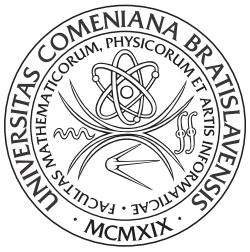
Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

Charge asymmetry in boosted top quark pair production in *pp* collisions at the ATLAS experiment

PhD Thesis

2020 Mgr. Oliver Majerský



Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

Charge asymmetry in boosted top quark pair production in pp collisions at the ATLAS experiment

PhD Thesis

Mgr. Oliver Majerský

Study programme:	Nuclear and Subnuclear Physics
Department:	Department of Nuclear Physics and Biophysics
Supervisor:	Prof. RNDr. Jozef Masarik, DrSc.
Consultant:	Prof. RNDr. Stanislav Tokár, DrSc.

Bratislava, 2020

Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

THESIS ASSIGNMENT

Name and Surname:	Mgr. Oliver Majerský
Study programme:	Nuclear and Subnuclear Physics (Single degree study, Ph.D. III. deg., full time form)
Field of Study:	Physics
Type of Thesis:	Dissertation thesis
Language of Thesis:	English
Secondary language:	Slovak

Title: Charge asymmetry in boosted top quark pair production in pp collisions at the ATLAS experiment

Annotation: This work deals with the top quark pair production. An investigation of differential cross-sections with focus on the charge-asymmetry as a function of different observables is expected, using the simulation and data samples of the ATLAS experiment at the centre-of-mass energy 13 TeV in proton-proton collisions. The top quark is an excellent perturbative object, allowing for theoretical calculations expected to reach up to NNLO (Next-to-Next-to-Leading Order) precision with soft gluon resummation at NNLL (Next-to-Next-to-Leading Logarithm). This allows the measurements of top quark properties to perform strict tests of the Standard Model (SM) as well as searches for the physics beyond the SM. Special focus will also be given to studies of very high-energy top-quark pairs, where dedicated reconstruction techniques are examined, and a study of possibility of charge asymmetry measurement in the all-hadronic decay channel will be investigated.

Aim: To measure the top-quark pair-production charge asymmetry inclusively and differentially as a function of different observables, with the focus on comparisons with the SM predictions. Additionally, to examine the challenges related to measurement of the charge asymmetry in the all-hadronic channel.

Literature:
1. M. Beneke et al., Top Quark Physics, Proc. of the Workshop on Standard Model Physics at LHC, CERN 2000-004
2. The ATLAS collaboration, Expected Performance of the ATLAS Experiment, CERN-OPEN-2008-020, Geneva 2009
3. J.F. Donoghue, E.Golowich and B.R. Holstein, Dynamics of the Standard Model, Cambridge University Press, New York 1992

Comment: computational skill and good knowledge of theory is welcome

Keywords: charge asymmetry, differential cross section, top quark, b-jet, QCD

Tutor:	prof. RNDr. Jozef Masarik, DrSc.
Consultant:	prof. RNDr. Stanislav Tokár, DrSc.
Department:	FMFI.KJFB - Department of Nuclear Physics and Biophysics
Head of department:	prof. RNDr. Stanislav Tokár, DrSc.

Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

Assigned: 24.02.2015

Approved: 24.02.2015 prof. RNDr. Jozef Masarik, DrSc.
Guarantor of Study Programme

.....
Student

.....
Tutor

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

21522464

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Mgr. Oliver Majerský
Študijný program: jadrová a subjadrová fyzika (Jednoodborové štúdium, doktorandské III. st., denná forma)
Študijný odbor: fyzika
Typ záverečnej práce: dizertačná
Jazyk záverečnej práce: anglický
Sekundárny jazyk: slovenský

Názov: Charge asymmetry in boosted top quark pair production in pp collisions at the ATLAS experiment
Nábojová asymetria v produkcii vysoko-energetických top-kvarkových párov v pp-zrážkach na experimente ATLAS

Anotácia: Táto práca sa venuje top-kvarkovej párovej produkcii. Očakáva sa štúdium diferenciálnych účinných prierezov so zameraním na nábojovú asymetriu ako funkciu rôznych kinematických premenných, použitím nasimulovaných a dátových vzoriek experimentu ATLAS pri energii v sústave hmotného stredu 13 TeV v protón-protónových zrážkach. Top kvark je vynikajúci perturbatívny objekt, umožňujúc teoretické výpočty ktoré sa očakávajú až s presnosťou NNLO (Next-to-Next-to-Leading Order) s gluónovými resumáciami v priblížení NNLL (Next-to-Next-to-Leading Logarithm). To umožní precízne porovnať merania vlastností top kvarku s predpoveďami Štandardného modelu (SM) ako aj hľadanie fyziky za SM. Špeciálny dôraz bude kladený štúdiu vysoko-energetických top-kvarkových párov, kde sú preskúmané špecializované techniky pre ich rekonštrukciu a bude preskúmaná možnosť merania nábojovej asymetrie v hadrónových rozpadoch top-kvarkových párov.

Cieľ: Zmerať nábojovú asymetriu v top-kvarkovej párovej produkcii inkluzívne ako aj diferenciálne ako funkciu rôznych kinematických pozorovateľných, so zameraním na porovnania s predpoveďou SM. Zároveň preskúmať hlavné výzvy merania nábojovej asymetrie v hadrónových rozpadoch top-kvarkových párov.

Literatúra:
1. M. Beneke et al., Top Quark Physics, Proc. of the Workshop on Standard Model Physics at LHC, CERN 2000-004
2. The ATLAS collaboration, Expected Performance of the ATLAS Experiment, CERN-OPEN-2008-020, Geneva 2009
3. J.F. Donoghue, E.Golowich and B.R. Holstein, Dynamics of the Standard Model, Cambridge University Press, New York 1992

Poznámka: výpočtové zručnosti a dobrá znalosť teórie sú vítané

Kľúčové slová: nábojová asymetria, diferenciálny účinný prierez, top kvark, b-jet, QCD

Školiteľ: prof. RNDr. Jozef Masarik, DrSc.
Konzultant: prof. RNDr. Stanislav Tokár, DrSc.

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

Katedra: FMFI.KJFB - Katedra jadrovej fyziky a biofyziky
Vedúci katedry: prof. RNDr. Stanislav Tokár, DrSc.

Dátum zadania: 24.02.2015

Dátum schválenia: 24.02.2015 prof. RNDr. Jozef Masarik, DrSc.
garant študijného programu

.....
študent

.....
školiteľ

Abstract

In this thesis, the charge asymmetry in top-quark pair ($t\bar{t}$) production in proton-proton collisions at the Large Hadron Collider at the centre-of-mass energy 13 TeV is studied using ATLAS Monte Carlo (MC) simulation and the Run-II data, with a focus on dedicated techniques for identification of high-energy (*boosted*) hadronically-decaying top quarks (boosted top-quark tagging).

A number of top-quark and W-boson tagging algorithms are investigated in a sample enriched in boosted hadronically-decaying top quarks and W bosons. The sample is obtained from $t\bar{t}$ events with a single isolated electron or muon in the final state (single-lepton channel). The taggers include simple selections on using two observables, up to the most complex algorithms using many observables via machine learning to identify top quarks and W bosons. The measurement of signal efficiency of the tagging algorithms in data collected in 2015–2016 with an integrated luminosity of 36.1 fb^{-1} is compared with simulations. The results show good agreement between the simulation and the data, demonstrating that it is in principle possible to calibrate the signal tagging efficiency of arbitrarily complex tagging algorithms to match the efficiency in data.

The charge asymmetry in the $t\bar{t}$ production is measured in the single-lepton channel, using the full Run-II ATLAS dataset with the integrated luminosity of 139 fb^{-1}. A combination of multiple regions is performed in the measurement, involving a so-called *resolved* topology, where individual particles from the $t\bar{t}$ decay are reconstructed and identified, and the boosted topology, where the hadronically decaying top quark is reconstructed using the aforementioned boosted tagging techniques. The asymmetry is measured both inclusively as well as differentially with respect to the mass and the longitudinal Lorentz boost of the $t\bar{t}$ pair. Corrections for the dilution of the asymmetry due to the limited detector and reconstruction acceptance and response are performed using a Fully-Bayesian unfolding method. For the first time at a hadron collider, the evidence of non-zero charge asymmetry is found with the significance of 4 σ. Both the inclusive and differential charge asymmetry measurements are found to agree with the Standard Model prediction.

Finally, a preliminary estimate of the expected statistical and systematic uncertainties of a charge asymmetry measurement in boosted all-hadronic $t\bar{t}$ events assuming full Run-II is obtained. Neural networks are employed to identify the top quark and top anti-quark electric charge using charged tracks of the $t\bar{t}$ decay products. The MC modelling of the neural network is compared with data in $t\bar{t}$ events in the single-lepton channel. An estimate of systematic uncertainty associated with the observed mismodelling of the neural network discriminant in the single-lepton channel is obtained and propagated into the charge asymmetry estimate in the all-hadronic channel. The obtained estimate suggests that a potential measurement would be statistically-limited. A combination of the boosted all-hadronic and single-lepton channel would lead to a sizeable reduction of the total uncertainty of measurement for $t\bar{t}$ mass above 1.5 TeV. Further options to reduce the uncertainty are discussed.

Abstrakt

V tejto práci je skúmaná nábojová asymetria v top-kvarkovej párovej ($t\bar{t}$) produkcii v protón-protónových zrážkach na Veľkom hadrónovom urýchľovači pri ťažiskovej energii 13 TeV použitím Monte Carlo (MC) simulácií a dát z experimentu ATLAS zozbieraných počas tzv. Run-II obdobia. Práca je zameraná na aplikácie techník identifikácie vysoko-energetických hadrónových rozpadov top kvarkov.

Rôzne algoritmy pre identifikáciu top kvarkov a W bozónov sú študované vo vzorke obohatenej o vysoko-energetické hadrónové rozpady top kvarkov a W bozónov. Vzorka je získaná z $t\bar{t}$ produkcie s izolovaným elektrónom alebo miónom vo finálnom stave (tzv. leptónový kanál). Spomenuté identifikačné algoritmy zahŕňajú jednoduchú selekciu aplikovanú na dve pozorovateľné až po najkomplikovanejšie algoritmy ktoré využívajú veľa pozorovateľných pomocou metód strojového učenia. Signálna účinnosť týchto identifikačných algoritmov je získaná z dátach zozbieraných v rokoch 2015–2016 s integrovanou luminozitou 36.1 fb^{-1} a porovnaná s predpoveďami z MC simulácii. Výsledky poukazujú na dobrú zhodu simulácii s nameranými dátami a demonštrujú že je v princípe možné vykonať kalibráciu simulovanej signálnej účinnosti ľubovoľne komplikovaného identifikačného algoritmu pomocou nameraných dát.

Nábojová asymetria v $t\bar{t}$ produkcii je určená v leptónovom kanáli použitím plnej dátovej vzorky z experimentu ATLAS počas obdobia Run-II s integrovanou luminozitou 139 fb^{-1}. V meraní je využitá kombinácia niekoľkých selekčných regiónov, ktoré využívajú aj nízko-energetické rozpady top-kvarkov, kde sú jednotlivé rozpadové produkty rekonštruované individuálne, ako aj vysoko-energetické rozpady top-kvarkov, kedy sú hadrónové rozpady top kvarkov identifikované pomocou vyššie uvedených techník. Nábojová asymetria je určená inkluzívne ako aj diferenciálne vzhľadom na invariantnú hmotnosť a pozdĺžny lorentzovský boost $t\bar{t}$ páru. V meraní sa aplikujú korekcie na odozvu a limitovanú akceptanciu detektora a rekonštrukcie použitím metódy dekonvolúcie využívajúcej Bayesovskú štatistiku. Toto meranie je prvým na hadrónovom urýchľovači ktoré zmeralo nenulovú nábojovú asymetriu s významnosťou na úrovni 4 σ. Inkluzívne ako aj diferenciálne merania sú kompatibilné s predpoveďou Štandardného modelu.

V poslednej časti práce je určený prvotný odhad očakávanej štatistickej a systematickej neistoty merania nábojovej asymetrie vo vysoko-energetickej $t\bar{t}$ produkcii v hadrónovom kanáli za predpokladu použitia plnej dátovej vzorky Run-II. Identifikácia elektrického náboja top kvarku a top anti-kvarku je určená pomocou neurónových sietí využitím vlastností stôp nabitých častíc z rozpadových produktov $t\bar{t}$ páru. Modelovanie neurónovej siete MC simuláciami je overené na dátach v $t\bar{t}$ produkcii v leptónovom kanáli. Je získaný odhad systematickej neistoty zahŕňajúcej pozorovaný nesúlad medzi simuláciou neurónovej siete a dátami, a táto systematická neistota je propagovaná v odhade neistoty merania nábojovej asymetrie v hadrónovom kanáli. Výsledky tohto odhadu naznačujú, že meranie je limitované štatistickou neistotou. Kombináciou hadrónového a leptónového kanálu by bolo možné značne znížiť neistotu merania pre invariantnú hmotnosť $t\bar{t}$ páru väčšiu ako 1.5 TeV. Ďalšie možnosti redukcie neistoty sú predmetom diskusie v práci.

Acknowledgements

The work encompassed in this thesis would not have not been possible without the many people, with whom I had a chance to work with, or who have helped me or supported me in any way over the course of my PhD studies, to reach that light at the end of the tunnel.

I would like to thank Stanislav Tokár, Tobias Golling and Jozef Masarik for their guidance during my studies and for their supervision of my thesis and also for giving me the opportunity to collaborate with many people in the large community of the ATLAS experiment at CERN. I would also like to thank them for helping me arrange the one-year stay at CERN and the Geneva university, and for the establishment of my cotutelle between the Comenius university and the Geneva university.

Many thanks goes to all the people who I had the chance to work together on the studies of boosted top-quark and W-boson tagging. This really felt initially like a dive in the pool without learning how to swim beforehand. I would like to thank Steven Schramm, Chris Delitzch, Christoph Anders and Sam Meehan for their insight and guidance of this large effort that resulted in the published boosted tagging paper. Also, thanks to Christoph and Sam for a comprehensive introduction into the internet culture of memes. I would like to thank Ece Akilli, Aparajita Dattagupta, Tobias Kupfer and Amal Vaidya, for all the time we spent together working on the boosted tagging studies. You have been a source of inspiration, motivation and quite importantly, an optimistic spirit, something that cannot be emphasized enough during the many stressful periods that a student goes through during their studies.

The milestone that is the public result of the charge asymmetry measurement in the single-lepton channel was only possible thanks to the hard work and collaboration with Pavol Bartoš, Báša Eckerová, Maťo Melo and Clement Helsens. I would also like to thank Tomáš Dado for his many insights when dealing with challenges in the analysis, and for all the discussion we've had, particularly those related to the profile-likelihood statistical analysis.

I would like to thank Ece, Marie Lanfermann and Dan Marley for helping me in my initial steps delving into the area of machine learning and its applications in high-energy physics. The lessons I have learnt from you were very helpful, especially for employing neural networks in the $t\bar{t}$ charge asymmetry study in the all-hadronic channel.

In addition to all the people above, the company of the UNIGE students Juno Nindhito, Joaco Hoya, Marco Valente, Riccardo Poggi and many more made my time spent in Geneva all the more fun, thank you for that. Same gratitude goes to my colleagues from Bratislava that I've already mentioned and also including Róbert Astaloš, Juraj Smieško, Dominik Babál, Michal Dubovský, Sofia Hyrych and others. Thanks for all the lunch discussions on whatever (often controversial) topic, and for all the time we have spent together.

Finally, a big thank you goes to my parents, Pavol and Darina, who have supported me all along from my bachelor studies through my PhD studies, never doubting me in my pursuit of this difficult path and encouraging me to keep pushing forward.

Contents

Introduction

The Standard Model (SM) of elementary particles encapsulates our current knowledge of the interactions of particles of the visible matter in the universe, successfully providing a huge amount of predictions which were experimentally confirmed over the course of its existence of roughly 50 years. Despite being in excellent agreement with experiments, the SM does not provide answers for several key questions on the nature of our universe, such as the explanation of the dark matter and dark energy, the baryon asymmetry, or the missing description of gravity at the quantum level. Therefore, many various experiments in the particle physics seek to find evidence for theories beyond the SM (BSM), which could explain (some of) these phenomena.

The ATLAS experiment is one of the four experiments using proton-proton collisions of the Large Hadron Collider (LHC), the largest and most powerful proton-proton collider ever built. The ATLAS detector is a multi-purpose detector designed for studying various processes to both provide stringent tests of the SM predictions as well as to search for clues of BSM physics.

From the year 2015 until 2018, the LHC operated at the unprecedented centre-of-mass energy $\sqrt{s} = 13\,\text{TeV}$ at a hadron collider. During this period, known as the Run-II, the ATLAS experiment has collected pp collision data of the approximate integrated luminosity of $140\,\text{fb}^{-1}$, surpassing the Run-I period of data taking in 2011-2012 by roughly a factor of seven. Combined with the increase of the collision energy from 8 TeV to 13 TeV, the LHC has allowed the ATLAS experiment to explore even more rare processes, and more extreme regions of phase space.

Many of the BSM theories predict the existence of massive particles decaying into top quarks or $W/Z/H$ bosons. The top quarks and the bosons are unstable particles further decaying into lighter particles. To study these decays, the particles must be reconstructed by the ATLAS experiment to infer the properties of their decaying parent particles. In the momenta range of hundreds of GeV and more, the massive top quarks and the bosons become boosted, and thus their decay products highly collimated. Dedicated reconstruction techniques are necessary to identify these decays, referred to as *boosted* tagging. One of the topics of this thesis is the investigation of boosted tagging of top-quarks and W bosons, specifically the performance and accuracy of Monte Carlo simulations of various algorithms designed to identify these particles.

The large ATLAS Run-II dataset allows to measure very rare SM phenomena, one of which is the so-called charge asymmetry in the top-quark pair ($t\bar{t}$) production, describing a phenomenon in the production of the $t\bar{t}$ pairs, which leads to angular asymmetry of the distributions of top-quark and top anti-quark production with respect to the initial-state particles producing the collision. A measurement of the $t\bar{t}$ charge asymmetry at the $\sqrt{s} = 13\,\text{TeV}$ is performed in this thesis, with dedicated studies of highly-boosted top quarks, employing the boosted tagging techniques. Traditionally, this measurement is performed in the single-lepton channel, where one of the top-quark decays includes an electron or a muon. This is because the measurement of the charge asymmetry requires distinguishing top quark from top anti-quark, and the electric charge of the electron or muon can be used to infer the electric charge of the decaying top (anti-)quark with very high precision.

A possibility of measuring $t\bar{t}$ charge asymmetry in the all-hadronic channel is investigated, which is the decay channel where both top-quarks decay hadronically, i.e. decay into quarks, thus having no prompt electrons or muons from the top-quark decay. In this study, alternative approaches to reconstructing the charge of the decaying top-quark are investigated, using the electric charge of the decay products of the hadronic top quark.

This thesis is organised as follows. In Chapter 1, the fundamentals of the Standard Model of elementary particles are introduced. In Chapter 2, the top quark pair-production and decay is discussed as well as the theoretical origin of the charge asymmetry and an overview of its previous measurements. In Chapters 3 and 4, the overview of the LHC and the ATLAS experiment is shown and the physics objects used in the thesis are defined. Chapter 5 introduces the concepts behind the identification of the boosted hadronically-decaying top-quarks and W bosons at the ATLAS experiment and the recent developments including the applications of machine learning. In Chapter 6, the measurement of signal efficiency of these identification algorithms using data is presented. The measurement of the charge asymmetry in the top-quark pair production in the single-lepton channel is presented in Chapter 7. Finally, Chapter 8 explores the possibility of a measurement of the charge asymmetry in boosted all-hadronic $t\bar{t}$ production.

Personal contributions of the author

Due to the complexity and the nature of cooperation within the ATLAS collaboration, it is often unclear what are the individual contributions of an author. To help the reader understand the individual contributions of the author of this thesis, this section outlines the contributions in a list with references to the respective chapters. In addition, those figures showing results which were already published, which the thesis author has not significantly contributed to, include a reference to the publication in which they appear. The contributions of the author in this thesis include:

- Performance comparisons and the measurement of signal efficiency of the boosted top-quark and W-boson identification techniques (Chapter 5 and 6): The results shown in these chapters are published in Ref. [1]. The author has been one of the two main analysers of the signal efficiency measurement. He has performed the comparisons of data and MC predictions, and contributed to the construction of the signal efficiency template fit, preparing the templates and performing the propagation of the systematic uncertainties in the signal efficiency measurement. In addition, the author has contributed to the paper writing and editing process and worked as the analysis team coordinator.
- Measurement of the $t\bar{t}$ charge asymmetry in the single-lepton channel (Chapter 7): The author has been one of the four contributing analysers to this measurement. His most notable contributions include the development of the Fully-Bayesian unfolding implementation used in the analysis, including the combination of resolved and boosted topologies, implementation of a method to asses the impact of systematic uncertainties (systematics ranking), and the assessment of uncertainties related to the unfolding method. Finally, the author has worked on the propagation of the systematic uncertainties into the measurement and the unfolding. The results of this measurement are published in Ref. [2].
- Study of top-quark charge identification and the estimate of charge asymmetry measurement sensitivity in the all-hadronic channel (Chapter 8): The author of the thesis is the sole analyser of this study, who performed the study of the neural network for the top-quark charge identification, the study of modelling of the predictions in data in the single-lepton $t\bar{t}$ channel, as well as the estimate of the charge asymmetry measurement precision in the all-hadronic channel.

1

The Standard Model of elementary particles

The Standard Model (SM) of elementary particles is a gauge-invariant relativistic quantum field theory (QFT) describing the known elementary particles and their interactions. The elementary particles in the SM can be divided into fundamental fermions, which are the particles of matter, and bosons, which are interaction mediators. They are further described in the next section and shown in Fig. 1.1.

The interactions include electromagnetic and weak interactions, unified into a single electroweak (EW) interaction, and the strong interaction. These are described in Sec. 1.3 and 1.5, respectively. The interactions are generated via a gauge-invariance principle, described in Sec. 1.2. Additionally, a mechanism of spontaneous symmetry breaking, described in Sec. 1.4, is used to generate the masses of the elementary particles in a manner that obeys the gauge symmetries. This mechanism predicts the existence of the Higgs boson.

Finally, a brief summary of some of the most important limitations of the SM are presented in Sec. 1.6.

1.1 Fundamental particles in Standard Model

The fundamental particles of matter in the SM are fermions, spin one-half particles. Firstly, these include three generations of charged leptons; electron, muon and tau. Each of these leptons has a charge of -1[(1)]. To each of the charged leptons corresponds a neutral lepton; electron, muon or tau neutrino. The charged leptons interact via electromagnetically and weakly. The neutrinos only interact weakly.

Additional fundamental fermions include quarks, which interact via strong interaction, as well as electromagnetic and weak interactions. As will be mention further, the nature of the strong interaction prohibits the existence of free quarks, only bound states commonly referred to as hadrons are observed. More specifically, mesons consisting of a pair of quark and anti-quark or baryons consisting of three quarks[(2)]. Similarly to leptons, three generations of quarks have been discovered so far, the first generation comprising of up and down quarks, the second generation of charm and strange quark and finally, the third generation of top and bottom quark. The up, charm and top quarks have an electric charge of $+2/3e$ and down, strange and bottom quarks have an electric charge of $-1/3e$. In contrast to the charged leptons, quarks carry additional quantum property commonly-referred to as colour or colour charge, a property of the strong interactions. All fundamental fermions in the SM are

[(1)] We use natural units where $\hbar = c = 1$. The electric charges are expressed as multiples of the elementary electric charge e.

[(2)] There are also experiments searching for other bound states of quarks, such as tetraquarks or pentaquarks.

assumed massive with the exception of neutrinos which are assumed to be massless[(3)]. The subsequent generations of leptons and quarks are heavier than the corresponding previous-generation particles.

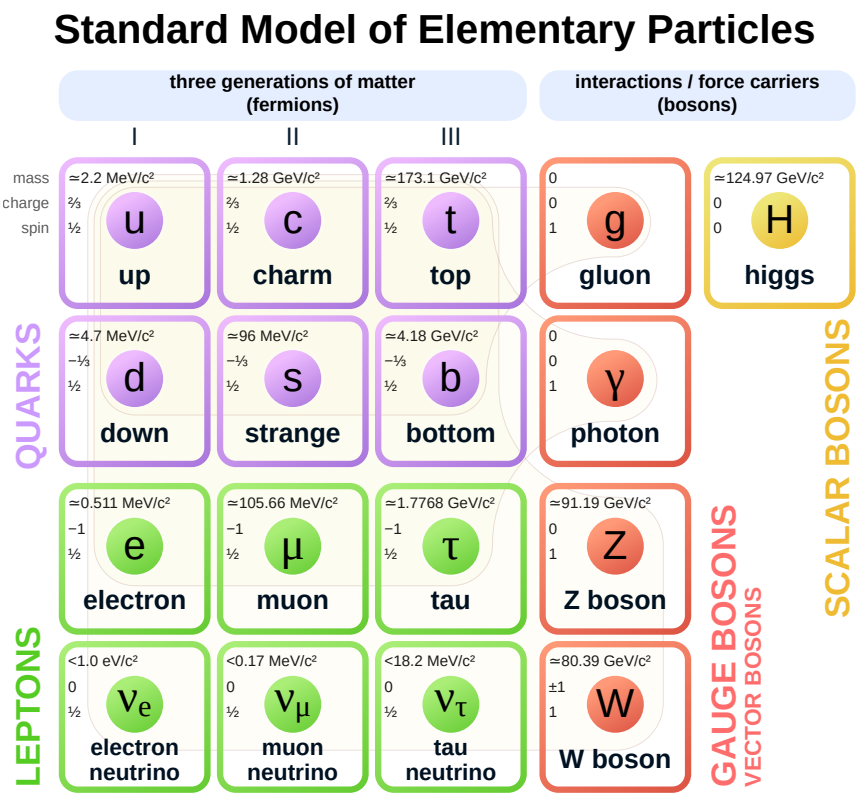


Fig. 1.1: The particles of the Standard Model [3]. The approximate mass, the electric charge, and the spin of the particles is shown. The approximate experimental upper limits on neutrino masses are shown, despite the SM assuming zero neutrino masses.

The interactions in the SM are mediated by spin-one gauge bosons. The mediator of the electromagnetic interaction is the massless photon. The mediators of the weak interactions are the massive $W^{\pm}$ and the massive Z^0 boson. The strong interaction is mediated by eight massless gluons. Finally, the SM also includes one additional particle, the massive spin-zero Higgs boson, whose existence is a manifestation of a mechanism of the spontaneous symmetry breaking. This mechanism generates masses of the $W^{\pm}$ and Z^0 bosons as well as the fermion masses in a manner that does not break the gauge invariance of the SM. We will discuss this mechanism in more detail in Sec. 1.4.

1.2 Local gauge invariance principle

The principle of local gauge invariance in the SM is used as a form of a recipe that allows to create interactions in the QFT framework, that are based on some symmetry described by a representation of a Lie group. The quantum electrodynamics (QED) was arguably the first theory that applied the gauge-invariance principle. It is based on the gauge calibrations of the classical electrodynamics in

[(3)]The SM predates the experimental evidence of neutrino flavour oscillations, which provide evidence of non-zero neutrino masses. In this chapter we simply refer to the SM as it was originally formulated Minimal SM extensions with massive neutrinos are briefly discussed in Sec 1.6.

the field formalism and a quantization of the classical fields to describe electromagnetic interactions between elementary particles.

The theories of particles and their interactions in high-energy physics are commonly formulated using the Lagrangian formalism, more specifically via the Lagrangian density in the classical field theory formalism. In this formalism, the Lagrangian density for spin one-half non-interacting particles is:

$$\mathcal{L} = \bar{\Psi}\left(i\partial\!\!\!/ - m\right)\Psi, \tag{1.1}$$

where Ψ is the Dirac bi-spinor representing the free field of spin one-half particles, m is the mass of the particle, $\bar{\Psi} = \Psi^{\dagger}\gamma_0$ [4], $\partial\!\!\!/ = \gamma^{\mu}\partial_{\mu}$ [5] and γ^{μ} are the Dirac gamma matrices. The bi-spinor $\Psi \equiv \Psi(x)$ is also a function of space-time coordinate in the Lagrangian formalism.

The Lagrangian in Eq. 1.1 is trivially invariant under a transformation:

$$\Psi(x) \mapsto e^{-i\alpha}\Psi(x), \tag{1.2}$$

where α is a fixed real value, independent of x. However, if $\alpha \equiv \alpha(x)$ is a function of space-time coordinate, $\mathcal{L}$ will no longer be invariant under such local gauge transformation due to extra terms stemming from the $\partial_{\mu}\alpha(x)$ derivative. It is possible to fix $\mathcal{L}$ to be invariant by adding extra terms, transforming the ∂_{μ} derivative into a gauge-covariant derivative D_{μ}:

$$\partial_{\mu} \mapsto D_{\mu} = \partial_{\mu} + iqA_{\mu}(x), \tag{1.3}$$

$$A_{\mu}(x) \mapsto A_{\mu}(x) - \frac{1}{q}\partial_{\mu}\alpha(x), \tag{1.4}$$

where the newly introduced field A_{μ} transforms according to Eq. 1.4. At this point one can notice, that the A_{μ} field transforms exactly like the four-potential in the Lorentz-covariant formulation of the classical electrodynamics. Indeed, for electrodynamics, the meaning of q in Eq. 1.3 is that of the electric charge, and in general for any interaction it is related to the coupling constant[6], expressing the relative strength of the interaction. An additional term needs to be added into the Lagrangian to allow for the free propagation of the interaction mediating bosons $-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$, where $F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu}$ is the field-strength tensor. In terms of electromagnetic interactions, this term represents the free photon field. It should be pointed out that the local gauge transformation in electrodynamics is a representation of the $U(1)$ group[7]. The full $U(1)$ gauge-invariant Lagrangian is:

$$\mathcal{L} = \bar{\Psi}\left(iD\!\!\!/ - m\right)\Psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{1.5}$$

Up to this point, we have been talking about building a relativistic gauge-invariant *classical* field theory. A quantization procedure is applied to this theory to create a QFT from the classical field theory. There are multiple ways how to perform the quantization, such as for example the generalisation of canonical quantization from non-relativistic quantum mechanics to field theory, or the path-integral

[4] For multi-component fields, the Hermitian conjugation $\Psi^{\dagger}$ is a combination of transposition and complex conjugation, i.e. $\Psi^{\dagger} = (\Psi^{\star})^{T}$.

[5] The partial derivative in Eq. 1.1 acts on both $\bar{\Psi}$ and Ψ, i.e. $\bar{\Psi}\partial\!\!\!/\Psi \equiv (\partial\!\!\!/\bar{\Psi})\Psi + \bar{\Psi}\partial\!\!\!/\Psi$.

[6] The relation between coupling constant α and q in natural units is ($\hbar = c = 1$) is $q = \sqrt{4\pi\alpha}$.

[7] The $U(n)$ group is a group of $n \times n$ unitary matrices. The $n = 1$ case is thus the group of all complex numbers with absolute value of one.

approach.

In the SM, the gauge invariance principle is also applied using more complicated symmetries than $U(1)$. In principle it is possible to create interactions with different representations of different Lie groups. In the context of SM, the representations of the non-Abelian $SU(2)$ and $SU(3)$ groups[(8)] are used to describe the weak and strong interactions, respectively. A recipe how to make the Lagrangian invariant under transformations of non-Abelian groups was determined by Yang and Mills [4]. The creation of a Lagrangian invariant under the $SU(N)$ local gauge transformations is achieved as follows. Firstly, the covariant derivative that ensures the local gauge invariance of $\mathcal{L}$ gains the form of:

$$\partial_\mu \mapsto D_\mu = \partial_\mu + igA^a_\mu T^a, \tag{1.6}$$

where A^a_μ are the interacting fields and T^a the generators of the representation of the $SU(N)$ group, where $a = 1 \ldots (N^2 - 1)$. Each interacting field corresponds to a new gauge boson. The field-strength tensor for non-Abelian gauge theories is also more complex:

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - g f^{abc} A^b_\mu A^c_\nu, \tag{1.7}$$

where the f^{abc} are the structure constants of the representation of the group, defined by the commutation relations of the generators: $[T^a, T^b] = f^{abc} T^c$ [(9)]. The defining characteristic of non-Abelian groups is that $f^{abc} \neq 0$. Consequently, the free-field term $F^a_{\mu\nu} F^{\mu\nu}_a$ now contains third- and fourth-power gauge field terms. These terms correspond to self-interactions of the gauge bosons, a phenomenon not present in electrodynamics.

The full SM Lagrangian is invariant under a combined $SU(3)_\mathrm{C} \otimes SU(2)_\mathrm{L} \otimes U(1)_\mathrm{Y}$ local gauge symmetry. The $SU(3)_\mathrm{C}$ stands for the colour symmetry of the Quantum Chromodynamics (QCD), the theory of strong interactions. The $SU(2)_\mathrm{L} \otimes U(1)_\mathrm{Y}$ is the symmetry of the unified EW theory. The requirement of the local gauge invariance gives rise to the corresponding interactions terms in the Lagrangian. We will further discuss these symmetries and the interactions in the following sections.

1.3 Electroweak interaction

The electromagnetic and weak interactions are unified in the SM into a single theory, foundations of which were formulated by Glashow [5], Weinberg [6] and Salam [7]. The idea of the *GWS* theory is that both weak and electromagnetic interactions are manifestations of a single underlying interaction. Let us first begin with how the weak interactions are embedded in the SM. Experimental evidence, such as the experiment by Wu, studying β decays of radioactive ^{60}Co isotope [8], showed that parity was violated in weak interactions. In this experiment, the electrons in the β decays were preferentially emitted in a direction opposite to the spin projection of the nuclei. The emitted electron spin projection was preferentially opposite to its momentum direction[(10)]. A hypothesis to explain the parity violation was formulated which suggested that the electron and neutrino in the decay were coupled via a *vector-axial*

[(8)] The $SU(n)$ is the group of unitary matrices of dimension $n \times n$ with determinant equal to one. The generators of representations of this group are $n \times n$ matrices that are anti-hermitian and with a trace equal to zero. The generators are the crucial piece to the formulation of a corresponding gauge-invariant Lagrangian and the induced interaction.

[(9)] $[T^a, T^b] = T^a T^b - T^b T^a$ is the commutator of two matrices.

[(10)] The projection of spin into momentum is commonly referred to as *helicity*.

(V-A) coupling [9]:

$$\bar{e}\gamma^{\mu}(\mathbb{1}-\gamma^{5})\nu_{e}, \tag{1.8}$$

with e and ν_e being the Dirac bi-spinors for electron and neutrino. This coupling can be rewritten into $\bar{e}_L\gamma^{\mu}\nu_{e,L}$, where e_L and $\nu_{e,L}$ are the left-component chirality projections of the electron and neutrino bi-spinor. The chirality states are defined by the projection operator $(\mathbb{1}-\gamma^5)$[11], i.e. the chirality projections[12] of a bi-spinor are defined as:

$$\Psi_L \equiv (\mathbb{1}-\gamma^5)\Psi, \qquad \Psi_R \equiv (\mathbb{1}+\gamma^5)\Psi. \tag{1.9}$$

From this point, chirality projections will be referred to as left-handed and right-handed components.

The V-A theory did explain the parity violation, however it suffered from the fact it was not a renormalisable theory. In the SM, the weak interaction is described by a renormalisable gauge theory based on SU(2) group, that also includes the vector-axial coupling of the fermions, however the left-handed fermion fields couple to gauge boson fields. In a low-momentum-transfer limit, such as in the β decays, this gauge theory yields the original V-A theory.

The coupling of left-handed weakly-interacting particles is reflected in the SM by the fact that the left-handed components of fermion fields are grouped in SU(2) doublets, while right-handed components are SU(2) singlets. For the leptons, the doublets and singlets are as follows:

$$\begin{pmatrix}\ell\\ \nu_\ell\end{pmatrix}_L, \qquad (\ell)_R, \qquad (\nu_\ell)_R, \tag{1.10}$$

where ℓ is one of the e,μ,τ bi-spinors and ν_ℓ is the neutrino bi-spinor from the corresponding lepton generation.

For quarks, the situation is more complicated due to the fact that eigenstates of the weak interaction are not mass eigenstates, leading to flavour violation in charged weak interactions. The relation between the weak eigenstates and the mass eigenstates is given by the *Cabibbo-Kobayashi-Maskawa* (CKM) matrix, a complex unitary 3×3 matrix:

$$\begin{pmatrix}d'\\ s'\\ b'\end{pmatrix} = \begin{pmatrix}V_{ud} & V_{us} & V_{ub}\\ V_{cd} & V_{cs} & V_{cb}\\ V_{td} & V_{ts} & V_{tb}\end{pmatrix}\begin{pmatrix}d\\ s\\ b\end{pmatrix}, \tag{1.11}$$

where the d', s' and b' are the weak eigenstates and the d, s, b are the mass eigenstates. The probability of a quark i transitioning to a quark j is proportional to $|V_{ij}|^2$. The matrix has four physical parameters[13], a common choice of parametrisation is via three mixing angles and one CP-violating complex phase. In the SM, the elements of the CKM matrix are free parameters and are measured experimentally. An interesting fact about the CKM matrix is that it was proposed before either of the third-generation quarks were discovered. This is because already in 1964, Cronin and Fitch observed violation of

[11] $\gamma^5 = i\gamma^1\gamma^2\gamma^3\gamma^4$, where γ^i are the Dirac gamma matrices. The $\mathbb{1}$ operator denotes the 4×4 unit matrix.

[12] A left-handed helicity projection of state in general contains non-zero contributions of both left-handed and right-handed chirality states. For massless particles and in ultra-relativistic limit, the helicity and chirality states coincide.

[13] A $N\times N$ unitary matrix does have N^2 real degrees of freedom. However a complex phase can be absorbed by each of the six quark fields, except for one overall phase. This yields $N^2-(2N-1)$ physical parameters.

combined charge-conjugation and parity (CP) violation in Kaon decays [10]. The quark mixing matrix is a natural construct to include such CP violation, however a 2-generation mixing 2×2 unitary matrix leads to only one real-value mixing angle, which cannot induce CP violation, hence why CKM matrix assumes at least three quark generations.

Going back to the coupling of the fermions in the EW sector, the up- and down-type quark bi-spinors are organised in left-handed doublets and right-handed singlets as follows:

$$\begin{pmatrix} u \\ d' \end{pmatrix}_L, \qquad (u)_R, \qquad (d)_R, \tag{1.12}$$

where the left-handed doublet mixes the quark fields according to the CKM matrix. No weak mixing is introduced in the right-handed singlets since no weak interaction is involved. Finally, the kinetic term of the Lagrangian is:

$$\mathcal{L} = \bar{\Psi}_L i \not{D}_L \bar{\Psi}_L + \bar{\Psi}_R i \not{D}_R \bar{\Psi}_R, \tag{1.13}$$

where Ψ_L (Ψ_R) are the left-handed doublets (right-handed singlets) from Eq. 1.10 and 1.12, for each of the three generations of quarks and leptons. The gauge-covariant derivative D_μ must be different for left-handed and for right-handed components, since the right-handed components do not interact via charged weak currents. At the same time, both left-handed and right-handed components of charged leptons and quarks must transform according to a non-trivial $U(1)$ representation since these particles do interact electromagnetically. The $SU(2)$ and $U(1)$ gauge symmetries give rise to the following covariant derivative for left-handed states:

$$D_\mu = \partial_\mu + i\frac{g}{2}\sigma_i W_\mu^i + i\frac{g'}{2} Y_W B_\mu, \tag{1.14}$$

where σ_i are Pauli's matrices, the representation of $SU(2)$ group, and W_μ^i are the corresponding gauge bosons. The $U(1)$ interaction is inserted via the last term, where B_μ is the gauge boson of this group, and an implicit 2x2 identity matrix is present. The $Y_W = 2(Q - I_3)$ is the weak hyper-charge defined via the charge of the particle Q and third component of the iso-spin I_3, which is $\pm\frac{1}{2}$ for left-handed doublets and zero for right-handed singlets. For right-handed components states, we expect trivial representation [(14)] for $SU(2)$, in other words, the W_μ^i fields should not enter the covariant derivative, and only a term with B_μ should be present. The immediate problem faced is that neither left-handed nor right-handed neutrino components should interact electromagnetically. For right-handed neutrinos, this is achieved trivially thanks to $Y_W = 0$. However, for left-handed neutrinos, this is not true and a coupling to B_μ is present. The solution in the GWS theory is that the fields describing real bosons of the weak and electromagnetic interactions are linear combinations of the W_μ^i and B_μ fields:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp W_\mu^2), \tag{1.15}$$

$$A_\mu = W_\mu^3 \sin\theta_W + B_\mu \cos\theta_W, \tag{1.16}$$

$$Z_\mu^0 = W_\mu^3 \cos\theta_W - B_\mu \sin\theta_W. \tag{1.17}$$

[(14)] A trivial representation of a group maps all group elements to unity. The generators of a trivial representation are zero.

The charged flavour-changing weak currents are mediated by the $W^{\pm}$ bosons and the neutral weak current and the electromagnetic current is mediated by the Z_μ and A_μ bosons, respectively. Finally, the coupling constants g and g' are not independent in the GWS theory, but are bound together via the weak mixing angle θ_W that also enters the mixing of fields of W^3 and B – expressing that the electromagnetic and weak interactions are manifestations of a single interaction with a single coupling constant:

$$g \sin\theta_W = g' \cos\theta_W = e \tag{1.18}$$

The angle θ_W is a free parameter of the SM and is determined experimentally. In addition, because the coupling constants in QFT are not constants but functions of an energy scale at which they are probed, the same applies to θ_W. The running of coupling constants is discussed in more detail in Sec. 1.5.

At this point, it is worth noting a very important limitation of the EW theory as presented up until this point. Experiments sensitive to the mass of the $W^{\pm}$ and Z^0 boson give evidence that these are massive particles. Adding mass terms for gauge fields to the Lagrangian will break the $SU(2)_\mathrm{L} \otimes U(1)_\mathrm{Y}$ symmetry. The problem of these contradictory requirements in the SM is resolved by mechanism of *spontaneous symmetry breaking* (SSB), described in the following section.

1.4 Spontaneous symmetry breaking mechanism

The idea of the SSB mechanism originates in solid-state material physics, where the interactions in the material may obey a certain symmetry, however the ground state of the material does not. In the SM, the SSB mechanism is used to introduce the mass terms for the massive $W^{\pm}$ and Z^0 bosons as well as masses of the fundamental fermions in a manner that does not break the gauge invariance. The mechanism was independently discovered in 1964–65 by Brout and Englert [11], by Higgs [12] and by Guralnik, Hagen and Kibble [13]. In the SM, the SSB mechanism is introduced via an addition of a new complex scalar doublet field that has four degrees of freedom. Since the field must transform according to a non-trivial $SU(2)_\mathrm{L}$ representation, a doublet structure is necessary:

$$\Phi = \begin{pmatrix} \Phi^+ \\ \Phi^0 \end{pmatrix}. \tag{1.19}$$

The dynamic term of this field is the most simple renormalisable theory that obeys the $SU(2)_\mathrm{L}$ symmetry and allows for a non-zero potential minimum:

$$\mathcal{L}_\mathrm{H} = (D^\mu \Phi)^\dagger (D_\mu \Phi) - V(\Phi), \qquad V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2, \tag{1.20}$$

where D_μ is the $SU(2)_\mathrm{L} \otimes U(1)_\mathrm{Y}$-covariant derivative from Eq. 1.14, and $\lambda > 0$ and $\mu^2 < 0$ are free parameters. Forcing the values of parameter to satisfy the aforementioned inequalities causes the potential $V(\Phi)$ to have a minimum $v = \sqrt{\mu^2/\lambda}$, referred to as the *vacuum expectation value* (VEV).

Upon the symmetry breaking, the field Φ is expanded from its VEV:

$$\Phi = \begin{pmatrix} 0 \\ v + H \end{pmatrix}, \tag{1.21}$$

where H is the scalar Higgs boson field. The other three degrees of freedom of the original unbroken

Φ doublet were transferred to the $W^{\pm}$ and Z^0 bosons, which by gaining mass, obtain an additional degree of freedom in the form of a possible longitudinal polarisation. Expanding the kinetic term in Eq. 1.20 using Eq. 1.21, leads to new terms that include interactions between the $W^{\pm}$, Z^0 and H bosons as well as the mass terms, where the masses of the bosons are:

$$m_W = \frac{gv}{2}, \qquad m_Z = \frac{gv}{2\cos\theta_W}, \qquad m_H = \sqrt{\frac{v\lambda}{2}}. \tag{1.22}$$

In addition, the mass terms $m\bar{\Psi}\Psi$ of fundamental fermions also break the $SU(2)_{\mathrm{L}}$ symmetry. This problem is cured in the SM by introducing $SU(2)_{\mathrm{L}}$-covariant *Yukawa* interaction terms $\sim g_f\bar{\Psi}_L\Phi\Psi_R$ terms, where the g_f are the *Yukawa* couplings that express the relative strength of the interaction of the fermions with the Higgs field. After the SSB, the Lagrangian includes terms $\sim g_f v\bar{\Psi}\Psi$, which are mass terms where the mass of the particle is given by the Yukawa coupling and the VEV. The Yukawa couplings as well as the masses of the $W^{\pm}$ and Z^0 bosons are free parameters, that are determined experimentally.

The Higgs boson[(15)] was discovered at the LHC by the ATLAS [14] and CMS [15] experiments in 2012, with a mass of approximately 125 GeV, concluding almost a 50-year period since the proposal of the SSB mechanism in the SM.

1.5 Strong interaction

The strong interaction within the SM is described by QCD, a theory based on the $SU(3)_{\mathrm{C}}$ colour gauge symmetry. The theory describes strong interactions between quarks, and the interaction mediators; gluons. It was formulated in 1973 by Fritzsch, Gell-Mann and Leutwyler [16], employing the previous work of Yang and Mills [4] on the non-Abelian gauge symmetries in QFT. In QCD, quarks carry a colour charge[(16)]. The strong interaction is mediated by eight massless gluons, determined by the dimension of representation of the $SU(3)_{\mathrm{C}}$ group. Similarly to $SU(2)$ non-Abelian group, gluons also interact with each other and carry colour charge.

The kinetic and gluon field terms in the Lagrangian are:

$$\mathcal{L} = \sum_q i\bar{\Psi}_q^j \not{D}_{jk}\Psi_q^k - \frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a, \tag{1.23}$$

with the sum over q denoting the sum over all quark flavours and indices $j, k = 1, 2, 3$ sum over colour states. The gauge-covariant derivative is defined as:

$$D^{\mu}_{jk} = \delta_{jk}\partial^{\mu} - ig_s(T^a)_{jk}G^{\mu}_a, \tag{1.24}$$

where g_s is the strong coupling constant and T^a are Gell-Mann's matrices, generators of the $SU(3)_{\mathrm{C}}$ representation, and the G^{μ}_a are the gluon fields. The field-strength tensor $G^{\mu\nu}_a$ for the gluon fields is defined analogously as in Eq. 1.7.

(15) The discovered particle appears to have properties consistent with the SM-predicted Higgs boson.

(16) The name *colour* is motivated by the fact that there are three unit colours in contrast to, for example, one unit electric charge.

Let us give a very brief overview of history of the strong interactions to highlight a particularly important aspect of the QCD theory. By the 1960s a large number of particles, today known as hadrons, were discovered. These particles had a large production cross-section, indicating that these particles were interacting via a strong interaction. It was hypothesised that this was the same interaction responsible for processes in atomic nuclei. It was noticed that the masses of the hadrons seemed to follow a hierarchy of some underlying symmetry. Gell-Mann [17] and Neeman [18] realised the particles formed mass multiplets of irreducible representations of the $SU(3)$ group, today referred to as the $SU(3)$ *flavour* group. This symmetry is today considered "accidental" because of the very similar masses of the up, down and strange quark. These quarks formed the bound states of all the known mesons and baryons discovered by the 1960s. Gell-Mann and Zweig [19] indeed argued, that the $SU(3)$ flavour symmetry had a physical meaning, that the hadrons were bound states of particles, that Gell-Mann referred to as quarks. The problem with this hypothesis at that time was that quarks were never observed. Studies of structure of the proton via deep inelastic scattering (DIS) experiments during 1960s and 1970s also suggested that proton is a composite particle containing more fundamental point-like constituents.

It was discovered by Gross, Wilczek [20] and Politzer [21] in 1973, that the QCD theory exhibits so-called *asymptotic freedom* – the strength of the interaction decreases with momentum transfer. The relative strength of an interaction in QFT is determined by its coupling constant. Despite the confusing name, the coupling constant is a function of an energy scale at which it is probed. This is a consequence of higher-order corrections in QFT which also impact the coupling constant itself. The so-called *bare* quantities that enter the Lagrangian do not correspond to the physical quantities at a particular energy scale of an experiment. The procedure of renormalisation is used to redefine the perturbative expansions in terms of measurable quantities. The dependence of the strong coupling α_S[(17)] in QCD on the energy scale Q^2 in the next-to-leading order of perturbative series is:

$$\alpha_S(Q^2) = \frac{\alpha_s(\mu_R^2)}{1 + \beta_0 \alpha_s(\mu_R^2) \log(Q^2/\mu_R^2)}, \qquad \beta_0 = \frac{11 n_C - 2 n_f}{12\pi}, \tag{1.25}$$

where $\alpha_S(\mu_R^2)$ is the coupling constant at a particular renormalisation scale μ_R^2, n_C is the number of colours (3 in SM) and n_f is the number of quark flavours (6 in SM). The equation relates the value of the coupling constant at some probed energy scale Q^2 with a value of the constant at some renormalisation scale[(18)] μ_R^2.

It can be shown that α_S diverges if the denominator goes to zero, which happens for a characteristic scale referred to as the $\Lambda_{\text{QCD}} \approx 200\,\text{MeV}$. Eq. 1.25 can be rewritten in terms of Λ_{QCD}^2

$$\alpha_S(Q^2) = \frac{1}{\beta_0 \log(Q^2/\Lambda_{\text{QCD}}^2)}. \tag{1.26}$$

Eq. 1.26 points at two features of the strong interactions. As mentioned, QCD exhibits asymptotic freedom, where in the $Q^2 \to \infty$ limit α_S converges to zero. Secondly, for low energies, neither quarks

[(17)] The α_S and the g_S in the Lagrangian are often both referred to as the coupling constant. The relation between the two in natural units ($\hbar = c = 1$) is $g_s = \sqrt{4\pi\alpha_S}$.

[(18)] This scale dependence is a result of the limited-order perturbative expansion. The cross-section of a process does not depend on the choice of some arbitrary scale, hence the scale dependence of the coupling constant and the scale dependence of the amplitude compensate each other.

nor gluons can be observed as free particles, because they are *confined* to bound states, which are colourless. Confinement is indeed supported by the experimental evidence, though exact theoretical proof does not yet exist. The divergence of α_S for $Q^2 \to \Lambda^2_{\text{QCD}}$ only demonstrates that the perturbative expansion for α_S is not valid in this kinematic region. At low energies, other non-perturbative approaches are generally used to describe quark and gluon interactions, such as the lattice QCD [22] which has also provided clues for the quark confinement.

The behaviour of the strong coupling implies further phenomena particularly relevant to high-energy collider experiments. High-energy quarks and gluons initiate *jets*, collimated showers of hadrons. Due to the strong coupling, hard quarks lose energy by radiating further gluons, and gluons split into $q\bar{q}$ pairs. This process can be described via perturbative QCD. At a sufficiently small energy, the showering stops. The produced quarks bind into mesons and baryons, a process referred to as the *hadronisation*. The hadronisation is a non-perturbative process, typically described by phenomenological model, such as the string model [23] or the cluster model [24].

1.5.1 Parton model and the structure of proton

The parton model was proposed by Feynman [24] to describe the structure of hadrons. It is based on the evidence from electron-proton DIS experiments which showed that at sufficiently high momentum transfer, electrons scattering on the proton appeared to scatter on point-like constituents of the proton.

In the parton model, the DIS on a proton is described as an elastic scattering on a parton which carries a fraction x of the momentum p of the proton. The probability that the parton entering the scattering carries momentum xp is given by a *structure function* $f(x)$. The observation that $f(x)$ does not depend on the momentum transfer Q^2 at sufficiently high energies, is called *Bjorken scaling*, discovered by Bjorken in 1968 [25].

After the discovery of the asymptotic freedom of QCD, it was thought that the partons in the proton are u and d quarks. With increasing energy of the DIS experiments, however, it was realised that the energy scale independence of the structure functions was violated. This is because the proton contains not only the uud quarks, but also a "sea" of other (anti-)quarks and gluons. From this point, when referring to partons in a proton, all of the quark/anti-quark types and gluons are meant.

To provide predictions about the structure of the proton requires non-perturbative approaches to QCD calculations. Because this is very challenging, the structure functions $f(x, Q^2)$, also referred to as the *parton density functions* (PDFs), are measured experimentally. It is, however, possible to calculate scale dependence of the PDFs in perturbative QCD via Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution equations [26–28]. These functions allow to extrapolate PDF predictions from measurements performed at some energy scale to a different energy scale. An illustration of the dependence of the PDFs on the energy scale is shown in Fig. 1.2.

In this thesis we focus on the studies of *hard-scattering* processes, in which partons from the pp collision interact and produce different particles X, for example a top quark–anti-quark pair. Therefore, the proton PDFs are a crucial ingredient because the cross-section of the $pp \to X$ production is a convolution of $x_1 x_2 \to X$ processes, where x_1 and x_2 are the incoming partons. The convolution is described by the *factorization theorem*:

$$\sigma_{pp\to X} = \sum_{i,j} \int dx_i \int dx_j f_i(x_i, \mu_F) f_j(x_j, \mu_F) \sigma_{ij\to X}(i, j, \mu_R, \mu_F). \tag{1.27}$$

The partonic cross-sections $\sigma_{ij\to X}$ is summed over all types of partons in the proton and over all possible momentum fractions x_i and x_j of the incoming partons, weighted by their corresponding PDFs f_i, f_j. The factorization theorem assumes that the partons are behaving as free particles, which is a valid assumption for processes involving high-momentum transfer based on the asymptotic freedom of QCD. The partonic cross-section $\sigma_{ij\to X}(i, j, \mu_R, \mu_F)$ depends on the momenta[(19)] of the incoming partons i, j and also on the choice of the renormalisation scale μ_R and the factorisation scale μ_F. The factorisation scale defines the scale which separates perturbative and non-perturbative parts in the evolution of the PDFs and in the partonic cross-section.

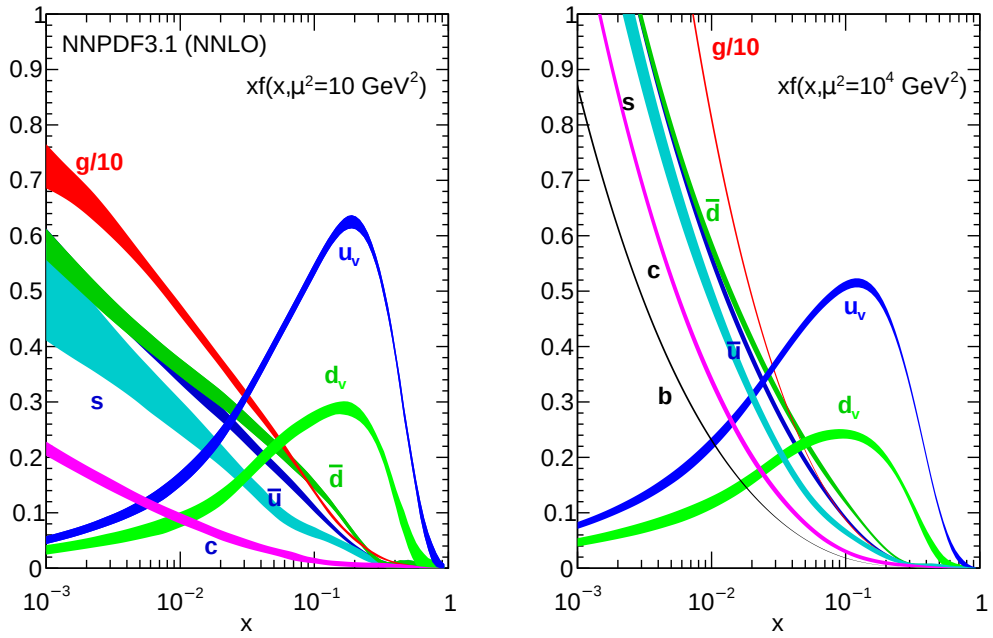


Fig. 1.2: Proton parton density functions from the NNPDF collaboration [29] at two different energy scales. At higher energy scale (right) the relative contribution of gluons (red line) and sea quarks at low momentum fraction x increases.

1.6 Limitations of the Standard Model

Up to this point, we have discussed the interactions included in the SM. The SM has been extensively tested over the past several decades, showing a remarkable agreement across a huge range of different processes. The predictive power of SM is illustrated in Fig. 1.3, which shows the comparison of ATLAS measurements and SM predictions of production cross-sections of processes spanning twelve orders of magnitude. There are, however, several limitations of the SM that prevent it from being a fundamental "theory of everything". Some of the most important omissions include:

- **No description of gravity.** SM does not include gravitational interaction, as quantization of general theory of relativity yields a non-renormalisable QFT.
- **No explanation of the mass range of the fermions**. The range of masses of fundamental particles in the SM spans twelve orders of magnitude. The SM gives no clues as to why this is

[(19)] There may be dependence on other properties of the partons if relevant, such as spin. It is implicitly assumed that sum over all relevant initial parton states is performed in the factorization theorem.

the case.

- **Properties of neutrinos in the SM.** The neutrinos in the SM are massless, effectively predicting existence of sterile right-handed neutrinos. However, experimental observations of neutrino flavour oscillations prove that at least two neutrino flavours must have non-zero mass to allow for the mixing of mass and weak eigenstates. It is possible to extend the SM by adding mass terms for neutrinos. However, due to the zero electric charge of neutrinos, it is not clear whether neutrinos are Dirac particles (distinct particle and anti-particle), or Majorana particles (particle is its own anti-particle).
- **Insufficient explanation of the baryon asymmetry in the universe.** It is generally accepted, that at the Big Bang, matter and anti-matter was created in equal amounts, however the visible universe appears to lack anti-matter. CP violation is necessary to introduce the asymmetry after the Big Bang, however the sources of CP violation in the SM are not sufficiently strong to explain this phenomenon.
- **No explanation of dark matter and dark energy.** The SM only describes about 5 % of the content of the universe, the visible matter. There are beyond-SM (BSM) theories, such as *supersymmetry* that predict neutral weakly interacting particles that could be dark matter candidates. However, no explanation of the dark energy is provided by SM or its extensions.

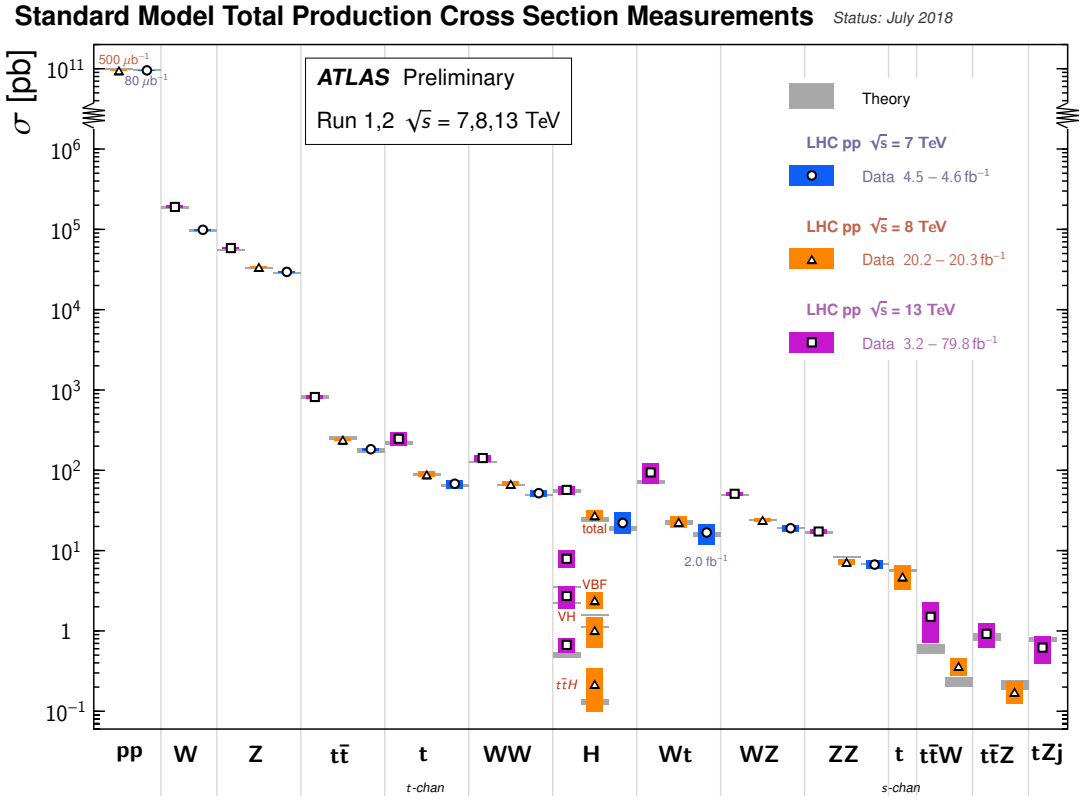


Fig. 1.3: Summary of several total cross-section measurements compared to the corresponding SM theoretical predictions [30]. The coloured bands show the uncertainties on the measurements while the grey bands show the uncertainties on the theoretical predictions. All of the theoretical predictions were calculated with NLO precision or higher.

2

Top quark and the charge asymmetry

The top quark is the third-generation up-type quark. It was discovered in 1995 at the Tevatron accelerator by the CDF and D0 experiments, completing the three generations of quarks predicted by the SM. It is the heaviest known elementary particle with a mass of approximately 173 GeV [31]. Because of the large mass, the top quark has a very large predicted decay width $\Gamma = 1.32\,\text{GeV}$ [32], resulting in an extremely short mean life-time[(1)] of $\approx 10^{-25}$ s. It is in fact order of magnitude less than the mean hadronisation time[(2)] $\approx 10^{-23}$ s. This means that the properties of the top quark are passed directly onto its decay products, allowing for precise measurements of properties of a "pseudo-bare" quark.

The top quark has the largest Yukawa coupling to Higgs boson $y_t \sim 1$ of all SM fermions, suggesting that it plays an important role in the EW sector. Studies of the top-Higgs Yukawa coupling are investigated in measurements of the associated production of the $t\bar{t}$ pair and the Higgs boson ($t\bar{t}H$) which can probe the consistency of the SM as well as search for anomalous Yukawa coupling. Both $t\bar{t}$ as well as single top production and top-quark decays are sensitive to various BSM physics, ranging from $t\bar{t}$ resonances, flavour changing neutral currents in top-quark decays, to anomalous couplings of top-quark to Higgs boson and to EW vector bosons. In addition, the top-quark production processes are an important background in many BSM searches involving top quarks as well as processes involving the Higgs boson production.

At hadron colliders, top quarks are most abundantly produced via $t\bar{t}$ pair production which is driven by strong interactions, though other production channels are also investigated, such as the single top quark production via weak interaction.

2.1 Top quark pair production

The pair production at hadron colliders occurs primarily through two initial states, either a quark-antiquark annihilation ($q\bar{q} \to t\bar{t}$), or a gluon fusion ($gg \to t\bar{t}$). The leading-order (LO) Feynman diagrams depicting these production channels are shown in Fig. 2.1. The relative contribution of $q\bar{q} \to t\bar{t}$ and $gg \to t\bar{t}$ depends on the PDFs, and more specifically on the collision energy as well as on the collided hadrons. The top quark was discovered at the Tevatron collider, which was a proton-antiproton ($p\bar{p}$) collider operating at centre-of-mass energy $\sqrt{s} = 1.96\,\text{TeV}$. At this energy regime and due to the $p\bar{p}$ collisions, the $q\bar{q} \to t\bar{t}$ is the dominant production channel ($\approx 85\,\%$), where the initial-state quark originates from proton and initial-state anti-quark from anti-proton. At the LHC, which is a proton-proton (pp) collider, the situation is dramatically different. For $q\bar{q} \to t\bar{t}$

[(1)] The mean life time of a particle is $\tau = \Gamma^{-1}$, where Γ is the decay width of the particle.

[(2)] The mean hadronisation time is estimated as $\Lambda_{\text{QCD}}^{-1}$.

production channel at least one of the quarks must come from the proton sea. The sea-quark PDFs at the LHC energies are more suppressed than the valence-quark or gluon PDFs. Additionally, due to the increase in $\sqrt{s}$, more $t\bar{t}$ pairs are produced via partons with low momentum fraction x. The gluon PDF dominates for small values of x. Due to these factors, the dominant $t\bar{t}$ production channel at the LHC is $gg \to t\bar{t}$. At $\sqrt{s} = 13\,\text{TeV}$, the contribution is $\approx 90\,\%$.

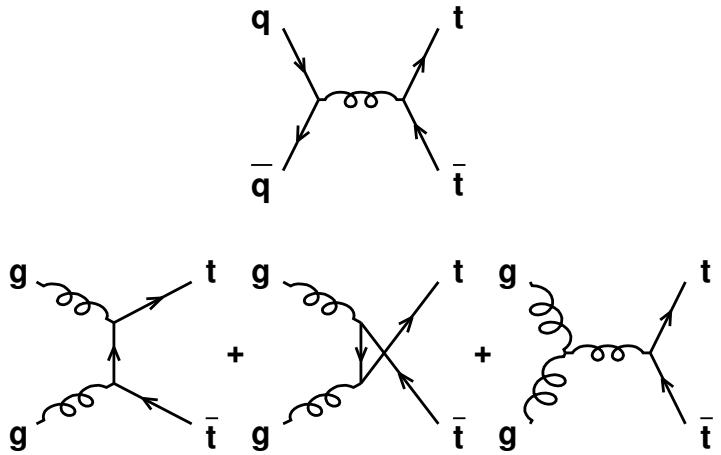


Fig. 2.1: Leading-order Feynman diagrams of $t\bar{t}$ pair production: $q\bar{q}$ annihilation (top), and gg fusion (bottom) [33].

The latest inclusive cross-section predictions [34–37] are calculated up to next-to-next-to-leading order (NNLO) accuracy in QCD with soft-gluon resummations up to next-to-next-to-leading logarithmic (NNLL) accuracy. These calculations are implemented in the Top++ v2.0 [38] software. The predictions for Tevatron and for LHC at various energies are shown in Table 2.1. The predictions are calculated assuming a top-quark mass of 172.5 GeV and use MSTW2008 NNLO PDF set [39]. A comparison of the predictions with latest cross-section measurements is shown in Fig. 2.2, showing good agreement between measurements and the SM.

Table 2.1: Predicted inclusive $t\bar{t}$ cross-sections based on NNLO QCD calculations [34–38] for the Tevatron $p\bar{p}$ collider and LHC pp collider at various centre-of-mass energies. The uncertainties include renormalisation and factorisation scale variations as well as uncertainty from PDF variations.

Accelerator ($\sqrt{s}$)	$\sigma_{t\bar{t}}$[pb]
Tevatron (1.96 TeV)	$7.35^{+0.20}_{-0.24}$
LHC (7 TeV)	$172\ ^{+6.4}_{-7.5}$
LHC (8 TeV)	253^{+15}_{-16}
LHC (13 TeV)	832^{+40}_{-46}

2.2 Top quark decay modes

The decay modes of a quark are proportional to the square amplitude of the corresponding CKM matrix elements. In the SM, the top quark decays almost exclusively into a W^+b pair [31]. Therefore, the decay modes of the top quark are essentially given by the decay modes of the W boson, which can

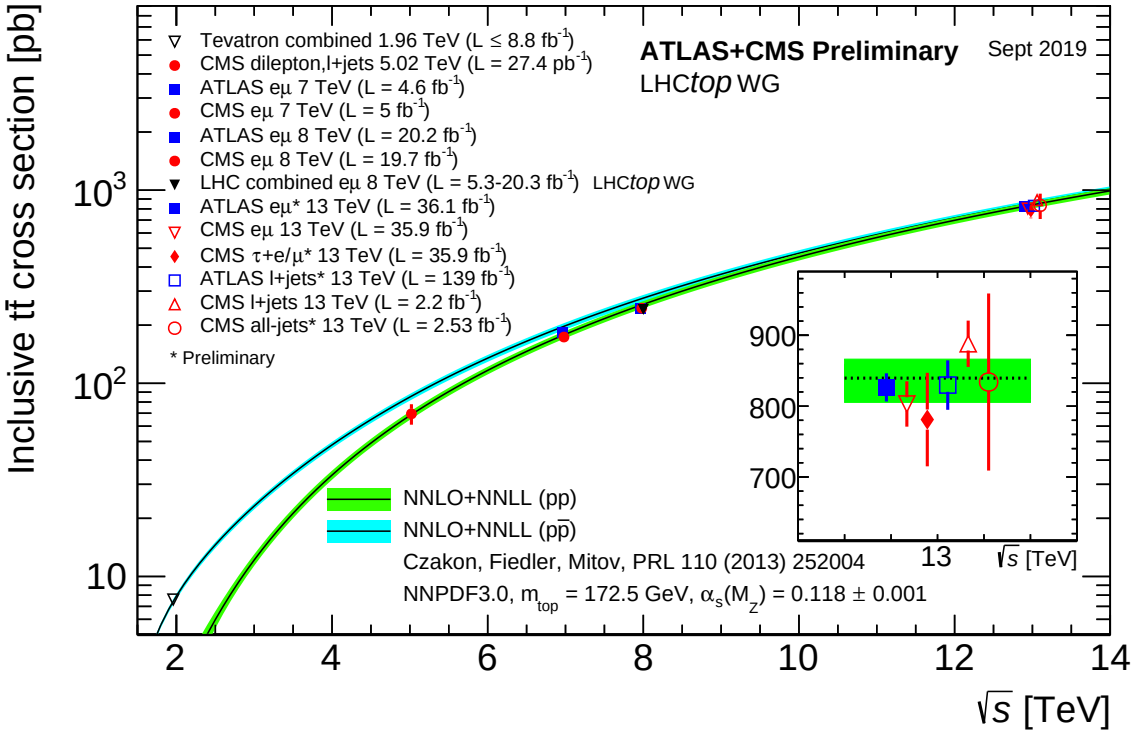


Fig. 2.2: The prediction of dependence of the inclusive $t\bar{t}$ production cross-section [34–38] vs $\sqrt{s}$ for pp (green line) and $p\bar{p}$ (cyan line) collisions. The points show a comparison with various recent Tevatron, ATLAS and CMS measurements. The summary is compiled by the LHC Top Working Group [40].

decay into a $q\bar{q}'$ pair ($\approx 67\,\%$) or a pair of $\ell\nu_\ell$[(3)] [31]. In the context of the $t\bar{t}$ pair production, the final state thus always contains two b quarks, and it is typical to distinguish three decay channels depending on the number of charged leptons in the final state, as shown in Fig. 2.3. The decay channels are as follows:

- *all-hadronic* channel: both of the W bosons decay into quarks,
- *single-lepton* channel: one of the W bosons decays into quarks, and the other into $\ell\nu_\ell$,
- *dilepton* channel: both of the W bosons decay into $\ell\nu_\ell$.

The single-lepton channel can be considered the most versatile decay channel for various measurements. It has sufficiently large branching fraction and good signal-to-background ratio. The presence of the lepton allows for good background suppression. The challenge in the reconstruction of the $t\bar{t}$ system in single-lepton channel is the presence of the single neutrino in the final state. It is typically inferred from missing energy in the detector, since the transverse momentum of the initial-state partons from the hadron-hadron collision is zero. However, only the transverse component of neutrino momentum can be determined unambiguously, because the longitudinal momentum of the initial state is unknown. The longitudinal momentum of the neutrino can be partially determined due to constraints on the W mass in the $W \to \ell\nu$ decay, however the direction of the longitudinal momentum is undetermined. The reconstruction of the $t\bar{t}$ system thus has an inherent ambiguity.

The dilepton channel has the smallest of the branching ratios of the three channels, but has the

[(3)] We will not further distinguish in the discussion the charges of the W bosons and which of the decay products are (anti)particles, but for completeness the W boson leptonic decays are $W^+ \to \bar{\ell}\nu_\ell$, $W^- \to \ell\bar{\nu}_\ell$.

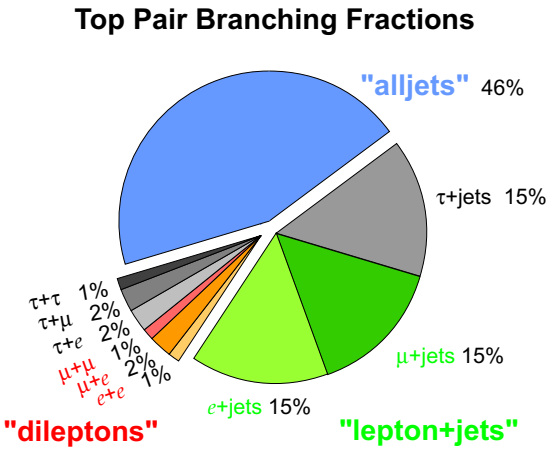


Fig. 2.3: The branching fractions of the $t\bar{t}$ decay channels [31]. Here "alljets" reffers to the all-hadronic final state, and "lepton+jets" to the single-lepton final state.

lowest background contribution, and the final state contains only two jets, which are typically measured less precisely than leptons. The major disadvantage of this channel is the presence of two neutrinos, which lead to an under-constrained system. A full $t\bar{t}$ system reconstruction is thus challenging.

Finally, the all-hadronic channel is the channel with largest branching fraction, and no neutrinos in the final state. This means that in principle the $t\bar{t}$ system can be fully reconstructed. In practice, due to the presence of six jets in final state, the correct assignment of jets to final-state partons is more complicated than for the single-lepton channel. The experimental resolution of observables sensitive to jet energy is also worse compared to single-lepton channel. Finally, the all-hadronic channel suffers from large multijet background, that is difficult to model, typically requiring complex data-driven techniques to accurately determine. As we will discuss in Ch. 5, for studying high-momentum (*boosted*) top-quarks, it is possible to reduce the complexity of the reconstruction of the jets from top quark decays as well as improve the suppression of the multijet background using dedicated techniques, thus improving the usefulness of the all-hadronic channel.

2.3 Charge asymmetry in heavy quark pair production

The charge asymmetry in the context of pair production of quarks describes an anisotropy in the angular distributions of final-state quark and anti-quark production. In order to better visualise this phenomenon, let us consider the $q\bar{q} \to Q\bar{Q}$ process, where $Q\bar{Q}$ is a pair of quarks such as top or bottom quark. Let us consider that Q is produced under some production angle ϑ with respect to q in the incoming $q\bar{q}$ pair rest frame, as shown in Fig. 2.4. One can then define a differential charge asymmetry as follows:

$$A_C^{Q\bar{Q}}(\cos\vartheta) = \frac{N_Q(\cos\vartheta) - N_{\bar{Q}}(\cos\vartheta)}{N_Q(\cos\vartheta) + N_{\bar{Q}}(\cos\vartheta)} \tag{2.1}$$

$$N_Q(\cos\vartheta) = \frac{d\sigma_{Q\bar{Q}}}{d\Omega(\cos\vartheta)} \tag{2.2}$$

$$N_{\bar{Q}}(\cos\vartheta) = \frac{d\sigma_{\bar{Q}Q}}{d\Omega(\cos\vartheta)} \tag{2.3}$$

Here $N_Q(\cos\vartheta)$ denotes the production rate of $Q\bar{Q}$, where Q is produced in the production angle interval $(\vartheta, \vartheta + d\vartheta)$ and $N_{\bar{Q}}(\cos\vartheta)$ is the production rate when Q and $\bar{Q}$ are swapped in the final state.

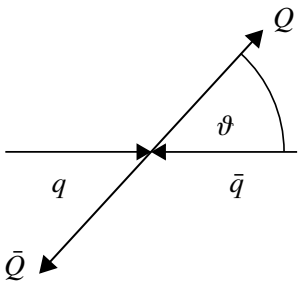


Fig. 2.4: Illustration of production angle ϑ in $q\bar{q} \to Q\bar{Q}$ process in the $q\bar{q}$ rest frame.

Assuming combined charge and parity (CP) conservation, it follows that in the $q\bar{q}$ rest frame the charge asymmetry can also be explained in terms of a forward-backward asymmetry with respect to the direction of the incoming q since:

$$N_{\bar{Q}}(\cos\vartheta) = \frac{d\sigma_{\bar{Q}Q}}{d\Omega(\cos\vartheta)} = \frac{d\sigma_{Q\bar{Q}}}{d\Omega(\cos(\pi-\vartheta))} = N_Q(-\cos\vartheta). \tag{2.4}$$

As was mentioned in Sec. 2.1, $Q\bar{Q}$ pairs at hadron collider are mostly produced via $gg \to Q\bar{Q}$ or $q\bar{q} \to Q\bar{Q}$. Since the charge asymmetry requires charge-asymmetric initial state, the $gg \to Q\bar{Q}$ production is always symmetric. In addition, at leading order, SM does not predict any charge asymmetry in $q\bar{q} \to Q\bar{Q}$ either.

However, at higher orders, several sources of corrections introduce an asymmetry. Firstly, it is the interference of several NLO QCD contributions (α_s^3) to scattering amplitude in $q\bar{q} \to Q\bar{Q}$ processes. The interference between initial-state (ISR) and final-state (FSR) radiation leads to a negative contribution (Fig. 2.5), while the interference between box and Born amplitudes leads to a positive contribution (Fig. 2.5). While the relative contribution of these amplitudes to the asymmetry depends on the region of phase space examined, inclusively, the total contribution from these corrections results in a positive asymmetry. In other words, on average Q is produced preferentially in direction of q. Another contribution to the asymmetry is due to the interference of amplitudes in the flavour excitation quark-gluon processes ($g + q \to Q\bar{Q} + q$, Fig. 2.6) [41, 42]. The relative asymmetric contribution of these processes is shown in Fig. 2.7.

Other contributions to the asymmetry include additional interference terms with electro-weak interactions at order of $\alpha\alpha_s^2$ [42–44], which can be basically obtained by replacing gluon propagator for Z or γ^* propagator in Fig. 2.5. In particular, the mixed QCD-QED contributions were found to be non-negligible corrections to the purely QCD α_s^3 asymmetry contribution in $q\bar{q} \to Q\bar{Q}$, enhancing the asymmetry by up to 20 % for $t\bar{t}$ production [44].

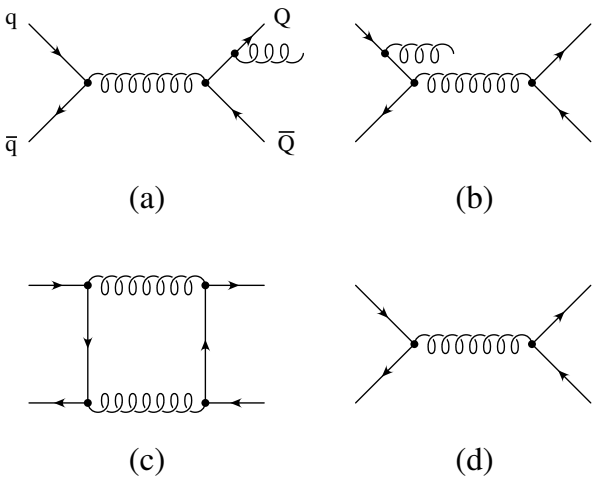


Fig. 2.5: Representative scattering amplitudes relevant for the interference terms inducing charge asymmetry in $q\bar{q} \to Q\bar{Q}$: gluon ISR (a), FSR (b), box (c) and leading-order (Born) (d) amplitudes [41].

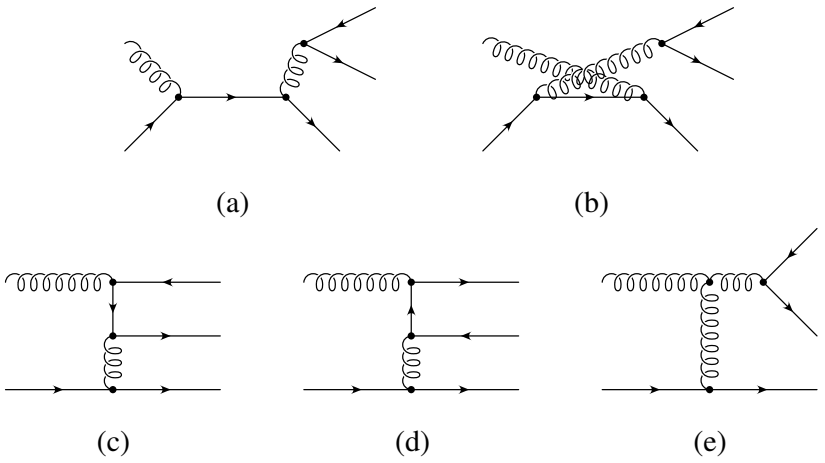


Fig. 2.6: Representative scattering amplitudes relevant for the interference terms inducing charge asymmetry in $q + g \to Q\bar{Q} + q$ [41].

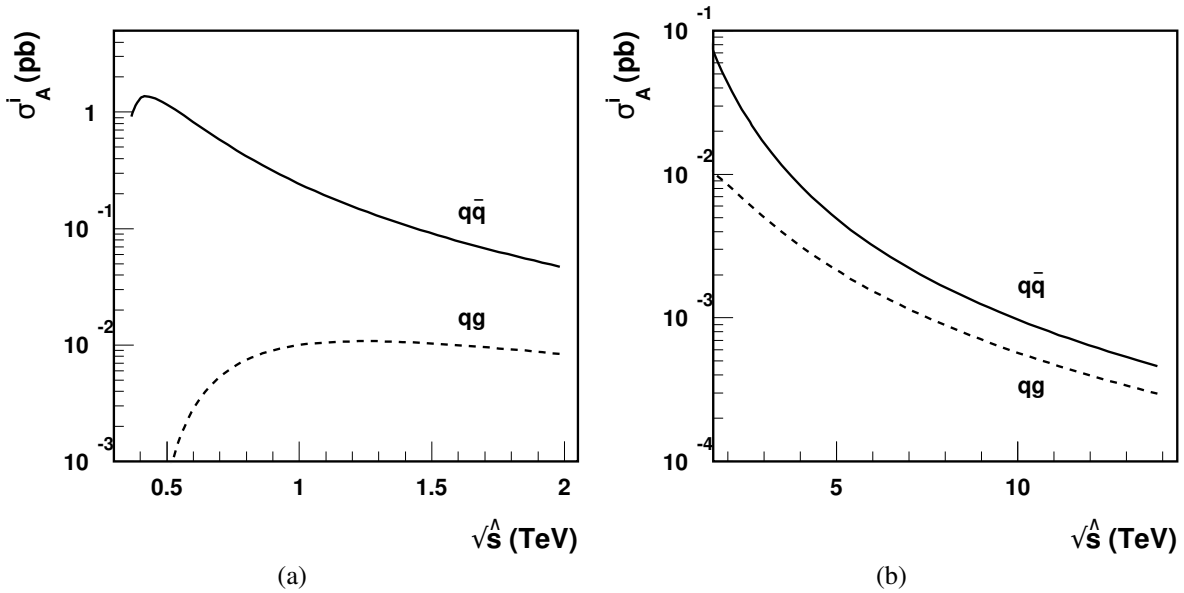


Fig. 2.7: Dependence of the asymmetric cross-section part on the partonic centre of mass energy [41], for $q\bar{q} \to t\bar{t}$ processes and for $qg \to t\bar{t}$.

2.4 Charge asymmetry measurements at the Tevatron

A substantial investigation of the charge asymmetry in heavy quark production was induced by measurements at the Tevatron accelerator by CDF and D0 experiments. Tevatron was a very suitable collider for studying $t\bar{t}$ charge asymmetry due to the dominant $q\bar{q} \to t\bar{t}$ production channel, thus suffering from only a small $gg \to t\bar{t}$ dilution.

