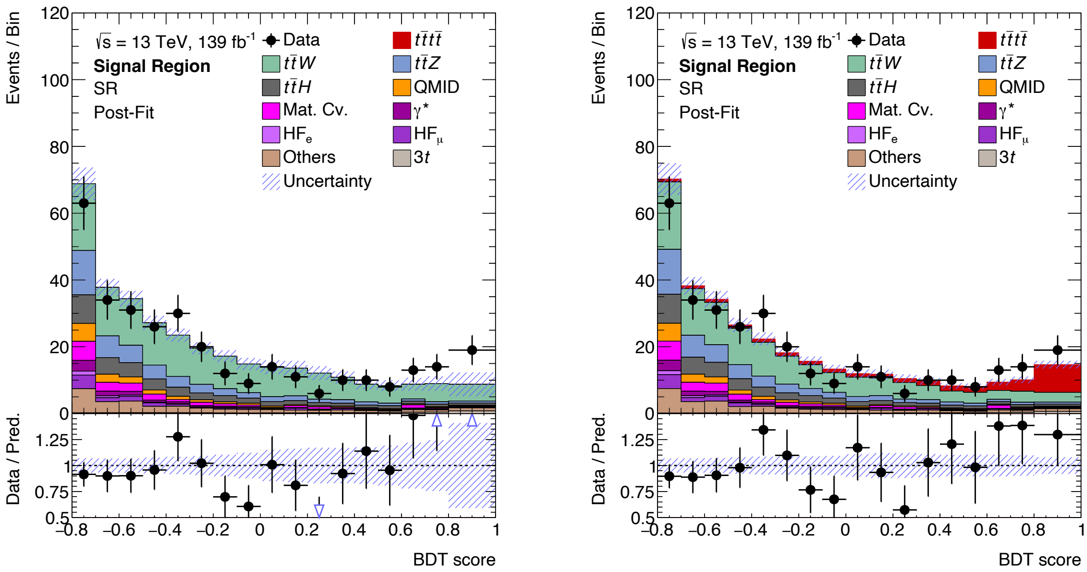
Figure 6.25: Post-fit distributions for the five SRs defined according to a splitting based on the previous ATLAS analysis, as detailed in Section 6.6.3. The shaded bands represent both statistical and systematic uncertainties.

Fit with fixed signal strengths

In the default measurement the signal strength μ is the parameter of interest, and thus a free parameter to be determined by the fit. It is of interest what type of outcome the fit to data would yield when the signal strength is instead fixed to its value predicted by the SM i.e. $\mu = 1.0$, or to zero which is equivalent to a lack of signal process. In both cases, the estimated NFs are very similar to the default fit as shown in Figure 6.27. This is expected, as the estimation of the signal strength mostly depends on the composition of the CRs, which is the same in all three fit setups. The distributions in the SRs show some differences, again as expected, given the fixed values for the signal strength. Among the three scenarios, it is seen that the default fit setup describes the data best, this is particularly evident in the last three bins of the BDT distributions where the signal fraction is highest. The post-fit SR distributions for the $\mu = 0$ and $\mu = 1$ fit models are shown in Figures 6.26(a) and 6.26(b), respectively.

Impact of $3t$ process cross-section

As previously discussed, the $3t$ process is the most signal-like among all background processes considered in this analysis. In absolute terms, it is a minor contribution, however this contribution is concentrated in the signal-like bins at the higher end of the BDT distribution. Given the absence of any experimental measurement of this process, the theoretical cross-section is assigned a large



(a) $\mu = 0$ (b) $\mu = 1$

Figure 6.26: Post-Fit distributions in the SR for (a) $\mu = 0$ and (b) $\mu = 1$ fit to data. The shaded bands represent both statistical and systematic uncertainties. The leftmost bin contains the underflow.

uncertainty of 100%. This warrants a further study on how the $t\bar{t}t\bar{t}$ measurement is influenced by various cross-section hypotheses for the $3t$ process. This is probed by repeating the fit with a cross-section for the $3t$ process scaled by the factors 1, 1.5, 2 and 5 of the nominal value. These varied fits show similar NF estimates when compared to the default fit as visualised in Figure 6.27. Due to its similarity to the signal process, the increased cross-section of the $3t$ process correlates with a decrease in the signal strength. However, the impact of this inverse correlation is small in the fit model and the measurement's sensitivity allows for claiming evidence for the $t\bar{t}t\bar{t}$ process for all assumed scale factors.

Differences between data-taking periods

The data used in this analysis has been collected by the ATLAS detector during the Run 2 period of LHC. This period is divided into two parts: 2015–2017 and 2018. This division reflects the different data-taking conditions during two time intervals, explained in Section 4.1. In order to study the possible dependence of the results on the accelerator conditions, the default fit is repeated separately for each period, using data and simulation files corresponding to each part. The results are found to be in agreement with each other within uncertainties. Despite not being significant, central values of the NFs are observed to differ between different periods when compared to the combined fit as shown in Figure 6.27. The estimate for the signal strength is found to be less affected by the different fit models.

In conclusion, no particular parameter has been found as source for the observed data excess. All checks are found to be compatible within the associated uncertainties.

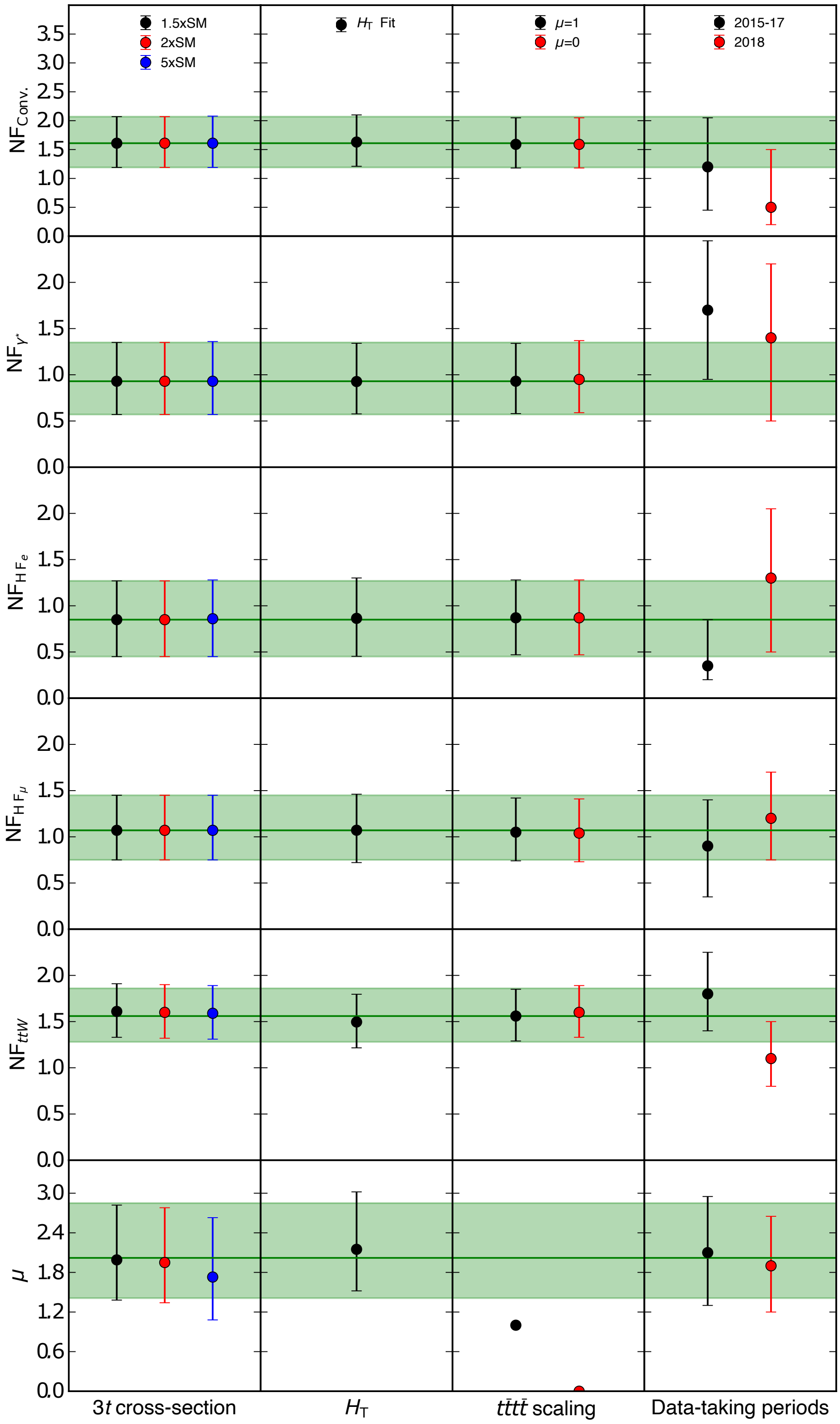


Figure 6.27: Summary plot of various post-unblinding checks. The green lines (bands) in each row show the values (total uncertainties) of the respective parameter in the nominal fit to data. Markers with error bars represent the values and total uncertainties of the corresponding parameter when the fit is repeated with the labeled variation.

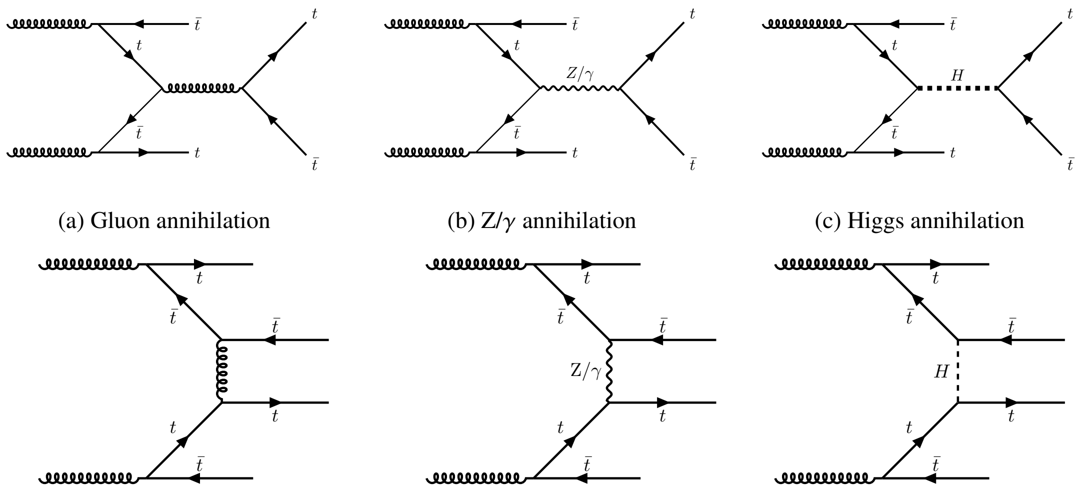
6.7 Result interpretation

As discussed and motivated in Sections 2.4 and 3.5.1, the $t\bar{t}t\bar{t}$ process can be used to measure the top-quark Yukawa coupling parameter, y_t. The CP-even case presented here follows the method used by the CMS collaboration [69], which is based on the calculations and arguments in Reference [74]. A second publication, Reference [167], considers in addition the CP-odd parametrisation and is used in this thesis to set a limit on the purely CP-odd top-quark Yukawa coupling. As explained in Chapter 1, the interpretation studies presented in this section are author's personal work except where it is explicitly mentioned, and does not represent the ATLAS collaboration. However, these interpretation results have been computed using the analysis setup and statistical fit model developed by the analysis team for the $t\bar{t}t\bar{t}$cross-section measurement described earlier in this Section.

The $t\bar{t}t\bar{t}$ cross-section can be parametrised as a function of y_t, assuming all other parameters are the same as their SM values. The equation reads [74]

$$\sigma(t\bar{t}t\bar{t}) = \sigma^{\mathrm{SM}}(t\bar{t}t\bar{t})_{g+Z/\gamma} + \kappa_t^2\sigma_{\mathrm{int}}^{\mathrm{SM}} + \kappa_t^4\sigma^{\mathrm{SM}}(t\bar{t}t\bar{t})_H. \tag{6.9}$$

The first and third term represent the gauge- (gluon/Z/γ) and Higgs-boson-mediated $t\bar{t}t\bar{t}$ processes, respectively. Example Feynman diagrams are shown in Figure 6.28. The second term accounts for the $t\bar{t}t\bar{t}$ contributions due to the interference between the processes included in the first and third terms.



(a) Gluon annihilation (b) Z/γ annihilation (c) Higgs annihilation

(d) Gluon exchange (e) Z/γ exchange (f) Higgs exchange

Figure 6.28: Example $gg \rightarrow t\bar{t}t\bar{t}$ Feynman diagrams corresponding to the processes represented by the first and third terms in Equation 6.9.

CP-even parametrisation

Three terms in Equation 6.9 have been calculated using the MadEvent event generator, computing the $t\bar{t}t\bar{t}$ cross-section at LO accuracy [74]. The associated theoretical uncertainties are estimated using the dynamical scale [168] in the same way as `tttt renorm./fact.scale` (see Section 6.5.2) are estimated. Nominal values and their respective scale uncertainties for the three cross-sections are listed in Table 6.16. Two assumptions are necessary for the interpretation of the result in terms of CP

Cross-section	μ_F, μ_R = 0.5 [fb]	Nominal [fb]	μ_F, μ_R = 2.0 [fb]
$\sigma^{\rm SM}(t\bar{t}t\bar{t})_{g+Z/\gamma}$	14.104	9.997	6.378
$\sigma^{\rm SM}_{\rm int}$	1.625	1.168	0.765
$\sigma^{\rm SM}(t\bar{t}t\bar{t})_{H}$	-2.152	-1.547	-0.999

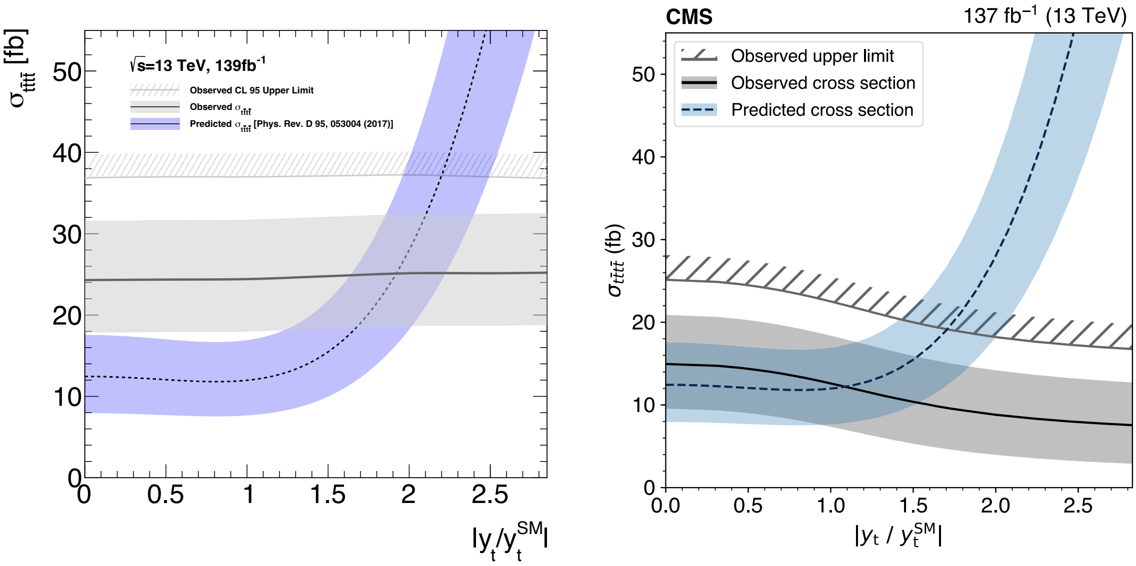
Table 6.16: $t\bar{t}t\bar{t}$ cross-section computed at LO accuracy for $\sqrt{s}$ = 13 TeV, and theoretical uncertainties estimated from scale variations. Nominal values are taken from Reference [74] and scale variations are taken from Reference [168].

properties. Both are related to the absence of dedicated MC simulation samples due to limitations of time and resources. First, the Higgs-mediated $t\bar{t}t\bar{t}$ MC simulation was not available as the used signal sample has QCD diagrams only. A small sample with Higgs-mediated four-top-quark processes was produced at a later stage of the analysis by the analysis team to check that the acceptance is similar. Second, there were no dedicated samples with modified top-quark Yukawa coupling values. As such, it was assumed that the acceptance is the same for samples with $y_t \neq y_t^{\rm SM}$.

