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

Dissertation

zur Erlangung des Doktorgrades der

Fakultät für Mathematik und Physik der

ALBERT-LUDWIGS-UNIVERSITÄT FREIBURG

vorgelegt von

Manuel Guth

Betreuer:

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

Henri Bachacou (Université Paris-Saclay)

Frédéric Déliot (Université Paris-Saclay)

Datum der mündlichen Prüfung: 25. März 2021

DEKAN:	Prof. Dr. Wolfgang Soergel (Albert-Ludwigs-Universität Freiburg)
BETREUER DER ARBEIT:	Prof. Dr. Gregor Herten (Albert-Ludwigs-Universität Freiburg)
	Henri Bachacou (Université Paris-Saclay)
	Frédéric Déliot (Université Paris-Saclay)
GUTACHTER:	Prof. Dr. Florencia Canelli (Universität Zürich)
	Priv.-Doz. Dr. Johannes Erdmann (TU Dortmund)
PRÜFER:	Prof. Dr. Harald Ita (Albert-Ludwigs-Universität Freiburg)
	Gautier Hamel de Monchenault (Université Paris-Saclay)

Thèse de doctorat

NNT: 2021UPASP023

Search for $\mathrm{t\bar{t}H}$ $(\mathrm{H} \to \mathrm{b\bar{b}})$ 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

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 576,
Particules, Hadrons, Énergie et Noyau:
Instrumentation, Image, Cosmos et Simulation (PHENIICS)
Spécialité de doctorat: Physique des particules
Unité de recherche: Université Paris-Saclay, CEA,
Département de Physique des Particules,
91191, Gif-sur-Yvette, France.
Référent: : Faculté des sciences d'Orsay

Thèse présentée et soutenue à Freiburg, le 25 Mars 2021, par

Manuel Guth

Composition du jury

Prof. Dr. Harald Ita Professeur, Albert-Ludwigs Universität Freiburg	Président
Prof. Dr. Florencia Canelli Professeur, Universität Zürich	Rapporteure & Examinatrice
Priv.-Doz. Dr. Johannes Erdmann Privatdozent, Technische Universität Dortmund	Rapporteur & Examinateur
Prof. Dr. Gregor Herten Professeur, Albert-Ludwigs Universität Freiburg	Rapporteur & Examinateur
Gautier Hamel de Monchenault Chercheur sénior, Université Paris-Saclay	Examinateur

Direction de la thèse

Henri Bachacou Chercheur sénior, Université Paris-Saclay	Directeur de thèse
Prof. Dr. Gregor Herten Professeur, Albert-Ludwigs Universität Freiburg	Directeur de thèse & Examinateur
Frédéric Déliot Chercheur sénior, Université Paris-Saclay	Co-Directeur de thèse

ABSTRACT

Since several decades, the predictions of the Standard Model (SM) of particle physics are being probed and validated. One major success of the Large Hadron Collider (LHC) at CERN was the discovery of the Higgs boson in 2012. With the increasing amount of proton-proton collisions recorded with the experiments located at the LHC, precise Higgs measurements are now possible and rare processes are accessible.

ATLAS and CMS recently discovered the production process of a Higgs boson in association with a pair of top quarks using LHC RUN II data. The $t\bar{t}H(H \to b\bar{b})$ process allows for a direct measurement of the Top-Yukawa coupling which is the strongest fermion-Higgs coupling in the Standard Model and plays therefore an important role in Higgs physics. The challenging final state with at least 4 b-jets requires an advanced analysis strategy as well as sophisticated b-jet identification methods. b-tagging is not only crucial in the $t\bar{t}H(b\bar{b})$ analysis, but most physics analyses within ATLAS are making use of it. The reoptimisation of the deep-learning-based heavy flavour tagger in ATLAS is shown in this thesis for two different jet collections. Various improvements were made resulting in a drastic performance increase up to a factor two in certain regions of the phase space. The $t\bar{t}H(b\bar{b})$ analysis is performed using $139\,fb^{-1}$ of RUN II ATLAS data at a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$. The signal strength, being the ratio of the measured cross-section over the predicted cross-section in the SM, was measured to be $0.43^{+0.20}_{-0.19}(\text{stat.})^{+0.30}_{-0.27}(\text{syst.})$ with an observed (expected) significance of 1.3 (3.0) standard deviations in the inclusive cross-section measurement. In addition, a simplified template cross-section (STXS) measurement in different Higgs p_T bins is performed which is possible because of the ability to reconstruct the Higgs boson. The measurement is limited by the capability to describe the challenging irreducible $t\bar{t} + b\bar{b}$ background and by systematic uncertainties.

KURZZUSAMMENFASSUNG

Seit mehreren Jahrzehnten werden die Vorhersagen des Standardmodells (SM) der Teilchenphysik erprobt und validiert. Mit der zunehmenden Anzahl von Proton-Proton-Kollisionen, die mit den Experimenten am LHC aufgezeichnet werden, sind nun präzise Higgs-Messungen möglich.
ATLAS und CMS haben kürzlich den $t\bar{t}H$-Produktionsprozess mit Hilfe von LHC RUN II-Daten entdeckt. Der $t\bar{t}H(H \to b\bar{b})$-Prozess ermöglicht eine direkte Messung der Top-Yukawa-Kopplung, welche die stärkste Fermion-Higgs-Kopplung ist und daher eine wichtige Rolle im SM einnimmt. Der anspruchsvolle Endzustand mit mindestens 4 b-Jets erfordert eine fortschrittliche Analysestrategie sowie elaborierte b-Jet-Identifikationsmethoden. b-Tagging ist nicht nur in der $t\bar{t}H(b\bar{b})$-Analyse von entscheidender Bedeutung, sondern die meisten Physik-Analysen innerhalb von ATLAS machen davon Gebrauch. Die Re-Optimierung des Deep-Learning-basierten Heavy-Flavour Taggers in ATLAS wird in dieser Arbeit für zwei verschiedene Jet-Definitionen gezeigt. Es wurden verschiedene Änderungen vorgenommen, die zu einer signifikanten Verbesserung von bis zu einem Faktor zwei in der Untergrundunterdrückung in bestimmten Phasenraumregionen führten. Die $t\bar{t}H(b\bar{b})$-Analyse wurde mit $139\,fb^{-1}$ RUN II ATLAS-Daten bei einer Schwerpunktsenergie von $\sqrt{s} = 13\,\text{TeV}$ durchgeführt. Die Signalstärke, d.h. das Verhältnis des gemessenen Wirkungsquerschnitts zum vorhergesagten Wirkungsquerschnitt im SM, wurde mit $0,43^{+0,20}_{-0,19}(\text{stat.})^{+0,30}_{-0,27}(\text{syst.})$ mit einer beobachteten (erwarteten) Signifikanz von $1,3$ $(3,0)$ Standardabweichungen für den inklusiven Wirkungsquerschnitt gemessen. Zusätzlich wurde zum ersten Mal eine vereinfachte differenzielle Wirkungsquerschnittsmessung in verschiedenen Higgs p_T-Bereichen durchgeführt. Die Messung wird durch systematische Unsicherheiten begrenzt, hauptsächlich im Zusammenhang mit dem anspruchsvollen irreduziblen $t\bar{t} + b\bar{b}$ Untergrund.

RESUMÉ

ATLAS et CMS ont récemment découvert le processus de production $t\bar{t}H$ en utilisant les données prises durant le RUN II du LHC. Le processus $t\bar{t}H(H \to b\bar{b})$ permet de mesurer directement le couplage de Yukawa du quark top, qui est le couplage fermion-Higgs le plus grand du modèle standard et joue donc un rôle important dans la physique du boson du Higgs. L'état final de ce processus contient au moins 4 jets provenant de quarks b ce qui nécessite d'établir une stratégie d'analyse avancée ainsi que de développer des méthodes sophistiquées pour l'identification des jets provenant de quarks b. L'étiquetage des quarks b n'est pas seulement crucial pour l'analyse $t\bar{t}H(b\bar{b})$, mais aussi pour la plupart des analyses de physique au sein de l'expérience d'ATLAS. La ré-optimisation de l'étiquetage des quarks de saveurs lourdes basé sur un apprentissage profond dans ATLAS est présentée dans cette thèse pour deux collections de jets différentes. Diverses améliorations ont été apportées, entraînant une augmentation importantes des performances allant jusqu'à un facteur deux dans certaines régions de l'espace des phases. L'analyse $t\bar{t}H(b\bar{b})$ est effectuée en utilisant $139\,fb^{-1}$ de données enregistrées par ATLAS durant le RUN II à une énergie dans le centre de masse de $\sqrt{s} = 13\,\text{TeV}$. L'intensité du signal, qui est le rapport entre la section efficace mesurée et la section efficace prédite par le modèle standard, a été mesurée à $0,43^{+0,20}_{-0,19}(\text{stat.})^{+0,30}_{-0,27}(\text{syst.})$ avec une signification observée (prévue) de $1,3$ $(3,0)$ déviations standard pour la mesure de la section efficace inclusive. En outre, une mesure simplifiée de la section efficace utilisant des gabarits Monte Carlo en fonction de l'impulsion transverse du boson de Higgs est effectuée. Cette mesure est limitée par la difficulté de simuler correctement le bruit de fond dominant $t\bar{t} + b\bar{b}$ ainsi que par de grandes incertitudes systématiques.

CONTRIBUTION BY THE AUTHOR

As member of the ATLAS collaboration, the author focused on different projects, the most significant contributions are described below.

HEAVY FLAVOUR TAGGING All the work related to heavy-flavour tagging described in the chapters 9 to 12 was almost exclusively done by the author of this thesis providing a new algorithm which is now used by almost all physics analyses in ATLAS.
The performance improvements achieved in this thesis are published as *public plots* in Ref. [1] for *PFlow* jets and for *VR Track* jets in Ref. [2]. Furthermore, a hyperparameter optimisation setup with GRID GPUs was developed in cooperation with ATLAS IT providing a tool for the collaboration which is documented in Ref. [3] together with publicly available plots in Ref. [4]. The development of an extended flavour tagger (adding a $b\bar{b}$-jet category) was initiated and several students were supervised by the author in this regime. The author also took a key role in the flavour tagging group, besides the algorithm optimisation, giving machine learning tutorials within ATLAS and being the liaison between the top-physics group and the flavour tagging group as well as the machine learning liaison of the flavour tagging group.

$t\bar{t}H(b\bar{b})$ ANALYSIS In the $t\bar{t}H(b\bar{b})$ analysis described in the chapters 13 to 15, the author of this dissertation was one of the main analysers. The author of this dissertation assumed a leading role. The author worked on the optimisation of the simplified cross-section analysis in the lepton+jets resolved channel. This includes the study of different fit models, the development of additional uncertainties as well as the close collaboration with the analysers of the boosted lepton+jets and resolved dilepton channels. He furthermore supervised an undergraduate student working on a related project. The results are documented in Ref. [5].

CONTENTS

1 INTRODUCTION

The field of high energy physics made major progress in the recent years probing the predictions of the Standard Model (SM) of particle physics which describes the interactions of the fundamental building blocks in Nature. The international research facility CERN is hosting the Large Hadron Collider (LHC). A milestone in the research program of the multi-purpose particle detectors ATLAS and CMS at the LHC was the discovery of the Higgs boson with a mass of about 125 GeV in 2012 [6, 7]. With the increasing amount of particle collisions delivered by the LHC, one major goal of the ATLAS experiment is to measure the properties of the Higgs boson more precisely. In particular, the Higgs boson and the heaviest fermion in the SM, the top quark, have a special relationship. The Higgs boson production mode in association with a pair of top quarks ($t\bar{t}H$) was recently observed [8–10], marking an important discovery. The measurement of this production channel is directly sensitive to the top Yukawa coupling, describing the interaction of the top quark with the Higgs boson. Since the Yukawa coupling increases proportionally to the fermion mass, the top Yukawa coupling is the strongest in the SM.

The two top quarks in the final state of the $t\bar{t}H$ process offer a distinct signature in the detector. Since top quarks decay almost exclusively into b-quarks and W-bosons, the identification of b-jets is an important tool to identify the signal and reject many background processes. Therefore, the ATLAS collaboration has developed sophisticated heavy-flavour tagging algorithms which are playing an important role in most physics analyses. Within this thesis, a deep-learning-based flavour tagger has been optimised for two different jet clustering algorithms. Various improvements were made, providing a new heavy-flavour tagger to the collaboration. In addition, an outlook towards a new machine learning design of the flavour tagger is presented, opening new opportunities for future improvements.

The Higgs boson decay channel to a pair of b-quarks ($H \to b\bar{b}$) has the largest branching fraction in the SM and also heavily benefits from good b-tagging performance. This decay channel has been observed by CMS and ATLAS [11, 12]; however, the $t\bar{t}H(b\bar{b})$ process itself has not yet been observed. Both experiments already published first results of the $t\bar{t}H(b\bar{b})$ process with a subset of the LHC RUN II dataset [13, 14]. In this thesis, the $t\bar{t}H(b\bar{b})$ analysis [5] is presented, performed with the full LHC RUN II dataset of $139\,\text{fb}^{-1}$ proton-proton collisions recorded with the ATLAS experiment at a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$.

Even though the SM is being tested with high precision, there are hints, especially from astrophysical observations, that there are physics phenomena beyond the SM. The coupling of the top quark to the Higgs boson is especially sensitive to the presence of new physics. Also, the measurement of the differential cross-section of the $t\bar{t}H$ process in bins of the Higgs transverse momentum in the Simplified Template Cross-Section (STXS) framework is sensitive to such effects. Such a measurement allows

to access the CP structure of the Higgs boson and to probe anomalous Higgs self-couplings with an increased sensitivity of the measurement.

