



Thèse

2023

Open Access

This version of the publication is provided by the author(s) and made available in accordance with the copyright holder(s).

Cross-section measurement of top-quark pair production with additional charm jets at the ATLAS experiment and Electron identification using machine learning at the ATLAS experiment and Reconstruction of all-hadronic top-quark pair decays using machine learning

Ehrke, Lukas

How to cite

EHRKE, Lukas. Cross-section measurement of top-quark pair production with additional charm jets at the ATLAS experiment and Electron identification using machine learning at the ATLAS experiment and Reconstruction of all-hadronic top-quark pair decays using machine learning. Doctoral Thesis, 2023. doi: 10.13097/archive-ouverte/unige:174210

This publication URL: <https://archive-ouverte.unige.ch/unige:174210>

Publication DOI: [10.13097/archive-ouverte/unige:174210](https://doi.org/10.13097/archive-ouverte/unige:174210)

**Cross-section measurement of top-quark pair
production with additional charm jets at the
ATLAS experiment**

and

**Electron identification using machine learning
at the ATLAS experiment**

and

**Reconstruction of all-hadronic top-quark pair
decays using machine learning**

THÈSE

présentée à la Faculté des sciences de l'Université de Genève
pour obtenir le grade de Docteur ès sciences, mention physique

par

Lukas Fabian Ehrke

de Heidelberg (Allemagne)

Thèse N° 5780

GENÈVE

Atelier d'impression ReproMail

2023



**UNIVERSITÉ
DE GENÈVE**

FACULTÉ DES SCIENCES

DOCTORAT ÈS SCIENCES, MENTION PHYSIQUE

Thèse de Monsieur Lukas Fabian EHRKE

intitulée :

**«Cross-Section Measurement of Top-Quark Pair
Production with Additional Charm Jets at the
ATLAS Experiment
and
Electron Identification Using Machine Learning
at the ATLAS Experiment
and
Reconstruction of All-Hadronic Top-Quark Pair
Decays Using Machine Learning»**

La Faculté des sciences, sur le préavis de Monsieur T. GOLLING, professeur associé et directeur de thèse (Département de physique nucléaire et corpusculaire), Madame A. SFYRLA, professeure associée (Département de physique nucléaire et corpusculaire) et Monsieur J. ERDMANN, professeur (Physics Institute III A, RWTH Aachen University, Aachen, Germany), autorise l'impression de la présente thèse, sans exprimer d'opinion sur les propositions qui y sont énoncées.

Genève, le 15 novembre 2023

Thèse - 5780 -

Le Décanat

Contents

Abstract	vii
Résumé	viii
Contributions by the author	x
Acknowledgements	xi
1 Introduction	1
I Background	3
2 The Standard Model of Particle Physics	5
2.1 Overview of the Standard Model of Particle Physics . .	5
2.2 Top quark physics	12
3 Experimental setup	15
3.1 The Large Hadron Collider	15
3.2 The ATLAS experiment	17
4 Object reconstruction	27
4.1 Interaction of particles with the detector	27
4.2 Base objects	28
4.3 Physics objects	30
4.4 Trigger	39
5 Simulation	41
5.1 Parton distribution functions	42
5.2 Matrix element generation	42
5.3 Decays of heavy particles	43
5.4 Parton shower	43
5.5 Underlying event	44
5.6 Detector simulation	44
5.7 Simulated samples	44
6 Machine learning	53
6.1 General introduction	53
6.2 Neural networks	56
6.3 Convolutional neural networks	59
6.4 Neural networks for point clouds	61
6.5 Performance measures	65
6.6 Preprocessing	66
6.7 Software	67
II Electron identification at the ATLAS experiment	69
7 Current approach	71
7.1 Likelihood algorithm	71

8	Identification of electrons using a deep neural network	75
8.1	Samples and selection	75
8.2	Classes	75
8.3	Inputs	77
8.4	Training procedure	84
8.5	Performance	85
9	Identification of electrons with low- and high-level information	97
9.1	Samples and classes	97
9.2	Input variables	97
9.3	Models	99
9.4	Performance	102
10	Performance in data	109
10.1	Signal selection	109
10.2	Background selection	110
10.3	Performance evaluation	111
III	Measurement of the fiducial and total cross-section of top-quark pairs produced in association with additional charm jets	115
11	Analysis strategy	117
11.1	Profile likelihood fit	117
11.2	Samples	119
11.3	Extraction of the cross-section	120
11.4	Truth-level classification of jets	123
11.5	bc -Tagger	123
12	Event selection	127
12.1	Categorisation of simulated events	127
12.2	Preselection	127
12.3	Fit regions	128
13	Systematic uncertainties	135
13.1	Templates in the profile likelihood fit	135
13.2	Experimental uncertainties	136
13.3	Modelling uncertainties	137
13.4	Pruning	142
14	Results	147
14.1	Fit results with the Asimov dataset	147
14.2	Fit results with ATLAS data	149
15	Combined results	157
15.1	Dilepton channel	157
15.2	Combined fit	159

IV Topological reconstruction of particle physics processes using graph neural networks	169
16 Introduction and motivation	171
16.1 Current approaches	171
16.2 The Topograph	173
17 Topographs for top-quark pair decays in the all-hadronic final state	175
17.1 Datasets	175
17.2 Implementation	178
17.3 Implementation of reference models	182
17.4 Results	183
V Conclusions	195
18 Summary and conclusions	197
VI Appendices	199
A Electron identification with GNNs: hyperparameters	201
B Electron identification: scale factor calculation	203
B.1 Working points	203
B.2 Efficiency determination	203
B.3 Results	206
C <i>bc</i>-Tagger calibration	209
C.1 Object definitions	209
C.2 Additional samples	210
C.3 MC-MC scale factors	211
C.4 Light-flavour jet calibration	213
C.5 <i>b</i> -jet calibration	218
C.6 <i>c</i> -jet calibration	227
C.7 Extrapolation uncertainty	234
D Topograph Hyperparameters	237
List of Figures	239
List of Tables	245
Acronyms	249
References	253

Abstract

This thesis covers three different topics: electron identification, a cross-section measurement of top-quark pair production with additional charm jets, and a novel reconstruction method of event topologies using machine learning. The first two topics are performed within the ATLAS experiment using data from Run 2 of the LHC. The last topic is not using ATLAS data but can easily be applied to physics analyses within the ATLAS experiment.

The first part studies electron identification in the ATLAS experiment using different machine learning techniques. In a first study, high-level variables were used with a deep neural network. The background rejection in simulated proton-proton collisions is increased by a factor of up to 5.5 over the current identification algorithm, while keeping the signal efficiencies the same. A better performance can also be observed in a subset of the data recorded in Run 2, corresponding to 2.4 pb^{-1} . The increased performance in simulated collisions can be further improved by using low-level information like the single cells of the calorimeter and the information of multiple track candidates. Additional improvements of factors between two and three can be achieved.

A measurement of the fiducial cross-section of top-quark pairs with additional charm jets is performed in the second part using the proton-proton collision dataset collected at the ATLAS experiment during Run 2 at $\sqrt{s} = 13 \text{ TeV}$ corresponding to 140 fb^{-1} . The focus is placed on final states with exactly one electron or muon and at least five jets. The final result is obtained by also including final states with two leptons and at least three jets. The cross-section is measured to be $\sigma_{t\bar{t}+\geq 2c}^{\text{fid}} = 1.33^{+0.28}_{-0.25} \text{ pb}$ for events with at least two additional c -jets and $\sigma_{t\bar{t}+1c}^{\text{fid}} = 6.78^{+1.02}_{-0.82} \text{ pb}$ for events with exactly one additional c -jet. These results are compared to the predictions of several generated samples. All samples predict a lower value, though within the uncertainties of the measured cross-sections. Extrapolating the fiducial cross-sections to the full phase space results in cross-sections of $\sigma_{t\bar{t}+\geq 2c}^{\text{total}} = 5.6^{+1.2}_{-1.1} \text{ pb}$ and $\sigma_{t\bar{t}+1c}^{\text{total}} = 41.3^{+6.9}_{-6.4} \text{ pb}$ for events with at least two c -jets and one c -jet, respectively. These values do not include all-hadronic final states.

Lastly, a new method to reconstruct underlying physics processes is introduced in the Topograph. The reconstruction includes intermediate particles by using prior knowledge of particle physics decays. This is done using either graph neural networks or transformer encoder layers. It not only solves the combinatorics of the final state objects but also directly predicts the kinematics of the intermediate particles. The achieved performance matches that of the state-of-the-art machine learning technique in its application to all-hadronic top-quark pair decays.

Résumé

Cette thèse couvre trois sujets différents: l'identification des électrons, la mesure de la section efficace de la production de paires de quarks top avec des jets de charme supplémentaires, et une nouvelle méthode de reconstruction des topologies d'événements à l'aide de l'apprentissage automatique. Les deux premiers sujets sont réalisés dans le cadre de l'expérience ATLAS en utilisant les données du Run 2 de LHC. Le dernier sujet n'utilise pas les données de ATLAS mais peut facilement être appliqué aux analyses de physique au sein de l'expérience ATLAS.

La première partie étudie l'identification des électrons dans ATLAS en utilisant différentes techniques d'apprentissage automatique. Dans une première étude, des variables de haut niveau ont été utilisées avec un réseau neuronal profond. Le rejet du bruit de fond dans les collisions proton-proton simulées est augmenté d'un facteur allant jusqu'à 5,5 par rapport à l'algorithme d'identification actuel, tout en maintenant l'efficacité du signal au même niveau. Une meilleure performance peut également être observée dans un sous-ensemble de données enregistrées dans Run 2, correspondant à 2.4 pb^{-1} . Les performances accrues dans les collisions simulées peuvent être encore améliorées en utilisant des informations de bas niveau telles que les cellules individuelles du calorimètre et les informations des candidats à des pistes multiples. Il est possible d'obtenir des améliorations supplémentaires de l'ordre de deux à trois facteurs.

Une mesure de la section efficace fiduciaire des paires de quarks top avec des jets de charme supplémentaires est effectuée dans la deuxième partie en utilisant l'ensemble des données de collisions proton-proton recueillies par l'expérience ATLAS pendant Run 2 à $\sqrt{s} = 13 \text{ TeV}$ correspondant à 140 fb^{-1} . L'accent est mis sur les états finaux avec exactement un électron ou un muon et au moins cinq jets. Le résultat final est obtenu en incluant également les états finaux avec deux leptons et au moins trois jets. La section efficace est mesurée comme étant $\sigma_{t\bar{t}+\geq 2c}^{\text{fid}} = 1.33^{+0.28}_{-0.25} \text{ pb}$ pour les événements avec au moins deux c -jets supplémentaires et $\sigma_{t\bar{t}+1c}^{\text{fid}} = 6.78^{+1.02}_{-0.82} \text{ pb}$ pour les événements avec exactement un c -jet supplémentaire. Ces résultats sont comparés aux prédictions de plusieurs échantillons générés. Tous les échantillons prédisent une valeur plus faible, bien que dans les limites des incertitudes des sections efficaces mesurées. L'extrapolation des sections transversales fiduciaires à l'ensemble de l'espace des phases donne des sections transversales de $\sigma_{t\bar{t}+\geq 2c}^{\text{total}} = 5.6^{+1.2}_{-1.1} \text{ pb}$ et $\sigma_{t\bar{t}+1c}^{\text{total}} = 41.3^{+6.9}_{-6.4} \text{ pb}$ pour les événements avec au moins deux c -jet et un c -jet, respectivement. Ces valeurs n'incluent pas les états finaux entièrement hadroniques.

Enfin, une nouvelle méthode de reconstruction des processus physiques sous-jacents est introduite dans le Topographe. La reconstruction inclut les particules intermédiaires en utilisant les connaissances préalables des désintégrations de la physique des particules. Pour ce faire, on utilise soit des réseaux neuronaux graphiques, soit

des couches d'encodeurs transformateurs. Elle permet non seulement de résoudre la combinatoire des objets de l'état final, mais aussi de prédire directement la cinématique des particules intermédiaires. Les performances obtenues sont comparables à celles de la technique d'apprentissage automatique de pointe dans son application aux désintégrations de paires de quarks supérieurs entièrement hadroniques¹.

¹ Translated with Ref. [1]

Contributions by the author

Three large topics are covered in this work.

ELECTRON IDENTIFICATION The identification using a DNN is a continuation of the author's master thesis. However, a lot of things have been changed to arrive at the final algorithm. For instance, the models trained during the master thesis were all using binary classification, while all models discussed in this work will use multinomial classification. These studies led to a publication detailing the developed algorithm as well as its performance in simulated events [2]. Additionally, the algorithm was implemented into the ATLAS software stack, making it available to the whole collaboration. This also included the calculation of identification SFs.

The author has not defined the single classes, which was first done with this publication. Colleagues in the collaboration have performed this task.

For the studies involving point cloud-based models, the author performed all the work detailed in this thesis. The CNN architecture was developed and optimised by colleagues in the ATLAS collaboration. This model was only retrained after a slight adjustment using the same dataset for a fair comparison. The work performed is currently being documented in a publication.

$t\bar{t}$ + CHARM ANALYSIS The author took a leading role in the $t\bar{t}$ + charm analysis. The author was the sole analyser in the lepton+jets channel. This includes the production of "ntuples", optimisation of the fit regions and the fit model, and the extraction of the fiducial and total cross-section. The dilepton channel was analysed by fellow PhD students. The author also contributed significantly to the combination of the lepton+jets and dilepton channel. The calibration of the bc -Tagger was almost exclusively done by the author, though it should be noted that only minor adjustments had to be made to already existing code to perform these calibrations. The only calibration step not performed by the author are the MC-MC SFs.

The analysis is still ongoing in the ATLAS experiment at the time of this writing. The shown results are a screenshot of the current status of the analysis. The final results, which will be published in the future, might therefore deviate from the results shown in this thesis.

TOPOGRAPHS Apart from the generation of the used samples and the original idea, all the work has been performed by the author. This includes the training and optimisation of the models, the design of the final network architecture, and the evaluation of the model. The author was also responsible for obtaining the results of the reference models. Ref. [3] details the results obtained with a model based on GNNs. Since the publication, additional work has been performed to adjust the architecture for transformers, which will also be shown in this work.

Acknowledgements

First and foremost, I have to thank my PhD supervisor, Tobias Golling. I am especially grateful for the freedom I had in choosing and pursuing projects.

A special thanks to Johnny, who was there, at least on zoom/skype, throughout my whole PhD thesis. Without your help and input on all projects, this thesis would certainly look very different. I also want to thank Knut and Manuel. Your help with top-quarks, TReX, and the c-jet calibration has been invaluable. And obviously, thanks for not losing touch with the German language.

I would like to thank the e/γ group for allowing me to pursue machine learning for electron identification and always providing feedback. This includes, especially, Stefan, Cyril, and Ximo.

I would like to thank the remaining $t\bar{t}$ + charm analysis group in Luisa, Lisa, Matthew, and Jelena for their input, feedback, and work in the dilepton channel. After lots of thinking, the analysis is hopefully soon ready for publication.

A large thanks to everyone involved in the RODEM project, especially all PhD students located at the AEM in Deb, Malte, Matt, Sam, and Tomke. I also want to thank the remaining PhD students sharing an office with me: Antti, Leon, and Mario. You all made lunch breaks and extracurricular activities much more enjoyable, although sometimes with questionable discussions. And apparently I have to highlight Deb out of these.

1 Introduction

With the discovery of the Higgs boson by the ATLAS and CMS collaborations [4, 5], the last missing piece of the Standard Model of Particle Physics (SM) has been found. The SM describes many observed phenomena of high-energy physics with incredible precision. However, certain phenomena, like dark matter, are not explained by the SM in its current form. The ATLAS collaboration, located at the Large Hadron Collider (LHC) at the European Organisation for Nuclear Research (CERN) in Geneva, Switzerland, is searching for new physics Beyond the Standard Model (BSM), but it is also measuring SM processes with high precision to probe the predictions of the SM. A good description of the SM background processes in these measurements or searches is important for finding new physics. With a better description of these, the sensitivity can be improved.

Top quark pair production with additional heavy-flavour jets is a common background for many interesting final states. These include, for instance, $t\bar{t}H(H \rightarrow b\bar{b})$ or four top-quark production. Top quark pair events with additional b -jets have the exact same final state as $t\bar{t}H(H \rightarrow b\bar{b})$ and only additional jets are needed to have the same final state as events with four top-quarks. As c -jets cannot be separated from b -jets as effectively as light-flavour jets, events with c -jets misidentified as b -jets also contribute to the non-reducible background of these final states. While dedicated measurements of the $t\bar{t}b\bar{b}$ cross-section exist for both ATLAS and CMS [6, 7], ATLAS has not measured the cross-section of $t\bar{t}c\bar{c}$ yet. Certain measurements have, however, included the $t\bar{t}c\bar{c}$ background and show an underprediction in simulation compared to data [8]. The CMS collaboration has measured the cross-section in dilepton final states [9]. This measurement also shows an underprediction of $t\bar{t}c\bar{c}$ in simulation.

In this work, the first measurement of the $t\bar{t}c\bar{c}$ cross-section at the ATLAS detector will be performed. The fiducial as well as the total cross-section will be measured. The focus of this work is placed on decays with one electron or muon. Apart from improving the modelling, this measurement could also lead the way to a search for $t\bar{t}H(H \rightarrow c\bar{c})$. However, due to several reasons, an observation even with the full dataset of the high luminosity LHC is unlikely: $t\bar{t}H(H \rightarrow b\bar{b})$ has not been observed by ATLAS using the full Run 2 dataset [8]; the branching fraction of $H \rightarrow b\bar{b}$ is around twenty times larger than of $H \rightarrow c\bar{c}$; and c -jets are harder to separate from the remaining jets than b -jets.

Another option for improving the sensitivity of measurements and searches is a better identification of physics objects like electrons. Due to the clean signature of electrons in the detector, they are often used in measurements or searches for new BSM physics. By increasing the rejection of objects mimicking the signature of electrons, the purity of a selection can be increased. This, in turn, leads to more

precise measurements and searches. In this thesis, an identification algorithm based on a deep neural network has been developed. The performance is improved over the current algorithm. Additional studies are performed to investigate how additional information, like cells in the calorimeter system, can be best incorporated using more advanced machine learning techniques.

Lastly, the reconstruction of the physics process can help in rejecting background as well as in differential measurements where the kinematics of intermediate particles are used. A new method, the Topograph, is introduced to solve this problem using advanced machine learning techniques like graph neural networks and transformers. It is tested in events with a top-quark pair. Both top-quarks are required to decay hadronically, leading to six quarks in the final state.

This thesis is divided into four parts: Part I covers necessary background knowledge. The SM is introduced in this part, as are the LHC and the ATLAS experiment. Additionally, the object reconstruction at the ATLAS experiment and the setup of samples with simulated events are described. Lastly, a short introduction to machine learning is provided. In Part II, the newly developed electron identification algorithms are discussed. This includes the training of the algorithms as well as estimating their performance in simulation and in data. The cross-section measurement of top-quark pairs with additional charm jets ($t\bar{t} + \text{charm}$) is introduced in Part III. The whole analysis is described for the lepton+jets channel, with results also being covered for a combination of the lepton+jets and dilepton channels. Lastly, Part IV introduces the Topograph. It covers the idea, the training, the evaluation, and the comparison to current approaches as well as state-of-the-art approaches in the all-hadronic channel of top-quark pair production. Additional studies are given in the Appendix in Part VI.

PART I

BACKGROUND

2 The Standard Model of Particle Physics

The theoretical framework describing fundamental interactions between particles will be introduced in this chapter. Utilising this framework, called the Standard Model of Particle Physics, predictions of interactions at the Large Hadron Collider can be made, which will be used throughout this work. The contents of this chapter are mainly based on Ref. [10].

Contents

2.1	Overview of the Standard Model of Particle Physics	5
2.1.1	Particle content	5
2.1.2	Quantum Chromodynamics	7
2.1.3	Electroweak Unification	8
2.1.4	Limitations of the SM	10
2.2	Top quark physics	12
2.2.1	Pair production of top-quarks	12
2.2.2	Top quark decays	13

2.1 Overview of the Standard Model of Particle Physics

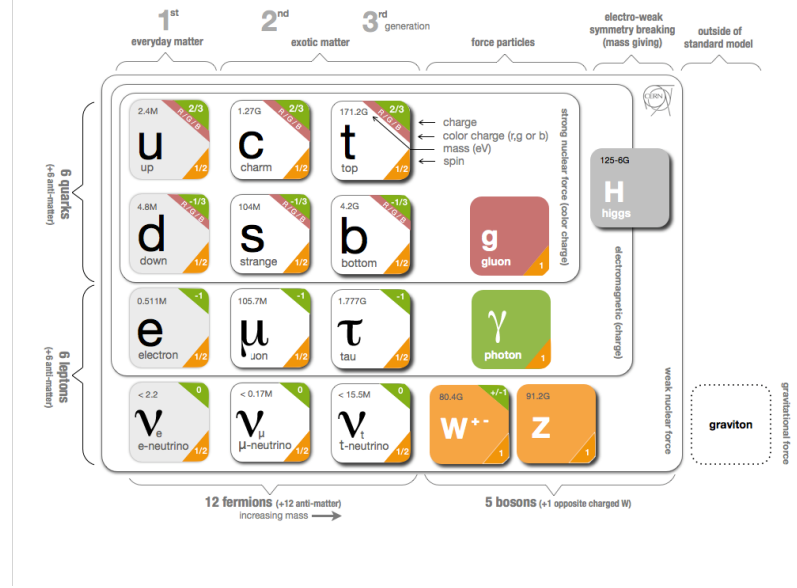
The Standard Model of Particle Physics (SM) is, to date, the best theory describing the fundamental interactions between elementary particles. It describes three of the four fundamental forces: the electromagnetic, weak, and strong forces, only missing gravity. It is a Quantum Field Theory (QFT) with internal symmetries following the $SU(3) \otimes SU(2) \otimes U(1)$ gauge groups, where each of the groups is responsible for one of the forces. The elementary particles have different quantum numbers, which allow them to interact through the different forces. For the strong force, this quantum number is called colour charge; for the weak force, weak isospin and weak hypercharge; and for the electromagnetic force, electric charge. A value different from zero in these quantum numbers makes an interaction through the respective force possible.

2.1.1 Particle content

All known elementary particles, depicted in Figure 2.1, are described by the SM. They can be separated into two groups: the fermions, which all have a spin of $1/2$, and the bosons, which have an integer spin.

The twelve fermions are arranged into three generations. They can also be distinguished into quarks and leptons based on the types of charges they carry. Each generation consists of two quarks and two leptons. The only difference between the particles in the different generations is the mass of the particles; all other quantum numbers are identical. As this is the only difference, particles from the second or third generation eventually decay into particles of the first generation,

Figure 2.1: Overview over the elementary particles of the SM. Taken from Ref. [11]



with particles of the first generation making up all ordinary matter. On top of the twelve depicted fermions, an antiparticle exists for each of the fermions. The antiparticle has the same properties as the regular particle but with inverted charges. As an example, the positron, the antiparticle of the electron and the only antiparticle with a dedicated name, has an electric charge of $+1$ instead of -1 . Typically, referring to a fermion also includes its corresponding antiparticle.

The six quarks interact through all three forces of the SM. They are split into up-type and down-type quarks based on their electric charge. The three up-type quarks, the up, charm, and top-quark, have an electric charge of $2/3$, while the three down-type quarks, the down, strange, and bottom quark, have an electric charge of $-1/3$. The six different types of quarks are also called flavours.

All leptons interact through the weak force. But only three of them interact through the electromagnetic force. These are the electron e , the muon μ , and the tau τ , which all carry an electric charge of -1 . The remaining three, also called neutrinos, do not have an electric charge. Additionally, they are considered massless in the SM. Quarks interact through all forces of the SM.

Four of the bosons are vector bosons with a spin of 1. These are also known as gauge bosons, as they are the mediators of the fundamental forces. Gluons (g) mediate the strong force, $W^\pm$ and Z bosons the weak force, and photons (γ) the electromagnetic force¹. Gluons are massless and carry colour charge. The $W^\pm$ and Z bosons both have a non-zero mass with values of $m_W = 80.377 \pm 0.012$ GeV and $m_Z = 91.1876 \pm 0.0021$ GeV [12]. The $W^\pm$ boson also has an electromagnetic charge. The W^+ and W^- are antiparticles of each other. Lastly, the photon is massless and does not carry any charge.

The last boson is the Higgs boson, which has a spin of zero and is therefore a scalar particle. It does not carry any charge but is

¹ Considering the electroweak unification, $W^\pm$ bosons, Z bosons, and photons all mediate the electroweak force.

responsible for the electroweak symmetry breaking. It has a non-zero mass of $m_H = 125.25 \pm 0.17$ GeV [12] and does not carry any charge. It can interact with all particles that have a non-zero mass.

Figure 2.1 also shows the graviton, which is a hypothetical boson. It is theorised to be a possible force mediator of gravity. However, no QFT of gravity exists so far.

2.1.2 Quantum Chromodynamics

Quantum Chromodynamics (QCD) is the QFT describing the strong interaction. Its underlying symmetry is the $SU(3)$ group. The quantum number associated with the strong interaction is the colour charge. Instead of a number, it is represented by actual colours: red, green, and blue. Antiparticles carry anticolour, so antired, antigreen, and antiblue. The combination of the three colours, the three anticolours, or a colour and its anticolour results in colourless particles.

Quarks and gluons carry colour charge, with the gluons being the force carriers of the strong interaction. Due to the nature of the $SU(3)$ group, a colour octet and a colour singlet are created. The colour octet represents the gluons with colour charge. Therefore, there are technically eight gluons with different combinations of colour charge, which cannot be distinguished as the colour charge cannot be measured. The colour singlet is colourless and has not been observed in nature.

The coupling constant of the strong interaction α_s is of order one at low energies. Therefore, perturbation theory cannot be used in this regime. This affects, for instance, the proton at rest, which consists of several quarks and gluons. The interactions between these objects inside the proton cannot be calculated from first principles.

The fact that gluons carry colour charge leads to several phenomena. When separating two colour-charged particles from each other, the energy stored in the gluon field between these particles will increase with the distance. At some point, it will be more energetically favourable to produce a new quark-antiquark pair than to store more energy in the gluon field. This leads to colour confinement and, in turn, hadronisation. Colour confinement states that colour-charged particles cannot be isolated. Therefore, quarks and gluons always appear in bound states, also called hadrons. To obtain a colourless object, three quarks, three antiquarks, or a quark and an antiquark are needed. States with three quarks are referred to as baryons, while quark-antiquark states are referred to as mesons¹. Hadronisation refers to the formation of hadrons from quarks and gluons.

As the value of the coupling constant decreases with rising energy transfer² and decreasing distance between quarks, quarks can be described as bare particles and not as bound states for large energy transfers. This feature is called asymptotic freedom [13, 14] and allows processes at the LHC to be calculated following perturbation theory.

¹ More exotic states with four or five quarks called tetra- and penta-quarks, respectively, have recently been observed by the LHCb collaboration.

² This energy transfer is also referred to as renormalisation scale μ_R

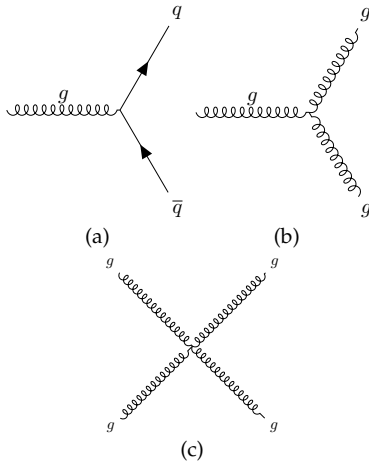


Figure 2.2: Feynman diagrams of gluon interactions. The interaction with (a) a quark-antiquark pair is shown, as well as its self-interaction of (b) three or (c) four gluons.

STRONG VERTICES Feynman diagrams are a pictorial visualisation of the interaction of particles with each other. Particles are represented as edges, while the interaction of particles is represented by vertices. In this work, the time flows from left to right, leaving the initial state on the left side of the diagram and the final state on the right side. Fermions are typically represented by lines with an arrow indicating whether the fermion is a particle or an antiparticle. Fermions travel forward in time, while antifermions travel backward in time. Gluons are represented by spirals, while W bosons, Z bosons, and photons are represented by waves. Scalar bosons, like the Higgs boson are represented by a dashed line. It should be noted that all vertices shown in the following can be rotated so that the initial and final state changes. That means a diagram with a gluon in the initial state and a quark-antiquark pair in the final state can be rotated to have a gluon and a quark in the initial state and a quark in the final state.

Besides being useful for visualisation, Feynman diagrams can also be used to calculate the matrix element of a process, as the arrangement of particles and interaction vertices can be translated into mathematical terms leading to the matrix element. All Feynman diagrams in this work were produced with the TikZ-Feynman package [15].

The gluon only couples to colour-charged particles: quarks and gluons. It can decay into a quark-antiquark pair, where the two quarks have to be of the same flavour. Since the gluon carries colour charge as well, it can also interact with itself, either in a vertex with three or four gluons. The corresponding vertices can be seen in Figure 2.2.

2.1.3 Electroweak Unification

The electroweak unification describes the remaining two forces: weak and electromagnetic. These two forces were initially described separately by the Fermi theory and Quantum Electrodynamics (QED). They are combined following the Glashow-Weinberg-Salam theory [16, 17]. The electroweak unification is described by the $SU(2) \times U(1)$ gauge group. A total of four gauge fields are introduced by this group: W_μ^i with $i = 1, 2, 3$ and B_μ with the W_μ^i originating from the $SU(2)$ group and B_μ from the $U(1)$ group. The W_μ^i generators are the weak isospin I , while the B_μ generator is the weak hypercharge Y with the weak hypercharge being defined as

$$Y = 2(Q - I_3), \quad (2.1)$$

with Q being the electrical charge and I_3 the third component of the weak isospin.

The $SU(2)$ group breaks parity, as the W_μ^i only couple to left-handed fermion fields, which transform as doublets. Right-handed fermion fields transform as singlets instead. Table 2.1 shows the different doublets and singlets, together with their electroweak quantum numbers. Right-handed neutrinos have not been observed yet and therefore are not shown in Table 2.1.

An important detail about the quarks is that the mass eigenstates

Handedness	Generation			Charge		
	I	II	III	Q	I_3	Y
Left	$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L$	$\begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L$	$\begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L$	0 -1	$\frac{1}{2}$ $-\frac{1}{2}$	-1 -1
Right	e_R	μ_R	τ_R	-1	0	-2
Left	$\begin{pmatrix} u \\ d' \end{pmatrix}_L$	$\begin{pmatrix} c \\ s' \end{pmatrix}_L$	$\begin{pmatrix} t \\ b' \end{pmatrix}_L$	$\frac{2}{3}$ $-\frac{1}{3}$	$\frac{1}{2}$ $-\frac{1}{2}$	$\frac{1}{3}$ $\frac{1}{3}$
Right	u_R d'_R	c_R s'_R	t_R b'_R	$\frac{2}{3}$ $-\frac{1}{3}$	0 0	$\frac{4}{3}$ $-\frac{2}{3}$

of the down-type quarks do not correspond to the weak eigenstates¹. This mixing is described by the Cabibbo-Kobayashi-Maskawa (CKM) matrix [18, 19]

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{\text{CKM}} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \quad (2.2)$$

with the mass eigenstates (d, s, b) , the weak eigenstates (d', s', b') , and the CKM matrix V_{CKM} . This mixing allows the W boson to decay into an up- and down-type quark of different generations. As an example, in the decay $\bar{B}^0 \rightarrow D^+ \pi^-$ at tree level, the b -quark decays into a W boson and a c -quark, going from the third to the second generation. The diagonal elements of the matrix are all close to one, such that decays with changing generations are heavily suppressed. This is one reason why b -hadrons typically have a longer lifetime than hadrons containing only up- and down-quarks.

The four gauge bosons of the electroweak interaction, the two W bosons, the Z boson, and the photon, are obtained from the gauge fields following

$$W^\pm = \frac{1}{\sqrt{2}} (W_\mu^1 \mp iW_\mu^2) \quad (2.3)$$

$$\begin{pmatrix} \gamma \\ Z \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix}, \quad (2.4)$$

where θ_W is the Weinberg angle, also called the weak mixing angle.

Due to the gauge invariance of the SM, all gauge bosons should be massless. The W and Z bosons, however, do have a non-zero mass. This is caused by spontaneous symmetry breaking, referred to as the Brout-Englert-Higgs mechanism [20, 21]. The result of this mechanism is, among other things, the Higgs boson.

In the Brout-Englert-Higgs mechanism, an additional complex scalar field doublet with a potential is introduced. This field is also called the Higgs field. By choosing the parameters of the potential accordingly, a non-zero vacuum expectation value can be achieved, which breaks electroweak gauge symmetry. After this breaking of symmetry, the neutral Higgs scalar of the doublet remains. A Higgs-boson-like particle was observed by the ATLAS and CMS collabora-

Table 2.1: The fermions grouped into left-handed doublets and right-handed singlets. Additionally, the quantum numbers of the electrical charge (Q), the third component of the weak isospin (I_3), and the weak hypercharge (Y) are given.

¹ As there are no decays of only up- or down-type quarks that change generation but only decays with one up- and one down-type quark involved, it is not possible to determine whether this mixing only appears for up- or down-type quarks or even both. For simplicity, the mixing is assumed to be solely in the down-type quarks.

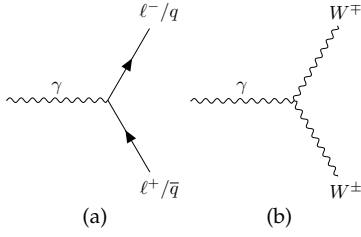


Figure 2.3: Feynman diagrams of photon interactions. The interaction with (a) a fermion-antifermion pair and (b) a W boson pair is shown.

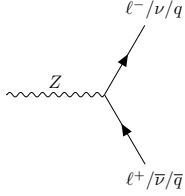


Figure 2.4: Feynman diagram of Z interactions. The interaction with a fermion-antifermion pair is shown.

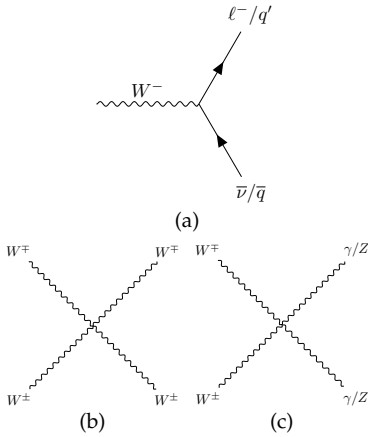


Figure 2.5: Feynman diagrams of W boson interactions. The interaction with (a) a fermion-antifermion pair is shown, as well as the interaction of (b) four W bosons or (c) two W bosons and two Z bosons.

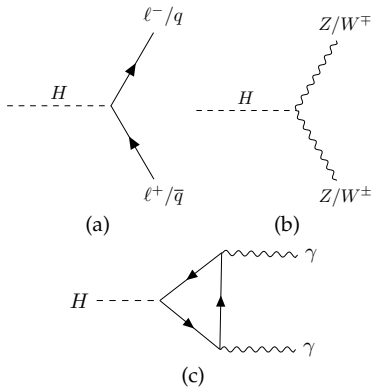


Figure 2.6: Feynman diagrams of Higgs interactions. The interaction with (a) a fermion-antifermion pair and (b) a W boson or Z boson pair is shown, as well as the decay into (c) a photon pair.

tions at CERN in 2012 [4, 5] leading to the Nobel Prize in Physics for Brout, Englert, and Higgs.

ELECTROWEAK VERTICES The photon interacts with every particle that has an electric charge. As it itself does not have an electric charge, and the electric charge has to be conserved, only a limited number of vertices are possible. The photon can decay into a fermion-antifermion pair, where the fermions have to be charged, so neutrinos are excluded. But, it can also decay into a pair of W bosons. These possible vertices are depicted in Figure 2.3. They need, however, to be a part of a larger Feynman diagram as the photon is massless.

The Z boson behaves very similar to the photon, but instead of the electric charge, it couples to the weak hypercharge. In fact, all vertices, including a photon, are also possible when replacing the photon by a Z boson. Additionally, neutrinos are included in the fermion-antifermion vertex, as shown in Figure 2.4, as they carry weak hypercharge.

W bosons also decay into fermion-antifermion pairs. Since the W boson carries an electric charge, these fermions cannot be particle-antiparticle pairs. They can either be a charged lepton and its neutrino or an up-type quark together with a down-type quark. Lastly, there are two possible vertices that only include electroweak bosons. Either two W bosons scatter off of each other, or two W bosons together create two photons or Z bosons. All of these vertices can be seen in Figure 2.5.

As the Higgs boson is introduced by the breaking of the electroweak gauge symmetry, its decay vertices will also be covered here. It can decay, as the photon and Z boson, into a fermion-antifermion pair, excluding neutrinos as they are considered massless in the SM. As the Higgs boson does not carry an electrical charge, the fermions have to be particle and antiparticle of each other. The Higgs boson can also decay into a pair of Z or W bosons. Lastly, it can also decay into a pair of photons or gluons. Since both of these are massless, the decay has to include a loop of fermions. For the decay into photons, the loop can also consist of W bosons. The Higgs can also couple to itself in a three- or four-Higgs vertex. Lastly, two Higgs bosons can interact in one vertex with two weak bosons. Some of these vertices can be seen in Figure 2.6.

2.1.4 Limitations of the SM

As already mentioned, the SM is an extremely successful theory in that its predictions match the observed values very well. Figure 2.7 shows a summary of cross-section measurements performed by the ATLAS Collaboration during Run 1 and Run 2 of the LHC which all agree with the SM prediction. However, there are several phenomena that cannot be explained by the SM.

Standard Model Total Production Cross Section Measurements

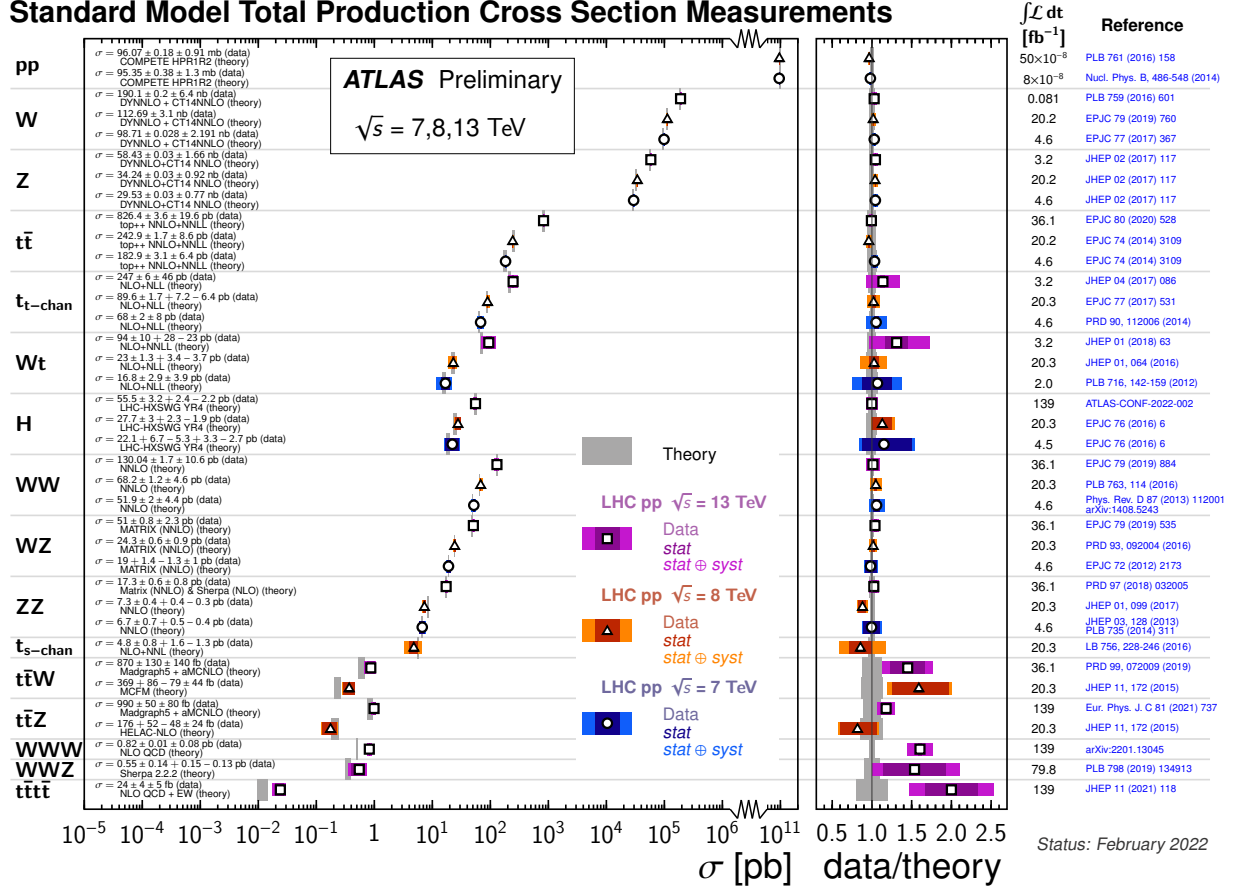


Figure 2.7: Summary plot of cross-section measurements of the ATLAS Collaboration during Run 1 and Run 2. Taken from Ref. [22].

NEUTRINO MASSEES The SM treats neutrinos as massless. However, neutrino oscillations have been measured by several experiments [23, 24]. For instance, an electron neutrino, which is produced in addition to an electron, can oscillate into a neutrino of a different flavour, such as a muon neutrino. These oscillations are only possible if the neutrinos have a non-zero mass. They are caused by a similar phenomenon as the CKM matrix. For neutrinos, the mass eigenstates and the flavour eigenstates are not the same. The interaction of neutrinos is determined by their flavour eigenstate, while their propagation is determined by their mass eigenstate. The mixing of these is described by the Pontecorvo-Maki-Nakagawa-Sakata matrix [25, 26].

DARK MATTER Several astrophysical measurements show that the rotation velocity of galaxies is too fast for the matter observed in these galaxies. A solution proposed to solve this problem is so-called dark matter. Dark matter is theorised to consist of so far unobserved particles that only interact through gravitational or weak interactions. Additionally, they have to have a high mass. One candidate for these would be right-handed massive neutrinos. However, dark matter is only one possible solution to this problem. Another solution is a modification of the theory of gravity.

UNIFICATION OF FUNDAMENTAL FORCES The electromagnetic interaction and the weak interaction were originally developed as separate interactions. They were only later combined following the electroweak unification. As two forces could be combined into one, there is a desire by many physicists to also include the strong interaction into the same force to leave only one force in the SM. For this unification to be possible, the gauge couplings of all forces need to intersect at some energy, which they do not according to the SM. By extending the SM with, for instance, supersymmetry, this unification can be made possible. However, there is no underlying reason for such a unification to be possible other than the desire of physicists for a “beautiful” theory.

2.2 Top quark physics

The top-quark is the last quark of the SM to be discovered. The DØ and CDF collaborations both announced its discovery in 1995 [27, 28]. It is the up-type quark of the third generation and forms a weak isospin doublet with the bottom quark. Its discovery was considerably later than for the remaining quarks, caused by the high mass of the top-quark of $m_t = 172.69 \pm 0.30$ GeV [12]. This high mass has additional effects. The Yukawa coupling of the top-quark with the Higgs field is close to unity. This large coupling has a considerable impact on the production modes and branching fractions of the Higgs boson through virtual loops. Additionally, the high mass leads to a decay width of $\Gamma_t = 1.42^{+0.19}_{-0.15}$ GeV, which in turn translates to a lifetime of order $\tau_t \sim 10^{-24}$ s. This lifetime is shorter than the timescale of the hadronisation process. Therefore, the top-quark uniquely does not form hadrons but decays before they can be formed. This also means that by studying the properties of the top-quark, the properties of a single quark that is not in a bound state can be studied.

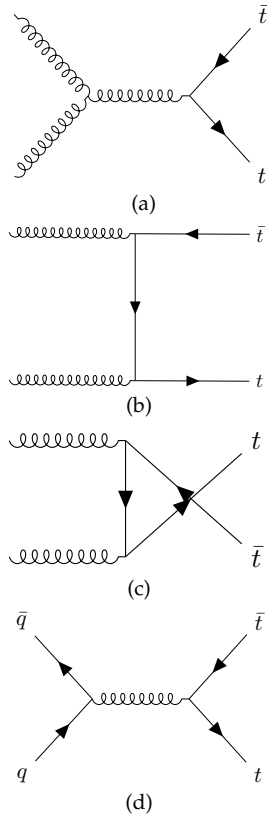


Figure 2.8: Leading order Feynman diagrams of top-quark pair production. The production through a gluon-gluon pair in the (a) s-, (b) t-, and (c) u-channel is shown, as well as the production through (d) quark-antiquark annihilation.

2.2.1 Pair production of top-quarks

Most top-quarks at hadron colliders are produced in top-antitop-quark pairs. The pairs can either be produced through gluon-gluon fusion or quark-antiquark annihilation. At the Tevatron, most pairs were produced by quark-antiquark annihilation, whereas at the LHC at a centre-of-mass energy of $\sqrt{s} = 13$ TeV around 90% were produced through gluon-gluon fusion. This difference has to do with the particles colliding and the centre-of-mass energy. The Tevatron collided protons with antiprotons. Therefore, an antiquark can be taken from the valence quarks of the antiproton, which carry on average a higher momentum fraction than the sea quarks. Additionally, due to the lower centre-of-mass energy at the Tevatron with $\sqrt{s} = 1.96$ TeV higher momentum fractions are needed to be able to produce two top-quarks. At the LHC, all antiquarks have to be sea quarks. Figure 2.8 shows all tree-level diagrams of the production of a top-quark pair at the LHC.

Using a top-quark mass of $m_t = 172.5$ GeV, the cross-section for top-quark-pair production at the LHC at $\sqrt{s} = 13$ TeV can be calculated to be $\sigma(t\bar{t}) = 832 \pm 51$ pb. This calculation is performed by Top++ 2.0 [29–35] at next-to-leading order (NLO) in QCD. Next-to-next-to-leading logarithmic (NNLL) soft gluon terms are included via resummation. The uncertainties of the calculated cross-section include Parton Distribution Function (PDF) and scale uncertainties. The PDF4LHC15 prescription [36] with the MSTW2008_{NNLO} [37, 38], CT10_{NNLO} [39, 40], and NNPDF2.3_{LO} PDF sets in the five flavour scheme (5FS) is used to estimate the PDF uncertainties.

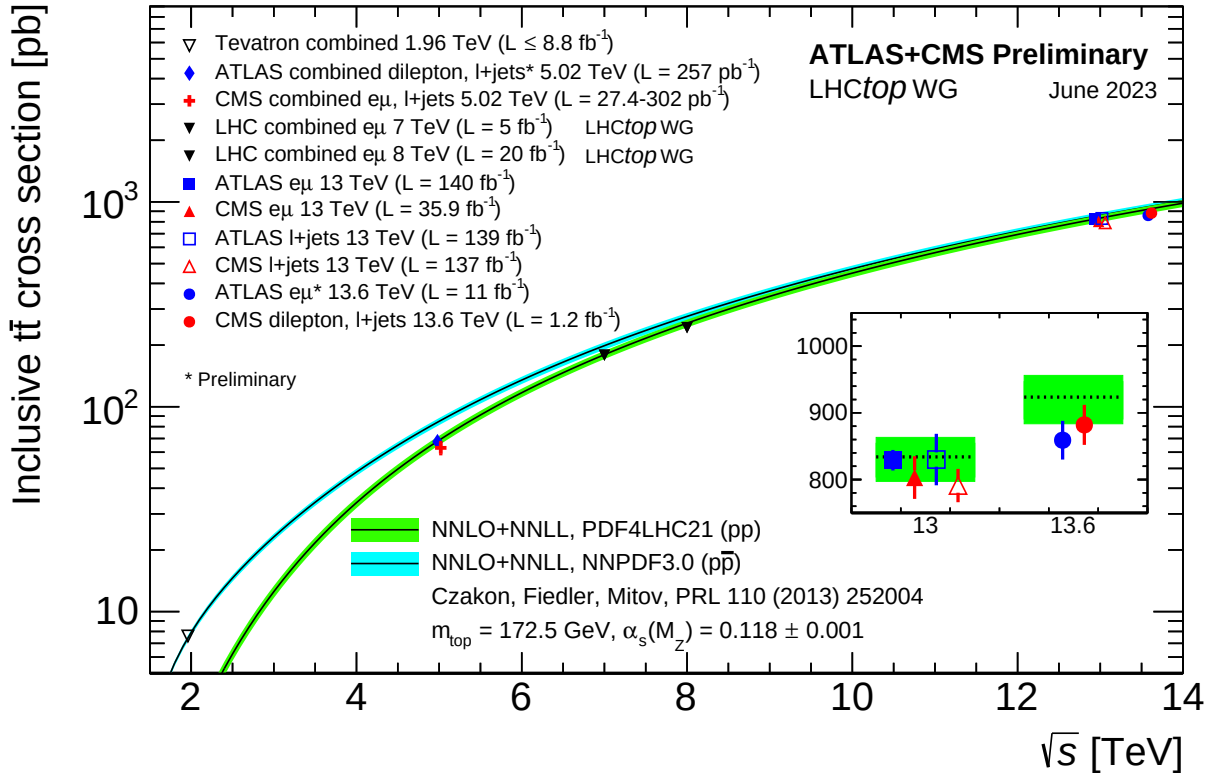


Figure 2.9: Summary plot of top-quark-pair cross-section measurements at the Tevatron and the LHC. Taken from Ref. [41].

The cross-section was measured at a wide variety of centre-of-mass energies at the Tevatron and the LHC in different channels. The results, as well as the theoretical predictions, can be seen in Figure 2.9. All measurements are in agreement with the predictions by the SM.

2.2.2 Top quark decays

Due to the values of the elements of the CKM matrix, the top-quark decays almost exclusively into a W boson and a b -quark. Decays into a W boson and a d - or s -quark have not been observed yet. The W boson then decays further. Around $2/3$ of W bosons decay into a quark-antiquark pair, also called hadronically decaying, while the remaining $1/3$ decay into a charged lepton and its corresponding

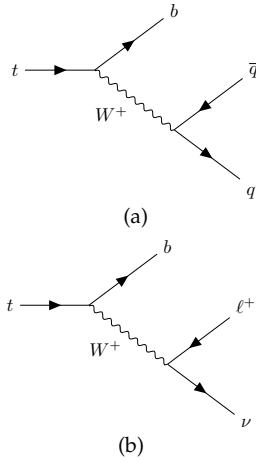


Figure 2.10: Feynman diagrams of the decay of a top-quark. The decay with a (a) hadronically and (b) leptonically decaying W boson are shown.

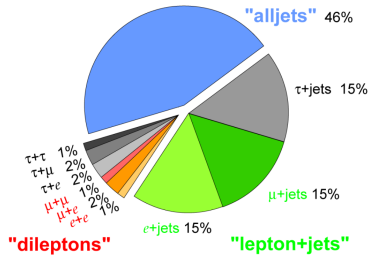


Figure 2.11: Decay modes of top-quark pairs with their branching fractions. Taken from Ref. [42]

neutrino, also called leptonically decaying. Feynman diagrams of these two decay modes of the top-quark can be seen in Figure 2.10.

For top-quark pairs, this leads to the distinction of three possible final states or channels: the all-hadronic channel, where both W bosons decay hadronically; the lepton+jets channel, where one W boson decays hadronically and the other leptonically; and the dilepton channel, where both W bosons decay leptonically. The corresponding branching fractions can be seen in Figure 2.11.

For the dilepton and lepton+jets channels, the decays without a tau lepton are highlighted specifically. This is because, contrary to the electron and muon, the tau decays within the detector volume, making the reconstruction more difficult. The benefits of the dilepton and lepton+jets channels are their clean signatures due to the electrons and muons in the final state. This would not be the case for decays including taus.

The cleanest signature with very little background contamination can be achieved in the dilepton channel with one electron and one muon. Due to the different flavours of the leptons, no Z boson decay has the same signature, which is the case for the ee or $\mu\mu$ channel. However, the two neutrinos are not detectable within the ATLAS detector, leading the $t\bar{t}$ system to be not reconstructable.

The all-hadronic channel can be fully reconstructed as there are no neutrinos in the final state. Additionally, it has the largest branching fraction. However, large background contributions from QCD and no clean trigger signature provide difficulties in this channel. Still, measurements in the all-hadronic channel show, in some regions of phase space, a better resolution compared to the lepton+jets channel [43].

3 Experimental setup

In this chapter, the experimental setup of the ATLAS experiment located at the LHC at CERN close to Geneva, Switzerland, will be discussed. First, the LHC will be described before turning attention to the ATLAS experiment.

Contents

3.1	The Large Hadron Collider	15
3.2	The ATLAS experiment	17
3.2.1	The Inner Detector	18
3.2.2	Calorimeter system	20
3.2.3	Muon Spectrometer	23
3.2.4	Magnet system	24
3.2.5	Trigger and Data Acquisition	24

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [44] is, to date, the largest and most powerful human-made particle accelerator. Located under the CERN research lab near Geneva, Switzerland, it crosses the Franco-Swiss border at a total of four points. It is a circular collider built on the principle of a synchrotron. The ring has a circumference of 26.7 km. The LHC was designed to reach a centre-of-mass energy of $\sqrt{s} = 14$ TeV which so far has not been reached. During Run 3², the highest energy so far of 13.6 TeV was reached. During Run 2³, a centre-of-mass energy of 13 TeV was reached, while during Run 1⁴ the accelerator ran at 7 TeV and 8 TeV.

Protons circulate in two beam pipes in opposite directions. They are brought to collision at four interaction points. To keep the protons on a quasi-circular path, some 1200 superconducting dipole magnets are located around the ring. The acceleration of the protons is performed by sixteen radio-frequency cavities, eight in each direction. To keep the beam focused and correct imperfections in the beam geometry, additional quadrupole magnets and magnets of higher orders are used.

Figure 3.1 shows the accelerator complex at CERN as of 2022. The several accelerators that the protons have to go through before the LHC can be seen. The protons are obtained from hydrogen atoms by stripping them of their electrons. The now bare protons are accelerated in the Linear Accelerator 4 (LINAC4)⁵ to 50 MeV. Following the LINAC4, the protons are injected into the BOOSTER, the first circular accelerator. Here, the protons reach an energy of 1.4 GeV. Afterwards, they are further accelerated by the Proton Synchrotron and the Super Proton Synchrotron (SPS) to 25 GeV and 450 GeV, respectively, before being injected into the LHC.

Due to this injection chain, no continuous beams are accelerated but rather single bunches of protons, each containing around 10^{11} protons.

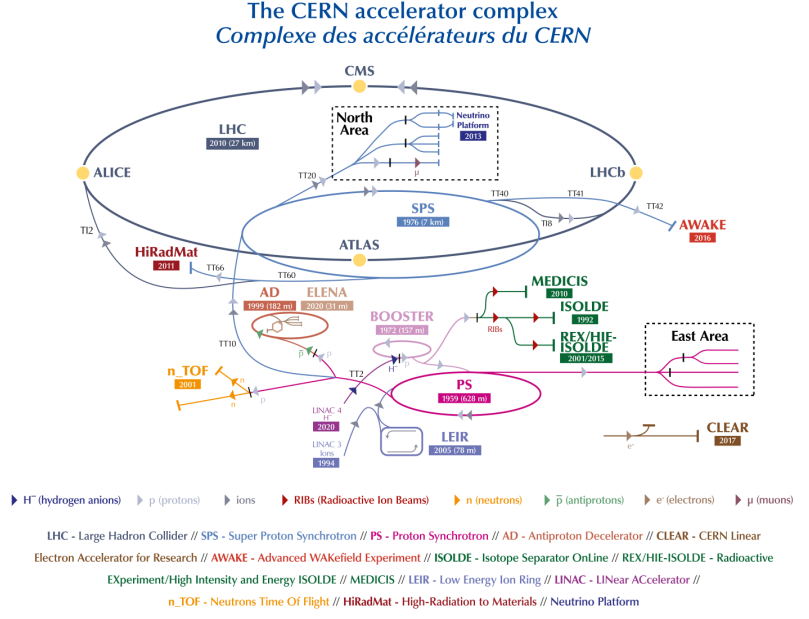
² The data-taking period, which started in 2022

³ Between 2015 and 2018

⁴ Between 2010 and 2012

⁵ The LINAC4 replaced the LINAC2 during the shutdown between 2018 and 2022. As all the data used in this thesis was collected during Run 2, none of the protons used for this thesis were accelerated by the LINAC4.

Figure 3.1: Current accelerator complex at CERN. Taken from Ref. [45]



Depending on the configuration, around 2800 bunches per beam can be filled into the LHC. This corresponds to a time difference between the bunches, and therefore the collisions, of 25 ns. The collisions happen at four interaction points, where four major experiments are located: ATLAS [46], Compact Muon Solenoid (CMS) [47], A Large Ion Collider Experiment (ALICE) [48], and LHCb [49]. ATLAS and CMS are large general-purpose detectors. They perform measurements of SM parameters and searches for new particles. The Higgs boson was discovered at these two experiments in 2012, completing the SM [4, 5]. ALICE is a detector specialised in collisions of heavy ions, whereas LHCb is a forward detector mainly studying b -hadrons.

The rate at which particles are produced depends on the instantaneous luminosity $\mathcal{L}$, which is a characteristic of the collider. The design instantaneous luminosity of the LHC is $\mathcal{L} = 10^{34} \text{ cm}^{-2}\text{s}^{-1}$. To obtain the number of times N a certain physics process happens during a period of time, the instantaneous luminosity has to be integrated over time to obtain the integrated luminosity L . The integrated luminosity then needs to be multiplied by the cross-section σ specific to this physics process

$$N = \sigma \cdot \int \mathcal{L} dt = \sigma \cdot L. \quad (3.1)$$

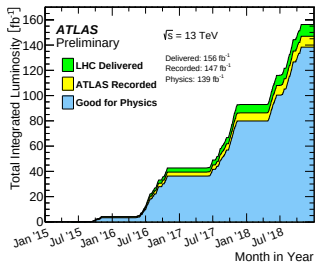


Figure 3.2: Integrated luminosity accumulated at the ATLAS experiment over the course of Run 2. The plot shows the total luminosity delivered by the LHC, the fraction recorded by ATLAS, and the fraction deemed good enough for physics analyses within ATLAS. Taken from Ref. [50].

The integrated luminosity in the ATLAS experiment during Run 2 can be seen in Figure 3.2. The ‘Good for Physics’ dataset is used throughout this thesis. Only events during stable data-taking conditions are considered in this dataset. Additionally, events during times in which the detector was not working properly are removed. The value for the integrated luminosity in Figure 3.2 is given with 139 fb^{-1} , though it will vary between 139 fb^{-1} and 140 fb^{-1} in this

work. This is caused by an updated calculation that was recently published [51], which leads to a value of 140 fb^{-1} , whereas a previous calculation yielded 139 fb^{-1} . If the value of 139 fb^{-1} is quoted, the values of the old calculation were used. This should, however, not impact these results.

To increase the instantaneous luminosity and therefore the rate of physics processes, not only one but multiple collisions happen in each bunch-crossing. Due to the low cross-section of most inelastic processes, almost all collisions will have at most one inelastic interaction. The remaining colliding protons are commonly referred to as in-time pileup. These usually leave many additional particles with low energies in the detector, which are considered noise. More noise is also caused by out-of-time pileup. These are signals in the detector that are caused by neighbouring bunch-crossings. Figure 3.3 shows the mean number of interactions per bunch-crossing $\langle\mu\rangle$ which is a way to quantify the effect of pileup. Different pileup profiles can be seen for the four years of data-taking. The overall mean number of interactions per crossing is $\langle\mu\rangle = 33.7$

3.2 The ATLAS experiment

The ATLAS experiment is one of two general-purpose detectors at the LHC, the other one being CMS. It is located at Point-1 of the LHC around 100 m below the surface. It is the largest detector at the LHC with a diameter of 25 m, a length of 44 m, and a weight of approximately 7000 t. The structure of the detector with its many subcomponents can be seen in Figure 3.4. It is built cylindrically around the beam pipe, covering almost the full 4π solid angle around the interaction point. Located closest to the beam line is the Inner Detector (ID), used to detect the trajectory of charged particles. The calorimeter system encloses the ID and is used to measure the energy of charged and neutral particles. The last layer of components is the Muon Spectrometer (MS) which detects muons and measures their

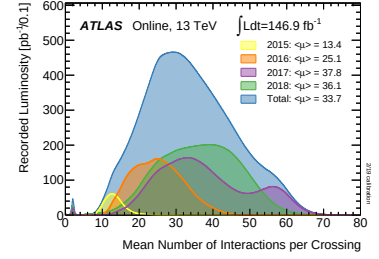


Figure 3.3: The mean number of interactions per bunch-crossing at ATLAS during Run 2. Taken from Ref. [50].

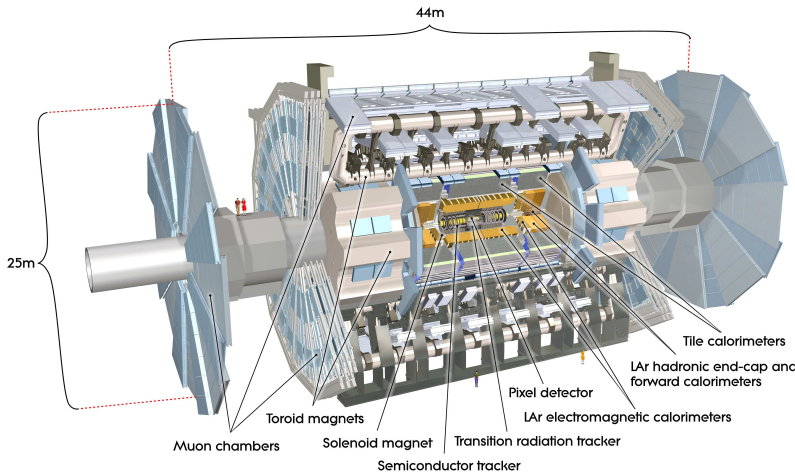


Figure 3.4: Overview of the ATLAS detector and its subcomponents. Taken from Ref. [46].

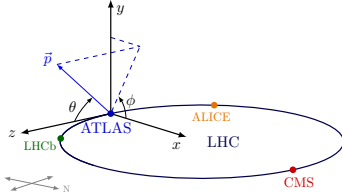


Figure 3.5: Coordinate system of the ATLAS detector in context of the LHC ring. Taken from Ref. [52].

transverse momentum. To obtain the transverse momentum of the particles in the ID and the muons in the MS, two magnetic systems are included in the detector: a solenoid magnet that encloses the ID and a system of toroid magnets embedded in the MS.

Coordinate system

Particles in ATLAS are described by an ATLAS-specific right-handed coordinate system. The origin is placed at the interaction point. The z -axis is defined along the beamline, the y -axis towards the surface, and the x -axis towards the centre of the LHC. This ensures that the x - y plane is transverse to the beam axis. A visualisation of the coordinate system can be seen in Figure 3.5.

Particles are usually described by their four momentum vector. The standard representation comprises four quantities: The transverse momentum of a particle is defined as $p_T = \sqrt{p_x^2 + p_y^2}$ with p_x and p_y being the x - and y -components of the momentum, respectively. As the protons are accelerated only in the z -direction, the initial transverse momentum of the collision is known to be zero. The second coordinate is the azimuthal angle ϕ with respect to the x -axis. As the third quantity, the rapidity

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) \quad (3.2)$$

of the particle with E the energy and p_z the momentum in the z -direction is used. The rapidity has one distinct advantage over the polar angle θ as it is invariant under Lorentz boosts along the beam axis. An approximation of the rapidity in the case of a small mass compared to the momentum of the particle ($m \ll |\vec{p}|$) is the pseudorapidity $\eta = -\ln(\tan(\theta/2))$ which can be calculated by measuring only the polar angle. For massless particles, rapidity and pseudorapidity are exactly the same. As the last quantity, the energy of the particle is most commonly used.

Due to the usage of azimuthal angle ϕ and pseudorapidity η , distances between two objects are calculated using

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}, \quad (3.3)$$

where $\Delta\phi$ and $\Delta\eta$ are the distances in ϕ and η , respectively. The distance can also be calculated using the actual rapidity. In this case, it will be denoted by ΔR_y .

3.2.1 The Inner Detector

The Inner Detector (ID) [53, 54] provides information of the trajectories of charged particles for the ATLAS experiment based on particle hits in its subcomponents. These hits are used to reconstruct the trajectories, also called tracks. It is located closest to the beam pipe out of all components of the ATLAS detector. Due to the magnetic field of the solenoid magnet, it also provides a measurement of the

transverse momentum of charged particles and an identification of the spatial origin of multiple particles. An overview of the ID can be seen in Figure 3.6. It consists of three subcomponents, ordered by the distance from the beam pipe: the Pixel detector, the Semiconductor Tracker (SCT), and the Transition Radiation Tracker (TRT). Each of them consists of a barrel component, which is built like a cylinder around the beam pipe, and two endcap components, which are built like discs or wheels perpendicular to the beam axis.

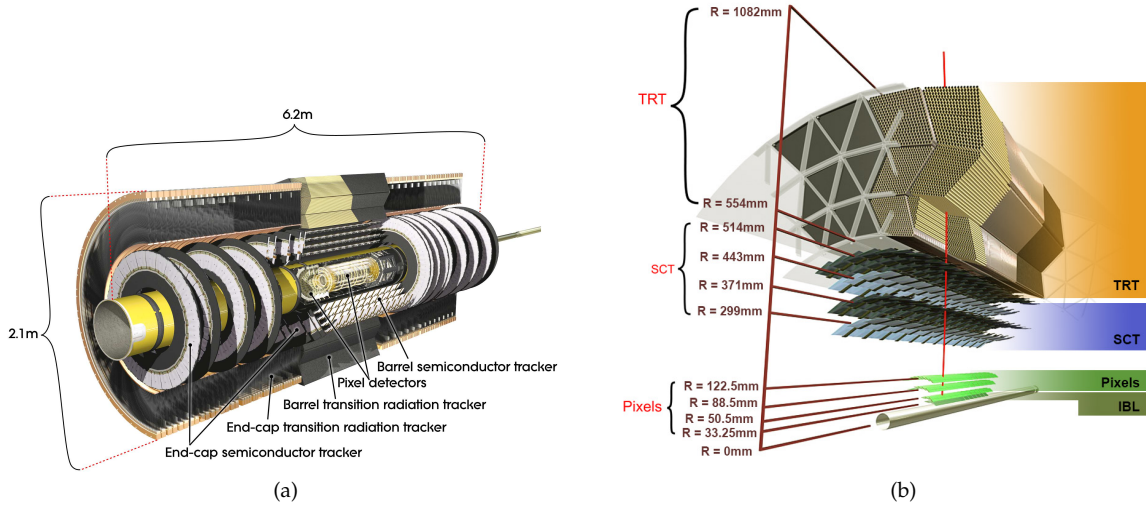


Figure 3.6: Overview of the ATLAS ID. The IBL which was installed before Run 2 is only included in (b). Taken from Ref. [46, 55].

Pixel detector

The Pixel detector consists of four layers in the barrel and three discs in each of the endcaps. The closest layer to the beam pipe, the Insertable b-Layer (IBL) [56] was installed between Run 1 and Run 2. It can be seen in Figure 3.6b though it is missing from Figure 3.6a. The Pixel detector is a silicon detector comprised of pixels. Located between 33.25 mm and 122.5 mm from the beam pipe, it covers a range up until $|\eta| < 2.5$. The pixels have a size of $50\ \mu\text{m}$ in the $R - \phi$ direction and $250\ \mu\text{m}$ or $400\ \mu\text{m}$ in the z -direction for the IBL and the remaining layers, respectively. Due to the small size of the pixels, the Pixel detector provides the best hit resolution of the tracking detectors with $8\ \mu\text{m}$ ($10\ \mu\text{m}$) in the $R - \phi$ direction for the IBL (remaining layers). In the z -direction, a resolution of $40\ \mu\text{m}$ and $115\ \mu\text{m}$ can be achieved. Because of this high resolution, the Pixel detector plays a crucial role in the reconstruction of tracks and decay vertices.

Semiconductor Tracker

Like the Pixel detector, the Semiconductor Tracker (SCT) is built using silicon sensors but consists of strip detectors instead of pixels. It comprises four layers in the barrel region and nine discs in each of the endcaps. Each of the layers incorporates two sets of strips, one set aligned with the beam axis and one set rotated by $40\ \text{mrad}$. The combination of hits in these two sets not only provides good

resolution in the ϕ coordinate but also in the z coordinate. In fact, this stereoscopic alignment is required for obtaining three-dimensional space points when using strip sensors. The strips have a size of $80\text{ }\mu\text{m} \times 12\text{ cm}$ leading to a resolution of $17\text{ }\mu\text{m}$ in the $R - \phi$ direction and $580\text{ }\mu\text{m}$ in the z -direction. In total, the SCT covers the same range as the Pixel detector up to $|\eta| < 2.5$.

Transition Radiation Tracker

The Transition Radiation Tracker (TRT) is, in contrast to the aforementioned silicon detector systems, a gaseous detector system. It consists of straw tubes filled with a gas mixture of Xe (70%), CO₂ (27%), and O₂ (3%)¹, which are placed parallel to the beam pipe in the barrel and radially in both endcaps. As the tubes only point in one direction, only $R - \phi$ information can be obtained from the hits with a resolution of $120\text{ }\mu\text{m}$ in the barrel and $130\text{ }\mu\text{m}$ in the endcap. Contrary to the silicon detector systems, the TRT only covers a range up until $|\eta| < 2$.

In addition to tracking information, the TRT provides electron identification by means of transition radiation. Particles with a high Lorentz factor ($\gamma > 1000$) entering the polypropylene, in which the straw tubes are enclosed, will create transition radiation photons. Xenon is efficient at absorbing these photons, leading to higher signal pulses being read out for highly relativistic particles. As electrons have a much higher Lorentz factor than, for instance, charged pions at typical energies observed in ATLAS, high threshold hits² can be used to distinguish electrons from charged pions. Figure 3.7 shows the probability of producing a high threshold hit based on the Lorentz factor. This identification ability is reduced in straws that use the argon gas mixture, as argon has a lower efficiency to absorb the transition radiation photons [57].

¹ Due to leaks in some tubes, the xenon has been replaced by argon for these tubes [57].

² Hits where the signal pulse exceeds a predefined amplitude

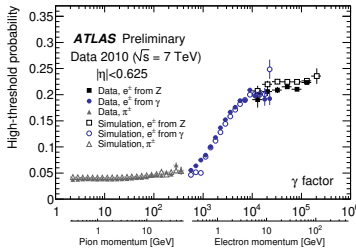


Figure 3.7: High threshold probability of a particle in the TRT depending on the Lorentz factor γ of the particle. Taken from Ref. [58].

³ Muons can lose some of their energy in the calorimeter system but will generally not be stopped by it, while neutrinos will always leave undetected.

3.2.2 Calorimeter system

The calorimeter system absorbs most particles³ while allowing for a measurement of their energy. Additionally, properties of their showers can be measured. These are important for particle identification, especially electron and photon identification. Showers arise when highly energetic particles interact with dense material. During the interaction, cascades of secondary particles are created, which in turn can create additional particles.

Both calorimeter systems in ATLAS, the electromagnetic [59] and the hadronic calorimeter [60], are sampling calorimeters. Figure 3.8 shows an overview of both calorimeters. Layers of active material (liquid argon or plastic scintillators) alternate with layers of passive material (copper, iron, tungsten, or lead). The active material is used to measure the energy deposited by particles, while the passive material is responsible for inducing the showers.

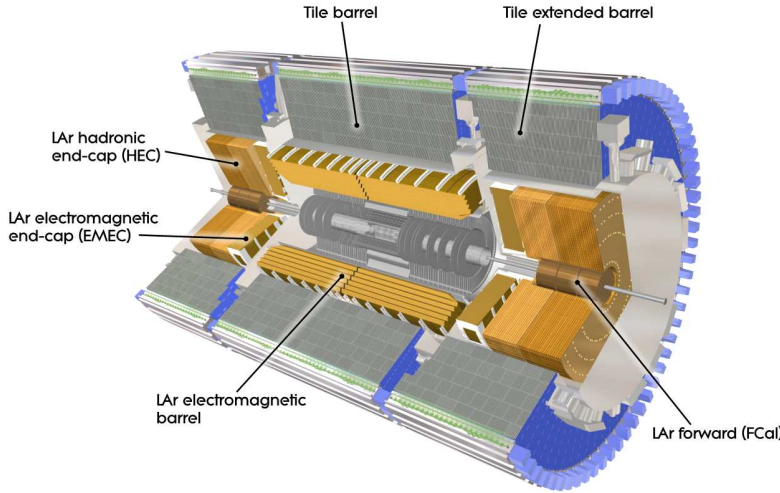


Figure 3.8: Overview of the ATLAS calorimeter system. Taken from Ref. [46].

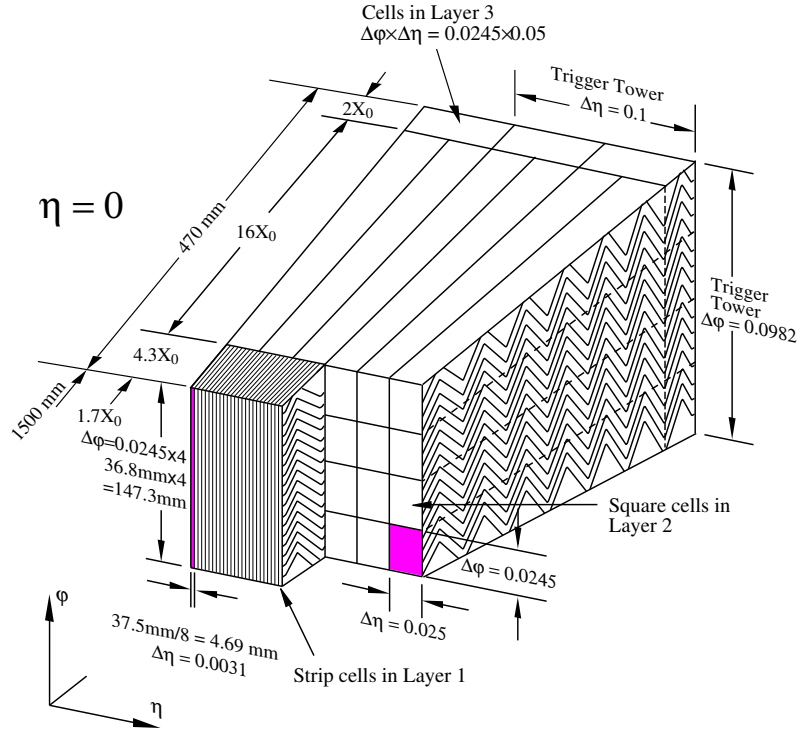
Electromagnetic Calorimeter

The Electromagnetic Calorimeter (ECAL) is designed to stop electrons and photons and measure their energies. It is built using Liquid Argon (LAr) as the active material and lead as the passive material. A slice of the ECAL at a pseudorapidity of zero can be seen in Figure 3.9. It has an accordion shape in ϕ to leave no dead space without any active material. In total, it has four layers. The very first layer, which is not shown in Figure 3.9 is called the presampler. Its main purpose is to correct for energy lost upstream of the ECAL. It has limited coverage in pseudorapidity ($|\eta| < 1.8$) and only consists of active material. The remaining three layers all cover the same range as the ID ($|\eta| < 2.5$) with the first and second layers extending further up until $|\eta| < 3.2$, however with reduced granularity. As can be seen from Figure 3.9, the three layers have different cell sizes. The first layer has very good η resolution to resolve photons coming from the decay of neutral pions. The cells of the second layer have almost a square size in $\Delta\eta \times \Delta\phi$ with 0.25×0.245 . It is also the deepest layer and absorbs most of the energy of electrons and photons. The third layer has a slightly lower resolution in η than the second layer and is mainly used to measure potential leakage of electromagnetic showers into the hadronic calorimeter. On top of the different cell sizes across the single layers, the cell sizes do not stay the same over the whole range of the calorimeter but change even inside a layer.

Similar to the ID, the ECAL comprises a barrel cylinder extending up to $|\eta| < 1.475$ ¹ and two endcap wheels covering $1.375 < |\eta| < 3.2$. Its thickness expressed in radiation lengths X_0 , that is, the distance over which an electron will reduce its energy by a factor of $1/e$, is at least $22 X_0$ and $24 X_0$ in the barrel and the endcaps, respectively.

¹ Not all layers extend to the same range in $|\eta|$

Figure 3.9: Overview of the ATLAS electromagnetic calorimeter at $\eta = 0$. The presampler is not shown. The accordion shape in ϕ is visible. Taken from Ref. [46].



Hadronic Calorimeter

The main purpose of the Hadronic Calorimeter (HCAL) is to measure the energy of hadronic particles. It consists of a barrel component, an extended barrel component, and two endcap components.

The barrel and the extended barrel are built of steel as passive material and plastic scintillator tiles as active material, arranged in three layers each. The first two layers have the same cell size of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$, while the cells of the third layer are twice as large in the η -direction. The barrel covers a range up until $|\eta| < 1.0$ while the extended barrel extends $0.8 < |\eta| < 1.7$.

The two endcaps use the same technology as the ECAL with LAr as active material, though copper is used as passive material. They incorporate four layers, all with cell sizes of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ for $1.5 < |\eta| < 2.5$. Further forward ($2.5 < |\eta| < 3.2$), the cell size doubles in the η -direction. Additionally, the Forward Calorimeter (FCAL) covers the range $3.2 < |\eta| < 4.9$. It also uses LAr as active material. The first layer uses copper as passive material, while the second and third layers use tungsten instead. Single electrodes are grouped together during the readout, leading to a granularity of about $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ for the first layer and $\Delta\eta \times \Delta\phi = 0.2 \times 0.2$ for the other two layers.

Lastly, single scintillators are placed in between the transition region of the ECAL barrel and endcaps and the HCAL barrel and extended barrel. In this transition region, also called the crack, particles have to traverse a larger amount of inactive material, potentially leading to large amounts of energy not being recorded. A total of six

plastic tile scintillators along $|\eta|$ are placed in each crack to estimate this lost energy. They cover a range of $0.8 < |\eta| < 1.6$ with cell sizes of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ and $\Delta\eta \times \Delta\phi = 0.2 \times 0.1$.

3.2.3 Muon Spectrometer

Muons typically traverse the calorimeter without losing the vast majority of their energy. Therefore, an additional detector system is placed around the calorimeter system, which identifies muons and measures their momentum: the Muon Spectrometer (MS) [61]. As can be seen in Figure 3.10, it consists of a total of four subsystems: two for precision tracking and two designated for triggering purposes. A total of three toroidal magnets are incorporated into the MS. They provide a magnetic field to measure the momentum of the muons in the MS.

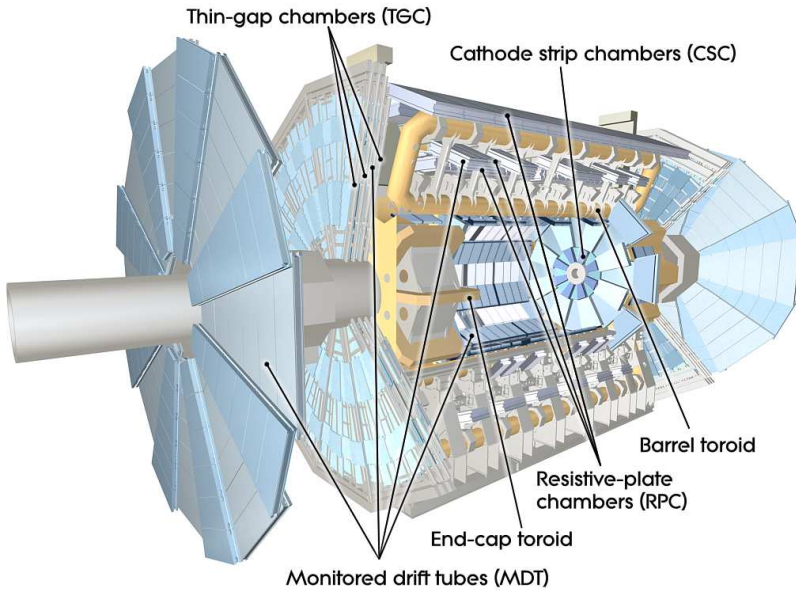


Figure 3.10: Overview of the ATLAS MS. Taken from Ref. [46].

Precision muon tracking chambers

Most of the high-precision tracking of muons is performed by the Monitored Drift Tubes (MDTs). They are installed in both the barrel and the endcaps and cover a range of $|\eta| < 2.7$. Similar to the TRT, the MDTs are gaseous detectors built of drift tubes. The gas mixture used for the MDTs is 93% Ar and 7% CO₂. Each chamber consists of multiple layers of drift tubes. Additionally, the Cathode Strip Chambers (CSCs) are placed in the endcaps covering $2.0 < |\eta| < 2.7$. They are multi-wire chambers filled with a gas mixture of 80% Ar and 20% CO₂. The MDTs reach a resolution of 35 μm in the z -direction, while the CSCs have a resolution of 40 μm in the R -direction. As the readout of these systems is too slow for triggering purposes, they are complemented by additional systems with a faster readout dedicated to triggering.

Muon trigger chambers

The Resistive Plate Chambers (RPCs) and the Thin Gap Chambers (TGCs) are used for triggering on muons. The RPCs cover the barrel region ($|\eta| < 1.05$) consisting of a total of three layers. They are made of parallel plates placed on a potential difference. The space between the plates is filled with a gas mixture of mostly $\text{C}_2\text{H}_2\text{F}_4$. The TGCs are located in the endcaps ($1.05 < |\eta| < 2.4$). They are built of multi-wire chambers filled with 55% CO_2 and 45% n-pentane. While being faster to read out than the chambers designed for precision tracking, the RPCs and the TGCs have worse spatial resolution than the precision chambers. The RPCs have a resolution of 10 mm in the z -direction, whereas the TGCs have a resolution between 2 mm and 6 mm in the R -direction.

3.2.4 Magnet system

Two separate subsystems create magnetic fields in the ATLAS detector. Through the Lorentz force, charged particles are bent in these magnetic fields, allowing a measurement of the charge and momentum of charged particles based on their trajectory.

A solenoid magnet [62] is placed around the ID. It is a superconducting magnet generating a homogeneous field of 2 T. The other magnetic system is the toroidal magnet system [63, 64]. Embedded in the MS, it consists of one barrel and two endcap toroids, each with eight coils. The provided field is inhomogeneous, with a magnetic field strength ranging from 0.15 T to 2.5 T in the barrel and from 0.2 T to 3.5 T in each of the endcaps. As the toroid magnets are embedded in the MS, they are visible in Figure 3.10.

3.2.5 Trigger and Data Acquisition

Collisions happen every 25 ns inside the ATLAS detector. This corresponds to a rate of 40 MHz which is too high for all events to be recorded. Additionally, most of these are elastic collisions. Studying these is not the primary goal of the ATLAS experiment. The Trigger and Data Acquisition (TDAQ) system is used to reduce this rate down to manageable levels while also recording all signals left by particles in the detector. The trigger flow used in Run 2 can be seen in Figure 3.11. It consists of two trigger systems: the hardware-based Level-1 (L1) trigger and the software-based High Level Trigger (HLT) [65].

The L1 trigger only processes information from the muon-trigger system (L1Muon) and the calorimeter (L1Calo). It defines regions of interest in η and ϕ corresponding to candidates for muons, electrons, photons, and more generally, large transverse momenta. The Central Trigger Processor makes the decision of whether an event passes the L1 trigger. The decision can be based on event-level quantities such as the total energy, multiplicities of certain objects, or topological requirements such as invariant masses or angular distances between objects. The L1Topo trigger is responsible for the topological require-

ments. If an event is accepted by the L1 trigger, it is passed on to the HLT. The rate is reduced to approximately 100 kHz after the L1 trigger.

The HLT performs some basic reconstruction of physics objects in the regions of interest defined by the L1 trigger. Information of all detector components is used for this reconstruction. All events that pass an HLT chain are sent to a computing centre for a full reconstruction to be performed and the data to be stored. This reduces the rate to around 1 kHz.

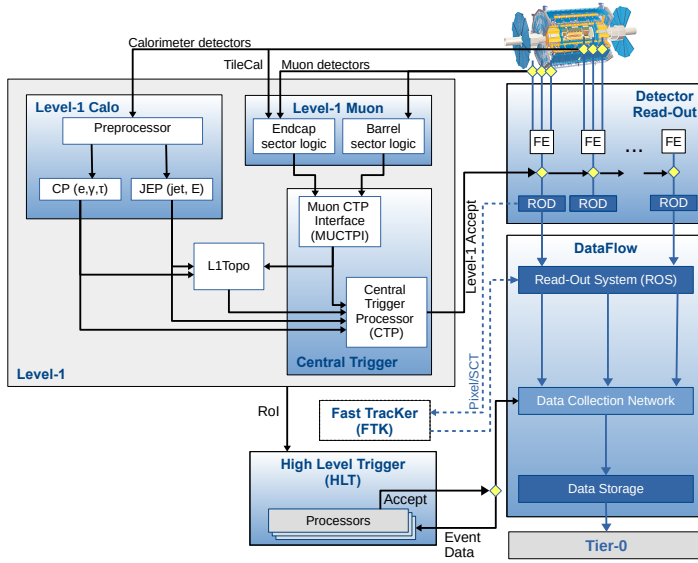


Figure 3.11: Overview of the ATLAS TDAQ system in Run 2. The FastTracker was not used during Run 2 due to it being commissioned. Taken from Ref. [65].

4 Object reconstruction

After having described how signals left by particles are detected in the ATLAS detector, the particles need to be reconstructed from the single hits in the detector. This chapter will first describe what signature the common final state particles leave in the detector before describing how “base” objects are reconstructed, which are used to reconstruct physics objects. More details will be given for electrons as they are further studied in Part II.

Contents

4.1	Interaction of particles with the detector	27
4.2	Base objects	28
4.2.1	Tracks	28
4.2.2	Vertices	29
4.2.3	Topo-clusters	29
4.3	Physics objects	30
4.3.1	Electrons	30
4.3.2	Muons	33
4.3.3	Jets	35
4.3.4	Missing transverse momentum	37
4.3.5	Overlap removal	38
4.4	Trigger	39

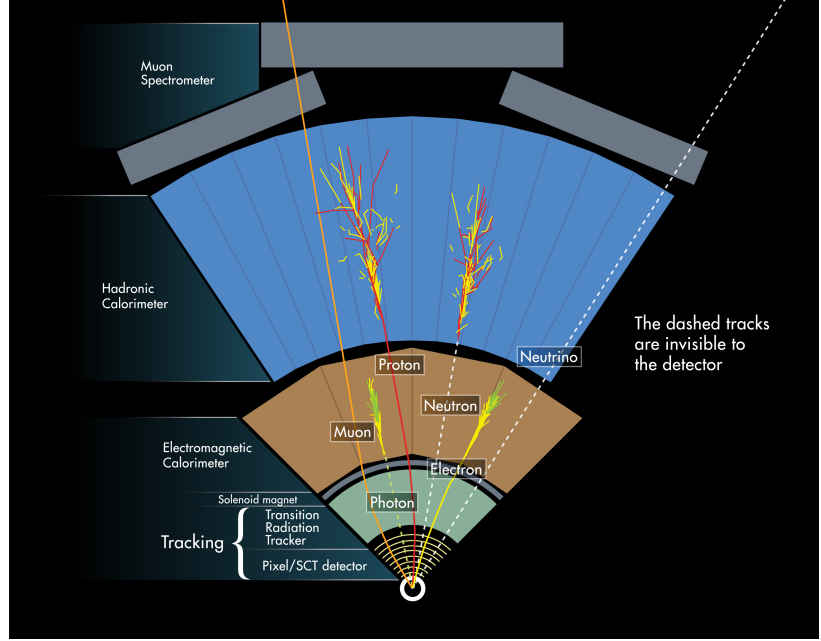
4.1 Interaction of particles with the detector

Although a multitude of different particles can be created in the collision of protons inside the ATLAS detector, most of them decay almost immediately into lighter particles. These may still have a very short lifetime and decay further until eventually stable particles are created. The particles that have a long enough lifetime to interact with the detector are: muons, electrons, photons, pions, kaons, protons, and neutrons². The first three are elementary particles, while the latter four consist of quarks. Due to the nature of colour confinement, these will produce collimated sprays of colourless particles.

Figure 4.1 shows a sketch of the interaction of common final state particles with the ATLAS detector. Charged particles - muons, electrons, and protons - leave a track in the ID, while neutral particles are invisible in the ID. Electrons and photons deposit their energy in the ECAL, protons and neutrons in the HCAL, while muons and neutrinos will traverse the detector without depositing energy. The muon, however, is detectable in the ID and the MS due to its charge, while the neutrino escapes undetected. It should be noted that this is an idealised sketch. In reality, particles can behave slightly differently; for instance, protons may start their shower earlier and deposit some amount of energy in the ECAL.

² Neutrinos also have a long enough lifetime to interact with the detector, but due to their weak interaction with matter, they cannot be detected by the ATLAS detector.

Figure 4.1: Idealized sketch of the interaction of different particles with the ATLAS detector. Taken from Ref. [66].



4.2 Base objects

As a first step, “base objects” need to be reconstructed. They combine several hits or signals in one detector system into one object, for instance, tracks in the ID. These “base objects” are then used to reconstruct physics objects that correspond to actual particles.

4.2.1 Tracks

TRACKS IN THE INNER DETECTOR A detailed description of the tracking in the ID during Run 2 can be found in Ref. [67].

Tracks are reconstructed over the whole coverage of the ID up to $|\eta| = 2.5$. In a first step, single signals in a sensor in the Pixel and SCT detector are grouped into clusters. These clusters are used to create three-dimensional space points. From sets of three space points, track seeds are created. These are filtered based on quality criteria to increase the purity of good-quality tracks. Additional space points are included iteratively by using a combinatorial Kalman filter [68] to build track candidates. These track candidates can have overlapping space points. Therefore, a track score is assigned to each track, which is used by an ambiguity solver to increase the fraction of correct tracks. The ambiguity solver also applies some quality criteria to the track candidates. They are listed in Table 4.1.

The tracks are extended into the TRT by extrapolating them into the TRT and adding additional hits [69]. Finally, a fit to the track candidates is performed to extract their parameters. This approach is called the *inside-out* approach. A secondary approach called *outside-in* is run afterwards. It starts by reconstructing tracks in the TRT in regions seeded by the ECAL. These TRT tracks can be extended back

$p_T > 400 \text{ MeV}$
$ \eta < 2.5$
Seven hits in the silicon detectors
Maximum of one shared Pixel cluster or two shared SCT clusters
Maximum of two holes in the silicon detectors
$ d_0 < 2.0 \text{ mm}$
$z_0 \sin(\theta) < 3.0 \text{ mm}$

Table 4.1: Quality criteria placed on track candidates by the ambiguity solver. A hole is a sensor that does not have a signal registered, but the track should have passed through the sensor and the neighbouring layers have registered a signal. The impact parameters d_0 and z_0 are measured with respect to the beam line.

into the silicon detectors, including hits not already used in the *inside-out* approach. Tracks with only TRT hits are also kept, as they can benefit photon conversion reconstruction.

TRACKS IN THE MUON SPECTROMETER Ref. [70] details the reconstruction of tracks in the MS. First, short track segments are formed from hits in individual MS stations. From these segments, track candidates are formed by combining different stations based on a parabolic trajectory. Here, information from the precision chambers and the trigger chambers is combined. A χ^2 fit is performed, including the effects of the magnetic field and potential interactions with the material. Based on the result, hits can be removed or added to the track. The removal or addition of hits requires a refitting of the track candidate. Similar to the ID tracks, an ambiguity solver is used to reduce the number of tracks that share hits among each other. Finally, the tracks are refitted and extrapolated back to the beam line.

4.2.2 Vertices

The origin of multiple tracks is a vertex. A Primary Vertex (PV) is defined as a vertex of a hard interaction of the collision, whereas secondary vertices are vertices from the decay of particles produced in the hard interaction. For the reconstruction of PVs [71], a set of tracks fulfilling certain quality criteria is created. A seed for a vertex is defined based on the parameters of the beams. The tracks and the vertex seed are used together in a fit to estimate the best vertex position. This fit is repeated multiple times, assigning weights to tracks after each fit based on how well they match the vertex seed. Tracks that are not compatible with the vertex after the final determination of its position are removed from it. Afterwards, this approach is repeated with the remaining tracks to iteratively reconstruct all PVs in the event.

Typically, at least one PV with at least two tracks with transverse momenta larger than 500 MeV is required. If there are multiple PVs, the one with the largest quadratic sum of the transverse momenta of the tracks is used as the first PV.

4.2.3 Topo-clusters

Three-dimensional clusters, so-called topo-clusters, are formed by combining single cells of the calorimeter system. A detailed descrip-

tion of the clustering algorithm can be found in Ref. [72].

The cell clustering is based on the cell's energy significance

$$\epsilon_{\text{cell}}^{\text{EM}} = \frac{E_{\text{cell}}^{\text{EM}}}{\sigma_{\text{cell}}^{\text{EM}}} \quad (4.1)$$

with $E_{\text{cell}}^{\text{EM}}$ the reconstructed energy in the cell and $\sigma_{\text{cell}}^{\text{EM}}$ the average noise in the cell. Both quantities are measured at the electromagnetic scale, so no corrections for the non-compensating nature of the calorimeter are taken into account. The clustering starts by finding a seed cell by requiring $|\epsilon_{\text{cell}}^{\text{EM}}| > 4$. Around each seed cell, all neighbouring cells with an absolute cell energy significance of at least two are added to the cluster. Lastly, all neighbouring cells with $|\epsilon_{\text{cell}}^{\text{EM}}| > 0$ are added.

To avoid combining signals from different particles decaying close to each other, an additional step called *cluster splitting* is performed. If a cluster has two or more local maxima, it is split into smaller clusters. A local maxima is defined as a cell with $E_{\text{cell}}^{\text{EM}} > 500$ MeV with at least four neighbours, all with a lower energy. If a cell is adjacent to two or more local maxima, its energy is divided between the two highest-energy clusters. The division of the energy is based on the distance of the cell to the centre of gravity of the split clusters. The final split clusters are called topo-clusters and are ready for use in the reconstruction of physics objects.

4.3 Physics objects

4.3.1 Electrons

Electrons leave a track in the ID and deposit most of their energy in the ECAL. Dedicated reconstruction techniques described in the following are used to account for their special properties, like their light mass compared to pions. A detailed description of the reconstruction and performance of electrons in Run 2 can be found in Ref. [73]

RECONSTRUCTION Topo-clusters are the basis for the reconstruction of electrons. To reduce the amount of electrons reconstructed from hadronic activity, only topo-clusters with at least half of their energy deposited in the ECAL are considered. In the following, they are referred to as EM topo-clusters.

The standard tracking algorithm is extended to account for electrons losing energy due to bremsstrahlung. If the pattern recognition fails for a silicon track seed in a region of interest defined by the EM topo-clusters, a modified version of the pattern recognition algorithm is used. It is based on a Kalman filter formalism [68] and allows for energy loss at each material intersection. The resulting track candidates are fit with the ATLAS global χ^2 fit. If this fit fails and the track can be loosely matched to an EM topo-cluster, the fit is retried, replacing the pion hypothesis with an electron hypothesis. Tracks with silicon hits matching EM topo-clusters obtained from these fits are once again

refitted using a Gaussian Sum Filter (GSF) [74]. The GSF better takes into account energy losses due to material interaction. These tracks are then matched to EM topo-clusters by extrapolating them into the ECAL and requiring proximity in η and ϕ . The ϕ matching criterion is chosen to be asymmetric depending on the charge to account for the potential energy loss due to bremsstrahlung. If multiple tracks can be matched to one EM topo-cluster, all are considered track candidates for the electron. They are sorted after track quality and afterwards after their proximity to the cluster centre in layer two of the ECAL. The first track is used to define the properties of the electron.