In addition to the dependence of the $t\bar{t}t\bar{t}$ process on y_t via Higgs-boson-mediated diagrams, the $t\bar{t}H$ background process has to be considered as well, as its contribution also depend on the y_t parameter. This is parametrised via

$$\frac{\sigma_{ttH}}{\sigma^{\rm SM}_{ttH}} = \left|\frac{y_t}{y_t^{\rm SM}}\right|^2 . \tag{6.10}$$

The fit to data is thus repeated, scaling the $t\bar{t}H$ contribution by $|y_t/y_t^{\rm SM}|^2$, where $|y_t/y_t^{\rm SM}|$ assumes values between 0 and 3 with a step-size of 0.5. Values above 3 are not considered as as that would not be compatible with the measurement. Theoretical predictions are plotted in the form of black dashed lines in Figure 6.29. The coloured bands around the dashed lines are the associated scaling uncertainties. As the theoretical prediction is computed at LO accuracy, it is scaled to the NLO level prediction used by the experimental measurement in order to provide a valid comparison. The measured cross-section is drawn by interpolating between the values at the measurement points. The result is shown in Figure 6.29 in grey where the solid line and the band represent the central $t\bar{t}t\bar{t}$ cross-section value and the total uncertainty of the measurement. The grey hatched line represents the 95% CL upper limit on $\sigma(t\bar{t}t\bar{t})$. The point where the (hatched) upper limit line intersects the (dashed) central value curve of the theoretical calculation marks the point of exclusion for $|y_t/y^{\rm SM}|$. The fit model in this work exclude the top-quark Yukawa coupling at 95% CL with $|y_t/y_t^{\rm SM}| < 2.2$. A recent ATLAS (CMS) combination reports $|y_t/y_t^{\rm SM}| = 0.95 \pm 0.07$ [169]($1.01^{+0.11}_{-0.10}$ [170]). Thus, existing limits are tighter than the one measured here. The corresponding limit from the CMS measurement is $|y_t/y_t^{\rm SM}| < 1.7$ and is shown in Figure 6.29(b). In comparison to the result of the CMS collaboration, the limit estimated employing the setup of the ATLAS analysis is found to be slightly weaker and less dependent on the cross-section scaling. The scale independence seems to be regulated by the compensation between the amount of $t\bar{t}W$ and $t\bar{t}H$ events in the ttW CR where an inverse proportionality can be seen from the values on Table 6.17. This results in the stabilisation of the overall outcome of the fit, and an almost unchanged $t\bar{t}t\bar{t}$ production cross-section. The CMS analysis does not feature a CR for the $t\bar{t}W$ normalisation and only offers a CR targeting the $t\bar{t}Z$ process. The weaker scale dependence has a weaker limit as a consequence of the cross-section measurement being much larger than that of the CMS measurement.



(a) Result of this thesis based on the ATLAS measurement

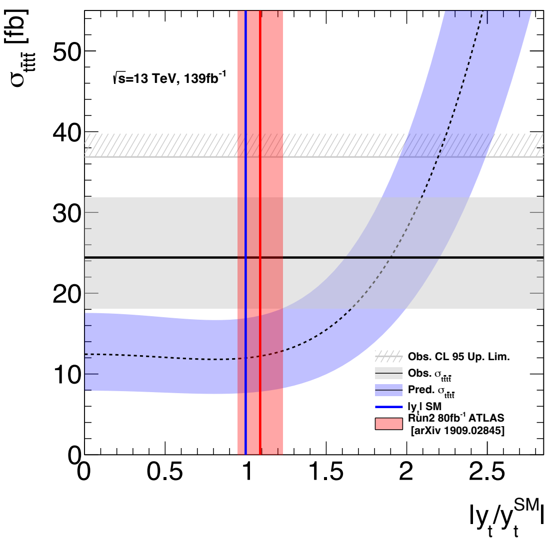
(b) Result of the CMS analysis

Figure 6.29: Comparison of y_t limits in the CP-even parametrisation of the analyses by the CMS collaboration and this thesis' results based on the ATLAS measurement. Central values are plotted in the form of black dashed lines. The coloured bands (purple on the left plot and blue on the right plot) around the dashed lines are the associated scale uncertainties. The experimental measurements are shown in grey where the solid line and the band represent the central $t\bar{t}t\bar{t}$ cross-section value and the total uncertainty of the measurement. The grey hatched line is the 95% CL exclusion upper limit of the measurement.

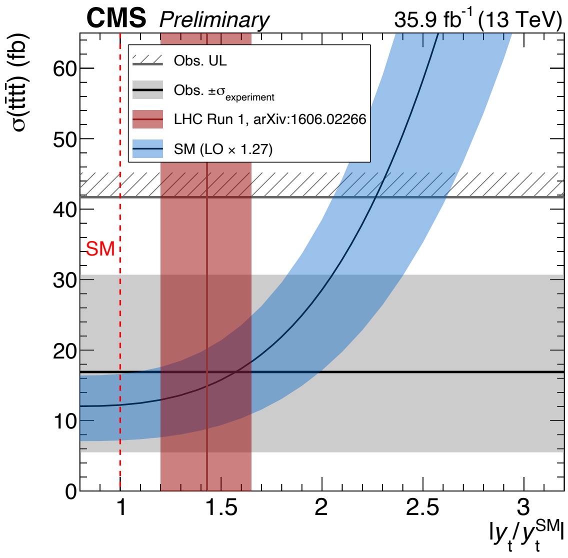
$\lvert y_t/y_t^{SM}\rvert$	$\sigma_{t\bar{t}H}$ scaled by $\lvert y_t/y_t^{SM}\rvert^2$	$\mu_{t\bar{t}t\bar{t}}$	$\mu_{t\bar{t}t\bar{t}}^{CL\,95}$	$\sigma_{t\bar{t}t\bar{t}}$	$\sigma_{t\bar{t}t\bar{t}}^{CL\,95}$	$\sigma_{t\bar{t}t\bar{t}}^{Theory}$	$NF_{t\bar{t}W}$	$t\bar{t}H$ in SR	
								Pre-Fit	Post-Fit
0.00	0.00	$2.01^{+0.81}_{-0.63}$	3.08	24.3	36.9	$12.4\ ^{+17.6}_{-7.9}$	$1.96^{+0.27}_{-0.26}$	0	0
0.50	0.25	$2.01^{+0.81}_{-0.62}$	3.08	24.3	36.9	$12.1\ ^{+17.0}_{-7.7}$	$1.87^{+0.28}_{-0.27}$	10 ± 2	10±2
1.00	1.00	$2.02^{+0.83}_{-0.62}$	3.09	24.4	36.9	$12.0\ ^{+16.9}_{-7.6}$	$1.62^{+0.28}_{-0.27}$	39 ± 9	37±9
1.50	2.25	$2.07^{+0.80}_{-0.62}$	3.10	24.8	37.1	$15.5\ ^{+21.8}_{-9.9}$	$1.26^{+0.29}_{-0.28}$	87 ± 21	74±16
2.00	4.00	$2.10^{+0.79}_{-0.61}$	3.11	25.1	37.2	$28.6\ ^{+39.2}_{-18.2}$	$0.89^{+0.31}_{-0.31}$	155 ± 37	107±20
2.50	6.25	$2.10^{+0.78}_{-0.60}$	3.09	25.1	36.9	$57.4\ ^{+79.9}_{-37.4}$	$0.55^{+0.34}_{-0.35}$	242 ± 58	132±21
3.00	9.00	$2.11^{+0.77}_{-0.58}$	3.07	25.3	36.7	$113\ ^{+157}_{-74}$	$0.25^{+0.36}_{-0.05}$	349 ± 84	151±19

Table 6.17: List of selected parameters and their values at each fit setup used for the estimation of the CP-even top-quark Yukawa coupling. CL 95 stands for the 95% confidence level.

The limit estimated above assumed variations of the on-shell top-quark Yukawa coupling between 0 and 3. Assuming that the on-shell y_t is SM-like, the off-shell coupling can be probed for deviations. In this case, no scaling is applied to the $t\bar{t}H$ process as it only has on-shell dependence on the y_t. Since the measurement yields an almost flat curve, no difference between the two cases is observed and the same limit is found. In the corresponding CMS analysis the estimated limit for the off-shell case is reported to be $|y_t/y_t^{\mathrm{SM}}| < 1.8$ [69]. The CMS analysis for the off-shell case exists only for a previous measurement using partial Run 2 data [171] whose result is shown in Figure 6.30(b). The measurement result using the ATLAS setup is shown in Figure 6.30(a).



(a) Result of this thesis based on the ATLAS measurement.



(b) Result of the preliminary CMS analysis [171].

Figure 6.30: Comparison of CP-even-only off-shell y_t parametrisation limits of the analyses by the CMS collaboration [171] and this thesis' results based on the ATLAS measurement. Central values are plotted in the form of black (a) solid (b) dashed lines. The coloured bands (purple on the left plot and dark blue on the right plot) around these lines are the associated scale uncertainties. The experimental measurements are shown in grey where the solid line and the band around it represents the central $t\bar{t}t\bar{t}$ cross-section value and the total uncertainty of the measurement. The grey hatched line is the 95% CL exclusion upper limit on $\sigma(t\bar{t}t\bar{t})$. The red lines and the same-coloured bands around them represent the results from another measurement reported in the plot legend. The SM prediction is shown by a red dashed line on the left, and by a blue solid line on the right plot. On the right plot, the experimental curve is drawn by interpolating between the values at the measurement points.

CP-odd allowing parametrisation

When also accounting for the CP-odd contributions, the $t\bar{t}t\bar{t}$ cross-section can be expressed using CP-even and CP-odd coupling parameters a_t and b_t [167]:

$$\sigma(t\bar{t}t\bar{t}) = 7.724 - 1.164a_t^2 + 2.434b_t^2 + 0.910a_t^4 + 2.183a_t^4b_t^2 + 1.424b_t^4 \text{ [fb]}. \tag{6.11}$$

Here, the cross-terms a_tb_t, a_t^3b and $a_tb_t^3$ are ignored due to their relatively small and thus negligible contributions. The terms of the equation can be broken down into the three processes distinguished in

Equation 6.9:

$$\begin{aligned}\sigma^{\rm SM}(t\bar{t}t\bar{t})_{g+Z/\gamma} &= 7.724\ [\text{fb}],\\ \sigma^{\rm SM}_{\rm int} &= -1.164a_t^2 + 2.434b_t^2\ [\text{fb}],\\ \sigma^{\rm SM}(t\bar{t}t\bar{t})_H &= +0.910a_t^4 + 2.183a_t^4b_t^2 + 1.424b_t^4\ [\text{fb}].\end{aligned} \tag{6.12}$$

It should be noted that setting $b_t = 0$ in Equation 6.12 does not completely recover the CP-even parametrisation of Equation 6.9. This is because the two publications used different event generation setups, leading to small differences. The parametrisation of the dependence of the $t\bar{t}H$ process on the top-quark Yukawa coupling expressed in Equation 6.10 picks up an additional CP-odd term and becomes

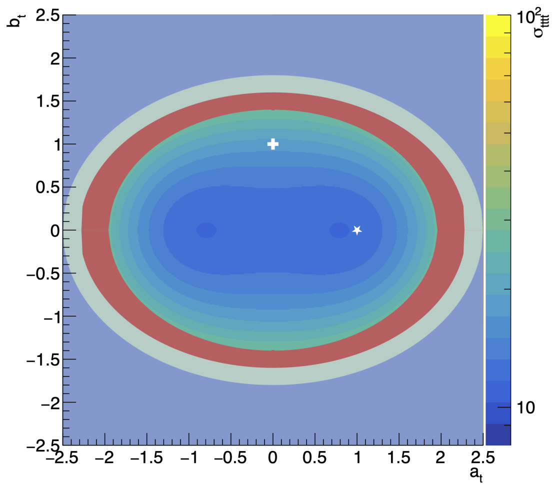
$$\frac{\sigma_{ttH}}{\sigma^{\rm SM}_{ttH}} = a^2 + 0.46b^2. \tag{6.13}$$

Repeating the fit to data using the setup of the ATLAS analysis, it can be noticed that the limit ($\approx 3.5 \times \sigma^{\rm SM}(t\bar{t}t\bar{t})$) is almost stable over the range chosen for the $t\bar{t}H$ parametrisation. Based on this observation, one limit is used for the whole range, allowing for an elliptical contour parametrisation. The ellipse equation is defined with a_t and b_t:

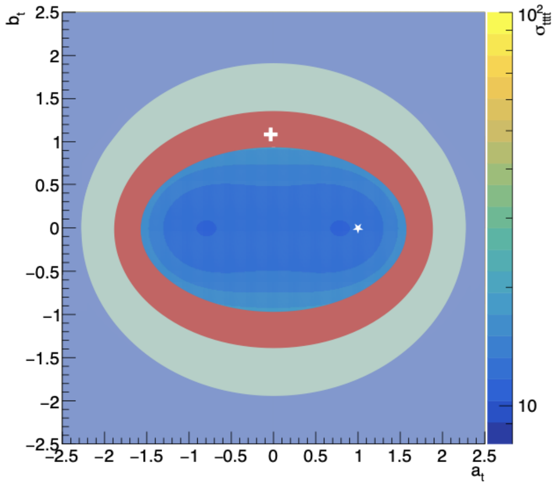
$$\frac{x^2}{a_{t0}^2} + \frac{y^2}{b_{t0}^2} = 1 \tag{6.14}$$

Here, a_{t0} and b_{t0} are the semi-major and semi-minor axes, respectively. Their values correspond to the absolute value of the purely CP-even and the purely CP-odd top-quark Yukawa coupling. The elliptic limit contour intersects the line defining the CP-odd-only coupling ($a_t = 0, b_t$) at $b_t = 1.6$. Thus, the measurement does not exclude the pure CP-odd coupling ($a_t = 0, b_t = 1$) at 95% CL. In comparison, ATLAS [172] and CMS [173] collaborations excluded pure CP-odd coupling in $t\bar{t}H \to \gamma\gamma$ channel with a significance larger than 3σ. The pure CP-even coupling ($a_t = 1, b_t = 0$) using this parametrisation is found to be slightly weaker than the results of the previous part and stands at 2.3. The results are plotted in Figure 6.31(a) where the solid green coloured area indicates the excluded region when the central value of the theoretical prediction is used. Purple and red areas correspond to the exclusion limits when using the values of lower and upper scale uncertainties of the theoretical prediction, respectively. The scale uncertainties used are the same as those reported in Table 6.16. When considering upper (lower) scale variation the limit is approximately 1.4 (1.8). The pure CP-odd coupling and SM prediction ($a_t = 1, b_t = 0$) are also shown.

For a comparison, the phenomenological result reported in Reference [167] is used. There, exclusion limits are provided for three scenarios with various integrated luminosity values: 80 fb^{-1}, 345 fb^{-1} and 680 fb^{-1}. Since the measurement is conducted using 139 fb^{-1}, the first scenario is used for comparison. The scale variations in this study are conservatively assumed to be 50%. The results are shown in Figure 6.31(b) where the solid green coloured area indicates the excluded region when the central value of the theoretical prediction is used. Purple and red areas correspond to the exclusion limits when using the 50% lower and upper scale uncertainties, respectively. The corresponding exclusion limit from the phenomenological study yields a value of approximately 1.1 for the pure CP-odd coupling in the case of using the central value of the theoretical prediction. The limit when considering upper (lower) scale variation is approximately 0.8 (1.8). Thus, in this study, pure CP-odd Yukawa coupling could be excluded when considering the theoretical prediction with the upper scale variation uncertainties.



(a) Mixed CP parametrisation result of this thesis. The light green coloured area indicates the excluded region when using the central value of the theoretical prediction. The purple and red areas correspond to the exclusion limits when using the values of lower and upper scale uncertainties, respectively.



(b) Result[4] of Reference [167] for 13 TeV corresponding to an integrated luminosity of 80 fb^{-1}. The light green coloured area indicates the excluded region when using the central value of the theoretical prediction. The purple and red areas correspond to the exclusion limits when using the 50% lower and upper scale uncertainties, respectively.

Figure 6.31: CP-mixed limit results of this thesis compared to predicted limits for an integrated luminosity of 80 fb^{-1} [167]. The contour lines represent the $t\bar{t}t\bar{t}$ production cross-section according to Equation 6.11. The + symbol represents the pure CP-odd coupling ($a_t = 0, b_t = 1$). The $\star$ symbol corresponds to the SM prediction ($a_t = 1, b_t = 0$). All limits are calculated at 95% CL.

6.8 Combination with 1/2LOS channel $t\bar{t}t\bar{t}$ measurement

The ATLAS Collaboration has published the $t\bar{t}t\bar{t}$ measurement in the 1L/2LOS channel, as well as the statistical combination of both measurements [174]. The combined measurement of the 1L/2LOS and 2LSS/3L channels is performed by a simultaneous profile-likelihood fit on the same dataset, where all fit regions are included. Due to the exclusive lepton selection, the two analyses are statistically independent. Both analyses feature the same definitions of objects and use the same dataset. The experimental uncertainties are therefore assumed to be fully correlated in the combined fit. Uncertainties related to the reducible backgrounds in the 2LSS/3L channel are treated as uncorrelated. With the exception of $t\bar{t}W$ and $t\bar{t}$+jets, theoretical modelling uncertainties for all backgrounds and the signal process are fully correlated. The exceptions are due to the relative importance and the different treatment in each of the analysis. In the 2LSS/3L analysis, the $t\bar{t}W$ process is the dominant background and its normalisation in the fit is left as a free parameter. The $t\bar{t}$+jets process in this final state is a minor contribution, most of which is from reducible background processes. In the 1L/2LOS

[4] Figure 4.a in Reference [167] is approximately redrawn here in order to match the style and colour scheme of the result of the thesis.

analysis $t\bar{t}W$ background is minor and its normalisation is directly taken from the simulation. The $t\bar{t}$+jets background is the dominant background process and subjected to data-driven corrections with various associated uncertainties. The treatment of correlations between systematic uncertainties of the two analyses in the combined fit are summarised in Table 6.18.

Uncertainty	Correlated	Comment
$t\bar{t}t\bar{t}$	Yes	
$t\bar{t}W$	No	Special treatment in 2LSS/3L, smaller in 1L/2LOS
$t\bar{t}$+jets	No	Special treatment in 1L/2LOS, smaller in 2LSS/3L
Other backgrounds	Yes	
Experimental	Yes	Same objects and dataset
Instrumental Backgrounds	No	Fake and Non-prompt lepton estimations
Data-Driven Corrections	No	QMID in 2LSS/3L and $t\bar{t}$+jets in 1L/2LOS

Table 6.18: Treatment of correlations between categories of uncertainties for the combined measurement of the $t\bar{t}t\bar{t}$ cross-section in the 1L/2LOS and 2LSS/3L channels.