2.4.1 Observables sensitive to charge asymmetry at the Tevatron

Several observables were studied in $t\bar{t}$ final states with one or two leptons (electrons and muons). All of the observables discussed in this section make use of the assumption that in the $q\bar{q} \to t\bar{t}$ process, the initial quark is from the proton and the anti-quark from anti-proton. Combining this fact with the assumption about CP conservation, it follows from Eq. 2.4, that at Tevatron it is possible to measure a forward-backward asymmetry. The most common definition of this observable uses kinematics of both the quarks from the $t\bar{t}$ pair:

$$A_{\mathrm{FB}} = \frac{N(\Delta y > 0) - N(\Delta y < 0)}{N(\Delta y > 0) + N(\Delta y < 0)}, \tag{2.5}$$

where $\Delta y = y_t - y_{\bar{t}}$ is the rapidity difference of top and antitop quark. The rapidity of a particle is defined as follows:

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z}, \tag{2.6}$$

where E is the energy and p_z the longitudinal momentum of the particle. This variable gives us an information about the direction of the movement of the particle with respect to the z-axis, which is identical to the direction of the proton beam. If the particle is moving forward with respect to proton beam, $y > 0$, and vice-versa. For $y = 0$ case, the particle is moving exactly perpendicular to the proton beam. In the $q\bar{q}$ rest frame, the t and $\bar{t}$ are produced back-to-back so it follows that $\Delta y > 0$ when $y_t > 0$ and vice-versa. The Δy is used because it is Lorentz-invariant with respect to boosts along the z-axis. This is a useful property since in hadron-hadron collisions, the longitudinal momenta of the interacting partons are random, and so the $q\bar{q}$ system is in general boosted along the z-axis direction. Using a non-invariant definition would otherwise lead to additional dilution of the observed asymmetry. Finally, measuring A_{FB} requires full reconstruction of the $t\bar{t}$ kinematics. The charge of the lepton from the leptonically-decaying (anti)top quark is used to distinguish the reconstructed top quark from antitop quark candidate.

Additionally, observables sensitive to the asymmetry using only the final-state leptons can be constructed. In the single-lepton channel, using rapidity distribution of the single lepton, it is possible to measure:

$$A_{\mathrm{FB}}^{\ell} = \frac{N(q_\ell y_\ell > 0) - N(q_\ell y_\ell < 0)}{N(q_\ell y_\ell > 0) + N(q_\ell y_\ell < 0)}, \tag{2.7}$$

where $q_\ell y_\ell$ is the product of lepton charge and its rapidity. This definition makes the assumption that if top quark is produced more abundantly in one direction with respect to the proton beam, then this property transfers to the lepton from the leptonic top decay.

Finally, in the dilepton channel, it is also possible to measure leptonic asymmetry using the rapidity difference of the two leptons from $t\bar{t}$ decay analogously to $A_{\mathrm{FB}}^{t\bar{t}}$ definition, since kinematics of both of

the leptons are correlated with the $t\bar{t}$ kinematics:

$$A_{\mathrm{FB}}^{\ell\ell} = \frac{N(\Delta y_{\ell\ell} > 0) - N(\Delta y_{\ell\ell} < 0)}{N(\Delta y_{\ell\ell} > 0) + N(\Delta y_{\ell\ell} < 0)}, \tag{2.8}$$

where $\Delta y_{\ell\ell} = y_{\ell^+} - y_{\ell^-}$ is the rapidity difference of the lepton from top and lepton from antitop, respectively. Both A_{FB}^{ℓ} and $A_{\mathrm{FB}}^{\ell\ell}$ asymmetries are affected by dilution due to neglecting other decay products from the top-quark decays. The limited coverage of the tracking system also further reduces the sensitivity to asymmetry which is pronounced more in the forward region.

2.4.2 Measurements at the Tevatron

The first $t\bar{t}$ forward-backward asymmetry measurements by CDF and D0 experiments performed in Run-II period of data-taking using dataset with integrated luminosity (L) of approximately 5 fb^{-1} showed an unexpected tension in comparison to the SM prediction calculated up to NLO accuracy in QCD [45–48]. In particular, the CDF measurement [49] of A_{FB} as a function of the invariant mass of the $t\bar{t}$ system ($m_{t\bar{t}}$)[(4)], showed a much stronger dependence than predicted (Fig. 2.8), reaching deviations of more than three standard deviations for high-$m_{t\bar{t}}$ region of the phase space. The D0 experiment also reported inconsistency with the SM prediction for the inclusive A_{FB} measurement [50], though no significant dependence with respect to $m_{t\bar{t}}$ was observed.

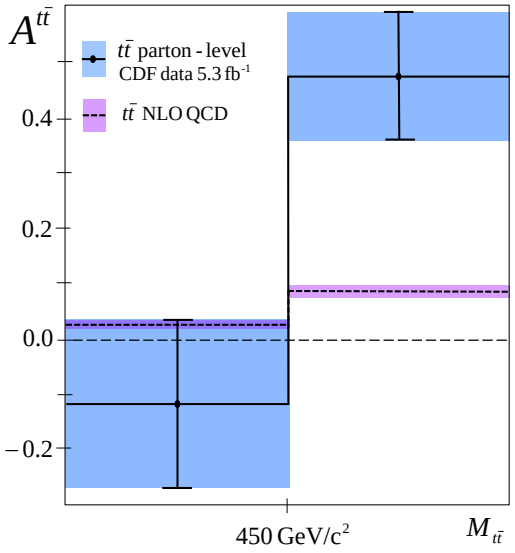


Fig. 2.8: Comparison of $A_{\mathrm{FB}}^{t\bar{t}}$ NLO QCD prediction [45–48] with measurement by the CDF experiment [49] in $t\bar{t}$ events with single lepton. The comparison is performed in bins of $m_{t\bar{t}} < 450$ GeV and $m_{t\bar{t}} > 450$ GeV. The shaded blue area on the measurement points shows the total uncertainty while the purple area shows the uncertainty on the theoretical prediction.

The initial partial-dataset measurements were followed up by full Run-II datasets measurements of $A_{\mathrm{FB}}^{t\bar{t}}$ [51–53], A_{FB}^{ℓ} [54, 55] and $A_{\mathrm{FB}}^{\ell\ell}$ [56, 57], which benefit from the almost factor of two increase in integrated luminosity. These measurements also improved the precision by combining single-lepton and dilepton channels and combining both CDF and D0 datasets.

[(4)] The invariant mass of $t\bar{t}$ system is defined as $\sqrt{(E_t + E_{\bar{t}})^2 - |\vec{p}_t + \vec{p}_{\bar{t}}|^2}$, where E_t ($E_{\bar{t}}$) is the energy of the t ($\bar{t}$) quark, and $\vec{p}_t$ ($\vec{p}_{\bar{t}}$) is the momentum vector of the t ($\bar{t}$) quark.

The results also motivated theory efforts to provide more precise SM predictions, which resulted in NLO EW and NNLO QCD calculations [58–60]. These predictions showed a sizeable increase in the asymmetry, reducing the tension between theory and experiment. A summary comparison of all the inclusive measurements with the SM predictions is shown in Fig. 2.9. The improved predictions also reduced the tension in the differential $A_{\mathrm{FB}}^{t\bar{t}}$ vs $m_{t\bar{t}}$ measurements, as is shown in Fig. 2.10. Additionally, the A_{FB} measurements prompted BSM interpretations, that are discussed in Sec. 2.6 for both the Tevatron and the LHC.

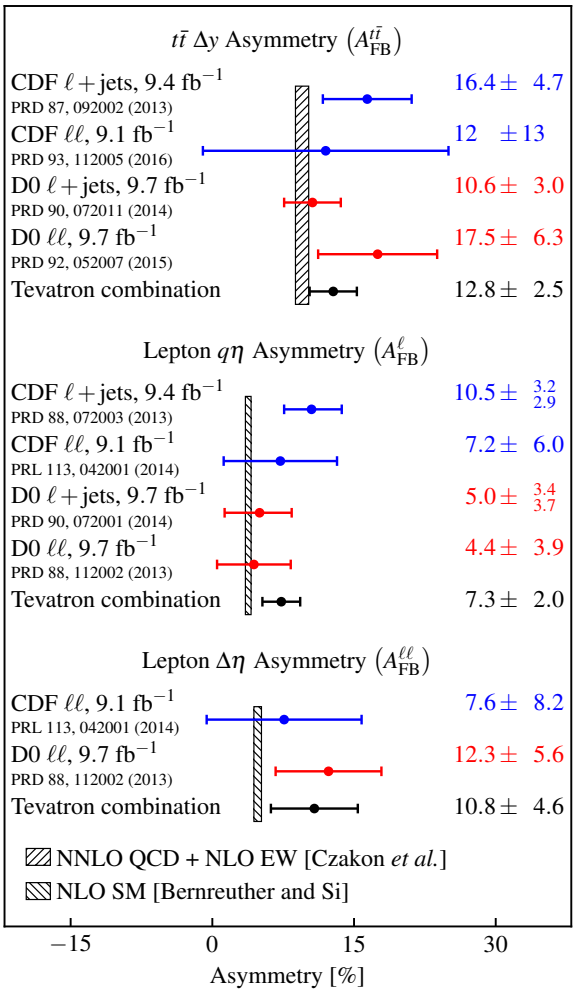


Fig. 2.9: Summary comparison of all inclusive A_{FB} measurements [53] by CDF and D0 experiments with NLO QCD+EW [58] and NNLO QCD + NLO EW [59, 60] SM predictions. The summary includes measurements using both $t\bar{t}$ reconstruction ($A_{\mathrm{FB}}^{t\bar{t}}$) [51–53] as well as using only observables related to the lepton(s) from $t\bar{t}$ decays (A_{FB}^{ℓ}, $A_{\mathrm{FB}}^{\ell\ell}$) [54–57]. The hash bands show the theoretical prediction and its uncertainty, while the points and the bars show the measurements and their total uncertainty.

In addition to the A_{FB} measurements in $t\bar{t}$ production, measurements of A_{FB} in $b\bar{b}$ production were also performed with the aim to investigate a different region of phase space, given the large difference between the top and bottom quark mass. The reconstruction of the $b\bar{b}$ pairs involves identifying events with two jets originating from B-hadrons. In addition, it is necessary to distinguish the charge of the b and $\bar{b}$ quark. One option is to consider only semileptonic B-hadron decays with a soft muon in the final state and using the soft muon charge as a proxy to the charge of the B-hadron. This approach has

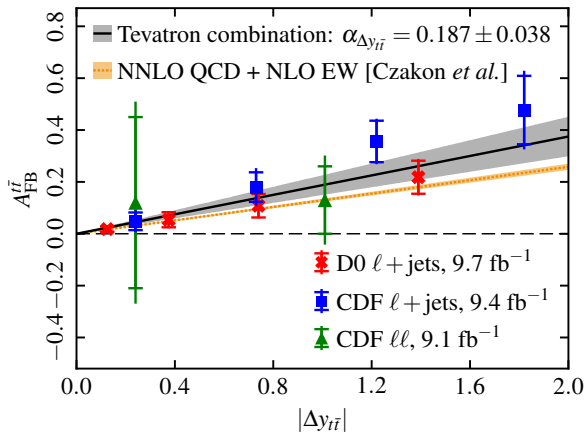
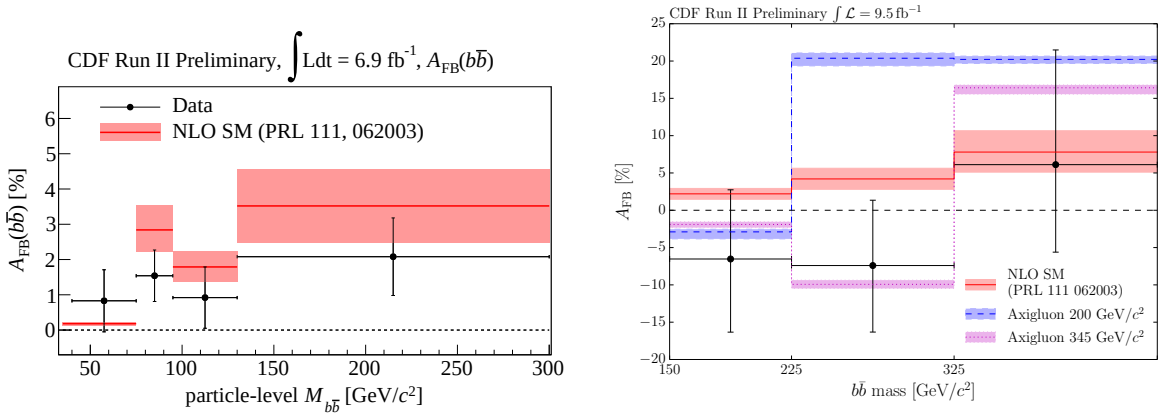


Fig. 2.10: Comparison of $A_{\mathrm{FB}}^{t\bar{t}}$ as a function of $m_{t\bar{t}}$, of CDF and D0 measurements [51, 52, 56] and their combination [53] with the NNLO QCD + NLO EW SM prediction [59, 60]. The inner error bars on data points show the statistical uncertainty while the outer bars show the total uncertainty. A one-parameter linear fit is performed to all the data points, shown by the black line, where the grey area shows the uncertainty on the fitted parameter. A similar band is shown for the theoretical prediction.

been used in the measurement of $A_{\mathrm{FB}}^{b\bar{b}}$ in events with low-$m_{b\bar{b}}$ [61] (Fig.2.11a). This approach suffers from the low branching ratio of the B-hadron decays to soft muons (approximately 11 % [31]), but at the same time the identification of the charge of the muon is very precise. A different approach was used in the CDF measurement of the $A_{\mathrm{FB}}^{b\bar{b}}$ in high-$m_{b\bar{b}}$ events [62] (Fig.2.11b), where the b-quark charge was inferred from the tracks matched to the calorimeter jet [63].

Both measurements of $A_{\mathrm{FB}}^{b\bar{b}}$ report no significant deviation with respect to SM prediction in Ref. [64]. In contrast to $t\bar{t}$ production, in the case of $b\bar{b}$ production the asymmetry is strongly diluted by the dominant, charge-symmetric gluon fusion production channel.



(a) (b)

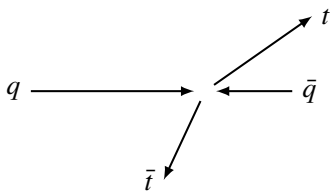
Fig. 2.11: Measurements of the $A_{\mathrm{FB}}^{b\bar{b}}$ as a function of $m_{b\bar{b}}$ in low-$m_{b\bar{b}}$ [61] (a) and high-$m_{b\bar{b}}$ [62] (b) region of phase space. Predictions for axigluon contributions with two different masses are shown in (b). The error-bars on data points show the total uncertainty on the measurement, while the bands show the theoretical predictions from Ref. [64] and their uncertainties.

2.5 Charge asymmetry measurements at the LHC

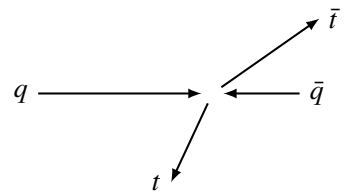
2.5.1 Observables sensitive to charge asymmetry at the LHC

Measurements of charge asymmetry at the LHC are significantly more challenging compared to Tevatron both due to the pp collisions as well as the energy regime. The asymmetry is strongly diluted due to the dominant, charge-symmetric $gg \to t\bar{t}$ production channel. In addition, in the pp collisions, there is no apparent significant direction and the asymmetry definition in Eq. 2.5 cannot be used. However, in case of $q\bar{q} \to t\bar{t}$ events in pp collisions, typically a valence quark q from one proton interacts with a sea anti-quark $\bar{q}$ from the other proton[(5)]. Valence quarks tend to have, on average, higher momentum compared to sea quarks, therefore the centre-of-mass of the $q\bar{q}$ interaction will be on average boosted in the direction of q. This effectively means, that if the top quark is produced in the direction of q, its rapidity magnitude $|y_t|$ will be larger compared to $|y_{\bar{t}}|$ (Fig. 2.12a). On the other hand, if anti-top quark is produced in the direction of q, then $|y_{\bar{t}}| > |y_t|$ (Fig. 2.12b). It follows, that one can use absolute difference of rapidities of top and anti-top quark $\Delta|y| = |y_t| - |y_{\bar{t}}|$ as an observable that is sensitive to the direction of the top quark with respect to q in the $q\bar{q}$ rest frame. The impact of a positive charge asymmetry on the rapidity distributions of top and anti-top quark is visualized in Fig. 2.12c. The charge asymmetry definition usable at LHC can be then defined as follows:

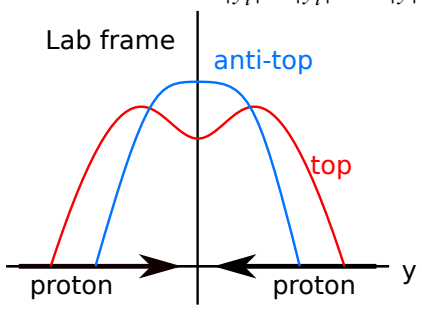
$$A_{\mathrm{C}}^{t\bar{t}} = \frac{N(\Delta|y| > 0) - N(\Delta|y| < 0)}{N(\Delta|y| > 0) + N(\Delta|y| < 0)} \tag{2.9}$$



(a) t produced in direction of q:
$|y_t| > |y_{\bar{t}}| \Rightarrow \Delta|y| > 0$



(b) $\bar{t}$ produced in direction of q:
$|y_t| < |y_{\bar{t}}| \Rightarrow \Delta|y| < 0$



(c)

Fig. 2.12: Illustration of the effect of $q\bar{q}$ longitudinal boost in $q\bar{q} \to t\bar{t}$ on the rapidity magnitudes of t (a) and $\bar{t}$ (b) quarks and their rapidity distributions in the laboratory frame at the LHC under positive asymmetry assumption (c). If the asymmetry was zero, the rapidity distributions of t and $\bar{t}$ would coincide.

[(5)] It is also possible that a sea quark interacts with sea antiquark, however such initial state is heavily suppressed by sea-quark PDFs.

In the case of dileptonic $t\bar{t}$ decays, it is also possible to measure so-called leptonic charge asymmetry, which is defined using pseudorapidities of the two final-state leptons:

$$A_{\mathrm{C}}^{\ell\ell} = \frac{N(\Delta|\eta| > 0) - N(\Delta|\eta| < 0)}{N(\Delta|\eta| > 0) + N(\Delta|\eta| < 0)} \tag{2.10}$$

where $\Delta|\eta| = |\eta_{\ell^+}| - |\eta_{\ell^-}|$ is the difference of absolute pseudorapidities of the leptons from top and anti-top quark, respectively. In contrast to $A_{\mathrm{C}}^{t\bar{t}}$, this observable does not require full reconstruction of $t\bar{t}$ system, and is less sensitive to detector-related systematic uncertainties, because in general leptons are more well-measured objects than jets from quarks, however the observable is more diluted due to the three-body top-quark decay.

While the observables in Eq. 2.9 and 2.10 are commonly referred to as charge asymmetries, it should be noted that they are rather "forward-central" asymmetries, and only have an indirect connection to the asymmetry defined in Eq. 2.1. In addition, the predicted magnitudes for these observables are suppressed by more than an order of magnitude compared to the Tevatron predictions. On the other hand, the increase in collision energy at the LHC allows to probe for evidence of BSM theories that predict seizable contribution to the charge asymmetry at energies previously inaccessible at the Tevatron. In addition, it is possible to impose additional kinematic criteria to select regions of phase where the charge asymmetry is enhanced. One such option is probing higher invariant masses of the $t\bar{t}$ pair ($m_{t\bar{t}}$). This is because higher-$m_{t\bar{t}}$ $t\bar{t}$ pairs are produced by incoming partons carrying higher momentum fraction x of the colliding protons. The probability that the interacting partons are $q\bar{q}$ rather than gg increases with x as was shown in Fig. 1.2. Another option to enhance the $q\bar{q} \to t\bar{t}$ contribution is to cut on the longitudinal boost of the $t\bar{t}$ pair $\beta_{\mathrm{z},t\bar{t}}$[(6)]. Finally, an option to directly impact the underlying asymmetry in $q\bar{q} \to t\bar{t}$ is to impose a cut on the maximum p_{T} of the $t\bar{t}$ pair ($p_{\mathrm{T},t\bar{t}}$). This constraints the amount of ISR/FSR in the $t\bar{t}$ production, reducing the negative contribution to the asymmetry from the interference term of the ISR/FSR amplitudes.

2.5.2 Measurements at the LHC

At the LHC, both ATLAS and CMS performed inclusive and differential measurements of the charge asymmetry, using Run-I collision data at $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV.

The observables investigated were very similar between ATLAS and CMS, measuring both inclusive $A_{\mathrm{C}}^{t\bar{t}}$ and $A_{\mathrm{C}}^{\ell\ell}$ as well as differential $A_{\mathrm{C}}^{t\bar{t}}$ as a function of $m_{t\bar{t}}$, $p_{\mathrm{T},t\bar{t}}$ and $|y_{t\bar{t}}|$ or $\beta_{\mathrm{z},t\bar{t}}$. The 7 TeV measurements [65–68] were largely statistically dominated due to the relatively small integrated luminosity of approximately 5 fb^{-1}. The 8 TeV measurements [69–73] achieved a significantly better precision thanks to the much larger dataset with integrated luminosity of approximately 20 fb^{-1}, allowing for differential measurements of $A_{\mathrm{C}}^{\ell\ell}$ as well. Further precision was gained by a combination of both 7 TeV as well as 8 TeV ATLAS and CMS measurements [74]. The combination results as well as comparisons with the individual inclusive asymmetry measurements are shown in Fig. 2.13 and 2.15. A comparison of the SM predictions with the combined ATLAS and CMS differential measurement of A_{C} vs $m_{t\bar{t}}$ is shown in Fig. 2.14. None of the measurements reported any significant excess with respect to the SM predictions.

(6) The longitudinal boost of a $t\bar{t}$ pair is defined as $\beta_{\mathrm{z},t\bar{t}} = \frac{p_{z,t\bar{t}}}{E_{t\bar{t}}}$, where $p_{z,t\bar{t}}$ and $E_{t\bar{t}}$ are the longitudinal momentum and energy of the $t\bar{t}$ pair, respectively.

Finally, ATLAS has also performed A_C measurement in the single-lepton channel using boosted top-quark pairs [75], using dedicated techniques we will discuss in Ch. 5. The measurement was focused on the region of $m_{t\bar{t}} > 750\,\text{GeV}$, but suffered from both low statistics as well as non-negligible systematic uncertainties, as can be seen in Fig. 2.15.

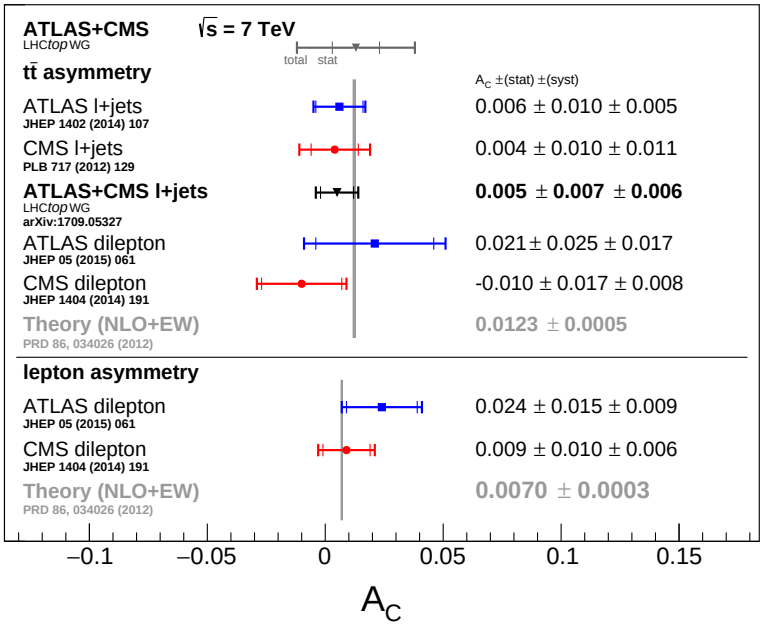


Fig. 2.13: Summary of the inclusive $A_C^{t\bar{t}}$ and $A_C^{\ell\ell}$ measurements at the LHC [74] using Run-I dataset at $\sqrt{s} = 7\,\text{TeV}$. Both standalone ATLAS [65, 66] and CMS [67, 68] measurements as well as their combination [74] is shown. The inner bars on the measurement points indicate the statistical uncertainty and the outer bars the total uncertainty. The theoretical prediction [58] with its uncertainty is shown by the grey bar.

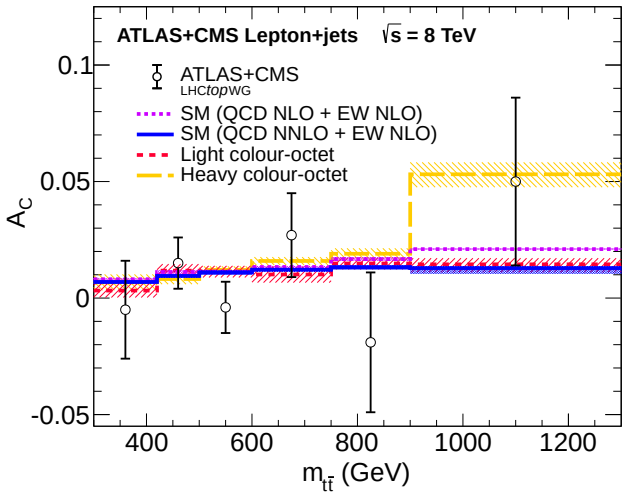


Fig. 2.14: Combination of the ATLAS and CMS results of the differential $A_C^{t\bar{t}}$ measurement as a function of $m_{t\bar{t}}$ using Run-I data at $\sqrt{s} = 8\,\text{TeV}$ [74]. The measurement is compared to the SM predictions [58, 60] and to the example predictions for light and heavy colour-octet models [76]. The bands around the BSM predictions show the statistical uncertainty of the simulations. The bands around the SM predictions are uncertainties dominated by scale variations.

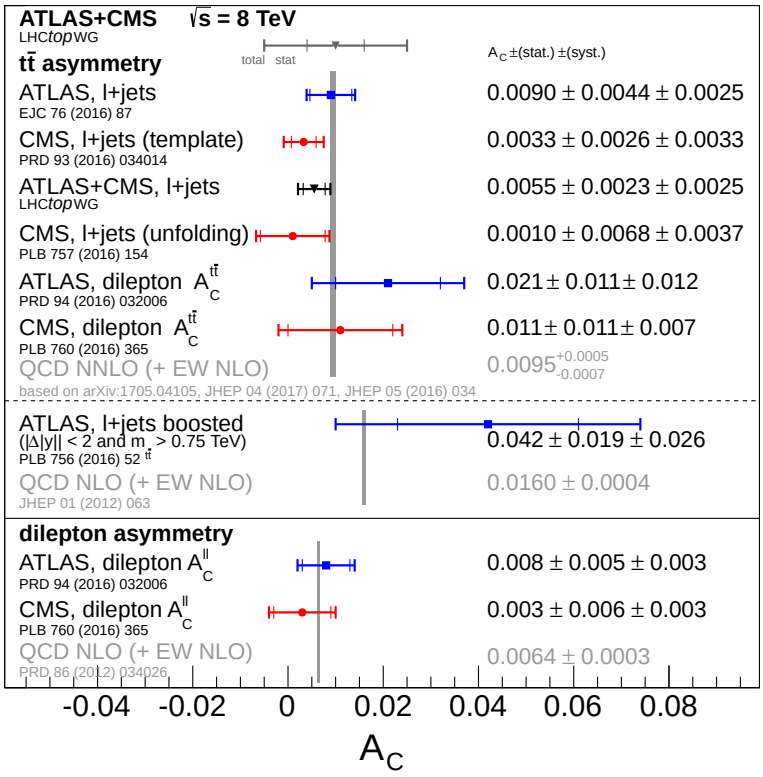


Fig. 2.15: Summary of the inclusive $A_C^{t\bar{t}}$ and $A_C^{\ell\ell}$ measurements at the LHC [74] using Run-I dataset at $\sqrt{s} = 8\,\text{TeV}$. Both standalone ATLAS [69, 70, 75] and CMS [71–73] measurements as well as their combinations [74] are presented. The SM prediction for $A_C^{t\bar{t}}$ features NNLO QCD + NLO EW calculation [59, 60, 77], while the $A_C^{\ell\ell}$ SM prediction is calculated up to NLO QCD + NLO EW [58]. A dedicated ATLAS measurement using boosted top quark identification techniques [75] is also shown (third measurement from the bottom) and compared to SM NLO QCD + NLO EW prediction [44]. The error bars follow the same convention as in Fig. 2.13.

2.6 BSM interpretations of the charge asymmetry measurements

The tension between the SM prediction and the initial CDF and D0 measurements prompted studies of potential beyond-SM (BSM) theories that could enhance A_{FB}. The higher energy of the LHC also allowed to explore more extreme regions of phase space to probe for BSM contributions.

The possible BSM contributions typically induce additional charge asymmetry contributions in $q\bar{q} \to t\bar{t}$ processes via exchange of a newly proposed particle, through interference terms of BSM and SM amplitudes or via BSM amplitude square module [78]. Higher-order contributions are typically neglected. From a theoretical perspective, the models must obey the SM $SU(3)_{\text{C}} \otimes SU(2)_{\text{L}} \otimes U(1)_{\text{Y}}$ symmetries, and be renormalisable. The models introduce new spin-0 or spin-1 particles. These requirements already limit the number of possible models to ten irreducible vector-boson representations and eight scalar-boson representations [78]. Additionally, the most prospective models typically have cancellations between interference terms of BSM-SM amplitudes and the LO BSM squared amplitudes. This reduces their impact on other differential distributions, which otherwise lead to tensions in other measurements.

The most recent BSM interpretations were performed using the Tevatron [53] combination and the

LHC 8 TeV [74] combination, as shown in Fig. 2.16 and in Fig. 2.14. Since no significant excess with respect to SM was observed, predictions for various BSM models were calculated, comparing them with the 95 % CL limits derived from the measurements.

In addition to interpretations in terms of concrete BSM models, searches for BSM physics can be also performed in a less model-dependent manner via the *effective field theory* (EFT) approach.

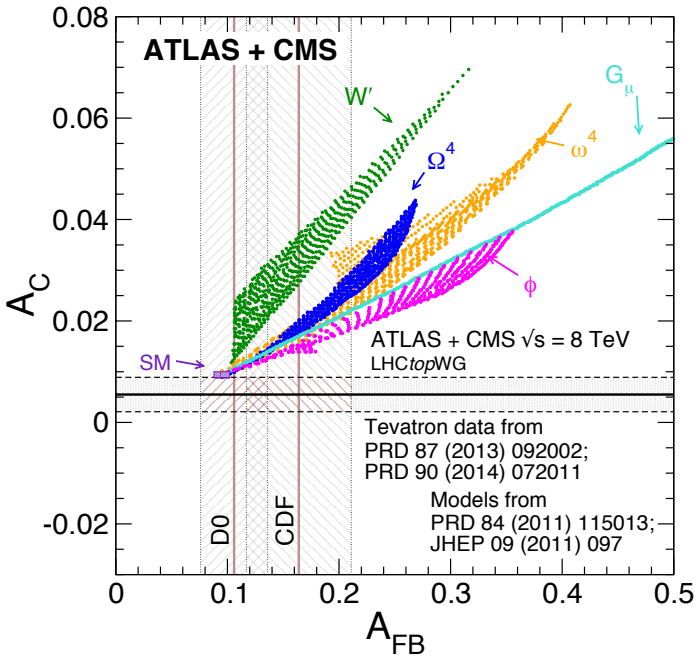


Fig. 2.16: Comparison of measured A_C at the LHC [74] and A_{FB} at the Tevatron [51, 52] with SM prediction at NNLO QCD + NLO EW [59, 60, 77] and various BSM predictions. The BSM predictions include W', heavy axigluons (G_μ), scalar isodoublet (φ), colour triplet scalar (ω^4) and a colour sextet scalar (Ω^4) [79, 80].

2.7 EFT interpretations of the charge asymmetry measurements

The EFT approach to BSM interpretations lies in the expansion of the SM Lagrangian by adding extra dimension > 4 operators that express additional corrections to vertices in process amplitudes. The lowest-order expansion includes dimension-six operators[(7)]:

$$\mathcal{L} = \mathcal{L}_{\text{SM}} + \sum_i \frac{C^i}{\Lambda^2} O_i + O(\Lambda^{-4}) \tag{2.11}$$

where O_i are the extra EFT operators, C_i their corresponding dimensionless Wilson coefficients and Λ is the scale of the new physics [82]. The operators are composed from the SM fields and are required to obey the SM $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ gauge symmetry. This approach assumes that new potential BSM particles may be extremely heavy, and thus their direct discovery is out of reach of the LHC, i.e.

[(7)] There are also odd-dimension operators, the lowest order being dimension-five, however these operators, if constructed only from SM fields, induce baryon or lepton number violation [81]. The baryon and lepton numbers are assumed conserved in the context of charge asymmetry EFT interpretations.

the energy scale Λ is much greater than energy scale probed at the LHC. Instead, the effects of these BSM contributions are manifested as corrections to SM amplitudes. Instead of constraining parameters of concrete BSM models, in the EFT approach, constraints on C^i/Λ are obtained.

For the $t\bar{t}$ charge asymmetry only a subset of the operators are relevant, that contribute to the $q\bar{q} \to t\bar{t}$ process. A visualisation of these corrections is shown in Fig. 2.17 which shows how the contributing operators affect amplitudes of this process at lowest order. The possible operators either impact the coupling of gluon to $t\bar{t}$ or they impact the whole $q\bar{q} \to t\bar{t}$ amplitude via four-fermion contributions. These can be further reduced into four operators [83] represented by four Wilson coefficients C_u^1, C_u^2, C_d^1 and C_d^2. The indices u and d indicate different contribution from coupling to either u or d quarks in the initial state[8]. In general, potential BSM physics could couple differently to u and d quarks, but the recent EFT interpretations of the charge asymmetry measurements typically assume same coupling to both u and d quarks, reducing the four operators into just two: $C_1 = C_u^1 = C_d^1$ and $C_2 = C_u^2 = C_d^2$ [84].

In this simplified two-operator basis, the $t\bar{t}$ cross-section is sensitive to BSM physics represented by the $C^1 + C^2$ coefficient combination, and the charge asymmetry to the $C^1 - C^2$ coefficient combination. Constraints on these operator combinations have been performed using inclusive $t\bar{t}$ cross-section measurements and charge asymmetry measurements from combinations of CDF and D0 measurements and combinations of ATLAS and CMS 8 TeV measurements[9] [84]. The resulting constraints of the coefficients are shown in Fig 2.18. The constraints due to cross-section from Tevatron measurements is much stronger than LHC 8 TeV measurements despite lower precision, because the $t\bar{t}$ production is dominated via gluon fusion, and the C^1, C^2 coefficients are only sensitive to $q\bar{q} \to t\bar{t}$.

Despite the gluon fusion dilution, it can be expected that the 13 TeV charge asymmetry will significantly improve the limits on the Wilson coefficients due to the significantly increased $t\bar{t}$ cross-section and the much larger dataset. Further improvements can be achieved by probing regions of phase space with enhancement of the $q\bar{q} \to t\bar{t}$, such as by probing higher $m_{t\bar{t}}$.

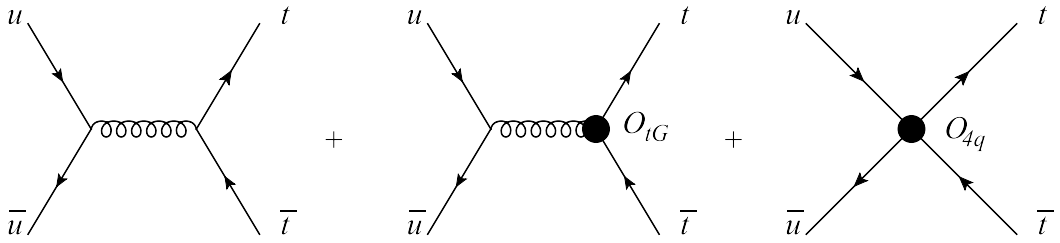


Fig. 2.17: EFT contributions to $u\bar{u} \to t\bar{t}$ [83]. The diagram (a) shows LO SM amplitude, the diagram (b) represents the correction to $gt\bar{t}$ vertex coupling and diagram (c) shows additional contribution via four-fermion interactions. The same diagrams for $d\bar{d} \to t\bar{t}$ are considered in the EFT contributions.

[8] Other quark types in the initial state of $q\bar{q} \to t\bar{t}$ are suppressed due to proton PDFs.

[9] There is no EFT interpretation of the combined ATLAS+CMS charge asymmetry yet, so the ATLAS and CMS results are interpreted standalone.

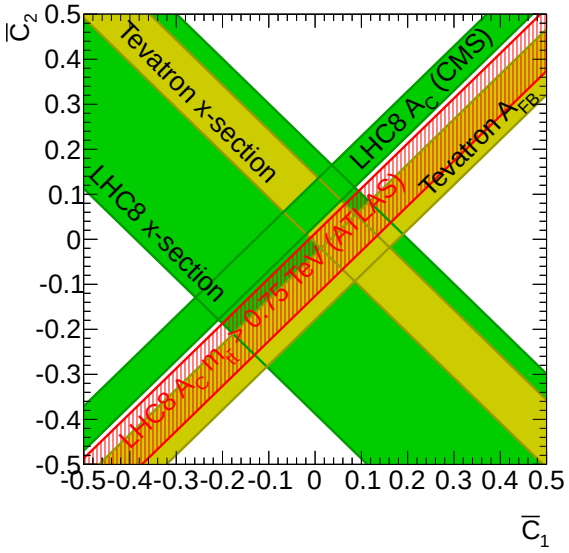


Fig. 2.18: The 68 % confidence level constraints on the linear combinations $\overline{C}^1 + \overline{C}^2$ and $\overline{C}^1 - \overline{C}^2$ of the Wilson coefficients from inclusive $t\bar{t}$ cross-section measurements and charge asymmetry measurements done at the Tevatron and the LHC at $\sqrt{s} = 8$ TeV [84]. The $\overline{C}^i = C^i \times v^2/\Lambda^2$ are normalised coefficients, where v is the SM Higgs vacuum expectation value.

3

The LHC and the ATLAS experiment

3.1 The Large Hadron Collider

The *Large Hadron Collider* (LHC) [85] is the largest hadron synchrotron accelerator to date, located at the *European Organisation for Nuclear Research* (CERN) near Geneva. Its circumference is roughly 27 km and it is located approximately 100 m underground. It was designed to collide protons at $\sqrt{s} = 14$ TeV and heavy ions with $\sqrt{s} = 2.8$ TeV per nucleon. The LHC has been operating in several periods with different setup and collision energy since 2010, with several planned technical stops for upgrades and maintenance. Currently, the LHC is undergoing upgrades and maintenance in preparation for the Run-III period of collisions, expected to begin in 2021.

3.1.1 The LHC accelerator complex

The LHC itself is a final collider in a series of linear and circular colliders. The overview of the accelerator complex is showed in Fig. 3.1. First, the protons are obtained by stripping electrons from Hydrogen atoms and injected in the first, linear accelerator (LINAC2) and are accelerated to 50 MeV. The protons then enter a series of circular colliders, each accelerating the particles up to a specific energy, the *Proton Synchrotron Booster* (PSB) to 1.4 GeV, the *Proton Synchrotron* (PS) to 26 GeV, and the *Super Proton Synchrotron* (SPS) to 450 GeV. The SPS is the final accelerator before the LHC, with a circumference of 7 km. Half of the protons from SPS enter LHC in one direction and the other half in opposite direction. The oppositely-moving protons are accelerated in separate acceleration tubes. Both proton beams are grouped into *bunches*, where each bunch contains the order of 1×10^{11} protons. The bunches are accelerated using radio-frequency (RF) cavities in a single region of the accelerator. The proton beam trajectory within the accelerator is contained by super-conducting niobium-titanium magnets operating at the temperature of 1.9 K, capable of generating magnetic field of 8.3 T. Dipole magnets are used to bend the trajectory within the LHC, and additional quadrupole magnets are used to focus the beam. Further corrections to the beam trajectory are achieved by higher multipole magnet systems.

There are four collision points along the LHC, where caverns with four detectors are installed. In these spots, the beam pipes are connected and the opposite beams are squeezed using focusing magnets and crossed using deflecting magnets to produce the collisions. The four detectors correspond to four main experiments using collisions from the LHC. Eth ATLAS (*A Toroidal LHC AparatuS*) and CMS (*Compact Muon Solenoid*) are the two largest, general-purpose detectors for testing SM predictions as well as searching for BSM physics. LHCb (*LHC beauty*) is a specialised experiment using a forward detector that focuses on physics of B-hadrons to study CP-violating processes. ALICE (*A Large Ion*

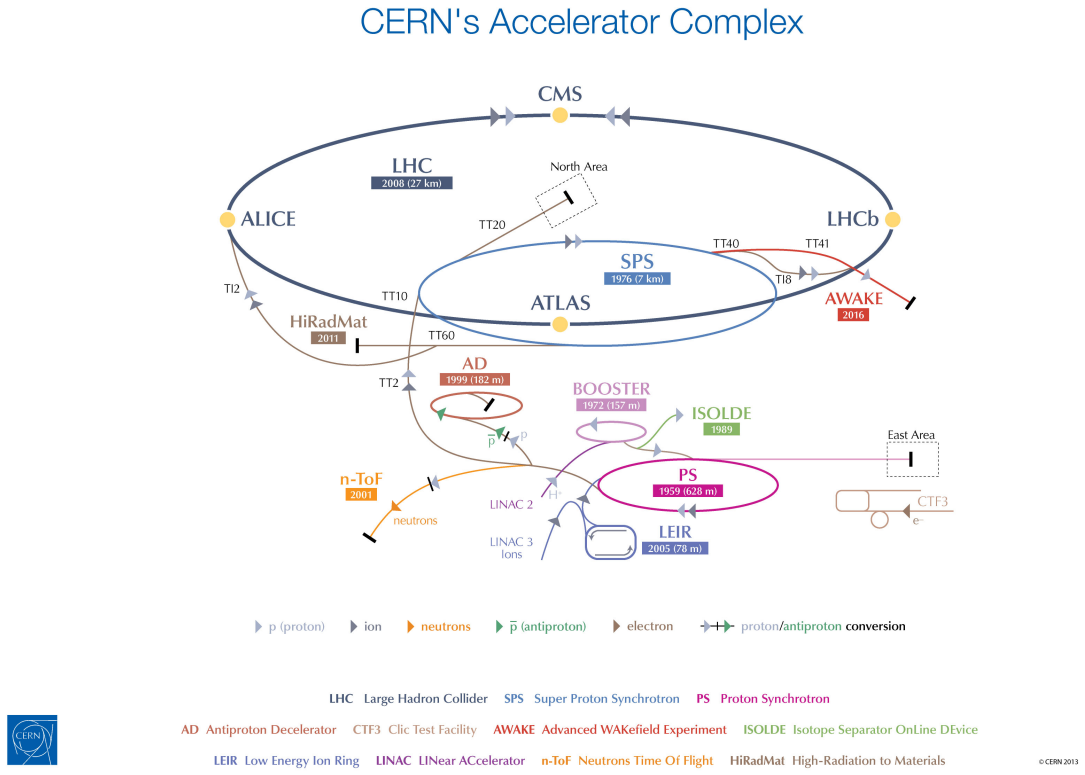


Fig. 3.1: The accelerator complex at CERN [86].

Collider Experiment) is another specialised experiment focused on studying quark-gluon plasma and hadronization processes in heavy ion (lead) collisions.

Other than unprecedented collision energy of the LHC, the other very important characteristic of a collider is its instantaneous luminosity, a quantity that describes the number of collisions per unit area and unit time:

$$\mathcal{L} = f \frac{N_c n_1 n_2}{A}. \tag{3.1}$$

The luminosity depends on the revolution frequency f of the proton beam, the number of bunches in the beam N_c, the number of particles per bunch in the colliding bunches n_1, n_2 and the overlapping area of the bunches A. The product of integrated luminosity over the period of collisions and the cross-section of a particular process gives us the prediction of how many times that process occurred during the time the LHC was colliding the beams. The operation of the LHC was divided into several periods of data-taking by the experiments.

In 2011, the LHC operated at $\sqrt{s} = 7\,\text{TeV}$ and both ATLAS and CMS experiments recorded approximately $5\,\text{fb}^{-1}$ of data. In 2012, the operation of LHC was restarted at $\sqrt{s} = 8\,\text{TeV}$, with ATLAS and CMS collecting approximately $20\,\text{fb}^{-1}$. These data taking periods concluded the Run-I period of data taking.

In summer of 2015, the LHC resumed operation after major upgrades for the Run-II period of data taking, achieving $\sqrt{s} = 13\,\text{TeV}$. During years 2015 to 2018, the Run-II dataset recorded by ATLAS and CMS reached almost $140\,\text{fb}^{-1}$. The large size of the dataset was achieved thanks to the outstanding instantaneous luminosity of the LHC that peaked during 2017 and 2018 data at over $2 \times 10^{-34}\,\text{cm}^{-2}\,\text{s}^{-1}$,

surpassing the original design by a factor of two. A visualisation of the cumulation of the integrated luminosity is shown in Fig. 3.2.

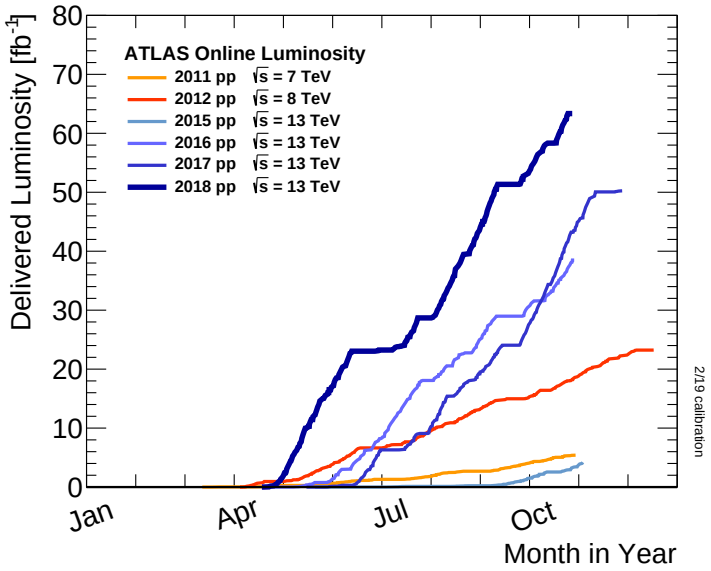


Fig. 3.2: Visualisation of the cumulative luminosity vs the day delivered to ATLAS during stable beams for *pp* collisions [87].

The large luminosity reached by LHC means that a single bunch crossing leads to multiple *pp* interactions, characterised by distribution of mean number of interactions per bunch crossing $\langle\mu\rangle$, shown in Fig. 3.3. This posses additional challenges in the reconstruction of the *pp* collisions due to in-time and out-of-time pile-up. In-time pile-up leads to signals in the ATLAS detector from multiple *pp* collisions in a single bunch crossing. Out-of-time pile-up is caused by limited read-out time of certain detector systems, which can be higher than spacing between individual bunches (25 ns) in the LHC. This means that slower detector systems can produce signals originating from more than one bunch crossing.

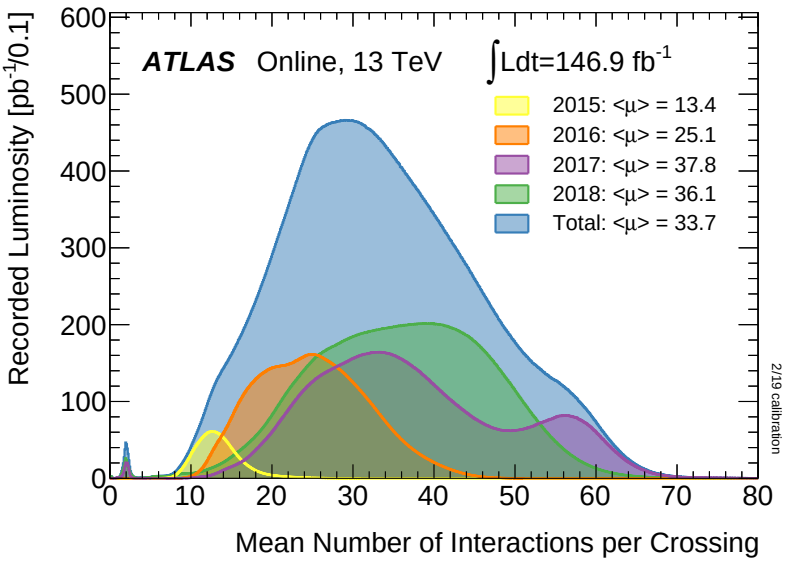


Fig. 3.3: The distribution of mean number of interactions per bunch crossing $\langle\mu\rangle$ for the individual data-taking periods in Run-II [87].

3.2 The ATLAS detector

The ATLAS detector is a cylindrical-shaped multi-purpose detector, with length of 44 m, diameter 25 m and weight of approximately 7000 tons. A longitudinal view of the detector is shown in Fig. 3.4. The purpose of ATLAS detector is to measure properties of particles. Due to the variability of particles produced in pp collisions and the manner of their interactions, ATLAS is composed of a system of different sub-detectors, a magnet system and a trigger system. There are three main sub-detectors: the inner detector, the calorimeter and the muon spectrometer. The ATLAS coordinate system and a brief description of the individual detectors is provided in the following sections based on Ref. [88].

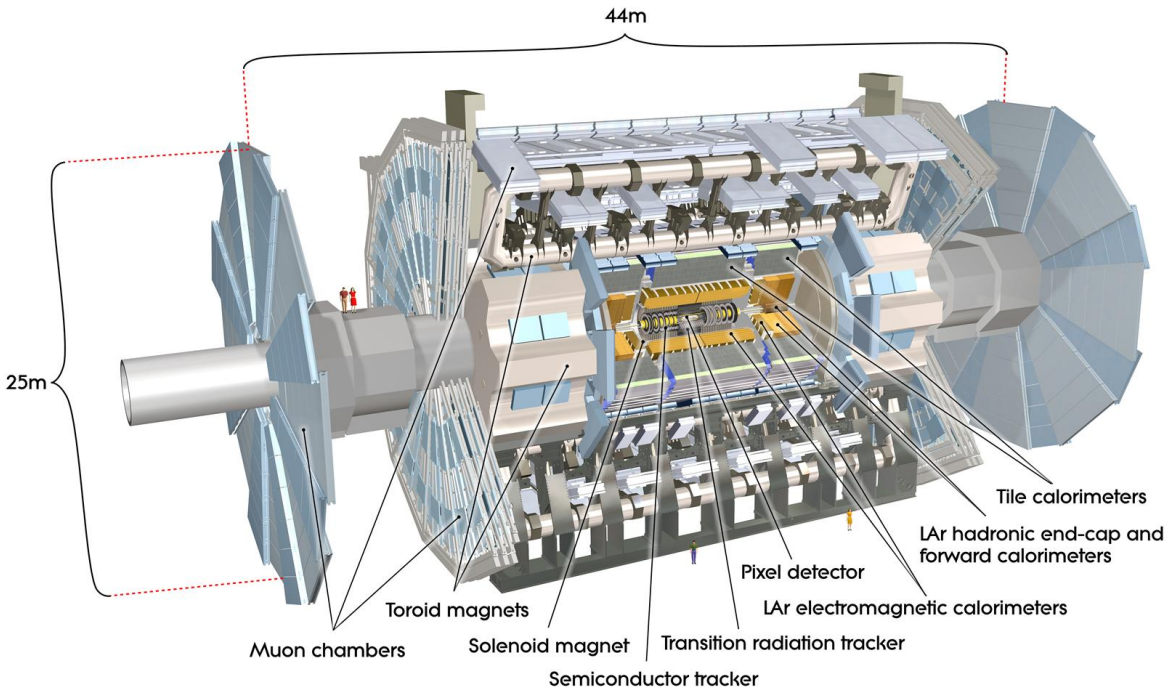


Fig. 3.4: Longitudinal view of the ATLAS detector [88].

3.2.1 ATLAS coordinate system

We first introduce the coordinate system and some of the quantities typically used in high-energy collider physics. ATLAS uses a right-handed coordinate system, where the z-axis points in the direction of the beam pipe. The $x-y$ plane is perpendicular to the beam pipe, with x coordinate pointing towards the centre of the LHC and y points upwards.

A spherical coordinate system is introduced, where ϕ is the azimuthal angle in the $x-y$ plane with respect to the x-axis direction. The direction with respect to the z-axis is denoted by angle θ. It is customary to use pseudorapidity η instead of θ for relativistic particles, due to $\Delta\eta$ being Lorentz-invariant under boosts along the z-axis. Pseudorapidity is defined as follows:

$$\eta = -\log\left(\tan\left(\frac{\theta}{2}\right)\right). \tag{3.2}$$

Distances between particles or other reconstructed objects are characterised via distance metric

$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$. This quantity is also Lorentz-invariant under z-axis boosts.

Due to the random nature of the total longitudinal momentum in the pp collision, it is typical to refer to *transverse* quantities, such as the *transverse momentum* p_{T}, defined as:

$$p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2} = |\vec{p}| \sin\theta \tag{3.3}$$

The sum of transverse momenta of all particles produced in a pp collision is expected to be zero due to momentum conservation and the interacting partons $p_{\mathrm{T}} \approx 0$.

We will now briefly describe the individual detector subsystems of the ATLAS and its magnet and trigger systems.

3.2.2 The inner detector

The inner detector (ID) is the closest detector to the interaction point, consisting of several types of detectors: semiconductor pixel detector, micro-strip semiconductor tracker (SCT) and a straw tube detector, the *Transition Radiation Tracker* (TRT). The purpose of the inner detector is the reconstruction of charged particle tracks and vertices, measurement of particle momentum and electron identification. The layout of the inner detector is shown in Fig. 3.5. The ID is placed within a 2 T magnetic field to allow measurement of electric charge and momentum of particles. The pseudorapidity coverage of the ID is $|\eta| < 2.5$. The momentum resolution is

$$\frac{\sigma}{p_{\mathrm{T}}} = 0.05\% \times p_{\mathrm{T}} \oplus 1\%^{(1)}, \tag{3.4}$$

where p_{T} is in GeV.

The pixel and SCT detectors are silicon-based semiconductor detectors, which are designed for reconstruction of particle tracks and vertices. The tracks are reconstructed from *hits*, electron-hole pairs produced by the passage of the charged particles in the detector's sensitive volume, and the electric signal is processed by the read-out electronics. To reduce the amount of noise from thermal electron-hole pairs, and to remove the heat produced by additional leakage currents from radiation damage, the semiconductor detectors are cooled down by an evaporative system.

The pixel detector is the closest detector to the interaction point. It was composed of three layers of silicon pixel detectors and for Run-II an additional layer called IBL was added for the purpose of improving identification of secondary vertices from B-hadrons. The nominal pixel size is 50×400 μm^2, each having a separate readout. The pixel detector thus contains more than 80 million readout channels. 90 % of the channels have spatial resolution 50 μm in the $x - y$ plane and 400 μm in z-axis direction.

Surrounding the pixel detector is the SCT, a micro-strip detector with similar purpose and technology as the pixel detector, however it is composed of four layers of longer rectangular strips rather than pixels. Each strip measures 80 μm × 6 − 12cm. To resolve ambiguities, hits are reconstructed from energy deposits in two layers. The SCT contains over 6 million SCT readout channels. The spatial resolution of SCT is 16 μm in $x - y$ plane and 580 μm in z.

Surrounding the silicon tracking detectors is the TRT detector composed of straw-tube detectors. Charged particles passing through the TRT pass through environments with different refraction indices,

(1) The ⊕ operator designates sum of operands in quadrature.

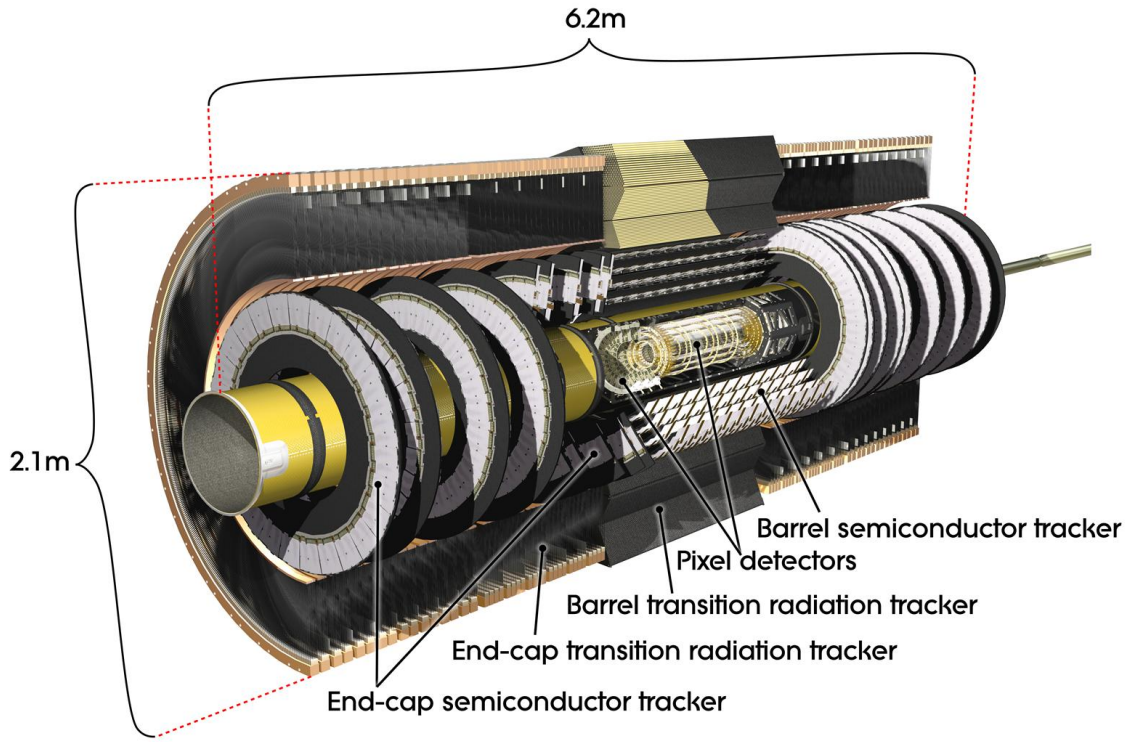


Fig. 3.5: The layout of the ATLAS inner detector. [89]

emitting transition-radiation photons. These photons are absorbed in a mixture of gasses inside the straw tubes. The amplitude of the signal depends on the Lorenz factor γ. This factor is typically very large for lightest particles like electrons and positrons, thus the signal from TRT is not only used for enhancing momentum resolution but also for electron/pion discrimination and fast triggering. In contrast to other components of the ID, the TRT covers region of only up to $|\eta| < 2.0$. The TRT consists of almost 30 000 straws, where each straw is 144 cm long (37 cm for the end caps), with a diameter of 4 mm. Each straw contains a 30 µm diameter gold-plated Wolfram-Rhenium wire that collects ionised electrons from the absorbed transition radiation and ionization produced by charged particles passing through the straw. The straws are filled with a gas mixture of 70% Xe, 27% CO_2 and 3% O_2. The electron collection time is 48 ns with a drift-time accuracy of 130 µm.

3.2.3 Calorimeters

In addition to momentum measurement for charged particles, the measurement of energy of both charged and neutral particles is necessary. The calorimeters in ATLAS are designed for this purpose. The calorimeter system covers large pseudorapidity range of $|\eta| < 4.9$ and is divided into electromagnetic (EM) calorimeter and a hadronic calorimeter that surrounds the EM calorimeter. The EM calorimeter has finer granularity than the hadronic calorimeter, allowing for high precision measurements of electron and photon energy. Both calorimeters are sampling-type, which means they contain layers of active medium sensitive to the energy of the particle, as well as passive absorbers. This design is a compromise between the size and cost of the calorimeters and the requirement to absorb the full energy of showers induced by the particles produced in pp collisions.

Two types of calorimeters are employed within ATLAS in terms of detection material, the liquid

Argon (LAr) calorimeter, which uses LAr as the active medium and lead absorber plates of an accordion shape. The particle-induced showers produced in the absorber ionize liquid argon atoms and the collected electrons produce a signal read out by electronics. The second type of calorimeter used in ATLAS is the Tile calorimeter, which uses alternating passive layers of iron and tiles of plastic scintillator as the active medium. The shower particles interacting with scintillator produce visible light which is converted in photomultiplier tubes into electric signal.

The EM calorimeter consists of the barrel calorimeter, which is a LAr calorimeter covering $|\eta| < 1.375$, and the end-cap LAr calorimeters covering $1.475 < |\eta| < 3.2$. The size of the EM calorimeter is such that it achieves absorbing capacity of approximately 22 radiation lengths[(2)]. The barrel EM calorimeter consists of three layers, where the first layer has increased granularity for discriminating isolated photons from non-prompt photons from hadron decays. Its granularity is $\Delta\eta \times \Delta\Phi = \frac{0.025}{8} \times 0.1$ for $|\eta| < 1.4$. The barrel and end-cap EM calorimeters have 101760 and 62208 readout channels, respectively.

The hadronic calorimeter consists of several parts surrounding the EM calorimeter, and is designed to measure the energy of jets produced by strongly-interacting particles. To absorb the showers, it provides the thickness of approximately 11 radiation lengths. The barrel and the extended barrel Tile calorimeters cover regions of $|\eta| < 1$ and $1 < |\eta| < 1.7$, respectively. The end-cap hadronic calorimeters use LAr technology with copper plate absorbers, and cover regions of $1.5 < |\eta| < 3.2$. The Tile calorimeter has 5769 readout channels in the long barrel and 4092 channels in the extended barrel and the granularity of 0.1×0.1.

In addition, the forward calorimeter covering $3.1 < |\eta| < 4.9$ is a LAr calorimeter made of three layers where the first layer is used as an EM calorimeter and the rest of the layers as hadronic calorimeter. The absorber layers are made of a single layer with copper and the rest of the layers made of tungsten. Both the forward and the end-cap calorimeters employ LAr as the active medium due to its larger radiation resistance, which is necessary due to the higher activity in forward regions of the detector.

Finally, the most important characteristic of a calorimeter is its energy resolution. For sampling calorimeters, it is typically parametrised as follows:

$$\frac{\sigma_E}{E} = \frac{S}{\sqrt{E}} \oplus \frac{N}{E} \oplus C, \tag{3.5}$$

where S, N and C are the stochastic, noise and constant terms, respectively, and E is in GeV. The stochastic term describes the fluctuations in the particle multiplicities of the shower in the active medium. The noise term describes the contributions of noise from the readout electronics, and is not energy-dependent. The constant term cumulates all of the systematic effects, which can range from imperfections in the detector structure or readout system, to radiation damage. The constant term is the dominant contribution to the energy resolution at high energies.

The EM calorimeter was designed to achieve $S = 10\,\%$, $N = 170\,\text{MeV}$, $C = 0.7\,\%$. The Tile calorimeter was designed to achieve $S = 50\,\%$ and $C = 3\,\%$, with a negligible noise term thanks to the precise photomultiplier readout. The end-cap and forward hadronic calorimeter achieve $S = 100\,\%$ and $C = 10\,\%$.

[(2)] The radiation length is the thickness of calorimeter that decreases the energy of the particle inducing the shower by a factor of e^{-1}.

3.2.4 Muon spectrometer

Due to their large mass compared to electrons, muons traverse calorimeters with minimum energy loss. A standalone system dedicated to their detection and measuring their momentum is installed, surrounding the calorimeters. The muon spectrometers are placed in a magnetic field generated by toroids in order to measure momentum based on the muon track curvature. They cover the region of $|\eta| < 2.7$. The momentum measurement is performed by combining information from both the muon detectors as well as the ID, achieving a combined momentum resolution of 2-3 %, except for very high momenta (> 1 TeV) where the resolution is 10 %. The muon spectrometer consists of several parts: the *cathode strip chambers* (CSC), the *monitored drift tube* (MDT) *chambers*, the *resistive plate chambers* (RPC) and the *thin gap chambers* (TGC) [90].

MDTs together with CSCs are used for precise measurement of particles momenta. MDT chambers cover range $|\eta| < 2$, and are composed of layers of drift tubes with a diameter of 3 cm. The drift tubes are proportional chambers filled with a mixture of Ar and CO_2 gasses and contain a central wire anode. The position of a crossing muon with respect to the anode is determined from the ionised electron cloud drift time. The average position resolution per tube is 80 µm and the maximum drift time is 500 ns.

CSCs cover the end-cap region of $2.0 < |\eta| < 2.7$ and are radially oriented in planes perpendicular to the beam pipe. CSCs are multi-wire proportional chambers with fast readout, with cathode planes segmented into strips in orthogonal directions. This enables CSC to measure both coordinates of charged particles and to cope with the larger particle multiplicities in the forward region. The resolution of chambers is 40 µm in the bending plane and 5 mm in the transverse plane.

RPCs and TGCs are used for fast trigger decisions due to their extremely fast operation. RPCs are very thin gas chambers with parallel Bakelite resistive plates, covering the barrel region of $|\eta| < 1.05$). Their fast gathering of ionised electron achieve readout time of 1.5 ns. The TGCs are multi-wire proportional chambers that cover the end-cap regions ($1.05 < |\eta| < 2.4$). Their fast operation is achieved by using a gas mixture with high quenching abilities, reducing the amount of ionised electrons that are collected by the anode, and thus achieving readout time < 25 ns. The fast readout of both RPCs and TGCs is also used to detect beam crossing.

3.2.5 Magnet systems

As we have mentioned, the magnet system forms a crucial part of the detector systems, for particle momenta and charge measurement. It consists of three parts: inner solenoid, barrel toroid, and end-cap toroids [91]. All of the magnet systems are based on super-conducting niobium-titanium cooled down by liquid Helium to a temperature of 4 K.

The *inner solenoid* surrounds the inner detector, creating mostly homogeneous 2 T magnetic field. The inner solenoid is very thin to absorb as little particle energy as possible, for that reason its width comprises only a 0.8-times radiation length. This is achieved partially by sharing the cryogenic system of the solenoid with the EM calorimeter.

The *barrel* and the *end-cap* air-filled toroids provide magnetic field for the muon spectrometers. Each of the toroids is built from 8 radially-arranged rectangular coils. The magnetic field varies in strength up to 2 T in the barrel region, and up to 4 T in the end-cap regions.

3.2.6 Triggers

The trigger system plays an important role in the data acquisition process, because the event rate at ATLAS is far beyond any technology available to process. The trigger system reduces the rate by picking events with interesting signatures. Compared to Run-I, the trigger system in Run-II underwent a significant upgrade to accommodate rates up to 5× higher due to the increased LHC luminosity [92]. The trigger systems consists of hardware-level L1 trigger and software high-level trigger (HLT).

L1 trigger is a hardware trigger that searches for high p_T muons, electrons, photons and jets. It combines the information from the fast muon spectrometer subsystems and reduced-granularity information from the calorimeter, while maintaining a latency of 2.5 µs. It reduces the data rate from approximately 40 MHz to around 100 kHz. In each event L1 also identifies one or more Regions-of-Interest (ROIs), which are coordinates in η and Φ with a potentially interesting activity. The ROI information is passed to the HLT. The L1 trigger does not attempt to use information from track reconstruction.

HLT is a software trigger using a large computer farm, which uses information from L1 trigger. In contrast to L1 trigger, the HLT also includes tracking information and all other detector information with full granularity within the ROIs, applying algorithms similar to the offline reconstruction of events. The HLT reduces the trigger rate to approximately 1 kHz with a latency of approximately 200 ms.

3.3 Simulations in the ATLAS experiment

The studies of various processes in the ATLAS experiment rely to a large extent on Monte Carlo (MC) simulations. These simulation largely follow the structure as follows.

1. **Simulation of the pp collisions.** Dedicated MC generators are used to simulate the various processes involved in the pp collisions. This includes simulation of the hard-scattering process, e.g. $pp \to t\bar{t}$ production with parton showering, hadronization and unstable particle decays. This simulation part requires as input the proton PDFs to simulate the relative contributions of various parton-parton interactions in the hard scattering. A simulation of the extra interactions from the proton remnants after a hard-scattering and multi-parton scattering is also performed. In addition, simulation of pp collisions that do not result in hard-scattering is necessary, e.g. elastic scattering. To simulate the multiple pp interactions per bunch crossing, multiple pp collisions are overlaid based on the measured probability distribution of the pp collisions multiplicity in the bunch crossing.

2. **Simulation of the detector response.** The particles produced in the previous simulation step propagate through the ATLAS detector, leaving signals in the various ATLAS detector sub-systems. The simulation of the detector response is performed using GEANT4 simulation software [93] using the ATLAS detector geometry [94]. The full simulation of the ATLAS detector is very demanding, in particular the simulation of the calorimeter response, and can reach order of 100 s/event. For this reason, fast simulation ATLFASTII is often used [94], which maintains the full simulation of the ID and muon systems, but uses a faster calorimeter simulation. The fast calorimeter simulation is based on replacing the simulation of individual particles in the showers from first principles with a simulation of single-particle showers using parametrisations

of longitudinal and lateral shower profile. This leads to a reduction of the simulation time per event by more than an order of magnitude [94].

3. **Digitization of the detector signals.** The simulated signals from the ATLAS detector are input to the simulation of the digitization, the output of which are voltages and currents of signals similar to those of the actual ATLAS detector.

The digitization output from simulations is processed using the same software that is used in the ATLAS trigger and data acquisition systems.

4

Object definitions

At the LHC, many particles are produced in the pp collisions. Most of these particles are unstable, and many of them have very short life-time, thus they decay before reaching the ATLAS detector. The particles with long-enough life-time to be reconstructed in the ATLAS detector include $e^{\pm}$, $\mu^{\pm}$, γ, $\pi^{\pm}$, $K^{\pm}$, K^{0}, $p^{\pm}$, n. Additionally, neutrinos produced in the pp collision also pass through the ATLAS detector, but in practice, without any interaction with the detector material.

In the pp collisions, we are interested in events where a hard scattering of partons from the protons occurs. The unstable particles produced in the hard-scattering are referred to as *parton level*, including also non-coloured particles such as leptons. The coloured particles form hadrons, at this stage we refer to *particle level*. Finally, after reconstructing those particles which are detected in the ATLAS detector, we refer to *reconstruction level* or *detector level*. The parton and particle level information is only accessible in dedicated *Monte Carlo* (MC) simulations, in the actual experiment only detector level information is available.

In this thesis, we perform studies involving $t\bar{t}$ pair production in the single-lepton and all-hadronic channel, which require reconstruction of electrons, muons, jets from quarks and neutrinos. Single-lepton $t\bar{t}$ decays are considered if there is an electron or a muon in the final state. These can include event where the leptonic decay produced a τ lepton, which decays before reaching the ATLAS detector, either hadronically or leptonically. Thus leptonic τ decays in $t\bar{t}$ production can contribute in the single-lepton channel.

In the following sections, the individual objects are described, how they are reconstructed, and what selection criteria are imposed to ensure efficient reconstruction with sufficiently low rate of background processes mimicking the objects of interest. For some of the objects, such as leptons, or jets multiple definitions are used, since different choices are made in the selection criteria in the various analyses included in this thesis. The selection criteria of the individual analyses within their respective chapters clarify the specific choices made.

4.1 Charged tracks

Charged tracks are reconstructed in the ID, their information being used in relation to other reconstructed objects in the ATLAS detector. A full description of the tracking reconstruction is provided in [95], nevertheless we illustrate briefly the method here. Firstly, clusters are reconstructed from hits in the pixel and SCT detectors. A single charge particle can produce hits in multiple adjacent cells, hence a dedicated algorithm [96] is used to correctly group these hits to produce *clusters* which correspond to the point in space where the charged particle passed through a layer of the ID (so called silicon space-point). Subsequently, track finding is performed to correctly assign clusters to the tracks. A

Kalman filter is seeded with cluster triplets satisfying basic criteria, iteratively adding more space-points to the track seeds in subsequent ID layers. The filter can produce multiple track candidates per seed. A fitting is performed to estimate five parameters of a track from the clusters of the track candidate: the charge-to-momentum ratio q/p, impact parameters d_0 and z_0, and the θ, ϕ angles (defined as in Sec. 3.2.1). The impact parameters d_0 and z_0 give the track transverse and longitudinal distances from the reconstructed primary vertex position, described in Sec. 4.1.1. To remove the ambiguities and poorly reconstructed tracks, all track candidates are assigned scores based on various criteria that penalize tracks with undesired properties. Track candidates crossing a sensitive part of the ID without producing a hit (a *hole*) or having a poor fit based on the χ^2 from the track fit are penalized. The p_{T} of the track candidates is also weighted in the score, favouring higher-p_{T} candidates. Out of the track candidates sharing a single cluster, the candidate with highest score is picked. Tracks sharing more clusters are removed. Finally, the main criteria all the tracks are required to fulfil, are the following:

- $p_{\mathrm{T}} > 400\,\mathrm{MeV}$
- $|\eta| < 2.5$
- Minimum of 7 hits,
- At most two holes in pixel and SCT detector combined,
- Not more than one hole in the pixel detector,
- Maximum one shared pixel cluster or two shared SCT clusters on the same layer.

4.1.1 Primary vertex reconstruction

Given the multiple pp interactions in a single bunch crossing, multiple interaction points are expected in the event. The vertices corresponding to these interaction points are reconstructed using dedicated algorithms using tracking information [97]. Collisions of interest must have at least one interaction vertex with at least two tracks with $p_{\mathrm{T,trk}} > 400\,\mathrm{MeV}$. The vertex with highest $\sum p^2_{\mathrm{T,trk}}$ is chosen as the primary vertex, where $p_{\mathrm{T,trk}}$ is the transverse momentum of tracks associated to the vertex.

4.2 Electrons

The electrons interact electromagnetically with the ATLAS detector, producing a track in the ID and depositing their energy by producing electromagnetic showers in the EM calorimeter. A dedicated energy clustering algorithm is used [98] which reconstructs variable-size clusters (so-called *super-clusters*) from energy deposits in the calorimeter cells. The clusters are matched to tracks in the ID. An electron candidate must have a single track matched to a super-cluster. The matched track must satisfy the impact-parameter criteria $|d_0|/\sigma(d_0) < 5$ and $|z_0 \sin\theta| < 0.5\,\mathrm{mm}$. The electron candidates are required to have $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $\eta_{\mathrm{cluster}} < 2.47$, excluding the region between the barrel and the end-cap calorimeter ($1.37 < |\eta_{\mathrm{cluster}}| < 1.52$). A multivariate algorithm is used to distinguish real electrons from other objects, such as photons and jets, referred to as *fake* electrons. The algorithm employs a likelihood-based discriminant d_L built from multiple variables encoding information about the electron candidate track and the EM shower shape [99]:

$$d_L = \frac{L_S}{L_S + L_B}, \qquad L_{S(B)}(\mathbf{x}) = \Pi_i^n P_{S(B),i}(x_i), \tag{4.1}$$

where the likelihood $L_{S(B)}$ is a product of probability density functions $P_{S(B),i}(x_i)$ of the individual discriminating variables x_i for the real (fake) electrons, estimated via MC simulation [98]. Several identification working points (WPs) are provided, differing in the value of the cut on the likelihood discriminant: *loose*, *medium*, *tight*. As the name suggests, the *tight* WP provides the smallest fake efficiency (*fake rate*) at the cost of reduced signal electron efficiency, starting from approximately 70 % at $p_{\mathrm{T}} = 25\,\mathrm{GeV}$ up to almost 90 % for $p_{\mathrm{T}} > 80\,\mathrm{GeV}$.

In order to suppress *non-prompt* electrons originating from decays of heavy hadrons and light hadrons misidentified as electrons, additional isolation criteria are imposed. Two quantities are defined to evaluate the amount of additional activity in the vicinity of the electron candidate:

$$E_{\mathrm{T}}^{R_{\mathrm{cut}}} = \sum_{\mathrm{cluster}\neq\ell}^{\Delta R(\mathrm{cluster},\ell)<R_{\mathrm{cut}}} E_{\mathrm{T}}^{\mathrm{cluster}} \tag{4.2}$$

$$p_{\mathrm{T}}^{\mathrm{var}R_{\mathrm{cut}}} = \sum_{\mathrm{track}\neq\ell}^{\Delta R(\mathrm{track},\ell)<R} p_{\mathrm{T}}^{\mathrm{track}}, \qquad \text{where } R = \min(10\,\mathrm{GeV}/p_{\mathrm{T}}^{\ell}, R_{\mathrm{cut}}). \tag{4.3}$$

The $E_{\mathrm{T}}^{R_{\mathrm{cut}}}$ defines the sum of transverse energies of the EM calorimeter clusters surrounding the candidate lepton cluster within a cone defined by radius R_{cut}. The $p_{\mathrm{T}}^{\mathrm{var}R_{\mathrm{cut}}}$ defines a similar quantity using transverse momenta of tracks surrounding the candidate lepton track within a p_{T}-dependent cone of maximum radius R_{cut}, that decreases with increasing lepton transverse momentum p_{T}^{ℓ}. Two different isolation definitions are used, the so-called *gradient* and *track-based fixed-cut* isolation.

The *gradient* isolation definition imposes an η-dependent cut on the $E_{\mathrm{T}}^{R_{\mathrm{cut}}}$ with $R_{\mathrm{cut}} = 0.2$ and $p_{\mathrm{T}}^{\mathrm{var}R_{\mathrm{cut}}}$ with $R_{\mathrm{cut}} = 0.3$. Both selections are adjusted to give constant real electron efficiency in η. The efficiency in p_{T}^{ℓ} starts from 90 % at $p_{\mathrm{T}}^{\ell} = 25\,\mathrm{GeV}$ up to 99 % at $p_{\mathrm{T}}^{\ell} = 60\,\mathrm{GeV}$ [98]. The *fixed-cut* isolation definition imposes the following cuts:

- $p_{\mathrm{T}}^{\mathrm{var}R_{\mathrm{cut}}}/p_T^{\ell} < 6\,\%$, $R_{\mathrm{cut}} = 0.2$,
- $E_{\mathrm{T}}^{R_{\mathrm{cut}}}/E_T^{\ell} < 6\,\%$, $R_{\mathrm{cut}} = 0.2$,

where p_T^{ℓ} and E_T^{ℓ} refer to the electron transverse momentum and transverse energy, respectively. The real electron efficiency is roughly 75 % at $p_{\mathrm{T}} = 25\,\mathrm{GeV}$ and reaches 99 % at 60 GeV [98].

The optimisation of the electron identification and isolation selection is performed via MC simulation. Because the simulation does not describe the detector response perfectly, comparisons of the predictions with data are performed in data using events with $Z \to e^+e^-$ and $J/\psi \to e^+e^-$ processes via *tag-and-probe* approach. This means that one of the electrons is selected using very tight criteria, ensuring high signal purity of the selected data. The other electron is then studied in both simulation and data. *Scale factors* are derived that evaluate the ratio between efficiencies measured in data and predicted efficiencies, for the reconstruction, identification, isolation and triggering [98]. These scale factors are then applied in the simulation to correct for the mismodelling. The same processes are used to perform calibration of the electron energy scale and resolution[(1)] [98, 100, 101].

[(1)] Unless, otherwise stated, the resolution of a quantity X is determined as the standard deviation of a distribution of $(X^{\mathrm{reco}} - X^{\mathrm{true}})/(X^{\mathrm{true}})$, where X^{reco} is the reconstructed value of the quantity and X^{true} is the true value of the quantity, determined from particle-level MC simulation, or from a well-measured reference object in case of in-situ methods.

4.3 Muons

Muons are reconstructed using tracks from both the ID and the muon spectrometer (MS), starting from the MS track and matching via extrapolation to an ID track. The ID and MS tracks are then refitted. Muon selection is performed by applying selection criteria to ensure robust momentum measurement and to suppress background processes mimicking prompt muons. These include mostly non-prompt muons from pion and kaon decays. The non-prompt muons have characteristic features in contrast to prompt muons, which may lead to poor fit quality of the MS and ID combined track, or incompatible momentum measurement from ID and MS. The *medium* WP selection criteria are imposed for muon identification [102]:

- ID track with at least one pixel hit, at least five SCT hits and less than three holes in the pixel and the SCT detector in total
- Combined MS and ID track satisfying quality cuts on the χ^2 of the global combined track fit, the relative difference in momentum measurement in ID and MS with respect to combined track momentum
- Combined track q/p significance less than seven
- The MS track must have at least three hits in at least two MDT layers, in $|\eta| < 0.1$, hits in a single layer is sufficient, with at most one hole
- Combined track $|\eta| < 2.5$ and $p_{\mathrm{T}}^{\ell} > 25\,\mathrm{GeV}$

In addition to the *medium* WP, *loose* WP is also used. The muon identification criteria for *loose* muons include muon tracks only reconstructed in the muon spectrometer, which are compatible with primary interaction point based on extrapolation of the track. Muon candidates with a track only in the ID are also included. These either satisfy criteria on the energy deposit in the calorimeter compatible with a minimally ionising particle, or the extrapolated track from ID matches a track segment in a single layer in MS [102].

Further suppression of non-prompt muons from hadron decays is achieved by imposing isolation criteria similar to those for electrons. The *gradient* and *fixed-cut* isolation definitions are used, which are defined using the same quantities $E_{\mathrm{T}}^{R_{\mathrm{cut}}}$ and $p_{\mathrm{T}}^{\mathrm{var}R_{\mathrm{cut}}}$ as for electron isolation. The *gradient* isolation uses η-dependent selection cut on $E_{\mathrm{T}}^{R_{\mathrm{cut}}}$ with $R_{\mathrm{cut}} = 0.2$ and $p_{\mathrm{T}}^{\mathrm{var}R_{\mathrm{cut}}}$ with $R_{\mathrm{cut}} = 0.3$. It achieves approximately 90 % (99 %) efficiency at $p_{\mathrm{T}}^{\ell} = 25\,\mathrm{GeV}(60\,\mathrm{GeV})$. The *fixed-cut* isolation for muons uses only track-based $p_{\mathrm{T}}^{\mathrm{var}R_{\mathrm{cut}}}$ isolation cut $p_{\mathrm{T}}^{\mathrm{var}R_{\mathrm{cut}}}/p_T^{\ell} < 6\,\%$ with $R_{\mathrm{cut}} = 0.3$. The achieved efficiency of this isolation is approximately 93 %(99 %) at $p_{\mathrm{T}}^{\ell} = 25\,\mathrm{GeV}(60\,\mathrm{GeV})$.

The calibration of the muon energy scale and resolution as well as measurement of the reconstruction, identification and isolation efficiencies, is performed in data using $Z \to \mu^+\mu^-$ and $J/\psi \to \mu^+\mu^-$ events using tag-and-probe approach [102] similar to that for the electron performance studies in data. The mismodelling of reconstruction, trigger, identification and isolation of muons is corrected by scale-factors derived using the measurements in data.

4.4 Jets

Strongly-interacting particles produced in the hadron-hadron collisions produce collimated showers of hadrons, jets. Jets are manifested by charged-particle tracks within the ID and showers induced by interacting hadrons in the ATLAS calorimeters. By reconstructing the jets, it is possible to infer

the properties of the quark/gluon that initiated the jet. This requires reconstructing the constituents which are combined into jets using a dedicated algorithm. In general, both calorimeter and tracking information can be used.

4.4.1 Topological calorimeter cell clustering

To construct jets from calorimeter information, the energy deposits in individual calorimeter cells are combined into topologically-connected clusters, referred to as *topo-clusters*. The cells of the topo-clusters are combined based on their signal-to-noise ratio [103], or cell significance $\zeta_{\text{cell}}^{\text{EM}}$, defined as:

$$\zeta_{\text{cell}}^{\text{EM}} = \frac{E_{\text{cell}}^{EM}}{\sigma_{\text{cell}}^{\text{noise}}} \tag{4.4}$$

where E_{cell}^{EM} is the cell energy measured at the electromagnetic (EM) scale, and $\sigma_{\text{cell}}^{\text{noise}}$ is the expected level of noise from electronics and pile-up. At the EM scale, the deposited energy in the calorimeter is inferred only from electromagnetic interactions of charged particles produced in hadronic showers in the absorbers. These charged particles either ionise the LAr or excite molecules in the plastic scintillators. The EM scale does not account for loss of signal due to the non-compensating character of the ATLAS calorimeters.