This dissertation is structured into four main parts: In Part I an overview of the Standard Model and the ATLAS experiment is given, along with a short introduction into the physics simulation at hadron colliders followed by the current status of $\mathrm{t\bar{t}H}$ measurements at the LHC. The object reconstruction and particle identification in ATLAS is discussed in Part II together with a short introduction to machine learning techniques. Part III focuses on the optimisation of the b-jet identification using a deep-learning-based approach for two different jet collections as well as an outlook into ongoing developments in flavour tagging. The $\mathrm{t\bar{t}H(b\bar{b})}$ analysis is finally introduced in Part IV, where a detailed description of the analysis strategy and the results for the inclusive cross-section and the Simplified Template Cross-Section (STXS) measurement are given.

Part I

OVERVIEW

2

THE STANDARD MODEL OF PARTICLE PHYSICS

The SM of elementary particle physics is the theoretical framework describing the known elementary particles and their interactions comprising all fundamental forces - the electromagnetic, the strong and the weak force - except the gravitational force. This theory has been probed over the last decades with enormous precision, although there are also hints for physics beyond its scope.

The SM is inspired by two main principles: simplicity and symmetries. It is a non-abelian gauge theory invariant under the gauge group

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

described in the framework of Lorentz invariant Quantum Field Theory (QFT) with the Lagrangian being renormalisable and invariant under local gauge transformation.
This chapter gives a brief overview of the particle content of the SM in Section 2.1 followed by the description of Quantum Electrodynamics (QED) and Quantum Chromodynamics (QCD) in Sections 2.2 and 2.3, respectively. Afterwards, the electroweak unification (sec. 2.4) and the Higgs sector (sec. 2.6) are introduced. The content of this chapter is mainly inspired by [15–17].

2.1 PARTICLE CONTENT OF THE STANDARD MODEL

The SM comprises all known elementary particles summarised in Figure 2.1. It consists of twelve fermions (half-integer spin particles), twelve vector bosons (spin-1 particles) and the Higgs boson, a scalar particle (spin 0).

Fermions are sorted into three generations comprising one charged lepton, one neutrino and two quarks each. The particles of a different generation have identical quantum numbers with the exception of their mass. In fact, the ordinary matter is only composed of the first generation fermions. In addition, every fermion[1] has also an associated anti-particle with opposite charge.
Quarks carry an electric and a colour charge and are therefore interacting weakly, electromagnetically and strongly. Each generation has an up-type quark (up-, charm- & top-quark) and a down-type quark (down-, strange & bottom-quark) with an electric charge of $Q = 2/3$ and $Q = -1/3$, respectively. In general, quarks can only occur in bound states due to the colour confinement [21]. The colour charge was introduced to maintain the Pauli principle and explain the coexistence of quarks in hadrons in otherwise identical quantum states. These bound states are called hadrons and they can be either fermions formed out of three quarks denoted as baryons or bosons composed of a quark and

1 In fact, the W-boson also has its anti-particle.

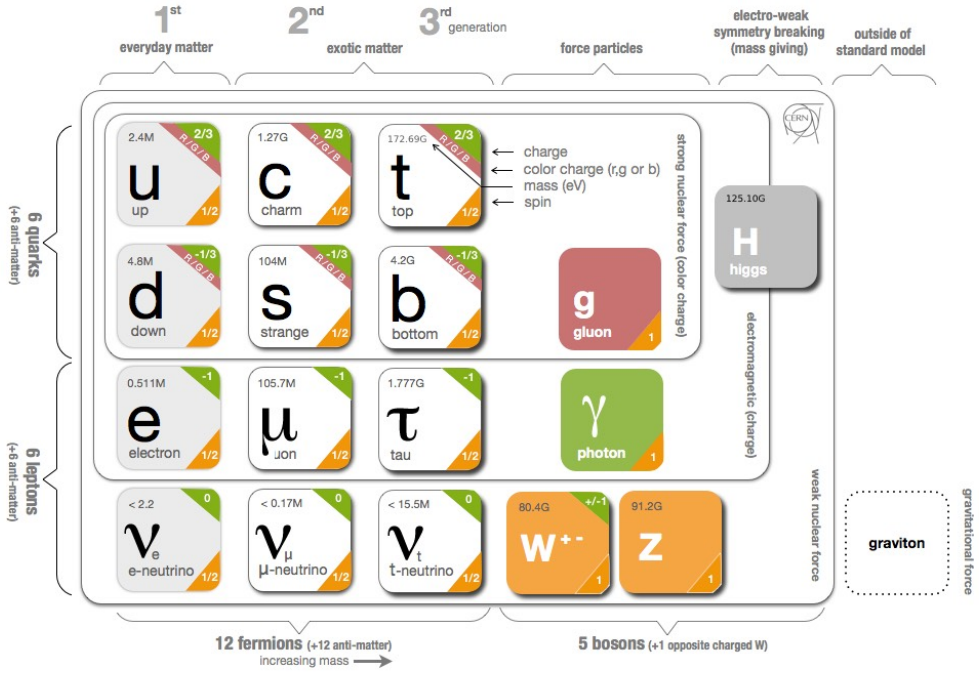


Figure 2.1.: Overview of the particles in the Standard Model [18]. Adapted the top quark mass according to Ref. [19] and the Higgs boson mass according to Ref. [20].

an anti-quark denoted as mesons[2].
Leptons are the electron e, muon μ and tau τ and their associated neutrinos ν_e, ν_μ and ν_τ, respectively. The neutrinos are considered massless in the SM. While the charged leptons (e, μ, τ) carry an electric charge $Q = -1$ and can interact electromagnetically, neutrinos carry neither electric nor colour charge and therefore are only interacting via the weak force.

The vector bosons are gauge bosons and act as force carriers. The massless photon γ is the mediator of the electromagnetic force while the massive Z and $W^{\pm}$ bosons are associated to the weak force. To be strictly accurate, they are all associated to the electroweak theory which unifies the electromagnetic and weak theory, described in Section 2.4. Furthermore, there are eight types of gluons g carrying the strong force.
The Higgs boson is the only scalar particle of the SM. The Higgs mechanism and the concept of electroweak symmetry breaking (EWSB) is discussed in more detail in Section 2.6.

In the following, using the terms electrons, muons and taus comprise always the particles and anti-particles if not stated differently. The same is valid for quarks and anti-quarks.

2 The LHCb collaboration discovered also penta- and tetra-quark states [22].

2.2 QUANTUM ELECTRODYNAMICS

The framework of QFT combines quantum mechanics and special relativity and thus particles are represented as fields. QED is the theoretical description of the electromagnetic interactions, based on the abelian $U(1)$ gauge group, being a generalisation of Maxwell's theory.
A freely propagating fermion field corresponding to a massive spin $1/2$ particle is described by the Dirac Lagrangian

$$\mathcal{L}_{\text{Dirac}} = \bar{\psi}(i\not{\partial} - m)\psi, \tag{2.2}$$

where $\not{\partial} = \gamma^\mu \partial_\mu$ denotes the contraction with the Dirac matrices γ^μ, the fermion mass m and a free spinor field ψ.
A local $U(1)$ gauge transformation would lead to an additional term $\bar{\psi}\not{\partial}\alpha\psi$ in the Lagrangian, with α being the electromagnetic coupling constant. It is enforced that the QED Lagrangian is invariant under this gauge transformation and thus a coupling between the Dirac fermion and the vector field A_μ (corresponding to the photon) is introduced in the form of a covariant derivative

$$D_\mu = \partial_\mu + ieA_\mu(x), \tag{2.3}$$

where $e = -|e|$ is the electron charge.
The QED lagrangian results in

$$\begin{aligned}
\mathcal{L}_{\text{QED}} &= \mathcal{L}_{\text{Dirac}} + \mathcal{L}_{\text{Maxwell}} + \mathcal{L}_{\text{interaction}} && (2.4)\\
&= \bar{\psi}\left(i\not{\partial} - m\right)\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} - e\bar{\psi}\gamma^\mu\psi A_\mu && (2.5)\\
&= \bar{\psi}\left(i\not{D} - m\right)\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}, && (2.6)
\end{aligned}$$

with $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ the field strength tensor. Consequently, the constructed QED lagragian is invariant under a local $U(1)$ gauge transformation

$$\psi(x) \to e^{-i\alpha(x)}\psi(x), \qquad A_\mu(x) \to A_\mu(x) - \frac{1}{e}\partial_\mu\alpha(x). \tag{2.7}$$

2.3 QUANTUM CHROMODYNAMICS

The strong interactions are described by QCD which is a non-abelian gauge theory based on the $SU(3)$ group. The Lagrangian can be retrieved in a similar manner as for QED. In this context the quark field can be written as colour triplets $\bar{q}_k = (\bar{q}_{\text{red}}, \bar{q}_{\text{blue}}, \bar{q}_{\text{green}})$, which transform under a local gauge transformation as

$$q_k(x) \to e^{i\alpha_a\lambda_a/2}q_k(x), \qquad \alpha \in \mathbb{R}, \quad a \in \{1,\dots,8\}, \tag{2.8}$$

with k the flavour index, α_a a local phase and λ_a the generators of the $SU(3)$ group called *Gell-Mann matrices* [23] and a the colour index. They follow the commutation rule

$$[\lambda_a, \lambda_b] = i f^c_{ab} \lambda_c, \tag{2.9}$$

with f^c_{ab} the completely anti-symmetric structure constant. The coupling between quarks and gluons is introduced analogous to QED as a covariant derivative

$$D_\mu = \partial_\mu - i g_s \frac{\lambda_a}{2} G^a_\mu, \tag{2.10}$$

where G^a_μ are the eight gluon field strength tensors and g_s is the strong coupling strength which can also be expressed as the coupling constant of the strong interaction

$$\alpha_s = \frac{g_s^2}{4\pi}. \tag{2.11}$$

The final QCD Lagrangian then reads

$$\mathcal{L}_{QCD} = \sum_k \bar{q}_k \left(i \not{D} - m_k \right) q_k - \frac{1}{4} G^a_{\mu\nu} G^{a\mu\nu} \tag{2.12}$$

$$= \sum_k \bar{q}_k \left(i \not{\partial} - m_k \right) q_k + \frac{g_s}{2} \bar{q}_k \left(i \gamma^\mu G^a_\mu \lambda_a \right) q_k - \frac{1}{4} G^a_{\mu\nu} G^{a\mu\nu}, \tag{2.13}$$

with

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f^a_{\beta\gamma} G^\beta_\mu G^\gamma_\nu. \tag{2.14}$$

Due to the non-abelian structure of QCD, the term $\frac{1}{4} G^a_{\mu\nu} G^{a\mu\nu}$ is needed in order to maintain local gauge invariance of the Lagrangian and results in a self-coupling of the gluons illustrated in the two Feynman diagrams on the right in Figure 2.2.

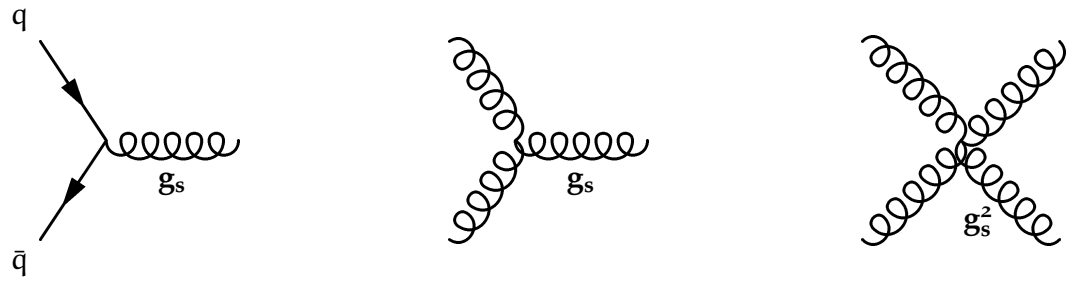


Figure 2.2.: Possible interactions of the gluon: the interaction with quarks (left), self-interaciton of three gluons (middle) and the self interaction of four gluons (right).

The self-coupling induces a different energy scaling behaviour compared to QED. The coupling constant α_s depends on the energy scale (renormalisation scale) μ_R^2 [24],

$$\alpha_s(\mu_R^2) = \frac{12\pi}{(33 - 2n_f) \ln(\mu_R^2 / \Lambda^2_{QCD})}, \tag{2.15}$$

with n_f the number of 'light' quark flavours (those whose mass is lower than μ) and Λ_{QCD} a non-perturbative constant indicating the scale at which the coupling diverges. The application of perturbation theory in order to calculate scattering amplitudes is only feasible for scales $\mu_R \gg \Lambda_{QCD}$, where $\alpha_s(\mu_R^2) \ll 1$. At low energy scales (larger distances) the effective coupling between two coloured particles increases and thus coloured objects cannot exist isolated and always form colourless bound states (hadrons), this effect is called colour confinement. At high energy scales (short distances) the coupling strength is decreasing, denoted as asymptotic freedom [21, 25].