The cross-section resulting from the combination is found to be [174]

$$\sigma_{t\bar{t}t\bar{t}} = 24 \pm 4\ (\text{stat.})\ ^{+5}_{-4}(\text{syst.})\ \text{fb} = 24^{+7}_{-6}\ \text{fb}, \tag{6.15}$$

corresponding to an observed (expected) significance of 4.7 (2.6[5]) standard deviations. The combination has increased the significance relative to individual channels' sensitivity. The excess in the observation is compatible with the signal hypothesis. This is demonstrated in Figure 6.32, where repeating the combined fit after increasing the expected signal prediction to the best-fit value ($\mu_{\text{fit}} = 2.2$) is shown to result in a post-fit distribution more compatible with data.

[5] The expected significance used here is different from what is used in the remainder of this work, and the same as the one used in the corresponding publication. Here, the 2LSS/3L channel alone would have an expected significance of 2.4σ.

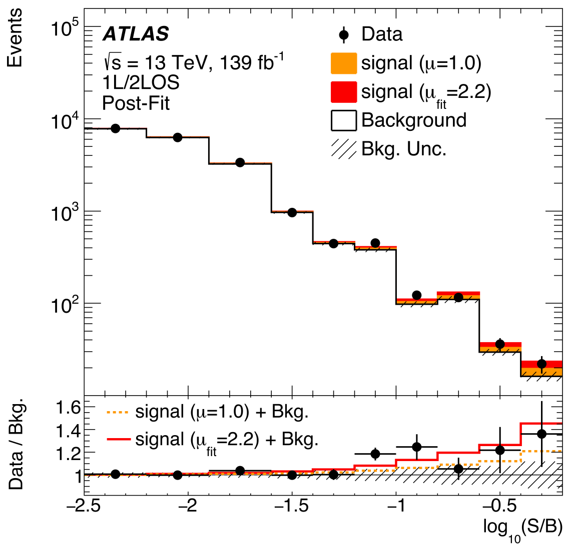


Figure 6.32: Expected post-fit event yields and observed number of events as a function of $\log_{10}$(S/B) where S (B) stands for signal (background). Events from all regions that are used in the fit are included. The case corresponding to the SM (best-fit) signal strength is shown with orange (red) colour on top of total background represented by the white colour. Uncertainty bands represent both statistical and systematical uncertainties of the background contribution. The Figure is taken from Reference [174].

CHAPTER 7

Summary and outlook

This thesis presents a measurement of the simultaneous production of four top-quarks ($t\bar{t}t\bar{t}$) in the final states with two same-charged leptons or at least three leptons. The analysis is conducted using the 139 fb^{-1} of data corresponding to full Run 2 dataset of the ATLAS detector, collected from proton–proton collisions provided by the LHC at a centre-of-mass energy of $\sqrt{s}$ = 13 TeV. The cross-section of the $t\bar{t}t\bar{t}$ process is measured and first evidence for the existence of this SM-predicted process was found.

The measured cross-section is

$$\sigma_{t\bar{t}t\bar{t}} = 24 \pm 5(\text{stat.})^{+5}_{-4}(\text{syst.})\ \text{fb} = 24^{+7}_{-6}\ \text{fb}, \tag{7.1}$$

where the reference theoretical prediction at NLO-accuracy is

$$\sigma^{\text{theory}}_{t\bar{t}t\bar{t}} = 12 \pm 2.4\ \text{fb}. \tag{7.2}$$

The measurement is found to be compatible with the SM prediction. With an observed (expected) significance of 4.3σ (2.9σ) over the background-only hypothesis, first evidence for the simultaneous production of four top-quarks is established.

To increase the sensitivity and the accuracy of the measurement, two categories of background processes are defined: irreducible processes with the same final state composition as the signal process, and reducible processes ending up in the same final state due to instrumental effects of the detector. Irreducible backgrounds are estimated from MC simulation except for the $t\bar{t}W$ process, due to the fact that this process is not well-modelled. The $t\bar{t}W$ process and two reducible background processes, namely photon conversion and non-prompt lepton contributions originating from heavy-flavour jets, are each estimated using the data assisted template fit method and constrained with control regions. Background processes due to leptons with mis-identified charges are estimated using a separate, fully data-driven method, again constrained by a dedicated control region.

Discrimination of the $t\bar{t}t\bar{t}$ process against the background processes is achieved by training a BDT, and inclusively performed in a single signal region. Two separate optimisations have been applied to the machine-learning setup. First, from a larger set of input variables, redundant variables that do not significantly contribute to the training performance are removed. The pseudo-continuous

b-tagging score is found to be the variable with the largest discrimination power. In total, twelve variables were used in the final setup. The second optimisation step tuned five selected hyperparameters of the BDT model through a grid-scan. For a set of predefined minimum and maximum values and step sizes, all possible combinations of five hyperparameters are tested in individual trainings and the best-performing setup is chosen.

Results are extracted using a binned profile likelihood fit, performed simultaneously on the inclusive signal region and the four control regions. Following the theoretical uncertainty in the $t\bar{t}t\bar{t}$ process, the largest contributing source to the measurement uncertainty is the modelling of the $t\bar{t}W$ process with 7 or more jets. The statistical combination of this measurement with another ATLAS analysis of the $t\bar{t}t\bar{t}$ process in the one lepton and two opposite-sign lepton final states improves the observed significance to 4.7σ.

The observed sensitivity of the analysis falls short of the 5σ statistical significance needed for claiming a discovery of the four top-quark production process. However, improvements to the analysis strategy and techniques together with the larger amounts of data and higher centre-of-mass energy, that will be accessible during the LHC Run 3 and the later High-Luminosity phase of the LHC (HL-LHC) give an optimistic outlook towards a future measurement. In Reference [175], several improvement scenarios have been considered for extrapolating the sensitivity of the current Run 2 analysis to the HL-LHC settings. Predictions from the current analysis are scaled-up according to expected increases in the total integrated luminosity from 139 fb^{-1} to 3000 fb^{-1}. As HL-LHC is also planned to operate at an increased centre-of-mass energy of 14 TeV, cross-section predictions of the current analysis are also scaled to account for the increase in the collision energy. Relative to 13 TeV, the cross-section of the $t\bar{t}t\bar{t}$ production process at 14 TeV is 30% larger. In comparison, important background processes such as $t\bar{t}V$+jets and $t\bar{t}$ production are enhanced by 20% only.

In the first extrapolation scenario considered in Reference [175], only the systematic uncertainty in the modelling of events from the $t\bar{t}W$ process with more than or equal to 7 jets has been modified. This systematic uncertainty is modified according to the post-fit values of the corresponding nuisance parameter found in the current analysis. In this scenario, an expected significance of 4.0σ is estimated. In the second extrapolation scenario, on top of the changes made in the first scenario, other systematic uncertainties are also modified based on assumptions on the possible improvements in the theoretical and experimental uncertainties. In this scenario, theoretical uncertainties are halved compared to the current analysis. Uncertainties associated with heavy-flavour jet and non-prompt lepton modelling have been scaled by the luminosity. Apart from the JVT uncertainties, which are halved based on the studies indicating the possibility for improvements, instrumental uncertainties are left unchanged due to the lack of studies about the performance of the detector after the upgrades. In this more detailed second scenario, a significance of 6.4σ is expected. For more details the reader is referred to Reference [175].

The result of the analysis developed for the measurement of the cross-section is interpreted in terms of the top-quark Yukawa coupling parameter strength (y_t). A simple cross-section prediction scaling is applied to the $t\bar{t}t\bar{t}$ and $t\bar{t}H$ processes since these depend on the top-quark Yukawa coupling. For a range of scaling values, the final fit of the analysis is repeated and a range of corresponding cross-section measurement results for different y_t values is obtained. The measured values are then compared to a theoretical prediction that parametrises the cross-section of the $t\bar{t}t\bar{t}$ process as a function of y_t. An upper limit is calculated from this comparison resulting in $|y_t| < 2.20$ with a 95% CL, somewhat weaker than the upper limit of $|y_t| < 1.70$ reported by the CMS collaboration. In Reference [166] it is projected that an upper limit of $|y_t| < 1.41$ might be achievable with the HL-LHC.

In a further study, the theoretical parametrisation of the $t\bar{t}t\bar{t}$ process taking CP-odd contributions into account has been used to set a limit on the purely CP-odd y_t. Having seen that adding a CP-odd parametrisation does not alter the estimates from the CP-even case, it is assumed that the relation between the two can be represented by an elliptic equation. Purely CP-odd couplings are excluded at $|y_t| < 1.62$ with a 95% CL. For comparison, in Reference [166], where the theoretical cross-section parametrisation is given, the authors estimated that purely CP-odd coupling can be excluded using 230 fb^{-1} of data at 13 TeV centre-of-mass energy. Alternatively, under the assumption of elliptic equations used in this dissertation, the projection for the CP-even limit reported in Reference [166] for the HL-LHC translates into a corresponding limit of $|y_t| < 1.18$ at 95% CL on the purely CP-odd coupling.

Top-quark Yukawa coupling parameter estimation studies could be improved by re-optimising the analysis so as to improve the outcome of the y_t limit estimation. For example, the introduction of CP-sensitive variables to the analysis optimisation or in the fit setup could improve the limits. This is also connected to the kinematic event reconstruction studies, as several important CP-sensitive variables are accessible once the event is fully reconstructed. Another idea to target the y_t parameter is to use MVAs to discriminate between BSM and SM y_t contributions. This could be combined with the existing SM signal versus background discriminant to create a 2D discriminant that would perform better.

An important simplification in this study was the use of SM-generated simulated events and scaling them in scenarios in which $y_t \neq y_t^{\mathrm{SM}}$. This approach assumed that there is no shape difference between distributions having $y_t = y_t^{\mathrm{SM}}$ and $y_t \neq y_t^{\mathrm{SM}}$ i.e. the acceptance being not affected. For a more precise study, dedicated simulated events with different y_t values need to be produced and used. This could not be done in the study presented here due to time and resource limitations.

A direct extension of the ATLAS $t\bar{t}t\bar{t}$ analysis is to include more final states, as ATLAS has e.g. not measured the $t\bar{t}t\bar{t}$ process in the all-hadronic final state. Another important aspect is the $3t$ background process, which is difficult to suppress due to close top-quark multiplicity to the $t\bar{t}t\bar{t}$ process. Since it has experimentally never been measured, a large systematic uncertainty on its cross-section had to be introduced in the present analysis. This uncertainty could be decreased once a measurement is made.

New MVA methods also hold the potential to improve the sensitivity of the analysis. In the current measurement, BDTs were used. As an alternative, deep neural networks might be able to increase the performance. New kinematic variables could help increasing the discrimination power in a direct cut or as additional input to MVAs. One way of accessing new kinematic variables is the kinematic reconstruction of the full final state of the event. In the analysis presented here, this would mean reconstructing two (three) leptonically decaying top-quarks and two (one) hadronically decaying top-quarks. The main challenge of such an endeavour is the missing information about the neutrinos. For final states with single neutrinos, an approximate reconstruction based on some assumptions can be achieved with a rather satisfactory performance. However, multiple neutrinos introduce an additional source of uncertainty as how the total missing momentum should be apportioned between the neutrinos is not known. Another important problem that requires further study is the large jet multiplicity of the $t\bar{t}t\bar{t}$ final state, even for the non all-hadronic final states such as the ones considered in this thesis. Finding the correct jet assignments with an MVA requires processing large number of possible jet combinations, pushing the limits of resources and bringing into question the applicability of such methods in practice. Symmetry preserving attention networks (SPA-net) [176] with their ability to reconstruct many-jet events without permuting over jet combinations might be useful in future studies.

In conclusion, the analysis described in this document has established the first evidence for the simultaneous production of four-top quarks. The measured production cross-section is in agreement with the theoretical prediction of the SM. The analysis model is also used to put exclusion limits on the pure CP-even and CP-odd top-quark Yukawa couplings. The established limits do not exclude the SM predictions. Future work in various directions such as probing new final states, collecting more data and utilising novel ML techniques holds the potential for improving these results.

Bibliography

[1] M. Planck, *Ueber das Gesetz der Energieverteilung im Normalspectrum*, Ann. Phys. **309** (1901) 553.

[2] ATLAS Collaboration, *Top working group cross-section summary plots, November 2022*, (2022), URL: `https://cds.cern.ch/record/2840591`.

[3] 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*, Eur. Phys. J. C **80** (2020) 1015.

[4] Ö. O. Öncel, *Measurement of the $t\bar{t}t\bar{t}$ production at 13 TeV with the ATLAS detector at the LHC*, SciPost Phys. Proc. **8** (2022) 116.

[5] CMS Collaboration, *Observation of four top quark production in proton-proton collisions at $\sqrt{s} = 13$ TeV*, Phys. Lett. B **847** (2023) 138290.

[6] ATLAS Collaboration, *Observation of four-top-quark production in the multilepton final state with the ATLAS detector*, Eur. Phys. J. C **83** (2023) 496.

[7] R. L. Workman et al. (Particle Data Group), *Review of Particle Physics*, PTEP **2022** (2022) 083C01.

[8] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory*, Addison-Wesley, 1995, URL: `https://doi.org/10.1201/9780429503559`.

[9] M. D. Schwartz, *Quantum Field Theory and the Standard Model*, Cambridge University Press, 2013, URL: `https://doi.org/10.1017/9781139540940`.

[10] S. Weinberg, *The Quantum Theory of Fields*, vol. 1, Cambridge University Press, 1995, URL: `https://doi.org/10.1017/CBO9781139644167`.

[11] D. Griffiths, *Introduction to elementary particles*, Wiley, 2008, URL: `https://doi.org/10.1002/9783527618460`.

[12] B. Andersson et al., *Parton Fragmentation and String Dynamics*, Phys. Rept. **97** (1983) 31.

[13] B. R. Webber, *A QCD Model for Jet Fragmentation Including Soft Gluon Interference*, Nucl. Phys. B **238** (1984) 492.

[14] J. Lykken and M. Spiropulu, *The future of the Higgs boson*, Phys. Today **66** (2013) 50.

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

[16] BaBar Collaboration, *Observation of CP Violation in the B^0 Meson System*, Phys. Rev. Lett. **87** (2001) 091801.

[17] LHCb Collaboration, *Observation of CP Violation in Charm Decays*, Phys. Rev. Lett. **122** (2019) 211803.

[18] LIGO Scientific Collaboration and Virgo Collaboration, *Observation of Gravitational Waves from a Binary Black Hole Merger*, Phys. Rev. Lett. **116** (2016) 061102.

[19] Super-Kamiokande Collaboration, *Evidence for Oscillation of Atmospheric Neutrinos*, Phys. Rev. Lett. **81** (1998) 1562.

[20] SNO Collaboration, *Measurement of the Rate of $\nu_e + d \to p + p + e^-$ Interactions Produced by ^8B Solar Neutrinos at the Sudbury Neutrino Observatory*, Phys. Rev. Lett. **87** (2001) 071301.

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

[22] A. de Gouvêa, *Neutrino Mass Models*, Annu. Rev. Nucl. Part. **66** (2016) 197.

[23] S. Mertens, *Direct Neutrino Mass Experiments*, Journal of Physics: Conference Series **718** (2016) 022013.

[24] Planck Collaboration, *Planck 2015 results. XIII. Cosmological parameters*, Astron. Astrophys. **594** (2016) A13.

[25] A. D. Sakharov, *Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe*, Pisma Zh. Eksp. Teor. Fiz. **34** (1967) 392.

[26] F. R. Klinkhamer and N. S. Manton, *A saddle-point solution in the Weinberg-Salam theory*, Phys. Rev. D **30** (1984) 2212.

[27] W. Bernreuther, *CP violation and baryogenesis*, Lect. Notes Phys. **591** (2002) 237.

[28] Planck Collaboration, *Planck 2013 results. XVI. Cosmological parameters*, Astron. Astrophys. **571** (2014) A16.

[29] A. G. Riess et al., *Observational evidence from supernovae for an accelerating universe and a cosmological constant*, Astron. J. **116** (1998) 1009.

[30] P. D. Mannheim, *Alternatives to dark matter and dark energy*, Prog. Part. Nucl. Phys. **56** (2005) 340.

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

[32] H. Harari, *A new quark model for hadrons*, Phys. Lett. B **57** (1975) 265.

[33] S. W. Herb et al., *Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV Proton-Nucleus Collisions*, Phys. Rev. Lett. **116** (1977) 252.

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

[35] DØ Collaboration, *Observation of the Top Quark*, Phys. Rev. Lett. **74** (1995) 2632.

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

[37] A. Giammanco and R. Schwienhorst,
Single top-quark production at the Tevatron and the LHC, Rev. Mod. Phys. **90** (2018) 035001.

[38] E. Boos and L. Dudko, *Triple top quark production in standard model*,
Int. J. Mod. Phys. A **37** (2021) 22500233.

[39] A. Deandrea and N. Deutschmann, *Multi-tops at the LHC*, JHEP **08** (2014) 134.

[40] J.A. Aguilar-Saavedra, *A minimal set of top-Higgs anomalous couplings*,
Nucl. Phys. B **821** (2009) 215.

[41] D. Buttazzo et al., *Investigating the near-criticality of the Higgs boson*, JHEP **12** (2013) 089.

[42] F. Bezrukov and M. Shaposhnikov,
Why should we care about the top quark Yukawa coupling?, JETP **120** (2015) 335.

[43] T. Plehn, *Lectures on LHC Physics*, Springer Berlin, Heidelberg, 2012,
URL: `https://doi.org/10.1007/978-3-642-24040-9`.

[44] Y. L. Dokshitzer, *Calculation of the Structure Functions for Deep Inelastic Scattering and e^+e^- Annihilation by Perturbation Theory in Quantum Chromodynamics*,
Sov. Phys. JETP **46** (1977) 1216.

[45] V. N. Gribov and L. N. Lipatov, *Deep inelastic electron scattering in perturbation theory*,
Phys. Lett. B **37** (1971) 78.

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

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

[48] ATLAS Collaboration, *The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim*, ATL-PHYS-PUB-2010-013 (2010),
URL: `https://cds.cern.ch/record/1300517`.

