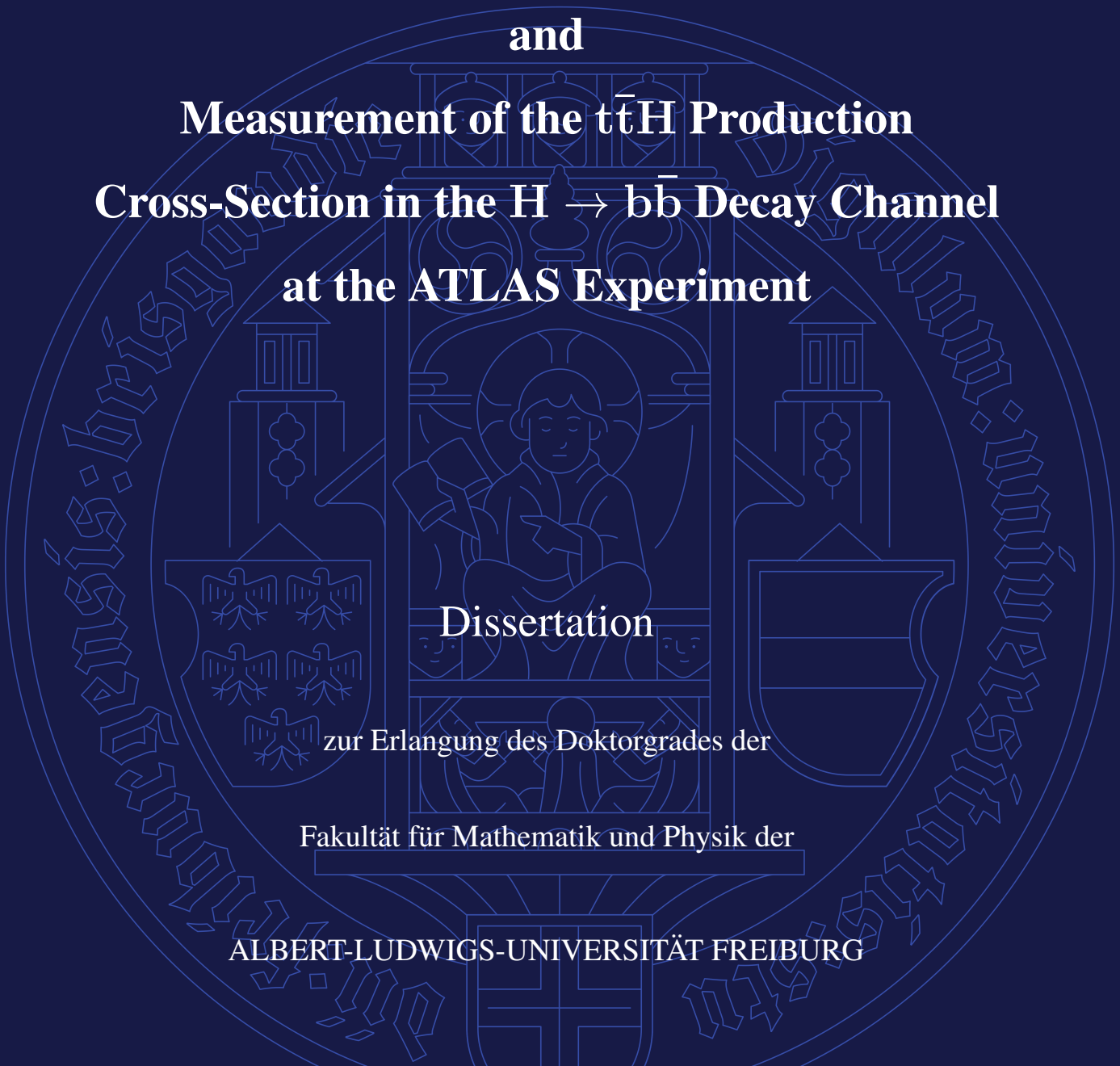
Heavy-Flavour Tagging and Measurement of the $\mathrm{t\bar{t}H}$ Production Cross-Section in the $\mathrm{H} \to \mathrm{b\bar{b}}$ Decay Channel at the ATLAS Experiment



Dissertation

zur Erlangung des Doktorgrades der

Fakultät für Mathematik und Physik der

ALBERT-LUDWIGS-UNIVERSITÄT FREIBURG

vorgelegt von

Alexander Froch

Betreuer:

Priv.-Doz. Dr. Andrea Helen Knue (Technische Universität Dortmund)

Prof. Dr. Gregor Herten (Albert-Ludwigs-Universität Freiburg)

CERN-THESIS-2024-213
16/05/2024

Heavy-Flavour Tagging and Measurement of the $\mathrm{t\bar{t}H}$ Production Cross-Section in the $\mathrm{H} \to \mathrm{b\bar{b}}$ Decay Channel at the ATLAS Experiment

Dissertation

zur Erlangung des Doktorgrades der

Fakultät für Mathematik und Physik der

ALBERT-LUDWIGS-UNIVERSITÄT FREIBURG

vorgelegt von

Alexander Froch

Betreuer:

Priv.-Doz. Dr. Andrea Helen Knue (Technische Universität Dortmund)

Prof. Dr. Gregor Herten (Albert-Ludwigs-Universität Freiburg)

Datum der mündlichen Prüfung: 16.05.2024

DEKAN:	Prof. Dr. Michael Růžička (Albert-Ludwigs-Universität Freiburg)
BETREUER DER ARBEIT:	Priv.-Doz. Dr. Andrea Helen Knue (Technische Universität Dortmund)
	Prof. Dr. Gregor Herten (Albert-Ludwigs-Universität Freiburg)
GUTACHTER:	Priv.-Doz. Dr. Andrea Helen Knue (Technische Universität Dortmund)
	Prof. Dr. Karl Jakobs (Albert-Ludwigs-Universität Freiburg)
PRÜFER:	Priv.-Doz. Dr. Andrea Helen Knue (Technische Universität Dortmund)
	Prof. Dr. Markus Schumacher (Albert-Ludwigs-Universität Freiburg)
	Prof. Dr. Stefan Dittmaier (Albert-Ludwigs-Universität Freiburg)

ABSTRACT

With the final missing piece of the Standard Model of particle physics (SM), the Higgs boson, discovered in 2012 by the ATLAS and CMS experiments at the Large Hadron Collider (LHC) at CERN, both experiments have shifted their focus towards a precise determination of the properties of the new boson. The large datasets collected by both experiments during the second run of the LHC, Run 2, enable such precision measurements of the Higgs boson properties. One of the properties under study is the Yukawa coupling of the top quark, the heaviest of all particles in the SM, to the Higgs boson. The top-Higgs Yukawa coupling, the largest in the SM, is varied in many postulated theoretical extensions of the SM and is sensitive to the effects and the possible presence of physics Beyond the SM (BSM). The coupling itself can be measured in an indirect way, e.g using $H \to \gamma\gamma$ measurements in which the coupling is involved via virtual loops, and in a direct way, which can be measured in the rare process where a Higgs boson is produced in association with a pair of top quarks, $t\bar{t}H$. Exploiting the dominant decay mode of the Higgs boson into a pair of b-quarks, $t\bar{t}H(H \to b\bar{b})$, this thesis investigates the direct measurement of the top-Higgs Yukawa coupling in a challenging yet promising final state with at least four b-quark initiated jets.

The $t\bar{t}H(H \to b\bar{b})$ legacy analysis presented in this thesis is part of the re-analysis of the complete Run 2 dataset of $140\,\text{fb}^{-1}$ at $\sqrt{s} = 13\,\text{TeV}$ recorded by the ATLAS experiment between 2015 and 2018. The focus lies on enhancing and integrating the latest physics object reconstruction algorithms, as well as employing an improved modelling of the main background process, $t\bar{t} + b\bar{b}$. Alongside state-of-the-art classification and reconstruction neural networks, a new estimate for the fake-lepton contribution in the single-lepton channel is derived. Further, the derivation of data-driven correction factors for the mis-modelled H_{T}, observed in various $t\bar{t} + \text{jets}$ processes, are presented. Finally, studies of the performance of profile likelihood fits as employed in the full analysis are presented with blinded data, yielding an expected significance of the $t\bar{t}H$ process at $5.5\,\sigma$.

Effective separation of the signal from the expected background requires dedicated and high-performing b-jet identification algorithms, the so-called b-taggers. In this thesis, the development and integration of a novel track-based b-tagger, DIPS, is presented. In addition, the DIPS tagger was added as a component to the ATLAS high-level b-tagger DL1r, replacing the previous track-based b-tagger, RNNIP. The resulting version, DL1d, is the new recommended b-tagger in ATLAS, thanks to its significantly improved performance compared to its predecessor. Furthermore, an extension of the DIPS b-tagger is introduced in this thesis, which is referred to as DIPS Tau which incorporates a new jet-class: τ-jets.

KURZZUSAMMENFASSUNG

Mit der Entdeckung des letzten fehlenden Teils des Standardmodells der Teilchenphysik (SM), dem Higgs-Boson, das 2012 von den Experimenten ATLAS und CMS am Large Hadron Collider (LHC) am CERN entdeckt wurde, haben beide Experimente ihren Schwerpunkt auf die genaue Bestimmung der Eigenschaften des neuen Bosons gelegt. Die von beiden Experimenten während des zweiten Laufs des LHCs, Run 2, gesammelten großen Datensätze ermöglichen Präzisionsmessungen der Eigenschaften des Higgs-Bosons. Eine der untersuchten Eigenschaften ist die Yukawa-Kopplung des Top-Quarks, des schwersten Teilchens im SM, an das Higgs-Boson. Die Top-Higgs-Yukawa-Kopplung, die größte im SM, wird in vielen postulierten theoretischen Erweiterungen des SM variiert und ist sensitiv auf Effekte und das mögliche Vorhandensein von Physik jenseits des SM (BSM). Die Kopplung selbst kann auf indirekte Weise gemessen werden, z.B. mit $H \to \gamma\gamma$ Messungen an denen die Kopplung über virtuelle Schleifen beteiligt ist, und auf direktem Weg, wo sie messbar ist im seltenen Produktionsprozess eines Higgs-Bosons in Verbindung mit einem Top-Quark Paar, $t\bar{t}H$. Unter Verwendung des dominanten Zerfalls des Higgs-Bosons in ein Paar von b-Quarks, $t\bar{t}H(H \to b\bar{b})$, untersucht diese Arbeit die direkte Messung der Top-Higgs Yukawa Kopplung in einem herausfordernden, aber vielversprechenden Endzustand mit mindestens vier durch b-Quarks initiierte Jets.

Die in dieser Arbeit vorgestellte $t\bar{t}H(H \to b\bar{b})$-Legacy-Analyse ist Teil der erneuten Analyse des kompletten Run 2-Datensatzes mit $140\,\text{fb}^{-1}$ bei $\sqrt{s} = 13\,\text{TeV}$, welcher vom ATLAS-Experiment zwischen den Jahren 2015 bis 2018 aufgezeichnet wurde. Der Fokus der Analyse liegt auf Verbesserung und Integration neuster Algorithmen zur Rekonstruktion von physikalischen Objekten und Anwendung einer verbesserten Modellierung vom Hauptuntergruprozess $t\bar{t} + b\bar{b}$. Neben State-of-the-Art neuronale Netzwerke zur Klassifizierung und Rekonstruktion von Ereignissen wurde ebenfalls eine neue Schätzung für den Fake-Lepton-Beitrag im Single-Lepton-Kanal entwickelt. Desweiteren wird die Entwicklung datengetriebenen Korrekturfaktoren für das, in einigen $t\bar{t} + \text{jets}$ Prozessen fehlerhaft modellierte, H_T präsentiert. Abschließend werden die Ergebnisse von Profile Likelihood Fits, wie sie in der vollständigen Analyse verwendet werden, mit blinden Daten vorgestellt, welche eine erwartete Signifikanz des $t\bar{t}H$-Prozesses von $5,5\,\sigma$ liefern.

Eine effektive Separation des Signals vom erwarteten Untergrund erfordert spezielle und leistungsstarke Algorithmen zur Identifizierung von b-Jets, die sogenannten b-Tagger. In dieser Arbeit wird die Entwicklung und Integration eines neuartigen spurbasierten b-Taggers, DIPS, vorgestellt. Darüber hinaus wurde der DIPS-Tagger als Komponente in den ATLAS-High-Level b-Tagger DL1r integriert, wo er den bisherigen spurbasierten b-Tagger RNNIP ersetzt. Die daraus resultierende Version, DL1d, ist der neue empfohlene b-Tagger in ATLAS, dank seine deutlich verbesserte Leistung im Vergleich zu seinem Vorgänger. Zudem wird eine Erweiterung des DIPS-Taggers vorgestellt, genannt DIPS Tau, welche eine zusätzliche Jet-Klasse einführt: τ-jets.

CONTRIBUTIONS BY THE AUTHOR

As a member of the ATLAS collaboration, the author of this thesis took part in multiple projects from which the most significant contributions, which are topics in this thesis, are listed below.

HEAVY-FLAVOUR TAGGING The work and results presented from Section 7.2 to Chapter 10 were nearly completely done by the author of this thesis. For the development of the full preprocessing chain for the training samples, including the development of a new resampling method, and the underlying framework for training, optimisation and visualisation, the author assumed a leading role. The presented results concerning preprocessing, training, and visualisation of the DIPS, DL1d, and DIPS Tau algorithms were exclusively done by the author. To effectively prepare the training samples, train the different taggers, and visualise their performance, different software packages where developed, including an all-in-one framework: UMAMI. From this all-in-one framework, multiple smaller packages were developed, e.g. the plotting package PUMA [1], for which the author assumed a leading role in their development. Both packages are the standards for training and plotting the results of the different mentioned taggers inside the Flavour-Tagging group of ATLAS.
Additionally to the provided software, the studies and developments on DIPS performed by the author were documented in *public plots* [2] and lead to the development of DL1d, the new recommended ATLAS b-tagging algorithm for Run 3, for which the author also assumed a leading role in development. Furthermore, the addition of the τ-jets into the training of the DIPS tagger showed great performance improvements, which lead to the usage of τ-jets in all trainings conducted by the Flavour-Tagging group with the new successor algorithms to DL1d, GN1 and GN2 [3, 4].
The author also took on important roles within the Flavour-Tagging group, preparing and giving tutorials within ATLAS about the developed software and taggers as well as being the liaison between the ATLAS Machine Learning forum and the Flavour-Tagging group. Further, the author also co-supervised a master student working on the addition of the bb-jets class into the mentioned algorithms.

$t\bar{t}H(H \rightarrow b\bar{b})$ ANALYSIS In addition to his work on heavy-flavour tagging, the author of this thesis is also one of the main analysers in the $t\bar{t}H(H \rightarrow b\bar{b})$ legacy analysis, presented in Chapters 11–14, which is currently ongoing. The author contributed to several aspects of this analysis, performing various studies to estimate the fake-lepton contributions in the single-lepton channel (Chapter 12), which include the implementation and derivation of the real and fake efficiencies for the electrons and muons for the matrix method. The derived fake efficiencies are also used in the fake factor method to estimate the systematic uncertainties. Furthermore, the application of the

matrix method estimation was performed by the author in addition to sanity checks to validate the results obtained using the matrix method.

The author also took part in the development of the data-driven H_{T} reweighting for both the single-lepton and dilepton channels (Chapter 13). In both channels, a significant mis-modelling of the H_{T} variable was observed for the $t\bar{t} + \mathrm{jets}$ background. The author derived and implemented the H_{T} reweightings for both channels for the $t\bar{t} + \geq 1c$ and $t\bar{t} +$ light processes using a dedicated, orthogonal control region and validated the effect of these reweightings.

Furthermore, the author took part in performing different fit studies, including the addition and testing of several improved systematic uncertainties. He also performed the baseline fits, provided to the other analysers, which are presented in Chapter 14 showing the current status of the analysis at the time of writing.

CONTENTS

1 INTRODUCTION

Since the postulation of the Higgs mechanism [5–7] and the discovery of its massive scalar boson, the Higgs boson [8], in 2012 by the ATLAS [9] and CMS [10] experiments at the Large Hadron Collider (LHC) [11] near Geneva, Switzerland, particle physicists all over the world are searching for the next major clue in the puzzle of High Energy Physics (HEP). Making use of the largest and most powerful particle collider in the world, these physicists are testing and probing one of the most successful physics theories ever developed: the Standard Model (SM) of particle physics. With the last missing piece, the Higgs boson, found, the conducted tests and probes of the SM aim at finding tiny deviations from its predictions, hinting possible phenomena of physics beyond the SM. To search for these deviations, physicists need to probe different SM predictions with unprecedented precision.

One of these many places where such a deviation from the prediction could emerge is the Yukawa coupling [12] between the Higgs boson and top-quarks (t-quarks), the heaviest elementary particles of the SM. With a value of ≈ 1, the t-quark has the largest Yukawa coupling of all fermions in the SM, making it ideally suited as a probe for physics Beyond Standard Model (BSM). Furthermore, many suggested theoretical extensions of the SM postulate a variation in the values of the Yukawa couplings [13]. Measurement of the top-Higgs Yukawa coupling is possible in an indirect way, e.g using $H \to \gamma\gamma$ measurements in which the coupling is involved via virtual loops, or in a direct way. Among the direct measurement methods, one significant process is the production of a Higgs boson in association with a t-quark pair ($t\bar{t}H$). However, it is crucial to note that this process is exceedingly rare when compared with other Higgs production modes and other processes taking place in proton-proton collisions.

To reject the large expected background contributions and isolate the signal process, dedicated physics object reconstruction and identification techniques are needed, especially making use of one of the t-quark decay products: b-quarks. Due to its unique decay signature, ATLAS developed dedicated algorithms to identify jets initiated by b-quarks, the so-called heavy-flavour tagging, or simply b-tagging. In this thesis, these specialised algorithms are introduced and developed further, improving their performance. The presented improvements have lead to a new b-tagging algorithm, which is recommended for use in ATLAS analyses for LHC Run 3: The Deep Learning 1 DIPS (DL1d) tagger.

The analysis regarding $t\bar{t}H$ focuses on the Higgs decay channel into two b-quarks ($H \to b\bar{b}$) and is part of the re-analyses focusing on this particular channel. This so-called legacy analysis uses the full LHC Run 2 dataset with an integrated luminosity of $140\,\text{fb}^{-1}$ of proton-proton collisions recorded by the ATLAS experiment between 2015 and 2018 at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$. Though the previous iteration of the analysis [14] has already been carried out using the full LHC Run 2 dataset, the legacy analysis provides an improved modelling of important background processes as well

as updated tools for identification and reconstruction of the physics objects involved. Additionally, new and dedicated event reconstruction and classification algorithms are used to better separate signal from background events. Similar to the previous analysis, inclusive and differential cross-section measurements are anticipated, making use of the Simplified Template Cross-Section (STXS) framework [15, 16].

This dissertation is structured as follows. Part I gives an overview of the SM of particle physics, the LHC, the ATLAS detector, and the event simulation of proton-proton collisions. In Part II, the necessary methodology is summarised and discussed, including the reconstruction of physics objects at the ATLAS detector, as well as the Machine Learning (ML) techniques employed in this thesis. Part III introduces the methods and algorithms used in ATLAS for the identification of b-quark initiated jets during Run 2 of the LHC. These techniques have been improved for LHC Run 3 as part of this dissertation, and these developments and the observed gain in performance are presented as well. Part IV covers the $t\bar{t}H(H \rightarrow b\bar{b})$ legacy analysis, including a review of the previous measurement, the analysis strategy, and improvements of background estimation and event reconstruction techniques. Finally, the blinded results are interpreted statistically using profile likelihood fits.

Part I

OVERVIEW

2

THE STANDARD MODEL OF PARTICLE PHYSICS

The SM of particle physics [17] is arguably one of the most thriving theories to date. It offers a comprehensive and elegant framework for understanding the fundamental constituents of our known universe and the forces that govern their behaviour. Upon closer examination, the SM is a Quantum Field Theory (QFT) that describes three of the four known fundamental forces in the universe. These three forces comprise the electromagnetic, weak, and strong forces. The fourth, the gravitational force, is not included in the SM. While gravity holds significance in our everyday lives, its strength, relative to the other three forces, is 37 orders of magnitude smaller. Hence, at the subatomic level of elementary particles, it can be neglected. In addition to the three fundamental forces, the SM also contains so-called elementary particles, which are point-like particles. All known matter in the universe (excluding dark matter) is constructed from a subgroup of these elementary particles known as fermions. Fermions are characterised by an intrinsic spin of $\frac{1}{2}$. In addition to fermions, the SM also includes force carriers, the gauge bosons, with an intrinsic spin of 1. A schematic overview of the various elementary particles described in the SM is provided in Figure 2.1. One of the long-standing challenges in particle physics was the determination of the origin of the mass of these particles. While some particles in the SM are massless, others possess mass, a feature which was originally not explained in the description of the SM. In 1964, Peter Higgs and the team of François Englert and Robert Brout proposed a solution to this issue, now known as the Higgs mechanism[1] [5–7]. From a mathematical perspective, the SM is a renormalisable gauge QFT. The symmetry of this QFT can be described by the gauge group, as depicted in Equation (2.1).

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y \tag{2.1}$$

Here, the $SU(3)_C$ symmetry group describes the strong force, which interacts with all colour-charged particles like quarks and gluons. The unification of the electromagnetic and weak forces, known as the Electroweak (EW) force, is described by the $SU(2)_L \otimes U(1)_Y$ term. The $SU(2)_L$ symmetry group interacts solely with left-handed particles, carrying weak isospin, while the $U(1)_Y$ interacts only with particles carrying weak hypercharge. To maintain the gauge invariance of the Lagrangian in the SM (as it is a gauge theory), new vector or gauge fields must be included, and their excitations correspond to the aforementioned gauge bosons. To account for particle masses, the EW symmetry is spontaneously broken through the Higgs mechanism, which not only results in particle masses but also in the emergence of a new boson with spin 0, known as the Higgs boson. This Higgs boson holds significant importance for this thesis and was discovered in 2012 by the ATLAS and CMS experiments [9, 10] at the LHC [11] at Conseil Européen pour la Recherche Nucléaire (CERN)

1 One needs to note here that multiple people proposed the spontaneous symmetry breaking [18], but Higgs was the only one postulating a massive scalar boson and also describing its properties [8].

near Geneva, Switzerland.

In the following chapter, a more detailed explanation of the elementary particles and the three fundamental forces described in the SM will be provided. Additionally, the Higgs mechanism will be examined in greater detail, given its importance for this thesis. To conclude, a brief summary of some of the current problems and limitations of the SM will be presented.

2.1 ELEMENTARY PARTICLES OF THE STANDARD MODEL

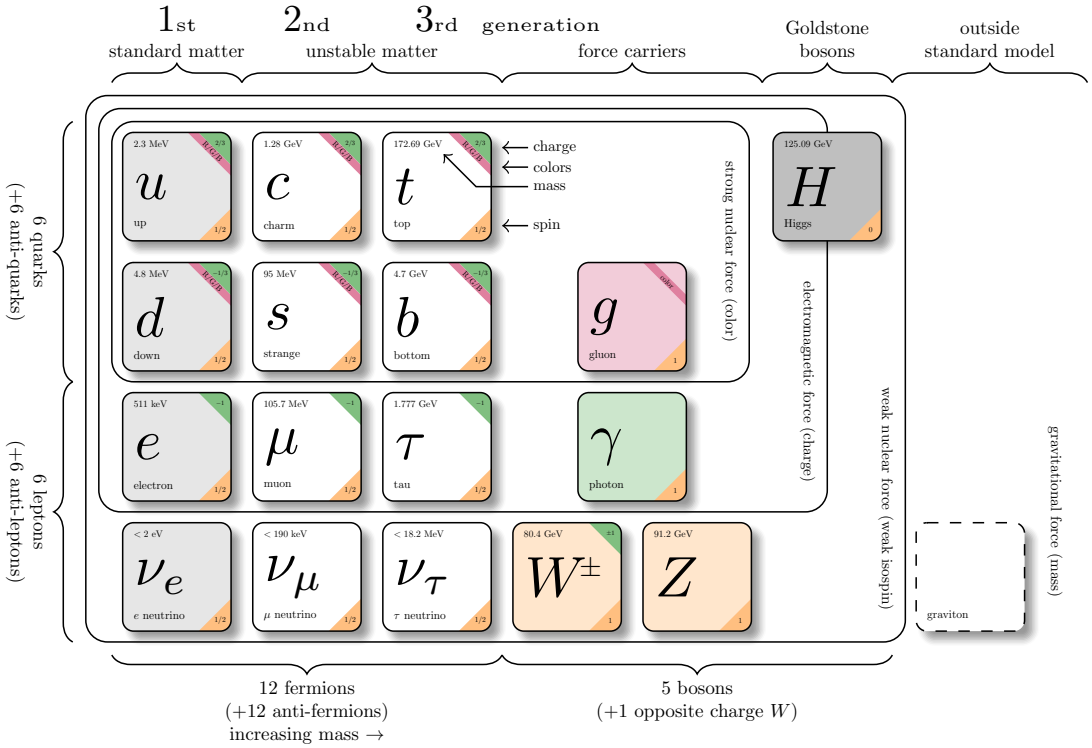


Figure 2.1.: Schematic view of the elementary particles in the SM. Adapted from [19].

2.1.1 *Fermions*

Fermions, and their corresponding anti-particles, are spin $\frac{1}{2}$ particles that make up all the visible matter in the universe. The 12 fermions (and 12 anti-fermions) of the SM are split into 2 groups, depending on whether they have colour charge or not. These two groups are called quarks [20, 21] and leptons, respectively. While the quarks have a colour charge and, therefore, interact via the strong force, the leptons do not carry colour charge, which excludes them from taking part in the strong interaction. Furthermore, the fermions are divided into three so-called generations or families and are ordered by their masses, with the first generation being the lightest and the third generation being

the heaviest. For both the quarks and the leptons, each generation has one up and one down-type particle. These two particles per generation per group form $SU(2)_L$ doublets.

Taking a closer look at the $SU(2)_L$ doublets, the up and down-type fermions have special properties, including the weak isospin (I), the third component of this weak isospin I_3, the hypercharge Y, and the electric charge Q. The different fermions and their respective quantum numbers are listed in Table 2.1, excluding all fermions, which do not interact with the SM.

Table 2.1.: Fermions who interact with the SM with their quantum numbers [22].

Fermion type	I	II	III	I	I_3	Y	Q
Quarks	u_L	c_L	t_L	$1/2$	$1/2$	$1/3$	$2/3$
	d_L	s_L	b_L	$1/2$	$-1/2$	$1/3$	$-1/3$
	u_R	c_R	t_R	0	0	$4/3$	$2/3$
	d_R	s_R	b_R	0	0	$-2/3$	$-1/3$
Leptons	$\nu_{e,L}$	$\nu_{\mu,L}$	$\nu_{\tau,L}$	$1/2$	$1/2$	-1	0
	e_L	μ_L	τ_L	$1/2$	$-1/2$	-1	-1
	e_R	μ_R	τ_R	0	0	-2	-1

In general, all (same-handed) up and down-type quarks and leptons have the same quantum numbers, respectively. While for all quarks their left and right-handed particles are listed, for the leptons, the right-handed up-type particles, the right-handed neutrinos, are excluded, due to not interacting with other particles of the SM.

2.1.2 *Gauge Bosons*

The gauge bosons are the force mediator particles of the SM. Carrying a spin of 1, each interaction has its own mediator particle(s). While the electromagnetic force is mediated via the massless photon and the strong force via the massless gluon(s), the weak force has two mediator particles, the $W^{\pm}$ [23, 24] and Z^0 [25, 26] bosons. In contrast to the other mediator particles, the $W^{\pm}$ and Z^0 bosons are massive with a mass of $m_{W^{\pm}} = 80.377 \pm 0.012\,\text{GeV}$ and $m_{Z^0} = 91.1876 \pm 0.0021\,\text{GeV}$ [27][2]. Due to these massive mediator particles, the range of the weak force is limited to short distances, while the electromagnetic force, due to its massless mediator particle, the photon, is long-ranged. One could assume similar properties for the strong force, but due to the gluons' colour charge and their self-interactions, the strong force is also short-ranged.

2.2 THE THREE FUNDAMENTAL FORCES

As mentioned in previous sections, the SM only encompasses three of the four known fundamental forces of nature. While the SM's Lagrangian must remain invariant under global and local gauge

2 The origin of this mass will be discussed in Section 2.3

transformations, the corresponding symmetries further constrain the Lagrangian and contribute to the development of the complete theory. In the following sections, the three fundamental forces described in the SM will be explained in more detail.

2.2.1 *Quantum Electrodynamics*

Quantum Electrodynamics (QED) [28–30] is the theory of the electromagnetic force. It describes the free propagations of the fermions, the photon, and is based on an underlying abelian $U(1)_{\text{EM}}$ symmetry with the electric charge as coupling constant. The interaction itself can be described using the Lagrangian $\mathscr{L}$ of QED which is given in Equation (2.2).

$$\mathscr{L}_{\text{QED}} = \bar{\Psi}\left(i\gamma^{\mu}\partial_{\mu} - m\right)\Psi + \bar{\Psi}\left(\gamma^{\mu}qA_{\mu}\right)\Psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} \tag{2.2}$$

Here, q is representing the electric charge, γ^{μ} are the Dirac or gamma matrices, A_{μ} is the photon vector field, Ψ is a massive fermion field with mass m, and the field-strength tensor $F^{\mu\nu} = \partial^{\mu}A^{\nu} - \partial^{\nu}A^{\mu}$. The $\bar{\Psi}$ denotes here the Dirac adjunct of Ψ. The indices μ and ν run over the spacetime dimensions (0, 1, 2, 3), where μ and $\nu = 0$ correspond to the time component, and μ and $\nu = 1, 2, 3$ correspond to the spatial components. This is commonly known as the Einstein notation [31]. The Lagrangian must be gauge invariant under a transformation under the $U(1)_{\text{EM}}$ symmetry. The transformation behaviour of the fermion fields is given by Equation (2.3).

$$\Psi\left(x\right) \rightarrow e^{iq\alpha(x)}\Psi\left(x\right) \tag{2.3}$$

Here, $\alpha\left(x\right)$ describes a local phase of the $U(1)_{\text{EM}}$ that would break the gauge invariance of the fermion field. To restore the gauge invariance, an additional interaction between the fermion field Ψ and photon vector field A_{μ} has to be introduced [Equation (2.4)]. This can be achieved by replacing the derivatives ∂ with covariant derivatives D, which are defined in Equation (2.5).

$$A_{\mu} \rightarrow A'_{\mu} = A_{\mu} + \partial_{\mu}\alpha\left(x\right) \tag{2.4}$$

$$\partial_{\mu} \rightarrow D_{\mu} = \partial_{\mu} + iqA_{\mu}\left(x\right) \tag{2.5}$$

Revisiting the local gauge-invariant Lagrangian for QED [Equation (2.2)], one can see that there is no mass term for the photon ($-\frac{1}{2}m^2A_{\mu}A^{\mu}$) present. Such a term would, again, violate gauge invariance. Additionally, due to local gauge invariance, a conserved quantity must be present, as derived from Noether's theorem [32]. This conserved quantity can be identified as electric charge q.

2.2.2 *Quantum Chromodynamics*

The strong force is described in the SM via a non-abelian $SU(3)_C$ gauge group. The theory of Quantum Chromodynamics (QCD) describes the interaction between colour-charged particles medi-

ated by the also colour-charged gluons. Colour charge appears in three different states: red, blue, and green, along with their corresponding anti-colours [33]. The Dirac spinor of the colour-charged quark field with flavour f, where $f \in \{u, d, c, s, t, b\}$, can, therefore, be represented as a vector of three spinors, one for each colour (see Equation (2.6)).

$$\Psi = \begin{pmatrix} \Psi_{\text{Red}}(x) \\ \Psi_{\text{Blue}}(x) \\ \Psi_{\text{Green}}(x) \end{pmatrix} \tag{2.6}$$

The Lagrangian for a free, massive quark field with mass m can be written as in Equation (2.7).

$$\mathscr{L}_{\text{Free}} = \sum_f \bar{\Psi}_f \left(i\partial\!\!\!/ - m_f\right) \Psi_f \tag{2.7}$$

The local $SU(3)_C$ transformation, which the Lagrangian needs to be invariant under, can be defined as in Equation (2.8),

$$\Psi_f^c(x) \rightarrow e^{ig_s \frac{\lambda_a}{2} \alpha_a(x)} \Psi_f^c(x) \tag{2.8}$$

where c denotes the colour charge, g_s is the strong coupling constant, a belongs to the set $\{1, ..., 8\}$ as the colour index in the adjoint representation, λ_a represents the Gell-Mann matrices [34] (generators of $SU(3)_C$), and the local phase $\alpha_a(x)$.

Similar to QED, local gauge invariance can be restored by replacing the derivatives ∂ with covariant derivatives D, as defined in Equation (2.9) where $G_\mu^a(x)$ represents the eight vector fields of the gluons.

$$\partial_\mu \rightarrow D_\mu(x) = \partial_\mu - ig_s \frac{\lambda_a}{2} G_\mu^a(x) \tag{2.9}$$

With these changes made, one can define now the local gauge invariant Lagrangian of QCD as shown in Equation (2.10).

$$\mathscr{L}_{\text{QCD}} = \sum_f \bar{\Psi}_f \left(iD\!\!\!/ - m_f\right) \Psi_f - \frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu} \tag{2.10}$$

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + g_s f_{bc}^a G_\mu^b G_\nu^c \tag{2.11}$$

Here, the gluon field strength tensor is denoted by $G_{\mu\nu}^a$ [Equation (2.11)], and the completely anti-symmetric structure constants are given by f_{bc}^a. a, b, and c belong to the set $\{1, ..., 8\}$, denoting the different Gell-Mann matrices. One distinctive feature of QCD is that when Equation (2.10) is written out with Equation (2.11), it becomes evident that gluons self-interact with each other. This self-interaction is also responsible for the finite range of the strong force. Similar to QED, the mediator boson of QCD is also massless, a requirement to maintain gauge invariance.

Another remarkable aspect of QCD is the phenomenon of the running coupling and asymptotic free-

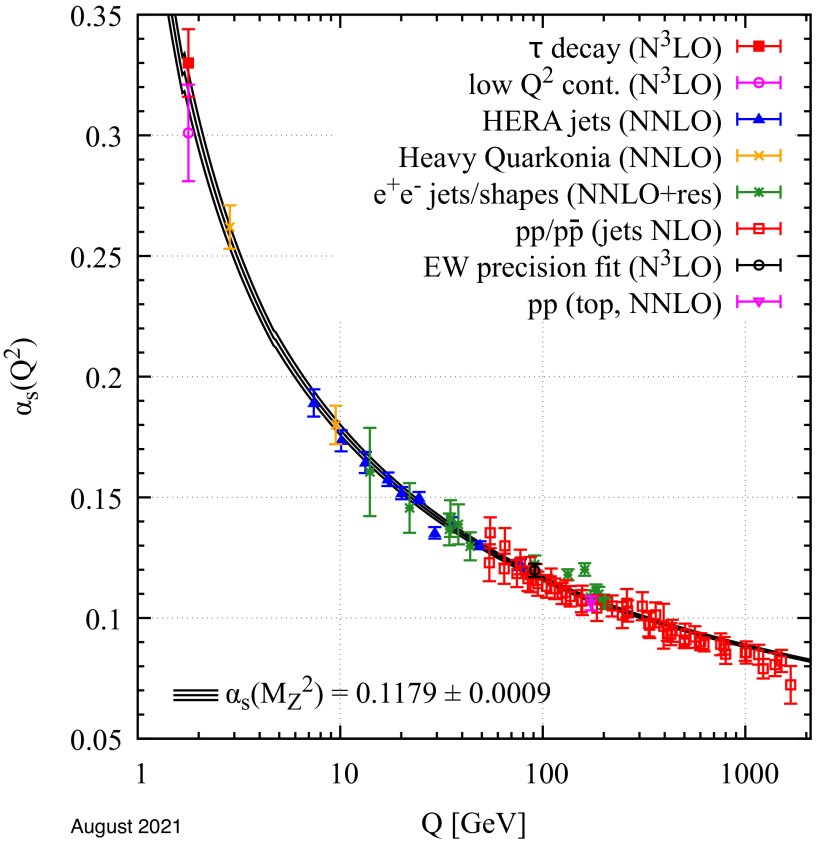


Figure 2.2.: Summary of different measurements of α_s as a function of the energy scale Q. The used degree of QCD perturbation theory is given in the brackets [27].

dom, which was discovered in 1973 by Politzer, Wilczek, and Gross [35, 36]. The strength of the strong coupling constant is dependent on the energy scale. In other words, when the characteristic energy scale of a process is increased, the strength of the strong coupling constant is decreased. This discovery is one of the most significant breakthroughs in QCD and was honoured with the Nobel Prize in 2004.

One of the consequences of asymptotic freedom is the ability to use a perturbative approach for QCD when dealing with large energy scales, where $Q \gg \Lambda_{\mathrm{QCD}}$ and $\alpha_s\left(Q^2\right) \ll 1$. In this context, the energy scale is denoted by Q, a non-perturbative constant Λ_{QCD}, and the strong coupling constant α_s. The running of the strong coupling is illustrated in Figure 2.2.

Another important mechanism in QCD is the so-called confinement [37]. It states that colour-charged particles do not exist as free particles, but rather as bound states known as hadrons. While gluons interact before such states can form, hadrons are composed of either a quark/anti-quark pair (referred to as mesons) or three (anti-)quarks (known as baryons). In addition to these two cases, according to the rules of confinement, additional bound states are possible, such as four- or five-quark bound states. These tetra- or pentaquarks were discovered by the LHCb experiment [38] in 2014 and 2015 [39, 40].

2.2.3 *Electroweak Unification*

The last of the three fundamental forces is the weak interaction. It was initiated in 1934 by Enrico Fermi with his proposal to address the continuous energy spectrum in β-decays by introducing a four-fermion vertex and a new particle, the neutrino [41]. Subsequently, the weak interaction was studied by various researchers. In 1957, Wu [42] and the team of Garwin, Lederman, and Weinrich [43] demonstrated that the weak interaction is parity-violating, as proposed in 1956 by Yang and Lee [44]. To explain this parity violation, Gell-Mann and Feynman proposed the vector-axial-coupling [45], which extended Fermi's existing theory. In 1964, after the discovery of the CP-violating nature of the weak interaction by Christenson, Cronin, Fitch, and Turlay [46], Cabibbo [47] and the team of Kobayashi and Maskawa [48] introduced what is now known as the Cabibbo-Kobayashi-Maskawa (CKM) matrix. With this underlying theory, the CKM mechanism explains all previously mentioned phenomena and the mixing of quarks via the weak interaction.
In addition to this well-described mechanism, Glashow [49], Salam [50], and Weinberg [12] independently proposed in the 1960s the unification of the weak and electromagnetic forces based on the non-abelian $SU(2)_L \otimes U(1)_Y$ symmetry group. While electric charge Q and weak isospin I are conserved quantities in the weak and electromagnetic forces, respectively, a relationship between the third component of weak isospin I_3, the hypercharge Y, and the electric charge Q can be found in this unified formalism. This is known as the Gell-Mann-Nishijima relation [51, 52], depicted in Equation (2.12).

$$Q = I_3 + \frac{Y}{2} \tag{2.12}$$

With the fermions introduced in Section 2.1, one can now define the states and their interactions via the EW force using their helicity. Right-handed fermions correspond to singlets in $SU(2)_L$, while left-handed fermions correspond to doublets in $SU(2)_L$. These doublets are the already-introduced generations. Right-handed singlets have a weak isospin of $I = 0$, whereas the left-handed doublets have $I = \frac{1}{2}$ and $I_3 = \pm\frac{1}{2}$, and, therefore, couple via the weak interaction. A special place in these interactions is occupied by the neutrinos, the up-type leptons. Within the SM, assuming massless neutrinos, although measurements have shown that at least two of the three neutrinos have non-zero masses [53–55], one can easily observe that right-handed massless neutrinos do not take part in any of the three fundamental forces due to being right-handed, massless, and having no colour or electric charge.
Similar to QED and QCD, the gauge invariance is enforced for the left-handed isospin doublets for a transformation under the $SU(2)_L$. This gauge transformation can be written as in Equation (2.13), where $\alpha_a(x)$ denotes the local phases, τ_a the generators of the $SU(2)_L$ (2×2 Pauli matrices) and g_1 the coupling strength.

$$\Psi_L(x) \rightarrow e^{i\frac{g_1}{2}\sum_{a=1}^{3}\tau_a\alpha_a(x)}\Psi_L(x) \tag{2.13}$$

Applying a similar transformation under the $U(1)_Y$ symmetry, one can find that the left-handed isospin doublets, denoted with an L, and the right-handed singlets, denoted with an R, undergo transformations as given by Equation (2.14), where $\beta(x)$ represents the local phase, Y the hypercharge, and g_2 the coupling strength of the $U(1)_Y$.

$$\Psi_{L/R}(x) \rightarrow e^{i\frac{g_2}{2}Y\beta(x)}\Psi_{L/R}(x) \tag{2.14}$$

As in QED and QCD, one can now replace the derivatives for both transformations with covariant derivatives to restore the gauge invariance. These are shown in Equations (2.15) and (2.16) for the $SU(2)_L$ and the $U(1)_Y$ respectively. In addition to the covariant derivatives, three fields W_μ^a with $a \in \{1,2,3\}$ and a field B_μ are introduced. While the three fields are added to restore the gauge invariance under the $SU(2)_L$ transformation, the gauge field B_μ is added to restore it under the $U(1)_Y$ transformation.

$$D_\mu^L = \partial_\mu + i\frac{g_1}{2}\sum_{a=1}^{3}\tau_\mu^a W_\mu^a + i\frac{g_2}{2}YB_\mu \tag{2.15}$$

$$D_\mu^R = \partial_\mu + i\frac{g_2}{2}YB_\mu \tag{2.16}$$

Taking into account all these changes, one can define now the EW Lagrangian as in Equation (2.17), where $W_a^{\mu\nu}$ and $B^{\mu\nu}$ are the field strength tensors as defined in Equations (2.18) and (2.19) respectively. ϵ_{abc} denotes the structure constant of the $SU(2)_L$.

$$\mathscr{L}_{\text{Electroweak}} = \underbrace{\bar{\Psi}_L\left(i\gamma^\mu D_\mu^L\right)\Psi_L + \bar{\Psi}_R\left(i\gamma^\mu D_\mu^R\right)\Psi_R}_{\text{Fermions}} \underbrace{-\frac{1}{4}\sum_{a=1}^{3}W_{a,\mu\nu}W_a^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}}_{\text{Gauge}} \tag{2.17}$$

$$W_a^{\mu\nu} = \partial^\mu W_a^\nu - \partial^\nu W_a^\mu - g_1\epsilon_{abc}W_b^\mu W_c^\mu \tag{2.18}$$

$$B^{\mu\nu} = \partial^\mu B^\nu - \partial^\nu B^\mu \tag{2.19}$$

Taking now a closer look at the Lagrangian defined in Equation (2.17), one can see that for both the fermions and the gauge particles, no mass terms are present. For the fermions, this would introduce mixing of the right- and left-handed particles, which would break the gauge invariance. For the gauge particles, an extra mass term would also break the gauge invariance. This stands in contrast to a lot of measurements of the masses of different fermions and bosons. To include masses for the SM particles, without breaking the gauge invariance, the Higgs mechanism is introduced.

2.3 THE HIGGS MECHANISM

The Higgs mechanism [5–7] is the final missing piece in the puzzle of the SM. It solves the problem of mass generation without breaking the local gauge invariance of the previously explained fundamental forces. It introduces direct mass terms for the later discovered $W^\pm$ and Z^0 bosons as well as an additional massive scalar boson, known as the Higgs boson [8].

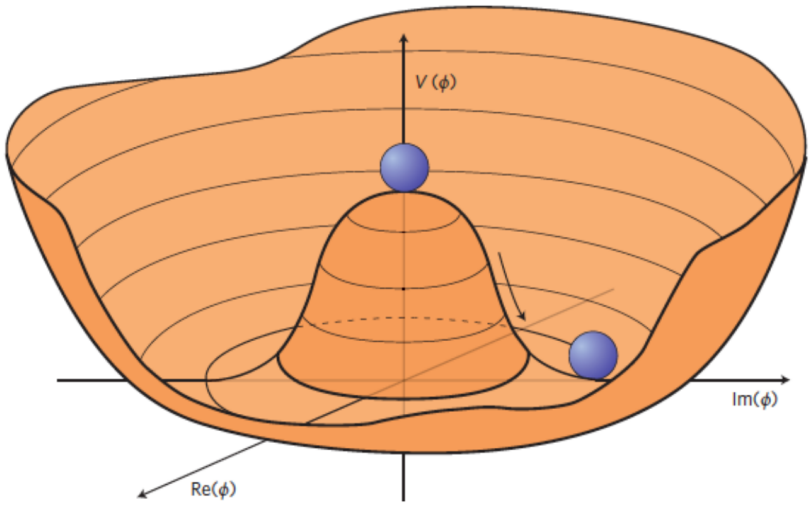


Figure 2.3.: Illustration of the Higgs potential with $\mu^2 < 0$. The minimum is at $|\phi|^2 = \frac{-\mu^2}{2\lambda}$ [56].

Focusing here first on the mass terms of the bosons and, therefore, on spontaneous symmetry breaking [18]. To spontaneously break the $SU(2)_L \otimes U(1)_Y$ symmetry, a new scalar field is introduced, which is a complex doublet of the $SU(2)_W$ symmetry [Equation (2.20)], and add it to the Lagrangian [Equation (2.21)], where $V\left(\phi^\dagger\phi\right)$ denotes the effective potential [Equation (2.22)] with $\mu^2 < 0$ and $\lambda > 0$.

$$\phi = \begin{pmatrix} \phi^+ \\ \phi \end{pmatrix} = \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{2.20}$$

$$\mathscr{L}_{\text{Higgs}} = (D_\mu\phi)^\dagger (D^\mu\phi) - V\left(\phi^\dagger\phi\right) \tag{2.21}$$

$$V\left(\phi^\dagger\phi\right) = \mu^2\left(\phi^\dagger\phi\right) + \lambda\left(\phi^\dagger\phi\right)^2 \tag{2.22}$$

The addition to the Lagrangian is accomplished by taking the covariant derivatives from Equations (2.15) and (2.16). The minimum of the effective potential corresponds to the ground state of $\phi = 0$ for $\mu^2 > 0$. In contrast, with $\mu^2 < 0$, non-zero ground states are achieved, as shown in Equation (2.23), where v is the Vacuum Expectation Value (VEV) [27] of the Higgs field and the resulting potential is known as the Higgs potential. Rewriting the formula leads to a definition of the VEV, given in Equation (2.24).

$$\phi^\dagger\phi = \frac{-\mu^2}{2\lambda} = \frac{v^2}{2} \tag{2.23}$$

$$v = \sqrt{\frac{-\mu^2}{\lambda}} \approx 246\,\text{GeV} \tag{2.24}$$

To demonstrate the effect of this in more detail, a simplified model of the Higgs potential is depicted in Figure 2.3. By selecting any of the infinite points in the potential's minimum, the rotational $U(1)$ symmetry is spontaneously broken.

With the chosen ground state, $Y = 1$ and $I_3 = -\frac{1}{2}$, the $SU(2)_L \otimes U(1)_Y$ symmetry is spontaneously broken down into a $U(1)_{EM}$ symmetry. Therefore, the gauge bosons of the QED, the

photons, remain massless.
Taking these chosen states and values into account, one can parametrise the Higgs field ϕ around the VEV [Equation (2.25)].

$$\phi(x) = \frac{1}{\sqrt{2}} e^{i \sum_{a=1}^{3} \frac{\tau_a N_a(x)}{v}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \tag{2.25}$$

$$\phi(x) \to e^{-i \sum_{a=1}^{3} \frac{\tau_a N_a(x)}{v}} \phi(x) \tag{2.26}$$

By parameterising the Higgs field around the VEV, three new fields $N_a(x)$ with $a \in \{1, 2, 3\}$ are introduced, which can be identified as the so-called Goldstone bosons [57–59]. However, with a simple gauge transform [Equation (2.26)], these bosons can be removed, leading to the Higgs doublet in the unitary gauge, shown in Equation (2.27).

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \tag{2.27}$$

The excitation of the remaining component is the massive Higgs boson. The other three degrees of freedom can be located in the kinetic term of the ϕ field when expanding around the minimum, introducing the two massive $W^{\pm}$ bosons and the massive Z^0 boson [Equation (2.28)], which is connected to the massless photon A_μ with θ_W being the Weinberg angle.

$$W_\mu^{\pm} = \frac{1}{\sqrt{2}} \left(W_\mu^1 \mp i W_\mu^2 \right), \quad \begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos(\theta_W) & \sin(\theta_W) \\ -\sin(\theta_W) & \cos(\theta_W) \end{pmatrix} \begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix} \tag{2.28}$$

$$m_{W^{\pm}} = v\frac{g}{2}, \quad m_{Z^0} = v\frac{\sqrt{g_1^2 + g_2^2}}{2} \tag{2.29}$$

$$\cos(\theta_W) = \frac{g_2}{\sqrt{g_1^2 + g_2^2}} \tag{2.30}$$

For the $W^{\pm}$ and Z^0 bosons, mass terms can be identified [Equations (2.29)] and a relation between their masses can also be found [Equation (2.30)]. The originally massless fields W^a and B acquired now a mass and the three degrees of freedom can be found as their longitudinal degrees of freedom. Taking all of this into account, one can define the Lagrangian for the EW symmetry breaking as shown in Equation (2.31).

$$\begin{aligned} \mathscr{L} = \; & \frac{1}{2} (\partial_\mu H) (\partial^\mu H) - \lambda v^2 H^2 - \lambda v H^3 - \frac{1}{4} \lambda H^4 \\ & + \frac{1}{2} \left(\frac{1}{2} v g \right)^2 W_\mu^+ W^{\mu,-} + \frac{1}{2} \left(\frac{v g}{2 \cos(\theta_W)} \right)^2 Z_\mu Z^\mu \\ & + g \left(\frac{v g}{2} \right) H W_\mu^+ W^{\mu,-} + g \frac{v g}{4 \cos^2(\theta_W)} H Z_\mu Z^\mu \\ & + \frac{g^2}{4} H^2 W_\mu^+ W^{\mu,-} + \frac{g^2}{4 \cos(\theta_W)} H^2 Z_\mu Z^\mu + \text{ const.} \end{aligned} \tag{2.31}$$

The Lagrangian provides multiple important relations between the massive weak gauge bosons and the Higgs boson, as well as interactions between multiple Higgs bosons (H^3 and H^4 terms). Additionally, a mass term for the new massive Higgs boson is generated [Equation (2.32)].

$$m_H = v\sqrt{2\lambda} \tag{2.32}$$

In addition to the new boson masses, fermion masses also need to be included. This is achieved by introducing the so-called Yukawa couplings [12] between the Higgs doublet and the left-handed $SU(2)_L$ fermion doublets and the right-handed fermion singlets. The Yukawa term for the quarks and leptons can be defined by introducing an additional charge conjugated Higgs doublet, which is given in Equation (2.33). The resulting Lagrangian for the Yukawa couplings to the first quark generation is given in Equation (2.34).

$$\phi_c(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} v + H(x) \\ 0 \end{pmatrix} \tag{2.33}$$

$$\mathscr{L}_{\text{Yukawa}}^{\text{Quarks}} = -\left(g_d \left(\bar{u}, \bar{d}\right)_L \phi d_R + g_u \left(\bar{u}, \bar{d}\right)_L \phi u_R + h.c\right) \tag{2.34}$$

$$\mathscr{L}_{\text{Yukawa}} = -\underbrace{v\frac{g_f}{2}}_{\text{Fermion mass}} \bar{\Psi}_f \Psi_f \left(1 + \frac{H}{v}\right) \tag{2.35}$$

Here, g_d and g_u denote the coupling constants to the down and up quark and $h.c$ refers to the hermitian conjugate term. The interaction to all other quarks and leptons can be described in a similar fashion and the resulting Lagrangian of the fermion Higgs interaction is given in Equation (2.35) with the coupling constant g_f and the fermion fields Ψ_f.

2.4 LIMITATIONS OF THE STANDARD MODEL

Although the SM shows unprecedented agreement with all measurements and is one of the most successful theories ever developed, it still lacks the explanation for certain empirical observations. One of these problematic observations is the mass of the already-mentioned neutrinos. In the SM, neutrinos are supposed to be massless particles, even though different experiments have measured the so-called neutrino oscillation [60]. This oscillation is a periodic effect in which the neutrinos change their flavour between the three generations and therefore need to have mass. Since this mass and therefore their influence on collider observables must be very small (the current upper limit is $0.8\,\text{eV}$ [61]), the predictions made by the SM are still valid due to the resolution restrictions of the available physics experiments.

Another significant problem of modern particle physics is the baryon asymmetry of the universe [62]. Postulated by Sakharov in 1967 [63], three conditions must be fulfilled to create the universe how we observe it today: baryon number violation, C and CP-symmetry violation, and interactions out of the thermal equilibrium.

While the SM in theory has the means to fulfill the first condition via so-called Sphaleron processes [64, 65], these processes cannot be the reason for the observed baryogenesis in the SM due to the high Higgs boson mass of $\approx 125\,\text{GeV}$, which would violate the third condition of Sakharov. To have a first order phase transition with Sphalerons, the Higgs boson mass must be below $73\,\text{GeV}$ [66]. For the the second condition, the observed CP-violation in weak interactions in the SM alone is insufficient to explain the observed disparity in the universe's current baryon asymmetry. According to the SM, during the early stages of the universe, the Big Bang [67] should have generated equal quantities of matter and antimatter. However, as matter and antimatter came into contact and annihilated each other, the universe, theoretically, should not have evolved into the matter-dominated state we observe today. Therefore, persistence of this matter-dominated universe remains an unsolved problem.

Another significant challenge is the naturalness problem concerning the mass of the Higgs boson, often referred to as the hierarchy problem [68]. While experimental measurements have determined the mass of the Higgs boson to $m_H = 125.11 \pm 0.11\,\text{GeV}$ [69], it is possible to calculate corrections to the bare mass of the Higgs boson. This bare mass is subject to various loop corrections. Only by an accidental cancellation of these corrections, the bare Higgs mass is not corrected to very high scales at the order of the Planck scale. Other bosons remain protected by local gauge symmetries and fermions by their chiral symmetries and are therefore not effected by these corrections. This is pure fine tuning in the case of the Higgs boson mass, which was initially pointed out by Susskind in 1979 [70].

While only a few issues with the SM are mentioned here, several other challenges exist that are not covered in this discussion. These include the enigmatic presence of dark matter [71, 72], as well as the dark energy [73, 74]. Additionally, the integration of gravity into the SM [75] remains a significant challenge for particle physicists.

3

THE LARGE HADRON COLLIDER AND THE ATLAS EXPERIMENT

To study the physics of the smallest objects, vast and powerful machines like the LHC [11] are needed. These enormous machines can accelerate particles to very high energies before colliding them to access and study the properties and interactions of elementary particles. The LHC is the world's largest and most powerful hadron accelerator and collider, being able to accelerate hadronic particles to velocities close to the speed of light and collide them. Its main purpose is to collide protons, although it is also capable of accelerating and colliding heavy ions, like lead. This thesis focuses on the proton-proton accelerator and collider properties of the LHC; therefore, only these properties of the LHC are explained further. The resulting collisions of the LHC are studied with enormous experiments like the ATLAS detector.
In the following chapter, the LHC is briefly described together with the ATLAS experiment and its sub-detectors. Due to the focus of this thesis on the LHC Run 2 dataset, which was recorded between the years 2015 and 2018, the layouts of the LHC and the ATLAS experiment in these years are depicted and discussed.

3.1 THE LARGE HADRON COLLIDER

The LHC is a circular particle accelerator and collider at the CERN near Geneva in Switzerland. The accelerator itself is nearly 27 km long and lies $100 - 150$ m underground. It consists of two ring acceleration pipes, in which the particles are circulated and accelerated in opposite directions. The pipes themselves are surrounded by 1232 superconducting dipole magnets, which can generate a magnetic field of up to 8.3 T to bend the proton beams and hold them on their circular track. In addition, 392 quadrupole magnets are installed to focus the beams. While the LHC is the largest of CERNs accelerators, it cannot work without its predecessors and other accelerators, which are used in an accelerator chain, delivering protons to the LHC. The Run 2 layout (status 2018) of all CERN accelerators is illustrated in Figure 3.1.

The LHC is fed from a chain of smaller accelerators, which extract and pre-accelerate the particles from a bottle of hydrogen to hundreds of GeV before being inserted into the LHC. Starting with the Duoplasmatron, which separates the electrons from the hydrogen atoms and delivers the resulting protons to the Linear Accelerator (LINAC) 2. While the electrons are discarded, the protons are accelerated to an energy of 50 MeV using radio-frequency cavities. After the LINAC, the protons enter the 157 m-long Proton Synchrotron Booster (PSB), or simply Booster, in which the protons are accelerated to 1.4 GeV and also bundled in so-called trains. These trains are then sent to the 628 m-long Proton Synchrotron (PS), which further accelerates the protons to an energy of 26 GeV. After multiple quality tests (and beam adaptations), the protons are sent to the Super Proton Synchrotron (SPS).

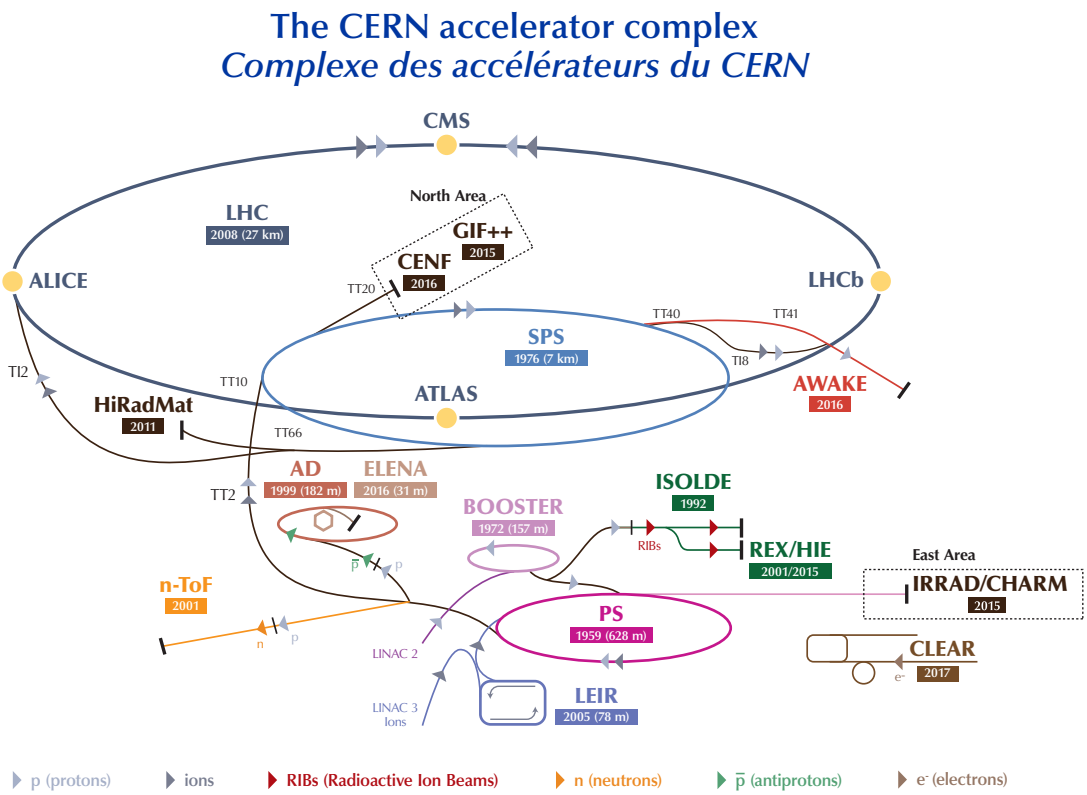


Figure 3.1.: 2018 layout of the CERN accelerator complex [76].

The 7 km-long SPS is one of the more prominent predecessors of the LHC, where both the $W^{\pm}$ [23, 24] and Z^0 [25, 26] bosons were discovered in 1983. It accelerates the proton trains to 450 GeV before sending them to their final destination, the LHC. In addition, the SPS also supplies other experimental facilities at CERN with protons and other particles. However, since these experiments are not related to this thesis, they are not further discussed here. It is herein filled with 2808 bunches of 10^{11} protons with a per-bunch spacing of 25 ns. The LHC is designed to accelerate the protons from 450 GeV to a collision (or center-of-mass) energy of $\sqrt{s} = 14$ TeV (7 TeV per beam) in nearly 15 minutes. However, the center-of-mass energy for Run 2 of the LHC was set to $\sqrt{s} = 13$ TeV (6.5 TeV per beam) due to technical considerations of the LHC magnets' condition. Once the protons have reached their target energy, the beams are brought to collision at four so-called interaction points around the ring. Around each of the interaction points, a particle detector is built to collect particles produced in the proton-proton collisions, which can be analysed. These four experiments are the ATLAS [77], CMS [78], LHCb [38], and ALICE detectors [79]. While the LHCb and ALICE experiments have very specific physics programs, the ATLAS and CMS experiments are designed as multi-purpose detectors with broader physics programs.

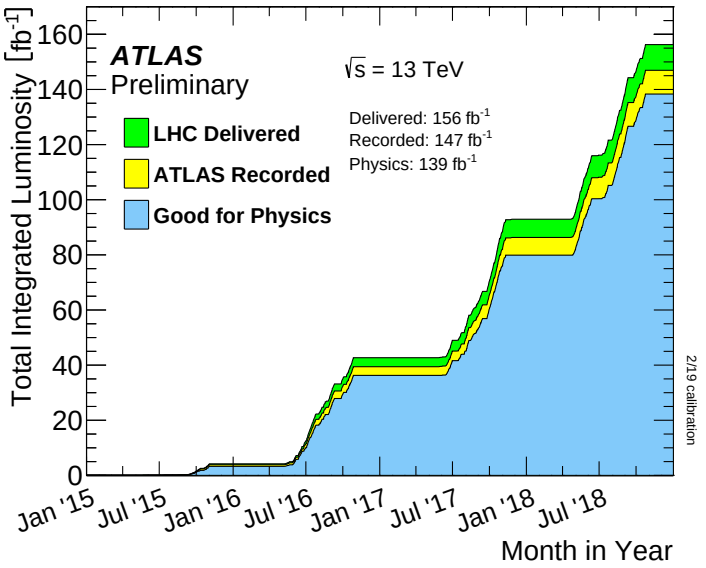


Figure 3.2.: Run 2 Luminosity delivered by LHC. In addition, the recorded and good for physics luminosity of the ATLAS experiment is presented [80].

While the high center-of-mass energy of the LHC is one of its main properties, another property is at least equally important: the instantaneous luminosity L, defined in Equation (3.1):

$$L = \frac{f_{\text{rev}} \cdot N^2}{4\pi\sigma_x \cdot \sigma_y} \cdot F \tag{3.1}$$

Here, L is determined by the revolution frequency f_{rev}, the total number of protons per beam N (assuming the same amount of protons in each beam), the Gaussian beam widths σ_x and σ_y, and the geometric luminosity reduction factor F. The latter factor accounts for corrections due to the beam crossing angle θ_c and other effects at the interaction point. The LHC was designed for an instantaneous luminosity of $10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$.
Using the instantaneous luminosity, one can define an absolute measure for the total number of interactions recorded by the different experiments: the integrated luminosity $\int L\,\mathrm{d}t$. The values of the integrated luminosity as provided by the LHC and recorded by the ATLAS detector in the years 2015 till 2018 are shown in Figure 3.2. The diagram also indicates the integrated luminosity corresponding to the fraction of the data recorded by ATLAS which is considered good for physics. The latter indicates if the collected data can be used for physics analyses, which is mainly related to the status of the detector and data-taking while recording the data.

With the high instantaneous luminosity of the LHC and the number of protons per bunch, the amount of interactions per bunch crossing needs to be taken into account when determining the integrated luminosity of the different experiments. This mean number of interactions per bunch crossing, denoted as $\langle\mu\rangle$, is shown in Figure 3.3 for the ATLAS Run 2 dataset.

In the Run 2 dataset of ATLAS, which is analysed in this thesis, the mean number of interactions per bunch crossing is $< \mu >= 33.7$. This means that for every bunch crossing that occurs, on

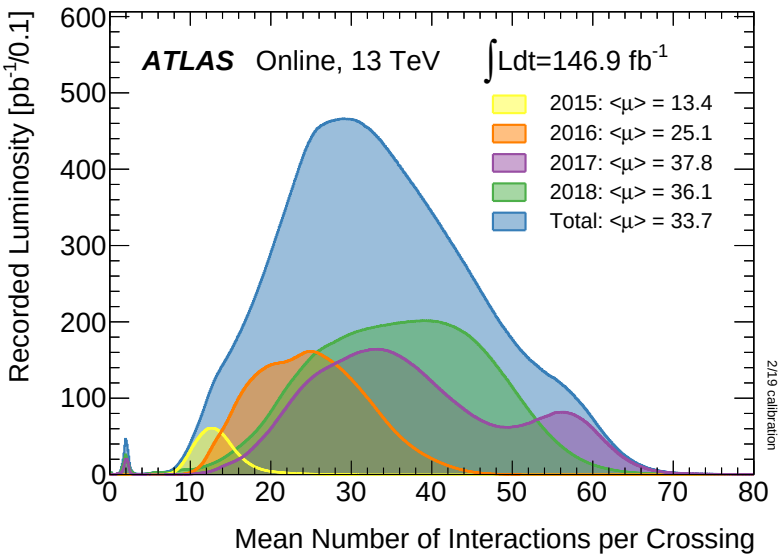


Figure 3.3.: Mean number of interactions per bunch crossing for the Run 2 of the LHC per year of the ATLAS recorded dataset [80].

average, 33.7 inelastic proton-proton interactions take place. This overlay of interactions per bunch crossing is called pile-up.

3.2 THE ATLAS EXPERIMENT

The already-mentioned ATLAS experiment [77] is one of the four large experiments at the LHC. It is a cylindrical multi-purpose detector with a length of 44 m and a diameter of 25 m and is, therefore, the largest of the four large experiments. The detector itself weighs 7000 metric tons and consists of multiple smaller sub-detectors, which are constructed in an onion-like manner around the interaction point 1 of the LHC. A schematic representation of the ATLAS detector is depicted in Figure 3.4.

ATLAS covers nearly the entire solid angle of 4π and employs a right-handed Cartesian coordinate system with the nominal interaction point as the center. The positive x-axis is oriented toward the center of the LHC ring, while the positive y-axis points upward to the top of the detector and the surface, defining the $x-y$ plane. The positive z-axis is directed along the beamline. Due to the shape of ATLAS, the $x-y$ plane is typically described using a cylindrical coordinate system with the $r-\phi$ plane. In cylindrical coordinates, the radius (or displacement from the beam axis) is defined as $r = \sqrt{x^2+y^2}$, and ϕ represents the azimuthal angle. The polar angle θ, which measures the displacement from the $y-z$ plane, is commonly not used. The reason for this is that in collider environments, a longitudinal Lorentz boost along the beam direction (z-axis) is present due to the very high energies of the particles in comparison to their mass. In this relativistic setting, where the momenta of incoming particles vary, the non-Lorentz invariant θ is replaced with pseudorapidity η, defined in Equation (3.2). With the relativistic approximation $m \ll |\vec{p}| \Rightarrow E \approx |\vec{p}|$, where $\vec{p}$ is the three-momentum vector and E is the energy of the particle, pseudorapidity converges to the definition of rapidity y, as shown in Equation (3.2).

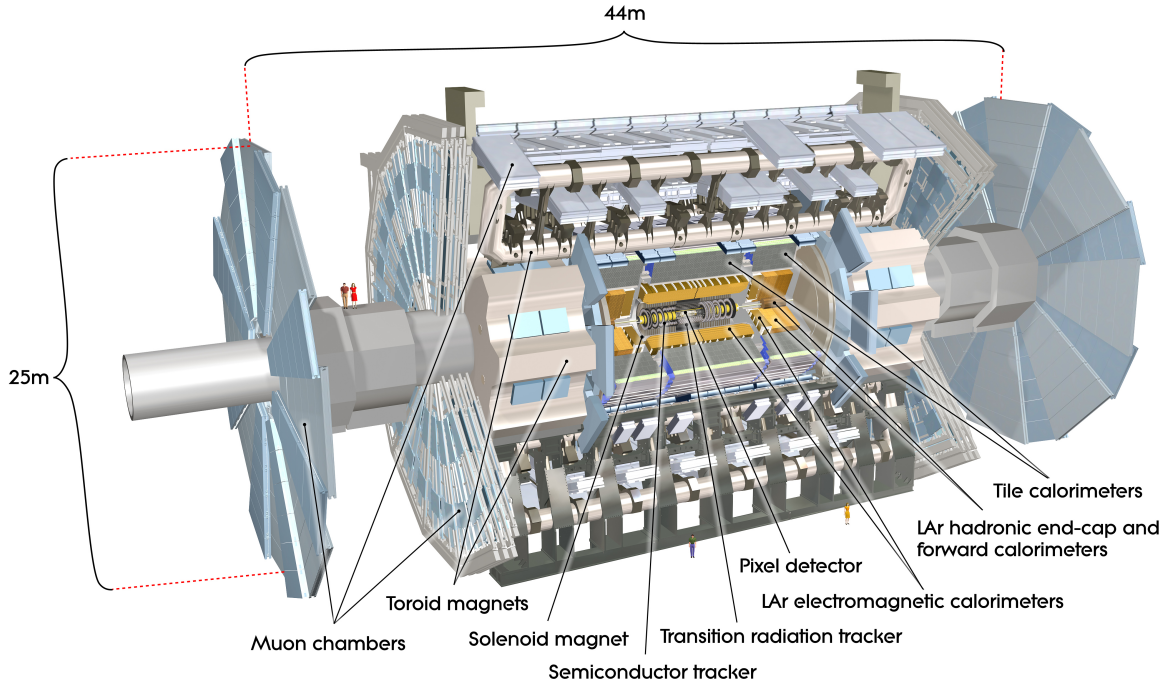


Figure 3.4.: Schematic of the ATLAS detector with its different sub-detectors [81]

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \underbrace{\Rightarrow}_{m \ll |\vec{p}|} y = \frac{1}{2}\ln\left(\frac{|\vec{p}| + p_z}{|\vec{p}| - p_z}\right) \tag{3.2}$$

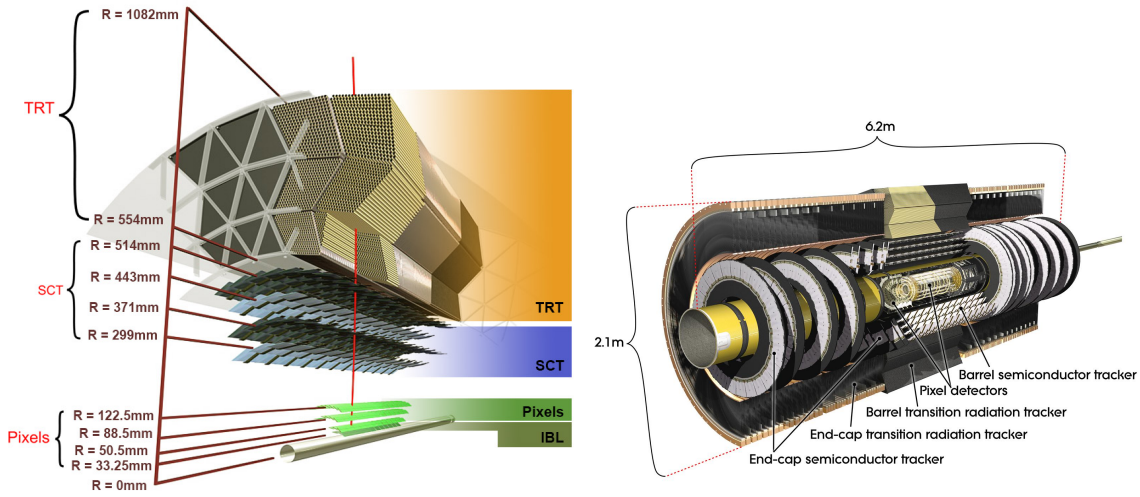
A difference in either η or y ($\Delta\eta$ or Δy) is invariant under a Lorentz boost in the z direction, making it the preferred geometric parameter in collider event physics. This becomes significant when defining, for example, the spatial distance between two objects in the detector in the $\eta - \phi$ plane, which is defined as:

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

This variable is essential for the reconstruction and identification of physics objects. Equally important for these tasks are the various sub-detectors of ATLAS and their functionalities. In the following sections, the sub-detectors of ATLAS will be described in more detail, starting from the innermost part of the detector and proceeding outwards.

3.2.1 *Inner Detector*

The Inner Detector (ID) [82, 83] serves as the heart of ATLAS and is one of the most important components of the detector itself, measuring a length of 6.2 m. It comprises three main components: the Pixel Detector (PD), the Semiconductor Tracker (SCT), and the Transition Radiation Tracker (TRT). Its primary goal is to precisely measure the momenta of particles and reconstruct their charges and



(a): Schematic cross-section of the ATLAS inner detector with its different sub-detectors [84].

(b): Schematic of the ATLAS inner detector [85].

Figure 3.5.: Schematic depictions of the ATLAS inner detector.

tracks within the ID. Additionally, this information is vital for the identification and reconstruction of different vertices (primary, secondary, tertiary), which play a pivotal role in the identification of physics objects. A schematic view and cross-section of the ID, excluding the solenoid magnet, are presented in Figure 3.5, and further details about its components are explained in their respective sections below.

Pixel Detector

The pixel detector (PD) [84, 86] consists of three barrel layers located in a $1.4\,\text{m}$ long cylinder with a diameter of $0.5\,\text{m}$ centred around the interaction point. The forward region is constructed with three discs positioned at each end of the cylinder, enabling the PD to provide coverage of $|\eta| < 2.5$. With a total of 80 million channels, the three barrels, placed at $50.5\,\text{mm}$, $88.5\,\text{mm}$, and $122.5\,\text{mm}$, respectively, contain a total of 1744 modules, each with 46080 pixels measuring typically $(R-\phi)\times z = 50\,\mu\text{m}\times 400\,\mu\text{m}$. A computer-generated image of the PD is shown in Figure 3.6.

In preparation for Run 2, the PD was extended with a fourth barrel layer, the Insertable b-Layer (IBL) [84, 88]. The IBL was inserted as the new innermost layer of the PD and is located at a mean radius of $33.2\,\text{mm}$ around the beam pipe. Containing 12 million pixels with a typical size of $(R-\phi)\times z = 50\,\mu\text{m}\times 250\,\mu\text{m}$, the IBL improves the tracking capabilities of the ID by providing an additional measurement point and mitigating the effects of radiation damage and the higher instantaneous luminosity of Run 2 compared to Run 1, for which the PD was designed. The benefit in transverse Impact Parameter (IP) resolution between Run 1 (without IBL) and Run 2 (with IBL) can be seen in Figure 3.7.

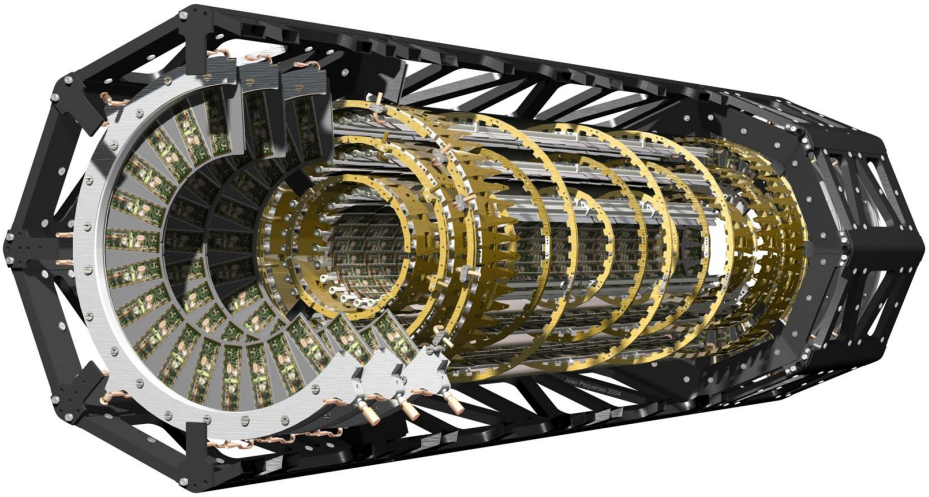


Figure 3.6.: Computer generated image of the pixel detector without the Insertable b-Layer (IBL) [87].

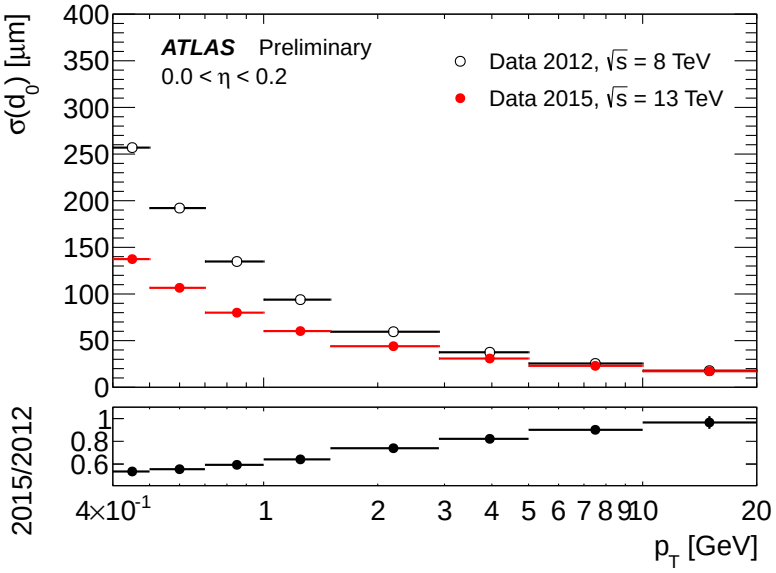


Figure 3.7.: Performance improvement with the IBL in terms of unfolded transverse IP resolution. The reference resolution (black) was calculated using data taken in 2012 with $\sqrt{s} = 8\,\text{TeV}$ with a mixture of jet, τ, and Missing Transverse Momentum ($E_{\text{T}}^{\text{miss}}$) triggers. The new resolution (red) was calculated using data from 2015 with $\sqrt{s} = 13\,\text{TeV}$ with a minimum bias trigger and the IBL. Both resolutions are depicted as functions of the track transverse momentum p_{T} with $0 < |\eta| < 0.2$ [89].

Semiconductor Tracker

The Semiconductor Tracker (SCT) [90] is a silicon-based microstrip detector that surrounds the PD and extends the tracking capabilities to radial distances of $299\,\text{mm} < r < 560\,\text{mm}$. The SCT comprises four layers of silicon-strip sensors in the barrel region and 9 disks in each forward region. Similar to the PD, the SCT covers a pseudorapidity range of $|\eta| < 2.5$, with a size of the detecting volume, in this case, the strips, measuring $(R - \phi) \times z = 80\,\mu\text{m} \times 12\,\text{cm}$. To accurately track particles passing through the layers, two of the four layers of each module are tilted by $40\,\text{mrad}$ relative to each other. To reduce electrical noise in the semiconductor-based detectors, the SCT is cooled to approximately $0°\,\text{C}$. Unlike the IBL and the PD, the SCT only records binary hit information for each strip with its 6.3 million readout channels.

Transition Radiation Tracker

The Transition Radiation Tracker (TRT) [91] is a gas-based tracking detector that uses so-called straw tubes [92]. The TRT comprises $300,000$ of these straw tubes, each with a diameter of $4\,\text{mm}$. It covers a radial range of $560\,\text{mm} < r < 1080\,\text{mm}$ in the barrel and $|z| < 720\,\text{mm}$ and $617\,\text{mm} < r < 1106\,\text{mm}$ in the forward region, with $827\,\text{mm} < |z| < 2774\,\text{mm}$. These tubes are filled with a gas mixture consisting of $70\,\%$ Xe, $27\,\%$ CO_2, and $3\,\%$ O_2. The TRT provides pseudorapidity coverage of $|\eta| < 2$. In the barrel area, each straw consists of a $144\,\text{cm}$ long Kapton tube with a $31\,\mu\text{m}$ thick gold-plated tungsten wire at its center. In the forward regions, $37\,\text{cm}$ long tubes are integrated into wheels. On average, each charged particle causes 36 ionisations in the straw tubes [77] as it traverses the materials between the tubes. These materials possess varying refractive indices, leading to the generation of transition radiation when charged particles traverse them.

Solenoid Magnet

Surrounding the ID is a $5.8\,\text{m}$-long and $2.56\,\text{m}$-wide central superconducting solenoid magnet. It generates a $2\,\text{T}$ magnetic field that permeates the ID and enables the precise measurement of particle momentum and charge since the magnetic field is parallel to the beam pipe. This results in a curved trajectory for charged particles in the $(R - \phi)$ plane. Using the bending radius of the charged particles, the transverse momentum p_T in GeV can be calculated with:

$$p_\text{T} \approx 0.3 \cdot z \cdot B \cdot r \tag{3.4}$$

Here, r is the bending radius in meters, B is the magnetic field in Tesla, and z is the particle charge in units of elementary charge. The 0.3 is a rounded factor for the speed of light, approximately $3 \times 10^8\,\text{m/s}$. The resulting momentum measurement suffers from higher uncertainties for higher momenta due to the smaller curvature of charged particle tracks and, consequently, higher uncertainty in radius measurement. The relative resolution of momentum measurements in the ID can be described as follows [77]:

$$\frac{\sigma\left(p_{\mathrm{T}}\right)}{p_{\mathrm{T}}} = \frac{0.05\,\%\,p_{\mathrm{T}}}{\mathrm{GeV}} \oplus 1\,\% \tag{3.5}$$

where $\oplus$ is the magnitude of two uncorrelated signals by taking the square root of the sum of squares.

3.2.2 *Calorimeter*

Surrounding the central solenoid magnet are the calorimeter systems of ATLAS. These systems are designed to measure the energy of the particles produced in the collisions and their subsequent decay by absorbing them. The calorimeter itself consists of three main parts: The Electromagnetic Calorimeter (ECal), the Hadronic Calorimeter (HCal), and the Forward Calorimeter (FCal). The different systems are designed for various tasks, but all calorimeter systems share a common feature: layering. The calorimeters consist of multiple layers of absorber and active material used to measure the energy of the particles. A more detailed description of each sub-calorimeter is provided in their respective following sections. The different calorimeter sub-systems are depicted in Figure 3.8.

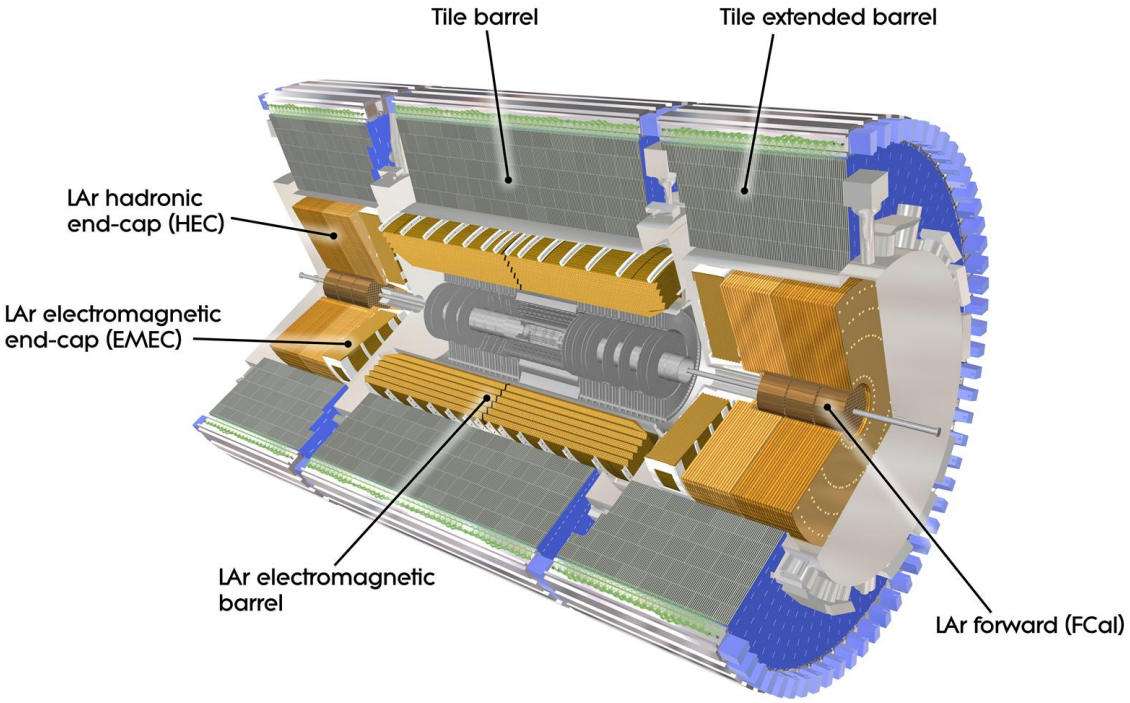


Figure 3.8.: Schematic representation of the ATLAS calorimeter system [93].

Electromagnetic Calorimeter

The ECal [77] is a high granularity sampling calorimeter that uses liquid argon (Liquid Argon (LAr)) as active and lead as absorber material. Its main task is the absorption and measurement of photons

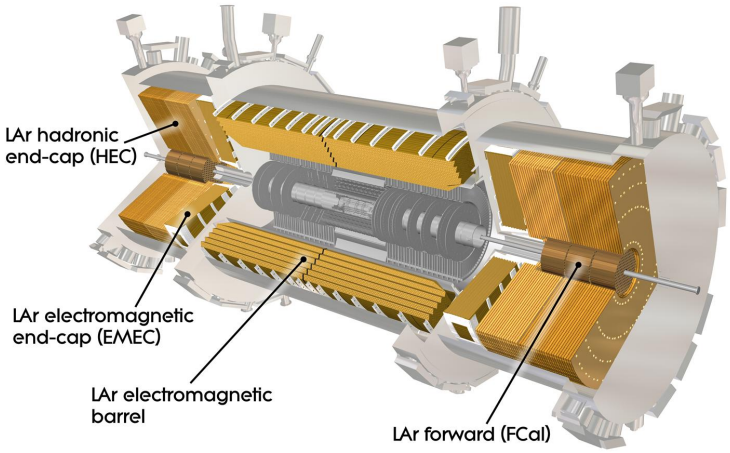


Figure 3.9.: Schematic representation of the ATLAS electromagnetic calorimeter [94].

and electrons and their respective energies. While heavier and more energetic particles pass through the ECal and deposit just a small fraction of their energy in the calorimeter cells due to their mass, (nearly) all electrons and photons are absorbed. The ECal electrodes and absorber plates are constructed in an accordion-shaped manner, ensuring complete symmetry in ϕ to achieve full coverage in that aspect. The ECal is split into one barrel and two Liquid Argon Electromagnetic End-Cap (EMEC) parts which are depicted in Figures 3.8 and 3.9, respectively. While the EMECs cover a range in pseudorapidity of $1.375 < |\eta| < 3.2$ with a slightly coarser granularity, the barrel is split again into two identical halves, each $6.4\,\mathrm{m}$ long and $53\,\mathrm{m}$ thick, covering a pseudorapidity range of $|\eta| < 1.475$ with a granularity of $(\Delta\eta \times \Delta\phi) = 0.025 \times 0.025$. The relative resolution of the energy measurement of the ECal is described with Equation (3.6) where E denotes the measured energy.

$$\frac{\sigma\,(E)}{E} = 10\,\%\sqrt{\frac{1}{E}} \oplus 0.7\,\% \tag{3.6}$$

In contrast to the momentum measurement described in Section 3.2.1, the relative resolution of the energy measurement increases with particle energy due to the statistical nature of electromagnetic showers.

Hadronic Calorimeter

Enclosing the ECal is the Hadronic Calorimeter (HCal) [77], which is, with its 2900 metric tons, one of the heaviest parts of ATLAS (depicted in Figure 3.8). Similar to the ECal, the HCal is a sampling calorimeter with alternating layers of absorber and active materials which is split into two parts, the barrel Tile Calorimeter (TileCal) [95] and two Liquid Argon Hadronic End-Caps (HECs). With a total length of $11\,\mathrm{m}$, subdivided in a central barrel ($5.8\,\mathrm{m}$) and two extended barrels ($2.6\,\mathrm{m}$ each), the TileCal uses iron as absorber and plastic scintillating tiles as active material. It covers a pseudorapidity range of $|\eta| < 1.7$, while the HECs cover a range of $1.5 < |\eta| < 3.2$ and are also

copper/LAr-based. In contrast to the ECal, the HCal is designed to absorb and measure all hadronic particles and therefore avoid the so-called punch-through of particles. The latter describes the effect, that hadronic showers of particles decaying inside the HCal reach the end of the calorimeter and access the muon system. In doing so, these punch-through particles could be misidentified as muons. The relative resolution of the energy measurement of the HCal is described with Equation (3.7) where E denotes the measured energy.

$$\frac{\sigma\left(E\right)}{E} = 50\,\%\sqrt{\frac{1}{E}} \oplus 3\,\% \tag{3.7}$$

Forward Calorimeter

The last part of the ATLAS calorimeter system is the Forward Calorimeter (FCal) [77, 96, 97]. It consists of three layers, one copper-based electromagnetic layer and two tungsten-based hadronic layers. Providing a coverage in pseudorapidity of $3.1 < |\eta| < 4.9$, the FCal was designed to provide energy measurements of forward particles and contribute, with the HCal, to the Missing Transverse Momentum ($E_{\mathrm{T}}^{\mathrm{miss}}$) measurement. The relative resolution of the energy measurement of the FCal is described with Equation (3.8) where E denotes the measured energy.

$$\frac{\sigma\left(E\right)}{E} = 100\,\%\sqrt{\frac{1}{E}} \oplus 10\,\% \tag{3.8}$$

3.2.3 *Toroid Magnets*

Outside of the calorimeter system, the second of the ATLAS magnet systems is placed. The toroid magnet [98], consisting of one barrel and two end-cap magnets, is 26 m long and 22 m in diameter. It is a superconducting magnet that provides, in addition to the solenoid magnet, a magnetic field for the measurement of the particle's momenta. Each of the magnets consists of eight coils creating a toroidal magnetic field of 0.5 T to the barrel and 1 T to the end-cap regions, respectively. A schematic drawing of the magnet system is presented in Figure 3.10.

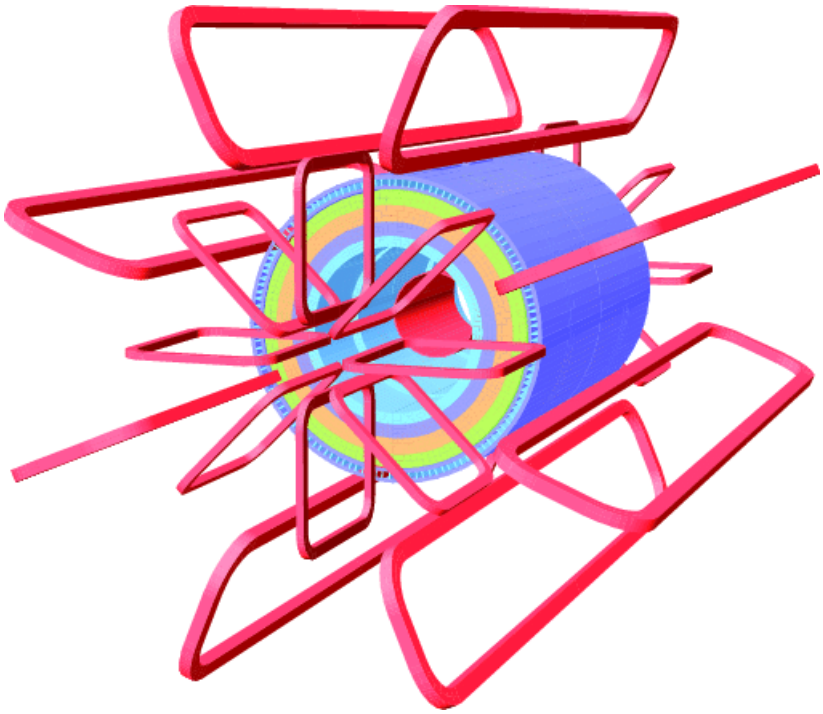


Figure 3.10.: Schematic representation of the ATLAS magnet system (red) [99].

3.2.4 *Muon Spectrometer*

The outermost and final sub-detector of the ATLAS experiment is the Muon Spectrometer (MS) [77, 100]. Its primary purpose is to track and measure muons, as they are normally the only particles, besides neutrinos, that exit the calorimeter system. Due to their minimally ionising nature, muons can easily traverse the entire detector without being stopped or absorbed. Offering a momentum resolution of $10\,\%$ for 1 TeV particles, the barrel detectors of the MS are arranged in three concentric layers around the beam pipe, located at radii of $5\,\text{m}$, $7.5\,\text{m}$, and $10.5\,\text{m}$, respectively. In the end-cap regions, three wheel-shaped detectors are positioned at $z = 7.4\,\text{m}$, $14\,\text{m}$, and $21.5\,\text{m}$ from the interaction point, respectively, providing coverage in the forward regions. The entire MS employs gas-based detector technologies. A schematic view of these sub-detector systems is displayed in Figure 3.11.

The MS is built around two types of chambers for separate corresponding use cases: precision measurement and trigger chambers. For precision measurements, Monitored Drift Tubes (MDTs) and Cathod Strip Chambers (CSCs) are used. The MDTs are aluminum tubes with various lengths and a $30\,\text{mm}$ diameter. They are filled with a gas mixture of $93\,\%$ Ar and $7\,\%$ CO_2 The CSCs are multiwire proportional chambers built in strips. Each of the strips consists of radially oriented wires and cathodes and is filled with a gas mixture of $80\,\%$ Ar and $20\,\%$ CO_2. While the MDTs provide better resolution in the barrel, the CSCs provide better spatial resolution to cope with the higher signal rates in the forward region. The MDTs are located at the three layers in the barrel and on each of the wheels in the end-cap regions. Each layer is fitted with three layers of MDTs beside the innermost barrel layer, which is equipped with four layers. On the innermost wheel in the forward

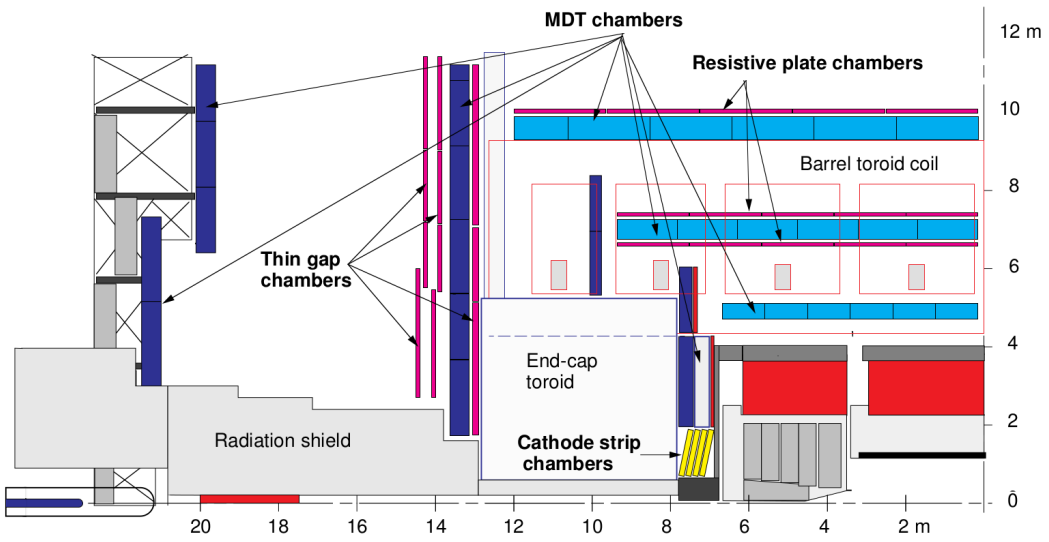


Figure 3.11.: Schematic representation of the ATLAS muon spectrometer [100].

region, the MDTs are replaced with CSCs to better cope with the high flux of particles.
For trigger purposes, the MS has two types of chambers: Resistive Plate Chambers (RPCs) and Thin Gap Chambers (TGCs). The RPCs are installed in the barrel region, stacked on the MDT chambers, providing crucial timing information for the trigger, due to their time resolution of 10 ns. The TGCs are found on the wheels of the MS in the forward region, also providing crucial timing information. In total, the MS covers a range in pseudorapidity of $|\eta| < 2.4$ for measurements and $|\eta| < 2.7$ for triggering.

3.2.5 *Trigger and Data Acquisition*

With the previously described bunch separation of 25 ns, proton-proton collisions inside the detector occur at a frequency of 40 MHz. Taking into account multiple interactions per bunch crossing (see Figure 3.3) and the average event size of 1.6 MB, the resulting data rate from the detector, ≈ 64 TB/s, is too high to be recorded and stored in its entirety. To control this data rate and the number of stored events, ATLAS employs a Data Acquisition (DAQ) system, consisting of a two-stage trigger [101], as shown in Figure 3.12. To efficiently exploit the available bandwidth, the trigger system identifies interesting events in real-time and decides whether to store or discard them. The criteria for what constitutes an interesting event are defined in the so-called trigger menu. In this menu, various data streams are designated for different purposes, such as the physics main stream, which, as the name implies, is the primary data stream for physics analyses. Each stream has its unique requirements and event selection criteria, and events meeting these criteria are retained. The first stage of the trigger system, the Level 1 (L1) trigger, is a hardware-based trigger that utilises custom-built Field Programmable Gate Arrays (FPGAs). These FPGAs use the information provided by the calorimeter and muon systems to decide within 2.5 μs regarding event acceptance or rejection. With a reduced data rate of only 100 kHz, events that pass the L1 trigger are directed to the High-

Level Trigger (HLT). The HLT is a software-based trigger that operates on a dedicated computer cluster, which has access to comprehensive event information from all sub-detectors. Leveraging fast reconstruction algorithms capable of promptly reconstructing and tracking physics objects, the HLT evaluates whether an event meets the criteria specified in the trigger menu to be considered within the Region of Interest (RoI). Passing through the HLT, the data rate is further reduced from 100 kHz to an average of 1 kHz, making it manageable for storage and subsequent analysis.

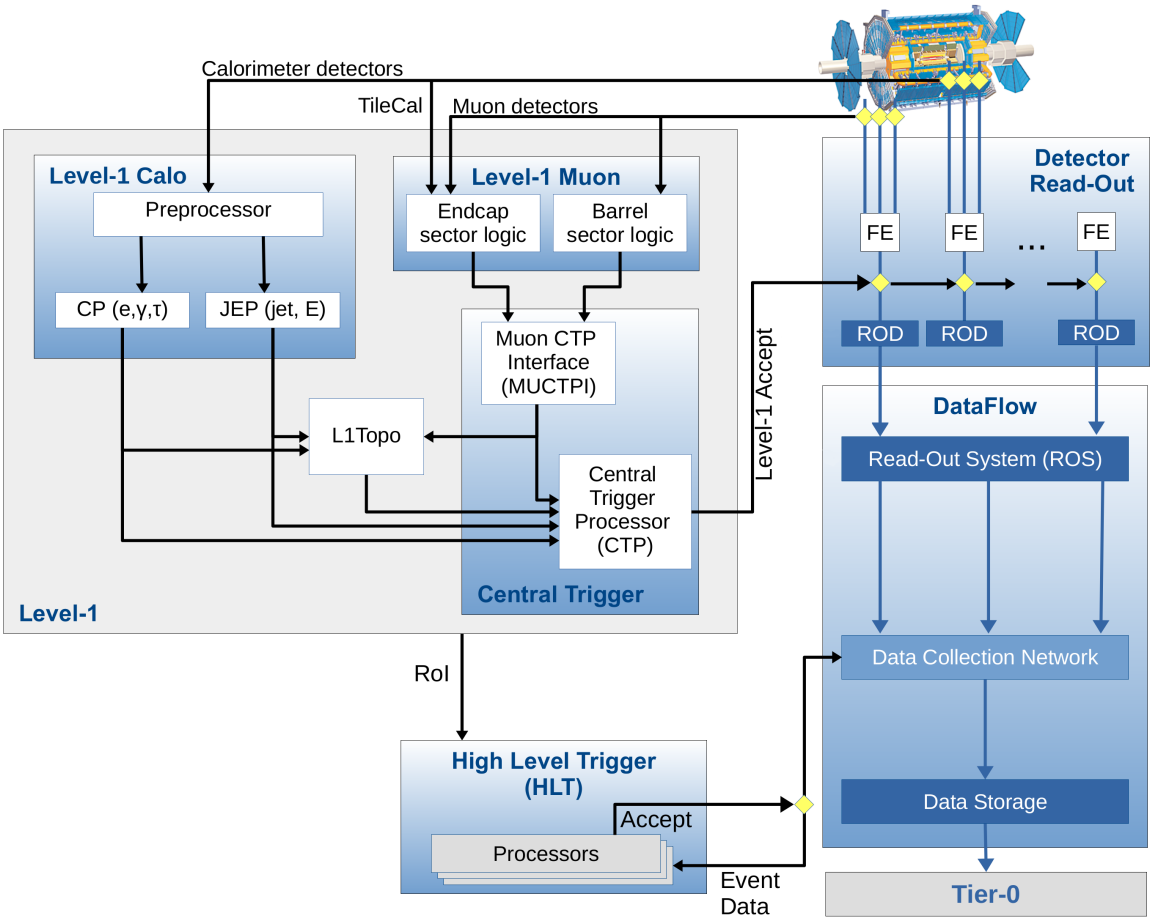


Figure 3.12.: Schematic overview of the trigger chain and DAQ of ATLAS. Taken and adapted from Ref. [101].

4

EVENT SIMULATION OF PROTON-PROTON COLLISIONS

Using the data taken with the ATLAS detector described in Chapter 3, one needs to compare the results from the data to expectations. These expectations stem from calculations of different theoretical models, usually the SM. To create simulations based on the theory calculations, different Monte Carlo (MC) techniques are used to simulate the proton-proton collisions. Using protons for the collisions makes the task of simulation very complex but thanks to the factorisation theorem [102] it is possible to separate different effects of the processes into hard and soft process. While the hard process or hard scattering describes the main interaction of the partons of the protons with large-scale momentum transfer and therefore with perturbation theory, the soft process describes all effects with smaller-scale momentum transfer which cannot be described with perturbation theory. Taking also into account other effects, like hadronisation, decay of the hadrons, pile-up, and the detector resolution, this chapter will focus on the simulation chain of ATLAS that is used to simulate the different processes taking place inside the detector.

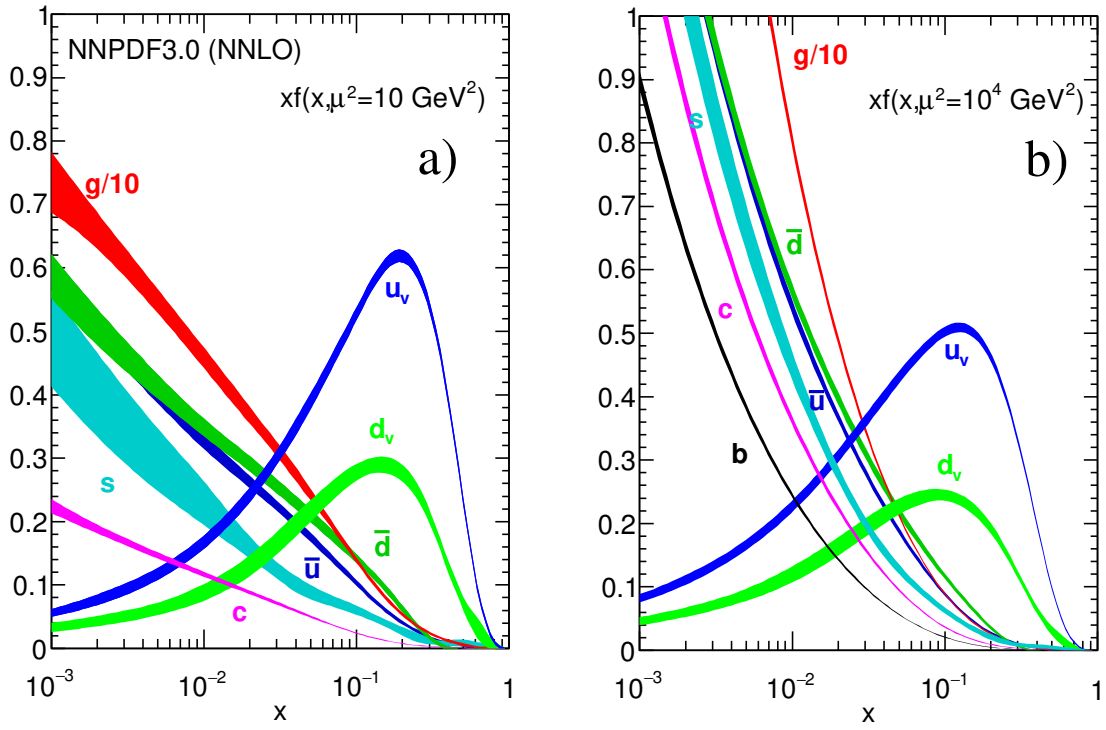
4.1 CROSS-SECTIONS AND PARTON DISTRIBUTION FUNCTIONS

The simulation chain starts with the hard scattering process. First, one needs to introduce a so-called cross-section σ which describes the probability for a certain process to take place. Using the integrated luminosity from Section 3.1, one can calculate the total number of events of this process with Equation (4.1) where N is the total number of events of the process.

$$N = \sigma \cdot \int \mathscr{L} \, \mathrm{d}t \tag{4.1}$$

When looking at a two-body collision producing a particle X, the cross-section is $\sigma_{p_1 p_2 \to X}$. The case of proton-proton collisions introduces some difficulties in the calculation of the cross-section given the composite nature of these particles. Due to the non-perturbative behaviour of the partons inside the proton, the composition of the proton constituents is described by so-called parton distribution functions. They describe the probability of finding a certain parton i with a momentum fraction x_i, contributing to the total momentum x of the proton when probed with a momentum transfer of Q^2. The parton distribution functions are extrapolated from multiple measurements provided by different collaborations, like CTEQ [103], MSTW [104] and NNPDF [105], to different scales of Q^2 using the Dokshitzer–Gribov–Lipatov–Altarelli–Parisi (DGLAP) equations [106–108]. The resulting functions are shown in Figure 4.1.

Using now the already mentioned factorisation theorem, the cross-section can be split into the different parton distribution functions and the partonic cross-section $\hat{\sigma}_{ab \to X}$, which can be calculated



(a): $\mu_F^2 = 10\,\mathrm{GeV}^2$ (b): $\mu_F^2 = 10\,\mathrm{TeV}^2$

Figure 4.1.: Parton distribution functions $xf\left(x, \mu_F^2\right)$ from the NNPDF3.0$_{\mathrm{NNLO}}$ analysis [105, 109].

using perturbation theory. The partonic cross-section is normally expressed with an expansion in the strong coupling

$$\sigma_{ab\hat{\to}X} = \hat{\sigma}_0 + \alpha_s\hat{\sigma}_1 + \alpha_s^2\hat{\sigma}_2 + ... \tag{4.2}$$

where $\hat{\sigma}_0$ describes the Leading Order (LO) or tree-level contribution, $\alpha_s\hat{\sigma}_1$ the Next-to-Leading Order (NLO) corrections and so on. The higher orders in α_s describe the radiative corrections (either real emissions or loops) due to gluon and quark radiation. The real emissions are grouped into two categories: Initial State Radiation (ISR) and Final State Radiation (FSR), describing radiations from the initial/final state partons which can introduce so-called infrared divergences. Virtual corrections or loops also lead to so-called ultraviolet divergences. Both divergences can be regularised by adding unphysical scales, the factorisation and renormalisation scales μ_F and μ_R, respectively. It should be noted that these scales are not intrinsic for QCD but are set to be equal to the momentum transfer of the hard process.

By integrating now all possible parton flavour combinations and their respective momenta, one can express the cross-section $\sigma_{p_1p_2\to X}$ as

$$\sigma_{p_1p_2\to X} = \sum_{a,b}\int \mathrm{d}x_a\mathrm{d}x_b\, f_a\left(x_a, \mu_F^2\right) f_b\left(x_b, \mu_F^2\right) \cdot \hat{\sigma}_{ab\to X}\left(x_a, x_b, \mu_F, \mu_R\right) \tag{4.3}$$

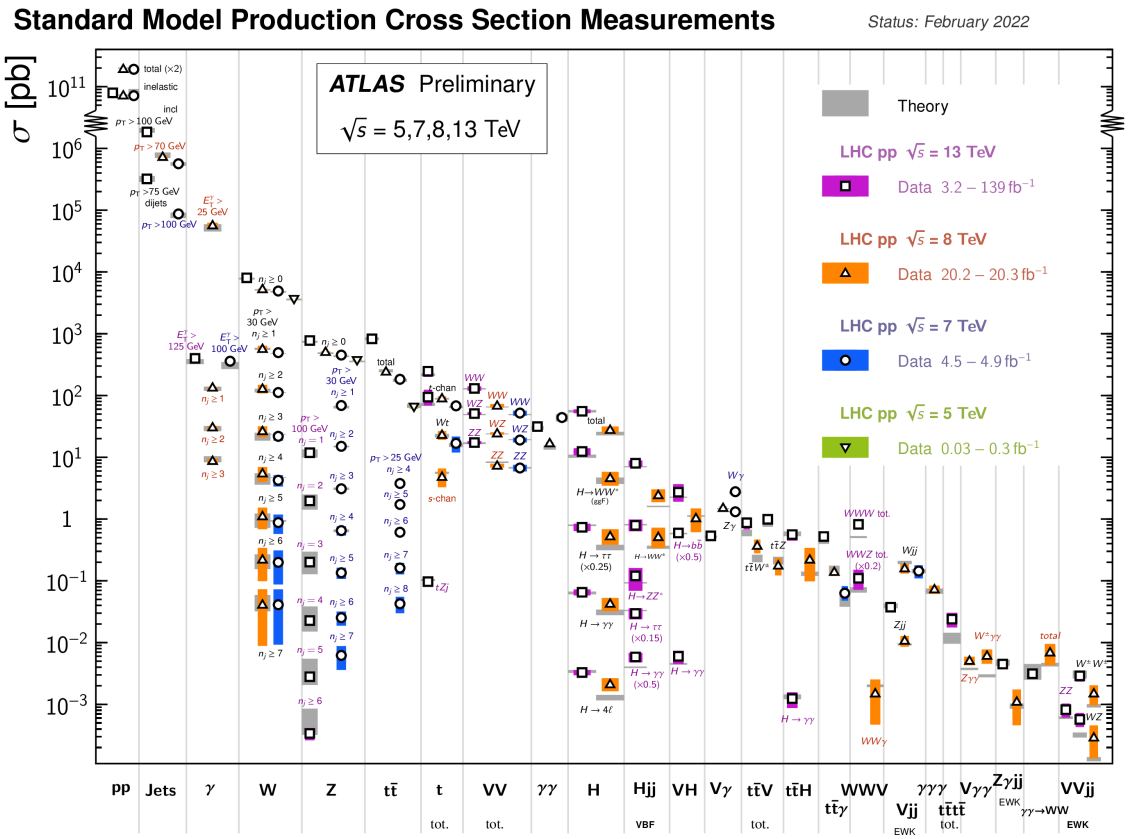


Figure 4.2.: Summary of the production cross-sections for various proton-proton collision processes [110]. Using the methods explained in this section, the theoretical predictions are calculated and compared to their respective measurement from ATLAS.

with the parton density functions f_a and f_b and the momentum x_a and x_b of the interacting partons a and b, respectively. The dependency of the cross-section from the unphysical factorisation and renormalisation scales is due to the finite order of the expansion used in the calculation of the partonic cross-sections in Equation (4.2). For $n \to \infty$, these dependencies disappear. A summary of the production cross-sections for various proton-proton collision processes is shown in Figure 4.2.