2.4 ELECTROWEAK UNIFICATION

The electroweak unification was introduced in the 1960s by Glashow, Salam and Weinberg [26–28]. It unifies electromagnetic and weak interactions within one theory based on the non-abelian gauge group $SU(2)_L \otimes U(1)_Y$. The electroweak Lagrangian is composed of several parts, the gauge, fermion, Higgs and Yukawa part

$$\mathcal{L}_{EW} = \mathcal{L}_{gauge} + \mathcal{L}_{fermion} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yukawa}. \tag{2.16}$$

Fermions (leptons and quarks) are represented as left-handed doublets ψ_L and right-handed singlets ψ_R, classified with the quantum numbers of the weak isospin I ($SU(2)$ generators), I_3 being the third component of the isospin, and the weak hypercharge Y ($U(1)$ generators) as shown in Table 2.1 where the doublets have $I_3 = 1/2$ and the singlets $I_3 = 0$. The Gell-Mann–Nishijima formula [29] relates these two quantum numbers with the electric charge Q

$$Q = I_3 + \frac{Y}{2}. \tag{2.17}$$

Fields	Generations			Charges		
	I	II	III	I_3	Y	Q
ψ_L, L_L	$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L$	$\begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L$	$\begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L$	$+1/2$ $-1/2$	-1 -1	0 -1
ψ_R, ℓ_R	e_R	μ_R	τ_R	0	-2	-1
ψ_L, Q_L	$\begin{pmatrix} u \\ d \end{pmatrix}_L$	$\begin{pmatrix} c \\ s \end{pmatrix}_L$	$\begin{pmatrix} t \\ b \end{pmatrix}_L$	$+1/2$ $-1/2$	$1/3$ $1/3$	$+2/3$ $-1/3$
ψ_R, u_R	u_R	c_R	t_R	0	$4/3$	$+2/3$
ψ_R, d_R	d_R	s_R	b_R	0	$-2/3$	$-1/3$

Table 2.1.: The fermions are grouped into right-handed singlets and in left-handed doublets. The shown quantum numbers are the third component of the weak isospin (I_3), the weak hypercharge (Y) and the electric charge (Q). The field column contains the definition of the different fields.

2.4.1 *Gauge Term*

Each generalised charge is associated to a vector field, $W_\mu^{1,2,3}$ to $I_{1,2,3}$ and the singlet field B_μ to Y. The field strength tensors of the vector fields are given as

$$W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a + g_2 \epsilon_{abc} W_\mu^b W_\nu^c, \tag{2.18}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \tag{2.19}$$

with ϵ_{abc} the totally asymmetric Levi-Civita tensor and g_2 the gauge coupling constant for the non-abelian factor $SU(2)$. The Lagrangian for the gauge part then reads

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4} W_{\mu\nu}^a W^{\mu\nu,a} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}. \tag{2.20}$$

With this formulation, mass terms for the gauge bosons would violate the gauge invariance. However, it is possible to introduce these mass terms with the mechanism of spontaneous electroweak symmetry breaking described below.

2.4.2 *Fermion Term*

As mentioned above, the fermions have different chiralities (left- & right-handed) on which also their representation depends (see Table 2.1). The covariant derivatives, describing the fermion-gauge field interaction, are slightly different for the right-handed R and left-handed L case:

$$D_\mu^L = \partial_\mu - i g_2 \frac{\sigma_a}{2} W_\mu^a + i g_1 \frac{Y}{2} B_\mu, \tag{2.21}$$

$$D_\mu^R = \partial_\mu + i g_1 \frac{Y}{2} B_\mu, \tag{2.22}$$

with $\vec{\sigma} = (\sigma_1 \quad \sigma_2 \quad \sigma_3)^T$ being the vector of Pauli matrices satisfying $[\sigma_i, \sigma_j] = 2i\epsilon_{ijk}\sigma_k$ and g_1 the gauge coupling constant for $U(1)_Y$ gauge group. Then, the fermionic part of the lagrangian is denoted as

$$\mathcal{L}_{\text{fermion}} = \sum_j \bar{\psi}_L^j i\gamma^\mu D_\mu^L \psi_L^j + \sum_{j,\xi} \bar{\psi}_{R\xi}^j i\gamma^\mu D_\mu^R \psi_{R,\xi}^j, \tag{2.23}$$

with the generation index j running over the three lepton and quark generations and ξ the index for up-type and down-type fermions. Similar to the gauge term, the fermion masses are also not described here since they are mixing left- and right-handed fields which would break the gauge symmetry. The fermion masses are introduced via the Yukawa term in Section 2.4.4.

2.4.3 *Higgs Term*

The Higgs mechanism, introduced in the 1960s [30–35], spontaneously breaks the gauge symmetry $SU(2)_L \otimes U(1)_Y$ down to the $U(1)_{EM}$ symmetry by introducing an isospin doublet of complex scalar fields

$$\phi(x) = \begin{pmatrix} \varphi^+(x) \\ \varphi^0(x) \end{pmatrix} \tag{2.24}$$

with the covariant derivative

$$D_\mu = \partial_\mu - i g_2 \frac{\sigma_a}{2} W_\mu^a + i \frac{g_1}{2} B_\mu, \tag{2.25}$$

which introduces three- and four-point interactions between the gauge bosons and the Higgs field in the lagrange density

$$\mathcal{L}_{\text{Higgs}} = (D_\mu \phi)^\dagger D^\mu \phi - V(\phi). \tag{2.26}$$

The Higgs potential

$$V(\phi) = -\mu^2 \phi^\dagger \phi + \frac{\lambda}{4} \left(\phi^\dagger \phi\right)^2, \tag{2.27}$$

contains in the first term the Higgs mass after the EWSB with the constant μ^2 and in the second term the Higgs field self-interaction with the constant $\lambda > 0$ guaranteeing a lower bound of the potential.

2.4.4 *Yukawa Term*

The last term of the electroweak Lagrangian is the Yukawa term, introducing fermion mass terms

$$\mathcal{L}_{\text{Yukawa}} = -G^\ell_{ij} \bar{L}^i_L \phi \ell^j_R - G^d_{ij} \bar{Q}^i_L \phi d^j_R - G^u_{ij} \bar{Q}^i_L \phi^c u^j_R + \text{h.c.}, \tag{2.28}$$

with $\phi^c = i\sigma_2 \phi$ the charge conjugate, h.c. the hermitian conjugated term and the Yukawa couplings $G^{\ell,d,u}_{ij}$ described as 3×3-matrices.

2.5 THE HIGGS MECHANISM

The Higgs mechanism induces the spontaneous EWSB as described in Section 2.4.3. This mechanism allows mass terms in the electroweak Lagrangian for gauge bosons $W^\pm$ and Z as well as for fermions via the Yukawa couplings.

The lowest energy state of the potential in Equation (2.27) is denoted as the vacuum expectation value (VEV) v. By choosing the parameter $\mu^2 < 0$, the minimum is located at $\phi^\dagger \phi = 0$ with all real scalar fields having zero VEV as shown in the left plot of Figure 2.3. This configuration would preserve the $SU(2)_L \otimes U(1)_Y$ symmetry also at the minimum. However, with $\mu^2 > 0$ the minimum

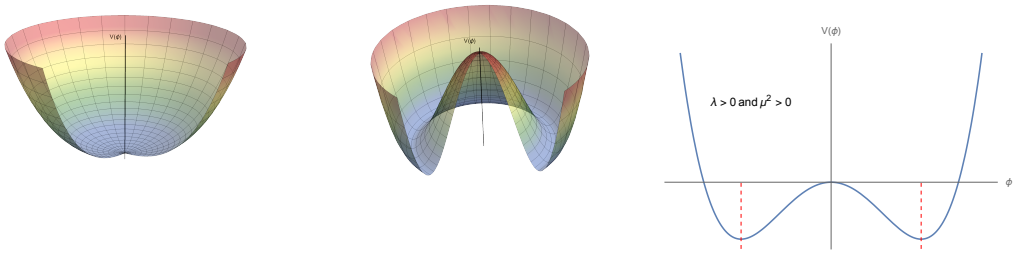


Figure 2.3.: Higgs potential following Equation (2.27). On the left with the parameter choice $\mu^2 < 0$ with only a minimum at 0. The middle plot shows the Higgs potential with $\mu^2 > 0$ as well as the right plot which is a projection of the middle plot, indicating the minima at $\phi^\dagger\phi = 2\mu^2/\lambda$.

is not located at 0, instead it is at $\phi^\dagger\phi = 2\mu^2/\lambda$, illustrated in the two right plots in Figure 2.3, resulting in the VEV

$$\langle\phi\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v \end{pmatrix} \quad \text{with} \quad v = \frac{2\mu}{\sqrt{\lambda}}, \tag{2.29}$$

picking ϕ as electrically neutral without loss of generality. In fact, the vacuum configuration $\langle\phi\rangle$ violates the $SU(2)_L \otimes U(1)_Y$ symmetry and spontaneously breaks it down to the electromagnetic (EM) subgroup $U(1)_{EM}$. Rewriting the potential in the unitary gauge one gets

$$\phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}, \tag{2.30}$$

where H is a scalar field depicting the Higgs boson and the potential can be written as

$$V = \mu^2 H^2 + \frac{\mu^2}{v}H^3 + \frac{\mu^2}{4v^2}H^4 = \frac{M_H^2}{2}H^2 + \frac{M_H^2}{2v}H^3 + \frac{M_H^2}{8v^2}H^4, \tag{2.31}$$

where the Higgs mass results in $M_H = \mu\sqrt{2}$. Also, the potential contains terms with triple and quartic self-interactions of the Higgs with couplings proportional to the Higgs mass M_H.
The kinematic term of Equation (2.26) describes the coupling of the Higgs field to the gauge fields, and can be expressed via (2.29) and (2.30) as

$$\begin{aligned}(D_\mu\phi)^\dagger D^\mu\phi =& \frac{g_2 v^2}{4} W^{+\mu}W^-_\mu \left(1 + \frac{H}{v}\right)^2 \\ &+ \frac{1}{2}\frac{\sqrt{g_1^2 + g_2^2}v^2}{4} Z^\mu Z_\mu \left(1 + \frac{H}{v}\right)^2 + \frac{1}{2}(\partial^\mu H)(\partial_\mu H),\end{aligned} \tag{2.32}$$

containing the physical fields

$$W^\pm_\mu = \frac{1}{\sqrt{2}}(W^1_\mu \mp i W^2_\mu), \tag{2.33}$$

depicting the massive charged gauge bosons $W^{\pm}$ and the neutral gauge bosons, given as

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} W^3_\mu \\ B_\mu \end{pmatrix}, \tag{2.34}$$

where the Z_μ field describes the massive Z-boson and the A_μ field the massless photon. As a consequence the masses are defined in terms of the couplings and the VEV as

$$M_W = \frac{1}{2} g_2 v, \quad M_Z = \frac{1}{2}\sqrt{g_1^2 + g_2^2}\, v. \tag{2.35}$$

The weak mixing angle θ_W is introduced in the rotation in Equation (2.34) and defined as

$$\cos\theta_W = \frac{g_2}{\sqrt{g_1^2 + g_2^2}} = \frac{M_W}{M_Z}, \tag{2.36}$$

which also allows to express the electric charge in terms of the gauge couplings as $e = g_2 \sin\theta_W$.

The fermions also acquire their masses via the interaction with the Higgs field described in the Yukawa term in Equation (2.28) which can be expressed in the unitary gauge:

$$\mathcal{L}_{\text{Yukawa}} = -\sum_f m_f \bar{\psi}_f \psi_f \left(1 + \frac{H}{v}\right), \tag{2.37}$$

with the mass defined as

$$m_f = G_f \frac{v}{\sqrt{2}} = y_f \frac{v}{\sqrt{2}}, \tag{2.38}$$

and y_f being the Yukawa coupling. The lepton mass matrix G^{ℓ}_{ij} has no off-diagonal entries since the lepton number is conserved for each generation in the SM and reduces with the assumption of massless neutrinos to

$$\begin{pmatrix} m_e & 0 & 0 \\ 0 & m_\mu & 0 \\ 0 & 0 & m_\tau \end{pmatrix}. \tag{2.39}$$

However, the quark mass matrix $G^{u,d}_{ij}$ has off-diagonal entries which can be diagonalised via four unitary matrices $V^{u,d}_{L,R}$ resulting in the mass eigenstates

$$\tilde{u}^i_{L,R} = (V^u_{L,R})_{ik} u^k_{L,R}, \quad \tilde{d}^i_{L,R} = (V^d_{L,R})_{ik} u^k_{L,R}. \tag{2.40}$$

By introducing the mass eigenstates in the Lagrange density, its structure is retained except for the flavour-changing quark interactions mediated by the charged vector bosons. The quark mixing matrix also denoted as Cabibbo-Kobayashi-Maskawa (CKM) matrix reads then as

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{CKM} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}, \tag{2.41}$$

with

$$V_L^u V_L^{d\dagger} \equiv V_{CKM}. \tag{2.42}$$

The diagonal elements of the CKM matrix are close to one, in particular the $|V_{tb}|$ term with a value of 0.999105 ± 0.000032 [20] which will be of special importance in the following chapters.

2.6 HIGGS BOSON PRODUCTION AND DECAY CHANNELS

In 2012, the ATLAS and CMS collaborations discovered a new particle compatible with the SM Higgs boson [6, 7], representing one of the major physics goals of the LHC. Further studies over the last years confirmed its properties being consistent with the SM predictions. Probing the EWSB sector is important for SM precision measurements as well as for investigations for physics beyond the SM. The Higgs boson is measured to have a mass of $(125.10 \pm 0.14)\,\text{GeV}$ [20].