[49] T. Sjostrand, S. Mrenna and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*,
Comp. Phys. Comm. **178** (2008) 852.

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

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

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

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

[54] S. Agostinelli et al., *GEANT4–a simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250.

[55] R. Frederix, D. Pagani and M. Zaro, *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.

[56] M. van Beekveld, A. Kulesza and L. M. Valero, *Threshold resummation for the production of four top quarks at the LHC*, (2022), arXiv: 2212.03259 [hep-ph].

[57] G. Bevilacqua and M. Worek, *Constraining BSM physics at the LHC: four top final states with NLO accuracy in perturbative QCD*, JHEP **07** (2012) 111.

[58] J. H. Kühn, A. Scharf and P. Uwer, *Weak interactions in top-quark pair production at hadron colliders: An update*, Phys. Rev. D **91** (2015) 014020.

[59] S. Gangal, M. Stahlhofen and F. J. Tackmann, *Rapidity-dependent jet vetoes*, Phys. Rev. D **91** (2015) 054023.

[60] D. de Florian et al., *Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector*, CERN, 2017, URL: https://doi.org/10.23731/CYRM-2017-002.

[61] ATLAS Collaboration, *Analysis of $t\bar{t}H$ and $t\bar{t}W$ production in multilepton final states with the ATLAS detector*, tech. rep., CERN, 2019, URL: https://cds.cern.ch/record/2693930.

[62] G. Bevilacqua et al., *The simplest of them all: $t\bar{t}W^{\pm}$ at NLO accuracy in QCD*, JHEP **08** (2020) 043.

[63] R. Frederix and I. Tsinikos, *Subleading EW corrections and spin-correlation effects in $t\bar{t}W$ multi-lepton signatures*, Eur. Phys. J. C **80** (2020) 803.

[64] E. Alvarez et al., *Topping-up multilepton plus b-jets anomalies at the LHC with a Z' boson*, JHEP **05** (2021) 125.

[65] CMS Collaboration, *Search for standard model production of four top quarks in the lepton + jets channel in pp collisions at $\sqrt{s}$ = 8 TeV*, JHEP **11** (2014) 154.

[66] ATLAS Collaboration, *Search for production of vector-like quark pairs and of four top quarks in the lepton-plus-jets final state in pp collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector*, JHEP **08** (2015) 105.

[67] ATLAS Collaboration, *Analysis of events with b-jets and a pair of leptons of the same charge in pp collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector*, JHEP **10** (2015) 150.

[68] CMS Collaboration, *Search for the production of four top quarks in the single-lepton and opposite-sign dilepton final states in proton-proton collisions at $\sqrt{s}$ = 13 TeV*, JHEP **11** (2019) 082.

[69] 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) 75.

[70] ATLAS Collaboration, *Measurement of the $t\bar{t}t\bar{t}$ production cross section in pp collisions at $\sqrt{s}$=13 TeV with the ATLAS detector*, JHEP **11** (2021) 118.

[71] C. Arina et al.,
A comprehensive approach to dark matter studies: exploration of simplified top-philic models,
JHEP **11** (2016) 111.

[72] D. Dicus, A. Stange and S. Willenbrock, *Higgs decay to top quarks at hadron colliders*,
Phys. Lett. B **333** (1994) 126.

[73] G. Bevilacqua and M. Worek, *Constraining BSM physics at the LHC: four top final states with NLO accuracy in perturbative QCD*, JHEP **07** (2012) 111.

[74] Q. Cao, S. Chen and Y. Liu,
Probing Higgs width and top quark Yukawa coupling from $t\bar{t}H$ and $t\bar{t}t\bar{t}$ productions,
Phys. Rev. D **95** (2017) 053004.

[75] ATLAS and CMS Collaborations,
Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s}$ =7 and 8 TeV,
JHEP **08** (2016) 045.

[76] ATLAS Collaboration, *Probing the CP nature of the top–Higgs Yukawa coupling in $t\bar{t}H$ and tH events with $H \to b\bar{b}$ decays using the ATLAS detector at the LHC*,
Phys. Lett. B **849** (2024) 138469.

[77] O. S. Brüning et al., *LHC Design Report*, CERN Yellow Reports: Monographs, CERN, 2004,
URL: `https://cds.cern.ch/record/782076`.

[78] C. Llewellyn Smith, *Genesis of the Large Hadron Collider*,
Phil. Trans. Roy. Soc. Lond. A **373** (2014) 20140037.

[79] E. Laface, *Selected issues for the LHC luminosity upgrade*,
PhD thesis: Clermont-Ferrand U., 2008, URL: `https://cds.cern.ch/record/1172173`.

[80] ATLAS Collaboration,
ATLAS data quality operations and performance for 2015–2018 data-taking,
JINST **15** (2020) P04003.

[81] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s}$ = 13 TeV using the ATLAS detector at the LHC*, ATLAS-CONF-2019-021 (2019),
URL: `http://cds.cern.ch/record/2677054`.

[82] ATLAS Collaboration, *Luminosity Public Results*, URL: `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2`
(visited on 27/05/2024).

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

[84] J. Pequenao, *Computer generated image of the whole ATLAS detector*, (2008),
URL: `https://cds.cern.ch/record/1095924`.

[85] ATLAS Collaboration,
Experiment Briefing: Keeping the ATLAS Inner Detector in perfect alignment,
(2020), General Photo, URL: `https://cds.cern.ch/record/2723878`.

[86] ATLAS Collaboration, *ATLAS pixel detector electronics and sensors*, JINST **3** (2008) P07007.

[87] M. Capeans et al., *ATLAS Insertable B-Layer Technical Design Report*, CERN-LHCC-2010-013, ATLAS-TDR-19 (2010), URL: `https://cds.cern.ch/record/1291633`.

[88] ATLAS Collaboration, *Operation and performance of the ATLAS semiconductor tracker in LHC Run 2*, JINST **17** (2022) P01013.

[89] J. Pequenao, "Computer Generated image of the ATLAS calorimeter", 2008, URL: `https://cds.cern.ch/record/1095927`.

[90] ATLAS Collaboration, *Commissioning of the ATLAS Muon Spectrometer with Cosmic Rays*, Eur. Phys. J. C **70** (2010) 875.

[91] W. P. Vazquez, *The ATLAS Data Acquisition System in LHC Run 2*, J. Phys. Conf. Ser. **898** (2017) 032017.

[92] A. Buckley et al., *Implementation of the ATLAS Run 2 event data model*, J. Phys. Conf. Ser. **664** (2015) 072045.

[93] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl. Instrum. Meth. A **389** (1997) 81.

[94] J. Catmore et al., *A new petabyte-scale data derivation framework for ATLAS*, J. Phys. Conf. Ser. **664** (2015) 072007.

[95] S. Snyder and A. Krasznahorkay, *Toolkit for data reduction to tuples for the ATLAS experiment*, ATL-SOFT-PROC-2012-023 (2012).

[96] ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2*, Eur. Phys. J. C **77** (2017) 673.

[97] ATLAS Collaboration, *ATLAS event at 13 TeV - First stable beam, 3 June 2015 - run: 266904, evt: 25884805*, (2015), URL: `https://cds.cern.ch/record/2022202`.

[98] ATLAS Collaboration, *Vertex Reconstruction Performance of the ATLAS Detector at $\sqrt{s}$ = 13 TeV*, ATL-PHYS-PUB-2015-026 (2015), URL: `https://cds.cern.ch/record/2037717`.

[99] G. Borissov et al., *ATLAS strategy for primary vertex reconstruction during Run-2 of the LHC*, J. Phys. Conf. Ser. **664** (2015) 072041.

[100] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data*, JINST **14** (2019) P12006.

[101] W. Lampl et al., *Calorimeter clustering algorithms: Description and performance*, ATL-LARG-PUB-2008-002, ATL-COM-LARG-2008-003 (2008), URL: `https://cds.cern.ch/record/1099735`.

[102] ATLAS Collaboration, *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*, ATLAS-CONF-2012-047 (2012), URL: `https://cds.cern.ch/record/1449796`.

[103] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at* $\sqrt{s}$ = 13 *TeV*, Eur. Phys. J. C **79** (2019) 639.

[104] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using 2012 LHC proton–proton collision data*, Eur. Phys. J. C **77** (2017) 195.

[105] ATLAS Collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at* $\sqrt{s}$ = 13 *TeV*, Eur. Phys. J. C **81** (2021) 578.

[106] G. P. Salam, *Towards jetography*, Eur. Phys. J. C **67** (2010) 637.

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

[108] ATLAS Collaboration, *Performance of b-jet identification in the ATLAS experiment*, JINST **11** (2016) P04008.

[109] 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) 970.

[110] 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) 903.

[111] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at 13 TeV with the ATLAS detector*, Phys. Rev. D **96** (2017) 072002.

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

[113] J. H. Friedman, *Data Analysis Techniques for High-Energy Particle Physics*, DOE-ER-13065-526 ; UR-1048 (1974) 271.

[114] J. Neyman and E. S. Pearson, *IX. On the problem of the most efficient tests of statistical hypotheses*, Phil. Trans. R. Soc. Lond. **231** (1933) 289.

[115] A. Wald, *Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large*, Trans. Amer. Math. Soc. **54** (1943) 426.

[116] A. Höcker et al., *TMVA - Toolkit for Multivariate Data Analysis*, (2007), arXiv: `physics/0703039`.

[117] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, Chapman and Hall, 1993, URL: `https://doi.org/10.1201/9780429246593`.

[118] R. Barlow, *Application of the Bootstrap resampling technique to Particle Physics experiments*, MAN-HEP-99-4 (2000).

[119] K. Cranmer, *Practical Statistics for the LHC*, 2015, arXiv: `1503.07622 [physics.data-an]`.

[120] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling for Top2016*, ATL-PHYS-PUB-2016-020 (2016), URL: https://cds.cern.ch/record/2216168/.

[121] C. Reißel, *Monte-Carlo-Simulation und Analyse des $t\bar{t}H$ Prozesses mit dem ATLAS Experiment bei*, MA thesis: University of Göttingen, 2016, URL: https://www.uni-goettingen.de/de/document/download/a57836b84faf94202790a093a200e643.pdf/Bachelorarbeit_Christina_Reissel.pdf.

[122] E. Bothmann et al., *Event generation with Sherpa 2.2*, SciPost Phys. **7** (2019) 034.

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

[124] A. Denner, S. Dittmaier and L. Hofer, *Collier: A fortran-based complex one-loop library in extended regularizations*, Comp. Phys. Comm. **212** (2017) 220.

[125] S. Hoeche et al., *A critical appraisal of NLO+PS matching methods.*, JHEP **09** (2012) 049.

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

[127] S. Hoeche et al., *QCD matrix elements and truncated showers.*, JHEP **05** (2009) 053.

[128] S. Catani et al., *QCD matrix elements+parton showers.*, JHEP **11** (2001) 063.

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

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

[131] D. de Florian et al., *Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector*, (2017).

[132] V. Barger, W. Keung, B. Yencho, *Triple-top signal of new physics at the LHC*, Phys. Lett. B **687** (2010) 70.

[133] S. Schumann and F. Krauss, *A Parton shower algorithm based on Catani-Seymour dipole factorisation*, JHEP **03** (2008) 38.

[134] ATLAS Collaboration, *Selection of jets produced in 13 TeV proton-proton collisions with the ATLAS detector*, ATLAS-CONF-2015-029 (2015), URL: https://cds.cern.ch/record/2037702.

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

[136] ATLAS Collaboration, *Recommendations of the Physics Objects and Analysis Harmonisation Study Groups 2014*, ATL-PHYS-INT-2014-018 (2014), URL: https://cds.cern.ch/record/1743654.

[137] ATLAS Collaboration, *Vertex Reconstruction Performance of the ATLAS Detector at $\sqrt{s}$ = 13 TeV*, ATL-PHYS-PUB-2015-026 (2015), URL: https://cds.cern.ch/record/2037717.

[138] ATLAS Collaboration,
Performance of electron and photon triggers in ATLAS during LHC Run 2,
Eur. Phys. J. C **80** (2019) 47.

[139] ATLAS Collaboration, *Performance of the ATLAS muon triggers in Run 2*,
JINST **15** (2020) P09015.

[140] ATLAS Collaboration, *MCTruthClassifier Tool*, (Accessed on: 26/3/2023), URL:
`https://twiki.cern.ch/twiki/bin/view/AtlasProtected/MCTruthClassifier`.

[141] ATLAS Collaboration,
Analysis of $t\bar{t}H$ and $t\bar{t}W$ production in multilepton final states with the ATLAS detector,
ATLAS-CONF-2019-045 (2019), URL: `https://cds.cern.ch/record/2693930`.

[142] 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) 39.

[143] E. K. Filmer, *Searching for Four Top Quarks with the ATLAS Detector*,
MA thesis: University of Adelaide, 2020, URL: `https://hdl.handle.net/2440/126066`.

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

[145] R. D. Field, Y. Kanev and M. Tayebnejad,
Topological analysis of the top quark signal and background at hadron colliders,
Phys. Rev. D **55** (1997) 5685.

[146] P. Glaysher, J. M. Katzy and S. An, *Iterative subtraction method for Feature Ranking*, 2019,
arXiv: `1906.05718 [physics.data-an]`.

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

[148] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s}$ = 13 TeV using the ATLAS detector at the LHC*, ATLAS-CONF-2019-021 (2019),
URL: `https://cds.cern.ch/record/2677054`.

[149] 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) 292.

[150] ATLAS Collaboration, *Alignment of the ATLAS Inner Detector in Run 2*,
Eur. Phys. J. C **80** (2020) 1194.

[151] ATLAS Collaboration, *Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s}$ = 13 TeV*, ATL-PHYS-PUB-2015-015 (2015),
URL: `https://cds.cern.ch/record/2037613`.

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

[153] ATLAS Collaboration, *Measurement of b-tagging Efficiency of c-jets in $t\bar{t}$ Events Using a Likelihood Approach with the ATLAS Detector*, ATLAS-CONF-2018-001 (2018),
URL: `https://cds.cern.ch/record/2306649`.

[154] ATLAS Collaboration, *Calibration of light-flavour b-jet mistagging rates using ATLAS proton-proton collision data at $\sqrt{s}$ = 13 TeV*, ATLAS-CONF-2018-006 (2018), URL: `https://cds.cern.ch/record/2314418`.

[155] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s}$ =8 TeV using the ATLAS detector*, Eur. Phys. J. C **79** (2019) 970.

[156] ATLAS Collaboration, *Measurement of the production cross-section of a single top quark in association with a Z boson in proton–proton collisions at 13 TeV with the ATLAS detector*, Phys. Lett. B **780** (2018) 557.

[157] F. Demartin et al., *tWH associated production at the LHC*, Eur. Phys. J. C **77** (2017) 34.

[158] ATLAS Collaboration, *Analysis of $t\bar{t}H$ and $t\bar{t}W$ production in multilepton final states with the ATLAS detector*, ATLAS-CONF-2019-045 (2019), URL: `https://cds.cern.ch/record/2693930`.

[159] 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 TeV with the ATLAS detector*, JHEP **04** (2019) 46.

[160] ATLAS Collaboration, *TRExFitter Framework*, URL: `https://trexfitter-docs.web.cern.ch/trexfitter-docs` (visited on 07/05/2023).

[161] D. Kirkby and W. Verkerke, *The RooFit Toolkit for Data Modeling*, URL: `http://roofit.sourceforge.net` (visited on 07/05/2023).

[162] L. Moneta et al., *The RooStats Project*, (2011), arXiv: `1009.1003 [physics.data-an]`.

[163] K. Cranmer et al., *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, (2012), URL: `https://cds.cern.ch/record/1456844`.

[164] F. James and M. Roos, *Minuit - a system for function minimization and analysis of the parameter errors and correlations*, Comp. Phys. Comm. **10** (1975) 343.

[165] T. P. Calvet, *Search for the production of a Higgs boson in association with top quarks and decaying into a b-quark pair and b-jet identification with the ATLAS experiment at LHC*, PhD thesis: Aix-Marseille U., 2017, URL: `https://cds.cern.ch/record/2296985`.

[166] F. Blekman et al., *Four-top quark physics at the LHC*, Universe **8** (2022) 638.

[167] Q. Cao et al., *Limiting top quark-Higgs boson interaction and Higgs-boson width from multitop productions*, Phys. Rev. D **99** (2019) 113003.

[168] Private communication with Qing-Hong, Shao-Long and Yandong Liu, (April, 2020).

[169] CMS Collaboration, *A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery*, Nature **607** (2022) 60.

[170] ATLAS Collaboration, *A portrait of the Higgs boson by the CMS experiment ten years after the discovery*, Nature **607** (2022) 52.

[171] CMS Collaboration, *Search for the standard model production of four top quarks with same-sign and multilepton final states in proton-proton collisions at* $\sqrt{s} = 13$ TeV, CMS-PAS-TOP-17-009 (2017), URL: `https://cds.cern.ch/record/2284599`.

[172] ATLAS Collaboration, *CP Properties of Higgs Boson Interactions with Top Quarks in the* $t\bar{t}H$ *and* tH *Processes Using* $H \to \gamma\gamma$ *with the ATLAS Detector*, Phys. Rev. Lett. **125** (2020) 061802.

[173] CMS Collaboration, *Measurements of* $t\bar{t}H$ *Production and the CP Structure of the Yukawa Interaction between the Higgs Boson and Top Quark in the Diphoton Decay Channel*, Phys. Rev. Lett. **125** (2020) 061801.

[174] ATLAS Collaboration, *Measurement of the* $t\bar{t}t\bar{t}$ *production cross-section in pp collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, JHEP **11** (2021) 118.

