

SILESIAN UNIVERSITY IN OPAVA

Faculty of Philosophy and Science in Opava

Santu Mondal

Study programme: Physics

Area: Theoretical Physics and Astrophysics

**Procesy zahrnující top kvarky ve srážkách proton-proton
a proton-olovo**

**Processes involving top quarks in proton-proton and
proton-lead collision**

Dissertation Thesis

Opava 2024

Dissertation supervisor:
prof. Ing. Peter Lichard, DrSc.
Consultant: Mgr. Jiří Kvita, Ph.D.

DOCTORAL THESIS DESCRIPTION

Academic year: 2023/2024

Submission institution:	Centrum interdisciplinárních studií
Student:	Santu Mondal, M.Sc.
Personal identification number:	47099
Programme:	Physics
Field of Study:	Theoretical Physics and Astrophysics
Thesis topic:	Processes involving top quarks in proton-proton and proton-lead collision
Thesis topic in English:	Processes involving top quarks in proton-proton and proton-lead collision
Description:	<p>This assignment involves two analyses in high-energy physics. The first part focuses on the ttH production in multilepton final states using proton-proton collisions at $\sqrt{s}=13$ TeV with an integrated luminosity of $80\sim 1/\text{fb}$. The primary objectives are to determine the cross-section for ttH production and estimate the fake rate for the specific final state comprising two same-sign leptons and one hadronic tau lepton. The second part centres on top-quark pair production in proton-lead collisions using $165\sim 1/\text{nb}$ of p+Pb data collected at $\sqrt{s}=8.16$ TeV. The goal is to measure the final cross-section by combining results from the lepton+jets and dilepton final states.</p>
Literature:	<ol style="list-style-type: none">1. ATLAS Collaboration, Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC, Phys. Lett. B, 716 (2012) 1, DOI:10.1016/j.physletb.2012.08.0202. CMS Collaboration, Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC, Phys. Lett. B, 716 (2012) 30, DOI:10.1016/j.physletb.2012.08.0213. 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 $\sqrt{s}=7, 8$ TeV, JHEP, 08 (2016) 045, DOI:10.1007/JHEP08(2016)0454. ATLAS Collaboration, Measurement of the production and lepton charge asymmetry of W bosons in Pb+Pb collisions at $\sqrt{s}=2.76$ TeV with the ATLAS detector, Eur. Phys. J. C, 75 (2015) 23, DOI:10.1140/epjc/s10052-014-3230-45. R. D. Ball et al., Parton distributions for the LHC Run II, JHEP, 04 (2015) 040, DOI:10.1007/JHEP04(2015)040

Thesis supervisor: prof. Ing. Peter Lichard, DrSc.

Thesis assignment date: 2024/04/26

I agree with the description (signature, date):

.....
RNDr. Tomáš Gráf, Ph.D.
department head

Declaration

Hereby I declare that I have prepared this Final Thesis independently using only the literature listed in the bibliography and on the basis of consultation with the supervisor. I have listed all the sources from which I drew when writing the work.

The author agrees with the publication of this final thesis in the sense of § 47b Act No. 111/1998 Coll. (On Higher Education Institutions and on Amendments to Other Acts) as amended.

Opava, June 24th 2024

.....

Santu Mondal

*Dedicated to my beloved uncles:
Late Shri Jahar Mondal and Shri Tapan Mondal*

Acknowledgement

To all those who have illuminated my path and fuelled my perseverance, your support has made this thesis a testament to our shared journey...

First and foremost, I would like to express my heartfelt gratitude to my supervisor, Prof. Peter Lichard. Thank you for believing in my decisions and instincts, and for giving me the freedom to navigate through the challenges and triumphs of my research journey. I am immensely grateful to Jiří Kvita, my co-supervisor, for his unwavering support throughout my doctoral studies. Jiří, your guidance has been instrumental in shaping my Ph.D. experience. Your encouragement to think creatively, tackle problems diligently and keep the big picture in mind has been invaluable. Our collaborative efforts, especially in heavy ion analysis and top quark discovery, have not only yielded significant results but also provided profound lessons in forward thinking.

I would like to express my sincere gratitude to Prof. Iwona Grabowska-Bold for her invaluable guidance and constant support. Her influential guidance helped me to complete our HI analysis, which provided me with crucial opportunities during my Ph.D. studies.

I am genuinely appreciative of Stanislav Hledík, Josef Juráň, Filip Blaschke, and Martin Blaschke from the Silesian University in Opava. Their guidance and support were irreplaceable throughout my studies. Not only did they help me navigate the challenging coursework, but they also played a crucial role in shaping my overall learning experience. I would also like to express my sincere thanks to Stanislav Hledík for his generosity in providing the thesis template and his continuous guidance throughout the process, which greatly contributed to the completion of my thesis.

I would like to express my earnest gratitude to the ATLAS group at the Institute of Experimental and Applied Physics (IEAP) of the Czech Technical University in Prague. I am deeply indebted to Ivan Štekl, Director of the IEAP, for his continuous support and efforts to secure funding and grants during these years.

Special thanks go to Karel Smolek, whose support has been immensely valuable in helping students like me find ways to achieve our dreams. His proactive support has always been a guiding light, always one step ahead at every turn. Above all, I'm deeply thankful to Babar Ali for his vital support during my doctoral pursuit. He's been more than a mentor—a trusted family-like figure who's always been there for me, both professionally and personally. I've learned so much from him, especially the importance of patience.

Thanks to all the people at IEAP who have been part of this Ph.D. experience. A heartfelt thank you to Andrés Melo and Catalina Lesmes Ramirez; since the beginning of our journey, we have shared so many memories to cherish as we progress towards our Ph.D. Next, I would like to express my sincere thanks to Petr Baroň and Patrycja Anna Potepa. This Ph.D. journey would have been very different without both of you. I have learnt so much from our time together and we have shared all the ups and downs of this journey. Thank you for always being there.

I would like to thank the members of the Higgs, Top and Heavy Ion groups for all the years of learning and understanding the working culture at CERN. For my involvement in flavour tagging at ATLAS, I would like to thank Chris Pollard and Marko Stamenkovic, together we tagged many b quarks. Thank you for all your support, it was always a joy working with you.

I am deeply grateful to Amartya Rej. Our journey from the IIT Bombay hostel to our first Prague visit was incredibly fruitful. Your concise yet profound suggestions have greatly influenced my Ph.D. journey, and the time you devoted to sharing crucial information has rightfully earned you the title of my virtual mentor.

I would like to thank Sarbojeet Bhowmick and Sudipta Hensh for all the wonderful moments I experienced in Prague and Opava. They both made my stay truly unforgettable. Sudipta, from our first meeting outside the lecture hall complex at IIT Bombay, India, to our time in Opava, Czech Republic, our journey together has been remarkable. Sarbojeet, your firm support and dedication have been unparalleled. From the tireless travelling back and forth between Prague and Opava to sharing the tasks of shopping and cooking, to those invaluable late-night discussions that helped us unwind and recharge with beer cheers and motivation for the next day's work, I couldn't have asked for better companions. I wish you both the very best in your future endeavours. I would also like to thank my college mate, Sudeb Ranjan Datta, who joined me for a short time in Prague, but made me rejoice in the memories of "B B College, Asansol" in Prague, far away from our homeland. Your presence brought a touch of home and nostalgia that I deeply appreciated.

A very special thanks to Aditi for her invaluable assistance. From helping me craft my first email to guiding me through replies and reviewing my SOP, your support has been truly invaluable. I deeply appreciate all that you have done.

Thank you to all my friends and well-wishers from my Bejdih village to Europe and around the world who have shared moments with me and stood by my side. Your support and encouragement means the world to me.

When it comes to friends and family, I'm deeply grateful to Priyanka. From the first day of my Ph.D. application to sharing the acceptance letter, to the final state examination, she has stood by my side and supported me like an iron lady. God bless her.

I want to thank my parents, brothers, sister, and my niece Shrinika from the bottom of my heart. Your support has meant everything to me on this journey. I'm so grateful for your constant love, help, and encouragement. I owe everything to my family for giving me the chances and lessons that made me who I am today. This journey wouldn't have been possible without you all by my side.

Abstract

This dissertation reports two analyses of Standard Model processes using proton-proton collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV and proton-lead collisions at a centre-of-mass energy of $\sqrt{s_{NN}} = 8.16$ TeV, recorded by the ATLAS detector at the Large Hadron Collider at CERN between 2015 and 2017 for pp collisions and in 2016 for p -Pb collisions.

The first analysis describes the search for $t\bar{t}H$ production in multilepton final states using proton-proton collisions at $\sqrt{s} = 13$ TeV with an integrated luminosity of 80 fb^{-1} . Six final states, categorised by charged-lepton number and flavour, and 25 event categories are defined to simultaneously detect the $t\bar{t}H$ signal and limit significant backgrounds. The $t\bar{t}W$ background normalisation is not constrained in the statistical analysis, resulting in a higher-than-expected normalisation. An excess of events consistent with $t\bar{t}H$ production is observed, corresponding to 1.8 standard deviations, compared to an expected 3.1 standard deviations. Assuming Standard Model branching fractions, the best-fit $t\bar{t}H$ production cross section is $\sigma_{t\bar{t}H} = 294^{+182}_{-162} \text{ fb}$, consistent with Standard Model predictions.

The second reports the observation of top-quark pair production in proton-lead collisions in the ATLAS experiment at the Large Hadron Collider. The measurement is performed using 165 nb^{-1} of p +Pb data collected at $\sqrt{s_{NN}} = 8.16$ TeV in 2016. Events are categorised in two analysis channels: events with exactly one lepton (electron or muon) and at least four jets, or events with two opposite-charge leptons and at least two jets. In both channels, at least one b -tagged jet is also required. Top-quark pair production is observed with a significance over five standard deviations in each channel. The top-quark pair production cross-section is measured to be $\sigma_{t\bar{t}} = 58.1 \pm 2.0$ (stat.) $^{+4.8}_{-4.4}$ (syst.) nb, with a total uncertainty of 9%, which makes this measurement the most precise $t\bar{t}$ cross-section determination in nuclear collisions to date. The measured cross-section is found to be in good agreement with a previous measurement by the CMS collaboration and with SM predictions. In addition, the nuclear modification factor is measured to be $R_{pA} = 1.090 \pm 0.039$ (stat.) $^{+0.094}_{-0.087}$ (syst.).

Contents

Introduction	1
Author's contribution	3
Chapter 1. Theoretical motivation	7
1.1. The Standard Model	7
1.2. The Elementary particles	8
1.3. Interactions	10
1.3.1. Quantum Electrodynamics - QED	11
1.3.2. Quantum Chromodynamics - QCD	12
1.3.3. Electroweak interaction	13
1.4. Top quark production and cross-section measurements	13
Chapter 2. The LHC and the ATLAS detector	17
2.1. The Large Hadron Collider	17
2.2. Luminosity and pileup	19
2.3. ATLAS detector	20
2.3.1. Coordinate system	21
2.3.2. The magnet system	22
2.3.3. The inner detector	23
2.3.4. The calorimeter system	24
2.3.5. The muon spectrometer	25
2.3.6. Trigger system	26
Part I. ATLAS Qualification task	
Chapter 3. Eigenvector decomposition study related to b-tagging pseudo-continuous data-based calibration	31
3.1. Introduction	31
3.2. Flavour tagging algorithms	32
3.2.1. Low-level b -tagging algorithms	32
3.2.2. High-level taggers	33
3.3. Tracking in ATLAS	34
3.4. Flavour tagging calibrations	35
3.4.1. $t\bar{t}$ calibration	35
3.4.2. Cumulative and pseudo-continuous calibration	35
3.4.3. b -tagging scale factors	36
3.5. Pseudo-continuous calibrations	37
3.5.1. Pseudo-continuous	37

3.5.2.	Selections and technical implementation	37
3.5.3.	Calibration data interface (CDI) structure	38
3.6.	Post-processing	39
3.6.1.	Eigenvector decomposition	40
3.7.	Scale factor re-scaling	41
3.7.1.	Scale factors for DL1 for EMPFlow jets	42
3.7.2.	Scale factors for DL1 for EMPFlow jets	43
3.7.3.	Scale factors for DL1r EMPFlow jets	45
3.8.	Scale factors for systematics	46
3.8.1.	Total systematics variation	47
3.8.2.	Individual systematic variation	51
3.9.	Summary	55

Part II. Proton-Proton analysis

Chapter 4.	Analysis of $t\bar{t}H$ and $t\bar{t}W$ production in multilepton final states with the ATLAS detector at 13 TeV	59
4.1.	Introduction	59
4.2.	Data samples	61
4.3.	Monte Carlo samples	61
4.3.1.	$t\bar{t}H$ signal	61
4.3.2.	$t\bar{t}W$ background	62
4.3.3.	Other backgrounds	64
4.4.	Object selection	65
4.4.1.	light leptons	65
4.4.2.	Jets and b -tagged jets	66
4.4.3.	Hadronically decaying tau leptons	66
4.4.4.	Missing transverse momentum	67
4.4.5.	Overlap removal	67
4.5.	Tight light leptons definition	67
4.5.1.	Isolation variables	68
4.5.2.	Non-prompt leptons	68
4.5.3.	Electron charge misidentification BDT	68
4.5.4.	Classification	68
4.6.	Event selection	69
4.7.	Categorisation	73
4.8.	Background overview and estimation	73
4.8.1.	Irreducible backgrounds	74
4.8.2.	Reducible backgrounds	77
4.9.	$2\ell SS1\tau_{\text{had}}$ channel	78
4.9.1.	Fake lepton estimation	79
4.10.	Systematic uncertainties	93
4.10.1.	Experimental uncertainties	93
4.10.2.	Background uncertainties	95
4.10.3.	Signal and background modelling uncertainties	97
4.11.	Results	99

4.12. Summary	108
-------------------------	-----

Part III. Proton-Lead analysis

Chapter 5. Observation of $t\bar{t}$ production in the lepton+jets and dilepton channels

in $p+\text{Pb}$ collisions at $\sqrt{s_{\text{NN}}} = 8.16$ TeV with the ATLAS detector	111
5.1. Introduction	111
5.2. Data samples	114
5.3. Monte Carlo samples	114
5.4. Object selections	116
5.4.1. Leptons	116
5.4.2. Jets	117
5.4.3. Missing transverse energy	117
5.4.4. Overlap removal	118
5.5. Backgrounds overview	118
5.5.1. W+jets background	119
5.5.2. Non-prompt/fake lepton backgrounds	119
5.6. Event selection	120
5.6.1. lepton+jets channel	120
5.6.2. Dilepton channel	121
5.6.3. Separation variable study	121
5.6.4. Yields	122
5.7. Systematic uncertainties	125
5.7.1. Luminosity	125
5.7.2. Modelling systematics	125
5.7.3. Background systematics	125
5.7.4. Lepton systematics	127
5.7.5. Jets systematics	128
5.7.6. Treatment of the uncertainties	129
5.7.7. Comparisons of distributions for important uncertainties	130
5.8. Statistical analysis	133
5.8.1. The likelihood function	133
5.8.2. Test statistic	135
5.9. Results	136
5.9.1. Results of the ℓ +jets and dilepton channels combined fit	137
5.9.2. Cross-section extraction	151
5.9.3. Nuclear Modification Factor (R_{pA})	153
5.10. Summary	154
Conclusion and outlook	155
List of papers	157
Conference contributions	159
Media appearance	161
Auxiliary plots: $p+\text{Pb}$	163
Bibliography	173

Introduction

*The LHC accelerates the protons and the lead
And the things that it discovers will rock you in the head.*

— Katherine McAlpine - “Large Hadron Rap”

The Higgs boson is a fundamental particle predicted by the Standard Model [1, 2] of particle physics. It is responsible for giving mass to fundamental particles and is the quantum of the Higgs field, which pervades all of space. The Higgs boson was discovered in 2012 by the ATLAS and CMS experiments at the Large Hadron Collider (LHC) at CERN. The Higgs boson is a scalar particle with a mass of about 125 GeV, produced by the interaction of fundamental particles. It decays into various particles, including photons, W and Z bosons, and fermions. The discovery of the Higgs boson has significant implications for our understanding of the universe and the fundamental forces that govern it.

This thesis reports the results of a search for $t\bar{t}H$ production at $\sqrt{s} = 13$ TeV using a dataset corresponding to an integrated luminosity of 80 fb^{-1} collected with the ATLAS detector during 2015-2017. The measurement uses six final states distinguished by the number and flavour of charged-lepton candidates (electron, muon, and hadronically decaying τ). In the following, the term light lepton, denoted by ℓ , refers to either electrons or muons. Multilepton signatures are primarily sensitive to the decays $H \rightarrow WW^*$ (with at least one W decaying to leptons), $H \rightarrow \tau^+\tau^-$, and $H \rightarrow ZZ^*$ (with a subsequent decay to $ll\nu\nu$ or $llqq$). The selection avoids overlap with the ATLAS searches for $t\bar{t}H$ production with $H \rightarrow b\bar{b}$, $H \rightarrow \gamma\gamma$, and $H \rightarrow ZZ^* \rightarrow 4\ell$ decays.

The main backgrounds to the $t\bar{t}H$ signal arise from $t\bar{t}W$, $t\bar{t}(Z/\gamma^*)$, and diboson (VV) production, as well as from $t\bar{t}$ production with additional light leptons from heavy-flavour(HF) hadron decays, misidentified jets, or photon conversions (collectively referred to as “non-prompt leptons”), and other processes where the electron charge is incorrectly assigned or where jets are misidentified as τ_{had} candidates. A binned maximum-likelihood fit is performed in all final states to extract the $t\bar{t}H$ cross-section and the background normalisations.

While the ATLAS experiment at CERN focuses primarily on proton-proton collisions, it also performs proton-lead (p -Pb) collisions. In 2016, the LHC underwent a dedicated p -Pb run, providing the ATLAS experiment with valuable data for this particular type of interaction. This data set makes it possible for the first time to search for top quark pairs decaying into pairs of electrons or muons (dilepton channel), as well as a single electron/muon accompanied by jets of particles (lepton+jet channel).

Top quarks and Higgs bosons were not observed in heavy-ion collisions with the ATLAS detector until September 2023, when top quarks were finally detected in such collisions, as documented in this thesis. Studying the production of top quarks in collisions between protons and lead nuclei offers a unique opportunity to study the properties of protons at high densities, which are difficult to probe by other experimental methods. This analysis aims to measure the top quark production rate, the quark-gluon plasma (QGP) dynamics and the nuclear modifications of the parton distribution functions, and to compare them with detailed theoretical predictions at a centre-of-mass energy of $\sqrt{s_{NN}} = 8.16$ TeV with the ATLAS experiment.

Furthermore, the ATLAS qualification task aimed at exploring the pseudo-continuous framework is described. The analysis focused on the format of pseudo-continuous inputs, their differences from cumulative inputs, the number of systematics, and their treatment. The procedure for using the tag weight distribution of b -tagging algorithms is detailed. In addition, a tool has been developed to plot the contents of pseudo-continuous calibrations directly from the CDI file. This tool serves as a basis for investigating pseudo-continuous scale factors for EV decomposition and smoothing of pseudo-continuous calibrations.

The aims of the presented thesis are:

- Develop the b -tagging calibration tool to fulfil the ATLAS authorship qualification task (QT).
- Perform an analysis of $t\bar{t}H$ and $t\bar{t}W$ production in multilepton final states with the ATLAS detector at 13 TeV.
- Measure the top pair cross-section in proton-lead collisions with the ATLAS detector at 8.16 TeV.

Outline of the thesis:

- Chapter 1: gives an overview of the Standard Model and the theory of top quark physics
- Chapter 2: introduces the LHC and the ATLAS detector
- Chapter 3: the b -tag calibration algorithm is discussed
- Chapter 4: describes the analysis of $t\bar{t}H$ and $t\bar{t}W$ production in multilepton final states with the ATLAS detector
- Chapter 5: describes the observation of $t\bar{t}$ production in lepton+jets and dilepton channels in p +Pb collisions with the ATLAS detector

Author's contribution

The author carried out the work described in this thesis as a member of the ATLAS collaboration. This document provides a summary of his significant contributions.

As part of the qualification task, the author contributed to the ATLAS Flavour-Tagging working group by performing specific tasks. The author maintained and developed the pseudo-continuous data-based calibration in Release 21, improved the software infrastructure for delivering these calibrations to physics analyses, and ensured seamless integration with the ATLAS workflow.

The primary objective of this work was to establish tools to directly plot the contents of pseudo-continuous calibrations from the Calibration Data Interface (CDI) file, corresponding to an integrated luminosity of 139 fb^{-1} at $\sqrt{s_{\text{NN}}} = 13 \text{ TeV}$. The code was extended to handle more complex uncertainty correlation schemes that are expected to be used in legacy Run 2 combinations. To ensure the long-term sustainability of this work, the code was streamlined and merged with the main CDI building infrastructure in use. In addition, these developments were thoroughly documented with the code on GitLab to ensure transparency and reproducibility. The performance of the ATLAS flavour-tagging algorithms in Run-2 was presented in Paper [4], as described in Chapter 3.

Chapters 4 and 5 represent the author's primary research. All information in these chapters is the original work of the author unless explicitly stated otherwise. For the fake-estimation of the $2\ell\text{SS}1\tau_{\text{had}}$ channel in the “ $t\bar{t}H$ and $t\bar{t}W$ production in multilepton final states” analysis, the author was the principal investigator. In addition, the author devised the analysis strategy and was the main contact for the fitting procedures across all channels in the “Analysis of top-quark pair production in proton-lead collisions”.

First, the author performed an analysis of the $2\ell\text{SS}1\tau_{\text{had}}$ channel using the proton-proton data set collected by the ATLAS experiment from 2015 to 2017, which had an integrated luminosity of 80 fb^{-1} . This analysis, outlined in chapter 4, was subsequently documented in Paper [1]. The author presented the results on behalf of the ATLAS Collaboration at the 40th International Conference on High Energy Physics (ICHEP 2020), with the detailed publication available in Paper [2]. In the context of the $2\ell\text{SS}1\tau_{\text{had}}$ channel analysis, the author undertook the following key tasks:

- Optimisation of event selection (Section 4.6)
- Estimation of the irreducible background (Section 4.8.1)
- Estimation of backgrounds with a fake τ_{had} candidate, as detailed in Paper [2] (Section 4.8.2)

- Estimation of the fake lepton background using the fake-factor (ABCD) method. This was a significant part of the analysis work, with detailed studies described in Section 4.9

The author was also responsible for the preparation of special input samples for the BDT optimisation. This included data and Monte Carlo simulations and optimisation.

The second analysis focuses on measuring the cross-section of top quark pair production in proton-lead collisions at $\sqrt{s_{NN}}=8.16$ TeV. In the context of heavy ion (HI) collisions, top quarks provide a unique probe of the nuclear modifications of the parton distribution functions (nPDFs), especially in a kinematic region that is not well constrained by other measurements. By comparing various observables from HI collisions with reference measurements from proton-proton (pp) collisions, these nuclear properties can be inferred. In this analysis, the author has made significant contributions towards meeting the thesis requirements by undertaking the following key tasks:

- Background overview (Section 5.5)
- Investigation of the separation power of different H_T variables to determine the most effective fit variable to separate the signal from the background (Section 5.6.3)
- Formulation of systematic uncertainties: normalisation, signal modelling and background-related systematics. These uncertainties are critical to ensure the accuracy and reliability of the final results. For a full understanding of these uncertainties, refer to Section 5.7
- Treatment of uncertainties Optimising pruning and symmetry processes are incorporated to stabilise fit results (Section 5.7.6)
- The statistical methods used in the analysis, including the fitting procedures, are described in detail in Section 5.8
- Profile likelihood fit of the analysis, comprising six channels (four ℓ +jets and two dilepton regions), is detailed in Section 5.9
- The results of the analysis (combined ℓ +jets and dilepton final states) are detailed in Section 5.9.1

In addition to the work mentioned above, the author has made significant contributions to pseudo-top reconstruction, ntuple validation, signal regions optimisation, and statistical and systematic fluctuations management to achieve the best possible significance. These efforts have been essential in extracting the signal strength, a key element in determining both the cross-section and R_{pA} . The results will be presented by the author on behalf of the ATLAS collaboration at the 42nd International Conference on High Energy Physics (ICHEP) on 18-24 July 2024.

RUPRECHT-KARLS-
UNIVERSITÄT
HEIDELBERG



Universität Heidelberg · Kirchhoff-Institut für Physik · Im Neuenheimer Feld 227 · 69120 Heidelberg

Kirchhoff-Institut für Physik
Im Neuenheimer Feld 227
69120 Heidelberg
Germany

apl. Prof. Dr. M. Dunford

Email:
monica.dunford@kip.uni-
heidelberg.de
Tel.: +49(0)6221/54 8038

To whom it may concern,

In my role as the physics coordinator of the ATLAS collaboration, I am writing to confirm that Santu Mondal from the Silesian University in Opava and the Czech Technical University in Prague has made a major contribution to ATLAS analysis of observing the top-antitop quark pairs in proton-lead collisions, a publication submitted by ATLAS to JHEP [1], for which he has been the key contributor and developer of the likelihood fit to the signal strength, took part in the internal note as well as the paper manuscript preparation, systematics evaluations and various checks. He has also made contributions to the analysis of ttH multi-lepton final states [2] in terms of the fake lepton background estimation, as well in performance for the b-tagging group [3] in terms of the pseudo-continuous data-based calibration [3].

Date
25. June 2024

Sincerely,

apl. Prof. Dr. Monica Dunford

[1] ttbar: https://atlas-glance.cern.ch/atlas/analysis/analyses/details.php?ref_code=ANA-TOPQ-2023-32, <https://arxiv.org/abs/2405.05078>, <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/TOPQ-2023-32/>

[2] ttH: https://atlas-glance.cern.ch/atlas/analysis/analyses/details.php?ref_code=ANA-HIGG-2018-05

[3] QT: <https://its.cern.ch/jira/browse/AFT-426>



Kirchhoff-Institut
für Physik

Chapter 1

Theoretical motivation

In 1897, Joseph J. Thomson unveiled the discovery of the electron, marking the inception of particle physics. Subsequently, a surge in discoveries and theoretical developments emerged over the ensuing four decades, underscoring the necessity for a comprehensive framework capable of elucidating and organizing the diverse range of observed phenomena. It was not until the 1960s and 1970s that a collaborative effort between theorists and experimentalists laid the groundwork for the SM of particle physics. This comprehensive theoretical structure encompassed all known elementary particles and interactions. Predictions made within the SM, such as the existence of the Higgs Boson, were later confirmed by experimental discoveries, including its detection in 2012. Within the SM, matter's fundamental constituents are organized into three generations of leptons and quarks. The identification of the top quark, the heaviest quark within the SM, in 1995 by the CDF and $D\bar{O}$ experiments [3, 4] marked a significant milestone. Unlike other quarks, the top quark does not form bound states, making it a focal point of research in high-energy physics. Its unique characteristics have spurred extensive investigations into the fundamentals of weak and strong interactions, contributing to a deeper understanding of particle physics.

1.1. The Standard Model

Particle physics endeavours to unravel the fundamental constituents of the cosmos, termed elementary particles, alongside elucidating their interactions, referred to as forces. Central to this pursuit is the Standard Model (SM) of particle physics, offering a unified framework wherein the forces governing particles are elucidated by intermediary particles. This model stands as the pinnacle of success in explaining experimental data amassed thus far. A primary objective of experimental particle physicists is to rigorously assess this theory with heightened precision through meticulous data acquisition and analysis. This endeavour entails scouring the data for unanticipated phenomena that could potentially illuminate new facets of our current understanding.

Within the realm of the SM, the excitations of fundamental fields give rise to two distinct classes of elementary particles: fermions (distinguished by their half-integer spin) and bosons (distinguished by their integer spin). Fermions, serving as the elemental constituents of matter, are further classified into leptons and quarks based on their interaction patterns. Presented in Figure 1.1, a schematic overview encapsulates the properties of the fundamental particles within

Three Generations of Matter (Fermions)				
	I	II	III	
mass→	2.4 MeV	1.27 GeV	171.2 GeV	0
charge→	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	0
spin→	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1
name→	u up	c charm	t top	γ photon
	4.8 MeV	104 MeV	4.2 GeV	0
	$-\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$	0
	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1
Quarks	d down	s strange	b bottom	g gluon
	<2.2 eV	<0.17 MeV	<15.5 MeV	91.2 GeV
	0	0	0	0
	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1
	ν_e electron neutrino	ν_μ muon neutrino	ν_τ tau neutrino	Z⁰ weak force
	0.511 MeV	105.7 MeV	1.777 GeV	80.4 GeV
	-1	-1	-1	± 1
	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1
Leptons	e electron	μ muon	τ tau	W[±] weak force

Figure 1.1. SM of elementary particles: the 12 fundamental fermions and 5 fundamental bosons.

the SM framework. Twelve elementary fermions compose the roster, with their interactions mediated by bosons. The electromagnetic, strong, and weak interactions constitute the three foundational forces in the SM, each facilitated respectively by photons, gluons, and the W and Z bosons. While general relativity provides insight into gravitational phenomena at macroscopic scales, a quantum theory of gravity remains indispensable for comprehensively integrating gravity into the SM framework. Lastly, the Higgs boson emerges as a pivotal element, interacting with elementary particles in proportion to their mass.

1.2. The Elementary particles

The groundwork for the SM was established in 1964 independently by Murray Gell-Mann and George Zweig [5, 6], who proposed the concept of nucleons comprising charged particles of spin $\frac{1}{2}$, termed quarks. Initially, only three quarks were postulated, representing the constituents of

particles known at that time. Subsequently, numerous other elementary particles predicted by the SM emerged in accelerator experiments in the ensuing years.

The discovery of the charm quark occurred in 1974 independently at both the Stanford Linear Accelerator Center [7] and the Brookhaven National Laboratory [8], through the observation of the J/ψ resonance, a $c\bar{c}$ bound state. Following this, in 1977 at the Tevatron (Fermilab), the observation of another heavy meson [9–11] state denoted as Υ , a $b\bar{b}$ bound state, led to the identification of the bottom quark. The final quark, the top, was ultimately observed in 1995 at the Tevatron.

The electron, the first discovered charged lepton, was credited to J.J. Thomson, while the muon and τ were respectively detected in cosmic rays in 1937 [12] and in accelerator experiments in 1975 [13]. Wolfgang Pauli suggested the neutrino's existence in 1930 to explain apparent energy and angular momentum nonconservation in beta decay. Clyde Cowan and Fred Reines discovered it (as the electron antineutrino) in 1956.

Both leptons and quarks are fermions, with spin $\frac{1}{2}$. They are described by the Dirac equation [14]. This equation yields solutions for both positive and negative energies, with the negative solutions interpreted as anti-particles, possessing identical mass and spin to their corresponding fermions but opposite charges. The classification of quarks and leptons into three families is outlined below:

Quarks			Leptons		
$\begin{pmatrix} u \\ d \end{pmatrix}$	$\begin{pmatrix} c \\ s \end{pmatrix}$	$\begin{pmatrix} t \\ b \end{pmatrix}$	$\begin{pmatrix} \nu_e \\ e \end{pmatrix}$	$\begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}$	$\begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}$

Table 1.1. Quarks and Leptons.

The quark species are characterized by additive internal quantum numbers called flavours. These include the third isospin component I_3 , strangeness S , charm C , bottomness (beauty) B , and topness (truth) T . Their values are shown in the Table 1.1. In addition, all quarks carry a baryon number of $\frac{1}{3}$. The flavour quantum numbers of antiquarks are the opposite. The flavour quantum numbers of quarks and antiquarks add to the quantum numbers of the hadrons they form. For both quarks and hadrons, the generalized Gell-Mann-Nishijima (Equation 1.1) formula holds.

$$Q = I_3 + \frac{1}{2}(B + S + C + B + T), \quad (1.1)$$

The analysis of the quark composition of baryons, which form multiplets with the same spin and nearly equal mass, showed that to obey the Pauli exclusion principle, the quarks must be assigned another quantum number with three possible values. This quantum number was called colour, with red, blue and green values. By analogy with the principles of colourimetry, baryons appear as the ‘white’ combination of quark colours. To make mesons ‘colourless’, antiquarks

are assigned anti-colours. Tripling the number of quarks by the colour degree of freedom helps to explain other phenomena, such as the decay of the neutral pion into two photons and the hadron/dimuon ratio in e^+e^- annihilation. The mathematically rigorous meaning of the colour concept is given in terms of the $SU(3)_c$ group to which the hadrons belong in the one-dimensional representation.

Quarks, as elementary fermions, combine to form singlet states in two fundamental configurations: either as mesons, comprising a quark-antiquark pair, or as baryons, consisting of three quarks. Mesons exhibit integral spin, while baryons demonstrate half-integral spin, consistent with Fermi-Dirac statistics. This statistical behavior governs the properties of atomic nuclei, which consist predominantly of protons (the sole stable hadrons) and neutrons. The neutrons outside the nucleus are unstable, with a relatively extended lifetime of approximately fifteen minutes. Composite states involving heavier quarks such as s , b , and c decay rapidly within a timescale of around 10^{-13} seconds, producing lighter particles and thus contributing minimally to ordinary matter.

The leptons encompass the electron (e), muon (μ), and tau (τ), with each family supplemented by three corresponding neutrinos. According to the convention, the electric charge of a particle is written as Qe , where e represents the charge of a proton. Additionally, each lepton carries an internal charge known as the leptonic number (l_e, l_μ, l_τ), unique to each family and defined based on the properties of weak interactions. Conservation of the leptonic quantum number is observed in numerous processes, except neutrino oscillation.

Interaction	Relative strength	Range (m)	Mediator
Strong	1	10^{-13}	gluons
Electromagnetic	10^{-2}	∞	photons
Weak	10^{-5}	$< 10^{-15}$	W^\pm, Z
Gravity	10^{-39}	∞	graviton

Table 1.2. Relative strength and range of various interactions [15]. The comparison of the strengths of the four fundamental forces is gauged by the force exerted between two particles separated by a specific distance and holds significance solely in relative terms. The graviton, denoted as the mediator of gravity, is a spin-2 boson introduced within quantum field theory but has not yet been incorporated into the SM.

1.3. Interactions

The electron, electron neutrino, up-quark, and down-quark are grouped as first-generation particles in the SM, illustrated in the leftmost columns of Figure 1.1. These particles are fundamental and form the basic building blocks of all observable matter in the universe. However, particle interactions become more complex when examined at high energies in particle colliders. Each elementary particle of the first generation has two versions with different masses. For instance,

the charm quark of the second generation is a heavier version of the up quark, with a mass approximately 400 times that of the up quark ($m_c \approx 400m_u$), while the top quark of the third generation is an even heavier variant, with a mass around 58000 times that of the up quark ($m_t \approx 58000m_u$). Despite these mass differences, the up, charm, and top quarks interact in the same manner.

Symmetries are integral to the Lagrangian density, extensively employed in describing particles within the SM. In four-dimensional space-time, elementary particles manifest as excitations on associated fields denoted as $\psi(x)$, with the Lagrangian density detailing their motion.

$$\mathcal{L}(x) = \mathcal{L}(\psi, \partial_\mu \psi), \quad (1.2)$$

In the SM, a key concept is how symmetries link up with conservation laws, as explained by the Noether theorem. For example, when something stays the same under certain changes like moving in time or space, it leads to conserving important stuff like energy and momentum. The Dirac Lagrangian density helps us describe fast-moving fermions while making sure energy and momentum are still conserved.

$$\mathcal{L} = \bar{\psi} (i\gamma^\mu \partial_\mu - m) \psi, \quad (1.3)$$

where γ^μ are the Dirac gamma matrices. The equations of motion are obtained by solving the Euler-Lagrange equations, resulting in the Dirac equation:

$$(i\gamma^\mu \partial_\mu - m) \psi = 0, \quad (1.4)$$

One notable thing is that there is an antiparticle with the same mass but a different charge for each of the twelve fermions. When there's no interaction happening, each tiny fermion's movement is well explained using this Lagrangian method. Global and local symmetries are two kinds of internal patterns that help us understand how interactions work. Local symmetries let changes happen differently from one spot to another in space and time. On the other hand, global symmetries apply changes consistently everywhere in space and time. With a global symmetry, if the Lagrangian stays the same, we get a constant flow, like keeping the electric charge the same in particle interactions. Gauge bosons, particles that manage different interactions by swapping a constant amount, come about because the Lagrangian stays the same under local symmetry, also called gauge symmetry.

1.3.1. Quantum Electrodynamics - QED

“Quantum electrodynamics (QED)” constitutes a vital domain within the SM that thoroughly outlines electromagnetic phenomena, such as electron-electron scattering. Renowned for its unparalleled precision in experimental validation, QED illuminates the intricate interplay of charged fermionic particles through the emission of photons. This framework mandates the

preservation of Lagrangian density under phase translations of fermionic fields, a principle encapsulated by the $U(1)$ symmetry group. Operating within the realms of an abelian gauge theory, QED harnesses the symmetry group $U(1)$ to expound upon electromagnetic interactions. The interaction dynamics between a spin- $\frac{1}{2}$ particle denoted as ψ and an electromagnetic field A_μ are succinctly encapsulated within the ensuing Lagrangian expression:

$$\mathcal{L}_{\text{QED}} = \bar{\psi} (i\gamma^\mu D_\mu - m) \psi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu}, \quad (1.5)$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the EM field tensor. $D_\mu = \partial_\mu + ieA_\mu + ieB_\mu$ is the gauge covariant derivative with B_μ the external field imposed by external source and γ^μ are Dirac matrices.

1.3.2. Quantum Chromodynamics - QCD

Quantum chromodynamics (QCD) serves as the framework for understanding strong interactions, characterized by its invariance under the symmetries of the local gauge group $SU(3)_c$. The conserved quantum attribute in QCD is the colour charge, which manifests in three distinct varieties: red, blue, or green. Despite the plethora of experimental evidence supporting the existence of quarks, the direct observation of free quarks remains elusive. This enigma finds resolution through the concept of colour confinement, stipulating that objects possessing colour charge are perpetually confined within colourless composite entities, thereby precluding the propagation of objects with non-zero colour charge as free particles.

Mesons and baryons, being composite particles devoid of colour charge, serve as observable manifestations of quarks, often observed as jets of these particles. Deriving interactions among quarks necessitates the Lagrangian density to uphold invariance under phase translations of the $SU(3)_c$ symmetry group. The resultant Lagrangian is then expressed as follows:

$$\mathcal{L}_{\text{QCD}} = \sum_q \bar{\psi}_q (i\gamma^\mu D_\mu - m_q) \psi_q - \frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu}, \quad (1.6)$$

where q denotes the sum of the quarks. The invariance under $SU(3)_c$ is ensured by using the the covariant derivative:

$$D_\mu = \partial_\mu + ig_s \frac{\lambda^a}{2} G_\mu^a, \quad (1.7)$$

where g_s is the strong interaction coupling constant, λ^a are the Gell-Mann matrices, and G_μ^a denotes the eight fields that mediate the strong interaction, the gluons, each corresponding to one generator of the $SU(3)$ symmetry. The strong interaction tensor is used to describe the Lagrangian density of these fields:

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a - g_s f^{abc} G_\mu^b G_\nu^c, \quad (1.8)$$

where f^{abc} are the structure constants of the $SU(3)$ group. Gluons are massless particles for the same reason that photons are. Because gluons always carry a colour charge, they also confine to colourless objects and, unlike photons, do not propagate over macroscopic distances.

1.3.3. Electroweak interaction

QED and QCD exhibit several parallels, notably in their utilization of massless neutral spin-1 bosons as mediators and in the uniformity of their interaction vertices. In contrast, the weak interaction, responsible for decay and nuclear fusion processes, diverges significantly. Mediated by massive bosons, particularly the W^\pm boson, which carries an electric charge facilitating charge exchange in particle interactions, the weak force stands apart. Although distinct phenomena, electromagnetic and weak forces can be amalgamated and conceptualized as facets of a unified force. This notion was concretized in 1979 by S. L. Glashow, A. Salam, and S. Weinberg [16–18], who formulated the electroweak theory, also known as the GSW theory, providing a singular theoretical framework encompassing both electromagnetic and weak interactions.

1.4. Top quark production and cross-section measurements

The top quark is a fundamental spin- $\frac{1}{2}$ fermion in the SM. It has a significantly large mass, $m_t = 173 \pm 0.4 \text{ GeV}$ [19], much heavier than any other quark or lepton (for example, the next heaviest quark, the bottom quark, is about 40 times lighter). This substantial mass suggests that the top quark may play a special role in Beyond the Standard Model (BSM) theories and interactions involving the top quark may be the first to reveal BSM physics [20, 21].

Due to its large mass, the top quark has a very short lifetime and does not hadronise, meaning that its properties are directly transferred to its decay products. Experimentally, this is advantageous because it allows the properties of the top quark to be studied without the complications of hadronisation, unlike other quarks [22].

The top quark has been observed in experiments at two accelerators: the Tevatron at Fermilab, USA, and the LHC at CERN, Switzerland. It was first observed in 1995 at the Tevatron during ‘Run I’ with proton-antiproton ($p\bar{p}$) collisions at a centre-of-mass energy of $\sqrt{s} = 1.8 \text{ TeV}$ by the CDF and DØ experiments [3, 4]. ‘Run I’ took place from 1992 to 1996, collecting about 100 pb^{-1} of integrated luminosity per experiment, resulting in a few dozen top-antitop ($t\bar{t}$) pair candidate events.

The second data period (Run II) at the Tevatron took place from 2001 to 2011 with a slightly higher energy of $\sqrt{s} = 1.96 \text{ TeV}$. Each experiment collected approximately 100 times more data, totalling 10 fb^{-1} . This extensive dataset enabled detailed measurements of top quark properties, although many measurements were still limited by statistics.

The LHC began operations in 2008, but following an incident shortly after, the first collisions at $\sqrt{s} = 7$ TeV occurred in 2010. The center-of-mass energy of the proton-proton collisions (and the luminosity) increased over time: from $\sqrt{s} = 7$ TeV in 2010 (5 fb^{-1}), to $\sqrt{s} = 8$ TeV in 2011-2012 (20 fb^{-1}), and then to $\sqrt{s} = 13$ TeV in 2015-2018 (150 fb^{-1}), with a shutdown occurring in 2013-2014. The period from 2010-2012 is known as “Run 1,” while 2015-2018 is referred to as “Run 2.” Currently, there is another planned shutdown from 2019 to 2021.

Thanks to the higher interaction energy and greater luminosity at the LHC compared to the Tevatron, significantly more top quarks have been produced, allowing for much more detailed measurements of their properties.

Since Kobayashi and Maskawa introduced the third quark family in 1973 [23], the search for the top quark has been crucial to explaining CP violation in weak interactions. Its considerable mass made the top quark difficult to detect, requiring more than two decades of intensive effort and significant advances in particle accelerator technology.

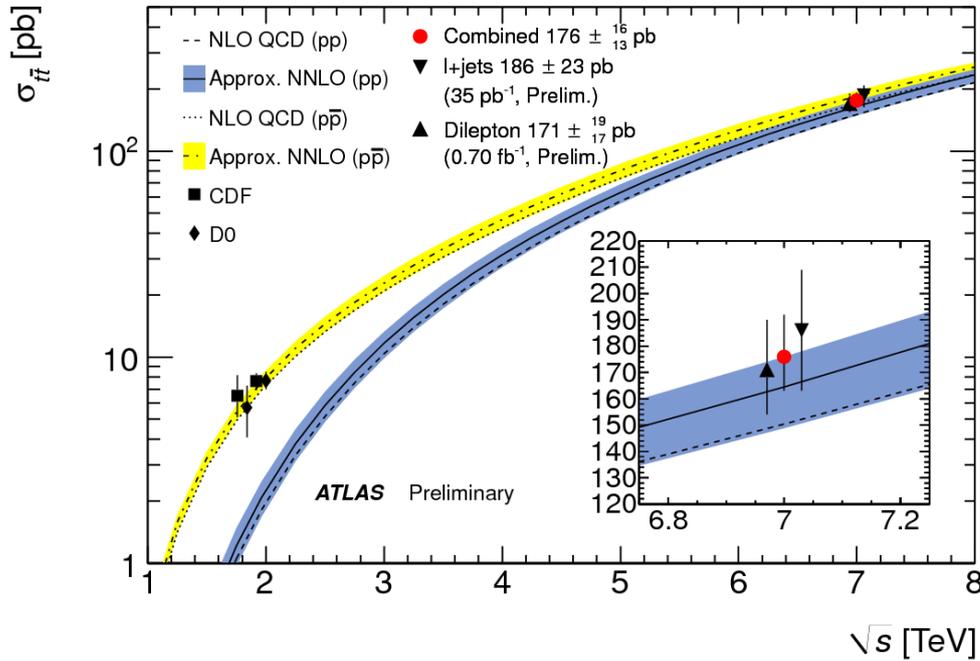


Figure 1.2. Dependence of σ_{tt} on \sqrt{s} from theoretical predictions based on a top mass of 172.5 GeV together with the dilepton, single lepton, and combined measurements from ATLAS in this note. Uncertainties on measurements are shown as vertical bars and include statistical, systematic, and luminosity contributions. Results obtained with the Tevatron are also shown. Measurements made at the same centre-of-mass energy are slightly offset for clarity [24].

With a mass measured at 173.0 ± 0.4 GeV, the top quark stands out as the heaviest quark within the SM. It uniquely exhibits a Yukawa coupling, denoted as y_t , to the Higgs boson, which is notably close to unity. This distinct characteristic underscores the pivotal role the top quark plays within the SM framework and various extensions thereof. Given its exceedingly

short lifetime, approximately on the order of 10^{-25} s, direct observation of the top quark within detectors remains unfeasible. Consequently, investigations into top quark properties necessitate the examination of its decay products. In approximately 99% of instances, the top quark decays into a W boson and a b quark. Hence, the characteristic signature of the $t\bar{t}$ process hinges on the decay products of the two W bosons in the event ($t\bar{t} \rightarrow bW^+\bar{b}W^-$). The final states resulting from $t\bar{t}$ decay are categorized into three distinct classes:

- **All-hadronic Channel:** In this channel, both the W boson decays into quark-antiquark pairs ($W \rightarrow q\bar{q}'$). The characteristic signature consists of six jets, primarily attributed to the hadronization of the quarks. The branching ratio of this final state is 45.7% [25]. However, discerning top quark pair production ($t\bar{t}$) from the substantial background of multi-jet events poses a significant challenge in this channel.
- **Di-leptonic Channel:** In this channel, both W bosons decay into leptons ($W \rightarrow l\nu$). The detector signature for these events comprises two high-energy leptons, accompanied by two jets originating from b -quark hadronization, and missing transverse energy due to the presence of undetectable neutrinos. This channel offers relatively straightforward identification as the presence of leptons aids in distinguishing the signal from background events.
- **Semi-leptonic Channel:** In this channel, one W boson decays into leptons while the other decays into hadrons. Detector signatures include an isolated high-energy lepton, missing transverse energy, and four jets. The branching ratio of this final state is 43.8% [25]. Referred to as the “golden channel,” it facilitates event triggering and signal-background separation due to the presence of a lepton. Moreover, approximately half of $t\bar{t}$ pairs decay in this channel, enhancing its statistical significance. By imposing the W mass constraint on the lepton-neutrino pair, the kinematics of the event can be fully constrained. Many analyses focus on final states containing two leptons due to the inherent challenges in identifying and reconstructing τ leptons compared to electrons and muons, which results in a reduction of branching ratio in the single lepton channel by approximately 15%.

Numerous precise measurements have been conducted to ascertain various properties of the top quark, encompassing factors such as the differential production cross-section, spin correlation, and forward-backwards or charge asymmetry. Further details on these measurements can be found in Ref. [19].

Top quarks are generated in a hadron collider via two distinct mechanisms: the formation of $t\bar{t}$ pairs stemming from strong interactions, or the production of single top quarks through weak interactions. The anticipated rate of top quark production is determined via the cross-section calculation [26]:

$$\sigma_{pp \rightarrow tX} \propto \int \frac{|M|^2}{v_i} dQ, \quad (1.9)$$

here, M represents the matrix element of the process, v_i denotes the velocity of the incoming particle, and dQ signifies the phase space element. In high-energy hadron collisions, the matrix element is computed solely considering the hard scattering among the constituents of the hadrons, excluding subsequent hadron formation. This separation is facilitated by specific conditions prevailing in the collisions. At energies characteristic of the LHC, the partons engage in collisions akin to free particles, permitting the application of the uncertainty principle to distinguish between the time intervals of parton scattering, $\tau_{sc} \propto \frac{1}{p_T}$, and hadronization, $\tau_{had} \propto p_T$.

Chapter 2

The LHC and the ATLAS detector

2.1. The Large Hadron Collider

The world's most powerful hadron-hadron collider, the Large Hadron Collider (LHC) [27], is located at the European Organization for Nuclear Research (CERN) site in Geneva, Switzerland. It consists of a 27-kilometer ring superconducting magnets and accelerating structures that accelerate two counter-rotating protons and heavy-ion (Pb) beams inside a ring of 26.7 km circular diameter and is located around 100 m underground. The LHC has four primary interaction points where the two beams are brought together to collide.

Before being injected into the LHC ring, the protons are accelerated to 450 GeV by several pre-accelerators. Protons are injected into the ring in bunches, each containing approximately 10^{11} protons. Two successive bunches are separated by a bunch gap of 25 ns or 50 ns. The LHC is designed to provide proton-proton (pp) collisions with a center-of-mass energy of $\sqrt{s} = 14$ TeV and an unprecedented luminosity of 10^{34} cm⁻² s⁻¹. In addition, the LHC can also collide heavy lead (Pb) ions with an energy of 2.8 TeV per nucleon and a peak luminosity of 10^{27} cm⁻² s⁻¹.

A schematic view of the LHC ring and the layout of the accelerator complex at CERN is shown in Figure 2.1. There are four main experiments installed at the LHC :

- A Toroidal LHC Apparatus (ATLAS) [28]
- Compact Muon Solenoid (CMS) [29]
- Large Hadron Collider beauty (LHCb) [30]
- A Large Ion Collider Experiment (ALICE) [31]

ATLAS and CMS are general-purpose detectors designed to cover the broadest possible range of physics at the LHC, including tests of the Standard Model and searches for new physics beyond the Standard Model. The LHCb focuses on the precision measurements of CP violation and rare decays of b-hadrons. ALICE is dedicated to heavy-ion physics. Additionally, three smaller, very specialized experiments are located near the ATLAS and CMS investigations:

- TOTal Elastic and diffractive cross-section Measurement (TOTEM) [32]
- Large Hadron Collider beauty (LHCb) [33]
- Monopole and Exotics Detector At the LHC (MoEDAL) [34]

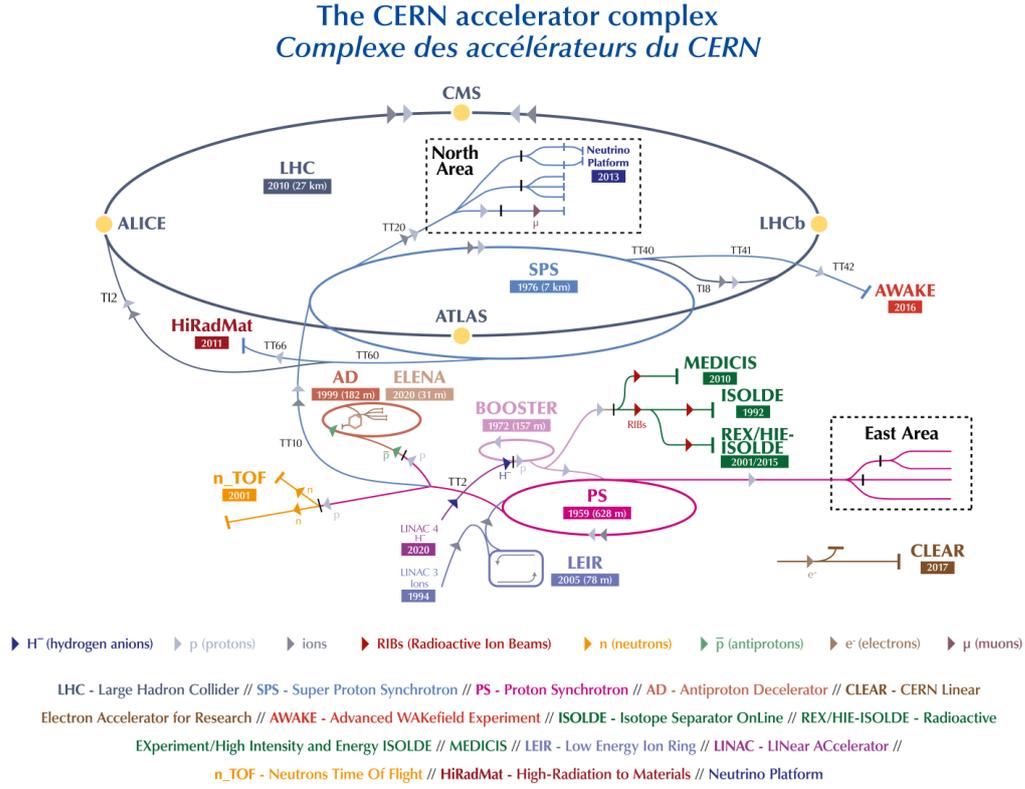


Figure 2.1. The CERN accelerator complex. The four main experiments are indicated as yellow dots.

The LHC delivers an extremely high instantaneous luminosity \mathcal{L} that is proportional to the rate of proton-proton interactions and therefore is directly correlated to the data volume recorded at the detectors. It is quoted in units of $\text{cm}^{-2} \text{s}^{-1}$ and evaluated as:

$$\mathcal{L} = f \frac{n_1 n_2}{4\pi\sigma_x\sigma_y}, \quad (2.1)$$

where f is the frequency with which the number of protons n_1 and n_2 in beams 1 and 2 collide. σ_x and σ_y denote the horizontal and vertical spread of the particle beams. The total integrated luminosity L over a data-taking period is therefore given as:

$$L = \int \mathcal{L} dt, \quad (2.2)$$

It is related to the total number of interactions N produced by a given physics process with cross-section σ as:

$$N = \sigma L, \quad (2.3)$$

Year	2015	2016	2017	2018
Total delivered integrated luminosity (fb^{-1})	4.0	38.5	50.2	63.4

Table 2.1. Year-wise integrated luminosity values.

The high instantaneous luminosity causes the number of proton-proton interactions in a single collision to be more than one, known as a pileup. However, due to the small bunch spacing, additional proton-proton interactions can also occur in a collision from neighbouring (previous or next) bunches. These pileup interactions consist mainly of low energy soft non-perturbative physics processes accompanying an energetic hard scattered interaction of interest.

Total integrated luminosity delivered by the LHC at $\sqrt{s} = 13$ TeV in Run-2. The dataset collected for the analyses presented in this thesis corresponds to the years 2015 and 2016 of the LHC operations.

2.2. Luminosity and pileup

The instantaneous luminosity generated by the proton beams dictates the collision rate of the underlying physics processes, while the total number of events is contingent upon the integrated luminosity accumulated over the operational timeframe. This primary luminosity assessment is conducted using the LUCID-2 detector situated alongside the LHC beam pipe, flanking the ATLAS detector.

Throughout Run 2 operations, the LHC delivered an integrated luminosity of 156 fb^{-1} , with ATLAS recording 147 fb^{-1} . Our analysis exclusively considers data from periods when the ATLAS detector was fully operational and the recorded data met quality standards, yielding an integrated luminosity of 138.97 fb^{-1} for the entirety of Run 2. The combined uncertainty in integrated luminosity for the 2015-2018 period is $\pm 1.7\%$. Details regarding the luminosity collected each year are presented in Table 2.2, while Figure 2.2(a) illustrates the cumulative luminosity throughout operations.

Year	Integrated Luminosity [fb^{-1}]
2015	3.2
2016	33.0
2017	44.3
2018	58.5
Total	$139 \pm 1.7\%$

Table 2.2. Summary of data luminosity in the years between 2015 and 2018.

Protons undergo acceleration in batches, with collisions occurring at intervals of 25 ns. Despite the high luminosity of data generated by the Large Hadron Collider (LHC), a significant

challenge arises due to the substantial number of proton-proton interactions per bunch crossing. The average number of interactions per crossing μ is depicted in Figure 2.2(b) for Run 2. Within these multiple interactions, those aligned with the physics objectives are denoted as in-time pileups. To discern the pertinent physics interaction, the interaction vertices are identified alongside their associated particles. Additionally, there may be overlapping detector signals stemming from interactions of adjacent bunches, known as out-of-time pileup.

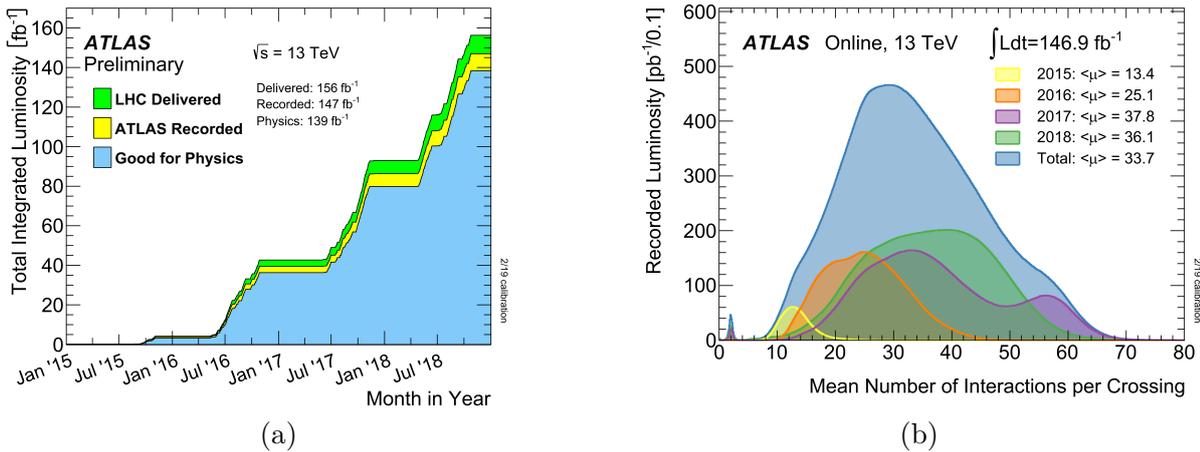


Figure 2.2. (a) Total Integrated Luminosity and Data Quality in 2015-2018: Cumulative luminosity versus time delivered to ATLAS (green), recorded by ATLAS (yellow), and certified to be good quality data (blue) during stable beams for pp collisions at 13 TeV centre-of-mass energy in 2015-2018. (b) Number of Interactions per Crossing: Shown is the luminosity-weighted distribution of the mean number of interactions per crossing for 2015-2018 pp collision data at 13 TeV centre-of-mass energy.

2.3. ATLAS detector

The ATLAS detector is located at one of the four interaction sites of the LHC collider. It's a multifunctional detector that can be used to study a wide range of physics processes, from high-precision measurements to the search for new phenomena. A schematic view of the detector is shown in Figure 2.3. It is 44 m long, 25 m in diameter and weighs over 7000 tonnes. To measure the momentum of the charged particles produced at the point of interaction, the detector relies on four magnets: a 2T central solenoid, an 8-coil barrel toroid arranged cylindrically around the detector, which produces a peak field of 4T, and two further 8-coil magnets at the detector end caps, which also produce a field of 4T. The weight of the detector is comparable to that of the Eiffel Tower.

The inner detector [35], the electromagnetic and hadron calorimeters, the magnet system, and the muon spectrometer are the four primary components of the detector. To record the trajectory, velocity, and energy of distinct particles, the subdetectors are placed in concentric layers around the beam pipe. Individual particles can therefore be identified.

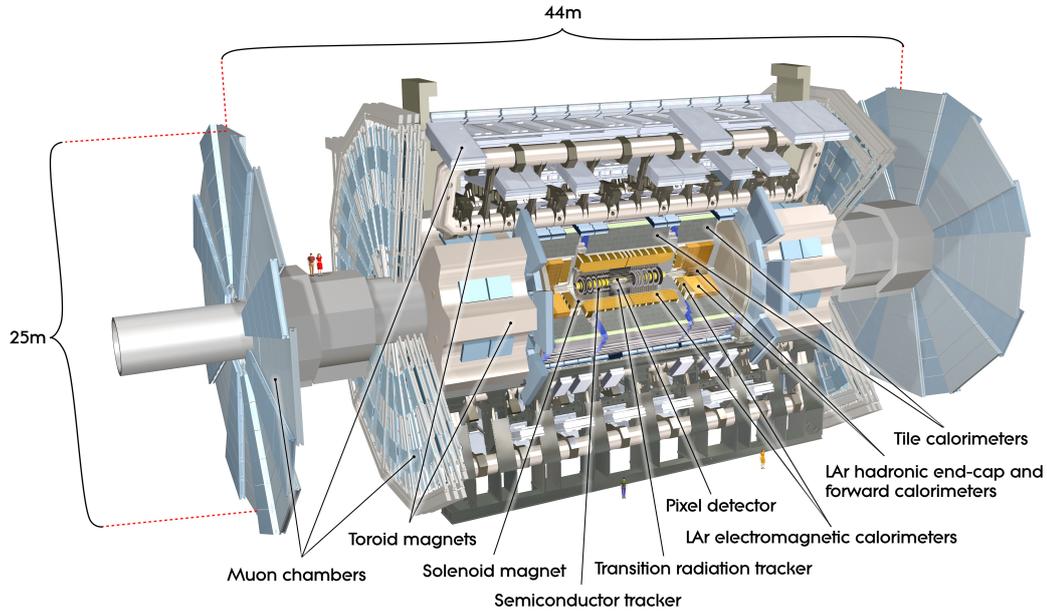


Figure 2.3. Schematic view of the ATLAS detector.

2.3.1. Coordinate system

The ATLAS experiment uses a right-handed coordinate system with its origin at the nominal interaction point at the centre of the detector. The positive x -axis points towards the centre of the LHC ring, the positive y -axis points vertically upwards, and the positive z -axis is along the beam direction. The cylindrical geometry of the detector makes it convenient to work with polar coordinates. The azimuthal angle ϕ is measured from the positive x -axis in the transverse xy -plane, whereas the polar angle θ is measured from the positive z -axis in the yz -plane.