To potentially capture the energy deposited in the ECAL by photons radiated by the electron via bremsstrahlung, so-called *superclusters* are formed out of the EM topo-clusters. Starting with all EM topo-clusters sorted after their transverse energy, EM topo-clusters can become a supercluster seed if they have sufficient energy and a matched track with at least four silicon detector hits. Additional topo-clusters can be added to the seed if they are within a 3×5 window defined by the cell size of layer 2 of the ECAL or within a 5×12 window and the primary track candidate is the same as the one of the seed cluster. A schematic overview of this addition can be seen in Figure 4.2.

If topo-clusters are added to a seed cluster, they cannot be reused as a seed cluster. Finally, the seed clusters, together with all their assigned additional clusters, are referred to as superclusters. A first calibration and position correction is applied before tracks are matched for the last time to the superclusters. The matching procedure is the same as before for topo-clusters.

IDENTIFICATION A likelihood algorithm is applied to all electron candidates to reject objects reconstructed as electrons that are not defined as signal electrons. This background consists primarily of hadrons. The *LHTight* Working Point (WP) and the *LHMedium* WP are utilised in Part III. The *LHTight* WP has the highest background rejection of the available WP and therefore also the lowest signal efficiency. The *LHMedium* WP has a lower background rejection than the *LHTight* WP and is a superset of it. A more detailed description of the likelihood algorithm and its WPs will be given in Part II.

ISOLATION On top of the identification algorithm, an additional requirement is placed on electrons to further suppress objects other than signal electrons. Signal electrons coming from decays such as $Z \rightarrow ee$ are isolated, meaning there is no other activity around the electron. Other electrons, such as those originating from the decay of a b -hadron, can have additional activity around them. Therefore, isolation WPs are applied. Certain thresholds need to be surpassed in a calorimeter and a track isolation variable for the WPs to be passed.

For the track isolation, two types of variables are defined: p_T^{coneXX} and $p_T^{\text{varconeXX}}$, where XX can take values of 20, 30, and 40. Both sum the p_T of all tracks surrounding the electron that are not associated

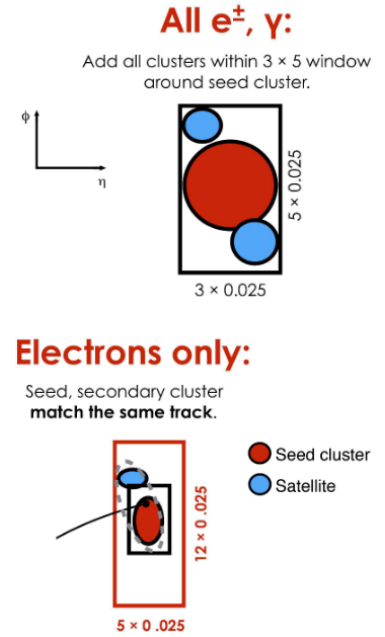


Figure 4.2: Schematic overview of the formation of superclusters during electron reconstruction. Adapted from Ref. [73].

with the electron. The tracks have to fulfil a vertex association requirement and have a p_T larger than 1 GeV or 0.5 GeV. For all electron isolation WPs used in this work, the minimum p_T is set to 1 GeV. For the p_T^{coneXX} variables, all tracks in a cone of radius $\Delta R = 0.XX$ are taken into consideration, while for the $p_T^{\text{varconeXX}}$ variables, a variable cone size of $\Delta R = \min(10/p_T, 0.XX)$ is used. So, p_T^{cone20} will sum the p_T of all tracks in a cone of radius $\Delta R = 0.2$.

A similar type of variable is defined for the calorimeter isolation: $E_T^{\text{topoconeXX}}$. For this variable, the transverse energies of additional topo-clusters within a radius of $\Delta R = 0.XX$ are summed. As for the track isolation, XX can take values of 20, 30, and 40. Corrections for leakage and pileup are applied to the calorimeter isolation variable.

All WPs used in Part III and Appendix C are listed in Table 4.2 together with their requirements. The *Tight* and *FCTight* WP have a fixed selection on track and calorimeter variables. The *Gradient* WP instead has a target efficiency that depends on p_T^{el} . Cuts are applied to match these efficiencies in both track and calorimeter isolation variables. In Part II, no WP are used but instead custom requirements are applied.

Table 4.2: Electron isolation WPs used throughout this work.

	Track isolation	Calorimeter isolation
<i>FCTight</i>	$p_T^{\text{varcone20}}/p_T^{el} < 0.06$	$E_T^{\text{topocone20}}/p_T^{el} < 0.06$
<i>Tight</i>	$p_T^{\text{varcone30}}/p_T^{el} < 0.06$	$E_T^{\text{topocone20}}/p_T^{el} < 0.06$
<i>Gradient</i>	$\epsilon = (0.1143 \times p_T^{el} + 92.14)\%$ using $p_T^{\text{varcone20}}$	$\epsilon = (0.1143 \times p_T^{el} + 92.14)\%$ using $E_T^{\text{topocone20}}$

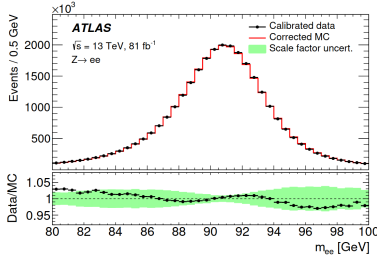


Figure 4.3: Invariant dielectron mass for $Z \rightarrow ee$ candidates for data and MC after the calibration is applied. The number of events in simulation is scaled to match the number of events in data. The band in the bottom panel indicates the impact of the uncertainties of the scale and resolution corrections. Taken from Ref. [73].

MOMENTUM CALIBRATION The calibration of the electron momentum consists of multiple steps. First, a multivariate algorithm is used to correct the detected energy to the truth energy. This algorithm is trained in Monte Carlo (MC) simulation using shower properties as inputs. Secondly, scale corrections are applied to data, while resolution corrections are applied to MC. The corrections are extracted simultaneously using $Z \rightarrow ee$ decays.

Figure 4.3 shows the invariant dielectron mass in a selection enriched in $Z \rightarrow ee$ decays for data and MC after the calibration.

In the $t\bar{t}$ + charm analysis, two types of electrons are used: a loose and a tight electron, with every tight electron also being a loose electron. The final selection is performed on tight electrons, while additional loose electrons are vetoed. Certain requirements are the same for both types. They have to be in the coverage of the ID with $|\eta| < 2.47$ and the crack region ($1.37 < |\eta| < 1.52$) is vetoed. Additionally, cuts on the impact parameters (IPs) are used to ensure that the electron is coming from the primary vertex. The significance of the transverse IP $|d_0|/\sigma(d_0)$ has to be below a value of 5, while the longitudinal requirement is $z_0 \sin(\theta) < 0.5$ mm. Table 4.3 lists the remaining requirements on the minimum transverse momentum, the

identification WP, and the isolation WP that differ between the two types of electrons.

	Min p_T^{el} [GeV]	Identification	Isolation
loose	10	<i>LHMedium</i>	None
tight	27	<i>LHTight</i>	<i>Tight</i>

Table 4.3: Requirements placed on the minimum transverse momentum, the identification WP, and the isolation WP of electrons for the $t\bar{t}$ + charm analysis. Work from Part II has specific requirements and is therefore not listed here.

4.3.2 Muons

As muons do not deposit a significant amount of their energy in the calorimeter, the reconstruction relies primarily on tracks in either the ID or the MS. A detailed description of the reconstruction and performance of muons in Run 2 can be found in Ref. [70].

RECONSTRUCTION Depending on the reconstruction, muons are grouped into five different types: combined (CB), inside-out combined (IO), muon-spectrometer extrapolated (MuE), segment-tagged (ST), and calorimeter-tagged (CT). For the reconstruction of CB muons, tracks in the ID and MS are matched to each other by extrapolating MS tracks into the ID. These tracks are combined and refitted with hits in both the ID and the MS. IO muons follow a similar approach; however, the ID tracks are extrapolated into the MS. No track has to be reconstructed in the MS for IO muons, but at least three loosely-aligned MS hits need to be present. MuE muons consist only of a track in the MS that is extrapolated to the beampipe. As no ID track is required, the range of these muons can be extended up to $|\eta| < 2.7$ compared to the coverage of the ID of $|\eta| < 2.5$. For ST muons, an ID track needs to be matched to at least one reconstructed MS segment. Lastly, CT muons are formed from ID tracks that can be extrapolated to energy deposits in the calorimeter that are consistent with those of a minimum-ionising particle.

IDENTIFICATION Additional identification criteria are applied to the reconstructed muons at the analysis level in order to prevent the use of particles that were incorrectly reconstructed as muons. In total, five WPs are provided: *Loose*, *Medium*, *Tight*, *High p_T* , and *Low p_T* . They all provide different efficiencies to cover different needs in physics analyses. In this thesis, *Loose*, *Medium* and *Tight* muons will be used.

For $|\eta| < 2.5$, *Medium* muons have to be either CB or IO muons. All tracks in the ID need at least one hit in the Pixel detector and at least five hits in the SCT. At most two holes are allowed in the Pixel detector and the SCT. The track in the MS needs to contain at least three hits in two stations of the MDT or the CSC, unless it is very central ($|\eta| < 0.1$) where one MDT or CSC station with three hits is sufficient if there is at most one MDT or CSC station with less than three hits and at least three more expected hits. Additionally, the tracks in the ID and the MS have to be compatible with each other. This is ensured by requiring a parameter called q/p -compatibility to be below a certain threshold.

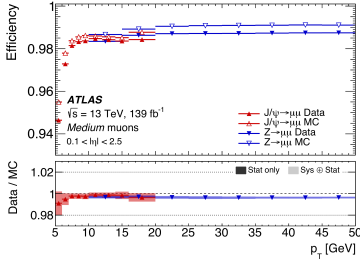


Figure 4.4: Reconstruction and identification efficiency as a function of the muon p_T for *Medium* muons with $0.1 < |\eta| < 2.5$. The values are measured in $Z \rightarrow \mu\mu$ and $J/\psi \rightarrow \mu^+\mu^-$ decays. Non-negligible statistical uncertainties are indicated by error-bars. The bottom panel shows the Data/MC ratio with statistical and systematic uncertainties included. Taken from Ref. [70].

Loose muons include all *Medium* muons. Additionally, CT and ST muons are included if they are in the range $|\eta| < 0.1$. In this region, a gap in the MS results in a loss of efficiency for CB muons. The inclusion of CT and ST muons can recover parts of this loss. To increase the efficiency at low transverse momenta, IO muons with $p_T < 7$ GeV and $|\eta| < 1.3$ are accepted if they can also be reconstructed as ST muons.

Tight muons are a subset of *Medium* muons and therefore need to pass the same requirements. On top of that, at least two stations of the MDT or the CSC with three hits are required even for $|\eta| < 0.1$. Harsher requirements are placed on the q/p -compatibility and an additional variable describing the track compatibility ρ' is used. It is defined as the absolute difference between the transverse momentum of the tracks in the ID and the MS normalised by the transverse momentum of the combined fit. Lastly, a requirement is placed on the normalised χ^2 of the combined track fit.

Figure 4.4 shows the reconstruction and identification efficiencies of *Medium* muons as measured in $Z \rightarrow \mu\mu$ and $J/\psi \rightarrow \mu^+\mu^-$ decays.

ISOLATION Similarly to electrons, muons also have to pass an isolation criterion to be considered in most analyses. Again, different WPs exist to match the needs of different analyses. All used WPs in this work are listed in Table 4.4 together with their requirements. The *TightTrackOnly* WP only applies a selection on the track isolation, while the *Tight* WP applies a selection on the calorimeter isolation as well. The last WP, *PFlowLoose*, uses an additional variable in E_{neflow}^{20} . It sums the transverse energy of neutral particles in a cone of $\Delta R = 0.2$. The decision about whether a deposit is from a charged or neutral particle is taken by a particle-flow algorithm. As only the energy of neutral particles is summed in this way, it is added to the regular track isolation variables with a weight of 0.4. Additionally, the $p_T^{\text{varcone30}}$ and p_T^{cone20} variables used for the *PFlowLoose* WP use tracks with a minimum of 0.5 GeV instead of 1 GeV.

	Track isolation	Calorimeter isolation
<i>Tight</i>	$p_T^{\text{varcone30}}/p_T^\mu < 0.04$	$E_T^{\text{topocone20}}/p_T^\mu < 0.15$
<i>TightTrackOnly</i>	$p_T^{\text{varcone30}}/p_T^\mu < 0.06$	-
<i>PFlowLoose</i>	$(p_T^{\text{varcone30}} + 0.4 \cdot E_{\text{neflow}}^{20})/p_T^\mu < 0.16$ for $p_T^\mu < 50$ GeV $(p_T^{\text{cone20}} + 0.4 \cdot E_{\text{neflow}}^{20})/p_T^\mu < 0.16$ for $p_T^\mu \geq 50$ GeV	

Table 4.4: Muon isolation WPs used throughout this thesis.

MOMENTUM CALIBRATION The muon momentum in simulation is corrected to match the measurements in data [75]. The correction combines the measured transverse momentum of the tracks in the ID and the MS, with the combined track. The correction factors are obtained using $Z \rightarrow \mu\mu$ and $J/\psi \rightarrow \mu^+\mu^-$ decays.

Figure 4.5 shows the invariant dimuon mass in a selection enriched in $Z \rightarrow \mu\mu$ decays for data and MC after the calibration.

As for electrons, a loose muon and a tight muon definition are used in the $t\bar{t}$ + charm analysis. Different requirements are placed on the minimum transverse momentum, the identification WP, and the isolation WP. An overview of the different selections can be seen in Table 4.5. On top of these, similar requirements as for electrons are placed on the IPs for all muons. A harsher cut is placed on the significance of the transverse IP $|d_0|/\sigma(d_0)$ with a value of 3. The longitudinal requirement is the same with $z_0 \sin(\theta) < 0.5$ mm.

	Min p_T^μ [GeV]	Identification	Isolation
loose	10	<i>Loose</i>	None
tight	27	<i>Medium</i>	<i>Tight</i>

4.3.3 Jets

As introduced earlier, quarks and gluons will create sprays of particles. To describe the original quark or gluon, all these particles need to be grouped together to be reconstructed as a *jet*, a cone-like structure in the detector. The definition of a jet depends on the clustering algorithm used.

JET CLUSTERING The clustering of jets in the ATLAS experiment is performed by the anti- k_t algorithm [76] implemented in the FASTJET package [77]. The anti- k_t algorithm is collinear and infrared safe, two important properties of jet clustering algorithms. Collinear safety ensures no change in jet configuration when replacing a single particle with two collinear particles. Infrared safety ensures stability against soft emissions. As inputs to the algorithm, four-vector objects are used. These are then recombined sequentially based on two distance measures

$$d_{i,j} = \min \left(p_{T,i}^{2n}, p_{T,j}^{2n} \right) \frac{\Delta R_{ij}^2}{R^2} \quad (4.2)$$

$$d_{i,B} = p_{T,i}^{2n} R^2, \quad (4.3)$$

where $d_{i,j}$ is the distance between two objects, $d_{i,B}$ is the distance of object i to the beam, $p_{T,i/j}$ the transverse momentum of the object i/j , and ΔR_{ij} the ΔR distance between objects i and j . The values for n and R can be set freely and will change the behaviour of the clustering algorithm. These distances are calculated for all objects. Finding the smallest of these, the two entities i and j are merged into one if $d_{i,j}$

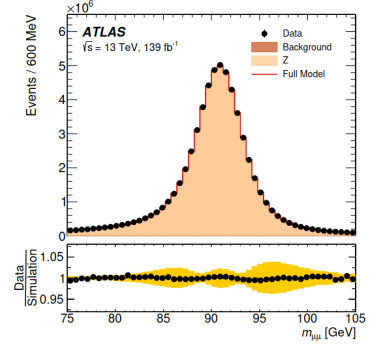


Figure 4.5: Invariant dimuon mass for $Z \rightarrow \mu\mu$ candidates for data and MC after the calibration is applied. The band in the bottom panel indicates the impact of the uncertainties of the momentum corrections. Taken from Ref. [75].

Table 4.5: Requirements placed on the minimum transverse momentum, the identification WP, and the isolation WP of muons for the $t\bar{t}$ + charm analysis.

¹ The k_t algorithm [78] as an alternative clustering algorithm uses $n = 1$

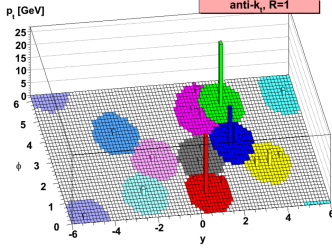


Figure 4.6: Jet clustering using the anti- k_t algorithm with a radius parameter of $R = 1$. Taken from Ref. [76].

² Jets clustered with a radius parameter of $R = 1.0$ also exist in ATLAS but are not used in this work.

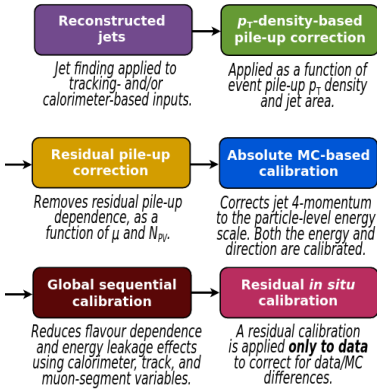


Figure 4.7: Single steps in the jet energy scale calibration. Taken from Ref. [80]

has a lower value than $d_{i,B}$. If $d_{i,B}$ has a lower value, the object i is classified as a jet and removed from further clustering. Afterwards, using the new set of objects, the distances are recalculated, and the clustering is repeated until no objects are left.

For the anti- k_t algorithm, the exponent n is set to -1 ¹. This gives higher importance to objects with a high p_T . Therefore, the anti- k_t algorithm typically ends up as a cone around the hardest emission, as shown in Figure 4.6.

JETS IN ATLAS Jets in the ATLAS experiment combine track and calorimeter information [79]. To combine the information, the tracks in the ID are matched to topo-clusters in the calorimeter. If a match can be found, the quantities of the topo-cluster are replaced by the track properties. The remaining topo-clusters and the tracks that match the primary vertex are used as input for the anti- k_t algorithm. A radius parameter of $R = 0.4$ is used².

JET ENERGY SCALE CALIBRATION As the ATLAS HCAL is a non-compensating calorimeter, the energy of hadronic objects will be underestimated. The calibration procedure [80] tries, among other things, to correct this underestimation.

Figure 4.7 shows the chain of the full calibration procedure of the jet energy scale. In all stages, the jet's momentum is corrected. The first two steps aim to remove excess energy due to pileup. The first of these two corrections is based on jet properties, while the second utilises event observables in the mean number of interactions per crossing and the number of reconstructed vertices. In the next step, the jets are corrected such that their energy and direction agree with truth jets from simulated dijet events. This step aims to recover lost energy due to the non-compensating nature of the HCAL. The following step is also derived from simulated dijet events. It further corrects the jet energy scale using calorimeter and tracking information from the ID and the MS. All these steps are applied in the same way to data and simulated events. The last step is only applied to data and aims to correct differences between simulation and data. Similar to the electron and muon calibrations, well-measured reference objects are used. For jets, this includes photons and Z bosons.

JET VERTEX Tagger The contribution of pileup to jets is reduced in the calibration detailed above. However, pileup can also create signals that are reconstructed as jets even though no hadron of the primary collision was present. To reduce the number of these pileup jets, a multivariate algorithm, the Jet Vertex Tagger (JVT) [81], is used. A two-dimensional likelihood is formed using the relative transverse momentum of the jet originating from jets associated with the primary vertex and a derived quantity called *jet vertex tagger fraction*. Scale Factors (SFs) to correct the efficiencies measured in simulation to the ones measured in data are obtained using $Z \rightarrow \mu\mu$ events.

In the $t\bar{t}$ + charm analysis, all jets have to have $p_T > 25$ GeV and $|\eta| < 2.5$. Jets with a transverse momentum smaller than 60 GeV and an absolute pseudorapidity smaller than 2.4 are vetoed if their JVT output is smaller than 0.5.

Flavour tagging

Jets have different properties based on the flavour of the original particle. Based on these differences, the flavour of the original quark, namely whether it was a b -quark, a c -quark, or a light quark¹, can be inferred. The identification of jets that are initiated by b -quarks is commonly referred to as flavour tagging. Due to the high lifetime of b -hadrons of around 1.5 ps, b -hadrons travel a significant distance before decaying. This can be exploited as the tracks inside the jet point towards a secondary vertex displaced from the primary vertex of the hard interaction. Additionally, b -hadrons have a higher decay multiplicity than light hadrons. Hadrons containing c -quarks behave slightly differently from both b -hadrons and light hadrons. Their lifetime is shorter than that of b -hadrons but, on average, longer than that of light hadrons. Similarly, their decay multiplicity is lower than that of b -hadrons but higher than that of light hadrons. Due to the mass of the quarks, b - and c -jets are also referred to as heavy-flavour (HF) jets.

In simulated samples, the flavour of a jet is defined by angular matching of the reconstructed jet to generator-level objects. Only objects with a p_T larger than 5 GeV are considered. A jet is labelled b -jet if the jet is matched to a b -hadron within a cone of $\Delta R = 0.3$. If the jet cannot be matched to a b -hadron, the process is iteratively repeated for c -hadrons and τ -leptons, resulting in labels of c -jet and τ -jet, respectively. Lastly, if none of these objects could be matched to the jet, it is labelled a light-flavour jet.

To identify b -jets in ATLAS, a two-step approach is used [82]. Several low-level algorithms are used to extract specific information from the tracks in the jet. IP2D and IP3D [83] exploit the large lifetime of b -hadrons by utilising the IPs of the tracks. RNNIP [84] also uses the IPs, but as an input to a recurrent neural network. SV1 [85] attempts to reconstruct secondary vertices, while JetFitter [86] attempts to reconstruct the whole decay chain of the b - or c -hadron. The outputs of these low-level algorithms are then combined into one high-level algorithm, DL1r. The algorithm used is a Neural Network (NN). The outputs of the DL1r NN are then used to identify b -jets.

As the outputs of the DL1r NN are used uniquely in the $t\bar{t}$ + charm analysis, more details will be given in Section 11.5.

¹ Up, down, and strange quarks are collectively called light quarks.

4.3.4 Missing transverse momentum

Neutrinos do not interact with the detector. However, their presence can be inferred from the remaining objects in the event. As the colliding protons only carry momentum in the z -direction, the transverse momentum before the collision is zero. Therefore, the

overall transverse momentum of each collision must balance to zero. The transverse momentum needed to obtain zero is called the missing transverse momentum or missing transverse energy E_T^{miss} .

To calculate E_T^{miss} , all other objects of the primary interaction are required [87]. Reconstructed and calibrated electrons, muons, photons, jets, and hadronically decaying τ -leptons are considered. A dedicated overlap removal is performed to avoid any double counting from tracks or clusters being used for multiple physics objects. An additional *soft term* is added to the calculation to account for particles that, for instance, have too little energy to be considered one of the physics objects mentioned before. This term includes tracks that pass some identification criteria in the ID but are not associated with any of the considered physics objects. To obtain the x - and y -components of E_T^{miss} , the negative sum of the momentum of the hard and soft objects is taken

$$E_{x,(y)}^{\text{miss}} = - \sum_{i \in \{\text{hard objects}\}} p_{x,(y)} - \sum_{i \in \{\text{soft tracks}\}} p_{x,(y)} \quad (4.4)$$

which allows to calculate the overall magnitude

$$E_T^{\text{miss}} = \sqrt{(E_x^{\text{miss}})^2 + (E_y^{\text{miss}})^2}. \quad (4.5)$$

$Z \rightarrow \mu\mu$, $W \rightarrow e\nu$, and $W \rightarrow \mu\nu$ events are used to calibrate the scale and resolution of E_T^{miss} . The events with a Z boson have no expected E_T^{miss} , while the events with a W boson have an expected E_T^{miss} due to the neutrinos.

4.3.5 Overlap removal

¹ And photons and taus which are not described here as they are not used in this work

All previously described physics objects¹ are built from the same set of tracks and clusters. No information exchange between the single reconstruction algorithms is performed, so the same tracks or clusters could be used to reconstruct multiple objects. To avoid double counting of tracks or clusters, a procedure called *overlap removal* is applied. It applies the following steps sequentially: Electron candidates sharing a track with muon candidates are removed. Jet candidates closer than $\Delta R_y = 0.2$ to electron candidates are removed, while electron candidates are removed if the distance between a jet candidate and the electron candidate is between $\Delta R_y = 0.2$ and $\Delta R_y = 0.4$. A similar procedure is used for the overlap of jets and muons. Jet candidates closer than $\Delta R_y = 0.2$ to muon candidates are removed if they have only two or fewer associated tracks. Muons candidates are removed if there is a jet candidate that has at least three associated tracks closer than $\Delta R_y = 0.4$.

This procedure is applied in the $t\bar{t}$ + charm analysis in Part III. No overlap removal is performed for the electron identification studies detailed in Part II.

4.4 Trigger

Different objects can be used as requirements for the trigger. Electrons and muons provide a clean signature to trigger on. Both can be identified with a high rejection of hadronic background, and both are indicators of an “interesting”¹ physics process. To further suppress the hadronic background, similar identification and isolation requirements can be applied as in the offline reconstruction described earlier. At lower transverse momenta, the rate of hadrons being identified as leptons is higher, which is why the identification and isolation requirements are tighter at lower p_T . To maintain good efficiencies over the full p_T spectrum, typically only one of multiple triggers needs to be passed, with additional triggers having higher p_T and looser identification requirements.

The nomenclature of the trigger chains is the following:

[Trigger Level]_[object multiplicity][object type][minimum transverse momentum in GeV].

The trigger level refers to HLT or L1. If only one object is required, the object multiplicity is usually omitted. As object types, muons (mu) and electrons (e) are considered in this work. Additional identification or isolation requirements are added at the end with additional underscores. If an HLT trigger chain was seeded by an L1 trigger, it might be added at the end as well. Therefore, the trigger HLT_e26_lhtight_nod0_ivarloose is for the HLT. It requires one electron with a p_T larger than 26 GeV. Additionally, a tight likelihood identification WP and a loose isolation WP need to be passed. However, no requirement is placed on the transverse IP d_0 .

Table 4.6 shows the single-lepton trigger used throughout most of this work. As can be seen for the period from 2016 to 2018, the electron trigger with the lowest p_T has a much tighter identification and isolation requirement than that with the higher p_T threshold. Different triggers were used for the period of 2015. Due to changes in the amount of pileup after 2015, the triggers needed to be adjusted to keep the trigger rate at acceptable levels.

	2015	2016–2018
Electrons	HLT_e24_lhmedium_L1EM20VH or HLT_e60_lhmedium or HLT_e120_lhloose	HLT_e26_lhtight_nod0_ivarloose or HLT_e60_lhmedium_nod0 or HLT_e140_lhloose_nod0
Muons	HLT_mu20_iloose_L1MU15 or HLT_mu50	HLT_mu26_ivarmedium or HLT_mu50

All the triggers in Table 4.6 are so-called unprescaled triggers. This means that every event that passes the trigger requirement is saved. Other triggers can be prescaled. These are only recorded every n times an event passes the trigger requirement. This allows the trigger requirements to be much looser. For instance, the prescaled trigger HLT_e5_etcut only needs one electron with a p_T of 5 GeV. Such triggers are, for instance, interesting for background studies regarding electron identification and will be used in Part II.

¹ Interesting is obviously subjective. Here, interesting is meant with respect to the physics programme of the ATLAS experiment.

Table 4.6: String of the single-lepton high-level triggers used for most analyses in this work.

5 Simulation

To estimate the contributions of different processes, simulated samples are used. Additionally, these simulated samples can be used to obtain data with labels to train Machine Learning (ML) algorithms. This chapter will introduce all samples produced by the ATLAS simulation infrastructure [88] used throughout Parts II and III.

Contents

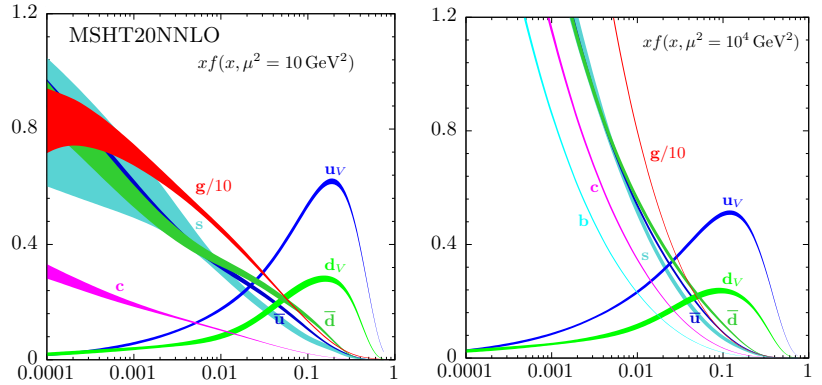
5.1	Parton distribution functions	42
5.2	Matrix element generation	42
5.3	Decays of heavy particles	43
5.4	Parton shower	43
5.5	Underlying event	44
5.6	Detector simulation	44
5.7	Simulated samples	44
5.7.1	Top-quark pairs	46
5.7.2	Top-quark pairs in association with a pair of b -quarks	47
5.7.3	Single top-quarks	48
5.7.4	Top-quark pairs in association with an electroweak boson	48
5.7.5	Top-quark pairs in association with a Higgs boson	49
5.7.6	Rare single top-quark decays	49
5.7.7	Single electroweak bosons	49
5.7.8	Two electroweak bosons	50
5.7.9	Z boson in association with a photon	50
5.7.10	J/ψ meson decaying into a pair of electrons	50
5.7.11	Two-to-two processes	50

A large part of the physics processes at the LHC are caused by the strong force and therefore need to be calculated using QCD. Matrix elements can be accurately calculated using perturbative QCD. However, the further interaction, including the hadronisation of the created partons, cannot be described by perturbative QCD. Instead, they are separated from the hard interaction following the factorisation theorem [89]. This introduces the factorisation scale μ_F which is the scale at which the two processes, matrix element generation with perturbative QCD and the further hadronisation are split. The generation of physics processes follows this structure. First, the process is simulated following the matrix element. For their calculation, the PDFs are important. The generated particles of the matrix element are handed over to a second algorithm according to the factorisation scale to generate the hadronisation and, in general, interactions within hadrons. Lastly, the interaction of the final state particles with the detector after the hadronisation is simulated.

5.1 Parton distribution functions

As protons are not point-like particles, not the protons themselves but their constituents, the partons, are colliding in the LHC. Protons consist of three valence quarks, two up- and one down-type quarks. These give the proton its electrical charge and other quantum numbers. On top of these three quarks, due to vacuum fluctuations, gluons and quark-antiquark pairs, also called sea quarks, are inside the proton. Each of these partons carries a specific fraction of the momentum of the proton, which is quantified by the Bjorken x . To simulate processes at the LHC, these values need to be known. The probability of having a parton of a specific type at a certain Bjorken x is given by the Parton Distribution Functions (PDFs). Due to the large value of α_s at low energies, they cannot be calculated from first principles. Instead, they have to be extracted from measurements. Different collaborations, like the CTEQ, MSTW, or NNPDF collaborations [37, 90, 91], provide sets of PDFs. In addition to not being calculable, the PDFs change depending on the factorisation scale μ_F . To extrapolate towards different scales, the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations [92–94] are used.

Figure 5.1: PDFs for two different choices of the factorisation scale: (left) $\mu_F^2 = 10 \text{ GeV}^2$ and (right) $\mu_F^2 = 10 \text{ TeV}^2$. The PDFs are obtained in the NNLO MHST20 analysis [95]. Taken from Ref. [12]



Two different schemes are used for the PDF sets in this work: five flavour scheme (5FS) and four flavour scheme (4FS). The number indicates the number of flavours of quarks included in the PDF of the proton, with both sets including gluons. 5FS PDFs contain all flavours except the top-quark, whereas a 4FS PDF will also not have any b -quark contribution. Hence, in 4FS samples, b -quarks need to be produced, for instance, by a gluon splitting into a $b\bar{b}$ pair and cannot be taken from the proton.

5.2 Matrix element generation

The Matrix Element (ME) generation simulates the interaction of two partons of the colliding protons, producing elementary particles. This part of the generation process can be represented by Feynman diagrams. For the calculation, perturbation theory is used. Including

higher orders in the generation generally leads to better results, but the computational cost also increases drastically. Most processes are generated at NLO. The cross-section, however, can often be calculated, including higher orders. Therefore, the generated events are usually rescaled to match the calculated cross-section.

Several algorithms are used for the generation of the ME including POWHEG BOX [96–100], PYTHIA 8 [101, 102], SHERPA [103], and MADGRAPH5_AMC@NLO [104].

5.3 Decays of heavy particles

In the simulation, W bosons, by default, cannot decay into any quark combination involving b -quarks. Due to the components of the CKM matrix, these decays are heavily suppressed. Dedicated samples are required if b -quarks from a W boson decay are needed.

For most samples, the decays of b - and c -hadrons were handled by the EVTGEN 1.6.0 program [105]. All samples generated with SHERPA were using the default SHERPA configuration instead. Other differences will be explicitly mentioned.

In all samples using MADGRAPH5_AMC@NLO for the ME generation as well as the t -channel single-top sample using POWHEG, MADSPIN [106, 107] was used to decay top-quarks, Z bosons, and W bosons to preserve spin correlations. The mass of the top-quark was always set to $m_{\text{top}} = 172.5$ GeV. The mass of b -quarks from a top-quark decay was set to $m_b = 4.95$ GeV for POWHEG BOX and MADSPIN, while it was set to 4.75 GeV for SHERPA. In samples containing a Higgs boson, the Higgs boson mass was set to $m_H = 125.0$ GeV and the mass of b -quarks from a Higgs boson decay was set to $m_b = 4.80$ GeV.

In samples using the 5FS, b -quarks not from top-quark or Higgs boson decays are treated massless. For the POWHEG BOX samples generated with a 4FS PDF, the b -quark mass for b -quarks from the ME is set to $m_b = 4.95$ GeV.

5.4 Parton shower

The Parton Shower (PS) is used to further propagate the particles coming from the ME generation. As the PS simulates the evolution of partons, it is also responsible for soft emissions. These can happen for both initial state and final state partons, also known as initial-state radiation (ISR) and final-state radiation (FSR), respectively.

Afterwards, the hadronisation process of the partons is simulated. The colour-charged partons hadronise into colourless hadrons following the colour confinement of QCD. As this process cannot be calculated from first principles, different methods exist to calculate the hadronisation with the two main ones being the Lund string model [108] and the cluster model [109]. In the Lund string model, quark-antiquark pairs are connected with strings. The energy stored in these strings increases linearly with the distance. If the energy

is large enough to create a new hadron, the string is split and the new hadron is generated, creating new strings between the original quarks and the newly generated ones. In the cluster-based model, the colour pre-confinement property of QCD is exploited. Here clusters of colour singlets are identified, which subsequently decay into the final state hadrons.

PS algorithms used throughout this work include PYTHIA 8, HERWIG 7 [110–112], and SHERPA. PYTHIA uses a Lund string model, while HERWIG and SHERPA use a cluster-based model.

5.5 Underlying event

The effect of additional pp collisions, either in the same bunch crossing or a neighbouring bunch crossing, was taken into account by adding so-called minimum-bias interactions to the generated samples. These minimum-bias events were generated with PYTHIA 8.186. A specific set of tuned parameters called A3 [113] and the NNPDF2.3LO PDF set [114] were used for the generation.

¹ 2015 and 2016 are treated as one year

Each sample was generated for each year¹ to reflect the data-taking conditions and to match the observed pileup profile. Different seeds were used for the different years, allowing analyses to combine the different years into one prediction for the overall period of Run 2. By applying weights to the single events, the pileup profile of the different years in Run 2 can be matched by the simulation.

5.6 Detector simulation

After the ME generation and the PS algorithm, only stable final state particles remain. However, the interaction of these particles with the detector needs to be simulated as well. This constitutes the last step of the full simulation chain. The most accurate detector simulation is obtained by using the GEANT4 toolkit [115]. Using detailed information of the detector, it propagates single particles through the detector, simulating the interactions of the particles with the detector at every step. This detector simulation with GEANT4 is also referred to as full simulation. It has, however, a major shortcoming in its computational cost. For this reason, several samples are not processed with the full simulation but instead use a faster algorithm called ATLFAST-II (AFII) [88], also referred to as fast simulation. It replaces the most expensive parts of the simulation with a parameterised simulation, for instance, the calorimeter response. It has, however, worse accuracy and does not describe data as well as the GEANT4 simulation.

5.7 Simulated samples

In the following, the generator setup of all samples generated with the ATLAS simulation infrastructure [88], which are used in this work, is detailed. This covers all samples used in Parts II and III. Table 5.1

lists all of these samples together with their use cases. The majority of samples are used for the $t\bar{t}$ + charm analysis to estimate the contributions from signal and background processes. Several processes are available not only in one generator setup but in several setups to estimate systematic uncertainties. Almost completely separate samples are used for the electron identification studies in Part II. These samples are not used to estimate signal and background contributions of a selection, but instead are used to obtain signal and background electron candidates.

Process	ME	PS	Additional parameters	Usage
$t\bar{t}$	POWHEG	PYTHIA	-	$t\bar{t}$ + charm, electron
	POWHEG	PYTHIA	h_{damp} variation	$t\bar{t}$ + charm
	POWHEG	PYTHIA	$p_{\text{T}}^{\text{hard}}$ variation	$t\bar{t}$ + charm
	POWHEG	HERWIG7.1.3	-	$t\bar{t}$ + charm
	MADGRAPH5_AMC@NLO	HERWIG	-	$t\bar{t}$ + charm
$t\bar{t}b\bar{b}$	POWHEG	PYTHIA	-	$t\bar{t}$ + charm
	POWHEG	PYTHIA	dipole recoil	$t\bar{t}$ + charm
	POWHEG	PYTHIA	h_{bzd} variation	$t\bar{t}$ + charm
	POWHEG	PYTHIA	$p_{\text{T}}^{\text{hard}}$ variation	$t\bar{t}$ + charm
	POWHEG	HERWIG7.1.6	-	$t\bar{t}$ + charm
	SHERPA2.2.10	SHERPA2.2.10	-	$t\bar{t}$ + charm
single top: Wt , s-, t-channel	POWHEG	PYTHIA	-	$t\bar{t}$ + charm
	POWHEG	PYTHIA	only Wt , DS	$t\bar{t}$ + charm
	POWHEG	HERWIG7.0.4	-	$t\bar{t}$ + charm
	MADGRAPH5_AMC@NLO	PYTHIA	-	$t\bar{t}$ + charm
$t\bar{t}V$	MADGRAPH5_AMC@NLO	PYTHIA	-	$t\bar{t}$ + charm
	SHERPA	SHERPA	-	$t\bar{t}$ + charm
$t\bar{t}H$	POWHEG Box	PYTHIA	-	$t\bar{t}$ + charm
tZ and tWZ	MADGRAPH5_AMC@NLO	PYTHIA	-	$t\bar{t}$ + charm
Z + jets	SHERPA	SHERPA	-	$t\bar{t}$ + charm
	POWHEG Box	PYTHIA	-	electron
W + jets	SHERPA	SHERPA	-	$t\bar{t}$ + charm
VV	SHERPA	SHERPA	-	$t\bar{t}$ + charm
$Z\gamma$	SHERPA	SHERPA	-	electron
$J/\psi \rightarrow e\bar{e}$	PYTHIA	PYTHIA	-	electron
JF17	PYTHIA	PYTHIA	-	electron

Many of the samples exist with different decay channel filters applied. For instance, for the nominal top-quark pair production setup, three different samples exist; one only containing events where both W bosons decay leptonically, one where at least one W boson decays leptonically, and one where both W bosons decay hadronically. These filters are used to increase the available statistics of the respective decay channel. Potential overlap between these samples is removed during the analysis.

For most samples, if PYTHIA was used for the PS and the hadronisation, the A14 set of tuned parameters [116] was used along with the NNPDF2.3Lo PDF set.

Table 5.1: List of all samples generated in the ATLAS framework used in this work. The "Usage" column describes in which parts of this work the sample is used. $t\bar{t}$ + charm refers to the analysis carried out in Part III and electron to the performance studies in Part II.

5.7.1 Top-quark pairs

The production of top-quark pairs was generated by a variety of different generator setups. All the following samples are generated in the 5FS setup. Figure 2.8 shows the tree-level diagrams for $t\bar{t}$ production at the LHC.

POWHEG BOX + PYTHIA Events with a top-quark pair were modelled using the POWHEG BOX v2 generator with the NNPDF3.0_{NLO} PDF set [91]. MEs were calculated at NLO in α_S . The factorisation and renormalisation scales were set to the default functional form of $\sqrt{m_{\text{top}}^2 + p_T^2}$. The h_{damp} parameter was set to $1.5 m_{\text{top}}$ [117]. This parameter controls the matching between the ME and PS generators. It effectively regulates high- p_T radiation. PYTHIA 8.230 was used for the PS and the hadronisation. This nominal sample, as well as all other alternative $t\bar{t}$ production samples, was normalised to the calculated cross-section of $\sigma(t\bar{t}) = 832 \pm 51$ pb.

To increase MC statistics for $t\bar{t}$ production with additional HF jets, dedicated filtered samples were generated. The filters require additional b - or c -hadrons apart from the ones created by the top-quark decays. They can be seen in Table 5.2. These filtered samples are combined with the inclusive sample, removing all overlap between the samples if additional statistics of top-quark pair decays with additional c - or b -hadrons is required. They are only used in the $t\bar{t}$ + charm analysis described in Part III. The electron identification studies in Part II only use the inclusive sample as additional statistics is not needed.

Table 5.2: Filters used for additional $t\bar{t}$ production samples.

Name	Hadron Requirement	Veto
bb -filter	≥ 2 b -hadrons with $p_T > 15$ GeV	
b -filter	≥ 1 b -hadron with $p_T > 5$ GeV	bb -filter passed
c -filter	≥ 1 c -hadron with $p_T > 15$ GeV	bb - or b -filter passed

The sample has several internal weight variations. By applying these weights, different effects, such as a doubling of the factorisation scale, can be emulated. These weight variations are used to estimate systematic uncertainties.

Two additional samples were generated using POWHEG BOX + PYTHIA. The first sample was generated while multiplying the h_{damp} parameter by a factor of two, setting it to $3.0 \cdot m_{\text{top}}$ with otherwise unchanged parameters. The second sample was simulated while changing the parameter p_T^{hard} which controls the “hardness” of allowed emissions to a value of one. This sample also uses a newer version of PYTHIA with PYTHIA8.306. No filtered samples are available for these two additional samples.

POWHEG BOX + HERWIG A sample using the same setup as for the POWHEG BOX + PYTHIA sample for the ME generation but HERWIG 7.1.3 for the PS was generated. The HERWIG 7.1 default set of

tuned parameters [111, 112] was used with the MMHT2014_{LO} PDF set [118].

MADGRAPH + HERWIG An additional set of events was modelled with the MADGRAPH5_AMC@NLO 2.6.0 generator with the NNPDF3.0_{NLO} PDF set for the ME calculation interfaced with HERWIG 7.1.3. The same settings for HERWIG as for the POWHEG BOX + HERWIG 7.1.3 sample were used with the HERWIG 7.1 default tune and the MMHT2014_{LO} PDF set. This sample is only used for a comparison of its predicted cross-section to the measured cross-section in the $t\bar{t}$ + charm analysis.

5.7.2 Top-quark pairs in association with a pair of b -quarks

Dedicated samples for modelling $t\bar{t}b\bar{b}$ were generated [119]. All of these samples use 4FS PDFs. An example tree-level Feynman diagram can be seen in Figure 5.2. Here, the $b\bar{b}$ pair is produced by an ISR gluon.

POWHEG BOX RES + PYTHIA Events of $t\bar{t}b\bar{b}$ production were simulated with the POWHEG BOX RES [120] generator and OPENLOOPS. A pre-release of this implementation provided by the authors of POWHEG BOX RES [121] was used. The NNPDF3.0_{NLO} PDF set was used. The generator is interfaced with PYTHIA 8.244. As the 4FS is used, the b -quark mass is set to 4.95 GeV. The POWHEG-internal parameter h_{bzd} is set to five. It controls the splitting of the real emission into the finite and singular parts in the NLO calculation. The decay of b - and c -hadrons is handled by the EVTGEN1.7.0 program.

Three additional samples were simulated while changing generation parameters. Two were generated with $h_{\text{bzd}} = 2$ instead of $h_{\text{bzd}} = 5$. All other parameters were left unchanged for the first sample. The second sample was simulated using the dipole recoil scheme instead of the global one in addition to the change in h_{bzd} . The last sample used $h_{\text{bzd}} = 5$ together with a changed $p_{\text{T}}^{\text{hard}}$ parameter.

POWHEG BOX RES + HERWIG An alternative sample was generated using the same setup of POWHEG BOX RES as the previously described sample. However, the events were interfaced with HERWIG 7.1.6 using the default HERWIG 7.1 tune with the MMHT2014_{LO} PDF set. As for the POWHEG BOX RES + PYTHIA sample, EVTGEN1.7.0 was used to handle the decays of b - and c -hadrons.

SHERPA Another sample was produced with SHERPA 2.2.10. Virtual corrections for the MEs at NLO precision were provided by the Comix [122] and OPENLOOPS [123–125] libraries, together with the MEPS@NLO prescription [126–129]. To merge the MEs with the SHERPA PS [130], a dedicated tune developed by the SHERPA authors

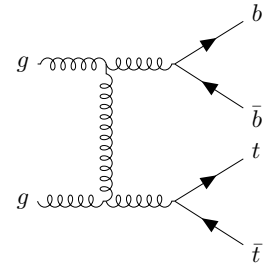


Figure 5.2: Example tree-level Feynman diagram of $t\bar{t}b\bar{b}$ production.

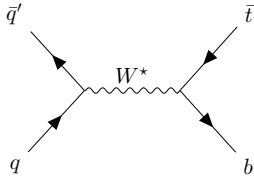


Figure 5.3: Example tree-level Feynman diagram of single top-quark s-channel production.

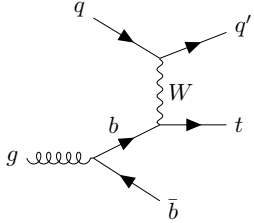


Figure 5.4: Example tree-level Feynman diagram of single top-quark t-channel production. As the sample is generated with a 4FS PDF set, no b -quark can be taken directly from the proton but instead is produced from a gluon.

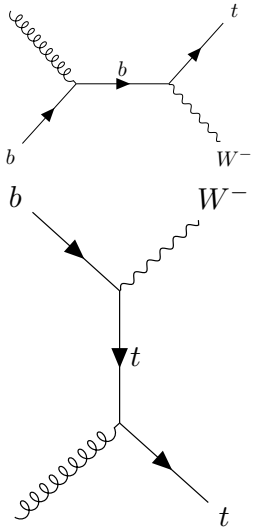


Figure 5.5: Example tree-level Feynman diagrams of tW production.

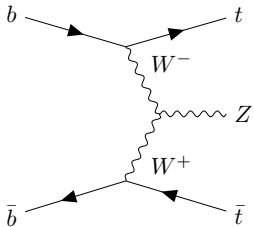


Figure 5.6: Example tree-level Feynman diagram of $t\bar{t}Z$ production.

was used. This sample is only used for a comparison of its predicted cross-section to the measured cross-section in the $t\bar{t}$ + charm analysis.

5.7.3 Single top-quarks

Three different production channels give rise to most of the single top-quark production at the LHC: s-channel production, t-channel production, and production in association with a W boson. Tree-level diagrams of all three production mechanisms can be seen in Figures 5.3 to 5.5. All three production channels were simulated with similar settings.

POWHEG BOX + PYTHIA A first set of events was modelled using the POWHEG Box v2 [131] generator with MEs at NLO precision in α_s . For the t-channel production, the 4FS corresponding NNPDF3.0_{NLO} PDF set was used, while for the s-channel production and the tW production, the 5FS NNPDF3.0_{NLO} PDF set was used. The events were interfaced with PYTHIA8.230 for the PS and hadronisation.

Single top-quark production in association with a W boson in NLO has overlap with $t\bar{t}$ production. This overlap was removed using the diagram removal (DR) scheme [132]. Another dedicated tW sample was generated, which instead uses the diagram subtraction (DS) scheme [117, 132] with otherwise unchanged parameters.

POWHEG BOX + HERWIG Additional samples were simulated using the same settings for the POWHEG Box generator but interfacing the events with HERWIG7.0.4. The H7UE set of tuned parameters [111] and the MMHT2014_{LO} PDF set [118] were used. The DR scheme was used for the tW production.

MADGRAPH + PYTHIA Additional events were modelled with the MADGRAPH5_AMC@NLO 2.6.2 generator. They were generated at NLO precision in QCD using the 5FS and the NNPDF2.3_{NLO} PDF set for all production channels. The events were interfaced to PYTHIA8.230.

5.7.4 Top-quark pairs in association with an electro-weak boson

Two different samples were generated for both, $t\bar{t}Z$ and $t\bar{t}W$ production. Both processes are produced with the same overall settings. An example tree-level Feynman diagram for $t\bar{t}Z$ production can be seen in Figure 5.6, while one for $t\bar{t}W$ production can be seen in Figure 5.7.

MADGRAPH + PYTHIA $t\bar{t}V$ events were modelled with the MADGRAPH5_AMC@NLO 2.3.3 generator. The events were modelled at NLO precision with the NNPDF3.0_{NLO} set of PDFs. They were interfaced to PYTHIA 8.210. For these samples, the decays of b - and c -hadrons were simulated using the EVTGEN 1.2.0 program instead of EVTGEN 1.6.0.

SHERPA Additional samples were produced with the SHERPA 2.2.0 generator. Final states with two leptons from the electroweak boson were produced at LO accuracy with the MEPS@NLO setup for up to one additional parton. The remaining final states were generated with LO accuracy for up to two additional partons. The default SHERPA 2.2.0 PS was used alongside the NNPDF3.0NNLO PDF set.

5.7.5 Top-quark pairs in association with a Higgs boson

An example Feynman diagram for the production of a Higgs boson in association with a top-quark pair can be seen in Figure 5.8.

The POWHEG Box v2 [133] generator was used, calculating the ME at NLO precision in the strong coupling constant α_s while using the NNPDF3.0NLO PDF set. As PS algorithm, PYTHIA 8.230 was used.

5.7.6 Rare single top-quark decays

Two rare single top final states were generated: tZq and tWZ . Example tree-level diagrams can be seen in Figure 5.9 and Figure 5.10 for tZq and tWZ , respectively.

Both samples were modelled at NLO accuracy using the MADGRAPH5_AMC@NLO 2.3.3 generator with the NNPDF3.0NLO PDF set. PYTHIA 8.230 was used for the PS and the hadronisation. For the tWZ sample, the EVTGEN 1.2.0 program was used for the decays of b - and c -hadrons, while for the tZq sample, the EVTGEN 1.6.0 program was used.

5.7.7 Single electroweak bosons

Two samples were generated for Z +jets production. For W +jets production, only one sample was generated with the same settings as for Z +jets production. LO Feynman diagrams for Z and W boson production can be seen in Figure 5.11 and Figure 5.12, respectively. Additional jets can be produced through ISR, FSR, or higher orders.

SHERPA Samples were produced using the SHERPA 2.2.1 generator with the NNPDF3.0NNLO PDF set. NLO precision in QCD was used for diagrams including up to two additional parton emissions, while LO precision was used for up to four additional parton emissions. MEs were merged with the SHERPA-internal PS using the MEPS@NLO formalism. Samples for Z +jets and W +jets were generated using SHERPA.

POWHEG Box + PYTHIA Additional events of Z +jets production were generated with the POWHEG Box v1 generator at NLO accuracy using the CT10NLO PDF set [39]. The events were interfaced to PYTHIA 8.186 with the AZNLO set of tuned parameters [134] and the CTEQ6L1 PDF set [135]. A dedicated simulation of QED FSR

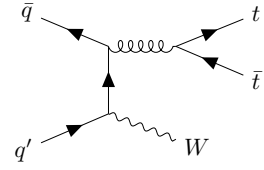


Figure 5.7: Example tree-level Feynman diagram of $t\bar{t}W$ production.

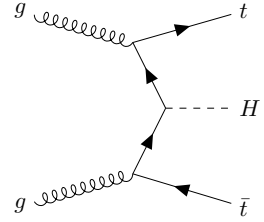


Figure 5.8: Example tree-level Feynman diagram of $t\bar{t}H$ production.

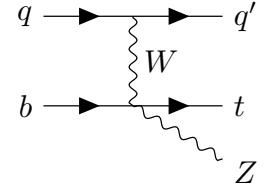


Figure 5.9: Example tree-level Feynman diagram of tZq production.

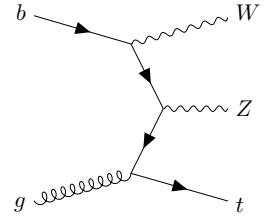


Figure 5.10: Example tree-level Feynman diagram of tWZ production.

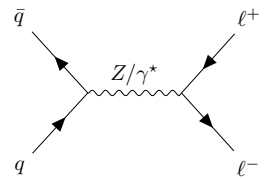


Figure 5.11: LO Feynman diagram of Z +jets production.

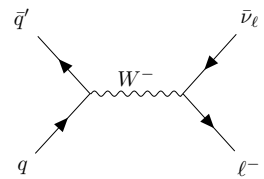


Figure 5.12: LO Feynman diagram of W +jets production.

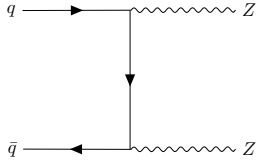


Figure 5.13: Example tree-level Feynman diagram of ZZ production.

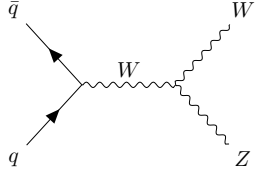


Figure 5.14: Example tree-level Feynman diagram of WZ production.

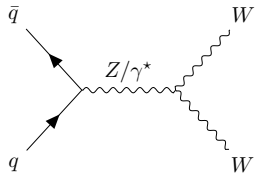


Figure 5.15: Example tree-level Feynman diagram of WW production.

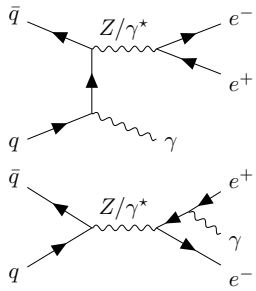


Figure 5.16: Example tree-level Feynman diagram of $Z\gamma$ production. Shown are diagrams with the photon being produced (top) in conjunction with the Z boson and (bottom) as FSR.

was performed with PHOTOS++ 3.52 [136, 137]. The decays of b - and c -hadrons were handled by the EVTGEN 1.2.0 package.

Only Z + jets events with the Z boson exclusively decaying into e^+e^- pairs were produced for this sample.

5.7.8 Two electroweak bosons

Figures 5.13 to 5.15 show example tree-level diagrams of ZZ, WZ and WW, so diboson production.

The SHERPA 2.2.1 or 2.2.2 generator was used to model samples of diboson final states. The choice of the version depends on the specific process; for instance, ZZ production from gluon-gluon fusion was generated using SHERPA 2.2.2, while quark-antiquark initiated ZZ production was generated using SHERPA 2.2.1. Off-shell effects and Higgs boson contributions were taken into account if appropriate. If at least one boson decays leptonically, NLO accuracy in QCD was used for up to one additional parton, while LO precision was used for up to three additional partons. LO precision was used for up to one additional parton if the process is loop-induced ($gg \rightarrow VV$). Using the MEPS@NLO prescription, the MEs are merged with the SHERPA PS. Virtual QCD corrections were provided by the OPENLOOPS library. The NNPDF3.0NNLO set of PDF was used together with a dedicated tune developed by the SHERPA authors.

5.7.9 Z boson in association with a photon

Example tree-level diagrams of the production of a Z boson in association with a photon can be seen in Figure 5.16. Only samples with the Z boson decaying into an e^+e^- pair were generated.

The SHERPA 2.2.4 generator was used to simulate the events. MEs were calculated at LO accuracy in QCD for up to three additional parton emissions. Using the MEPS@LO setup, they were matched and merged with the SHERPA-internal PS. A dedicated set of tuned PS parameters developed by the SHERPA authors was used for the generation of this sample alongside the NNPDF3.0NNLO PDF set.

5.7.10 J/ψ meson decaying into a pair of electrons

Events of promptly produced J/ψ mesons decaying into e^+e^- pairs were generated using PYTHIA 8.186 with the CTEQ6L1 PDF set. QED FSR was handled by PHOTOS++, while decays of b - and c -hadrons were handled by PYTHIA.

5.7.11 Two-to-two processes

The PYTHIA 8.186 generator was used to simulate generic two-to-two processes. These include multijet production, W and Z boson production, $qg \rightarrow q\gamma$ and $q\bar{q} \rightarrow g\gamma$ interactions, and top-quark pair production. The bosons can decay into any particles as long as the decay is allowed by the SM. A filter was applied to veto events in

which the sum of the transverse energy of all particles of the hard scatter¹ was below 17 GeV in all $\Delta\eta \times \Delta\phi$ windows of 0.1×0.1 . This sample will be referred to as JF17 for the remainder of this work.

¹ Muons and neutrinos are excluded from the sum

6 Machine learning

The topic of machine learning becomes increasingly important in today's research in many scientific fields, not just particle physics. This chapter will first give a brief introduction to supervised machine learning, before describing neural networks, and specifically convolutional neural networks and graph neural networks.

Contents

6.1	General introduction	53
6.1.1	Datasets	53
6.1.2	Task	54
6.1.3	Loss function	55
6.2	Neural networks	56
6.2.1	Activation function	56
6.2.2	Regularisation	57
6.2.3	Optimiser and backpropagation	58
6.2.4	Training of neural networks	58
6.3	Convolutional neural networks	59
6.4	Neural networks for point clouds	61
6.5	Performance measures	65
6.6	Preprocessing	66
6.7	Software	67

6.1 General introduction

Machine Learning (ML) is a very broad term, covering lots of different algorithms, methods, and tools. An ML algorithm can be defined as an algorithm that learns to fulfil a specific task based on some input data [138]. The input data usually populates a multidimensional phase space. Supervised algorithms are a subset of ML algorithms, which are trained using some set of n input data points $X = \{\vec{x}_0, \vec{x}_1, \dots, \vec{x}_n\}$ where each data point consists of m input features $\vec{x}_i = (x_{i,0}, x_{i,1}, \dots, x_{i,m})$. Each of these data points has a known true label y_i associated with it. The goal is to infer this label based on the input data.

6.1.1 Datasets

During the training process of any supervised ML model, the optimisation is done based on some data referred to as training data. However, the best performance on previously unseen data is wanted. Choosing the best model purely based on its performance on the training data will easily lead to a model that learns specific properties of the training data that are not present in the data that the model will eventually be evaluated on. This means that the complexity of the model is too high. With increasing model complexity, the model can better represent the training data and improve its performance on

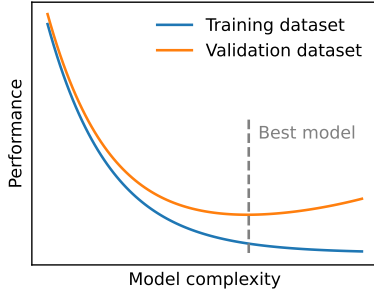


Figure 6.1: Sketch of the performance of a ML model with different complexities.

it. On unseen data, however, performance will eventually deteriorate. This problem is also known as overfitting.

To reduce this problem and obtain the best performance on previously unseen data, a so-called validation dataset is used. Instead of choosing the model performing best on the training dataset, the model performing best on the validation dataset is chosen. Figure 6.1 shows a sketch of the performance of a model with different model complexities for the training and the validation datasets.

To get a final, unbiased evaluation of the ML model, the validation dataset cannot be used as it was used during the training process. Therefore, a third dataset, the test dataset, is used to evaluate the final performance of the model. Typically, the training, validation, and testing datasets are obtained from the same initial dataset by splitting data points at random.

6.1.2 Task

Several tasks can be performed by ML. Among others, they include synthesis and sampling, anomaly detection, denoising, and machine translation [138]. Some of these are also utilised in High Energy Physics (HEP). As an example, the simulation of the detector response is a major bottleneck regarding computation time in the ATLAS experiment. Generative models are studied to replace the full detector simulation and speed up the simulation of events drastically without sacrificing accuracy [139–142].

A single algorithm can also perform multiple tasks at once. In this work, two of the most basic tasks will be used: regression and classification.

REGRESSION During regression, a ML model is asked to predict a numerical value based on the input data. The most typical use of regression models inside ATLAS is to predict the properties of some particle. This can, for instance, be used to correct the measured energy of electrons and photons to their true energy [143] or to predict the missing transverse momentum [144]. A single numerical value or a vector of values, like the four momentum of a particle, can be the target.

CLASSIFICATION Classification is the task of predicting which of k predefined classes a data point belongs to. Usually, classification models predict probabilities of being in each of the k output classes. If there are only two classes, it is also called *binary* classification, while for more than two classes, it is called *multinomial* classification. In binary classification, a single output is returned, giving the probability of being one of the two classes. The probability of being the other can easily be obtained by subtracting the output of the algorithm from one. In HEP, the class for which the probability is predicted is usually referred to as signal, while the other is called background. The output, or score, can be used to select samples enriched in signal

or background events.

For multinomial classification, one output is returned for each of the k classes. These can be combined into a single discriminant using a likelihood ratio following the Neyman-Pearson lemma [145]

$$\frac{p_0}{\sum_{i=1}^k f_i p_i}, \quad (6.1)$$

where p_i is the output probability of the i -th class and f_i a weight for each of the classes in the denominator. With a likelihood ratio like Equation (6.1), class 0 is considered as signal, while all other classes are considered background. Using the f_i , the importance of the single classes can be changed. A higher value will give greater weight to this class. To choose the f_i , an additional requirement of $\sum_{i=1}^k f_i = 1$ can be imposed. This will not change the discrimination power of the discriminant but only scale it. If desired, multiple classes can be placed in the numerator. The most typical application of classification in ATLAS is the identification of particles like electrons [2, 146], identifying the flavour of a jet [82, 147, 148], or the separation of signal and background in a physics analysis [8, 149–152].

6.1.3 Loss function

To train a supervised ML model, the model needs to know how well it is currently performing. This is evaluated by the loss function¹. During training, the goal of the model is to minimise the loss function. The loss function of a dataset is calculated by averaging the loss value of the single data points

$$\mathcal{L}(\hat{\vec{y}}, \vec{y}) = \frac{1}{N} \sum_{i=1}^N \mathcal{L}(\hat{y}_i, y_i), \quad (6.2)$$

where $\hat{\vec{y}}$ is the set of predictions of the model, $\vec{y}$ the corresponding true labels, N the number of data points in the set, and $\hat{y}_i$ and y_i single data points of $\hat{\vec{y}}$ and $\vec{y}$, respectively. Different loss functions will result in different solutions, which might not be ideal depending on the problem. For instance, for regression problems, two common loss functions are the Mean Absolute Error (MAE)² and the Mean Squared Error (MSE)³. When the model has to predict the same value for all data points, the MAE is minimised by setting this prediction to the median of $\vec{y}$, while for the MSE, the mean minimises the loss. Depending on the goal of the model, one or the other can be more beneficial. Both loss functions are visualised in Figure 6.2.

For binary classification, the most commonly used loss function is the binary cross-entropy

$$\mathcal{L}(\hat{y}, y) = -y \log(\hat{y}) + (1 - y) \log(1 - \hat{y}). \quad (6.3)$$

Its generalisation to multinomial classification is given by the categorical cross-entropy

$$\mathcal{L}(\hat{y}, y) = - \sum_c y_c \log(\hat{y}_c), \quad (6.4)$$

¹ Sometimes also called the cost or objective function

² $\mathcal{L}(\hat{y}, y) = |\hat{y} - y|$

³ $\mathcal{L}(\hat{y}, y) = (\hat{y} - y)^2$

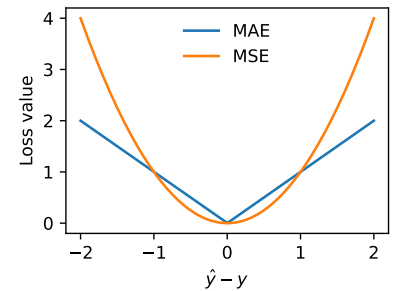


Figure 6.2: MAE (blue) and MSE (orange) loss functions

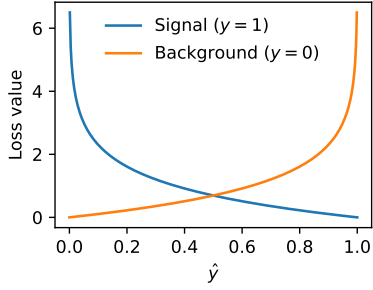


Figure 6.3: Binary cross-entropy loss function for signal ($y = 1$) in blue and background ($y = 0$) in orange

¹ The naming of NNs is not always consistent. Most of the time, any NN with multiple hidden layers is referred to as DNN where the 'D' stands for 'Deep'. In this work, only feed-forward NNs will be referred to as DNN

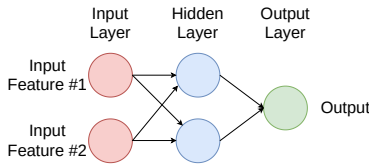


Figure 6.4: Sketch of a neural network with two input features, one hidden layer with two nodes, and one output layer with a single node.

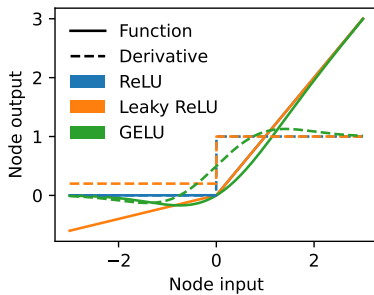


Figure 6.5: ReLU (blue), Leaky ReLU (orange), and GELU (green) activation function and their derivatives around zero

² In TensorFlow, a value of 0.3 is the default value. Other values can be chosen as well. A value of one would result in an identity function while a value of zero would result in the standard ReLU

with c referring to the single classes. Figure 6.3 shows a visualisation of the binary cross-entropy.

6.2 Neural networks

One of the most prominent algorithms for ML are Neural Networks (NNs). They cover a wide range of different types of algorithms. The most basic NN is a Feed-Forward Neural Network (DNN)¹. DNNs are mostly used for tabular input data, whereas more complicated network architectures like Convolutional Neural Networks (CNNs) or Graph Neural Networks (GNNs) are used for structured data in the form of images or sets.

The main building block of a NN is a linear layer. It applies an affine transformation to all the input data. This can be written as

$$\vec{x}_{\text{out}} = W\vec{x}_{\text{in}} + b, \quad (6.5)$$

where W is a weight matrix, b is a bias vector, and $\vec{x}_{\text{in,out}}$ are the input and output feature vectors. The weight matrix and the bias vector are both learnable parameters of the layer. To build a NN, one or multiple of these layers can be applied in sequence, as depicted in Figure 6.4. However, as a single layer is an affine transformation, the combined output will also be an affine transformation. This does not allow the NN to learn any non-linearities. To be able to learn non-linearities, *activation functions* are utilised. Including these non-linearities, DNNs are a class of universal approximators [153].

6.2.1 Activation function

To be able to learn non-linear functions, NNs require some non-linearity to be injected. This is achieved by passing the output of the linear layers through an activation function.

A commonly used activation function is the Rectifier Linear Unit (ReLU) defined by

$$f(x) = \begin{cases} x, & \text{if } x \geq 0. \\ 0, & \text{otherwise.} \end{cases} \quad (6.6)$$

It maps all negative inputs to zero while applying the identity function to positive values. This definition provides a non-linearity and a derivative that is easy to calculate. Figure 6.5 shows the behaviour of the ReLU as well as its derivative. Modifications to the ReLU often lead to performance improvements of the final NN. Such modifications are the Leaky Rectifier Linear Unit (Leaky ReLU) [154], modifying the negative part to have a slope of some value², and the Gaussian Error Linear Unit (GELU) [155] defined by

$$f(x) = x \cdot \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x}{\sqrt{2}} \right) \right] \quad (6.7)$$

with $\operatorname{erf}(x)$ being the Gauss error function. Both modifications and their derivatives can be seen in Figure 6.5 as well. The Leaky ReLU,

contrary to the ReLU, does not remove all outputs below zero but instead keeps them scaled by a factor. The GELU behaves like the ReLU at values far from zero but has a continuous derivative at zero.

Another commonly used activation function is the sigmoid function

$$f(x) = \frac{1}{(1 + \exp(-x))}. \quad (6.8)$$

It is bound between zero and one. Used as an activation of the last layer of a NN for binary classification, the outputs can be interpreted as a probability of being of the signal class. Figure 6.6 shows the sigmoid function and its derivative around zero. For NNs with multiple outputs for multinomial classification, the softmax function can be used as a generalisation. It is defined as

$$f(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)}, \quad (6.9)$$

ensuring that the outputs of all classes sum up to one $\sum_j f(x_j) = 1$. Therefore, the single outputs can be interpreted as the probability of belonging to the specific class of that output. For regression tasks, typically no activation function is applied to the output of the final layer. If only positive outputs are desired, the softplus activation function¹ can be applied.

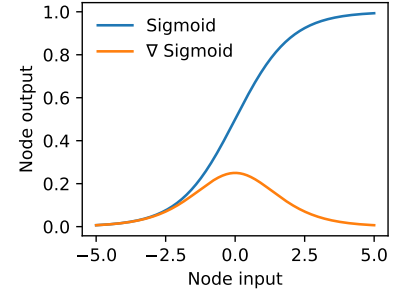


Figure 6.6: Sigmoid activation function and its derivative around zero

¹ $f(x) = \log(1 + \exp(x))$

6.2.2 Regularisation

To reduce the problem of overfitting for NNs, regularisation techniques such as L2 regularisation, dropout, or normalisation layers can be used.

L2 REGULARISATION L2 regularisation, also called weight decay, is a regularisation technique that adds a term to the loss function. This term is the sum of the squared weights of the NN multiplied by a parameter λ . The value of λ needs to be chosen before the training, with a larger value leading to a stronger regularisation effect. The additional loss term penalises large values of weights, forcing the weights towards smaller values, hence the name weight decay. Smaller weights will lead to smaller deviations when input features are changed, making the model more robust to slight differences in the input data.

DROPOUT Dropout [156] is another way of combating overfitting. During the training, random nodes are “dropped” out of the NN, meaning they are temporarily removed from the NN. This introduces a noise element to the training. In an NN without dropout, single nodes can be used to fix the mistakes of other nodes. This co-adaptation of nodes can easily lead to overfitting. When randomly dropping nodes, the node that fixes the mistake of another node might be dropped, forcing the original node to correct the mistake. Therefore, single nodes are forced to give more meaningful representations of the data, which generalises better to previously unseen data.

Additionally, the same information needs to be able to be passed down multiple paths through the network, making it more robust. Utilising dropout can also be interpreted as training many NNs on the training dataset with different architectures.

NORMALISATION Another means of regularisation is normalisation. It normalises and rescales the outputs of a layer in a NN. The two most popular normalisation operations are batch normalisation [157] and layer normalisation [158]. Batch normalisation subtracts the mean and divides by the standard deviation, taking the mean and standard deviation separately for each feature over all data points in the batch. To be able to do the same while evaluating the model on potentially a single data point, a moving average of the mean and the standard deviation is recorded during training for use during evaluation. Layer normalisation instead takes the mean and the standard deviation over all features in a layer for each data point separately. This makes the operation exactly the same between training and evaluation.

6.2.3 Optimiser and backpropagation

To update the learnable parameters of a NN, an optimiser is utilised. Many optimisers exist, with many being based on Stochastic Gradient Descent (SGD). The SGD optimiser evaluates the gradient of the loss function with respect to the learnable parameters of the NN. The learnable parameters are then updated, taking a step in the direction of a negative gradient to eventually find a minimum in the loss function. A single parameter is updated by

$$\theta_{i+1} = \theta_i - \eta \nabla_{\theta} \mathcal{L}_{\theta}(\hat{\mathbf{y}}, \mathbf{y}), \quad (6.10)$$

where θ_i is the learnable parameter at step i and η is the learning rate (LR) which defines the step size. Extensions to SGD utilised in this thesis include the *Adam* [159] and *AdamW* [160] optimisers.

As the gradients are used for the update, they need to be calculated. However, this cannot be done analytically in a feasible way. Therefore, an algorithm called backpropagation [161] is used. It gradually calculates the gradients starting from the last layer and propagating backwards through the NN.

6.2.4 Training of neural networks

During the training of a NN, the learnable parameters are optimised by passing a small set of data points called a *batch* through the NN. The value of the loss function of this batch is calculated and used to update the learnable parameters of the NN to reduce the loss value. This is repeated for all batches in the training dataset. After all batches are processed, also called an *epoch*, the performance of the NN on the validation dataset is evaluated. These training steps are repeated for multiple epochs. In the end, the best model during the training based on the validation loss can be picked.

6.3 Convolutional neural networks

Convolutional Neural Networks (CNNs) are commonly used for image processing. To exploit the spatial relationship between the single pixels, a simple linear layer as described in Section 6.2 cannot be used. Instead, so-called filters are used. The input to a CNN is an $a \times b \times c$ image, where a is the width in pixels, b the height in pixels, and c the number of channels, like red, green, and blue for regular coloured images. To reduce the dimensions of the images, pooling operations are applied.

FILTER The filter is a three-dimensional matrix, $m \times n \times c$, where the channel dimension has to be the same as in the input images. Figure 6.7 shows an example of a convolutional operation with an input dimension of 4×4 and a filter dimension, also called kernel size, of 2×2 . Here, the channel dimension is, for simplicity, equal to one. The filter slides across the image, performing a pixel-wise multiplication with the image at each step, and placing the sum of the multiplication in the output image. In the sketch, the filter slides one pixel at a time, though larger values could be chosen. This value, called stride size, is a hyperparameter of the model and needs to be fixed before the training. Figure 6.7b shows the first multiplication, highlighting the position of the filter on the input image in green, whereas Figure 6.7c and Figure 6.7d show the status one step further and at the end, respectively. For a channel dimension larger than one, the single channels are processed in the same way, summing the outputs of the single channels after this operation pixel-wise. The values of the filter matrix are the learnable parameters of the operation. Usually, multiple filters are used at each convolutional operation, similar to multiple hidden nodes being used in each layer of a DNN. As can be seen from Figure 6.7, the dimension of the output image is smaller than that of the input image. This can be avoided by using padding. The simplest version is zero-padding, where the dimension of the input image is extended with pixels having values of zero such that after the convolutional operation, the output image has the same dimension as the input image before any padding.

POOLING The convolutional operation described in the previous paragraph reduces the dimension of the input image. Another option for reducing it is the use of pooling operations. For a pooling operation, a $k \times l$ window is defined. This window slides across the input image just like the filter. At each step, a pooling operation is applied to the pixels overlapping this window. Possible pooling operations include, for instance, taking the maximum or the mean. Figure 6.8 shows a maximum pooling with a 2×2 pooling size and a stride length of one on a 3×3 input.

To further process the output of the convolutional operations with linear layers, the image can be flattened. That means the single pixels are unrolled into a vector, removing spatial information but introduc-

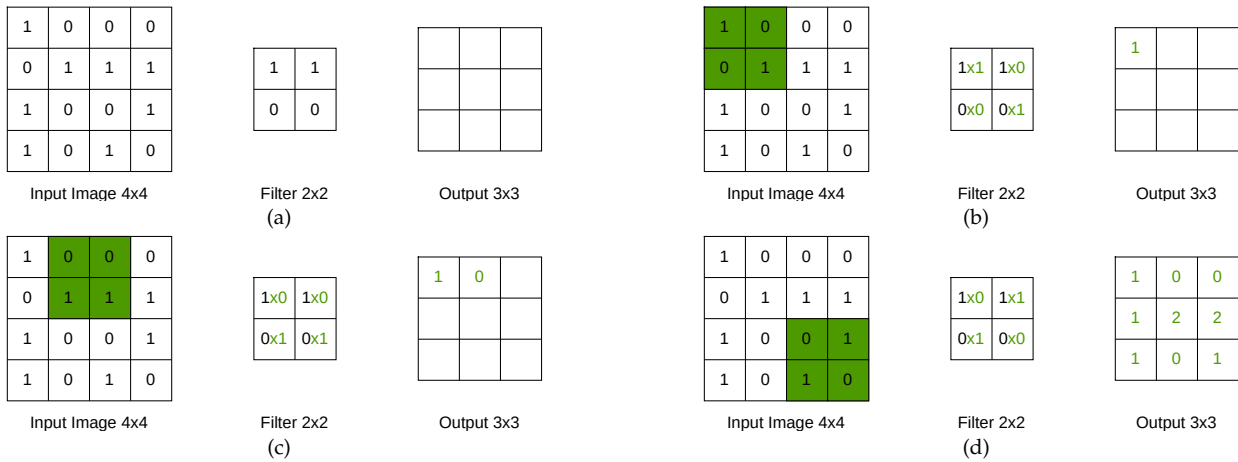
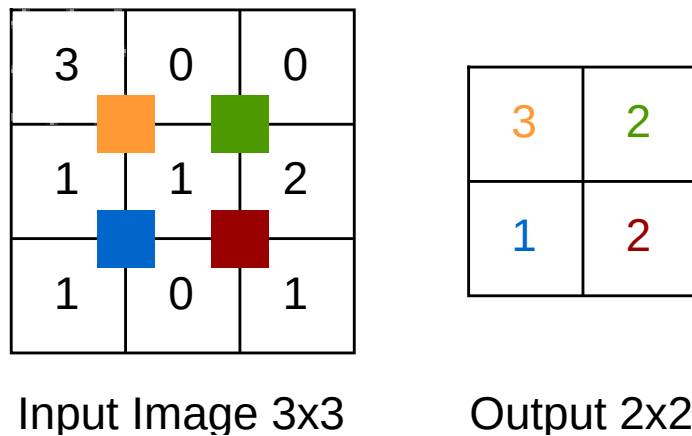


Figure 6.7: Overview of a convolutional operation. A 4×4 image is processed by a 2×2 filter resulting in a 3×3 output. Single steps are taken through (a) to (d).

Figure 6.8: Overview of a pooling operation using the maximum. A 3×3 image is pooled using a 2×2 window. The window for the different pooling operations is indicated by the coloured squares.



ing a structured order.

Applications in ATLAS mostly involve particle identification [146, 162] interpreting the cells of the calorimeter as pixels of an image. Outside ATLAS, CNNs are also explored in terms of jet tagging [163], where jet tagging refers to identifying the particle that originated a jet. The decay products of said particle are inside the jet as it is highly boosted; for instance, the W boson and the b -quark can be contained in one top-quark-initiated jet. This task is not to be confused with flavour tagging. For jet tagging, the images are created based on the jet constituents. However, CNNs are outperformed in this task by graph neural networks [164], which will be introduced in the following.

6.4 Neural networks for point clouds

While CNNs work very well for image data, there is no actual image data being recorded in the ATLAS detector. The closest thing to being an image is the calorimeter when interpreting the single cells as pixels. However, the size of the cells is not constant between the different layers and sometimes changes even inside a layer. Therefore, the creation of any images at ATLAS has to make some compromises. It has to enforce a structure and might introduce artificial padding. Especially the enforced structure can cause issues, for instance, if one wants to represent all jets in an event. Typically, the η and ϕ coordinates are used to determine which pixel the jet falls in. These two variables are, however, continuous, and by forcing them into a grid structure like an image, information will be lost.

A representation that is not restricted to the rigid grid structure of images are sets of points, also called point clouds. These point clouds are made of n data points that all have the same type of features $\vec{x}$. Furthermore, there is no intrinsic ordering in this set. This representation of the data is more natural for most of the data collected at the ATLAS detector.

Forming connections between single data points, so-called edges, creates a graph. In the following, Graph Neural Networks (GNNs), DeepSets, and transformers will be introduced, which can operate on point clouds or graphs. A major point of all of these architectures is that they are invariant under permutations of the single points in the point cloud.

DEEPSETS One of the simplest architectures for point clouds is the DeepSets architecture [165]. The basic idea is to first transform all elements of the point cloud through one function ϕ , apply a permutation-invariant pooling operation like summing, and transform the output by another function ρ . The final output is therefore given by $\rho(\sum_i(\phi(\vec{x}_i)))$ where i refers to the single instances of the set. This ensures permutation invariance as long as the pooling operation, in this case the sum, is permutation invariant. Both of the functions ϕ

and ρ can be expressed by simple DNNs.

DeepSets architectures have become more prominent inside the ATLAS experiment over the last couple of years, with them being utilised primarily for particle identification [148, 162, 166].

GRAPH NEURAL NETWORKS A general GNN [167, 168] has three types of input data: global features G which is tabular data; nodes N which are the set of instances; and edges E which are connections between nodes. The edges can be a part of the actual data with their own features, or they can be defined from the existing nodes and only used as a device to exchange information between the single nodes. Figure 6.9a shows an overview of a general graph operation.

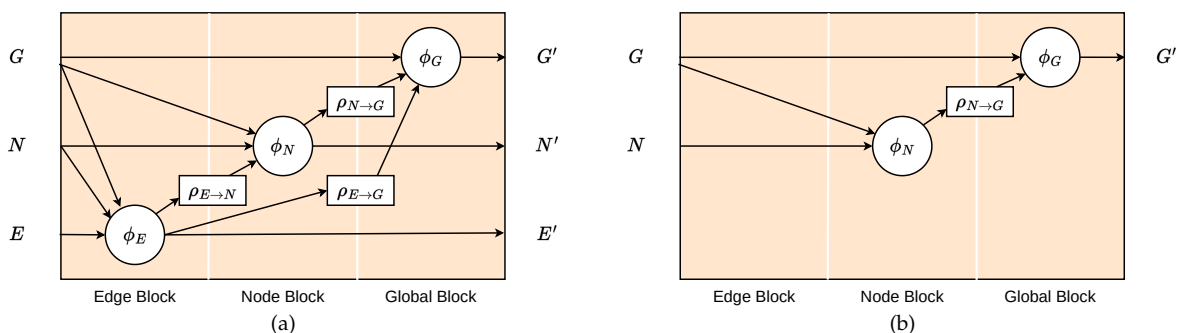


Figure 6.9: Schematic of the processing of edges E , nodes N , and global features G of (a) a general GNN operation and (b) a DeepSets. Inspired by Ref. [167]

The ϕ represent learnable functions or NNs, whereas the ρ represent some kind of pooling operation. The operation starts in the edge block. The edges are processed by a learnable function ϕ_E with the nodes and the global features being included in the function as well. The output of this operation is pooled by two different functions, one for further usage in the node block ($\rho_{E \rightarrow N}$) and another for further usage in the global block ($\rho_{E \rightarrow G}$). Furthermore, the updated edges E' are used as input for the next graph operation, if there is one. After the processing of the edges, it is time for the nodes. The nodes are, similar to the edges, updated by a learnable function ϕ_N . The input consists of the nodes, the global features, and the pooled edges. The output is once again pooled ($\rho_{N \rightarrow G}$) for further usage in the global block or passed to the next graph operation (N'). Lastly, the global features are updated (ϕ_G) with the inclusion of the pooled nodes and the pooled edges. Again, the output can be utilised further in the next graph operation.

The DeepSets architecture can also be represented in this way, as can be seen in Figure 6.9b. In this case, there are no edges at all, and the updated nodes are not kept.

USED GRAPH OPERATION Often, edges are defined by the structure of the data. For instance, if a molecule is represented as a graph, edges can represent the bonds within the molecule [169]. In all datasets utilised in this work, there are no predefined edges. Instead, edges need to be defined by hand. This can be done, for instance, by finding nodes in the vicinity using a k-Nearest Neighbours (kNN) algorithm

and connecting these nodes. Another option is to connect all nodes to each other. If no predefined edges are available that can have features associated with them, the edge features are usually obtained by concatenating the features of the two nodes attached to the edge.

In this work, all edges are directed edges, meaning they send information from one node to another with another edge corresponding to the other direction. The mainly used graph operation for all GNNs in this work can be seen in Figure 6.10. It shows how the node n_1 is updated based on the remaining nodes. The remaining nodes are updated in the same way, though this is not shown for visualisation purposes. This operation is based on the Interaction Network (IN) [170], though other architectures like GRAVNET and GARNET [171] exist.

First, in Figure 6.10a, edges are formed between the nodes by concatenating the node features of the attached nodes. These edges are passed through the edge network ϕ_E , see Figure 6.10b. If global features are available, these are concatenated with the edge features before being passed through the edge network. As can be seen in Figure 6.10c, the outputs of the edge network are pooled using the pooling function $\rho_{E \rightarrow N}$ before being concatenated to the original node features, creating a first updated node $\tilde{n}_1$. Lastly, this updated node is passed through the node network ϕ_N in Figure 6.10d, potentially concatenating global information onto it beforehand. The output of this operation is the new updated node n'_1 . Global features are never updated by the graph operation. If a single output is required, a pooling operation is applied to the nodes ($\rho_{N \rightarrow G}$). The pooled nodes are concatenated with the global features and any additional scalar variables before being processed by a simple DNN.

Few results using GNNs have been published by the ATLAS experiment. Most of the published results focus on particle identification [147, 162] with only a few applications in physics analyses [172]. However, many papers have been published using publicly available or privately produced datasets spanning a wide variety of topics, from jet tagging to generation, tracking, and more [3, 173–189].

TRANSFORMERS Transformers were originally developed for natural language processing [190], but they have also gained traction in other domains like HEP [191–194]. Often, only a part of the full transformer, as introduced in Ref. [190] is utilised with the transformer encoder layer. At the heart of the transformer concept is the concept of attention. The basis is the so-called scaled dot-product attention. As inputs, three sets of instances are used: queries (Q), keys (K), and values (V). The attention is calculated by

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V, \quad (6.11)$$

where d_k is the number of features of the queries and the keys. The scaling with $\sqrt{d_k}$ is motivated by reducing the size of the input of the softmax activation function, as for large input values, the softmax

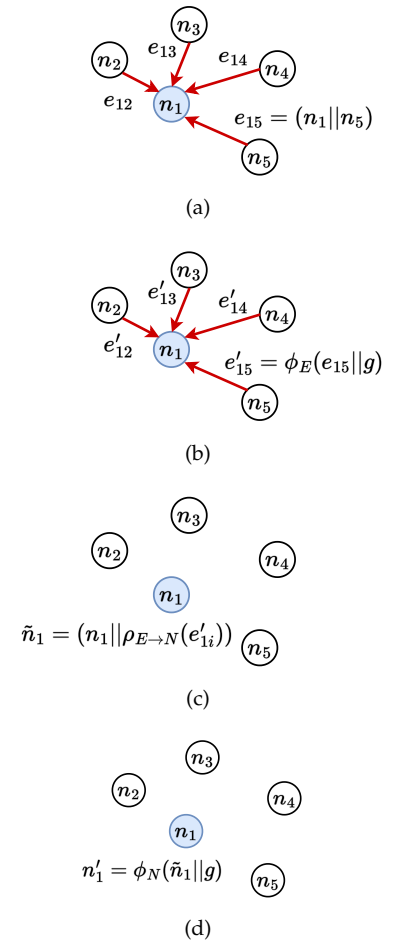


Figure 6.10: Overview over the graph operation used in this work. The “||” represents a concatenation of the two vectors.

activation function has small gradients. Instead of calculating one single attention score, multiple are calculated by splitting the features of queries, keys, and values into multiple heads. Each of the sets of subfeatures is multiplied by a different weight matrix. Afterwards, the attention score is calculated for each head separately. Finally, the different attention scores are concatenated and multiplied with another weight matrix. This procedure is called Multi-Head Attention. Its calculation can be written as

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_n)W^O \quad (6.12)$$

$$\text{with head}_i = \text{Attention}(Q_iW_i^Q, K_iW_i^K, V_iW_i^V). \quad (6.13)$$

The weight matrices W are learnable parameters of the model.

For this work, two types of Multi-Head Attention are used: self-attention and cross-attention. Self-attention refers to the queries, keys, and values being the same set before being passed into the Multi-Head Attention operation. For each single instance of the set, it retrieves a representation based on all instances of the set. For cross-attention, the queries are a different set from the keys and values. Keys and values are still the same. In this case, for each instance of the queries, a representation of all instances of the keys and values is obtained.

The queries can also be interpreted as a question that should be answered by the operation. The keys represent all possible answers, which can also be combined. By computing the attention matrix, a value is assigned to each query-key pair about how well they match together, meaning how well the question is answered by the key. Lastly, given the attention, the answer to the query is extracted from the values, with the values being all possible answers as the keys.

A full transformer encoder layer as used in this work can be seen in Figure 6.11. As keys and values are always the same for this work, they are represented by K/V . Queries, keys, and values are first normalised before being passed into a Multi-Head Attention operation. The output of this operation is again normalised before being added to the original queries. The sum is then passed through a DNN to introduce non-linearities, with the output being added to the output of the Multi-Head Attention to obtain the output of the full transformer encoder layer. This layer can be used multiple times in sequence. In this layer, two skip connections are used, adding a previous output to the current values. They ensure some locality of the information. The Multi-Head Attention can learn arbitrary information from all input nodes. The skip connection reintroduces the original features, sort of “reminding” the operation about the original state.

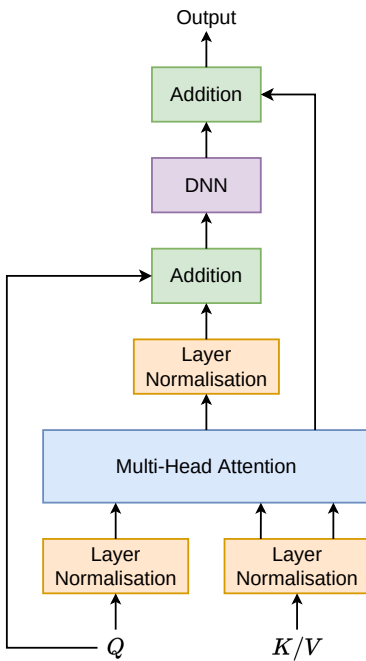


Figure 6.11: Overview of a transformer encoder layer as utilised in this work.

ATTENTION POOLING Similar to the transformer attention mechanism, one can also use the concept of attention for pooling over nodes or edges in DeepSets, or GNNs. In attention pooling, part of the pooling operation is learned by the network. Instead of just summing the outputs of the ϕ network, the sum is weighted $\sum_i w_i \phi(\vec{x}_i)$. The weights are learned by an additional small DNN. Additionally, the

weights are passed through a softmax activation function, such that all the weights sum up to one ($\sum_i w_i = 1$). This kind of pooling operation allows the network to focus more on important instances of the set while ignoring unimportant ones.

6.5 Performance measures

The performance of many ML algorithms can be evaluated by the value of the loss function, where a lower value typically corresponds to better performance. However, specific metrics and figures of merit can be beneficial in investigating the performance of these algorithms further. Depending on the problem that should be solved, the best metric can vary.

REGRESSION For regression tasks, precision and accuracy are important. Accuracy is given by the distance of the predicted value from the true value, whereas precision refers to predicting the same value for the same object. Accuracy can also be interpreted as a scale, while precision represents the resolution of the prediction. For the evaluation of both, it is beneficial to look at the residuals of the prediction given by $\hat{y} - y$ which can additionally be normalised by $\hat{y}$ as long as the values of $\hat{y}$ are not close to zero.

To measure the accuracy of the model, the mean of the residuals can be used. If the mean is zero, there is no bias in the predictions of the algorithms. Alternatively, one could also use the median.

For precision, the interquartile range (IQR) can be used. It is defined as the difference between the 75th percentile and 25th percentile of the residuals, effectively giving the range the middle 50% of the distribution span. A lower value corresponds to a better resolution. By only taking the middle 50% of the distribution, this metric is robust to outliers, which are predictions that are far off the true value. The standard deviation is an alternative that is more sensitive to outliers.

CLASSIFICATION For most classification algorithms used in this work, it is only important to reject as much background as possible at high signal efficiencies. A visualisation of the performance of a binary classifier is the receiver operating characteristic (ROC) curve. To obtain a ROC curve, the signal efficiency and background efficiency are calculated at different values of the discriminant, where efficiency is defined as the number of candidates with a discriminant value larger than the chosen value divided by the total number of candidates. In this work, instead of using the background efficiency, typically the background rejection, defined as the inverse of the background efficiency, is used. By plotting the signal efficiency against the background rejection, the best classifier can be determined, as can be seen in Figure 6.12. The better a classifier is, the further the curve will be located towards the top-right corner of the plot.

A step-like behaviour is clearly visible at low signal efficiencies for the curve with the highest background rejection. This is caused by

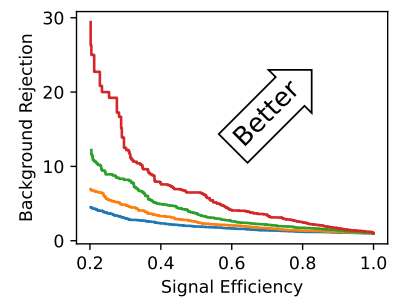


Figure 6.12: Schematic overview of ROC curves for different classifiers.

low statistics. Several signal candidates are removed when lowering the threshold on the discriminant score, but no additional background candidates are removed, such that the signal efficiency is reduced while the background rejection remains constant.

6.6 Preprocessing

Using available data just as it is during the training of NNs might not lead to the best performance possible with the given data. Additionally, it might introduce features that are undesirable, such as a dependence on a feature that is not wanted. To mitigate these, the data needs to be preprocessed.

SCALING OR TRANSFORMING NNs perform best when all input features are of order one. This will improve the stability of the training process and might potentially improve the final performance of the model. To visualise this, one can consider two inputs, $x_1 = 1$ and $x_2 = 100$. In a NN, these will be combined by $w_1x_1 + w_2x_2$, where w_1 and w_2 are the weights of a hidden layer. These weights will typically have similar values at the start of the training process. This, however, leads to x_1 not contributing much, if at all, to the output of this hidden layer. Instead, it will be dominated by x_2 . There is also no guarantee that with an optimisation like gradient descent, the weights will change such that x_1 will be considered more by the NN. Scaling or transforming of the input features before passing them into the NN can mitigate these issues.

Three different algorithms, all implemented in the scikit-learn package [195], are considered in this thesis: StandardScaler, RobustScaler, and QuantileTransformer. Each algorithm acts on each feature separately. The StandardScaler subtracts the mean of the input feature and divides it by its standard deviation. This centres the distribution at zero with a standard deviation of one. This algorithm is, however, not robust to outliers. For distributions with many outliers, the RobustScaler can be used instead. It implements the same scaling but uses the median instead of the mean and the IQR instead of the standard deviation. These properties are more robust towards outliers, and therefore the scaling is more robust. The QuantileTransformer does not implement an affine transformation like the StandardScaler or the RobustScaler but the transformation is still monotonic. It transforms each variable separately using the quantiles of the input distributions. The single quantiles are matched to the quantiles between zero and one, leading to an output distribution that is rectangular between zero and one.

Figure 6.13 shows the effect of these different preprocessing algorithms on a distribution with 100 000 entries following a Gaussian distribution with a mean of one and a standard deviation of 0.5. Additionally, 400 events with a value of 40 are added to emulate the effect of outliers. Both the StandardScaler and the RobustScaler centre the distribution around zero, with a slight shift towards lower values for

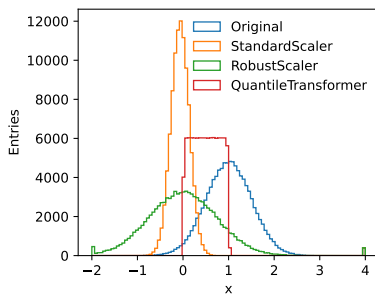


Figure 6.13: Gaussian distribution with some outliers before and after three different preprocessing algorithms. The first and last bin contain under- and overflow events, respectively.

the StandardScaler due to the outliers. A large difference can be seen in the standard deviation of these two output distributions. While the RobustScaler gives an output distribution with a standard deviation of around one, the standard deviation of the output distribution of the StandardScaler is much smaller. Both algorithms keep the outliers at rather larger values, which can be seen by the contribution to the overflow bin. The QuantileTransformer does not suffer from this problem. It maps all events between zero and one. However, as the transformation is not affine, correlations between the variables can be significantly changed.