The Feynman diagrams of the four major Higgs production modes at the LHC are shown in Figure 2.4. The most dominant production mode is the gluon fusion (ggF) process, comprising a fermion loop dominated by the heaviest fermion, the top quark, followed by the vector boson fusion (VBF) and the Higgs Strahlung (VH) which both probe the Higgs coupling to the heavy gauge bosons. The fourth most dominant Higgs production mode is the associated production with a top quark pair ($t\bar{t}H$) allowing a direct measurement of the top Yukawa coupling. The tH production mode is sensitive to both the magnitude and the sign of the Yukawa coupling. Due to negative inference of the Feynman diagrams, its cross-section is about one order of magnitude lower than the $t\bar{t}H$ cross-section. Figure 2.5 illustrates the cross-sections of important processes, including Higgs productions, at hadron colliders. Compared to many other SM processes, or even compared to the $t\bar{t}$ cross-section, the Higgs production modes are orders of magnitudes lower and therefore sophisticated techniques are necessary to extract the Higgs signals in analyses. For the $t\bar{t}H$ production mode, the main background is coming from $t\bar{t}$ processes whose cross-section is more than two orders of magnitudes larger.

Figure 2.6 shows the different decay modes for a Higgs with a mass of $m_H = 125\,\text{GeV}$. By far the largest branching ratio is the $H \to b\bar{b}$ decay mode with 58.2%. The second most probable decay mode with 21.4% is the decay to WW^*. Even though both of these decay modes have a large branching ratio, they are challenging to access due to difficulties to distinguish them from

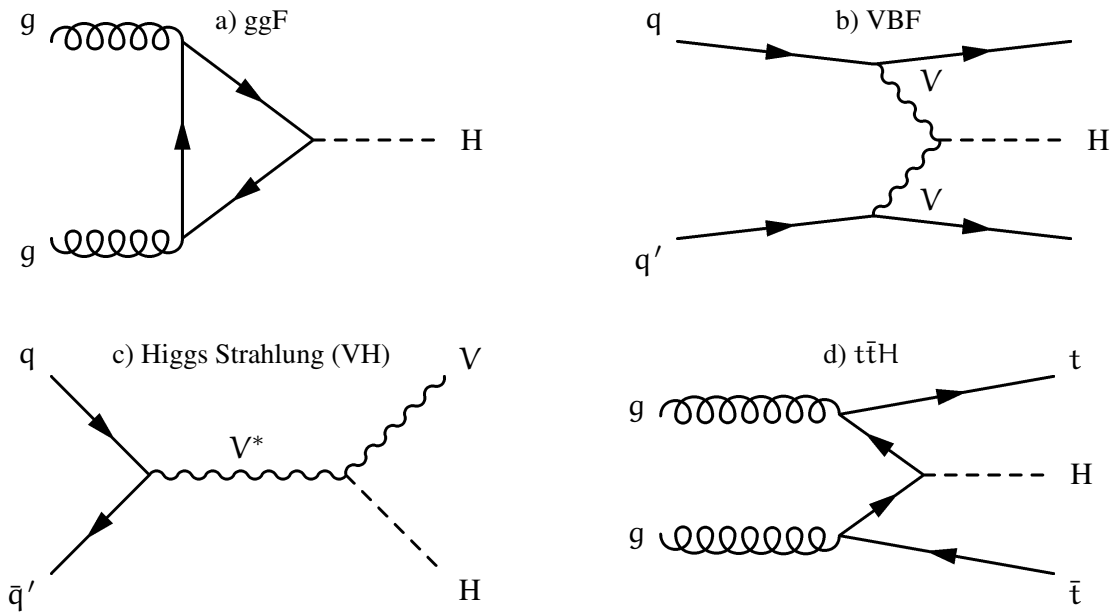


Figure 2.4.: Feynman diagrams of the four major Higgs production modes at the LHC: (a) gluon fusion (ggF), (b) vector boson fusion (VBF), (c) Higgs Strahlung (VH) and (d) associated production with a top quark pair ($t\bar{t}H$).

background processes. The so-called *Golden Channels*, having the cleanest final state signatures, are the $\gamma\gamma$ and the $ZZ^* \to 4\ell$ decay modes. Despite their very low branching ratios: 0.2% and 2.6%, respectively, they are easier to access in physics analyses.

The $t\bar{t}H$ production combined with the $b\bar{b}$ decay mode is analysed in this thesis in Part IV and a more detailed description and motivation for this analysis can be found in Chapter 5.

2.7 THE TOP QUARK

The top quark, with its mass of $(172.69 \pm 0.25(\text{stat.}) \pm 0.41(\text{syst.}))$ GeV [19], is the heaviest particle in the SM. Due to its large fermion mass, also its Yukawa coupling to the Higgs is the strongest with $y_t \simeq 1$ shown in Equation (2.38). This strong coupling induces the dominance in the loop contributions in Higgs productions (ggF) and decays ($H \to \gamma\gamma$). Since these loop effects involve heavy virtual particles they are also sensitive to physics beyond the SM.
Another implication of the high mass is that the top quark can decay into a W-boson and a b-quark. In fact, the top quark almost exclusively decays via this mode indicated by the value of the CKM matrix element $|V_{tb}|$ being close to 1, as already mentioned above. Its decay width of $1.42^{+0.19}_{-0.15}$ GeV [20] leads to a mean lifetime of $\sim 5 \cdot 10^{-25}$ s which is uniquely shorter than the time scale of hadronisation processes, making the top quark decay as an almost free particle. These properties create a very recognisable decay signature in the detector.

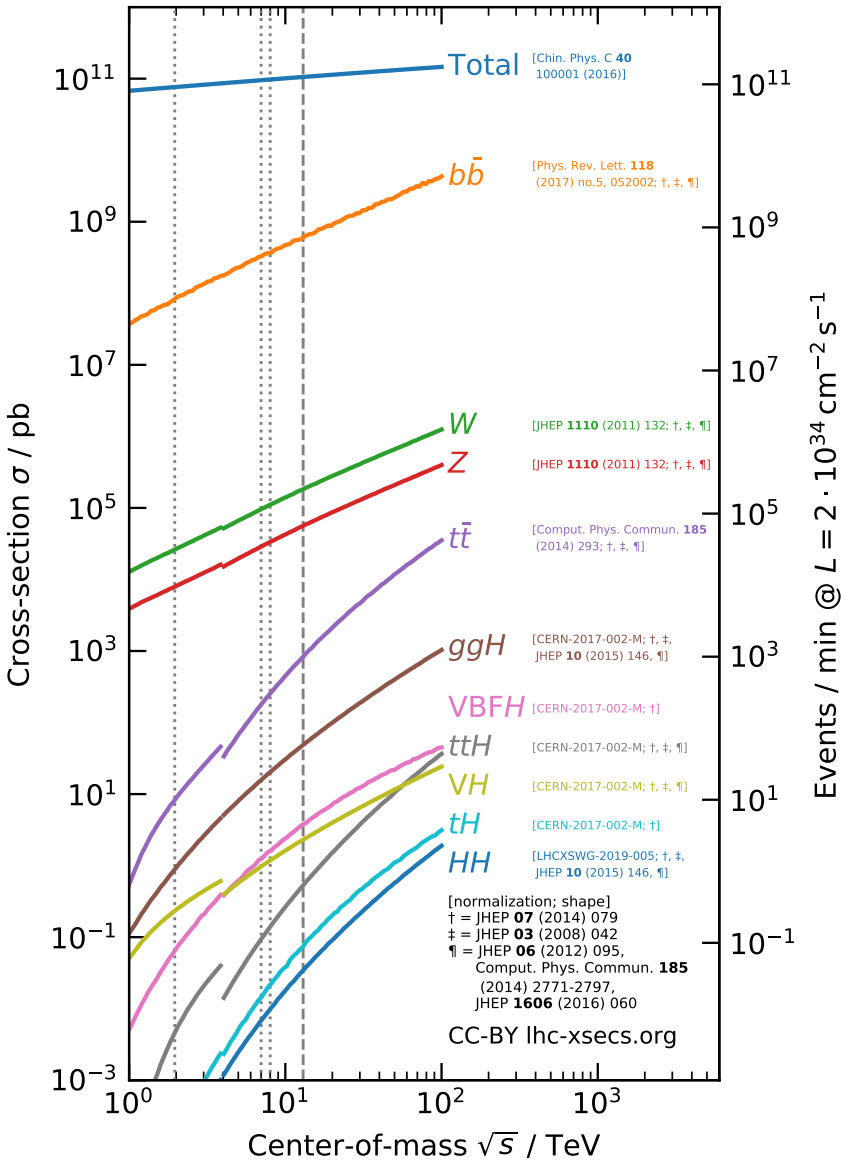


Figure 2.5.: Cross-section of different processes as a function of the centre-of-mass energy in proton-proton collisions above $\sqrt{s} = 4$ TeV and below for proton-antiproton collisions. The dashed line indicates a centre-of-mass energy of $\sqrt{s} = 13$ TeV [36].

At hadron colliders, top quarks are predominantly produced in $t\bar{t}$ pairs. The decay modes are typically classified according to the decay of the two involved W-bosons. The pie chart in Figure 2.7 shows the individual decay rates of $t\bar{t}$. The all-hadronic final state has the largest branching ratio with 45.7% where both W-bosons decay hadronically. Followed by the lepton+jets (semi-leptonic) decay channel (43.8%) in which one W-boson decays hadronically and one leptonically. The smallest fraction of 10.5% is allocated to the dileptonic mode (both W-bosons decay leptonically).

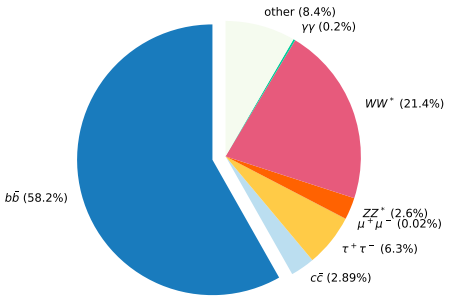


Figure 2.6.: Branching ratio of the Higgs decay with a Higgs mass of $m_H = 125\,\text{GeV}$. The numbers are taken from [20].

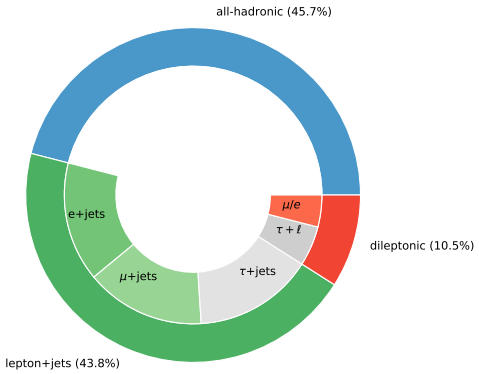


Figure 2.7.: Pie chart showing the individual $t\bar{t}$ decay modes. The label e/μ comprises the dileptonic decay modes $\mu\mu$, μe, ee and $\tau + \ell$ the decay modes $\tau + \tau/\mu/e$. Also hadronically decaying τs are included in the leptonic decay modes. The numbers are taken from [20].

2.8 LIMITATIONS OF THE STANDARD MODEL

Figure 2.8 shows different SM cross-section measurements in comparison to their theoretical predictions. They are all in a good agreement with theoretical expectations. However, despite the success of the SM, it cannot describe all experimental observations sufficiently. Several astrophysical observations saw that only a small fraction (around 17%) of the matter in the universe is made of the SM components. Thus, another source of matter has to be present making up 25% of the universe, corresponding to 83% of all matter in the universe, denoted as dark matter [37] which could consist of weakly interacting massive particles. There are, however, no suitable candidates within the SM. Also, astrophysical observations show a stronger expansion of the universe than predicted by cosmological theories [38]. This phenomenon appears to be caused by a non-detectable energy, called dark energy (being the last missing 70%). Moreover, neutrino experiments observed neutrino oscillations [39, 40], indicating that neutrinos do have a mass. However, they are assumed to be massless in the SM because at the time of the formulation of the SM their mass was not observed. If neutrinos would have a Dirac mass, it could be easily added to the SM. However, there are no indications yet for a right-handed neutrino and it is therefore not evident that neutrinos have a Dirac mass. The fact that the universe consists of far more matter than anti-matter reveals another unexplained physics behaviour.

Furthermore, there are theory implications suggesting physics beyond the SM e.g. the unification of the electroweak and strong force as well as the missing gravity in the theory.

Several of these beyond SM physics processes predict new particles that could be found in particle collider experiments. There is, however, no evidence yet for these beyond SM particles. The search for these new particles is also one of the main goals of the LHC in the future.