In the high energy particle collisions, the rapidity y is preferred over θ as the difference in rapidity δy remains Lorentz-Invariant under longitudinal boosts along the beam direction. It is defined as:

$$y = \frac{1}{2} \ln \left[\frac{E + p_z}{E - p_z} \right], \quad (2.4)$$

where E is the particle's energy and p_z corresponds to the z -component of its three-momentum vector \vec{p} . However, it is not easy to measure the energy and momentum of highly relativistic particles. In such a scenario, another useful quantity pseudorapidity $\eta = -\ln \left(\text{tg} \frac{\theta}{2} \right)$ is considered that only depends on the polar angle θ . In the high relativistic limit with the approximation that the mass of a particle is negligible, the pseudorapidity becomes equivalent to the rapidity. The distance in the $\eta - \phi$ plane is defined as:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}, \quad (2.5)$$

In hadron collider physics, transverse quantities are often used to describe the kinematics of objects. Such quantities are the transverse momentum (p_T) and energy (E_T), which are the components of a particle's momentum and energy in the transverse $x - y$ plane, and can be calculated using the following relations:

$$p_T = \sqrt{p_x^2 + p_y^2}, \quad (2.6)$$

$$E_T = \sqrt{p_T^2 + m^2}, \quad (2.7)$$

where p_x and p_y are the momentum components in the x - and y -direction, respectively.

2.3.2. The magnet system

The magnet system is required to accurately measure the momentum of charged particles. It consists of a central solenoid, a barrel toroid and two endcap toroids, all designed to generate magnetic fields that bend the trajectories (tracks) of charged particles. The curvature of the trajectories of charged particles can be used to calculate their momentum. The stronger the magnetic field, the more dramatic the curvature and the more accurate the momentum measurement.

The solenoid is aligned with the beam axis and generates a 2 T axial magnetic field for the inner detector, which bends the particles in that direction. The solenoid is 5.8 m long and has a radius of 1.28 m metres. The solenoid and the calorimeter share the same cooling cryostat to reduce the amount of material in front of the calorimeter. The calorimeter and its supporting structure act as the return yoke for the solenoid.

In the barrel and endcap sections, the three massive superconducting air-core toroids provide a magnetic field of about 0.5 T and 1 T respectively for the muon spectrometer. Eight coils are arranged radially around the beam axis in each of the three toroids. The barrel coils have a length of 25.3 m metres with inner and outer diameters of 9.4 metres and 20.1 metres respectively. They are contained in a stainless steel vacuum vessel shaped like a racetrack. The end cap coils are five metres long with inner and outer diameters of 1.65 m and 10.7 m respectively. To provide radial overlap and greater magnetic field homogeneity, the end cap toroids are rotated by 22.5° with respect to the barrel toroid.

2.3.3. The inner detector

The Internal Detector (ID) [36] is placed closest to the point of interaction to capture charged particles produced by a collision and accurately measure their momenta up to $|\eta| < 2.5$. A 2T solenoidal magnetic field parallel to the rod hub operates throughout the ID, bending charged particles according to their electrical charge. The ID consists of three distinct and complementary subsystems: the pixel finder, the semiconductor tracker (SCT) and the transition radiation tracker (TRT). Each subsystem advance contains several layers of detection materials stacked on top of each other. The R-z cross-sectional view of the ID subsystems is shown in Figure 2.4.

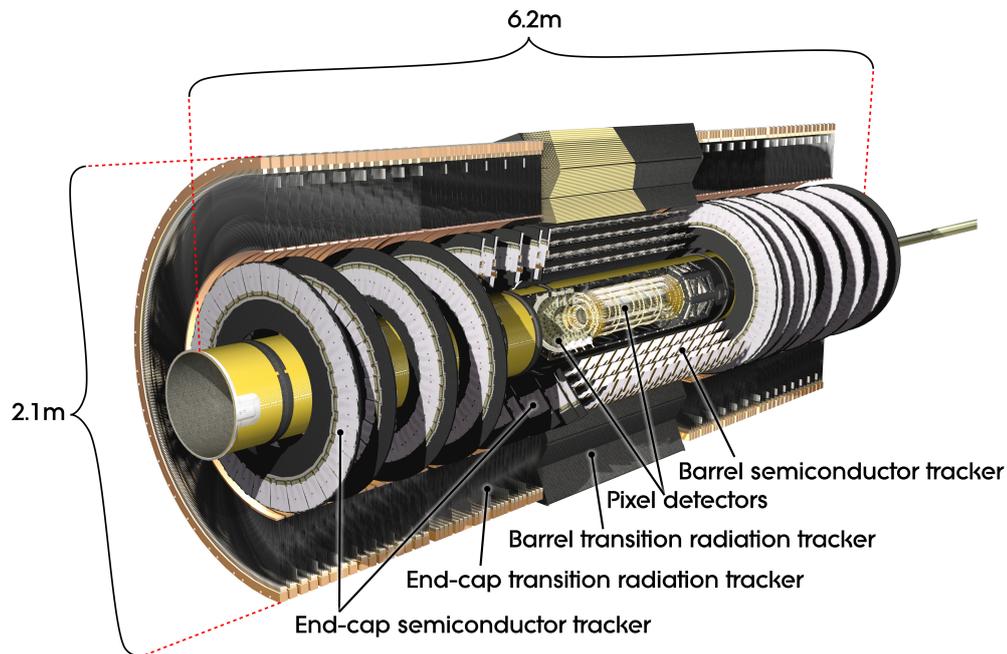


Figure 2.4. Cut-away view of the ATLAS inner detector.

Pixel detector

The pixel detector [37] is the ID subsystem closest to the beam pipe that gives the maximum resolution for detecting charged particle trajectories (tracks). Four layers of silicon pixel sensors are embedded in the device. As a charged particle travels through it, each pixel layer contributes to a three-dimensional (R, ϕ, z) space point measurement (hit). During LS1, an additional tracking layer - the Insertable B-Layer (IBL) [38, 39] was inserted as the innermost pixel barrel layer (Figure 2.4). Even in the face of a significant number of particle tracks created in greater brightness and centre-of-mass energy collisions, the IBL enhances the overall tracking efficiency of the pixel detector.

Three outer pixel layers are placed as three cylindrical barrel layers and three perpendicular discs in each end cap area to complement the IBL. With only the barrel region, position mea-

surement in IBL sensors of size $50 \times 250 \mu\text{m}^2$ has an inherent azimuthal ($R - \phi$ plane) resolution of $8 \mu\text{m}$ and an axial (z) resolution of $40 \mu\text{m}$. The last three pixel layers are $50 \times 400 \mu\text{m}^2$ sensors, giving $10 \mu\text{m}$ azimuthal resolution in the $R - \phi(z - \phi)$ plane and $115 \mu\text{m}$ axial resolution in the $z(R)$ direction over the barrel (end caps) unit.

SemiConductor Tracker

The Semiconductor Tracker (SCT) [40, 41] has four barrel layers and nine endcap discs of silicon microstrip sensor modules. Each module consists of four strip sensors, two at the top and two at the bottom. To capture a full three-dimensional position measurement, the top and bottom pair of sensors on opposite sides of a single module are arranged at a relative angle of 40 mrad . In the barrel and end cap components, the strip sensors will have an azimuthal resolution of 17 m and an axial resolution of 580 m .

Transition Radiation Tracker

The Transition Radiation Tracker (TRT) [42, 43] consists of 4 mm diameter drift tubes (straws) filled with an active xenon-based gas mixture. Prior to LS1, sites affected by gas leaks were fumigated with argon gas as a cheaper alternative to xenon 24. During the LS1 shutdown phase, the LHC and the ATLAS detector will be switched off. The gap between the tubes is filled with polypropylene polymer, which acts as a transition radiation material. The ionisation of the TRT gas occurs when charged particles cross the straw tubes, resulting in two-dimensional hits in the azimuthal plane with an inherent resolution of $130 \mu\text{m}$ per straw. The TRT is estimated to contribute ~ 36 hits to precision tracking within $|\eta| < 2.0$ due to its coarser resolution than the pixel and SCT detectors. As the particles pass through the polymer material, they produce low-energy transition radiation photons in addition to ionisation impacts. The probability of producing these photons is determined by the relativistic Lorentz factor $\gamma = \frac{E}{m}$ of the incoming particle, so electrons are more likely to emit than charged pions. This property allows us to distinguish electrons from charged pions.

2.3.4. The calorimeter system

Calorimeters [44] are devices that measure the energy of particles and are used to reconstruct electrons, photons, and jets. Layers of passive and active materials make up calorimeters. The passing particles interact with the passive substance, causing a showering effect. The active material is utilized to calculate the flow of particles created in the shower, which is proportional to the energy of the particle that started it.

The calorimeter utilized in the ATLAS experiment is made up of two detectors: an electromagnetic and a hadronic calorimeter, as illustrated in Figure 2.5. Both detectors are intended to detect and stop particles by measuring their energy. The electromagnetic calorimeter, made of lead and liquid argon, measures the energy of electrons and photons, whereas the hadronic calorimeter, made of steel and scintillators, measures the energy of hadrons that the thinner but higher-resolution electromagnetic calorimeter could not record.

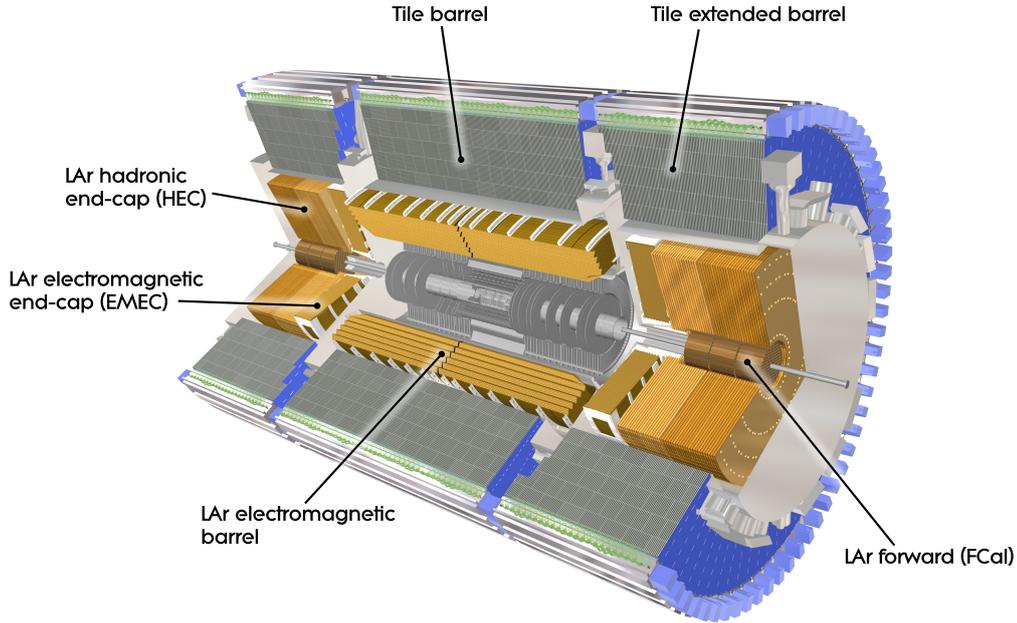


Figure 2.5. Cut-away view of the ATLAS calorimeters.

2.3.5. The muon spectrometer

To monitor muon trajectories bent by magnetic deflection, the Muon Spectrometer (MS) [45] is entirely integrated into three huge air-core toroids. One barrel ($|\eta| < 1.0$) and two end-cap ($1.4 < |\eta| < 2.7$) toroids have eight coils each positioned radially around the beam pipe. The area between $1.0 < |\eta| < 1.4$ represents the transition region of the barrel and end-cap detector components. The air core reduces the amount of material density that muons must travel through. Multiple scattering is reduced as a result, and the tracking resolution improves. Due to the overlapping of the barrel and end-cap fields in the transition zone, magnetic deflection occurs, and performance has been significantly improved by adding extra detectors during LS1. The muon detector is shown in Figure 2.6.

To estimate the momentum and location of muon tracks in the bending ($R - z$) plane, the high-precision muon tracking system employs two different detection technologies. The Monitored Drift Tube (MDT) chambers, which consist of three to eight layers of drift tubes filled with pressurized argon gas, are primarily responsible for precise muon tracking. Muon hits in the ionised gas are detected with a single-hit spatial resolution of $80 \mu\text{m}$. The MDT chambers are replaced by Cathode Strip Chambers (CSCs), multiwire proportional chambers with better rate capability and increased spatial resolution of $60 \mu\text{m}$, due to high radiation levels in the area of $2.0 < |\eta| < 2.7$. The MS uses precise trigger chambers called Resistive Plate Chambers (RPCs) and Thin Gap Chambers (TGCs) to trigger muon tracks in the area ($1.05 < |\eta| < 2.4$). Both RPCs ($|\eta| < 1.05$) and TGCs ($1.05 < |\eta| < 2.4$) add to the tracking chambers' capabilities

by measuring and coordinating muon tracks.

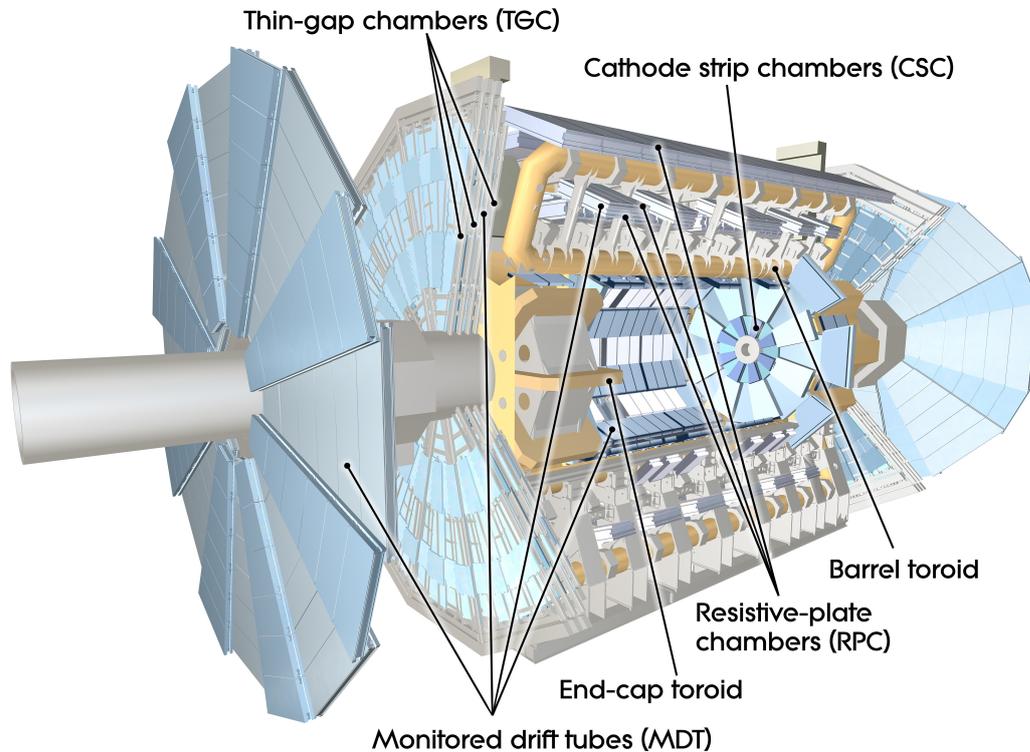


Figure 2.6. Cut-away view of the ATLAS muon spectrometer.

2.3.6. Trigger system

The LHC is planned to deliver a collision rate of 40 MHz with a nominal bunch spacing of 25 ns. For each event recorded by the ATLAS detector, tens of thousands of read-out channels are used, which amounts to about 1.5 MB of information to write to disk per event. Therefore, a 60 TB/s rate would be necessary to record all the events produced in the collisions, which is limited by the current capabilities of $\mathcal{O}(1)$ GB/s. Due to technological and computer resource restrictions, only a percentage of the entire event rate can be recorded. At a tolerable readout rate of 1 kHz, the ATLAS Trigger and Data Acquisition System TDAQ (Figure 2.7) [46–48] captures events created in pp collisions. Furthermore, it records occurrences that are relevant to physics research. The trigger system consists of two types of triggers: a hardware-based Level-1 (L1) trigger and a single software-based High-Level Trigger (HLT).

The L1 trigger uses customised electronics and is instrumented directly in the detector to select events based on the calorimeters and muon detectors and is implemented using specific hardware. This trigger level can identify events containing electrons, muons, τ -jets, photons, and missing transverse energy, and it's aimed to cut the rate of events recorded to 75 thousand

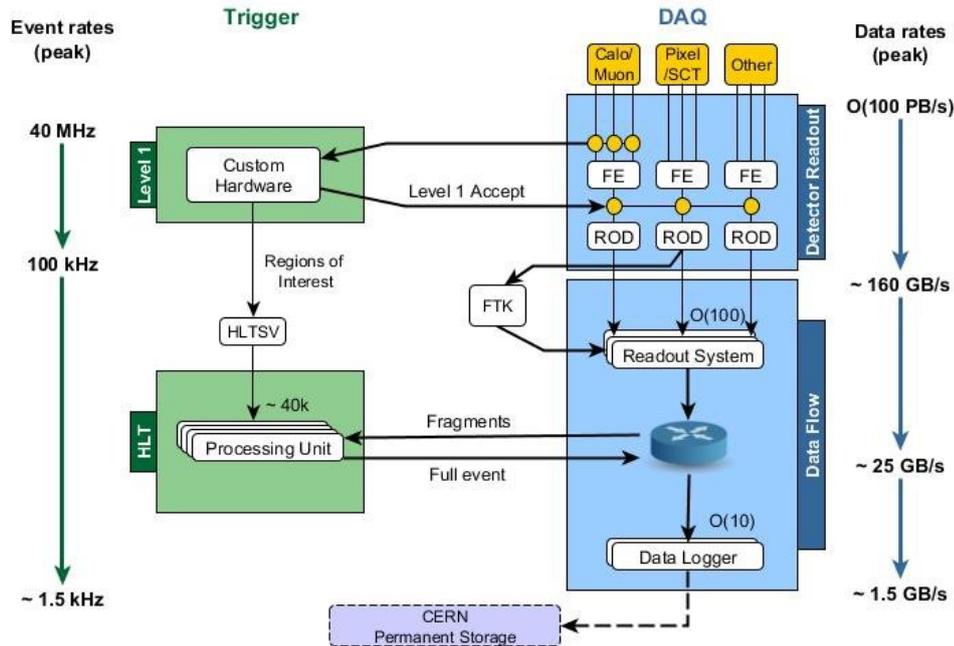


Figure 2.7. The ATLAS DAQ System in LHC Run 2. Events passing the Level 1 hardware trigger (top left) are passed to the HLT (bottom left) via the farm supervisor node (HLTSV), now including assembly of Regions of Interest. Simultaneously, event data from the detector front-end electronics systems are sent to the Readout System (ROS) via optical links from the Readout Drivers (RODs) in response to a Level 1 trigger accept signal. These data are then buffered in the ROS and made available for sampling by algorithms running in the HLT. Once the HLT accepts an event it is sent to permanent storage via the Data Logger.

per second. It reduces the bunch crossing rate of 40 MHz to the readout event rate of 100 kHz with a decision latency of less than $2.5 \mu\text{s}$.

Events that pass the first level trigger are temporarily kept to allow the second level trigger to reconstruct the events and determine whether or not they should be recorded to disk based on event attributes indicated as a region of interest. The second level trigger, also known as the Event Filter, uses a large computational cluster near the detector to reconstruct all the events passed by the level 1 trigger and further reduces the readout event rate to 1 kHz. Eventually, 1000 events are written to disk per second, and the entire trigger system and the data acquisition process will take about 4 seconds for the event to finally be selected.

Part I

ATLAS Qualification task

Chapter 3

Eigenvector decomposition study related to b -tagging pseudo-continuous data-based calibration

3.1. Introduction

The identification of jets containing b -hadrons (b -jets), c -hadrons (c -jets) or neither b - nor c -hadrons (light-flavour jets) is an important part of the physics programme of the ATLAS experiment at the Large Hadron Collider (LHC) [49]. The ATLAS Collaboration has developed several algorithms to identify b -jets, known as b -tagging algorithms, using data from LHC Run 2 (2015-2018). These algorithms exploit several properties of b -hadrons, including their long lifetime, high mass, and the charged-track multiplicity of their decays. The probability that a b -tagging algorithm correctly identifies a b -jet is called the b -tagging efficiency $\epsilon(b)$.

The performance of a b -tagging algorithm is evaluated using b -jet tagging single-cut operating points (OPs). These are based on a fixed selection requirement on the output distribution of a b -tagging algorithm such that a given b -tagging efficiency, $\epsilon(b)$, is evaluated for the b -jets present in simulated top quark pair events. Four single-cut OPs are currently defined at 60%, 70%, 77% and 85%. Simulation-to-data scale factors (SFs) [50] are derived as a function of the jet transverse momentum (p_T) for each OP as the ratio of the measured b -jet tagging efficiency to the b -jet tagging efficiency evaluated from the simulation. The SFs are then applied in the physical analysis as a per-jet weight to the jets in the simulation.

Given a lifetime of the order of 1.5 ps ($\langle c\tau \rangle \approx 450 \mu m$), measurable b -hadrons [51] have a significant mean flight length $\langle l \rangle = \beta\gamma\tau c$ in the detector before decaying, generally leading to at least one vertex displaced from the hard scattering collision point. The strategy developed by the ATLAS collaboration is based on a two-step approach. First, low-level algorithms reconstruct the characteristic features of the b -jets using two complementary approaches: one uses the individual properties of the charged particle tracks, the so-called tracks associated with a hadronic jet. The second combines the tracks into explicitly reconstructed displaced vertices. These algorithms, first introduced in Run 1, have been improved and re-tuned for Run 2. Second, to maximise b -tagging performance, the results of the low-level b -tagging algorithms are combined with high-level algorithms consisting of multivariate classifiers. The performance of a b -tagging

algorithm is characterised by the probability of tagging a b -jet (b -jet tagging efficiency, ϵ_b) and the probability of misidentifying a c -jet or a light flavour jet as a b -jet, denoted ϵ_c (ϵ_l).

The primary aim of this analysis is to explore the pseudo-continuous framework, focusing on the format of pseudo-continuous inputs, their differences from cumulative inputs, the number of systematics, and their treatment. By comparing the pseudo-continuous approach with cumulative inputs, we also highlight the activity of different high-level taggers [52]. In addition, this study investigates how Eigenvector (EV) decomposition works within pseudo-continuous inputs. Many systematics are close to zero after EV decomposition, which significantly improves the smoothing process of the tagweight bin.

This note is structured as follows: Section 3.2 gives an overview of flavour tagging algorithms. Sections 3.3 and 3.4 cover the details of the ATLAS tracking system and flavour tagging calibrations respectively. Section 3.5 explains the pseudo-continuous calibration and the structure of the Calibration Data Interface (CDI). The Section 3.6 discusses post-processing along with eigenvector decomposition. Scale factor re-scaling plots for tag-weight bins and systematics can be found in Sections 3.7 and 3.8. Finally, conclusions are drawn in Section 3.9.

3.2. Flavour tagging algorithms

This section describes the different algorithms used for b -jet identification [53] and the evaluation of their performance in simulation. Flavour tagging uses the latest developments in machine learning and neural networks [54]. Low-level b -tagging algorithms fall into two broad categories. The first approach, implemented in the IP2D [55] and IP3D [56] algorithms, uses the signed impact parameter [57]. The second approach explicitly reconstructs displaced vertices. The SV1 [58] algorithm attempts to reconstruct an inclusive secondary vertex [59], while the JetFitter [60] algorithm aims to reconstruct the complete b - to c -hadron decay chain. To maximise b -tagging performance, the results of the low-level algorithms are combined using multivariate classifiers. Three high-level tagging algorithms have been developed. The first, MV2, is based on a boosted decision tree (BDT) [61] discriminant, while the second, DL1, is based on a deep feed-forward neural network (NN) [62], and the last is DL1r (RNNIP).

3.2.1. Low-level b -tagging algorithms

The identification of jets containing b -hadrons relies on distinctive features that are leveraged through various algorithmic strategies, as described below:

Impact parameter based algorithms (IP2D and IP3D):

The IP2D and IP3D algorithms use the signed impact parameter significance of the tracks associated with the jet. IP3D uses both the transverse (d_0) and longitudinal (z_0) impact parameter significances, taking into account their correlations, while IP2D uses only the transverse (d_0) significance. The probability density functions (pdfs) for the signed impact parameter significance



Figure 3.1. Low-level b -tagging algorithms.

of these tracks are used to define ratios of the b , c and light flavour jet hypotheses, which are then combined into three log-likelihood ratio discriminants (LLRs): b/LF , b/c and c/LF .

Secondary vertex finding algorithm (SV1):

SV1 aims to explicitly reconstruct an inclusive displaced secondary vertex within the jet. The first step is to reconstruct two-track vertices from the candidate tracks. Tracks are discarded if they form a secondary vertex that can be identified as likely originating from the decay of a long-lived particle (e.g. K_s or Λ), photon transformations or hadronic interactions with the detector material. A new vertex is then fitted with all tracks that survive this selection, and outlier tracks are iteratively removed from this set of tracks.

Decay chain multi-vertex algorithm (JetFitter):

JetFitter exploits the topological structure of weak b - and c -hadron decays inside the jet and tries to reconstruct the entire b -hadron decay chain. A Kalman filter is used to find a common line on which the primary vertex [63] and the heavy flavour (HF) vertices lie, approximating the b -hadron flight path as well as their positions. Hence, HF vertices can be resolved even when only a single track is attached to them whenever the resolution allows. Fig 3.1 schematically represents the working procedure of Low-level b -tagging algorithms.

3.2.2. High-level taggers

The high-level taggers combine the inputs from several low-level taggers using a multivariate classifier to maximise the b -tagging performance. In the 2017 configuration [64], ATLAS has

adopted two variants of high-level taggers. The first is an evolution of the Boosted Decision Tree (BDT) discriminant already used in the past (MV2), and the second is a new tagger (DL1) based on Deep Learning NN.

MV2 tagger

The first high-level tagger is a BDT discriminant that combines the output of the low-level taggers. It is trained using the ROOT Toolkit for Multivariate Data Analysis (TMVA) [65] on the hybrid sample. The b -jets are considered to signal, and the c - and light-flavour jets are the background. The performance is evaluated separately on the $t\bar{t}$ and Z +jets samples. The list of input variables includes the p_T and η of the jets, as they give useful information in interpreting the separation power of the variables from the low-level taggers.

The effect of the finite number of events available for the MC samples for the MV2 training is estimated by performing cross-training tests where independent sub-samples are used as input to the BDT classifier.

DL1 tagger

The second high-level tagger (DL1) [66] is based on an Artificial Deep Neural Network. DL1 is trained using Keras with the Theano backend and the Adam optimiser. The DL1 NN has a multidimensional output corresponding to the probabilities for a jet to be a b -, c - or light-flavour jet. Its topology consists of a mixture of fully-connected hidden layers and max-out layers. The parameters of DL1 are optimised, including the architecture of the NN, the number of training epochs, the learning rates and training batch sizes.

DL1r tagger

In addition to DL1 features, RNNIP [48] and network architecture optimisation significantly improve light and charm rejection, which results in DL1r. Fig 3.2 schematically explains the complex process of High-level b -tagging algorithms.

3.3. Tracking in ATLAS

Tracks used for b -tagging are reconstructed using the ATLAS Inner Detector (ID) immersed in a 2T magnetic field (essential for measuring track momentum). The innermost layer is the insertable B layer with a radius of ~ 30 mm (silicon pixels), which greatly improves the track resolution. Subsequent layers (radially outwards), pixel layers, semiconductor tracker and transition radiation tracker, which tracks charged particles, are reconstructed from hits in the ID associated with jets using ΔR (track, jet) matching (jet p_T dependent).

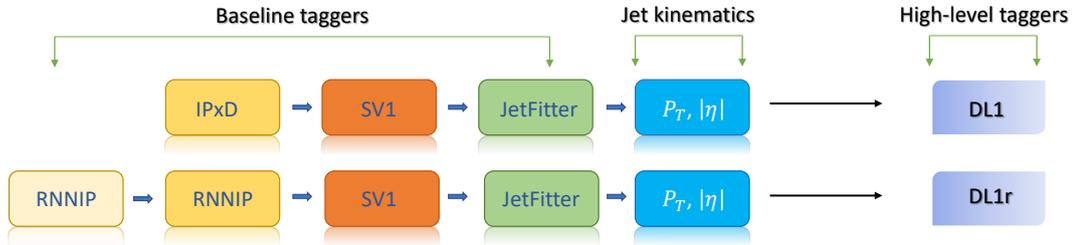


Figure 3.2. High-level b -tagging algorithms.

3.4. Flavour tagging calibrations

Calibration conceptualises the process of verifying the performance of the data. Correction factors related to detector and modelling effects are applied here. The b -jet tagging efficiency measurements are provided in the form of jet transverse momentum-dependent scale factors that correct the b -tagging performance in the simulation to that observed in the data.

3.4.1. $t\bar{t}$ calibration

The b -jets tagging efficiency is measured in data using di-leptonic $t\bar{t}$ events [67] which are 70% pure in b -jets, with almost no c -jet contamination. The b -jet efficiency is extracted from fit to data. It is dominated by systematic uncertainties from modelling of $t\bar{t}$ background. Systematics arising from b -tagging is a key factor in many analyses. The range of uncertainties varies between 2-5% over the majority of the spectrum. The largest experimental systematic uncertainties have been observed in $VH(\rightarrow bb)$ analysis.

3.4.2. Cumulative and pseudo-continuous calibration

A cut at different values is applied to the MV2c10/DL1/DL1r one-dimensional distribution. After passing this cut, the events are considered tagged. There are two ways of b -tagging calibration:

In cumulative calibration, SFs are computed for different working points (WP_s) (efficiency cuts) and fixed cuts. Cumulative working points, corresponding to different efficiencies:

- FixedCutBEff_60 < tagweight
- FixedCutBEff_70 < tagweight
- FixedCutBEff_77 < tagweight
- FixedCutBEff_85 < tagweight

In pseudo-continuous calibration, SFs are computed all together over the 5-bin b -tagging weight distribution. The b , c , and light jets are calibrated using their specific methods. In the pseudo-continuous tagger, a different approach is used:

- FixedCutBEff_60 < tagweight
- FixedCutBEff_70 < tagweight < FixedCutBEff_60
- FixedCutBEff_77 < tagweight < FixedCutBEff_70
- FixedCutBEff_85 < tagweight < FixedCutBEff_77
- tagweight < FixedCutBEff_85

The information is redundant between cumulative and pseudo-continuous tagging. For FixedCutBEff_60 < tagweight, the SFs should be identical in both implementations. All calibrations are derived from the pseudo-continuous calibrations, and the cumulative working points are extracted from them, so FixedCutBEff_60 < tagweight is the only bin that needs to be consistent between them. This information can be plotted in several ways. In this note, it is plotted from the pseudo-continuous data and compared with the cumulative working points for FixedCutBEff_60 < tagweight.

3.4.3. b -tagging scale factors

Flavour-tagging scale factors are ratios of efficiencies (ϵ) applied to simulated events to correct the tagging rate in the simulation to data. They are provided in bins of jet p_T and $|\eta|$ and defined as:

$$SF_{\text{Flavour}}(p_T, \eta) = \frac{\epsilon_{\text{Flavour}}^{\text{Data}}(p_T, \eta)}{\epsilon_{\text{Flavour}}^{\text{MC}}(p_T, \eta)}, \quad (3.1)$$

$$w_{\text{jet}} = SF_{\text{Flavour}}(p_T, \eta), \quad (3.2)$$

$$w_{\text{jet}} = \frac{1 - \epsilon_{\text{Flavour}}^{\text{Data}}(p_T, \eta)}{1 - \epsilon_{\text{Flavour}}^{\text{MC}}(p_T, \eta)} = \frac{1 - SF_{\text{Flavour}}(p_T, \eta) \epsilon_{\text{Flavour}}^{\text{MC}}(p_T, \eta)}{1 - \epsilon_{\text{Flavour}}^{\text{MC}}(p_T, \eta)}, \quad (3.3)$$

To avoid boundary effects and prevent distortions in the distributions of interest for the analysis relying on flavour tagging, the scale factors are smoothed. The method of smoothing is local polynomial kernel estimation [68] and the Eigen Vector decomposition method.

3.5. Pseudo-continuous calibrations

The calibration does not need to be truly continuous; for most applications, it is sufficient to calibrate the tagweight distribution in a limited number of bins. The continuous b -tagging calibration in ATLAS has been implemented by calibrating the tagweight distribution in five bins, whose boundaries correspond to the following b -jet efficiencies [100%; 85%; 77%; 70%; 60%; 0%].

Pseudo-continuous calibrations use 3D textual input provided by calibration groups. It also includes an extra dimension for the tagweight, which indicates where they live in the pseudo-continuous regime.

3.5.1. Pseudo-continuous

b -tagging calibration is the process of correcting the MC after applying the b -tagging algorithm. It is called pseudo-continuous because it lies between a “continuous” calibration (using fine binning) and a cumulative calibration using fixed efficiency working points (only 2 bins, corresponding to tagged and untagged cases).

The transverse momentum axis has 9 bins, the tagweight axis has 5 bins and η has one bin. p_T , η and tagweight contain the SFs information. It is possible to retrieve all the details from them.

3.5.2. Selections and technical implementation

This section focuses on the calibration of particle jets, which are thought to be produced by collisions between fundamental particles. The aim is to improve the precision of the measurements. Here’s a detailed overview of the key components involved:

Taggers: These sophisticated algorithms, including MV2c10, DL1 and DL1r, are used to identify the type of particle responsible for producing a particular jet. This could be bottom quarks, charm quarks or lighter quarks

Jet Collections: The PFlowJets algorithm serves as the primary tool for reconstructing jets from the particles detected in the ATLAS experiment

Binning: To effectively categorise the jets, they are grouped into five different categories or “bins”. These bins are determined based on a tagging efficiency threshold, providing insight into the effectiveness of the taggers in identifying specific particle types of jets

Inputs: The calibration process is driven by a variety of inputs:

- Contributions from various teams involved in flavour calibration to determine the particle type from which a jet originates

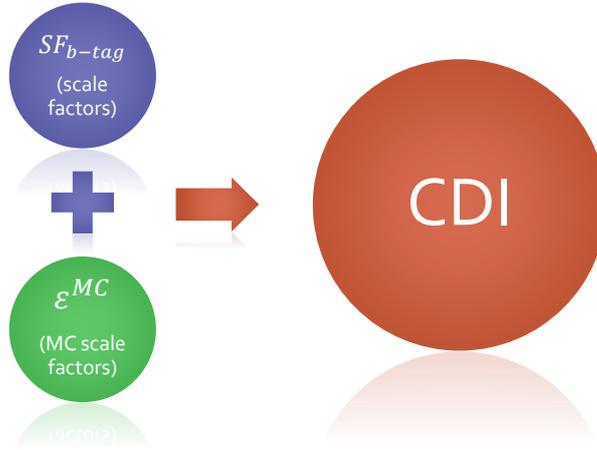


Figure 3.3. CDI structure.

- Cumulative calibration details for four labelling efficiencies (60%, 70%, 77% and 85%) applied to jets likely to be derived from lighter quarks
- Pseudo-continuous calibration data specifically tailored for bottom and charm quarks. This data is presented in a 3D format with an additional tagweight axis

At the end of the calibration process, text files containing pseudo-continuous calibration data are converted into Calibration Data Interface (CDI) files. By calibrating the jets based on their inferred origin (flavour), the aim is to refine the measurement accuracy and thereby contribute to a deeper understanding of fundamental particles and their interactions. These files are designed for seamless integration into subsequent analyses.

3.5.3. Calibration data interface (CDI) structure

The CDI (Figure 3.3) file is a lookup table containing all the information needed to calibrate taggers. It defines tagger-specific operating points and fractions. It also contains or calculates calibrations for (non-)tagged jets. CDI accounts for differences in MC generators. The physics analysis receives a CDI file with the following recommendations. The CDI workflow is summarised in Figure 3.4.

- Definition of working points
- Flavour fractions for building the taggers discriminant
- Efficiency maps for the inefficiency SFs
- Efficiency maps for the MC/MC SFs
- Simulation to data SFs

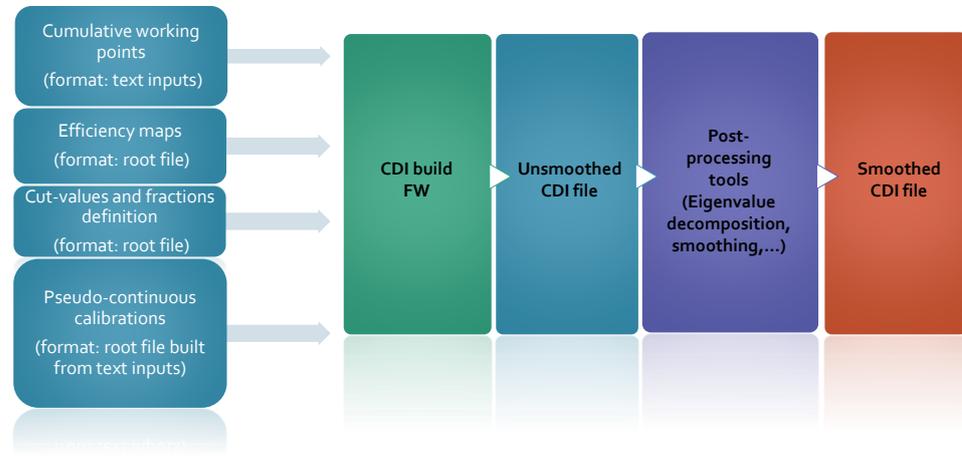


Figure 3.4. CDI workflow.

3.6. Post-processing

Fitting a curve of a particular functional form to a data set is a fundamental and crucial task in analysis. These curves are essential for interpolating or extrapolating into important regions of phase space. Therefore, choosing the correct functional form for the fitted curve is of great importance.

Local Polynomial Kernel Estimation: The non-parametric method used in the calibration file described is Local Polynomial Kernel Estimation (LPKE). This method can be conceptualised as fitting a polynomial by weighted least squares, where the weights are functions of the covariates. These weight functions, called kernels, are typically chosen to be uniform and integrated to unity. For regression, however, the requirement of integrability can be relaxed. A common kernel in physics, and the one used here, is a normalised Gaussian kernel.

Eigenvalue Analysis: Eigenvalue decomposition is performed on the covariance matrix of the systematic and statistical variations. The resulting number of variations is equal to the number of bins in the scale factor distribution. This process is performed dynamically at the software level.

3.6.1. Eigenvector decomposition

Each scale factor distribution in a CDI dataset has up to 40 systematic and statistical variations that need to be taken into account. Some physical analyses do not require the full tagweight binning provided but can work with a coarser binning. In such a case, reducing the number of bins will simplify the use of the calibration. Eigenvalue decomposition on the covariance matrix of the systematic and statistical variations preserves the bin-to-bin correlations and the total error. After smoothing, most EVs are very small in magnitude and can be removed without significantly affecting the correlations or total error. These systematics are reduced by isolating kinematic regions where certain systematics dominate and merging (by quadrature sum) the least dominant systematics in these regions. This method suffers from highly biased correlations, but only in isolated regions.

The correct inclusion of scale factor uncertainties in a physical analysis requires the construction of the full-scale factor covariance matrix V from the individual uncertainty covariance matrices:

$$V_{ij} = \sum_{u=1}^{N_{\text{unc}}} V_{ij}^u, \quad (3.4)$$

here the bins i and j represent combinations of kinematic and tagweight bins. The size of the covariance matrix, N , is equal to $\eta_{p_T} \times \eta_{\eta}$ for the b and c jet calibrations, and $\eta_{p_T} \times \eta_{\text{tag}} \times \eta_{\eta}$ for the light flavour jet calibrations. Note that for the tagweight variable, only $\eta_{\text{tag}} - 1$ bins are independent for any given kinematic bin, due to the normalisation constraint.

The covariance matrix is then inverted: this gives the weight matrix V^{-1} . In principle, at this point, the uncertainties associated with the b -tagging efficiency measurements can be expressed in terms of an auxiliary likelihood term such as:

$$\mathcal{L} \propto \exp \left[- \sum_{ij} (SF_i - SF_{\text{nom},i}) V_{ij}^{-1} (SF_j - SF_{\text{nom},j}) / 2 \right], \quad (3.5)$$

The weight matrix is diagonalised and a change of variables from the original scale factor variations to the basis of the eigenvectors ($SF - SF_{\text{nom}} \rightarrow \vec{y}$) is applied so that the auxiliary likelihood term becomes a product of N one-dimensional Gaussians:

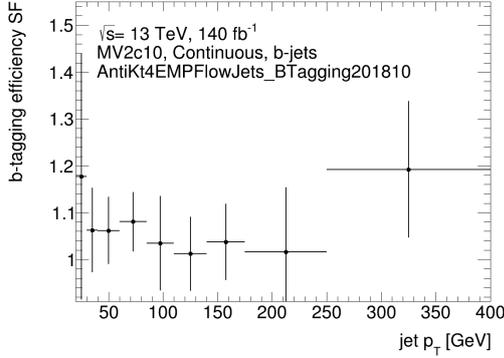
$$\mathcal{L} \propto \prod_{k=1}^N \exp \left[- \frac{y_k^2}{2\sigma_k^2} \right], \quad (3.6)$$

where $1/\sigma_k^2$ are the eigenvalues of the diagonalised weight matrix. This gives a set of N -independent scale factor variations. Any physics analysis can apply the scale factor variations independently sum them in quadrature and find their effect on the final measurement.

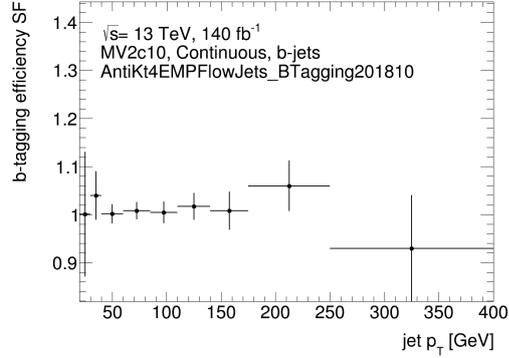
3.7. Scale factor re-scaling

When calibrating the b -tagging in the Monte Carlo (MC) simulation to reflect the measurement of b -jet efficiency in the data, event weights corresponding to MC-to-data efficiency scale factors are applied to our Monte Carlo samples. The calibration scale factors depend on the tagweight bin and the jet p_T . However, light jets also depend on jet η , but cannot account for possible topological and process-related dependencies. When the calibration scale factor is applied to a different Monte Carlo sample or region of phase space from where it was originally derived, it is assumed that the Monte Carlo correctly reproduces the ratio of efficiencies between the sample of interest and the original calibration sample. In this section, the scale factor corresponding to each tagweight bin for the pseudo-continuous calibration has been plotted as a function of transverse momentum.

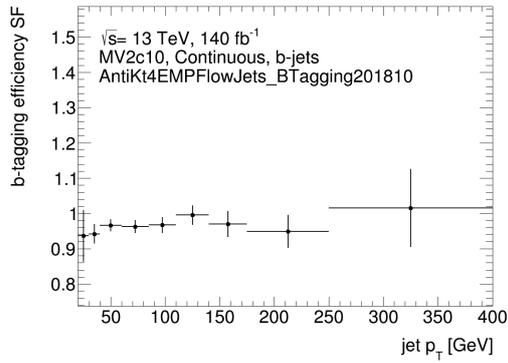
3.7.1. Scale factors for DL1 for EMPFlow jets



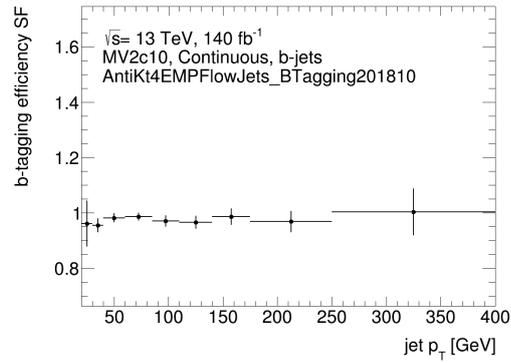
(a) tagweight < FixedCutBEff_85



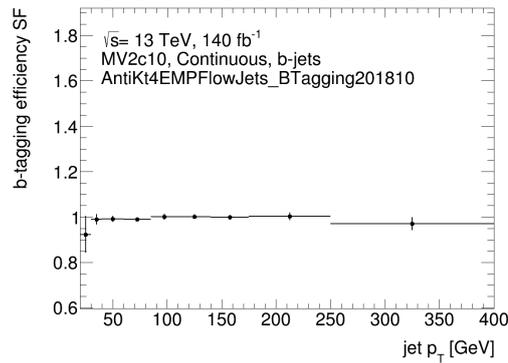
(b) FixedCutBEff_85 < tagweight < FixedCutBEff_77



(c) FixedCutBEff_77 < tagweight < FixedCutBEff_70



(d) FixedCutBEff_70 < tagweight < FixedCutBEff_60



(e) FixedCutBEff_60 < tagweight

Figure 3.5. MC-to-data calibration scale factors for b -jets for each tagweight bins. The binned SFs, and their total uncertainties, are shown by the data points which are located at the geometric centre of each bin.

3.7.2. Scale factors for DL1 for EMPFlow jets

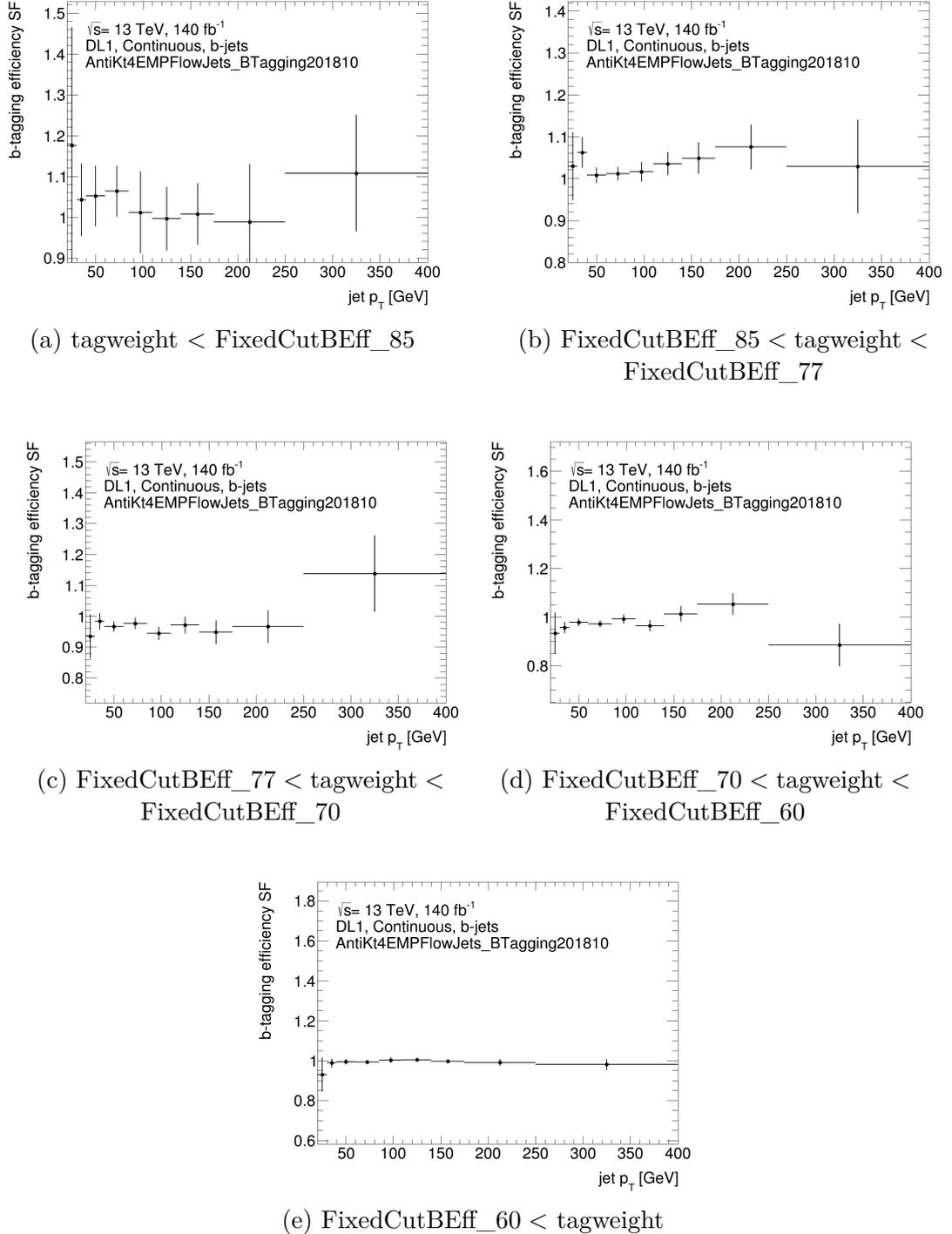
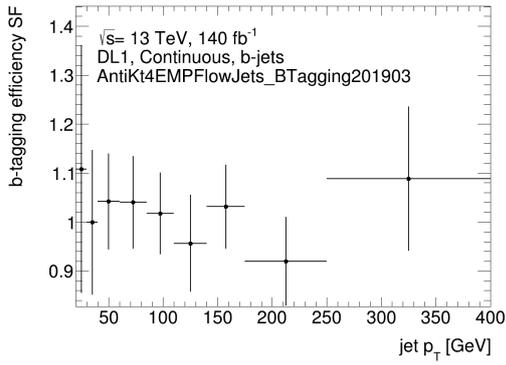
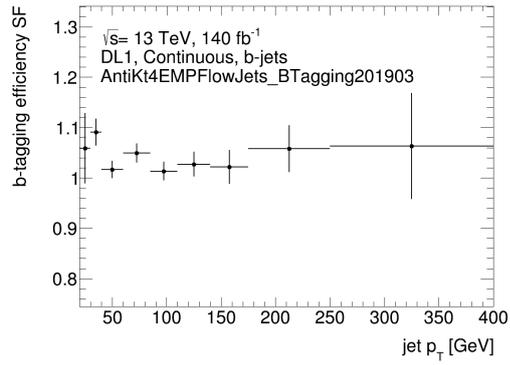


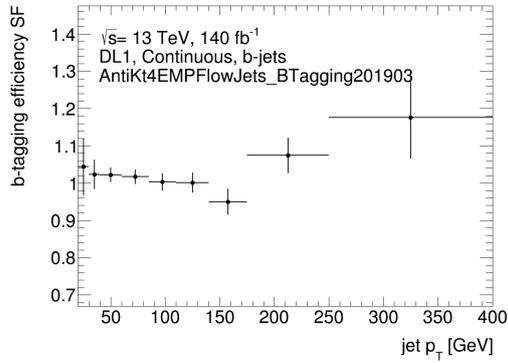
Figure 3.6. MC-to-data calibration scale factors for b -jets for each tagweight bins. The binned SFs, and their total uncertainties, are shown by the data points which are located at the geometric centre of each bin.



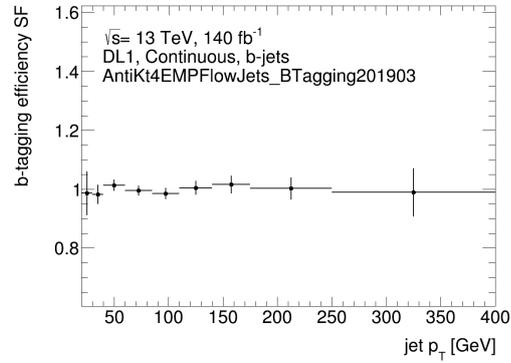
(a) tagweight < FixedCutBEff_85



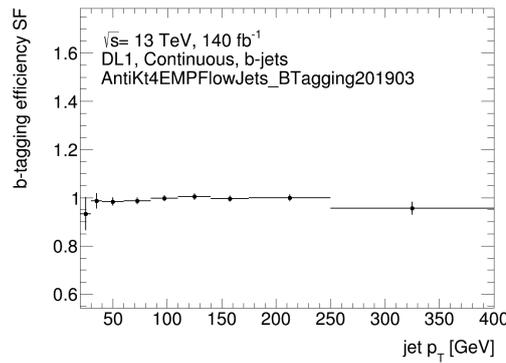
(b) FixedCutBEff_85 < tagweight < FixedCutBEff_77



(c) FixedCutBEff_77 < tagweight < FixedCutBEff_70



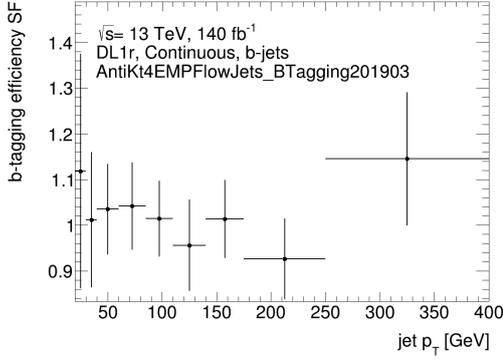
(d) FixedCutBEff_70 < tagweight < FixedCutBEff_60



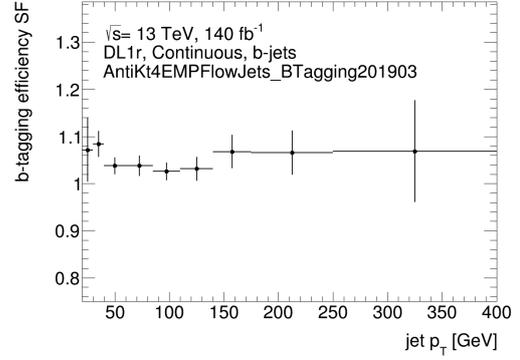
(e) FixedCutBEff_60 < tagweight

Figure 3.7. MC-to-data calibration scale factors for b -jets for each tagweight bins. The binned SFs, and their total uncertainties, are shown by the data points which are located at the geometric centre of each bin.

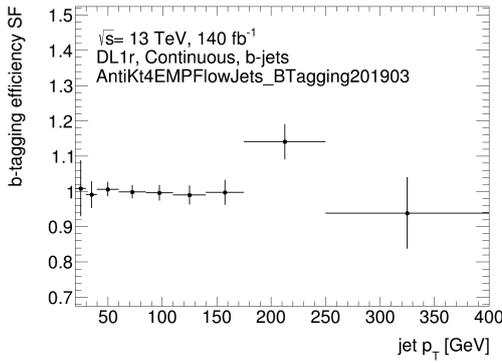
3.7.3. Scale factors for DL1r EMPFlow jets



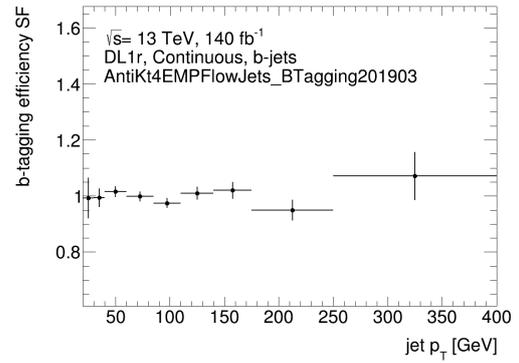
(a) tagweight < FixedCutBEff_85



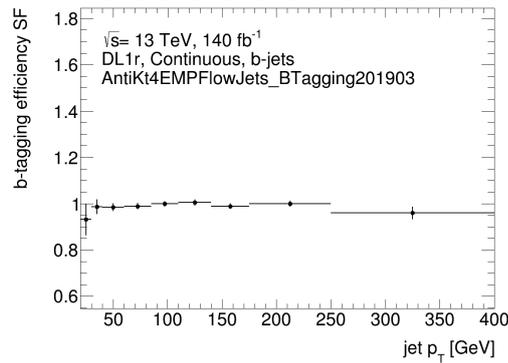
(b) FixedCutBEff_85 < tagweight < FixedCutBEff_77



(c) FixedCutBEff_77 < tagweight < FixedCutBEff_70



(d) FixedCutBEff_70 < tagweight < FixedCutBEff_60



(e) FixedCutBEff_60 < tagweight

Figure 3.8. MC-to-data calibration scale factors for b -jets for each tagweight bins. The binned SFs, and their total uncertainties, are shown by the data points which are located at the geometric centre of each bin.

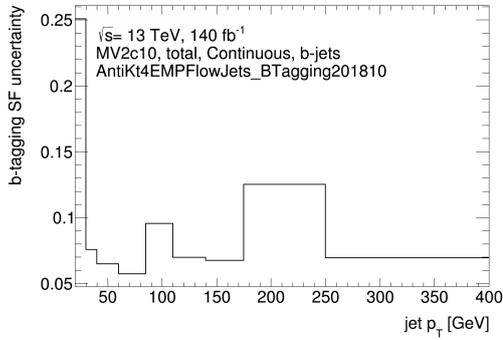
3.8. Scale factors for systematics

Uncertainties affecting the measurement which originate from statistical sources are considered together with systematic uncertainties related to the detector calibration and physics modelling. Several sources of systematic uncertainties can impact the measurement. These can broadly be grouped into four categories: experimental, multi-jet background estimation, the modelling of the simulated background and signal samples.

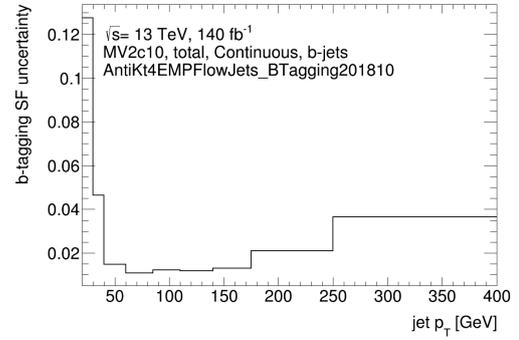
In this section, scale factors are plotted to represent systematic uncertainties for pseudo-continuous calibration as a function of transverse momentum.

3.8.1. Total systematics variation

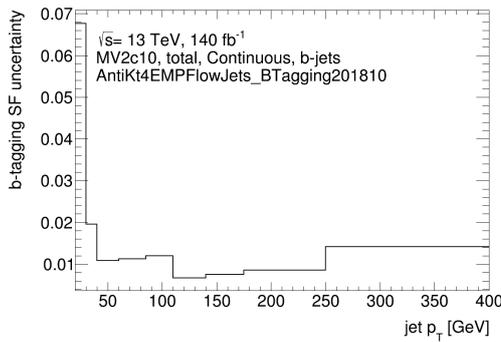
Scale factors for MV2c10 EMPFlow jets



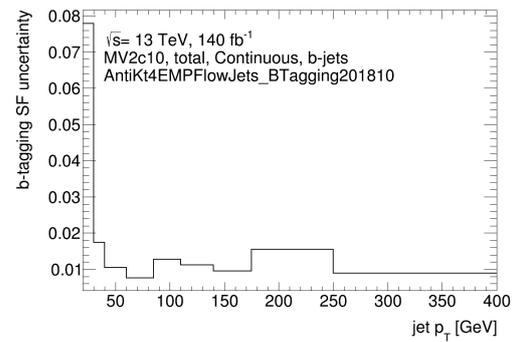
(a) tagweight < FixedCutBEff_85



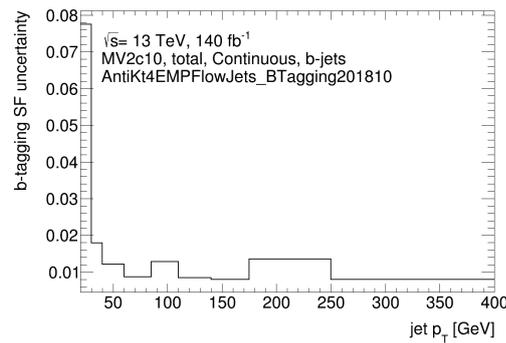
(b) FixedCutBEff_85 < tagweight < FixedCutBEff_77



(c) FixedCutBEff_77 < tagweight < FixedCutBEff_70



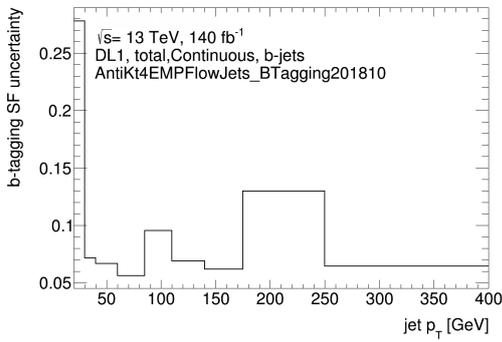
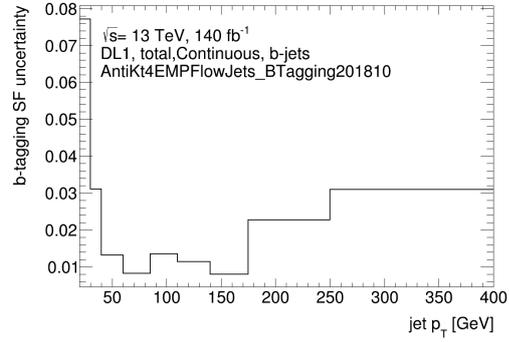
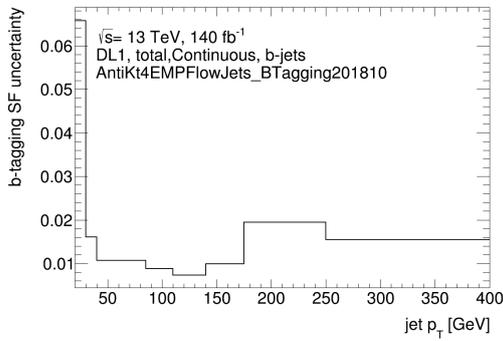
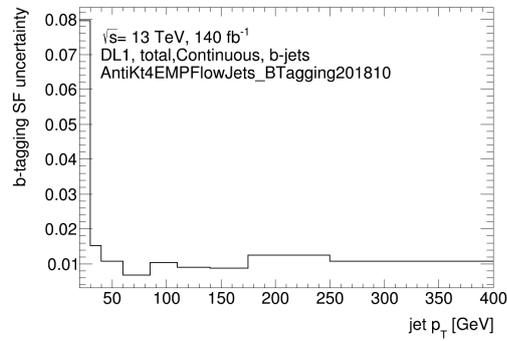
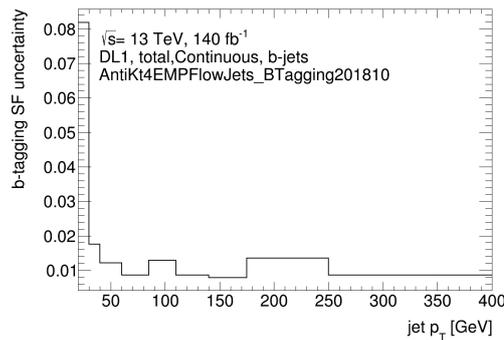
(d) FixedCutBEff_70 < tagweight < FixedCutBEff_60

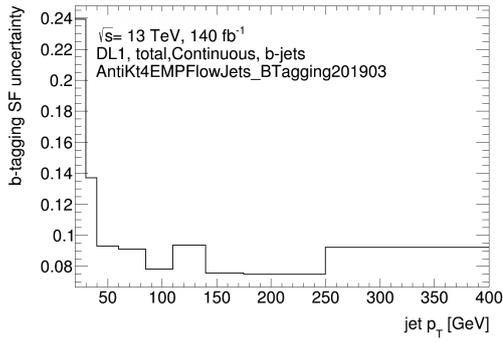


(e) FixedCutBEff_60 < tagweight

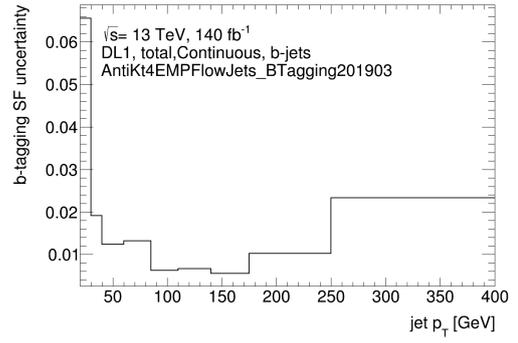
Figure 3.9. Total systematic uncertainties for b -jets for MV2c10 2018 tagger.