One important aspect of the factorisation scheme used is the appearance of b-quarks in such processes. In QCD, this can be handled in two different manners: One can assume that the b-quarks are massive with $m_b > m_{\text{Proton}}$ and therefore do not appear in the initial state and do not have a dedicated parton distribution function. Under this assumption, the b-quarks decouple from α_s running and are not taken into account in its calculation. This becomes very important when calculating processes at lower scales. This scheme is called the Four Flavour Scheme (4FS).
On the other hand, at higher scales, these mass effects become negligible. Therefore, the initial b-quarks can be assumed massless, and b-quarks can be treated similarly to the other lighter quarks. This leads to a dedicated parton distribution function for b-quarks. This scheme is called Five Flavour Scheme (5FS).

4.2 MATRIX ELEMENT AND PARTON SHOWER SIMULATION

In MC generators, the hard process is described with a so-called Matrix Element (ME). This ME is generated by the different generators, which can calculate those automatically, using perturbation theory. In addition to the ME calculation, which covers the already mentioned effects of the hard process and parts of the Parton Shower (PS), namely ISR and FSR, the generators also simulate interactions involving soft gluons taking place in the PS. These soft gluons split, producing more gluons, or fuse to create quarks. With this cycle of splitting and fusing, a cascade-like shower of partons is created. This non-perturbative effect is approximated using different PS models like the CS shower [111] employed by SHERPA [112], or VINCIA [113] used by PYTHIA [114]. Having now two separate simulations, one needs to combine these approaches with a so-called matching or merging scheme. By setting an energy threshold, where all particles above/below are described by the ME/PS respectively, any overlap of particles in the given phase space can be resolved. Other particles like beam remnants or parts of the initial proton can also produce additional, mostly soft, processes. These processes can produce a so-called underlying event which consists of multiple interactions between the remnants. Due to their soft nature, the products of these processes are normally not reconstructed as physics objects but need to be taken into account for specific measurements like the $E_{\mathrm{T}}^{\mathrm{miss}}$ measurement and are modelled using phenomenological models.

4.3 HADRONISATION AND DECAY

After the parton splitting and gluon emissions, the so-called hadronisation sets in at a momentum transfer between the partons of $Q^2 = 1\,\mathrm{GeV}^2$. Due to the confinement of colour-charged particles [37], the hadronisation combines the different particles into colour-neutral mesons and baryons. Being at the non-perturbative energy scale, the hadronisation is simulated using effective models like the Lund string [115] or cluster [116] fragmentation. Both models are based on the principle of the local Parton-Hadron Duality (PHD) [117] which supposes that the flow of momentum and quantum numbers at the hadron level (after hadronisation) tends to follow the flow of the parton level (before hadronisation). In more detail, the cluster fragmentation model, used in SHERPA, applies a split of gluons into quark pairs until the lower cut-off scale is reached. Using these new quark pairs, colourless clusters are formed which decay into hadrons. In contrast, the Lund string model, used by PYTHIA, applies an increasing potential energy between the spatially separated colour-charged particles. A threshold for the upper limit of this potential energy, a new quark pair is created once this limit is crossed. This procedure is repeated iteratively until all particle energies are low enough so that only colourless hadrons remain. In addition, these hadrons can be unstable, resulting in decays into lighter hadrons until all hadrons are stable.

4.4 PILE-UP

Once the simulation of the main process at parton level (before any detector simulation) is done, one needs to also take pile-up into account. Pile-up, shortly introduced already in Section 3.1, describes the effects and interactions of other protons in the same bunch (in-time) and protons of neighbouring bunches (out-of-time), which have not taken part in the main interaction process. The pile-up simulation [118] is done using PYTHIA [114] by including single, double, and non-diffractive components and overlaying multiple interactions at the same time. For the in-time pile-up, this is done similarly for all sub-detectors, while for out-of-time pile-up, the simulation is done separately per sub-detector due to the different detector responses for this type of pile-up. Given the dependency of the effect on the beam conditions and the number of interactions per bunch crossing μ, the respective simulations are reweighted to the μ of the corresponding data.
Another way to estimate pile-up effects is by considering so-called zero-bias runs. In these types of runs, data is taken with a random trigger without any pile-up suppression techniques making this a purely data-driven approach. These recorded background events are then overlaid on top of the simulation of signal events to recreate the environment in which the corresponding data will be taken.
In ATLAS, the simulation approach is used for proton-proton collisions, while the data-driven approach is used in heavy ion collisions due to their enormous amount of background particles from pile-up.

4.5 DETECTOR SIMULATION

The detector simulation of the ATLAS experiment and its sub-detectors [119] is one of the most important, but also the most computing-intense step of the simulation chain. ATLAS makes use of two different types of detector simulations: the so-called full simulation using GEANT4 [120, 121] and a fast simulation using the Fast ATLAS simulation Atlfast-II (AF2) [122].
GEANT4 is a C++-based simulation software that can model the passing of particles through matter. By precisely simulating the passing of the particles through different materials and their energy loss in those, GEANT4 can model the pass-through of all particles coming from the simulations through the full ATLAS detector. To properly simulate the traversing process, a very detailed model of the ATLAS detector, including all materials and the magnetic fields, is given to GEANT4 which enables the proper simulation.
While the full GEANT4 simulation is very precise and describes the passage of particles through ATLAS exceptionally well, the computing cost for simulating a full event is very high. Due to the large number of events that have to be simulated, a fast simulation algorithm with less thorough modelling is applied to generate the number of simulated events needed. The largest part of the computing costs arises from the simulation of particle showers in the calorimeters and the tracking in the ID and the MS. By replacing the particle shower simulation with the FastCaloSim [123] utilising parametrised calorimeters, a huge gain in computing costs can be achieved. In addition,

another order of magnitude in computing time can be gained when using the fast-tracking simulation FATRAS [124]. It employs simplified parametrisations of physics processes as well as a simplified geometry of the ATLAS track reconstruction algorithms.

The output of both detector simulations is in the same format as the data taken by the detector so they can be easily fed into the trigger and DAQ pipeline described in Section 3.2.5. With the same format, simulation and data can be directly compared and are sent to the object reconstruction algorithms to reconstruct physics objects using the information collected at the so-called detector level.

Part II

METHODOLOGY

5

OBJECT RECONSTRUCTION AT THE ATLAS DETECTOR

After the DAQ, the collected data is used in reconstruction algorithms. The detector signals created by the passage of particles through the different sub-detectors are key for the reconstruction of interesting events and the identification of their constituent particles. One can split the signals of the sub-detectors into two groups: tracking and calorimetry information. While the former data are mainly produced by the ID and the MS, the calorimetry information is collected by the ATLAS calorimetry system. The raw information from the sub-detectors can be utilised in various dedicated reconstruction algorithms, which aim to rebuild different physics objects from the measured detector signals. In addition to electrons and muons, the list of physics objects also includes particle tracks, the Primary Vertex (PV), jets, and $E_{\mathrm{T}}^{\mathrm{miss}}$. Furthermore, photons are also physics objects that are reconstructed, but will not be covered here due to their negligibility for this thesis.

This chapter introduces the different object reconstruction algorithms employed in ATLAS.

5.1 TRACKS AND PRIMARY VERTICES

Track reconstruction is among the first reconstruction algorithms applied to the data/simulation, for the reason that nearly all other reconstruction processes rely on the tracks. The primary track algorithm [125, 126] utilises the full ID, starting from the PD and SCT and extending outwards to the TRT. When charged particles traverse the ID, they deposit energy that can be measured as a signal. The signals (or hits) from the PD and SCT are then organised into clusters employing a Connected Component Analysis (CCA) [127]. Using these clusters, three-dimensional space-points are created, indicating where the charged particle passed through the detector's active material. In the PD, each cluster corresponds to one space-point, while in the SCT, a combination of clusters from both sides of the strip layer must be used to determine the space-points.
Track seeds are generated by grouping three space-points. This approach optimises the available potential combinations while also allowing for an initial rough estimation of the particle momentum. To assess the Impact Parameters (IPs)[1] of a track seed relative to the center of the interaction region, an assumption of a flawless helical trajectory within a uniform magnetic field is made. Subsequently, a combinatorial Kalman filter [128] is employed to construct potential tracks from the seeds. This involves incorporating additional spatial data from the remaining layers of the pixel and silicon detectors, provided they align with the initial track trajectory. Utilising these available tracks and the

1 The transversal IP d_0 refers to the minimum distance between a track and the PV in the transverse plane, while the longitudinal IP z_0 represents the longitudinal separation between the PV and the point on the track from which d_0 is measured.

supplementary spatial information results in a notably high reconstruction efficiency for primary particle tracks, concurrently eliminating a substantial number of tracks formed solely from random groupings of space-points. Once all realistic combinations of space-points have been identified, multiple track candidates can occur due to overlapping or incorrectly assigned space-points. This ambiguity is resolved [129] by ranking the track candidates using a so-called track score. A track is given a score according to the alignment accuracy of hits with the selected trajectory, as well as the fulfilment of specific quality criteria. For tracks passing these criteria, corresponding compatible TRT hits are identified to further extend the track. Subsequently, a global χ^2 fit is employed to accurately determine the track parameters. The outcome of this tracking sequence is a set of definitive tracks, with each track representing the path of an individual charged particle originating from the interaction point.

In addition to this inside-out tracking approach, another distinct tracking algorithm known as the outside-in tracking is used. This reversed approach starts with the TRT and expands inwards towards the PD and SCT. This method is particularly effective in identifying tracks primarily arising from distant secondary decays which makes it a vital technique for detecting particle tracks that could be oriented towards the interaction point, even if they do not necessarily originate from there. TRT tracks are automatically excluded if they do not have an extension to the PD and SCT.

Making use of the reconstructed tracks, PVs can be reconstructed using an algorithm [130] which involves two main steps: vertex finding and vertex fitting. Initially, the set of fully reconstructed tracks is projected inwards to locate a common and distinct origin within the beam spot. Once this shared vertex point is identified, an iterative χ^2 minimising fit is applied to accurately determine its position. After obtaining the vertex position, any tracks that do not align with the vertex are removed and used to identify another vertex. Ultimately, the primary vertex is determined based on the highest sum of squared transverse momenta (p_{T}^2) from its associated tracks. The remaining reconstructed vertices are classified as pile-up vertices. This criterion relies on the assumption that charged particles resulting from hard scattering interactions generally possess larger average transverse momenta in comparison to those generated in pile-up collisions.

5.2 ELECTRONS

Electrons are among the few particle types that are stable and primarily interact through the electromagnetic force. In this section, a brief explanation of electron reconstruction, calibration, identification, and isolation is provided, highlighting their significance in the $t\bar{t}H$ analysis and the estimation of fake lepton backgrounds. A more detailed description, including the exact definition of Working Points (WPs) can be found in Ref. [131].

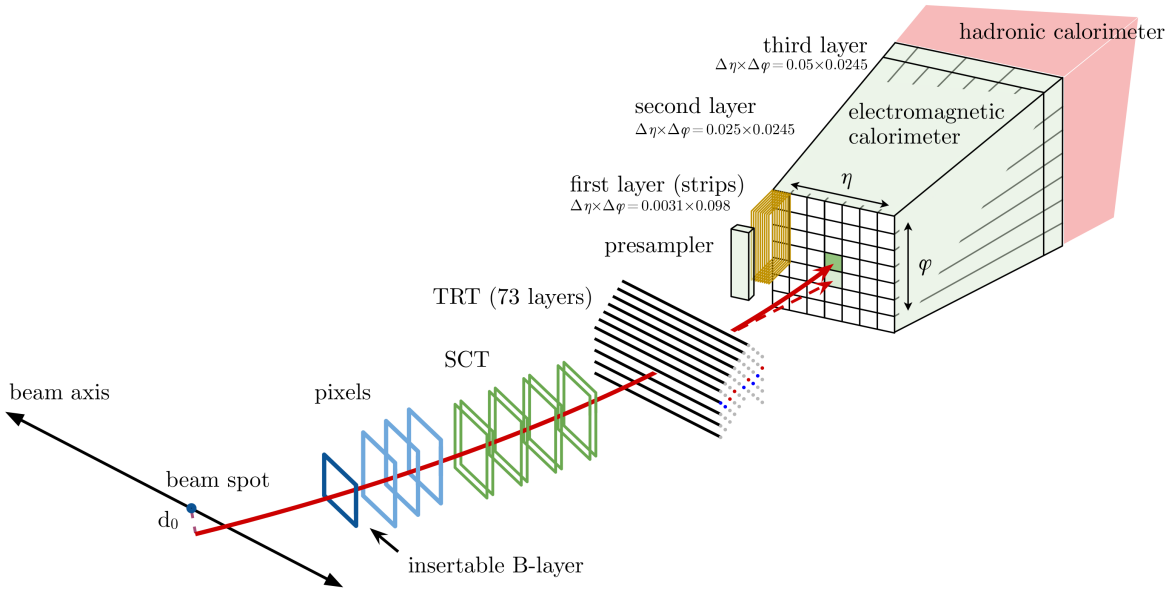


Figure 5.1.: Schematic representation of the path of an electron through the ATLAS detector. The electron path is shown with the solid red line, traversing first the ID and entering then the ECal. The dotted red line shows the path of a photon produced due to bremsstrahlung inside of the TRT [132].

Reconstruction

When an electron traverses a material, it has the potential to lose a considerable portion of its energy through bremsstrahlung. This process involves the emission of a photon, which can further convert into an electron-positron pair, subsequently interacting with the detector material. These emitted positrons, electrons, and photons are typically located closely around the main trajectory of the electron and are often reconstructed as components of a single electromagnetic cluster, which is a clustered group of energy depositions inside the ECal. These interactions can occur within the volume of the ID or even within the beam pipe, resulting in the generation of multiple tracks. Alternatively, these interactions might occur downstream of the ID, affecting the energy shower within the calorimeter. Consequently, a scenario arises where multiple tracks can be produced and correlated with the same electromagnetic cluster, all stemming from the same primary electron source. Using the high-granularity ECal and the tracks provided by the ID, the electron reconstruction is based on three key elements that define the signature of electrons. These elements include localised clusters of energy deposits detected in the ECal, charged-particle tracks within the ID, and a precise alignment of those in the $\eta - \phi$ plane. The latter is used to build the final electron candidates. Due to the lower granularity of the ECal outside the barrel region and the geometric acceptance of the ID, the electron reconstruction explained above is used only in the central region ($|\eta| < 2.47$) of ATLAS. A schematic representation of the path of an electron through the detector is given in Figure 5.1.

The reconstruction of an electron begins with the selection of clusters of energy depositions in topologically connected ECal and HCal cells, so-called Topological Clusters (Topo-Clusters). These Topo-Clusters are built in an iterative way, combining so-called proto-clusters with surrounding cells if a certain significance threshold is reached. If two proto-clusters contain the same cell with a

significance above the threshold, they are merged. If a proto-cluster contains two cells with $E_{\text{cell}}^{\text{EM}} > 500\,\text{MeV}$ and each has four neighbours, all with a smaller energy deposition, the proto-cluster is split. If a Topo-Cluster is found, the algorithm attempts to match it to reconstructed tracks from the ID in the $\eta - \phi$ plane.
Based on the Topo-Clusters and the associated reconstructed tracks, so-called superclusters are built. The reconstruction starts from corresponding supercluster seeds, which are found by iterating the Topo-Clusters and selecting those, which have $E_{\text{T}} > 1\,\text{GeV}$ and a matched track with at least four hits in the ID. Afterwards, all Topo-Clusters around the seed are used to find satellite clusters, stemming from bremsstrahlung, by matching them to the seeds if found within $\Delta\eta \times \Delta\phi = 0.075 \times 0.125$. The resulting combination of seeds and satellite clusters is than reconstructed as supercluster, from which the electrons are built.

Energy Calibration

To improve the energy resolution of electrons and to correct for differences in the electron energy scale in data and MC simulation, a chain of calibrations is applied. The calibrations account for energy loss in the tracker material, energy deposition in neighbouring cells relative to the energy cluster, and general energy loss outside the ECal.
In the first step, the energy of the superclusters is corrected, which takes into account the properties of the shower development, determined using a multivariate regression algorithm [133]. The regression algorithm is trained on simulated electrons and serves to improve the energy resolution while simultaneously reducing the impact of the material in front of the calorimeter, which is applied to both data and MC.
Another correction that is needed is for the energy scale. This correction can be described as a constant factor for the simulated electron energy which is applied to match the simulated electron energy with the one measured in data, derived using $Z \to e^+e^-$ and $J/\Psi \to e^+e^-$.

Identification

Differentiating electrons from the prompt decay of particles, e.g. $W^{\pm}$ boson, from other background sources, such as mis-identified hadrons, electrons from photon conversions or heavy flavour decays, plays an important role in the $t\bar{t}H$ measurement to suppress the background. The identification is based on a Likelihood (LH) approach which uses different quantities as input, e.g track conditions and calorimeter information, as well as a Probability Density Function (PDF). These PDFs are derived using simulated $Z \to e^+e^-$ and $J/\Psi \to e^+e^-$ events. A LH is built for both the signal (L_S) and the background (L_B), each with its own PDF.
Using the resulting LHs for the electron candidate, a discriminant d_L is defined as

$$d_L = \frac{L_S}{L_S + L_B} \tag{5.1}$$

where L_S and L_B are the signal and background LHs for the electron candidate. With a sharp peak at 1 for the signal and at 0 for the background, the discriminant is transformed using an inverse sigmoid function to flatten out the distribution for an easier definition of the electron identification WPs. More details about this can be found in Ref. [131].

Isolation

Prompt electrons are typically separated from other physics objects due to their characteristic signature of low activity in the surrounding area in η and ϕ. To separate the signal prompt electrons from the hard-scattering process from other backgrounds, such as the semileptonic decay of heavy quarks and hadrons faking electrons, so-called isolation variables are constructed. These variables quantify the amount of activity in the region around the reconstructed electron. Two types of isolation techniques are considered: calorimeter-based and track-based isolation.
The calorimeter-based isolation employs a cone approach. Around the electron, a cone is defined in the $\eta - \phi$ plane with a certain radius denoted as ΔR [Equation (3.3)][2]. Additionally, Topo-Clusters are constructed inside and around the cone. These clusters originate from cells with energy deposits exceeding four times their noise-level threshold. Subsequently, cells around the center of the Topo-Cluster are examined. If these cells have energy deposits exceeding twice their expected noise-level threshold, they are added to the Topo-Cluster. This process is repeated in all three spatial directions, encompassing both the ECal and HCal, until no further cell satisfies this requirement. In the final step, a final shell of cells around the cluster is incorporated. Topo-Clusters with a barycenter inside the previously defined cone are included in the calculation of isolation variables, such as the isolation transverse energy $E_{\mathrm{T}}^{\mathrm{Iso}}$. It should be noted that the electron's contribution is subtracted from these variables.
For the calculation of track-based isolation variables, tracks with at least $p_{\mathrm{T}} > 1\,\mathrm{GeV}$ are used. Additionally, the tracks need to satisfy basic track-quality criteria, and $|z_0 \sin\theta| < 3\,\mathrm{mm}$ is required to reduce the effect of pile-up. Here, z_0 denotes the longitudinal IP of the track from the primary vertex. Using the track of the reconstructed electron as the main axis, another cone with a radius of ΔR is placed around it. All tracks inside this cone are considered when calculating track-based isolation variables, such as the isolation transverse momentum $p_{\mathrm{T}}^{\mathrm{Iso}}$.

5.3 MUONS

In the ATLAS detector, muons are unique in their ability to penetrate and the outer layer of the calorimeter system and still be detectable as they are minimum-ionising particles.[3]. This distinctive property of muons leads to the presence of clear tracks in the ID and MS and characteristic energy deposits in the calorimeter. For these muon candidates, identification and isolation algorithms are

2 The chosen ΔR depends on the isolation WPs being used.
3 In rare instances, the shower of highly energetic hadrons can extend into the MS.

employed to effectively distinguish prompt muons originating from the hard-scattering process from other sources of muons, such as those produced by the decay of heavy quarks. A more detailed description of the full process of muon reconstruction, calibration, identification, and isolation can be found in Ref. [134].

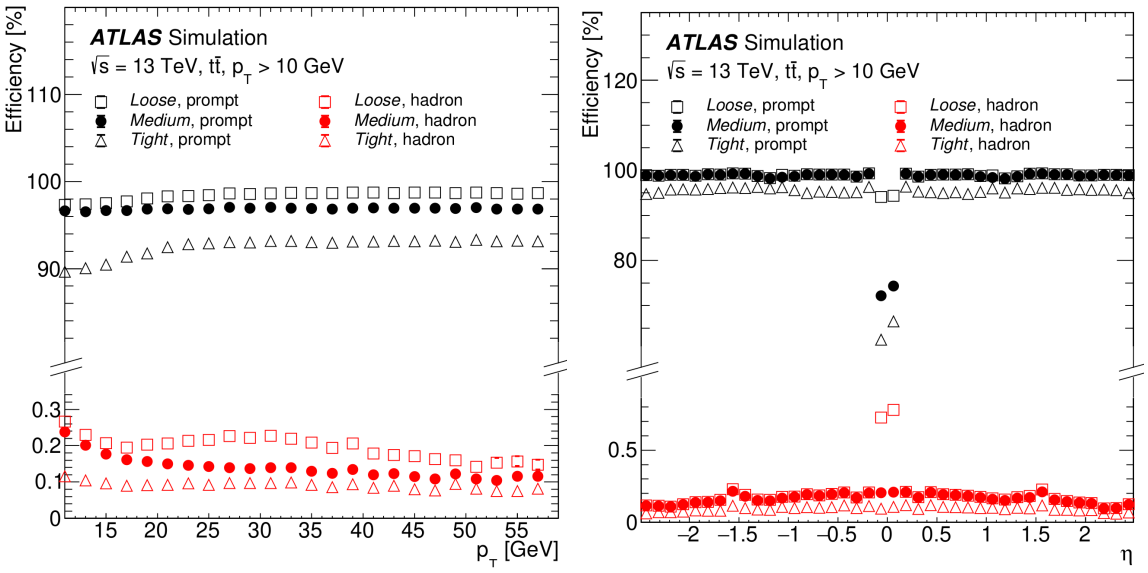
Reconstruction

Muon reconstruction requires not only the tracks provided by the ID but also tracks from the MS. The process begins with the identification of hits in the individual MS stations, which are used to reconstruct straight local track segments. These segments are identified using a Hough transform [135] and are used to build preliminary track candidates, leveraging precision measurements of the track's bending radius and information from the trigger detectors.
As a final step in creating the muon track candidates, a χ^2 fit is applied to the muon's trajectory within the toroidal magnetic field, accounting for possible effects of misalignment and material interactions in the MS. Subsequently, any outlier hits along the track trajectory are removed, and the fit is recalculated. Ambiguities are resolved by removing tracks that share hits with higher-quality tracks. These final muon tracks are then refitted, taking into account the possible energy loss in the calorimeter system.

For the actual reconstruction of muons, five different types of muons are considered, each with its respective reconstruction strategy. These five muon types are: combined, inside-out combined, segment-tagged, calorimeter-tagged, and MS extrapolated.
The first type of muons, combined muons, is constructed by matching tracks in the MS and ID using a fit-based on their underlying hits in the respective sub-detector. Similar to the reconstruction of muon candidates, a possible loss of energy in the calorimeter system is taken into account.
In contrast to combined muons, inside-out muons are reconstructed by extrapolating tracks from the ID into the MS. If at least three MS hits can be aligned with the track, the track is reconstructed as a muon.
Segment-tagged muons are identified using a track in the ID that tightly matches in angular distance with at least one MS segment. Calorimeter-tagged muons do not rely on the MS. They are identified using an ID track with an energy deposition in the calorimeter system consistent with a minimum-ionising particle.
Finally, a track is reconstructed as an MS-extrapolated muon if the track in the MS cannot be matched to an ID track.

Identification

After the initial reconstruction, the muon candidates are selected for each type via a dedicated set of requirements, special for each of the types. These requirements are given in tight, medium, and loose



(a): Identification efficiency vs p_{T}.

(b): Identification efficiency vs η.

Figure 5.2.: Efficiency of the muon identification WPs for muons with an ID track of at least $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ from $t\bar{t}$ events as a function of p_{T} (a) and η (b) [134]. In black, the efficiency for prompt muons is shown, while in red, the efficiency for muons from light hadron decays are shown. A significant drop for all WPs can be observed for $\eta < 0.1$ due to a gap in the coverage of the MS.

WPs. More WPs are defined to match the vastly different needs of analyses in terms of prompt muon identification efficiency, momentum resolution, and background rejection (non-prompt muons). For non-prompt muons, special WPs are defined to separate in addition the muons from light hadron decays (typically with low-quality muon tracks) from muons from a heavy-flavour decay (typically with high-quality muon track). A detailed description of all WPs can be found in Ref. [134]. The efficiency for the aforementioned WPs are depicted in Figure 5.2

Isolation

Similar to prompt electrons, prompt muons exhibit low hadronic activity in their vicinity. While muons originating from light hadron decays are already rejected by the identification criteria, the isolation process targets non-prompt muons arising from heavy hadron decays. These non-prompt muons also meet high-quality identification requirements, as previously mentioned. In the case of a muon resulting from a heavy hadron decay, the muon is referred to as soft and typically occurs with multiple energy depositions in the calorimeter near it.

Much like the identification process, isolation also employs different WPs, each with variations in their acceptance of prompt muons, rejection of non-prompt muons, and performance near other objects. Furthermore, similar to electrons, isolation criteria can be defined based on both track-based and calorimeter-based measurements.

The track-based isolation variables for muons are comparable to the isolation of electrons. Around the muon track, a cone in the $\eta - \phi$ plane is placed with a radius of ΔR. The isolation variable is calculated by summing up all transverse momenta of all ID tracks inside the cone, excluding the muon track. Depending on the selected isolation, ΔR is either

$$p_{\mathrm{T}}^{\mathrm{cone20}}: \ \Delta R = 0.2 \quad \text{or} \quad p_{\mathrm{T}}^{\mathrm{varcone30}}: \ \Delta R = \min\left(\frac{10\,\mathrm{GeV}}{p_{\mathrm{T}}^{\mu}}, 0.3\right) \tag{5.2}$$

where p_{T}^{μ} is the transverse momentum of the muon. Both of these variables are utilised across various WPs. The latter variable is optimised for scenarios where jets and leptons within the event are expected to be close to each other. Additionally, the minimum p_{T}^{μ} of the isolation variable can be adjusted to enhance the rejection of hadronic activity. For all isolation variables, a minimum track transverse momentum of either $500\,\mathrm{MeV}$ or $1\,\mathrm{GeV}$ is required. Because these variables are largely independent of pile-up effects, all isolation WPs include at least one of the track-based isolation variables.

Calorimeter-based isolation functions in the same way as described for electrons in Section 5.2. In a cone around the muon track with $\Delta R = 0.2$, the energy deposited in all Topo-Clusters in the calorimeter, with their barycenters inside the cone, is taken into account and summed. This new isolation variable is referred to as $E_{\mathrm{T}}^{\mathrm{topocone20}}$. The energy deposition from the muon is subtracted to account for contributions from pile-up and the underlying event, necessitating the estimation and proper subtraction of their effects. This can result in a decreased energy resolution due to the substantial applied relative correction. Consequently, calorimeter-based isolation tends to be more sensitive to pile-up effects compared to track-based isolation. The advantages of calorimeter-based isolation variables lie in the ability to measure neutral particles and contributions from particles whose p_{T} is below the threshold of the ID track.

To improve performance and exploit the advantages of both types of muon isolation, the variables are normally combined in the different WPs. With the particle-flow algorithm[4] [136] it is also possible to remove overlapping contributions which results in a smaller correlation of the variables.

A more detailed description of all the available WPs and their definition can be found in Ref. [134].

5.4 JETS

Colour-charged particles such as quarks and gluons cannot be observed as free particle in the detector due to the confinement of QCD. Therefore, they hadronise, forming colourless bound states. Some of these bound states, or hadrons, have a limited lifetime after which they decay into further particles, which can also decay further. Combined with other sources of particles, like fragmentation, this avalanche of collimated particles can be seen in the detector and reconstructed as hadronic jets by using both Topo-Clusters from the calorimeter systems and the tracks from the ID.

The definition of jets is an important aspect to consider. They are not unique objects but are highly dependent on the algorithms employed for their reconstruction. ATLAS utilises various jet types,

4 The particle-flow algorithm is also used in the jet reconstruction. A description can be found in Section 5.4

such as EMTopo, PFlow, and VR track. However, given the specific focus of this thesis on PFlow jets, the following explanation of their reconstruction and calibration will be exclusively provided for them.

Particle Flow Algorithm

The foundation of PFlow jet reconstruction is the particle flow algorithm [136]. This algorithm, a cell-based subtraction algorithm, is employed to remove overlaps in the measurements of energy and momentum between the ID and the calorimeter systems. The algorithm starts with tracks and Topo-Clusters in the calorimeter. Using a specific track selection, the tracks are matched to Topo-Clusters within the calorimeter. Making use of the position of the Topo-Cluster and the momentum of the track, the expected energy deposition of the same particle that generated the track in the calorimeter is estimated. One needs to take into account that single particles often deposit energy in more than one Topo-Cluster. To incorporate this effect, the algorithm checks the probability for this to happen and decides if more Topo-Clusters need to be added to the track/Topo-Cluster pair. After building these pairs/systems of tracks and Topo-Clusters, the estimated deposited energy of the track in the Topo-Clusters is subtracted cell by cell. As the last step, the algorithm checks if the remaining energy in the Topo-Clusters is consistent with the expected energy of the shower fluctuations of a single particle. If this is the case, the Topo-Clusters remnants are removed. A more detailed explanation is provided in Ref. [136].

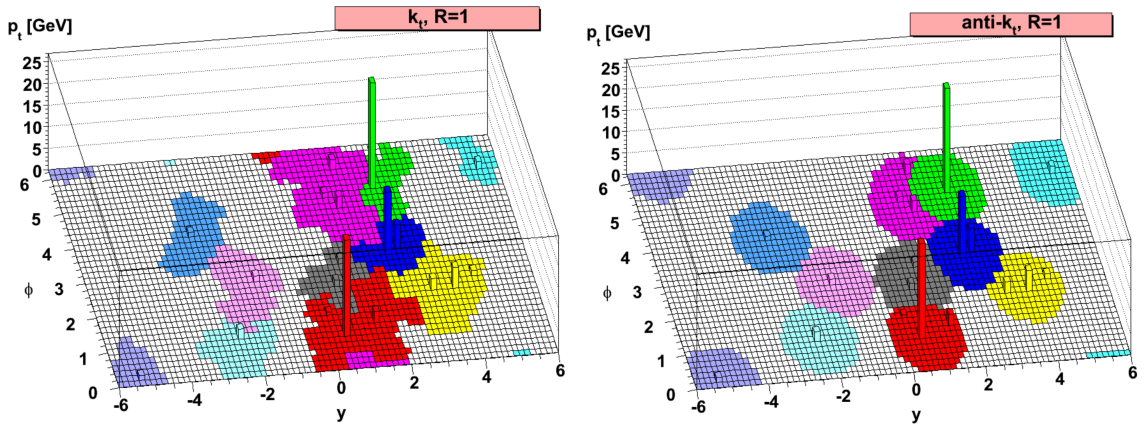
Reconstruction

The actual jet reconstruction builds on top of the already-explained particle flow algorithm, using the anti-k_t algorithm [137] to cluster jets, which is a Infrared and Collinear (IRC) safe sequential recombination algorithm. The algorithm itself first defines two distances: d_{ij} [Equation (5.3)] and d_{iB} [Equation (5.4)], where d_{ij} defines the distance between objects i and j, d_{iB} defines the distance between an object i and the beam B. Using these two measures, the algorithm recursively identifies the smallest distance between all objects (including the beam B). If the smallest distance is between two objects, so $d_{ij} < d_{iB}$, the two objects are combined while, if $d_{iB} < d_{ij}$, the object is called a jet and removed from the list of objects. This procedure continues until no further objects are available.

$$d_{ij} = \min\left(k_{ti}^{2p}, k_{tj}^{2p}\right) \frac{\Delta_{ij}^2}{R^2} \tag{5.3}$$

$$d_{iB} = k_{ti}^{2p} \tag{5.4}$$

with $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$. Here, k_{ti} denotes the transverse momentum of the particle i, y_i the rapidity, and ϕ_i the azimuthal angle. Further, R is the radius parameter which defines the size of the cone and p is added to steer the relative power of the energy concerning Δ_{ij}. Different values for p yield different results: For $p = 1$, the standard k_t algorithm [138] is constructed while for $p = 0$



(a): k_T algorithm using $R = 1$.

(b): Anti-k_T algorithm using $R = 1$.

Figure 5.3.: An example parton-level event generated with HERWIG [140] in the $y - \phi$ plane. As an addition, many random soft particles, or "ghosts", are added. Depicted are the calorimeter cells and the deposited energy. The k_t (a) and the anti-k_t algorithm (b) are used to build jets. Using a radius parameter of $R = 1$, one can see the difference in shapes between the two algorithms. While the jet shape of the k_t jets is affected by the soft particles, the anti-k_t jet shape is dominated by the hard particles [137].

the Cambridge-Aachen algorithm [139] is constructed. In general, positive values of p yield a similar behaviour for soft radiation as the k_t algorithm. In the case of negative values for p, the behaviour is similar for all values of $p < 0$ and will be shown in the following. The new algorithm with $p = -1$ is the anti-k_t algorithm.

Using an event with a few hard (well-separated) and soft particles as an example, one can observe the benefits of the new anti-k_t algorithm. Due to the distance metrics used and its dependence on transverse momentum, the soft particles tend to cluster primarily with nearby hard particles rather than with other soft particles. If only one hard particle is within a distance of $2R$, the soft particles will accumulate around this particle, resulting in a perfectly conical jet. If a second hard particle is present, it is still possible for two perfectly conical jets to be clustered as long as their separation is large enough. When both jets are close to each other, one jet can still maintain its conical shape while the other one is truncated. When the distance is equal to both hard particles, the resulting jets will both be truncated by straight lines due to the overlap of the jets. The key takeaway here is the shape independence of the jet form from the soft particles, while hard particles can still affect the form of the jet. These effects are visualised in Figure 5.3.

Utilising the anti-k_t algorithm, ATLAS employs two jet collections with radius parameters: $R = 0.4$, referred to as small-R jets, and $R = 1.0$, referred to as large-R jets. Here, the focus is set primarily on the small-R jets due to their greater importance to this thesis. For the small-R jets, the inputs for the jet reconstruction algorithm include the Topo-Clusters from the calorimeter, which survive the particle flow algorithm's subtraction phase (Section 5.4), and the selected tracks.

The selection of tracks is based on a criterion related to the longitudinal IP z_0. By imposing the requirement $|z_0 \sin(\theta)| < 2\,\text{mm}$, a significant portion of pile-up tracks and their associated energy deposits in the calorimeter are filtered out while retaining the tracks originating from the hard scatter-

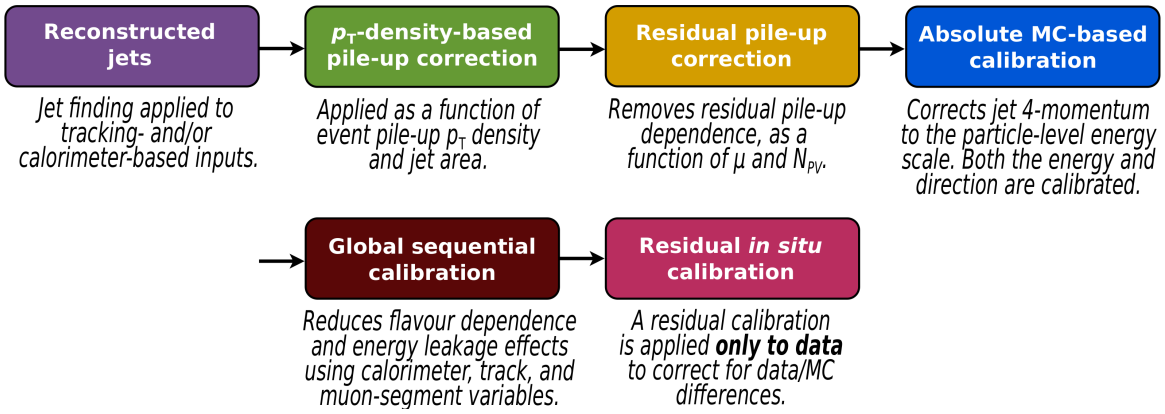


Figure 5.4.: Schematic view of the different stages of the JES calibration [143].

ing process. To ensure the accurate positioning of the calorimeter cells, the η and ϕ positions of the cells are recomputed concerning the hard scattering vertex before the jet-finding procedure begins. The association of tracks to a jet is achieved using ghost association [141], which treats each track as a four-vector with infinitesimally magnitude during the jet reconstruction and assigns it to the corresponding clustered jet. This algorithm is implemented in the FASTJET package [142].
As a last requirement after the jet reconstruction, the jets that are taken further for calibration need to have at least $p_{\mathrm{T}} > 7\,\mathrm{GeV}$ and $|\eta| < 4.5$. After calibration, the jets must have at least $p_{\mathrm{T}} > 20\,\mathrm{GeV}$. In addition to the reconstructed jets, so-called truth jets are clustered. These truth jets are built on the particle level using stable[5] final-state particles (excluding muons, neutrinos, and soft particles from pile-up). Using the same final selection criteria as for the reconstructed jets, the truth jets are angular distance matched to reconstructed jets with $\Delta R < 0.3$. Truth jets are used in the calibration of the absolute MC-based calibration.

Calibration

Following the jet reconstruction, multiple steps of calibrations are needed to take different effects into account, like pile-up corrections. The calibration of jets differentiates here between the Jet Energy Scale (JES) and Jet Energy Resolution (JER) calibration. While both calibrations are rather complex, a short explanation will be given here for the JES calibration. A more detailed explanation of both calibrations is given in Ref. [143]
For the JES calibration, a schematic of the different stages is shown in Figure 5.4.

In the first step, a p_{T}-density-based pile-up correction is applied. Using the jet area in the calorimeter, one can estimate the relative number of particles from pile-up that are clustered together with the jet. To remove the contributions of these ghost particles, the median p_{T} density, ρ, is calculated in the $y - \phi$ plane using the k_t algorithm, which, in comparison to anti-k_t, uses $p = 1$ is more sensitive to soft particle contributions. After removing these contributions from the jets, one can observe a residual dependence on the pile-up for the corrected jets which is also removed. This residual pile-up

5 Stable refers here to the colourless bound states originating from the hadronisation.

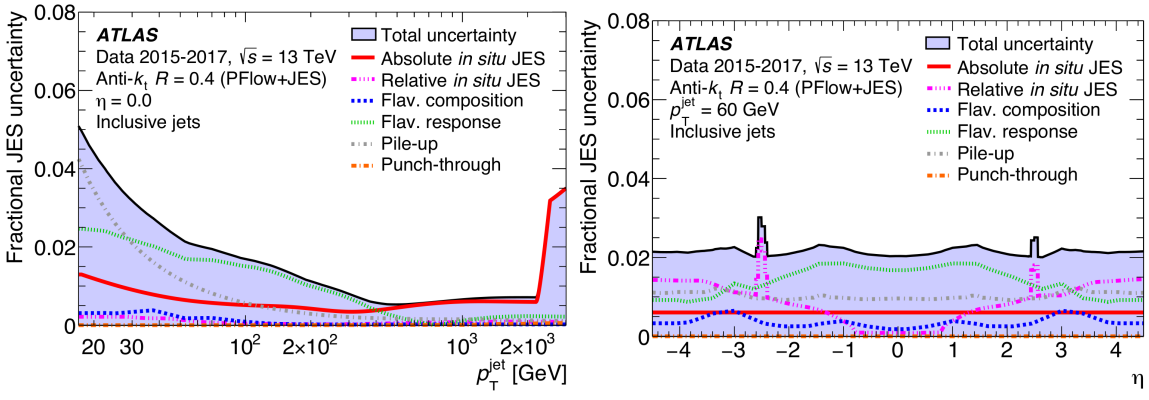
effect can be parametrised as a function of the number of reconstructed PVs, N_{PV}, and the average number of collisions per bunch crossing μ, and is estimated using MC simulation.

The next step in the calibration is the absolute Monte Carlo Jet Energy Scale (MCJES) calibration. It corrects the absolute JES and η of the reconstructed jet four-momentum on average to their matched truth jets. This takes into account different effects, e.g. energy losses in the passive material, and reconstruction issues in the jet η reconstruction. For this calibration, MC-simulated dijet events from PYTHIA8 [114] are used, excluding jets from dense areas. This isolation is achieved by applying additional ΔR criteria on the distance between the jets.

Although the MCJES adjusts the jet p_{T} and η, the calibration is implemented in p_{T} and η bins, without considering the variations caused by different jet flavours. These fluctuations arise from disparities in the kinematics of the jet constituents. Quark-initiated jets often feature hadrons with a higher fraction of the jet's p_{T}, whereas gluon-initiated jets tend to include more particles with softer p_{T}. To address these differences, the Global Sequential Calibration (GSC) is employed. This calibration method also employs MC-simulated dijet events to improve the jet p_{T} resolution and diminish associated uncertainties. It achieves this by no longer relying on observables constructed from information provided by the ID, the calorimeters, and the MS to built the response.

The final step in the JES calibration is the *in-situ* calibration. While all the previous calibrations are applied to data and MC, the *in-situ* calibration takes residual differences between data and MC into account and is only applied on data. These final corrections are derived from events with objects in the final state, which are well defined and can be reconstructed with high precision. Typically, Z + jets or γ + jets events are employed.

After all these calibrations, the uncertainties on the reconstructed jet p_{T} and η can be estimated and are depicted in Figure 5.5.



(a): Uncertainties with respect to jet p_{T} at $\eta = 0$.

(b): Uncertainties with respect to jet η at $p_{\mathrm{T}} = 60\,\mathrm{GeV}$.

Figure 5.5.: Fractional JES systematic uncertainty and uncertainties from the different components for anti-k_t $R = 0.4$ PFlow jets [143]. The uncertainties are shown as a function of jet p_{T} with $\eta = 0$ (a) and as a function of jet η for $p_{\mathrm{T}} = 60\,\mathrm{GeV}$ (b).

5.5 MISSING TRANSVERSE MOMENTUM

$E_{\mathrm{T}}^{\mathrm{miss}}$ is an important variable for many ATLAS analyses looking for new exotic or BSM particles that do not interact with the detector material and, therefore, cannot be measured. A SM particle with such properties is the neutrino. For example, when produced in the decay of a $W^{\pm}$ boson, the neutrino interacts weakly with the detector material and is, therefore, not measured. In the following, a short explanation of the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction is given. A more detailed explanation can be found in Ref. [144].

The reconstruction of $E_{\mathrm{T}}^{\mathrm{miss}}$ is based on the momentum conservation in the $x-y$ plane, which must in total zero due to no momentum in this plane before collision. To correctly reconstructed the difference from zero measured in the event, two contributions need to be taken into account: hard contributions, reconstructed and calibrated physics objects like jets, electrons, τ-leptons, and muons; and soft contributions, such as reconstructed charged-particle tracks that are not associated with a hard object but with the hard-scattering vertex.
Making use of the momentum vectors of these contributions, one can calculate the $E_{\mathrm{T}}^{\mathrm{miss}}$ in the $x-y$ plane, depicted in Equation (5.5), as well as the vectorial sum of the $E_{\mathrm{T}}^{\mathrm{miss}}$, $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$, shown in Equation (5.6).

$$E_{x(y)}^{\mathrm{miss}} = -\sum_{i\in\text{Hard contributions}} p_{x(y),i} - \sum_{j\in\text{Soft contributions}} p_{x(y),j} \tag{5.5}$$

$$\vec{E}_{\mathrm{T}}^{\mathrm{miss}} = \underbrace{\underbrace{\underbrace{\underbrace{\underbrace{-\sum_{\text{Selected electrons}} \vec{p}_{\mathrm{T}}^{\,e}}_{\vec{E}_{\mathrm{T}}^{\mathrm{miss},e}} - \underbrace{\sum_{\text{Accepted } \tau\text{-leptons}} \vec{p}_{\mathrm{T}}^{\,\tau_{\mathrm{had}}}}_{\vec{E}_{\mathrm{T}}^{\mathrm{miss},\tau_{\mathrm{had}}}}}_{\text{Hard contributions}} - \underbrace{\sum_{\text{Selected muons}} \vec{p}_{\mathrm{T}}^{\,\mu}}_{\vec{E}_{\mathrm{T}}^{\mathrm{miss},\mu}}}_{\text{Hard contributions}} - \underbrace{\sum_{\text{Accepted jets}} \vec{p}_{\mathrm{T}}^{\,\mathrm{jet}}}_{\vec{E}_{\mathrm{T}}^{\mathrm{miss,jet}}}}_{\text{Hard contributions}} - \underbrace{\sum_{\text{Unused tracks}} \vec{p}_{\mathrm{T}}^{\,\mathrm{track}}}_{\vec{E}_{\mathrm{T}}^{\mathrm{miss,soft}}}}_{\text{Soft contributions}} \tag{5.6}$$

Here, $\vec{p}_{\mathrm{T}}^{\,o}$ is $\vec{p}_{\mathrm{T}} = (p_x, p_y)$ of the accepted objects $o \in \{e, \gamma, \tau_{\mathrm{had}}, \mu, \mathrm{jet}\}$. The $\vec{E}_{\mathrm{T}}^{\mathrm{miss,soft}}$ denotes the corresponding soft contributions from unused tracks.
One persistent issue is the reconstruction of the physics objects themselves. In the dedicated reconstructions of the physics objects, the different detector signals can be used multiple times, for instance, in the electron reconstruction and the jet reconstruction. This would introduce double counting if $E_{\mathrm{T}}^{\mathrm{miss}}$ is calculated as in Equation (5.5). To avoid this double counting, the object-based $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction uses a signal ambiguity resolution, which forces the contributing objects to be reconstructed from mutually exclusive detector signals. To achieve this, a specific order in which contributions are taken into account and which are rejected needs to be implemented. The most commonly used order for the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction starts with the hard contributions, in this case, the selected electrons e. Afterwards, the hadronically decaying τ leptons (τ_{had}) are taken into account, and lastly, the jets. Muons (μ) are somewhat unique because they are reconstructed from ID and MS tracks, and the energy loss in the calorimeter is already taken into account, leading to little or no signal overlap with other reconstructed objects. To illustrate this behaviour further, one can start by

taking all selected electrons and adding their contributions to $E_{\mathrm{T}}^{\mathrm{miss}}$. Now, all lower-priority objects that share their calorimeter signal with the higher-priority objects will be removed. This procedure is repeated for all other objects down the hierarchy.
Using Equations (5.8) and (5.9), one can construct the corresponding $E_{\mathrm{T}}^{\mathrm{miss}}$ and ϕ^{miss}.

$$\vec{E}_{\mathrm{T}}^{\mathrm{miss}} = \frac{\vec{E}_x^{\mathrm{miss}}}{\vec{E}_y^{\mathrm{miss}}} \tag{5.7}$$

$$E_{\mathrm{T}}^{\mathrm{miss}} = \left|\vec{E}_{\mathrm{T}}^{\mathrm{miss}}\right| = \sqrt{\left(E_x^{\mathrm{miss}}\right)^2 + \left(E_y^{\mathrm{miss}}\right)^2} \tag{5.8}$$

$$\phi^{\mathrm{miss}} = \tan^{-1}\left(\frac{E_y^{\mathrm{miss}}}{E_x^{\mathrm{miss}}}\right) \tag{5.9}$$

In addition to the already described procedure, ATLAS also employs a flexible method to recalculate the $E_{\mathrm{T}}^{\mathrm{miss}}$ based on the analysis requirements. These requirements might exclude certain objects or use different collections of objects, which need to be considered and, therefore, invalidate the already described method. In the flexible method, the excluded objects are simply dropped from the $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ calculation [Equation (5.6)].

6

MACHINE LEARNING

Over the last couple of years, the amount of data produced and processed in HEP increased continuously to unprecedented numbers. With the upcoming High-Luminosity LHC (HL-LHC) phase, this number will grow even further, beyond the capability of the standard algorithms utilised for Run 2 of the LHC. To resolve this issue, multiple new techniques in different areas of the LHC need to be successfully implemented. One of these techniques, which will be used more heavily, is ML. The basics of ML are nothing new, stated and first implemented already in the 1990s, but recent developments in the field show huge improvements in performance and usability [145]. While the LHC collaborations are already using ML extensively in their analyses and combined performance groups, most of the ML techniques are either outdated or could be upgraded to make use of the improvements achieved outside of HEP. An example for this, which will be later shown in this thesis, is the usage of the Deep Sets architecture [146] over the Long Short-Term Memory (LSTM) architecture [147] in the context of heavy-flavour tagging.
To prepare for the ML techniques employed in this thesis, the following chapter will provide a comprehensive explanation of fundamental ML concepts for supervised learning. This includes Neural Networks (NNs) [148], various architectures of NNs, and advanced methods which are essential for constructing and training state-of-the-art ML techniques.

6.1 MULTILAYER PERCEPTRONS

The basic explanation of ML and NNs will be given in the context of so-called Multilayer Perceptrons (MLPs), one of the most basic NN architectures possible.

6.1.1 *Structure and Inference*

MLPs are fully-connected feedforward NNs composed of multiple neurons organised into layers. These layers are categorised into three classes: Input, hidden, and output layers. A simplified schematic representation of a MLP is depicted in Figure 6.1. In addition to being a MLP, the NN depicted can also be classified as a Deep Neural Network (DNN) due to its usage of multiple hidden layers. In general, a NN is classified as a DNN when it employs more than one hidden layer.
In MLPs, each neuron in one layer is connected to every neuron in the subsequent layer, except for the output neurons, which have no further connections. This connection scheme, also known as fully-connected, together with the one-directional flow of information from the input to the output layer, known as feedforward, are the basic properties of MLP NNs.

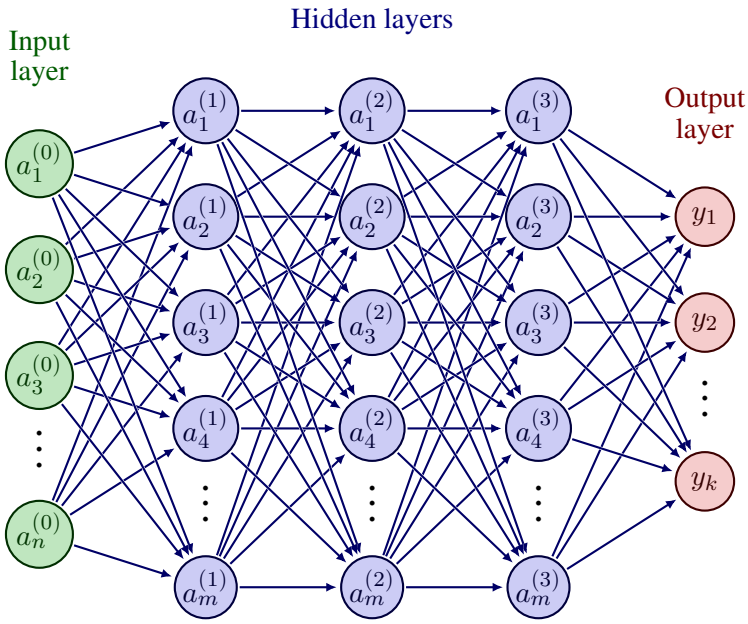


Figure 6.1.: Schematic representation of a MLP neural network with three hidden layers. The different layers and neurons are shown with their connections to the next layer. Taken and adapted from Ref. [149].

The starting point of all MLPs is the neurons in the input layer, $a_n^{(0)}$. These neurons have no connections to previous neurons and directly receive the inputs to the NN. In the context of particle physics, these inputs may consist of variables from objects such as jet p_{T} and η, or other event-related information like $E_{\mathrm{T}}^{\mathrm{miss}}$. The information provided to the input neurons is processed and transmitted to the next layer of neurons, the hidden layer.

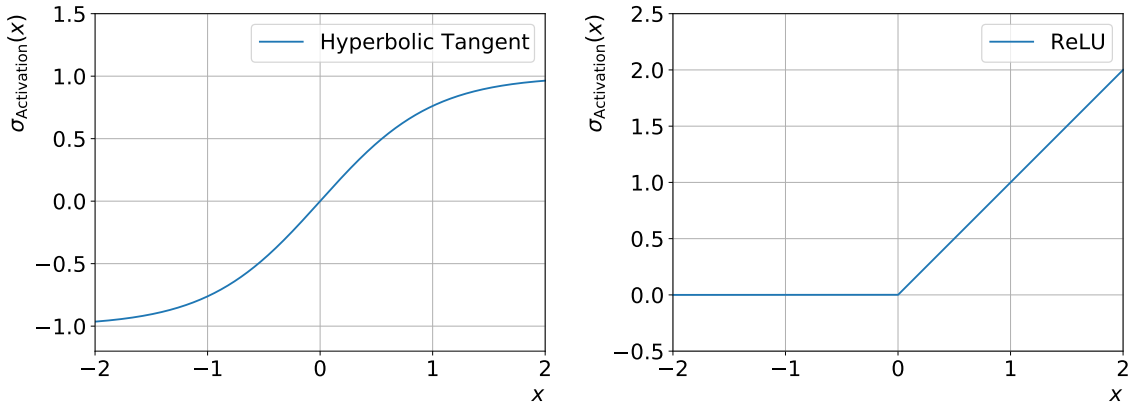
For a single neuron of any layer, the input is computed using Equation (6.1). In this equation, a_i represents the output of the i-th neuron in the previous layer, w_i denotes the weight of the connection to the i-th neuron in the previous layer, and N is the total number of neurons in the previous layer[1].

$$I = \sum_{i=1}^{N} w_i a_i \tag{6.1}$$

The weight mentioned here is one of the first trainable parameter of the NN. A more detailed explanation of how weight values are determined will be provided in Section 6.1.2.

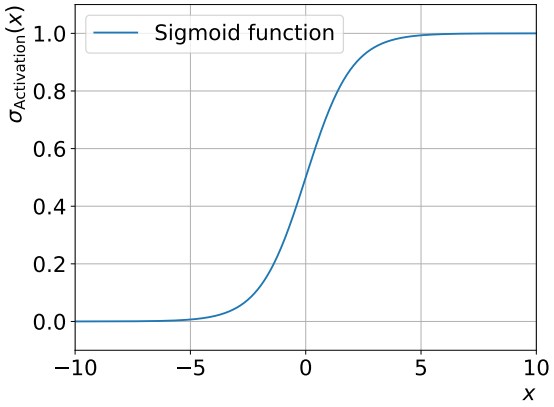
To calculate the output of a neuron, an activation function is introduced, denoted as $\sigma_{\mathrm{Activation}}$. This function, based on the input, determines the neuron's output and can significantly influence the network's behaviour. By setting different activation functions for each layer independently, the behaviour can adapted as needed and adds flexibility to the network. Typically, these functions are chosen based on the layer type and problem specific, with one for the hidden layer and another for the output layer. In general, non-linear activation functions are used for the different layers, which enables the NN to solve non-trivial problems with a small number of neurons. It also differentiates NNs from simple, linear models. As an example, three commonly used activation functions are shown in Figure 6.2. The three functions, the hyperbolic tangent (tanh) (a), the Rectified Linear

1 For the input layer, $N = 1$ which simply is the input value to this neuron.



(a): Hyperbolic tangent (tanh) function

(b): ReLU function



(c): Sigmoid function

Figure 6.2.: Example activation functions used in hidden layer neurons.

Unit (ReLU) [150] (b), and the Sigmoid (c) are also defined in Equations (6.2), (6.3), (6.4), respectively.

$$\tanh(x) = 1 - \frac{2}{e^{2x} + 1} \tag{6.2}$$

$$\text{ReLU}(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases} \tag{6.3}$$

$$\text{Sigmoid}(x) = \frac{1}{1 + e^{-x}} \tag{6.4}$$

To regulate the activation functions not only by the weights of the connections and to prevent the output of a neuron from being zero if the input is zero, a trainable bias term b_i is introduced, which can be interpreted as an offset to the network's input. Making use of the introduced activation functions, the input definition from Equation (6.1), and the bias term b, one can define the output of a single neuron using Equation (6.5).

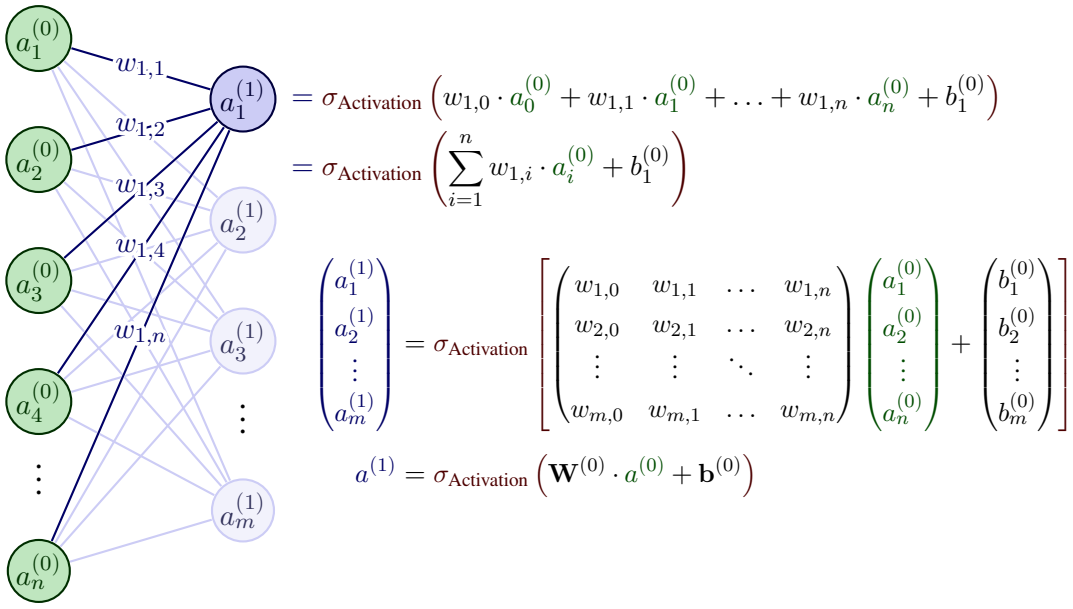


Figure 6.3.: Calculation of the output of a single hidden layer neuron and the combined calculation for all neurons in one hidden layer, written as a matrix equation. Taken and adapted from Ref. [149].

$$O = \sigma_{\text{Activation}} \left[\left(\sum_{i=1}^{N} w_i a_i \right) + b \right] \tag{6.5}$$

Expanding upon this concept, one can now consider all the neurons in the hidden layer, leading to a matrix equation for the outputs of all hidden layer neurons. This concept is illustrated in Figure 6.3.

For the final output layer, depending on the task which the NN should solve, the output and the corresponding activation functions are unique. One of these noteworthy activation function, which plays a prominent role in this thesis, is the softmax function [151], as defined in Equation (6.6).

$$\text{softmax}\,(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{N} e^{z_j}} \tag{6.6}$$

Here, z represents the vector of inputs for all N output layer neurons and i is the index of the output neuron for which the output is currently being computed. The softmax function is unique in its application, as it is used exclusively for the output layer. Unlike other activation functions that operate on individual neurons, the softmax function works on the entire output layer, simultaneously computing the output for each neuron based on the complete input information. Due to its definition, the outputs of the softmax function are always positive and normalised along the layer. To elaborate, the sum of all outputs of the output layer neurons will always be one. Therefore, the outputs can be interpreted as probabilities, which is an important feature for a classification NN.

While the softmax function is used for classification tasks, other activation functions achieve different results and are therefore chosen depending on the task which is to be solved. As an example, using a single neuron in the output layer and choosing the linear activation function $y = x$ results in a regression NN, which can be used to regress a certain value, like the invariant mass of a $t\bar{t}$ system [152].

One of the key reasons to use NNs is their speed. The so-called inference time, the time needed to calculate the output for a given input, is very small in comparison to standard algorithms, which makes them a useful tool in time-critical applications. The reason for this is the speed at which matrix equations can be solved. This efficiency is largely attributed to the remarkable optimisation of programming languages for handling matrix computations. Furthermore, due to the parallelisability of the calculations, Graphics Processing Units (GPUs) can be utilised to further speed up the inference.

6.1.2 *Training a Neural Network*

With the structure of the NN now defined, the next step is to optimise the network for a given problem. In the case of supervised learning, this optimisation is achieved using labelled training simulations[2]. In the context of particle physics, this optimisation may involve a classification task, such as categorising an event into a specific category, or a regression task, like predicting the invariant mass of a $t\bar{t}$ system. The labelled data consists of simulated MC events for which the true classification or mass is already known.

To optimise or train the network for a given problem within the framework of maximum LH, the typical objective is to maximise the LH and therefore minimise the difference between two probability distributions. In the case of a supervised classification task, these distributions are the true classification distribution of the training simulations and the predicted classification distribution generated by the network. Hence, network optimisation involves minimising the discrepancy between these distributions. To accomplish this, it is essential to define a metric that accurately quantifies this difference. In ML, given the minimisation nature of this metric, it is often referred to as a loss or cost function. In the context of maximum LH, particularly for a classification task, a commonly chosen metric is the cross-entropy function.

To illustrate and enhance the understanding of these functions in the context of ML, one can start with a simple binary classification problem. In this scenario, there is only one neuron in the output layer with a Sigmoid activation function. The training sample consists of data points of two classes: Background (label "0") and signal (label "1"). The binary cross-entropy loss function $f_{\text{BCE Loss}}$, a normalised negative Log-Likelihood (LLH), can therefore be defined as shown in Equation (6.7).

$$f_{\text{BCE Loss}} = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot \log\left(a_i\right) + \left(1 - y_i\right) \cdot \log\left(1 - a_i\right) \tag{6.7}$$

Here, N denotes the number of training data points, y_i the correct label of the i-th data point, and a_i the predicted label of the i-th data point. To now minimise this binary cross-entropy loss, the network needs to be changed in such a way that it solves the underlying problem better. To achieve this, the already mentioned trainable parameters of the NN come into play. The connection weights and bias terms of all the neurons are parameters, which are initialised randomly when the network is created but can be adapted during the training to optimise the network's performance and

2 Note that while real labelled data can be used for training, in the field of HEP, this is often impractical due to the uncertain nature of the events themselves. Therefore, HEP typically relies on simulated events where the truth is known.

minimise the loss function. To do so, so-called gradient descent methods are used. Although many possible gradient descent methods are available and valid, the focus here will be on the stochastic gradient descent method [153]. To minimise the loss function using this method, the approximation in Equation (6.8) plays a pivotal role.

$$\nabla f_{\text{Loss}} \approx \frac{\partial f_{\text{Loss}}}{\partial x_i} \nabla x_i \tag{6.8}$$

This equation establishes a relationship between how the loss function changes (∇f_{Loss}) when the variable x_i changes by a certain amount (∇x_i). Here, $\partial f_{\text{Loss}} / \partial x_i$ represents the gradient of the loss function concerning the altered variable. Using this approximation, one can calculate the change in the loss function, which depends on all the trainable parameters of the neural network, independently at any neuron in the network. To minimise the loss function, each weight can be adjusted by moving it in the direction opposite to the gradient, as shown in Equation (6.9), where w_j represents the trainable parameter to be modified.

$$w_j = w_j - \text{LR} \cdot \frac{\partial f_{\text{Loss}}}{\partial w_j} \tag{6.9}$$

Another parameter in the equation is the so-called Learning Rate (LR), a non-trainable parameter set by the user before training, which gives, as the name suggests, the rate at which the weights are adapted. It can also be interpreted as a step size for the movement towards the direction opposite to the gradient. There are other such network parameters that need to be configured before training commences, and these are typically referred to as hyperparameters.
As an example, the stochastic gradient descent method is shown schematically in Figure 6.4 for a loss function with one parameter w.

As one can see with a properly chosen LR, the loss function converges towards a local minimum (marked by the black cross). Although this method will ensure convergence to a minimum, one can never be sure if a local or a global minimum is reached, while the global minimum is the optimal training result. Therefore, testing out multiple LRs and checking the performance of the network is an important optimisation step when training a network. These kinds of optimisations, which need to be done by hand, are referred to as hyperparameter optimisations.
While Equations (6.8) and (6.9) offer accurate descriptions that can be used for the network's training, in practice, adjusting the network's parameters after each data point is computationally expensive and time-consuming. Moreover, it only provides marginal improvements to the network's performance. To address this issue, an adaptation of the stochastic gradient descent method is employed, resulting in what is known as mini-batch gradient descent. With mini-batch gradient descent, exact derivatives are not computed for each data point and each weight. Instead, an approximation of the derivatives is used, which is derived from a batch of training data points rather than individual data points. This technique resolves the computational inefficiency in weight adaptation, although it introduces its own set of challenges. While per-data point gradient descent guarantees a decrease in the loss function, mini-batch gradient descent only assures a decrease in the loss function over time. Therefore, the

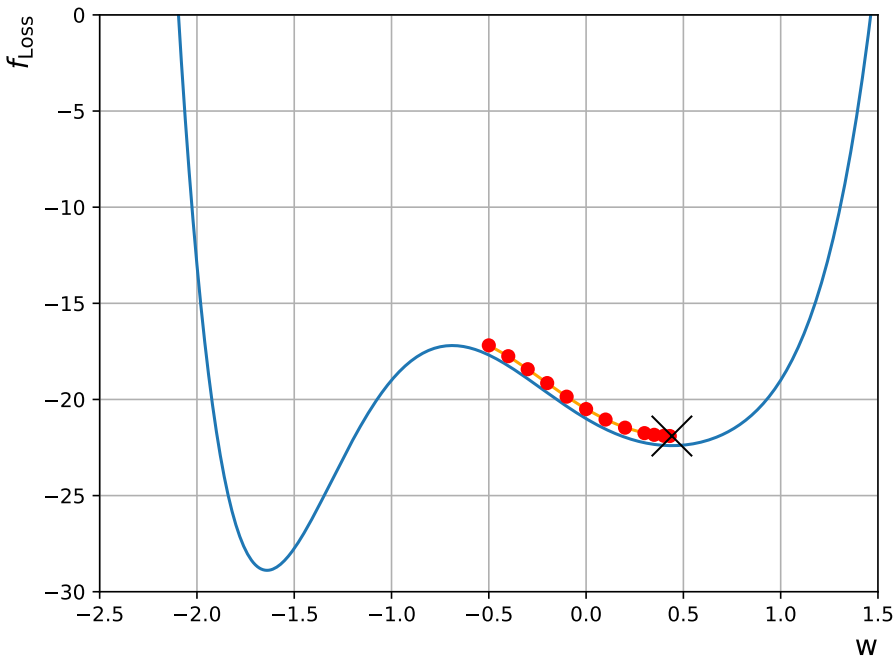


Figure 6.4.: Schematic loss function (red) and the behaviour of infinitesimally small changes in w to minimise the value of the loss function f_{Loss}.

loss function will still converge with an appropriate learning rate, but it may exhibit minor fluctuations. Furthermore, in addition to these fluctuations, mini-batch gradient descent introduces another hyperparameter: the batch size. This parameter determines the number of data points used in each approximation, or, in other words, how many data points are utilised to calculate an update of the network's trainable parameters.

To adapt the weights, the technique known as backpropagation [154] is employed. Backpropagation is a complex method that adjusts the network's trainable parameters. Starting from the last layer and working towards the first layer, backpropagation utilises the chained derivatives to optimise the parameters of the network, which gives the technique its name.

Once the training of all batches is completed, the network may not be perfectly trained yet. To further improve the network's performance, all batches are trained again. This periodic training of all available batches is what optimises the network to properly solve a specific problem. Similar to how humans benefit from revisiting material to enhance their learning, this iterative process improves the network's performance. A cycle of training all available batches is referred to as an epoch.

6.1.3 *Sample Handling*

To train the NNs, the training data points mentioned earlier play a pivotal role. Although one might intuitively consider using all available data points for training NNs to benefit from a larger number of training data points, it is generally correct to split data points into three categories: training, validation, and testing. While various conventions exist for naming these categories, in this thesis, the

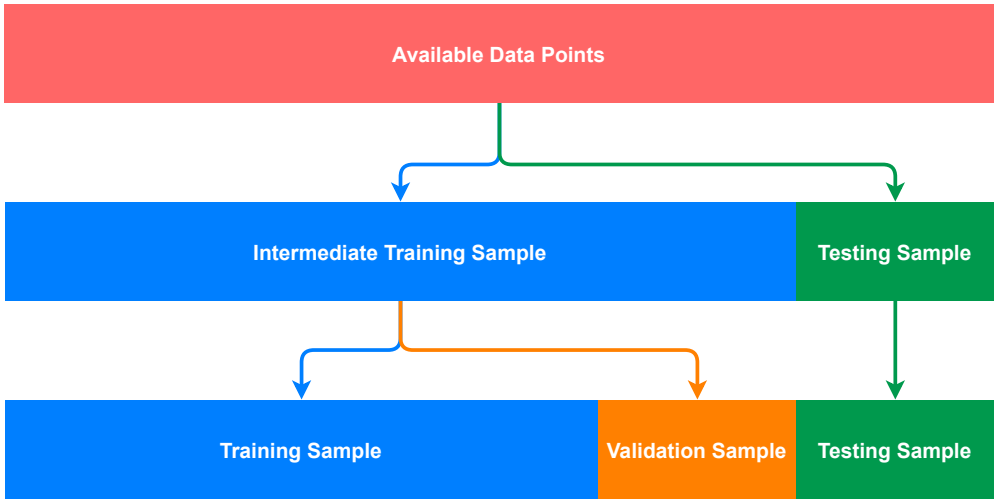


Figure 6.5.: Schematic representation of the splitting of the available data points for training a NN.

terms training, validation, and testing will be used. The full splitting procedure is illustrated in Figure 6.5.

Beginning with the full available dataset, the data points are initially separated into two groups: the intermediate training sample and the testing sample. The intermediate training sample is adapted for training the NNs, whereas the testing sample is never used or modified for network training. It exclusively serves as a final testing sample to assess the performance of the trained NNs for the predefined task. While the testing sample is separated and left untouched, the intermediate training sample is further divided into the training sample and the validation sample. The training sample and the data points it contains are employed for the actual training of the NNs, as described in Section 6.1.2, while the validation sample is used to validate the network's performance during training. Since the validation and testing samples are primarily used for network evaluation, their sizes are often relatively small, however each of these samples need to be statistically independent and representative with respect to the full available data points. A common data point split is $80\,\%$ for training, $10\,\%$ for validation, and $10\,\%$ for testing

The split into three categories is primarily intended to identify and address problems during or after training the NNs. Some of these challenges and how to address them are discussed in the following Section 6.1.4.

6.1.4 *Challenges in the Training of a Neural Network*

With the given setup from Section 6.1.2 and the samples from Section 6.1.3, it is now possible to train the supervised NN for a given problem. While the network will adapt to better solve the given problem, another important effect comes into play which is shown in Figure 6.6.

It illustrates the loss values of the loss function for both the training and validation samples plotted against the number of epochs trained. While the training loss was previously explained, the validation loss is computed using the validation sample. The validation loss serves as a metric to evaluate

the network's performance on data points it has never encountered during training and therefore the generalisability of the network to new, not trained data points.

As clearly visible in the later epochs, starting from epoch 14 onwards, the validation loss begins to increased, while the training loss remains at a minimum and may even decrease further. This phenomenon is known as overfitting or overtraining. While the network continues to learn the features of the training data points, it starts to acquire highly specific properties that are not generalisable to all data points. This can be compared to the network memorising the training dataset. Although overfitting should not be prevalent in a properly sized training sample with suitable hyperparameters, an overfitted network will not perform as effectively on new, unseen samples as it does on the training sample itself. With smaller training samples, which do not properly reflect the general features of the data points or wrong hyperparameters, this effect can start even earlier or compromise the whole training.

To encounter such overfitting and restrict the network to not learn the training sample by heart, one can introduce so-called regularisation terms. These terms are constraints, which are added to the loss function, to penalise the network based on different criteria. Two of the most used regularisation methods are the L1 and L2 regularisations, for which the full loss functions are shown in Equations (6.10) and (6.11), respectively.

$$f_{\text{L1 Loss}} = f_{\text{Loss}} + \frac{\lambda}{N} \sum_{i=1}^{N_w} |w_i| \tag{6.10}$$

$$f_{\text{L1 Loss}} = f_{\text{Loss}} + \frac{\lambda}{2N} \sum_{i=1}^{N_w} w_i^2 \tag{6.11}$$

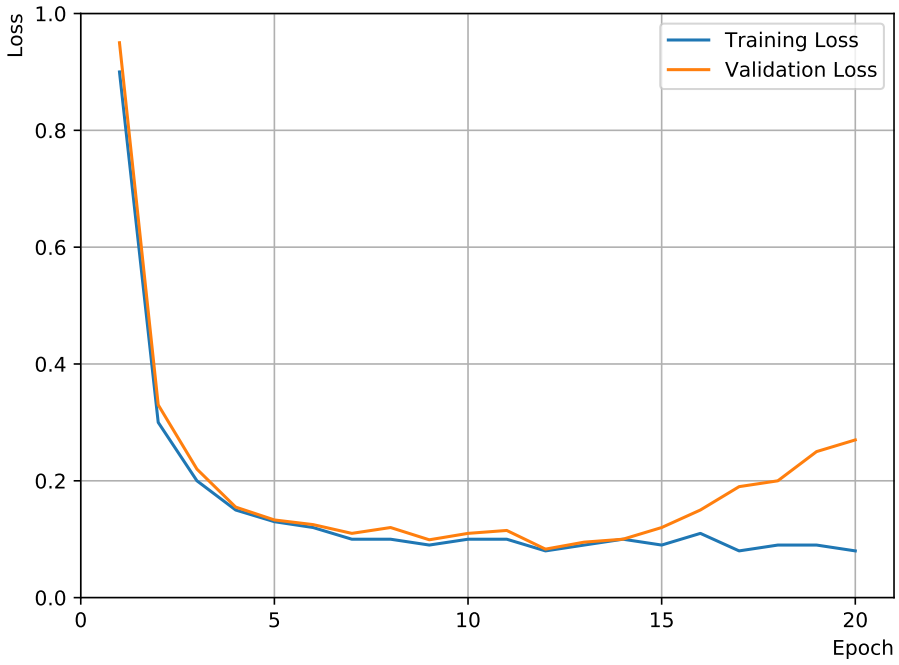


Figure 6.6.: Schematic training and validation loss as a function of the number of epochs trained.

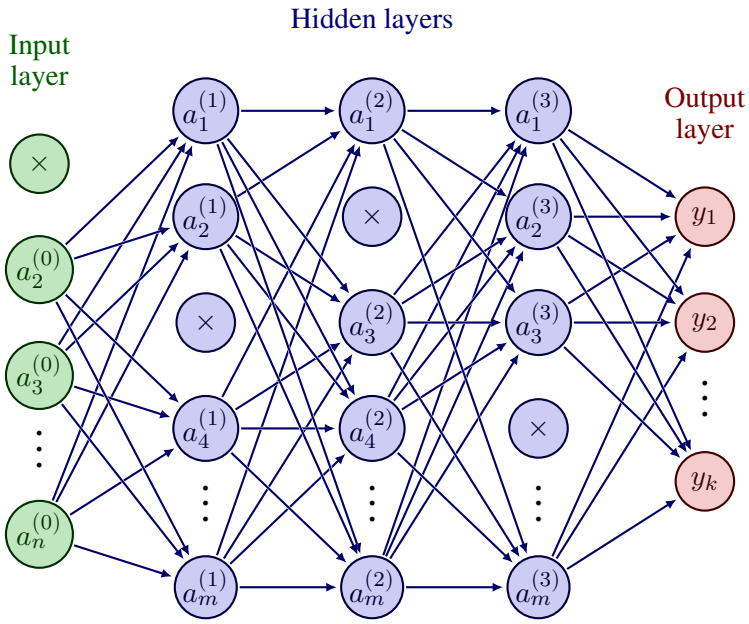


Figure 6.7.: Schematic representation of a MLP with Dropout. The neurons marked with $\times$ are dropped and therefore all connections to previous and next layer neurons are discarded. Taken and adapted from Ref. [149].

Here, w_i denotes the i-th weight, N the number of data points used for the loss calculation and weight adaptation, N_w the number of weights in the network, and λ a new hyperparameter, which controls the relative contribution to the loss. While the L1 regularisation constraints the sum of the absolute values of the weights in the network, the L2 regularisation constraints the sum of the squared weights in the network. For both types of regularisation, large weights are penalised while an equal magnitude, or size, of the weights is favoured.

In addition to the previously mentioned regularisation techniques, there are various other methods available to further mitigate overfitting in NNs. One such technique, utilised in this thesis, is Dropout [155]. When overfitting occurs, some neurons develop complex co-adaptations with other neurons to better address specific training data points, potentially ignoring more general features that are crucial for generalisation to unseen samples. Dropout addresses this issue by probabilistically 'dropping' or deactivating certain neurons in the input and hidden layers with a specified probability p, also known as the dropout rate. By excluding these neurons and their connections, Dropout compels the different layers to take varying degrees of responsibility for their input. This probabilistic approach ensures that the model focuses more on learning the general features, thereby reducing the overfitting problem. A schematic of how the neurons are 'dropped' is depicted in Figure 6.7.

An intriguing effect of Dropout becomes evident when examining the loss curves for training and validation samples. As depicted in Figure 6.8, the validation loss is higher than the training loss during the initial epochs, which is expected. However, from epoch seven onward, the validation loss begins to be lower than the training loss. This peculiar behaviour of Dropout can be attributed to how the training and validation losses are calculated. The training loss for an epoch is an average computed over the per-batch losses during training, while the validation loss is calculated at the end

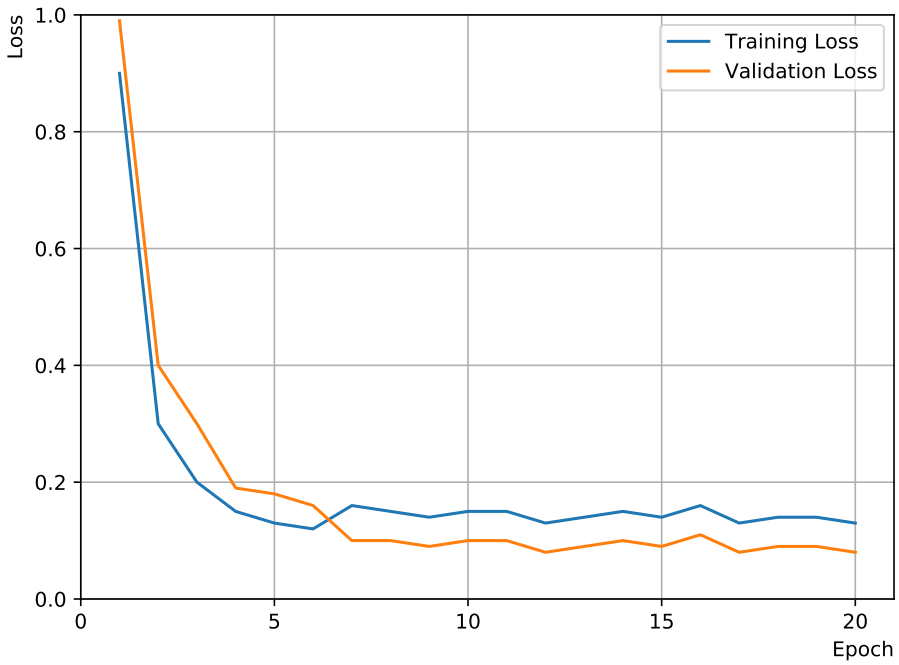


Figure 6.8.: Schematic training and validation loss as a function of the number of epochs trained. In the training of the network, Dropout was used as a regularisation.

of the epoch and is not an average. The unusual feature arises from the 'dropping' of neurons during batch training. When certain neurons with co-adaptations are dropped, the loss temporarily increases for that batch iteration. This increase is due to the significance of the dropped neuron in influencing the network's final output. Consequently, the average training loss for the epoch rises, while the epoch validation loss remains unaffected.

Although this behaviour is demonstrated for Dropout, various regularisation techniques produce similar effects on the loss curves due to the same underlying reason.

Another technique for a faster convergence of the network's training worth mentioning is the so-called Batch Normalisation [156]. In DNNs, due to the complex chains of connected neurons, a small change in one of the parameters of the network can already cause a huge change in the input to another later layer. Therefore, in each parameter adaptation, the layers need to adapt to new incoming distributions, which slows down the final convergence of the network. This so-called covariance shift can be removed by using the already mentioned Batch Normalisation. Calculating the expectation value and variance of the input to a layer for a batch of data points $\mathcal{B}$, one can retrieve the normalisation for the inputs x with Equation (6.12), where $\hat{x}^{(k)}$ denotes the normalised k-th input to the layer, $x^{(k)}$ the un-normalised k-th input to the layer, $E\left[x^{(k)}\right]_{\mathcal{B}}$ the expectation value for the k-th input, and $\text{Var}\left[x^{(k)}\right]_{\mathcal{B}}$ the variance for the k-th input.

$$\hat{x}^{(k)} = \frac{x^{(k)} - E\left[x^{(k)}\right]_{\mathcal{B}}}{\sqrt{\text{Var}\left[x^{(k)}\right]_{\mathcal{B}}}} \tag{6.12}$$

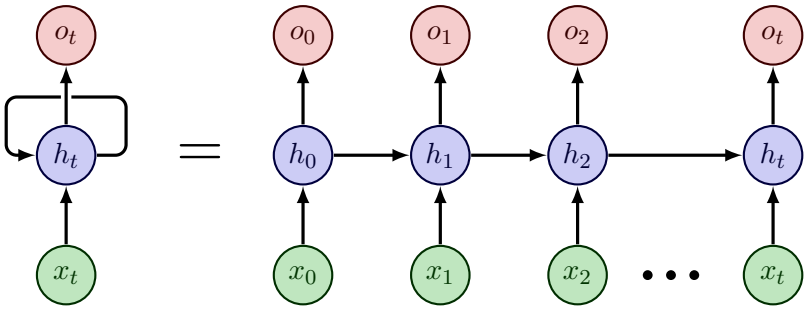


Figure 6.9.: Simplified architecture of a RNN layer with a single neuron.

However, merely normalising all inputs to a layer can significantly alter the layer's representational capabilities, especially given the presence of non-linear activation functions. To address this issue, Batch Normalisation ensures that the transformation incorporated in the network can represent the identity transform by introducing two new trainable parameters, denoted as $\gamma^{(k)}$ and $\beta^{(k)}$. The updated input to the layer can be expressed as shown in Equation (6.13), where $y^{(k)}$ represents the new input to the layer.

$$y^{(k)} = \gamma^{(k)} \hat{x}^{(k)} + \beta^{(k)} \tag{6.13}$$

Although the mentioned techniques are very helpful in training a basic NN or a MLP, certain problems are not solvable by regularisation or training techniques. One of these problems, which will play a prominent role in this thesis, is the processing of sequences of data. While the MLP has a fixed size of input neurons and is therefore bound to the number of inputs possible, problems like language processing are not fixed in size. For example, one sentence that is to be processed can have different amounts of words and punctuation marks and also can have pivotal information in the ordering of the inputs. These problems require not only new techniques of how to process the input sequence but also new architectures that can make use of the order of the sequence. The first network type addressing these problems is the Recurrent Neural Network (RNN), which will be presented in the Section 6.2.

6.2 RECURRENT AND LONG SHORT-TERM MEMORY NEURAL NETWORKS

Recurrent Neural Networks (RNNs) [157] are another type of NN architecture used in ML. In contrast to MLPs, RNNs possess the unique ability to process sequences as input, making them one of the standard types of neural network architecture in various fields, including natural language processing, speech recognition, and time series forecasting. A schematic and simplified representation of an RNN is shown in Figure 6.9.

While the output of the RNN layer behaves similarly to MLPs, the input and hidden functionalities are different. Given an input sequence x with t entries, the RNN begins processing the entries in x from the start to the end. This processing is carried out by the so-called hidden state h, which simultaneously acts as memory. In contrast to a standard hidden neuron, the hidden state uses a

different formula to compute its output, considering their output from the previous input entry $\mathbf{h}_{t-1}$, as shown in Equation (6.14). Similar to hidden neurons, the inputs to the neuron are weighted by their respective connection weights $\mathbf{U}$ in addition to a bias term $\mathbf{b}$. Moreover, the output of the previous hidden states is added, weighted with their own trainable connection weights $\mathbf{W}$. This combined input is then processed, akin to hidden neurons, by an activation function $\sigma_{\text{Activation}}$. However, this process is not identical to that of standard hidden neuron. For RNNs, the output is not exactly $\mathbf{h}_t$ but rather calculated using Equation (6.15), where $\sigma_{\text{Output Activation}}$ denotes the used activation function, $\mathbf{V}$ the connection weights, $\mathbf{h}_t$ the hidden state output, and $\mathbf{c}$ the bias term.

$$\mathbf{h}_t = \sigma_{\text{Activation}} \left[\mathbf{U}\mathbf{x}_t + \mathbf{W}\mathbf{h}_{t-1} + \mathbf{b}\right] \tag{6.14}$$

$$\mathbf{o}_t = \sigma_{\text{Output Activation}} \left[\mathbf{V}\mathbf{h}_t + \mathbf{c}\right] \tag{6.15}$$

Similar, but slightly more complex than for the MLPs, is the training of RNNs. While similar techniques are employed, a distinct method of backpropagation must be incorporated to effectively train RNNs. This specific method is known as Backpropagation Through Time (BPTT) [158–160], which optimises the network's weights in an unfolded manner (illustrated in Figure 6.9). Similar to MLPs, BPTT starts at the final entry or time step t of the network, applying the standard backpropagation method to optimise the trainable parameters. However, in contrast to MLPs, where the weights are unique and adapted only once per batch training, the same neurons are adapted repeatedly for each time step in RNNs. This might initially seem problematic, but it is precisely the intended behaviour. In RNNs, the weights for each time step of a hidden state neuron are shared. This weight sharing implies that, for example, the same $\mathbf{W}$ is used in time step one as in time step t.
While RNNs are highly versatile and find utility in numerous applications, the training process is relatively slow and cannot be accelerated beyond the $O(t)$ limit. This limitation is due to their sequential processing of inputs and their dependence on previous hidden state outputs. Consequently, RNNs are unable to leverage parallelised training using GPUs. In addition, the order of the entries in the input sequence is very important for RNNs, which makes them not favoured when working with unsorted sequences.

A specific form of RNNs are the so-called Long Short-Term Memory (LSTM) networks. Building upon the architecture of RNNs, LSTM networks enhance the memory capabilities of sequence-processing NNs by incorporating additional memory cells and so-called gates. These gates enable the network to remember or forget information from previous sequence entries. A schematic representation of an LSTM cell is depicted in Figure 6.10.

Similar to basic RNNs, LSTM networks operate on the input x at time step t, denoted as x_t, and the previous time step's hidden state h_{t-1}. In addition to these two inputs, the previous so-called cell state c_{t-1} is included. Serving as additional memory for storing long-term information, the cell state, along with the hidden state (also referred to as short-term memory), constitutes the memory mechanism of the LSTM cell. The LSTM network incorporates three of the previously mentioned gates, which can manipulate and alter the memory stored within the network. To comprehend the

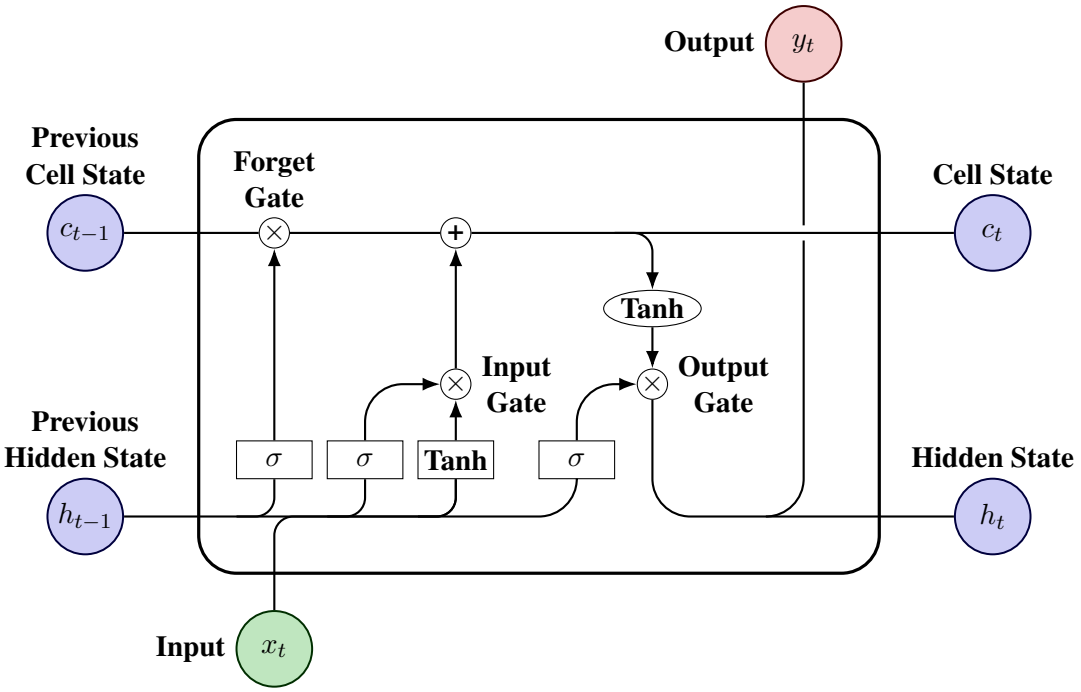


Figure 6.10.: Schematic representation of a LSTM cell. Taken and adapted from Ref. [161].

function of these gates, it is necessary to examine how the output of each gate is calculated.
For the forget gate, the input x_t and the previous hidden state h_{t-1} are concatenated and processed with an activation function, much like a simple neuron. This process is illustrated in Equation (6.16), where $\mathbf{W}_f$ represents the weights for the forget gate, $\mathbf{b}_f$ stands for the bias term of the forget gate, and σ denotes the activation function.

$$f_t = \sigma\left(\mathbf{W}_f \cdot [h_{t-1}, x_t] + \mathbf{b}_f\right) \tag{6.16}$$

While multiple activation functions are available for use, the forget gate in an LSTM network consistently employs the Sigmoid activation function. The output f_t is then multiplied in an element-wise way with the previous cell state c_{t-1}. Because of the characteristics of the Sigmoid function, the yielded values are always between zero and one. This property enables the forget gate to precisely control what is retained and what is discarded from the previous cell state.
Similar to the forget gate, the input gate takes the input and the previous hidden state as inputs. However, unlike the forget gate, the input gate processes the inputs a second time using a different activation function, weights, and bias term, resulting in the creation of the candidate cell state, denoted as $c_t^{\text{Candidate}}$. Both of these processed representations are then multiplied in an element-wise way, forming the output of the input gate. This output is added to the cell state, which has already been modified by the forget gate. The calculation of the input gate's output is illustrated in Equation (6.17), where $\mathbf{W}_i$ and $\mathbf{W}_c$ represent the weights, $\mathbf{b}_i$ and $\mathbf{b}_c$ denote the bias terms, and σ is the activation function, once again utilising the Sigmoid function.

$$i_t = \sigma\left(\mathbf{W}_i \cdot [h_{t-1}, x_t] + \mathbf{b}_i\right) \times \underbrace{\tanh\left(\mathbf{W}_c \cdot [h_{t-1}, x_t] + \mathbf{b}_c\right)}_{c_t^{\text{Candidate}}} \tag{6.17}$$

The final gate of the LSTM network is the output gate. This gate takes, as input, the now fully adapted cell state c_t, which has been processed by a hyperbolic tangent function. It also considers the concatenated and processed input x_t and the previous hidden state h_{t-1}. The output gate simply multiplies these inputs element-wise, resulting in the new hidden state h_t, which also serves as the cell's output, denoted as y_t. The formula for the output is provided in Equation (6.18), where $\mathbf{W}_o$ represents the weights, $\mathbf{b}_o$ signifies the bias term, and σ denotes the Sigmoid function.

$$o_t = \sigma\left(\mathbf{W}_o \cdot [h_{t-1}, x_t] + \mathbf{b}_o\right) \times \underbrace{\tanh\left(c_t\right)}_{h_t, y_t} \tag{6.18}$$

While LSTM networks are well-suited for extracting long-term correlations within input sequences, making them highly useful for complex sequence-related tasks, they also exhibit robustness to missing data or high levels of noise in the input sequence. Despite these favourable features, LSTM networks face several challenges. Similar to basic RNNs, they are computationally expensive in both training and inference due to their sequential data processing, a drawback that becomes more pronounced with larger input sequences. Additionally, they are highly susceptible to overfitting when working with small training samples due to their complex architecture. This drawback also necessitates the use of very large training samples to achieve full convergence for certain problems.

Although RNNs and LSTM networks are widely adopted and powerful when dealing with sequential data, the associated computational costs, particularly during training, can overshadow their success. In recent years, several new architectures have emerged to address the challenges of processing sequential inputs. One of these notable architectures, which holds particular significance for this thesis, is the Deep Sets architecture, which will be discussed in Section 6.3.

6.3 DEEP SETS NETWORKS

Similar to RNNs, Deep Sets networks work with a given set, consisting of inputs x_M, $\chi \in \mathbb{R}^M$. These inputs are processed by a set function denoted as $f(\chi)$. While for RNNs, this function is simply the cell responsible for processing the given input set, for Deep Sets networks, the processing of the entries within the set is separated from the final processing, as illustrated in Equation (6.19) and Figure 6.11.

$$f(\chi) = \rho\left(\sum_{x \in \chi} \phi(x)\right) \tag{6.19}$$

The fundamental idea behind Deep Sets is that a certain process function f, which operates on the input set χ, can be divided into two distinct processing functions, denoted as ϕ and ρ. To initially

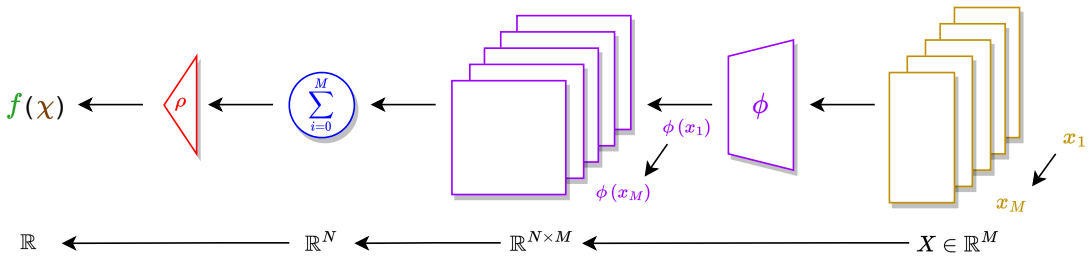


Figure 6.11.: Schematic representation of the functionality of the Deep Sets architecture.

process the input set and transform it into a different representation, the ϕ function is employed, working on each of the entries of x individually. The resulting representations of the inputs are then aggregated into a representation that is independent of the size of the set χ. While various aggregation functions are possible, the simplest and most commonly used one is the summation across the entries. This new, input-size-independent representation is subsequently processed by the second processing function ρ, which yields the overall processed results of the input set.

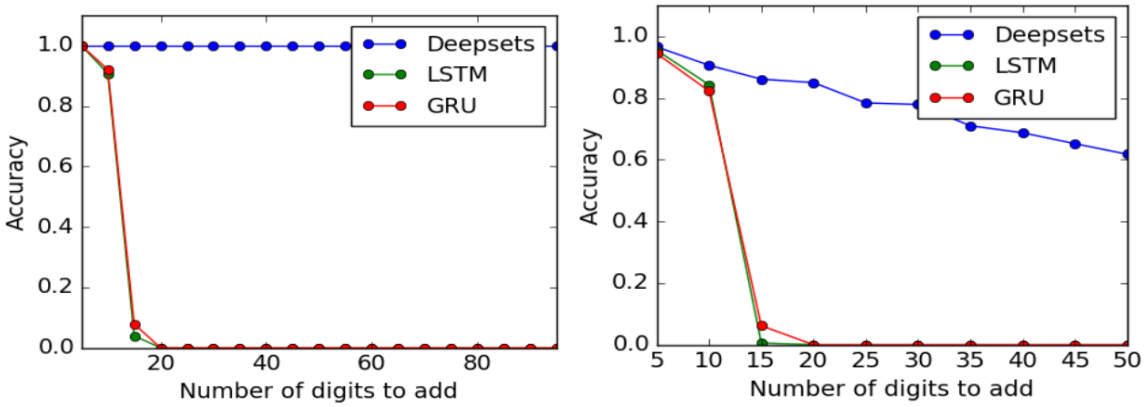
Because both ϕ and ρ operate on fixed-size inputs, ϕ on the entries of x and ρ on the aggregated representation of χ, both functions can be parametrised by simple NNs like MLPs, with ϕ playing a particular role. When processing the input entries x with the ϕ NN, multiple copies of ϕ are instantiated to process the entries in parallel. Similar to RNNs, these copies represent the same network, although, during training, they are treated as distinct instances. To accommodate this behaviour, the different networks share a set of weights.

While this approach may appear minimalistic and straightforward, it offers significant advantages over RNNs. RNNs shine at processing language because they can extract information based on the order of the inputs. However, they are not optimal for problems where no specific order of the entries in the set is provided. An example is the addition of digits, where three networks are used: Deep Sets, LSTM, and Gated Recurrent Unit (GRU)[3] networks [162]. The input provided to these networks is a set of digits, presented either as text or image[4]. The training samples include various sets with up to ten digits. Subsequently, the trained networks are evaluated using test samples, either in text or image format, depending on what the network was trained for. For these test samples, the sets were created differently, containing up to 100 digits per set. This was done to demonstrate the network's performance within the range of the number of digits for which they were trained and to assess their generalisation ability. The results of this evaluation are presented in Figure 6.12, with (a) displaying the results for text-based training and (b) showing the results for image-based training. The charts illustrate the accuracy of the digit sum as a function of the number of digits in the sets.

As one can observe, the accuracy of both the LSTM and GRU networks is comparable but slightly lower for sets with up to ten digits. However, for sets with more than ten digits, the accuracy rapidly deteriorates for both the LSTM and GRU networks, while the accuracy remains consistently high for the Deep Sets. Deep Sets networks excel at precisely these types of problems. Their independence

3 A GRU network is an adapted LSTM network with fewer parameters and no output gate. It has been shown that gating is generally a helpful feature for these problems.

4 For each of the possible inputs (text or images), separate networks are trained.



(a): Text-based set input

(b): Image-based set input

Figure 6.12.: Accuracy for the summation of digits as a function of number of digits in the input set. Different ways to provide these digits are presented: (a) text-based and (b) image-based. Of the three networks shown, the Deep Sets network outperforms both, the LSTM and the GRU network [146].

from the order of input sets, enabled by their aggregation function and the shared weights in the ϕ network, along with their ability to process inputs in parallel, gives them the potential to outperform various RNN-based architectures for such tasks. This advantage also includes the crucial ability to leverage GPUs. Due to the parallelisable processing of inputs in Deep Sets, GPU cards can be utilised during training, significantly reducing training and inference times.

6.4 TRANSFORMER NETWORKS

In recent years, alongside the already high-performing architectures, a new and highly powerful architecture for processing sequence inputs has emerged. These new types of NNs, known as transformers, are based on an attention mechanism [163] and large language models. This mechanism enables the network to process its inputs more efficiently and establish connections between different entries in the input set.

6.4.1 *Architecture*

The architecture of this new model is depicted in Figure 6.13.

The architecture used herein is based on an encoder-decoder approach. The left side of the architecture handles input encoding, while the right side handles decoding and produces output probabilities. The encoder-decoder structure shown in Figure 6.13 illustrates a single encoder and a single decoder layer[5]. These can be stacked multiple times to increase the size and complexity of the transformer network and thus increasing its learning capabilities.

Starting with the input to the encoder/decoder[6], a learnable embedding layer is used to convert the

5 The encoder and decoder layers are composed of multiple other layers but are typically referred to as a 'layer' due to their recurring structure in NNs.

6 The input to the decoder is the last entry output, similar to the hidden state in RNNs.

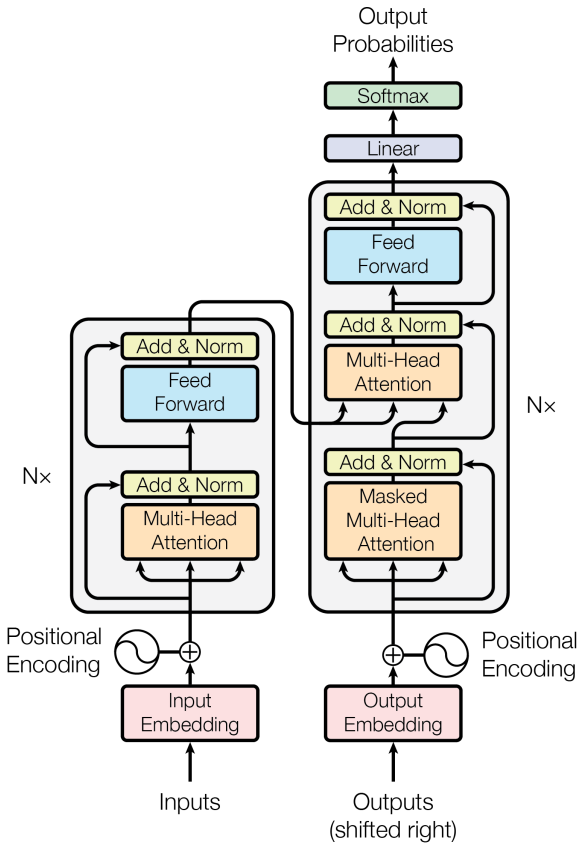


Figure 6.13.: Schematic representation of the transformer architecture [163].

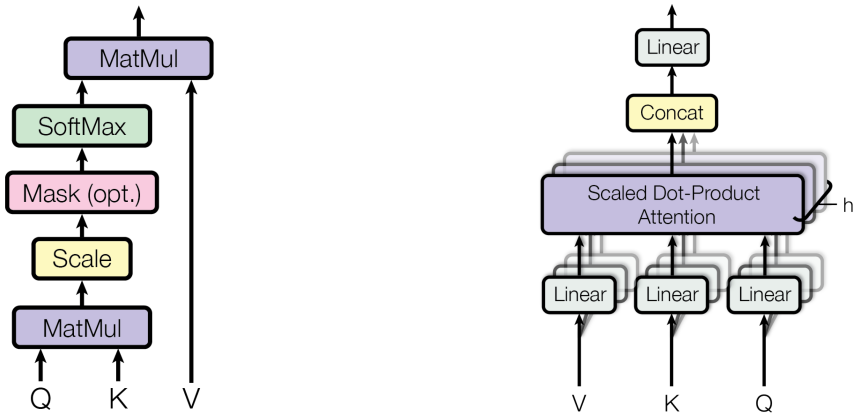
input entry of the sequence into vectors of a certain dimension. Similar to the ϕ network in Deep Sets, the learnable embedding layers share their weights for both the encoder and decoder. Next, positional encoding is added to the embedded inputs of the encoder/decoder layers to inject information about the relative or absolute position of the entry in the sequence. This positional encoding has the same dimension as the embedded inputs, allowing for simple addition after embedding itself. Different types of positional encodings, such as learned and fixed, can be used [164].

Moving on to the actual architecture of the encoder part of the transformer (on the left), it is primarily composed of two different types of sub-layers: a Multi-Head Attention (MHA)[7] layer and a simple position-wise fully connected feedforward layer[8]. For both layer outputs, the input to the layers itself is added and normalised using a layer normalisation approach [165]. Reusing the inputs together with the outputs or simply bypassing the inputs around the actual layer is referred to as residual connections [166]. These connections allow the network, especially for the MHA layer, to combine the original input with the layer's output to further extract specific correlations.

The second part of the transformer architecture is the decoder (on the right), which is similar to the encoder and, in general, consists of a (masked) MHA layer and a simple position-wise fully connected feedforward layer. The MHA layer used herein is masked, which hides future positions of the input to the decoder before running the actual MHA process. Similar to the encoder, residual

7 This will be explained in more detail in Section 6.4.2

8 This position-wise fully connected feedforward layer has an architecture and usage identical to the ϕ network used in Deep Sets.



(a): Scaled Dot Product Attention layer (b): Multi-Head Attention layer

Figure 6.14.: Schematic representation of the Scaled Dot Product Attention layer (a) and the Multi-Head Attention layer (b) [163].

connections are used in combination with layer normalisation. In addition to the two sub-layers, the decoder introduces a third sub-layer, a basic MHA layer, which utilises the output of the encoder. The output of the decoder is then processed with a learned linear transformation, which also shares its weights with the learned input embeddings of the encoder/decoder. It also includes a softmax activation function to transform the decoder output into probabilities for the next entry in the sequence.

With the herein-described architecture, the transformer can learn, extract, and correctly solve many different problems. One of these problems, which is the example of the underlying paper [163] is the translation of texts from one language to another. Although quite powerful, in physics, mainly one of the two parts is used for solving different problems, and this is the encoder.

6.4.2 *Multi-Head Attention Mechanism*

In order to understand how the MHA layer is operating, one needs to first understand the attention mechanism, or more specifically self-attention, in more detail. Self-attention plays a crucial role in understanding the relationships between different entries in the input sequence. It enables the model to weight the importance of each entry in the sequence concerning every other entry. As shown before, the self-attention mechanism is at the heart of how transformers process and capture contextual information in different tasks. Using a simple self-attention layer as an example, one can split the different processes inside this layer into different steps, which will be explained in the following. This is also depicted in Figure 6.14 (a). The depicted type of layer is a specific form of a self-attention layer, called the Scaled Dot Product Attention layer. While multiple other types of self-attention are available, the visualised form is the one generally used in transformer models.

Starting with the input to the self-attention layer, each entry of the embedded input sequence is further transformed into three different vectors by multiplying it with three different weight matrices: the Query, Key, and Value matrices. These three new vectors are called Query Q, Key K, and Value

V, respectively. While the Q vectors represent the contribution of each entry to the attention scores, the K vectors help identify the relevance of each entry to the Q vectors. The final V vectors store the information about the entries themselves. They contain the actual content associated with the entries[9].

The second step involves calculating the attention scores, achieved by taking the dot product of the Q vector of one entry with the K vector of all other entries in the sequence. This creates a set of attention scores, one for each pair of entries in the sequence. This operation is denoted in Figure 6.14 (a) with the MATMUL layer.

Next, to ensure that the magnitude of the attention scores does not become too high or too low, they are divided by the square root of the dimension of the K vectors d_k. This scaling of the attention scores helps control the gradients during transformer training, improving and maintaining numerical stability. In Figure 6.14 (a), this is denoted with the SCALE layer.

The fourth step, denoted with the MASK (OPT.) layer, is optional and only used in the decoder's masked MHA layer. It involves masking certain illegal connections that could introduce incorrect information flow to subsequent layers.

In the fifth step, to make the scaled attention scores interpretable and ensure that they sum up to one, the softmax function is applied. As mentioned earlier, the softmax function converts the scores into probabilities, reflecting their importance or relevance. Similar to before, this is denoted by the SOFTMAX layer in Figure 6.14 (a).

Finally, the calculated probabilities are used to create a weighted sum with the V vectors, as denoted by the MATMUL layer. These probabilities determine how important each entry is to other entries during processing and how much attention is given to them when processing the entry. This, coupled with the fact that the attention for the entries is calculated using the entries themselves, gives a more complete understanding of self-attention. The output of the weighted sum serves as the output of the full self-attention layer $O_{\text{Attention}}$, which can also be described using Equation (6.20), where K^T denotes the transposed K vector.

$$O_{\text{Attention}}\left(Q, K, V\right) = \text{softmax}\left(\frac{Q \cdot K^T}{\sqrt{d_k}}\right) \cdot V \tag{6.20}$$

With the newly defined self-attention layer, the actual MHA layer can be constructed. As the name suggests, MHA layers are essentially self-attention layers with multiple so-called attention heads. In simple terms, the previously explained functionality of self-attention corresponds to one attention head. This is also depicted in Figure 6.14 (b). By processing the embedded inputs X through the self-attention layer multiple times, each time with different and independent attention heads (using different Q, K, and V weight matrices), multiple additional outputs Z_1-Z_3 are created. These outputs are then concatenated and multiplied by an additional weight matrix W^O, which is also adjusted during training. The resulting representation Z serves as the output of the MHA layer. An example with three attention heads is depicted in Figure 6.15.

9 These introduced vectors are abstractions that are helpful when calculating or conceptualising attention.

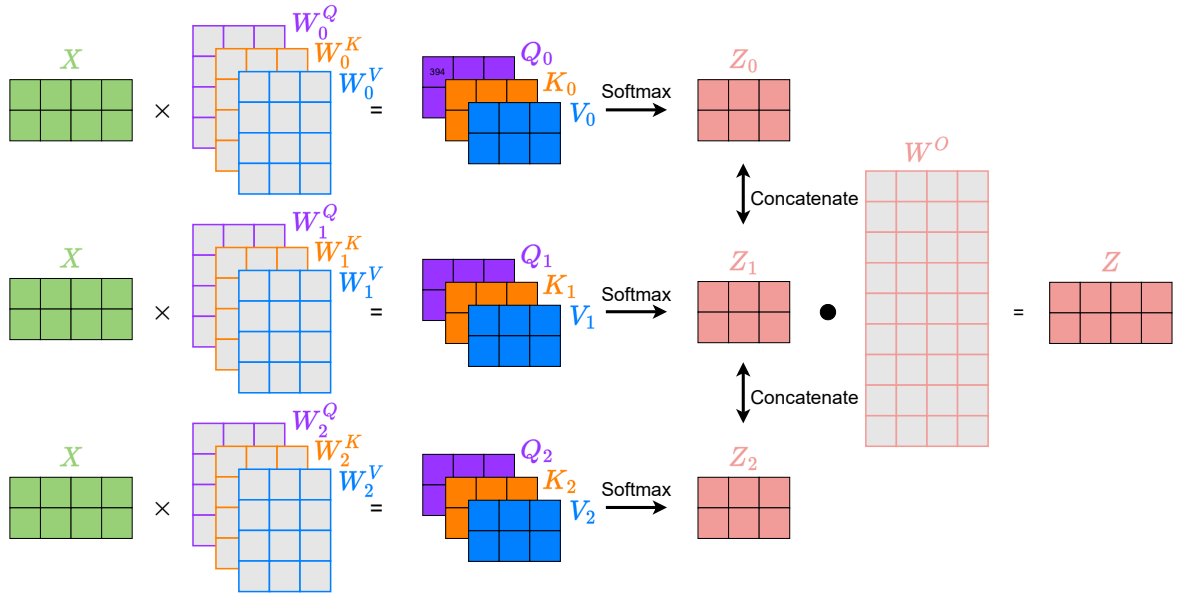


Figure 6.15.: Schematic representation of the inner workings of a MHA layer with three attention heads.

Part III

HEAVY-FLAVOUR TAGGING

7

INTRODUCTION TO HEAVY-FLAVOUR TAGGING

Heavy-flavour tagging, or simply flavour tagging, is the process of identifying jets originating from the decay of heavy-flavour particles, namely charm (c) and bottom (b) quarks. This powerful tool plays a pivotal role for the ATLAS experiment to discriminate between signal and background. While flavour tagging includes the identification of c-quark initiated jets, the primary goal of the algorithms discussed in this chapter is the tagging of b-quark initiated jets, referred to as b-jets. This task, also known as b-tagging, holds particular significance for the physics programme of the ATLAS experiment, which largely revolves around measurements involving top (t) quarks and Higgs bosons. Both of these particles decay predominantly into final states involving b-quarks.
In t-quark decays, (nearly) all t-quarks decay into a $W^{\pm}$ boson and a b-quark. Additionally, the Higgs boson mainly decays into a pair of b-quarks. The branching ratio for the latter is depicted in Figure 7.1. Furthermore, searches for new physics phenomena, like BSM particles, where a new particle decays into one or multiple b-quarks, vastly benefit from an excellent b-tagging. Even in analyses that do not include b-quarks in the final state, e.g. searches for Supersymmetry (SUSY) particles, b-tagging is employed to veto b-jets.