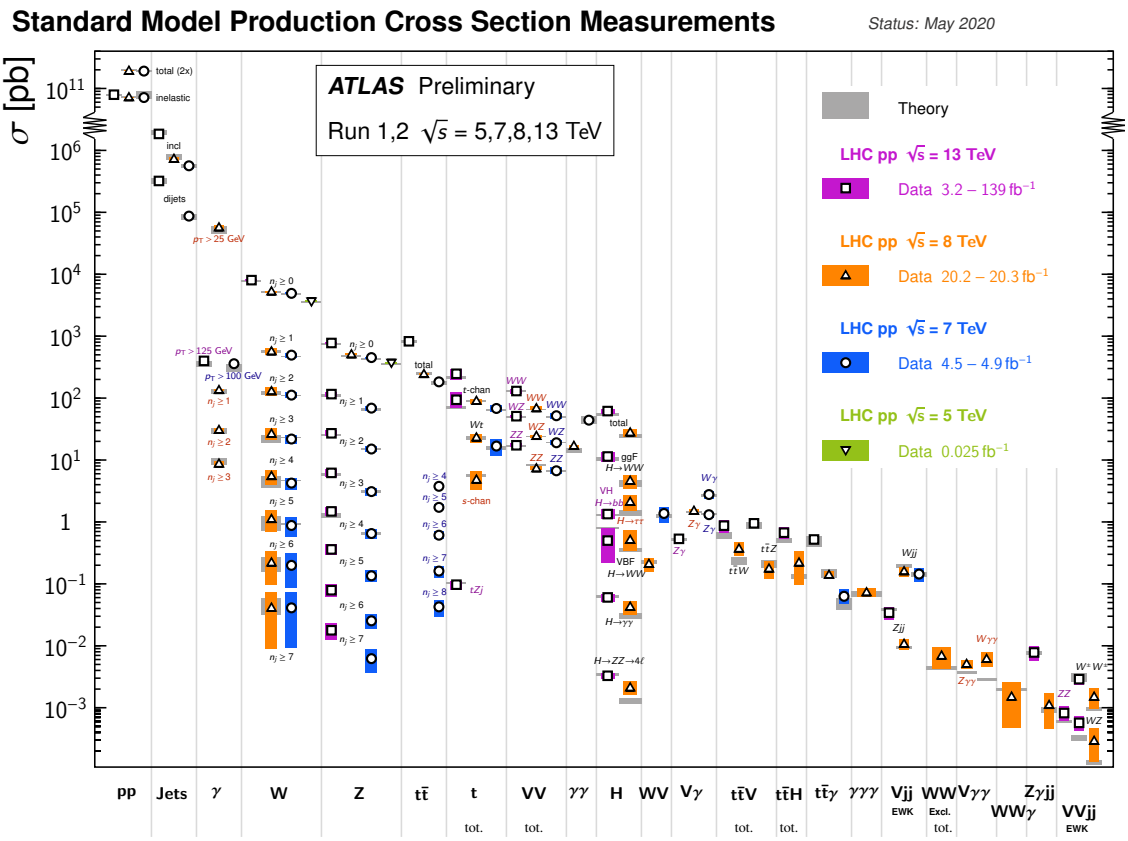


Figure 2.8.: A summary of different Standard Model production cross-section measurements (total and fiducial) performed with the ATLAS experiment. They are corrected for leptonic branching fractions and compared to the corresponding theoretical expectations [41].

3

THE ATLAS EXPERIMENT AT THE LARGE HADRON COLLIDER

Particle physics explores scales in the order of $< 10^{-15}$ m which requires large and complex machines to explore physics at the TeV scale. Such a machine is the LHC [42–44]. Experiments in these dimensions are only possible within international collaborations. Thus, the Organisation européenne pour la recherche nucléaire (CERN) is an optimal environment to host such experiments.
The work in this thesis is based on the $\sqrt{s} = 13$ TeV proton-proton collision data collected with the ATLAS experiment at the LHC.
This chapter gives a short overview of the LHC and describes the different sub-detector systems of the ATLAS detector including its magnetic and trigger system.

3.1 THE LARGE HADRON COLLIDER

The largest and most powerful hadron accelerator ever built, the LHC, is situated near Geneva, Switzerland, with a circumference of 27 km. It consists of 1232 super-conducting dipole magnets designed to reach a centre-of-mass energy of $\sqrt{s} = 14$ TeV and currently, during RUN II, operating at a centre-of-mass energy of $\sqrt{s} = 13$ TeV ($\sqrt{s} = 7/8$ TeV in RUN I).
Apart from protons, the LHC can also operate with heavy ions. This dissertation is, however, entirely based on proton-proton collisions.

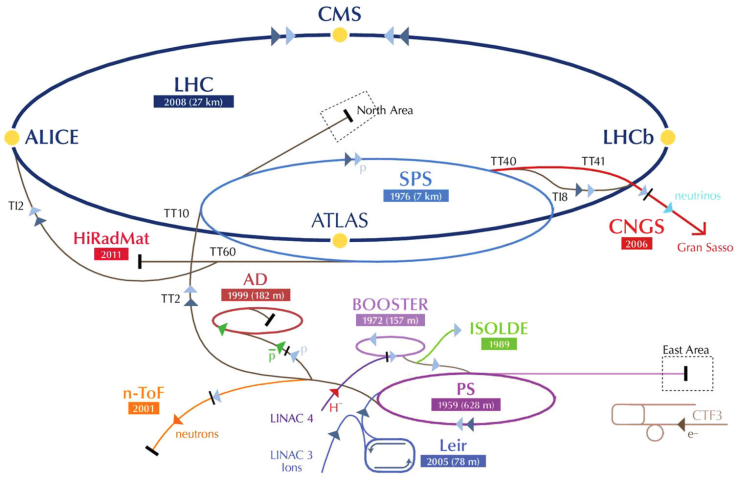


Figure 3.1.: Current CERN accelerator complex with Large Hadron Collider (LHC), Super Proton Synchrotron (SPS), Proton Synchrotron (PS), Booster, Antiproton Decelerator (AD), Low Energy Ion Ring (LEIR), Linear Accelerators (LINAC 3 & 4), CLIC Test Facility (CTF3), CERN to Gran Sasso (CNGS), Isotopes Separation on Line (ISOLDE), neutrons Time of Flight (n-ToF) and High-Radiation to Materials High-Radiation to Materials (HiRadMat) Facility (HiRadMat) [45].

As shown in Figure 3.1, the LHC is part of a large accelerator complex being the final element in this accelerator chain. The pre-acceleration of protons is performed in several steps beginning with a linear accelerator (LINAC 2 in RUN II and from RUN III on LINAC 4) where the protons are injected as hydrogen gas and are accelerated to 50 MeV followed by the Booster where they reach 1.4 GeV. The next step is the Proton Synchrotron (PS) which accelerates the protons to 25 GeV and the final pre-acceleration is done in the Super Proton Synchrotron (SPS), which injects the protons into the LHC with an energy of 450 GeV.
The LHC is using eight radiofrequency cavities per beam operating at 400 MHz to accelerate the protons which are brought to collision at four different points, each hosting an experiment. Two of them are the multi-purpose experiments ATLAS [46] and CMS [47] pursuing a wide range of physics, comprising SM precision measurements as well as searches for beyond the SM phenomena such as Supersymmetry, Exotic particles or Dark Matter searches. These two collaborations are the largest ones at CERN comprising around 3000 scientist each [48, 49]. The LHCb experiment [50] is specialised in exploring hadrons containing b- or c-quarks especially investigating CP-violating processes. The ALICE experiment [51] is the only experiment fully focusing on heavy-ion collisions[1] and therefore particularly specialised on QCD physics.

Apart from the centre-of-mass energy, the instantaneous luminosity is a main characteristic of a particle collider. For a circular collider with a Gaussian-shaped effective beam area $A = 4\pi\sigma_x\sigma_y$, where $\sigma_{x,y}$ are the Gaussian beamwidths in the x- and y-direction[2], the instantaneous luminosity can be written as

$$\mathcal{L} = f_{rev} \cdot \frac{N_1 \cdot N_2}{4\pi\sigma_x\sigma_y} F(\theta_c), \tag{3.1}$$

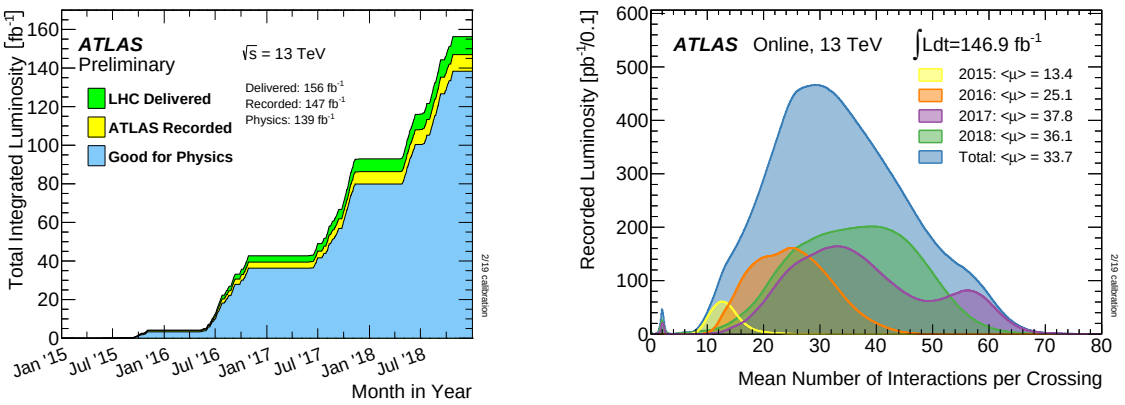
with f_{rev} the revolution frequency, $N_{1,2}$ the total number of protons in each beam and $F(\theta_c)$ scoping for geometric effects caused by the crossing angle θ_c of the two beams since the beams of the LHC are not colliding exactly head-on. The revolution frequency of the LHC is $f_{rev} = c/27\,\mathrm{km} = 11\,\mathrm{kHz}$ with nominally 2808 proton bunches. The protons are organised in bunches which can contain up to 10^{11} protons. This leads to an instantaneous luminosity of $\mathcal{O}(10^{34}\,\mathrm{cm^{-2}s^{-1}})$. Consequently, the produced events by the LHC, which are the number of collisions, can be retrieved by integrating the instantaneous luminosity $\mathcal{L}$

$$N = \sigma \cdot \int \mathcal{L}\,dt = \sigma \cdot \mathscr{L}, \tag{3.2}$$

where σ is the event cross-section for a given physics process. The evolution of the integrated luminosity $\mathscr{L}$ of the LHC RUN II delivered to the ATLAS experiment is shown in Figure 3.2 (a) yielding in total an integrated luminosity of $\mathscr{L} = 139\,\mathrm{fb^{-1}}$ good for physics which means that this data can be used in analyses. In fact, this data will be used in this thesis.

Due to the large number of protons within a bunch, more than one collision of interest can occur within a bunch crossing, which is called in-time pile-up. In addition, there are interactions coming

1 ALICE uses proton-proton collisions only for calibration purposes.
2 The coordinate system is defined in Section 3.2.



(a) (b)

Figure 3.2.: The development of the cumulative luminosity collected by the ATLAS experiment during RUN II (a) and the mean number of interactions per bunch crossing splitted into the different data taking periods (b) [52].

from neighbouring bunch crossings which cannot be resolved fast enough by the detector. These are called out-of-time pile-up. The mean interactions per crossing is a measure to quantify the pile-up. Clearly, the suppression of pile-up effects is quite a challenge for physics analyses. The distribution of the mean interactions per crossing is shown in Figure 3.2 (b). The pile-up profiles differ for each data taking period since the instantaneous luminosity (corresponding to the slope in Fig. 3.2 (a)) constantly increased, reaching a plateau in 2017/2018, and therefore more interactions per bunch crossing occur indicated in the plot legend as the average number of interactions per crossing $\langle\mu\rangle$.

3.2 THE ATLAS DETECTOR

The ATLAS (**A** **T**oroidal **L**HC **A**pparatu**S**) detector [46] is a multi-purpose particle detector, used to study a wide range of physics topics. It is situated 100 m below ground at Point-1 of the LHC. With its large dimensions of 25 m in diameter, a length of 44 m and a weight of 7000 t, it is the largest detector located at a collider. The detector has a cylindrical structure composed of several detector layers with an almost full solid angle coverage of 4π schematically illustrated in Figure 3.3.

All detector systems are designed such that they provide optimal performance for the different physics analyses. Hence it is important that the detector satisfies the following criteria: fast electronics for the readout, high granularity, good object reconstruction efficiency and resolution.

Coordinate System

In order to describe the particles recorded with the ATLAS detector, a right-handed coordinate system is used as illustrated in Figure 3.4 with its origin at the centre of the detector which is also the nominal interaction point. The z-axis is defined along the beamline while the y-axis points towards the surface

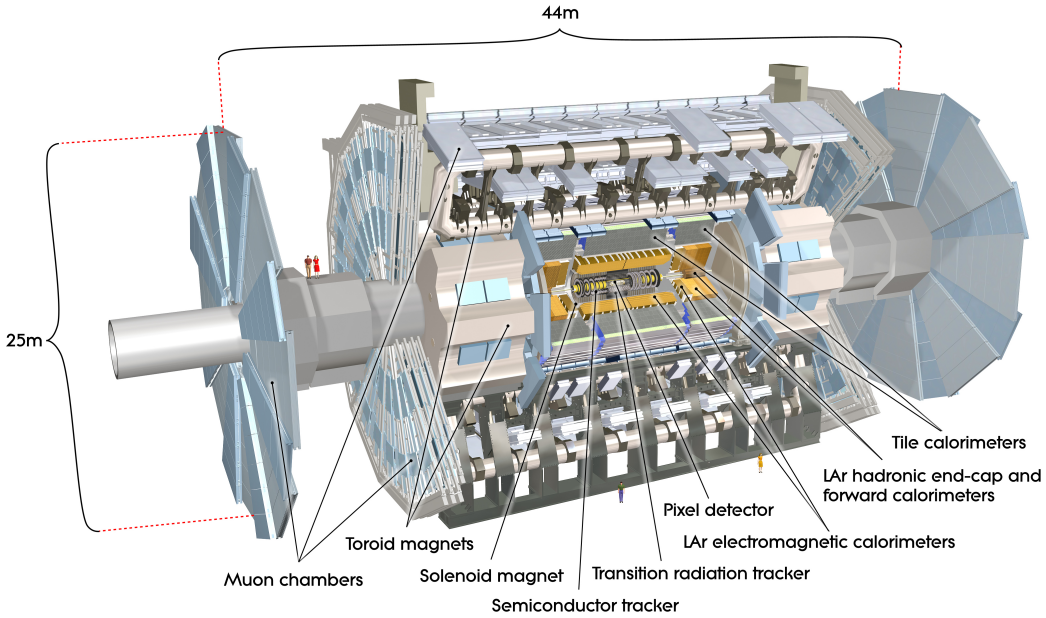


Figure 3.3.: Schematic overview of the ATLAS detector [53].