Scale factors for DL1 EMPFlow jets

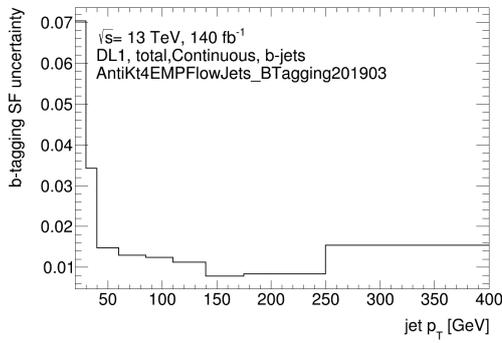
(a) $\text{tagweight} < \text{FixedCutBEff}_{85}$ (b) $\text{FixedCutBEff}_{85} < \text{tagweight} < \text{FixedCutBEff}_{77}$ (c) $\text{FixedCutBEff}_{77} < \text{tagweight} < \text{FixedCutBEff}_{70}$ (d) $\text{FixedCutBEff}_{70} < \text{tagweight} < \text{FixedCutBEff}_{60}$ (e) $\text{FixedCutBEff}_{60} < \text{tagweight}$ **Figure 3.10.** Total systematic uncertainties for b -jets DL1 2018 tagger.



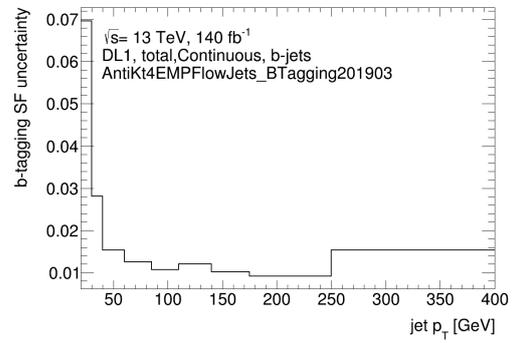
(a) tagweight < FixedCutBEff_85



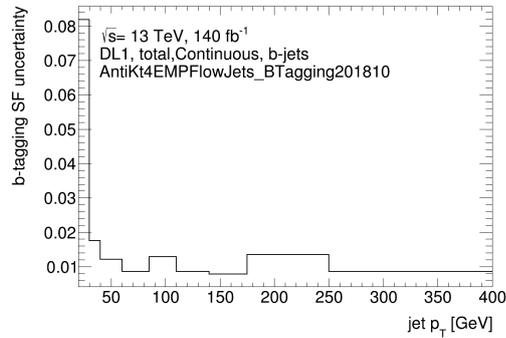
(b) FixedCutBEff_85 < tagweight < FixedCutBEff_77



(c) FixedCutBEff_77 < tagweight < FixedCutBEff_70



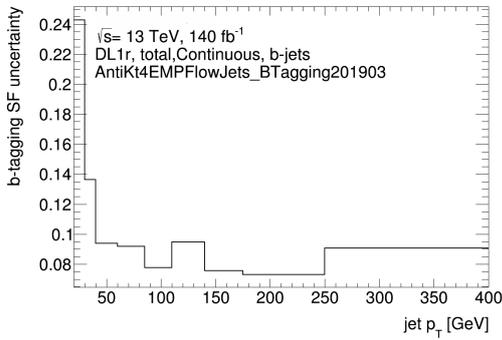
(d) FixedCutBEff_70 < tagweight < FixedCutBEff_60



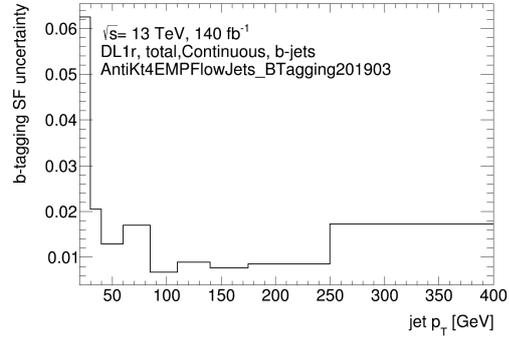
(e) FixedCutBEff_60 < tagweight

Figure 3.11. Total systematic uncertainties for b -jets DL1 2019 tagger.

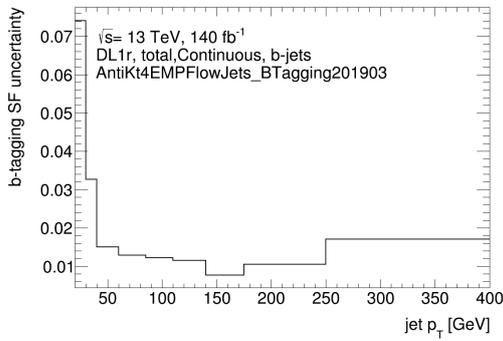
Scale factors for DL1r EMPFlow jets



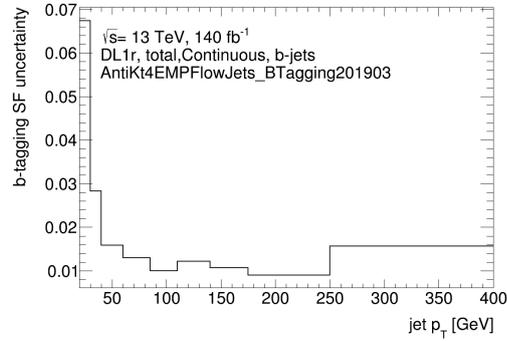
(a) tagweight < FixedCutBEff_85



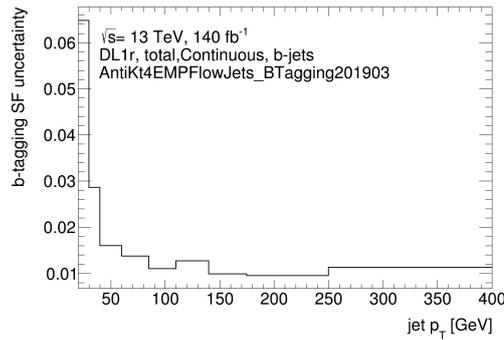
(b) FixedCutBEff_85 < tagweight < FixedCutBEff_77



(c) FixedCutBEff_77 < tagweight < FixedCutBEff_70



(d) FixedCutBEff_70 < tagweight < FixedCutBEff_60

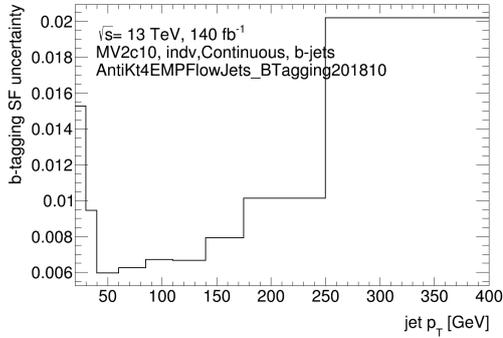


(e) FixedCutBEff_60 < tagweight

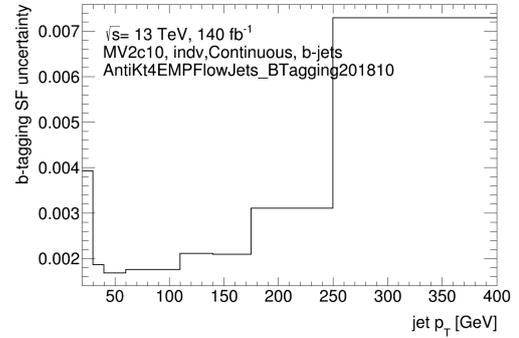
Figure 3.12. Total systematic uncertainties for b -jets DL1r 2019 tagger.

3.8.2. Individual systematic variation

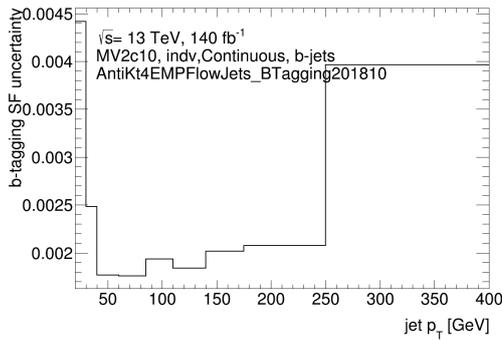
Scale factors for MV2c10 EMPFlow jets



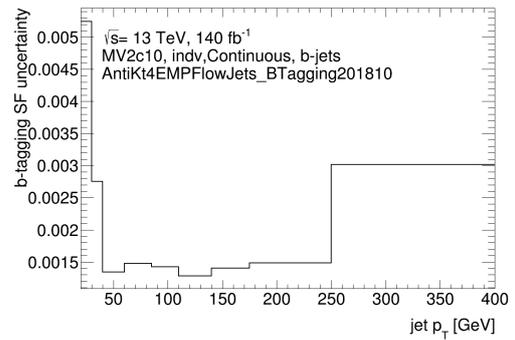
(a) tagweight < FixedCutBEff_85



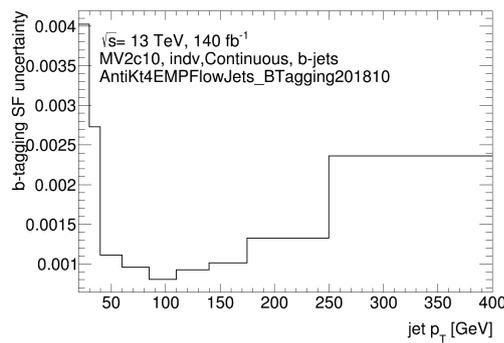
(b) FixedCutBEff_85 < tagweight < FixedCutBEff_77



(c) FixedCutBEff_77 < tagweight < FixedCutBEff_70



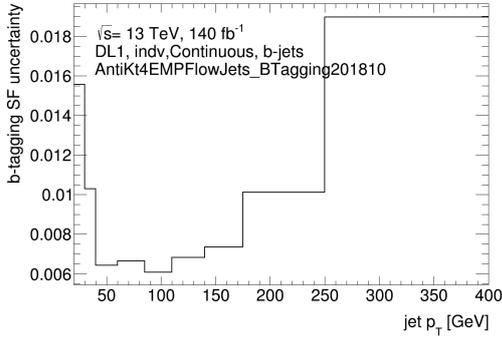
(d) FixedCutBEff_70 < tagweight < FixedCutBEff_60



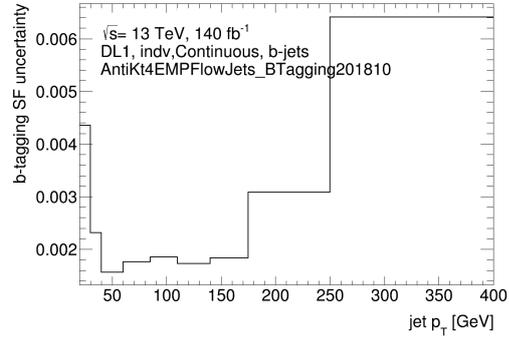
(e) FixedCutBEff_60 < tagweight

Figure 3.13. Total systematic uncertainties for b -jets for individual systematics. The variation of **FT_EFF_MC_stat_nominal** has been shown.

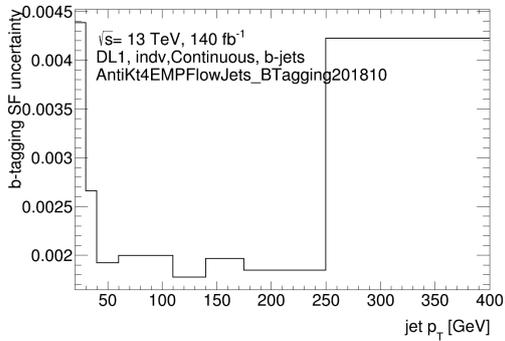
Scale factors for DL1 EMPFlow jets



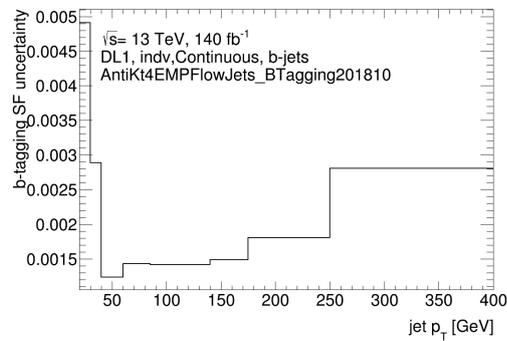
(a) tagweight < FixedCutBEff_85



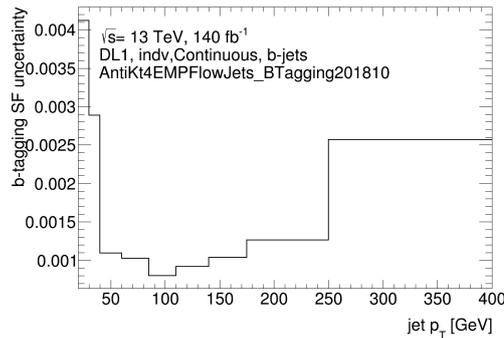
(b) FixedCutBEff_85 < tagweight < FixedCutBEff_77



(c) FixedCutBEff_77 < tagweight < FixedCutBEff_70

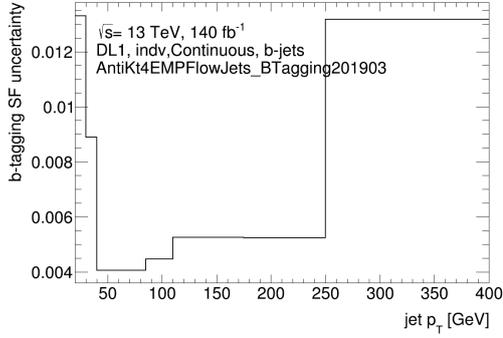


(d) FixedCutBEff_70 < tagweight < FixedCutBEff_60

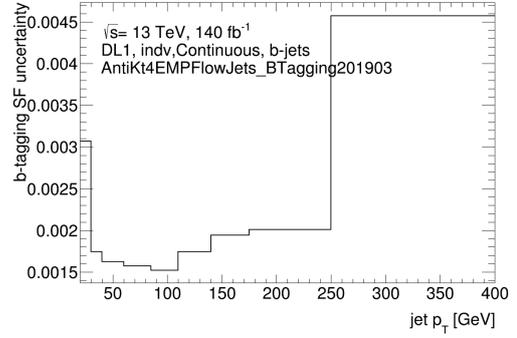


(e) FixedCutBEff_60 < tagweight

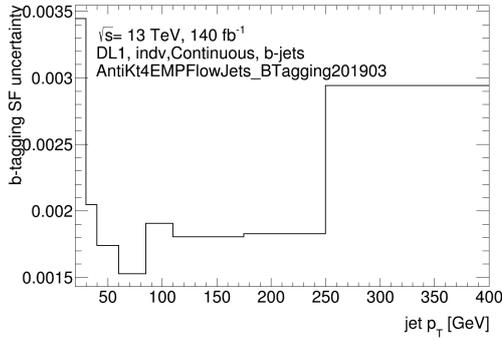
Figure 3.14. Total systematic uncertainties for b -jets for individual systematics. The variation of FT_EFF_MC_stat_nominal has been shown.



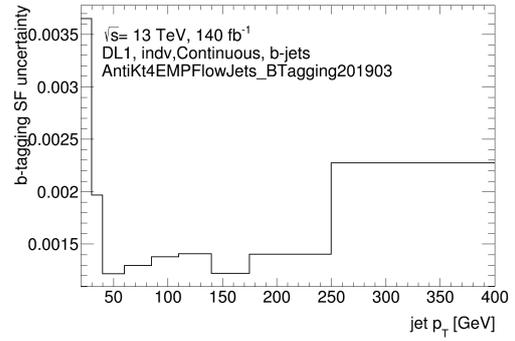
(a) tagweight < FixedCutBEff_85



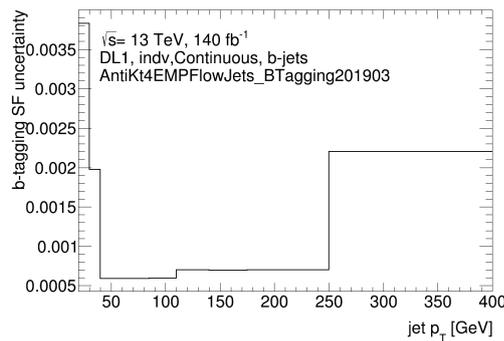
(b) FixedCutBEff_85 < tagweight < FixedCutBEff_77



(c) FixedCutBEff_77 < tagweight < FixedCutBEff_70



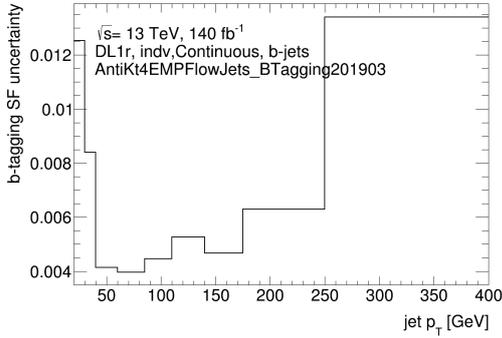
(d) FixedCutBEff_70 < tagweight < FixedCutBEff_60



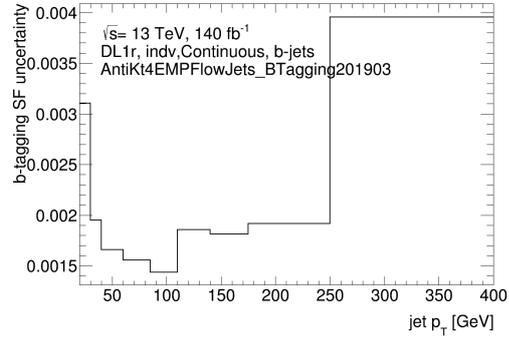
(e) FixedCutBEff_60 < tagweight

Figure 3.15. Total systematic uncertainties for b -jets for individual systematics. The variation of **FT_EFF_MC_stat_nominal** has been shown.

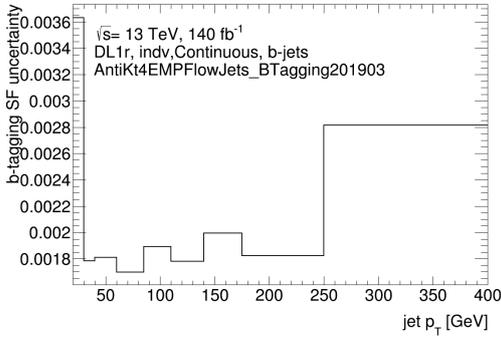
Scale factors for DL1r EMPFlow jets



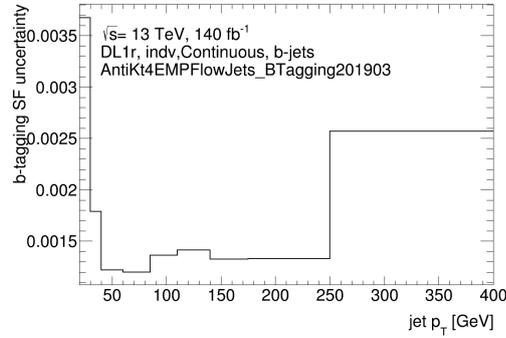
(a) tagweight < FixedCutBEff_85



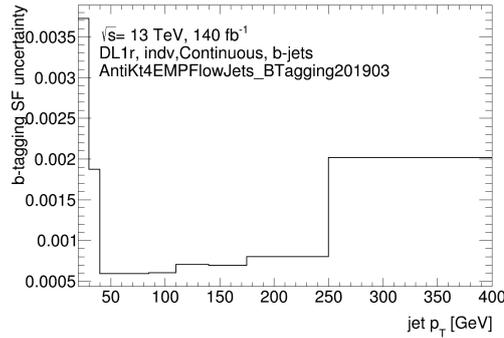
(b) FixedCutBEff_85 < tagweight < FixedCutBEff_77



(c) FixedCutBEff_77 < tagweight < FixedCutBEff_70



(d) FixedCutBEff_70 < tagweight < FixedCutBEff_60



(e) FixedCutBEff_60 < tagweight

Figure 3.16. Total systematic uncertainties for b -jets for individual systematics. The variation of FT_EFF_MC_stat_nominal has been shown.

3.9. Summary

The b -jet identification strategy integrates the results of low-level algorithms (IP2D, IP3D, SV1, JetFitter) with high-level multivariate classifiers (MV2, DL1, DL1r). The transition from cumulative to pseudo-continuous calibration exploits tagweight bin information, resulting in a significant improvement in the precision of the b -jet tagging efficiency measurement for the ATLAS experiment. Simulation-to-data scale factors are calculated by comparing the efficiency measured in the collision data with that observed in the simulations.

This chapter describes the procedure for using the tagweight distribution of b -tagging algorithms, known as continuous b -tagging. The study shows that even coarsely binned tagweight distributions are sufficient to achieve most of the expected improvements. Calibrations for continuous b -tagging are presented, as well as the procedures for accounting for statistical and systematic uncertainties.

A tool has been developed to plot the contents of pseudo-continuous calibrations directly from the CDI file. This tool serves as a basis for investigating pseudo-continuous scale factors for EV decomposition and smoothing of pseudo-continuous calibrations. The plotting scripts can automatically generate eigenvector plots corresponding to each tagweight bin and systematic uncertainty for all possible combinations of jet collections, flavours and taggers. This process is automated as part of the git validation procedure.

Part II

Proton-Proton analysis

Chapter 4

Analysis of $t\bar{t}H$ and $t\bar{t}W$ production in multilepton final states with the ATLAS detector at 13 TeV

In 2012, the Large Hadron Collider (LHC) at CERN reached a groundbreaking milestone in particle physics with the discovery of the Higgs boson. This particle long theorised but never observed, is crucial to understanding how particles acquire mass. The discovery not only confirmed a key component of the Standard Model but also opened up new avenues for exploring the fundamental nature of the universe. This monumental achievement is the culmination of decades of theoretical and experimental work and marks a new era in the quest to unravel the deepest mysteries of matter and energy.

The chapter describes the search for $t\bar{t}H$ production in multilepton final states, outlines the general technique of multilepton analysis for $t\bar{t}H$ and provides a detailed description of the estimation of fake leptons in the $2\ell\text{SS}1\tau_{\text{had}}$ (containing two light leptons with same-sign electric charges and a hadronically decaying tau lepton) final state. The analysis is based on proton-proton collision data collected by the ATLAS detector from 2015 to 2017 at a centre-of-mass energy of $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 80 fb^{-1} .

4.1. Introduction

The Yukawa coupling between the Higgs boson and the top quark is one of the fundamental parameters of the Standard Model (SM). This coupling can be inferred from the ratio of the top quark mass to the vacuum expectation value of the Higgs field. Alternatively, it can be inferred from the cross-sections of certain processes: the production of the Higgs boson associated with a pair of top quarks, $gg/qq \rightarrow t\bar{t}H$, which occurs at tree levels and represents the lowest order in perturbation theory.

The ATLAS and CMS collaborations have extensively studied the $t\bar{t}H$ production in proton-proton (pp) collisions at the Large Hadron Collider (LHC), using data collected during LHC Run 1 at centre-of-mass energies of 7 and 8 TeV [69–72]. The analyses were tuned to detect decays of the Higgs boson, including $H \rightarrow WW^*$, $\tau\tau$, $b\bar{b}$ and $\gamma\gamma$ [73–77]. Combining the results from ATLAS and CMS during Run 1, the ratio of observed to SM cross-sections, denoted as $\mu_{t\bar{t}H} = \sigma/\sigma_{\text{SM}}$, is determined to be $2.3^{+0.7}_{-0.6}$. This excess over the SM prediction ($\mu_{t\bar{t}H} = 1$) is

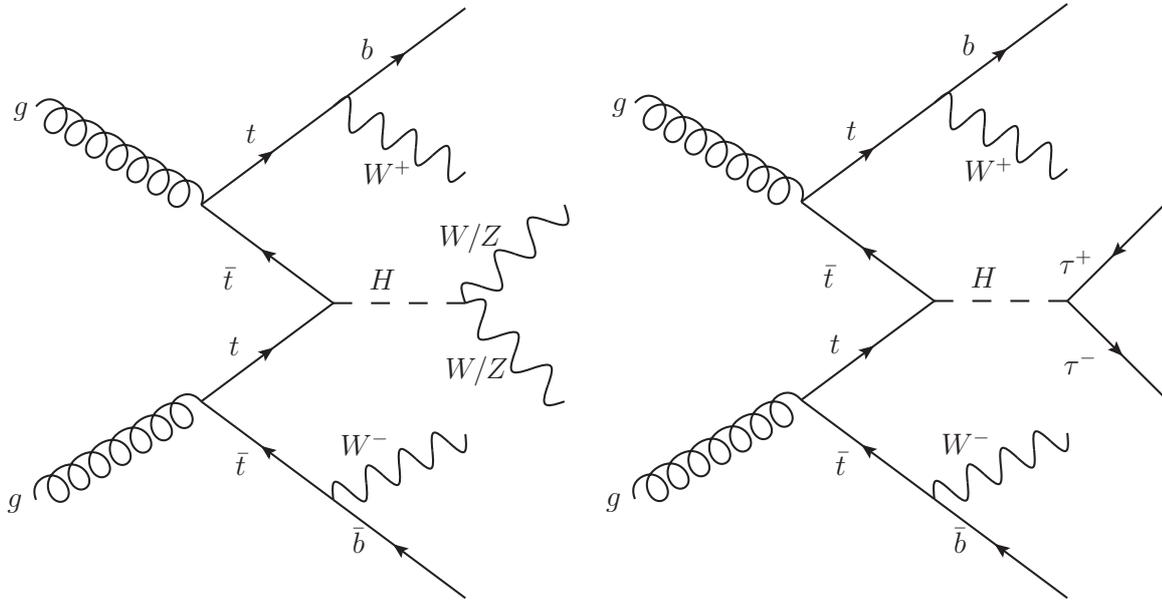


Figure 4.1. Tree-level Feynman diagrams for the production of the Higgs boson in association with a pair of top quarks. Higgs boson decays to WW/ZZ (top) or $\tau\tau$ (bottom) are shown.

mainly attributed to multi-lepton final states, which are particularly sensitive to $H \rightarrow WW^*$ and $\tau\tau$ decays [78]. It's also noteworthy that the cross-section for $t\bar{t}H$ production increased by a factor of 3.9 at a centre-of-mass energy of 13 TeV [79, 80].

ATLAS reported evidence of $t\bar{t}H$ production based on the analysis of 36.1 fb^{-1} of data collected in 2015 and 2016 [81]. The observed (expected) significance of this evidence is 4.2σ (3.8σ). The analysis yielded a best-fit value for the ratio $\mu_{t\bar{t}H}$, representing the $t\bar{t}H$ production rate relative to expectation, as 1.2 ± 0.2 (stat) $^{+0.3}_{-0.2}$ (syst). In addition, the $t\bar{t}H$ production cross-section at a centre-of-mass energy of $\sqrt{s} = 13 \text{ TeV}$ was determined to be $790^{+230}_{-210} \text{ fb}$, in agreement with the SM prediction. Notably, the multilepton channel showed the highest sensitivity, with an observed (expected) significance of 4.1σ (2.8σ).

This chapter presents the results of a search for $t\bar{t}H$ production using 80 fb^{-1} data collected by the ATLAS detector at a centre-of-mass energy of $\sqrt{s} = 13 \text{ TeV}$ between 2015 and 2017. The study delineates six different final states, categorised by the number and type of leptons: two same-charge light leptons (e or μ) with no hadronic decaying τ lepton candidates ($2\ell\text{SS}$); two light leptons of the same charge with one hadronic decaying τ lepton candidate ($2\ell\text{SS}1\tau_{\text{had}}$); three light leptons (3ℓ); four light leptons (4ℓ); three light leptons with one hadronically decaying τ lepton candidate ($3\ell 1\tau_{\text{had}}$); and a light lepton with two hadronic decaying τ lepton candidates ($1\ell 2\tau_{\text{had}}$). Illustrations of the Feynman diagrams of the signal processes are shown in Figure 4.1. These defined signatures predominantly probe $H \rightarrow WW^*$ (with subsequent decay to $\ell\nu\ell\nu$ or $\ell\nu jj$) and $H \rightarrow \tau\tau$ decays. Background estimates use a mixture of simulation and data-driven techniques, and a global fit across all final states is employed to derive the most accurate estimate of the $t\bar{t}H$ production rate.

This chapter is structured as follows: Section 4.2 describes the recorded data and Section 4.3 describes the Monte Carlo sample used. The object definitions are outlined in Section 4.4. Section 4.6 defines the event selection. The categorisation and background overview are summarised in sections 4.7 and 4.8. A dedicated background study for the $2\ell SS1\tau_{\text{had}}$ channel using the ABCD method is presented in Section 4.9. Theoretical and experimental uncertainties are discussed in Section 4.10. Finally, the results for all final states combined are presented in Section 4.11, followed by the conclusions of the analysis in Section 4.12.

4.2. Data samples

A dataset of proton-proton collisions at a centre-of-mass energy of 13 TeV, collected by the ATLAS experiment during 2015–2017 corresponding to an integrated luminosity of $79.9 \pm 1.6 \text{ fb}^{-1}$ is used. The uncertainty in the integrated luminosity [82] is obtained using the LUCID-2 detector [83] for the primary luminosity measurements. The number of additional pp interactions per bunch crossing (pileup) in this dataset varies from approximately 8 to 70 interactions, with an average of 34. Only events recorded under stable beam conditions and for which all detector subsystems were known to be in a satisfactory operational state are included.

4.3. Monte Carlo samples

Monte Carlo simulation samples were generated to represent signal and background processes, utilizing the configurations detailed in Table 4.1, with the samples used to estimate the systematic uncertainties indicated in parentheses. Pileup is modelled using events from minimum-bias interactions generated with PYTHIA 8.186 [84] with the A3 set of tuned parameters [85] and overlaid onto the simulated hard-scatter events according to the luminosity profile of the recorded data. The generated events were processed through a simulation [86] of the ATLAS detector geometry and response using GEANT4 [87], and through the same reconstruction software as the data. Corrections were carefully implemented on the simulated events to ensure consistency with the selection efficiencies, energy scales, and energy resolutions determined from data control samples. Subsequently, the simulated samples underwent normalization to their respective cross-sections, computed within perturbation theory.

4.3.1. $t\bar{t}H$ signal

The nominal sample used to model the $t\bar{t}H$ signal was generated using the next-to-leading-order (NLO) generator POWHEG BOX v2 [88, 89] with the NNPDF3.0 NLO [90] parton distribution function (PDF) set. For setting the renormalization and factorization scales, denoted as μ_R and μ_F respectively, they were established as the geometric mean of the transverse energies of the top quark, the antitop quark, and the Higgs boson.

The parameter h_{damp} within the POWHEG-BOX model, crucial for controlling the matrix element (ME) to parton shower (PS) matching and effectively managing high- p_T radiation, is established at $1.5 \times (2m_t + m_H)/2 = 352.5$ GeV.

The PS and hadronization processes were simulated utilizing PYTHIA 8.2 [91], employing the A14 tune [92], and incorporating Higgs decay branching ratios calculated using HDECAY [93, 94]. The normalization of the simulated sample is performed utilizing a cross-section of 507_{-50}^{+35} fb, computed at the next-to-leading order (NLO) in quantum chromodynamics (QCD) with the leading NLO electroweak corrections, i.e., $\mathcal{O}(\alpha_s^2\alpha^2)$ [95–100]. Uncertainties include $_{-9.2\%}^{+5.8\%}$ estimated by varying the QCD factorisation and renormalisation scales and $\pm 3.6\%$ due to uncertainties on the PDFs and the coupling α_s .

Factors contributing to uncertainties in the modelling of acceptance and event kinematics include variations in QCD factorization and renormalization scales, the choice of PS and hadronization models, representation of initial state radiation (ISR), and uncertainties related to PDFs. The theoretical uncertainties due to the choice of the QCD scale are estimated by independently varying the renormalisation and factorisation scales by a factor of 0.5 and 2.0 with respect to the central value. The largest variation is obtained when both scales are varied simultaneously in the same direction. The uncertainties arising from the selection of PS and hadronization models are assessed by comparing the nominal prediction with an alternative sample generated using POWHEG-BOX interfaced to HERWIG7 [101].

The uncertainty related to the modelling of ISR is determined by examining the Var3c A14 tune variation, which entails a variation of α_s within the A14 tune. The uncertainty associated with the choice of PDF set is evaluated utilizing the PDF4LHC15 prescription [102], incorporating 30 eigenvector shifts derived from fits to multiple Next-to-Next-to-Leading Order (NNLO) PDF sets. Additionally, uncertainties stemming from the predicted Higgs-boson branching ratios [93] are also taken into consideration.

4.3.2. $t\bar{t}W$ background

The simulated dataset for $t\bar{t}W$ production was generated using the SHERPA 2.2.1 [110] generator, incorporating the NNPDF3.0 NLO PDF set. ME calculations were performed for up to one additional parton at NLO and up to two partons at LO using COMIX [121] and OPENLOOPS [122]. These were then merged with the SHERPA PS [123] utilizing the MEPS@NLO prescription [124], with a merging scale set to 30 GeV. The choice of renormalization and factorization scales is $\mu_R = \mu_F = H_T^1/2$.

The nominal cross-section for the simulated $t\bar{t}W$ sample is 601 ± 76 fb, calculated at NLO in QCD with leading NLO electroweak corrections (i.e., $\mathcal{O}(\alpha_s^2\alpha^2)$) [93, 125, 126]. The uncertainties arising from the QCD scale and PDF+ α_s variations are $\pm 12\%$ and $\pm 4\%$, respectively. This cross-section value was employed in a prior analysis [81].

¹ H_T represents the scalar sum of the transverse masses $\sqrt{p_T^2 + m^2}$ of all final state particles.

Process	Generator (alternative)	PS (alternative)	PDF (alternative)	Tune (alternative)
$t\bar{t}H$	POWHEG-BOX [103, 104] (POWHEG-BOX)	PYTHIA 8 (HERWIG7)	NNPDF 3.0 NLO [90]/ NNPDF 2.3 LO [105] (NNPDF3.0 NLO/MMHT2014 LO)	A14 (H7-UE-MMHT)
$tHqb$	MG5_AMC	PYTHIA 8	CT10 [106]	A14
tHW	MG5_AMC	HERWIG++	CT10/CTEQ6L1 [107, 108]	UE-EE-5 [109]
$t\bar{t}W$	SHERPA 2.2.1 [110] (MG5_AMC)	SHERPA (PYTHIA 8)	NNPDF 3.0 NNLO (NNPDF3.0 NLO/NNPDF2.3 LO)	SHERPA default (A14)
$t\bar{t}(Z/\gamma^*)$	MG5_AMC (SHERPA 2.2.0)	PYTHIA 8 (SHERPA)	NNPDF 3.0 NLO/NNPDF2.3 LO (NNPDF3.0 NLO)	A14 (Sherpa default)
$t\bar{t} \rightarrow W^+bW^-\bar{b}l^+l^-$	MG5_AMC	PYTHIA 8	NNPDF3.0 LO	A14
tZ	MG5_AMC	PYTHIA 6	CTEQ6L1	Perugia2012 [111]
tWZ	MG5_AMC	PYTHIA 8	NNPDF 2.3 LO	A14
$t\bar{t}t, t\bar{t}\bar{t}$	MG5_AMC	PYTHIA 8	NNPDF 2.3 LO	A14
$t\bar{t}W^+W^-$	MG5_AMC	PYTHIA 8	NNPDF 2.3 LO	A14
$t\bar{t}$	POWHEG-BOX	PYTHIA 8	NNPDF3.0 NLO/NNPDF2.3 LO	A14
Single top	POWHEG-BOX [112, 113]	PYTHIA 8	NNPDF3.0 NLO/NNPDF2.3 LO	A14
$VV, qqVV, VVV$	SHERPA 2.2.2 [110]	SHERPA	NNPDF 3.0 NNLO	SHERPA default
$Z \rightarrow \ell^+\ell^-$	SHERPA 2.2.1	SHERPA	NNPDF 3.0 NLO	SHERPA default

Table 4.1. The configurations used for event generation of signal and background processes. “V” refers to the production of an electroweak boson (W or Z/γ^*). “Tune” refers to the underlying event tune of the PS generator. “MG5_AMC” refers to MADGRAPH5_AMC@NLO 2.2.1 [114]; “PYTHIA 6” refers to version 6.427 [115]; “PYTHIA 8” refers to version 8.2 [116]; “HERWIG++” refers to version 2.7 [117]. Samples using PYTHIA 6 or PYTHIA 8 have heavy-flavour hadron decays modelled by EVTGEN 1.2.0 [118]. All samples include leading-logarithm photon emission, either modelled by the PS generator or by PHOTOS [119] [120].

In this study, additional scaling factors are applied to the $t\bar{t}W$ cross-section to incorporate missing QCD and electroweak corrections. Earlier investigations [114] have demonstrated significant NLO QCD corrections to $t\bar{t}W+1$ -jet production. An inclusive scaling factor of 1.11 has been determined using dedicated samples generated with SHERPA 2.2.5, employing the MEPS@NLO prescription, and cross-validated with the NLO generator MADGRAPH5_AMC@NLO 2.2.1 using the FxFX prescription [127].

Furthermore, recent findings [128] indicate that sub-leading NLO electroweak corrections for $t\bar{t}W$ production exceed expectations, primarily due to the sizable NLO₃ term influenced by $t\bar{t}W+1$ -jet diagrams featuring a Higgs boson exchanged in the t -channel. The corresponding estimated scaling factor is 1.09. Consequently, upon applying these two scaling factors, the inclusive cross-section used to normalize the $t\bar{t}W$ sample becomes 727 ± 92 fb.² Since the $t\bar{t}W$ normalization will be determined from data (as discussed in Section 4.11), this uncertainty is not treated as a systematic uncertainty. This cross-section is henceforth referred to as the “updated $t\bar{t}W$ theoretical cross-section”.

Systematic uncertainties arising from the absence of higher-order QCD corrections are estimated by simultaneously adjusting the factorization and renormalization scales in the nominal sample by factors of 0.5 and 2.0 relative to the central value. Uncertainties associated with the modelling of additional QCD radiation are assessed by comparing the nominal $t\bar{t}W$ prediction with an alternative sample generated at NLO using the MADGRAPH5_AMC@NLO 2.2.1 generator. This alternative sample, generated with the same scale choice and PDF set as the nominal sample, was interfaced to PYTHIA 8.2 combined with the A14 tune. This alternative sample served as the nominal sample in a previous analysis [81]. Finally, the uncertainty stemming from the choice of PDF set is evaluated using the PDF4LHC15 prescription.

4.3.3. Other backgrounds

The simulated samples for $t\bar{t}(Z/\gamma^*)$, VV and $t\bar{t}$ production are based on the methods described in Ref. [129–131].

For the $t\bar{t}(Z/\gamma^*)$ sample, the inclusive $t\bar{t}l^+l^-$ ME is computed at the NLO, including the off-shell Z and γ^* contributions with $m(l^+l^-) > 1$ GeV. In addition, a specialised $t\bar{t}$ sample includes rare $t \rightarrow Wb\gamma^*(\rightarrow l^+l^-)$ radiative decays, $t\bar{t} \rightarrow W^+bW^-\bar{b}l^+l^-$, is generated using a leading order (LO) ME, which also requires $m(l^+l^-) > 1$ GeV. In this example, the photon can come from the top quark, the W boson or the b quark. Both the $t\bar{t}(Z/\gamma^*)$ and the $t\bar{t} \rightarrow W^+bW^-\bar{b}l^+l^-$ samples are combined to form the $t\bar{t}(Z/\gamma^*)$ (high mass) sample.

The simulation of internal photon conversions ($\gamma^* \rightarrow l^+l^-$) with $m(l^+l^-) < 1$ GeV is performed using QED multiphoton radiation through the PS in an inclusive $t\bar{t}$ sample, referred to as “ $t\bar{t}\gamma^*$ (low mass)”. The generation settings for these samples are detailed in Table 4.1. To ensure accurate results, steps have been taken to avoid double counting and to cover the entire

²The theoretical uncertainties are not adjusted based on the additional corrections considered but are scaled proportionally to the applied scaling factors.

phase space when combining different simulated samples. The cross-section for the inclusive $t\bar{t}l^+l^-$ production, with $m(l^+l^-) > 1$ GeV, is 162 ± 21 fb, calculated at NLO for both QCD and electroweak interactions [93, 114, 126].

The uncertainties due to variations in the QCD scale and PDF+ α_s are $\pm 12\%$ and $\pm 4\%$ respectively. To address the differences between NNLO+ Next-to-Leading (NLL) [132–136] and LO cross-sections for $t\bar{t}$ production, the LO cross-section from $t\bar{t} \rightarrow W^+bW^-\bar{b}l^+l^-$ sample is multiplied by a factor of 1.54 and subjected to a normalisation uncertainty of 50%.

Uncertainties affecting the modelling of the acceptance and event kinematics for the $t\bar{t}l^+l^-$ sample include variations in the QCD scale and tune, PDF variations using the PDF4LHC15 prescription, and comparison with an alternative LO multileg sample (see Table 4.1). For the inclusive $t\bar{t}$ sample, uncertainties in the modelling of additional QCD radiation are addressed by two alternative samples generated with settings that adjust the amount of radiation [137].

The normalisation of the diboson backgrounds is based on cross-section calculations performed by SHERPA 2.2.2, with an assigned normalisation uncertainty of 50%, treated independently for different subprocesses (WZ +light jets, $WZ+\geq 1c$, $WZ+\geq 1b$, and ZZ +jets). Rare background contributions (tZ , $t\bar{t}t$, $t\bar{t}W$, WtZ , VVV , $t\bar{t}t$, $tHjb$, and WtH) are scaled according to their NLO theoretical cross-sections, with a 50% uncertainty in the normalisation.

4.4. Object selection

This section outlines the selection criteria applied to the reconstructed physics objects. These criteria ensure that the physics objects meet certain kinematic standards, satisfy trigger requirements and potentially increase the sensitivity of the analysis. The discussion includes the working points (WPs) for lepton reconstruction and identification criteria, isolation requirements and kinematic criteria for both leptons and jets. In addition, the section deals with the reconstruction and identification of the hadronically decayed τ leptons. Finally, the procedure for removing the overlap is described in detail. Tracks are reconstructed inside the inner detector. Interaction vertices resulting from the pp collisions are reconstructed from at least two tracks with p_T above 400 MeV, which are consistent with originating from the beam collision region in the x - y plane. In scenarios where multiple primary vertex candidates are detected, the candidate whose associated tracks yield the highest sum of squared p_T [138] is selected as the hard scatter primary vertex.

4.4.1. light leptons

To select leptons from the primary vertex, requirements on the longitudinal (z_0) and transverse (d_0) impact parameters are applied. For both electrons and muons, the requirement is $|z_0 \sin \theta| < 0.5$ mm. Additionally, the transverse impact parameter significance must be $|d_0|/\sigma_{d_0} < 5$ for electrons and $|d_0|/\sigma_{d_0} < 3$ for muons.

Muon candidates [139] are reconstructed by correlating track segments across different layers of the muon spectrometer with tracks identified in the inner detector. These muon candidates

then undergo a refitting process using the comprehensive track information from both detector systems. They must satisfy $p_T > 10$ GeV and $|\eta| < 2.5$, while also meeting loose identification requirements [139].

Electrons are reconstructed by associating energy clusters in the electromagnetic calorimeter with inner detector tracks [140]. They must have a $p_T > 10$ GeV and $|\eta_{\text{cluster}}| < 2.47$, excluding the transition region between the barrel and endcap calorimeters, $1.37 < |\eta_{\text{cluster}}| < 1.52$.

4.4.2. Jets and b -tagged jets

In the calorimeters [141, 142], jets are reconstructed from clusters formed by energy deposition using the anti- k_t algorithm with a radius parameter $R = 0.4$ [143, 144]. The jets are calibrated using simulations with adjustments from in situ³ methods [145]. Both $|\eta| < 2.5$ and $p_T > 25$ GeV must be satisfied by the jets. To eliminate jets associated with pileup vertices with $p_T < 60$ GeV and $|\eta| < 2.4$ [146], a Jet Vertex Tagger (JVT) is used. Depending on the jet p_T , the total uncertainty of the jet energy scale (JES) can range from 1% to 2%. The uncertainty of the JES is decomposed into uncorrelated components [147, 148]. The JES takes into account variations in the fraction of jets containing quarks and gluons in different physics processes.

Using a multivariate discriminant (MV2c10) combining data from track impact parameters and secondary vertices [149–151], jets containing b -hadrons are identified as “ b -tagged”. A working point is selected with an average efficiency in $t\bar{t}$ events of 70% for b quark jets and rejection factors of 380, 12 and 55 respectively against light quark/gluon jets, c quark jets and hadronically decaying τ leptons.

The uncertainties in the b -tagging efficiencies [149] are decomposed into uncorrelated components. For b -jets, the relative uncertainty of the b -tagging efficiency is about 2%; for c -jets and τ -leptons, it is about 10%; and for light jets.

4.4.3. Hadronically decaying tau leptons

Clusters in the calorimeters and associated inner detector tracks are used to reconstruct hadronically decaying τ lepton candidates (τ_{had}) [152]. These candidates must have a total charge of ± 1 for one or three related tracks. They must also have $|\eta| < 2.5$ and $p_T > 25$ GeV, excluding the electromagnetic calorimeter transition zone. Also, the primary vertex is where τ_{had} candidates must come from. The τ_{had} candidates are identified and the jet backgrounds are discriminated using a Boosted Decision Tree (BDT) discriminant that includes variables from both the calorimeter and the tracking systems [153].

The medium working point is set to achieve an efficiency of 55% (40%) for one (three) prong τ_{had} decays, while the *tight* working point aims for an efficiency of 40% (30%) for one (three) prong τ_{had} decays. Electrons that are reconstructed as one-prong are removed using a BDT with an efficiency (rejection factor) of 95% (30–100%) for true (false) τ_{had} candidates, with its

³ In situ techniques exploit the transverse momentum (p_T) balance between a jet and a well-measured reference object

effectiveness varying according to their p_T . The uncertainty associated with the identification efficiency for τ_{had} candidates is about 6% [153].

4.4.4. Missing transverse momentum

The missing \vec{p}_T^{miss} , denoted as E_T^{miss} can be defined as the negative vector sum of the p_T of all selected and calibrated objects within the event. It is further enhanced by an additional term that includes momentum contributions from soft particles that are not associated with any of the selected objects [154]. This soft term is computed from inner detector tracks matched to the selected primary vertex.

4.4.5. Overlap removal

To distinguish between electrons and jets, both loose and *tight* electron identification criteria are applied [155], using a likelihood discriminant that incorporates calorimeter, track, and combined variables. For *tight* electrons, further requirements are imposed on the associated track p_T and the ratio of electron calorimeter energy to track momentum to better suppress material conversion.

Muons must be separated from any selected jets by $\Delta R > \min(0.4, 0.04 + (10 \text{ GeV})/p_{T,\mu})$. If two electrons are within $\Delta R = 0.1$ of each other, only the electron with the higher p_T is considered. Moreover, any electron within $\Delta R = 0.1$ of a muon is excluded.

Jets within $\Delta R = 0.3$ of a selected electron or a τ lepton candidate that decays hadronically are excluded. Furthermore, τ_{had} candidates must be separated by $\Delta R > 0.2$ from any selected electron or muon candidate. Any τ_{had} candidate that is also tagged as a b -jet is ignored.

Keep	Remove	Cone size (ΔR)
electron	electron (low p_T)	0.1
muon	electron	0.1
electron	jet	0.3
jet	muon	$\min(0.4, 0.04 + 10[\text{GeV}]/p_T \text{ (muon)})$
electron	tau	0.2
muon	tau	0.2
tau	jet	0.3

Table 4.2. Summary of the overlap removal procedure between electrons, muons, hadronically decaying taus, and jets.

4.5. Tight light leptons definition

To further reduce the contribution of fake light leptons, a tighter selection criteria is applied. This selection relies on isolation parameters and the results of two multivariate algorithms. The specifics of this stringent selection process are elaborated below.

4.5.1. Isolation variables

The selection includes isolation variables to ensure adequate separation of light leptons from other particles. Isolation criteria are set based on the energy deposited in the detector environment relative to the energy of the lepton. Both calorimetric and track-based isolation criteria are used for electrons and muons, achieving 99% efficiency for $Z \rightarrow \ell^+\ell^-$ events. Calorimetric isolation involves the summation of the transverse energies of the clusters within $\Delta R = 0.3$ around the light-lepton candidate, excluding the cluster of the electron candidate. Track isolation involves the summation of the transverse momenta of tracks with $p_T > 1$ GeV emanating from the primary vertex, but excluding the light lepton candidate track, within $\Delta R = \min(0.3, 10 \text{ GeV}/p_T(\ell))$.

4.5.2. Non-prompt leptons

The BDT discriminator, constructed from isolation and b -tagging variables and referred to as the non-prompt lepton BDT [81], is used to identify non-prompt leptons produced at some displacement from the interaction point. At the designated operating point, muons (electrons) satisfying the calorimeter- and track-based isolation criteria show an efficiency of about 80% (65%) at $p_T \sim 20$ GeV, rising to a steady 95% (90%) at $p_T \sim 45$ GeV.

4.5.3. Electron charge misidentification BDT

The removal of *tight* electrons with incorrect charge assignments is achieved by a BDT discriminator that relies on calorimeter and tracking characteristics [140]. This procedure achieves a 95% efficiency for electrons with correct charge assignments and a rejection factor of approximately 17 for electrons with incorrect charge assignments.

4.5.4. Classification

The *tight* electron candidates are categorized into three groups: material conversions, internal conversions, and very *tight*.

- Material conversion candidates exhibit a displaced vertex with a radius $r > 20$ mm, encompassing the electron-associated track. The invariant mass of the associated track and the nearest opposite-charge track, reconstructed in the silicon detector at the conversion vertex, must be less than 100 MeV
- Internal conversion candidates do not meet the material conversion criteria. The invariant mass of the di-track system, calculated at the primary vertex, must also be less than 100 MeV.
- Conversely, very *tight* electron candidates are those that fail both material and internal conversion criteria, with $|\eta| < 2$. This criterion excludes a small fraction of electrons with a high rate of charge misidentification due to the limited number of hits used in track reconstruction

	e				μ		
	L	L*	T	T*	L	L*	T/T*
Identification	Loose		Tight		Loose		Medium
Isolation	No	Yes			No	Yes	
Non-prompt lepton veto	No		Yes		No		Yes
Charge misidentification veto	No		Yes		N/A		
Material/internal conversion veto	No		Yes		N/A		
Lepton $ \eta $	< 2.47		< 2		< 2.5		
$ d_0 /\sigma_{d_0}$	< 5				< 3		
$ z_0 \sin \theta $	< 0.5 mm						

Table 4.3. The requirements applied to select loose (L), loose and minimally-isolated (L*), *tight* (T) and very *tight* (T*) light leptons [120].

Table 4.3 outlines the different types of light leptons utilized in the analysis. Although uncertainties in their identification, isolation, reconstruction, and trigger efficiencies are considered, their overall effect on the analysis remains minimal. Loose and tight electron identification operating points are used as described in Ref. [71].

4.6. Event selection

Six final states (channels) shown in Figure 4.2, are analysed and classified according to the number and flavour lepton candidates. Some channels are further subdivided to increase the significance of the analysis. The selection criteria are orthogonal, meaning that each event contributes to only one channel. The six channels are described in Table 4.4.

The analysis uses a dilepton trigger or a single lepton trigger to select events. The dielectron and dimuon triggers are designed for channels with two or more light leptons. The p_T threshold for dielectron triggers is set at 24 GeV. For dimuon triggers, the leading muon has a p_T threshold of 22 GeV and the sub-leading muon has a threshold of 8 GeV. The electron+muon trigger thresholds have remained consistent across all datasets, with electrons at 17 GeV and muons at 14 GeV.

Table 4.5 summarises the selection criteria for each channel, and Table 4.6 lists the strategies for the six analysis channels. Multivariate techniques are used to separate the $t\bar{t}H$ signal from the background in the $2\ell SS$, 3ℓ and $1\ell 2\tau_{\text{had}}$ channels, with additional selection criteria used in the 4ℓ channel. No additional selection is used in the $2\ell SS 1\tau_{\text{had}}$ and $3\ell 1\tau_{\text{had}}$ channels. BDTs trained with the TMVA package are used in the $2\ell SS$ and $1\ell 2\tau_{\text{had}}$ channels, while XGBoost is used in the 3ℓ channel.

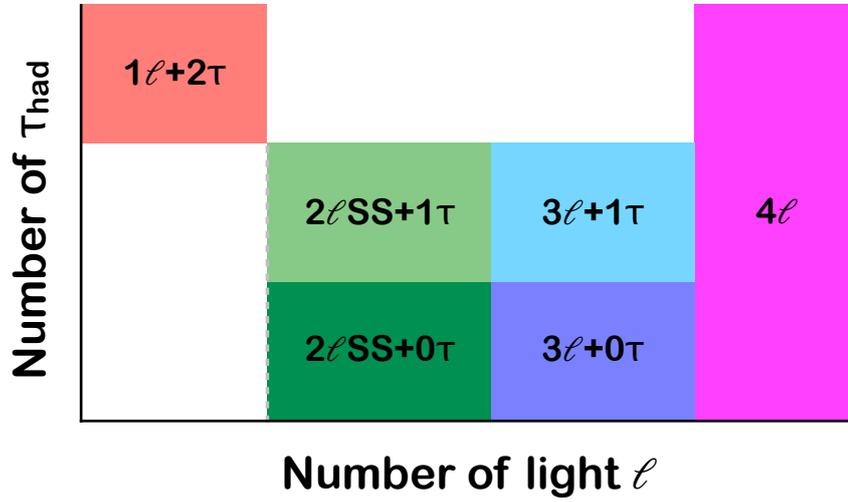


Figure 4.2. $t\bar{t}H$ multilepton channels in terms of light lepton multiplicity and τ_{had} multiplicity.

Category	Description
$2\ell\text{SS}$	two same-charge light leptons and no hadronically decaying τ -lepton candidates
3ℓ	three light leptons and no hadronically decaying τ -lepton candidates
4ℓ	four light leptons
$1\ell 2\tau_{\text{had}}$	one light lepton and two opposite-charge hadronically decaying τ -lepton candidates
$2\ell\text{SS}1\tau_{\text{had}}$	two same-charge light leptons and one hadronically decaying τ -lepton candidate
$3\ell 1\tau_{\text{had}}$	three light leptons and one hadronically decaying τ -lepton candidate

Table 4.4. Lepton categories and their descriptions.

Channel	Selection criteria
Common	$N_{\text{jets}} \geq 2$ and $N_{b\text{-jets}} \geq 1$
2ℓSS	Two same-charge (SS) very <i>tight</i> (T*) leptons, $p_T > 20$ GeV No τ_{had} candidates $m(\ell^\pm\ell^\pm) > 12$ GeV 13 categories: enriched with $t\bar{t}H$, $t\bar{t}W$, $t\bar{t}$, mat. conv, int. conv., split by lepton flavour, charge, jet and b -jet multiplicity
3ℓ	Three loose (L) leptons with $p_T > 10$ GeV; sum of light-lepton charges = ± 1 Two SS very <i>tight</i> (T*) leptons, $p_T > 15$ GeV One OS (w.r.t the SS pair) loose-isolated (L*) lepton, $p_T > 10$ GeV No τ_{had} candidates $m(\ell^+\ell^-) > 12$ GeV and $ m(\ell^+\ell^-) - 91.2$ GeV > 10 GeV for all SFOS pairs $ m(3\ell) - 91.2$ GeV > 10 GeV 7 categories: enriched with $t\bar{t}H$, $t\bar{t}W$, $t\bar{t}Z$, VV , $t\bar{t}$, mat. conv, int. conv
4ℓ	Four loose-isolated (L*) leptons; sum of light lepton charges = 0 $m(\ell^+\ell^-) > 12$ GeV and $ m(\ell^+\ell^-) - 91.2$ GeV > 10 GeV for all SFOS pairs $m(4\ell) < 115$ GeV or $m(4\ell) > 130$ GeV 2 categories: Z _{enr} (Z -enriched; 1 or 2 SFOS pairs) or Z _{dep} (Z -depleted; 0 SFOS pairs)
1ℓ2 τ_{had}	One <i>tight</i> (T) lepton, $p_T > 27$ GeV Two OS τ_{had} candidates At least one <i>tight</i> τ_{had} candidate $N_{\text{jets}} \geq 3$
2ℓSS1 τ_{had}	2ℓSS selection, except: One medium τ_{had} candidate $N_{\text{jets}} \geq 4$
3ℓ1 τ_{had}	3ℓ selection, except: One medium τ_{had} candidate, of opposite charge to the total charge of the light leptons Two SS <i>tight</i> (T) leptons

Table 4.5. Selection criteria applied to the channels. The common selection criteria for all channels are listed in the first line under the title “Common”. Same-charge (opposite-charge) lepton pairs are also referred to as same-sign (opposite-sign) with the abbreviation SS (OS). Same-flavour (SF), OS lepton pairs are referred to as SFOS pairs. In the categories for conversions, the selection requirements on one of the leptons are loosened as discussed in Section 4.4 [120].

	Non-tau channels			Tau channels		
	$2\ell SS$	3ℓ	4ℓ	$1\ell + 2\tau_{had}$	$2\ell SS + 1\tau_{had}$	$3\ell + 1\tau_{had}$
Light lepton	$2 T^{**}$	$1 L^*, 2 T^*$	$4 L^*$	$1 L^*$	$2 T^{**}$	$1 L^*, 2 T$
τ_{had}	0M	0M	-	$\geq 1 T$	1M	1M
N_{jets}, N_{b-jets}	$\geq 2, \geq 1$	$\geq 2, \geq 1$	$\geq 2, \geq 1$	$\geq 3, \geq 1$	$\geq 4, \geq 1$	$\geq 2, \geq 1$
Non-prompt lepton strategy	semi-DD (TF)	semi-DD (TF)	semi-DD (SF)	MC	semi-DD (TF)	MC
Fake tau strategy	-	-	-	DD (SS data)	semi-DD (SF)	semi-DD (SF)
BDT trained against Discriminant in SRS	Fakes and $t\bar{t}V$ 2D BDT	$t\bar{t}, t\bar{t}W, t\bar{t}Z, VV$ 5D BDT	$t\bar{t}Z/-$ Event count	$t\bar{t}$ BDT	-	-
Number of bins in SRS	3/3	3	1/1	3	Event count	Event count
Number of CRS	11	6	-	-	1	1

Table 4.6. Summary of basic characteristics and strategies of the six analysis channels. In lepton selection, T stands for *tight* lepton definition as defined in Tables 6.1, L stands for Loose lepton definition, and L^* stands for Loose lepton definition with an additional requirement to pass “FixedCutLoose” isolation. For the fake lepton and τ_{had} background estimations, DD means data-driven, from which TF is the template fit method and SF refers to the fake scale factor method [120].

4.7. Categorisation

The $2\ell\text{SS}$ channel defines five event categories for background determination. Two require an electron from internal or material conversion. The other three, called “low jet multiplicity (LJ)”, require two or three jets and are grouped by the flavour of the sub-leading lepton, with further subdivisions based on the number of b -jets.

Two BDTs are trained on events with four or more jets to separate the signal from the non-prompt background and $t\bar{t}W$ events. These are divided into categories enriched in signal, $t\bar{t}W$ and $t\bar{t}$ events, further subdivided by lepton charge, resulting in six additional background categories and two signal categories, known as “high jet multiplicity (HJ)” categories. In total, there are 13 categories in the $2\ell\text{SS}$ channel.

The 3ℓ channel uses a five-dimensional multinomial BDT targeting five classifications: $t\bar{t}H$, $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}$ and diboson, resulting in five categories. Two additional categories are defined for transformations by relaxing the electron requirements, giving a total of seven categories.

Cross-check analysis using cut-and-count categorisation is developed for the $2\ell\text{SS}$ and 3ℓ channels, following the same event selection and defining categories based on variables such as jet multiplicity and lepton flavour.

In the $1\ell 2\tau_{\text{had}}$ channel, three categories are defined using a BDT to distinguish signal from background.

A total of 877 events are selected in 25 categories with signal-to-background ratios ranging from 0.3% to 104% as shown in Figure 4.3. The total expected number of reconstructed signal events is 170, or 0.42% of all events produced. In the $2\ell\text{SS}1\tau_{\text{had}}$ and $3\ell 1\tau_{\text{had}}$ channels, about 20% of the selected events are non-prompt leptons or fake τ_{had} candidates, treated as signal and corrected with specific normalisation factors. The full details can be found in Ref. [120].

4.8. Background overview and estimation

Backgrounds are divided into irreducible and reducible types. Irreducible backgrounds include all prompt leptons produced in W/Z boson decays, leptonic τ decays or internal conversions. Reducible backgrounds include prompt leptons with misassigned charge (QMisID), non-prompt light leptons, or jets misidentified as τ_{had} candidates. QMisID and fake τ_{had} backgrounds are estimated using data-driven techniques, while others are estimated using simulations.