In the topo-clustering procedure [103], each cell with $\zeta_{\text{cell}}^{\text{EM}} > 4$, is iteratively connected with neighbouring cells with $\zeta_{\text{cell}}^{\text{EM}} > 2$, and neighbours with neighbours of $\zeta_{\text{cell}}^{\text{EM}} > 2$. This "proto-cluster growth" is terminated when only $2 > \zeta_{\text{cell}}^{\text{EM}} > 0$ neighbours are left, and these are included in the proto-cluster. The whole procedure is repeated, until no seed cells are left. After this procedure, topo-clusters including multiple local maxima are split up. The direction of a single topo-cluster in (η, ϕ) is calculated as barycentre weighted by the signal of the constituting cells. The total energy of a cluster is the sum of energy deposits in the constituting cells, accounting for cells shared by split clusters, where the cell energy is divided among the two clusters.

4.4.2 Sequential recombination jet algorithms

To reconstruct jets from the constituents, various algorithms have been developed in the past. Both ATLAS and CMS experiments employ sequential recombination algorithms. These algorithms iteratively combine the constituents reconstructed from the parton shower induced by a strongly-interacting particle into jets. The anti-k_t algorithm [104] is a default choice of a recombination algorithm used in ATLAS. It is defined as follows:

1. For each pair of constituents i,j a distance measure is calculated

$$d_{ij} = \min\left(p_{\text{T},i}^{k}, p_{\text{T},j}^{k}\right) \frac{\Delta R^2}{R^2}, \tag{4.5}$$

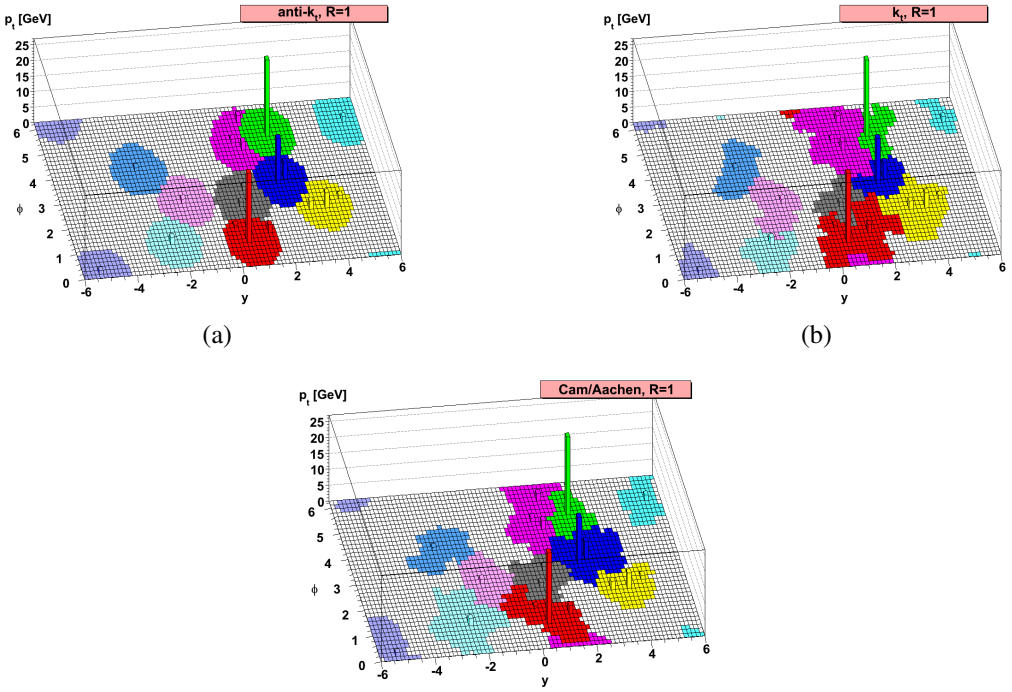
where ΔR is the distance between the constituents i,j in the $\eta - \phi$ space, the constant R is the radius parameter of the jet and exponent $k = -2$. For each constituent i a "distance to the beamline" is calculated

$$d_{i,\text{beam}} = p_{\text{T},i}^{k}. \tag{4.6}$$

2. If $d_{ij} < d_{i,\text{beam}}$, the constituents i,j are combined into a *proto-jet* which is again input in subsequent iterations of the algorithm. Otherwise consider i a jet and remove it from the list of constituents.

The procedure is repeated until all constituents and proto-jets are combined into jets. The four-vector of a jet is the sum of the four-vectors of its constituents.

Additionally to the anti-k_t algorithm, there are two other recombination algorithms, which both use the exact same sequential procedure, however a different power of k is used. The k_t algorithm [105] uses $k = 2$ and the Cambridge-Aachen algorithm [106] (C/A) uses $k = 0$, i.e. d_{ij} depends only on the angular separation ΔR. The choice of the distance metric affects the shape of the jets as illustrated in Fig. 4.1. The anti-k_t algorithm starts by clustering hard constituents first. In the busy environment of the LHC, this makes the algorithm less susceptible to soft particles from pile-up and underlying event[(2)]. Additionally, anti-k_t jets are typically cone-like, the R parameter can thus be reasonably well interpreted as a radius of the jet. In contrast the k_t jets lead to very irregular shape of jets. All of these factors make anti-k_t a more robust jet definition and easier to calibrate.



(a) (b) (c)

Fig. 4.1: Comparison of shapes of jets built by different recombination algorithms: anti-k_t (a), k_t (b) and Cambridge-Aachen (c) [104].

The aforementioned recombination algorithms are simple in definition, and robust computational efficient implementations exist [107]. They are also theoretically well-founded, since they are *infrared-safe* and *collinear-safe* [108]. This means that the jet clustering does not yield different results under a soft or collinear particle emission. This ensures proper treatment of infra-red and collinear divergences encountered in perturbative QCD.

[(2)] Underlying event includes any extra processes in the pp collision other than the hard-scattering of the interacting partons. This includes, e.g. additional multi-parton scattering in the same pp collision and other processes involving the proton remnants, inducing additional soft particle production.

Having described the possibilities for jet reconstruction, the following jet definitions are employed in analyses in this thesis and they are calibrated as described.

4.4.3 Small-radius jets

Anti-k_t R=0.4 (*small-R*) jets are used to reconstruct individual partons in events. These are build from EM-scale topo-clusters. A multi-step calibration procedure is used to correct the four-momentum of the jet. The calibration includes origin correction to point the jet to the primary vertex, energy correction for effects of pile-up [109] and an energy correction based on both MC simulation and in-situ calibration [110]. The MC-based calibration corrects the jet energy as well as direction to the particle-level energy. The in-situ calibration corrects for differences in jet response between data and MC simulation. First, the dijet η inter-calibration is applied to extend the jet calibration in the forward region ($0.8 < |\eta| < 2.5$) by comparing differences in momentum balance of a forward and central jet between MC and data. Subsequently, additional calibration in central region ($|\eta| < 0.8$) is performed by measuring momentum of jets recoiling against a well-measured reference object in processes such as ($Z \to \ell\ell$)+jet and γ+jet. The precise measurement of lepton and photon four-momenta is used to compare reference object and jet energy in both MC and data and correct the MC response to match data. Finally, a multijet-balance calibration is performed in events where a single high-p_{T} jet recoils against multiple lower-p_{T} jets, where the lower-p_{T} jets can be calibrated via the Z/γ+jets techniques. This allows to extend the in-situ calibration to higher energies ($\sim 2\,\mathrm{TeV}$) [110]. The uncertainty on the jet energy scale (JES) after the in-situ calibration is $\approx 3-4\,\%$ for $p_{\mathrm{T}} = 25\,\mathrm{GeV}$ down to 1 % at $p_{\mathrm{T}} < 2\,\mathrm{TeV}$.

The fully-calibrated small-R jets are required to have $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 2.5$. Additionally, jets with $p_{\mathrm{T}} < 60\,\mathrm{GeV}$ must be tagged by the *jet vertex tagger* (JVT) as jets not originating from pile-up. The JVT is a multivariate technique [111] that discriminates jets reconstructed from pile-up activity from hard-scattering jets. It is based on the fact that pile-up jets originate from a different vertex, and thus use tracking information and information about primary vertex to estimate whether a jet is originating from the hard-scattering or pile-up.

4.4.4 Large-radius jets

To study the full kinematics of unstable decaying particles, it is necessary to reconstruct their decay products four-momenta. For hadronic decays of particles such as the top quark, or W or Z boson for example, traditional approaches rely on the reconstruction of small-R jets and correctly matching them to the decay products that initiated them. The high energy of the LHC has enabled studies of these unstable particles at p_{T} much larger than their masses. In this kinematic regime, the decay products become collimated along the direction of the decaying particle. Their angular separation R in the $(\eta - \phi)$ space is:

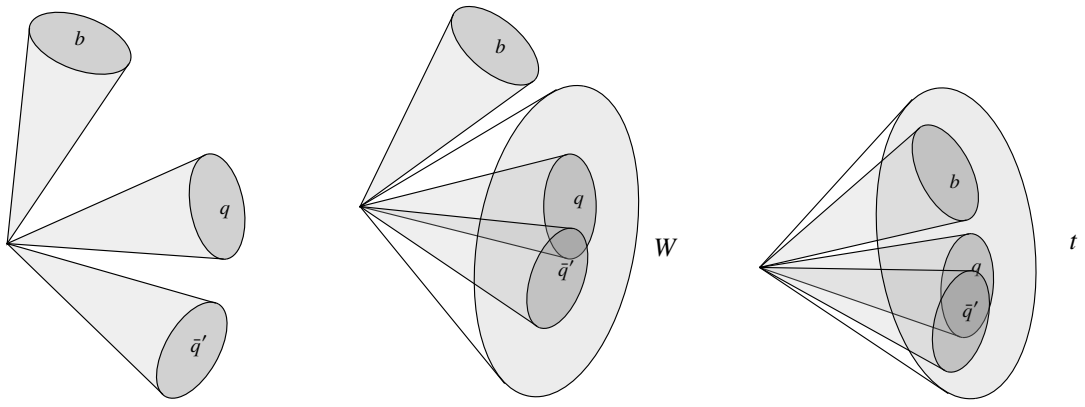
$$R \approx \frac{2m_X}{p_{\mathrm{T}}}, \tag{4.7}$$

with m_X being the mass of the decaying particle, and p_{T} its momentum. At sufficiently high p_{T} the showers from the decay products begin to overlap, reducing the efficiency of reconstruction approaches which rely on resolving the individual jets. An alternative approach developed during Run-I is used here, to reconstruct a single large-R jet that can contain the decay products (Fig. 4.2). In ATLAS,

dedicated optimisation studies [112] lead to the choice of R=1.0 jets. The challenge in the increase of the jet radius is that this increases sensitivity to pile-up and underlying event contamination. Therefore, techniques to reduce these undesired contributions have been developed [113], commonly referred to as *jet grooming*.

The grooming technique used commonly in ATLAS based on previous studies [112] is *trimming* [114]. The trimming procedure removes soft constituents of the jet that are more likely to originate from pile-up, rather than from the hard-scattering. The exact procedure is as follows, also illustrated in Fig. 4.3. Firstly, the constituents of the anti-k_t $R = 1.0$ jet are re-clustered into small sub-jets using k_t algorithm[(3)] with radius $R = 0.2$. Sub-jets failing the criterion of $p_{\mathrm{T,subj}}/p_{\mathrm{T}} > f_{\mathrm{cut}}$ are removed, where f_{cut} is a minimum fraction of transverse momentum that the sub-jet must carry. The chosen value of f_{cut} is 5 % [112]. Taking an ideal case example of a large-R jet containing hadronically-decaying top quark, after the trimming procedure, the jet should only contain three hard sub-jets corresponding to the decay products. To illustrate how grooming improves the stability of properties of jets with respect to pile-up, the effects of various grooming techniques on the invariant mass of the jet[(4)] and its pile-up resistivity are illustrated for large-R jets originating from W boson decays in Fig. 4.4.

The calibration of large-R jets is performed using a three step approach. Firstly, the clusters reconstructed at EM scale are individually calibrated to correct for non-compensating response of the calorimeter using MC simulation of single-hadron calorimeter response [103]. The large-R jets are subsequently built from this locally-calibrated clusters and groomed. Secondly, the jet energy, direction, and mass are corrected using MC-driven calibration to a particle-level scale. Finally, in-situ techniques similar to those for small-R jets are used to correct for differences in jet response between imperfect MC simulation and data [115]. The JES uncertainty after the in-situ calibration is $\approx 1 - 2$ % for 200 GeV $< p_{\mathrm{T}} <$ 2 TeV.



(a) Resolved topology (b) Semi-boosted topology (c) Boosted topology

Fig. 4.2: Illustration of common topologies of reconstructed top quarks. With increasing p_{T} of the large-R jet, it is possible to capture decay products from W boson from the top decay and reconstruct an isolated close-by b-jet, as shown in (b). At sufficiently high p_{T}, all of the top quark decay products are contained within the jet as shown in (c).

(3) The k_t algorithm is preferred for the sub-jet clustering, because it distributes the clusters to close-by sub-jets in a more balanced manner than the anti-k_t or the C/A algorithms. The other algorithms produce more sub-jets that are artificially soft in p_{T}, leading to undesirably aggressive trimming [114].

(4) The invariant mass of a jet is defined as $M = \sqrt{\left(\sum_i E_i\right)^2 - \left(\sum_i \vec{p}_i\right)^2}$, where E_i and $\vec{p}_i$ are the energy and momentum of individual constituents.

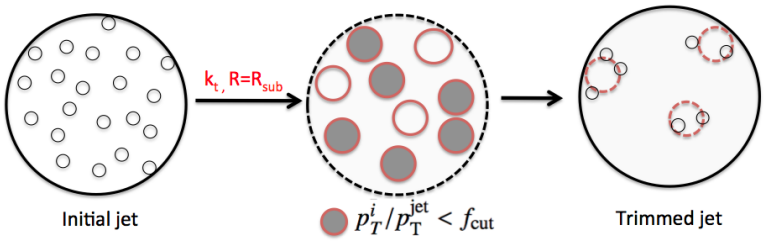
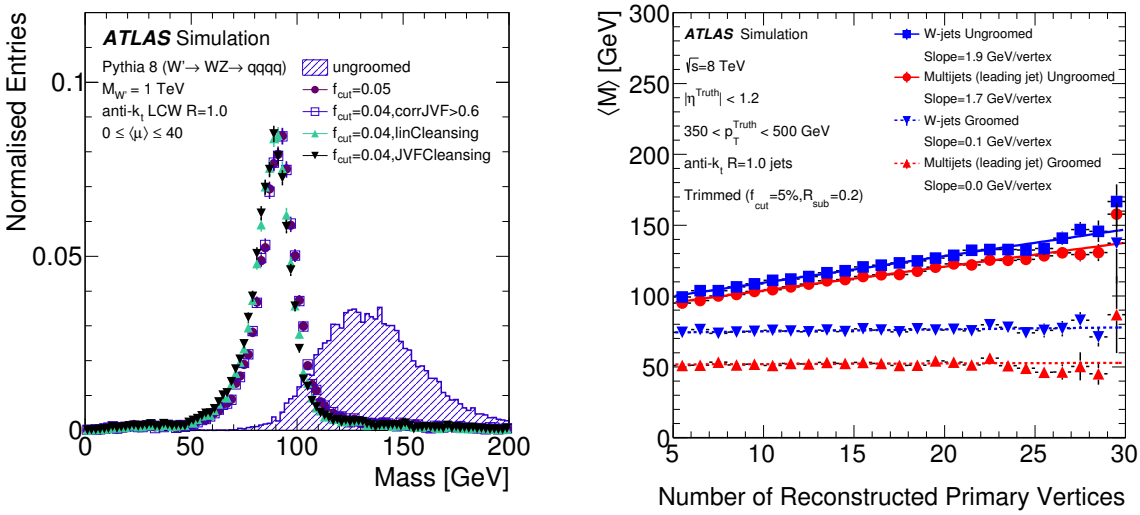


Fig. 4.3: Illustration of the trimming procedure [113].



(a) Comparison of invariant mass of ungroomed vs groomed jet from W boson. The groomed jet mass clearly shows a W peak, whereas the ungroomed jet mass peak scale and resolution are severely affected by pile-up.

(b) Comparison of dependence of mean mass $\langle M \rangle$ on the number of primary vertices in event for ungroomed and groomed jet. Groomed jets show relative insensitivity to pile-up.

Fig. 4.4: Illustration of properties of an ungroomed and a groomed jet mass [112].

The reconstruction of boosted hadronic decays of top quarks and other particles using large-R jets is mimicked by processes of multijet production. Therefore dedicated identification techniques, commonly known as *boosted tagging* have been developed to suppress the multijet background. They are discussed in Ch. 5.

4.4.5 Flavour tagging

The $t\bar{t}$ production leads to the presence of two jets initiated by b-quarks in all decay channels. Therefore their identification presents a very useful information to distinguish $t\bar{t}$ from background processes.

The bound states including b quarks (B hadrons) have relatively long life-time ($\approx 10^{-12}$ s [31]) despite the large b quark mass of ≈ 4 GeV thanks to the small $|V_{cb}|$ and $|V_{ub}|$ CKM matrix elements. Due to this, it is possible to reconstruct secondary vertex from the B hadron decay and discriminate b-jets from other flavour jets using information of the secondary vertex. A multivariate technique referred to as *MV2c10* [116, 117] based on a Boosted Decision Tree (BDT) is used to build a discriminant (Fig. 4.5) from a set of 21 input track-related variables from dedicated algorithms *IP3D* [116], *SV1* [116] and *JetFitter* [118]. The *IP3D* algorithm uses tracks matched to jet to calculate their impact parameter significance with respect to the primary vertex. The significance is larger for tracks from B-hadron

fragmentation. The *SV1* algorithm attempts to reconstruct a single displaced secondary vertex (referred to as *SV* vertex) and uses various kinematic parameters of the vertex such as track multiplicity, mass and energy fraction with respect to the whole jet to discriminate b-jets from light jets. Finally, the *JetFitter* is an algorithm based on Kalman filter [119] that attempts to reconstruct the whole decay chain of a B-hadron including the tertiary charm hadron vertex using similar quantities as the *SV1* to discriminate b-jets from light-flavour jets.

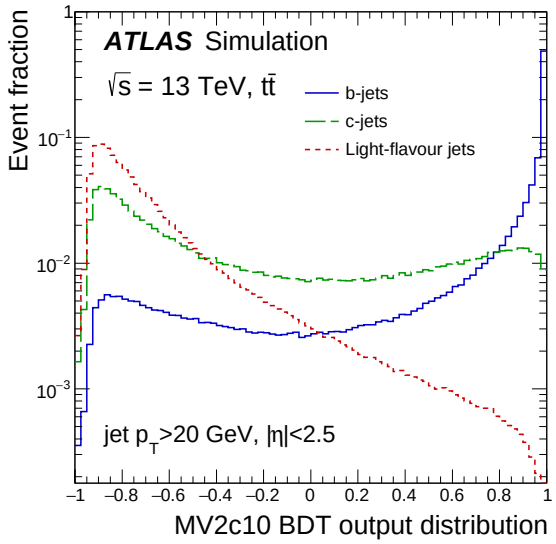


Fig. 4.5: The *MV2c10* discriminant for b-jets (blue solid line), c-jets (green dashed line) and light-flavour jets (red dotted line) in $t\bar{t}$ MC-simulated events [117]. The jets are built using the anti-k_t R=0.4 from topo-clusters calibrated at EM scale.

In this thesis, two approaches to b-tagging are used, depending on the jet collection used. In the first approach, tracks are matched to small-R calorimeter-based jets and the tracks are used by the *MV2c10* tagger to perform the b-tagging. The selection on the *MV2c10* classifier is performed, that inclusively achieves 77 % efficiency to select jet from b-hadron as b-tagged (signal efficiency) and suppresses light (charm) jets by a factor of approximately 100 (6).

The second approach used is to construct *track jets*, jets with radius R=0.2 built from ID tracks only, and use the track jet constituents as inputs for the *MV2c10* tagger. This strategy is typically employed to tag the B-hadron(s) inside a large-R jet used to reconstruct boosted decays of particles such as the top quark or Higgs boson decaying to $b\bar{b}$ pair. The small radius and use of tracks only leads to higher reconstruction efficiency due to smaller overlap of jets in the boosted topologies and good pile-up resistance. Due to differences in the performance of the b-tagging for calorimeter and track-based jets, a slightly different selection on the *MV2c10* classifier is performed, which results in an inclusive 70 % signal efficiency with similar background rejection compared to the calorimeter jet approach above. The b-tagging efficiency of track jets compared to calorimeter-based jets improves with increasing p_{T}. The track-jets are required to have $p_{\mathrm{T}} > 10\,\mathrm{GeV}$, $|\eta| < 2.5$ and at least two tracks.

4.5 Missing transverse momentum

The single-lepton and dilepton channels of $t\bar{t}$ production contain neutrino(s) in the final state. Given that the transverse momentum of the interacting partons in a pp collision is approximately zero, it

follows that the sum of transverse momenta of all particles produced in the collision must be zero. A non-zero total transverse momentum suggests the presence of an undetected particle, such as the neutrino. The *missing transverse momentum* is defined as follows:

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} = -\left(\sum_{i}^{\ell} \vec{p}_{\mathrm{T,i}} + \sum_{i}^{\gamma} \vec{p}_{\mathrm{T,i}} + \sum_{i}^{\mathrm{jets}} \vec{p}_{\mathrm{T,i}} + \sum_{i}^{\mathrm{soft}} \vec{p}_{\mathrm{T,i}} \right) \tag{4.8}$$

where the four sum terms represent the vectorial sum of transverse momenta in the $x - y$ plane. The first three terms encapsulate the transverse momentum contribution of tracks and calorimeter clusters matched to electrons, muons, taus [(5)], photons and the small-R jets. All of these objects are defined as outlined in previous sections. Finally, the *soft* term accounts for contribution of soft particles not assigned to any of the aforementioned reconstructed objects, and is calculated only using unmatched ID tracks. The tracking and vertexing information is used to distinguish contributions from pile-up [120].

The quantity upon which a selection criterion is typically imposed, is the *missing transverse energy*, $E_{\mathrm{T}}^{\mathrm{miss}}$, which is the magnitude of the missing transverse momentum vector. The only other piece of information is the angle ϕ of the missing transverse momentum direction in the $x - y$ plane. The longitudinal component of the missing transverse energy is undetermined due to the randomness of the momentum carried by the interacting partons in the pp collision.

The $E_{\mathrm{T}}^{\mathrm{miss}}$ scale and resolution in MC simulation is compared to data and calibrated, using $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ events [120].

4.6 Overlap removal of reconstructed objects

Due to the various ATLAS detector sub-systems being used for reconstruction of various particles, ambiguities can arise when the same particle-level object is reconstructed using multiple algorithms as multiple detector-level objects. To avoid double-counting in such cases, overlapping objects are removed in a step-by-step procedure as follows:

1. Any electron found with an ID track overlapping with another electron is removed.
2. Any electron sharing an ID track with muon is removed.
3. Any jet within distance $\Delta R < 0.2$ of an electron candidate is removed. If multiple jets overlap with an electron candidate based on this criterion, only the one closest in ΔR is removed.
4. Any electron subsequently found within $\Delta R < 0.4$ of a jet is removed.
5. If a jet is found within $\Delta R < 0.2$ of a muon, it is removed unless it has more than three associated ID tracks.
6. If a jet with less than three associated ID tracks has a muon ID track matched to it, the jet is removed.
7. Any muon subsequently found within $\Delta R < \Delta R_{\mu,\mathrm{jet}}$ of a jet is removed. Two strategies are employed in this thesis, differing in the choice of $\Delta R_{\mu,\mathrm{jet}}$:

[(5)] In case of $\tau \to \ell\nu_\ell\nu_\tau$, only the charged lepton is accounted for in the total transverse momentum sum.

(a) A fixed cut $\Delta R_{\mu,\text{jet}} = 0.4$ is used.

(b) A cut dependent on the muon transverse momentum p_{T}^{μ} is used: $\Delta R_{\mu,\text{jet}} = 0.04 + 10\,\text{GeV}/p_{\text{T}}^{\mu}$. This approach leads to a looser overlap removal for high-p_{T} muons, and is targeted to improve the event selection efficiency for boosted semileptonic top-quark decays $t \to \mu\nu_{\mu}b$.

5

Identification of boosted top quarks and *W* bosons

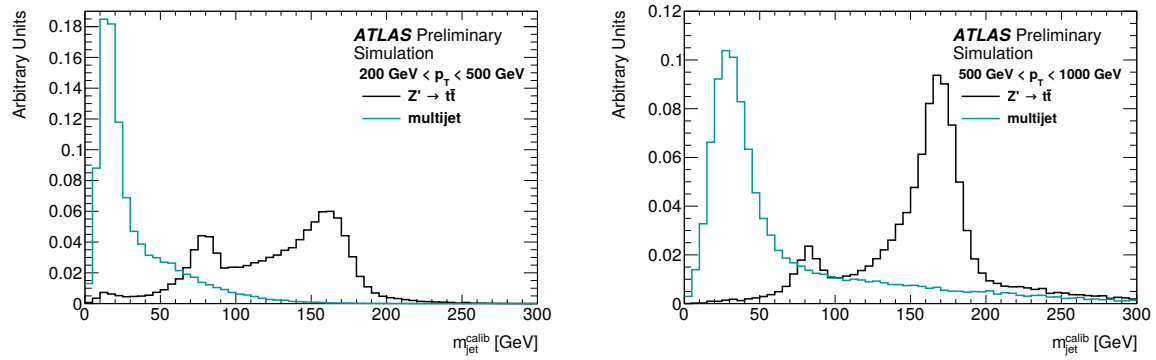
In this chapter, the techniques currently employed in the ATLAS experiment for identifying high-p_{T} hadronically-decaying top quarks and W bosons, commonly referred to as *boosted object tagging*, are introduced.

5.1 Substructure quantification using jet mass and *N*-subjettiness

The reconstruction of hadronically-decaying top quarks and W bosons as large-R jets is impacted by the multijet production, which is the dominant background process mimicking the signal particles' final state. Using groomed large-R jets, one can exploit substructure of jets to discriminate between signal and multijet background (light jets[(1)]). This is because jets from top-quark or W-boson decays have a different topology compared to light jets. Light jets are characterised by a single group of close-by clusters carrying the majority of jet energy, accompanied by soft, wide-angle emissions reconstructed as additional clusters. On the other hand, in simplistic terms, a jet from a hadronic top quark (W boson), will have three (two) groups of clusters where the jet energy is concentrated. These structures correspond to the quarks from the decay.

Different properties of the groomed light and top/W jets substructure above can be expressed by many different observables introduced by theorists, that discriminate between signal (top/W) and background (light) jets. One of the most illustrative variables is the jet invariant mass [112], shown in Fig. 5.1. Top-quark and W-boson jet mass distributions have peaks around the respective particle masses. The light-jet mass distribution is a steeply falling distribution with a p_{T}-dependant mass peak. As shown in Fig. 5.1a, the light-jet mass peak is well below the W-boson mass peak for low jet p_{T}. It is therefore possible to impose a cut on the mass (a minimum mass cut, or a mass window cut) to suppress a significant fraction of multijet production. However, with increasing light-jet p_{T}, as shown in Fig. 5.1b, the peak position shifts towards higher mass values and the tail of the distribution is also more pronounced. This means that the light jet rejection based on a mass cut becomes less effective with increasing jet p_{T}.

[(1)] The multijet background is the production of mostly light-quark and gluon jets. For simplicity, we refer to jets originating from this process as *light jets*.



(a) Jet $p_{\mathrm{T}} \in [200, 500]$ GeV (b) Jet $p_{\mathrm{T}} \in [500, 1000]$ GeV

Fig. 5.1: Comparison of the large-R jet mass distribution for light jets from multijet background and top-quark jets [121]. The mass peak around 80 GeV in (a) corresponds to jets containing decay products of the W boson from top quark. With increasing jet p_{T} the fraction of W-boson jets from top quarks decreases as shown in (b).

Nevertheless, there are many other variables quantifying the substructure properties to provide further discriminating power. An example how to quantify these substructure differences is illustrated using the *N-subjettiness* variable. This has been the canonical choice of an observable for top-quark tagging in both Run-I [122] and Run-II [121]. The N-subjettiness τ_N function [123] is defined in Eq. 5.1– 5.4, with an illustrative definition of the inputs to the τ_N calculation in Fig. 5.2.

$$\tau_0(\beta) = \sum_{i \in J} p_{T_i} \Delta R^{\beta}, \tag{5.1}$$

$$\tau_1(\beta) = \frac{1}{\tau_0(\beta)} \sum_{i \in J} p_{T_i} \Delta R^{\beta}_{a_1,i}, \tag{5.2}$$

$$\tau_2(\beta) = \frac{1}{\tau_0(\beta)} \sum_{i \in J} p_{T_i}\, min\left(\Delta R^{\beta}_{a_1,i}, \Delta R^{\beta}_{a_2,i}\right), \tag{5.3}$$

$$\vdots$$

$$\tau_N(\beta) = \frac{1}{\tau_0(\beta)} \sum_{i \in J} p_{T_i}\, min\left(\Delta R^{\beta}_{a_1,i}, \ldots, \Delta R^{\beta}_{a_N,i}\right). \tag{5.4}$$

The meaning of the equations can be understood as follows. For any $N \geq 1$, the large-R jet is re-clustered using exclusive k_t algorithm with N sub-jets[(2)]. The N-subjettiness is then a normalised[(3)] sum over all jet constituents, of the product of constituent p_{T} and its distance to the closest sub-jet axis, where the distance between the cluster i and a particular sub-jet axis k is given by $\Delta R_{a_k,i}$. The sum in Eq. 5.1– 5.4 runs over all of the large-R jet constituents. The exponent β in the ΔR distance terms changes the weight with which clusters away from the axes are penalised. For the τ_N definition used in ATLAS, $\beta = 1$, and the *winner-takes-all* (*wta*) sub-jet axis definition is used, as illustrated in Fig. 5.2, where the *wta* axis points in the direction of the hardest cluster momentum vector in a given sub-jet, instead of the reconstructed sub-jet axis. The τ_N function describes, how likely does the jet appear to contain N sub-jets. Jets which have their radiation aligned with direction of axes of

[(2)] The difference between the exclusive k_t and the standard k_t algorithm is that the exclusive k_t algorithm is terminated when a specified number of proto-jets are left during the clustering.

[(3)] The normalisation of τ_N ($N = 1 \ldots$) is given by $1/\tau_0$, defined in Eq. 5.1.

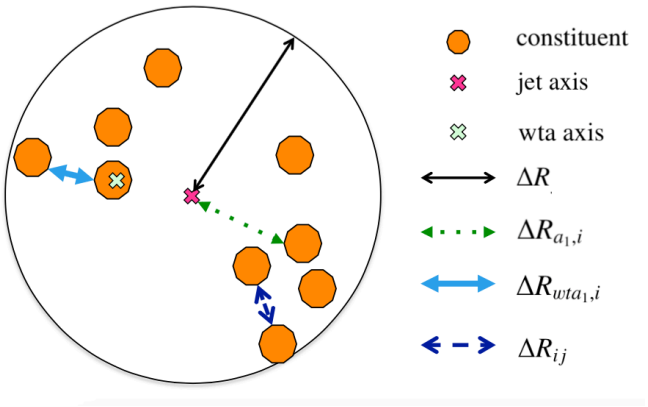
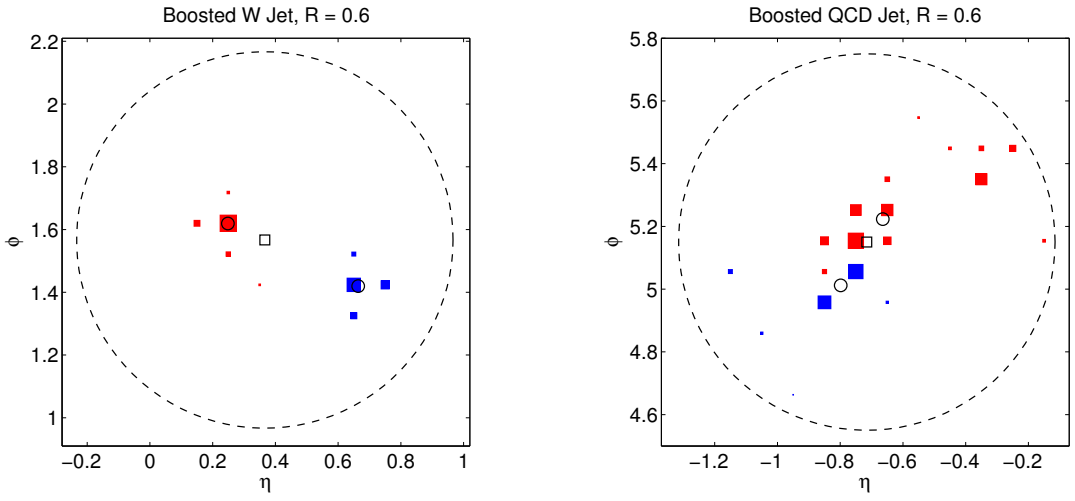


Fig. 5.2: Definitions of distances and axes in a single sub-jet for the N-subjettiness variables [112]. The *winner takes all* (*wta*) axis is the axis in the direction of the hardest cluster in the sub-jet.

individual sub-jets have $\tau_N \approx 0$ and thus have N or fewer sub-jets. Jets which have a large fraction of their energy distributed far away from the candidate sub-jet directions have $\tau_N \geq 0$ and therefore have at least N+1 sub-jets. This is illustrated on example distribution of τ_2 and τ_1 for W and light jets in Fig. 5.4b and 5.4a respectively. For a W jet, the constituents with highest p_{T} are close to axes of the two respective sub-jets (see Fig. 5.3a for illustration of the jet), and so the τ_2 variable yields a relatively small value, compared to τ_1, where many high-p_{T} constituents are far away from the single sub-jet axis. The same logic applies to light jets for τ_1 due to the fact that there is typically soft but wide-angle radiation. Note, that light jets can also have small τ_2 value. Therefore individual τ_N variables do not have such a strong discrimination power. However, for light jets, there is a correlation between high τ_1 and high τ_2 value, whereas W jets typically have high τ_1 regardless of value τ_2, as shown in Fig. 5.4c. As a consequence, ratio $\tau_{21} = \tau_2/\tau_1$ is commonly used, because this ratio for light jets has a tendency

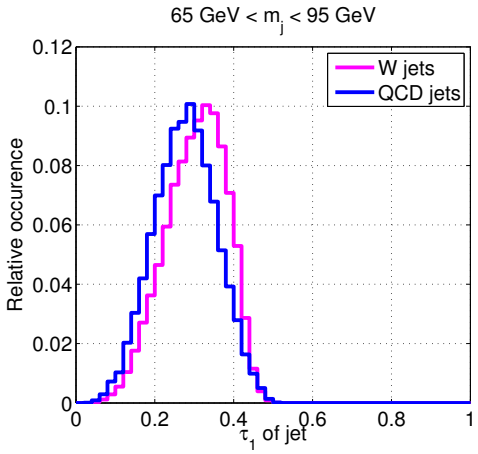


(a) Example W jet. The clusters with non-negligible deposited E_{T} are aligned with sub-jet axes, resulting in small τ_2 value.

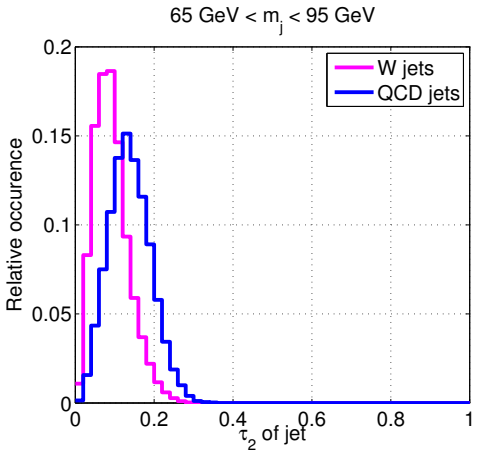
(b) Example light jet. A non-negligible amount of E_{T} is deposited by clusters far away from the sub-jets' axes. This contributes to a larger τ_2 value.

Fig. 5.3: Example event displays of a W jet and a light jet in $\eta - \phi$ space. The red and blue colors identify clusters matched to the two sub-jets. Size of each square is proportional to $\log(E_{\mathrm{T}})$. The small circles overlapping with the clusters in each picture denote the axes of these sub-jets [123].

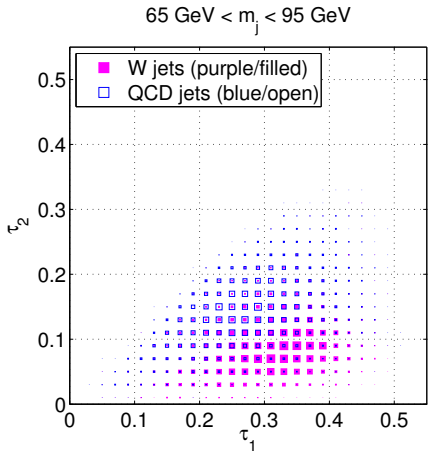
to be close to unity, whereas for W jets it has a tendency to be close to zero (Fig. 5.4d). Similarly, for top-quark tagging, $\tau_{32} = \tau_3/\tau_2$ variable can be used to discriminate top-quark jets from light jets [124].



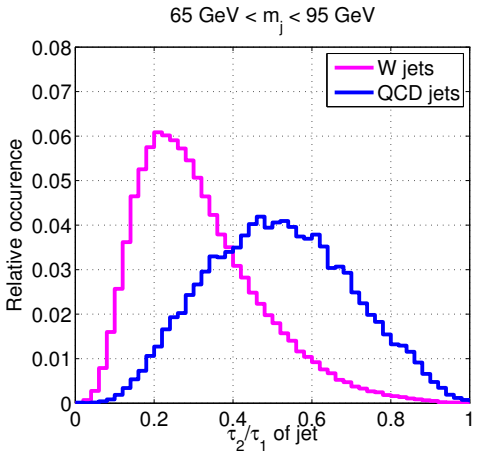
(a) Comparison of τ_1 variable for W jets and light jets.



(b) Comparison of τ_2 variable for W jets and light jets.



(c) Two-dimensional plot of τ_2 vs τ_1 for W jets and light jets.



(d) Comparison of $\tau_{21} = \tau_2/\tau_1$ ratio for W jets and light jets.

Fig. 5.4: Distributions of τ_1 and τ_2 variables for W and light jets with mass around the W peak [123]. Notice the correlation between τ_2 and τ_1 for light jets in Fig. (c), which is the motivation for introducing τ_2/τ_1 variable.

The N-subjettiness is only a single example from a long list of sub-structure variables that have been introduced and applied for identification of boosted top quarks and W bosons. In Sec. 5.3, a few examples of taggers examined using ATLAS simulation are described, that are subsequently studied in data in the signal efficiency measurement presented in Ch. 6.

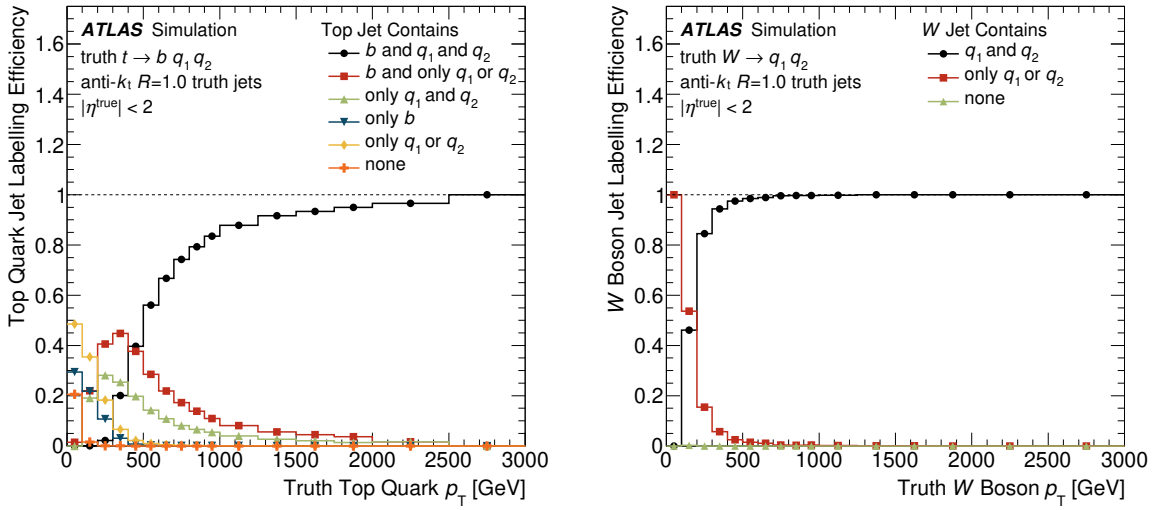
5.2 Optimisation of tagging algorithms

All of the taggers to be discussed are optimised using MC simulations, where the signal jets are obtained from simulated decays of BSM vector bosons, specifically the $Z' \to t\bar{t}$ for top-quark taggers and $W' \to WZ$ [125] for W-boson taggers. Both processes are simulated using Pythia8 with the NNPDF3.0LO [126] PDF set and the A14 [127] set of tuned parameters. The sample of background jets is obtained from dijet simulations, using the same generator setup as for the signal samples.

An exact definition of what is a signal and a background jet is necessary for optimisation and study of any tagging algorithm. The procedure of defining the origin of a jet with respect to a signal and background jet definition is referred to as jet *labelling*. The jet labelling is typically based on information accessible only in MC simulations. Top-quark jets and W-boson jets are labelled based on the containment of a top-quark or a W-boson particle and their decay products within the jet radius, at the parton level before hadronisation. Two definitions are encountered in this thesis:

- *Inclusive* – the top-quark or W-boson particle must be within $\Delta R < 0.75$ of the large-R jet axis, to consider the jet as originating from a top quark or a W boson, respectively.
- *Contained* – the partonic decay products of the top quark ($q\bar{q}'b$) or W boson ($q\bar{q}'$) are contained within $\Delta R < 0.75$ of the large-R jet axis.

The Fig. 5.5 illustrates the fraction of jets that are labelled as *contained* as a function of the p_{T} of the decaying top quark and W boson. The decay products of a W boson are contained within the jet with an almost 100 % efficiency at a $p_{\mathrm{T}} \approx 500\,\mathrm{GeV}$. For top quarks, the more complex decay topology leads to a much higher probability of not containing the decay products within the large-R jet. The contained definition is thus a compromise between the labelling efficiency and the resolution of the top-quark and W-boson jet mass peak. Unless explicitly stated, it is assumed that the tagger discussed uses the *contained* signal definition.



(a) (b)

Fig. 5.5: The fraction of particle-level anti-k_t $R = 1.0$ jets containing top-quark (a) and W-boson (b) decay products (jet labelling efficiency) as a function of the p_{T} of the decaying particle [1].

For background jets, no special criteria of their origin are used in simulation. Instead, the two highest-p_{T} jets from dijet MC simulation are used directly. These are jets originating from gluons and mostly light quarks, and hence are all treated as background jets.

For any tagging algorithm in general, the following quantities are typically examined as the figure of merit, the signal efficiency ϵ_{sig} and the background efficiency ϵ_{bckg}, defined as follows:

$$\epsilon_{\mathrm{sig}} = \frac{N^{\mathrm{tagged}}_{\mathrm{signal}}}{N^{\mathrm{tagged\ \&\ untagged}}_{\mathrm{signal}}}, \qquad \epsilon_{\mathrm{bckg}} = \frac{N^{\mathrm{tagged}}_{\mathrm{background}}}{N^{\mathrm{tagged\ \&\ untagged}}_{\mathrm{background}}}, \tag{5.5}$$

with $N^{\text{tagged}}_{\text{signal}}$ $\left(N^{\text{tagged}}_{\text{background}}\right)$ denoting the number of tagged jets which are labelled as signal (background) and $N^{\text{tagged \& untagged}}_{\text{signal}}$ $\left(N^{\text{tagged \& untagged}}_{\text{background}}\right)$ denoting the total number of jets labelled as signal (background), both tagged and untagged. In addition to ϵ_{bckg}, it is common to use the reciprocal quantity $1/\epsilon_{\text{bckg}}$, referred to as background rejection.

5.3 Examples of boosted tagging algorithms used in the ATLAS experiment

High-level substructure observables considered

Historically, the most frequent approach to boosted jet tagging is based on applying selection cuts on jet mass and jet moments sensitive to substructure, which are calculated using the jet constituents. We refer to these observables as *high-level* features, in contrast to the properties of the jet constituents, which are often referred to as *low-level* features. We introduce additional observables which are employed in the examined taggers.

The previously introduced jet mass suffers from degradation of resolution because of merging of clusters with increasing jet p_{T}, given by the finite granularity of the ATLAS calorimeter system. To mitigate this issue, the *track-assisted* mass (m_{TA}) [128] can be used, defined as follows:

$$m_{\text{TA}} = \frac{p_{\text{T}}^{\text{calo}}}{p_{\text{T}}^{\text{track}}} m^{\text{track}}, \tag{5.6}$$

where $p_{\text{T}}^{\text{calo}}$ is the jet transverse momentum and $p_{\text{T}}^{\text{track}}$ is the transverse momentum of a four-vector sum of tracks associated to the large-R jet, and m^{track} is the invariant mass of the four-vector sum of the tracks. The mass of individual tracks in the definition is set to the mass of the pion. The calorimeter-to-track transverse momentum ratio is used to correct for the neutral component of the jet energy deposits. The advantage of m_{TA} stems from the high angular resolution of the ID. A combination of both m_{calo} and m_{TA} resulted in the combined mass m_{comb}, which is an average of m_{TA} and m_{calo}, weighted by the inverse of their resolutions [128]. The combined mass is the large-R jet mass definition used in all of the analyses presented in this thesis.

The next group of substructure observables, commonly referred to as *jet shapes*, use relative position and momenta of jet constituents with respect to each other, without defining sub-jets as in the case of N-subjettiness. The jet mass is one example of a jet shape, other shapes include energy correlation ratios C_2, D_2 and e_3 [129, 130], the mass-normalised angularity a_3 [131] and the planar flow $\mathcal{P}$ [132].

The characteristics of a boosted top-quark and W-boson decay can also be quantified via the history of jet clustering, for example via the splitting scales $\sqrt{d_{12}}$ and $\sqrt{d_{23}}$. The d_{ij} in the definition of the splitting scales is the same distance measure d_{ij} as in the recombination jet algorithms in Eq. 4.5. The $\sqrt{d_{12}}$ and $\sqrt{d_{23}}$ scales correspond to the last and next-to-last merging step in the jet recombination algorithm, respectively. A slightly modified definition of the $\sqrt{d_{12}}$, referred to as z_{cut}, normalised by the mass of the boosted decaying particle, is studied in [133] and employed for boosted W-boson tagging, as well as a generalised splitting scale invariant under longitudinal boosts, $KtDR$ [134]. Finally, for top-quark tagging, an observable sensitive to the mass of the W boson in the top-quark decay, Q_w, is

defined through as the minimum mass of a pair of sub-jets of the large-R jet, where the sub-jets are clustered using exclusive k_t algorithm to form three sub-jets from the constituents of the large-R jet.

Finally, the last group of observables, particularly useful for tagging two-body decays of $W/Z/H$ bosons, are the centre-of-mass jet shapes. These observables quantify the anisotropy of the distribution of jet constituents in the jet rest frame, and include observables previously used in ATLAS measurements [135], such as the Fox-Wolfram moment R_2^{FW} [136, 137] and the aplanarity A [137].

Machine-learning based taggers using high-level substructure observables

The approach used in the Run-I for identifying boosted particles using a set of two variables, such as jet mass and a jet shape variable, can be extended to more variables. A particular approach gaining in popularity is the application of supervised machine learning (ML) techniques, such as BDTs and neural networks (NNs). These techniques combine several observables to classify the jets as signal or background. The output classifier is a continuous function giving a fixed range of values, that characterise the probability of the jet being a signal or background[(4)]. The internal architecture is different, but both techniques employ a non-linear transformation of the inputs into the classifier, where a cut on the classifier effectively defines a non-linear boundary in the input parameter space separating the jets into two sets, where one of them is considered background-like and the other signal-like. The parameters of this non-linear transformation are given by the architecture of the algorithm, and they are optimised during what is referred to as *training* where a representative set of signal and background jets are provided to the BDT or NN and a minimisation of a *loss function* is performed to find optimal values of the parameters[(5)]. The loss function represents a penalty for incorrect identification of the jet class [138].

The ML algorithms can exploit the discrimination power of the individual observables as well as their correlations for signal and background jets. The focus here is given to a deep neural network (DNN) top-quark and W-boson tagger, employing variables listed in Table 5.1. The taggers are trained using jets with p_{T} from 200 GeV up to 2 TeV and with $m_{\mathrm{comb}} > 40$ GeV. For the DNN top-quark tagger the jet p_{T} threshold is increased to 350 GeV to mitigate the larger separation of the top-quark decay products compared to W-boson decays. Further details on the training and the exact architecture of the ML-based taggers can be found in Ref. [1]. Once the DNN classifier is trained, two WPs are defined for each of the top-quark and W-boson taggers. The WPs are designed to give 80 % or 50 % efficiency, constant in large-R jet p_{T}. This is achieved by the fixed $m_{\mathrm{comb}} > 40$ GeV cut and a p_{T}-dependent lower cut on the DNN classifier.

Table 5.1: Summary of the large-R jet variables used by the DNN top-quark and W-boson taggers [1].

Tagger	Variables used	
Top-quark W-boson	p_{T}, m_{comb}, C_2, D_2, τ_{21}, $\sqrt{d_{12}}$	e_3, τ_1, τ_2, τ_3, τ_{32}, $\sqrt{d_{23}}$, Q_w R_2^{FW}, $\mathcal{P}$, a_3, A, z_{cut}, $KtDR$

(4) For BDT, the output classifier is typically from interval $[-1, 1]$, and for NN from interval $[0, 1]$, where the lower bound represents background-like jet and upper bound represents signal-like jet.

(5) A much more thorough introduction of NNs is provided in Ch. 8.

Taggers using lower-level jet information

The previously described ML-based taggers use high-level features. A different approach using low-level features is also examined in the signal efficiency measurement, a DNN top-quark tagger using topo-cluster information directly as inputs [1, 139], referred to as *TopoDNN*. This algorithm takes the vectors of the ten highest-p_{T} clusters in the jet. The tagger is trained on jets with $p_{\mathrm{T}} \in [450, 2400]$ GeV, using the same MC simulations for signal and background jets as the DNN top-quark tagger. Further details on the architecture and the training of TopoDNN can be found in Refs. [1, 139].

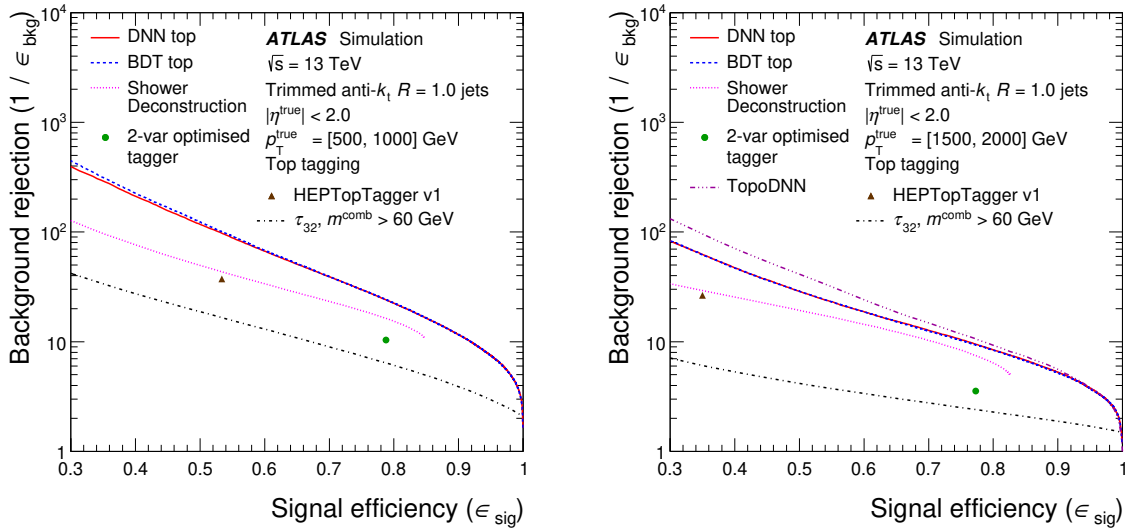
Finally, the *Shower Deconstruction* [140, 141] is an example of a boosted top-quark tagging algorithm that uses an approach similar to *matrix element method* [142]. In this approach, a likelihood ratio χ of the probability of signal hypothesis over the probability of background hypothesis is computed. The signal hypothesis is modelled via a simplified simulation of the top-quark decay and the subsequent parton shower to obtain a set of signal-like shower histories. Similarly, a set of shower histories is obtained for background hypothesis of a gluon splitting into $q\bar{q}$ pair followed by parton shower. To calculate the probability of signal hypothesis and background hypothesis, the constituents of the jet that is being tagged are re-clustered using the exclusive k_t algorithm into sub-jets. The sub-jet four-momenta are the observables used for the signal and background hypothesis probability calculation. To calculate χ, the large-R jet must contain at least three sub-jets obtained from the k_t re-clustering. Up to a maximum of six leading-p_{T} sub-jets are considered in the χ calculation due to computational complexity. In the matrix element simulation, the mass of the decaying top quark and the mass of the W boson from top quark are smeared using Breit-Wigner function with a width of 10 GeV and 25 GeV, respectively, to emulate the impact of limited detector resolution. A minimum-value cut on the log-likelihood ratio, $log\,\chi$, is imposed to tag top-quark jets and reject light jets. Similarly to the DNN and TopoDNN top-quark taggers, a WP is designed to maintain constant 80 % signal efficiency in jet p_{T}, which in addition to a jet p_{T}-dependent cut on $log\,\chi$, imposes $m_{\mathrm{comb}} > 60$ GeV selection on the large-R jet.

Expected performance of the taggers

A comparison of expected performance in MC simulations of these taggers is illustrated by the *receiver operator characteristic* (ROC) curves in Fig. 5.6 and 5.7, showing the dependence of background rejection of a tagger as a function of a picked signal efficiency. The comparisons are performed in two different kinematic regions for top-quark and W-boson tagging, respectively. The regions are defined by the transverse momentum of a particle-level anti-k_t $R = 1.0$ jet ($p_{\mathrm{T}}^{\mathrm{true}}$), associated with the reconstructed large-R jet. The particle-level jets are associated with large-R jets based on the distance of their axes, requiring $\Delta R < 0.75$. The choice of $p_{\mathrm{T}}^{\mathrm{true}}$ as the quantity defining the kinematic range is motivated by the fact that in the ROC curves, comparisons are made between taggers based on anti-k_t $R = 1.0$ jets and one particular tagger using C/A $R = 1.5$ jets, the *HepTopTagger*[(6)] [143, 144]. A common particle-level jet definition is used in the matching with both the reconstructed jet definitions to provide a consistent definition of kinematic regions in the tagger comparisons. Additionally, the ROC curves show the performance of a two-variable $m_{\mathrm{comb}} + \tau_{32}$ top-quark tagger and $m_{\mathrm{comb}} + D_2$ W-boson tagger. These taggers have cuts on both observables optimised simultaneously as a function

[(6)] The HepTopTagger is aimed at tagging boosted top quarks at lower transverse momenta ($p_{\mathrm{T}} > 200$ GeV), where the decay products are on average too separated to be contained within a $R = 1.0$ jet.

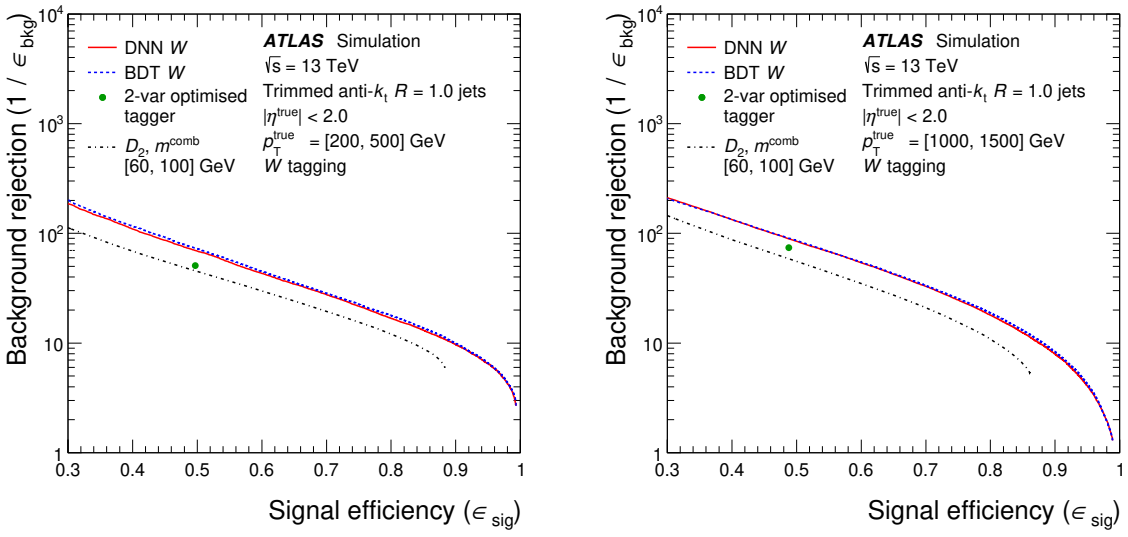
of jet p_{T}. This simultaneous cut optimisation leads to better exploitation of correlations between the observables for signal and background jets, yielding larger rejection than the combination of a fixed m_{comb} cut and a p_{T} dependent cut on τ_{32} or D_2 observables. The ROC curves clearly show the dominance of ML-based taggers in terms of higher background rejection for a given signal efficiency. As is shown in the figures, both DNN-based and BDT-based top-quark and W-boson taggers were trained in Ref. [1], with the primary goal to examine whether a DNN can learn additional features from the high-level sub-structure observables in comparison to BDTs. The similar performance of the BDT and DNN shown in the ROC curves indicate that this does not appear to be the case. In addition, the performance difference between ML-based and other taggers for top-quark tagging is significantly larger than for W-boson tagging. This attributed to the fact that the top-quark decay has a richer, more complex topology, thus having more distinct features that a ML-based technique can exploit to differentiate signal from background jets. Finally, for the highest-p_{T} interval for top-quark tagging in Fig. 5.6b, the comparisons include also the TopoDNN tagger. This tagger was trained using the *inclusive* signal definition, rather than the *contained* signal definition used for the rest of the taggers examined here. For jets with $p_{\mathrm{T}} > 1500\,\mathrm{GeV}$ shown in Fig. 5.6b, the labelling definition differences become negligible[(7)]. The TopoDNN achieves better background rejection at smaller signal efficiencies than the DNN top-quark tagger, suggesting that by training on low-level inputs, the TopoDNN can learn additional features not contained within the high-level observables included in the DNN top-quark tagger.



(a) (b)

Fig. 5.6: Expected performance of selected boosted top-quark taggers as a function of transverse momentum of a particle-level jet ($p_{\mathrm{T}}^{\mathrm{true}}$) associated with the reconstructed large-R jet, for $p_{\mathrm{T}}^{\mathrm{true}} \in [500, 1000]$ GeV (a) and $p_{\mathrm{T}}^{\mathrm{true}} \in [1500, 2000]$ GeV (b) [1]. The lines indicate the dependence of background rejection for a given signal efficiency for the DNN and BDT taggers, the Shower Deconstruction tagger and for a tagger based on a fixed cut of $m_{\mathrm{comb}} > 60\,\mathrm{GeV}$ and a variable cut on τ_{32} to achieve constant efficiency. In (b) the TopoDNN tagger is also included. Additionally, the two points in the plot show WPs for a 2-variable optimised $m_{\mathrm{comb}} + \tau_{32}$ tagger and the HepTopTagger.

[(7)]The fraction of inclusive signal jets that are also contained signal jets is significantly less than unity for jets with $p_{\mathrm{T}} \in [500, 1000]$ GeV. For this reason, it is not possible to make fair comparisons between TopoDNN and other taggers in this kinematic region.



(a) (b)

Fig. 5.7: Expected performance of selected boosted W-boson taggers as a function of transverse momentum of a particle-level jet ($p_{\mathrm{T}}^{\mathrm{true}}$) associated with the reconstructed large-R jet, for $p_{\mathrm{T}}^{\mathrm{true}} \in [200, 500]$ GeV (a) and $p_{\mathrm{T}}^{\mathrm{true}} \in [1000, 1500]$ GeV (b) [1]. The lines indicate the dependence of background rejection for a given signal efficiency for the DNN and BDT taggers and for a tagger based on a fixed mass-window cut of $m_{\mathrm{comb}} \in [60, 100]$ GeV and a variable cut on D_2 to achieve constant efficiency. Additionally, the single point in the plot shows an optimised WP for a 2-variable $m_{\mathrm{comb}} + D_2$ tagger.

Studies of taggers in data

Finally, when using boosted jet taggers in MC simulations, assumptions are made about correct modelling of the involved observables and their correlations. For this reason, studies of performance of the taggers in data are vital to assess the accuracy of the MC modelling. Previous studies in data were performed using tag-and-probe approach by selecting signal samples enriched in top-quark and W-boson jets or background samples enriched in light-quark and gluon jets. To study the modelling of taggers on top-quark and W-boson jets and to measure tagger signal efficiency, $t\bar{t}$ events can be used, as was done in previous Run-I studies [112, 122]. Several approaches have been used to assess the modelling of taggers on background jets and to measure background rejection. One option is to select events using jet triggers as was done in the Run-I W-boson tagging study [112] and Run-II study in Ref. [1]. This yields a very pure sample of jets which are either quark- or gluon-initiated. An alternative option is to select events using single-electron triggers, which have p_{T} thresholds starting at 20 GeV, and subsequently veto events containing a tight isolated muon, as was done in Ref. [122]. This selection yields a sample enriched with multijet events where one of the jets was either mis-reconstructed as an electron or contained a non-prompt electron that fired the trigger. Finally, in Run-II, the study of background jets is also performed in the photon+jet events [1], where a single photon trigger with photon $p_{\mathrm{T}} > 150$ GeV threshold is used.

In the following chapter, a study of modelling of boosted top-quark and W-boson taggers and their signal efficiency measurement in $t\bar{t}$ events using Run-II data is presented.

6

Measurement of signal efficiency of boosted top-quark and W-boson taggers

In this analysis, a measurement of signal efficiency of several top-quark and W-boson taggers is performed using ATLAS Run-II data and compared to MC predictions. The study presented here was published as a part of Ref. [1].

The efficiency measurement uses a tag-and-probe approach. A sample enriched in $t\bar{t}$ events with single lepton in the final state is selected using criteria outlined in Sec. 6.3. The hadronically-decaying top quark is reconstructed as a single large-R jet, which is used as the probe to study the boosted taggers in data. Top-quark and W-boson taggers are studied in two respective sub-samples, defined using additional selection criteria, to enhance contribution of large-R jet containing either all of the top-quark decay products, or the decay products of the W boson from the top quark. The MC generators used for the simulation of signal and background estimates and data-driven background estimates are discussed in Sec. 6.1 and Sec. 6.2. Comparisons of data and prediction for a number of substructure-related observables are shown in Sec. 6.5 and general control distributions for various objects used in the analysis showing the level of agreement between data and prediction are included in App. A. The strategy to measure the signal efficiency of the taggers is described in Sec. 6.6. Finally, the results of the measurement are shown in Sec. 6.7.

6.1 Data and simulation samples

The measurement is performed using data collected during 2015 and 2016 period of Run-II data-taking at $\sqrt{s} = 13$ TeV with an integrated luminosity of 36.1 fb^{-1}.

A number of background processes mimic the $t\bar{t}$ production, including single top-quark production, associated production of W/Z bosons and jets (V+jets), and vector-boson pair (*diboson*) production. All of these processes are modelled using MC simulations. Finally, the $t\bar{t}$ production is also mimicked by multijet production via fake and non-prompt leptons. This background is estimated using the data-driven *matrix method*, described in Sec. 6.2.

For all the MC simulations, the simulation of in-time and out-of-time pile-up is performed. The hard scattering events are overlaid with a set of minimum-bias interactions generated using Pythia8 [145] with the MSTW2008LO [146] PDF set with the A2M [127] tuned parameter settings.

6.1.1 $t\bar{t}$ and single top-quark modelling

For the $t\bar{t}$ and single top-quark production, several generator setups are used for the simulation and estimation of $t\bar{t}$ modelling uncertainties, as described below. All of the $t\bar{t}$ simulations are normalised to the NNLO QCD cross-section quoted in Sec. 2.1, and the single top-quark simulations are normalised to the approximate NNLO QCD cross-section [147–150]. All of the $t\bar{t}$ and single top-quark simulations use the top-quark mass set to a value of 172.5 GeV.

The nominal $t\bar{t}$ as well as single top-quark Wt- and s- channels MC samples are simulated using Powheg-Box v2 generator [151–153] for the matrix element (ME) generation at the NLO QCD accuracy in production, using CT10 [154] NLO PDF set. The single top-quark t-channel production is simulated using Powheg-Box v1 generator with the same choice of the PDF set. To simulate the parton shower (PS), hadronisation and the underlying event, for all of the aforementioned processes, the ME generation is interfaced with Pythia6 (v6.428) [155] generator with the CTEQ6L1 [156] PDF set and the corresponding Perugia 2012 tune (P2012) [157]. The h_{damp} parameter[(1)] is set to the mass of the top quark.

To estimate the uncertainties on the $t\bar{t}$ modelling, several alternative MC samples are used. The uncertainty on the initial-state radiation (ISR) modelling is estimated by comparing the nominal sample with samples generated using the same generator setup, but with a modified value of μ_R and μ_F scales ($\times 2$ or $\times 0.5$) and simultaneously varied h_{damp} value ($h_{\text{damp}} = m_{\text{top}}$ or $h_{\text{damp}} = 2 \times m_{\text{top}}$) [158].

The estimate of the uncertainty on the PS and hadronisation modelling is derived by comparing results obtained with the Powheg-Box v2 generator interfaced to Herwig++ [159] instead of Pythia6. To estimate the uncertainty on the choice of matching scheme between the ME and PS, the MadGraph5_aMC@NLO generator [160] interfaced with Herwig++ is compared to the prediction of Powheg-Box v2 interfaced with Herwig++.

6.1.2 Modelling of additional backgrounds

Both V+jets production as well as diboson ($WW/WZ/ZZ$) production is simulated using Sherpa [161] v2.1.1 and v2.2.1, respectively. The ME is calculated at NLO accuracy for up to two jets and LO accuracy for up to four jets produced, using the CT10 PDF set, and the default set of tuned parameters from the Sherpa authors. The V+jets samples are normalised to the NNLO cross-section prediction [162], while the diboson samples are normalised to NLO QCD cross-section prediction [163].

6.2 Estimation of non-prompt and fake leptons background

Among the background processes mimicking the $t\bar{t}$ production in the single-lepton channel is the multijet production. Prompt electron candidates can be faked by jets or non-prompt decays of hadrons, while the prompt muon candidates are most-frequently faked by non-prompt muons from decays of hadrons. Comparisons of simulations with data suggest that the modelling of these processes is very challenging. In addition, the selection requirements for an isolated lepton lead to low acceptance of fake or non-prompt leptons, however the multijet production rate is orders of magnitude higher than $t\bar{t}$ production. Therefore, the MC estimation of this background are not only difficult to model accurately

[(1)] The h_{damp} parameter controls the p_{T} of the first gluon emission beyond LO in the ME generation, effectively controlling the ME cut-off in the matching with PS.

but also impractical due to the necessity to simulate a large number of events. Therefore, a data-driven *matrix method* is used to estimate the contribution.

Table 6.1: Comparison of the tight and loose identification and isolation WPs for leptons, used to estimate the multijet background. The description of the WPs is described in Sec. 4.2 and Sec. 4.3.

Lepton selection	Criterion	Electrons	Muons
tight	Identification WP	*tight*	*medium*
	Isolation WP	*track-based*	*fixed-cut*
loose	Identification WP	*medium*	*loose*
	Isolation WP	None	

The matrix method employs two different lepton selections, a *tight* and a *loose* selection, specified in Table 6.1, differing in the choice of identification and isolation WP. All the other event selection criteria are the same for both lepton definitions. It follows that the total yields in the loose (N^{loose}) and tight (N^{tight}) selections are:

$$\begin{aligned} N^{\text{tight}} &= N^{\text{tight}}_{\text{real}} + N^{\text{tight}}_{\text{fake}}, \\ N^{\text{loose}} &= N^{\text{loose}}_{\text{real}} + N^{\text{loose}}_{\text{fake}}, \end{aligned} \tag{6.1}$$

where $N^{\text{tight}}_{\text{real}}$ ($N^{\text{loose}}_{\text{real}}$) denotes the number of events with a prompt lepton satisfying tight (loose) criteria and $N^{\text{tight}}_{\text{fake}}$ ($N^{\text{loose}}_{\text{fake}}$) the number of events with fake/non-prompt lepton satisfying tight (loose) criteria. The so-called real (ϵ^{real}) and fake (ϵ^{fake}) efficiencies are defined, which give the probability that a lepton passing loose selection also passes tight selection:

$$\begin{aligned} \epsilon_{\text{real}} &= \frac{N^{\text{tight}}_{\text{real}}}{N^{\text{loose}}_{\text{real}}}, \\ \epsilon_{\text{fake}} &= \frac{N^{\text{tight}}_{\text{fake}}}{N^{\text{loose}}_{\text{fake}}}. \end{aligned} \tag{6.2}$$

Combining Eq. 6.1 and 6.2, it is possible to extract the estimated number of events from fake/non-prompt leptons background passing the tight selection, assuming the knowledge of N^{tight}, N^{loose}, ϵ_{real} and ϵ_{fake}:

$$N^{\text{tight}}_{\text{fake}} = \frac{\epsilon_{\text{fake}}}{\epsilon_{\text{real}} - \epsilon_{\text{fake}}} \left(\epsilon_{\text{real}} N^{\text{loose}} - N^{\text{tight}} \right). \tag{6.3}$$

The ϵ_{real} and ϵ_{fake} efficiencies are determined using data in suitable regions enriched with prompt and fake/non-prompt leptons. The ϵ_{real} efficiency is derived in $Z \to \ell^+\ell^-$ events using the tag-and-probe method, where one of the leptons is required to pass tight criteria to ensure a pure sample of prompt leptons and the other lepton is used for the ϵ_{real} measurement. The ϵ_{fake} efficiency is derived in a selection similar to that of the single-lepton $t\bar{t}$, but with inverted $E_{\text{T}}^{\text{miss}}$ and M_{T}^{W} cuts, enhancing the contribution of the fake and non-prompt leptons. The residual prompt-lepton contributions in this

region are estimated from MC simulation and subtracted from data, thus obtaining a "pure" sample of fake/non-prompt leptons.

In practice, it is necessary to estimate distributions of various observables of the fake/non-prompt leptons background instead of just the total yield. This is done by constructing the distributions from loose events in data by applying a per-event weight:

$$w = \frac{\epsilon_{\text{fake}}(\epsilon_{\text{real}} - c)}{\epsilon_{\text{real}} - \epsilon_{\text{fake}}}, \tag{6.4}$$

where $c = 1$ for loose leptons that pass the tight criteria, and $c = 0$ for loose leptons that do not pass tight criteria (*loose not-tight*).

In this analysis, the ϵ_{real} and ϵ_{fake} efficiencies used were measured in Ref. [164]. The efficiencies are parametrised as a function of multiple variables related to lepton kinematics and isolation:

- Electron efficiencies are parametrised as a function of electron p_{T} and the $E_{\text{T}}^{R_{\text{cut}}}$ as defined for the *fixed-cut* isolation in Sec. 4.2.
- Muon efficiencies are parametrised as a function of muon p_{T}, $E_{\text{T}}^{R_{\text{cut}}}$ and the distance to the closest jet.

The ϵ_{fake} efficiency for electrons ranges from 18 % up to 92 % for high-p_{T} electrons and for muons from 4 % up to 94 % for high-p_{T} muons [164].

6.3 Event selection

The selection used in this analysis is based on the selection optimised in the search for heavy particles decaying into $t\bar{t}$ pairs in [164], which includes a dedicated boosted region to reconstruct the hadronically decaying top quark as a large-R jet and the leptonically decaying top quark from single lepton in event, a close-by small-R jet, and missing transverse momentum.

Only events recorded under stable beam conditions and with all ATLAS detector sub-systems operations are considered. A reconstructed primary vertex is necessary and a single-electron or single-muon trigger must fire. The trigger selections are shown in Table 6.2, and include selection on minimum lepton p_{T}, identification WP and isolation. The trigger requirements during the 2016 data-taking period are tighter than in 2015. The differences arise from the increased instantaneous luminosity of the LHC and limits on the throughput of the ATLAS trigger and data acquisition system. The electron identification WPs used in the triggers correspond to those described in Sec 4.2 and the isolation WPs are defined similarly as the fixed-cut track isolation in Sec 4.2 and 4.3, but with a looser selection[(2)]. The trigger requirements are imposed in both data as well as MC simulation.

The events are then required to pass the following selection criteria:

- At least one *tight* electron or *medium* muon with $p_{\text{T}} > 30\,\text{GeV}$, passing the *fixed-cut* isolation selection, is required. These criteria are tighter than the trigger criteria, ensuring the trigger is fully-efficient in this kinematic regime. Events containing additional leptons with $p_{\text{T}} > 25\,\text{GeV}$ are rejected.

[(2)] The fraction of momentum carried by tracks in the cone surrounding the electron or muon candidate $p_{\text{T}}^{\text{var}R_{\text{cut}}}/p_{\text{T}}^{\ell}$ is allowed to be higher for looser isolation requirements of the triggers.

Table 6.2: Single-lepton triggers employed in the analysis, including their p_{T} thresholds, identification WP, isolation WP or any other criteria if applicable.

	Electron triggers		Muon triggers	
Data period	p_{T} threshold	extra criteria	p_{T} threshold	extra criteria
2015	24 GeV	*medium* WP	20 GeV	
	60 GeV	*medium* WP	50 GeV	
	120 GeV	*loose* WP		
2016 and later	26 GeV	*tight* WP, no $\lvert d_0\rvert/\sigma(d_0)$ cut, *loose* isolation	26 GeV	*medium* isolation
	60 GeV	*medium* WP	50 GeV	
	140 GeV	*loose* WP		

- The events must have $E_{\mathrm{T}}^{\mathrm{miss}} > 20\,\mathrm{GeV}$ and $E_{\mathrm{T}}^{\mathrm{miss}} + M_{\mathrm{T}}^{W} > 60\,\mathrm{GeV}$[(3)]. This selection enhances the contribution of leptons originating from W-boson decays.
- To ensure a topology consistent with semileptonically-decaying top quark, at least one small-R calorimeter jet with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 2.5$ is required, and must be close to the lepton; $\Delta R(\mathrm{jet}, \ell) < 1.5$.
- At least one $R = 0.2$ track jet in the event is required, that is identified as originating from B-hadron (b-tagged) based on the b-tagging selection described in Sec. 4.4.5.
- At least one large-R jet with $p_{\mathrm{T}} > 200\,\mathrm{GeV}$ and $|\eta| < 2.0$ is required. It is expected, that the hadronically-decaying top quark is well separated from the semileptonically-decaying top-quark, hence the following angular separation criteria are required for the large-R jet:
  - Distance from the small-R jet close to the lepton is required to be $\Delta R(\text{small-}R\text{ jet}, \text{large-}R\text{ jet}) > 1.5$.
  - The transverse plane opening angle between the large-R jet and the isolated lepton $\Delta\phi(\text{large-}R\text{ jet}, \ell) > 2.3$.

  If multiple jets satisfy this selection, the highest-p_{T} jet is considered.

The yields for $t\bar{t}$, the individual backgrounds and the data are shown in Table 6.3, separately for the electron and muon channel. The pre-selected sample is divided into two sub-samples based on the proximity of a b-tagged track-jet and the selected large-R jet:

- The top-quark-enriched sample is selected by requiring $\Delta R(b\text{-jet}, \text{large-}R\text{ jet}) < 1.0$, enhancing the likelihood that the jet contains top-quark decay products. To further increase the fraction of fully-contained[(4)] top-quark jets, the large-R jet must satisfy $p_{\mathrm{T}} > 350\,\mathrm{GeV}$ requirement.
- The W-boson-enriched sample is selected by requiring $\Delta R(b\text{-jet}, \text{large-}R\text{ jet}) > 1.0$. This selection enhances the selection of $t\bar{t}$ events, where the B-hadron from top-quark decay is not clustered in the large-R jet, thus the large-R jet contains only the W-boson decay products,

[(3)] $M_{\mathrm{T}}^{W} = \sqrt{2p_{\mathrm{T}}^{\ell}E_{\mathrm{T}}^{\mathrm{miss}}(1-\cos\Delta\phi)}$, where $\Delta\phi$ is the opening angle between the lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$ direction in the plane perpendicular to the beam pipe.

[(4)] Fully-contained here refers to having all the top-quark decay products clustered within the large-R jet.

and can be used as a probe to study W-tagging algorithms. The W-boson-enriched sample is kinematically limited to large-R jets with p_{T} up to approximately 600 GeV. This is due to the selection acceptance decreasing with large-R jet p_{T} due to the decrease in the ΔR separation between the B-hadron and the large-R jet.

The total yield of $t\bar{t}$, the individual backgrounds and the data, are shown in Table 6.4 for the two sub-samples, with the electron and muon channel yields combined. Finally, the $t\bar{t}$ and single top-quark MC samples are split according to the *contained* labelling definition of the selected large-R jet. The $t\bar{t}$ MC sample is split into three categories, *top*, W and *other*, depending on whether $q\bar{q}'b$ partons are contained within the jet, or only $q\bar{q}'$, or neither of the two options, respectively. Similarly, the single top-quark MC sample is split into W and *other* categories based on the same labelling criteria. The W category contains events from the single top-quark tW production, where the top-quark decays leptonically and the W-boson decays hadronically and is reconstructed as a large-R jet.

Table 6.3: Yields for the individual predicted processes and data after the pre-selection, shown separately for the electron channel (e+jets column) and the muon channel (μ+jets column). Only the statistical uncertainties on the predicted yield corresponding to $L = 36.1\,\mathrm{fb}^{-1}$ are shown[(5)].

Process	e+jets	μ+jets
$t\bar{t}$	112460 ± 230	121790 ± 230
Single top-quark	11160 ± 70	11370 ± 60
W + jets	11050 ± 100	21850 ± 130
Z + jets	2525 ± 24	2278 ± 22
Diboson	809 ± 14	858 ± 14
Fake/non-prompt lep.	19400 ± 500	8400 ± 600
Total prediction	157400 ± 600	166500 ± 800
Data	146300	153400

A normalisation discrepancy is observed when comparing data with prediction, attributed to the known mismodelling of the p_{T} of the top quark, where the MC simulation predicts harder p_{T} spectrum than that observed in data, as documented in various comparisons of $t\bar{t}$ generator predictions with unfolded differential cross-section measurements [158, 165, 166]. This mismodelling translates for example into the p_{T} of the leading large-R jet, hence the selection on the minimum leading jet p_{T} leads to a normalisation acceptance effect. To assess the possible impact of the large-R jet p_{T} mismodelling on the shape of substructure observables, a data-driven reweighting test to compensate the leading-p_{T} large-R jet mismodelling in $t\bar{t}$ is performed and documented in App. B. The impact of the reweighting on the shape of the investigated sub-structure observables is found to be small in the relevant regions of phase space. The normalisation mismodelling in different intervals of large-R jet p_{T} is mitigated by the design of the template fit performed in the signal efficiency measurement, described in Sec. 6.6.

[(5)] Unless, otherwise stated, it is assumed that the MC-predicted yields are always normalised to the integrated luminosity of the corresponding dataset. Therefore, for weighted predictions, i.e. the MC-simulated samples and the matrix-method-estimated fake/non-prompt leptons background, the statistical uncertainty of the prediction is calculated as the $\sqrt{\sum_{i}^{\text{Events}} w_i^2}$, where w_i is the total weight of the i^{th} event. The statistical uncertainty in particular for MC samples is typically smaller than the statistical uncertainty of data, since a larger number of events are generated than the number of data events.

Basic control distributions showing the level of agreement between data and prediction after the pre-selection and for the top-quark and W-boson-enriched sub-sample selections are shown in App. A.

Table 6.4: Yields for the individual predicted processes and data for the sub-samples selected for top-quark tagging (Top-quark selection) and W-boson tagging (W-boson selection). The yields in the electron and muon channel are combined together.

Process	W-boson selection	Top-quark selection
$t\bar{t}$	112680 ± 220	30610 ± 120
Single top-quark	12380 ± 70	2562 ± 31
W + jets	13140 ± 90	4180 ± 40
Z + jets	2011 ± 19	557 ± 8
Diboson	812 ± 14	297 ± 9
Fake/non-prompt lep.	14800 ± 600	2690 ± 180
Total prediction	155900 ± 800	40900 ± 290
Data	146400	35250

6.4 Systematic uncertainties

The data-to-prediction comparisons as well as the modelling of signal efficiency in data are both impacted by a number of systematic uncertainties. These include both uncertainties due to theoretical assumptions made within the MC predictions as well as uncertainties on the reconstruction, calibration and detector response of the physics objects employed in this measurement. For each systematic uncertainty, a systematically-shifted prediction of the distribution is obtained, for both up and down variation of the uncertainty. One-sided systematic uncertainties are symmetrised by mirroring the shift in both up and down direction. In the data-to-prediction comparisons in Sec. 6.3, the two-sided variation are also symmetrised, taking the maximum of the up/down shifts. The individual uncertainties are summed in quadrature.

In the signal efficiency fit, systematic uncertainties are propagated by repeating the fit with a systematically-shifted set of templates. The systematic uncertainty on the signal efficiency is estimated as the difference between the fit result obtained from the systematically-shifted templates and the nominal set of templates. The uncertainties considered are described in the following paragraphs.

6.4.1 Experimental uncertainties

In this section, the uncertainties related to object reconstruction, identification, pile-up modelling and the luminosity measurement are outlined.

Lepton-related uncertainties

The lepton efficiencies of trigger, reconstruction, identification and isolation differ between data and the MC predictions. These differences are corrected using scale factors applied to MC simulations, which are obtained in dedicated tag-and-probe measurements in $Z \to \ell^+\ell^-$ and $J/\psi \to \ell^+\ell^-$ events,

as described in Sec. 4.2 and 4.3. The uncertainties on these scale-factor measurements [99, 102] are propagated in the analysis by varying the nominal scale factors by the systematically-shifted ones. Similarly, the aforementioned event topologies are used to calibrate the lepton momentum scale and resolution in MC simulation to data, where the uncertainties on these measurements [100, 102] are propagated by shifting the scale and smearing the resolution in MC simulations by the systematic variations.

Small-R calorimeter jet uncertainties

The calibration of jet energy described in Sec. 4.4.3 contains many sources of uncertainties, which are constructed as an uncorrelated set of almost hundred eigenvectors. The signal efficiency measurement was found to be insensitive to the JES of small-R jets, and therefore a strongly-reduced set of four NPs is used [110]. The jet energy resolution (JER) in MC simulation is compared to data in a dedicated in-situ measurement [167] in dijet events. The uncertainty on the JER calibration, when averaged over the central calorimeter η, ranges approximately from 20 % for jets with $p_{\mathrm{T}} \sim 30\,\mathrm{GeV}$ down to 10 % for jets with $p_{\mathrm{T}} \sim 500\,\mathrm{GeV}$ [167]. A single nuisance parameter is used to parametrise the JER uncertainty.