and the x-axis in the direction of the centre of the LHC. To describe the physics objects within the detector, spherical coordinates are the best choice where the polar angle θ is the angle between the z-axis and the direction considered while the azimuthal angle ϕ is measured in the x-y plane with respect to the x-axis. In fact, the polar angle is usually stated as the pseudorapidity η which is a high-energy approximation of the rapidity y:

$$y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right) \xrightarrow{m \ll E} -\ln\left(\tan\frac{\theta}{2}\right) = \eta, \tag{3.3}$$

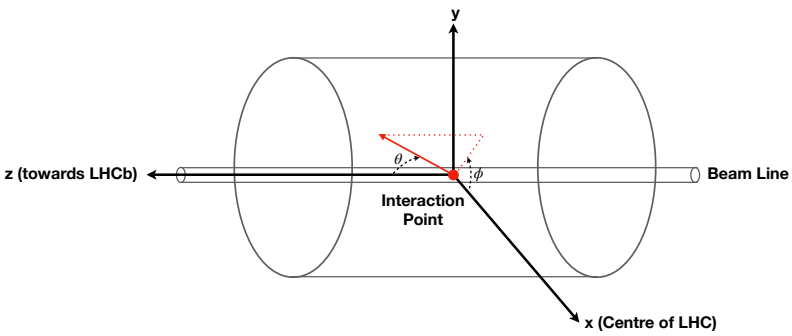


Figure 3.4.: Coordinate system of the ATLAS detector.

with E the energy and p_z the z component of the momentum vector. The advantage of the pseudorapidity is that the difference $\Delta\eta$ is invariant under Lorentz transformation. Thus the distances of two objects are calculated via

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}. \tag{3.4}$$

In some cases, the distance is also calculated using the rapidity instead of the pseudorapidity which will be denoted as ΔR_y. Furthermore, the x-y-plane defines the transverse plane where the transverse momentum is an important quantity denoted as

$$\vec{p}_T = \begin{pmatrix} p_x \\ p_y \end{pmatrix} \qquad p_T = \sqrt{(p_x)^2 + (p_y)^2}. \tag{3.5}$$

Since protons are composite particles and the centre-of-mass system moves with respect to the lab system, only the transverse momentum component of the initial partons is known to be zero in the lab system at the time of the collision.

3.2.1 *Inner Detector*

The innermost detector system is the Inner Detector (ID) [54, 55] enclosing the beam pipe shown in Figure 3.5. This detector system provides precise tracking information of charged particles. It is structured into three sub-detectors: the pixel detector, the semiconductor tracker (SCT) and the transition radiation tracker (TRT).

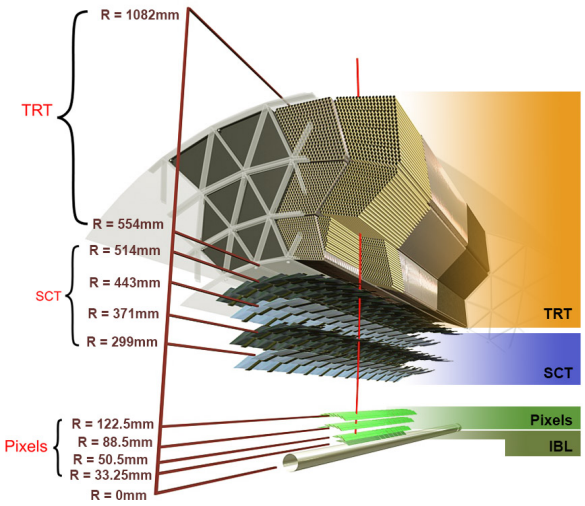


Figure 3.5.: Overview of the inner detector of the ATLAS experiment divided into three sub-detectors (Pixels, SCT and TRT) [56].

Pixel Detector

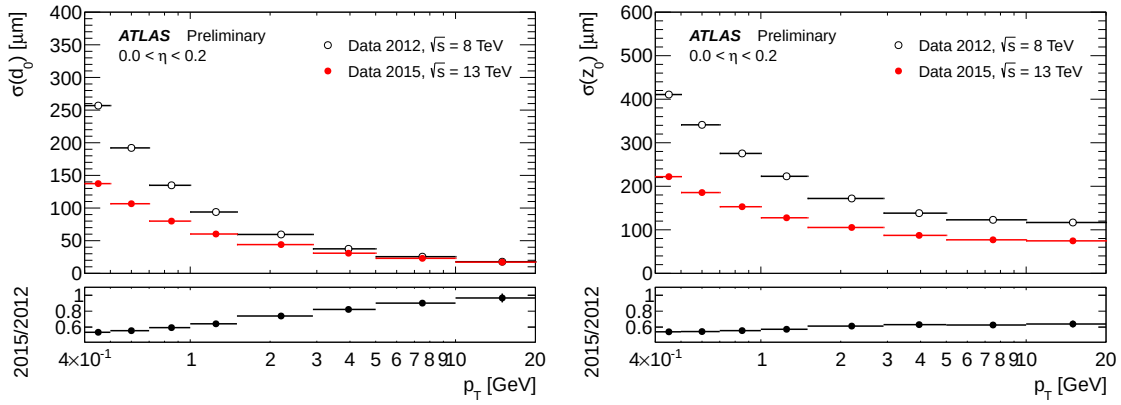
The first part of the ID is the silicon pixel detector comprising 4 cylindrical layers and 2 end-caps with 3 disc layers each. The layers are located between 33.25 mm to 122.5 mm around the beam pipe

with a coverage of $|\eta| < 2.5$. The pixel detector is especially important for the track reconstruction, the primary vertex reconstruction as well as for secondary vertex finding.

The insertable b-layer (IBL) [57] is the innermost layer, installed in-between RUN I and RUN II, having the highest granularity with a pixel size of 50 µm in R-ϕ-direction and 250 µm in z-direction. In particular, the IBL plays a crucial role for b-tagging. Figure 3.6 compares the resolution of the transverse and longitudinal impact parameters, which are important variables for b-tagging, with and without the IBL installed in ATLAS. All flavour-tagging studies in Part III of this thesis are performed for RUN II and therefore include the IBL.

Furthermore, the three remaining layers have a pixel size of 50 µm in the R-ϕ-direction and 400 µm in the z-direction. This gives an expected hit resolution of 8 µm & 10 µm in the direction of R-ϕ and 40 µm & 115 µm in the z-direction for the IBL and the three remaining layers, respectively.

In total, the pixel detector contains 86 M pixels providing a good spatial resolution which make up around 50% of all ATLAS readout channels.



(a) (b)

Figure 3.6.: Unfolded transverse impact parameter resolution (a) and longitudinal impact parameter resolution (b) measured from data in 2015 (red) with $\sqrt{s} = 13\,\text{TeV}$ with the Inner Detector including the IBL, as a function of pT, compared to that measured from data in 2012, $\sqrt{s} = 8\,\text{TeV}$ (without the IBL) [58].

Semiconductor Tracker

The semiconductor tracker (SCT) is a silicon strip detector comprising 4 double layers in the barrel region and nine planar end-cap discs on each side. The strips have a size of 80 µm$\times$12 cm and cover a region up to $|\eta| < 2.5$. The two layers within one layer-module are rotated by a stereo angle of 40 mrad. In general, the semiconductor-based detectors in ATLAS operate at a temperature between -10 °C and -5 °C to suppress different types of electronic noise.

Overall the SCT has a resolution of 17 µm in the R-ϕ direction and 580 µm in the z-direction with a total of 6.3 M readout channels.

Transition Radiation Tracker

The outermost part of the ID is the transition radiation tracker (TRT). In contrast to the other ID detectors, the TRT is not based on silicon but is a gaseous detector system. It consists of around 300k straw tubes with a diameter of 4 mm filled with a gas mixtures[3] of Xe (70 %), CO_2 (27 %) and O_2 (3 %) and a gold-plated tungsten wire in the tube centre with a potential different to the tube surface of 1.5 kV. The straws have a length of 144 cm in the barrel region and 37 cm in the end cap. The single hit resolution is 120 μm in the barrel and 130 μm in the end-cap. In fact, the TRT provides besides the tracking information also a particle ID. This is achieved with emitted transition radiation at the material boundaries since the straws are interleaved with polypropylene. Especially electrons can be distinguished from charged pions due to their larger transition radiation.
However, the TRT will be replaced for the High Luminosity LHC (HL-LHC) by a new, fully silicon-based Inner Tracker (ITk) [60] which will in fact replace the full ID.

3.2.2 *Calorimeter System*

The calorimeter system is responsible for the precise measurement of the particle energies by absorbing them as well as measuring the shower properties to allow for particle identification. Showers are cascades of secondary particles which are formed when a highly energetic particle interacts with dense material. ATLAS uses sampling calorimeters which consist of alternating layers of active material (liquid argon & plastic scintillators) and passive detector material (copper, iron, tungsten and lead). While the active material measures the energy deposit of the particles, the passive material induces the shower creation. The calorimeter system is composed of two main sub-systems, the electromagnetic [61, 62] and the hadronic calorimeter [63, 64] as shown in Figure 3.7. The calorimeter covers an η range up to a far forward region of $|\eta| < 4.9$.

Electromagnetic Calorimeter

The EM calorimeter encloses the ID and is a high granularity sampling calorimeter based on liquid argon (LAr) technology with absorber plates made out of lead. To provide full coverage in ϕ, the EM calorimeter has an accordion-shaped structure where the active material is placed in the gaps between the lead absorber plates and the Kapton electrodes. The detector operates at -183 °C with a total of 170k readout channels. The barrel region of the EM calorimeter, consisting of two parts with a 4 mm gap between them and a length of 3.2 m each, covers $|\eta| < 1.475$ with its granularity of $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$ in the second layer (middle-layer) and the two end caps cover $|\eta| < 3.2$ with a slightly coarser granularity.
In general, the absorption power at high energies of a calorimeter can be quantified in a material-independent way by using the radiation length X_0 of its medium. It is defined as the distance over which the particle energy is reduced via radiation losses by a factor $1/e$. The thickness of the barrel

3 In RUN II, a second gas mixture of Ar (70 %), CO_2 (27 %) and O_2 (3 %) was used for straw tubes belonging to modules with large gas leaks [59].

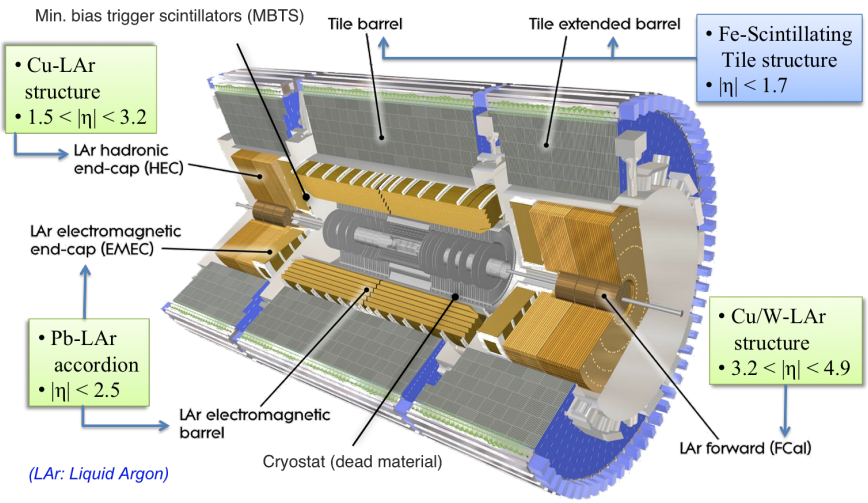


Figure 3.7.: Cut-away view of the calorimeter system of the ATLAS experiment [46] (and adpated by [65]).

region, given in terms of the radiation length, is 22 X_0 and 24 X_0 for the end-caps. Moreover, the intrinsic energy resolution of the EM calorimeter is $\sigma_E/E = 10\%/\sqrt{E} \oplus 0.7\%$ [62, p. 9], where the first term is the stochastic part and the second one the constant part.

Hadronic Calorimeter

The second calorimeter system is the hadronic calorimeter located around the EM calorimeter consisting of three components with two different detector technologies providing roughly 19,000 readout channels.

Firstly, the tile calorimeter is made out of alternating layers of steel as absorber material and scintillator plastic tiles as active material being read out via photomultiplier tubes. Out of its three layers, the first two have the highest granularity with $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$. The barrel part of the tile calorimeter covers a region with $|\eta| < 1.0$ and the two extended barrels a range of $0.8 < |\eta| < 1.7$. The resolution of the tile calorimeter is $\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$ [66, p. 3].

Secondly, the end-cap calorimeters, which are directly outside the EM calorimeter, and the forward calorimeter are based on the LAr technology. The end-caps use copper as passive material and cover a region of $1.5 < |\eta| < 3.2$ with their highest granularity of 0.1×0.1 $(\Delta\eta \times \Delta\phi)$ within $|\eta| < 2.5$. Also, the first layer of the forward calorimeter uses copper as absorber scoping for EM activities. The other two layers make use of tungsten as absorber which is better suitable for hadronic measurements. In total the forward calorimeter covers a region of $3.2 < |\eta| < 4.9$. The overall resolution of the LAr based hadronic calorimeters is $\sigma_E/E = 100\%/\sqrt{E} \oplus 10\%$ [67, p. 2].