Reducible backgrounds contain at least one charge-flip electron or a non-prompt lepton (hereafter referred to as a fake lepton) or a fake hadronic tau, depending on the channel. These reducible backgrounds originate mainly from the $t\bar{t}$ production in the HF decay. The procedure for estimating the light lepton fakes is the same for the three relevant channels - $2\ell\text{SS}$, $2\ell\text{SS} + 1\tau_{\text{had}}$, and 3ℓ , and involves a template fit. A similar strategy is used for the 4ℓ channel.

To capture the $0\tau_{\text{had}} \rightarrow 1\tau_{\text{had}}$ extrapolation, the $2\ell\text{SS} + 1\tau_{\text{had}}$ channel also uses the alternative fake factor ABCD method for systematic comparison.

The estimate for fake τ_{had} background events is estimated in the $2\ell\text{OS} + 1\tau_{\text{had}}$ channel and used in the $3\ell + 1\tau_{\text{had}}$ and $2\ell\text{SS} + 1\tau_{\text{had}}$ analyses. The $1\ell + 2\tau_{\text{had}}$ channel uses a fully data-driven

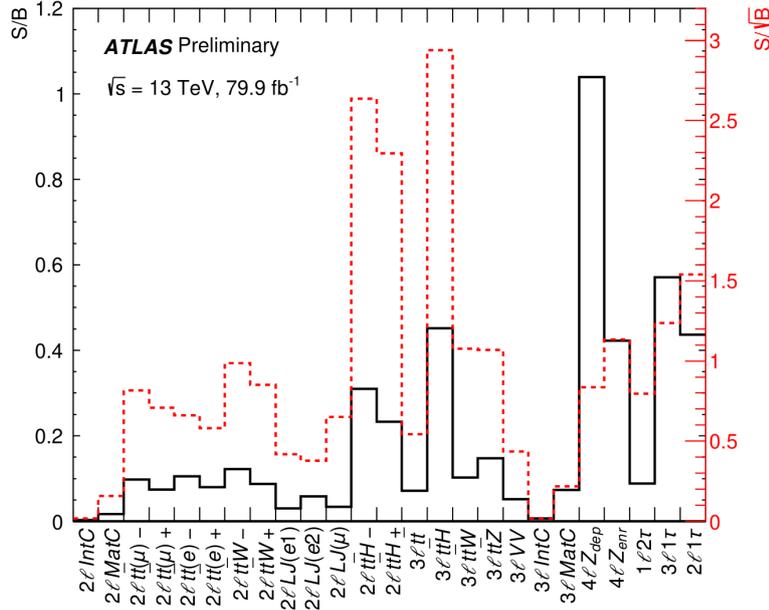


Figure 4.3. Pre-fit S/B (black line) and S/\sqrt{B} (red dashed line) ratios for each analysis category. The background prediction methods are described in Section 4.8 [120].

approach to estimate the fakes. Table 4.6 summarises the non-prompt lepton strategy in each channel.

4.8.1. Irreducible backgrounds

Background contributions with prompt leptons arise from different processes, with the relative importance of each process varying between different channels. The primary irreducible backgrounds arise from $t\bar{t}W$ and $t\bar{t}(Z/\gamma^*)$ production, followed by VV production (in particular WZ), which exhibits final states and kinematic properties similar to those of the $t\bar{t}H$ signal. Smaller contributions come from rare processes, including tZ , tW , WtZ , $t\bar{t}WW$, VVV , $t\bar{t}t$ and $t\bar{t}t\bar{t}$ production.

The associated production of single top quarks with a Higgs boson, tH , accounts for a maximum of 2% in any SR, while other Higgs boson production mechanisms contribute insignificantly $< 0.2\%$ in any SR. They are therefore considered to be background processes and are fixed to the predictions of the Standard Model. Backgrounds with prompt leptons are estimated from simulations using the samples described in Section 4.3.1, which also addresses the systematic uncertainties in modelling these processes.

$t\bar{t}W$ background

The $t\bar{t}W$ background emerges as the dominant background, particularly in the $2\ell\text{SS}$ and 3ℓ channels. Despite the use of advanced simulations, accurate modelling of the additional QCD

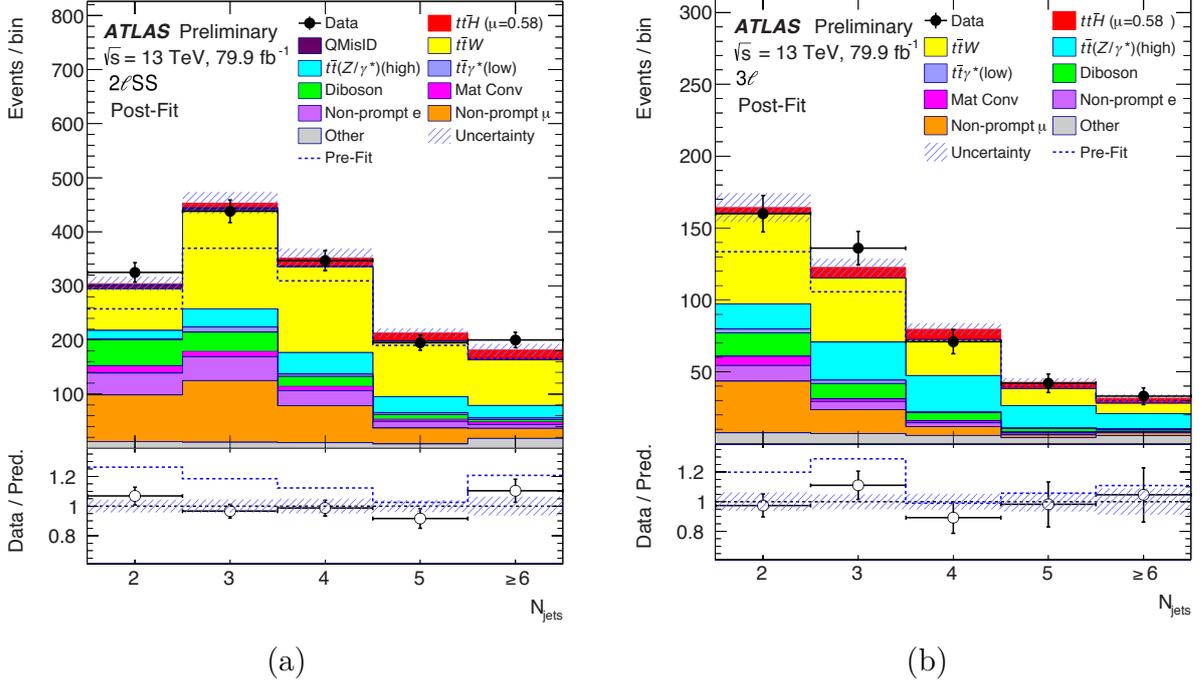


Figure 4.4. Comparison between data and signal-plus-background prediction for the distribution of jet multiplicity in (a) the 2ℓ SS channel and (b) the 3ℓ channel after event selection and before further event categorisation (Section 4.6). The background contributions after the likelihood fit (“Post-Fit”) are shown as filled histograms. The total signal-plus-background prediction before the fit (“Pre-Fit”) is shown as a dashed blue histogram. The $t\bar{t}H$ signal, scaled according to the results of the fit, is shown as a filled red histogram added to the post-fit background. The size of the combined statistical and systematic uncertainty in the signal-plus-background prediction is indicated by the blue-hatched band. The ratios of the data to the total pre- and post-fit predictions are shown in the lower panel. The last bin in each figure contains the overflow [120].

radiation in $t\bar{t}W$ production remains a challenge. To address this, categories sensitive to the $t\bar{t}W$ background are included in the analysis to study and constrain this background.

Figure 4.4 shows the jet multiplicity distributions in the 2ℓ SS and 3ℓ channels after event selection, after a likelihood fit (5.8) to the data, revealing discrepancies between the data and the predictions. To reduce the dependence of the $t\bar{t}H$ signal extraction on the $t\bar{t}W$ prediction, three independent normalisation factors for the $t\bar{t}W$ background are introduced into the likelihood fit, two corresponding to the LJ and HJ categories of the 2ℓ SS channel, and one for the 3ℓ channel categories.

The measured normalisation factors are $\hat{\lambda}_{t\bar{t}W}^{2\ell LJ} = 1.56^{+0.30}_{-0.28}$, $\hat{\lambda}_{t\bar{t}W}^{2\ell HJ} = 1.26^{+0.19}_{-0.18}$, and $\hat{\lambda}_{t\bar{t}W}^{3\ell} = 1.68^{+0.30}_{-0.28}$. After applying background corrections from the likelihood fit, in particular the $t\bar{t}W$ normalisation factors, the agreement with the data is improved (Figure 4.4).

Additional uncertainties associated with the modelling of the b -jet multiplicity and the W boson charge asymmetry in the $t\bar{t}W$ background are introduced to account for the observed discrepancies in the shape of these distributions between the data and pre-fit background predictions

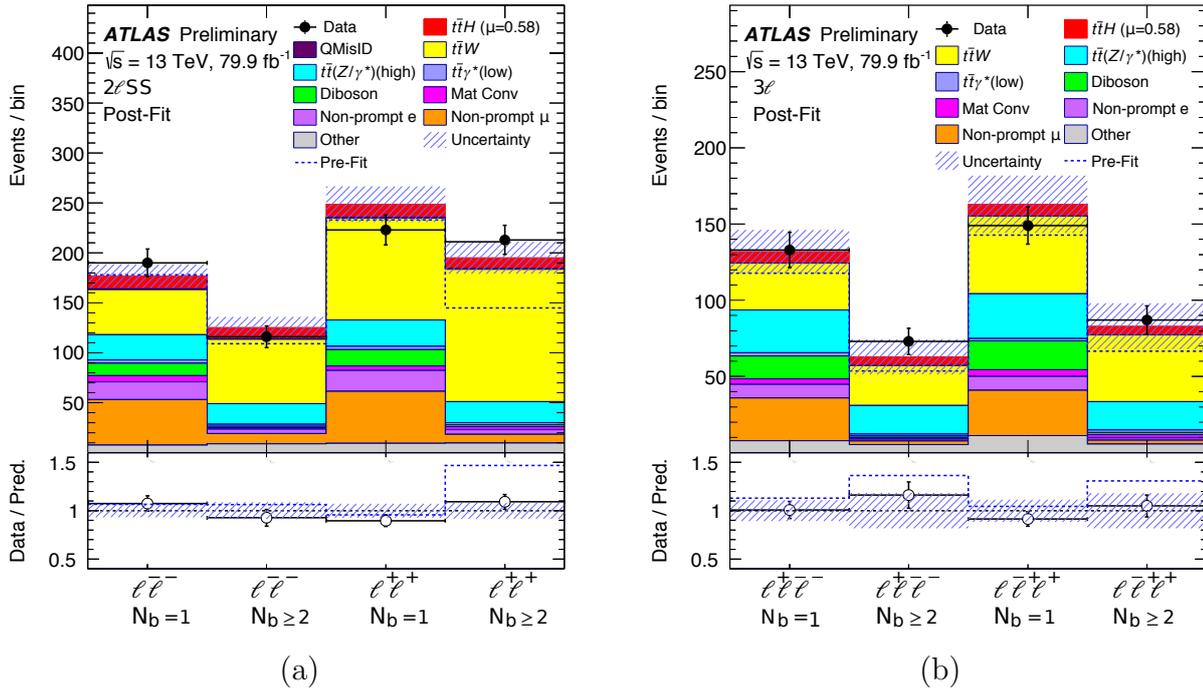


Figure 4.5. Comparison between data and signal-plus-background prediction for the event yields in (a) the 2ℓ SS channel and (b) the 3ℓ channel after event selection and before further event categorisation (Section 4.6), split into four separate categories depending on the total charge and b -jet multiplicity. Additionally, the events in (a) are required to have four or more jets. The background contributions are shown as filled histograms after the likelihood fit (“Post-Fit”). The total signal-plus-background prediction before the fit (“Pre-Fit”) is shown as a dashed blue histogram. The $t\bar{t}H$ signal, scaled according to the fit results, is shown as a filled red histogram added to the post-fit background. The size of the combined statistical and systematic uncertainty in the signal-plus-background prediction is indicated by the blue-hatched band. The ratios of the data to the total pre- and post-fit predictions are shown in the lower panel [120].

in the 2ℓ SS and 3ℓ channels (Figure 4.5). The uncertainty associated with the b jet multiplicity distribution is $\pm 25\%$ ($\mp 35\%$) for events with exactly one (at least two) b jets. Meanwhile, the uncertainty associated with the total charge distribution is $\pm 20\%$ ($\mp 35\%$) for events with positive (negative) total charge. These additional uncertainties, called “extrapolation” uncertainties, are treated as uncorrelated between the 2ℓ SS and 3ℓ channels. In total, 41 uncertainties are included in the statistical model to describe the $t\bar{t}W$ background.

Other irreducible backgrounds

The total yields in the $3\ell t\bar{t}Z$ and $3\ell VV$ control regions play a crucial role in the likelihood fit, improving the accuracy of background estimates from $t\bar{t}(Z/\gamma^*)$ and VV processes. Background estimates from internal conversions with $m(e^+e^-) < 1$ GeV are based on two dedicated control regions (2ℓ IntC and 3ℓ IntC). The total yield of each category is included in the likelihood fit to obtain the normalisation factor $\hat{\lambda}_e^{\text{IntC}} = 0.83 \pm 0.32$, where statistical uncertainty dominates.

The validity of the internal conversion background normalisation is confirmed by a comparison between data and scaled simulation in a validation region enriched in $Z \rightarrow \mu^+\mu^-\gamma^*(\rightarrow e^+e^-)$ candidate events. This region requires two oppositely charged muons and one electron to satisfy the internal conversion criteria. The observed and predicted yields agree to within 25% [120] which is the systematic uncertainty associated with extrapolating the estimate from the internal conversion control regions to the other event categories.

4.8.2. Reducible backgrounds

The reducible backgrounds arise from events in which at least one of the light leptons does not originate from the decays of electroweak bosons, but rather from alternative sources. These sources typically include decays of heavy-flavoured hadrons, photon conversions, and misidentification or misconstruction of particles, such as hadronic jets or prompt light leptons with misidentified charges. Light leptons originating from such sources are subsequently called fake light leptons.

Charge misassignment

Backgrounds involving leptons with incorrectly assigned charges mainly affect the $2\ell SS$ and $2\ell SS1\tau_{\text{had}}$ channels, mainly due to the $t\bar{t}$ production. This occurs when an electron undergoes hard bremsstrahlung or asymmetric conversions ($e^\pm \rightarrow e^\pm\gamma^* \rightarrow e^\pm e^+e^-$), or when a track curvature is mismeasured. The muon charge misassignment rate is negligible in the p_T range relevant to this analysis.

In the data, the electron charge misassignment rate is measured using $Z \rightarrow e^+e^-$ events reconstructed as both same-charge and opposite-charge pairs, with background subtraction via a sideband method. This rate is measured separately for three types of *tight* electrons (internal conversion, material conversion and very *tight*), parameterised by the electron p_T and $|\eta|$. For very *tight* electrons, it ranges from about 10^{-5} for low p_T electrons ($15 \leq p_T \leq 90$) GeV at $|\eta| \leq 1.37$ to about 3×10^{-4} for high p_T electrons ($p_T \geq 90$) GeV at $1.52 \leq |\eta| \leq 2$. The measured charge mismatch rate is then applied to data events meeting the requirements of the $2\ell SS$ and $2\ell SS1\tau_{\text{had}}$ channels, additionally, both leptons must have opposite charges, to estimate the QMisID background in each corresponding event category.

The total systematic uncertainty in the estimate of the charge misalignment background for very *tight* electrons is about 30%, with the main contribution coming from closure tests at low p_T and from statistical uncertainties at high p_T .

Non-prompt light leptons

Non-prompt leptons originate from material conversions, decays of heavy flavour hadrons, and improper reconstruction of other particles, and vary depending on the event category. The main contribution to the non-prompt lepton background comes from $t\bar{t}$ production, followed by much smaller contributions from V +jets and single-top quark processes. This background is estimated

from simulations with data-driven corrections in the $2\ell\text{SS}$, 3ℓ , 4ℓ and $2\ell\text{SS}1\tau_{\text{had}}$ channels. The detailed analysis can be found in Ref. [120]

Fake τ_{had} candidates

In both the $2\ell\text{SS}1\tau_{\text{had}}$ and $3\ell 1\tau_{\text{had}}$ channels, the primary source of the fake τ_{had} background arises from events involving $t\bar{t}$ and $t\bar{t}V$, where a jet is misidentified as a τ_{had} candidate. A control region is defined that requires two oppositely charged leptons, at least three jets (with at least one b -tagged), and at least one τ_{had} candidate. This control region, which is particularly enriched in dileptonic $t\bar{t}$ events, ensures that the selected τ_{had} candidates are predominantly from misidentified jets. A normalisation factor is derived which is applied to correct the possible mismodeling of the fake τ_{had} simulation.

The normalisation factor is determined based on $p_T(\tau_{\text{had}})$ and applied categorically separately for single and triple τ_{had} candidates. For one-pronged (three-pronged) τ_{had} candidates, the normalisation factors vary from 1.05 ± 0.06 (1.25 ± 0.42) for $p_T(\tau_{\text{had}})$ in the range of 25–45 (25–50) GeV to 0.64 ± 0.12 (0.52 ± 0.71) for $p_T(\tau_{\text{had}}) \geq 70$ (75) GeV. The systematic uncertainties include statistical uncertainty, uncertainties associated with the contribution of true τ_{had} candidates subtracted from data, and deviations in the normalisation factors between the control region and a validation region enriched in Z +jets events. The cumulative systematic uncertainty is on average 13% (60%) for one-pronged (three-pronged) τ_{had} candidates and depends on $p_T(\tau_{\text{had}})$.

Simulated events with both a non-prompt light lepton and a fake τ_{had} candidate are scaled by the product of their respective per-lepton normalisation factors. The fraction of fake τ_{had} background with an electron misidentified as a τ_{had} candidate is about 10% and is estimated from the simulation.

The primary background in the $1\ell 2\tau_{\text{had}}$ channel comes from $t\bar{t}$ production, often with one or two fake τ_{had} candidates. As a jet in $t\bar{t}$, events are equally likely to be reconstructed as positively or negatively charged τ_{had} candidates. A systematic uncertainty of 30% for the estimated fake τ_{had} background in the SR is obtained by a closure test of the method in simulation.

4.9. $2\ell\text{SS}1\tau_{\text{had}}$ channel

In this channel, about 60% of the $t\bar{t}H$ signal events originate from $H \rightarrow \tau^+\tau^-$ decays, while the remaining 40% originate from $H \rightarrow WW^*$ decays. A typical $t\bar{t}H$ signal event is characterised by the presence of two prompt⁴ light leptons of the same charge, a hadronically decaying tau lepton, missing energy from four neutrinos, high jet multiplicity and b quark jets. Examples of tree-level Feynman diagrams illustrating the $t\bar{t}H$ signal in the $2\ell\text{SS}1\tau_{\text{had}}$ channel are shown in Figure 4.6.

Light leptons of the same charge, regardless of flavour ($e^\pm e^\pm$, $e^\pm \mu^\pm$, $\mu^\pm \mu^\pm$), are typically produced by the successive decays of a τ lepton from the Higgs boson and a W boson from a

⁴ An isolated lepton resulting from the decay of W , Z or Higgs bosons is called a “prompt” or “real” lepton.

top quark, both of which carry electric charges of the same sign, especially in the $H \rightarrow \tau^+\tau^-$ decay process. In particular, the hadronically decaying tau lepton usually carries an electric charge opposite to that of the light leptons. In such cases, four reconstructed jets are expected, not counting additional jets from initial and final state radiation, two of which are initiated by b quarks.

The $2\ell\text{SS}1\tau_{\text{had}}$ channel exhibits the highest sensitivity among the channels with τ_{had} candidates, a consequence of numerous improvements specifically tailored to this channel. Due to the presence of two same-charge light leptons and stringent criteria for a τ_{had} candidate, together with a high jet multiplicity, relatively low events are expected in this channel.

4.9.1. Fake lepton estimation

The accurate modelling of fake leptons is challenging and unreliable. Data-driven approaches are commonly used to estimate the contributions of fake leptons.

Estimation of fake with fake factor method

Background contributions from non-prompt light leptons and photon conversion, mainly from $t\bar{t}$ and $t\bar{t}\gamma$ processes, are estimated from data. Estimation of fake light leptons⁵ background, the fake factor (FF) method is used.

The FF method is a data-driven (DD) approach used to evaluate both the normalisation and the shape of backgrounds with fake light leptons. It uses a data set enriched in events with fake light leptons in regions orthogonal to the SR to estimate the contribution of fake light leptons to the SR. This method considers backgrounds with either single or multiple fake light leptons. The basic procedure of the FF method is outlined below.

First, a *anti-tight* selection is defined by reversing the identification and isolation variables for the light lepton. A light lepton that passes this selection is referred to as an *anti-tight* light lepton. A FF is computed using the *tight* and *anti-tight* criteria in dedicated control regions, called “FF regions”, which are enriched with fake light leptons having similar kinematics and composition to those in the SR where the background estimation is performed. The FF denotes the probability that a fake light lepton satisfies the *tight* criteria and is identified as a signal lepton. It is expressed as:

$$F = \frac{N_{\ell\ell}}{N_{\ell f}}, \quad (4.1)$$

where N_{ℓ} and N_f are the number of events with *tight* and *anti-tight* lepton selections respectively. The FF can depend on the light lepton kinematics such as p_T , and the additional activity in the event. Thus, the FF can be calculated as a function of the light lepton p_T .

⁵ Herein, the term “fake light leptons” refers exclusively to non-prompt light leptons and photon conversion. In the figures, it is labelled “fake lepton”.

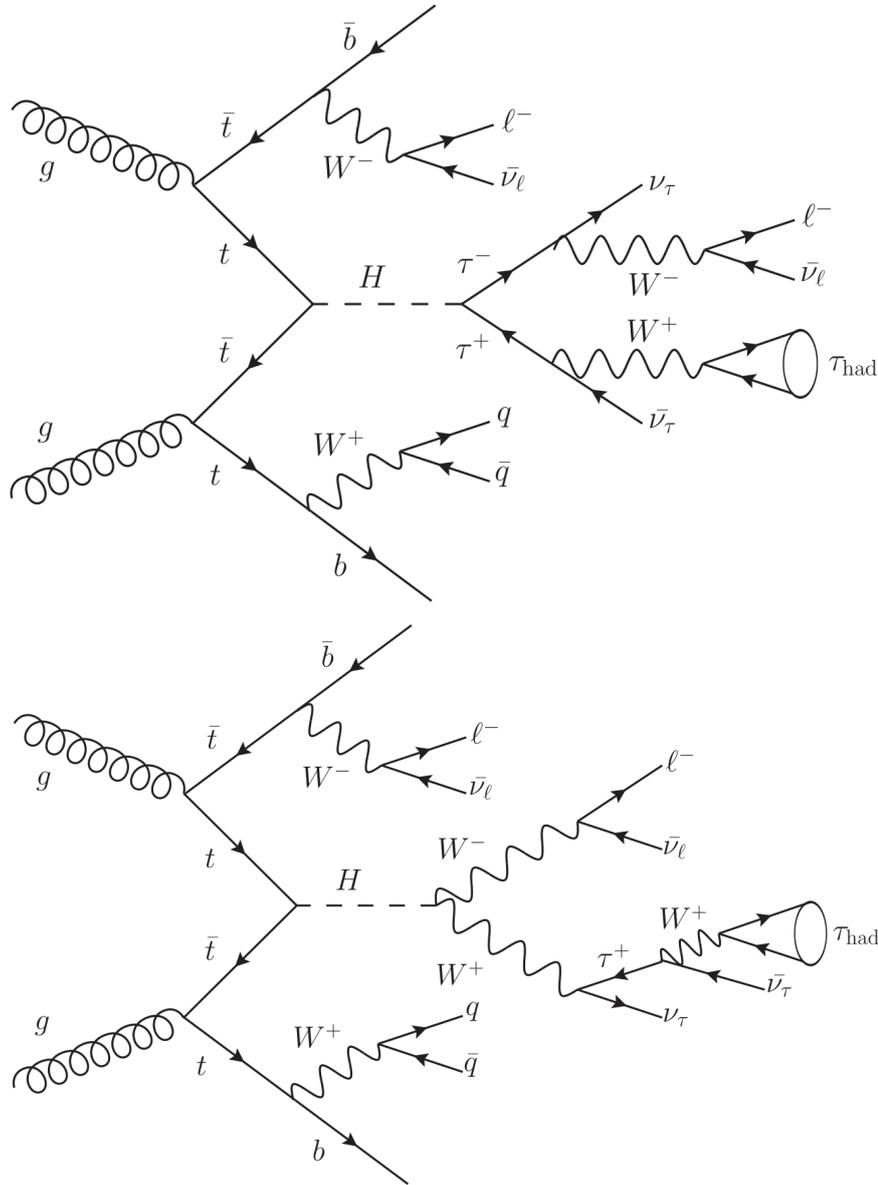


Figure 4.6. Examples of tree-level Feynman diagrams of a $t\bar{t}H$ signal with $2\ell\text{SS1 } \tau_{\text{had}}$ final state, where the Higgs boson decays into a pair of (a) τ leptons and (b) W bosons. The hadronically decaying tau lepton, τ_{had} , is either 1 - or 3 -track candidate with an opposite charge to that of the light leptons.

$$f_{\ell}(i) = \frac{N_{\ell}(i)}{N_{\not{\ell}}(i)}, \quad (4.2)$$

where i refers to the bin i^{th} p_{T} . The FF is applied to events in a control region similar to that of the SR, except that the light lepton is required to satisfy the *anti-tight* selection. The background contribution of fake light leptons in the SR is estimated as:

$$N_{\text{SR}}^{\text{fake}} = \sum_i f_{\ell}(i) \times N_{\not{\ell}}^{\text{CR}}(i), \quad (4.3)$$

where $N_{\text{SR}}^{\text{fake}}$ is the estimated number of events with fake light leptons in the SR, and the sum is over i^{th} p_{T} bins of the FF. $f_{\ell}(i)$ is the FF of the given p_{T} of the light lepton and $N_{\not{\ell}}^{\text{CR}}(i)$ is the number of events in the extrapolation CR.

Estimation strategy in the $2\ell\text{SS}1\tau_{\text{had}}$ channel

To estimate the contribution of fake lepton background events in the $2\ell\text{SS}1\tau_{\text{had}}$ channel, a data-driven FF method is employed. The sideband control regions, denoted B, C, and D, are defined in a two-dimensional space using two variables: Number of Jets (N_{jets})⁶ and the number of light leptons passing the “anti-tight” selection \mathcal{X} ⁷

A schematic representation of these control regions is provided in Figure 4.7. The control regions are designed to mimic the characteristics of the SR while enriching the composition with fake leptons. Regions C and D correspond to the low jet multiplicity regions $2 \leq N_{\text{jets}} \leq 3$, while region B follows the same jet selection criteria as the SR. A.

- *Region C*: This region requires exactly two *tight* light leptons, ensuring high purity for real leptons while maintaining sufficient statistics for background estimation.
- *Region D*: This region relaxes the lepton identification criteria compared to region C. It requires one *tight* and one *anti-tight* light lepton and is enriched in the number of events containing fake leptons.

To improve the statistical power of these control regions and leverage the similarity in fake lepton composition between the $2\ell\text{SS}$ and $2\ell\text{SS}1\tau_{\text{had}}$ channels, both regions C and D require either zero or one reconstructed τ_{had} candidate. A complete definition of all control regions, including the specific lepton identification criteria and τ_{had} selection, is provided in Table 4.7.

The FF for electrons and muons is calculated as follows:

$$f_{\ell}(i) = \frac{N_{\text{TT}}^{\text{C}}(i)}{N_{\text{TX}}^{\text{D}}(i)}, \quad (4.4)$$

⁶ This variable helps differentiate between signal and background processes with different jet production rates.

⁷ The “anti-tight” selection is a looser identification criterion compared to the SR, allowing for the enrichment of events containing fake leptons.

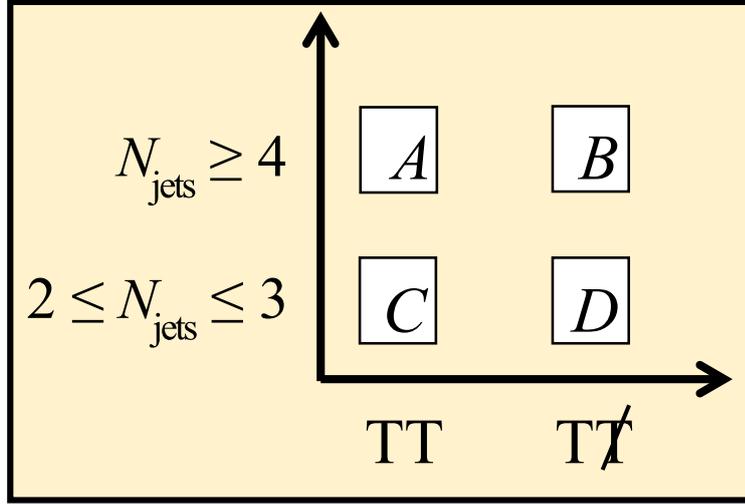


Figure 4.7. Schematic representation of the regions used in the FF method. The FF f_ℓ is derived from the “FF regions” C and D. This FF is then applied to the region B to estimate the fake light lepton background in the SR A. $T(\bar{T})$ represents the *tight* (anti-tight) light lepton. All the control regions are orthogonal to the SR [156].

	Common (CR C and CR D)
Number of light leptons	2, with same-charge
Lepton flavor	$ee, \mu\mu$
Light leptons p_T	$\geq 20\text{GeV}$
Z-veto	$ m(ee) - 91.2\text{GeV} > 10\text{GeV}$
Trigger requirement	SLT OR DLT
Trigger-matched leptons	≥ 1
Number of jets	2 AND 3
Number of b -tagged jets	≥ 1
Number of τ_{had} candidate	0 OR 1

Table 4.7. Definition of the control regions used for measuring the electron and the muon fake factors. The common selection criteria for regions C and D are listed under the title “Common”.

where N_{TT}^C and $N_{T\mathcal{X}}^D$ ⁸ are the number of events in regions C and D, respectively, and i refers to the i^{th} p_T bin. The contributions of charge misidentification background and the backgrounds with prompt light leptons and the $t\bar{t}H$ signal are subtracted from control regions C and D.

The fake factors are calculated only as a function of lepton p_T and found to be stable with respect to the η of the light lepton.

The region B is required to satisfy the same event selection as the SR (Table 4.5) except that one of the light lepton must be *anti-tight*. The fake factors are applied to the events in the region B event-by-event as a global event weight to estimate the fake light lepton background in the SR. The final number of events with fake light leptons in the SR is estimated as:

$$N_A^{\text{fake}} = \sum_i f_\ell(i) \times N_{T\mathcal{X}}^B(i), \quad (4.5)$$

where $N_{T\mathcal{X}}^B$ is the number of events in region B, and the sum is over p_T bins of the light lepton FF⁹. The contribution of the backgrounds with prompt light leptons, QMisID, and $t\bar{t}H$ are subtracted¹⁰ from $N_{T\mathcal{X}}^B$. In the case of opposite-flavour $e\mu, \mu e$ leptons, both the pair of *tight* electron and *anti-tight* muon ($e\mu$) and the *tight* muon and *anti-tight* electron (μe) are considered, separately. Furthermore, the fake τ_{had} candidates which arise from the $t\bar{t}$ process, as described later, are not accounted for separately as it is assumed that the fraction of real and fake τ_{had} candidates is similar in the signal and the extrapolation region B. Thus, the estimated number of events with fake light leptons in the SR A, N_A^{fake} , also include events with both fake light lepton and fake τ_{had} candidates.

The FF procedure is validated by checking that it correctly reproduces the SR event yield and the shape expected in $t\bar{t}$ simulation, as detailed in the following.

Assumptions of the method

In order for the FF procedure to work in the $2\ell\text{SS}1\tau_{\text{had}}$ channel, the following assumptions are made:

- The sources of the fake light leptons have the same composition in the SR and the control regions
- The backgrounds with prompt light leptons from different processes are known and their modelling by MC simulation is reliable, as the contribution of these backgrounds is subtracted from the control regions
- The major source of fake light leptons is the non-prompt light leptons from the $t\bar{t}$ process, or that the fake light leptons from different sources have the same FF

⁸ The events with both *anti-tight* leading and subleading light lepton ($N_{T\mathcal{X}}^D$) are considered. For brevity, only the symbol $N_{T\mathcal{X}}^D$ is used in the text and equation 4.4. The same applies to region B as well in equation 4.5.

⁹ In practice, the p_T of the *anti-tight* lepton in the event in region B is used to select the corresponding p_T bin of the light lepton FF.

¹⁰ Due to event-by-event application of the fake factors, the subtraction is done only at the end.

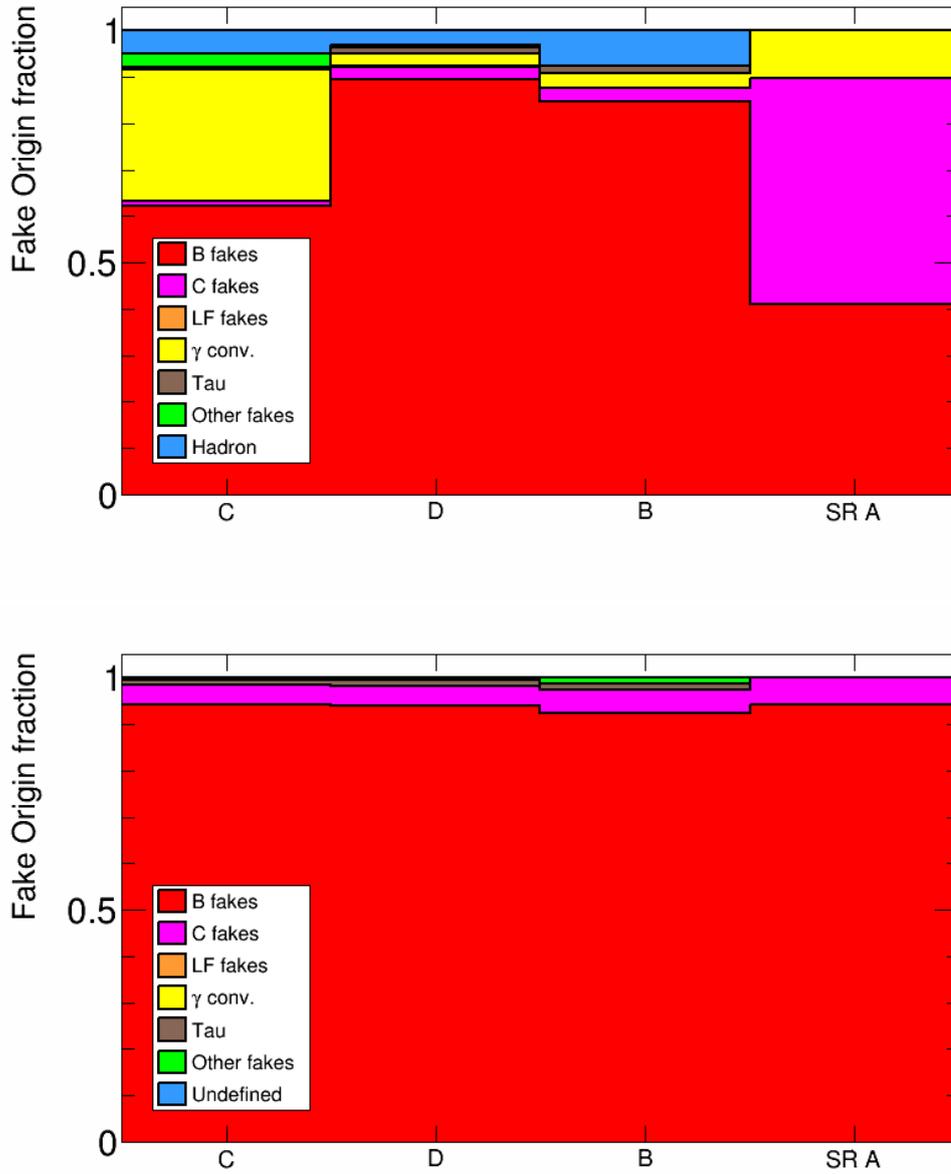


Figure 4.8. Sub-leading fake electron (top) and muon (bottom) origin fraction in $t\bar{t}$. The last bin corresponds to the SR.

- The FF is stable with respect to the jet multiplicity, η , and $\Delta R(\ell, \text{jet})$
- The fraction of real and fake τ_{had} candidates is similar in both SR and the extrapolation region B
- The ratio of the number of events in the SR and the control regions fulfil: $\frac{N_A}{N_B} = \frac{N_C}{N_D}$

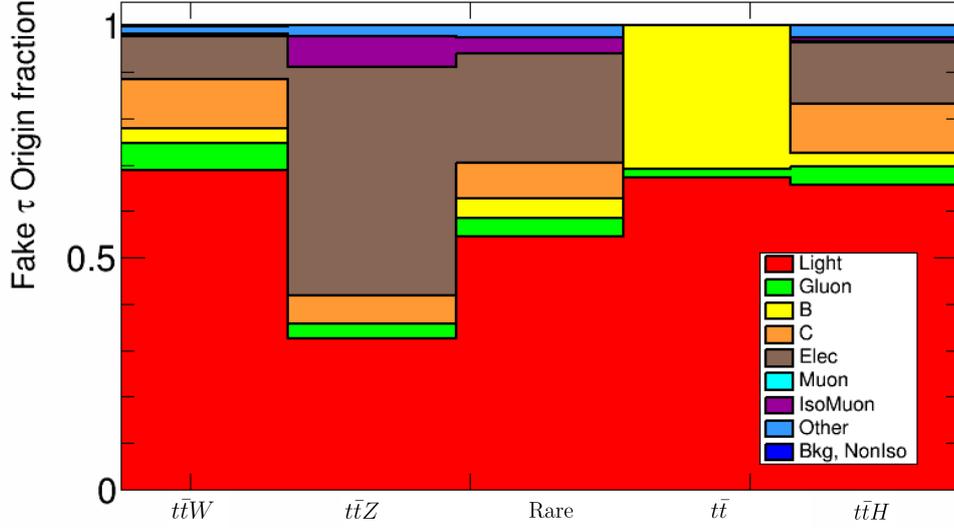


Figure 4.9. Fake τ origin fraction in SR for all the processes.

Composition of the fake light leptons

In the $2\ell\text{SS}1\tau_{\text{had}}$ channel, the main fake background is $t\bar{t}$ where one of the leptons comes from (HF) b -hadron, and c -hadron decays aka non-prompt leptons. Only thirteen raw events are observed in $t\bar{t}$ after passing the *tight* SR cuts (Table 4.5). The origin of the fake electron and muon is shown in Figure 4.8 in the SR and in the CR C and D as well as in the CR B.

For muons, the dominant fake source is heavy flavour b , c jets produced in $t\bar{t}$ semileptonic processes. Other fake sources are light-flavoured, τ lepton (excluding when τ originated from gauge bosons or top quarks in which case leptons are classified as prompt). In case the truth matching algorithm fails to find the origin of the lepton, they are labelled as “Unknown”.

For electrons, the dominant fake sources are heavy flavour and photon conversions. These photons can originate from ISR/FSR or π_0 decay products or radiate off the top, anti-top quark and interact with the detector material to produce a pair of electrons of which only one gets reconstructed. In SR three HF non-prompt electron events are contributing in $t\bar{t}$, and two raw electron events pass the SR selection for the $t\bar{t}\gamma$ process. Other spurious sources are light, τ leptons and hadrons.

Composition of the fake τ

In the $t\bar{t}$ background, a jet can fake a τ and enter the SR or any of the control regions. Figure 4.9 shows the origin of fake taus for each process in the SR. Mostly light jets are faking taus. It can be seen that the processes like $t\bar{t}W$, $t\bar{t}Z$, VV and even $t\bar{t}H$ also have fake taus.

In summary, in $t\bar{t}$ one of the leptons is always fake, and τ is fake as well. There are only these two cases with fake lepton and τ :

- One lepton is fake and comes from HF decays. The other lepton is real and comes from top to W boson decays. The τ_{had} is real and comes from top to W boson decays. Only one jet is b -tagged i.e. additional three jets are required to fulfil the SR criteria.
- One lepton is fake and comes from HF decays. The other lepton is real and comes from top to W boson decays. The other top-to-W boson decays hadronically into two jets where one of the jets is a fake τ . There are two jets left, out of which one is b -tagged i.e. additional two jets are required to fulfil the SR criteria.

There is no such event found where both leptons and τ_{had} are all fake, and no events where τ_{had} is real while both leptons are fake.

Measurement of electron and muon fake factors

Two CRs namely C and D are constructed for the e and μ FF calculation.

- CR C: Two leptons pass the *tight* selection, same-sign, $2 \leq \text{Jets} \leq 3$, zero OR exactly one τ_{had}
- CR D: One *tight* and one *anti-tight* lepton, same-sign, $2 \leq \text{Jets} \leq 3$, zero OR exactly one τ_{had}

All the other event selection cuts in the CR C and D are the same as in the SR. In the following, the *anti-tight* e or μ is denoted as $\bar{\mathcal{T}}$. The $2\ell\text{SS}$ requirement is applied to increase the statistics in these CRs.

The processes $t\bar{t}W$, $t\bar{t}Z$, diboson which produces two real prompt leptons are subtracted from the CR C and D, and there is a non-negligible amount of charge flip. These charge flip yields are estimated from data using OS leptons and then subtracted. All the $t\bar{t}H$ signal yields are also subtracted. In the CR C, where both leptons are *tight*, the sub-leading lepton p_{T} is taken for FF calculation while in the CR D the p_{T} of the *anti-tight* lepton is used.

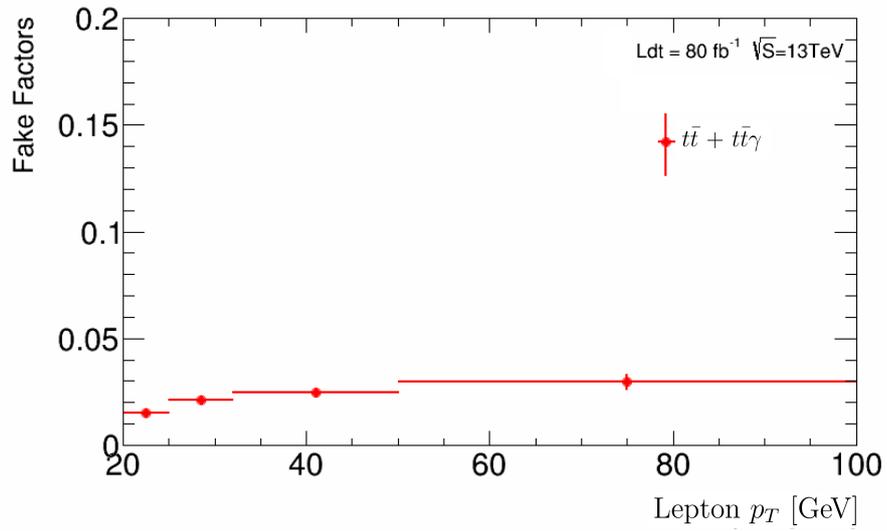
The charge flip yields are usually estimated from data by inverting the same charge to the opposite charge. In CR D one of the electrons is *anti-tight* and both electrons are required to be OS, which results in poor statistics. Because of the OS requirement, both leptons are real, and thus they do not pass the *anti-tight* selection. Therefore, the charge flip yield estimation is unreliable for this CR and is taken from MC.

The CR B is an extrapolation CR used for the estimation of the fakes in the SR. The definition of this CR is:

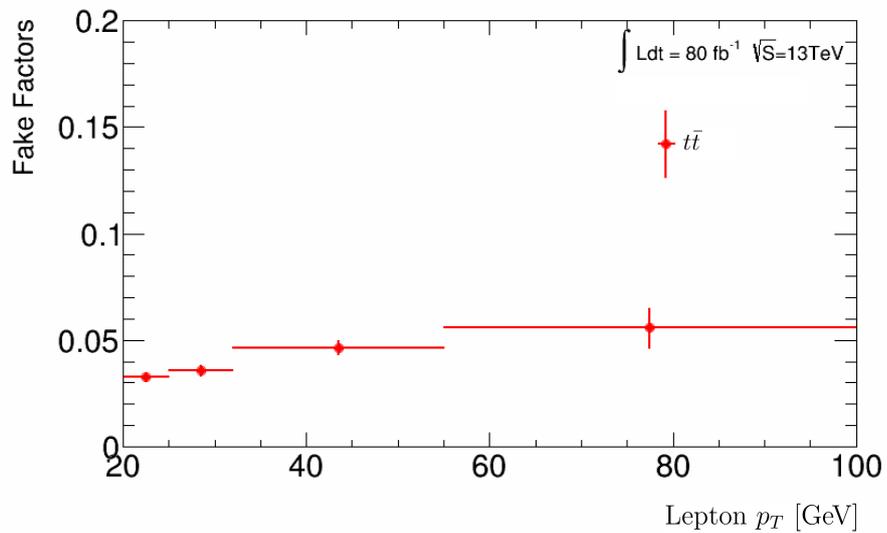
- CR B: At least one *anti-tight* lepton, exactly one τ_{had} , all the other event selection cuts are the same as in SR 4.5.

The e and μ fake factors are then applied to CR B using the p_{T} information of *anti-tight* lepton to estimate the e and μ fakes in the SR. In OF events, the electron fake factors are applied after requiring electrons to be anti-tight, and muon fake factors after requiring muons to be anti-tight.

The final fake estimation in the SR is estimated according to the equation 4.5. The inclusive yields in CR B,C,D are given in Table 4.8. Figure 4.10 shows the e and μ DD FFs before and after background subtraction.



Electron fake factors



Muon fake factors

Figure 4.10. The electron and muon fake factors as a function of p_T using $t\bar{t}$ and $t\bar{t}\gamma$ samples.

	B	C	D
$t\bar{t}H$	2.04 ± 0.05	23.87 ± 0.18	10.63 ± 0.12
$t\bar{t}$	51.44 ± 3.11	232.55 ± 6.84	8171.63 ± 40.95
$t\bar{t}W$	1.29 ± 0.13	262.48 ± 1.81	57.00 ± 0.87
$t\bar{t}Z$	1.48 ± 0.15	29.72 ± 0.38	10.94 ± 0.36
Di-boson	0.35 ± 0.08	84.43 ± 1.19	37.08 ± 1.55
Rare	1.19 ± 0.10	45.07 ± 0.98	43.53 ± 1.70
$t\bar{t}\gamma$	0.66 ± 0.16	21.92 ± 1.01	87.68 ± 2.13
Total Background	56.73 ± 3.13	684.40 ± 7.40	8407.29 ± 41.08
Data	65.00	768.00	9176.00

Table 4.8. Inclusive yields in B , C , and D regions.

Closure test

A closure test is performed on the $t\bar{t}$ MC sample to validate the FF method by comparing the prediction in the SR with the fake estimate obtained by the FF method. Figure 4.11 shows the electron and muon fake factors in $t\bar{t}$ used for the fake estimation. The $t\bar{t}\gamma$ sample is also used to estimate the electron FF. An overlap removal is applied between $t\bar{t}$ and $t\bar{t}\gamma$ to avoid double-counting events with real γ .

The non-closure and its uncertainty are defined as:

$$Non - closure = \frac{N_{MC} - N_{Fake}}{N_{Fake}}, \quad (4.6)$$

$$Non - closure\ uncertainty = \frac{N_{MC}}{N_{Fake}} \times \sqrt{\left(\frac{\sigma_{MC}}{N_{MC}}\right)^2 + \left(\frac{\sigma_{Fake}}{N_{Fake}}\right)^2}, \quad (4.7)$$

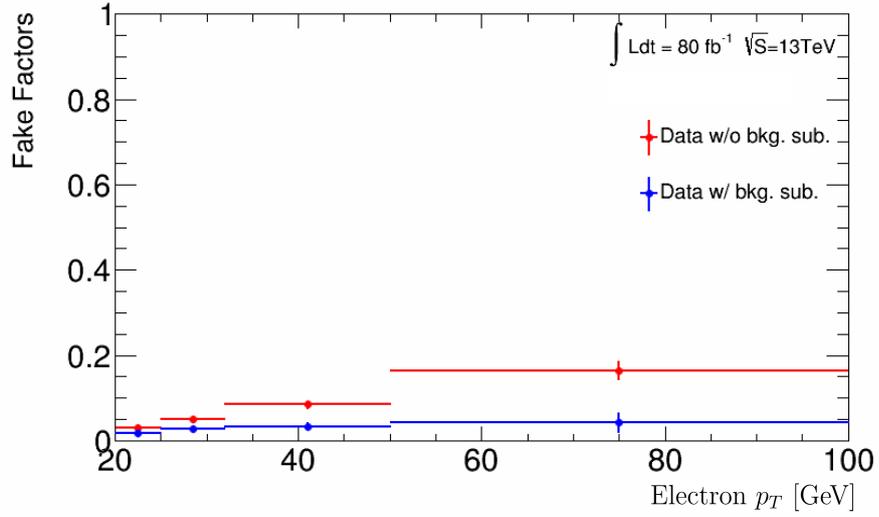
The closure test is performed inclusively. The events with charge flip electrons are vetoed for the closure test. The results of the closure test are summarized in Table 4.9.

The final fake estimation yields is:

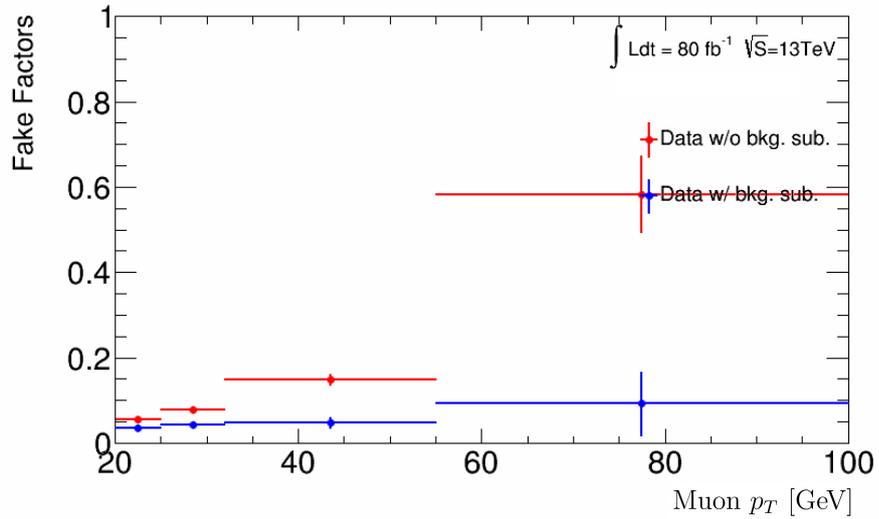
$$N_{Fake} = 2.10 \pm 0.30(stat.) \pm 0.76(syst.) \quad (4.8)$$

Dependence on N_{jets} , η , and ΔR

The assumption that fake factors are stable with respect to N_{jets} , η , and ΔR was checked. Figure 4.12 shows the fake factors as a function of N_{Jet} , and no dependence was observed.



Electron fake factors



Muon fake factors

Figure 4.11. The electron and muon as function of p_T fake factors before and after prompt background subtraction.

	SR yields	FF estimation	non-closure in [%]
Electron $t\bar{t}+t\bar{t}\gamma$	0.77 ± 0.36	0.58 ± 0.05	32.65 ± 63.41
Muon $t\bar{t}$	1.05 ± 0.40	0.94 ± 0.85	11.99 ± 43.73
Inclusive $t\bar{t}$	1.71 ± 0.53	1.51 ± 0.09	25.85 ± 35.61
Inclusive $t\bar{t}+t\bar{t}\gamma$	1.93 ± 0.54	1.53 ± 0.09	25.58 ± 36.35

Table 4.9. Closure test results using $t\bar{t}+t\bar{t}\gamma$ MC samples.

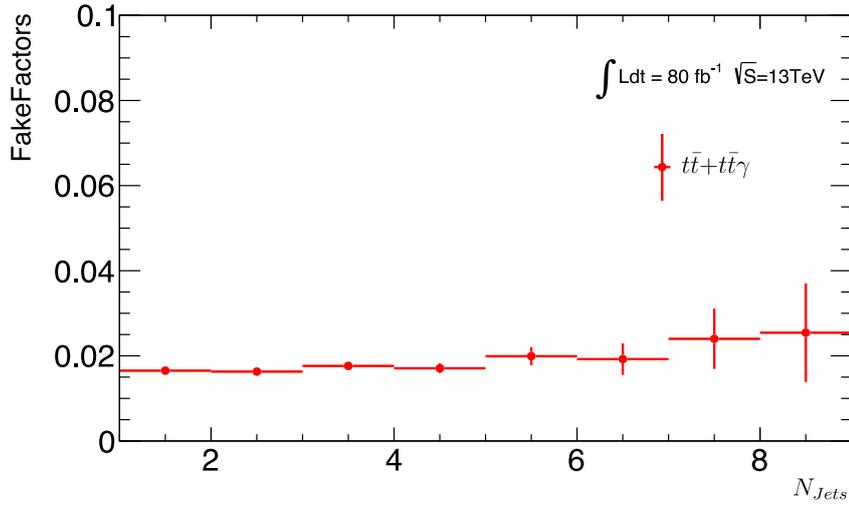
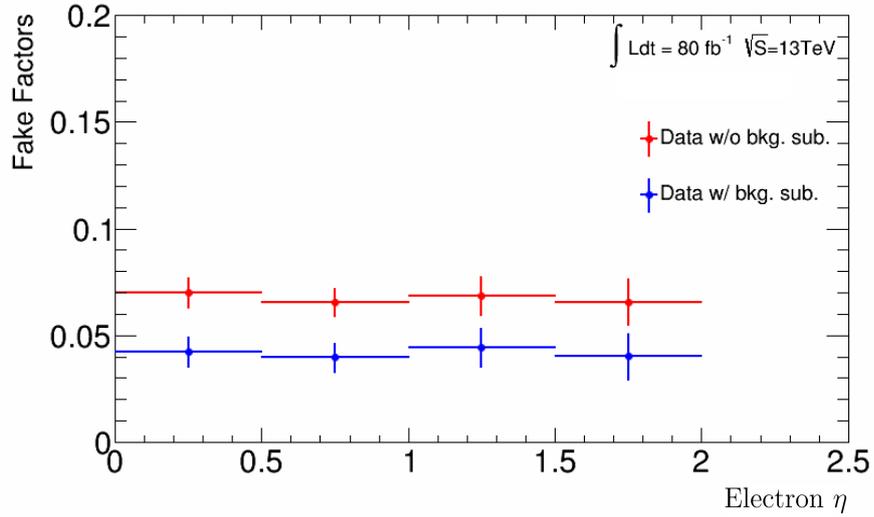
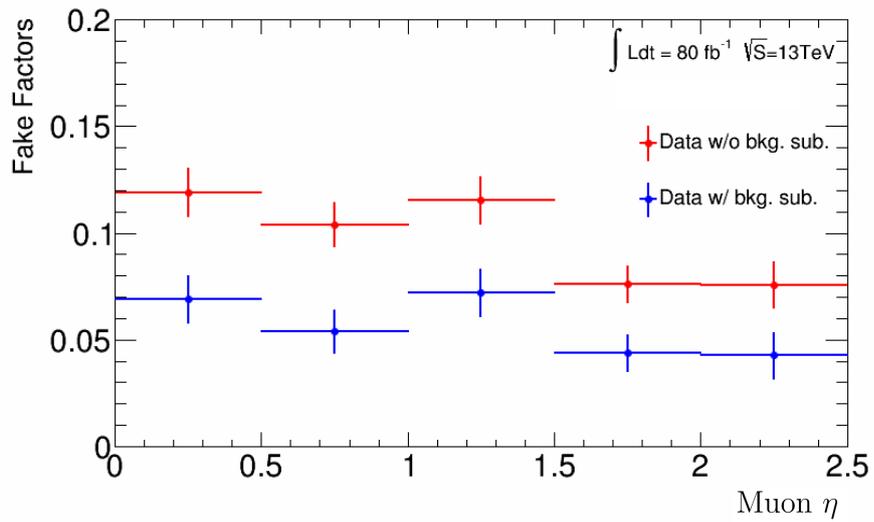


Figure 4.12. The fake factors as a function of N_{Jets} using $t\bar{t}$ and $t\bar{t}\gamma$ MC samples (no data is used because the high jet multiplicity region is blinded).

The CR C and D selections were applied, except for the jet multiplicity cut, for the FF estimation as a function of N_{Jet} . To avoid overlap in region C with the blinded signal data, $t\bar{t}$ and $t\bar{t}\gamma$ simulation samples were used.

In the FF method, the extrapolation from low jet multiplicity to high jet multiplicity was valid, and there was no bias due to this extrapolation. Furthermore, the dependence of fake factors as a function of lepton η was also checked using the same CR C and D selection. Figure 4.13 shows the e and μ fake factors as a function of η before and after background subtraction. The fake factors were stable with respect to lepton η .

Finally, the fake factors as a function of ΔR between the leading lepton and the closest jet, before and after background subtraction, are shown in Figure 4.14, and no dependence was observed.

Electron η fake factorsMuon η fake factors**Figure 4.13.** The DD electron and muon η fake factors before and after background subtraction.

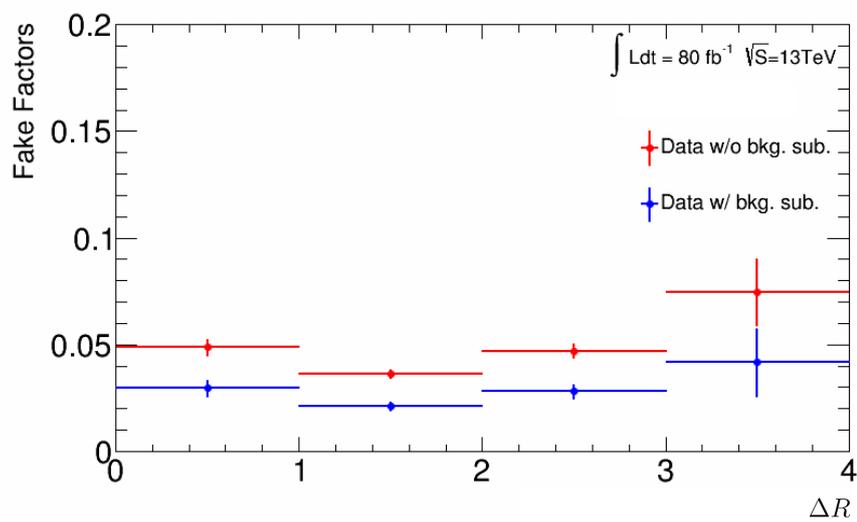


Figure 4.14. The fake factors as a function of ΔR between leading lepton and closest jet before and after background subtraction.

4.10. Systematic uncertainties

Statistical and systematic uncertainties limit the sensitivity of an analysis. Systematic uncertainties can arise from an incomplete understanding of the various sources of mismeasurement associated with physical objects, together with relevant theoretical uncertainties. These uncertainties are commonly divided into three categories: experimental uncertainties, modelling uncertainties related to background processes, and theoretical uncertainties related to signal processes.

A summary of systematic uncertainties is given in this section. An overview can be found in Table 4.10.

Systematic uncertainty	Components	Systematic uncertainty	Components
Luminosity (N)	1	$t\bar{t}H$ modelling	
Pileup modelling	1	Renormalisation and factorisation scales	3
Physics objects		PS and hadronisation model	1
Electron	8	Higgs boson branching ratio	4
Muon	11	Shower tune	1
Tau	7	PDF	32
JES and resolution	28	$t\bar{t}W$ modelling	
Jet vertex fraction	1	Radiation	1
Jet flavour tagging	17	Generator	1
E_T^{miss}	3	PDF	32
Total (Experimental)	77	Extrapolation	4
Data-driven background estimates		$t\bar{t}(Z/\gamma^*)$ (high mass) modelling	
Non-prompt light-lepton estimates ($3\ell, 3\ell 1\tau_{\text{had}}$)	1	cross-section (N)	2
Fake τ_{had} estimates	6	Generator	1
Electron charge misassignment	2	Renormalisation and factorisation scales	3
Total (Data-driven reducible background)	9	Shower tune	1
Template fit uncertainties		$t\bar{t}$ modelling	
Material conversions	1	Radiation	1
Internal conversions	1	WZ modelling	
HF non-prompt leptons	18	HF composition (N)	3
LF non-prompt leptons	2	Shower tune	1
Total (Template fit)	22	Other background modelling	
		cross-section (N)	22
		Total (Signal and background modelling)	120
		Total (Overall)	218

Table 4.10. Sources of systematic uncertainty considered in the analysis. “N” means that the uncertainty is taken as normalisation only for all processes and channels affected. Some of the systematic uncertainties are split into several components, as indicated by the number in the rightmost column.

The PDF uncertainties are correlated between the $t\bar{t}H$ signal and the $t\bar{t}W$ background [120].

4.10.1. Experimental uncertainties

Experimental uncertainties are related to the trigger efficiency, reconstruction and identification of the physics objects described in Section 4.4 and the global event activities.

Event-wide uncertainties

Event-wide uncertainties, which are independent of specific processes, affect the overall normalisation of the processes and are influenced by both luminosity and pileup measurements. In the $t\bar{t}H$ analysis, the uncertainty in the combined 2015-2017 integrated luminosity is 2%. This value is obtained by calibrating the luminosity scale using x-y beam separation scans, following a methodology similar to that described in Ref. [157], and relying on the LUCID-2 detector for the baseline luminosity measurements [83]. This luminosity uncertainty is applied uniformly to each MC simulated process. In addition, an uncertainty associated with MC scale factors is incorporated to account for differences in the pile-up distributions between data and MC simulations.

Light lepton uncertainties

The performance of electron and muon reconstruction, identification, isolation, and triggering differs between data and MC. To correct for these differences, scale factors are applied, and estimated using the tag-and-probe method [158, 159]. The associated systematic uncertainties resulting from these corrections are then incorporated into the final distributions used in the analysis. Furthermore, additional uncertainties related to lepton kinematics, arising from the resolution and scale of the electron energy (muon momentum) measurement, are also taken into account.