[175] ATLAS Collaboration, *Extrapolation of ATLAS sensitivity to the measurement of the Standard Model four top quark cross section at the HL-LHC*, ATL-PHYS-PUB-2022-004 (2022), URL: `https://cds.cern.ch/record/2801400`.

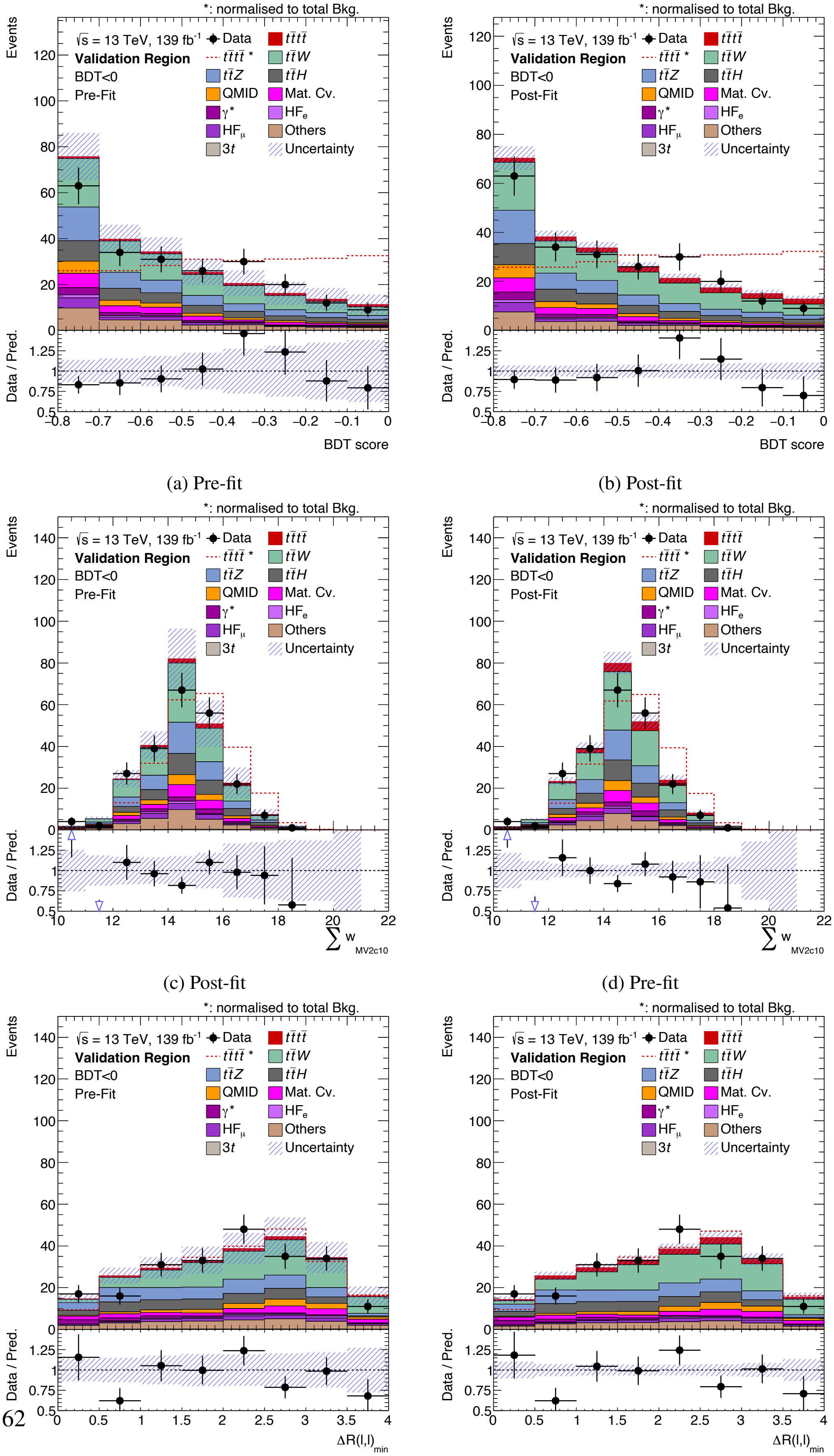
[176] M. J. Fenton et al., *Permutationless many-jet event reconstruction with symmetry preserving attention networks*, Phys. Rev. D **105** (2022) 112008.

Appendix

APPENDIX **A**

Input variable modelling in the low and high BDT validation regions

In this Appendix, modelling of input variables (described and listed in Table 6.12) used in the training of the MVA discriminant are shown in the low (BDT<0) and high (BDT>0) BDT validation regions.

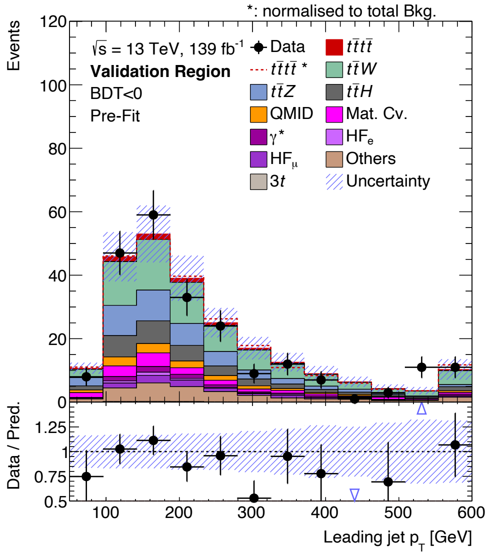


(a) Pre-fit (b) Post-fit

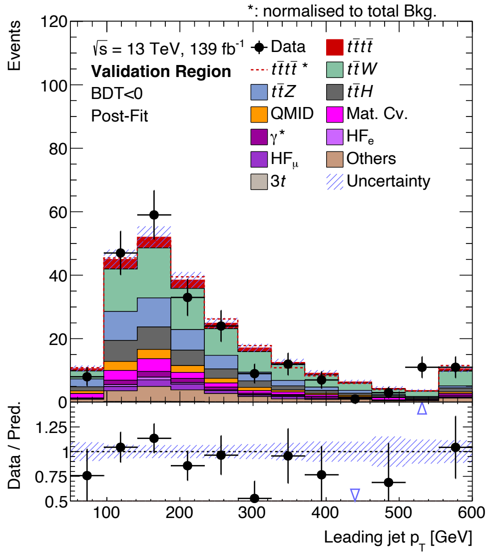
(c) Post-fit (d) Pre-fit

(e) Post-fit (f) Post-fit

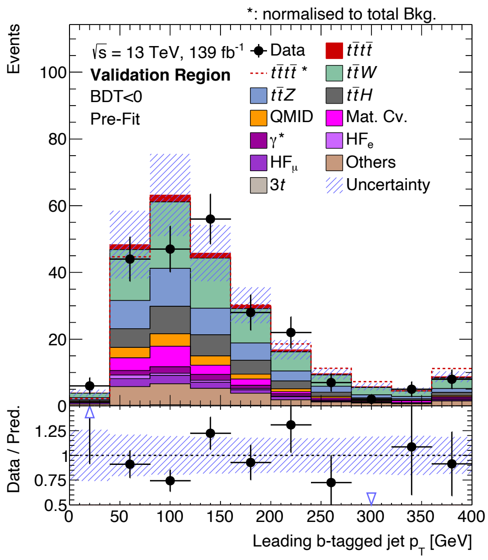
Figure A.1: Pre-fit and Post-Fit distributions in the Low BDT VR using the data fit. Dashed red line shows the signal distribution normalised to total event yield. Shaded bands represent both, statistical and systematic uncertainties.



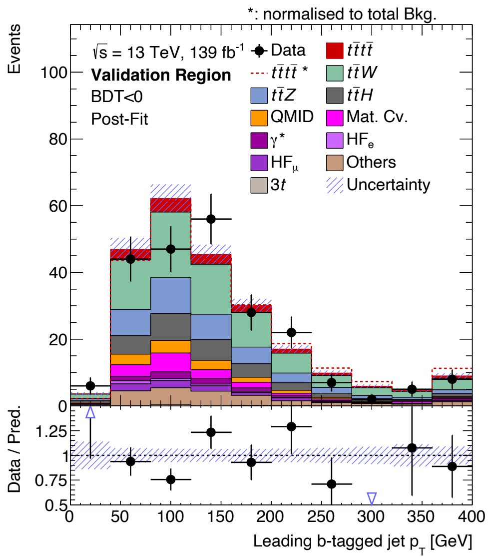
(a) Pre-fit



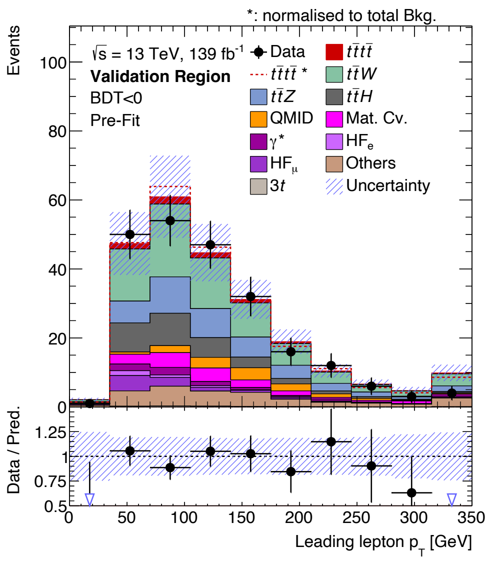
(b) Post-fit



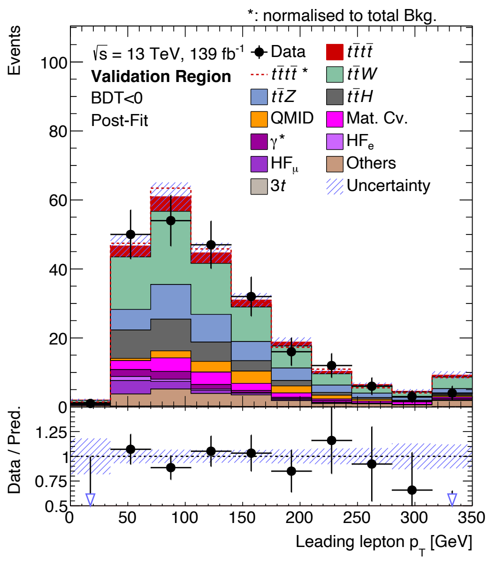
(c) Post-fit



(d) Pre-fit

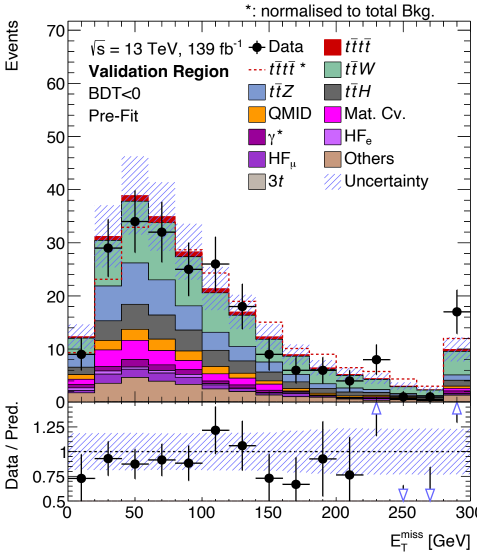


(e) Post-fit

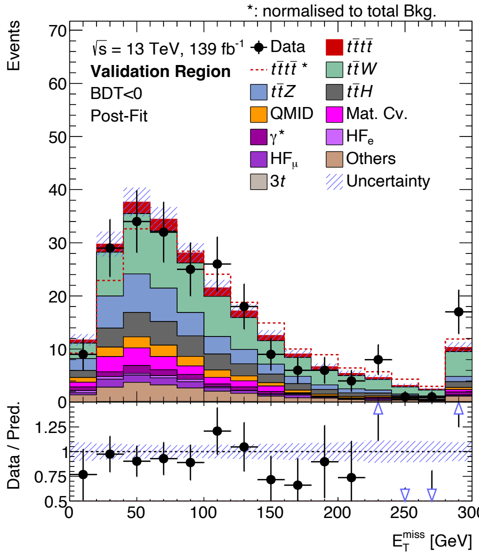


(f) Post-fit

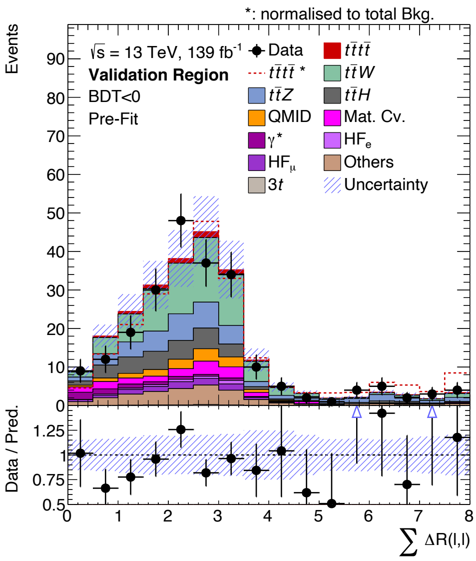
Figure A.2: Pre-fit and Post-Fit distributions in the Low BDT VR using the data fit. Dashed red line shows the signal distribution normalised to total event yield. Shaded bands represent both, statistical and systematic uncertainties.



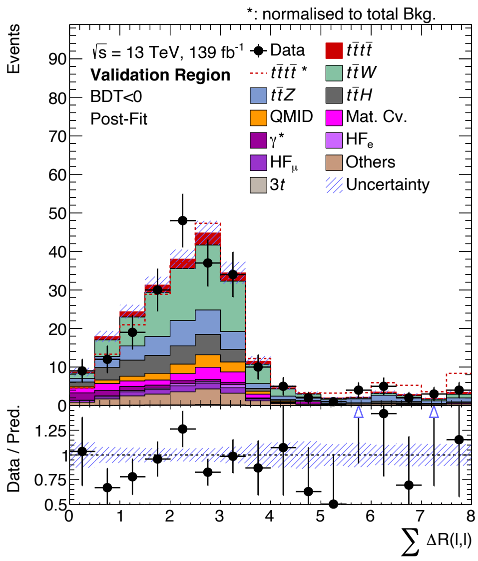
(a) Pre-fit



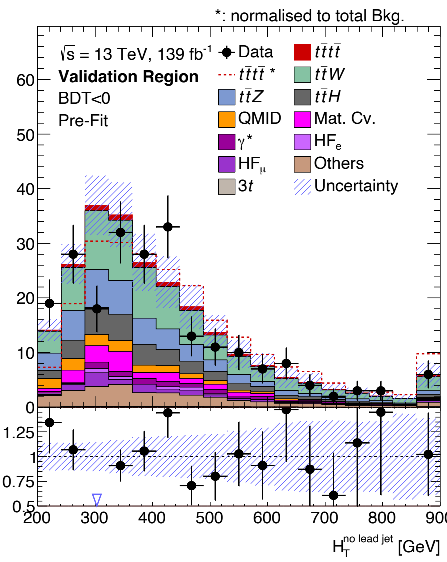
(b) Post-fit



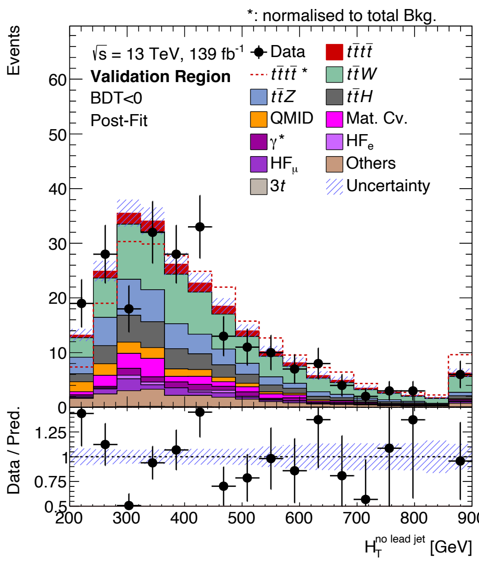
(c) Post-fit



(d) Pre-fit

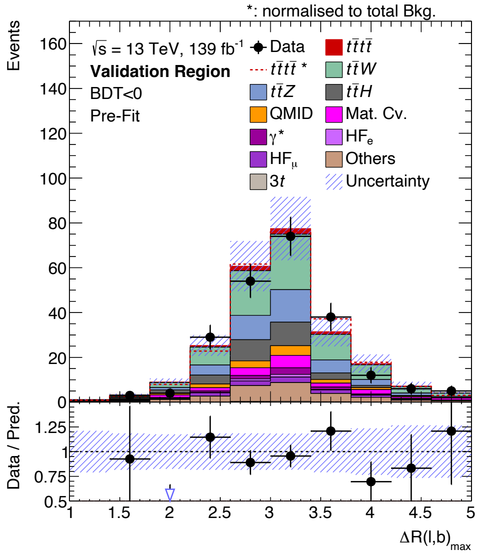


(e) Post-fit

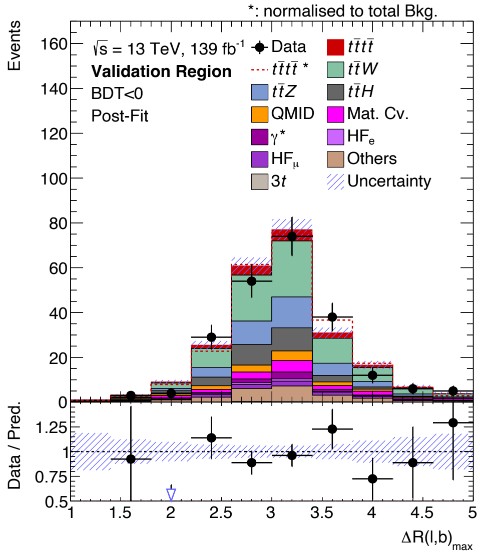


(f) Post-fit

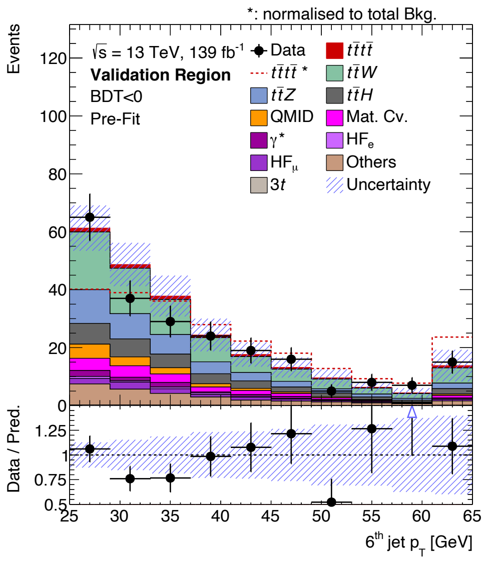
Figure A.3: Pre-fit and Post-Fit distributions in the Low BDT VR using the data fit. Dashed red line shows the signal distribution normalised to total event yield. Shaded bands represent both, statistical and systematic uncertainties.