3.2.3 *Muon Spectrometer*

Muons mostly traverse the detector without losing energy. Hence they are identified in the muon spectrometer (MS) [68] which is the outermost detector system of ATLAS with a distance to the

beam of 5-10 m (see Figure 3.8). It consists of four detector systems grouped into trigger and precision muon tracking chambers. In total the MS has more than one million readout channels and is embedded in three superconducting toroidal magnets (one in the barrel and one at each end-cap), providing a magnetic field in ϕ-direction (typically perpendicular to the muon trajectory). The muon system is not entirely symmetric in ϕ due to some gaps for detector services and support structure (detector feet). The momentum resolution of the MS is around 10% for 1 TeV muons and around 3% for 10-200 GeV muons.

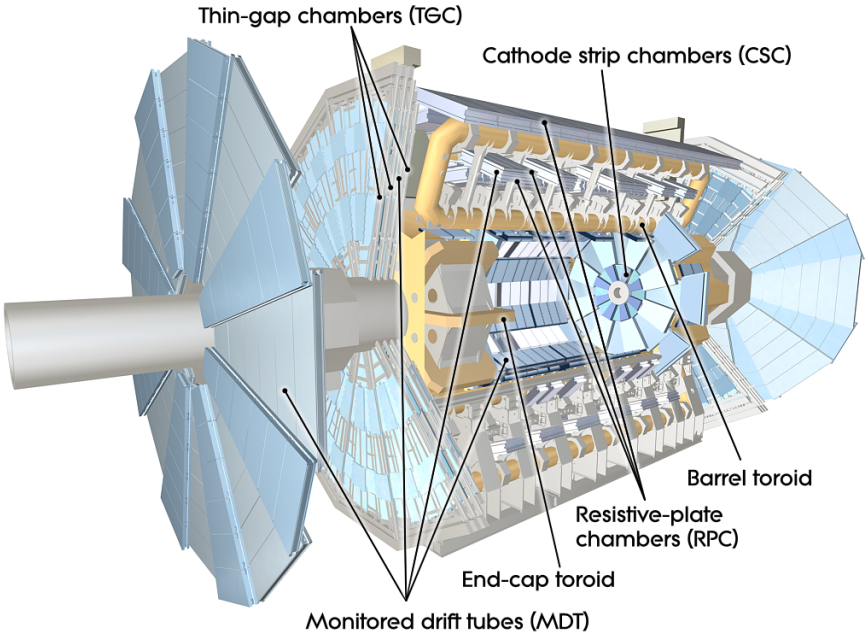


Figure 3.8.: Layout of the ATLAS muon spectrometer [69].

Muon Trigger Chambers

The muon trigger chambers are designed for a fast readout to provide trigger information. In the barrel region with $|\eta| < 1.05$, three layers of resistive plate chambers (RPCs) are used. The RPCs are made out of parallel plates with a high resistivity and a potential difference between them where the gap is filled with a gas mixture (94.7% $C_2H_2F_4$, 5% Iso-C_4H_{10}, 0.3% SF_6). Besides the trigger information the RPCs also provide an $\eta - \phi$ measurement with a spatial resolution of 10 mm.
In the end-caps ($1.05 < |\eta| < 2.4$) multi-wire chambers filled with a gas mixture of 55% CO_2 and 45% n-C_5H_{12} are used. These are called thin gap chambers (TGCs). They use graphite-coated cathodes and the wires are separated by 1.8 mm. Apart from the trigger information, the TGCs provide ϕ information with a resolution of 5 mm.

Precision Muon Tracking Chambers

While the muon trigger chambers are always read out, the precision muon tracking chambers are only read out when a trigger decision was made since the detector technology is slower but provides a high resolution and precision tracking information.

The monitored drift tubes (MDTs) are installed in the barrel and end-cap region covering $|\eta| < 2.7$. MDTs are aluminium drift tubes with a diameter of 3 cm filled with an Ar/CO_2 (93/7%) gas mixture and a wire in their centres. Typically, each chamber contains 3-8 layers of drift tubes resulting in a spatial resolution of 35 μm [46, p. 165].
Cathod strip chambers (CSCs) are installed in the forward region ($2.0 < |\eta| < 2.7$) and are proportional multi-wire chambers (as the TGCs), providing a radial resolution of 40 μm and a resolution in ϕ of 5 mm [46, p. 165].

3.2.4 *Magnet System*

Besides the detector systems, the magnet system is of major importance to allow momenta and charge measurements. It bends the trajectory of charged particles via the Lorentz force depending on their momentum and charge and consists of two sub-systems.
Firstly, the central solenoid magnet located between the ID and the calorimetry generates a constant magnetic field of 2 T. The superconducting magnet made out of NbTi is cooled via liquid helium to a temperature of 1.8 K.
Secondly, the vast toroidal magnet system embedded in the MS comprising one barrel toroid and two end-cap toroids with eight coils each. The toroidal magnets deliver an inhomogeneous magnetic field of roughly 0.5 T and 1 T in the central and end-cap regions, respectively.

3.2.5 *Trigger System and Data Acquisition*

In order to handle the high event rates, which are expected to be 40 MHz for ATLAS corresponding to more than 40 TB/s of data, a trigger system is required to reduce the amount of data to be recorded without losing important information. Since RUN II the trigger system is structured into two parts, the Level-1 (L1) hardware trigger and the software-based high level trigger (HLT) [70, 71] sketched in Figure 3.9.
The L1 trigger uses information from the RPCs, TGCs and the calorimeter to identify high p_T electrons, muons, photons, jets and high missing transverse momentum. It has a very fast latency of 2.5 μs and reduces the rate to 100 kHz. The L1 trigger identifies regions of interest (RoIs) in η and ϕ and passes this information to the HLT.
The HLT is fully software based and uses the full detector information within the RoIs to reduce the event rate down to approximately 1 kHz with a latency of 200 ms.
Afterwards, the data is transferred to a computing centre for further processing and storage.

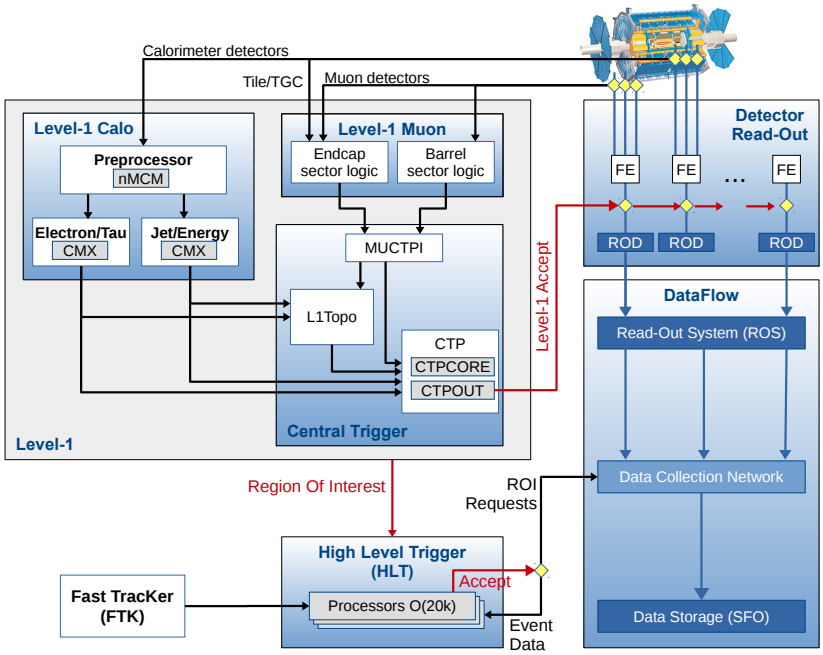


Figure 3.9.: Schematic view of the ATLAS trigger and data acquisition system in Run II [71].

4

PHYSICS SIMULATION AT HADRON COLLIDERS

As described in Chapter 2, most elementary particles are unstable and have short lifetimes. To investigate their properties in a controlled environment particle colliders are used. In particular, hadron colliders allow to reach higher centre-of-mass energies than e^+-e^--colliders. However, the initial state in hadron collisions is not well defined since hadrons are composite particles and carry only a fraction of the hadron momentum.
This chapter will give an introduction to the event simulation with Parton Distribution Functions (PDFs) and Monte Carlo (MC) generators and a quick discourse on the detector simulation.

4.1 EVENT SIMULATION

In order to analyse the data from collider experiments, it is important to have a reliable simulation of the underlying processes. Simulation is the basis for each physics analysis performed at collider experiments.
In particle physics, the simulation is based on MC generators which are a stochastic tool incorporating theoretical predictions, which are well-suited to describe the statistical processes.
The cross-section of a hard scattering-event at hadron colliders $\sigma_{A,B\to X}$ can be factorised into two components using the factorisation theorem [72]. The PDFs f_a^A and f_a^B describe the colliding partons a, b which are contained in the hadrons A, B while the cross-section of the hard scattering itself, $\hat{\sigma}_{a,b}$, can be usually calculated with perturbation theory. The cross-section can be written as

$$\sigma_{A,B\to X} = \sum_{a,b} \int_0^1 dx_1 dx_2 f_a^A(x_1, \mu_F^2) f_b^B(x_2, \mu_F^2) \hat{\sigma}_{a,b\to X}(\alpha_s(\mu_R^2), \mu_R^2), \tag{4.1}$$

where μ_F is the factorisation scale chosen such that it usually corresponds to a characteristic momentum transfer of the selected process and $x_{1,2}$ the *Bjorken* x described in more detail below.

4.1.1 *Parton Distribution Functions*

The PDFs are crucial for the description of proton-proton collisions since protons are not point-like particles but consist of so-called *partons*. The first type of partons are the valence quarks which determine the quantum numbers (charge, etc.) of the proton. In addition, gluons and virtual quark-antiquark pairs (sea-quarks) coming from vacuum fluctuations are also a part of the proton. A PDF $f_a^A(x, Q^2)$ describes the probability density of a parton a inside a hadron A to carry a certain momentum fraction $x = p_a/p_A$ also betoken as *Bjorken* x evaluated at a specific momentum transfer

Q^2. In general, PDFs cannot be directly predicted[1] thus they are extracted from several measurements using a complex fit, performed at a specific scale. Several collaborations such as the CTEQ, MSTW and NNPDF collaborations [74–77] determine the PDFs and provide them for physics analyses. With the help of the Dokshitzer–Gribov–Lipatov–Altarelli–Parisi (DGLAP) Equations [78–80], the PDFs can be extrapolated to different scales Q^2 and do not have to be measured at each scale individually. Figure 4.1 shows the proton PDFs for two different factorisation scales.

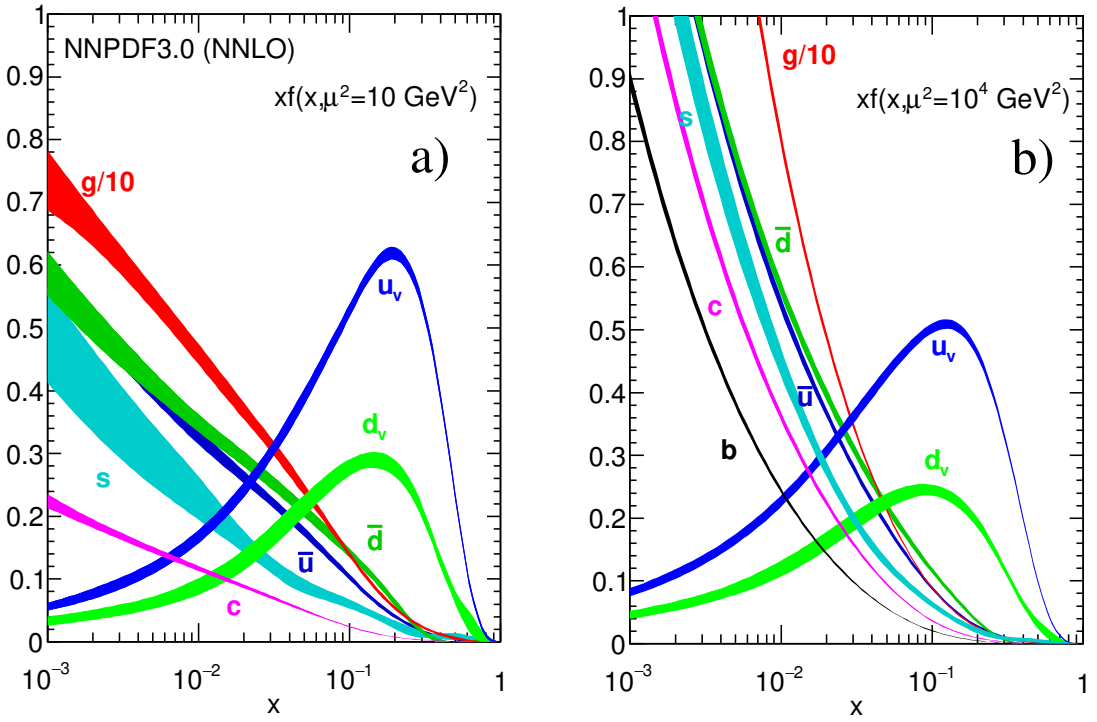


Figure 4.1.: The parton distribution functions $xf(x, \mu_F^2)$ are shown for two different factorisation scales μ_F^2: left: $\mu_F^2 = 10\,\mathrm{GeV}^2$ and right: $\mu_F^2 = 10\,\mathrm{TeV}^2$. They are obtained with the `NNPDF3.0NNLO` global analysis [77]. Plots were taken from [20].