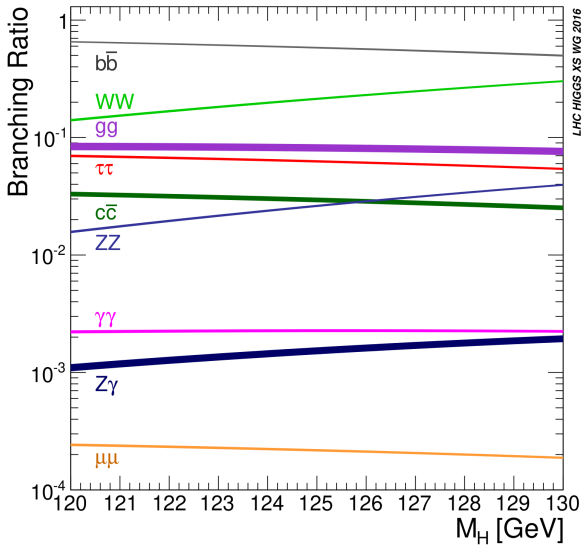


Figure 7.1.: Branching ratios of the Higgs boson as a function of the Higgs mass [15].

The extended lifetime of hadrons containing either b or c-quarks, which is significantly longer than that of other hadrons, enables the identification of b and c-jets. This unique characteristic is due to the weak decay of b and c-quarks, as well as the CKM matrix, where the matrix entries describe the probability of a quark transitioning to a quark of another flavour. With the dominant elements on its diagonal, transitions within one generation are favoured. While this poses no issue for the first generation and the up-type quarks of the second and third generation, the transition of

b and s-quarks to the up-type counterparts of the same generation is not possible due to the higher mass of their generation partners, violating energy conservation. Notably, s-quarks have a mass of $m_s = \left(93.4^{+8.6}_{-3.4}\right)$ MeV, while their generation counterpart, the c-quarks, have a mass of $m_c = (1.27 \pm 0.02)$ GeV. A similar situation holds for b and t-quarks. With the highest mass in the SM, $m_t = (172.69 \pm 0.30)$ GeV, t-quarks are much heavier than b-quarks, which have a mass of $\left(4.18^{+0.03}_{-0.02}\right)$ GeV[1].

Due to the extended lifetime of these hadrons, b-jets have unique properties in their topology, which can be exploited for identification purposes. To illustrate this, the topologies of a b-jet and two non-b-jets are shown in Figure 7.2.

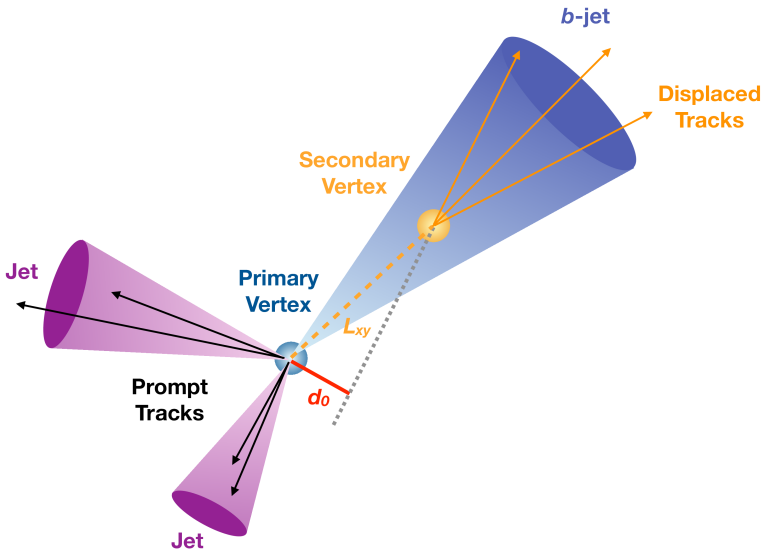


Figure 7.2.: Topologies of a b-jet (blue) and two non-b-jets (violet) [168].

The most crucial among these properties is the presence of secondary and tertiary decay vertices inside the jets and the displaced charged particle tracks, or simply displaced tracks, of their decay products. With a lifetime of around 1.5 ps at $p_{\mathrm{T}} = 50$ GeV, b-hadrons can travel approximately $L_{xy} = 5$ mm within the detector before decaying. This distance from the PV can be resolved, and the resulting IPs of the tracks[2] from the decay products, like d_0 shown in Figure 7.2, can be used to reconstruct additional vertices. With these new decay vertices comes also an increased number of tracks inside the jets, which arises from subsequent decays of the B meson decay products, which is another essential property of b-jets.

To leverage all these properties of the b-jets, multiple key ingredients are needed, including the previously mentioned tracks, PVs, and the hadronic jets clustered from the previous two elements. For the Run 2 algorithms and the b-tagging algorithms discussed in this thesis, these reconstructed

1 Mass values are extracted from Ref. [27]. Note that the s and b-quark masses are given in the Modified Minimal Subtraction ($\overline{\mathrm{MS}}$) renormalisation scheme [167].

2 See Section 5.1.

objects need to satisfy certain quality criteria.

Tracks are reconstructed using the method explained in Section 5.1, where, in addition to the usual track selection criteria, the tracks need to fulfil $p_{\mathrm{T}} > 0.5\,\mathrm{GeV}$. This basic selection can be enhanced by tighter/extra criteria, depending on the algorithm used, to enhance the efficiency for tracks from heavy-flavour hadron decays. PVs are reconstructed using the method explained in Section 5.1, playing a pivotal role in b-tagging since they define the reference points for the calculation of the IPs of the tracks. The jet collection used for b-tagging in Run 2 and Run 3 is the small-R PFlow jet collection, its elements being reconstructed as described in Section 5.4, with a radius parameter of $\Delta R = 0.4$ and an additional selection of $p_{\mathrm{T}} > 20\,\mathrm{GeV}$. Furthermore, for jets with $20\,\mathrm{GeV} < p_{\mathrm{T}} < 60\,\mathrm{GeV}$ and $|\eta| < 2.4$, a Jet Vertex Tagger (JVT) [169] criterion with JVT > 0.5 is applied to suppress pile-up.

To determine the flavour of the quark that initiated the jet, the so-called jet flavour, a ΔR-based labelling procedure is employed in MC. This labelling procedure uses truth information from the particles inside the jet to determine the truth flavour/label. Two labelling schemes are utilised for this, a standard and an extended labelling scheme. While the standard labelling scheme is applied for b-tagging, the extended labelling scheme can be used to further understand the classification of the different algorithms. Both schemes are presented in Figure 7.3.

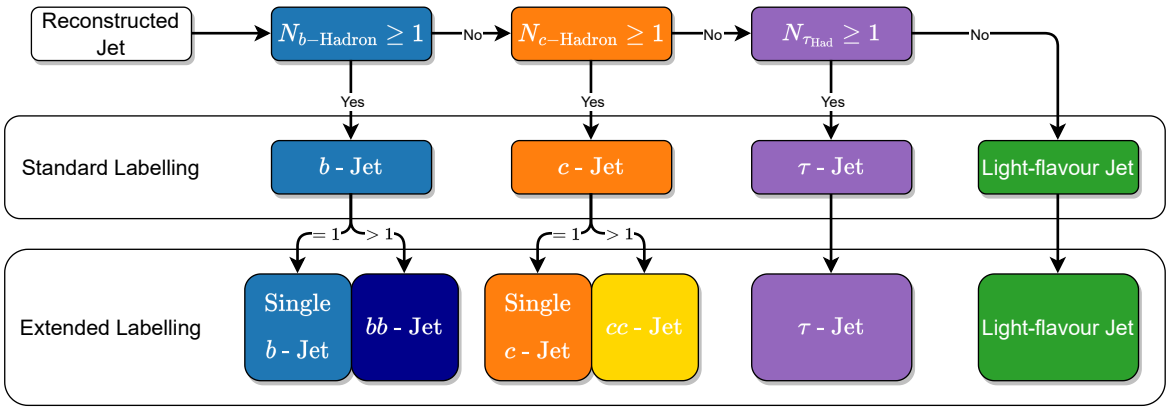


Figure 7.3.: Standard and extended labelling scheme used to classify the truth jet flavour. The number of hadrons/leptons used corresponds to the number of weakly decaying hadrons/leptons within $\Delta R = 0.3$ around the jet axis. If no weakly decaying hadron/lepton is found, the jet is labelled as light-flavour jet, or simply light-jet.

For the standard labelling scheme, jets are labelled as b-jets if at least one weakly decaying b-hadron with $p_{\mathrm{T}} \geq 5\,\mathrm{GeV}$ is located within $\Delta R = 0.3$ around the jet axis. If no b-hadron is found, a similar matching attempt is conducted first for c-hadrons and then for hadronically decaying τ leptons. If a match is found, the jet is labelled respectively as a c or a τ-jet, and the labelling procedure is stopped. If none of the previous searches is successful, the jet is labelled as a light-flavour or simply light jet, originating from u, d, or s-quark, or a gluon.

The extended labelling scheme further categorises b and c-jets into single and double b and c-jets based on the number of weakly decaying hadrons inside the jet. However, the τ and light-flavour jets

are not subjected to further separation[3].

In the following sections, the different algorithms, also called taggers, and their functionality are presented. Starting with an overview of the algorithms used in Run 2, including the Recurrent Neural Network Impact Parameter (RNNIP) [170], Deep Learning 1 (DL1) [171], and Deep Learning 1 RNNIP (DL1r) [172] taggers, also the basics of b-tagging in ATLAS will be explained in more detail. Then, the preparation and processing of the training sample will be discussed, which involves jets from different processes as well as different techniques of preprocessing to prepare the training sample for the training procedure. Finally, the changes from the Run 2 algorithms to the Run 3 algorithms will be discussed, including the new Deep Impact Parameter Sets (DIPS) and DL1d taggers. The latter one is the new and currently recommended ATLAS b-tagging algorithm for Run 3 of the LHC.

7.1 HEAVY-FLAVOUR TAGGING IN RUN 2

The b-tagging algorithms employed in Run 2 of the LHC are designed to identify b-jets while rejecting all other types of jets efficiently. Different algorithms, each exploiting different features of the b-jet topology are combined to achieve this goal. There are two main classes of algorithms: low-level algorithms, which focus on one or two specific properties of b-jets, and high-level algorithms, which combine the outputs of various low-level algorithms to predict the jet's flavour.
This section will provide a detailed presentation of the low-level track-based algorithm RNNIP and the final Run 2 high-level b-tagging algorithm, DL1r. Both algorithms hold significant roles as reference for the DIPS and DL1d algorithms, which will be explained in detail in Chapter 8 and 9, respectively.[4]. A more comprehensive overview of all algorithms used in Run 2, including the various not-mentioned low-level algorithms, can be found in Ref. [172]. Furthermore, a detailed description of the calibrations performed for the high-level tagger can be found in References [171, 173–175].

7.1.1 *The Recurrent Neural Network Impact Parameter Tagger*

The most important of the many low-level algorithms used in Run 2 is the RNNIP tagger. While officially being a low-level algorithm, RNNIP is, similar to the high-level algorithms, a DNN based on the RNN architecture (Section 6.2). The input sequence for RNNIP is composed of the tracks associated with the jet. The architecture of RNNIP is depicted in Figure 7.4. RNNIP aims to exploit the large IPs of tracks stemming from secondary vertices. While other low-level algorithms use LH templates to extract conditional LHs of the flavours, a process that requires very large event samples to compute those templates. In contrast, RNNIP can be trained with a significantly smaller event sample by focusing on the inter-sequence correlations between different tracks and their variables.

3 It is worth noting that further separation of both the τ- and light-flavour jets is possible, but not implemented or used.
4 It is worth noting that all performance plots in Section 7.1 were generated using the PUMA package [1].

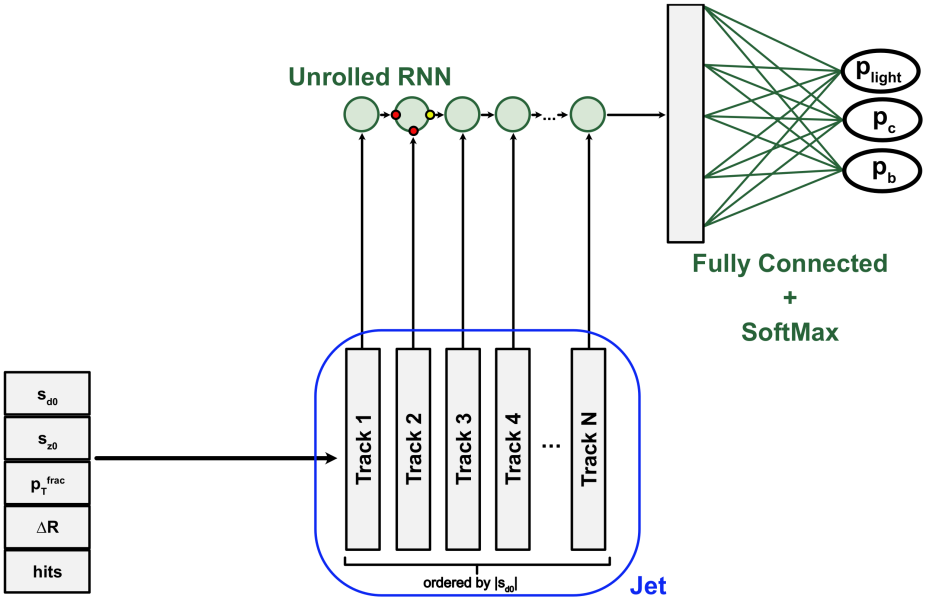


Figure 7.4.: Schematic depiction of the architecture of the RNNIP tagger [172]. The RNN architecture is shown here in an unrolled manner. The output classes p_b, p_c and p_{light} depict the output probabilities of the RNNIP, where each of them can be interpreted as a probability that the processed jet belongs to the respective class.

These correlations are also one of the biggest improvements of RNNIP over the other low-level algorithms, which assume in the LH-based approach that the variables of each track are uncorrelated.

RNNIP takes the lifetime signed transverse and longitudinal IP significances (S_{d_0} and S_{z_0}) as input variables for each track. It also considers the fraction of transverse momentum of the track with respect to the jet ($p_{\text{T}}^{\text{frac}}$), the angular distance between the track and the jet axis ($\Delta R\,(\text{track, jet})$), and the hit multiplicities of the track with different ID layers. The signed IP significances are defined in Equations (7.1) and (7.2). The sign of the IP significances is determined by the position of the point of closest approach from which the IP is measured. If the point is in front of the primary vertex with respect to the jet direction, a positive sign is applied. In the other case, where the point is behind, a negative sign is applied. This signature is done for both IP significances.

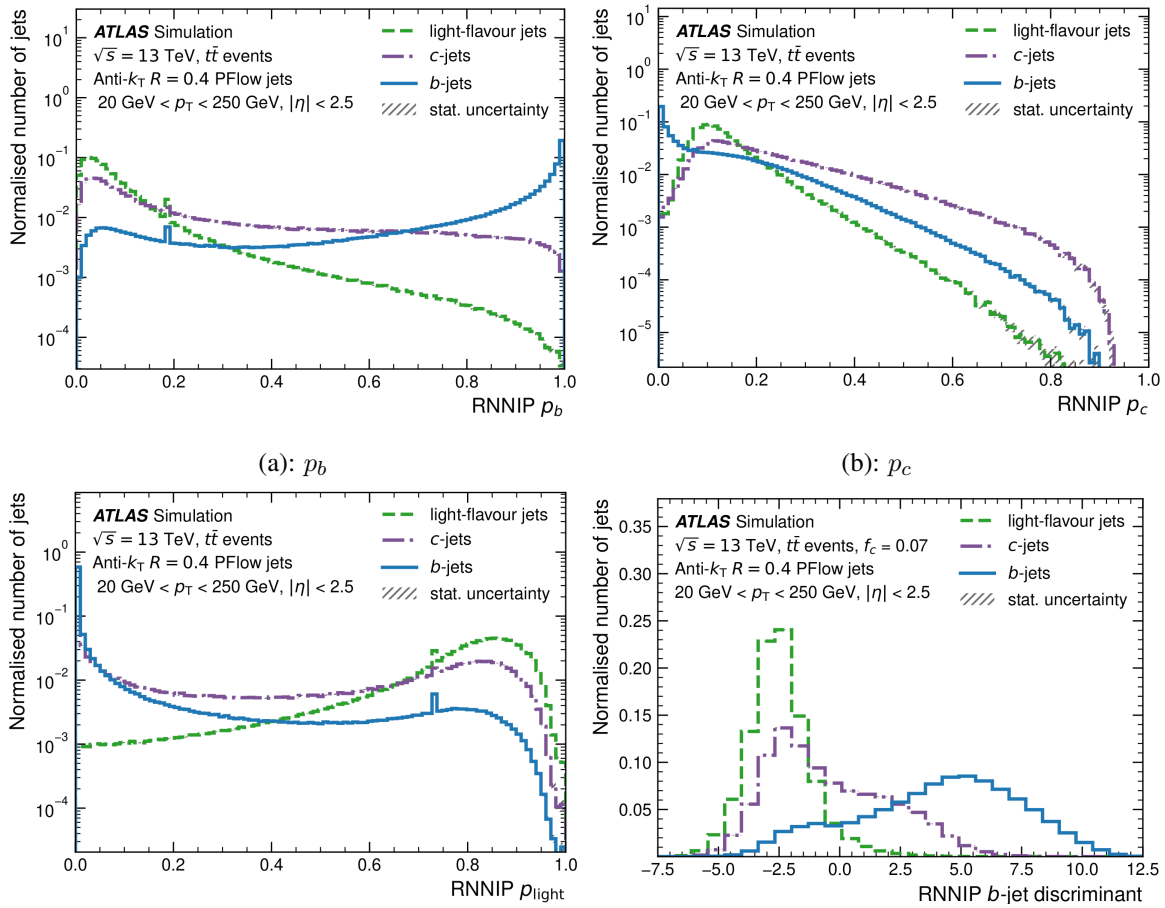
$$S_{d_0} = \begin{cases} +\frac{d_0}{\sigma_{d_0}} & \text{if in front of primary vertex} \\ -\frac{d_0}{\sigma_{d_0}} & \text{if behind of primary vertex} \end{cases} \tag{7.1}$$

$$S_{z_0} = \begin{cases} +\frac{z_0}{\sigma_{z_0}} & \text{if in front of primary vertex} \\ -\frac{z_0}{\sigma_{z_0}} & \text{if behind of primary vertex} \end{cases} \tag{7.2}$$

$$p_{\text{T}}^{\text{frac}} = \frac{p_{\text{T}}^{\text{track}}}{p_{\text{T}}^{\text{jet}}} \tag{7.3}$$

$$\Delta R\,(\text{track, jet}) = \sqrt{(\phi_{\text{track}} - \phi_{\text{jet}})^2 + (\eta_{\text{track}} - \eta_{\text{jet}})^2} \tag{7.4}$$

As depicted in Figure 7.4, RNNIP produces three values as its output prediction. With the softmax activation function applied in the output layer, these values sum up to one, representing probabilities



(a): p_b (b): p_c

(c): p_{light} (d): b-tagging discriminant

Figure 7.5.: Output probabilities for the jet being either a b (a), c (b), or a light-flavour jet (c). In addition the b-tagging discriminant built from these three probabilities is shown in (d). The used fraction values are $f_c = 0.07$ and $f_{\text{light}} = 0.93$. The spikes in the p_b and p_{light} distributions stem from jets, where no jet survives the track selection. The jets shown in these plots are taken from $t\bar{t}$ events [172].

that the processed jet originated from a b, c, or light-flavour quark[5]. These output probabilities are then combined using a LH-inspired discriminant known as the b-tagging discriminant, defined in Equation (7.5).

$$D_{\text{RNNIP}} = \ln\left(\frac{p_b}{f_c \cdot p_c + f_{\text{light}} \cdot p_{\text{light}}}\right) \quad \text{with } f_c + f_{\text{light}} \stackrel{!}{=} 1 \tag{7.5}$$

Here, p_b, p_c, and p_{light} represent the output probabilities for their respective class. The parameters f_c and f_{light} act as relative importance parameters, often referred to as fraction values, influencing the background rejection[6]. By definition, these values must add up to one, allowing f_{light} to be replaced with $(1 - f_c)$. This approach allows for a tuning of the rejections based on the fraction values. For RNNIP, the values for f_c and f_{light} were optimised and set to $f_c = 0.07$ and $f_{\text{light}} = 0.93$. The output probabilities and the b-tagging discriminant derived from them are illustrated in Figure 7.5.

5 Light-flavour quark denotes here u, d, s-quarks, and gluons

6 Background rejection is defined as $\frac{1}{\epsilon_{\text{Bkg}}}$, where ϵ_{Bkg} denotes the background efficiency or mis-tag efficiency for one class.

Table 7.1.: Additional track selection requirements which are applied for the tracks used by RNNIP. Taken and adapted from Ref. [2].

Variable	Requirement
Max N_{Tracks}	25
p_{T}	$> 1\,\text{GeV}$
$\lvert d_0 \rvert$	$< 1\,\text{mm}$
$\lvert z_0 \sin(\theta) \rvert$	$< 1.5\,\text{mm}$
$\lvert \eta \rvert$	< 2.5
$N_{\text{Pixel holes}}$	< 2
$N_{\text{Silicon hits}}$	≥ 7
$N_{\text{Silicon shared hits}}$	< 2
$N_{\text{Silicon holes}}$	< 3

RNNIP is trained using a combined dataset of three million jets, comprising $70\,\%$ $t\bar{t}$ jets and $30\,\%$ jets extracted from $Z' \rightarrow q\bar{q}$ events[7]. The inclusion of Z' jets aimed to enhance training statistics in the high-p_{T} regime. To prevent a kinematic bias during training, the different flavours were reweighted in p_{T} to align with the distribution of light-flavour jets. RNNIP utilised up to 25 tracks per jet[8], ordered in decreasing order based on S_{d_0}. In addition to the already applied quality track criteria, extra requirements to further enhance the quality of the tracks were applied. These requirements are listed in Table 7.1. The different pixel and silicon variables mentioned therein define the number of hits, holes[9] and multiple hits in one module for the given track in the respective ID part.

The network is trained for 50 epochs using KERAS [176] with the THEANO [177] backend, employing the ADAM optimiser [178]. The trainable parameters are initialised using a Glorot Uniform distribution [179]. The trained model is implemented for inference within the ATLAS reconstruction software [180] using the Lightweight Trained Neural Network (LWTNN) package [181].
In the subsequent parts, particularly in Chapter 8, RNNIP serves as a reference for the newly developed algorithms.

7.1.2 *High-Level Algorithms*

Although the output of low-level algorithms can be directly used for tagging b and c-jets, the correlation among their outputs depends on the jet flavour and the kinematic region. Generally, for b and c-jets, significant correlations exist between the outputs of various low-level algorithms, while this is not the case for light-flavour jets. In the high-p_{T} regime, these correlations further diminish due to the increased number of tracks from other sources and the decreased precision of track reconstruction. Additionally, RNNIP contributes its output probabilities, which are not highly correlated with

7 More detail on the Z' sample is provided in Section 7.2
8 While the original version of RNNIP was trained with pure $t\bar{t}$ jets and up to 15 tracks per jet, the retrained version here uses up to 25 tracks per jet.
9 Holes are described as missing hits in one of the ID layers, which the track should have transversed through.

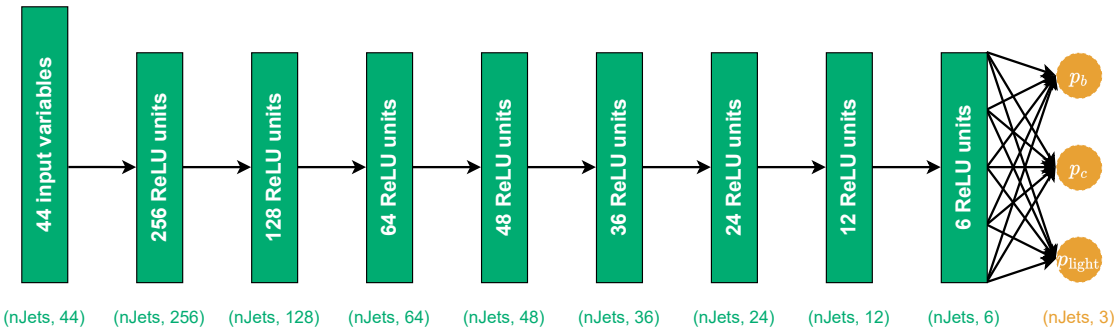


Figure 7.6.: Schematic depiction of the architecture of the DL1r tagger. The output classes p_b, p_c and p_{light} depict the output probabilities of DL1r, where each of them can be interpreted as a probability that the processed jet belongs to the respective class.

the outputs of the other low-level algorithms. The goal of high-level algorithms is to exploit these correlation differences and achieve the best possible performance in tagging b-jets.

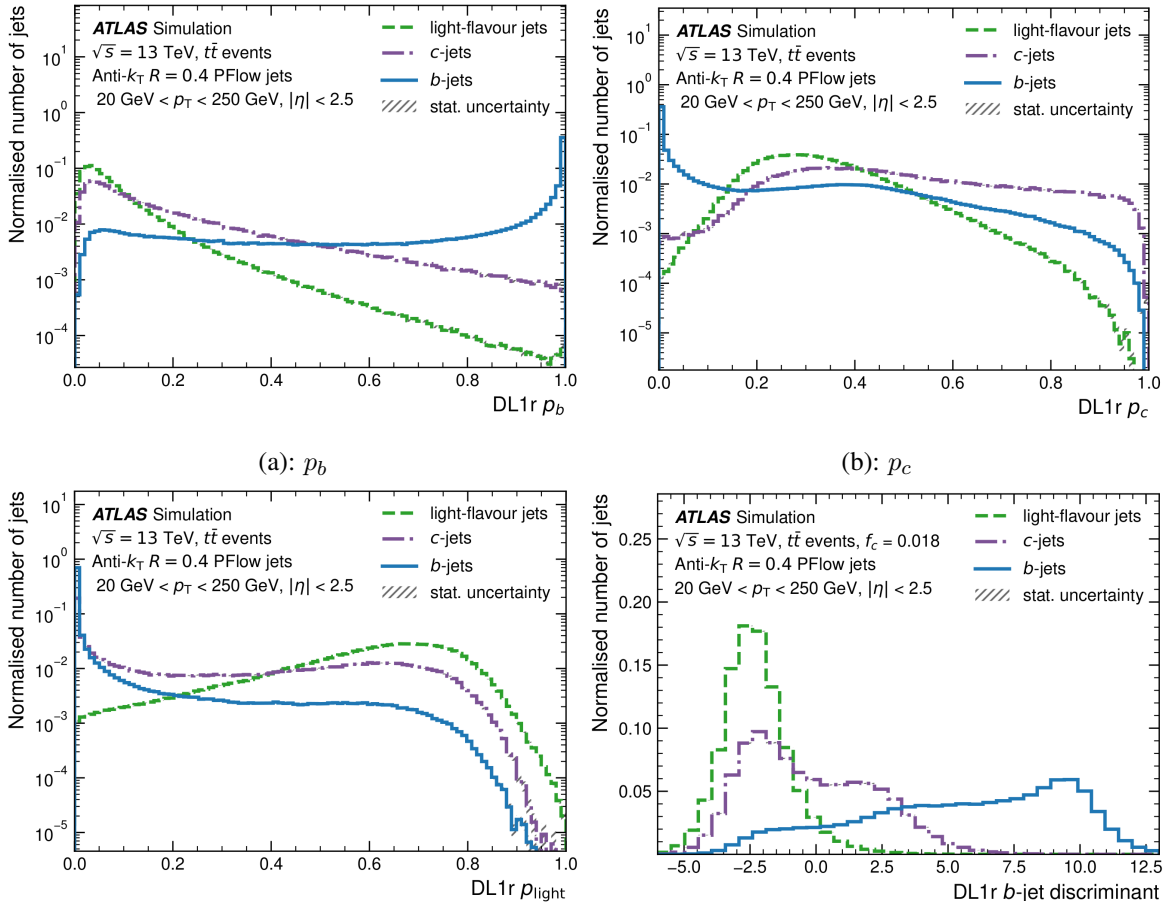
These high-level algorithms are fully connected feedforward DNNs, following the architecture of MLPs and undergo supervised training. Two types of high-level algorithms will be discussed in this section, both belonging to the same algorithm family, the DL1 family. Both algorithms, DL1 and DL1r, take as input the jet p_{T} and $|\eta|$ values, as well as the output of different low-level algorithms. A comprehensive list of input variables for both algorithms can be found in Table A.1.

Both algorithms were trained on a dataset of 22 million jets, comprising 70 % $t\bar{t}$ jets and 30 % Z' jets. To mitigate a potential kinematic bias, a two-dimensional resampling in jet p_{T} and $|\eta|$ is applied. The full preprocessing chain is explained in more detail in Section 7.2. Both DL1 and DL1r are implemented and trained in the UMAMI framework, utilising TENSORFLOW with the KERAS front-end and ADAM as the optimiser for training. The architecture is schematically depicted in Figure 7.6, while a more detailed description of the hyperparameters can be found in Table A.2. DL1 uses the same hyperparameters and architecture as DL1r, excluding the RNNIP probabilities as input. After training, network inference within the ATLAS reconstruction software is performed using the LWTNN package [181].

Similar to RNNIP, the DL1 and DL1r output can be interpreted as class probabilities due to the usage of the softmax activation function in the output layer. These probabilities are then combined into the b-tagging discriminant using Equation (7.6), which is akin to the b-tagging discriminant used in RNNIP [Equation (7.5)].

$$D_{\text{DL1r}} = \ln\left(\frac{p_b}{f_c \cdot p_c + f_{\text{light}} \cdot p_{\text{light}}}\right) \quad \text{with } f_c + f_{\text{light}} \stackrel{!}{=} 1 \tag{7.6}$$

The fraction values for DL1 and DL1r were optimised to achieve a good compromise between c and light-flavour jets rejection across the entire b-jets efficiency range. This optimisation was performed considering multiple analyses in ATLAS, in particular, the measurements of $VH\,(H \to b\bar{b})$ [182] and $t\bar{t}H\,(H \to b\bar{b})$ [14] production. The optimal fraction values were found to be $f_c = 0.018$ and $f_{\text{light}} = 0.982$. The results of the training, including the output distributions for the class probabilities and the b-tagging discriminant, are shown in Figure 7.7.



(a): p_b (b): p_c

(c): p_{light} (d): b-tagging discriminant

Figure 7.7.: Output probabilities for the jet being either a b (a), c (b), or a light-flavour jet (c) [172]. In addition the b-tagging discriminant built from these three probabilities is shown in (d). The used fraction values are $f_c = 0.018$ and $f_{\text{light}} = 0.982$. The jets shown in these plots are taken from $t\bar{t}$ events.

Comparing Figures 7.7 (d) and 7.5 (d), the expectation of an improved discriminating power of DL1r over RNNIP is confirmed. This improvement is due to DL1r utilising more information from all the low-level algorithms, including RNNIP.

In analysis applications, a binary decision must be made whether a jet is tagged as a b-jet or not-tagged. This decision is based on Working Points (WPs)[10], defined by the b-tagging discriminant. In ATLAS, the provided and supported WPs are the 60 %, 70 %, 77 %, and 85 % WPs. Their definition is closely related to their naming, representing the percentage of b-jets that are tagged when the WPs are applied. To define these WPs, one starts integrating the b-jet distribution from the right side of the discriminant plot to the left. After integrating a specific percentage of the entire b-jet distribution, the discriminant value at this threshold is adopted as the cut value for the WP. This cut value determines the limit for whether a jet is tagged with the given WP or not. For example, assume the cut value for the 70 % WP might be 3. If a jet has a b-jet discriminant above 3, it is tagged as a b-jet; otherwise, it is not tagged. The same logic applies to all other WPs. With this simple definition comes also that a jet that is tagged with 60 % WP is also tagged with all other WPs. In application, these WPs and

10 It is worth noting here that in newer publications, the term Operation Point (OP) is used.

their cut values are defined using jets from simulated $t\bar{t}$ events[11].
Based on these so-called fixed-cut WPs are the so-called Pseudo-Continuous b-Tagging (PCBT) bins, which split the given b-tagging discriminant into different tag-weight bins, using the fixed-cut WPs as bin edges. The so-defined bins are the $100 - 85\,\%$, $85 - 77\,\%$, $77 - 70\,\%$, $70 - 60\,\%$ and $60 - 0\,\%$ bins. To simplify this a bit further, a jet that is tagged with the $70\,\%$ fixed-cut WP but fails to be tagged by the $60\,\%$ fixed-cut WP would be situated in the $70 - 60\,\%$ bin. This, as the name already suggests, pseudo-continuous binning can be very helpful when working on analyses involving c-jets, due to the relatively large population of c-jets in the central PCBT bins. In addition, the PCBT bins are often used to train analysis-specific ML algorithms due to the higher amount of information in the binning when compared to the simple binary tagged or not-tagged information from the fixed-cut WPs.

The performance of the DL1 and DL1r taggers is now being compared. So-called Receiver Operating Characteristic (ROC) curves are utilised in an adapted way, wherein the mis-tag efficiency of the background classes ϵ_{Bkg} is plotted against the tagging efficiency of the signal class ϵ_{Signal}. In b-tagging, the inverse of the mis-tag efficiency, denoted as $\frac{1}{\epsilon_{\text{Bkg}}}$ or 'rejection' is typically employed due to the very small values of ϵ_{Bkg} for smaller ϵ_{Signal}. The resulting ROC curves for DL1 and DL1r are displayed in Figure 7.8, with the former recommended ATLAS b-tagging algorithm, MV2c10 [171], which is a Boosted Decision Tree (BDT) [183]-based algorithm, serving as the reference[12].

The fixed-cut WPs previously mentioned are positioned on the x-axis, with 0.60 corresponding to the $60\,\%$ WP, and so forth. Upon closer examination of the $60\,\%$ WP, the performance gain for both DL1 and DL1r over MV2c10 is evident in both c and light-flavour jets rejection. DL1r exhibits an increase in c-jets rejection over MV2c10 by a factor of 1.76 and for light-flavour jets rejection by a factor of 2.06.
In addition to the inclusive performance comparison with ROC curves, the background rejections for a specific fixed-cut WP can be compared as a function of p_{T}, which is crucial for analyses targeting high-p_{T} regions, due to the notably more challenging reconstruction of certain features such as secondary vertices. The c and light-flavour jets rejections as a function of p_{T} for the $77\,\%$ WP of the three high-level algorithms are illustrated in Figure 7.9. In each bin, the b-jets efficiency is set to $77\,\%$, ensuring that $77\,\%$ of all b-jets in the corresponding bin are tagged. This flat efficiency per bin procedure slightly differs from the overall inclusive definition of the fixed-cut WPs, as their definition relies on the inclusive b-jet p_{T} distribution, encompassing all p_{T} bins.

These plots, depicting flat efficiency per bin, offer valuable insights into evaluating the networks' performance within specific p_{T} bins, avoiding the averaging of performance across the entire p_{T} range. DL1r consistently outperforms both DL1 and the reference MV2c10 across the entire p_{T} spectrum, exhibiting even greater performance gains over MV2c10 in the high-p_{T} regime.
Furthermore, the same bins are presented again in Figure 7.10, but with the inclusive procedure

11 The usage of $t\bar{t}$ jets can be explained with the focus of the ATLAS physics program towards processes, which involve t-quarks. Also, differences between various processes and data and simulation will be covered by the calibration of these WPs, which rely on $t\bar{t}$ jets.

12 Due to its negligible meaning for this thesis, no explanation for BDTs or MV2c10 is provided.

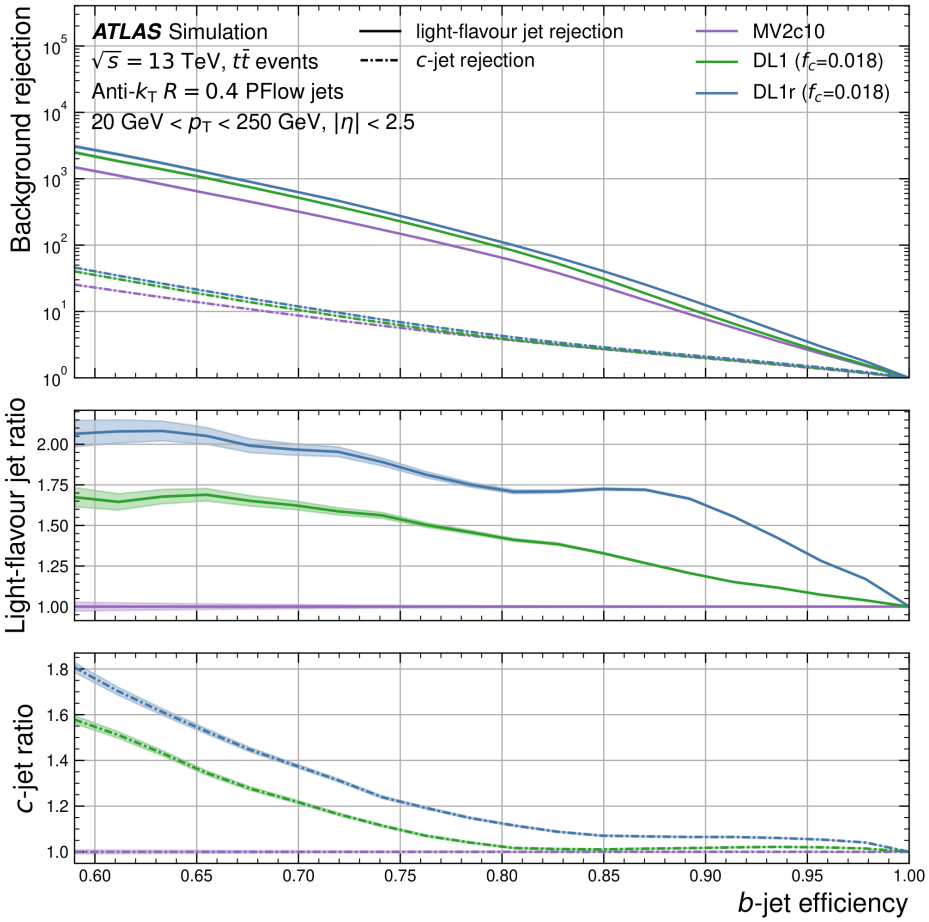
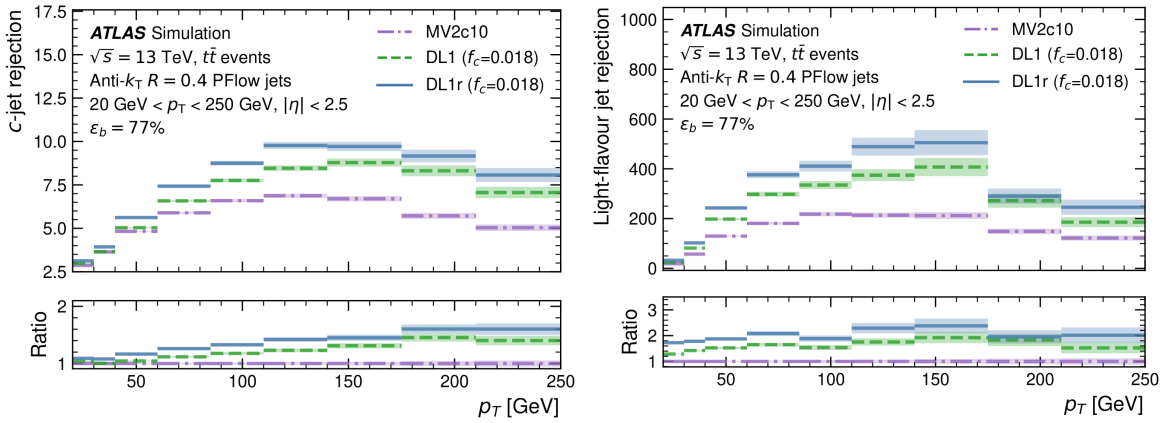
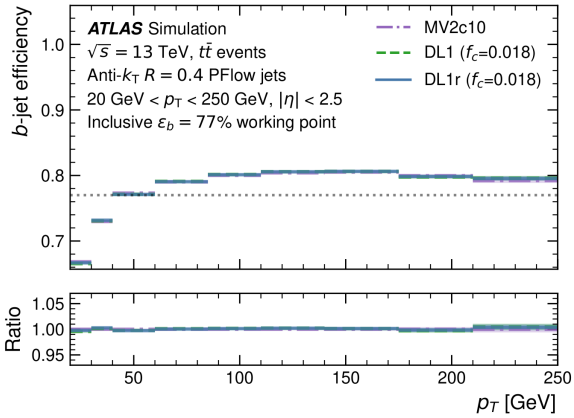


Figure 7.8.: ROC plot with the c and light-flavour jets rejections as a function of the b-jet efficiency ϵ_b for the MV2c10, DL1 and DL1r high-level algorithms. The upper subplot shows both rejections for all three algorithms, while the lower panels show the ratio of the high-level algorithms with MV2c10 as a baseline for certain rejections. Statistical uncertainties were derived using binomial uncertainties. They are shown as coloured bands around the main plotted line [172].

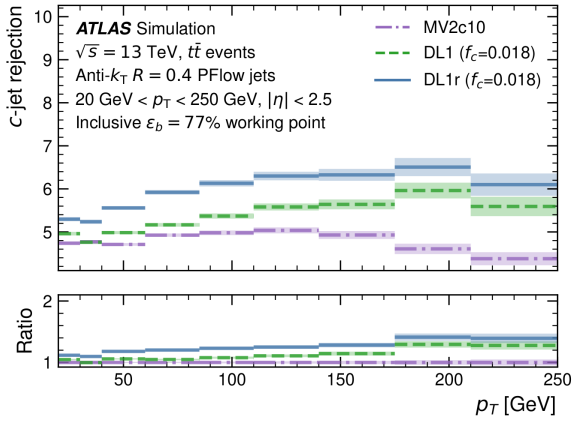


(a): c-jets rejection (b): Light-flavour jets rejection

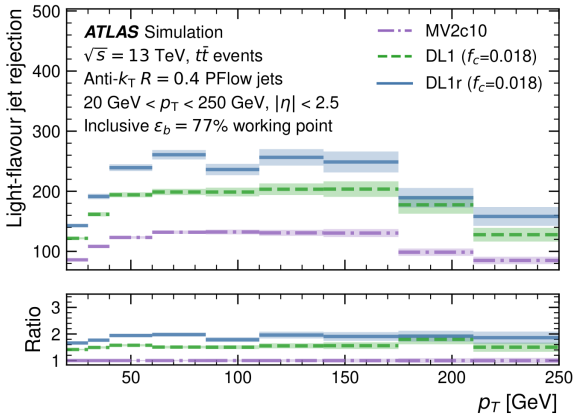
Figure 7.9.: The c (a) and light-flavour (b) jets rejections as a function of p_{T} for the 77 % WP. The flat efficiency per bin procedure is applied, so that in each p_{T} bin, the b-jets efficiency is at 77 %. The upper panel shows the overall rejections, while the lower one shows the ratio between the three high-level algorithms with MV2c10 as a baseline. Statistical uncertainties were derived using binomial uncertainties. They are shown as coloured bands around the main plotted line [172].



(a): b-jets efficiency



(b): c-jets rejection



(c): Light-flavour jets rejection

Figure 7.10.: The b-jets efficiency (a) and the c (b) and light-flavour (c) jets rejections as a function of p_{T} for the 77 % WP [172]. The inclusive efficiency WP is used, so that the inclusive b-jet efficiency is at 77 %. The upper panel shows the overall rejections/efficiencies, while the lower one shows the ratio between the three high-level algorithms with MV2c10 as baseline. Statistical uncertainties were derived using binomial uncertainties. They are shown as coloured bands around the main plotted line.

for defining the fixed-cut WPs. These plots showcase the averaged performance of the algorithms, which is also later employed in the analyses. Due to the inclusive determination of the WPs, the b-jets efficiency represents an average over the full p_{T} range, a factor that needs consideration. To illustrate this behaviour, the b-tagging efficiency as a function of p_{T} is also presented.

In contrast to the flat efficiency per bin plots, these plots illustrate the actual performance of the algorithms for a given WP as a function of p_{T}. As depicted in Figure 7.10 (a), the b-jets efficiency, on average, stands at 77 %, although the actual efficiencies for each bin deviate from this value. In the lower p_{T} bins, the b-jets efficiency tends to fall below the WP, while in the higher p_{T} bins, the efficiencies tend to be higher.

Overall, DL1r exhibits significant improvement over the reference MV2c10 and the DL1 algorithms, establishing itself as the most effective algorithm during Run 2. DL1r is the recommended b-tagging algorithm for Run 2.

7.2 PREPARATION OF THE TRAINING SAMPLE

Training a b-tagging algorithm, or b-tagger, based on supervised NNs requires a training sample. These training samples, utilised for all NN based taggers elucidated in this thesis, comprise jets from two distinct processes. This so-called 'hybrid' sample is constructed by combining jets from SM $t\bar{t}$ and $Z' \rightarrow q\bar{q}$ events. The Z' resonance used here is a non-SM particle, whose decay width is modified such that the mass spectrum is approximately flat. This approach of using jets from two processes ensures that a high number of jets are available for training the taggers, even in the very high-p_{T} regime, ranging from above 1 TeV up to 5 TeV.

The next sections delve into a detailed explanation of the $t\bar{t}$ and Z' MC samples. Subsequently, a more in-depth exploration of kinematic resampling and the final preparation of the training sample will be presented. Lastly, additional insights will be provided on the creation of the validation and test samples, crucial for evaluating and testing the performance of the networks.

7.2.1 *Monte Carlo Samples*

The $t\bar{t}$ simulation sample is generated using POWHEG BOX V2 [184–186], which provides MEs at NLO for t-quark pair production. The cut-off scale parameter for the first-gluon-emission, denoted as h_{damp}, was set to $1.5\,m_t$, where m_t represents the t-quark mass, set to $m_t = 172.5\,\text{GeV}$. For the interface of POWHEG BOX V2, PYTHIA 8.230 [187] was employed with the A14 set of tuned parameters [188] and NNPDF3.0NLO (NNPDF2.3LO) [189, 190] parton distribution functions in the MEs (PS). Only $t\bar{t}$ events where at least one of the subsequent $W^{\pm}$ bosons decays leptonically[13] were considered. This choice ensures the presence of sufficient b, c, and light-flavour jets for training. Additionally, hadronically decaying τ-leptons are present, which are reclustered as jets, known as τ-jets, and will be further explored in Chapter 10.
The Z' simulation sample employed was generated using PYTHIA 8.212 with the A14 set of tuned parameters and NNPDF2.3$_{\text{LO}}$ parton distribution functions. The cross-section of the hard-scattering process was artificially adjusted by applying a per-event weight that broadens the natural decay width of the Z' resonance, resulting in the creation of the so-called flat mass Z' sample. Branching fractions were also adjusted, as indicated in Table 7.2, to ensure an adequate representation of b, c, and light-flavour jets in the final sample used for training.

Similar to the $t\bar{t}$ events, the Z' events also contain a small amount of τ-jets, which will also play an important role in Chapter 10. The resulting Z' events have a relatively flat jet p_{T} spectrum from 250 GeV to 1.5 TeV with a falling tail up to 6 TeV.

The EVTGEN [191] package was employed to simulate the decays of heavy-flavour hadrons, and PYTHIA 8.160 with the A3 set of tuned parameters [192] and NNPDF2.3$_{\text{LO}}$ parton distribution

13 This indicates that the $W^{\pm}$ boson decays into a charged lepton and a neutrino.

Table 7.2.: Branching factions of the Z' resonance used in training the different taggers [172].

Decay	Branching fraction
$b\bar{b}$	30 %
$c\bar{c}$	30 %
$s\bar{s}$	10 %
$d\bar{d}$	10 %
$u\bar{u}$	10 %
$\tau^-\tau^+$	5 %
e^-e^+	5 %

functions was used to additionally overlay all simulated events with minimum-bias interactions, simulating the effect of pile-up. The resulting events are reweighted to match the distribution of the average number of interactions per bunch-crossing observed in the corresponding data samples. Subsequently, the events undergo full detector simulation, using GEANT4 [120, 121]. It is worth noting that interactions between b and c-hadrons and τ-leptons with the detector material, particularly significant in the high-p_{T} regime, were not simulated but were addressed through dedicated correction factors and associated uncertainties.

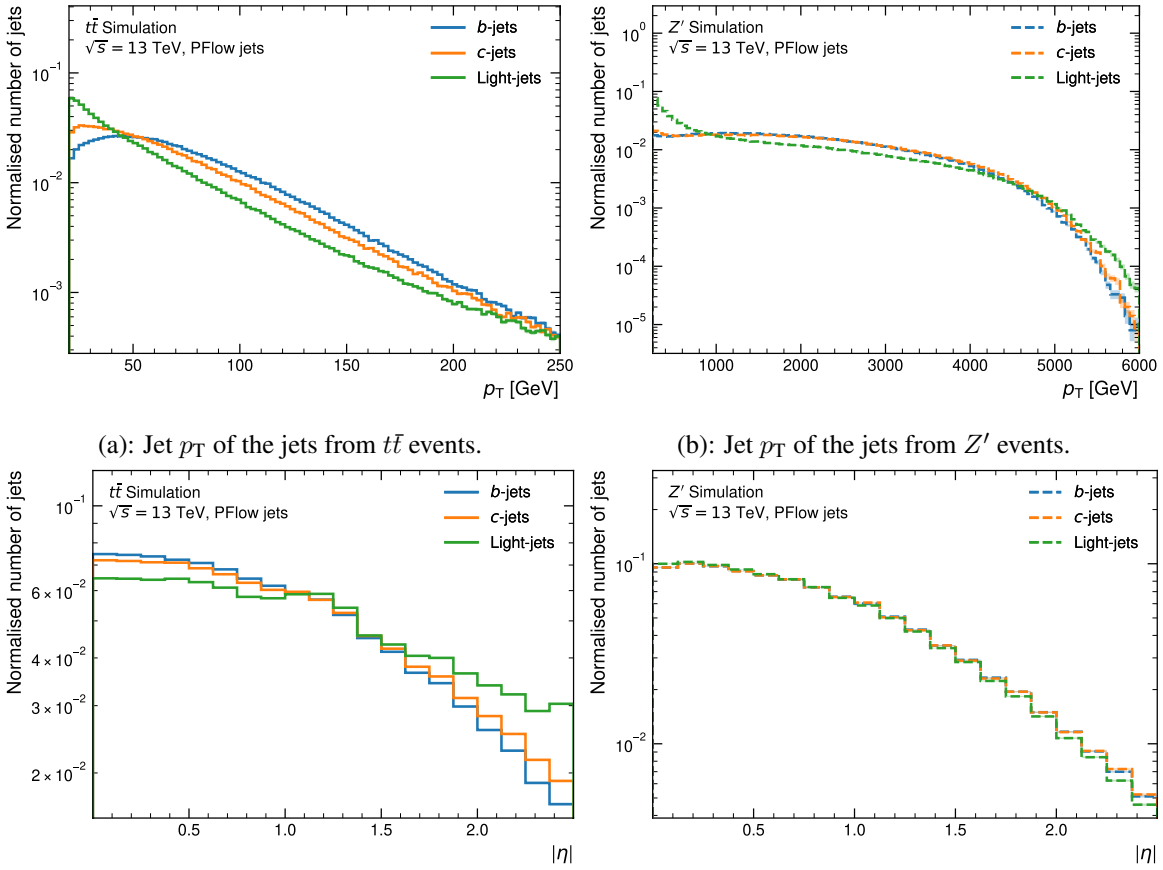
7.2.2 *Kinematic Resampling*

The hybrid training sample is constructed by combining jets from all employed jet flavours/classes of $t\bar{t}$ and Z' events[14]. For the low-p_{T} regime, $t\bar{t}$ jets, selected if $20\,\mathrm{GeV} \leq p_{\mathrm{T}} \leq 250\,\mathrm{GeV}$, are used and no Z' jets due to simulation issues in the true b-meson p_{T} for low-p_{T} Z' jets. In contrast, for the high-p_{T} regime, Z' jets are considered, taking into account those with p_{T} exceeding $250\,\mathrm{GeV}$. No $t\bar{t}$ jets are used in this regime due to a lack of statistics. The p_{T} and $|\eta|$ distributions for the used jet classes for both $t\bar{t}$ and Z' jets, considered for the hybrid training sample, are depicted in Figure 7.11.

The distributions for b and c-jets closely resemble each other, while the light-flavour jets distributions deviate from the other distributions. In addition to that, the overall number of jets per jet class is also drastically different, which would resemble a bias in the training. These differences in distributions underscore the necessity for a preprocessing step before commencing the actual training. While one might intuitively consider leveraging these distribution differences for better jet flavour separation, it rather introduces biases than a performance improvement.

To mitigate such biases in training, a two-dimensional resampling in jet p_{T} and $|\eta|$ for both $t\bar{t}$ and Z' jets is derived and applied. Unlike the count-based resampling method used for Run 2 algorithms, which downsampled all jet classes to the lowest available one, as detailed in Ref. [193], the new taggers (Chapters 8–10) use a training sample created with the so-called PDF resampling method. This resampling method, developed in ATLAS, is not unique and can be conceptualised as an importance

14 For the standard b-tagging case, this means b, c, and light-flavour jets, labelled with the standard labelling procedure (Figure 7.3).



(a): Jet p_T of the jets from $t\bar{t}$ events.

(b): Jet p_T of the jets from Z' events.

(c): Jet $|\eta|$ of the jets from $t\bar{t}$ events.

(d): Jet $|\eta|$ of the jets from Z' events.

Figure 7.11.: Jet p_T and $|\eta|$ distributions of the jets stemming from $t\bar{t}$ and Z' events. The jets shown are the ones which are used in the training of the taggers.

sampling with replacement. A schematic representation of the derivation of the PDF resampling is depicted in Figure 7.12.

To begin, one must designate the target class to which everything will be resampled. For the taggers discussed in this thesis, the target was defined as the b-jet p_T and $|\eta|$ distributions. The subsequent steps involve normalising all distributions and establishing a two-dimensional binning in jet p_T and $|\eta|$, which will serve as the basis for resampling. For each resulting bin, the ratios between the target and the other jet classes are computed. These ratios for each jet class are then interpolated using a two-dimensional spline, which is saved and subsequently employed for the actual resampling process.

Once the PDFs for each non-target (background) class to the target class are derived, the resampling algorithm initiates by populating the jet p_T and $|\eta|$ bins with the target jets. The number of jets filled at this stage is user-determined and can be adjusted to create samples of different sizes. Following the target filling, the background classes are filled iteratively. By looping over the available number of jets for a background class, these jets are distributed into the bins based on their p_T and $|\eta|$, using the previously derived PDFs. Once iterated over the full available amount, the process starts again iterating until enough jets are filled in each bin to match the number of target jets. This is depicted in Figure 7.13.

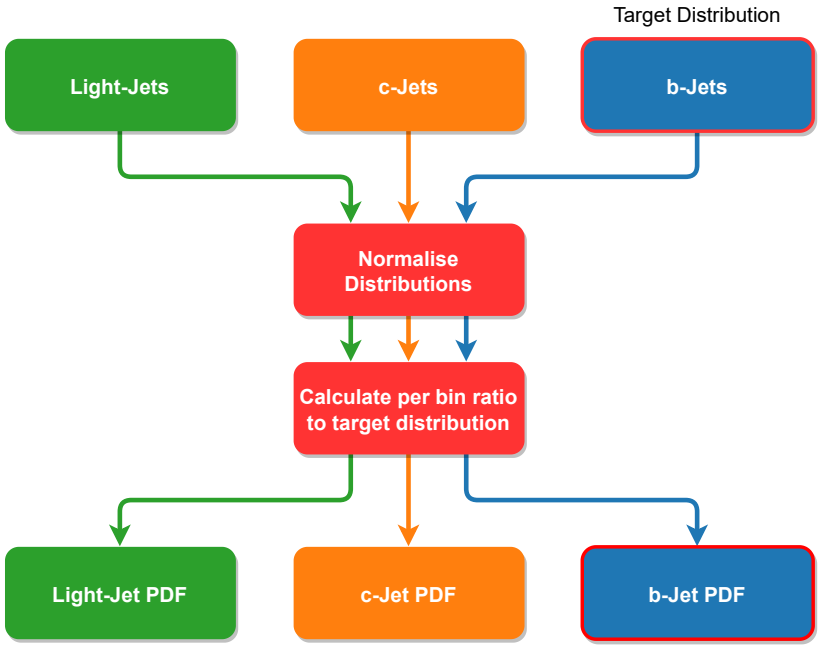


Figure 7.12.: Schematic depiction of the process to calculate the PDFs for the different jet flavours.

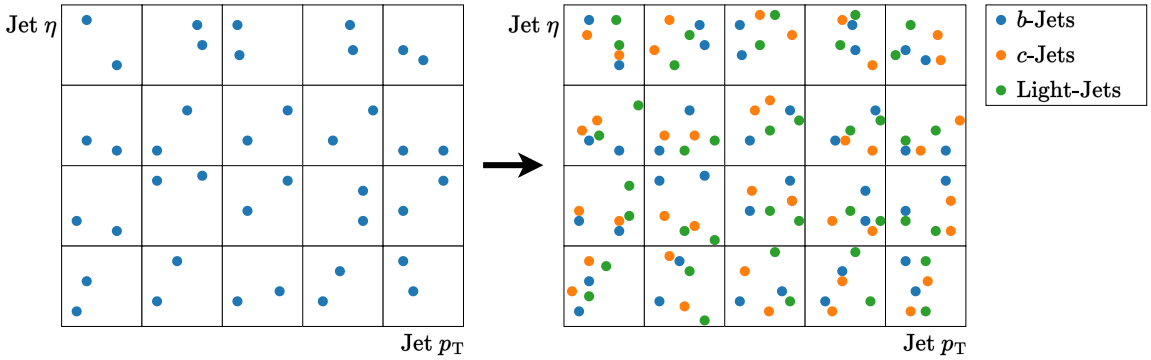


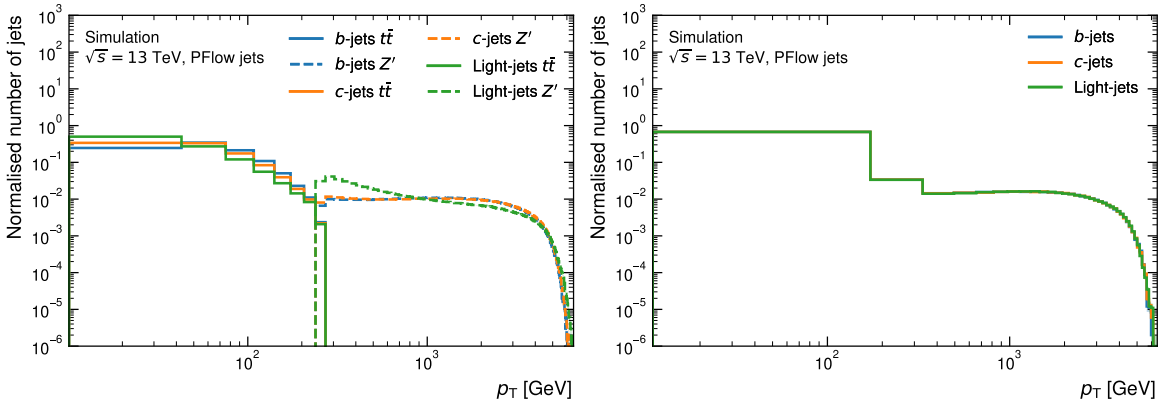
Figure 7.13.: Schematic representation of the PDF resampling method and how the jets are filled in the two-dimensional bins. Taken and adapted from Ref. [193].

Once the resampling process is concluded, the two resampled training samples for both $t\bar{t}$ and Z' jets are combined. The un-resampled and resampled merged training samples are illustrated in Figure 7.14.

In addition to the newly matched distributions, also the number of jets for each jet class are now matching, which is schematically shown in Figure 7.15.

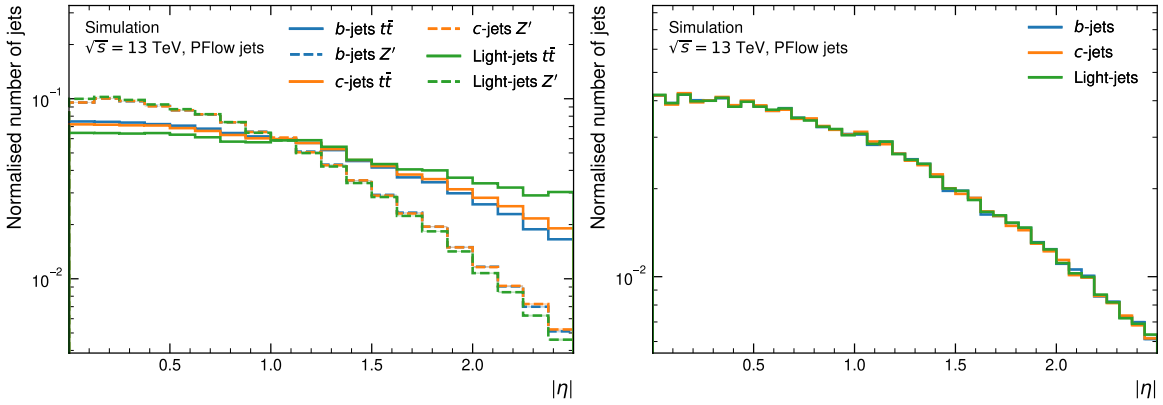
7.2.3 *Scaling and Shifting*

The resampled training sample comprises of jets and their associated tracks with different variables, each having different ranges and magnitudes. For instance, jet p_{T} ranges from 20 GeV to 6 TeV, while jet $|\eta|$ spans from 0 to 2.5. These differences in magnitudes among input variables can significantly impact the convergence of training for NNs optimised using gradient descent. In the early



(a): Jet p_T of the un-resampled training sample.

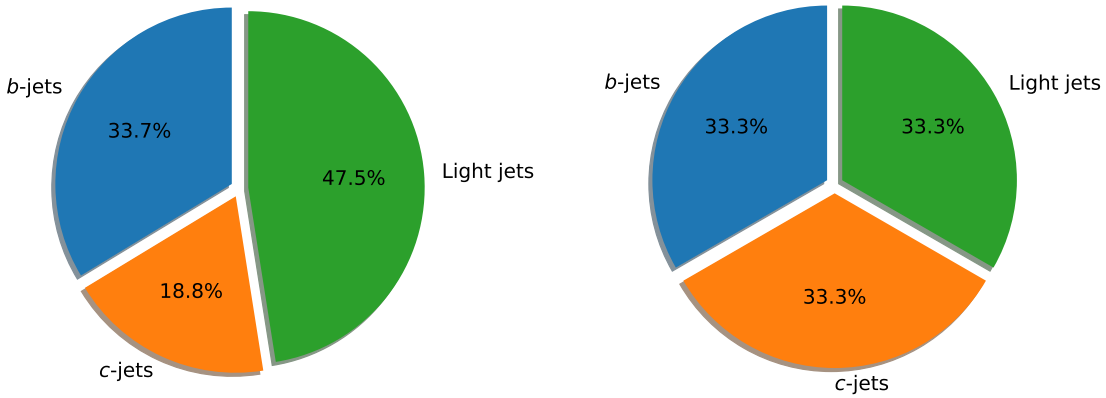
(b): Jet p_T of the resampled training sample.



(c): Jet $|\eta|$ of the un-resampled training sample.

(d): Jet $|\eta|$ of the resampled training sample.

Figure 7.14.: Jet p_T and $|\eta|$ distributions of the un-resampled (a), (c) and resampled (b), (d) merged training sample. The resulting resampled merged training sample has the same amount of jets for each of the jet classes.



(a): Before resampling.

(b): After resampling.

Figure 7.15.: Schematic representation of the jet class composition of the training sample before (a) and after resampling (b) in terms of percentage of total jets.

stages of training, the networks tend to prioritise variables with higher values, potentially overlooking the significance of variables with smaller magnitudes that may contain crucial information for the given problem.

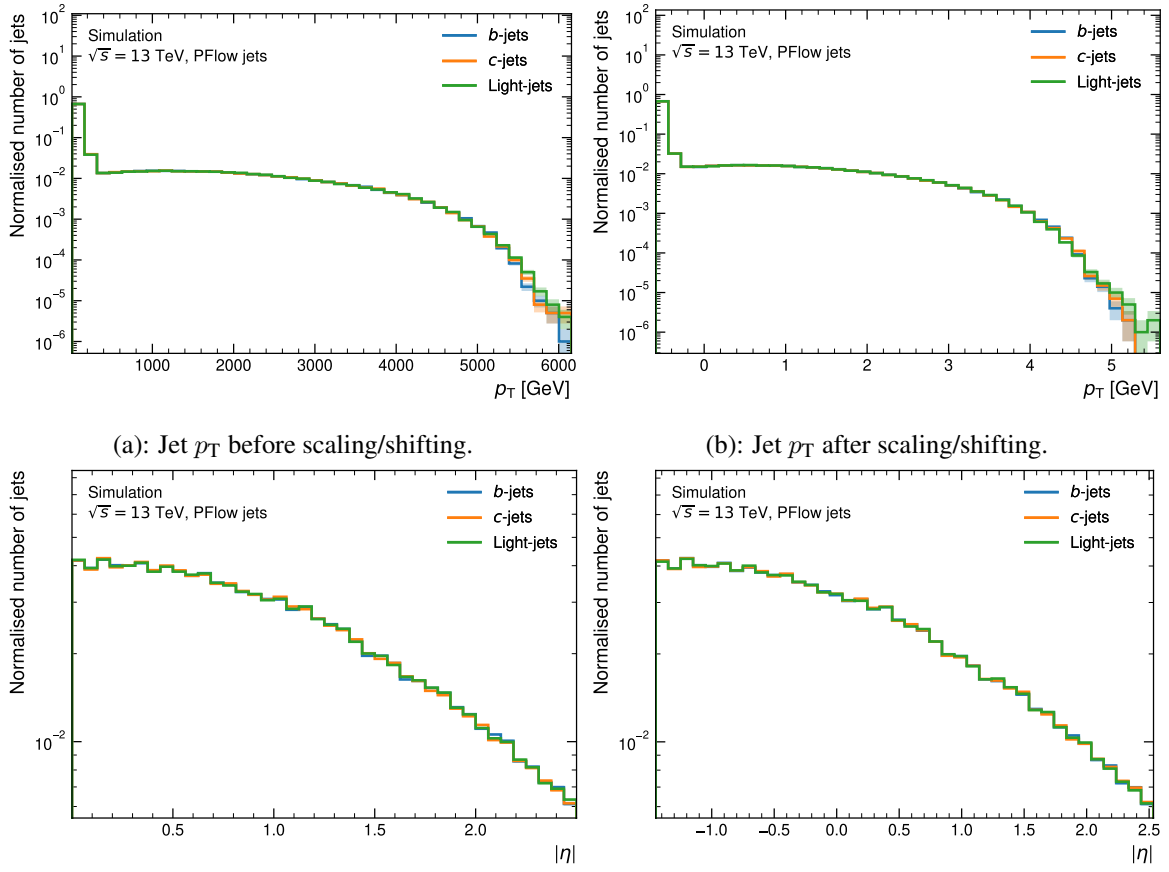
To address this concern, a strategy similar to that employed in Batch Normalisation (refer to Section 6.1.4) is applied: scaling and shifting each input variable to a mean of zero and a standard deviation of one. The mean $\bar{x}$ and standard deviation σ_x for a variable x are determined from the distributions of each variable, calculated using Equations (7.7) and (7.8), where N denotes the number of jets or tracks[15]. The resulting means and standard deviations for each variable are then utilised to scale and shift the variables according to Equation (7.9).

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{7.7}$$

$$\sigma_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{7.8}$$

$$x_i = \frac{x_i - \bar{x}}{\sigma_x} \tag{7.9}$$

To visualise the effect of the scaling and shifting, the jet p_{T} and $|\eta|$ distributions of the merged training sample before and after scaling and shifting are depicted in Figure 7.16.



(a): Jet p_{T} before scaling/shifting.

(b): Jet p_{T} after scaling/shifting.

(c): Jet $|\eta|$ before scaling/shifting.

(d): Jet $|\eta|$ after scaling/shifting.

Figure 7.16.: Jet p_{T} and $|\eta|$ distributions of the merged and resampled training sample before and after applying the scaling/shifting as denoted in Equation (7.9). The scaling is visible for the jet p_{T}, while the shifting is more visible in the jet $|\eta|$ distribution.

15 This depends on the type of variable used. If a jet variable is used, N denotes the number of jets; likewise for tracks.

7.2.4 *Validation and Test Samples*

Validation and test samples are needed to evaluate the performance during and after training the NNs. The validation sample is created similarly to the training sample, undergoing the same routine of kinematic resampling and merging. For the scaling and shifting, the mean and standard deviation values are not explicitly re-derived but simply taken from the training sample due to the approximately same values for both samples. The resulting hybrid validation sample is therefore, in terms of the preprocessing steps applied, identical to the training sample. In addition, the un-resampled non-merged $t\bar{t}$ and Z' validation samples are also taken into account, to validate the performance on non-resampled physics samples.

Two test samples are created, one sample $t\bar{t}$ and Z' each, skipping the kinematic resampling and merging step. While the validation samples are used for direct validation during training, the test samples are used to evaluate the performance of the taggers after training on non-resampled physics process samples. This ensures that the actual performance delivered to the analyses is checked and that no unwanted feature is learned. While no explicit scaling and shifting is derived, to process the test samples through the network and have comparable results, each test sample is scaled and shifted on the fly while loading using the scaling and shifting values derived for the training sample. The same holds for the inference of the taggers inside the ATLAS reconstruction software [180]. When interfacing the taggers using LWTNN [181], the inputs to the taggers are scaled and shifted with the frozen and stored values inside the LWTNN models.

8

THE DEEP IMPACT PARAMETER SETS TAGGER

The Deep Impact Parameter Sets (DIPS) [194] tagger succeeds the RNNIP tagger for Run 3 of the LHC. Similar to RNNIP, DIPS operates on the associated tracks and their variables of the targeted jet, addressing the same task. However, RNNIP relies on a high-performing yet relatively slow sequence-based RNN architecture. Sequential processing, potent for ordered entries such as sentences, proves unfavourable for flavour tagging. This is because products of hadron decays lack intrinsic ordering, diminishing the effectiveness of sequence-based processing.
DIPS, on the contrary, treats the associated tracks of the jet as an unordered, variable-sized set. This permutation invariance allows DIPS to process inputs in parallel while leveraging correlations between different set entries. Moreover, this permutation invariance is more logically aligned with the physics compared to the ordered sequence-based approach of RNNIP.

This chapter delves deeper into the architecture and final training sample of DIPS. The latest and most performant version is presented herein, utilising the samples and preprocessing methods outlined in Section 7.2. The RNNIP tagger is included as reference due to its role as the predecessor of DIPS.

8.1 TRAINING SAMPLE

The training, validation, and test samples for DIPS are derived using the techniques explained in Section 7.2. The classifier employs three output classes: b, c, and light-flavour jets. The final training sample comprises 40 million jets per jet class, $70\,\%$ $t\bar{t}$ and $30\,\%$ Z' jets, yielding a total of 120 million jets for training. The validation sample mirrors a similar fraction of $t\bar{t}$ and Z' jets, totalling three million jets, with one million per jet class. For testing purposes, three million $t\bar{t}$ and Z' jets each are used without any resampling or merging.
Further, 15 track variables serve as input to DIPS, each listed in Table 8.1 along with a brief description, an indication of whether they undergo scaling and shifting, and if their logarithm is considered as input or the plain variable[1]. These variables are further categorised into kinematic, hit, and IP-based variables, as visually represented in the table by their separation into distinct blocks. Moreover, Figure A.1 and Figure A.2 illustrate the distributions of these variables for $t\bar{t}$ and Z' jets, respectively. DIPS uses the input variables of RNNIP and two additional variables: the direct values of the transverse and longitudinal IPs. Moreover, the track selection criteria and the maximum number of tracks per jet processed are changed, due to the substantial improvements in computing

1 The logarithm of certain variables is taken due to their vast range of possible values.

Table 8.1.: Input variables to the DIPS b-tagging algorithms.

Category	Variable	Description	S&S[2]	Log
Kinematics	$p_{\mathrm{T}}^{\mathrm{frac}}$	Fraction of transverse momentum of the track with respect to the jet [Equation (7.3)]	✓	✓
	$\Delta R\,(\text{track, jet})$	Angular distance between the track and the jet axis [Equation (7.4)]	✓	✓
Hits	$N_{\text{Pixel hits}}$	Number of PD hits	✓	
	$N_{\text{SCT hits}}$	Number of SCT hits	✓	
	$N_{\text{Innermost pixel hits}}$	Number of innermost pixel layer hits		
	$N_{\text{Next to innermost pixel hits}}$	Number of next to innermost pixel layer hits		
	$N_{\text{Innermost pixel shared hits}}$	Number of innermost pixel layer shared hits		
	$N_{\text{Innermost pixel split hits}}$	Number of innermost pixel layer split hits		
	$N_{\text{Pixel shared hits}}$	Number of PD shared hits		
	$N_{\text{Pixel split hits}}$	Number of PD split hits		
	$N_{\text{SCT shared hits}}$	Number of SCT shared hits		
IP-based	d_0	Transversal IP	✓	
	$z_0 \sin(\theta)$	Longitudinal IP times the sine of θ	✓	
	S_{d_0}	Lifetime signed d_0 significance [Equation (7.1)]		
	S_{z_0}	Lifetime signed z_0 significance [Equation (7.2)]		

efficiency, training, and inference time. This updated track selection, compared with the RNNIP track selection in Table 7.1, is outlined in Table 8.2.

The differences in track selection criteria between DIPS and RNNIP, along with the maximum number of tracks per jet, are visible in the track p_{T} and IPs thresholds. While RNNIP adheres to a relatively tight selection, DIPS employs a significantly looser track selection strategy. These relaxed criteria enable DIPS to incorporate a greater number of associated tracks per jet, enabling a more comprehensive characterisation of the jet class compared to RNNIP. This distinction is particularly crucial when analysing jets from Z' events. To illustrate this, the average number of tracks per jet class and sample is depicted in Figure 8.1.

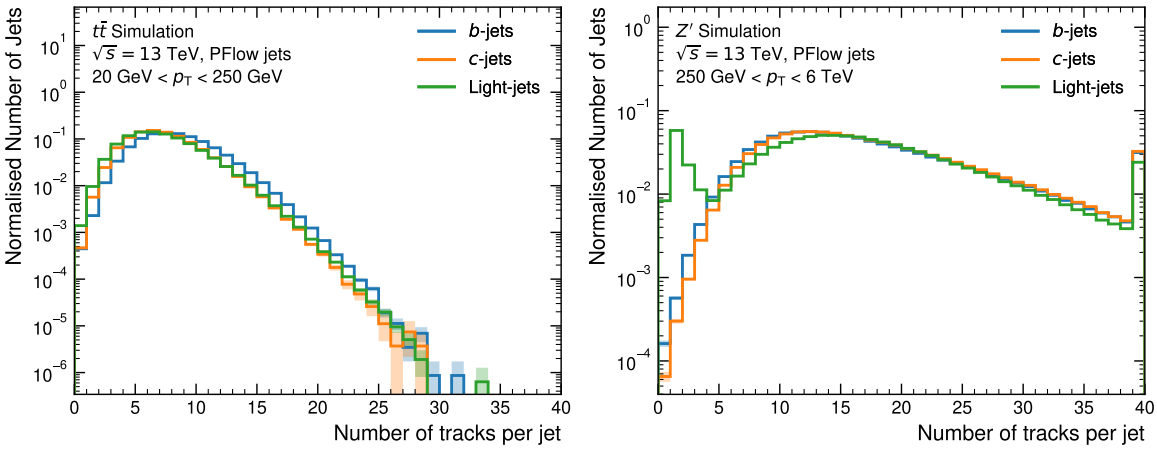
As shown in Figure 8.1 (a), the increased maximum number of tracks per jet used in the training of DIPS in comparison to RNNIP allows the network to take all tracks from $t\bar{t}$ jets into account. Conversely, Figure 8.1 (b) illustrates a notably higher number of tracks for Z' jets, particularly attributable to their higher p_{T} regime.

Opposite to that, one specific feature of the Z' jets is a large amount of light-flavour jets with less than five tracks, with a sharp peak at one track per jet. These jets were identified as electrons, which

2 Scaled and shifted

Table 8.2.: Track selection cuts for the RNNIP and DIPS tagger. Taken and adapted from Ref. [2].

Variable	Description	RNNIP	DIPS
Max N_{Tracks}	Maximum number of tracks used	25	40
$p_{\text{T}}^{\text{track}}$	Track p_{T}	$> 1\,\text{GeV}$	$> 0.5\,\text{GeV}$
$\lvert d_0 \rvert$	Transversal IP	$< 1\,\text{mm}$	$< 3.5\,\text{mm}$
$\lvert z_0 \sin(\theta) \rvert$	Longitudinal IP	$< 1.5\,\text{mm}$	$< 5\,\text{mm}$
$\lvert \eta^{\text{track}} \rvert$	Track $\lvert \eta \rvert$	< 2.5	
$N_{\text{Silicon hits}}$	Number of hits in the pixel and SCT layers	≥ 7	
$N_{\text{Silicon shared hits}}$	Number of space points shared by multiple tracks in pixel and SCT layers	< 2	
$N_{\text{Pixel holes}}$	Number of missing hits in the pixel layers	< 2	
$N_{\text{Silicon holes}}$	Number of missing hits in the SCT layers	< 3	

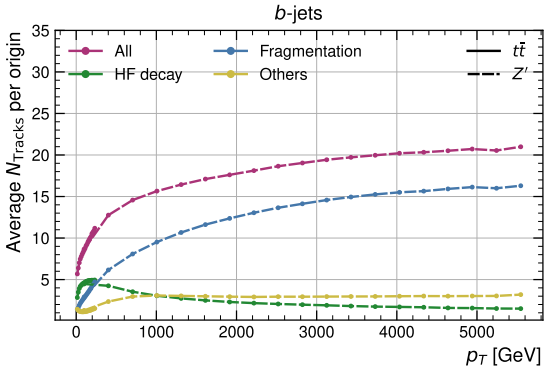


(a): $t\bar{t}$ (b): Z'

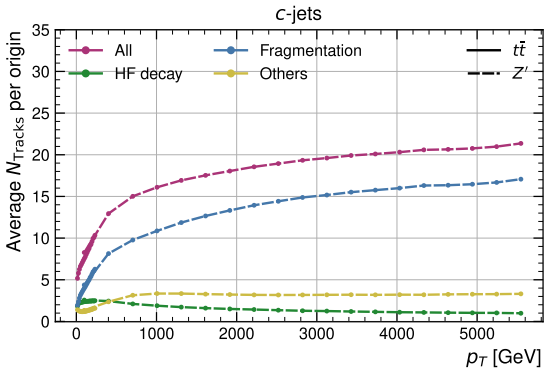
Figure 8.1.: Number of tracks per jet for both $t\bar{t}$ (a) and Z' (b) jets. For the Z' distribution, an overflow bin is used in the last bin, indicating that more than 40 tracks are associated to some jets.

are faking a light-flavour jet, and are classified in the labelling as light-flavour jets. This is due to the inner workings of the labelling algorithm, making these electron-initiated light-flavour jets rather unique and easy for the algorithms to identify due to their low number of associated tracks in the jet. To delve deeper, Figure 8.2 showcases the average number of tracks per jet across all three jet classes for both $t\bar{t}$ and Z' jets, plotted against jet p_{T}.

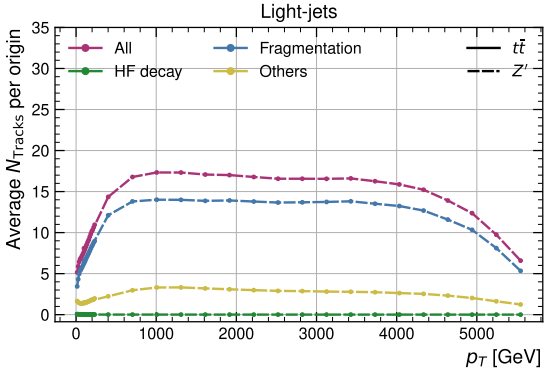
Despite Z' jets averaging around 10 more tracks per jet compared to $t\bar{t}$ jets, most of these tracks do not originate from heavy-flavour decays but rather from other sources, such as fragmentation or secondary interactions. While these additional tracks are not particularly beneficial for the actual training, they still need consideration to ensure the inclusion of all tracks stemming from b or c-hadron decays within the selected 40 tracks per jet for processing. While the maximum number of tracks of 40 is sufficient for $t\bar{t}$ jets, for Z' jets, which can have more than 40 tracks associated to a jet, this is insufficient. To determine which tracks are taken into account for processing, although no explicit ordering of the tracks is needed for DIPS, the same sorting as for RNNIP is applied: sorting in decreasing order after S_{d_0} and taking the 40 tracks with the highest S_{d_0}.



(a): b-jets



(b): c-jets



(c): Light-flavour jets

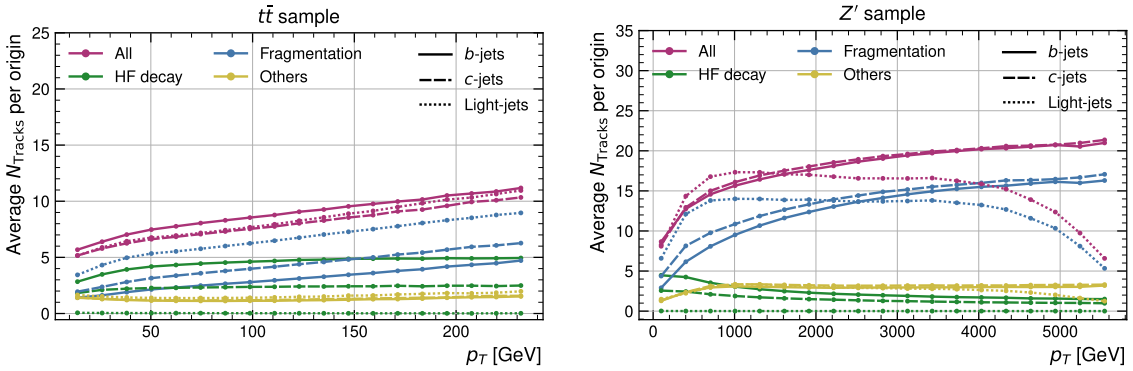
Figure 8.2.: Number of tracks per jet as a function of jet p_{T} for b (a), c- (b) and light-flavour (c) jets for both $t\bar{t}$ and Z' jets. The chosen binning gives 20 bins each for $t\bar{t}$ between $20 - 250\,\text{GeV}$ and Z' between $250 - 6000\,\text{GeV}$. The statistical uncertainty of the bins is indicated with a coloured but is too small to be seen.

To give further insight into the composition of the origin of the tracks of $t\bar{t}$ and Z' jets, the average number of tracks per origin per jet class are listed in Table 8.3 and the number of tracks per jet as a function of jet p_{T} separated by sample are shown in Figure 8.3.

Z' jets generally contain fewer tracks originating from heavy-flavour decays and notably more tracks from fragmentation. This trend is explicable due to the higher jet p_{T}. Also, the number of tracks stemming from heavy-flavour decays remains consistent for $t\bar{t}$ b and c-jets across the entire p_{T} range, while it decreases for Z' jets due to being more collimated at high-p_{T} and harder to reconstructed.

Table 8.3.: Average number of tracks per jet class for both $t\bar{t}$ and Z' jets, split into the track origins. Here, Fragmentation (Frag) denotes tracks from the primary vertex and fakes, where the latter stem from hits of multiple particles. Heavy-flavour decay (HF) denotes tracks that originate from the decay of either a b or a c-hadron. The Others category is comprised of all other track origins, like pile-up, τ decays, or other secondary interactions. The statistical uncertainties for these values are negligible and therefore not shown.

Sample	Jet Flavour	$\bar{N}^{\text{All}}_{\text{Tracks}}$	$\bar{N}^{\text{HF}}_{\text{Tracks}}$	$\bar{N}^{\text{Frag}}_{\text{Tracks}}$	$\bar{N}^{\text{Others}}_{\text{Tracks}}$
$t\bar{t}$	b	7.97	4.29	2.49	1.19
	c	6.93	2.27	3.42	1.25
	Light-flavour	6.78	0.04	5.26	1.49
Z'	b	16.07	2.82	10.54	2.71
	c	15.79	1.79	11.20	2.80
	Light-flavour	10.77	0.02	8.83	1.93



(a): $t\bar{t}$ (b): Z'

Figure 8.3.: Number of tracks per jet per origin as a function of jet p_{T} for $t\bar{t}$ (a) and Z' (b) jets. The statistical uncertainty of the bins is indicated with a coloured band but is too small to be seen.

8.2 ARCHITECTURE

DIPS is a DNN based on the Deep Sets architecture, introduced in Section 6.3. It comprises two distinct sub-networks, ϕ and ρ, intentionally designed with slight differences. A schematic depiction of the architecture is shown in Figure 8.4.

The ϕ network consists of a fully-connected feedforward MLP with shared weights comprising one input and four hidden layers. It takes m variables or features per track as input for all n tracks of a jet. Though multiple ϕ networks are depicted in Figure 8.4, it is one network used in multiple instances with shared weights, working simultaneously on all input tracks n at the same time. The four hidden layers have varying numbers of neurons, and each hidden layer neuron uses the ReLU activation function. Batch Normalisation layers are employed between the hidden layers to ensure rapid and stable training convergence of the network.

Once all tracks are processed with the ϕ network, the outputs from the last hidden layer are concatenated and aggregated using summation as the aggregation function. Subsequently, the aggregated representation is fed into the ρ network, also a fully-connected feedforward MLP, comprising six

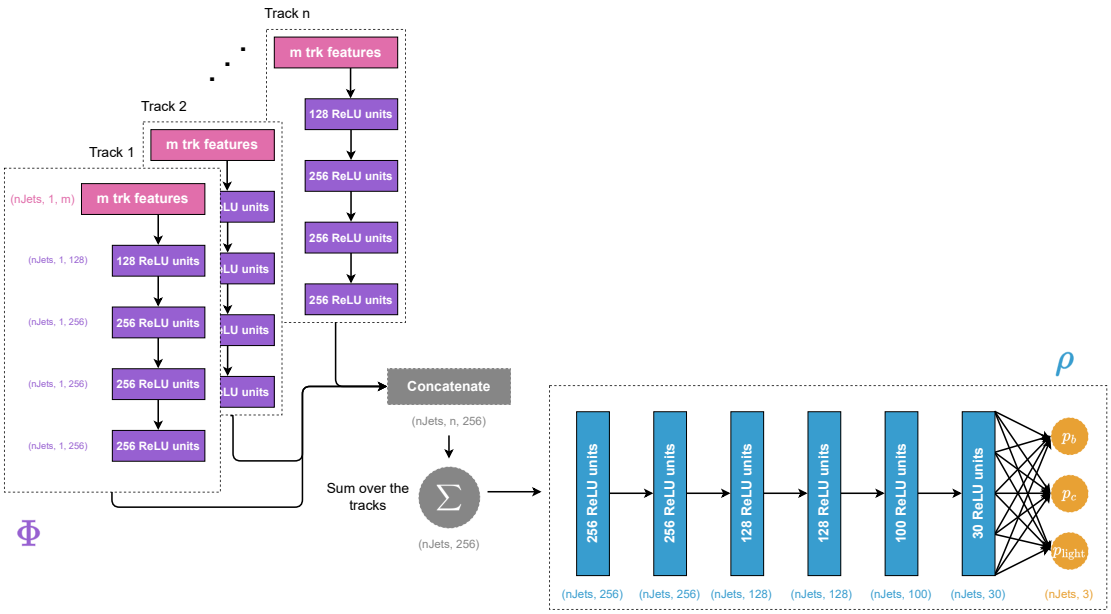


Figure 8.4.: Schematic representation of the latest and best performing DIPS architecture.

hidden and one output layer. The hidden layers have a varying number of neurons, each using the ReLU activation function. The final output layer, with three neurons, uses the softmax function as the activation function. The resulting outputs, interpretable as probabilities due to the usage of softmax, contribute to building the b-tagging discriminant D_b. Similar to RNNIP and DL1r, the b-tagging discriminant is defined using Equation (7.6), where f_c and f_{light} denote the fraction values, and p_b, p_c and p_{light} represent the output probabilities of DIPS indicating whether the evaluated jet originates from a b, c or light-flavour quark. DIPS was trained for 200 epochs using TensorFlow with the Keras front-end. For optimisation, the Adam optimiser was employed, and the trainable parameters were initialised using a Glorot Uniform distribution. The final trained model was integrated for inference within the ATLAS reconstruction software using LWTNN. All hyperparameters for DIPS are detailed in Table 8.4. A minor hyperparameter optimisation was conducted, varying the number of hidden layers and neurons per layer, revealing no substantial gains by further expanding the architecture of DIPS.

Although DIPS was trained for 200 epochs, the model state was saved after every epoch for a more detailed examination of the training progress. No early stopping or other regularisation techniques were employed to mitigate potential overfitting of the model. A comprehensive explanation of the training sample and the corresponding input variables used in training will be provided in Section 8.1.

8.3 RESULTS

The training of the DIPS tagger was done using the samples explained in Section 8.1. The tagger was trained for 200 epochs in total using one NVIDIA V100 GPU. The resulting accuracy and loss curves are depicted in Figure 8.5 and Figure 8.6 respectively.

Table 8.4.: Hyperparameters of the DIPS tagger.

Hyperparameter	Value
Number of input variables	15
ϕ $N_{\text{Hidden layers}}$	4
ϕ $N_{\text{Neurons/layer}}$	[128, 256, 256, 256]
ρ $N_{\text{Hidden layers}}$	6
ρ $N_{\text{Neurons/layer}}$	[256, 256, 128, 128, 100, 30]
Number of output variables	3
Loss function	Categorical Crossentropy
Optimiser	ADAM
LR	0.001
Batch size	15000
Aggregation function	Summation
Activation function (hidden neurons)	ReLU
Activation function (output neurons)	softmax
Number of epochs trained	200
Free (trainable) parameters	367259
Fixed parameters	3588
Total parameters	370847

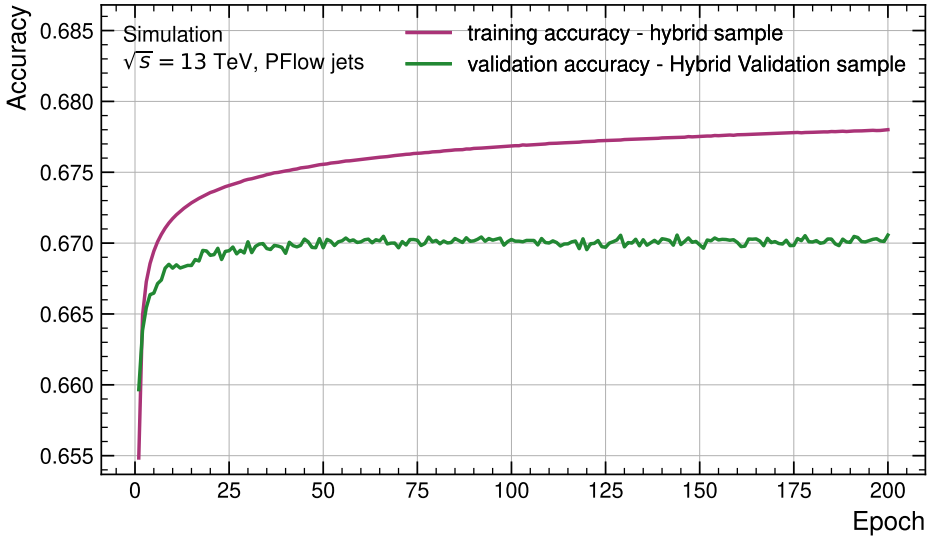


Figure 8.5.: Training and validation accuracy as a function of the training epoch for DIPS. Both accuracies were derived using the respective resampled hybrid samples. While the training accuracy was derived as an average over all batches for one epoch, the validation accuracy was derived after training, using all batches at once for calculation.

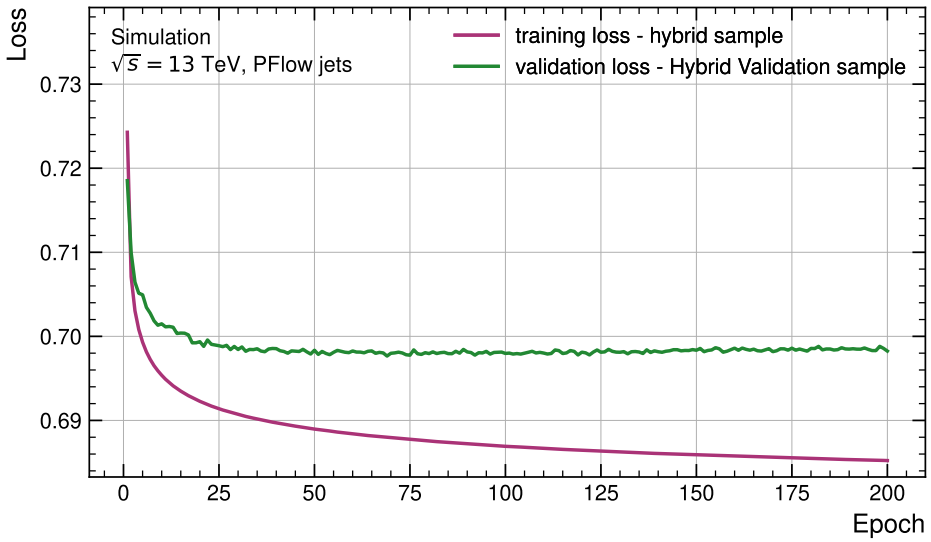


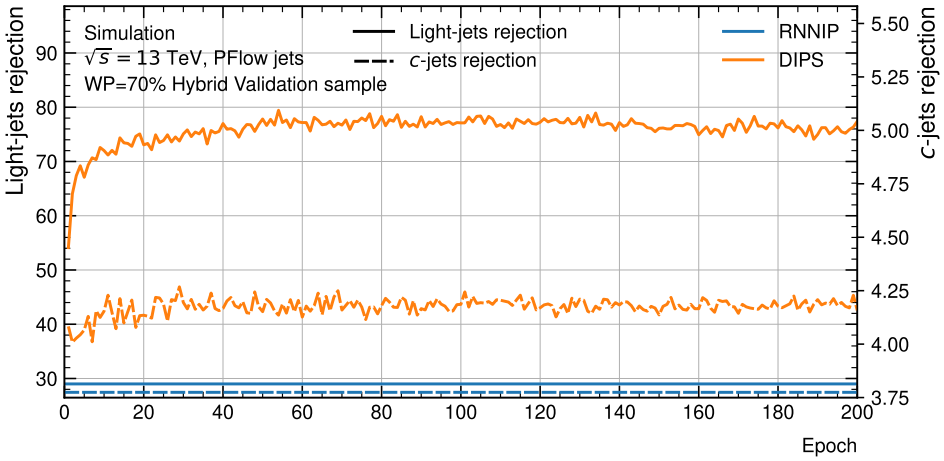
Figure 8.6.: Training and validation loss as a function of the training epoch for DIPS. Both losses were derived using the respective resampled hybrid samples. While the training loss was derived as an average over all batches per epoch, the validation accuracy was derived after training, using all batches at once for calculation.

Both plots exhibit a noticeable divergence between the validation and training curves starting from the fifth epoch. On initial observation, this might suggest overfitting, given the considerable gap between these curves. However, this divergence alone is not a conclusive sign. The decisive factor in this context would be a decrease or increase in validation accuracy or loss. What one witnesses here is not an overall performance drop but rather a plateauing of the performance increase. Therefore, the constant validation accuracy and loss rather suggest that the network has reached its limit to learning the general features of the different jet classes, indicating the achievement of a (local) minimum in the optimisation.

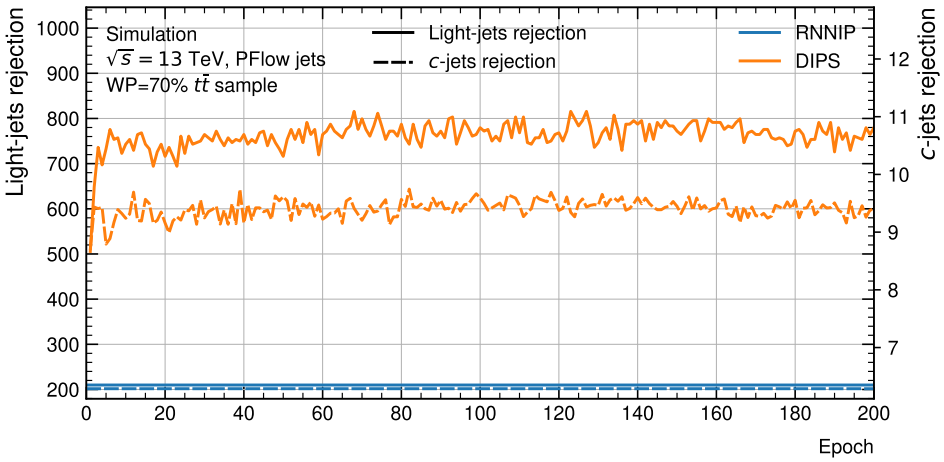
Figure 8.7 illustrates the resulting rejections for the hybrid, $t\bar{t}$, and Z' validation samples using the $70\,\%$ WP. Similar to the validation accuracy and loss curves, the rejections ascend initially over the first epochs before stabilising and fluctuating around a stable value. For the Z' validation sample, rejections peak after around 20 epochs, while the corresponding hybrid and $t\bar{t}$ validation samples peak after roughly 60 to 80 epochs. The model state of epoch 150 is used for further evaluation with the testing samples

For the evaluation process, this selected model state is used to compute the output probabilities for all jets within the test samples. The resulting output probability distributions for both the $t\bar{t}$ and Z' jets are visualised in Figure 8.8.

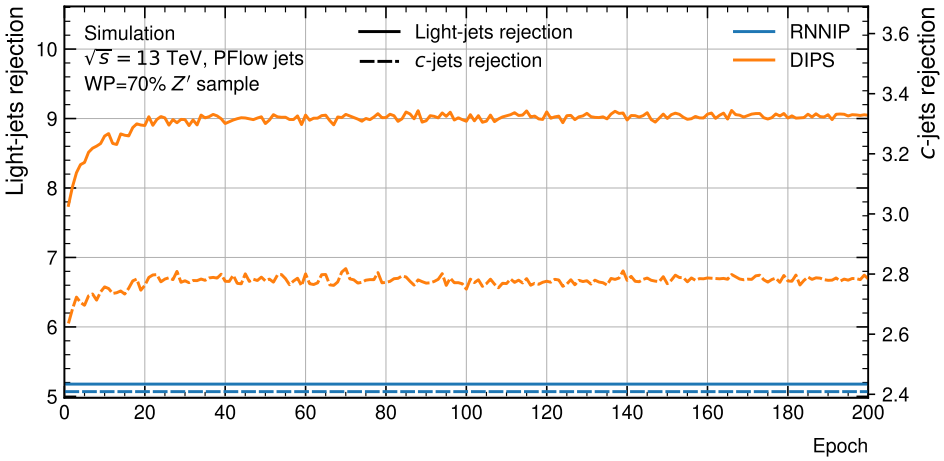
For the $t\bar{t}$, the different probabilities already exhibit a clear separation between the b-jets distribution and the c and light-flavour jets distributions. Additionally, for each output probability, the respective truth jet class achieves the highest scores, indicating a good separation among the three classes. This effect is most pronounced for b and light-flavour jets, as they are relatively easier to distinguish from each other. The c-jets, which share similar properties with b-jets and sometimes with light-flavour jets, typically fall between them and exhibit slightly lower peaks in their respective output probability.



(a): Resampled hybrid validation sample

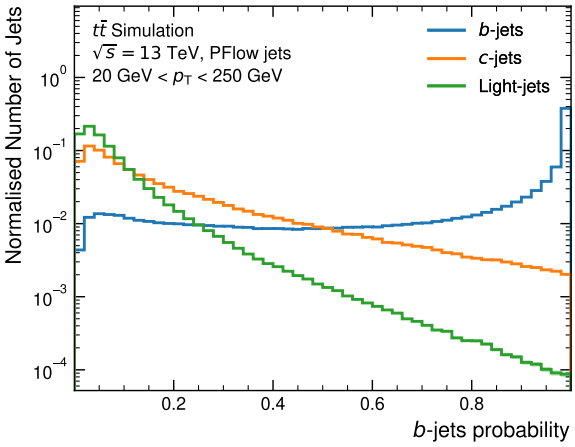


(b): Un-resampled $t\bar{t}$ validation sample

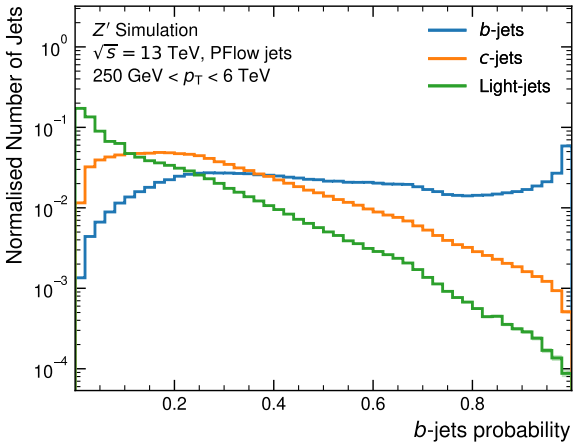


(c): Un-resampled Z' validation sample

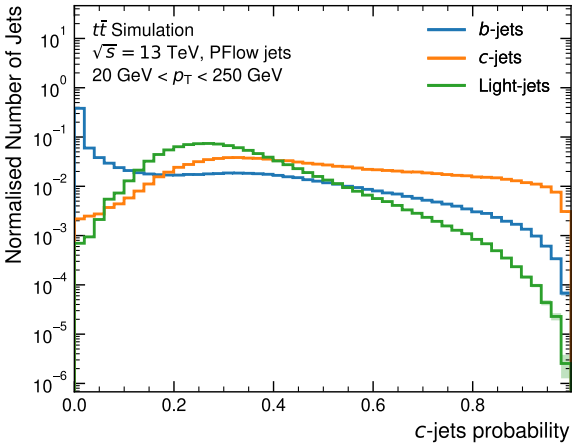
Figure 8.7.: c and light-flavour jets rejections for DIPS for the hybrid (a), $t\bar{t}$ (b) and Z' (c) validation samples as a function of training epoch. As reference, the c and light-flavour jets rejections for RNNIP are shown in blue.



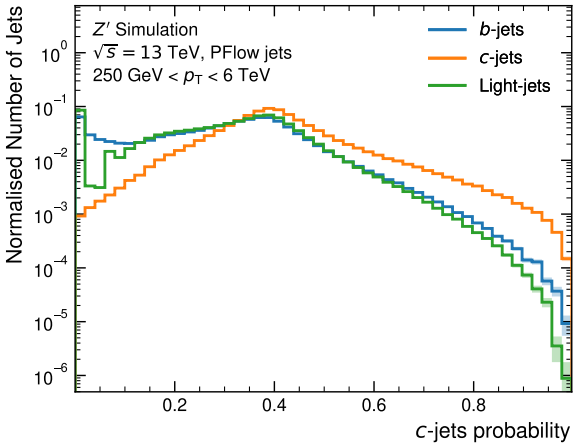
(a): b-jets probability output p_b for $t\bar{t}$



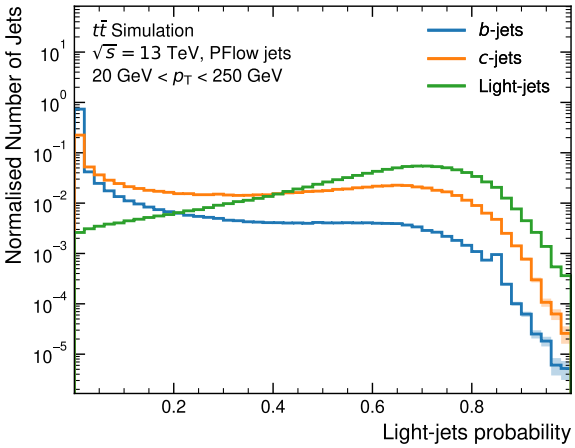
(b): b-jets probability output p_b for Z'



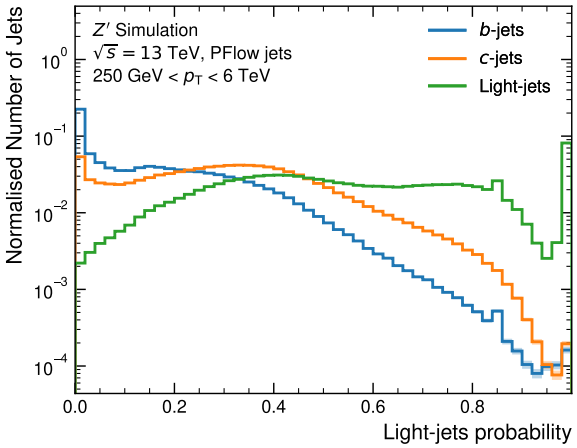
(c): c-jets probability output p_c for $t\bar{t}$



(d): c-jets probability output p_c for Z'

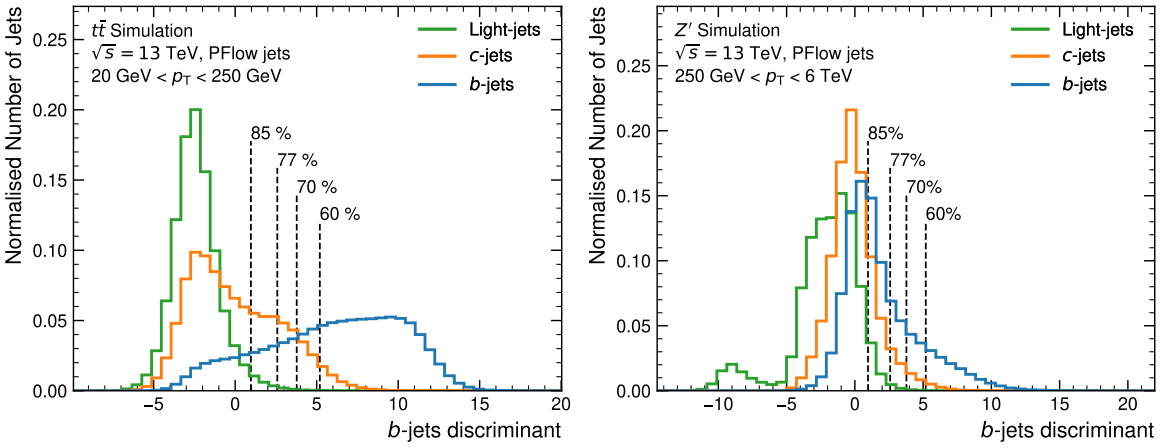


(e): Light-flavour jets probability output p_{light} for $t\bar{t}$



(f): Light-flavour jets probability output p_{light} for Z'

Figure 8.8.: Output probability distributions of DIPS for the $t\bar{t}$ and Z' test samples.



(a): $t\bar{t}$ jets (b): Z' jets

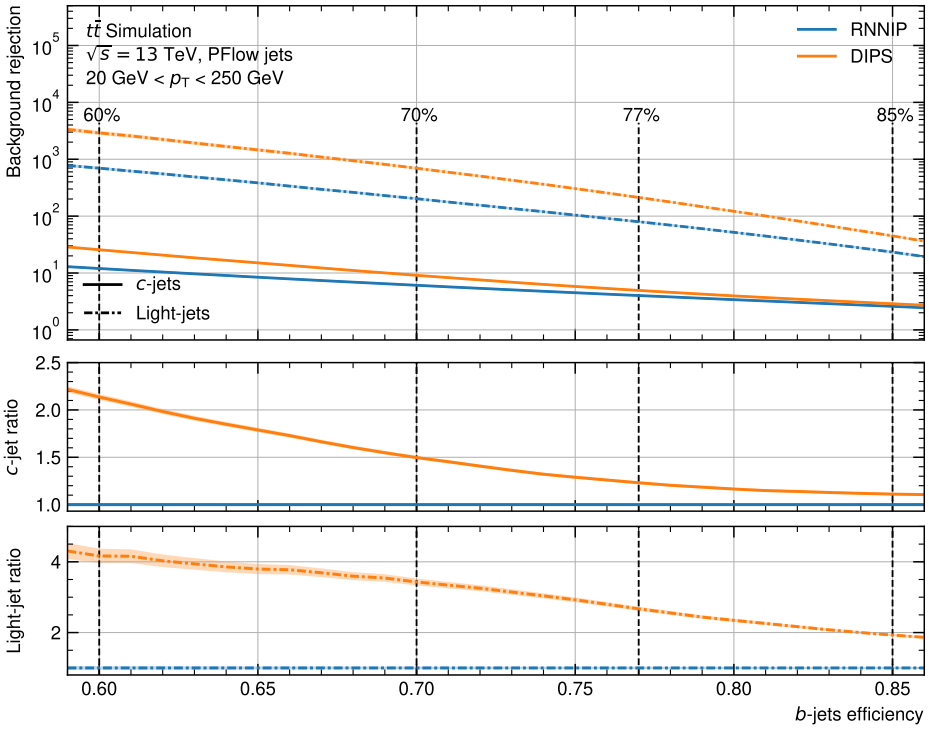
Figure 8.9.: b-tagging discriminant distributions of the DIPS tagger for the $t\bar{t}$ and Z' test samples. The WPs shown in both plots were derived on the $t\bar{t}$ test sample.

Distinctive characteristics of the Z' light-flavour jets, previously noted in Figure 8.1 (b), are the significant peaks in Figures 8.8 (d) and 8.8 (f). These sharp peaks stem from jets with fewer than five tracks per jet, primarily originating from electron-initiated fake jets, making them easily identifiable by DIPS. The impact of these jets on the training was studied and found to be negligible.

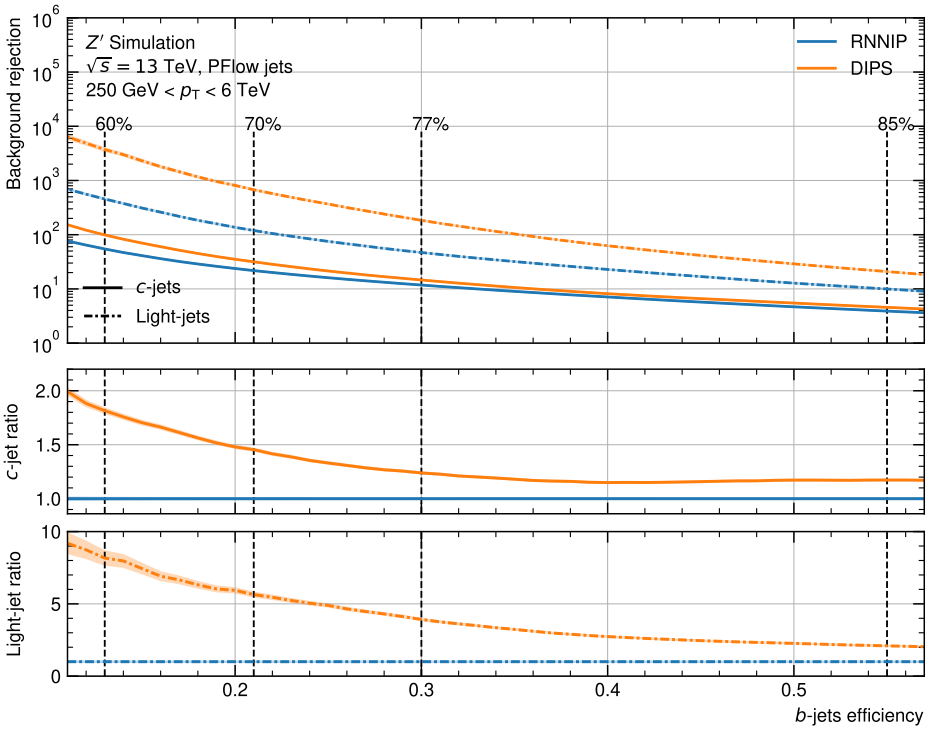
The probability values for the jets are used to construct the b-tagging discriminant, as defined in Equation (7.6) with $f_c = 0.005$ and $f_{\text{light}} = 0.995$. The resultant b-tagging discriminant distributions for both $t\bar{t}$ and Z' jets are depicted in Figure 8.9.