Finally, the JVT algorithm is used to suppress contribution of small-R jets originating from pile-up. The efficiency of the JVT is calibrated in $Z \to \mu^+\mu^-$ events with additional production of jets [168]. Per-jet SFs are used to correct the JVT efficiency in MC simulations to match that in data. The SFs are varied in the analysis within the uncertainties of the JVT efficiency measurement in data.

Large-R jet uncertainties

In contrast to the small-R jets, for the large-R jets uncertainties are derived for both JES as well as jet mass scale (JMS) and the scale for relevant substructure observables. In this analysis, the in-situ calibration of large-R jets described in Sec. 4.4.4 was not yet available, only the MC-based calibration of energy scale and mass scale is performed. A simplified in-situ method to estimate the uncertainty on this calibration was used, based on comparing calorimeter response to ID response [112]. In this method a double ratio of $\left\langle X^{\mathrm{jet}}/X^{\mathrm{ref}}\right\rangle_{\mathrm{data}} / \left\langle X^{\mathrm{jet}}/X^{\mathrm{ref}}\right\rangle_{\mathrm{MC}}$ for a quantity X is measured, where the reference value X^{ref} is calculated using ID tracks associated to the calorimeter jet. Using this method, the uncertainties on JES, JMS and the scale of the substructure observables are estimated.

The JER and jet mass resolution (JMR) uncertainties are based on previous studies. The JER uncertainty is estimated using the same techniques as those for small-R jets, and is found to be approximately 10 % for the jet $p_{\mathrm{T}} > 200\,\mathrm{GeV}$ [112]. The JMR uncertainty is estimated from the data/MC variations in the width of the mass peak of large-R jets identified as W bosons in $t\bar{t}$ events [112, 113], thus the JMR is propagated by an additional 20 % smearing of the nominal JMR.

Missing transverse energy uncertainties

The $E_{\mathrm{T}}^{\mathrm{miss}}$ quantity is computed from several sources of objects described in Sec. 4.5. The uncertainties on these objects, described in this sub-section, are propagated into the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation. Additionally, $E_{\mathrm{T}}^{\mathrm{miss}}$ includes the soft term, which has an uncertainty estimated in-situ using $Z \to \mu^+\mu^-$ events [169]. This event topology is suitable due to the high signal-to-noise ratio, the precise muon kinematics

reconstruction and the expected true $E_{\mathrm{T}}^{\mathrm{miss}}$ value being zero. It is therefore primarily sensitive to the $E_{\mathrm{T}}^{\mathrm{miss}}$ soft-term effects.

Flavour tagging uncertainties

The modelling of the MV2c10 b-tagging algorithm is judged based on the level of agreement of the efficiency to tag b, c and light jets between MC prediction and data. The differences between the efficiencies in data and MC simulations are corrected using scale-factors determined in-situ in dedicated measurements [116, 117, 170], as a function of the jet p_{T}. The b-tagging efficiency is calibrated using di-lepton $t\bar{t}$ events. The c-tagging efficiency is calibrated using single-lepton $t\bar{t}$ events, exploiting the fact that the W-boson decay branching fraction of final states including a charm hadron is approximately 33 % [31]. Finally, the light-jet efficiency is calibrated in multijet events. The uncertainties on the scale-factors are constructed as a set of de-correlated nuisance parameters.

Luminosity uncertainty

The estimates of processes modelled using MC simulations are normalised to the product of cross-section and integrated luminosity. The luminosity is measured experimentally using the LUCID-2 detector [171]. The uncertainty on the integrated luminosity of the 2015-16 dataset is 2.1 % [172].

Pile-up modelling uncertainty

The modelling of the additional pp collisions per bunch crossing multiplicity μ via the overlay of minimum-bias events onto the hard-scattering simulation is not sufficiently accurate to describe the data. Therefore, the MC simulations are reweighted event-by-event to match the pile-up profile in data based on μ. The per-event weights are varied within their uncertainties, which are derived from the luminosity measurements of μ profile in data [172] .

6.4.2 Signal and background modelling uncertainties

A number of assumptions enter the MC-simulated predictions, which have associated theoretical uncertainties, discussed in this section.

Normalisation uncertainties of MC-predicted processes

In addition to the luminosity uncertainty, the inclusive yield of the MC-simulated signal and background estimates is impacted by the uncertainty on the respective theoretical cross-section predictions. The theoretical uncertainties on normalisation of $t\bar{t}$ as well as single-top and W+jets backgrounds are listed in Table 6.5. In the signal efficiency fit, the $t\bar{t}$ normalisation has no impact, since the normalisation of $t\bar{t}$ is determined in the fit. The normalisation uncertainty of the diboson production is neglected due to the very small contribution of this background.

Modelling uncertainties of the W+jets background

The W+jets is the dominant MC-simulated background in the signal efficiency fit and as such, additional modelling uncertainties impact the acceptance and the shape of distributions. Firstly, the CKKW scale,

Table 6.5: The uncertainties on the normalisation of the individual MC-simulated processes considered in the boosted tagging efficiency measurement. The uncertainty values quoted here are based on the theory predictions cited in Sec. 6.1. The uncertainties on the cross-section typically include variation of the μ_R and μ_F scales as well uncertainties related to the PDF set.

Process	Normalisation uncertainty (%)
$t\bar{t}$	5.5
Single-top	5.3
W+jets	5.0

controlling the matching of the ME and PS, is varied by a factor ×0.75 and ×1.5 with respect to the nominal value of 30 GeV, and the soft-gluon resummation scale QSF is varied by a factor ×0.5 and ×2.0 [173, 174]. Finally, factor ×0.5 and ×2.0 variations of μ_R and μ_F scales in the ME are considered, both in an uncorrelated as well as correlated and anti-correlated way. The maximum of the various correlation scenarios of μ_R and μ_F scale variations is considered as the final uncertainty.

Uncertainties on the fake and non-prompt leptons estimate

The fake and non-prompt leptons background is estimated using data-driven matrix method. To account for limitations of this estimate, two source of uncertainties are considered. Firstly, a conservative normalisation uncertainty of 50 % is propagated in the analysis, based on previous studies of non-prompt and fake lepton estimates [175]. Secondly, the ϵ_{real} and ϵ_{fake} efficiencies are varied within their statistical uncertainties in a manner that maximises the impact on the matrix-method weight in Eq. 6.4. This is done by simultaneously varying the ϵ_{real} efficiency up and ϵ_{fake} efficiency down within their statistical uncertainties obtaining an "up" variation, and vice-versa for a "down" variation. This uncertainty induces a variation in shape as well as normalisation.

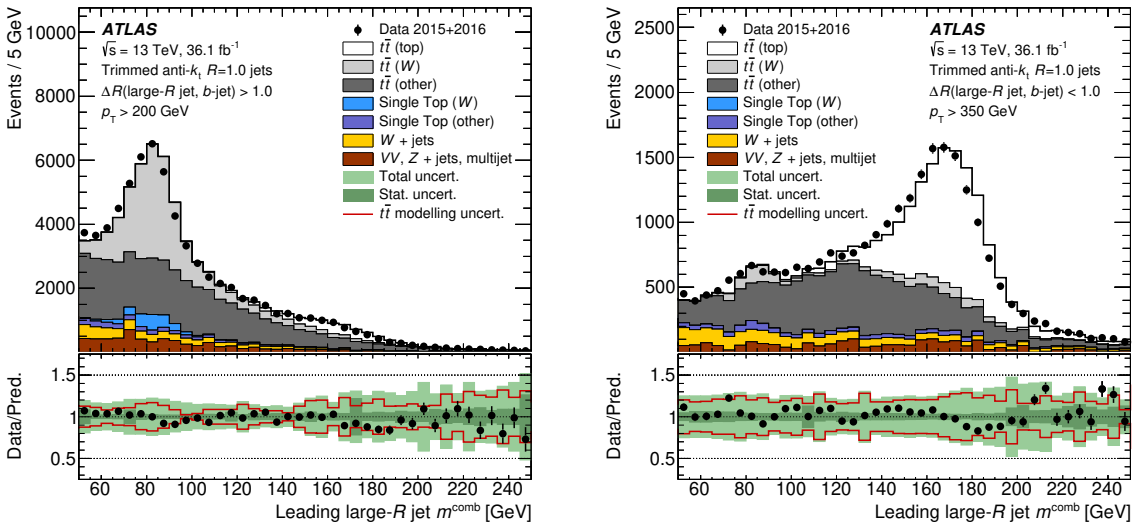
Signal modelling uncertainties

A number of systematic uncertainties on the signal modelling are considered, all of which are based on comparing various generator setups described in Sec. 6.1.1. Firstly, the uncertainty on the modelling of the PS and hadronisation is estimated as the difference between predictions from the nominal Powheg+Pythia sample and the sample generated using Powheg+Herwig. The uncertainty on the matching of the ME and PS is estimated by taking the difference between the nominal sample and the sample generated using MadGraph5_aMC@NLO+Pythia8. An uncertainty on the modelling of additional QCD radiation is estimated by comparing the nominal sample with a sample with simultaneously varied μ_R and μ_F scales and variation of h_{damp} parameter, as previously described in Sec. 6.1.1.

Finally, an uncertainty on the choice of PDF set in the signal sample is assessed using the PDF4LHC15 prescription [176] which combines several PDF sets with their uncertainties into a set of 30 one-sided uncorrelated variations, which are symmetrised into two-sided variations.

6.5 Modelling of jet substructure observables in data

A number of observables related to the boosted tagging techniques described in Sec. 5.3 are shown in Fig. 6.2–6.4. The systematic uncertainties shown in these figures in the uncertainty band are described in Section 6.4. For the distribution of the jet mass (Fig. 6.1) and substructure observables τ_{32} (Fig. 6.2a) and D_2 (Fig. 6.2b) the uncertainties on the τ_{32} and D_2 scale is also shown in the respective distributions. For simple taggers based on a selection on jet mass, τ_{32} and D_2 observables, it is possible to derive uncertainties on their signal efficiency and background rejection by propagating the uncertainties on the inputs. The jet mass distributions in Fig. 6.1 show a tension in the peak position between data and the prediction, which is attributed to a mismodelling of the MC jet mass scale as studied in in Ref. [115].

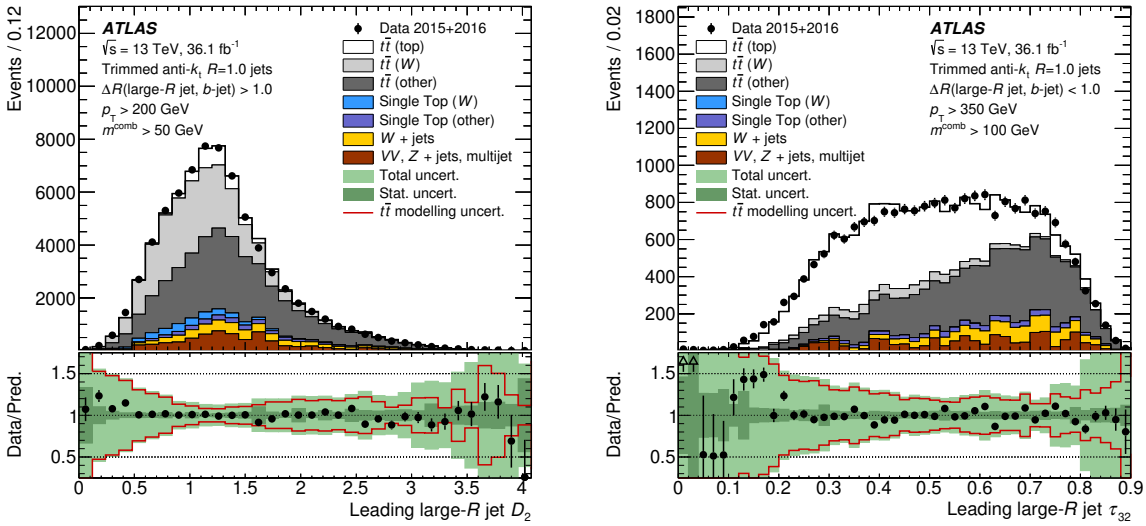


(a) (b)

Fig. 6.1: Comparison of data and prediction for the distribution of the selected large-R jet mass for the W-boson enriched (a) and top-quark-enriched (b) sub-samples. The MC-based predictions are normalised to data. The $t\bar{t}$ and single-top-quark predictions are further split according to the matching of the large-R jet to hadronically decaying top quark or W boson, or other (not matched to either top-quark or a W boson), based on the contained definition from Sec. 6.3. The data-to-prediction ratio shows the statistical uncertainty of the prediction (filled dark green band), the total uncertainty (filled light green band) combining systematic and statistical uncertainties in quadrature, and the impact of $t\bar{t}$ modelling systematic uncertainties (empty red band). The error bars on data points (black) show the statistical uncertainty of data.

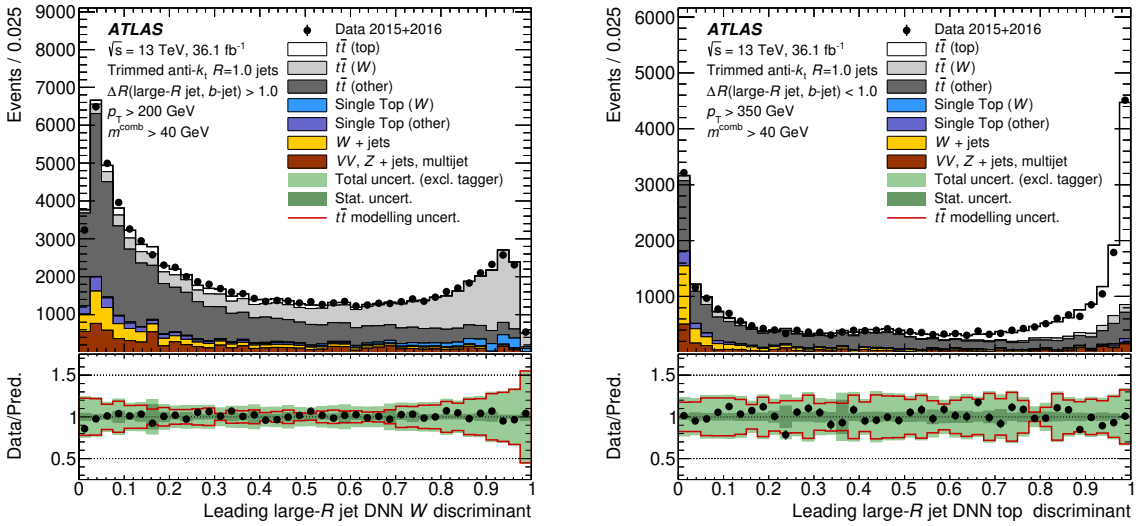
Additionally, more complicated observables are investigated, such as the high-level variable-based DNN taggers (Fig. 6.3), the TopoDNN (Fig. 6.4a) and the Shower Deconstruction's $log\,\chi$ discriminant (Fig. 6.4b). These approaches use a combination of a larger number of either high-level or low-level features, where no uncertainties on the inputs are propagated. The derivation of the uncertainties on the inputs is potentially challenging, and the correlations of a larger number of input observables may not be well known[(6)]. Instead, in Sec. 6.6, the signal efficiency of taggers is measured in-situ, which provides information about the accuracy of modelling of the inputs and their correlations on the studied tagger.

[(6)] The correlations can be studied in MC simulation, but there is no guarantee that the modelling of these correlations is sufficient without validation in real data.



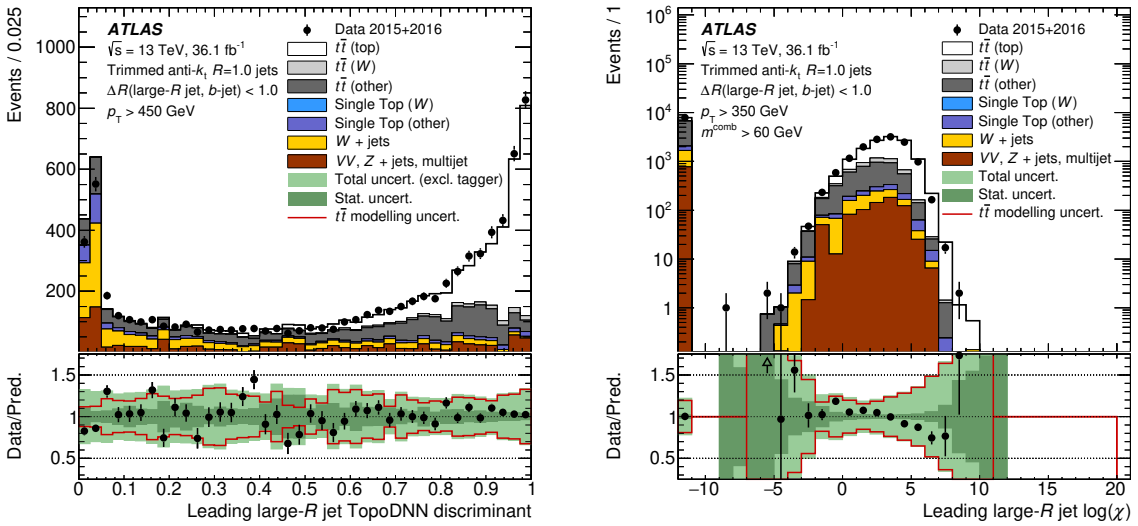
(a) (b)

Fig. 6.2: Comparison of data and prediction for the distribution of the selected large-R jet D_2 for the W-boson enriched (a) and τ_{32} for the top-quark-enriched (b) sub-samples. The normalisation of the predictions and uncertainty bands follow the same convention as in Fig. 6.1. Additionally, the uncertainties on the scale of D_2 (a) and τ_{32} (b) are propagated in these distributions.



(a) (b)

Fig. 6.3: Comparison of data and prediction for the distribution of the selected large-R jet W-tagger DNN discriminant in W-boson enriched sub-sample (a) and the top-tagger DNN discriminant in the top-quark-enriched sub-sample (b). The normalisation of MC predictions and uncertainty bands follow the same convention as in Fig. 6.1.



(a) (b)

Fig. 6.4: Comparison of data and prediction for the distribution of the selected large-R jet topocluster-based DNN top-tagger discriminant (a) and the shower deconstruction top-tagger discriminant $log\,\chi$ (b), in the top-quark-enriched sub-sample. The selection of large-R jet $p_{\mathrm{T}} > 450\,\mathrm{GeV}$ in (a) is imposed as the topocluster-based tagger was optimised for higher-p_{T} jets than other top-taggers investigated. The jets in the first bin with $log\,\chi < -10$ in (b) end up with less than three re-clustered k_t subjets, thus failing one of the initial conditions of the Shower Deconstruction tagger. The normalisation of MC predictions and uncertainty bands follow the same convention as in Fig. 6.1.

The distributions are nevertheless found to show reasonably good agreement between the prediction and the data. The uncertainties are mostly dominated by theory uncertainties on the $t\bar{t}$ modelling. Additionally, uncertainties on the flavour tagging mis-identification rates are found to be non-negligible in the background-enriched regions, such as large-R jet mass below 100 GeV in Fig. 6.1 and the low-discriminant values of machine-learning-based taggers in Fig. 6.3 and 6.4a. Finally, the scale uncertainties on the large-R jet mass, τ_{32} and D_2 are non-negligible particularly in the tails of these distributions.

6.6 Signal efficiency template fit

The relatively high purity of the top-quark- and W-boson-enriched samples obtained by the selection in Sec. 6.3 allows for a measurement of the signal-tagging efficiency in data in these topologies. The efficiency is obtained from a measurement of the number of jets passing and failing the criteria of a particular tagging algorithm in data and is compared to the predicted efficiency from MC simulation. From this data-to-MC comparison, it is possible to obtain a calibration scale factor, along with the uncertainties associated with the measured efficiency. These uncertainties can be then used in an independent analysis. This approach of uncertainty propagation is often referred to as *top-down*. The advantage of this approach is that the calibration corrects for the mismodelling of the input observables and their correlations on the signal efficiency.

For any tagger, the numbers of signal-like events in data that pass and fail the tagger selection, are obtained from a chi-square template fit of signal and background predicted distributions to the data. The templates are distributions of the selected large-R jet mass for events that pass (*pass* template) and

events that fail (*fail* template) the tagger selection, respectively. The templates are estimated using MC simulations and data-driven techniques outlined in Sec. 6.1.

The $t\bar{t}$ and single top-quark contributions to the templates are split according to the selected large-R jet label using the *contained* definition, as described in Sec. 6.3. To increase the stability of the fit, various background contributions with similar shapes are merged into a single template. For top-tagging, a six-template fit is performed, where the pass and fail templates are constructed for three different signal and background compositions:

- the signal template – $t\bar{t}$ events with the selected large-R jet matched to a hadronically-decaying top quark and its decay products,
- the $t\bar{t}$ background template – $t\bar{t}$ events failing the top-quark matching from previous point,
- the non-$t\bar{t}$ background template – all non-$t\bar{t}$ background processes considered.

In the case of the W-boson tagging, a fit of four templates is performed, where the pass and fail templates are constructed for a single signal and background composition:

- the signal template – $t\bar{t}$ events with the selected large-R jet matched to a hadronically-decaying W boson and its decay products, where the W boson is from the top-quark decay,
- the background template – $t\bar{t}$ events failing the matching from previous point and all other non-$t\bar{t}$ background processes considered.

In the fit, several constraints are imposed on the normalisation of the templates. The pass and fail template normalisations are fitted simultaneously. Additionally, the fitted number of background events that pass (fail) the tagger selection $N^{\text{pass}}_{\text{fit bckg}}$ ($N^{\text{fail}}_{\text{fit bckg}}$) are bound by their tagging efficiency that is kept fixed to the predicted value based on the pre-fit predicted number of background events that pass (fail) the tagger selection $N^{\text{pass}}_{\text{bckg}}$ ($N^{\text{fail}}_{\text{bckg}}$):

$$\epsilon^{\text{bckg}}_{\text{MC}} \equiv \frac{N^{\text{pass}}_{\text{bckg}}}{N^{\text{pass}}_{\text{bckg}} + N^{\text{fail}}_{\text{bckg}}} = \frac{N^{\text{pass}}_{\text{fit bckg}}}{N^{\text{pass}}_{\text{fit bckg}} + N^{\text{fail}}_{\text{fit bckg}}}. \tag{6.5}$$

The number of signal events that pass (fail) the tagger selection $N^{\text{pass}}_{\text{fit sig}}$ ($N^{\text{fail}}_{\text{fit sig}}$) are extracted from the fit. The signal efficiency in data is then defined as follows:

$$\epsilon_{\text{data}} = \frac{N^{\text{pass}}_{\text{fit sig}}}{N^{\text{pass}}_{\text{fit sig}} + N^{\text{fail}}_{\text{fit sig}}}. \tag{6.6}$$

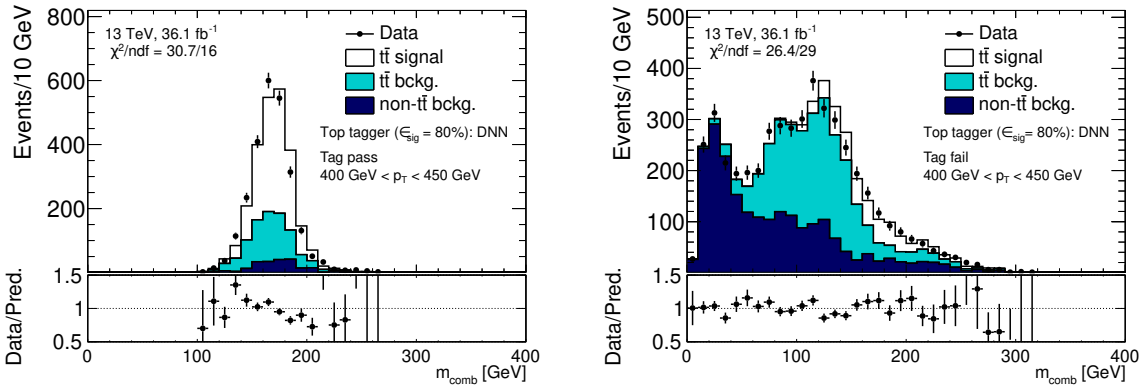
A comparison with the predicted signal efficiency from MC simulation can be performed, calculated from the pre-fit number of events where the selected large-R jet is labelled as signal, passing (failing) the tagger selection $N^{\text{pass}}_{\text{sig}}$ ($N^{\text{fail}}_{\text{sig}}$):

$$\epsilon_{\text{MC}} = \frac{N^{\text{pass}}_{\text{sig}}}{N^{\text{pass}}_{\text{sig}} + N^{\text{fail}}_{\text{sig}}}. \tag{6.7}$$

An example comparison of data and prediction after the template fit is shown in Fig. 6.5 for the high-level DNN top-quark tagger in the $p_{\text{T}} \in [400, 450]$ GeV bin. The disagreement in the ratio of the top-quark peak in Fig. 6.5a is of the same origin as in Fig. 6.1, due to missing in-situ calibration of the jet mass scale.

The method of extracting signal efficiency in data presented here relies on the assumption that the modelling of tagging efficiency for the non-$t\bar{t}$ backgrounds, is sufficiently accurate. Attempts to

design a dedicated control region for in-situ determination of the non-$t\bar{t}$ background tagging efficiency are outside of the scope of this work.



(a) (b)

Fig. 6.5: Distribution of jet mass of data and the fitted templates, as used in the χ^2 fit, for selected large-R jets that pass (a) and fail (b) the selection of the high-level DNN top-quark tagger, with $p_{\mathrm{T}} \in [400, 450]$ GeV. Only the statistical uncertainties of the data are shown in the bottom ratio plot.

6.7 Results

In the following figures, the efficiencies for a selected set of taggers measured in data ared compared to the MC prediction. The efficiencies are measured as a function of the selected large-R jet p_{T} as well as the average number of interactions per bunch crossing (μ) to asses the accuracy of modelling of the dependence of the taggers on various pile-up conditions. The emphasis is placed on the best-performing taggers based on the previous comparisons in MC simulation. This includes the DNN top-quark tagger (Fig. 6.6) and W-boson tagger (Fig. 6.7), the TopoDNN (Fig. 6.8) and the ShowerDeconstruction (Fig. 6.9) taggers.

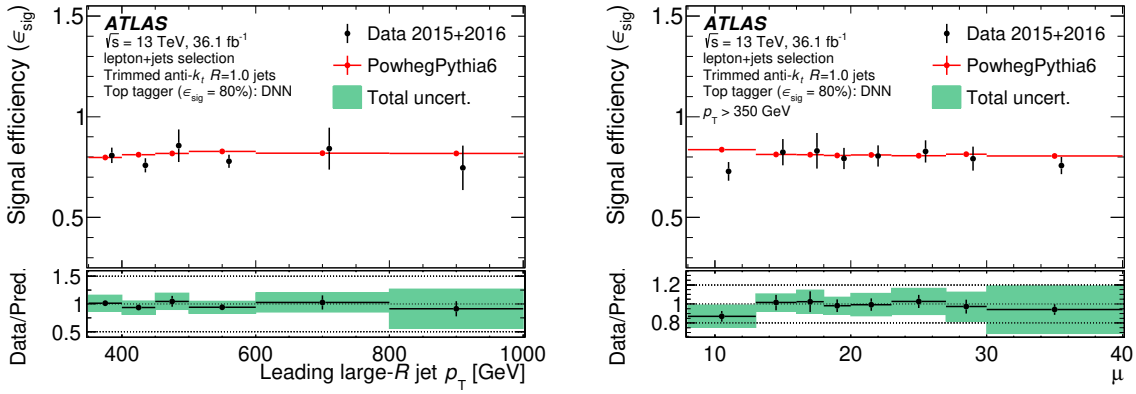
The signal efficiencies measured in data as a function of large-R jet p_{T} are found to be compatible with the predictions based on the $t\bar{t}$ MC simulation within the uncertainties. For W-boson tagging, a discrepancy is observed between the target 50 % efficiency, as designed in the optimisation of the tagger working points using the W' simulations, and the efficiency obtained from $t\bar{t}$ MC simulation as well as the measurement in data. The source of this discrepancy is attributed to event topology differences between the W-boson jets from $t\bar{t}$ and W-boson jets from decay of W', also observed in previous Run-I studies of W-tagging in data [112]. In particular, the W bosons from W' decays were found to be mostly longitudinally polarised, resulting in the $q\bar{q}'$ decay products being more balanced in momentum. In contrast, transversally polarised W bosons have a larger momentum imbalance of the decay products, making the topology more resembling background jets, which typically have a momentum imbalance in the radiation patterns due to the soft wide-angle emissions [177]. Approximately 31 % of the W bosons from top-quark decays are transversally polarised [178]. In addition, the W bosons from W' decays are well isolated, whereas the W bosons obtained from $t\bar{t}$ decays are potentially close to additional hadronic activity from the b-jet from top-quark decay which could impact the distributions sub-structure distributions of these jets.

The modelling of the pile-up dependency of the signal efficiencies is also found to be reasonably

good, even with respect to statistical uncertainty of the data only. This is an important positive result, given that that the systematic uncertainties are mostly correlated across the bin in μ and as such do not give an accurate message about the level of agreement. Most of the taggers are found to be robust against pile-up, with a somewhat more pronounced reduction in efficiency with increasing pile-up for the DNN W-boson tagger.

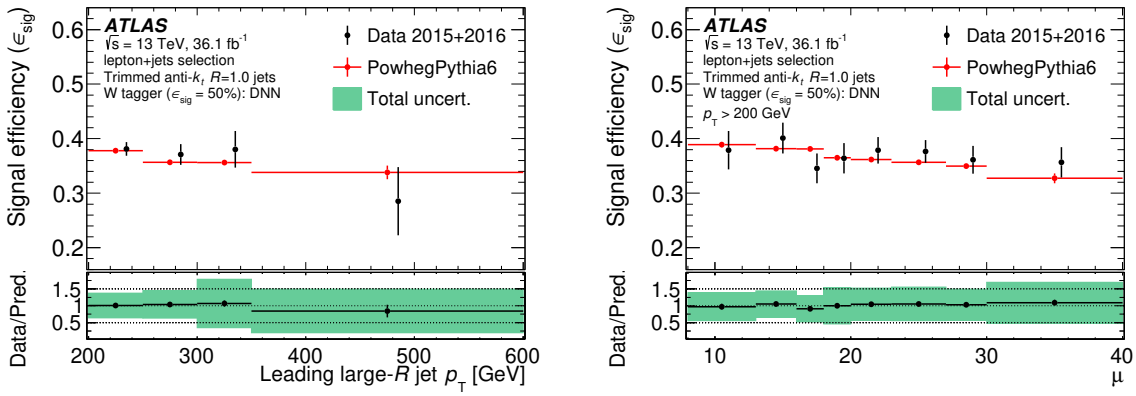
The systematic uncertainties reach up to approximately 20 % for top-quark tagging efficiencies and up to 50 % for W-boson tagging. They are dominated by modelling uncertainties arising from the implicit subtraction of non-signal $t\bar{t}$ backgrounds in the template fit, in particular due to the parton showering and hadronisation modelling uncertainties. This limitation is related to the contained signal definition which is based on the parton level information. This result motivates further work to define a different definition of top-quark and W-boson jets which would show lesser dependency on the choice of a particular $t\bar{t}$ generator. Experimental uncertainties found to have impact on the input distributions, such as b-tagging and large-R jet uncertainties, were found to impact the efficiencies up to 2 % and 5 % respectively, for highest-p_{T} bin considered.

To conclude, a study of the modelling of boosted top-quark and W-boson tagging techniques has been performed, using $\sqrt{s} = 13\,\mathrm{TeV}$ Run-II data with the integrated luminosity of $36.1\,\mathrm{fb}^{-1}$. Comparisons prediction with data of various sub-structure-related observables show good agreement within the considered uncertainties. The measurement of the signal efficiency of several different tagging algorithms shows good agreement between data and MC simulation prediction. The results presented demonstrate, that it is principle possible to derive in-situ calibration for arbitrarily complex boosted tagging algorithms by providing correction factors and their associated uncertainties, to account for mismodelling in the MC simulation. The precision of the efficiency measurement is largely limited by the modelling of $t\bar{t}$ simulations, prompting work on determining signal definition that is less generator-dependent. Finally, the signal efficiency is found to be quite robust against additional pile-up contamination in the event and the dependence is well described by the MC simulation.



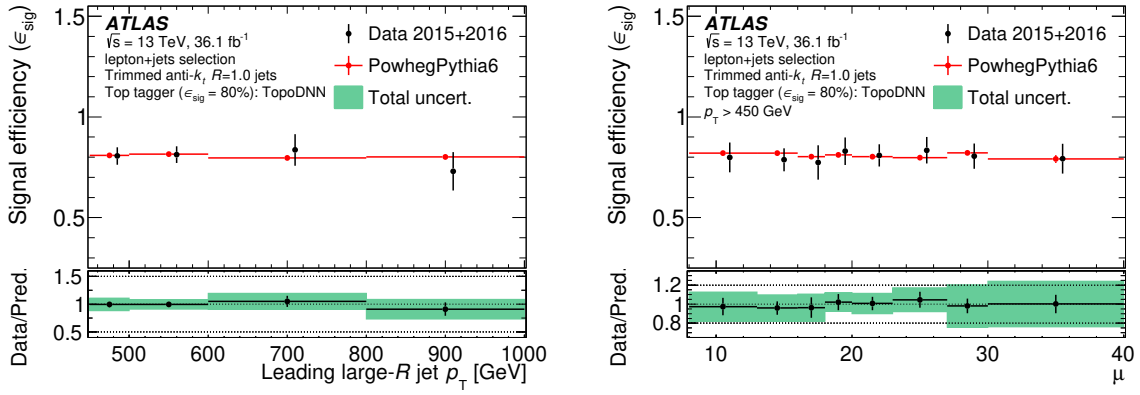
(a) (b)

Fig. 6.6: The signal efficiency measured in data and predicted by MC simulation, of the high-level DNN top-quark tagger, as a function of large-R jet p_{T} (a) and the average number of interactions per bunch crossing μ (b). The tagger is designed to give 80 % signal efficiency for contained top-quark jets, constant in large-R jet p_{T}. In the top plot, the error bars on the points show the statistical uncertainty of the data (black points) and the MC simulation (red points). The bottom plot shows the ratio of measured signal efficiency in data to the prediction from MC simulation, where the errors bars on black points show the statistical uncertainty of the ratio and the green band the total uncertainty. The total uncertainty is the statistical and systematic uncertainty summed in quadrature.



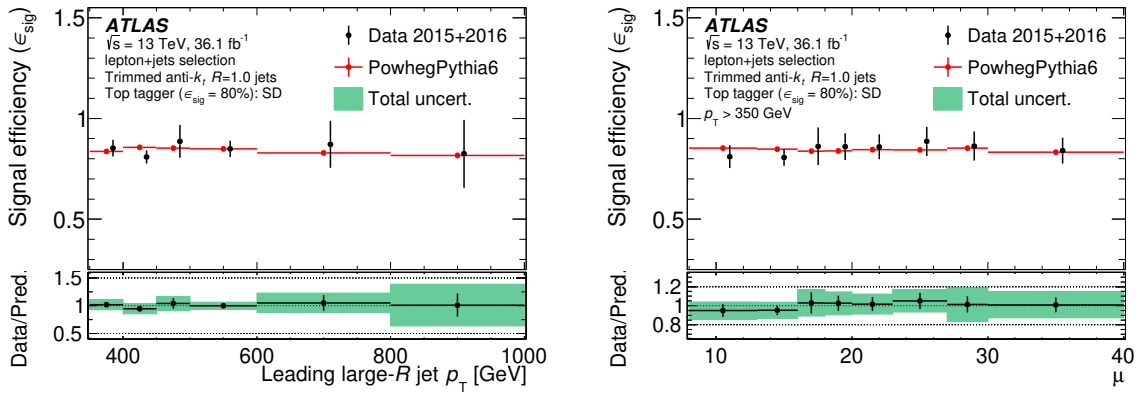
(a) (b)

Fig. 6.7: The signal efficiency measured in data and predicted by MC simulation, of the high-level DNN W-boson tagger, as a function of large-R jet p_{T} (a) and the average number of interactions per bunch crossing μ (b). The tagger is designed to give 50 % signal efficiency constant in large-R jet p_{T}. The definition of signal jets as well as the convention on the error bars in the plots is the same as in Fig. 6.6.



(a) (b)

Fig. 6.8: The signal efficiency measured in data and predicted by MC simulation, of the TopoDNN top-quark tagger, as a function of large-R jet p_{T} (a) and the average number of interactions per bunch crossing μ (b). The tagger is designed to give 80 % signal efficiency constant in large-R jet p_{T}. The definition of signal jets as well as the convention on the error bars in the plots is the same as in Fig. 6.6.



(a) (b)

Fig. 6.9: The signal efficiency measured in data and predicted by MC simulation, of the TopoDNN top-quark tagger, as a function of large-R jet p_{T} (a) and the average number of interactions per bunch crossing μ (b). The tagger is designed to give 80 % signal efficiency constant in large-R jet p_{T}. The definition of signal jets as well as the convention on the error bars in the plots is the same as in Fig. 6.6.

7

Charge asymmetry measurement in $t\bar{t}$ single-lepton channel

In this analysis, we measure the $A_C^{t\bar{t}}$ defined in Eq. 2.9 inclusively as well as differentially as a function of the $\beta_{z,t\bar{t}}$ and $m_{t\bar{t}}$. The actual observable measured is the binned distribution of $\Delta|y|$ from which the $A_C^{t\bar{t}}$ is calculated. For the differential $A_C^{t\bar{t}}$ measurements, the double-differential $\Delta|y|$ vs $\beta_{z,t\bar{t}}$ and $\Delta|y|$ vs $m_{t\bar{t}}$ observables are measured. The $\Delta|y|$ observable is corrected for background contributions, detector response and limited acceptance using the Fully-Bayesian unfolding method [179] described in detail in Sec. 7.4.

In contrast to the standalone resolved and boosted 8 TeV ATLAS measurements [69, 75], a combination of channels corresponding to the resolved and boosted topologies is performed, where the resolved topology refers to the standard method of reconstructing $t\bar{t}$ pair by matching individual jets to the corresponding quarks from the $t\bar{t}$ decay. The reconstruction of resolved and boosted $t\bar{t}$ pairs is detailed in Sec. 7.3. Additionally, based on the b-jet multiplicity, the events are split into $1b$-exclusive and $2b$-inclusive regions. The method of combination of the four regions in total is outlined in Sec. 7.4.2.

The analysis requires an optimisation of the $\Delta|y|$ observable binning to achieve unbiased unfolding response, which is briefly discussed in Sec 7.5. Systematic uncertainties impacting the measurement are specified in Sec. 7.6 and finally, the results are presented in Sec. 7.7.

7.1 Data and simulation samples

This analysis is performed using the full ATLAS Run-II dataset at $\sqrt{s} = 13$ TeV with an integrated luminosity of 139 fb^{-1}, collected in 2015-2018.

The MC and data-driven estimates of backgrounds mimicking the $t\bar{t}$ production are very similar to those in Ch. 6. The MC generators used for the simulations have been updated, however, and are described again in the following section. Additionally, the simulations are performed separately for the 2015-2016, 2017 and 2018 periods of data-taking, due to differences in the distribution of in-time pile-up (Fig. 3.3) and other luminosity-related beam conditions. The fake and non-prompt leptons background is estimated using the data-driven matrix-method described in Sec. 7.2.

All the generators in this analysis are interfaced with EvtGen v1.6.0 [180] generator to model the decays of heavy hadrons, with the exception of the samples generated with Sherpa [181]. For the simulation of multiple pp interactions per bunch crossing, the hard scattering events are overlaid with a set of minimum-bias interactions generated using Pythia8 [182] with the MSTW2008LO [146] PDF set and the A3 [183] set of tuned parameters.

7.1.1 Signal modelling

All of the $t\bar{t}$ samples are normalised to the inclusive production cross-section quoted in Sec. 2.1 and assume a top-quark mass of $m_{\text{top}} = 172.5\,\text{GeV}$[(1)]. The nominal $t\bar{t}$ MC sample is generated with the Powheg-Box v2 [151–153, 184] generator which calculates the production ME at NLO QCD precision, with the NNPDF3.0NLO [126] PDF set and the h_{damp} parameter set to $1.5\,m_{\text{top}}$ [165]. The ME generator is interfaced with Pythia8 for the PS and hadronisation, using the A14 [127] set of tuned parameters.

To estimate the uncertainties on the $t\bar{t}$ modelling, several alternative MC samples are used, based on prescriptions derived in Ref. [174]. Firstly, the uncertainty on the modelling of the initial-state radiation (ISR) is estimated from the comparison of the nominal sample with two variations. The "up" variation is obtained from a sample with the simultaneous variation of the μ_R and μ_F scales in the ME and the μ_R scale of the Var3c parameter in the A14 tune by a factor of 0.5, setting the h_{damp} parameter to $3\,m_{\text{top}}$. The down variation is obtained by simultaneously varying the ME μ_R and μ_F scales and the Var3c parameter μ_R scale by a factor of 2.0, keeping the h_{damp} parameter at the nominal $1.5m_{\text{top}}$ value [174]. The impact of the final-state radiation (FSR) is evaluated by varying the μ_R scale related to QCD-induced emissions in the PS, by a factor of 0.5 and 2.0 respectively.

The impact of the choice of PS and hadronisation model is evaluated by comparing the nominal $t\bar{t}$ signal sample with the nominal ME generator setup interfaced with Herwig7.0.4 [159, 185], using the Herwig7 default set of tuned parameters [185] and the MMHT2014LO PDF set [186].

To assess the uncertainty due to the choice of the matching scheme between the ME and the PS, MadGraph5_aMC@NLO [160] is used with the NNPDF3.0NLO PDF set, interfaced with Pythia8, using the A14 tune.

For estimation of the uncertainty on the value of the top-quark mass, samples are generated using the same setup as nominal sample, but with the top-quark mass set to 172 GeV, thus varying the top-mass by 0.5 GeV with respect to nominal value.

Finally, for alternative non-SM charge asymmetry predictions, the PROTOS generator [187] is used to simulate heavy axigluon models. The generator provides ME at LO accuracy in QCD, with the CTEQ6L1 [156] PDF set. The simulated samples contain only the parton-level information, which is used to reweight the nominal $t\bar{t}$ sample to induce BSM charge asymmetries. The reweighting is performed according to the ratio of the parton-level $\Delta|y|$ observable between the PROTOS and the nominal $t\bar{t}$ sample, where both $\Delta|y|$ predictions are normalised to the same yield, to only induce a change in the parton-level $\Delta|y|$ distribution shape.

7.1.2 Single top-quark modelling

The single top-quark tW production and the s- and t-channel production is modelled using the Powheg-Box v2 [151–153, 188–190] generator which calculates the ME up to the NLO QCD accuracy and uses the NNPDF3.0NLO PDF set. The ME generation is interfaced with Pythia8 using the A14 tune. The samples are normalised to the approximate NNLO cross-sections [147–150].

In this analysis, modelling uncertainties on the single top-quark production are also evaluated, using similar alternative generator variations as for the $t\bar{t}$ simulations. Additionally, specifically for

[(1)] The only exceptions are the top-quark mass-varied samples, which are generated with different values of the top-quark mass.

the tW production modelling, an overlap between tW and $t\bar{t}$ amplitudes must be removed. For this purpose, the diagram removal (DR) scheme [165, 191] is employed. An uncertainty on the $t\bar{t}$ overlap treatment is evaluated by comparing the sample using DR with a sample that uses diagram subtraction (DS) scheme [191].

7.1.3 Modelling of additional backgrounds

Both the V+jets as well as diboson production is simulated with the Sherpa v2.2 [181] generator. This generator includes both ME and PS generation [192]. The ME is calculated at NLO accuracy for up to two jets and LO accuracy for up to four jets [193–195]. The generator uses a set of tuned parameters from the Sherpa authors and the NNPDF3.0NNLO PDF set. The V+jets samples are normalised to the NNLO cross-section prediction [162], while the diboson samples are normalised to NLO QCD cross-section prediction [163].

The production of $t\bar{t}V$ and $t\bar{t}H$ events is modelled using the MadGraph5_aMC@NLO and Powheg-Box generators, respectively. The generators provide ME at NLO QCD, with the NNPDF3.0NLO PDF set. The ME generators are interfaced to Pythia8 using the A14 tune. The samples are normalised to NLO QCD+EW cross-section predictions [196, 197].

7.2 Estimation of the non-prompt and fake leptons background

The estimate of the fake and non-prompt leptons is performed using the same data-driven approach as in Sec. 6.2, with different ϵ_{real} and ϵ_{fake} efficiencies, measured using the 2015–2017 data. The efficiencies for 2017 are used for the fake estimate in the 2018 data as well. Different lepton isolation definitions were used in this updated efficiency measurement. The tight leptons use *gradient* isolation and loose leptons use no isolation criteria. Various efficiency parametrisations in two variables were tested, and those parametrisations that lead to smallest data-to-MC disagreement were preferred. Additionally, the number of negative $\Delta|y|$ bins of the estimated background is also considered, where efficiencies leading to smaller number of negative-yield bins are preferred. The best parametrisations are the same for resolved and boosted and ϵ_{real} and ϵ_{fake} efficiencies. The parametrisations are as follows:

- p_{T} of the leading jet, $\Delta\Phi(\ell, E_{\text{T}}^{\text{miss}})$ for the electrons,
- p_{T} of the lepton, p_{T} of the leading jet for the muons.

Additionally, alternative efficiency parametrisations, which were found to be similarly satisfactory in terms of the criteria outlined above, are used to derive a systematic uncertainty on the background estimate:

- p_{T} of the lepton, $\Delta\Phi(\ell, E_{\text{T}}^{\text{miss}})$ for the electrons,
- p_{T} of the leading jet, $\Delta\Phi(\ell, E_{\text{T}}^{\text{miss}})$ for the muons.

In the muon channel $2b$-inclusive boosted region, the contribution of fake and non-prompt leptons background is neglected, as its contribution is as small as $\approx 0.4\,\%$ of the total prediction, as shown in Table 7.2.

7.3 Event selection and reconstruction

Since the measurement in this analysis is performed in the single-lepton channel, a presence of a single isolated lepton is mandatory. Both the resolved and boosted topologies have a common pre-selection aimed to ensure a topology consistent with leptonic top-quark decay, outlined below. The selection criteria are similar to those used in Ch. 6, with differences in the lepton isolation criteria and selections imposed on $E_{\mathrm{T}}^{\mathrm{miss}}$ and M_{T}^{W} to ensure compatibility with the updated lepton efficiencies for the data-driven fake and non-prompt lepton background estimate.

Firstly, only events recorded under stable beam conditions and with all detector sub-systems operational are considered. A reconstructed primary vertex is necessary, and a single-electron or a single-muon trigger must be fired. The triggers used are the same as in Sec. 6.3, and the triggers used in 2016 are also used for the 2017 and 2018 period of data-taking. Following the trigger selection, the events are required to satisfy the following selection criteria:

- At least one *tight* electron or *medium* muon with $p_{\mathrm{T}} > 28\,\mathrm{GeV}$ and gradient isolation requirement is required. The p_{T} cut is slightly higher than the thresholds of the lowest lepton trigger to ensure that the trigger is fully efficient. Events containing additional leptons with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ are rejected in order to suppress multi-lepton backgrounds.
- In the electron channel, $E_{\mathrm{T}}^{\mathrm{miss}} > 30\,\mathrm{GeV}$ and the transverse mass of the W boson, $M_{\mathrm{T}}^{W} > 30\,\mathrm{GeV}$ is required. In the muon channel, $E_{\mathrm{T}}^{\mathrm{miss}} + M_{\mathrm{T}}^{W} > 60\,\mathrm{GeV}$ is required.
- At least one small-R calorimeter jet is required to be identified as originating from B-hadron (b-tagged) according to the selection described in Sec. 4.4.5.

After this pre-selection, the resolved and boosted topology selections are defined as follows. In the resolved topology selection, the following criteria are imposed:

- At least four small-R calorimeter jets with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 2.5$ are required.
- The $t\bar{t}$ system is reconstructed using a *boosted* decision tree (BDT) multivariate technique, which attempts to find the correct assignment of the four jets to the partons from $t\bar{t}$ decay in the single-lepton channel. The events are required to pass a cut on the output BDT classifier (> 0.3), which is related to the probability that the picked permutation of the jet-to-parton assignments is the correct one. The reconstruction is described in more detail in Sec. 7.3.1.
- If an event passes both the resolved topology selection as well as the boosted topology selection (below), it is removed from the resolved topology selection. This ensures the orthogonality of the selections, and is motivated by the fact that at high top-quark momenta, the boosted topology selection is more reliable at reconstruction compared to resolved topology reconstruction approach.

The boosted topology selection criteria are as follows:

- The presence of at least one small-R calorimeter jet with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 2.5$ is required, where the jet must be in the vicinity of the lepton candidate defined by $\Delta R(\mathrm{jet}, \ell) < 1.5$. In case of multiple jet matches, the highest-p_{T} jet is considered in further relevant selection criteria.
- At least one large-R jet with $p_{\mathrm{T}} > 350\,\mathrm{GeV}$ and $|\eta| < 2.0$ is required. The jet must be tagged as originating from top quark using a two-variable tagger applying selection on the jet mass and the τ_{32} substructure observable. The selection cuts are optimised to yield a constant 80 %

efficiency as a function of jet p_{T} to tag jets labelled as top-quark jets using the *inclusive* definition introduced in Ch. 5.

- The large-R jet candidate must be isolated from the small-R jet close-by the lepton, and must be back-to-back to the lepton. Specifically, ΔR(large-R jet, small-R jet) > 1.5 and $\Delta\phi$(large-R jet, ℓ) > 2.3.

The reconstruction of the $t\bar{t}$ system in the boosted topology is predominantly a task to reconstruct the leptonically-decaying top quark due to the presence of the neutrino. The hadronic top-quark four-momentum is the four-momentum of the isolated, top-tagged large-R candidate passing the boosted topology selection criteria. The reconstruction of the leptonic top-quark candidate four-momentum is performed using the four-momentum of the lepton and the close-by jet and $E_{\mathrm{T}}^{\mathrm{miss}}$ magnitude and direction in the transverse plane.

In both resolved and boosted topology, the neutrino is reconstructed using the constraint on the W boson mass $m_W = (p_\ell + p_\nu)^2$, where p_ℓ and p_ν are the four-momenta of the electron and neutrino, respectively. The p_ν^x and p_ν^y components of the neutrino momentum are determined from the $E_{\mathrm{T}}^{\mathrm{miss}}$ magnitude and the azimuthal angle. The neutrino mass can be neglected, and thus the m_W constraint leads to a quadratic equation where the longitudinal p_ν^z component is the unknown. If there are two solutions, the solution yielding a value of reconstructed leptonic top-quark mass closer to the value of 172.5 GeV is chosen. If no solution exists, the $E_{\mathrm{T}}^{\mathrm{miss}}$ value is decreased iteratively by a small amount to yield a solution.

7.3.1 Reconstruction of $t\bar{t}$ pairs in resolved topology

The reconstruction of $t\bar{t}$ pairs in the resolved topology is challenging due to the combinatorial complexity of assigning individual jets correctly to the corresponding partons from the $t\bar{t}$ decay. Various approaches to $t\bar{t}$ system reconstruction have been tested in the past. A popular choice in the ATLAS top physics precision measurements is the *KLFitter* [198]. The KLFitter is a likelihood-based algorithm for identification of correct jet-to-parton assignment. The likelihood uses constraints on the mass of the top quarks and the W bosons assuming a Breit-Wigner distribution for the constraints, accounting for detector response by using transfer functions that map the quark transverse energy distribution to the transverse energy of the jet. The discriminant to assess the permutations is the *event probability* which combines the kinematic-based likelihood with information about the b-tagging signal efficiency and background rejection to account for the probability of mis-identifying the b-jet candidates in the event.

In this analysis, the reconstruction is taken one step forward, by using a techniques based on BDT designed in Ref. [199] and re-trained for the selection used in this analysis. The BDT uses variables related to event kinematics, including the KLFitter event probability. The problem is formulated as a binary classification problem, where the individual permutations are evaluated by the BDT. A correct permutation assignment is considered signal, and incorrect permutations background. The BDT classifier can be understood as a likelihood of the permutation to be the correct one. The implementation of the BDT in the TMVA [200] package is used. Individual jet-to-parton permutations are evaluated using the BDT and the permutation with the highest BDT discriminant value is used for the $t\bar{t}$ reconstruction. Since the number of permutations increases roughly as $n!$ where n is the jet multiplicity in event, permutations are built using only up to five highest-p_{T} jets in event.

The training of the BDT is performed using SM $t\bar{t}$ events, using the signal and background

permutations, defined as follows. Each of the quarks from $t\bar{t}$ decay is geometrically matched to jet using $\Delta R < 0.3$ criterion. The signal permutation must have each of the quarks matched to exactly one jet. To ensure the BDT is not over-trained on statistical fluctuations of the training set, the sample of signal and combinatorial $t\bar{t}$ background events is split randomly into an equal-size training and a testing set, where only the training set is used to train the BDT. The trained BDT is evaluated on the testing set, assessing if significant performance differences are observed, such as different distribution of the BDT discriminant between training and the testing set, or a different separation power[(2)] observed in the two sets. The total number of signal permutations used in training and testing is 2×10^6 in $1b$-exclusive region and 1×10^6 in the $2b$-inclusive region. The number of background permutations used in the training and testing is 16×10^6 in $1b$-exclusive region and 20×10^6 in the $2b$-inclusive region.

The following observables are used as an inputs to the BDT:

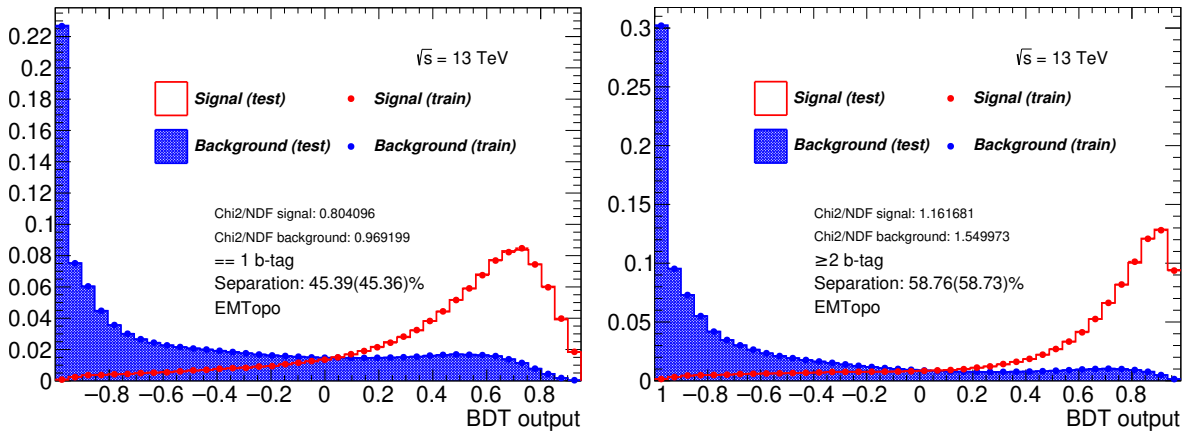
- the reconstructed mass of the hadronically-decaying top quark,
- the KLFitter event probability,
- the reconstructed mass of the hadronically-decaying W boson,
- b-tagging decision[(3)] for the candidate b-jet from the semileptonically-decaying top quark,
- b-tagging decision for the candidate b-jet from the hadronically-decaying top quark,
- b-tagging decision for the light jet with the highest b-tagging score from the hadronically-decaying W boson,
- the reconstructed mass of the semileptonically-decaying top quark,
- the ΔR between b-jet from the semileptonically-decaying top quark and the lepton,
- the ΔR between the two light jets from the hadronically-decaying W decay,
- the p_{T} of the lepton and b-jet from the semileptonically-decaying top quark,
- the number of jets in the event,
- the pseudorapidity of the hadronically-decaying top quark,
- and the ΔR between the two b-jets from the $t\bar{t}$ decay.

Further details on the input variables, the training and testing of the BDT are outlined in Ref. [199].

The distribution of the BDT discriminant for the signal and $t\bar{t}$ combinatorial background as determined using the matching of partons to jets, is shown in Fig. 7.1. In the analysis, a cut on the BDT discriminant is performed to suppress non-$t\bar{t}$ backgrounds as well as the combinatorial $t\bar{t}$ background. The choice of cut on BDT discriminant > 0.3 was optimised to minimise the statistical uncertainty of the unfolding. Systematic uncertainties were neglected in the optimisation due to the computational complexity of the full analysis chain required. The chosen cut on the BDT discriminant results in a signal to non-$t\bar{t}$ background ratio enhancement by a factor of ≈ 2 and in 75 % results in correct jet-to-parton assignment in the $t\bar{t}$ reconstruction.

[(2)] The separation power quantifies the discrimination power of an observable to distinguish two classes of events, referred to here as signal and background. For binned distributions, it is defined as $\frac{1}{2}\sum_i \frac{(s_i - b_i)^2}{(s_i + b_i)}$, where s_i and b_i are the signal and the background yields in i^{th} bin of the observable.

[(3)] For each permutation, the assigned jet is checked whether it passed the *MV2c10* b-tagging criterion or not.



(a) (b)

Fig. 7.1: Comparison of BDT score for correct (signal) and incorrect (background) jet-to-parton assignment in the 1b-exclusive (a) and 2b-inclusive (b) region, determined using the parton-level matching. The distributions are shown for both the training and the testing set, showing good agreement without signs of overtraining.

7.3.2 Signal and background contributions and comparisons to data

The event yields of the individual signal and background processes and the data yields obtained after the selection described in previous section are shown in Table 7.1 and 7.2 for the resolved and boosted topologies, respectively, split in b-tagging multiplicity and shown separately for the electron and muon channel.

In Fig. 7.2- 7.4, the comparison of data to prediction is shown for the observables that are subsequently unfolded, the $\Delta|y|$ and the double-differential $\Delta|y|$ vs $m_{t\bar{t}}$ and $\Delta|y|$ vs $\beta_{z,t\bar{t}}$ respectively. The data to prediction ratio in boosted regions shows the previously-discussed mismodelling resulting in an approximately 20 % over-estimation of prediction compared to data. A treatment for the normalisation discrepancy is discussed in Sec. 7.6.3.

Table 7.1: Event yields in the lepton+jets resolved channels, split by the lepton flavour (e, μ) and b-jet multiplicity (1b-excl., 2b-incl.). The total uncertainty on the prediction is shown, including the impact of systematic uncertainties.

Process:	Electron channel		Muon channel	
	1b-excl.	2b-incl.	1b-excl.	2b-incl.
$t\bar{t}$	700000 ± 60000	850000 ± 70000	820000 ± 70000	1000000 ± 80000
Single top	41000 ± 6000	22000 ± 4000	48000 ± 7000	27000 ± 5000
W + jets	90000 ± 50000	10000 ± 6000	120000 ± 60000	13000 ± 8000
Z + jets	21000 ± 11000	3300 ± 1800	16000 ± 9000	2800 ± 1600
Diboson	4400 ± 2300	600 ± 400	5200 ± 2800	800 ± 500
$t\bar{t}V, t\bar{t}H$	2500 ± 1300	3300 ± 1800	2800 ± 1400	3700 ± 2000
Fake lep.	71000 ± 35000	36000 ± 18000	18000 ± 9000	11000 ± 6000
Total pred.	920000 ± 100000	920000 ± 80000	1030000 ± 110000	1050000 ± 90000
Data (139 fb^{-1})	898484	922567	1046553	1086959

Table 7.2: Event yields in the lepton+jets boosted topology, split by the lepton flavour (e, μ) and b-jet multiplicity (1b-excl., 2b-incl.). The total uncertainty on the prediction is shown, including the impact of systematic uncertainties.

Process:	Electron channel		Muon channel	
	1b-excl.	2b-incl.	1b-excl	2b-incl.
$t\bar{t}$	24000 ± 4000	36000 ± 5000	25000 ± 4000	39000 ± 6000
Single top	1800 ± 500	1400 ± 600	1900 ± 700	1500 ± 700
W + jets	4400 ± 2400	800 ± 500	5200 ± 2800	1000 ± 500
Z + jets	550 ± 300	120 ± 70	580 ± 320	130 ± 80
Diboson	410 ± 240	100 ± 70	430 ± 250	100 ± 90
$t\bar{t}V$, $t\bar{t}H$	300 ± 180	490 ± 260	280 ± 170	510 ± 290
Fake lep.	3000 ± 1500	2300 ± 1200	~0.7%	~0.4%
Total pred.	35000 ± 6000	41000 ± 6000	34000 ± 6000	42000 ± 6000
Data (139 fb^{-1})	26999	32155	27711	34427

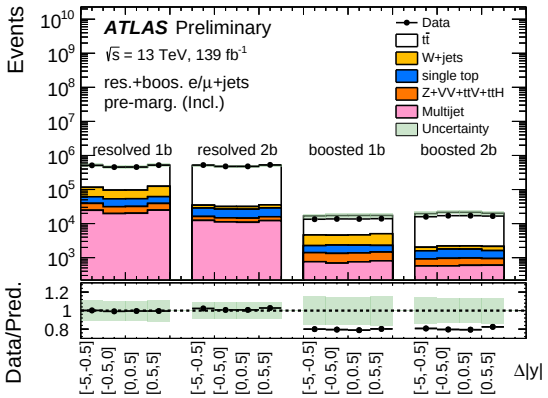
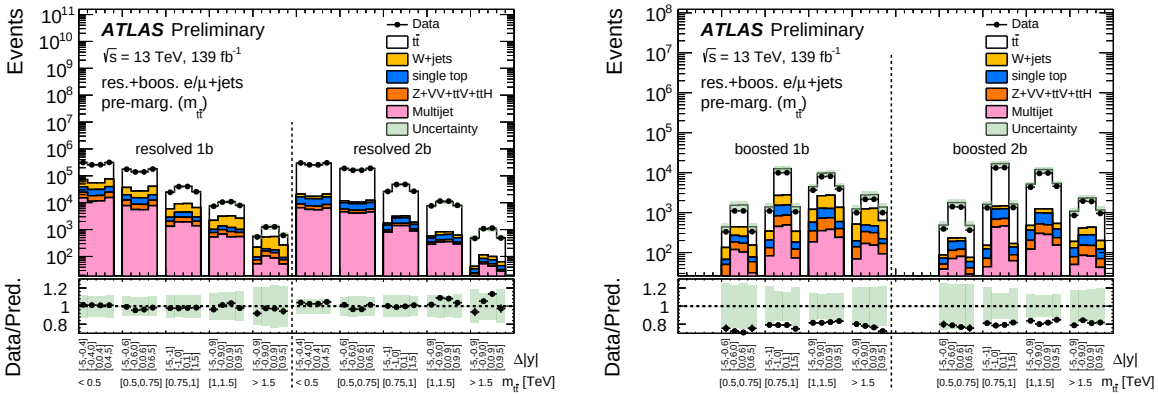
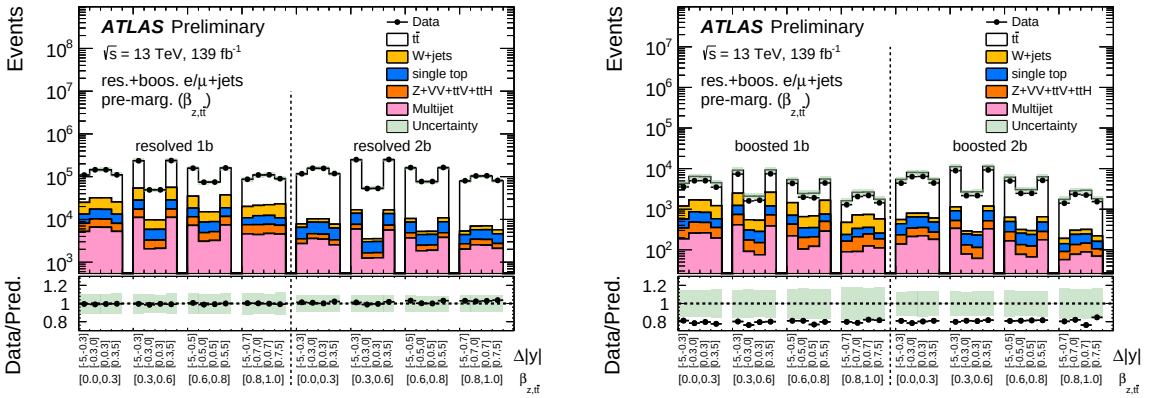


Fig. 7.2: The comparison of data to prediction for the $\Delta|y|$ distribution [2]. The x-axis labels show the bin edges of the $\Delta|y|$ bins as optimised for the unfolding. All four regions are shown in the plot, denoted by the labels above the distributions. The black points in the bottom plot show the ratio of data to total prediction, with the green band showing the total uncertainty, including both statistical and systematic uncertainties on the prediction in quadrature. The Z+VV+ttV+ttH entry in the legend corresponds to a merged prediction of Z+jets, diboson, $t\bar{t}V$ and $t\bar{t}H$ processes.



(a) (b)

Fig. 7.3: The comparison of data to prediction for the double-differential $\Delta|y|$ vs $m_{t\bar{t}}$ distribution [2], for resolved regions (a) and boosted regions (b). The upper x-axis labels show the bin edges of the $\Delta|y|$ bins, while the bottom x-axis labels show the $m_{t\bar{t}}$ binning. The plot style, content and uncertainties follow the convention from Fig. 7.2.



(a) (b)

Fig. 7.4: The comparison of data to prediction for the double-differential $\Delta|y|$ vs $\beta_{z,t\bar{t}}$ distribution [2], for resolved regions (a) and boosted regions (b). The upper x-axis labels show the bin edges of the $\Delta|y|$ bins, while the bottom x-axis labels show the $\beta_{z,t\bar{t}}$ binning. The plot style, content and uncertainties follow the convention from Fig. 7.2.

7.4 Fully-Bayesian unfolding

7.4.1 Motivation of the unfolding problem

The goal of this measurement is to measure the charge asymmetry and to compare the results with fixed-order theory predictions. However, the true charge asymmetry in the $t\bar{t}$ pair production is diluted due to detector response and limited acceptance as well as contributions from background processes. It is therefore necessary to correct the measured asymmetry for these effects to make a direct comparison with the theory prediction. Let us first formulate the general problem at hand. The objective is to determine a binned true distribution of the observable of interest, from an observed binned detector-level distribution of the observable of interest. The relation between the true and the detector-level binned distributions can be related through the following equation:

$$\mathbf{D} = \mathbf{S} + \mathbf{B} \equiv \mathcal{M} \cdot \mathbf{T} + \mathbf{B}, \tag{7.1}$$

where the $\mathbf{D}$[(4)] is the detector-level (reconstructed) data distribution, $\mathbf{B}$ is the distribution of the detector-level background processes contribution, $\mathbf{S} = \mathcal{M} \cdot \mathbf{T}$ is the detector-level signal distribution, $\mathbf{T}$ is the *unknown true distribution* of interest, and $\mathcal{M}$ is the response matrix that encodes the detector response and acceptance. In the equation above, $\mathbf{T}$ is the distribution that we want to determine from data $\mathbf{D}$. The solution to this problem also requires estimation of the backgrounds $\mathbf{B}$ as well as the response matrix $\mathcal{M}$.

The response matrix is typically estimated from MC simulation, where both information of the underlying true distribution as well as the reconstructed distribution is available, giving the ability to simulate the migrations and the acceptance. The response matrix encodes the following information:

$$\mathcal{M}_{ij} = \mathfrak{M}_{ij}\epsilon_j \text{ (no summation over index } j \text{ is performed here)}, \tag{7.2}$$

where:

- $\mathfrak{M}_{ij}$ is the normalised migration matrix, that gives the probability that a reconstructed event from true distribution bin j migrated to reconstructed distribution bin i
- ϵ_j is the acceptance; the probability that event from true distribution bin j is reconstructed and passes event selection criteria.

Assuming we have the estimate of background processes, and the response matrix, the task of unfolding is to recover the true distribution from the reconstructed distribution. A naive solution by using matrix inversion in Eq. 7.1 can rarely be used, as the original problem is ill-defined. The inversion of the matrix can amplify statistical fluctuations in data as well as in the response matrix, resulting in a wildly oscillating solution. There are many unfolding methods that take different approaches to solving this numerical problem. We will focus on the method, that is used in this thesis, the Fully-Bayesian unfolding (FBU) [179].

The FBU is a method that performs unfolding by applying Bayes' theorem, yielding a full posterior probability distribution of the unfolded spectra. It employs likelihood formalism, allowing for a natural inclusion of systematic uncertainties as nuisance parameters. Via Bayesian marginalisation, the inclusion of systematic uncertainties allows to reduce total uncertainty by accounting for correlations between the nuisance parameters and by constraining the nuisance parameters.

7.4.2 Application of Bayesian inference

The unfolding problem illustrated in previous section can be written down in terms of the Bayes' theorem. We have measured the data distribution $\mathbf{D}$ of the observable of interest. Let us assume that we have an estimate of total background contribution $\mathbf{B}$, composed of N_b background processes, and we have the knowledge of detector and reconstruction response encoded in a response matrix $\mathcal{M}$. We would like to know, what is the probability of having a true distribution $\mathbf{T}$ given the observed data $\mathbf{D}$. In terms of the Bayes' theorem:

$$P(\mathbf{T}|\mathbf{D}, \mathcal{M}) \propto \mathcal{L}(\mathbf{D}|\mathbf{S}(\mathbf{T}, \mathcal{M}), \mathbf{B})\pi(\mathbf{T}), \tag{7.3}$$

where $P(\mathbf{T}|\mathbf{D}, \mathcal{M})$ is the posterior probability of the true distribution $\mathbf{T}$, $\mathcal{L}(\mathbf{D}|\mathbf{S}(\mathbf{T}, \mathcal{M}), \mathbf{B})$ is the likelihood function of $\mathbf{D}$ given $\mathbf{S}(\mathbf{T}, \mathcal{M})$ and $\mathbf{B}$, and $\pi(\mathbf{T})$ is the prior probability density for the true

[(4)] A bold symbol denotes a one-dimensional vector, such as a one-dimensional binned distribution.

spectrum $\mathbf{T}$. The spectra are represented by binned histograms with N_r reconstructed bins and N_t true bins, respectively, i.e. $\mathbf{D}, \mathbf{S} \in \mathbb{R}^{N_r}$, $\mathbf{T} \in \mathbb{R}^{N_t}$ and $\mathcal{M} \in \mathbb{R}^{N_t} \times \mathbb{R}^{N_r}$. By sampling the prior probability distribution $\pi(\mathbf{T})$ and propagating the values of the sampled $\mathbf{T}$ through the likelihood, it is possible to obtain the full posterior distribution of the true spectrum $\mathbf{T}$. Note that no matrix inversion is actually performed in this approach. In the following sections, we will explain how the $\pi(\mathbf{T})$ is chosen, what is the exact definition of the likelihood $\mathcal{L}(\mathbf{D}|\mathbf{S}(\mathbf{T}, \mathcal{M}), \mathbf{B})$, and how the sampling of the priors is performed.

Likelihood

The likelihood in Eq. 7.3 is constructed under the assumption that the data is distributed according to Poisson probability:

$$\mathcal{L}(\mathbf{D}|\mathbf{S}(\mathbf{T}, \mathcal{M}), \mathbf{B}) = \prod_{i=1}^{N_r} \text{Poisson}(d_i|s_i + b_i), \tag{7.4}$$

$$b_i = \sum_{k=1}^{N_b} b_i^k, \tag{7.5}$$

where s_i is the expected signal yield, b_i^k the expected k^{th} background yield, d_i the observed data yield, in i^{th} bin, respectively. The expected detector-level signal distribution $\mathbf{S}$ is given by the (unknown) true distribution $\mathbf{T}$ and the response matrix $\mathcal{M}$:

$$s_i(\mathbf{T}, \mathcal{M}) = \sum_{j=1}^{N_r} \mathcal{M}_{ij} t_j f_j^{-1}, \tag{7.6}$$

where t_j is the j^{th} bin yield of the true distribution, and the $f_j \leq 1$ is an out-of-fiducial correction, accounting for events which are reconstructed, but do not pass the selection imposed at the true level (*fiducial* cuts). In this measurement, the unfolding is performed to full phase space, in other words there are no fiducial cuts and therefore $f_j = 1$.

Prior probability

The prior probability density $\pi(\mathbf{T})$ should be based on what we know about $\mathbf{T}$ before the measurement is performed. The simplest possible choice is a flat, so-called "uninformative" prior, which is a bounded uniform distribution given by minimum and maximum vale t_i^{min} and t_i^{max} in each of the bins of the true distribution:

$$\pi(\mathbf{T}) \propto \begin{cases} 1 & \text{if } t_i \in [t_i^{\text{min}}, t_i^{\text{max}}], \forall i \in [1, N_t] \\ 0 & \text{otherwise} \end{cases} . \tag{7.7}$$

The "uninformative" prior thus makes no preference for any choice of $\mathbf{T}$ as long as it is within the specified bounds. It is possible to extend this choice of prior with additional information, effectively

introducing further regularisation[(5)], defined by a function $S(\mathbf{T})$:

$$\pi(\mathbf{T}) \propto \begin{cases} e^{\alpha S(\mathbf{T})} & \text{if } t_i \in [t_i^{\min}, t_i^{\max}], \forall i \in [1, N_t] \\ 0 & \text{otherwise} \end{cases}, \tag{7.8}$$

where α is a free parameter impacting the strength of the regularisation. This enables the use of additional a priori made assumptions about $\mathbf{T}$ to constrain the parameter space, potentially decreasing the variance of the unfolding, however at the cost of an additional bias.

In this analysis, an uninformative prior from Eq. 7.7 is used, where $t_i^{\min} = 0$ and $t_i^{\max} = 2t_i^{\text{MC}}$, where t_i^{MC} is the predicted true yield in i^{th} bin of the $\Delta|y|$ distribution, based on the SM $t\bar{t}$ MC prediction used in this analysis. In other words, the bounds of the uniform are given by a $\pm 100\,\%$ interval around the MC-based prediction. It is found that no additional regularisation function is necessary to eliminate spurious fluctuations in the unfolded true distribution. The choice of the uninformative prior is also motivated by the fact that there is no previous unfolded measurement of the $\Delta|y|$ observable at 13 TeV. Therefore the prior knowledge of the observable is limited to MC predictions which have been compared to unfolded data distributions of other observables but not the $\Delta|y|$ observable. Other intervals of the choice of the prior interval were also tested, both smaller and larger, and were found to have no impact on the unfolded result.

Inclusion of nuisance parameters in FBU

The likelihood formalism employed in FBU allows for a natural inclusion of systematic uncertainties as nuisance parameters (NPs), that encode the imperfect knowledge of various parameters in the model of the analysis. The systematic uncertainties affect the detector-level distributions of both signal and background contributions, in other words their impact is quantified by an alternative distribution of the detector-level observable with generally different shape and/or yield. In FBU, the likelihood $\mathcal{L}(\mathbf{D}|\mathbf{S}(\mathbf{T}, \mathcal{M}), \mathbf{B})$ is extended by the NPs into a marginal likelihood defined as:

$$\mathcal{L}(\mathbf{D}|\mathbf{S}(\mathbf{T}, \mathcal{M}), \mathbf{B}) = \int \mathcal{L}(\mathbf{D}|\mathbf{S}(\mathbf{T}, \mathcal{M}; \boldsymbol{\theta}), \mathbf{B}(\boldsymbol{\theta}))\pi(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}, \tag{7.9}$$

where $\boldsymbol{\theta}$ denotes all the NPs and $\pi(\boldsymbol{\theta})$ their priors. The priors are probability distribution functions of the parameters of the model that encompass the information about *auxiliary* measurements in which the unknown parameters of the model are determined with a limited precision that defines the systematic uncertainty. There are many various NPs that can be implemented in the marginal likelihood. The individual systematic uncertainties affecting the measurement are discussed in detail in Sec. 7.6. Here we discuss the types of NPs that are considered.

Background normalisations $\boldsymbol{\theta}_b$, which only affect the respective background prediction. The prior for these uncertainties is Gaussian with $\mu = 0$ and $\sigma = 1$, truncated to prevent negative yield. These uncertainties introduce only a shift in total yield of the background contribution.

Uncertainties due to object reconstruction, identification and calibration, $\boldsymbol{\theta}_r$, impacting both signal $\mathbf{S}(\mathbf{T}, \mathcal{M}; \boldsymbol{\theta}_r)$ and background $\mathbf{B}(\boldsymbol{\theta}_r, \boldsymbol{\theta}_b)$. These uncertainties are usually assumed to be Gaussian, providing a $\pm 1\sigma$ variations on the detector-level observable of interest. The prior is a Gaussian with

[(5)] It should be noted, that the bounded uniform prior contains in itself regularisation, since it by definition puts constraints on the possible values of the true spectrum.

$\mu = 0$ and $\sigma = 1$. Theoretical uncertainties on the MC predictions are also commonly introduced in this manner.

The bins of the detector-level signal distribution $\mathbf{S}(\mathbf{T}, \mathcal{M}; \boldsymbol{\theta}_r)$ with included NPs are defined as follows:

$$s_i(\mathbf{T}, \mathcal{M}; \boldsymbol{\theta}_r) = s_i(\mathbf{T}, \mathcal{M}; \mathbf{0}) \left(1 + \sum_k \theta_r^k \Delta s_i^k\right), \tag{7.10}$$

where $s_i(\mathbf{T}, \mathcal{M}; \mathbf{0})$ is defined in (7.6) and Δs_i^k is the relative systematic uncertainty variation on the signal yield in the i^{th} bin corresponding to the k^{th} nuisance parameter θ_r^k. The $\theta_r^k \Delta s_i^k$ terms assume a linear interpolation of the uncertainty variation in-between the $\pm 1\sigma$ uncertainty bands and a linear extrapolation beyond this interval.

For each background, the respective background prediction is altered in a similar manner:

$$b_i^k(\boldsymbol{\theta}_r, \boldsymbol{\theta}_b) = b_i^k(\mathbf{0})(1 + \theta_b^k \Delta b^k) \left(1 + \sum_j \theta_r^j \Delta b_i^{k,j}\right), \tag{7.11}$$

where $b_i^k(\mathbf{0})$ is the predicted yield of the k^{th} background in i^{th} bin, Δb^k is the relative uncertainty on the background normalisation and $\Delta b_i^{k,j}$ is the relative systematic uncertainty variation on the k^{th} background yield in the i^{th} bin corresponding to j^{th} nuisance parameter θ_r^j.

Statistical uncertainties on the background prediction. The prime example of this uncertainty is the MC prediction uncertainty due to limited number of generated events, where the true prediction in general differs from the generated prediction due to statistical fluctuations. To account for this effect, an approach inspired by a proposal from Barlow and Beeston in [201] is used, where each bin i of the total background prediction $\mathbf{B}$ receives an additional nuisance parameter γ_i, which has a flat prior and a Poisson constraint on the statistical uncertainty of $\mathbf{B}$. The full likelihood term is defined as:

$$\mathcal{L}(\mathbf{D}|\mathbf{S}(\mathbf{T}, \mathcal{M}; \boldsymbol{\theta_r}), \mathbf{B}(\boldsymbol{\theta_r}; \boldsymbol{\theta_b}), \boldsymbol{\gamma}) = \prod_i^{N_r} \text{Poisson}(d_i | s_i(\mathbf{T}, \mathcal{M}; \boldsymbol{\theta_r}) + \gamma_i b_i(\boldsymbol{\theta_r}, \boldsymbol{\theta_b})) \text{Poisson}(\tau_i | \gamma_i \tau_i) \tag{7.12}$$

$$\gamma_i b_i(\boldsymbol{\theta_r}, \boldsymbol{\theta_b}) = \gamma_i \sum_{k=1}^{N_b} b_i^k(\boldsymbol{\theta_r}, \theta_b^k)$$

with the $s_i(\mathbf{T}, \mathcal{M}; \boldsymbol{\theta_r})$ defined in Eq. 7.10 and $b_i^k(\boldsymbol{\theta_r}, \theta_b^k)$ defined in Eq. 7.11. In the constraint term, $\tau_i = (b_i/\delta b_i)^2$, where b_i is defined in Eq. 7.5 and δb_i is the statistical uncertainty of b_i. In this configuration, the $\boldsymbol{\gamma}$ NPs are distributed according to Gamma probability distribution function, and the total background yield $\gamma_i b_i$ in i^{th} bin is allowed to fluctuate around the nominal value b_i at a penalty introduced by the Poisson constraint.

Finally, the nuisance parameters are not the parameters of interest, only the true distribution $\mathbf{T}$ is. Hence, in the Bayesian marginalisation, the nuisance parameters are integrated out, or marginalised. The marginal likelihood is defined as:

$$\mathcal{L}(\mathbf{D}|\mathbf{S}(\mathbf{T},\mathcal{M}),\mathbf{B}) = \int \mathcal{L}(\mathbf{D}|\mathbf{S}(\mathbf{T},\mathcal{M};\boldsymbol{\theta}_r),\mathbf{B}(\boldsymbol{\theta}_r,\boldsymbol{\theta}_b);\boldsymbol{\gamma})\ \pi(\boldsymbol{\theta}_r)\pi(\boldsymbol{\theta}_b)\pi_{\text{flat}}(\boldsymbol{\gamma})\mathrm{d}\boldsymbol{\theta}_r\mathrm{d}\boldsymbol{\theta}_b\mathrm{d}\boldsymbol{\gamma}. \quad (7.13)$$

Combining Eq. 7.3 and 7.13, the marginalised posterior probability density for a true distribution, $\mathbf{T}$, is obtained numerically using random sampling in the $N_t + N_{\text{NP}}$ parameter space and projecting the samples into the $\mathbf{T}$-parameter space, where N_{NP} is the total number of nuisance parameters.

Similarly, marginalised posterior distributions of the individual NPs can also be obtained, by projecting all other parameters into the one-dimensional space of the NP of interest. An example marginalised posterior of a NP from unfolding of the Asimov dataset[6] is shown in Fig. 7.5. Typically, the mean and the variance of the marginalised posterior of the NP are interpreted as the marginalised NP pull and constraint, and commonly, though not necessarily, the marginalised posterior distribution is Gaussian. The meaning of marginalised pull and constraint should not however be assumed to be equivalent to the meaning of pull and constraint in profile-likelihood formalism, because the marginalisation process includes the effect of correlations with the other marginalised parameters, which is not the case of profile-likelihood formalism, where a pull or a constraint of a parameter is simply the projected value of the parameter corresponding to the maximum of the likelihood. A Bayesian estimate equivalent to the profile likelihood is the *maximum a posteriori* (MAP) if the choice of the true distribution prior is uniform, which is the case in this analysis. In this case, the MAP estimate is basically the mode of the likelihood, obtained via minimization similar to profile likelihood formalism. Thus, for posterior distributions or results, where MAP estimate is quoted, such estimate is obtained by minimising the negative logarithm of the likelihood in Eq. 7.12. If the observable in the likelihood is sensitive to a particular uncertainty, the corresponding nuisance parameter can be further constrained from the data. The constraint manifests itself by narrower width of the posterior distribution compared to the prior distribution. Constraints reduce the uncertainty on the observable of interest, one of the main motivations to use the Bayesian marginalisation. Additionally, likelihood formalism allows to take into account the NP correlations in the marginalisation, leading to further reduction of the total uncertainty. An estimate of the expected constraints can be obtained by unfolding the Asimov dataset.

Relation between A_{C} and $\Delta|y|$ observables in FBU

In the charge asymmetry measurement, the mean of the posterior for each bin of $\mathbf{T}$ corresponds to the central value in the corresponding bin of the unfolded $\Delta|y|$ distribution. However, the quantity of interest is the charge asymmetry A_{C}. Formally, the relation between A_{C} posterior and the $\mathbf{T}$ posterior is defined by:

$$p(A_{\text{C}}|\mathbf{D}) = \int d\mathbf{T}\,\delta(A_{\text{C}} - A_{\text{C}}(\mathbf{T}))\,P(\mathbf{T}|\mathbf{D}), \quad (7.14)$$

where $A_{\text{C}}(\mathbf{T})$, is the definition of charge asymmetry from Eq. 2.9, δ denotes the Dirac δ-distribution and $P(\mathbf{T}|\mathbf{D})$ is the N_t-dimensional posterior distribution of $\Delta|y|$ estimated by the FBU method from data $\mathbf{D}$. In practice, the posterior defined in Eq. 7.14 is obtained by calculating the A_{C} value for every

[6] The *Asimov dataset* is the total predicted distribution, i.e. the sum of predicted signal and background distributions of the observable of interest.