Jet uncertainties

Systematic uncertainties associated with the reconstructed jet objects are related to the JVT, the JES and the Jet Energy Resolution (JER).

JVT systematic uncertainties are determined by varying the JVT cut up and down using the JetVertexTaggerTool [160]. This uncertainty takes into account the residual contamination from pile-up jets after the application of pile-up suppression and the choice of MC generator. The modelling of JVT is validated using $Z(\rightarrow \mu\mu)+$ jet events and semileptonic $t\bar{t}$ events.

Several sources contribute to the uncertainties associated with JES and JER, each associated with specific aspects of the JES and JER calibration measurements are considered in Section 4.4.2. These uncertainties arise from the techniques and corrections derived in MC and include statistical, detector and modelling effects, jet flavour composition, pileup corrections and η dependence effects.

JES-related uncertainties consist of 30 (27) components used in the analysis, each with up/down variations and potentially different dependencies on jet p_T and η .

JER is measured independently for both data and MC using in situ techniques similar to those described in Ref. [161]. The expected fractional jet p_T resolution is derived using the JERSmearingTool, which is a function of the jet p_T and velocity. The systematic uncertainty is quantified as the quadratic difference between the JER for the data and the MC simulation. In the $t\bar{t}H$ analysis, a single comprehensive uncertainty is used to account for all JER-related effects.

Flavour tagging uncertainties

Flavour tagging uncertainties are associated with b -tagging efficiencies, as well as c - and light-jet mis-tagging efficiencies. These uncertainties comprise a mixture of statistical, experimental and modelling uncertainties arising from the flavour tagging calibration procedures. They affect the analyses through their influence on the scaling factors. In this analysis [120], where fewer b -jets are required in the SR the corresponding uncertainties have a reduced impact on the result.

Missing transverse energy

Systematic variations in E_T^{miss} arise as a result of the systematic variations associated with the objects (jets and leptons) provided as input to the E_T^{miss} calculation. Additional uncertainties associated with the scale and resolution of the soft term in the E_T^{miss} calculation are also taken into account. The uncertainty is estimated using events without real E_T^{miss} , namely $Z \rightarrow \ell\ell$ [162]. In this case, the soft term should be balanced against the object-based term to satisfy the no E_T^{miss} requirement, but resolution effects spoil the equivalence between soft and hard terms. This non-compensation between soft and hard terms is measured in the data and compared with the MC simulation. The difference is considered as uncertainty.

4.10.2. Background uncertainties

Modelling uncertainties related to background processes need to be considered. These include variations in normalisation or kinematics (shape). Normalisation uncertainties include cross-section uncertainties derived from theoretical calculations or other measurements. Shape uncertainties are typically assessed by data-driven methods or by comparison with MC simulations. By contrasting the nominal MC simulation with an alternative MC simulation in which theoretical or phenomenological parameters are varied, the effect of the underlying assumptions made in the MC simulation can be estimated.

$t\bar{t}W$ production

Given the importance of the $t\bar{t}W$ production, the modelling uncertainties are carefully derived. The absence of higher-order corrections and the discrepancies found between data and MC simulation for the $t\bar{t}W$ process necessitates a novel approach to its modelling. The systematic uncertainties associated with the $t\bar{t}W$ process include:

- The uncertainty related to the comparison of AMC@NLO+PYTHIA8 (0j@NLO) with SHERPA 2.2.1 (0j@NLO), both inclusive NLO setups, referred to as generator uncertainties
- The renormalization and factorization scale variations of the SHERPA2.2.1 multileg NLO (0, 1j@NLo + 2j@LO) sample by varying both scales by a factor 2 and 1/2

Data-driven systematics are also assessed for the $t\bar{t}W$ process to account for disagreement observed between data and MC simulation (Section 4.8.1)

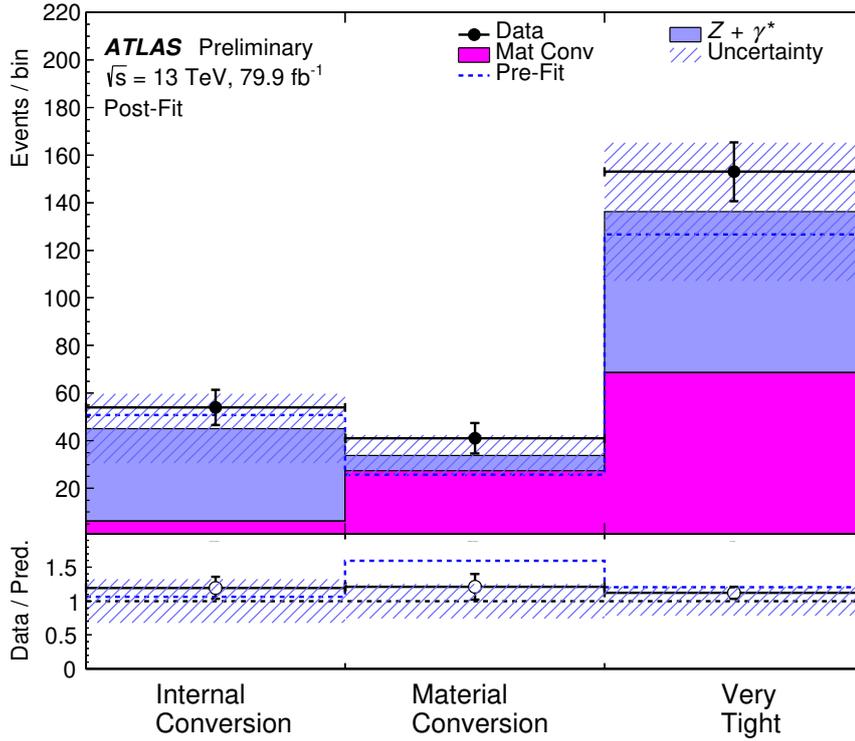


Figure 4.15. Comparison between data and prediction for the event yields in the trilepton validation region enriched in $Z \rightarrow \mu^+ \mu^- \gamma^* (\rightarrow e^+ e^-)$ candidate events, divided into three categories depending on the requirements satisfied by the electron: internal conversion, material conversion, or very *tight* [120].

Uncertainties related to fakes

The template fit method is the nominal method to estimate contributions from fake leptons. Shape systematic uncertainties are considered for the fake templates in $t\bar{t}H$ analysis, and are derived as follows. By inverting the lepton isolation requirements, regions with dominating fake contributions are obtained. The fake contribution is calculated after subtracting other prompt backgrounds from data in these inverted isolation regions. A systematic uncertainty for fake templates (especially for heavy flavour electrons and heavy flavour muons) is derived by comparing the calculated fake contribution with the MC simulation in all bins used in the template fit.

A special $Z \rightarrow \mu^+ \mu^- \gamma^* (\rightarrow e^+ e^-)$ sample, defined by requiring two opposite-charge muons and one electron, is selected to validate the template fit results on conversions shown in Figure 4.15.

Uncertainties for the internal and material conversions are obtained by comparing the data and the POWHEG+PYTHIA8 simulation of $Z (\rightarrow \mu^+ \mu^-) + \text{jets}$ production.

Charge mis-identification

In the $t\bar{t}H$ analysis, three sources of systematic uncertainties are considered for the charge flip rates: statistical uncertainties from the likelihood fit, differences between the rate extracted by the likelihood method and the truth-matching method on the simulated $Z \rightarrow ee$ events, and

systematic uncertainties from the variation of the dielectron invariant mass window used to define the Z region.

4.10.3. Signal and background modelling uncertainties

Theoretical uncertainties are outlined in this section. Table 4.11 gives an overview of the uncertainties, which include cross-section variations, PS effects and generator variations for primary backgrounds derived from MC predictions. Further details on the $t\bar{t}W$ systematics are given in the sections 4.10.2. For the $t\bar{t}Z/\gamma^*$ sample, uncertainties affecting the acceptance modelling and event kinematics include QCD scale and tuning variations, as well as a comparison with an alternative sample (see table 4.11). For the diboson backgrounds, a normalisation uncertainty of 50% is assigned and treated as uncorrelated between different subprocesses ($WZ+$ light jets, $WZ+ \geq 1c$, $WZ+ \geq 1b$ and $ZZ+$ jets). Rare background contributions (tZ , $t\bar{t}t$, $ttWW$, WtZ , VVV , $t\bar{t}$, $tHjb$, and WtH) are subject to an overall normalisation uncertainty of 50%.

$t\bar{t}H$ modeling uncertainties

The uncertainties affecting the $t\bar{t}H$ cross-section are $^{+5.8\%}_{-9.2\%}$, estimated by varying the QCD factorisation and renormalisation scales, and $\pm 3.6\%$ due to uncertainties in the PDFs and the strong coupling constant α_s . The uncertainties affect the measured signal strength and do not affect the measured cross-section.

Regarding the modelling of the acceptance and event kinematics, uncertainties arise from variations in the QCD factorisation and renormalisation scales, the choice of the PS and hadronisation model, the modelling of the ISR and the PDF uncertainties. Theoretical uncertainties due to the choice of the QCD scale are assessed by independently varying the renormalisation and factorisation scales by a factor of 0.5 and 2.0 with respect to the central value. The largest variation occurs when both scales are varied simultaneously in the same direction. Uncertainties associated with the choice of PS and hadronisation model are evaluated by comparing the nominal prediction with that derived from an alternative sample 4.1. The uncertainty associated with the ISR modelling is estimated by considering the Var3c A14 tune variation [92], which corresponds to a variation of α_s in the A14 tune. The uncertainty due to the choice of PDF set is determined using the PDF4LHC15 recipe [163], using 32 eigenvector shifts derived from fits to several NNLO PDF sets. Finally, uncertainties associated with the predicted Higgs boson branching ratios [164] are also taken into account.

Process (default)	X-section [%]	Generator (alternative)	PS (alternative)
$t\bar{t}H$ (POWHEG+PYTHIA8)	QCD Scale: $^{+5.8}_{-9.2}$ PDF ($+\alpha_s$): ± 3.6	- (AMC@NLO+PYTHIA8)	(A14 shower tune)
$t\bar{t}Z$ (AMC @NLO+PYTHIA8)	QCD Scale: $^{+9.6}_{-11.3}$ PDF ($+\alpha_s$): ± 4	- (SHERPA)	(A14 shower tune)
$t\bar{t}W$ (SHERPA NLO multileg)	-	AMC@NLO+PYTHIA8(0j@NLO) vs. SHERPA2.2.1 (0j@NLO)	

Table 4.11. Summary of theoretical uncertainties for $t\bar{t}H$, $t\bar{t}Z/\gamma^*$, and $t\bar{t}W$ MC predictions [120].

4.11. Results

A maximum likelihood fit (detailed in Section 5.8) is performed over all bins within the 25 event categories, outlined in Section 4.7, to determine the $t\bar{t}H$ cross-section, the normalisation factors of the $t\bar{t}W$ process, and other background contributions. The $t\bar{t}H$ acceptance within each category is projected by the simulation under the SM assumptions. Seventeen categories from the $2\ell SS$ and 3ℓ channels are designated as control regions. These regions serve to either establish or constrain various backgrounds, including material transformations, internal transformations, $t\bar{t}$ with non-prompt electrons and muons, $t\bar{t}W$, $t\bar{t}Z$ and VV .

Thirteen of the control regions use the total event yield (i.e, a single bin), while the remaining four control regions use different kinematic variables to distinguish between $t\bar{t}$ and $t\bar{t}W$ backgrounds (Section 4.8.1). The remaining eight categories serve as signal regions for measuring the $t\bar{t}H$ cross-section. A BDT discriminant is used for the $t\bar{t}H$ categories in the $2\ell SS$, 3ℓ and $1\ell 2\tau_{\text{had}}$ channels, while the total event yield is used in the remaining four signal regions.

The likelihood function $\mathcal{L}(\mu, \vec{\lambda}, \vec{\theta})$ is defined in terms of the signal strength parameter μ , which acts as a scaling factor for the $t\bar{t}H$ signal yield normalised to the SM prediction. It also depends on $\vec{\lambda}$, the normalisation factors for different backgrounds and $\vec{\theta}$, a set of nuisance parameters (NP) representing systematic uncertainties in the signal and backgrounds. These systematic uncertainties (as summarised in Table 4.10) can affect the estimated signal and background rates, the migration of events between categories, and the shape of the fitted discriminants. In the likelihood fit, both μ and $\vec{\lambda}$ are treated as free parameters. The NPs $\vec{\theta}$ allow for variations in signal and background expectations due to systematic uncertainties, constrained by Gaussian or log-normal distributions in the fit. Their fitted values indicate deviations from the nominal expectations that best fit the data. Statistical uncertainties in each bin, arising from the finite size of the simulated samples, are incorporated using dedicated parameters following the Beeston-Barlow technique [165].

The test statistic q_0 is defined as the profile likelihood ratio:

$$q_0 = -2 \ln(\mathcal{L}(0, \hat{\lambda}_0, \hat{\theta}_0) / \mathcal{L}(\hat{\mu}, \hat{\lambda}_{\hat{\mu}}, \hat{\theta}_{\hat{\mu}})), \quad (4.9)$$

where $\hat{\mu}$, $\hat{\lambda}_{\hat{\mu}}$, and $\hat{\theta}_{\hat{\mu}}$ are the values of the parameters that maximise the likelihood function, and $\hat{\lambda}_0$ and $\hat{\theta}_0$ are the values of the parameters that maximise the likelihood function when fixing μ to zero. The test statistic is evaluated with the RooFit package [166, 167] and is used to quantify how well the observed data agrees with the background-only hypothesis.

The fitted value, $\hat{\mu}$, is obtained by maximising the likelihood function over all parameters. The total uncertainty, $\Delta\hat{\mu}$, is determined by the variation of q_0 by one unit from its minimum. The systematic uncertainty is estimated by subtracting the statistical uncertainty (determined by refitting the data with all nuisance parameters fixed at their best-fit values) in quadrature from $\Delta\hat{\mu}$. The contribution from the background normalisation factors is included in the statistical uncertainty. Expected results are derived similarly to observed results, using predictions from

simulations and data-driven estimates with all nuisance parameters set to their best-fit values from the data fit. Significance is calculated using the test statistic and asymptotic formulae from Ref. [168].

Figures 4.16 (a) and 4.16 (b) illustrates the comparison between the data and the yields after adjustment by the fitted predictions for the 25 event categories considered. For all categories, the observed yields are in close agreement with the predictions derived from the fit, taking into account the uncertainties. The background distribution before fitting is shown as a dashed blue histogram. In addition, Figure 4.17 shows the distributions of the BDT discriminants in $2\ell\text{SS}$, 3ℓ , and $1\ell 2\tau$ channels.

In the remaining categories, the fit only includes event yields. Discrepancies between the data and the predictions are evident before fitting, but overall satisfactory agreement is achieved after fitting. Figure 4.18 shows the data, background and $t\bar{t}H$ signal yields, where the bins of the final discriminant overall event categories are merged into logarithmic bins of $\log_{10}(S/B)$, where S is the expected $t\bar{t}H$ signal yield and B is the fitted background yield. The cumulative background before fitting is shown as a dashed blue histogram.

The significance of the observed (expected) excess over the background-only prediction ($\mu = 0$) is 1.8 (3.1) standard deviations. The best fit value for μ in the fit is:

$$\hat{\mu} = 0.58_{-0.25}^{+0.26} \text{ (stat.)}_{-0.15}^{+0.19} \text{ (exp.)}_{-0.11}^{+0.13} \text{ (bkg. th.)}_{-0.07}^{+0.08} \text{ (sig. th.)} = 0.58_{-0.33}^{+0.36}, \quad (4.10)$$

Figure 4.19 shows the best-fit value of μ for each specific channel and the combination of all channels. The results for individual channels are derived from a simultaneous fit with the signal-strength parameter for each channel floating independently.

The fitting procedure produces normalisation factors for several significant irreducible and reducible backgrounds. Of particular importance are the three normalisation factors measured for the $t\bar{t}W$ background over the $2\ell\text{SS}$ and 3ℓ event categories: $\hat{\lambda}_{t\bar{t}W}^{2\ell\text{LJ}} = 1.56_{-0.28}^{+0.30}$, $\hat{\lambda}_{t\bar{t}W}^{2\ell\text{HJ}} = 1.26_{-0.18}^{+0.19}$, and $\hat{\lambda}_{t\bar{t}W}^{3\ell} = 1.68_{-0.28}^{+0.30}$. These values are consistent with each other and systematically exceed unity, indicating a data preference for a higher $t\bar{t}W$ cross-section compared to the updated theoretical estimate (Section 4.3.1). Since the modelling uncertainties for $t\bar{t}W$ are designed to affect only the shape of the distribution while maintaining a fixed total yield, the normalisation factors serve as scaling parameters for the $t\bar{t}W$ events chosen in this analysis. However, uncertainties in extrapolating the $t\bar{t}W$ scaling factor to the inclusive phase space are not considered.

The detailed breakdown of the uncertainties affecting the measured signal strength is given in Table 4.12. The most significant uncertainty arises from the scale and resolution of the jet energy. In addition, systematic uncertainties from theoretical predictions, in particular in the modelling of the $t\bar{t}W$ production and the normalisation of the $t\bar{t}(Z/\gamma^*)$ background, also contribute significantly. In addition, the uncertainty associated with the τ_{had} background estimate is considerable.

Figure 4.20 shows the ranking of NPs is highly ranked. The NPs associated with the b -jet multiplicity and total charge extrapolation uncertainties on the $t\bar{t}W$ background in the $2\ell\text{SS}$ channel have been adjusted by +0.33 and +0.75 pre-fit standard deviations, respectively, with

their uncertainties reduced by factors of 3 and 2, respectively. Conversely, the NP associated with the uncertainty in the closure of the non-prompt lepton estimate in the $1\ell 2\tau_{\text{had}}$ channel has been adjusted by -0.56 pre-fit standard deviations, with its uncertainty reduced by a factor of 1.7. However, this uncertainty does not significantly affect the $t\bar{t}H$ cross-section measurement.

An extrapolation to the inclusive phase space, assuming SM $t\bar{t}H$ kinematics, is made and the measured $t\bar{t}H$ production cross-section is:

$$\hat{\sigma}(t\bar{t}H) = 294_{-127}^{+132} \text{ (stat.)}_{-74}^{+94} \text{ (exp.)}_{-56}^{+73} \text{ (bkg. th.)}_{-39}^{+41} \text{ (sig. th.) fb} = 294_{-162}^{+182} \text{ fb}, \quad (4.11)$$

The SM cross-section is $\sigma(t\bar{t}H) = 507_{-50}^{+35}$ fb computed at NLO in QCD and electroweak couplings [93]. The measured cross-section is consistent with the SM prediction within uncertainties.

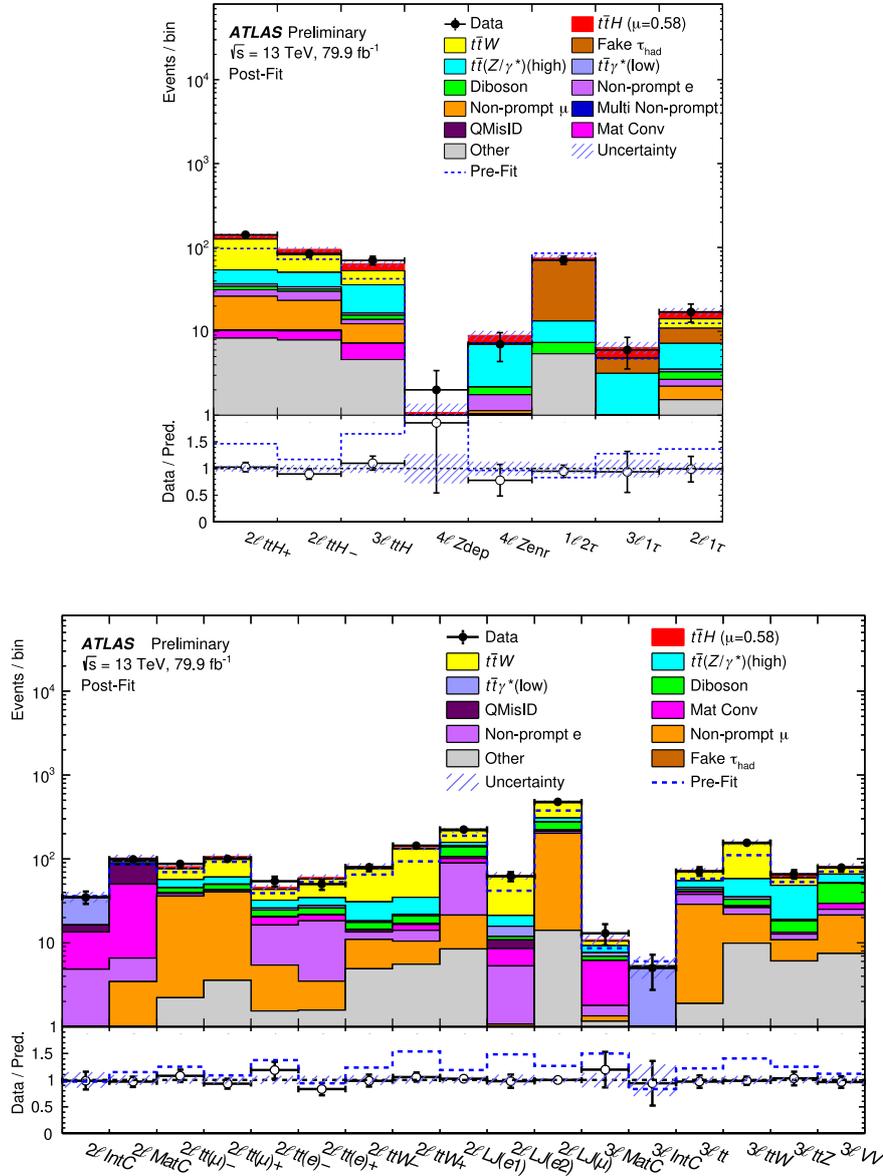


Figure 4.16. Comparison between data and prediction for the event yields in (top) the eight $t\bar{t}H$ categories and (bottom) the 17 control-region categories. The background contributions after the likelihood fit (“Post-Fit”) are shown as filled histograms. The total background before the fit (“Pre-Fit”) is shown as a dashed blue histogram. The $t\bar{t}H$ signal, scaled according to the results of the fit, is shown as a filled red histogram added to the post-fit background. The size of the combined statistical and systematic uncertainty in the total signal-plus-background prediction is indicated by the blue-hatched band. The ratio of the data to the total post-fit prediction is shown in the lower panel [120].

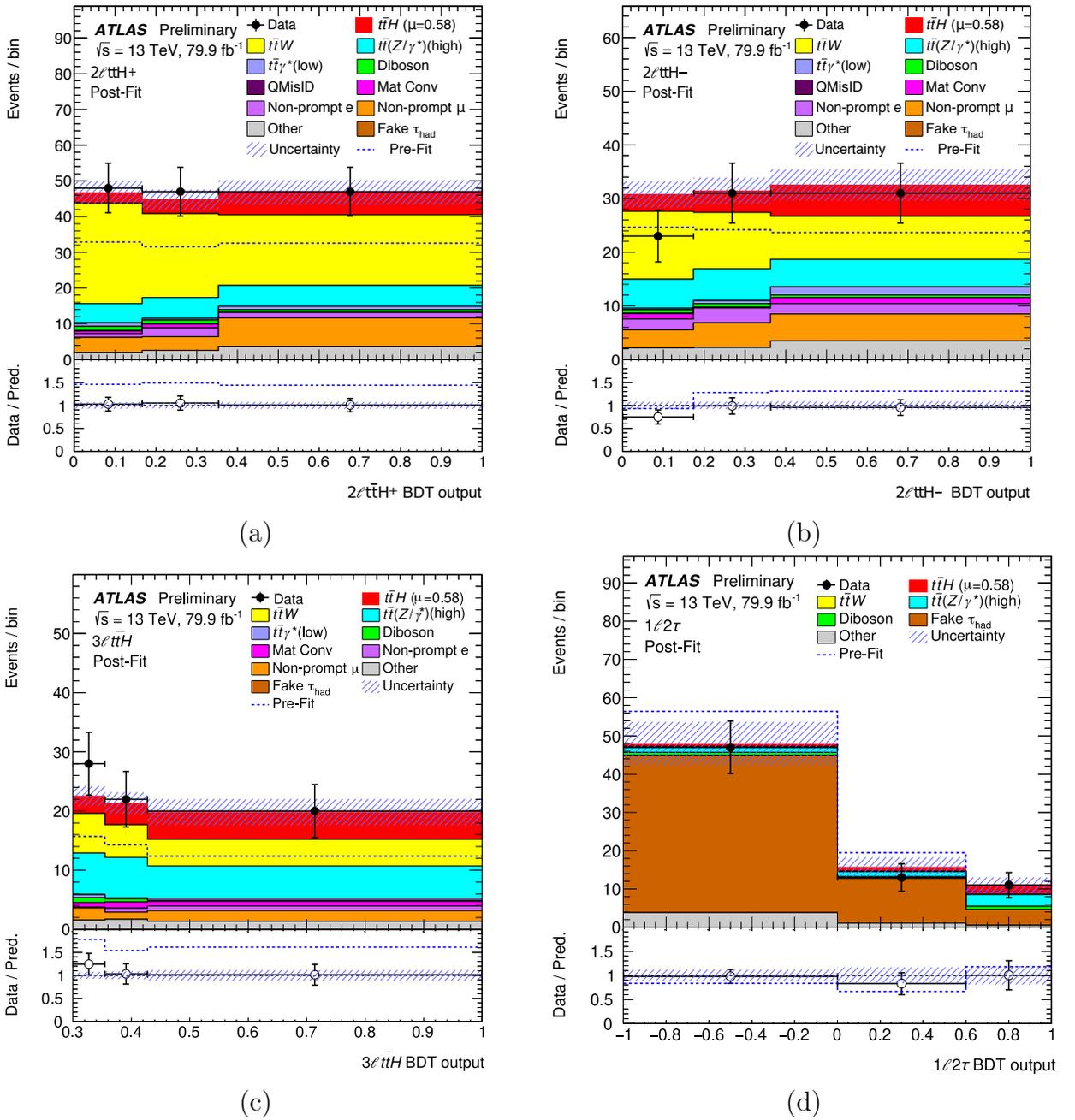


Figure 4.17. Comparison between data and prediction for the BDT discriminants used in different $t\bar{t}H$ categories: (a) $2l t\bar{t}H+$, (b) $2l t\bar{t}H-$, (c) $3l t\bar{t}H$, and (d) $1l 2\tau_{\text{had}}$. The background contributions after the likelihood fit (“Post-Fit”) are shown as filled histograms. The total background before the fit is shown as a dashed blue histogram. The total background before the fit (“Pre-Fit”) is shown as a dashed blue histogram. The $t\bar{t}H$ signal, scaled according to the results of the fit, is shown as a filled red histogram added to the post-fit background. The size of the combined statistical and systematic uncertainty in the total signal-plus-background prediction is indicated by the blue-hatched band. The ratio of the data to the total post-fit prediction is shown in the lower panel [120].

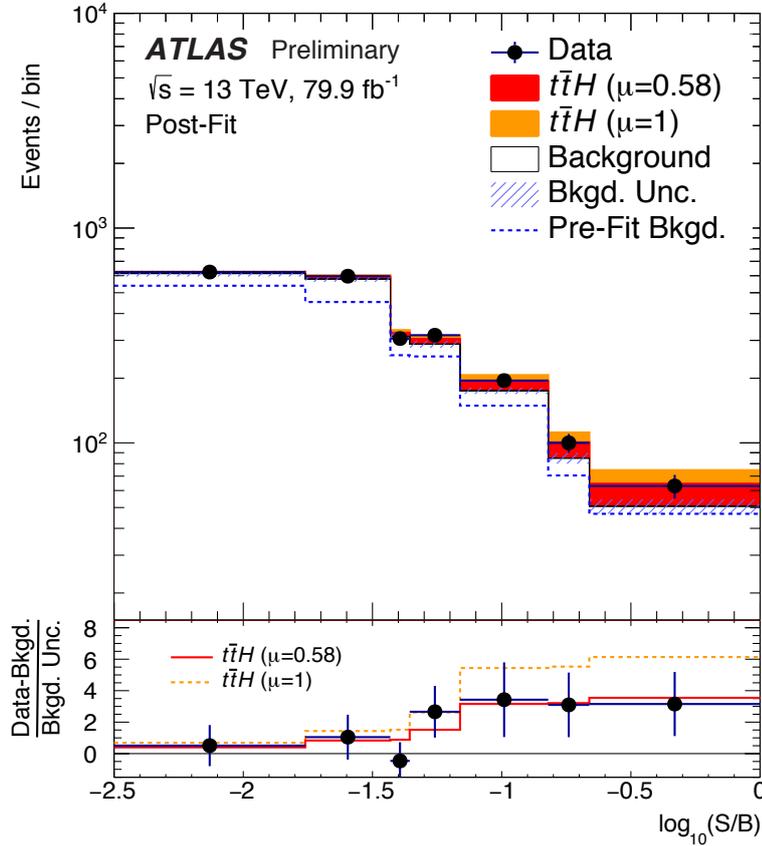


Figure 4.18. Event yields as a function of $\log_{10}(S/B)$ for data, background and a Higgs boson signal with $m_H = 125\text{GeV}$. The discriminant bins in all categories are combined into bins of $\log_{10}(S/B)$, where S is the expected $t\bar{t}H$ signal yield and B the background yield from the unconditional fit. The background yields are shown as the fitted values, while the $t\bar{t}H$ signal yields are shown for the fitted value ($\mu = 0.58$) and the SM prediction ($\mu = 1$). The total background before the fit is shown as a dashed blue histogram. The size of the combined statistical and systematic uncertainty in the background prediction is indicated by the blue-hatched band. The pull (residual divided by its uncertainty) of the data relative to the background-only prediction is shown in the lower panel, where the full red line (dashed orange line) indicates the pull of the prediction for the signal with $\mu = 0.58$ ($\mu = 1$) and background relative to the background-only prediction. [120].

Uncertainty source	$\Delta\hat{\mu}$	
Jet energy scale and resolution	+0.13	-0.13
$t\bar{t}(Z/\gamma^*)$ (high mass) modelling	+0.09	-0.09
$t\bar{t}W$ modelling (radiation, generator, PDF)	+0.08	-0.08
Fake τ_{had} background estimate	+0.07	-0.07
$t\bar{t}W$ modelling (extrapolation)	+0.05	-0.05
$t\bar{t}H$ cross-section	+0.05	-0.05
Simulation sample size	+0.05	-0.05
$t\bar{t}H$ modelling	+0.04	-0.04
Other background modelling	+0.04	-0.04
Jet flavour tagging and τ_{had} identification	+0.04	-0.04
Other experimental uncertainties	+0.03	-0.03
Luminosity	+0.03	-0.03
Diboson modelling	+0.01	-0.01
$t\bar{t}\gamma^*$ (low mass) modelling	+0.01	-0.01
Charge misassignment	+0.01	-0.01
Template fit (non-prompt leptons)	+0.01	-0.01
Total systematic uncertainty	+0.25	-0.22
Intrinsic statistical uncertainty	+0.23	-0.22
$t\bar{t}W$ normalisation factors	+0.10	-0.10
Non-prompt leptons normalisation factors (HF, material conversions)	+0.05	-0.05
Total statistical uncertainty	+0.26	-0.25
Total uncertainty	+0.36	-0.33

Table 4.12. Breakdown of the contributions to the uncertainties in $\hat{\mu}$. The contribution of the different sources of uncertainty is evaluated after the fit described in Section 4.11. The total statistical uncertainty is evaluated, by fixing all the nuisance parameters in the fit except for the free-floating background normalisation factors. The contribution from the uncertainty in those normalisation factors is then included in the quoted total statistical uncertainty rather than in the systematic uncertainty component. The statistical uncertainty evaluated after fixing the background normalisation factors is then indicated as “intrinsic statistical uncertainty”. Statistical uncertainties from data-driven background estimates are included within the experimental uncertainties. The other quoted numbers are obtained by repeating the fit after having fixed a certain set of nuisance parameters corresponding to a group of systematic uncertainty sources, and subtracting in quadrature the resulting total uncertainty of μ from the uncertainty from the full fit. The same procedure is followed for quoting the individual effects of background normalisation factors. Due to rounding effects and small correlations between the different sources of uncertainty, the total systematic uncertainty differs from the sum in quadrature of the individual sources [120].

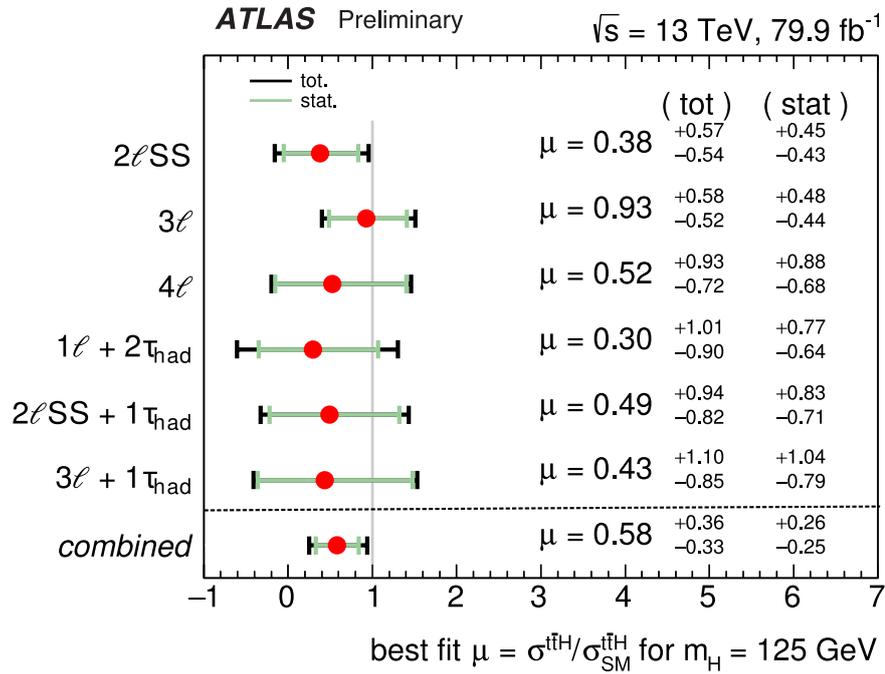


Figure 4.19. The observed best-fit values of the $t\bar{t}H$ signal strength μ and their uncertainties by analysis channel and combined. The individual μ values for the channels are obtained from a simultaneous fit with the signal-strength parameter for each channel floating independently. The SM prediction corresponds to $\mu=1$ [120].

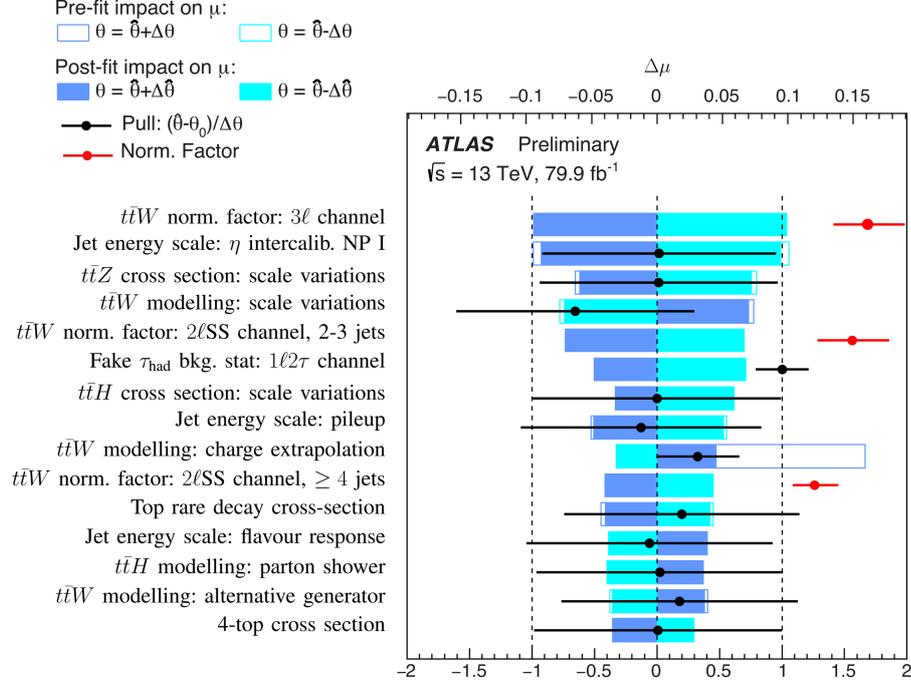


Figure 4.20. Ranking of the parameters included in the fit according to their impact on the signal strength μ . Only the 15 most highly ranked parameters are shown. The empty blue rectangles correspond to the pre-fit impact on μ and the filled blue ones to the post-fit impact on μ , both referring to the upper x -axis scale. The impact of each nuisance parameter (NP), $\Delta\mu$, is computed by comparing the nominal best-fit value of μ with the result of the fit when fixing the considered NP to its best-fit value, $\hat{\theta}$, shifted by its pre-fit (post-fit) uncertainties $\pm\Delta\theta$ ($\pm\Delta\hat{\theta}$). The black points show the pulls of the NPs relative to their nominal values, θ_0 . The nominal value for all NPs is $\theta_0 = 0$, with the exception of the NP associated with the limited sample size in the estimation of the fake τ_{had} background in the 1 ℓ 2 τ_{had} channel, for which the nominal value is $\theta_0 = 1$. These pulls and their relative post-fit errors, $\Delta\hat{\theta}/\Delta\theta$, refer to the lower x -axis scale. The *t* \bar{t} W normalisation factors (red points) also refer to the lower x -axis scale and correspond to the floating normalisations of the *t* \bar{t} W background, for which the pre-fit impact on μ is not defined. The nominal value of the *t* \bar{t} W normalisation factors is 1, which corresponds to the *t* \bar{t} W prediction based on the “updated theoretical cross-section” discussed in Section 4.2. For experimental uncertainties decomposed into several independent sources, “NP I” corresponds to the first nuisance parameter, ordered by its impact on μ Ref. [120].

4.12. Summary

A search for $t\bar{t}H$ production in multilepton final states has been performed using 80fb^{-1} of proton-proton collision at $\sqrt{s}=13$ TeV recorded by the ATLAS experiment at the LHC. Six final states are considered $2\ell\text{SS}$, 3ℓ , 4ℓ , $1\ell 2\tau_{\text{had}}$, $2\ell\text{SS}1\tau_{\text{had}}$ and $3\ell 1\tau_{\text{had}}$. Six final states were analysed, targeting Higgs boson decays to WW^* , $\tau\tau$ and ZZ^* . An excess of events over the SM background was observed (1.8 standard deviations significance) but below the SM expectation (3.1 standard deviations). The measured cross section ($\sigma(t\bar{t}H) = 294_{-162}^{+182}$ fb) agrees with the SM prediction.

Part III

Proton-Lead analysis

Chapter 5

Observation of $t\bar{t}$ production in the lepton+jets and dilepton channels in $p+\text{Pb}$ collisions at $\sqrt{s_{\text{NN}}} = 8.16$ TeV with the ATLAS detector

5.1. Introduction

Heavy-ion (HI) collisions at the TeV-scale energies of CERN's Large Hadron Collider (LHC) have provided the opportunity to detect several elementary particles for the first time in lead-lead (Pb+Pb) and proton-lead ($p+\text{Pb}$) systems. With the first sightings of the W^\pm boson [169, 170], the Z boson [171–173], bottom quark jets (b quark jets) [174, 175], and the τ lepton [176, 177], only two Standard Model (SM) particles remain to be directly observed in Pb+Pb collisions: the Higgs boson and the top quark. Although the production cross-section of the Higgs boson is insufficient for a definitive observation in Pb+Pb collisions at the LHC, the detection of top quark production is within the realm of possibility.

In heavy-ion (HI) collisions, the presence of top quarks provides a novel probe of nuclear modifications of the parton distribution functions (nPDFs) [178, 179] in a kinematic region ($x \approx 5 \cdot 10^{-3} - 0.05$) that lacks strong constraints from other measurements. In addition, top quarks will serve as a distinctive tool for assessing the properties of the strongly interacting quark-gluon plasma (QGP) [180] formed during ultra-relativistic Pb+Pb collisions at the LHC. These properties can be inferred from changes in various observables observed in HI collisions compared to reference measurements in the proton-proton (pp) system.

By studying the top quark yields through the kinematics of the electrons and muons resulting from their decays, precise insights can be gained [181] into the underlying nuclear gluon distribution function in the unexplored high Bjorken- x region. In this regime, expected adjustments due to anti-shadowing and EMC effects [179] could reshape their profile relative to the free proton scenario, potentially leading to a remarkable increase of up to 10% in the $t\bar{t}$ production cross-section, in contrast to measurements in pp collisions.

The gluon nPDF, crucial for perturbative calculations in quantum chromodynamics (QCD) at LHC energies, remains insufficiently constrained due to limited experimental data sensitive

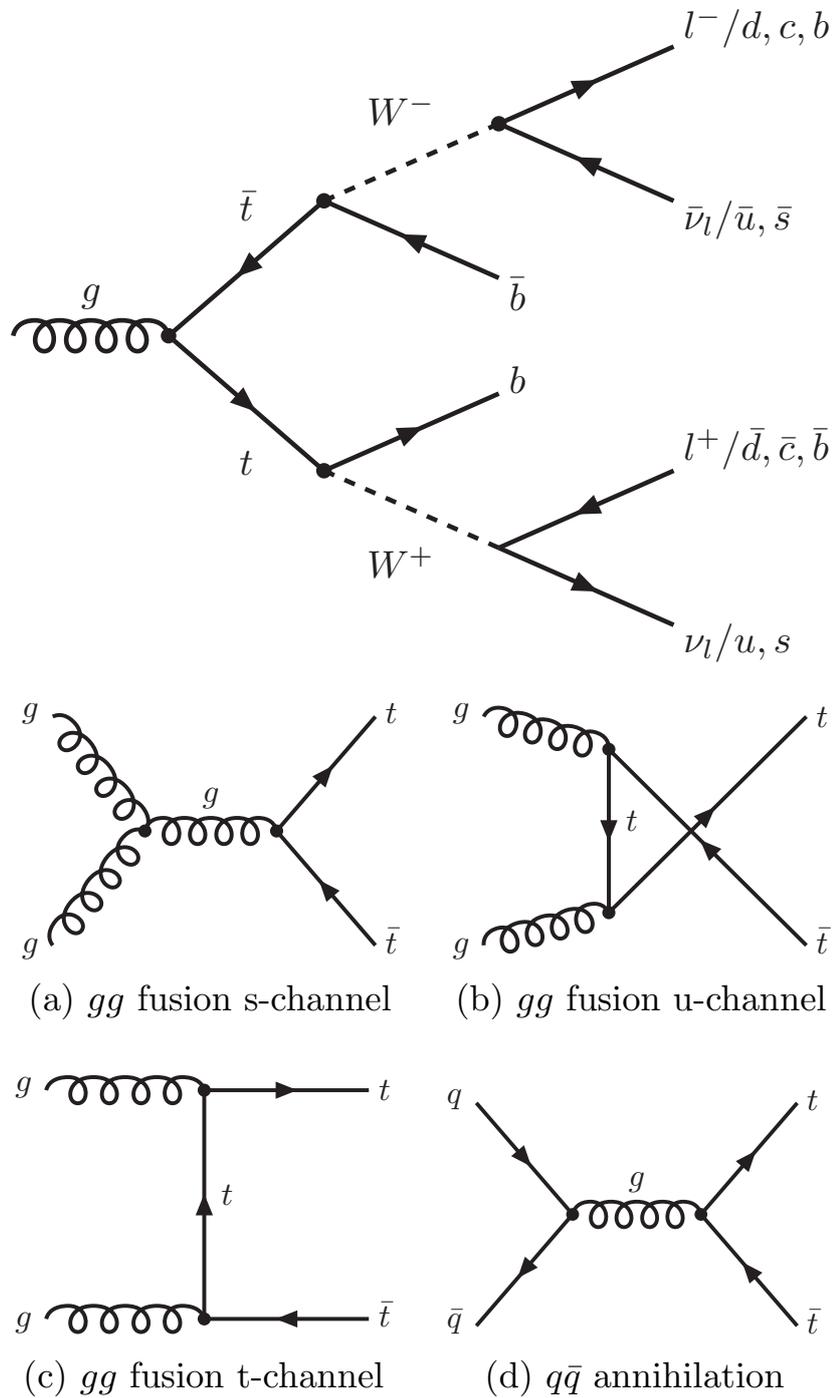


Figure 5.1. Feynman diagrams for top-quark pair production (top) and decay process (bottom) at leading order.

to gluon nPDFs at perturbative scales [179, 182, 183]. Nevertheless, changes in event yields in HI collisions compared to the pp reference system can be attributed to both initial state effects (e.g. different parton distribution functions for heavy nuclei compared to nucleons) and final state effects arising from QGP formation. Understanding the nPDFs in heavy nuclei is therefore crucial for the accurate extraction of QGP properties from experimental data.

The top quark, known as the heaviest elementary particle, has a fleeting existence, decaying predominantly via $t \rightarrow Wb$ with an almost 100% branching ratio. The subsequent decay of the W boson can take two primary paths: leptonically ($W \rightarrow \ell\nu_\ell$) or hadronically ($W \rightarrow q\bar{q}'$) [184].

At the LHC, top quarks are primarily produced in top quark-antiquark pairs ($t\bar{t}$) by gluon-gluon fusion, overshadowing single top quark production [185]. With extensive integrated luminosities of $p+\text{Pb}$ and $\text{Pb}+\text{Pb}$ data sets from Run 2, the detection of the $t\bar{t}$ process in heavy ion (HI) collisions at the LHC has become feasible for the first time. Specifically, $t\bar{t}$ events reconstructed in the ℓ +jets ($t\bar{t} \rightarrow WbW\bar{b} \rightarrow \bar{\ell}\nu_\ell b q \bar{q}' \bar{b}$) and dileptonic channels ($t\bar{t} \rightarrow WbW\bar{b} \rightarrow \bar{\ell}\nu_\ell b \ell' \bar{\nu}_{\ell'} \bar{b}$), which are expected to have relatively small background contributions, can now be experimentally investigated [186].

The CMS experiment detected $t\bar{t}$ production through the ℓ +jets decay channel in $p+\text{Pb}$ collisions at a centre-of-mass energy of $\sqrt{s_{\text{NN}}} = 8.16$ TeV [187]. In addition, the CMS Collaboration reported evidence for $t\bar{t}$ production in $\text{Pb}+\text{Pb}$ collisions at $\sqrt{s_{\text{NN}}} = 5.02$ TeV with a significance of 4.0σ using the dilepton channel [188].

This chapter presents a study of $t\bar{t}$ production using $p+\text{Pb}$ collision data collected at a centre-of-mass energy of $\sqrt{s_{\text{NN}}} = 8.16$ TeV with the ATLAS experiment. Top quark pairs are reconstructed in both ℓ +jets and dilepton channels, using final states consisting of electrons, muons and jets.

While the dilepton mode is less abundant compared to the ℓ +jets channels, it has a significantly higher purity. Lower transverse momentum requirements for individual leptons and jets, coupled with improvements in detector calibration accounted for in the data analysis, contribute to a notable improvement in measurement precision compared to the results reported in Ref. [187].

The measurement is compared with calculations up to next-to-next-to-leading order (NNLO) in the strong coupling constant α_s , using the most recent sets of nuclear parton distribution functions (nPDFs).

The chapter is structured as follows: Section 5.2 outlines the data used in the analysis, and Section 5.3 covers the Monte Carlo sampling. The object selection process is detailed in Section 5.4. The background overview is presented in Section 5.5. Section 5.6 explains the event selection for both final states (lepton+jets and dilepton channels). Systematic uncertainties and their treatment are discussed in Section 5.7, while Section 5.8 covers the statistical theory applied in the data analysis. Section 5.9 presents the results of the combined channels, with the cross-section and R_{pA} results presented in Sections 5.9.2 and 5.9.3. Finally, Section 5.10 concludes the chapter.

5.2. Data samples

In 2016, the ATLAS detector recorded data from p +Pb collisions with a centre-of-mass energy of $\sqrt{s_{\text{NN}}} = 8.16$ TeV. The data corresponds to an integrated luminosity of 164.6 nb^{-1} and consists of 30 runs: 19 in Pb+ p (107.79 nb^{-1}) and 11 in p +Pb (56.76 nb^{-1}) beam configurations.

The proton and lead beams had energies of 6.5 TeV and 2.56 TeV per nucleon, respectively, resulting in nucleon-nucleon centre-of-mass collision energy of 8.16 TeV and a rapidity boost of ± 0.465 units relative to the ATLAS laboratory frame, depending on the direction of the proton beam.

The data collection included two beam-direction configurations: p +Pb and Pb+ p , with approximately twice the integrated luminosity in the latter configuration (Pb beam in + z direction). The average number of hadronic interactions per bundle crossing was 0.18. The data were collected under stable beam conditions with all sub-detectors operating at full capacity. Poor quality data were screened using the Good Runs List (GRL), which excludes luminosity blocks with subsystem errors.

5.3. Monte Carlo samples

All kinematic distributions needed for both the signal and background processes are simulated by Monte Carlo (MC) event generators. The event selection criteria and the optimisation of the analysis strategy are made based on these MC samples. The MC samples are also used to estimate the experimental and modelling-related systematic uncertainties. All MC samples are normalised to the most accurate theoretical prediction available of the corresponding process cross-section.

MC simulation samples are used to model the expected signal and background distributions. Dedicated simulated samples were generated at $\sqrt{s_{\text{NN}}} = 8.16$ TeV based on the generator setups used for $\sqrt{s_{\text{NN}}} = 13$ TeV analyses. Two isospin combinations were used for pp and pn collisions. Simulated events for signal and background processes for individual isospin combinations were embedded into real-data events collected during the 2016 p +Pb run, resulting in data overlay samples. The embedding technique is commonly used in the ATLAS Heavy Ion measurements involving hard processes to describe the underlying event of a p +Pb collision. All simulation samples were normalized according to the most accurate theoretical cross-sections and k -factor calculations.

Monte Carlo (MC) simulated event samples play a central role in refining analysis methods, assessing signal and background contributions, estimating signal efficiencies and providing predictions for comparison with actual data. All samples are extensively processed by the full ATLAS detector simulation [189], run within the GEANT 4 framework [190]. Signal and simulated background samples are meticulously generated separately for two isospin configurations (proton-proton and proton-neutron). These samples are then embedded in actual p Pb or Pb p data events to ensure accurate modelling of the underlying event (UE). These merged events,

called “data overlay events”, then pass through the identical reconstruction and analysis pipeline as the real data.

Given the minimal discrepancies in cross-sections between the two isospin configurations, typically less than 0.1%, adjustments are made to the rates of the simulated events. This process involves scaling by the mass number $A_{\text{Pb}} = 208$ of the Pb nucleus and by a ratio of the integrated luminosities between the data and the MC simulation.

The top quark mass remains fixed at 172.5 GeV in all top quark samples. The decay of b - and c -flavoured hadrons in samples simulated with POWHEG BOX v2 and MADGRAPH MC generators are managed by the EVTGEN programme [118].

The nominal simulated $t\bar{t}$ sample is generated using the next-to-leading-order (NLO) event generator POWHEG BOX v2, using the NNPDF3.0NLO Parton Distribution Function (PDF) set [191]. It is then linked to the PYTHIA 8.243 generator [192], using the NNPDF2.3LO PDF set [105] and the A14 tune [92] for parton shower and hadronisation modelling. The resummation damping parameter h_{damp} in POWHEG, which governs the matrix element to parton-shower fit and effectively regulates high p_{T} radiation, is set to 1.5 times the top quark mass. All signal samples are normalised using the NNLO+next-to-next-to-leading logarithmic (NNLL) $t\bar{t}$ cross-section prediction from the TOP++ v2 program [193].

Additional $t\bar{t}$ simulation samples are generated to assess the systematic uncertainties associated with the signal modelling. One example uses the POWHEG BOX v2 MC generator coupled with the HERWIG v7.2 parton shower and hadronisation model [101], using the H7.2 default tune [101, 194]. Another example is generated using the MADGRAPH generator in conjunction with the NNPDF3.0NLO PDF set, using the PYTHIA 8 parton shower and hadronisation model. Uncertainties due to the amount of parton shower radiation are evaluated by producing POWHEG BOX v2+PYTHIA 8 samples with an increased cut-off scale for the first gluon emission, represented by the parameter h_{damp} , set to three times the top quark mass.

The backgrounds evaluated by MC simulation originate from W and Z bosons produced in conjunction with jets, single top quark production and diboson production. Events involving Z +jets and W +jets are simulated using the SHERPA v2.2.10 generator [195] in conjunction with the NNPDF3.0NNLO PDF set [191]. These simulations use NLO matrix elements for up to two partons and LO matrix elements for up to four partons, as described in Ref. [196]. The V +jets ($V = W, Z$) samples are normalised to NNLO cross-sections [197] and additionally filtered to categorise the light, c and b quark content, labelled as W +light/ Z +light, W + c / Z + c and W + b / Z + b .

Single top quark production processes in the t and tW channels are simulated using the POWHEG BOX v2 generator [198, 199], with the PDF sets NNPDF3.04fNLO and NNPDF3.0NLO. The parton shower and hadronisation modelling is done using PYTHIA8 with the A14 tune. To deal with interference between the $t\bar{t}$ and tW final states, the diagram removal scheme [200] is used. Diboson production backgrounds (WW , WZ and ZZ) with additional jets are simulated using the SHERPA v2.2.11 generator with the NNPDF3.0NNLO PDF set.

5.4. Object selections

In the ATLAS experiment, particle collisions produce a variety of secondary particles. These interactions are electronically captured by subdetector systems, and reconstruction algorithms are used to identify the particle types and their kinematic properties, such as momentum and direction, from the collected signals.

The decay of a $t\bar{t}$ pair leads to a complex final state containing charged leptons, neutrinos and both light and heavy quarks. After subsequent hadronisation and decay, these particles interact in different sections of the ATLAS detector and are reconstructed using special algorithms. The results of this process are called objects, each of which is characterised by specific signatures. The definition of each object is closely related to the characteristics of the detector; for example, the $|\eta|$ region in which an electron is defined depends on the acceptance of the detector.

This chapter delves into the reconstruction methods for the primary objects of interest in the analysis, encompassing charged leptons (electrons and muons), jets (potentially comprising different quark flavours), and missing transverse momentum (E_T^{miss}).

5.4.1. Leptons

Electron candidates are reconstructed from energy deposits (clusters) within the central region of the electromagnetic calorimeter (EMC), along with corresponding reconstructed tracks from the Inner Detector (ID). The identification of electrons employs thresholds set by the likelihood-based (LH) algorithm, which utilizes probability density functions of signal and background from various discriminant variables [159]. These candidates must satisfy the MediumLH criteria and exhibit a transverse momentum $p_T > 18$ GeV with $|\eta| < 2.5$, excluding the EMC's transition (crack) region $1.37 < |\eta| < 1.52$. Additionally, their transverse energy, denoted as E_T , should surpass 18 GeV. Electrons must also originate from the primary vertex, necessitating constraints on their transverse impact parameter significance $|d_0|/\sigma_{d_0} < 5$ and longitudinal impact parameter $|\Delta z_0 \sin \theta| < 0.5$ mm. To mitigate background contributions from misidentified QCD jets or semileptonic decays of heavy-flavour hadrons, electrons must exhibit isolation within both the EM calorimeter and the inner detector (ID), using the Gradient isolation criterion. In the EM calorimeter, isolation is quantified by summing the transverse energy of topo-clusters within a cone of radius $\Delta R = 0.2$ centred on the electron's direction, divided by its p_T . Meanwhile, in the ID, the isolation variable sums the p_T of tracks within a cone around the electron track, with the cone size adjusted to $\Delta R = \min[0.2, 10/p_T]$.

Muon candidates were required to have a reconstructed track in the muon spectrometer (MS) combined with a track in the inner detector (ID) called “combined muon”, selected with the MuID algorithm. Muon candidates are required to pass Medium ID criteria and have a transverse momentum of $p_T > 18$ GeV and $|\eta| < 2.5$ and to be associated with the primary vertex with a transverse impact parameter significance $|d_0|/\sigma_{d_0} < 3$, and with a longitudinal impact parameter of $|\Delta z_0 \sin \theta| < 0.5$ mm. To reduce background from muons originating from heavy-flavour decays inside jets, muon candidates are required to be isolated using the

track-quality and isolation criteria similar to those applied to electrons, except for the maximum cone size used in the ID isolation, which is 0.3 for muons as opposed to 0.2 for electrons.

5.4.2. Jets

Jets are reconstructed from calorimeter energy deposits [201], employing the anti- k_t algorithm [143, 144] with a radius parameter $R = 0.4$. To account for contributions from the underlying event (UE), the jet kinematics are corrected event-by-event and calibrated using simulations of the calorimeter response [202], along with in situ measurements of the absolute energy scale. These in situ measurements are conducted during pp collisions and cross-calibrated to the p +Pb system. The jets reconstructed using this method are denoted as HI jets. Kinematic variables are computed utilizing these jets. However, the b -tagging information is absent for HI jets, necessitating the incorporation of a second type of jet in the analysis.

The alternative set of jets is reconstructed from particle-flow (PF) objects, which amalgamate data from topological clusters of calorimeter energy deposits and ID tracks [203]. These PF jets are formed utilizing the anti- k_t algorithm with a radius parameter $R = 0.4$ and are calibrated following the same procedure as in scenarios characterized by high pile-up pp collisions at $\sqrt{s_{\text{NN}}} = 13$ TeV [202]. PF jets containing b -hadrons are identified using the DL1r algorithm [204], which employs a multivariate discriminator grounded in deep-learning techniques utilizing track impact parameters and reconstructed secondary vertices. Employing a tagger working point with 85% efficiency (assessed in simulated $t\bar{t}$ events) in pp collisions for identifying b -quark jets stemming from top-quark decays, this corresponds to rejection factors of approximately three against c -quark jets and 40 against light-quark and gluon jets.

HI jets with transverse momentum (p_{T}) exceeding 20 GeV and pseudorapidity ($|\eta|$) less than 2.5 are paired with PF jets within the same event using a geometric criterion based on minimal ΔR . The b -tagging information is transferred from matched PF jets if the angular separation ΔR between a HI jet and a PF jet is less than 0.3. HI jets without a corresponding PF jet are classified as non- b -tagged.

To avoid the double-counting of electron energy deposits as jets, the nearest jet to an electron candidate is excluded if it lies within $\Delta R \leq 0.2$ of the electron. Additionally, to mitigate the contribution of leptons originating from decays of heavy-flavour hadrons within jets, leptons found within $\Delta R \leq 0.4$ of selected jets are discarded, except when the lepton is a muon and the jet possesses fewer than three associated tracks, in which case the jet itself is discarded. This methodology is independently applied to both types of jets before the matching between HI and PF jets.

5.4.3. Missing transverse energy

Conservation of momentum dictates that the vectorial sum of transverse momenta (\vec{p}_{T}) for all outgoing particles must vanish, reflecting the initial zero \vec{p}_{T} in the beam direction. The full azimuthal coverage (φ) of the ATLAS detector in the transverse plane exploits this principle

to identify weakly interacting particles. The missing transverse momentum (MET), denoted as E_T^{miss} , quantifies the vector imbalance between the reconstructed \vec{p}_T (summed over electrons, muons, photons, hadronically decaying τ leptons and small-R jets within the pseudorapidity range of $|\eta| < 4.9$) and the expected zero value. An additional “soft term” accounts for low \vec{p}_T tracks originating from the primary vertex but not associated with any reconstructed object. In this analysis, MET serves as a crucial control variable, providing insight into the overall transverse energy flow of the collision without imposing a specific E_T^{miss} cut on the signal region. MET offers valuable insights into the collision’s energy distribution and serves as an indicator of the neutrino’s transverse momentum (p_T).

$$E_{x(y)}^{\text{miss}} = E_{x(y)}^{\text{miss},e} + E_{x(y)}^{\text{miss},\mu} + E_{x(y)}^{\text{miss},\gamma} + E_{x(y)}^{\text{miss},\tau} + E_{x(y)}^{\text{miss,jets}} + E_{x(y)}^{\text{miss,soft}} \quad (5.1)$$

5.4.4. Overlap removal

The reconstruction of objects previously defined, at times, generates two different objects from the same deposit of energy, which could lead to double counting of energy and objects in a misreconstructed state. To mitigate these issues, the concept of overlap removal between object collections is applied. Following the definition of selections for electrons, muons, and jets, objects are subject to further criteria aimed at avoiding overlap:

- Electrons sharing a track with a selected muon are excluded
- Any heavy ion (HI) jet found within a distance of $\Delta R < 0.2$ cone from a reconstructed electron is excluded
- Electrons within a radius of $\Delta R < 0.4$ from a HI jet are excluded to mitigate the influence of non-prompt leptons
- HI jets with fewer than three tracks and located within $\Delta R < 0.4$ from a muon are excluded
- The selected muon is removed if it is within $\Delta R < 0.4$ of a HI jet which has three or more tracks

5.5. Backgrounds overview

Three distinct event categories are defined, each imposing an additional criterion on the presence of b -jets:

- Events with 0 b -jets
- Events with precisely 1 b -jet
- Events with at least 2 b -jets

Categories (I) and (II) are anticipated to be predominantly governed by background processes, serving as crucial control regions for validating background estimations. In contrast,

Category (III) displays a significant enrichment in signal events, indicative of its potential sensitivity to phenomena of interest or new physics.

The primary sources of the background in all three b -jet categories are W +jets and the false lepton background. The normalisation of the W +jets background is determined using cross-sections from Monte Carlo (MC) simulations. The fake lepton contribution is estimated using a data-driven matrix method (MM) [205] based on the selection of two categories of events with loose and tight lepton selection requirements [206]. Both the shape and normalisation of the spurious lepton backgrounds are estimated from the data by weighting each selected event by the probability of containing a fake lepton. The Z +jet, the single top quark and diboson contributions are taken from MC simulations and normalised to theoretical cross sections.

The main background contributions in the dilepton channel are attributed to the Z +jets and single-top tW processes. In particular, up to 5% of the Z +jets events originate from the $Z \rightarrow \tau^+\tau^-$ process. This fraction increases significantly, reaching about 99% for selected events in the $e\mu$ channel.