BIAS MITIGATION Often, variables can be important for an algorithm, although the algorithm should not base its decision on them. An example is the pseudorapidity and the transverse momentum of an electron candidate. The features of a signal electron will vary significantly with respect to these variables; however, the algorithm should not prefer electrons at specific values of these. It should be unbiased regarding η and p_T .

To mitigate biases introduced by such variables, the training data can be modified such that the input distributions of signal and background for binary classification or, more generally, all classes in multinomial classification are the same. This can be achieved through a variety of options. One can remove candidates from classes that have more candidates in a specific range of these variables. Additionally, one can apply weights to single candidates. Lastly, one can also sample from the original biased training dataset during the training procedure. In this case, an epoch cannot be defined as passing every candidate once through the network. These single methods can also be combined.

Another bias that can occur in a classification algorithm is bias towards a certain class. If the training dataset is dominated by a specific class, this class will by default get a higher score than the other classes, therefore biasing the classifier towards this class. To mitigate this impact, the same techniques as described before can be employed. The weighting procedure, however, simplifies significantly as only a single weight is needed for each class, such that each class has the same sum of weights in the training dataset.

6.7 Software

Many libraries exist to perform ML. Two of the most popular libraries are TensorFlow [196] and PyTorch [197]. Most of the models trained for this thesis use TensorFlow. Different minor versions of version 2 were used. The transformer models in Part IV were trained using PyTorch version 2.

These two libraries cannot easily be integrated into the ATLAS software stack. To combat this, the lwttn [198] package was developed to include NNs into the ATLAS software. More recently, the open-source format ONNX using ONNX Runtime [199] has been integrated.

ONNX aims to provide a format that is easily exchangeable between different tools. As lwttn does not support GNNs, these are interfaced with the ATLAS software utilising the ONNX format.

PART II

ELECTRON IDENTIFICATION AT THE ATLAS EXPERIMENT

7 Current approach

Many objects can mimic the signature of electrons in the detector, leading to them being reconstructed as electrons. In a physics analysis, these other objects should not be considered. Additionally, not every electron should be considered. Only prompt, isolated electrons are of interest for most analyses. This excludes, for instance, electrons that originate from the decay of heavy-flavour objects, as long as they were not produced promptly. Therefore, an identification algorithm is applied to all reconstructed electrons. It aims to reduce the contributions of all background types while maintaining a high amount of actual, prompt, isolated electrons.

Contents

7.1 Likelihood algorithm	71
------------------------------------	----

7.1 Likelihood algorithm

The identification of electrons at the ATLAS experiment is based on a likelihood (LH) approach [73, 200]. A wide variety of information from different subsystems is used. A detailed description of the input variables can be found in Ref. [200]. For all input variables, one-dimensional Probability Density Functions (PDFs) are created separately for signal and background. These do not take correlations between the input variables into account. Each of the PDFs is obtained by applying an adaptive kernel density estimator as implemented in the TMVA toolkit [201] to histograms of the input variables. To obtain the likelihood of a candidate being signal (L_S) or background (L_B), the values of the respective PDFs are multiplied together:

$$L_{S(B)}(\mathbf{x}) = \prod_i P_{S(B),i}(x_i),$$

with $P_{S(B),i}$ being the signal (background) PDF and x_i the value of the i -th input variable. These two likelihoods are combined into one discriminant d_L similar to Equation (6.1)

$$d_L = \frac{L_S}{L_S + L_B}.$$

An additional monotonic transformation is applied, which does not change the separation power

$$d_L = -\frac{1}{15} \ln(d_L^{-1} - 1).$$

This separates the signal and background distributions into two clear peaks. An example of the final discriminant can be seen in Figure 7.1.

As the distribution of some input variables changes considerably with E_T and $|\eta|$, different PDFs are obtained in bins of E_T and $|\eta|$. The bin edges used can be seen in Table 7.1. Additionally, it shows a different binning in E_T , which is used to define thresholds on the final discriminant. This gives additional flexibility in defining target signal efficiencies over the whole phase space. However, using different cut

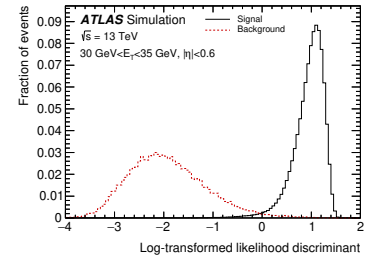


Figure 7.1: Discriminant of the LH algorithm for signal and background taken from simulation. An older version of the PDFs obtained using simulated samples is used. Taken from Ref. [200].

values for different E_T ranges will lead to discontinuous efficiencies at the bin edges. To avoid this, the cut value is interpolated along E_T , leading to a smooth behaviour of the signal efficiency with respect to E_T . No interpolation is done along $|\eta|$.

Bin boundaries in $ \eta $										
PDFs	0.0		0.6	0.8	1.15	1.37	1.52	1.81	2.01	2.37 2.47
Discriminant cuts	0.0	0.1	0.6	0.8	1.15	1.37	1.52	1.81	2.01	2.37 2.47

Bin boundaries in E_T [GeV]										
PDFs	4.5	7	10	15	20		30		40	∞
Discriminant cuts	4.5	7	10	15	20	25	30	35	40	45 ∞

Table 7.1: Bin edges in $|\eta|$ and E_T used to obtain the PDFs and to define different thresholds on the discriminant.

The selection based on the discriminant is not the only requirement placed on candidates to pass a certain WP. Additional requirements are placed on the number of hits in different parts of the tracking detectors and on the ambiguity type, a variable assigned during the reconstruction that gives information about whether the same topo-cluster was also reconstructed as a photon. These requirements change for the different WPs. Three WPs are considered in this work: *LHLoose*, *LHMedium*, and *LHTight*. They are all subsets of each other, meaning every candidate passing the *LHTight* WP also passes the *LHMedium* and *LHLoose* WP. To account for the effects of pileup, the final discriminant is further modified for the *LHMedium* and *LHTight* WP such that the background rejection is constant with respect to pileup.

To obtain the best performance in real data, the PDFs are obtained from data directly instead of using simulated samples. For this, two selections need to be applied. One to obtain signal candidates and one to obtain background candidates. During these selections, a LH algorithm obtained from simulated events is used.

SIGNAL SELECTION The signal selection applies a tag-and-probe method to select $Z \rightarrow ee$ or $J/\psi \rightarrow ee$ decays. Candidates obtained from $Z \rightarrow ee$ decays are used for PDFs with $E_T > 15$ GeV while candidates from $J/\psi \rightarrow ee$ decays are used for lower values of E_T ($E_T < 15$ GeV).

The tag-and-probe method is based on the fact that the decay of the Z boson and the J/ψ meson results in two electrons. One electron, the tag, is required to pass tight identification and isolation requirements. If now a second electron, the probe, can be found such that, together with the tag, it gives an invariant mass close to the mass of the Z boson or the J/ψ meson, there is a high probability that the probe is an actual electron. Since no identification or isolation requirements are applied to the probe, this gives an unbiased set of signal electrons. In practice, very loose identification requirements are applied to the probe electron to improve purity without distorting the shape of the PDFs too much. For the $J/\psi \rightarrow ee$ tag-and-probe selection, additional

requirements are applied, targeting the lifetime of the J/ψ meson to veto decays of J/ψ mesons originating from b -hadron decays.

BACKGROUND SELECTION The selection of background candidates is simpler than the selection of signal candidates. As the cross-section of multi-jet events is much higher than that of any process producing signal electrons, just selecting any reconstructed electron will give a large portion of background electrons. To improve purity, the most common decays that result in signal electrons are vetoed. These are $Z \rightarrow ee$ and $W \rightarrow e\nu$ decays.

The veto on $Z \rightarrow ee$ decays works similarly to the tag-and-probe method for the signal selection. If an additional electron that passes some identification requirements is in the event and, together with the electron candidate, gives an invariant mass close to the mass of the Z boson, the event is vetoed. To veto $W \rightarrow e\nu$ decays, additional requirements are placed on the missing transverse momentum and the transverse mass of the W boson defined as

$$m_T = \sqrt{2p_T^\ell E_T^{\text{miss}}(1 - \cos \Delta\phi)}, \quad (7.1)$$

with $\Delta\phi$ being the azimuthal angle between the lepton and the missing transverse momentum. As all electrons from top-quark decays originate from W bosons, this will also remove contributions from all decays involving top-quarks.

8 Identification of electrons using a deep neural network

The LH algorithm described in the previous chapter provides excellent rejection of backgrounds while selecting signal electrons with high efficiency. However, it does not take any correlations between the input variables into account. This chapter introduces an algorithm based on a DNN with similar input variables to, among other things, exploit the information contained in these correlations. The training setup and the performance on simulated events will be shown. An overview of this algorithm can be seen in Ref. [2].

Contents

8.1	Samples and selection	75
8.2	Classes	75
8.3	Inputs	77
8.3.1	Correction factors	79
8.3.2	Preprocessing	79
8.4	Training procedure	84
8.4.1	Multinomial classification	84
8.5	Performance	85
8.5.1	Identification performance	87
8.5.2	Changing fractions in the discriminant	92
8.5.3	Difference in selected electrons	93

8.1 Samples and selection

To obtain the training, validation, and test datasets, several simulated samples are used. The $Z \rightarrow ee$ sample generated with the POWHEG BOX + PYTHIA setup is used as one of the main samples. The other main sample is the JF17 sample. To increase statistics, additional samples containing $J/\psi \rightarrow ee$, $Z\gamma$, and $t\bar{t}$ events are used. For $t\bar{t}$ events, the POWHEG BOX + PYTHIA sample is chosen. Only events with at least one lepton in the final state are considered for this sample.

All reconstructed electrons with a transverse energy of at least 4 GeV that have at least one hit in the Pixel detector and at least seven hits in the silicon detector systems are selected. Not all types of electrons are taken from each sample; instead, there are differences based on the origin of the candidate, which splits the candidates into a total of six classes.

8.2 Classes

The signature of prompt electrons can be mimicked by different types of particles. As the similarity to prompt electrons is inherently different between these types of particles, any identification algorithm will have varying rejection power depending on the type of background. Therefore, the electron candidates are split into six classes.

Class	Description	Label	Sample
Prompt Electrons	Prompt isolated electrons, e.g. electrons from $Z \rightarrow ee$, $W \rightarrow ev$, $J/\psi \rightarrow ee$ decays with the J/ψ being produced in the hard scatter. Electrons from a final state radiation photon or bremsstrahlung are also considered here if the origin is a prompt electron. Furthermore, the reconstructed charge is the same as the true charge.	El	$Z \rightarrow ee$, $J/\psi \rightarrow ee$
Charge Flips	Same as prompt electrons, but the reconstructed charge is the opposite of the true charge. In case of an electron originating from bremsstrahlung, the charge of the original prompt electron is considered the true charge.	CF	$Z \rightarrow ee$, $J/\psi \rightarrow ee$
Photon Conversion	Electrons from prompt photons which convert into an e^+e^- pair. Prompt photons which are reconstructed as an electron are also considered for this class.	PC	JF17, $t\bar{t}$, $Z\gamma$
Heavy-Flavour	Electrons coming from a decay of a b - or c -hadron. Prompt quarkonium decays such as $J/\psi \rightarrow ee$ where the J/ψ is produced in the hard scatter event are not included here but rather as prompt electrons.	HF	JF17, $t\bar{t}$
Light-Flavour e/γ	Electrons and photons from a decay of a u -, d -, or s -hadron. This also includes for instance electrons from intermediate photon conversions: $\pi^0 \rightarrow \gamma\gamma$ with $\gamma \rightarrow ee$.	LFEG	JF17
Light-Flavour Hadrons	Undecayed hadrons	LFH	JF17

Table 8.1: Different classes used for electron candidates throughout the thesis based on truth information of the simulation. A short description, the label used in figures or equations, and the samples used to obtain the respective class are given as well. Adapted from Ref. [2]

A list of all classes as used in this thesis can be seen in Table 8.1. It additionally lists their definition as well as the samples from which the candidates of the respective class are obtained. The two signal-like classes, prompt electrons and charge flip electrons, are obtained from the $Z \rightarrow ee$ and the $J/\psi \rightarrow ee$ samples, with the $Z \rightarrow ee$ sample providing most of the statistics. The $J/\psi \rightarrow ee$ sample increases the statistics mostly at low E_T . Candidates of the other four classes are taken primarily from the JF17 sample. Additional electrons from heavy-flavour decays are taken from the $t\bar{t}$ sample, while additional electrons from photon conversions are taken from the $t\bar{t}$ and the $Z\gamma$ sample. This is done to increase the overall statistics of these classes.

Prompt electrons are always considered as signal, while electrons from photon conversions, electrons from heavy-flavour decays, electrons or photons from light-flavour hadron decays, and hadronic background are always considered as background. The assignment of charge flip electrons is not as straightforward. They are prompt electrons, but the charge was not reconstructed correctly. Therefore, they can be considered signal or background, depending on the physics

analysis. If there is only one lepton in the final state, the charge of the electron does not matter to the measurement, and the inclusion of charge flip electrons into the signal will increase the overall statistics. Even for measurements that depend on the reconstructed charge of the lepton, like a $t\bar{t}$ charge asymmetry measurement [202], charge flip electrons are not specifically considered as the rate of misidentification of the charge is small enough to not affect the measurement. However, if there are two leptons with the same charge in the final state, charge flip electrons are generally considered background to suppress SM processes, which typically have a much higher cross-section than the targeted final state. For instance, in the observation of four top-quark production in the channel selecting two leptons with the same charge [172], top-quark pair events with a charge flip electron are considered as background.

As the rejection power varies vastly between the different background classes, ROC curves will be shown using only one background class at a time as well as all background classes at the same time. Due to their special status, charge flip electrons are not included in the signal or background definition. Instead, an additional ROC curve using prompt electrons as signal and charge flip electrons as background will be shown.

8.3 Inputs

Information from all subsystems of the detector apart from the MS is exploited to perform electron identification. In addition to information from the tracking system and the calorimeter system, the matching of the track and the cluster is also used. In Table 8.2, all variables used for the DNN are shown. The shown variables, however, do not all enter the DNN as input. Several are only used as an additional selection criterion when defining WPs for potential usage in a physics analysis. These variables are labelled with a "C" in the "Usage" column of Table 8.2. The cuts include requirements on the number of hits used to reconstruct the track in different subsystems as well as the ambiguity type. These requirements are chosen to be the same as for the LH WPs.

The choice of variables differs slightly from the ones used by the LH algorithm [200]. Firstly, E_T and $|\eta|$ are used as inputs to the DNN to give the network context. The LH does not use these kinematic variables directly. Instead, separate PDFs are obtained in bins of E_T and $|\eta|$. Therefore, the kinematic information is used, but in an indirect way. This results in several LH algorithms, which might not be an issue for a simple algorithm like the LH but for something like a DNN this would result in a considerable overhead.

Secondly, only non-correlated variables can be used for the LH algorithm due to its definition. The DNN on the other hand, can handle correlated variables fine. Therefore, R_{had} and R_{had1} can both be used as input over the whole range of the detector, whereas the LH algorithm uses only one depending on the geometry of the detector.

Table 8.2: Description of the quantities used in the electron identification based on the DNN. In the column labelled “Usage,” a “DNN” indicates that the quantity is used as an input to the DNN and a “C” indicates that this quantity is used as an additional selection criterion. If the quantity is used as an input to the DNN and as an additional selection criterion, it is denoted by “DNN/C”. In the description of the quantities formed using the second layer of the calorimeter, 3×3 , 3×5 , 3×7 , and 7×7 refer to areas of $\Delta\eta \times \Delta\phi$ space in units of 0.025×0.0245 . Adapted from Ref. [2, 200].

Description	Name	Usage
Ratio of E_T in the first layer of the HCAL to E_T of the EM cluster.	R_{had1}	DNN
Ratio of E_T in the HCAL to E_T of the EM cluster.	R_{had}	DNN
Ratio of the energy in the third layer to the total energy in the ECAL. Due to known mismodelling at high $ \eta $, this variable is set to a default value for candidates with $ \eta > 2.01$.	f_3	DNN
Lateral shower width in the second layer of the ECAL $\sqrt{(\sum E_i \eta_i^2)/(\sum E_i) - ((\sum E_i \eta_i)/(\sum E_i))^2}$, where E_i is the energy and η_i is the pseudorapidity of cell i and the sum is calculated within a window of 3×5 cells.	$w_{\eta 2}$	DNN
Ratio of the energy in 3×3 cells in the second layer of the ECAL over the energy in 3×7 cells centred at the electron cluster position.	R_ϕ	DNN
Ratio of the energy in 3×7 cells in the second layer of the ECAL over the energy in 7×7 cells centred at the electron cluster position.	R_η	DNN
Shower width in the first layer of the ECAL, $\sqrt{(\sum E_i (i - i_{\text{max}})^2)/(\sum E_i)}$, where i runs over all strips in a window of $\Delta\eta \times \Delta\phi \approx 0.0625 \times 0.2$, corresponding typically to 20 strips in η , and i_{max} is the index of the highest-energy strip.	w_{stot}	DNN
Ratio of the energy difference between the maximum energy deposit and the energy deposit in a secondary maximum in the cluster to the sum of these energies in the first layer of the ECAL.	E_{ratio}	DNN
Ratio of the energy in the first layer to the total energy in the ECAL.	f_1	DNN
Number of hits in the innermost Pixel layer. This variable is only used as a boolean whether there was at least one hit in the innermost Pixel layer or not.	$n_{\text{Innermost}}$	C
Number of hits in the Pixel detector.	n_{Pixel}	DNN/C
Number of hits in the SCT detectors.	n_{SCT}	DNN
Total number of hits in the Pixel and SCT detectors.	n_{Si}	C
Transverse impact parameter relative to the beam-line.	d_0	DNN
Significance of transverse impact parameter defined as the ratio of d_0 to its uncertainty.	$ d_0/\sigma(d_0) $	DNN
Momentum lost by the track between the perigee and the last measurement point divided by the momentum at perigee.	$\Delta p/p$	DNN
Likelihood probability based on transition radiation in the TRT. This variable is set to a default value for candidates with $ \eta > 2.01$ due to the limited coverage of the TRT.	TRTPID	DNN
$\Delta\eta$ between the cluster position in the first layer and the extrapolated track.	$\Delta\eta_1$	DNN
$\Delta\phi$ between the cluster position in the second layer of the ECAL and the momentum-rescaled track, extrapolated from the perigee, times the charge q .	$\Delta\phi_{\text{res}}$	DNN
Ratio of the cluster energy to the track momentum.	E/p	DNN
Transverse energy of the electron measured by the calorimeter system. This variable is not used for discrimination purposes but to give the DNN additional information.	E_T	DNN
Absolute value of the pseudorapidity of the electron as measured by the calorimeter system. This variable is not used for discrimination purposes but to give the DNN additional information.	$ \eta $	DNN
Output of an ambiguity resolution algorithm to distinguish objects that are reconstructed as both electrons and photons [73].	Amb type	C

Lastly, the number of hits of the primary electron track is not used in the LH algorithm but just as additional rectangular cuts, whereas it enters the DNN as input.

8.3.1 Correction factors

To obtain the best performance in data, variables showing large data-MC disagreement are corrected in the simulation to better match data. Ref. [200] details these corrections. Each variable is corrected separately by applying an affine transformation dependent on E_T and $|\eta|$. Electron candidates are collected with a $Z \rightarrow ee$ tag-and-probe selection in both data and simulation. By minimising the χ^2 between the simulation and data, the parameters of the affine transformations are obtained. Figure 8.1 shows an example of these corrections for two variables: f_3 and R_{had} .

Especially for f_3 in Figure 8.1a, a clear linear shift of the simulation compared to the data can be seen, which is corrected after applying the affine transformation. Even with these corrections, f_3 still experiences large differences between data and simulation at large values of $|\eta|$. For this reason, the variable is effectively removed from the inputs of the DNN for candidates with $|\eta| > 2.01$ by setting it to a constant value.

8.3.2 Preprocessing

DOWNSAMPLING AND REWEIGHTING Due to the use of physics processes to obtain electron candidates, the distribution of E_T will vary considerably between the different classes. For instance, as most prompt electrons are obtained from a $Z \rightarrow ee$ sample, a majority of electrons will have a transverse energy around 45 GeV corresponding to half the Z boson mass. This can be seen in Figure 8.2a. Additionally, it shows the E_T distribution for all other classes. Using this distribution in the training while feeding E_T to the network will result in a large bias in the output score. In fact, E_T and $|\eta|$ should not be used as discriminatory variables but instead as context variables, as the response of the detector to a prompt electron will change with E_T and $|\eta|$. Therefore, the input distributions of E_T and $|\eta|$ should be the same for all six classes, such that the network cannot distinguish between the classes based solely on E_T and $|\eta|$. To match the distributions, a combination of downsampling and reweighting is used.

To reduce the range of weights needed to match the distributions of the single classes, first the prompt electron class is downsampled. Prompt electrons are randomly removed in bins of E_T such that the number of prompt electrons in this bin does not exceed five times the number of background candidates. Applying this downsampling reduces the peak around 45 GeV as can be seen in Figure 8.2b. As the overall statistics is much smaller for charge flip electrons, no downsampling is applied to this class, even though a similar but not

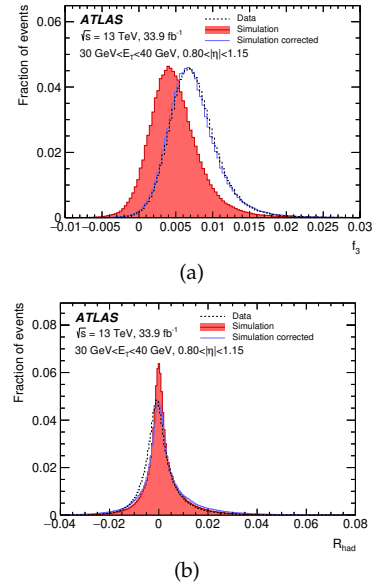


Figure 8.1: Distributions of (a) f_3 and (b) R_{had} of prompt electrons in data, simulation, and corrected simulation. Only electrons fulfilling $30 < E_T [\text{GeV}] < 40$ and $0.80 < |\eta| < 1.15$ are considered. Taken from Ref. [200].

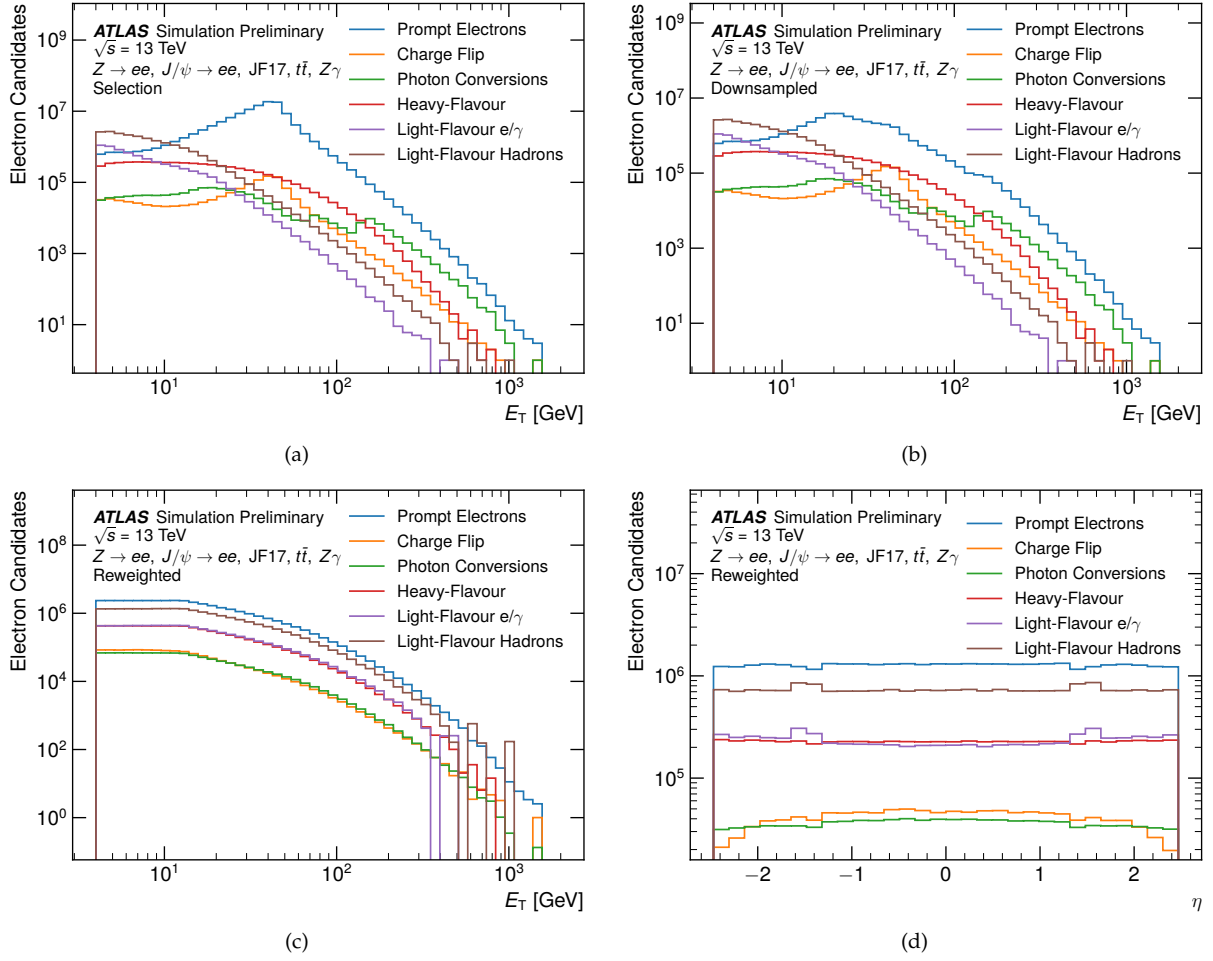


Figure 8.2: E_T distribution of the different classes in the training dataset after (a) the selection, (b) the downsampling, and (c) the kinematic reweighting. (d) shows the η distribution after the kinematic reweighting. The weights making all classes equiprobable are not applied. Taken from Ref. [2].

as pronounced peak at 45 GeV can be seen.

Using the downsampled distributions, weights are obtained to reweight E_T and $|\eta|$ separately for each class. To obtain a final weight to be applied to each candidate, the two weights are multiplied together. As long as the two quantities are uncorrelated, this will result in matching distributions in both quantities. The weights are normalised such that the sum of weights of the single classes remains the same after the application of the weights. The resulting distributions can be seen in Figures 8.2c and 8.2d for E_T and $|\eta|$, respectively.

The reweighting in $|\eta|$ is performed to obtain a flat distribution, while for E_T a flat distribution until 15 GeV with a falling distribution at higher values of E_T following the light-flavour hadron class of the $t\bar{t}$ sample is the target. This shape in E_T has been found to reduce the bias considerably while maintaining good performance. As E_T and $|\eta|$ have slight correlations, some differences can be observed in the reweighted distributions, especially for the charge flip class. A two-dimensional reweighting approach might be considered to mitigate these effects. This would, however, suffer from low statistics, especially at high values of E_T .

To reduce any bias from differences in the available statistics for

the single classes, an additional weight is applied. This weight is the same for all candidates in one class and is multiplied by the weight obtained from the reweighting procedure. With this weight applied, all classes have the same overall sum of weights, but the sum of weights of all classes combined remains the same. Figures 8.2c and 8.2d do not include these weights for visualisation purposes.

Figure 8.3 shows the distribution of all input variables for all six classes after the downsampling and reweighting, including the class weights, were applied.

TRANSFORMATION As discussed in Section 6.6, feeding the network the variables as shown in Figure 8.3 would lead to a difficult training process and potentially bad performance. Instead, all input variables are transformed using a QuantileTransformer. An addition to the Scikit-learn implementation is the inclusion of the weights described in the previous paragraph when calculating the quantiles.

Figure 8.4 shows all input variables after this transformation is applied. All variables are now bound between zero and one and follow a rectangular distribution for most variables. If multiple candidates have the exact same value before the transformation, it can happen that multiple quantiles are mapped to the same value, leading to spikes in the distribution after the transformation with no other candidates close by. This is most obvious and visible for the number of hits in the Pixel and the SCT detector, but the same can also be observed for other variables such as f_3 , w_{stot} , E_{ratio} , f_1 , and TRTPID.

The transformation is monotonic, such that the overall relations between the single classes are not changed. For instance, the R_{had} variable has on average lower values for prompt electrons than for the hadronic background before the transformation. This relationship is still intact after the transformation is applied.

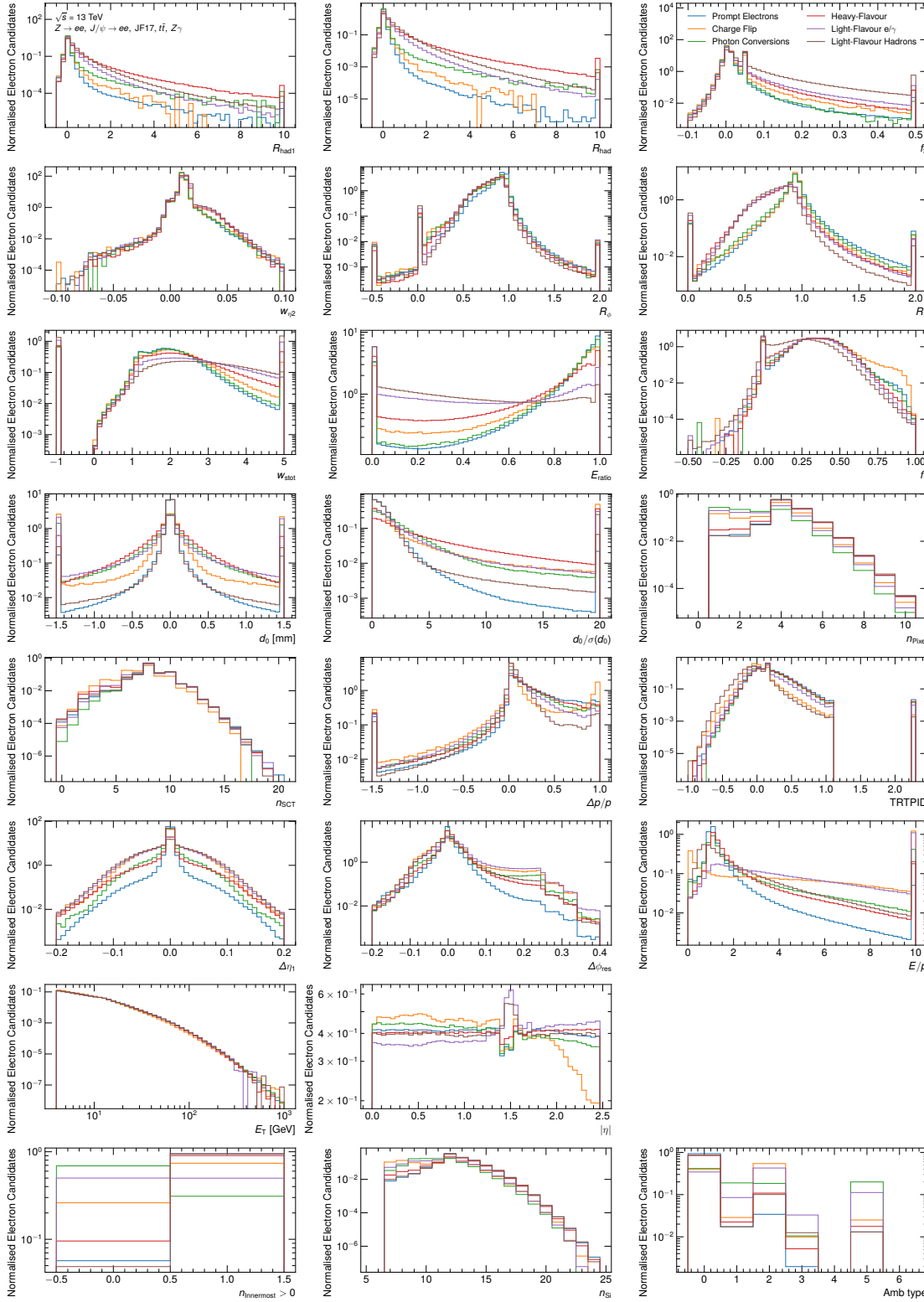


Figure 8.3: Distributions of all input variables after the downsampling and reweighting is applied. The last row shows the variables which are only used as additional selection criteria. The first and last bin contain under- and overflow candidates, respectively.

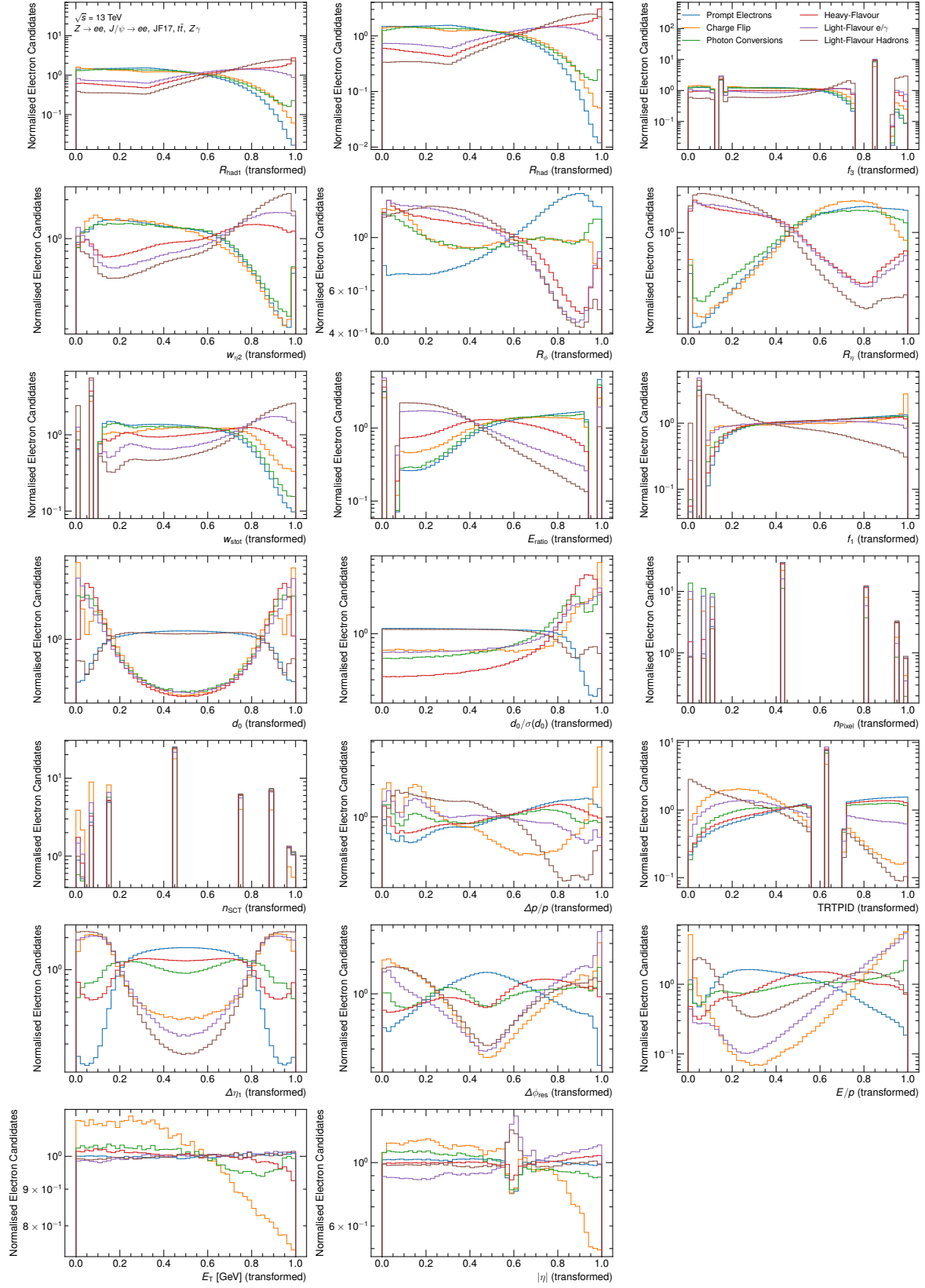


Figure 8.4: Distributions of all input variables after the QuantileTransformer is applied.

8.4 Training procedure

The TensorFlow library version 2.4 is used to train the DNN. The network is built of five hidden layers, each with 256 nodes. The output layer has six output nodes, one for each of the classes. The Leaky ReLU activation function is used after each hidden layer, while the softmax activation is used for the outputs to be able to interpret them as probabilities. A sketch of the architecture can be seen in Figure 8.5.

Figure 8.5: Schematic overview of the architecture of the DNN. The amount of layers and number of nodes shown corresponds to the final optimised architecture. Taken from Ref. [2].

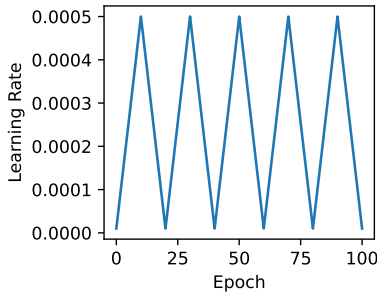
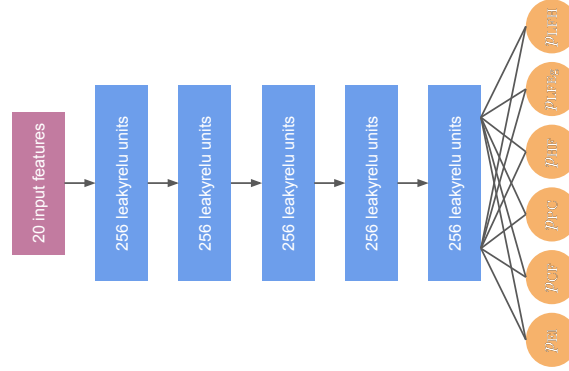


Figure 8.6: Evolution of the LR during the DNN training.

The *Adam* optimiser was used during the training with a cyclical learning rate schedule with a triangular pattern [203]. A sketch of the learning rate throughout the training can be seen in Figure 8.6. Different choices of hyperparameters were tested with this set leading to the best performance on the validation dataset. The final model is trained for 100 epochs, taking the set of weights resulting in the lowest loss on the validation dataset. The interface to the ATLAS analysis framework [204] is performed with the *lwttn* package.

8.4.1 Multinomial classification

As can be seen in Figure 8.4, the distributions of the input variables vary significantly between the different classes, even between only the background classes. One example is the transverse impact parameter d_0 . Electrons from heavy-flavour quark decays have a wider distribution than light-flavour hadrons, caused by the long lifetime of b - and c -hadrons. To exploit these differences, the DNN is trained using multinomial classification.

However, to define WPs, a binary decision is needed. Therefore, a single discriminant is formed out of the six outputs following Equation (6.1). As prompt electrons and charge flip electrons are considered as signal for a generic electron identification, the final discriminant is given by

$$\mathcal{D}_{\text{el}} = \ln \left(\frac{f_{\text{EI}} p_{\text{EI}} + (1 - f_{\text{EI}}) p_{\text{CF}}}{\sum_{\text{PC, HF, LFEg}} f_X p_X + (1 - \sum_{\text{PC, HF, LFEg}} f_X) p_{\text{LFH}}} \right). \quad (8.1)$$

where the f_X are tunable parameters and the p_X are the outputs of the DNN. As prompt electrons and charge flip electrons are considered as

signal, they are both placed in the numerator. To reduce the number of free parameters to optimise after the training, the f_X in the numerator and denominator have to sum up to one.

By changing the f_X , the importance of the single classes to the discriminant can be changed, therefore changing the overall rejection of the algorithm. An example of this change can be seen in Section 8.5.2.

However, training a multinomial classification model does not only have benefits. Training directly in data, as is done for the LH, becomes much more challenging and potentially impossible as pure samples of all six classes have to be obtained with an unbiased data selection. A solution might be to take some classes from data while others are taken from simulated samples. However, this poses a significant risk that the network will primarily learn differences between the classes based on mismodelling of the simulation.

8.5 Performance

To evaluate the performance of the DNN, a test dataset is used. It is statistically independent of the training and validation datasets and was not used during any point of the training procedure. One important difference from the training and validation datasets is that the four background classes are only taken from the JF17 sample. No additional background candidates from the top-quark pair or the $Z\gamma$ sample are used. This is done, as the JF17 sample alone should more closely resemble data. However, the sample is not ideal either. Especially below a transverse energy of 17 GeV, potential differences that might appear due to the filtering of events have not been studied yet. Nevertheless, this sample should still provide a good comparison of the DNN and the LH across the whole E_T spectrum.

Additionally, it should be considered that the LH was optimised using data, whereas the DNN used simulated samples. While the correction factors are applied to the test dataset to minimise the data-MC mismodelling, the differences in performance between the LH and the DNN will change when evaluating on collision data.

The downsampling and reweighting procedure used for the training and validation datasets are not applied to the test dataset. Instead, the performance will be evaluated in bins of E_T and $|\eta|$. The resulting contributions of the single background classes in one such bin and the inclusive dataset can be seen in Table 8.3.

Fraction [%] of	Inclusive	$15 < E_T \text{ [GeV]} \leq 20$ $0.0 < \eta \leq 0.8$
Photon Conversions	0.142 ± 0.001	0.30 ± 0.01
Heavy-Flavour	0.528 ± 0.002	1.24 ± 0.02
Light-Flavour e/γ	24.12 ± 0.01	18.59 ± 0.08
Light-Flavour Hadrons	75.21 ± 0.01	79.87 ± 0.08

Table 8.3: Composition of the background in the dataset used to evaluate the performance. Adapted from Ref. [2].

To obtain the best rejection against the combined background, the free parameters f_X of Equation (8.1) have to be tuned. As the different

classes have different spectra in $|\eta|$ and E_T , different f_X can be chosen in bins of $|\eta|$. The same f_X will be used along E_T to obtain a smooth discriminant and therefore smooth efficiencies along E_T . The f_X are chosen by maximising the area under the ROC curve between 70% and 95% signal efficiency. This efficiency window covers most of the WPs commonly used for electron identification.

To define the DNN WPs, different thresholds are placed on the discriminant based on E_T and $|\eta|$ to obtain the desired efficiencies. The chosen binning in E_T and $|\eta|$ can be seen in Tables 8.4 and 8.5. Additionally, the choice of binning for the optimisation of the f_X is shown as well. A coarser binning is chosen for this optimisation to retain good statistics for all background classes.

		Bin boundaries in $ \eta $										
Discriminant cuts	0.0	0.1	0.6	0.8	1.15	1.37	1.52	1.81	2.01	2.37	2.47	
f_X optimisation	0.0			0.8		1.37	1.52		2.01		2.47	

Table 8.4: Bin edges in $|\eta|$ to define different cut values on the discriminant and to optimise the f_X comprising the discriminant. Adapted from Ref. [2].

Table 8.5: Bin edges in E_T to define different cut values on the discriminant. Adapted from Ref. [2].

Bin boundaries in E_T [GeV]					
4	7	10	15	20	25
30	35	40	45	∞	

The resulting discriminant $\mathcal{D}_{el}$ can be seen in Figure 8.7 for one $(E_T, |\eta|)$ bin. Both signal classes, prompt and charge flip electrons, have higher discriminant values than the remaining classes, with prompt electrons having higher values than charge flip electrons. There is a considerable overlap of electrons from heavy-flavour decays and photon conversions with the two signal classes. Especially the heavy-flavour background has more candidates with very high discriminant values than the charge flip class. Furthermore, it has a wide spread of values. This can potentially be explained by the lifetimes of b - and c -hadrons. If the decay time is short, the resulting response of the detector will be very similar to that of a prompt electron, whereas for longer decay times, the response is more different. Very good separation of prompt electrons from both light-flavour backgrounds can be seen. These two background classes differ the most from prompt electrons. However, as these backgrounds are the dominant ones before any identification requirement, good separation is needed.

In Figure 8.7c, the discriminant is shown for prompt electrons and the combined background. Charge flip electrons are not included in this figure. For the combination of the background, no weights are used, such that the distribution is similar to what one could expect in a physics analysis.

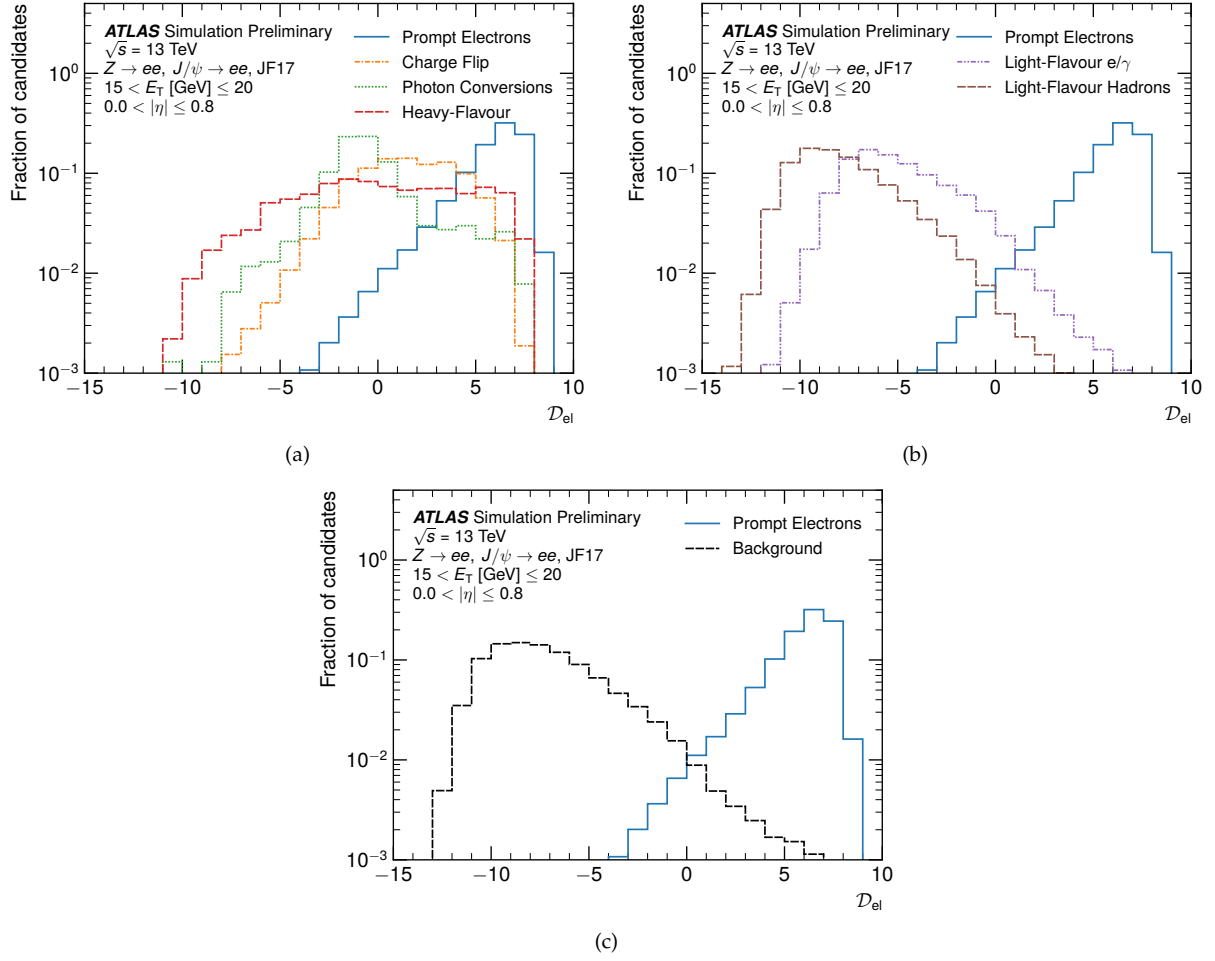


Figure 8.7: Final discriminant $\mathcal{D}_{el}$ of the DNN for (a) prompt electrons, charge flip electrons, electrons from photon conversions, and electrons from heavy-flavour quark decays, (b) prompt electrons, e/γ objects from light-flavour decays, and light-flavour hadrons, (c) prompt electrons and the combined background consisting of all classes except prompt electrons and charge flip electrons. Only electron candidates satisfying $15 < E_T [\text{GeV}] \leq 20$ and $|\eta| \leq 0.8$ are shown. Taken from Ref. [2].

8.5.1 Identification performance

To compare the performance of the DNN algorithm to that of the LH algorithm, ROC curves are used. In Figure 8.8, ROC curves can be seen for one $(E_T, |\eta|)$ bin. It shows the ROC curves against the combined background and all four background classes separately. The signal always only constitutes the prompt electron class. The LH is only represented by single points, corresponding to the single WPs. To have a comparison between the DNN and the LH that is as fair as possible, the same requirements, apart from the discriminant, are used. An overview can be seen in Table 8.6. On top of the listed requirements, all WPs require at least seven hits in the silicon detectors and one hit in the pixel detector. This selection was already applied during the selection of the training, validation, and test datasets. To apply these rectangular cuts for the DNN, the discriminant of candidates not passing these requirements is set to a large negative value. Because of the different additional requirements, three different ROC curves can be seen for the combined background. For the single backgrounds, only the ROC curve with the cuts corresponding to the *Tight* WP are shown, as the uncertainties become larger and the plots become very cluttered otherwise.

Table 8.6: Additional selection on top of the discriminant for the different WPs.

	$n_{\text{Innermost}}$	Amb type
Loose	–	< 5
Medium	> 0	< 5
Tight	> 0	< 2

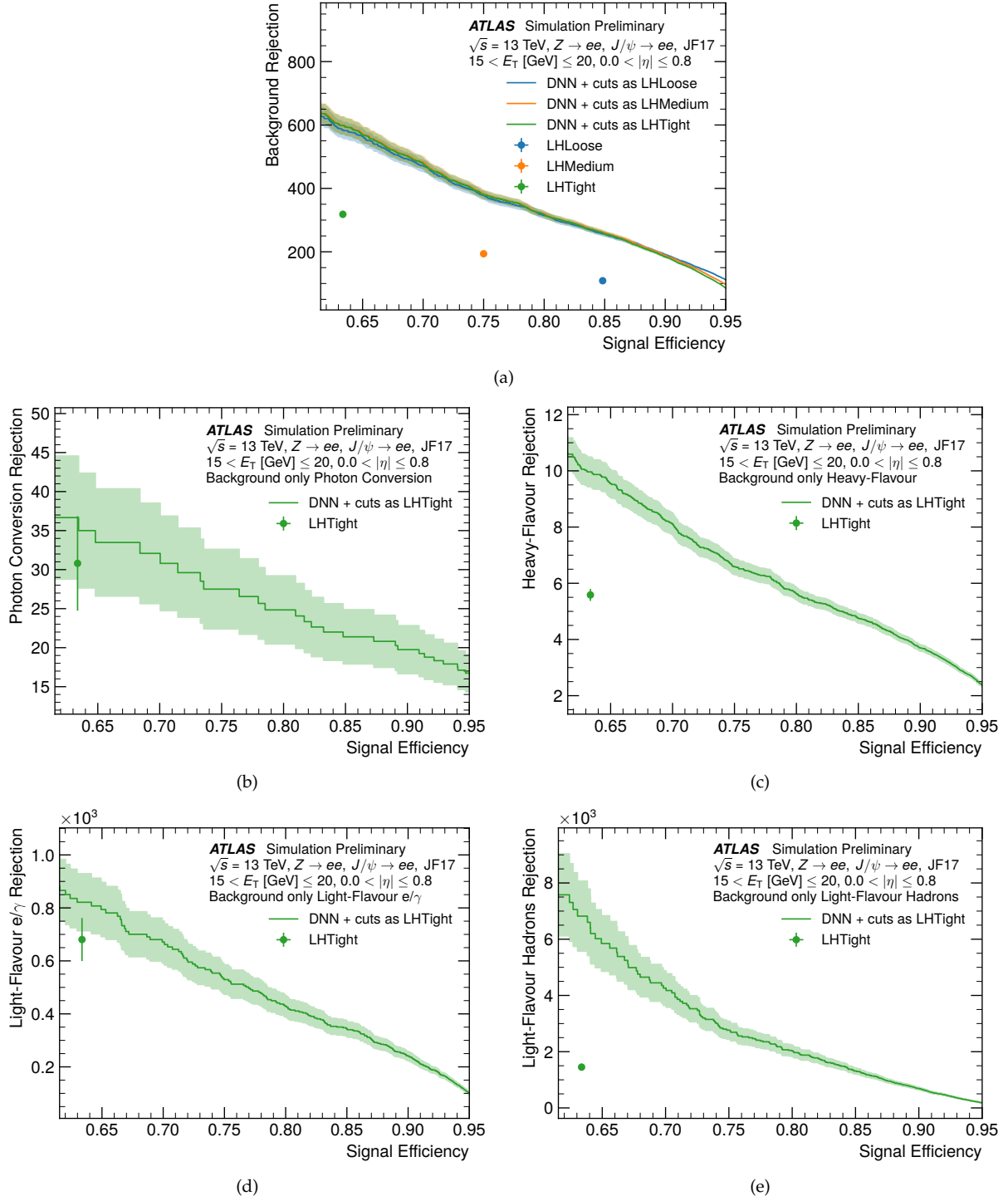


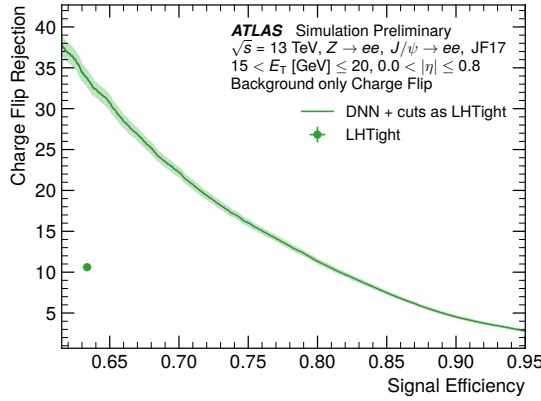
Figure 8.8: ROC curves for (a) the combined background, (b) only electrons from photon conversions, (c) only electrons from heavy-flavour quark decays, (d) only electrons and photons from light-flavour hadrons, and (e) only light-flavour hadrons. Only candidates with $15 < E_T [\text{GeV}] \leq 20$ and $|\eta| \leq 0.8$ are shown. Statistical uncertainties of the background rejection are shown as bands. Taken from Ref. [2].

As can be seen in Figure 8.8a, the DNN has a better overall background rejection than the LH. A larger relative increase in background rejection can be seen at higher signal efficiencies. The different additional selections do not have a significant impact on the rejection power of the combined background. All three ROC curves lie on top of each other. However, for certain background classes, these cuts are very beneficial. Figure 8.9 shows the three LH WPs for only the electron from photon conversions background. Here, a significant

increase in background rejection can be seen between the *LHLoose* and *LHMedium* WP caused by the addition of the requirement of having at least one hit in the innermost tracking layer.

The rejection against the single background classes is shown in Figures 8.8b to 8.8e. The DNN outperforms the LH for every class. But, the improvement depends heavily on the class. For electrons from heavy-flavour decays, the improvement is similar to what can be seen for the combined background. For electrons from photon conversions and electrons or photons from light-flavour decays, the improvement is small, while for light-flavour hadrons, the rejection is increased substantially.

Figure 8.10 shows the rejection of charge flip electrons. As discussed previously, for a generic electron identification, these should be accepted, and therefore a lower background rejection is better. The DNN has a higher rejection compared to the LH. However, as much more prompt electrons than charge flip electrons are expected during any selection, the overall signal efficiency when including charge flip electrons in the signal definition would not be affected.



These ROC curves only show one $(E_T, |\eta|)$ bin, which provides only a limited view of the performance over the whole phase space. Therefore, the efficiencies and rejections of defined WPs are shown with respect to E_T , $|\eta|$, and the average number of interactions per bunch crossing $\langle \mu \rangle$ in Figures 8.11 and 8.12. The DNN is tuned such that, in the finer binning introduced in Tables 8.4 and 8.5, the DNN has the same signal efficiency as the LH, while applying the additional rectangular selection listed in Table 8.6. A single cut value is chosen for each bin for the DNN, instead of doing an interpolation along E_T .

The *Loose* WP is shown in Figure 8.11, while the *Tight* one is shown in Figure 8.12. By design, the signal efficiencies in E_T and $|\eta|$ are the same for the DNN and the LH for both WPs. The signal efficiencies rise with E_T , while for $|\eta|$ they heavily depend on the detector geometry. In the crack region ($1.37 < |\eta| < 1.52$) and very high $|\eta|$ values ($2.37 < |\eta| < 2.47$), the efficiencies tend to be lower than in the remaining detector. The largest improvements in background rejection can be seen for the *Loose* WP in the endcap region ($|\eta| > 1.52$)¹ and lower values of E_T . Improvements up to a factor of 5.5 can be

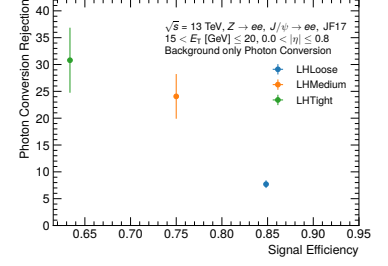


Figure 8.9: Prompt electron efficiency and photon conversion rejection for three LH WPs. Only candidates with $15 < E_T$ [GeV] ≤ 20 and $|\eta| \leq 0.8$ are shown

Figure 8.10: ROC curves with charge flip electrons as background. Only candidates with $15 < E_T$ [GeV] ≤ 20 and $|\eta| \leq 0.8$ are shown. Statistical uncertainties of the background rejection are shown as bands. Taken from Ref. [2].

¹ Large improvements can also be seen in the crack region. However, this region is commonly vetoed in physics analyses due to poorer performance in this region. Therefore, the main focus is on the remaining detector.

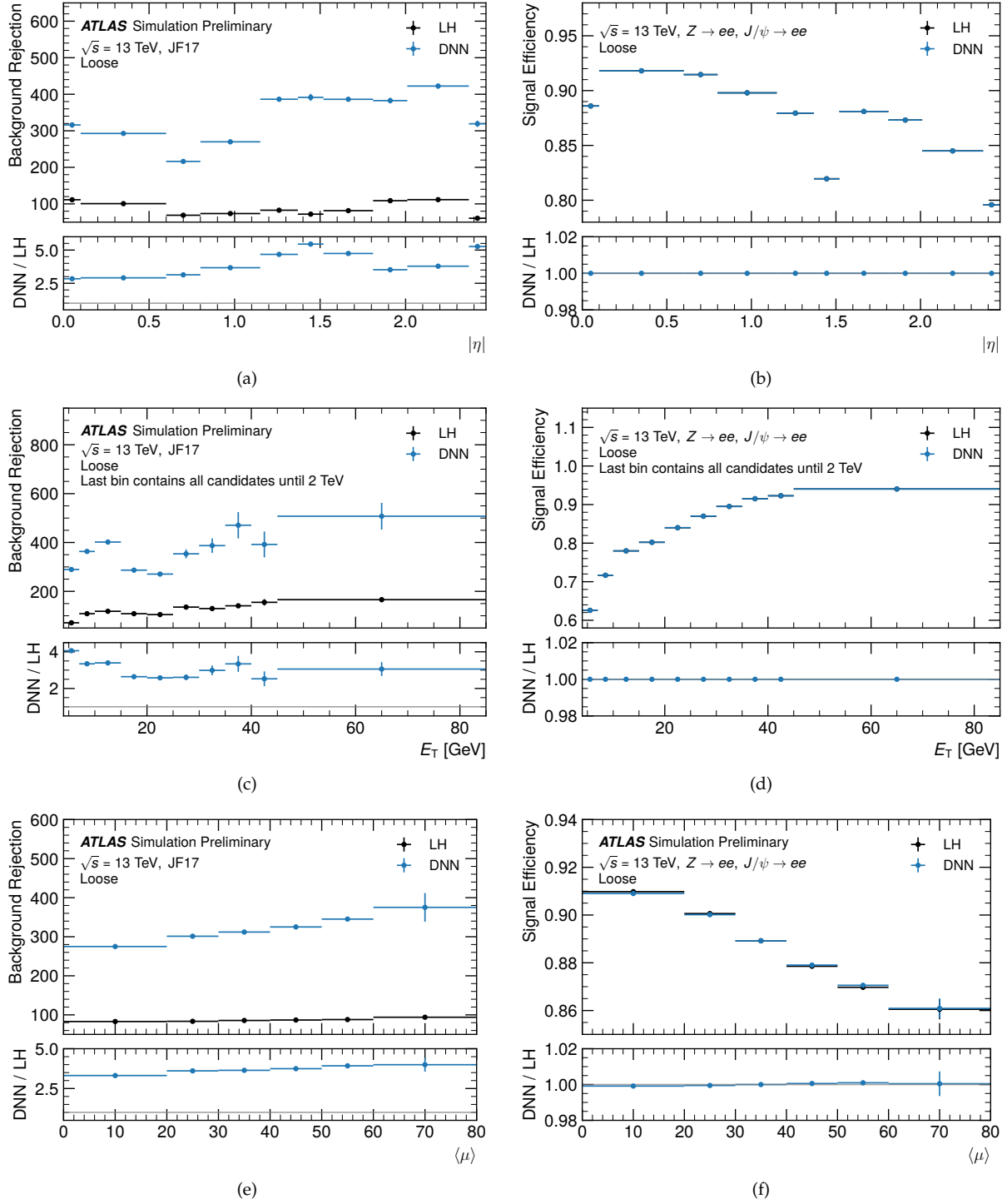


Figure 8.11: Performance plots matching the DNN signal efficiency to that of the Loose LH WP. Plots (a), (c), and (e) show the background rejection with respect to E_T , $|\eta|$, and $\langle\mu\rangle$. Plots (b), (d), and (f) show the signal efficiency with respect to the same quantities. The ratio is shown with respect to the LH WP. The signal efficiencies with respect to E_T and $|\eta|$ are the same by design. Partially taken from Ref. [2].

observed. Overall, the improvement never decreases below 2.5 for the *Loose* WP. Using the *Tight* WP, the conclusions are similar. While the improvement is generally smaller compared to the *Loose* WP, an improvement can still be seen over the whole phase space, ranging between values of 1.7 and 2.5.

Some differences between the *Loose* and *Tight* WPs can be seen in the behaviour with respect to pileup. For the *Loose* WP, the signal efficiencies of the DNN and the LH drop in the same way, whereas for

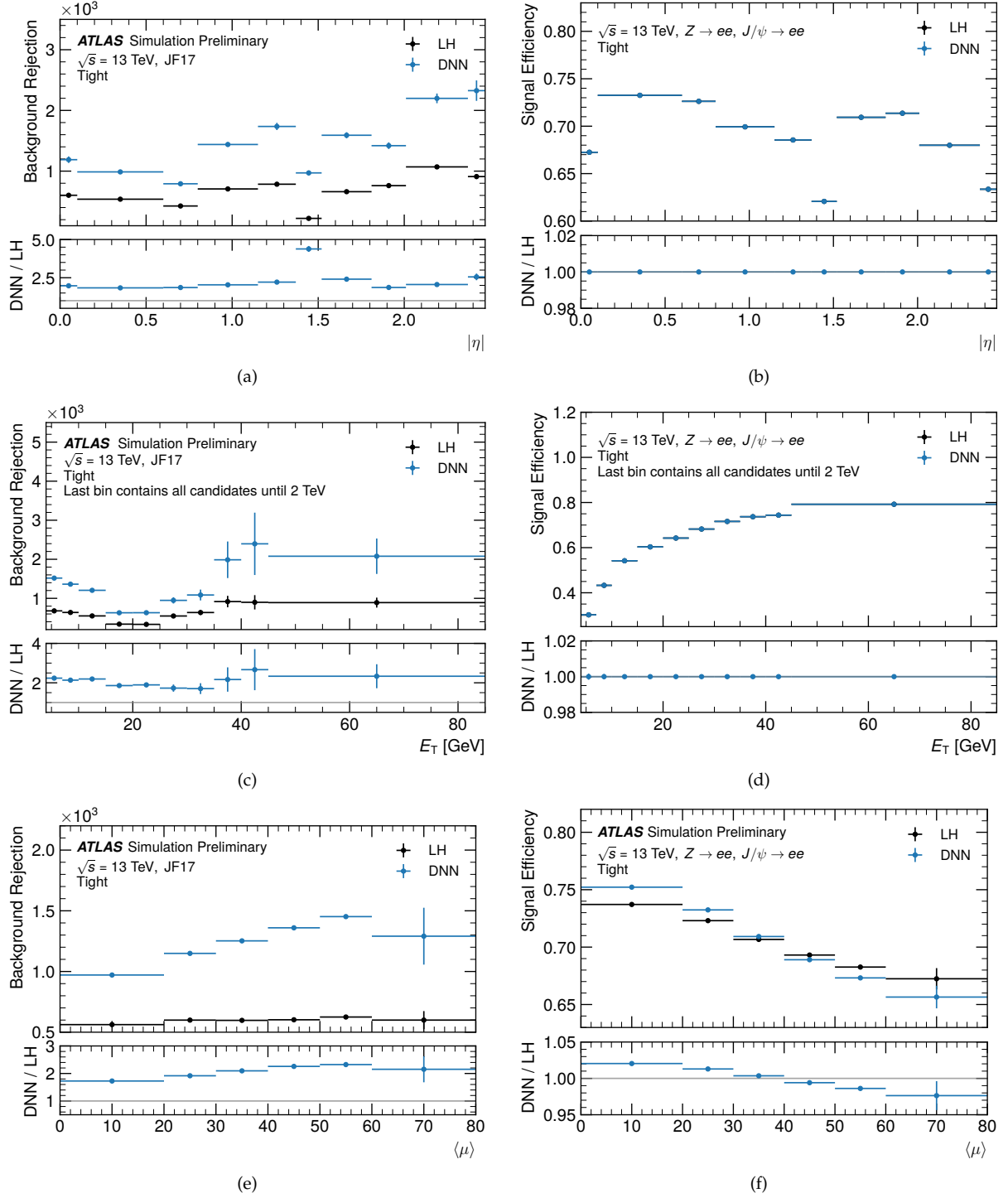


Figure 8.12: Performance plots matching the DNN signal efficiency to that of the Tight LH WP. Plots (a), (c), and (e) show the background rejection with respect to E_T , $|\eta|$, and $\langle\mu\rangle$. Plots (b), (d), and (f) show the signal efficiency with respect to the same quantities. The ratio is shown with respect to the LH WP. The signal efficiencies with respect to E_T and $|\eta|$ are the same by design. Partially taken from Ref. [2]

the *Tight* WP, the DNN efficiencies drop more than the LH ones. For both WPs, the background rejection of the DNN is larger compared to the LH at high values of $\langle\mu\rangle$. Due to the changes in signal efficiency and background rejection at the same time, it is difficult to conclude which algorithm has more resilience to pileup. Important to note is that the discriminant is modified for the *LHTight* WP depending on the number of reconstructed primary vertices, which is another proxy for pileup [200]. No such modification is applied to the DNN, in fact,

the DNN has no direct notion of pileup whatsoever. This could be the reason for the different behaviour of the signal efficiencies between the *Loose* and *Tight* WPs.

8.5.2 Changing fractions in the discriminant

As the DNN algorithm performs multinomial classification, the discriminant defined in Equation (8.1) is not the only possible arrangement. A multitude of different discriminants are possible, adding a lot of flexibility to the algorithm. It should be noted that this is not only true for a DNN based algorithm. The LH algorithm could easily be extended to a multinomial classification, making it benefit from the same advantages.

Two additional discriminants will be explored in this section. The first one only changes the values of the f_X , thus resulting in a discriminant that pays more attention to a specific background class. Here, the value of f_{HF} is multiplied by a factor of 20. To keep the sum of all f_X in the denominator at one, the remaining f_X are divided by 1.22. The second additional discriminant changes the definition to

$$\mathcal{D}'_{el} = \ln \left(\frac{p_{El}}{\sum_{CF,PC,HE,LFEg} f'_X p_X + \left(1 - \sum_{CF,PC,HE,LFEg} f'_X\right) p_{LFH}} \right).$$

Here, the output probability of being a charge flip electron is placed in the denominator instead of the numerator. Therefore, this discriminant will reject charge flip electrons instead of accepting them.

Figure 8.13 shows ROC curves for these additional two discriminants as well as for the standard discriminant. The performance on the combined background does not change between the standard and the charge flip rejecting discriminant. In fact, the rejection of all four standard background classes is very similar for these two discriminants. However, the rejection of charge flip electrons is increased significantly when placing the charge flip probability in the denominator.

Switching from the standard discriminant to the heavy-flavour rejecting discriminant, the rejection of the combined background drops slightly. This is caused by a significant drop in the rejection of light-flavour hadrons. At 65% signal efficiency, the rejection is around three times lower for the heavy-flavour rejecting discriminant. However, at the same time, the rejection of electrons from heavy-flavour decays increases significantly. At 65% signal efficiency, an increase of around 25% in rejection can be seen. Additionally, the heavy-flavour rejecting discriminant has a lower rejection of charge-flip electrons than the standard discriminant. This might be explained by the reduced value of f_{PC} . Electrons from photon conversions can be very similar to charge flip electrons, so a reduction in this parameter might result in a reduction in the rejection of charge flip electrons.

Overall, the added flexibility is clearly visible. Changing the values of the parameters f_X changes the behaviour of the discriminant signif-

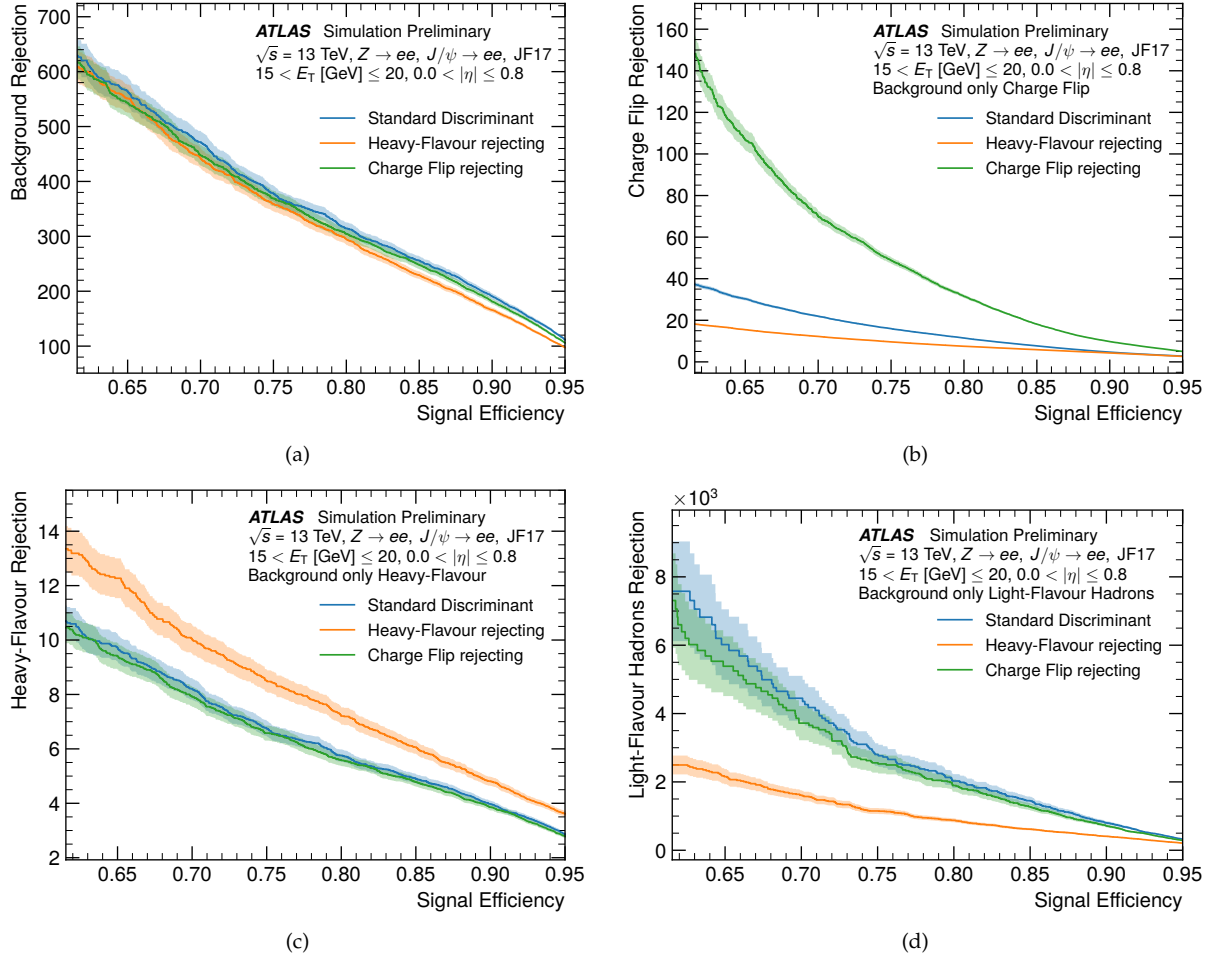


Figure 8.13: ROC curves for (a) the combined background, (b) only charge flip electrons, (c) only electrons from heavy-flavour quark decays, and (d) only light-flavour hadrons comparing different versions of the discriminant $\mathcal{D}_{el}$. Only candidates with $15 < E_T [\text{GeV}] \leq 20$ and $|\eta| \leq 0.8$ are shown. Statistical uncertainties of the background rejection are shown as bands. Taken from Ref. [2].

icantly. While on this dataset, the combined rejection drops, analyses that suffer mostly from electrons from heavy-flavour decays might benefit from such a modified discriminant. Additionally, placing the charge flip output probability in the denominator can significantly improve the charge flip rejection while maintaining very similar rejection levels for the remaining backgrounds.

However, it should be considered that a change in the discriminant will lead to different electrons being selected, which can cause issues as detailed in the following section.

8.5.3 Difference in selected electrons

As electrons have a clean signature, they are often used for trigger purposes. In fact, all studies in this work utilise at least one trigger requiring an electron. However, triggering on all reconstructed electrons would lead to large rates, mostly due to hadronic background. Therefore, an identification algorithm is already applied at the trigger level. For Run 2, this algorithm follows the same LH approach detailed in Chapter 7. Different PDFs are used due to differences in the input variables at trigger level and after the full reconstruction, also referred to as online and offline, respectively. The use of the same

algorithm still provides good overlap between the selected electrons. Most electrons that pass the online *LHTight* WP also pass the offline *LHTight* WP. However, some electrons that pass the offline *LHTight* WP do not pass the online *LHTight* WP. One cause of these inefficiencies is differences in the resolution of variables in the online and offline reconstruction. Switching to a DNN based identification at the offline level might introduce additional inefficiencies. To quantify these inefficiencies, the selected electrons of the DNN are compared with the selected ones of the offline LH which is used as a proxy for the online LH.

Figures 8.14 and 8.15 show a comparison for prompt electrons and the combined background, respectively. The *Loose* and *Tight* WPs are shown. The majority of prompt electrons behave the same for the DNN and the LH. However, $(5.72 \pm 0.02)\%$ ($(9.67 \pm 0.03)\%$) are accepted by the DNN but not the LH for the *Loose* (*Tight*) WP. As the two algorithms are tuned to the same signal efficiency, the same amount is accepted by the LH but rejected by the DNN.

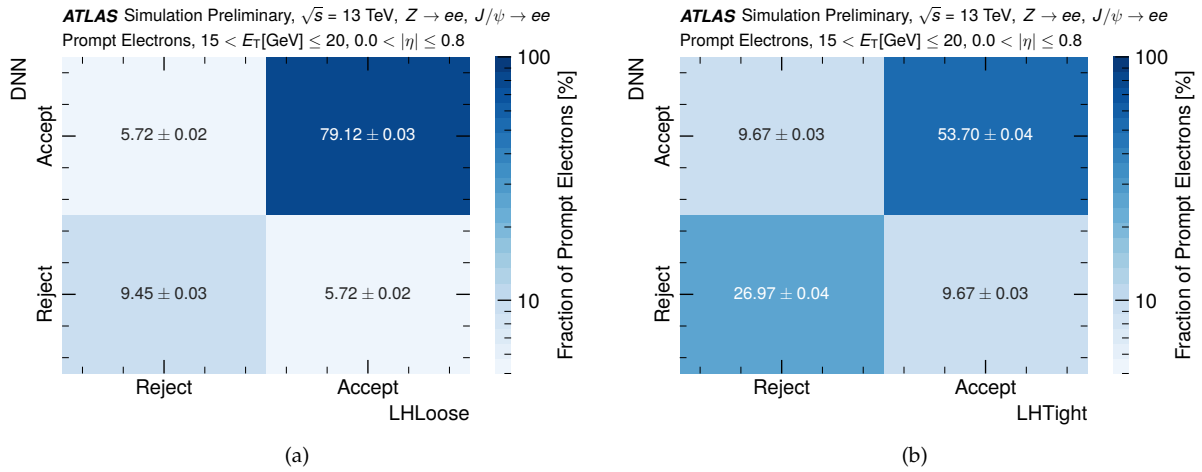


Figure 8.14: Comparison of the selection of prompt electrons based on the DNN and the LH. (a) shows a *Loose* WP, while (b) shows a *Tight* WP. Only candidates fulfilling $15 < E_T [\text{GeV}] \leq 20$ and $|\eta| \leq 0.8$ are shown. Taken from Ref. [2].

Over 99% of the background is rejected by both algorithms for the *Loose* and the *Tight* WPs. Additionally, most of the background that is accepted by the DNN is also accepted by the LH. The DNN only accepts $(0.069 \pm 0.005)\%$ ($(0.042 \pm 0.004)\%$) of the background that is rejected by the LH for the *Loose* (*Tight*) WP. The other way around, so background being accepted by the LH but not the DNN, the value increases to $(0.60 \pm 0.02)\%$ ($(0.19 \pm 0.01)\%$). This shows again the improved rejection power of the DNN compared to the LH.

A considerable difference in the selected electrons can be seen. Therefore, the DNN is not suited for analyses that require low inefficiencies at the trigger level. For other analyses, the improvement in rejection power might be reduced due to these differences. To combat this and increase rejection at the trigger level, a DNN based identification is being developed for the trigger level. However, so far, it has not been deployed into production.

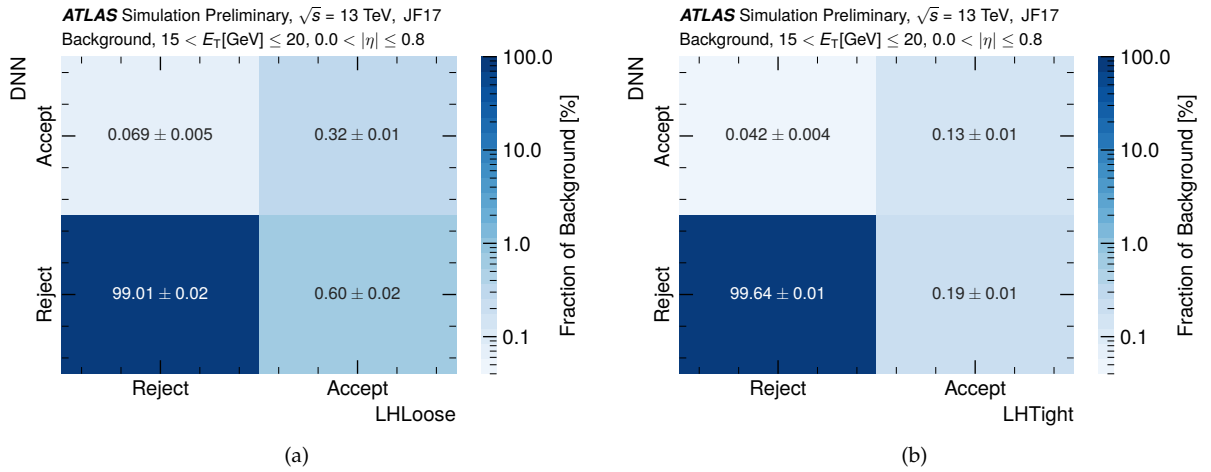


Figure 8.15: Comparison of the selection of the inclusive background based on the DNN and the LH. (a) shows a *Loose* WP, while (b) shows a *Tight* WP. Only candidates fulfilling $15 < E_T [\text{GeV}] \leq 20$ and $|\eta| \leq 0.8$ are shown. Taken from Ref. [2].

9 Identification of electrons with low- and high-level information

As could be seen in the previous chapter, a simple DNN using high-level features can improve the performance of electron identification over the currently used LH algorithm. However, there also exists the option to use low-level data, like the cells in the calorimeter or the multiple candidate tracks assigned to a reconstructed electron. This chapter will cover different ways of including this information in a ML algorithm. For the cells, a representation as images or as point clouds is explored, while for the track information, only a point cloud is considered. The performance in simulated events will be shown as well.

Contents

9.1	Samples and classes	97
9.2	Input variables	97
9.2.1	High-level features	97
9.2.2	Cells	98
9.2.3	Tracks	98
9.2.4	Further preprocessing	99
9.3	Models	99
9.3.1	Comparison of operations on cells	99
9.3.2	Comparison of inputs	101
9.4	Performance	102
9.4.1	Comparison of methods	103
9.4.2	Comparison of inputs	106

9.1 Samples and classes

The selection of generated samples is very similar to the previous chapter. The $Z \rightarrow ee$ sample generated with POWHEG + PYTHIA is used to obtain signal electrons and charge flip electrons. The JF17 sample is used to obtain candidates from all four background classes, while the POWHEG BOX + PYTHIA sample for top-quark pair production is used to increase the available statistics for electrons from photon conversions and electrons from heavy-flavour decays. No $J/\psi \rightarrow ee$ or $Z\gamma$ samples are used as they were not available with the calorimeter cell information stored. The same classes as in the previous chapter are used.

9.2 Input variables

9.2.1 High-level features

The same high-level features as used for the DNN described in Table 8.2 are used for all models trained in this chapter, though the correction factors are not applied. This choice was made because no equivalent corrections exist for the cells of the calorimeter. The same QuantileTransformer as in Chapter 8 is applied to the input features.

9.2.2 Cells

Each topo-cluster consists of multiple cells in the calorimeter. For electrons, only cells in the ECAL are considered for the building of the topo-cluster. All cells typically fall into a 7×11 window in $\Delta\eta \times \Delta\phi$ around the cluster centre, using cells in layer two of size 0.025×0.0245 as the unit. As a large fraction of the background is hadronic background, knowledge about the leakage into the HCAL is beneficial for the algorithm. Therefore, all cells within the same window in the HCAL are considered for the algorithms as well.

IMAGES To create images out of the cells, each cell can be thought of as a pixel. However, the cell size is not constant along $|\eta|$ and across layers. To treat each cell as a pixel of an image, each cell has to have the same dimension in η and ϕ . To achieve a constant cell dimension, cells are split into smaller cells if they are too large. The energy is distributed evenly to the smaller cells, not changing the overall sum of energy. A detailed description of the image creation can be found in Ref. [146].

POINT CLOUDS To represent the cells as point clouds, the cells are sorted by their energy in each layer of the calorimeter system. In each layer, a maximum number of cells considered needs to be defined due to technical reasons. Due to the varying resolutions of the different layers, this number will vary. For the presampler, layer three of the ECAL, and all hadronic layers combined, only a maximum of ten cells are considered. For the first layer of the ECAL, the highest number of cells is considered with 80, while for the second layer only 40 cells are taken into account. The relative energy of the cluster and the relative η and ϕ positions are used as features of each cell. The relative energy is standardised with a StandardScaler before passing it into the network. No preprocessing is applied to the relative positions.

9.2.3 Tracks

During the reconstruction, multiple tracks can be assigned to an electron. Every track within a certain window in η and ϕ of the topo-cluster used for the reconstruction is considered. The tracks are sorted by their quality, which is judged by the tracking detectors that registered hits for the specific track. For instance, a track with hits in the Pixel detector and the SCT will be ranked higher than a track with hits only in the SCT. Tracks with the same quality are sorted by their proximity to the cluster centre.

Up to five tracks assigned to the electron are represented as a point cloud. Six features are used for each track. The features used are the transverse momentum of the track scaled by the transverse momentum of the first track, the relative η and ϕ positions with respect to the first track, the transverse impact parameter d_0 of each track as well as its uncertainty, and the transverse momentum of the

track scaled by the transverse energy of the cluster. Each of the track features is separately standardised using a StandardScaler.

9.2.4 Further preprocessing

As for the DNN, no bias should be introduced with respect to E_T and $|\eta|$ or a specific class. Instead of using the same approach of downsampling and reweighting, a different method is used that does not depend on weights. During the training, candidates are sampled from the training dataset. This means that for every batch, a subset of the training dataset is selected, allowing for repeats. A single candidate might therefore be used multiple times while another is not used at all. The sampling is done probabilistically such that the same E_T and $|\eta|$ distributions as for the DNN training are obtained over several batches. Additionally, all classes are equiprobable, so for each class, the same number of candidates are passed through the model.

As each batch can now contain candidates multiple times, an epoch cannot be defined as passing all candidates once through the network. Instead, an epoch is defined as a certain number of batches being processed by the network. This number can be chosen arbitrary.

9.3 Models

All models trained in this chapter extract information from the low-level data before concatenating the extracted information with the high-level features. The result of the concatenation is passed to a final network which ultimately classifies the candidates. To make the comparison as fair as possible, the architecture of this final network is the same for all models. It is chosen to be the same as used for the DNN in Chapter 8. Hyperparameters concerning the remaining parts of the models were optimised for the best performance. They are listed in Table A.1.

The training is performed with TensorFlow version 2.10. Each model is trained for 40 epochs using the *AdamW* optimiser. The same learning rate schedule as for the DNN in the previous chapter is used. It follows a triangular pattern with a starting value of 10^{-5} . The learning rate rises linearly up to $5 \cdot 10^{-3}$ before decreasing back down to the starting value. One such cycle is performed over the 40 epochs, adjusting the learning rate after each batch.

For best comparability, the same batch size and the same definition of an epoch are used.

9.3.1 Comparison of operations on cells

Different options exist to process the cells in the calorimeter. They can be arranged into images or point clouds. Additionally, this arrangement is not unique, so different arrangements can be thought of. For instance, for the point cloud representation, two different arrangements will be used. In the first, each layer of the calorimeter

is treated as a separate point cloud, whereas in the second, all cells end up in one single point cloud after an initial embedding step. Depending on the cell representation, different network architectures have to be used. These different architectures will be compared.

As the high-level features provide important information for electron identification and not all features can be recalculated by only the cells, all high-level features are used for the training of each network in addition to the cells. For these models, no track information is used.

IMAGES When representing the cells as images, the chosen network architecture is a CNN. Ref. [146] details the architecture used.

So-called coarse images with a resolution of 7×11 cells in $\eta \times \phi$ in the second layer of the ECAL are created for all layers. Additional fine images are created only for the first layer of the barrel and the endcap. These layers have excellent resolution along η . Therefore, the fine images have eight times more pixels along η compared to the coarse images, leading to 56×11 pixels. Both sets of images are processed by separate convolutional operations. The output of both of these is eventually flattened and, together with the high-level features, passed to the final DNN.

MULTIPLE POINT CLOUDS The main comparison of different algorithms operating on point clouds is performed when representing the cells as multiple point clouds. For this representation, a variety of algorithms are trained: a DeepSets, a GNN, and a transformer. In total, the cells are arranged in five point clouds. Cells in the barrel and endcap in corresponding layers are placed into the same point cloud to reduce the number of point clouds. For similar reasons, all cells of the HCAL comprise one point cloud. Therefore, four point clouds represent the ECAL in the presampler and layers one through three, and one the HCAL. Each point cloud is processed separately. The output of the single graph operations are concatenated and passed through a small DNN before being concatenated with the high-level features as input for the final DNN. A sketch of this structure can be seen in Figure 9.1. The GNN and the transformer have additional embedding networks for each layer before applying the graph-based operation. The embedding network is similar to a DeepSets in that it processes each cell in the point cloud with the same DNN. Especially for the transformer architecture, this is necessary to increase the number of features, as the used transformer encoder layer cannot change it. For the GNN, this embedded space is used to define edges between the cells. An edge is created between each cell and its four nearest neighbours. As the behaviour of the candidates in the calorimeter system depends on E_T and $|\eta|$, the embedding networks and the DeepSets networks are conditioned on E_T and $|\eta|$ by concatenating their values to the cells before passing them through the network.

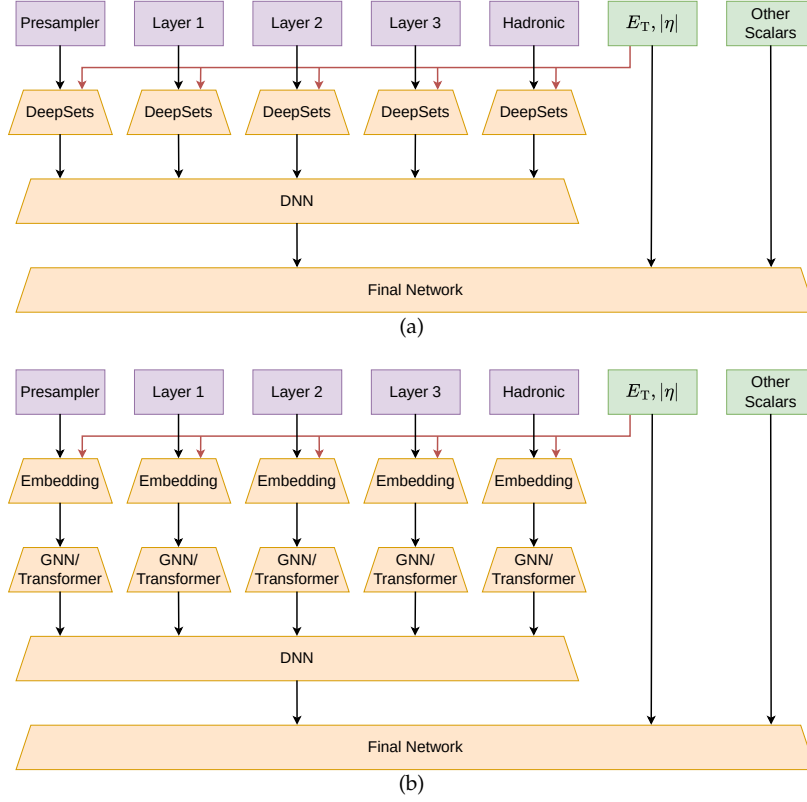


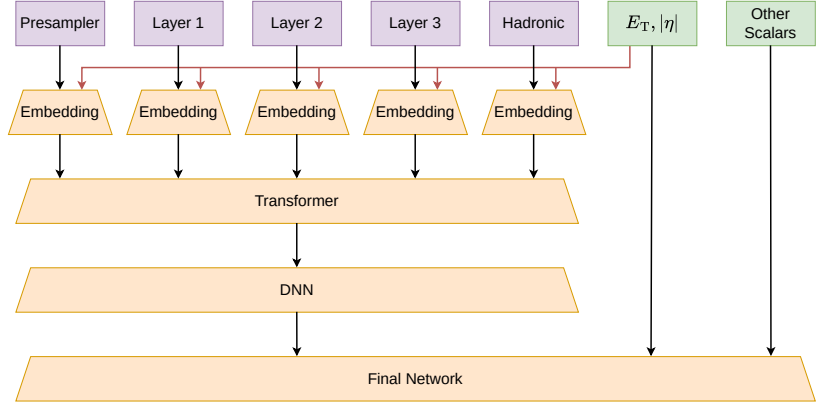
Figure 9.1: Schematic overview over the architecture of the networks processing the different point clouds with separate networks. (a) shows the DeepSets architecture, while (b) shows the architecture of the transformer and GNN based model.

SINGLE POINT CLOUD One drawback of representing the cells as multiple point clouds is the limited information exchange between the single layers. In a combined point cloud, cells could receive information from other layers in addition to their own, potentially improving performance. As the different layers will behave differently, each layer is processed by a separate embedding network. This ensures that the network can treat the single layers differently. Additionally, a one-hot encoded vector is concatenated with the embedded features. This vector encodes the layer the cell is in. As the different point cloud-based architectures are already compared in the multiple point cloud setup, only one architecture in the transformer is trained for this representation. This is further motivated by the fact that, using the multiple point cloud representation, no large differences between the different models are observed. The choice of the transformer is taken as for the GNN, additional considerations about the formation of the edges have to be taken. Specifically, how to connect cells from different layers. For the DeepSets algorithm, having a single point cloud has no benefits as there is, by design, no information exchange between the single cells. A sketch of the used architecture can be seen in Figure 9.2. As for the multiple point cloud models, E_T and $|\eta|$ are used to condition the embedding networks.

9.3.2 Comparison of inputs

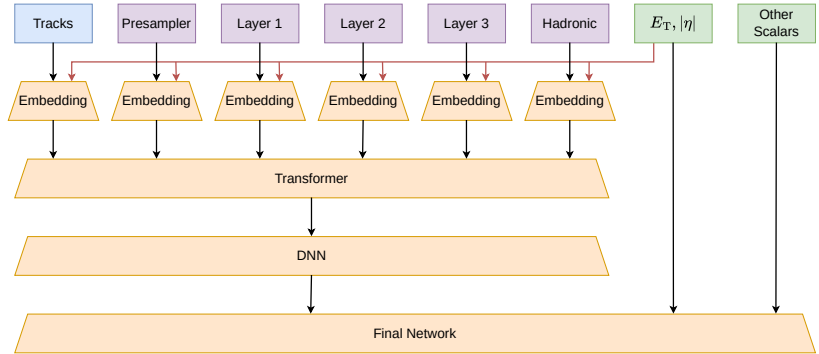
To estimate the increase in performance by including low-level information on top of the high-level features, three different models

Figure 9.2: Schematic overview over the architecture of the transformer-based network processing the different point clouds within the same point cloud after the embedding.



are trained. Firstly, a DNN with the same architecture as described in Chapter 8 using only high-level features; secondly, a model including cell information; and lastly, a model including cell and track information. The DNN is retrained on the same data with the same training procedure as the point cloud-based models to ensure a fair comparison. For the model with cell information and high-level features, the transformer model trained on a single point cloud from the previous comparison is used. No additional training is performed. For the last model, this architecture is extended to include track information. The tracks are treated as an additional point cloud, which is passed through an additional embedding network and then included in the single point cloud. A sketch of this architecture can be seen in Figure 9.3.

Figure 9.3: Schematic overview over the architecture of the transformer-based network processing the different point clouds including the track point cloud within the same point cloud after the embedding.



9.4 Performance

To compare the performance of the different models, a test dataset that is statistically independent of the training dataset is used. Candidates are only taken from the $Z \rightarrow ee$ and the JF17 samples. The top-quark pair production sample is not considered for the test dataset. The final discriminant is formed following Equation (8.1) with the same optimisation strategy of the f_X being used.

9.4.1 Comparison of methods

ROC curves of the models using cell information and high-level features are shown in Figure 9.4. These models are: a CNN, a DeepSets, a GNN, a transformer trained on multiple point clouds (MPC Transformer), and a transformer trained on a single point cloud (SPC Transformer). Some small differences in the performance can be seen. The point cloud-based models outperform the CNN by up to 15% on the combined background. Including some form of information exchange between the cells, as in the GNN and the transformers, increases the performance over the DeepSets model. None of the models including direct information exchange performs clearly better than the remaining ones. Depending on the signal efficiency, a different model has the highest combined background rejection.

Looking at the single backgrounds, the behaviour is different. The performance for the photon conversion and heavy-flavour backgrounds is similar for all models. Larger improvements of the point cloud-based models over the CNN can be seen for the two light-flavour backgrounds. An increase of up to 40% can be seen for the light-flavour e/γ background, while it is a bit smaller for the hadronic background with an increase between 10% and 20%. Especially at higher signal efficiencies, better performance can be seen for the point cloud-based models. For lower signal efficiencies, the point cloud-based models still tend to improve over the CNN but mostly to a smaller degree. Only the single point cloud transformer with an increase in rejection of almost 50% has a higher increase at low signal efficiencies compared to high signal efficiencies. For all backgrounds, the ordering of the point cloud-based models is similar. The DeepSets is usually the worst. For the remaining three models, depending on the background, one or the other performs best. For instance, for the heavy-flavour background, the GNN tends to be best, while for the light-flavour e/γ background, it is the transformer trained on multiple point clouds.