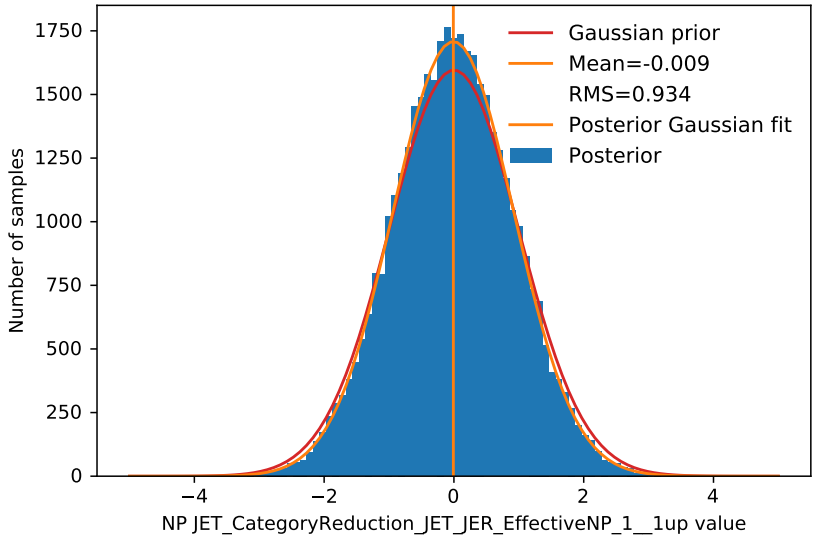


Fig. 7.5: Example of a marginalised posterior distribution of a NP corresponding to JER uncertainty. The blue histograms shows the distribution of the samples from the FBU sampling. The Gaussian prior (red line) as well as Gaussian fit to the posterior (orange line) is shown, highlighting a small ($\approx 7\,\%$) constraint of the NP obtained in the inclusive Asimov A_{C} unfolding.

sample from the N_t-dimensional $\Delta|y|$ posterior distribution. The result of this per-sample calculation is a one-dimensional posterior distribution of A_{C}, where the mean and the variance of the distribution is interpreted as the central value and the uncertainty on A_{C}. This approach correctly accounts for correlations between the bins of the unfolded $\Delta|y|$ distribution in the uncertainty on A_{C}.

Combination of multiple regions in the unfolding

The likelihood formalism allows for a natural combination of multiple regions with different signal and background composition by taking a product of per-region likelihoods, with the exception of the nuisance parameter priors which are shared among the regions. Combining multiple regions allows to constrain certain systematic uncertainties in regions sensitive to the source of the uncertainty. It also is useful if regions differ in their sensitivity to the observable of interest, potentially reducing sources of dilution. Multiple regions may also enhance the information on the correlations of the systematic uncertainties, which in general impact multiple regions simultaneously. The marginalised likelihood with multiple channels is defined as:

$$\mathcal{L}(\mathbf{D}_1 \dots \mathbf{D}_\nu | \mathbf{S}_1(\mathcal{M}_1, \mathbf{T}) \dots \mathbf{S}_\nu(\mathcal{M}_\nu, \mathbf{T}), \mathbf{B}_1 \dots \mathbf{B}_\nu) = \int d\boldsymbol{\theta} \prod_{i=1}^{\nu} \mathcal{L}(\mathbf{D}_i | \mathbf{S}_i(\mathcal{M}_i, \mathbf{T}; \boldsymbol{\theta}), \mathbf{B}_i(\boldsymbol{\theta}))\ \pi(\boldsymbol{\theta}). \tag{7.15}$$

In the unfolding, a single true distribution $\mathbf{T}$ is unfolded, therefore for each region i, a separate response matrix $\mathcal{M}_i$ describes the migration and acceptance of events from the region into the common $\mathbf{T}$ distribution.

In this analysis, a combination of four regions is done, split in $1b$-exclusive and $2b$-inclusive b-jet multiplicity and in resolved and boosted topologies.

Sampling

To obtain the marginalised posterior distribution of the true spectrum $\mathbf{T}$, it is necessary to sample the $(N_t + N_{np})$-dimensional parameter space, where N_{np} is the number of NPs, and for each point evaluate the $\mathcal{L}(\mathbf{D}|\mathbf{T}, \mathcal{M}, \mathbf{B})\pi(\mathbf{T})$ term. This is effectively equivalent to performing a numerical integration of the likelihood over the NPs, and obtaining N_t dimensional posterior distribution of $\mathbf{T}$, which can subsequently used to obtain the A_C posterior. In order to sample this high-dimensional space efficiently, Markov Chain Monte Carlo (MCMC) techniques [202] are used. The FBU implementation used in this analysis is based on the PyMC3 [203] probabilistic programming library, and the MCMC sampling algorithm employed is the *No-U-turn sampler* (NUTS) [204], based on Hamiltonian Monte Carlo methods, which exploit gradient information in the sampling.

The NUTS sampler works in two stages. In the first stage, *tuning sampling* is performed, during which the parameter space is scanned to adapt the sampling step size, to estimate initial parameter values, and to estimate the covariance matrix of the parameters. The tuning steps are discarded afterwards. At the second stage, the sampler is set up with the previously determined parameters of the sampler to obtain samples used for posteriors. In this analysis, the NUTS sampler is used to sample four chains in parallel, which brings two major advantages. Firstly, this allows to leverage multi-core processors to parallelise and thus speed up the unfolding. Secondly, before the tuning, each of the chains is initialised with random independent parameter values, and after the full sampling the Gelman-Rubin test statistic [205] is calculated to check that good convergence across the chains is obtained, ensuring larger robustness of the sampling in contrast to using a single chain.

The number of tuning and sampling steps are free parameters, that need to be determined. Typically at least 500 tuning steps are needed, but the actual number depends on the problem being solved. The PyMC3 library provides robust diagnostics if the amount of tuning steps is insufficient. In this analysis, 2500 tuning steps per each chain were found to be more than sufficient. The number of sampling steps determines the accuracy of the posterior estimates. The nuisance parameters and the $\Delta|y|$ bin posterior means and A_C posterior means were found to be stable at the order of 10000 sampling steps. In the analysis, 40000 steps are sampled in total (10000 per chain). A systematic uncertainty on the statistical precision due to finite number of sampling steps is assessed in Sec. 7.6.3, showing that the chosen number of tuning and sampling steps leads to negligible statistical fluctuations of the A_C posterior mean.

7.5 Optimisation of binning

The choice of $\Delta|y|$ binning in this analysis can be divided into two steps. Firstly, for differential measurements of A_C, binning of the double-differential X vs $\Delta|y|$ distribution must be chosen. This means appropriate choice of the binning for $m_{t\bar{t}}$ and $\beta_{z,t\bar{t}}$ observables, here referred to as *differential binning*. The second step involves the choice of binning for the $\Delta|y|$ observable.

The choice of binning of the differential distributions is a compromise between fine binning and sufficient amount of statistics. Finer binning allows to probe the dependence of the A_C in more detail, allowing for more precise comparisons with SM predictions and larger constraining power for BSM interpretations. In addition, both high-$m_{t\bar{t}}$ and high-$\beta_{z,t\bar{t}}$ regions of phase space are interesting due to the enhancement of the charge-asymmetric $q\bar{q} \to t\bar{t}$ contribution. The following differential bins are

used:

- $m_{t\bar{t}}$ bins: $< 500\,\text{GeV}$, $[500, 750]$, $[750, 1000]$, $[1000, 1500]\,\text{GeV}$, $> 1500\,\text{GeV}$ (in boosted regions, the $m_{t\bar{t}} < 500\,\text{GeV}$ bin is removed due to negligible acceptance given the selection criteria),
- $\beta_{z,t\bar{t}}$ bins: $[0, 0.3]$, $[0.3, 0.6]$, $[0.6, 0.8]$, $[0.8, 1]$.

The $\Delta|y|$ distribution requires at least two bins, $\Delta|y| < 0$ and $\Delta|y| > 0$ to calculate the A_C, however previous measurements [69] show that two bins are insufficient to accurately describe the migrations between true and detector-level distributions, leading to biased response of the unfolding. Therefore, four $\Delta|y|$ bins are used instead with bin edges $[-5.0, x, 0, x, 5.0]$, where x denotes the inner bin edge, that must be chosen for each differential bin separately. The contribution of $t\bar{t}$ pairs produced at $|\Delta|y||| > 5.0$ is negligible.

The figure of merit in the choice of the $\Delta|y|$ binning is the linearity of the response of the unfolding. This is tested by injecting asymmetry into the SM $t\bar{t}$ MC prediction based on light axigluon[(7)] BSM models generated using Protos generator. Both the true-level and detector-level signal $\Delta|y|$ distributions are reweighted event-by-event according to $\Delta|y|$ ratio of axigluon to SM prediction. In an ideal case, the A_C unfolded from the Asimov dataset built using the reweighted detector level signal distribution should match the A_C of the reweighted true signal distribution. Eight points with different choice of axigluon couplings are used for this test, varying in the asymmetry from approximately ± 1 % up to ± 5 %. For each reweighting point, the unfolding test is repeated using 300 pseudo-experiments, where the axigluon-reweighted Asimov dataset prediction is smeared according to the expected uncertainty, assuming Poisson statistics[(8)]. The dependence of mean unfolded A_C from the pseudo-experiments as a function of the true A_C is fitted using linear function. For an unbiased response, we expect slope of the fit equal to one and offset equal to zero.

Due to the computational complexity of the pseudo-experiments, the binning optimisation is performed using stat-only unfolding[(9)]. Additionally, optimising the inner $\Delta|y|$ bin edge for multiple differential bins leads to a sizeable dimensionality of the optimisation problem. Due to this, a random scan of binning is performed initially, and final binning is obtained iteratively by tuning the bin edges to minimise the offset and obtain slope close to unity, until binning with linearity is found. While the procedure is to some extent arbitrary and not guaranteed to find the global best binning in terms of slope and offset, we assign in Sec. 7.6.3 an uncertainty due to the residual non-closure in linearity and demonstrate in Sec. 7.7 that this uncertainty is negligible. For the optimised $\Delta|y|$ binning, the linearity test is repeated with the inclusion of full set of systematic uncertainties as NPs in the unfolding. This is done to ensure that the injected asymmetry is not absorbed by pulls of nuisance parameters, thus deteriorating the linearity. The binning choice used for both inclusive and differential observables as well as the slope and offset values from the full-systematics linearity tests are shown in Table 7.3. In addition, the migration matrices for the chosen binning are shown in App. C.1.

(7) The mass of the axigluons in this model are set to 250 GeV.

(8) The expected uncertainty of an Asimov dataset for i^{th} bin of the distribution is $\sqrt{\mu_i}$, where μ_i is the respective bin yield.

(9) No systematic uncertainties are included as NPs in the unfolding.

Table 7.3: The $\Delta|y|$ binning for inclusive and differential A_C unfolding. Both the slope and the offset from the linearity test are shown, along with their uncertainties from the fit.

Differential bin	$\Delta\|y\|$ binning	Linearity	
		slope	offset
Inclusive	[-5,**-0.5**, 0, **0.5**, 5]	0.998 ± 0.001	0.0000 ± 0.0000
$m_{t\bar{t}} < 500\,\text{GeV}$	[-5,**-0.4**, 0, **0.4**, 5]	1.001 ± 0.003	0.0001 ± 0.0001
$m_{t\bar{t}} \in [500, 750]\,\text{GeV}$	[-5,**-0.6**, 0, **0.6**, 5]	0.986 ± 0.001	0.0001 ± 0.0000
$m_{t\bar{t}} \in [750, 1000]\,\text{GeV}$	[-5,**-1.0**, 0, **1.0**, 5]	0.997 ± 0.003	0.0002 ± 0.0001
$m_{t\bar{t}} \in [1000, 1500]\,\text{GeV}$	[-5,**-0.9**, 0, **0.9**, 5]	0.998 ± 0.004	-0.0004 ± 0.0002
$m_{t\bar{t}} > 1500\,\text{GeV}$	[-5,**-0.9**, 0, **0.9**, 5]	1.027 ± 0.016	-0.0009 ± 0.0005
$\beta_{z,t\bar{t}} \in [0, 0.3]$	[-5,**-0.3**, 0, **0.3**, 5]	1.015 ± 0.006	-0.0001 ± 0.0001
$\beta_{z,t\bar{t}} \in [0.3, 0.6]$	[-5,**-0.3**, 0, **0.3**, 5]	1.007 ± 0.002	0.0002 ± 0.0001
$\beta_{z,t\bar{t}} \in [0.6, 0.8]$	[-5,**-0.5**, 0, **0.5**, 5]	1.012 ± 0.002	-0.0005 ± 0.0001
$\beta_{z,t\bar{t}} \in [0.8, 1.0]$	[-5,**-0.7**, 0, **0.7**, 5]	0.995 ± 0.002	0.0007 ± 0.0001

7.6 Systematic uncertainties

A wide range of systematic uncertainties are considered in the measurement of the charge asymmetry and described in this section. Many of the uncertainties are similar in approach to those described in Sec. 6.4. However some differences are expected due to the expansion of this analysis to the full Run-II dataset usage and due to updated, more precise measurements of the various object-related properties, from which the systematic uncertainties arise. The relevant differences are highlighted in the respective uncertainty descriptions.

The individual systematic uncertainties are included as nuisance parameters in the unfolding, with a Gaussian prior, unless otherwise specified. A technical limitation of the FBU implementation used in this analysis is that it does not allow for treatment of asymmetrical uncertainties, therefore the individual up and down variations of uncertainties are symmetrised according to the following formula:

$$\sigma_{\text{symm}}(\mathbf{X}) = \frac{1}{2}\left|\sigma_{\text{up}}(\mathbf{X}) - \sigma_{\text{down}}(\mathbf{X})\right|, \tag{7.16}$$

where $\sigma_{\text{up}}(\mathbf{X})$ and $\sigma_{\text{down}}(\mathbf{X})$ are the up and down variations of a binned observable $\mathbf{X}$ and $\sigma_{\text{symm}}(\mathbf{X})$ is the symmetrised variation. The individual two-sided variations are checked for strongly asymmetric behaviour, which would lead to a potentially significant underestimation in the symmetrised variation. Special cases of uncertainties, where asymmetric behaviour is expected, are treated individually, and explicitly mentioned. One-sided variations are symmetrised to create two-sided variations, where both up and down variation have the same magnitude as the original one-sided variation.

7.6.1 Experimental uncertainties

The pile-up, lepton, large-R jet, $E_{\text{T}}^{\text{miss}}$ and flavour tagging uncertainties use the same methodology as those in Sec. 6.4.1, with dedicated validations performed for the different pile-up and other collision

conditions in 2017 and 2018 year of data-taking. Below we highlight the differences in the uncertainties, which mostly included updated small-R jet JES and JER uncertainties, and reduction in the luminosity uncertainty.

Small-R jet uncertainties

In this analysis, the JES uncertainty is propagated using a more detailed parametrisation using 29 NPs in total, which are decorrelated using eigenvector decomposition. The NPs comprise of multiple characteristic sources of uncertainties. Firstly, 16 NPs describe the uncertainties associated with the individual in-situ calibration methods described in Sec. 4.4.3. Five NPs are related to the uncertainties on the η inter-calibration. Four NPs arise from the uncertainties on the correction of jet energy to pile-up effects. Additionally, in the calibration process, assumptions are made about the flavour composition of jets (fraction of quark- and gluon-induced jets), and an uncertainty on the composition is added, where by default it is assumed that 50 % of jets are quark-initiated, with a 50 % uncertainty on this fraction. Furthermore, the detector and reconstruction response differs for quark- and gluon-initiated jets due to difference in properties such as shower shape in the calorimeter system and the charge particle multiplicity in the jet. As such, one NP is added to characterise the uncertainty on the flavour dependence. Finally, a single NP is added to account for the uncertainty on the JES calibration due to punch-through[(10)] and a single NP accounts for high-p_{T} extrapolation of the JES calibration [110].

The JER uncertainty is derived in dijet events in the same manner, as in Sec. 6.4.1, via techniques similar to those in [167], but is propagated via a finer parametrisation using 8 decorrelated NPs in total.

Luminosity uncertainty

The luminosity for full Run-II dataset is determined with an uncertainty of 1.7 % [172], propagated to all MC simulated processes in the analysis.

7.6.2 Signal and background modelling uncertainties

Signal modelling uncertainties

The signal modelling uncertainties are estimated via comparisons of several different generator setups described in Sec. 7.1.1 designed to probe a particular aspect of the $t\bar{t}$ modelling. The uncertainty on the amount of ISR is estimated by comparing the nominal sample with two samples with alternative μ_R, μ_F and h_{damp} settings in the ME, as specified in 7.1.1, yielding a sample with enhanced and suppressed ISR. The FSR variations are obtained by varying the μ_R scale of the Var2c parameter of the A14 tune by a factor of 0.5 and 2.0 with respect to the nominal scale [174]. To judge, which of the up and down variations for both ISR and FSR has larger impact on the charge asymmetry, the unfolding is repeated using either of the variations, taking the one yielding a larger total uncertainty on the unfolded asymmetry.

The uncertainty on the modelling of PS and hadronisation is estimated by replacing the nominal Powheg+Pythia8 generator with Powheg+Herwig7. In a similar manner, the uncertainty

[(10)] Punch-through denotes the phenomenon, when a jet passes through the whole calorimeter system without being fully absorbed.

on the matching of PS and ME is estimated by comparing the nominal POWHEG+PYTHIA8 with MADGRAPH5_AMC@NLO +PYTHIA8.

An uncertainty on the top-quark mass is considered by comparing the nominal sample prediction which uses $m_{\text{top}} = 172.5\,\text{GeV}$, with predictions with a mass of 172 GeV[(11)].

Finally, an uncertainty on the choice of PDF set in the signal sample is assessed using the PDF4LHC15 prescription [176] similarly as in Sec. 6.4.2. Each of the 30 variations is assigned to a separate NP in the unfolding.

Since the charge asymmetry is a ratio of production rates, signal normalisation variations do not impact the asymmetry nor its uncertainty. Due to this signal normalisation uncertainty and generally normalisation effects of any uncertainties which only impact the signal distribution have no effect on the asymmetry. Therefore, signal modelling uncertainties are normalised to nominal prediction, thus only impacting the shape of distributions. This is done also to avoid potential constraints of the corresponding NPs due to differences in the normalisation of the various generator predictions.

The high purity of the selection combined with small data statistical uncertainty with the full Run-II dataset and the large $t\bar{t}$ modelling uncertainties estimated from comparisons of two generator setups result in constraints of these uncertainties. In order to reduce potential over-constraints, all of the uncertainties with the exception of PDF variations are decorrelated between resolved and boosted regions, resulting in two NPs per uncertainty, each affecting either resolved or boosted regions independently. This approach reduces the constraints, in particular the transfer of constraint from one region into another, and is thus a more conservative approach. This avoids the questionable assumption that the modelling across both resolved and boosted kinematics is sufficiently accurate.

Normalisation uncertainties of MC-predicted processes

Each of the MC simulated backgrounds has an associated normalisation uncertainty, that is added as a nuisance parameter with a Gaussian prior truncated at a value of $-1/\sigma_b$, where σ_b is the respective background normalisation uncertainty. For example, for a 50 % uncertainty, the Gaussian prior is truncated at value -2. This is done to disallow negative background yields.

The individual background normalisation uncertainties are listed in Table 7.4. For W+jets background, a conservative 50 % normalisation uncertainty is assigned and is decorrelated across $1b$-exclusive and $2b$-inclusive regions. This is done as a conservative approach to reduce the transfer of normalisation constraint from $1b$-exclusive into $2b$-inclusive region. The magnitude of the uncertainty is motivated by acceptance effect of μ_R and μ_F scale variations leading to 50 % (30 %) variations in $2b$-inclusive ($1b$-exclusive) regions. Finally, the smallest backgrounds in the analysis, comprising Z+jets, diboson, $t\bar{t}V$ and $t\bar{t}H$ production, are combined into a single background, with a conservative 50 % normalisation uncertainty.

Modelling uncertainties of the W+jets background

The W+jets is one of the most-important backgrounds, and is charge asymmetric. In $1b$-exclusive region, is the dominant background contribution. Therefore, uncertainties on its modelling are considered, by considering several scale variations. The CKKW and QSF two-sided scale variations are defined in

(11) Other alternative mass variations were not generated at the time of this analysis.

Table 7.4: The uncertainties on the normalisation of the individual MC-simulated backgrounds considered in the charge asymmetry measurement.

Process	Normalisation uncertainty (%)
Single-top	5.3
W+jets	50.0
Z+jets, diboson, $t\bar{t}V$, $t\bar{t}H$	50.0

the same manner as in Sec. 6.4.2, each of the scale variations added as a single NP. The μ_R and μ_F variations are also added as three NPs either shifting the μ_R scale, or the μ_F scale, or both μ_R and μ_F in a correlated manner. The magnitude of the scale variations are a factor of $\times 0.5$ ("up" variation) and $\times 2.0$ ("down" variation).

Modelling uncertainties of the single-top-quark background

The single-top background, in particular the tW production, is the dominant background in the $2b$-inclusive region. A number of uncertainties on the modelling are considered using comparisons of predictions between various generator setups, described in Sec. 7.1.2. The uncertainty prescriptions are the same as those for $t\bar{t}$ modelling uncertainties, with the exception of the top-quark mass uncertainty which is not considered for single-top-quark production. In addition, an additional uncertainty on the approach to removal of overlap between $t\bar{t}$ and tW production amplitudes in the ME generation is considered. This uncertainty is estimated by comparing predictions between the diagram subtraction (DS) and diagram removal (DR) scheme, yielding a one-sided systematic variation.

Uncertainties in the fake and non-prompt leptons estimate

Two uncertainties are imposed on the data-driven estimate the fake and non-prompt leptons. Firstly, a conservative normalisation uncertainty of 50 % is assigned to the estimate. Additionally, an alternative parametrisation to the default choice, described in Sec. 7.2 is used to obtain a systematically varied distribution of the background estimate. The alternative parametrisation uncertainty yield is normalised to the nominal parametrisation choice, making this variation shape-only. Additionally, both the shape uncertainty as well as normalisation uncertainty are decorrelated across all four regions, resulting in four NPs for normalisation uncertainty and four NPs for shape uncertainty, respectively. This step is motivated by the fact that the underlying physics processes contributing to the background estimate are different in the individual regions.

7.6.3 Other systematic uncertainties

Treatment of $t\bar{t}$ normalisation mismodelling in boosted region

The top-quark p_{T} mismodelling observed in $t\bar{t}$ MC simulation has a large impact in the boosted regions. In the resolved region, the predicted yields agree well with the data as shown in Table 7.1. However in the boosted region, the MC predicts approximately 20 % more events than data. Even though the true-level signal distribution in the unfolding is allowed to float freely, the normalisation discrepancy

at detector-level cannot be compensated in both resolved and boosted regions simultaneously. This is because the combined resolved+boosted response matrix binds the migrations between the resolved and boosted regions and the acceptances according to the $t\bar{t}$ MC simulation. Consequently, it is expected that unfolding of data would lead to large pulls of NPs attempting to compensate the boosted regions' normalisation discrepancy in data.

We introduce a single additional NP with a flat prior, that affects only the signal normalisation in boosted regions, expecting the pull of this NP to absorb the observed normalisation discrepancy. It is expected that this NP should be pulled to $\theta_{\text{boosted norm.}} \approx -1$, leading a normalisation correction factor $k_{\text{boosted}} = 0.8$ in the unfolding of data. For the Asimov dataset unfolding the pull should be $\theta_{\text{boosted norm.}} = 0$, corresponding to $k_{\text{boosted}} = 1$. The Table 7.5 shows the observed values of $\hat{\theta}_{\text{boosted norm.}}$ and k_{boosted}, as obtained from the unfolding of both Asimov and data distribution, for $\Delta|y|$ as well as the double-differential $\Delta|y|$ vs $m_{t\bar{t}}$ and $\Delta|y|$ vs $\beta_{z,t\bar{t}}$ observables. The observed normalisation corrections are compatible with the expected values and also compatible between the individual unfolded measurements.

Table 7.5: The pull of the boosted normalisation NP $\hat{\theta}_{\text{boosted norm.}}$ and the corresponding normalisation correction k_{boosted} of $t\bar{t}$ prediction in the boosted region, obtained from Asimov dataset unfolding and unfolding of data.

	Asimov		Data	
Unfolding	$\hat{\theta}_{\text{boosted norm.}}$	k_{boosted}	$\hat{\theta}_{\text{boosted norm.}}$	k_{boosted}
Inclusive $\Delta\|y\|$	0.00 ± 0.60	1.00 ± 0.12	-1.10 ± 0.60	0.78 ± 0.12
$\Delta\|y\|$ vs $m_{t\bar{t}}$	0.00 ± 0.28	1.00 ± 0.06	-1.00 ± 0.28	0.80 ± 0.06
$\Delta\|y\|$ vs $\beta_{z,t\bar{t}}$	0.00 ± 0.43	1.00 ± 0.09	-1.12 ± 0.43	0.78 ± 0.09

Uncertainty due to limited signal MC sample statistics

Due to the limited number of generated events in the signal MC sample, the response matrix is estimated with a limited statistical precision. The resulting uncertainty is estimated by performing pseudo-experiments, where the bins of the response matrix are smeared and the Asimov dataset is unfolded by the smeared response matrix. The bins of the response matrix are smeared by a Gaussian with mean equal to the bin content and the standard deviation equal to the MC statistical uncertainty of the bin. A distribution of the unfolded A_{C} of pseudo-experiments is obtained, where the standard deviation of this distribution is taken as the actual systematic uncertainty on A_{C}.

Uncertainty due to unfolding non-closure

The linearity tests in Sec. 7.5 show small deviations from an ideal slope of one and offset of zero. A bias uncertainty is defined in the following, which quantifies by how much does the true value of A_{C} differ from the unfolded $A_{\text{C}}^{\text{meas.}}$ value, given the observed slope a and offset b in the linearity test:

$$\sigma_{\text{bias}}(A_{\text{C}}^{\text{meas.}}) = A_{\text{C}}^{\text{meas.}} - \frac{A_{\text{C}}^{\text{meas.}} - b}{a} \tag{7.17}$$

Uncertainty due to limited number of sampling steps in FBU

Due to the random nature of the MCMC sampling employed in FBU, repeating the unfolding each time yields a slightly different result. To quantify, how much the results can vary due to the limited number of MCMC steps, the Asimov unfolding is repeated 500 times. A distribution of the unfolded A_C values is obtained, where the standard deviation of this distribution expresses the uncertainty on the unfolded A_C due to the limited number of sampling steps. In a similar manner, the distribution of the unfolded A_C uncertainty from the 500 unfolding repetitions gives an estimate of the variance of the unfolding uncertainty. The results of this test are summarised in Table 7.6. In general, the variations in the unfolded A_C central value reach up to order of 1 %, with the lowest $\beta_{z,t\bar{t}}$ bin showing a 15 % variation. However the predicted asymmetry in this bin is very close to zero, therefore the relative variation is not a meaningful metric, and in contrast to the unfolded A_C uncertainty in this bin, the variation is negligible. The test also shows that the MCMC fluctuations impact the unfolded uncertainty by less than 0.5 % for any of the inclusive or differential A_C measurements. The conclusion therefore is that the variations are negligible and the amount of MCMC steps used is sufficient. No uncertainty due to the variations is assigned to the final result.

Table 7.6: Stability test of the FBU with respect to the amount of sampling steps performed. The test is performed by 500 repetitions of the Asimov unfolding with all NPs. The unfolded A_C columns show the mean value μ_{PE} and the standard deviation σ_{PE} of the unfolded A_C obtained from the ensemble, and the resulting relative standard deviation σ_{PE}/μ_{PE}. In a similar manner, the unfolded $\sigma(A_C)$ columns show the mean value and the standard deviation of the unfolded A_C uncertainty obtained from the ensemble. The rows show the individual inclusive and differential unfolded A_C bin results.

	Unfolded A_C				Unfolded $\sigma(A_C)$			
	$\mu_{PE} \pm \sigma_{PE}$ [$\times 10^{-2}$]			$\frac{\sigma_{PE}}{\mu_{PE}}$ [%]	$\mu_{PE} \pm \sigma_{PE}$ [$\times 10^{-2}$]			$\frac{\sigma_{PE}}{\mu_{PE}}$ [%]
Inclusive	0.363	±	0.001	0.22	0.161	±	0.001	0.44
$m_{t\bar{t}} < 500$ GeV	0.290	±	0.002	0.76	0.396	±	0.002	0.38
$m_{t\bar{t}} \in [500, 750]$ GeV	0.420	±	0.001	0.28	0.253	±	0.001	0.42
$m_{t\bar{t}} \in [750, 1000]$ GeV	0.476	±	0.003	0.69	0.605	±	0.002	0.38
$m_{t\bar{t}} \in [1000, 1500]$ GeV	0.639	±	0.004	0.68	0.793	±	0.003	0.38
$m_{t\bar{t}} > 1500$ GeV	1.143	±	0.015	1.27	2.800	±	0.012	0.45
$\beta_{z,t\bar{t}} \in [0.0, 0.3]$	0.020	±	0.003	15.31	0.585	±	0.002	0.39
$\beta_{z,t\bar{t}} \in [0.3, 0.6]$	0.153	±	0.002	1.13	0.383	±	0.002	0.41
$\beta_{z,t\bar{t}} \in [0.6, 0.8]$	0.224	±	0.003	1.17	0.421	±	0.002	0.42
$\beta_{z,t\bar{t}} \in [0.8, 1.0]$	0.902	±	0.002	0.22	0.403	±	0.002	0.38

7.6.4 Bootstrapping of systematic uncertainties

Due to the limited amount of generated MC events, the systematic uncertainties are also impacted by statistical fluctuations. In this analysis the *bootstrap method* [206] is used for the $\Delta|y|$ distribution to correctly estimate the statistical uncertainty of a systematic variation, which in general is correlated with the nominal prediction. This method is based on generating multiple event replicas with weights

generated from a Poisson distribution with $\mu = 1$. The random seed for the replicas is set for each event based on a unique event number to ensure correlated smearing of the nominal distribution and systematically-shifted distribution. For each bin of the distribution of interest, for each Poisson replica the difference between the smeared systematic shift and the smeared nominal value is taken, obtaining a distribution of systematic-shift replicas. The mean of this distribution is taken as the systematic uncertainty and the standard deviation as its statistical component. In total, 500 Poisson replicas are generated.

Additionally, statistically insignificant uncertainties are removed based on the significance[(12)] of the mean systematic shift of a distribution using the following procedure:

- If the systematic shift in at least one of the bins is more than 2σ significant, the uncertainty is preserved.
- If none of the bins is more than 2σ significant, a check is performed, whether the normalisation shift of the systematic on the distribution is at least 2σ significant. If yes, the normalisation effect of the uncertainty is preserved, while the shape is removed.
- Finally, if the normalisation shift of the uncertainty is not more than 2σ significant either, the systematic variation is removed, i.e. is set to nominal distribution.

This procedure is applied independently on signal and individual background contributions, and independently for each region. Additionally, for differential $\Delta|y|$ vs X observables, the procedure is applied independently for individual bins of X. For two-sided variations, the up and down variations are treated separately, and symmetrised afterwards.

In summary, the bootstrapping method helps to reduce statistical noise in the uncertainties, potentially reducing non-physical constraints of NPs, and removing insignificant sources of systematic uncertainties. Significance thresholds lower than 2σ were also tested with the bootstrap method, showing negligible variations in the total uncertainty, ensuring that the procedure is not too aggressive in terms of uncertainties removal.

Finally, the bootstrap method is not used for systematic uncertainties derived using weights, given that these variations are 100 % correlated and typically have a small spread of weights with respect to nominal weight values. A single exception to this rule is the estimate of the uncertainty on the amount of FSR in $t\bar{t}$ and single-top-quark simulations, where a large spread of weights is observed. Therefore, the bootstrap method is used for this uncertainty. In general, it is observed that the FSR up variation which has μ_R decreased by factor 0.5, leading to larger α_s, yields much larger statistical uncertainties on the estimate. Therefore, the FSR down variation is used instead and symmetrised in the measurement.

7.6.5 Ranking of nuisance parameters in the FBU

Due to the nature of the profile-likelihood and Bayesian marginalisation methods, the impact of individual systematic uncertainties and the statistical uncertainty of the data are not evaluated individually. The following approach is used in this analysis to qualitatively estimate the relative importance of the individual uncertainties, referred to as *ranking* of systematic uncertainties. For each NP ranked, the corresponding systematic up or down shift is superimposed onto the Asimov

[(12)] Here, the significance is equal to the ratio of μ/σ, where μ is the mean value and σ the standard deviation of the distribution of systematically-shifted Poisson replicas.

dataset prediction. This systematically-shifted distribution is then unfolded using model with all of the NPs with the exception of the ranked NP. The result is the unfolded A_C corresponding to an up or down shift of the ranked NP, which is compared to Asimov unfolded A_C. This ranking is referred to as the *pre-marginalisation* ranking. In a similar manner, the *post-marginalisation* ranking is performed by considering constraints of the NPs. For each NP ranked, the up or down variation of a constrained systematic uncertainty is superimposed onto the Asimov dataset prediction and the rest of the procedure is identical to the pre-marginalisation ranking. The constrained variation is obtained by scaling the unconstrained original systematic uncertainty by the constraint factor C obtained from Asimov unfolding, where $0 < C \leq 1$. The results of the systematic ranking are discussed in Sec. 7.7.1.

7.7 Results

7.7.1 Impact of statistical and systematic uncertainties on the measurement

In this section the results of the A_C measurement are presented. The unfolded A_C values from data and their uncertainties are summarised in Table 7.7. Additionally, Table 7.8 shows an approximate decomposition of the total unfolded asymmetry into statistical and systematic component. With the full Run-II dataset, the A_C measurement is reaching a point where systematic uncertainties start to dominate the measurement, in particular for the inclusive A_C and differential A_C vs $\beta_{z,t\bar{t}}$ measurements. The differential measurement of A_C vs $m_{t\bar{t}}$ in particular for the two highest-$m_{t\bar{t}}$ differential bins is still statistically dominated.

Fig. 7.6 shows the ranking of top 20 largest systematic uncertainties in the inclusive A_C measurement, using the approach described in Sec. 7.6.5. The ranking of systematic uncertainties for the differential A_C measurements is included in App. C.2. The post-marginalisation correlation matrices of the NPs obtained from the unfolding of data are included in App. C.3.

The systematic uncertainties limiting the precision of the measurement mostly include the signal and background modelling uncertainties and some specific detector-related uncertainties. Among the most dominant systematic uncertainties are the $t\bar{t}$ modelling uncertainties, which were found to be limiting the previous ATLAS measurements as well [69, 75]. In this measurement, their impact is reduced by including them in the Bayesian marginalisation, as the corresponding NPs are constrained and the total uncertainty can be reduced by accounting for the correlations of the NPs in the likelihood. The most important background modelling uncertainties include the W+jets background normalisation and scale variations, due to the fact that the W+jets production at the LHC is charge-asymmetric, and the uncertainty on the scheme of $t\bar{t}$ and Wt overlap removal. Finally, some of the experimental uncertainties, particularly jet energy scale and E_T^{miss} soft-term uncertainties account for differences in detector response of the respective object reconstruction, as a function of $|\eta|$. Because of its definition, A_C is sensitive to differences in detector response for central and forward rapidity.

Table 7.7: Results of the inclusive and differential A_C measurements in data. The central value with the marginalised stat.+syst. uncertainty is shown in the first two columns. The additional uncertainties included are the uncertainty due to limited number of MC events (MC stat. column) and uncertainty due to the unfolding bias. The total uncertainty is the sum-in-quadrature of the aforementioned uncertainties. The SM predictions are calculated at NNLO QCD + NLO EWK precision [207, 208].

	Data 139 fb^{-1}					SM prediction
	A_C	Marg.	MC stat.	Bias	Total unc.	
Inclusive	0.0060	0.0014	0.0005	0.0001	0.0015	$0.0064^{+0.0005}_{-0.0006}$
$m_{t\bar{t}}$						
< 500 GeV	0.0045	0.0044	0.0013	0.0001	0.0045	$0.0055^{+0.0007}_{-0.0005}$
[500, 750] GeV	0.0051	0.0029	0.0009	<0.0001	0.0031	$0.0072^{+0.0006}_{-0.0006}$
[750, 1000] GeV	0.0100	0.0067	0.0021	0.0001	0.0070	$0.0079^{+0.0003}_{-0.0005}$
[1000, 1500] GeV	0.0169	0.0077	0.0029	0.0004	0.0083	$0.0096^{+0.0009}_{-0.0009}$
> 1500 GeV	0.0121	0.0315	0.0092	0.0005	0.0329	$0.0094^{+0.0015}_{-0.0011}$
$\beta_{z,t\bar{t}}$						
[0, 0.3]	0.0007	0.0051	0.0020	0.0001	0.0055	$0.0011^{+0.0004}_{-0.0004}$
[0.3, 0.6]	0.0085	0.0040	0.0013	0.0003	0.0042	$0.0023^{+0.0006}_{-0.0004}$
[0.6, 0.8]	0.0014	0.0044	0.0015	0.0004	0.0047	$0.0042^{+0.0003}_{-0.0003}$
[0.8, 1.0]	0.0100	0.0049	0.0013	0.0007	0.0051	$0.0146^{+0.0012}_{-0.0014}$

Table 7.8: Estimate of the systematic component in the FBU marginalisation. The unfolding is performed without nuisance parameters (Stat. column) and with full nuisance parameter marginalisation (Stat.+Syst. column). The Syst. column is calculated as $(\sigma^2_{\text{tot.}} - \sigma^2_{\text{stat.}})^{1/2}$ and gives an approximate estimate of the relative impact of statistical and systematic uncertainties.

	A_C	Stat.	Syst.	Stat.+Syst.
Inclusive	0.0060	0.0011	0.0009	0.0014
$m_{t\bar{t}}$				
< 500 GeV	0.0045	0.0028	0.0034	0.0044
[500, 750] GeV	0.0051	0.0020	0.0021	0.0029
[750, 1000] GeV	0.0100	0.0049	0.0046	0.0067
[1000, 1500] GeV	0.0169	0.0072	0.0027	0.0077
> 1500 GeV	0.0121	0.0277	0.0150	0.0315
$\beta_{z,t\bar{t}}$				
[0, 0.3]	0.0007	0.0040	0.0032	0.0051
[0.3, 0.6]	0.0085	0.0031	0.0025	0.0040
[0.6, 0.8]	0.0014	0.0029	0.0033	0.0044
[0.8, 1.0]	0.0100	0.0026	0.0042	0.0049

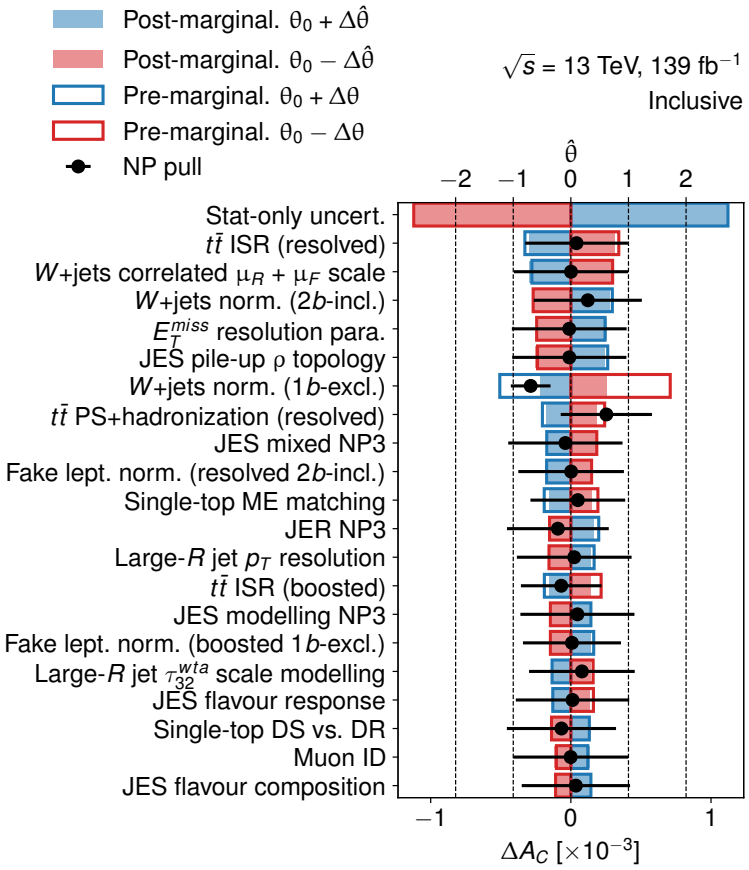


Fig. 7.6: The ranking of top 20 systematic uncertainties and the pulls of the corresponding NPs in the inclusive A_C measurement. The blue (red) colour shows the impact of up (down) variation of a NP on the A_C. The empty rectangles show the pre-marginalisation impact (un-constrained), while the filled rectangles show the post-marginalisation impact of NP (constrained). The ranking of NPs is ordered by the average of post-marginalisation up and down shift impact. The first row shows the statistical uncertainty on A_C.

7.7.2 Comparison of unfolded data with SM theory predictions

In Fig. 7.7, the unfolded posterior distribution is shown for the inclusive A_C measurement. The posterior distributions of A_C are found to be symmetric, having a Gaussian shape, suggesting that the mean and the standard deviations of the posterior are a good metric to evaluate the A_C and it's uncertainty. Fig. 7.8 and 7.9 show the comparison of the unfolded inclusive A_C as well as differential A_C as a function of $m_{t\bar{t}}$ and $\beta_{z,t\bar{t}}$, with the prediction from Powheg+Pythia8 nominal $t\bar{t}$ MC simulation as well as the NNLO QCD + NLO EWK theory prediction [207, 208]. The unfolded results are found to be compatible with both the MC prediction as well as the NNLO QCD + NLO EWK theory prediction. The additional corrections included in the theory prediction are found to enhance the asymmetry, in contrast to the nominal MC simulation with NLO QCD accuracy. The unfolded inclusive A_C result is also the first evidence of a non-zero charge asymmetry at a pp collider, with the observed significance of 4σ.

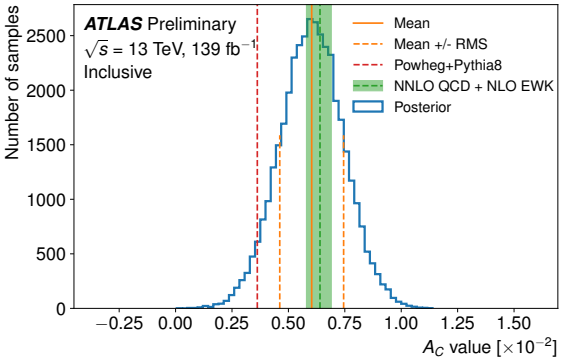


Fig. 7.7: Posterior distribution of unfolded A_C in data. The blue line shows the posterior histogram, where the orange solid line and dashed lines show the mean and the standard deviation (RMS) of the posterior. The left-most red dashed line shows the Powheg+Pythia8 prediction and the green band the NNLO QCD + NLO EWK theory prediction [207, 208].

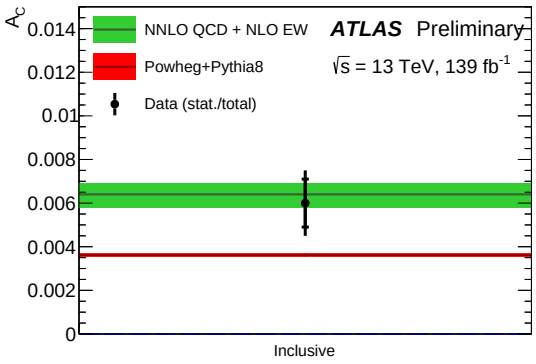
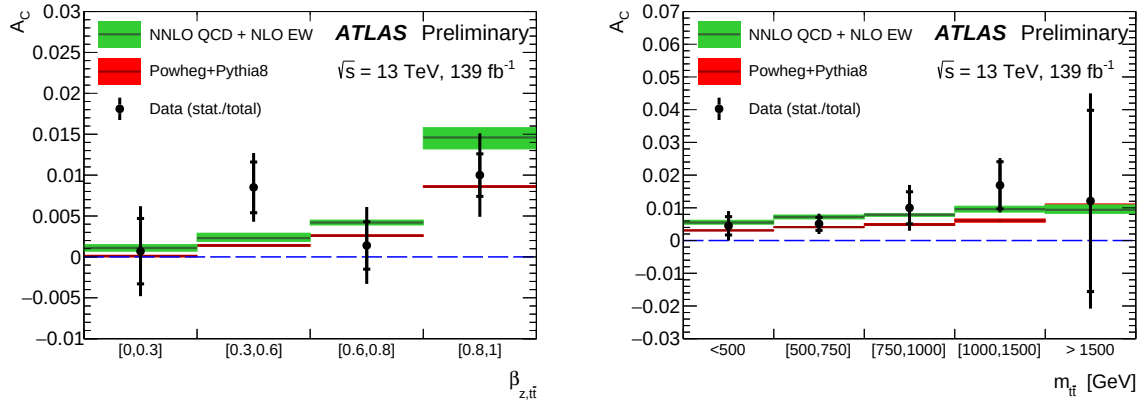


Fig. 7.8: Comparison of the unfolded A_C in data with the Asimov prediction obtained from Powheg+Pythia8 simulation (dark red line) and with the NNLO QCD + NLO EWK prediction [207, 208] (green band).



(a) (b)

Fig. 7.9: Comparison of the unfolded A_C in data with the Asimov prediction obtained from Powheg+Pythia8 simulation (dark red line) and with the NNLO QCD + NLO EWK prediction [207, 208] (green band), as a function of $\beta_{z,t\bar{t}}$ (a) and $m_{t\bar{t}}$ (b).

7.7.3 EFT interpretation of the results

The measured inclusive A_C and the A_C as a function of $m_{t\bar{t}}$ are used for an EFT interpretation, similarly to the interpretation discussed in Sec. 2.7 and using the same parametrisation of the interpretation as for the 8 TeV LHC A_C EFT interpretation in Ref. [84]. The 68 % confidence-level bounds on the C^-/Λ^2 Wilson coefficient are shown in Fig. 7.10, for the inclusive A_C measurement as well as individual $m_{t\bar{t}}$ bins. A comparison with the LHC 8 TeV [74] and Tevatron A_{FB} combination [53] is made. The measurement presented here shows substantial improvements in the sensitivity to the respective EFT operator in contrast to previous measurements, benefiting from the much smaller statistical and systematic uncertainties, despite the larger gluon-fusion dilution of the asymmetry. The high-$m_{t\bar{t}}$ region of phase space is particularly sensitive to the C^-. The A_C in the $m_{t\bar{t}} \in [1000, 1500]$ GeV bin

shows the largest constraint on the C^- coefficient, even larger than the much more precise inclusive A_C measurement. This result is an interplay of two effects; on one hand, $t\bar{t}$ pairs produced with large $m_{t\bar{t}}$ are more abundantly produced via the $q\bar{q} \to t\bar{t}$ process, which is sensitive to C^-. On the other hand, the precision of the measurement decreases with high $m_{t\bar{t}}$. Further constraints of C^- could be achieved by combining the results of the individual $m_{t\bar{t}}$ bins, as well as aiming to improve the precision of the measurement for high-$m_{t\bar{t}}$. The next chapter follows up on the second approach, by investigating a possibility to measure A_C in boosted all-hadronic $t\bar{t}$ production as a function of $m_{t\bar{t}}$. To conclude, the obtained values of the C^- Wilson coefficient are compatible with zero, showing no tension with the SM prediction.

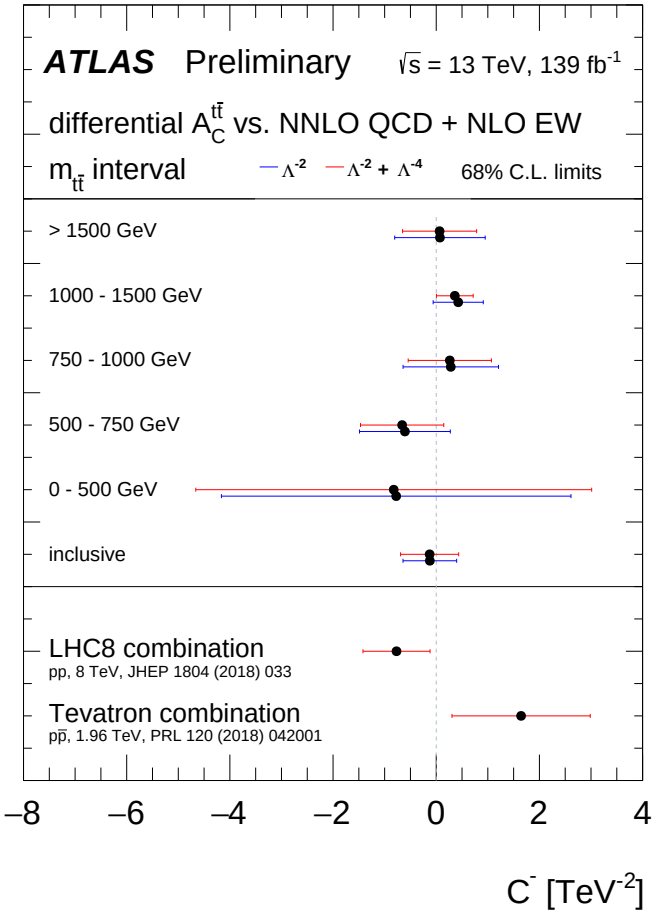


Fig. 7.10: Bounds on the C^-/Λ^2 Wilson coefficient using 68 % confidence level interval (blue lines), obtained from the Run-II ATLAS A_C measurement [2]. An approximate estimate of high-order EFT contributions proportional to Λ^{-4} is calculated according to Ref. [84] by considering $(C^-)^2$ coefficient with the square amplitude of the dimension-six four-fermion operator, obtaining an alternative bound shown by red lines. The constraints are calculated for the inclusive A_C measurement as well as the individual A_C vs $m_{t\bar{t}}$ bins. The previously obtained bounds from the LHC 8 TeV A_C combination [74] as well as the Tevatron A_{FB} combination [53] are shown at the bottom of the plot.

8

Prospect of charge asymmetry measurement in boosted all-hadronic $t\bar{t}$ events

The measurements of the $t\bar{t}$ charge asymmetry, were traditionally performed in the single-lepton and the dilepton channel, since the final-state lepton charge can be used as the proxy for the charge of the reconstructed top quark. In this chapter, a possibility of a measurement of A_C in the all-hadronic channel is investigated. There are two challenges related to this task. Firstly, it is necessary to reconstruct the $t\bar{t}$ system consisting of six jets, a combinatorially very difficult task. We narrow down the problem by focusing only on a boosted topology, where each of the top quarks is reconstructed as a single large-R jet. This avoids the combinatorial complexity and allows to use boosted tagging techniques to suppress the dominant multijet background. The boosted all-hadronic channel is particularly suitable for probing high-$m_{t\bar{t}}$ region of phase space, where the charge-asymmetric $q\bar{q} \to t\bar{t}$ contribution is enhanced. This region of phase space is thus interesting for improving the constraints on the EFT operators sensitive to charge asymmetry.

The second, more important challenge is related to the absence of the lepton from top-quark decay, thus a different method of distinguishing top quark and top anti-quark is needed. Alternative methods of inferring charge of partons initiating jets were investigated in the past, based on calculating charge-sensitive observables using tracks matched to the jet. In this study, we attempt to employ NNs, making use of charge-sensitive observables of large-R jets, to build a discriminant that is sensitive to the charge of the top quark producing the large-R jet. The modelling of the NN discriminant is validated in boosted $t\bar{t}$ events in the single-lepton channel, benefiting from its high signal purity.

Finally, we perform Asimov unfolding of the differential A_C vs $m_{t\bar{t}}$ measurement to obtain an preliminary estimate of the expected A_C sensitivity. The analysis presented here design does not include a complete selection optimisation, and thus the results should be understood as a proof-of-a-concept study of the possibility of this measurement.

The study is organised as follows. The description of the MC-based signal and background estimates is summarised in Sec. 8.1. All of the selection criteria for the NN studies in both MC simulations as well as the validation in data, and the selection criteria for the all-hadronic charge asymmetry sensitivity estimate, are described in Sec. 8.2. In Sec. 8.3, the development of the neural network top-quark charge discriminator is discussed, and in Sec. 8.4 the modelling of the NN is checked in data in $t\bar{t}$ events in the single-lepton channel. The A_C estimate method in the all-hadronic channel is described in Sec. 8.5, with the results presented in Sec. 8.6.

8.1 Data and simulation samples

For the study of the NN modelling in data in the single-lepton channel and for the $t\bar{t}$ charge asymmetry study in the all-hadronic channel, the full ATLAS Run-II dataset with an integrated luminosity of $139\,\text{fb}^{-1}$ is used.

A number of processes modelled by MC simulations are considered for this analysis. In the all-hadronic channel, the background processes include non-all-hadronic $t\bar{t}$, single-top-quark and multijet production as well as small contribution from $t\bar{t}V$ and $t\bar{t}H$ processes. In this study, we consider only a subset of the background processes, neglecting the contributions from $t\bar{t}V$ and $t\bar{t}H$ processes. Based on previous studies [209], these backgrounds contributed by approximately 0.75 %.

The non-all-hadronic $t\bar{t}$ background consists of events where one of the top quarks decays hadronically, and the other top quark decays semi-leptonically, into a b-quark, τ lepton and a τ neutrino. The τ lepton decays hadronically and thus the large-R jet is mis-tagged as originating from a top-quark.

For the study of the modelling of the NN classifier in the boosted $t\bar{t}$ single-lepton channel, the single top-quark and the W+jets backgrounds are considered. Since the study is performed in $2b$-inclusive muon channel, characteristic of its very high purity as shown in Table 7.2, the small backgrounds, which include fake/non-prompt leptons, Z+jets, diboson, $t\bar{t}V$ and $t\bar{t}H$ production, are omitted. Based on the yields in Table 7.2 it is estimated that approximately 2 % of background contributions are neglected.

For all of the MC simulated processes considered, the same generator setups as those described in Sec. 7.1 are used. Finally, the multijet background in the all-hadronic channel is modelled using data-driven technique, described in Sec. 8.5.

8.2 Event selection and reconstruction

8.2.1 All-hadronic channel selection

The event selection for the charge asymmetry sensitivity in all-hadronic $t\bar{t}$ events is based on the boosted all-hadronic $t\bar{t}$ differential cross-section measurement [209]. The analysis selection consists of a common pre-selection and additional selection criteria based on top-tagging and b-tagging of the large-R jets. The additional criteria are used to define the signal region as well as regions for data-driven multijet background estimation and validation.

Only events recorded under stable beam conditions in data and with all detector sub-systems operational are considered. A reconstructed primary vertex is necessary, and a single large-R jet trigger must fire, where the imposed trigger p_{T} threshold is slightly different for different years of data taking, as listed in Table 8.1. Additionally, the following event pre-selection criteria are imposed:

- Events are required to contain zero isolated electrons or muons with $p_{\text{T}} > 25\,\text{GeV}$, with the lepton identification and isolation WPs similar to those in Sec. 6.3.
- Events must contain at least two large-R jets with $p_{\text{T}} > 350\,\text{GeV}$ and $|\eta| < 2.0$. Additionally, the highest-p_{T} large-R jet is required to have $p_{\text{T}} > 500\,\text{GeV}$, to ensure the triggers are fully-efficient.
- The mass of each of the two leading-p_{T} large-R jets must satisfy $|m_{\text{jet}} - m_{\text{top}}| < 50\,\text{GeV}$, where $m_{\text{top}} = 172.5\,\text{GeV}$ is the top-quark mass.

The yields for observed data and predicted MC-simulated contributions after the pre-selection and in the signal region are summarised in Table 8.2.

Table 8.1: The single large-R jet trigger used in the analysis for the individual periods of data-taking.

Period of data-taking	Trigger jet p_{T} threshold
2015	320 GeV
2016	360 GeV
2017+2018	460 GeV

Table 8.2: Event yields in the boosted all-hadronic channel for the pre-selection and the signal region. The statistical uncertainties on the predicted yields corresponding to $L = 139\,\mathrm{fb}^{-1}$ are shown[(1)].

Process	Yield	
	Pre-selection	Signal region
$t\bar{t}$ (all-hadronic)	56780 ± 40	13465 ± 24
$t\bar{t}$ (non-all-hadronic)	10190 ± 50	683 ± 14
Single-top	1653 ± 15	99 ± 4
Multijet	-	2912 ± 41
Total	-	17159 ± 49
Data	2231100	14010

Following the pre-selection, 16 regions are defined based on whether the leading and/or sub-leading large-R jet passes one or both of the top-tagging and b-tagging criteria. The choice of top-tagging and b-tagging algorithms are the main differences in selection compared to both the all-hadronic $t\bar{t}$ differential cross-section measurement as well the charge asymmetry measurement in the single-lepton channel, as outlined in the next paragraphs.

The large-R jet is top-tagged if it passes p_{T}-dependent selection on the discriminant of a DNN top-tagger using high-level substructure observables, introduced in Ch. 5. The WP of this DNN top-tagger has been re-optimised to use a less modelling-dependent signal definition. The large-R jet is labelled as signal, if it passes the *inclusive* definition from Sec. 5.2 and additionally the mass of a matched particle-level jet is greater than 140 GeV and a B-hadron is ghost-matched[(2)] with the particle-level jet. The WP for the DNN top-tagger is designed to give constant-in-p_{T} 80 % signal efficiency with respect to the aforementioned definition.

The b-tagging of a large-R jet is based on b-tagging of track jets as described in Sec. 4.4.5. If a b-tagged track jet is ghost-matched with the large-R jet[(3)], the large-R jet is considered as b-tagged. In the signal region, both of the two highest-p_{T} large-R jets are required to be top-tagged and b-tagged. The additional regions with relaxed b-tagging and/or top-tagging selection are used for the data-driven multijet background estimate.

[(1)] The MC samples were generated with much more events than the number of events in data. In particular, the all-hadronic $t\bar{t}$ sample was generated with approximately 30-times more events.

[(2)] Ghost matching between a track/particle and a jet is performed by repeating the jet clustering algorithm on the constituents of the jet and the track/particle which is added as another constituent with infinitesimal momentum (to not change the jet definition). If the track/particle is clustered into the jet in the re-clustering, it is considered matched. This approach performs better for close-by jets where irregularities in the jet shape are expected. [210]

[(3)] The tracks of the track jets are ghost-matched with the untrimmed large-R jet. After trimming, the track jets matched to the k_t sub-jets passing the trimming criteria are considered matched to the trimmed large-R jet [211].

8.2.2 Selection for top-quark charge neural network training

The NN is trained using a slightly more inclusive sample of simulated all-hadronic $t\bar{t}$ events in contrast to the all-hadronic channel pre-selection in Sec. 8.2.1, specifically:

- Only events where the $t\bar{t}$ is produced via gluon fusion are considered. This removes the charge-asymmetric contribution of $q\bar{q} \to t\bar{t}$ events, to ensure the NN does not bias the A_C estimate.
- Events are required to contain zero isolated electrons or muons with $p_T > 25\,\text{GeV}$, with the lepton identification and isolation WPs same as in Sec. 6.3.
- At least two large-R jets with $p_T > 300\,\text{GeV}$ and $|\eta| < 2.0$ are required.
- The large-R jets must contain a ghost-matched b-tagged track jet.
- The hadronically-decaying top quark and top anti-quark particles at parton level are matched to large-R jets. A large-R jet is considered originating from top (anti-)quark if the distance from jet axis to the parton is $\Delta R < 0.75$. In case of multiple jet matches, the jet with the smallest ΔR is considered.
- Only the large-R jets matched to top (anti-)quark are used for the training.

Approximately 5.6 million $t\bar{t}$ events pass this selection, yielding approximately twice as many large-R jets matched to a top quark or top anti-quark[(4)].

8.2.3 Selection for neural network modelling validation in data

The modelling of the NN input variables and discriminant is studied in boosted $t\bar{t}$ events with single isolated lepton in final state. The same requirements on beam conditions, detector status, primary vertex and trigger requirements are imposed as in Sec. 7.3. However, only the muon channel is considered in this analysis, therefore only one of the single-muon triggers listed in Table 6.2 must fire. The additional selection criteria are similar to those employed in Ch. 6 and 7 for boosted topology:

- At least one *medium* muon with $p_T > 28\,\text{GeV}$, passing the *fixed-cut* isolation selection is required. Events containing additional electrons or muons with $p_T > 25\,\text{GeV}$ are rejected.
- The event is required to satisfy $E_T^{\text{miss}} + M_T^W > 60\,\text{GeV}$.
- At least one small-R calorimeter jet with $p_T > 25\,\text{GeV}$ and $|\eta| < 2.5$ is required, and must be close to the lepton; $\Delta R(\text{jet}, \ell) < 1.5$.
- At least one large-R jet with $p_T > 350\,\text{GeV}$ and $|\eta| < 2.0$ is required. Additionally at least one of the large-R satisfying these conditions must pass the following isolation criteria:
  - The distance between the large-R jet and the small-R jet close to the lepton is required to be $\Delta R(\text{small-}R\text{ jet}, \text{large-}R\text{ jet}) > 1.5$.
  - The transverse plane opening angle between the large-R jet and the isolated lepton must satisfy $\Delta\phi(\text{large-}R\text{ jet}, \ell) > 2.3$.

[(4)] It is possible that events with only one of the top quarks matched to a large-R jet pass the selection. Nevertheless, the fraction of top-quark jets in the sample is 0.01 % higher than top-anti-quark jets, thus both classes of jets are represented in a balanced manner.

- At least one of the large-R jets passing the isolation criteria must be top-tagged using the inclusive 80 % WP DNN top-tagger and the mass of the large-R jet must satisfy $|m_{\text{jet}} - m_{\text{top}}| < 50\,\text{GeV}$. In addition, a b-tagged $R = 0.2$ track-jet must be ghost-matched to the large-R jet. The highest-p_{T} large-R jet passing these selection criteria is considered the hadronic top-quark candidate.
- Finally, the events must contain at least two b-tagged track jets, including the b-tagged track jet matched to the large-R jet top-quark candidate.

The selected events are divided into two regions, based on the sign of the electric charge of the single muon in event. The regions are used to study the modelling of the NN input variables and discriminant separately for jets initiated by top quarks and top anti-quarks, respectively. The event yields for the individual signal and background predictions and the observed data are shown in Table. 8.3.

Finally, for certain observables, such as the $m_{t\bar{t}}$, full kinematic reconstruction of the $t\bar{t}$ system is required. In such a case, the hadronically-decaying top-quark four-vector is assumed to be the four vector of the large-R jet candidate. The leptonic top-quark four vector is reconstructed from the isolated lepton, close-by small-R jet and the neutrino, determined using the same approach as in Sec. 7.3.

Table 8.3: The event yields in the single-lepton boosted $t\bar{t}$ channel, split into two regions based the sign of the electric charge of the (anti-)muon in the event. The statistical uncertainties on the predicted yields are shown.

Process	Region	
	2b-incl. μ^-	2b-incl. μ^+
$t\bar{t}$	14380 ± 40	14390 ± 40
Single top	380 ± 10	423 ± 11
W + jets	182 ± 5	279 ± 5
Total pred.	14950 ± 50	15090 ± 50
Data	12560	13140

8.3 Top-quark charge discrimination using neural networks

8.3.1 Previously designed jet charge observables

Several measurements in the past have used observables based on weighted average of charge of tracks matched to a jet, to infer the charge of the parton initiating the jet. The measurement of top-quark charge at the CDF experiment [63] used the following observable to calculate the charge of a b-tagged jet:

$$Q_{\text{jet}} = \frac{\sum_i q_i \left(\vec{p}_i \cdot \vec{p}_{\text{jet}}\right)^{\kappa}}{\sum_i \left(\vec{p}_i \cdot \vec{p}_{\text{jet}}\right)^{\kappa}}, \tag{8.1}$$

where q_i is the charge of the individual tracks, $\vec{p}_i$ their momentum vector, and $\vec{p}_{\text{jet}}$ is the momentum vector of the jet[(5)]. The κ exponent in the numerator is a free parameter. The sums in numerator and denominator run overall tracks matched to the jet. The observable in Eq. 8.1 was also used in

[(5)]The jet momentum vector is sometimes replaced with a unit vector in the momentum direction. Given the normalisation term in the denominator, both such definitions are equivalent.

the measurement of the $b\bar{b}$ forward-backward asymmetry at the CDF experiment [62], as well as the measurement of the top-quark charge at the ATLAS experiment [212]. The observable was used to distinguish the b-jet from $\bar{b}$-jet, where a jet with negative Q_{jet} was assigned to originate from b quark, and a jet with positive Q_{jet} was assigned to originate from $\bar{b}$ quark. The κ parameter was chosen to be 0.5, optimised to maximise the separation power of Q_{jet}. In general $\kappa > 0$ values are used, because such values guarantee the infra-red safety of the definition [213].

A slightly different definition of the jet charge was used for measuring the charge of quark and gluon jets at the ATLAS experiment [213]:

$$Q_{\text{jet}} = \frac{1}{p_{\text{T,jet}}^{\kappa}} \sum_i q_i \left(p_{\text{T},i}\right)^{\kappa}, \tag{8.2}$$

which weights the charge of individual tracks q_i by their transverse momentum $p_{\text{T},i}$ instead of projection of momentum into jet axis. The charge is normalised by $(p_{\text{T,jet}})^{\kappa}$, where $p_{\text{T,jet}}$ is the transverse momentum of the jet.

An extension of the jet charge approach has been investigated within a study in Ref. [214] based on ATLAS $\sqrt{s} = 13$ TeV MC simulations, by exploiting additional information contained within the B-hadron decays in the b-jet combined into a NN classifier, referred to as *jet vertex charge* (JVC). This algorithm uses tracks ghost-matched to a small-R calorimeter jet to define a set of high-level observables used in the NN. In addition to the charge of all tracks matched to a b-tagged jet, a charge of tracks originating from displaced secondary and tertiary vertices reconstructed by *JetFitter* algorithm introduced in Sec. 4.4.5 is calculated. A slightly modified version of the charge definition in Eq. 8.2 is used in this approach:

$$Q_{\text{jet}} = \frac{\sum_i q_i \left(p_{\text{T},i}\right)^{\kappa}}{\sum_i \left(p_{\text{T},i}\right)^{\kappa}}, \tag{8.3}$$

where the normalisation is provided by the sum of track $(p_{\text{T},i})^{\kappa}$ instead. This definition is used for both the charge of all the tracks associated with the b-jet, as well as for calculating charge of the tracks matched to the displaced vertices. Additionally, the charge of a soft-muon from semi-leptonic decay of the B-hadron within the b-jet is used. Multiple NN classifiers are trained, with different sets of inputs variables, since not all of them are always defined. For example, only approximately 11 % of b-jets contain a soft muon, therefore a standalone NN is trained using events which contain soft muon.

8.3.2 Observables used in the neural network

The previously mentioned measurements of top-quark charge relied on using the b-jet charge as the proxy. Nevertheless, the hadronically-decaying W boson also carries extra charge information and its charge of ± 1 is three times larger in magnitude than that of the b quark. Previous measurements have not exploited this option, presumably because the correct matching of light-quark jets to the corresponding hadronically-decaying W-boson is challenging. However, for boosted top-quark decays, the decay products are highly-collimated and can be clustered within a single large-R jet. The identification of tracks from individual quarks contained within the large-R jet is attempted using anti-k_t track jets with $R = 0.2$, ghost-matched with the large-R jet. Since top quark has three decay products, in an ideal case, it is expected that there should be three track jets within the large-R jet, one of which is b-tagged.

Therefore the charges of the leading-p_{T} b-tagged track jet and two leading-p_{T} non-b-tagged track jets are considered as inputs for the NN.

The choice of using three track jets, out of which one is b-tagged, is motivated by Fig. 8.1, showing the true origin of the track jets based on the matching with partonic decay products of the top (anti-)quark. The b-tagged track jet is in approximately 90 % of the cases originating from a b-quark, and most of the background jets are mistagged charm jets. The two highest-p_{T} non-b-tagged track jets are most frequently matched to non-b-quarks from the top (anti-)quark decay, with the sub-leading non-b-jet not being matched to any of the decay products in approximately 22 %. Including more track jets as inputs is discouraged based on Fig. 8.1d, which shows that in more than 60 % of cases, the 3$^{\mathrm{rd}}$-leading-p_{T} non-b-tagged track jet is not matched to any of the decay products of a top (anti-)quark.

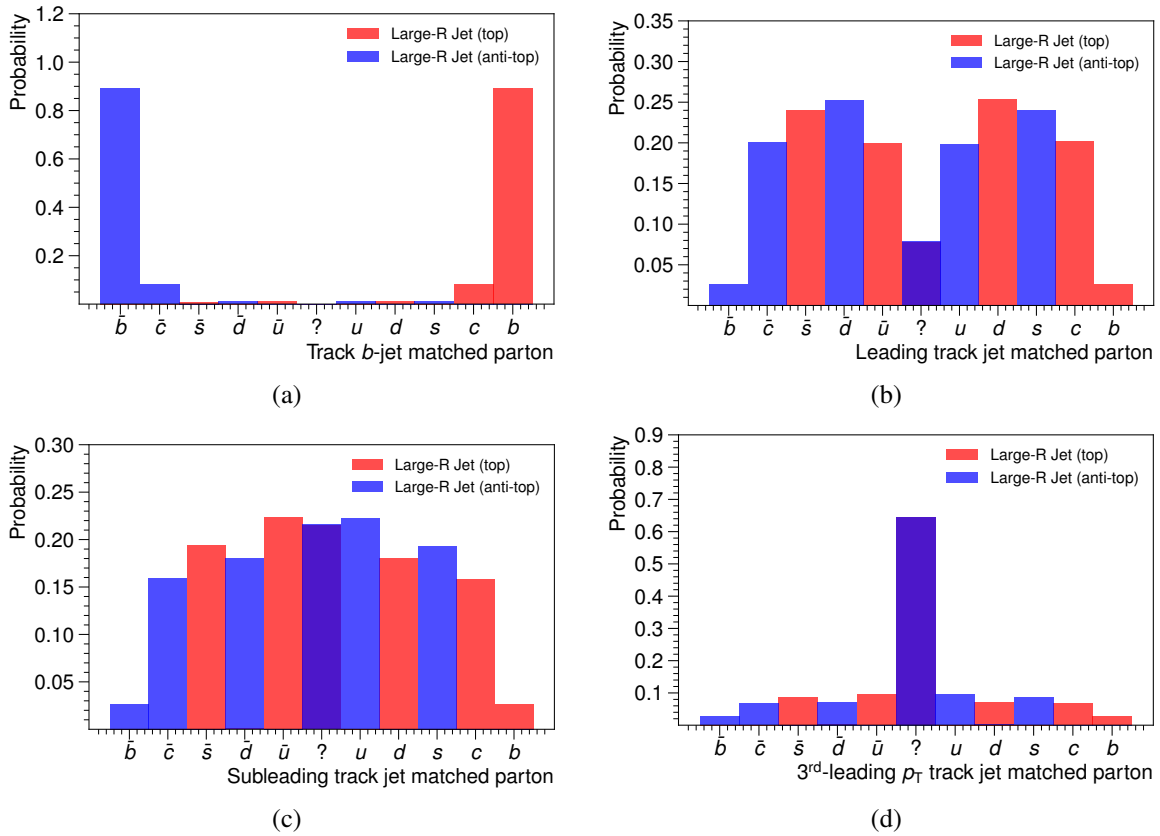


Fig. 8.1: The classification of the origin b-tagged track jet (a), the leading (b), 2$^{\mathrm{nd}}$-leading (c) and 3$^{\mathrm{rd}}$-leading (d) non-b-tagged track jet matched to a candidate large-R jet. The classification is based on matching a decay product from hadronically decaying top (anti-)quark with the track jet based on $\Delta R(\text{parton, track jet}) < 0.2$. The middle column in the plots, labelled as "?", marks track jets which were not matched to any of the decay products.

The JVC discriminant shows that exploiting information from displaced vertices of a B-hadron decay can provide additional discrimination power. The charges of secondary and tertiary vertices reconstructed by JetFitter for b-tagged track jets and relevant auxiliary variables are thus also considered for inclusion as NN inputs. The charge of the vertices reconstructed by the SV inclusive vertex-finding algorithm described in Sec. 4.4.5 is also considered. The aim is to examine whether there is additional information not extracted by JetFitter observables, that can be exploited for further discrimination of the charge of the B-hadrons initiating the b-tagged track-jets contained within the large-R jet candidates. The JVC discriminant itself is not used as an input for the NN, primarily because it was optimised on an

inclusive sample of $t\bar{t}$ events. Such a sample typically does not contain sufficient amount of generated evens for outlier regions of phase space, such as the high-p_{T} region examined in the all-hadronic channel. It is argued here, that adding some of the input observables directly into the NN trained in this study is more preferable, than nesting the JVC NN output as an input into another NN.

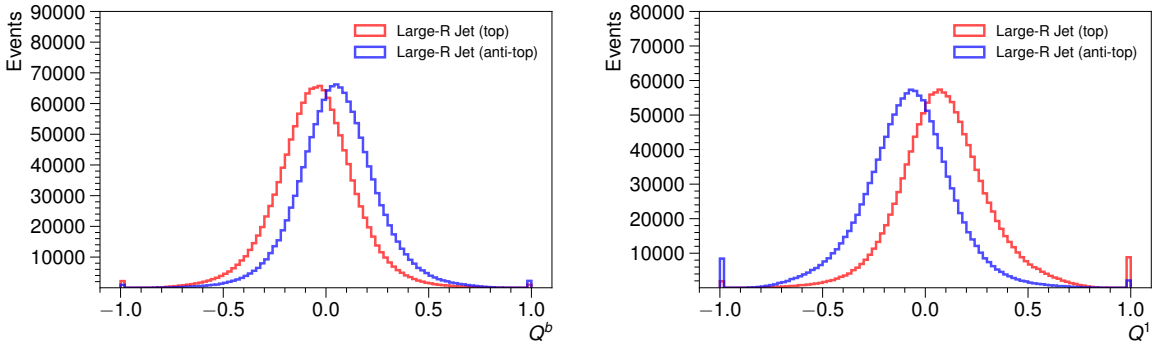
Table 8.4: List of observables of the track-jets matched to a large-R jet, which are considered as the NN input observables.

Variable(s)	Description
	b-tagged track jet variables
Q^{b}	Charge of the b-tagged track jet
p_{T}^{b}	p_{T} of the b-tagged track jet
N_{trk}^{b}	Track multiplicity of the b-tagged track jet
Q^{SV}	Charge of the SV vertex of the b-tagged track jet
$N_{\mathrm{trk}}^{\mathrm{SV}}$	Track multiplicity of the SV vertex
$L_{\mathrm{3D}}^{\mathrm{SV}}$	Distance of the SV vertex from primary vertex
$\Delta L_{\mathrm{3D}}^{\mathrm{SV}}$	Uncertainty on the distance
Q^{JFS}	Charge of the JetFitter secondary vertex
$N_{\mathrm{trk}}^{\mathrm{JFS}}$	Track multiplicity of the JetFitter secondary vertex
$L_{\mathrm{3D}}^{\mathrm{JFS}}$	Distance of the JetFitter secondary vertex from primary vertex
$\Delta L_{\mathrm{3D}}^{\mathrm{JFS}}$	Uncertainty on the distance
Q^{JFT}	Charge of the JetFitter tertiary vertex
$N_{\mathrm{trk}}^{\mathrm{JFT}}$	Track multiplicity of the JetFitter tertiary vertex
$L_{\mathrm{3D}}^{\mathrm{JFT}}$	Distance of the JetFitter tertiary vertex from primary vertex
$\Delta L_{\mathrm{3D}}^{\mathrm{JFT}}$	Uncertainty on the distance
	Non-b-tagged track jet variables
Q^{1}, Q^{2}	Charge of the leading and sub-leading non-b-tagged track jet
p_{T}^{1}, p_{T}^{2}	p_{T} of the leading and sub-leading non-b-tagged track jet
N_{trk}^{1}, N_{trk}^{2}	Track multiplicity of the leading and sub-leading non-b-tagged track jet
	Binary-decision variables
d^{1}	Whether the large-R jet has at least one non-b-tagged track jet
d^{2}	Whether the large-R jet has at least two non-b-tagged track jets
d^{JFS}	Whether JetFitter secondary vertex is reconstructed
d^{JFT}	Whether JetFitter tertiary vertex is reconstructed
d^{SV}	Whether SV vertex is reconstructed

The full list of observables investigated for the NN is listed in Table 8.4. Since not all of the observables are always defined, the binary-decision variables d^{1}, d^{2}, d^{JFS}, d^{JFT} and d^{SV} are added, which have a value of zero if the respective object is undefined, or one, if it is defined. The fraction of large-R jets in the sample, where the variables are defined, is listed in Table 8.5. For undefined objects, the respective variables such as p_{T}, track multiplicity or vertex position are set to zero. The corresponding charges are also set to zero, thus giving no discrimination for the class of the considered large-R jet. This is an alternative, simplified approach, where it is possible to train a single NN with

inputs that are not always defined, in contrast to the approach employed for the JVC, where for each combination of defined variables, a separate NN was trained.

For the charge observables, the method of how the charges of the tracks should be weighted, and the value of κ must be chosen. The definitions of jet charge which are considered, are defined in Eq. 8.1 and 8.3. The figure of merit in the optimisation is the separation power of the charge variables. The κ value is optimised separately for the individual charge observables. It is found that the two charge definitions give very similar separation for equal κ values, yielding no preference of one over the other. As such, the definition in Eq. 8.3 is chosen. The optimised values of κ from a scan in the range of $\kappa = 0.1$ up to 1.0 are shown in Table. In all cases a maximum in the scanned interval was found, corresponding to separation power quoted in the table. For illustration, the Q^b and Q^1 jet charge variables using the respective optimised κ values are shown in Fig. 8.2. All of the distributions of the NN input variables are shown in App. D.1.



(a) (b)

Fig. 8.2: Distribution of the Q^b (a) and Q^1 (b) track jet charges for large-R jets originating from top quark (red line) and top anti-quark (blue line). The spikes at values ±1 correspond to track jets, where all tracks have the same charge. Most frequently, these track jets are constituted by two tracks, which is the minimum track multiplicity requirement.

Table 8.5: Fraction of large-R jets for which respective features are defined, denoted by the binary-decision variables listed in the table.

Feature	d^1	d^2	d^{JFS}	d^{JFT}	d^{SV}
[%] of cases defined	95	59	95	54	95

Table 8.6: Optimised values of κ parameter for the charge definition of the individual charge observables. The separation power S for the optimal value of κ is also shown.

Variable	Q^b	Q^1	Q^2	Q^{SV}	Q^{JFS}	Q^{JFT}
κ	0.6	0.4	0.4	0.4	0.5	0.5
S [%]	8.8	15.2	9.2	9.6	5.1	1.4

Except for the charge observables in Table 8.6, the other variables are *auxiliary* in the sense that they do not have discrimination power itself, but may provide additional information to the NN through

correlations. Of particular interest are observables which have different correlations for each of the two classes of large-R jets, because such observables bring additional discrimination power that the NN can exploit. The correlations and anti-correlations of track multiplicities with charge are an example of this property, as shown in Fig. 8.4 and 8.5, motivating why the track multiplicity variables are included as the NN inputs. The distances of vertices from primary vertex and corresponding fitted distance errors are tested as NN input observables since they could be expected to provide some metric of the quality of the reconstructed vertex. Similarly the p_{T} of the individual jets are tested as NN inputs under the assumption that the kinematic information may be relevant since the tracking performance is expected to deteriorate with high p_{T}. Whether these assumptions transfer into a better discrimination of the trained NN is not clear, therefore a test needs to be designed to assess the gain of inclusion of these observables in the NN. The approach used in this study is to train multiple NN setups with different choices of input observables, and compare the performance of these networks. First, the NN architecture and training is discussed in the following section, and subsequently the NN setups are compared in Sec. 8.3.4.

8.3.3 Neural network architecture and training

The task of top-quark charge discrimination falls within the domain of a classification problem, where the goal is to design a function $f(\mathbf{x}, \mathbf{w})$, which maps the input observables $\mathbf{x}$ into a single discriminant, optimising the free parameters $\mathbf{w}$ to achieve the best possible separation between the two classes of large-R jets, where one class of jets is initiated by a top-quark and the other class of jets is initiated by top anti-quark. The optimisation is performed using a training set of large-R jets, for which the true origin is known, and the parameters are optimised based on a *loss* function [138] which evaluates how well the function separates the two classes. This type of optimisation is usually referred to as *supervised learning* [138]. The type of loss function used here is described further in this section.

Neural networks are one particular example of a representation of a basis of functions which can be used to construct arbitrary function for classification. The name of this ML technique draws inspirations from how the neuron cells inside a brain function. Individual neurons are interconnected, where the neuron connections (synapses) store a piece of information, impacting how the network of neurons responds to different signals. An illustration of a fully-connected feed-forward NN is shown in Fig. 8.3, the architecture used in this study. A general linear combination of the N inputs $\mathbf{x}$ is fed as input to each of the M hidden-layer neurons, each having an activation function $\phi(\theta)$, returning value a_j. While N is given by the number of input variables, M is a free parameter. The linear combination of $\mathbf{x}$ is given by the weights w_{ij} and biases b_j where $i = 1..N$, $j = 1..M$. Various different activation functions exist, some of which are discussed in this section, but they all have in common that they are *non-linear*. Finally, a linear combination given by weights w_j and bias b, of the outputs a_j of the M hidden-layer activation functions is fed to the single output neuron with its own activation function $\phi'(\theta)$. The network architecture is called fully-connected since each output of each neuron from previous layer in general enters as input of each neuron in the next layer. The architecture of the NN is called feed-forward, when the information in the NN is propagated only forwards, with no back-propagation into previous layers. The NN architecture can have many hidden layers, where the outputs of each layer are fed as a linear combination given by weights and biases into activation functions of the neurons of the next layer.

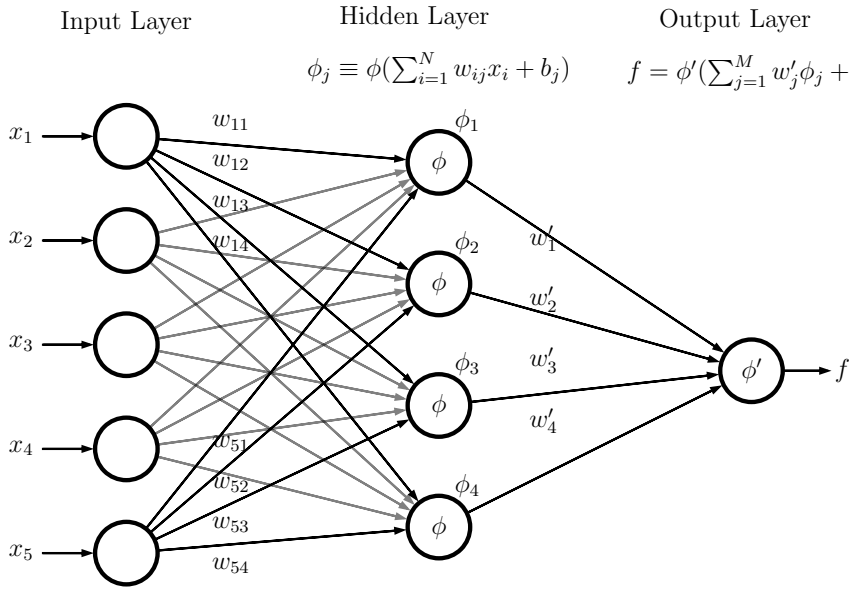


Fig. 8.3: Illustration of NN with a single hidden layer [215].