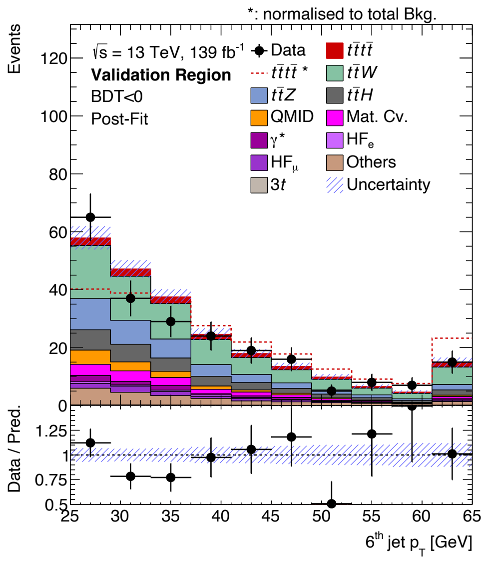
(a) Pre-fit



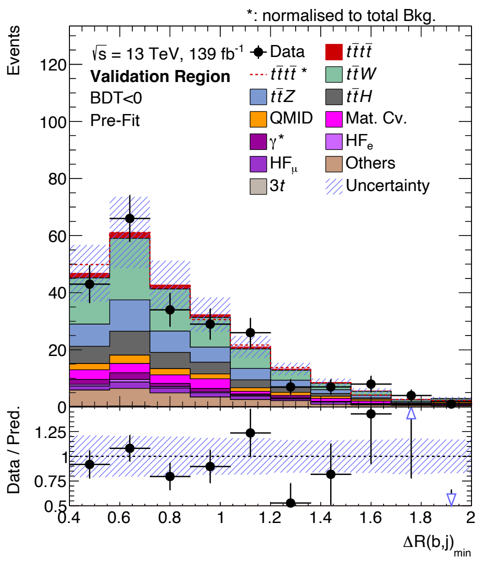
(b) Post-fit



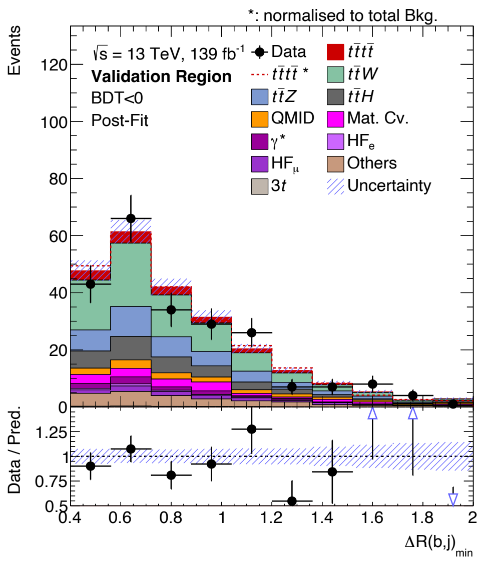
(c) Post-fit



(d) Pre-fit

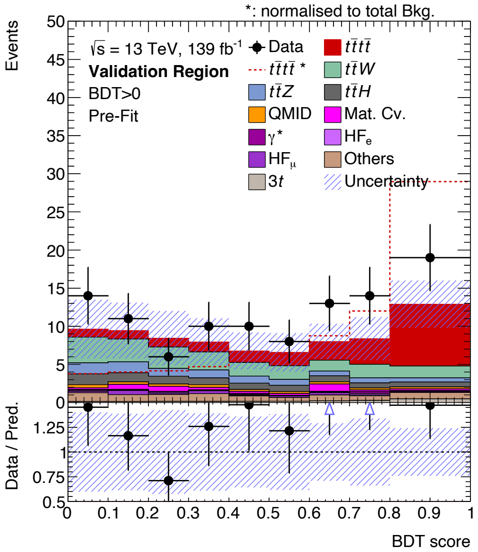


(e) Post-fit

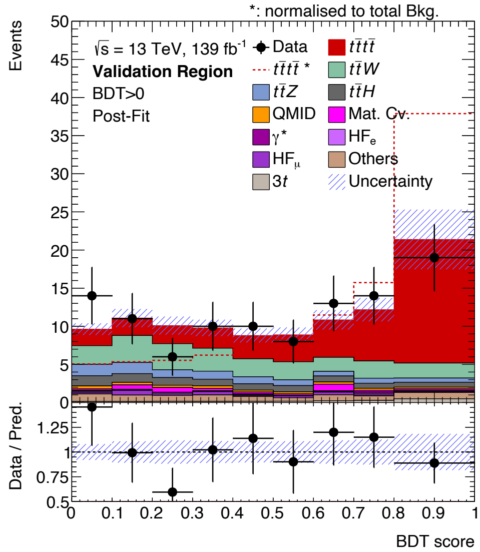


(f) Post-fit

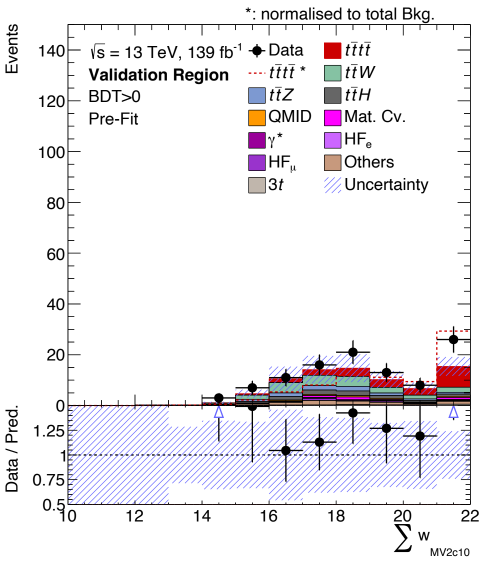
Figure A.4: Pre-fit and Post-Fit distributions in the Low BDT VR using the data fit. Dashed red line shows the signal distribution normalised to total event yield. Shaded bands represent both, statistical and systematic uncertainties.



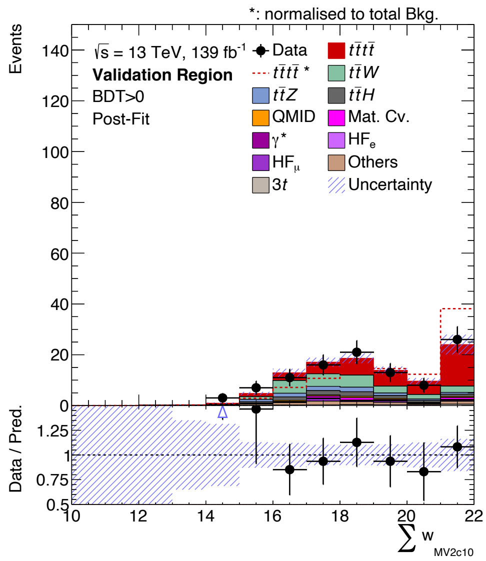
(a) Pre-fit



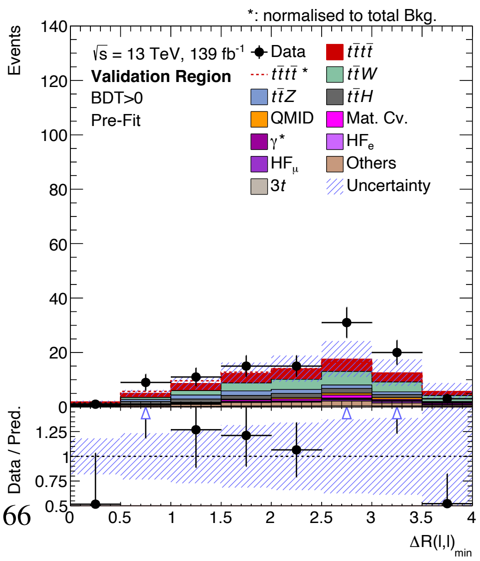
(b) Post-fit



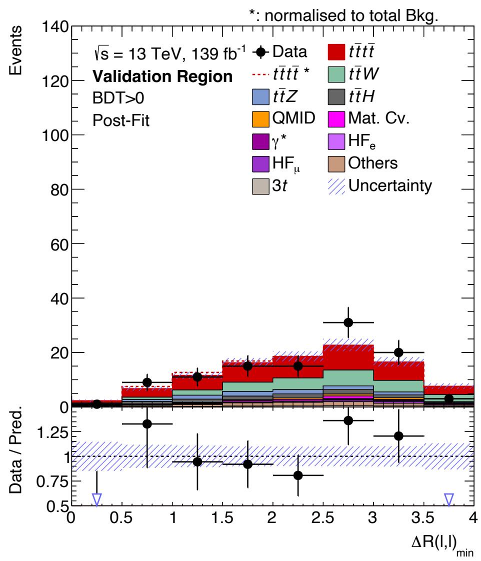
(c) Post-fit



(d) Pre-fit

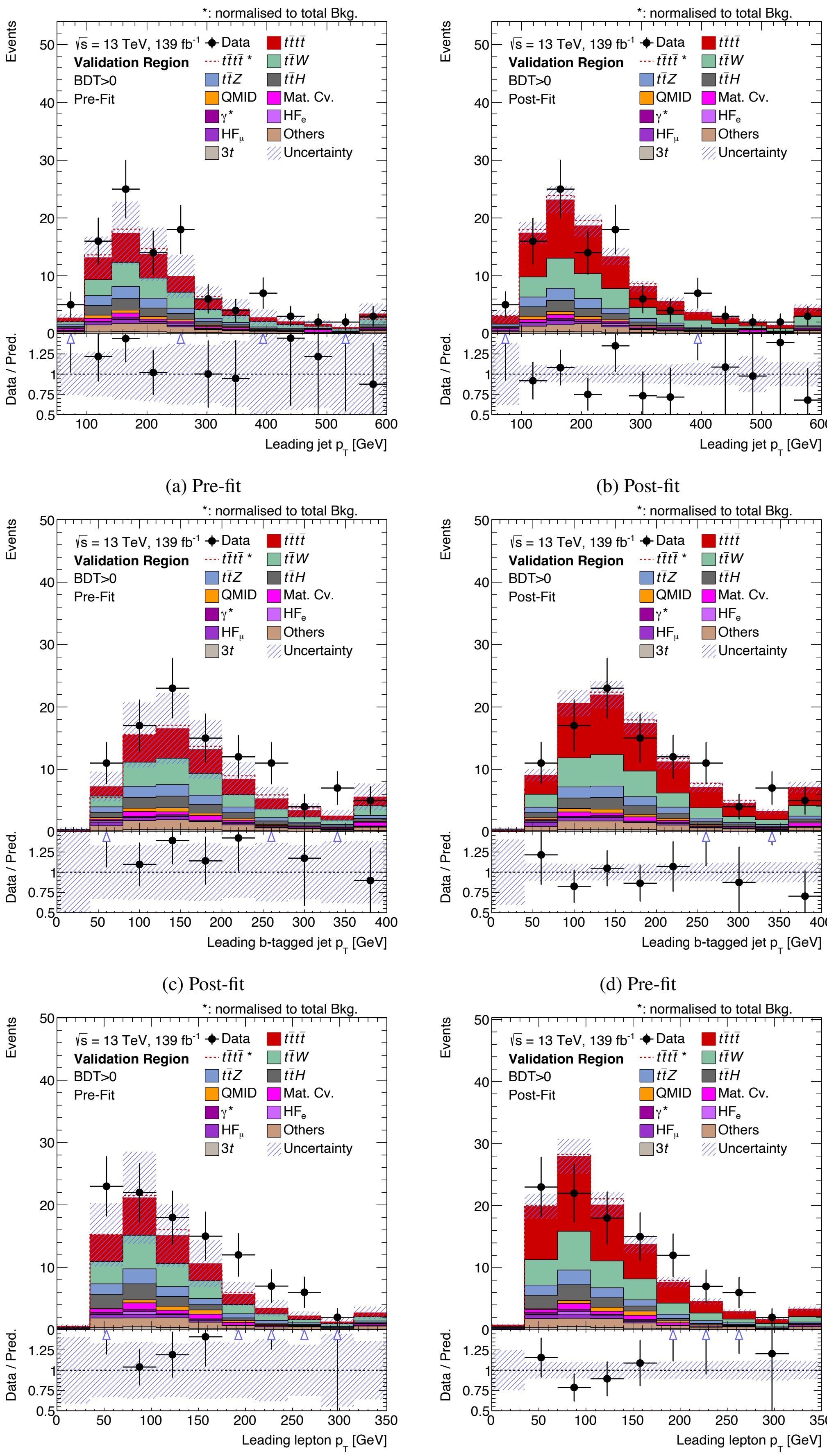


(e) Post-fit



(f) Post-fit

Figure A.5: Pre-fit and Post-Fit distributions in the High BDT VR using the data fit. Dashed red line shows the signal distribution normalised to total event yield. Shaded bands represent both, statistical and systematic uncertainties.

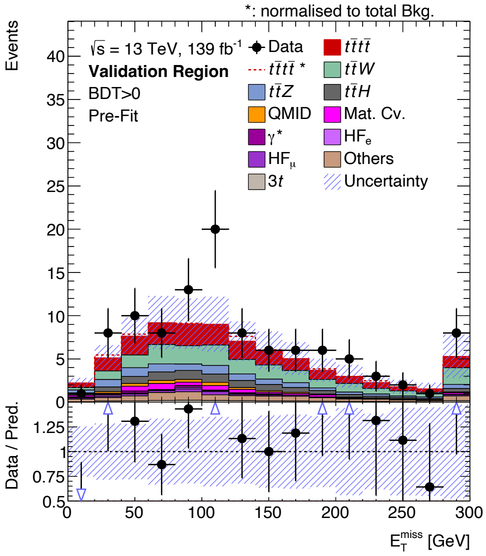


(a) Pre-fit (b) Post-fit

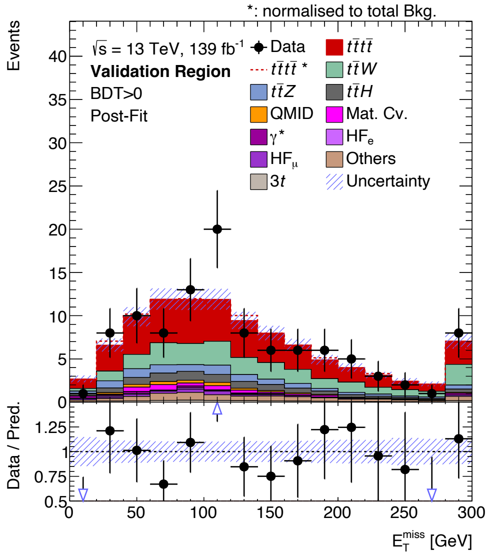
(c) Post-fit (d) Pre-fit

(e) Post-fit (f) Post-fit

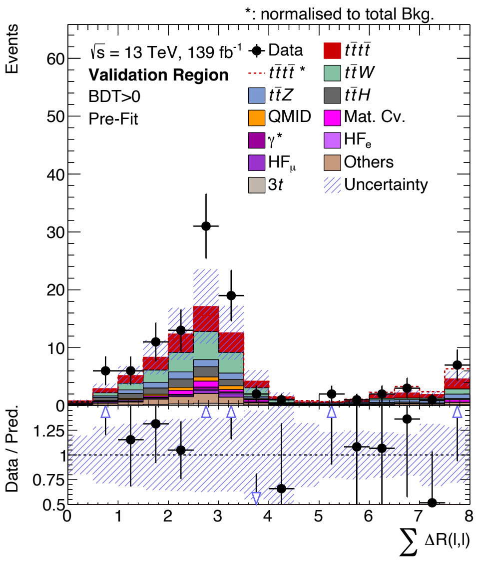
Figure A.6: Pre-fit and Post-Fit distributions in the High BDT VR using the data fit. Dashed red line shows the signal distribution normalised to total event yield. Shaded bands represent both, statistical and systematic uncertainties.