The CNN has the worst performance for charge flip electrons since it has the highest rejection, as can be seen in Figure 9.5. The best model for charge flip electrons is the transformer trained on a single point cloud.

The statistical uncertainties often cover the differences between the different models, at least partially. Also, the choice of the f_X can change the behaviour of the discriminant significantly, as seen in Section 8.5.2. Therefore, differences in the rejection of single backgrounds might not be caused by inherent differences in the methods but a difference in the choice of the f_X .

The rejection plots in Figure 9.6 show the behaviour of the different models along E_T , $|\eta|$, and $\langle\mu\rangle$. All models are tuned to have 80% signal efficiency in the fine bins given in Tables 8.4 and 8.5. As these are therefore the same for all models, they are not shown. The DeepSets model, together with the CNN, have overall the worst rejection. The three models with information exchange between single

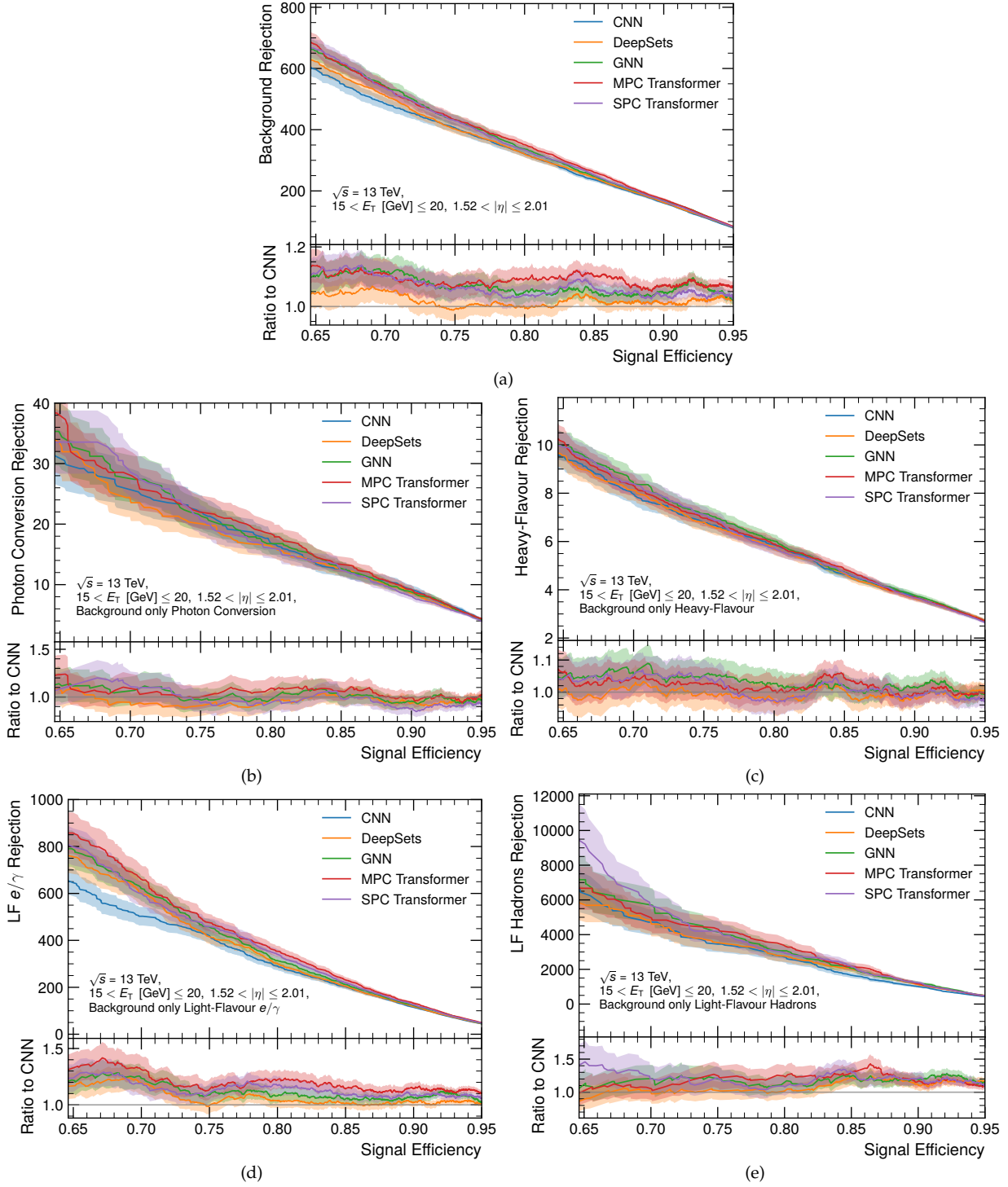


Figure 9.4: ROC curves for (a) the combined background, (b) only electrons from photon conversions, (c) only electrons from heavy-flavour quark decays, (d) only electrons and photons from light-flavour hadrons, and (e) only light-flavour hadrons. Only candidates with $15 < E_T [\text{GeV}] \leq 20$ and $1.52 < |\eta| \leq 2.01$ are shown. The models comparing different methods of processing the cells of the calorimeter system are shown. Statistical uncertainties of the background rejection are shown as bands.

cells outperform these two models in all regions of phase space. The two transformer-based models are slightly more performant than the GNN. As the GNN only exchanges information between a small amount of neighbours when updating the nodes, it is not unexpected that it performs worse than the transformer-based models. In the transformer architecture, information is exchanged between all nodes in each step. By increasing the number of nearest neighbours used in the GNN, the performance might be improved to similar levels as

the transformer-based models. However, this will also slow down the training significantly.

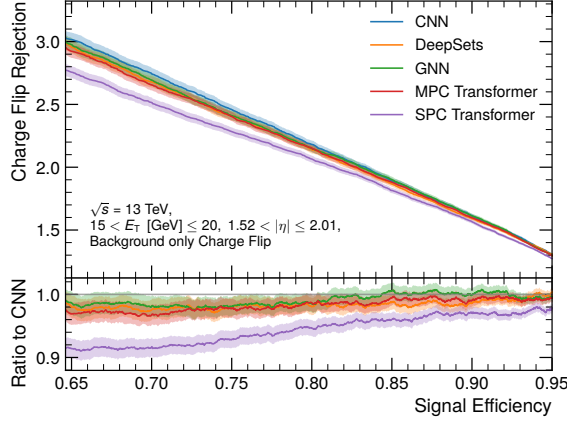


Figure 9.5: ROC curves with charge flip electrons as background. Only candidates with $15 < E_T [\text{GeV}] \leq 20$ and $1.52 < |\eta| \leq 2.01$ are shown. The models comparing different methods of processing the cells of the calorimeter system are shown. Statistical uncertainties of the background rejection are shown as bands.

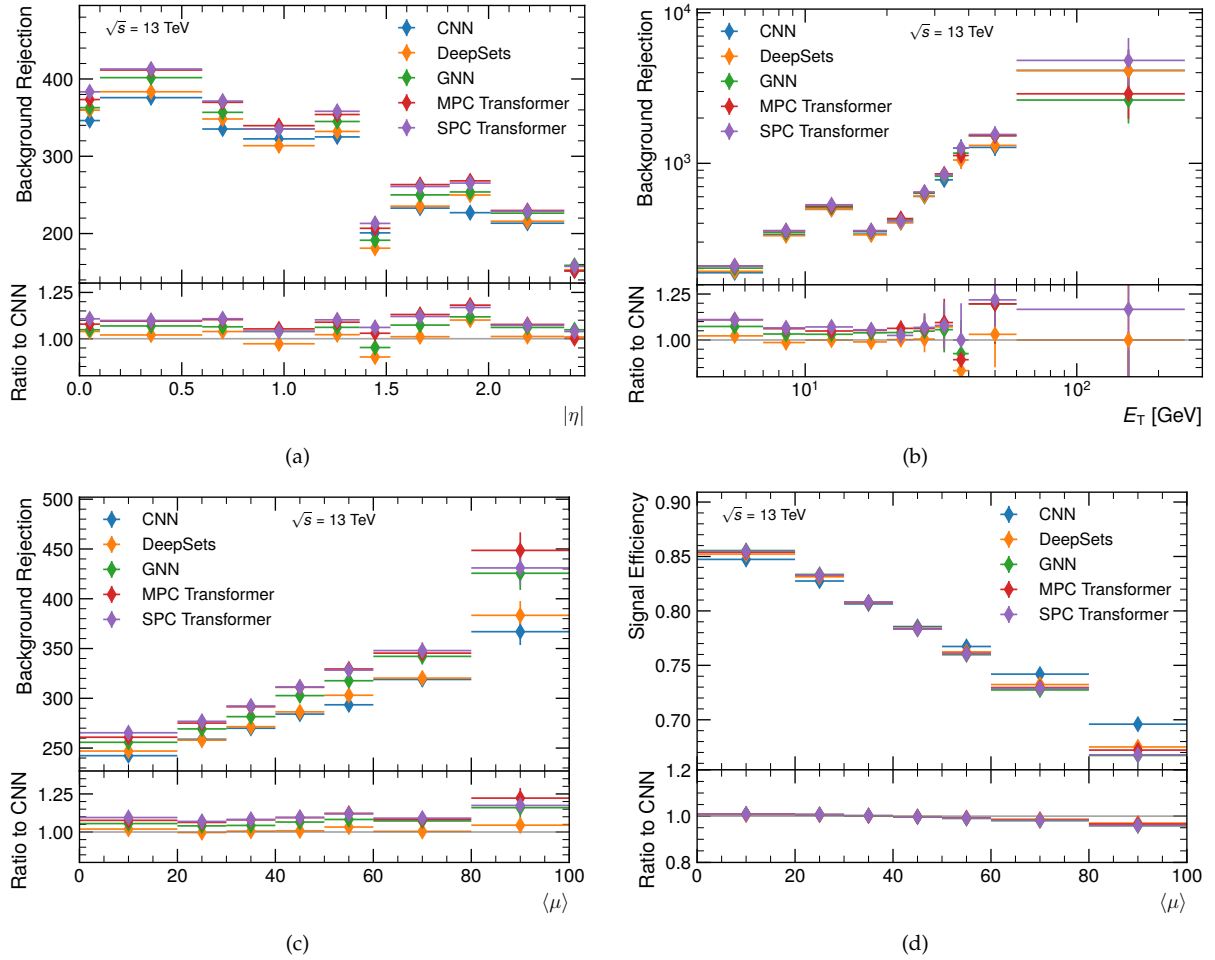


Figure 9.6: Performance plots of the models comparing different methods of processing the cells of the calorimeter system. (a), (b), and (c) show the background rejection with respect to E_T , $|\eta|$, and $\langle \mu \rangle$. (d) shows the signal efficiency with respect to $\langle \mu \rangle$. The last bin in (b) contains all electron candidates up until $E_T = 2$ TeV. The signal efficiencies with respect to E_T and $|\eta|$ are not shown since they are constant at 80% for all models by design.

An interesting feature can be seen in the rejection along E_T . The rejection for all models rises with increasing E_T , except around 15 GeV where the rejection drops. Afterwards, it rises similarly to before the dip. This feature might be caused by the use of the JF17 sample. As it has a filter requirement at 17 GeV, the background composition might differ below and above 17 GeV.

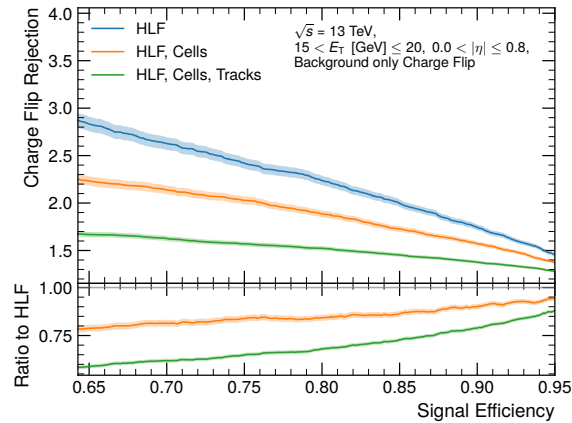
Similar trends as for the DNN in Figures 8.11 and 8.12 can be seen for the behaviour with respect to $\langle\mu\rangle$. The signal efficiencies drop with rising $\langle\mu\rangle$ while the background rejection rises. No significant difference in the behaviour of the background rejection can be seen, but for the signal efficiencies, the CNN seems to be less impacted by increasing $\langle\mu\rangle$ than the point cloud-based models. As this difference is visible for all point cloud-based models, this feature is likely a consequence of the different representations.

While background rejection is the most important metric, the time it takes to train the models is also of interest. As it depends on multiple factors, such as the number of parameters and the hardware used, only a qualitative comparison will be done. Out of the shown models, the DeepSets model trains the fastest by a large margin. The transformer-based models are faster than the CNN and the GNN, with the GNN being slower than the CNN. No large difference can be seen between the transformer-based models trained on a single point cloud and multiple point clouds.

9.4.2 Comparison of inputs

By including additional information, the rejection of charge flip electrons is reduced, as can be seen in Figure 9.7. As charge flip electrons are treated as signal in the discriminant, this is an increase in performance. The model with cell and track information even has close to the same efficiencies for signal and charge flip electrons.

Figure 9.7: ROC curves with charge flip electrons as background. Only candidates with $15 < E_T [\text{GeV}] \leq 20$ and $|\eta| \leq 0.8$ are shown. The three models comparing different inputs are shown. Statistical uncertainties of the background rejection are shown as bands.



ROC curves for the different backgrounds are shown in Figure 9.8. Unsurprisingly, the inclusion of more information increases performance. Using both cell and track information in addition to the high-level features, the performance can be improved by a factor of around two over using only high-level features. By only including the cell information, the improvement over the simple DNN is around 20%.

Large differences can be seen between the different background classes. The performance of the photon conversion background does not improve much when including only the cell information. But adding the track information on top boosts the performance by a

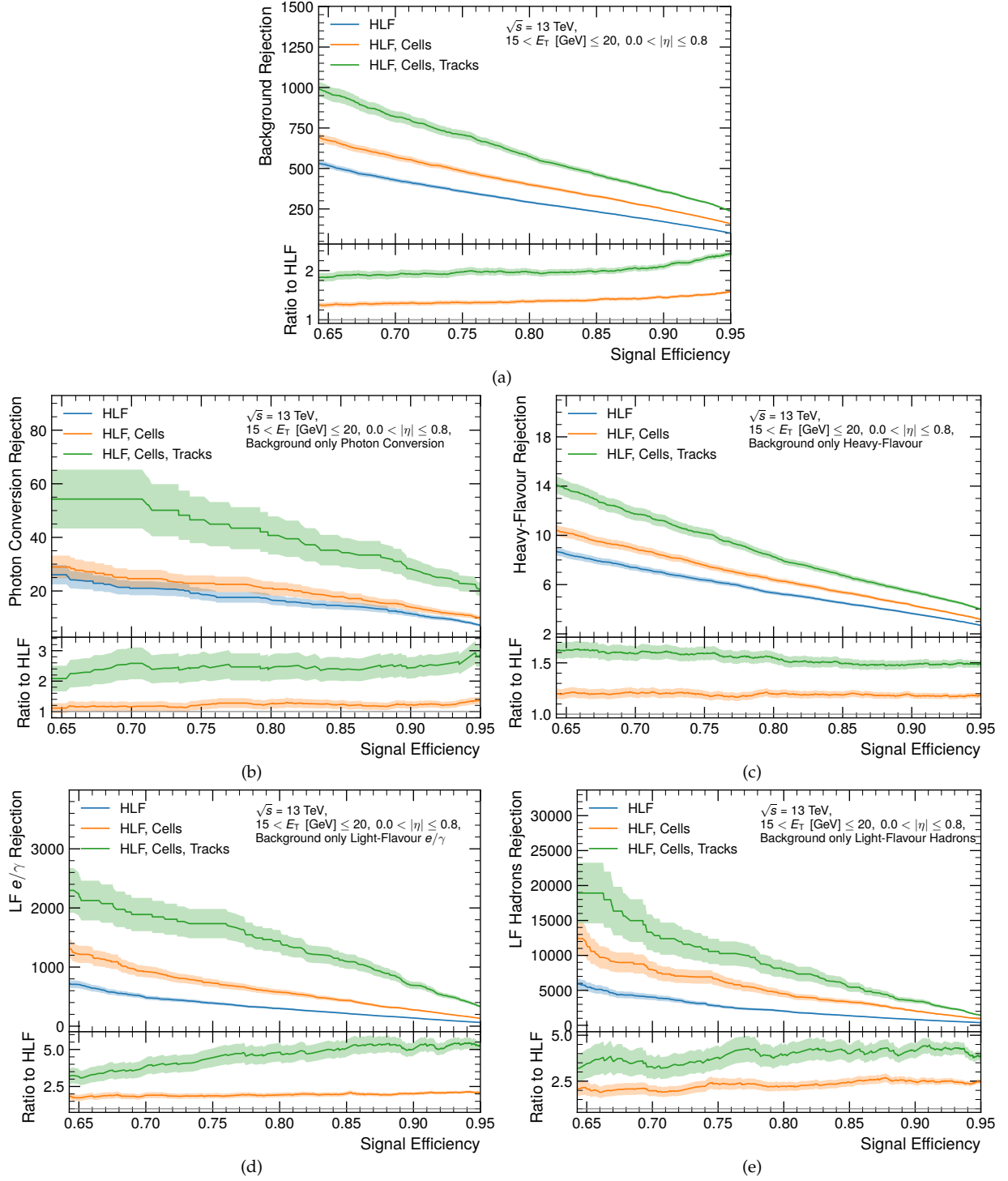


Figure 9.8: ROC curves for (a) the combined background, (b) only electrons from photon conversions, (c) only electrons from heavy-flavour quark decays, (d) only electrons and photons from light-flavour hadrons, and (e) only light-flavour hadrons. Only candidates with $15 < E_T [\text{GeV}] \leq 20$ and $|\eta| \leq 0.8$ are shown. The three models comparing different inputs are shown. Statistical uncertainties of the background rejection are shown as bands.

factor of over two. This can be explained by the nature of the photon conversion background. In the calorimeter, it consists mostly of actual electrons, so the cell information cannot help much. The main difference to signal electrons happens in the tracking detector, which is why the track information boosts the performance considerably. For the remaining backgrounds, a considerable increase in performance can be seen even when not including the track information. As electrons from heavy-flavour decays are also actual electrons in the

calorimeter, the main increase in performance most likely originates from the additional hadronic activity. The increase in performance is largest for the light-flavour backgrounds. Including cell and track information, the rejection can be increased by a factor of up to five (four) for the light-flavour e/γ (light-flavour hadron) background.

The rejection across $|\eta|$, E_T , and $\langle\mu\rangle$ as well as the signal efficiency across $\langle\mu\rangle$ are shown in Figure 9.9. Again, the signal efficiencies of the different models are tuned to 80% in bins of E_T and $|\eta|$. A larger improvement from the additional information can be seen closer to the crack region. The largest improvement can be seen in the crack region itself, with an increase in rejection of over five. Interesting to see is also that the rejection in the crack is higher than in the forward region ($|\eta| > 2.37$) for the models with low-level information, which is not the case for the simple DNN.

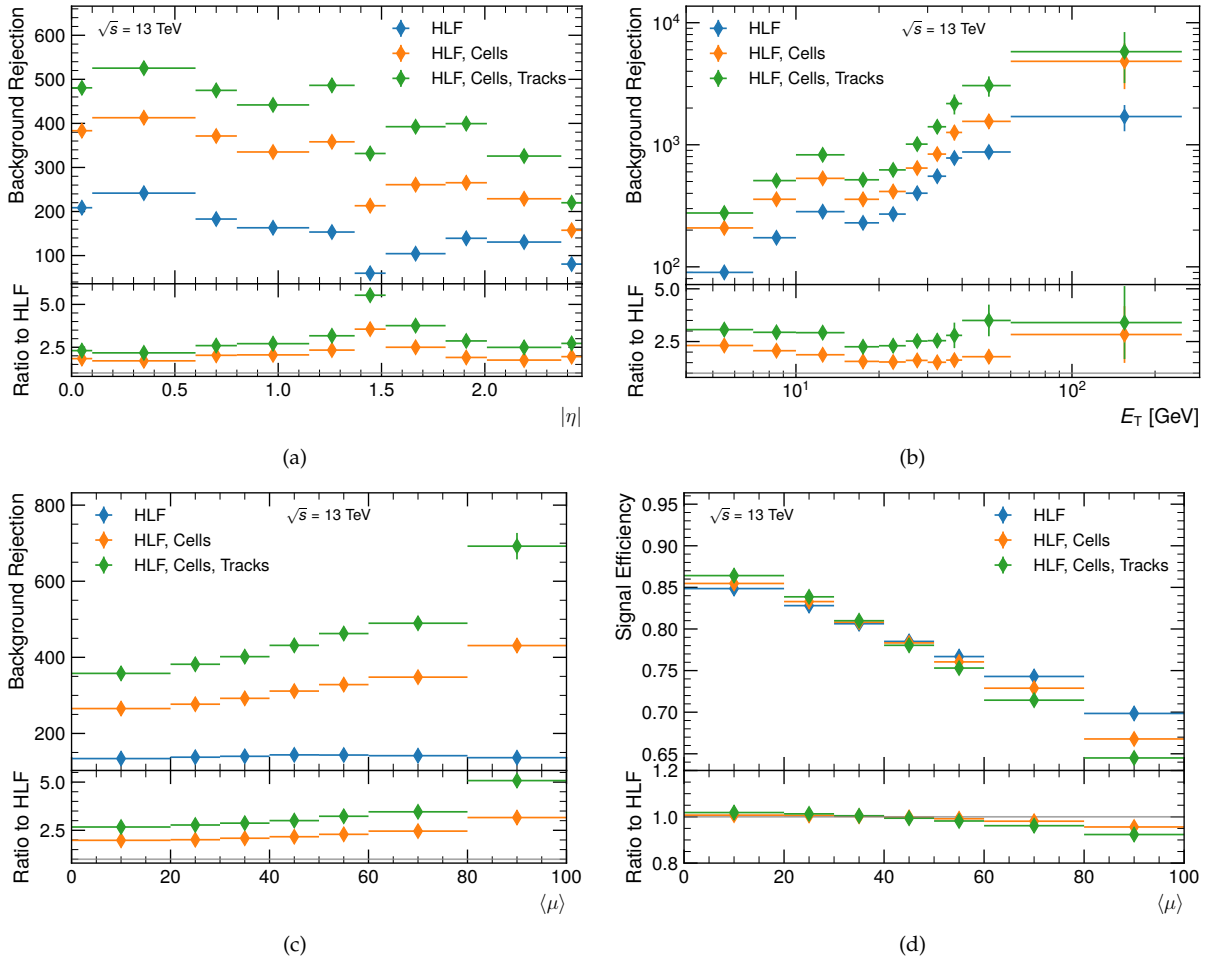


Figure 9.9: Performance plots of the three models comparing different inputs. (a), (b), and (c) show the background rejection with respect to E_T , $|\eta|$, and $\langle\mu\rangle$. (d) shows the signal efficiency with respect to $\langle\mu\rangle$. The last bin in (b) contains all electron candidates up until $E_T = 2$ TeV. The signal efficiencies with respect to E_T and $|\eta|$ are not shown since they are constant at 80% for all models by design.

Across E_T , the improvement is fairly constant. The lowest improvement can be seen between $15 \text{ GeV} < E_T < 45 \text{ GeV}$. The simple DNN has a constant background rejection with respect to $\langle\mu\rangle$ which is not the case when including cell and track information. The models with more information have an increasing rejection with rising $\langle\mu\rangle$. However, this rising rejection comes at the cost of a stronger decrease in signal efficiency with increasing $\langle\mu\rangle$.

10 Performance in data

As mentioned previously, the input variables of the DNN are corrected to better match the observed distribution in data. However, these corrections will not be perfect, leading to differences in the signal efficiencies and background rejections between MC and data. Additionally, the point cloud-based models do not have any corrections applied to the training data. To estimate the performance in data, a signal and background rejection is performed. Based on these, ROC curves will be shown. Background contamination of the signal selection is not taken into account in this estimate. A more precise calculation of signal efficiencies and therefore SFs can be found in Appendix B.

Contents

10.1 Signal selection	109
10.2 Background selection	110
10.3 Performance evaluation	111

Only a fraction of events recorded in 2017 corresponding to an integrated luminosity of 2.4 pb^{-1} are used for this study.

10.1 Signal selection

A similar selection to the one used to obtain electron candidates to construct the PDFs for signal electrons in data is used. Only events passing one of the single electron triggers defined in Table 4.6 are considered. The tag electron has to have sufficiently high transverse energy $E_T > 27 \text{ GeV}$. It needs to be outside the crack region $1.37 < |\eta| < 1.52$ and within the ID acceptance $|\eta| < 2.47$. It needs to pass the *LHTight* WP and a custom isolation requirement of $p_T^{\text{topocone20}}/p_T < 0.1$. Lastly, it needs to be the electron that caused the trigger to fire.

No requirements are placed on the probe electron apart from its transverse energy, which needs to fulfil $E_T > 15 \text{ GeV}$.

The remaining requirements are all placed on the tag-and-probe pair. The two electrons have to have opposite electrical charges, and their invariant mass has to be within $75 \text{ GeV} < m_{ee} < 105 \text{ GeV}$. If multiple possible tag-and-probe pairs exist, for instance, by exchanging tag and probe, all possible probes are considered for the measurement.

Figure 10.1 shows the discriminant of the point cloud-based models and the DNN trained in Chapter 9 in multiple E_T bins. Only the very central region with $|\eta| < 0.8$ is shown. All models have similar behaviour. In all bins, a clear peak is visible at high values of the discriminant. In this peak, most of the candidates will be actual signal electrons. At lower values of the discriminant, an additional peak can be seen, especially at low values of E_T . With increasing energy, this peak reduces in size. Most of the candidates in this peak will be background. The reduction in size with rising transverse energy has several reasons. Firstly, the amount of hadronic background will decrease with rising E_T due to the cross-section of the production of

hadrons. Secondly, the amount of signal electrons increases due to electrons from a Z boson decay being targeted. The distribution in E_T of these electrons peaks around 45 GeV and drops off to both sides. Lastly, above 26 GeV, an additional increase in signal electrons will occur due to the E_T threshold of the trigger used.

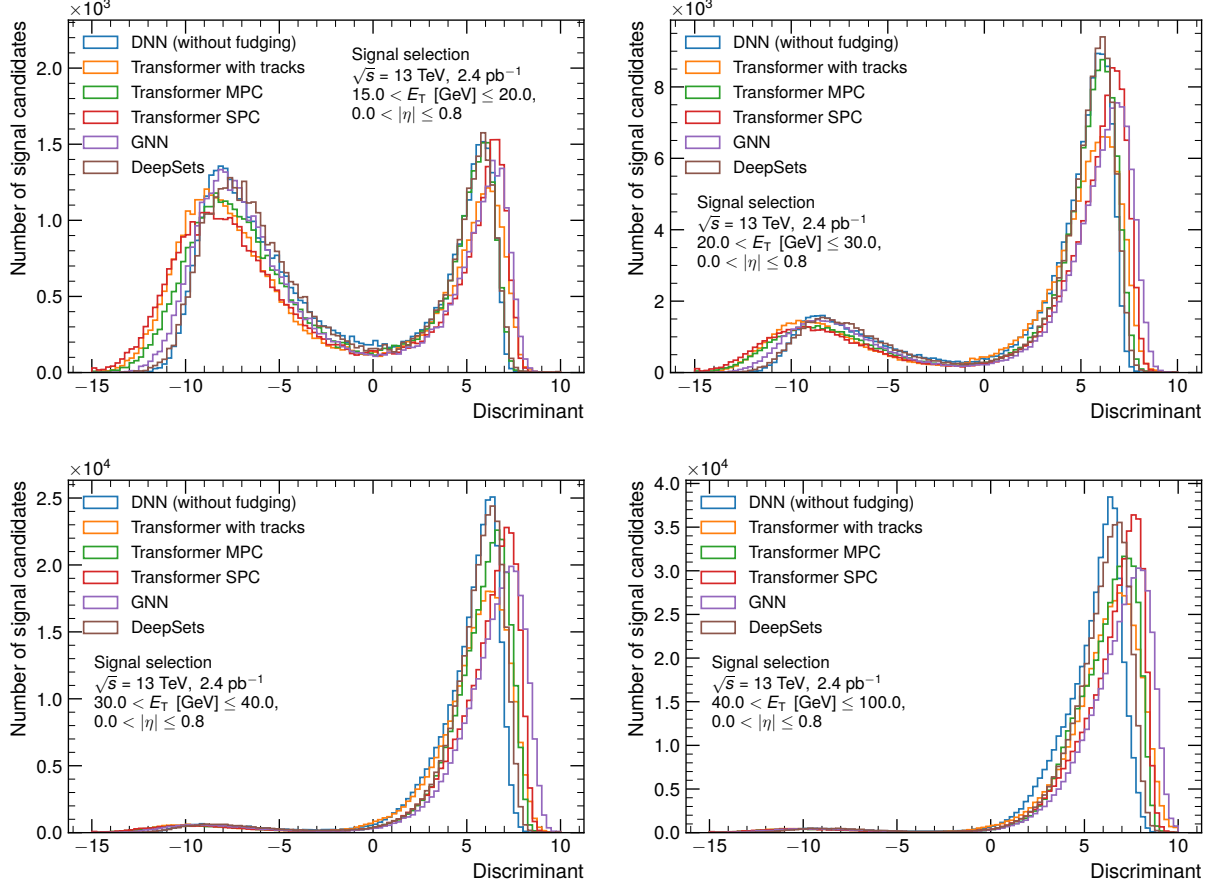


Figure 10.1: Discriminant of candidates passing the signal selection split into bins in E_T and $|\eta|$. The first and last bin contain under- and overflow events, respectively.

10.2 Background selection

To pass the background selection, the candidate has to have sufficient transverse energy $E_T > 15$ GeV. Additionally, it needs to have fired one of the prescaled triggers listed in Table 10.1. None of these have any identification requirements included to not bias the selection. To veto events containing a W boson decaying into an electron, the missing transverse momentum has to be $E_T^{\text{miss}} < 25$ GeV and the transverse mass of the lepton and the missing transverse momentum has to be $m_T < 40$ GeV. To veto events containing a Z boson decaying into an electron-positron pair, no additional electron passing the *LHMedium* WP with $E_T > 4$ GeV is allowed in the event such that the invariant mass of the candidate electron and the additional electron is within 10 GeV of the Z boson mass of 91 GeV.

Figure 10.2 shows the discriminant of the point cloud-based models and the DNN trained without correction factors for candidates

Table 10.1: Prescaled single electron triggers used for the background selection.

HLT_e5_etcut
HLT_e10_etcut_L1EM7
HLT_e15_etcut_L1EM7
HLT_e20_etcut_L1EM12
HLT_e25_etcut_L1EM15
HLT_e30_etcut_L1EM15
HLT_e40_etcut_L1EM15
HLT_e50_etcut_L1EM15
HLT_e60_etcut
HLT_e70_etcut
HLT_e80_etcut
HLT_e100_etcut
HLT_e120_etcut

passing the background selection. Again, all models show similar behaviour. In all bins, only one peak is clearly visible at low values of the discriminant, hinting towards the background selection having very low signal contamination.

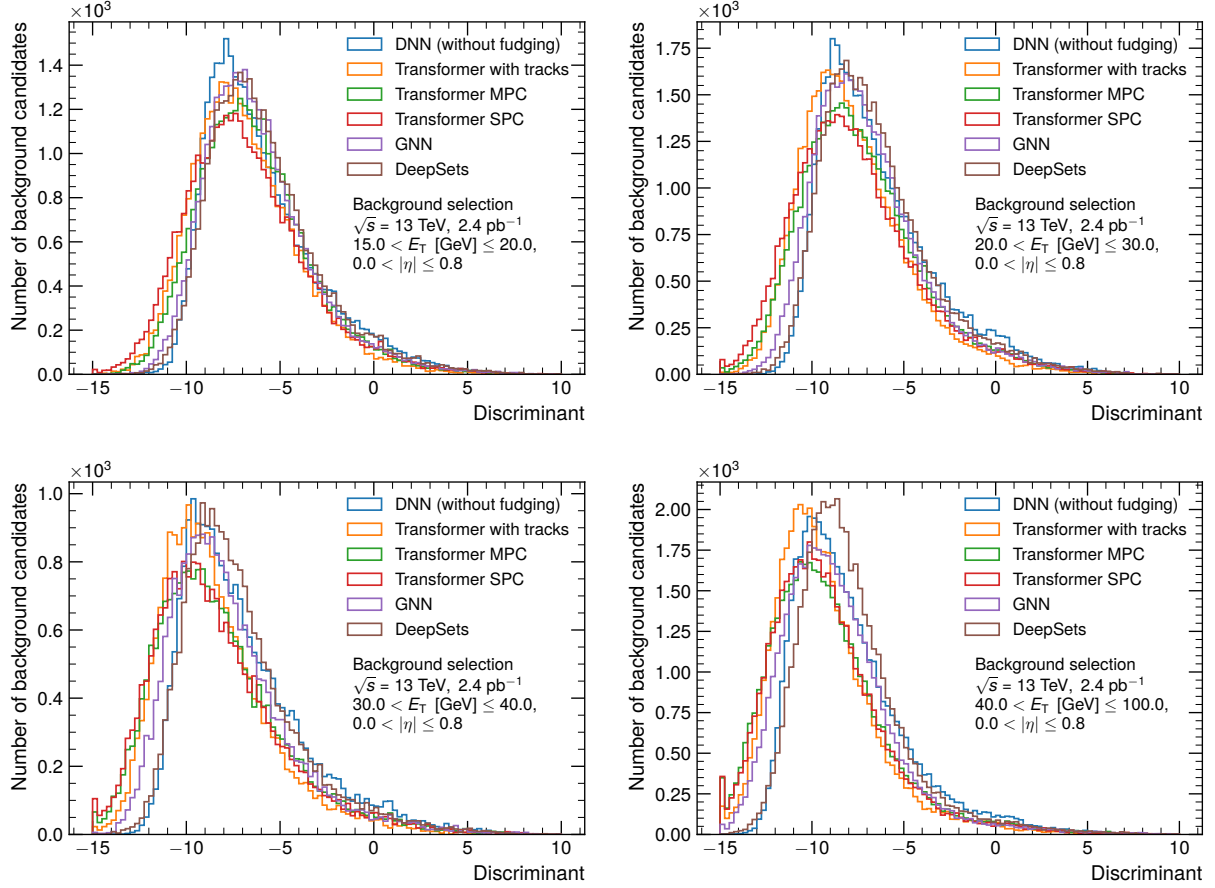


Figure 10.2: Discriminant of candidates passing the background selection split into bins in E_T and $|\eta|$. The first and last bin contain under- and overflow events, respectively.

10.3 Performance evaluation

To estimate the performance in data, ROC curves will be used. As the signal selection is contaminated with background, special considerations have to be made to compare the results. As long as the background that contaminates the signal selection behaves similarly to the background in the background selection, the ROC curves can be interpreted in the same way as for pure selections. If one considers a signal selection with 90% signal and 10% background and a pure background selection, two different algorithms with a background rejection of 100 and 200 at 80% true signal efficiency will result in background rejections of 100 and 200 at 72.1% and 72.05% of the signal selection, respectively. So the better algorithm will still look better using the ROC curve but with a slightly smaller difference between the two algorithms.

As the contamination of the signal selection is quite high for the lowest E_T bin, no ROC curves will be shown for this bin. Figure 10.3

shows the ROC curves for the LH algorithm and two DNNs, one trained with the correction factors applied and one without. They are referred to as "with fudging" and "without fudging", respectively. For the LH and the corrected DNN, only WPs are shown, whereas the uncorrected DNN has the actual ROC curve shown.

From the ROC curve, the background contamination of the signal selection can be estimated. The background rejection rises very little with decreasing signal efficiency until most of the background in the signal is rejected. At lower signal efficiencies, more typical behaviour can be seen. The DNN trained on corrected MC has a higher background rejection at a similar level of signal efficiency than the LH. The performance gains seen in Chapter 8 therefore translate into data, although maybe with a lower overall improvement. It can also be clearly seen, that the correction factors are needed. The DNN trained without them applied performs much worse than both the LH and the other DNN.

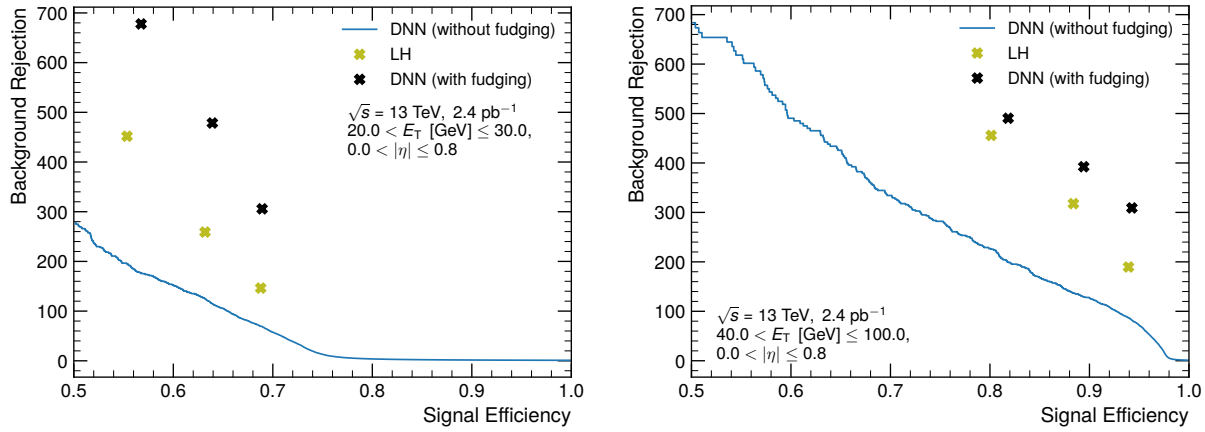


Figure 10.3: ROC curves using the signal and background selection in data. Shown are the LH algorithm and two DNNs, one trained with corrections applied to MC and one without these corrections applied.

Figure 10.4 shows the same results for the two DNNs but also adds all point cloud-based models. The same trend as in MC can be observed. The DNN trained without corrections performs worst. The inclusion of cells boosts the performance, with the DeepSets algorithm performing worse than the fully graph-based or transformer-based models. Including track information further improves the performance. However, even though the performance is improved in comparison to the DNN trained without corrections, the transformer with cell and track information still performs worse than the DNN trained with corrections.

Furthermore, Figure 10.4 only shows the central region of the barrel, which tends to have small amounts of mismodelling. Therefore, Figure 10.5 shows results over the whole η range of the detector. In the barrel region, the same observations as before can be made, though the improvements from the inclusion of track information are smaller. In the crack region, the transformer, including the track information, even performs worse than some models with only the cell information and the high-level features. Turning to the endcap region, the gap between the best point cloud-based model and the corrected DNN

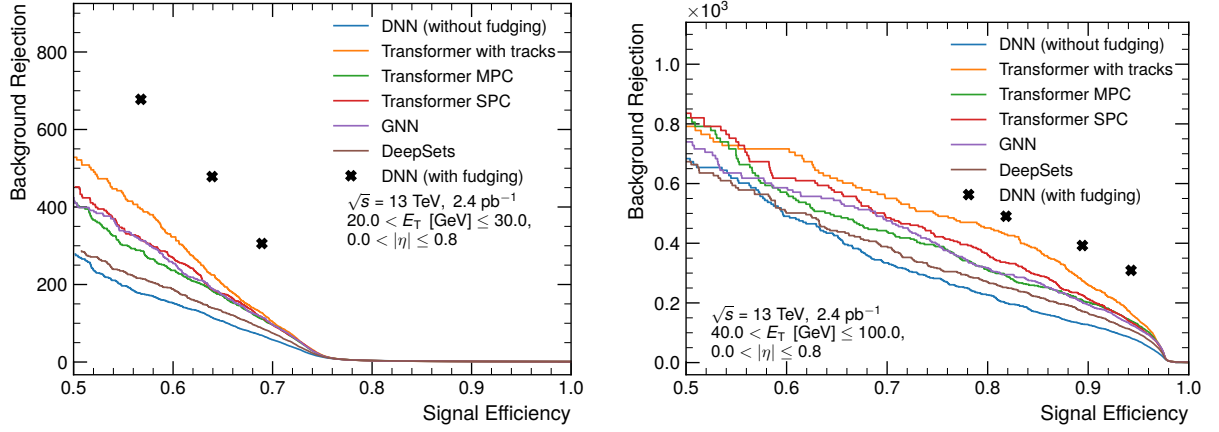


Figure 10.4: ROC curves using the signal and background selection in data. Shown are the point cloud-based models and two DNNs, one trained with corrections applied to MC and one without these corrections applied.

grows larger, especially in the highest $|\eta|$ bin with $2.01 < |\eta| < 2.47$. A few additional interesting features can be observed in the endcap. The DeepSets algorithm performs worse than the simple DNN. And in the highest $|\eta|$ bin, the inclusion of tracks results in the largest improvement over the models using only cell information and high-level features. Both of these observations can also be made at different energies.

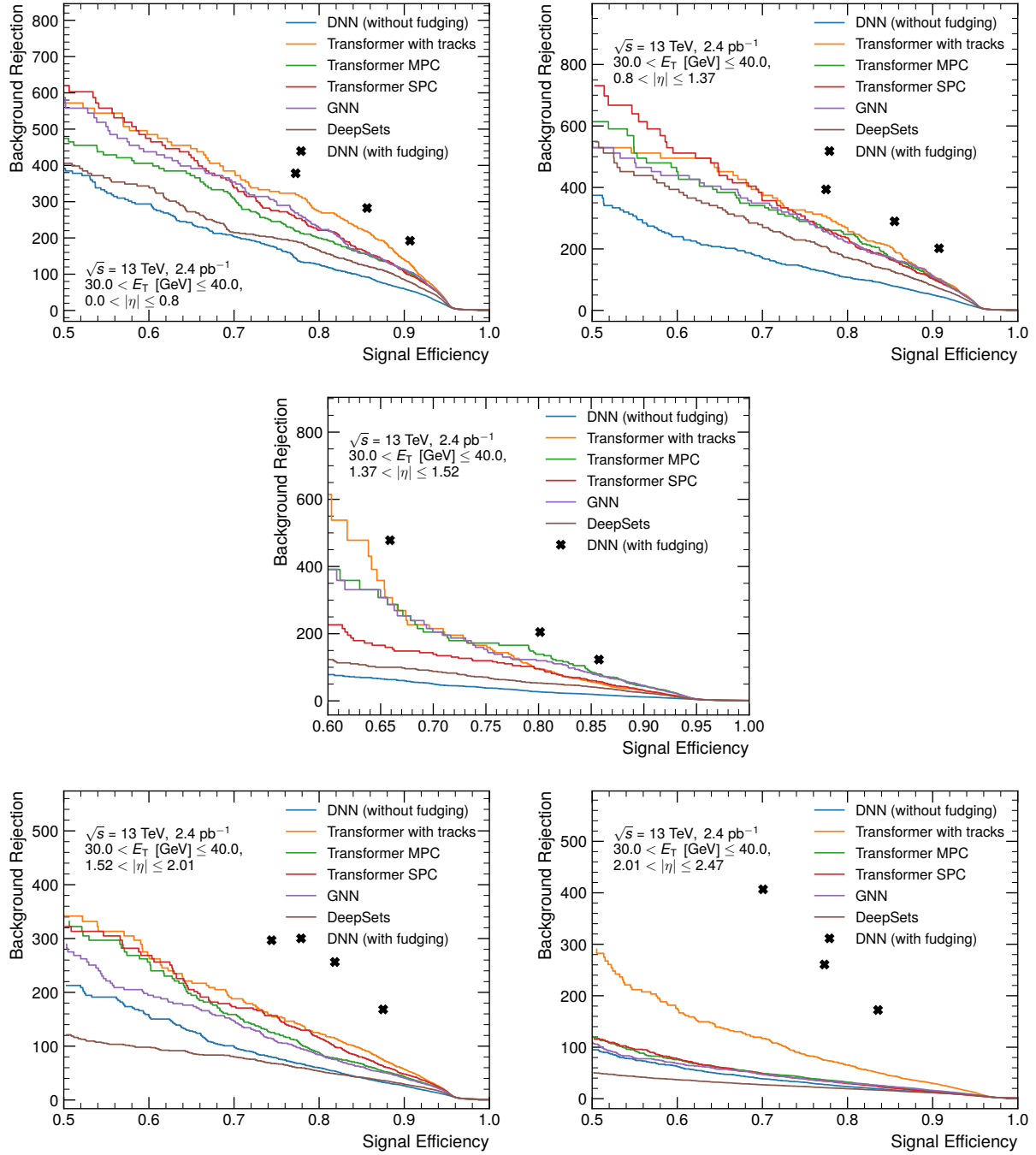


Figure 10.5: ROC curves using the signal and background selection in data. Shown are the point cloud-based models and two DNNs, one trained with corrections applied to MC and one without these corrections applied. Different $|\eta|$ ranges are shown in the different subfigures.

PART III

MEASUREMENT OF THE FIDUCIAL AND TOTAL CROSS-SECTION OF TOP-QUARK PAIRS PRODUCED IN ASSOCIATION WITH ADDITIONAL CHARM JETS

11 Analysis strategy

In this part, the analysis measuring the fiducial and total cross-section of top-quark pair production in association with additional charm jets will be described. This chapter will cover a variety of topics related to the analysis strategy, such as the extraction method and a dedicated flavour tagger developed specifically for this analysis.

Contents

11.1	Profile likelihood fit	117
11.2	Samples	119
11.3	Extraction of the cross-section	120
11.3.1	Definition of the fiducial phase space . .	121
11.3.2	Normalisation	123
11.4	Truth-level classification of jets	123
11.5	<i>bc</i> -Tagger	123
11.5.1	Standard flavour-tagging in ATLAS . . .	124
11.5.2	<i>bc</i> -Tagger definition	124

The analysis is performed in the lepton+jets and dilepton channels. The dilepton channel has lower available statistics but a purer final state. In the lepton+jets channel, an additional *c*-jet can originate from the hadronically decaying *W* boson. A reconstruction of the top-quark pair as introduced in Part IV could help reduce the complications introduced by these *c*-jets; however, it was chosen not to employ such a reconstruction in this first analysis. As the author worked mainly on the lepton+jets channel, the analysis will be described focusing on this channel. Combined results will be shown as well in Chapter 15.

11.1 Profile likelihood fit

The cross-section is extracted using a profile likelihood fit. The basic setup is an extended binned maximum likelihood fit, which can be written as

$$\log \mathcal{L}(N) = -\nu_{\text{tot}} + \sum_i^M n_i \log \nu_i, \quad (11.1)$$

after dropping constant terms, with ν_{tot} (N) the total number of expected (observed) events and ν_i (n_i) the number of expected (observed) events in the bin i of all M bins. This likelihood has, however, no adjustable parameters. As the cross-section of the signal should be estimated, a norm factor, also called signal strength μ is introduced, which is used to scale the signal contribution

$$\log \mathcal{L}(N|\mu_S) = -(\mu_S \nu_S + \nu_B) + \sum_i^M n_i \log(\mu_S \nu_{S,i} + \nu_{B,i}), \quad (11.2)$$

with $\nu_{S(B)}$ being the number of predicted signal (background) events. As the signal strength scales the signal prediction, it scales the cross-section. To obtain the measured cross-section, the measured signal

strength has to be multiplied by the predicted cross-section

$$\sigma_S^{meas} = \mu_S \sigma_S^{pred}. \quad (11.3)$$

Multiple signal strengths can easily be introduced by further splitting the single contributions.

So far, no systematic uncertainties have been considered. Systematic uncertainties are included using the template method and profiling. Each systematic uncertainty is represented by a nuisance parameter (NP) $\theta \in \Theta$. Including the dependence on systematic uncertainties, the likelihood has to be adjusted to

$$\log \mathcal{L}(N|\mu_S, \Theta) = -(\mu_S \nu_S(\Theta) + \nu_B(\Theta)) \quad (11.4)$$

$$+ \sum_i^M n_i \log(\mu_S \nu_{S,i}(\Theta) + \nu_{B,i}(\Theta)). \quad (11.5)$$

Systematic uncertainties lead to a different prediction than the nominal one, also called variation. In this work, each systematic uncertainty has either one or two variations. Uncertainties with two variations, an up- and a down-variation, are referred to as three-point systematics, whereas uncertainties with only one variation¹ are referred to as two-point systematics. In cases where there is only one variation, the variation is always symmetrised to obtain an up- and a down-variation. To note, the up-variation typically represents varying the uncertainty up, which might not result in more events. The difference between the nominal prediction and the variation is evaluated for each sample by filling histograms for each sample and each variation. These histograms are also referred to as templates. During the fit, each uncertainty is represented by a NP. If the uncertainty acts on multiple samples, each sample is modified by the same NP. If the NP has a value of zero, the nominal prediction is used. For a value of one, the up-variation of the uncertainty is used instead. For any values not equal to zero, one, or negative one, an interpolation or extrapolation between the nominal prediction and the templates is performed.

The nominal prediction includes the results of many dedicated measurements, for instance, a measurement of the luminosity. If the uncertainty of the luminosity is treated as a NP, the fit could change the value of the luminosity freely. This is, however, not wanted, as the dedicated measurement should have better precision than this measurement. To encode this prior knowledge, each nuisance parameter adds a penalty term to the likelihood. The penalty is parametrised by a Gaussian PDF with a mean corresponding to the nominal prediction and a standard deviation corresponding to the up- and down-variation. With this penalty, the NP will be forced to stay close to zero unless there is strong MC-data disagreement. A NP deviating from the nominal value after the fit is also referred to as pulled. A pull can appear if the varied template describes the observed data better than the nominal template. Similarly, a NP whose post-fit uncertainty is smaller than the prior uncertainty is called constrained. A constrained NP shows that the measurement performed

¹ This single variation will always be considered as the up-variation.

leads to a reduced uncertainty compared to the one initially assigned. For certain uncertainties, like the luminosity, this should never happen. The luminosity calibration is performed including information from additional detectors, whereas measurements in physics analysis have no information to determine the luminosity themselves. One reason for a constrained NP can be large opposing normalisation effects in different bins. Lastly, NPs can also be underconstrained, meaning the post-fit uncertainty is larger than the prior uncertainty. An underconstrained NP usually hints towards problems in the fit setup.

Another type of NPs will be used in this analysis to treat uncertainties due to the limited statistics of the generated samples. These are taken into account using so-called γ parameters. To reduce the number of parameters, the uncertainty of all predicted events is parametrised by one NP for each bin. They are assumed to be uncorrelated between the single bins pre-fit. Each of the γ parameters has a constraint placed on it following a Poisson distribution.

The implementation of the fit uses the HistFactory [205] software. It makes use of the RooStats [206] and RooFit [207] libraries implemented in the ROOT [208] software framework.

ASIMOV DATASET The fitting procedure is tested using a so-called Asimov dataset. The Asimov dataset is built from the predicted yields in each bin. Therefore, the prediction matches the pseudodata that is used for the fit perfectly. Hence, the fit result for all norm factors and all NP should be their nominal value. Furthermore, the uncertainties of the norm factors give an estimate of the expected uncertainty in the fit to real data. Therefore, this dataset can also be used to modify the fit setup, for instance, the event selection, to get the best uncertainty possible. At the same time, it can be used to investigate possible constraints on the NP.

11.2 Samples

A variety of different processes can have a final state similar to the targeted one. For all of these processes, simulated samples are used to predict their contribution. Table 11.1 lists all samples used, which are described in more detail in Section 5.7. For many processes, not only one sample but many samples are used. One of them is always used as a nominal prediction, while the remaining are used to estimate systematic uncertainties. More details about systematic uncertainties are given in Chapter 13. Z+jets events with at least one HF jet are weighted with a factor of 1.25, increasing their yield by 25% to improve the modelling. A similar scaling was used in previous analyses [8].

Samples for $t\bar{t}$ and $t\bar{t}b\bar{b}$ production are used. The $t\bar{t}$ sample also implicitly contains events with additional $b\bar{b}$ pairs, which would lead to a double counting of $t\bar{t}b\bar{b}$ events if samples of both processes were to be used as they are. Instead, only parts are used from each sample,

depending on the HF classification introduced in Section 11.4.

Other processes, like four top-quark production or the production of a single top-quark with an additional Higgs boson, were found to have negligible contributions to the event selection and are therefore not considered.

Process	Generator	Parton Shower	Additional	Usage
$t\bar{t}$	POWHEG BOX	PYTHIA		Nominal
	POWHEG BOX	PYTHIA	h_{damp} variation	Alternative
	POWHEG BOX	PYTHIA	$p_{\text{T}}^{\text{hard}}$ variation	Alternative
	POWHEG BOX	HERWIG		Alternative
$t\bar{t}b\bar{b}$	POWHEG BOX RES	PYTHIA		Nominal
	POWHEG BOX RES	PYTHIA	dipole recoil	Alternative
	POWHEG BOX RES	PYTHIA	h_{bzd} variation	Alternative
	POWHEG BOX RES	PYTHIA	$p_{\text{T}}^{\text{hard}}$ variation	Alternative
	POWHEG BOX RES	HERWIG		Alternative
single top (s- and t-channel, tW)	POWHEG BOX	PYTHIA	tW with DR	Nominal
	POWHEG BOX	HERWIG		Alternative
tW	POWHEG BOX	PYTHIA	DS	Alternative
	MADGRAPH5_AMC@NLO	PYTHIA		Alternative
$t\bar{t}V$	MADGRAPH5_AMC@NLO	PYTHIA		Nominal
	SHERPA	SHERPA		Alternative
$t\bar{t}H$	POWHEG BOX	PYTHIA		Nominal
tZq	MADGRAPH5_AMC@NLO	PYTHIA		Nominal
tWZ	MADGRAPH5_AMC@NLO	PYTHIA		Nominal
Z + jets	SHERPA	SHERPA		Nominal
W + jets	SHERPA	SHERPA		Nominal
Diboson	SHERPA	SHERPA		Nominal

Table 11.1: Overview of samples used in the $t\bar{t}$ + charm analysis and their generator settings. Samples only used as a comparison to the result and not in the fit are not included.

11.3 Extraction of the cross-section

The primary goal of the analysis is to extract the $t\bar{t} + \geq 2c$ and $t\bar{t} + 1c$ cross-sections, so of events with at least two additional c -jets and exactly one additional c -jet, respectively. Additionally, the ratio of $t\bar{t} + \geq 2c$ or $t\bar{t} + 1c$ production to the overall $t\bar{t}$ + jets production is interesting. These ratios have the additional benefit that systematic uncertainties that act the same on the numerator and denominator will be cancelled out in the measurement. To measure the ratios, the overall $t\bar{t}$ + jets production has to be measured. Therefore, four signal strengths are measured in a first fit: $t\bar{t} + \geq 2c$, $t\bar{t} + 1c$, $t\bar{t} + \geq 1b$, and $t\bar{t} + \text{light}$. From these signal strengths, the $t\bar{t}$ + jets cross-section and the ratios can be calculated. To correctly propagate uncertainties while forming the ratios, a second fit is performed where the signal strengths are reexpressed as the overall $t\bar{t}$ + jets cross-section and the ratios. The fit is exactly the same and converges to the same minimum. This is just used as a means of propagating uncertainties while correctly taking correlations between the signal strengths into account. To express the overall $t\bar{t}$ + jets signal strength with the signal strengths of the single $t\bar{t}$ component, they have to be summed, weighting each

signal strength with their relative contribution as predicted by MC

$$\begin{aligned} \mu(t\bar{t} + \text{jets}) &= \mu(t\bar{t} + \geq 2c) \cdot \frac{N^{\text{MC}}(t\bar{t} + \geq 2c)}{N^{\text{MC}}(t\bar{t} + \text{jets})} \\ &\quad + \mu(t\bar{t} + 1c) \cdot \frac{N^{\text{MC}}(t\bar{t} + 1c)}{N^{\text{MC}}(t\bar{t} + \text{jets})} \\ &\quad + \mu(t\bar{t} + \geq 1b) \cdot \frac{N^{\text{MC}}(t\bar{t} + \geq 1b)}{N^{\text{MC}}(t\bar{t} + \text{jets})} \\ &\quad + \mu(t\bar{t} + \text{light}) \cdot \frac{N^{\text{MC}}(t\bar{t} + \text{light})}{N^{\text{MC}}(t\bar{t} + \text{jets})}. \end{aligned} \quad (11.6)$$

On top of the $t\bar{t} + \text{jets}$ cross-section, the ratio of $t\bar{t} + \geq 2c$, $t\bar{t} + 1c$, and $t\bar{t} + \geq 1b$ production to $t\bar{t} + \text{jets}$ production are extracted in the second fit. They are named $R_{\geq 2c}$, R_{1c} , and $R_{\geq 1b}$ and can be calculated as

$$R_{\geq 2c} = \frac{\sigma_{t\bar{t} + \geq 2c}}{\sigma_{t\bar{t} + \text{jets}}} \quad (11.7)$$

$$R_{1c} = \frac{\sigma_{t\bar{t} + 1c}}{\sigma_{t\bar{t} + \text{jets}}} \quad (11.8)$$

$$R_{\geq 1b} = \frac{\sigma_{t\bar{t} + \geq 1b}}{\sigma_{t\bar{t} + \text{jets}}}, \quad (11.9)$$

with their norm factors being calculated as

$$\mu(R_{\geq 2c}) = \frac{\mu(t\bar{t} + \geq 2c)}{\mu(t\bar{t} + \text{jets})} \quad (11.10)$$

$$\mu(R_{1c}) = \frac{\mu(t\bar{t} + 1c)}{\mu(t\bar{t} + \text{jets})} \quad (11.11)$$

$$\mu(R_{\geq 1b}) = \frac{\mu(t\bar{t} + \geq 1b)}{\mu(t\bar{t} + \text{jets})}. \quad (11.12)$$

When comparing the results of theoretical predictions and experimental measurements of the total cross-section, all generated events have to be considered. However, only a small part of the generated events will be selected in the experiment due to acceptance effects of the detector, requiring extrapolation to compare theory and experiment. To reduce these extrapolation effects, a so-called fiducial phase space can be used. In this analysis, it is defined using the so-called particle level. Particle-level objects are created in simulated events from true particles that have a lifetime greater than 30 ps. Using these objects, a selection can be performed, emulating the selection that is applied to data recorded with the detector. From the events passing through the fiducial phase space selection, the predicted fiducial cross-section can be calculated, which is then measured in the fit. In the end, the measured fiducial cross-section can be extrapolated back to the full phase space. The advantage of measuring a fiducial cross-section is a potential reduction in the size of systematic uncertainties as no or only little extrapolation has to be performed.

11.3.1 Definition of the fiducial phase space

Electrons and muons have to originate from W boson decays or from intermediate tau decays. To account for possible bremsstrahlung

Table 11.2: Fiducial phase space selection

Leptons	Exactly one $ \eta < 2.5$ $p_T > 27 \text{ GeV}$
Leptons veto	No additional $ \eta < 2.5$ $p_T > 10 \text{ GeV}$
Jets	$N_{\text{jets}} \geq 5$ $ \eta < 2.5$ $p_T > 25 \text{ GeV}$ ≥ 2 with ghost matched b -hadron
Overlap removal	$\Delta R(e, \text{jet}) < 0.4 \rightarrow \text{remove } e$ $\Delta R(\mu, \text{jet}) < 0.4 \rightarrow \text{remove } \mu$

effects, the four-momentum of photons within a radius of $\Delta R = 0.1$ about their direction is added to the lepton's four-momentum. Photons from hadronic decays are not considered for this addition. Only leptons with $p_T > 10 \text{ GeV}$ and $|\eta| < 2.5$ are used. Particle-level jets are clustered with the anti- k_t algorithm as are reconstructed jets. The radius parameter is set to $\Delta R = 0.4$. Electrons, muons, and photons already used in the lepton definition are not considered for clustering. It is performed on the remaining stable particles. For the definition of the fiducial phase space, jets are required to have $p_T > 25 \text{ GeV}$ and $|\eta| < 2.5$, though particle-level jets with transverse momenta as low as 15 GeV will be used to classify events. Using a procedure called ghost-matching, the jets can be labelled as b - or c -jets. B - and c -hadrons typically decay before 30 ps and are therefore not considered in the regular clustering. They are, however, included to perform the ghost-matching. For this, the simulated b - and c -hadrons are assigned an infinitesimal transverse momentum without changing the angular coordinates. Therefore, they will not change the properties of the resulting jets. If a hadron is included in the clustering of the jet, it is matched to the jet. Only hadrons with at least $p_T > 5 \text{ GeV}$ are considered for labelling the particle-level jets as b - or c -jets. If a b -hadron is matched to a jet, it is labelled b -jet. If no b -hadron is matched to a jet but a c -hadron, it is labelled c -jet.

As for the reconstruction level, overlap removal has to be performed. Leptons that are within a cone of size $\Delta R = 0.4$ of a jet are removed.

For the definition of the fiducial phase space, exactly one lepton with $p_T > 27 \text{ GeV}$ with no additional leptons is required. Additionally, at least five jets have to be in the event, at least two of which have to be labelled as b -jet. An overview of the selection can be seen in Table 11.2.

11.3.2 Normalisation

The different samples used for systematic uncertainties will all have different selection efficiencies for the fiducial phase space, or, in other words, their predicted yields will be different. These differences should, however, not be included in the systematic uncertainties. Therefore, the predicted yield of all samples is determined separately for the $t\bar{t} + 1c$, $t\bar{t} + \geq 2c$, $t\bar{t} + \geq 1b$, and $t\bar{t} + \text{light}$ contributions. The fiducial yield of each $t\bar{t}$ modelling variation, so alternative samples or the nominal sample with a reweighting applied, is scaled to the predicted fiducial yield of the nominal sample. As a single norm factor is used for the $t\bar{t} + \geq 1b$ contribution, the combined $t\bar{t} + \geq 1b$ contribution is scaled to the same yield instead of separate scales for $t\bar{t} + 1b$ and $t\bar{t} + \geq 2b$. The nominal yields for the $t\bar{t} + 1c$, $t\bar{t} + \geq 2c$, and $t\bar{t} + \text{light}$ components are always taken from the nominal $t\bar{t}$ sample generated in the 5FS, whereas for the $t\bar{t} + \geq 1b$ component, the nominal $t\bar{t}b\bar{b}$ sample generated in the 4FS is used.

11.4 Truth-level classification of jets

As no dedicated samples exist for $t\bar{t} + 1c$, $t\bar{t} + \geq 2c$, $t\bar{t} + \text{light}$, $t\bar{t} + 1b$, and $t\bar{t} + \geq 2b$ production, each $t\bar{t} + \text{jets}$ event is classified based on a “HF classification”. The procedure is the same as detailed in Ref. [8, 209]. Particle-level jets with $p_T > 15$ GeV and $|\eta| < 2.5$ are considered for this classification. Jets that originate from a W boson decay or a top-quark decay are not considered. This matching is done by matching the hadrons, which are ghost-matched to the jets, to the partons. Jets with a parton that is a top-quark or a quark of the W boson decay are vetoed. The HF multiplicities of the remaining jets are used to iteratively assign the events to the single $t\bar{t}$ components:

1. Events with two or more b -jets are labelled $t\bar{t} + \geq 2b$.
2. Events with exactly one b -jet are labelled $t\bar{t} + 1b$.
3. Events with two or more c -jets and no b -jets are labelled $t\bar{t} + \geq 2c$.
4. Events with one c -jet and no b -jets are labelled $t\bar{t} + 1c$.
5. All remaining events, so events with no c -jets and no b -jets, are labelled $t\bar{t} + \text{light}$.

As there is the dedicated $t\bar{t}b\bar{b}$ sample generated with a 4FS PDF set, the $t\bar{t} + 1b$ and $t\bar{t} + \geq 2b$ events from this sample are used, while the remaining components, $t\bar{t} + 1c$, $t\bar{t} + \geq 2c$, and $t\bar{t} + \text{light}$, are taken from the inclusive $t\bar{t}$ sample.

11.5 bc -Tagger

In $t\bar{t} + 1c$ and $t\bar{t} + \geq 2c$ events, b - and c -jets are present. For this analysis, it is beneficial to tag both of these jets with high efficiency at the same time: b -jets to enrich the selection in $t\bar{t}$ events in general,

and c -jets to reduce the contribution of $t\bar{t} + \geq 1b$ and $t\bar{t} + \text{light}$ events. The standard flavour-tagging provided in ATLAS is, however, not well suited for this problem.

11.5.1 Standard flavour-tagging in ATLAS

As introduced in Section 4.3.3, flavour tagging in ATLAS is performed using the output of the DL1r NN. The task of the NN is multinomial classification, resulting in an output probability of being a light-flavour, c -, or b -jet. As most physics analyses are only interested in selecting b -jets, these three output scores are combined into one discriminant through a likelihood ratio

$$\mathcal{D}_{\text{DL1r}} = \log \left(\frac{p_b}{0.018 \cdot p_c + 0.982 \cdot p_u} \right), \quad (11.13)$$

where p_b , p_c , and p_u are the output probabilities for being a b -, a c -, or a light-flavour jet, respectively. WPs with inclusive b -jet efficiencies of 85%, 77%, 70%, and 60% are defined based on $\mathcal{D}_{\text{DL1r}}$. Commonly used is the pseudo-continuous WP. It uses the same efficiency cuts but treats them as bin edges, leading to a total of five bins. With the pseudo-continuous WP, it is possible to require jets in different tagging bins; for instance, one might require one jet to be in the tightest bin while an additional jet has to be at least in the second-highest bin.

This discriminant is not optimised for tagging c -jets. In fact, the output probability of being a c -jet is treated the same as the output probability of being a light-flavour jet while forming the discriminant, albeit with a lower weight. Therefore, these flavour tagging WPs are not ideal for the $t\bar{t} + \text{charm}$ analysis. Instead, a custom tagger called bc -Tagger is defined.

11.5.2 bc -Tagger definition

As the DL1r NN has output probabilities for all three jet flavours, no new training of a NN is required. The same outputs can be used for simultaneous b - and c -tagging by combining the probabilities differently. Instead of forming only a b -tagging discriminant, a second discriminant is formed, targeting c -jets as a signal. The two discriminants are therefore

$$\mathcal{D}_{b/c} = \log \left(\frac{p_{b/c}}{f_{b/c} \cdot p_{c/b} + (1 - f_{b/c}) \cdot p_u} \right), \quad (11.14)$$

where the $f_{b/c}$ are scaling factors between the backgrounds for the b - and c -discriminants, respectively.

Using these two discriminants, a two-dimensional plane is formed with the b -tagging discriminant on the y -axis, and the c -tagging discriminant on the x -axis. Figure 11.1 shows this plane together with densities of b -, c -, and light-flavour jets from a simulated top-quark pair production sample. For visualisation purposes, both discriminants are passed through a sigmoid function to have the axes

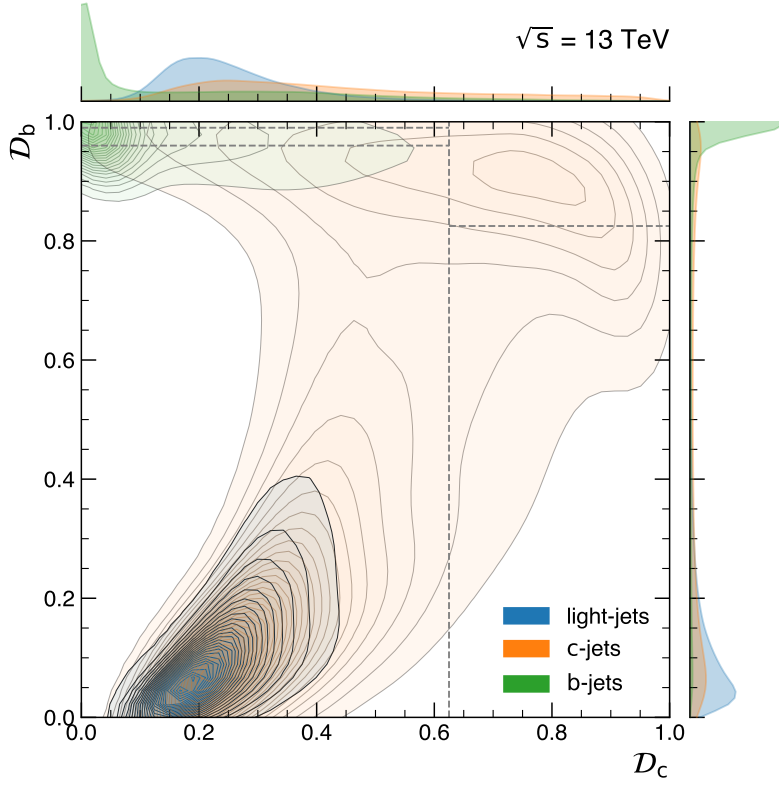


Figure 11.1: The two-dimensional bc -Tagger used in this analysis. The grey dashed lines show the bin edges of the five calibrated bins. Only areas with a density above a certain threshold value are plotted, such that empty areas do not correspond to areas without jets of the respective flavour.

bound by zero and one. The values for the scaling factors of the background are chosen to be $f_b = 0.018$ and $f_c = 0.4$. For simplicity reasons, the f_b is chosen to be the same as the standard DL1r WPs, so $\mathcal{D}_b = \mathcal{D}_{\text{DL1r}}$.

Similar to the standard DL1r WP, five bins are defined in this two-dimensional plane¹. The bins were chosen to have two bins targeting b -jets with inclusive efficiencies of 60% and 70% following the standard DL1r WPs. A vertical cut was chosen to obtain an inclusive c -jet efficiency of 22% for larger values of $\mathcal{D}_c$. This large c -jet bin is further split into two, using a cut on the b -tagging discriminant, splitting them into equal efficiencies of 11%.

In the analysis, four tagging requirements are utilised based on these five bins: two b -tagging WPs, $b@60\%$ and $b@70\%$, corresponding to the upmost left bin and the two upmost left bins in the two-dimensional plane, respectively, and two c -tagging WPs, $c@11\%$ and $c@22\%$. The $c@11\%$ WP covers the top right bin, whereas the $c@22\%$ WP covers both bins to the right side of the two-dimensional plane. Table 11.3 shows the efficiencies and rejections of these four WPs, as well as their cut-values on the b - and c -discriminant. To refer to the bin that is included in the $c@22\%$ WP but not in the $c@11\%$ WP, $c@22\%e$ will be used. Similarly, $b@70\%e$ will be used for the bin containing around 10% of the b -jets.

The calibration of the bc -Tagger and its results are detailed in Appendix C.

¹ The standard WP defines five bins along the y -axis only

	$b@60\%$	$b@70\%$	$c@11\%$	$c@22\%$
b/c -efficiency [%]	60.3	70.0	11.3	22.4
c/b -rejection	37.1	12.2	28.7	18.9
Light rejection	2321	573	1051	104
$\text{sigmoid}(\mathcal{D}_b)$	≥ 0.990	≥ 0.963	≥ 0.825	–
$\text{sigmoid}(\mathcal{D}_c)$	< 0.625	< 0.625	≥ 0.625	≥ 0.625

Table 11.3: The b - and c -jet efficiencies with their corresponding rejection factors for the five bins of the bc -Tagger. The cut values to separate the plane into the five bins are also shown. The ‘–’ denotes no cut being applied on this discriminant.

12 Event selection

Based on the introduced *bc*-Tagger, a selection to enrich the sample in $t\bar{t} + \geq 1c$ events can be performed. First, a preselection targetting top-quark pair decays in the lepton+jets channel with at least one additional jet is performed, before the *bc*-Tagger is used to define regions for the fit.

Contents

12.1	Categorisation of simulated events	127
12.2	Preselection	127
12.3	Fit regions	128

12.1 Categorisation of simulated events

Several smaller backgrounds are grouped into larger contributions meaning they are placed in the same histogram, also called channel. This has the practical advantage of not listing all individual processes but also benefits the fit process as all channels will have sufficient statistics in all bins. The $t\bar{t}$ +jets events are grouped based on the HF classification scheme described in Section 11.4. This results in five different processes: $t\bar{t} + 1b$, $t\bar{t} + \geq 2b$, $t\bar{t} + 1c$, $t\bar{t} + \geq 2c$, and $t\bar{t} + \text{light}$. The previously introduced signal strength $\mu(t\bar{t} + \geq 1b)$ scales both the $t\bar{t} + 1b$ and $t\bar{t} + \geq 2b$ components. The remaining processes are grouped into two contributions. Diboson and V +jets events are grouped together in “ VV , V +jets”, while all remaining processes are grouped under “Other top”. The latter consists of all single top production samples, the $t\bar{t}H$ sample, and the $t\bar{t}V$ samples.

Additionally, events with fake leptons can contribute to the selection. Fake leptons refer to objects reconstructed as leptons that are not prompt leptons but some background, for instance, a hadron. To estimate this potential contribution, all-hadronic $t\bar{t}$ events are used. The highest contribution of this sample was found to be 0.25% in SR_{loose} with at least six jets. In the remaining regions, the contribution was below 0.1%. Because of this low contribution, no fake lepton contribution is considered for this analysis.

12.2 Preselection

Before the final selection of fit regions, a preselection is performed. For this preselection, the *bc*-Tagger is not used. Instead, a selection on the DL1r WPs is performed. This has technical and historical reasons.

Section 4.3 introduced the definitions of loose and tight electrons and muons. Exactly one tight lepton has to be in the event, with an additional veto on loose leptons. At least five jets are required to have at least one jet not coming from the $t\bar{t}$ system. At least two of the jets have to be tagged with the DL1r $b@70\%$ WP. Additionally, at least three jets have to pass the DL1r $b@85\%$ WP. The requirement

of the third tagged jet greatly increases the $t\bar{t} + \geq 1c$ and $t\bar{t} + \geq 1b$ contributions compared to the $t\bar{t} + \text{light}$ contribution. The data is split into events with exactly five jets and events with at least six jets. All preselection requirements can be seen in Table 12.1.

Selection	e +jets	μ +jets
Leptons	Exactly one electron, no muon	Exactly one muon, no electron
	$ d_0 /\sigma(d_0) < 5$	$ d_0 /\sigma(d_0) < 3$
	$ \Delta z_0 \sin \theta < 0.5 \text{ mm}$	$ \Delta z_0 \sin \theta < 0.5 \text{ mm}$
	<i>LHTight</i> identification	<i>Medium</i> identification
	<i>Tight</i> isolation	<i>TightTrackOnly</i> isolation
	$ \eta < 1.37$ or $1.52 < \eta < 2.47$	$ \eta < 2.5$
	$p_T > 27 \text{ GeV}$	$p_T > 27 \text{ GeV}$
Lepton veto	No additional electron or muon	
	Electrons with:	Muons with:
	$ d_0 /\sigma(d_0) < 5$	$ d_0 /\sigma(d_0) < 3$
	$ \Delta z_0 \sin \theta < 0.5 \text{ mm}$	$ \Delta z_0 \sin \theta < 0.5 \text{ mm}$
	<i>LHMedium</i> identification	<i>Loose</i> identification
	$ \eta < 1.37$ or $1.52 < \eta < 2.47$	$ \eta < 2.5$
	$p_T > 10 \text{ GeV}$	$p_T > 10 \text{ GeV}$
Jets	$N_{\text{jets}} \geq 5$	
	$ \eta < 2.5$	
	$p_T > 25 \text{ GeV}$	
	JVT cut if $p_T < 60 \text{ GeV}$ and $ \eta < 2.4$	
	≥ 2 b-tagged jets with DL1r@70%	
	≥ 3 b-tagged jets with DL1r@85%	
Overlap Removal	If an electron and a muon share a track: electron removed	
	$\Delta R_y(e, \text{jet}) < 0.2$: jet removed	
	then $\Delta R_y(e, \text{jet}) < 0.4$: electron removed	
	$\Delta R_y(\mu, \text{jet}) < 0.4$ and $n_{\text{tracks}}^{\text{jet}} \geq 3$: μ removed	
	$\Delta R_y(\mu, \text{jet}) < 0.4$ and $n_{\text{tracks}}^{\text{jet}} < 3$: jet removed	

Table 12.1: Summary of the requirements on detector-level objects to pass the preselection.

Figure 12.1 shows the multiplicities of the different bc -Tagger tagging bins after the preselection. It can be clearly seen that with rising b -jet multiplicity, the contribution of $t\bar{t} + 1b$ and $t\bar{t} + \geq 2b$ becomes dominant, whereas with rising c -jet multiplicity, the contribution of $t\bar{t} + 1c$ and $t\bar{t} + \geq 2c$ increases. Events with zero jets in the bc -Tagger $b@70\%$ bin are possible as the bin definition for the bc -Tagger is tighter than the DL1r one. Meaning not every DL1r $b@70\%$ jet will be a bc -Tagger $b@70\%$ jet, but every bc -Tagger $b@70\%$ jet will be a DL1r $b@70\%$ jet.

12.3 Fit regions

Based on the tagged jet multiplicities of the bc -Tagger, fit regions can be defined. This selection has to solely depend on the bc -Tagger WPs and cannot include any DL1r WPs. Otherwise, a model for the

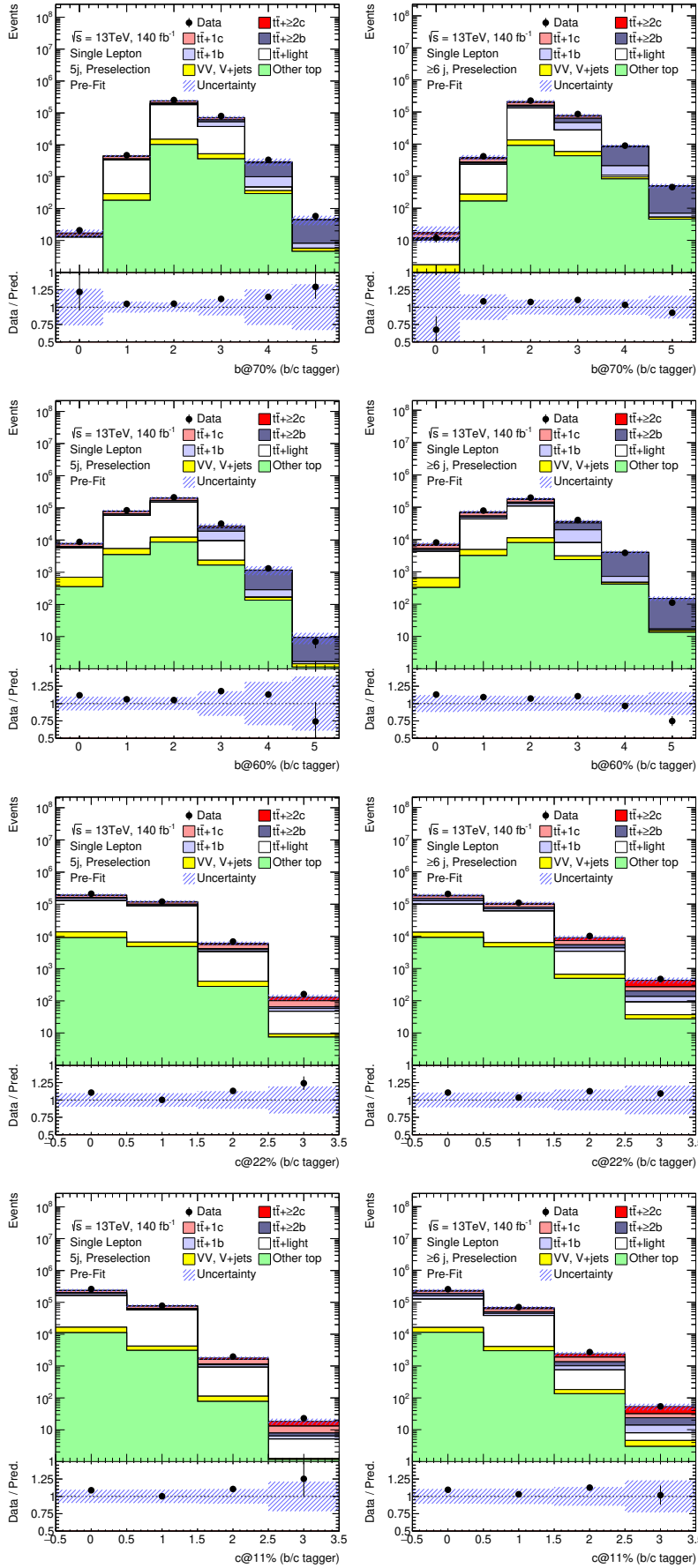


Figure 12.1: Data-MC agreement for tagged jets multiplicities after the pre-selection. Events with (left) five jets and (right) at least six jets are shown separately. The hashed bands show the uncertainty originating from finite statistics in the MC modelling and systematic uncertainties. The first and last bin contain underflow and overflow events, respectively.

combination of the uncertainties of these two taggers would need to be developed. Therefore, the selection has to be tighter than the preselection purely using bc -Tagger multiplicities. As two $b@70\%$ jets for DL1r are required in the preselection, at least two $b@70\%$ jets for the bc -Tagger need to be required for each region. To take care of the third tagged jet at $b@85\%$, two options exist for the bc -Tagger. Either an additional $b@70\%$ jet needs to be required or a $c@11\%$ jet as the threshold on the b -discriminant for the $c@11\%$ tagging bin is higher than for the $b@85\%$ tagging bin of DL1r.

Following these restrictions, three control regions and two signal regions are defined, each further separated into events with five jets and events with at least six jets, leading to ten regions in total. The selection requirements can be seen in Table 12.2. The first control region (CR_1) requires two $b@70\%$ jets and a $c@11\%$ jet with a veto on additional $c@22\%$ jets. The second control region (CR_2) requires at least three $b@60\%$ jets with no additional c -tagged jets. The last control region (CR_3) requires exactly three $b@60\%$ jets and a $c@11\%$ jet. Again, additional $c@22\%$ jets are vetoed. The signal regions all require two $b@70\%$ jets and at least two additional c -tagged jets. The first signal region (SR_{loose}) requires exactly one of these jets to be a $c@11\%$ jet, while the second signal region (SR_{tight}) requires at least two $c@11\%$ jets. No veto on additional $c@22\%$ jets is used for SR_{tight} .

Table 12.2: Definition of the fit regions based on the bc -Tagger working points. A dash refers to no requirement on this category of jets. $b@70\%$ and $c@22\%$ are inclusive of their respective tighter working points. A number with no additional “ $\geq$ ” implies an exact requirement, e.g. $c@22\% = 1, c@11\% = 1$ vetoes additional $c@22\%$ jets. The selection criteria are identical for the regions with exactly five jets and regions with at least six jets.

Region	$b@70\%$	$b@60\%$	$c@22\%$	$c@11\%$
CR_1	2	—	1	1
CR_2	—	≥ 3	0	—
CR_3	—	3	1	1
SR_{loose}	2	—	≥ 2	1
SR_{tight}	2	—	—	≥ 2

Figure 12.2 shows the expected background composition in the ten fit regions. The background in CR_1 is dominated by $t\bar{t}$ + light. Most of this $t\bar{t}$ + light contribution has an actual c -jet from the W boson decay tagged as $c@11\%$. CR_2 has a mixture of $t\bar{t}$ + light and $t\bar{t}$ + $\geq 1b$ as background, while CR_3 has mostly $t\bar{t}$ + $\geq 1b$ as background. Both signal regions have a mixture of $t\bar{t}$ + light and $t\bar{t}$ + $\geq 1b$ with varying fractions of $t\bar{t}$ + light and $t\bar{t}$ + $\geq 1b$. SR_{tight} has more $t\bar{t}$ + $\geq 1b$ background, as expected. As both c -tagged jets have to be tagged as $c@11\%$, it is much less likely that light-flavour jets are mistagged. But more b -jets will be mistagged, leading to a relative increase in the $t\bar{t}$ + $\geq 1b$ background.

It can also be seen that the $t\bar{t}$ + $\geq 1b$ background has a larger contribution to regions with at least six jets compared to regions with five jets. For the $t\bar{t}$ + light background, the opposite behaviour can be observed. As $t\bar{t}$ + light includes every $t\bar{t}$ event that is not classified

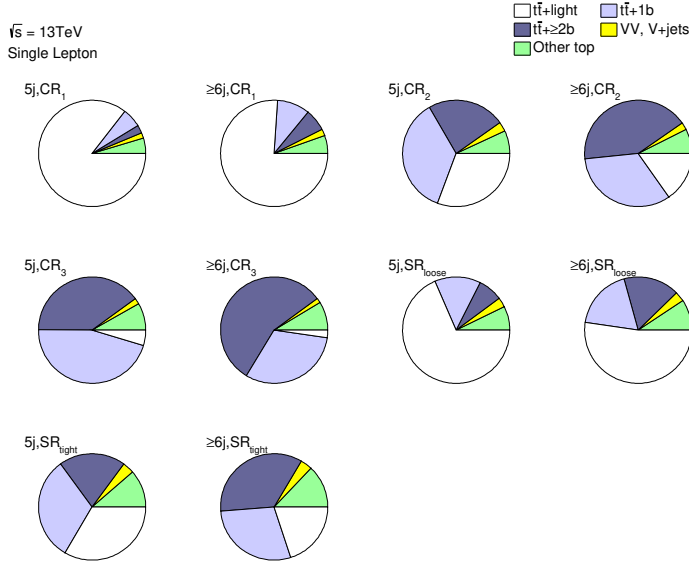


Figure 12.2: Expected background composition in the ten fit regions. The pie charts show the fractional contributions of the single backgrounds to the total background. Everything apart from $t\bar{t} + 1c$ and $t\bar{t} + \geq 2c$ is considered as background.

as $t\bar{t} + \geq 1c$ or $t\bar{t} + \geq 1b$, this includes all events where there are only four jets from the $t\bar{t}$ system. Therefore, the distribution of the jet multiplicity will tend to have lower values than the distributions for the $t\bar{t} + \geq 1c$ and $t\bar{t} + \geq 1b$ components. For both of these, at least one, but more often than not, two additional jets are present. The same effect can be seen when comparing the relative contribution of $t\bar{t} + 1b$ to $t\bar{t} + \geq 2b$. The relative contribution of $t\bar{t} + \geq 2b$ is larger in regions with a higher jet multiplicity.

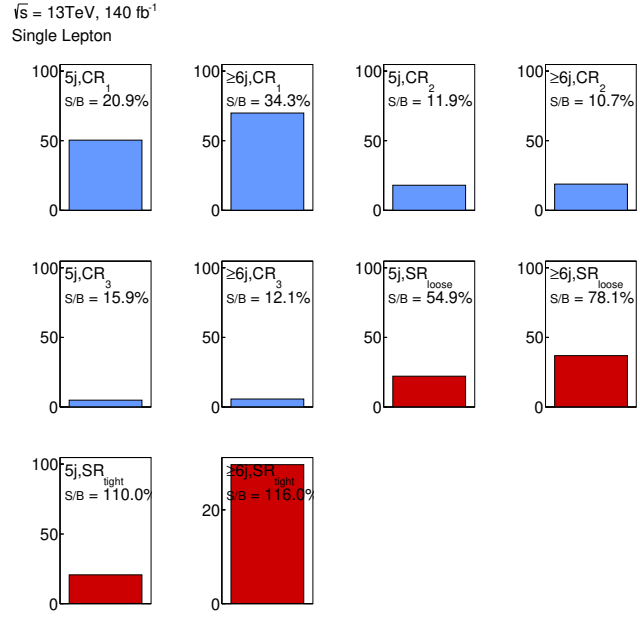
Figure 12.3 shows the fraction of signal compared to background in the ten regions with signal being defined as $t\bar{t} + \geq 2c$ and $t\bar{t} + 1c$. The relative contribution of $t\bar{t} + \geq 1c$ to the signal regions is larger in regions with exactly six jets compared to regions with five jets. This is similar to the behaviour of the $t\bar{t} + \geq 1b$ background. The values of S/B in the control regions range from 10.7% up to 34.3%, while in the signal regions they range from 54.9% up to 116.0%. SR_{tight} with at least six jets has the highest value.

The fit observables can be seen in Figures 12.4 and 12.5. For the control regions, only the yield is used for the fit to reduce potential constraints in the fit. Especially CR_1 with its large statistics, can lead to large constraints. The two signal regions use the same fit observables, though different ones are used for the regions with five jets and at least six jets. In regions with at least six jets, the jet multiplicity is used. For the two regions with five jets, the invariant mass of the closest¹ pair of jets that are c -tagged is used. For SR_{tight} , only jets tagged as $c@11\%$ are considered, as only these are considered for the selection. This variable has good discriminating power between the $t\bar{t} + \geq 1c$ signal and the $t\bar{t} + \text{light}$ background, especially in SR_{loose} .

Most of the $t\bar{t} + \text{light}$ events include a true c -jet originating from the W boson decay. This jet is also often tagged as one of the c -jets. For the $t\bar{t} + \text{light}$ background, the other jet from the W boson decay

¹ Proximity is measured by ΔR .

Figure 12.3: Expected signal contribution in the ten fit regions. The y -axes of the plots show the signal significance $S/\sqrt{B}$, the quoted values in the plots the signal-to-background ratio S/B . S (B) is the expected number of signal (background) events. $t\bar{t} + 1c$ and $t\bar{t} + \geq 2c$ are considered as signal.



is often tagged as the second c -jet, leading the invariant mass of the two c -tagged jets to be located around the W boson mass. As there is an additional true c -jet for $t\bar{t} + \geq 1c$ events, this occurs less often for these events. Due to the better light-flavour rejection of the $c@11\%$ tagging bin compared to the $c@22\%$ tagging bin, this effect is reduced for SR_{tight} . This feature is also reduced when going to regions with at least six jets. As there are now at least three other jets apart from the two b -jets and the c -jet from the W boson decay, the probability that the second jet from the W boson decay is mistagged is much reduced. Therefore, this is not used as a fit observable in regions with at least six jets.

No large shape differences between MC and data can be seen in the signal regions, but a normalisation difference can. The prediction is always too low for the observed data. This is also the case for CR_2 and CR_3 , which have a large fraction of $t\bar{t} + \geq 1b$ events. CR_1 on the other hand, has a very good agreement of yields between MC and data. CR_1^{5j} is the only region where slightly more events are predicted than observed. This observation is in agreement with previous measurements from the ATLAS experiment. Previous analyses, such as in Ref. [6], see a similar underprediction of MC for events with top-quark pairs with additional b -jets. Ref. [8] shows a similar result for the $t\bar{t} + \geq 1b$ background, but in this analysis, the post-fit yields of the $t\bar{t} + \geq 1c$ contribution are also considerably larger than the pre-fit yields.

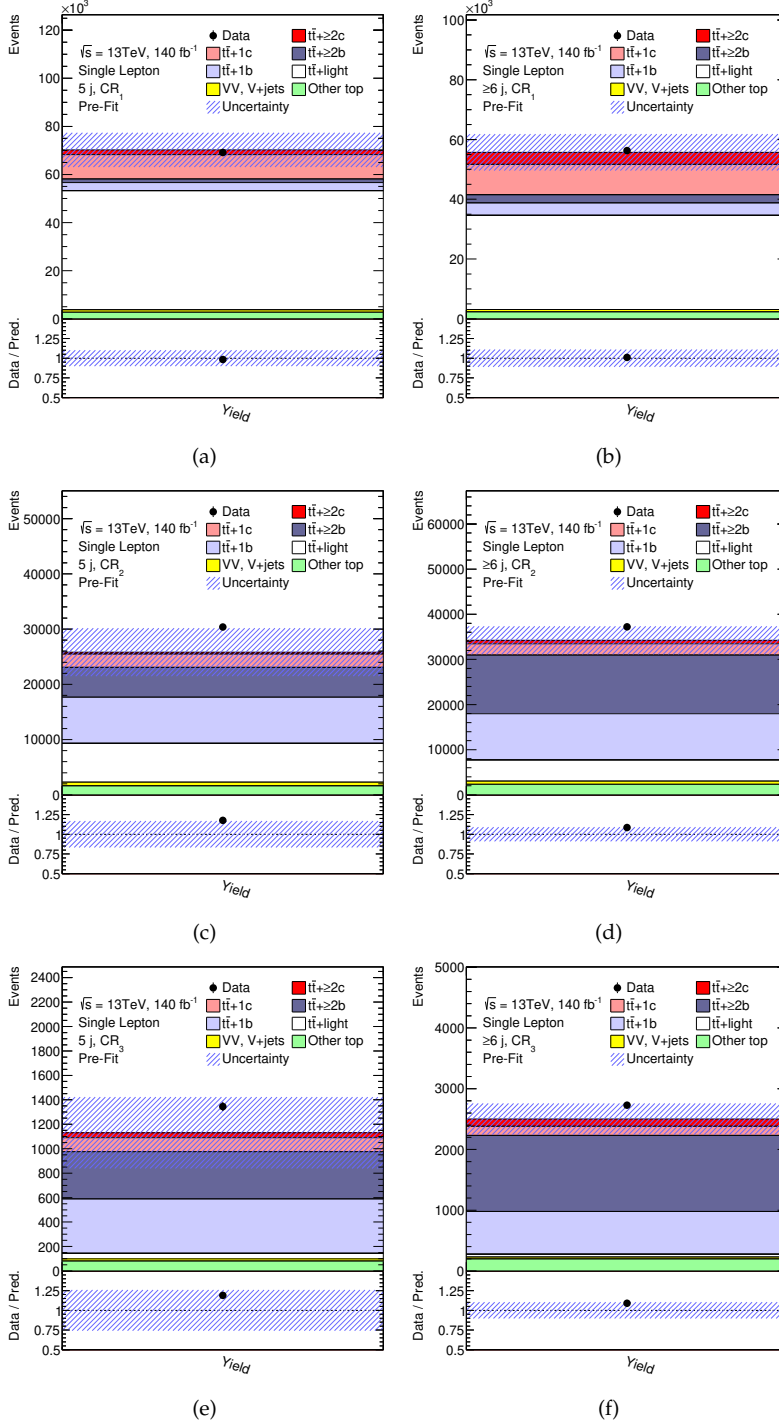


Figure 12.4: Yields used in the fit for the three control regions. All regions are split into (a,c,e) events with exactly five jets and (b,d,f) events with at least six jets. (a,b) CR₁, (c,d) CR₂, and (e,f) CR₃ are shown. The hashed bands show the uncertainty originating from finite statistics in the MC modelling and systematic uncertainties.