In the ℓ +jets channel, the dominant background sources are W +jets events and the fake lepton contribution. The expected signal fractions in the signal regions (SRs) are 21% and 73% in the ℓ +jets $1b$ and $\geq 2b$ regions, respectively. Similarly, in the dilepton channel, the signal fractions are 53% and 91% in the $1b$ and $\geq 2b$ regions, respectively.

The contributions from each source of background are described in detail in the following sections.

5.5.1. W +jets background

The main contributions to the background come from W + jets events and are simulated using SHERPA-2.2.10, which includes $W \rightarrow \ell\nu_\ell$ decay modes with leptons in three flavours (e^\pm , μ^\pm , and τ^\pm). In addition, filters are applied to partons in three configurations to enhance events containing (I) light jets, (II) c quark jets, and (III) b quark jets. The normalisation of the W +jets background is estimated from Monte Carlo (MC) simulations based on the product of the production cross-section and the filter efficiency. In addition, weighted events from MC simulations are used to refine the background estimate.

5.5.2. Non-prompt/fake lepton backgrounds

Non-prompt leptons, hadrons and photons that meet the lepton selection criteria are sources of the non-prompt and misidentified lepton background, commonly referred to as fake-lepton background. The sources of the fake-lepton background are sources of uncertainty in normalization and shape, and they are usually estimated from data. In this measurement, the method applied is called the Matrix Method (MM) [205]. The MM technique is based on the differences in the lepton identification characteristics between prompt, isolated leptons coming from W and Z boson decays (so-called “real lepton”) and leptons that are either not isolated or originating from the misidentification of photons or jets (so-called “fake lepton”). The method relies on the

fact that the number of events selected in each sample may be expressed as a linear combination of the picked counts for the natural and fake leptons. ‘Tight’ and ‘Loose’ selections are used to compute the efficiencies for real and counterfeit leptons. The ‘Tight’ selection consists of the total list of tight lepton identification criteria. The loose lepton selection is obtained from the ‘tight’ selection by loosening the identification requirements and removing the lepton isolation requirements [207, 208].

The lepton efficiencies, ϵ_r , are estimated in MC samples (Z +jets, $t\bar{t}$) as the ratio of the numbers of simulated/prompt leptons passing the ‘Tight’ requirements and simulated/prompt leptons passing the ‘Loose’ requirements. The fake-lepton efficiency is estimated in a control region (CR) with one lepton passing ‘Loose’ identification and isolation requirements, and missing transverse energy $E_T^{\text{miss}} < 20$ GeV. The limited statistical precision in the two-lepton CR is therefore considered when deriving the fake lepton contribution in the central SR for events in the dilepton channel. The real- and fake-lepton lepton efficiencies are evaluated for electrons and muons, bins in p_T , and $|\eta|$, for the event categories with zero, one, and at least two b -tagged jets.

The efficiencies of fake leptons, which range from 15% to 22% for electrons and from 0.5% to 10% for muons, tend to be higher for lower lepton p_T . Data events that meet the criteria of the baseline analysis for ‘Loose’ leptons are weighted based on both prompt and fake lepton efficiencies. To verify the approach, the predictions are compared with data in a control region (CR) characterised by a higher proportion of fake lepton candidates than expected in the main signal region (SR). This comparison shows satisfactory agreement. The CR, defined by implementing the dilepton and ℓ +jets selection criteria while excluding b -tagged jets, is predominantly influenced by W/Z +jets processes.

5.6. Event selection

The event selection process involves the application of a set of criteria aimed at the kinematic properties of the reconstructed objects, designed to match the topology of the signal process.

For a lepton to be considered, it must exceed the specified p_T thresholds for the relevant data collection period and be associated with the appropriate trigger. Events are selected using either single-lepton electron or muon triggers, requiring a minimum transverse momentum (p_T) threshold of 15 GeV [209, 210]. Furthermore, they must contain at least one reconstructed vertex with at least two good-quality charged particle tracks with $p_T > 0.1$ GeV.

5.6.1. lepton+jets channel

The signature of the $t\bar{t} \rightarrow \ell$ +jets channel is characterized by one isolated charged lepton with relatively high transverse momentum, E_T^{miss} —arising from the neutrino from the leptonic W boson decay, two b -jets, and two light jets from the hadronic W boson decay. The transverse mass of the W boson is defined as $m_T^W \equiv \sqrt{2E_T^{\text{miss}} p_T^\ell (1 - \cos \Delta\phi_{E_T^{\text{miss}}, \ell})}$. A requirement of at least four

jets has to be passed. Also, a lepton has to be matched to a trigger lepton. No requirements are imposed on missing transverse energy E_T^{miss} or m_T^W .

5.6.2. Dilepton channel

In the dilepton channel, both W bosons decay into either an electron or a muon together with the corresponding neutrino. Consequently, the event configurations require exactly 2 oppositely charged leptons (either electron or muon) with $E_T > 18$ GeV (for electron) or $p_T > 18$ GeV (for muon) and at least two jets. One of these leptons must correspond to the object indicated by the triggered (HLT) event. For the dilepton invariant mass $m_{\ell\ell}$, it should exceed 15 (45) GeV in the $e\mu$ (ee and $\mu\mu$) channels to reduce contributions from low-mass resonances away from the Monte Carlo (MC) threshold. In addition, it should be away from the Z boson mass, i.e. $|m_{\ell\ell} - m_Z| > 10$ GeV in the ee and $\mu\mu$ channels to mitigate the background from Z +jet events. There are no constraints on the missing transverse energy E_T^{miss} or H_T .

5.6.3. Separation variable study

The following types of scalar sum of transverse momentum ($p_T \equiv |\vec{p}_T|$) associated with the different objects (which also contribute to the E_T^{miss} reconstruction) have been used for various performance studies:

- $H_T^{l,j} = \sum_{j,l} p_{T,j,l}$, sum of all jets and leptons in the ℓ +jets channel
- $H_T^{b,l} = \sum_{b,l} p_{T,b,l}$, sum of only b -jets and all leptons in the dilepton channel
- $H_T^{b,l,\text{MET}} = \sum_{j,l,\text{MET}} p_{T,b,l,\text{MET}}$, sum of only b -jets, all leptons, and MET
- $H_T^{\text{had}} = \sum_j p_{T,j}$, sum of all jets only in the ℓ +jets channel
- $H_T^{\text{all}} = \sum_{j,l,\text{MET}} p_{T,j,l,\text{MET}}$, sum of all jets, leptons, and MET
- H_T^{pseudo} , sum of transverse momenta of all objects forming the pseudo top quarks

The MET carries useful information on the hardness of the collision and is a footprint of the neutrino p_T . Therefore we also defined the $H_T^{b,l,\text{MET}}$ or H_T^{all} to explore the usage in the fit optimization in the dilepton channel. The signal-background separation is computed as:

$$\text{Separation} \equiv \frac{|\mu_{\text{sig}} - \mu_{\text{bg}}|}{\sqrt{\sigma_{\text{sig}}^2 + \sigma_{\text{bg}}^2}}, \quad (5.2)$$

which is designed to behave like a significance, and where σ_{sig} and σ_{bg} are standard deviations of the signal and background distributions and μ_{sig} and μ_{bg} are the means of the distributions.

The separation power for the key studied H_T variables is presented in Table 5.1. Although not the most effective variable, particularly in the dilepton channel where MET is expected to be larger and better separate the signal from the background, the $H_T^{\ell,j}$ variable was ultimately chosen. This is primarily due to technical limitations, as it is not feasible to calculate MET using HI jets as input.

For the control plots and to construct variables such as m_T^W , MET from the PFlow collection is used for matched events. However, this approach would also require all jet-related MET systematics to use PFlow jets as input, which is not correct. It was therefore decided to exclude MET from both the event selection and the fit variable.

Variable	= 1 b -jet	\geq 2 b -jets	Variable	= 1 b -jet	\geq 2 b -jets
H_T^{had}	0.75	0.65	H_T^{had}	0.58	0.36
H_T^{all}	0.81	0.69	H_T^{all}	0.82	0.50
$H_T^{b,\ell}$	0.60	0.62	$H_T^{b,\ell}$	0.56	0.44
$H_T^{b,j}$	0.72	0.62	$H_T^{\ell,j}$	0.36	0.11
$H_T^{b,\ell,\text{MET}}$	0.75	0.73	$H_T^{b,\ell,\text{MET}}$	0.83	0.62
H_T^{pseudo}	0.79	0.66			

Table 5.1. Signal and background separation of the H_T -like variables in the ℓ +jets (left) and dilepton (right) channels.

5.6.4. Yields

The tables showing the event yields in the ℓ +jets and dilepton channels can be found in the tables 5.2 and 5.3. Three event categories are defined based on an additional requirement regarding the number of b -jets: events with 0 b -jets, events with 1 b -jets, and events with at least 2 b -jets. In the categories with 0 and 1 b -jets, background processes dominate, with expected signal-to-background ratios of 0.01 and 0.26 in the ℓ +jets channel, respectively. Conversely, the category with at least 2 b -jets is signal-enhanced.

= 0 <i>b</i> -jet	Yield	= 1 <i>b</i> -jet	Yield
Fakes	2126.2 ± 24.0	Fakes	791.5 ± 13.3
<i>Wlvb</i>	42.8 ± 0.8	<i>Wlvb</i>	81.9 ± 1.2
<i>Wlvc</i>	538.0 ± 7.1	<i>Wlvc</i>	260.9 ± 5.4
<i>Wlv light</i>	1626.2 ± 29.2	<i>Wlv light</i>	205.8 ± 11.7
<i>Zllb</i>	12.1 ± 0.4	<i>Zllb</i>	22.8 ± 0.6
<i>Zllc</i>	79.6 ± 1.7	<i>Zllc</i>	35.4 ± 1.2
<i>Zll light</i>	321.6 ± 9.4	<i>Zll light</i>	52.5 ± 3.5
diboson	2.0 ± 0.0	diboson	0.6 ± 0.0
single top	11.1 ± 0.1	single top	46.5 ± 0.3
<i>t</i> \bar{t}	71.7 ± 0.2	<i>t</i> \bar{t}	403.9 ± 0.5
Data	4795	Data	1804
Prediction	4831	Prediction	1902
Ratio	0.99	Ratio	0.95
Exp. <i>S/B</i>	0.02	Exp. <i>S/B</i>	0.27
Exp. <i>S/(S + B)</i>	0.01	Exp. <i>S/(S + B)</i>	0.21

$\geq 2b$ -jets	Yield
Fakes	140.2 ± 6.1
<i>Wlvb</i>	35.2 ± 0.8
<i>Wlvc</i>	31.4 ± 2.2
<i>Wlv light</i>	13.7 ± 2.3
<i>Zllb</i>	11.1 ± 0.5
<i>Zllc</i>	6.5 ± 0.6
<i>Zll light</i>	2.8 ± 1.3
diboson	0.1 ± 0.0
single top	41.0 ± 0.2
<i>t</i> \bar{t}	753.1 ± 0.6
Data	1034
Prediction	1035
Ratio	1.00
Exp. <i>S/B</i>	2.67
Exp. <i>S/(S + B)</i>	0.73

Table 5.2. Event yields in the ℓ +jets channel with the statistical uncertainties in predictions.

= 0b-jet	Yield	= 1b-jet	Yield
Fakes	8.2 ± 2.2	Fakes	1.9 ± 1.0
$Zllb$	5.4 ± 0.2	$Zll\ b$	10.4 ± 0.3
$Zllc$	43.1 ± 0.9	$Zll\ c$	14.1 ± 0.6
$Zlllight$	221.6 ± 5.3	$Zll\ light$	14.1 ± 1.2
diboson	4.3 ± 0.0	diboson	0.5 ± 0.0
single top	1.3 ± 0.0	single top	4.9 ± 0.1
$t\bar{t}$	9.5 ± 0.1	$t\bar{t}$	52.7 ± 0.2
Data	306	Data	90
Prediction	293	Prediction	99
Ratio	1.04	Ratio	0.91
Exp. S/B	0.03	Exp. S/B	1.15
Exp. $S/(S+B)$	0.03	Exp. $S/(S+B)$	0.53
	$\geq 2b$ -jets	Yield	
	Fakes	0.5 ± 0.6	
	$Zll\ b$	2.6 ± 0.2	
	$Zll\ c$	1.0 ± 0.1	
	$Zll\ light$	0.5 ± 0.1	
	diboson	0.0 ± 0.0	
	single top	2.3 ± 0.1	
	$t\bar{t}$	73.8 ± 0.2	
	Data	97	
	Prediction	81	
	Ratio	1.20	
	Exp. S/B	10.68	
	Exp. $S/(S+B)$	0.91	

Table 5.3. Event yields in the dilepton channel with the statistical uncertainties in prediction.

5.7. Systematic uncertainties

Several sources of systematic uncertainty are accounted for in the analysis that can affect the signal and background yields. These are related to luminosity, the reconstruction and identification of physics objects, and the modelling of signal and background processes. Each systematic uncertainty can affect either the normalization of a process, or the shape of the final distribution used in the profile likelihood fit (Section 5.8.1), or both the normalization and shape of the distribution simultaneously. The effect of the systematics is evaluated as $\pm 1\sigma$ variations around the nominal value of a given quantity. The systematic uncertainties considered in the analysis are described in the following.

Systematic uncertainties affecting the measurement arise from the reconstruction of leptons and jets, b -tagging, fake-lepton background, the signal and background modelling, and integrated luminosity. In total, 137 systematic uncertainties have been processed, as detailed in Table 5.4, depicting both their grouped and individual contributions.

5.7.1. Luminosity

The uncertainty in the integrated luminosity of the combined data sample is 2.4% [211]. It is derived from the calibration of the luminosity scale using x - y beam-separation scans, following a methodology similar to that detailed in Ref. [212], and using the LUCID-2 detector for the baseline luminosity measurements [83].

5.7.2. Modelling systematics

Uncertainties due to the choice of the parton-shower and hadronisation models in addition to the matrix-element matching to the parton shower are estimated by using the alternative $t\bar{t}$ MC samples. The uncertainty due to initial-state radiation (ISR) is estimated by variations of α_s for ISR in the A14 tune [92]. Further effects on the ISR are evaluated by varying the renormalisation (μ_r) and factorisation scales (μ_f) in the matrix-element calculation as well as the h_{damp} parameter. The μ_r and μ_f are varied independently by factors of 0.5 and 2.0 avoiding same-side variations of the scales. The effect of final-state radiation (FSR) uncertainties is evaluated by modifying the μ_r for emissions from the parton shower by factors of 0.5 and 2.0.

The PDF uncertainties affecting the $t\bar{t}$ signal are evaluated using the PDF4LHC15 Hessian uncertainties [102], using a set of 30 components. The PDF variations are propagated by using alternative MC generator weights corresponding to the PDF4LHC15 variations.

5.7.3. Background systematics

V +jets

The normalization systematics of the W +jet samples are evaluated as described in [213–215] by summing in quadrature a baseline 5% uncertainty with an additional 24% uncertainty per jet, added in quadrature up to a jet multiplicity of 5. This is performed as an additional event

Systematic uncertainty	Number of components
Luminosity	
Luminosity	1
$t\bar{t}$ modelling	
$t\bar{t}$ PhH7EG	2
$t\bar{t}$ aMCPy8EG	2
$t\bar{t} h_{\text{damp}}$	2
$t\bar{t}$ PDF	30
Background modelling	
W+jets	6
Z+jets	3
Single top	2
Diboson	1
Fake lepton	1
Electron	
Calibration	3
Scale factors	4
Muon	
Calibration	4
Scale factors	10
Fake lepton	
Fake lepton	7
Shape	4
Jet	
JES	25
JER	9
PF-to-HI matching	2
b -tagging	19

Table 5.4. Summary of the sources of systematic uncertainty considered in the fit, some of the systematic uncertainties are split into several components, as indicated by the number in the rightmost column.

weight on an event-by-event basis. Varied samples are prepared and treated independently in the likelihood fitter for the three different jet flavor-filtered sub-samples and are kept uncorrelated in the 1 and $\geq 2b$ regions but correlated in the e jets and μ jets channels. An additional flat normalization uncertainty of 50% is assigned to the $W + b$ and $W + c$ samples.

In the treatment of the Z +jets background, a normalization uncertainty of 50% is employed for the light jet samples, drawing upon the theoretical framework delineated in [215] and subsequently integrated within the likelihood fitter. Conversely, a universal uncertainty of 100% is attributed to both Z and heavy-flavor jet samples. Importantly, the Z +jet uncertainties manifest correlations across all analyzed regions, comprising both single-lepton channels (ℓ jets) and dilepton channels.

Single-top background

Single-top-quark diagram removal and diagram subtraction variation samples are used to assess the uncertainties from the interference between the $t\bar{t}$ and tW processes [199]. A conservative uncertainty of 9.5% is considered for the normalisation of both the tW and t -channel single-top-quark processes [216].

VV background

The diboson background normalisation is allowed to vary by 50% [217].

Fake-lepton background

Conservative normalisation uncertainties of 100% in the μ jets and 50% in the e jets and the dilepton SRs are imposed as uncorrelated uncertainties.

5.7.4. Lepton systematics

Electrons

Uncertainties associated with electrons arise from the reconstruction, identification, trigger and isolation efficiencies. These efficiencies and uncertainties have been studied using the Tag-and-Probe method with $Z \rightarrow e^+e^-$ decays in 2016 p +Pb data and MC simulation at $\sqrt{s_{\text{NN}}} = 8.16$ TeV as described here Ref. [218]

In total, there are seven components of uncertainties associated with electrons. The recommended egamma calibration uncertainties are used, while scale-factor uncertainties are derived specifically for this analysis.

Muons

Uncertainties in the muon momentum scale and resolution follow those in Ref. [208]. The analysis includes uncertainties in the data-to-MC correction factors applied to simulated samples for muon reconstruction, isolation, track-to-vertex association and trigger efficiencies evaluated using $Z \rightarrow \mu^+\mu^-$ events in p +Pb collisions.

In total, there are 14 uncertainties associated with muons. The recommended 13 TeV pp high-pileup calibration uncertainties are used, while the scale factor uncertainties are derived specifically for this analysis.

Fake lepton

Systematic uncertainties in the estimation of fake-lepton backgrounds in both ℓ +jets and dilepton channels stem from statistical and systematic fluctuations in real- and fake-lepton efficiencies. These uncertainties are evaluated using the Matrix Method (MM) [205] technique. The magnitudes of systematic variations in the normalization of the fake-lepton background are determined based on the agreement between data and prediction in a zero b -tagged jets ($0b$) control region (CR).

Additional shape variations of this background in the ℓ +jets channel are assessed within the $0b$ CR. To derive these variations, all background contributions except for the fake-lepton events are subtracted from the data, and the difference is normalized to the number of fake-lepton events. Subsequently, a ratio is constructed between the subtracted and scaled data and the fake-lepton contribution as a function of the azimuthal angle $\Delta\phi(E_T^{\text{miss}}, \ell)$ between the lepton and MET. This ratio is then fitted using a second-order polynomial.

Shape variations of the fake-lepton background in $1b$ and $\geq 2b$ ℓ +jets signal regions (SRs) are defined as upward and downward fit shape variations, utilizing the uncertainties of the fit parameters. These variations exhibit a range from 0.5 to 3.5 in bins of $\Delta\phi(E_T^{\text{miss}}, \ell)$. Notably, the shape of the $\Delta\phi(E_T^{\text{miss}}, \ell)$ variable is independent of the shape of the fit variable.

5.7.5. Jets systematics

The calibration of the jet energy scale (JES) is used to get the energy of the jet corrected to the particle level. The jet-related uncertainties are derived from in situ studies of the calorimeter response [202] and their application to the jets used in HI data [219] and from comparisons of the simulated response in samples from different generators. The JES uncertainty is decomposed into 18 uncorrelated components, addressing factors such as single-particle response and the influence of jets extending beyond the calorimeter's containment. For the cross-calibration of heavy-ion (HI) jets, a sophisticated approach is adopted, utilizing EMTopo jets as a reference, as described in Ref. [220]. This procedure introduces a dedicated systematic uncertainty on cross-calibration to ensure accurate jet energy measurements.

Ensuring precise measurements of jet energy resolution (JER) is crucial, particularly for the accurate reconstruction of top quarks. To achieve this, the JER in Monte Carlo (MC) simulations is calibrated to match experimental data. The analysis incorporates the recommended uncertainty of the jet energy resolution (JER), which encompasses nine components, addressing jet transverse momentum (p_T) and pseudorapidity (η)-dependent differences between MC simulations and experimental data [221].

Two systematic variations are introduced to assess uncertainties in the HI-PF jet matching process. The first variation involves adjusting the matching distance ΔR to ± 0.1 from the default

value of $\Delta R = 0.3$ for matching b -tagged jets. The second variation addresses events where HI jets lack a PF counterpart, which accounts for 18% of jets in the data and 15% in the signal and background MC samples. In such cases, HI jets are randomly treated as b -tagged based on the light-flavour jet mistag rate [222]. This systematic uncertainty arises from differences in the energy calibration, as HI jets undergo a special calibration optimised for low pile-up environments. However, this systematic variation has a negligible impact on the final result.

The uncertainties considered for the b -tagging scale factors used to weight events arise from statistical uncertainties as well as modelling and experimental systematic uncertainties [223]. These uncertainties are calculated by varying the data-to-MC correction factors within their uncertainties [222, 224, 225], and they are determined separately for b -jets, c -jets and light flavour jets [204]. This process involves a breakdown into 19 components (11 for b -jets, 4 for c -jets and 4 for light-flavour jets) to account for differences between data and simulation. For light flavour jets, these uncertainties depend on p_T and η , while for b and c jets, they depend only on p_T .

5.7.6. Treatment of the uncertainties

The systematic uncertainties corresponding to individual sources are treated as uncorrelated, while their variations over different regions (SR) are treated as correlated. Each systematic uncertainty is associated with a nuisance parameter (NP) governed by a Gaussian prior. A single NP represents the upward and downward variations of a given systematic uncertainty. Different approaches symmetrize the effect of these uncertainties on the analysed distribution. Sources of systematic uncertainty with minimal effects are excluded from the fitting procedure - symmetrization and pruning processes described in this section. In addition, a smoothing technique is used to improve the stability of the systematic variation.

Pruning

Due to various uncertainties, some of which have minimal impact, a pruning process is applied to increase computational efficiency during the fitting process. This pruning is performed separately for shape and normalisation differences between each variation and its corresponding nominal histograms, using a threshold of 0.5% for both shape and normalisation-based pruning. The b tag, lepton scale factor and electron systematics are mostly pruned. PDF, normalisation and a few fake-related systematics are not present in some signal regions and are represented by the grey blocks.

The pruned and retained uncertainties are shown in Appendix 22. Different colours are used to indicate whether the systematic variation aspects are retained or pruned.

Symmetrization

Since a Gaussian distribution constrains the nuisance parameter (NP) within the likelihood function, it mandates that the upward and downward fluctuations of the associated uncertainty maintain symmetry about the nominal value of a histogram bin, thus necessitating symmetriza-

Systematic Uncertainty Group	Symmetrization
Jet energy scale	Two-sided
$t\bar{t}$ generator	One-sided
Fake-lepton background (charge-flipped)	Two-sided (One-Sided)
Background	Two-sided
Muon uncertainty	Two-sided
W +jets	Two-sided
b -tagging	Two-sided
Electron uncertainties	Two-sided
Jet energy resolution	One-sided
$t\bar{t}$ PDF	Two-sided
Instrumental (t diagram subtraction + removal)	One-Sided

Table 5.5. Summary list of systematic uncertainties group (individual component) and their symmetrization methods.

tion. When both upward and downward fluctuations are present, resulting in three-point uncertainties, the two-sided symmetrization option is used; conversely, in scenarios where only a single fluctuation uncertainty is present, resulting in two-point uncertainties, a one-sided symmetrization approach is used.

Table 5.5 provides a comprehensive overview of all systematic uncertainties considered in the analysis.

Two-Sided: This technique adjusts the disparity between the upward and downward fluctuations to achieve symmetry about the nominal value while preserving their original signs. The positive ('+') and negative ('-') signs of the following expression are interpreted as symmetrized upward and downward fluctuations respectively.

$$\text{symmetrised up/down} = \text{nominal} \pm \frac{(\text{up} - \text{down})}{2}, \quad (5.3)$$

One-Sided: In the case of a single variation uncertainty, the variation is mirrored around the nominal value to achieve symmetrized variation.

5.7.7. Comparisons of distributions for important uncertainties

This section presents a comprehensive analysis of the effects of the most important nuisance parameters considered in the fits, as shown in Figures 5.2, 5.3 and 5.4. The main effects on the signal strength parameter μ are shown for each region included in the fit.

The nominal distributions for each sample are shown as black lines. Dashed red and blue lines show the effects of systematic variations before any smoothing or symmetrization is applied. Solid red and blue lines show the effects of the symmetrization algorithms. The fit is then

performed using these symmetrized distributions, which may be one- or two-sided depending on the availability of the sample systematics. The nuisance parameters associated with the fake systematics and the $t\bar{t}$ modelling generator are shown for the fitted bins.

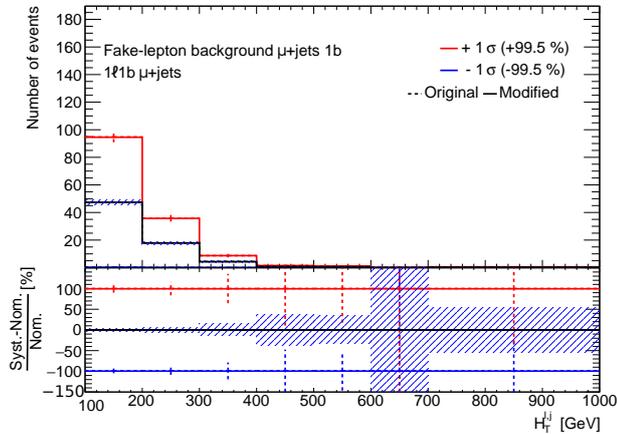


Figure 5.2. Impact of the Fake-lepton background μ +jets with one b -jet systematic uncertainty on the distribution of the $H_T^{l,j}$ in the signal region.

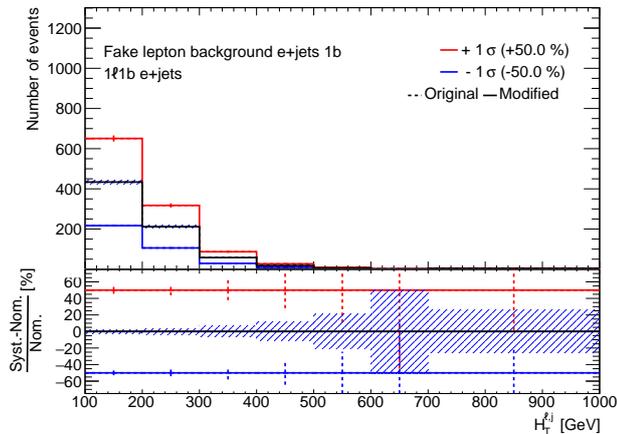


Figure 5.3. Impact of the Fake-lepton background e +jets with one b -jet systematic uncertainty on the distribution of the $H_T^{l,j}$ in the signal region.

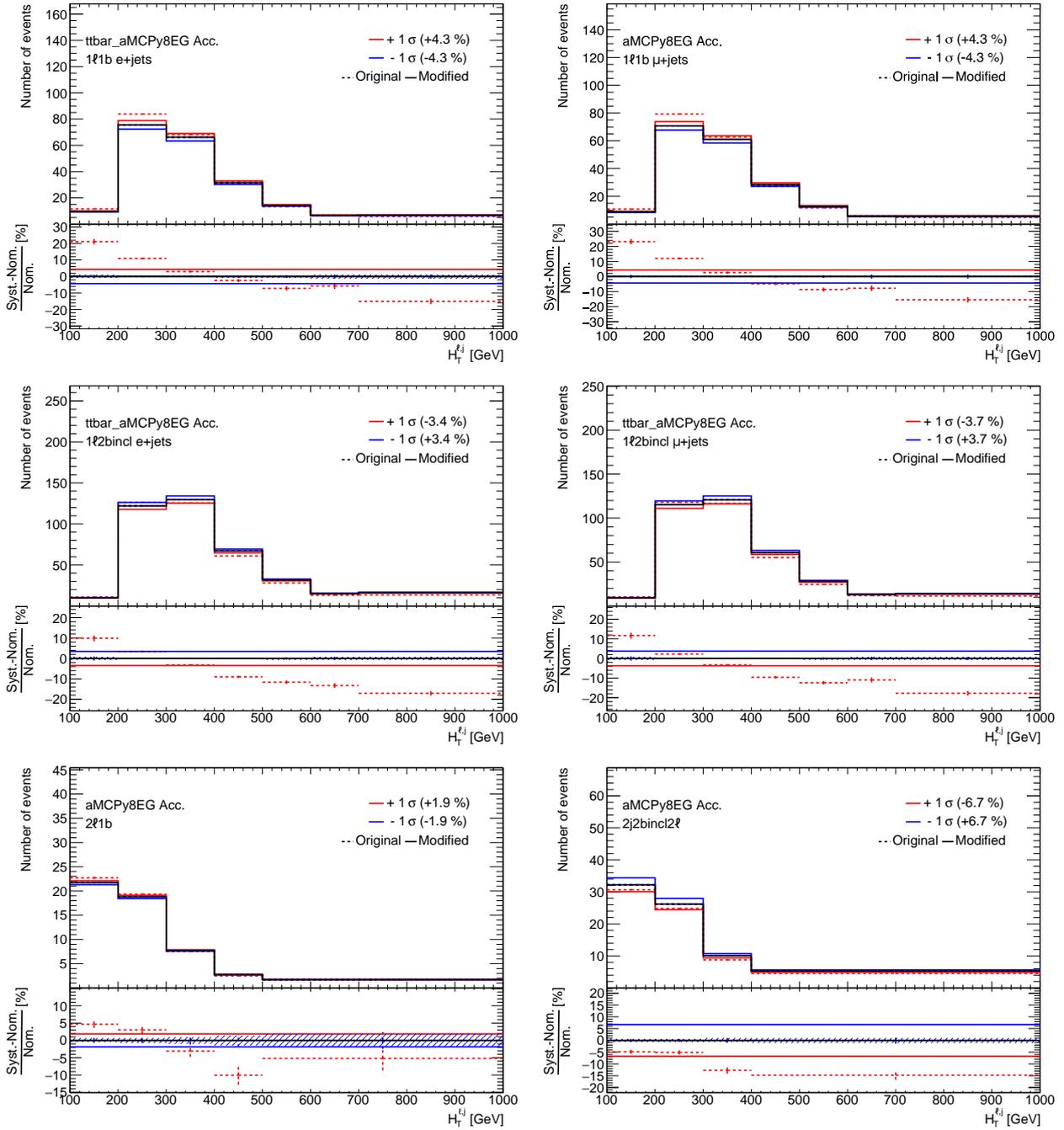


Figure 5.4. Impact of the $t\bar{t}$ acc. aMC@NLO - modelling normalization systematic uncertainty on the distribution of the $H_T^{l,j}$ in the signal region.

5.8. Statistical analysis

Statistical analysis is performed to extract the signal cross-section normalized to the Standard Model's prediction and quantify the expected significance. Then, the signal and background predictions and the observed data are incorporated into a statistical model. This model is implemented using the `TRExFitter` [226] package, based on the `HistFactory` [227] tool in the `Roostats` framework [228]. In the following, the description of the statistical model used in this analysis is presented.

The statistical model is built up with a binned likelihood function. To obtain the cross-section, this likelihood fit is performed for the number of events in the signal regions and the control regions of all sub-channels simultaneously. For a blinded analysis, the number of observed data events in the signal regions is taken from the sum of the expected MC for SM processes. The likelihood is constructed as follows:

$$\mathcal{L} = \prod_{c \in \text{channels}} \prod_{b \in \text{bins}} \text{Poisson} \left(n_{c,b}^{\text{obs}} \mid n_{c,b}^S, n_{c,b}^B \right) \times \prod_{s \in \mathbb{S}} \mathcal{G} \left(0 \mid \theta_s, 1 \right), \quad (5.4)$$

where c stands for the channel index, b is the bin index for each channel. The Poisson terms Poisson are given by:

$$\text{Poisson} \left(n_{c,b}^{\text{obs}} \mid n_{c,b}^S, n_{c,b}^B \right) = \frac{1}{n_{c,b}^{\text{obs}}!} \left(\mu * S_{c,b} + n_{c,b}^B \right)^{n_{c,b}^{\text{obs}}} \exp^{-\left(\mu * S_{c,b} + n_{c,b}^B \right)}, \quad (5.5)$$

where the number of events observed in each bin is marked as $n_{c,b}^{\text{obs}}$, and the expected numbers of signal and background in the corresponding bin is $n_{c,b}^S$ and $n_{c,b}^B$, respectively. The Parameter-of-interest (POI) μ is the signal strength, which is shared among different bins and channels. To represent the nuisance parameter (NP) constraint terms, a Gaussian function, $\mathcal{G}(0 \mid \theta_s, 1)$ is considered, where θ_s is the NP term.

5.8.1. The likelihood function

The maximum-likelihood method is widely used in high-energy particle physics in searches for new processes. The likelihood is defined as the probability of observed data consistent with a given hypothesis. In the search for the $t\bar{t}$ process, the hypothesis is represented by a *signal strength* parameter, μ , defined as the ratio of the observed $t\bar{t}$ cross-section to the Standard Model predicted cross-section:

$$\mu = \frac{\sigma_{t\bar{t}}^{\text{obs}}}{\sigma_{t\bar{t}}^{\text{SM}}}, \quad (5.6)$$

A value of $\mu = 0$ represents the background-only hypothesis (null hypothesis) where no $t\bar{t}$ signal is present, while $\mu = 1$ represents the signal-plus-background (S + B) hypothesis which coincides with the SM Higgs boson production.

Consider a binned distribution of events in the signal region, e.g., the $H_T^{l,j}$ output histogram. The expected number of events in the bin i of the histogram can be expressed as:

$$E[n_i] = \mu \cdot s_i + b_i, \quad (5.7)$$

To define the likelihood function, the ATLAS experiment is assumed to be a counting experiment and the measured data is expected to follow a Poisson probability distribution. Furthermore, each bin of the distribution is statistically independent from the others. The likelihood as a function of the signal strength is then given by:

$$L(\mu) = \prod_i^{N_{\text{bins}}} \frac{(\mu \cdot s_i + b_i)^{n_i}}{n_i!} e^{-(\mu \cdot s_i + b_i)}, \quad (5.8)$$

where n_i is the number of observed data events [229].

The impact of the systematic uncertainties on the signal and background expectations is described by the nuisance parameters (NPs), $\boldsymbol{\theta}$. The NPs are constrained by the probability density functions (PDFs). The choice of the PDF for NP depends on the type of uncertainty associated with it. The Gaussian PDF is a frequent choice for NPs, as it describes the uncertainties on parameters that can be both positive and negative. The nuisance parameter is expected to be distributed as a Gaussian with a width equal to the uncertainty on the parameter, as estimated from auxiliary measurements. However, the Gaussian PDF is not suited for positively defined observables such as luminosity, cross-sections, and efficiencies. The log-normal PDF is used for nuisance parameters associated with these observables, which only affect normalization to ensure that they are always positive. The statistical uncertainty in the simulated background predictions and the fake light lepton background are included as bin-by-bin NPs using the technique described in Ref. [230]. The gamma distribution is used for describing these statistical uncertainties. The PDFs are included in the likelihood function as an additional term. The full likelihood as a function of μ and the set of nuisance parameters $\boldsymbol{\theta}$ is written as:

$$L(\mu, \boldsymbol{\theta}) = \prod_i^{N_{\text{bins}}} \frac{(\mu \cdot s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta}))^{n_i}}{n_i!} e^{-(\mu \cdot s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta}))} \prod_{\theta_k \in \boldsymbol{\theta}} \rho(\theta_k), \quad (5.9)$$

where $\rho(\theta)$ represents the prior PDFs for each nuisance parameter. Thus, the expected number of signal and background events in a given bin is a function of $\boldsymbol{\theta}$.

The signal strength μ is chosen as the parameter of interest (PoI) and represents a free parameter (an unconstrained parameter). The best estimate for μ and $\boldsymbol{\theta}$ is obtained by maximizing the likelihood or equivalently minimizing the negative logarithm of the likelihood (negative log-likelihood).

5.8.2. Test statistic

By creating a test statistic, the compatibility of the observed data with the prediction for the provided hypothesis can be quantified. The likelihood function is defined in equation 5.10, and the test statistics utilized in the study are based on the profile likelihood ratio. As a function of μ , the profile likelihood ratio is represented as:

$$\lambda(\mu) = \frac{L(\mu, \hat{\theta})}{L(\hat{\mu}, \hat{\theta})}, \quad (5.10)$$

where the $\hat{\theta}$ at the numerator denotes the value of θ that maximizes the likelihood for the specified μ , it is called the conditional maximum-likelihood (ML) estimator of θ , and thus it is a function of μ . The $\hat{\mu}$ and $\hat{\theta}$ at the denominator are the unconditional ML estimators of μ and θ , respectively. The presence of the nuisance parameters broadens the profile likelihood as a function of μ relative to what one would have if their values were fixed. This reflects the loss of information about μ due to systematic uncertainties.

The profile likelihood ratio $\lambda(\mu)$ varies between 0 and 1, with $\lambda \approx 1$ signifying good agreement between data and the hypothesized value of μ . The presence of a new signal can only increase the mean event rate beyond the background expectation. Thus, the signal strength μ has physical meaning only when it is positive $\mu > 0$ and any physical estimator for μ must be non-negative. However, it is more suitable to use an effective estimator $\hat{\mu}$ as the value of μ that maximizes the likelihood, which is allowed to take on negative values ($\hat{\mu} < 0$) but providing that the Poisson mean values ($\mu \cdot s + b$) remain non-negative. If data results in $\hat{\mu} < 0$ due to downward fluctuations, then the best agreement between data and any physical value of μ occurs for $\mu = 0$. Therefore, the profile likelihood ratio is modified as:

$$\lambda(\mu) = \begin{cases} \frac{L(\mu, \hat{\theta}(\mu))}{L(\hat{\mu}, \hat{\theta})} & \hat{\mu} \geq 0 \\ \frac{L(\mu, \hat{\theta}(\mu))}{L(0, \hat{\theta}(0))} & \hat{\mu} < 0 \end{cases}, \quad (5.11)$$

where $\hat{\theta}(\mu)$ and $\hat{\theta}(0)$ are the conditional ML estimators of θ given a strength parameter of μ and 0, respectively. It is convenient to define the test statistics as:

$$t_\mu = -2 \ln \lambda(\mu) = \begin{cases} -2 \ln \frac{L(\mu, \hat{\theta}(\mu))}{L(\hat{\mu}, \hat{\theta})} & \hat{\mu} \geq 0 \\ -2 \ln \frac{L(\mu, \hat{\theta}(\mu))}{L(0, \hat{\theta}(0))} & \hat{\mu} < 0 \end{cases}, \quad (5.12)$$

Thus, the larger values of t_μ correspond to increasing incompatibility between data and μ . The level of disagreement is quantified by computing the p -value:

$$p_\mu = \int_{t_{\mu, \text{obs}}}^{\infty} f(t_\mu | \mu) dt_\mu, \quad (5.13)$$

where $t_{\mu, \text{obs}}$ is the value of the test statistics observed from the data and $f(t_\mu | \mu)$ represents the PDF of t_μ under the assumption of μ . For the discovery of a new signal with non-negative strength ($\mu \geq 0$), such as $t\bar{t}$, the statistic t_μ is used to reject the background-only hypothesis, i.e., $\mu = 0$. Using equation 5.12 with $\mu = 0$, the test statistics (q_0) for discovery is defined as:

$$q_0 = \begin{cases} -2 \ln \lambda(0) & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases}, \quad (5.14)$$

where $\lambda(0)$ is the profile likelihood ratio for $\mu = 0$ as defined in equation 5.10. The p -value is computed using the observed value of q_0 similarly as done with t_μ to quantify the level of disagreement between data and the null hypothesis, namely,

$$p_0 = \int_{q_{0, \text{obs}}}^{\infty} f(q_0 | 0) dq_0, \quad (5.15)$$

where $f(q_0 | 0)$ denotes the PDF of the statistic q_0 under the assumption of the background only hypothesis.

In particle physics, when performing searches, it is common to convert the p -value into a significance Z . The significance is defined such that a Gaussian distributed variable which is Z standard deviations above its mean has an upper-tail probability equal to p -value

$$Z = \Phi^{-1}(1 - p), \quad (5.16)$$

where Φ^{-1} is the inverse function of Φ , i.e., the quantile (inverse of the cumulative distribution) of the standard Gaussian. In the particle physics community, the significance required to reject the background-only hypothesis is set to $Z = 5$ as an appropriate level (5σ) to constitute a discovery. This corresponds to a value of $p = 2.87 \cdot 10^{-7}$. If the significance of $Z \geq 3$ is found, evidence is reported for a new process. In order to exclude a signal hypothesis, a p -value of 0.05 (i.e., 95% confidence level) is used that corresponds to $Z = 1.64$ [229].

5.9. Results

This section outlines the analysis performed to determine the signal strength, focusing on both the ℓ +jets and dilepton channels. It also includes a cross-sectional measurement of top quark pair production in proton-lead collisions at 8.16 TeV, accompanied by the nuclear modification factor R_{pA} .

5.9.1. Results of the ℓ +jets and dilepton channels combined fit

The Asimov data set is an artificial representative data set used in particle physics to evaluate the performance of statistical methods and estimators. It is designed to mimic the behaviour of real data and is particularly useful for testing the stability of the fitting procedure and the effect of systematic uncertainties on the measurement. Appendix 5.10 provides a more detailed analysis of the Asimov dataset.

In this section, the results of the combined six SRs (e +jets: $1\ell 1b$ and $1\ell 2bincl$, μ +jets: $1\ell 1b$ and $1\ell 2bincl$, dilepton: $2\ell 1b$ and $2\ell 2bincl$.) are presented. Table 5.7 shows a comparison of the predicted event yields to data in the signal region before fitting the observed data (pre-fit). The dominant backgrounds are W +jets (ℓ +jets), Z +jets (dilepton) and non-prompt and fake lepton backgrounds as shown in Figure 5.5.

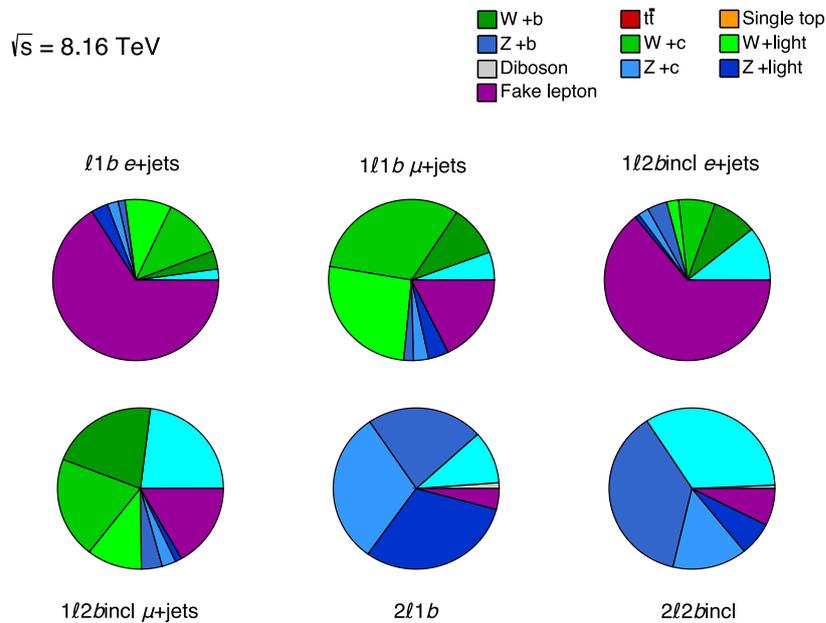


Figure 5.5. Pie plots representing the background composition in the six signal regions.

A maximum-likelihood ¹ fit is performed to extract the signal strength, $\mu_{t\bar{t}}$. The transverse momentum sum of all jets and selected isolated leptons in the event ($H_T^{\ell,j}$) is used as a key discriminating variable for the fit. Figure 5.7 shows the pre-fit distribution of the $H_T^{\ell,j}$ for data and prediction in the signal region.

¹ The method is widely used in high-energy particle physics to search for new processes. Here, the likelihood represents the probability of observed data being consistent with a given hypothesis. The hypothesis is represented by a signal strength parameter, μ , defined as the ratio of the observed cross-section to the Standard Model predicted cross-section.

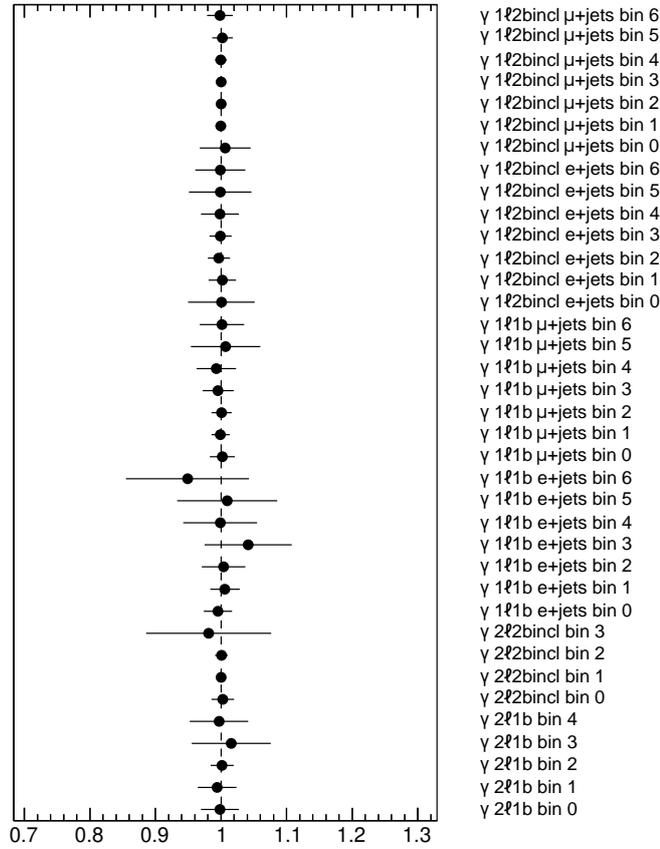


Figure 5.6. γ Parameters after the fit with data.

In the $4j$ regions, this analysis incorporates seven discrete bins, while within the dilepton region, there are four and five distinct bins in the $H_T^{\ell,j}$ distribution. To effectively manage the statistical uncertainties inherent in the MC samples, a number of bins were selected, optimizing the uncertainties' representation in the gamma plot shown in Figure 5.6. In instances of bins with limited event statistics, the bins were merged.

To ensure the chosen variable did not significantly impact the ability to distinguish between signal and background events (separation power), a study was conducted detailed in Section 5.6.3. This study concluded that there was only a minimal difference in separation power between the various H_T variables.

The NP plot presented in Figure 5.9 depicts the integration of all systematic sources into the TRExFitter framework. Notably, substantial constraints are absent except for the fake-related systematics, which is subject to a pre-applied conservative normalization. This approach offers enough flexibility to the fitting procedure, to avoid constraints and provide enough degrees of freedom to the fit.

A value of $\mu = 0$ represents the background-only hypothesis (null hypothesis) where no signal

is present, while $\mu = 1$ represents the signal-plus-background (S+B) hypothesis which coincides with the Standard Model.

The fit uses templates constructed from the predicted event yields in the bins of the input distribution. The systematic uncertainties are treated as nuisance parameters. As described in Section 5.7, several systematic uncertainties affect the final results. In total 140 NPs are considered, most having experimental origin. Systematic uncertainties that have a negligible impact on the final results are discarded to improve the speed of the fit. Normalization or a shape uncertainty is not applied if the associated variation is below 0.5% in all bins. Most of the neglected NPs are related to flavour tagging and the lepton scale factor. Moreover, the correlations between the uncertainties are also considered.

Table 5.8 also compares the data to the yields after the fit to the observed data (post-fit). The post-fit $t\bar{t}$ signal event yield and the uncertainty reflect the best-fit production rate of $1.04_{-0.9}^{+0.9}$ times the SM prediction, and it is obtained using a maximum-likelihood global fit to the data yields of all the signal and control regions of proton-lead collision analysis. Figure 5.8 shows the post-fit $H_T^{\ell,j}$ distribution. The level of agreement is improved in the post-fit as the background predictions were adjusted by the fit. The summary of all six SRs are shown in Figure 5.10.

The observed and expected best-fit values of μ are shown in Table 5.6. The systematics uncertainty contributes more than the statistical component to the overall uncertainty of the measurement. The statistical uncertainty is determined by fixing all NPs to their best-fit values and redoing the fit. The systematic uncertainty is obtained by subtracting in quadrature the statistical uncertainty from the total uncertainty. The expected results are obtained in the same way as the observed results by replacing the data in each bin of the input distribution by the prediction from simulation and the data-driven fake lepton estimates and with all NPs set to their best-fit values obtained from the fit to data.

Best-fit μ	
Expected	$1.00_{-0.035}^{+0.034}$ (stat.) $_{-0.077}^{+0.085}$ (syst.) = $1.0_{-0.085}^{+0.092}$
Observed	$1.04_{-0.035}^{+0.034}$ (stat.) $_{-0.087}^{+0.081}$ (syst.) = $1.04_{-0.094}^{+0.088}$

Table 5.6. Expected and observed best-fit values of the signal strength μ in six combined signal regions.

The background-only hypothesis is rejected with a significance of more than five standard deviations, establishing the observation of the $t\bar{t}$ process in p +Pb collisions by ATLAS. Figure 5.12 presents the signal strength $\mu_{t\bar{t}}$ obtained in each region separately and in the combined fit. The fitted $\mu_{t\bar{t}}$ values in individual channels are consistent within uncertainties and the SM prediction. The precision of the $\mu_{t\bar{t}}$ value is limited by systematic uncertainties in the ℓ +jets SRs while the statistical uncertainties dominate in the dilepton SRs. The significance is extracted using separate fits of $\mu_{t\bar{t}}$ to the combined four ℓ +jets and combined two dilepton SRs, and exceeds in both cases 5 standard deviations. This establishes the observation of $t\bar{t}$ production in the individual ℓ +jets and dilepton channels. The latter is reported for the first time in p +Pb collisions at the LHC.

It is important to assess the impact of each systematic uncertainty on the final result. In particular, to investigate how far the nuisance parameters are pulled away from their nominal values and how their uncertainties are constrained. For this purpose, a ranking of nuisance parameters is introduced. For each nuisance parameter, the fit is performed again with the parameter fixed to its fitted value varied up or down by its fitted uncertainty, with all the other parameters allowed to vary. The ranking obtained for those nuisance parameters with the largest contribution to the uncertainty in the signal strength is shown in Figure 5.11. The top 15 nuisance parameters with the largest pull from the uncertainty are considered. A good agreement is found between the pre-fit and post-fit values of the nuisance parameters and neither large pulls nor large constraints are observed. The systematic uncertainty in the $t\bar{t}$ cross-section due to Fake-lepton background μ +jets has the largest impact on μ . This pull is mainly due to the assigned quite larger fake normalization uncertainties but motivated by actually looking into fake-dominated regions in the independent 0b category, in missing transverse energy especially. Other systematic uncertainties with a large impact are those associated with the signal modelling, the luminosity, and the jet energy scale.

Maintaining a correlation threshold of 30% between NPs, a correlation matrix shown in Figure 5.13 reveals a selected number of moderately large correlations, primarily originating from the interplay between matching-related and fake reweighted systematics. Importantly, all remaining correlations between NPs remain well within acceptable bounds, attesting to the overall stability and controlled nature of the systematic uncertainties.

The observed excess of events is quantified by calculating the background-only p -value using a profile likelihood-ratio test statistic [168], resulting in an observed statistical significance of 16.3 standard deviations for six signal regions combined. The observed significance in the ℓ +jets channel (four regions) is 14.1 standard deviations, while in the dilepton channel (two regions) it amounts to 8.6 standard deviations. The significance is extracted using separate fits of $\mu_{t\bar{t}}$ to the combined four ℓ +jets and combined two dilepton signal regions, each exceeding five standard deviations. This establishes the observation of $t\bar{t}$ production in the individual ℓ +jets and dilepton channels, with the latter reported for the first time in p +Pb collisions at the LHC.

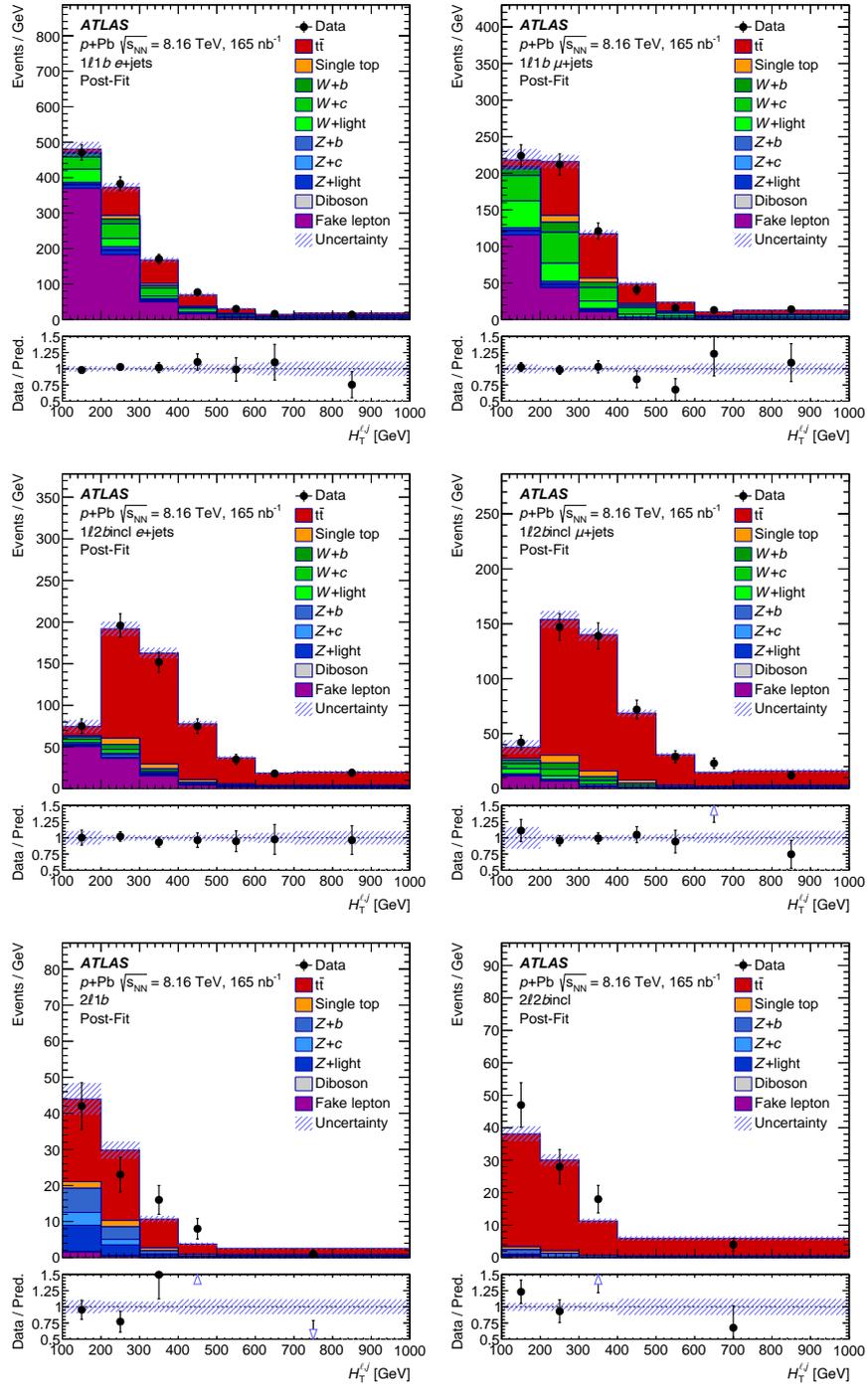


Figure 5.7. Data pre-fit plots representing the $H_T^{\ell,j}$ variable in the six SRs (e +jets: $1l1b$ and $1l2bincl$, μ +jets: $1l1b$ and $1l2bincl$, dilepton: $2l1b$ and $2l2bincl$), with uncertainties represented by the hatched area. The full markers in the bottom panels show a ratio between data and a sum of predictions. Open triangles indicate bins with entries which are outside the ratio range.

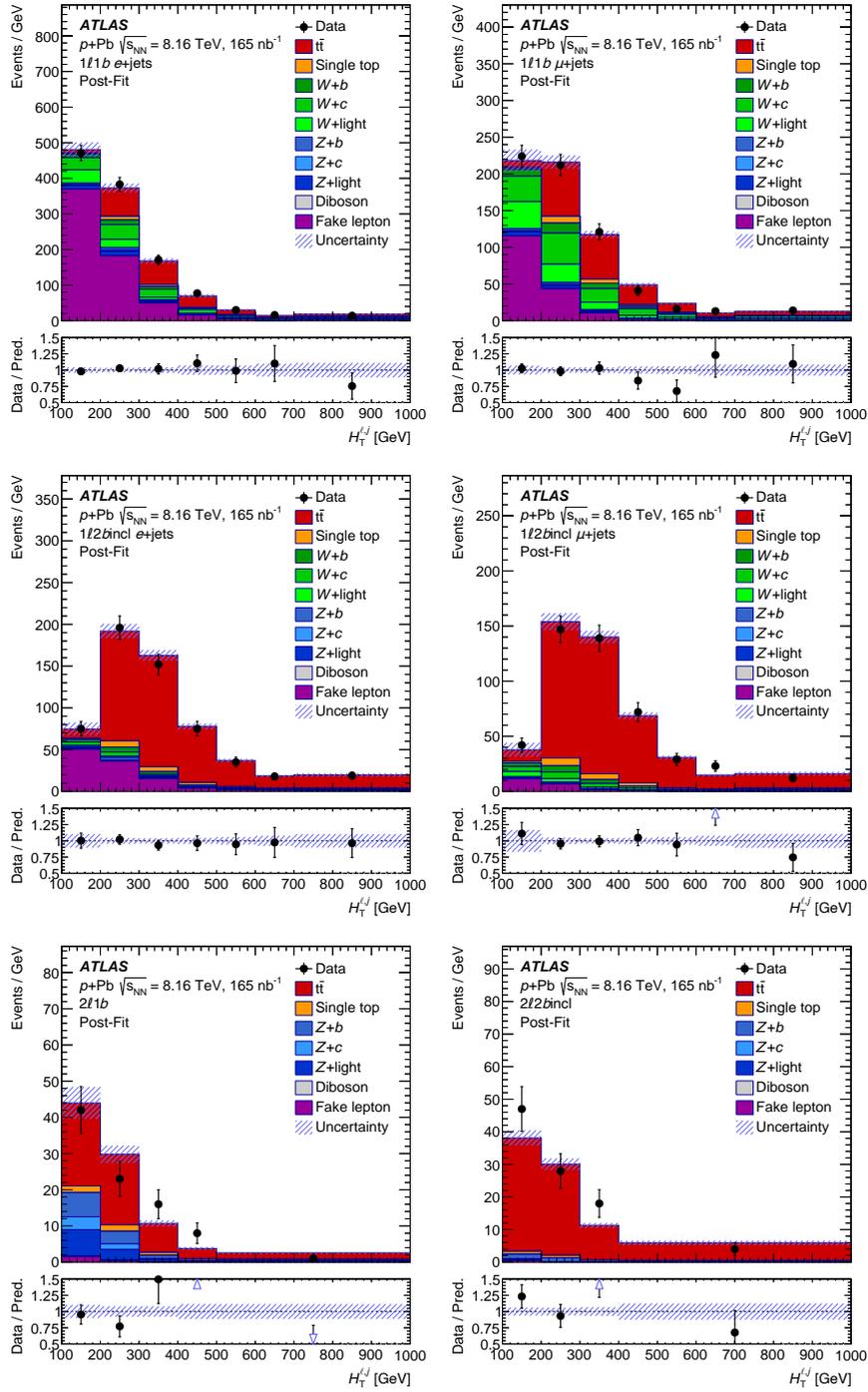


Figure 5.8. Data post-fit plots representing the $H_T^{\ell,j}$ variable in the six SRs ($e+jets$: $1l1b$ and $1l2bincl$, $\mu+jets$: $1l1b$ and $1l2bincl$, dilepton: $2l1b$ and $2l2bincl$), with uncertainties represented by the hatched area. The full markers in the bottom panels show a ratio between data and a sum of predictions. Open triangles indicate bins with entries which are outside the ratio range.