Given a fixed NN architecture, the task is to find optimal values of its weights and biases. This is performed by minimising the loss function using the jets from the training set. The loss function used in this study is the *binary cross-entropy* [216], defined as follows:

$$L(\mathbf{w}) = \sum_{i}^{\text{jets}} \left[f^i_{\text{true}} \log(f(\mathbf{x}^i, \mathbf{w})) + (1 - f^i_{\text{true}}) \log(1 - f(\mathbf{x}^i, \mathbf{w})) \right]. \tag{8.4}$$

The sum in the loss function runs over all large-R jets in the set, where both classes of jets are represented equally in the set in this study. The f^i_{true} denotes the true class of the large-R jet, denoted as zero for jet originating from top anti-quark and unity for jet from top quark. The $f(\mathbf{x}^i, \mathbf{w})$ is the NN output value from interval $(0, 1)$ for the inputs $\mathbf{x}$ of the i^{th} jet. The weights and biases of the NN are further jointly referred to as weights, and are symbolically labelled as $\mathbf{w}$ in the equation. The binary cross-entropy is based on the Bernoulli distribution that quantifies what is the probability of observing a jet of class f^i_{true}, given the observed features $\mathbf{x}^i$ and the separation function $f(\mathbf{x}, \mathbf{w})$. For a set of training jets, the corresponding likelihood employing the Bernoulli distribution is:

$$L(\mathbf{w}) = \prod_{i}^{\text{jets}} P(f^i_{\text{true}}, \mathbf{x}^i) = \prod_{i}^{\text{jets}} f(\mathbf{x}^i, \mathbf{w})^{f^i_{\text{true}}} \left[1 - f(\mathbf{x}^i, \mathbf{w}) \right]^{(1 - f^i_{\text{true}})}. \tag{8.5}$$

The binary cross-entropy in Eq. 8.4 is thus the log-likelihood of Eq. 8.5.

The minimisation of the loss function is a challenging problem, due to the fact that the weights are highly correlated and the function in general has many local minima. Typically, methods based on stochastic gradient descent are used. For a gradient descent, the weights are initially selected randomly and the value of the loss function is calculated using the initial weights and the values of $\mathbf{x}^i$ of the jets from the training set. After that, the gradient of the loss function is calculated, and the weights are shifted by $\mathbf{w} \to \mathbf{w} + \eta \nabla_w$, where η is the learning rate, a free parameter impacting the size of the gradient step in the minimisation. It can be shown that the gradient of the loss function is obtained

via derivative chain-rule, which effectively results in a back-propagation procedure, where first the change in the weights of the output neuron is calculated and subsequently this change is propagated backwards into the neurons of the previous layer [138]. The procedure is recursively repeated through all the layers until all weights are updated. The *stochastic* nature of the gradient descent is achieved by randomly selecting a small subset of the training set used for gradient calculation, referred to as a *batch*. The aim of this approach is two-fold, the randomness of the batch reduces the likelihood of the gradient descent getting stuck in a local minimum, and it also results in a faster gradient update calculation, given that the gradient is calculated for a small sub-set of jets from the training-set. A single period in the training, i.e. when all of the training jets are looped over in the batches, is referred to as an *epoch*. The NN training typically involves many epochs, and the necessary amount of epochs for convergence is a problem-dependent task.

In this study, the minimisation of the loss function is performed using the ADAptive Moments (ADAM) minimising algorithm [217], which is based on the stochastic gradient descent, but also takes into account the size of the update of weights from previous iteration in the current update of weights. This further improves the speed of convergence and robustness against falling into a local minimum [217, 218]. The code for the NN training uses a combination of the *Keras* [219] and *Tensorflow* [220] Python-based libraries.

In the following, we summarise all of the free parameters related to the NN architecture and the training, which require a-priori choice as they are not determined in the training of the NN. They are commonly referred to as *hyper-parameters*.

The choice of learning rate and batch size. The choice of learning rate impacts the size of the steps of the stochastic gradient descent. Too small values increase the training time. Large values may lead to oscillations around a local minimum or divergences. The size of the batch affects the stochastic nature of the gradient descent, impacting the representativeness of the batches in contrast to the full training set, as well as the convergence of the minimisation. The choice of the batch size is in general problem-dependent.

The number of hidden layers and the number of neurons in each hidden layer. In particular, NNs with more hidden layers may learn more features at different levels of abstraction. The success in generalisation however strongly depends on the other aspects such as what features are provided as inputs, what is the NN architecture, and the size and representativeness of the training set.

The activation function used for the neurons. Activation functions define the behaviour of response of the neurons to the input. The following activation functions are investigated for the NN architecture, the *sigmoid*, *hyperbolic tangent* (tanh), *rectified linear unit* (ReLU) and the *exponential linear unit* (ELU) [138]. The sigmoid function is defined as:

$$\phi(\theta) = \frac{1}{1 + e^{-\theta}}. \tag{8.6}$$

The sigmoid function has some nice properties, such as having a smooth derivative and being bounded to $(0, 1)$ interval, therefore large values due to outlier values are not encountered. On the other hand, it suffers from vanishing gradients for θ far from zero. The computational complexity of calculation of the function combined with the vanishing gradient problem lead to slower convergence and thus longer training times compared to alternatives. Sigmoid, however, is a popular choice for the activation function of the output node of a binary-classification NNs [138], and is chosen as such in this study as

well. The tanh function is basically a scaled sigmoid, with the difference of a steeper gradient and the output range of (-1, 1):

$$\tanh(\theta) = 2\phi_{\text{sig}}(2\theta) - 1, \tag{8.7}$$

where ϕ_{sig} is the sigmoid function from Eq. 8.6. The ReLU function is defined as:

$$\phi(\theta) = \max(0, \theta). \tag{8.8}$$

The advantage of ReLU is its computational simplicity compared to sigmoid and tanh and no vanishing gradient problem. On the other hand, it is potentially susceptible to undesirably large activation, because the function is unbounded for $\theta > 0$. The zero gradient for negative θ also potentially poses a problem in the optimisation, since no update of weight can happen, however the occurrence of this situation is mitigated by the stochastic nature of the training in batches. Finally, the ELU function is defined as:

$$\phi(\theta) = \begin{cases} (e^{\theta} - 1) & \text{if } \theta \leq 0 \\ \theta & \text{if } \theta > 0 \end{cases}. \tag{8.9}$$

Unlike ReLU, ELU is differentiable for all values of θ including zero, and has non-zero (though vanishing) gradient for negative values. While not as computationally efficient as ReLU, ELU has been found to converge faster and provide better NN classification performance than ReLU for various architectures including the feed-forward networks [221] employed in this study.

The regularisation choice and strength. During the training, it is important to ensure the NN is not overtrained, meaning that it does not fit statistical noise in the training set. One of the symptoms of NN overtraining is that the trained NN will have large values of weights due fitting statistical fluctuations in the training set. Therefore, the use of regularisation is investigated, by adding a penalty into the loss function. The L1 and L2 regularisations are investigated, which are defined as:

$$R_{\text{L1}}(\mathbf{w}) = \sum_{i=1}^{n} |w_i|, \qquad R_{\text{L2}}(\mathbf{w}) = \sum_{i=1}^{n} w_i^2, \tag{8.10}$$

where $\mathbf{w}$ is the vector of all the weights and biases of all neurons in the network, and n is the size of the w vector. The loss function is modified as follows:

$$L(\mathbf{w}) = L_0(\mathbf{w}) + \lambda R(\mathbf{w}), \tag{8.11}$$

where the $L_0(\mathbf{w})$ is the loss function without regularisation defined in Eq. 8.4, $R(\mathbf{w})$ is the introduced regularisation and the free parameter $\lambda > 0$ is the regularisation strength.

The NN is trained using jets from the signal all-hadronic $t\bar{t}$ sample, which pass the selection criteria in Sec. 8.2.2. During the training, the value of the loss function after each epoch is calculated on an independent set of jets, referred to as the validation set. This is done to check for overtraining, manifesting itself by a decrease in the value of the training loss[(6)] accompanied by increasing validation loss[(7)], hinting that the NN is fitting statistical noise in the training set, not present in the validation

[(6)] It is common to refer to the value of a loss function simply as *loss*. The value of the loss function averaged over all batches within an epoch is referred to as the training loss.

[(7)] The validation loss is the value of the loss function for the entire validation set.

set. Additionally, after the training, the NN classifier and the ROC curve are compared to another independent set of jets, referred to as the testing set. Therefore, the initial sample of jets needs to be split into three parts. However, the jets from the same simulation are also used for the all-hadronic charge asymmetry analysis. It is in general discouraged to use the same jets for both training and evaluation of the NN, therefore a different strategy is required.

A so-called *cross-training* approach is used in this study, where two independent NNs with the same architecture are trained on two independent sets of jets. In the subsequent analysis employing the NNs, it is ensured that the NN trained on one set of jets is not evaluated on the same set of events. The initial sample of jets is split into two equal-size samples by using a unique per-event number, where one sample contains odd-number events, and the other event-number events. After that, each of the two samples is further randomly split into two equal-size sub-samples. The four sets are then used as training, testing and validation sets for training of the two networks NN_1 and NN_2, according to an assignment scheme described in Table 8.7. Both of the networks are trained using the same hyper-parameter values. Each of the networks is trained on 5.6 million jets. The testing and validation set each have additional 2.8 million jets.

Table 8.7: Split of jets into the four samples labelled as $\mathcal{T}_1$, $\mathcal{T}_2$, $\mathcal{V}_1$ and $\mathcal{V}_2$. The samples $\mathcal{T}_1$ and $\mathcal{V}_1$ are used as the training set of NN_2, while $\mathcal{T}_2$ and $\mathcal{V}_2$ are used as the testing and validation set of NN_2, respectively. Analogously, NN_1 is trained on $\mathcal{V}_2$ and $\mathcal{T}_2$ and tested on $\mathcal{T}_1$, while $\mathcal{V}_1$ is used as the validation set. Each of the four samples contains 2.8 million jets.

	Event-number split	
	odd number	even number
Random split	$\mathcal{T}_1$ $\mathcal{V}_1$	$\mathcal{T}_2$ $\mathcal{V}_2$

Before the training, an input variable pre-processing step is executed, where each variable is shifted by an offset and scaled by a scale, such that the observables have zero mean and a standard deviation of unity. The scale and offset are thus the original distributions' mean and standard deviation. This procedure limits the range of all of the input observables into approximately the same range, thus preventing variables with large value range from dominating the NN weights. For activation functions with vanishing gradient problem, such variables may even prevent the training from converging.

The training is performed with a maximum number of 200 epochs. However, the plateau of the achievable training loss may be reached much sooner, after which there is no point in continuing the training, as this can potentially lead to overtraining. The validation set can be used to gauge whether the training has reached a saturation point, where the validation loss no longer decreases. Therefore, the training is stopped early, if the validation loss does not improve during 10 training epochs, by at least a value of 0.001. The number of epochs and the minimum improvement required were adjusted based on the observed training and validation loss evolution for NNs trained using a set of randomly picked hyper-parameter configurations.

A hyper-parameter grid search is performed, where NNs are trained with different values of parameters as listed in Table 8.8 to decide their optimal values. For each trained NN, the initial

performance assessment is made using the testing-set *accuracy*, which is defined as:

$$A = \frac{1}{N_{\text{jets}}} \sum_{i}^{\text{jets}} \delta\left(f^i_{\text{true}} - f(\mathbf{x}^i)\right), \tag{8.12}$$

$$\delta\left(f^i_{\text{true}} - f(\mathbf{x}^i)\right) = \begin{cases} 1 & \text{if } \left(f^i_{\text{true}} = 1 \wedge f(\mathbf{x}^i) > 0.5\right) \vee \left(f^i_{\text{true}} = 0 \wedge f(\mathbf{x}^i) < 0.5\right) \\ 0 & \text{if } \left(f^i_{\text{true}} = 1 \wedge f(\mathbf{x}^i) < 0.5\right) \vee \left(f^i_{\text{true}} = 0 \wedge f(\mathbf{x}^i) > 0.5\right) \end{cases}. \tag{8.13}$$

The accuracy is thus a mean rate of successfully classifying jets, where the binary classification of a jet is based on whether it's score is greater or smaller than 0.5. This metric is a reasonable choice if both classes of jets are equally represented, which is the case for the $t\bar{t}$ events used. Hyper-parameters resulting in higher NN accuracy are preferred, however it is also required that the NN does not exhibit signs of overtraining.

Table 8.8: Summary of the hyper-parameters and the architecture setup in the NN grid search.

Hyper-parameter	Scanned values of the parameter
Optimiser	Adam
Number of hidden layers	1 – 5
Number of neurons per hidden layer	5 – 30
Hidden layer activation function	ReLU, ELU, tanh
Output layer activation function	sigmoid
Learning rate	10^{-5}, 5×10^{-5}, 10^{-4}
Batch size	100, 200, 300, 400, 500
L1 regularisation strength	10^{-6}, 5×10^{-6}, 10^{-5}, 5×10^{-5}, 10^{-4}
L2 regularisation strength	10^{-6}, 5×10^{-6}, 10^{-5}, 5×10^{-5}, 10^{-4}

8.3.4 Choice of input variables

The NNs are trained for five different sets of input variables, and their accuracy is compared in Table 8.9. This approach is a compromise between the number of different combinations of input variables and the amount of time it takes to train the NNs. In general, different sets of input variables require full repetition of the hyper-parameter search. The five groups differ in the choice of whether SV and/or JetFitter vertex charge and corresponding N_{trk} variables are added, whether the p_{T} of individual track jets are added, and whether the corresponding vertex L_{3D}, ΔL_{3D} are added.

The motivations for the structure of the variable groups stems from the input variable linear correlations, shown in Fig. 8.4 and 8.5. These show that the charges of jets and vertices are correlated with track multiplicities, and as expected, the correlations are opposite for jets from top quark and top anti-quark, making these variables useful for additional discrimination. The track-jet transverse momenta and the distances of the vertices from primary vertex show some degree of correlations between each other and the track multiplicities, but no or negligible correlations with the relevant charge observables. This prompts the test of adding and removing these variables from the NNs. Finally, while Q^{SV} shows relatively strong separation according to Table 8.6, it is approximately 50 % correlated with the Q^b, potentially reducing the additional information this variable brings. The Q^{JFS}

and Q^{JFT} are found to be less correlated with Q^{b}. Therefore, setups where either SV or JetFitter or both vertex-related variables are removed from the NN are also tested.

Table 8.9: Trained NN setups with various combinations of inputs variables. The variable group 1 includes the track-jet charges, track multiplicities and the binary-decision variables for the non-b-tagged track jets. The rest of the groups add track-jet p_{T}, JetFitter vertex variables, SV vertex variables, and the corresponding vertices' fitted distance from primary vertex and the distance error, respectively. For every NN setup, the best accuracy achieved is quoted. The "∘" symbol marks a group of variables included in a given setup. The chosen setup is highlighted in bold.

		NN inputs setup				
	Group of variables	1	2	3	**4**	5
1	Q^{b}, Q^{1}, Q^{2}, N^{b}_{trk}, N^{1}_{trk}, N^{2}_{trk}, d^{1}, d^{2}	∘	∘	∘	**∘**	∘
2	p^{b}_{T}, p^{1}_{T}, p^{2}_{T}	∘	∘	∘		∘
3	Q^{JFS}, Q^{JFT}, $N^{\mathrm{JFS}}_{\mathrm{trk}}$, $N^{\mathrm{JFT}}_{\mathrm{trk}}$, d^{JFS}, d^{JFT}			∘	**∘**	∘
4	Q^{SV}, $N^{\mathrm{SV}}_{\mathrm{trk}}$, d^{SV}		∘		**∘**	∘
5	Corresponding vertex L_{3D}, ΔL_{3D}		∘	∘		∘
Total number of input variables		8	16	21	**17**	26
Accuracy		71.3	71.8	72.4	**72.9**	73.0

8.3.5 Optimised NN architecture and comparisons with alternative approaches

The setup labelled as number 4 in Table 8.9 is used for further studies and the charge asymmetry sensitivity estimate, as it reaches almost the same accuracy as the setup including all observables, but using a much smaller set of input variables (17 vs 26). The optimised values of hyper-parameters of the trained NN are listed in Table 8.10. The comparison of the NN discriminant for the training and testing set is shown in Fig. 8.6 and the evolution of the training loss and validation loss for both cross-trained NNs in Fig. 8.8. The validation loss is found to be systematically smaller than that of the training loss, which is not unexpected. Firstly, the training loss is evaluated continuously after each batch within an epoch, and the average training loss is quoted, whereas the validation loss is calculated only after the full epoch. The training loss average thus includes higher loss values from early batches in the epoch, increasing the value of the loss average. Additionally, the regularisation term is only added to the training loss function and not the validation loss, further increasing the training loss in contrast to validation loss. Nevertheless, a cross-check was performed by randomly swapping the four samples in Table 8.7 and repeating the NN training. The same trend in training and validation loss was reproduced, ruling out the issue of jets with outlier properties not being evenly represented in all four samples. In Fig. 8.7, the ROC curves for both cross-trained NNs are shown, showing very similar training set and testing set accuracy. An approximately 0.1 % accuracy difference is observed between the two cross-trained NNs. This is attributed to slightly different convergence of the NN training and validation loss due to the early stopping, and not due to overtraining, since both training and validation loss in Fig. 8.8b show similar trend.

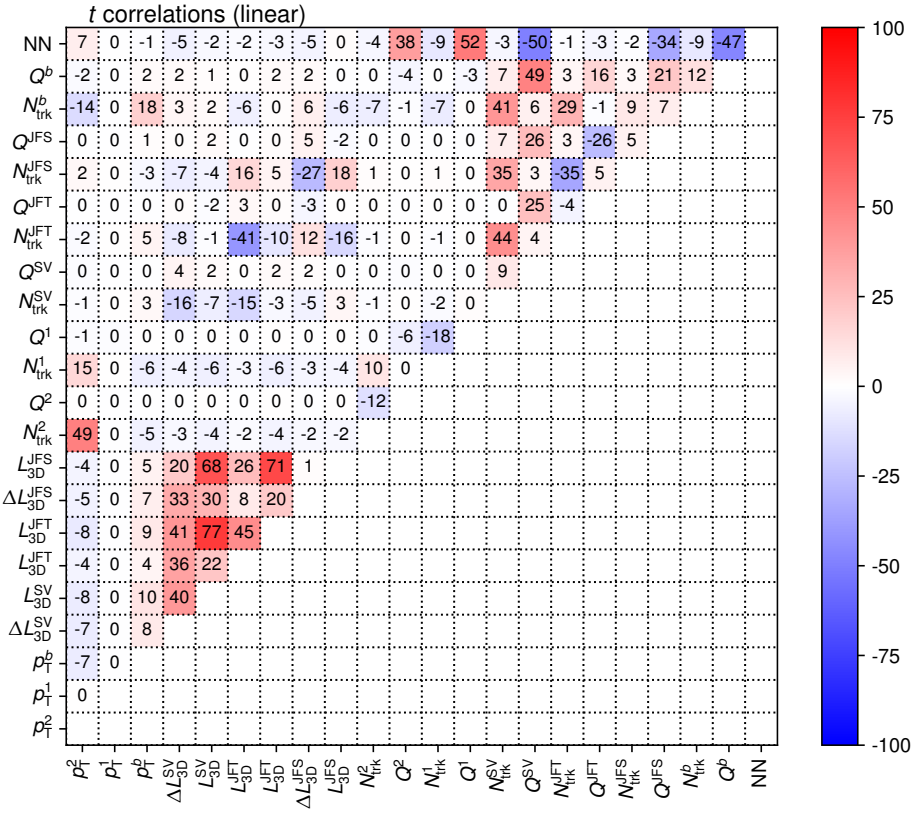


Fig. 8.4: Linear correlations for the NN input variables and the discriminant for top-quark jets, expressed in %.

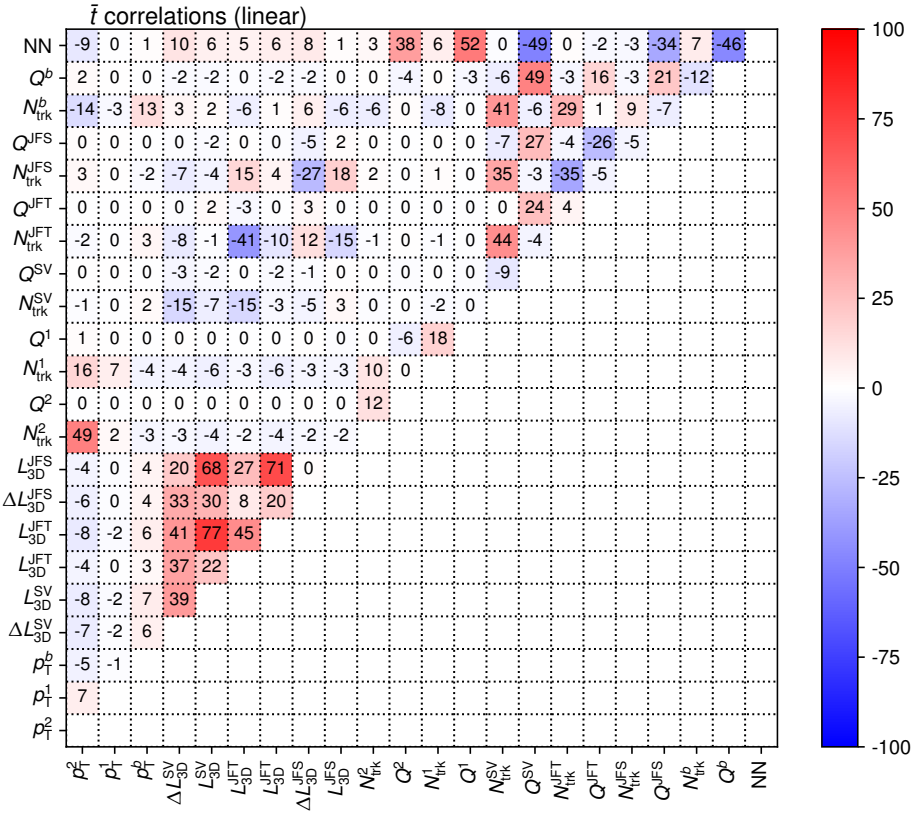
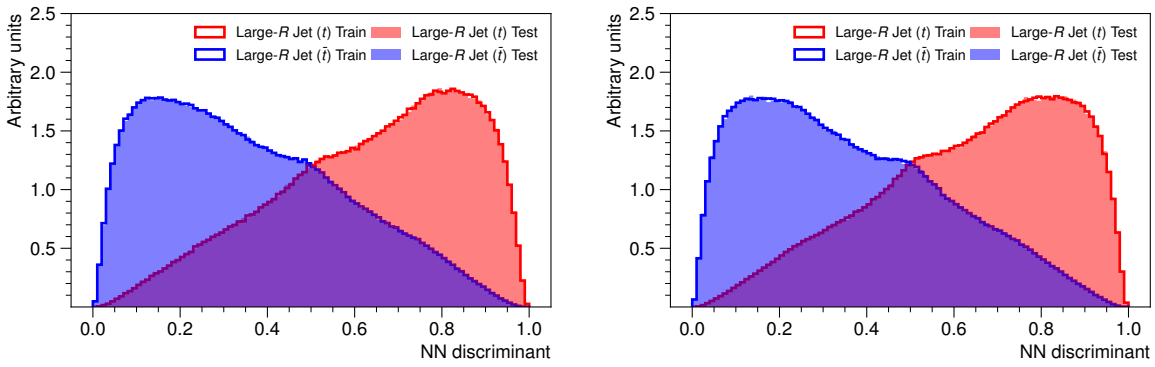


Fig. 8.5: Linear correlations for the NN input variables and the discriminant for top-anti-quark jets, expressed in %.

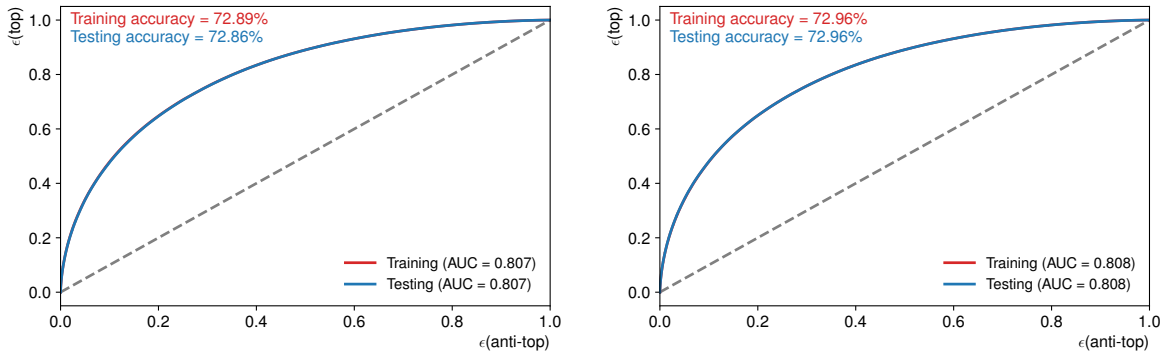
Table 8.10: The chosen hyper-parameters and the architecture of the optimised NN.

Hyper-parameter	Value
Optimiser	Adam
Number of hidden layers	2
Number of neurons in the hidden layers	15, 10
Hidden layer activation function	ELU
Output layer activation function	sigmoid
Learning rate	10^{-4}
Batch size	100
L2 regularisation strength	5×10^{-5}



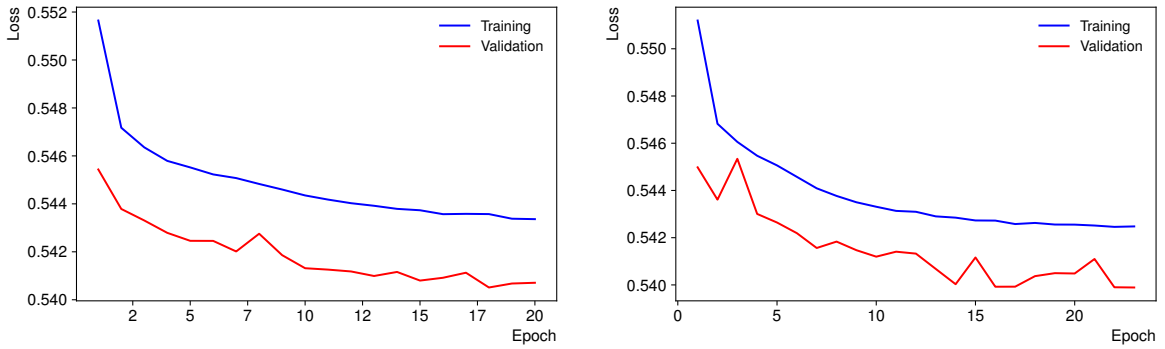
(a) (b)

Fig. 8.6: The comparison of the NN discriminant for the training and testing set, for the two cross-training configurations in (a) and (b) respectively, as defined in Table 8.7.



(a) (b)

Fig. 8.7: The comparison of ROC curves for the training and testing set, for the two cross-training configurations in (a) and (b) respectively.



(a) (b)

Fig. 8.8: The evolution of the training and validation loss as a function of the training epoch, for the two cross-training configurations in (a) and (b) respectively.

In the all-hadronic charge-asymmetry measurement, the NN is applied to the two candidate large-R jets passing the signal region criteria defined in Sec. 8.2.1. The two jets are denoted as J_1 and J_2, with their NN input values $\mathbf{x}_{J_1}$ and $\mathbf{x}_{J_2}$, respectively. They are identified as the top-quark jet J_t and top anti-quark jet $J_{\bar{t}}$ by comparing the NN scores $f(\mathbf{x}_{J_1})$ and $f(\mathbf{x}_{J_2})$:

$$\begin{aligned} f(\mathbf{x}_{J_1}) > f(\mathbf{x}_{J_2}) \Rightarrow J_t \equiv J_1, \quad J_{\bar{t}} \equiv J_2, \\ f(\mathbf{x}_{J_1}) < f(\mathbf{x}_{J_2}) \Rightarrow J_t \equiv J_2, \quad J_{\bar{t}} \equiv J_1, \end{aligned} \tag{8.14}$$

where the cross-training NN setup is employed as discussed in Sec. 8.3.3. The main source of dilution in the all-hadronic channel A_C measurement is expected to be the mis-identification of the jets in Eq. 8.14, resulting in wrong determination of the sign of $\Delta|y| = |y_t| - |y_{\bar{t}}|$, where $|y_t|$ is the absolute rapidity of J_t and $|y_{\bar{t}}|$ is the absolute rapidity of $J_{\bar{t}}$. Therefore, the figure of merit when comparing the NN with alternative approaches, is the $\Delta|y|$ sign assignment purity, labelled as $\mathcal{P}_{\Delta|y|}$. It is defined as the fraction of events where the sign of reconstructed $\Delta|y|$ matches the sign of $\Delta|y|^{\text{parton}} = |y_t^{\text{parton}}| - |y_{\bar{t}}^{\text{parton}}|$, calculated from the true absolute rapidities of the top-quark and top anti-quark at parton level, and is evaluated in the signal region using $t\bar{t}$ MC simulated events.

For the comparison with alternative approaches, the $\mathcal{P}_{\Delta|y|}$ is also calculated using only the b-tagged track jets matched to the large-R jets, either by using the inclusive jet charge Q^b, or by using the aforementioned JVC discriminant. In these approaches, the charge classification of the candidate large-R jets is performed based on the assumption that the top-quark jet should contain a b-jet with smaller value of Q^b or JVC value than the b-jet matched to top-anti-quark jet. Finally, the $\mathcal{P}_{\Delta|y|}$ is also calculated for the charge asymmetry measurement in the single-lepton channel, where the charge assignment is performed based on the single isolated lepton in final state.

Since the all-hadronic channel targets a specific kinematic region, the $\mathcal{P}_{\Delta|y|}$ comparisons are performed in the two highest-$m_{t\bar{t}}^{\text{parton}}$ bins corresponding to the single-lepton channel A_C measurement in Ch. 7, where the $m_{t\bar{t}}^{\text{parton}}$ is calculated from the parton-level top quark and anti-quark four-vectors. The true $m_{t\bar{t}}$ definition is used instead of reconstructed $m_{t\bar{t}}$ in order to remove the differences in $t\bar{t}$ system reconstruction between all-hadronic and single-lepton channel in the comparison. The $\mathcal{P}_{\Delta|y|}$ results are summarised in Table 8.11.

Table 8.11: Comparison of the $\Delta|y|$ charge assignment purity $\mathcal{P}_{\Delta|y|}$ in the all-hadronic channel obtained using the NN and by using b-tagged track-jet charge only, and in the single-lepton channel, in bins of parton-level $m_{t\bar{t}}$.

Channel, method	$\mathcal{P}_{\Delta\|y\|}$ [%] vs $m_{t\bar{t}}^{\text{parton}}$ [GeV]	
	[1000, 1500]	> 1500
All-hadronic, NN discriminant	**82.3**	**82.3**
All-hadronic, b-jet charge	65.9	64.8
All-hadronic, JVC discriminant	67.2	66.7
Single-lepton combined (resolved+boosted)	91.4	90.2
Single-lepton resolved	90.5	85.6
Single-lepton boosted	92.6	93.0

The $\mathcal{P}_{\Delta|y|}$ comparisons show that the NN approach adopted here significantly outperforms the approach of using just the b-tagged jet charge, as was traditionally done in resolved $t\bar{t}$ topologies. It is difficult to draw conclusions from the comparison with the observed JVC performance, since it is unclear whether the JVC discriminant can be used as-is for the kinematic regime examined in this study, due to the reasons outlined in Sec. 8.3.2 on the representativeness of the training set used for the JVC training. The $\mathcal{P}_{\Delta|y|}$ purity in the single-lepton channel is still higher, than in the all-hadronic channel. This is not unexpected, since the charge of the isolated lepton is a very good proxy to the charge of the corresponding top quark, and so the dilution of $\Delta|y|$ sign is driven primarily by mis-reconstruction of object kinematics, such as incorrect neutrino longitudinal component determination, or incorrect jet-to-parton assignment. The boosted single-lepton topology shows the highest $\mathcal{P}_{\Delta|y|}$, since the reconstruction of a hadronically-decaying top-quark as a single large-R jet is more robust with respect to $\mathcal{P}_{\Delta|y|}$ than the jet-to-parton assignment in resolved topology.

8.4 Validation of the neural network in data

In order to use the NN in any measurement relying on MC simulations, it is necessary to compare the modelling of the input observables and the NN discriminant in MC simulation and in data, and assess potential related systematic uncertainties. The all-hadronic channel is dominated by two processes, the MC simulated $t\bar{t}$ signal and the data-driven multijet background. Thus, the modelling study should be performed in a sample of boosted hadronically-decaying $t\bar{t}$ events in data. The boosted all-hadronic channel suffers from a substantial multijet background contribution of the order of 20 % of total prediction and thus is not a very clean channel for this purpose. Instead, a tag-and-probe approach in the boosted $t\bar{t}$ production in the single-lepton channel is used, specifically, the muon $2b$-inclusive channel defined by selection criteria in Sec. 8.2.3. The selection yields a very pure sample of $t\bar{t}$ events, as shown in Table 8.3, where the hadronically-decaying top quark is reconstructed as a large-R jet, and is used as the probe for the modelling study. Further split of the muon $2b$-inclusive channel into μ^+ and μ^- channels based on the muon charge is done, in order to examine the modelling of the relevant observables separately for top-quark jets and top-anti-quark jets.

8.4.1 Systematic uncertainties

The systematic uncertainties evaluated in this study are very similar to the experimental and modelling uncertainties outlined in Sec. 7.6.1 and 7.6.2. Below, we highlight the differences, otherwise it is assumed that the same prescriptions of the corresponding uncertainties are considered. The differences in treatment arise primarily from newer calibrations with improved uncertainties or newer ATLAS recommendations on treatment of MC modelling uncertainties at the time of this analysis.

Experimental uncertainties

The same experimental uncertainties as in Sec. 7.6.1 are evaluated in this analysis, with the exception of updated large-R jet uncertainties.

The large-R jet calibration is improved by including the in-situ calibration [115] using similar methodology as for the small-R jets. In total, 24 NPs related to JES are propagated in the analysis, and a single NP for the uncertainty on JER. The uncertainties on the large-R jet JMS and JMR are the same as those outlined in Sec. 7.6.1. No calibration and related uncertainties on the large-R jet DNN top-tagger were available at the time of this analysis. Based on the previous studies presented in Ch. 6 and specifically Fig. 6.6, it is expected that the uncertainties on the top-quark tagging signal efficiency will mostly impact the total $t\bar{t}$ yield in this study, and will have negligible shape effects on the top-quark charge NN inputs and the NN discriminant. Given the small background contribution, uncertainties on the top-quark tagging background rejection can be neglected.

Signal modelling uncertainties

The $t\bar{t}$ modelling uncertainties considered in this analysis are slightly updated with respect to Sec. 7.6.2, namely the uncertainty on the initial-state radiation and the matching of ME to PS. The initial-state radiation uncertainty is decomposed into three NPs. The first two NPs correspond to factor 0.5 and factor 2.0 variations of the μ_R scale and the μ_F scale in the ME, respectively. The third NP corresponds to equally-sized variations of μ_R scale of the Var3c parameter of the A14 tune [174] in the parton shower. Finally, the uncertainty on the matching of ME to PS is estimated by varying the nominal value of $h_{\text{damp}} = 1.5\ m_{\text{top}}$ to a value of $h_{\text{damp}} = 3\ m_{\text{top}}$. The PS and hadronisation uncertainty, the FSR uncertainty, and the PDF uncertainties are the same as in Sec. 7.6.2.

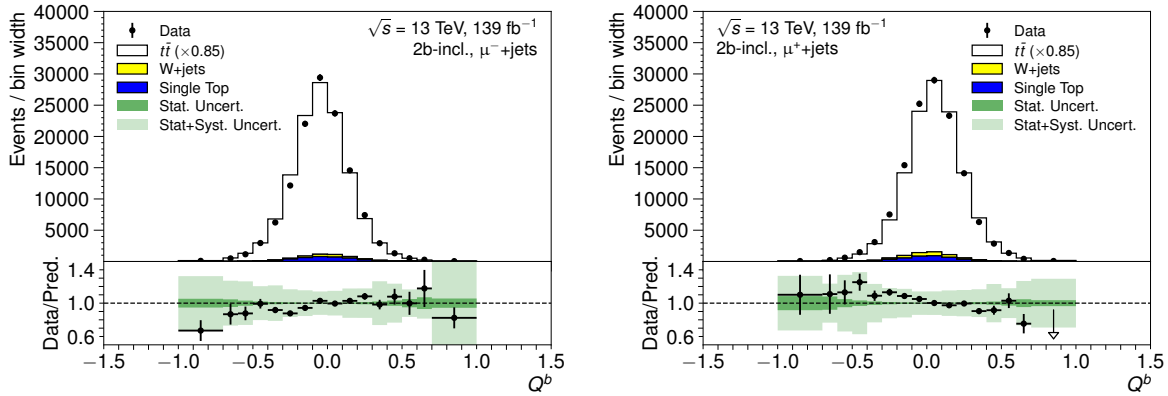
Given the small contribution of the backgrounds in the 2b-inclusive muon channel, background modelling uncertainties are neglected.

8.4.2 Control distributions of the NN inputs and discriminant

In Fig. 8.9– 8.14, the comparisons of data and predictions are shown for the NN input variables and the NN discriminant of the probed isolated b-tagged and top-tagged large-R jet, referred to as the hadronic top-quark large-R jet candidate. Due to the previously-discussed top-quark p_{T} mismodelling, the $t\bar{t}$ contribution in all of the figures is scaled by a factor of 0.85 to normalise the prediction to data, in order to highlight shape differences between data and the prediction. The track-jet charge and the large-R jet NN discriminant plots are shown separately depending on the sign of the isolated muon electric charge, which determines whether the selected large-R jet probes a hadronically-decaying top quark or top anti-quark. The inclusive NN discriminant distribution of $t\bar{t}$ in Fig. 8.13 shows a decomposition of the

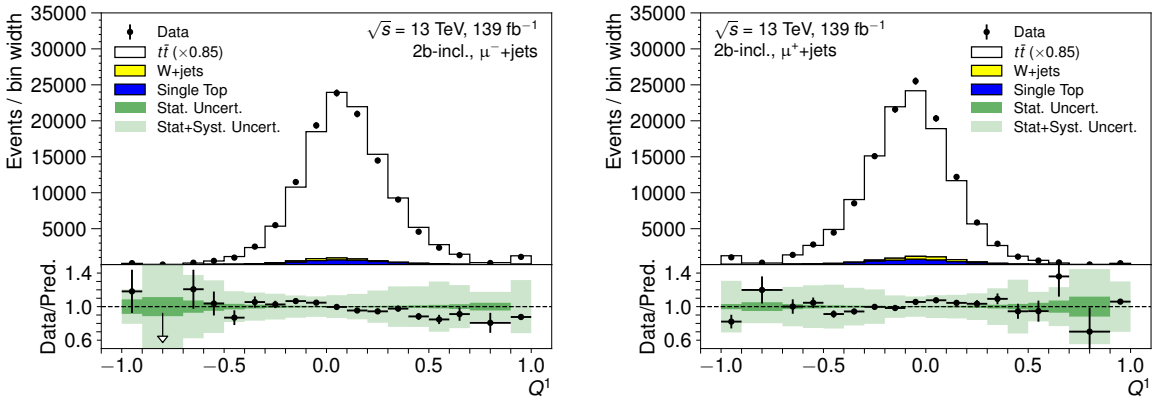
$t\bar{t}$ sample into categories based on whether the large-R jet is matched to a top quark, top anti-quark, or no match is successful. The expectation is that in the μ^- (μ^+) channel, the selected large-R jet should contain the decay products of a hadronically-decaying top quark (top anti-quark). The contribution of events where the matching fails is approximately 3 %, and the contribution of events where opposite matching is found in contrast to the expectation is less than 0.02 %. Therefore, the charge of the isolated muon is a good quantity for the selection of a pure sample of large-R jets initiated by the corresponding top (anti-)quark.

In general, the agreement between the data and the prediction for some of the charge observables and the NN discriminants shows slight tension with respect to the systematic uncertainties. Most of the systematic uncertainties are correlated across the bins of distributions, introducing a normalisation effect, that does not cover the slopes in the data-to-prediction ratios. The shape in the uncertainty bands observed is dominated by the uncertainty on the $t\bar{t}$ modelling of PS and hadronisation. The tension in the data-to-prediction agreement is particularly evident for the charge observables calculated from all tracks of the track jets, as shown in Fig. 8.9 and 8.10. The MC-predicted charges of the JetFitter and the SV vertices are found to agree well with data, as shown for example in Fig. 8.11, with the rest of the observables shown in App. D.2. Similarly, the MC simulations predict smaller track-jet track multiplicities than data, whereas the track multiplicity for vertices is found to be modelled significantly better, as shown for example in Fig. 8.12. The jet charge observables in MC appear to be slightly more offset from zero than in data, and this offset results in a slightly higher separation power of these variables in MC simulation compared to data. The same trend is observed for the NN discriminant in Fig. 8.13 and 8.14.



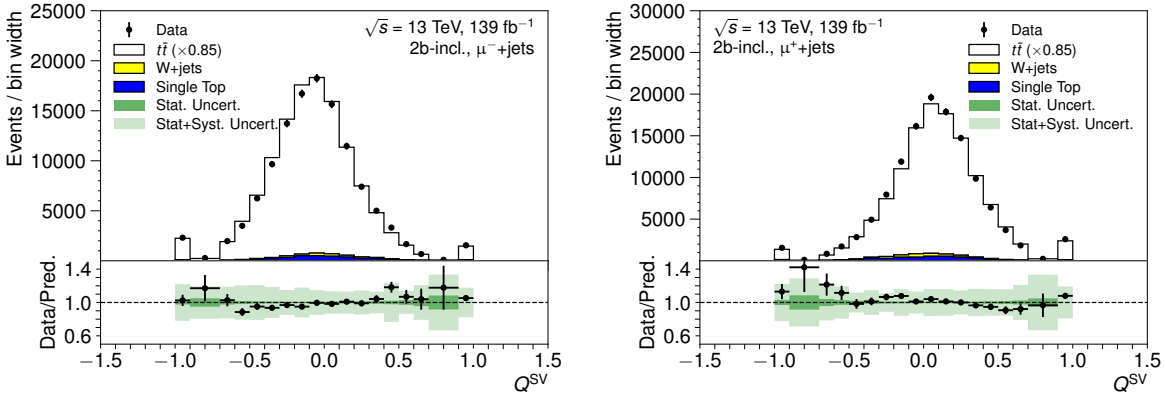
(a) (b)

Fig. 8.9: Comparison of data and prediction for the charge of the b-tagged track jet Q^b, in the μ^- region (a) and μ^+ region (b). The dark green band shows the statistical uncertainty of the total prediction, and the light green band shows the total prediction statistical and systematic uncertainty summed in quadrature. The error bars on black points show the statistical uncertainty of data. The $t\bar{t}$ prediction is scaled to normalise the total prediction yield to data. The bin yields are divided by the bin width.



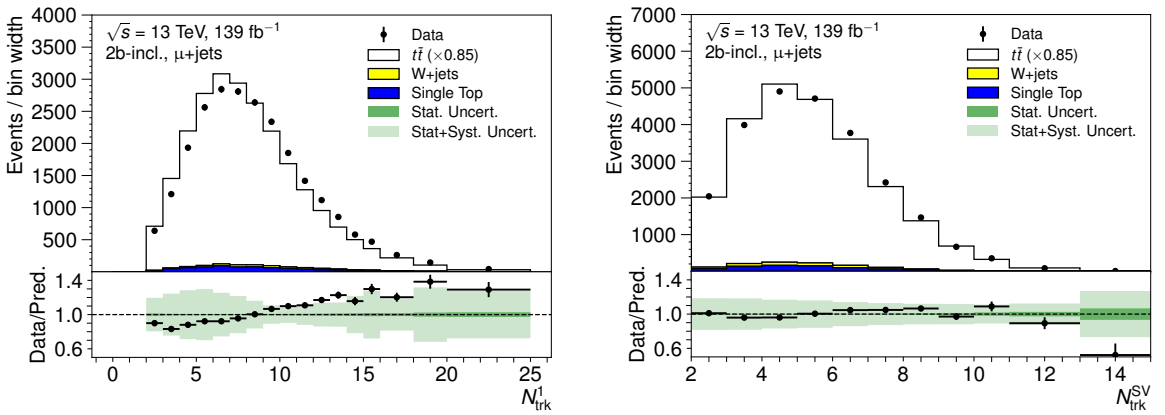
(a) (b)

Fig. 8.10: Comparison of data and prediction for the charge of the leading-p_T non-b-tagged track jet Q^1, in the μ^- region (a) and μ^+ region (b). The plot style follows the convention of Fig. 8.9. The bin yields are divided by the bin width.



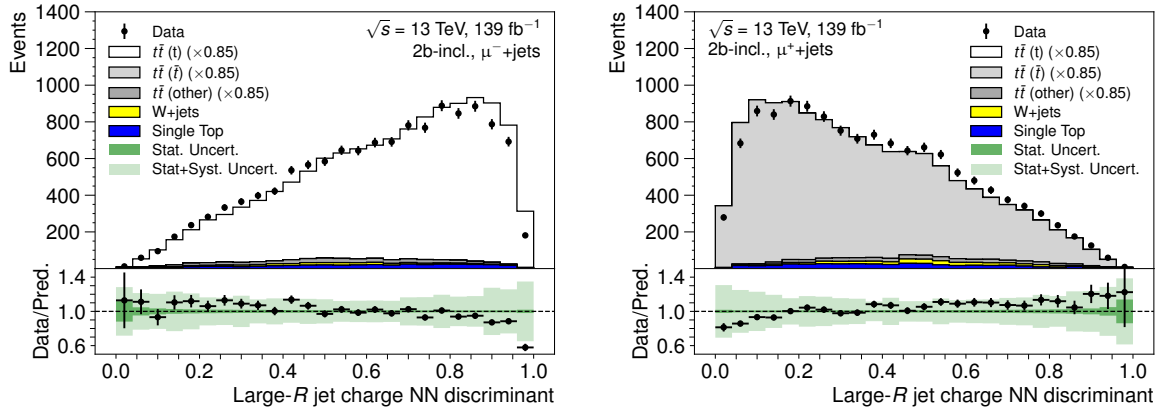
(a) (b)

Fig. 8.11: Comparison of data and prediction for the charge of the SV vertex of the b-tagged track jet Q^{SV}, in the μ^- region (a) and μ^+ region (b). The plot style follows the convention of Fig. 8.9. The bin yields are divided by the bin width.



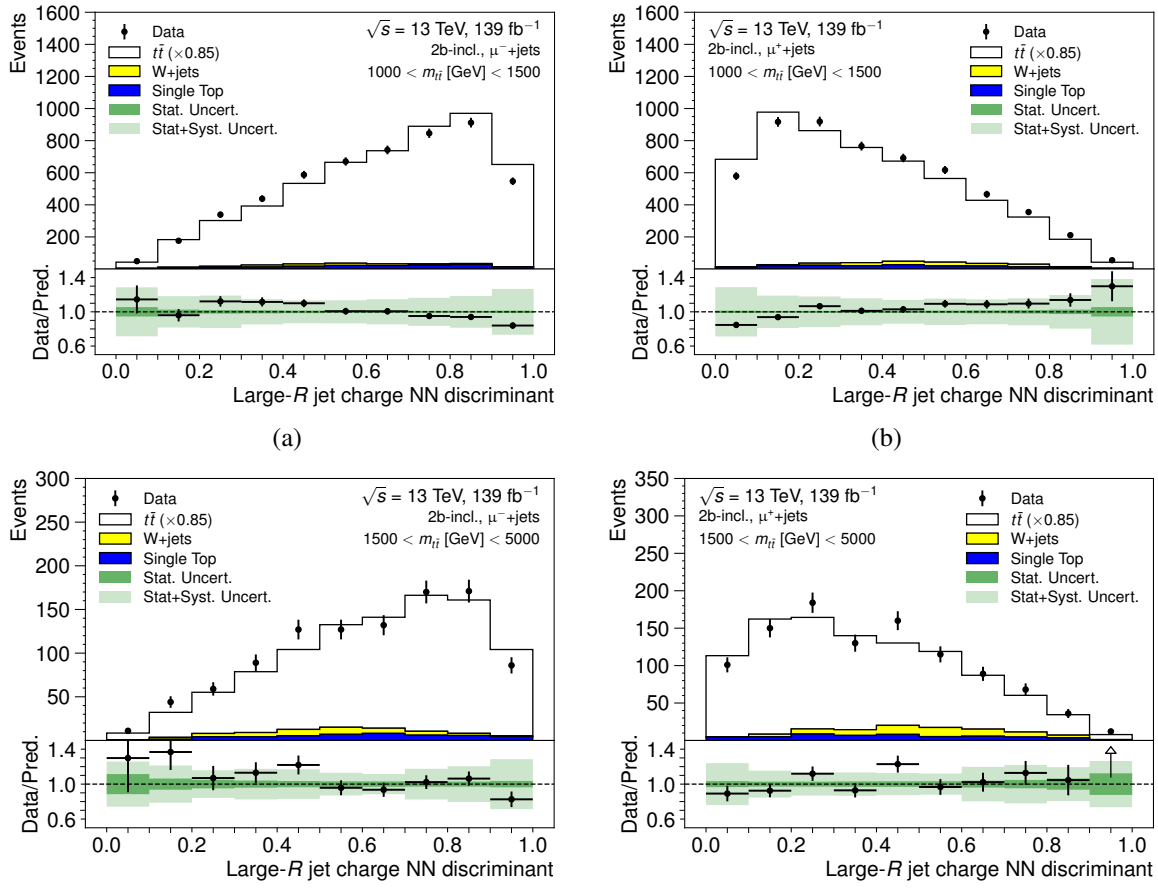
(a) (b)

Fig. 8.12: Comparison of data and prediction for the rate of the b-tagged track jet containing a SV vertex d^{SV} (a) and the SV vertex track multiplicity N^{SV}_{trk} (b). The plot style follows the convention of Fig. 8.9. The bin yields in (a) are divided by the bin width.



(a) (b)

Fig. 8.13: Comparison of data and prediction for the NN discriminant of the hadronic top-quark large-R jet candidate, in the μ^- region (a) and μ^+ region (b). The $t\bar{t}$ contribution is split according to the true charge of the top quark matched to the large-R jet, into t, $\bar{t}$ and other (failed matching) category. The plot style follows the convention of Fig. 8.9.



(a) (b)

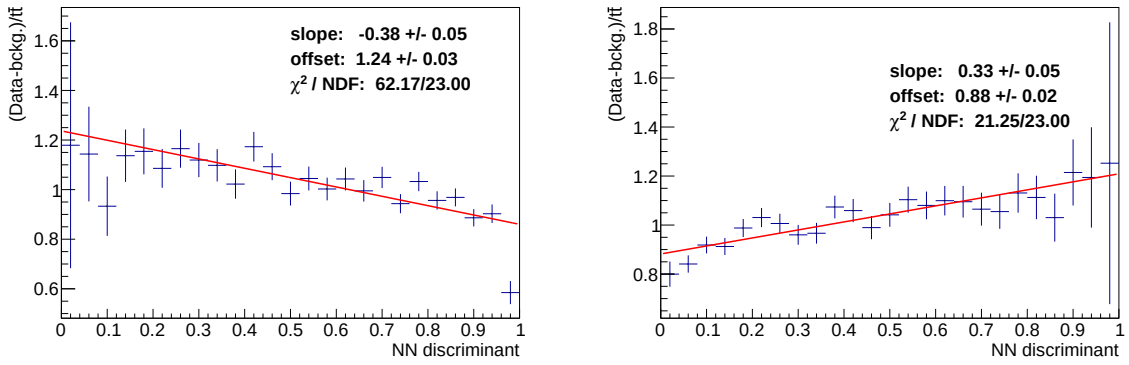
(c) (d)

Fig. 8.14: Comparison of data and prediction for the NN discriminant of the hadronic top-quark large-R jet candidate, in the μ^- region (left) and μ^+ region (right), for two intervals of $m_{t\bar{t}}$ (top/bottom). The plot style follows the convention of Fig. 8.9.

8.4.3 Uncertainty due to the NN mismodelling

Based on the NN discriminant data-to-prediction comparisons, the paramount question of interest is, how does the mismodelling impact the $\Delta|y|$ observable to be used for the A_C measurement in the all-hadronic channel. The sign of the $\Delta|y|$ observable depends on the comparison of the NN discriminant values of the two large-R jet candidates, as shown in Eq. 8.14. The observations of the mismodelling suggest that the rate of correctly assigned $\Delta|y|$ sign is potentially lower in data than in $t\bar{t}$ MC simulation. The uncertainty on the modelling of $t\bar{t}$ PS and hadronisation in this study is also propagated in the all-hadronic A_C estimate, thus partial covering of the shape effects is expected from this uncertainty.

An additional systematic uncertainty due to the NN discriminant mismodelling is estimated in the single-lepton channel by reweighting of the $t\bar{t}$ distribution of the NN discriminant in Fig. 8.13 to match data, separately for top-quark jets and for top-anti-quark jets. In the μ^- (μ^+) channel, the non-$t\bar{t}$ background as well as the $t\bar{t}$ contribution of large-R jet candidates not matched or incorrectly matched[(8)] are subtracted from data. The $t\bar{t}$ contribution is normalised to the background-subtracted data, same as in the data-to-prediction comparison figures. Subsequently, the ratio of distribution of the NN discriminant for large-R jets matched to top (anti-)quark in $t\bar{t}$ to the background-subtracted distribution is obtained in μ^- (μ^+) regions. The ratios in the respective regions are fitted with a linear function fit. Each of the fit functions corresponds to one of the two classes of large-R jets. The fit is shown in Fig. 8.15. It is found that the high χ^2 per degrees of freedom for the μ^- region fit is caused by the last outlier bin in the ratio. Removing the bin from the fit, reduces the χ^2/NDF to 19.3/22, yielding a slope of -0.28 ± 0.05 and offset of 1.19 ± 0.03. For the uncertainty estimate, the fit to full distribution is taken, based on the fact that the obtained slope is larger, thus leading to a more conservative uncertainty.



(a) (b)

Fig. 8.15: The linear fit of ratio of background-subtracted data distribution to distribution of signal jets from $t\bar{t}$ sample, of the large-R jet candidate NN discriminant. In (a) the fit is performed in μ^- region, where the signal jets originate from top quarks. In (b) the fit is performed in μ^+ channel, where the signal jets originate from top anti-quarks.

Finally, the two obtained linear functions for top-quark jets and top-anti-quark jets, respectively, are used to perform event-by-event reweighting in the all-hadronic channel. The event weight is a product of two weights corresponding to the two candidate large-R jets, where each weight is calculated

[(8)] Cases when a top-quark is matched to the large-R jet candidate in the μ^+ channel, or the top anti-quark is matched to the large-R jet candidate in the μ^- channel, are considered as incorrect matches.

based on the value of the NN discriminant, using the fit function corresponding to the top (anti-)quark matched to the large-R jet candidate. For large-R jet candidates failing truth matching, no weight is applied. The described reweighting is used to derive a systematically-shifted $\Delta|y|$ distribution and calculate the resulting systematic uncertainty on A_C in the all-hadronic channel in Sec. 8.6. The effect of the systematic variation on the NN discriminant is illustrated in Fig. 8.16.

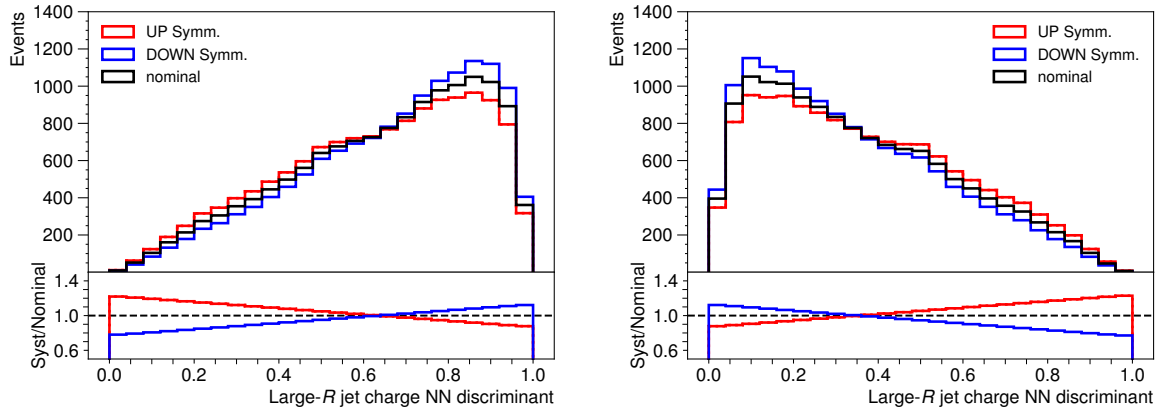


Fig. 8.16: Impact of the NN reweighting uncertainty on the NN discriminant in the single-lepton channel $t\bar{t}$ MC prediction, for candidate large-R jets matched to top quark in μ^- region (a) and top anti-quark in μ^+ region (b).

8.5 Expected sensitivity of the all-hadronic $t\bar{t}$ charge asymmetry

In this section, we obtain an approximate estimate of sensitivity of the boosted all-hadronic channel to the A_C. Due to the kinematic selection in this analysis, the A_C sensitivity is estimated and compared with the single-lepton channel for $m_{t\bar{t}} \in [1000, 1500]$ GeV and $m_{t\bar{t}} > 1500$ GeV bins only. All of the signal and background predictions are obtained from MC simulations with the exception of the multijet background, which is estimated using a data-driven ABCD method similar to the method in Ref. [209], described in Sec. 8.5.1.

The A_C vs $m_{t\bar{t}}$ estimate is obtained by unfolding the Asimov dataset, using the FBU method previously described in Sec. 7.4. An optimisation of the $\Delta|y|$ binning is performed in Sec. 8.5.3. In Sec. 8.5.4, the systematic uncertainties impacting the A_C estimate are discussed. Finally, in Sec. 8.6, the results of the A_C vs $m_{t\bar{t}}$ sensitivity estimate are presented.

8.5.1 Multijet background estimate

The motivation for using a data-driven technique for the multijet background estimate is that this background is notoriously difficult to model accurately by MC simulations, and also expensive in terms of the time required to simulate sufficient amount of events due to the very high background rejection imposed by the top-quark tagging and b-tagging selection criteria.

The ABCD method is based on factorising the event selection into regions typically based on two-dimensional selection and extrapolating from a background enriched region into signal region, hence the origin of the name ABCD, where the letters denote the individual regions. Let us assume

for simplicity a four-region scenario, defined by two different selections on variable X and variable Y respectively. Let us denote that region D is the signal region where high signal purity is achieved by cuts on X and Y. Control regions C and B are defined by imposing a looser cut on one of the variables, and region A by imposing a looser cut on both of the variables. If the variables X and Y are uncorrelated, the shape of observable X is independent of selection on observable Y and vice-versa. It follows that the signal region D yield can be scaled using yields in the control region A and regions B and C via relation:

$$D = \frac{B \cdot C}{A}. \tag{8.15}$$

In practice, contamination by other backgrounds and signal is expected in the regions A, B and C, therefore these are subtracted, typically using MC based estimates. The ABCD approach can be extended to binned distributions, where Eq. 8.15 is used in each bin of the distribution of interest. The assumption that the variables X and Y are uncorrelated are often difficult to satisfy, finding uncorrelated observables is not guaranteed, and therefore corrections to account for the correlations may be necessary.

The ABCD multijet estimate in this analysis uses 16 regions in total, defined by leading and sub-leading large-R jet top-quark tagging and b-tagging selection (tagged/un-tagged), as defined in Sec. 8.2. The region definitions are illustrated in Table 8.12, along with the expected signal contamination. The control region A is defined by having neither the leading nor the sub-leading large-R jet top- or b-tagged, and contains only 0.1 % of signal events. In all of the regions, the signal and non-multijet background contributions are subtracted from data. The extrapolation to signal region S is performed using regions O and J via $O \cdot J/A$. As discussed in [209], the rate of top-tagging and b-tagging for the leading large-R jet is not uncorrelated to the sub-leading large-R jet tag rates. The correlation corrections are extracted from ratios of additional regions, for example ratio (F/E) gives the leading jet top-quark tag rate when the sub-leading jet is also top-tagged, and ratio (C/A) gives the leading jet top-tag rate when the sub-leading jet is not top-tagged. If the ratio of these two rates is different from unity, this indicates that the rates are correlated. Four correction factors are derived to account for these correlations:

- The dependence of the leading jet top-tag rate on the sub-leading jet top-tag rate:

$$\frac{F/E}{C/A} = \frac{F \cdot A}{E \cdot C}, \tag{8.16}$$

- the dependence of the leading jet top-tag rate on the sub-leading jet b-tag rate

$$\frac{D/B}{C/A} = \frac{D \cdot A}{B \cdot C}. \tag{8.17}$$

- the dependence of the leading jet b-tag rate on sub-leading jet b-tag rate

$$\frac{H/B}{C/A} = \frac{H \cdot A}{B \cdot C}. \tag{8.18}$$

- the dependence of the leading jet b-tag rate on sub-leading jet top-tag rate

$$\frac{H/B}{I/A} = \frac{H \cdot A}{B \cdot I}. \tag{8.19}$$

Combining the bin yields of an observable from regions A, J and O and the correction factors in Eq. 8.16- 8.19, the multijet distribution of the observable in signal region S are obtained from bin yields of other regions as:

$$S = \frac{J \cdot O}{A} \cdot \frac{D \cdot A}{B \cdot C} \cdot \frac{G \cdot A}{E \cdot I} \cdot \frac{F \cdot A}{E \cdot C} \cdot \frac{H \cdot A}{B \cdot I}. \tag{8.20}$$

Table 8.12: Definition of the regions used for data-driven ABCD multijet estimate, along with the expected signal contribution, defined as the ratio of the all-hadronic $t\bar{t}$ yield and the data yield. The regions are parametrised in terms of top-tagging and b-tagging decisions of leading and sub-leading large-R jet in p_{T}. The notation 0t0b denotes whether the respective jet is top-tagged and/or b-tagged, where 0 means un-tagged and 1 means tagged.

Sub-leading large-R jet \ Leading large-R jet	0t0b	1t0b	0t1b	1t1b
1t1b	J (6.4%)	K (33.3%)	L (39.2%)	**S**
0t1b	B (0.9%)	D (4.3%)	H (5.4%)	N (48.2%)
1t0b	E (0.6%)	F (2.8%)	G (3.8%)	M (40.7%)
0t0b	A (0.1%)	C (0.6%)	I (0.8%)	O (8.0%)

Regions K, L, M, N all contain various mixtures of signal and background composition, and hence can be used as validation regions. The region K and N are used for a data-to-prediction comparisons with a very different mixture of signal and background contributions to validate the multijet background estimate. The multijet contribution in these regions is estimated using the following set of regions:

$$K = \frac{D \cdot F}{C}, \quad N = \frac{H \cdot D}{B}. \tag{8.21}$$

8.5.2 Data-to-prediction comparisons in the validation and signal regions

The data-to-prediction comparisons for the validation regions K and N and the signal region, for the observables related to the two leading large-R jets, the charge NN discriminant and the $\Delta|y|$ vs $m_{t\bar{t}}$ double-differential distribution are shown in Fig. 8.18– 8.21. Further signal and validation region plots are shown in App. D.3. The systematic uncertainties included in the distributions are described in Sec. 8.5.4.

In general, the agreement between the distribution shape of the data and the prediction is reasonable for most of the observables, mostly within the statistical and systematic uncertainties, with small tension in shape of the validation-region $\Delta|y|$ vs $m_{t\bar{t}}$ observable and the NN discriminants. Some tension is observed in the modelling of the large-R jet p_{T}, both in signal and validation regions, showing a slope at the low p_{T} region. In general the predicted contribution of $t\bar{t}$ is over-estimated as observed in the other data-to-prediction comparisons in the measurements presented in this thesis. It is also observed, that the signal normalisation factor differs slightly between the individual regions. One of the possible contributions to these differences could be the missing top-quark tagging calibration, in particular for the differences between the region K and N yields, since these regions differ in the multiplicity of the top-tagged large-R jets. Nevertheless, this phenomenon is still under investigation. In the unfolding,

the signal normalisation is a free parameter not impacting the A_C value, since A_C is insensitive to the signal normalisation by definition. However the normalisation of the multijet background does impact the A_C result. Further studies of the multijet background are indeed necessary for a full measurement. In the unfolding, uncertainties on both the shape as well as the normalisation of the multijet background are added, as discussed in Sec. 8.5.4.

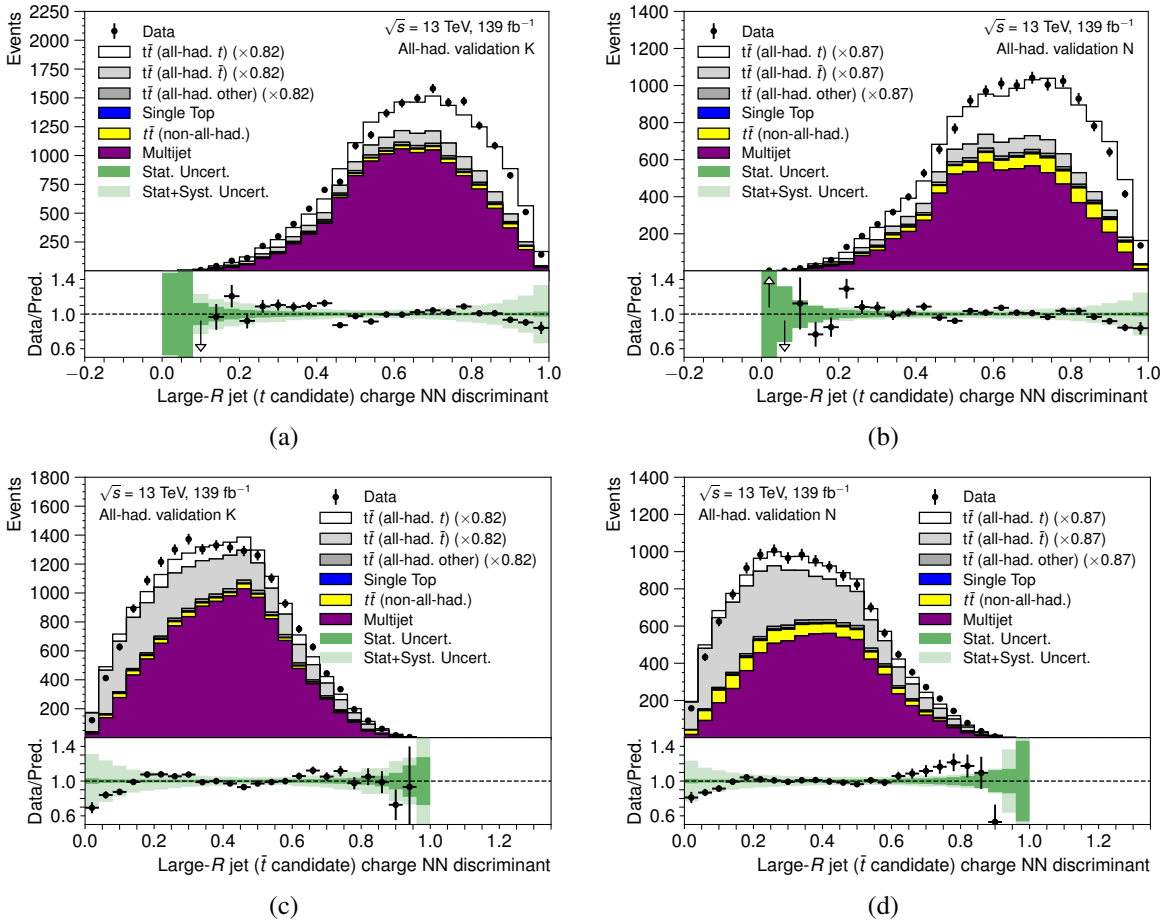
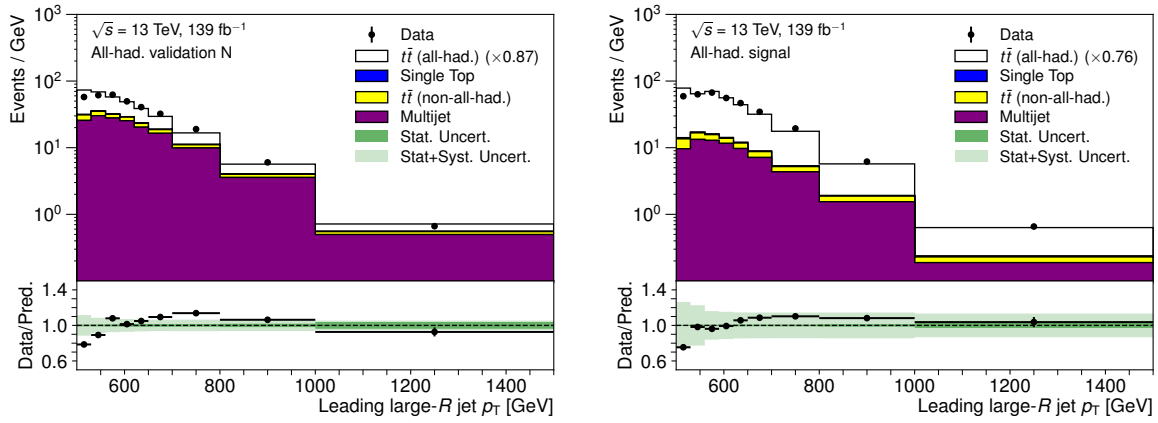
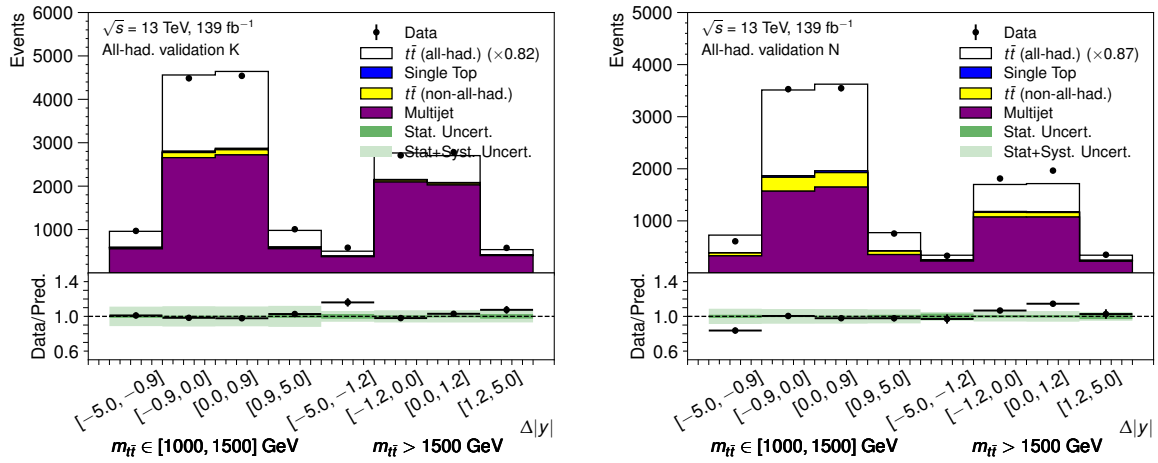


Fig. 8.17: The comparison of data to prediction of the distribution of top-quark candidate (top row) and top-anti-quark (bottom row) large-R jet NN discriminant, in the validation regions K (left) and N (right). In the figures, the all-hadronic $t\bar{t}$ contribution is scaled to match the prediction yield to data. The dark-green band in the ratio plot shows the statistical uncertainty on the prediction, while the light-green band shows the statistical and systematic uncertainty on the prediction, summed in quadrature. The error bars on the black points show the statistical uncertainty of data. Additionally, the all-hadronic $t\bar{t}$ contribution is split according to the true charge of top quark matched to the large-R jet, into t, $\bar{t}$ and other (failed matching) category.



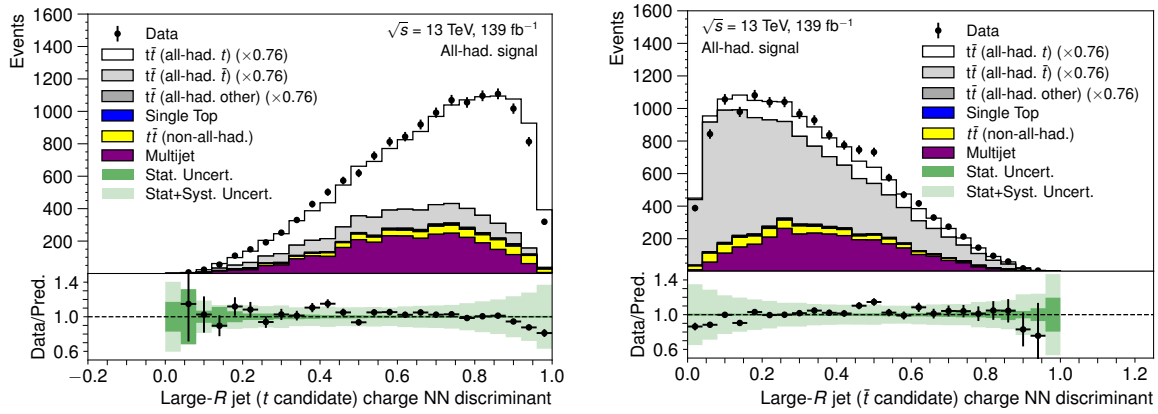
(a) (b)

Fig. 8.18: The comparison of data to prediction of the leading large-R jet p_{T} in the validation region N (a) and in the signal region (b). The uncertainties and the signal normalisation follows the same convention as Fig. 8.17. The bin yields are divided by the bin width.



(a) (b)

Fig. 8.19: The comparison of data to prediction of the $\Delta|y|$ vs $m_{t\bar{t}}$ distribution, in the validation regions K (a) and N (b). The uncertainties and the signal normalisation follows the same convention as Fig. 8.17.



(a) (b)

Fig. 8.20: The comparison of data to prediction of the distribution of top-quark candidate large-R jet NN discriminant, in the validation regions K (a) and N (b). The uncertainties and the signal normalisation follows the same convention as Fig. 8.17.

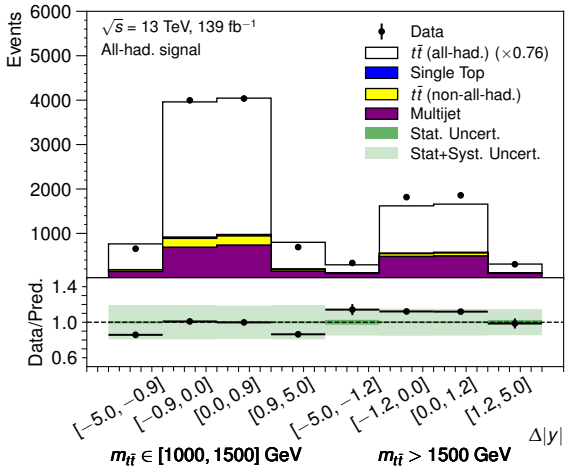


Fig. 8.21: The comparison of data to prediction of the $\Delta|y|$ vs $m_{t\bar{t}}$ distribution in the signal region. The uncertainties and the signal normalisation follows the same convention as Fig. 8.17.

8.5.3 Binning optimisation and unfolding

The unfolding of A_C vs $m_{t\bar{t}}$ requires optimisation of the $\Delta|y|$ binning, which is performed in a similar manner as the optimisation described in Sec. 7.5, using linearity test with injected asymmetry predicted by light-mass axigluon models. Four $\Delta|y|$ bins are used in a similar manner as in the single-lepton channel, to map the migrations in a sufficiently-accurate manner, and the same $m_{t\bar{t}}$ bins are used as the two highest-$m_{t\bar{t}}$ bins used in the single-lepton channel. The $\Delta|y|$ binning is obtained by performing the linearity test for various values of the inner $\Delta|y|$ bin edge, x, in the $[-5, -x, 0, x, 5]$ binning. The $\Delta|y|$ inner bin edge is scanned in steps of 0.1 for both $m_{t\bar{t}}$ bins independently, performing a two-dimensional scan. The figure of merit for the choice of $\Delta|y|$ binning is the R-factor:

$$R = \sum_{i}^{m_{t\bar{t}}\text{ bins}} \left[(1 - s_i)^2 + o_i^2\right], \tag{8.22}$$

where s_i and o_i are the slope and offset from the linearity tests, for the i^{th} unfolded $m_{t\bar{t}}$ bin. The metric is designed to penalise binning choices that yield non-unity slopes and non-zero offsets.

It is observed, that no choice of binning can achieve a slope compatible with one. The problem arises from the extrapolation to full phase space in $\Delta|y|$ at parton level. The event selection criteria limit the large-R jets to satisfy $|\eta| < 2.0$. This effectively means, that approximately beyond $|\Delta|y|| > 2.0$, no events are reconstructed. When injecting axigluon-induced BSM asymmetries, the charge asymmetry contribution is more enhanced in the tails of the $\Delta|y|$ distribution, reaching beyond the detector acceptance. At detector level, the enhancement of asymmetry is therefore smaller than at parton level due to the insensitivity to $|\Delta|y|| > 2.0$, resulting in an underestimation of the magnitude of the unfolded asymmetry.

A possible solution to the linearity non-closure is to impose fiducial cuts at the parton level on $\Delta|y|$ observable. The linearity test is repeated with $|\Delta|y|| < 2.0$ fiducial cut, yielding linearity with slope compatible with one. The best-linearity binning for fiducial and for full-phase-space unfolding is shown in Table 8.13. Restricting the parton-level $\Delta|y|$ using fiducial cuts yields an out-of-fiducial $t\bar{t}$ background, constituted by signal events which do not pass parton-level fiducial cuts, but migrate

into the signal region at detector level. It is found however, that the contribution of this background is approximately 0.03 % and is therefore neglected.

The disadvantage of the fiducial unfolding is that it introduces differences in the $\Delta|y|$ definition with respect to the single-lepton channel A_C measurement, complicating a potential combination. If a combination of multiple channels directly at the level of unfolding is desired, similarly to how resolved and boosted regions are combined in the single-lepton channel A_C measurement, it is necessary to ensure that the parton-level $\Delta|y|$ binning is the same. Both full phase space and fiducial unfolding is studied further in this analysis to asses the impact of the extrapolation to full phase space on the uncertainty of the unfolded A_C and also on the magnitude of the extracted asymmetry.

In Fig. 8.22, the migration matrix for the $\Delta|y|$ vs $m_{t\bar{t}}$ double-differential distribution is shown. In general, the migrations between the differential bins reach up to 15 % in the highest-$m_{t\bar{t}}$ bin. This is a significantly lower value than in case of the single-lepton channel, where the migrations between the two highest-$m_{t\bar{t}}$ bins reach up to 25 % in resolved and 30 % in boosted regions, as can be seen in $m_{t\bar{t}}$ vs $\Delta|y|$ migration matrices in App. C.1. The migrations between inner and outer $\Delta|y|$ bin for the same $\Delta|y|$ sign are also smaller in the all-hadronic channel than in the single-lepton channel. The resolved single-lepton channels suffer from migrations due to mis-measurement of neutrino momentum as well as wrong jet-to-parton assignment. In boosted single-lepton channel the latter challenge is circumvented by reconstructing the hadronic top quark as a single large-R jet, but the neutrino reconstruction remains an under-constrained problem.

Finally, the implicit $m_{t\bar{t}} > 1000\,\text{GeV}$ restriction in the parton-level binning in the all-hadronic channel leads to a non-fiducial contribution of events below this $m_{t\bar{t}}$ threshold, which are reconstructed with $m_{t\bar{t}} > 1000\,\text{GeV}$, passing the signal region selection. The contribution of this background is 6.5 % in the $m_{t\bar{t}} \in [1000, 1500]$ GeV bin, and is negligible in the $m_{t\bar{t}} > 1500\,\text{GeV}$ bin.

Table 8.13: The $\Delta|y|$ binning for inclusive and differential A_C unfolding. Both the slope and the offset from the linearity test are shown, along with their uncertainties from the fit.

Differential bin	$\Delta\|y\|$ binning	Linearity	
		slope	offset
Full phase space unfolding			
$m_{t\bar{t}} \in [1000, 1500]$ GeV	[-5,**-0.9**, 0, **0.9**, 5]	0.880 ± 0.010	-0.0000 ± 0.0003
$m_{t\bar{t}} > 1500\,\text{GeV}$	[-5,**-1.1**, 0, **1.1**, 5]	0.914 ± 0.014	-0.0007 ± 0.0005
Fiducial unfolding ($\|\Delta\|y\|\| < 2.0$)			
$m_{t\bar{t}} \in [1000, 1500]$ GeV	[-5,**-0.9**, 0, **0.9**, 5]	0.993 ± 0.011	-0.0006 ± 0.0003
$m_{t\bar{t}} > 1500\,\text{GeV}$	[-5,**-1.2**, 0, **1.2**, 5]	0.982 ± 0.015	-0.0004 ± 0.0005

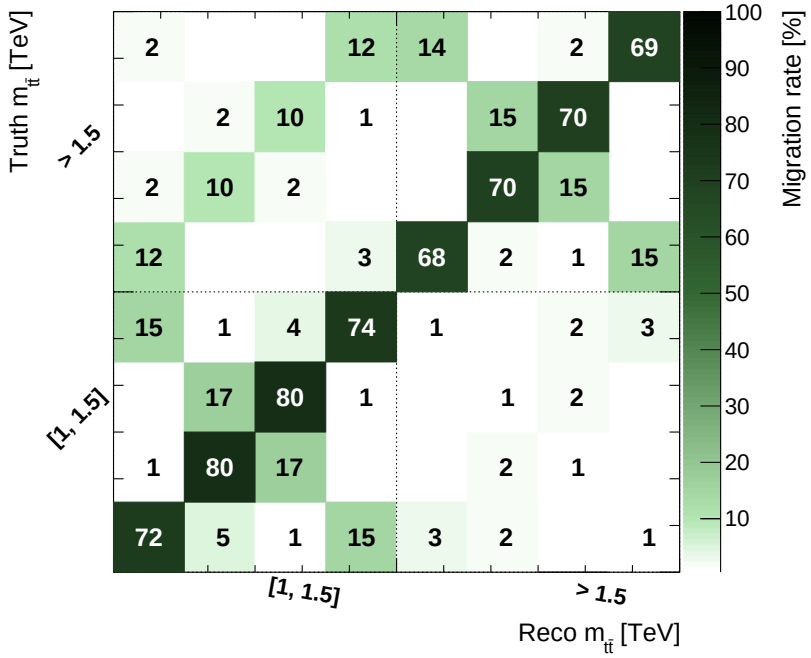


Fig. 8.22: Migration matrix of the $\Delta|y|$ vs $m_{t\bar{t}}$ double-differential distribution in the boosted all-hadronic channel, using the binning for full-phase-space unfolding. The groups of four bins correspond to the four $\Delta|y|$ bins for a given $m_{t\bar{t}}$ bin specified by the axis labels. The precise $\Delta|y|$ binning is listed in Table 8.13.

8.5.4 Systematic uncertainties

The systematic uncertainties evaluated for the A_C Asimov estimate are the same as those in Sec. 8.4.1, with an addition of uncertainties on the multijet background estimate, the uncertainty on the NN discriminant modelling described in Sec. 8.4.3, and the uncertainties related to unfolding.

The uncertainty on the NN discriminant, derived in Sec. 8.4.3, is propagated to the all-hadronic $t\bar{t}$ events, by applying weights to the two hadronic top-quark large-R jet candidates, where depending on whether the jet is matched to a parton-level top quark or top anti-quark, the weight is calculated from the corresponding fitted dependence of the weight on NN discriminant previously shown in Fig. 8.15.

The systematic uncertainties on the MC modelling of the non-all-hadronic $t\bar{t}$ and single top-quark backgrounds are neglected. The normalisation uncertainties previously quoted in Table 6.5 are considered.