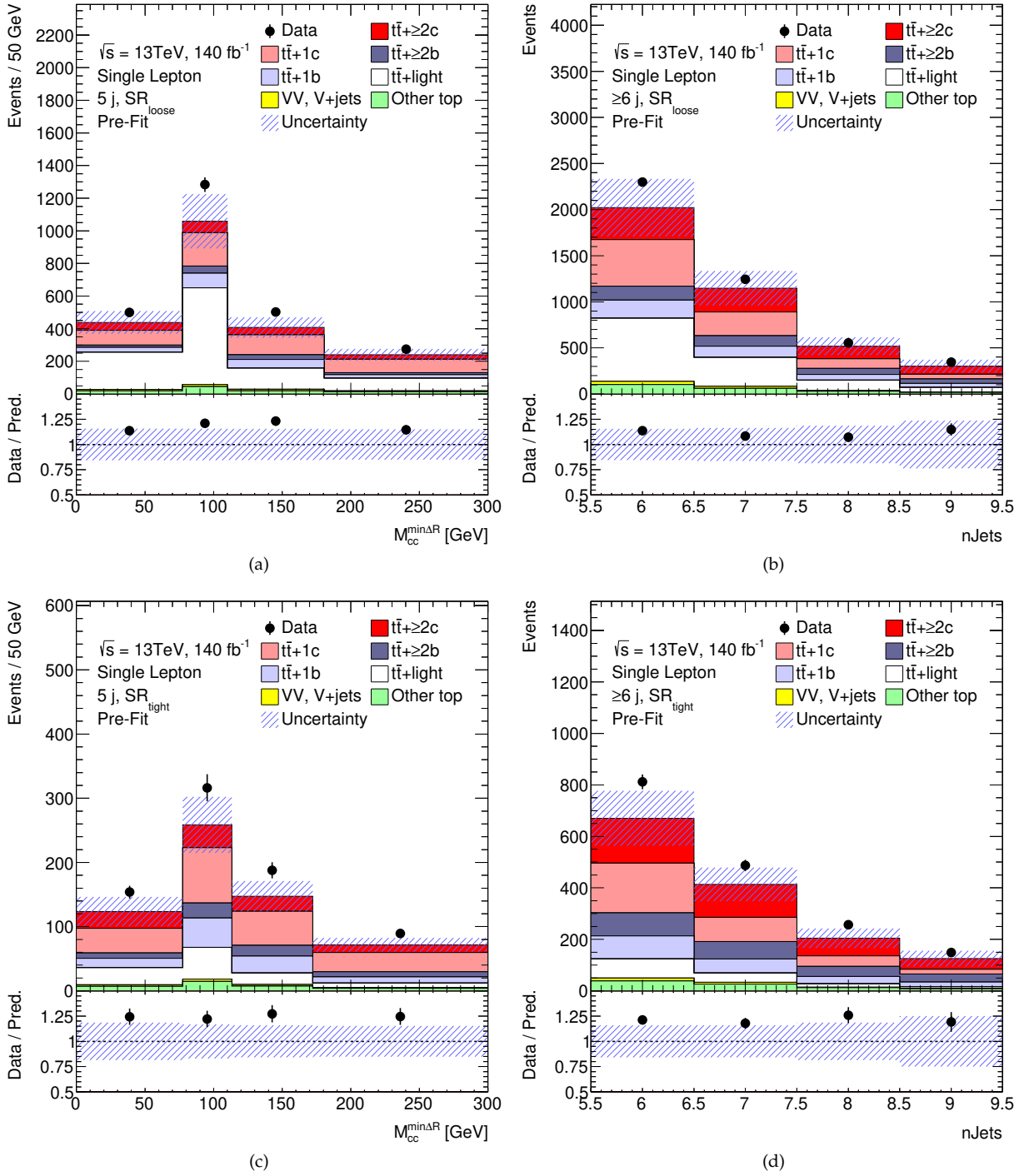


Figure 12.5: Observables used in the fit for the two signal regions. Both regions are split into (a,c) events with exactly five jets and (b,d) events with at least six jets. For both regions, (a,b) SR_{loose} and (c,d) SR_{tight} , the regions with five jets show the invariant mass of the two c -tagged jets with the lowest ΔR distance in the event. For SR_{tight} only tight c -tagged jets are considered. The regions with at least six jets show the jet multiplicity. The hashed bands show the uncertainty originating from finite statistics in the MC modelling and systematic uncertainties. The first and last bin contain underflow and overflow events, respectively.

13 Systematic uncertainties

Several sources of uncertainties have to be considered for the $t\bar{t}$ + charm cross-section measurement. Not only the statistical uncertainty due to the limited number of events at the LHC has to be considered but also systematic uncertainties. This chapter will first introduce how these uncertainties are derived before detailing all uncertainties considered.

Contents

13.1	Templates in the profile likelihood fit	135
13.1.1	Smoothing	135
13.1.2	Symmetrisation	136
13.2	Experimental uncertainties	136
13.3	Modelling uncertainties	137
13.3.1	$t\bar{t}$ modelling	138
13.3.2	$t\bar{t}b\bar{b}$ modelling	140
13.3.3	Background modelling	140
13.4	Pruning	142

Systematic uncertainties include all uncertainties that are not statistical uncertainties of the recorded data. They can be separated into experimental and modelling uncertainties.

Experimental uncertainties range from uncertainties related to the efficiencies of certain selections to the calibration of particle energies or the calibration of the luminosity. As they are uncertainties originating from the experimental setup, they act on all processes at the same time, resulting in one NP acting on all processes.

Modelling uncertainties contain all uncertainties with respect to the modelling of the signal and background processes. These can be simple uncertainties in the calculated cross-section or choices of certain parameters like the factorisation scale.

Lastly, the uncertainty due to the available statistics of simulated samples is also taken into account using γ parameters.

13.1 Templates in the profile likelihood fit

To include systematic uncertainties, the template method as described in Section 11.1 is used. All variations related to $t\bar{t}$ modelling uncertainties are normalised to the same fiducial cross-section as the nominal prediction, such that the difference only arises from selection efficiencies or migration effects between the single regions.

13.1.1 Smoothing

As some variations can have large statistical fluctuations, the obtained templates can be smoothed. Two different smoothing algorithms are used: The default algorithm, *MaxVariation*, limits the number of slope variations in the ratio of the systematic variation to the nominal template. Neighbouring bins are merged until the statistical

uncertainty of the merged bins is below a pre-set tolerance value. If the number of slope variations is greater than a given threshold, the tolerance is halved. This process is repeated iteratively until the number of slope variations reaches the threshold or lower. After the merging of bins, the "353QH twice" algorithm [210] is applied to the ratio. This method averages bin contents based on the integral of the histogram and neighbouring bin information. For some systematic variations, *parabolic smoothing* is used instead. It only merges bins and does not perform any dedicated smoothing. Bins are merged until the relative variation has up to three maxima, where maxima are defined as bins that are different from the neighbouring ones. Therefore, this rebinning can at most result in a parabolic shape.

13.1.2 Symmetrisation

As already mentioned, a symmetrisation process has to be applied to two-point systematics in order to obtain an up- and a down-variation. However, even three-point systematics can have asymmetric variations. With symmetrisation applied, the two variations can be centred around the nominal prediction.

For three-point systematics, two-sided symmetrisation can be applied. The two absolute variations are divided by the nominal histogram to obtain the relative variations, and then their difference is divided by two. The resulting value is taken as new relative up and relative down variation

$$\text{new relative up/down} = \pm \left| \frac{\text{absolute up} - \text{absolute down}}{2 \times \text{nominal}} \right|. \quad (13.1)$$

Therefore, the variations are centred around the nominal prediction after the symmetrisation. If the variations are symmetric before the symmetrisation is applied, the variations will not be changed by construction.

For two-point systematics, one-sided symmetrisation is applied. It simply mirrors the provided variation on the other side of the nominal prediction.

13.2 Experimental uncertainties

Several uncertainties originating from the experimental setup are considered in the analysis. All experimental uncertainties are provided with an up- and a down-variation.

FLAVOUR TAGGING The calibration of the flavour tagging SFs with its uncertainties is described in Appendix C. The resulting uncertainties are decorrelated using a principal component analysis, leading to a total of 85 NP, 45 for b -jets and 20 for light flavour and c -jets each. Figure 13.1 shows an example of the obtained templates for one NP in SR_{loose} for the $t\bar{t}$ + light sample.

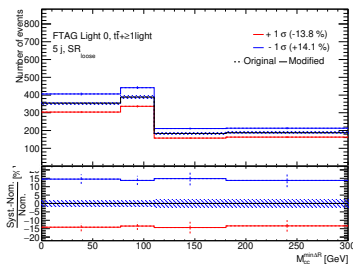


Figure 13.1: Systematic templates of one flavour tagging NP. No smoothing or symmetrisation is applied.

JETS Several types of uncertainties are considered for jets. The first one is related to the selection based on the JVT. The SFs applied to the events are varied within their uncertainties to obtain the varied templates.

The second type of uncertainty is related to the jet energy scale (JES). For the calibration of the JES, test-beam data, LHC collision data, and simulation have been used. Different SM processes are used. As there are considerable differences between different PS algorithms, MC-MC SFs are applied to correct the JES of different samples. In total, 35 NP are assigned to the JES. Figure 13.2 shows one of these NP.

The last type of uncertainty is related to the jet energy resolution (JER). Using an eigenvector decomposition, the number of NP for JER can be reduced to twelve. Instead of just smearing MC, additional pseudo-data is used for regions in which data has a better resolution than MC. In the end, a total of thirteen NP are used, as an additional NP comparing data and MC is added.

CHARGED LEPTONS Two different types of uncertainties exist for charged leptons. The first relates to the uncertainties of their SFs for reconstruction, identification, isolation, and triggers. The SFs are measured using tag-and-probe methods using $Z \rightarrow \ell\ell$ and $J/\psi \rightarrow \ell\ell$ decays. A more detailed description of the methodology for identification SFs for electrons is given in Appendix B. The uncertainties on the SFs are included by constructing templates while varying the values of the SFs within their uncertainties.

The other uncertainties affect the energy and momentum of the electrons and muons. Here, no SFs are varied, but the values of the energy and momentum themselves. They are varied within the uncertainties obtained from the calibration of energy and momentum. An uncertainty model where single NP are merged is used.

PILEUP As mentioned in Section 5.5, weights are applied to each event to represent the observed pileup profile. Their uncertainty is propagated to the fit by varying them within their uncertainties.

LUMINOSITY A dedicated measurement of the luminosity yields a result for the integrated luminosity of $140.1 \pm 1.2 \text{ fb}^{-1}$, corresponding to a relative uncertainty of 0.83% [51]. This uncertainty is applied to all processes in all regions. It only has a normalisation component and no shape contribution.

13.3 Modelling uncertainties

Three types of modelling uncertainties will be used throughout this work. The first one is a normalisation uncertainty. This type can be used, for instance, for uncertainties in the theoretically calculated cross-section. The second one is based on internal weight variations of the respective sample, which can be applied to emulate a difference in

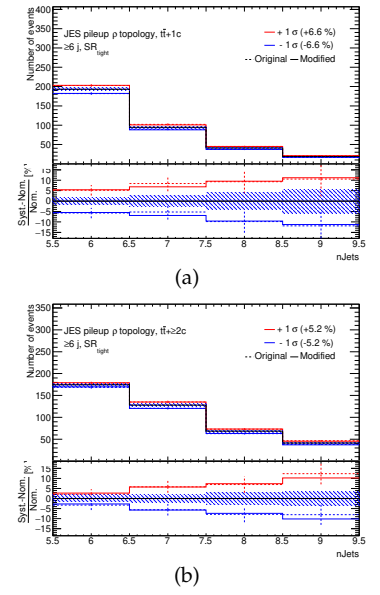


Figure 13.2: Systematic templates of the JES ρ topology uncertainty for the (a) $t\bar{t} + 1c$ and (b) $t\bar{t} + \geq 2c$ samples. Two-sided symmetrisation is used.

the generation process. For instance, the resulting distribution might represent what happens if the renormalisation scale is doubled. Lastly, alternative samples can be used. Typically, the difference between the nominal and alternative samples is used as the systematic uncertainty. All uncertainties using an alternative sample in this work are two-point systematics, as only one alternative sample is available. All two-point systematics are symmetrised and smoothed. If the alternative sample was only produced with the fast detector simulation, it will be compared to the nominal sample generated with the fast detector simulation as well, so as not to increase or decrease the uncertainties due to differences in the fast and full detector simulation.

Ideally, modelling uncertainties should only vary one source of possible uncertainties at a time. For instance, a comparison between a MADGRAPH5_AMC@NLO + PYTHIA and a SHERPA sample is not ideal as multiple parameters are changed at once. A different ME generator and a different PS algorithm are used.

13.3.1 $t\bar{t}$ modelling

The three components $t\bar{t} + \text{light}$, $t\bar{t} + 1c$, and $t\bar{t} + \geq 2c$ are all taken from the same simulated samples. Therefore, the same modelling uncertainties are applied to all of them. By default, the NP describing the uncertainties for the $t\bar{t} + 1c$ and $t\bar{t} + \geq 2c$ components are treated as completely correlated. Another way of thinking about this is that the same set of NPs is used for both of them. For some uncertainties, the $t\bar{t} + 1c$ and $t\bar{t} + \geq 2c$ components will be decorrelated, leading to two separate sets of NPs. The $t\bar{t} + \text{light}$ component is always treated separately from the two other components. Table 13.1 shows an overview of all uncertainties associated with these three components.

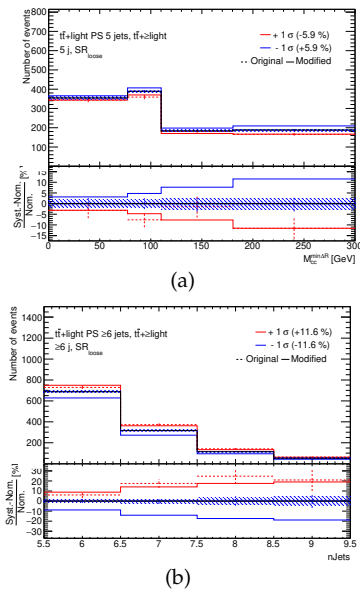


Figure 13.3: Systematic templates of the PS uncertainty for the $t\bar{t} + \text{light}$ sample in SR_{loose} . While the normalisation effect of the alternative sample in (a) five jet regions is positive, it is negative in (b) regions with at least six jets. Smoothing and one-sided symmetrisation are applied.

PS AND HADRONISATION To estimate the impact of the choice of the PS and hadronisation algorithm, the alternative sample produced with POWHEG Box + HERWIG7 is used. This also includes a comparison of the choice of the hadronisation model, as PYTHIA uses a Lund string model, whereas HERWIG uses a cluster-based model. The $t\bar{t} + 1c$ and $t\bar{t} + \geq 2c$ components are treated as decorrelated pre-fit to reduce observed constraints. Similarly, two NP are introduced for the $t\bar{t} + \text{light}$ component, one acting on all regions with exactly five jets and one for all regions with at least six jets. As the normalisation effect of the variation is opposite in these two sets of regions, using a single NP leads to a large constraint. This difference in the normalisation effect can be seen in Figure 13.3. By decorrelating these regions, the constraint can be reduced.

NLO MATCHING The POWHEG Box + PYTHIA sample with the p_T^{hard} parameter set to one instead of zero is used as an alternative sample to estimate an uncertainty related to the NLO matching. This parameter is an internal PYTHIA parameter. As for the PS, the $t\bar{t} + 1c$ and $t\bar{t} + \geq 2c$ components are treated as decorrelated pre-fit.

Source	Description	Component
PS & hadronisation	POWHEG BOX + PYTHIA8 vs POWHEG BOX + HERWIG7	$t\bar{t}$ + light 5j, $t\bar{t}$ + light ≥ 6 j, $t\bar{t}$ + 1c, $t\bar{t}$ + ≥ 2 c
NLO matching	varied p_T^{hard} parameter in POWHEG BOX + PYTHIA8	$t\bar{t}$ + light, $t\bar{t}$ + 1c, $t\bar{t}$ + ≥ 2 c
h_{damp}	varied h_{damp} parameter in POWHEG BOX + PYTHIA8	$t\bar{t}$ + light, $t\bar{t}$ + ≥ 1 c
FSR	$\mu_R^{\text{FSR}} \times 2, \mu_R^{\text{FSR}} \times 0.625$ in POWHEG BOX + PYTHIA8 (internal weights)	$t\bar{t}$ + light, $t\bar{t}$ + ≥ 1 c
ISR	Variation of A14 tune (internal weights)	$t\bar{t}$ + light, $t\bar{t}$ + ≥ 1 c
Renormalisation scale	$\mu_R \times 2, \mu_R \times 0.5$ in POWHEG BOX + PYTHIA8 (internal weights)	$t\bar{t}$ + light, $t\bar{t}$ + ≥ 1 c
Factorisation scale	$\mu_F \times 2, \mu_F \times 0.5$ in POWHEG BOX + PYTHIA8 (internal weights)	$t\bar{t}$ + light, $t\bar{t}$ + ≥ 1 c
PDF	PDF4LHC15 error set (internal weights)	$t\bar{t}$ + light, $t\bar{t}$ + ≥ 1 c

h_{damp} VARIATION The POWHEG BOX + PYTHIA sample with the h_{damp} parameter set to three times the top mass is compared to the nominal sample for which the parameter is set to 1.5 times the top mass. The h_{damp} parameter is an internal parameter of POWHEG BOX. It regulates the threshold of the hardest emission.

FSR An internal reweighting of the nominal sample is used to estimate an uncertainty related to FSR modelling. The nominal value of the renormalisation scale factorised for FSR is multiplied by a value of 2.0 or 0.625 for the up- and down-variation, respectively. The choice of 0.625 instead of 0.5 as the down variation was taken as the 0.625 variation provides a more symmetric variation with respect to the 2.0 variation. Additionally, some trends that are consistent between the 0.875, 0.75, and 0.625 variations change when going to the 0.5 variation. Lastly, the 0.5 variation always has some bins that deviate from the remaining bins. This is most likely caused by a tail towards large weights, which results in large statistical uncertainties. Using the 0.625 variation, the obtained templates are more consistent with each other. Two resulting templates can be seen in Figure 13.4.

ISR The Var3c up- and down-variation of the A14 set of tuned parameters is used to obtain uncertainties related to the modelling of ISR. This variation roughly correlates to changes in the strong coupling for ISR. Internal weights of the nominal sample are used for this variation.

SCALE CHOICES The choice of the renormalisation and factorisation scale is separately varied by a factor of 2.0 and 0.5. This leads to two different NP, one for the renormalisation scale and one for the factorisation scale. As for the changes in the renormalisation scale for FSR, these variations are obtained using an internal reweighting.

PDF The choice of the used PDF is evaluated using the PDF4LHC15 error set. It consists of 30 variations, each representing one NP. These are the only variations that are obtained using an internal reweighting, but only one variation is given. Therefore, all 30 variations are symmetrised using one-sided symmetrisation. Additionally, they are not compared to the actual nominal prediction but to a nominal prediction of the PDF4LHC15 error set.

Table 13.1: Overview of the uncertainties associated to the $t\bar{t}$ + light and $t\bar{t}$ + ≥ 1 c process. If $t\bar{t}$ + 1c and $t\bar{t}$ + ≥ 2 c are listed in the component column, two separate NPs are used for these components. If only $t\bar{t}$ + ≥ 1 c is listed, the same NP is used for $t\bar{t}$ + 1c and $t\bar{t}$ + ≥ 2 c. The $t\bar{t}$ + light component always has a different NP than the $t\bar{t}$ + ≥ 1 c components. For the PS, the $t\bar{t}$ + light component uses two NPs, one for regions with five jets and one for regions with at least six jets.

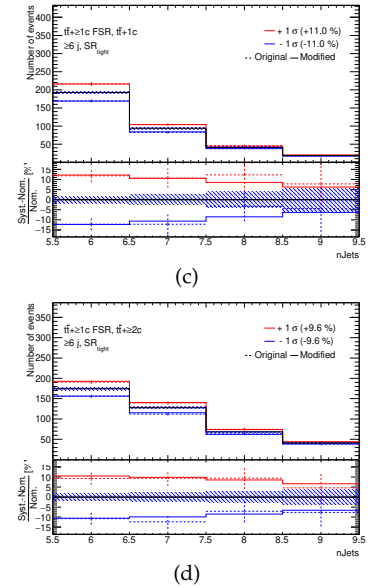


Figure 13.4: Systematic templates of the FSR uncertainty for the (a) $t\bar{t}$ + 1c and (b) $t\bar{t}$ + ≥ 2 c samples. Smoothing and two-sided symmetrisation are applied.

13.3.2 $t\bar{t}b\bar{b}$ modelling

Table 13.2 shows an overview of the systematic uncertainties associated with the $t\bar{t} + \geq 1b$ process. The same internal weight variations exist in the dedicated 4FS $t\bar{t}b\bar{b}$ sample as in the inclusive 5FS $t\bar{t}$ sample. Therefore, several uncertainties are the same as for $t\bar{t}$ modelling, and their description will not be repeated here. This includes the FSR, ISR, renormalisation scale, factorisation scale, and PDF uncertainties. Additionally, the PS and p_T^{hard} uncertainties are obtained using the same variations. For these two uncertainties, the $t\bar{t} + 1b$ and $t\bar{t} + \geq 2b$ components are treated as decorrelated pre-fit, just as the $t\bar{t} + 1c$ and $t\bar{t} + \geq 2c$ components.

Source	Description	Component
PS & hadronisation	POWHEG BOX RES + PYTHIA8 vs POWHEG BOX RES + HERWIG7	$t\bar{t} + 1b, t\bar{t} + \geq 2b$
NLO matching	varied p_T^{hard} parameter in POWHEG BOX RES + PYTHIA8	$t\bar{t} + 1b, t\bar{t} + \geq 2b$
Recoil scheme	varied recoil scheme in POWHEG BOX RES + PYTHIA8	$t\bar{t} + \geq 1b$
FSR	$\mu_R^{\text{FSR}} \times 2, \mu_R^{\text{FSR}} \times 0.625$ in POWHEG BOX RES + PYTHIA8 (internal weights)	$t\bar{t} + \geq 1b$
ISR	Variation of A14 tune (internal weights)	$t\bar{t} + \geq 1b$
Renormalisation scale	$\mu_R \times 2, \mu_R \times 0.5$ in POWHEG BOX RES + PYTHIA8 (internal weights)	$t\bar{t} + \geq 1b$
Factorisation scale	$\mu_F \times 2, \mu_F \times 0.5$ in POWHEG BOX RES + PYTHIA8 (internal weights)	$t\bar{t} + \geq 1b$
PDF	PDF4LHC15 error set (internal weights)	$t\bar{t} + \geq 1b$

Table 13.2: Overview of the uncertainties associated to the $t\bar{t} + \geq 1b$ process. If $t\bar{t} + 1b$ and $t\bar{t} + \geq 2b$ are listed in the component column, two separate NPs are used for these components. If only $t\bar{t} + \geq 1b$ is listed, one NP is used for this systematic variation.

RECOIL SCHEME To estimate the effect of the choice of the recoil scheme, an alternative sample that uses the dipole recoil scheme is used. The nominal sample uses the global recoil scheme instead. As the alternative sample was generated with the h_{bzd} parameter set to two, the template is not obtained from the difference between the alternative and the nominal sample. Instead, the difference of the varied sample to a reference sample is used. The reference sample is the same as the nominal sample, but the h_{bzd} parameter was set to two during the generation. With this setup, only differences from the recoil scheme are probed.

13.3.3 Background modelling

The modelling of all remaining processes has a considerably lower impact on the measurement than the already discussed uncertainties. Table 13.3 summarises all uncertainties associated with these samples as well as their description.

SINGLE TOP The choice of the strategy of the overlap removal strategy between tW at NLO and $t\bar{t}$ at LO is included as a systematic uncertainty. The nominal setup uses the sample generated with the DR scheme, while for the systematic variation, the sample generated with the DS scheme is used.

For the three single top samples generating tW , the s - and t -channel, the PS algorithm uncertainty is assessed in the same way as for the $t\bar{t}$ and $t\bar{t}b\bar{b}$ processes. The nominal sample is replaced by the sample generated with POWHEG BOX + HERWIG7 resulting in three NP.

Process	Source	Description
single top production only s - and t -channel	PS & hadronisation norm uncertainty	POWHEG BOX + PYTHIA8 vs POWHEG BOX + HERWIG7 $\pm 50\%$
only tW	NLO matching	POWHEG BOX + PYTHIA8 vs MADGRAPH5_AMC@NLO + PYTHIA8
only tW	cross-section	$\pm 5\%$
only tW	tW and $t\bar{t}$ interference	DR vs DS in POWHEG BOX + PYTHIA8
$t\bar{t}V$	PS & hadronisation & NLO matching cross-section	MADGRAPH5_AMC@NLO + PYTHIA8 vs SHERPA $+10.4\%$ ($t\bar{t}Z$) and $+13.3\%$ ($t\bar{t}W$) -12.0% ($t\bar{t}Z$) and -12.0% ($t\bar{t}W$)
$t\bar{t}H$	norm uncertainty	$\pm 50\%$
tZq & tWZ	norm uncertainty	$\pm 50\%$
W + jets	norm uncertainty	$\pm 40\%$
W + HF jets	norm uncertainty	$\pm 40\%$ decorrelated between = 2 HF jets and > 2 HF jets
Z + jets	norm uncertainty	$\pm 35\%$
Diboson	norm uncertainty	$\pm 50\%$

The uncertainty related to the choice of the ME generator and therefore the NLO matching is estimated by replacing the nominal sample with the one produced with MADGRAPH5_AMC@NLO + PYTHIA. For the s - and t -channel, this alternative sample has large statistical uncertainties and leads to several bins with a negative prediction for this process. Therefore, this uncertainty is not considered for these two samples. Two examples can be seen in Figure 13.5. Instead, a 50% norm uncertainty is used for both samples.

Lastly, the uncertainty in the calculated cross-section is considered with 5% for the tW process. As there is already the 50% uncertainty for the s - and t -channel, no additional cross-section uncertainty is assigned to these two processes.

TOP QUARK PAIRS WITH AN ADDITIONAL WEAK BOSON The nominal $t\bar{t}V$ samples generated with MADGRAPH5_AMC@NLO + PYTHIA are replaced by the sample generated with SHERPA. This results in an uncertainty that has effects not only on the PS algorithm but also on the ME generator. Additional uncertainties for the cross-section are applied with $+10.4\%$ and $+13.3\%$ for $t\bar{t}Z$ and $t\bar{t}W$, respectively. Both uncertainties are treated separately for $t\bar{t}Z$ and $t\bar{t}W$, leading to four NP in total.

OTHER TOP For the remaining processes including a top-quark, $t\bar{t}H$, tZq , and tWZ , only a normalisation uncertainty of 50% is considered. One NP is used for the $t\bar{t}H$ process, while tZq and tWZ are described by one NP.

ELECTROWEAK BACKGROUND A normalisation uncertainty of $\pm 40\%$ (W +jets), $\pm 35\%$ (Z +jets), and $\pm 50\%$ (diboson) is used for the electro-weak background. They are treated as decorrelated between the separate processes, leading to three NP. For the W +jets background, two additional NP are included in two separate norm uncertainties of 40% for W +jets events with exactly two HF jets and W +jets events

Table 13.3: Overview of the uncertainties associated to processes other than $t\bar{t}$ + light, $t\bar{t} + \geq 1c$, and $t\bar{t} + \geq 1b$.

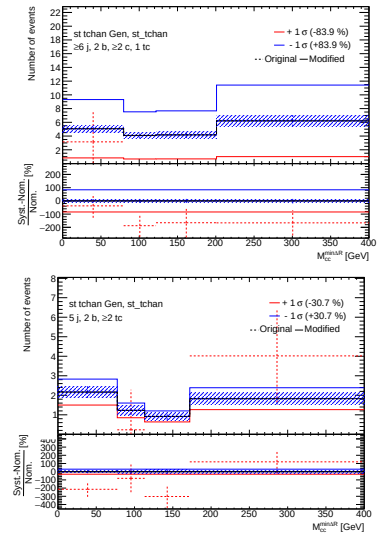


Figure 13.5: Systematic templates of the ME generator uncertainty for the single top t -channel. Several bins are predicted to be negative.

with at least three HF jets.

13.4 Pruning

Table 13.4 lists all systematic uncertainties with their number of NPs before any pruning, together with whether smoothing and symmetrisation are applied. As can be seen, numerous NPs are included in the fit. But most of the variations only have a small impact, meaning the varied templates are very similar to the nominal templates. By removing NPs with a small impact, the fitting procedure can be sped up tremendously without a large impact on the final result. This dropping of NPs is referred to as pruning. A NP can be dropped completely or just the norm or the shape component. If a NP affects multiple samples or multiple regions, it can also be dropped for single samples or single regions. To define which NPs are pruned, a pruning threshold is defined. For this work, a threshold of 0.5% was set for both the normalisation and the shape component. If the impact of the variation on the sample is smaller than the threshold, it is pruned.

Figures 13.6 and 13.7 show the effect of the pruning. Each square corresponds to an uncertainty in one region for one sample. The contributions of the modelling uncertainties are shown uncorrelated. For instance, there is an uncertainty for $t\bar{t} + 1c$ FSR and $t\bar{t} + \geq 2c$ FSR. In the fit, however, these two uncertainties will be treated as correlated, so a single NP will be used to describe them.

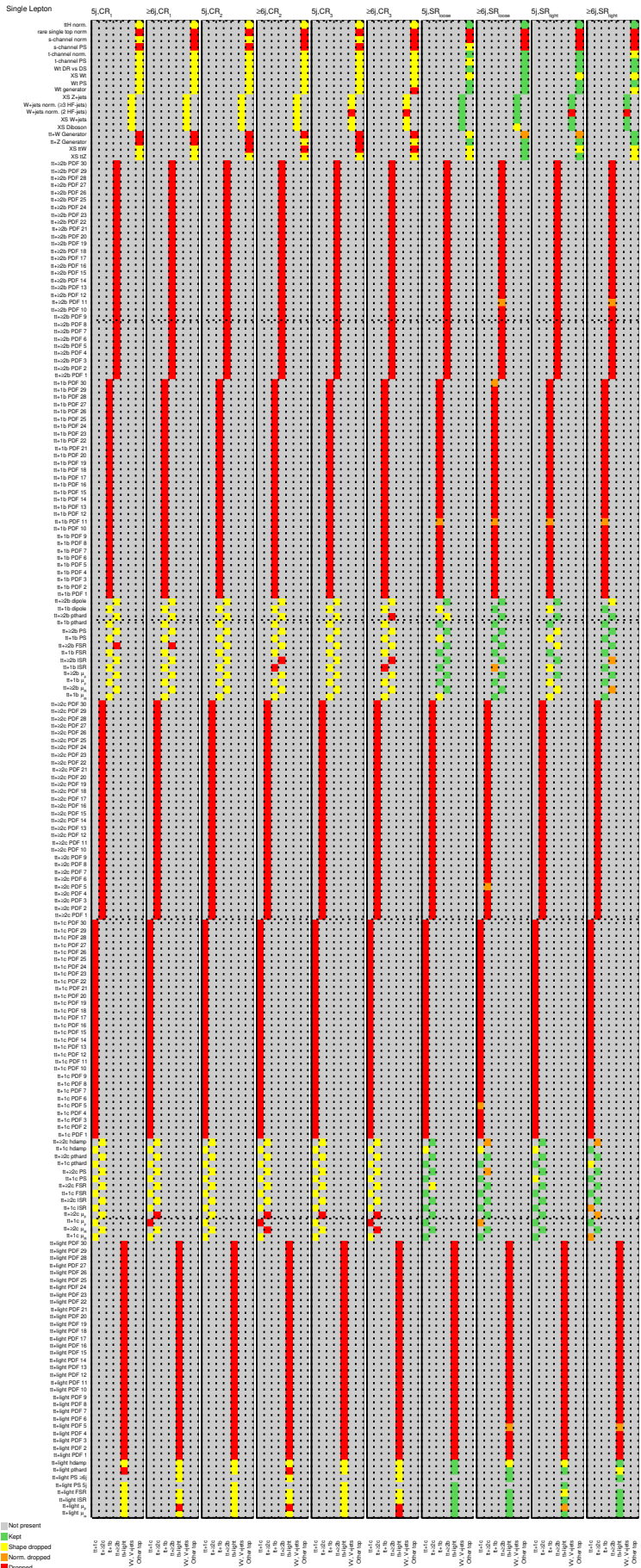
Grey squares represent samples in regions where an uncertainty does not influence them. For instance, an uncertainty for $t\bar{t} + \text{light}$ modelling will not affect the “ $V + \text{jets}, VV$ ” sample. Red squares represent uncertainties that are completely pruned, while green squares represent uncertainties that are fully kept. Yellow (orange) squares are uncertainties for which only the shape (norm) component is dropped. From the figure, it can be seen that many uncertainties are pruned. Especially $t\bar{t}$ PDF uncertainties, lepton uncertainties, and several flavour tagging uncertainties are pruned. Additionally, it can be seen that in the control regions, all shape components are dropped. As there is only one bin in every control region, there is no shape information, and therefore it is automatically dropped.

As NPs are left out of the fit, pruning will typically decrease the uncertainty on the parameters of interest. During the development of the analysis, a test about the impact of the pruning was performed using the Asimov dataset. The uncertainty on the parameters of interest had a relative decrease of around 1% when comparing the case of no pruning to a pruning threshold of 0.5%.

Systematic	Number	Symmetrisation	Smoothing
Luminosity	1	-	-
Pile-up	1	-	-
Electrons (trigger, reco, ID, isolation)	4	-	-
Electrons (resolution, scale)	3	Two-sided	-
Muons (trigger, reco, ID, isolation)	5	-	-
Muons (resolution, scale)	4	Two-sided	-
JVT	1	-	-
JES	35	Two-sided	-
JER	13	Two-sided	-
Flavour tagging	85	-	-
$t\bar{t}$ PS and hadronisation	4	One-sided	X
$t\bar{t}$ p_T^{hard}	3	One-sided	X
$t\bar{t}$ h_{damp}	2	One-sided	X
$t\bar{t}$ FSR	2	Two-sided	X
$t\bar{t}$ Var3c	2	-	-
$t\bar{t}$ μ_R	2	-	-
$t\bar{t}$ μ_F	2	-	-
$t\bar{t}$ PDF	60	One-sided	-
$t\bar{t} + \geq 1b$ PS and hadronisation	2	One-sided	X
$t\bar{t} + \geq 1b$ p_T^{hard}	2	One-sided	X
$t\bar{t} + \geq 1b$ dipole recoil	1	One-sided	X
$t\bar{t} + \geq 1b$ FSR	1	Two-sided	X
$t\bar{t} + \geq 1b$ Var3c	1	-	-
$t\bar{t} + \geq 1b$ μ_R	1	-	-
$t\bar{t} + \geq 1b$ μ_F	1	-	-
$t\bar{t} + \geq 1b$ PDF	30	One-sided	-
Wt DR vs DS	1	One-sided	X
Single top PS and hadronisation	3	One-sided	X
Wt ME generator	1	One-sided	X
Single top norm uncertainty	3	-	-
$t\bar{t}V$ hadronisation and ME	2	One-sided	-
$t\bar{t}V$ cross-section uncertainties	2	-	-
Other top	1	-	-
$t\bar{t}H$	1	-	-
W +jets	3	-	-
Z +jets	1	-	-
Diboson	1	-	-

Table 13.4: Complete list of systematics considered before pruning is applied. The number column describes how many independent NPs are used for this uncertainty in the fit before potential pruning. The last two columns define the treatment of the variations with symmetrisation and smoothing. A "-" indicates no symmetrisation or smoothing, while an "X" indicates smoothing. Systematics for $t\bar{t} + \geq 1b$ are listed separately from $t\bar{t}$ as $t\bar{t} + \geq 1b$ is taken from the dedicated 4FS sample.

Figure 13.6: Experimental uncertainties after the pruning.



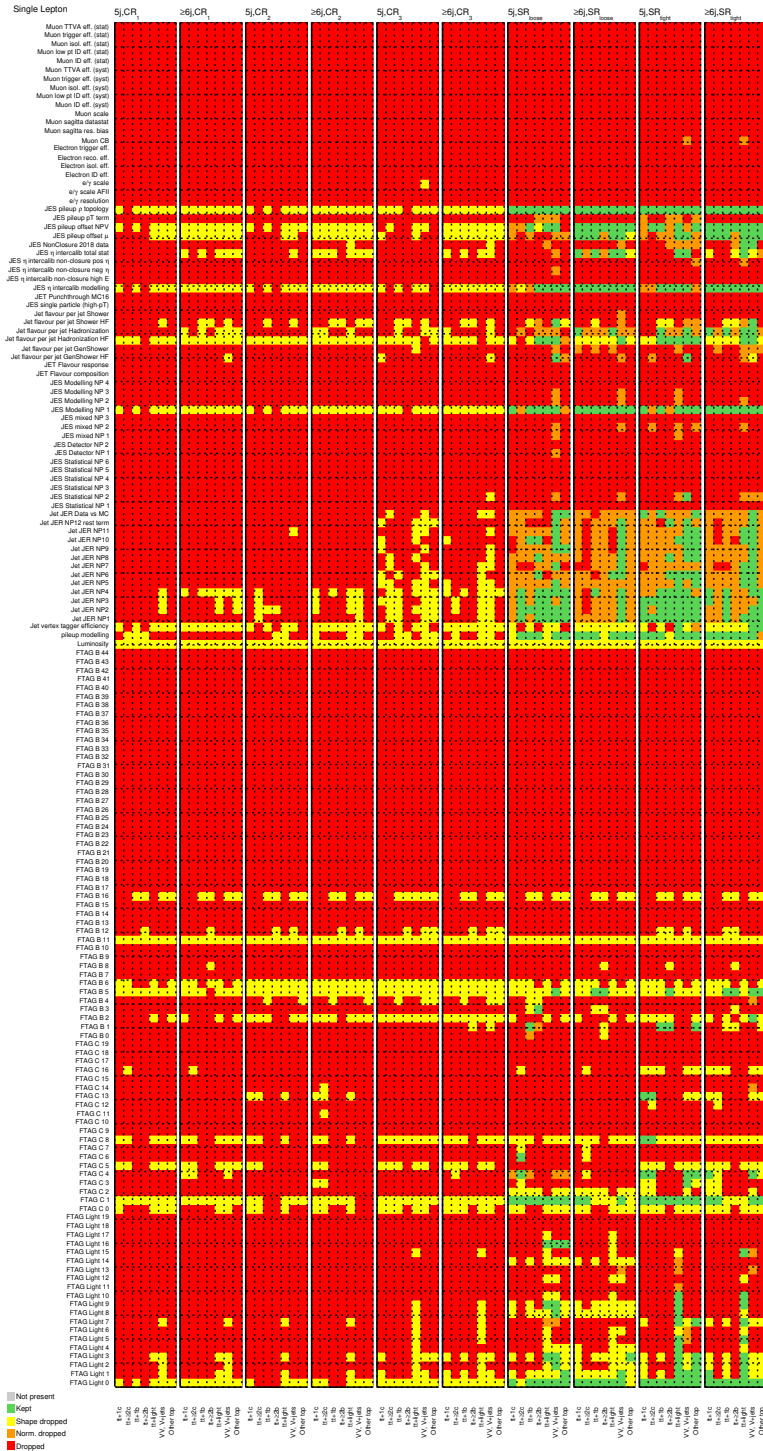


Figure 13.7: Modelling uncertainties after the pruning.

14 Results

After having described the extraction of the parameters, the event selection, and the systematic uncertainties considered, this chapter will discuss the results of the analysis. First, the results of a fit to the Asimov dataset will be shown, before turning attention to ATLAS data. This includes the fit results of the norm factors with their uncertainties, the post-fit values of the NPs, and the impact of single NPs and groups of NPs. Additionally, the “goodness” of the fit is evaluated.

Contents

14.1 Fit results with the Asimov dataset	147
14.2 Fit results with ATLAS data	149

14.1 Fit results with the Asimov dataset

First, the fit model is optimised using the Asimov dataset. One such optimisation can be seen in Figure 14.1. It shows two so-called pull plots, which show $\hat{\theta} - \theta_0 / \Delta\theta$ for NPs for two different fit setups. Here, θ_0 is the nominal value of the NP, $\hat{\theta}$ the value estimated by the fit, and $\Delta\theta$ the prior uncertainty of the NP. The error bar of the single NP corresponds to the post-fit uncertainty $\Delta\hat{\theta}$. In the Asimov fit, all NPs should be at exactly zero, which is the case for the two shown plots. However, the post-fit uncertainty can be smaller than the prior uncertainty for constrained NPs. This is the case in Figure 14.1. The top plot shows the pull for the $t\bar{t} + \geq 1b$ PS NP leading to a normalised post-fit uncertainty around 70%. This constraint is, at least partially, caused by tension between the $t\bar{t} + 1b$ and $t\bar{t} + \geq 2b$ components of this NP. The bottom plot shows the pulls in case one NP is used for the $t\bar{t} + 1b$ component and one for the $t\bar{t} + \geq 2b$ component. Both NPs show a relaxed constraint, as the normalised post-fit uncertainty is now around 90%. Therefore, the decorrelated fit setup was chosen for the analysis. Similar considerations were taken for the $t\bar{t} + \geq 1c$ PS and $t\bar{t} + \text{light PS}$, and the $t\bar{t} + \geq 1b$ and $t\bar{t} + \geq 1c$ NLO matching uncertainty. For the $t\bar{t} + \geq 1c$ PS uncertainty and the two NLO matching uncertainties, the component with a single c/b is treated as decorrelated from the one with at least two c/b . For the $t\bar{t} + \text{light PS}$ uncertainty, the regions with five jets and the regions with at least six jets are described by two separate NP

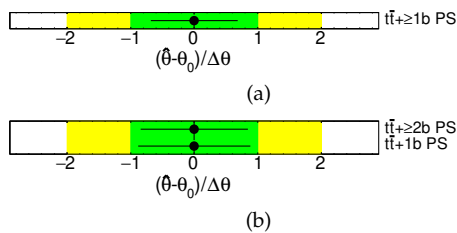


Figure 14.1: Pull plot of the $t\bar{t} + \geq 1b$ PS uncertainty for two different fit setups in the fit to the Asimov dataset. Either (a) a single NP is used or (b) two, one for $t\bar{t} + 1b$ and one for $t\bar{t} + \geq 2b$.

Table 14.1 shows the values of the norm factors in the fit to the Asi-

¹ For the fit to the Asimov dataset, this corresponds to their nominal value, but for the fit to ATLAS data, these will be different.

Table 14.1: Fitted free-floating parameters in the fit to the Asimov dataset. Statistical, systematic, and total uncertainties are shown.

Norm factor	Value	Uncertainties		
		statistical	systematic	total
$\mu(t\bar{t} + \geq 2c)$	1.000	+0.114 -0.114	+0.306 -0.259	+0.327 -0.283
$\mu(t\bar{t} + 1c)$	1.000	+0.108 -0.108	+0.235 -0.194	+0.259 -0.222
$\mu(t\bar{t} + \geq 1b)$	1.000	+0.008 -0.008	+0.078 -0.074	+0.078 -0.074
$\mu(t\bar{t} + \text{light})$	1.000	+0.020 -0.020	+0.105 -0.086	+0.107 -0.098
$\mu(t\bar{t} + \text{jets})$	1.000	+0.007 -0.007	+0.087 -0.079	+0.087 -0.079
$\mu(R_{\geq 2c})$	1.000	+0.109 -0.109	+0.288 -0.248	+0.308 -0.271
$\mu(R_{1c})$	1.000	+0.114 -0.112	+0.262 -0.209	+0.286 -0.237
$\mu(R_{\geq 1b})$	1.000	+0.008 -0.008	+0.086 -0.079	+0.086 -0.079

The total uncertainties for $t\bar{t} + \text{light}$ and $t\bar{t} + \geq 1b$ are considerably smaller than for $t\bar{t} + 1c$ and $t\bar{t} + \geq 2c$. For the first two, they are around 10% or below, while for the latter two, they range from 22% up to 33%. This propagates to the measured ratios and the overall $t\bar{t} + \text{jets}$ cross-section, as $\mu(R_{\geq 2c})$ and $\mu(R_{1c})$ have considerably larger uncertainties compared to $\mu(R_{\geq 1b})$ and $\mu(t\bar{t} + \text{jets})$.

The relative uncertainty on the ratios is only smaller than the relative uncertainty of its cross-section counterpart for $t\bar{t} + \geq 2c$, as only $\mu(R_{\geq 2c})$ has lower uncertainties than $\mu(t\bar{t} + \geq 2c)$. One motivation for measuring ratios is often the cancellation of uncertainties. However, no large effect of this cancellation can be seen here. This can be explained by the fact that only uncertainties that are correlated between the numerator and the denominator will cancel out. The easiest example is the uncertainty on the luminosity. Increasing the luminosity will reduce the fit result of all cross-section norm factors to the same degree, so all cross-section norm factors are positively correlated with each other. In the results in Section 14.2, it can be seen that the impact of the luminosity uncertainty is reduced for the ratios. However, the luminosity uncertainty is 0.83% which is small compared to other uncertainties. Other uncertainties with larger impacts on the final result are not correlated between the numerator and the denominator and therefore do not cancel out. Additionally, uncertainties with a

large impact on only one signal strength will affect all ratios, since all signal strengths always enter in the denominator.

The considerably larger uncertainty of $\mu(t\bar{t} + \geq 2c)$, $\mu(t\bar{t} + 1c)$, $\mu(R_{\geq 2c})$, and $\mu(R_{1c})$ can partially be explained by the choice of using separate norm factors for $t\bar{t} + 1c$ and $t\bar{t} + \geq 2c$, as these are much more difficult to separate from each other than, for instance, $t\bar{t} + \geq 1b$ from $t\bar{t} + \text{light}$. During the development of the analysis, it was also considered to use separate norm factors for $t\bar{t} + 1b$ and $t\bar{t} + \geq 2b$. This led to a considerably larger uncertainty on the single components than on the combined component. As the cross-section of $t\bar{t} + \geq 1b$ is not the primary goal of the analysis, it was chosen to use a combined norm factor. On the other side, the measurement of the cross-section of $t\bar{t} + 1c$ and $t\bar{t} + \geq 2c$ is the goal of this analysis, and, therefore, they are kept separate.

14.2 Fit results with ATLAS data

To assess the quality of the fit, a saturated model [211] is used. The fit is performed twice, once with the nominal fit model and once with a saturated model. The saturated model can perfectly describe the data without introducing penalties to the likelihood due to NPs having non-zero values. Technically, this is implemented through the γ factors. In the saturated model fit, these are used to match the data prediction without pulling any NP. The ratio of the likelihoods obtained from these two fits follows the χ^2 distribution asymptotically. Using this method, a probability of 79% is obtained, indicating that the fit describes the data well.

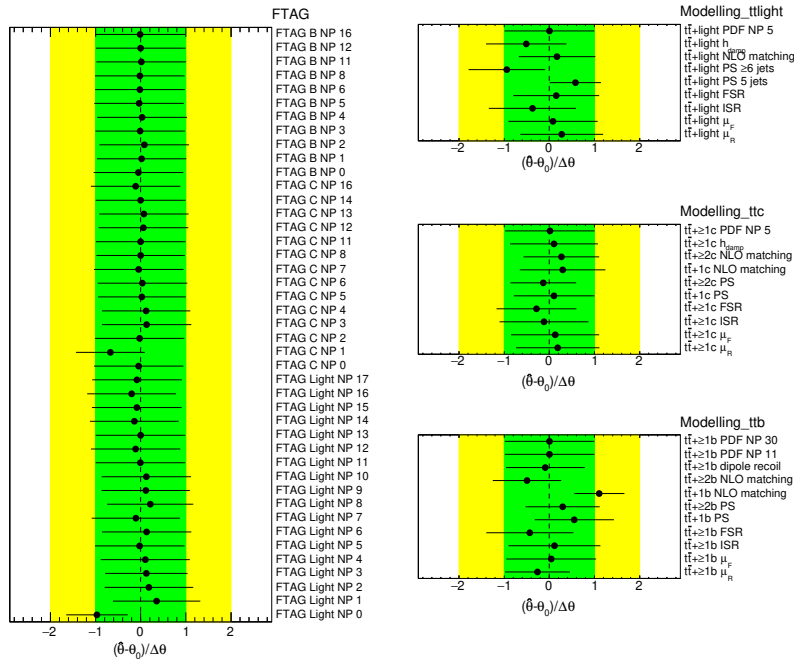


Figure 14.2: Pull distributions of some NPs in the fit to ATLAS data. The NPs are grouped into different origins of the uncertainty.

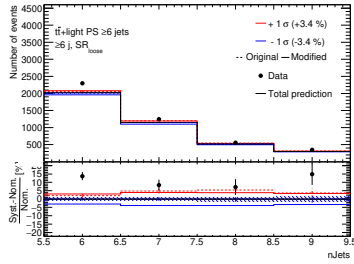


Figure 14.3: Systematic template of the PS uncertainty for the $t\bar{t}$ + light sample with ATLAS data overlaid in $\text{SR}_{\text{loose}}^{\geq 6j}$.

The pulls of all NPs related to flavour-tagging or the modelling of $t\bar{t}$ can be seen in Figure 14.2. Two of the flavour-tagging uncertainties are strongly pulled and constrained, one NP for light-flavour jets and one for c -jets. Several pulls and constraints can be seen for the $t\bar{t}$ modelling uncertainties. Overall, the $t\bar{t} + \geq 1c$ modelling uncertainties are less pulled and constrained than the $t\bar{t} + \geq 1b$ and $t\bar{t}$ + light modelling uncertainties. All pulls and constraints have been studied for their origin. No large pulls or constraints can be seen for the remaining NPs.

The observed pull for the $t\bar{t}$ + light PS NP for regions with at least six jets originates from $\text{SR}_{\text{loose}}^{\geq 6j}$. Figure 14.3 shows the up- and down-variation of this uncertainty on the whole prediction together with data. From the ratio plot, it can be seen that the data has sort of a parabolic shape compared to the prediction, exactly as the down-variation. By pulling the NP to negative values, the shape of the prediction gets corrected to the observed shape of data, therefore explaining the origin of the pull.

Pulls and constraints can also be influenced by correlations between different NPs. Figure 14.4 shows the correlation between all NPs and the four signal strengths, if there is at least one correlation that is larger than 30% or smaller than -30% . The largest correlation can be seen between the $t\bar{t} + 1b$ NLO matching NP and the $t\bar{t} + \geq 1b$ μ_R NP with -77% . The first of these is strongly pulled in the fit. Other large correlations mostly include at least one of the signal strengths. A large correlation to the signal strength will typically result in a large uncertainty on the signal strength due to this NP. For instance, $\mu(t\bar{t} + \geq 2c)$ has a strong anticorrelation with the first flavour-tagging NP for c -jets with -46% . As will be later discussed, this NP has the largest contribution to the uncertainty of this signal strength.

Table 14.2 shows the post-fit values and uncertainties for all eight free-floating parameters. The uncertainties behave the same way as for the fit to the Asimov dataset. The parameters $\mu(t\bar{t} + \geq 2c)$, $\mu(t\bar{t} + 1c)$, $\mu(R_{\geq 2c})$, and $\mu(R_{1c})$ have much larger uncertainties than the remaining ones. All uncertainties are systematically dominated, though the uncertainties for $\mu(t\bar{t} + \geq 2c)$, $\mu(t\bar{t} + 1c)$, $\mu(R_{\geq 2c})$, and $\mu(R_{1c})$ also have sizeable statistical uncertainties. Again, only $\mu(R_{\geq 2c})$ has a lower relative uncertainty than $\mu(t\bar{t} + \geq 2c)$. For $t\bar{t} + 1c$ and $t\bar{t} + \geq 1b$, the ratio has a larger relative uncertainty than the respective cross-section norm factor. For the uncertainties to cancel out in the ratio, ideally all cross-section norm factors are positively correlated. From Figure 14.4, it can be seen that especially $\mu(t\bar{t} + 1c)$ and $\mu(t\bar{t} + \geq 1b)$ have a large negative correlation. As for $\mu(R_{\geq 1b})$ and $\mu(R_{1c})$, one of them is placed in the numerator of the ratio; the uncertainties are not reduced but rather increased.

As expected from previous analyses [6, 8], $\mu(t\bar{t} + \geq 2c)$, $\mu(t\bar{t} + 1c)$, and $\mu(t\bar{t} + \geq 1b)$ are all larger than one. At the same time, $\mu(t\bar{t} + \text{light})$ is smaller than one. The overall $t\bar{t}$ + jets cross-section norm-factor $\mu(t\bar{t} + \text{jets})$ is very close to one which is in agreement with previous measurements [212]. As the ratios are the cross-section

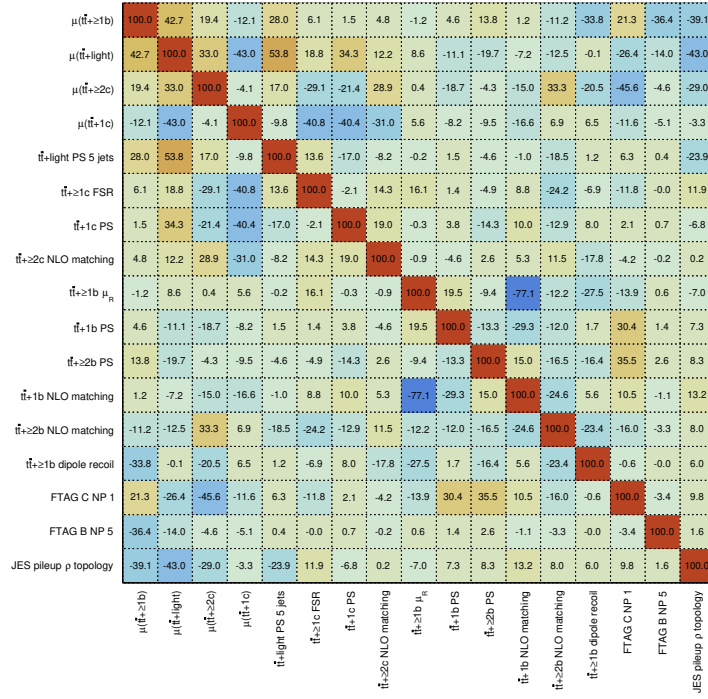


Figure 14.4: Correlation matrix between the NPs and cross-section norm factors in the fit to ATLAS data. Only quantities which have a correlation of at least 30%, positive or negative, with any other quantity are shown.

Norm factor	Value	Uncertainties		
		statistical	systematic	total
$\mu(t\bar{t} + \geq 2c)$	1.547	+0.141 -0.140	+0.419 -0.358	+0.442 -0.384
$\mu(t\bar{t} + 1c)$	1.224	+0.127 -0.127	+0.281 -0.236	+0.308 -0.268
$\mu(t\bar{t} + \geq 1b)$	1.179	+0.009 -0.009	+0.087 -0.082	+0.088 -0.082
$\mu(t\bar{t} + \text{light})$	0.922	+0.023 -0.023	+0.103 -0.097	+0.106 -0.100
$\mu(t\bar{t} + \text{jets})$	0.986	+0.008 -0.008	+0.084 -0.077	+0.084 -0.077
$\mu(R_{\geq 2c})$	1.569	+0.133 -0.134	+0.375 -0.336	+0.398 -0.362
$\mu(R_{1c})$	1.242	+0.139 -0.137	+0.315 -0.257	+0.344 -0.291
$\mu(R_{\geq 1b})$	1.196	+0.010 -0.010	+0.097 -0.090	+0.098 -0.091

norm factors normalised by $\mu(t\bar{t} + \text{jets})$, the values of the cross-section and ratio norm factors are very similar. While the post-fit values of $\mu(t\bar{t} + \geq 2c)$, $\mu(t\bar{t} + 1c)$, and $\mu(t\bar{t} + \geq 1b)$ are all larger than one, they are all compatible with one within uncertainties. The largest deviation is observed for $\mu(t\bar{t} + \geq 1b)$ with 2.2σ , while for $\mu(t\bar{t} + 1c)$ it is only 0.8σ .

The impact of individual NPs on the norm factors $\mu(t\bar{t} + \geq 2c)$, $\mu(t\bar{t} + 1c)$, $\mu(R_{\geq 2c})$ and $\mu(R_{1c})$ can be seen in Figure 14.5. Only the

Table 14.2: Fitted free-floating parameters in the fit to ATLAS data. Statistical, systematic, and total uncertainties are shown.

ten uncertainties with the highest post-fit impact are shown. The impact is evaluated by running the fit while fixing the value of the NP to its fit result varied by its prior uncertainty $\hat{\theta} \pm \Delta\theta$ to obtain its pre-fit impact. For the post-fit impact, the value is instead fixed to the upper and lower post-fit estimate $\hat{\theta} \pm \Delta\hat{\theta}$. The difference in the result of the signal strengths to the nominal fit result can be interpreted as the impact of the NP on the signal strength. In addition to pre- and post-fit impact, the pull $\hat{\theta} - \theta_0 / \Delta\theta$ is shown as well.

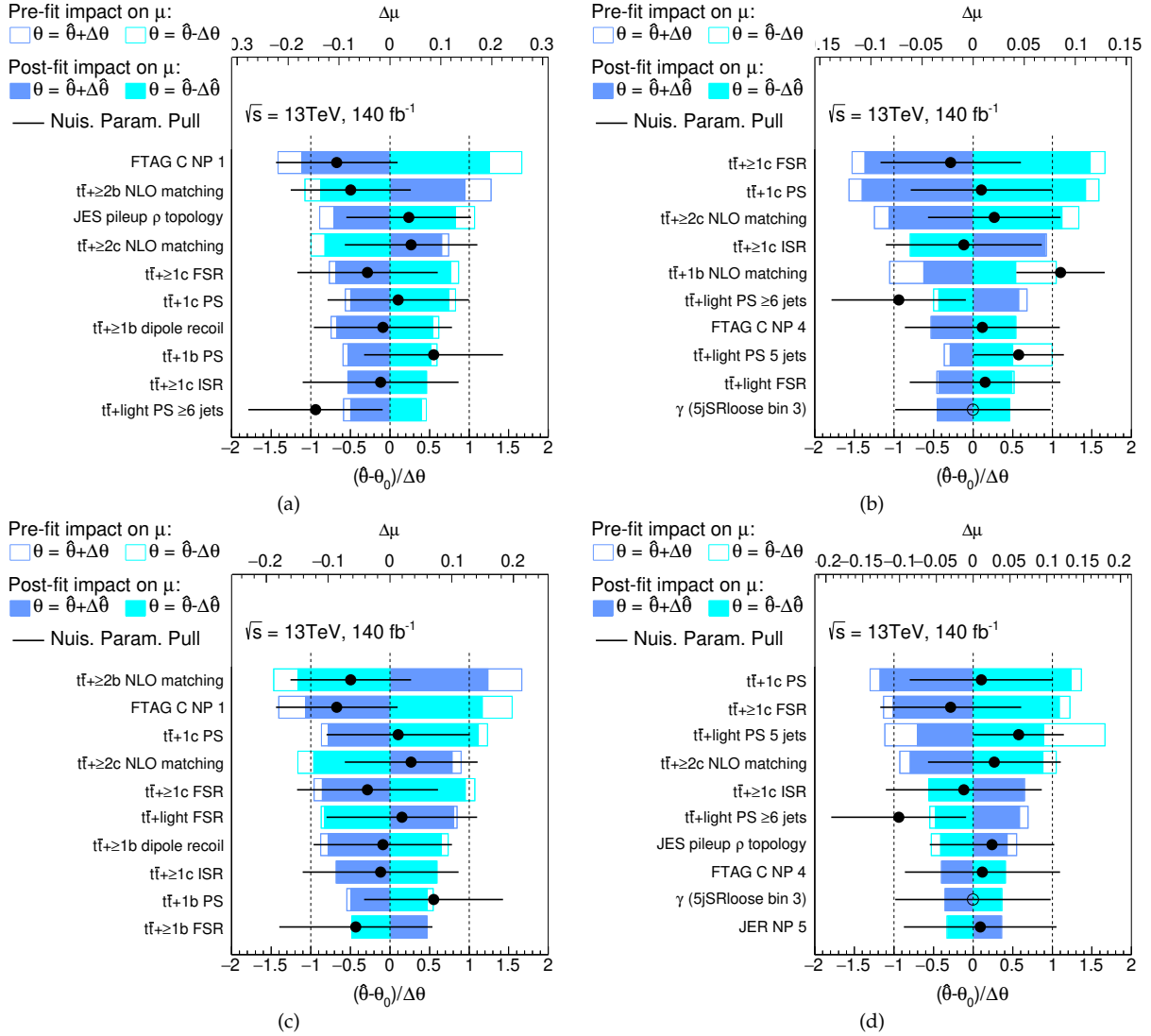


Figure 14.5: Ranking plots for the fit to ATLAS data. The top row shows the impact of each systematic uncertainty on (a) $\mu(t\bar{t} + \geq 2c)$ and (b) $\mu(t\bar{t} + 1c)$. The bottom row shows the impact of each systematic uncertainty on (c) $\mu(R_{\geq 2c})$ and (d) $\mu(R_{1c})$. Only the 10 NP with the highest impact are shown.

Most of the highly ranked NPs concern $t\bar{t}$ modelling, though jet modelling and flavour-tagging uncertainties are also ranked fairly high. For $\mu(t\bar{t} + 1c)$ and $\mu(R_{1c})$ even one γ parameter in $\text{SR}_{\text{loose}}^{5j}$ is among the ten highest ranked NPs with rank ten and nine, respectively. Not only $t\bar{t} + \geq 1c$ modelling has a high impact on the results, but also $t\bar{t} + \geq 1b$ and $t\bar{t} + \text{light}$ modelling. The highly ranked $t\bar{t}$ modelling uncertainties are mostly related to the PS, the NLO matching, or FSR. The scale variations in μ_R and μ_F or the PDF uncertainties are not among the ten highest ranked NP for any $t\bar{t}$ component. For the

flavour tagging uncertainties, only uncertainties with respect to c -jets are ranked highly. For $\mu(t\bar{t} + \geq 2c)$, one flavour-tagging uncertainty is even the highest-ranked one.

Another way of estimating the impact of a NP or a group of NPs is to fix these NPs to their post-fit value. The uncertainty of the norm factors should be reduced after fixing these. Subtracting the nominal uncertainty and the uncertainty obtained after fixing the NPs from each other in quadrature gives the estimated impact of this NP or group of NPs. It should be noted that the uncertainties of the single groups do not have to add up to the overall uncertainty due to correlations between NPs of different groups.

Table 14.3 lists the impact of several groups on all eight extracted parameters. Due to the size of the luminosity uncertainty of 0.83%, the luminosity alone cannot have a smaller impact on any of the non-ratio parameters. Indeed, the minimal impact of the “Luminosity & pileup” observed for cross-section norm factors is 0.8%. For the ratios, however, the luminosity uncertainty cancels out, resulting in no impact due to the luminosity uncertainty. The uncertainty related to the treatment of pileup can still have a non-zero impact, which is why not all impacts of the “Luminosity & pileup” group are zero for the ratios.

Group	Relative uncertainty [%]							
	$\mu(t\bar{t} + \geq 2c)$		$\mu(t\bar{t} + 1c)$		$\mu(t\bar{t} + \geq 1b)$		$\mu(t\bar{t} + \text{light})$	
Luminosity & pileup	+1.6	−1.2	+1.2	−1.0	+1.0	−0.9	+1.0	−0.8
Flavour-tagging light-flavour jets	+5.4	−5.0	+5.4	−3.5	+2.1	−1.9	+1.5	−1.7
Flavour-tagging c -jets	+14.3	−10.7	+6.2	−4.6	+2.0	−1.7	+3.4	−2.8
Flavour-tagging b -jets	+2.1	−1.2	+1.8	−1.1	+3.3	−3.0	+2.0	−1.6
Leptons	+0.7	—	+0.4	—	—	—	+0.1	—
Jet modelling	+12.9	−8.7	+7.4	−5.8	+4.4	−4.0	+7.6	−6.6
$t\bar{t}$ + light modelling	+8.3	−8.1	+6.1	−6.2	+3.6	−3.3	+7.9	−6.8
$t\bar{t} + \geq 1c$ modelling	+17.0	−14.1	+17.3	−13.0	+1.3	−1.2	+4.7	−5.0
$t\bar{t} + \geq 1b$ modelling	+16.1	−12.7	+9.4	−5.9	+5.4	−4.9	+6.0	−4.8
Background modelling	+3.1	−2.4	+2.5	−2.3	+2.3	−2.4	+1.1	−1.0
Total systematic uncertainty	+27.1	−23.1	+23.0	−19.3	+7.4	−7.0	+11.2	−10.5
Total	+28.6	−24.9	+25.2	−21.9	+7.4	−7.0	+11.5	−10.8
	$\mu(R_{\geq 2c})$		$\mu(R_{1c})$		$\mu(R_{\geq 1b})$		$\mu(t\bar{t} + \text{jets})$	
Luminosity & pileup	—	−1.1	—	−1.2	—	—	+1.0	−0.8
Flavour-tagging light-flavour jets	+5.5	−5.0	+6.3	−3.7	+1.7	−1.5	+0.9	−0.8
Flavour-tagging c -jets	+10.7	−8.6	+4.7	−4.0	+5.0	−4.4	+3.1	−2.8
Flavour-tagging b -jets	+1.3	−1.2	—	−1.3	+1.6	−1.5	+1.9	−1.7
Leptons	—	—	+0.4	−0.2	—	—	—	—
Jet modelling	+5.5	−4.4	+8.3	−7.0	+2.5	−2.2	+6.5	−5.9
$t\bar{t}$ + light modelling	+8.2	−7.7	+10.6	−9.0	+4.2	−3.9	+5.8	−5.1
$t\bar{t} + \geq 1c$ modelling	+17.6	−15.2	+18.9	−13.7	+1.7	−1.4	+2.0	−1.8
$t\bar{t} + \geq 1b$ modelling	+13.0	−11.2	+5.6	−3.1	+6.0	−5.4	+5.6	−5.0
Background modelling	+2.1	−1.8	+1.4	−1.5	+2.1	−1.9	+1.2	−1.2
Total systematic uncertainty	+23.9	−21.4	+25.4	−20.7	+8.1	−7.5	+8.5	−7.8
Total	+25.4	−23.1	+27.7	−23.4	+8.2	−7.6	+8.5	−7.8

Different groups have the highest impact on different extracted parameters. Generally, the modelling of the measured process, for instance, $t\bar{t} + \geq 1c$ modelling for $\mu(t\bar{t} + \geq 2c)$, has one of the largest

Table 14.3: Groups of NP and their relative impact on the eight extracted norm factors. A — represents a negligible impact.

impacts. Additionally, the $t\bar{t} + \geq 1b$ modelling has also a sizeable impact on $\mu(t\bar{t} + \geq 2c)$, $\mu(R_{\geq 2c})$, $\mu(t\bar{t} + \text{light})$, and $\mu(t\bar{t} + \text{jets})$. The flavour-tagging category with the highest impact is, most of the time, the uncertainty on c -jets. The uncertainties concerning b -jets have almost always a negligible impact compared to the other flavour-tagging uncertainties, except for $\mu(t\bar{t} + \geq 1b)$. Jet uncertainties have the largest impact of all groups for $\mu(t\bar{t} + \text{jets})$ and almost the largest impact for $\mu(t\bar{t} + \text{light})$. The modelling of all non- $t\bar{t}$ processes has a small impact on all extracted parameters, while lepton uncertainties are completely negligible.

Finally, the value of the likelihood is scanned with respect to $\mu(t\bar{t} + \geq 2c)$ and $\mu(t\bar{t} + 1c)$. The fit is repeated a total of 900 times. In each of the fits, the values of $\mu(t\bar{t} + \geq 2c)$ and $\mu(t\bar{t} + 1c)$ are fixed to a different value following a grid-like structure around the nominal fit result. At each point, the log-likelihood value is recorded and subtracted from the log-likelihood of the nominal fit. The resulting two-dimensional histogram is shown in Figure 14.6. The histogram has a minimum at the nominal fit result and smoothly rising values in all directions, showing good fit stability. No local minima or other artefacts can be seen. From this log-likelihood scan, the uncertainties of the parameters can be obtained as well. The 1σ and 2σ contours are defined by the values where $-\Delta \ln \mathcal{L} = 0.5$ and $-\Delta \ln \mathcal{L} = 2$, respectively. Both contours are highlighted by the ellipses in the plot. As can be seen, the ellipses are slightly tilted due to the slight correlation of -5% of the two parameters.

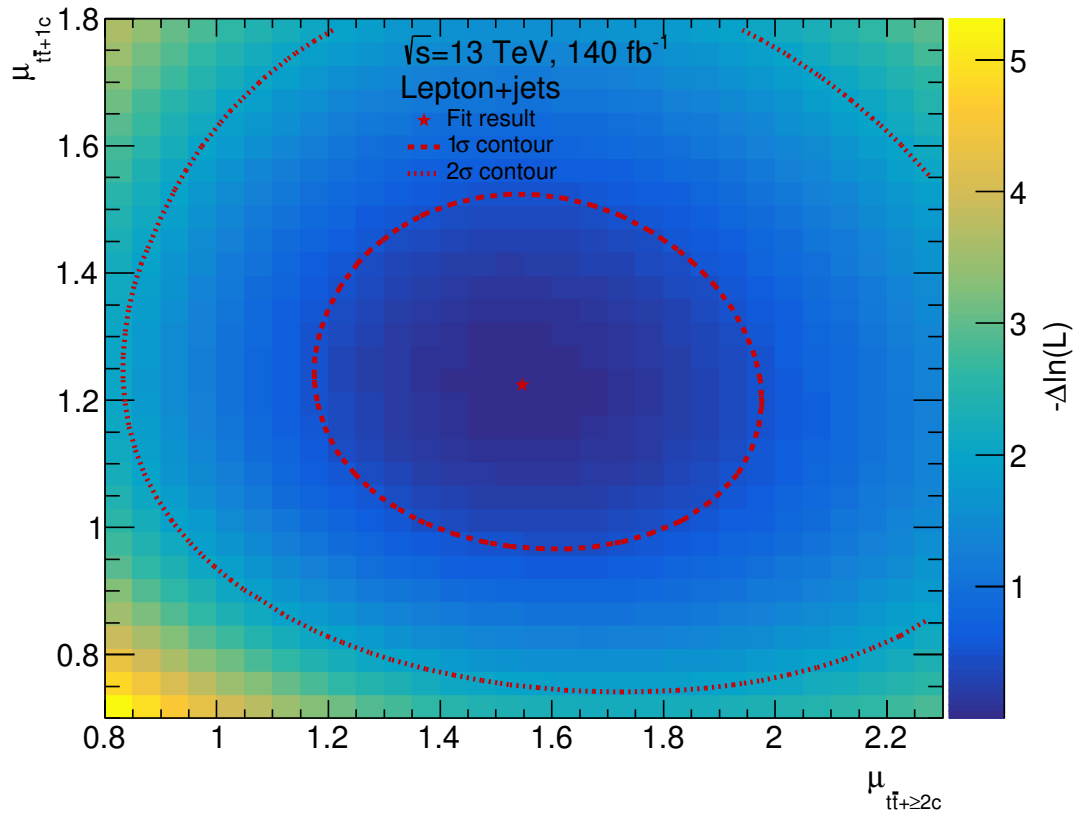


Figure 14.6: Two-dimensional likelihood scan of $\mu(t\bar{t} + \geq 2c)$ and $\mu(t\bar{t} + 1c)$ for the fit to ATLAS data. The 1σ and 2σ contour are highlighted with ellipses, while the nominal fit result is highlighted with a star.

15 Combined results

The final result of the analysis will be given by a combined fit that includes the lepton+jets and the dilepton channel. In the combined fit, regions from both channels are used at the same time. Additionally, the systematic uncertainties are, by default, treated as correlated between the two channels. In certain cases, they can also be treated as decorrelated pre-fit. Results from this combination will be quoted here after briefly describing the analysis in the dilepton channel.

Contents

15.1	Dilepton channel	157
15.1.1	Event selection	157
15.1.2	Fiducial phase space	158
15.1.3	Systematics uncertainties	158
15.2	Combined fit	159

15.1 Dilepton channel

15.1.1 Event selection

The overall setup for the event selection is very similar to the lepton+jets event selection. The definition of the physics objects is the same. A similar preselection with DL1r tagging bins is performed before creating fit regions based on the multiplicities of the bc -Tagger. For the preselection, one of the single lepton triggers has to be passed. Exactly two leptons with opposite electrical charges are required, with one having $p_T > 27$ GeV. If the two leptons are electrons, the second electron has to have $p_T > 15$ GeV, while no additional p_T requirement is applied for other flavour combinations. Both leptons have to be, apart from the p_T requirement, tight leptons, with no additional loose leptons in the event. If the leptons have the same flavour, the invariant dilepton mass has to be above 15 GeV and outside a window around the Z boson mass ranging from 83 GeV to 99 GeV. At least three jets are required, with at least two being in the DL1r $b@70\%$ tagging bin.

Similar to the lepton+jets channel, the fit regions are split into regions with exactly three jets and regions with at least four jets. Three control regions are formed for both jet multiplicities. One signal region is used for events with three jets, while two are used for events with at least four jets. In the control regions, only the yield is used in the fit. In the signal region with three jets, the invariant mass of the tagged c -jet and the closest other jet is used, while in the signal regions with at least four jets, the jet multiplicity is used as a fit observable. Table 15.1 shows the selection based on the bc -Tagger.

An important difference to the lepton+jets channel is the treatment of fake leptons. All simulated events passing the selection that have at least one fake lepton are collected in a separate sample describing fakes. Certain processes require at least one fake lepton to pass the event selection, like W +jets production. These processes do, therefore,

Table 15.1: Definition of the fit regions in the dilepton channel based on the bc -Tagger working points. A dash refers to no requirement on this category of jets. $b@70\%$ and $c@22\%$ are inclusive of their respective tighter working points. A number with no additional “ $\geq$ ” implies an exact requirement, e.g. $b@70\% = 3$, $b@60\% = 3$ vetoes additional $b@70\%$ jets.

Region	No. jets	$b@70\%$	$b@60\%$	$c@22\%$	$c@11\%$
$CR_1^{2\ell,3j}$	3	2	—	0	—
$CR_2^{2\ell,3j}$	3	3	3	—	—
$CR_3^{2\ell,3j}$	3	3	<3	—	—
$SR_{\text{loose}}^{2\ell,3j}$	3	2	—	1	—
$CR_1^{2\ell,\geq 4j}$	≥ 4	2	—	0	—
$CR_2^{2\ell,\geq 4j}$	≥ 4	≥ 3	≥ 3	—	—
$CR_3^{2\ell,\geq 4j}$	≥ 4	≥ 3	<3	—	—
$SR_{\text{loose}}^{2\ell,\geq 4j}$	≥ 4	2	—	1	—
$SR_{\text{tight}}^{2\ell,\geq 4j}$	≥ 4	2	—	≥ 2	—

not contribute by themselves but only through the fake sample.

15.1.2 Fiducial phase space

Similar to the event selection at the reconstruction level, the fiducial phase space selection has to be adjusted for the dilepton channel. Two leptons have to be in the event. A p_T requirement of $p_T > 27$ GeV is placed on one lepton. The second lepton has to have $p_T > 15$ GeV if both leptons are electrons; otherwise, it has to have $p_T > 10$ GeV. A veto on additional leptons with $p_T > 10$ GeV is used as well. The two leptons have to have opposite electrical charges. If the two leptons have the same flavour, the same requirements for the invariant dilepton mass as in the reconstruction level selection are used. Three jets are required, with at least two of them being ghost-matched to a b -hadron.

15.1.3 Systematics uncertainties

Most systematic uncertainties are used in the same way as in the lepton+jets channel. Uncertainties associated with samples with only events that cannot pass the event selection because they have too few genuine leptons are removed. Instead, the fake sample receives an additional systematic uncertainty of 25%. The same decorrelation scheme as for the lepton+jets channel for the $t\bar{t}$ modelling uncertainties is used. For instance, the $t\bar{t} + 1c$ and $t\bar{t} + \geq 2c$ p_T^{hard} uncertainties are treated as decorrelated pre-fit. Similar to the decorrelation of regions with five jets and at least six jets for the $t\bar{t} + \text{light PS NP}$ in the lepton+jets channel, regions with three jets and exactly four jets are decorrelated for the $t\bar{t} + \text{light PS NP}$ in the dilepton channel. In the combined fit, all NPs that have an equivalent NP in the other channel are correlated with each other except the $t\bar{t}$ PS NPs and the $t\bar{t}$ p_T^{hard}

NPs.

As the fiducial phase space is different, the reweighting factors for the $t\bar{t}$ modelling uncertainties have to be rederived for the dilepton channel. The same applies for the fractions used in the expression of the overall $t\bar{t}$ +jets norm factor. For the combined fit, both the reweighting factors and the fractions are derived from the combined fiducial phase space. The results of the combined fit will also be compared to the prediction of the combined fiducial phase space.

15.2 Combined fit

As for the fit in only the lepton+jets channel, the goodness of the fit is evaluated using the saturated model approach. A p-value of one is obtained, showing a very good description of the data by the fit model.

Figure 15.1 shows a comparison of the eight extracted parameters from the dilepton, lepton+jets, and combined fit, while Table 15.2 lists the values of the combined fit. Additionally, the statistical, systematic, and total uncertainties are listed.

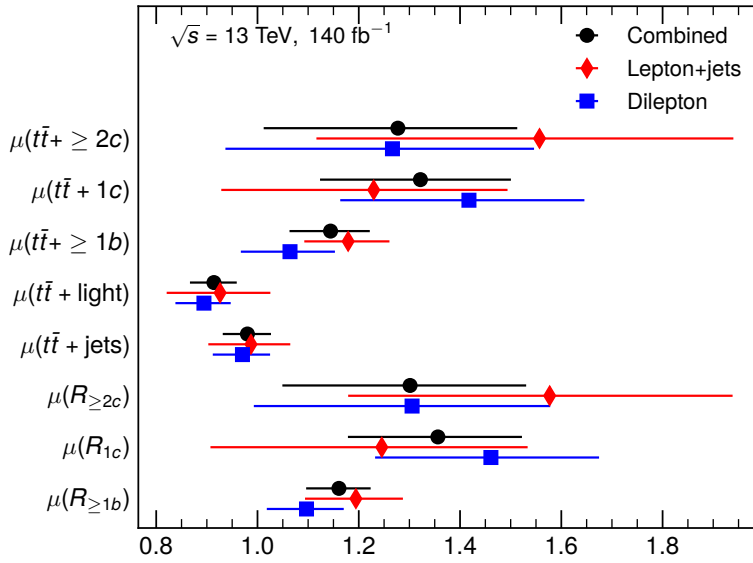


Figure 15.1: Comparison of the fitted free-floating parameters in the dilepton, lepton+jets, and combined fit.

The results of all three fits are compatible with each other within uncertainties, though there are some deviations in the values of the norm factors. In the dilepton fit, the norm factor for $t\bar{t} + \geq 2c$ is smaller than the one for $t\bar{t} + 1c$, while the opposite is the case for the lepton+jets fit. Both the $t\bar{t} + \geq 1b$ and $t\bar{t} + \text{light}$ norm factors are slightly smaller in the dilepton fit than in the lepton+jets fit. The overall $t\bar{t}$ +jets norm factor does change only slightly between the two fits.

The results of the combined fit always lie between the results of the single fits. Often, the result of the combined fit is pulled more towards the fit result with the lower uncertainty. For instance, for $t\bar{t} + \geq 2c$,

Table 15.2: Fitted free-floating parameters in the combined fit to ATLAS data. Statistical, systematic, and total uncertainties are shown.

Norm factor	Value	Uncertainties		
		statistical	systematic	total
$\mu(t\bar{t} + \geq 2c)$	1.277	+0.073 -0.072	+0.255 -0.224	+0.265 -0.236
$\mu(t\bar{t} + 1c)$	1.322	+0.037 -0.037	+0.195 -0.175	+0.198 -0.179
$\mu(t\bar{t} + \geq 1b)$	1.144	+0.007 -0.007	+0.081 -0.077	+0.081 -0.078
$\mu(t\bar{t} + \text{light})$	0.914	+0.003 -0.003	+0.047 -0.045	+0.047 -0.045
$\mu(t\bar{t} + \text{jets})$	0.980	+0.002 -0.002	+0.049 -0.046	+0.049 -0.047
$\mu(R_{\geq 2c})$	1.301	+0.075 -0.074	+0.241 -0.217	+0.252 -0.229
$\mu(R_{1c})$	1.356	+0.037 -0.037	+0.174 -0.162	+0.178 -0.166
$\mu(R_{\geq 1b})$	1.161	+0.007 -0.007	+0.064 -0.062	+0.065 -0.063

the dilepton channel has a lower uncertainty, and the combined fit result is closer to the dilepton result than the lepton+jets result. For $t\bar{t} + \geq 1b$, the opposite is the case.

The uncertainties are reduced with respect to the lepton+jets fit but also the dilepton fit. Additional constraints in the combined fit and changes in the correlations cause the reduction in systematic uncertainties. For instance, the relative uncertainties of the ratios are all smaller than their respective cross-section norm factors in the combined fit due to a positive correlation of all cross-section norm factors. This was not the case in the lepton+jets fit.

For the statistical uncertainties, the reduction is caused not only by the additional statistics of the dilepton channel but also by lower statistical uncertainties of the dilepton channel alone, especially for the $t\bar{t} + \geq 2c$ and $t\bar{t} + 1c$ norm factors. A reason for this is a purer selection, mostly caused by the additional c -jet from the W boson decay in the lepton+jets channel. A selection of two c -tagged jets leads to mostly $t\bar{t} + \geq 2c$ in the dilepton channel, whereas in the lepton+jets channel, a considerable $t\bar{t} + 1c$ contribution will be observed due to the additional c -jet. This could be reduced by requiring three c -tagged jets, but this reduces the available statistics too much.

Table 15.3 shows the impact of groups of NPs on the measured norm factors. The same categories as in the lepton+jets fit are the dominant uncertainties. Overall, the impact of most categories is reduced in the combined fit. A large reduction can, for instance, be seen in the impact of the modelling of $t\bar{t} + \geq 1b$ on $\mu(t\bar{t} + \geq 2c)$. Similar impacts as in the lepton+jets fit are obtained for ‘luminosity & pileup’ and the modelling of background processes. An increase in impact can be seen for uncertainties related to leptons, though they are still negligible compared to the $t\bar{t}$ modelling uncertainties.

Figures 15.2 and 15.3 show post-fit plots of the control and signal regions of the lepton+jets selection, respectively. Here, the single samples are scaled by their norm factors, and all NPs are fixed to

Group	Relative uncertainty [%]							
	$\mu(t\bar{t} + \geq 2c)$		$\mu(t\bar{t} + 1c)$		$\mu(t\bar{t} + \geq 1b)$		$\mu(t\bar{t} + \text{light})$	
Luminosity & pileup	+0.6	-1.3	+0.9	-0.9	+1.0	-0.9	+1.1	-1.0
Flavour-tagging light-flavour jets	+2.8	—	+2.7	-3.0	+2.0	-1.8	+0.8	-0.7
Flavour-tagging c -jets	+10.7	-8.1	+5.9	-4.8	+2.1	-1.7	+0.7	-0.7
Flavour-tagging b -jets	+1.8	-1.8	+1.3	-1.5	+3.4	-3.1	+2.2	-2.0
Leptons	+1.2	-0.9	+0.4	—	+0.4	-0.3	+0.8	-0.7
Jet modelling	+8.0	-5.1	+6.2	-3.8	+4.3	-4.0	+3.8	-3.5
$t\bar{t}$ + light modelling	+10.1	-7.6	+7.0	-5.6	+3.8	-3.5	+4.3	-4.0
$t\bar{t} + \geq 1c$ modelling	+13.9	-11.4	+12.0	-9.8	+1.4	-1.7	+2.1	-1.9
$t\bar{t} + \geq 1b$ modelling	+9.6	-8.0	+8.9	-7.3	+5.4	-5.1	+3.0	-2.7
Background modelling	+2.1	-2.5	+2.0	-2.0	+2.5	-2.7	+1.3	-1.2
Total systematic impact	+20.0	-17.5	+14.8	-13.1	+7.0	-6.8	+5.2	-4.9
Total	+20.8	-18.5	+15.0	-13.5	+7.1	-6.8	+5.2	-4.9
	$\mu(R_{\geq 2c})$		$\mu(R_{1c})$		$\mu(R_{\geq 1b})$		$\mu(t\bar{t} + \text{jets})$	
Luminosity & pileup	—	-1.1	+0.9	—	+0.5	-0.5	+1.0	-1.0
Flavour-tagging light-flavour jets	+2.0	-1.6	+2.6	-2.8	+1.4	-1.3	+0.7	-0.7
Flavour-tagging c -jets	+10.1	-8.2	+5.3	-4.4	+2.3	-2.1	+0.4	-0.5
Flavour-tagging b -jets	+1.7	—	+1.0	-0.8	+1.5	-1.4	+2.1	-2.0
Leptons	+0.8	-0.3	+0.8	-0.4	+0.3	-0.3	+0.6	-0.6
Jet modelling	+2.8	-3.0	+3.0	-1.9	+1.4	-1.2	+3.9	-3.7
$t\bar{t}$ + light modelling	+8.3	-6.7	+6.3	-5.7	+2.9	-2.6	+3.8	-3.6
$t\bar{t} + \geq 1c$ modelling	+13.2	-11.7	+10.0	-8.6	+2.1	-2.1	+2.2	-2.1
$t\bar{t} + \geq 1b$ modelling	+8.5	-7.5	+6.7	-5.8	+4.2	-4.0	+3.3	-3.1
Background modelling	—	-2.2	+1.8	-1.0	+2.1	-2.2	+1.2	-1.2
Total systematic impact	+18.4	-16.7	+12.9	-11.9	+5.5	-5.4	+5.0	-4.7
Total	+19.4	-17.6	+13.1	-12.2	+5.6	-5.4	+5.0	-4.7

Table 15.3: Groups of NP and their relative impact on the eight extracted norm factors in the combined fit. A — represents a negligible impact.

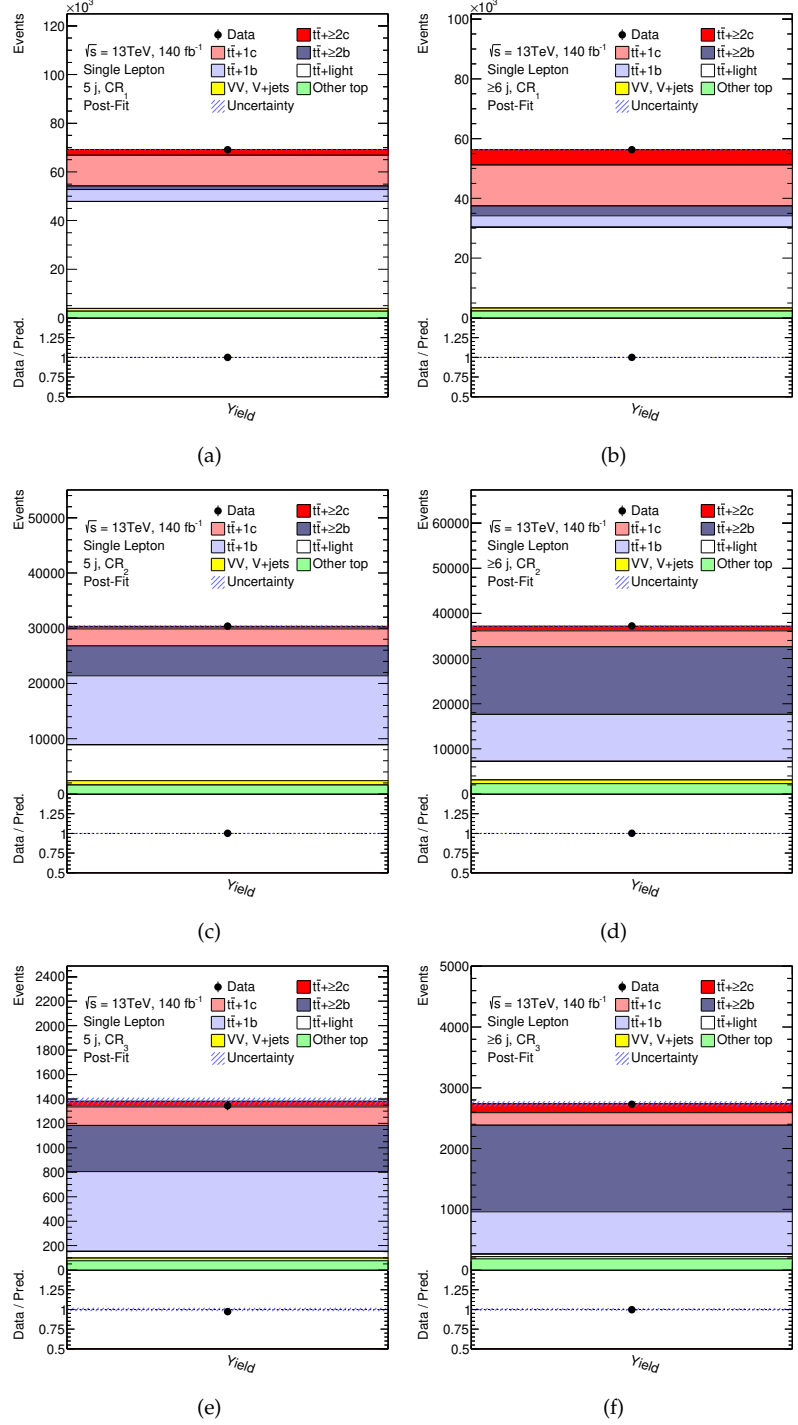
their estimated value. It can be clearly seen that much better data-MC agreement is achieved using the post-fit values. Additionally, the uncertainty band of the prediction is reduced. This has two reasons: constraints and correlations. Obtained constraints reduce the impact of single NPs, while correlations between NPs can reduce the combined impact.

Table 15.4 shows expected and observed yields in the single regions pre- and post-fit. The expected yields are further split into the single samples. Pre-fit, large differences between the overall predicted and observed yield can be seen in nearly every region. No large deviations can be seen for the post-fit yields. As expected from the norm factors, the $t\bar{t} + \geq 2c$ and $t\bar{t} + 1c$ yields are always increased post-fit compared to pre-fit.

The same is not true for $t\bar{t} + \geq 2b$ and $t\bar{t} + 1b$. In regions with five jets, the $t\bar{t} + \geq 2b$ yield is typically reduced, while it is increased in regions with at least six jets. For $t\bar{t} + 1b$, the behaviour is exactly the opposite; a reduction in yields can be seen in regions with at least six jets, while an increase can be seen in regions with five jets. This is caused by pulls introduced in the fit. For instance, the p_T^{hard} variation has a large impact on the shape of the number of jets. For $t\bar{t} + 1b$ and $t\bar{t} + \geq 2b$, the p_T^{hard} variation increases the expected number of events in each region with five jets while reducing it in regions with at least six jets. Pulls in these NPs can therefore increase yields in one jet multiplicity but decrease it for the other.

The $t\bar{t}$ + light post-fit yield does not follow any particular pattern. In some regions, it is reduced compared to the pre-fit yield; in others, it is increased. The yield of the electroweak background increases in all regions post-fit. This is caused by a slight pull on the normalisation of W +jets events. Lastly, the yields of all other processes involving top-quarks stay fairly constant.

Figure 15.2: Yields in the three control regions post-fit. All regions are split into (a,c,e) events with exactly five jets and (b,d,f) events with at least six jets. (a,b) CR_1 , (c,d) CR_2 , and (e,f) CR_3 are shown.



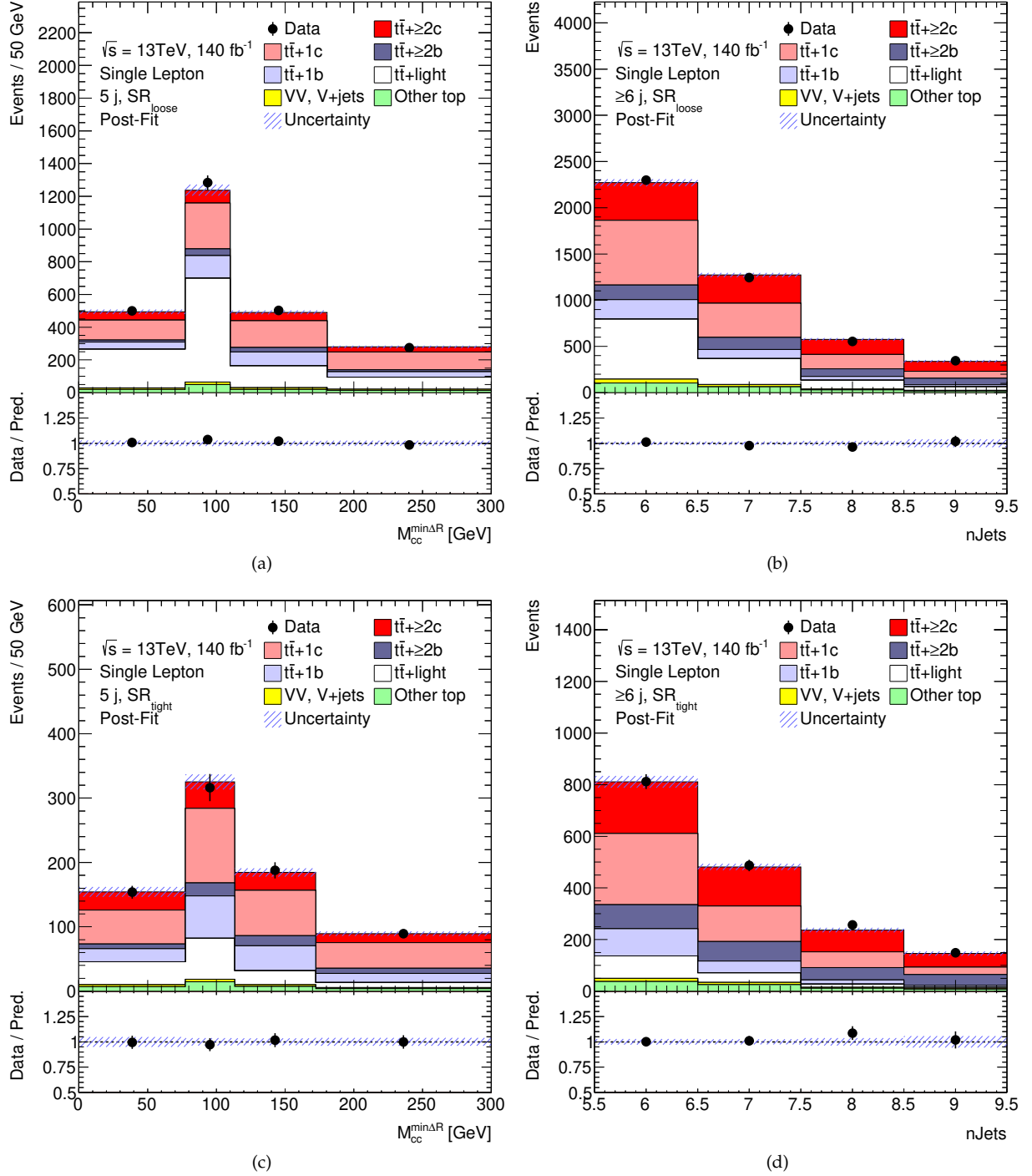


Figure 15.3: Observables used in the fit for the two signal regions, (a,b) SR_{loose} and (c,d) SR_{tight} , post-fit. Both regions are split into (a,c) events with exactly five jets and (b,d) events with at least six jets.

Table 15.4: Expected and observed number of events in the ten fit regions pre- and post-fit. The uncertainties include statistical and systematic uncertainties.