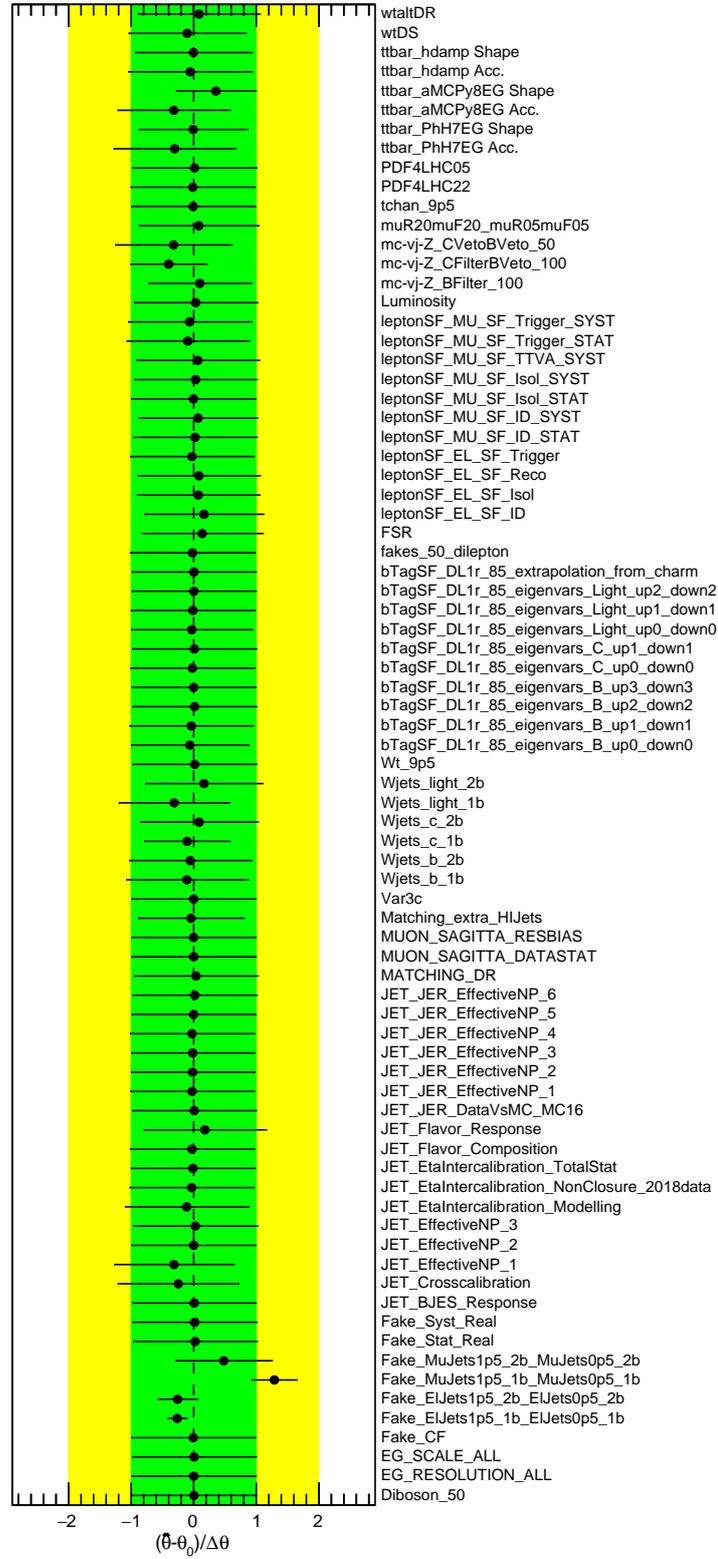


Figure 5.9. The fitted NPs ($\hat{\theta}$) with their pulls and the constraints in terms of their pre-fit values.

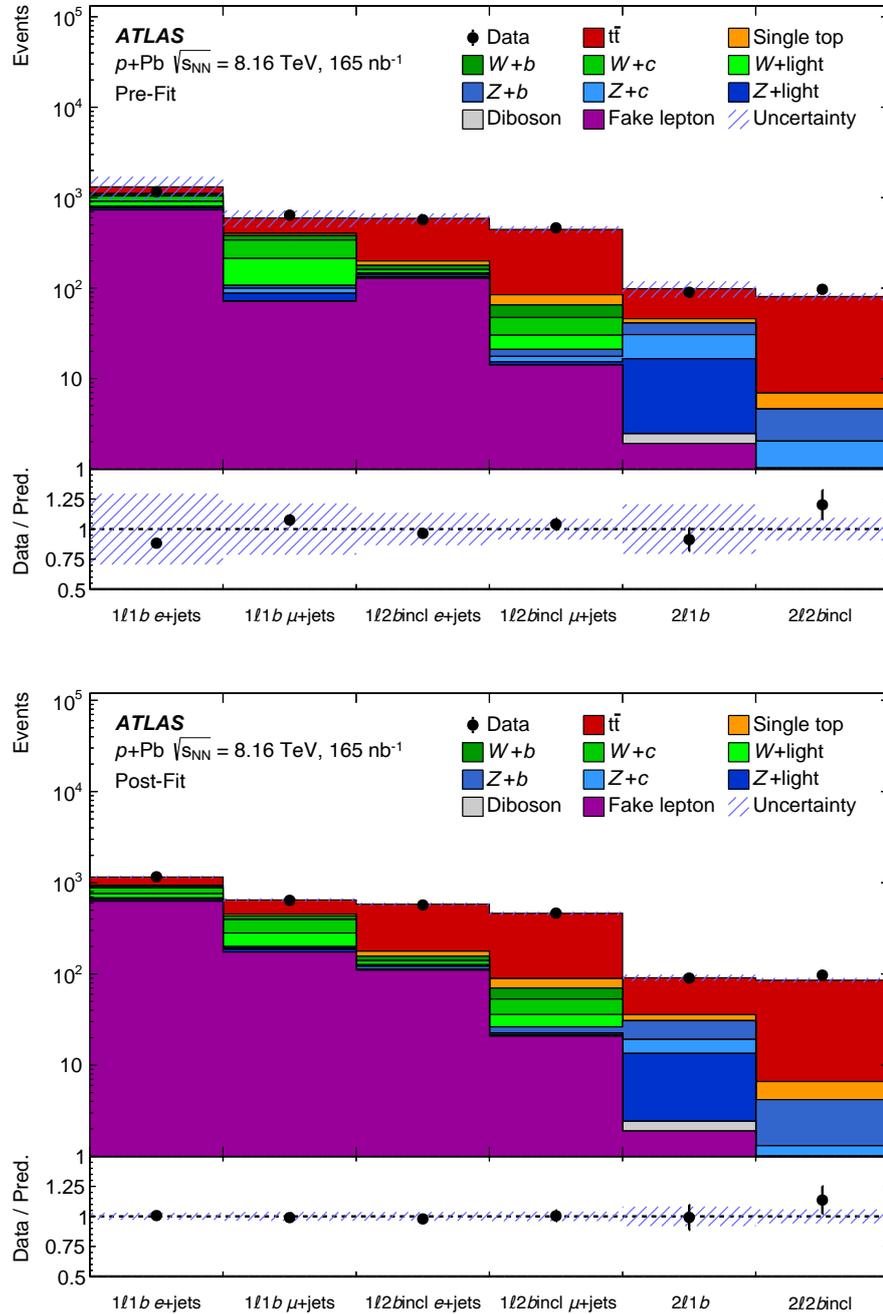


Figure 5.10. Composition of pre- and post-fit events in signal regions for both signal and background, alongside data.

	1 ℓ 1b e+jets	1 ℓ 1b μ +jets	1 ℓ 2bincl e+jets	1 ℓ 2bincl μ +jets	2 ℓ 1b	2 ℓ 2bincl
$t\bar{t}$	211 \pm 18	193 \pm 16	392 \pm 30	361 \pm 27	52.7 \pm 3.5	74 \pm 7
t -channel	7.1 \pm 1.0	6.7 \pm 1.1	7.9 \pm 1.0	7.2 \pm 0.9	0.0014 \pm 0.0034	0.000004 \pm 0.000011
$W + b$	41 \pm 21	41 \pm 21	17 \pm 9	18 \pm 9	0 (0)	0 (0)
$W + c$	130 \pm 70	130 \pm 70	14 \pm 8	17 \pm 9	0 (0)	0 (0)
$W + \text{light}$	100 \pm 60	110 \pm 60	5 \pm 4	9 \pm 6	0 (0)	0 (0)
$Z + b$	15 \pm 15	8 \pm 8	8 \pm 8	3.4 \pm 3.5	10 \pm 10	2.6 \pm 2.6
$Z + c$	24 \pm 24	12 \pm 12	4 \pm 4	2.3 \pm 2.4	14 \pm 14	1.0 \pm 1.0
$Z + \text{light}$	36 \pm 20	16 \pm 9	1.7 \pm 2.4	1.1 \pm 0.8	14 \pm 8	0.5 \pm 0.4
diboson	0.33 \pm 0.17	0.30 \pm 0.15	0.055 \pm 0.030	0.040 \pm 0.022	0.54 \pm 0.28	0.048 \pm 0.026
tW	17.1 \pm 3.1	15.6 \pm 2.9	13.7 \pm 3.4	12.2 \pm 3.1	4.9 \pm 2.0	2.3 \pm 1.2
Fake lepton	700 \pm 400	70 \pm 70	130 \pm 70	14 \pm 14	1.9 \pm 1.4	0.5 \pm 0.9
Total	1300 \pm 400	600 \pm 130	590 \pm 80	450 \pm 40	99 \pm 20	81 \pm 8
Data	1163	641	570	464	90	97

Table 5.7. Data prefit yields of the analysis using 6 regions fit. The total uncertainty is a quadrature sum of statistical and systematic uncertainties.

Table 5.8. Data post-fit yields of the analysis using 6 regions fit. The total uncertainty is a quadrature sum of statistical and systematic uncertainties.

	$1\ell 1b$ e+jets	$1\ell 1b$ μ +jets	$1\ell 2b \text{incl}$ e+jets	$1\ell 2b \text{incl}$ μ +jets	$2\ell 1b$	$2\ell 2b \text{incl}$
$t\bar{t}$	214 ± 24	195 ± 21	405 ± 21	373 ± 19	55 ± 6	79 ± 5
t -channel	6.9 ± 1.0	6.4 ± 1.0	7.8 ± 1.0	7.1 ± 0.9	0.0013 ± 0.001	0.000003 ± 0.000009
$W + b$	37 ± 19	37 ± 19	16 ± 8	16 ± 9	0 (0)	0 (0)
$W + c$	120 ± 40	110 ± 40	14 ± 7	17 ± 8	0 (0)	0 (0)
$W + \text{light}$	80 ± 40	80 ± 40	5.3 ± 3.4	10 ± 5	0 (0)	0 (0)
$Z + b$	16 ± 13	8 ± 7	8 ± 7	3.6 ± 3.1	11 ± 9	2.8 ± 2.4
$Z + c$	9 ± 14	5 ± 7	1.7 ± 2.6	0.9 ± 1.4	6 ± 9	0.4 ± 0.6
$Z + \text{light}$	29 ± 16	13 ± 7	1.4 ± 1.6	1.0 ± 0.6	11 ± 6	0.38 ± 0.28
di-boson	0.32 ± 0.17	0.29 ± 0.15	0.056 ± 0.029	0.039 ± 0.021	0.54 ± 0.28	0.05 ± 0.026
tW	17.1 ± 3.0	15.4 ± 2.7	13.7 ± 3.2	12.2 ± 2.9	5.1 ± 2.0	2.4 ± 1.2
Fake lepton	620 ± 60	170 ± 40	109 ± 19	21 ± 12	1.9 ± 1.0	0.6 ± 0.6
Total	1150 ± 40	648 ± 25	583 ± 21	463 ± 18	91 ± 7	85 ± 5
Data	1163	641	570	464	90	97

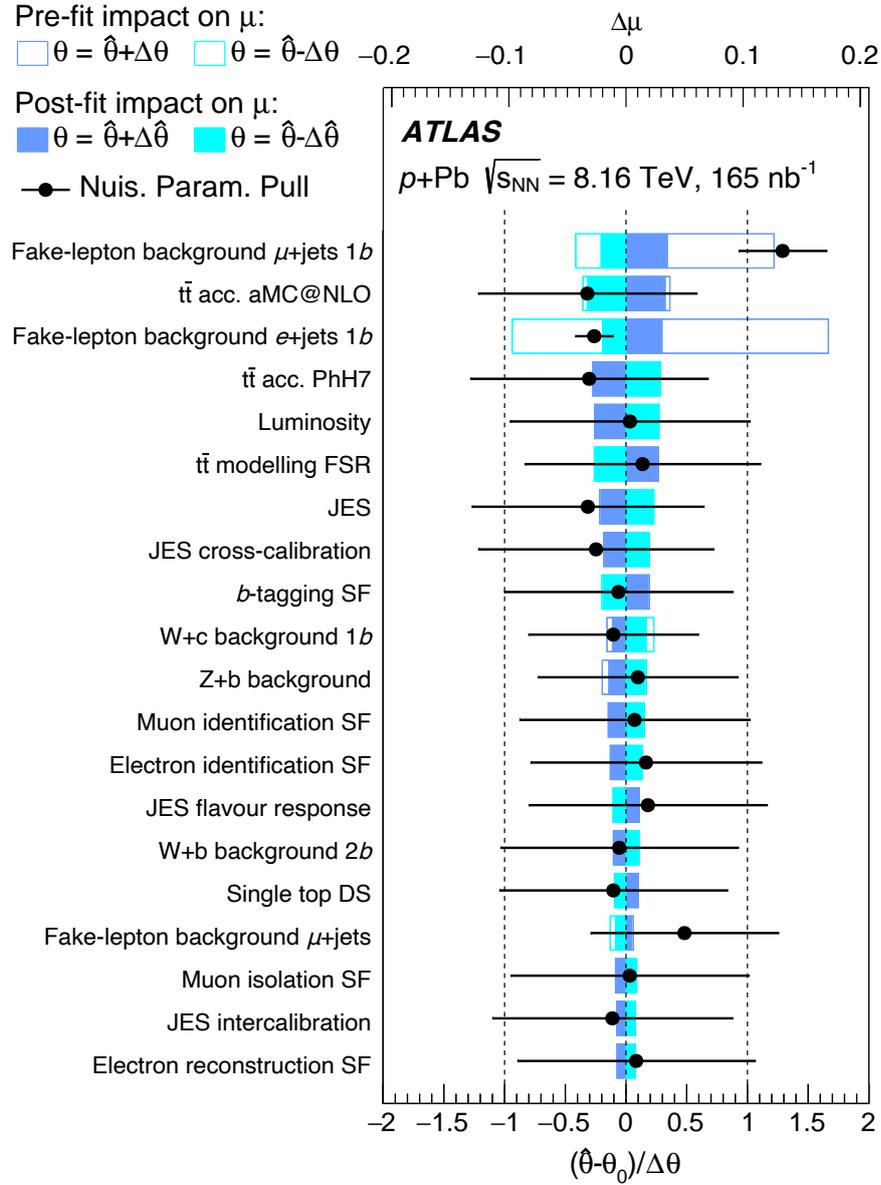


Figure 5.11. The impact of systematic uncertainties on the fitted signal-strength parameter $\hat{\mu}$ for the combined fit of all channels. Only the 15 most significant systematic uncertainties are shown and listed in decreasing order of their impact on μ on the y -axis. The empty (filled) blue/cyan boxes correspond to the pre-fit (post-fit) impact on μ , referring to the upper x -axis. The impact of each systematic uncertainty, $\Delta\mu$, is calculated by comparing the nominal best-fit value of μ with the result of the fit when fixing the corresponding nuisance parameter θ to its best-fit value $\hat{\theta}$ shifted by its pre-fit (post-fit) uncertainties $\hat{\theta} \pm \Delta\theta$ ($\hat{\theta} \pm \Delta\hat{\theta}$). The black points, which refer to the lower x -axis, show the pulls of the fitted nuisance parameters, i.e., the deviations of the fitted parameters $\hat{\theta}$ from their nominal values θ_0 , normalized to their nominal uncertainties $\Delta\theta$. The black lines show the post-fit uncertainties of the nuisance parameters, relative to their nominal uncertainties, which are indicated by the dashed lines.

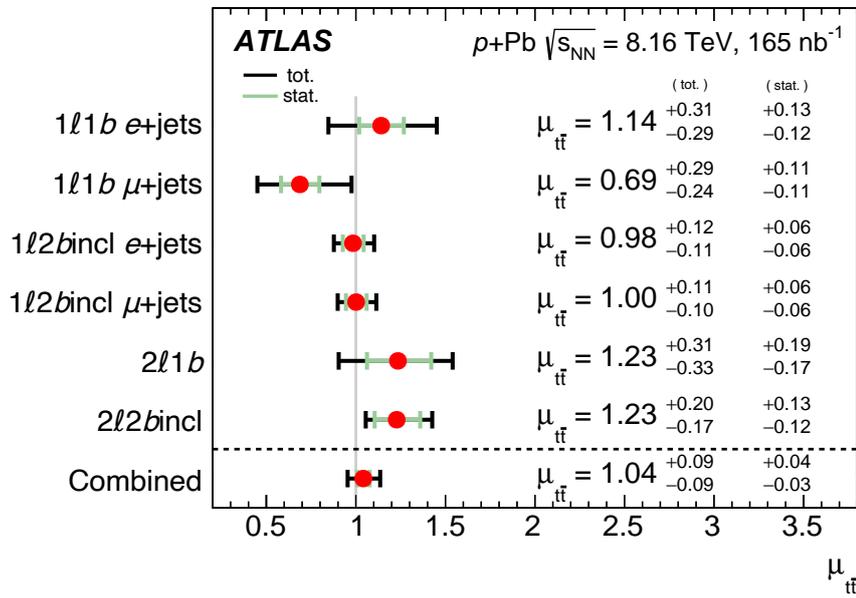


Figure 5.12. The observed best-fit values of the signal strength $\mu_{t\bar{t}}$ and their uncertainties by final-state category and combined. The horizontal green lines represent the statistical uncertainties. The vertical black line at $\mu_{t\bar{t}} = 1$ indicates the SM prediction. The individual $\mu_{t\bar{t}}$ values for the channels are obtained from a simultaneous fit with the signal-strength parameter for each channel floating independently. The SM prediction is $\mu_{t\bar{t}} = 1$.

ATLAS

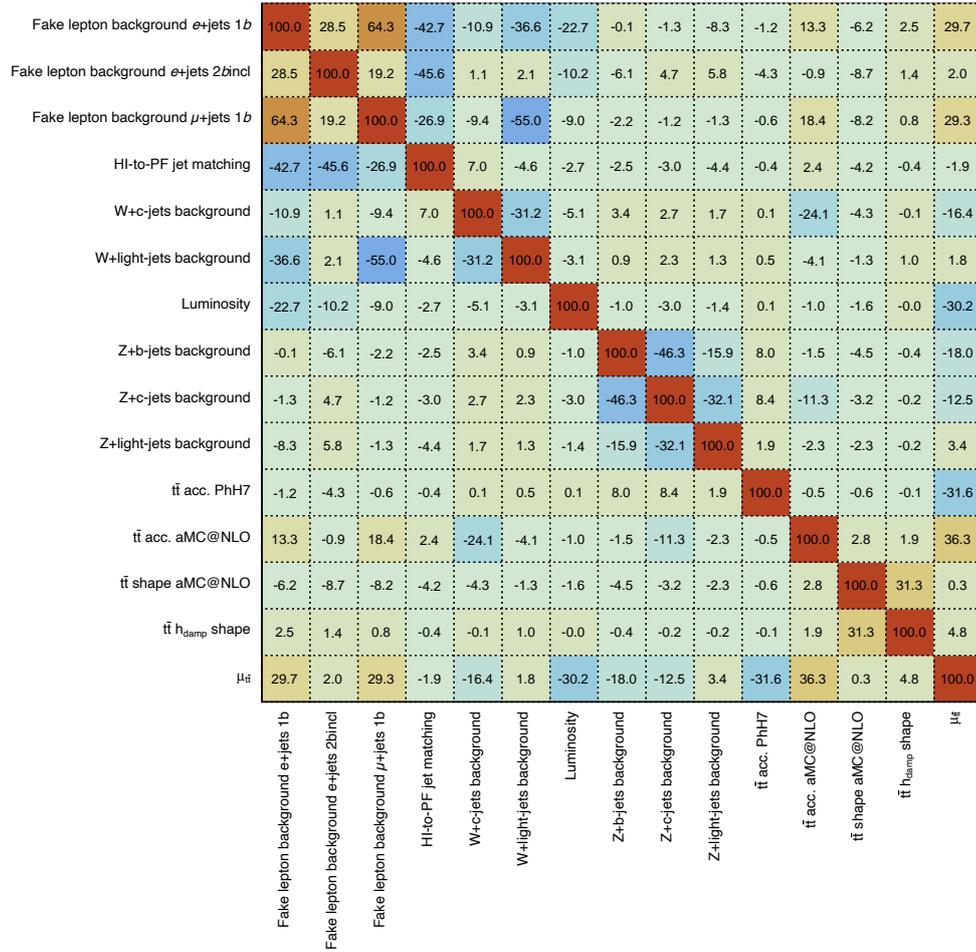


Figure 5.13. Correlation matrix of the fit parameters for the combined fit to data: only parameters with at least one correlation number greater than 30% are shown.

Source	$\Delta\sigma_{t\bar{t}}/\sigma_{t\bar{t}}$	
	unc. up [%]	unc. down [%]
Jet energy scale	+4.6	-4.1
$t\bar{t}$ generator	+4.5	-4.0
Fake-lepton background	+3.1	-2.8
Background	+3.1	-2.6
Luminosity	+2.8	-2.5
Muon uncertainties	+2.3	-2.0
W +jets	+2.2	-2.0
b -tagging	+2.1	-1.9
Electron uncertainties	+1.8	-1.5
MC statistical uncertainties	+1.1	-1.0
Jet energy resolution	+0.4	-0.4
$t\bar{t}$ PDF	+0.1	-0.1
Systematic uncertainty	+8.3	-7.6

Table 5.9. Breakdown of relative systematic uncertainties in the measured $t\bar{t}$ cross-section. The quoted uncertainties are obtained by repeating the fit with a group of nuisance parameters fixed to their fitted values and subtracting in quadrature the resulting total uncertainty from the uncertainty of the complete fit. However, the total uncertainty is not the quadratic sum of the grouped impacts, as this approach neglects the correlation among the different groups.

5.9.2. Cross-section extraction

The signal strength $\mu_{t\bar{t}}$, defined as the ratio of the observed signal for the combined ℓ +jets and dilepton final states to the SM expectation with no nPDF effects included, is measured using a binned profile-likelihood method [227]. The parameter $\mu_{t\bar{t}}$ is determined by the fit to the $H_T^{l,j}$ data distributions in the six SRs, where the $H_T^{l,j}$ variable is defined as the scalar sum of the transverse momenta of the leptons and HI jets.

In the fit, systematic uncertainties are represented by nuisance parameters, which are additional fit parameters constrained by a Gaussian-distributed probability density. By allowing the nuisance parameters to shift from their expected values of zero, the best global fit to the data is achieved. This procedure permits an improved description of the data by combined signal and background contributions, considering their modelling in terms of shapes and normalisation, and the effects of experimental uncertainties, which leads to a reduction of the total systematic uncertainty in the parameter of interest. The measured $\mu_{t\bar{t}}$ value is translated to the inclusive $t\bar{t}$ production cross-section ($\sigma_{t\bar{t}}$) using the formula:

$$\sigma_{t\bar{t}} = \mu_{t\bar{t}} \cdot A_{\text{Pb}} \cdot \sigma_{t\bar{t}}^{\text{th}}, \quad (5.17)$$

where $A_{\text{Pb}} = 208$ is the lead mass number and $\sigma_{t\bar{t}}^{\text{th}}$ is the predicted $t\bar{t}$ production cross-section in nucleon-nucleon collisions derived at the NNLO precision used to normalize the signal $t\bar{t}$ samples in ℓ +jets and dilepton decay modes [193, 231].

The measured inclusive $t\bar{t}$ cross-section for p +Pb collisions is

$$\sigma_{t\bar{t}} = 58.1 \pm 2.0 \text{ (stat.) } {}^{+4.8}_{-4.4} \text{ (syst.) nb} = 58.1 {}^{+5.2}_{-4.9} \text{ (tot.) nb.} \quad (5.18)$$

The combined relative uncertainty amounts to 9% and is dominated by the systematic contribution.

Figure 5.14 shows a comparison of the observed $\sigma_{t\bar{t}}$ with the measurement by CMS in p +Pb collisions at $\sqrt{s_{\text{NN}}} = 8.16$ TeV [187]. The two results are in agreement within 1.4 standard deviations. Also the most precise measurement of the $t\bar{t}$ production cross-section in pp collisions at $\sqrt{s_{\text{NN}}} = 8$ TeV from the ATLAS and CMS combination [232] is shown. The cross-section value is extrapolated to the centre-of-mass energy of this measurement using the TOP++ v2 prediction and scaled by A_{Pb} to the p +Pb system. The extrapolated cross-section has a 2.5% relative uncertainty and does not involve any dependence on nPDF. The extrapolation factor amounts to 1.0528 ± 0.0005 (PDF) ${}^{+0.0001}_{-0.0013}$ (scale). The measured cross-section is also compared with NLO calculations obtained with the MCFM generator [233] scaled to the NNLO precision in QCD using the K -factor ($K = 1.139$) derived using the TOP++ v2 generator. Four nPDF sets are used as input to the MCFM calculations: EPPS21 [235], nCTEQ15HQ [236, 237], nNNPDF30 [238, 239] and TUJU21 [240]. The largest discrepancy is found for the nNNPDF30 nPDF set which does not include the recent Run 2 LHC data for heavy-flavour production from p +Pb collisions [241]. The remaining nPDF sets are in good agreement with the measured cross-section value.

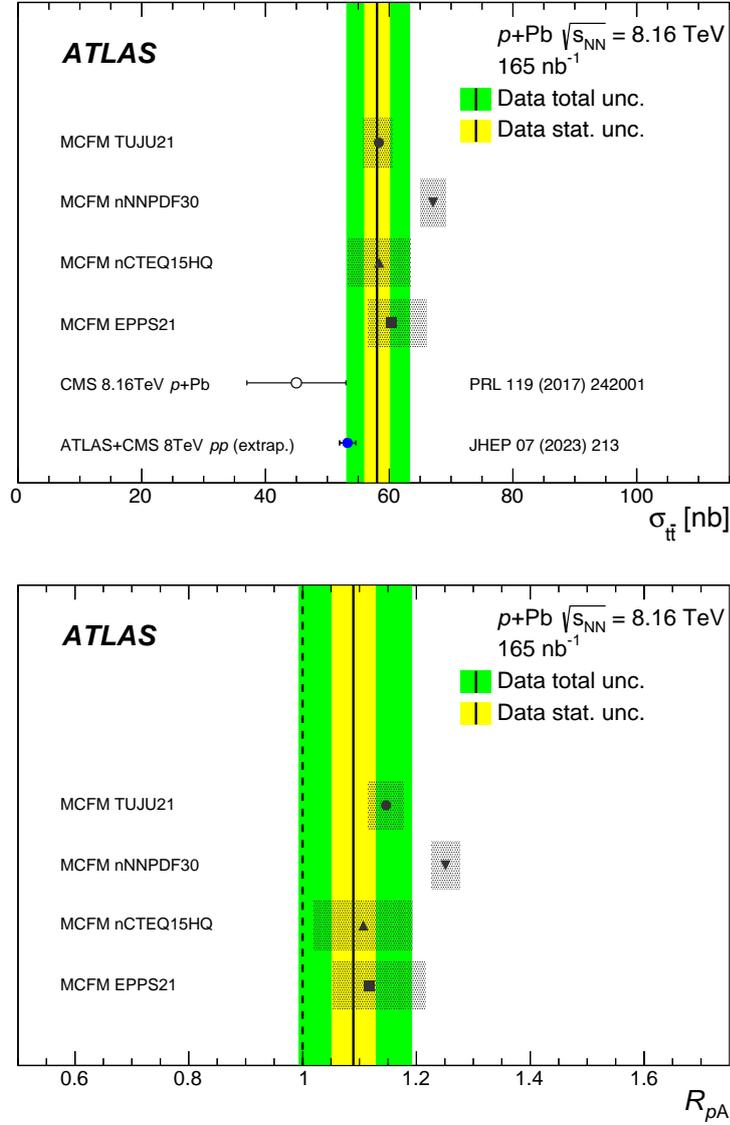


Figure 5.14. Comparison between measured and predicted values of (a) $\sigma_{t\bar{t}}$ and (b) R_{pA} . $\sigma_{t\bar{t}}$ is also compared with the existing measurement in $p+\text{Pb}$ collisions at $\sqrt{s_{\text{NN}}} = 8.16 \text{ TeV}$ [187], and the combined measurement of $t\bar{t}$ production cross-section in pp collisions at $\sqrt{s_{\text{NN}}} = 8 \text{ TeV}$ from ATLAS and CMS collaborations [232]. The latter is extrapolated to the centre-of-mass energy of this measurement and is using the A_{Pb} factor. Predictions are calculated at NNLO precision using the MCFM code [233] scaled to the $p+\text{Pb}$ system and given for different nPDF sets. The uncertainty in the predictions represents the internal PDF uncertainty. The solid black line indicates the measured value. The combined statistical and systematic uncertainty of the measurement is represented by the outer band around the central value, while the statistical component is depicted as the inner band [234].

5.9.3. Nuclear Modification Factor (R_{pA})

During the initial phase of collisions, certain partons undergo scattering with significant momentum transfer as they pass through the produced medium, resulting in energy loss. In Pb-Pb collisions, this phenomenon manifests itself as a suppression of high-pT particles in the final state, a phenomenon known as jet quenching [242, 243], in contrast to the hadron yields expected from an incoherent superposition of pp collisions. The degree of suppression observed in charged hadrons exceeds that observed at RHIC [244], mainly due to the increased energy density achieved at the LHC.

In this context, p +Pb collisions serve as a crucial control experiment, providing insight into whether initial state effects contribute to the observed suppression of hadron production in Pb-Pb collisions. By analysing the nuclear modification factors (R_{pA}) and comparing the identified hadron yields in different collision systems, one can elucidate the particle production mechanisms and discern the impact of nuclear effects.

A nuclear modification factor is defined as

$$R_{pA} = \frac{\sigma_{t\bar{t}}^{p+Pb}}{A_{Pb} \cdot \sigma_{t\bar{t}}^{pp}} , \quad (5.19)$$

is extracted using the measured $t\bar{t}$ cross-sections in p +Pb collisions at $\sqrt{s_{NN}} = 8.16$ TeV and pp collisions at $\sqrt{s_{NN}} = 8$ TeV [232]. The latter is extrapolated to the centre-of-mass energy of the p +Pb system. All uncertainties are assumed to be uncorrelated in the p +Pb and pp cross-section measurements.

The nuclear modification factor is measured to be

$$R_{pA} = 1.090 \pm 0.039 \text{ (stat.) } {}_{-0.087}^{+0.094} \text{ (syst.)} = 1.090 \pm 0.100 \text{ (tot.)} \quad (5.20)$$

Figure 5.14 shows a comparison between the measured R_{pA} in data and theory. Within the uncertainty, the measured value is found to be consistent with unity.

R_{pA} is also calculated at NNLO precision using the MCFM code [233], scaled to the p +Pb system for four different nPDF sets. The uncertainty associated with the baseline PDF for pp interactions is treated as fully correlated in the predictions and cancelled out in the ratio. The resulting uncertainty reflects the uncertainty in the nPDF. All nPDF calculations give R_{pA} values greater than unity. There is good agreement between the measured and predicted R_{pA} . The largest deviation, more than one standard deviation above the measured R_{pA} value, is observed for the nNNPDF30 prediction.

5.10. Summary

This study reports a measurement of top-quark pair production in p +Pb collisions at the centre-of-mass energy $\sqrt{s_{\text{NN}}} = 8.16$ TeV per nucleon pair with the ATLAS experiment. Top-quark pairs are observed in the individual ℓ +jets and dilepton channels with electrons and muons in the final state. The top-quark pair production in the dilepton channel is observed with a significance exceeding 5 standard deviations for the first time in the p +Pb system at the LHC. The cross-section is measured with a relative uncertainty of 9% which makes this measurement the most precise $t\bar{t}$ cross-section determination in nuclear collisions to date.

The measured cross-section is found to be in good agreement with the previous measurement and SM predictions. A measurement of the nuclear modification factor is reported using a NNLO pQCD-based extrapolation of the previously published pp data at $\sqrt{s_{\text{NN}}} = 8.16$ TeV. A good agreement is found between the measured and predicted R_{pA} values involving most of the state-of-the-art nPDF sets. The largest discrepancy of more than one sigma is identified for the nNNPDF30 set. This measurement paves a new way to constrain nPDFs in the high Bjorken- x region.

Conclusion and outlook

A fraction of a second after the Big Bang, the Universe was a hot, dense plasma of quarks and gluons (QGP), a state recreated at the LHC by colliding heavy nuclei. The top quarks produced in these collisions play a crucial role in exploring the properties of QGP. As the heaviest particle in the Standard Model (172 GeV), top quarks are produced only during the initial hard scattering, providing insights into the dynamics of the early universe. Studying their production in proton-proton collisions helps to refine our understanding of the gluon content of parton distribution functions (PDFs) in the unique nuclear environment, facilitating comparisons with proton-proton collisions.

The top quark, the heaviest particle in the Standard Model, interacts strongly with the Higgs boson. This interaction, known as the “top quark Yukawa coupling”, is crucial because of the Higgs boson’s stronger affinity for heavier particles. The direct observation of processes such as $t\bar{t}H$ production offers new perspectives on the Higgs mechanism and possible deviations from expected behaviour, which could point to physics beyond the Standard Model.

$t\bar{t}H$ production, in which a pair of top quarks emits a Higgs boson, is a rare event but crucial for understanding fundamental particle interactions. The low production rate of $t\bar{t}H$, which accounts for only 1% of the total Higgs boson rate, poses significant challenges for detection. Using data collected at 13 TeV between 2015 and 2017, ATLAS conducted multiple searches targeting different Higgs boson decay modes, such as decays into pairs of W bosons or τ leptons, which occur with probabilities of 22% and 6%, respectively. Detection methods included searching for events with same-sign charged leptons or multiple charged leptons, resulting in analyses focusing on different final lepton combinations.

This thesis presents results from the analysis of $t\bar{t}H$ and $t\bar{t}W$ production in multilepton final states collected by the ATLAS detector at a centre-of-mass energy of $\sqrt{s} = 13$ TeV. The observed production cross section $\sigma_{t\bar{t}H}$ was measured to be 294^{+182}_{-162} fb, in agreement with the Standard Model prediction of 507^{+35}_{-50} fb, with a significance of 3.1 standard deviations.

The observed production rate of the $t\bar{t}W$ background process in the selected phase space exceeds the Standard Model expectation by 1.3 to 1.7 times, as validated by several analyses and confirmed by the CMS experiment. This discrepancy suggests the potential influence of Beyond the Standard Model (BSM) physics, such as the production of charged Higgs, and requires further comprehensive studies to explain the observed increase.

The thesis also sheds light on the proton-lead collision, focusing on the prompt decay of top quarks that bypass hadronization. These quarks typically decay into a W boson and a b quark due to their short lifetimes. The W boson then decays into either a pair of lighter quarks or

a lepton and a neutrino. When two top quarks are produced, two dominant signals appear in the detector: one with leptons and jets (lepton+jet) and another with only leptons (dilepton) accompanied by b -quarks.

In this work, proton collisions with heavy lead ions at 8.16 TeV are studied. Top quark production was observed separately in the lepton+jets and dilepton channels, as well as in the combined data set with a significance well above 5 standard deviations. This is the first observation of top quark pair production in the dilepton channel in proton-lead collisions. Combining both channels, the top quark pair production cross section is measured to be $\sigma_{t\bar{t}} = 58.1 \pm 2.0$ (stat.) ${}_{-4.4}^{+4.8}$ (syst.) nb, with a total relative uncertainty of 9%, making this the most precise measurement of the $t\bar{t}$ cross-section in nuclear collisions to date. The total uncertainty is dominated by systematics in the lepton+jets channel and by statistical precision in the dilepton channel. The high precision was achieved by using a tagging technique to distinguish b -quark jets and more comprehensive lepton selection criteria.

This result confirms and extends an earlier observation made by the CMS collaboration into a new decay channel. The findings open avenues for physicists to investigate Parton Distribution Functions, which detail how a proton's momentum distributes among its constituent quarks and gluons, within a novel kinematic domain.

List of papers

This thesis is based on the following papers, to which the author directly contributed.

- [1] **ATLAS Collaboration.** (2019). Analysis of $t\bar{t}H$ and $t\bar{t}W$ production in multilepton final states with the ATLAS detector. *ATLAS-CONF-2019-045*. DOI: [10.22323/1.345.6789](https://doi.org/10.22323/1.345.6789)
- [2] **Mondal, S.** (2020). Tau-lepton Fake-Rate determination for the $t\bar{t}H$ coupling measurement using the ATLAS detector at the LHC. *PoS ICHEP2020*, 100. DOI: [10.22323/1.390.0100](https://doi.org/10.22323/1.390.0100)
- [3] **ATLAS Collaboration.** (2023). Observation of $t\bar{t}$ production in lepton+jets and dilepton channels in $p+Pb$ collisions at $\sqrt{s_{NN}} = 8.16\text{TeV}$ with the ATLAS detector. *ATLAS-CONF-2023-063*. DOI: [10.22323/1.345.6789](https://doi.org/10.22323/1.345.6789)
- [4] **ATLAS Collaboration.** (2023). ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset. *The European Physical Journal C*, **83**(7), 681. DOI: [10.1140/epjc/s10052-023-11699-1](https://doi.org/10.1140/epjc/s10052-023-11699-1)
- [5] **ATLAS Collaboration.** (2024). Observation of $t\bar{t}$ production in the lepton+jets and dilepton channels in $p+Pb$ collisions at $\sqrt{s_{NN}} = 8.16\text{ TeV}$ with the ATLAS detector. *arXiv:2405.05078 [nucl-ex]*. URL: [arXiv.org/abs/2405.05078](https://arxiv.org/abs/2405.05078)
Submitted to Journal of High Energy Physics (JHEP)

Conference contributions

PLENARY TALK

Top production in Heavy Ion (PbPb and pPb) collisions

UNIVERSITY OF LISBON, LISBON, PORTUGAL

ATLAS Week 2022
2022

INVITED TALK

TRExFitter: A framework for binned template profile likelihood fits in ATLAS

POLISH ACADEMY OF SCIENCES, AGH UNIVERSITY OF SCIENCE AND TECHNOLOGY, KRAKÓW, POLAND

New Vistas in Photon Physics in Heavy-Ion Collisions
2022

Fake-Rate Determination for the $t\bar{t}H$ Coupling Measurement with a Signature of Two Same Electric Charge Light Leptons Associated with a Tau Using the ATLAS Detector at the LHC

RWTH AACHEN UNIVERSITY, AACHEN, GERMANY

DPG Conference Aachen
2019

ATLAS TALK

Towards the observation of top-quark pair production in proton-lead collisions at 8.16 TeV

INSTITUTO DE FÍSICA CORPUSCULAR (IFIC): VALENCIA, SPAIN

ATLAS top workshop 2022
2022

Towards the observation of top-quark pair production in proton-lead collisions at 8.16 TeV

CZECH REPUBLIC - SLOVAKIA

CZ+SK HEP Workshop
2019, 2023, 2024

POSTER

Observation of top-quark pair production in p +Pb collisions in the ATLAS experiment

PRAGUE, CZECH REPUBLIC

42th International Conference on High Energy Physics (ICHEP)
2024

Fake-Rate determination for the $t\bar{t}H$ coupling measurement with a signature of two same electric charge light leptons associated with a tau using the ATLAS detector at the LHC

VIRTUAL

40th International Conference on High Energy Physics (ICHEP)
2020

Media appearance

- **Press release:** <https://atlas.cern/Updates/Briefing/HI-Top-Observation>
- **Video briefing:** <https://videos.cern.ch/record/2298651>
- **Facebook:** <https://www.facebook.com/ATLASexperiment/videos/827977738818553>
- **Twitter/X:** <https://x.com/i/status/1707105631910396348>

Auxiliary plots: $p+\text{Pb}$

Asimov fit results

The fit to Asimov’s data shows that the nuisance parameters (NPs) are fitted to a mean of 0. Ideally, when the signal strength is $\mu = 1$, no shifts in the nuisance parameters (NPs) are expected. However, they can still be constrained, with $\Delta\hat{\theta}$ being less than 1 standard deviation.

Figures 16 and 17 display the pre-fit and post-fit distributions of the variable fitted to the Asimov data set.

The uncertainty in the fitted γ factors is primarily due to Monte Carlo (MC) statistical effects. These γ factors are fitted close to 1, and consequently, their uncertainties are generally negligible (Figure 18).

Only a few nuisance parameters (NPs), such as those related to signal modelling, normalisation, and fake-related systematics, show constraints². Figure 19 illustrates the pull (shift of $\hat{\theta}$ from 0) and the constraints of all NPs. This indicates that the expected number of events in each bin of the fitted distribution had larger a priori systematic uncertainties. Alternatively, these uncertainties could be constrained through (anti-)correlation with other systematic uncertainties.

Fitting the Asimov data results in the NPs also being correlated with each other. The correlation between the NPs among themselves and also with the POI. The correlation with at least one other NP or POI is greater than 30% is shown in Figure 20. The significant correlations are mainly between jet matches and fake μ jets in the $1b$ region, followed by W +jets systematics with the same mentioned.

The most important systematics influencing the final uncertainties are ranked in Figure 21. Modelling, fake systematics and luminosity are among the most important contributing systematics. Table 10 shows the Asimov POI (μ_{tt}) value extracted from the fit.

²In these figures, the green region represents $\Delta\hat{\theta} < 1$ standard deviation, while the yellow region represents $\Delta\hat{\theta} < 2$ standard deviations, both centred around a mean of zero.

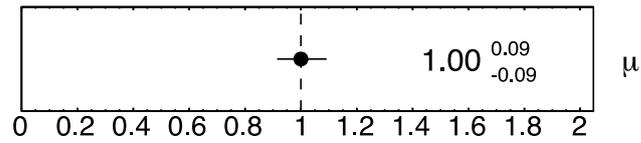


Figure 15. Asimov best-fit signal strength value.

Fit Result	Value
Total Uncertainties	$\mu = 1.000^{+0.092}_{-0.085}$
Statistical Uncertainties	$\mu = 1.000^{+0.035}_{-0.035}$

Table 10. Asimov fit results with total and statistical uncertainties.

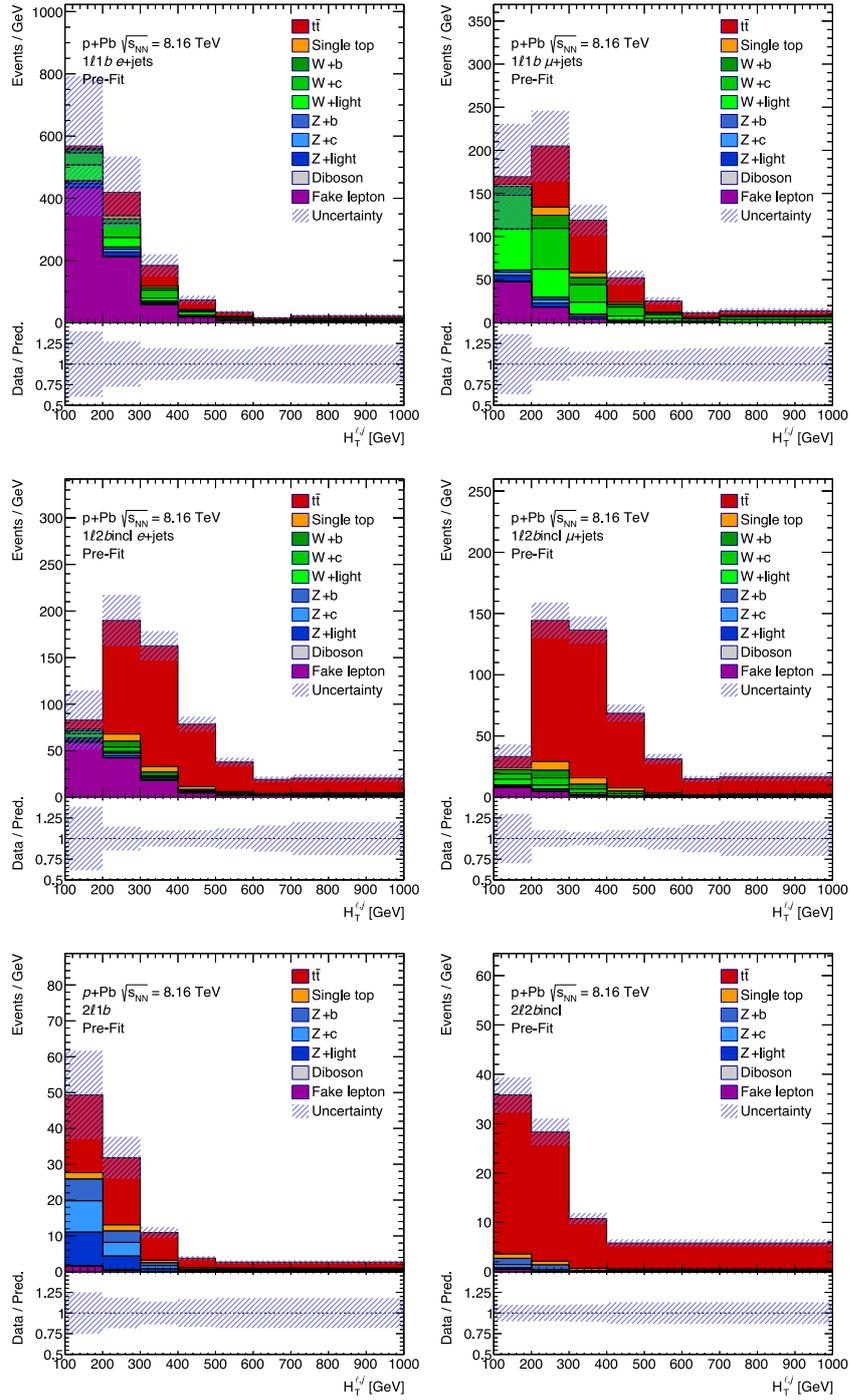


Figure 16. Asimov pre-fit plots representing the $H_T^{\ell,j}$ variable in the six SRs (e +jets: $1\ell 1b$ and $1\ell 2bincl$, μ +jets: $1\ell 1b$ and $1\ell 2bincl$, dilepton: $2\ell 1b$ and $2\ell 2bincl$), with uncertainties represented by the hatched area. The full markers in the bottom panels show a ratio between data and a sum of predictions. Open triangles indicate bins with entries which are outside the ratio range.

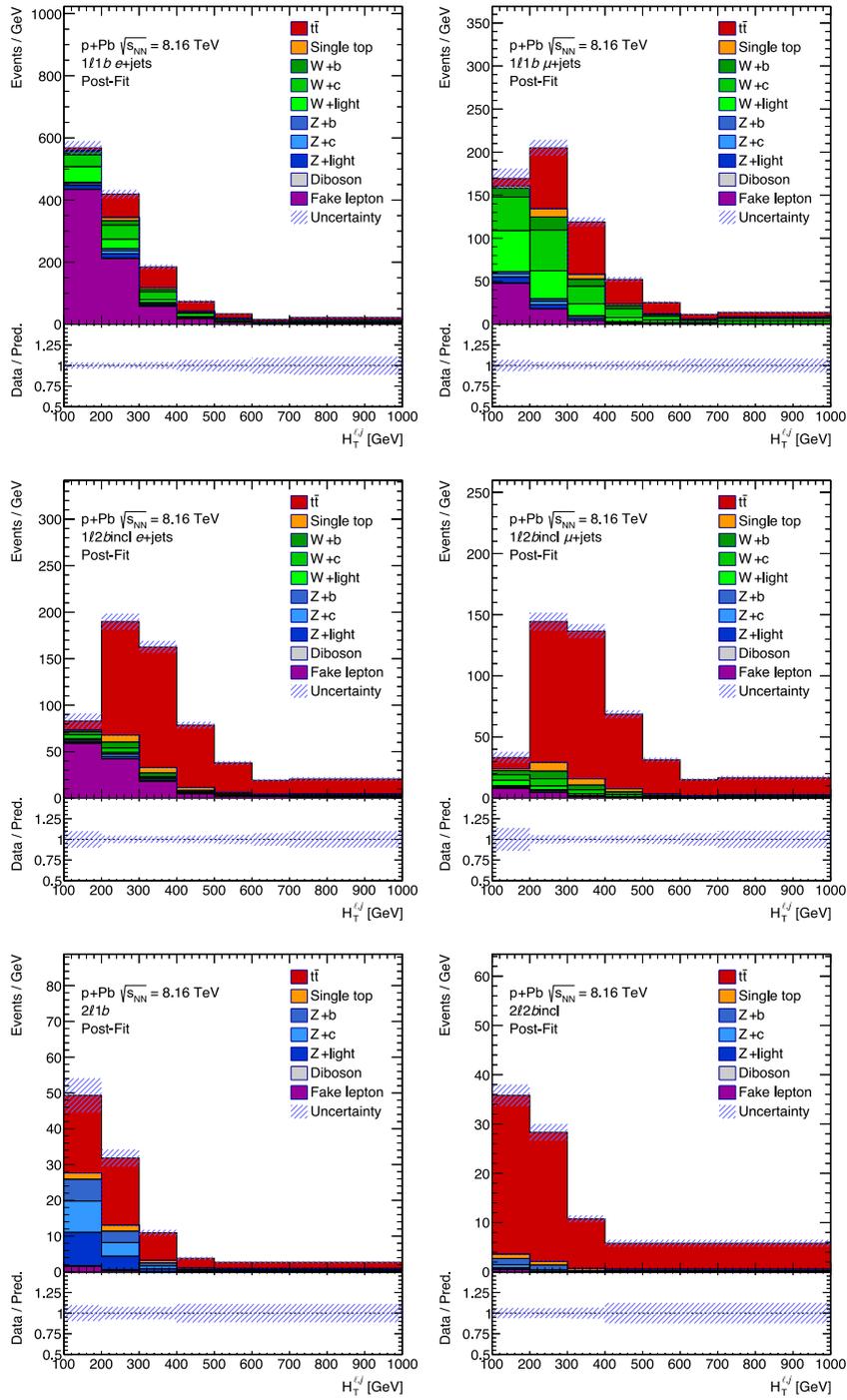


Figure 17. Asimov post-fit plots representing the $H_T^{\ell,j}$ variable in the six SRs ($e+jets$: $1\ell 1b$ and $1\ell 2bincl$, $\mu+jets$: $1\ell 1b$ and $1\ell 2bincl$, dilepton: $2\ell 1b$ and $2\ell 2bincl$), with uncertainties represented by the hatched area. The full markers in the bottom panels show a ratio between data and a sum of predictions. Open triangles indicate bins with entries which are outside the ratio range.

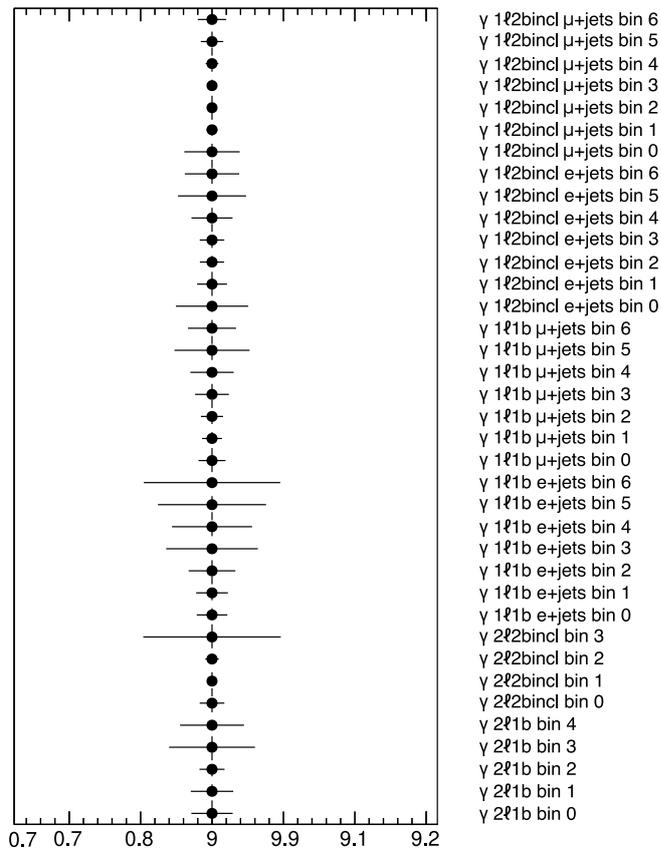


Figure 18. γ parameters after the fit with Asimov data.

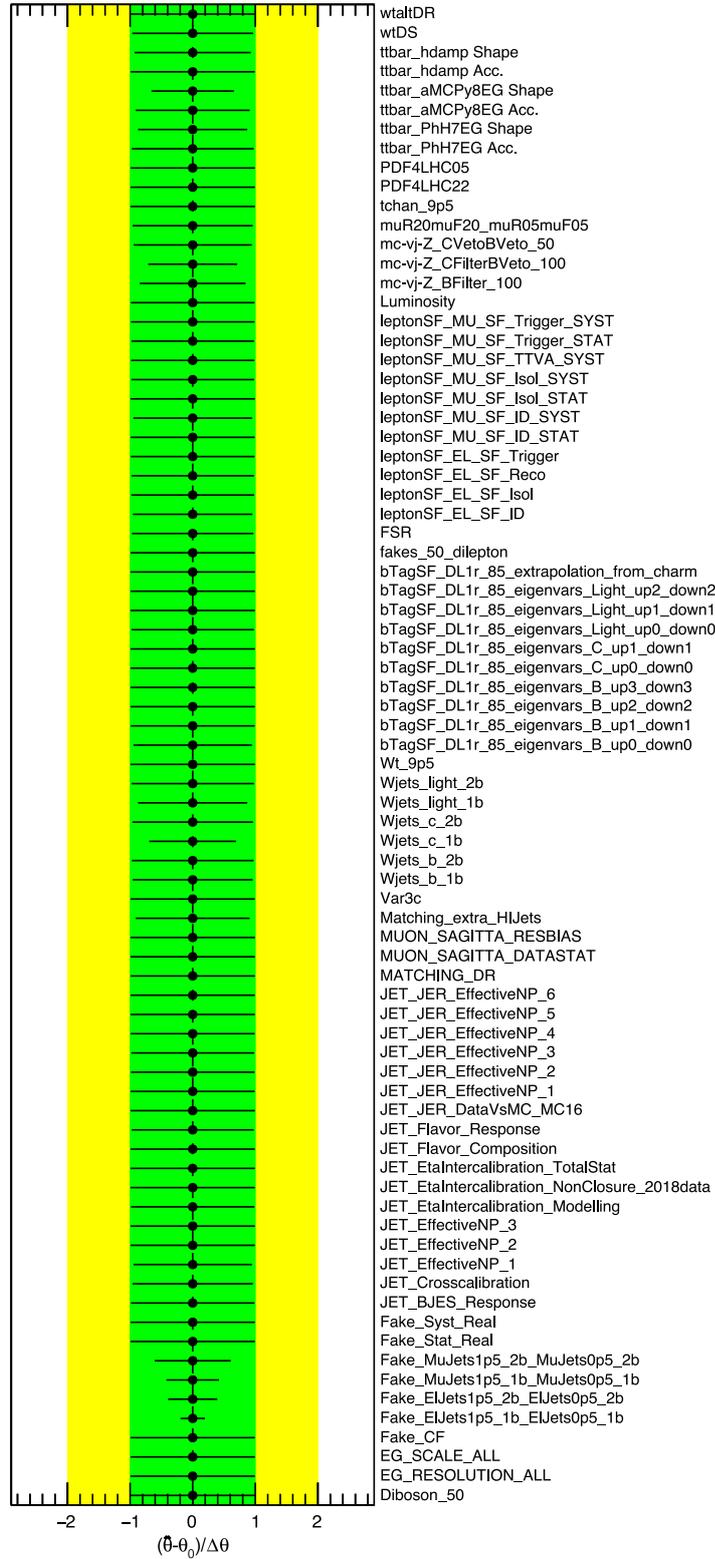


Figure 19. The fitted NPs ($\hat{\theta}$) with their pulls and the constraints in terms of their pre-fit values.

Fake lepton background $e+jets$ 1b	100.0	41.3	59.2	21.8	-57.7	-4.5	-25.0	3.5	1.1	-1.1	10.3	-4.6	2.1	28.3
Fake lepton background $e+jets$ 2bnd	41.3	100.0	21.1	30.2	-57.7	4.2	5.3	-4.0	7.8	-3.9	-2.1	-6.1	1.6	2.3
Fake lepton background $\mu+jets$ 1b	59.2	21.1	100.0	13.6	-25.4	-5.6	-49.5	-0.0	1.2	-0.8	19.2	-8.4	0.7	30.8
Fake lepton background $\mu+jets$ 2bnd	21.8	30.2	13.6	100.0	-27.7	1.2	4.2	-3.5	3.6	-0.5	-0.2	-10.7	1.4	6.2
Hto-PF jet matching	-57.7	-57.7	-25.4	-27.7	100.0	0.6	-7.8	-5.4	-4.2	-0.0	3.5	-5.1	-0.3	-2.2
W+c-jets background	-4.5	4.2	-5.6	1.2	0.6	100.0	-32.3	4.4	3.7	0.5	-23.2	-4.4	0.6	-14.8
W+light-jets background	-25.0	5.3	-49.5	4.2	-7.8	-32.3	100.0	1.5	2.8	0.3	-3.1	-0.6	0.9	2.8
Z+b-jets background	3.5	-4.0	-0.0	-3.5	-5.4	4.4	1.5	100.0	-53.0	7.5	-0.7	-5.2	-0.3	-16.7
Z+c-jets background	1.1	7.8	1.2	3.6	-4.2	3.7	2.8	-53.0	100.0	7.0	-10.6	-3.1	0.1	-7.9
$t\bar{t}$ acc. PHH7	-1.1	-3.9	-0.8	-0.5	-0.0	0.5	0.3	7.5	7.0	100.0	-0.9	-0.9	-0.3	-31.2
$t\bar{t}$ acc. aMC@NLO	10.3	-2.1	19.2	-0.2	3.5	-23.2	-3.1	-0.7	-10.6	-0.9	100.0	2.9	1.8	37.3
$t\bar{t}$ shape aMC@NLO	-4.6	-6.1	-8.4	-10.7	-5.1	-4.4	-0.6	-5.2	-3.1	-0.9	2.9	100.0	30.1	2.0
$t\bar{t}$ h_{amp} shape	2.1	1.6	0.7	1.4	-0.3	0.6	0.9	-0.3	0.1	-0.3	1.8	30.1	100.0	4.4
μ_H	28.3	2.3	30.8	6.2	-2.2	-14.8	2.8	-16.7	-7.9	-31.2	37.3	2.0	4.4	100.0
	Fake lepton background $e+jets$ 1b	Fake lepton background $e+jets$ 2bnd	Fake lepton background $\mu+jets$ 1b	Fake lepton background $\mu+jets$ 2bnd	Hto-PF jet matching	W+c-jets background	W+light-jets background	Z+b-jets background	Z+c-jets background	$t\bar{t}$ acc. PHH7	$t\bar{t}$ acc. aMC@NLO	$t\bar{t}$ shape aMC@NLO	$t\bar{t}$ h_{amp} shape	μ_H

Figure 20. Correlation matrix of the fit parameters for the combined fit to Asimov data: only parameters with at least one correlation number greater than 30% are shown.

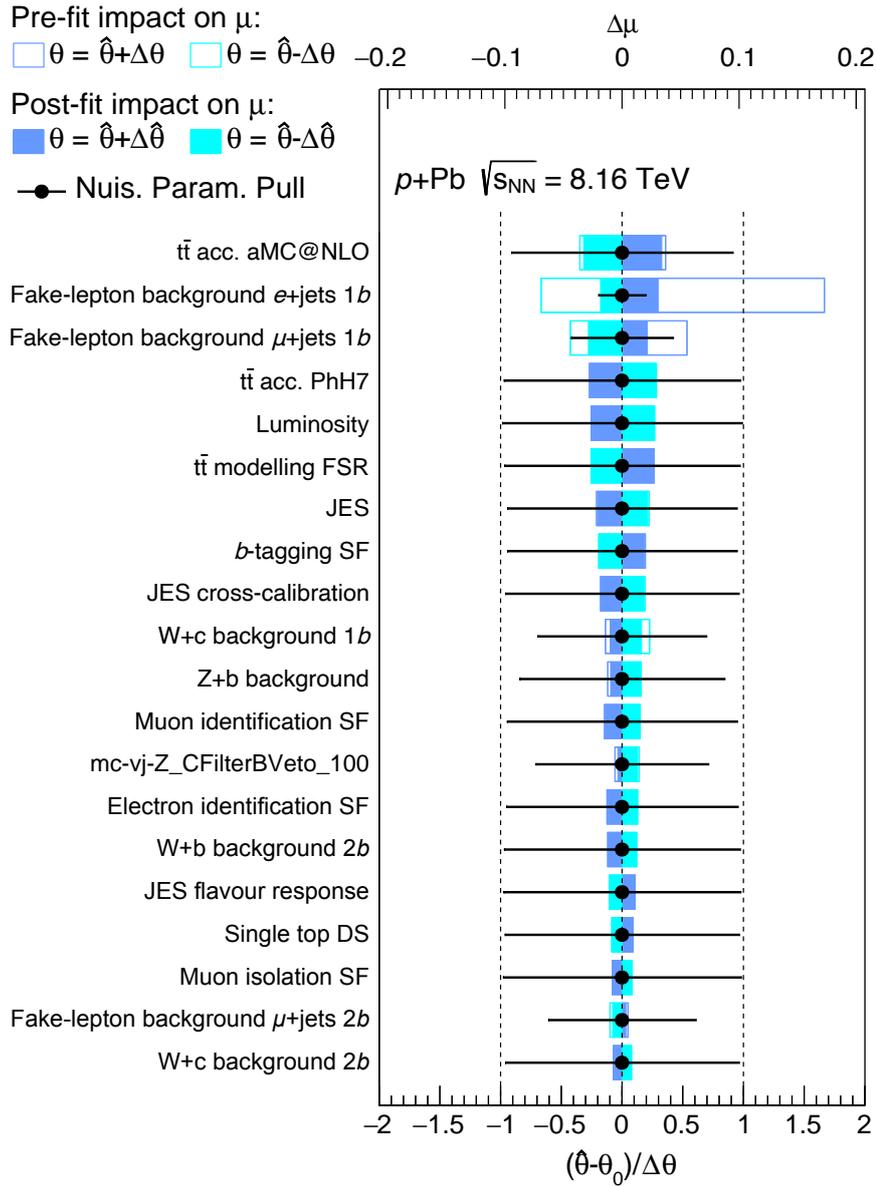


Figure 21. The impact of systematic uncertainties on the fitted signal-strength parameter $\hat{\mu}$ for the combined fit of all channels. Only the 15 most significant systematic uncertainties are shown and listed in decreasing order of their impact on μ on the y -axis. The empty (filled) blue/cyan boxes correspond to the pre-fit (post-fit) impact on μ , referring to the upper x -axis. The impact of each systematic uncertainty, $\Delta\mu$, is calculated by comparing the nominal best-fit value of μ with the result of the fit when fixing the corresponding nuisance parameter θ to its best-fit value $\hat{\theta}$ shifted by its pre-fit (post-fit) uncertainties $\hat{\theta} \pm \Delta\theta$ ($\hat{\theta} \pm \Delta\hat{\theta}$). The black points, which refer to the lower x -axis, show the pulls of the fitted nuisance parameters, i.e., the deviations of the fitted parameters $\hat{\theta}$ from their nominal values θ_0 , normalized to their nominal uncertainties $\Delta\theta$. The black lines show the post-fit uncertainties of the nuisance parameters, relative to their nominal uncertainties, which are indicated by the dashed lines.