For the $t\bar{t}$ jet distribution, a clear separation is observable between the target class, b-jets, and the background classes, c and light-flavour jets. It is important to note that for both $t\bar{t}$ and Z' jets, the depicted WPs were derived solely using the $t\bar{t}$ jet distribution illustrated in Figure 8.9 (a). This choice is motivated by the primary focus of the algorithms on tagging $t\bar{t}$ jets, with Z' jets exclusively employed during training to boost the statistics in the high-p_T regime. Furthermore, the physics programme of ATLAS is more focused on $t\bar{t}$ jets and Z' jets are an artificial auxillary support, only for training and to fix the low statistics for $t\bar{t}$ jets in the high-p_T region. Generally, WPs used for the analyses are obtained using the same method. To maintain consistency across WPs, this approach is employed to calculate all b-tagging WPs in this chapter, including those for Z' jets.
Regarding the Z' jet distribution, the separation among the three classes is not as distinct as observed for the $t\bar{t}$ jets. This discrepancy is attributed to the increased number of tracks per jet that do not originate from heavy-flavour decays, as previously discussed in Section 8.1. Similar to observations in the output probability distributions, a notable distinct peak for Z' light-flavour jets is also visible in Figure 8.9 (b), identifiable as originating from jets with fewer than two tracks.

For comparison purposes, different types of plots are utilised to compare the new DIPS tagger with the well-established RNNIP tagger. One of the primary tools for performance comparison involves the ROC curves, which are depicted in Figures 8.10 (a) and 8.10 (b) for $t\bar{t}$ and Z' jets, respectively.



(a): $t\bar{t}$ test sample



(b): Z' test sample

Figure 8.10.: ROC curves for the DIPS and RNNIP tagger. The depicted efficiencies and rejections for both taggers were calculated using the $t\bar{t}$ and Z' test samples, each with in total three million jets. The shown b-jets efficiencies on the x-axis denote the b-jets efficiency on the respective test sample, while the depicted WPs were derived from the $t\bar{t}$ test sample and correspond to the WPs which would be used in analyses. The shaded areas denote only the statistical uncertainties, derived by using binomial uncertainties.

For the $t\bar{t}$ jets, a substantial increase in rejection for DIPS over RNNIP is noticeable for both c and light-flavour jets. At the $70\,\%$ WP, DIPS outperforms RNNIP by a factor of 1.5 in c-jets rejection and by a factor of 3.5 in light-flavour jets rejection. This notable performance improvement becomes more pronounced with stricter WPs, increasing to factors of 2.25 for c-jets rejection and 4.05 in light-flavour jets rejection at the $60\,\%$ WP.
In addition to the considerable performance gains for $t\bar{t}$ jets, even more substantial performance improvements are evident for the Z' jets. When applying the WPs and their respective cut values derived from the $t\bar{t}$ jet test sample, the performance gains for Z' jets surpass those observed for $t\bar{t}$ jets. At the $70\,\%$ WP, DIPS outperforms RNNIP by a factor of 1.45 in c-jets rejection and by a factor of 6 in light-flavour jets rejection. Similar to $t\bar{t}$ jets, for stricter WPs, the performance continues to increase, reaching factors of 1.8 for c-jets rejection and 8 in light-flavour jet rejection at the $60\,\%$ WP.

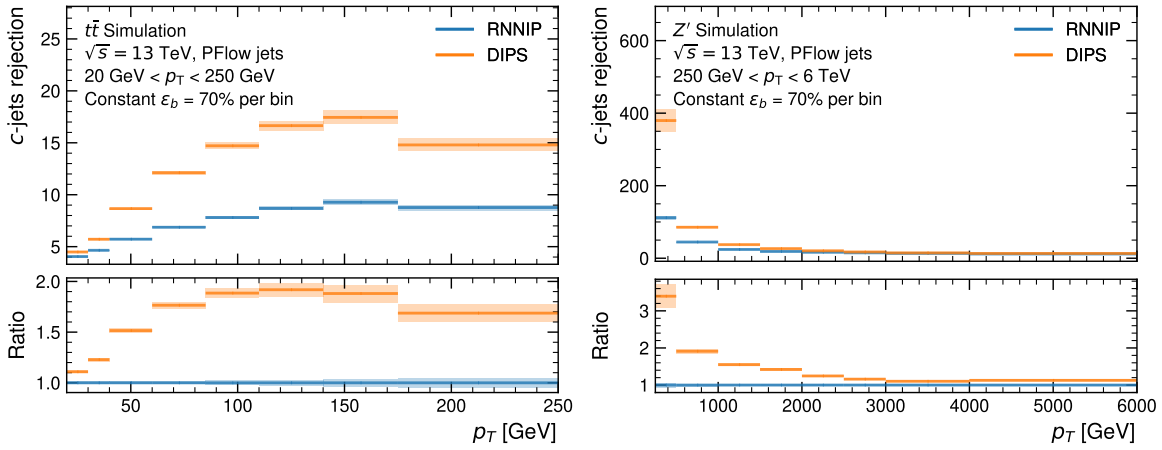
To further assess the performance of DIPS, the c and light-flavour jet rejections as a function of jet p_{T} for both the $t\bar{t}$ and Z' samples at the $70\,\%$ WP are illustrated in Figure 8.11. This representation utilises the flat efficiency per bin procedure[3].

For both $t\bar{t}$ and Z' c-jets rejections, a distinct enhancement of DIPS over RNNIP is noticeable across the entire jet p_{T} range of the test samples. The most significant performance gain in $t\bar{t}$ is observed in the $110 - 140\,\mathrm{GeV}$ bin, with an improvement of a factor of 1.9, while for Z', this enhancement is seen in the first bin, ranging from $250 - 500\,\mathrm{GeV}$, showing an improvement factor of 2.6.
Moreover, alongside the substantial performance improvements in c-jets rejection, the light-flavour jets rejection displays a similar trend. In both test samples, DIPS clearly outperforms RNNIP, with the most significant improvement in $t\bar{t}$ observed in the $140 - 175\,\mathrm{GeV}$ jet p_{T} bin, increasing by a factor of 6. In the case of Z', the most notable improvement occurs once again in the initial bin from $250 - 500\,\mathrm{GeV}$, demonstrating an improvement factor of 8.

Furthermore, the c and light-flavour jets rejections as a function of jet p_{T} can also be plotted inclusively, where the WP and its respective cut value on the b-tagging discriminant are calculating using the full $t\bar{t}$ test sample. This is depicted for both test samples in Figure 8.12.

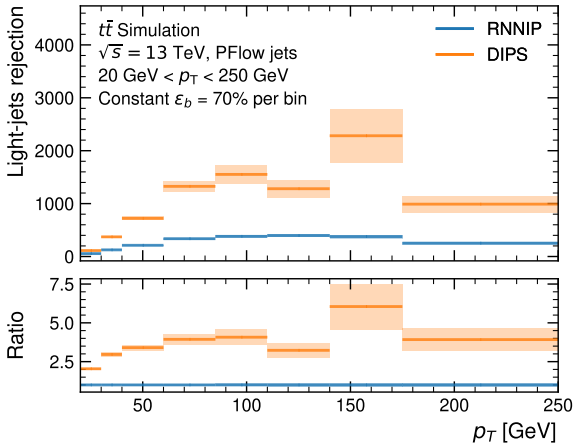
In addition to rejections, in the inclusive case, the b-jet efficiency per bin can also be plotted as it is calculated over the entire jet p_{T} range, which is important for the final performance of the taggers in analyses. By definition, if all bins were summed, the same performance would be observed for both DIPS and RNNIP in terms of b-jet efficiency. Therefore, the only noticeable difference in these plots would be the shifting of performance between jet p_{T} bins.
For $t\bar{t}$, DIPS and RNNIP demonstrate similar performances in nearly all bins, except for the initial bin from $20 - 30\,\mathrm{GeV}$, where a $2\,\%$ drop for DIPS is observed. In the case of Z', a shift in performance from the higher p_{T} regime to the lower p_{T} regime is noticeable, which is expected. This shift is due to the main training samples for DIPS being $t\bar{t}$ and low-p_{T} Z' jets (see Figure 7.14 (b)), and the decreasing proportion of tracks from heavy-flavour decays concerning the total number of tracks

3 As detailed in Section 7.1.2, this procedure calculates the cut values for the WP not inclusively, but per jet p_{T} bin. Thus, within each bin, a ϵ_b of $70\,\%$ is ensured.

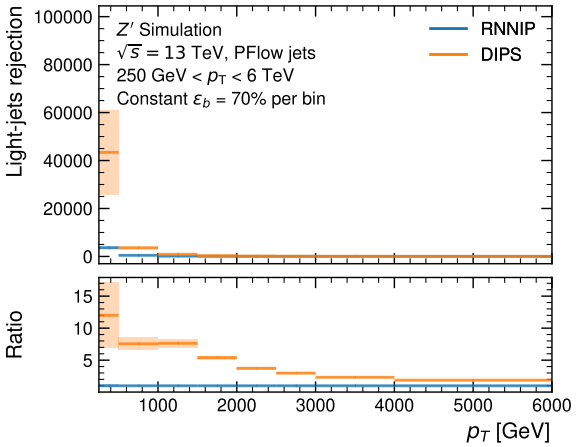


(a): $t\bar{t}$ c-jets rejection

(b): Z' c-jets rejection



(c): $t\bar{t}$ light-flavour jets rejection



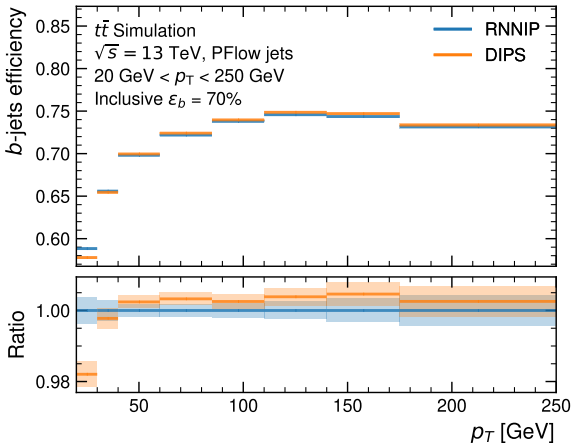
(d): Z' light-flavour jets rejection

Figure 8.11.: c and light-flavour jets rejections for the $t\bar{t}$ and Z' test samples for the 70 % WP using the flat efficiency per bin procedure for the DIPS and RNNIP taggers. For both test samples, $t\bar{t}$ and Z', the same WP was used with the same cut value on the b-tagging discriminant, which was derived on the $t\bar{t}$ test sample.

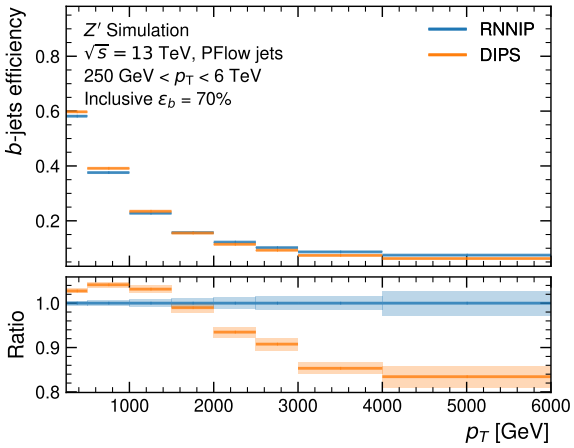
in the jet as p_{T} increases. This rise in the number of tracks, carrying no information about heavy-flavour decays, poses slight challenges for DIPS due to its looser track selection and consideration of an increased maximum number of tracks per jet.

Regarding rejections, the outcomes closely resemble those of the flat efficiency per bin procedure. For both $t\bar{t}$ and Z', DIPS notably outperforms RNNIP in both c and light-flavour jets rejection, showcasing impressive improvement factors of up to 3.8 in $t\bar{t}$ and 8 in Z', even though the trend of improvements is inverted for Z'. While the largest enhancements for the flat efficiency per bin procedure were observed in the lower p_{T} regime, in the inclusive case, the performance gains increase from low to high-p_{T}. This particular behaviour can be explained by the shift in performance in the b-jets efficiency. A smaller b-jets efficiency leads to higher rejection rates due to the reduced number of jets passing the inclusively calculated WP.

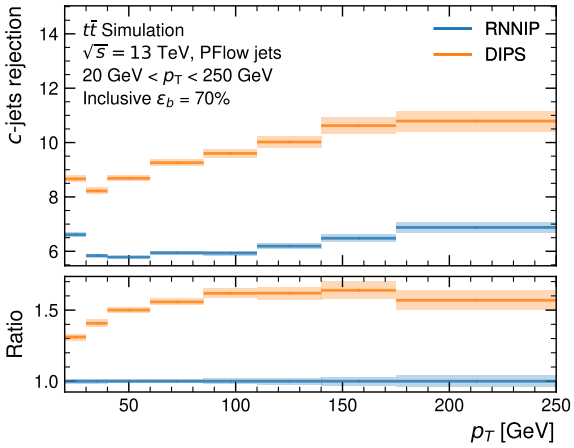
Similar to RNNIP and DL1r, an optimisation of the fraction values for Equation (7.6) is required. To achieve this, the c and light-flavour jets rejections as a function of different fraction value combinations are calculated for a given WP. Due to its significance, the optimisation is performed using the



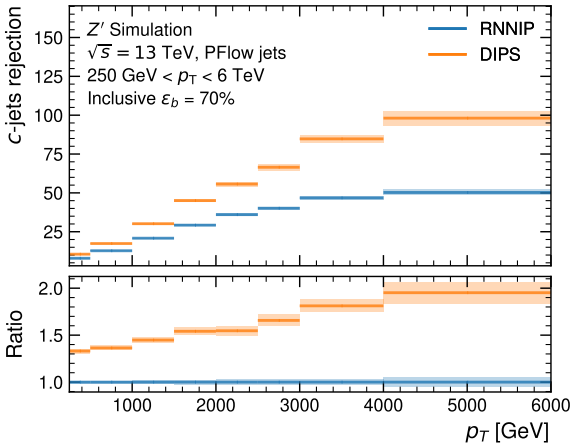
(a): $t\bar{t}$ b-jets efficiency



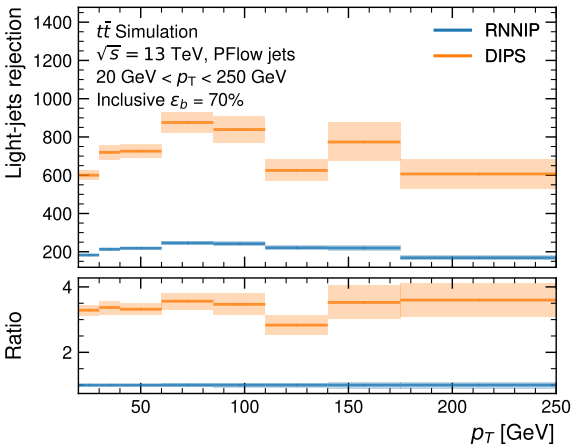
(b): Z' b-jets efficiency



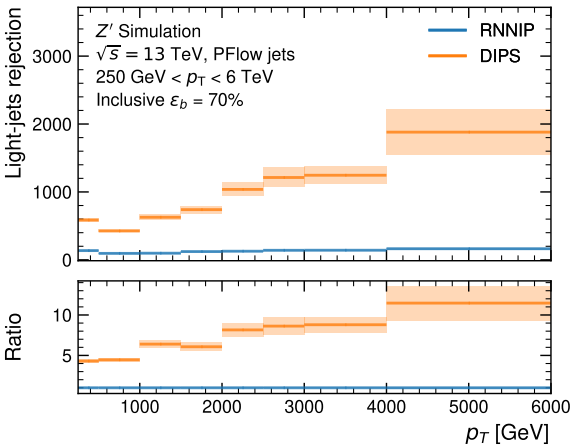
(c): $t\bar{t}$ c-jets rejection



(d): Z' c-jets rejection



(e): $t\bar{t}$ light-flavour jets rejection



(f): Z' light-flavour jets rejection

Figure 8.12.: b-jets efficiency and c and light-flavour jets rejections for the $t\bar{t}$ and Z' test samples for the inclusive 70 % WP for the DIPS and RNNIP taggers. For both test samples, $t\bar{t}$ and Z', the same WP was used with the same cut value on the b-tagging discriminant, which was derived on the $t\bar{t}$ test sample.

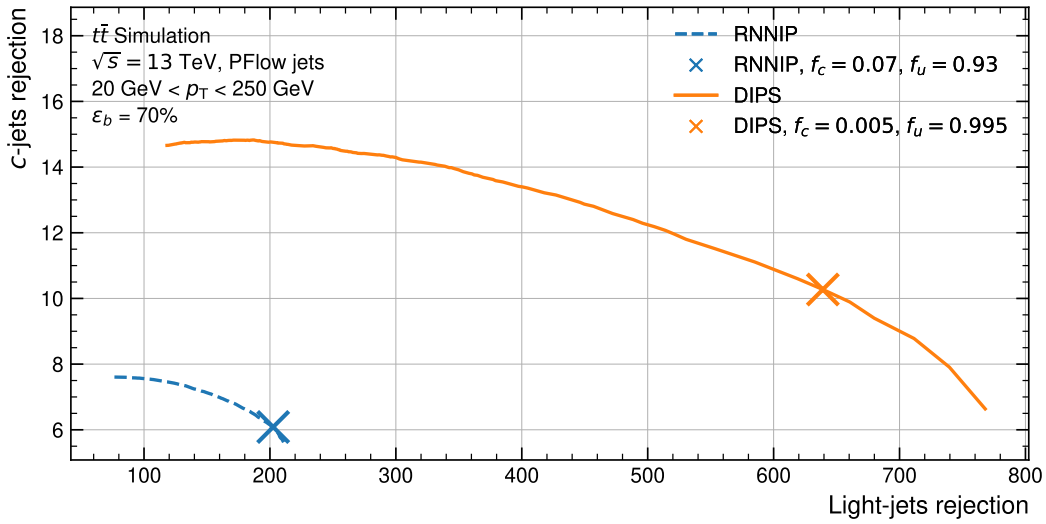


Figure 8.13.: Fraction value scan for DIPS and RNNIP. Depicted are the c- and light-flavour jets rejections as a function of the fraction values at the $70\,\%$ WP. The curves here were calculated by changing the fraction values by 0.005 in each step, ranging from 0 to 1. The chosen fraction values for both RNNIP and DIPS are marked with explicit markers and were chosen to maintain a high light-flavour jets rejection.

$70\,\%$ WP. The resulting rejections for both RNNIP, considered as a reference, and DIPS for the $t\bar{t}$ sample are plotted and depicted in a two-dimensional manner in Figure 8.13.

Both RNNIP and DIPS utilise the selected fraction values displayed in the legend and are marked uniquely in the plot. Both combinations were chosen to sustain high light-flavour jets rejection while maintaining satisfactory c-jets rejection.

In summary, the newly trained DIPS tagger significantly outperforms RNNIP, establishing it as the optimal choice for the new low-level, track-based algorithm intended for integration into the high-level tagger.

9

THE DEEP LEARNING 1 DIPS TAGGER

The high-level tagger for Run 3 of the LHC is the Deep Learning 1 DIPS (DL1d) tagger. DL1d makes use of multiple low-level tagger outputs alongside the jet's kinematic variables as inputs. Despite being relatively similar to its predecessor, DL1d introduces a significant alteration: the replacement of RNNIP with DIPS. This transformation is visually represented in Figure 9.1.

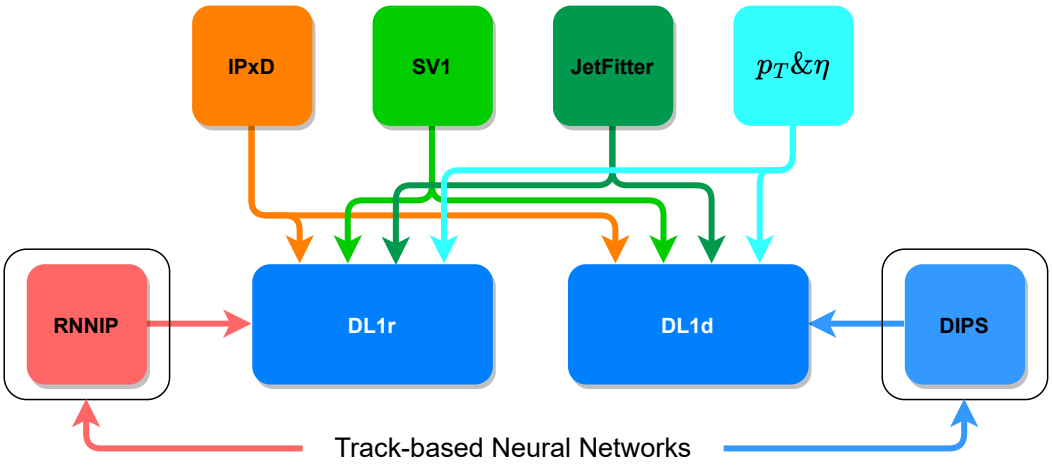


Figure 9.1.: Schematic representation of the inputs to DL1r and DL1d. Here, IPxD denotes the two IP-based low-level algorithms IP2D and IP3D.

This chapter delves into the architecture, training sample, and results of the new DL1d tagger in greater detail. For reference, the predecessor, DL1r, will be used.

9.1 TRAINING SAMPLE

The training, validation, and testing samples used remain the same as described in Section 8.1, except for employing different variables. While the DIPS training incorporated the associated tracks of the jets, DL1d training utilised outputs from the low-level algorithm alongside other jet-level variables. A brief explanation of the variables used is listed in Table 9.1.
All input variables, excluding the "is default" variables, underwent scaling and shifting before being fed into the network. No logarithmic transformation was applied to any input variable, and no additional variable selection was applied.

Apart from substituting RNNIP with DIPS, DL1d incorporates additional input variables, including "is default" variables. These indicators ascertain whether the input variables from their respective low-level algorithms contain proper values or default values. These boolean variables are zero for proper values and one for default values. Default values are necessary when the underlying low-level algorithm fails to execute its task, resulting in the unavailability of output values for the given vari-

Table 9.1.: Input Variables to the DL1d tagger. Extended and adapted from Ref. [172].

Input	Variable	Description
Kinematics	p_{T}	Jet p_{T}
	$\lvert\eta\rvert$	Jet $\lvert\eta\rvert$
IP2D	is default	Indicates if the variables are filled with default values or with proper values
	$\log(p_b/p_{\text{light}})$	LH ratio of the b-jet to light-flavour jet hypotheses
	$\log(p_b/p_c)$	LH ratio of the b-jet to c-jet hypotheses
	$\log(p_c/p_{\text{light}})$	LH ratio of the c-jet to light-flavour jet hypotheses
IP3D	is default	Indicates if the variables are filled with default values or with proper values
	$\log(p_b/p_{\text{light}})$	LH ratio of the b-jet to light-flavour jet hypotheses
	$\log(p_b/p_c)$	LH ratio of the b-jet to c-jet hypotheses
	$\log(p_c/p_{\text{light}})$	LH ratio of the c-jet to light-flavour jet hypotheses
DIPS	p_b	b-jet probability
	p_c	c-jet probability
	p_{light}	Light-flavour jet probability
Secondary-Vertex-Tagging (SV1)	is default	Indicates if the variables are filled with default values or with proper values
	m(SV)	Invariant mass of tracks at the secondary vertex assuming pion mass
	f_E(SV)	Jet energy fraction of the tracks associated with the secondary vertex
	N_{TrkAtVtx}(SV)	Number of tracks used in the secondary vertex
	N_{2TrkVtx}(SV)	Number of two-track vertex candidates
	L_{xy}(SV)	Transverse distance between the primary and secondary vertices
	L_{xyz}(SV)	Distance between the primary and secondary vertices
	S_{xyz}(SV)	Distance between the primary and secondary vertices divided by its uncertainty
	$\Delta R(\vec{p}_{\text{jet}}, \vec{p}_{\text{vtx}})$(SV)	ΔR between the jet axis and the direction of the secondary vertex relative to the primary vertex
JetFitter	is default	Indicates if the variables are filled with default values or with proper values
	m(JF)	Invariant mass of tracks from displaced vertices
	f_E(JF)	Jet energy fraction of the tracks associated with the displaced vertices
	$\Delta R(\vec{p}_{\text{jet}}, \vec{p}_{\text{vtx}})$(JF)	ΔR between the jet axis and the vectorial sum of momenta of all tracks attached to displaced vertices
	S_{xyz}(JF)	Significance of the average distance between PV and displaced vertices
	N_{TrkAtVtx}(JF)	Number of tracks from multi-prong displaced vertices
	N_{2TrkVtx}(JF)	Number of two-track vertex candidates (prior to decay chain fit)
	$N_{\text{1-trk vertices}}$(JF)	Number of single-prong displaced vertices
	$N_{\geq\text{2-trk vertices}}$(JF)	Number of multi-prong displaced vertices
JetFitter (2^{ND} vertex)	is default	Indicates if the variables are filled with default values or with a proper values
	$m_{\text{Trk}}(2^{\text{nd}})$(JF)	Invariant mass of tracks associated with the 2^{nd} vertex
	$E(2^{\text{nd}})$(JF)	Energy of the tracks associated with the 2^{nd} vertex
	$f_E(2^{\text{nd}})$(JF)	Jet energy fraction of the tracks associated with the 2^{nd} vertex
	$N_{\text{TrkAtVtx}}(2^{\text{nd}})$(JF)	Number of tracks associated with the 2^{nd} vertex
	$L_{xy}(2^{\text{nd}})$(JF)	Transverse displacement of the 2^{nd} vertex
	$L_{xyz}(2^{\text{nd}})$(JF)	Distance of 2^{nd} vertex from PV
	$\eta_{\text{trk}}^{\text{min, max, avg}}(2^{\text{nd}})$ (JF)	Min., max. and avg. pseudorapidity of tracks at the 2^{nd} vertex
	$\eta_{\text{trk}-\text{jet}}^{\text{min, max, avg}}(2^{\text{nd}})$ (JF)	Min., max. and avg. pseudorapidity of tracks at the 2^{nd} vertex to the jet

ables.

Three additional variables, termed JetFitter (2^{ND} vertex) variables characterise the minimum, maximum, and average pseudorapidity of the tracks from the second vertex to the jet. Figure A.3 showcases the distributions of all input variables.

9.2 ARCHITECTURE

The DL1d tagger is a DNN based fully-connected feedforward MLP with one input and output layer and multiple hidden layers. A schematic representation is shown in Figure 9.2.

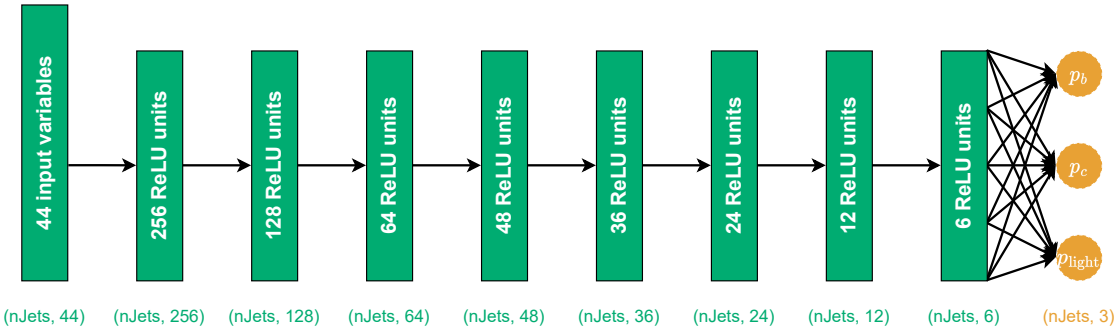


Figure 9.2.: Schematic representation of the architecture of the DL1d tagger. The output classes p_b, p_c and p_{light} depict the output probabilities of DL1d, where each of them can be interpreted as a probability that the processed jet belongs to the respective class.

The architecture of DL1d closely mirrors that of its predecessor, DL1r. It comprises an input layer that receives 44 variables as inputs for training, transmitting them directly to the first hidden layer. It includes eight hidden layers with varying neuron counts per layer and culminates in an output layer producing three values, one for each output class. These three outputs are the three probabilities, which correspond to each jet class. The hidden layer neurons use the ReLU activation function, while the output layer neurons employ the softmax function to compute these probabilities.
Similar to RNNIP, DIPS, and DL1r, these output probabilities serve in computing the b-tagging discriminant D_b. The formula for D_b is defined in Equation (7.6), where f_c and f_{light} represent the fraction values, and p_b, p_c, and p_{light} denote the output probabilities of DL1d, indicating whether the evaluated jet originates from a b, c, or light-flavour quark. DL1d were trained for 50 epochs using TENSORFLOW with the KERAS front-end and employed the ADAM optimiser for optimisation. All trainable parameters were initialised using a Glorot Uniform distribution. The final trained model was exported to the ATLAS reconstruction software using LWTNN.
All hyperparameters underwent optimisation across various versions of DL1d, culminating in a dedicated hyperparameter optimisation phase. The best-found hyperparameters were incorporated into the model presented in this chapter. Table 9.2 outlines all hyperparameters and their values.

Similar to the approach taken for DIPS, the model state of DL1d was saved after each epoch for a more comprehensive examination following the completion of the training process. Furthermore, these preserved model states can serve as the final model if signs of overfitting emerge in subsequent model states. To counter overfitting and co-adaptations, the Dropout regularisation technique was applied in the middle two hidden layers with a factor of 0.2.

Table 9.2.: Hyperparameters of the DL1d tagger.

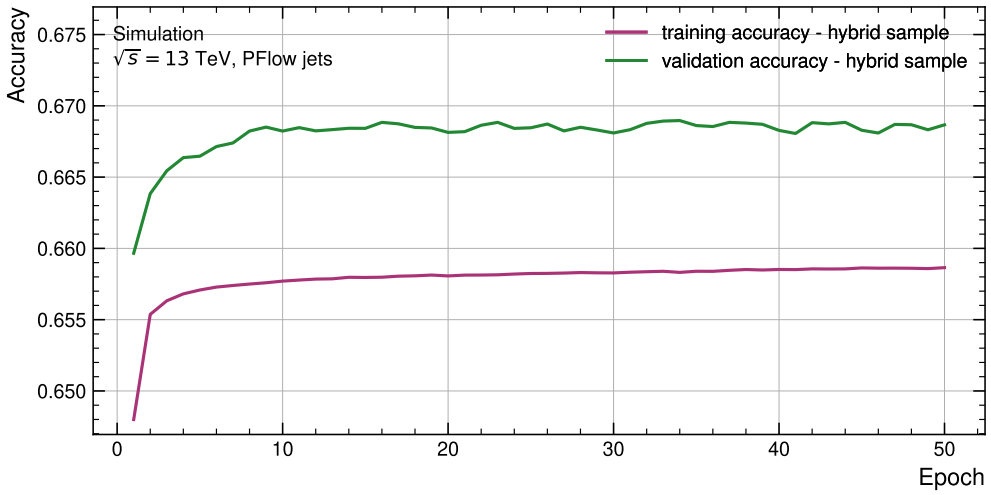
Hyperparameter	Value
Number of input variables	44
Number of hidden layers	8
Number of neurons per layer	[256, 128, 60, 48, 36, 24, 12, 6]
Dropout factor in hidden layers	[0, 0, 0, 0.2, 0.2, 0, 0, 0]
LR	0.001
Batch size	15000
Activation function (hidden neurons)	ReLU
Number of epochs trained	50
Free (trainable) parameters	57227
Fixed parameters	1140
Total parameters	58367

9.3 RESULTS

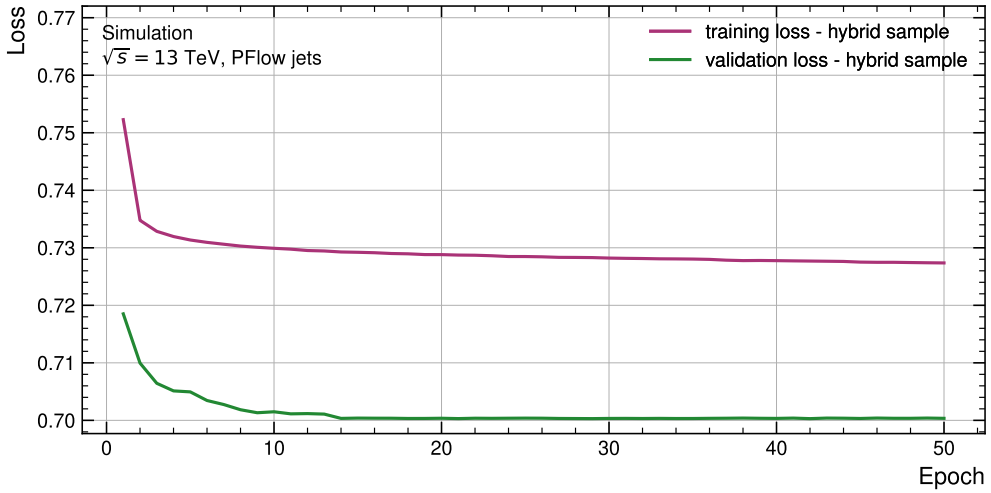
Using the preprocessed samples and variables outlined in Section 9.1, DL1d completed successful training across 50 epochs, leveraging a single NVIDIA V100 GPU. The resulting accuracy and loss curves for both the training and validation samples are illustrated in Figures 9.3 (a) and 9.3 (b), respectively.

As evident from the accuracy and loss curves of DL1d (Figures 9.3 (a) and 9.3 (b)), a notable distinction is observed where the validation accuracy and loss exceed or fall below the training accuracy and loss, respectively. This effect can be attributed to the implementation of Dropout as a regularisation method. Prior trainings highlighted a tendency towards overfitting, primarily attributed to the high performance of DIPS used as input. To counter this behaviour, regularisation methods were introduced. For an in-depth understanding of Dropout and its function, refer to Section 6.1.4.
Employing these strategies effectively mitigated overfitting in DL1d, stabilising the training and validation curves around the 14th epoch. The consistent validation accuracy and loss suggest that the network has reached a threshold in learning the general characteristics of the different jet classes, indicative of achieving a (local) minimum in optimisation. To further scrutinise the network's performance concerning trained epochs, the c and light-flavour jets rejections for the $70\,\%$ WP across hybrid, $t\bar{t}$ and Z' validation samples are depicted in Figures 9.4 (a), 9.4 (b) and 9.4 (c) correspondingly. Notably, the DL1r c and light-flavour jets rejections are presented as vertical lines, reflecting their independence from the number of trained epochs, and are used as reference.

Consistent with the trends observed in the accuracy and loss plots, both c and light-flavour jets rejections in the hybrid validation sample stabilise after around ten epochs of training. A similar behavior is evident in the $t\bar{t}$ validation sample, whereas the Z' validation sample displays a consistent plateau followed by a slight decline in performance for both rejections after epoch 15. Based on these results, the model state after 14 epochs of training was selected for further evaluation.
To conduct a more comprehensive assessment of network performance, the $t\bar{t}$ and Z' testing samples



(a): Accuracy

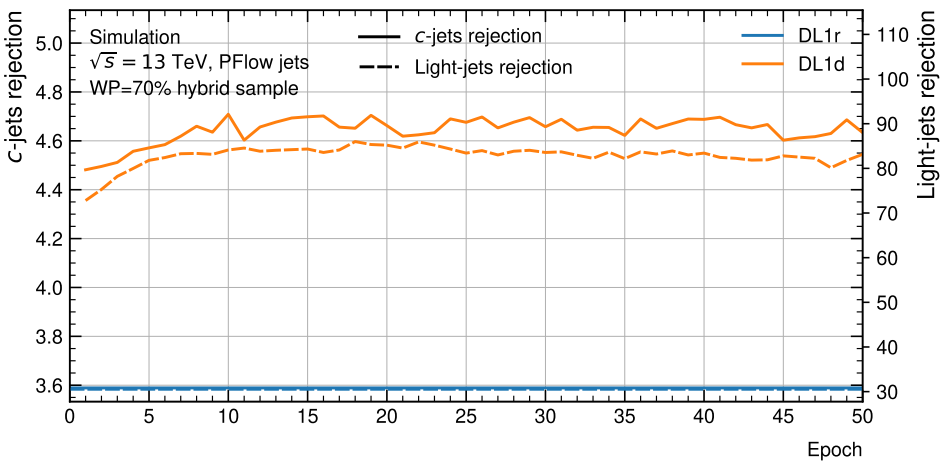


(b): Loss

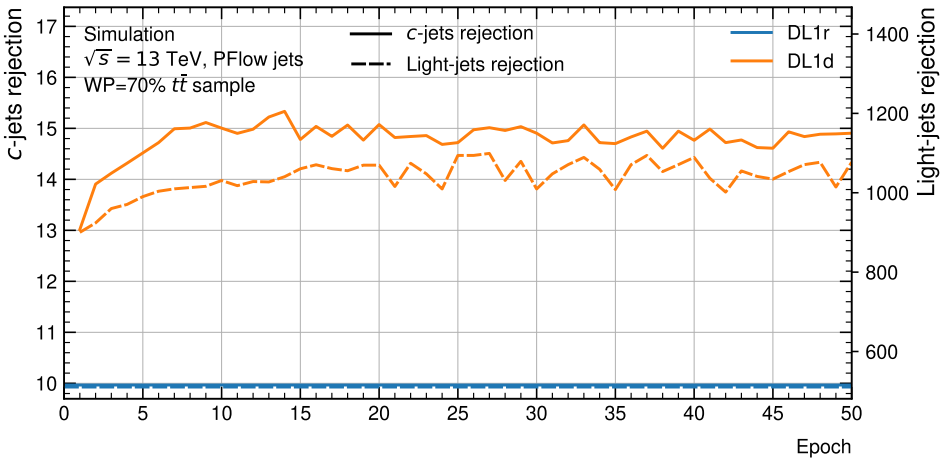
Figure 9.3.: Accuracy (a) and loss (b) curves for the training and validation samples for the training of DL1d. DL1d shows a higher/lower validation accuracy/loss compared to the training accuracy/loss. This is an effect of the usage of Dropout during training.

were utilised for evaluation, comparing DL1d against a reference, namely DL1r. Commencing with the probability output distributions for both test samples, depicted in Figure 9.5, it is apparent that all probability output distributions demonstrate significant separations between various jet classes. This indicates the network's capability to effectively differentiate between jet classes. While the most prominent disparities between distributions are observed in the $t\bar{t}$ test sample, noteworthy separations are also visible in the Z' test sample. Similar to DIPS, the most notable separation is between the b and light-flavour jets distributions, owing to the distinct topological differences among these jet classes. Discriminating c-jets is relatively more challenging due to their similarities with both b and light-flavour jets.

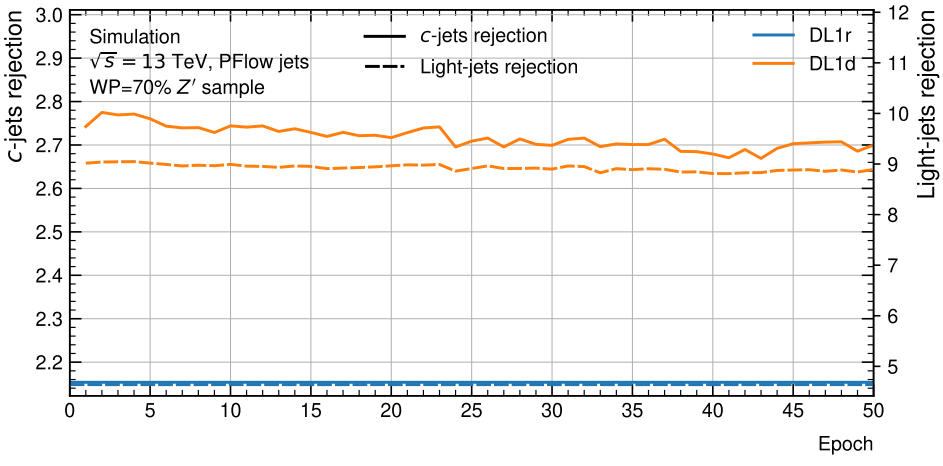
The b-tagging discriminant is computed using Equation (7.6). To determine the optimal fraction values, a fraction scan was conducted employing the $t\bar{t}$ test sample for the 70 % WP, as illustrated in Figure 9.6. The depicted graph portrays the c and light-flavour jets rejections as a function of varying



(a): Resampled hybrid validation sample

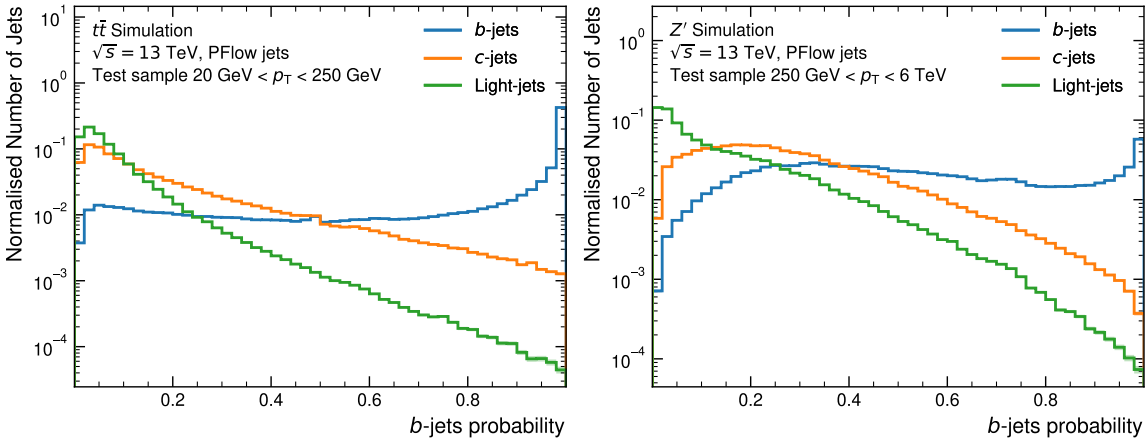


(b): Un-resampled $t\bar{t}$ validation sample



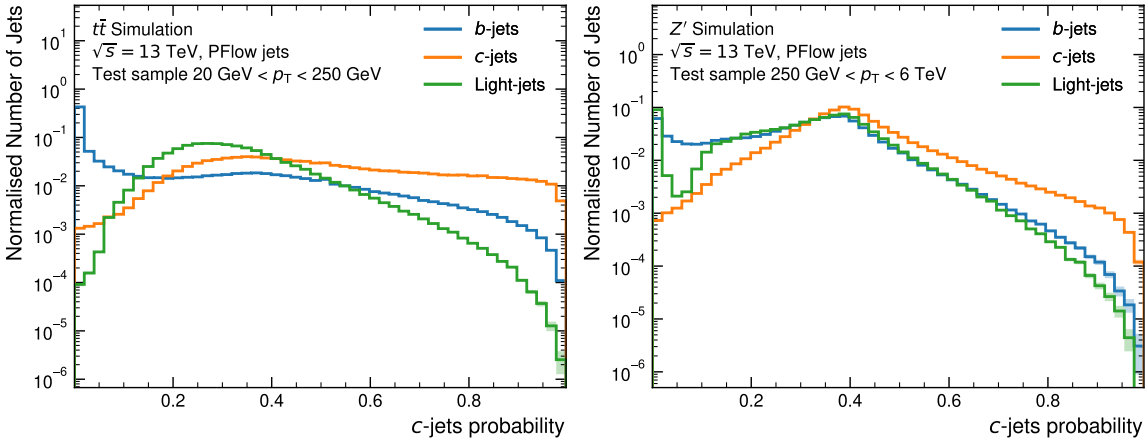
(c): Un-resampled Z' validation sample

Figure 9.4.: c and light-flavour jets rejections for DL1d for the hybrid (a), $t\bar{t}$ (b) and Z' (c) validation samples as a function of training epoch. As reference, the c and light-flavour jets rejections for DL1r are shown in blue.



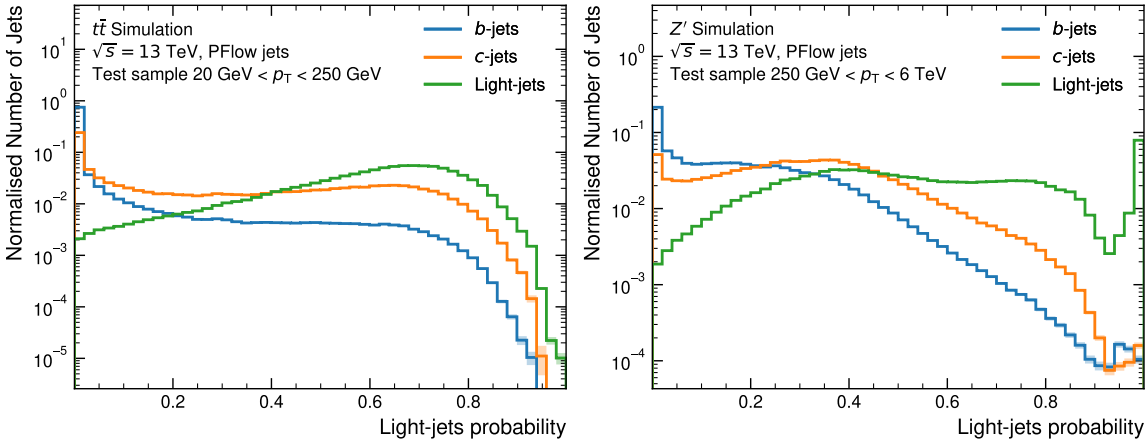
(a): b-jets probability output p_b for $t\bar{t}$

(b): b-jets probability output p_b for Z'



(c): c-jets probability output p_c for $t\bar{t}$

(d): c-jets probability output p_c for Z'



(e): Light-flavour jets probability output p_{light} for $t\bar{t}$

(f): Light-flavour jets probability output p_{light} for Z'

Figure 9.5.: Output probability distributions for both $t\bar{t}$ and Z' jets of the DL1d tagger.

fraction values. Rejections were calculated by iteratively adjusting the fraction values by increments of 0.002 and re-evaluating rejections for each combination. With the requirement that $f_c + f_{\text{light}} \stackrel{!}{=} 1$, 500 potential fraction value combinations within the zero to one range were tested.

Similar to DL1r, which uses $f_c = 0.018$ and f_{light}, DL1d maintains a commendable balance between c and light-flavour jets rejections with these values, thus retaining them for DL1d. Notably, Figure 9.6

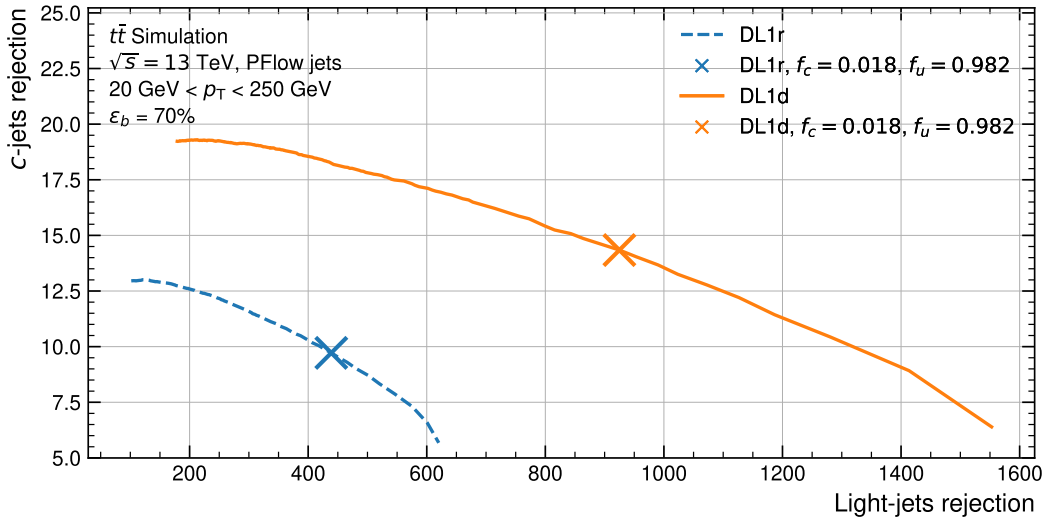


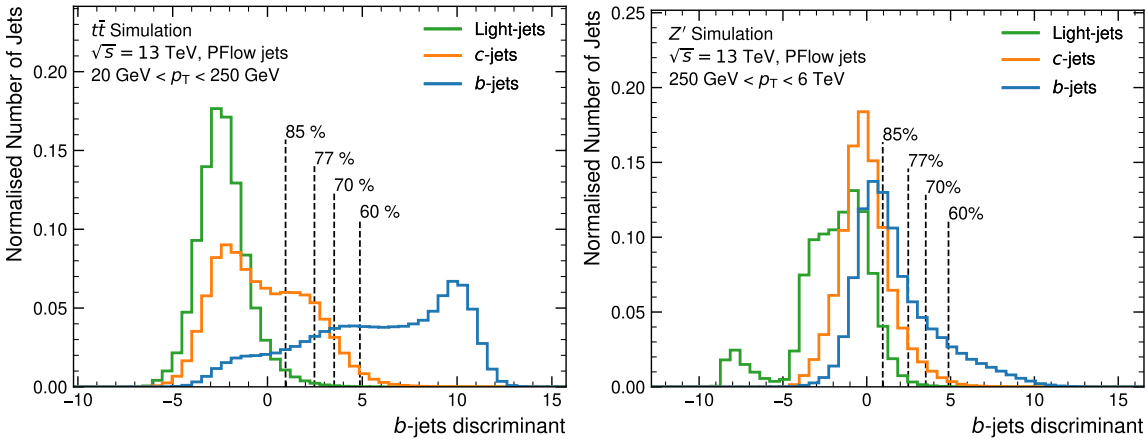
Figure 9.6.: Fraction value scan for DL1d and DL1r. Depicted are the c- and light-flavour jets rejections as a function of the fraction values at the 70 % WP. The curves here were calculated by changing the fraction values by 0.002 in each step, ranging from zero to one. The chosen fraction values for both DL1d and DL1r are marked with explicit markers and were chosen to balance out the resulting c and light-flavour jets rejections.

also presents initial indications of performance enhancements in DL1d over DL1r. The fraction value optimisation solely relies on the $t\bar{t}$ test sample due to the defined procedure used for determining the WPs. As observed in DIPS, the WPs for DL1d are defined based solely on the $t\bar{t}$ test sample, aligning with the primary focus of the ATLAS physics program.

With the optimal fraction values determined, the final b-tagging discriminant for DL1d can be computed for both test samples. The resulting distributions are presented in Figure 9.7. The illustrated WPs and their corresponding cut values are computed based on the $t\bar{t}$ test sample and subsequently applied to both test samples. A clear separation is evident between the to-be-tagged b-jets and all other jet classes in the $t\bar{t}$ test sample. In contrast, a less pronounced separation trend is observed for the Z' test sample. This distinction in separation between $t\bar{t}$ and Z' can be attributed to the increased count of tracks per jet in Z', not stemming from heavy-flavour decays as explained in Section 8.1. This effect notably impacts DL1d, given that the output probabilities of DIPS serve as the primary and performance-driving inputs of DL1d. Moreover, a distinct peak is observable for Z' light-flavour jets in Figure 9.7 (b), identifiable as originating from jets with fewer than two tracks.

The ROC curves are shown in Figures 9.8 (a) and 9.8 (b) for the $t\bar{t}$ and Z' test samples, respectively. Regarding the $t\bar{t}$ ROC curves, a significant enhancement in c and light-flavour jets rejection is evident for DL1d compared to DL1r, with improvement factors reaching up to 1.65 for c-jets rejection and 2.6 for light-flavour jets rejection. Across all provided WPs, DL1d consistently outperforms DL1r.

Notably, substantial performance gains are also witnessed in Z'. It is crucial to note that the x-axis of the plots represents the b-jet efficiency on the respective test sample, while the highlighted WPs were derived exclusively from the $t\bar{t}$ test sample. As a result, the depicted 60 % WP in Figure 9.8 (b) translates to a b-jets efficiency of 11 % in the Z' test sample. This explains the remarkable increase



(a): $t\bar{t}$ jets (b): Z' jets

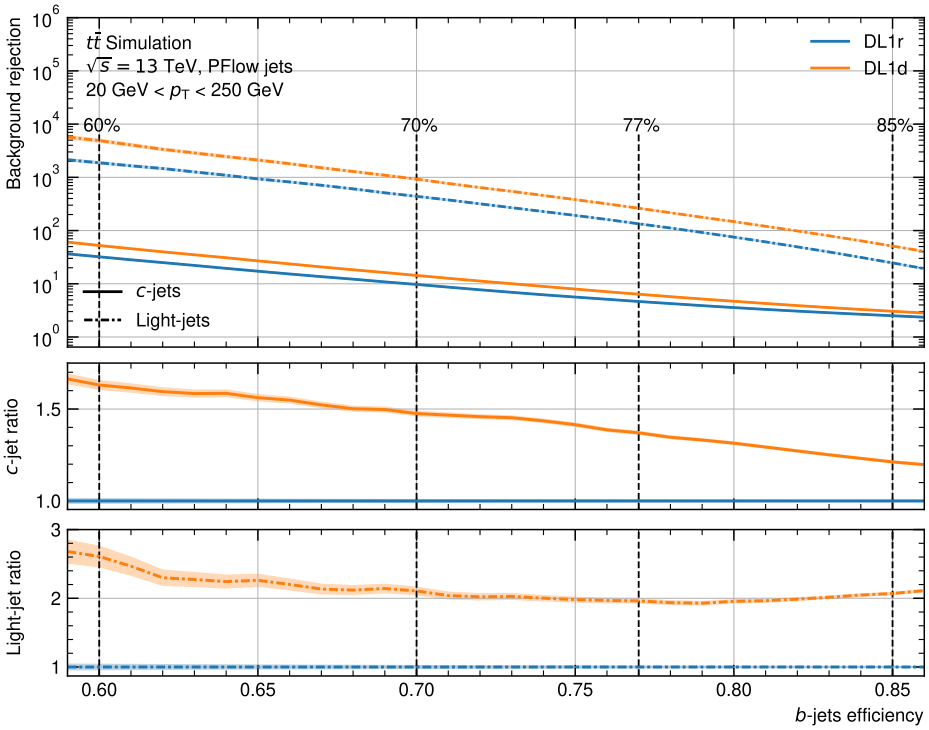
Figure 9.7.: b-tagging discriminant distributions of the DL1d tagger for the $t\bar{t}$ and Z' test samples. The WPs shown in the both plots were derived on the $t\bar{t}$ test sample.

in light-flavour jets rejection of DL1d over DL1r. Moreover, the performance improvements in c and overall light-flavour jets rejections are striking, ranging from around 2 for c-jets rejection to approximately 12 for light-flavour jets rejection at the 70 % WP. These results underscore the compelling advancements in performance achieved by DL1d over DL1r across both test samples and various rejection metrics.

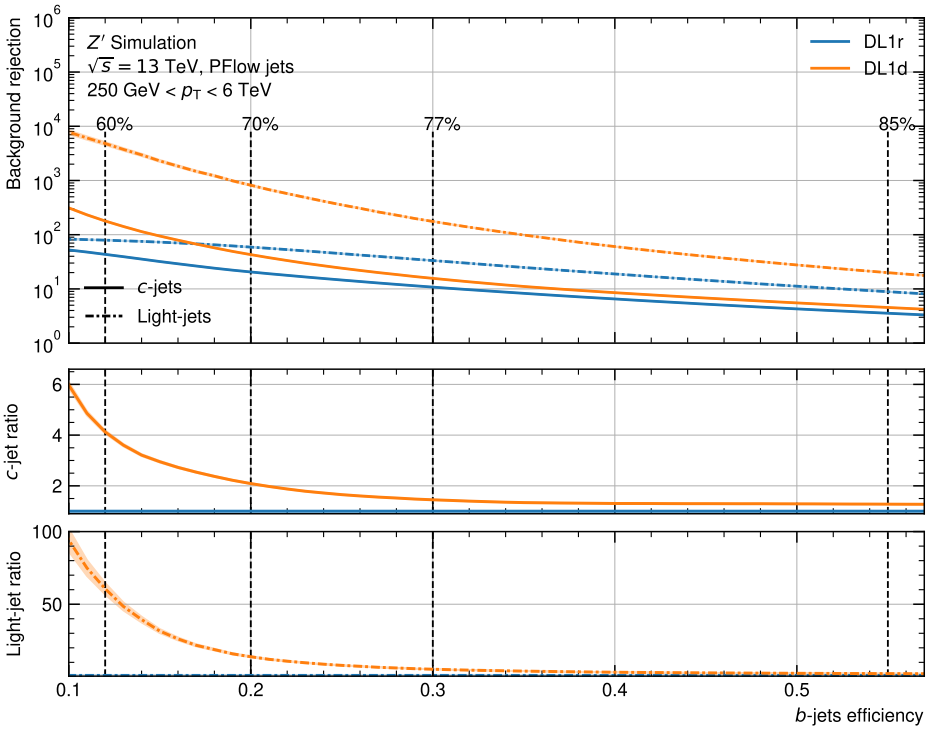
The rejections are depicted as a function of jet p_T using the two previously introduced procedures: Flat efficiency per bin and inclusive. For the former approach, the resulting rejections for the 70 % WP for both test samples are illustrated in Figure 9.9.

Similar to what was observed in the ROC plots, DL1d consistently outperforms DL1r in both test samples for both rejections. While an impressive performance improvement is evident in $t\bar{t}$ for both rejections, with improvement factors of up to 2.4 in c-jets rejection for the $175 - 250$ GeV bin and 4.3 for light-flavour jets rejection in the same bin, even higher improvement factors are evident in Z'. In c-jets rejection, an improvement factor of around 2.2 is achieved in the first bin, ranging from $250 - 500$ GeV, while an improvement factor of around 8 is attained in the $1 - 1.5$ TeV bin.
To further assess performance using the inclusive procedure, the corresponding plots are presented in Figure 9.10.

In addition to the rejections, in the inclusive procedure, also the b-jets efficiencies can be plotted as a function of jet p_T. Although differences in these can show a shift in the performance of the taggers, by definition, the overall efficiency averaged over all bins will be 70 %, therefore only showing shifts in performance. For $t\bar{t}$, a slight shift of up to 5 % towards the higher jet p_T regions is observed. In the case of Z', a drastic shift of up to 40 % from the high to the low jet p_T region is observed. Although these shifts in performance seem drastic, the effect on the overall performance is rather small, due to the relatively small amount of jets above 2 TeV. Also note here, that the overall averaged b-jets efficiency for the Z' test sample is different to 70 %, due to the method used to calculate the WPs. The 70 % WP, calculated on the $t\bar{t}$ test sample, corresponds to an inclusive 20 % b-jets efficiency on

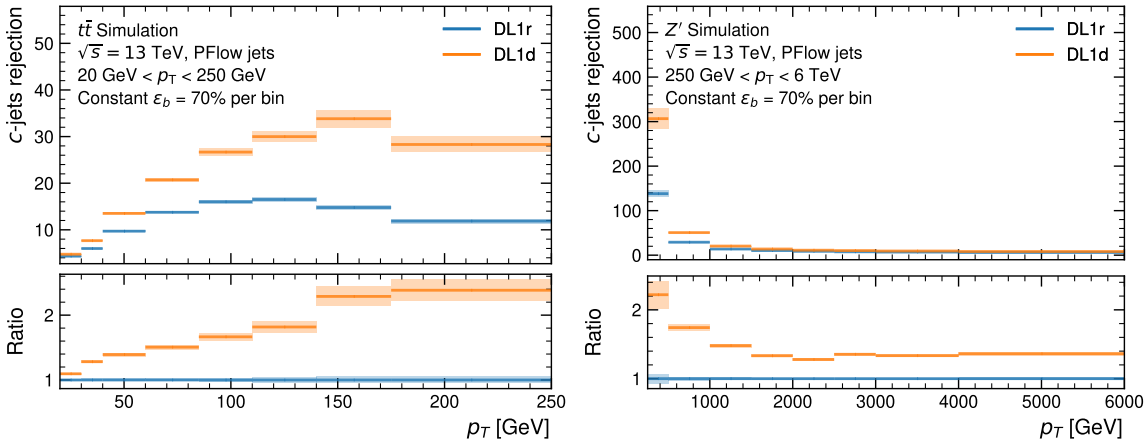


(a): $t\bar{t}$ test sample



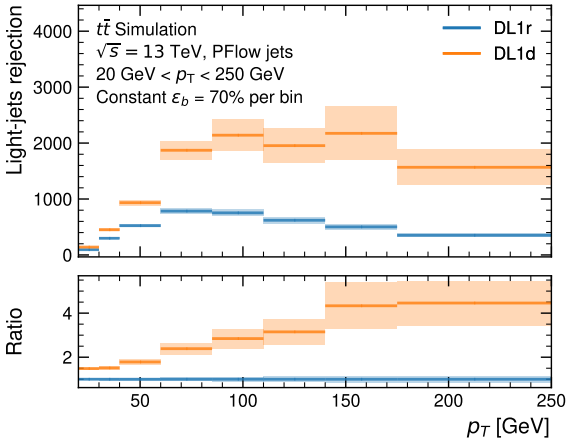
(b): Z' test sample

Figure 9.8.: ROC curves for the DL1d and DL1r tagger. The depicted b-jets efficiencies and rejections for both taggers were calculated using the $t\bar{t}$ and Z' test samples, each containing in total three million jets. The shown b-jets efficiencies on the x-axis denote the b-jets efficiency on the respective test sample, while the depicted WPs were derived from the $t\bar{t}$ test sample and correspond to the WPs which would be used in analyses. The shaded areas denote only the statistical uncertainties, derived by using binomial uncertainties.

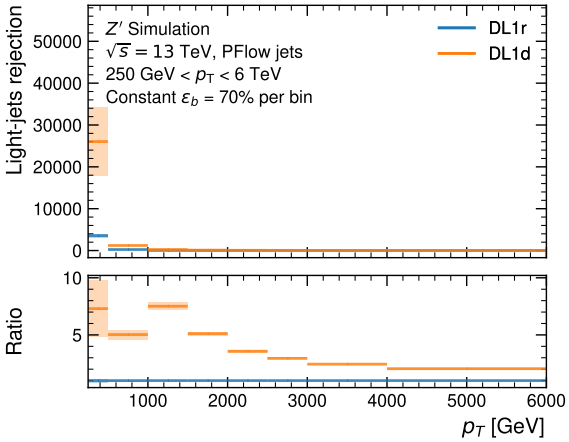


(a): $t\bar{t}$ c-jets rejection

(b): Z' c-jets rejection



(c): $t\bar{t}$ light-flavour jets rejection



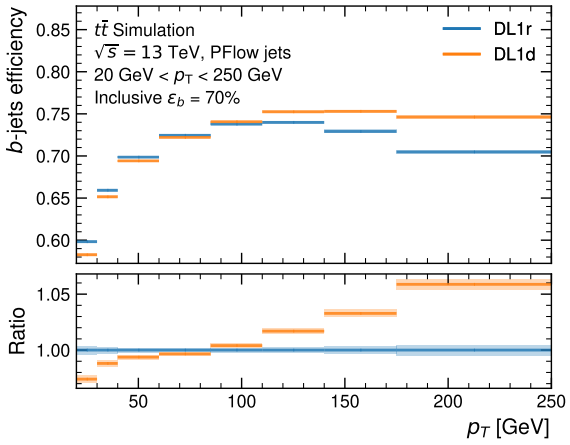
(d): Z' light-flavour jets rejection

Figure 9.9.: c and light-flavour jets rejections for the $t\bar{t}$ and Z' test sample for the 70 % WP using the flat efficiency per bin procedure for the DL1d and DL1r taggers. For both test samples, $t\bar{t}$ and Z', the same WP was used with the same cut value on the b-tagging discriminant, which was derived on the $t\bar{t}$ test sample.

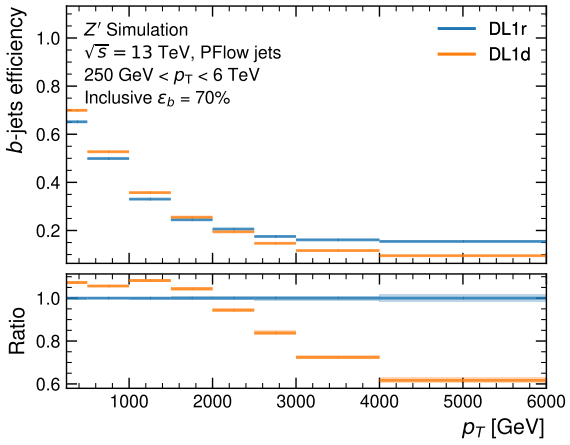
the Z' test sample.

In terms of c-jets rejection, for both test samples, DL1d shows an impressive improvement over DL1r with improvement factors of up to 1.6 for $t\bar{t}$ and around 3.0 for Z'. For the light-flavour jets rejections, improvement factors of up to 2.6 for $t\bar{t}$ and around 13.5 for Z' are achieved. One needs to note here that the large gains in the high-p_T region of Z' are to be taken with care, due to the at the same time reduced b-jets efficiency in these bins, indicating that overall more jets are rejected in these regions. To have a complete picture of the improvements, one needs to take into account plots from both procedures.

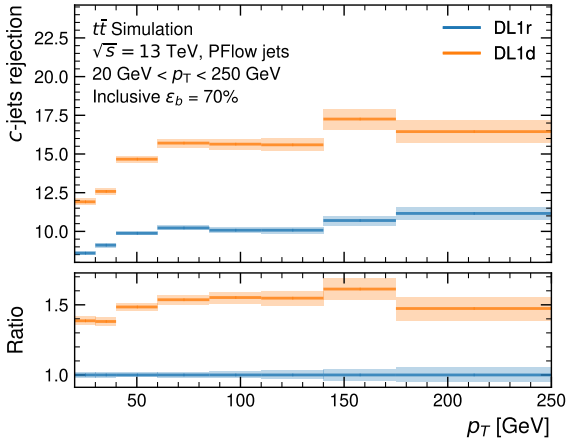
In summary, the new DL1d tagger outperforms the currently recommended and for Run 2 used ATLAS b-tagging algorithm, DL1r. Making use of the new DIPS tagger, which yields much greater classification performance concerning its predecessor RNNIP, DL1d is able to significantly boost the c and light-flavour jets rejections for both test samples. Therefore, DL1d is set to be the new pre-recommended high-level b-tagging algorithm of ATLAS for Run 3.



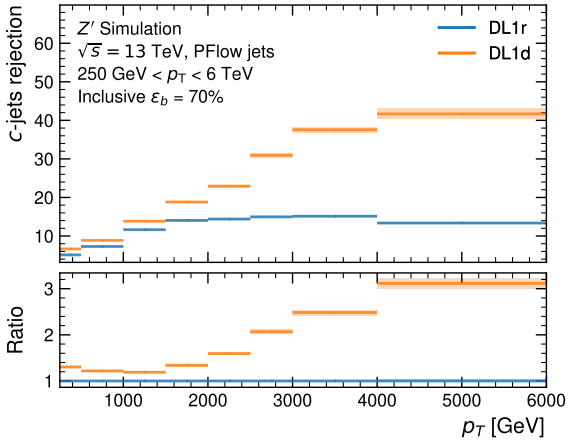
(a): $t\bar{t}$ b-jets efficiency



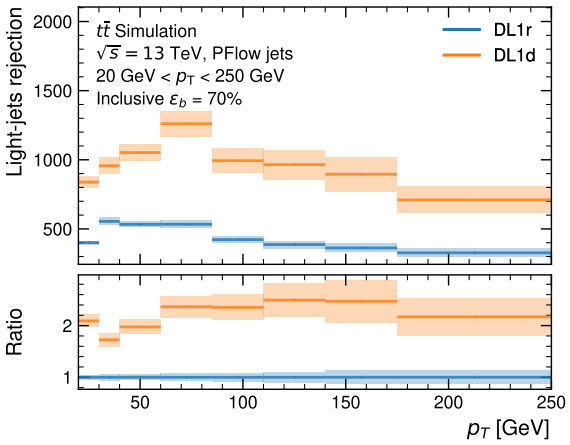
(b): Z' b-jets efficiency



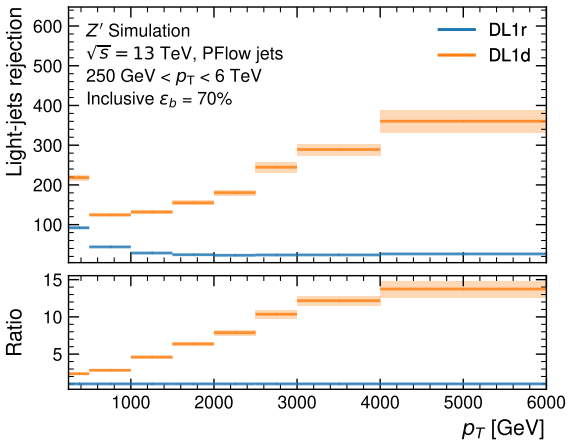
(c): $t\bar{t}$ c-jets rejection



(d): Z' c-jets rejection



(e): $t\bar{t}$ light-flavour jets rejection



(f): Z' light-flavour jets rejection

Figure 9.10.: b-jets efficiency and c and light-flavour jets rejections for the $t\bar{t}$ and Z' test sample for the inclusive 70 % WP for the DL1d and DL1r taggers. For both test samples, $t\bar{t}$ and Z', the same WP was used with the same cut value on the b-tagging discriminant, which was derived on the $t\bar{t}$ test sample.

10

LOOKING AHEAD: DIPS TAU

With the remarkable results achieved by DIPS and the latest high-level algorithm DL1d, the next step for these taggers involves their implementation as c-tagging algorithms. While b-tagging assumes a more pivotal role in identifying heavy-flavour jets, the relevance of c-tagging is rising, particularly in the context of $H \to c\bar{c}$ searches. Although the taggers and their outputs are designed to be easily adapted to serve as c-taggers, a persistent challenge remains: τ-jets.

While the quantity of τ-jets, representing hadronically decaying τ-leptons reconstructed as jets, is not notably large compared to other jet flavours, they pose an issue for c-tagging due to the similarity in tagging discriminants assigned to them by the algorithms. One possible explanation for this particular behaviour is the similar lifetime of τ-leptons and D mesons, resulting in comparable topologies.

For illustration purposes, Figure 10.1 showcases the b-tagging discriminants for DIPS within the $t\bar{t}$ test sample containing τ-jets. It is important to note that the exhibited version of DIPS was not trained using τ-jets.

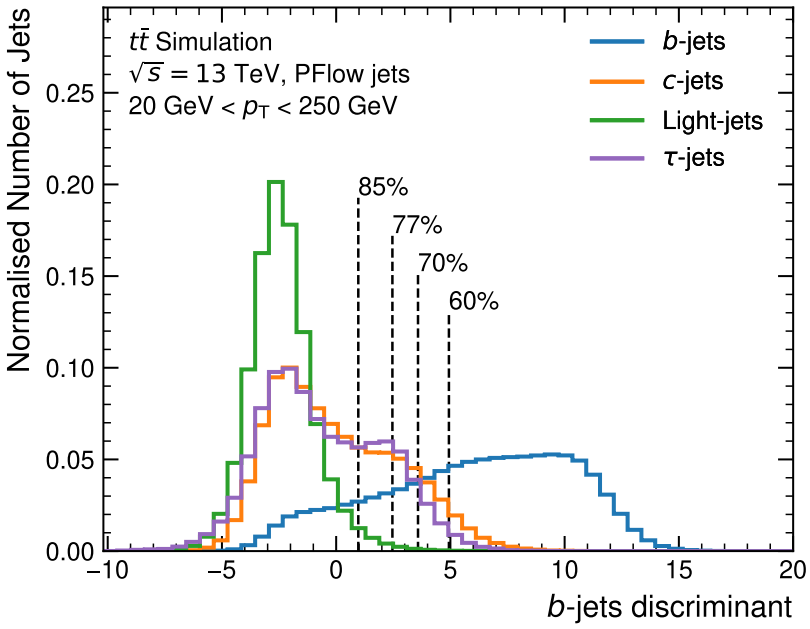


Figure 10.1.: b-tagging discriminant scores for three-classes DIPS (not trained on τ-jets) for the $t\bar{t}$ test sample.

The similarity in shape and overlap between the c and τ-jets distributions explains why simply redefining the taggers as c-taggers is not sufficient. The most straightforward and elegant solution involves adding τ-jets to the training of the taggers as a separate jet class. Although this issue predominantly affects c-tagging, b-tagging, with its outstanding performance, serves as an ideal testing ground to validate the integration of an additional jet class.

The upcoming chapter discusses the incorporation of τ-jets in the training of DIPS. While minimal adaptations were made to the architecture of DIPS to accommodate the new jet class, adjustments in the creation and preprocessing of the training sample were necessary for the same purpose. Subsequently, a training using the new training sample comprising four jet classes was carried out to evaluate the performance of the updated DIPS version, referred to as DIPS Tau. Two approaches were explored in using DIPS Tau: incorporating τ-jets as an additional class in training, assigning a non-zero fraction value for τ-jets in the b-tagging discriminant, and masking the τ-jets output probability by setting the respective fraction value to zero. The latter approach allows us to study the impact of adding τ-jets to the training in the context of standard three-class b-tagging, facilitating a fair comparison with the standard DIPS

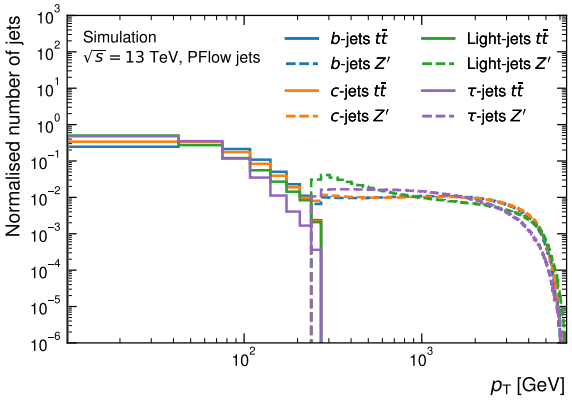
10.1 TRAINING SAMPLE

The training, validation, and test samples for DIPS Tau are created following the preprocessing steps and techniques from Section 7.2. The preprocessing yielded 40 million jets per jet class for the training sample, except for τ-jets, where only 15 million jets are used. This difference in training statistics between the jet classes is due to the low available amount of τ-jets for resampling. In the case of a $t\bar{t}$ event, the only possibility to produce a τ-lepton is by a $W^{\pm}$ boson decay, while all other jet classes can also be produced via additional, QCD-induced, radiations. The resampled jet p_{T} and $|\eta|$ distributions, including the τ-jets, of the training sample are depicted in Figure 10.2, showing the matching p_{T} and $|\eta|$ distributions for all jet classes.

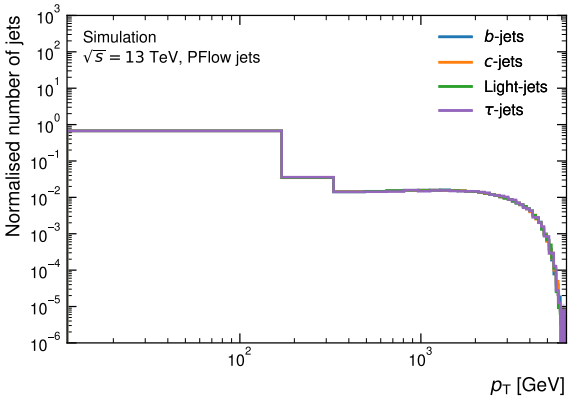
The training, validation, and test samples for DIPS Tau are generated closely following the preprocessing steps and techniques outlined in Section 7.2. The preprocessing resulted in 40 million jets per jet class for the training sample, except for τ-jets, for which only 15 million jets are used. This discrepancy in training statistics between the jet classes stems from the limited availability of τ-jets for resampling purposes. The resampled jet p_{T} and $|\eta|$ distributions, encompassing the τ-jets, in the training sample are illustrated in Figure 10.2, demonstrating the aligned p_{T} and $|\eta|$ distributions across all jet classes. For the validation samples, one million jets per jet class are used, in total four million jets per sample. For the test samples, similar amounts are used: Four million for the $t\bar{t}$ and Z' test samples each.

Furthermore, a schematic representation illustrating the composition of jet classes in the merged training sample before and after resampling is shown in Figure 10.3.

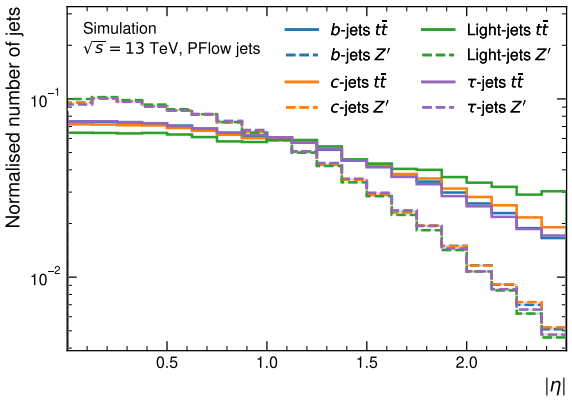
Similarly, the variables and cuts outlined in Table 8.2 were utilised, maintaining consistency with the standard DIPS approach. Additionally, the same preprocessing methodology for input variables, listed in Table 8.1, was employed. The resulting distributions of input variables for $t\bar{t}$ jets, accounting for the presence of τ-jets, are illustrated in Figure A.4.



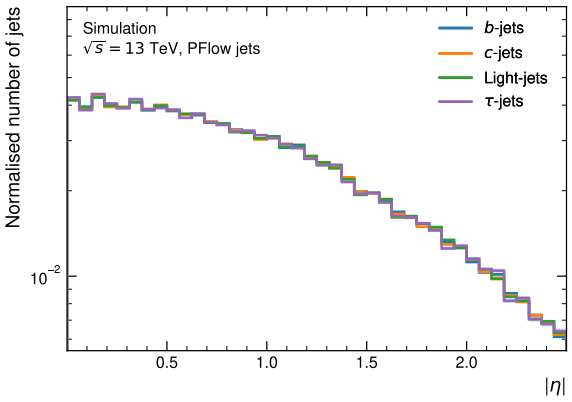
(a): Jet p_T of the un-resampled training sample.



(b): Jet p_T of the resampled training sample.

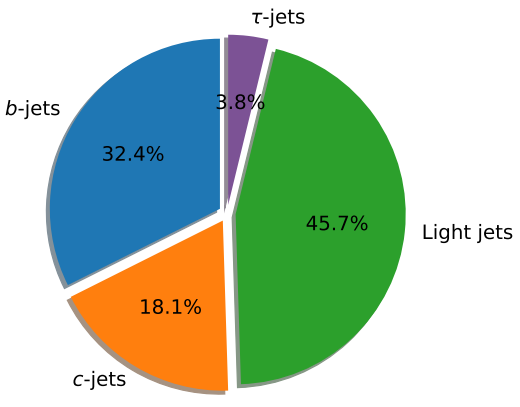


(c): Jet $|\eta|$ of the un-resampled training sample.

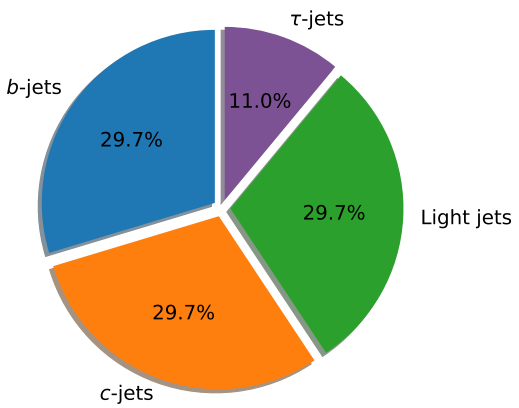


(d): Jet $|\eta|$ of the resampled training sample.

Figure 10.2.: Jet p_T and $|\eta|$ distributions of the un-resampled (a), (c) and resampled (b), (d) merged training sample. The resulting resampled merged training sample has the same amount of jets for each of the jet classes, except for the τ-jets.

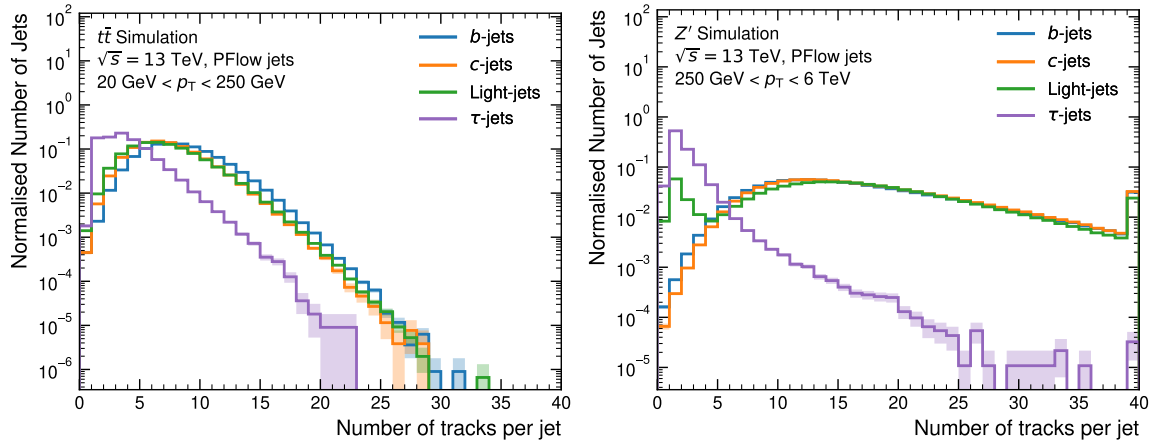


(a): Before resampling.



(b): After resampling.

Figure 10.3.: Schematic representation of the jet class composition of the training sample before (a) and after resampling (b) in terms of percentage of total jets.



(a): $t\bar{t}$ (b): Z'

Figure 10.4.: Number of tracks per jet for both $t\bar{t}$ (a) and Z' (b) jets. For the Z' distribution, an overflow bin is used in the last bin, indicating that more than 40 tracks are associated to some jets.

An important difference between τ-jets and the other jet classes lies in the number of associated tracks. Figure 10.4 showcases the number of tracks per jet for both $t\bar{t}$ and Z', considering all jet classes employed during the training of DIPS Tau.

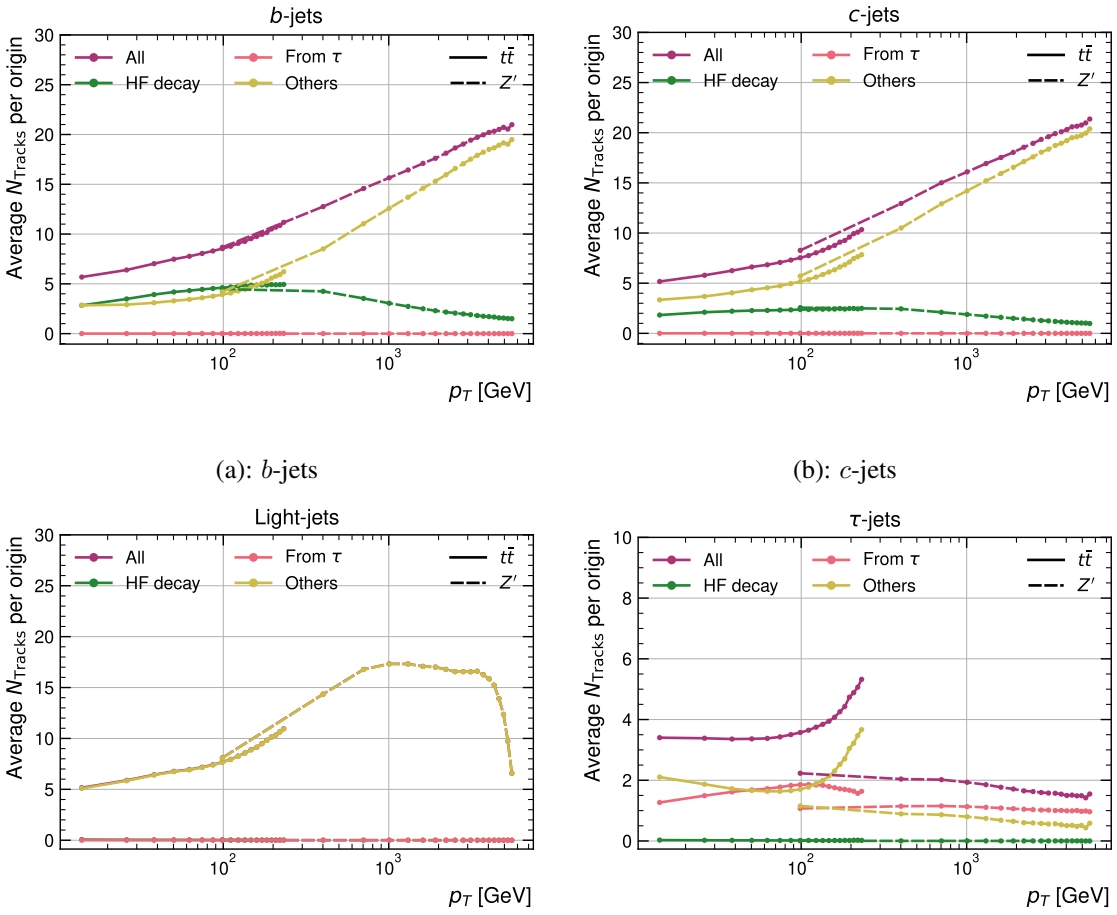
A clear and striking difference can be observed for the new τ-jet class, with a drastically smaller averaged number of tracks per jet for both, $t\bar{t}$ and Z'. This overall decrease in tracks per jet can be explained by the missing hadronic interaction of the τ lepton, which therefore decays relatively clean, meaning without any hadronic interactions besides the actual decay. Although this is true for both $t\bar{t}$ and Z', there is still a slight difference observable between them. This difference, where a smaller but longer tail is present in Z', is due to the origin of the τ lepton itself. While in $t\bar{t}$ processes, the τ lepton originates from the decay of $W^{\pm}$ boson, in Z' it originates from the direct decay of the artificial Z', making the latter environment less affected by hadronic interactions[1]. The longer tail in Z' can be explained by the overall higher p_T of the Z' jets, enabling the hadronic decay productions of the τ lepton to further decay or interact and therefore create more tracks.

Further investigation into these nuances and disparities involves plotting the number of associated tracks per jet origin against jet p_T. These plots, distinct for each jet class and combining jets from $t\bar{t}$ and Z', are presented in Figure 10.5, enabling a detailed exploration of the differences.

Unlike the plots used for DIPS (Figure 8.2), the track origins in this representation have been slightly redefined to categorise 'From τ' as a separate origin. 'HF decay' here represents tracks stemming from the decay of either a b or a c-hadron, while 'Others' includes tracks from sources like pile-up, fakes, primary vertices, and other secondary interactions.

While the standard jet classes exhibit the anticipated trend of higher track counts from other sources in the high jet p_T range, τ-jets display distinctly different behaviour. The evident reduction in the overall number of tracks per jet for τ-jets is apparent, with no observable increase in the number

1 Although the $W^{\pm}$ boson decay is also free from hadronic interactions, the dense environment of $t\bar{t}$ processes usually leads to overlaps in the reconstruction of the jets, explaining the higher number of τ-jets with three or four tracks.



(a): b-jets

(b): c-jets

(c): Light-flavour jets

(d): τ-jets

Figure 10.5.: Number of tracks per jet as a function of jet p_{T} for b (a), c- (b), light-flavour (c), and τ-jets (d) for both $t\bar{t}$ and Z' jets. The chosen binning gives 20 bins each for $t\bar{t}$ between $20 - 250\,\text{GeV}$ and Z' between $250 - 6000\,\text{GeV}$. The statistical uncertainty of the bins is indicated with a coloured band but is too small to be seen. HF decay denotes tracks that are originating from the decay of either a b or a c-hadron, while "Others" is comprised of all other track origins, like pile-up, fakes, primary vertex, and other secondary interactions. "From τ" denotes tracks, that directly originate from the decay of the τ.

of tracks from other sources in the high jet p_{T} regime. This characteristic behaviour is due to the absence of hadronic interactions, owing to the leptonic nature of the τ. While this accounts for the low average number of tracks per jet in Z', an increase in the average number of tracks in the higher p_{T} range of $t\bar{t}$ is observed. This behaviour is expected, considering the higher chance of overlapping jets with increased p_{T}, a phenomenon specific to $t\bar{t}$ due to its denser process environment.

10.2 ARCHITECTURE

To accurately assess the impact of incorporating τ-jets, the architecture of DIPS Tau remains consistent with the original structure outlined in Section 8.2 (Figure 8.4). The sole adjustment made

Table 10.1.: Hyperparameters of the DIPS Tau tagger.

Hyperparameter	Value
Number of input variables	15
ϕ $N_{\text{Hidden layers}}$	4
ϕ $N_{\text{Neurons/layer}}$	[128, 256, 256, 256]
ρ $N_{\text{Hidden layers}}$	6
ρ $N_{\text{Neurons/layer}}$	[256, 256, 128, 128, 100, 30]
Number of output variables	4
Loss function	Categorical Crossentropy
Optimiser	ADAM
LR	0.001
Batch size	15000
Aggregation function	Summation
Activation function (hidden neurons)	ReLU
Activation function (output neurons)	softmax
Number of epochs trained	200
Free (trainable) parameters	367290
Fixed parameters[2]	3588
Total parameters	370878

involves adding an extra neuron in the output layer dedicated to the new τ-jet output probability. Table 10.1 provides a comprehensive list of the hyperparameters specifically chosen for DIPS Tau.

No major changes were introduced to the hyperparameters of DIPS Tau. Given the modified architecture and the additional output probability values generated by DIPS Tau for τ-jets, adjustments are required in defining the b-tagging discriminant. This adaptation involves incorporating the new τ-jet probability by adding it in the denominator, multiplied by the new fraction value for the τ-jet class, denoted as f_τ. The exact formulation is provided in Equation (10.1).

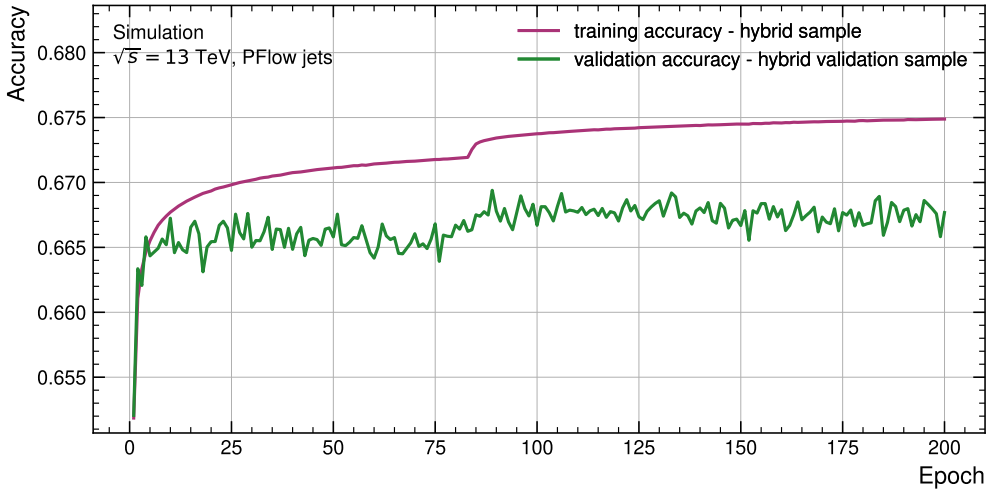
$$D_b = \ln\left(\frac{p_b}{f_c \cdot p_c + f_{\text{light}} \cdot p_{\text{light}} + f_\tau \cdot p_\tau}\right) \quad \text{with} \quad f_c + f_{\text{light}} + f_\tau \stackrel{!}{=} 1 \tag{10.1}$$

The redefined b-tagging discriminant allows for a comparison between the performance of DIPS Tau and DIPS using two distinct procedures: employing a non-zero value for f_τ and using zero as the value for f_τ. The former leverages the newly incorporated τ-jets output probability, while the latter masks this value, essentially using DIPS Tau as a standard three-class DIPS. Although the benefit of setting $f_\tau = 0$ might not be immediately evident, incorporating τ-jets in training, even if their output is unused, could potentially positively impact the overall performance of the tagger, including the other class rejections.

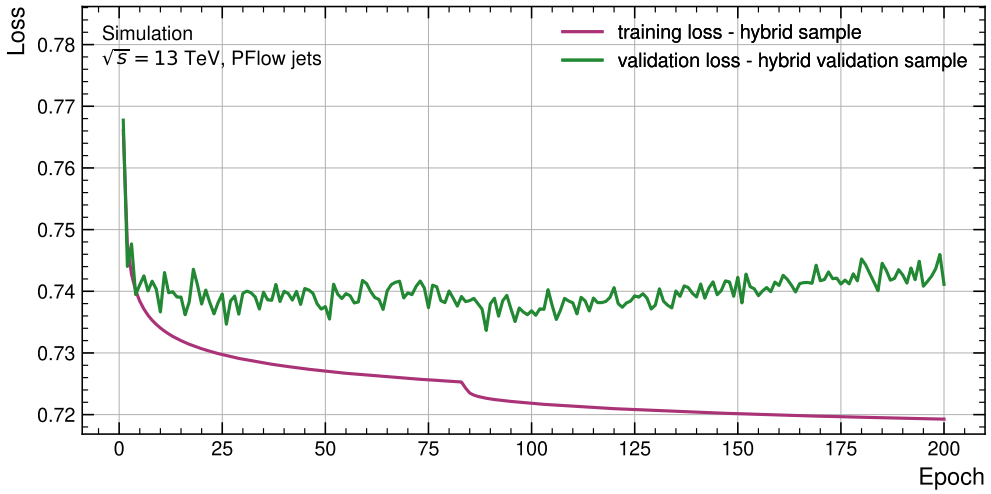
2 Non-trainable parameters are an effect from using Batch Normalisation in the network.

10.3 RESULTS

DIPS Tau was trained for 200 epochs using a single NVIDIA V100 GPU. The resulting accuracy and loss curves for both the training and validation samples are illustrated in Figures 10.6 (a) and 10.6 (b), respectively.



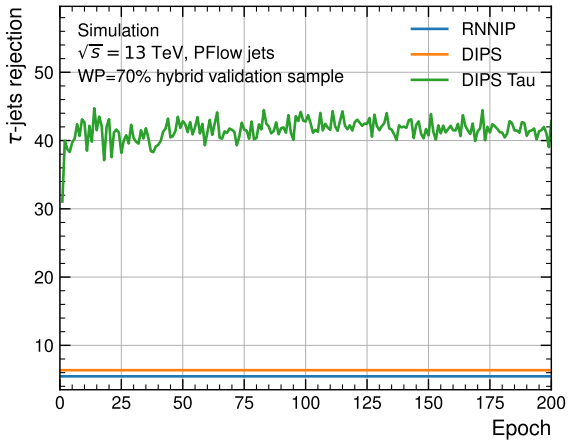
(a): Accuracy



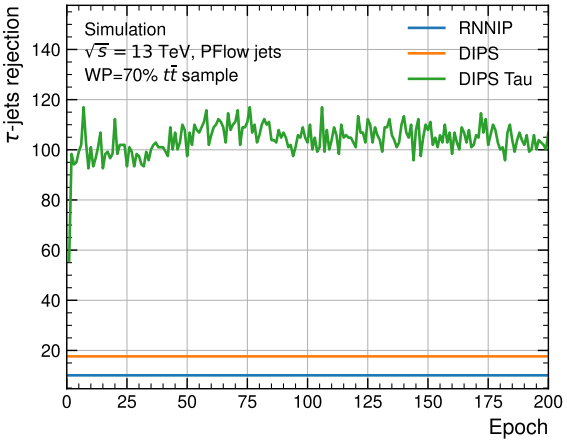
(b): Loss

Figure 10.6.: Accuracy (a) and loss (b) curves for the training and validation samples for the training of DIPS Tau.

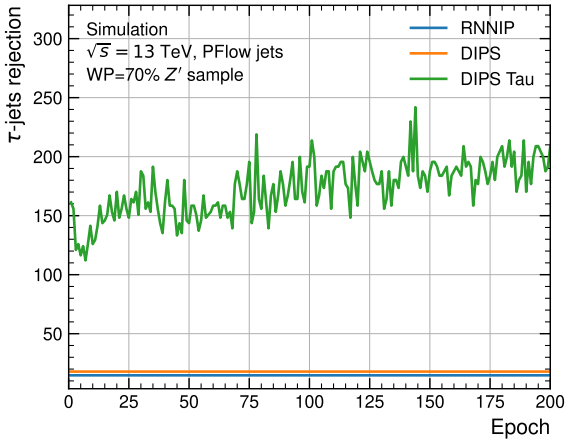
Similar to DIPS, DIPS Tau exhibits continuously improving training accuracy and loss curves, while their validation counterparts quickly reach a plateau after a few epochs. The validation curves never reach the performance levels of the training curves but remain at a consistent value.
In both the training and validation curves, a significant performance surge occurs after approximately 80 epochs, resembling the network 'jumping' out of a local loss minimum and discovering a better-performing minimum. Following this notable leap, the validation curves begin to slightly decline, indicating a potential risk of overfitting in this scenario. To prevent this, the model state after the 89th training epoch is marked for further evaluation.



(a): Resampled hybrid validation sample



(b): Un-resampled $t\bar{t}$ validation sample



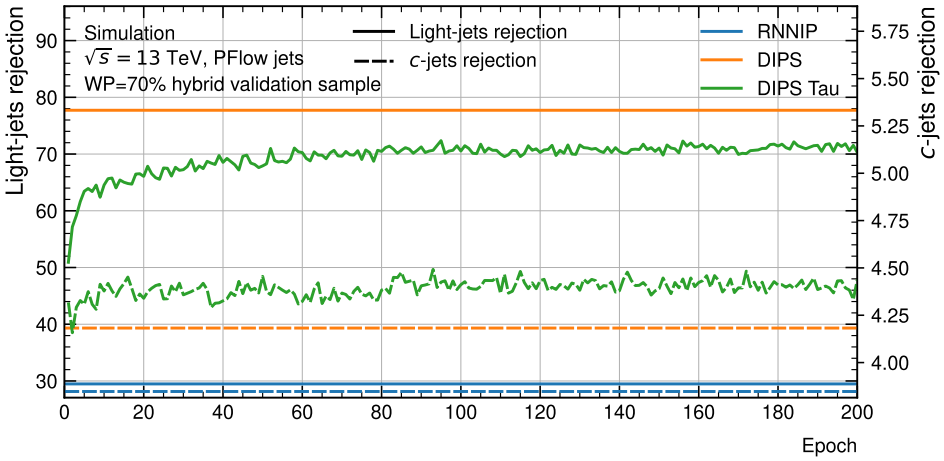
(c): Un-resampled Z' validation sample

Figure 10.7.: τ-jets rejections for DIPS Tau for the hybrid (a), $t\bar{t}$ (b) and Z' (c) validation samples as a function of the training epoch. As reference, the τ-jets rejections for RNNIP and DIPS are shown in blue and orange.

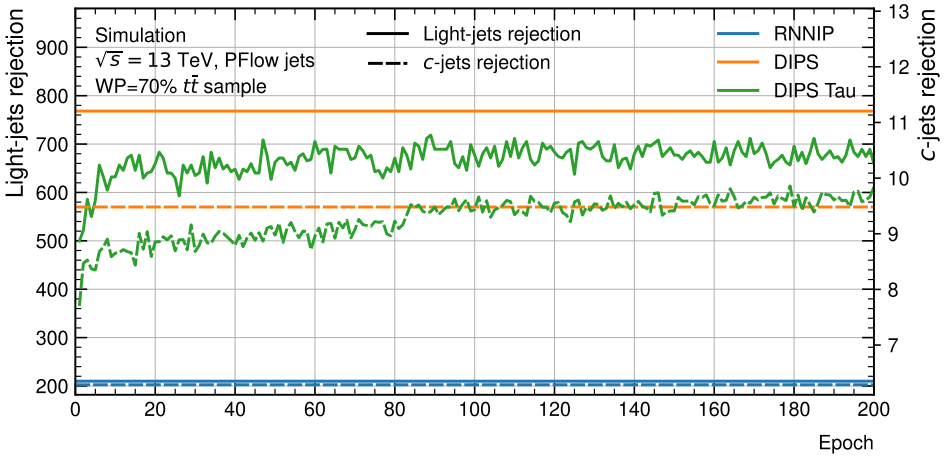
To verify the selected model state's performance and suitability for further evaluation, the rejections for c and light-flavour jets, as well as the τ-jets rejections, are assessed using the hybrid, $t\bar{t}$, and Z' validation samples. The τ-jets rejections are illustrated in Figure 10.7, while the c and light-flavour jet rejections are presented in Figure 10.8.

Analysing the τ-jets rejections, it is evident that DIPS Tau significantly enhances τ-jets rejection compared to both DIPS and RNNIP across all three validation samples. While the c-jets rejections show slightly higher or comparable performance to that of DIPS, there is an overall decrease in light-flavour jet rejections observed in the hybrid and $t\bar{t}$ validation samples. This reduction can be attributed to the adjustments in the fraction values, notably a decrease in the value of f_{light} to accommodate the usage of f_τ. Further details on this adjustment will be presented in the forthcoming fraction value optimisation.

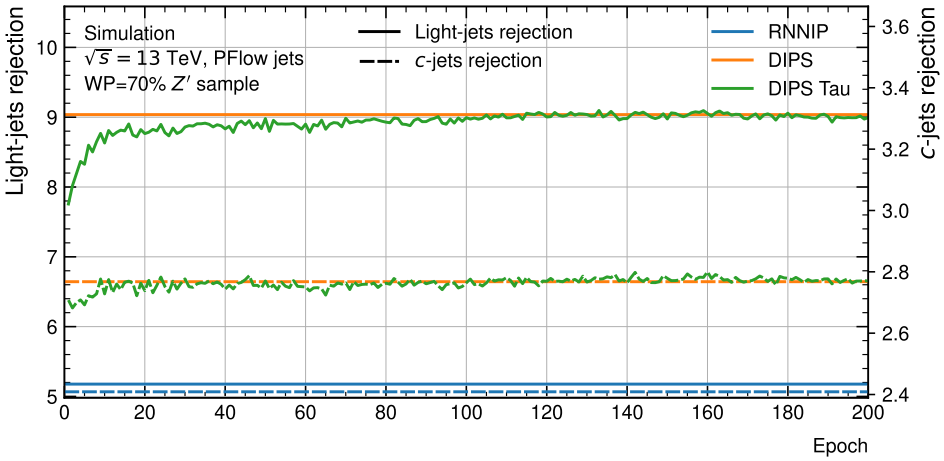
Overall, the identified model state after epoch 89 demonstrates strong performance across all rejection types.



(a): Resampled hybrid validation sample

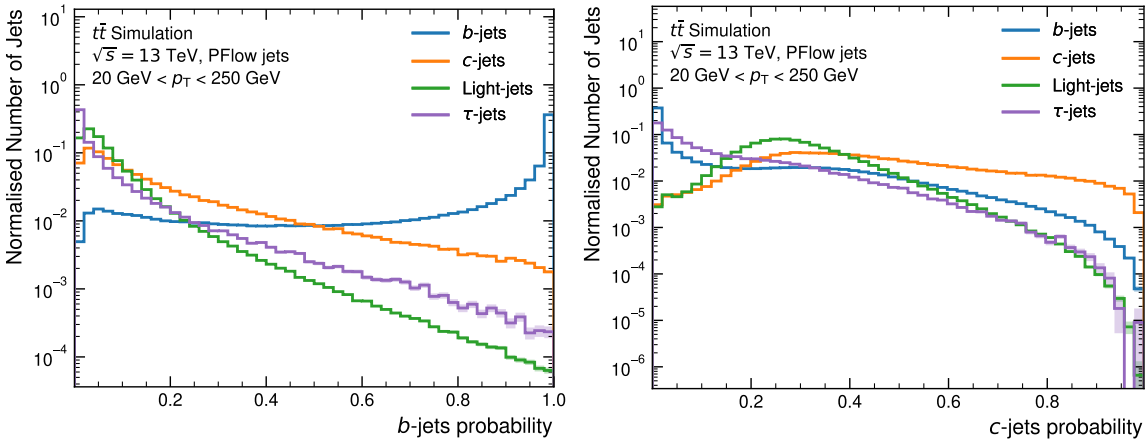


(b): Un-resampled $t\bar{t}$ validation sample



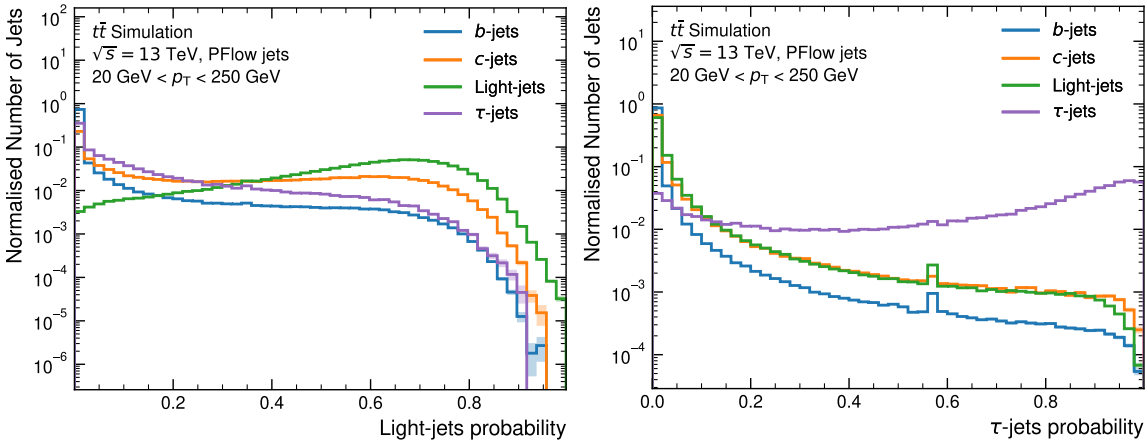
(c): Un-resampled Z' validation sample

Figure 10.8.: c and light-flavour jets rejections for DIPS Tau for the hybrid (a), $t\bar{t}$ (b) and Z' (c) validation samples as a function of training epoch. As reference, the c and light-flavour jets rejections for RNNIP and DIPS are shown in blue and orange.



(a): b-jets probability output p_b

(b): c-jets probability output p_c



(c): Light-flavour jets probability output p_{light}

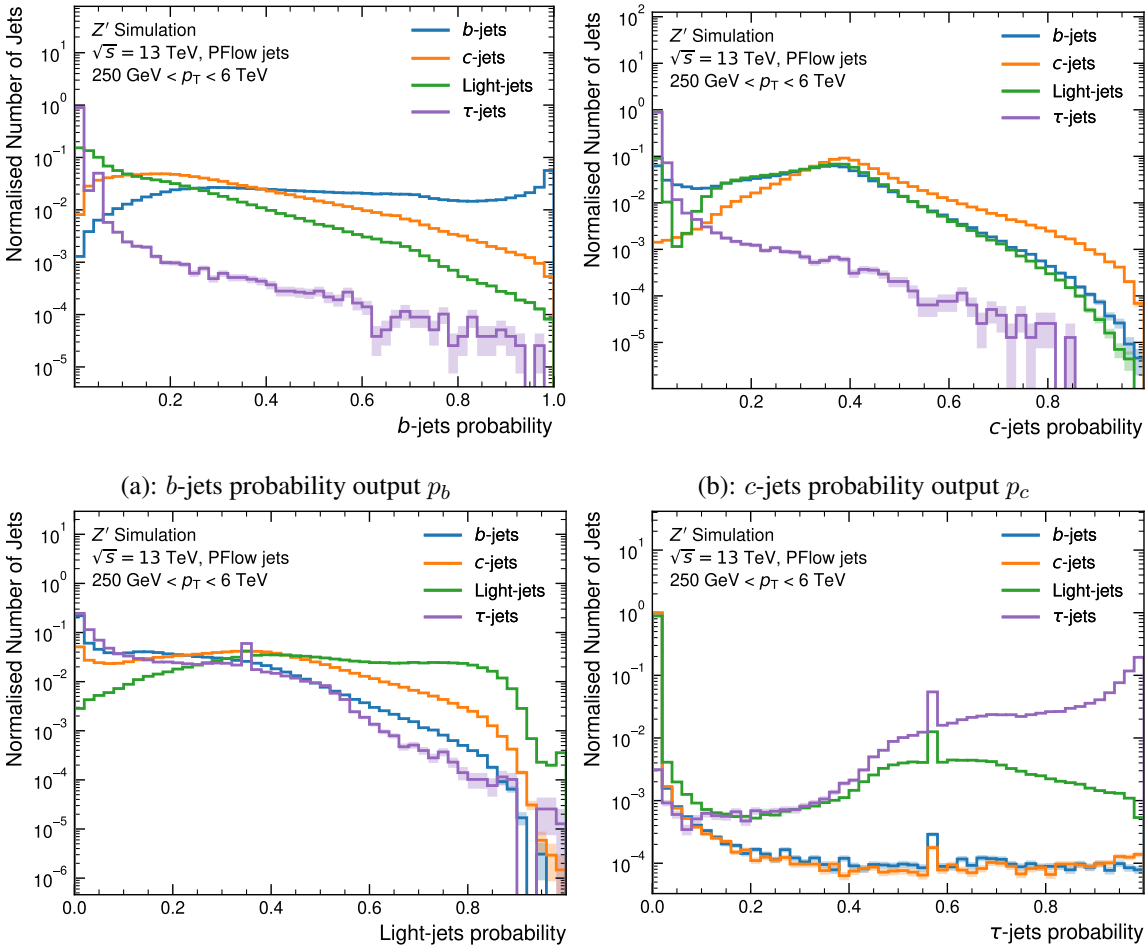
(d): τ-jets probability output p_τ

Figure 10.9.: Output probability distributions of DIPS Tau for the $t\bar{t}$ test sample. The spikes in the distributions originate from jets where no associated track survives the track selection. Therefore, DIPS Tau receives only masked input values for which it returns one fixed set of values. The shaded areas denote only the statistical uncertainties, derived by using binomial uncertainties.

The resulting output probability distributions for the selected model state, referred to as DIPS Tau from now on, are illustrated in Figures 10.9 and 10.10 for the $t\bar{t}$ and Z' test samples, respectively.

In both test samples, a noticeable separation exists among the distributions in each of the output probabilities. The most prominent separations are observed between the b and light-flavour jets distributions in the b-jets probability, as well as between the τ and b-jets distributions in the τ-jets probability. These distinct separations in output probability distributions demonstrate the impressive capability of DIPS Tau in effectively distinguishing and categorising τ-jets from other jet classes. Leveraging these output probabilities alongside Equation (10.1), the b-tagging discriminant can be calculated, and further optimisation of the fraction values becomes feasible. To accomplish this, the two-dimensional plotting of c and light-flavour jet rejections, deemed more crucial than τ-jets rejection, against the fraction values is illustrated in Figure 10.11.

The fraction value scan involved incremental changes of 0.005 for f_c and f_{light} and 0.2 for f_τ in each step. In the case of DIPS Tau, the chosen values are $f_c = 0.005$, $f_{\text{light}} = 0.795$, and $f_\tau = 0.2$. These values were selected to maintain a similar c-jets rejection as compared to DIPS, while not



(a): b-jets probability output p_b

(b): c-jets probability output p_c

(c): Light-flavour jets probability output p_{light}

(d): τ-jets probability output p_τ

Figure 10.10.: Output probability distributions of DIPS Tau for the Z' test sample. The spikes in the distributions originate from jets where no associated track survives the track selection. Therefore, DIPS Tau returns default values. The shaded areas denote only the statistical uncertainties, derived by using binomial uncertainties.

significantly compromising the light-flavour jets rejection.