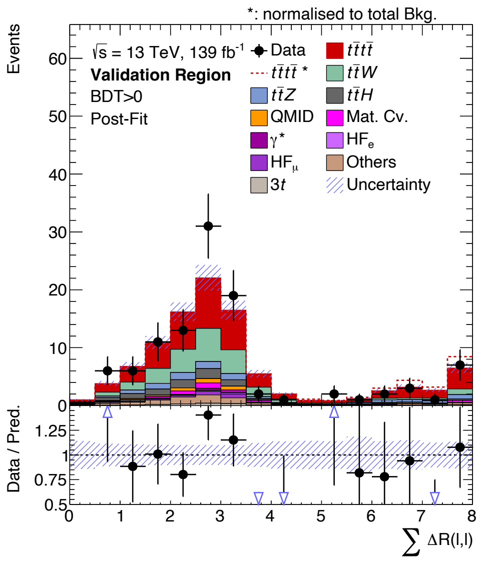
(a) Pre-fit



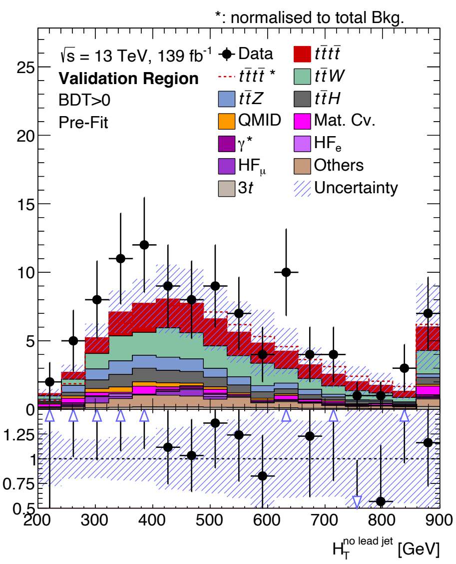
(b) Post-fit



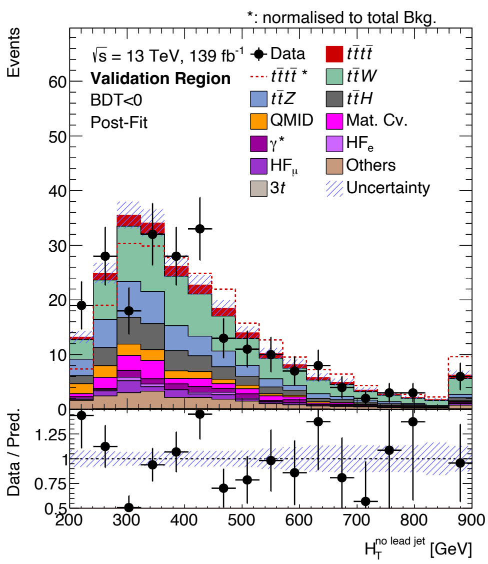
(c) Post-fit



(d) Pre-fit

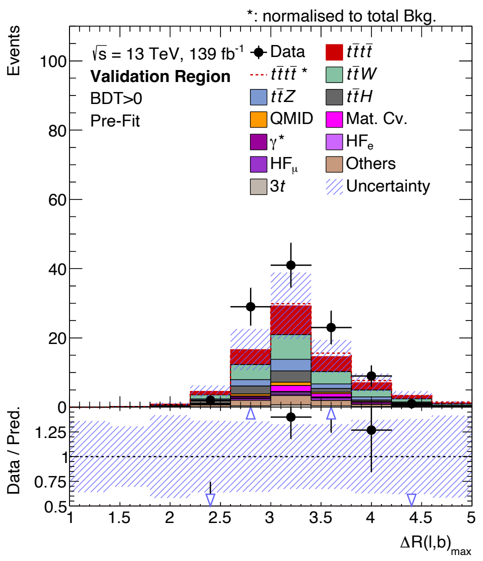


(e) Post-fit

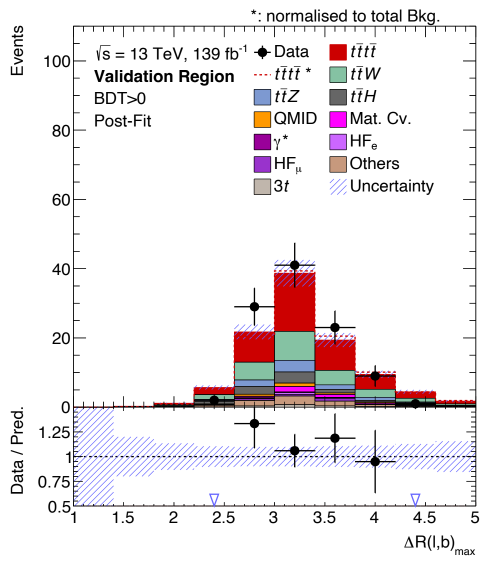


(f) Post-fit

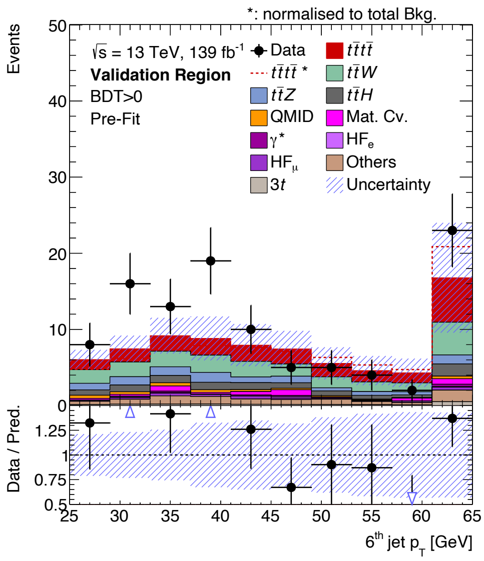
Figure A.7: Pre-fit and Post-Fit distributions in the High BDT VR using the data fit. Dashed red line shows the signal distribution normalised to total event yield. Shaded bands represent both, statistical and systematic uncertainties.



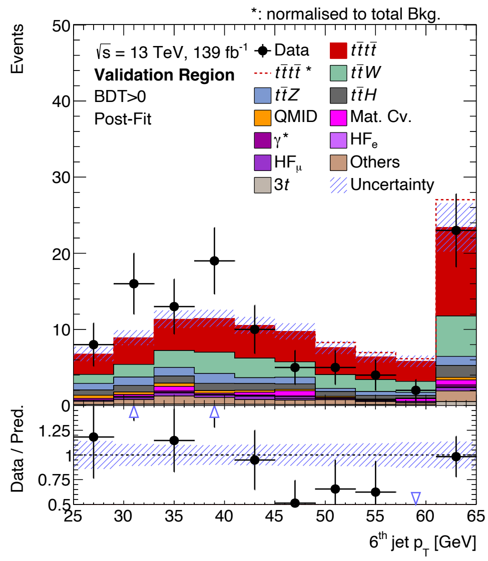
(a) Pre-fit



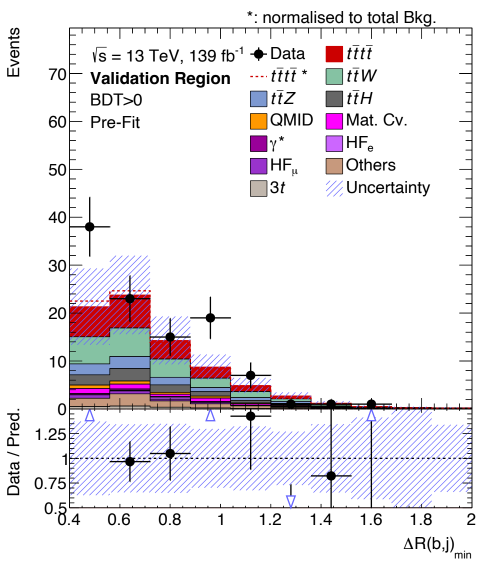
(b) Post-fit



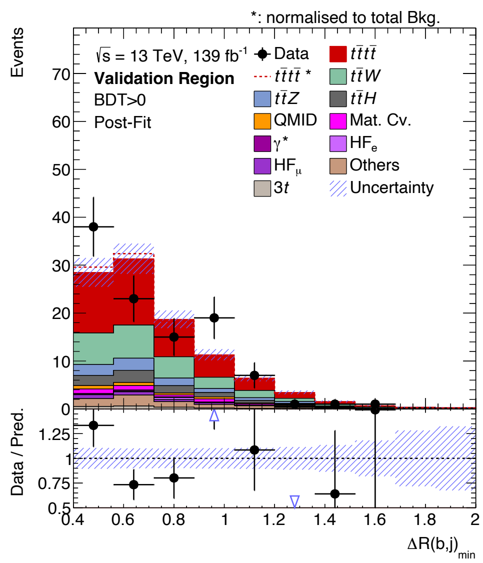
(c) Post-fit



(d) Pre-fit



(e) Post-fit



(f) Post-fit

Figure A.8: Pre-fit and Post-Fit distributions in the High BDT VR using the data fit. Dashed red line shows the signal distribution normalised to total event yield. Shaded bands represent both, statistical and systematic uncertainties.

APPENDIX B

Detailed parameters of the fit to data

In this Appendix, full set of nuisance parameters consider in the fit to data are shown. For the nuisance parameters with a correlation coefficient larger than 0.1, correlation factors are also reported in the form of a matrix.

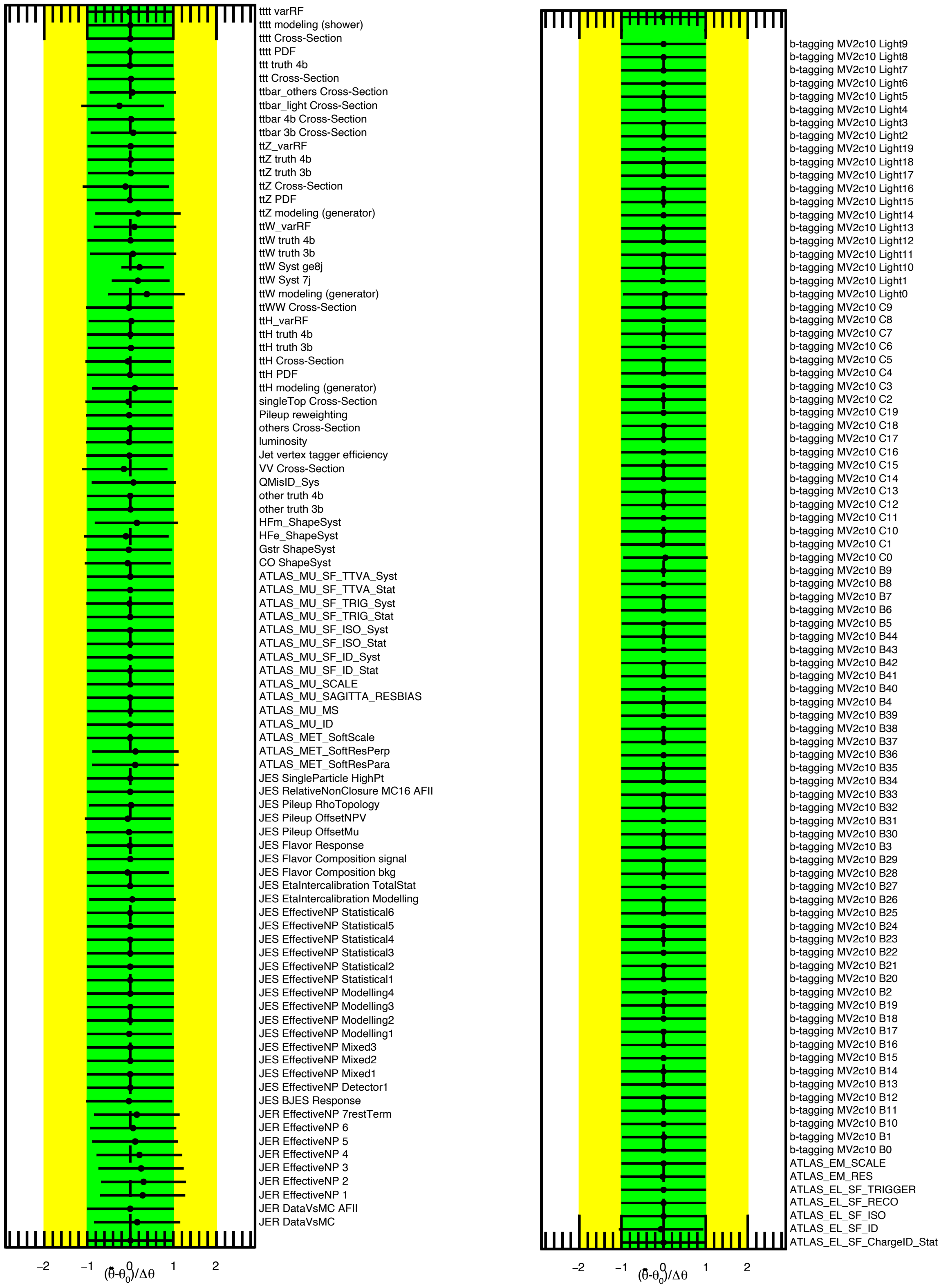


Figure B.1: Nuisance parameter central values and pulls for the data fit.

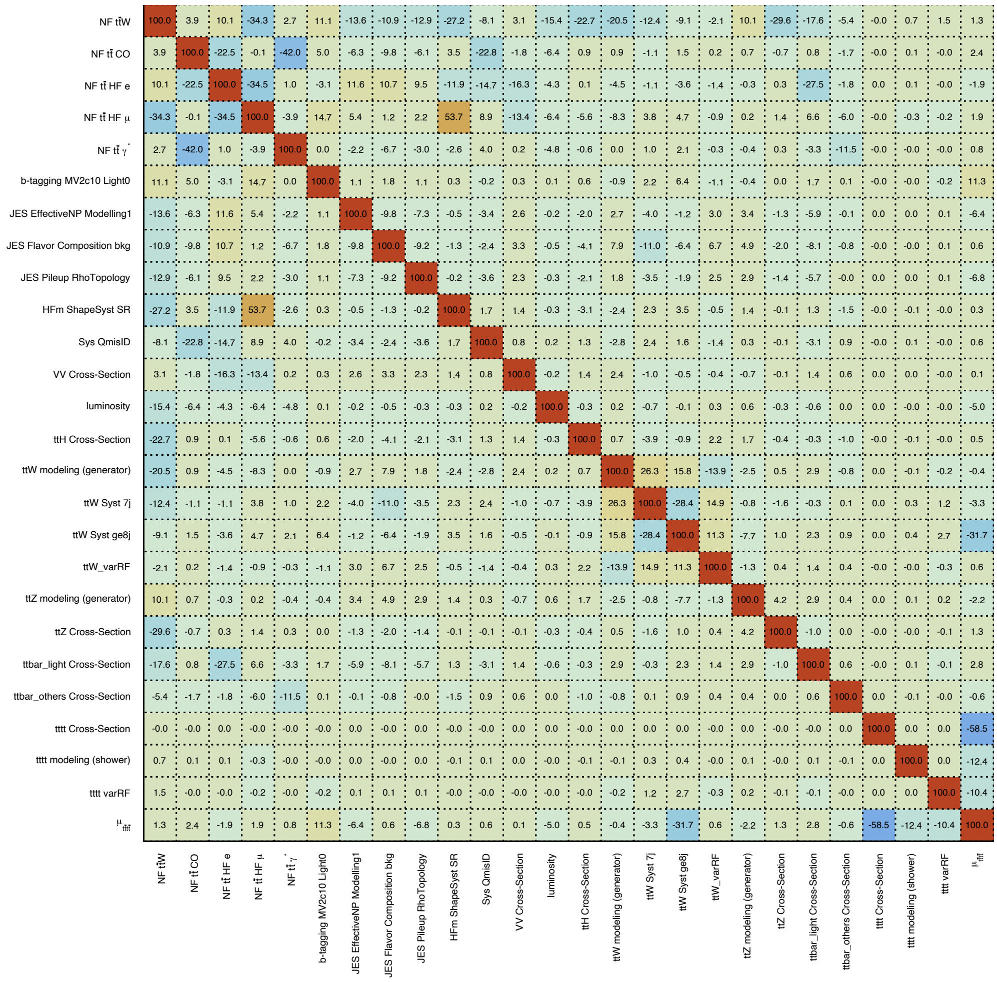
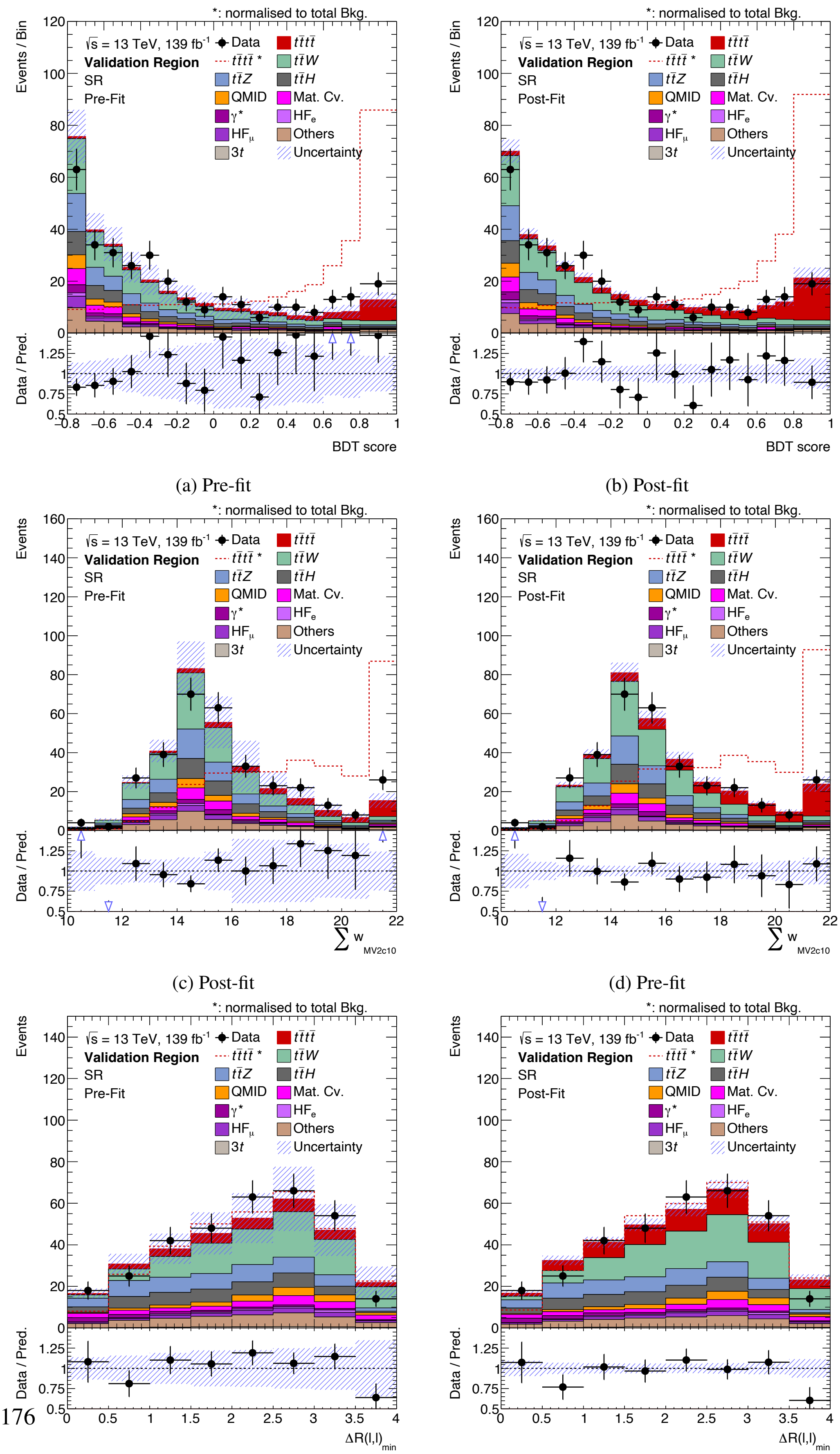


Figure B.2: Correlation matrix of the nuisance parameters for the data fit. Only parameters having a correlation coefficient larger than 0.1 are shown.