Pruning

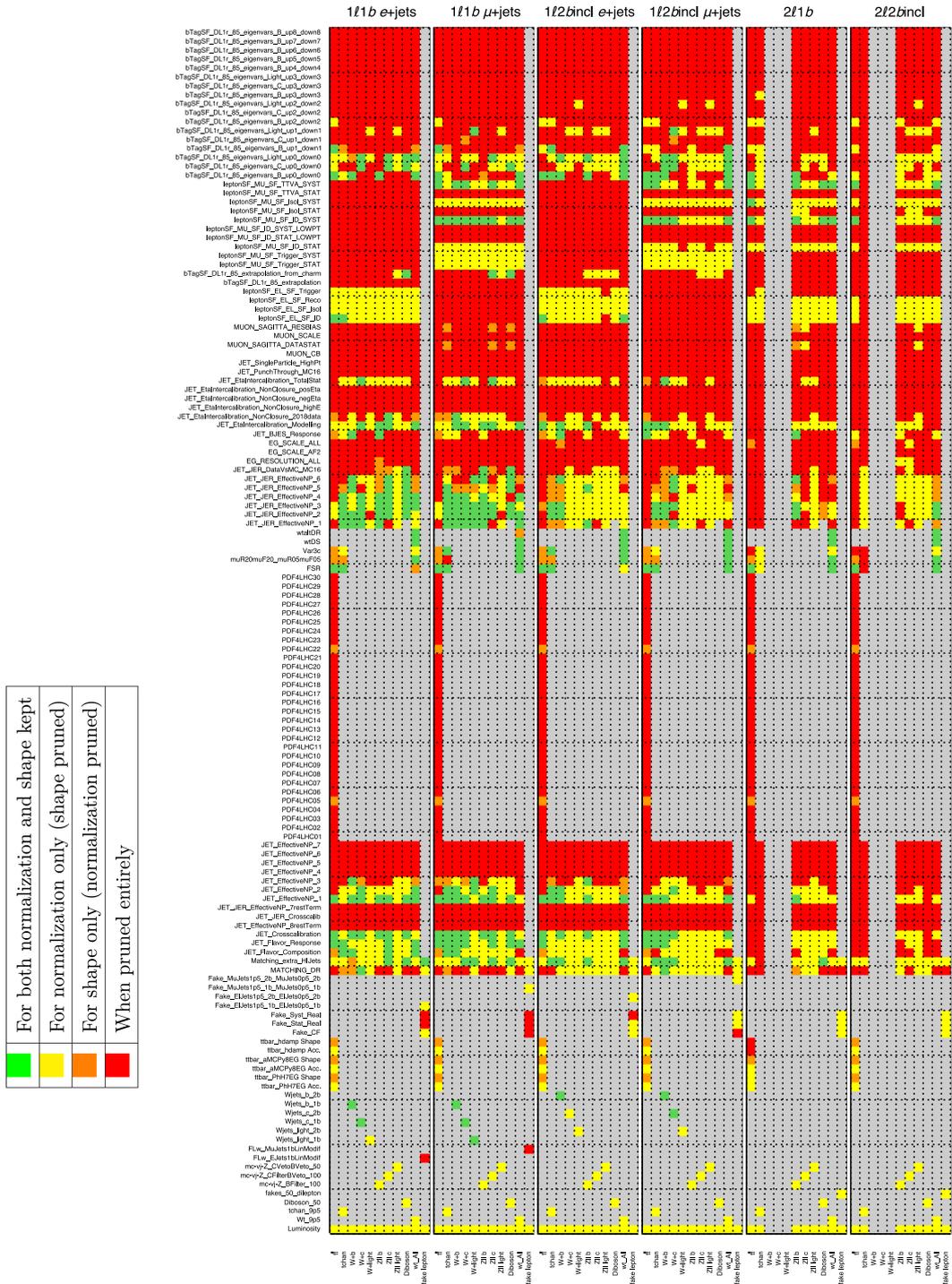


Figure 22. The nuisance parameters considered for normalization and shape uncertainties and whether they are kept or pruned in the fit.

Event display

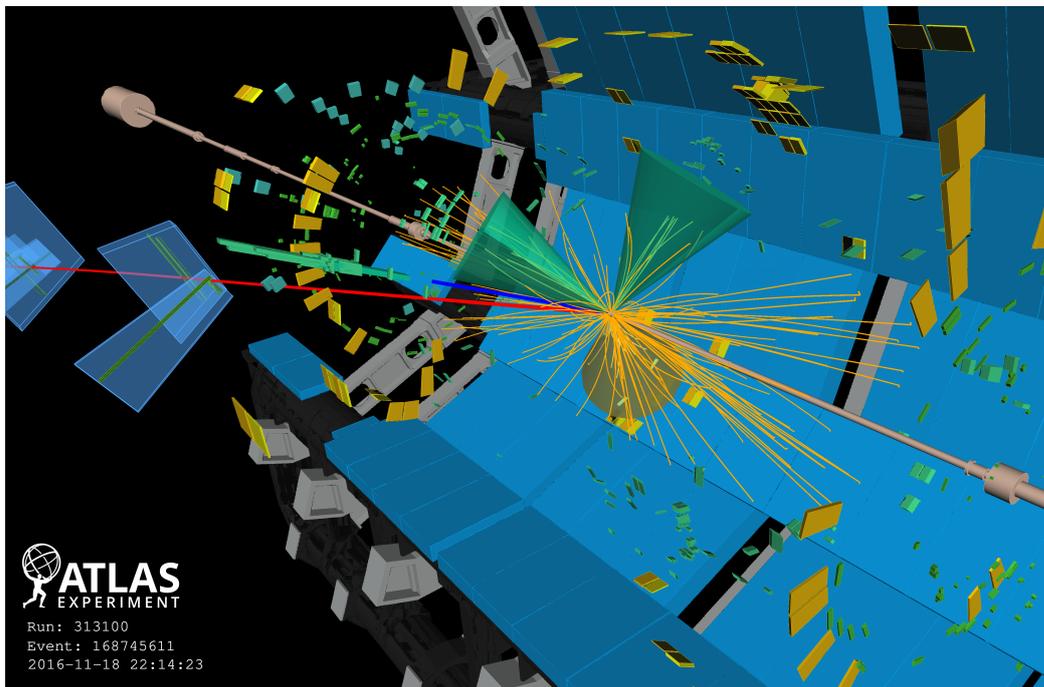


Figure 23. Event candidate for $t\bar{t}$ production in proton-lead ion collision decaying in the dilepton channel with the reconstructed electron track (blue line) and its associated energy deposit in the electromagnetic calorimeter (green boxes), the reconstructed muon track (red line) and its associated muon chambers (blue boxes; the bars are the related measurements), one b -tagged jet (yellow cone) and two non- b -tagged jets (green cones), and tracks in the inner detector (orange lines). Yellow rectangles correspond to energy deposits in cells of the hadron calorimeter.

Bibliography

- [1] COLLABORATION, A.: Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Physics Letters B*, **716**(1), pp. 1–29, 2012.
- [2] CMS COLLABORATION: Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Physics Letters B*, **716**(1), pp. 30–61, 2012.
- [3] Observation of top quark production in $\bar{p}p$ collisions with the collider detector at fermilab. *Phys. Rev. Lett.*, **74**, pp. 2626–2631, Apr 1995, URL <https://link.aps.org/doi/10.1103/PhysRevLett.74.2626>.
- [4] Observation of the top quark. *Phys. Rev. Lett.*, **74**, pp. 2632–2637, Apr 1995, URL <https://link.aps.org/doi/10.1103/PhysRevLett.74.2632>.
- [5] A schematic model of baryons and mesons. *Physics Letters*, **8**(3), pp. 214–215, 1964, ISSN 0031-9163, URL <https://www.sciencedirect.com/science/article/pii/S0031916364920013>.
- [6] *An SU(3) model for strong interaction symmetry and its breaking. Version 2*, pp. 22–101, 2 1964.
- [7] Discovery of a narrow resonance in e^+e^- annihilation. *Phys. Rev. Lett.*, **33**, pp. 1406–1408, Dec 1974, URL <https://link.aps.org/doi/10.1103/PhysRevLett.33.1406>.
- [8] Experimental observation of a heavy particle j . *Phys. Rev. Lett.*, **33**, pp. 1404–1406, Dec 1974, URL <https://link.aps.org/doi/10.1103/PhysRevLett.33.1404>.
- [9] Observation of a dimuon resonance at 9.5 gev in 400-gev proton-nucleus collisions. *Phys. Rev. Lett.*, **39**, pp. 252–255, Aug 1977, URL <https://link.aps.org/doi/10.1103/PhysRevLett.39.252>.
- [10] Observation of structure in the Υ region. *Phys. Rev. Lett.*, **39**, pp. 1240–1242, Nov 1977, URL <https://link.aps.org/doi/10.1103/PhysRevLett.39.1240>.
- [11] Evidence for the Υ'' and a search for new narrow resonances. *Phys. Rev. Lett.*, **42**, pp. 486–489, Feb 1979, URL <https://link.aps.org/doi/10.1103/PhysRevLett.42.486>.
- [12] Note on the nature of cosmic-ray particles. *Phys. Rev.*, **51**, pp. 884–886, May 1937, URL <https://link.aps.org/doi/10.1103/PhysRev.51.884>.
- [13] Evidence for anomalous lepton production in $e^+ - e^-$ annihilation. *Phys. Rev. Lett.*, **35**, pp. 1489–1492, Dec 1975, URL <https://link.aps.org/doi/10.1103/PhysRevLett.35.1489>.
- [14] The quantum theory of the electron. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, **117**(778), pp. 610–624, 1928, arXiv:<https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.1928.0023>, URL <https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1928.0023>.

- [15] SCHNEIDER, J.: *Introduction to Hydraulics and Pneumatics*. Springer Netherlands, 2012, ISBN 978-94-007-2464-8.
- [16] SALAM, A.: Weak and Electromagnetic Interactions. *Conf. Proc. C*, **680519**, pp. 367–377, 1968.
- [17] HOOFT, G.: Renormalizable lagrangians for massive yang-mills fields. *Nuclear Physics B*, **35**(1), pp. 167–188, 1971, ISSN 0550-3213, URL <https://www.sciencedirect.com/science/article/pii/0550321371901398>.
- [18] WEINBERG, S.: A model of leptons. *Phys. Rev. Lett.*, **19**, pp. 1264–1266, Nov 1967, URL <https://link.aps.org/doi/10.1103/PhysRevLett.19.1264>.
- [19] PARTICLE DATA GROUP: Review of particle physics. *Phys. Rev. D*, **98**, p. 030001, 2018, URL <https://link.aps.org/doi/10.1103/PhysRevD.98.030001>.
- [20] POMAROL, A. & SERRA, J.: Top quark compositeness: Feasibility and implications. *Phys. Rev. D*, **78**, p. 074026, Oct 2008, URL <https://link.aps.org/doi/10.1103/PhysRevD.78.074026>.
- [21] RANDALL, L. & SUNDRUM, R.: Large mass hierarchy from a small extra dimension. *Phys. Rev. Lett.*, **83**, pp. 3370–3373, Oct 1999, URL <https://link.aps.org/doi/10.1103/PhysRevLett.83.3370>.
- [22] LYSÁK, R.: Charge asymmetry in top quark pair production. *Symmetry*, **12**(8), p. 1278, 2020, ISSN 2073-8994.
- [23] MATSUNO, S.: Quantum theory of macroscopic systems and its application to the electromagnetic field. *Progress of Theoretical Physics*, **49**(2), pp. 652–662, 1973.
- [24] Measurement of the top quark pair production cross-section based on a statistical combination of measurements of dilepton and single-lepton final states at $s = 7$ TeV with the ATLAS detector. Technical Report, CERN, Geneva, 2011, all figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2011-108>, URL <https://cds.cern.ch/record/1373410>.
- [25] PATRIGNANI, C.: Review of particle physics. *Chinese Physics C*, **40**(10), p. 100001, oct 2016, URL <https://dx.doi.org/10.1088/1674-1137/40/10/100001>.
- [26] BRAIBANT, S.; GIACOMELLI, G. & SPURIO, M.: *Particles and Fundamental Interactions: An Introduction to Particle Physics*. Springer, Dordrecht, 2012.
- [27] EVANS, L. & BRYANT, P.: Lhc machine. *Journal of Instrumentation*, **3**, p. S08001, 2008, URL <https://dx.doi.org/10.1088/1748-0221/3/08/S08001>.
- [28] COLLABORATION, A.: The atlas experiment at the cern large hadron collider. *Journal of Instrumentation*, **3**, p. S08003, 2008, URL <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08003>.
- [29] COLLABORATION, C.: The cms experiment at the cern lhc. *Journal of Instrumentation*, **3**, p. S08004, 2008, URL <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S0004>.
- [30] COLLABORATION, L.: The lhcb detector at the lhc. *Journal of Instrumentation*, **3**, p. S08005, 2008, URL <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08005>.
- [31] COLLABORATION, A.: The alice experiment at the cern lhc. *Journal of Instrumentation*, **3**, p. S08002, 2008, URL <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08002>.

- [32] COLLABORATION, T.: The totem experiment at the cern large hadron collider. *Journal of Instrumentation*, **3**, p. S08007, 2008, URL <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08007>.
- [33] COLLABORATION, L.: The lhcf detector at the cern large hadron collider. *Journal of Instrumentation*, **3**, p. S08006, 2008, URL <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08006>.
- [34] COLLABORATION, M.: Technical design report of the moedal experiment. Technical Report, CERN, 2009.
- [35] COLLABORATION, A.: The atlas inner detector commissioning and calibration. *European Physical Journal C*, **70**, pp. 787–821, 2010.
- [36] COLLABORATION, A.: The atlas inner detector commissioning and calibration. *European Physical Journal C*, **70**, pp. 787–821, 2010.
- [37] AAD, G. E. A. A. C.: Atlas pixel detector electronics and sensors. *Journal of Instrumentation*, **3**, p. P07007, 2008.
- [38] CAPEANS, M. ET AL.: Atlas insertable b-layer technical design report. Technical Report CERN-LHCC-2010-013, ATLAS-TDR-19, CERN, 2010, URL <https://cds.cern.ch/record/1303611>.
- [39] COLLABORATION, A. I.: Production and integration of the atlas insertable b-layer. *Journal of Instrumentation*, **13**(05), p. T05008, 2018, URL <https://iopscience.iop.org/article/10.1088/1748-0221/13/05/T05008>.
- [40] AHMAD, A. E. A.: The silicon microstrip sensors of the atlas semiconductor tracker. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **578**, pp. 98–118, 2007.
- [41] COLLABORATION, A.: Operation and performance of the atlas semiconductor tracker. *Journal of Instrumentation*, **9**, p. P08009, 2014, URL <https://iopscience.iop.org/article/10.1088/1748-0221/9/08/P08009>.
- [42] ABAT, E. E. A. A. T. C.: The atlas trt barrel detector. *Journal of Instrumentation*, **3**, p. P02014, 2008, URL <https://iopscience.iop.org/article/10.1088/1748-0221/3/02/P02014>.
- [43] COLLABORATION, A. T.: The atlas trt end-cap detectors. *Journal of Instrumentation*, **3**, p. P10003, 2008, URL <https://iopscience.iop.org/article/10.1088/1748-0221/3/10/P10003>.
- [44] COLLABORATION, A.: The atlas experiment at the cern large hadron collider. *Journal of Instrumentation*, **3**, p. S08003, 2008, URL <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08003>.
- [45] COLLABORATION, A.: Atlas muon spectrometer: Technical design report. Technical Report, CERN, 1997, URL <https://cds.cern.ch/record/331068>.
- [46] COLLABORATION, A.: Performance of the atlas trigger system in 2010. *European Physical Journal C*, **72**, p. 1849, 2012, URL <https://link.springer.com/article/10.1140/epjc/s10052-011-1849-1>.
- [47] COLLABORATION, A.: The atlas data acquisition and high level trigger system. *Journal of Instrumentation*, **11**(06), p. P06008, 2016, URL <https://iopscience.iop.org/article/10.1088/1748-0221/11/06/P06008>.

- 1088/1748-0221/11/06/P06008.
- [48] COLLABORATION, A.: Performance of the atlas trigger system in 2015. *European Physical Journal C*, **77**, p. 317, 2017, URL <https://link.springer.com/article/10.1140/epjc/s10052-017-4852-3>.
- [49] LHC Machine. *Journal of Instrumentation*, **3**, p. S08001, 2008, URL <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08001>.
- [50] Monte Carlo to Monte Carlo scale factors for flavour tagging efficiency calibration. Technical Report ATL-PHYS-PUB-2020-009, CERN, Geneva, May 2020, URL <https://cds.cern.ch/record/2718610>.
- [51] Flavor Tagging with Track Jets in Boosted Topologies with the ATLAS Detector. Technical Report ATL-PHYS-PUB-2014-013, CERN, Geneva, Aug 2014, URL <https://cds.cern.ch/record/1750681>.
- [52] AABOUD, M.; AAD, G.; ABBOTT, B.; ABDALLAH, J.; ABDINOV, O.; ABELOOS, B.; ABEN, R.; ABOUZEID, O. S.; ABRAHAM, N. L. & ET AL.: Performance of the atlas trigger system in 2015. *The European Physical Journal C*, **77**(5), May 2017, ISSN 1434-6052, URL <http://dx.doi.org/10.1140/epjc/s10052-017-4852-3>.
- [53] ATLAS COLLABORATION: **79**(2), p. 120, Feb 2019, URL <https://doi.org/10.1140/epjc/s10052-019-7450-8>.
- [54] AAD, G. ET AL. (ATLAS): A neural network clustering algorithm for the ATLAS silicon pixel detector. *JINST*, **9**, p. P09009, 2014, arXiv:1406.7690.
- [55] COLLABORATION, A.: Expected performance of the atlas b-tagging algorithms in run-2. Technical Report, CERN, 2015, URL <https://cds.cern.ch/record/2037697>.
- [56] ON BEHALF OF THE ATLAS COLLABORATION, M. P.: Machine learning algorithms for b-jet tagging at the atlas experiment. *Journal of Physics: Conference Series*, **1085**, 2018, URL <http://cds.cern.ch/record/2273281>.
- [57] CAPEANS, M.; DARBO, G.; EINSWEILLER, K.; ELSING, M.; FLICK, T.; GARCIA-SCIVERES, M.; GEMME, C.; PERNEGGER, H.; ROHNE, O. & VUILLERMET, R. (ATLAS Collaboration): ATLAS Insertable B-Layer Technical Design Report. Technical Report CERN-LHCC-2010-013. ATLAS-TDR-19, Sep 2010, URL <https://cds.cern.ch/record/1291633>.
- [58] Performance of b-jet identification in the atlas experiment. *Journal of Instrumentation*, **11**(04), p. P04008–P04008, Apr 2016, ISSN 1748-0221, URL <http://dx.doi.org/10.1088/1748-0221/11/04/P04008>.
- [59] Secondary vertex finding for jet flavour identification with the ATLAS detector. Technical Report ATL-PHYS-PUB-2017-011, CERN, Geneva, Jun 2017, URL <https://cds.cern.ch/record/2270366>.
- [60] Performance and Calibration of the JetFitterCharm Algorithm for c-Jet Identification. Technical Report ATL-PHYS-PUB-2015-001, CERN, Geneva, Jan 2015, URL <https://cds.cern.ch/record/1980463>.
- [61] Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run. Technical Report ATL-PHYS-PUB-2016-012, CERN, Geneva, Jun 2016, URL <https://cds.cern.ch/record/2160731>.

- [62] Identification of Jets Containing b -Hadrons with Recurrent Neural Networks at the ATLAS Experiment. Technical Report ATL-PHYS-PUB-2017-003, CERN, Geneva, Mar 2017, URL <https://cds.cern.ch/record/2255226>.
- [63] Performance of primary vertex reconstruction in proton-proton collisions at $\sqrt{s} = 7$ TeV in the ATLAS experiment. Technical Report ATLAS-CONF-2010-069, CERN, Geneva, Jul 2010, URL <https://cds.cern.ch/record/1281344>.
- [64] Optimisation and performance studies of the ATLAS b -tagging algorithms for the 2017-18 LHC run. Technical Report ATL-PHYS-PUB-2017-013, CERN, Geneva, Jul 2017, URL <https://cds.cern.ch/record/2273281>.
- [65] HOCKER, A. & ET AL: TMVA - Toolkit for Multivariate Data Analysis with ROOT: Users guide. TMVA - Toolkit for Multivariate Data Analysis. Technical Report physics/0703039, CERN, Geneva, Mar 2007, tMVA-v4 Users Guide: 135 pages, 19 figures, numerous code examples and references, URL <https://cds.cern.ch/record/1019880>.
- [66] LANFERMANN, M. C. (ATLAS): Deep Learning in Flavour Tagging at the ATLAS experiment, 2017, URL <https://cds.cern.ch/record/2274065>.
- [67] Calibration of the ATLAS b -tagging algorithm in $t\bar{t}$ semi-leptonic events. Technical Report ATLAS-CONF-2018-045, CERN, Geneva, Sep 2018, URL <http://cds.cern.ch/record/2638455>.
- [68] Optimisation of the smoothing of b -jet identification efficiency and mistag rate simulation-to-data scale factors in ATLAS. Technical Report ATL-PHYS-PUB-2020-004, CERN, Geneva, 2020, URL <https://cds.cern.ch/record/2710598>.
- [69] COLLABORATION, A.: Measurements of the higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s}=7$ and 8 tev in the atlas experiment. *Phys. Rev. D*, **76**, 2016.
- [70] COLLABORATION, C.: Precise determination of the mass of the higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 tev. *Phys. Rev. Lett.*, **115**, 2015.
- [71] COLLABORATION, A.: Electron and photon performance measurements with the atlas detector using the 2015–2017 lhc proton-proton collision data, 2019, arXiv:[1908.00005](https://arxiv.org/abs/1908.00005).
- [72] COLLABORATION, C.: Combined measurements of higgs boson couplings in proton–proton collisions at $\sqrt{s} = 13$ tev. *Eur. Phys. J. C*, **79**, 2019.
- [73] ATLAS COLLABORATION: Search for the associated production of the Higgs boson with a top quark pair in multilepton final states with the ATLAS detector. *Phys. Lett. B*, **749**, p. 519, 2015, arXiv:[1506.05988](https://arxiv.org/abs/1506.05988).
- [74] ATLAS COLLABORATION (ATLAS): Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Eur. Phys. J. C*, **75**, p. 349, 2015, arXiv:[1503.05066](https://arxiv.org/abs/1503.05066).
- [75] ATLAS COLLABORATION: Search for $h \rightarrow \gamma\gamma$ produced in association with top quarks and constraints on the yukawa coupling between the top quark and the higgs boson using data taken at 7 tev and 8 tev with the atlas detector. *Phys. Lett. B*, **740**, p. 222, 2015, arXiv:[1409.3122](https://arxiv.org/abs/1409.3122).
- [76] CMS COLLABORATION (CMS): Search for the associated production of the higgs boson with a

- top-quark pair. *JHEP*, **09**, p. 087, 2014, arXiv:[1408.1682](#).
- [77] CMS COLLABORATION (CMS): Search for a Standard Model Higgs Boson Produced in Association with a Top-Quark Pair and Decaying to Bottom Quarks Using a Matrix Element Method. *Eur. Phys. J. C*, **75**(6), p. 251, 2015, arXiv:[1502.02485](#).
- [78] ATLAS AND CMS COLLABORATIONS (ATLAS, CMS): 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, 2016, arXiv:[1606.02266](#).
- [79] HEINEMEYER, S. ET AL. (LHC Higgs Cross Section Working Group): Handbook of lhc higgs cross sections: 3. higgs properties, 2013, arXiv:[1307.1347](#).
- [80] LHC HIGGS CROSS SECTION WORKING GROUP: , 2016, URL <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt13TeV>.
- [81] COLLABORATION, A.: Measurement of the $t\bar{t}w$ production cross-section. *ATLAS-CONF-2017-050*, 2017.
- [82] Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC, 6 2019.
- [83] AVONI, G.; BRUSCHI, M.; CABRAS, G.; CAFORIO, D.; DEGHANIAN, N.; FLODERUS, A. & ... ZOCCOLI, A. T. O. . A.: The new lucid-2 detector for luminosity measurement and monitoring in atlas. *Journal of Instrumentation*, **13**(07), 2018.
- [84] SJÖSTRAND, T.; MRENNNA, S. & SKANDS, P.: A brief introduction to pythia 8.1. *Computer Physics Communications*, **178**(11), p. 852–867, June 2008, ISSN 0010-4655, URL <http://dx.doi.org/10.1016/j.cpc.2008.01.036>.
- [85] Further ATLAS tunes of PYTHIA6 and Pythia 8. Technical Report, CERN, Geneva, 2011, all figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2011-014>, URL <https://cds.cern.ch/record/1400677>.
- [86] ATLAS COLLABORATION: The ATLAS Simulation Infrastructure. *The European Physical Journal C*, **70**(3), p. 823–874, September 2010, ISSN 1434-6052, URL <http://dx.doi.org/10.1140/epjc/s10052-010-1429-9>.
- [87] AGOSTINELLI, S. ET AL. (GEANT4): GEANT4: A Simulation toolkit. *Nucl. Instrum. Meth.*, **A506**, pp. 250–303, 2003.
- [88] FRIXIONE, S.; RIDOLFI, G. & NASON, P.: A positive-weight next-to-leading-order monte carlo for heavy flavour hadroproduction. *JHEP*, **09**, p. 126, 2007, arXiv:[0707.3088](#).
- [89] HARTANTO, H.; JÄGER, B.; REINA, L. & WACKEROTH, D.: Higgs boson production in association with top quarks in the powheg box. *Physical Review D*, **91**(9), May 2015, ISSN 1550-2368, URL <http://dx.doi.org/10.1103/PhysRevD.91.094003>.
- [90] BALL, R. D. ET AL. (NNPDF): Parton distributions for the LHC Run II. *JHEP*, **04**, p. 040, 2015, arXiv:[1410.8849](#).
- [91] SJÖSTRAND, T.; ASK, S.; CHRISTIANSEN, J. R.; CORKE, R.; DESAI, N.; ILTEN, P.; MRENNNA, S.; PRESTEL, S.; RASMUSSEN, C. O. & SKANDS, P. Z.: An introduction to pythia 8.2. *Computer Physics Communications*, **191**, p. 159–177, June 2015, ISSN 0010-4655, URL <http://dx.doi.org/10.1016/j.cpc.2015.01.024>.
- [92] ATLAS Pythia 8 tunes to 7 TeV data. Technical Report, CERN, Geneva, 2014, URL <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2014-014>.

- [//cds.cern.ch/record/1966419](https://cds.cern.ch/record/1966419).
- [93] DE FLORIAN ET AL., D.: Higgs boson production cross sections at the lhc through nnlo qcd and nlo ew corrections. *Phys.Lett.B*, **783**, pp. 349–359, 2018, arXiv:[1711.00822](https://arxiv.org/abs/1711.00822).
- [94] DJOUADI, A.; KALINOWSKI, J. & SPIRA, M.: Hdecay: a program for higgs boson decays in the standard model and its supersymmetric extension. *Computer Physics Communications*, **108**(1), p. 56–74, January 1998, ISSN 0010-4655, URL [http://dx.doi.org/10.1016/S0010-4655\(97\)00123-9](http://dx.doi.org/10.1016/S0010-4655(97)00123-9).
- [95] BEENAKKER, W.; DITTMAYER, S.; KRAMER, M.; PLUMPER, B.; SPIRA, M. & ZERWAS, P. M.: Higgs radiation off top quarks at the Tevatron and the LHC. *Phys. Rev. Lett.*, **87**, p. 201805, 2001, arXiv:[hep-ph/0107081](https://arxiv.org/abs/hep-ph/0107081).
- [96] BEENAKKER, W.; DITTMAYER, S.; KRAMER, M.; PLUMPER, B.; SPIRA, M. & ZERWAS, P. M.: NLO QCD corrections to t anti-t H production in hadron collisions. *Nucl. Phys. B*, **653**, p. 151, 2003, arXiv:[hep-ph/0211352](https://arxiv.org/abs/hep-ph/0211352).
- [97] DAWSON, S.; ORR, L. H.; REINA, L. & WACKEROTH, D.: Associated top quark Higgs boson production at the LHC. *Phys. Rev. D*, **67**, p. 071503, 2003, arXiv:[hep-ph/0211438](https://arxiv.org/abs/hep-ph/0211438).
- [98] DAWSON, S.; JACKSON, C.; ORR, L. H.; REINA, L. & WACKEROTH, D.: Associated Higgs production with top quarks at the large hadron collider: NLO QCD corrections. *Phys. Rev. D*, **68**, p. 034022, 2003, arXiv:[hep-ph/0305087](https://arxiv.org/abs/hep-ph/0305087).
- [99] ZHANG, Y.; MA, W.-G.; ZHANG, R.-Y.; CHEN, C. & GUO, L.: QCD NLO and EW NLO corrections to $t\bar{t}H$ production with top quark decays at hadron collider. *Phys. Lett. B*, **738**, p. 1, 2014, arXiv:[1407.1110](https://arxiv.org/abs/1407.1110).
- [100] FRIXIONE, S.; HIRSCHI, V.; PAGANI, D.; SHAO, H. S. & ZARO, M.: Weak corrections to Higgs hadroproduction in association with a top-quark pair. *JHEP*, **09**, p. 065, 2014, arXiv:[1407.0823](https://arxiv.org/abs/1407.0823).
- [101] BELLM, J. ET AL.: Herwig 7.0/Herwig++ 3.0 release note. *Eur. Phys. J. C*, **76**(4), p. 196, 2016, arXiv:[1512.01178](https://arxiv.org/abs/1512.01178).
- [102] BUTTERWORTH, J. ET AL.: PDF4LHC recommendations for LHC Run II. *J. Phys. G*, **43**, p. 023001, 2016, arXiv:[1510.03865](https://arxiv.org/abs/1510.03865).
- [103] FRIXIONE, S.; RIDOLFI, G. & NASON, P.: A positive-weight next-to-leading-order monte carlo for heavy flavour hadroproduction. *JHEP*, **09**, 2007, arXiv:[arXiv:0707.3088](https://arxiv.org/abs/0707.3088)[\[hep-ph\]](https://arxiv.org/abs/0707.3088).
- [104] HARTANTO, H. B.; JÄGER, B.; REINA, L. & WACKEROTH, D.: Higgs boson production in association with top quarks in the powheg box. *Phys. Rev. D*, **91**, 2015, arXiv:[arXiv:1501.04498](https://arxiv.org/abs/1501.04498)[\[hep-ph\]](https://arxiv.org/abs/1501.04498).
- [105] BALL, R. D. ET AL.: Parton distributions with LHC data. *Nucl. Phys. B*, **867**, pp. 244–289, 2013, arXiv:[1207.1303](https://arxiv.org/abs/1207.1303).
- [106] LAI, H.-L. ET AL.: New parton distributions for collider physics. *Phys. Rev. D*, **82**, p. 074024, 2010, arXiv:[1007.2241](https://arxiv.org/abs/1007.2241).
- [107] PUMPLIN, J.; STUMP, D. R.; HUSTON, J.; LAI, H.-L.; NADOLSKY, P. M. & TUNG, W. K.: New generation of parton distributions with uncertainties from global qcd analysis. *JHEP*, **07**, p. 012, 2002, arXiv:[hep-ph/0201195](https://arxiv.org/abs/hep-ph/0201195).
- [108] NADOLSKY, P. M. ET AL.: Implications of cteq global analysis for collider observables. *Phys. Rev. D*, **78**, p. 013004, 2008, arXiv:[0802.0007](https://arxiv.org/abs/0802.0007).

- [109] SEYMOUR, M. H. & SIODMOK, A.: Constraining MPI models using σ_{eff} and recent Tevatron and LHC Underlying Event data. *JHEP*, **10**, p. 113, 2013, arXiv:[1307.5015](#).
- [110] GLEISBERG, T. ET AL.: Event generation with sherpa 1.1. *JHEP*, **02**, p. 007, 2009, arXiv:[0811.4622](#).
- [111] SKANDS, P.: Tuning monte carlo generators: The perugia tunes. *Phys. Rev. D*, **82**, p. 074018, 2010, arXiv:[1005.3457](#).
- [112] RE, E.: Single-top wt-channel production matched with parton showers using the powheg method. *Eur. Phys. J. C*, **71**, p. 1547, 2011, arXiv:[1009.2450](#).
- [113] ALIOLI, S.; NASON, P.; OLEARI, C. & RE, E.: Nlo single-top production matched with shower in powheg: s - and t -channel contributions. *JHEP*, **09**, p. 111, 2009, arXiv:[0907.4076](#).
- [114] ALWALL, J. ET AL.: The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, **07**, p. 079, 2014, arXiv:[1405.0301](#).
- [115] SJÖSTRAND, T. ET AL.: High-energy-physics event generation with pythia 6.1. *Comput. Phys. Commun.*, **135**, p. 238, 2001, arXiv:[hep-ph/0010017](#).
- [116] SJÖSTRAND, T.; MRENNNA, S. & SKANDS, P.: A brief introduction to pythia 8.1. *Comput. Phys. Commun.*, **178**, p. 852, 2008, arXiv:[0710.3820](#).
- [117] BAHR, M. ET AL.: Herwig++ Physics and Manual. *Eur. Phys. J. C*, **58**, pp. 639–707, 2008, arXiv:[0803.0883](#).
- [118] LANGE, D. J.: The EvtGen particle decay simulation package. *Nucl. Instrum. Meth. A*, **462**, pp. 152–155, 2001.
- [119] GOLONKA, P. & WAS, Z.: PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays. *Eur. Phys. J.*, **C45**, pp. 97–107, 2006, arXiv:[hep-ph/0506026](#).
- [120] Analysis of $t\bar{t}H$ and $t\bar{t}W$ production in multilepton final states with the ATLAS detector, 10 2019, URL <https://inspirehep.net/literature/1759582>.
- [121] GLEISBERG, T. & HÖCHE, S.: Comix, a new matrix element generator. *Journal of High Energy Physics*, **2008**(12), p. 039–039, December 2008, ISSN 1029-8479, URL <http://dx.doi.org/10.1088/1126-6708/2008/12/039>.
- [122] CASCIOLI, F.; MAIERHÖFER, P. & POZZORINI, S.: Scattering amplitudes with open loops. *Physical Review Letters*, **108**(11), March 2012, ISSN 1079-7114, URL <http://dx.doi.org/10.1103/PhysRevLett.108.111601>.
- [123] SCHUMANN, S. & KRAUSS, F.: A parton shower algorithm based on catani-seymour dipole factorisation. *Journal of High Energy Physics*, **2008**(03), p. 038–038, March 2008, ISSN 1029-8479, URL <http://dx.doi.org/10.1088/1126-6708/2008/03/038>.
- [124] HÖCHE, S.; KRAUSS, F.; SCHÖNHERR, M. & SIEGERT, F.: Qcd matrix elements + parton showers. the nlo case. *Journal of High Energy Physics*, **2013**(4), April 2013, ISSN 1029-8479, URL [http://dx.doi.org/10.1007/JHEP04\(2013\)027](http://dx.doi.org/10.1007/JHEP04(2013)027).
- [125] CAMPBELL, J.; ELLIS, K. & RÖNTSCH, R.: Single top production in association with a z boson at the lhc. *Phys.Rev.D*, **87**, p. 094007, 2013, arXiv:[1212.4888](#).
- [126] FRIXIONE, S. & PASSARINO, G.: Heavy flavors in higgs production. *Phys.Lett.B*, **728**, pp.

- 328–336, 2014, arXiv:[1308.1032](#).
- [127] FREDERIX, R. & FRIXIONE, S.: Merging meets matching in mc@nlo. *JHEP*, **12**, p. 061, 2012, arXiv:[1209.6215](#).
- [128] FREDERIX, R. & FRIXIONE, S.: Sub-leading nlo electroweak corrections for $t\bar{t}w$ production. *JHEP*, **09**, p. 028, 2017, arXiv:[1707.08204](#).
- [129] MCFAYDEN, J. A.: *ATLAS searches for additional scalars and exotic Higgs boson decays with the LHC Run 2 dataset*. Ph.D. thesis, University of Oxford, 2020, cERN-THESIS-2020-050.
- [130] GUTSCHOW, C.: *Measurements of electroweak-boson plus jets production using the ATLAS detector*. Ph.D. thesis, University of London, 2020, cERN-THESIS-2020-048.
- [131] NACHMAN, B. P.: *Understanding Hadronic Activity in the ATLAS Experiment*. Ph.D. thesis, Stanford University, 2020, cERN-THESIS-2020-046.
- [132] CACCIARI, M.; CZAKON, M.; MANGANO, M.; MITOV, A. & NASON, P.: Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation. *Phys. Lett. B*, **710**, pp. 612–622, 2012, (cit. on p. 6), arXiv:[1111.5869](#).
- [133] BÄRNREUTHER, P.; CZAKON, M. & MITOV, A.: Percent-level-precision physics at the tevatron: Next-to-next-to-leading order qcd corrections to $qq \rightarrow tt + x$. *Phys. Rev. Lett.*, **109**, p. 132001, 2012, (cit. on p. 6), arXiv:[1204.5201](#).
- [134] CZAKON, M. & MITOV, A.: Nnlo corrections to top-pair production at hadron colliders: the all-fermionic scattering channels. *JHEP*, **12**, p. 054, 2012, (cit. on p. 6), arXiv:[1207.0236](#).
- [135] CZAKON, M. & MITOV, A.: Nnlo corrections to top pair production at hadron colliders: the quark-gluon reaction. *JHEP*, **01**, p. 080, 2013, (cit. on p. 6), arXiv:[1210.6832](#).
- [136] CZAKON, M.; FIEDLER, P. & MITOV, A.: Total top-quark pair-production cross section at hadron colliders through $\mathcal{O}(\alpha_s^4)$. *Phys. Rev. Lett.*, **110**, p. 252004, 2013, (cit. on p. 6), arXiv:[1303.6254](#).
- [137] COLLABORATION, A.: Improved luminosity determination in pp collisions at $\sqrt{s} = 8$ tev using the atlas detector at the lhc. Technical Report ATL-PHYS-PUB-2016-004, CERN, 2016.
- [138] ATLAS COLLABORATION: ATLAS Inner Detector Tracking Performance. Technical Report, CERN, 2015, URL <https://cds.cern.ch/record/2005781>.
- [139] ATLAS COLLABORATION: Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, **76**(5), p. 292, 2016, arXiv:[1603.05598](#).
- [140] 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**, p. 639, 2019, arXiv:[1902.04655](#).
- [141] ATLAS COLLABORATION: Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1. *Eur. Phys. J. C*, **77**, p. 490, 2017, arXiv:[1603.02934](#).
- [142] Properties of Jets and Inputs to Jet Reconstruction and Calibration with the ATLAS Detector Using Proton-Proton Collisions at $\sqrt{s} = 13$ TeV. Technical Report, CERN, Geneva, 2015, URL <https://cds.cern.ch/record/2044564>.
- [143] CACCIARI, M.; SALAM, G. P. & SOYEZ, G.: The anti- k_t jet clustering algorithm. *JHEP*, **04**,

- p. 063, 2008, arXiv:[0802.1189](#).
- [144] CACCIARI, M.; SALAM, G. P. & SOYEZ, G.: FastJet User Manual. *Eur. Phys. J. C*, **72**, p. 1896, 2012, arXiv:[1111.6097](#).
- [145] ATLAS COLLABORATION: Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev. D*, **96**, p. 072002, 2017, arXiv:[1703.09665](#).
- [146] COLLABORATION, A.: Performance of pile-up mitigation techniques for jets in pp collisions at $s = 8$ tev using the atlas detector. *Eur. Phys. J. C*, **76**(10), p. 581, 2016, (cit. on p. 8), arXiv:[1510.03823](#).
- [147] ATLAS COLLABORATION: Jet energy measurement and its systematic uncertainty in proton–proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Eur. Phys. J. C*, **75**, p. 17, 2015, arXiv:[1406.0076](#).
- [148] ATLAS COLLABORATION: Jet energy resolution in proton–proton collisions at $\sqrt{s} = 7$ TeV recorded in 2010 with the ATLAS detector. *Eur. Phys. J. C*, **73**, p. 2306, 2013, arXiv:[1210.6210](#).
- [149] COLLABORATION, A.: Jet energy scale uncertainties and correlations for jets reconstructed in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. Technical Report ATLAS-CONF-2012-116, CERN, 2012.
- [150] COLLABORATION, A.: Measurement of the top quark pair production cross-section with ATLAS in the single lepton channel at $\sqrt{s} = 8$ TeV. Technical Report ATL-PHYS-PUB-2015-022, CERN, 2015.
- [151] COLLABORATION, A.: Jet energy scale and its systematic uncertainty in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. Technical Report ATL-PHYS-PUB-2016-012, CERN, 2016.
- [152] ATLAS COLLABORATION: Identification of hadronically decaying τ leptons in 2015 data. Technical Report ATL-PHYS-PUB-2015-045, CERN, 2015.
- [153] ATLAS COLLABORATION: Performance of τ lepton reconstruction and identification in the ATLAS experiment using 2015–2016 data. Technical Report ATLAS-CONF-2017-029, CERN, 2017.
- [154] 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**, p. 903, 2018, arXiv:[1802.08168](#).
- [155] ATLAS COLLABORATION: Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data. *JINST*, **14**, p. P12006, 2019, arXiv:[1908.00005](#).
- [156] ALI, B.: *Search for the Standard Model Higgs boson produced in association with a top quark pair in multilepton final states at $\sqrt{s} = 13$ TeV with the ATLAS detector*. Ph.D. thesis, CTU, Prague, 12 2019.
- [157] COLLABORATION, A.: Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC. *The European Physical Journal C*, **76**(12), November 2016, ISSN 1434-6052, URL <http://dx.doi.org/10.1140/epjc/s10052-016-4466-1>.
- [158] COLLABORATION, A.: Muon reconstruction performance of the ATLAS detector in

- proton-proton collision data at $\sqrt{s} = 13$ TeV. *The European Physical Journal C*, **76**(5), May 2016, ISSN 1434-6052, URL <http://dx.doi.org/10.1140/epjc/s10052-016-4120-y>.
- [159] 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. *The European Physical Journal C*, **79**(8), August 2019, ISSN 1434-6052, URL <http://dx.doi.org/10.1140/epjc/s10052-019-7140-6>.
- [160] Tagging and suppression of pileup jets, 5 2014.
- [161] COLLABORATION, A.: Jet energy measurement with the atlas detector in proton-proton collisions at $\sqrt{s} = 7$ TeV. *The European Physical Journal C*, **73**(3), March 2013, ISSN 1434-6052, URL <http://dx.doi.org/10.1140/epjc/s10052-013-2304-2>.
- [162] E_T^{miss} performance in the ATLAS detector using 2015-2016 LHC pp collisions, 6 2018.
- [163] BUTTERWORTH, J. M.; DISSERTORI, G.; GREINER, N.; KIRSCHENMANN, H.; SALAM, G. P. & SOYEZ, G.: PDF4LHC recommendations for LHC Run II. *Journal of Physics G: Nuclear and Particle Physics*, **43**(2), p. 023001, 2016, URL <http://dx.doi.org/10.1088/0954-3899/43/2/023001>.
- [164] CERN: Cern yellow reports: Monographs, vol 2 (2017): Handbook of lhc higgs cross sections: 4. deciphering the nature of the higgs sector, 2017, URL <https://e-publishing.cern.ch/index.php/CYRM/issue/view/32>.
- [165] BARLOW, R. & BEESTON, C.: Fitting using finite monte carlo samples. *Computer Physics Communications*, **77**(2), pp. 219–228, 1993, ISSN 0010-4655, URL <https://www.sciencedirect.com/science/article/pii/001046559390005W>.
- [166] VERKERKE, W. & KIRKBY, D.: The roofit toolkit for data modeling, 2003, arXiv:[physics/0306116](https://arxiv.org/abs/physics/0306116).
- [167] VERKERKE, W. & KIRKBY, D.: *RooFit Users Manual*, 2010, (cit. on p. 20), URL <http://roofit.sourceforge.net>.
- [168] COWAN, G.; CRANMER, K.; GROSS, E. & VITELLS, O.: Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, **71**, p. 1554, 2011, [Erratum: *Eur.Phys.J.C* **73**, 2501 (2013)], arXiv:[1007.1727](https://arxiv.org/abs/1007.1727).
- [169] ATLAS COLLABORATION: Measurement of the production and lepton charge asymmetry of W bosons in Pb+Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV with the ATLAS detector. *Eur. Phys. J. C*, **75**(1), p. 23, 2015, arXiv:[1408.4674](https://arxiv.org/abs/1408.4674).
- [170] CMS COLLABORATION: Study of W Boson Production in PbPb and pp Collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *Phys. Lett. B*, **715**, pp. 66–87, 2012, arXiv:[1205.6334](https://arxiv.org/abs/1205.6334).
- [171] ATLAS COLLABORATION (ATLAS): Measurement of the centrality dependence of J/ψ yields and observation of Z production in lead-lead collisions with the ATLAS detector at the LHC. *Phys. Lett. B*, **697**, pp. 294–312, 2011, arXiv:[1012.5419](https://arxiv.org/abs/1012.5419).
- [172] ATLAS COLLABORATION: Measurement of Z boson Production in Pb+Pb Collisions at $\sqrt{s_{NN}} = 2.76$ TeV with the ATLAS Detector. *Phys. Rev. Lett.*, **110**(2), p. 022301, 2013, arXiv:[1210.6486](https://arxiv.org/abs/1210.6486).
- [173] CMS COLLABORATION: Study of Z boson production in PbPb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *Phys. Rev. Lett.*, **106**, p. 212301, 2011, arXiv:[1102.5435](https://arxiv.org/abs/1102.5435).
- [174] ATLAS COLLABORATION: Measurement of the nuclear modification factor of b -jets in 5.02 TeV

- Pb+Pb collisions with the ATLAS detector. *Eur. Phys. J. C*, **83**(5), p. 438, 2023, arXiv:[2204.13530](#).
- [175] CMS COLLABORATION: Evidence of b-Jet Quenching in PbPb Collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *Phys. Rev. Lett.*, **113**(13), p. 132301, 2014, [Erratum: *Phys.Rev.Lett.* 115, 029903 (2015)], arXiv:[1312.4198](#).
- [176] ATLAS COLLABORATION: Observation of the $\gamma\gamma\rightarrow\tau\tau$ Process in Pb+Pb Collisions and Constraints on the τ -Lepton Anomalous Magnetic Moment with the ATLAS Detector. *Phys. Rev. Lett.*, **131**(15), p. 151802, 2023, arXiv:[2204.13478](#).
- [177] CMS COLLABORATION: Observation of τ lepton pair production in ultraperipheral lead-lead collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *Phys. Rev. Lett.*, **131**, p. 151803, 2023, arXiv:[2206.05192](#).
- [178] KHANPOUR, H. & ATASHBAR TEHRANI, S.: Global Analysis of Nuclear Parton Distribution Functions and Their Uncertainties at Next-to-Next-to-Leading Order. *Phys. Rev. D*, **93**(1), p. 014026, 2016, arXiv:[1601.00939](#).
- [179] ETHIER, J. J. & NOCERA, E. R.: Parton Distributions in Nucleons and Nuclei. *Ann. Rev. Nucl. Part. Sci.*, **70**, pp. 43–76, 2020, arXiv:[2001.07722](#).
- [180] APOLINÁRIO, L.; MILHANO, J. G.; SALAM, G. P. & SALGADO, C. A.: Probing the time structure of the quark-gluon plasma with top quarks. *Phys. Rev. Lett.*, **120**(23), p. 232301, 2018, arXiv:[1711.03105](#).
- [181] D’ENTERRIA, D.; KRAJCZÁR, K. & PAUKKUNEN, H.: Top-quark production in proton–nucleus and nucleus–nucleus collisions at LHC energies and beyond. *Phys. Lett. B*, **746**, pp. 64–72, 2015, arXiv:[1501.05879](#).
- [182] PAUKKUNEN, H.: Nuclear PDFs in the beginning of the LHC era. *Nucl. Phys. A*, **926**, pp. 24–33, 2014, arXiv:[1401.2345](#).
- [183] PAUKKUNEN, H.: Status of nuclear PDFs after the first LHC p–Pb run. *Nucl. Phys. A*, **967**, pp. 241–248, 2017, arXiv:[1704.04036](#).
- [184] WORKMAN, R. L. ET AL. (Particle Data Group): Review of Particle Physics. *PTEP*, **2022**, p. 083C01, 2022.
- [185] GIAMMANCO, A. & SCHWIENHORST, R.: Single top-quark production at the Tevatron and the LHC. *Rev. Mod. Phys.*, **90**(3), p. 035001, 2018, arXiv:[1710.10699](#).
- [186] D’ENTERRIA, D.: Top-quark and Higgs boson perspectives at heavy-ion colliders. *Nucl. Part. Phys. Proc.*, **289-290**, pp. 237–240, 2017, arXiv:[1701.08047](#).
- [187] CMS COLLABORATION: Observation of top quark production in proton-nucleus collisions. *Phys. Rev. Lett.*, **119**(24), p. 242001, 2017, arXiv:[1709.07411](#).
- [188] CMS COLLABORATION: Evidence for Top Quark Production in Nucleus-Nucleus Collisions. *Phys. Rev. Lett.*, **125**(22), p. 222001, 2020, arXiv:[2006.11110](#).
- [189] ATLAS COLLABORATION: The ATLAS Simulation Infrastructure. *Eur. Phys. J. C*, **70**, pp. 823–874, 2010, arXiv:[1005.4568](#).
- [190] AGOSTINELLI, S. ET AL. (GEANT4): GEANT4—a simulation toolkit. *Nucl. Instrum. Meth. A*, **506**, pp. 250–303, 2003.
- [191] BALL, R. D. ET AL. (NNPDF): Parton distributions for the LHC Run II. *JHEP*, **04**, p. 040,

- 2015, arXiv:[1410.8849](#).
- [192] SJOSTRAND, T. ET AL.: An Introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, **191**, pp. 159–177, 2015, arXiv:[1410.3012](#).
- [193] CZAKON, M. & MITOV, A.: Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders. *Comput. Phys. Commun.*, **185**, p. 2930, 2014, arXiv:[1112.5675](#).
- [194] BELLM, J. ET AL.: Herwig 7.2 release note. *Eur. Phys. J. C*, **80**(5), p. 452, 2020, arXiv:[1912.06509](#).
- [195] BOTHMANN, E. ET AL. (Sherpa): Event Generation with Sherpa 2.2. *SciPost Phys.*, **7**(3), p. 034, 2019, arXiv:[1905.09127](#).
- [196] HÖCHE, S.; KRAUSS, F.; SCHÖNHERR, M. & SIEGERT, F.: W+ n -Jet predictions at the Large Hadron Collider at next-to-leading order matched with a parton shower. *Phys. Rev. Lett.*, **110**(5), p. 052001, 2013, arXiv:[1201.5882](#).
- [197] ANASTASIOU, C. ET AL.: Inclusive Higgs boson cross-section for the LHC at 14 TeV. *Phys. Lett. B*, **572**, pp. 232–240, 2003, URL <https://doi.org/10.1016/j.physletb.2003.07.006>.
- [198] FREDERIX, R.; RE, E. & TORRIELLI, P.: Single-top t-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO. *JHEP*, **09**, p. 130, 2012, arXiv:[1207.5391](#).
- [199] RE, E.: Single-top Wt-channel production matched with parton showers using the POWHEG method. *Eur. Phys. J. C*, **71**, p. 1547, 2011, arXiv:[1009.2450](#).
- [200] FRIXIONE, S.; LAENEN, E.; MOTYLINSKI, P.; WEBBER, B. R. & WHITE, C. D.: Single-top hadroproduction in association with a W boson. *JHEP*, **07**, p. 029, 2008, arXiv:[0805.3067](#).
- [201] ATLAS COLLABORATION: Dijet azimuthal correlations and conditional yields in pp and p+Pb collisions at $\sqrt{s}=5.02\text{TeV}$ with the ATLAS detector. *Phys. Rev. C*, **100**(3), p. 034903, 2019, arXiv:[1901.10440](#).
- [202] ATLAS COLLABORATION: Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13\text{ TeV}$ with the ATLAS detector. *Eur. Phys. J. C*, **81**(8), p. 689, 2021, arXiv:[2007.02645](#).
- [203] ATLAS COLLABORATION: Jet reconstruction and performance using particle flow with the ATLAS Detector. *Eur. Phys. J. C*, **77**(7), p. 466, 2017, arXiv:[1703.10485](#).
- [204] ATLAS COLLABORATION: ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13\text{ TeV}$. *Eur. Phys. J. C*, **79**(11), p. 970, 2019, arXiv:[1907.05120](#).
- [205] ATLAS COLLABORATION: Tools for estimating fake/non-prompt lepton backgrounds with the ATLAS detector at the LHC. *JINST*, **18**(11), p. T11004, 2023, arXiv:[2211.16178](#).
- [206] CARDILLO, F.; DUCU, O. A.; MAURER, J.; MERLASSINO, C.; REICHERT, J.; VARNES, E.; BAKKER, P. J. & MORGENSTERN, M.: Tools for estimating fake/non-prompt lepton backgrounds in ATLAS. Technical Report, CERN, Geneva, Sep 1999, URL <https://cds.cern.ch/record/2686919>.
- [207] ATLAS COLLABORATION: Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data. *JINST*, **14**(12), p. P12006, 2019, arXiv:[1908.00005](#).

- [208] ATLAS COLLABORATION (ATLAS): 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**(7), p. 578, 2021, arXiv:2012.00578.
- [209] ATLAS COLLABORATION (ATLAS): Performance of electron and photon triggers in ATLAS during LHC Run 2. *Eur. Phys. J. C*, **80**(1), p. 47, 2020, arXiv:1909.00761.
- [210] ATLAS COLLABORATION (ATLAS): Performance of the ATLAS muon triggers in Run 2. *JINST*, **15**(09), p. P09015, 2020, arXiv:2004.13447.
- [211] ATLAS COLLABORATION: LuminosityForPhysics, 2016 8.16 TeV proton-lead final January 2019, 2019, URL https://twiki.cern.ch/twiki/bin/viewauth/Atlas/LuminosityForPhysics#2016_8_16_TeV_proton_lead_final.
- [212] COLLABORATION, A.: Luminosity determination in pp collisions at $\sqrt{s} = 8$ tev using the atlas detector at the lhc. *Phys. Rev. D*, 2016.
- [213] ACHARYA, B. E. A.: Estimation of the W+Jets Background for Top Quark Re-Discovery in the Single Lepton+Jets Channel. Technical Report, CERN, Geneva, Oct 2010, supporting document for top paper, URL <https://cds.cern.ch/record/1298781>.
- [214] On the production of a w and jets at hadron colliders. *Nucl. Phys. B*, **357**(1), pp. 32–64, 1991.
- [215] W's, Z's and jets. *Phys. Lett. B*, **154**(5), pp. 435–440, 1985.
- [216] ATLAS COLLABORATION (ATLAS): Measurement of the $t\bar{t}$ production cross-section in pp collisions at $\sqrt{s} = 5.02$ TeV with the ATLAS detector. *JHEP*, **06**, p. 138, 2023, arXiv:2207.01354.
- [217] ATLAS COLLABORATION: Measurement of the $t\bar{t}$ cross section and its ratio to the Z production cross section using pp collisions at $\sqrt{s}=13.6$ TeV with the ATLAS detector. *Phys. Lett. B*, **848**, p. 138376, 2024, arXiv:2308.09529.
- [218] POTEPA, P. A. (ATLAS): Measurement of high- p_T electron performance in proton-lead collisions in the ATLAS experiment. *PoS, LHCP2022*, p. 314, 2023.
- [219] ATLAS COLLABORATION: Jet energy scale and its uncertainty for jets reconstructed using the ATLAS heavy ion jet algorithm. ATLAS-CONF-2015-016, Geneva, 2015, URL <https://cds.cern.ch/record/2008677>.
- [220] KRIVOS, M.; RYBAR, M. & SPOUSTA, M.: Numerical inversion and cross-calibration for heavy-ion jet collections in 2017 pp and 2018 Pb+Pb . Technical Report, CERN, Geneva, 2020, URL <https://cds.cern.ch/record/2706256>.
- [221] ATLAS COLLABORATION: Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7$ TeV. *Eur. Phys. J. C*, **73**(3), p. 2304, 2013, arXiv:1112.6426.
- [222] ATLAS COLLABORATION: Calibration of the light-flavour jet mistagging efficiency of the b -tagging algorithms with Z +jets events using 139 fb^{-1} of ATLAS proton–proton collision data at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, **83**, p. 728, 2023, arXiv:2301.06319.
- [223] AAD, G. E. A. A. C.: 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**(11), 2019.
- [224] 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**, p. 970, 2019, arXiv:1907.05120.

- [225] ATLAS COLLABORATION: Measurement of the c -jet mistagging efficiency in $t\bar{t}$ events using pp collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS detector. *Eur. Phys. J. C*, **82**, p. 95, 2022, arXiv:[2109.10627](#).
- [226] TREXFITTER COLLABORATION: Trexfitter documentation, Accessed 2024, URL <https://trexfitter-docs.web.cern.ch/trexfitter-docs/>.
- [227] CRANMER, K.; LEWIS, G.; MONETA, L.; SHIBATA, A. & VERKERKE, W. (ROOT): HistFactory: A tool for creating statistical models for use with RooFit and RooStats, 2012, URL <https://cds.cern.ch/record/1456844>.
- [228] MONETA, L.; BELASCO, K.; CRANMER, K.; KREISS, S.; LAZZARO, A.; PIPARO, D.; SCHOTT, G.; VERKERKE, W. & WOLF, M.: The roostats project, 2011, arXiv:[1009.1003](#).
- [229] COWAN, G.; CRANMER, K.; GROSS, E. & VITELLS, O.: Asymptotic formulae for likelihood-based tests of new physics. *The European Physical Journal C*, **71**(2), February 2011, ISSN 1434-6052, URL <http://dx.doi.org/10.1140/epjc/s10052-011-1554-0>.
- [230] BARLOW, R. J. & BEESTON, C.: Fitting using finite Monte Carlo samples. *Comput. Phys. Commun.*, **77**, pp. 219–228, 1993.
- [231] ALIOLI, S.; NASON, P.; OLEARI, C. & RE, E.: A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, **06**, p. 043, 2010, arXiv:[1002.2581](#).
- [232] ATLAS, CMS COLLABORATIONS: Combination of inclusive top-quark pair production cross-section measurements using ATLAS and CMS data at $\sqrt{s} = 7$ and 8 TeV. *JHEP*, **07**, p. 213, 2023, arXiv:[2205.13830](#).
- [233] CAMPBELL, J. M.; WACKEROTH, D. & ZHOU, J.: Study of weak corrections to Drell-Yan, top-quark pair, and dijet production at high energies with MCFM. *Phys. Rev. D*, **94**(9), p. 093009, 2016, arXiv:[1608.03356](#).
- [234] COLLABORATION, A.: Observation of $t\bar{t}$ production in the lepton+jets and dilepton channels in p + p collisions at $\sqrt{s_{NN}} = 8.16$ tev with the atlas detector, 2024, arXiv:[2405.05078](#).
- [235] ESKOLA, K. J.; PAAKKINEN, P.; PAUKKUNEN, H. & SALGADO, C. A.: Towards EPPS21 nuclear PDFs. *SciPost Phys. Proc.*, **8**, p. 033, 2022, arXiv:[2106.13661](#).
- [236] KOVARIK, K. ET AL.: nCTEQ15 - Global analysis of nuclear parton distributions with uncertainties in the CTEQ framework. *Phys. Rev. D*, **93**(8), p. 085037, 2016, arXiv:[1509.00792](#).
- [237] DUWENTÄSTER, P.; JEŽO, T.; KLASSEN, M.; KOVAŘÍK, K.; KUSINA, A.; MUZAKKA, K. F.; OLNES, F. I.; RUIZ, R.; SCHIENBEIN, I. & YU, J. Y.: Impact of heavy quark and quarkonium data on nuclear gluon PDFs. *Phys. Rev. D*, **105**(11), p. 114043, 2022, arXiv:[2204.09982](#).
- [238] UBIALI, M. (NNPDF): Towards the NNPDF3.0 parton set for the second LHC run. *PoS, DIS2014*, p. 041, 2014, arXiv:[1407.3122](#).
- [239] ABDUL KHALEK, R.; GAULD, R.; GIANI, T.; NOCERA, E. R.; RABEMANANJARA, T. R. & ROJO, J.: nNNPDF3.0: evidence for a modified partonic structure in heavy nuclei. *Eur. Phys. J. C*, **82**(6), p. 507, 2022, arXiv:[2201.12363](#).
- [240] HELENIUS, I.; WALT, M. & VOGELANG, W.: TUJU21: nuclear PDFs with electroweak-boson data at NNLO. *29th International Workshop on Deep-Inelastic Scattering and Related Subjects*, 7 2022, arXiv:[2207.04654](#).

-
- [241] KLASSEN, M. & PAUKKUNEN, H.: Nuclear PDFs After the First Decade of LHC Data, 11 2023, arXiv:[2311.00450](https://arxiv.org/abs/2311.00450).
- [242] LÉVAI, P.; PAPP, G.; FAI, G.; GYULASSY, M.; BARNAFÖLDI, G.; VITEV, I. & ZHANG, Y.: Discovery of jet quenching at rhic and the opacity of the produced gluon plasma. *Nuclear Physics A*, **698**(1–4), p. 631–634, February 2002, ISSN 0375-9474, URL [http://dx.doi.org/10.1016/S0375-9474\(01\)01445-2](http://dx.doi.org/10.1016/S0375-9474(01)01445-2).
- [243] GYULASSY, M.; LÉVAI, P. & VITEV, I.: Jet quenching in thin quark-gluon plasmas i: formalism. *Nuclear Physics B*, **571**(1–2), p. 197–233, April 2000, ISSN 0550-3213, URL [http://dx.doi.org/10.1016/S0550-3213\(99\)00713-0](http://dx.doi.org/10.1016/S0550-3213(99)00713-0).
- [244] Suppression of hadrons with large transverse momentum in central $au + au$ collisions at $\sqrt{s_{NN}} = 130\text{GeV}$. *Phys. Rev. Lett.*, **88**, p. 022301, Dec 2001, URL <https://link.aps.org/doi/10.1103/PhysRevLett.88.022301>.