When comparing standard DIPS and DIPS Tau with $f_\tau = 0$, an interesting observation can be made. The addition of τ-jets to the training sample, even when masking the respective output probability values, results in similar, slightly improved performances in terms of c and light-flavour jet rejections. To further illustrate this, the fraction scans for light-flavour and τ-jets rejections are plotted for both DIPS and DIPS Tau with $f_\tau = 0$, as shown in Figure 10.12.

As evident, even without using the τ output probability, one observes an improvement in light-flavour jets rejection while maintaining a similar level of τ-jets rejection. When considering this alongside Figure 10.11, it becomes evident that the inclusion of the τ-jet class enhances the performance of the DIPS taggers, even in scenarios where the respective output probabilities are not utilised.

Having defined the fraction values, the b-tagging discriminant for different jets can be calculated using Equation (10.1) and then plotted. The resulting b-tagging discriminant distributions for the $t\bar{t}$ and Z' test samples are presented in Figures 10.13 (a) and 10.13 (b).

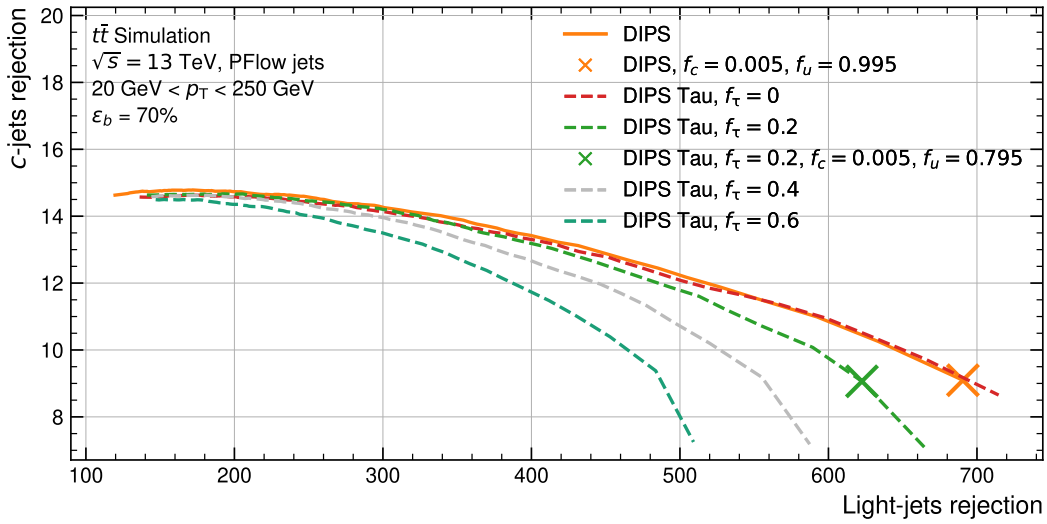


Figure 10.11.: Fraction value scan for DIPS Tau and DIPS. Depicted are the c and light-flavour jets rejections as a function of the fraction values at the 70 % WP. The curves here were calculated by changing the values of f_c and f_{light} by 0.005 and f_τ by 0.2 in each step, ranging from zero to one. The chosen fraction values for both DIPS Tau and DIPS are marked with explicit markers and were chosen to balance out the resulting c and light-flavour jets rejections. The DIPS Tau f_τ value of 0.2 was chosen to retain a relatively high light-flavour jet rejection but also having a good τ-jets rejection.

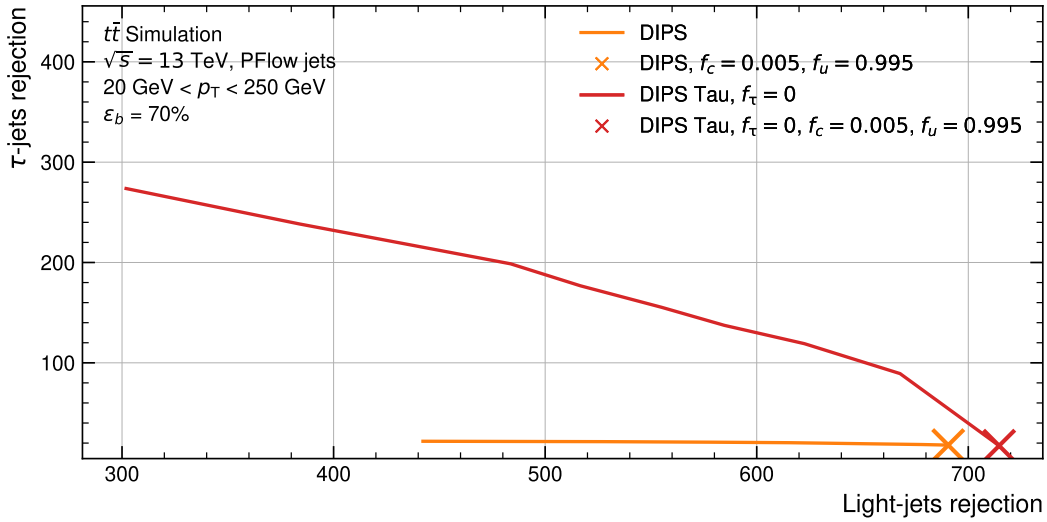
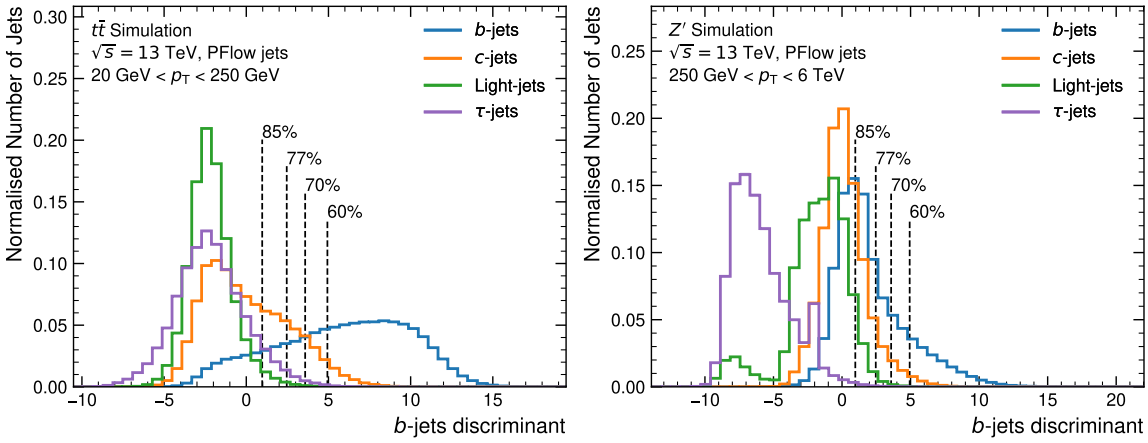


Figure 10.12.: Fraction value scan for DIPS Tau and DIPS. Depicted are the light-flavour and τ-jets rejections as a function of the fraction values at the 70 % WP. The curves here were calculated by changing the values of f_c and f_{light} by 0.005 with $f_\tau = 0$ in each step, ranging from zero to one. The chosen fraction values for both DIPS Tau and DIPS are marked with explicit markers.

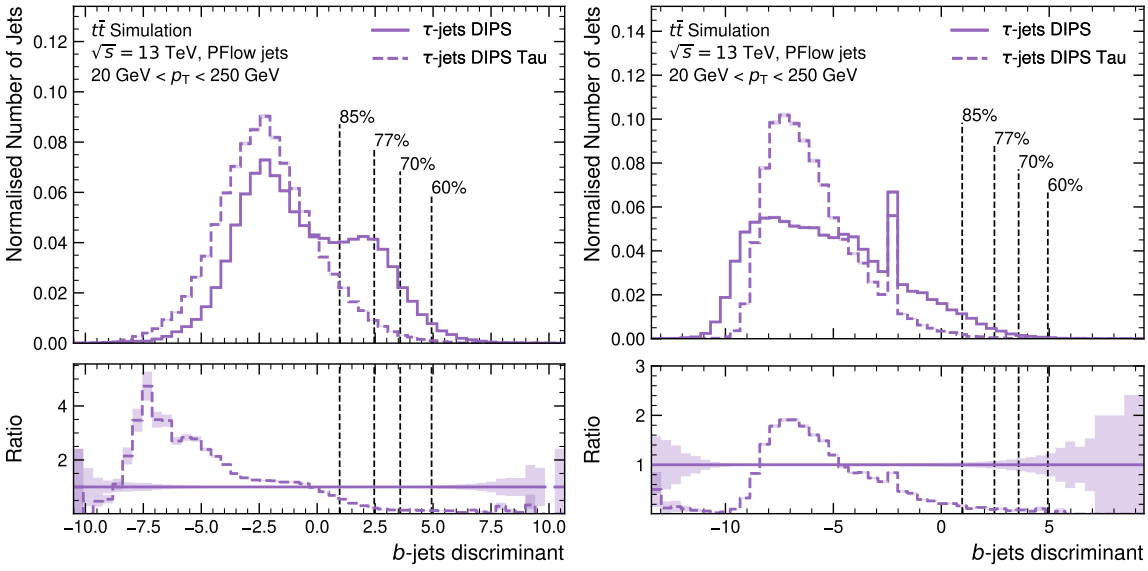
In both test samples, there is a clear separation among the different jet classes. As anticipated, the largest separation appears in the $t\bar{t}$ test sample, but there is also a discernible seperation between the τ-jets and other jet classes in the Z' test sample. Moreover, a consistent separation of b-jets from other jet classes is observed. Figures 10.14 (a) and 10.14 (b) portray both the $t\bar{t}$ and Z' test samples, respectively, showcasing the distributions for both τ-jets in DIPS and in DIPS Tau.

A clear and distinct improvement is evident in both test samples, highlighting the significantly better performance of DIPS Tau over DIPS in terms of τ-jets rejection.



(a): $t\bar{t}$ (b): Z'

Figure 10.13.: b-tagging discriminant distributions of the DIPS Tau tagger for the $t\bar{t}$ (a) and Z' (b) test samples. The WPs shown in the both plots were derived on the $t\bar{t}$ test sample.

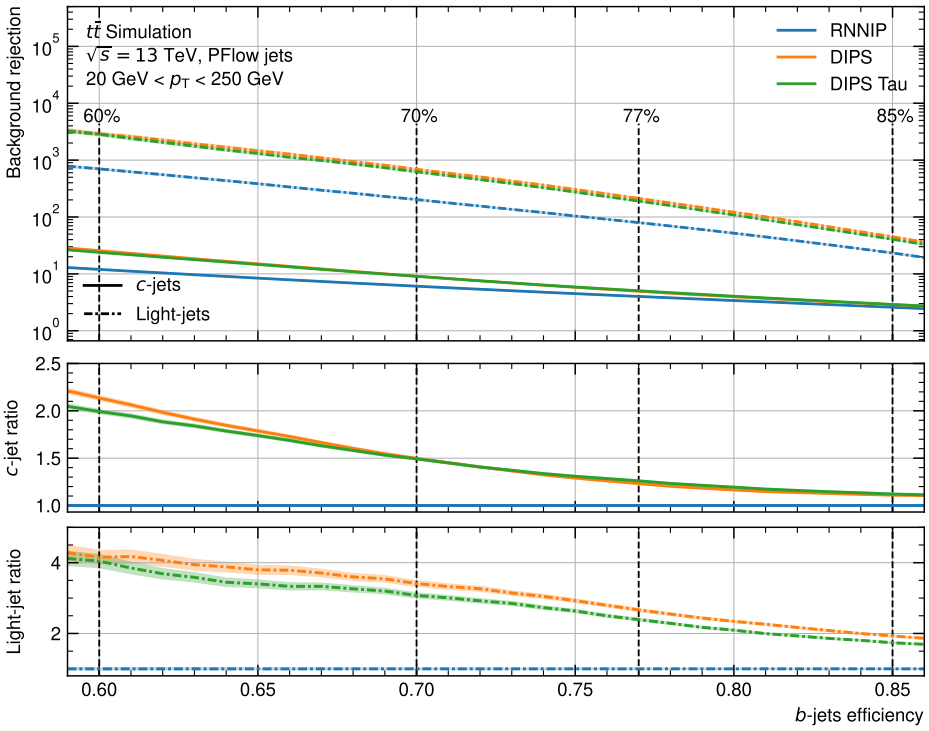


(a): $t\bar{t}$ (b): Z'

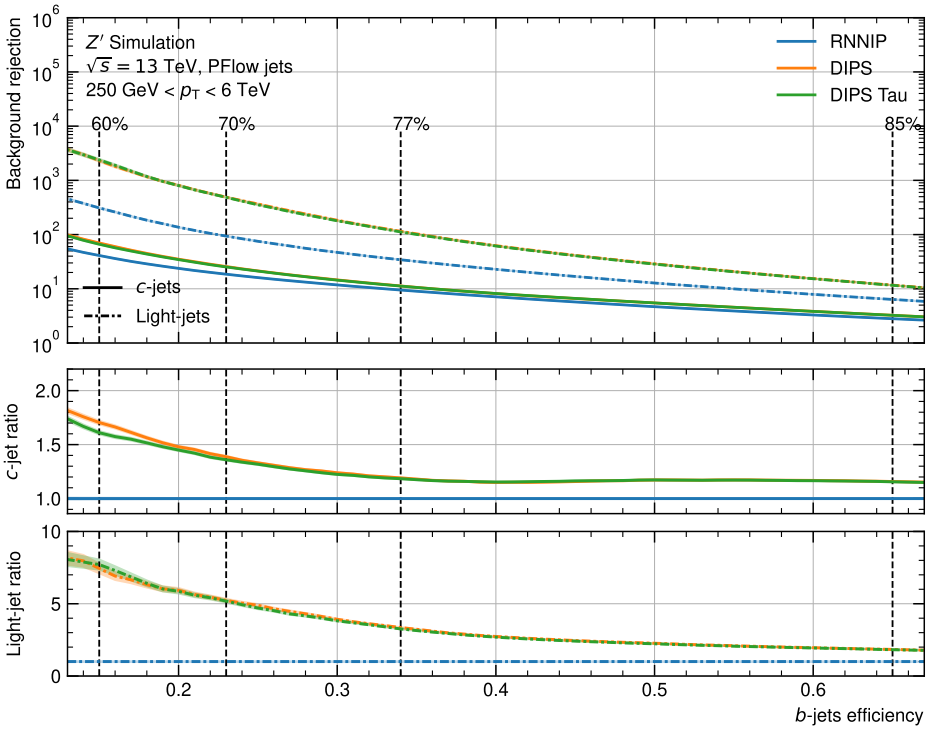
Figure 10.14.: b-tagging discriminant distributions of the DIPS and DIPS Tau taggers for τ-jets for the $t\bar{t}$ (a) and Z' (b) test samples. The WPs shown in both plots were derived on the $t\bar{t}$ test sample. The shaded areas denote only the statistical uncertainties, derived by using binomial uncertainties.

Furthermore, to evaluate the overall performance comprehensively, ROC curves are employed. These curves are presented in Figures 10.15 (a) and 10.15 (b) for the $t\bar{t}$ and Z' test samples respectively, illustrating the rejections of c and light-flavour jets in relation to b-jets efficiency. Additionally, the ROC curves showcasing the τ-jets rejections are displayed in Figures 10.16 (a) and 10.16 (b).

In terms of $t\bar{t}$ light-flavour jets rejection, a minor performance decrease is noticed for DIPS Tau compared to DIPS across the complete b-jets efficiency spectrum. However, there is still a notable improvement in performance when compared to RNNIP. This decline in performance is an anticipated outcome due to the adjustments made in the fraction values. To accommodate the setting of f_τ to 0.2, an equivalent amount was deducted from f_{light}, reducing the significance of the light-flavour

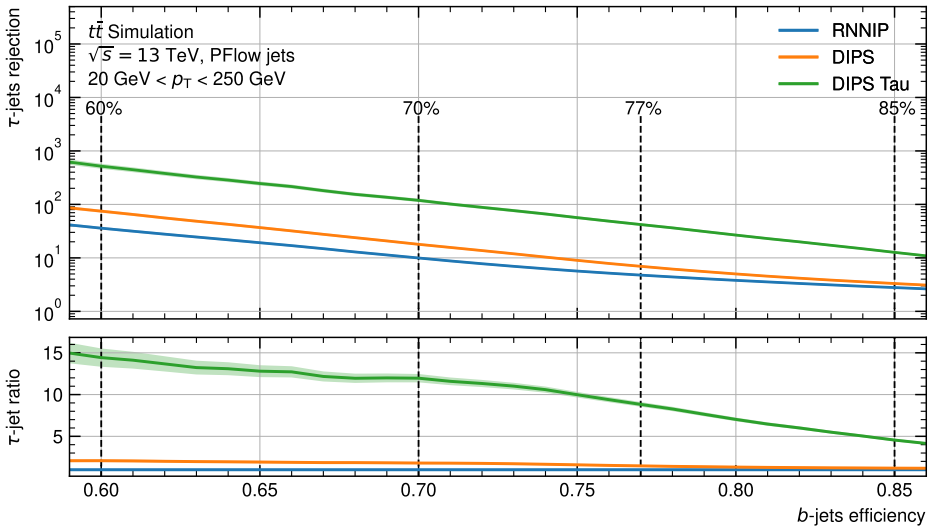


(a): $t\bar{t}$

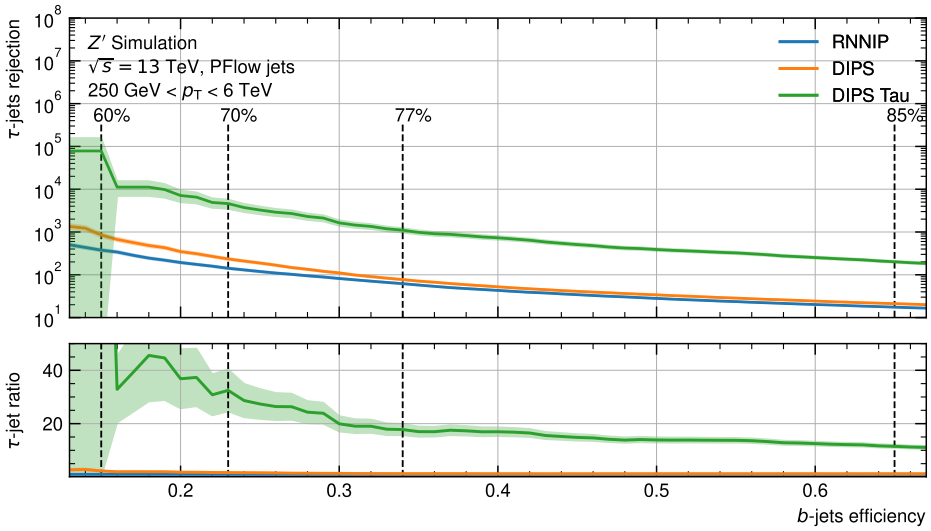


(b): Z'

Figure 10.15.: ROC curves for the DIPS Tau, DIPS and RNNIP tagger. The depicted b-jets efficiencies and c and light-flavour jets rejections for all taggers were calculated using the respective test sample, each with in total three million jets. The shown b-jets efficiencies on the x-axis denote the b-jets efficiency on the respective test sample, while the depicted WPs were derived from the $t\bar{t}$ test sample and correspond to the WPs which would be used in analyses. The shaded areas denote only the statistical uncertainties, derived by using binomial uncertainties.



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



(b): Z'

Figure 10.16.: ROC curves for the DIPS Tau, DIPS and RNNIP tagger. The depicted efficiencies and τ-jets rejections for all taggers were calculated using the respective test sample, each with a total of three million jets. The shown b-jets efficiencies on the x-axis denote the b-jets efficiency on the respective test sample, while the depicted WPs were derived from the $t\bar{t}$ test sample and correspond to the WPs which would be used in analyses. The shaded areas denote only the statistical uncertainties, derived by using binomial uncertainties.

output probability. Regarding c-jets rejection, a slight performance dip is observed in the lower b-jets efficiency regions, while maintaining a comparable c-jets rejection in the higher b-jets efficiency region.

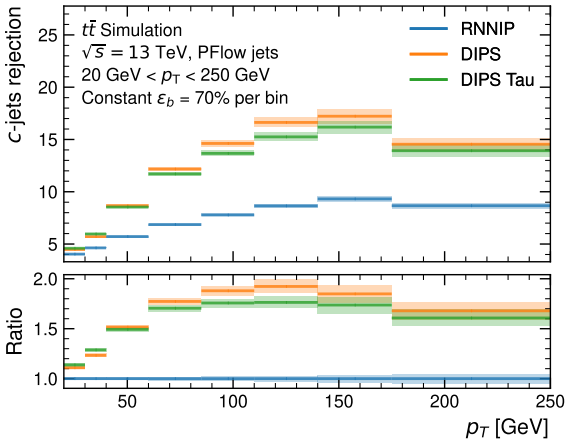
In the case of Z', no significant increase or decrease in light-flavour rejection for DIPS Tau, in comparison to DIPS, is evident. There is a slight decrease observed in c-jets rejection within the lower b-jets efficiency region. In terms of τ-jets rejection, DIPS Tau significantly outperforms both RNNIP and DIPS, showcasing an improvement factor of around 13 for the 70 % WP on the $t\bar{t}$ test sample. A more substantial improvement is noted for Z', with an enhancement factor of up to 30 for the 70 % WP. However, it is important to note that the associated uncertainties are relatively large

due to the limited number of τ-jets still passing the WPs. Despite a minor decrease in light-flavour jets rejection, the remarkable gains in terms of τ-jets rejection demonstrate a significant enhancement in overall performance for DIPS Tau over both DIPS and RNNIP.
For further assessment, rejections are plotted as functions of jet p_{T}, utilising the inclusive and flat efficiency per bin procedures. The respective plots for the $t\bar{t}$ and Z' test samples are presented in Figure 10.17.

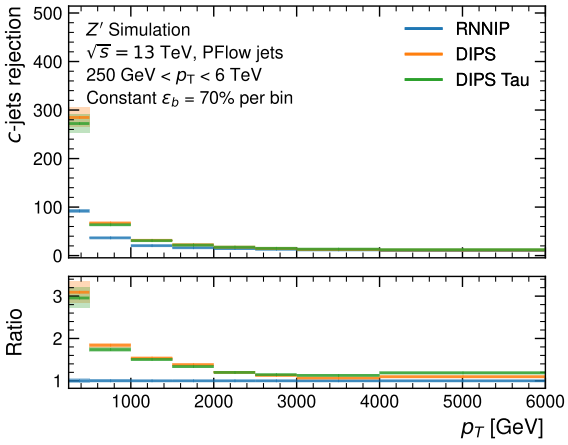
There is a minor observed decrease in performance for both $t\bar{t}$ c and light-flavour jets rejections in DIPS Tau compared to DIPS. However, noteworthy improvement factors of up to 25 in τ-jets rejection are found for the $t\bar{t}$ test sample. For the Z' sample, even more substantial improvement factors in τ-jets rejection are noted, reaching up to a factor of 60. It is important to note that in some jet p_{T} bins in the Z' τ-jets rejection plot, there are no entries for DIPS Tau. This occurs when no τ-jet in that particular jet p_{T} bin passes the specified WP, resulting in an infinite rejection.
In the inclusive procedure, the b-jets efficiency and rejections for other jet classes are illustrated in Figures 10.18 and 10.19 for the $t\bar{t}$ and Z' test samples, respectively.

In the inclusive procedure for $t\bar{t}$, a noticeable trend towards the higher jet p_{T} range is observed for DIPS Tau concerning b-jets efficiency. Conversely, the $t\bar{t}$ c-jets rejection demonstrates a shift towards the lower jet p_{T} range, showing decreased performance in higher jet p_{T} bins. Similarly to the observations in the flat efficiency per bin procedure, only marginal decreases in $t\bar{t}$ light-flavour jet rejections are detected, accompanied by a pronounced increase in τ-jets rejection.

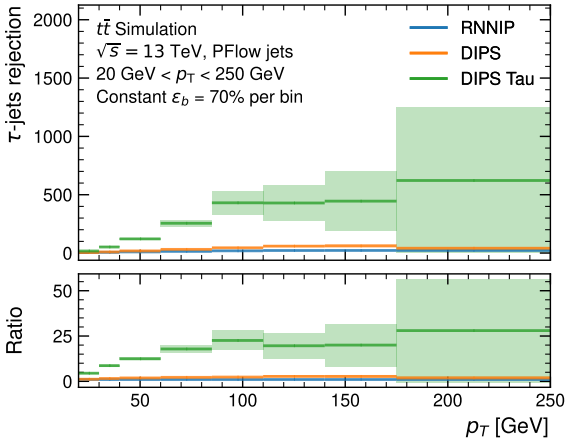
In the Z' test sample, a more pronounced trend is noticed in b-jets efficiency, indicating a shift towards higher jet p_{T} bins compared to DIPS. Regarding c-jets rejection, an enhancement in performance is observed in higher jet p_{T} bins, juxtaposed with a slight decrease in lower p_{T} bins. Similar trends are observed in light-flavour jets rejection, while τ-jets rejection exhibits significant enhancements, especially in the initial three bins up to 1.5, TeV, showing improvements of up to 40. However, in the higher jet p_{T} bins, no evaluation of τ-jets rejection is feasible due to the absence of τ-jets meeting the WP, resulting in infinite rejections in these bins.
The new DIPS Tau tagger has surpassed expectations in τ-jets rejection while maintaining acceptable levels of rejection for c and light-flavour jets. Surprisingly, even with the τ-jets output probabilities masked ($f_\tau = 0$), the performance in light-flavour jets rejection improves compared to the standard DIPS. This indicates that incorporating τ-jets in the training offers benefits, even without using the τ-jets output probabilities.



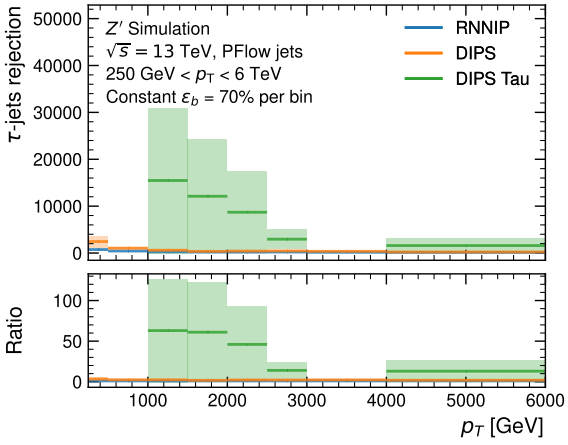
(a): $t\bar{t}$ c-jets rejection



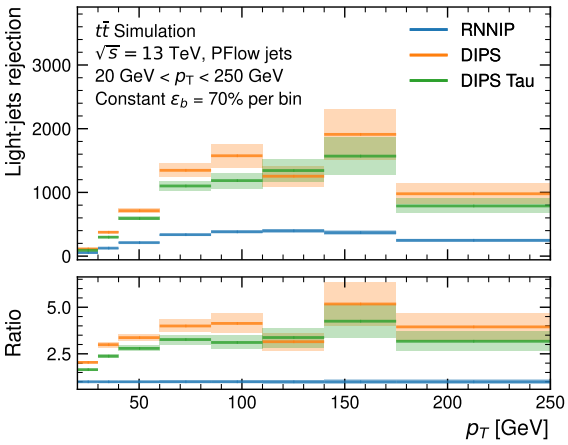
(b): Z' c-jets rejection



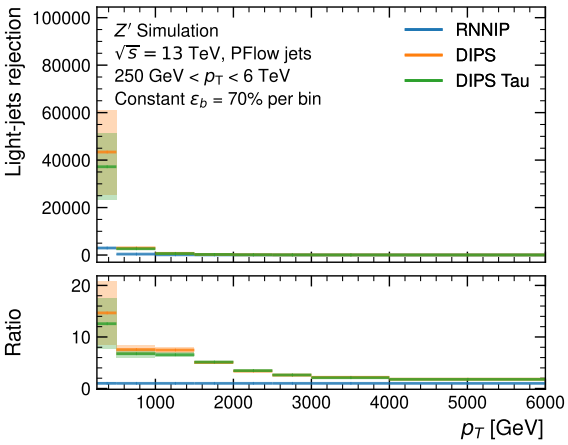
(c): $t\bar{t}$ τ-jets rejection



(d): Z' τ-jets rejection

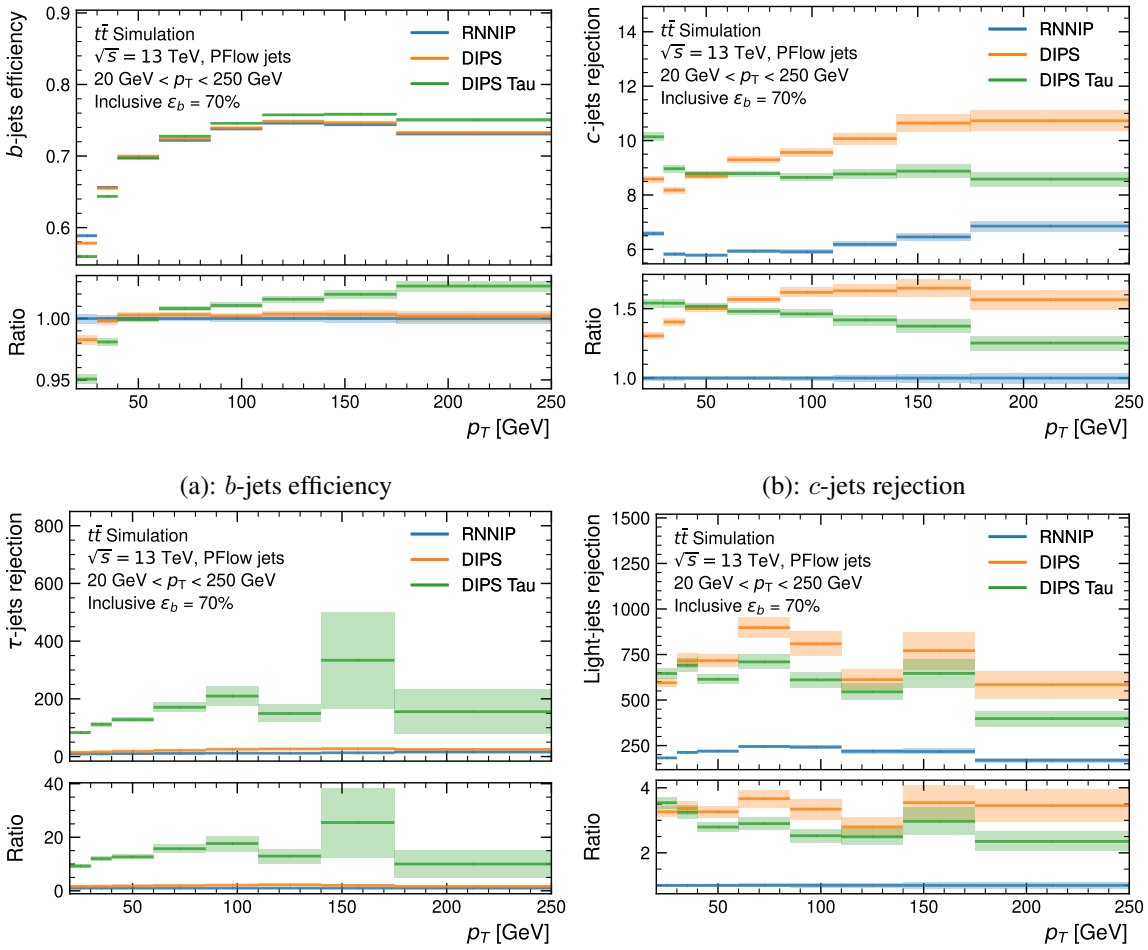


(e): $t\bar{t}$ light-flavour jets rejection



(f): Z' light-flavour jets rejection

Figure 10.17.: c, τ, and light-flavour jets rejections for the $t\bar{t}$ and Z' test samples for the 70 % WP using the flat efficiency per bin procedure for the DIPS Tau, DIPS and RNNIP taggers. For both test samples, $t\bar{t}$ and Z', the same WP was used with the same cut value on the b-tagging discriminant, which was derived using the $t\bar{t}$ test sample.



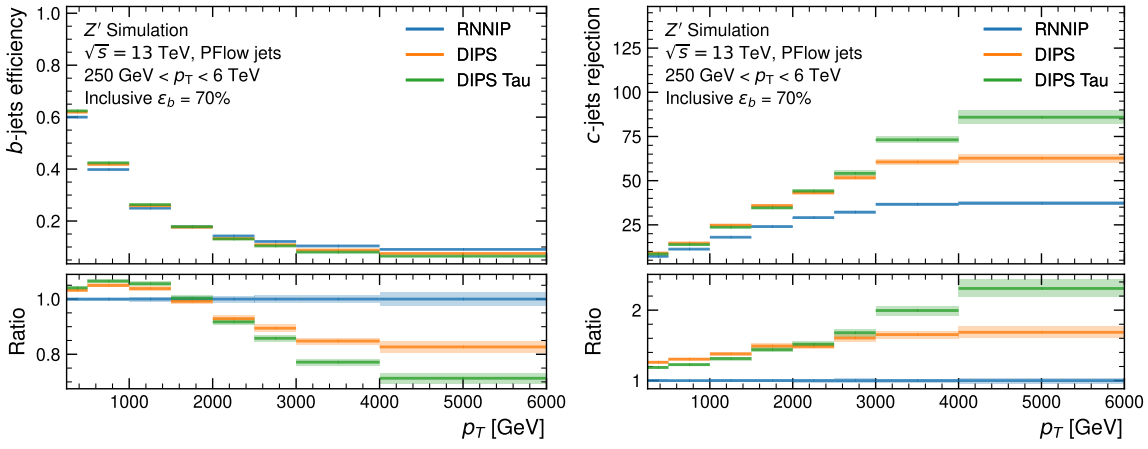
(a): b-jets efficiency

(b): c-jets rejection

(c): τ-jets rejection

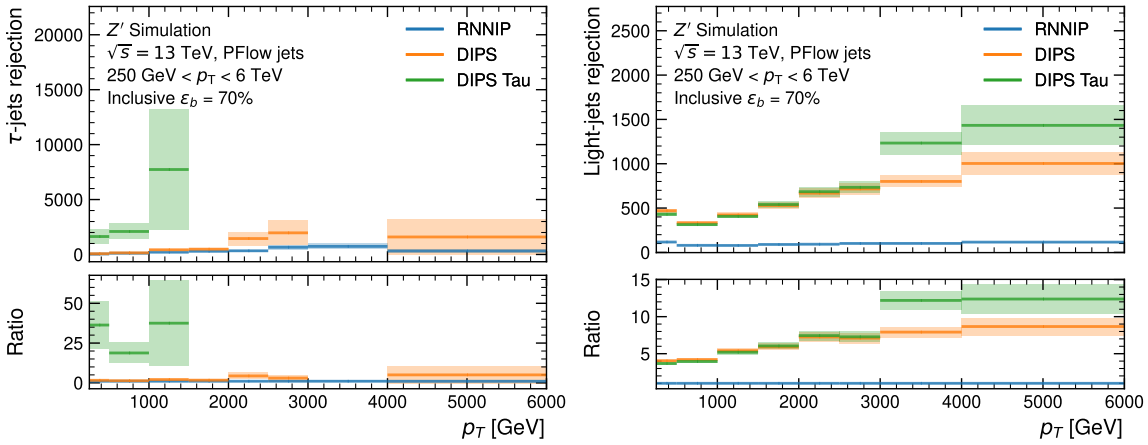
(d): Light-flavour jets rejection

Figure 10.18.: b-jets efficiency and c, τ, and light-flavour jets rejections for the $t\bar{t}$ test sample for the inclusive 70 % WP for the DIPS Tau, DIPS and RNNIP taggers. The WP used was derived on the same sample.



(a): b-jets efficiency

(b): c-jets rejection



(c): τ-jets rejection

(d): Light-flavour jets rejection

Figure 10.19.: b-jets efficiency and c, τ, and light-flavour jets rejections for the Z' test sample for the inclusive $70\,\%$ WP for the DIPS Tau, DIPS and RNNIP taggers. The WP used was derived on the $t\bar{t}$ test sample.

Part IV

MEASUREMENT OF THE $t\bar{t}H$ PRODUCTION CROSS-SECTION IN THE $H \to b\bar{b}$ DECAY CHANNEL

11

ANALYSIS STRATEGY AND OVERVIEW

The second major topic of this thesis involves the measurement of the production cross-section of the associated production of a t-quark pair and a Higgs boson ($t\bar{t}H$) in the final state, where the Higgs boson decays into a pair of b-quarks ($t\bar{t}H(H \to b\bar{b})$). The final state where the Higgs boson decays into a pair of b-quarks is employed due to its dominant branching ratio over other Higgs decay modes (as depicted in Figure 7.1), which maximises the potential for accurately measuring the elusive $t\bar{t}H$ process[1]. Although predicted to be relatively rare with a production cross-section of 507^{+35}_{-50} fb [15] (around $1\,\%$ of the total Higgs cross-section), the $t\bar{t}H$ process offers a direct measurement of the Yukawa coupling of the t-quark to the Higgs boson. Given the substantial mass of the t-quark and its large expected coupling to the Higgs boson, any deviation from the anticipated value could indicate the existence of new particles or interactions. Therefore, with a precise determination of this coupling, various BSM theories could be probed and tested, which include possible explanations for many of the unresolved problems of the SM, e.g. the fermion mass hierarchy. Furthermore, a more precise estimation of the metastability of the EW vacuum [196–198] can be achieved.
The analysis, referred to as $t\bar{t}H(H \to b\bar{b})$ legacy, uses the complete Run 2 dataset collected by the ATLAS experiment. The term 'legacy' indicates that this analysis constitutes the re-analysis of the full dataset and involves the adoption of new, optimised, and most performant techniques along with refined object definitions, all aiming at enhancing the precision of measuring the $t\bar{t}H(H \to b\bar{b})$ process. Furthermore, the $t\bar{t}H(H \to b\bar{b})$ final states not only facilitate an overall measurement but also enable a cross-section measurement as a function of the reconstructed Higgs p_{T} (p_{T}^{H}) using the STXS formalism.

This chapter provides an overview of the previous analysis results from both ATLAS and CMS and introduces the structure and strategies used in the ATLAS legacy analysis. It delves deeper into the novel techniques employed, offering insights into the analysis approach and justifying the subsequent studies detailed in the following chapters. Furthermore, alongside an exploration of the analysis uncertainties, a brief introduction to Profile-Likelihood Fits (PLFs) will be provided, given their central role as the primary statistical tool in this analysis.

11.1 PREVIOUS MEASUREMENTS FROM ATLAS AND CMS

The $t\bar{t}H(H \to b\bar{b})$ process was already measured twice by the ATLAS [14, 199] and CMS [200, 201] experiments using the partial or full Run 2 dataset recorded by them between 2015 and 2018.

1 While the $t\bar{t}H$ process was already discovered in the combination of multiple Higgs decay channels [195], the direct discovery in the $H \to b\bar{b}$ channel is not yet achieved.

The newest measurements available were performed using the full Run 2 dataset, corresponding to an integrated luminosity of $139\,\text{fb}^{-1}$ ($138\,\text{fb}^{-1}$) for ATLAS (CMS) at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$. Both experiments have measured the $t\bar{t}H$ signal strength inclusively as well as in terms of p_{T}^{H} using the STXS formalism.

11.1.1 *Inclusive Measurements*

The ATLAS analysis [14] focused on two different final states of the $t\bar{t}H(H \to b\bar{b})$ process, which differ by the subsequent decay of the $W^{\pm}$ bosons from the decays of the t-quarks. If one $W^{\pm}$ boson decays hadronically and one leptonically, the respective channel is referred to as the single-lepton channel, while if both $W^{\pm}$ bosons decay leptonically, the channel is referred to as the dilepton channel. The case of both $W^{\pm}$ bosons decaying hadronically is not considered, due to the large expected QCD-induced background in this channel. Based on these channels, the analysis is constructed around an invested event categorisation based on the previously mentioned channels, number of jets, number of b-jets, and number of boosted Higgs boson candidates. Making use of BDTs for classification and reconstruction of the events, the analysis was optimised to achieve the highest possible sensitivity when measuring the signal strength. The final measurement of this signal strength yielded $\mu_{t\bar{t}H} = 0.35^{+0.36}_{-0.34}$ with an observed (expected) significance of 1.0 (2.7) standard deviations. This combined measurement, together with the signal strengths of the respective channels, is depicted in Figure 11.1.

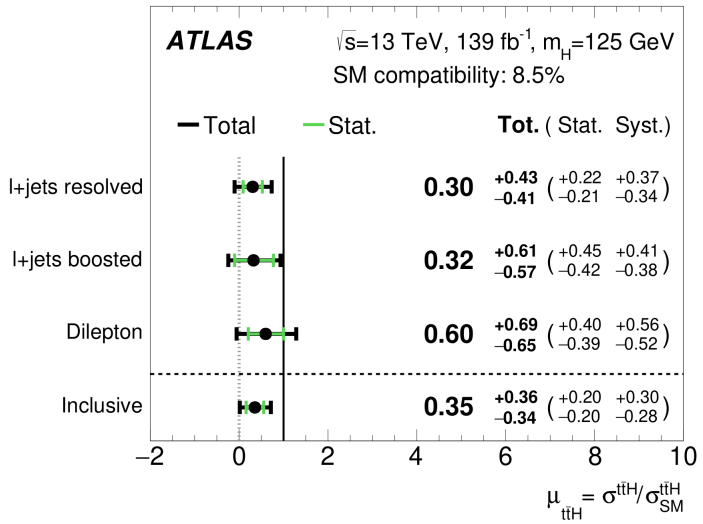


Figure 11.1.: Measured $t\bar{t}H$ signal strength parameters $\mu_{t\bar{t}H}$ for the single-lepton (l+jets) and dilepton channels and the combination of those measured by ATLAS [14].

The high overall uncertainty on the result is dominated by systematic uncertainties, mainly arising from the modelling of the $t\bar{t} + \geq 1b$ contributions. The full breakdown of uncertainties and their effect on $\mu_{t\bar{t}H}$ are shown in Figure 11.1.

Table 11.1.: Contributions to the uncertainties of the measurement of the signal strength $\mu_{t\bar{t}H}$ by ATLAS [14]. The overall uncertainty is dominated by systematic contributions, which largest contributor is the modelling of the $t\bar{t} + \geq 1b$ process.

Uncertainty group	Uncertainty source	$\Delta\mu$	
Process modelling	$t\bar{t}H$ modelling	$+0.13$	-0.05
	$t\bar{t} + \geq 1b$ NLO matching	$+0.21$	-0.20
	$t\bar{t} + \geq 1b$ fractions	$+0.12$	-0.12
	$t\bar{t} + \geq 1b$ FSR	$+0.10$	-0.11
	$t\bar{t} + \geq 1b$ PS and hadronisation	$+0.09$	-0.08
	$t\bar{t} + \geq 1b$ p_{T}^{bb}	$+0.04$	-0.04
	$t\bar{t} + \geq 1b$ ISR	$+0.04$	-0.04
	$t\bar{t} + \geq 1c$ modelling	$+0.03$	-0.04
	$t\bar{t}$ + light modelling	$+0.03$	-0.03
	tW modelling	$+0.08$	-0.07
b-tagging efficiency and mis-tag rates	b-tagging efficiency	$+0.03$	-0.02
	c-mis-tag rates	$+0.03$	-0.03
	Light-flavour mis-tag rates	$+0.02$	-0.02
JES and JER	b-jet energy scale	$+0.00$	-0.01
	JES (flavour)	$+0.01$	-0.01
	JES (pile-up)	$+0.00$	-0.01
	JES (remaining)	$+0.01$	-0.01
	JER	$+0.02$	-0.02
Luminosity		$+0.01$	-0.00
Other sources		$+0.03$	-0.03
Total systematic uncertainty		$+0.30$	-0.28
$t\bar{t} + \geq 1b$ normalisation		$+0.04$	-0.07
Background-model statistical uncertainty		$+0.04$	-0.05
Total statistical uncertainty		$+0.20$	-0.20
Total uncertainty		$+0.36$	-0.34

The corresponding CMS analysis [201] was performed slightly different to the ATLAS measurement, as not only the $t\bar{t}H$ process but also the tH process was measured. While ATLAS only makes use of channels with leptons in the final state, to not be influenced by the QCD background, CMS also uses the full hadronic channel, where both $W^{\pm}$ bosons from the t-quark decay hadronically. For each of the channels, selection criteria are defined, similar to ATLAS, based on the number of jets and b-jets in the final state. Afterwards, reconstruction and classification are performed using state-of-the-art ML techniques, like transformers. CMS also measures $\mu_{t\bar{t}H}$ in each of the considered channels/years as well as in combination, which was measured to be $\mu_{t\bar{t}H} = 0.33^{+0.26}_{-0.26}$ with an observed (expected) significance of 1.3 (4.1) standard deviations. The results are shown in Figure 11.2.

Again similar to ATLAS, the measurement is limited by systematic uncertainties, where the most dominant contributions come from the modelling of the $t\bar{t} + \geq 1b$ process. Both measurements from

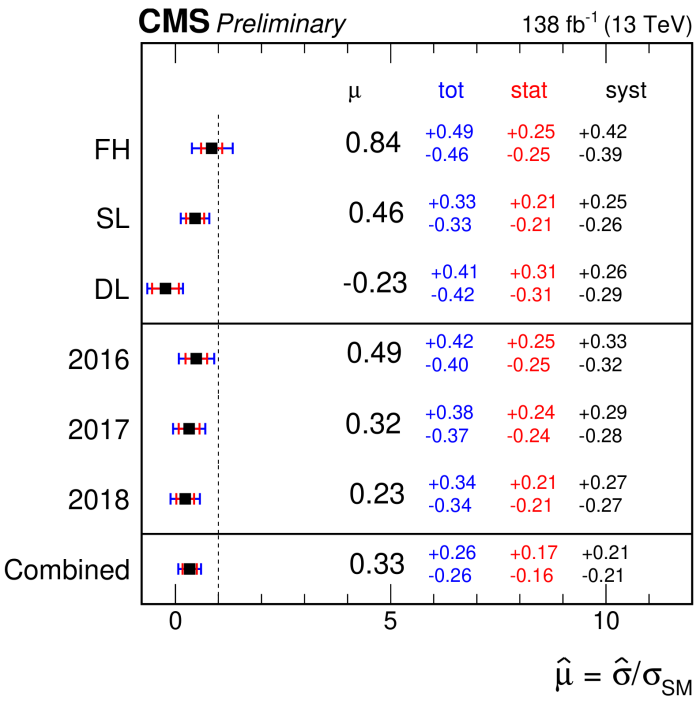


Figure 11.2.: Measured $t\bar{t}H$ signal strength parameters $\mu_{t\bar{t}H}$ for the full-hadronic (FH), single-lepton (SL), and dilepton (DL) channels as well as for the different years and the overall combination measured by CMS [201].

ATLAS and CMS show a similar central value with slightly smaller uncertainties for the CMS measurement but both below the SM prediction. This is due to the different treatment of certain uncertainties in CMS and ATLAS. While ATLAS takes a more conservative approach towards modelling uncertainties in the signal and dominant background processes, taking into account various variations and possible uncertainties, CMS has overall smaller modelling uncertainties due to choosing a different systematic model where many of the variations used by ATLAS are evaluated and taken into account differently. This difference in uncertainty treatment leads to a difference in the sensitivity of both measurements.

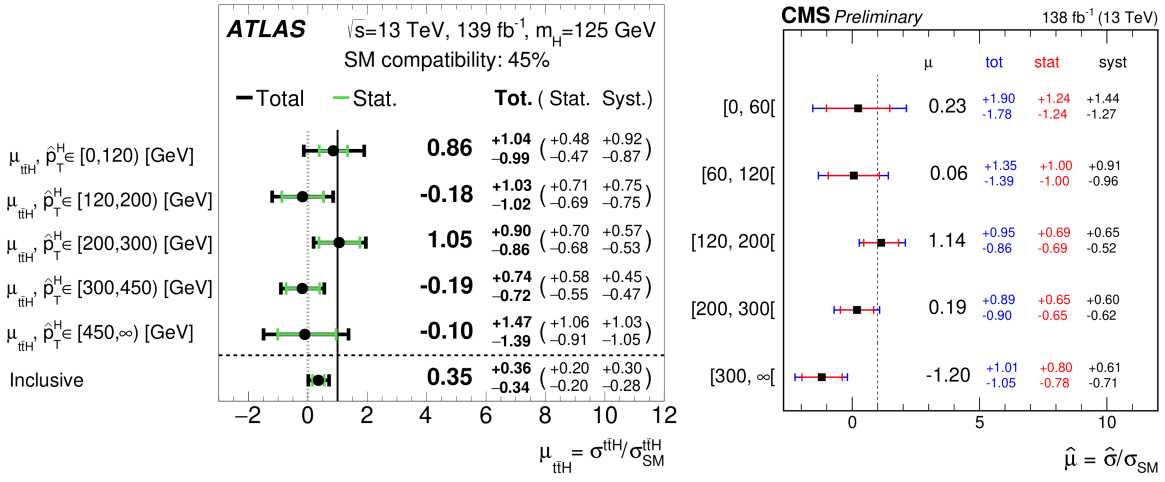
11.1.2 *Simplified Template Cross-Section Measurements*

As already mentioned, both experiments also measured the signal strength in terms of p_{T}^{H} using the STXS framework [15, 16], which was developed by some of the LHC experiments in cooperation with theorists. It provides the possibility to measure the signal strength of certain processes in exclusive kinematic phase space regions, which allow a straightforward combination of the results from different experiments and different channels. These phase space regions, also called STXS bins, are designed to reduce theory modelling dependences, which are folded into the measurements. For most of the Higgs-related measurements, this framework can be applied for an easy combination. Since most of the Higgs decay channels can be fully reconstructed and therefore the p_{T}^{H} can be measured, the STXS bins are defined as a function of p_{T}^{H}. Usually, to ensure compatible results, the so-called

truth p_{T}^{H} ($\hat{p}_{\mathrm{T}}^{H}$) is used for the classification of the MC prediction over the reconstructed Higgs p_{T} to be independent of the reconstruction. For the measurement of $t\bar{t}H$, the commonly agreed on STXS bins are defined as depicted in Table 11.2.

Table 11.2.: Commonly by the different LHC experiments agreed on STXS bins for measurements of the $t\bar{t}H$ process.

Bin	p_{T}^{H} range
STXS 1	$0 < p_{\mathrm{T}}^{H} \leq 60\,\mathrm{GeV}$
STXS 2	$60 < p_{\mathrm{T}}^{H} \leq 120\,\mathrm{GeV}$
STXS 3	$120 < p_{\mathrm{T}}^{H} \leq 200\,\mathrm{GeV}$
STXS 4	$200 < p_{\mathrm{T}}^{H} \leq 300\,\mathrm{GeV}$
STXS 5	$300 < p_{\mathrm{T}}^{H} \leq 450\,\mathrm{GeV}$
STXS 6	$450 < p_{\mathrm{T}}^{H} \leq \infty\,\mathrm{GeV}$



(a): ATLAS [14]. (b): CMS [201].

Figure 11.3.: Measurements of the $t\bar{t}H$ signal strength parameters in terms of p_{T}^{H}. The results of ATLAS are given in terms of $\hat{p}_{\mathrm{T}}^{H}$ and also the inclusive result is shown. For the CMS result, the last two STXS bins are merged.

While the lower STXS bins can be measured more precisely in other decay channels of the Higgs boson, namely $t\bar{t}H\,(H \to \gamma\gamma)$, the $t\bar{t}H(H \to b\bar{b})$ channel allows for better measurement of the higher STXS bins, due to the increased statistics for this process over $t\bar{t}H\,(H \to \gamma\gamma)$ and the usage of reclustered jets[2].

The results of the STXS measurements of both ATLAS and CMS are depicted in Figures 11.3 (a) and 11.3 (b), respectively. While ATLAS measures the first two STXS bins combined as one due to the large migrations between these bins, CMS measures the last two bins combined as one due to a lack in statistics. While the measurement strategies slightly differ, an overall trend of $\mu < 1$ can be observed for both measurements.

2 Reclustered jets are built using the jet reclustering technique from Ref. [202] resulting in fully calibrated jets with $R = 1.0$.

11.2 STRATEGY AND OVERVIEW OF THE LEGACY ANALYSIS

The legacy measurement conducted by the ATLAS experiment represents a refined re-analysis of the previously mentioned full Run 2 analysis. New techniques, concerning object definition and event classification, are employed to further enhance the sensitivity of the measurement. Similar to the preceding analysis, the legacy analysis concentrates on two of the three possible channels, defined by the decay modes of the subsequent $W^{\pm}$ bosons from the $t\bar{t}$ decay.

The first and highest channel in terms of branching ratio ($45.7\,\%$) is the full hadronic channel, where both $W^{\pm}$ bosons decay hadronically. Despite having the highest branching ratio, this channel is the most challenging to measure due to the overwhelming QCD-induced multi-jet background. Therefore, similar to the previous analysis, this channel is not considered in the measurement and is excluded.

The second channel is the single-lepton channel, where one of the two $W^{\pm}$ bosons decays hadronically and the other leptonically. This channel is easier to distinguish from the background due to the presence of a final-state lepton and has a similarly high branching ratio as the full hadronic channel ($43.8\,\%$). With only one neutrino in the final state, determinable from $E_{\mathrm{T}}^{\mathrm{miss}}$, this channel provides a clean topology for reconstruction and measurement. Due to the relatively large branching ratio, a higher signal-to-background ratio and reasonable statistics are expected even in the higher-p_{T} regime, which is exploited by further segregating this channel into a resolved and a boosted regime. Thanks to this split, it is possible to employ the already-mentioned reclustered jets, which can be used to capture both b-quarks from the Higgs decay in one jet[3].

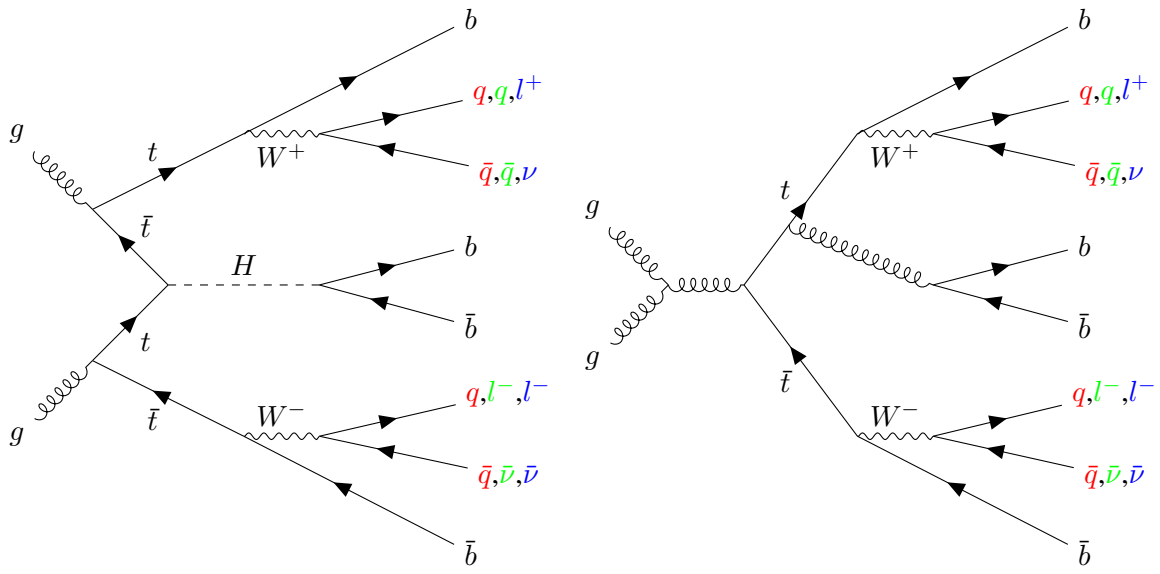
The last channel is the dilepton channel, where both $W^{\pm}$ bosons decay leptonically. With the smallest branching ratio of $10.5\,\%$ and two neutrinos in the final state, this channel is not optimal for reconstruction and measurement. Although relatively small and challenging to reconstruct, the dilepton channel is considered due to the presence of two leptons in the final state, effectively suppressing contributions from multi-jet background and offering the cleanest signature among all channels. Similar to the single-lepton channel, this channel is also divided into a resolved and a boosted regime.

A Feynman diagram representing all three final states is presented in Figure 11.4 (a). Given the overall complex topology of the considered final states, one of the major challenges is the large, irreducable background from $t\bar{t} + \mathrm{jets}$ events, where the additional jets are identified as b-jets ($t\bar{t} + b\bar{b}$). This specific background shares the same final state as the signal process and entails significant theoretical modelling uncertainties, impacting the final measurement. The corresponding Feynman diagram is depicted in Figure 11.4 (b).

11.2.1 *Event Selection And Categorisation*

For the single-lepton channel, a total of six jets are anticipated, with four originating from final-state b-quarks, exactly one lepton, and $E_{\mathrm{T}}^{\mathrm{miss}}$. In the dilepton channel, a total of four jets are expected, with

3 In the high-p_{T} regime, the two b-quarks from the Higgs decay are usually very close together in terms of ΔR which enables the usage of jets with $R = 1.0$.



(a): $t\bar{t}H$ (b): $t\bar{t}+b\bar{b}$

Figure 11.4.: Representative Feynman diagrams of the three main decay channels of the $t\bar{t}H(H \to b\bar{b})$, (a), and the $t\bar{t}+b\bar{b}$, (b), processes. This includes the full-hadronic (red), single-lepton (green), and dilepton (blue) channels.

all four stemming from final-state b-quarks, two opposite-sign leptons, and higher $E_{\mathrm{T}}^{\mathrm{miss}}$ compared to the single-lepton channel.

Based on these characteristics of the final states, dedicated triggers and channel-specific pre-selections are chosen. Subsequently, ML-based reconstruction and classification algorithms are utilised to categorise the events and reconstruct the p_{T}^{H}.

For both channels, the lowest un-prescaled single-lepton triggers are employed to identify candidate events. The dedicated triggers are combined with a logical "OR", ensuring a higher selection efficiency. The triggers used are outlined in Table 11.3 and categorised by lepton type and the year of data-taking. Events with the specified lepton type must either meet a low lepton p_{T} threshold and stricter lepton identification/isolation WPs or a higher lepton p_{T} threshold and more relaxed lepton identification/isolation WPs[4].

Only events with electrons and muons are taken into consideration, while events containing one or more (two or more) τ_{had} candidates, are vetoed in the single-lepton (dilepton) channel to maintain orthogonality with other $t\bar{t}H$ analyses.

After the events are selected by the trigger, objects are reconstructed as described in Chapter 5. For the leptons, special requirements need to be satisfied. Reconstructed electrons are required to have $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ and $|\eta| < 2.47$, excluding the transition region of the barrel to forward ECal ($1.37 < |\eta| < 1.52$). In addition, electrons need to pass the tight LH identification WP and the tight isolation WP. Reconstructed muons are required to have $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ and $|\eta| < 2.5$ and pass the medium identification and tight isolation WPs. To avoid any double-counting of a single detector re-

4 Note here that the leptons considered in the trigger are not as precisely identified and calibrated as the final reconstructed and calibrated leptons used later on in the analysis.

Table 11.3.: List of the triggers and their lepton p_{T} thresholds and isolation criteria for the different data-taking periods. Note here that trigger leptons are used, not the final reconstructed and calibrated leptons used later on in the analysis.

Lepton flavour	p_{T} (GeV)		Identification WP		Isolation WP	
	2015	2016–2018	2015	2016–2018	2015	2016–2018
Electrons	24	26	Medium	Tight	-	Loose
	60	60	Medium	Medium	-	-
	120	140	Loose	Loose	-	-
Muons	20	26	-	-	Loose	Medium
	50	50	-	-	-	-

sponse, an overlap removal procedure is applied. For electrons, the closest jet in $\Delta R_y < 0.2$ around the electron is removed. If the nearest jet is within $0.2 < \Delta R_y < 0.4$, the electron is removed. Muons are removed if a jet is present in $\Delta R_y < 0.4$ around the respective muon to avoid muons from heavy-flavour decays. However, if the jet has less than three tracks, the muon is kept and the jet is removed. τ_{had} candidates are removed when within $\Delta R_y < 0.2$ any electron or muon is found.
After the object reconstruction stage, a dedicated pre-selection for each channel is applied, rejecting events with drastically different final states. In the single-lepton resolved channel, a minimum of five jets are required, with at least three of them tagged as b-jets using the $70\,\%$ DL1r WP. Additionally, exactly one lepton is required, either a muon or an electron with at least $p_{\mathrm{T}} > 27\,\mathrm{GeV}$, aligning with the trigger lepton within $\Delta R = 0.15$.
For the single-lepton boosted channel, at least four jets are required, with at least three of them tagged as b-jets using the $85\,\%$ DL1r WP. Furthermore, at least one reclustered jet is required with $p_{\mathrm{T}} \geq 200\,\mathrm{GeV}$ and $m_{\mathrm{RC\,Jet}} \geq 50\,\mathrm{GeV}$.
In the dilepton resolved channel, a minimum of three jets are required, with at least three of them tagged as b-jets using the $85\,\%$ DL1r WP. Additionally, two of the jets must be tagged as b-jets using the $70\,\%$ DL1r WP. Exactly two opposite sign leptons are required. If the two leptons have the same flavour, the invariant mass of the leptons needs to be above $15\,\mathrm{GeV}$ and outside of the Z-boson mass window of $83 - 99\,\mathrm{GeV}$.
The dilepton boosted channel has similar requirements as the dilepton resolved channel but necessitates at least one additional reclustered jet with $p_{\mathrm{T}} \geq 200\,\mathrm{GeV}$ and $m_{\mathrm{RC\,Jet}} \geq 50\,\mathrm{GeV}$.

With these pre-selections applied to their respective channels, the data sample is mainly dominated by $t\bar{t} + \mathrm{jets}$ events, which also includes the main background, $t\bar{t} + b\bar{b}$. To increase the signal-to-background ratio, the events are further classified into so-called Signal Regions (SRs) and Control Regions (CRs). While the SRs should contain most of the signal events and the least amount of background events possible, maximising the signal-to-background ratio, the CRs are used to provide stringent constraints on the normalisation of the different backgrounds. These resulting regions are treated equally and fitted simultaneously in a combined fit to achieve the best possible estimate of the Parameters of Interest (POIs).

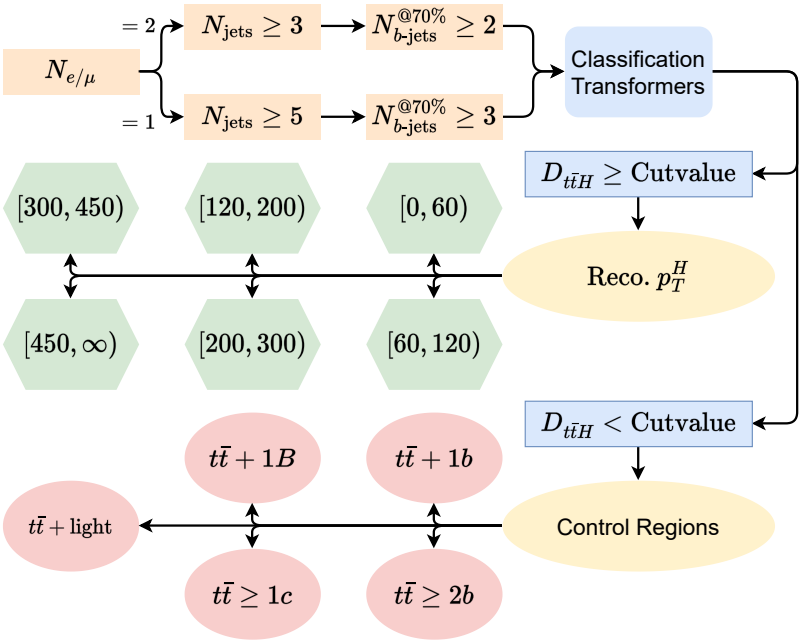


Figure 11.5.: Schematic representation of the pre-selection and the classification into resolved SRs and CRs based on the classification transformers.

The resolved channels have one $t\bar{t}H$ SR and multiple CRs, one for each $t\bar{t} + \text{jets}$ sub-category: $t\bar{t} + b$, $t\bar{t} + B$, $t\bar{t} + \geq 2b$, $t\bar{t} + \geq 1c$, and $t\bar{t} + \text{light}$. While $t\bar{t} + b$ describes $t\bar{t}$ events with one additional b-jet, $t\bar{t} + B$ describes events where the two additional b-jets are merged and reconstructed as one jet. $t\bar{t} + \geq 2b$ simply describes the events with at least two additional b-jets, and $t\bar{t} + \geq 1c$ describes events with additional c-jets. The final sub-category, $t\bar{t} + \text{light}$, describes all other events where the additional jets are light-flavour jets[5]. In addition to this split in SR and CRs, the SR is further split into the already introduced STXS bins based on the reconstructed p_{T}^H.

For the single-lepton boosted channel, the events are separated into either a SR or a CR, and the reconstructed p_{T}^H is used as fitting variable in these regions. No additional split into $t\bar{t} + \text{jets}$ sub-category or STXS bin is applied.

In the dilepton boosted channel, the SR is split into two STXS-based regions, boosted STXS 5 and 6, where the classification is again based on the reconstructed p_{T}^H. No additional CR as for the single-lepton channel is used.

The classification into SRs and CRs is performed by dedicated multi-class classification transformer networks, which predict probabilities that the event is a $t\bar{t}H$ or one of the background events. Combining resolved and boosted channels for training, one transformer is trained for the single-lepton and one for the dilepton channel. A schematic representation showing the pre-selection and the classification into the different resolved SRs and CRs is depicted in Figure 11.5.

After pre-selection, the events are classified with the respective transformer network, using a variety of inputs, including the four-momenta of the jets and the leptons as well as the PCBT scores of the jets and $E_{\text{T}}^{\text{miss}}$. The probability outputs of the network are then combined into different so-called

5 The definition of light-flavour jets follows the explanation given in Figure 7.3.

discriminants, one for the SR[6] and one for each CR. These discriminants, based on the Neyman-Pearson lemma [203] which is also the base for PLFs, are calculated using Equation (11.1), where p_i is the output probability for region i, and N_j is the number of events in sub-category j, which are used to train the network.

$$D_i = \frac{p_i}{\sum_{j \neq i} (p_j \cdot w_{ij})} \quad \text{with} \quad w_{ij} = \frac{N_j}{\sum_{k \neq i} N_k} \tag{11.1}$$

In simpler terms, the discriminants are calculated by dividing the probability that the event is from sub-category i by the sum of the weighted probabilities of all other possible sub-categories. This approach takes into account the imbalance of the training dataset.
Before sorting the events based on these discriminants into the specific SRs and CRs, a pre-sorting is done to determine if the event is going to one of the SRs or one of the CRs. This decision is based on the $t\bar{t}H$ discriminant, $D_{t\bar{t}H}$, using an optimised cut value. The cut value is set to be 4.072 for the single-lepton and 9.031 for the dilepton channels. These cut values are optimised to maximise the signal-to-background ratio while maintaining reasonable statistics in the SRs. If the discriminant value for an event is above the cut value, it is sorted in one of the STXS regions based on p_{T}^{H}, which is reconstructed using a dedicated transformer network. In the case that the event has a discriminant value below the cut value, the event is sorted into the CR, for which the respective discriminant value is the largest. For the single-lepton boosted region, only the pre-sorting is applied while for the dilepton boosted region, the classification into STXS regions is also applied.
After sorting and categorising the events in the different SRs and CRs, a PLF is used to extract the signal strength of the $t\bar{t}H(H \rightarrow b\bar{b})$ process. The variable fitted in each region is the respective class discriminant, except for the boosted regions. While in the single-lepton boosted SR and CR, the boosted reconstructed Higgs p_{T} is fitted, in the dilepton boosted regions, the invariant mass of the leading b-subjet of the leading reclustered jet is used as fitted variable.

11.2.2 *Signal and Background Modelling*

To model the signal and background events, MC generators implemented in the ATLAS simulation software are utilised. The samples generated using these methods are simulated as described in Chapter 4. A table providing an overview of all samples and their settings is shown in Table 11.4.

In the following, to provide more insight about the selected samples and their differences in terms of simulation, additional details are provided on the signal and background modelling, along with the settings used to simulate them.

Signal Modelling

To simulate the $t\bar{t}H$ process, the POWHEG BOX V2 generator [184–186, 204] is employed at NLO with the NNPDF3.0$_{\mathrm{NLO}}$ [105] set of parton distribution functions, using the 5FS. The h_{damp} para-

6 Although multiple SRs are present, only one discriminant is used for all SRs. The final sorting into the STXS regions follows afterwards.

Table 11.4.: List of MC samples used in the $t\bar{t}H(H \to b\bar{b})$ legacy analysis. The nominal samples are always listed first, afterwards the alternative samples are listed, which are used as variations to estimate systematic uncertainties. Here, "nf4" indicates the usage of the 4FS set of parton distribution functions, "pthard" the setting of the pthard parameter in PYTHIA 8 to one, "dipole" the usage of the dipole recoil setting instead of the global recoil setting, and h_{damp} the variation of the respective setting.

Process	ME Generator	ME PDF	PS	Normalisation
$t\bar{t}H$ and tH				
$t\bar{t}H$	POWHEG BOX V2	NNPDF3.0$_{\text{NLO}}$	PYTHIA 8.230	NLO + NLO (EW)
	POWHEG BOX V2	NNPDF3.0$_{\text{NLO}}$	HERWIG 7.04	NLO + NLO (EW)
	POWHEG BOX V2	NNPDF3.0$_{\text{NLO}}$	PYTHIA 8.230 pthard	NLO + NLO (EW)
$tHjb$	MADGRAPH5_AMC@NLO 2.6.2	NNPDF3.0$_{\text{NLO}}$ nf4	PYTHIA 8.230	-
tWH	MADGRAPH5_AMC@NLO 2.6.2 [DR]	NNPDF3.0$_{\text{NLO}}$	PYTHIA 8.235	-
$t\bar{t}+$ jets Background				
$t\bar{t}$	POWHEG BOX V2	NNPDF3.0$_{\text{NLO}}$	PYTHIA 8.230	NNLO + NNLL
	POWHEG BOX V2	NNPDF3.0$_{\text{NLO}}$	HERWIG 7.04	NNLO + NNLL
	POWHEG BOX V2	NNPDF3.0$_{\text{NLO}}$	PYTHIA 8.230 pthard	NNLO + NNLL
	POWHEG BOX V2	NNPDF3.0$_{\text{NLO}}$	PYTHIA 8.230 h_{damp}	NNLO + NNLL
$t\bar{t}+b\bar{b}$	POWHEG BOX RES	NNPDF3.1$_{\text{NLO}}$ nf4	PYTHIA 8.244	-
	POWHEG BOX V2	NNPDF3.1$_{\text{NLO}}$ nf4	HERWIG 7.16	NNLO + NNLL
	POWHEG BOX RES	NNPDF3.1$_{\text{NLO}}$ nf4	PYTHIA 8.244 pthard	-
	POWHEG BOX RES	NNPDF3.1$_{\text{NLO}}$ nf4	PYTHIA 8.244 dipole	-
Single-t Backgrounds				
tW	POWHEG BOX V2 [DR]	NNPDF3.0$_{\text{NLO}}$	PYTHIA 8.230	NNLO + NNLL
	POWHEG BOX V2 [DS]	NNPDF3.0$_{\text{NLO}}$	PYTHIA 8.230	NNLO + NNLL
	POWHEG BOX V2 [DR]	NNPDF3.0$_{\text{NLO}}$	HERWIG 7.04	NNLO + NNLL
t-channel	POWHEG BOX V2	NNPDF3.0$_{\text{NLO}}$ nf4	PYTHIA 8.230	NLO
	POWHEG BOX V2	NNPDF3.0$_{\text{NLO}}$ nf4	HERWIG 7.04	NLO
s-channel	POWHEG BOX V2	NNPDF3.0$_{\text{NLO}}$	PYTHIA 8.230	NLO
Other Backgrounds				
$W+$jets	SHERPA 2.2.1 (NLO [2j], LO [4j])	NNPDF3.0$_{\text{NNLO}}$	SHERPA	NNLO
$Z+$jets	SHERPA 2.2.1 (NLO [2j], LO [4j])	NNPDF3.0$_{\text{NNLO}}$	SHERPA	NNLO
VV (had)	SHERPA 2.2.1	NNPDF3.0$_{\text{NNLO}}$	SHERPA	-
VV (lep)	SHERPA 2.2.2	NNPDF3.0$_{\text{NNLO}}$	SHERPA	-
VV (lep) + jj	SHERPA 2.2.1 (LO [EW])	NNPDF3.0$_{\text{NNLO}}$	SHERPA	-
$t\bar{t}W$	MADGRAPH5_AMC@NLO 2.3.3	NNPDF3.0$_{\text{NLO}}$	PYTHIA 8.210	NLO + NLO (EW)
	SHERPA 2.0.0 (LO [2J])	NNPDF3.0$_{\text{NNLO}}$	SHERPA	NLO + NLO (EW)
$t\bar{t}Z$	MADGRAPH5_AMC@NLO 2.3.3	NNPDF3.0$_{\text{NLO}}$	PYTHIA 8.210	NLO + NLO (EW)
	SHERPA 2.0.0 (LO [2J])	NNPDF3.0$_{\text{NNLO}}$	SHERPA	NLO + NLO (EW)
$t\bar{t}t\bar{t}$	MADGRAPH5_AMC@NLO 2.3.3	NNPDF3.0$_{\text{NLO}}$	PYTHIA 8.210	NLO + NLO (EW)
tZq	MADGRAPH5_AMC@NLO 2.3.3 (LO)	CTEQ6LI	PYTHIA 8.210	-
tWZ	MADGRAPH5_AMC@NLO 2.3.3 [DR]	CTEQ6LI	PYTHIA 8.230	-

meter[7] is set to $3/4 \cdot (m_t + m_{\bar{t}} + m_H) = 352.5\,\text{GeV}$, and the functional form of the renormalisation and factorisation scales, μ_{R} and μ_{F}, is set to $\sqrt[3]{m_{\text{T}}(t) \cdot m_{\text{T}}(\bar{t}) \cdot m_{\text{T}}(H)}$, where m_{T} denotes the transverse mass. The PS and hadronisation are simulated using PYTHIA 8.230 [187].

Two additional samples are simulated to evaluate the impact of the used PS and hadronisation model as well as the choice of matching scheme between the ME and the PS. For the former, POWHEG BOX V2 is used a generator with similar settings as the nominal sample but interfaced with HERWIG 7.04 [205]. For the latter, POWHEG BOX V2 is employed as generator with similar settings as the

7 This parameter controls the p_{T} of the first additional NLO emission in the PS.

nominal sample and interfaced with PYTHIA 8.230 with the pthard parameter set to 1, varying and recalculating the hardness value of POWHEG BOX V2 [206].
All possible final states of $t\bar{t}H$ are simulated, but only a negligible fraction of events with other Higgs decay modes pass the selection and is found in some regions. After simulation, the samples are normalised to the predicted $t\bar{t}H$ cross-section $\sigma_{t\bar{t}H} = 507^{+35}_{-50}$ fb [15], calculated including NLO QCD and EW corrections.

$t\bar{t} + b\bar{b}$ Modelling

To mitigate the dominant modelling uncertainties on the $t\bar{t} + \geq 1b$ processes encountered in earlier instances of the analysis (see Figure 11.1), dedicated $t\bar{t} + \geq 1b$ samples were developed where events were simulated at NLO in the 4FS framework using the POWHEG BOX RES generator [207] and OPENLOOPS [208, 209]. The parton distribution functions utilised were taken from the $\text{NNPDF3.1}_{\text{NLO}}$ nf4 set, interfaced by PYTHIA 8.244 for PS simulation and hadronisation. Specialised settings are applied in both generators to further refine and reduce modelling uncertainties.
For scale adjustments, the μ_{R} and μ_{F} parameters are set according to Equation (11.2) and Equation (11.3), respectively. The h_{damp} parameter is specifically defined according to Equation (11.4).

$$\mu_{\text{R}} = \frac{1}{2} \cdot \sum_{i=t,\bar{t},b,\bar{b},j} m_{\text{T},i} \tag{11.2}$$

$$\mu_{\text{F}} = \frac{1}{2} \cdot \sqrt[4]{m_{\text{T}}(t) \cdot m_{\text{T}}(\bar{t}) \cdot m_{\text{T}}(b) \cdot m_{\text{T}}(\bar{b})} \tag{11.3}$$

$$h_{\text{damp}} = \frac{1}{2} \cdot \sum_{i=t,\bar{t},b,\bar{b}} m_{\text{T},i} \tag{11.4}$$

Furthermore, the mass attributed to the b-quarks is set to $m_b = 4.95\,\text{GeV}$ to be consistent with the b-quarks from the $t\bar{t}$ decay. A more detailed description of the dedicated $t\bar{t} + b\bar{b}$ samples can be found in Ref. [210]. Additionally to the nominal sample, three variation samples are produced to to examine the uncertainties arising from the choice of PS and hadronisation model, the choice of matching scheme between the ME and the PS, and the recoil setting used. For the first two, a similar approach is chosen as for the signal samples, either interfacing the nominal generator with HERWIG 7.04 (PS and hadronisation model) or changing the pthard setting in PYTHIA 8.244 to one (matching scheme). The last sample varies the recoil setting in PYTHIA 8.244 from global recoil setting used in the nominal sample to the dipole recoil setting in the variation.

$t\bar{t}$ Inclusive Modelling

In addition to the dedicated $t\bar{t} + b\bar{b}$ samples, the remaining $t\bar{t} + \text{jets}$ backgrounds are obtained from inclusive 5FS $t\bar{t}$ samples, excluding any $t\bar{t} + \geq 1b$ event. The inclusive $t\bar{t}$ processes are simulated using POWHEG BOX V2 at NLO QCD level interfaced with PYTHIA 8.230. Generally, similar settings as for the POWHEG+PYTHIA 8 signal samples are applied. Using $h_{\text{damp}} = 1.5\,m_t$ and $\mu_{\text{R}} = \mu_{\text{F}} = \sqrt{m_t^2 + p_{\text{T}}^2}$ as renormalisation and factorisation scale. Three additional samples are simulated to evaluate the impact of the used PS and hadronisation model, the employed matching

scheme between ME and PS, and the setting of the h_{damp} parameter. While the first two samples have the same variations applied as for the alternative samples for the signal process, the third sample concerning the variation of the h_{damp} parameter applies $h_{\text{damp}} = 3\,m_t$.

Other Processes Modelling

In addition to the $t\bar{t}$ processes, multiple other contributions from various processes are taken into account, including single-t, $t\bar{t}V$, and several other processes involving vector bosons. The generator and the respective interface software employed are listed in Table 11.4.

Fake Lepton Estimation

To estimate the contributions from events with fake or non-prompt leptons, different approaches are employed for the distinct channels. In the dilepton channel, the contribution is estimated using MC. However, for the single-lepton channel, a dedicated data-driven estimation is necessary due to the origin of the fake leptons. The techniques applied and the resulting estimate of the fake contribution will be explained in more detail in Chapter 12.

11.2.3 *Expected Contributions in the Analysis Regions*

Based on the event selection and subsequent classification, the events are sorted into different SRs and CRs. Before the application of a PLF, pre-fit checks are performed to validate that no major mis-modelling issues are present. This is accomplished through data/MC comparisons in the different regions. In this section, the single-lepton plots are presented, while the corresponding dilepton plots can be found in Appendix A.5.
To examine the overall yields, the data/MC yields for each region are depicted in Figure 11.6, and the background composition in each region is further illustrated in Figure 11.7. Due to the blinded status of the $t\bar{t}H(H \to b\bar{b})$ legacy analysis at the time of writing, all bins with a signal-to-background ratio of more than $7.7\,\%$ are blinded.

While all SRs are blinded due to the blinding policy, all CRs show good pre-fit agreement between data and MC. The main background processes are listed separately, while the minor background processes, including di-boson and single-t processes, are grouped in the "Other" category. The separate distributions of each region, as well as the different yield tables, are shown in Appendix A.5.
As clearly visible in Figure 11.7, the dominant background in all SRs is the $t\bar{t} + \geq 2b$ process, which is expected due to the similar final state. Additionally, in most of the CRs, the respective process is also dominant, showing an overall good performance of the classification algorithm based on the expected contributions.

In addition to the overall yield and background composition checks, the performance of the classification of the signal process into the different STXS bins is shown in Figure 11.8.

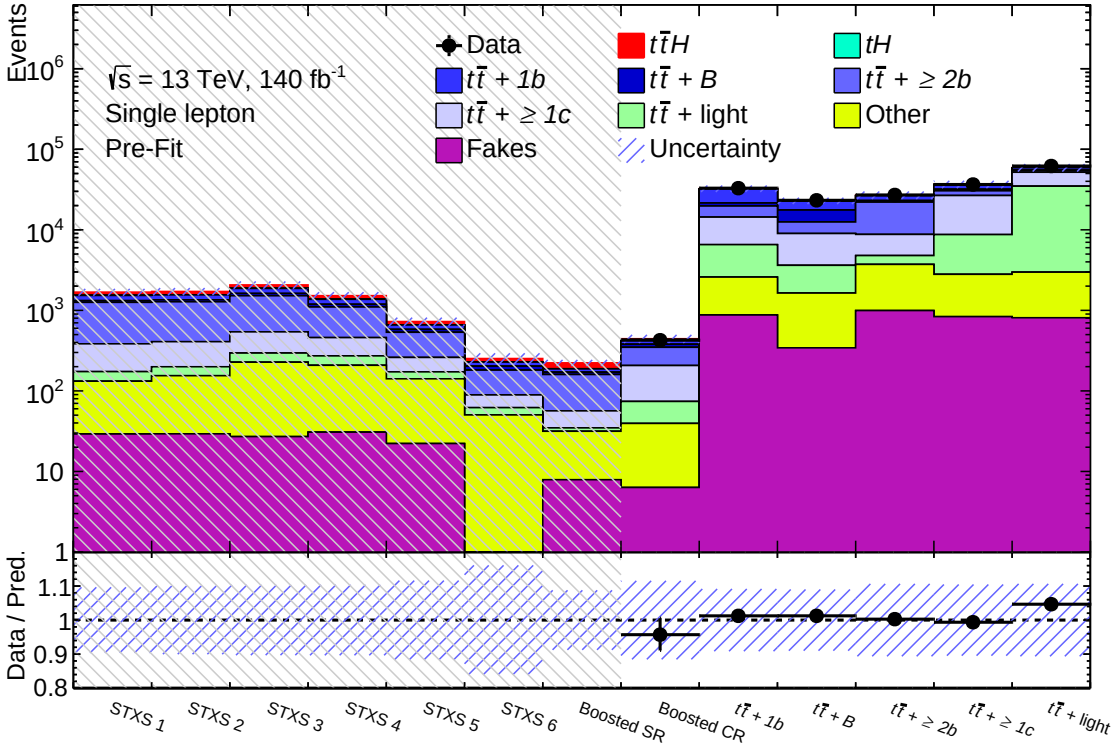


Figure 11.6.: Pre-fit summary plot for the single-lepton channel showing the overall MC and data yields in the different SRs and CRs. All regions with a data to MC ratio of more than 7.7 % are blinded, indicated by the grey shaded area.

The classification based on the reconstructed p_{T}^{H}, is particularly important for the differential cross-section measurement in the STXS framework, demonstrates great separation power, as indicated by the dominant contributions in each region from the correct $t\bar{t}H$ sub-category. While the classification performs very well for the lower STXS regions, its performance starts to degrade in the higher regions. This behaviour is expected due to the more challenging reconstruction of physics objects and the lower available training statistics for the classification algorithm.

Finally, validation plots are created, showcasing different sensitive variables and their pre-fit modelling. One of these variables is the scalar sum of all reconstructed jets and leptons, H_{T}, which is displayed in Figure 11.9 for the $t\bar{t}+\geq 1c$ and $t\bar{t}+$ light CRs.

In both regions, there is a noticeable mis-modelling of the shape between data and MC. While a simple offset could be corrected inside the PLF, mis-modelling in general, but mis-modelled shapes especially, require additional pre-fit correction. To address this, a data-driven correction for the $t\bar{t}+\geq 1c$ and $t\bar{t}+$ light MC samples is derived and applied. The details of this correction and its derivation will be explained in Chapter 13. Attempts were made to correct the mis-modelled shape of the $t\bar{t}+\geq 1b$ processes, but with the new classification algorithm, the mis-modelling was found to be negligible, and therefore, no correction is derived.

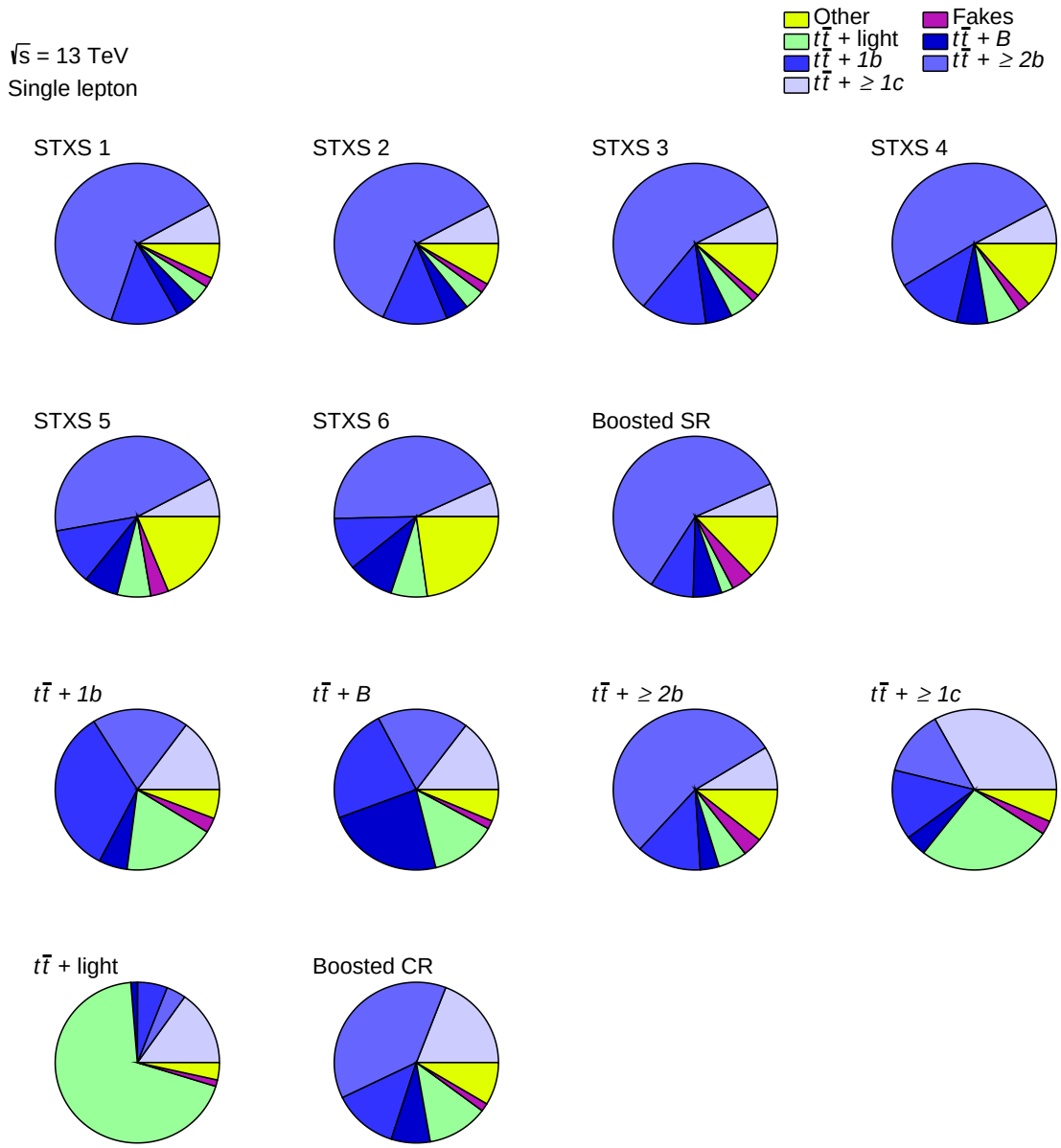


Figure 11.7.: Pre-fit background composition of the different SRs and CRs for the single-lepton channel. The upper two rows show the various SRs and the lower two show the CRs.

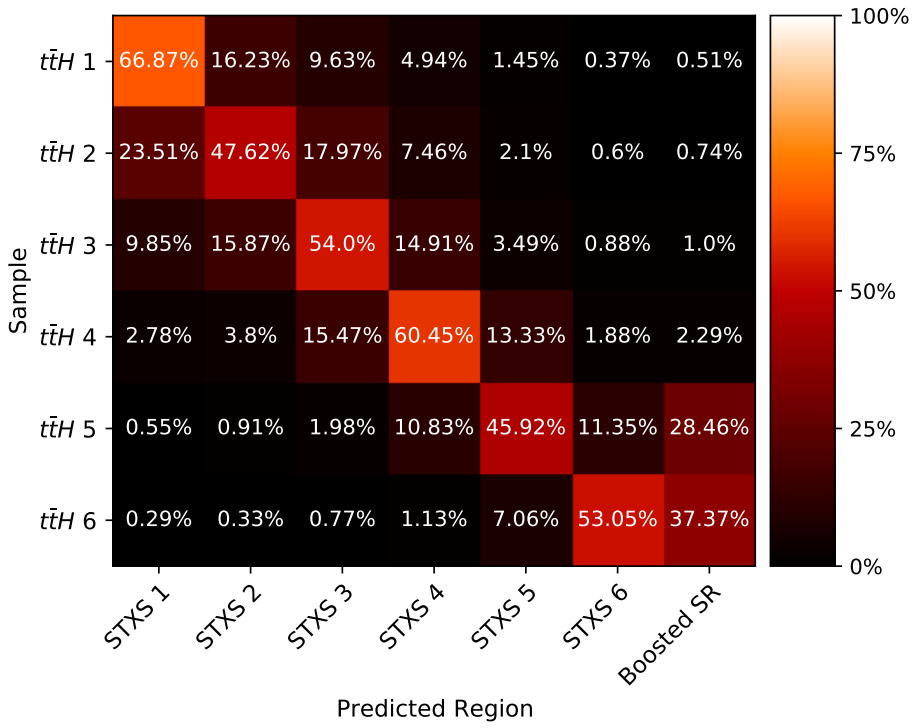
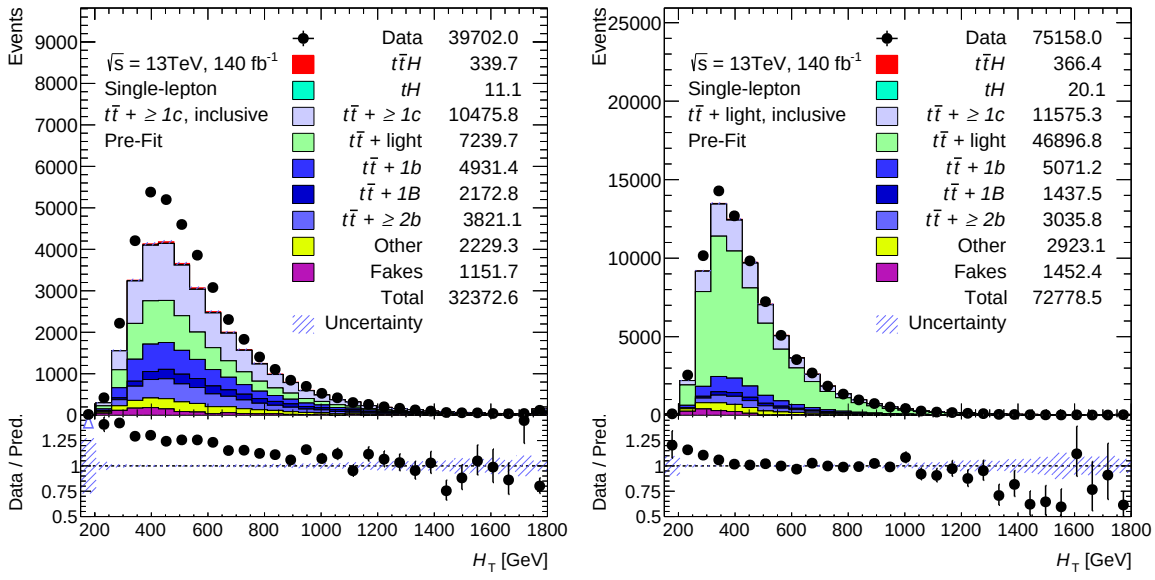


Figure 11.8.: Pre-fit expected signal contributions in the different STXS SRs in the single-lepton channel. The different $t\bar{t}H$ process indices indicate the true STXS bin, determined based on the $\hat{p}_{\mathrm{T}}^{H}$ and Table 11.2.



(a): $t\bar{t} + \geq 1c$ region

(b): $t\bar{t} +$ light region

Figure 11.9.: H_{T} distributions in the $t\bar{t} + \geq 1c$ and $t\bar{t} +$ light CRs.

11.3 PROFILE LIKELIHOOD FITS

The statistical analysis conducted for the $t\bar{t}H(H \rightarrow b\bar{b})$ legacy analysis is based on a binned PLF [211]. Such a fit relies on the Neyman-Pearson lemma [203], which provides a statistical test for the presence of the signal in the data. The lemma states that the ratio of the LHs of two hypotheses is the best test to reject one of the two hypotheses. These LHs can be generally described with Equation (11.5).

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} f(x_i, \Theta) \tag{11.5}$$

Here, x_i describes the i-th measurement of the variable x, n is the total number of measurements, and $f(x_i, \Theta)$ is the PDF of x with unknown parameters Θ. For the $t\bar{t}H(H \rightarrow b\bar{b})$ legacy analysis, these unknown parameters are the signal strengths μ, the normalisation factors k, and the set of Nuisance Parameter (NP) θ. The latter describes the effects of systematic uncertainties in the expectations, treated as extra degrees of freedom inside the fit.
For a binned distribution, the number of expected events for bin i can be compared to the experimentally observed number of events in bin i, N_i^{exp}, using Equation (11.6).

$$N_i^{\text{exp}}(\mu, k, \theta) = \sum_s \mu_s \cdot N_{\text{Sig},i}^s(\theta) + \sum_b k_b \cdot N_{\text{Bkg},i}^b(\theta), \tag{11.6}$$

Here, $N_{\text{Sig},i}^s$ denotes the number of expected signal events from process s in bin i, and $N_{\text{Bkg},i}^b$ denotes the number of expected background events from process b in bin i. Furher, μ_s describes the strength of the signal process s, and k_b describes the normalisation factor for the background process b. k_b can be either set to one, assuming the expected SM cross-section, or it can be kept free-floating in the fit with an initial start value of one. This means that the value can be changed by the fit to adapt for differences between expected and observed events. For the signal, μ_s can be defined using Equation (11.7).

$$\mu_s = \frac{\sigma^s}{\sigma_{\text{SM}}^s}, \tag{11.7}$$

Here, σ^s describes the observed cross-section of the signal process s, and σ_{SM}^s denotes the corresponding expected SM cross-section. For inclusive analyses, usually, one signal strength parameter is used, while for STXS measurements, multiple signal strength parameters are used, one for each STXS region. The NPs have a central value of $\theta = 0$, which describes the best value associated with a systematic uncertainty, while a shift in this parameter describes a so-called pull, which is a variation from zero. By definition, a ± 1 deviation corresponds to a $1\,\sigma$ deviation. In the case of continuous parameters, these are defined by extrapolation (interpolation) for $|\theta| > 1$ ($|\theta| < 1$).
To include the NPs in the LH from Equation (11.5), they are added as Poissonian or Gaussian priors, as shown in Equation (11.8).

$$\mathcal{L}(\mu, k, \theta) = \prod_{i}^{M} \frac{N_i^{\text{exp}}(\mu, k, \theta)^{N_i}}{N_i!} \cdot e^{-N_i^{\text{exp}}(\mu,k,\theta)}, \tag{11.8}$$

Here, M is the total number of bins and N_i is the number of events in bin i. To extract the best estimates for the unknown parameters, a LH ratio is built and maximised. For computational reasons, this is usually rewritten to be a negative LLH ratio that is then minimised. This is done due to the more numerically stable behaviour of minimisation algorithms compared to maximisation algorithms. The resulting test statistic λ_μ is defined as in Equation (11.9).

$$\lambda_\mu = -2\ln\left(\frac{\mathcal{L}\left(\mu, \hat{\hat{k}}, \hat{\hat{\theta}}\right)}{\mathcal{L}\left(\hat{\mu}, \hat{k}, \hat{\theta}\right)}\right), \tag{11.9}$$

Here, the double-hat parameters correspond to NPs which maximise the LH function for a given μ, and the single-hat parameters denote values which simply maximise the LH. Based on this statistic, the background-only hypothesis ($\mu = 0$) is tested by measuring the compatibility of the observed data with the expected background events. The deviation from the background-only hypothesis is measured using the significance S, which is defined as given in Equation (11.10).

$$S = \sqrt{\lambda_0} \tag{11.10}$$

The significance describes how important or significant a deviation is with respect to the background-only hypothesis and is usually given in terms of standard deviations σ. In particle physics, two major limits are used. While a significance of at least $3\,\sigma$ usually is referred to as "evidence", which corresponds to a p-value of $0.3\,\%$, a significance of $5\,\sigma$ is referred to as a "discovery". At the $5\,\sigma$-level, the p-value is no larger than $3 \cdot 10^{-7}\,\%$. The p-value describes here the probability to observe such a deviation or a larger effect under the background-only hypothesis.

All statistical tools used in the analysis were provided by the RooStat framework [212, 213] and applied using the TRExFitter framework, which is based on the HistFactory framework [214]. The minimisation is performed using the Minuit [215] library implemented in ROOT [216].

For the application, a simultaneous fit to the data is performed in both the single-lepton and dilepton channels under the background-only hypothesis in a binned fashion. All bins with a signal-to-background ratio above $7.7\,\%$ are blinded and therefore not considered in this fit, due to the blinded status of the analysis at the time of writing. Fully unblinded fits are therefore not presented in this thesis.
In addition to the background-only fits, fits to Asimov data [211] instead of real data are performed in both channels. These so-called Asimov fits set the data yields in each bin to the expected yields based on the chosen model parameters. They are used to estimate the expected uncertainties on both the signal strength and the background normalisations in addition to providing an estimate of the expected significance.

Various studies using background-only and Asimov fits were performed for both the single-lepton and dilepton channels as well as for a combination of both channels. These studies are presented and discussed in Chapter 14, which also reflects the current status of the $t\bar{t}H(H \to b\bar{b})$ legacy analysis.

11.4 SYSTEMATIC UNCERTAINTIES

Various sources of systematic uncertainties are taken into account in the $t\bar{t}H(H \to b\bar{b})$ legacy analysis. These include experimental uncertainties arising from the reconstruction and identification of physics objects, as well as modelling uncertainties, which are estimated based on alternative samples using different generators or generator settings.
In this section, the systematic uncertainties considered in the analysis will be briefly discussed, with a greater emphasis on modelling uncertainties due to their dominant effect on the analysis.

11.4.1 *Experimental Uncertainties*

Luminosity and Pile-Up Modelling

The luminosity of the full Run 2 dataset, collected between 2015 and 2018, is measured to be $140.1 \pm 1.2\,\text{fb}^{-1}$, corresponding to a relative uncertainty of $0.83\,\%$ [217].
The uncertainty in pile-up modelling is estimated using a reweighting-based procedure, achieved by varying the nominal Scale Factors (SFs) applied to the data pile-up distribution [218].

Leptons

For leptons, systematic uncertainties arise from various sources, including the trigger, reconstruction, identification, and isolation procedures. Both electrons and muons exhibit efficiency differences between simulation and data, which are corrected using specific SFs derived using the tag-and-probe method on $Z^0 \to l^+l^-$ events [132, 219]. The SFs are taken into account using in four (ten) independent NPs for electrons (muons).
Moreover, additional systematic uncertainties stem from corrections applied to the energy and momentum scale and resolution of the leptons. These corrections are determined using $Z^0 \to l^+l^-$ and $J/\Psi \to l^+l^-$ events, with additional corrections for electrons derived from the E/p ratio measured in $W^\pm \to e\nu$ events [131, 219]. In total, three (four) NPs are considered for electrons (muons).

Jets and b-Tagging

Various sources of jet uncertainties are taken into account, including the efficiency of pile-up rejection by the JVT and effects from the JES and JER calibrations. Dedicated SFs for the JVT efficiency corrections are derived from $Z^0 \to \mu^+\mu^-$ events in both data and simulation. These factors are then propagated to MC via event weights applied at the analysis level [220].

Table 11.5.: List with the sources and the number of independent NPs from the source, respectively.

Category	Number of NPs
Statistical	6
Modelling and theory	4
Detector components	2
Mixed detector and modelling	3
Jet flavour	3
Pile-up	4
η dependence	6
High-p_{T} jets	1
Punch-through	2

JES uncertainties are obtained by combining information from test-beam data, actual collision data, and simulation. These uncertainties are parametrised in 31 independent NPs, depicted in Table 11.5.

The three jet flavour related uncertainties include the b-jet energy scale, flavour response and composition. A more detailed description of all sources and their respective NPs can be found in Ref. [143].

JER uncertainties are determined from data and simulation using dijet events, where the balance in jet p_{T} and η is measured [221]. 13 independent sources of uncertainties are considered, which are parametrised using 12 effective terms and one term for comparison of data/MC.

Uncertainties in b-tagging arise from the b-jets efficiency and c and light-flavour jets mis-tag calibrations of the DL1r algorithm. Multiple eigenvalues to estimate the uncertainties are derived using a principal component analysis, which yields in total 45 b-jets, 20 c-jets, and 20 light-flavour jets eigenvalues. Additionally, due to the limited p_{T} range of the calibrations, extrapolation uncertainties at high-p_{T} are derived. These uncertainties involve extrapolating the scale factors from the last-p_{T} calibrated bin to higher-p_{T} bins, using a sample of artificial heavy Z' resonances[8] that produce a sufficient number of high-p_{T} jets for the extrapolation.

Missing Transverse Momentum

All systematic uncertainties, concerning the energy scale and resolution of the reconstructed physics objects mentioned above, are propagated to the $E_{\mathrm{T}}^{\mathrm{miss}}$ algorithms. In addition to these hard contributions, additional uncertainties for the soft terms are incorporated to address discrepancies between data and simulation [144], yielding in total three independent NPs, which are considered.

11.4.2 *Signal Modelling Uncertainties*

The impact of parton distribution function and α_s variations on the total $t\bar{t}H$ cross-section is $\pm 3.6\,\%$ [15, 222–226]. The effect on the $\hat{p}_{\mathrm{T}}^{H}$ distribution is considered negligible. Theoretical uncertainties on

8 The Z' sample used is similar to the one described in Section 7.2.1.

the $t\bar{t}H$ cross-section arising from migration between STXS bins are also considered.
To evaluate and estimate the uncertainty on the ISR, weights are used in the ME and PS simulation, varying the renormalisation and factorisation scales, μ_{R} and μ_{F}, by a factor of 0.5 while using the Var3c up variation from the A14 tune [188] to simulate higher parton radiation. For lower parton radiation, the scales are varied by a factor of 2 while using the Var3c down variation. For the FSR uncertainty evaluation, a similar approach is employed, varying the renormalisation scale for the QCD emission in the PS for the FSR by factors of 0.5 and 2.
The uncertainties on the total cross-section due to higher-order corrections in QCD are centrally evaluated for the $t\bar{t}H$ process by the LHC Higgs Cross-Section (LHCHXS) working group [14] and are estimated to be $9.2\,\%$. For the branching fraction of the Higgs boson into b-quarks, an uncertainty of $(+1.2/-1.3)$ % is used. Furthermore, for the branching fraction of the Higgs boson to a pair of $W^{\pm}$ bosons (other processes), an uncertainty of $(+1.6/-1.5)$ % ($\pm 5\,\%$) is taken into account.
To assess the uncertainty due to the NLO matching, the nominal sample is compared to a POWHEG BOX V2 sample interfaced with PYTHIA 8, where the pthard parameter is set to 1 (Table 11.4). A similar approach is used to evaluate the effect of the chosen PS and hadronisation model, where the nominal sample is compared to a POWHEG BOX V2 sample interfaced with HERWIG 7.04.

11.4.3 *Background Modelling Uncertainties*

$t\bar{t}+\mathrm{jets}$

For all $t\bar{t}+\mathrm{jets}$ sub-processes, different types of uncertainties are assigned to each sub-process based on their individual properties. While $t\bar{t}+$ light is relatively precisely measured in data due to differences in diagrams compared to $t\bar{t}+\geq 1b$ and $t\bar{t}+\geq 1c$, the $t\bar{t}+\geq 1b$ and $t\bar{t}+\geq 1c$ processes can have additional diagrams depending on the used flavour scheme in the parton distribution functions. Furthermore, the difference in mass between the c and b-quarks for the $t\bar{t}$+HF processes adds additional distinctions for the latter ones.
To assess uncertainties on acceptance and shapes, a comparison between the nominal setup and alternative setups is made. Due to the use of a PLF to estimate the normalisation factors for the $t\bar{t}+\geq 1b$ processes in data, such comparisons would drastically alter the fraction of each $t\bar{t}+\geq 1b$ component in the analysis phase space. To address this issue, the alternative predictions are reweighted so that the same fraction of $t\bar{t}+\geq 1b$ events as in the nominal sample is present in the analysis phase space.
Similar to the signal process, the uncertainties on ISR and FSR are estimated with a weight-based approach by varying the used renormalisation and factorisation scales by 0.5 and 2. To assess the impact of the PS and hadronisation model, the nominal sample is compared to a POWHEG BOX V2 sample with the same settings, interfaced with HERWIG 7.16. Furthermore, to evaluate the effect of the recoil handling in the initial state PS, an additional sample was generated, changing the global recoil setting in PYTHIA 8 to a dipole recoil.
For the NLO matching uncertainties, the nominal samples are compared to a POWHEG BOX V2 interfaced with PYTHIA 8, where the pthard parameter for the latter is set to 1. Additional information

on the estimation of uncertainties for the $t\bar{t} + \geq 1b$ samples can be found in Ref. [210].
To examine the uncertainties arising from the choice of the h_{damp} parameter, an alternative sample with a varied h_{damp} parameter is used as variation (see Section 11.2.2).

H_T Reweighting

In both channels, two variations are considered. Firstly, alternative H_{T} reweightings, derived with adapted post-fit normalisation factors for $t\bar{t} + \geq 1c$ and $t\bar{t} +$ light, are employed, varying the central value of the factors by their respective uncertainties (central value $\pm$ uncertainty). Secondly, a simple non-application of the reweighting (on/off) is considered. While the latter is applied for each N_{jets} reweighting respectively, the variation with adapted post-fit normalisation factors is applied to the overall reweighting.

Fake/Non-prompt Leptons

For the fake and non-prompt leptons in the single-lepton channel, an overall normalisation uncertainty of $50\,\%$ is taken into account together with a variation based on an adapted parametrisation for the 2D fake efficiencies, where the efficiencies are parametrised in bins of lepton p_{T} and lepton $|\eta|$. For the dilepton channel, an overall normalisation uncertainty of $50\,\%$ is taken into account.

Other Backgrounds

For all other background processes, the sources of systematic uncertainties and their description can be found in Table 11.6.

Table 11.6.: List with all sources of systematic uncertainties considered for the other processes modelling.

	Source of Systematic Uncertainty	Description
$t\bar{t}V$ [227]	Cross-section	$\pm 15\,\%$
	NLO matching	MADGRAPH5_AMC@NLO + PYTHIA 8 vs SHERPA
	PS and hadronisation	MADGRAPH5_AMC@NLO + PYTHIA 8 vs SHERPA
Single-t [228–230]	NLO cross-section	$\pm 5\,\%$
	NLO matching	POWHEG BOX V2 + PYTHIA 8 vs POWHEG BOX V2 + PYTHIA 8 pthard
	PS and hadronisation	POWHEG BOX V2 + PYTHIA 8 vs POWHEG BOX V2 + HERWIG 7
	tW and $t\bar{t}$ interference [231]	Diagram Removal (DR) vs Diagram Subtraction (DS) scheme
W+jets	Cross-section	$\pm 40\,\%$
	Cross-section	$\pm 30\,\%$
Z+jets	Normalisation	$\pm 35\,\%$
Diboson [232]	Cross-section and add. jet production	$\pm 50\,\%$
tZq [233]	Cross-section (PDFs)	$\pm 0.9\,\%$
	Factorisation and normalisation	$\pm 7.9\,\%$
tWZ [233]	Cross-section	$\pm 50\,\%$
$tHjb$ [15]	Cross-section (PDFs)	$\pm 3.7\,\%$
	Factorisation and normalisation	$+6.5\,\%/-14.9\,\%$
tWH [15]	Cross-section (PDFs)	$\pm 3.7\,\%$
	Factorisation and normalisation	$+6.5\,\%/-6.7\,\%$
$t\bar{t}t\bar{t}$	Normalisation	$\pm 50\,\%$

12

ESTIMATION OF THE FAKE-LEPTON CONTRIBUTION IN THE SINGLE-LEPTON CHANNEL

In addition to the background contributions estimated with MC as introduced in Section 11.2.2, other processes with mis-identified physics objects can pass the trigger and pre-selection stage, contributing to the overall yield in the SRs and CRs. In many of these events, a jet or a photon may be mis-identified as a lepton. These are referred to as fake leptons and result from imperfect object reconstruction and identification. Additionally, leptons from secondary particle decays, known as non-prompt leptons, identified as prompt leptons, which stem from the primary vertex, can also contribute. To mitigate the impact of these particles on analyses, dedicated and well-established methods have been developed to identify and reject these objects. However, these algorithms may also reject a non-negligible portion of the final statistics, making analysis in a challenging kinematic phase space very difficult.
One of these analyses with such a challenging phase space is the $t\bar{t}H(H \to b\bar{b})$ analysis. To ensure that enough events pass the trigger and pre-selection stage, relatively loose identification and isolation criteria are required for candidate leptons, making it easier to mis-identify a photon or jet as a lepton. In the previous round of the analysis, due to even tighter requirements on the number of jets and b-tagged jets, the contribution from events with fake leptons was estimated to be negligible. With the new legacy analysis, a relaxation of these requirements was introduced, necessitating a re-estimation of the contributions from events with fake leptons. While this effect can be estimated in the dilepton channel using MC simulations, in the single-lepton channel, a data-driven approach is chosen. The reason for this choice is the origin of fake leptons in the single-lepton channel, mainly originating from multijet QCD events, where one jet is mis-identified as a lepton. Although large MC samples for these processes are available, the amount of statistics and computational power needed for an acceptable estimation is still too large. Additionally, these processes are subject to large modelling uncertainties, which can be avoided with a data-driven approach.

Two data-driven methods are considered for the estimation of fake-lepton contributions: the matrix method and the fake factor method [234]. Although both methods follow a similar approach, they differ in the application of the estimation.
To estimate the contributions from events with fake leptons to the overall yield, both the matrix and fake factor methods use two different definitions for leptons. These definitions or selections, namely *loose* and *tight*, differ in their identification and isolation requirements and are applied after the trigger but before the pre-selection stage. As the names suggest, the *loose* definition employs looser requirements, while the *tight* definition enforces more stringent criteria. The identification and isolation requirements for both selections are listed in Table 12.1.

Table 12.1.: Selection requirements for the two lepton definitions *loose* and *tight*.