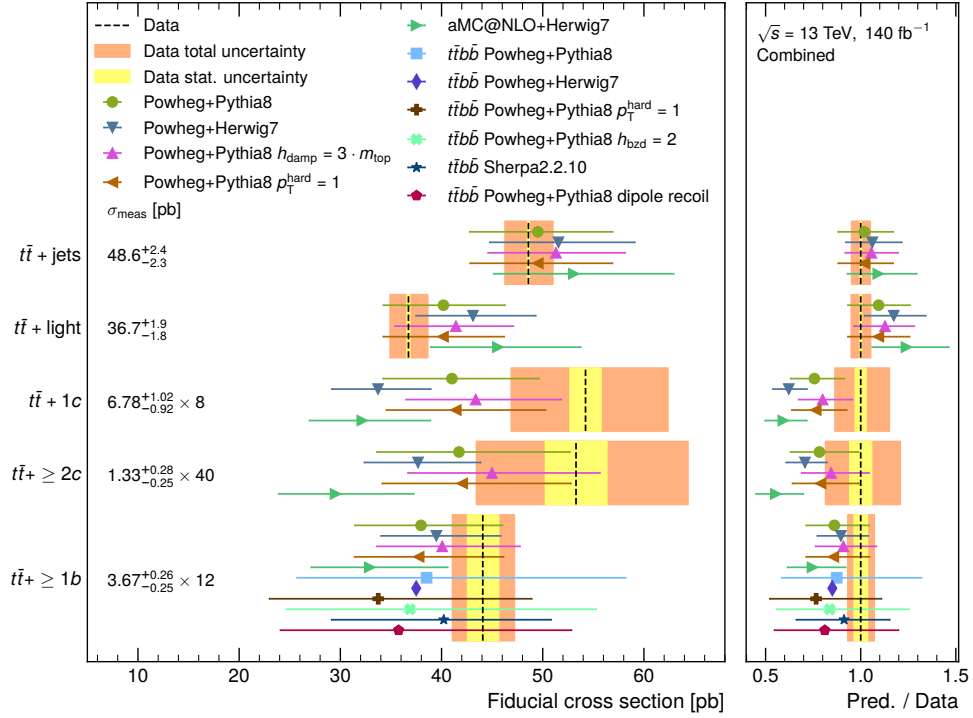
	pre-fit		post-fit		pre-fit		post-fit	
	CR_1^{5j}				$CR_1^{\geq 6j}$			
$t\bar{t} + \geq 2c$	1920±	180	2340±	470	4060±	370	5150±	860
$t\bar{t} + 1c$	10200±	1000	12700±	1800	10200±	1400	13700±	1700
$t\bar{t} + \geq 2b$	1420±	510	1410±	260	2760±	350	3300±	270
$t\bar{t} + 1b$	3400±	1300	4910±	480	4100±	1200	3780±	330
$t\bar{t} + \text{light}$	49600±	6200	44000±	1800	31500±	4100	27100±	1500
$VV, V+\text{jets}$	980±	370	1130±	400	840±	330	990±	360
Other Top	2710±	560	2740±	550	2260±	530	2290±	520
Total	70200±	7100	69170±	270	55700±	6100	56300±	240
Data	69140		69140		56282		56282	
	CR_2^{5j}				$CR_2^{\geq 6j}$			
$t\bar{t} + \geq 2c$	369±	37	480±	100	785±	95	1040±	190
$t\bar{t} + 1c$	2380±	200	3040±	410	2520±	480	3540±	460
$t\bar{t} + \geq 2b$	5400±	2000	5400±	1000	13000±	1500	14980±	660
$t\bar{t} + 1b$	8300±	3400	12500±	1200	10200±	2400	10360±	640
$t\bar{t} + \text{light}$	7070±	890	6520±	300	4730±	730	4160±	330
$VV, V+\text{jets}$	670±	250	770±	270	680±	260	800±	290
Other Top	1610±	370	1630±	390	2320±	600	2290±	610
Total	25800±	4400	30330±	190	34200±	3300	37170±	200
Data	30390		30390		37209		37209	
	CR_3^{5j}				$CR_3^{\geq 6j}$			
$t\bar{t} + \geq 2c$	40±	10	46±	11	120±	31	146±	22
$t\bar{t} + 1c$	115±	25	152±	18	150±	42	207±	28
$t\bar{t} + \geq 2b$	390±	150	378±	73	1250±	180	1426±	80
$t\bar{t} + 1b$	440±	220	652±	75	700±	180	693±	67
$t\bar{t} + \text{light}$	45±	15	55±	11	51±	31	44±	18
$VV, V+\text{jets}$	22±	8	28±	11	32±	13	37±	14
Other Top	80±	22	77±	22	195±	53	185±	52
Total	1130±	300	1382±	28	2500±	270	2738±	44
Data	1345		1345		2728		2728	
	SR_{loose}^{5j}				$SR_{\text{loose}}^{\geq 6j}$			
$t\bar{t} + \geq 2c$	250±	53	279±	58	830±	160	980±	150
$t\bar{t} + 1c$	640±	140	860±	97	920±	250	1300±	180
$t\bar{t} + \geq 2b$	121±	54	114±	27	379±	61	443±	39
$t\bar{t} + 1b$	230±	110	353±	42	410±	130	363±	37
$t\bar{t} + \text{light}$	1120±	220	1152±	89	1170±	290	1070±	140
$VV, V+\text{jets}$	47±	19	57±	21	66±	27	80±	30
Other Top	119±	35	123±	35	211±	50	216±	47
Total	2520±	380	2936±	46	3990±	630	4458±	58
Data	2976		2976		4443		4443	
	SR_{tight}^{5j}				$SR_{\text{tight}}^{\geq 6j}$			
$t\bar{t} + \geq 2c$	124±	29	142±	31	412±	76	489±	75
$t\bar{t} + 1c$	260±	54	348±	38	350±	120	503±	77
$t\bar{t} + \geq 2b$	70±	31	65±	16	227±	36	258±	22
$t\bar{t} + 1b$	110±	45	159±	16	188±	54	173±	19
$t\bar{t} + \text{light}$	117±	38	148±	29	131±	50	144±	37
$VV, V+\text{jets}$	12±	5	15±	5	23±	9	27±	10
Other Top	40±	9	40±	9	85±	24	82±	23
Total	730±	120	917±	24	1410±	220	1675±	36
Data	913		913		1706		1706	

FIDUCIAL CROSS-SECTIONS The actual extracted quantities, so the cross-sections and ratios can be seen in Figure 15.4. Additionally, the predicted cross-sections and ratios of several generated samples are shown. The uncertainties on the values of the generated samples include variations of the renormalisation and factorisation scale, variations of the Var3c parameters, variations of the renormalisation scale for FSR, and PDF variations. Not all uncertainties are available for all samples. Only samples using PYTHIA as PS algorithm have an FSR or Var3c uncertainty assigned to them. Additionally, no uncertainties are available for the $t\bar{t}b\bar{b}$ POWHEG BOX + HERWIG sample. The $t\bar{t}b\bar{b}$ samples are only compared to the $t\bar{t} + \geq 1b$ cross-section, as the remaining $t\bar{t}$ components are not generated in these samples.

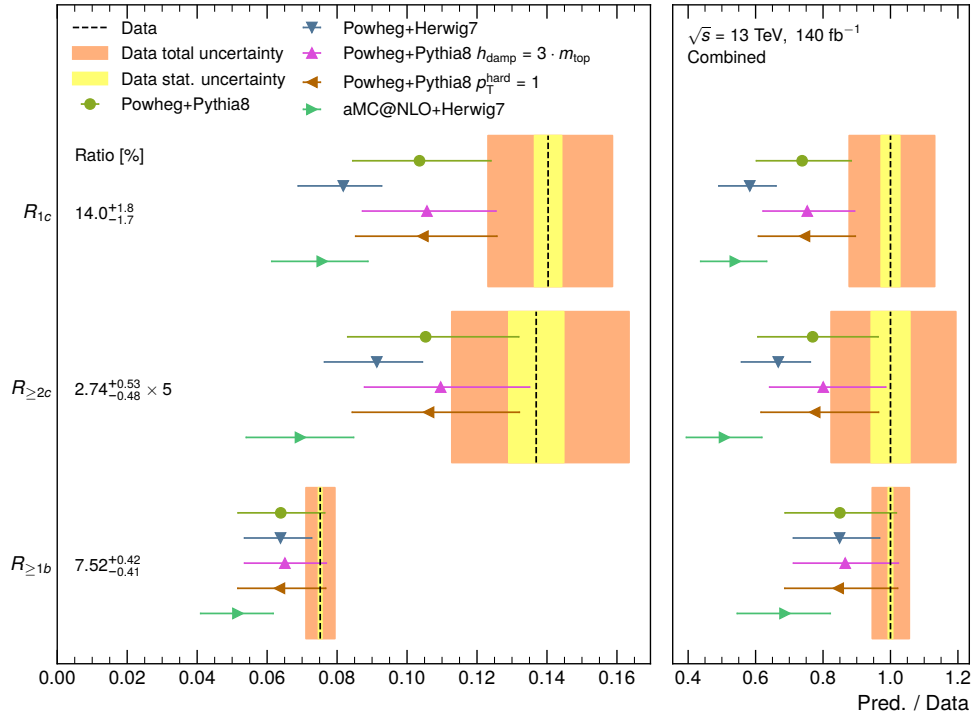
All samples predict a larger cross-section for $t\bar{t} + \text{jets}$ and $t\bar{t} + \text{light}$ than the measured one, while for the remaining six parameters, all samples predict a lower value than the measured one. The sample produced with MADGRAPH5_AMC@NLO + HERWIG consistently deviates the most out of any samples for all cross-sections and ratios. The $t\bar{t}$ POWHEG BOX + HERWIG sample also deviates considerably for the $t\bar{t} + \geq 2c$ and $t\bar{t} + 1c$ cross-sections and ratios, though it is one of the best samples for the $t\bar{t} + \geq 1b$ cross-section and ratio. The nominal $t\bar{t}$ POWHEG BOX + PYTHIA sample is always one of the better samples. The h_{damp} variation and the p_T^{hard} variation do not change the predicted cross-sections significantly. In fact, the h_{damp} -varied sample predicts the $t\bar{t} + \geq 2c$ and $t\bar{t} + 1c$ cross-sections and ratios closest to the measured ones. For the $t\bar{t} + 1b$ cross-section, it is the second closest, with only the $t\bar{t}b\bar{b}$ SHERPA sample being slightly closer. In general, the uncertainties cover the differences between the measured cross-section and the predicted ones.

EXTRAPOLATION TO THE FULL PHASE SPACE Lastly, the obtained fiducial cross-sections are extrapolated to the full phase space. For this purpose, the fiducial cross-section is divided by the acceptance of the fiducial phase space, where the acceptance is the fraction of events entering the fiducial phase space out of all generated events. Only generated events with at least one lepton are considered. To obtain values for all $t\bar{t}$ final states, including all-hadronic final states, the obtained cross-sections need to be divided by the branching fraction of events with one lepton in the final state of around 54%. The ratios in the full phase space are obtained using the extrapolated cross-sections.

As the fiducial cross-sections of all systematic variations are normalised to the nominal one for all $t\bar{t}$ components, an additional extrapolation uncertainty has to be assigned. The relative uncertainty of a single variation is usually obtained by $\frac{A_{\text{alt}} - A_{\text{nom}}}{A_{\text{nom}}}$, where $A_{\text{alt(nom)}}$ is the fiducial acceptance of the alternative (nominal) prediction. The single uncertainties are summed in quadrature to receive the uncertainty on the cross-section. However, the different samples will also predict different cross-sections for the full phase space. These differences should not be taken into account. Therefore, $A_{\text{alt(nom)}}$ is additionally



(a)



(b)

Figure 15.4: Comparison of the measured fiducial (a) cross-sections and (b) ratios to the values predicted by several generated samples.

normalised such that the full phase-space cross-sections of $t\bar{t} + \geq 2c$, $t\bar{t} + 1c$, $t\bar{t} + \geq 1b$, and $t\bar{t} + \text{light}$ are the same for all variations. To obtain the extrapolation uncertainty of the ratios, no acceptance can be defined. Instead, the actual values of the ratios are used, so $\frac{R_{\text{alt}} - R_{\text{nom}}}{R_{\text{nom}}}$. The extrapolation uncertainty is added to the measured uncertainty in quadrature to obtain the uncertainty on the full phase space quantities. It is negligible for the $t\bar{t} + \geq 2c$ and $t\bar{t} + 1c$ cross-sections and ratios, while it is larger than the measured uncertainty for the remaining norm factors.

The extrapolated $t\bar{t} + \text{jets}$ cross-section does not only contain events with additional jets but all $t\bar{t}$ events due to the classification scheme applied. All events with no additional HF jets are labelled $t\bar{t} + \text{light}$, including events with only jets from the $t\bar{t}$ system. Table 15.5 shows the extrapolated cross-sections and ratios with their uncertainties. Using the branching fraction of all-hadronic top-quark pair events, the total cross-section of $t\bar{t}$ production is measured to be $\sigma_{t\bar{t}}^{\text{meas}} = 816_{-63}^{+70}$ pb in good agreement with the theoretically calculated cross-section of $\sigma_{t\bar{t}}^{\text{theory}} = 832 \pm 51$ pb.

Cross-section	Value [pb]	Ratio	Value [%]
$\sigma_{t\bar{t}+\geq 2c}$	$5.6_{-1.1}^{+1.2}$	$R_{\geq 2c}$	$1.27_{-0.24}^{+0.26}$
$\sigma_{t\bar{t}+1c}$	$41.3_{-6.4}^{+6.9}$	R_{1c}	$9.4_{-1.2}^{+1.2}$
$\sigma_{t\bar{t}+\geq 1b}$	$13.3_{-1.2}^{+1.9}$	$R_{\geq 1b}$	$2.97_{-0.29}^{+0.38}$
$\sigma_{t\bar{t}+\text{light}}$	371_{-33}^{+36}		
$\sigma_{t\bar{t}}$	444_{-34}^{+38}		

Table 15.5: Measured parameters extrapolated to the full phase space.

PART IV

TOPOLOGICAL RECONSTRUCTION OF PARTICLE PHYSICS PROCESSES USING GRAPH NEURAL NETWORKS

16 Introduction and motivation

The reconstruction of the top-quark pair system is helpful for certain tasks, like the c-jet calibration within ATLAS [213]. Reconstruction here refers to matching the reconstructed jets to the quarks of the top-quark pair decay. Such a reconstruction would also be beneficial for the $t\bar{t}$ + charm analysis. Identifying the jets originating from the W boson can reduce the contamination of c-jets originating from the W boson in the selection. This part will introduce a new ML approach, the Topograph, to reconstruct the top-quark pair system in the all-hadronic decay channel. However, the approach can also be extended to other decay topologies.

Contents

16.1 Current approaches	171
16.2 The Topograph	173

16.1 Current approaches

In many physics analyses involving top-quarks, the reconstruction of the decay system is a key part of the measurement [43, 212, 214–227]. Common approaches are the χ^2 method [212, 214, 221] and the Kinematic Likelihood Fitter (KLFFitter) [228]. Both methods test all possible permutations of matching the reconstructed jets to the partons of the decay. A score is assigned to each permutation, taking kinematic constraints based on the masses of the reconstructed W bosons and top-quarks into account. Flavour tagging information can be included to enhance the efficiency of matching all jets correctly to the partons. KLFFitter additionally applies so-called transfer functions to remove effects of the PS, the hadronisation, and the detector response. These functions are analytical and approximate.

While good performance can be achieved with these approaches, they suffer from a high computational cost due to a combinatoric explosion. For a top-quark pair event with both top-quarks decaying hadronically, six jets need to be matched to the correct partons. With six reconstructed jets, $6! = 720$ permutations are possible. However, multiple symmetries can be applied to reduce the number of permutations. The decay is symmetric under exchange of the two top-quarks and under exchange of the two jets of the W boson decays. Therefore, only $6!/8 = 90$ permutations need to be tested. However, many events have more than just six reconstructed jets. With seven reconstructed jets, the number of permutations to test increases to $7!/8 = 630$, and to $8!/16 = 2520$ with eight reconstructed jets².

Apart from the possible inclusion of flavour tagging information, these approaches only use information of the mass of the jet doublets (triplets) forming the W bosons (top-quarks). Due to the implementation, combinations with masses closest to the mass of the W boson and the top-quark are preferred. This introduces a bias towards the mass of the W boson and the top-quark while also assuming that the top-quarks and W bosons are on-shell.

² An additional symmetry in the exchange of the jets not assigned to any parton of the top-quark pair system can be employed.

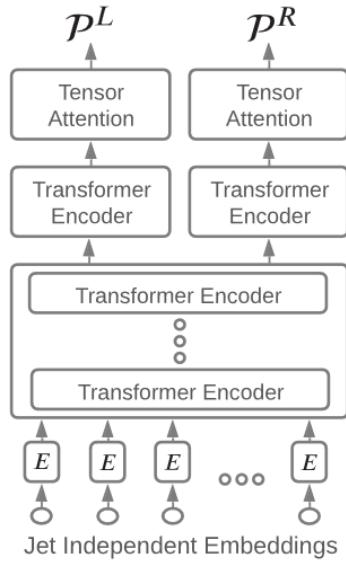


Figure 16.1: High-level overview of the SPA-net architecture. Taken from Ref. [191]

Instead of using just the masses of the doublets (triplets), ML can be used to exploit more underlying information [229–233]. However, these approaches still test every possible permutation and choose the permutation that is most likely. Therefore, these approaches still suffer from the combinatoric explosion. A different approach considers only events with a fixed jet multiplicity [234]. It includes so-called Lorentz Layers [232, 235], which employ Lorentz-invariance when using four vector momenta as inputs.

An approach exploiting a physics prior is the Matrix Element Method [219, 233, 236, 237]. It calculates the likelihood of observing an event based on the ME of a process. Again, all possible permutations need to be tested before choosing the best one as the result. On top of the combinatoric explosion, the calculation of the likelihood of an event based on a ME is exceedingly slow since an integral must be computed over the entire phase space of possible final state particle momenta. Additionally, the effects of the PS, the hadronisation, and the detector have to be removed before the calculation by applying transfer functions similar to KLFitter.

State-of-the-art results are achieved using transformer-inspired architectures [191, 192, 238]. The model compared to in this work is SPA-net [191, 192].

SPA-NET A high-level overview of the SPA-net architecture can be seen in Figure 16.1. It uses transformer encoder layers and so-called tensor attention. In a first step, the jets are passed through an embedding network before being processed by several transformer encoder layers. Two branches of the same architecture receive the output of these layers, one to predict a triplet for the top-quark and the other to predict a triplet for the anti-top-quark. It is not fixed which triplet, $\mathcal{P}^L$ or $\mathcal{P}^R$, represents the top-quark and the anti-top-quark. Instead, a loss that is symmetric under the exchange of the two triplets is used. This is motivated by the symmetry under the exchange of the two top-quarks in top-quark pair events, with both top-quarks decaying hadronically. The branches start with an additional transformer encoder layer before the main component, the tensor attention layer. The tensor attention outputs a score for each triplet of jets, with the score being invariant under the exchange of the two jets assigned to the quarks of the W boson decay. The triplet with the highest score is assigned to the respective top-quark.

Each tensor attention layer consists of one three-dimensional tensor $\theta \in \mathbb{R}^{D \times D \times D}$ where D is the number of features of each jet. The values of this tensor are learnable parameters of the model. This tensor is symmetrised by

$$S^{ijk} = \frac{1}{2} (\theta^{ijk} + \theta^{jik}) \quad (16.1)$$

to encode the symmetry under the exchange of the two jets of the W boson. Using this tensor, the three-dimensional attention O^{ijk} is formed:

$$O^{ijk} = X_n^i X_m^j X_l^k S^{nml} \quad (16.2)$$

with $X \in \mathbb{R}^{N \times D}$ being the embedded jets, where N is the number of jets. Applying a softmax activation to this attention produces the output scores of the model. It can be shown that the complexity of the problem can be reduced to $\mathcal{O}(N^2)$ [192]. This approach is easily extendable to other decays like $t\bar{t}H$ or $t\bar{t}t\bar{t}$ by adding additional branches, as shown in Ref. [192].

16.2 The Topograph

While GNNs have only recently gained traction within the HEP community, graph-like structures have been used in HEP for a long time. The Feynman diagram of a process can be understood as a graph with vertices representing particle interactions and edges representing particles themselves. This structure can easily be converted into one in which vertices represent particles and edges represent particle interactions instead. An example of a top-quark pair decay can be seen in Figure 16.2.

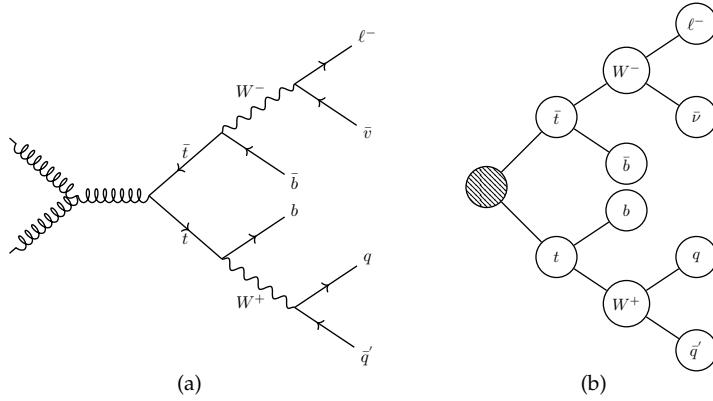
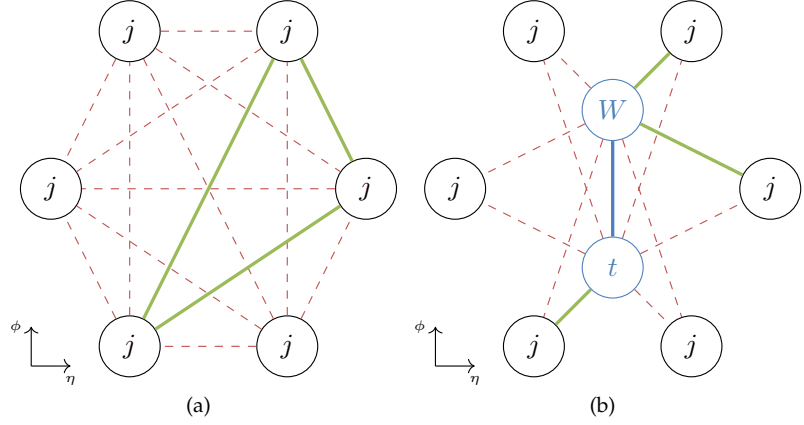


Figure 16.2: Graph representations of top-quark pair production with one top-quark decaying semi-leptonically and the other hadronically. Taken from Ref. [3].

In the Topograph algorithm, a graph is built based on the graph representation of the Feynman diagram in Figure 16.2b. This requires the injection of intermediate particles into the graph. Their initial values can be obtained either by using learnable weights or by utilising information from the observed final state particles. Edges are formed between the final state particles (daughter particles) and the intermediate particles (mother particles). By only predicting whether an edge between a daughter and mother particle is a true edge, the number of predictions is given by $M \cdot N$ where M is the number of injected intermediate particles and N the number of final state particles. Using a fully-connected graph with edges to determine whether two final state particles share a common origin instead, leads to $N(N - 1)$ predictions. As long as there is at least one more final state particle than there are injected particles ($N > M + 1$) this leads to a lower number of overall predictions using the Topograph approach. A comparison of these two approaches for a single top-quark decay and six reconstructed jets can be seen in Figure 16.3.

Additionally, the injected nodes can be used to predict true proper-

Figure 16.3: Comparison of the edges in (a) fully-connected GNNs and (b) Topographs for a single top-quark decay. The injected nodes in the Topograph for the top-quark (t) and the W boson (W), as well as the predefined connection between them, is coloured in blue. True edges, either edges between jets sharing an origin or edges between true mother and daughter particles are coloured in green. All other edges are dashed red lines. Taken from Ref. [3].



ties, like the four momentum vector of the particles they represent. The results of this auxiliary regression task can potentially be used to enhance the resolution and accuracy of differential measurements with respect to intermediate particle kinematics.

If multiple different types of final state objects like jets, muons, or electrons are present in the event, they can be embedded into a common space using an embedding network for each object type. With such an embedding, all embedded objects can be treated the same and be connected to all possible mother particles. However, this could lead to physically impossible results, like a jet and an electron originating from the same W boson. An alternative is to pre-define fixed connections before the training. For example, in top-quark pair decays with one top-quark decaying semi-leptonically and one top-quark decaying hadronically, the lepton has to come from the W boson of the semi-leptonically decaying top-quark. This can be enforced by a fixed connection between the lepton and the W boson. Similarly, the missing transverse momentum could be represented as a node with a fixed connection to the W boson. As an alternative to using the missing transverse momentum directly, a neutrino estimate could be obtained using a method like ν -Flows [239]. In principle, this could even be included in the training process, training ν -Flows and Topograph in parallel.

17 Topographs for top-quark pair decays in the all-hadronic final state

To test the Topograph algorithm, it is applied to top-quark pair production with both top-quarks decaying hadronically. It is compared to the χ^2 method and SPA-net. The χ^2 method is a non-ML approach used in many top-quark analyses, while SPA-net reaches state-of-the-art performance.

Contents

17.1	Datasets	175
17.2	Implementation	178
17.2.1	Topograph	180
17.2.2	Topotransformer	181
17.3	Implementation of reference models	182
17.3.1	χ^2 method	182
17.3.2	SPA-net	183
17.4	Results	183
17.4.1	Jet parton assignment	183
17.4.2	Interpretation of edge scores	189
17.4.3	Regression tasks	191

17.1 Datasets

GENERATION To compare the different methods on the same data, a custom sample² of 20 million top-quark pair events was generated at a centre-of-mass energy of $\sqrt{s} = 13$ TeV. The sample was generated with the MADGRAPH5_AMC@NLO 3.1.0 generator with decays of top-quarks and W bosons being handled by MADSPIN. Both W bosons were required to decay hadronically. The events were interfaced to PYTHIA 8.243 for the PS and hadronisation. The modelling of the detector response was performed by DELPHES 3.4.2 [241] with the default ATLAS configuration provided by DELPHES. Just as in the ATLAS experiment, the anti- k_t algorithm of the FASTJET package was used for jet clustering using a radius parameter of $\Delta R = 0.4$. A binary decision on whether a jet was tagged as a b -jet or not is provided with a p_T dependence similar to that of a previous ATLAS b -tagging WP [242]. Overall the efficiency of tagging b -jets is around 70%. No actual information of the jet is used, but instead the decision is based on random numbers.

An additional smaller sample was generated using PYTHIA 8.307 for both the ME generation and the PS evolution to evaluate the models on a sample with a different ME generator. The Monash tuned set of parameters [243] was used at LO accuracy to generate a total of 10 million events.

Only events with at least six reconstructed jets with $p_T > 28$ GeV and $|\eta| < 2.5$ are considered. Up to 16 jets ordered by their transverse momentum are kept for each event. Jets are truth matched to the

²The dataset is available at <https://zenodo.org/record/7737248> [240]

partons in the hard scatter using a cone of $\Delta R < 0.4$. If multiple jets are matched to a single parton, or multiple partons are matched to a single jet, the event is vetoed.

PARTIAL EVENTS The training of the ML models is performed using either only fully-reconstructable events or including so-called partial events. An event is classified as fully-reconstructable if all six partons originating from the top-quark pair decay are matched unambiguously to a jet. Partial events also include events with fewer matched partons. For the SPA-net model only events where all partons from one of the top-quark decays are matched to a jet can be used, while for Topographs events with all partons from one W boson decay are also considered. This difference is introduced by the values the models predict. As SPA-net predicts a score for triplets of jets, all three of the jets need to be truth-matched to a parton to be considered a true triplet, whereas for Topographs, technically only a single parton needs to be matched to a jet as Topographs makes predictions on an edge level basis. For simplicity, in this work only events with at least one reconstructable W boson are considered in the training for Topographs.

TRAINING STATISTICS For training with partial events, 4 550 000 (4 150 000) events are available for Topographs (SPA-net) with an additional 241 000 (220 000) events for validation. When only considering fully-reconstructable events, the number of events reduces to 1 340 000 training and 71 000 validation events. Additionally, 298 000 events are reserved for testing. For these, no requirement on having a certain number of jets matched to a parton is applied. 76 000 of these events are fully-reconstructable. Out of all events for testing, 147 000 have at least two jets tagged as b -jets with 44 000 being fully-reconstructable.

The additional dataset generated with PYTHIA for the ME generation and the PS evolution is only used for testing. After the selection, 1 490 000 events remain of which 226 000 are fully reconstructable while also having at least two b -jets.

DATASET DISTRIBUTIONS To train the ML models, the four momentum vector (p_x, p_y, p_z, E) is used together with a boolean flag indicating the result of the b -tagging algorithm. The energy is scaled by applying the logarithm. Afterwards, each component is normalised separately using a StandardScaler. No preprocessing is applied to the boolean flag. Additional features of the jets, like a quark-gluon tagging score, could easily be added to all ML models, as jets are represented by single nodes in a point cloud.

Figure 17.1 shows the number of jets and the number of b -tagged jets in both samples. For the MADGRAPH5_AMC@NLO + PYTHIA sample, the training dataset is used, whereas for the PYTHIA sample, all generated events are considered. For both samples, only events that are fully-reconstructable are shown. Figure 17.2 shows the invariant mass of the W bosons and the top-quark mass when using

the reconstructed truth-matched jets to calculate the invariant mass. The true three momentum (p_x, p_y, p_z) of the particles after subtracting their mean and dividing by the standard deviation is shown as well. These values are used as targets in the regression part of the Topographs architecture. Lastly, Figure 17.3 shows all input distributions of the jets. The jets are split into jets truth-matched to the b -quark of a top-quark decay (top-jets), jets truth-matched to a quark originating from a W boson decay (W -jets), and additional jets which are not truth-matched to any parton. A double-peak structure can be observed for the p_x and p_y distribution caused by the requirement of $p_T > 28$ GeV. Comparing the distributions of the two samples, no large differences can be observed. The largest difference can be seen for the reconstructed invariant mass, for which the top-quark mass is shifted towards lower values for the PYTHIA sample, whereas the W boson mass is shifted towards larger values.

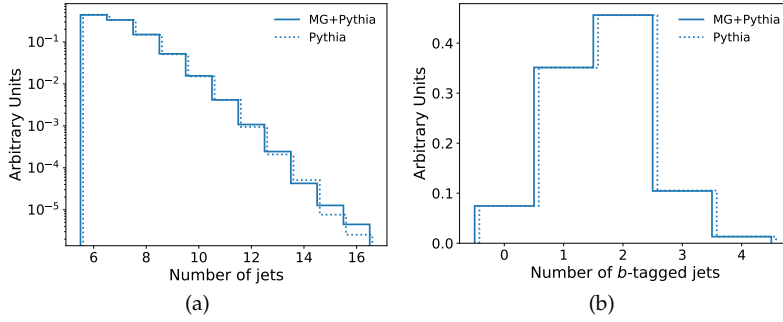


Figure 17.1: Distribution of (a) the jet and (b) b -jet multiplicity in the two generated samples. The first and last bin contain under- and overflow events, respectively. For visualisation purposes, the histogram of the PYTHIA sample is shifted.

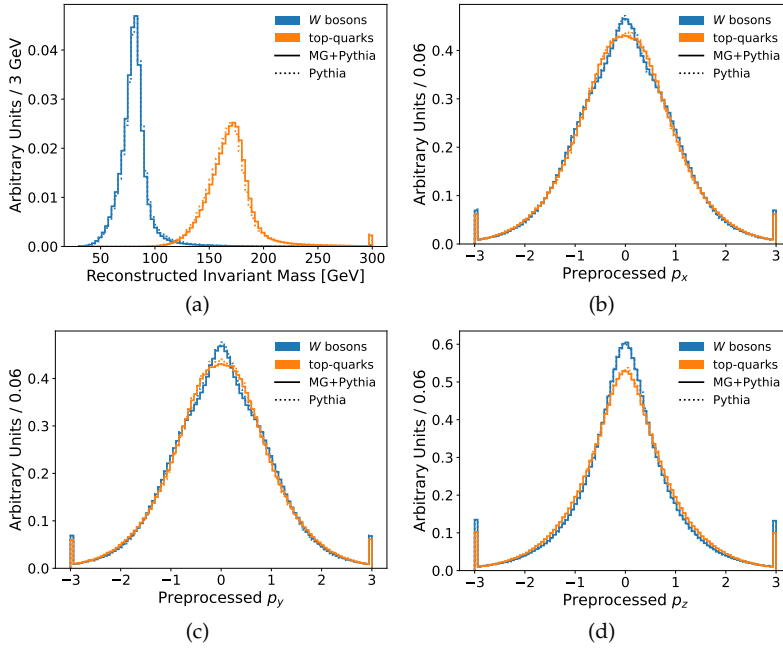
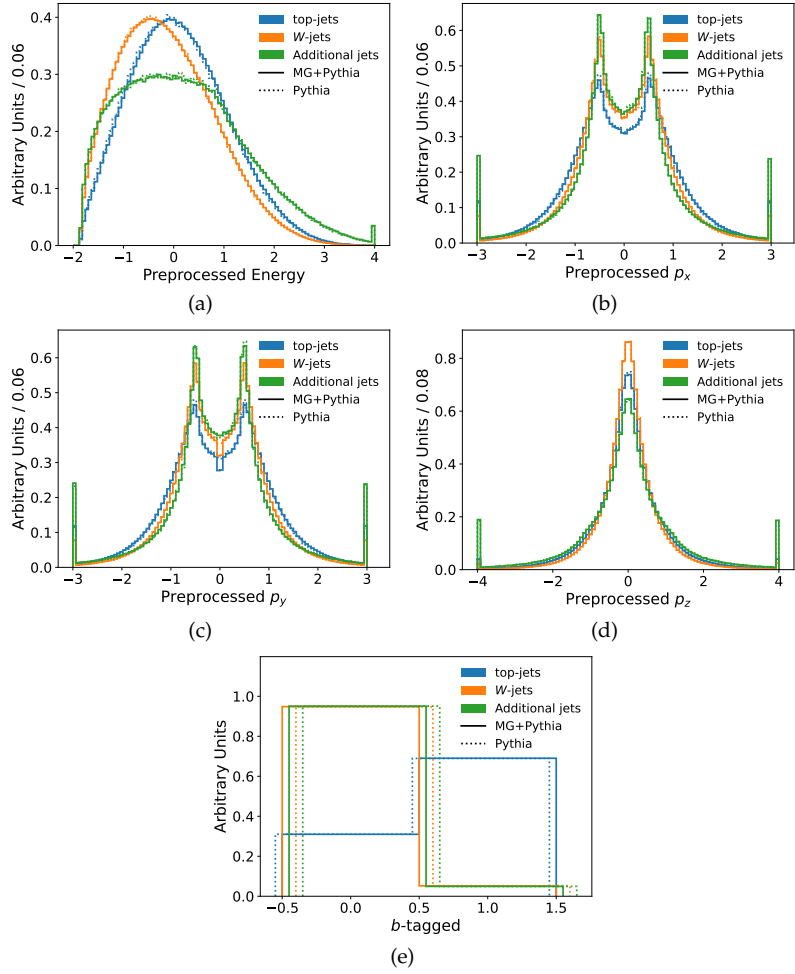


Figure 17.2: Distribution of (a) the reconstructed invariant mass, (b) p_x , (c) p_y , and (d) p_z of the W bosons and top-quarks in both generated samples. The values for p_x , p_y , and p_z are shown after the preprocessing is applied. The first and last bin contain under- and overflow events, respectively.

Figure 17.3: Distribution of (a) the energy, (b) p_x , (c) p_y , (d) p_z and (e) the b -tagging decision for all jet in the two generated samples. The energy, p_x , p_y , and p_z are shown after the pre-processing is applied. For the b -tagging decision, the histograms are slightly shifted versus each other for visualisation purposes. The first and last bin contain under- and overflow events, respectively.



17.2 Implementation

Two different versions of the Topograph algorithm are tested: the first one uses INs to exchange information between nodes, while the second uses transformer encoder layers instead. The first one will be referred to as Topograph (TG), the second one as Topotransformer (TT). For both models, information is exchanged in an initial step between only the jets. Afterwards, the nodes representing the W bosons and top-quarks are initialised. As the top-quark always decays into a W and a b -quark, a connection between each W node and a top node is enforced. This uniquely assigns one W boson to a top-quark. Using the initialisation representation, the true properties of the W bosons and the top-quarks are predicted. The same DNN is used for the two top (W) nodes, but different DNNs are used for the top nodes and the W nodes. Information is now exchanged between all three types of nodes, the jets, W bosons, and top-quarks, for a total of n times. After each information exchange, the edges are classified and the true particle properties of the W bosons and top-quarks are predicted. This results in a total of $2n + 1$ loss terms which are summed together.

EDGE CLASSIFICATION To classify the edges between jets and W bosons (top-quarks), the features of a jet node and a W (top) node are concatenated. This is done for all jets and all W (top) nodes. These edges are passed through a DNN resulting in a single output value for each edge. The output is passed through a sigmoid activation function such that the output score can be interpreted as the probability of the jet originating from the particle of the connected node. The binary cross-entropy is used as the loss function.

An edge is labelled as true if the jet originates from the connected respective particle. Therefore, edges between a jet and a top node are labelled as true if the jet is the top-jet of this top-quark. Similarly, edges between a jet and a W node are labelled as true if the jet is a W -jet of this W boson. All other edges are labelled as false. As this labelling results in a lot more false edges than true edges, the contribution of each edge to the loss is weighted such that the overall contribution of true and false edges is the same. This is done to avoid any bias towards lower values of output.

PARTICLE REGRESSION The W and top nodes are used to regress towards the true properties of the particle they represent. As the target, the three momentum (p_x, p_y, p_z) with additional preprocessing is used as can be seen in Figure 17.2. The energy is not included to reduce possible sculpting towards the mass of the respective particle. The MAE is used as loss function.

SYMMETRIC LOSS FUNCTIONS The overall loss is calculated by taking the sum of the loss values of each edge classification and regression task with potential weighting between them. Since the decay of the top-quark pair is symmetric¹, meaning the two top-quarks can be exchanged, the loss function has to reflect this behaviour. Otherwise an implicit ordering would be introduced by the network, which might hinder its ability to solve the problem. For this purpose, the loss value is calculated for both possible configurations, the first top node being the top-quark or the first top node being the anti-top-quark. The lower of these two values is taken as the loss value used to optimise the weights of the NN. As each W boson node has a fixed connection to a single top node, the contribution of the W bosons does not require additional symmetrisation. A single contribution to the loss can therefore be written as

$$\begin{aligned} \mathcal{L}_{\text{symm}} = \min(&\mathcal{L}(t_1, t) + \mathcal{L}(W_1, W^+) \\ &+ \mathcal{L}(t_2, \bar{t}) + \mathcal{L}(W_2, W^-), \\ &\mathcal{L}(t_1, \bar{t}) + \mathcal{L}(W_1, W^-) \\ &+ \mathcal{L}(t_2, t) + \mathcal{L}(W_2, W^+)), \end{aligned} \quad (17.1)$$

where $\mathcal{L}(p_i, p)$ is the loss value of either the regression or the edge classification task with nodes p_i and truth particles p .

To train on partial events, the loss terms including missing particles are excluded from the loss calculation. For instance, if there is no

¹ The production of top-quark pairs in pp collisions is symmetric for gluon-gluon fusion at LO. Production through quark-antiquark annihilation as well as NLO diagrams can introduce asymmetries. However, for the purpose of the reconstruction, the production can be considered completely symmetric.

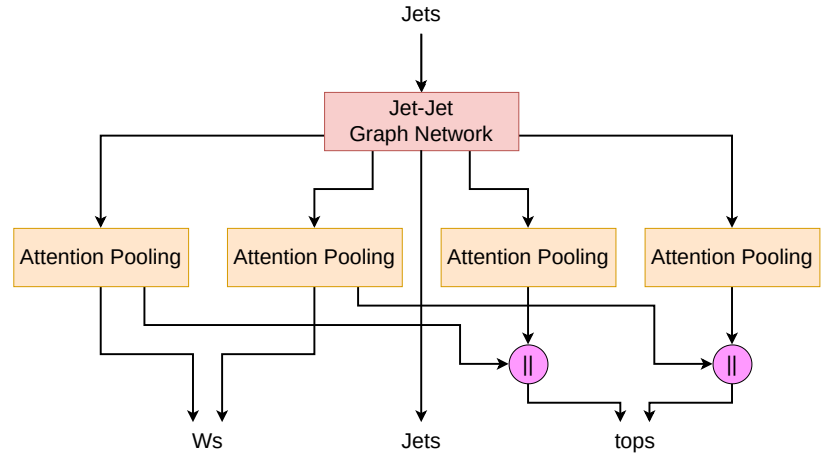
top-jet of the anti-top-quark, the terms containing the anti-top-quark, $\mathcal{L}(t_2, \bar{t})$ and $\mathcal{L}(t_1, \bar{t})$ in Equation (17.1), are set to zero.

PARTON ASSIGNMENT The edge classification results in $6 \cdot N$ output scores. Out of these, the best assignment has to be extracted. Many options could be considered. For simplicity, an iterative approach was chosen. The edge with the overall highest score of the $6 \cdot N$ scores is selected, meaning the jet is taken as originating from the connected particle. Afterwards, the edge with the next highest score is selected. To avoid ambiguities, each jet can only be assigned to one particle and each W boson (top-quark) can only have two (one) jets associated to it.

17.2.1 Topograph

The initialisation step for the TG model can be seen in Figure 17.4. As a first step, m iterations of message passing between the jets is performed using an IN connecting all jets with each other. The output of this is used in four different networks used to obtain weights for attention pooling: two for the two W nodes, and two for the two top nodes. The pooled W nodes are concatenated to the pooled top nodes in order to obtain the initialised top nodes. The first W node is concatenated to the first top node, while the second W node is concatenated to the second top node.

Figure 17.4: Sketch of the initialisation part of the TG. The "||" represents a concatenation operation.



The resulting jet, W , and top nodes are further processed as shown in Figure 17.5. To update the jet nodes, three INs are used: one between the jets themselves, one between jets and W s, and one between jets and tops. The output of the three networks is concatenated and passed through a DNN. For the W nodes and top nodes, only two INs are used: one between the particle node and the jets, and one between the particle node and the other particle node. No message passing is performed between the W nodes or top nodes themselves, as it was found not to improve the performance. Messages between the W nodes and the top nodes are only passed along the previously

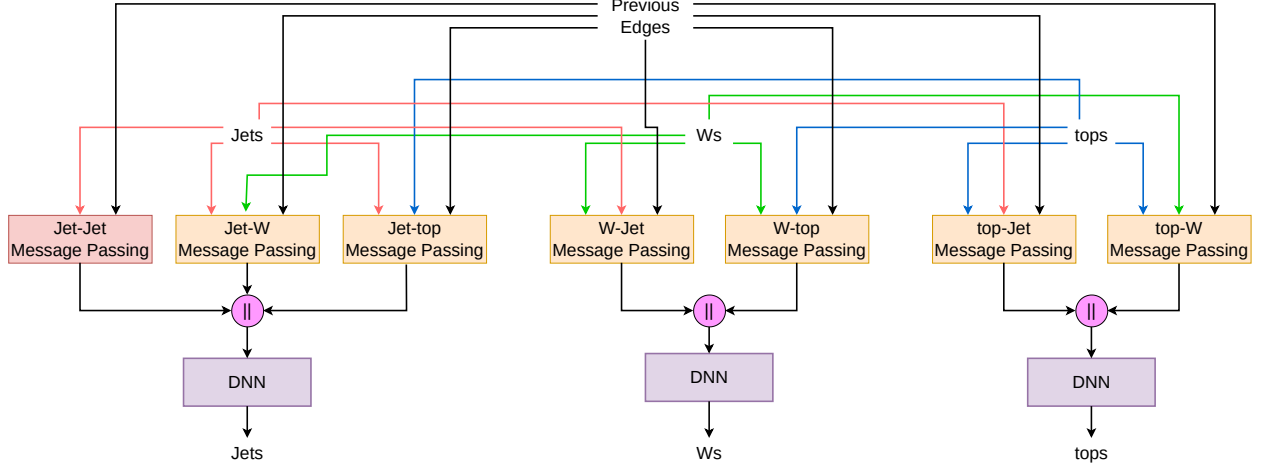


Figure 17.5: Sketch of the updates during the TG training. For message passing operations with different inputs, the receiving nodes enter to the left, the sending nodes in the middle, and potential previous edges to the right. If sending and receiving nodes are the same, they enter in the middle. The "||" represents a concatenation operation.

described fixed connection. As for the jets, the outputs are concatenated and passed through a DNN. For all INs, the values of the same edges from previous iterations are concatenated onto the edges. This block is repeated n -times, classifying edges and regressing towards the truth particle properties with each iteration.

The TG models are trained for 100 epochs with the *AdamW* optimizer, choosing the weights of the epoch with the lowest validation loss. A cosine annealing schedule with warm restarts [244] is used for the LR. The evolution of the LR over the training can be seen in Figure 17.6. A list of all hyperparameters used can be seen in Table D.1. The models are trained using TensorFlow v2.10.

17.2.2 Topotransformer

For the TT, the jets are first passed through an embedding network, see Figure 17.7. The output of the embedding network is passed through m encoder layers to exchange information between the jets. These updated jets are used to initialise the W nodes and top nodes by using cross-attention with learnable parameters being used as the query. The output of the cross-attention layer for the W nodes is taken as is, whereas for the top nodes the W nodes are concatenated onto it. As the feature dimension doubles by this concatenation, an additional DNN is used to reduce the feature dimension of the top nodes to the same size as the W nodes and the jet nodes.

The processing after the initialisation step can be seen in Figure 17.8. Jets are further updated with three different transformer encoder layers: the first one uses self-attention with only jets, the second (third) one uses cross-attention with the W nodes (top nodes) used as keys and values. The outputs of the three networks are concatenated and passed through a DNN. To update the W nodes (top nodes) a single transformer encoder layer using cross-attention with the jets is used. The output of this layer is concatenated with the original top nodes (W nodes) and passed through a DNN.

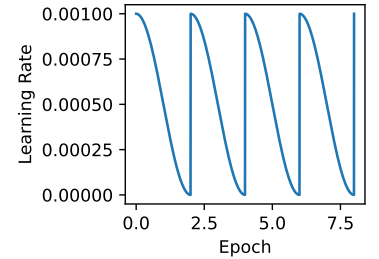


Figure 17.6: Evolution of the LR during the TG training. The periodic part is continuously repeated for the following epochs.

Figure 17.7: Sketch of the initialisation part of the TT. For cross-attention encoder layers, the query always enters in the middle, whereas the key and value enter to the right. The "||" represents a concatenation operation.

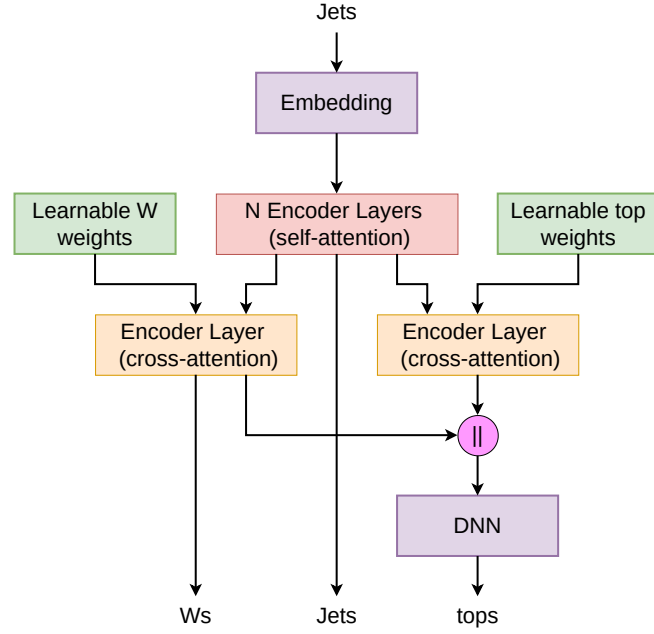


Figure 17.8: Sketch of the updates during the TT. For cross-attention encoder layers, the query always enters in the middle, whereas the key and value enter to the right. The "||" represents a concatenation operation.

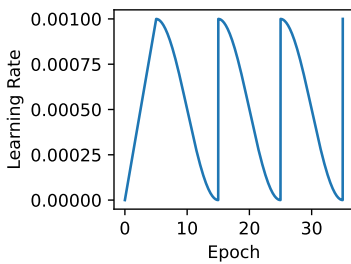
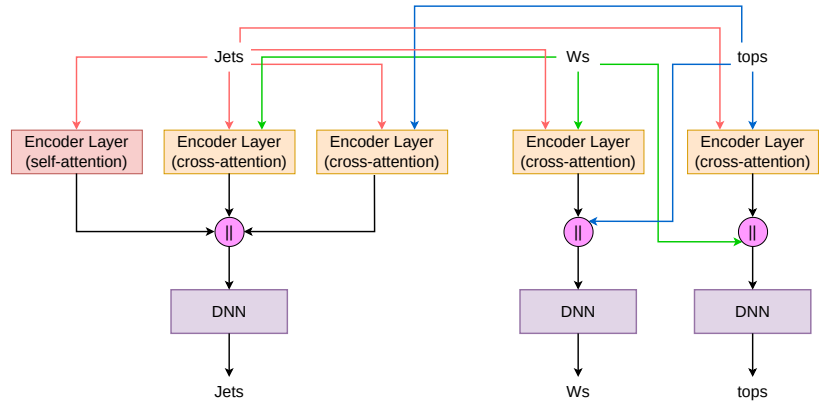


Figure 17.9: Evolution of the LR during the TT training. The periodic part is continuously repeated for the following epochs.

As for the TG models, the TT models are trained for 100 epochs using the *AdamW* optimizer. A similar LR schedule is used, though in the first five epochs a linear warm-up is added. Additionally, the period is chosen to be longer. The evolution can be seen in Figure 17.9. A list of all hyperparameters used can be seen in Table D.1. The models are trained using Pytorch v2.0.

17.3 Implementation of reference models

17.3.1 χ^2 method

To obtain the assignment using the χ^2 method, the χ^2 score is calculated for all possible permutations of jets taking the permutation with

the lowest score as the assignment. It is defined as

$$\chi^2 = \frac{(m_{b_1q_1q_2} - m_t)^2}{\sigma_t^2} + \frac{(m_{b_2q_3q_4} - m_t)^2}{\sigma_t^2} + \frac{(m_{q_1q_2} - m_W)^2}{\sigma_W^2} + \frac{(m_{q_3q_4} - m_W)^2}{\sigma_W^2}. \quad (17.2)$$

Here $m_{b_iq_jq_k}$ is the invariant mass of the jets b_i , q_j , and q_k , m_t (m_W) is the mass of the top-quark (W boson), and σ_t (σ_W) is the width of the top-quark (W boson). The mean of the reconstructed invariant masses of the top-quark (W boson) is used for m_t (m_W), while the standard deviation is used for σ_t (σ_W). The values obtained from the MADGRAPH5_AMC@NLO + PYTHIA dataset are: $m_t = 169.8$ GeV, $m_W = 81.0$ GeV, $\sigma_t = 29.0$ GeV, and $\sigma_W = 18.5$ GeV. The same values are used for the evaluation on the PYTHIA dataset.

To improve the accuracy of the jet assignment, only permutations in which the top-jets are tagged with the b -tagging algorithm are considered. No b -tagging requirements are placed on W -jets.

17.3.2 SPA-net

The implementation provided by the SPA-net authors is used to train models using the SPA-net architecture¹. A small hyperparameter scan was performed using only fully-reconstructable events. The values given in Ref. [191] while using v1.0 of the software led to the highest performance on the nominal MADGRAPH5_AMC@NLO + PYTHIA dataset. The same hyperparameters are used for the training including partial events.

¹ <https://github.com/Alexanders101/SPANet/tree/v1.0>

As for the TG and TT models, all models were trained for a maximum of 100 epochs. For the predictions on the test set, the weights resulting in the lowest validation loss are used.

17.4 Results

If not specified otherwise, only events with at least two b -tagged jets are considered for the evaluation of the performance, since this is a common requirement for top-quark physics analyses in order to suppress the multi-jet background. For the training, this requirement was not imposed in order to maximise the available statistics.

17.4.1 Jet parton assignment

The event reconstruction efficiencies for different jet and b -jet multiplicities of the different models are shown in Table 17.1. They are calculated by dividing the number of events in which all jets are correctly assigned to their partons by the total number of all events. For each of the ML methods, SPA-net, TG, and TT, five models are trained, only differing by the random initialisation of the weights. The mean efficiencies of the models are shown as well as the best model, with best defined as the model with the highest efficiency in events with at

least six jets and at least two b -jets. The uncertainties shown are the uncertainty on the mean, and the binomial uncertainties, respectively.

The three different ML methods have very similar performance with differences only on the percent level. However, all consistently outperform the χ^2 method by around 10%. Larger improvements can be seen for higher jet multiplicities. The TG method consistently has the best performance with the two transformer based methods having slightly lower performance, with larger differences observed at higher jet multiplicities.

N_{jets}		6		7		≥ 6	
$N_{b\text{-jets}}$		2	≥ 2	2	≥ 2	2	≥ 2
Mean	SPA-net	81.43 $\pm$ 0.03	79.45 $\pm$ 0.07	64.47 $\pm$ 0.08	62.38 $\pm$ 0.12	68.48 $\pm$ 0.06	65.76 $\pm$ 0.08
	TG	81.54 $\pm$ 0.12	79.76 $\pm$ 0.14	65.46 $\pm$ 0.20	63.51 $\pm$ 0.25	69.12 $\pm$ 0.19	66.59 $\pm$ 0.22
	TT	81.51 $\pm$ 0.05	79.58 $\pm$ 0.07	64.66 $\pm$ 0.12	62.30 $\pm$ 0.13	68.44 $\pm$ 0.09	65.60 $\pm$ 0.12
Best	SPA-net	81.47 $\pm$ 0.31	79.56 $\pm$ 0.30	64.69 $\pm$ 0.44	62.86 $\pm$ 0.40	68.65 $\pm$ 0.25	66.05 $\pm$ 0.23
	TG	81.70 $\pm$ 0.31	79.95 $\pm$ 0.29	65.90 $\pm$ 0.44	63.94 $\pm$ 0.40	69.54 $\pm$ 0.25	67.03 $\pm$ 0.22
	TT	81.70 $\pm$ 0.31	79.81 $\pm$ 0.29	64.92 $\pm$ 0.44	62.55 $\pm$ 0.40	68.73 $\pm$ 0.25	65.96 $\pm$ 0.23
	χ^2	72.73 $\pm$ 0.36	70.94 $\pm$ 0.33	54.28 $\pm$ 0.46	52.11 $\pm$ 0.41	58.57 $\pm$ 0.26	55.90 $\pm$ 0.24

Table 17.1: Event reconstruction efficiencies [%] for the SPA-net model, the TG model, the TT model, and the χ^2 method in different jet and b -jet multiplicities. The highest efficiency is highlighted in bold. Adapted from Ref. [3].

Figure 17.10 shows the invariant mass of the reconstructed W boson and top-quark using the jet assignment of the different methods. Events are classified according to whether all jets were correctly matched to the respective W boson or top-quark (*correct*). If not all jets were correctly matched, the event is classified as either *incorrect*, if all partons of the W boson or the top-quark are reconstructed as a jet, or as *impossible*, if at least one of the partons could not be reconstructed. *Incorrect* top-quarks also contain assignments where the three jets from the top-quark were assigned to the correct triplet but the b -jet was assigned to the W boson and one of W -jets to the top-quark. This results in the correct mass being reconstructed but the event being labelled as *incorrect*. These events make up around 5% for the SPA-net, TG, and TT models, and around 3% for the χ^2 method. All methods produce peaks at the W boson or top-quark mass for all event categories. The distributions of the three ML methods are very similar, indicating no difference in potential sculpting of the mass distribution towards the W boson and top-quark mass. The distribution of the χ^2 method are slightly shifted towards higher values compared to the three ML methods. However, no stronger sculpting effect can be seen. To properly evaluate the potential sculpting of the mass distribution, simulated samples of background events such as multi-jet events would be needed.

The ability of the models to correctly assign jets in events in which not all partons are reconstructed as a jet is also of interest. The efficiencies of correctly assigning only the partons present in the event are shown in Table 17.2. The TG model consistently has the highest performance, except for events where only three (four) partons are matchable and at least one (two) of those is (are) not being correctly

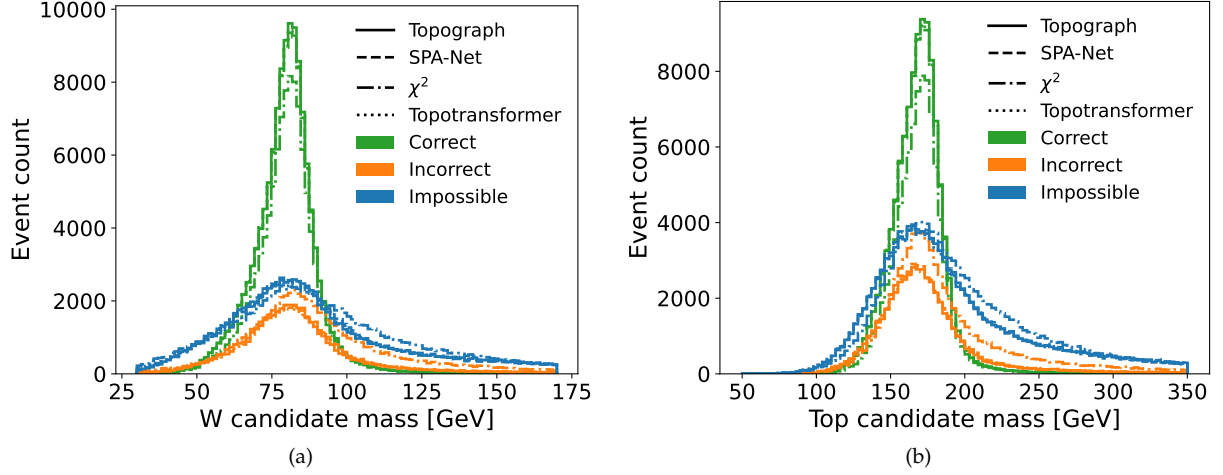


Figure 17.10: Reconstructed invariant mass of (a) the W boson and (b) the top-quark using the jet assignment by TG (solid), SPA-net (dashed), χ^2 (dash-dotted), and TT (dotted). Events are categorised into *correct* (green), *incorrect* (orange), and *impossible* (blue) events. Adapted from Ref. [3].

identified. The three ML methods again have very similar performance, while the χ^2 method tends to be worse. It is interesting to see that the TT model tends to be much closer to the TG model than the SPA-net model, even though the efficiency for correctly identifying fully-reconstructable events is higher for SPA-net. This points to the models using the Topograph structure being better at identifying jets of not fully reconstructable particles. It is worth noting that the efficiencies of correctly identifying all possible partons reduce the fewer partons are matchable.

N_{partons}		Incorrectly matched partons				
matched	Model	0	≤ 1	≤ 2	≤ 3	≤ 4
3	SPA-net	34.34 ± 0.65	74.88 ± 0.59	98.03 ± 0.19	100.00 ± 0.00	-
	TG	36.57 ± 0.66	77.80 ± 0.57	98.70 ± 0.15	100.00 ± 0.00	-
	TT	35.86 ± 0.65	77.89 ± 0.57	98.81 ± 0.15	100.00 ± 0.00	-
	χ^2	34.30 ± 0.65	80.12 ± 0.54	99.05 ± 0.13	100.00 ± 0.00	-
4	SPA-net	35.69 ± 0.28	64.53 ± 0.28	90.99 ± 0.17	99.41 ± 0.04	100.00 ± 0.00
	TG	37.74 ± 0.28	67.06 ± 0.27	92.21 ± 0.16	99.53 ± 0.04	100.00 ± 0.00
	TT	37.30 ± 0.28	66.64 ± 0.28	92.40 ± 0.15	99.55 ± 0.04	100.00 ± 0.00
	χ^2	31.01 ± 0.27	63.84 ± 0.28	92.19 ± 0.16	99.73 ± 0.03	100.00 ± 0.00
5	SPA-net	44.74 ± 0.19	64.87 ± 0.18	86.92 ± 0.13	97.91 ± 0.05	99.94 ± 0.01
	TG	46.57 ± 0.19	66.88 ± 0.18	88.58 ± 0.12	98.45 ± 0.05	99.96 ± 0.01
	TT	45.90 ± 0.19	66.47 ± 0.18	88.23 ± 0.12	98.38 ± 0.05	99.95 ± 0.01
	χ^2	32.36 ± 0.18	58.27 ± 0.19	84.70 ± 0.14	98.21 ± 0.05	99.94 ± 0.01
6	SPA-net	66.05 ± 0.23	74.43 ± 0.21	90.06 ± 0.14	96.22 ± 0.09	99.52 ± 0.03
	TG	67.03 ± 0.22	75.71 ± 0.20	91.33 ± 0.13	97.03 ± 0.08	99.70 ± 0.03
	TT	65.96 ± 0.23	74.71 ± 0.21	90.84 ± 0.14	96.74 ± 0.08	99.67 ± 0.03
	χ^2	55.90 ± 0.24	64.93 ± 0.23	84.14 ± 0.17	93.97 ± 0.11	99.43 ± 0.04

Table 17.3 shows the efficiencies of correctly assigning a jet to its parton split by the flavour of the quark initiating the jet. The truth flavour is determined using a ΔR matching to c - and b -hadrons. Jets

Table 17.2: Percentage of events with at most N incorrectly matched partons using the SPA-net model, the TG model, the TT model, and the χ^2 method. Events are categorised based on the number of partons matched to jets at truth level. The highest efficiency is highlighted in bold. Adapted from Ref. [3].

matched to a b -hadron are labelled as a b -jet. If the jet is not matched to any b -hadron but to a c -hadron, it is labelled as a c -jet, while jets not matched to any c - or b -hadron are labelled as light-flavour jets. Only events with at least one jet incorrectly assigned are considered to observe whether there are differing dependences on the jet flavour. A similar pattern as before can be observed: the three ML methods have similar performance, with the TG model performing best. The χ^2 efficiencies are a bit lower than the three ML methods, however the drop is much less significant than for the other comparisons. The largest reduction in efficiencies for the χ^2 method can be seen for c -jets. This can be explained by the larger mis-tagging efficiency for c -jets compared to light-flavour jets. If they are mis-identified as b -jets, they are more likely to be assigned to one of the b -quarks from the top-quark decay, especially if one of the actual b -jets is not tagged as a b . As the ML methods do not have a hard requirement of the top-jets being b -tagged, this effect is reduced.

Table 17.3: Efficiencies [%] of correctly associating a jet to its corresponding parton for the SPA-net model, the TG model, the TT model, and the χ^2 method. Jets are categorised into their truth flavour using parton matching. Only jets coming from the top-quark pair decay in events with at least one incorrect assignment are considered. The highest efficiency is highlighted in bold. Adapted from Ref. [3].

	b -jets	c -jets	Light jets
SPA-net	58.90 ± 0.29	55.08 ± 0.39	70.42 ± 0.22
TG	59.99 ± 0.29	55.81 ± 0.39	72.01 ± 0.22
TT	59.94 ± 0.28	55.33 ± 0.39	71.54 ± 0.22
χ^2	57.68 ± 0.25	50.38 ± 0.34	68.51 ± 0.20

Table 17.4 shows the association efficiencies of each parton in *incorrect* events. The three ML methods all have similar behaviour, with the TG model performing best overall. The SPA-net model has slightly lower efficiencies for all partons. This is in line with the lower event reconstruction efficiency observed for *incorrect* events in Table 17.2. The χ^2 method has mostly lower association efficiencies for jets with higher p_T with the leading b -jet and both of the leading W -jets. In the case of the jet which most often has the lowest p_T of the top-quark pair decay products, the subleading W -jet from the top-quark with the leading b -jet, the χ^2 method even achieves the highest association efficiency.

	Correctly identified	SPA-net	TG	TT	χ^2
top-quark (leading b)	b -quark	61.83 ± 0.40	62.94 ± 0.40	62.74 ± 0.40	59.17 ± 0.35
	Leading quark (W)	67.79 ± 0.38	68.92 ± 0.39	68.69 ± 0.38	61.80 ± 0.35
	Subleading quark (W)	63.22 ± 0.40	64.70 ± 0.40	64.12 ± 0.39	65.09 ± 0.34
top-quark (subleading b)	b -quark	56.05 ± 0.41	57.03 ± 0.41	57.15 ± 0.41	56.20 ± 0.36
	Leading quark (W)	67.49 ± 0.38	68.22 ± 0.39	67.93 ± 0.38	61.30 ± 0.35
	Subleading quark (W)	66.41 ± 0.39	68.31 ± 0.39	67.61 ± 0.38	65.94 ± 0.34

Table 17.4: Efficiencies [%] of correctly associating a jet to the corresponding parton for events with at least one incorrect assignment for the SPA-net model, the TG model, the TT model, and the χ^2 method. Top quarks are classified by the momentum of the corresponding b -quark. The quarks from the W boson decays are ordered by their p_T . The highest efficiency is highlighted in bold. Adapted from Ref. [3].

RESILIENCE TO SYSTEMATIC VARIATIONS To estimate the effect of systematic uncertainties, three potential sources are explored. Effects of using a different generator setup are evaluated by comparing efficiencies when evaluating the models on the sample generated

with PYTHIA only. Other typically large uncertainties concern the jet energy resolution and the jet energy scale. Effects of the jet energy resolution are investigated by smearing the jet energy by sampling from a Gaussian with a standard deviation of 0.05 and a mean of one and scaling the energy by the sampled value. The difference of the smeared and unsmeared energies after the preprocessing of top-jets can be seen in Figure 17.11a. Negative energies after the smearing are set to an energy of zero. As the values of the three momentum are not set to zero, this does not remove the jet from the event. To mimic uncertainties related to the jet energy scale, all jet energies are scaled by 2.5%. The resulting distribution of energies with and without the scaling applied after the preprocessing can be seen in Figure 17.11b. The choice of the values used for the smearing and the scaling are motivated by the typical size of uncertainties in the ATLAS experiment [80].

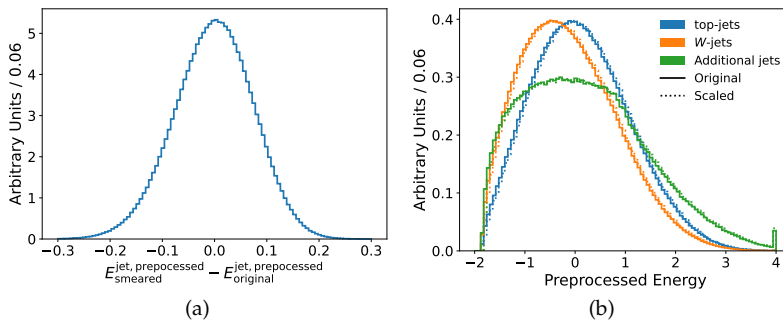


Figure 17.11: (a) difference of the top-jet energies before and after smearing and (b) the distribution of jet energies before and after scaling. The preprocessing of the energy is applied in both cases.

Table 17.5 shows the effect of the systematic variations on the reconstruction efficiency. The difference between the efficiency on the varied dataset and the nominal dataset is shown. The three ML methods behave very similar. For the evaluation on the PYTHIA sample, all experience a drop in efficiency of around 2%. The TT model tends to have a slightly larger reduction in efficiency than the other two models. The χ^2 method also drops in performance, hinting towards the PYTHIA sample being inherently harder to solve. However, the reduction in efficiency is lower for the χ^2 method at below 1% for events with at least six jets.

When smearing the jet energy, the SPA-net and TG model stay constant in performance, in fact for some jet multiplicities the efficiencies are even slightly larger than without smearing applied. The TT model has a slight drop in efficiencies of around 0.2%. This is however a much smaller reduction than for the χ^2 method. Here, the efficiencies drop between 2.82% and 12.69% depending on the jet multiplicity. The lowest reduction is seen for events with exactly six jets.

A similar overall behaviour can be seen for changes in the jet energy scale. All three ML methods are nearly unaffected by the systematic variation while the efficiency of the χ^2 method suffers dropping even for events with exactly six jets by 12.79%. For events with seven jets the reduction in efficiency is even larger with 21.52%.

N_{jets}		6		7		≥ 6	
$N_{b\text{-jets}}$		2	≥ 2	2	≥ 2	2	≥ 2
PYTHIA	SPA-net	-2.13	-2.28	-2.04	-2.36	-1.64	-1.86
	TG	-2.00	-2.13	-2.14	-2.37	-1.53	-1.89
	TT	-2.32	-2.60	-2.27	-2.23	-1.77	-2.01
	χ^2	-0.94	-1.33	-1.66	-1.62	-0.72	-0.94
Resolution	SPA-net	+0.04	+0.07	± 0.00	-0.02	+0.04	+0.06
	TG	± 0.00	+0.03	± 0.00	+0.04	+0.01	+0.05
	TT	-0.21	-0.24	-0.15	-0.10	-0.16	-0.18
	χ^2	-2.82	-3.63	-12.30	-12.69	-8.23	-9.02
Scale	SPA-net	+0.08	+0.04	-0.05	-0.10	+0.02	-0.01
	TG	± 0.00	+0.01	-0.06	-0.07	+0.02	+0.03
	TT	± 0.00	+0.01	± 0.00	± 0.00	± 0.00	+0.02
	χ^2	-12.79	-14.30	-21.52	-21.63	-17.18	-17.96

Table 17.5: Difference in reconstruction efficiencies [%] when evaluating the different methods on systematic variations. The difference is taken with respect to Table 17.1. Adapted from Ref. [3].

Overall, the ML models seem to be more affected by a change of the generator setup than the χ^2 method. However, they are very resilient to changes to the jet energy which is not true for the χ^2 method.

TRAINING ON PARTIAL EVENTS In addition to training on only complete events, models are trained including partial events, adjusting the loss accordingly. Again, five models are trained for each ML method choosing the one with the highest efficiency in events with at least six jets and at least two b -jets. The same hyperparameters as when training on complete events are used. Table 17.6 shows the reconstruction efficiencies on complete events. As the χ^2 method does not change, the efficiencies are the same as in Table 17.1.

N_{jets}	6		7		≥ 6	
$N_{b\text{-jets}}$	2	≥ 2	2	≥ 2	2	≥ 2
SPA-net	81.03 $\pm$ 0.32	79.11 $\pm$ 0.30	65.35 $\pm$ 0.44	63.40 $\pm$ 0.40	68.93 $\pm$ 0.25	66.33 $\pm$ 0.23
TG	80.77 $\pm$ 0.32	78.93 $\pm$ 0.30	65.18 $\pm$ 0.44	63.13 $\pm$ 0.40	68.75 $\pm$ 0.25	66.21 $\pm$ 0.23
TT	79.56 $\pm$ 0.32	77.57 $\pm$ 0.31	63.24 $\pm$ 0.45	60.90 $\pm$ 0.40	66.97 $\pm$ 0.25	64.09 $\pm$ 0.23
χ^2	72.73 $\pm$ 0.36	70.94 $\pm$ 0.33	54.28 $\pm$ 0.46	52.11 $\pm$ 0.41	58.57 $\pm$ 0.26	55.90 $\pm$ 0.24

Table 17.6: Event reconstruction efficiencies [%] for the SPA-net model, the TG model, the TT model, and the χ^2 method in different jet and b -jet multiplicities. The ML models are trained including partial events. The highest efficiency is highlighted in bold. Adapted from Ref. [3].

Here, the SPA-net model has the highest reconstruction efficiencies for all jet multiplicities. Comparing it with the model trained on only complete events, the performance is slightly lower for events with exactly six jets, but improves for higher jet multiplicities and inclusive jet multiplicities. For both TG and TT, the efficiencies are reduced compared to the model trained on only complete events. The efficiencies of the TG model drop by around 1%, while the efficiencies of the TT model drop by around 2%. As for the SPA-net model, the reduction is more significant in events with exactly six jets.

Table 17.7 shows the efficiencies of correctly assigning different amounts of jets in events where not all partons can be matched. All

N_{partons}		Incorrectly matched partons				
matched	Model	0	≤ 1	≤ 2	≤ 3	≤ 4
3	SPA-net	36.68 ± 0.66	76.94 ± 0.57	98.55 ± 0.16	100.00 ± 0.00	-
	TG	37.27 ± 0.66	77.89 ± 0.57	98.77 ± 0.15	100.00 ± 0.00	-
	TT	42.75 ± 0.67	82.70 ± 0.52	98.98 ± 0.14	100.00 ± 0.00	-
	χ^2	34.30 ± 0.65	80.12 ± 0.54	99.05 ± 0.13	100.00 ± 0.00	-
4	SPA-net	37.78 ± 0.28	66.29 ± 0.28	91.78 ± 0.16	99.46 ± 0.04	100.00 ± 0.00
	TG	38.08 ± 0.28	67.53 ± 0.27	92.57 ± 0.15	99.62 ± 0.04	100.00 ± 0.00
	TT	43.80 ± 0.29	71.66 ± 0.26	94.18 ± 0.14	99.68 ± 0.03	100.00 ± 0.00
	χ^2	31.01 ± 0.27	63.84 ± 0.28	92.19 ± 0.16	99.73 ± 0.03	100.00 ± 0.00
5	SPA-net	48.14 ± 0.19	67.24 ± 0.18	88.18 ± 0.12	98.14 ± 0.05	99.95 ± 0.01
	TG	48.22 ± 0.19	69.27 ± 0.18	89.52 ± 0.12	98.68 ± 0.04	99.96 ± 0.01
	TT	52.97 ± 0.19	71.33 ± 0.17	90.78 ± 0.11	98.79 ± 0.04	99.96 ± 0.01
	χ^2	32.36 ± 0.18	58.27 ± 0.19	84.70 ± 0.14	98.21 ± 0.05	99.94 ± 0.01
6	SPA-net	66.33 ± 0.23	74.38 ± 0.21	89.86 ± 0.14	96.18 ± 0.09	99.54 ± 0.03
	TG	66.20 ± 0.23	74.87 ± 0.21	91.11 ± 0.14	97.04 ± 0.08	99.73 ± 0.02
	TT	64.09 ± 0.23	73.22 ± 0.21	90.24 ± 0.14	96.69 ± 0.09	99.70 ± 0.03
	χ^2	55.90 ± 0.24	64.93 ± 0.23	84.14 ± 0.17	93.97 ± 0.11	99.43 ± 0.04

models show increased efficiencies in events where not all six partons are matched in comparison to the models trained on complete events only. In particular the TT model has much better performance. For instance, the efficiency of reconstructing all five partons in events with only five partons available, increases by over 7%. It also performs significantly better than the other models for almost all event categories in Table 17.7. With a more careful tuning of the loss function, the model could probably maintain similar performance on complete events while losing some gains on partial events, making the efficiencies more similar to what is seen for the TG and the SPA-net model.

Table 17.7: Percentage of events with at most N incorrectly matched partons using the SPA-net model, the TG model, the TT model, and the χ^2 method. Events are categorised based on the number of partons matched to jets at truth level. The ML models are trained including partial events. The highest efficiency is highlighted in bold. Adapted from Ref. [3].

17.4.2 Interpretation of edge scores

INDIVIDUAL EDGE SCORES The edge output scores can be used to understand certain features of the models. Figures 17.12 and 17.13 show the edge scores of the TG model between jets and the nodes identified as the W^+ boson and the top-quark node, respectively. Figures 17.12a and 17.13a show the scores for all jets in the event. The distribution of scores of the jets that originate from the particle of the corresponding node peak at one, while all other distributions peak at zero. The jets that originate from the same top-quark, so b -jets from the top-quark for the W^+ node and jets from the W^+ boson for the top node, receive on average higher scores than the other false edges. For the W node, jets from the other W boson and the additional jets receive on average higher scores than the b -jets from the anti-top-quark, while for the top-quark node, the b -jets from the anti-top-quark

receive higher scores. This can be explained by differences in the b -tagging performance of these jets. Most of the jets from the W bosons and most of the additional jets are not b -tagged, but most jets from the top-quarks are b -tagged. Therefore, the network prefers non- b -tagged jets for jets connected to the W boson nodes, while for top-quark nodes it is the other way around.

This effect can also be seen in Figures 17.12b and 17.13b. Here, only the scores of top-jets are shown split on whether the jet was b -tagged or not. It can be seen that top-jets receive higher scores to the W node if they are not b -tagged. Even jets from the anti-top-quark which are not b -tagged receive on average higher scores than the jets from the top-quark which are b -tagged. For the scores to the top node, the b -tagging score seems to be not as important. Here, jets from the top-quark receive higher scores even if they are not b -tagged compared to b -tagged jets from the anti-top-quark. This can be explained by the higher likelihood of b -jets not being b -tagged than non- b -jets being b -tagged.

Figure 17.12: Edge scores of jets to the helper node representing the W^+ boson of the TG model. (a) shows all jets, while (b) only shows the b -jets from the two tops. They are further split into whether the jet was tagged as a b -jet or not. No requirement is placed on the number of b -tags. Adapted from Ref. [3].

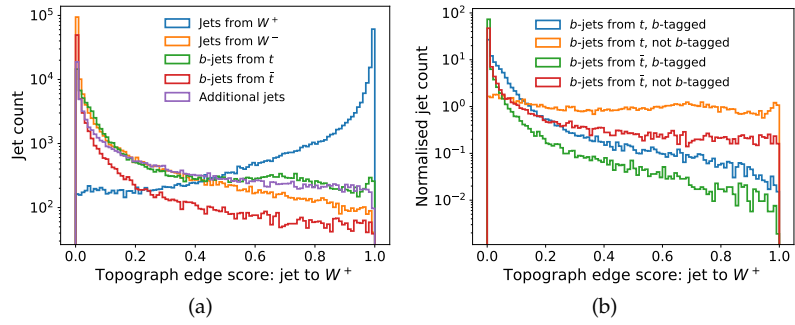
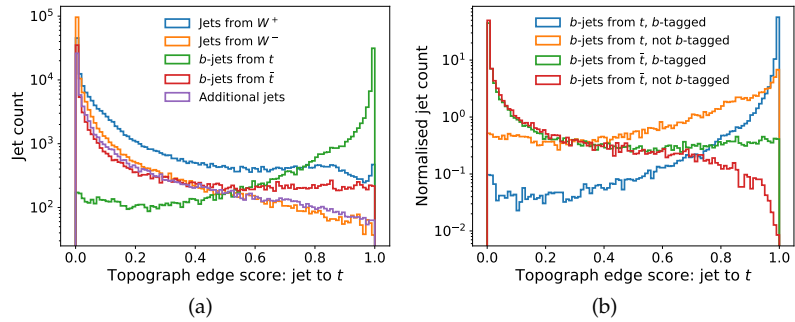


Figure 17.13: Edge scores of jets to the helper node representing the top-quark of the TG model. (a) shows all jets, while (b) only shows the b -jets from the two tops. They are further split into whether the jet was tagged as a b -jet or not. No requirement is placed on the number of b -tags. Adapted from Ref. [3].



EVENT LEVEL SCORE The individual edge scores of all jets can be aggregated into one score. This score could for instance be used to reject events which likely do not have the correct reconstruction or are not possible to reconstruct. This would especially be helpful to reduce the contributions of other processes passing an event selection, similar to the cut on the KLFitter output in the c -jet calibration, see Appendix C.6.

One such possible aggregation is multiplying the scores of all assigned jets. The score obtained in this way is shown in Figure 17.14 for the TG and TT model. For visualisation purposes, the logarithm of the product is shown. In both cases, the events which are assigned correctly receive higher total scores on average. *Incorrect* and *impossible* events have a peak at a lower but similar value, though *impossible* events also have a shoulder visible towards higher values.

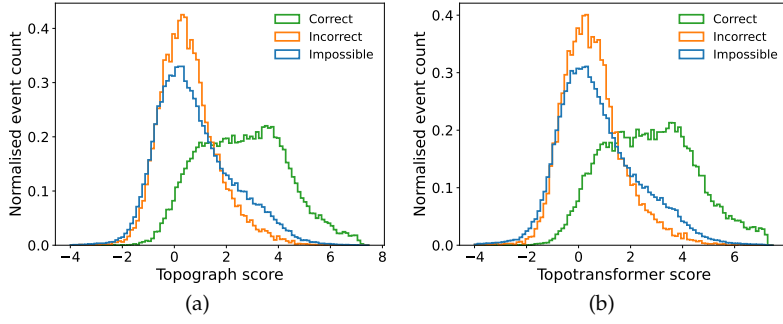


Figure 17.14: Score of (a) the TG and (b) TT model for *correct*, *incorrect*, and *impossible* events. The score is taken by multiplying the edge scores of the six edges assigned as true edges. (a) taken from Ref. [3]

Figure 17.15 shows all the *impossible* events split on how many jets are correctly identified. It can be seen that the shoulder is caused by events where only one jet is wrongly assigned. Fewer jets being correctly identified leads to lower scores on average.

Whether a selection on such a score would benefit any downstream task, such as the calibration of c -jets detailed in Appendix C.6 remains an open question.

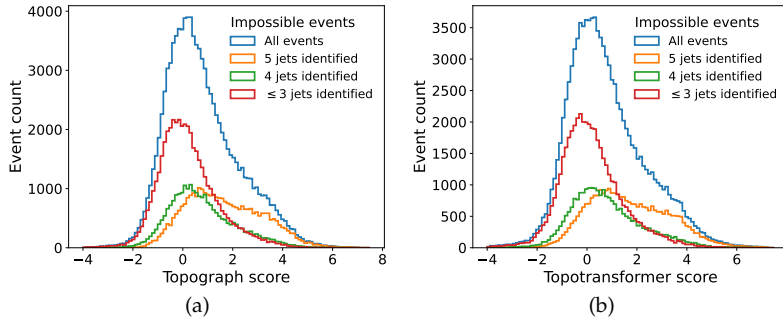


Figure 17.15: Score of (a) the TG and (b) TT model for *impossible* events. The events are separated by how many jets are correctly identified. The score is taken by multiplying the edge scores of the six edges assigned as true edges. (a) taken from Ref. [3]

17.4.3 Regression tasks

Two options are available to assess the kinematics of the W bosons and the top-quarks using the Topograph algorithm: using the assigned jets and their four-momentum to reconstruct the kinematics or using the output of the regression network. The regression network has the advantage that it tries to predict the kinematics of the particles without any detector effects. Figure 17.16 shows the resolution of p_T , η , and ϕ for the W bosons and the top-quarks for the TG model. For p_T , the difference between the true and the predicted value is

normalised by the true value. This is not done for η and ϕ as both quantities cross zero.

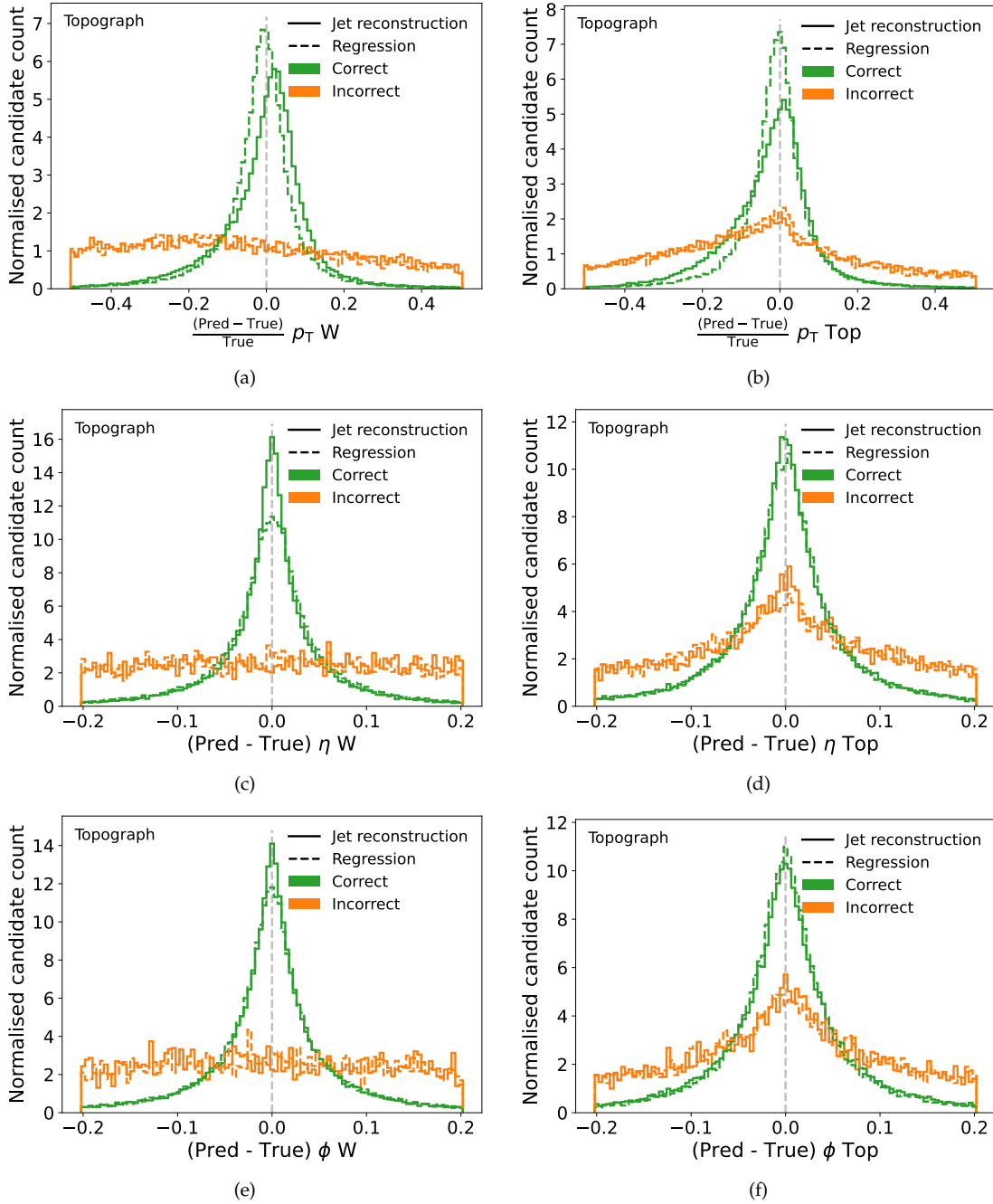


Figure 17.16: Resolution of (a,c,e) W boson and (b,d,f) top-quark kinematics either reconstructing the particles using the assigned jets or using the regression output of the TG model. The events are separated into *correct* and *incorrect* events. The kinematics include (a,b) p_T , (c,d) η , and (e,f) ϕ . Adapted from Ref. [3].

The regression network shows a better resolution for p_T for both W bosons and top-quarks, if the event is correctly assigned. This can be seen by the narrower peak of the distribution. For η and ϕ , the reconstruction tends to have better resolution than the regression network, apart from ϕ of top-quarks. As η and ϕ are defined by angles which can be measured much more accurately than the p_T of a jet, it is not unexpected that the reconstruction performs comparably better for these two quantities.

For incorrectly reconstructed events, both the reconstruction and

the regression are much worse, often not peaking around zero. However, it should be noted that even when choosing the true assignment of jets, the resolution of the reconstructed kinematics is worse for *incorrect* events compared to *correct* events, as can be seen in Figure 17.17. This points towards larger detector resolution effects in these events, which has a negative impact on the reconstruction performance of the algorithms.

Table 17.8 quantifies the accuracy and the precision of the reconstruction of the kinematics and the regression. As a measure for the accuracy, the mean is used. For the resolution, the IQR is used. As could be seen from Figure 17.16, the regression has a better resolution for the p_T of W bosons and top-quarks and ϕ of top-quarks for *correct* events, while the reconstruction has a better resolution for the other quantities. For *incorrect* events, the regression always has a better resolution, albeit much worse than for *correct* events. This is not unexpected as the reconstruction is forced into a wrong value by using the wrong jets, whereas the regression can still migrate towards the true properties.

For *correct* events, the mean of all predictions is relatively close to zero. The same can be said for η and ϕ for *incorrect* events. However, p_T for *incorrect* events has a large positive bias, with the regression output having a lower bias than the reconstruction. This is different to what could be expected from Figure 17.16, where, especially for the W boson, a negative bias can be seen. This is explained by a large portion of events not being contained in the range shown in Figure 17.16. Figure 17.18 shows the same figure as Figure 17.16 for the p_T of the top-quark but including an under- and an overflow bin. As the amount of events in these bins is large for *incorrect* events, they are not included in Figure 17.16.

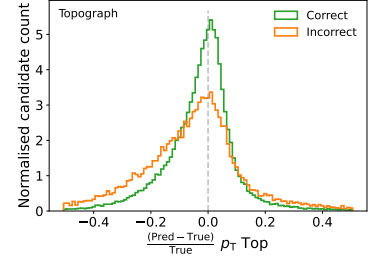


Figure 17.17: Reconstructed top-quark p_T resolution using true jet assignment. The events are separated into *correct* and *incorrect* events based on the TG model.

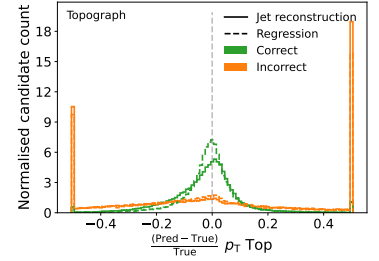


Figure 17.18: Reconstructed top-quark p_T resolution using the model assignment or the regression output. The events are separated into *correct* and *incorrect* events based on the TG model. The first and last bin include under- and overflow events.

Particle	Assigned	Model	Method	p_T		η		ϕ	
				Mean	IQR	Mean	IQR	Mean	IQR
W boson	Correct	TG	Reg	-0.006	0.091	0.000	0.061	- 0.002	0.064
			Reco	0.008	0.114	0.000	0.048	- 0.002	0.058
		TT	Reg	- 0.004	0.093	0.000	0.062	- 0.002	0.066
			Reco	0.007	0.114	0.000	0.048	- 0.002	0.059
	Incorrect	TG	Reg	0.393	0.826	-0.013	0.845	0.004	1.244
			Reco	0.568	1.002	-0.013	0.973	0.008	1.518
		TT	Reg	0.388	0.805	- 0.007	0.831	0.010	1.222
			Reco	0.577	1.002	-0.011	0.982	0.010	1.511
top-quark	Correct	TG	Reg	-0.003	0.084	0.000	0.066	0.000	0.066
			Reco	-0.016	0.123	-0.001	0.065	0.000	0.072
		TT	Reg	- 0.001	0.087	0.000	0.068	0.002	0.067
			Reco	-0.017	0.122	-0.001	0.065	0.000	0.072
	Incorrect	TG	Reg	0.196	0.484	-0.002	0.519	0.000	0.638
			Reco	0.253	0.570	-0.007	0.609	-0.002	0.807
		TT	Reg	0.200	0.477	-0.005	0.517	0.000	0.651
			Reco	0.263	0.572	0.000	0.614	0.006	0.825

Table 17.8: Mean and IQR of the normalised error for p_T and the error for η and ϕ , for the regression outputs of the TG and TT model, as well as the reconstructed values using the jet assignment of the same models. The events are split into *correct* and *incorrect* events.

PART V

CONCLUSIONS

18 Summary and conclusions

Three topics have been presented in this work: an improved electron identification using either only high-level features or including low-level information on top; the measurement of the fiducial and total cross-section of top-quark pairs with additional charm jets; and a new method of reconstructing all-hadronic top-quark pair final states. The $t\bar{t}$ + charm analysis could benefit from a reconstruction of the final state to perform c -tagging only on jets not originating from the top-quark pair system. The new method developed for this purpose uses GNNs or transformers, which are also tested in the context of electron identification.

ELECTRON IDENTIFICATION Using similar high-level features as the current method, the background rejection in MC of certain WPs for electron identification can be increased by a factor of up to five. The relative increase in rejection is larger for higher signal efficiencies or looser WPs. On top of providing improved rejection, the trained DNN is more versatile due to being a multinomial classification algorithm. This, however, also limits the option to train with recorded data directly and makes the training completely dependent on MC. As there is known mismodelling in MC, correction factors need to be applied to the training data to obtain competitive performance. Lastly, it would also introduce additional trigger inefficiencies in the Run 2 dataset due to a LH algorithm being run at trigger level during this time. A trigger algorithm based on a DNN could, however, mitigate this issue.

Including low-level information can further boost the performance of the identification algorithm. Representing the cells of the calorimeter as a point cloud instead of images leads to better performance. The best algorithm trained in this work is a transformer-based model. It can boost the performance in MC by a factor of around two over a DNN. The inclusion of both calorimeter information and information on additional tracks is beneficial. However, no correction factors exist for the low-level information, leading to a much reduced performance in data. Nevertheless, the trained algorithms that include low-level information outperform a DNN trained without correction factors applied, indicating that if correction factors were to be applied to the low-level information and the high-level features, an improved performance could also be seen in data.

$t\bar{t}$ + CHARM ANALYSIS The first measurement of the cross-section of top-quark pairs with additional charm jets has been performed. It was carried out with the full Run 2 dataset of proton-proton collisions at $\sqrt{s} = 13$ TeV corresponding to an integrated luminosity of 140 fb^{-1} . Results are presented for the lepton+jets channel and a combination of the lepton+jets and dilepton channels. For this analysis, a dedicated bc -Tagger was developed to simultaneously tag c - and

b -jets. The extracted fiducial cross-section of events with at least two additional c -jets is $\sigma_{t\bar{t}+\geq 2c}^{\text{fid}} = 1.33_{-0.25}^{+0.28}$ pb, while the one for events with only one additional c -jet is $\sigma_{t\bar{t}+1c}^{\text{fid}} = 6.78_{-0.82}^{+1.02}$ pb. Both values quote the results of the combined fit setup. They are compared to the predicted cross-sections of several generated samples. All samples predict lower cross-sections, though within uncertainties, they agree with the measured values. The best agreement for these two cross-sections is obtained with the generated $t\bar{t}$ POWHEG BOX + PYTHIA sample, where the h_{damp} parameter was set to three times the top-quark mass. The worst agreement can be seen for the generated $t\bar{t}$ MADGRAPH5_AMC@NLO + HERWIG sample. Extrapolated to the full phase space while only including events with one leptonic W boson decay, the results are $\sigma_{t\bar{t}+\geq 2c}^{\text{total}} = 5.6_{-1.1}^{+1.2}$ pb and $\sigma_{t\bar{t}+1c}^{\text{total}} = 41.3_{-6.4}^{+6.9}$ pb, respectively. The extrapolated total cross-section of top-quark pair production is measured to be $\sigma_{t\bar{t}}^{\text{total}} = 816_{-63}^{+70}$ pb including all-hadronic top-quark pair events. This is in agreement with the theoretically calculated cross-section. The dominant uncertainties are systematic uncertainties, more specifically uncertainties related to the modelling of top-quark pair events. With the combination, these uncertainties can be reduced.

To improve the analysis, a reconstruction algorithm like Topograph could be used to reduce the contribution of c -jets from W boson decays. This would also be most likely necessary for the potential inclusion of the all-hadronic final state, where two c -jets can originate from W boson decays. Instead of using simple event-level observables, machine learning algorithms could be trained to separate signal from background and use the resulting discriminant as a fit observable. Additionally, recent advancements in flavour-tagging [147] algorithms can further improve the analysis by selecting purer regions of $t\bar{t} + \geq 1c$, $t\bar{t} + \geq 1b$, and $t\bar{t} + \text{light}$. Lastly, differential cross-section measurements are an obvious next step to further probe and improve the modelling of top-quark pairs with additional charm jets.

TOPOGRAPHS State-of-the-art results can be achieved in the reconstruction of all-hadronic top-quark pair decays using Topographs. This new method represents jets as a point cloud and injects helper nodes representing intermediate particles of the decay chain. This is used to reconstruct a graph inspired by the Feynman diagram of the decay, introducing an inductive bias into the network. The helper nodes can additionally be used to regress towards the true properties of the intermediate particles, which could lead to improved results in differential measurements or something similar. An application was only shown for all-hadronic top-quark decays, though it can easily be applied to other final states, such as the lepton+jets final state. This reconstruction could then be applied to the $t\bar{t} + \text{charm}$ analysis to potentially increase the sensitivity in the lepton+jets channel.

PART VI

APPENDICES

A Electron identification with GNNs: hyperparameters

Hyperparameter	DeepSets	GNN	MPC Transformer	SPC Transformer
Embedding nodes	-	[128, 32]	[128, 128, 64]	[256, 256, 123/122]
Embedding activation	-	GELU	GELU	GELU
Embedding norm	-	layer	layer	layer
Embedding dropout	-	0	0.1	0.1
Edges nodes	-	[128, 128, 32]	-	-
Edges activation	-	GELU	-	-
Edges norm	-	layer	-	-
Edges neighbours	-	4	-	-
Edges pooling	-	attention	-	-
Edges att. nodes	-	[32, 32, 1]	-	-
Edges att. activation	-	GELU	-	-
Edges att. norm	-	layer	-	-
Transformer nodes	-	-	[128, 64]	[256, 128]
Transformer activation	-	-	GELU	GELU
Transformer norm	-	-	layer	layer
Transformer dropout	-	-	0.1	0.1
Number of heads	-	-	8	8
Encoder layers	-	-	3	4
Nodes nodes	[128, 128, 32]	[128, 128, 32]	-	-
Nodes activation	GELU	GELU	-	-
Nodes norm	layer	layer	-	-
Nodes pooling	attention	attention	attention	attention
Nodes att. nodes	[32, 32, 1]	[32, 32, 1]	[32, 32, 1]	[32, 32, 1]
Nodes att. activation	GELU	GELU	GELU	GELU
Nodes att. norm	layer	layer	layer	layer
Downscaling nodes	[16]	[16]	[16]	[16]
Downscaling activation	Leaky ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU
Downscaling norm	batch	batch	batch	batch
Batch size	256	256	256	256
Updates per epoch	100000	100000	100000	100000

Table A.1: Hyperparameters of the point cloud-based models. A dash indicates a hyperparameter that is not applicable to the architecture. The architecture for the single point cloud transformer is the same regardless of the inclusion of the track information apart from the number of nodes in the last layer of the embedding network. Here, both values are given separated by a “/”.