A number of assumptions are made in the multijet background estimate. The regions used for the estimate, including the regions used to correct for correlations between large-R jet top-quark tagging and b-tagging, include a small contamination by non-multijet processes. No systematic uncertainties are propagated on this contamination except for the statistical uncertainties of both data and the contributions of the non-multijet processes. Strictly speaking, the systematic uncertainties on the non-multijet MC-simulated contributions in the individual regions should be propagated. For this preliminary estimate, omitting this systematic uncertainties propagation is motivated by the fact that the non-multijet contamination in the individual regions is only up to 10 %. Because of the different shape of the multijet background $\Delta|y|$ distribution compared to the signal $\Delta|y|$ distribution, both different normalisation as well as shape effects in the modelling of the multijet background can alter the unfolded A_C. As such, a 50 % normalisation uncertainty on the multijet background is added

in the unfolding, not included in the data-to-prediction comparison plots. The choice of the value is admittedly ad-hoc, its purpose is mainly to check in the ranking of the systematic uncertainties, whether this uncertainty yields an important effect on A_C.

Finally, to induce shape uncertainties on the total background for the $\Delta|y|$ distribution, an approach inspired by the $\boldsymbol{\gamma}$ NPs defined in Sec. 7.4.2 is used. In this approach, the individual bins of the total backgrounds are allowed to fluctuate independently based on the per-bin statistical uncertainty of the background prediction. In other words, for each bin of the total background $\Delta|y|$ distribution, a separate systematic uncertainty is added. The magnitude of this uncertainty for each bin is given by a sum-in-quadrature of the MC statistical uncertainty of the single top-quark and non-all-hadronic $t\bar{t}$ contributions and the statistical uncertainty of the multijet background in the given bin. The multijet background statistical uncertainty is obtained by propagating the statistical uncertainties of data and the subtracted non-multijet contributions in the individual ABCD regions in the Eq. 8.20. The per-bin yield variations range from 2 % to 6 %.

The following uncertainties on the unfolding are considered, previously described in Sec. 7.6.3. The uncertainty due to the limited signal MC sample is propagated via PEs by smearing the response matrix. The linearity non-closure in the binning optimisation is propagated into a bias uncertainty according to Eq. 7.17.

In contrast to the single-lepton channel A_C measurement, the bootstrapping of systematics is omitted in this estimate. The justification for omitting this procedure is that the statistical uncertainties on the MC predictions in the all-hadronic channel are much smaller than those in the single-lepton channel due to the larger number of generated MC events relative to the total predicted yield. Specifically, the signal $t\bar{t}$ sample as well as the alternative samples for modelling uncertainties are generated with approximately 30-times more events than the expected number of signal events in the data. This is an order of magnitude more than the number of generated events in the $t\bar{t}$ sample in the single-lepton channel, where bootstrapping was applied. The single top-quark and the non-all-hadronic $t\bar{t}$ backgrounds have seven- and four-times more generated events compared to expected yield in data, respectively, and only constitute approximately 4.5 % of the total expected yield.

Finally, the systematic uncertainties in this estimate are not marginalised within FBU, but instead their impact is obtained by performing a statistical-only unfolding the Asimov dataset shifted by the corresponding systematic uncertainty. The total uncertainty is the sum-in-quadrature of the individual systematic uncertainties and the statistical uncertainty of the unfolding of Asimov dataset. Marginalisation of systematic uncertainties is left for further studies, as to first order it is not expected that significant gains will be achieved, since the Asimov estimate is dominated by statistical uncertainty, as presented in the next section.

8.6 Results and discussion of further possible improvements

In Table 8.14, the unfolded Asimov dataset A_C vs $m_{t\bar{t}}$ in the all-hadronic channel is shown, for both the fiducial and the full-phase-space unfolding, along with a decomposition of the total uncertainty into statistical uncertainty and groups of systematic uncertainties. The A_C estimate is found to be statistically limited. The largest systematic uncertainties include $t\bar{t}$ modelling, followed by background modelling and the statistical uncertainty on the response matrix MC estimate due to limited number of

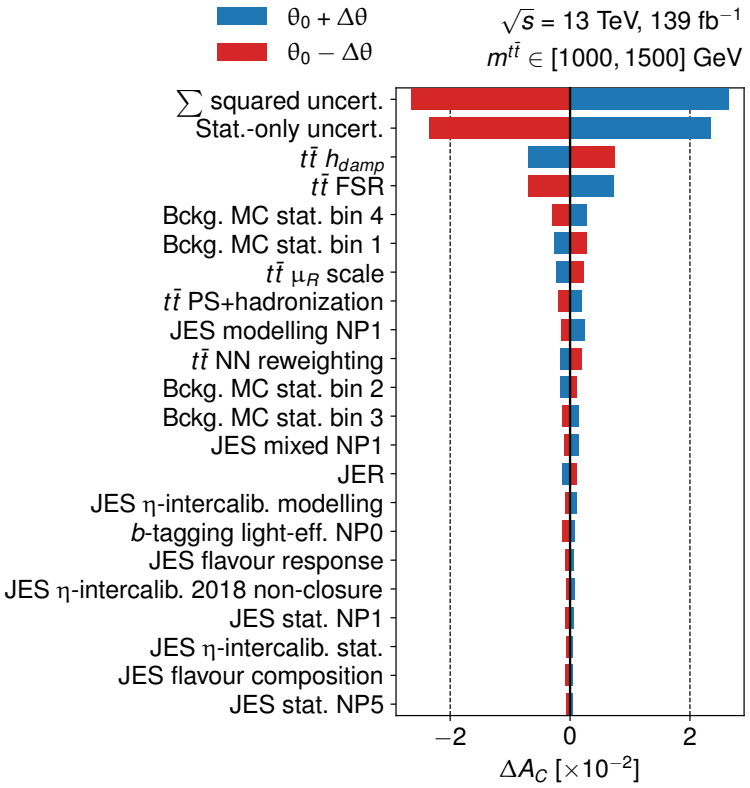
generated signal MC events. The background modelling uncertainty is dominated by uncertainty on the multijet normalisation and the per-bin statistical uncertainty of the total background contribution. The impact of top 20 systematic uncertainties is shown in Fig. 8.23 for the fiducial unfolding and in Fig. 8.24 for the full-phase-space unfolding. Among the dominant $t\bar{t}$ modelling uncertainties are the uncertainty on the FSR modelling and the PS + hadronisation modelling, similarly to the dominant systematic uncertainties in the single-lepton. The multijet normalisation and shape uncertainties (labelled as "Bckg. MC stat. bin") are among the highest-ranking systematic uncertainties, suggesting that further studies of this background are necessary for a proper measurement on data.

Table 8.14: Impact of statistical and systematic uncertainties on the Asimov dataset unfolding of A_C vs $m_{t\bar{t}}$. Both fiducial and full-phase-space unfolding in $\Delta|y|$ is shown. The total uncertainty is the sum-in-quadrature of statistical uncertainty and the systematic uncertainties. For the unfolding to full phase space, the unfolding bias is evaluated using the NNLO QCD + NLO EWK SM prediction [207, 208] and added to the total uncertainty instead of the Asimov-calculated bias.

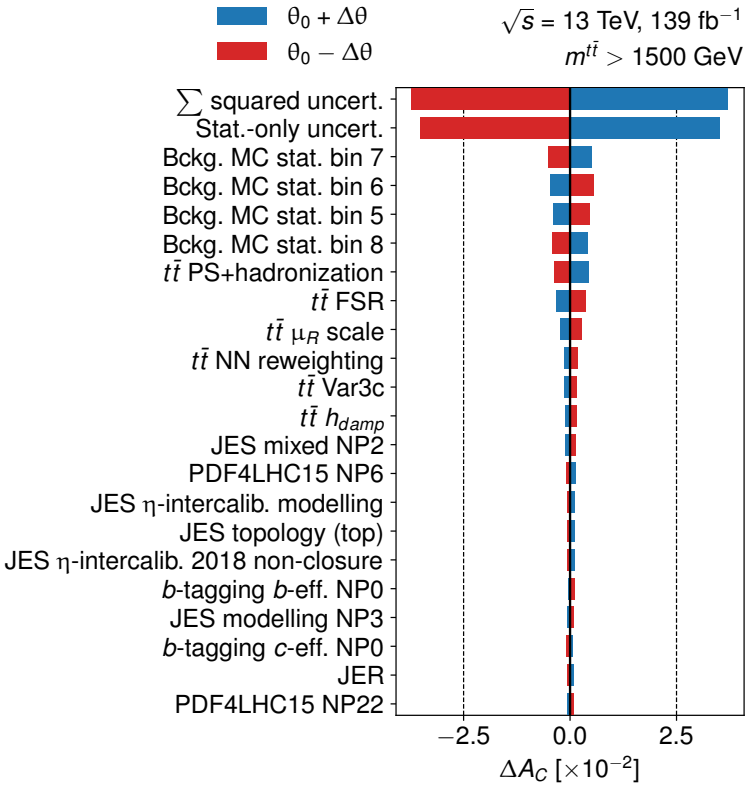
	Fiducial		Full phase space	
Bin ($m_{t\bar{t}}$ [GeV])	[1000, 1500]	> 1500	[1000, 1500]	> 1500
A_C value	0.0039	0.0039	0.0047	0.0045
Statistical uncertainty	0.0231	0.0351	0.0258	0.0366
$t\bar{t}$ modelling	0.0107	0.0063	0.0131	0.0077
Background modelling	0.0052	0.0097	0.0054	0.0103
Response matrix stat.	0.0043	0.0040	0.0049	0.0042
Large-R jet JES/JER, E_T^{miss}	0.0033	0.0026	0.0039	0.0033
b-tagging	0.0014	0.0016	0.0014	0.0017
$t\bar{t}$ PDF	0.0011	0.0020	0.0013	0.0019
Unfolding bias (Asimov)	0.0005	0.0005	0.0007	0.0012
Unfolding bias (SM)	–	–	0.0014	0.0017
Total uncertainty	0.0266	0.0374	0.0302	0.0393

Table 8.15: The comparison of the unfolded full-phase-space and true (parton-level) A_C and its uncertainty for the Asimov dataset, between the all-hadronic and the single-lepton channel. The parton-level true A_C uncertainty is statistical-only. The unfolded A_C uncertainty is the total statistical and systematic uncertainty. Additionally, the NNLO QCD + NLO EWK SM theory prediction [207, 208] is shown for reference of a higher-order prediction.

	A_C vs $m_{t\bar{t}}$[GeV]	
	[1000, 1500]	> 1500 GeV
All-hadronic unfolded	0.0047 ± 0.0302	0.0045 ± 0.0393
All-hadronic true	0.0047 ± 0.0008	0.0047 ± 0.0014
Single-lepton unfolded	0.0063 ± 0.0084	0.0114 ± 0.0297
Single-lepton true	0.0061 ± 0.0006	0.0100 ± 0.0013
NNLO QCD + NLO EWK SM	$0.0096^{+0.0009}_{-0.0009}$	$0.0094^{+0.0015}_{-0.0011}$

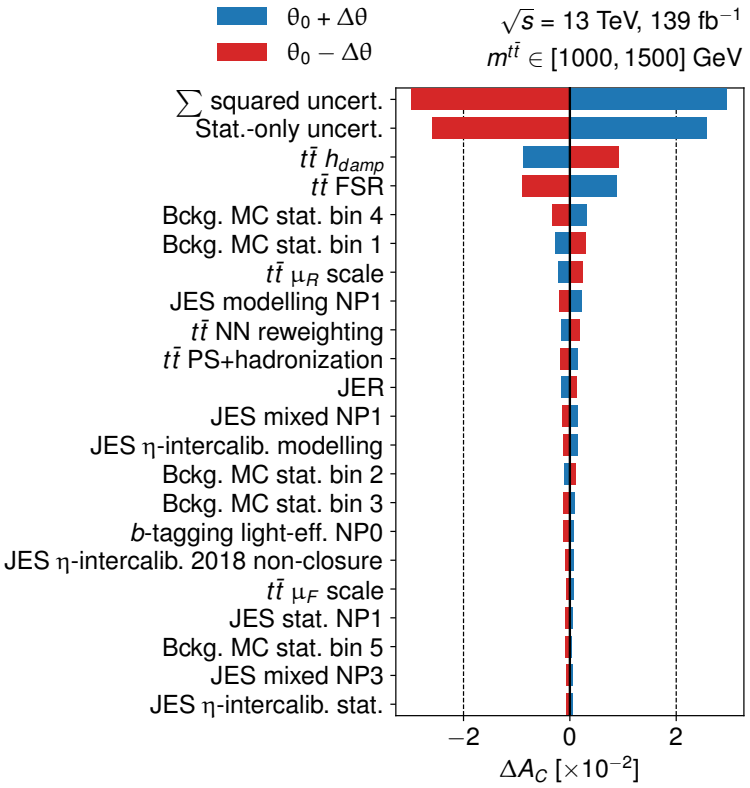


(a)

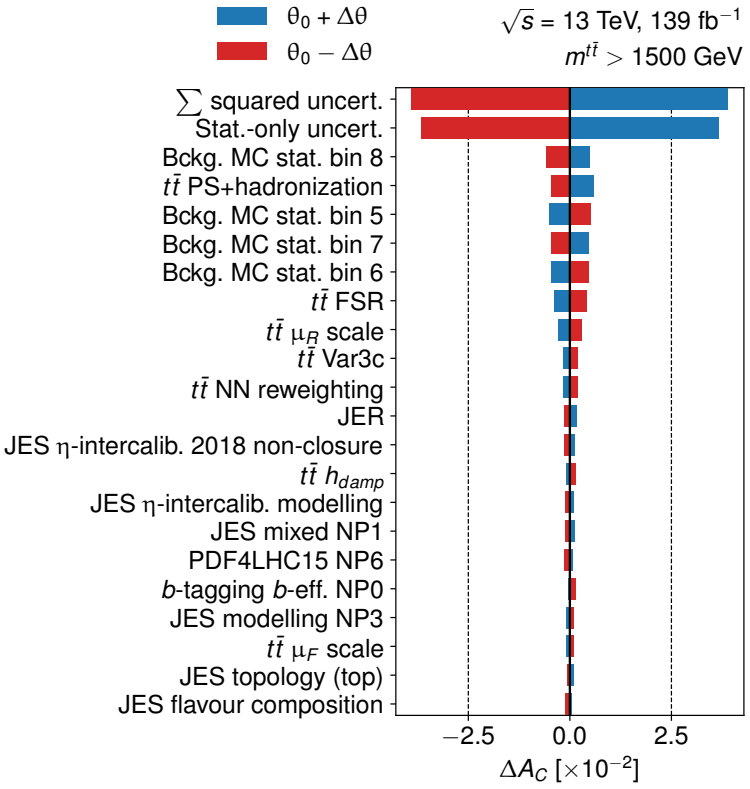


(b)

Fig. 8.23: Ranking of top 20 systematic uncertainties and the total (first row) and the statistical (second row) uncertainty of the A_C vs $m_{t\bar{t}}$ unfolding to fiducial phase space in $\Delta|y|$, for $m_{t\bar{t}} \in [1000, 1500]$ GeV (a) and $m_{t\bar{t}} > 1500$ GeV (b). The blue (red) bars show the impact of up-variation (down-variation) of a systematic uncertainty on the A_C.



(a)



(b)

Fig. 8.24: Ranking of top 20 systematic uncertainties and the total (first row) and the statistical (second row) uncertainty of the A_C vs $m_{t\bar{t}}$ unfolding to full phase space in $\Delta|y|$, for $m_{t\bar{t}} \in [1000, 1500]$ GeV (a) and $m_{t\bar{t}} > 1500$ GeV (b).

The uncertainty on the NN modelling is among the top 20 ranking uncertainties, however its contribution is small compared to other $t\bar{t}$ and multijet background uncertainties. The NN reweighting uncertainty should be interpreted as a test rather than a proper uncertainty derivation, but the result at least suggests, that the impact of the NN discriminant reweighting on the $\Delta|y|$ sign assignment is not limiting the measurement. Further refinement of the reweighting into a proper correction with associated uncertainties is possible. A full treatment would require deriving the correction, in a form of a fit or per-bin SFs based on the data-to-prediction disagreement, applying the correction, and additionally to propagate the uncertainties on the correction factors, e.g. all of the detector-related uncertainties and $t\bar{t}$ and background modelling uncertainties. Nevertheless, this results suggests that the machine-learning approach to top-quark charge identification is feasible for the measurement.

Finally, a comparison of the unfolded Asimov dataset A_C between the all-hadronic and the single-lepton channel is shown in Table 8.15, where both results are unfolded to full phase space. The Asimov A_C values differ between the single-lepton and all-hadronic channel, due to the differences in the parton-level A_C value in the non-all-hadronic[(9)] and all-hadronic $t\bar{t}$ MC samples. At the parton-level, the A_C does not depend on the decay channel, and thus the differences in the values are of statistical nature. Regardless of the value of the Asimov parton-level A_C, the linear response of the unfolding to injected asymmetries is more important to ensure that different A_C values in data can be unfolded without significant bias. The total uncertainty includes the bias uncertainty. Since according to Eq. 7.17 its magnitude depends on the unfolded A_C value, the bias uncertainty is evaluated for both the Asimov unfoled value as well as the SM NNLO QCD + NLO EWK prediction [207, 208]. The higher-order prediction yields larger bias uncertainty due to higher A_C value compared to the Asimov result, thus leading to a more conservative bias uncertainty. In both cases the bias is a sub-dominant systematic uncertainty.

Not surprisingly, the single-lepton channel achieves higher precision, since it has more statistics in both of the $m_{t\bar{t}}$ bins and its uncertainty is also largely statistically dominated in this kinematic region. The lower statistics in the all-hadronic channel are given by the necessity to employ stricter selection in the all-hadronic signal region to suppress the multijet background. As mentioned in Sec. 8.2.1, the selection in this estimate has not undergone a rigorous optimisation, and thus there is potential for improvement. The kinematic regime is primarily determined by the leading large-R jet minimum p_T due to the ATLAS jet trigger throughput limits. In this estimate, single-jet triggers have been used, however for 2017-2018 period of data-taking, ATLAS has also employed single-jet and di-jet triggers which include a selection on the minimum jet mass, resulting in p_T thresholds as low as 330 GeV, in contrast to the 460 GeV of the highest-threshold single jet trigger employed in this study. Using these triggers, it may be possible to lower the leading large-R jet p_T cut, increasing the statistics particularly in the $m_{t\bar{t}} \in [1000, 1500]$ GeV bin, and possibly extending the measurement to even lower $m_{t\bar{t}}$.

The second, easier-to-achieve step, is the incorporation of improved b-tagging algorithms. In this study, the anti-k_t $R = 0.2$ track jets and the MV2c10 b-tagging algorithm were employed. ATLAS has since examined more advanced machine-learning techniques, namely a deep neural network tagger [222] which uses a recurrent neural network impact-parameter tagger [223] for identification of displaced vertices from B-hadrons, leading up to a factor of two improvement in background rejection [224] for the same signal efficiency. The improved b-tagging background rejection could

[(9)] The $t\bar{t}$ MC simulation in the single-lepton channel includes both single-lepton and dilepton events, since dilepton events can pass the single-lepton selection.

give additional room for achieving higher selection efficiency to reduce the statistical uncertainty, while keeping sufficiently low multijet background rates. Similarly, the optimisation of the choice of the top-quark tagging efficiency WP and the choice of inclusive or contained definition is subject to further optimisation studies.

Given the presented results and the further room for improvement available, the results show promise that an application of machine learning to precision measurements for top-quark charge determination is possible, and there is a reasonable prospect for a combination of the single-lepton and all-hadronic channel, that would improve the A_{C} measurement precision in the region of high $m_{t\bar{t}}$ for further enhanced sensitivity to EFT operators in this kinematic regime.

9

Conclusions

The Run-II of LHC operation was characteristic of not only the unprecedented amount of collected data at the highest achieved pp collision energy of 13 TeV at an accelerator, but also marked many interesting milestones in terms of development of new techniques to push the reach of BSM physics searches as well as precision of measurements. This thesis focused on an intersection of the field of top-quark physics precision measurements, advanced object identification techniques for boosted hadronic decays of massive particles, and applications of machine learning.

During the ATLAS Run-II period, the machine learning techniques were applied in the field of identification of boosted hadronically-decaying top quarks and W bosons, showing substantial improvements in the multijet background suppression, especially for top-quark tagging, where the rich decay structure provides many features exploitable in a multivariate technique. It has been demonstrated in this thesis, that the ATLAS MC simulation is capable of modelling these complex machine learning discriminants in agreement with the data within the systematic uncertainties. The signal efficiency measurement is a proof-of-concept of a possibility to perform calibrations of these techniques with estimates of uncertainties, allowing for their application in BSM searches and measurements on the collected data. All of these studies were published in the paper in Ref. [1].

The benefits of the large Run-II dataset have greatly helped push the precision top-quark physics frontier for rare phenomena, such as the charge asymmetry. Despite the large gluon fusion dilution at $\sqrt{s} = 13$ TeV, for the first time at a hadron collider, an evidence of non-zero A_{C} in $t\bar{t}$ production was observed in the inclusive measurement with a significance of $4\,\sigma$, as published in a ATLAS conference note in Ref. [2], where previous measurements were both compatible with zero asymmetry and the SM prediction. In contrast to the separate resolved and boosted topology Run-I A_{C} measurements, the Run-II measurement was designed from the beginning with a consideration of the combination of these two topologies. Both the inclusive and differential A_{C} measurements, show good agreement with the SM prediction, at a precision that is factor 3 to 8 times improved compared to previous Run-I measurements. The improvements in the precision translate also into the sensitivity to BSM physics. An EFT interpretation of the inclusive A_{C} and differential A_{C} vs $m_{t\bar{t}}$ measurements was done, improving the constraints on the C^{-} Wilson coefficient by almost a factor of two.

Finally, a study of the possibility of A_{C} measurement in the boosted all-hadronic channel has been performed. This study was motivated by the fact that the boosted all-hadronic channel would potentially contribute in the region of $m_{t\bar{t}} > 1$ TeV, where the A_{C} sensitivity to EFT is higher due to the smaller gluon-fusion dilution. In this channel, the full $t\bar{t}$ system is reconstructed, with no ambiguities due to presence of no neutrinos in the final state. The signal-to-noise ratio in this channel can be improved compared to different past measurements in this $t\bar{t}$ topology by using the previously examined neural network boosted top-quark tagging techniques. The main challenge of this measurement is the necessity

to distinguish reconstructed top quark from top anti-quark using other means than the clean signature of a final state prompt lepton in the single-lepton channel. In this thesis, it is demonstrated, that using neural networks and tracking information from the inner detector, it is possible to perform this charge identification with a precision much higher than previously applied approaches for top-quark charge measurement not relying on semi-leptonic quark decays. It is also demonstrated, that while some tension in the modelling is observed, it is plausible that accounting for the uncertainties on the observed mis-modelling of the NN discriminant would not limit the A_{C} measurement in the all-hadronic channel. The first estimate of A_{C} in this channel is presented, and is found to be statistically limited, however, with a room for further improvement in the selection efficiency at a small to no drop in the background rejection. The estimation of the multijet background uncertainties requires further studies, since these uncertainties were found to be among the largest systematic uncertainties. Given the presented results, a combination with the single-lepton channel would most benefit the $m_{t\bar{t}} > 1.5\,\mathrm{TeV}$ region. However, with the further studies and selection efficiency optimisations, it may be possible to attempt to combine the channels with a non-negligible improvement for $m_{t\bar{t}}$ as low as 1 TeV.

Bibliography

[1] ATLAS Collaboration, *Performance of top-quark and W-boson tagging with ATLAS in Run 2 of the LHC*, Eur. Phys. J. C **79** (2019) 375.

[2] ATLAS Collaboration, *Inclusive and differential measurement of the charge asymmetry in $t\bar{t}$ events at 13 TeV with the ATLAS detector*, ATLAS-CONF-2019-026, 2019, https://cds.cern.ch/record/2682109.

[3] *Standard Model of Elementary Particles*, https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg. Figure version from September 17, 2019. Accessed 2020-03-01.

[4] C. N. Yang and R. L. Mills, *Conservation of Isotopic Spin and Isotopic Gauge Invariance*, Phys. Rev. **96** (1954) 191–195.

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

[6] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264–1266.

[7] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. C **680519** (1968) 367–377.

[8] C. S. Wu, et al., *Experimental Test of Parity Conservation in Beta Decay*, Phys. Rev. **105** (1957) 1413–1415.

[9] R. P. Feynman and M. Gell-Mann, *Theory of the Fermi Interaction*, Phys. Rev. **109** (1958) 193–198.

[10] J. H. Christenson, et al., *Evidence for the 2π Decay of the K_2^0 Meson*, Phys. Rev. Lett. **13** (1964) 138–140.

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

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

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

[14] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1.

[15] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B **716** (2012) 30.

[16] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, *Advantages of the Color Octet Gluon Picture*, Phys. Lett. B **47** (1973) 365–368.

[17] M. Gell-Mann, *The Eightfold way: A theory of strong interaction symmetry*, (1961).

[18] Y. Neeman, *Derivation of strong interactions from a gauge invariance*, Nucl. Phys. **26** (1961) 222–229.

[19] G. Zweig, *An SU_3 model for strong interaction symmetry and its breaking; Version 2*, https://cds.cern.ch/record/570209.

[20] D. J. Gross and F. Wilczek, *Ultraviolet Behavior of Non-Abelian Gauge Theories*, Phys. Rev. Lett. **30** (1973) 1343–1346.

[21] H. D. Politzer, *Reliable Perturbative Results for Strong Interactions?*, Phys. Rev. Lett. **30** (1973) 1346–1349.

[22] K. G. Wilson, *Confinement of quarks*, Phys. Rev. D **10** (1974) 2445–2459.

[23] B. Andersson, et al., *Parton Fragmentation and String Dynamics*, Phys. Rept. **97** (1983) 31–145.

[24] B. R. Webber, *A QCD Model for Jet Fragmentation Including Soft Gluon Interference*, Nucl. Phys. B **238** (1984) 492–528.

[25] J. D. Bjorken, *Asymptotic Sum Rules at Infinite Momentum*, Phys. Rev. **179** (1969) 1547–1553.

[26] Y. L. Dokshitzer, *Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.*, Sov. Phys. JETP **46** (1977) 641–653.

[27] V. Gribov and L. Lipatov, *Deep inelastic e p scattering in perturbation theory*, Sov. J. Nucl. Phys. **15** (1972) 438–450.

[28] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, Nucl. Phys. B **126** (1977) 298–318.

[29] L. Del Debbio, *Parton distributions in the LHC era*, EPJ Web of Conferences **175** (2018) 01006.

[30] *Summary plots from the ATLAS Standard Model physics group*, https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SM/index.html. Accessed: 2020-04-24.

[31] P. D. G. Collaboration, *Review of Particle Physics*, Phys. Rev. D **98** (2018) 030001.

[32] J. Gao, C. S. Li, and H. X. Zhu, *Top-Quark Decay at Next-to-Next-to-Leading Order in QCD*, Phys. Rev. Lett. **110** (2013) 042001.

[33] D0 Collaboration, *Measurement of the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV using secondary vertex b-tagging*, Physical Review D **74** (2006).

[34] P. Bärnreuther, M. Czakon, and A. Mitov, *Percent Level Precision Physics at the Tevatron: First Genuine NNLO QCD Corrections to $q\bar{q} \to t\bar{t} + X$*, Phys. Rev. Lett. **109** (2012) 132001.

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

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

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

[38] M. Czakon and A. Mitov, *Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders*, Comput. Phys. Commun. **185** (2014) 2930.

[39] A. D. Martin, et al., *Parton distributions for the LHC*, Eur. Phys. J. C **63** (2009) 189–285.

[40] *LHC Top Working Group summary plots*, `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TopPairCrossectionSqrtsHistory`. Figure version from 2019-05. Accessed: 2020-04-05.

[41] J. H. Kühn and G. Rodrigo, *Charge asymmetry of heavy quarks at hadron colliders*, Physical Review D **59** (1999) 054017.

[42] W. Bernreuther and Z.-G. Si, *Top quark and leptonic charge asymmetries for the Tevatron and LHC*, Phys. Rev. D **86** (2012) 034026.

[43] W. Hollik and D. Pagani, *The electroweak contribution to the top quark forward-backward asymmetry at the Tevatron*, Phys. Rev. D **84** (2011) 093003.

[44] J. H. Kuhn and G. Rodrigo, *Charge asymmetries of top quarks at hadron colliders revisited*, JHEP **01** (2012) 063.

[45] S. Frixione and B. R. Webber, *Matching NLO QCD computations and parton shower simulations*, Journal of High Energy Physics **2002** (2002) 029–029.

[46] L. G. Almeida, G. Sterman, and W. Vogelsang, *Threshold resummation for the top quark charge asymmetry*, Phys. Rev. D **78** (2008) 014008.

[47] O. Antuñano, J. H. Kühn, and G. Rodrigo, *Top quarks, axigluons, and charge asymmetries at hadron colliders*, Phys. Rev. D **77** (2008) 014003.

[48] M. T. Bowen, S. D. Ellis, and D. Rainwater, *Standard model top quark asymmetry at the Fermilab Tevatron*, Phys. Rev. D **73** (2006) 014008.

[49] CDF Collaboration, *Evidence for a mass dependent forward-backward asymmetry in top quark pair production*, Phys. Rev. D **83** (2011) 112003.

[50] D0 Collaboration, *Forward-backward asymmetry in top quark-antiquark production*, Phys. Rev. D **84** (2011) 112005.

[51] CDF Collaboration, *Measurement of the top quark forward-backward production asymmetry and its dependence on event kinematic properties*, Physical Review D **87** (2013) 092002.

[52] D0 Collaboration, *Measurement of the forward-backward asymmetry in top quark-antiquark production in $p\bar{p}$ collisions using the lepton+jets channel*, Phys. Rev. D **90** (2014) 072011.

[53] CDF Collaboration and D0 Collaboration, *Combined Forward-Backward Asymmetry Measurements in Top-Antitop Quark Production at the Tevatron*, Phys. Rev. Lett. **120** (2018) 042001.

[54] CDF Collaboration, *Measurement of the inclusive leptonic asymmetry in top-quark pairs that decay to two charged leptons at CDF*, Phys. Rev. Lett. **113** (2014) 042001, [Erratum: Phys. Rev. Lett.117,no.19,199901(2016)].

[55] D0 Collaboration, *Measurement of the forward-backward asymmetry in the distribution of leptons in $t\bar{t}$ events in the lepton+jets channel*, Phys. Rev. D **90** (2014) 072001.

[56] CDF Collaboration, *Measurement of the Leptonic Asymmetry in $t\bar{t}$ Events Produced in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV*, Phys. Rev. D **88** (2013) 072003, [Erratum: Phys. Rev.D94,no.9,099901(2016)].

[57] D0 Collaboration, *Simultaneous measurement of forward-backward asymmetry and top polarization in dilepton final states from $t\bar{t}$ production at the Tevatron*, Phys. Rev. D **92** (2015) 052007.

[58] W. Bernreuther and Z.-G. Si, *Top quark and leptonic charge asymmetries for the Tevatron and LHC*, Phys. Rev. D **86** (2012) 034026.

[59] M. Czakon, et al., *NNLO QCD predictions for fully-differential top-quark pair production at the Tevatron*, JHEP **05** (2016) 034.

[60] M. Czakon, et al., *Top-pair production at the LHC through NNLO QCD and NLO EW*, JHEP **10** (2017) 186.

[61] CDF Collaboration, *Measurement of the forward-backward asymmetry in low-mass bottom-quark pairs produced in proton-antiproton collisions*, Phys. Rev. D **93** (2016) 112003.

[62] CDF Collaboration, *First measurement of the forward-backward asymmetry in bottom-quark pair production at high mass*, Phys. Rev. D **92** (2015) 032006.

[63] CDF Collaboration Collaboration, CDF Collaboration, *Exclusion of exotic top-like quarks with $-4/3$ electric charge using jet-charge tagging in single-lepton $t\bar{t}$ events at CDF*, Phys. Rev. D **88** (2013) 032003.

[64] B. Grinstein and C. W. Murphy, *Bottom-Quark Forward-Backward Asymmetry in the Standard Model and Beyond*, Phys. Rev. Lett. **111** (2013) 062003.

[65] ATLAS Collaboration, *Measurement of the top quark pair production charge asymmetry in proton–proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, JHEP **02** (2014) 107.

[66] ATLAS Collaboration, *Measurement of the charge asymmetry in dileptonic decays of top quark pairs in pp collisions at $\sqrt{s}$ = 7 TeV using the ATLAS detector*, JHEP **05** (2015) 061.

[67] CMS Collaboration, *Inclusive and differential measurements of the $t\bar{t}$ charge asymmetry in proton–proton collisions at $\sqrt{s}$ = 7 TeV*, Phys. Lett. B **717** (2012) 129.

[68] CMS Collaboration, *Measurements of the $t\bar{t}$ charge asymmetry using the dilepton decay channel in pp collisions at $\sqrt{s}$ = 7 TeV*, JHEP **04** (2014) 191.

[69] 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*, Eur. Phys. J. C **76** (2016) 87.

[70] ATLAS Collaboration, *Measurements of the charge asymmetry in top-quark pair production in the dilepton final state at $\sqrt{s}$ = 8 TeV with the ATLAS detector*, Phys. Rev. D **94** (2016) 032006.

[71] CMS Collaboration, *Inclusive and differential measurements of the $t\bar{t}$ charge asymmetry in pp collisions at $\sqrt{s}$ = 8 TeV*, Phys. Lett. B **757** (2016) 154.

[72] CMS Collaboration, *Measurement of the charge asymmetry in top quark pair production in pp collisions at $\sqrt{s}$ = 8 TeV using a template method*, Phys. Rev. D **93** (2016) 034014.

[73] CMS Collaboration, *Measurements of $t\bar{t}$ charge asymmetry using dilepton final states in pp collisions at $\sqrt{s}$ = 8 TeV*, Phys. Lett. B **760** (2016) 365.

[74] ATLAS and CMS Collaborations, *Combination of inclusive and differential $t\bar{t}$ charge asymmetry measurements using ATLAS and CMS data at $\sqrt{s}$ = 7 TeV and 8 TeV*, JHEP **04** (2018) 033.

[75] ATLAS Collaboration, *Measurement of the charge asymmetry in highly boosted top-quark pair production in $\sqrt{s}$ = 8 TeV pp collision data collected by the ATLAS experiment*, Phys. Lett. B **756** (2016) 52.

[76] J. A. Aguilar-Saavedra, *Portrait of a colour octet*, JHEP **08** (2014) 172.

[77] M. Czakon, D. Heymes, and A. Mitov, *Dynamical scales for multi-TeV top-pair production at the LHC*, JHEP **04** (2017) 071.

[78] J. A. Aguilar-Saavedra, et al., *Asymmetries in top quark pair production at hadron colliders*, Rev. Mod. Phys. **87** (2015) 421–455.

[79] J. Aguilar-Saavedra and M. Perez-Victoria, *Asymmetries in $t\,\bar{t}$ production: LHC versus Tevatron*, Phys. Rev. D **84** (2011) 115013.

[80] J. Aguilar-Saavedra and M. Perez-Victoria, *Simple models for the top asymmetry: Constraints and predictions*, JHEP **09** (2011) 097.

[81] A. Kobach, *Baryon Number, Lepton Number, and Operator Dimension in the Standard Model*, Phys. Lett. B **758** (2016) 455–457.

[82] B. Grzadkowski, et al., *Dimension-Six Terms in the Standard Model Lagrangian*, JHEP **10** (2010) 085.

[83] C. Zhang and S. Willenbrock, *Effective-Field-Theory Approach to Top-Quark Production and Decay*, Phys. Rev. D **83** (2011) 034006.

[84] M. P. Rosello and M. Vos, *Constraints on four-fermion interactions from the $t\bar{t}$ charge asymmetry at hadron colliders*, Eur. Phys. J. C **76** (2016) 200.

[85] L. Evans and P. Bryant, *LHC Machine*, Journal of Instrumentation **3** (2008) S08001–S08001.

[86] *CERN accelerator complex*, http://cds.cern.ch/record/1621583/files/CERN's-accelerator-complex2013.jpg?version=1. Accessed 18. 2. 2017.

[87] *ATLAS Luminosity public results*, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2. Accessed: 2020-04-07.

[88] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[89] J. Pequenao, *Computer generated image of the ATLAS inner detector*, Mar, 2008, https://cds.cern.ch/record/1095926.

[90] ATLAS Collaboration Collaboration,, *ATLAS muon spectrometer: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1997. https://cds.cern.ch/record/331068.

[91] ATLAS Collaboration Collaboration,, *ATLAS magnet system: Technical Design Report, 1*. Technical Design Report ATLAS. CERN, Geneva, 1997. https://cds.cern.ch/record/338080.

[92] ATLAS Collaboration, *The Run-2 ATLAS Trigger System*, Tech. Rep. ATL-DAQ-PROC-2016-003, CERN, Geneva, Feb, 2016. https://cds.cern.ch/record/2133909.

[93] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4—a simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250.

[94] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823–874.

[95] ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2*, Eur. Phys. J. C **77** (2017) 673.

[96] ATLAS Collaboration, *A neural network clustering algorithm for the ATLAS silicon pixel detector*, JINST **9** (2014) P09009.

[97] ATLAS Collaboration, *Vertex Reconstruction Performance of the ATLAS Detector at* $\sqrt{s}$ = 13 TeV, Tech. Rep. ATL-PHYS-PUB-2015-026, Geneva, Jul, 2015. https://cds.cern.ch/record/2037717.

[98] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015-2017 LHC proton–proton collision data*, arXiv:1908.00005 [hep-ex].

[99] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at* $\sqrt{s}$ = 13 *TeV*, Eur. Phys. J. C **79** (2019) 639.

[100] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using data collected in 2015 at* $\sqrt{s}$ = 13 *TeV*, ATL-PHYS-PUB-2016-015, 2016, https://cds.cern.ch/record/2203514.

[101] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton–proton collision data*, JINST **14** (2019) P03017.

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

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

[104] M. Cacciari, G. P. Salam, and G. Soyez, *The Anti-k(t) jet clustering algorithm*, JHEP **04** (2008) 063.

[105] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, Phys. Rev. D **48** (1993) 3160–3166.

[106] Y. L. Dokshitzer, et al., *Better jet clustering algorithms*, JHEP **08** (1997) 001.

[107] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet user manual*, Eur. Phys. J. C **72** (2012) 1896.

[108] G. P. Salam, *Towards Jetography*, Eur. Phys. J. C **67** (2010) 637–686.

[109] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, Phys. Lett. B **659** (2008) 119–126.

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

[111] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, ATLAS-CONF-2014-018, 2014, http://cdsweb.cern.ch/record/1700870.

[112] ATLAS Collaboration, *Identification of boosted, hadronically decaying W bosons and comparisons with ATLAS data taken at* $\sqrt{s}$ = 8 *TeV*, Eur. Phys. J. C **76** (2016) 154.

[113] ATLAS Collaboration, *Performance of jet substructure techniques for large-R jets in proton–proton collisions at* $\sqrt{s}$ = 7 *TeV using the ATLAS detector*, JHEP **1309** (2013) 076.

[114] D. Krohn, J. Thaler, and L.-T. Wang, *Jet Trimming*, JHEP **02** (2010) 084.

[115] ATLAS Collaboration, *In situ calibration of large-radius jet energy and mass in* 13 *TeV proton–proton collisions with the ATLAS detector*, Eur. Phys. J. C **79** (2019) 135.

[116] ATLAS Collaboration, *Performance of b-jet identification in the ATLAS experiment*, JINST **11** (2016) P04008.

[117] ATLAS Collaboration, *Measurements of b-jet tagging efficiency with the ATLAS detector using* $t\bar{t}$ *events at* $\sqrt{s}$ = 13 *TeV*, JHEP **08** (2018) 089.

[118] ATLAS Collaboration Collaboration,, *Topological b-hadron decay reconstruction and identification of b-jets with the JetFitter package in the ATLAS experiment at the LHC*, Tech. Rep. ATL-PHYS-PUB-2018-025, CERN, Geneva, Oct, 2018. https://cds.cern.ch/record/2645405.

[119] R. Fruhwirth, *Application of Kalman filtering to track and vertex fitting*, Nucl. Instrum. Meth. A **262** (1987) 444–450.

[120] ATLAS Collaboration, E_T^{miss} *performance in the ATLAS using 2015-2016 LHC p-p collisions*, ATLAS-CONF-2018-023, 2018, http://cdsweb.cern.ch/record/2625233.

[121] ATLAS Collaboration, *Boosted hadronic top identification at ATLAS for early* 13 *TeV data*, ATL-PHYS-PUB-2015-053, 2015, http://cdsweb.cern.ch/record/2116351.

[122] ATLAS Collaboration, *Identification of high transverse momentum top quarks in pp collisions at* $\sqrt{s}$ = 8 *TeV with the ATLAS detector*, JHEP **06** (2016) 093.

[123] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with N-subjettiness*, JHEP **03** (2011) 015.

[124] J. Thaler and K. Van Tilburg, *Maximizing boosted top identification by minimizing N-subjettiness*, JHEP **02** (2012) 093.

[125] G. Altarelli, B. Mele, and M. Ruiz-Altaba, *Searching for New Heavy Vector Bosons in* $p\bar{p}$ *Colliders*, Z. Phys. C **45** (1989) 109, [Erratum: Z.Phys.C 47, 676 (1990)].

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

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

[128] ATLAS Collaboration, *Jet mass reconstruction with the ATLAS Detector in early Run 2 data*, ATLAS-CONF-2016-035, 2016, https://cds.cern.ch/record/2200211.

[129] A. J. Larkoski, G. P. Salam, and J. Thaler, *Energy correlation functions for jet substructure*, JHEP **06** (2013) 108.

[130] A. J. Larkoski, I. Moult, and D. Neill, *Power counting to better jet observables*, JHEP **12** (2014) 009.

[131] ATLAS Collaboration, *ATLAS measurements of the properties of jets for boosted particle searches*, Phys. Rev. D **86** (2012) 072006.

[132] L. G. Almeida, et al., *Top quark jets at the LHC*, Phys. Rev. **D79** (2009) 074012.

[133] J. Thaler and L.-T. Wang, *Strategies to identify boosted tops*, JHEP **07** (2008) 092.

[134] S. Catani, et al., *Longitudinally invariant K_t clustering algorithms for hadron hadron collisions*, Nucl. Phys. **B406** (1993) 187–224.

[135] ATLAS Collaboration, *Measurement of the cross-section of high transverse momentum vector bosons reconstructed as single jets and studies of jet substructure in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, New J. Phys. **16** (2014) 113013.

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

[137] C. Chen, *New approach to identifying boosted hadronically-decaying particle using jet substructure in its center-of-mass frame*, Phys. Rev. D **85** (2012) 034007.

[138] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[139] J. Pearkes, et al., *Jet Constituents for Deep Neural Network Based Top Quark Tagging*, `arXiv:1704.02124 [hep-ex]`.

[140] D. E. Soper and M. Spannowsky, *Finding top quarks with shower deconstruction*, Physical Review D **87** (2013).

[141] D. E. Soper and M. Spannowsky, *Finding physics signals with shower deconstruction*, Phys. Rev. D **84** (2011) 074002.

[142] S. Wertz, *The Matrix Element Method in the LHC era*, EPJ Web of Conferences **137** (2017) 11010.

[143] T. Plehn, G. P. Salam, and M. Spannowsky, *Fat Jets for a Light Higgs*, Phys. Rev. Lett. **104** (2010) 111801.

[144] T. Plehn, et al., *Stop reconstruction with tagged tops*, JHEP **10** (2010) 078.

[145] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852–867.

[146] A. D. Martin, et al., *Parton distributions for the LHC*, Eur. Phys. J. C **63** (2009) 189–285.

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

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

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

[150] P. Kant, et al., *HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions*, Comput. Phys. Commun. **191** (2015) 74–89.

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

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

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

[154] H.-L. Lai, et al., *New parton distributions for collider physics*, Phys. Rev. D **82** (2010) 074024.

[155] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) 026.

[156] J. Pumplin, et al., *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP **07** (2002) 012.

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

[158] ATLAS Collaboration, *Simulation of top-quark production for the ATLAS experiment at* $\sqrt{s}$ = 13 *TeV*, ATL-PHYS-PUB-2016-004, 2016, https://cds.cern.ch/record/2120417.

[159] M. Bahr et al., *Herwig++ Physics and Manual*, Eur. Phys. J. C **58** (2008) 639.

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

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

[162] C. Anastasiou, et al., *High precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at NNLO*, Phys. Rev. D **69** (2004) 094008.

[163] J. M. Campbell and R. K. Ellis, *An Update on vector boson pair production at hadron colliders*, Phys. Rev. D **60** (1999) 113006.

[164] ATLAS Collaboration, *Search for heavy particles decaying into top-quark pairs using lepton-plus-jets events in proton–proton collisions at* $\sqrt{s}$ = 13 *TeV with the ATLAS detector*, Eur. Phys. J. C **78** (2018) 565.

[165] ATLAS Collaboration Collaboration, ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling for Top2016*, Tech. Rep. ATL-PHYS-PUB-2016-020, CERN, Geneva, Sep, 2016. https://cds.cern.ch/record/2216168.

[166] ATLAS Collaboration, G. Aad et al., *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*, Eur. Phys. J. C **79** (2019) 1028.

[167] ATLAS Collaboration, *Jet energy resolution in proton–proton collisions at $\sqrt{s}$ = 7 TeV recorded in 2010 with the ATLAS detector*, Eur. Phys. J. C **73** (2013) 2306.

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

[169] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at $\sqrt{s}$ = 13 TeV*, Eur. Phys. J. C **78** (2018) 903.

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

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

[172] A. Collaboration, *Luminosity determination in pp collisions at $\sqrt{s}$ = 13 TeV using the ATLAS detector at the LHC*, Tech. Rep. ATLAS-CONF-2019-021, CERN, Geneva, Jun, 2019. https://cds.cern.ch/record/2677054.

[173] ATLAS Collaboration, *ATLAS simulation of boson plus jets processes in Run 2*, ATL-PHYS-PUB-2017-006, 2017, https://cds.cern.ch/record/2261937.

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

[175] ATLAS Collaboration, *Estimation of non-prompt and fake lepton backgrounds in final states with top quarks produced in proton–proton collisions at $\sqrt{s}$ = 8 TeV with the ATLAS Detector*, ATLAS-CONF-2014-058, 2014, http://cdsweb.cern.ch/record/1951336.

[176] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, J. Phys. G **43** (2016) 023001.

[177] Y. Cui, Z. Han, and M. D. Schwartz, *W-jet tagging: Optimizing the identification of boosted hadronically-decaying W bosons*, Physical Review D **83** (2011).

[178] A. Czarnecki, J. G. Körner, and J. H. Piclum, *Helicity fractions ofWbosons from top quark decays at next-to-next-to-leading order in QCD*, Physical Review D **81** (2010).

[179] G. Choudalakis, *Fully Bayesian Unfolding*, arXiv:1201.4612 [hep-ex].

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

[181] E. Bothmann et al., *Event Generation with Sherpa 2.2*, arXiv:1905.09127 [hep-ph].

[182] T. Sjöstrand, et al., *An Introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159–177.

[183] 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, https://cds.cern.ch/record/2206965.

[184] S. Frixione, P. Nason, and G. Ridolfi, *A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, JHEP **09** (2007) 126.

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

[186] L. A. Harland-Lang, et al., *Parton distributions in the LHC era: MMHT 2014 PDFs*, Eur. Phys. J. C **75** (2015) 204.

[187] J. Aguilar-Saavedra, *Single top quark production at LHC with anomalous Wtb couplings*, Nucl. Phys. B **804** (2008) 160–192.

[188] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, Eur. Phys. J. C **71** (2011) 1547.

[189] R. Frederix, E. Re, and P. Torrielli, *Single-top t-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO*, JHEP **09** (2012) 130.

[190] S. Alioli, et al., *NLO single-top production matched with shower in POWHEG: s- and t-channel contributions*, JHEP **09** (2009) 111, [Erratum: JHEP 02 (2010) 011].

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

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

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

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

[195] A. Denner, S. Dittmaier, and L. Hofer, *Collier: a fortran-based Complex One-Loop LIbrary in Extended Regularizations*, Comput. Phys. Commun. **212** (2017) 220–238.

[196] S. Frixione, et al., *Electroweak and QCD corrections to top-pair hadroproduction in association with heavy bosons*, arXiv:1504.03446 [hep-ph].

[197] LHC Higgs Cross Section Working Group Collaboration, D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, arXiv:1610.07922 [hep-ph].

[198] J. Erdmann, et al., *A likelihood-based reconstruction algorithm for top-quark pairs and the KLFitter framework*, Nucl. Instrum. Meth. A **748** (2014) 18–25.

[199] T. Dado, *Direct measurement of the top-quark decay width with the ATLAS detector*. PhD thesis, Georg-August Universität, Göttingen and Comenius University, Bratislava, 2019. http://hdl.handle.net/21.11130/00-1735-0000-0003-C15B-D.

[200] A. Höcker, et al., *TMVA - Toolkit for Multivariate Data Analysis with ROOT: Users guide.*, Tech. Rep. physics/0703039, CERN, Geneva, Mar, 2007. https://cds.cern.ch/record/1019880.

[201] R. J. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*, Comput. Phys. Commun. **77** (1993) 219–228.

[202] P. Diaconis, *The Markov Chain Monte Carlo Revolution*, Bull. Amer. Math. Soc. (N.S.) **46** (2009) 179–205.

[203] J. Salvatier, T. V. Wiecki, and C. Fonnesbeck, *Probabilistic programming in Python using PyMC3*, PeerJ Computer Science **2** (2016) e55.

[204] M. D. Hoffman and A. Gelman, *The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo*, arXiv:1111.4246 [stat.CO].

[205] A. Gelman and D. B. Rubin, *Inference from Iterative Simulation Using Multiple Sequences*, Statist. Sci. **7** (1992) 457–472.

[206] G. Bohm and G. Zech, *Introduction to Statistics and Data Analysis for Physicists; 2nd rev. ed.* Verlag Deutsches Elektronen-Synchrotron, Hamburg, 2014. https://bib-pubdb1.desy.de/record/169869.

[207] M. Czakon, et al., *Top-quark charge asymmetry at the LHC and Tevatron through NNLO QCD and NLO EW*, Phys. Rev. D **98** (2018) 014003.

[208] M. Czakon, et al., *Differential distributions for top-pair production in NNLO QCD + NLO EW*, http://www.precision.hep.phy.cam.ac.uk/results/ttbar-nnloqcd-nloew/. Accessed: 2019-06-03.

[209] ATLAS Collaboration, *Measurements of $t\bar{t}$ differential cross-sections of highly boosted top quarks decaying to all-hadronic final states in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector*, Phys. Rev. D **98** (2018) 012003.

[210] M. Cacciari, G. P. Salam, and G. Soyez, *The Catchment Area of Jets*, JHEP **04** (2008) 005.

[211] ATLAS Collaboration, *Flavor Tagging with Track-Jets in Boosted Topologies with the ATLAS Detector*, ATL-PHYS-PUB-2014-013, 2014, https://cds.cern.ch/record/1750681.

[212] ATLAS Collaboration, *Measurement of the top quark charge in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, JHEP **11** (2013) 031.

[213] ATLAS Collaboration, *Measurement of jet charge in dijet events from $\sqrt{s} = 8$ TeV pp collisions with the ATLAS detector*, Phys. Rev. D **93** (2016) 052003.

[214] ATLAS Collaboration, *A new tagger for the charge identification of b-jets*, Tech. Rep. ATL-PHYS-PUB-2015-040, CERN, Geneva, Sep, 2015. https://cds.cern.ch/record/2048132.

[215] Ž. Ivezić, et al., *Statistics, Data Mining and Machine Learning in Astronomy*. Princeton University Press, 2014.

[216] O. Behnke, et al., *Data Analysis in High Energy Physics: A Practical Guide to Statistical Methods*. Wiley-VCH, 1st ed., 2013.

[217] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, arXiv:1412.6980 [cs.LG].

[218] S. Ruder, *An overview of gradient descent optimization algorithms*, arXiv:1609.04747 [cs.LG].

[219] F. Chollet, *Keras*, 2015. https://github.com/fchollet/keras.

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

[221] G. Klambauer, et al., *Self-Normalizing Neural Networks*, arXiv:1706.02515 [cs.LG].

[222] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*, ATL-PHYS-PUB-2017-013, 2017, https://cds.cern.ch/record/2273281.

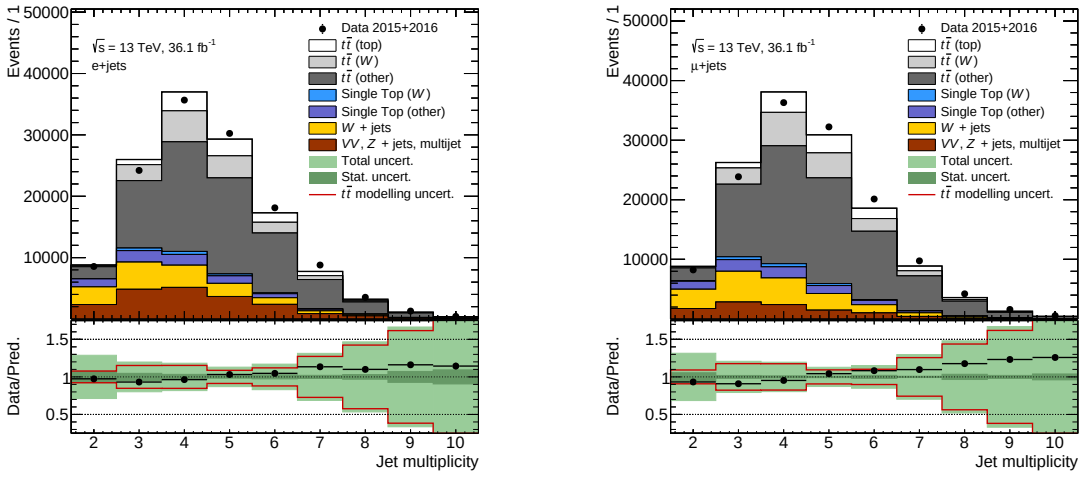
[223] ATLAS Collaboration, *Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment*, ATL-PHYS-PUB-2017-003, 2017, https://cds.cern.ch/record/2255226.

[224] ATLAS Collaboration, *Performance of 2019 recommendations of ATLAS Flavor Tagging algorithms with Variable Radius track jets*, ATLAS-FTAG-2019-006, 2019, https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2019-006, Fig. 01a.

A

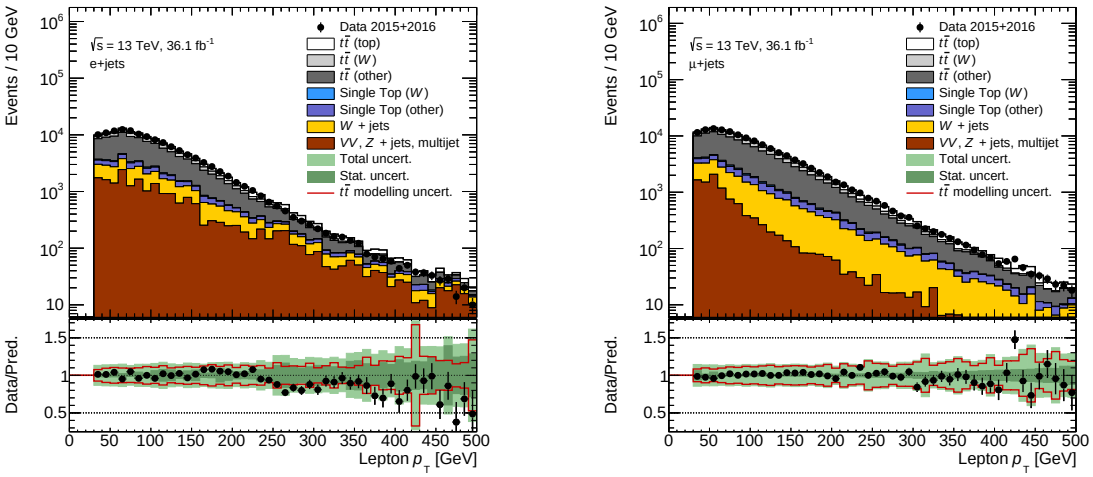
Additional control distributions in the boosted tagging signal efficiency measurement

Here we present additional comparisons between data and the prediction for the signal efficiency measurement of the boosted top-quark and W-boson taggers, shown in Fig. A.1– A.6.

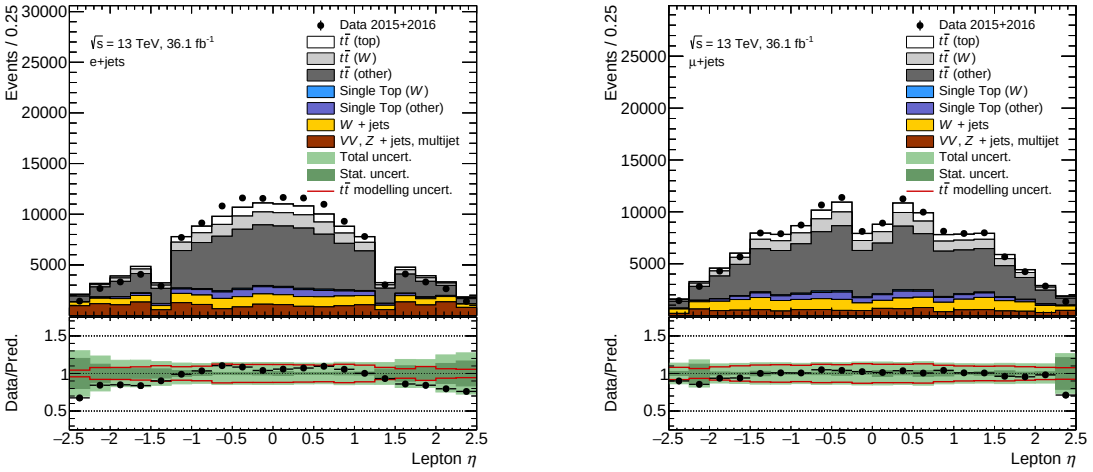


(a) (b)

Fig. A.1: The comparison of data and prediction for the multiplicity of small-R jets in event, for electron channel (left) and muon channel (right). The MC-simulated predictions are normalised to data. The Data/Pred. ratio shows the statistical uncertainty of the prediction (filled dark green band), the total uncertainty (filled light green band) combining statistical and systematic uncertainties in quadrature, and the impact of the $t\bar{t}$ modelling systematic uncertainties (empty red band). The error bars on data points (black) show the statistical uncertainty of data.

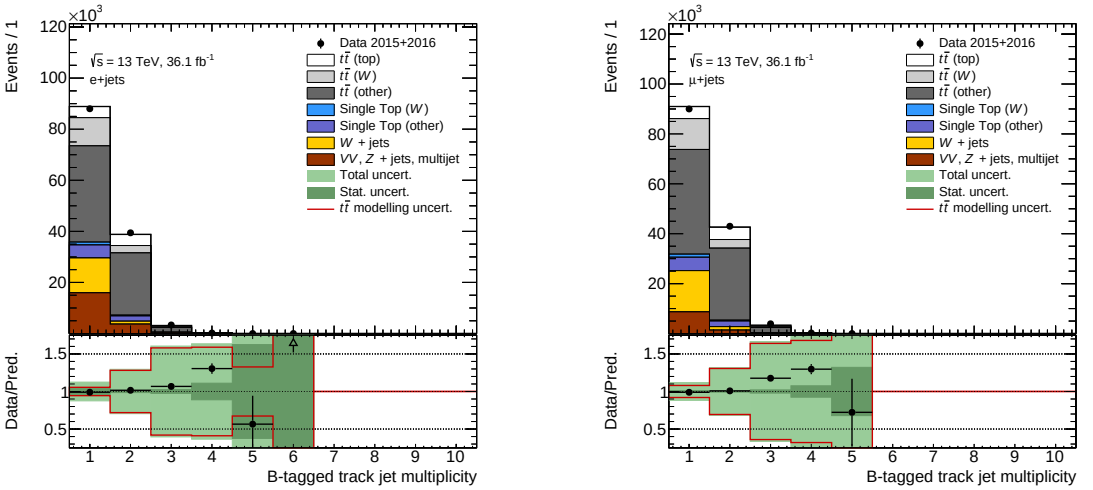


(a) (b)



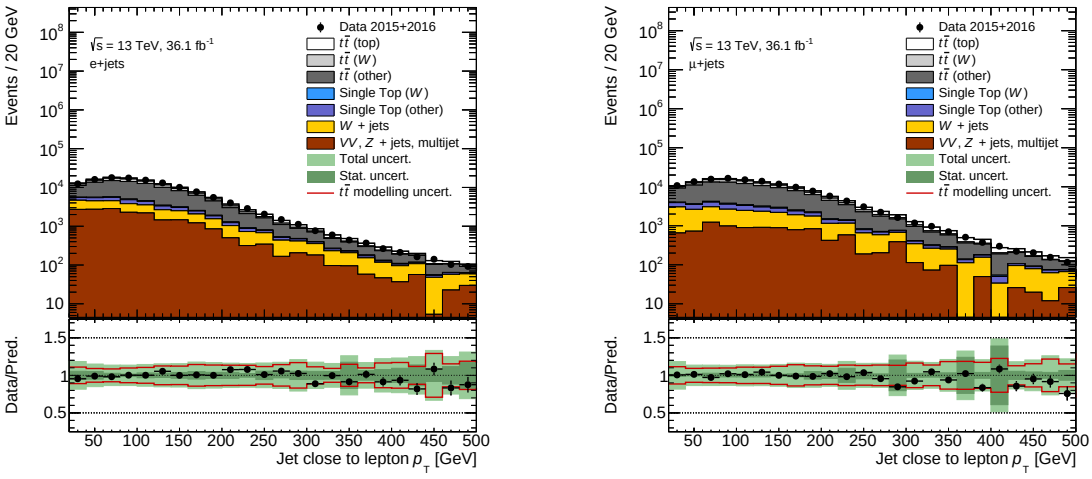
(c) (d)

Fig. A.2: The comparison of data and prediction for the p_{T} (top) and η (bottom) of the isolated lepton, for electron channel (left) and muon channel (right). The plot style follows the convention from Fig. A.1.

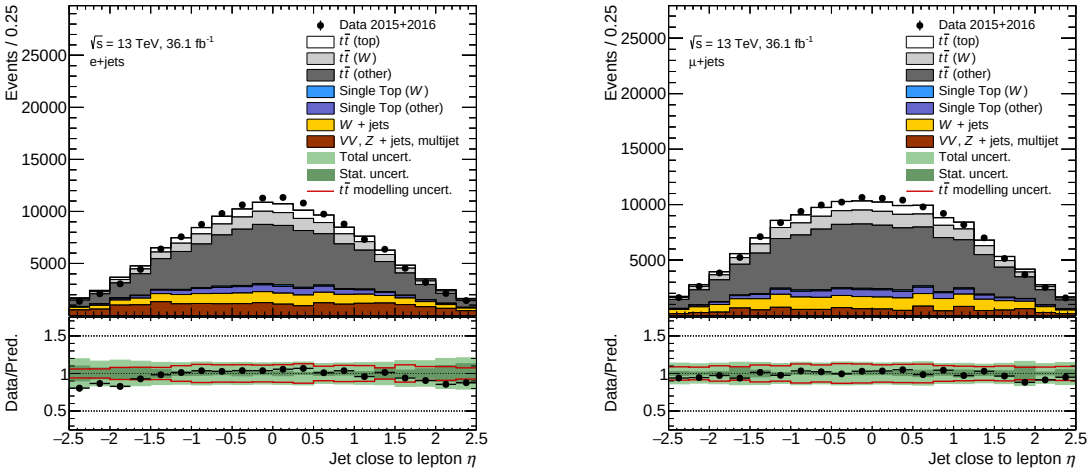


(a) (b)

Fig. A.3: The comparison of data and prediction for the multiplicity of b-tagged track jets in event, for electron channel (left) and muon channel (right). The plot style follows the convention from Fig. A.1.

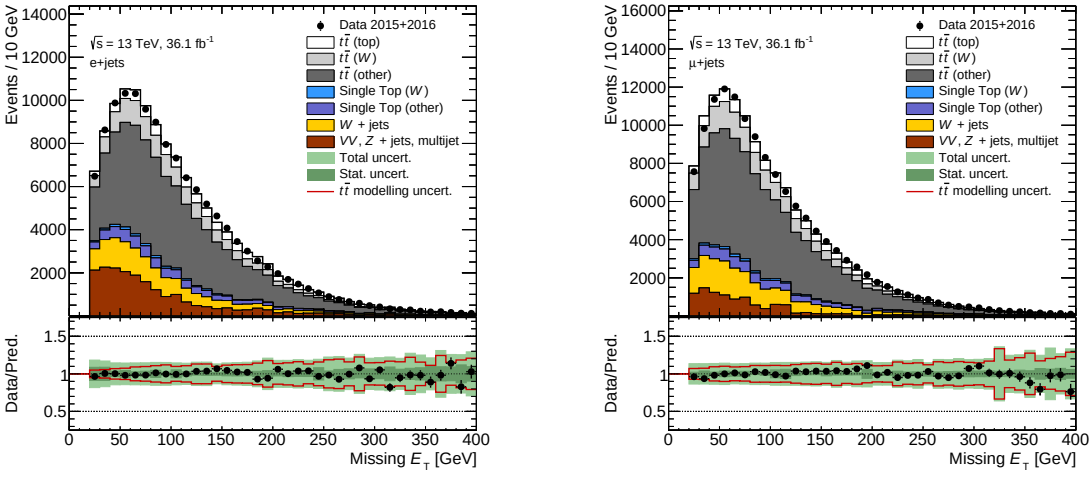


(a) (b)



(c) (d)

Fig. A.4: The comparison of data and prediction for the p_{T} (top) and η (bottom) of the selected small-R jet close to the isolated lepton, for electron channel (left) and muon channel (right). The plot style follows the convention from Fig. A.1.



(a) (b)

Fig. A.5: The comparison of data and prediction for the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ in the event, for electron channel (left) and muon channel (right). The plot style follows the convention from Fig. A.1.

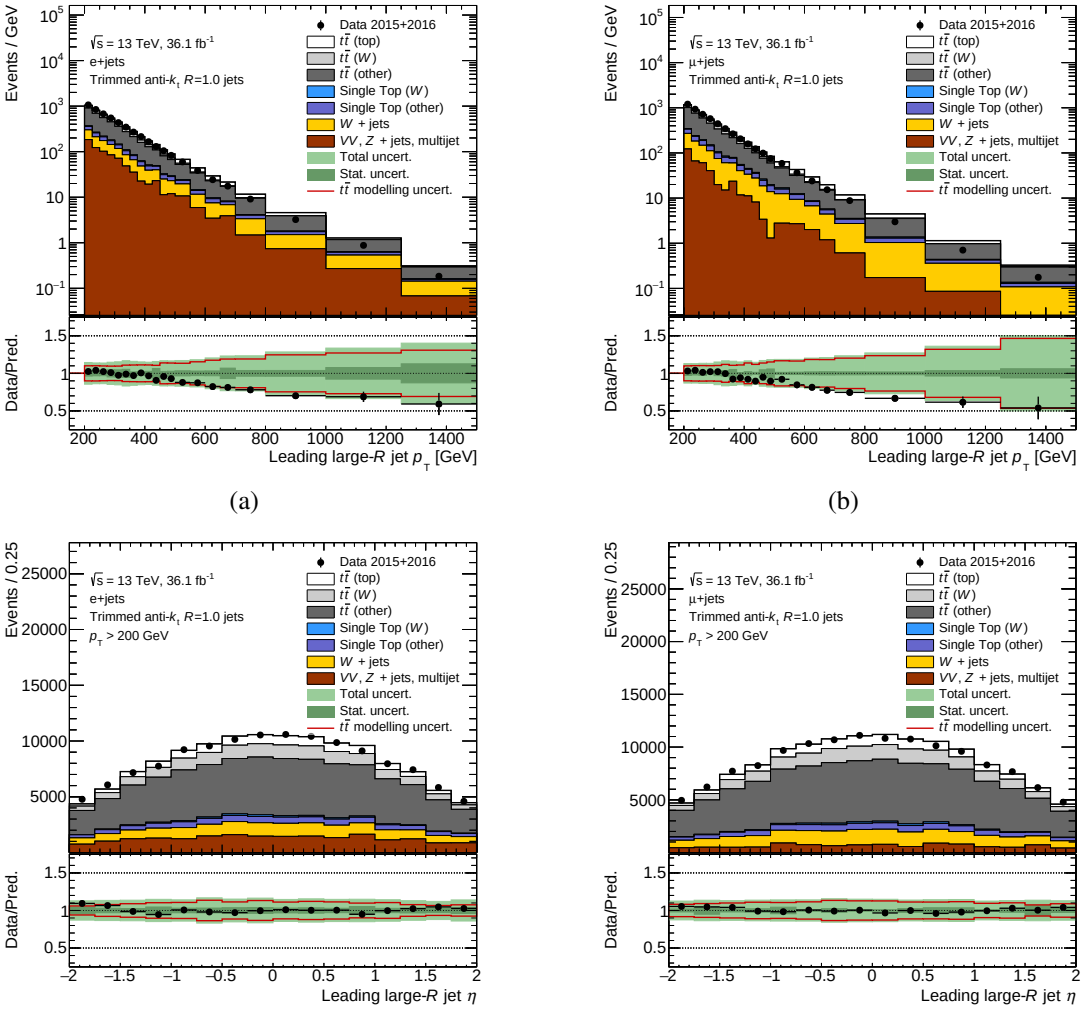


Fig. A.6: The comparison of data and prediction for the p_{T} (top) and η (bottom) of the selected large-R jet isolated from the lepton and the jet close to lepton, for electron channel (left) and muon channel (right). The plot style follows the convention from Fig. A.1.

B

Impact of large-R jet p_{T} mismodelling on substructure observables

A normalisation discrepancy between prediction and data in $t\bar{t}$-enriched topologies is expected due to $t\bar{t}$ process mismodelling discussed in Sec. 6.3. In Fig. B.1a the comparison of data and prediction for the selected large-R jet p_{T} is shown. A significant slope in the ratio plot is observed. A reweighting procedure is performed to check the impact of the large-R jet p_{T} mismodelling on the shape of jet substructure observables. For each event a correction factor is assigned based on the ratio of (data - background) / ($t\bar{t}$) in the corresponding bin of large-R jet p_{T} distribution. In Fig. B.1b the comparison of data and prediction for the selected large-R jet p_{T} after applying the reweighting is shown. Fig. B.2b- B.3b show the original comparison of data and prediction for a number of jet substructure observables, as well as comparisons of prediction before and after applying the reweighting. The data to prediction comparisons include the same normalisation convention and uncertainties as in App. A. In the comparisons of reweighted with non-reweighted predictions, both distributions are normalised to unity.

Variables such as the jet mass (Fig. B.2a, B.2b) and the $\sqrt{d_{23}}$ splitting scale (Fig. B.3a) are found to be sensitive to the reweighting. Particularly large impact is observed in the case of the W-boson mass, where the largest effect is observed at high-mass tail. However, this effect is well outside of the region of the W-boson mass peak, where cuts would typically be imposed by W-boson taggers, reducing the impact of this mismodelling on tagging efficiency. The $\sqrt{d_{23}}$ is another example of a strongly momentum-dependent observable, and is an input variable of the DNN top-quark tagger, however the actual DNN classifier (Fig. B.3b) is found to be quite insensitive to the reweighting. Finally, observables such as τ_{32} (Fig. B.4a) and D_2 (Fig. B.4b) are largely scale-independent by construction and are found to be insensitive to the reweighting.

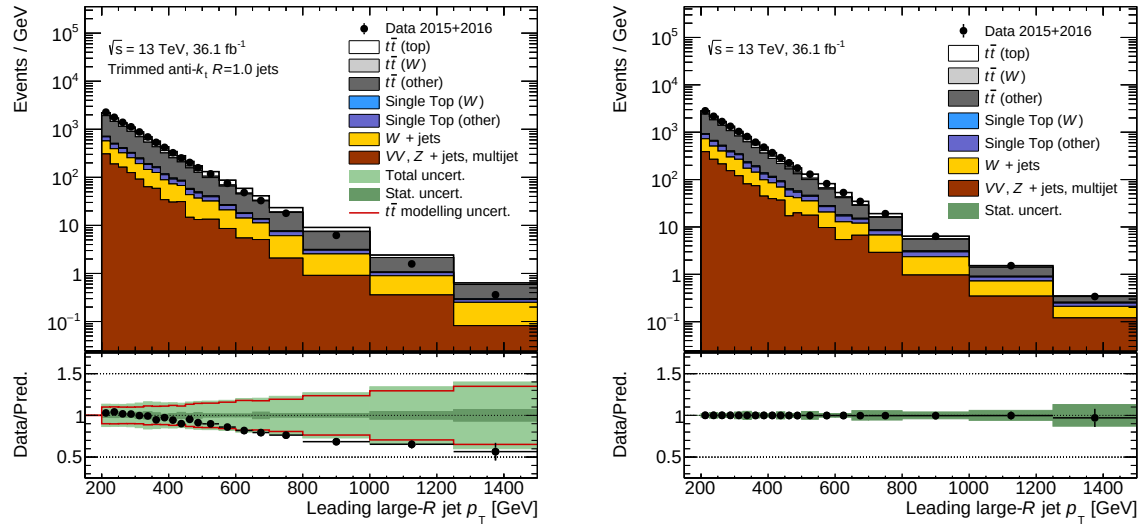


Fig. B.1: Comparison of data and prediction for the distribution of the selected large-R jet p_{T} before (a) and after (b) applying the data-driven reweighting.

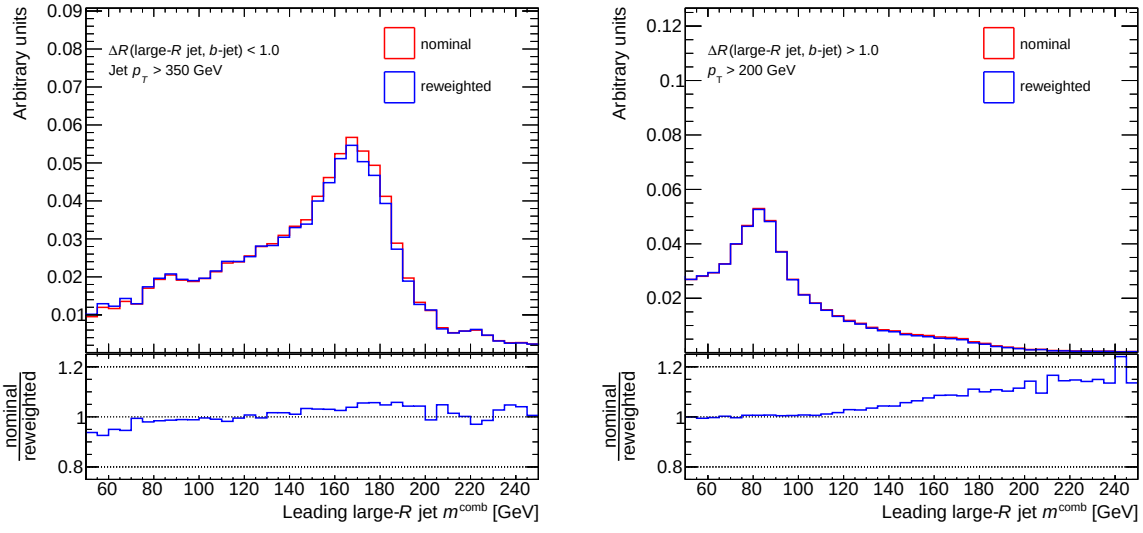
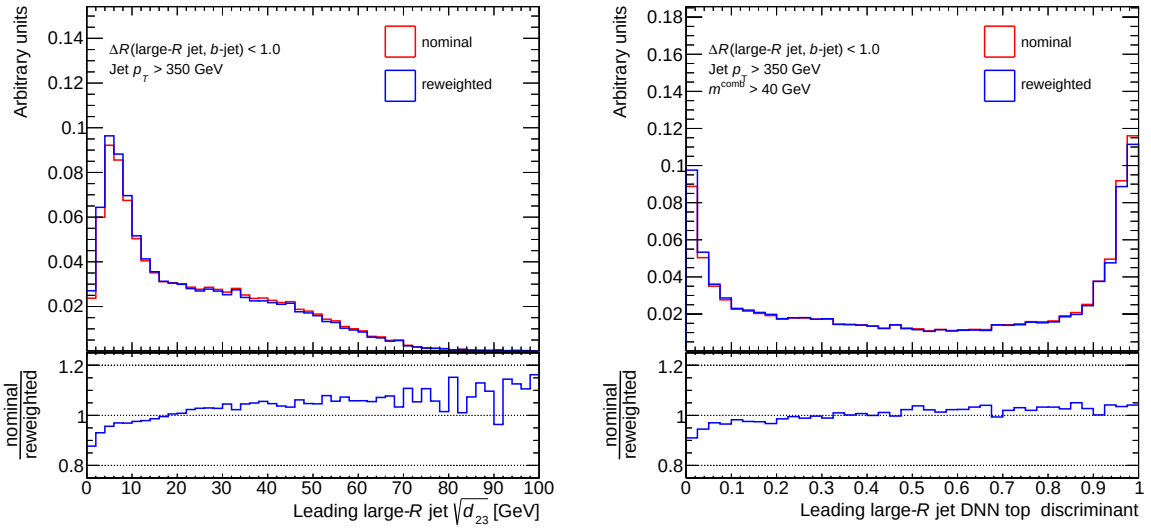
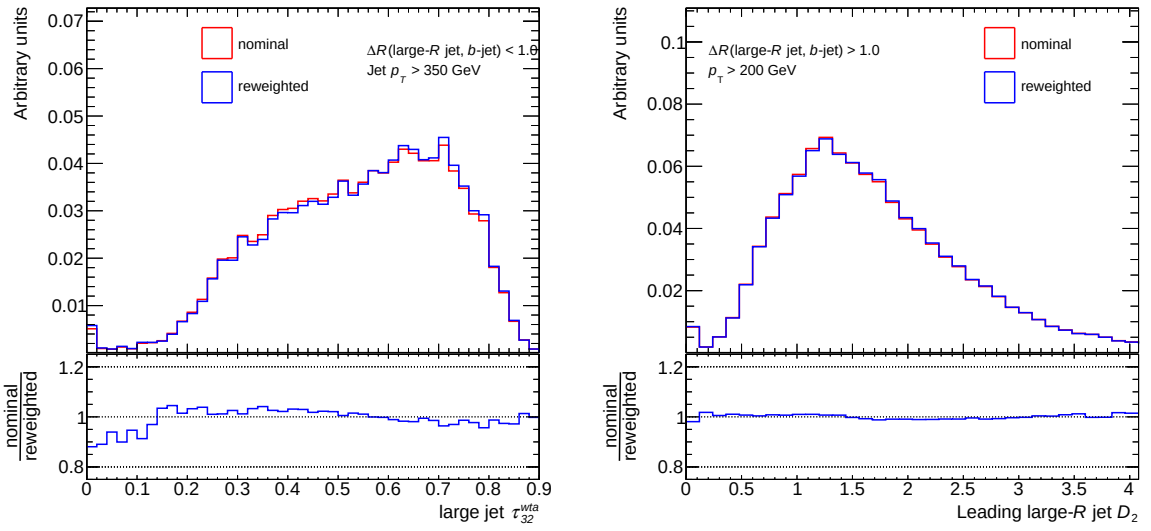


Fig. B.2: Comparison of the predicted distribution before and after applying the data-driven reweighting, of the selected large-R jet mass in the top-quark enriched sub-sample (a) and the W-boson enriched sub-sample (b).



(a) (b)

Fig. B.3: Comparison of the predicted distribution before and after applying the data-driven reweighting, of the selected large-R jet $\sqrt{d_{23}}$ (a) and DNN classifier (b) in the top-quark enriched sub-sample.



(a) (b)

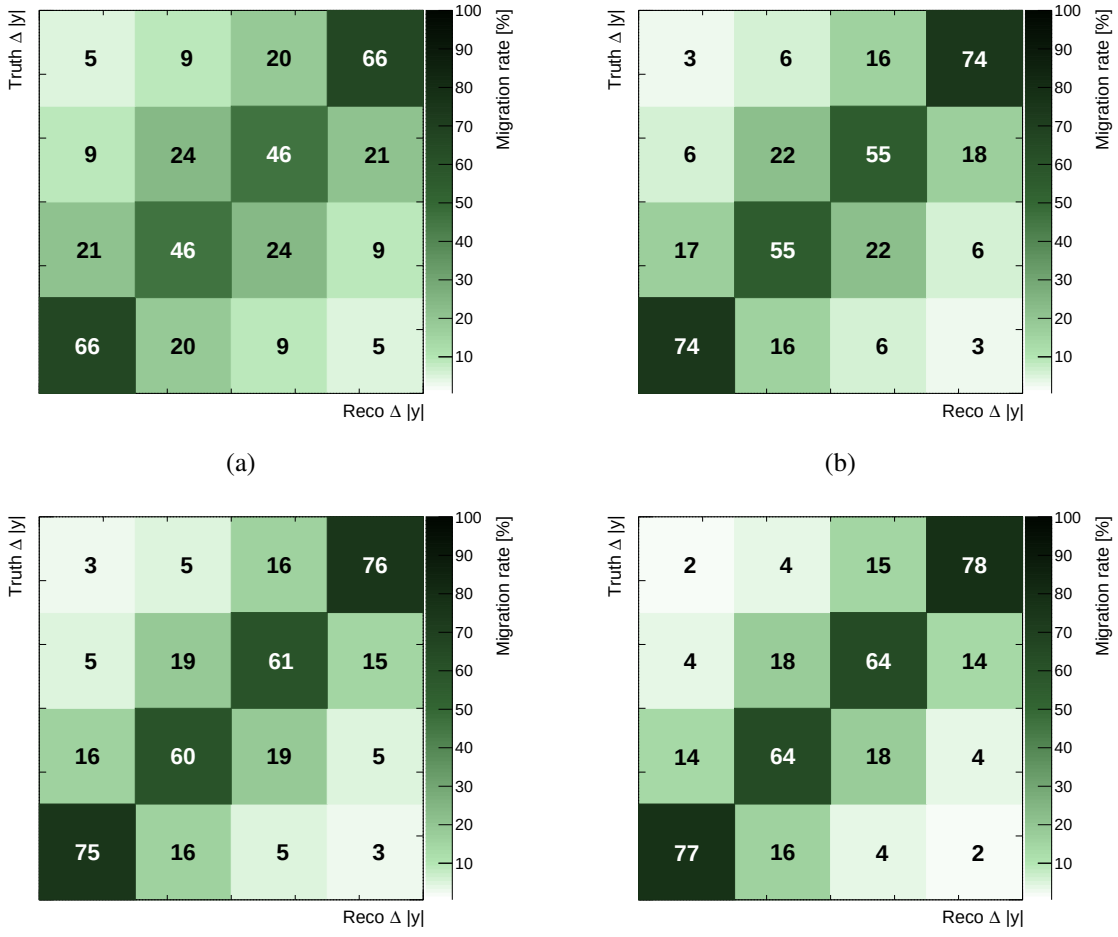
Fig. B.4: Comparison of the predicted distribution before and after applying the data-driven reweighting, of the selected large-R jet τ_{32} in the top-quark enriched sub-sample (a) and the selected large-R jet D_2 in the W-boson enriched sub-sample (b).