Lepton Type	Definition	Identification/Quality WP	Isolation WP
Electrons	*tight*	*TightLH* [235]	*FCTight*
	loose	*MediumLH* [235]	-
Muons	*tight*	*medium* [219]	*TightTrackOnlyVarRad*
	loose	*loose* [219]	-

While the *tight* definition employs the same lepton criteria as are present in the event selection for the SRs, the *loose* definition adopts more lenient lepton criteria, incorporating no additional isolation requirement. It is important to note that, owing to the employed triggers, specific isolation requirements (Table 11.3) are already imposed on the lepton candidates, which do not significantly differ from the isolation criteria required by the *tight* definition, making the distinction between *loose* and *tight* relatively minor. Additionally, it should be mentioned that, by definition, all events where the lepton passes *tight* also satisfy *loose*. For the fake factor method, an additional definition is used, referred to as *loose-not-tight*, which encompasses all events with a lepton passing *loose* but not *tight*, as the name suggests.

12.1 REAL AND FAKE EFFICIENCIES

For both the matrix and fake factor methods, so-called efficiencies are employed. While the matrix method requires two efficiencies, namely the real ϵ_{real} and fake ϵ_{fake} efficiencies, the fake factor method only uses one, ϵ_{fake}. These efficiencies are defined by Equation (12.1):

$$\epsilon_{\text{real}} = \frac{N_{\text{real}}^{tight}}{N_{\text{real}}^{loose}}, \quad \epsilon_{\text{fake}} = \frac{N_{\text{fake}}^{tight}}{N_{\text{fake}}^{loose}} \tag{12.1}$$

Here, $N_{\text{real}}^{tight} / N_{\text{real}}^{loose}$ represents the number of events with a real lepton passing the *tight*/*loose* definition, and $N_{\text{fake}}^{tight} / N_{\text{fake}}^{loose}$ denotes the number of events with a fake lepton passing the *tight*/*loose* definition.

To accurately determine these numbers, truth information about the origin of the leptons is required to estimate the number of events correctly. While this information is readily available for MC simulations, it is not present in the data. To address this challenge, a slightly different method is employed to measure these efficiencies. While the real efficiency can be easily estimated using only MC by matching the reconstructed leptons to prompt leptons, the fake efficiency needs to be derived using both MC and data.
To achieve this, two regions are utilised, each enriched in either real or fake leptons. The definitions of these two regions are listed in Table 12.2.

The real-enriched region is designed to be close to the Z-peak and capture opposite-sign leptons from the latter decay. The real efficiency can then be easily calculated using Equation (12.2),

Table 12.2.: Definition of the fake and real lepton enriched regions in which the respective efficiencies are measured.

Requirement	Fake enriched region	Real enriched region
N_{Jets}	≥ 2	≥ 1
Jet p_{T}	$\geq 25\,\text{GeV}$	$\geq 25\,\text{GeV}$
$N_{b-\text{tag}}$	≥ 2	$\stackrel{!}{=} 0$
N_{Leptons}	$\stackrel{!}{=} 1$	$\stackrel{!}{=} 2$, opposite sign
Lepton p_{T}	$\geq 27\,\text{GeV}$	At least one lepton $\geq 27\,\text{GeV}$
$E_{\text{T}}^{\text{miss}}$	$\leq 20\,\text{GeV}$	-
M_{T}^{W}	$\leq 20\,\text{GeV}$	-
m_{ll}	-	$90\,\%m_Z < m_{ll} < 110\,\%m_Z$

$$\epsilon_{\text{real}} = \frac{N^{tight}}{N^{loose}} \cdot \frac{SF_{tight}}{SF_{loose}} \tag{12.2}$$

where N^{tight} / N^{loose} denotes the number of leptons passing the *tight*/*loose* definition, and SF_{tight} / SF_{loose} denotes the recommended lepton scale factors for their respective lepton definition.

For the fake efficiency, the respective region was designed to be enriched in fake leptons and depleted in real leptons, but also close to the final analysis phase space. To determine the fake efficiency, MC and data are used. The number of fake leptons, which are not modelled in MC but present in data, can be extracted by subtracting the expected number of prompt *tight* and *loose* leptons from the *tight* and *loose* leptons present in data. This results in the definition of the fake efficiency as shown in Equation (12.3),

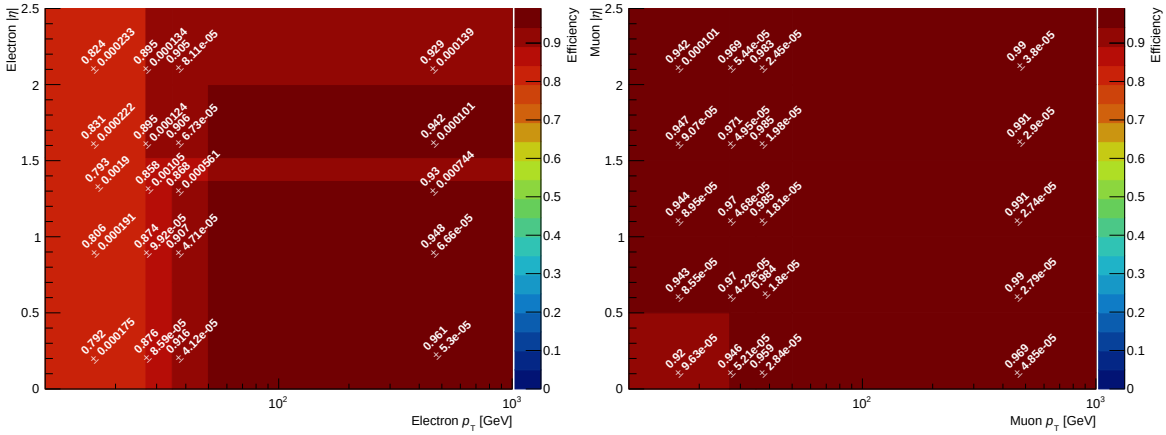
$$\epsilon_{\text{fake}} = \frac{N_{\text{data}}^{tight} - N_{\text{MC}}^{tight}}{N_{\text{data}}^{loose} - N_{\text{MC}}^{loose}} \tag{12.3}$$

where N_{data}^{tight} / N_{data}^{loose} denotes the number of *tight*/*loose* leptons in data and N_{MC}^{tight} / N_{MC}^{loose} denotes the number of *tight*/*loose* leptons in MC.

Due to the high dependence of the efficiencies on different event and object kinematics, the efficiencies are derived in a binned 2D plane using the lepton p_{T} and $|\eta|$ as the variables that span the plane. Additionally, the efficiencies are derived separately for each lepton flavour. The binning of these efficiency maps for both the real and fake efficiencies are the same but differ between the lepton flavours to avoid the crack region of the ECal for the electrons. The binning is shown in Table 12.3.

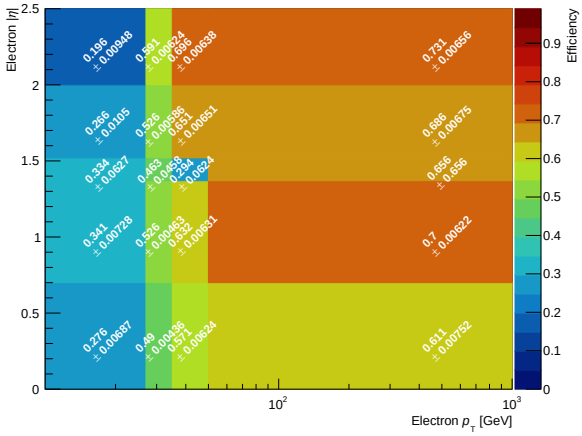
Table 12.3.: Binning of the efficiency maps for electrons and muons. The $|\eta|$ binning differs between the two lepton flavours to avoid the crack region of the ECal for the electrons.

	Electrons	Muons
Lepton p_{T} (in GeV)	[10, 27, 35, 50, 1000]	
Lepton $\lvert\eta\rvert$	[0, 0.7, 1.37, 1.52, 2, 2.5]	[0, 0.5, 1, 1.5, 2, 2.5]

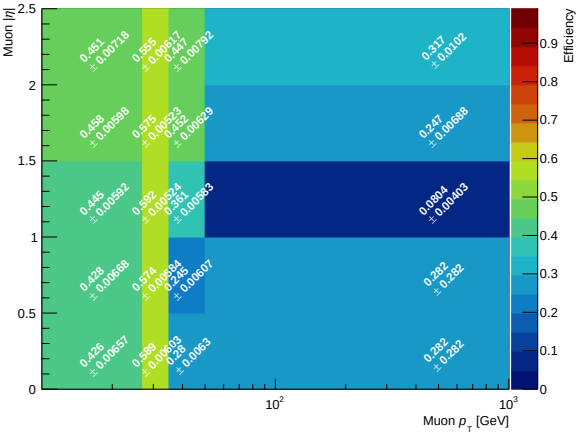


(a): Electron real efficiency map

(b): Muon real efficiency map



(c): Electron fake efficiency map



(d): Muon fake efficiency map

Figure 12.1.: Real and fake efficiency maps for electrons and muons.

The resulting efficiency maps for both electrons and muons are visualised in Figure 12.1. For both electrons and muons, high real efficiencies are achieved across nearly all bins, with performance degrading towards the lower p_{T} region, as expected. In contrast to the very promising real efficiencies, the derived fake efficiencies show high values in the higher p_{T} region for electrons as well as for muons in the lower p_{T} region. Both lepton flavours exhibit large fluctuations and high uncertainties, which can be explained by the difference in the definitions of *loose* and *tight*. As mentioned earlier, the triggers used already apply isolation criteria to all leptons. These criteria, which are relatively similar to the criteria used in *tight*, are the source of the large fake efficiencies observed here. Due to the resulting small difference between the definitions of *loose* and *tight*, it is relatively likely that leptons passing the *loose* definition also pass the *tight* definition, leading to large values when using Equation (12.3).

With the newly derived real and fake efficiencies, both methods can be used to estimate the contribution of events with fake leptons.

12.2 MATRIX METHOD

The matrix method relies on both real and fake efficiencies to estimate the number of events with a fake lepton passing the *tight* definition, denoted as N^{tight}_{fake}. The total number of events with a lepton passing the *loose*/*tight* definition can be defined as $N^{loose}\,/\,N^{tight}$, using Equation (12.4):

$$N^{loose} = N^{loose}_{\text{real}} + N^{loose}_{\text{fake}}, \quad N^{tight} = N^{tight}_{\text{real}} + N^{tight}_{\text{fake}} \tag{12.4}$$

Applying Equation (12.1) to the above, one obtains an equation for N^{tight}_{fake}:

$$N^{tight}_{\text{fake}} = \frac{\epsilon_{\text{fake}}}{\epsilon_{\text{real}} - \epsilon_{\text{fake}}} \cdot \left(\epsilon_{\text{real}} \cdot N^{loose} - N^{tight}\right) \tag{12.5}$$

With Equation (12.5), it is now possible to estimate the overall number of events with a fake lepton passing the *tight* definition, contributing to the overall event yields, using the derived efficiencies and the total number of data events with a lepton passing *loose*/*tight*.

Application of the Matrix Method

To apply this to a binned distribution, a per-event weight w_i is calculated for all data events using Equation (12.6):

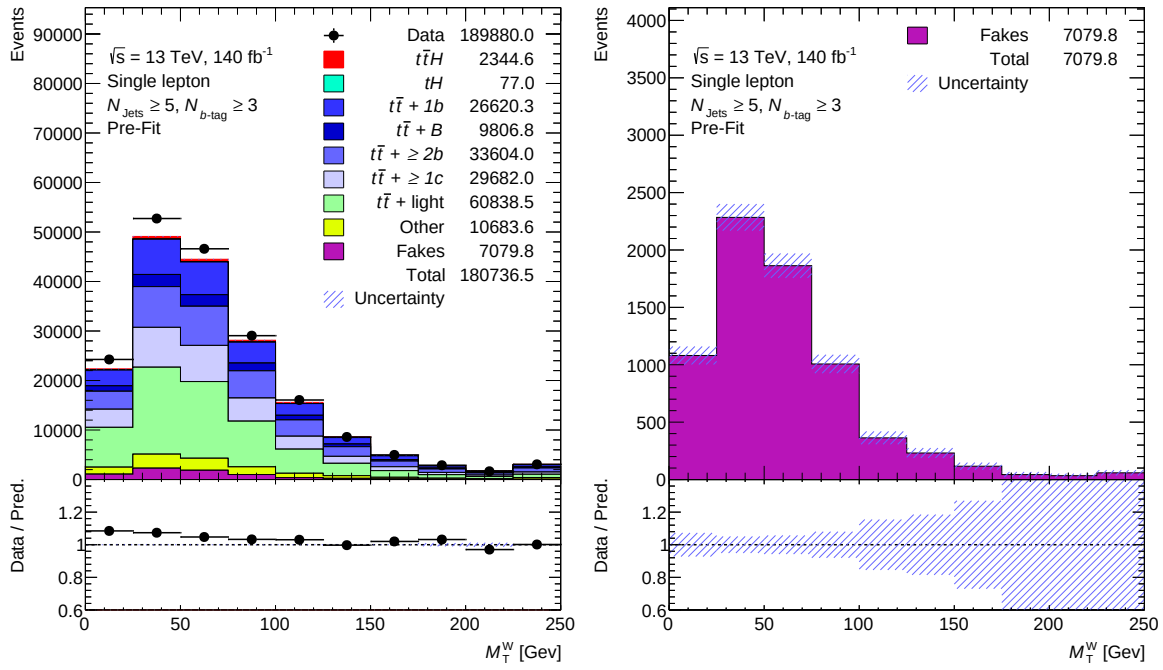
$$w_i = \frac{\epsilon_{\text{fake}}}{\epsilon_{\text{real}} - \epsilon_{\text{fake}}} \cdot (\epsilon_{\text{real}} - P_i) \tag{12.6}$$

Here, P_i describes whether the lepton in the event passes the *tight* definition or not. By definition, $P_i = 1$ if the lepton passes the *tight* definition and $P_i = 0$ if not. This approach is advantageous as the per-event weight also enables the extraction of the shape of the resulting fake distributions.
To visualise and examine the derived estimate, the transverse $W^{\pm}$ mass, m^W_{T}, distribution for all contributions and only the fakes contribution are depicted in Figure 12.2.

A similar shape of the fake contribution compared to all other contributions can be observed. While a significant part of the fakes contribution is located at lower values, a tail up to 250 GeV is visible.

Limitations of the Matrix Method

Despite its effectiveness and well-established nature, the matrix method faces challenges due to the closely aligned definitions of *loose* and *tight*. As mentioned earlier, these definitions are not significantly different. Upon closer examination of the first part of Equation (12.6), it becomes apparent that if both the fake and real efficiencies are high, the absolute value of the weight will be large, potentially leading to unphysically large fake contributions. Additionally, with similar definitions, most fake leptons passing *loose* will also pass *tight*, resulting in an overall negative event weight. Proper and substantial separation between the definitions is essential to counterbalance these negat-



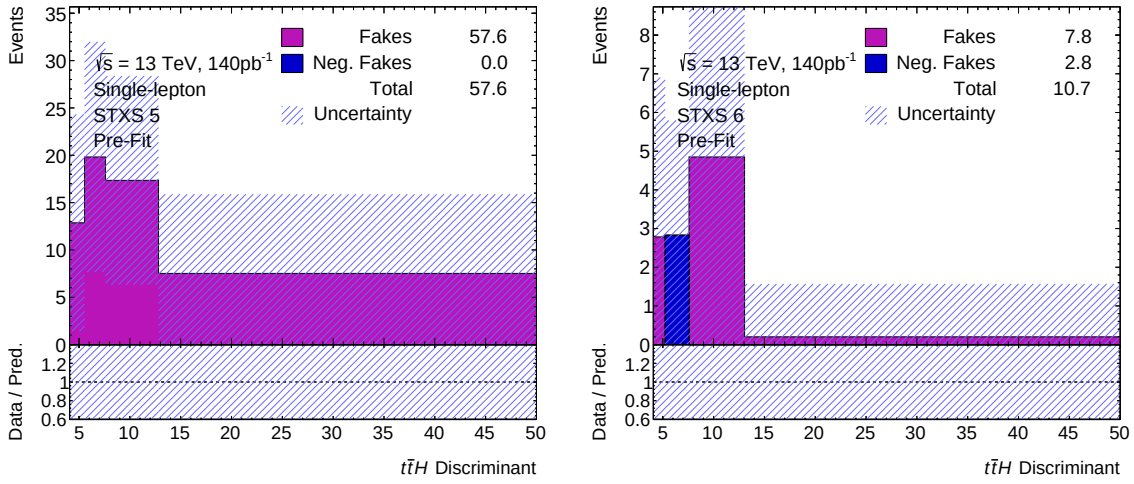
(a): Including all contributions

(b): Only fakes contribution

Figure 12.2.: m_{T}^{W} distributions with all contributions (a) and only with the fakes contribution (b). The fakes contribution was derived using the matrix method.

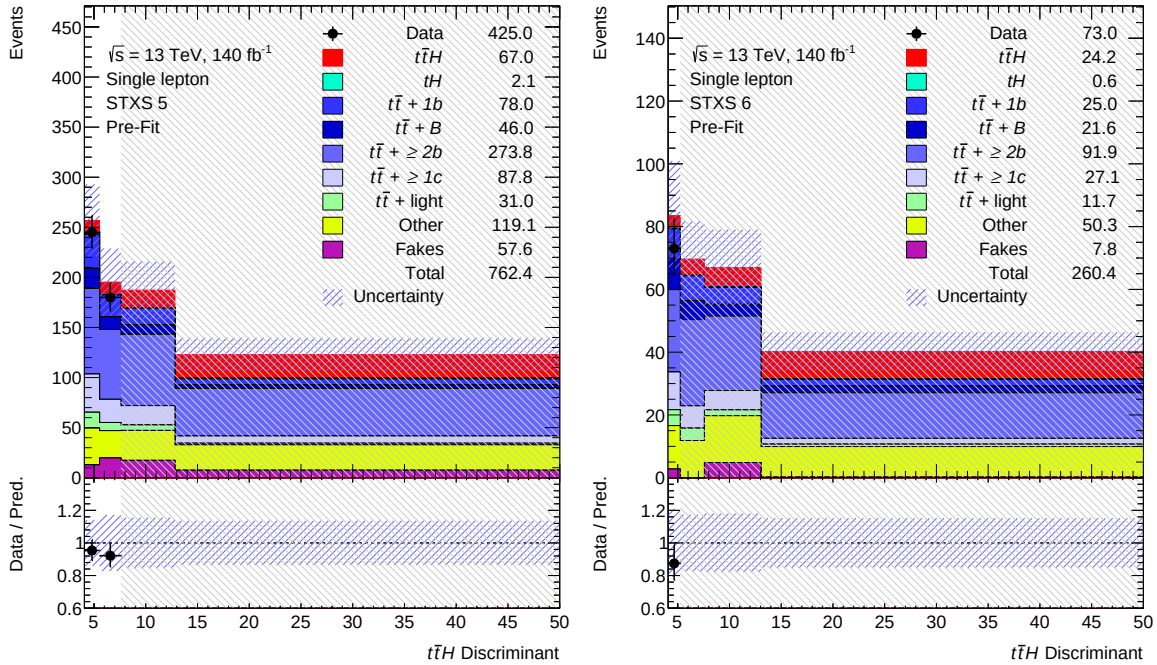
ive weights with positive weights, ensuring their overall cancellation. Both issues can be exacerbated in statistically limited regions, such as is the case for two of the STXS regions in the SR. To illustrate this, the fake contributions to STXS regions 5 and 6 are depicted in Figure 12.3 together with the full pre-fit plots for STXS regions 5 and 6, which are presented in Figure 12.4.

As visible, the contributions in STXS regions are relatively large with even larger uncertainties. The overall contribution is comparable to the $t\bar{t}+$ light contribution expected in this region, which is known to be too large for the fakes. In the STXS 6 region, the contributions are at an acceptable level, but the second and third bins have negative fake contributions, which are unphysical.



(a): STXS 5 (b): STXS 6

Figure 12.3.: Positive and negative fakes contribution, estimated using the matrix method, in the STXS regions 5 and 6 in the signal region.



(a): STXS 5 (b): STXS 6

Figure 12.4.: Pre-fit plots of STXS regions 5 and 6, including the estimated fakes contribution using the matrix method. All bins with a signal-to-background ratio of above 7.7 % are blinded, indicated by the grey marked area.

12.3 FAKE FACTOR METHOD

To address the limitations and challenges associated with the matrix method, an alternative method for estimating fake contributions in the single-lepton channel was explored. The fake factor method, as mentioned earlier, is known to be more stable against closely aligned definitions for *loose* and *tight* and is expected to handle fluctuations and negative contributions more effectively.
In contrast to the matrix method, the fake factor method solely relies on the fake efficiency ϵ_{fake} while assuming that real leptons are generally well-modelled in MC. Under this assumption, it becomes possible to estimate the number of events with a fake lepton passing the *tight* definition by using only the number of events with a lepton passing the *loose-not-tight* criteria. Utilising the definition of ϵ_{fake} from Equation (12.1) and introducing the *loose-not-tight* definition, $N_{\text{fake}}^{\textit{tight}}$ can be expressed as given in Equation (12.7).

$$N_{\text{fake}}^{\textit{tight}} = N_{\text{fake}}^{\textit{loose}} \epsilon_{\text{fake}} = \left(N_{\text{fake}}^{\textit{loose-not-tight}} + N_{\text{fake}}^{\textit{tight}} \right) \cdot \epsilon_{\text{fake}} \tag{12.7}$$

Here, $N_{\text{fake}}^{\textit{loose-not-tight}}$ represents the number of events with a fake lepton passing the *loose* but not the *tight* definition. By rearranging this equation, a fake factor $\mathbf{F}$ can be extracted, allowing the estimation of $N_{\text{fake}}^{\textit{tight}}$ using only $N_{\text{fake}}^{\textit{loose}}$. This relationship is illustrated in Equation (12.8).

$$N_{\text{fake}}^{\textit{tight}} = \underbrace{\frac{\epsilon_{\text{fake}}}{1 - \epsilon_{\text{fake}}}}_{:=\mathbf{F}} \cdot N_{\text{fake}}^{\textit{loose-not-tight}} = \mathbf{F} \cdot \left(N^{\textit{loose-not-tight}} - N_{\text{real}}^{\textit{loose-not-tight}} \right) \tag{12.8}$$

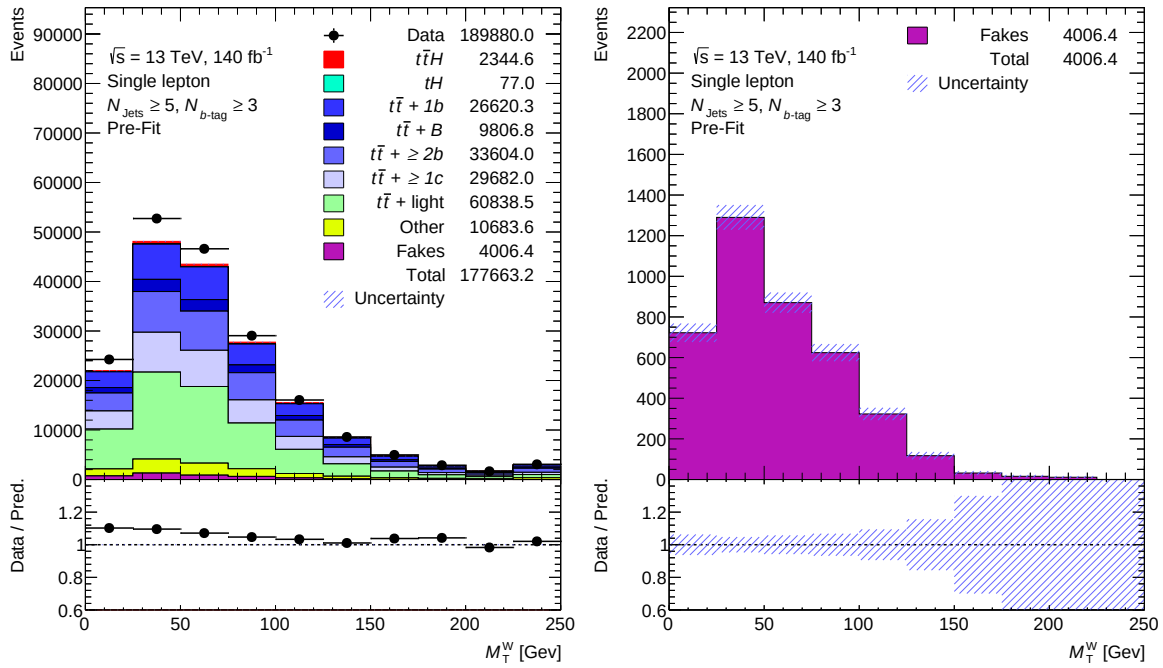
In this equation, $N^{\textit{loose-not-tight}}$ denotes the total number of events with a lepton passing the *loose* but not the *tight* definition, and $N_{\text{real}}^{\textit{loose-not-tight}}$ represents the number of events with a real lepton passing the *loose* but not the *tight* definition.

Application of the Fake Factor Method

Similar to the matrix method, the fake factor method needs to be applied to a binned distribution. To achieve this, a per-event weight is calculated using the fake factor per event $\mathbf{F}_i$, considering the different fake efficiencies based on the leading jet and lepton kinematics in event i. While the matrix method relies on fake efficiencies parametrised by lepton p_{T} and $|\eta|$, the fake factor method re-derived the fake efficiencies, using the leading jet p_{T} and the lepton $|\eta|$ as parametrisation. With this, the final estimate of the fakes contribution can be obtained using Equation (12.9).

$$N_{\text{fake}}^{\textit{tight}} = \sum_{i}^{\textit{loose-not-tight}\ \text{data}} \mathbf{F}_i - \sum_{j}^{\textit{loose-not-tight}\ \text{MC}} w_{\text{MC}}^{j} \mathbf{F}_j \tag{12.9}$$

Here, w_{MC}^{j} denotes the MC weight for MC event j. This formula involves subtracting all contributions from events with prompt leptons from the given data. The estimation for these events comes from MC, based on the assumption that it is well-modelled.



(a): Including all contributions

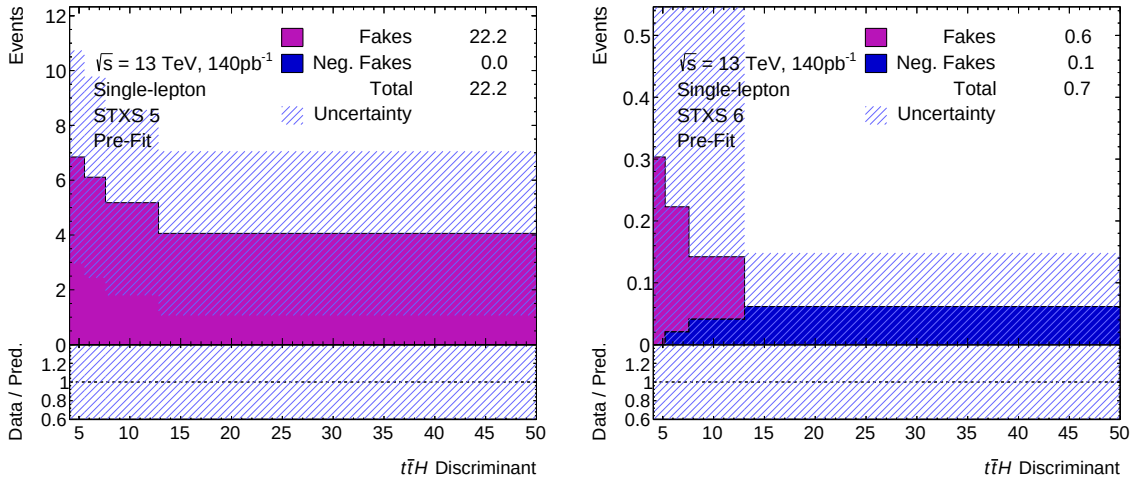
(b): Only fakes contribution

Figure 12.5.: m_{T}^{W} distributions with all contributions (a) and only with the fakes contribution (b). The fakes contribution was derived using the fake factor method.

To examine the derived estimate, the m_{T}^{W} distribution for all contributions and only the fakes contribution are depicted in Figure 12.5.

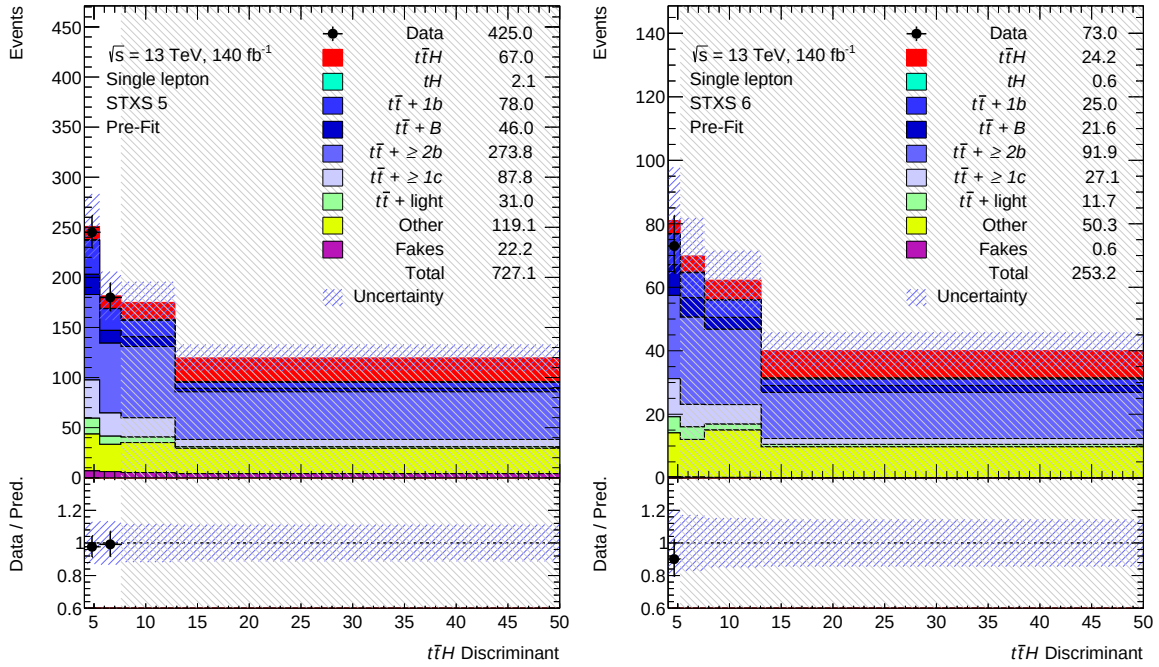
Compared to the fakes contribution yielded by the matrix method (Figure 12.2), a significant drop in size can be observed. Although similar shapes are evident, no tail for the fake factor fakes contribution is present and also a slight shift of the distribution towards lower values is observed.
To examine the good performance of the fake factor method also in the critical regions of the matrix method, the fakes contribution in the signal region for the STXS regions 5 and 6 are illustrated in Figure 12.6.

In comparison to the results obtained from the matrix method, the overall fakes contribution is considerably smaller with the fake factor estimation, exhibiting a significantly improved shape of the distribution. Moreover, negative contributions in STXS 6 are minor and entirely negligible in comparison to other contributions.
Considering these factors, the decision was made to utilise the fakes estimation from the fake factor method, providing a more accurate and reasonable overall estimate. To illustrate the impact that the fakes will have, the pre-fit plots for STXS regions 5 and 6 are presented in Figure 12.7, encompassing all other contributions.
When comparing the shapes of the fake contributions in Figure 12.6 with the overall shape in Figure 12.7, it becomes evident that, in addition to the overall improved yields, the estimation from the fake factor method also provides a better representation of the shape of the fakes contribution. Furthermore, considering the negligible impact and agreement with zero within uncertainties, the decision was made to exclude the fakes contribution from STXS region 6.



(a): STXS 5 (b): STXS 6

Figure 12.6.: Positive and negative fakes contribution, estimated using the fake factor method, in the two STXS regions 5 and 6 in the signal region.



(a): STXS 5 (b): STXS 6

Figure 12.7.: Pre-fit plots of STXS regions 5 and 6, including the estimated fakes contribution using the fake factor method. All bins with a signal-to-background ratio of above 7.7 % are blinded, indicated by the grey marked area.

13

H_{T} REWEIGHTING

With the latest inclusion of the new and improved $t\bar{t}+b\bar{b}$ samples to the $t\bar{t}$ inclusive samples, the used MC simulations describe the recorded data reasonably well (see Figure 11.6). Besides some minor normalisation issues, which can be easily addressed in the PLF, a non-negligible shape mis-modelling is observed in the H_{T} distributions of multiple $t\bar{t}+\mathrm{jets}$ processes when comparing data and MC. The origin of these shape mis-modellings can be traced back to missing higher order corrections in the simulation of the respective $t\bar{t}+\mathrm{jets}$ processes. While a data-driven correction or reweighting of the miss-modelled $t\bar{t}+\geq 1b$ shape cannot be easily achieved due to the similar final state compared to the signal, the $t\bar{t}+$ light and $t\bar{t}+\geq 1c$ shape mis-modelling can be corrected by reweighting, making use of the different final state and an orthogonal, $t\bar{t}+\mathrm{jets}$-enriched CR. This is done for both the single-lepton and dilepton channels, respectively.

In the following chapter, the procedure to derive these reweightings and their application and effect on H_{T}-sensitive variables will be discussed.

13.1 DERIVATION PROCEDURE OF THE REWEIGHTING FACTORS

To derive the correction factors in a data-driven way, an orthogonal CR, which does not overlap with the analysis phase space, needs to be defined. Additionally, the region should be dominated by the processes that are to be corrected, $t\bar{t}+$ light and $t\bar{t}+\geq 1c$. To achieve this, different requirements on the number of b-tagged jets are chosen to select an orthogonal but $t\bar{t}+$ light and $t\bar{t}+\geq 1c$ enriched region. The requirements for both the single-lepton and dilepton channels are listed in Table 13.1.

Table 13.1.: Definition of the orthogonal CRs for the derivation of the H_{T} correction factors. To define if a jet is b-tagged or not, the 70 % DL1r WP is used.

Channel	N_{Jets} Requirement	$N_{b-\mathrm{Tag}}$ Requirement	N_{Jets} Sub-Categories
Single-lepton	≥ 5	$\stackrel{!}{=} 2$	$\stackrel{!}{=} 5$ $\stackrel{!}{=} 6$ $\stackrel{!}{=} 7$ ≥ 8
Dilepton	≥ 3	$\stackrel{!}{=} 2$	$\stackrel{!}{=} 3$ $\stackrel{!}{=} 4$ $\stackrel{!}{=} 5$ ≥ 6

Table 13.2.: Post-fit normalisation factors used for deriving the H_{T} reweighting factors. The normalisation factors were derived using a dedicated background-only fit per channel of the analysis CRs and unblinded SRs.

Channel	$t\bar{t}+\mathrm{jets}$ Process	Post-Fit Normalisation Factor
Single-lepton	$t\bar{t}+\mathrm{light}$	$0.74\,(+0.10/-0.10)$
	$t\bar{t}+\geq 1c$	$1.79\,(+0.25/-0.23)$
	$t\bar{t}+b$	$1.08\,(+0.14/-0.13)$
	$t\bar{t}+B$	$1.08\,(+0.14/-0.13)$
	$t\bar{t}+\geq 2b$	$0.96\,(+0.14/-0.12)$
Dilepton	$t\bar{t}+\mathrm{light}$	$0.85\,(+0.17/-0.15)$
	$t\bar{t}+\geq 1c$	$1.59\,(+0.22/-0.20)$
	$t\bar{t}+b$	$1.29\,(+0.22/-0.20)$
	$t\bar{t}+B$	$1.29\,(+0.22/-0.20)$
	$t\bar{t}+\geq 2b$	$0.93\,(+0.15/-0.13)$

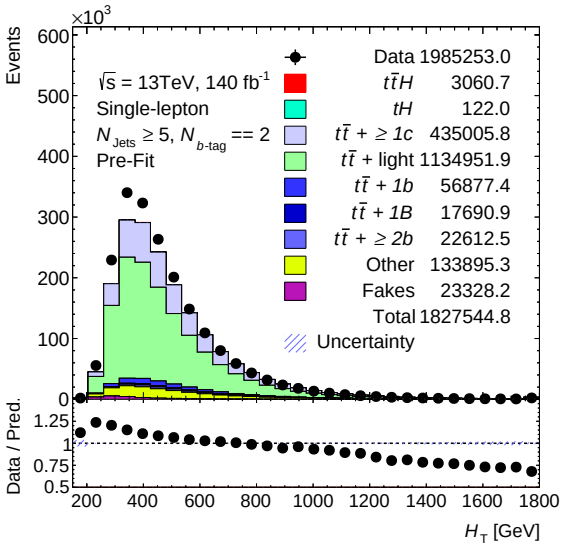
Both regions are selected to be relatively close to their analysis counterparts to ensure that the derived reweightings are effective in correcting the shape mis-modelling in the analysis region. Additionally, the regions are further split into sub-categories based on the number of jets, N_{Jets}, in the events. This is necessary due to the very sensitive nature of H_{T} towards this variable.

With these new CRs, one binning per sub-category is calculated using the $t\bar{t}+\mathrm{light}$ and $t\bar{t}+\geq 1c$ MC samples. The binning needs to adhere to two main criteria: Firstly, the number of events per bin must be at least 5000, and secondly, the bins must be at least $50\,\mathrm{GeV}$ wide. These criteria were chosen to minimise the effect of fluctuations and to maximise the stability of the reweightings.
Afterwards, to avoid correcting the normalisation issues mentioned at the beginning of this chapter, post-fit normalisation factors are applied to each of the $t\bar{t}+\mathrm{jets}$ contributions. These factors are derived using a dedicated background-only fit per channel in the analysis CRs and unblinded SRs. The resulting normalisation factors for each channel and $t\bar{t}+\mathrm{jets}$ contribution are listed in Table 13.2.

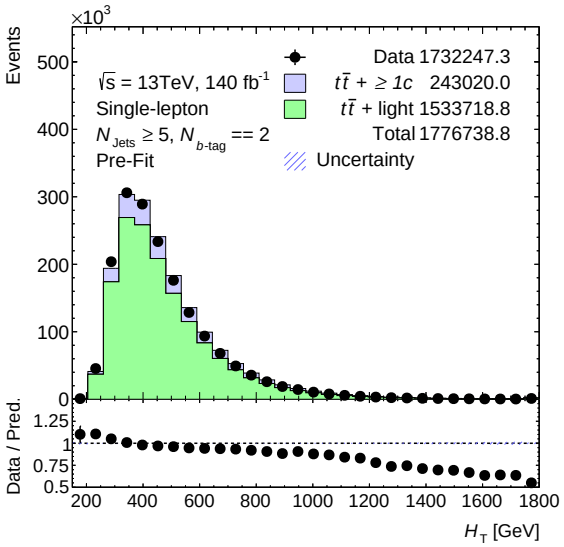
Afterwards, all non-to-be reweighted MC samples, including $t\bar{t}+b$, $t\bar{t}+B$, $t\bar{t}+\geq 2b$, and all other contributions besides $t\bar{t}+\mathrm{light}$ and $t\bar{t}+\geq 1c$, are subtracted bin-by-bin from data according to their corresponding MC prediction. The resulting histograms with data and MC, which should only contain contributions from $t\bar{t}+\mathrm{light}$ and $t\bar{t}+\geq 1c$, should have the same shape mis-modelling as before subtraction. To visualise this, the H_{T} distributions before and after subtracting all non-to-be reweighted contributions are shown in Figure 13.1.

As clearly visible, the mis-modelled shape in the distribution is similar before and after subtracting all non-to-be reweighted MC contributions, further consolidating the assumption that the mis-modelling is arising from the $t\bar{t}+\mathrm{light}$ and $t\bar{t}+\geq 1c$ contributions.

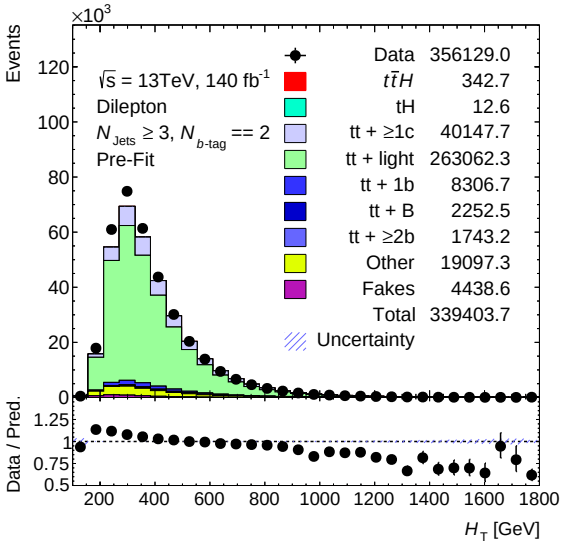
Following the subtraction, the resulting histograms are normalised, and the reweighting factors k_i are calculated using Equation (13.1):



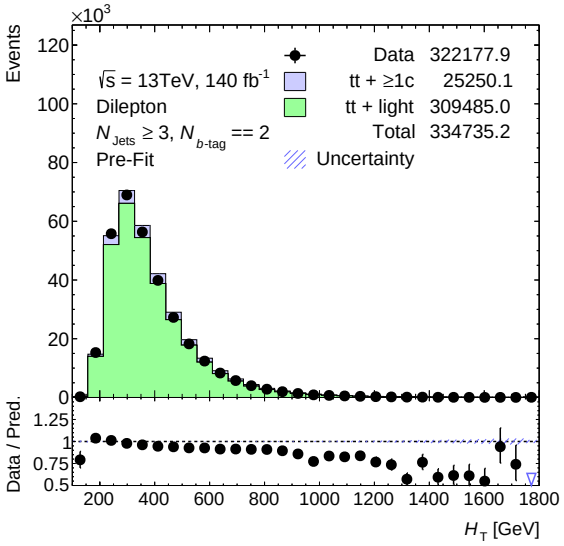
(a): Single-lepton before subtraction



(b): Single-lepton after subtraction



(c): Dilepton before subtraction



(d): Dilepton after subtraction

Figure 13.1.: H_{T} distributions before and after subtracting all non-to-be reweighted MC contributions for both the single-lepton and dilepton channel. No post-fit normalisation factors are applied.

$$k_i = \frac{N_i^{\mathrm{data}}}{N_i^{\mathrm{MC}}} \tag{13.1}$$

where k_i is the reweighting factor for the i-th bin, and $N_i^{\mathrm{data/MC}}$ is the number of data/MC events in the i-th bin. Additionally, to achieve a smooth reweighting and match transitions between consecutive bins, a smoothing function $f\left(H_{\mathrm{T}}\right)$ is used to derive per-event correction weights instead of constant factors. The smoothing function, defined in Equation (13.2), is a spline composed of 2nd order polynomials, one for each bin:

$$f\left(H_{\mathrm{T}}\right) = a_i H_{\mathrm{T}}^2 + b_i H_{\mathrm{T}} + c_i \tag{13.2}$$

a_i, b_i, and c_i are the polynomial coefficients for the i-th bin, which need to adhere to the condition that each function is continuous and has a continuous first derivative. Furthermore, the number of MC events in each bin should remain the same as when using constant reweighting factors, as expressed in Equation (13.3):

$$k_i \sum_j w_j = \sum_j \left(a_i H_{\mathrm{T}}^2 + b_i H_{\mathrm{T}} + c_i \right) w_j \tag{13.3}$$

where the sum runs over all events j in the i-th bin. Under these conditions, a system of linear equations is solved to determine the polynomial coefficients for each bin. The resulting splines for each jet multiplicity and channel are then used to calculate the per-event weights for the $t\bar{t}+$ light and $t\bar{t}+\geq 1c$ MC events.

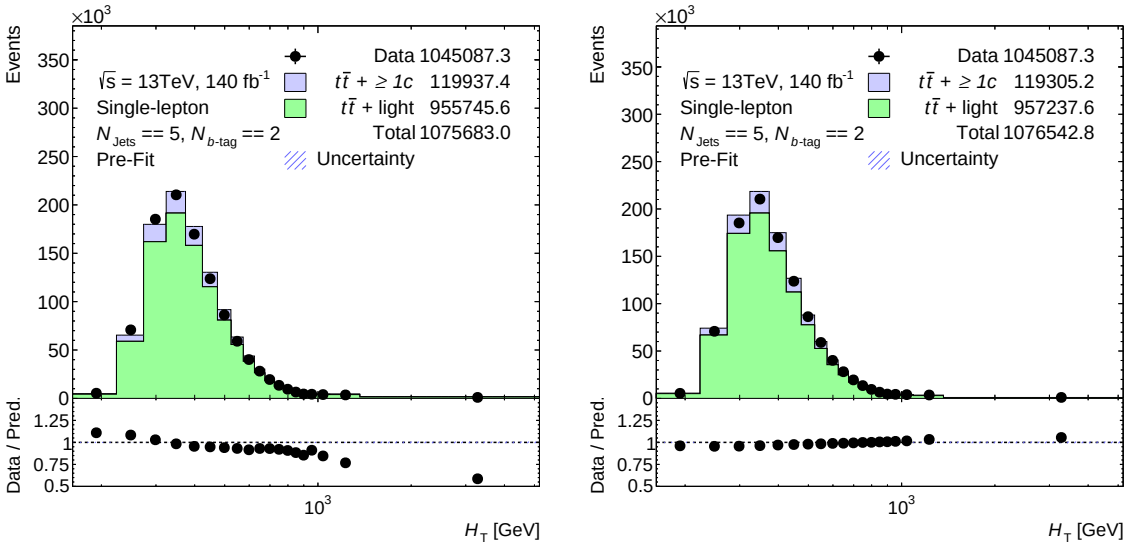
To visualise the different steps and their effects, the resulting distributions are shown in Figure 13.2 for the single-lepton and in Figure 13.3 for the dilepton channel.

When looking at Figure 13.2 (a), a clear shape mis-modelling can be observed in the ratio between data and MC. After rescaling (Figure 13.2 (c)), this mis-modelling becomes even stronger. However, the applied reweighting is performing as expected, correcting the resulting slope from the shape mis-modelling and leaving a simple yield offset, which can easily be addressed in the PLF using normalisation factors. In addition to the final distributions, the reweighting spline can also be visualised, as shown in Figure 13.4 for the single-lepton channel and in Figure 13.5 for the dilepton channel, together with different systematic variations.

The resulting splines show larger weights for the lower H_{T} regions, while lower function values are observed in the higher H_{T} regions. No significant deviation from the nominal reweighting is observed for any of the systematic variations, indicating that the usage of the post-fit normalisation factors is not as important as expected. However, these variations are still used to examine and estimate the systematic uncertainties of the reweighting procedure.

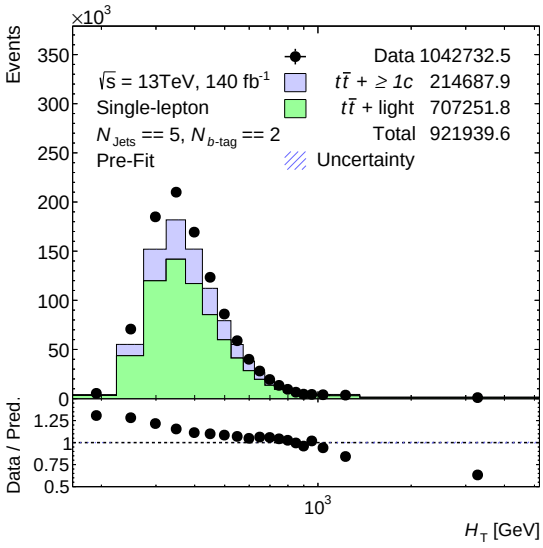
With the per-jet multiplicity reweightings now derived, one can visualise the effect of the reweighting on the jet multiplicity inclusive distributions. These distributions are shown in Figure 13.6 for the single-lepton channel and in Figure 13.7 for the dilepton channel.

While a noticeable effect of removing the shape mis-modelling for the single-lepton channel can be observed, the resulting effect of the reweighting in the dilepton channel is not as strong as for the single-lepton channel. However, an improved shape is still observed.

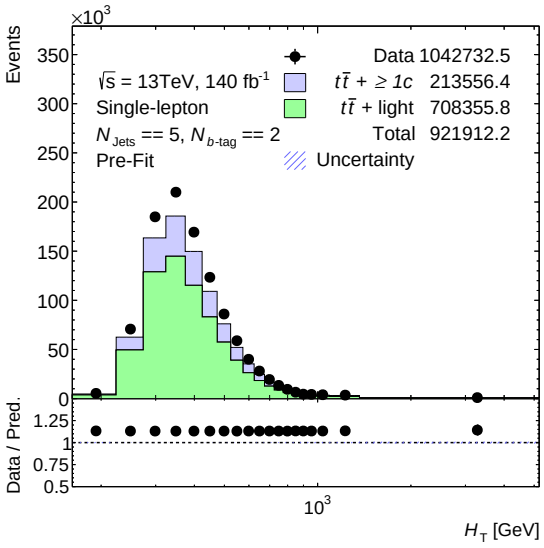


(a): Without rescaling and reweighting

(b): Without rescaling but with reweighting

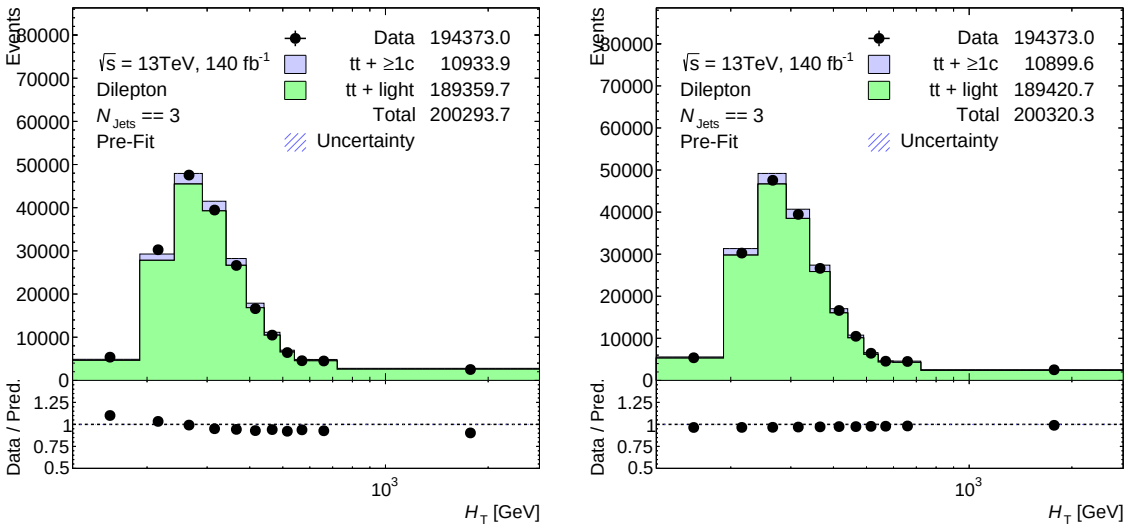


(c): With rescaling but without reweighting



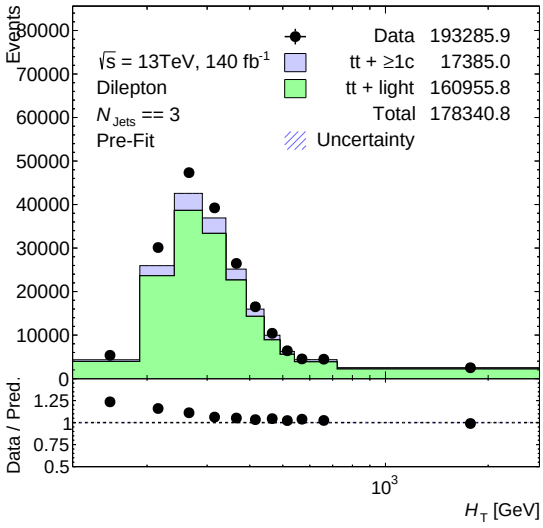
(d): With rescaling and reweighting

Figure 13.2.: H_{T} distributions for the single-lepton channel orthogonal control region for the case of $N_{\mathrm{Jets}} \stackrel{!}{=} 5$. The different figures show the effect of the rescaling with the post-fit normalisation factors and the reweighting factors derived using the smoothed reweighting spline. The yields of all other MC contributions are subtracted from the data yields in a per-bin fashion.

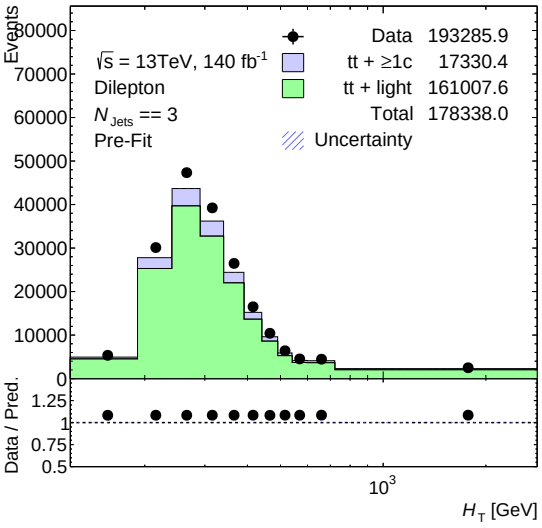


(a): Without rescaling and reweighting

(b): Without rescaling but with reweighting



(c): With rescaling but without reweighting



(d): With rescaling and reweighting

Figure 13.3.: H_T distributions for the dilepton channel orthogonal control region for the case of $N_{\text{Jets}} \stackrel{!}{=} 3$. The different figures show the effect of the rescaling with the post-fit normalisation factors and the reweighting factors derived using the smoothed reweighting spline. The yields of all other MC contributions are subtracted from the data yields in a per-bin fashion.

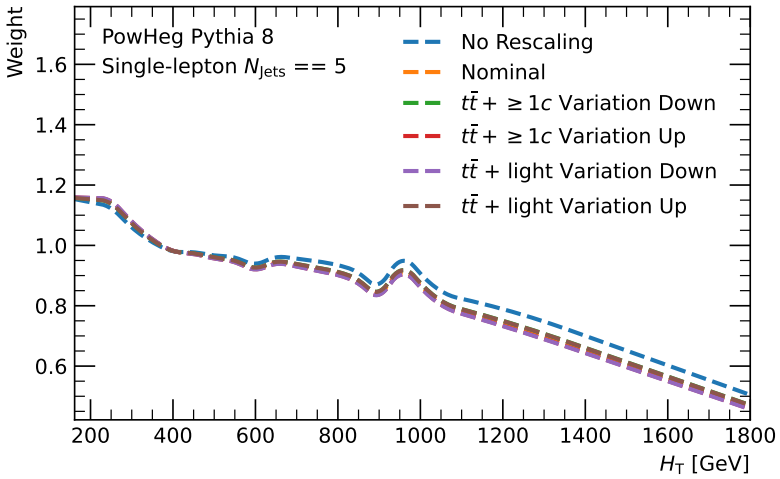


Figure 13.4.: Reweighting spline for the single-lepton channel with $N_{\text{Jets}} \stackrel{!}{=} 5$ reweighting. The nominal reweighting (orange) is compared to different systematic variations, including using no post-fit normalisation factors (blue) and up and down variations, where the respective post-fit normalisation factor is varied.

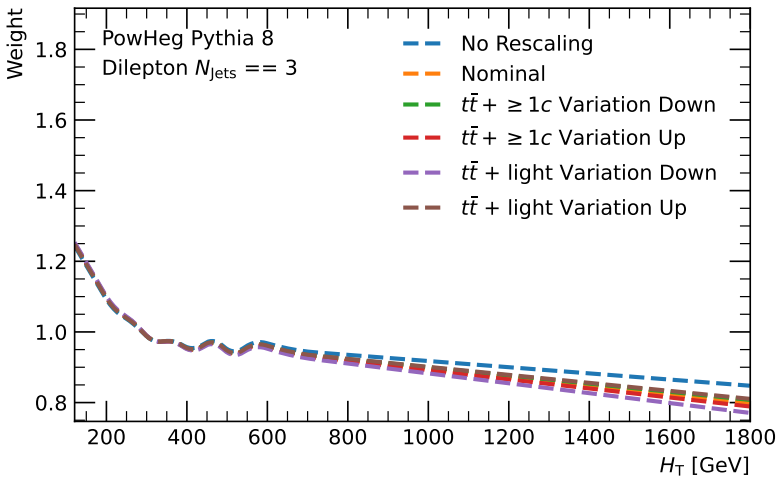
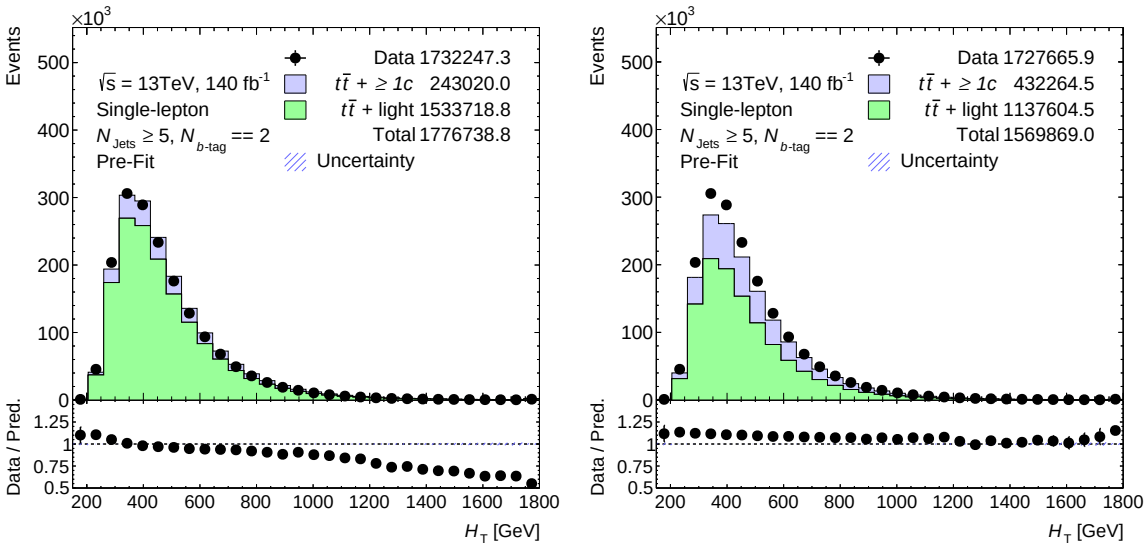
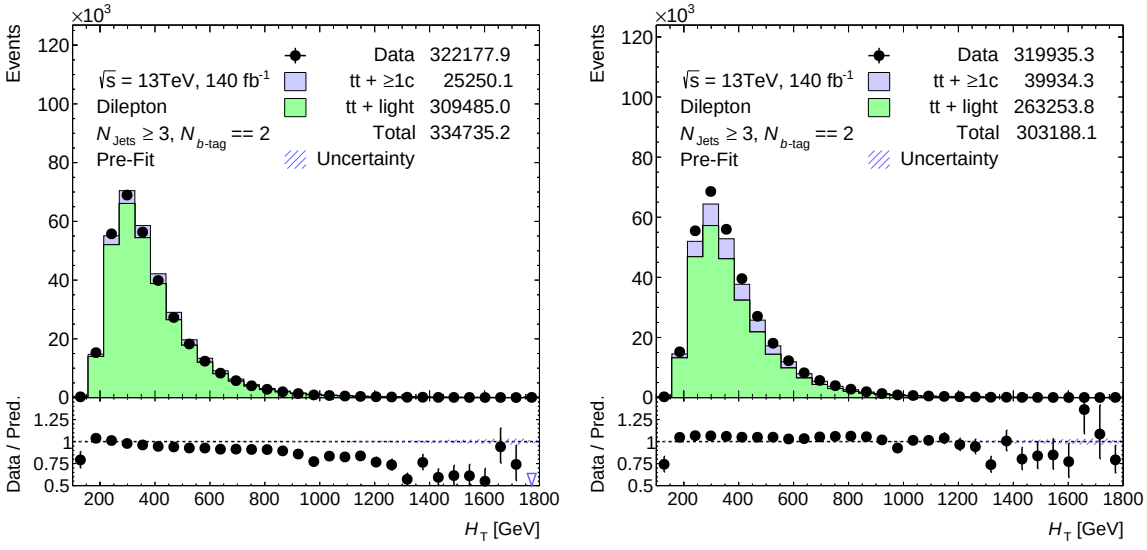


Figure 13.5.: Reweighting spline for the dilepton channel with $N_{\text{Jets}} \stackrel{!}{=} 3$ reweighting. The nominal reweighting (orange) is compared to different systematic variations, including using no post-fit normalisation factors (blue) and up and down variations, where the respective post-fit normalisation factor is varied.



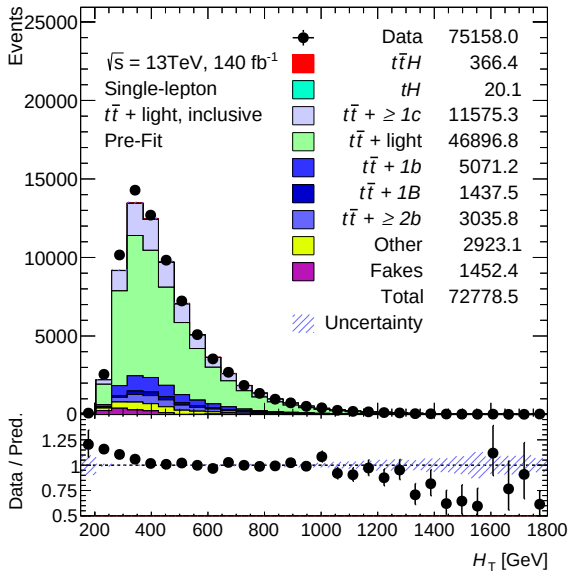
(a): Without rescaling and reweighting (b): With rescaling and reweighting

Figure 13.6.: Jet multiplicity inclusive H_T distributions without rescaling and reweighting (a) and with rescaling and reweighting (b) for the single-lepton channel. The yields of all other MC contributions are subtracted from the data yields in a per-bin fashion.

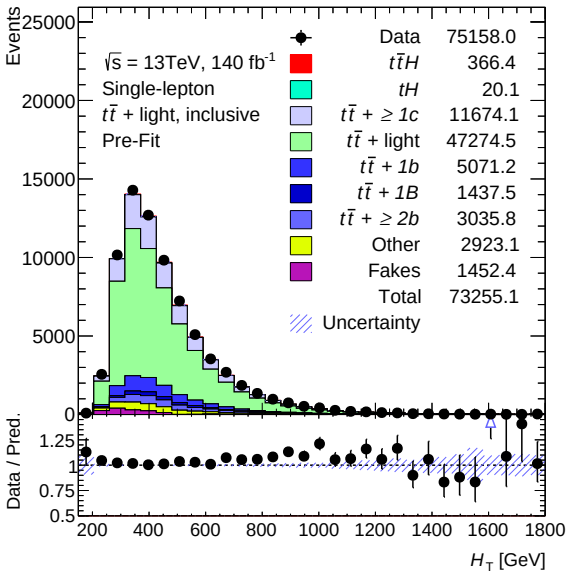


(a): Without rescaling and reweighting (b): With rescaling and reweighting

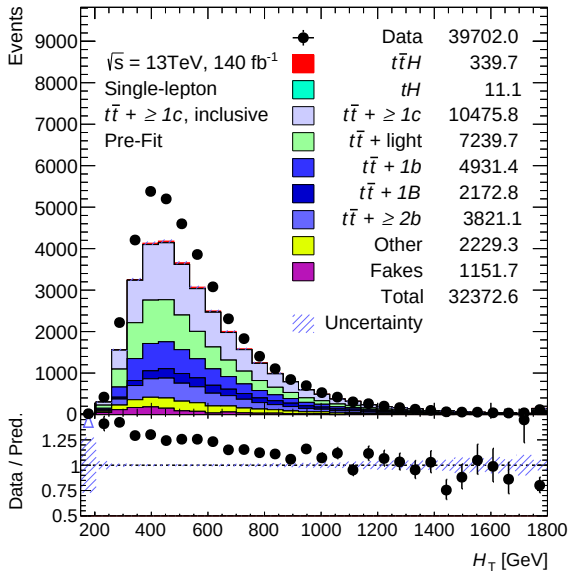
Figure 13.7.: Jet multiplicity inclusive H_T distributions without rescaling and reweighting (a) and with rescaling and reweighting (b) for the dilepton channel. The yields of all other MC contributions are subtracted from the data yields in a per-bin fashion.



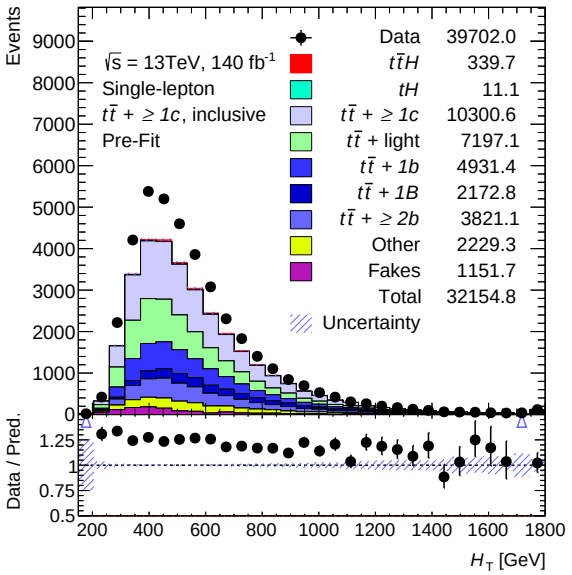
(a): $t\bar{t}+$ light CR without reweighting



(b): $t\bar{t}+$ light CR with reweighting



(c): $t\bar{t}+\geq 1c$ CR without reweighting



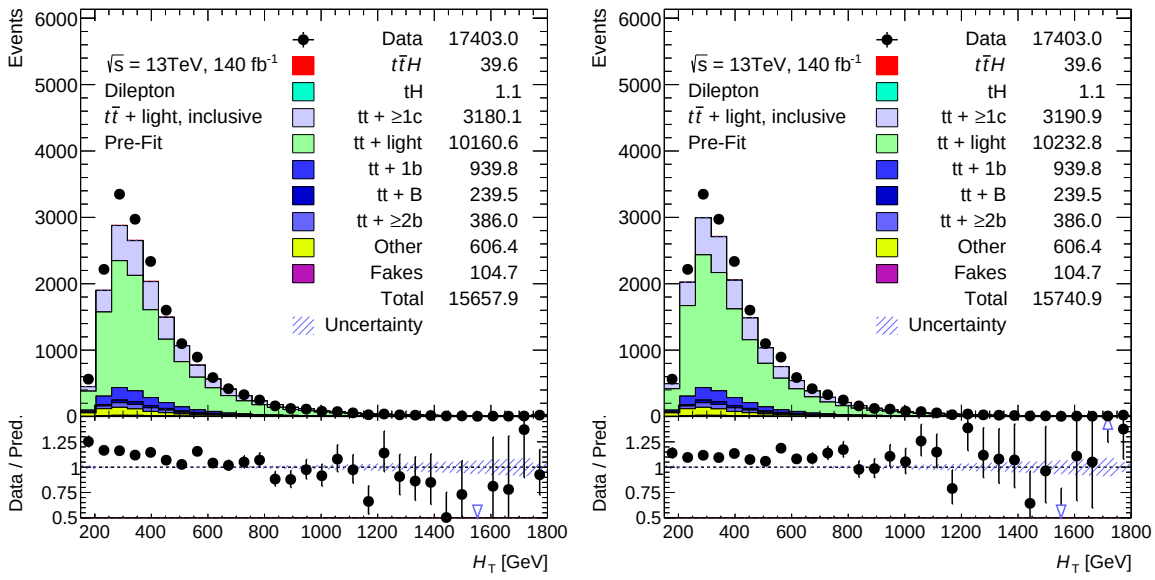
(d): $t\bar{t}+\geq 1c$ CR with reweighting

Figure 13.8.: H_{T} distributions for the single-lepton channel $t\bar{t}+$ light and $t\bar{t}+\geq 1c$ CRs before and after reweighting. No post-fit normalisation factors are applied.

13.2 APPLICATION AND EFFECT IN THE ANALYSIS REGIONS

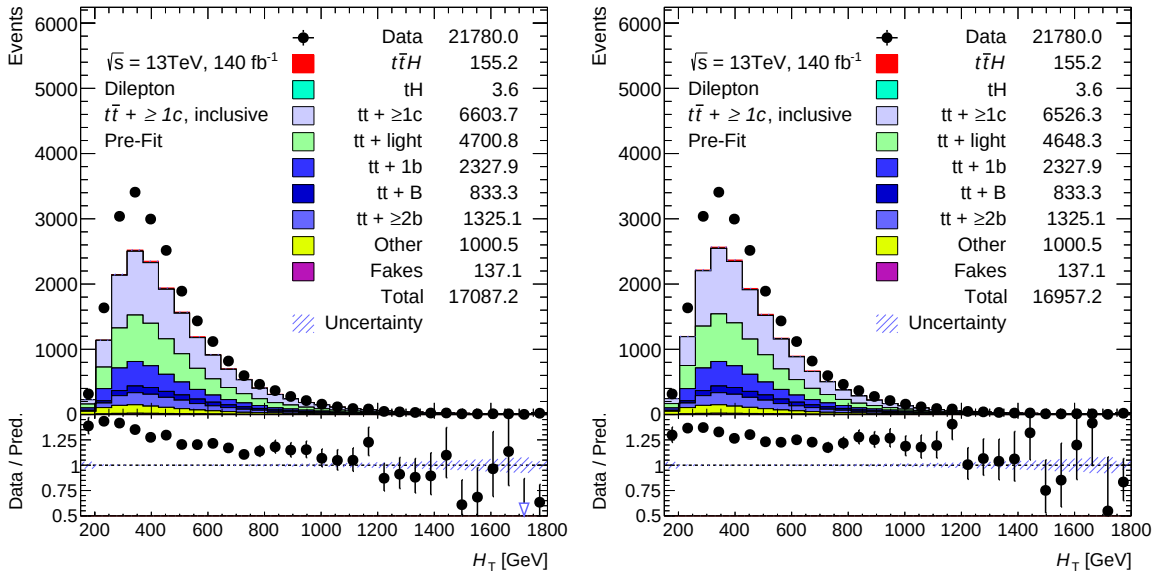
After the derivation of the reweighting, the respective factors can be applied, and their effect observed in the analysis region. The non-reweighted and reweighted distributions of the $t\bar{t}+$ light and $t\bar{t}+\geq 1c$ CRs are shown in Figure 13.8 for the single-lepton channel and in Figure 13.9 for the dilepton channel.

While no significant change is observed for both channels, improvement can still be seen in the lower H_{T} bins. For the single-lepton $t\bar{t}+$ light CR, the first bins up to $400\,\mathrm{GeV}$ have a reduced



(a): $t\bar{t}$ + light CR without reweighting

(b): $t\bar{t}$ + light CR with reweighting



(c): $t\bar{t} + \geq 1c$ CR without reweighting

(d): $t\bar{t} + \geq 1c$ CR with reweighting

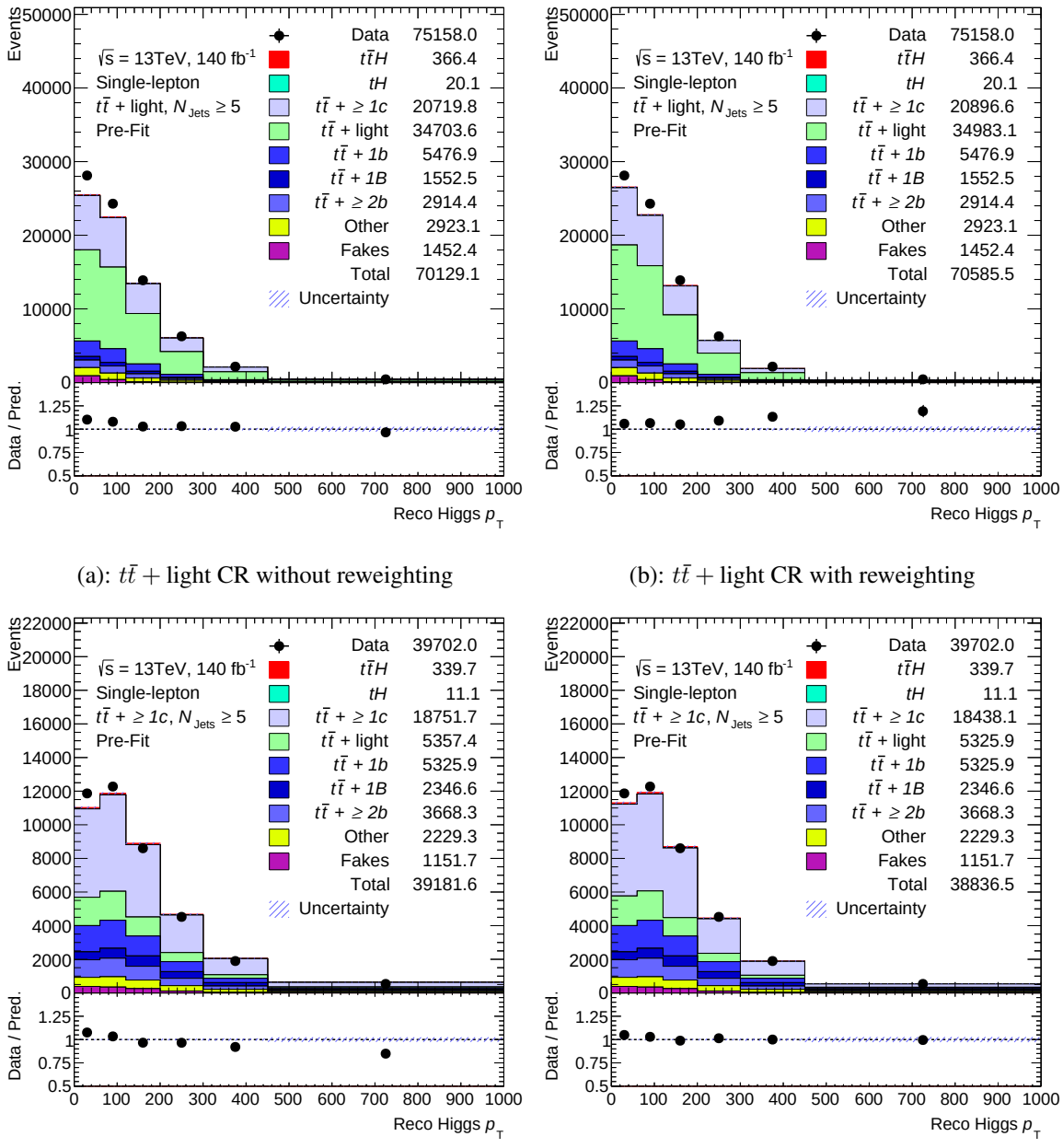
Figure 13.9.: H_{T} distributions for the dilepton channel $t\bar{t}$ + light and $t\bar{t} + \geq 1c$ CRs before and after reweighting. No post-fit normalisation factors are applied.

shape and are more aligned with the data. Similar improvements are observed for the $t\bar{t} + \geq 1c$ CR, particularly for the first bins. Additionally, a shift towards one is observed in the higher H_{T} regions for the $t\bar{t}$ + light CR, suggesting an improvement due to the reweighting.

For the dilepton channel, similar observations can be made for the lower H_{T} bins for both CRs, although much smaller than in the single-lepton channel. In the higher H_{T} region, the shape improvement is insignificant.

13.3 CLOSURE TEST OF THE REWEIGHTING

In addition to the visible effect on the H_{T} distributions in the analysis regions, a closure test is performed to evaluate the effect of the reweighting on other H_{T}-sensitive variables. For this purpose, the reconstructed Higgs p_{T}, p_{T}^{H}, is chosen. Similar to what was shown before, the resulting distributions for p_{T}^{H} for the $t\bar{t}+$ light and $t\bar{t}+\geq 1c$ CRs are shown in Figure 13.10 for the single-lepton channel and in Figure 13.11 for the dilepton channel.



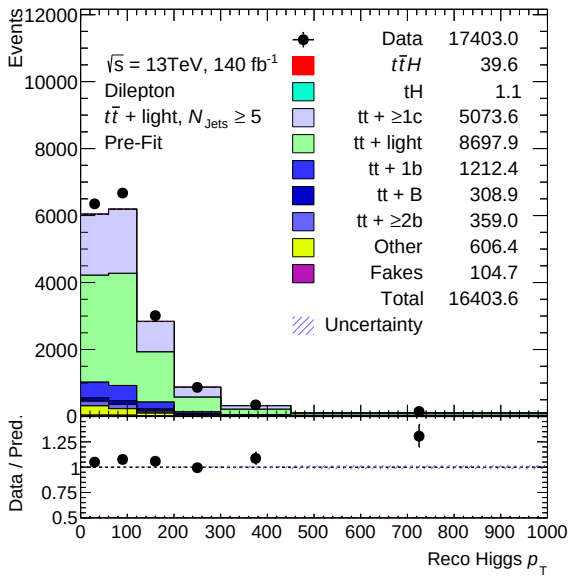
(a): $t\bar{t}+$ light CR without reweighting

(b): $t\bar{t}+$ light CR with reweighting

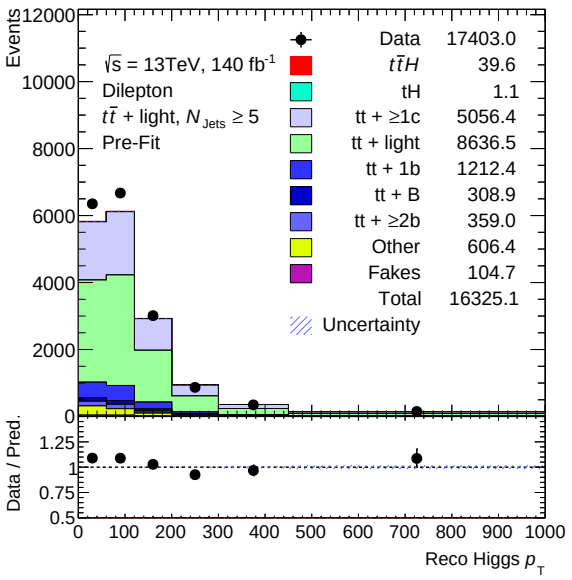
(c): $t\bar{t}+\geq 1c$ CR without reweighting

(d): $t\bar{t}+\geq 1c$ CR with reweighting

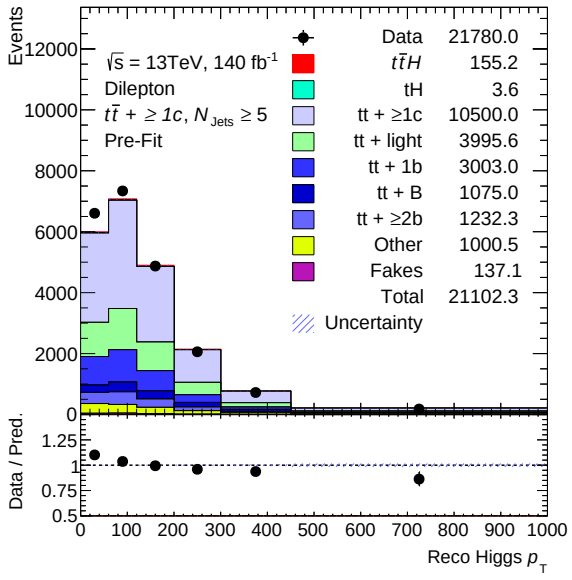
Figure 13.10.: p_{T}^{H} distributions for the single-lepton channel $t\bar{t}+$ light and $t\bar{t}+\geq 1c$ CRs before and after reweighting. No post-fit normalisation factors are applied.



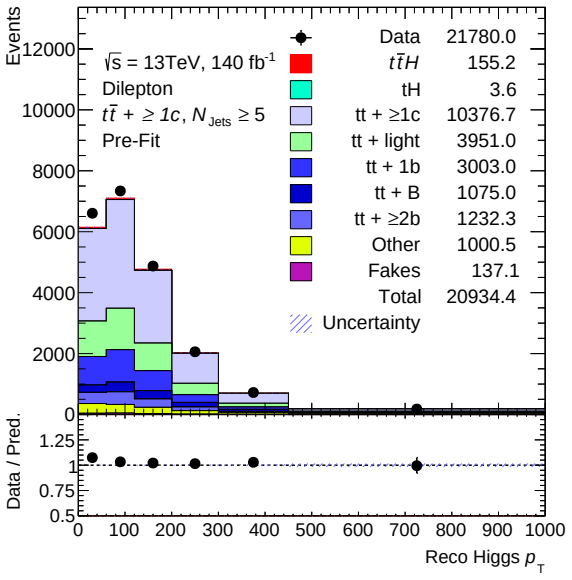
(a): $t\bar{t}$ + light CR without reweighting



(b): $t\bar{t}$ + light CR with reweighting



(c): $t\bar{t}$ + $\geq$1c CR without reweighting



(d): $t\bar{t}$ + $\geq$1c CR with reweighting

Figure 13.11.: p_T^H distributions for the dilepton channel $t\bar{t}$ + light and $t\bar{t}$ + $\geq$1c control regions before and after reweighting. No post-fit normalisation factors are applied.

For the single-lepton $t\bar{t}$ + light CR, a slight improvement is visible for the first bins, while a worsening is observed in the higher p_T^H bins. This can be explained by the missing post-fit normalisation factors. When looking at Figure 13.2 (b), a similar effect is visible when only applying the reweighting but not the corresponding post-fit normalisation factors, which were used to derive the reweighting. These factors are not applied here due to their re-derivation and application in the PLF. However, due to this observation and the improvement of the shape shown in Figure 13.2 (d), the reweighting has a beneficial impact on the shape of p_T^H. For the $t\bar{t}$ + $\geq$1c CR, the observed impact is even greater, nearly completely removing the shape mis-modelling when applying the reweighting.

For the dilepton channel, although the H_{T} distributions showed no significant improvement, the effect of the reweighting on p_{T}^{H} is visible in both the $t\bar{t}+$ light and $t\bar{t}+\geq 1c$ CRs. In both CRs, shape mis-modelling is mitigated and nearly completely removed when applying the reweighting.

In total, the derived reweightings for $t\bar{t}+$ light and $t\bar{t}+\geq 1c$ seem to have a positive impact on the H_{T}-sensitive variables, improving the observed mis-modelling and partly removing the resulting slope.

14

ANALYSIS RESULTS

After introducing the general idea, strategy and methodology behind the $t\bar{t}H(H \rightarrow b\bar{b})$ legacy analysis, two types of fits were performed. These fit types, already mentioned in Section 11.3, are background-only and Asimov fits. Both fit types are performed blinded, due to the blinded status of the $t\bar{t}H(H \rightarrow b\bar{b})$ legacy analysis at the time of writing. The blinding threshold of the signal-to-background ratio was set to $7.7\,\%$, denoting that for the background-only fits, only bins with a signal-to-background ratio below this value are considered.

In the following chapter, the results of the per-channel and combined Asimov and background-only fits will be presented.

14.1 ASIMOV FITS

To assess the expected sensitivity and constraints on NPs, Asimov fits were conducted using the complete statistical model, encompassing the signal, applied to an Asimov dataset. These fits incorporated all systematic uncertainties outlined in Section 11.4 as NPs. While various NPs were considered, a pruning procedure was implemented. Any NPs with an impact below $0.5\,\%$ were pruned and excluded from further consideration. This streamlined approach significantly decreased the computational time needed for the fit to converge. Additionally, different smoothing methods are applied to MC samples with low statistics in some regions to avoid large statistical fluctuations.
To estimate the expected uncertainties on the signal strength parameters and the resulting significances, two types of Asimov fits were performed. Inclusive fits, combining the signal strength POIs of all STXS regions into one POI, $\mu_{t\bar{t}H}$, and STXS fits, where for each of the STXS regions, one signal strength parameter is fitted.

Figure 14.1 shows the best-fit results for the different normalisation factors for all three fits[1]. While both individual channels have higher uncertainties, the combined channel, as expected, yields the smallest uncertainties on the normalisation factors, ranging from $\approx 8\,\%$ for $t\bar{t}+$ light up to $\approx 14\,\%$ for $t\bar{t}+B$ processes. In comparison to the last analysis, which only used a free-floating normalisation factor for $t\bar{t}+\geq 1b$ with an uncertainty of ± 0.08, $100\,\%$ normalisation uncertainty for $t\bar{t}+\geq 1c$, and $\pm 6\,\%$ cross-section uncertainty for $t\bar{t}+$ light, similar or smaller uncertainties for the normalisation factors are expected, even when further dividing the $t\bar{t}+\geq 1b$ processes.

1 By construction, the best-fit value for Asimov fits is 1.

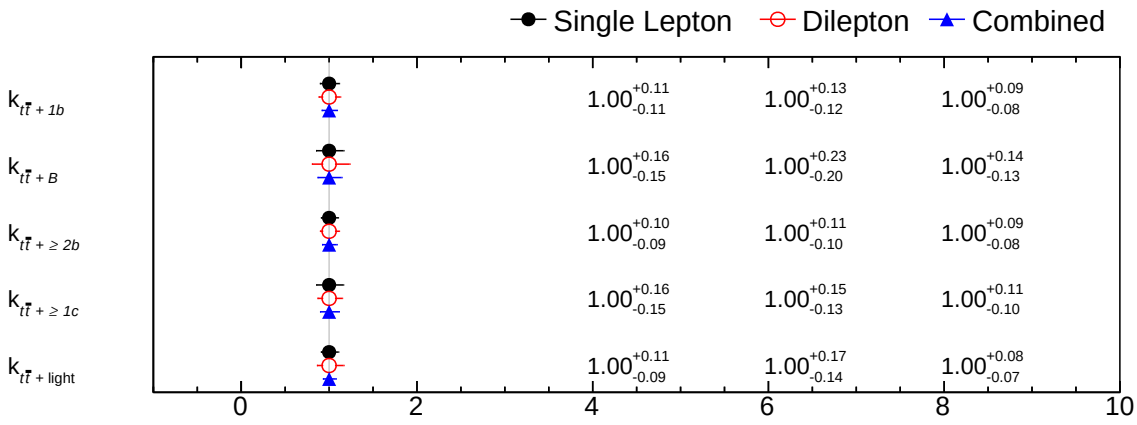


Figure 14.1.: Best-fit results from Asimov fits for the free-floating normalisation factors. The three fits shown, single-lepton (black), dilepton (red), and combined (black) show their respective result.

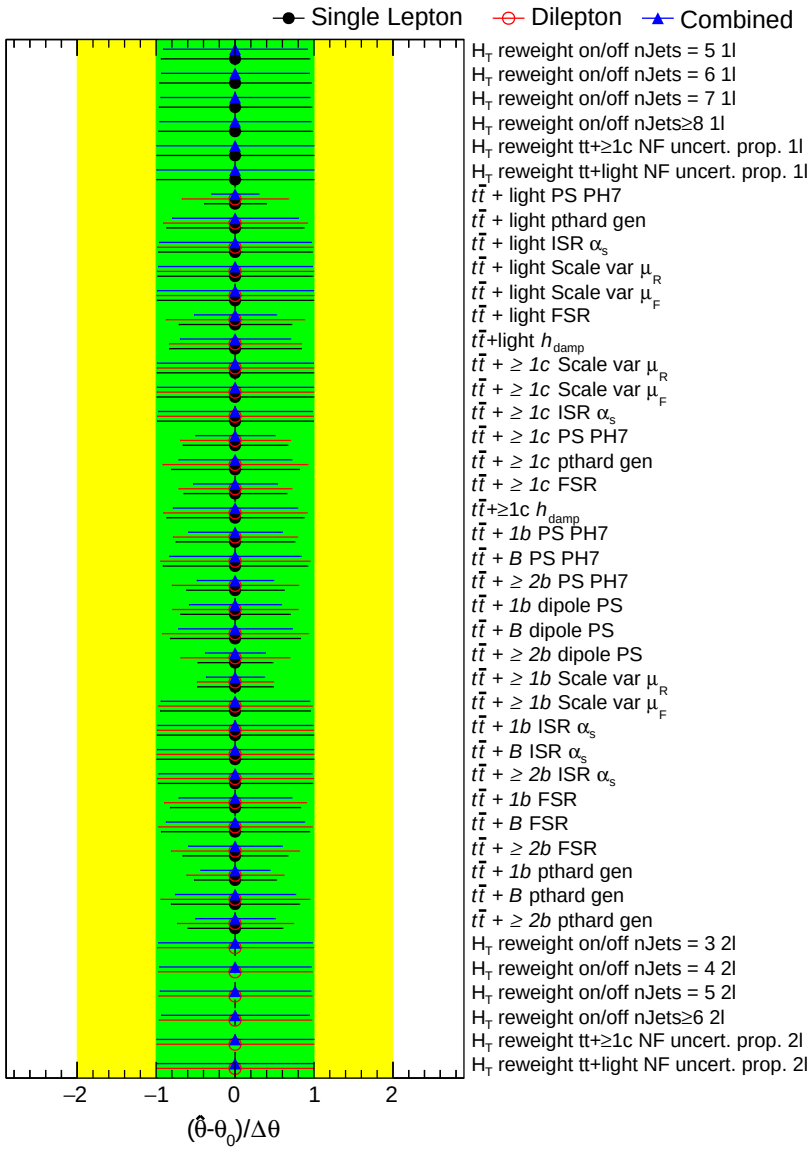


Figure 14.2.: $t\bar{t}$ modelling NPs constraints for the different Asimov fits.

Due to their dominant impact on the systematic uncertainty of the last analysis result, the fitted $t\bar{t}$ modelling NPs and their constraints are depicted in Figure 14.2. No significant constraints are visible for any of the H_{T} rewriting systematic variations, while larger constraints can be observed for the different $t\bar{t}$ PS and FSR variations. For all $t\bar{t}$ sub-categories, the corresponding NPs are clearly more constrained by the different fits than others.
In addition to these, further constraints can be observed for the $t\bar{t} + \geq 1b$ renormalisation scale variation and different NLO matching systematics, denoted with "pthard gen". All other modelling-related NPs show no significant constraints and are therefore depicted in Appendix A.5. A similar assessment holds true for the instrumental or experimental NPs, which are also shown in Appendix A.5.

14.1.1 *STXS Results*

The expected uncertainties for the STXS signal strength POIs for the combined fit are visualised in Figure 14.3.

For all STXS regions, the statistical uncertainties are dominant with respect to the systematic uncertainties, which shows the improvement provided by the changes introduced for the legacy analysis when comparing to the previous analysis results (Figure 11.3 (a)). For all systematic and total uncertainties in the STXS regions, an improvement can be observed with respect to the previous analysis, even with the split of the previous first STXS region ($0 - 120\,\mathrm{GeV}$) into two regions (STXS 1 and 2). The largest uncertainties are observed in the STXS 6 region, while the smallest uncertainties are

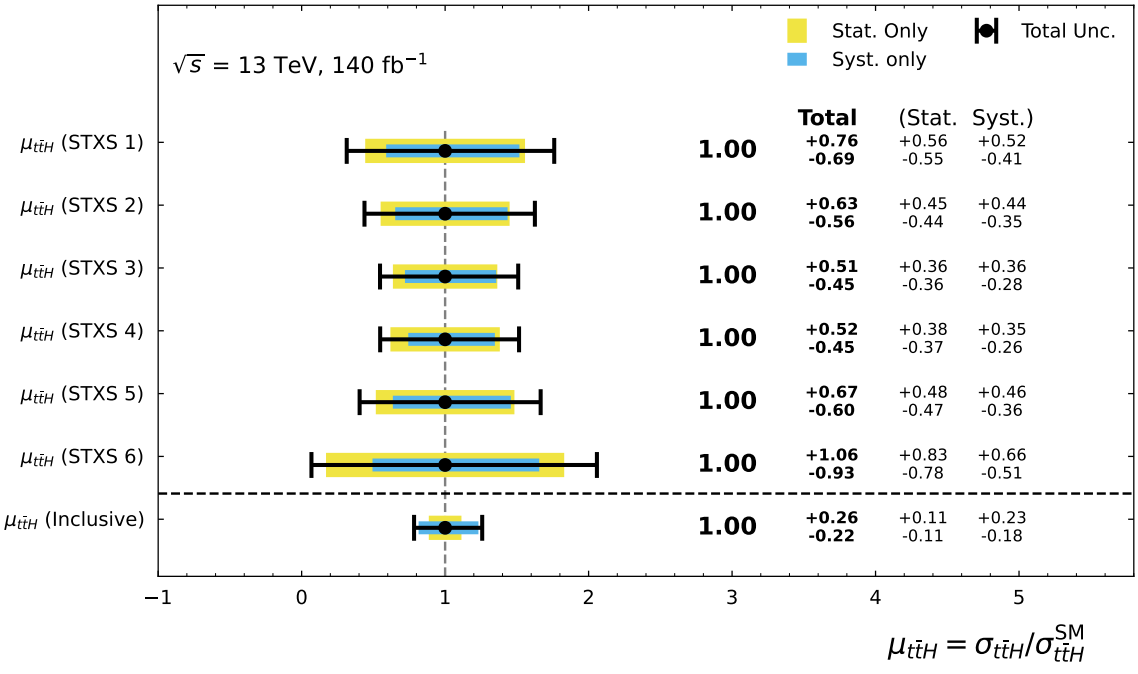


Figure 14.3.: Best-fit signal strength POIs for the different STXS regions and the inclusive signal strength and their expected uncertainties for the combined fit.

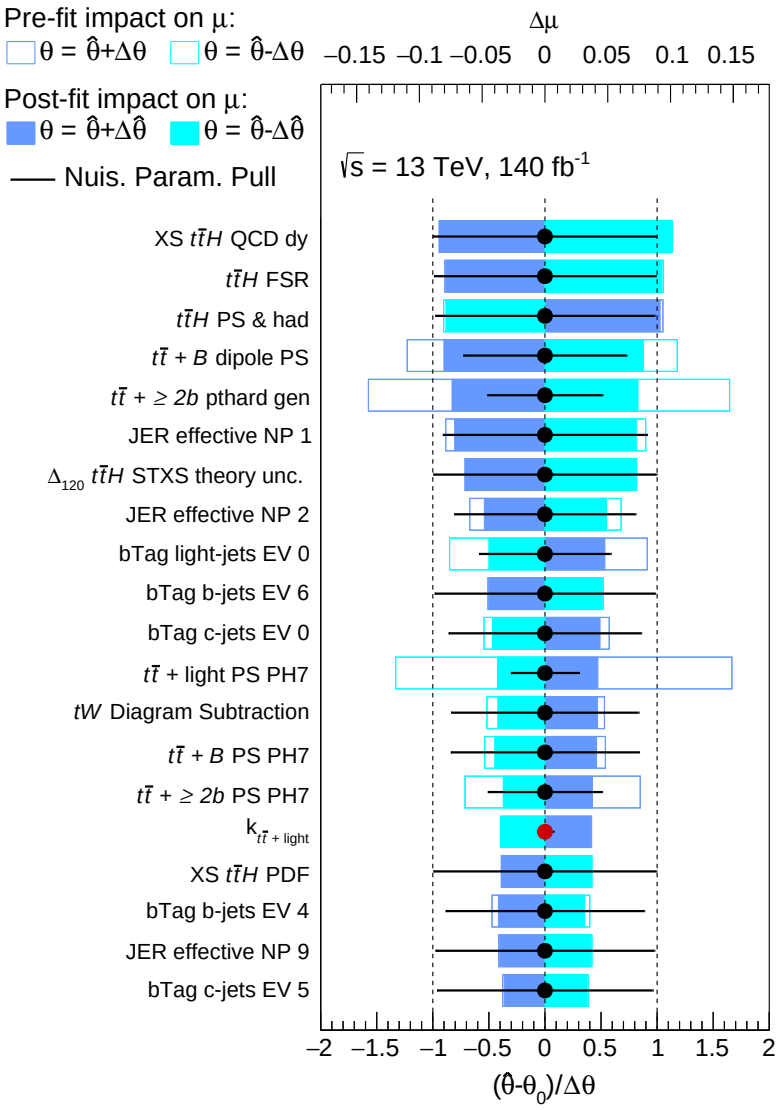


Figure 14.4.: Ranking of the 20 NPs with the largest post-fit impact on $\mu_{t\bar{t}H}$ (STXS 3). The empty boxes visualise the pre-fit, and the filled boxes the post-fit impact on $\mu_{t\bar{t}H}$ (STXS 3). While black dots are used for NPs, red dots are employed for normalisation factors.

expected for the STXS 3 and 4 regions. The results for the per-channel fits are shown in Figure A.22.

To evaluate and examine the origins of the systematic uncertainties for each STXS region further, the 20 NPs contributing the most to the overall systematic uncertainty are shown in Figure 14.4 for $\mu_{t\bar{t}H}$ (STXS 3) due to its lowest expected uncertainties. The so-called ranking plots for the other signal strength parameters are depicted Figures A.23–A.27. The rankings are calculated by comparing the nominal best-fit value of the respective μ to the result when fixing the NP to its best-fit value $\hat{\theta}$, shifted by its pre-fit value $\Delta\theta$. A similar procedure is done for the post-fit $\Delta\hat{\theta}$.

For $\mu_{t\bar{t}H}$ (STXS 3), the largest impacts stem from the different modelling systematics for the signal process, followed by modelling systematics of different $t\bar{t} + \geq 1b$ processes. In addition to modelling systematics, also experimental uncertainties are listed, like different b-tagging eigenvalue, which originate from the b-tagging calibration, the normalisation factor for $t\bar{t} +$ light, $k_{t\bar{t}+\text{light}}$, and

multiple JER-related systematics. Furthermore, other modelling and experimental systematics are listed, including the tW diagram subtraction variation.

14.1.2 *Inclusive Results*

For the inclusive fits, the expected uncertainties for the inclusive signal strength $\mu_{t\bar{t}H}$ for the combined fit are visualised also in Figure 14.3. For the inclusive case, the systematic uncertainty is dominant with respect to the statistical uncertainty. With an overall uncertainty of $+0.26/-0.22$ on $\mu_{t\bar{t}H}$ compared to the previous result of $+0.36/-0.34$, a clear improvement of the expected uncertainty is achieved. Further, the expected significance compared the previous expected significance of $2.7\,\sigma$ is drastically enhanced, yielding $5.5\,\sigma$, which is above the $5\,\sigma$ threshold for a discovery.

The 20 NPs contributing most to the overall systematic uncertainty on $\mu_{t\bar{t}H}$ are shown in Figure 14.5 to further examine the origins of the dominant systematic uncertainty.

In comparison to the systematic uncertainties of the previous version of this analysis, shown in Figure 11.1, the previously dominant modelling uncertainties on $t\bar{t}+\geq 1b$ are reduced and only placed fourth and lower in the ranking. The now dominant uncertainties, placed first to third, are signal modelling uncertainties, arising from cross-section and FSR variations. The listed instrumental uncertainties are related mainly to b-tagging, including different eigenvalues 0 stemming from the corresponding calibrations. Other modelling systematics with large post-fit impacts include tW diagram subtraction, which is dominant in the higher STXS regions (see Figure A.27), and ISR ($t\bar{t}+\geq 1c$ h_{damp}), FSR ($t\bar{t}+\geq 1c$ FSR), and PS ($t\bar{t}+\geq 1c$ PS PH7) variations for the $t\bar{t}+\geq 1c$ process.

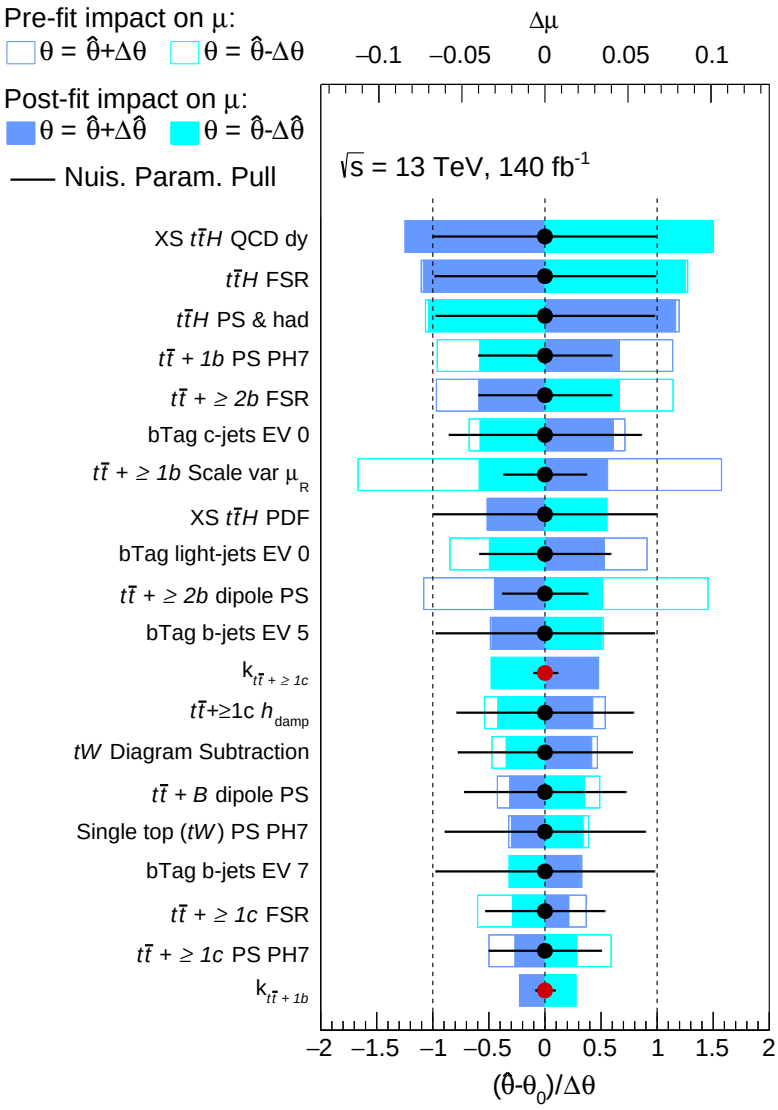


Figure 14.5.: Ranking of the 20 NPs with the largest post-fit impact on $\mu_{t\bar{t}H}$. The empty boxes visualise the pre-fit, and the filled boxes the post-fit impact on $\mu_{t\bar{t}H}$. While black dots are used for NPs, red dots are employed for normalisation factors.

14.2 BACKGROUND-ONLY FITS

Dedicated background-only fits are conducted to test and validate the background modelling in all considered analysis regions. As only background contributions are considered and no signal strength parameters are extracted, there is no need for a split in STXS or inclusive measurement.
Figure 14.6 shows the resulting normalisation factors for three different fits, including the two per-channel fits, single-lepton and dilepton, and the combined fit.

Acceptable agreement between the different channels is observed for all normalisation factors, although some high values emerge for the $t\bar{t} + \geq 1c$ normalisation, exceeding the expectations in the single-lepton channel. Within the uncertainties, the $t\bar{t} + \geq 2b$ and $t\bar{t} + B$ normalisation factors are in agreement with one, while the $t\bar{t} + \geq 1c$, $t\bar{t} + b$, and $t\bar{t} +$ light normalisation factors deviate slightly. Compared to the previous fully unblinded fit results, similar values for $k_{t\bar{t}+b}$ are observed while for

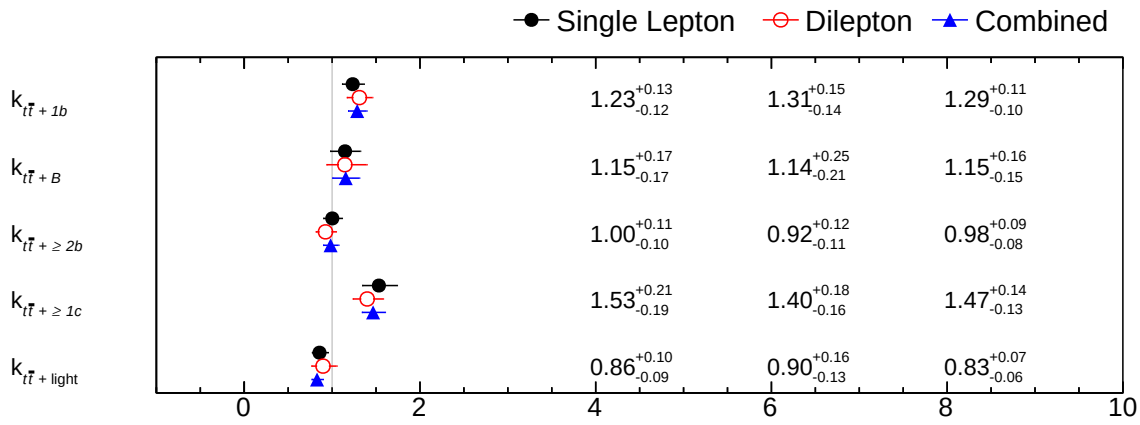


Figure 14.6.: Best-fit results from background-only fits for the free-floating normalisation factors. Three fits, including single-lepton (black), dilepton (red) and combined (black) are shown with their respective results.

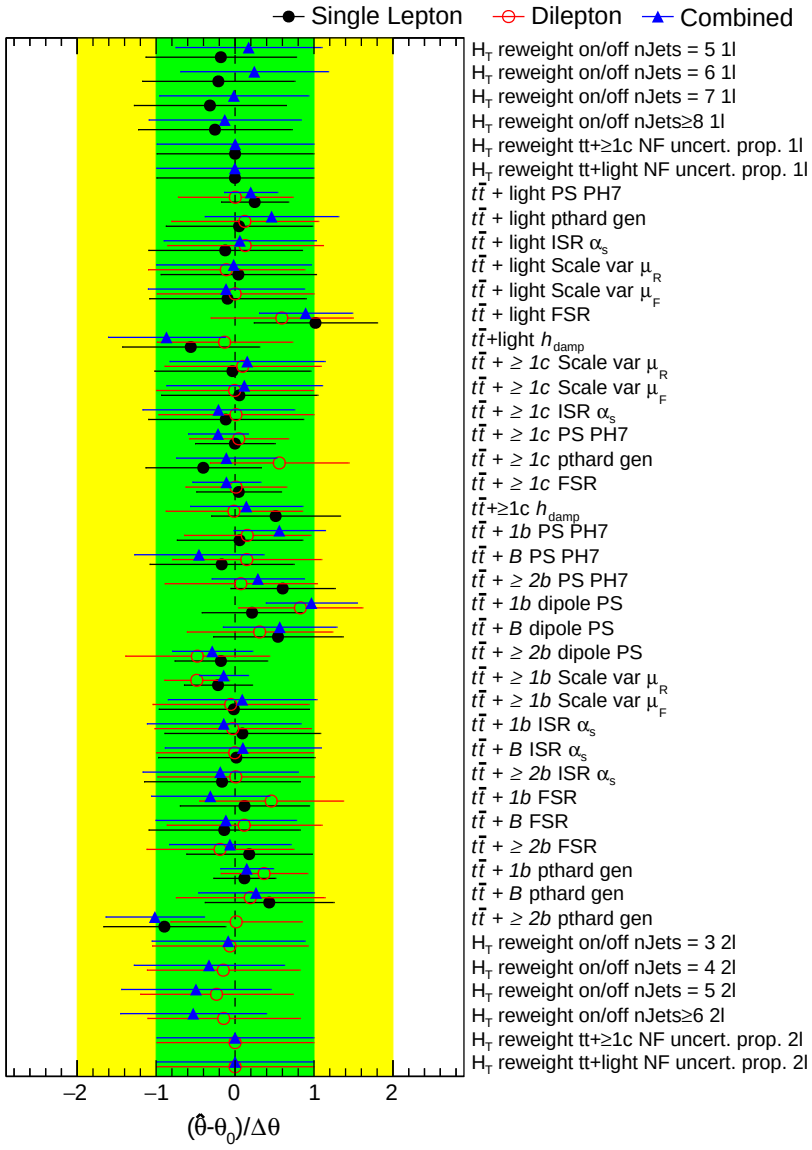


Figure 14.7.: $t\bar{t}$ modelling NPs pulls for the different background-only fits.

both $k_{t\bar{t}+B}$ and $k_{t\bar{t}+\geq 2b}$ smaller values are observed. Both values are in better agreement with one as the previous results.

In Figure 14.7, the fitted $t\bar{t}$ modelling NPs and their pulls are depicted. The most concerning pulls are observed for the different PS and FSR NPs, including $t\bar{t}+$ light FSR and $t\bar{t}+b$ dipole PS. Additionally, the NLO matching systematic for $t\bar{t}+\geq 2b$ ($t\bar{t}+\geq 2b$ pthard gen) is pulled in both the single-lepton and combined fit, but not in the dilepton fit. These NPs are also ranked high in the different ranking plots from the Asimov fits (Figures 14.4 and 14.5) and need to be studied in more detail to fully understand the origin of the pulls.

In addition to the pulls, large constraints are observed also for the $t\bar{t}+$ light PS variation, which were already observed in the Asimov fits and are currently under investigation.

For the H_{T} reweighting, pulls of up to $0.5\,\sigma$ are observed in both channels with no significant constraints. While for the dilepton H_{T} reweighting, both the individual fit and the combined fit are pulled towards negative values, the single-lepton H_{T} reweightings are pulled in different directions, but with no significant effect.

To validate the post-fit yields in each of the regions, the data/MC comparison is shown in Figure 14.8 for the single-lepton and dilepton channels.

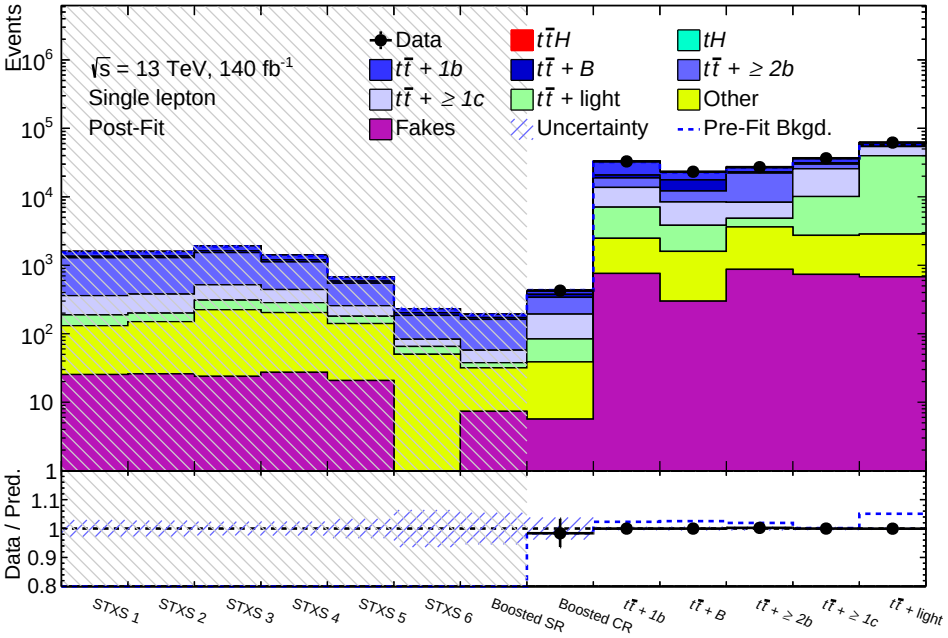
Good post-fit agreement of the MC with the data is observed in both channels, with a slightly worse post-fit agreement compared to the pre-fit agreement in the boosted regions.

In Figure 14.9, the fitted other process modelling NPs and their pulls are depicted. No significant pulls or constraints are observed. However, for the fakes normalisation uncertainty, pulls of up to $-0.5\,\sigma$ are observed. All other modelling-related NPs show no significant constraints or pulls and are therefore depicted in Figures A.28 and A.29.