B Electron identification: scale factor calculation

In Chapter 10, the performance of the developed electron identification algorithms is estimated. The applied signal selection has, however, a considerable background contamination. For an estimate of the performance, this is not an issue. However, to accurately calculate signal efficiencies and therefore SFs this contamination has to be taken into account. For the identification of electrons with $E_T > 15$ GeV, this is done with two methods: Z_{mass} and Z_{iso} .

B.1 Working points

To measure efficiencies, WPs have to be defined. Three WPs are defined for the DNN: *Loose*, *Medium*, and *Tight*. Additional rectangular cuts are applied on top of the selection based on the discriminant, as introduced in Table 8.6. The DNN WPs are tuned to give the same signal efficiencies as the respective LH WPs in simulation. For the evaluation of the DNN WPs, the correction factors are applied to simulated electron candidates. This choice was made to keep the evaluation as close to the training data as possible.

An interpolation of the threshold along E_T as for the LH WPs is applied to the DNN WPs as well.

B.2 Efficiency determination

A detailed description of the following methods can be found in Ref. [245]. To determine the efficiencies of the WPs in data, a selection of electron candidates has to be performed. This selection should be as pure as possible in signal electrons to reduce any effects of the necessary background estimation. As simulated samples do not describe the background contribution well, background templates have to be created with selections enriched in background. For electrons with $E_T > 15$ GeV, this is done by two methods: Z_{mass} and Z_{iso} . Both rely on a $Z \rightarrow ee$ tag-and-probe selection to obtain a pure sample of signal electrons. The background templates are constructed using the invariant mass of the tag-and-probe pair for the Z_{mass} method, while for the Z_{iso} method, a calorimeter isolation variable is used.

For electrons with lower E_T ($E_T < 20$ GeV), a method relying on a $J/\psi \rightarrow ee$ tag-and-probe selection is used by the ATLAS experiment. As there is an overlap from 15 GeV to 20 GeV, the three methods are combined into one set of SFs in this range. This method will, however, not be discussed in this work.

Excluding the low E_T region, the SFs are obtained in 210 bins of E_T and η . The bin edges can be seen in Table B.1. For the highest E_T bin with $E_T > 150$ GeV, the positive and negative η bins are combined to increase the available statistics.

Bin boundaries in η											
	-2.47	-2.37	-2.01	-1.81	-1.52	-1.37	-1.15	-0.8	-0.6	-0.1	
0.0	0.1	0.6	0.8	1.15	1.37	1.52	1.81	2.01	2.37	2.47	
Bin boundaries in E_T [GeV]											
15	20	25	30	35	40	45	50	60	80	150	250

Table B.1: Bin edges in η and E_T used for the SF calculation.

COMMON SELECTION For both methods, only a single simulated sample is used, with the POWHEG BOX + PYTHIA sample generating $Z \rightarrow ee$ events. The same selection as the signal selection described in Chapter 10 is used as a baseline for the two methods.

B.2.1 Z_{mass} method

The Z_{mass} method utilises the spectrum of the invariant mass of the tag-and-probe pair to estimate the background contribution close to the Z boson peak. In the first step, background templates are generated in a background control region. Probe electrons have to fail a relaxed version of the *LHLoose* WP and have to fulfil a calorimeter isolation requirement of $E_T^{\text{topocone30}}/p_T > 0.1$. To get a pure background template, contributions from the signal are subtracted using simulated electrons. This is done by performing a fit using the signal as predicted from MC and parametrizing the background by a polynomial function. The resulting signal contribution is subtracted from the template. The obtained background template is then fitted to data in the signal region together with a signal component from MC. The fitted background template is subtracted from the data to obtain the yields of signal electrons. As mismodelling of the Z boson peak can be observed after the subtraction, the subtracted distribution of probes passing the *LHTight* WP is used to reweight the signal contribution. The *LHTight* WP is used to reduce the impact of background, as this WP gives the purest selection of signal electrons. The obtained reweighting factors are applied to all simulation templates and the background template generation. Afterwards, the cleaning of the template and the fit in the signal region are repeated with the reweighted signal.

Several systematic uncertainties are considered for this measurement. They are estimated by varying one part of the method at a time. The difference between the efficiencies of the variation and the nominal measurement is taken as the systematic uncertainty of this variation. The uncertainties of all sources are combined by summing them in quadrature.

The requirements to obtain the background template are changed by either using an even looser LH WP that the probe is not allowed to pass or alternatively varying the calorimeter isolation requirement to $E_T^{\text{topocone40}}/p_T > 0.07$. The normalisation of the template is varied by the normalisation uncertainty obtained from the fit to account for

uncertainties in the background estimate. To estimate the effects of the potential background contribution to the tag electron, its selection is tightened to include an additional calorimeter isolation requirement of $E_T^{\text{topocone40}} < 5$ GeV. To probe the effects of the choice of the m_{ee} window, two uncertainties are extracted. Firstly, the fit to obtain the normalisation in the signal region is performed only with values of m_{ee} above or below the Z peak. And secondly, the yields considered for determining the efficiencies are taken for different ranges of m_{ee} , 70 GeV to 110 GeV and 80 GeV to 100 GeV.

B.2.2 Z_{iso} method

Calorimeter isolation variables are used to estimate the background contribution in the Z_{iso} method. To define background templates, a separate selection is performed in which both the tag and the probe are required to have the same charge. Additionally, the probe has to fail certain cuts on identification variables. To not bias the measurement, these are chosen to be uncorrelated with the calorimeter isolation variables. The signal contamination in the background template is subtracted using the simulated sample. Different templates are chosen within each (E_T, η) bin. Two templates are considered for each bin, one as a nominal measurement and one as a systematic variation.

To extract the number of electrons before any cut and after the selection of each of the WPs, the background template is normalised using the region with $E_T^{\text{topocone30}} > 12.5$ GeV for the probe. Any signal contamination in this region is subtracted using simulated events. Using this normalisation, the background is subtracted in the signal region ($E_T^{\text{topocone30}} < 12.5$ GeV). The yields obtained after the subtraction are taken as the yield of signal electrons after each of the respective cuts.

Several systematic uncertainties are considered. As already mentioned, a different background template is used. To account for modelling uncertainties, the normalisation of the signal contamination, which is subtracted from either the background template generation or the normalisation in the high-tail $E_T^{\text{topocone30}}$ region, is scaled up and down by 30%. Instead of $E_T^{\text{topocone30}}$, $E_T^{\text{topocone40}}$ is used to split the regions into normalisation of background and extraction of yields to estimate the effect of using a different calorimeter isolation variable. The effect of the chosen value of the cut on the isolation variable is estimated by changing the boundary of this cut to 10 GeV and 15 GeV. To estimate uncertainties related to the selection of the tag electron, it has to pass an additional calorimeter isolation requirement of $E_T^{\text{topocone30}} < 7.5$ GeV. Lastly, the selection of the m_{ee} window is changed to range from 70 GeV to 110 GeV or from 80 GeV to 100 GeV.

The treatment of the different systematic variations is the same as for the Z_{mass} method.

B.2.3 Combination

The final set of SFs to be used by physics analyses is obtained by combining the results of the two methods. The results are averaged while taking their uncertainties into account. Different uncertainties are treated as uncorrelated or correlated, depending on the source of the uncertainty. For instance, the uncertainty from varying the invariant mass window is treated as correlated, while specific changes to the background estimation are treated as uncorrelated. This combination has not been performed by the author, and therefore no combined results will be shown. Only the results of the single methods will be shown and compared.

B.3 Results

Figures B.1 and B.2 show the resulting SFs for the Z_{mass} and Z_{iso} methods, respectively. Results are only shown for data collected during 2018, as the results for the remaining years are similar.

Figure B.1: Efficiencies in MC and data as well as the SFs using the DNN as measured by the Z_{mass} method for data collected in 2018 with respect to η .

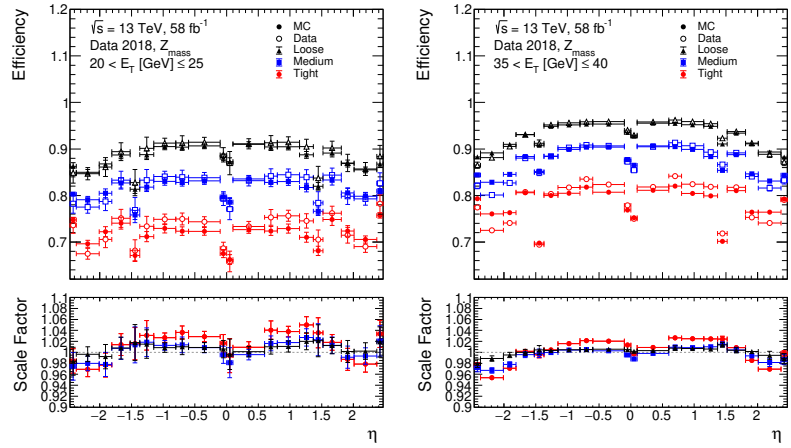
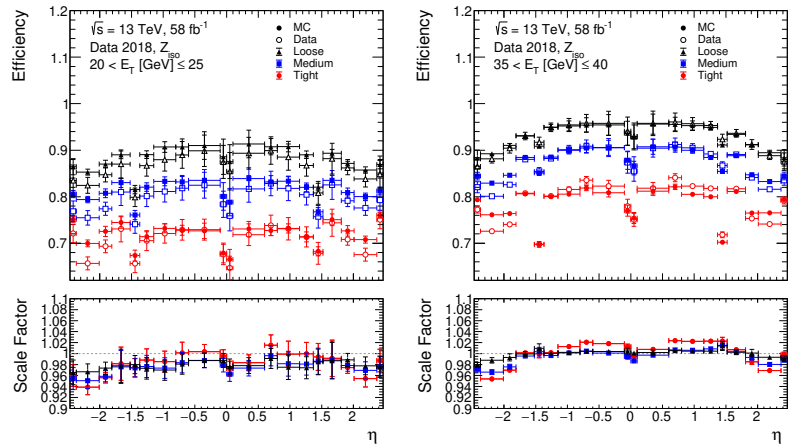


Figure B.2: Efficiencies in MC and data as well as the SFs using the DNN as measured by the Z_{iso} method for data collected in 2018 with respect to η .



The efficiencies are tuned to match the efficiencies of the LH al-

gorithm on corrected MC. Efficiencies in the very centre ($|\eta| < 0.1$) and in the crack region ($1.37 < |\eta| < 1.52$) are on average lower than in the remaining detector. Additionally, the efficiencies tend to drop towards larger values of $|\eta|$.

Uncertainties are higher at low E_T , where a large fraction of background is selected. At higher values of E_T , the background contamination is reduced, leading to a more consistent background estimate and therefore lower uncertainties.

Overall, all SFs are close to one. Larger deviations can be seen for the *Tight* WP compared to the *Loose* WP. This is expected as harsher cuts need to be made, and therefore, potential mismodelling will affect the efficiencies more. The SFs tend to be larger than one at low E_T for the Z_{mass} method. Most of the barrel region also has SFs slightly larger than one at higher values of E_T . The values of the SFs always drop towards higher values of $|\eta|$.

The results for the Z_{iso} method have a slightly different behaviour. At low values of E_T , the SFs tend to be around one instead of above. However, uncertainties are also larger for the Z_{iso} method at these ranges of E_T . The remaining trends and observations are the same as for the Z_{mass} method.

For both methods, a slight asymmetry in the SFs with respect to η can be seen. Figure B.3 overlays the SFs in one positive η bin with the respective negative η bin showing results for data from 2017 with the Z_{iso} method. Only the *Tight* WP is shown. Clearly, larger SFs are obtained for positive values of η . The difference is relatively consistent across the whole E_T spectrum, with a slight drop towards larger energies. The assigned uncertainties do not cover this asymmetry. However, this asymmetry does not seem to be caused by the DNN algorithm, as Figure B.3b shows the same behaviour for the LH algorithm. Additionally, a large difference in the values of the SFs can be seen. The LH algorithm produces lower SFs than the DNNs algorithm. The primary reason is, however, not the choice of the algorithm but the choice of using corrected MC in the evaluation of the efficiencies in simulation for the DNN, while the LH is evaluated on uncorrected MC. If corrected MC was to be used for the LH algorithm, the value of the SFs would most likely increase.

Lastly, Figure B.4 shows a comparison of the results of the Z_{mass} and the Z_{iso} methods. Overall, good agreement, especially at higher values of E_T , can be observed. At low values of E_T , the Z_{iso} method results in lower SFs than the Z_{mass} method. However, the uncertainties at lower values of E_T are also significantly larger, such that the differences are covered by the uncertainties.

Figure B.3: Comparison of the SFs for positive and negative η values using the Z_{iso} method for data collected in 2017. The SFs for (a) the DNN algorithm as well as (b) the LH algorithm are shown.

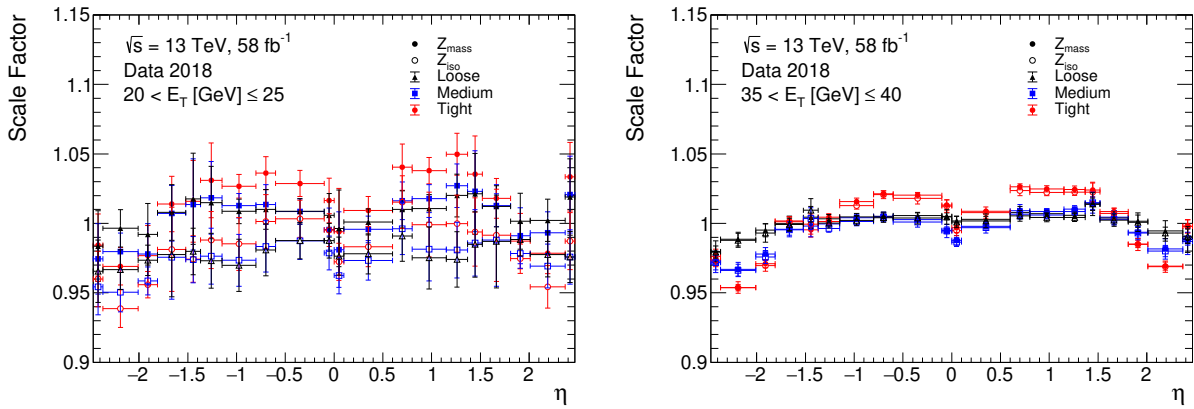
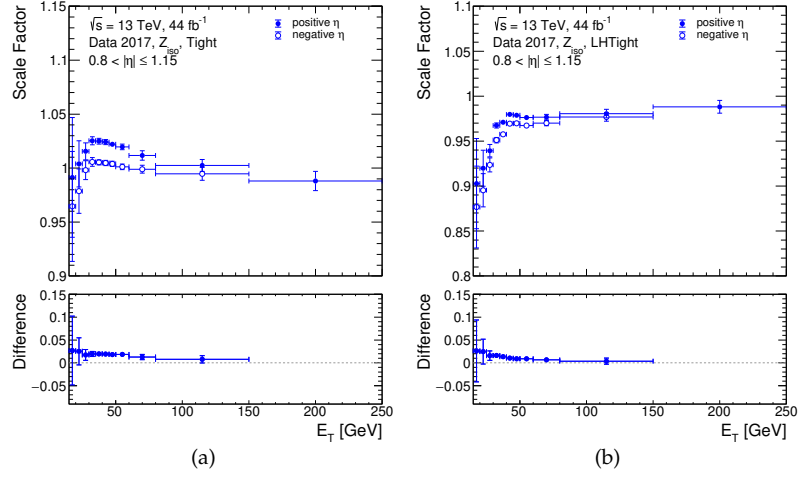


Figure B.4: Comparison of the SFs obtained with the Z_{mass} and Z_{iso} method for data collected in 2018.

C *bc*-Tagger calibration

The simulation of jets is not perfect. Differences between quantities in MC and data can arise from many sources, like the PS algorithm or the detector simulation. To mitigate these differences in a physics analysis, SFs are used. The SF is defined as the ratio between the efficiency in data events and MC events. The efficiencies in MC are easily obtained, however, the efficiencies in data are more challenging. Dedicated calibration procedures to measure these, and therefore the SFs are performed. In the following, the calibration of the *bc*-Tagger for light-flavour, *b*-, and *c*-jets is detailed. No dedicated calibration is performed for τ -jets. Instead, the SFs obtained for *c*-jets are applied to τ -jets as well.

The first calibration performed is the light-flavour jet calibration. The resulting SFs, together with their uncertainties, are utilised in the *b*-jet calibration. Lastly, the *c*-jet calibration uses the results of both the light-flavour and *b*-jet calibrations. This calibration chain is chosen as the light-flavour jet calibration simultaneously extracts the efficiencies of light-flavour and *b*-jets with a conservative estimate of the efficiencies of *c*-jets. The *b*-jet calibration has very little *c*-jet contamination but a considerable amount of light-flavour jet contamination, especially for the looser tagging bins where most of the light-flavour jets end up. To reduce systematic uncertainties from a conservative estimate of the light-flavour jet efficiencies, the corresponding SFs with their uncertainties are used instead. Lastly, the *c*-jet calibration has some contamination from both light-flavour and *b*-jets, and therefore their respective SFs are used to reduce uncertainties.

As all calibrations suffer from some degree of contamination of the other jet flavours and no single calibration is able to extract efficiencies for all three jet flavours at the same time, one could think about an iterative approach, performing the whole chain again utilising the latest SFs for each flavour except the one being calibrated. This would, however, have quite a large computational cost and, most likely, little performance gain.

C.1 Object definitions

LEPTONS The transverse and longitudinal IP requirements and the $|\eta|$ requirements on electrons and muons are the same as in the $t\bar{t}$ + charm analysis. The remaining requirements which are different and can also differ between the different parts of the calibration are listed in Table C.1

JETS Jets are reconstructed in the same way as for the $t\bar{t}$ + charm analysis. The same $|\eta|$ requirement and the same JVT cut is used. The requirement on the transverse momentum is, however, lowered to $p_T > 20$ GeV.

Table C.1: Requirements placed on the minimum transverse momentum and the isolation WP of leptons for the different parts of the bc -Tagger calibration.

	Min p_T [GeV]	Identification	Isolation
Electrons			
Light calibration	18	<i>LHTight</i>	<i>Gradient</i>
c -calibration	27	<i>LHTight</i>	<i>Gradient</i>
b -calibration	28	<i>LHTight</i>	<i>FCTight</i>
Muons			
Light calibration	18	<i>Tight</i>	<i>PFlowLoose</i>
c -calibration	27	<i>Medium</i>	<i>TightTrackOnly</i>
b -calibration	28	<i>Tight</i>	<i>Tight</i>

C.2 Additional samples

Additional samples to the ones described in Section 5.7 are needed for the calibration of the bc -Tagger, though a lot of the samples used in the $t\bar{t}$ + charm analysis are also used for the calibration procedure. Table C.2 lists all samples used together with the calibration it is used for.

Process	ME	PS	Additional parameters	Usage
$t\bar{t}$	POWHEG	PYTHIA	-	MC-MC, light, c , b
	POWHEG	PYTHIA	h_{damp} variation	c , b
	POWHEG	PYTHIA	$W \rightarrow cb$	c
	POWHEG	HERWIG7.0.4	-	MC-MC, c , b
	POWHEG	HERWIG7.1.3	-	MC-MC
	MADGRAPH5_AMC@NLO	PYTHIA	-	MC-MC
	SHERPA2.2.1	SHERPA2.2.1	-	MC-MC
single top: Wt , s -, t -channel	POWHEG	PYTHIA	-	c , b
	POWHEG	PYTHIA	only Wt , DS	b
	POWHEG	HERWIG7.0.4	-	b
$t\bar{t}V$	MADGRAPH5_AMC@NLO	PYTHIA	-	c
Z + jets	SHERPA	SHERPA	-	light, c , b
	MADGRAPH5_AMC@NLO	PYTHIA	-	light
W + jets	SHERPA	SHERPA	-	c , b
VV	SHERPA	SHERPA	-	light, c , b

Table C.2: List of samples generated in the ATLAS framework used in the bc -Tagger calibration. The "Usage" column describes in which calibration the sample is used in.

C.2.1 Top-quark pairs

POWHEG BOX + PYTHIA An additional sample was generated using the nominal POWHEG BOX + PYTHIA setup. In this sample, however, one W boson was forced to decay leptonically, while the other had to decay into a c -quark and a b -quark.

POWHEG BOX + HERWIG A sample using the same setup as for the POWHEG BOX + PYTHIA sample for the ME generation but the HERWIG 7.0.4 PS was generated. The H7UE tune was used for HERWIG 7.0.4. The MMHT2014LO PDF set was used for the PS.

MADGRAPH + PYTHIA An additional set of events was modelled with the MADGRAPH5_AMC@NLO 2.6.0 generator with the

NNPDF3.0_{NLO} PDF set for the ME calculation. The events were interfaced with PYTHIA 8.230.

SHERPA An additional sample of $t\bar{t}$ production was produced using the SHERPA 2.2.1 generator. The ME is accurate to NLO for up to one additional parton and to LO for up to four additional partons calculated with the Comix and OPENLOOPS libraries. The MEPS@NLO prescription together with the set of tuned parameters developed by the SHERPA authors to match the NNPDF3.0_{NNLO} PDF set was used to interface them with the SHERPA-internal PS.

C.2.2 Single electroweak bosons

MADGRAPH + PYTHIA An additional set of events was produced using the MADGRAPH5_AMC@NLO event generator. Only Z + jets events were modelled using MADGRAPH5_AMC@NLO. Up to four partons were generated at ME level with LO precision. For decays of the Z boson into tau leptons, MADGRAPH 2.2.2 with the NNPDF2.3_{LO} PDF set was used. The events were interfaced to PYTHIA 8.186 for the PS, hadronisation, and underlying event. For decays into electrons or muons, MADGRAPH 2.2.3 with PYTHIA 8.210 and the NNPDF3.0_{NLO} PDF set was used instead.

C.3 MC-MC scale factors

Before determining the efficiencies of all flavours in data, MC-MC SFs are derived. As the PS algorithm has a considerable impact on the distribution of the variables that are used for flavour tagging, different PS algorithms will have different efficiencies. SFs are derived to correct these differences. Top-quark pair production samples are used to obtain these. The different samples for which efficiencies are derived are given in Table C.3. To obtain the MC-MC SFs, the efficiencies of the POWHEG BOX + PYTHIA 8 sample are divided by the respective sample, adjusting the efficiencies of all samples to be the same as the ones of the POWHEG BOX + PYTHIA 8 sample. The SFs derived for SHERPA 2.2.1 are also applied to samples generated with SHERPA 2.2.0 and SHERPA 2.2.2. Efficiencies of all samples are calculated by selecting all jets of the corresponding flavour. They are obtained for two different choices of binning in p_T and $|\eta|$: the calibration binning and the fine binning. The bins in p_T do not differ between these two choices, but the calibration binning is inclusive in $|\eta|$, while the fine binning has four bins in $|\eta|$. The bins in p_T are different for the different flavours, with only c - and τ -jets having the same binning. The bin edges for all flavours in p_T and the bin edges of the fine binning in $|\eta|$ can be seen in Table C.4. The p_T bins are also later used to derive efficiencies in data.

The resulting SFs for all samples for light-flavour, c -, and b -jets can be seen in Figure C.1. The largest deviations for light-flavour jets are about 15%. No ME generator PS algorithm pair deviates

Table C.3: $t\bar{t}$ samples used for MC-MC SFs

POWHEG BOX + PYTHIA 8
POWHEG BOX + HERWIG 7.0
POWHEG BOX + HERWIG 7.1
MADGRAPH5_AMC@NLO + PYTHIA 8
SHERPA 2.2.1

Table C.4: Bin edges for the SFs in p_T and $|\eta|$.

Bin edges in p_T [GeV]				
Light-flavour jets				
20	50	100	150	300
b -jets				
20	30	40	60	85
110	140	175	250	400
c - and τ -jets				
20	40	65	140	250
Fine bin edges in $ \eta $				
0	0.6	1.2	1.8	2.5

clearly more than the others. For c -jets, this is different. Here, the SFs for SHERPA 2.2.1 are larger than unity by sometimes over 40% for the tagged bins. As a result, the SFs in the untagged bins for the SHERPA 2.2.1 sample are below one. For b -jets, SHERPA 2.2.1 has much better agreement with POWHEG Box + PYTHIA 8. Instead, the POWHEG Box + HERWIG 7.0 sample has the largest deviations from unity, however, to a much smaller degree than the largest deviation for c -jets.

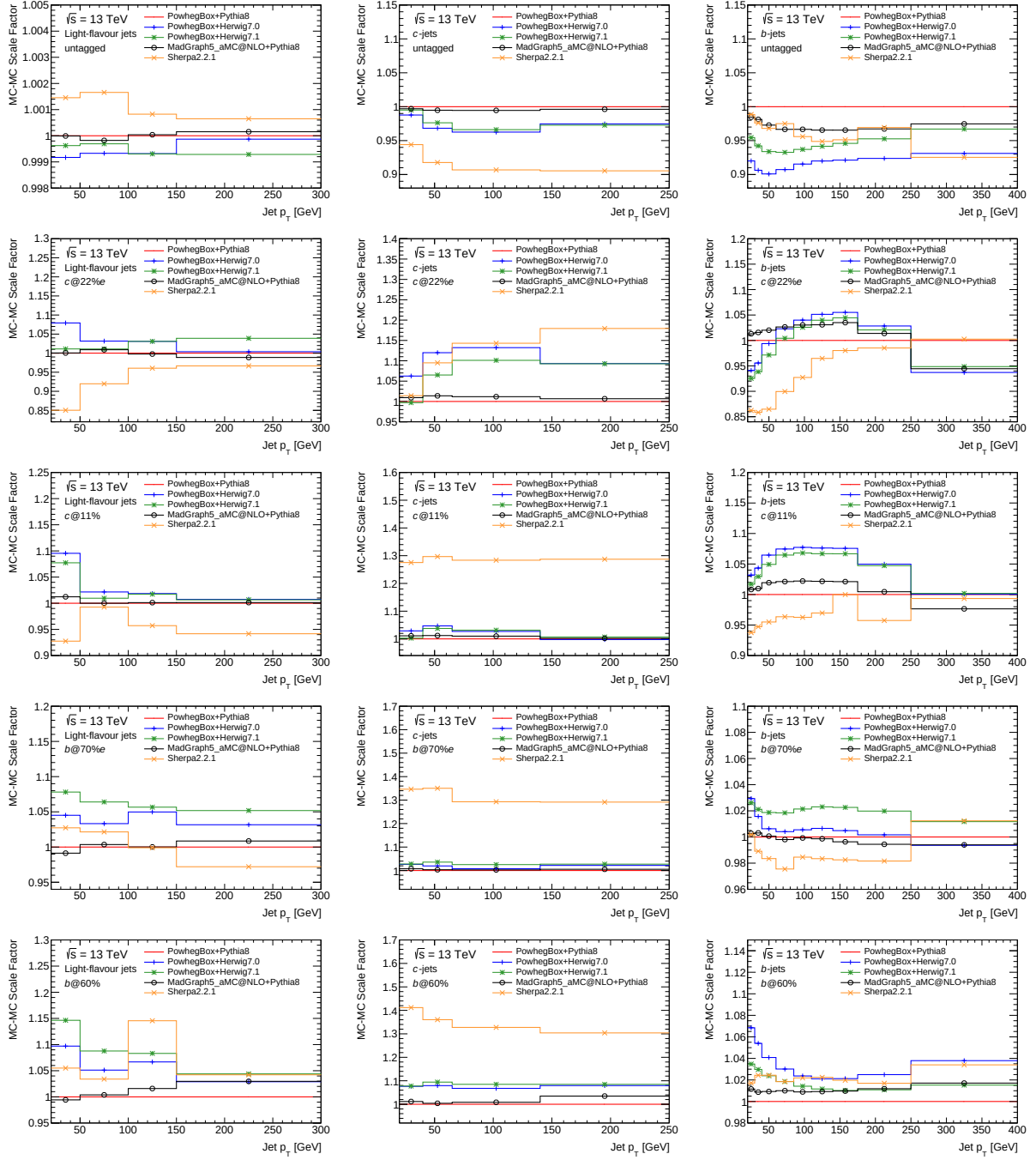


Figure C.1: MC-MC SFs for (left) light-flavour jets, (middle) c -jets, and (right) b -jets. The single rows correspond to the different tagging bins in order from top to bottom: untagged, $c@22\%$, $c@11\%$, $b@70\%$, and $b@60\%$.

C.4 Light-flavour jet calibration

The procedure for the calibration of light-flavour jets is detailed in Ref. [246]. It uses jets produced in conjunction with a Z boson. These jets will most of the time be light-flavour jets. To have a pure sample of Z + jets production, only decays into e^+e^- or $\mu^+\mu^-$ pairs are selected. As a fit observable, the mass of the secondary vertex of the jets, as found by the Single Secondary Vertex Finder algorithm [85], is used. If no vertex can be found, a negative mass is assigned to the jet.

DATA AND SIMULATED SAMPLES To perform the calibration, the dataset collected during Run 2 at the ATLAS detector corresponding to 139 fb^{-1} is used. Events need to pass the single lepton triggers given in Table 4.6. SM processes expected to contribute to the selection are generated using MC simulation. Only samples generated with the full simulation of the detector response using GEANT4 are used.

The main process contributing is Z + jets. The sample produced with SHERPA is used as the nominal sample, while the sample produced with MADGRAPH is used as an alternative sample. Diboson production and top-quark pair production constitute additional backgrounds. The diboson samples generated with the SHERPA generator are used, while the POWHEG BOX + PYTHIA top-quark pair production sample is used. Only the top-quark pair production sample with decays into two charged leptons is used. More details on the samples are given in Section 5.7. An overview of all simulated samples that are used can be seen in Table C.5.

Process	Generator
Z + jets	SHERPA 2.2.1
ZZ from gluon-gluon fusion	SHERPA 2.2.2
ZW and $q\bar{q}$ -initiated ZZ	SHERPA 2.2.1
$t\bar{t}$	POWHEG BOX v2
Z + jets (Z decays into e^+e^- and $\mu^+\mu^-$)	MADGRAPH 2.2.3
Z + jets (Z decays into $\tau^+\tau^-$)	MADGRAPH 2.2.2

Table C.5: Overview of samples used in the light-flavour jet calibration and their generator settings. Adapted from [246]

EVENT SELECTION To enrich the selection in Z + jets events, exactly two leptons are required. They need to be of the same flavour and have opposite electrical charges. The leading (subleading) lepton is required to have $p_T > 28 \text{ GeV}$ ($p_T > 18 \text{ GeV}$). The invariant mass of the pair of leptons has to be within 10 GeV of the Z boson mass ($81 \text{ GeV} < m_{\ell\ell} < 101 \text{ GeV}$). Additionally, the p_T of the pair has to be larger than 50 GeV. At least one jet with $p_T > 20 \text{ GeV}$ needs to be present in the event. The overlap removal described in Section 4.3.5 is used. An overview of the event selection can be seen in Table C.6.

NEGATIVE TAG METHOD Due to the high light rejection of most of the tagging bins, the calculation of SFs for light-flavour jets is difficult. Only a few light-flavour jets are selected, while a lot of b - or c -jets

Selection	e^+e^-	$\mu^+\mu^-$
Leptons	Two same-flavour opposite-sign leptons	
	leading ℓ : $p_T > 28$ GeV	
	subleading ℓ : $p_T > 18$ GeV	
	$81 \text{ GeV} < m_{\ell\ell} < 101 \text{ GeV}$	
	$p_T^{\ell\ell} > 50 \text{ GeV}$	
	$ \eta < 1.37$ or $1.52 < \eta < 2.47$	$ \eta < 2.5$
	$ d_0 /\sigma(d_0) < 5$	$ d_0 /\sigma(d_0) < 3$
	$ \Delta z_0 \sin \theta < 0.5 \text{ mm}$	$ \Delta z_0 \sin \theta < 0.5 \text{ mm}$
	<i>LHTight</i> identification	<i>Tight</i> identification
	<i>Gradient</i> isolation	<i>PflowLoose</i> isolation
Jets	≥ 1 jet	
	$ \eta < 2.5$	
	$p_T > 20 \text{ GeV}$	
	JVT cut if $p_T < 60 \text{ GeV}$ and $ \eta < 2.4$	
Overlap Removal	If an electron and a muon share a track: electron removed	
	$\Delta R_y(e, \text{jet}) < 0.2$: jet removed	
	then $\Delta R_y(e, \text{jet}) < 0.4$: electron removed	
	$\Delta R_y(\mu, \text{jet}) < 0.4$ and $n_{\text{tracks}}^{\text{jet}} \geq 3$: μ removed	
	$\Delta R_y(\mu, \text{jet}) < 0.4$ and $n_{\text{tracks}}^{\text{jet}} < 3$: jet removed	

Table C.6: Summary of requirements placed on objects to pass the event selection

are selected. To increase the ratio of light-flavour to b - and c -jets, a modified discriminant is used for the calibration.

Due to the high lifetime of b -jets, they have on average large positive IP values. Light-flavour jets have a more symmetric distribution around zero, where differences from zero are mostly due to resolution effects¹. By inverting the sign of the IPs of all tracks and passing the resulting jets through the same algorithms, a modified discriminant, the ‘flipped’ discriminant, can be obtained. The Negative Tag method [246–248] assumes that the efficiency of selecting a light-flavour jet does not change after this inversion, based on the assumption that the main cause for misidentifying light-flavour jets as b -jets is due to resolution effects in the track reconstruction. These resolution effects lead to tracks with large positive IP values, but at the same time should lead to tracks with large negative IP values, which after the inversion can mimic the large IP values of tracks inside b -jets. While the efficiency of selecting light-flavour jets remains similar, the efficiency of selecting b -jets drops. In the $b@60\%$ bin, around 0.9% of the jets are light-flavour jets when using the standard discriminant, while this fraction increases to 5.9% with the flipped discriminant. $\mathcal{D}_b$ and $\mathcal{D}_c$ can be seen for both the standard jets and the jets with flipped IP values in Figure C.2.

Efficiencies in MC are derived using the Z + jets sample generated with MADGRAPH5_AMC@NLO. This sample is chosen for consistency as it uses PYTHIA 8 as PS algorithm, which is the PS algorithm used for the other flavours. The prediction of the Z + jets sample generated with SHERPA is used for the shape of the mass of the

¹ Some long-lived particles with only light-flavour quarks, like the K_S^0 meson or the Λ baryon, exist, leading to a slight positive bias.

secondary vertex m_{SV} as the SHERPA sample has a better agreement with data than the MADGRAPH5_AMC@NLO sample.

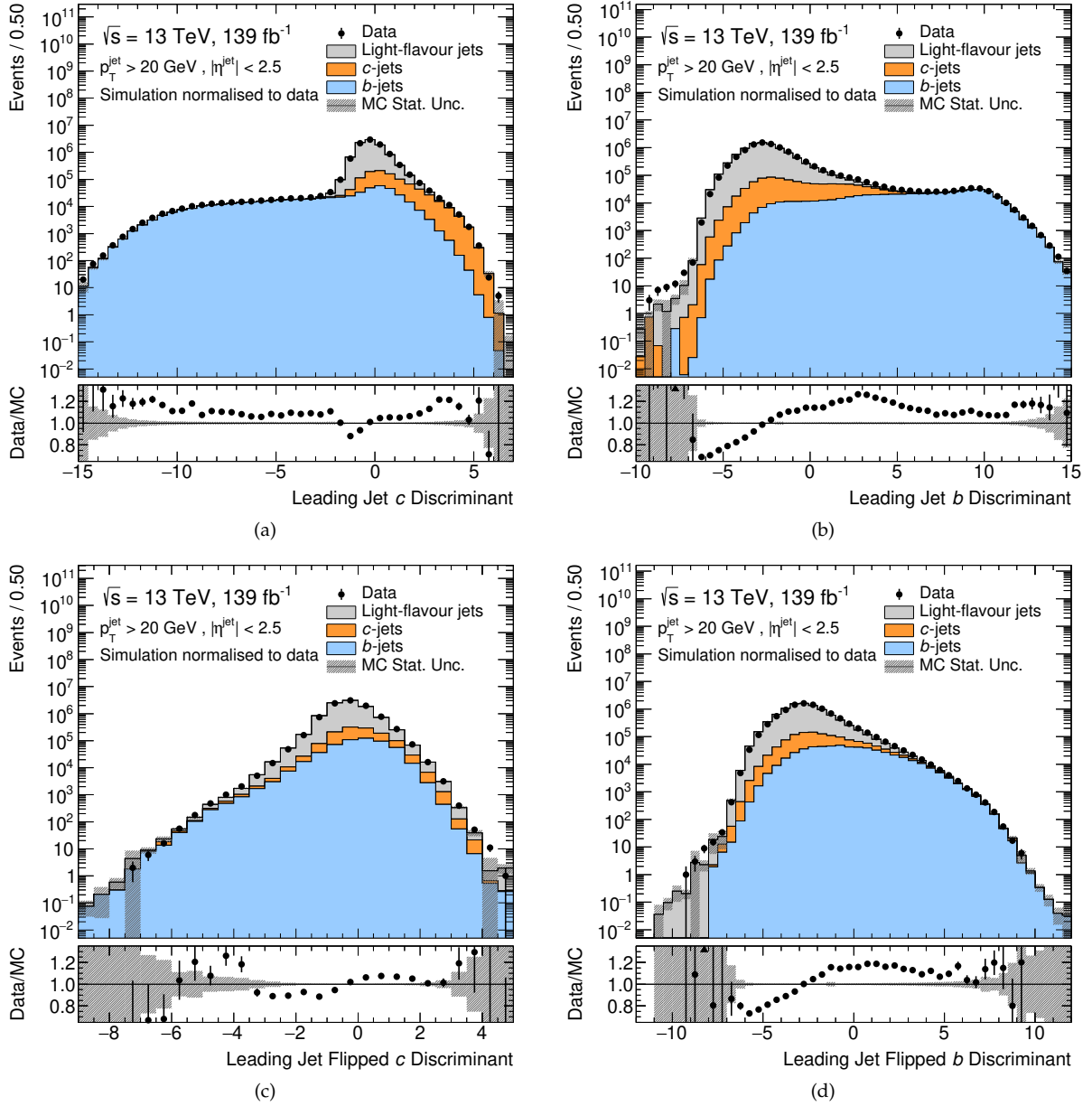


Figure C.2: Different discriminants after the event selection. (a) and (b) show the discriminants of the standard jets, whereas (c) and (d) show the discriminant when inverting the IP values of all tracks inside the jet. $\mathcal{D}_c$ is shown in (a) and (c), $\mathcal{D}_b$ is shown in (b) and (d).

The efficiencies in data of both b -jets and light-flavour jets are extracted during the fitting procedure. It is not possible to extract the c -jet efficiencies as well due to the sensitivity of the fit. While the b -jet efficiencies are extracted, SFs are obtained in a different, dedicated measurement that uses the standard discriminant. The SFs for c -jets are fixed to unity, and an additional uncertainty accounts for potential deviations. The efficiencies are determined independently in four p_T bins, with the bin edges shown in Table C.4. The mass of the secondary vertex m_{SV} is fitted simultaneously in all five tagging bins. No inversion of the IPs is performed for the determination of m_{SV} . Three bins in this variable are used in each of the tagging bins, leading to a total of 15 bins for each p_T bin. One of the bins represents

all jets where no secondary vertex has been found ($m_{SV} < 0$). The final fit function is given by

$$N_i(m_{SV}) = N_{i,MC} \times C \times \left[\sum_{f=\text{light},c,b} F^f \times SF_i^f \times \epsilon_{i,MC}^f \times P_i^f(m_{SV}) \right]. \quad (\text{C.1})$$

Here, i always represents the tagging bin. $N_{i,MC}$ is the number of events in each tagging bin as predicted by MC, C is an overall normalisation factor, F^f are the contributions of each jet-flavour given as fractions, SF_i^f are the SFs for each flavour, $\epsilon_{i,MC}^f$ are the efficiencies in MC for each flavour, and $P_i^f(m_{SV})$ are the PDFs of m_{SV} as predicted by MC.

The free-floating parameters of the fit are: C , F^f , and $SF_i^{b/\text{light}}$. $N_{i,MC}$ and $\epsilon_{i,MC}^f$ are fixed to the values obtained from simulated samples, while the SFs for c -jets SF^c are fixed to unity. Additional constraints are placed on the parameters. The total efficiency of each flavour in data as well as the jet-flavour fractions have to sum up to unity¹. This reduces the number of free parameters to a total of ten for each tagging bin.

¹ $\sum_i \epsilon_{i,MC}^f \times SF_i^f = 1$ and $F^{\text{light}} + F^c + F^b = 1$

Due to low statistics of light-flavour jets and a large contamination of HF jets, the light-flavour SFs for the $b@60\%$ tagging bin are set to unity. This is motivated by the SFs being consistent with unity for looser tagging bins: for the bc -Tagger the $b70\%e$ bin, and for DL1r the $77\% - 70\%$ and $70\% - 60\%$ tagging bins. The total uncertainties of the $b@70\%e$ bin are doubled and applied as the total uncertainties for the $b@60\%$ bin. Due to the very high rejection of light-flavour jets in this bin, small deviations from the true SFs will not affect the physics analysis by much. Figure C.3a shows the distribution of the single jet flavours in the single tagging bins for $50 \text{ GeV} < p_T < 100 \text{ GeV}$, while Figure C.3b shows the m_{SV} distribution in the same p_T bin for the $b@70\%e$ bin. Both distributions are shown post-fit.

UNCERTAINTIES Various systematic uncertainties are considered for the final result. They cover experimental effects and the modelling of the Z + jets process. To evaluate the effect of each uncertainty, a separate fit is performed, with the difference between the result and the nominal fit being taken as uncertainty. Additionally, an uncertainty is assigned due to the use of the flipped discriminant during the calibration. The final uncertainty is obtained by adding the single uncertainties in quadrature and taking the square root.

To estimate the effect of the modelling of the IPs, data-driven corrections are applied to the IPs d_0 and z_0 in a sample of simulated top-quark pair production. Utilising this sample, the effect of the corrections on the m_{SV} PDFs is calculated. This effect is propagated to the PDFs used in the fit.

The nominal result is obtained by utilising the SHERPA Z + jets sample. An alternative fit replacing the SHERPA Z + jets sample with the alternative MADGRAPH Z + jets sample is performed. The difference obtained in this fit is taken as the MC modelling uncertainty.

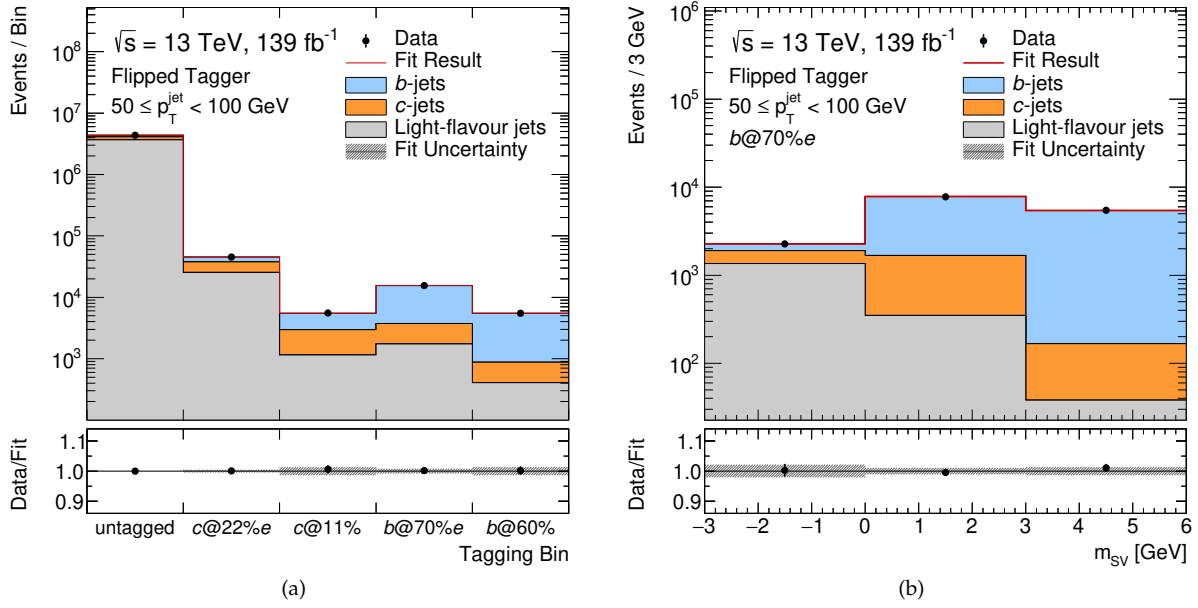


Figure C.3: Post-fit distribution of (a) the single jet flavours in the single tagging bins and (b) the fit variable m_{SV} in the $b@70\%e$ tagging bin. Both distributions are shown for the 50 GeV to 100 GeV p_T interval.

The effect of the uncertainty on the charm SF is estimated by varying the value of the SF ^{c} by 30% to 0.7 or 1.3 and repeating the fit. The mean impact is taken as the final uncertainty.

To account for differences between the flipped discriminant used to determine the efficiencies and the standard discriminant used in analyses, an additional uncertainty is added. The same uncertainty as derived for the DL1r WPs is used. No dedicated calculation has been performed for the bc -Tagger. The effect of using the flipped discriminant is estimated by comparing the efficiencies of the flipped discriminant to those of the standard discriminant on smeared and non-smeared MC samples. The smearing contains data-driven corrections for mostly track parameters. This uncertainty is derived for each tagging bin.

The statistical uncertainties of both data and MC simulated events are taken into account. For the uncertainty regarding MC simulated events, the Beeston-Barlow method is used [205].

RESULTS The obtained SFs can be seen in Figure C.4. All SFs are compatible with unity within uncertainties. Large uncertainties can be observed for the c@11% tagging bin. This bin suffers from similar problems as the b@60% tagging bin, namely low statistics of light-flavour jets and a large contamination of HF jets, but to a lesser degree. Since this bin has very high light rejection, these large uncertainties will not have a major effect on the analysis.

Table C.7 shows a breakdown of the single uncertainties in the different p_T bins for all tagging bins, as well as the total uncertainty. The leading uncertainty changes for the different tagging and p_T bins. The choice of the MC generator tends to lead to large uncertainties at low p_T but relatively small uncertainties at higher p_T . The flip uncertainty and the uncertainty due to the change in the charm SFs

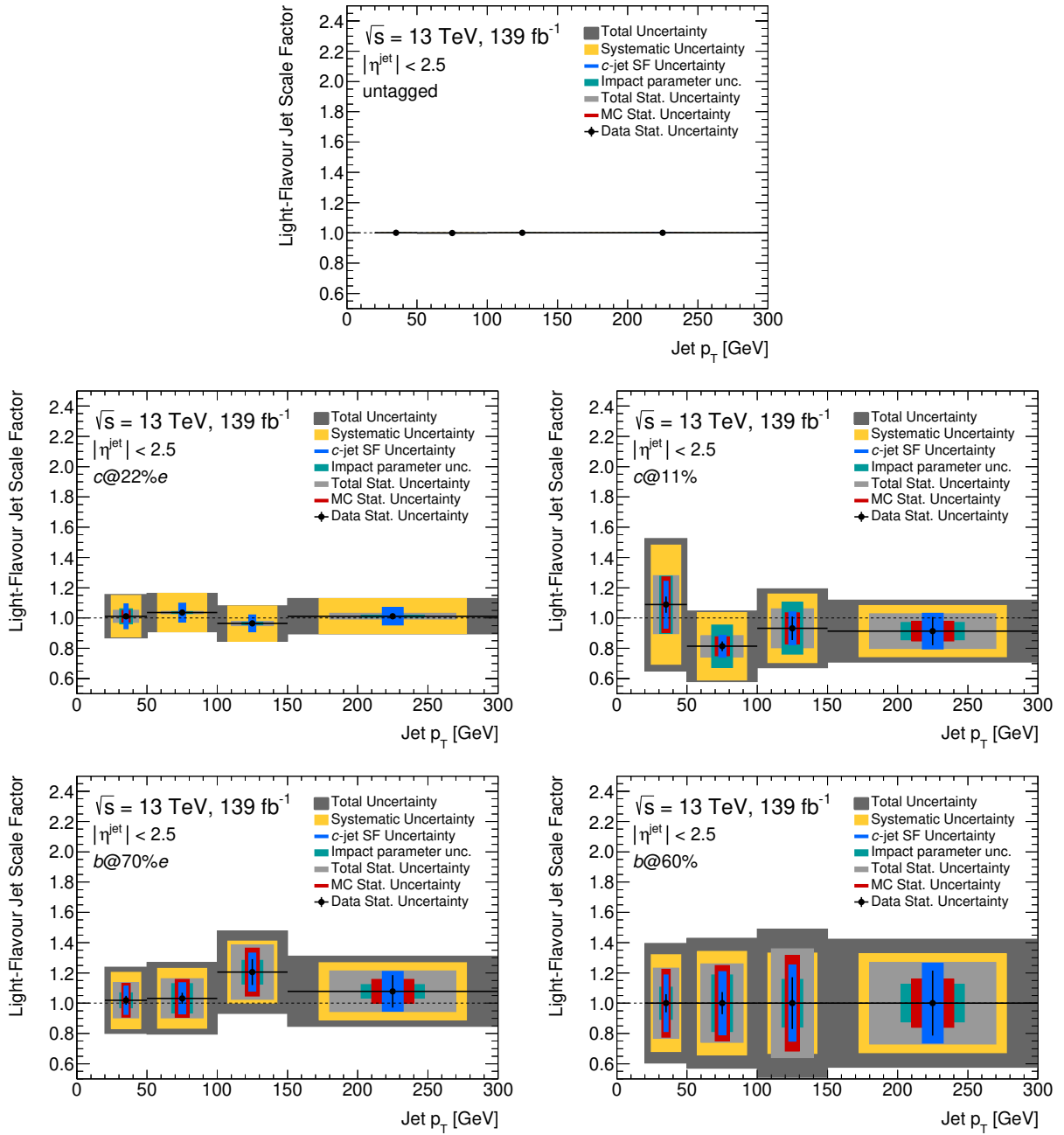


Figure C.4: Results of light-flavour jet SFs for all tagging bins. The different widths of the uncertainty bands is for visualisation purposes only and has no deeper meaning.

tend to have similar magnitudes, while the track IP uncertainty tends to be on the smaller side. In most bins, the systematic uncertainty is the dominant one, although the statistical uncertainties are also sizeable.

As almost all light-flavour jets end up in the untagged bin, the uncertainties in this bin are much smaller than in the other bins.

C.5 *b*-jet calibration

An overview of the calibration procedure for *b*-jets can be found in Ref. [249]. The calibration focuses on jets originating from top-quark

p_T bin [GeV]	[20, 50]	[50, 100]	[100, 150]	[150, 300]
Tagging Bin: untagged				
<i>c</i> -jet SF	0.0554	0.0488	0.0457	0.0641
Flipped tagger	0	0	0	0
MC generator	0.005 22	0.0398	0.0287	0.0193
IP modelling	0.0357	0.007 22	0.004 49	0.006 86
Data statistics	0.006 61	0.005 59	0.0101	0.0174
MC statistics	0.0228	0.005 99	0.006 32	0.004 41
Total (syst)	0.0661	0.0634	0.0541	0.0673
Total (stat)	0.0239	0.0083	0.0123	0.0195
Total	0.0703	0.064	0.0555	0.0701

p_T bin [GeV]	[20, 50]	[50, 100]	[100, 150]	[150, 300]	[20, 50]	[50, 100]	[100, 150]	[150, 300]
Tagging Bin: <i>c</i> @22% <i>e</i>					Tagging Bin: <i>c</i> @11%			
<i>c</i> -jet SF	8.17	5.98	5.81	5.72	14.2	8.87	11.7	13
Flipped tagger	9.7	9.7	9.7	9.7	10.6	10.6	10.6	10.6
MC generator	0.99	4.53	4.11	1.64	26.4	16.2	1.11	4.88
IP modelling	4.99	0.202	0.505	0.902	16	14.8	17.4	12.5
Data statistics	1.13	0.75	1.44	1.76	4.89	3.9	8.37	10.2
MC statistics	3.82	0.055	0.099	0.81	16.9	7.7	11	7.1
Total (syst)	13.7	12.3	12	11.4	35.5	25.9	23.5	21.5
Total (stat)	3.99	0.77	1.49	1.88	17.6	8.63	13.8	12.5
Total	14.2	12.3	12.1	11.6	39.6	27.3	27.3	24.9

p_T bin [GeV]	[20, 50]	[50, 100]	[100, 150]	[150, 300]	[20, 50]	[50, 100]	[100, 150]	[150, 300]
Tagging Bin: <i>b</i> @70% <i>e</i>					Tagging Bin: <i>b</i> @60%			
<i>c</i> -jet SF	9.15	10.1	10.4	12.2	18.7	20.8	25.2	26.3
Flipped tagger	11.1	11.1	11.1	11.1	11.9	11.9	11.9	11.9
MC generator	10.1	7.56	3.69	4.29	20.6	15.6	8.89	9.25
IP modelling	5.2	9.06	6.52	5.76	10.6	18.7	15.7	12.4
Data statistics	2.92	3.53	7.09	9.93	5.94	7.29	17.1	21.4
MC statistics	11	12.1	13.1	7.35	22.4	24.9	31.6	15.9
Total (syst)	18.3	19.1	17	18	32	34.2	33.2	32.7
Total (stat)	11.4	12.6	14.9	12.5	23.2	26	36	26.9
Total	21.6	22.8	22.6	21.9	39.5	42.9	49	42.4

Table C.7: Absolute uncertainties [%] of the light-flavour jet scale factors.

decays where both *W* bosons decay leptonically. Therefore, the only jets from the hard scatter at leading order are *b*-jets. To have the purest sample with enough statistics, events with an electron and a muon are targeted.

DATA AND SIMULATED SAMPLES The dataset of *pp* collisions collected during Run 2 at the ATLAS detector, corresponding to an integrated luminosity of 139 fb^{-1} is used. The single lepton triggers described in Table 4.6 need to be passed by every event. Simulated events of all SM processes expected to contribute to the selection are used.

As nominal samples for top-quark pair production and for all single top samples, the samples generated with POWHEG BOX + PYTHIA are used. For the remaining processes, *Z* + jets, *W* + jets, and diboson, the SHERPA samples are used. All nominal samples are produced with

the full detector simulation. The top-quark pair production sample as well as the single top samples are additionally produced with AFII, as alternative samples are mostly only available with AFII.

As alternative samples, the top-quark pair production sample produced with POWHEG BOX + PYTHIA with a modified h_{damp} parameter and the sample produced with POWHEG BOX + HERWIG 7.0.4 top-quark pair production sample are used. The POWHEG BOX + HERWIG samples are also used for the single top processes. Lastly, the tW sample using the DS scheme is used. This sample is the only alternative sample that uses the full detector simulation.

Process	Generator	Parton Shower	Detector Simulation	Additional
$t\bar{t}$	POWHEG BOX v2	PYTHIA 8	Full	
	POWHEG BOX v2	PYTHIA 8	AFII	
	POWHEG BOX v2	PYTHIA 8	AFII	h_{damp} variation
	POWHEG BOX v2	HERWIG 7.0.4	AFII	
single top	POWHEG BOX v2	PYTHIA 8	Full	tW with DR
	POWHEG BOX v2	PYTHIA 8	AFII	tW with DR
	POWHEG BOX v2	HERWIG 7	AFII	
tW	POWHEG BOX v2	PYTHIA 8	Full	DS
Z + jets	SHERPA 2.2.1	SHERPA	Full	
W + jets	SHERPA 2.2.1	SHERPA	Full	
Diboson	SHERPA 2.2.2 or 2.2.1	SHERPA	Full	

Table C.8: Overview of samples used in the b -jet calibration and their generator settings.

EVENT SELECTION To enrich the selection in dileptonic top-quark pair decays, exactly one electron and exactly one muon are required. The contribution of Z boson decays is reduced by requiring different flavours for the two leptons. An additional cut on the invariant mass of the $e\mu$ system of 50 GeV is applied to reduce the contribution of $Z \rightarrow \tau\tau$ decays. Due to the additional neutrinos in $Z \rightarrow \tau\tau$ decays, the mass of the potential $e\mu$ pair will typically be lower.

Exactly two jets are required, enriching the available jets in b -jets. To further separate the two jets based on their flavour, the mass of a jet and a lepton $m_{\text{lep},j}$ is formed. As there are two leptons and two jets, a pairing has to be fixed. It is chosen such that it minimises $m_{\text{lep},j1}^2 + m_{\text{lep},j2}^2$ with $j1$ being the jet with the higher p_T and $j2$ the jet with the lower p_T . To reduce events with soft jets being close to soft leptons, a requirement of $m_{\text{lep},j} > 20$ GeV is applied.

The standard overlap removal described in Section 4.3.5 is applied. An overview of all selection requirements can be seen in Table C.9.

FIT SETUP The events are classified based on the flavours of the two jets. If both jets are b -jets, the event is labelled as bb , while if both jets are non- b -jets, it is labelled as ll . For events with one b -jet and one non- b -jet, it is labelled as either bl or lb depending on the ordering in p_T of the jets, with the flavour of the higher p_T jet being written first.

For top-quark pair events with both b -jets being selected, the $m_{\text{lep},j}$

Selection	Electrons	Muons
Two different-flavour opposite-sign leptons		
$p_T > 28 \text{ GeV}$		
$m_{e\mu} > 50 \text{ GeV}$		
Leptons	$ \eta < 1.37 \text{ or } 1.52 < \eta < 2.47$	$ \eta < 2.5$
	$ d_0 /\sigma(d_0) < 5$	$ d_0 /\sigma(d_0) < 3$
	$ \Delta z_0 \sin \theta < 0.5 \text{ mm}$	$ \Delta z_0 \sin \theta < 0.5 \text{ mm}$
	<i>LHTight</i> identification	<i>Tight</i> identification
	<i>FCTight</i> isolation	<i>FCTight</i> isolation
Jets	Exactly two jets	
	$ \eta < 2.5$	
	$p_T > 20 \text{ GeV}$	
	JVT cut if $p_T < 60 \text{ GeV}$ and $ \eta < 2.4$	
Event Level	$m_{\text{lep},j1} > 20 \text{ GeV}, m_{\text{lep},j2} > 20 \text{ GeV}$	
Overlap Removal	If an electron and a muon share a track: electron removed	
	$\Delta R_y(e, \text{jet}) < 0.2$: jet removed	
	then $\Delta R_y(e, \text{jet}) < 0.4$: electron removed	
	$\Delta R_y(\mu, \text{jet}) < 0.4$ and $n_{\text{tracks}}^{\text{jet}} \geq 3$: μ removed	
	$\Delta R_y(\mu, \text{jet}) < 0.4$ and $n_{\text{tracks}}^{\text{jet}} < 3$: jet removed	

have an upper limit around the top mass $m_{\text{top}} = 172.5 \text{ GeV}$. Usually, they are even lower due to the missing neutrino. This behaviour can be seen in Figure C.5. Most of the events of the *bb* class have values of $m_{\text{lep},j}$ below the top mass. Additionally, for most events, the $m_{\text{lep},j}$ with the first (second) jet in the *bl* (*lb*) class is below the top mass as well.

Table C.9: Summary of requirements placed on objects to pass the event selection

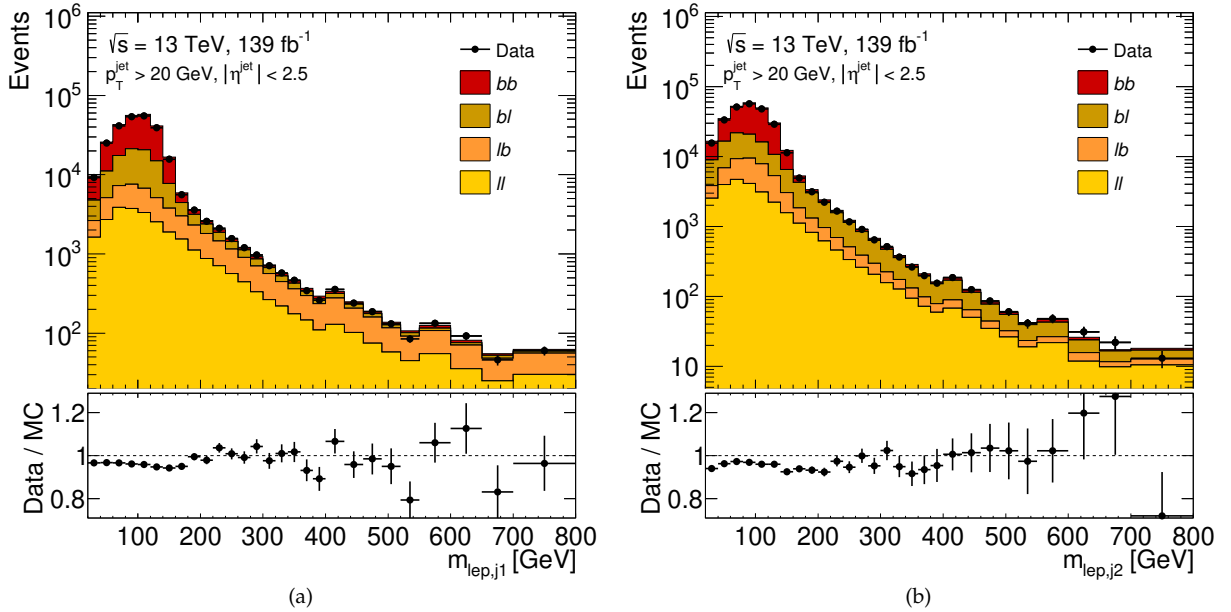


Figure C.5: Distribution of the mass of a jet and its associated lepton for (a) the leading and (b) subleading jet after the event selection.

Based on this feature, the events are further separated. A signal region (SR) is formed by requiring $m_{\text{lep},j}$ of both pairs to be below

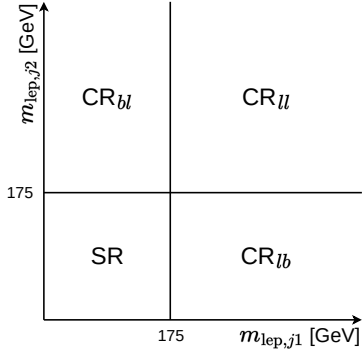


Figure C.6: Selection of the SR and CRs for the b -jet calibration.

175 GeV. This region will be dominated by events labelled as bb . The remaining events are further separated into three control regions (CRs) enriching the selection in bl , lb , or ll events. For the bl CR, $m_{lep,j1} < 175$ GeV and $m_{lep,j2} \geq 175$ GeV is required. The lb CR flips these requirements ($m_{lep,j1} \geq 175$ GeV and $m_{lep,j2} < 175$ GeV), while all remaining events constitute the ll CR with $m_{lep,j}$ being larger than 175 GeV for both pairs. A sketch of this selection can be seen in Figure C.6.

In each of the regions, the event is placed into a bin based on the p_T of the two jets. As there are nine p_T bins, as can be seen in Table C.4, the events can end up in 45 bins based on $p_{T,j1}, p_{T,j2}$. In the SR, the events are further binned based on the tagging bin of both jets, leading to 25 bins. Therefore, there are 1260 bins in total.

An extended log-likelihood fit is used to extract the b -jet tagging efficiencies in data. After dropping terms that do not depend on the parameters to be estimated, the fit function can be written as

$$\log \mathcal{L}(\nu_{\text{tot}}, \Theta) = -\nu_{\text{tot}} + \sum_i^N n_i \log \nu_i(\nu_{\text{tot}}, \Theta). \quad (\text{C.2})$$

Here, ν_{tot} is the total number of expected events, Θ all parameters to be estimated, and ν_i (n_i) the number of expected (observed) events in the bin i of all N bins. The b -jet tagging efficiencies in data $\mathcal{P}_b$ are determined for each tagging bin (O^k) $_{k=1\dots5}$ conditionally on the p_T -bin of the jet (T^m) $_{m=1\dots9}$. The expected number of events in the single CRs can be written as the sum of the expected number of events of the bb , bl , lb , and ll contributions. Correction factors dependent on the p_T of the jets are included to correct for data-MC mismodelling. Therefore, the yields in each CR for the single p_T bins can be written as

$$\nu_{\text{CR}}(T^m, T^n) = c_{bb}^{m,n} \nu_{\text{CR},bb}^{m,n} + c_{bl}^{m,n} \nu_{\text{CR},bl}^{m,n} + c_{lb}^{m,n} \nu_{\text{CR},lb}^{m,n} + c_{ll}^{m,n} \nu_{\text{CR},ll}^{m,n} \quad (\text{C.3})$$

with c being the correction factor for the respective jet flavour combination in the single p_T bin combinations. The correction factors in the CRs and the SR are the same. In the SR, the events are additionally binned in the single tagweight bin. Therefore, the yield in the SR can be written as

$$\begin{aligned} \nu_{\text{SR}}(T^m, T^n, O^k, O^\ell) = & c_{bb}^{m,n} \nu_{\text{SR},bb}^{m,n} \cdot \mathcal{P}_b(O^k|T^m) \cdot \mathcal{P}_b(O^\ell|T^n) \\ & + c_{bl}^{m,n} \nu_{\text{SR},bl}^{m,n} \cdot \mathcal{P}_b(O^k|T^m) \cdot \mathcal{P}_l(O^\ell|T^n) \\ & + c_{lb}^{m,n} \nu_{\text{SR},lb}^{m,n} \cdot \mathcal{P}_l(O^k|T^m) \cdot \mathcal{P}_b(O^\ell|T^n) \\ & + c_{ll}^{m,n} \nu_{\text{SR},ll}^{m,n} \cdot \mathcal{P}_l(O^k|T^m) \cdot \mathcal{P}_l(O^\ell|T^n). \end{aligned} \quad (\text{C.4})$$

Here, the $\mathcal{P}_l$ are the tagging efficiencies of the mixture of c - and light-flavour jets obtained after the selection. The b -jet tagging efficiencies and the correction factors are left to float during the minimisation procedure. The remaining parameters are obtained from simulation. The SFs for light-flavour jets, both data-MC and MC-MC, are applied to calculate the expected number of events in each bin.

UNCERTAINTIES To evaluate systematic uncertainties, a fit is performed for each systematic uncertainty by varying one parameter at a time. When an up and a down variation are provided or can be defined, half of the difference between these two is taken as symmetrised uncertainty. If only one variation is available, the difference from the nominal setup is taken as the resulting symmetrised systematic uncertainty.

Several uncertainties related to detector effects are considered. Uncertainties of the energy scale and resolution of hadronic jets, uncertainties of the energy scale of electrons and muons, SF uncertainties of electron and muon reconstruction, trigger, identification, and isolation SFs, SF uncertainties of the JVT SF, and uncertainties related to the modelling of pileup are separately taken into account. As the event selection has some light-flavour jet contamination, especially for the untagged and the $c@22\%e$ bin, the uncertainties obtained in the light-flavour jet calibration are propagated to this measurement.

Several sources of top-quark pair modelling uncertainties are considered. The choice of the PS algorithm is included by replacing the nominal sample with the sample generated with POWHEG BOX + HERWIG 7.0.4. To estimate uncertainties related to FSR, the value of the renormalisation scale is varied by a factor 2 and 0.5 for the FSR using the provided internal weights. Evaluating the effect of other radiation effects is done by using the sample with the modified h_{damp} parameter together with the up variation of the Var3c parameters, while the down variation is defined by the nominal sample with the down variation of the Var3c parameters. PDF uncertainties are included following the PDF4LHC prescription [36].

Similar to top-quark pair modelling, the nominal samples are replaced with the POWHEG BOX + HERWIG samples to estimate the impact of the choice of the PS algorithm for single top modelling. Additionally, the effect of the overlap removal strategy of tW production and top-quark pair production is estimated by replacing the tW sample generated with the DR scheme with the sample generated with the DS scheme.

Variations of the renormalisation and factorisation scales are used as uncertainty for both Z + jets and diboson production. Additionally, the total cross-section is changed by $\pm 6\%$ ($\pm 5\%$) for Z + jets (diboson) production. PDF uncertainties are evaluated using bootstrap replicas of the NNPDF set provided within the samples.

The contribution of events with a non-prompt lepton is found to be non-negligible. As non-prompt leptons do not have dedicated SFs and their efficiencies are generally not well modelled in simulation, a data-driven correction for the number of events with one non-prompt electron is obtained. A fit with this correction applied is used as a systematic variation. Due to the lower contribution of events with a non-prompt muon, no correction is done for muons.

Statistical uncertainties of the data are extracted from the error matrix of the fit returned by MINUIT [250]. As they are correlated between various parameters due to the usage of two-dimensional

binning in p_T , a principal component analysis is performed to obtain several decorrelated components for each parameter.

RESULTS Figure C.7 shows the b -jet tagging efficiency in data and MC as well as the SFs in all tagging bins. The $b@60\%$ bin has efficiencies around 60%, as it was designed to. Similarly, the $b@70\%$ bin has an efficiency of around 10%, such that combined with the $b@60\%$ bin, it achieves an efficiency of around 70%. Both c -targeted bins have much lower efficiencies, with the largest being around 4%. As the $c@11\%$ bin contains jets with higher $\mathcal{D}_b$ than the $c@22\%$ bin, it is expected that it will have a higher b -jet tagging efficiency, as observed.

All SFs are close to unity. Although the definitions of the $b@60\%$ and $b@70\%$ bins are slightly different from the standard DL1r WPs, the SFs and their uncertainties are in agreement with the standard WPs.

Table C.10 lists the overall uncertainties of the SFs as well as the uncertainty associated with single groups of uncertainties. For jets with low p_T , the systematic uncertainties are always the dominant ones, whereas at higher p_T , the statistical uncertainty becomes dominant. Especially in the two c -tagging bins, the statistical uncertainties become large, with up to 17.6%. This is expected due to the low b -jet tagging efficiency in these two bins. The systematic uncertainties with the largest impact are mostly related to the modelling of top-quark pair production or the modelling of jets. Lepton, non-prompt lepton, and pileup uncertainties are never the dominant uncertainty. The uncertainties associated with the light-flavour jet SFs have the largest impact at low p_T for the $c@22\%$ bin. This bin suffers from a large light-flavour jet contamination; therefore, this large uncertainty is expected. The light-flavour jet SF uncertainties do not have a large impact on the untagged bin, which has even greater contamination, as the uncertainties obtained in the light-flavour jet calibration for this bin are also much smaller.

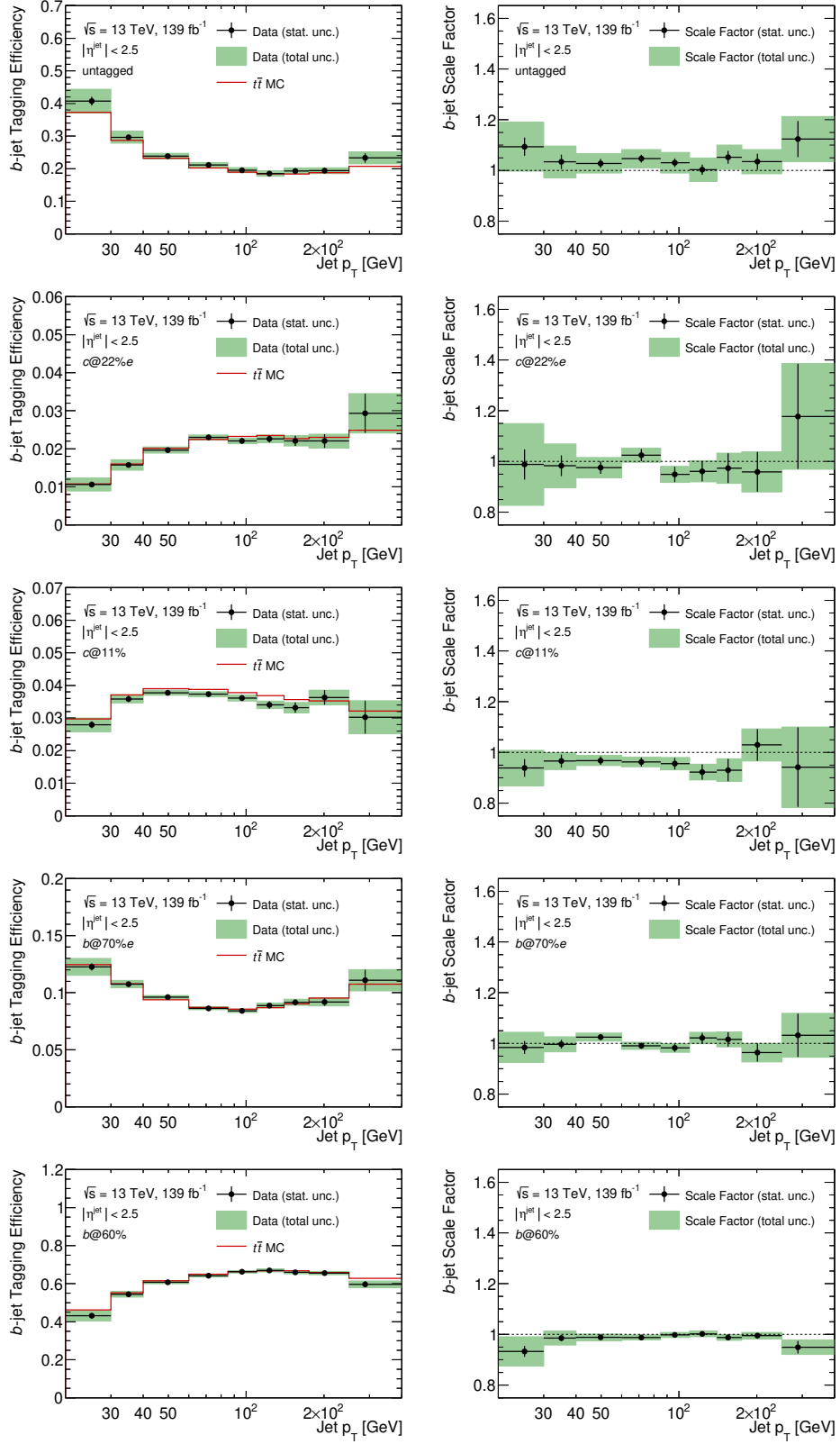


Figure C.7: *b*-jet (left) tagging efficiencies in MC and data, and (right) SFs for all tagging bins.

p_T [GeV]	[20, 30]	[30, 40]	[40, 60]	[60, 85]	[85, 110]	[110, 140]	[140, 175]	[175, 250]	[250, 400]
untagged									
Non-prompt leptons, pileup	1.08	0.23	0.22	0.16	0.19	0.16	0.16	0.12	0.05
Light-flavour SFs	0.39	0.33	0.36	0.12	0.11	0.13	0.11	0.1	0.15
Leptons	1.13	0.61	0.38	0.24	0.28	0.19	0.12	0.12	0.2
Jet modelling	5.76	3.38	1.06	0.69	0.52	0.41	0.41	0.75	0.75
$t\bar{t}$ modelling	4.62	3.31	2.27	2.24	3.02	3.81	3.54	2.97	2.19
Background modelling	2.4	1.88	1.93	1.92	1.71	1.58	1.37	1.51	4.03
Data statistics	3.26	2.78	1.66	1.45	1.66	1.99	2.41	3.06	6.32
MC statistics	2.11	1.58	0.63	0.6	0.49	0.5	0.46	0.55	0.86
Total (stat)	3.88	3.2	1.77	1.57	1.73	2.05	2.45	3.11	6.38
Total (syst)	7.93	5.14	3.21	3.05	3.53	4.15	3.82	3.42	4.65
Total	8.83	6.06	3.67	3.43	3.93	4.63	4.54	4.62	7.9
$c@22\%e$									
Non-prompt leptons, pileup	1.17	0.67	0.25	0.15	0.14	0.17	0.18	0.13	0.09
Light-flavour SFs	13.57	7.07	3.01	1.07	0.77	1.12	0.66	0.78	1.1
Leptons	2.38	0.26	0.21	0.37	0.05	0.07	0.16	0.13	0.12
Jet modelling	3.63	1.94	0.49	0.32	0.27	0.35	0.37	0.72	1.16
$t\bar{t}$ modelling	2.73	0.68	0.61	0.44	0.28	0.21	0.22	0.43	0.56
Background modelling	0.42	0.72	0.35	0.25	0.34	0.34	0.34	0.34	0.53
Data statistics	6.05	4.15	2.5	2.35	3.17	4.13	6.03	8.15	17.56
MC statistics	4.06	1.85	0.93	0.58	0.6	0.7	0.6	0.79	1.28
Total (syst)	14.56	7.43	3.15	1.29	0.94	1.25	0.89	1.21	1.78
Total (stat)	7.29	4.54	2.67	2.42	3.23	4.19	6.06	8.19	17.61
Total	16.28	8.71	4.13	2.74	3.36	4.37	6.12	8.28	17.7
$c@11\%$									
Non-prompt leptons, pileup	0.46	0.06	0.04	0.05	0.06	0.03	0.04	0.02	0.12
Light-flavour SFs	1.07	0.54	0.28	0.17	0.17	0.18	0.21	0.18	0.27
Leptons	0.72	0.24	0.1	0.08	0.16	0.04	0.04	0.04	0.18
Jet modelling	4.75	1.57	0.48	0.32	0.26	0.2	0.2	0.27	0.42
$t\bar{t}$ modelling	3.02	1.05	0.53	0.46	0.6	0.63	0.63	0.45	0.46
Background modelling	2.11	0.7	0.6	0.51	0.4	0.38	0.34	0.22	1.71
Data statistics	3.71	2.66	1.73	1.8	2.41	3.25	4.76	6.11	16.64
MC statistics	1.76	0.76	0.36	0.34	0.36	0.38	0.37	0.36	0.82
Total (syst)	6.17	2.1	0.98	0.78	0.8	0.78	0.77	0.6	1.85
Total (stat)	4.1	2.77	1.77	1.83	2.44	3.27	4.77	6.12	16.66
Total	7.41	3.47	2.02	1.99	2.57	3.36	4.84	6.15	16.76
$b@70\%e$									
Non-prompt leptons, pileup	0.76	0.08	0.07	0.02	0.03	0.03	0.03	0.04	0.06
Light-flavour SFs	0.38	0.13	0.11	0.04	0.04	0.03	0.04	0.03	0.05
Leptons	0.77	0.26	0.12	0.06	0.07	0.05	0.03	0.03	0.07
Jet modelling	3.86	1.72	0.34	0.19	0.13	0.13	0.14	0.24	0.38
$t\bar{t}$ modelling	2.82	1.21	0.69	0.58	0.72	0.87	0.84	0.73	0.94
Background modelling	1.79	0.84	0.64	0.53	0.41	0.34	0.31	0.37	1.11
Data statistics	2.57	1.75	1.12	1.18	1.55	1.96	2.77	3.71	8.26
MC statistics	1.59	0.7	0.22	0.15	0.15	0.18	0.15	0.13	0.35
Total (syst)	5.23	2.28	1.02	0.81	0.84	0.95	0.91	0.85	1.51
Total (stat)	3.02	1.88	1.14	1.19	1.56	1.97	2.77	3.71	8.27
Total	6.04	2.96	1.53	1.44	1.77	2.18	2.92	3.81	8.4
$b@60\%$									
Non-prompt leptons, pileup	0.8	0.1	0.06	0.03	0.04	0.03	0.03	0.03	0.02
Light-flavour SFs	0.12	0.04	0.08	0.01	0.01	0.01	0.02	0.01	0.01
Leptons	0.8	0.26	0.13	0.07	0.07	0.05	0.03	0.03	0.06
Jet modelling	4.42	1.75	0.47	0.24	0.14	0.1	0.11	0.19	0.22
$t\bar{t}$ modelling	3.02	1.38	0.81	0.67	0.77	0.89	0.84	0.74	0.8
Background modelling	0.84	0.86	0.66	0.54	0.44	0.38	0.34	0.4	1.21
Data statistics	2.32	1.23	0.57	0.44	0.48	0.56	0.74	0.97	2.61
MC statistics	1.42	0.75	0.21	0.16	0.13	0.12	0.1	0.13	0.28
Total (syst)	5.54	2.41	1.16	0.9	0.9	0.97	0.91	0.86	1.47
Total (stat)	2.72	1.44	0.61	0.47	0.5	0.57	0.75	0.98	2.63
Total	6.17	2.8	1.31	1.01	1.03	1.13	1.18	1.3	3.01

Table C.10: Absolute uncertainties [%]
of the b -jet scale factors.

C.6 c -jet calibration

The c -jet calibration procedure is detailed in Ref. [213]. It focuses on using c -jets from W boson decays. To get a pure sample, a selection of top-quark pair events with one W boson decaying leptonically and the other W boson decaying hadronically is applied. A reconstruction algorithm is used to identify the two jets originating from the W boson.

DATA AND SIMULATED SAMPLES The full pp collision dataset collected during Run 2 at the ATLAS experiment, corresponding to a total integrated luminosity of 139 fb^{-1} is used. Only events passing one of the single lepton triggers listed in Table 4.6 are considered. Different processes expected to contribute significantly to the selection were simulated, with the effects of pileup being applied by overlaying minimum bias events, as described in Section 5.5. These processes comprise top-quark pair production, single top-quark production, Z + jets and W + jets production, diboson production, and $t\bar{t} + V$ production. For top-quark pair production as well as single top-quark production, the samples generated with POWHEG BOX + PYTHIA are used as nominal prediction. An additional sample is used for the nominal prediction of top-quark pair production, in which the hadronic W boson is required to decay into a cb pair. All electroweak processes were generated with SHERPA 2.2.1 or 2.2.2. The MADGRAPH + PYTHIA sample is used for the $t\bar{t} + V$ prediction.

Alternative top-quark pair samples are used to estimate systematic uncertainties. The sample with a modified h_{damp} parameter as well as the POWHEG BOX + HERWIG 7.0.4 sample are used for this purpose. All used samples are given in Table C.11. More details of the single samples are given in Section 5.7.

Process	Generator	Parton Shower	Detector Simulation	Additional
$t\bar{t}$	POWHEG BOX v2	PYTHIA 8	Full	h_{damp} variation
	POWHEG BOX v2	PYTHIA 8	AFII	
	POWHEG BOX v2	HERWIG 7.0.4	AFII	$W \rightarrow cb$
	POWHEG BOX v2	PYTHIA 8	AFII	
single top	POWHEG BOX v2	PYTHIA 8	Full	tW with DR
Z + jets	SHERPA 2.2.1	SHERPA	Full	
W + jets	SHERPA 2.2.1	SHERPA	Full	
Diboson	SHERPA 2.2.2 or 2.2.1	SHERPA	Full	
$t\bar{t}V$	MADGRAPH5_AMC@NLO 2.3.3	PYTHIA 8	Full	

Table C.11: Overview of samples used in the c -jet calibration and their generator settings.

EVENT SELECTION To enrich the selection in $t\bar{t}$ decays in the lepton+jets, exactly one lepton is required. As there is a neutrino present in the event, the missing transverse momentum $E_{\text{T}}^{\text{miss}}$ is required to be larger than 20 GeV. Additionally, the transverse mass m_{T} of the missing transverse momentum and the lepton has to be larger than

40 GeV, with m_T being defined as

$$m_T = \sqrt{2p_T^\ell E_T^{\text{miss}}(1 - \cos \Delta\phi)}, \quad (\text{C.5})$$

where $\Delta\phi = \phi(E_T^{\text{miss}}) - \phi(\ell)$ is the azimuthal angle between the missing transverse momentum and the lepton.

Exactly four jets are required, with three fulfilling $p_T > 25$ GeV, while the fourth one can have a transverse momentum as low as 20 GeV. No events with more than four jets are used to avoid any overlap with the event selection to determine the $t\bar{t}$ + charm cross-section.

Using the four-momentum vectors of the four jets and the lepton together with the missing transverse momentum, a likelihood-based reconstruction of the top-quark pair system called KLFitter [228] is performed. Based on a likelihood function, it assigns a jet to the b -jet of the leptonically decaying top-quark, a jet to the b -jet of the hadronically decaying top-quark, and two jets to the hadronically decaying W boson. No flavour tagging information is used to perform the assignment. Henceforth, the jets associated with one of the b -jets are referred to as top-jets, while the other two jets are referred to as W -jets.

To further increase the purity of the selection, only events where the best permutation of the KLFitter algorithm has an output larger than -48 are considered. Additionally, both top-jets need to be in the $b@60\%$ tagging bin. Lastly, the invariant mass of the W boson reconstructed from the two W -jets has to be between $30 \text{ GeV} < m_W < 150 \text{ GeV}$. The standard overlap removal described in Section 4.3.5 is applied. All requirements for the event selection can be seen in Table C.12.

Since a recent measurement of the ATLAS experiment indicates an underestimation of top-quark pair production with additional b -jets in the simulation compared to data [6], events where at least one of the W -jets is a b -jet and both top-jets are b -jets are scaled by a value of 1.25. The SFs obtained from the light-flavour jet and b -jet calibration are applied to the top- and W -jets, if they are light-flavour jets or b -jets. However, no MC-MC SFs are applied. Since most of the events passing the selection are generated with POWHEG BOX + PYTHIA8, this will not have a large impact as all MC-MC SFs for these samples are equal to one.

FIT SETUP Jets are referred to as untagged if they are in the untagged tagging bin. All other jets are referred to as tagged. Events where both W -jets are tagged are excluded from the determination of the c -jet tagging efficiencies in data for simplicity. These make up less than a percent of the total events. The efficiencies are derived in four bins in p_T with the bin edges given in Table C.4. Events in which both W -jets are untagged are denoted by $N_{\text{data}}^{\text{untag}}(i, j)$ ($N_{\text{MC}}^{\text{untag}}(i, j)$) for data (MC) with i and j being the p_T bin of the two W -jets. The higher p_T bin is always taken as j such that $j \geq i$. For events where one W -jet is tagged, the number of events is denoted as $N_{\text{data}}^t(i, j)$. Here,

Selection	Electrons	Muons
Leptons	Exactly one lepton	
	$p_T > 27 \text{ GeV}$	
	$ \eta < 1.37$ or $1.52 < \eta < 2.47$	$ \eta < 2.5$
	$ d_0 /\sigma(d_0) < 5$	$ d_0 /\sigma(d_0) < 3$
	$ \Delta z_0 \sin \theta < 0.5 \text{ mm}$	$ \Delta z_0 \sin \theta < 0.5 \text{ mm}$
	<i>LHTight</i> identification	<i>Medium</i> identification
	<i>Gradient</i> isolation	<i>TightTrackOnly</i> isolation
Jets	Three jets with $p_T > 25 \text{ GeV}$ and one jet with $p_T > 20 \text{ GeV}$	
	$ \eta < 2.5$	
	JVT cut if $p_T < 60 \text{ GeV}$ and $ \eta < 2.4$	
Event Level	$E_T^{\text{miss}} > 20 \text{ GeV}$	
	$m_T > 40 \text{ GeV}$	
	KLFitter output > -48	
	$30 \text{ GeV} < m_W < 150 \text{ GeV}$	
	Two top-jets <i>b</i> -tagged at 60%	
Overlap Removal	If an electron and a muon share a track: electron removed	
	$\Delta R_y(e, \text{jet}) < 0.2$: jet removed	
	then $\Delta R_y(e, \text{jet}) < 0.4$: electron removed	
	$\Delta R_y(\mu, \text{jet}) < 0.4$ and $n_{\text{tracks}}^{\text{jet}} \geq 3$: μ removed	
	$\Delta R_y(\mu, \text{jet}) < 0.4$ and $n_{\text{tracks}}^{\text{jet}} < 3$: jet removed	

t is the tagging bin, i and j are again the p_T bins of the W -jets, with i being the bin of the tagged jet. In MC, these events have to be further separated. If the tagged W -jet is a c -jet, the number of events is denoted as $N_C^t(i, j)$, whereas if no c -jet is present in the event, it is denoted as $N_J^t(i, j)$. The last contribution are events in which one of the top-jets is a c -jet. As this top-jet will be in a different p_T bin, this number of events depends on a third p_T bin k , the bin of the top-jet. Therefore, the number of events in this class is given by $N_X^t(i, j, k)$.

A total of 32 free parameters are included in the fit. They are: the c -jet SFs in each p_T bin for each tagging bin $c^t(i)$, apart from the untagged bin, as well as an overall normalisation in each W -jet p_T bin combination $p(i, j)$. In the fit, the χ^2 given by

$$\chi^2 = \sum_{t=1}^4 \sum_{i=1}^4 \sum_{j=1}^4 \left[N_{\text{data}}^t(i, j) - p(i, j) \left(c^t(i) N_C^t(i, j) + N_J^t(i, j) + \sum_{k=1}^4 c^{t=4}(k) N_X^t(i, j, k) \right) \right]^2 / N_{\text{data}}^t(i, j) + \sum_{i=1}^4 \sum_{j=i}^4 \left(N_{\text{data}}^{\text{untag}}(i, j) - p(i, j) N_{\text{MC}}^{\text{untag}}(i, j) \right)^2 / N_{\text{data}}^{\text{untag}}(i, j), \quad (\text{C.6})$$

is minimised. As all top-jets need to be tagged in the $b@60\%$ bin, only $c^{t=4}(k)$ contributes to the term for events with a c -jet as a top-jet. The efficiencies in data for the untagged bins, and therefore the SFs, can be retrieved by subtracting the efficiencies of the remaining bins from 100%. Technically, this is implemented as an additional term in the

Table C.12: Summary of requirements placed on objects to pass the event selection.

χ^2 definition, with $c^{\text{untag}}(i)$ being a free parameter with 1% tolerance:

$$\sum_{i=1}^4 \left(1 - c^{\text{untag}}(i) \epsilon_{\text{MC}}^{\text{untag}}(i) - \sum_{t=1}^4 c^t(i) \epsilon_{\text{MC}}^t(i) \right)^2 / (1\%)^2. \quad (\text{C.7})$$

Here, $\epsilon_{\text{MC}}^{\text{untag}}(i)$ ($\epsilon_{\text{MC}}^t(i)$) are the c -jet tagging efficiencies obtained from simulation in the untagged (t tagging) bin in the i -th p_T bin.

UNCERTAINTIES A variety of different systematic uncertainties are taken into account. For each uncertainty, an additional fit is performed. If only one variation exists, for instance, the $t\bar{t}$ PS uncertainty, the resulting difference to the nominal fit is symmetrised around the nominal result. If there are two variations, for instance, the normalisation of all non- $t\bar{t}$ contributions, the average of the two differences is taken as symmetrised uncertainty. All systematic uncertainties are added in quadrature to obtain the total systematic uncertainty.

The uncertainties of light-flavour jet and b -jet tagging obtained in the previous two sections are propagated to this measurement. If the uncertainties are the same as the uncertainties considered in this analysis, they are treated as correlated. This concerns, for instance, uncertainties related to the jet energy scale and resolution. Additional uncertainties introduced by physics objects are related to additional uncertainties of the jet energy scale and resolution, JVT efficiencies, the track-based soft-term of the transverse missing momentum, electron and muon reconstruction, electron and muon momentum scale, and the modelling of pileup.

The remaining uncertainties all relate to the modelling of the generated samples. The choice of the PS algorithm for top-quark pair production is evaluated by replacing the nominal sample with the alternative POWHEG BOX + HERWIG 7.0.4 sample. The effects related to the emission of FSR are included by multiplying or dividing the value of the renormalisation scale for the FSR by a factor of two. An additional uncertainty related to radiation in general is included. For the up variation, the top-quark pair sample with the modified h_{damp} parameter is used. Additionally, the Var3c up parameters while also halving the factorisation and renormalisation scale are applied. For the down variation, the nominal top-quark pair sample is used with the Var3c down parameters and doubling the factorisation and renormalisation scale. The scaling of the $t\bar{t} + b\bar{b}$ component by a factor of 1.25 is scaled by $\pm 25\%$. As a last uncertainty, all non- $t\bar{t}$ components are also scaled by $\pm 25\%$.

RESULTS The measured c -jet tagging efficiencies as well as the SFs can be seen in Figure C.8. More than 60% of the c -jets end up in the untagged bin for all p_T bins. The two c -focused bins, $c@22\%$ and $c@11\%$, have overall efficiencies in data around 11% each but with large differences with respect to p_T . Especially the $c@22\%$ has an efficiency around 6% for the lowest p_T bin but 17% in the highest two bins. A good c -jet rejection can be observed for the two b -focused bins: $b@70\%$ and $b@60\%$.

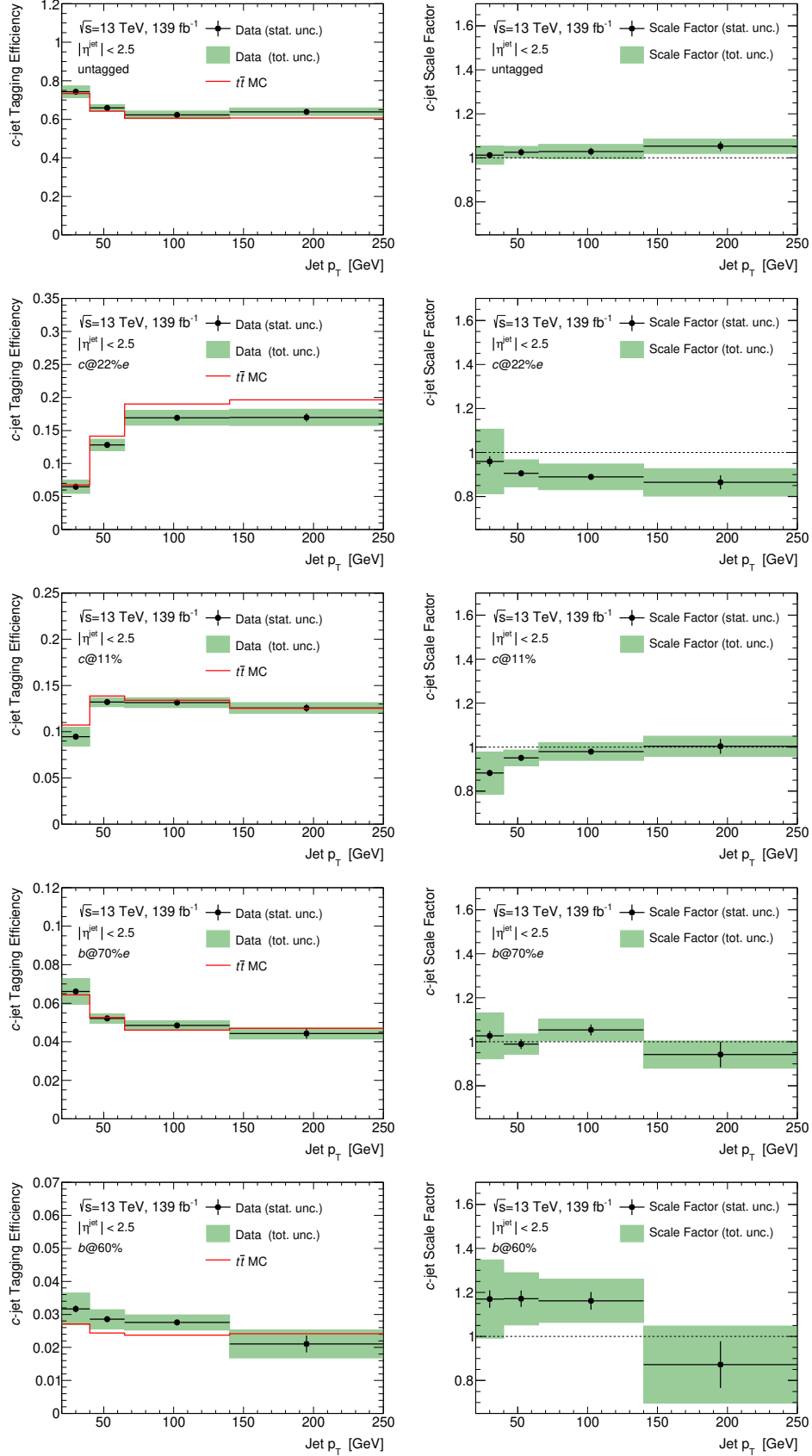


Figure C.8: c -jet (left) tagging efficiencies in MC and data, and (right) SFs for all tagging bins.

Deviations of the SFs from unity not covered by the uncertainty can be seen in many bins. Especially the $c@22\%$ bin has SFs which are significantly lower than one. Similar trends have been observed for an alternative definition of a bc -Tagger.