APPENDIX C

Input variable modelling in the signal region

In this Appendix, modelling of input variables (described and listed in Table 6.12) used in the training of the MVA discriminant are shown in the signal region which corresponds to the full BDT distribution without any selection on the MVA discriminant value.



(a) Pre-fit

(b) Post-fit

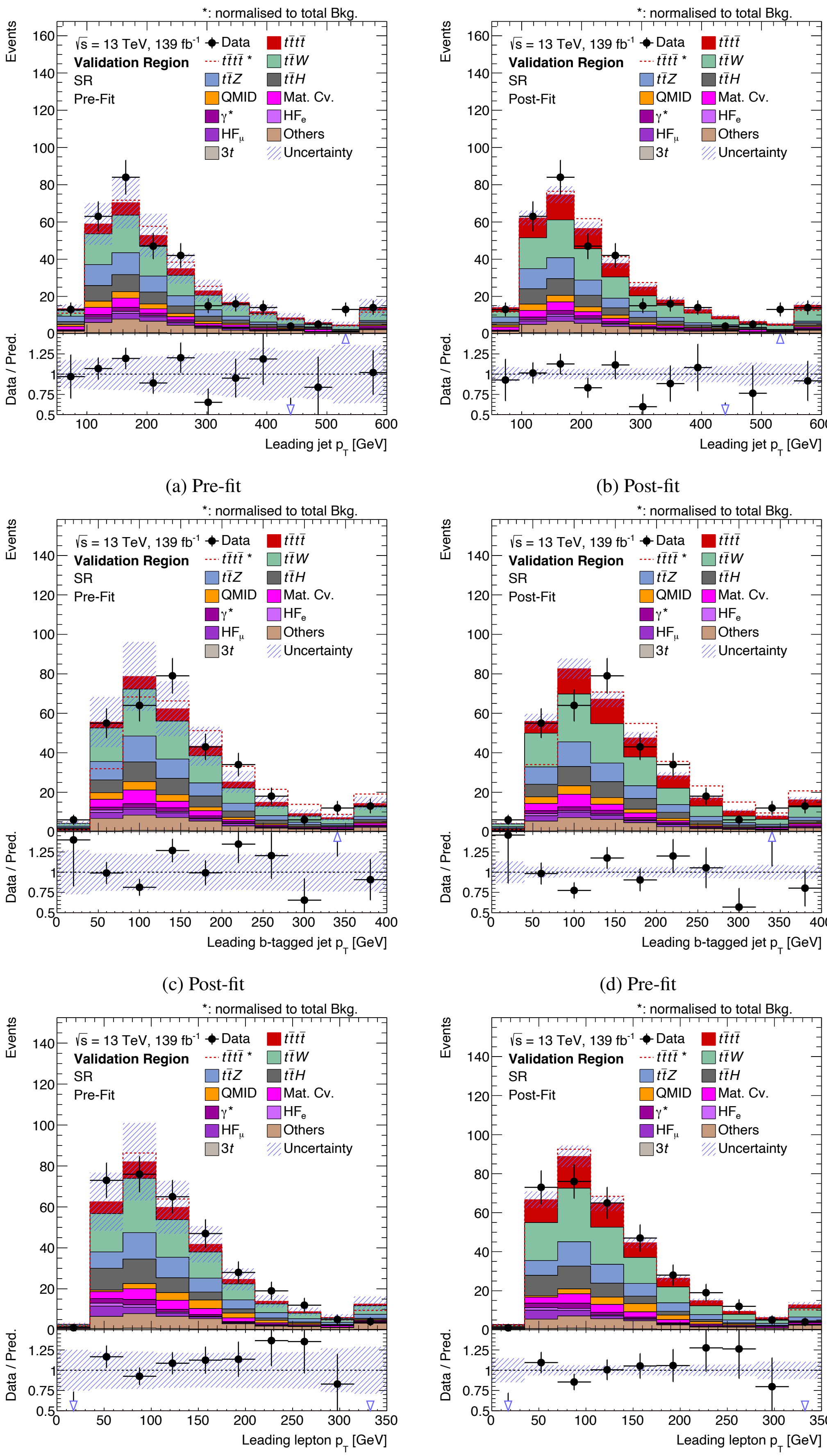
(c) Post-fit

(d) Pre-fit

(e) Post-fit

(f) Post-fit

Figure C.1: Pre-fit and Post-Fit distributions in the Full BDT SR using the data fit. Dashed red line shows the signal distribution normalised to total event yield. Shaded bands represent both statistical and systematic uncertainties.



(a) Pre-fit

(b) Post-fit

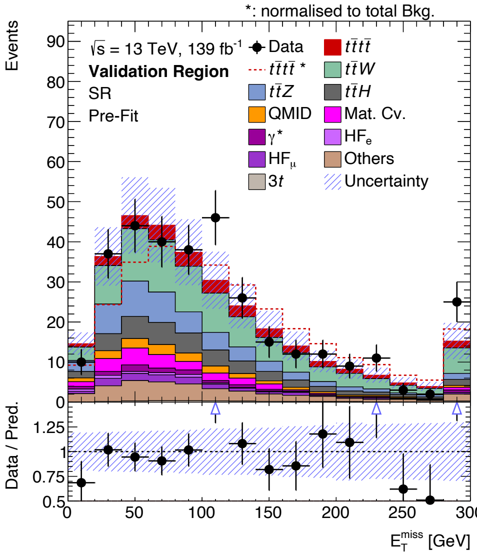
(c) Post-fit

(d) Pre-fit

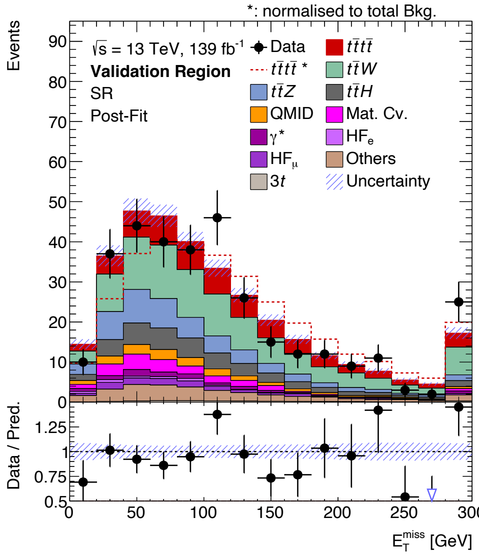
(e) Post-fit

(f) Post-fit

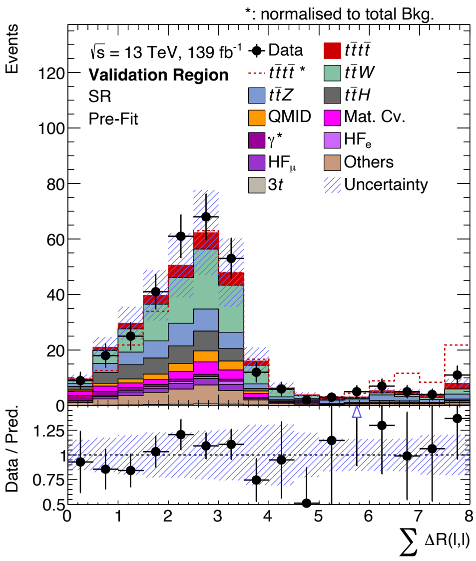
Figure C.2: Pre-fit and Post-Fit distributions in the Full BDT SR using the data fit. Dashed red line shows the signal distribution normalised to total event yield. Shaded bands represent both statistical and systematic uncertainties.



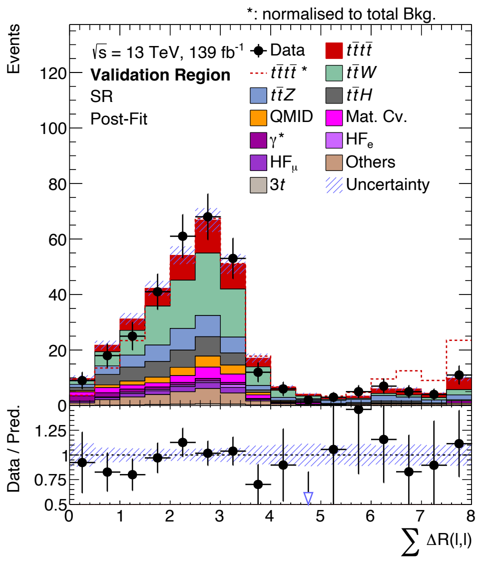
(a) Pre-fit



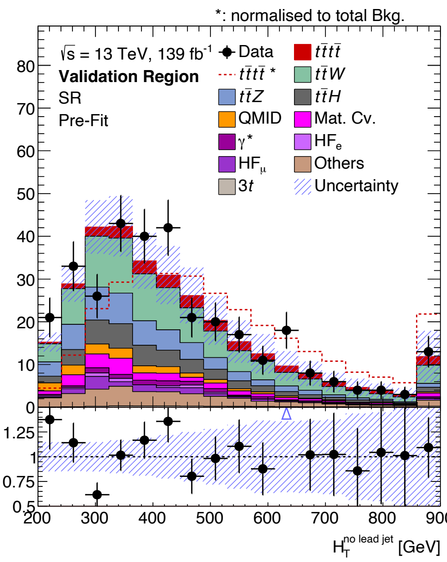
(b) Post-fit



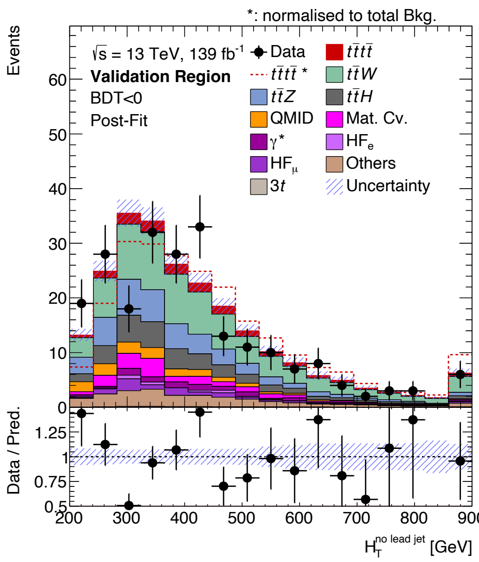
(c) Post-fit



(d) Pre-fit

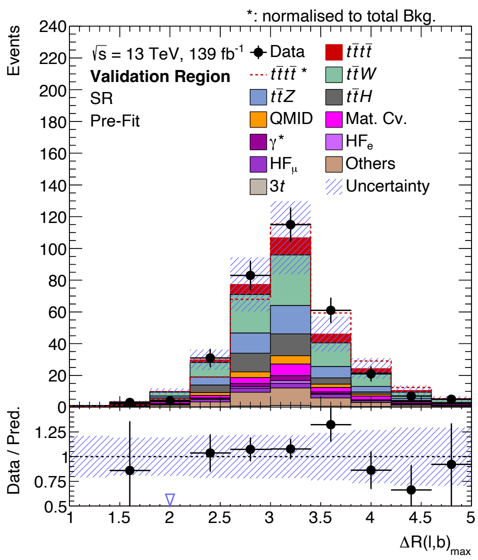


(e) Post-fit

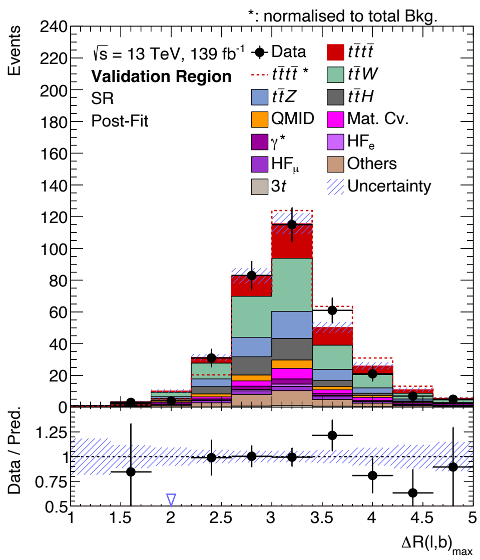


(f) Post-fit

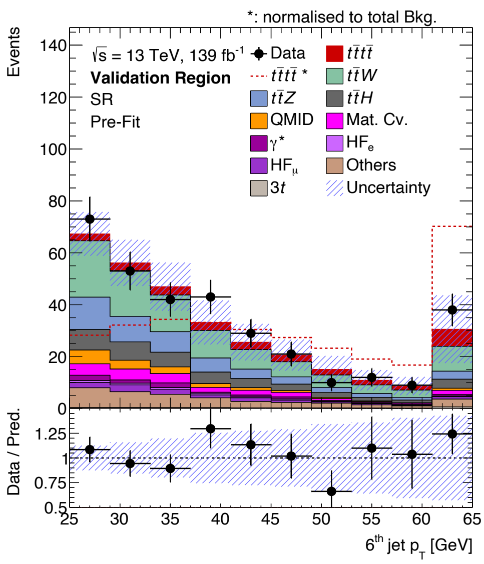
Figure C.3: Pre-fit and Post-Fit distributions in the Full BDT SR using the data fit. Dashed red line shows the signal distribution normalised to total event yield. Shaded bands represent both statistical and systematic uncertainties.



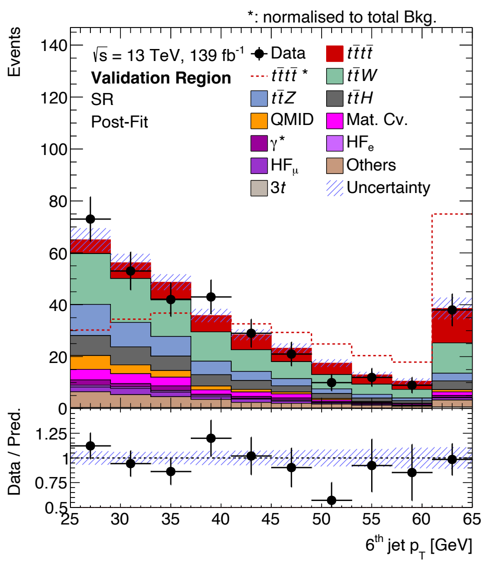
(a) Pre-fit



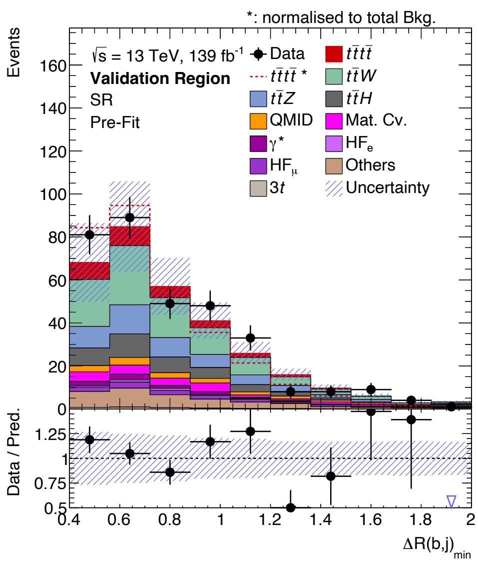
(b) Post-fit



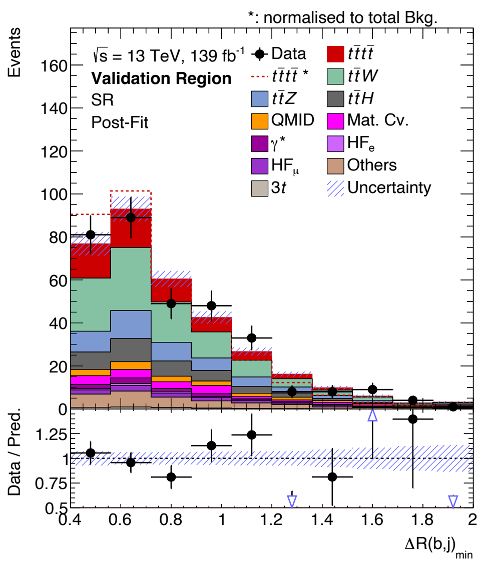
(c) Post-fit



(d) Pre-fit



(e) Post-fit



(f) Post-fit

Figure C.4: Pre-fit and Post-Fit distributions in the Full BDT SR using the data fit. Dashed red line shows the signal distribution normalised to total event yield. Shaded bands represent both statistical and systematic uncertainties.