C

Additional plots for charge asymmetry measurement in single-lepton channel

C.1 Migration matrices used in the unfolding

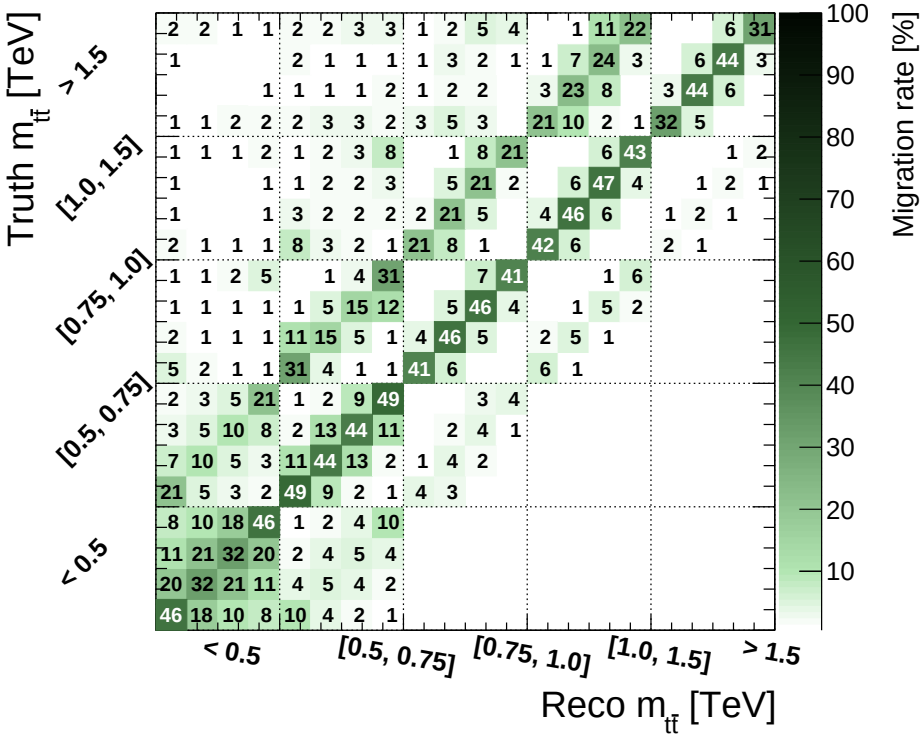
In Fig. C.1– C.5, the migration matrices for the inclusive A_C as well as differential A_C vs $m_{t\bar{t}}$ and $\beta_{z,t\bar{t}}$ measurements are shown. All of the migration rates are expressed in %, using the $\Delta|y|$ bin edges specified in Table 7.3.



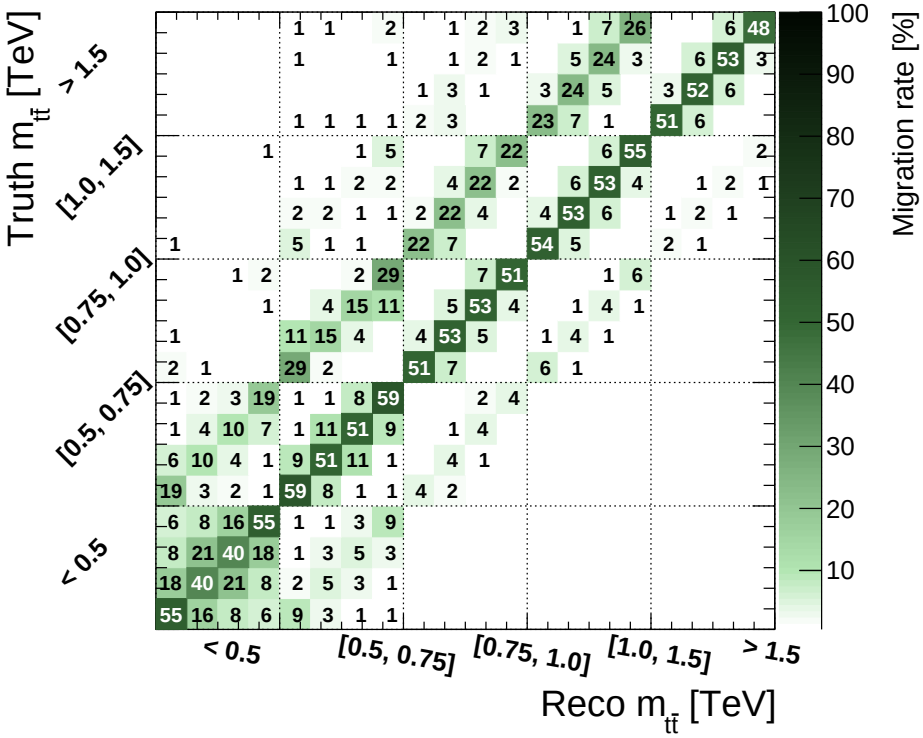
(a) (b)

(c) (d)

Fig. C.1: Migration matrices of the $\Delta|y|$ distribution for the resolved 1b-exclusive (a), 2b-inclusive (b) and boosted 1b-exclusive (c) and 2b-inclusive (d) regions.

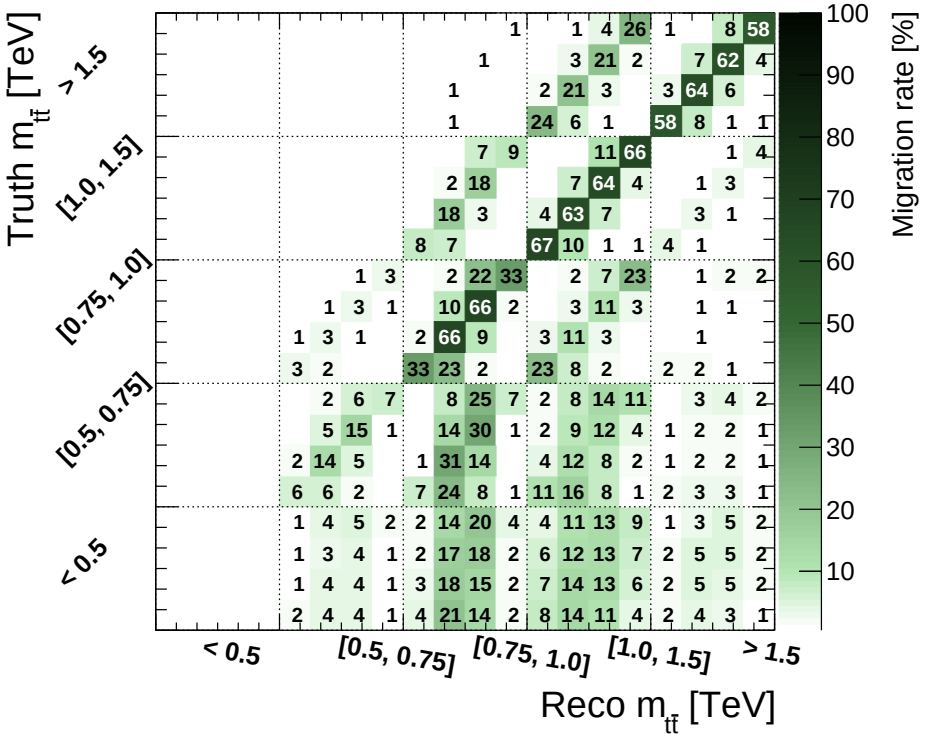


(a)

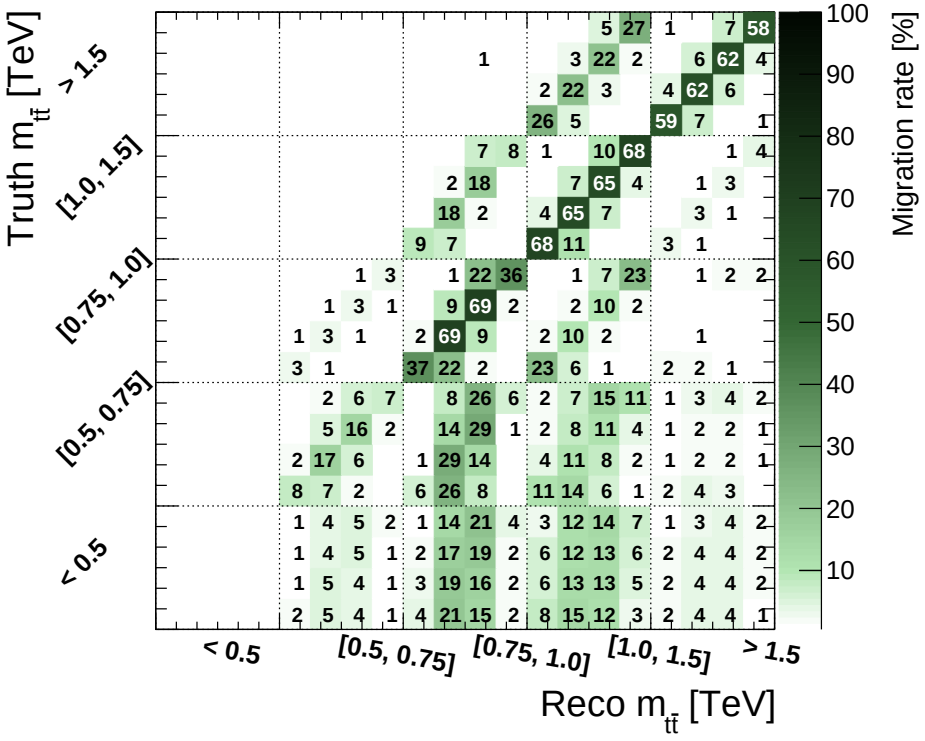


(b)

Fig. C.2: Migration matrices of the $\Delta|y|$ vs $m_{t\bar{t}}$ distribution for the resolved 1b-exclusive (a), 2b-inclusive (b).

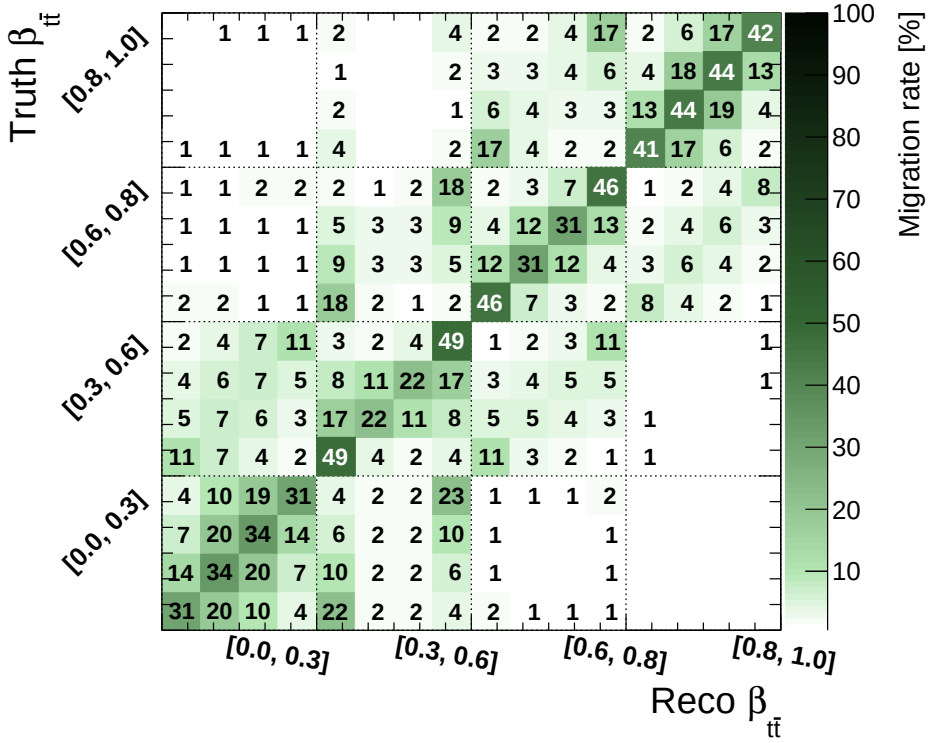


(a)

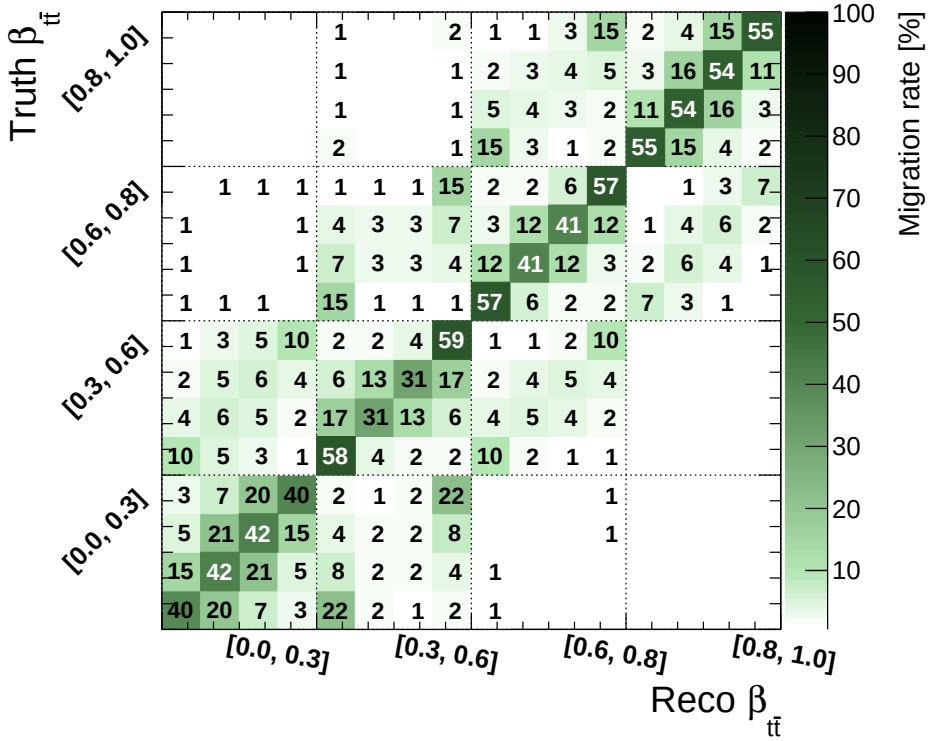


(b)

Fig. C.3: Migration matrices of the $\Delta|y|$ vs $m_{t\bar{t}}$ distribution for the boosted $1b$-exclusive (a), $2b$-inclusive (b).

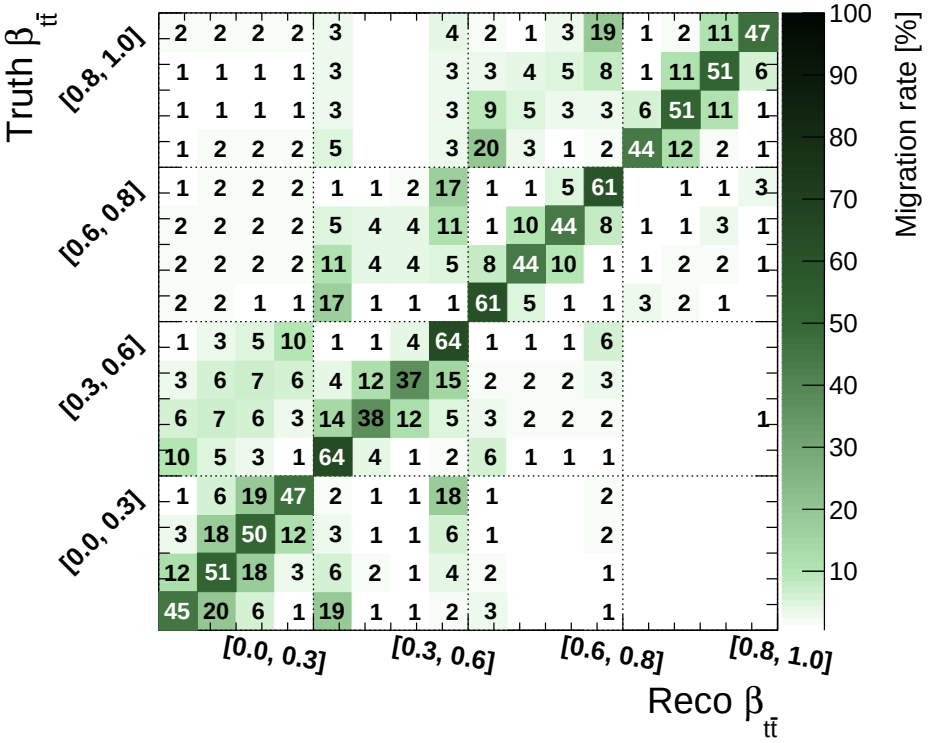


(a)

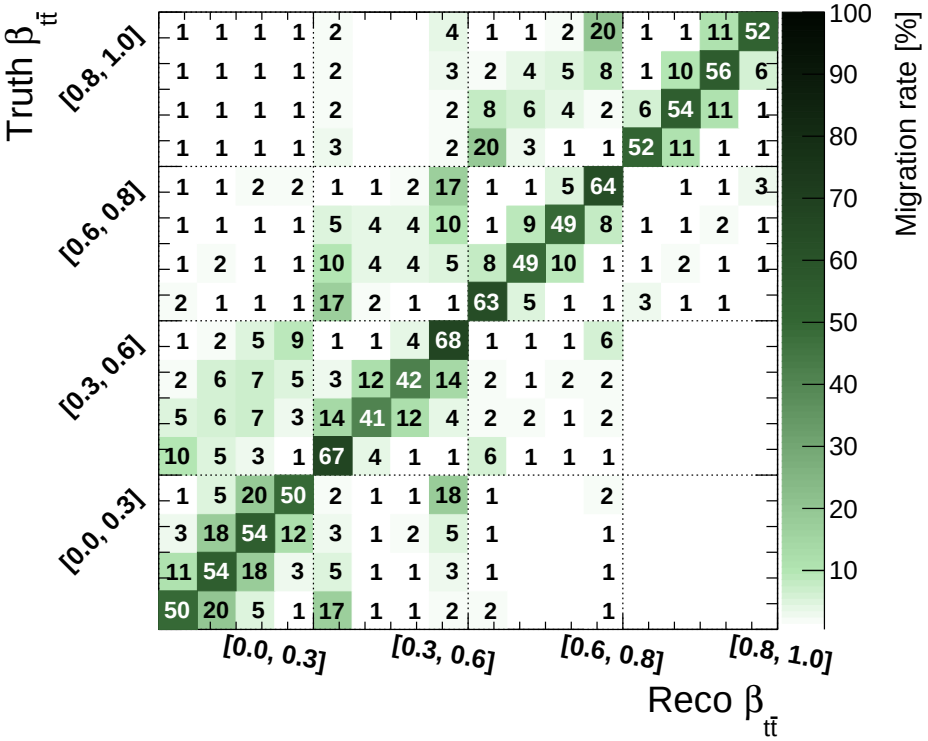


(b)

Fig. C.4: Migration matrices of the $\Delta|y|$ vs $\beta_{z,t\bar{t}}$ distribution for the resolved 1b-exclusive (a), 2b-inclusive (b) regions.



(a)

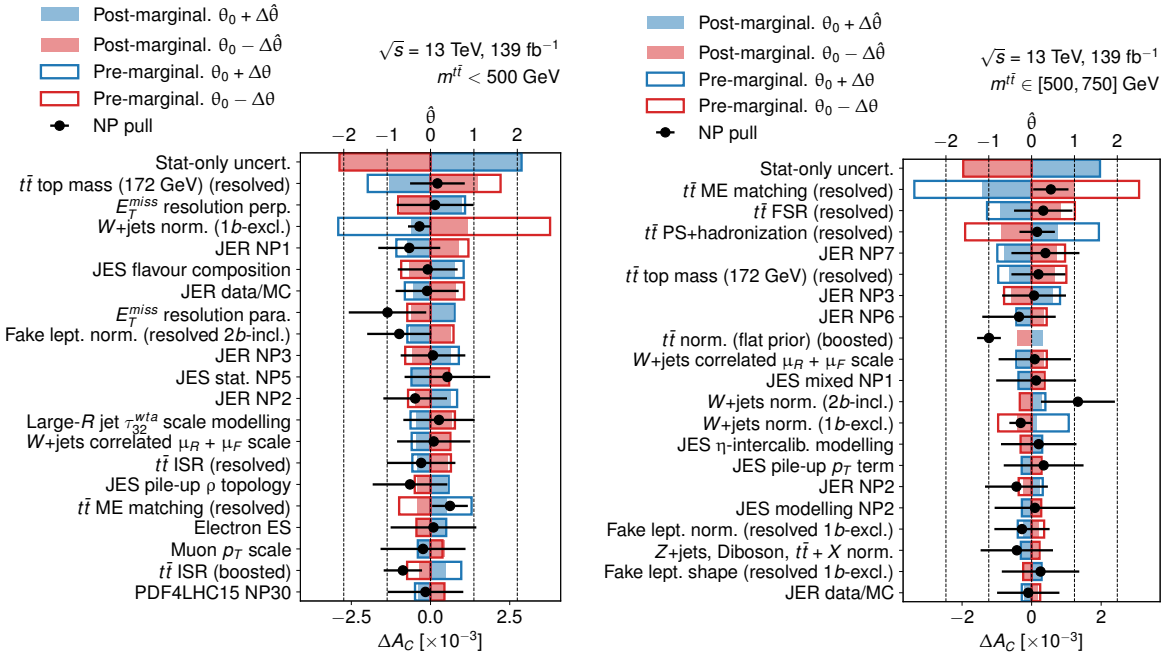


(b)

Fig. C.5: Migration matrices of the $\Delta|y|$ vs $\beta_{z,t\bar{t}}$ distribution for the boosted $1b$-exclusive (a) and $2b$-inclusive (b) regions.

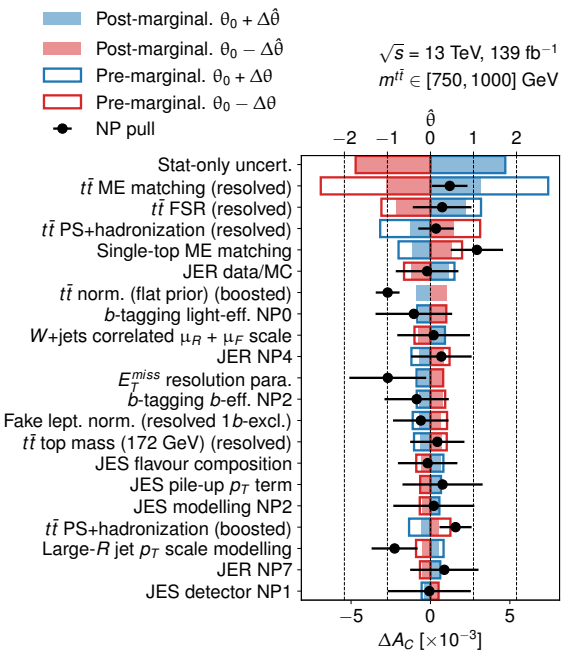
C.2 Ranking of top systematic uncertainties

In Fig. C.6- C.9 the ranking of impact of systematic uncertainties on A_C for the individual differential bins is shown, along with the pull of the corresponding NP obtained in the unfolding of data distributions.

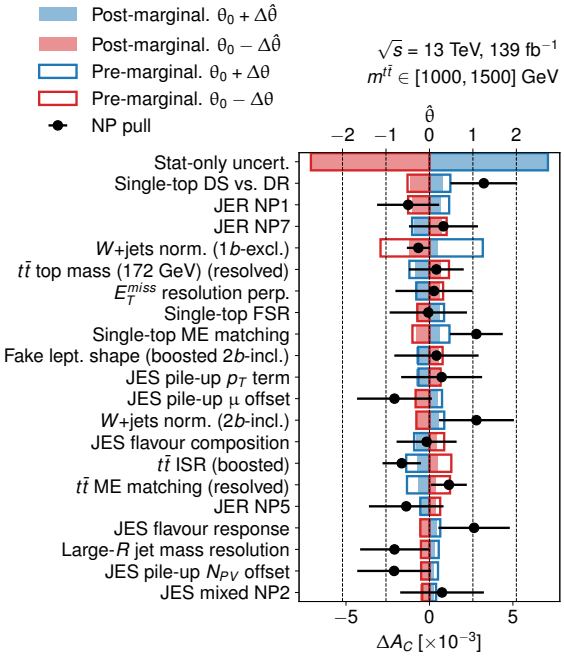


(a) $m_{t\bar{t}} < 500$ GeV (b) $m_{t\bar{t}} \in [500, 750]$ GeV

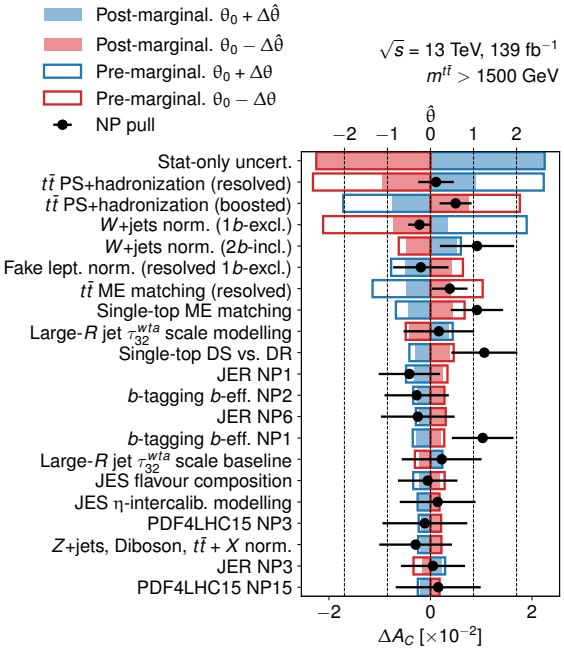
Fig. C.6: The ranking of top 20 systematic uncertainties and the pulls of the corresponding NPs in the differential A_C vs $m_{t\bar{t}}$ measurement. The blue (red) colours shows the impact of up (down) variation of a NP on the A_C. The empty rectangles show the pre-marginalisation impact (un-constrained), while the filled rectangles show the post-marginalisation impact of NP (constrained). The ranking of NPs is ordered by the average of post-marginalisation up and down shift impact. The first row shows the statistical uncertainty on A_C.



(a) $m_{t\bar{t}} \in [750, 1000]$ GeV

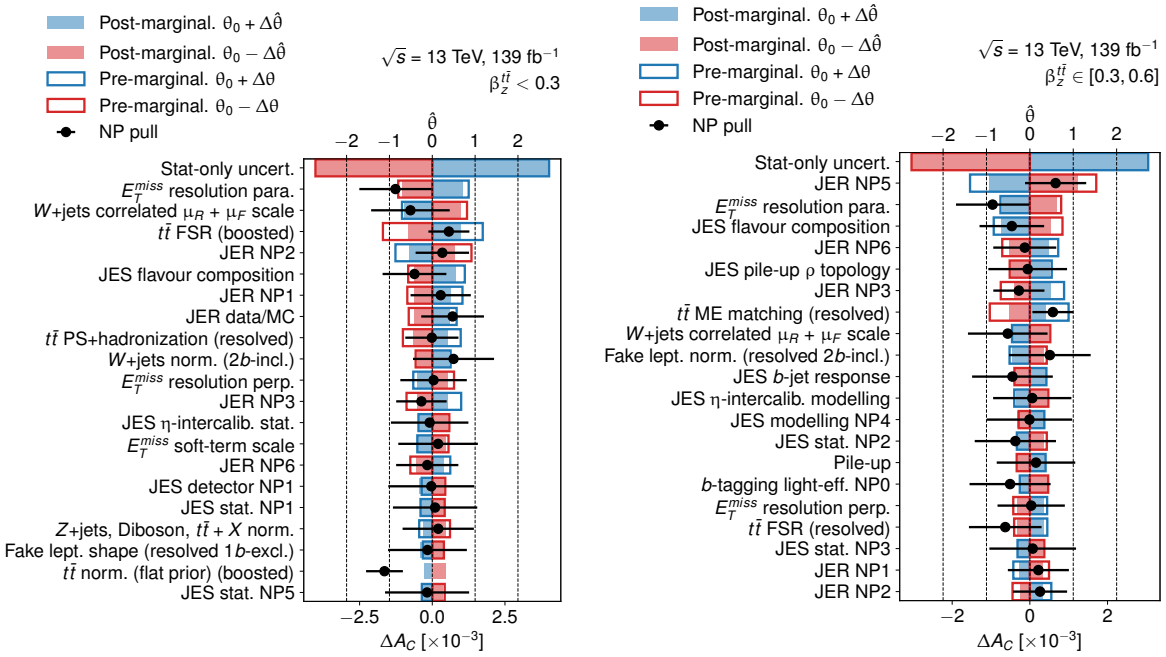


(b) $m_{t\bar{t}} \in [1000, 1500]$ GeV



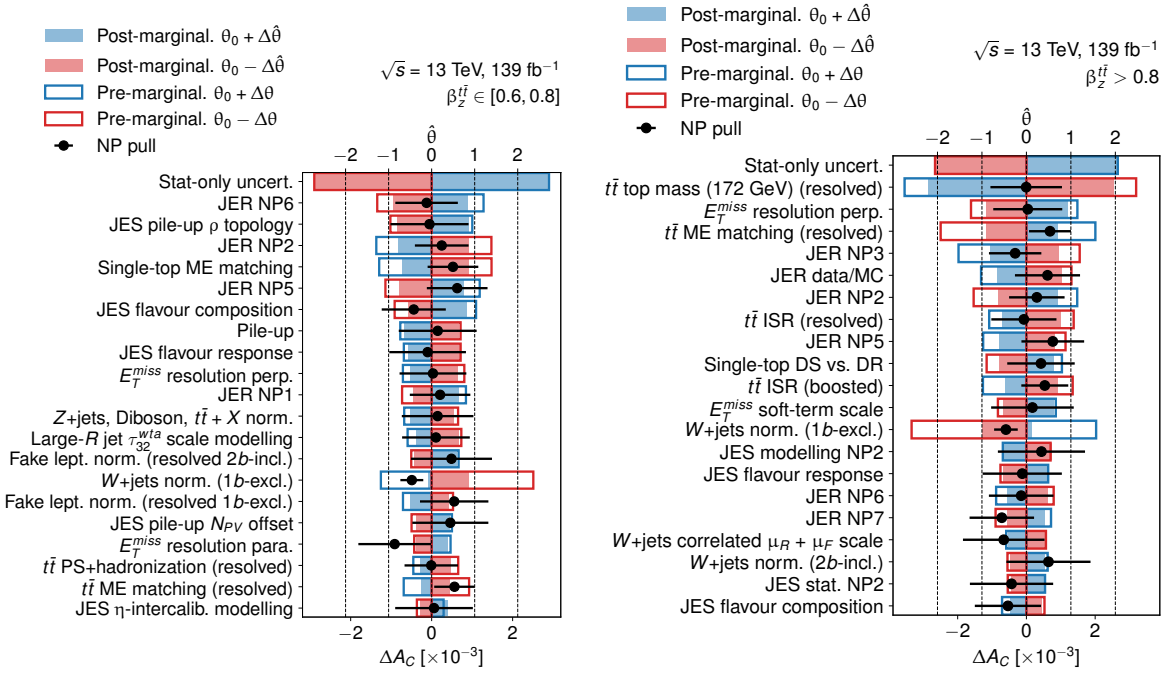
(c) $m_{t\bar{t}} > 1500$ GeV

Fig. C.7: The ranking of top 20 systematic uncertainties and the pulls of the corresponding NPs in the differential A_C vs $m_{t\bar{t}}$ measurement. The plot content follows the same convention as Fig. C.6.



(a) $\beta_{z,t\bar{t}} \in [0, 0.3]$ (b) $\beta_{z,t\bar{t}} \in [0.3, 0.6]$

Fig. C.8: The ranking of top 20 systematic uncertainties and the pulls of the corresponding NPs in the differential A_C vs $\beta_{z,t\bar{t}}$ measurement. The plot content follows the same convention as Fig. C.6.



(a) $\beta_{z,t\bar{t}} \in [0.6, 0.8]$ (b) $\beta_{z,t\bar{t}} \in [0.8, 1.0]$

Fig. C.9: The ranking of top 20 systematic uncertainties and the pulls of the corresponding NPs in the differential A_C vs $\beta_{z,t\bar{t}}$ measurement. The plot content follows the same convention as Fig. C.6.

C.3 Post-marginalisation correlation matrices of nuisance parameters

In Fig. C.10– C.12, the post-marginalisation linear correlation matrices are shown, which are obtained from the unfolding of data,for both inclusive and differential A_C measurements. The correlations are calculated from the ensemble of samples, obtained in the MCMC sampling in the unfolding.

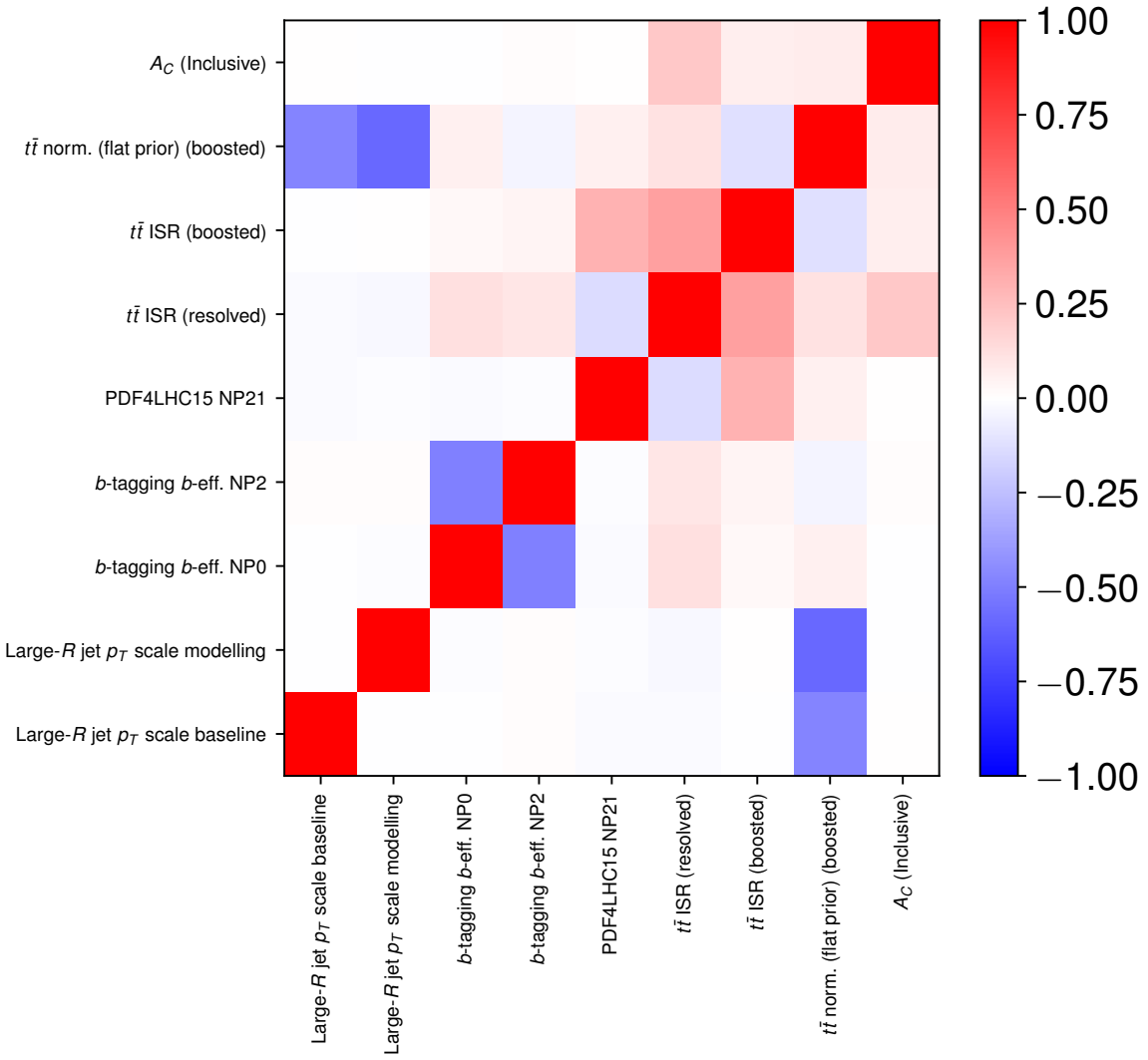


Fig. C.10: The post-marginalisation correlation matrix for NPs, obtained from the unfolding of data in the inclusive A_C measurement. Only correlations greater than 30 % are shown. Additionally, the correlations between A_C and the NPs are also shown.

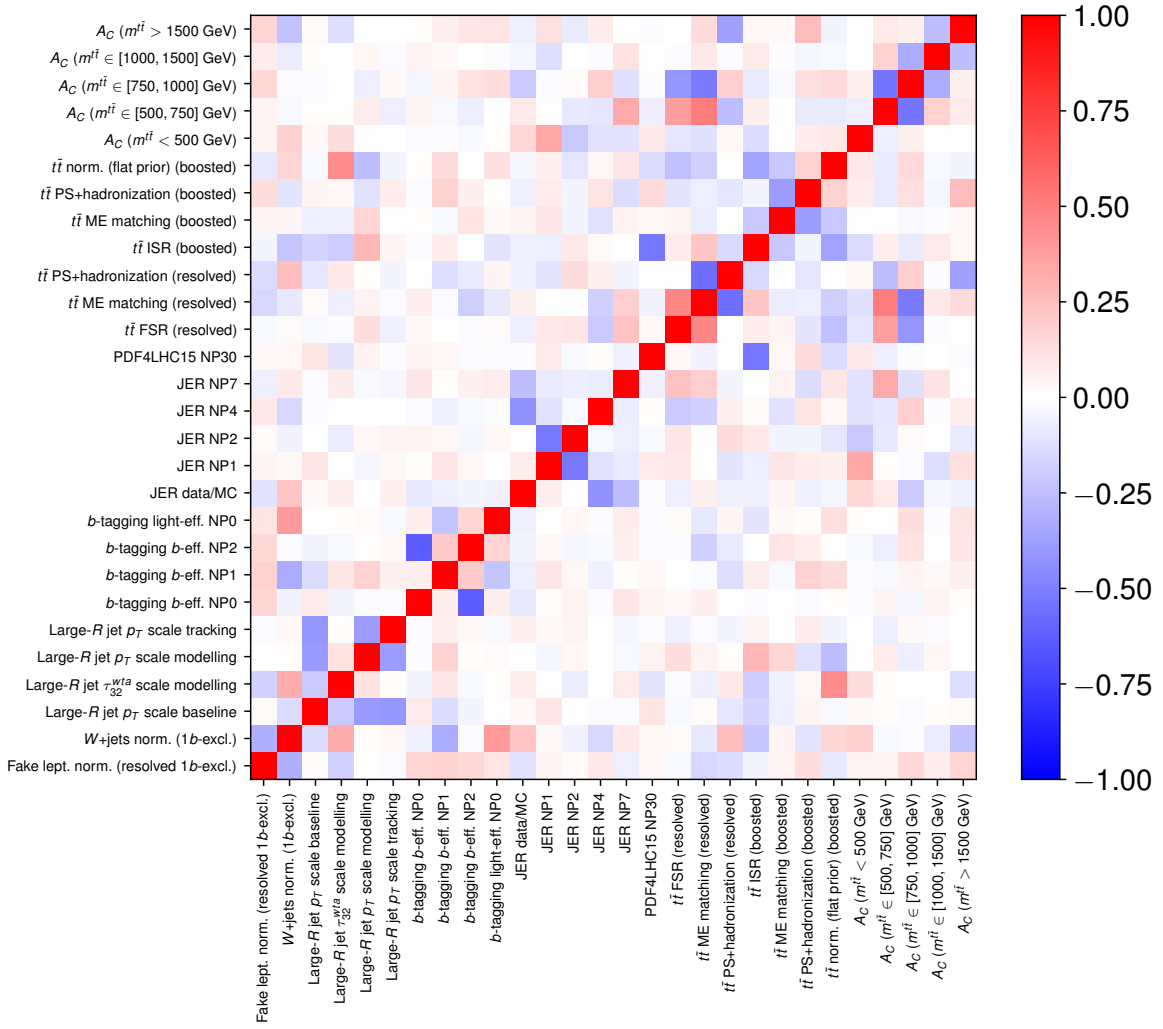


Fig. C.11: The post-marginalisation correlation matrix for NPs, obtained from the unfolding of data in the differential A_C vs $m_{t\bar{t}}$ measurement. Only correlations greater than 30 % are shown. Additionally, the correlations between A_C and the NPs are also shown.

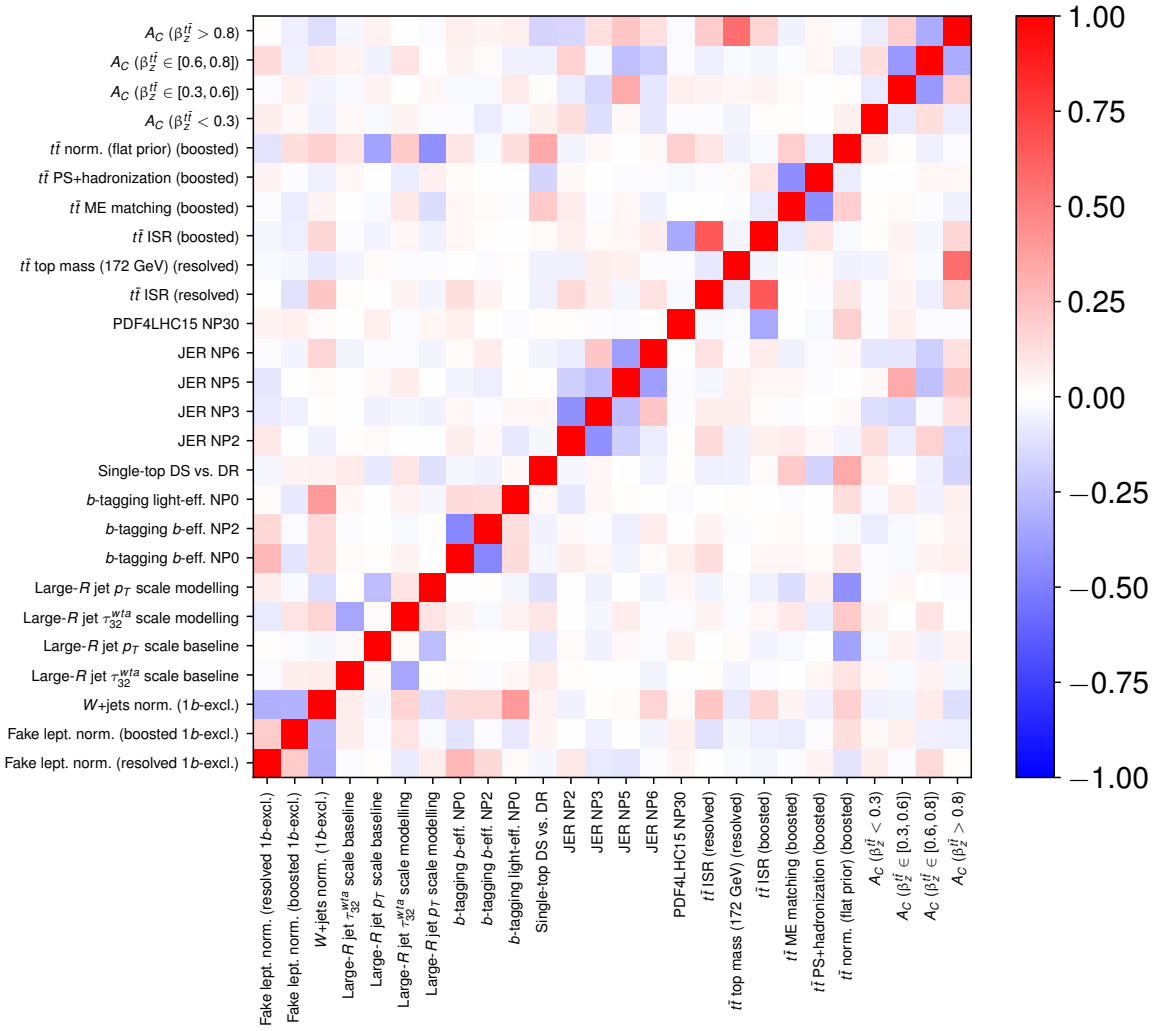


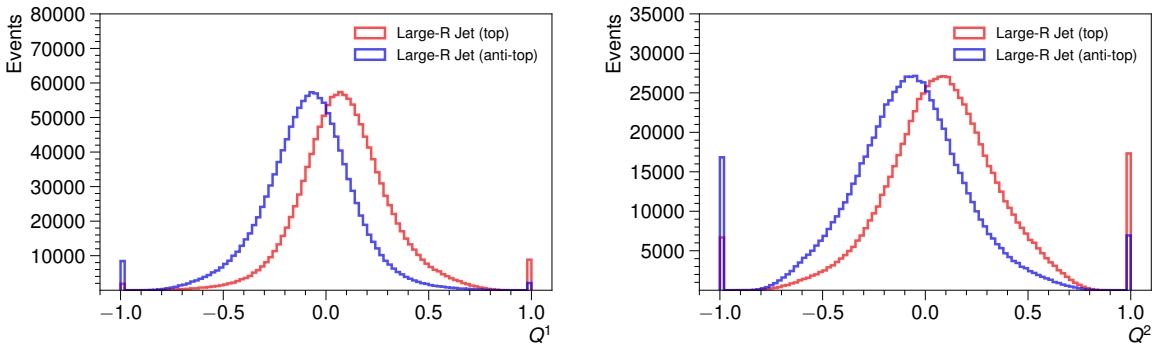
Fig. C.12: The post-marginalisation correlation matrix for NPs, obtained from the unfolding of data in the differential A_C vs $\beta_{z,t\bar{t}}$ measurement. Only correlations greater than 30 % are shown. Additionally, the correlations between A_C and the NPs are also shown.

D

Additional plots for the charge asymmetry in $t\bar{t}$ all-hadronic channel

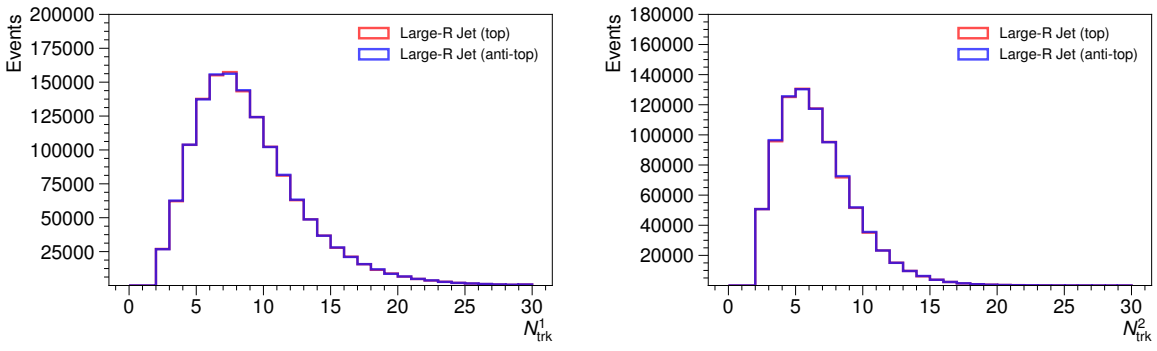
D.1 Neural network input variables

In this section, the distributions of the variables used as inputs for the NN are compared for both large-R jets originating from top quarks (red line in figures) and top-anti-quarks (blue line in figures) in Fig. D.1– D.7.



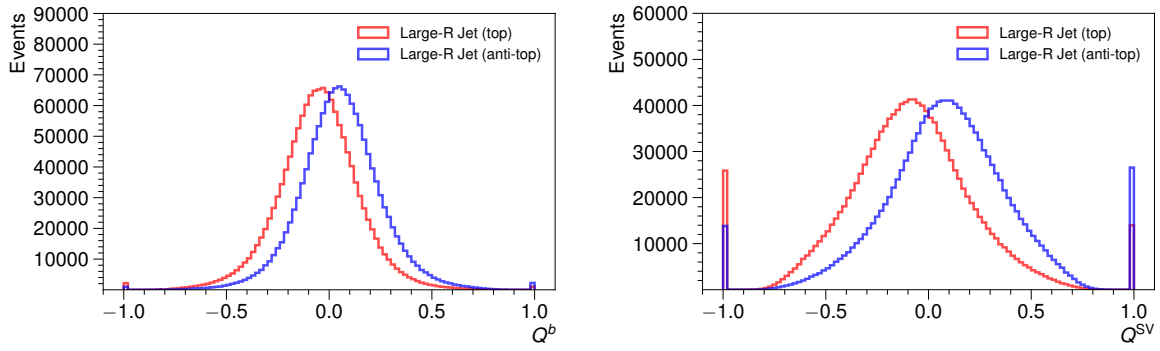
(a) (b)

Fig. D.1: The distribution of the leading (a) and the sub-leading (b) non-b-tagged track-jet charge, Q^1 and Q^2, respectively.



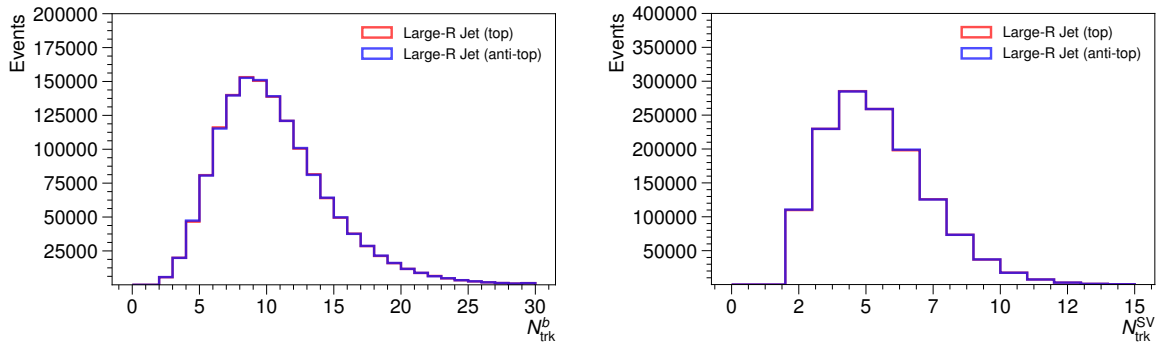
(a) (b)

Fig. D.2: The distribution of the leading (a) and the sub-leading (b) non-b-tagged track-jet track multiplicity, N^1_{trk} and N^2_{trk}, respectively.



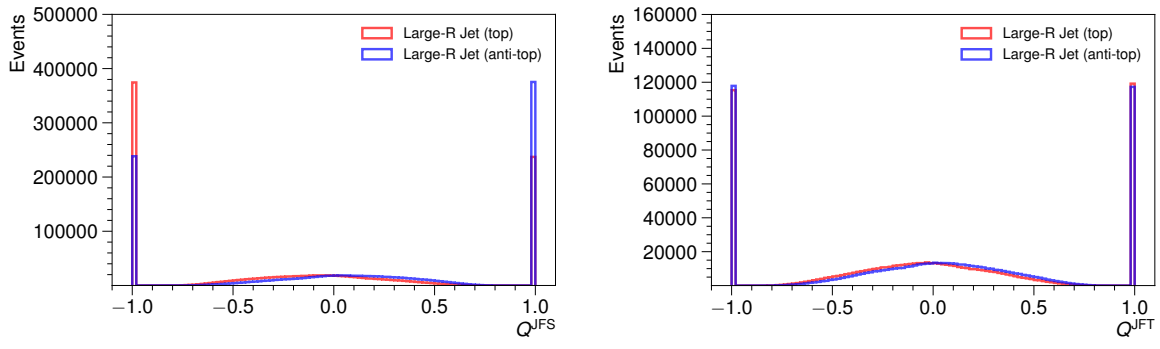
(a) (b)

Fig. D.3: The distribution of the b-tagged track jet charge, Q^b (a) and the b-tagged track jet SV vertex charge, Q^{SV} (b).



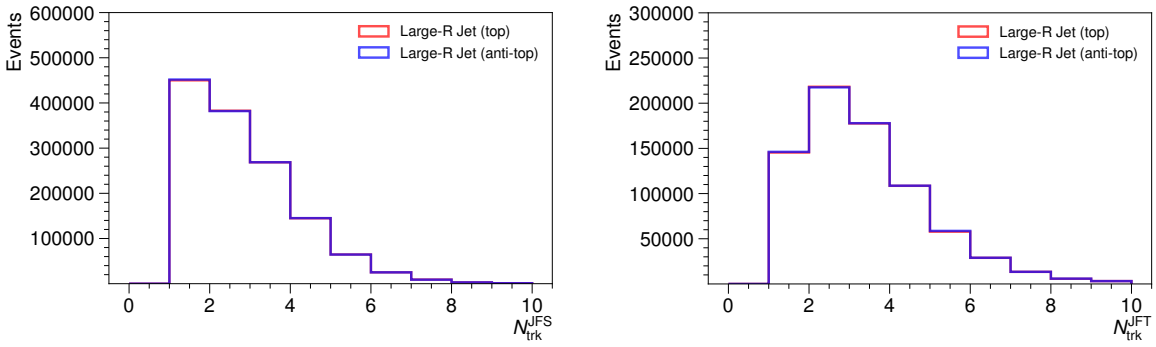
(a) (b)

Fig. D.4: The distribution of the b-tagged track jet track multiplicity, N^b_{trk} (a) and the b-tagged track jet SV vertex track multiplicity, N^{SV}_{trk} (b).



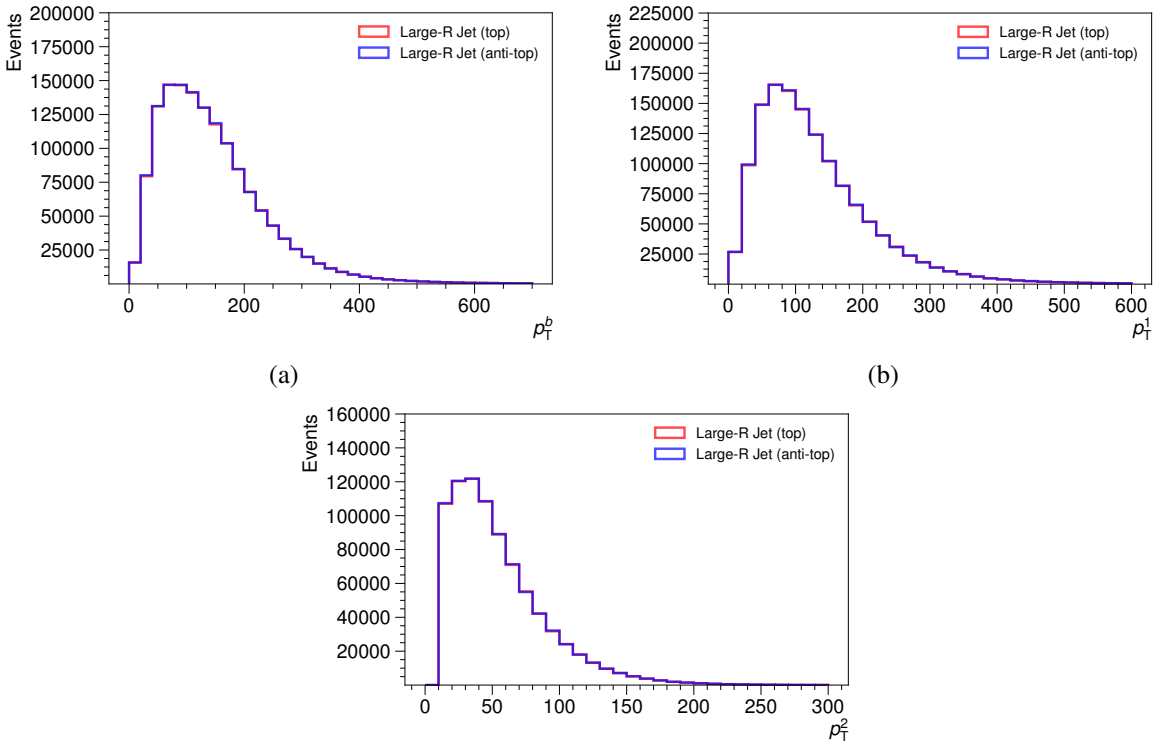
(a) (b)

Fig. D.5: The distribution of the b-tagged track jet JetFitter secondary (a) and tertiary (b) vertex charge, Q^{JFS} and Q^{JFT}, respectively.



(a) (b)

Fig. D.6: The distribution of the b-tagged track jet JetFitter secondary (a) and tertiary (b) vertex track multiplicity, N_{trk}^{JFS} and N_{trk}^{JFT}, respectively.



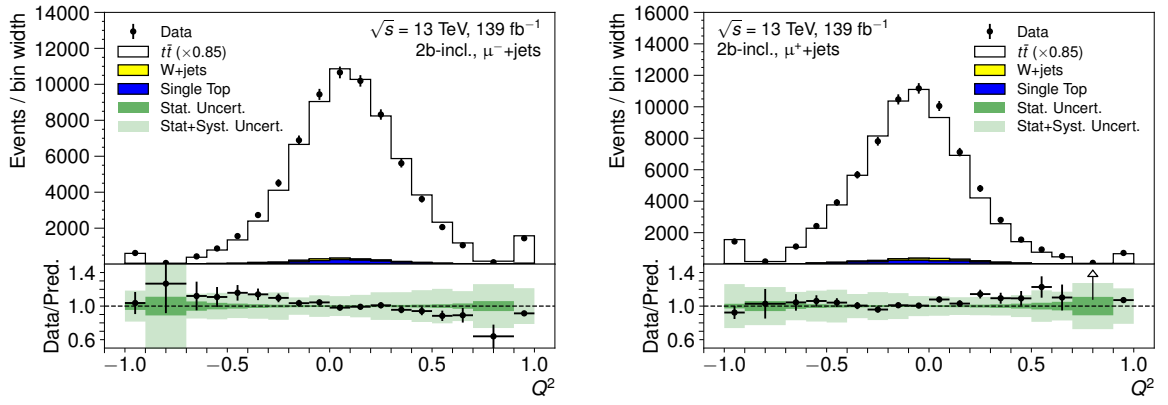
(a) (b)

(c)

Fig. D.7: The distribution of the transverse momentum of the b-tagged track jet (a), and the leading (b) and sub-leading (c) non-b-tagged track jet, p_T^b, p_T^1 and p_T^2, respectively.

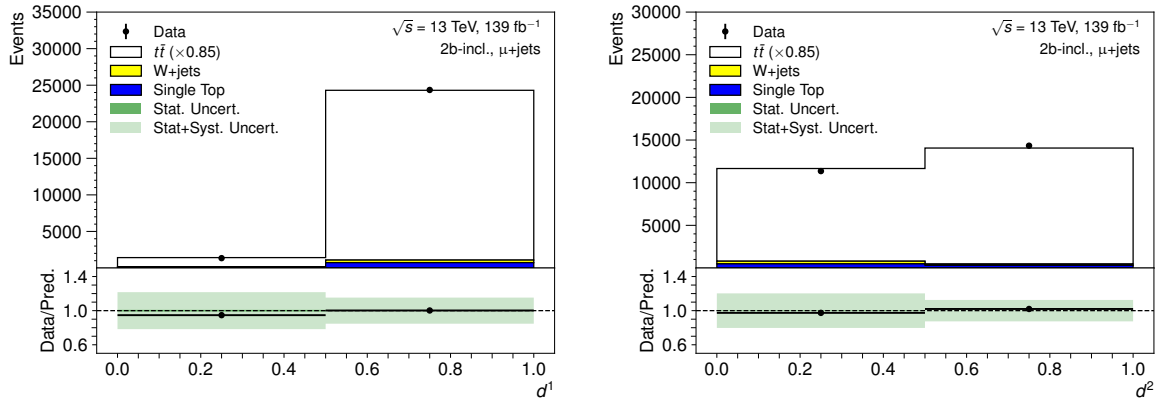
D.2 Comparisons of data and prediction for the neural network in the boosted $t\bar{t}$ single-lepton channel

In Fig. D.8– D.15, the data-to-prediction comparisons are shown for track-jet observables related to the top-quark charge NN. All of the track jets are matched to the hadronic top-quark large-R jet candidate. Charge observables are shown separately for μ^+ and μ^- regions, while auxiliary observables such as p_{T}, N_{trk} and the probability rates of input variables being defined are shown inclusively.



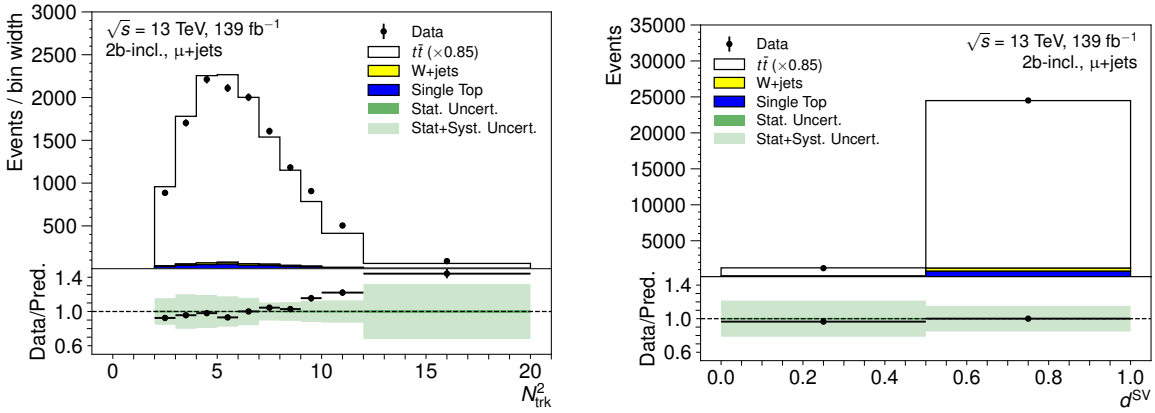
(a) (b)

Fig. D.8: Comparison of data and prediction for the charge of the sub-leading-p_{T} non-b-tagged track jet, Q^2, in the μ^- region (a) and μ^+ region (b). The dark green band shows the statistical uncertainty of the total prediction, and the light green band shows the total prediction statistical and systematic uncertainty summed in quadrature. The $t\bar{t}$ prediction is scaled to normalise the total prediction yield to data. The bin yields are divided by the bin width.



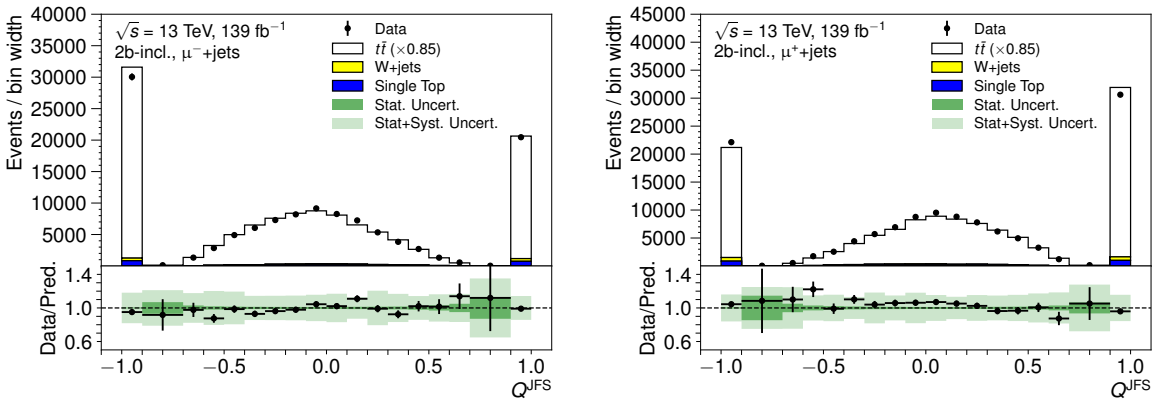
(a) (b)

Fig. D.9: Comparison of data and prediction for the rate of the hadronic top quark large-R jet candidates containing at least one associated non-b-tagged track jet (a) and at least two associated non-b-tagged track jets (b), d^1 and d^2, respectively. The plot style follows the convention of Fig. D.8.



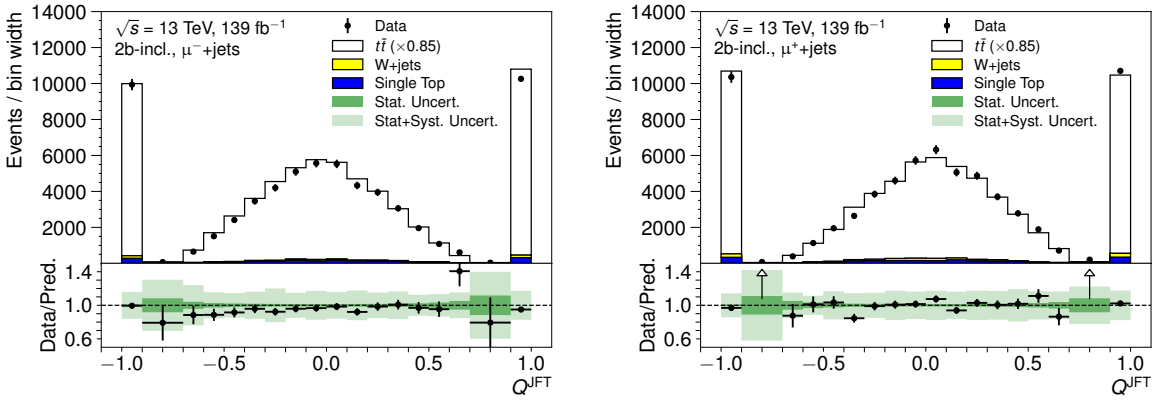
(a) (b)

Fig. D.10: Comparison of data and prediction for the track multiplicity of the sub-leading track jet, N^1_{trk} (a) and the rate of the b-tagged track jet containing a SV vertex, d^{SV} (b). The plot style follows the convention of Fig. D.8. The bin yields in (a) are divided by the bin width.



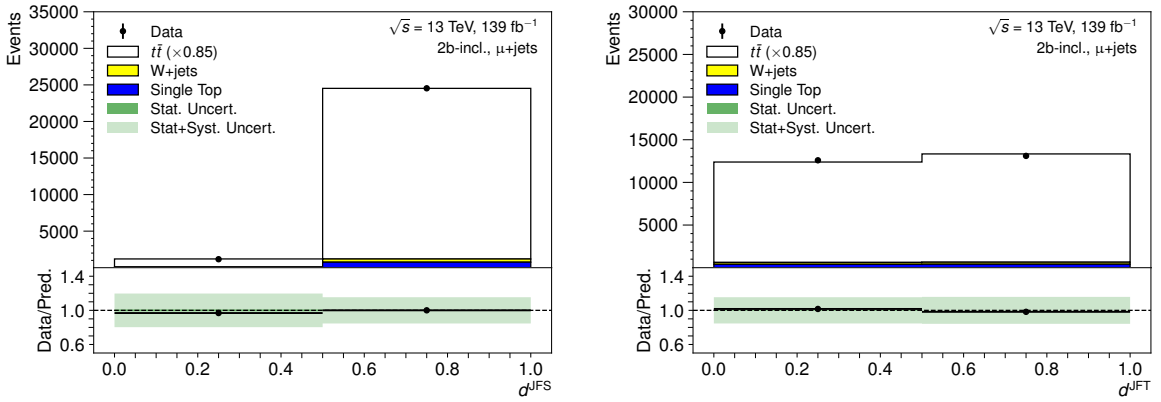
(a) (b)

Fig. D.11: Comparison of data and prediction for the charge of the JF secondary vertex of the b-tagged track jet, Q^{JFS}, in the μ^- region (a) and μ^+ region (b). The plot style follows the convention of Fig. D.8. The bin yields are divided by the bin width.



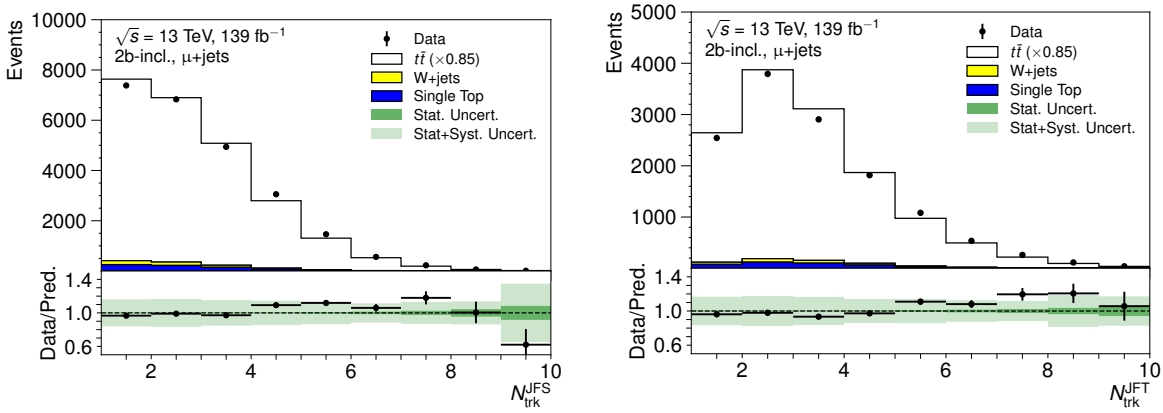
(a) (b)

Fig. D.12: Comparison of data and prediction for the charge of the JF tertiary vertex of the b-tagged track jet, Q^{JFT}, in the μ^- region (a) and μ^+ region (b). The plot style follows the convention of Fig. D.8.



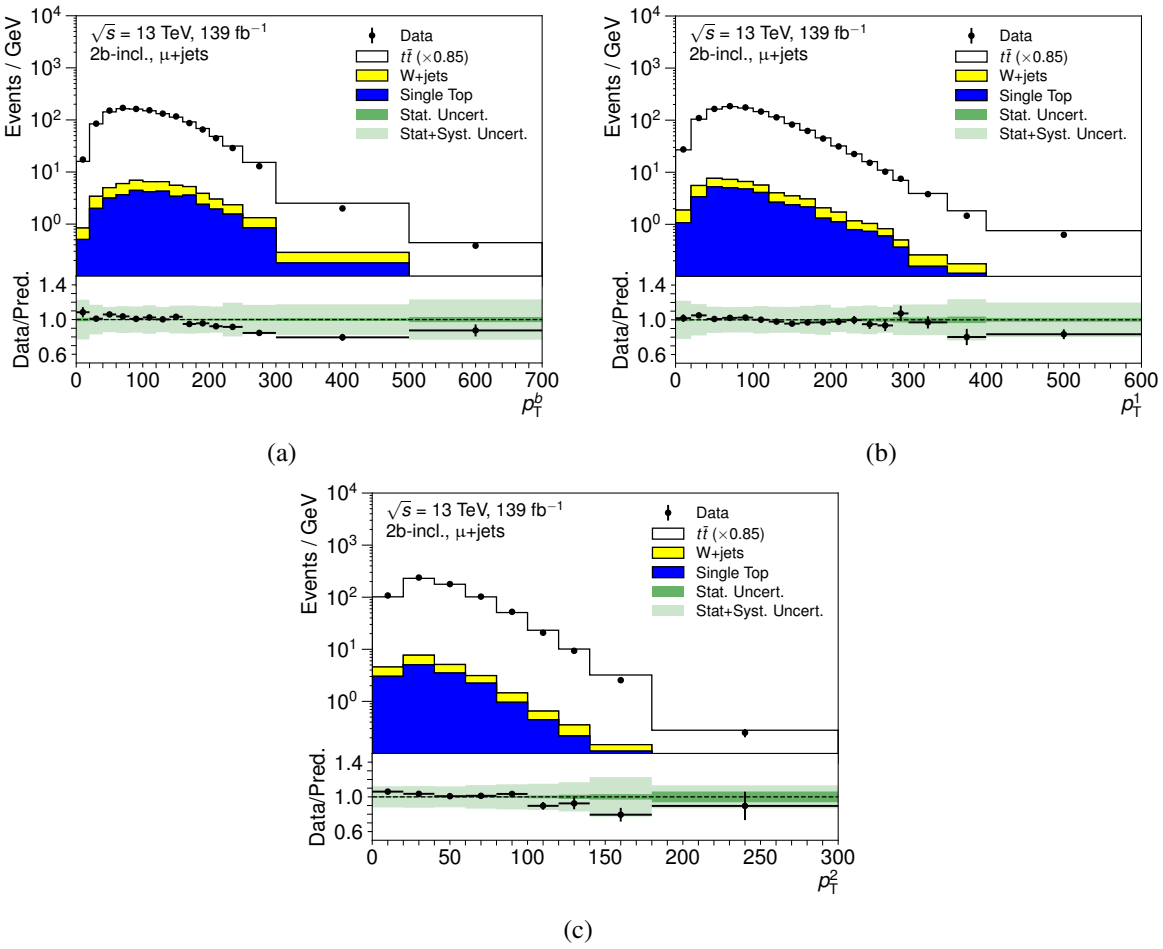
(a) (b)

Fig. D.13: Comparison of data and prediction for the rate of the b-tagged track jet containing a JetFitter secondary vertex (a) and JetFitter tertiary vertex (b), d^{JFS} and d^{JFT}, respectively. The plot style follows the convention of Fig. D.8.



(a) (b)

Fig. D.14: Comparison of data and prediction for the b-tagged track jet JetFitter secondary (c) and tertiary (d) vertex track multiplicities, N_{trk}^{JFS} and N_{trk}^{JFT}, respectively. The plot style follows the convention of Fig. D.8.



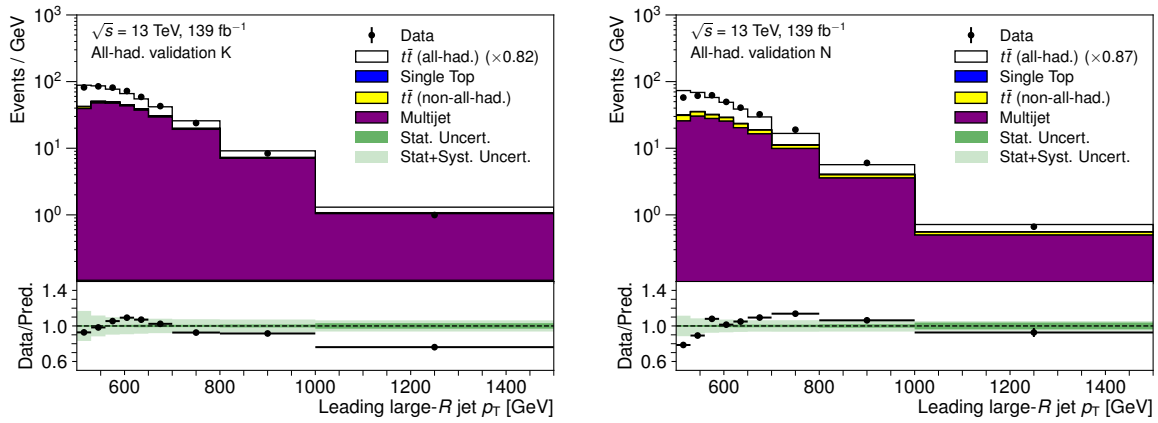
(a) (b)

(c)

Fig. D.15: Comparison of data and prediction for the p_{T} of the b-tagged track jet (a), the leading non-b-tagged track jet (b) and the sub-leading non-b-tagged track jet (c), p_{T}^{b}, p_{T}^{1} and p_{T}^{2}, respectively. The plot style follows the convention of Fig. D.8. The bin yields are divided by the bin width.

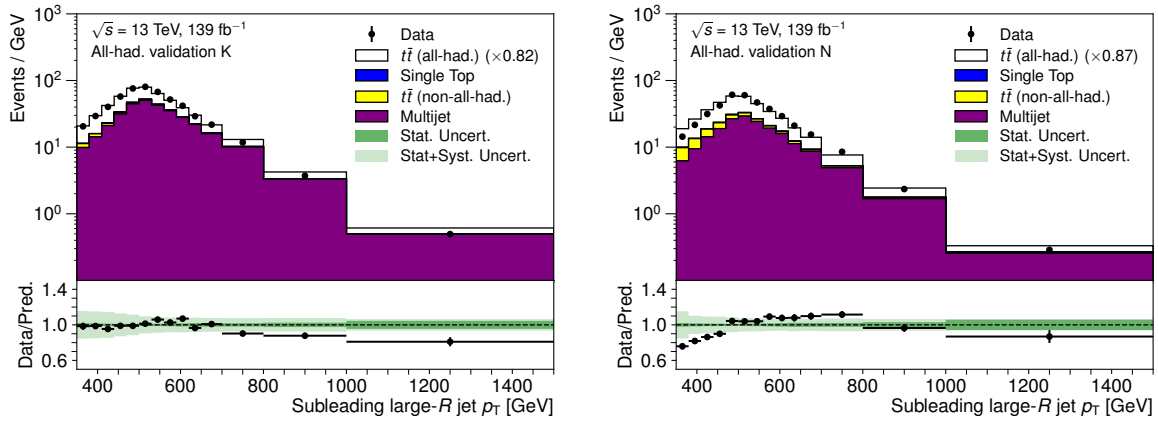
D.3 Comparisons of data and prediction in the all-hadronic channel

In Fig. D.16– D.19, additional comparisons of data and prediction are shown for the leading and sub-leading large-R jet p_{T} and η distributions in the validation regions. In Fig. D.20 and D.21, the leading and sub-leading large-R jet p_{T} and η comparisons of data to prediction are shown in the signal region. In all of the figures, both statistical and systematic uncertainties described in Sec. 8.5.4 are included in the ratio plots. In all of the figures, the signal contribution is scaled to match the prediction yield with data.



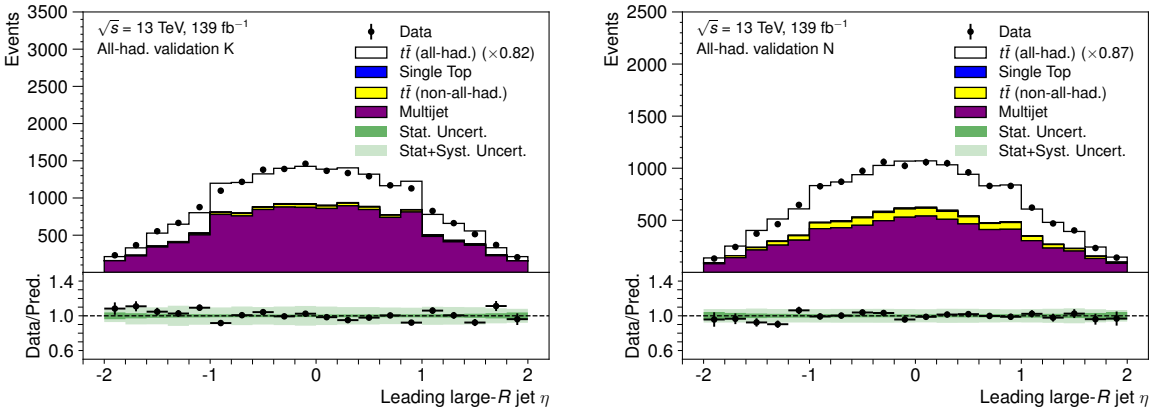
(a) (b)

Fig. D.16: The comparison of data to prediction of the leading large-R jet p_{T} distribution, in the validation regions K (a) and N (b). The all-hadronic $t\bar{t}$ contribution is scaled to match the prediction to data. The bin yields are divided by the bin width.



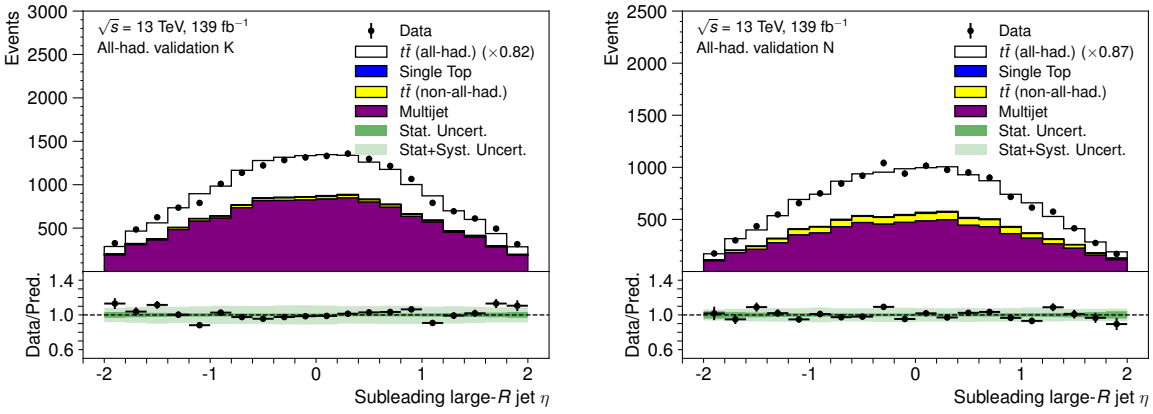
(a) (b)

Fig. D.17: The comparison of data to prediction of the sub-leading large-R jet p_{T} distribution, in the validation regions K (a) and N (b). The plot follows the convention of Fig. D.16. The bin yields are divided by the bin width.



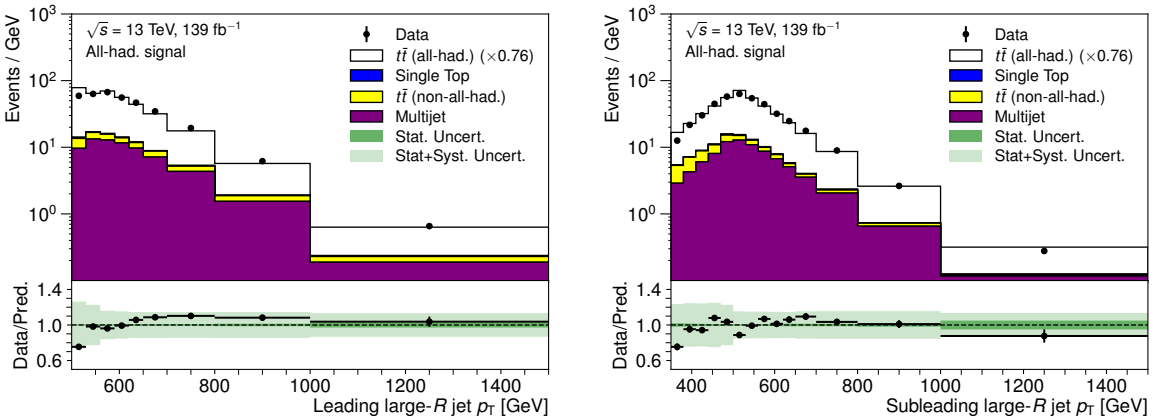
(a) (b)

Fig. D.18: The comparison of data to prediction of the leading large-R jet η distribution, in the validation regions K (a) and N (b). The plot follows the convention of Fig. D.16.



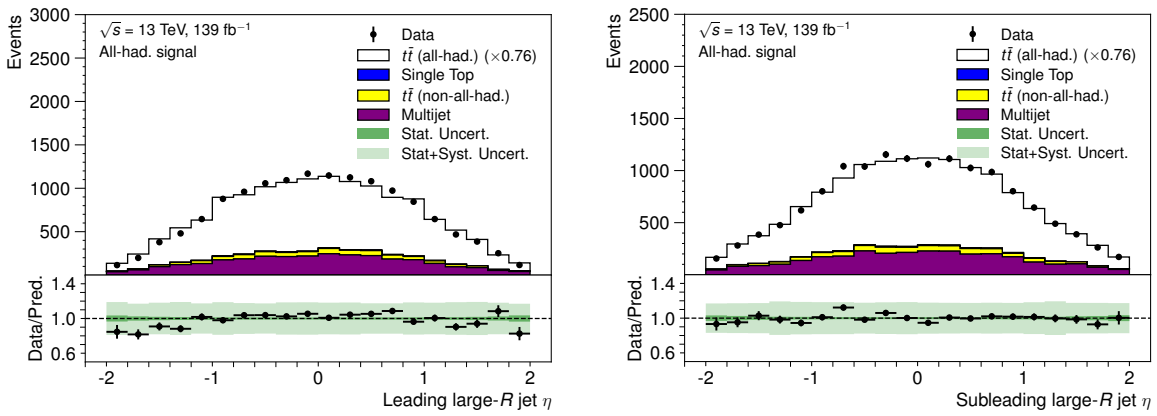
(a) (b)

Fig. D.19: The comparison of data to prediction of the sub-leading large-R jet η distribution, in the validation regions K (a) and N (b). The plot follows the convention of Fig. D.16.



(a) (b)

Fig. D.20: The comparison of data to prediction of the leading (a) and sub-leading (b) large-R jet p_{T} distribution, in the signal region. The plot follows the convention of Fig. D.16. The bin yields are divided by the bin width.



(a) (b)

Fig. D.21: The comparison of data to prediction of the leading (a) and sub-leading (b) large-R jet η distribution, in the signal region. The plot follows the convention of Fig. D.16.