An overview of all uncertainties can be seen in Table C.13. Uncertainties that are correlated between this measurement and the light-flavour jet or b -jet calibration are grouped in their respective groups and not in the flavour tagging uncertainty. The dominant sources of uncertainties are always related to top-quark pair modelling, more precisely due to the choice of the PS algorithm. Uncertainties related to the modelling of jets and missing transverse momentum are mostly relevant in the $b@60\%$ tagging bin. Flavour tagging uncertainties have a large impact in the $c@22\%$ tagging bin, which has a large light-flavour jet contamination while also having sizeable light-flavour jet uncertainties. A similar behaviour can be observed for the uncertainties of the b -jet SFs. Lepton, pileup, and background uncertainties are always subdominant. The measurement is systematically limited for all tagging bins for the lower p_T bins. In the highest p_T bin, the size of the statistical uncertainties gets comparable to the size of the systematic uncertainties. In a few cases, the statistical uncertainties are even larger than the systematic uncertainties.

p_T bin [GeV]	[20, 40]	[40, 65]	[65, 140]	[140, 250]
untagged				
Pileup	0.16	0.07	0.02	0.05
Light-flavour and b SFs	0.94	1.0	0.93	0.81
Leptons	0.02	0.04	0.01	0.05
Jet modelling	0.5	0.27	0.28	0.45
$t\bar{t}$ modelling	3.77	1.77	2.52	2.16
Background modelling	0.09	0.05	0.07	0.09
Total (syst)	3.92	2.05	2.7	2.35
Total (stat)	1.34	1.56	1.74	2.18
Total	4.14	2.58	3.21	3.21

p_T bin [GeV]	[20, 40]	[40, 65]	[65, 140]	[140, 250]	[20, 40]	[40, 65]	[65, 140]	[140, 250]
$c@22\%$					$c@11\%$			
Pileup	0.97	0.29	0.06	0.06	0.27	0.05	0.06	0.02
Light-flavour and b SFs	6.91	4.22	3.02	2.58	2.19	0.79	0.39	0.23
Leptons	0.25	0.05	0.06	0.18	0.16	0.08	0.01	0.18
Jet modelling	2.88	0.81	0.78	1.12	1.23	0.8	0.57	1.05
$t\bar{t}$ modelling	13.05	4.87	5.55	5.55	10.41	3.34	3.78	2.91
Background modelling	0.65	0.23	0.19	0.23	0.53	0.22	0.19	0.16
Total (syst)	15.09	6.51	6.37	6.23	10.73	3.53	3.85	3.11
Total (stat)	2.38	1.43	1.23	3.17	1.54	1.3	1.36	3.39
Total	15.28	6.66	6.49	6.99	10.84	3.76	4.08	4.6

p_T bin [GeV]	[20, 40]	[40, 65]	[65, 140]	[140, 250]	[20, 40]	[40, 65]	[65, 140]	[140, 250]
$b@70\%$					$b@60\%$			
Pileup	0.24	0.0	0.12	0.01	0.34	0.06	0.06	0.62
Light-flavour and b SFs	1.37	0.88	0.75	1.06	0.48	0.34	0.49	1.05
Leptons	0.04	0.06	0.15	0.12	0.24	0.45	0.36	1.71
Jet modelling	0.93	0.56	0.7	1.43	3.56	3.37	3.03	7.88
$t\bar{t}$ modelling	9.8	3.96	3.91	1.33	14.32	8.91	7.04	13.73
Background modelling	0.29	0.13	0.1	0.15	1.44	1.02	0.6	0.36
Total (syst)	9.95	4.1	4.05	2.23	14.84	9.6	7.71	15.97
Total (stat)	2.21	2.23	2.47	5.83	3.96	3.76	4.0	10.61
Total	10.19	4.67	4.74	6.24	15.36	10.31	8.69	19.18

Table C.13: Absolute uncertainties [%]
of the c -jet scale factors.

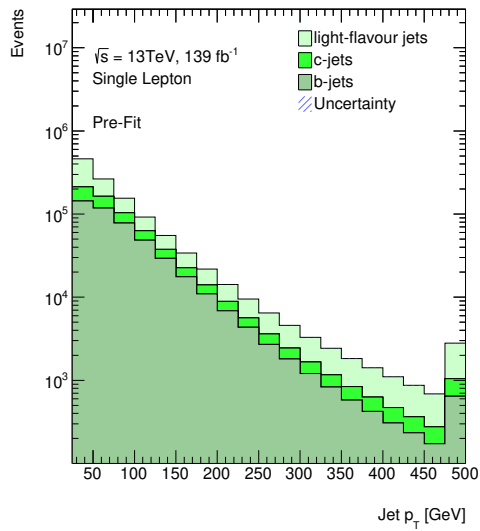
C.7 Extrapolation uncertainty

The range of p_T values that are calibrated for the single flavours does not extend to infinity. Instead, there is an upper bound until which the calibration is valid. This upper bound changes for the different flavours. To still use events with jets above this upper bound, an additional uncertainty is applied to these jets, the extrapolation uncertainty. As nominal values for the SFs, the values of the highest p_T bin that is still calibrated are used.

To determine how sophisticated the extrapolation uncertainty has to be for the $t\bar{t}$ + charm analysis, the number of jets outside the calibration region is estimated. Only events selected in the lepton + jets channel of the analysis are considered for this estimation. The p_T of every jet split by jet flavour can be seen in Figure C.9. For light-flavour jets, 1.6% of the jets are outside the calibration region, while for b -jets (c -jets), the value is 0.3% (2%). As single events can have multiple jets outside the calibration region, it is also interesting to estimate the number of events with at least one jet outside the calibration region. Figure C.10 shows the maximum p_T of all jets in each event split by flavours. The first bin of each plot corresponds to the calibration region of that specific flavour. Events with at least one light-flavour jet outside the calibration region make up around 4% of the total selection. For b -jets (c -jets), this value is smaller with 1% (2%)

As these values are relatively small, it was decided that no sophisticated derivation of an uncertainty has to be performed. Instead, the uncertainties of the highest calibrated p_T bin are doubled.

Figure C.9: p_T distribution of all jets split by jet flavour. Only events in one of the CRs or SRs of the lepton+jets selection are considered.



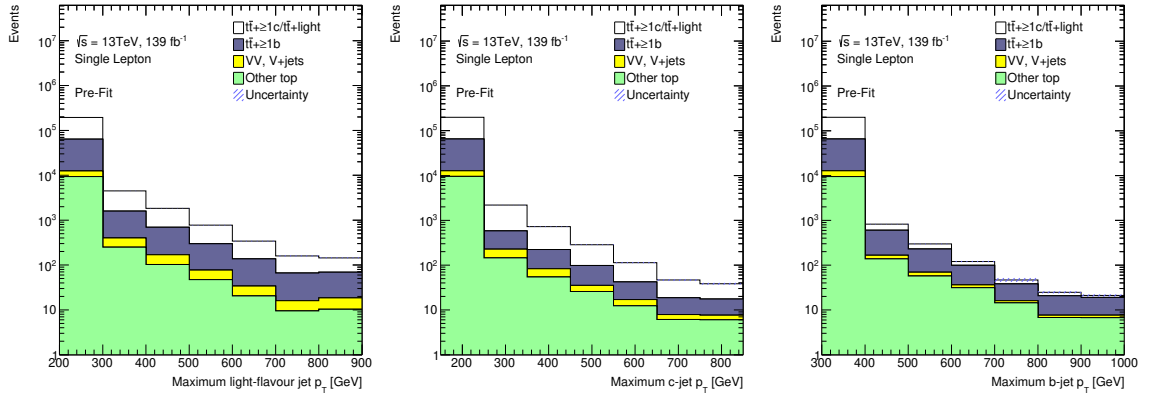


Figure C.10: Maximum p_T of (a) light-flavour, (b) c -, and (c) b -jets in each event. Only events in one of the CRs or SRs of the lepton+jets selection are considered.

D Topograph Hyperparameters

Hyperparameter	Value	
	TG	TT
Epochs	100	
Batch size	256	512
Optimizer	<i>AdamW</i>	
LR schedule	Cosine annealing	Cosine annealing with linear warm-up
Maximum LR	0.001	
Decay steps	2 Epochs	10 Epochs
Warm-up Epochs	-	5
N initial jet updates	2	4
N updates	4	4
W /top initialisation	Attention pooling	Cross-attention with learnable tokens
Activation	GELU	
Normalisation	Layer normalisation	
Graph processing units	[256, 256, 64]	-
Encoder Units	-	[128, 256, 128]
Multihead attention dropout	-	0.1
Attention Heads	-	16
Embedding Units	-	[128, 128]
Persistent edges	True	-
Pooling	Attention	-
Attention units	[32, 32, 1]	-
Regression units	[64, 64, 3]	[128, 128, 3]
Edge classification units	[128, 128, 128, 1]	
Classification loss	Weighted binary crossentropy	
Regression loss	MAE	
Classification loss weight	1	2
Library	TensorFlow v2.10	PyTorch v2.0
Number of parameters	6.5M	8M

Table D.1: Hyperparameters used for the training of TG and TT

List of Figures

2.1	Overview over the elementary particles of the SM	6
2.2	Feynman diagrams of gluon interactions	8
2.3	Feynman diagrams of photon interactions	10
2.4	Feynman diagram of Z boson interactions	10
2.5	Feynman diagrams of W boson interactions	10
2.6	Feynman diagrams of Higgs interactions	10
2.7	Summary plot of cross-section measurements of the ATLAS Collaboration during Run 1 and Run 2	11
2.8	Feynman diagram of top-quark pair production	12
2.9	Summary plot of top-quark-pair cross-section measure- ments at the Tevatron and the LHC	13
2.10	Feynman diagrams of the decay of a top-quark	14
2.11	Decay modes of top-quark pairs with their branching fractions	14
3.1	Current accelerator complex at CERN	16
3.2	Integrated luminosity at ATLAS during Run 2	16
3.3	Mean number of interactions at ATLAS during Run 2 . .	17
3.4	Overview of the ATLAS detector	17
3.5	Coordinate system of the ATLAS detector	18
3.6	Overview of the ATLAS ID	19
3.7	High threshold probability in the TRT	20
3.8	Overview of the ATLAS calorimeter system	21
3.9	Overview of the ATLAS electromagnetic calorimeter . .	22
3.10	Overview of the ATLAS MS	23
3.11	The ATLAS TDAQ system in Run 2	25
4.1	Interaction of different particles with the ATLAS detector	28
4.2	Formation of superclusters during electron reconstruction	31
4.3	Invariant dielectron mass after calibration	32
4.4	Reconstruction and identification efficiency of <i>Medium</i> muons	34
4.5	Invariant dimuon mass after calibration	35
4.6	Jet clustering using the anti- k_t algorithm	36
4.7	Single steps in the jet energy scale calibration	36
5.1	PDFs for two different choices of the factorisation scale .	42
5.2	Feynman diagram of $t\bar{t}b\bar{b}$ production	47
5.3	Feynman diagram of single top-quark s-channel production	48
5.4	Feynman diagram of single top-quark t-channel production	48
5.5	Feynman diagram of $t\bar{t}W$ production	48
5.6	Feynman diagram of $t\bar{t}Z$ production	48
5.7	Feynman diagram of $t\bar{t}W$ production	49
5.8	Feynman diagram of $t\bar{t}H$ production	49
5.9	Feynman diagram of $t\bar{t}Zq$ production	49
5.10	Feynman diagram of $t\bar{t}WZ$ production	49

5.11	Feynman diagram of Z + jets production	49
5.12	Feynman diagram of W + jets production	49
5.13	Feynman diagram of ZZ production	50
5.14	Feynman diagram of WZ production	50
5.15	Feynman diagram of WW production	50
5.16	Feynman diagram for $Z\gamma$ production	50
6.1	Sketch of the performance of a ML model with different complexities	54
6.2	MAE and MSE loss functions	55
6.3	Binary cross-entropy loss function	56
6.4	Sketch of a neural network	56
6.5	ReLU, Leaky ReLU, and GELU activation function and their derivatives around zero	56
6.6	Sigmoid activation function and its derivative around zero	57
6.7	Overview of a convolutional operation	60
6.8	Overview of a pooling operation	60
6.9	Schematic of a GNN operation	62
6.10	Overview over the graph operation used in this work	63
6.11	Overview of a transformer encoder layer	64
6.12	Schematic overview of ROC curves for different classifiers	65
6.13	Gaussian distribution with some outliers before and after three different preprocessing algorithms	66
7.1	Discriminant of the LH algorithm for signal and background taken from simulation	71
8.1	Distributions of (a) f_3 and (b) R_{had} of prompt electrons in data, simulation, and corrected simulation	79
8.2	E_T and η distribution of the different classes in the training dataset	80
8.3	Distributions of all input variables after the downsampling and reweighting is applied	82
8.4	Distributions of all input variables after the Quantile-Transformer is applied	83
8.5	Schematic overview of the architecture of the DNN	84
8.6	Evolution of the LR during the DNN training	84
8.7	Final discriminant $\mathcal{D}_{\text{el}}$ of the DNN for the six classes	87
8.8	ROC curves for different backgrounds for the DNN	88
8.9	Prompt electron efficiency and photon conversion rejection for three LH WPs	89
8.10	ROC curves with charge flip electrons as background for the DNN	89
8.11	Performance plots matching the DNN signal efficiency to that of the Loose LH WP	90
8.12	Performance plots matching the DNN signal efficiency to that of the Tight LH WP	91
8.13	ROC curves for different background comparing different versions of the discriminant $\mathcal{D}_{\text{el}}$	93

8.14	Comparison of the selection of prompt electrons based on the DNN and the LH	94
8.15	Comparison of the selection of the inclusive background based on the DNN and the LH	95
9.1	Schematic overview over the architecture of the networks processing the different point clouds with separate networks	101
9.2	Schematic overview over the architecture of the transformer-based network processing the different point clouds within the same point cloud after the embedding.	102
9.3	Schematic overview over the architecture of the transformer-based network processing the different point clouds including the track point cloud within the same point cloud after the embedding.	102
9.4	ROC curves for different backgrounds for the comparison of cell processing	104
9.5	ROC curves with charge flip electrons as background for the comparison of cell processing	105
9.6	Performance plots for the comparison of cell processing	105
9.7	ROC curves with charge flip electrons as background for the comparison of different amounts of information used by the networks	106
9.8	ROC curves for different backgrounds for the comparison of different amounts of information used by the networks	107
9.9	Performance plots for the comparison of different amounts of information used by the networks	108
10.1	Discriminant of candidates passing the signal selection	110
10.2	Discriminant of candidates passing the background selection	111
10.3	ROC curves using the signal and background selection in data for the LH and the two DNNs	112
10.4	ROC curves using the signal and background selection in data for the two DNNs and the point cloud-based models in different E_T ranges	113
10.5	ROC curves using the signal and background selection in data for the two DNNs and the point cloud-based models in different $ \eta $ ranges	114
11.1	Overview of the two-dimensional bc -Tagger	125
12.1	Data-MC agreement for tagged jets multiplicities after the preselection	129
12.2	Expected background composition in the ten fit regions	131
12.3	Expected signal contribution in the ten fit regions	132
12.4	Yields used in the fit for the three control regions	133
12.5	Observables used in the fit for the two signal regions	134
13.1	Systematic templates of one flavour tagging NP	136

13.2	Systematic templates of the JES ρ topology uncertainty for the $t\bar{t} + 1c$ and $t\bar{t} + \geq 2c$ samples	137
13.3	Systematic templates of the PS uncertainty for the $t\bar{t} +$ light sample in SR_{loose}	138
13.4	Systematic templates of the FSR uncertainty for the $t\bar{t} + 1c$ and $t\bar{t} + \geq 2c$ samples	139
13.5	Systematic templates of the ME generator uncertainty for the single top t -channel	141
13.6	Experimental uncertainties after the pruning	144
13.7	Modelling uncertainties after the pruning	145
14.1	Pull plot of the $t\bar{t} + \geq 1b$ PS uncertainty for two different fit setups in the fit to the Asimov dataset	147
14.2	Pull distributions of some NPs in the fit to ATLAS data	149
14.3	Systematic template of the PS uncertainty for the $t\bar{t} +$ light sample with ATLAS data overlaid in $SR_{\text{loose}}^{\geq 6j}$ with at least six jets	150
14.4	Correlation matrix between the NPs and cross-section norm factors in the fit to ATLAS data	151
14.5	Ranking plots for the fit to ATLAS data	152
14.6	Two-dimensional likelihood scan of $\mu(t\bar{t} + \geq 2c)$ and $\mu(t\bar{t} + 1c)$ for the fit to ATLAS data	155
15.1	Comparison of the fitted free-floating parameters in the dilepton, lepton+jets, and combined fit	159
15.2	Yields in the three control regions post-fit	162
15.3	Observables used in the fit for the two signal regions post-fit	163
15.4	Comparison of the measured fiducial cross-sections and ratios to the values predicted by several generated samples	166
16.1	High-level overview of the SPA-net architecture	172
16.2	Graph representations of top-quark pair production with one top-quark decaying semi-leptonically and the other hadronically	173
16.3	Comparison of the edges in fully-connected GNNs and Topographs for a single top-quark decay	174
17.1	Distribution of the jet and b -jet multiplicity in the two generated samples	177
17.2	Distribution of W boson and top-quark properties	177
17.3	Distribution of jet properties	178
17.4	Sketch of the initialisation part of the TG	180
17.5	Sketch of the updates during the TG training	181
17.6	Evolution of the LR during the TG training	181
17.7	Sketch of the initialisation part of the TT	182
17.8	Sketch of the updates during the TT	182
17.9	Evolution of the LR during the TT training	182

17.10	Reconstructed invariant mass of the W boson and the top-quark using the jet assignment by TG, SPA-net, χ^2 , and TT	185
17.11	Difference of the top-jet energies before and after smearing and the distribution of jet energies before and after scaling	187
17.12	Edge scores of jets to the helper node representing the W^+ boson of the TG model.	190
17.13	Edge scores of jets to the helper node representing the top-quark of the TG model.	190
17.14	Score of the TG and TT model for <i>correct</i> , <i>incorrect</i> , and <i>impossible</i> events	191
17.15	Score of the TG and TT model for <i>impossible</i> events	191
17.16	Resolution of W boson and top-quark kinematics either reconstructing the particles using the assigned jets or using the regression output of the TG model	192
17.17	Reconstructed top-quark p_T resolution using true jet assignment	193
17.18	Reconstructed top-quark p_T resolution with over- and underflow bins	193
B.1	Efficiencies in MC and data as well as the SFs using the DNN as measured by the Z_{mass} method for data collected in 2018 with respect to η	206
B.2	Efficiencies in MC and data as well as the SFs using the DNN as measured by the Z_{iso} method for data collected in 2018 with respect to η	206
B.3	Comparison of the SFs for positive and negative η values using the Z_{iso} method for data collected in 2017	208
B.4	Comparison of the SFs obtained with the Z_{mass} and Z_{iso} method for data collected in 2018	208
C.1	MC-MC SFs	212
C.2	b and c discriminants for the standard and flipped tagger after the event selection for the light-flavour jet calibration	215
C.3	Post-fit distributions of the single jet flavours in the single tagging bins and of the fit variable m_{SV} for the light-flavour jet calibration	217
C.4	Results of light-flavour jet SFs	218
C.5	Distribution of the mass of a jet and its associated lepton after the event selection of the b -jet calibration	221
C.6	Selection of the SR and CRs for the b -jet calibration	222
C.7	b -jet tagging efficiencies in MC and data, and SFs for all tagging bins	225
C.8	c -jet tagging efficiencies in MC and data, and SFs for all tagging bins	231
C.9	p_T distribution of all jets split by jet flavour	234
C.10	Maximum p_T of light-flavour, c -, and b -jets in each event	235

List of Tables

2.1	The fermions grouped into left-handed doublets and right-handed singlets and their quantum numbers related to the electroweak unification	9
4.1	Quality criteria placed on track candidates by the ambiguity solver	29
4.2	Electron isolation WPs used throughout this work	32
4.3	Requirements placed on electrons for the $t\bar{t}$ + charm analysis	33
4.4	Muon isolation WPs used throughout this thesis	34
4.5	Requirements placed on muons for the $t\bar{t}$ + charm analysis	35
4.6	String of the single-lepton high-level triggers used for most analyses in this work.	39
5.1	List of all samples generated in the ATLAS framework used in this work	45
5.2	Filters used for additional $t\bar{t}$ production samples	46
7.1	Bin edges in $ \eta $ and E_T used to obtain the PDFs and to define different thresholds on the discriminant	72
8.1	Different classes used for electron candidates throughout the thesis based on truth information of the simulation	76
8.2	Description of the quantities used in the electron identification based on the DNN	78
8.3	Composition of the background in the dataset used to evaluate the performance	85
8.4	Bin edges in $ \eta $ to define different cut values on the discriminant and to optimise the f_X comprising the discriminant	86
8.5	Bin edges in E_T to define different cut values on the discriminant	86
8.6	Additional selection on top of the discriminant for the different WPs	87
10.1	Prescaled single electron triggers used for the background selection	110
11.1	Samples used in the $t\bar{t}$ + charm analysis	120
11.2	Fiducial phase space selection	122
11.3	Efficiencies, rejections, and cut values for the single bins of the bc -Tagger	126
12.1	Summary of the requirements on detector-level objects to pass the preselection	128
12.2	Definition of the fit regions based on the bc -Tagger working points	130

13.1	Overview of the uncertainties associated to the $t\bar{t}$ + light and $t\bar{t} + \geq 1c$ process	139
13.2	Overview of the uncertainties associated to the $t\bar{t} + \geq 1b$ process	140
13.3	Overview of the uncertainties associated to processes other than $t\bar{t}$ + light, $t\bar{t} + \geq 1c$, and $t\bar{t} + \geq 1b$	141
13.4	Complete list of systematics considered before pruning is applied	143
14.1	Fitted free-floating parameters in the fit to the Asimov dataset	148
14.2	Fitted free-floating parameters in the fit to ATLAS data	151
14.3	Groups of NP and their relative impact on the eight extracted norm factors	153
15.1	Definition of the fit regions in the dilepton channel based on the bc -Tagger working points	158
15.2	Fitted free-floating parameters in the combined fit to ATLAS data	160
15.3	Groups of NP and their relative impact on the eight extracted norm factors in the combined fit	161
15.4	Expected and observed number of events in the ten fit regions pre- and post-fit	164
15.5	Measured parameters extrapolated to the full phase space	167
17.1	Event reconstruction efficiencies for the SPA-net model, the TG model, the TT model, and the χ^2 method in different jet and b -jet multiplicities	184
17.2	Percentage of events with at most N incorrectly matched partons using the SPA-net model, the TG model, the TT model, and the χ^2 method	185
17.3	Efficiencies of correctly associating a jet to its corresponding parton for the SPA-net model, the TG model, the TT model, and the χ^2 method	186
17.4	Efficiencies of correctly associating a jet to the corresponding parton for events with at least one incorrect assignment for the SPA-net model, the TG model, the TT model, and the χ^2 method	186
17.5	Difference in reconstruction efficiencies when evaluating the different methods on systematic variations	188
17.6	Event reconstruction efficiencies for the SPA-net model, the TG model, the TT model, and the χ^2 method in different jet and b -jet multiplicities. The ML models are trained including partial events.	188
17.7	Percentage of events with at most N incorrectly matched partons using the SPA-net model, the TG model, the TT model, and the χ^2 method. The ML models are trained including partial events.	189

17.8	Mean and IQR of the normalised error for p_T and the error for η and ϕ , for the regression outputs of the TG and TT model, as well as the reconstructed values using the jet assignment of the same models	194
A.1	Hyperparameters of the point cloud-based models	201
B.1	Bin edges in η and E_T used for the SF calculation	204
C.1	Requirements placed on leptons for the different parts of the bc -Tagger calibration	210
C.2	List of samples generated in the ATLAS framework used in the bc -Tagger calibration	210
C.3	$t\bar{t}$ samples used for MC-MC SFs	211
C.4	Bin edges for the SFs in p_T and $ \eta $	211
C.5	Samples used in the light-flavour jet calibration	213
C.6	Event selection requirements of the light-flavour jet calibration	214
C.7	Absolute uncertainties of the light-flavour jet scale factors	219
C.8	Samples used in the b -jet calibration	220
C.9	Event selection requirements of the b -jet calibration . . .	221
C.10	Absolute uncertainties of the b -jet scale factors	226
C.11	Samples used in the c -jet calibration	227
C.12	Event selection requirements of the c -jet calibration . . .	229
C.13	Absolute uncertainties of the c -jet scale factors	233
D.1	Hyperparameters used for the training of TG and TT . .	237

Acronyms

4FS	four flavour scheme
5FS	five flavour scheme
AFII	ATLFAST-II
ALICE	A Large Ion Collider Experiment
BSM	Beyond the Standard Model
CB	combined
CERN	European Organisation for Nuclear Research
CKM	Cabibbo-Kobayashi-Maskawa
CMS	Compact Muon Solenoid
CNN	Convolutional Neural Network
CR	control region
CSC	Cathode Strip Chamber
CT	calorimeter-tagged
DGLAP	Dokshitzer-Gribov-Lipatov-Altarelli-Parisi
DNN	Feed-Forward Neural Network
DR	diagram removal
DS	diagram subtraction
ECAL	Electromagnetic Calorimeter
FCAL	Forward Calorimeter
FSR	final-state radiation
GELU	Gaussian Error Linear Unit
GNN	Graph Neural Network
GSF	Gaussian Sum Filter
HCAL	Hadronic Calorimeter

HEP	High Energy Physics
HF	heavy-flavour
HLT	High Level Trigger
IBL	Insertable b-Layer
ID	Inner Detector
IN	Interaction Network
IO	inside-out combined
IP	impact parameter
IQR	interquartile range
ISR	initial-state radiation
JER	jet energy resolution
JES	jet energy scale
JVT	Jet Vertex Tagger
KLfitter	Kinematic Likelihood Fitter
kNN	k-Nearest Neighbours
L1	Level-1
LAr	Liquid Argon
Leaky ReLU	Leaky Rectifier Linear Unit
LH	likelihood
LHC	Large Hadron Collider
LINAC2	Linear Accelerator 2
LINAC4	Linear Accelerator 4
LO	leading order
LR	learning rate
MAE	Mean Absolute Error
MC	Monte Carlo
MDT	Monitored Drift Tube
ME	Matrix Element
ML	Machine Learning

MS	Muon Spectrometer
MSE	Mean Squared Error
MuE	muon-spectrometer extrapolated
NLO	next-to-leading order
NN	Neural Network
NNLL	next-to-next-to-leading logarithmic
NP	nuisance parameter
PDF	Parton Distribution Function
PDF	Probability Density Function
PS	Parton Shower
PV	Primary Vertex
QCD	Quantum Chromodynamics
QED	Quantum Electrodynamics
QFT	Quantum Field Theory
ReLU	Rectifier Linear Unit
ROC	receiver operating characteristic
RPC	Resistive Plate Chamber
SCT	Semiconductor Tracker
SF	Scale Factor
SGD	Stochastic Gradient Descent
SM	Standard Model of Particle Physics
SPS	Super Proton Synchrotron
SR	signal region
ST	segment-tagged
TDAQ	Trigger and Data Acquisition
TG	Topograph

TGC	Thin Gap Chamber
TRT	Transition Radiation Tracker
TT	Topotransformer
WP	Working Point

References

- [1] DeepL. Accessed on Sep. 28, 2023. URL: <https://www.deepl.com/translator> (cit. on p. ix).
- [2] ATLAS Collaboration. *Identification of electrons using a deep neural network in the ATLAS experiment*. ATL-PHYS-PUB-2022-022. 2022. URL: <https://cds.cern.ch/record/2803878> (cit. on pp. x, 55, 75, 76, 78, 80, 84–91, 93–95).
- [3] Lukas Ehrke et al. “Topological reconstruction of particle physics processes using graph neural networks”. In: *Phys. Rev. D* 107 (11 June 2023), p. 116019. DOI: [10.1103/PhysRevD.107.116019](https://doi.org/10.1103/PhysRevD.107.116019) (cit. on pp. x, 63, 173, 174, 184–186, 188–192).
- [4] ATLAS Collaboration. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Phys. Lett. B* 716 (2012), p. 1. DOI: [10.1016/j.physletb.2012.08.020](https://doi.org/10.1016/j.physletb.2012.08.020) (cit. on pp. 1, 10, 16).
- [5] CMS Collaboration. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. In: *Phys. Lett. B* 716 (2012), p. 30. DOI: [10.1016/j.physletb.2012.08.021](https://doi.org/10.1016/j.physletb.2012.08.021) (cit. on pp. 1, 10, 16).
- [6] ATLAS Collaboration. “Measurements of inclusive and differential fiducial cross-sections of $t\bar{t}$ production with additional heavy-flavour jets in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *JHEP* 04 (2019), p. 046. DOI: [10.1007/JHEP04\(2019\)046](https://doi.org/10.1007/JHEP04(2019)046) (cit. on pp. 1, 132, 150, 228).
- [7] CMS Collaboration. “Measurement of the $t\bar{t}b\bar{b}$ production cross section in the all-jet final state in pp collisions at $\sqrt{s} = 13$ TeV”. In: *Phys. Lett. B* 803 (2020), p. 135285. DOI: [10.1016/j.physletb.2020.135285](https://doi.org/10.1016/j.physletb.2020.135285) (cit. on p. 1).
- [8] ATLAS Collaboration. “Measurement of Higgs boson decay into b -quarks in associated production with a top-quark pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *JHEP* 06 (2021), p. 097. DOI: [10.1007/JHEP06\(2021\)097](https://doi.org/10.1007/JHEP06(2021)097) (cit. on pp. 1, 55, 119, 123, 132, 150).
- [9] CMS Collaboration. “First measurement of the cross section for top quark pair production with additional charm jets using dileptonic final states in pp collisions at $\sqrt{s} = 13$ TeV”. In: *Phys. Lett. B* 820 (2020), p. 136565. DOI: [10.1016/j.physletb.2021.136565](https://doi.org/10.1016/j.physletb.2021.136565) (cit. on p. 1).
- [10] Mark Thomson. *Modern Particle Physics*. Cambridge University Press, 2013. DOI: [10.1017/CB09781139525367](https://doi.org/10.1017/CB09781139525367) (cit. on p. 5).
- [11] CERN Bulletin. *Go on a particle quest at the first CERN webfest*. Accessed on Feb. 24, 2023. 2012. URL: <http://cds.cern.ch/journal/CERNBulletin/2012/35/News%20Articles/1473657> (cit. on p. 6).

- [12] R. L. Workman et al. “Review of Particle Physics”. In: *PTEP* 2022 (2022), p. 083C01. DOI: [10.1093/ptep/ptac097](https://doi.org/10.1093/ptep/ptac097) (cit. on pp. 6, 7, 12, 42).
- [13] H. David Politzer. “Reliable Perturbative Results for Strong Interactions?”. In: *Phys. Rev. Lett.* 30 (26 June 1973), pp. 1346–1349. DOI: [10.1103/PhysRevLett.30.1346](https://doi.org/10.1103/PhysRevLett.30.1346) (cit. on p. 7).
- [14] David J. Gross and Frank Wilczek. “Ultraviolet Behavior of Non-Abelian Gauge Theories”. In: *Phys. Rev. Lett.* 30 (26 June 1973), pp. 1343–1346. DOI: [10.1103/PhysRevLett.30.1343](https://doi.org/10.1103/PhysRevLett.30.1343) (cit. on p. 7).
- [15] Joshua P. Ellis. “TikZ-Feynman: Feynman diagrams with TikZ”. In: *Computer Physics Communications* 210 (2017), pp. 103–123. ISSN: 0010-4655. DOI: <https://doi.org/10.1016/j.cpc.2016.08.019> (cit. on p. 8).
- [16] Sheldon L. Glashow. “Partial-symmetries of weak interactions”. In: *Nuclear Physics* 22.4 (1961), pp. 579–588. ISSN: 0029-5582. DOI: [https://doi.org/10.1016/0029-5582\(61\)90469-2](https://doi.org/10.1016/0029-5582(61)90469-2) (cit. on p. 8).
- [17] Steven Weinberg. “A Model of Leptons”. In: *Phys. Rev. Lett.* 19 (21 Nov. 1967), pp. 1264–1266. DOI: [10.1103/PhysRevLett.19.1264](https://doi.org/10.1103/PhysRevLett.19.1264) (cit. on p. 8).
- [18] Nicola Cabibbo. “Unitary Symmetry and Leptonic Decays”. In: *Phys. Rev. Lett.* 10 (12 June 1963), pp. 531–533. DOI: [10.1103/PhysRevLett.10.531](https://doi.org/10.1103/PhysRevLett.10.531) (cit. on p. 9).
- [19] Makoto Kobayashi and Toshihide Maskawa. “CP-Violation in the Renormalizable Theory of Weak Interaction”. In: *Progress of Theoretical Physics* 49.2 (Feb. 1973), pp. 652–657. ISSN: 0033-068X. DOI: [10.1143/PTP.49.652](https://doi.org/10.1143/PTP.49.652) (cit. on p. 9).
- [20] P.W. Higgs. “Broken symmetries, massless particles and gauge fields”. In: *Physics Letters* 12.2 (1964), pp. 132–133. ISSN: 0031-9163. DOI: [https://doi.org/10.1016/0031-9163\(64\)91136-9](https://doi.org/10.1016/0031-9163(64)91136-9) (cit. on p. 9).
- [21] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. In: *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323. DOI: [10.1103/PhysRevLett.13.321](https://doi.org/10.1103/PhysRevLett.13.321) (cit. on p. 9).
- [22] *Standard Model Summary Plots February 2022*. Geneva, 2022. URL: <https://cds.cern.ch/record/2804061> (cit. on p. 11).
- [23] “Evidence for Oscillation of Atmospheric Neutrinos”. In: *Phys. Rev. Lett.* 81 (8 Aug. 1998), pp. 1562–1567. DOI: [10.1103/PhysRevLett.81.1562](https://doi.org/10.1103/PhysRevLett.81.1562) (cit. on p. 11).
- [24] “Measurement of the Rate of $\nu_e + d \rightarrow p + p + e^-$ Interactions Produced by ^8B Solar Neutrinos at the Sudbury Neutrino Observatory”. In: *Phys. Rev. Lett.* 87 (7 July 2001), p. 071301. DOI: [10.1103/PhysRevLett.87.071301](https://doi.org/10.1103/PhysRevLett.87.071301) (cit. on p. 11).

- [25] Ziro Maki, Masami Nakagawa, and Shoichi Sakata. “Remarks on the Unified Model of Elementary Particles”. In: *Progress of Theoretical Physics* 28.5 (Nov. 1962), pp. 870–880. ISSN: 0033-068X. DOI: [10.1143/PTP.28.870](https://doi.org/10.1143/PTP.28.870) (cit. on p. 11).
- [26] B. Pontecorvo. “Inverse beta processes and nonconservation of lepton charge”. In: *Zh. Eksp. Teor. Fiz.* 34 (1957), p. 247 (cit. on p. 11).
- [27] “Observation of Top Quark Production in $\bar{p}p$ Collisions with the Collider Detector at Fermilab”. In: *Phys. Rev. Lett.* 74 (14 Apr. 1995), pp. 2626–2631. DOI: [10.1103/PhysRevLett.74.2626](https://doi.org/10.1103/PhysRevLett.74.2626) (cit. on p. 12).
- [28] “Observation of the Top Quark”. In: *Phys. Rev. Lett.* 74 (14 Apr. 1995), pp. 2632–2637. DOI: [10.1103/PhysRevLett.74.2632](https://doi.org/10.1103/PhysRevLett.74.2632) (cit. on p. 12).
- [29] M. Beneke et al. “Hadronic top-quark pair production with NNLL threshold resummation”. In: *Nucl. Phys. B* 855 (2012), pp. 695–741. DOI: [10.1016/j.nuclphysb.2011.10.021](https://doi.org/10.1016/j.nuclphysb.2011.10.021) (cit. on p. 13).
- [30] Matteo Cacciari et al. “Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation”. In: *Phys. Lett. B* 710 (2012), pp. 612–622. DOI: [10.1016/j.physletb.2012.03.013](https://doi.org/10.1016/j.physletb.2012.03.013) (cit. on p. 13).
- [31] Peter Bärnreuther, Michal Czakon, and Alexander Mitov. “Percent-Level-Precision Physics at the Tevatron: Next-to-Next-to-Leading Order QCD Corrections to $q\bar{q} \rightarrow t\bar{t} + X$ ”. In: *Phys. Rev. Lett.* 109 (2012), p. 132001. DOI: [10.1103/PhysRevLett.109.132001](https://doi.org/10.1103/PhysRevLett.109.132001) (cit. on p. 13).
- [32] Michal Czakon and Alexander Mitov. “NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels”. In: *JHEP* 12 (2012), p. 054. DOI: [10.1007/JHEP12\(2012\)054](https://doi.org/10.1007/JHEP12(2012)054) (cit. on p. 13).
- [33] Michal Czakon and Alexander Mitov. “NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction”. In: *JHEP* 01 (2013), p. 080. DOI: [10.1007/JHEP01\(2013\)080](https://doi.org/10.1007/JHEP01(2013)080) (cit. on p. 13).
- [34] Michal Czakon, Paul Fiedler, and Alexander Mitov. “Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_s^4)$ ”. In: *Phys. Rev. Lett.* 110 (2013), p. 252004. DOI: [10.1103/PhysRevLett.110.252004](https://doi.org/10.1103/PhysRevLett.110.252004) (cit. on p. 13).
- [35] Michal Czakon and Alexander Mitov. “Top++: A program for the calculation of the top-pair cross-section at hadron colliders”. In: *Comput. Phys. Commun.* 185 (2014), p. 2930. DOI: [10.1016/j.cpc.2014.06.021](https://doi.org/10.1016/j.cpc.2014.06.021) (cit. on p. 13).
- [36] Jon Butterworth et al. “PDF4LHC recommendations for LHC Run II”. In: *J. Phys. G* 43 (2016), p. 023001. DOI: [10.1088/0954-3899/43/2/023001](https://doi.org/10.1088/0954-3899/43/2/023001) (cit. on pp. 13, 223).

- [37] A. D. Martin et al. “Parton distributions for the LHC”. In: *Eur. Phys. J. C* 63 (2009), p. 189. DOI: [10.1140/epjc/s10052-009-1072-5](https://doi.org/10.1140/epjc/s10052-009-1072-5) (cit. on pp. 13, 42).
- [38] A. D. Martin et al. “Uncertainties on α_S in global PDF analyses and implications for predicted hadronic cross sections”. In: *Eur. Phys. J. C* 64 (2009), pp. 653–680. DOI: [10.1140/epjc/s10052-009-1164-2](https://doi.org/10.1140/epjc/s10052-009-1164-2) (cit. on p. 13).
- [39] H.-L. Lai et al. “New parton distributions for collider physics”. In: *Phys. Rev. D* 82 (2010), p. 074024. DOI: [10.1103/PhysRevD.82.074024](https://doi.org/10.1103/PhysRevD.82.074024) (cit. on pp. 13, 49).
- [40] J. Gao et al. “CT10 next-to-next-to-leading order global analysis of QCD”. In: *Phys. Rev. D* 89 (2014), p. 033009. DOI: [10.1103/PhysRevD.89.033009](https://doi.org/10.1103/PhysRevD.89.033009) (cit. on p. 13).
- [41] *Top cross section summary plots - June 2023*. Geneva, 2023. URL: <https://cds.cern.ch/record/2861926> (cit. on p. 13).
- [42] Geoffrey Gilles. “Top quark production at the LHC”. In: (2018). URL: <https://cds.cern.ch/record/2314658> (cit. on p. 14).
- [43] ATLAS Collaboration. “Measurements of top-quark pair single- and double-differential cross-sections in the all-hadronic channel in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector”. In: *JHEP* 01 (2021), p. 033. DOI: [10.1007/JHEP01\(2021\)033](https://doi.org/10.1007/JHEP01(2021)033) (cit. on pp. 14, 171).
- [44] Lyndon Evans and Philip Bryant. “LHC Machine”. In: *JINST* 3 (2008), S08001. DOI: [10.1088/1748-0221/3/08/S08001](https://doi.org/10.1088/1748-0221/3/08/S08001) (cit. on p. 15).
- [45] Ewa Lopienska. “The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022”. In: (2022). General Photo. URL: <https://cds.cern.ch/record/2800984> (cit. on p. 16).
- [46] ATLAS Collaboration. “The ATLAS Experiment at the CERN Large Hadron Collider”. In: *JINST* 3 (2008), S08003. DOI: [10.1088/1748-0221/3/08/S08003](https://doi.org/10.1088/1748-0221/3/08/S08003) (cit. on pp. 16, 17, 19, 21–23).
- [47] CMS Collaboration. “The CMS Experiment at the CERN LHC”. In: *JINST* 3 (2008), S08004. DOI: [10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004) (cit. on p. 16).
- [48] ALICE Collaboration. “The ALICE experiment at the CERN LHC”. In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08002. DOI: [10.1088/1748-0221/3/08/S08002](https://doi.org/10.1088/1748-0221/3/08/S08002) (cit. on p. 16).
- [49] LHCb Collaboration. “The LHCb Detector at the LHC”. In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08005. DOI: [10.1088/1748-0221/3/08/S08005](https://doi.org/10.1088/1748-0221/3/08/S08005) (cit. on p. 16).

- [50] ATLAS Collaboration. *Luminosity Public Results Run 2*. Accessed on Feb. 24, 2023. URL: <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2> (cit. on pp. 16, 17).
- [51] ATLAS Collaboration. “Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC”. In: (2022). arXiv: 2212.09379 [hep-ex] (cit. on pp. 17, 137).
- [52] Izaak Neutelings. *CMS coordinate system*. Accessed on Mar. 05, 2023. URL: <https://tikz.net/axis3d-cms/> (cit. on p. 18).
- [53] ATLAS Collaboration. *ATLAS Inner Detector: Technical Design Report, Volume 1*. ATLAS-TDR-4; CERN-LHCC-97-016. 1997. URL: <https://cds.cern.ch/record/331063> (cit. on p. 18).
- [54] ATLAS Collaboration. *ATLAS Inner Detector: Technical Design Report, Volume 2*. ATLAS-TDR-5, CERN-LHCC-97-017. 1997. URL: <https://cds.cern.ch/record/331064> (cit. on p. 18).
- [55] ATLAS Collaboration. “Alignment of the ATLAS Inner Detector in Run-2”. In: *Eur. Phys. J. C* 80 (2020), p. 1194. DOI: 10.1140/epjc/s10052-020-08700-6 (cit. on p. 19).
- [56] ATLAS Collaboration. *ATLAS Insertable B-Layer: Technical Design Report*. ATLAS-TDR-19; CERN-LHCC-2010-013. 2010. URL: <https://cds.cern.ch/record/1291633> (cit. on p. 19). Addendum: ATLAS-TDR-19-ADD-1; CERN-LHCC-2012-009. 2012. URL: <https://cds.cern.ch/record/1451888>.
- [57] ATLAS Collaboration. “Performance of the ATLAS Transition Radiation Tracker in Run 1 of the LHC: tracker properties”. In: *JINST* 12 (2017), P05002. DOI: 10.1088/1748-0221/12/05/P05002 (cit. on p. 20).
- [58] ATLAS Collaboration. *Particle Identification Performance of the ATLAS Transition Radiation Tracker*. ATLAS-CONF-2011-128. 2011. URL: <https://cds.cern.ch/record/1383793> (cit. on p. 20).
- [59] ATLAS Collaboration. *ATLAS Liquid Argon Calorimeter: Technical Design Report*. ATLAS-TDR-2; CERN-LHCC-96-041. 1996. URL: <https://cds.cern.ch/record/331061> (cit. on p. 20).
- [60] ATLAS Collaboration. *ATLAS Tile Calorimeter: Technical Design Report*. ATLAS-TDR-3; CERN-LHCC-96-042. 1996. URL: <https://cds.cern.ch/record/331062> (cit. on p. 20).
- [61] ATLAS Collaboration. *ATLAS Muon Spectrometer: Technical Design Report*. ATLAS-TDR-10; CERN-LHCC-97-022. CERN, 1997. URL: <https://cds.cern.ch/record/331068> (cit. on p. 23).
- [62] ATLAS Collaboration. *ATLAS Central Solenoid: Magnet Project Technical Design Report, Volume 4*. ATLAS-TDR-9; CERN-LHCC-97-021. CERN, 1997. URL: <https://cds.cern.ch/record/331067> (cit. on p. 24).

- [63] ATLAS Collaboration. *ATLAS Barrel Toroid: Magnet Project Technical Design Report, Volume 2*. ATLAS-TDR-7; CERN-LHCC-97-019. 1997. URL: <https://cds.cern.ch/record/331065> (cit. on p. 24).
- [64] ATLAS Collaboration. *ATLAS End-Cap Toroids: Magnet Project Technical Design Report, Volume 3*. ATLAS-TDR-8; CERN-LHCC-97-020. CERN, 1997. URL: <https://cds.cern.ch/record/331066> (cit. on p. 24).
- [65] ATLAS Collaboration. “Operation of the ATLAS trigger system in Run 2”. In: *JINST* 15 (2020), P10004. DOI: [10.1088/1748-0221/15/10/P10004](https://doi.org/10.1088/1748-0221/15/10/P10004) (cit. on pp. 24, 25).
- [66] Joao Pequeno and Paul Schaffner. “How ATLAS detects particles: diagram of particle paths in the detector”. Accessed on Mar. 08, 2023. 2013. URL: <https://cds.cern.ch/record/1505342> (cit. on p. 28).
- [67] ATLAS Collaboration. “Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2”. In: *Eur. Phys. J. C* 77 (2017), p. 673. DOI: [10.1140/epjc/s10052-017-5225-7](https://doi.org/10.1140/epjc/s10052-017-5225-7) (cit. on p. 28).
- [68] R. Frühwirth. “Application of Kalman filtering to track and vertex fitting”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 262.2 (1987), pp. 444–450. ISSN: 0168-9002. DOI: [https://doi.org/10.1016/0168-9002\(87\)90887-4](https://doi.org/10.1016/0168-9002(87)90887-4) (cit. on pp. 28, 30).
- [69] ATLAS Collaboration. *Performance of the reconstruction of large impact parameter tracks in the inner detector of ATLAS*. ATLAS-PUB-2017-014. 2017. URL: <https://cds.cern.ch/record/2275635> (cit. on p. 28).
- [70] ATLAS Collaboration. “Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV”. In: *Eur. Phys. J. C* 81 (2021), p. 578. DOI: [10.1140/epjc/s10052-021-09233-2](https://doi.org/10.1140/epjc/s10052-021-09233-2) (cit. on pp. 29, 33, 34).
- [71] ATLAS Collaboration. “Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC”. In: *Eur. Phys. J. C* 77 (2017), p. 332. DOI: [10.1140/epjc/s10052-017-4887-5](https://doi.org/10.1140/epjc/s10052-017-4887-5) (cit. on p. 29).
- [72] ATLAS Collaboration. “Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1”. In: *Eur. Phys. J. C* 77 (2017), p. 490. DOI: [10.1140/epjc/s10052-017-5004-5](https://doi.org/10.1140/epjc/s10052-017-5004-5) (cit. on p. 30).
- [73] ATLAS Collaboration. “Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data”. In: *JINST* 14 (2019), P12006. DOI: [10.1088/1748-0221/14/12/P12006](https://doi.org/10.1088/1748-0221/14/12/P12006) (cit. on pp. 30–32, 71, 78).

- [74] ATLAS Collaboration. *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*. ATLAS-CONF-2012-047. 2012. URL: <https://cds.cern.ch/record/1449796> (cit. on p. 31).
- [75] ATLAS Collaboration. “Studies of the muon momentum calibration and performance of the ATLAS detector with pp collisions at $\sqrt{s} = 13$ TeV”. In: (2022). arXiv: [2212.07338 \[hep-ex\]](https://arxiv.org/abs/2212.07338) (cit. on p. 35).
- [76] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. “The anti- k_t jet clustering algorithm”. In: *JHEP* 04 (2008), p. 063. DOI: [10.1088/1126-6708/2008/04/063](https://doi.org/10.1088/1126-6708/2008/04/063) (cit. on pp. 35, 36).
- [77] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. “FastJet user manual”. In: *Eur. Phys. J. C* 72 (2012), p. 1896. DOI: [10.1140/epjc/s10052-012-1896-2](https://doi.org/10.1140/epjc/s10052-012-1896-2) (cit. on p. 35).
- [78] Gavin P. Salam. “Towards jetography”. In: *The European Physical Journal C* 67.3-4 (May 2010), pp. 637–686. DOI: [10.1140/epjc/s10052-010-1314-6](https://doi.org/10.1140/epjc/s10052-010-1314-6) (cit. on p. 36).
- [79] ATLAS Collaboration. “Jet reconstruction and performance using particle flow with the ATLAS Detector”. In: *Eur. Phys. J. C* 77 (2017), p. 466. DOI: [10.1140/epjc/s10052-017-5031-2](https://doi.org/10.1140/epjc/s10052-017-5031-2) (cit. on p. 36).
- [80] ATLAS Collaboration. “Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *Eur. Phys. J. C* 81 (2020), p. 689. DOI: [10.1140/epjc/s10052-021-09402-3](https://doi.org/10.1140/epjc/s10052-021-09402-3) (cit. on pp. 36, 187).
- [81] ATLAS Collaboration. *Tagging and suppression of pileup jets with the ATLAS detector*. ATLAS-CONF-2014-018. 2014. URL: <https://cds.cern.ch/record/1700870> (cit. on p. 36).
- [82] ATLAS Collaboration. “ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset”. In: (2022). arXiv: [2211.16345 \[physics.data-an\]](https://arxiv.org/abs/2211.16345) (cit. on pp. 37, 55).
- [83] ATLAS Collaboration. *Optimisation and performance studies of the ATLAS b -tagging algorithms for the 2017-18 LHC run*. ATL-PHYS-PUB-2017-013. 2017. URL: <https://cds.cern.ch/record/2273281> (cit. on p. 37).
- [84] ATLAS Collaboration. *Identification of Jets Containing b -Hadrons with Recurrent Neural Networks at the ATLAS Experiment*. ATL-PHYS-PUB-2017-003. 2017. URL: <https://cds.cern.ch/record/2255226> (cit. on p. 37).
- [85] ATLAS Collaboration. *Secondary vertex finding for jet flavour identification with the ATLAS detector*. ATL-PHYS-PUB-2017-011. 2017. URL: <https://cds.cern.ch/record/2270366> (cit. on pp. 37, 213).

- [86] ATLAS Collaboration. *Topological b -hadron decay reconstruction and identification of b -jets with the JetFitter package in the ATLAS experiment at the LHC*. ATL-PHYS-PUB-2018-025. 2018. URL: <https://cds.cern.ch/record/2645405> (cit. on p. 37).
- [87] ATLAS Collaboration. “Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13$ TeV”. In: *Eur. Phys. J. C* 78 (2018), p. 903. DOI: [10.1140/epjc/s10052-018-6288-9](https://doi.org/10.1140/epjc/s10052-018-6288-9) (cit. on p. 38).
- [88] ATLAS Collaboration. “The ATLAS Simulation Infrastructure”. In: *Eur. Phys. J. C* 70 (2010), p. 823. DOI: [10.1140/epjc/s10052-010-1429-9](https://doi.org/10.1140/epjc/s10052-010-1429-9) (cit. on pp. 41, 44).
- [89] John C. Collins, Davison E. Soper, and George Sterman. “FACTORIZATION OF HARD PROCESSES IN QCD”. In: *Perturbative QCD*, pp. 1–91. DOI: [10.1142/9789814503266_0001](https://doi.org/10.1142/9789814503266_0001) (cit. on p. 41).
- [90] Sayipjamal Dulat et al. “New parton distribution functions from a global analysis of quantum chromodynamics”. In: *Phys. Rev. D* 93.3 (2016), p. 033006. DOI: [10.1103/PhysRevD.93.033006](https://doi.org/10.1103/PhysRevD.93.033006) (cit. on p. 42).
- [91] Richard D. Ball et al. “Parton distributions for the LHC run II”. In: *JHEP* 04 (2015), p. 040. DOI: [10.1007/JHEP04\(2015\)040](https://doi.org/10.1007/JHEP04(2015)040) (cit. on pp. 42, 46).
- [92] Yuri L. Dokshitzer. “Calculation of the Structure Functions for Deep Inelastic Scattering and e^+e^- Annihilation by Perturbation Theory in Quantum Chromodynamics.” In: *Sov. Phys. JETP* 46 (1977), pp. 641–653 (cit. on p. 42).
- [93] V. N. Gribov and L. N. Lipatov. “Deep inelastic $e p$ scattering in perturbation theory”. In: *Sov. J. Nucl. Phys.* 15 (1972), pp. 438–450 (cit. on p. 42).
- [94] G. Altarelli and G. Parisi. “Asymptotic freedom in parton language”. In: *Nuclear Physics B* 126.2 (1977), pp. 298–318. ISSN: 0550-3213. DOI: [https://doi.org/10.1016/0550-3213\(77\)90384-4](https://doi.org/10.1016/0550-3213(77)90384-4) (cit. on p. 42).
- [95] S. Bailey et al. “Parton distributions from LHC, HERA, Tevatron and fixed target data: MSHT20 PDFs”. In: *The European Physical Journal C* 81.4 (Apr. 2021). DOI: [10.1140/epjc/s10052-021-09057-0](https://doi.org/10.1140/epjc/s10052-021-09057-0) (cit. on p. 42).
- [96] Stefano Frixione, Giovanni Ridolfi, and Paolo Nason. “A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction”. In: *JHEP* 09 (2007), p. 126. DOI: [10.1088/1126-6708/2007/09/126](https://doi.org/10.1088/1126-6708/2007/09/126) (cit. on p. 43).
- [97] Paolo Nason. “A new method for combining NLO QCD with shower Monte Carlo algorithms”. In: *JHEP* 11 (2004), p. 040. DOI: [10.1088/1126-6708/2004/11/040](https://doi.org/10.1088/1126-6708/2004/11/040) (cit. on p. 43).

- [98] Stefano Frixione, Paolo Nason, and Carlo Oleari. “Matching NLO QCD computations with parton shower simulations: the POWHEG method”. In: *JHEP* 11 (2007), p. 070. doi: [10.1088/1126-6708/2007/11/070](https://doi.org/10.1088/1126-6708/2007/11/070) (cit. on p. 43).
- [99] Simone Alioli et al. “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”. In: *JHEP* 06 (2010), p. 043. doi: [10.1007/JHEP06\(2010\)043](https://doi.org/10.1007/JHEP06(2010)043) (cit. on p. 43).
- [100] Simone Alioli et al. “NLO vector-boson production matched with shower in POWHEG”. In: *JHEP* 07 (2008), p. 060. doi: [10.1088/1126-6708/2008/07/060](https://doi.org/10.1088/1126-6708/2008/07/060) (cit. on p. 43).
- [101] T. Sjöstrand, S. Mrenna, and P. Skands. “A brief introduction to PYTHIA 8.1”. In: *Comput. Phys. Commun.* 178 (2008), pp. 852–867. doi: [10.1016/j.cpc.2008.01.036](https://doi.org/10.1016/j.cpc.2008.01.036) (cit. on p. 43).
- [102] Torbjörn Sjöstrand et al. “An introduction to PYTHIA 8.2”. In: *Comput. Phys. Commun.* 191 (2015), p. 159. doi: [10.1016/j.cpc.2015.01.024](https://doi.org/10.1016/j.cpc.2015.01.024) (cit. on p. 43).
- [103] Enrico Bothmann et al. “Event generation with Sherpa 2.2”. In: *SciPost Phys.* 7.3 (2019), p. 034. doi: [10.21468/SciPostPhys.7.3.034](https://doi.org/10.21468/SciPostPhys.7.3.034) (cit. on p. 43).
- [104] J. Alwall et al. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”. In: *JHEP* 07 (2014), p. 079. doi: [10.1007/JHEP07\(2014\)079](https://doi.org/10.1007/JHEP07(2014)079) (cit. on p. 43).
- [105] D. J. Lange. “The EvtGen particle decay simulation package”. In: *Nucl. Instrum. Meth. A* 462 (2001), p. 152. doi: [10.1016/S0168-9002\(01\)00089-4](https://doi.org/10.1016/S0168-9002(01)00089-4) (cit. on p. 43).
- [106] Stefano Frixione et al. “Angular correlations of lepton pairs from vector boson and top quark decays in Monte Carlo simulations”. In: *JHEP* 04 (2007), p. 081. doi: [10.1088/1126-6708/2007/04/081](https://doi.org/10.1088/1126-6708/2007/04/081) (cit. on p. 43).
- [107] Pierre Artoisenet et al. “Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations”. In: *JHEP* 03 (2013), p. 015. doi: [10.1007/JHEP03\(2013\)015](https://doi.org/10.1007/JHEP03(2013)015) (cit. on p. 43).
- [108] B. Andersson et al. “Parton fragmentation and string dynamics”. In: *Physics Reports* 97.2 (1983), pp. 31–145. issn: 0370-1573. doi: [https://doi.org/10.1016/0370-1573\(83\)90080-7](https://doi.org/10.1016/0370-1573(83)90080-7) (cit. on p. 43).
- [109] B.R. Webber. “A QCD model for jet fragmentation including soft gluon interference”. In: *Nuclear Physics B* 238.3 (1984), pp. 492–528. issn: 0550-3213. doi: [https://doi.org/10.1016/0550-3213\(84\)90333-X](https://doi.org/10.1016/0550-3213(84)90333-X) (cit. on p. 43).
- [110] M. Bähr et al. “Herwig++ physics and manual”. In: *Eur. Phys. J. C* 58 (2008), p. 639. doi: [10.1140/epjc/s10052-008-0798-9](https://doi.org/10.1140/epjc/s10052-008-0798-9) (cit. on p. 44).

- [111] Johannes Bellm et al. “Herwig 7.0/Herwig++ 3.0 release note”. In: *Eur. Phys. J. C* 76.4 (2016), p. 196. DOI: [10.1140/epjc/s10052-016-4018-8](https://doi.org/10.1140/epjc/s10052-016-4018-8) (cit. on pp. 44, 47, 48).
- [112] Johannes Bellm et al. “Herwig 7.1 Release Note”. In: (2017). arXiv: [1705.06919 \[hep-ph\]](https://arxiv.org/abs/1705.06919) (cit. on pp. 44, 47).
- [113] ATLAS Collaboration. *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model*. ATL-PHYS-PUB-2016-017. 2016. URL: <https://cds.cern.ch/record/2206965> (cit. on p. 44).
- [114] Richard D. Ball et al. “Parton distributions with LHC data”. In: *Nucl. Phys. B* 867 (2013), p. 244. DOI: [10.1016/j.nuclphysb.2012.10.003](https://doi.org/10.1016/j.nuclphysb.2012.10.003) (cit. on p. 44).
- [115] A. Ribon et al. “Status of Geant4 hadronic physics for the simulation of LHC experiments at the start of the LHC physics program”. In: (2010). URL: <https://lcgapp.cern.ch/project/docs/noteStatusHadronic2010.pdf> (cit. on p. 44).
- [116] ATLAS Collaboration. *ATLAS Pythia 8 tunes to 7 TeV data*. ATL-PHYS-PUB-2014-021. 2014. URL: <https://cds.cern.ch/record/1966419> (cit. on p. 45).
- [117] ATLAS Collaboration. *Studies on top-quark Monte Carlo modelling for Top2016*. ATL-PHYS-PUB-2016-020. 2016. URL: <https://cds.cern.ch/record/2216168> (cit. on pp. 46, 48).
- [118] L. A. Harland-Lang et al. “Parton distributions in the LHC era: MMHT 2014 PDFs”. In: *Eur. Phys. J. C* 75.5 (2015), p. 204. DOI: [10.1140/epjc/s10052-015-3397-6](https://doi.org/10.1140/epjc/s10052-015-3397-6) (cit. on pp. 47, 48).
- [119] ATLAS Collaboration. *Study of $t\bar{t}b\bar{b}$ and $t\bar{t}W$ background modelling for $t\bar{t}H$ analyses*. ATL-PHYS-PUB-2022-026. 2022. URL: <https://cds.cern.ch/record/2810864> (cit. on p. 47).
- [120] Tomáš Ježo et al. “New NLOPS predictions for $t\bar{t} + b$ -jet production at the LHC”. In: *Eur. Phys. J. C* 78.6 (2018), p. 502. DOI: [10.1140/epjc/s10052-018-5956-0](https://doi.org/10.1140/epjc/s10052-018-5956-0) (cit. on p. 47).
- [121] Tomas Ježo. *Powheg-Box-Res $t\bar{t}b\bar{b}$ source code*. 2019. URL: <https://gitlab.cern.ch/tjezo/powheg-box-res-ttbb/> (cit. on p. 47).
- [122] Tanju Gleisberg and Stefan Höche. “Comix, a new matrix element generator”. In: *JHEP* 12 (2008), p. 039. DOI: [10.1088/1126-6708/2008/12/039](https://doi.org/10.1088/1126-6708/2008/12/039) (cit. on p. 47).
- [123] Federico Buccioni et al. “OpenLoops 2”. In: *Eur. Phys. J. C* 79.10 (2019), p. 866. DOI: [10.1140/epjc/s10052-019-7306-2](https://doi.org/10.1140/epjc/s10052-019-7306-2) (cit. on p. 47).
- [124] Fabio Cascioli, Philipp Maierhöfer, and Stefano Pozzorini. “Scattering Amplitudes with Open Loops”. In: *Phys. Rev. Lett.* 108 (2012), p. 111601. DOI: [10.1103/PhysRevLett.108.111601](https://doi.org/10.1103/PhysRevLett.108.111601) (cit. on p. 47).

- [125] Ansgar Denner, Stefan Dittmaier, and Lars Hofer. “COLLIER: A fortran-based complex one-loop library in extended regularizations”. In: *Comput. Phys. Commun.* 212 (2017), pp. 220–238. doi: [10.1016/j.cpc.2016.10.013](https://doi.org/10.1016/j.cpc.2016.10.013) (cit. on p. 47).
- [126] Stefan Höche et al. “A critical appraisal of NLO+PS matching methods”. In: *JHEP* 09 (2012), p. 049. doi: [10.1007/JHEP09\(2012\)049](https://doi.org/10.1007/JHEP09(2012)049) (cit. on p. 47).
- [127] Stefan Höche et al. “QCD matrix elements + parton showers. The NLO case”. In: *JHEP* 04 (2013), p. 027. doi: [10.1007/JHEP04\(2013\)027](https://doi.org/10.1007/JHEP04(2013)027) (cit. on p. 47).
- [128] S. Catani et al. “QCD Matrix Elements + Parton Showers”. In: *JHEP* 11 (2001), p. 063. doi: [10.1088/1126-6708/2001/11/063](https://doi.org/10.1088/1126-6708/2001/11/063) (cit. on p. 47).
- [129] Stefan Höche et al. “QCD matrix elements and truncated showers”. In: *JHEP* 05 (2009), p. 053. doi: [10.1088/1126-6708/2009/05/053](https://doi.org/10.1088/1126-6708/2009/05/053) (cit. on p. 47).
- [130] Steffen Schumann and Frank Krauss. “A parton shower algorithm based on Catani–Seymour dipole factorisation”. In: *JHEP* 03 (2008), p. 038. doi: [10.1088/1126-6708/2008/03/038](https://doi.org/10.1088/1126-6708/2008/03/038) (cit. on p. 47).
- [131] Emanuele Re. “Single-top Wt -channel production matched with parton showers using the POWHEG method”. In: *Eur. Phys. J. C* 71 (2011), p. 1547. doi: [10.1140/epjc/s10052-011-1547-z](https://doi.org/10.1140/epjc/s10052-011-1547-z) (cit. on p. 48).
- [132] Stefano Frixione et al. “Single-top hadroproduction in association with a W boson”. In: *JHEP* 07 (2008), p. 029. doi: [10.1088/1126-6708/2008/07/029](https://doi.org/10.1088/1126-6708/2008/07/029) (cit. on p. 48).
- [133] Heribertus B. Hartanto et al. “Higgs boson production in association with top quarks in the POWHEG BOX”. In: *Phys. Rev. D* 91.9 (2015), p. 094003. doi: [10.1103/PhysRevD.91.094003](https://doi.org/10.1103/PhysRevD.91.094003) (cit. on p. 49).
- [134] ATLAS Collaboration. “Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector”. In: *JHEP* 09 (2014), p. 145. doi: [10.1007/JHEP09\(2014\)145](https://doi.org/10.1007/JHEP09(2014)145) (cit. on p. 49).
- [135] J. Pumplin et al. “New Generation of Parton Distributions with Uncertainties from Global QCD Analysis”. In: *JHEP* 07 (2002), p. 012. doi: [10.1088/1126-6708/2002/07/012](https://doi.org/10.1088/1126-6708/2002/07/012) (cit. on p. 49).
- [136] Piotr Golonka and Zbigniew Was. “PHOTOS Monte Carlo: a precision tool for QED corrections in Z and W decays”. In: *Eur. Phys. J. C* 45 (2006), pp. 97–107. doi: [10.1140/epjc/s2005-02396-4](https://doi.org/10.1140/epjc/s2005-02396-4) (cit. on p. 50).

- [137] N. Davidson, T. Przedzinski, and Z. Was. “PHOTOS Interface in C++: Technical and physics documentation”. In: *Comput. Phys. Commun.* 199 (2016), pp. 86–101. DOI: [10.1016/j.cpc.2015.09.013](https://doi.org/10.1016/j.cpc.2015.09.013) (cit. on p. 50).
- [138] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press, 2016 (cit. on pp. 53, 54).
- [139] ATLAS Collaboration. “Deep generative models for fast photon shower simulation in ATLAS”. In: (2022). arXiv: [2210.06204 \[hep-ex\]](https://arxiv.org/abs/2210.06204) (cit. on p. 54).
- [140] ATLAS Collaboration. “AtlFast3: The Next Generation Of Fast Simulation in ATLAS”. In: *Comput. Softw. Big Sci.* 6 (2021), p. 7. DOI: [10.1007/s41781-021-00079-7](https://doi.org/10.1007/s41781-021-00079-7) (cit. on p. 54).
- [141] ATLAS Collaboration. *Deep generative models for fast shower simulation in ATLAS*. ATL-SOFT-PUB-2018-001. 2018. URL: <https://cds.cern.ch/record/2630433> (cit. on p. 54).
- [142] ATLAS Collaboration. *Fast simulation of the ATLAS calorimeter system with Generative Adversarial Networks*. ATL-SOFT-PUB-2020-006. 2020. URL: <https://cds.cern.ch/record/2746032> (cit. on p. 54).
- [143] ATLAS Collaboration. “Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton–proton collision data”. In: *JINST* 14 (2019), P03017. DOI: [10.1088/1748-0221/14/03/P03017](https://doi.org/10.1088/1748-0221/14/03/P03017) (cit. on p. 54).
- [144] ATLAS Collaboration. *METNet: A combined p_T^{miss} working point using a neural network with the ATLAS detector*. ATL-PHYS-PUB-2021-025. 2021. URL: <https://cds.cern.ch/record/2776653> (cit. on p. 54).
- [145] Jerzy Neyman and E.S. Pearson. “On the Problem of the Most Efficient Tests of Statistical Hypotheses”. In: *Philosophical Transactions of the Royal Society, A* 231 (Jan. 1933), pp. 289–337. DOI: [10.1098/rsta.1933.0009](https://doi.org/10.1098/rsta.1933.0009) (cit. on p. 55).
- [146] ATLAS Collaboration. *Electron Identification with a Convolutional Neural Network in the ATLAS Experiment*. ATL-PHYS-PUB-2023-001. 2023. URL: <https://cds.cern.ch/record/2850666> (cit. on pp. 55, 61, 98, 100).
- [147] ATLAS Collaboration. *Graph Neural Network Jet Flavour Tagging with the ATLAS Detector*. ATL-PHYS-PUB-2022-027. 2022. URL: <https://cds.cern.ch/record/2811135> (cit. on pp. 55, 63, 198).
- [148] ATLAS Collaboration. *Deep Sets based Neural Networks for Impact Parameter Flavour Tagging in ATLAS*. ATL-PHYS-PUB-2020-014. 2020. URL: <https://cds.cern.ch/record/2718948> (cit. on pp. 55, 62).

- [149] ATLAS Collaboration. “Measurement of the $t\bar{t}t\bar{t}$ production cross section in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *JHEP* 11 (2021), p. 118. DOI: [10.1007/JHEP11\(2021\)118](https://doi.org/10.1007/JHEP11(2021)118) (cit. on p. 55).
- [150] ATLAS Collaboration. “Evidence for $t\bar{t}t\bar{t}$ production in the multilepton final state in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *Eur. Phys. J. C* 80 (2020), p. 1085. DOI: [10.1140/epjc/s10052-020-08509-3](https://doi.org/10.1140/epjc/s10052-020-08509-3) (cit. on p. 55).
- [151] ATLAS Collaboration. “Measurement of single top-quark production in association with a W boson in the single-lepton channel at $\sqrt{s} = 8$ TeV with the ATLAS detector”. In: *Eur. Phys. J. C* 81 (2020), p. 720. DOI: [10.1140/epjc/s10052-021-09371-7](https://doi.org/10.1140/epjc/s10052-021-09371-7) (cit. on p. 55).
- [152] ATLAS Collaboration. “Observation of the associated production of a top quark and a Z boson in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *JHEP* 07 (2020), p. 124. DOI: [10.1007/JHEP07\(2020\)124](https://doi.org/10.1007/JHEP07(2020)124) (cit. on p. 55).
- [153] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. “Multilayer feedforward networks are universal approximators”. In: *Neural Networks* 2.5 (1989), pp. 359–366. ISSN: 0893-6080. DOI: [https://doi.org/10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8) (cit. on p. 56).
- [154] A.L. Maas, A.Y. Hannun, and A.Y. Ng. “Rectifier Nonlinearities Improve Neural Network Acoustic Models”. In: *Proceedings of the International Conference on Machine Learning*. Atlanta, Georgia, 2013. URL: https://ai.stanford.edu/~amaas/papers/relu_hybrid_icml2013_final.pdf (cit. on p. 56).
- [155] Dan Hendrycks and Kevin Gimpel. *Gaussian Error Linear Units (GELUs)*. 2020. arXiv: [1606.08415](https://arxiv.org/abs/1606.08415) [cs.LG] (cit. on p. 56).
- [156] Nitish Srivastava et al. “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”. In: *Journal of Machine Learning Research* 15.56 (2014), pp. 1929–1958. URL: <http://jmlr.org/papers/v15/srivastava14a.html> (cit. on p. 57).
- [157] Sergey Ioffe and Christian Szegedy. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. 2015. arXiv: [1502.03167](https://arxiv.org/abs/1502.03167) [cs.LG] (cit. on p. 58).
- [158] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton. *Layer Normalization*. 2016. arXiv: [1607.06450](https://arxiv.org/abs/1607.06450) [stat.ML] (cit. on p. 58).
- [159] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: [1412.6980](https://arxiv.org/abs/1412.6980) [cs.LG] (cit. on p. 58).
- [160] Ilya Loshchilov and Frank Hutter. *Decoupled Weight Decay Regularization*. 2019. arXiv: [1711.05101](https://arxiv.org/abs/1711.05101) [cs.LG] (cit. on p. 58).

- [161] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. “Learning representations by back-propagating errors.” In: *Nature* 323.6088 (1986), pp. 533–536 (cit. on p. 58).
- [162] ATLAS Collaboration. *Point Cloud Deep Learning Methods for Pion Reconstruction in the ATLAS Experiment*. ATL-PHYS-PUB-2022-040. 2022. URL: <https://cds.cern.ch/record/2825379> (cit. on pp. 61–63).
- [163] Gregor Kasieczka et al. “Deep-learning top taggers or the end of QCD?” In: *Journal of High Energy Physics* 2017.5 (May 2017). DOI: [10.1007/jhep05\(2017\)006](https://doi.org/10.1007/jhep05(2017)006) (cit. on p. 61).
- [164] Gregor Kasieczka et al. “The Machine Learning landscape of top taggers”. In: *SciPost Phys.* 7 (2019), p. 014. DOI: [10.21468/SciPostPhys.7.1.014](https://doi.org/10.21468/SciPostPhys.7.1.014) (cit. on p. 61).
- [165] Manzil Zaheer et al. *Deep Sets*. 2018. arXiv: [1703.06114](https://arxiv.org/abs/1703.06114) [cs.LG] (cit. on p. 61).
- [166] ATLAS Collaboration. *DeXTer: Deep Sets based Neural Networks for Low- p_T $X \rightarrow b\bar{b}$ Identification in ATLAS*. ATL-PHYS-PUB-2022-042. 2022. URL: <https://cds.cern.ch/record/2825434> (cit. on p. 62).
- [167] Peter W. Battaglia et al. “Relational inductive biases, deep learning, and graph networks”. 2018. arXiv: [1806.01261](https://arxiv.org/abs/1806.01261) [cs.LG] (cit. on p. 62).
- [168] Zonghan Wu et al. “A comprehensive survey on graph neural networks”. In: *IEEE transactions on neural networks and learning systems* 32.1 (2020), pp. 4–24 (cit. on p. 62).
- [169] Yuyang Wang, Zijie Li, and Amir Barati Farimani. *Graph Neural Networks for Molecules*. 2023. arXiv: [2209.05582](https://arxiv.org/abs/2209.05582) [cs.LG] (cit. on p. 62).
- [170] Peter W. Battaglia et al. “Interaction Networks for Learning about Objects, Relations and Physics”. In: (Dec. 2016). arXiv: [1612.00222](https://arxiv.org/abs/1612.00222) [cs.AI] (cit. on p. 63).
- [171] Shah Rukh Qasim et al. “Learning representations of irregular particle-detector geometry with distance-weighted graph networks”. In: *The European Physical Journal C* 79.7 (July 2019). DOI: [10.1140/epjc/s10052-019-7113-9](https://doi.org/10.1140/epjc/s10052-019-7113-9) (cit. on p. 63).
- [172] ATLAS Collaboration. “Observation of four-top-quark production in the multilepton final state with the ATLAS detector”. In: (2023). arXiv: [2303.15061](https://arxiv.org/abs/2303.15061) [hep-ex] (cit. on pp. 63, 77).
- [173] Frédéric A. Dreyer and Huilin Qu. “Jet tagging in the Lund plane with graph networks”. In: *JHEP* 03 (2021), p. 052. DOI: [10.1007/JHEP03\(2021\)052](https://doi.org/10.1007/JHEP03(2021)052) (cit. on p. 63).
- [174] Gage DeZoort et al. “Charged Particle Tracking via Edge-Classifying Interaction Networks”. In: *Comput. Softw. Big Sci.* 5.1 (2021), p. 26. DOI: [10.1007/s41781-021-00073-z](https://doi.org/10.1007/s41781-021-00073-z) (cit. on p. 63).

- [175] Ali Hariri, Darya Dyachkova, and Sergei Gleyzer. “Graph Generative Models for Fast Detector Simulations in High Energy Physics”. Apr. 2021. arXiv: [2104.01725 \[hep-ex\]](#) (cit. on p. 63).
- [176] Oliver Atkinson et al. “Anomaly detection with convolutional Graph Neural Networks”. In: *JHEP* 08 (2021), p. 080. DOI: [10.1007/JHEP08\(2021\)080](#) (cit. on p. 63).
- [177] Huilin Qu and Loukas Gouskos. “ParticleNet: Jet Tagging via Particle Clouds”. In: *Phys. Rev. D* 101.5 (2020), p. 056019. DOI: [10.1103/PhysRevD.101.056019](#) (cit. on p. 63).
- [178] “Identification of highly Lorentz-boosted heavy particles using graph neural networks and new mass decorrelation techniques”. CMS-DP-2020-002. 2020. URL: <https://cds.cern.ch/record/2707946> (cit. on p. 63).
- [179] Jannicke Pearkes et al. “Jet Constituents for Deep Neural Network Based Top Quark Tagging”. 2017. arXiv: [1704.02124 \[hep-ex\]](#) (cit. on p. 63).
- [180] Huilin Qu and Loukas Gouskos. “Jet tagging via particle clouds”. In: *Physical Review D* 101.5 (Mar. 2020). ISSN: 2470-0029. DOI: [10.1103/physrevd.101.056019](#) (cit. on p. 63).
- [181] Jonathan Shlomi et al. “Secondary vertex finding in jets with neural networks”. In: *The European Physical Journal C* 81.6 (June 2021). ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-021-09342-y](#) (cit. on p. 63).
- [182] Jonathan Shlomi, Peter Battaglia, and Jean-Roch Vlimant. “Graph neural networks in particle physics”. In: *Machine Learning: Science and Technology* 2.2 (2020), p. 021001 (cit. on p. 63).
- [183] Javier Duarte and Jean-Roch Vlimant. “Graph neural networks for particle tracking and reconstruction”. In: *Artificial intelligence for high energy physics*. World Scientific, 2022, pp. 387–436 (cit. on p. 63).
- [184] Savannah Thais et al. “Graph Neural Networks in Particle Physics: Implementations, Innovations, and Challenges”. In: *2022 Snowmass Summer Study*. Mar. 2022. arXiv: [2203.12852 \[hep-ex\]](#) (cit. on p. 63).
- [185] Eric A. Moreno et al. “JEDI-net: a jet identification algorithm based on interaction networks”. In: *Eur. Phys. J. C* 80.1 (2020), p. 58. DOI: [10.1140/epjc/s10052-020-7608-4](#) (cit. on p. 63).
- [186] Shiqi Gong et al. “An efficient Lorentz equivariant graph neural network for jet tagging”. In: *JHEP* 07 (2022), p. 030. DOI: [10.1007/JHEP07\(2022\)030](#) (cit. on p. 63).
- [187] Elias Bernreuther et al. “Casting a graph net to catch dark showers”. In: *SciPost Phys.* 10.2 (2021), p. 046. DOI: [10.21468/SciPostPhys.10.2.046](#) (cit. on p. 63).

- [188] Xiangyang Ju and Benjamin Nachman. “Supervised Jet Clustering with Graph Neural Networks for Lorentz Boosted Bosons”. In: *Phys. Rev. D* 102.7 (2020), p. 075014. DOI: [10.1103/PhysRevD.102.075014](https://doi.org/10.1103/PhysRevD.102.075014) (cit. on p. 63).
- [189] Jun Guo et al. “Boosted Higgs boson jet reconstruction via a graph neural network”. In: *Phys. Rev. D* 103.11 (2021), p. 116025. DOI: [10.1103/PhysRevD.103.116025](https://doi.org/10.1103/PhysRevD.103.116025) (cit. on p. 63).
- [190] Ashish Vaswani et al. *Attention Is All You Need*. 2017. arXiv: [1706.03762 \[cs.CL\]](https://arxiv.org/abs/1706.03762) (cit. on p. 63).
- [191] Michael James Fenton et al. “Permutationless many-jet event reconstruction with symmetry preserving attention networks”. In: *Phys. Rev. D* 105.11 (2022), p. 112008. DOI: [10.1103/PhysRevD.105.112008](https://doi.org/10.1103/PhysRevD.105.112008) (cit. on pp. 63, 172, 183).
- [192] Alexander Shmakov et al. “SPANet: Generalized permutationless set assignment for particle physics using symmetry preserving attention”. In: *SciPost Phys.* 12.5 (2022), p. 178. DOI: [10.21468/SciPostPhys.12.5.178](https://doi.org/10.21468/SciPostPhys.12.5.178) (cit. on pp. 63, 172, 173).
- [193] Matthew Leigh et al. *PC-JeDi: Diffusion for Particle Cloud Generation in High Energy Physics*. 2023. arXiv: [2303.05376 \[hep-ph\]](https://arxiv.org/abs/2303.05376) (cit. on p. 63).
- [194] Huilin Qu, Congqiao Li, and Sitian Qian. *Particle Transformer for Jet Tagging*. 2022. arXiv: [2202.03772 \[hep-ph\]](https://arxiv.org/abs/2202.03772) (cit. on p. 63).
- [195] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830 (cit. on p. 66).
- [196] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: <https://www.tensorflow.org/> (cit. on p. 67).
- [197] Adam Paszke et al. *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. 2019. arXiv: [1912.01703 \[cs.LG\]](https://arxiv.org/abs/1912.01703) (cit. on p. 67).
- [198] Daniel Hay Guest et al. *lwttn/lwttnn: Release for Athena v21*. Version v2.0. Feb. 2017. DOI: [10.5281/zenodo.290682](https://doi.org/10.5281/zenodo.290682) (cit. on p. 67).
- [199] ONNX Runtime developers. *ONNX Runtime*. <https://onnxruntime.ai/>. 2021 (cit. on p. 67).
- [200] 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”. In: *Eur. Phys. J. C* 79 (2019), p. 639. DOI: [10.1140/epjc/s10052-019-7140-6](https://doi.org/10.1140/epjc/s10052-019-7140-6) (cit. on pp. 71, 77–79, 91).
- [201] Andreas Hoecker et al. *TMVA - Toolkit for Multivariate Data Analysis*. 2007. arXiv: [physics/0703039 \[physics.data-an\]](https://arxiv.org/abs/physics/0703039) (cit. on p. 71).

- [202] ATLAS Collaboration. “Measurement of the charge asymmetry in top-quark pair production in the lepton-plus-jets final state in pp collision data at $\sqrt{s} = 8$ TeV with the ATLAS detector”. In: *Eur. Phys. J. C* 76 (2016), p. 87. DOI: [10.1140/epjc/s10052-016-3910-6](https://doi.org/10.1140/epjc/s10052-016-3910-6) (cit. on p. 77). Erratum: in: *Eur. Phys. J. C* 77 (2017), p. 564. DOI: [10.1140/epjc/s10052-017-5089-x](https://doi.org/10.1140/epjc/s10052-017-5089-x).
- [203] Leslie N. Smith. *Cyclical Learning Rates for Training Neural Networks*. 2017. arXiv: [1506.01186 \[cs.CV\]](https://arxiv.org/abs/1506.01186) (cit. on p. 84).
- [204] ATLAS Collaboration. *The ATLAS Collaboration Software and Firmware*. ATL-SOFT-PUB-2021-001. 2021. URL: <https://cds.cern.ch/record/2767187> (cit. on p. 84).
- [205] Kyle Cranmer et al. *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*. Tech. rep. New York: New York U., 2012. URL: <https://cds.cern.ch/record/1456844> (cit. on pp. 119, 217).
- [206] Lorenzo Moneta et al. “The RooStats Project”. In: *PoS ACAT2010* (2010). Ed. by T. Speer et al., p. 057. DOI: [10.22323/1.093.0057](https://doi.org/10.22323/1.093.0057) (cit. on p. 119).
- [207] Wouter Verkerke and David P. Kirkby. “The RooFit toolkit for data modeling”. In: *eConf C0303241* (2003). Ed. by L. Lyons and Muge Karagoz, MOLT007. arXiv: [physics/0306116](https://arxiv.org/abs/physics/0306116) (cit. on p. 119).
- [208] Rene Brun and Fons Rademakers. “ROOT — An object oriented data analysis framework”. In: *Nucl. Inst. & Meth. in Phys. Res. A* 389.1 (1997), pp. 81–86. ISSN: 0168-9002. DOI: [https://doi.org/10.1016/S0168-9002\(97\)00048-X](https://doi.org/10.1016/S0168-9002(97)00048-X) (cit. on p. 119).
- [209] ATLAS Collaboration. “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”. In: *Eur. Phys. J. C* 75 (2015), p. 349. DOI: [10.1140/epjc/s10052-015-3543-1](https://doi.org/10.1140/epjc/s10052-015-3543-1) (cit. on p. 123).
- [210] *Proceedings of the 1974 CERN School of Computing: Godøysund, Norway 11 - 24 Aug 1974. 3rd CERN School of Computing*. CERN. Geneva: CERN, 1974. DOI: [10.5170/CERN-1974-023](https://doi.org/10.5170/CERN-1974-023) (cit. on p. 136).
- [211] Robert D. Cousins. *Generalization of Chisquare Goodness-of-Fit Test for Binned Data Using Saturated Models, with Application to Histograms*. 2013. URL: https://www.physics.ucla.edu/~cousins/stats/cousins_saturated.pdf (cit. on p. 149).
- [212] ATLAS Collaboration. “Measurements of top-quark pair differential and double-differential cross-sections in the ℓ +jets channel with pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector”. In: *Eur. Phys. J. C* 79 (2019), p. 1028. DOI: [10.1140/epjc/s10052-019-7525-6](https://doi.org/10.1140/epjc/s10052-019-7525-6). [Erratum: *Eur. Phys. J. C* 80 (2020) 1092] (cit. on pp. 150, 171).

- [213] 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”. In: *Eur. Phys. J. C* 82 (2022), p. 95. DOI: [10.1140/epjc/s10052-021-09843-w](https://doi.org/10.1140/epjc/s10052-021-09843-w) (cit. on pp. 171, 227).
- [214] ATLAS Collaboration. “Top-quark mass measurement in the all-hadronic $t\bar{t}$ decay channel at $\sqrt{s} = 8$ TeV with the ATLAS detector”. In: *JHEP* 09 (2017), p. 118. DOI: [10.1007/JHEP09\(2017\)118](https://doi.org/10.1007/JHEP09(2017)118) (cit. on p. 171).
- [215] CMS Collaboration. “Measurement of the top quark mass with lepton+jets final states using pp collisions at $\sqrt{s} = 13$ TeV”. In: *Eur. Phys. J. C* 78 (2018), p. 891. DOI: [10.1140/epjc/s10052-018-6332-9](https://doi.org/10.1140/epjc/s10052-018-6332-9) (cit. on p. 171).
- [216] ATLAS Collaboration. “Measurement of the top quark mass in the $t\bar{t} \rightarrow$ lepton+jets channel from $\sqrt{s} = 8$ TeV ATLAS data and combination with previous results”. In: *Eur. Phys. J. C* 79 (2019), p. 290. DOI: [10.1140/epjc/s10052-019-6757-9](https://doi.org/10.1140/epjc/s10052-019-6757-9) (cit. on p. 171).
- [217] ATLAS Collaboration. “Measurement of the top-quark mass in $t\bar{t} + 1$ -jet events collected with the ATLAS detector in pp collisions at $\sqrt{s} = 8$ TeV”. In: *JHEP* 11 (2019), p. 150. DOI: [10.1007/JHEP11\(2019\)150](https://doi.org/10.1007/JHEP11(2019)150) (cit. on p. 171).
- [218] ATLAS Collaboration. “Evidence for the charge asymmetry in $pp \rightarrow t\bar{t}$ production at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: (2022). arXiv: [2208.12095 \[hep-ex\]](https://arxiv.org/abs/2208.12095) (cit. on p. 171).
- [219] ATLAS Collaboration. “Evidence for single top-quark production in the s -channel in proton–proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector using the Matrix Element Method”. In: *Phys. Lett. B* 756 (2016), p. 228. DOI: [10.1016/j.physletb.2016.03.017](https://doi.org/10.1016/j.physletb.2016.03.017) (cit. on pp. 171, 172).
- [220] CMS Collaboration. “Measurement of spin correlations in $t\bar{t}$ production using the matrix element method in the muon+jets final state in pp collisions at $\sqrt{s} = 8$ TeV”. In: *Phys. Lett. B* 758 (2016), p. 321. DOI: [10.1016/j.physletb.2016.05.005](https://doi.org/10.1016/j.physletb.2016.05.005) (cit. on p. 171).
- [221] CMS Collaboration. “Measurement of differential $t\bar{t}$ production cross sections in the full kinematic range using lepton+jets events from proton–proton collisions at $\sqrt{s} = 13$ TeV”. In: *Phys. Rev. D* 104 (2021), p. 092013. DOI: [10.1103/PhysRevD.104.092013](https://doi.org/10.1103/PhysRevD.104.092013) (cit. on p. 171).
- [222] ATLAS Collaboration. “Measurements of normalized differential cross-sections for $t\bar{t}$ production in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector”. In: *Phys. Rev. D* 90 (2014), p. 072004. DOI: [10.1103/PhysRevD.90.072004](https://doi.org/10.1103/PhysRevD.90.072004) (cit. on p. 171).

- [223] CMS Collaboration. “Measurement of the top quark mass using proton–proton data at $\sqrt{s} = 7$ and 8 TeV”. In: *Phys. Rev. D* 93 (2016), p. 072004. DOI: [10.1103/PhysRevD.93.072004](https://doi.org/10.1103/PhysRevD.93.072004) (cit. on p. 171).
- [224] ATLAS Collaboration. “Determination of the top-quark pole mass using $t\bar{t} + 1$ -jet events collected with the ATLAS experiment in 7 TeV pp collisions”. In: *JHEP* 10 (2015), p. 121. DOI: [10.1007/JHEP10\(2015\)121](https://doi.org/10.1007/JHEP10(2015)121) (cit. on p. 171).
- [225] CMS Collaboration. “Measurement of differential cross sections for top quark pair production using the lepton+jets final state in proton–proton collisions at 13 TeV”. In: *Phys. Rev. D* 95 (2017), p. 092001. DOI: [10.1103/PhysRevD.95.092001](https://doi.org/10.1103/PhysRevD.95.092001) (cit. on p. 171).
- [226] CMS Collaboration. “Measurement of the top quark mass in the all-jets final state at $\sqrt{s} = 13$ TeV and combination with the lepton+jets channel”. In: *Eur. Phys. J. C* 79 (2019), p. 313. DOI: [10.1140/epjc/s10052-019-6788-2](https://doi.org/10.1140/epjc/s10052-019-6788-2) (cit. on p. 171).
- [227] CMS Collaboration. “Measurement of differential cross sections for the production of top quark pairs and of additional jets in lepton+jets events from pp collisions at $\sqrt{s} = 13$ TeV”. In: *Phys. Rev. D* 97 (2018), p. 112003. DOI: [10.1103/PhysRevD.97.112003](https://doi.org/10.1103/PhysRevD.97.112003) (cit. on p. 171).
- [228] Johannes Erdmann et al. “A likelihood-based reconstruction algorithm for top-quark pairs and the KLFitter framework”. In: *Nucl. Instrum. Meth. A* 748 (2014), pp. 18–25. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2014.02.029> (cit. on pp. 171, 228).
- [229] Tomas Dado. “Direct measurement of the top-quark decay width with the ATLAS detector”. Presented 12 Jun 2019. 2019. URL: <https://cds.cern.ch/record/2680836> (cit. on p. 172).
- [230] Dirk Neumann. “Deep Neural Network Extension to Kinematic Likelihood Fitter. Erweiterung des Kinematic Likelihood Fitters durch tiefe neuronale Netze”. Presented 25 Sep 2020. 2020. URL: <https://cds.cern.ch/record/2744706> (cit. on p. 172).
- [231] M. Erdmann, B. Fischer, and M. Rieger. “Jet-parton assignment in $t\bar{t}H$ events using deep learning”. In: *Journal of Instrumentation* 12.08 (Aug. 2017), P08020–P08020. DOI: [10.1088/1748-0221/12/08/p08020](https://doi.org/10.1088/1748-0221/12/08/p08020) (cit. on p. 172).
- [232] J. Erdmann et al. “From the bottom to the top—reconstruction of $t\bar{t}$ events with deep learning”. In: *Journal of Instrumentation* 14.11 (Nov. 2019), P11015–P11015. DOI: [10.1088/1748-0221/14/11/p11015](https://doi.org/10.1088/1748-0221/14/11/p11015) (cit. on p. 172).

- [233] ATLAS Collaboration. “Search for the standard model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *Phys. Rev. D* 97 (2018), p. 072016. DOI: [10.1103/PhysRevD.97.072016](https://doi.org/10.1103/PhysRevD.97.072016) (cit. on p. 172).
- [234] Anthony Badea et al. “Solving combinatorial problems at particle colliders using machine learning”. In: *Physical Review D* 106.1 (July 2022). DOI: [10.1103/physrevd.106.016001](https://doi.org/10.1103/physrevd.106.016001) (cit. on p. 172).
- [235] Anja Butter et al. “Deep-learned Top Tagging with a Lorentz Layer”. In: *SciPost Physics* 5.3 (Sept. 2018). DOI: [10.21468/scipostphys.5.3.028](https://doi.org/10.21468/scipostphys.5.3.028) (cit. on p. 172).
- [236] Frank Fiedler et al. “The matrix element method and its application to measurements of the top quark mass”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 624.1 (Dec. 2010), pp. 203–218. ISSN: 0168-9002. DOI: [10.1016/j.nima.2010.09.024](https://doi.org/10.1016/j.nima.2010.09.024) (cit. on p. 172).
- [237] CMS Collaboration. “Measurement of the production cross sections for a Z boson and one or more b jets in pp collisions at $\sqrt{s} = 7$ TeV”. In: *JHEP* 06 (2014), p. 120. DOI: [10.1007/JHEP06\(2014\)120](https://doi.org/10.1007/JHEP06(2014)120) (cit. on p. 172).
- [238] Jason Sang Hun Lee et al. “Zero-Permutation Jet-Parton Assignment using a Self-Attention Network”. Dec. 2020. arXiv: [2012.03542 \[hep-ex\]](https://arxiv.org/abs/2012.03542) (cit. on p. 172).
- [239] Matthew Leigh et al. “ ν -Flows: Conditional Neutrino Regression”. In: *SciPost Phys.* 14 (2023), p. 159. DOI: [10.21468/SciPostPhys.14.6.159](https://doi.org/10.21468/SciPostPhys.14.6.159) (cit. on p. 174).
- [240] Knut Zoch et al. *Top quark pair production at the LHC, with all hadronic resolved decays for solving event combinatorics*. Version 1.0. Zenodo, Mar. 2023. DOI: [10.5281/zenodo.7737248](https://doi.org/10.5281/zenodo.7737248) (cit. on p. 175).
- [241] J. de Favereau et al. “DELPHES 3, A modular framework for fast simulation of a generic collider experiment”. In: *JHEP* 02 (Feb. 2014), p. 057 (cit. on p. 175).
- [242] ATLAS Collaboration. *Expected performance of the ATLAS b -tagging algorithms in Run-2*. ATL-PHYS-PUB-2015-022. 2015. URL: <https://cds.cern.ch/record/2037697> (cit. on p. 175).
- [243] Peter Skands, Stefano Carrazza, and Juan Rojo. “Tuning PYTHIA 8.1: the Monash 2013 Tune”. In: *Eur. Phys. J. C* 74.8 (2014), p. 3024. DOI: [10.1140/epjc/s10052-014-3024-y](https://doi.org/10.1140/epjc/s10052-014-3024-y) (cit. on p. 175).
- [244] Ilya Loshchilov and Frank Hutter. *SGDR: Stochastic Gradient Descent with Warm Restarts*. 2017. arXiv: [1608.03983 \[cs.LG\]](https://arxiv.org/abs/1608.03983) (cit. on p. 181).

- [245] ATLAS Collaboration. “Electron and photon efficiencies in LHC Run 2 with the ATLAS experiment”. In: (2023). arXiv: [2308.13362 \[hep-ex\]](#) (cit. on p. 203).
- [246] 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 $\sqrt{s} = 13\text{ TeV}$ ”. In: (2023). arXiv: [2301.06319 \[hep-ex\]](#) (cit. on pp. 213, 214).
- [247] ATLAS Collaboration. *Calibration of the performance of b -tagging for c and light-flavour jets in the 2012 ATLAS data*. ATLAS-CONF-2014-046. 2014. URL: <https://cds.cern.ch/record/1741020> (cit. on p. 214).
- [248] D0 Collaboration. “Measurement of the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96\text{ TeV}$ using secondary vertex b tagging”. In: *Phys. Rev. D* 74 (11 Dec. 2006), p. 112004. DOI: [10.1103/PhysRevD.74.112004](#) (cit. on p. 214).
- [249] ATLAS Collaboration. “ATLAS b -jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13\text{ TeV}$ ”. In: *Eur. Phys. J. C* 79 (2019), p. 970. DOI: [10.1140/epjc/s10052-019-7450-8](#) (cit. on p. 218).
- [250] F James. “MINUIT: Function Minimization and Error Analysis Reference Manual”. In: (1998). CERN Program Library Long Writeups. URL: <http://cds.cern.ch/record/2296388> (cit. on p. 223).