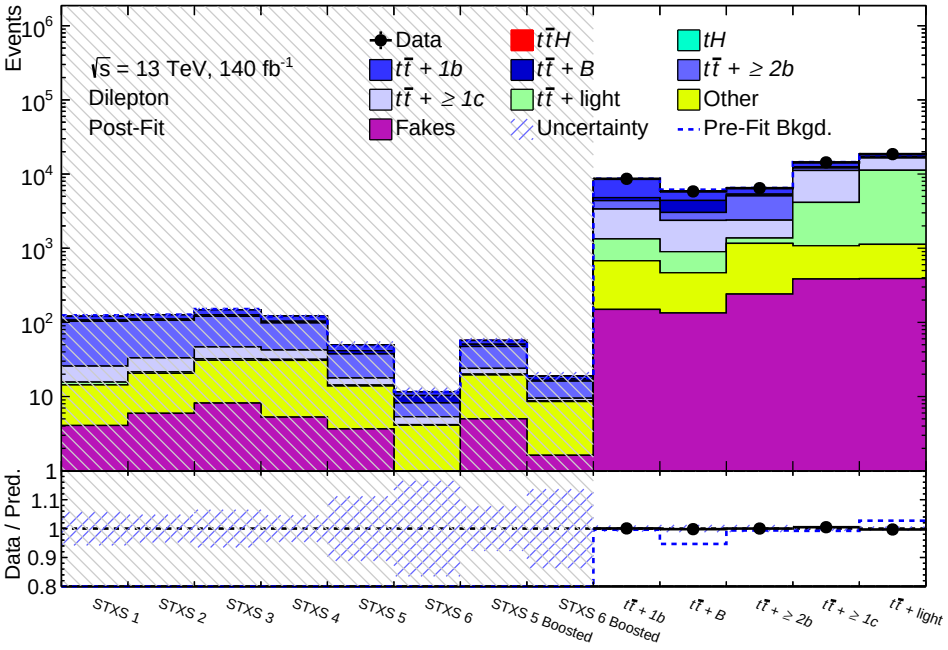
For the experimental NPs, the b-tagging related NPs are depicted in Figure 14.10.

For most NPs, no significant pull or constraint is visible except for the light-flavour jets eigenvalue 0 and c-jet eigenvalue 7. While for the light-flavour jets eigenvalue 0 both individual channels show a slight pull with no significant constraints, the combined fit shows a large pull with a notable constraint. For the c-jet eigenvalue 7, similar holds true although no serious constraint is observed for the combined fit. Both NPs are currently under investigation.

All other experimental NPs show no significant pulls or constraints and are depicted in Figures A.30–A.33.



(a): Single-lepton



(b): Dilepton

Figure 14.8.: Background-only post-fit summary plot for both the single-lepton and dilepton channels showing the overall MC and data yields in the different SRs and CRs. All regions with a data to MC ratio of more than 7.7 % are blinded, indicated by the grey shaded area.

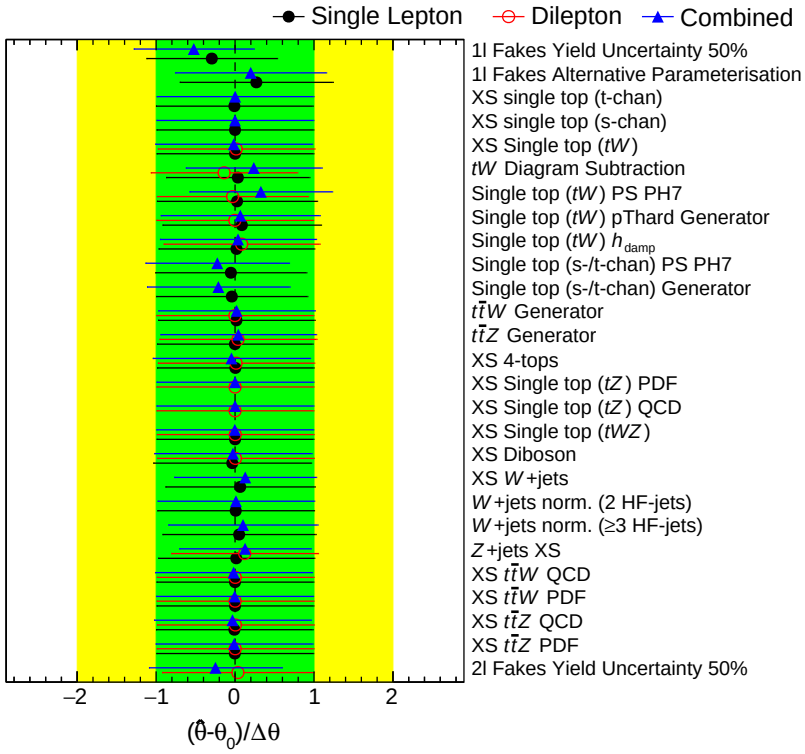


Figure 14.9.: Other processes modelling NPs pulls for the different background-only fits.

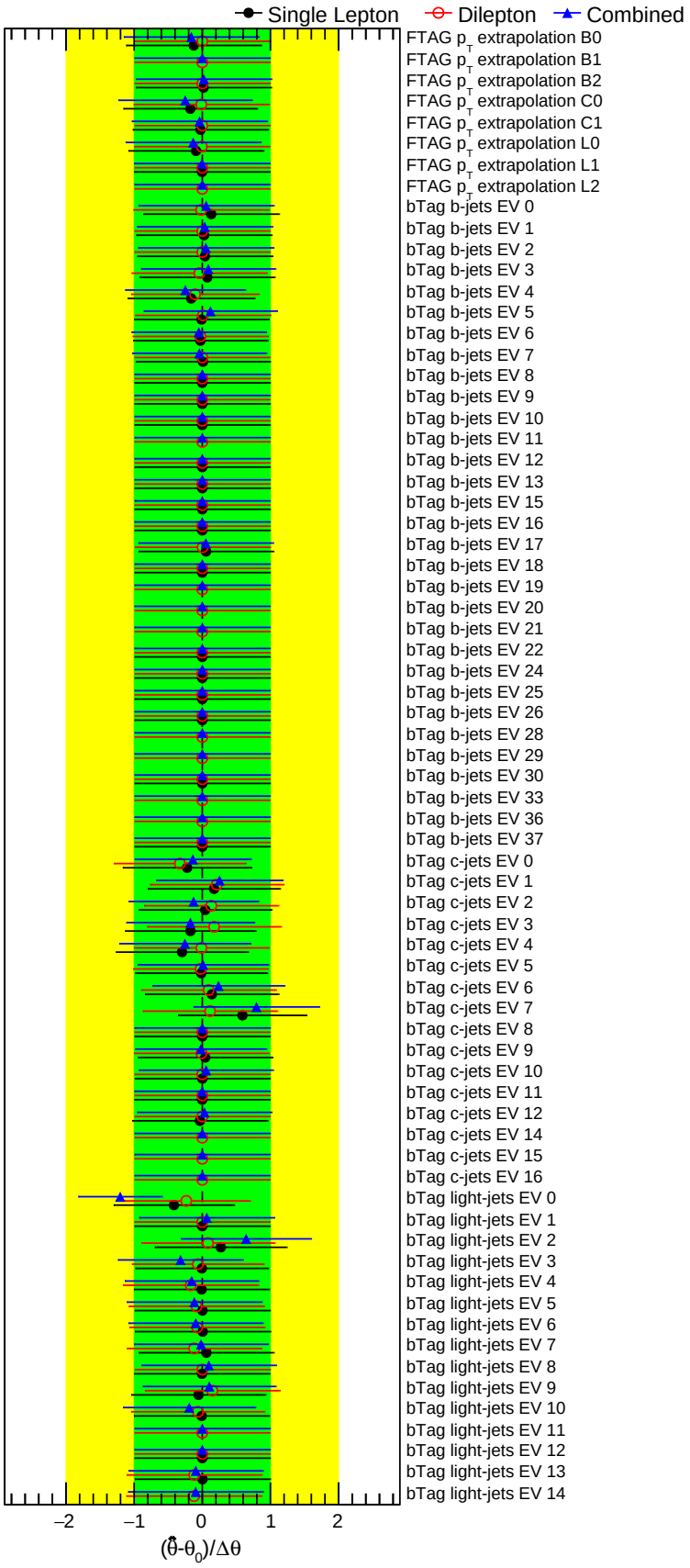


Figure 14.10.: *b*-tagging NPs pulls for the different background-only fits.

Part V

CONCLUSION

15

SUMMARY

In this thesis, two main topics were presented: development of new heavy-flavour tagging algorithms in ATLAS and the current status of the ATLAS $t\bar{t}H(H \to b\bar{b})$ legacy analysis using the full Run 2 dataset of $140\,\text{fb}^{-1}$ at $\sqrt{s} = 13\,\text{TeV}$.

The first topic was the identification of b-quark induced jets, b-tagging, relying on new and state-of-the-art ML techniques. The previously recommended high-level ATLAS b-tagging algorithm DL1r, using the low-level track-based algorithm RNNIP, was introduced and discussed. While the overall performance gain of DL1r over its predecessor DL1, thanks to the addition RNNIP, was astonishing, RNNIP was found to be not the optimal solution for a track-based b-tagging algorithm due to its sequential processing of the given tracks. Therefore, RNNIP was replaced with the new DIPS algorithm, which is based on the Deep Sets architecture and allows for parallel processing of the input tracks. In addition to the overall impressive gains in computing time, the usage of the correlations among different tracks and their features enabled DIPS to surpass RNNIP in terms of performance, yielding unprecedented rejection rates for c and light-flavour jets. With the combination of DIPS and the DL1 high-level architecture, the resulting new high-level tagger DL1d was able to achieve even larger rejection rates, outperforming the previously recommended DL1r algorithm clearly.
After the successful demonstration of the enhanced performance of DIPS and DL1d, the extension of both algorithms was studied. Thanks to the modular structure and training of all b-tagging algorithms, a simple redefinition of the output discriminant transforms the b-tagging algorithms into c-tagging algorithms, which can be used in dedicated analyses focusing on c-jets in the final state. Unfortunately, the newly transformed c-taggers suffer from large mis-identification efficiencies of hadronically decaying τ-jets, which motivates the extension of the DIPS and DL1d to incorporate τ-jets in the training of the very same. First studies were performed with DIPS, showing an astonishing improvement in τ-jets rejection capabilities while maintaining acceptable c and light-flavour rejection rates. Furthermore, similar studies were performed with DL1d, also showing impressive enhancements in the τ-jets rejection rate while maintaining acceptable c and light-flavour rejection rates. As a result of these studies, the upcoming state-of-the-art transformer-based tagger, GN2, will be trained with the additional τ-jet class.

The second topic of this thesis was the search and measurement of the $t\bar{t}H(H \to b\bar{b})$ process using the full Run 2 data collected by the ATLAS experiment at $\sqrt{s} = 13\,\text{TeV}$. Compared with an earlier iteration of this analysis using this dataset, the $t\bar{t}H(H \to b\bar{b})$ legacy analysis presented herein incorporates a number of improvements in terms of ML-based classification and reconstruction of events, as well as an improved modelling of the dominant $t\bar{t} + \geq 1b$ background processes. Further-

more, updated object definitions and identification algorithms were used, namely the particle-flow algorithm for jet reconstruction and the DL1r algorithm for b-jet identification. Thanks to these improvements and updates, the pre-selection of the analysis could be loosened, allowing more events to be considered in the signal region.

While the previous analysis, owing to the stricter pre-selection, found the fake and non-prompt lepton contributions to be negligible, with the change to a looser pre-selection, the fake estimation for the single-lepton channel was revisited and a new estimation was derived. Two methods for estimation, the matrix method and the fake factor method, were tested. While the matrix method is the standard technique used within ATLAS, it suffers under certain object definition issues, leading to unstable and negative estimates in different analysis regions. Therefore, it was decided to employ the fake factor method, which is in general more stable against the object definition issue seen within the matrix method.

Several tests for pre-fit background modelling were done, yielding a mis-modelling of the H_{T} distributions for the $t\bar{t}+\mathrm{jets}$ processes. While a data-driven correction of the $t\bar{t}+\geq 1b$ processes was not achievable due to the similar final state with the signal process, correction factors were derived for the $t\bar{t}+\geq 1c$ and $t\bar{t}+$ light processes using dedicated orthogonal CRs. While the overall effect of the reweighting on the H_{T} distributions is relatively small, positive effects in the shape of H_{T}-sensitive variables, like p_{T}^{H}, were observed, improving the mis-modelling and partly removing the resulting shape effect.

Due to the currently blinded status of the analysis, only background-only and Asimov profile-likelihood fits were performed to estimate and examine the expected uncertainties on different parameters of interest and the effect of various nuisance parameters. While a reasonable pre-fit and post-fit modelling was observed for background-only fits, resulting in acceptable post-fit normalisation factors for the main backgrounds, Asimov fits gave an estimate of the expected uncertainties and significances of the measurement. For both the STXS and inclusive measurement, clear improvements in the expected uncertainties on the different parameters of interest were observed. Additionally, an expected significance of the inclusive measurement was derived, yielding $5.5\,\sigma$ and therefore surpassing the required $5\,\sigma$ threshold for a discovery. Although a high expected significance is observed, comparing this results to the results of the previous analysis, which yielded an observed (expected) significance of $1.0\,\sigma$ ($2.7\,\sigma$), no clear conclusion can be drawn from this besides the fact that a lower observed significance can be expected. Therefore, the here presented expected significance is no clear evidence for a successful measurement of the $t\bar{t}H(H\rightarrow b\bar{b})$ process.

With these fit results, the analysis is currently preparing for an unblinding request, which enables unblinded fits to data, marking the final phase of the analysis and hinting at the upcoming results.

Part VI

APPENDICES

A

ADDITIONAL MATERIAL

A.1 ADDITIONAL MATERIAL FOR THE INTRODUCTION TO HEAVY-FLAVOUR TAGGING

Table A.1.: Input Variables to the DL1 and DL1r high-level b-tagging algorithms. Adapted from [172].

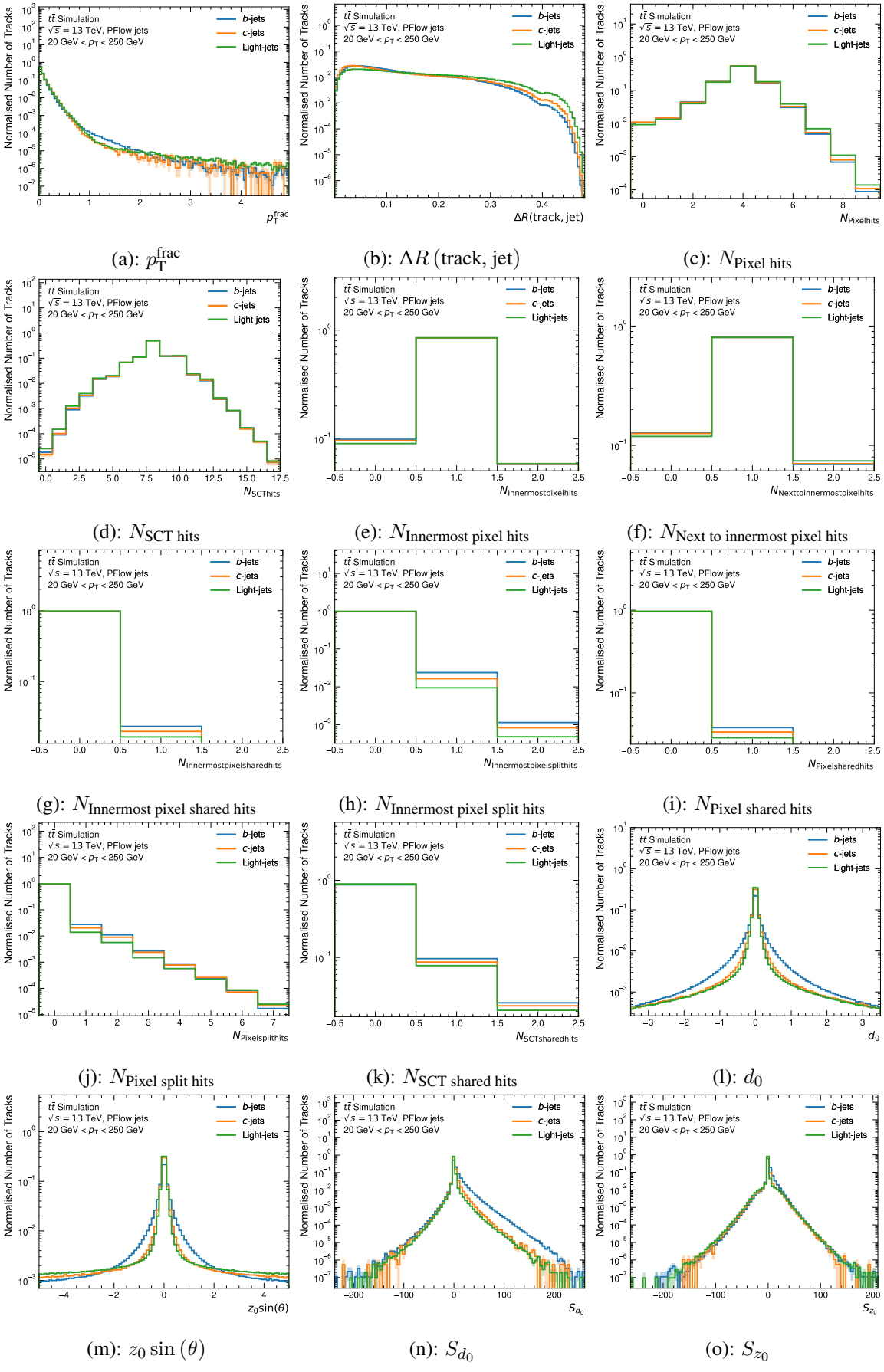
Input	Variable	Description	DL1	DL1r
Kinematics	p_{T}	Jet p_{T}	✓	✓
	η	Jet $\lvert\eta\rvert$	✓	✓
IP2D, IP3D	$\log(p_b/p_{\mathrm{light}})$	IP2D and IP3D LH ratio of the b-jet to light-flavour jet hypotheses	✓	✓
	$\log(p_b/p_c)$	IP2D and IP3D LH ratio of the b-jet to c-jet hypotheses	✓	✓
	$\log(p_c/p_{\mathrm{light}})$	IP2D and IP3D LH ratio of the c-jet to light-flavour jet hypotheses	✓	✓
RNNIP	p_b	b-jet probability		✓
	p_c	c-jet probability		✓
	p_{light}	Light-flavour jet probability		✓
SV1	m(SV)	Invariant mass of tracks at the secondary vertex assuming pion mass	✓	✓
	f_E(SV)	Jet energy fraction of the tracks associated with the secondary vertex	✓	✓
	N_{TrkAtVtx}(SV)	Number of tracks used in the secondary vertex	✓	✓
	N_{2TrkVtx}(SV)	Number of two-track vertex candidates	✓	✓
	L_{xy}(SV)	Transverse distance between the primary and secondary vertices	✓	✓
	L_{xyz}(SV)	Distance between the primary and secondary vertices	✓	✓
	S_{xyz}(SV)	Distance between the primary and secondary vertices divided by its uncertainty	✓	✓
	$\Delta R(\vec{p}_{\mathrm{jet}}, \vec{p}_{\mathrm{vtx}})$(SV)	ΔR between the jet axis and the direction of the secondary vertex relative to the primary vertex.	✓	✓
JETFITTER	m(JF)	Invariant mass of tracks from displaced vertices	✓	✓
	f_E(JF)	Jet energy fraction of the tracks associated with the displaced vertices	✓	✓
	$\Delta R(\vec{p}_{\mathrm{jet}}, \vec{p}_{\mathrm{vtx}})$(JF)	ΔR between the jet axis and the vectorial sum of momenta of all tracks attached to displaced vertices	✓	✓
	S_{xyz}(JF)	Significance of the average distance between PV and displaced vertices	✓	✓
	N_{TrkAtVtx}(JF)	Number of tracks from multi-prong displaced vertices	✓	✓
	N_{2TrkVtx}(JF)	Number of two-track vertex candidates (prior to decay chain fit)	✓	✓
	$N_{\mathrm{1\text{-}trk\ vertices}}$(JF)	Number of single-prong displaced vertices	✓	✓
	$N_{\geq\mathrm{2\text{-}trk\ vertices}}$(JF)	Number of multi-prong displaced vertices	✓	✓
JETFITTER (2^{ND} VERTEX)	$L_{xyz}(2^{\mathrm{nd}})$(JF)	Distance of 2^{nd} vertex from PV	✓	✓
	$L_{xy}(2^{\mathrm{nd}})$(JF)	Transverse displacement of the 2^{nd} vertex	✓	✓
	$m_{\mathrm{Trk}}(2^{\mathrm{nd}})$(JF)	Invariant mass of tracks associated with the 2^{nd} vertex	✓	✓
	$E(2^{\mathrm{nd}})$(JF)	Energy of the tracks associated with the 2^{nd} vertex	✓	✓
	$f_E(2^{\mathrm{nd}})$(JF)	Jet energy fraction of the tracks associated with the 2^{nd} vertex	✓	✓
	$N_{\mathrm{TrkAtVtx}}(2^{\mathrm{nd}})$(JF)	Number of tracks associated with the 2^{nd} vertex	✓	✓
	$\eta_{\mathrm{trk}}^{\mathrm{min,\ max,\ avg}}(2^{\mathrm{nd}})$(JF)	Min., max. and avg. pseudorapidity of tracks at the 2^{nd} vertex	✓	✓

1 Non-trainable parameters are an effect from using Batch Normalisation in the network.

Table A.2.: Hyperparameters of the DL1r tagger trained for Run 2. Adapted from Ref. [172].

Hyperparameter	Value
Number of input variables	36
Number of hidden layers	8
Number of neurons per layer	[256, 128, 60, 48, 36, 24, 12, 6]
LR	0.01
Batch size	15000
Activation function (hidden neurons)	ReLU
Number of epochs trained	200
Free (trainable) parameters	57227
Fixed parameters[1]	1140
Total parameters	58367

A.2 ADDITIONAL MATERIAL FOR THE DIPS TAGGER



(a): $p_{\mathrm{T}}^{\mathrm{frac}}$ (b): $\Delta R\,(\mathrm{track, jet})$ (c): $N_{\mathrm{Pixel\ hits}}$

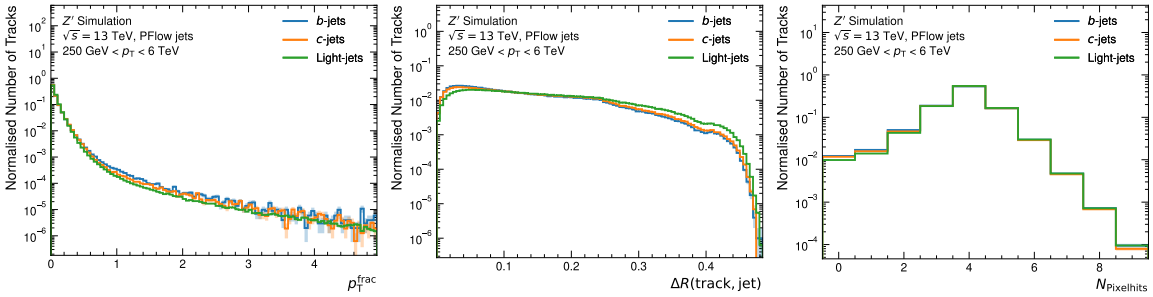
(d): $N_{\mathrm{SCT\ hits}}$ (e): $N_{\mathrm{Innermost\ pixel\ hits}}$ (f): $N_{\mathrm{Next\ to\ innermost\ pixel\ hits}}$

(g): $N_{\mathrm{Innermost\ pixel\ shared\ hits}}$ (h): $N_{\mathrm{Innermost\ pixel\ split\ hits}}$ (i): $N_{\mathrm{Pixel\ shared\ hits}}$

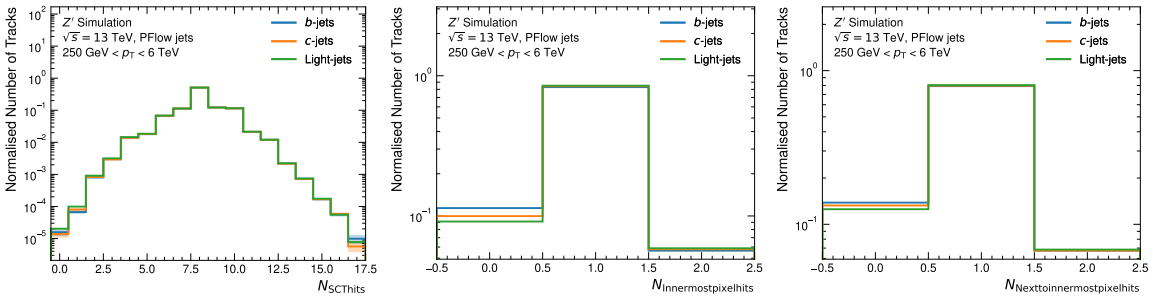
(j): $N_{\mathrm{Pixel\ split\ hits}}$ (k): $N_{\mathrm{SCT\ shared\ hits}}$ (l): d_0

(m): $z_0 \sin(\theta)$ (n): S_{d_0} (o): S_{z_0}

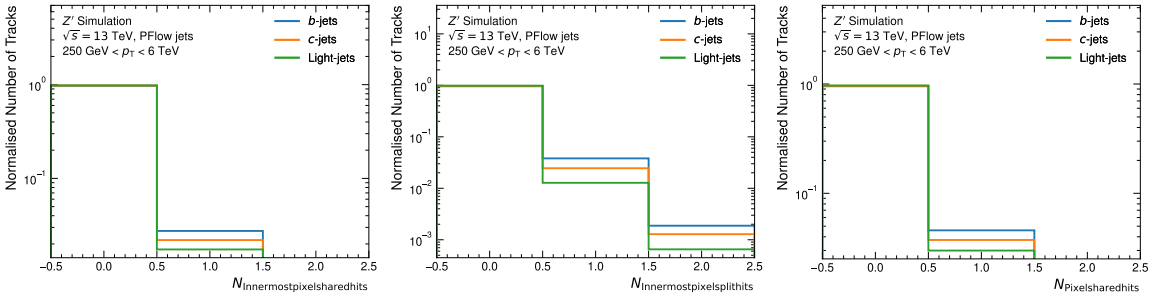
Figure A.1.: Distributions of the track variables for $t\bar{t}$ jets used in the training of DIPS.



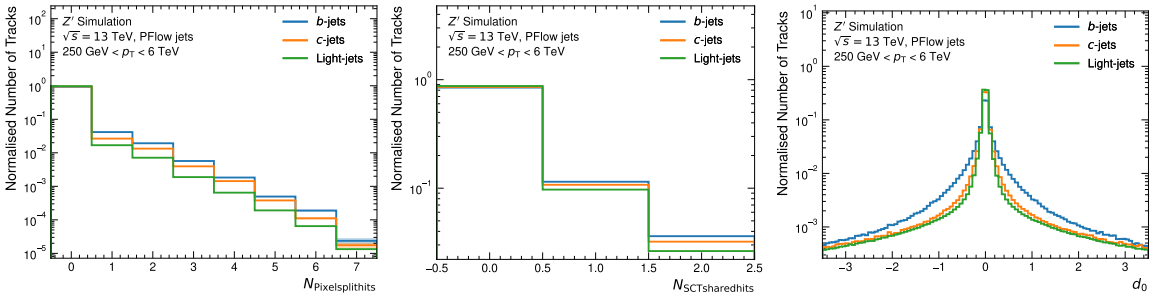
(a): $p_{\mathrm{T}}^{\mathrm{frac}}$ (b): $\Delta R\,(\mathrm{track,\,jet})$ (c): $N_{\mathrm{Pixel\ hits}}$



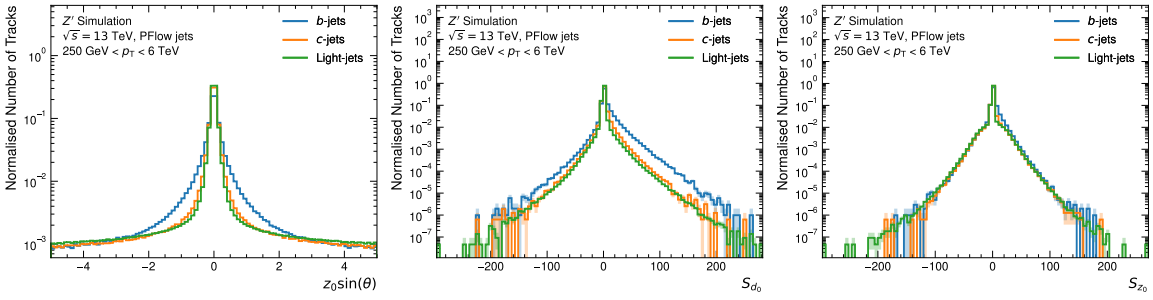
(d): $N_{\mathrm{SCT\ hits}}$ (e): $N_{\mathrm{Innermost\ pixel\ hits}}$ (f): $N_{\mathrm{Next\ to\ innermost\ pixel\ hits}}$



(g): $N_{\mathrm{Innermost\ pixel\ shared\ hits}}$ (h): $N_{\mathrm{Innermost\ pixel\ split\ hits}}$ (i): $N_{\mathrm{Pixel\ shared\ hits}}$



(j): $N_{\mathrm{Pixel\ split\ hits}}$ (k): $N_{\mathrm{SCT\ shared\ hits}}$ (l): d_0



(m): $z_0 \sin(\theta)$ (n): S_{d_0} (o): S_{z_0}

Figure A.2.: Distributions of the track variables for Z' jets used in the training of DIPS.

A.3 ADDITIONAL MATERIAL FOR THE DL1D TAGGER

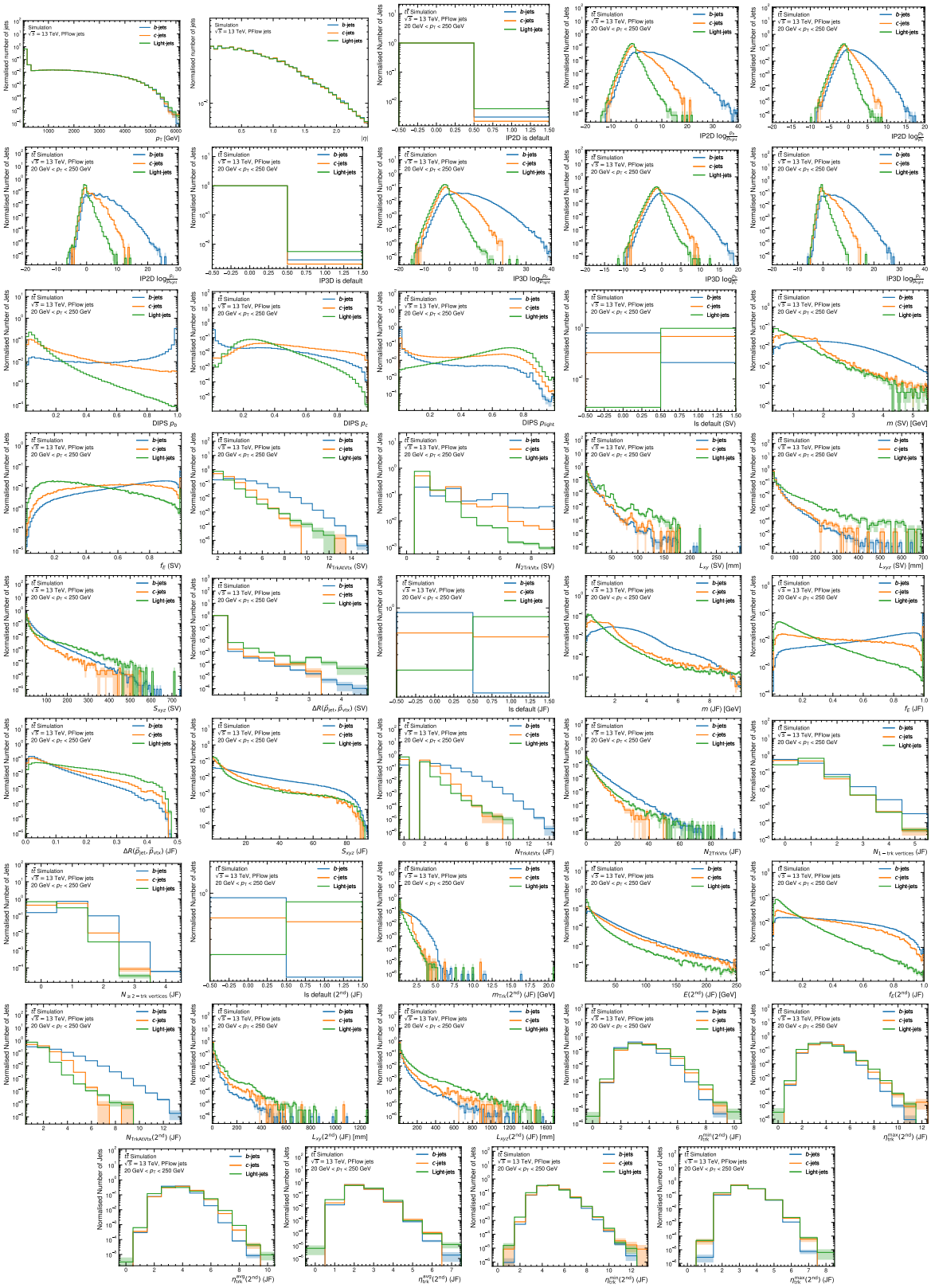
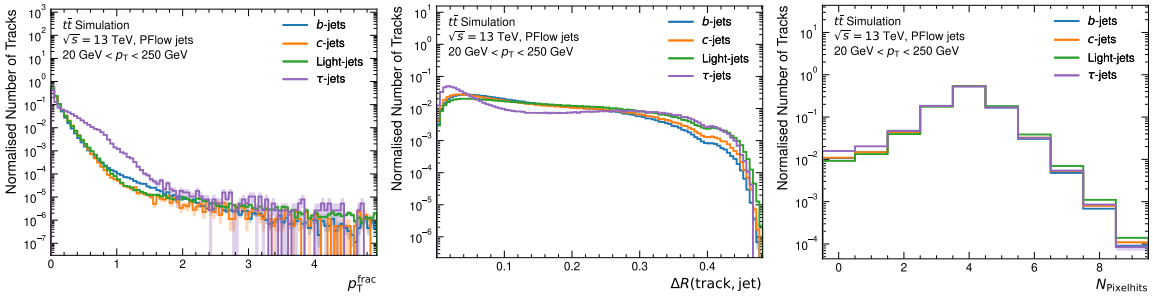
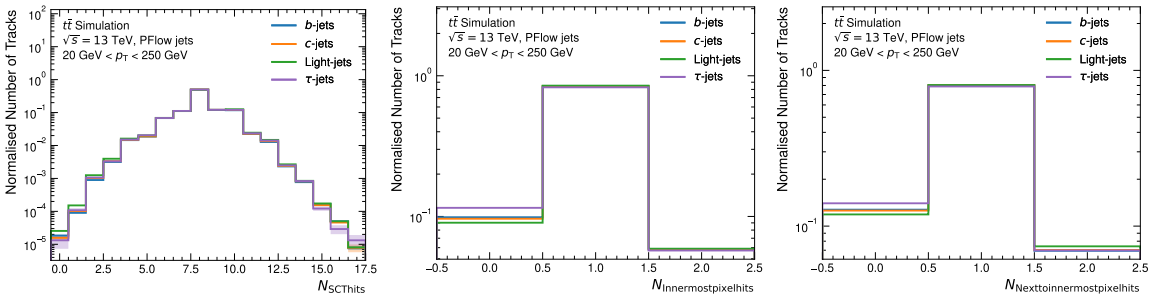


Figure A.3.: Input variable distributions for the DL1d tagger.

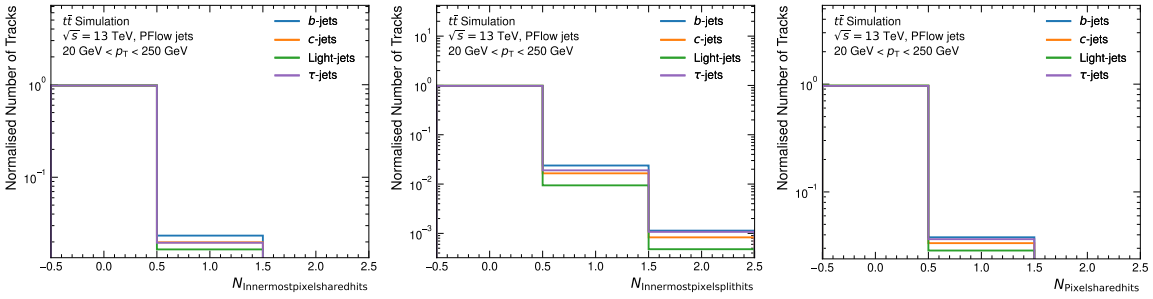
A.4 ADDITIONAL MATERIAL FOR THE DIPS TAU TAGGER



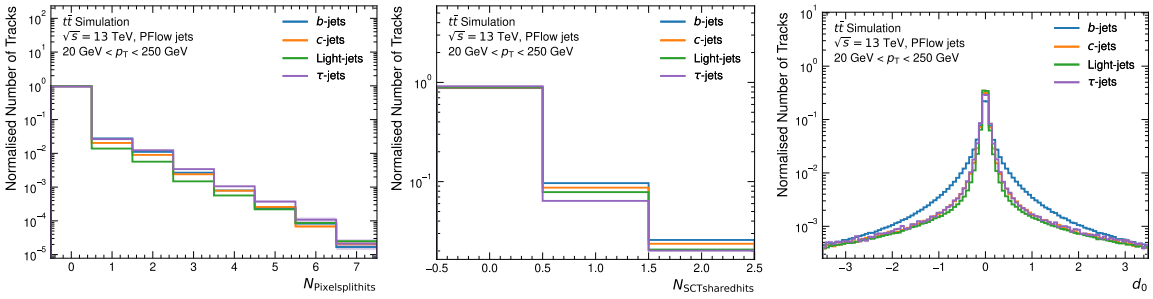
(a): $p_{\mathrm{T}}^{\mathrm{frac}}$ (b): $\Delta R\,(\mathrm{track,\,jet})$ (c): $N_{\mathrm{Pixel\ hits}}$



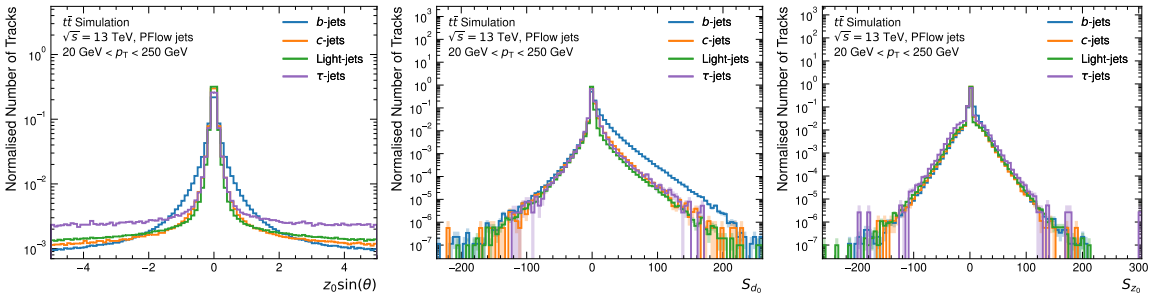
(d): $N_{\mathrm{SCT\ hits}}$ (e): $N_{\mathrm{Innermost\ pixel\ hits}}$ (f): $N_{\mathrm{Next\ to\ innermost\ pixel\ hits}}$



(g): $N_{\mathrm{Innermost\ pixel\ shared\ hits}}$ (h): $N_{\mathrm{Innermost\ pixel\ split\ hits}}$ (i): $N_{\mathrm{Pixel\ shared\ hits}}$



(j): $N_{\mathrm{Pixel\ split\ hits}}$ (k): $N_{\mathrm{SCT\ shared\ hits}}$ (l): d_0



(m): $z_0 \sin(\theta)$ (n): S_{d_0} (o): S_{z_0}

Figure A.4.: Distributions of the track variables for $t\bar{t}$ jets used in the training of DIPS Tau.

A.5 ADDITIONAL MATERIAL FOR THE $t\bar{t}H(H \to b\bar{b})$ LEGACY ANALYSIS

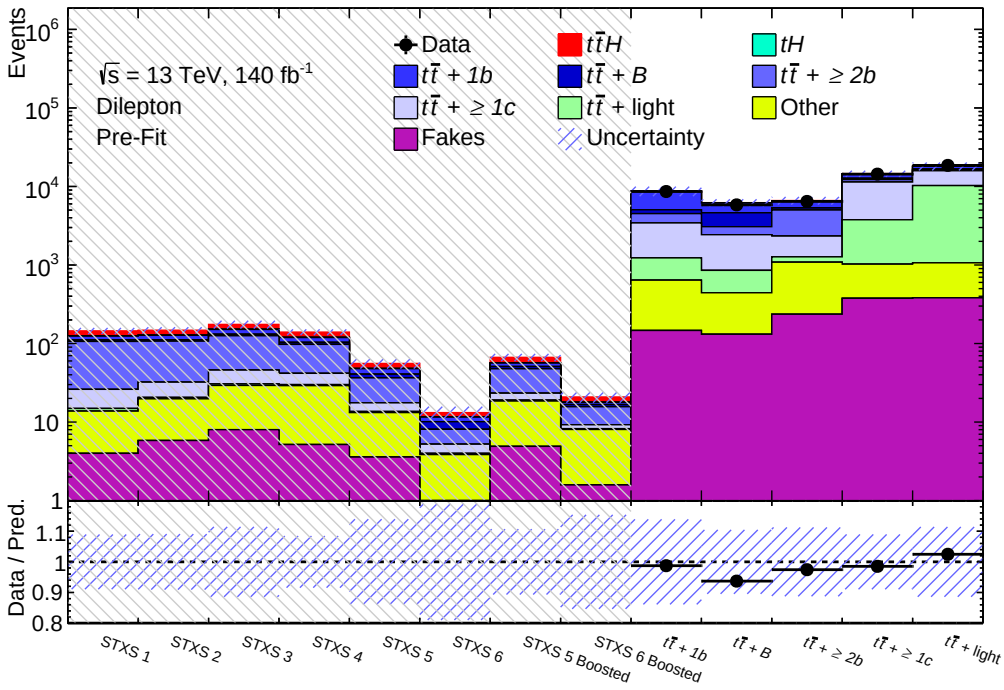


Figure A.5.: Pre-fit summary plot for the dilepton channel showing the overall MC and data yields in the different SRs and CRs. All regions with a data to MC ratio of more than 7.7 % are blinded, indicated by the grey shaded area.

Table A.3.: Pre-fit yields of the single-lepton channel SRs. The errors include all systematic uncertainties. No data yields are shown due to the blinded status of the analysis.

	STXS 1	STXS 2	STXS 3	STXS 4	STXS 5	STXS 6	Boosted SR
$t\bar{t}H$	153.215 ± 24.982	160.846 ± 26.559	197.264 ± 34.031	151.300 ± 25.966	67.038 ± 11.958	24.227 ± 4.800	34.650 ± 6.589
tH	1.656 ± 0.306	2.050 ± 0.277	3.521 ± 0.480	3.395 ± 0.372	2.121 ± 0.313	0.637 ± 0.138	1.051 ± 0.157
$t\bar{t}+b$	194.886 ± 53.977	190.188 ± 53.161	232.827 ± 73.129	168.550 ± 57.051	72.206 ± 26.760	23.171 ± 8.508	15.537 ± 4.541
$t\bar{t}+B$	59.286 ± 17.909	67.437 ± 22.111	94.241 ± 35.729	80.187 ± 27.693	42.596 ± 18.001	20.040 ± 10.049	10.175 ± 3.449
$t\bar{t}+\geq 2b$	908.455 ± 80.873	905.224 ± 93.412	1017.480 ± 101.143	668.794 ± 65.311	285.257 ± 41.651	95.779 ± 22.319	107.481 ± 8.275
$t\bar{t}+\geq 1c$	117.658 ± 25.859	116.612 ± 24.558	137.134 ± 27.496	104.298 ± 25.634	49.044 ± 11.857	15.156 ± 7.400	12.156 ± 3.900
$t\bar{t}+$ light	56.238 ± 21.613	60.860 ± 16.785	91.787 ± 29.486	86.762 ± 25.770	41.860 ± 14.297	15.744 ± 14.422	4.351 ± 4.130
Other t	11.284 ± 10.276	11.010 ± 11.236	14.443 ± 9.151	11.857 ± 11.395	7.253 ± 7.218	2.425 ± 0.710	2.216 ± 0.820
Single-t tW	17.637 ± 9.151	27.492 ± 17.145	52.046 ± 28.293	44.918 ± 28.786	31.291 ± 22.547	15.084 ± 13.996	5.271 ± 3.922
$t\bar{t}Z$	38.834 ± 5.854	41.323 ± 6.523	56.283 ± 8.679	48.715 ± 8.270	29.090 ± 6.575	7.226 ± 1.159	6.942 ± 1.609
$t\bar{t}W$	3.222 ± 0.620	4.516 ± 0.755	6.974 ± 1.221	5.939 ± 1.040	4.378 ± 0.706	2.779 ± 0.489	0.688 ± 0.141
V+jets	21.908 ± 9.467	26.626 ± 11.172	47.517 ± 19.547	46.376 ± 19.994	35.964 ± 15.416	17.206 ± 7.634	5.909 ± 2.641
Diboson	1.273 ± 0.776	2.527 ± 1.393	4.374 ± 2.293	4.659 ± 2.701	3.085 ± 1.615	1.352 ± 0.715	0.715 ± 0.369
$t\bar{t}t\bar{t}$	8.582 ± 2.563	11.638 ± 3.461	18.383 ± 5.464	14.613 ± 4.352	8.061 ± 2.394	4.228 ± 1.260	1.895 ± 0.569
Fakes	29.164 ± 15.066	29.308 ± 15.886	27.078 ± 14.454	30.753 ± 16.780	22.197 ± 14.604	0 ± 0	7.895 ± 5.149
Total	1623.3 ± 152.224	1657.66 ± 163.372	2001.35 ± 202.764	1471.12 ± 148.386	701.44 ± 80.890	245.055 ± 39.796	216.931 ± 18.434
Data	0	0	0	0	0	0	0

Table A.4.: Pre-fit yields of the single-lepton channel CRs. The errors include all systematic uncertainties. No data yields are shown due to the blinded status of the analysis.

	Boosted CR	$t\bar{t}+b$	$t\bar{t}+B$	$t\bar{t}+\geq 2b$	$t\bar{t}+\geq 1c$	$t\bar{t}+$ light
$t\bar{t}H$	11.044 ± 2.277	326.137 ± 47.863	286.427 ± 47.606	426.529 ± 64.683	278.122 ± 39.001	234.163 ± 30.489
tH	0.902 ± 0.223	16.492 ± 2.091	7.070 ± 0.791	14.972 ± 2.138	9.150 ± 1.144	14.498 ± 1.651
$t\bar{t}+b$	49.773 ± 17.136	9843.130 ± 1391.940	4695.960 ± 543.288	3233.820 ± 919.655	4268.160 ± 711.541	3651.570 ± 381.631
$t\bar{t}+B$	29.534 ± 6.124	1645.380 ± 251.943	4716.630 ± 347.232	944.731 ± 348.692	1287.050 ± 246.850	820.410 ± 76.995
$t\bar{t}+\geq 2b$	146.850 ± 11.036	5587.960 ± 589.140	3653.340 ± 396.796	13854.700 ± 1194.110	3991.210 ± 407.142	2414.550 ± 285.907
$t\bar{t}+\geq 1c$	74.074 ± 17.545	4371.880 ± 594.317	3000.860 ± 640.728	2219.870 ± 413.095	10010.200 ± 1176.310	9497.620 ± 1968.540
$t\bar{t}+$ light	46.595 ± 11.002	5344.400 ± 809.098	2677.850 ± 634.740	1443.570 ± 426.690	8005.030 ± 1567.450	42998.000 ± 4698.230
Other-t	3.865 ± 1.119	303.342 ± 190.826	227.628 ± 143.477	317.495 ± 245.125	368.032 ± 238.218	365.344 ± 222.509
Single-t tW	9.194 ± 6.211	750.930 ± 247.570	514.736 ± 184.610	769.205 ± 397.843	671.038 ± 224.234	943.468 ± 178.188
$t\bar{t}Z$	6.588 ± 1.411	155.801 ± 21.241	153.412 ± 22.441	294.477 ± 39.731	186.408 ± 25.115	154.600 ± 19.473
$t\bar{t}W$	1.688 ± 0.296	30.660 ± 5.984	37.505 ± 6.379	46.619 ± 7.735	82.624 ± 13.453	74.212 ± 12.493
V+jets	8.279 ± 3.636	436.404 ± 171.456	309.801 ± 121.805	1119.980 ± 470.063	576.840 ± 232.565	589.842 ± 242.200
Diboson	1.099 ± 0.729	23.297 ± 12.040	21.103 ± 10.839	57.312 ± 29.276	38.795 ± 20.148	37.060 ± 19.285
$t\bar{t}t\bar{t}$	2.418 ± 0.724	11.950 ± 3.545	32.492 ± 9.570	119.843 ± 35.638	40.958 ± 12.081	4.874 ± 1.450
Fakes	6.321 ± 3.598	875.761 ± 451.472	342.607 ± 178.762	993.947 ± 524.182	835.528 ± 445.432	809.799 ± 405.036
Total	398.224 ± 42.244	29723.500 ± 2716.880	20677.400 ± 1771.180	25857.000 ± 2693.770	30649.100 ± 3341.660	62610.000 ± 6237.900
Data	0	0	0	0	0	0

Table A.5.: Pre-fit yields of the dilepton channel SRs. The errors include all systematic uncertainties. No data yields are shown due to the blinded status of the analysis.

	STXS 1	STXS 2	STXS 3	STXS 4	STXS 5	STXS 6	Boosted STXS 5	Boosted STXS 6
$t\bar{t}H$	20.657 ± 3.377	20.460 ± 3.340	23.863 ± 3.991	20.201 ± 3.583	7.609 ± 1.399	1.608 ± 0.391	10.074 ± 1.759	3.024 ± 0.659
tH	0.074 ± 0.036	0.249 ± 0.089	0.306 ± 0.075	0.364 ± 0.055	0.307 ± 0.065	0.030 ± 0.011	0.262 ± 0.042	0.073 ± 0.025
$t\bar{t}+b$	10.549 ± 3.397	11.812 ± 4.185	15.415 ± 5.926	12.239 ± 3.023	5.537 ± 3.268	1.124 ± 0.591	4.202 ± 2.050	0.849 ± 0.469
$t\bar{t}+B$	3.371 ± 2.038	3.627 ± 2.439	4.342 ± 1.825	4.603 ± 1.740	3.230 ± 1.449	1.568 ± 0.969	2.543 ± 1.043	0.782 ± 0.611
$t\bar{t}+\geq 2b$	86.238 ± 8.595	80.620 ± 7.477	85.706 ± 13.461	59.979 ± 4.036	20.393 ± 2.893	3.073 ± 0.884	26.191 ± 2.757	6.913 ± 1.834
$t\bar{t}+\geq 1c$	7.044 ± 1.894	7.324 ± 1.547	9.673 ± 2.688	7.386 ± 1.825	2.476 ± 1.233	0.778 ± 0.647	2.714 ± 1.513	0.627 ± 0.337
$t\bar{t}+$ light	1.308 ± 0.745	0.964 ± 0.560	1.489 ± 1.071	1.064 ± 0.554	0.503 ± 0.371	0.162 ± 0.154	0.562 ± 0.415	0.232 ± 0.230
Other t	0.059 ± 0.024	0.056 ± 0.043	0.279 ± 0.140	0.378 ± 0.171	0.104 ± 0.044	0.116 ± 0.073	0.162 ± 0.069	0.193 ± 0.081
Single-t tW	1.263 ± 0.809	3.245 ± 1.877	5.528 ± 4.723	6.554 ± 4.377	2.909 ± 2.310	0.892 ± 0.737	3.047 ± 2.655	1.604 ± 1.657
$t\bar{t}Z$	4.276 ± 0.723	4.963 ± 0.768	6.476 ± 1.068	7.225 ± 1.424	2.961 ± 0.739	0.403 ± 0.345	4.336 ± 0.795	1.456 ± 0.398
$t\bar{t}W$	0.561 ± 0.198	0.964 ± 0.201	1.588 ± 0.251	1.745 ± 0.295	0.869 ± 0.150	0.438 ± 0.123	1.416 ± 0.322	1.050 ± 0.408
V+jets	2.232 ± 0.822	2.933 ± 1.157	4.672 ± 1.804	5.570 ± 2.065	1.584 ± 0.846	0.855 ± 0.342	3.238 ± 1.200	1.328 ± 0.577
$t\bar{t}t\bar{t}$	1.391 ± 0.421	1.773 ± 0.528	2.700 ± 0.801	2.431 ± 0.722	1.183 ± 0.354	0.500 ± 0.153	1.418 ± 0.422	0.829 ± 0.250
Fakes	3.994 ± 1.998	5.822 ± 2.913	8.001 ± 4.004	5.175 ± 2.589	3.585 ± 1.795	0.648 ± 0.326	4.903 ± 2.456	1.582 ± 0.794
Total	143.019 ± 12.451	144.812 ± 12.595	170.038 ± 19.267	134.913 ± 10.985	53.249 ± 6.851	12.196 ± 2.149	65.069 ± 6.557	20.543 ± 3.162
Data	0	0	0	0	0	0	0	0

Table A.6.: Pre-fit yields of the dilepton channel CRs. The errors include all systematic uncertainties. No data yields are shown due to the blinded status of the analysis.

	$t\bar{t}+b$	$t\bar{t}+B$	$t\bar{t}+\geq 2b$	$t\bar{t}+\geq 1c$	$t\bar{t}+$ light
$t\bar{t}H$	72.323 ± 11.078	63.001 ± 10.187	123.840 ± 18.101	96.856 ± 13.805	43.595 ± 5.979
tH	2.875 ± 0.306	1.227 ± 0.139	2.507 ± 0.296	2.070 ± 0.205	1.170 ± 0.121
$t\bar{t}+b$	2800.95 ± 728.378	1187.71 ± 188.788	903.01 ± 285.33	1380.65 ± 107.796	1090.11 ± 105.447
$t\bar{t}+B$	415.701 ± 160.433	1202.41 ± 183.646	241.215 ± 84.320	352.797 ± 32.6289	240.953 ± 28.1899
$t\bar{t}+\geq 2b$	1158.34 ± 218.672	690.961 ± 103.553	2881.07 ± 212.71	918.106 ± 63.1281	528.825 ± 36.9595
$t\bar{t}+\geq 1c$	1383.8 ± 140.862	991.78 ± 252.053	670.371 ± 128.568	4790.8 ± 599.402	3518.04 ± 389.99
$t\bar{t}+$ light	694.754 ± 159.348	483.737 ± 87.8478	216.952 ± 80.2643	3227.23 ± 494.41	10818.7 ± 1711.01
Other-t	4.411 ± 1.945	2.135 ± 0.742	12.415 ± 5.159	8.062 ± 3.060	5.605 ± 2.35
Single-t tW	198.086 ± 64.475	132.473 ± 48.32	270.899 ± 143.152	192.285 ± 67.698	266.855 ± 58.702
$t\bar{t}Z$	34.043 ± 5.189	37.717 ± 6.065	97.942 ± 14.08	86.476 ± 11.284	43.899 ± 8.615
$t\bar{t}W$	9.698 ± 1.437	13.653 ± 2.855	22.692 ± 3.698	55.927 ± 8.32	31.014 ± 4.924
V+jets	246.518 ± 89.556	117.827 ± 42.338	398.399 ± 148.292	289.294 ± 106.322	333.468 ± 129.4
$t\bar{t}t\bar{t}$	1.637 ± 0.496	7.964 ± 2.347	46.817 ± 13.878	15.947 ± 4.701	2.628 ± 0.778
Fakes	146.654 ± 73.337	131.238 ± 65.628	236.044 ± 118.038	376.886 ± 188.469	382.093 ± 191.073
Total	7169.79 ± 978.008	5063.83 ± 491.433	6124.17 ± 646.069	11793.4 ± 1035.87	17307 ± 2132.02
Data	0	0	0	0	0

Table A.7.: Background-only post-fit yields of the single-lepton channel signal regions. The errors include all systematic uncertainties.

	STXS 1	STXS 2	STXS 3	STXS 4	STXS 5	STXS 6	Boosted SR
$t\bar{t}H$	0	0	0	0	0	0	0
tH	1.627 ± 0.274	2.053 ± 0.241	3.650 ± 0.459	3.378 ± 0.347	2.159 ± 0.292	0.647 ± 0.129	1.051 ± 0.149
$t\bar{t} + b$	233.293 ± 30.911	227.600 ± 31.753	281.873 ± 40.861	200.100 ± 32.619	75.584 ± 15.991	26.067 ± 4.740	17.038 ± 3.543
$t\bar{t} + B$	73.310 ± 21.923	78.879 ± 25.236	116.655 ± 41.192	95.874 ± 30.010	44.538 ± 16.295	20.374 ± 8.583	11.500 ± 3.829
$t\bar{t} + \geq 2b$	924.487 ± 51.250	899.864 ± 53.132	1012.890 ± 57.047	683.708 ± 46.616	297.678 ± 26.105	106.870 ± 13.855	103.623 ± 6.304
$t\bar{t} + \geq 1c$	174.592 ± 26.191	169.618 ± 17.513	204.831 ± 30.385	154.454 ± 19.239	74.847 ± 12.204	19.633 ± 5.926	19.148 ± 4.078
$t\bar{t} +$ light	60.031 ± 10.026	53.613 ± 7.515	85.037 ± 14.706	83.392 ± 11.275	40.775 ± 6.807	16.509 ± 6.320	6.708 ± 2.779
Other t	14.666 ± 12.058	14.332 ± 13.081	17.596 ± 10.137	15.550 ± 13.397	9.816 ± 8.762	2.862 ± 0.774	2.366 ± 0.763
Single-t tW	14.546 ± 7.427	20.672 ± 11.990	42.992 ± 20.252	35.615 ± 20.374	25.128 ± 17.148	9.838 ± 8.383	4.893 ± 3.569
$t\bar{t}Z$	38.987 ± 5.625	41.769 ± 6.094	56.289 ± 8.156	48.832 ± 7.979	28.651 ± 6.286	7.397 ± 1.154	6.996 ± 1.557
$t\bar{t}W$	3.402 ± 0.597	4.644 ± 0.720	7.158 ± 1.161	6.148 ± 1.016	4.501 ± 0.700	2.843 ± 0.482	0.710 ± 0.129
V+jets	25.246 ± 9.671	29.849 ± 10.980	50.991 ± 18.297	50.637 ± 18.924	39.552 ± 14.784	19.375 ± 7.499	6.806 ± 2.635
Diboson	1.239 ± 0.727	2.294 ± 1.269	4.771 ± 2.492	4.420 ± 2.514	3.079 ± 1.607	1.413 ± 0.750	0.737 ± 0.379
$t\bar{t}t\bar{t}$	8.720 ± 2.620	11.769 ± 3.525	18.434 ± 5.517	14.797 ± 4.431	8.159 ± 2.439	4.285 ± 1.284	1.958 ± 0.589
Fakes	21.335 ± 11.064	21.691 ± 11.803	19.969 ± 10.706	22.799 ± 12.483	17.150 ± 11.138	0	6.077 ± 3.914
Total	1595.480 ± 41.544	1578.650 ± 40.416	1923.130 ± 44.046	1419.710 ± 31.762	671.620 ± 21.512	238.113 ± 14.065	189.611 ± 9.108
Data	0	0	0	0	0	0	0

Table A.8.: Background-only post-fit yields of the single-lepton channel control regions. The errors include all systematic uncertainties.

	Boosted CR	$t\bar{t}+b$	$t\bar{t}+B$	$t\bar{t}+\geq 2b$	$t\bar{t}+\geq 1c$	$t\bar{t}+$ light
$t\bar{t}H$	0	0	0	0	0	0
tH	0.860 ± 0.186	16.436 ± 1.804	7.048 ± 0.713	15.008 ± 1.795	9.064 ± 1.026	14.513 ± 1.498
$t\bar{t}+b$	54.415 ± 12.822	12767.4 ± 662.055	5780.35 ± 545.811	3991.85 ± 515.869	5363.34 ± 533.069	5031.11 ± 359.877
$t\bar{t}+B$	35.741 ± 5.500	1927.97 ± 180.687	5551.8 ± 594.794	1020.13 ± 357.138	1440.85 ± 320.019	933.972 ± 117.724
$t\bar{t}+\geq 2b$	146.846 ± 9.285	4925.62 ± 385.802	3715.65 ± 392.671	13844.1 ± 759.21	4269.79 ± 383.966	2283.02 ± 257.088
$t\bar{t}+\geq 1c$	111.27 ± 12.885	6405.68 ± 617.127	4333.41 ± 445.047	3328.58 ± 366.514	14965.9 ± 904.288	14537.3 ± 1831.73
$t\bar{t}+$ light	49.791 ± 7.450	4578.55 ± 298.162	2275.28 ± 162.779	1278.29 ± 114.151	7776.78 ± 696.42	36353.7 ± 1748.28
Other-t	3.982 ± 0.988	364.462 ± 208.123	274.687 ± 157.818	404.056 ± 282.012	463.698 ± 271.869	445.726 ± 246.296
Single-t tW	7.332 ± 4.547	663.861 ± 191.931	455.937 ± 137.706	630.922 ± 284.971	613.936 ± 171.84	904.614 ± 144.917
$t\bar{t}Z$	6.656 ± 1.391	156.729 ± 21.253	153.249 ± 21.775	296.509 ± 38.527	190.34 ± 24.265	155.409 ± 19.282
$t\bar{t}W$	1.767 ± 0.291	31.106 ± 5.885	37.699 ± 5.983	48.003 ± 7.311	86.584 ± 13.045	75.251 ± 12.167
V+jets	9.493 ± 3.569	465.222 ± 156.463	332.251 ± 113.721	1231.85 ± 441.852	642.271 ± 219.173	633.59 ± 221.152
Diboson	1.156 ± 0.728	22.96 ± 11.805	21.028 ± 10.74	56.689 ± 28.909	40.773 ± 21.019	36.048 ± 18.464
$t\bar{t}t\bar{t}$	2.452 ± 0.737	11.857 ± 3.536	32.403 ± 9.639	120.701 ± 36.085	41.189 ± 12.262	4.892 ± 1.46
Fakes	4.707 ± 2.684	638.44 ± 330.391	250.888 ± 131.468	730.419 ± 386.9	616.003 ± 329.817	574.643 ± 285.949
Total	436.467 ± 16.338	32976.3 ± 179.841	23221.7 ± 149.041	26997.1 ± 180.703	36520.6 ± 195.879	61983.8 ± 254.758
Data	426	32887	23245	27089	36528	61954

Table A.9.: Background-only post-fit yields of the dilepton channel SRs. The errors include all systematic uncertainties.

	STXS 1	STXS 2	STXS 3	STXS 4	STXS 5	STXS 6	Boosted STXS 5	Boosted STXS 6
$t\bar{t}H$	0	0	0	0	0	0	0	0
tH	0.070 ± 0.030	0.258 ± 0.081	0.295 ± 0.058	0.359 ± 0.051	0.308 ± 0.056	0.032 ± 0.010	0.263 ± 0.039	0.065 ± 0.019
$t\bar{t} + b$	13.109 ± 2.633	14.190 ± 2.934	18.968 ± 3.123	117.376 ± 2.458	8.306 ± 3.762	0.894 ± 0.272	6.613 ± 1.487	1.106 ± 0.348
$t\bar{t} + B$	3.063 ± 1.680	4.520 ± 2.651	5.782 ± 2.126	5.243 ± 1.571	3.660 ± 1.484	1.977 ± 0.982	2.859 ± 0.981	1.108 ± 0.663
$t\bar{t} + \geq 2b$	83.251 ± 4.984	78.354 ± 5.036	80.422 ± 5.789	559.062 ± 3.766	21.712 ± 2.025	3.041 ± 0.472	24.985 ± 2.083	7.636 ± 1.596
$t\bar{t} + \geq 1c$	10.396 ± 1.565	10.689 ± 1.649	14.153 ± 1.991	110.590 ± 1.648	3.127 ± 0.883	0.859 ± 0.344	3.473 ± 0.861	0.906 ± 0.341
$t\bar{t} +$ light	1.580 ± 0.764	1.092 ± 0.458	1.094 ± 0.739	0.827 ± 0.373	0.453 ± 0.278	0.046 ± 0.043	0.589 ± 0.356	0.220 ± 0.185
Other t	0.061 ± 0.024	0.058 ± 0.037	0.286 ± 0.136	0.384 ± 0.166	0.102 ± 0.042	0.106 ± 0.064	0.174 ± 0.072	0.214 ± 0.088
Single-t tW	1.012 ± 0.610	2.605 ± 1.361	3.873 ± 3.157	5.165 ± 3.121	2.216 ± 1.676	0.673 ± 0.562	2.108 ± 1.704	0.987 ± 1.001
$t\bar{t}Z$	4.329 ± 0.711	5.114 ± 0.772	6.265 ± 0.966	7.277 ± 1.406	3.037 ± 0.727	0.403 ± 0.336	4.357 ± 0.754	1.447 ± 0.384
$t\bar{t}W$	0.582 ± 0.204	1.009 ± 0.201	1.687 ± 0.258	1.821 ± 0.281	0.903 ± 0.149	0.458 ± 0.125	1.483 ± 0.328	1.101 ± 0.422
V+jets	2.341 ± 0.707	3.261 ± 1.069	5.347 ± 1.638	6.174 ± 1.884	1.904 ± 0.868	1.003 ± 0.338	3.583 ± 1.067	1.353 ± 0.491
$t\bar{t}t\bar{t}$	1.408 ± 0.427	1.816 ± 0.544	2.723 ± 0.814	2.440 ± 0.730	1.179 ± 0.354	0.501 ± 0.152	1.448 ± 0.433	0.840 ± 0.253
Fakes	3.453 ± 1.627	5.034 ± 2.372	6.918 ± 3.260	4.475 ± 2.109	3.099 ± 1.461	0.560 ± 0.264	4.239 ± 1.998	1.368 ± 0.645
Total	124.656 ± 5.197	127.999 ± 5.013	147.812 ± 6.321	121.194 ± 4.254	50.007 ± 4.696	10.553 ± 1.322	56.173 ± 2.932	18.350 ± 1.963
Data	0	0	0	0	0	0	0	0

Table A.10.: Background-only post-fit yields of the dilepton channel control regions. The errors include all systematic uncertainties.

	$t\bar{t}+b$	$t\bar{t}+B$	$t\bar{t}+\geq 2b$	$t\bar{t}+\geq 1c$	$t\bar{t}+$ light
$t\bar{t}H$	0	0	0	0	0
tH	2.867 ± 0.289	1.269 ± 0.135	2.567 ± 0.265	2.110 ± 0.195	1.221 ± 0.119
$t\bar{t}+b$	3768.92 ± 196.113	1488.69 ± 135.123	1058.46 ± 148.469	1796.11 ± 137.705	1449.67 ± 112.554
$t\bar{t}+B$	490.018 ± 49.080	1443.08 ± 113.182	268.962 ± 86.720	405.792 ± 68.077	278.025 ± 32.003
$t\bar{t}+\geq 2b$	936.765 ± 105.153	677.209 ± 85.004	2880.35 ± 177.919	936.925 ± 79.846	523.991 ± 44.637
$t\bar{t}+\geq 1c$	2036.71 ± 160.922	1383.78 ± 123.638	986.86 ± 92.870	7041.6 ± 361.509	5300.06 ± 343.444
$t\bar{t}+$ light	689.684 ± 66.164	434.969 ± 36.103	224.101 ± 28.188	3129.78 ± 246.549	10001.2 ± 352.454
Other-t	4.567 ± 1.985	2.184 ± 0.732	12.944 ± 5.319	8.583 ± 3.183	6.126 ± 2.450
Single-t tW	182.162 ± 52.713	121.352 ± 37.211	225.243 ± 103.472	178.278 ± 56.754	275.555 ± 47.655
$t\bar{t}Z$	34.203 ± 5.194	38.113 ± 6.045	99.194 ± 13.853	89.989 ± 11.406	46.074 ± 8.551
$t\bar{t}W$	9.898 ± 1.424	13.916 ± 2.851	23.416 ± 3.556	58.471 ± 8.415	33.17 ± 4.755
V+jets	266.158 ± 77.604	129.025 ± 37.514	432.695 ± 125.206	327.564 ± 95.406	373.347 ± 110.088
$t\bar{t}t\bar{t}$	1.622 ± 0.487	7.947 ± 2.366	46.959 ± 14.014	16.046 ± 4.774	2.674 ± 0.798
Fakes	126.803 ± 59.760	113.473 ± 53.478	204.094 ± 96.186	325.872 ± 153.578	330.374 ± 155.7
Total	8550.37 ± 84.782	5855 ± 61.647	6465.84 ± 75.743	14317.1 ± 107.757	18621.5 ± 135.087
Data	8624	5830	6448	14361	18557

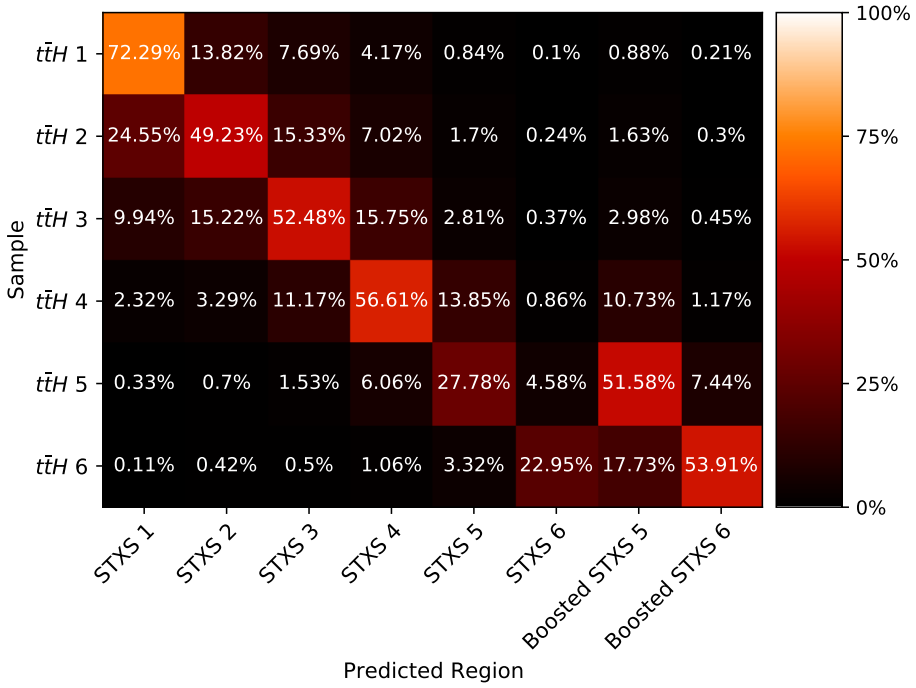


Figure A.6.: Pre-fit expected signal contributions in the different STXS SRs in the dilepton channel. The different $t\bar{t}H$ process indices indicate the true STXS bin, determined based on the $\hat{p}_{\mathrm{T}}^{H}$ and Table 11.2.

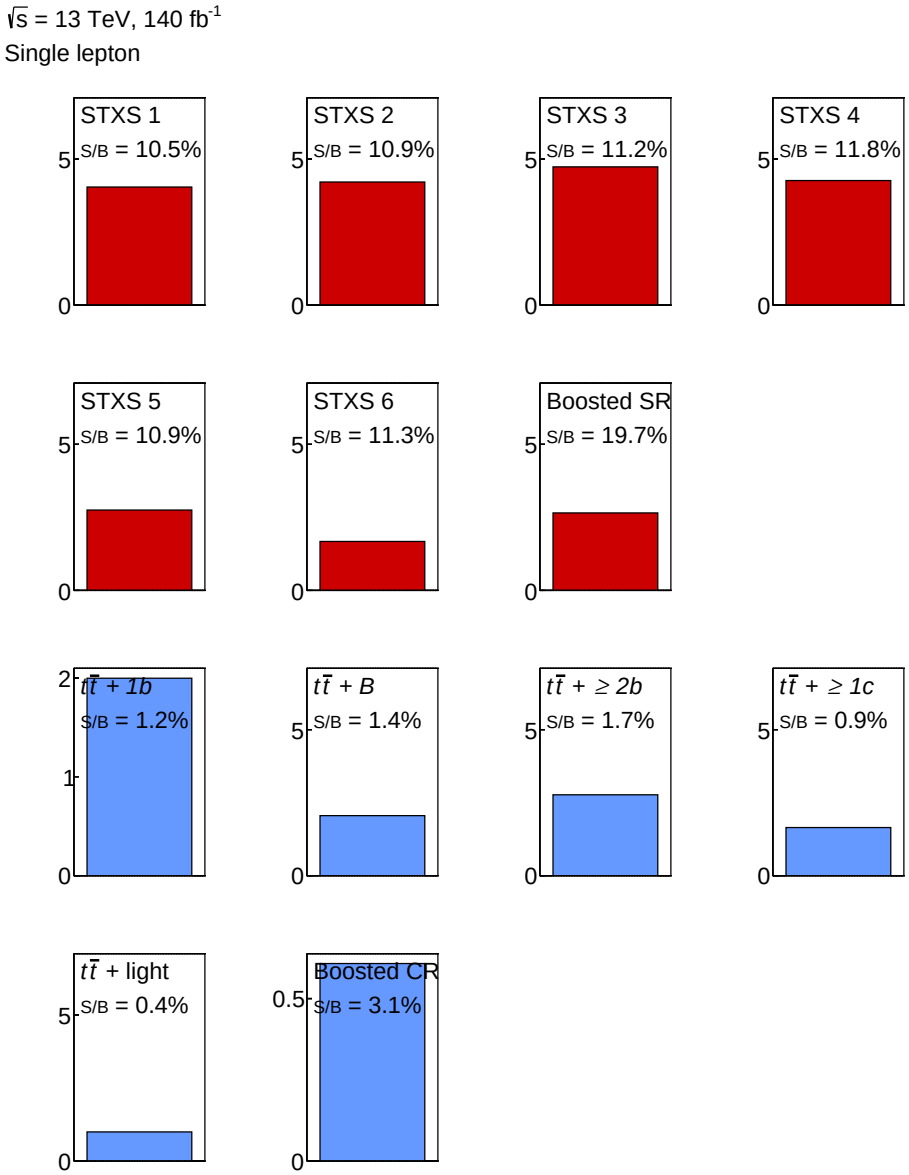


Figure A.7.: Pre-fit signal to background ratio in the different single-lepton SRs and CRs.

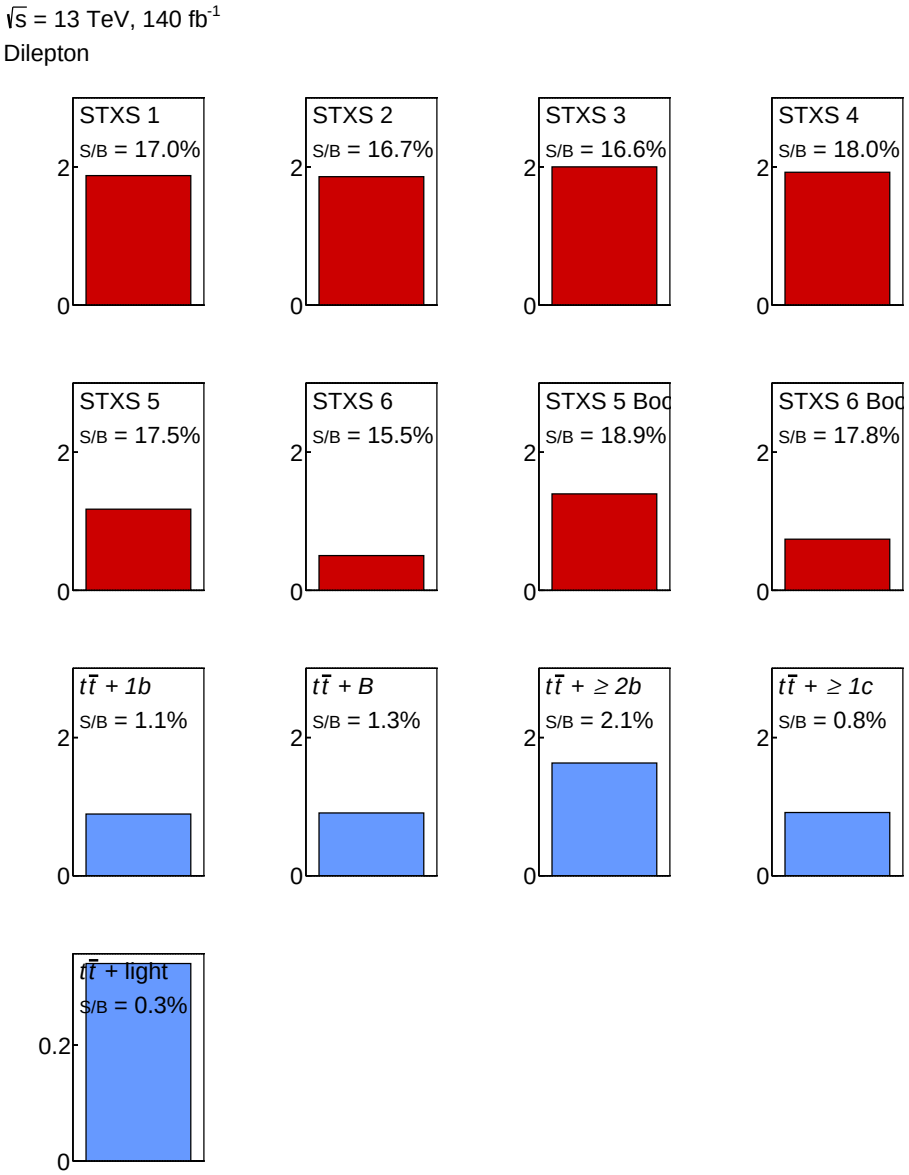
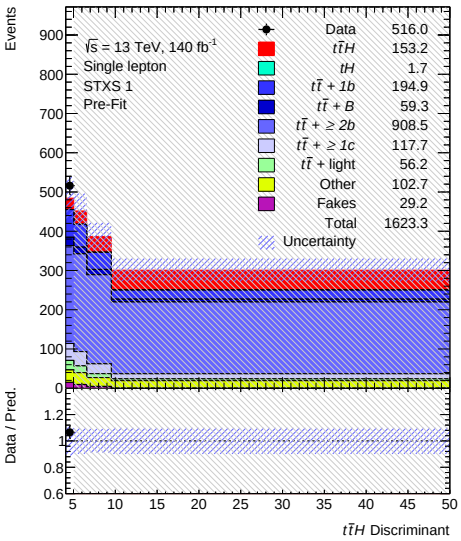
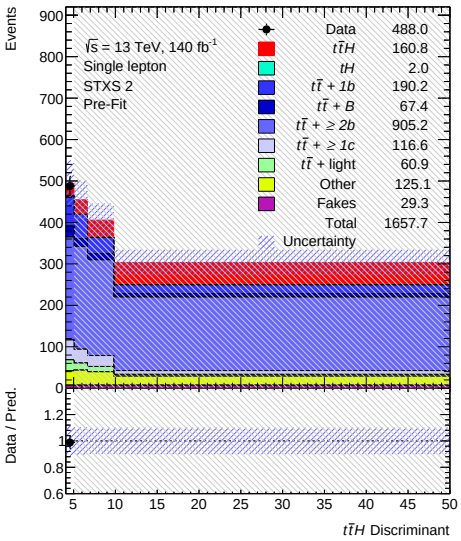


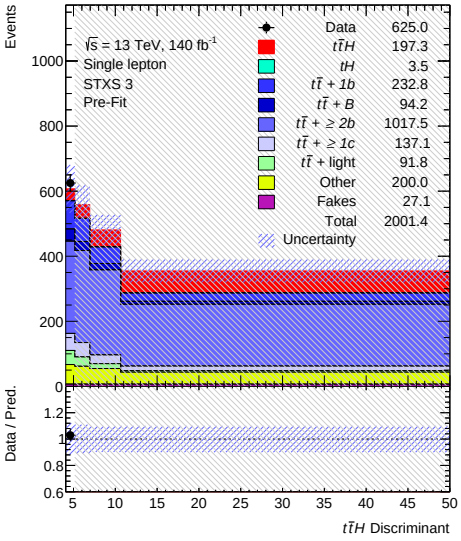
Figure A.8.: Pre-fit signal to background ratio in the different dilepton SRs and CRs.



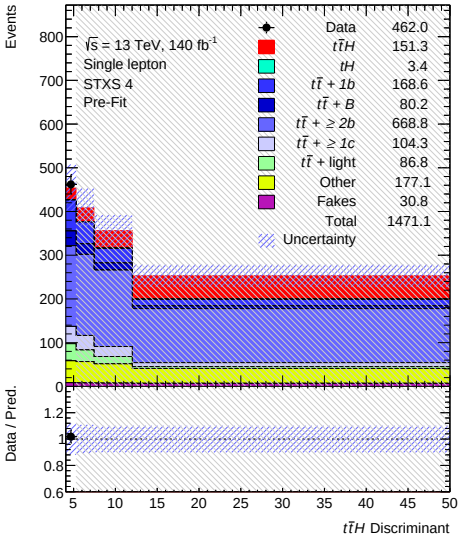
(a): STXS 1



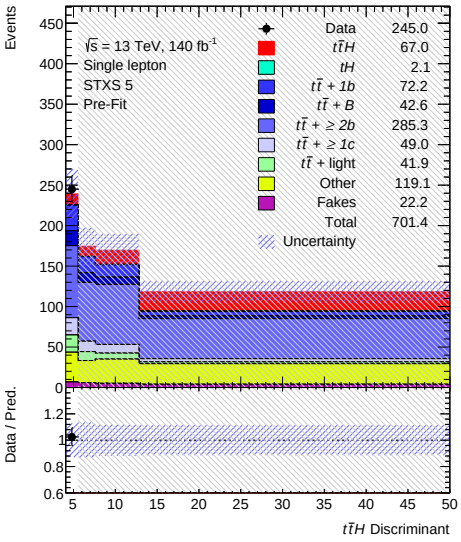
(b): STXS 2



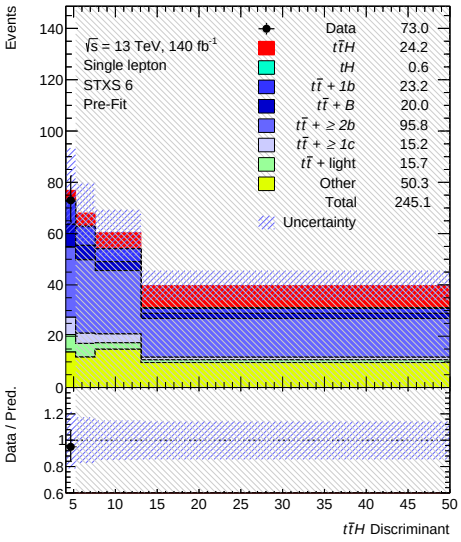
(c): STXS 3



(d): STXS 4

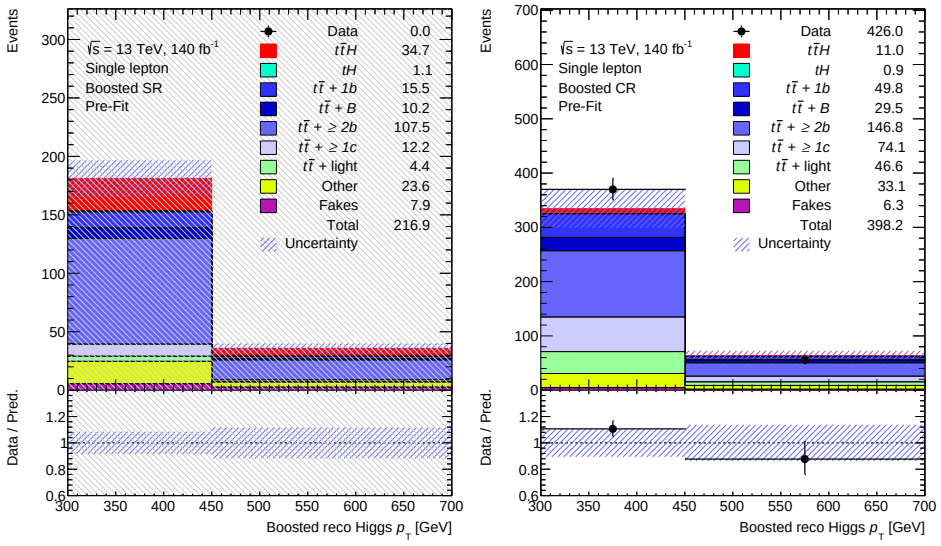


(e): STXS 5



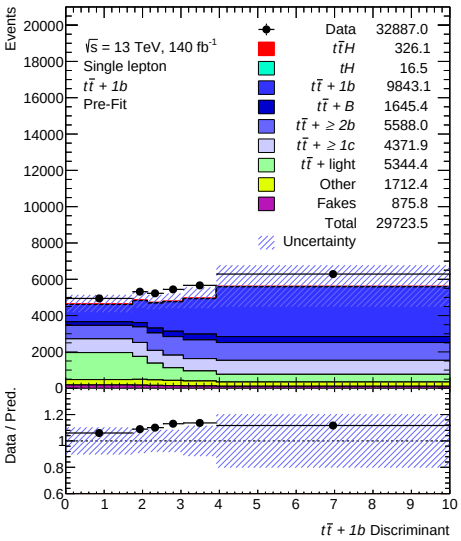
(f): STXS 6

Figure A.9.: Pre-fit single-lepton resolved SRs.

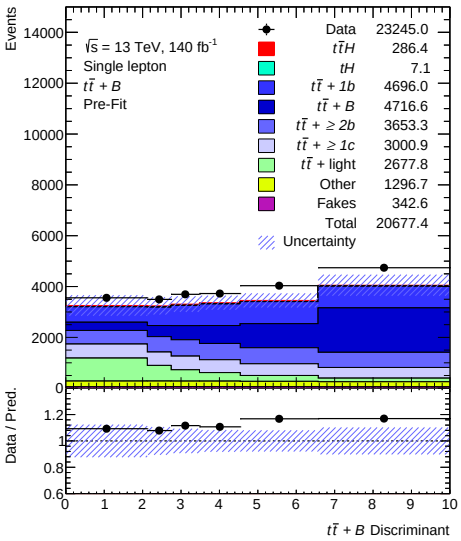


(a): Boosted SR (b): Boosted CR

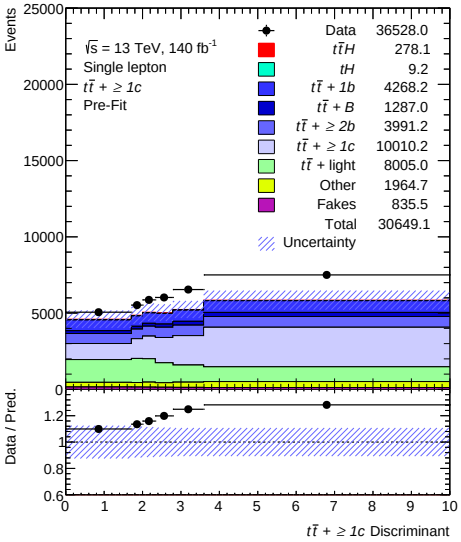
Figure A.10.: Pre-fit single-lepton boosted SR and CR.



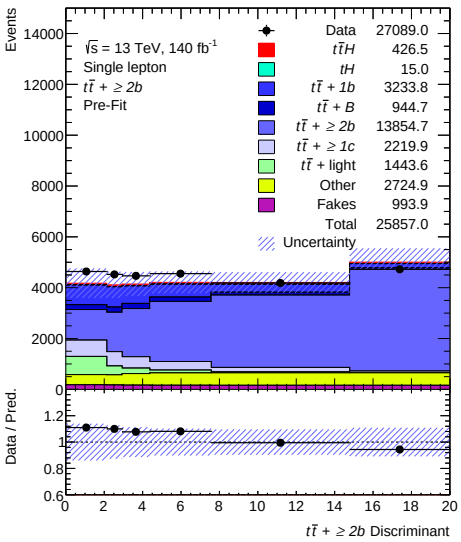
(a): $t\bar{t} + b$ region



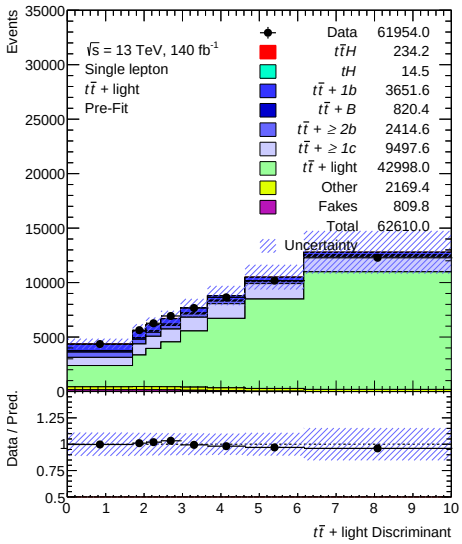
(b): $t\bar{t} + B$ region



(c): $t\bar{t} + \geq 1c$ region

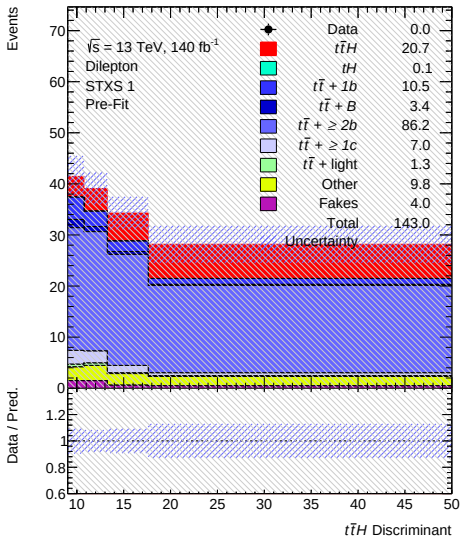


(d): $t\bar{t} + \geq 2b$ region

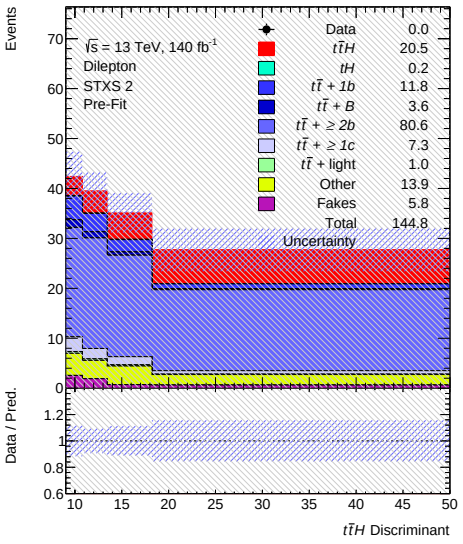


(e): $t\bar{t} +$ light region

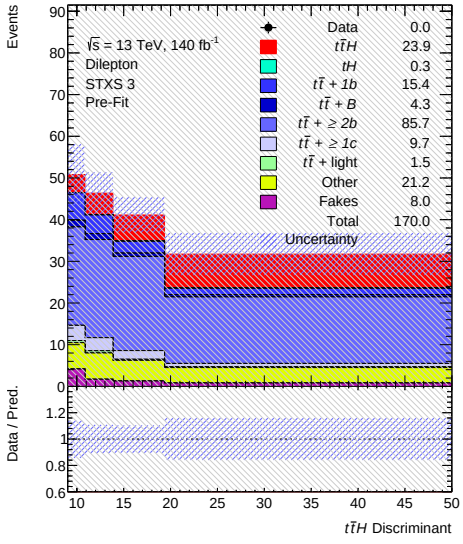
Figure A.11.: Pre-fit single-lepton resolved CRs.



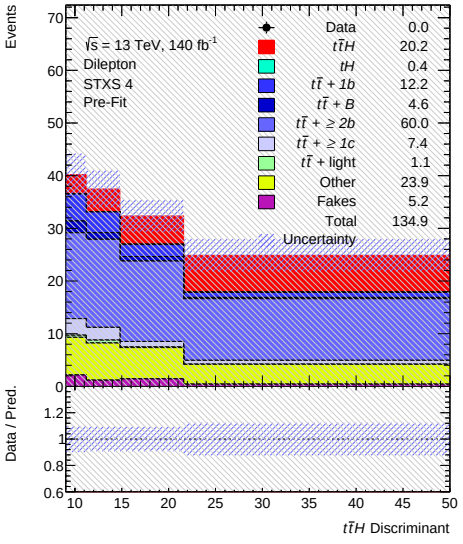
(a): STXS 1



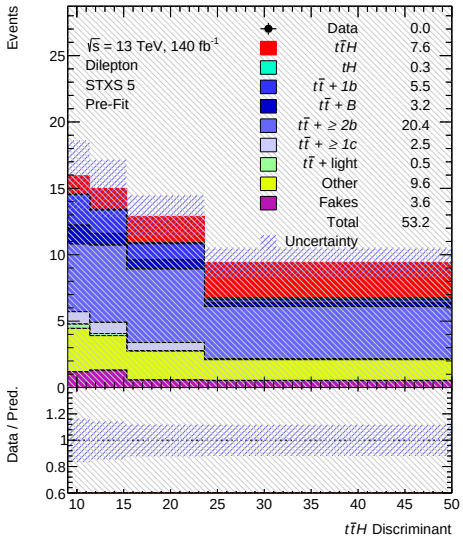
(b): STXS 2



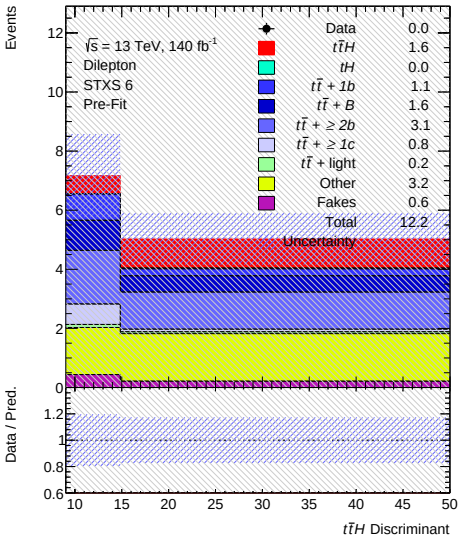
(c): STXS 3



(d): STXS 4

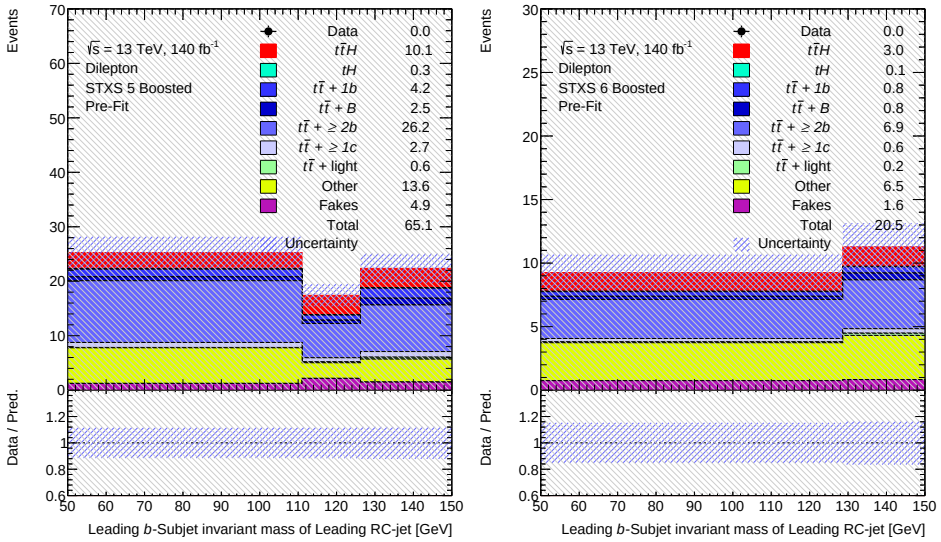


(e): STXS 5



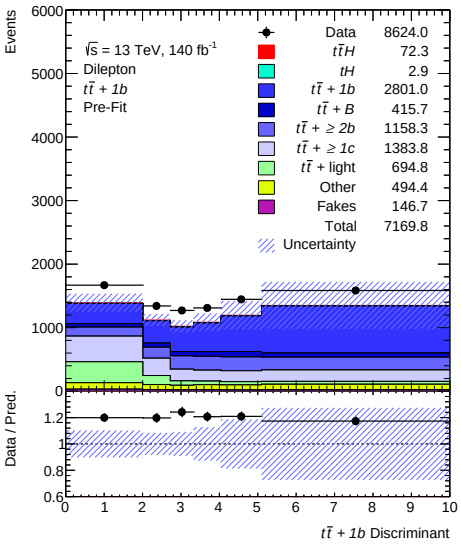
(f): STXS 6

Figure A.12.: Pre-fit dilepton resolved SRs.

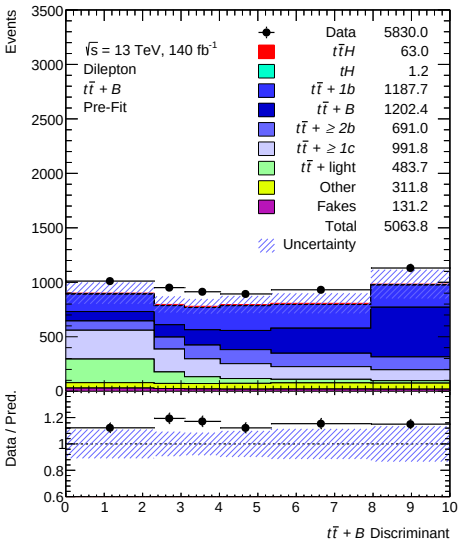


(a): Boosted STXS 5 (b): Boosted STXS 6

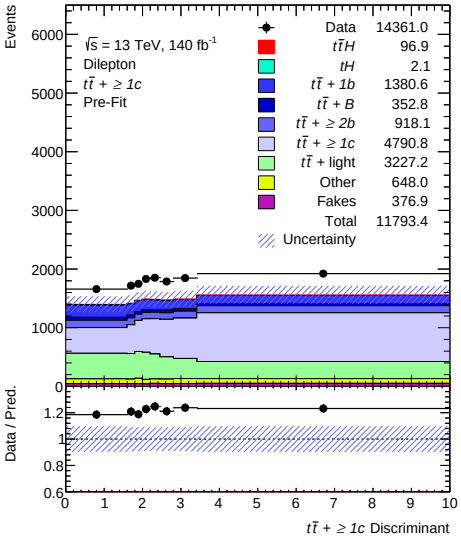
Figure A.13.: Pre-fit dilepton boosted SRs.



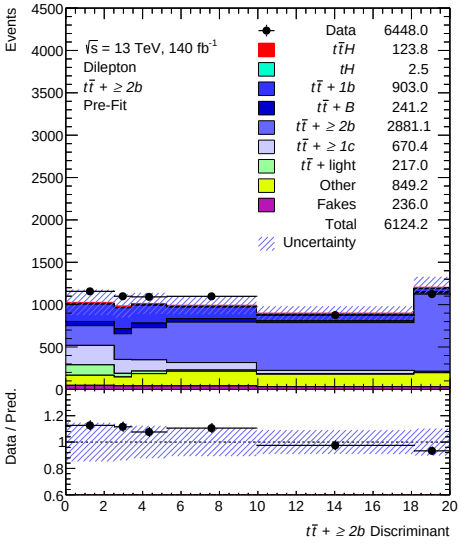
(a): $t\bar{t} + b$ region



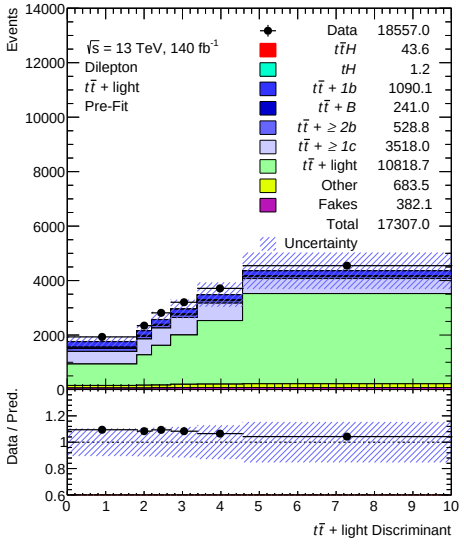
(b): $t\bar{t} + B$ region



(c): $t\bar{t} + \geq 1c$ region



(d): $t\bar{t} + \geq 2b$ region



(e): $t\bar{t} +$ light region

Figure A.14.: Pre-fit dilepton resolved CRs.

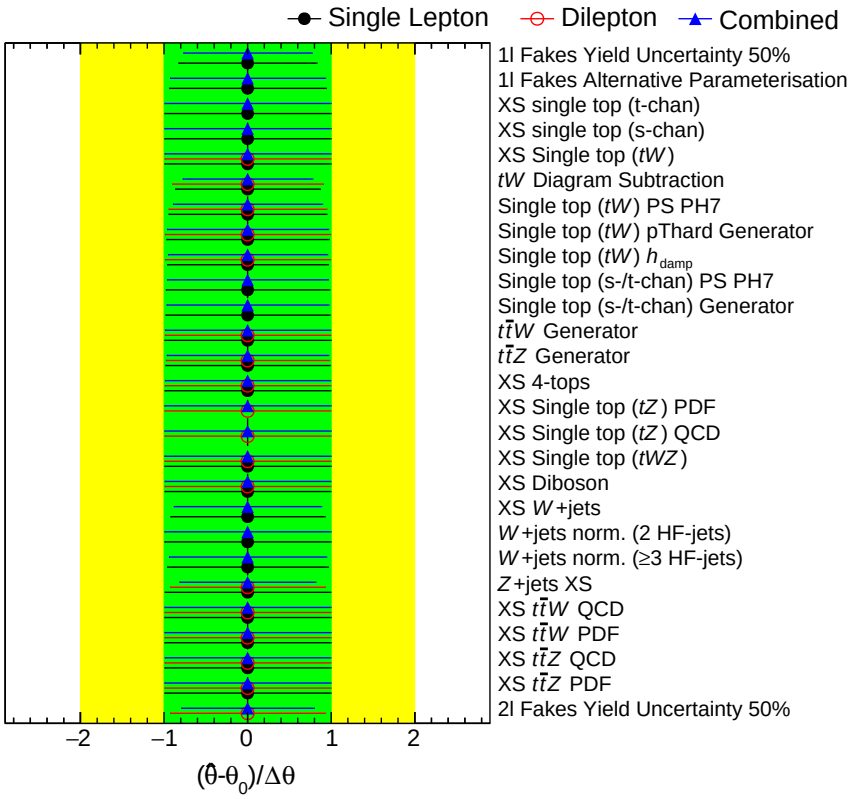


Figure A.15.: Other processes modelling NPs pulls for the different Asimov fits.

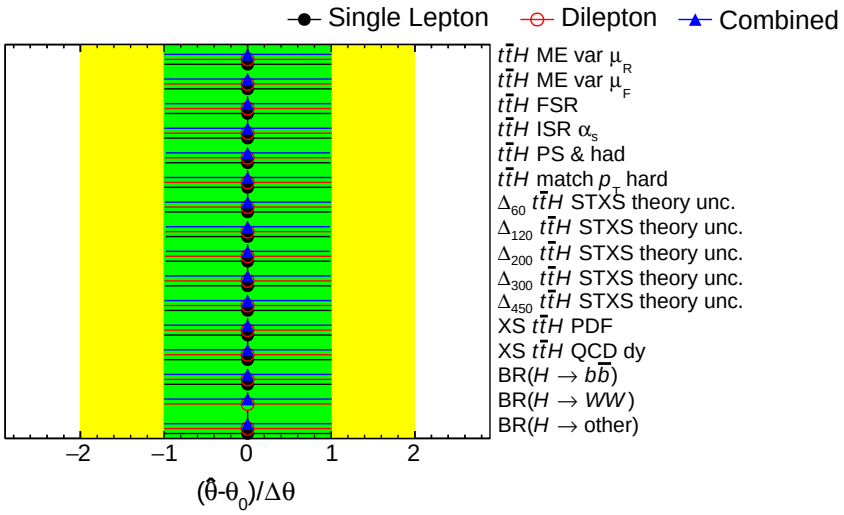


Figure A.16.: Signal process modelling NPs pulls for the different Asimov fits.

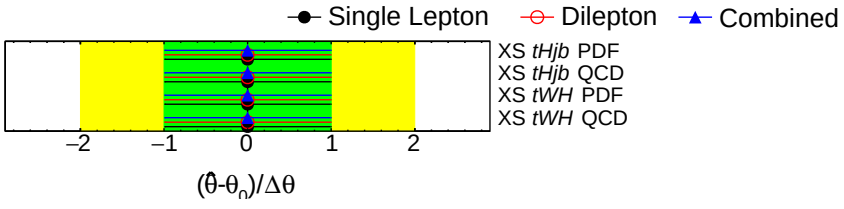


Figure A.17.: tH process modelling NPs pulls for the different Asimov fits.

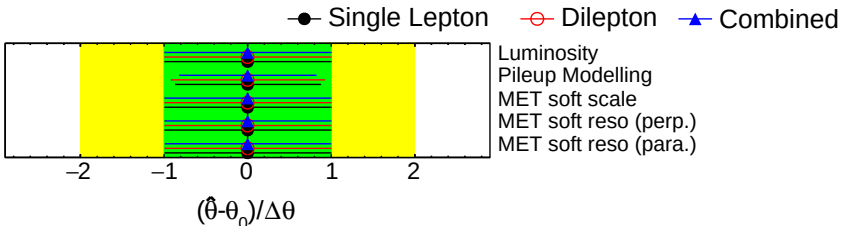


Figure A.18.: Instrumental/Experimental NPs pulls for the different Asimov fits.

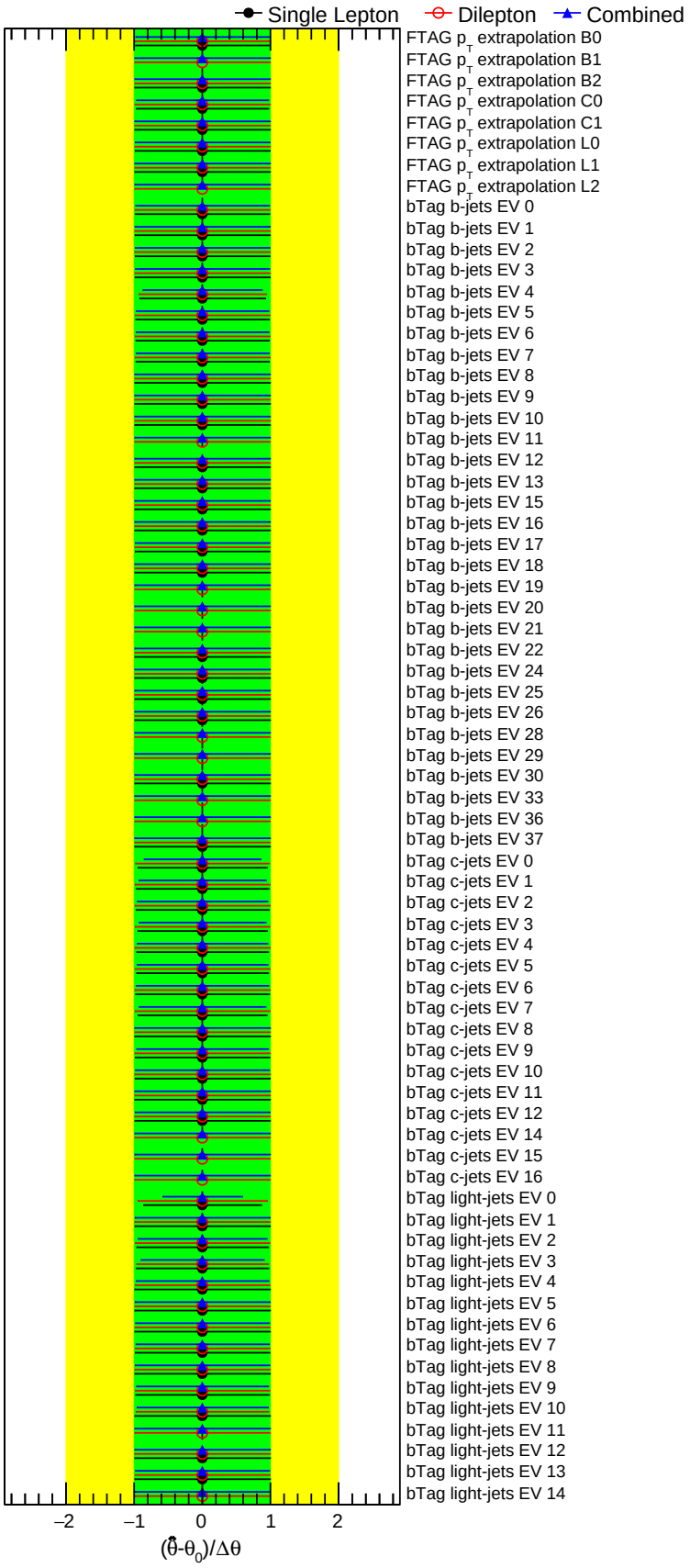


Figure A.19.: b-tagging NPs pulls for the different Asimov fits.

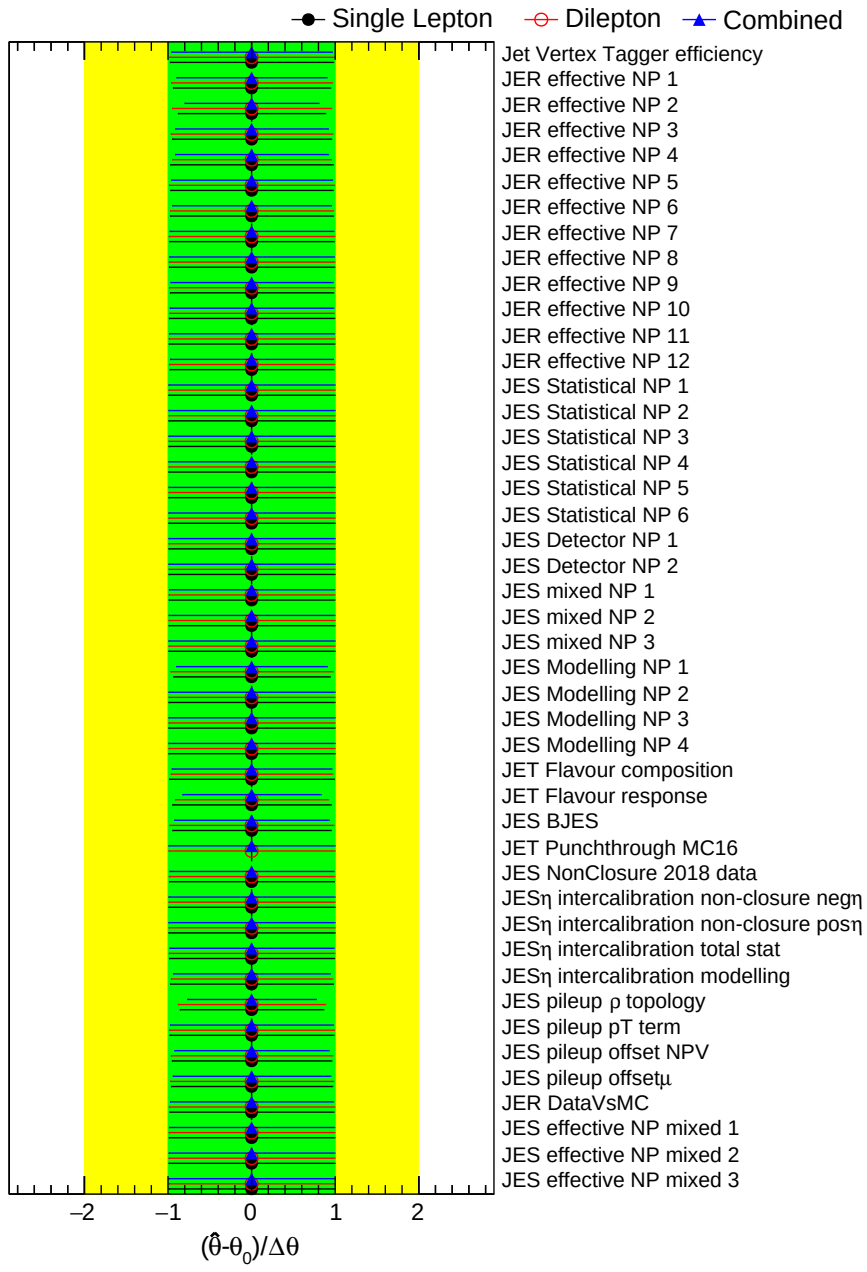


Figure A.20.: Jets related NPs pulls for the different Asimov fits.

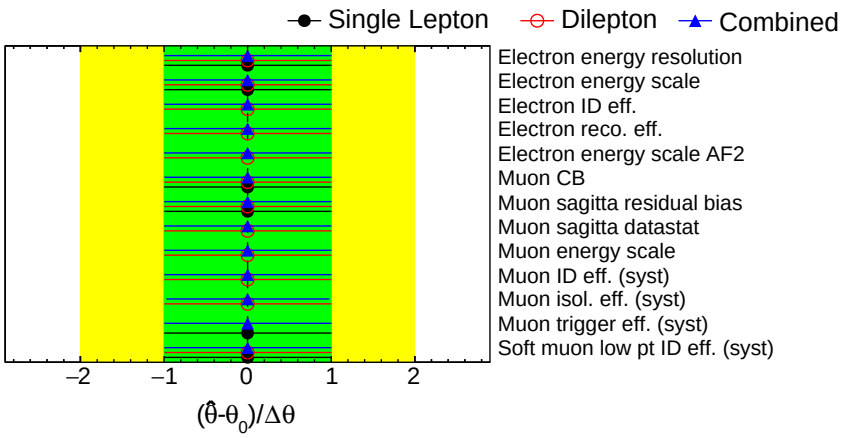
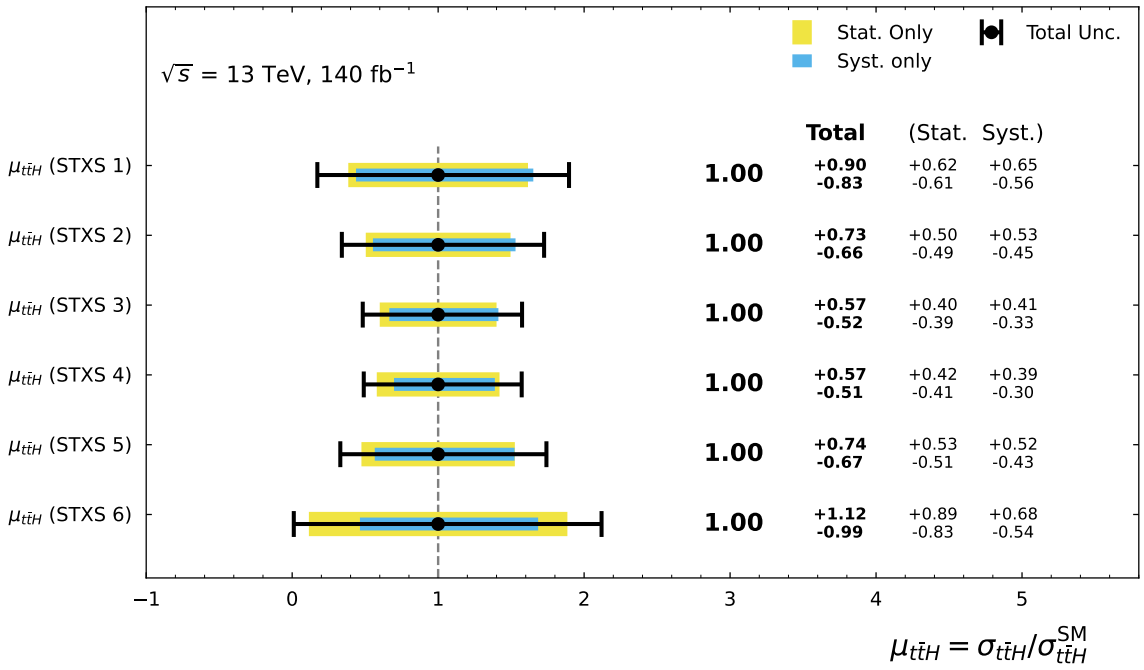
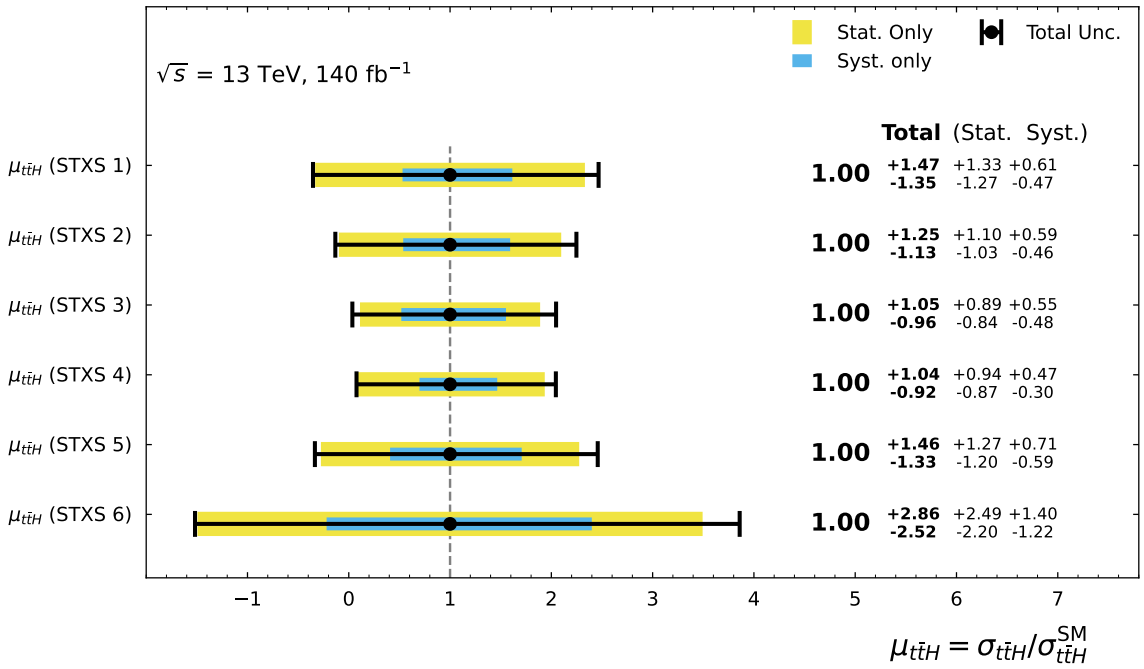


Figure A.21.: Leptons related NPs pulls for the different Asimov fits.



(a): Single-lepton fit



(b): Dilepton fit

Figure A.22.: Best-fit signal strength POIs for the different STXS regions and their expected uncertainties for the different per-channel fits. For both fits, the first and last (1 and 6) STXS bins show the largest uncertainties, while the expected uncertainties of the dilepton channel are larger compared to the single-lepton channel. This is expected due to the difference in statistics between the single-lepton and the dilepton channel.

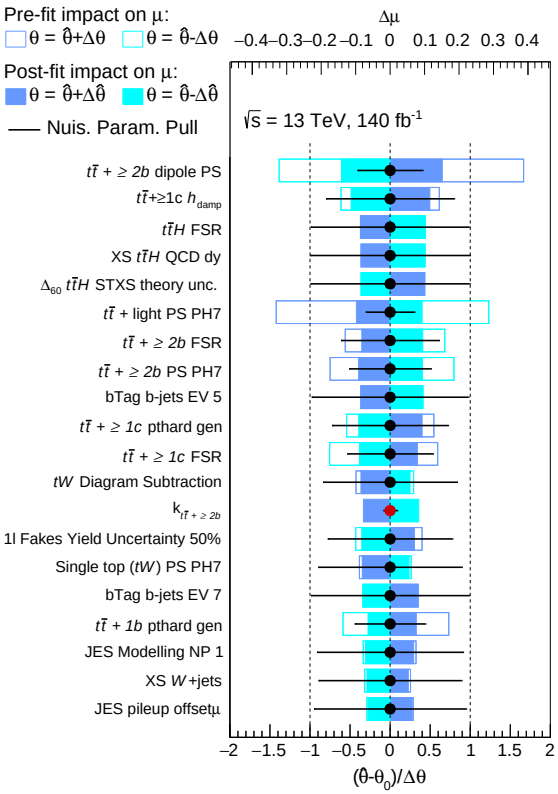


Figure A.23.: Ranking of the 20 NPs with the largest post-fit impact on $\mu_{t\bar{t}H}$ (STXS 1). The empty boxes visualise the pre-fit, and the filled boxes the post-fit impact on $\mu_{t\bar{t}H}$ (STXS 1). While black dots are used for NPs, red dots are employed for normalisation factors.

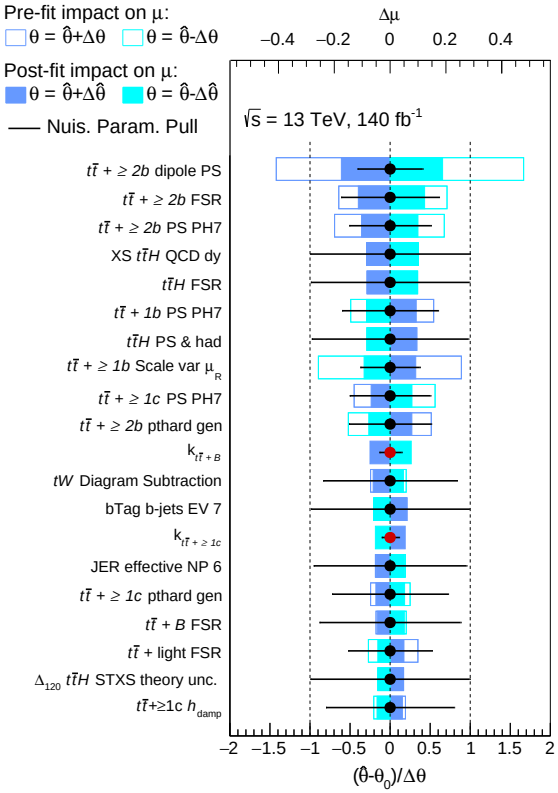


Figure A.24.: Ranking of the 20 NPs with the largest post-fit impact on $\mu_{t\bar{t}H}$ (STXS 2). The empty boxes visualise the pre-fit, and the filled boxes the post-fit impact on $\mu_{t\bar{t}H}$ (STXS 2). While black dots are used for NPs, red dots are employed for normalisation factors.

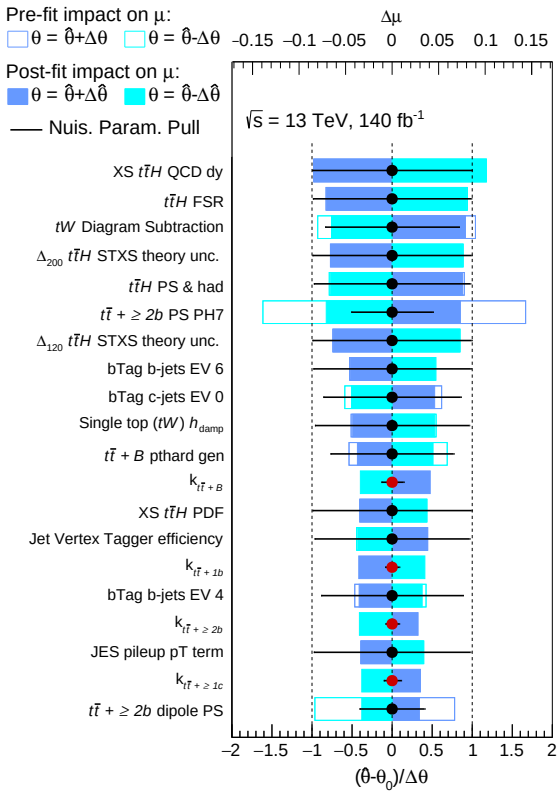


Figure A.25.: Ranking of the 20 NPs with the largest post-fit impact on $\mu_{t\bar{t}H}$ (STXS 4). The empty boxes visualise the pre-fit, and the filled boxes the post-fit impact on $\mu_{t\bar{t}H}$ (STXS 4). While black dots are used for NPs, red dots are employed for normalisation factors.

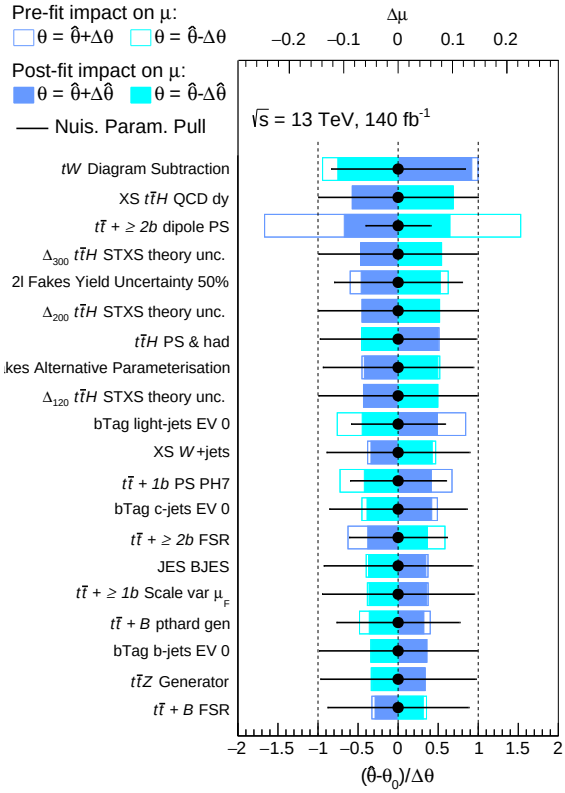


Figure A.26.: Ranking of the 20 NPs with the largest post-fit impact on $\mu_{t\bar{t}H}$ (STXS 5). The empty boxes visualise the pre-fit, and the filled boxes the post-fit impact on $\mu_{t\bar{t}H}$ (STXS 5). While black dots are used for NPs, red dots are employed for normalisation factors.

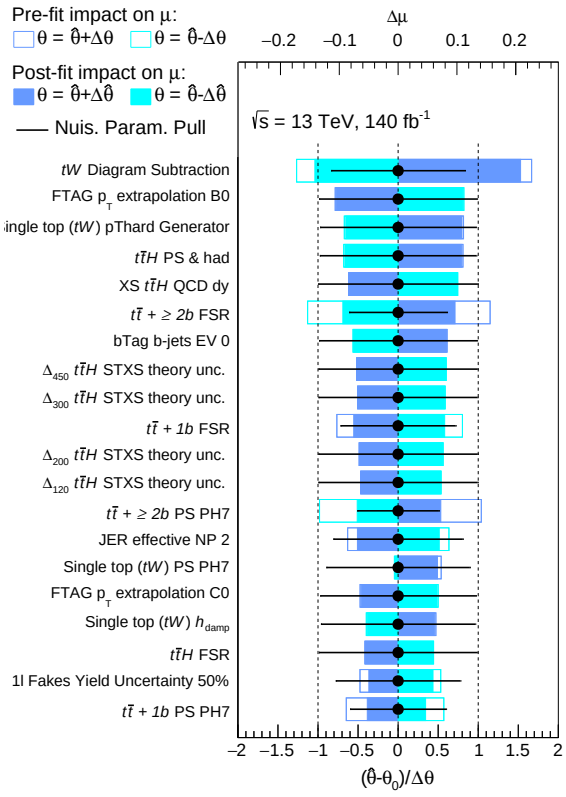


Figure A.27.: Ranking of the 20 NPs with the largest post-fit impact on $\mu_{t\bar{t}H}$ (STXS 6). The empty boxes visualise the pre-fit, and the filled boxes the post-fit impact on $\mu_{t\bar{t}H}$ (STXS 6). While black dots are used for NPs, red dots are employed for normalisation factors.

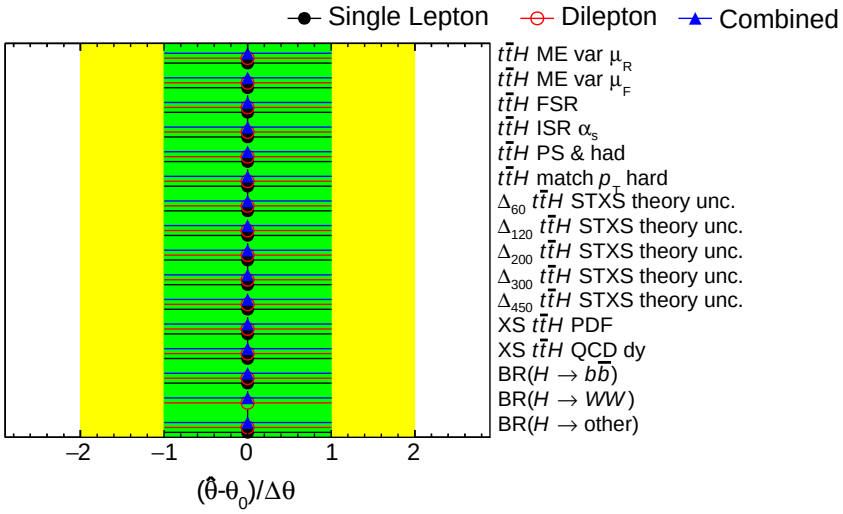


Figure A.28.: Signal process modelling NPs pulls for the different background-only fits.

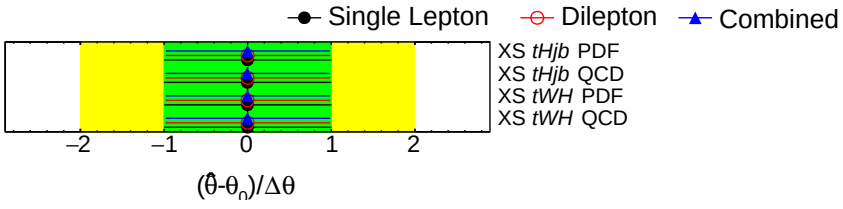


Figure A.29.: tH process modelling NPs pulls for the different background-only fits.

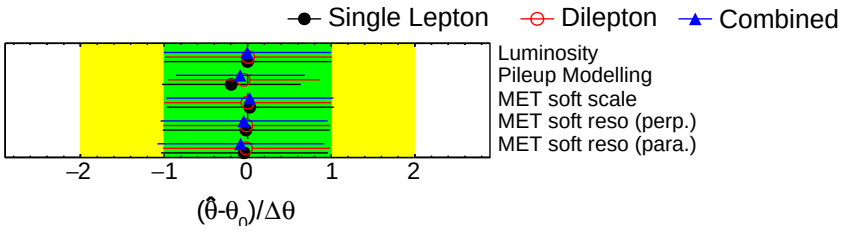


Figure A.30.: Instrumental/Experimental NPs pulls for the different background-only fits.

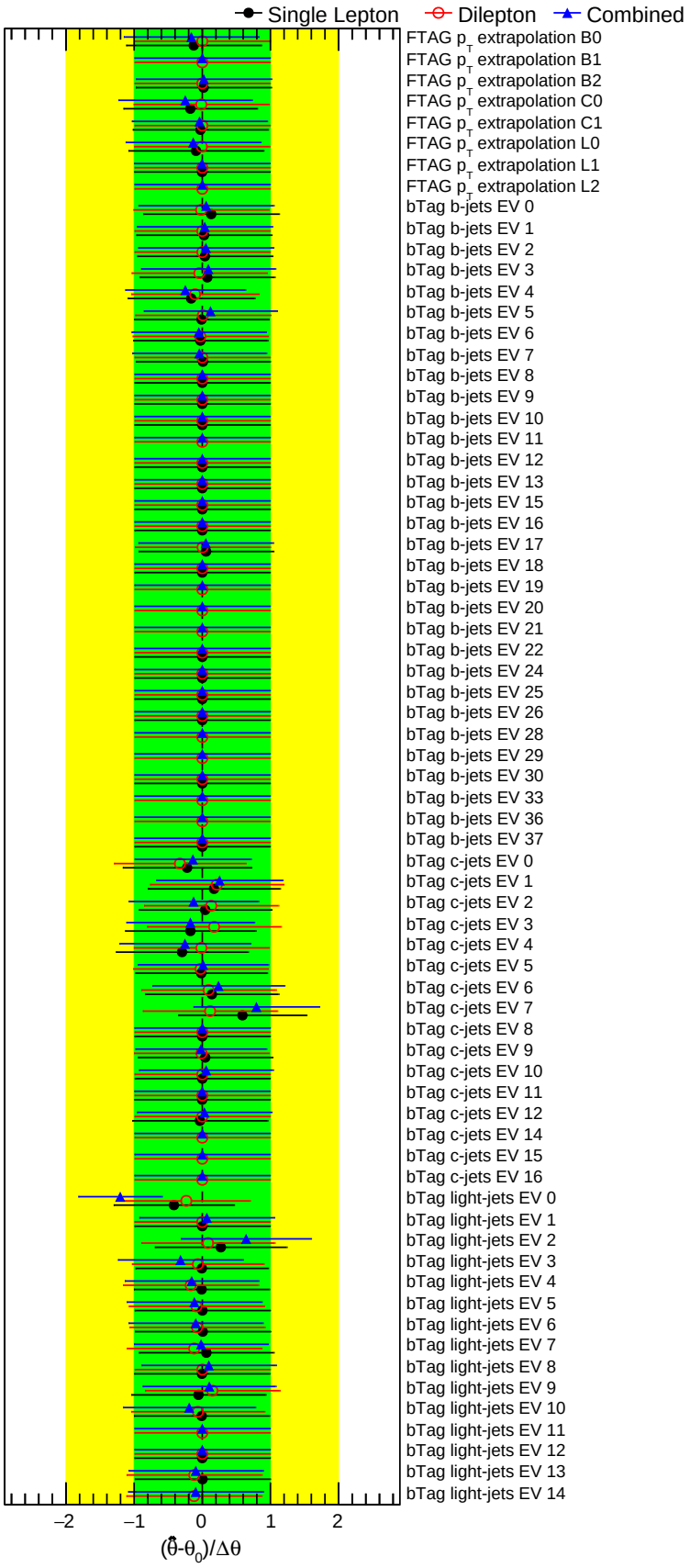


Figure A.31.: b-tagging NPs pulls for the different background-only fits.

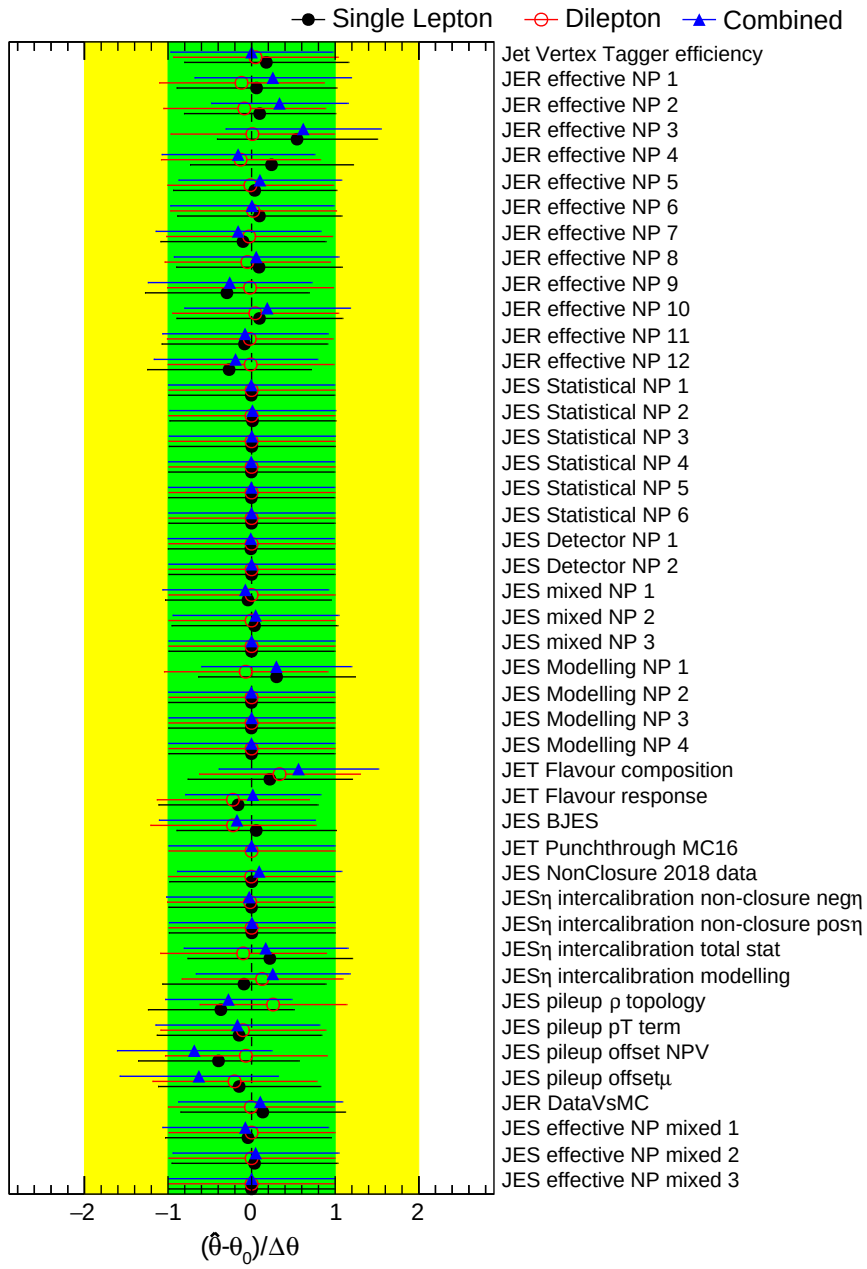


Figure A.32.: Jets related NPs pulls for the different background-only fits.

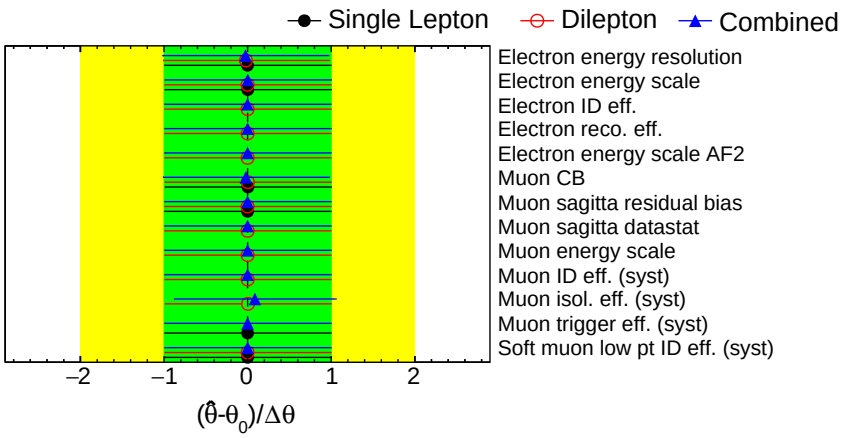
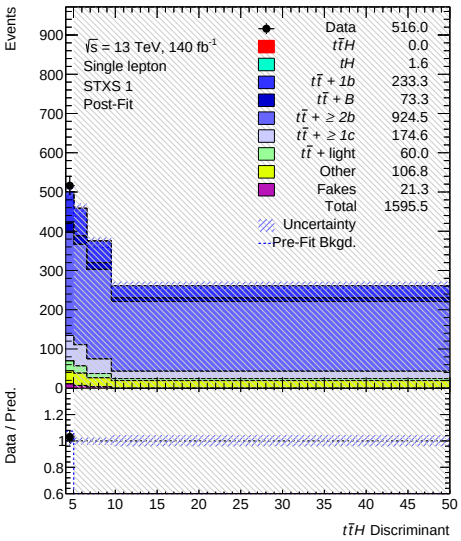
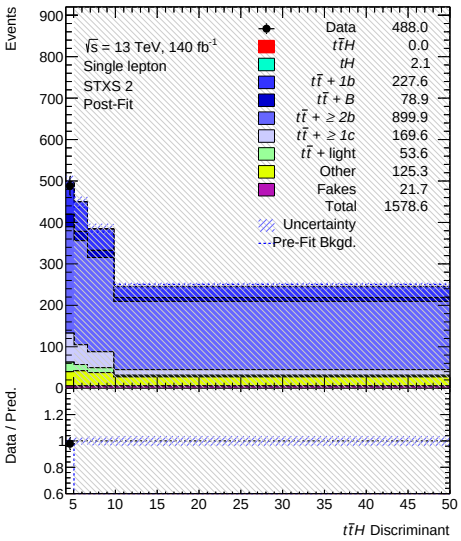


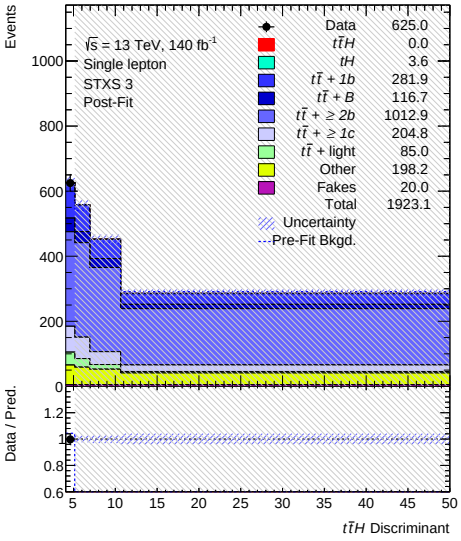
Figure A.33.: Leptons related NPs pulls for the different background-only fits.



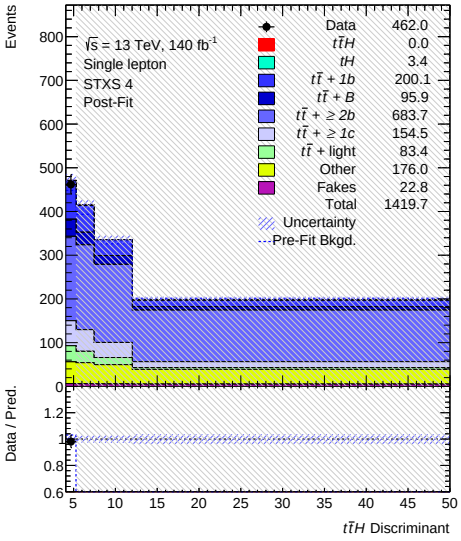
(a): STXS 1



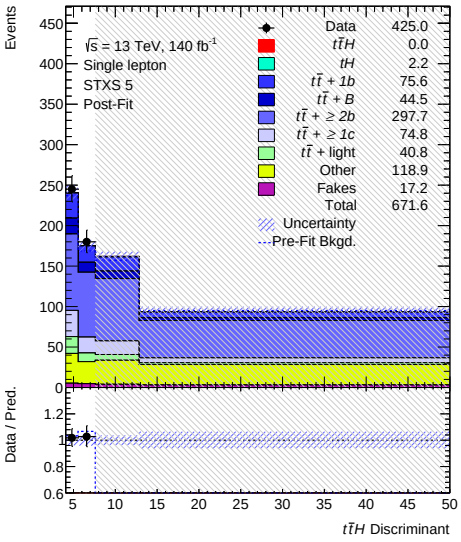
(b): STXS 2



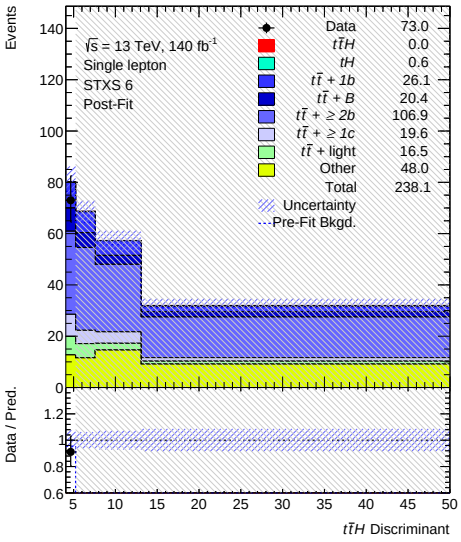
(c): STXS 3



(d): STXS 4

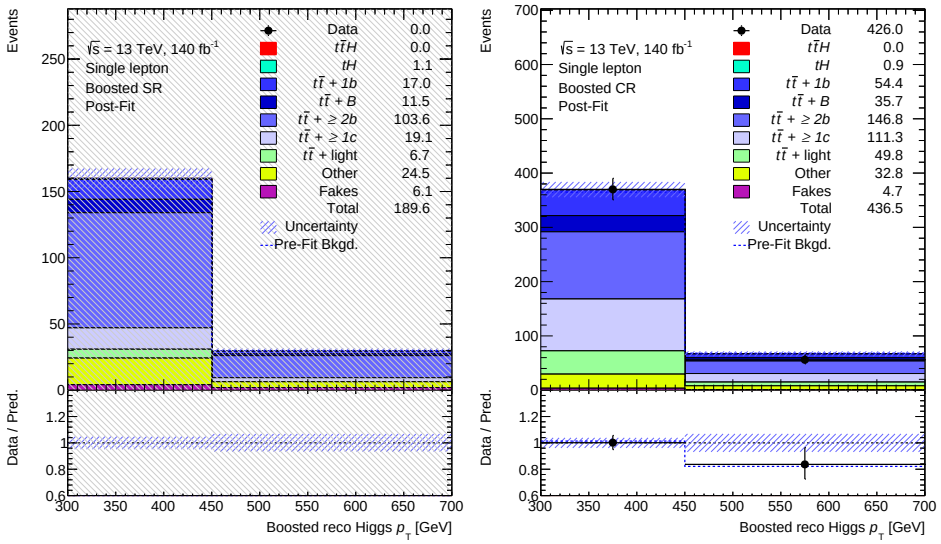


(e): STXS 5



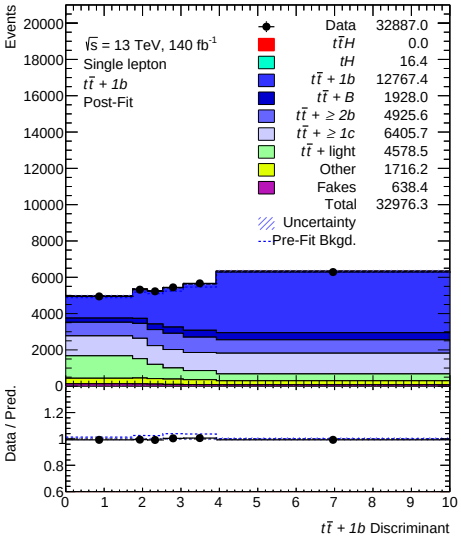
(f): STXS 6

Figure A.34.: Background-only post-fit single-lepton resolved SRs.

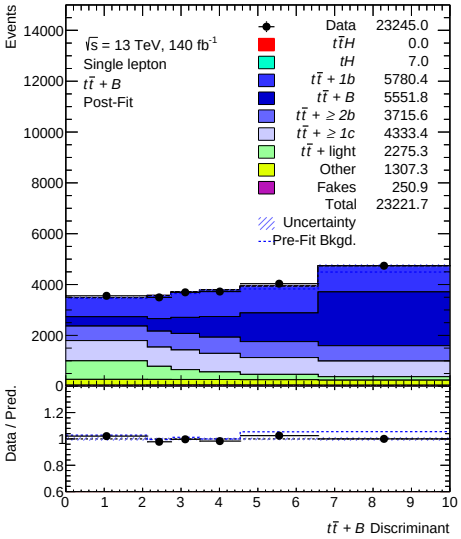


(a): Boosted SR (b): Boosted CR

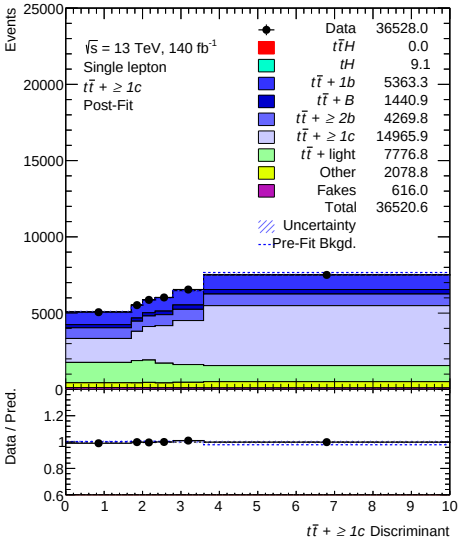
Figure A.35.: Background-only post-fit single-lepton boosted SR and CR.



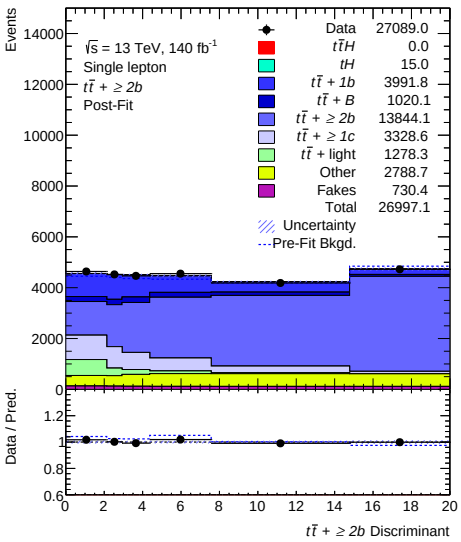
(a): $t\bar{t} + b$ region



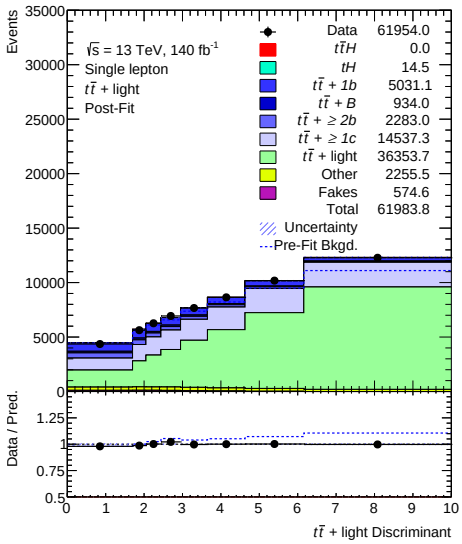
(b): $t\bar{t} + B$ region



(c): $t\bar{t} + \geq 1c$ region

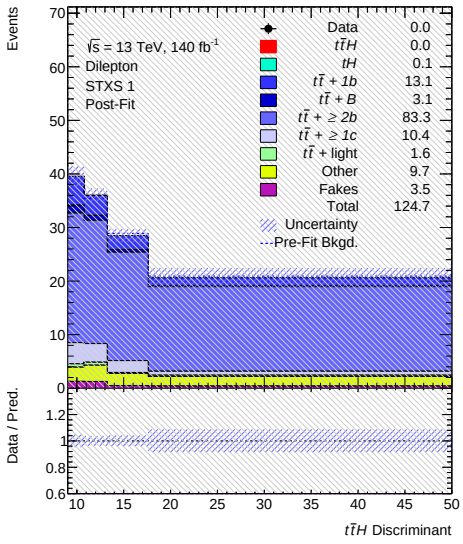


(d): $t\bar{t} + \geq 2b$ region

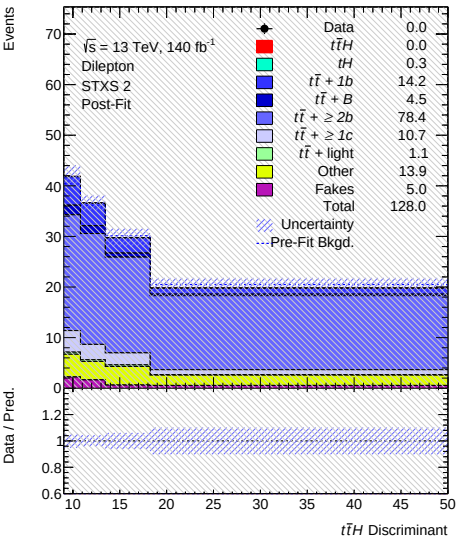


(e): $t\bar{t} +$ light region

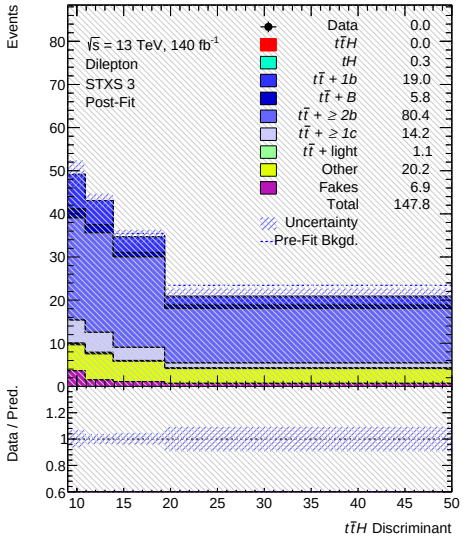
Figure A.36.: Background-only post-fit single-lepton resolved CRs.



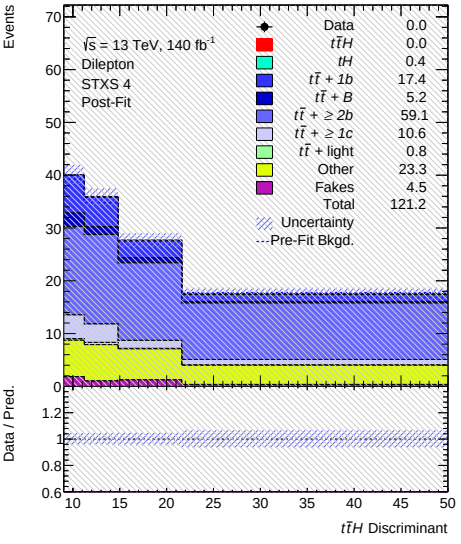
(a): STXS 1



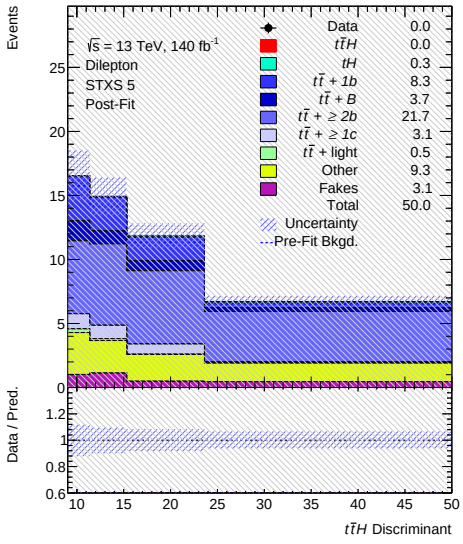
(b): STXS 2



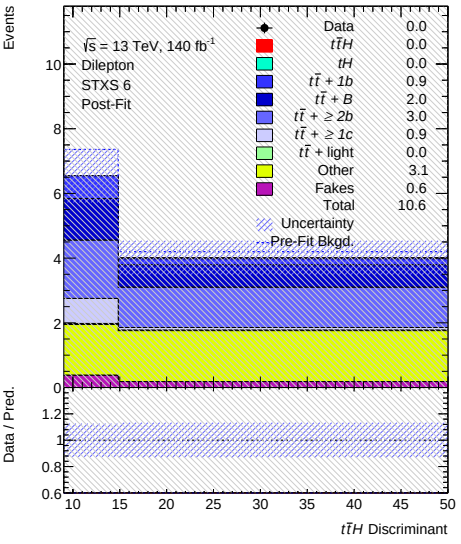
(c): STXS 3



(d): STXS 4

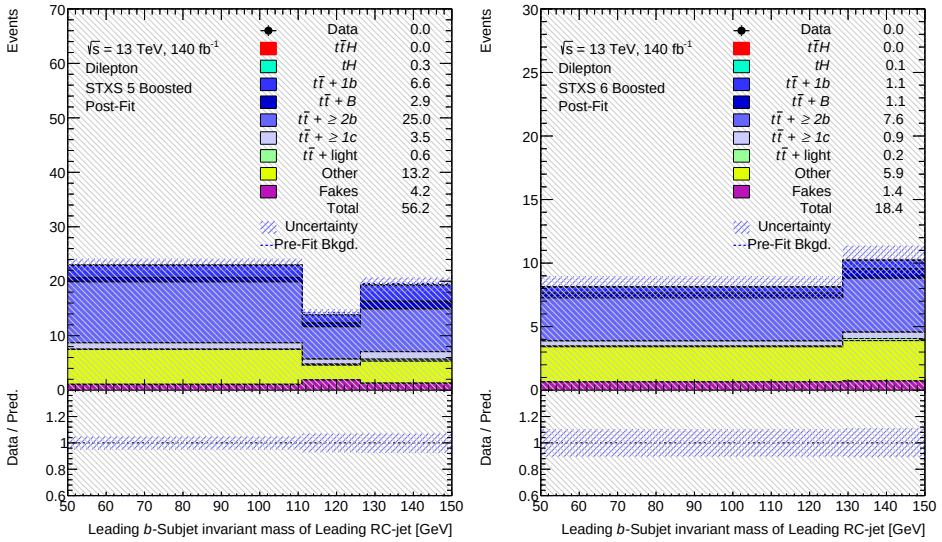


(e): STXS 5



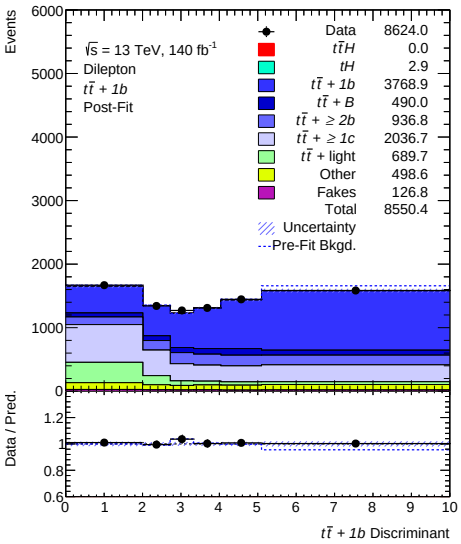
(f): STXS 6

Figure A.37.: Background-only post-fit dilepton resolved SRs.

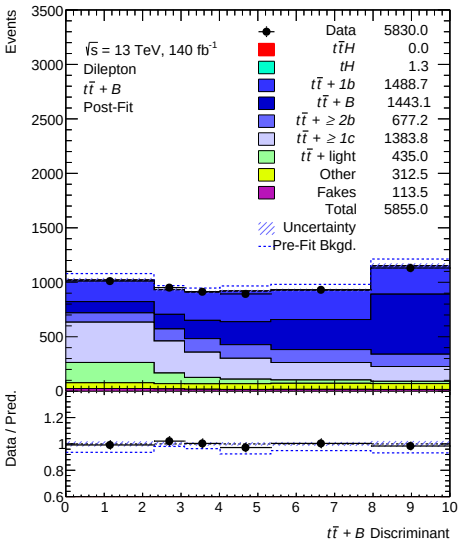


(a): Boosted STXS 5 (b): Boosted STXS 6

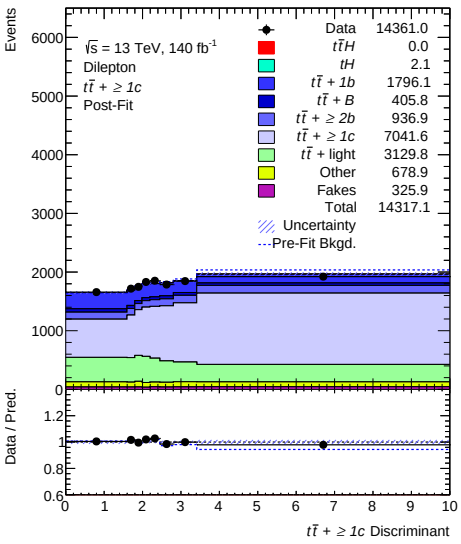
Figure A.38.: Background-only post-fit dilepton boosted SRs.



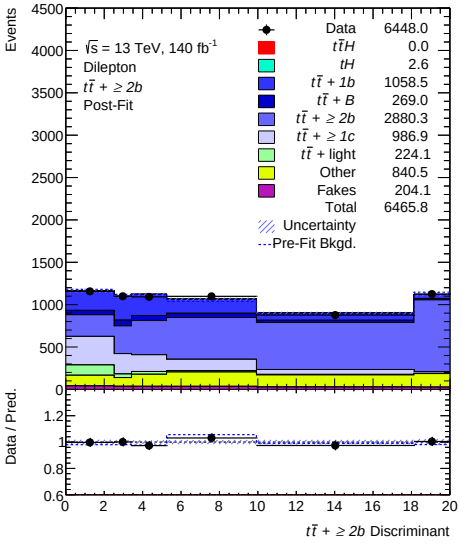
(a): $t\bar{t} + b$ region



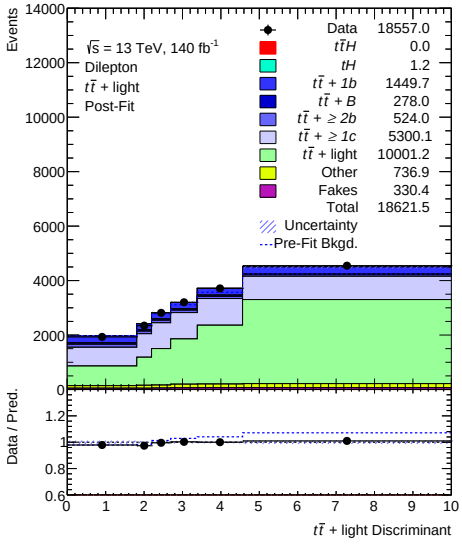
(b): $t\bar{t} + B$ region



(c): $t\bar{t} + \geq 1c$ region



(d): $t\bar{t} + \geq 2b$ region



(e): $t\bar{t} +$ light region

Figure A.39.: Background-only post-fit dilepton resolved CRs.

B

ACRONYMS

CERN Conseil Européen pour la Recherche Nucléaire

SM Standard Model

EW Electroweak

BSM Beyond Standard Model

QFT Quantum Field Theory

QED Quantum Electrodynamics

QCD Quantum Chromodynamics

CKM Cabibbo-Kobayashi-Maskawa

VEV Vacuum Expectation Value

LHC Large Hadron Collider

LINAC Linear Accelerator

PSB Proton Synchrotron Booster

PS Proton Synchrotron

SPS Super Proton Synchrotron

ID Inner Detector

PD Pixel Detector

IP Impact Parameter

SCT Semiconductor Tracker

TRT Transition Radiation Tracker

IBL Insertable b-Layer

ECal Electromagnetic Calorimeter

HCal Hadronic Calorimeter

FCal Forward Calorimeter

LAr Liquid Argon

EMEC Liquid Argon Electromagnetic End-Cap

TileCal Tile Calorimeter

HEC Liquid Argon Hadronic End-Cap

MS Muon Spectrometer

MDT Monitored Drift Tube

CSC Cathod Strip Chamber

RPC Resistive Plate Chamber

TGC Thin Gap Chamber

L1 Level 1

FPGA Field Programmable Gate Array

HLT High-Level Trigger

RoI Region of Interest

DAQ Data Acquisition

MC Monte Carlo

DGLAP Dokshitzer–Gribov–Lipatov–Altarelli–Parisi

LO Leading Order

NLO Next-to-Leading Order

NNLO Next-to-Next-to-Leading Order

NNLL Next-to-Next-to-Leading Log

ISR Initial State Radiation

FSR Final State Radiation

4FS Four Flavour Scheme

5FS Five Flavour Scheme

ME Matrix Element

PS Parton Shower

PHD Parton-Hadron Duality

AF2 Atlfast-II

$E_{\mathrm{T}}^{\mathrm{miss}}$ Missing Transverse Momentum

PV Primary Vertex

CCA Connected Component Analysis

WP Working Point

LH Likelihood

LLH Log-Likelihood

PDF Probability Density Function

Topo-Cluster Topological Cluster

IRC Infrared and Collinear

JES Jet Energy Scale

JER Jet Energy Resolution

MCJES Monte Carlo Jet Energy Scale

GSC Global Sequential Calibration

HEP High Energy Physics

HL-LHC High-Luminosity LHC

ML Machine Learning

NN Neural Network

DNN Deep Neural Network

MLP Multilayer Perceptron

LSTM Long Short-Term Memory

ReLU Rectified Linear Unit

LR Learning Rate

RNN Recurrent Neural Network

GPU Graphics Processing Unit

BPTT Backpropagation Through Time

GRU Gated Recurrent Unit

MHA Multi-Head Attention

$\overline{\text{MS}}$ Modified Minimal Subtraction

SUSY Supersymmetry

RNNIP Recurrent Neural Network Impact Parameter

DL1 Deep Learning 1

DL1r Deep Learning 1 RNNIP

DIPS Deep Impact Parameter Sets

DL1d Deep Learning 1 DIPS

JVT Jet Vertex Tagger

SV1 Secondary-Vertex-Tagging

LWTNN Lightweight Trained Neural Network

ROC Receiver Operating Characteristic

PCBT Pseudo-Continuous b-Tagging

BDT Boosted Decision Tree

SF Scale Factor

OP Operation Point

PLF Profile-Likelihood Fit

STXS Simplified Template Cross-Section

DS Diagram Subtraction

DR Diagram Removal

SR Signal Region

CR Control Region

NP Nuisance Parameter

LHCHXS LHC Higgs Cross-Section

POI Parameter of Interest

ACKNOWLEDGEMENTS

During my time preparing, working, and finally writing this PhD thesis, several people supported and accompanied me, which I would like to thank here.
First I want to thank Priv.-Doz Dr. Andrea Knue and Prof. Dr. Gregor Herten for the opportunity to work and finalise my PhD under their supervision. Thanks for the wonderful and always supportive guiding through this crazy time. I want to specially thank Andrea here again for the constant and sometimes non-physics related conversations we had, which were always cheerful and sometimes decorated by well cooled drinks.
Further, I want to thank Prof. Dr. Karl Jakobs for the excellent supervision and constant support I received during my time in Freiburg. I really enjoyed our biannual interviews, where I always got very helpful feedback.
In addition to my supervisors, I want to thank in advance the examiners of my defence for accepting this important role, Prof. Dr. Markus Schuhmacher and Prof. Dr. Stefan Dittmaier.
I also want to thank the whole Freiburg group and the other PhD students from the GRK, especially Ksenia, Patrick, José, Prasham, Sascha, Dimbi and Vlad. One special thanks here goes to one former Freiburg student, who supervised me the full first year in ATLAS in the FTAG group, Manuel Guth. Thanks for taking me under your wing and teaching me all the needed skills to survive in the group and in ATLAS. I will never forget the fun evening we had at your place in Geneva.
Furthermore, a very big thanks goes to Konstantin (Connie) Heidegger for proof-reading my whole thesis. I owe you a lot and I will never forget an Oxford comma again. In addition to Connie, I want to thank all the other people who proof-read my thesis, including Johnny, Phillip, Mathieu, Jonas, and Felix.
A special thanks goes to the ATLAS FTAG group. It was a great joy to work and also discuss some topics over one or more beers with you. Again here I need to explicitly mention a few people who I owe a big thanks: Valerio, Bing, Francesco, Dan, and again Manuel and Philipp.

Nach meinem Master in Dortmund bin ich ohne wirklich Leute zu kennen nach Freiburg gekommen, kurz bevor der erste Lockdown verhängt wurde. In dieser Zeit habe ich es trotzdem irgendwie es geschafft, neue Leute kennenzulernen und gleichzeitig den Kontakt zu meinen Freunden in Dortmund zu halten. An dieser Stelle möchte ich mich bei all diesen Leuten bedanken, die mir auch außerhalb meiner Arbeit geholfen haben. Darunter Daniel und Fabian all die anderen, die ich hier jetzt nicht aufzähle.
Von den Dortmundern geht ein großer, ganz spezieller Dank an Felix und Jannine, die mich stets aufgemuntern und mich gelegentlich auch mal von nicht so tollen Ergebnissen abgelenkt haben.
Der letzte, ganz besondere Dank geht an meine Familie, Katrin, Uwe und Kristin für ihre dauerhafte Unterstützung.

REFERENCES

[1] J. Birk et al. *umami-hep/puma:* 2023.
URL: https://doi.org/10.5281/zenodo.6607414 (cit. on pp. v, 80).

[2] ATLAS Collaboration. *Performance Plots of the new trained Deep Impact Parameter Sets (DIPS) Heavy Flavour Tagger*.
URL: http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2021-004/ (visited on 13/11/2023) (cit. on pp. v, 83, 99).

[3] ATLAS Collaboration.
Graph Neural Network Jet Flavour Tagging with the ATLAS Detector. Tech. rep.
All figures including auxiliary figures are available at
https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2022-027. Geneva: CERN, 2022. URL: https://cds.cern.ch/record/2811135
(cit. on p. v).

[4] ATLAS Collaboration. *Jet Flavour Tagging With GN1 and DL1d. Generator dependence, Run 2 and Run 3 data agreement studies*.
URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2023-01/ (visited on 02/02/2024) (cit. on p. v).

[5] Peter W. Higgs. 'Broken symmetries, massless particles and gauge fields'.
In: *Phys. Lett.* 12 (1964), pp. 132–133 (cit. on pp. 1, 5, 12).

[6] Peter W. Higgs. 'Broken symmetries and the masses of gauge bosons'.
In: *Phys. Rev. Lett.* 13 (1964), pp. 508–509 (cit. on pp. 1, 5, 12).

[7] F. Englert and R. Brout. 'Broken Symmetry and the Mass of Gauge Vector Mesons'.
In: *Phys. Rev. Lett.* 13 (9 1964), pp. 321–323 (cit. on pp. 1, 5, 12).

[8] Peter W. Higgs. 'Spontaneous symmetry breakdown without massless bosons'.
In: *Phys. Rev.* 145 (1966), pp. 1156–1163 (cit. on pp. 1, 5, 12).

[9] ATLAS Collaboration. 'Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC'. In: *Phys. Lett. B* 716 (2012), pp. 1–29. arXiv: 1207.7214 [hep-ex] (cit. on pp. 1, 5).

[10] CMS Collaboration.
'Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC'.
In: *Phys. Lett. B* 716 (2012), pp. 30–61. arXiv: 1207.7235 [hep-ex] (cit. on pp. 1, 5).

[11] Lyndon Evans and Philip Bryant. 'LHC Machine'.
In: *Journal of Instrumentation* 3.08 (2008), S08001–S08001 (cit. on pp. 1, 5, 17).

[12] Steven Weinberg. 'A Model of Leptons'. In: *Phys. Rev. Lett.* 19 (21 1967), pp. 1264–1266. URL: https://link.aps.org/doi/10.1103/PhysRevLett.19.1264 (cit. on pp. 1, 11, 15).

[13] A. Arbey, F. Mahmoudi, O. Stal and T. Stefaniak. 'Status of the Charged Higgs Boson in Two Higgs Doublet Models'. In: *Eur. Phys. J. C* 78.3 (2018), p. 182. arXiv: 1706.07414 [hep-ph] (cit. on p. 1).

[14] ATLAS Collaboration. 'Measurement of Higgs boson decay into b-quarks in associated production with a top-quark pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector'. In: *JHEP* 06 (2022), p. 097. arXiv: 2111.06712 [hep-ex] (cit. on pp. 1, 84, 147–149, 151, 167).

[15] LHCHiggsCrossSectionWorkingGroup. *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*. CERN Yellow Reports: Monographs. 869 pages, 295 figures, 248 tables and 1645 citations. Working Group web page: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2227475 (cit. on pp. 2, 77, 147, 150, 158, 166, 168).

[16] Nicolas Berger et al. 'Simplified Template Cross Sections - Stage 1.1'. In: (June 2019). arXiv: 1906.02754 [hep-ph] (cit. on pp. 2, 150).

[17] Mary K. Gaillard, Paul D. Grannis and Frank J. Sciulli. 'The Standard model of particle physics'. In: *Rev. Mod. Phys.* 71 (1999), S96–S111. arXiv: hep-ph/9812285 (cit. on p. 5).

[18] M. Bustamante, L. Cieri and John Ellis. 'Beyond the Standard Model for Montaneros'. In: *5th CERN - Latin American School of High-Energy Physics*. Nov. 2009. arXiv: 0911.4409 [hep-ph] (cit. on pp. 5, 13).

[19] Carsten Burgard. *Example: Standard model of physics*. 1999. URL: https://texample.net/tikz/examples/model-physics/ (visited on 12/10/2023) (cit. on p. 6).

[20] Murray Gell-Mann. 'A Schematic Model of Baryons and Mesons'. In: *Phys. Lett.* 8 (1964), pp. 214–215 (cit. on p. 6).

[21] G. Zweig. 'An SU(3) model for strong interaction symmetry and its breaking. Version 2'. In: *DEVELOPMENTS IN THE QUARK THEORY OF HADRONS. VOL. 1. 1964 - 1978*. Ed. by D. B. Lichtenberg and Simon Peter Rosen. Feb. 1964, pp. 22–101 (cit. on p. 6).

[22] David J Griffiths. *Introduction to elementary particles; 2nd rev. version*. Physics textbook. New York, NY: Wiley, 2008. URL: https://cds.cern.ch/record/111880 (cit. on p. 7).

[23] G. Arnison et al. ‘Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV’.
In: *Phys. Lett. B* 122 (1983), pp. 103–116 (cit. on pp. 7, 18).

[24] M. Banner et al. ‘Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider’.
In: *Phys. Lett. B* 122 (1983), pp. 476–485 (cit. on pp. 7, 18).

[25] G. Arnison et al. ‘Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c**2 at the CERN SPS Collider’. In: *Phys. Lett. B* 126 (1983), pp. 398–410 (cit. on pp. 7, 18).

[26] P. Bagnaia et al. ‘Evidence for $Z^0 \to e^+e^-$ at the CERN $\bar{p}p$ Collider’.
In: *Phys. Lett. B* 129 (1983), pp. 130–140 (cit. on pp. 7, 18).

[27] R. L. Workman and Others. ‘Review of Particle Physics’. In: *PTEP* 2022 (2022), p. 083C01. ISSN: 2050-3911. eprint: https://academic.oup.com/ptep/article-pdf/2022/8/083C01/49175539/ptac097.pdf.
URL: https://doi.org/10.1093/ptep/ptac097 (cit. on pp. 7, 10, 13, 78).

[28] R. P. Feynman. ‘Space-Time Approach to Quantum Electrodynamics’.
In: *Phys. Rev.* 76 (6 1949), pp. 769–789.
URL: https://link.aps.org/doi/10.1103/PhysRev.76.769 (cit. on p. 8).

[29] S. Tomonaga.
‘On a Relativistically Invariant Formulation of the Quantum Theory of Wave Fields*’.
In: *Progress of Theoretical Physics* 1.2 (Aug. 1946), pp. 27–42. ISSN: 0033-068X.
eprint: https://academic.oup.com/ptp/article-pdf/1/2/27/24027031/1-2-27.pdf.
URL: https://doi.org/10.1143/PTP.1.27 (cit. on p. 8).

[30] Julian Schwinger. ‘CXXIV. A note on the quantum dynamical principle’.
In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 44.357 (1953), pp. 1171–1179.
eprint: https://doi.org/10.1080/14786441008520377.
URL: https://doi.org/10.1080/14786441008520377 (cit. on p. 8).

[31] A. Einstein. ‘Die Grundlage der allgemeinen Relativitätstheorie’.
In: *Annalen der Physik* 354.7 (1916), pp. 769–822. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.19163540702.
URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.19163540702 (cit. on p. 8).

[32] E. Noether. ‘Invariante Variationsprobleme’. ger.
In: *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* 1918 (1918), pp. 235–257.
URL: http://eudml.org/doc/59024 (cit. on p. 8).

[33] V. V. Ezhela, S. B. Lugovsky and O. V. Zenin. ‘Hadronic part of the muon g-2 estimated on the sigma**2003(tot)(e+ e- —> hadrons) evaluated data compilation’. In: (Dec. 2003).
arXiv: hep-ph/0312114 (cit. on p. 9).

[34] Murray Gell-Mann. ‘Symmetries of Baryons and Mesons’.
In: *Phys. Rev.* 125 (3 1962), pp. 1067–1084 (cit. on p. 9).

[35] David J. Gross and Frank Wilczek. ‘Ultraviolet Behavior of Non-Abelian Gauge Theories’.
In: *Phys. Rev. Lett.* 30 (26 1973), pp. 1343–1346.
URL: https://link.aps.org/doi/10.1103/PhysRevLett.30.1343
(cit. on p. 10).

[36] H. David Politzer. ‘Reliable Perturbative Results for Strong Interactions?’
In: *Phys. Rev. Lett.* 30 (26 1973), pp. 1346–1349.
URL: https://link.aps.org/doi/10.1103/PhysRevLett.30.1346
(cit. on p. 10).

[37] Kenneth G. Wilson. ‘Confinement of quarks’. In: *Phys. Rev. D* 10 (8 1974), pp. 2445–2459.
URL: https://link.aps.org/doi/10.1103/PhysRevD.10.2445
(cit. on pp. 10, 34).

[38] LHCb Collaboration. ‘The LHCb Detector at the LHC’.
In: *Journal of Instrumentation* 3.08 (2008), S08005 (cit. on pp. 10, 18).

[39] LHCb Collaboration. ‘Observation of the resonant character of the $Z(4430)^-$ state’.
In: *Phys. Rev. Lett.* 112.22 (2014), p. 222002. arXiv: 1404.1903 [hep-ex]
(cit. on p. 10).

[40] LHCb Collaboration. ‘Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays’. In: *Phys. Rev. Lett.* 115 (2015), p. 072001.
arXiv: 1507.03414 [hep-ex] (cit. on p. 10).

[41] E. Fermi. ‘An attempt of a theory of beta radiation. 1.’ In: *Z. Phys.* 88 (1934), pp. 161–177
(cit. on p. 11).

[42] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes and R. P. Hudson.
‘Experimental Test of Parity Conservation in Beta Decay’.
In: *Phys. Rev.* 105 (4 1957), pp. 1413–1415.
URL: https://link.aps.org/doi/10.1103/PhysRev.105.1413
(cit. on p. 11).

[43] Richard L. Garwin, Leon M. Lederman and Marcel Weinrich.
'Observations of the Failure of Conservation of Parity and Charge Conjugation in Meson Decays: the Magnetic Moment of the Free Muon'.
In: *Phys. Rev.* 105 (4 1957), pp. 1415–1417.
URL: https://link.aps.org/doi/10.1103/PhysRev.105.1415
(cit. on p. 11).

[44] T. D. Lee and C. N. Yang. 'Question of Parity Conservation in Weak Interactions'.
In: *Phys. Rev.* 104 (1 1956), pp. 254–258.
URL: https://link.aps.org/doi/10.1103/PhysRev.104.254
(cit. on p. 11).

[45] R. P. Feynman and M. Gell-Mann. 'Theory of the Fermi Interaction'.
In: *Phys. Rev.* 109 (1 1958), pp. 193–198.
URL: https://link.aps.org/doi/10.1103/PhysRev.109.193
(cit. on p. 11).

[46] J. H. Christenson, J. W. Cronin, V. L. Fitch and R. Turlay.
'Evidence for the 2π Decay of the K_2^0 Meson'. In: *Phys. Rev. Lett.* 13 (4 1964), pp. 138–140.
URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.138
(cit. on p. 11).

[47] Nicola Cabibbo. 'Unitary Symmetry and Leptonic Decays'.
In: *Phys. Rev. Lett.* 10 (12 1963), pp. 531–533.
URL: https://link.aps.org/doi/10.1103/PhysRevLett.10.531
(cit. on p. 11).

[48] Makoto Kobayashi and Toshihide Maskawa.
'CP-Violation in the Renormalizable Theory of Weak Interaction'.
In: *Progress of Theoretical Physics* 49.2 (Feb. 1973), pp. 652–657. ISSN: 0033-068X.
eprint: https://academic.oup.com/ptp/article-pdf/49/2/652/5257692/49-2-652.pdf.
URL: https://doi.org/10.1143/PTP.49.652 (cit. on p. 11).

[49] Sheldon L. Glashow. 'Partial-symmetries of weak interactions'.
In: *Nuclear Physics* 22.4 (1961), pp. 579–588. ISSN: 0029-5582. URL: https://www.sciencedirect.com/science/article/pii/0029558261904692
(cit. on p. 11).

[50] Abdus Salam. 'Weak and Electromagnetic Interactions'.
In: *Conf. Proc. C* 680519 (1968), pp. 367–377 (cit. on p. 11).

[51] M. Gell-Mann. 'The interpretation of the new particles as displaced charge multiplets'.
In: *Nuovo Cim.* 4.S2 (1956), pp. 848–866 (cit. on p. 11).

[52] Kazuhiko Nishijima. 'Charge Independence Theory of V Particles'.
In: *Prog. Theor. Phys.* 13.3 (1955), pp. 285–304 (cit. on p. 11).

[53] Ziro Maki, Masami Nakagawa and Shoichi Sakata.
'Remarks on the unified model of elementary particles'.
In: *Prog. Theor. Phys.* 28 (1962), pp. 870–880 (cit. on p. 11).

[54] M. H. Ahn et al. 'Measurement of Neutrino Oscillation by the K2K Experiment'.
In: *Phys. Rev. D* 74 (2006), p. 072003. arXiv: hep-ex/0606032 (cit. on p. 11).

[55] B. Pontecorvo.
'Neutrino Experiments and the Problem of Conservation of Leptonic Charge'.
In: *Zh. Eksp. Teor. Fiz.* 53 (1967), pp. 1717–1725 (cit. on p. 11).

[56] John Ellis. 'Higgs Physics'. In: *2013 European School of High-Energy Physics*. 2015, pp. 117–168. arXiv: 1312.5672 [hep-ph] (cit. on p. 13).

[57] Yoichiro Nambu.
'Quasi-Particles and Gauge Invariance in the Theory of Superconductivity'.
In: *Phys. Rev.* 117 (3 1960), pp. 648–663.
URL: https://link.aps.org/doi/10.1103/PhysRev.117.648
(cit. on p. 14).

[58] J. Goldstone. 'Field Theories with Superconductor Solutions'.
In: *Nuovo Cim.* 19 (1961), pp. 154–164 (cit. on p. 14).

[59] Jeffrey Goldstone, Abdus Salam and Steven Weinberg. 'Broken Symmetries'.
In: *Phys. Rev.* 127 (3 1962), pp. 965–970.
URL: https://link.aps.org/doi/10.1103/PhysRev.127.965
(cit. on p. 14).

[60] Y. Fukuda et al. 'Evidence for Oscillation of Atmospheric Neutrinos'.
In: *Phys. Rev. Lett.* 81 (8 1998), pp. 1562–1567.
URL: https://link.aps.org/doi/10.1103/PhysRevLett.81.1562
(cit. on p. 15).

[61] M. Aker et al. 'Direct neutrino-mass measurement with sub-electronvolt sensitivity'.
In: *Nature Phys.* 18.2 (2022), pp. 160–166. arXiv: 2105.08533 [hep-ex]
(cit. on p. 15).

[62] Planck Collaboration et al.
'Planck 2018 results - VI. Cosmological parameters (Corrigendum)'.
In: *Astronomy and Astrophysics* 652 (2021), p. C4.
URL: https://doi.org/10.1051/0004-6361/201833910e (cit. on p. 15).

[63] A. D. Sakharov.
'Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe'.
In: *Pisma Zh. Eksp. Teor. Fiz.* 5 (1967), pp. 32–35 (cit. on p. 15).

[64] N. S. Manton. 'Topology in the Weinberg-Salam Theory'.
In: *Phys. Rev. D* 28 (1983), p. 2019 (cit. on p. 16).

[65] Frans R. Klinkhamer and N. S. Manton.
'A Saddle Point Solution in the Weinberg-Salam Theory'.
In: *Phys. Rev. D* 30 (1984), p. 2212 (cit. on p. 16).

[66] Werner Bernreuther. 'CP violation and baryogenesis'.
In: *Lect. Notes Phys.* 591 (2002), pp. 237–293. arXiv: hep-ph/0205279 (cit. on p. 16).

[67] G. Gamow. 'Expanding Universe and the Origin of Elements'.
In: *Phys. Rev.* 70 (7-8 1946), pp. 572–573.
URL: https://link.aps.org/doi/10.1103/PhysRev.70.572.2
(cit. on p. 16).

[68] James D. Wells. 'Higgs naturalness and the scalar boson proliferation instability problem'.
In: *Synthese* 194.2 (2017), pp. 477–490. arXiv: 1603.06131 [hep-ph] (cit. on p. 16).

[69] ATLAS Collaboration.
'Combined measurement of the Higgs boson mass from the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ell$ decay channels with the ATLAS detector using $\sqrt{s}$ = 7, 8 and 13 TeV pp collision data'.
In: *Phys. Rev. Lett.* 131 (2023), p. 251802. arXiv: 2308.04775 [hep-ex] (cit. on p. 16).

[70] Leonard Susskind.
'Dynamics of Spontaneous Symmetry Breaking in the Weinberg-Salam Theory'.
In: *Phys. Rev. D* 20 (1979), pp. 2619–2625 (cit. on p. 16).

[71] V. C. Rubin, N. Thonnard and W. K. Ford Jr.
'Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 /R = 4kpc/ to UGC 2885 /R = 122 kpc/'. In: *Astrophys. J.* 238 (1980), p. 471 (cit. on p. 16).

[72] Jose M. Diego et al.
'A free-form lensing grid solution for A1689 with new multiple images'.
In: *Mon. Not. Roy. Astron. Soc.* 446 (2015), pp. 683–704.
arXiv: 1402.4170 [astro-ph.CO] (cit. on p. 16).

[73] Adam G. Riess et al. 'Observational evidence from supernovae for an accelerating universe and a cosmological constant'. In: *Astron. J.* 116 (1998), pp. 1009–1038.
arXiv: astro-ph/9805201 (cit. on p. 16).

[74] S. Perlmutter et al. 'Measurements of Ω and Λ from 42 high redshift supernovae'.
In: *Astrophys. J.* 517 (1999), pp. 565–586. arXiv: astro-ph/9812133 (cit. on p. 16).

[75] Lance J. Dixon. 'Ultraviolet Behavior of $\mathcal{N} = 8$ Supergravity'.
In: *Subnucl. Ser.* 47 (2011). Ed. by Antonino Zichichi, pp. 1–39.
arXiv: 1005.2703 [hep-th] (cit. on p. 16).

[76] Esma Mobs. 'The CERN accelerator complex - August 2018. Complexe des accélérateurs du CERN - Août 2018'. In: (2018). General Photo.
URL: http://cds.cern.ch/record/2636343 (cit. on p. 18).

[77] ATLAS Collaboration. 'The ATLAS Experiment at the CERN Large Hadron Collider'. In: *Journal of Instrumentation* 3.08 (2008), S08003 (cit. on pp. 18, 20, 24–28).

[78] CMS Collaboration. 'The CMS Collaboration experiment at the CERN LHC'. In: *Journal of Instrumentation* 3.08 (2008), S08004 (cit. on p. 18).

[79] ALICE Collaboration. 'The ALICE experiment at the CERN LHC'. In: *Journal of Instrumentation* 3.08 (2008), S08002 (cit. on p. 18).

[80] ATLAS Collaboration. *Luminosity Public Results for Run-2*. URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2 (visited on 07/08/2023) (cit. on pp. 19, 20).

[81] ATLAS Collaboration. 'Computer generated image of the whole ATLAS detector'. 2008. URL: https://cds.cern.ch/record/1095924 (cit. on p. 21).

[82] ATLAS Collaboration. *ATLAS inner detector: Technical Design Report, 1*. Technical Design Report ATLAS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/331063 (cit. on p. 21).

[83] S Haywood, L Rossi, R Nickerson and A Romaniouk. *ATLAS inner detector: Technical Design Report, 2*. Technical Design Report ATLAS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/331064 (cit. on p. 21).

[84] Karolos Potamianos. 'The upgraded Pixel detector and the commissioning of the Inner Detector tracking of the ATLAS experiment for Run-2 at the Large Hadron Collider'. In: *PoS* EPS-HEP2015 (2015), p. 261. arXiv: 1608.07850 [physics.ins-det] (cit. on p. 22).

[85] ATLAS Collaboration. 'Computer generated image of the ATLAS inner detector'. 2008. URL: https://cds.cern.ch/record/1095926 (cit. on p. 22).

[86] ATLAS Collaboration. *Technical Design Report for the ATLAS Inner Tracker Pixel Detector*. Tech. rep. CERN-LHCC-2017-021. ATLAS-TDR-030. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2285585 (cit. on p. 22).

[87] ATLAS Collaboration. 'Computer generated images of the Pixel, part of the ATLAS inner detector.' 2008. URL: https://cds.cern.ch/record/1095925 (cit. on p. 23).

[88] ATLAS Collaboration. *ATLAS Insertable B-Layer Technical Design Report*. Tech. rep. CERN-LHCC-2010-013. ATLAS-TDR-19. 2010. URL: https://cds.cern.ch/record/1291633 (cit. on p. 22).

[89] ATLAS Collaboration. *Impact Parameter Resolution*. URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/IDTR-2015-007/ (visited on 16/11/2023) (cit. on p. 23).

[90] ATLAS Collaboration.
‘Operation and performance of the ATLAS semiconductor tracker in LHC Run 2’.
In: *JINST* 17.01 (2022), P01013. arXiv: 2109.02591 [physics.ins-det]
(cit. on p. 24).

[91] ATLAS Collaboration. *ATLAS Transition Radiation Tracker (TRT): Straw tubes for tracking and particle identification at the Large Hadron Collider*. Tech. rep. Geneva: CERN, 2017.
URL: https://cds.cern.ch/record/2139567 (cit. on p. 24).

[92] ATLAS Collaboration. ‘The ATLAS Transition Radiation Tracker (TRT) proportional drift tube: Design and performance’. In: *JINST* 3 (2008), P02013 (cit. on p. 24).

[93] ATLAS Collaboration. ‘Computer Generated image of the ATLAS calorimeter’. 2008.
URL: https://cds.cern.ch/record/1095927 (cit. on p. 25).

[94] ATLAS Collaboration. ‘Computer generated image of the ATLAS Liquid Argon’. 2008.
URL: https://cds.cern.ch/record/1095928 (cit. on p. 26).

[95] ATLAS Collaboration. ‘The tile hadronic calorimeter for the ATLAS experiment’.
In: *IEEE Trans. Nucl. Sci.* 48 (2001), pp. 367–371 (cit. on p. 26).

[96] ATLAS Collaboration. ‘The ATLAS forward calorimeters’. In: *JINST* 3 (2008), P02010
(cit. on p. 27).

[97] ATLAS Collaboration. ‘The ATLAS Forward Calorimeter’. In: (2011).
URL: https://cds.cern.ch/record/1386649 (cit. on p. 27).

[98] ATLAS Collaboration. *ATLAS magnet system: Technical Design Report, 1*.
Technical design report. ATLAS. Geneva: CERN, 1997.
URL: https://cds.cern.ch/record/338080 (cit. on p. 27).

[99] Steffen Kaiser. ‘Search for the Higgs Boson in the Process $pp \to Hqq$, $H \to WW$ with the ATLAS Detector’. Presented on 22 Feb 2010. 2010.
URL: https://cds.cern.ch/record/1254897 (cit. on p. 28).

[100] ATLAS Collaboration. *ATLAS muon spectrometer: Technical Design Report*.
Technical Design Report ATLAS. Geneva: CERN, 1997.
URL: https://cds.cern.ch/record/331068 (cit. on pp. 28, 29).

[101] ATLAS Collaboration. ‘Operation of the ATLAS trigger system in Run 2’.
In: *JINST* 15.10 (2020), P10004. arXiv: 2007.12539 [physics.ins-det]
(cit. on pp. 29, 30).

[102] John C. Collins, Davison E. Soper and George F. Sterman.
‘Factorization of Hard Processes in QCD’.
In: *Adv. Ser. Direct. High Energy Phys.* 5 (1989), pp. 1–91. arXiv: hep-ph/0409313
(cit. on p. 31).

[103] Sayipjamal Dulat et al.
'New parton distribution functions from a global analysis of quantum chromodynamics'.
In: *Phys. Rev. D* 93 (3 2016), p. 033006 (cit. on p. 31).

[104] M.R. Whalley, D. Bourilkov and R.C. Group.
'The Les Houches accord PDFs (LHAPDF) and LHAGLUE'. In: *HERA and the LHC: A Workshop on the Implications of HERA and LHC Physics (Startup Meeting, CERN, 26-27 March 2004; Midterm Meeting, CERN, 11-13 October 2004)*. Aug. 2005, pp. 575–581.
arXiv: hep-ph/0508110 (cit. on p. 31).

[105] Richard D. Ball et al. 'Parton distributions for the LHC Run II'. In: *JHEP* 04 (2015), p. 040.
arXiv: 1410.8849 [hep-ph] (cit. on pp. 31, 32, 156).

[106] V.N. Gribov and L.N. Lipatov. 'Deep inelastic e p scattering in perturbation theory'.
In: *Sov. J. Nucl. Phys.* 15 (1972), pp. 438–450 (cit. on p. 31).

[107] G. Altarelli and G. Parisi. 'Asymptotic freedom in parton language'.
In: *Nuclear Physics B* 126.2 (1977), pp. 298 –318. ISSN: 0550-3213 (cit. on p. 31).

[108] Yuri L. Dokshitzer. 'Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.'
In: *Sov. Phys. JETP* 46 (1977), pp. 641–653 (cit. on p. 31).

[109] M. Tanabashi et al. 'Review of Particle Physics'. In: *Phys. Rev. D* 98 (3 2018), p. 030001.
URL: https://doi.org/10.1103/PhysRevD.98.030001 (cit. on p. 32).

[110] ATLAS Collaboration. *Standard Model Summary Plots February 2022*. Tech. rep.
All figures including auxiliary figures are available at
https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2022-009. Geneva: CERN, 2022. URL: https://cds.cern.ch/record/2804061
(cit. on p. 33).

[111] Steffen Schumann and Frank Krauss.
'A parton shower algorithm based on Catani–Seymour dipole factorisation'.
In: *JHEP* 03 (2008), p. 038. arXiv: 0709.1027 [hep-ph] (cit. on p. 34).

[112] Enrico Bothmann et al. 'Event Generation with Sherpa 2.2'.
In: *SciPost Phys.* 7.3 (2019), p. 034. arXiv: 1905.09127 [hep-ph] (cit. on p. 34).

[113] Peter Skands, Nadine Fischer, Stefan Prestel and Mathias Ritzmann.
'The VINCIA Antenna Shower for Hadron Colliders'. In: Sept. 2016.
arXiv: 1609.07205 [hep-ph] (cit. on p. 34).

[114] Torbjörn Sjöstrand et al. 'An introduction to PYTHIA 8.2'.
In: *Comput. Phys. Commun.* 191 (2015), p. 159. arXiv: 1410.3012 [hep-ph]
(cit. on pp. 34, 35, 50).

[115] Bo Andersson, G. Gustafson, G. Ingelman and T. Sjöstrand. 'Parton fragmentation and string dynamics'. In: *Phys. Rept.* 97 (1983), pp. 31–145 (cit. on p. 34).

[116] B. R. Webber. 'A QCD Model for Jet Fragmentation Including Soft Gluon Interference'. In: *Nucl. Phys. B* 238 (1984), pp. 492–528 (cit. on p. 34).

[117] Yakov I. Azimov, Yuri L. Dokshitzer, Valery A. Khoze and S. I. Troian. 'The String Effect and QCD Coherence'. In: *Phys. Lett. B* 165 (1985), pp. 147–150 (cit. on p. 34).

[118] Zachary Marshall. 'Simulation of Pile-up in the ATLAS Experiment'. In: *J. Phys. Conf. Ser.* 513 (2014). Ed. by D. L. Groep and D. Bonacorsi, p. 022024 (cit. on p. 35).

[119] ATLAS Collaboration. 'The ATLAS Simulation Infrastructure'. In: *Eur. Phys. J. C* 70 (2010), pp. 823–874. arXiv: 1005.4568 [physics.ins-det] (cit. on p. 35).

[120] S. Agostinelli et al. 'GEANT4–a simulation toolkit'. In: *Nucl. Instrum. Meth. A* 506 (2003), pp. 250–303 (cit. on pp. 35, 90).

[121] P.J. Clark for the ATLAS Collaboration. 'The ATLAS Detector Simulation'. In: *Nuclear Physics B - Proceedings Supplements* 215.1 (2011). Proceedings of the 12th Topical Seminar on Innovative Particle and Radiation Detectors (IPRD10), pp. 85 –88. ISSN: 0920-5632. URL: http://www.sciencedirect.com/science/article/pii/S092056321100212X (cit. on pp. 35, 90).

[122] W Lukas. *Fast Simulation for ATLAS: Atlfast-II and ISF*. Tech. rep. Geneva: CERN, 2012. URL: https://cds.cern.ch/record/1458503 (cit. on p. 35).

[123] T Yamanaka. *The ATLAS calorimeter simulation FastCaloSim*. Tech. rep. Geneva: CERN, 2011. URL: https://cds.cern.ch/record/1322202 (cit. on p. 35).

[124] K. Edmonds et al. 'The fast ATLAS track simulation (FATRAS)'. In: (Mar. 2008) (cit. on p. 36).

[125] T. Cornelissen et al. 'The new ATLAS track reconstruction (NEWT)'. In: *J. Phys. Conf. Ser.* 119 (2008). Ed. by Randall Sobie, Reda Tafirout and Jana Thomson, p. 032014 (cit. on p. 39).

[126] ATLAS Collaboration. 'Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2'. In: *Eur. Phys. J. C* 77.10 (2017), p. 673. arXiv: 1704.07983 [hep-ex] (cit. on p. 39).

[127] Azriel Rosenfeld and John L. Pfaltz. 'Sequential Operations in Digital Picture Processing'. In: *J. ACM* 13.4 (1966), 471–494. ISSN: 0004-5411. URL: https://doi.org/10.1145/321356.321357 (cit. on p. 39).

[128] R. Frühwirth. 'Application of Kalman filtering to track and vertex fitting'. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 262.2 (1987), pp. 444–450. ISSN: 0168-9002. URL: https://www.sciencedirect.com/science/article/pii/0168900287908874 (cit. on p. 39).

[129] D Wicke. 'A New Algorithm for Solving Tracking Ambiguities'. In: (1998). URL: https://cds.cern.ch/record/2625731 (cit. on p. 40).

[130] ATLAS Collaboration. 'Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC'. In: *Eur. Phys. J. C* 77.5 (2017), p. 332. arXiv: 1611.10235 [physics.ins-det] (cit. on p. 40).

[131] ATLAS Collaboration. 'Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data'. In: *JINST* 14.12 (2019), P12006. arXiv: 1908.00005 [hep-ex] (cit. on pp. 40, 43, 165).

[132] ATLAS Collaboration. 'Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s}$ = 13 TeV'. In: *Eur. Phys. J. C* 79.8 (2019), p. 639. arXiv: 1902.04655 [physics.ins-det] (cit. on pp. 41, 165).

[133] ATLAS Collaboration. 'Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton-proton collision data'. In: *JINST* 14.03 (2019), P03017. arXiv: 1812.03848 [hep-ex] (cit. on p. 42).

[134] ATLAS Collaboration. 'Muon reconstruction and identification efficiency in ATLAS using the full Run 2 *pp* collision data set at $\sqrt{s} = 13$ TeV'. In: *Eur. Phys. J. C* 81.7 (2021), p. 578. arXiv: 2012.00578 [hep-ex] (cit. on pp. 44–46).

[135] J. Illingworth and J. Kittler. 'A survey of the hough transform'. In: *Computer Vision, Graphics, and Image Processing* 44.1 (1988), pp. 87–116. ISSN: 0734-189X. URL: https://www.sciencedirect.com/science/article/pii/S0734189X88800331 (cit. on p. 44).

[136] ATLAS Collaboration. 'Jet reconstruction and performance using particle flow with the ATLAS Detector'. In: *Eur. Phys. J. C* 77.7 (2017), p. 466. arXiv: 1703.10485 [hep-ex] (cit. on pp. 46, 47).

[137] Matteo Cacciari, Gavin P Salam and Gregory Soyez. 'The anti-k_t jet clustering algorithm'. In: *Journal of High Energy Physics* 2008.04 (2008), p. 063. URL: https://doi.org/10.1088%2F1126-6708%2F2008%2F04%2F063 (cit. on pp. 47, 48).

[138] Stephen D. Ellis and Davison E. Soper. 'Successive combination jet algorithm for hadron collisions'. In: *Phys. Rev. D* 48 (1993), pp. 3160–3166. arXiv: hep-ph/9305266 (cit. on p. 47).

[139] Yuri L. Dokshitzer, G. D. Leder, S. Moretti and B. R. Webber. 'Better jet clustering algorithms'. In: *JHEP* 08 (1997), p. 001. arXiv: hep-ph/9707323 (cit. on p. 48).

[140] G. Corcella et al. 'HERWIG 6.5 release note'. In: (Oct. 2002). arXiv: hep-ph/0210213 (cit. on p. 48).

[141] Matteo Cacciari and Gavin P. Salam. 'Pileup subtraction using jet areas'. In: *Phys. Lett. B* 659 (2008), pp. 119–126. arXiv: 0707.1378 [hep-ph] (cit. on p. 49).

[142] Matteo Cacciari, Gavin P. Salam and Gregory Soyez. 'FastJet user manual'. In: *Eur. Phys. J. C* 72 (2012), p. 1896. arXiv: 1111.6097 [hep-ph] (cit. on p. 49).

[143] ATLAS Collaboration. 'Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector'. In: *Eur. Phys. J. C* 81.8 (2021), p. 689. arXiv: 2007.02645 [hep-ex] (cit. on pp. 49, 50, 166).

[144] ATLAS Collaboration. 'Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s}$ = 13 TeV'. In: *Eur. Phys. J. C* 78.11 (2018), p. 903. arXiv: 1802.08168 [hep-ex] (cit. on pp. 51, 166).

[145] Michael R. Douglas. 'Large Language Models'. In: (July 2023). arXiv: 2307.05782 [cs.CL] (cit. on p. 53).

[146] Manzil Zaheer et al. *Deep Sets*. 2018. arXiv: 1703.06114 [cs.LG] (cit. on pp. 53, 69).

[147] Sepp Hochreiter and Jürgen Schmidhuber. 'Long Short-Term Memory'. In: *Neural Comput.* 9.8 (1997), pp. 1735–1780 (cit. on p. 53).

[148] Ian H. Witten, Eibe Frank, Mark A. Hall and Christopher J. Pal. 'Data Mining (Fourth Edition), Chapter 10 - Deep learning'. In: (2017), pp. 417–466. URL: https://www.sciencedirect.com/science/article/pii/B9780128042915000106 (cit. on p. 53).

[149] *TikZ.net Neural Networks*. https://tikz.net/neural_networks/. Accessed: 2023-10-20 (cit. on pp. 54, 56, 62).

[150] Abien Fred Agarap. 'Deep learning using rectified linear units (relu)'. In: *arXiv preprint arXiv:1803.08375* (2018) (cit. on p. 55).

[151] John Scott Bridle. ‘Probabilistic Interpretation of Feedforward Classification Network Outputs, with Relationships to Statistical Pattern Recognition’. In: *NATO Neurocomputing*. 1989. URL: https://api.semanticscholar.org/CorpusID:59636530 (cit. on p. 56).

[152] G Guerrieri. ‘Regression Deep Neural Networks for top-quark-pair resonance searches in the dilepton channel’. In: *Nuovo Cimento C* 45.5 (2022), p. 110. URL: https://cds.cern.ch/record/2837854 (cit. on p. 56).

[153] Shun ichi Amari. ‘Backpropagation and stochastic gradient descent method’. In: *Neurocomputing* 5.4 (1993), pp. 185–196. ISSN: 0925-2312. URL: https://www.sciencedirect.com/science/article/pii/092523129390006O (cit. on p. 58).

[154] David E. Rumelhart, Geoffrey E. Hinton and Ronald J. Williams. ‘Learning representations by back-propagating errors’. In: *Nature* 323.6088 (Oct. 1986), pp. 533–536. URL: http://dx.doi.org/10.1038/323533a0 (cit. on p. 59).

[155] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever and Ruslan Salakhutdinov. ‘Dropout: A Simple Way to Prevent Neural Networks from Overfitting’. In: *Journal of Machine Learning Research* 15.56 (2014), pp. 1929–1958. URL: http://jmlr.org/papers/v15/srivastava14a.html (cit. on p. 62).

[156] Sergey Ioffe and Christian Szegedy. ‘Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift’. In: *CoRR* abs/1502.03167 (2015). arXiv: 1502.03167. URL: http://arxiv.org/abs/1502.03167 (cit. on p. 63).

[157] Alex Sherstinsky. ‘Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network’. In: *CoRR* abs/1808.03314 (2018). arXiv: 1808.03314. URL: http://arxiv.org/abs/1808.03314 (cit. on p. 64).

[158] Michael Mozer. ‘A Focused Backpropagation Algorithm for Temporal Pattern Recognition’. In: *Complex Systems* 3 (Jan. 1995) (cit. on p. 65).

[159] A. J. Robinson and Frank Fallside. *The Utility Driven Dynamic Error Propagation Network*. Tech. rep. CUED/F-INFENG/TR.1. Cambridge, UK: Engineering Department, Cambridge University, 1987 (cit. on p. 65).

[160] Paul J. Werbos. ‘Generalization of backpropagation with application to a recurrent gas market model’. In: *Neural Networks* 1.4 (1988), pp. 339–356. ISSN: 0893-6080. URL: https://www.sciencedirect.com/science/article/pii/089360808890007X (cit. on p. 65).

[161] J. Leon. *How to Draw a LSTM Network in TikZ?*
URL: https://tex.stackexchange.com/questions/432312/how-do-i-draw-an-lstm-cell-in-tikz/432344#432344 (visited on 25/10/2023) (cit. on p. 66).

[162] Kyunghyun Cho et al. 'Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation'. In: *CoRR* abs/1406.1078 (2014). arXiv: 1406.1078. URL: http://arxiv.org/abs/1406.1078 (cit. on p. 68).

[163] Ashish Vaswani et al. 'Attention Is All You Need'. In: *CoRR* abs/1706.03762 (2017). arXiv: 1706.03762. URL: http://arxiv.org/abs/1706.03762 (cit. on pp. 69–71).

[164] Jonas Gehring, Michael Auli, David Grangier, Denis Yarats and Yann N. Dauphin. *Convolutional Sequence to Sequence Learning*. 2017. arXiv: 1705.03122 [cs.CL] (cit. on p. 70).

[165] Jimmy Lei Ba, Jamie Ryan Kiros and Geoffrey E. Hinton. 'Layer Normalization'. In: (2016). arXiv: 1607.06450. URL: https://arxiv.org/abs/1607.06450 (cit. on p. 70).

[166] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun. 'Deep Residual Learning for Image Recognition'. In: *CoRR* abs/1512.03385 (2015). arXiv: 1512.03385. URL: http://arxiv.org/abs/1512.03385 (cit. on p. 70).

[167] William A. Bardeen, A. J. Buras, D. W. Duke and T. Muta. 'Deep-inelastic scattering beyond the leading order in asymptotically free gauge theories'. In: *Phys. Rev. D* 18 (11 1978), pp. 3998–4017. URL: https://link.aps.org/doi/10.1103/PhysRevD.18.3998 (cit. on p. 78).

[168] ATLAS Collaboration. 'Configuration and performance of the ATLAS b-jet triggers in Run 2'. In: *Eur. Phys. J. C* 81.12 (2021), p. 1087. arXiv: 2106.03584 [hep-ex] (cit. on p. 78).

[169] ATLAS Collaboration. *Tagging and suppression of pileup jets with the ATLAS detector*. ATLAS-CONF-2014-018. 2014. URL: https://cds.cern.ch/record/1700870 (cit. on p. 79).

[170] ATLAS Collaboration. *Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment*. Tech. rep. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2017-003. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2255226 (cit. on p. 80).

[171] ATLAS Collaboration. 'ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV'. In: *Eur. Phys. J. C* 79.11 (2019), p. 970. arXiv: 1907.05120 [hep-ex] (cit. on pp. 80, 86).

[172] ATLAS Collaboration. 'ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset'. In: *Eur. Phys. J. C* 83.7 (2023), p. 681. arXiv: 2211.16345 [physics.data-an] (cit. on pp. 80–82, 85, 87, 88, 90, 114, 211, 212).

[173] ATLAS Collaboration. *Flavour-tagging efficiency corrections for the 2019 ATLAS PFlow jet and VR track jets b-taggers with the full LHC Run II dataset*. URL: http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2021-001/ (visited on 07/11/2023) (cit. on p. 80).

[174] ATLAS Collaboration. 'Measurement of the c-jet mistagging efficiency in $t\bar{t}$ events using pp collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS detector'. In: *Eur. Phys. J. C* 82.1 (2022), p. 95. arXiv: 2109.10627 [hep-ex] (cit. on p. 80).

[175] ATLAS Collaboration. 'Calibration of the light-flavour jet mistagging efficiency of the b-tagging algorithms with Z+jets events using 139 fb^{-1} of ATLAS proton–proton collision data at $\sqrt{s} = 13$ TeV'. In: *Eur. Phys. J. C* 83.8 (2023), p. 728. arXiv: 2301.06319 [hep-ex] (cit. on p. 80).

[176] François Chollet et al. *Keras*. 2015. URL: https://keras.io (cit. on p. 83).

[177] The Theano Development Team and Rami Al-Rfou et al. *Theano: A Python framework for fast computation of mathematical expressions*. 2016. arXiv: 1605.02688 [cs.SC] (cit. on p. 83).

[178] Diederik P Kingma and J Adam Ba. 'A method for stochastic optimization'. In: (2017). arXiv: 1412.6980. URL: https://arxiv.org/abs/1412.6980 (cit. on p. 83).

[179] Xavier Glorot and Yoshua Bengio. 'Understanding the difficulty of training deep feedforward neural networks'. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. Ed. by Yee Whye Teh and Mike Titterington. Vol. 9. Proceedings of Machine Learning Research. Chia Laguna Resort, Sardinia, Italy: PMLR, 2010, pp. 249–256. URL: https://proceedings.mlr.press/v9/glorot10a.html (cit. on p. 83).

[180] ATLAS Collaboration and Jovan Mitrevski. *Preparing ATLAS reconstruction software for LHC's Run 2*. Tech. rep. 7. Geneva: CERN, 2015. URL: https://cds.cern.ch/record/2016146 (cit. on pp. 83, 95).

[181] Daniel Hay Guest et al. *lwtnn/lwtnn*. Version v2.13. 2022. URL: https://doi.org/10.5281/zenodo.597221 (cit. on pp. 83, 84, 95).

[182] ATLAS Collaboration. 'Measurement of VH, $\mathrm{H} \to \mathrm{b\bar{b}}$ production as a function of the vector-boson transverse momentum in 13 TeV pp collisions with the ATLAS detector'. In: *JHEP* 05 (2019), p. 141. arXiv: 1903.04618 [hep-ex] (cit. on p. 84).

[183] Yann Coadou. 'Boosted decision trees'. In: *Artificial Intelligence for High Energy Physics*. Mar. 2022. Chap. Chapter 2, pp. 9–58. arXiv: 2206.09645 [physics.data-an]. URL: https://arxiv.org/abs/1412.6980 (cit. on p. 86).

[184] Paolo Nason. 'A new method for combining NLO QCD with shower Monte Carlo algorithms'. In: *JHEP* 11 (2004), p. 040. arXiv: hep-ph/0409146 (cit. on pp. 89, 156).

[185] Stefano Frixione, Paolo Nason and Carlo Oleari. 'Matching NLO QCD computations with parton shower simulations: the POWHEG method'. In: *JHEP* 11 (2007), p. 070. arXiv: 0709.2092 [hep-ph] (cit. on pp. 89, 156).

[186] Simone Alioli, Paolo Nason, Carlo Oleari and Emanuele Re. 'A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX'. In: *JHEP* 06 (2010), p. 043. arXiv: 1002.2581 [hep-ph] (cit. on pp. 89, 156).

[187] T. Sjöstrand, S. Mrenna and P. Skands. 'A brief introduction to PYTHIA 8.1'. In: *Comput. Phys. Commun.* 178 (2008), pp. 852–867. arXiv: 0710.3820 [hep-ph] (cit. on pp. 89, 157).

[188] ATLAS Collaboration. *ATLAS Pythia 8 tunes to 7 TeV data*. ATL-PHYS-PUB-2014-021. 2014. URL: https://cds.cern.ch/record/1966419 (cit. on pp. 89, 167).

[189] Richard D. Ball et al. 'Parton distributions for the LHC run II'. In: *JHEP* 04 (2015), p. 040. arXiv: 1410.8849 [hep-ph] (cit. on p. 89).

[190] Richard D. Ball et al. 'Parton distributions with LHC data'. In: *Nucl. Phys. B* 867 (2013), p. 244. arXiv: 1207.1303 [hep-ph] (cit. on p. 89).

[191] D. J. Lange. 'The EvtGen particle decay simulation package'. In: *Nucl. Instrum. Meth. A* 462 (2001), p. 152 (cit. on p. 89).

[192] ATLAS Collaboration. *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model*. ATL-PHYS-PUB-2016-017. 2016. URL: https://cds.cern.ch/record/2206965 (cit. on p. 89).

[193] Manuel Guth. 'Search for $t\bar{t}H$ $(H \to b\bar{b})$ Production in the Lepton + Jets Channel and Quark Flavour Tagging with Deep Learning at the ATLAS Experiment. Recherche de la production ttH(bb) dans le canal lepton+jets et étiquetage de quarks de saveur lourde par apprentissage profond dans l'expérience ATLAS'. Presented 25 Mar 2021. Albert-Ludwigs-Universität Freiburg, Université Paris-Saclay, 2021. URL: https://cds.cern.ch/record/2765038 (cit. on pp. 90, 92).

[194] ATLAS Collaboration.
Deep Sets based Neural Networks for Impact Parameter Flavour Tagging in ATLAS.
ATL-PHYS-PUB-2020-014. 2020. URL: https://cds.cern.ch/record/2718948
(cit. on p. 97).

[195] ATLAS Collaboration. 'Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector'.
In: *Physics Letters B* 784 (2018), pp. 173 –191. ISSN: 0370-2693 (cit. on p. 147).

[196] H. David Politzer and Stephen Wolfram.
'Bounds on Particle Masses in the Weinberg-Salam Model'.
In: *Phys. Lett. B* 82 (1979). [Erratum: Phys.Lett.B 83, 421 (1979)], pp. 242–246
(cit. on p. 147).

[197] H. David Politzer and Stephen Wolfram.
'Erratum for: Bounds on particle masses in the Weinberg-Salam model'.
In: *Physics Letters B* 83.3 (1979), pp. 421–422. ISSN: 0370-2693. URL: https://www.sciencedirect.com/science/article/pii/0370269379911444
(cit. on p. 147).

[198] Pham Quang Hung. 'Vacuum Instability and New Constraints on Fermion Masses'.
In: *Phys. Rev. Lett.* 42 (1979), p. 873 (cit. on p. 147).

[199] ATLAS Collaboration.
'Search for the standard model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector'.
In: *Phys. Rev. D* 97 (2018), p. 072016. arXiv: 1712.08895 [hep-ex] (cit. on p. 147).

[200] CMS Collaboration. *Measurement of* $\mathrm{t\bar{t}H}$ *production in the* $\mathrm{H} \to \mathrm{b\bar{b}}$ *decay channel in* $41.5\,\mathrm{fb}^{-1}$ *of proton-proton collision data at* $\sqrt{s} = 13\,\mathrm{TeV}$.
Tech. rep. CMS-PAS-HIG-18-030. Geneva: CERN, 2019.
URL: http://cds.cern.ch/record/2675023 (cit. on p. 147).

[201] CMS Collaboration. *Measurement of the* $\mathrm{t\bar{t}H}$ *and* tH *production rates in the* $\mathrm{H} \to \mathrm{b\bar{b}}$ *decay channel with* $138\,\mathrm{fb}^{-1}$ *of proton-proton collision data at* $\sqrt{s} = 13\,\mathrm{TeV}$. Tech. rep.
Geneva: CERN, 2023. URL: https://cds.cern.ch/record/2868175
(cit. on pp. 147, 149–151).

[202] Benjamin Nachman, Pascal Nef, Ariel Schwartzman, Maximilian Swiatlowski and Chaowaroj Wanotayaroj. 'Jets from jets: re-clustering as a tool for large radius jet reconstruction and grooming at the LHC'. In: *JHEP* 02 (2015), p. 075.
arXiv: 1407.2922 [hep-ph] (cit. on p. 151).

[203] J. Neyman and E. S. Pearson.
'On the Problem of the Most Efficient Tests of Statistical Hypotheses'.
In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers*

of a Mathematical or Physical Character 231 (1933), pp. 289–337. ISSN: 02643952. URL: http://www.jstor.org/stable/91247 (cit. on pp. 156, 163).

[204] Heribertus B. Hartanto, Barbara Jäger, Laura Reina and Doreen Wackeroth. 'Higgs boson production in association with top quarks in the POWHEG BOX'. In: *Phys. Rev. D* 91.9 (2015), p. 094003. arXiv: 1501.04498 [hep-ph] (cit. on p. 156).

[205] Johannes Bellm et al. 'Herwig 7.0/Herwig++ 3.0 release note'. In: *Eur. Phys. J. C* 76.4 (2016), p. 196. arXiv: 1512.01178 [hep-ph] (cit. on p. 157).

[206] Paolo Nason and Carlo Oleari. 'Generation cuts and Born suppression in POWHEG'. In: (Mar. 2013). arXiv: 1303.3922 [hep-ph] (cit. on p. 158).

[207] Tomáš Ježo, Jonas M. Lindert, Niccolo Moretti and Stefano Pozzorini. 'New NLOPS predictions for $t\bar{t} + b$-jet production at the LHC'. In: *Eur. Phys. J. C* 78.6 (2018), p. 502. arXiv: 1802.00426 [hep-ph] (cit. on p. 158).

[208] Fabio Cascioli, Philipp Maierhöfer and Stefano Pozzorini. 'Scattering Amplitudes with Open Loops'. In: *Phys. Rev. Lett.* 108 (2012), p. 111601. arXiv: 1111.5206 [hep-ph] (cit. on p. 158).

[209] Federico Buccioni et al. 'OpenLoops 2'. In: *Eur. Phys. J. C* 79.10 (2019), p. 866. arXiv: 1907.13071 [hep-ph] (cit. on p. 158).

[210] ATLAS Collaboration. *Studies of Monte Carlo predictions for the $t\bar{t}b\bar{b}$ process*. Tech. rep. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2022-006. Geneva: CERN, 2022. URL: https://cds.cern.ch/record/2802806 (cit. on pp. 158, 168).

[211] Glen Cowan, Kyle Cranmer, Eilam Gross and Ofer Vitells. 'Asymptotic formulae for likelihood-based tests of new physics'. In: *Eur. Phys. J. C* 71 (2011), p. 1554. arXiv: 1007.1727 [physics.data-an] (cit. on pp. 163, 164).

[212] Wouter Verkerke and David Kirkby. *The RooFit toolkit for data modeling*. 2003. arXiv: physics/0306116 [physics.data-an] (cit. on p. 164).

[213] Lorenzo Moneta et. al. *The RooStats Project*. 2011. arXiv: 1009.1003 [physics.data-an] (cit. on p. 164).

[214] Kyle Cranmer, George Lewis, Lorenzo Moneta, Akira Shibata and Wouter Verkerke. *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*. Tech. rep. New York: New York U., 2012. URL: https://cds.cern.ch/record/1456844 (cit. on p. 164).

[215] F. James and M. Roos. 'Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations'. In: *Comput. Phys. Commun.* 10 (1975), pp. 343–367 (cit. on p. 164).

[216] Rene Brun and Fons Rademakers. 'ROOT – An object oriented data analysis framework'. In: *Nucl. Instrum. Meth. A* 389.1 (1997), pp. 81 –86. ISSN: 0168-9002 (cit. on p. 164).

[217] ATLAS Collaboration. 'Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC'. In: *Eur. Phys. J. C* 83.10 (2023), p. 982. arXiv: 2212.09379 [hep-ex] (cit. on p. 165).

[218] ATLAS Collaboration. 'Measurement of the Inelastic Proton–Proton Cross Section at $\sqrt{s} = 13$ TeV with the ATLAS Detector at the LHC'. In: *Phys. Rev. Lett.* 117 (2016), p. 182002. arXiv: 1606.02625 [hep-ex] (cit. on p. 165).

[219] ATLAS Collaboration. 'Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$ =13 TeV'. In: *Eur. Phys. J. C* 76.5 (2016), p. 292. arXiv: 1603.05598 [hep-ex] (cit. on pp. 165, 170).

[220] ATLAS Collaboration. 'Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector'. In: *Eur. Phys. J. C* 76 (2016), p. 581. arXiv: 1510.03823 [hep-ex] (cit. on p. 165).

[221] ATLAS Collaboration. 'Determination of jet calibration and energy resolution in proton–proton collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector'. In: (2019). arXiv: 1910.04482 [hep-ex] (cit. on p. 166).

[222] Risto Raitio and Walter W. Wada. 'Higgs-boson production at large transverse momentum in quantum chromodynamics'. In: *Phys. Rev. D* 19 (1979), p. 941 (cit. on p. 166).

[223] W. Beenakker et al. 'NLO QCD corrections to $t\bar{t}H$ production in hadron collisions'. In: *Nucl. Phys. B* 653 (2003), pp. 151–203. arXiv: hep-ph/0211352 (cit. on p. 166).

[224] S. Dawson, C. Jackson, L. H. Orr, L. Reina and D. Wackeroth. 'Associated Higgs boson production with top quarks at the CERN Large Hadron Collider: NLO QCD corrections'. In: *Phys. Rev. D* 68 (2003), p. 034022. arXiv: hep-ph/0305087 (cit. on p. 166).

[225] Yu Zhang, Wen-Gan Ma, Ren-You Zhang, Chong Chen and Lei Guo. 'QCD NLO and EW NLO corrections to $t\bar{t}H$ production with top quark decays at hadron collider'. In: *Phys. Lett. B* 738 (2014), pp. 1–5. arXiv: 1407.1110 [hep-ph] (cit. on p. 166).

[226] S. Frixione, V. Hirschi, D. Pagani, H.-S. Shao and M. Zaro. 'Electroweak and QCD corrections to top-pair hadroproduction in association with heavy bosons'. In: *JHEP* 06 (2015), p. 184. arXiv: 1504.03446 [hep-ph] (cit. on p. 166).

[227] John M. Campbell and R. Keith Ellis. '$t\bar{t}W^{+-}$ production and decay at NLO'. In: *JHEP* 07 (2012), p. 052. arXiv: 1204.5678 [hep-ph] (cit. on p. 168).

[228] M. Aliev et al. 'HATHOR – HAdronic Top and Heavy quarks crOss section calculatoR'. In: *Comput. Phys. Commun.* 182 (2011), pp. 1034–1046. arXiv: 1007.1327 [hep-ph] (cit. on p. 168).

[229] P. Kant et al. 'HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions'. In: *Comput. Phys. Commun.* 191 (2015), pp. 74–89. arXiv: 1406.4403 [hep-ph] (cit. on p. 168).

[230] A. D. Martin, W. J. Stirling, R. S. Thorne and G. Watt. 'Uncertainties on α_S in global PDF analyses and implications for predicted hadronic cross sections'. In: *Eur. Phys. J. C* 64 (2009), pp. 653–680. arXiv: 0905.3531 [hep-ph] (cit. on p. 168).

[231] Stefano Frixione, Eric Laenen, Patrick Motylinski, Chris White and Bryan R. Webber. 'Single-top hadroproduction in association with a W boson'. In: *JHEP* 07 (2008), p. 029. arXiv: 0805.3067 [hep-ph] (cit. on p. 168).

[232] ATLAS Collaboration. *Multi-boson simulation for* 13 *TeV ATLAS analyses*. ATL-PHYS-PUB-2016-002. 2016. URL: https://cds.cern.ch/record/2119986 (cit. on p. 168).

[233] J. Alwall et al. 'The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations'. In: *JHEP* 07 (2014), p. 079. arXiv: 1405.0301 [hep-ph] (cit. on p. 168).

[234] ATLAS Collaboration. 'Tools for estimating fake/non-prompt lepton backgrounds with the ATLAS detector at the LHC'. In: *JINST* 18.11 (2023), T11004. arXiv: 2211.16178 [hep-ex] (cit. on p. 169).

[235] ATLAS Collaboration. *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton–proton collision data*. ATLAS-CONF-2016-024. 2016. URL: https://cds.cern.ch/record/2157687 (cit. on p. 170).