Processes involving b-quarks can be described in QCD in two different factorisation schemes arising from the b-quark mass $\Lambda_{QCD} < m_b \ll \nu$: the four-flavour scheme (4FS) and the five-flavour scheme (5FS). The 4FS treats the b-quarks massive and since $m_b > m_{proton}$, they do not appear in the initial state. Consequently, the b-quarks do not have dedicated PDFs, so they decouple from the QCD perturbative evolution and therefore decouple from the α_s running and the number of 'light' flavour quarks is set to $n_f = 4$ in Equation (2.15). Considering the b-quarks as massive is especially impacting calculations at lower scales, around the production threshold. On the other hand, at high scales the mass effects are negligible. This case is described by the 5FS in which the initial state b-quarks are considered massless and they are treated in the same manner as the other light quarks comprising a b-quark PDF and $n_f = 5$.

1 In principle lattice QCD could be used to calculate PDFs [73], however, this is very computationally intensive.

4.1.2 *Monte Carlo Generators*

Typically, the event generation is divided into two parts: first the matrix element (ME) generation describing the hard scattering and secondly the parton shower (PS) evolution and hadronisation modelling including initial state radiation (ISR) and final state radiation (FSR). While the ME and most parts of the PS can be calculated perturbatively, the other processes are non-perturbative. A simplified illustration of this full simulation process is shown in Figure 4.2. For the modelling of the hadronisation, there are different models, the most widely used models are: the Lund string model [81] and the cluster model [82]. In the Lund string model, the colour connection of a quark-antiquark pair is described as a string and the potential between them is assumed to be linearly increasing with their distance. The strings then split according to a fragmentation function forming new quark-antiquark pairs which continues until only hadrons with on-shell mass remain. The cluster model is based on QCD pre-confinement, where neighbouring partons build colour-singlet clusters, these clusters then decay into two hadrons and they then decay further until the final state hadrons are formed.

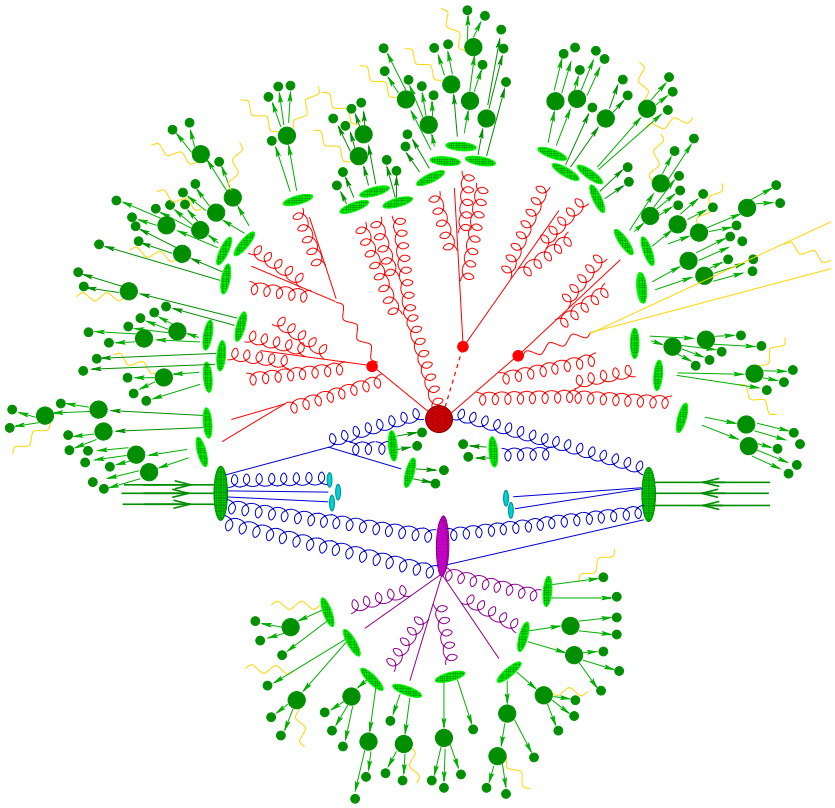


Figure 4.2.: Illustration of a hadron-hadron collision event simulated with a MC event generator. In the centre, the red circle represents the hard collision while the purple oval depicts the secondary hard scattering process (underlying event) with multi parton interaction. Both are surrounded by a tree-like structure describing the QCD bremsstrahlung simulated by the PS. The other elements in the sketch are the hadronisation (light green), hadron decays (dark green) and photon radiation (yellow) [83].

The full process involving matrix element generation, parton shower, underlying event, hadronisation and fragmentation can be simulated by MC generators like PYTHIA8 [84], HERWIG7 [85, 86] or SHERPA [87]. However, PYTHIA8 provides mainly leading order calculations which are often not sufficient since the next-to-leading order (NLO) corrections can be fairly large. HERWIG7 provides many MEs also at NLO. Since the fraction of negative event weights can be quite large (up to $\sim 40\%$ for certain generator setups), the generator is only used as parton shower in this thesis. In fact, there are other generators like POWHEGBOX [88–92] or MADGRAPH5_aMC@NLO [93] providing higher-order calculations which can be interfaced with PYTHIA8 or HERWIG7 for the simulation of PS and hadronisation.
Furthermore, the models used to describe the non-perturbative processes have parameters that can be tuned using collision data. The most common tunes used by the ATLAS experiment are the A14 parameters [94] for PYTHIA8 or the H7UE set of tuned parameters [86] for HERWIG7.

4.1.3 *Common Generator Setup of used Samples*

Throughout this thesis the physics processes for proton-proton collisions at a centre-of-mass energy $\sqrt{s} = 13\,\text{TeV}$ are modelled using various combinations of MC generators and settings. The specific details are stated in the dedicated chapters. Nevertheless, all MC samples using PYTHIA8 or HERWIG7 to model the multi-parton interaction (MPI), hadronisation and PS use the same settings if not differently stated. The mass of the top quark is set to $m_t = 172.5\,\text{GeV}$, the Higgs boson mass to $m_H = 125\,\text{GeV}$ and the mass of the b-quark to $m_b = 4.8\,\text{GeV}$ for PYTHIA8, to $m_b = 4.5\,\text{GeV}$ for HERWIG7 and to $m_b = 4.75\,\text{GeV}$ for SHERPA. The simulation of b- and c-hadron decays is performed via the EVTGEN v1.6.0 program [95] with the exception of SHERPA. As mentioned above the two tunes A14 combined with the `NNPDF2.3LO` PDF set [96] and H7UE together with the set of `MMHT2014LO` PDFs [97] are used for PYTHIA8 and HERWIG7, respectively.

4.2 DETECTOR SIMULATION

The last step in the simulation chain is the detector simulation. The MC generators, as described in Section 4.1, provide information about stable particles in the final state, not taking into account the detector response. The full ATLAS detector simulation [98] is performed in two steps. The first step is based on GEANT4 [99] incorporating the geometry of the detector and providing highly precise modelling of the particle interactions with the detector matter. However, it comes with the shortcoming of using a large fraction of the available computing power of ATLAS. As an alternative, fast calorimeter simulation algorithms [100–102] are developed and already used in practice. They mimic the GEANT4 results, based on thousands of individual parametrisations of the calorimeter response, using significantly less computing resources with a trade-off in precision. A comparison of the necessary CPU time for the different detector simulations are shown in Figure 4.3. In practice,

the fast simulation algorithms are widely used in ATLAS and are called *AtlFast-II*. In the second step, the readout electronics and digitisation is simulated which is adjusted for the different detector systems.

Taking advantage of the latest machine learning developments in the last years, deep generative algorithms such as Generative Adversarial Networks (GANs) and Variational Auto-Encoders (VAEs) are studied to improve the fast calorimeter simulation [103] showing already promising results.

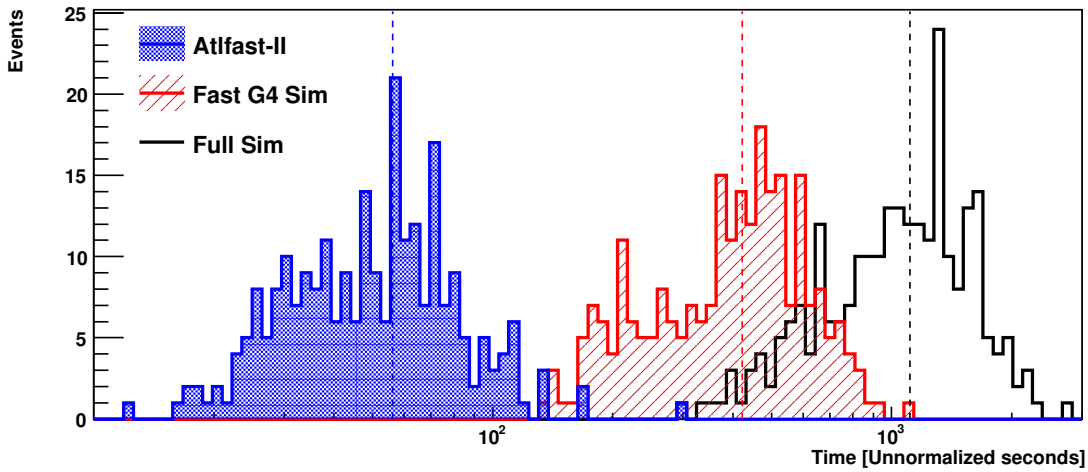


Figure 4.3.: Comparison of the CPU time distributions for the full GEANT4 (black), fast GEANT4 (red) and the fast calorimeter simulation (blue) for 250 $t\bar{t}$ events. The vertical dotted lines indicate the average of the distributions [100].

5

CURRENT STATUS OF TTH MEASUREMENTS AT THE LHC

The discovery of the Higgs boson in 2012 was a huge success for the LHC. The four main Higgs production channels at the LHC were described in Section 2.6. This thesis focuses on the $t\bar{t}H$ production channel which was observed by ATLAS and CMS [8–10] combining different Higgs decay channels. The $t\bar{t}H$ production allows a direct measurement of the top-Yukawa coupling y_t which is the largest Yukawa coupling within the SM. The measurement of y_t is an essential validity test of the SM in particular for the Higgs mechanism and is important for both new physics searches and Higgs precision measurements. A direct y_t measurement can be compared to indirect measurements, e.g. form the loop induced ggF production or the $H \to \gamma\gamma$ decay mode, giving hints for possible effects beyond the SM. Compared to the total Higgs production cross-section, the $t\bar{t}H$ production only contributes around 1% at the LHC, as shown in Figure 2.5, but has a recognisable detector signature with the two associated top quarks.
The $H \to b\bar{b}$ decay mode has the largest branching fraction of 58% (see Fig. 2.6). It was observed by both ATLAS and CMS [11, 12] and is sensitive to the second-largest Yukawa coupling in the SM, the b-quark Yukawa coupling y_b. Besides the large branching fraction, the $H \to b\bar{b}$ decay mode also allows the kinematic reconstruction of the Higgs boson. Therefore, it is possible to further explore the properties of the Higgs boson in $t\bar{t}H(b\bar{b})$ events.

In this thesis, the $t\bar{t}H$ production channel is being investigated together with the $H \to b\bar{b}$ decay mode. This particular process is not yet discovered but CMS already sees an evidence [104].
In this chapter, a short overview of the $t\bar{t}H$ discovery and of the latest $t\bar{t}H(b\bar{b})$ results is given, based on the latest analysis results from ATLAS [13] and CMS [14], followed by the introduction to Simplified Template Cross-Section (STXS) measurements.

5.1 LATEST RESULTS

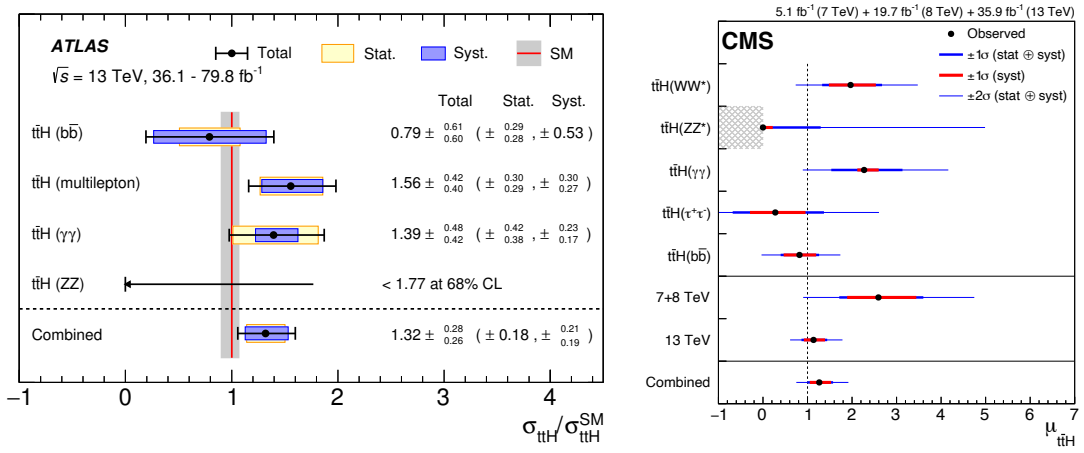
5.1.1 $t\bar{t}H$ *Observation*

The observation of the $t\bar{t}H$ production of ATLAS [10] and CMS [8] was an important achievement for the two experiments. The CMS combination considered five different Higgs decay channels as shown in Figure 5.1 (b) using the data recorded during RUN I and II of the LHC, resulting in an observed (expected) significance of 5.2 (4.2) standard deviations. The ATLAS analysis used four different Higgs decay channels in the $t\bar{t}H$ combination with RUN II data indicated in Figure 5.1 (a) with an observed (expected) significance of 5.8 (4.9) standard deviations and a cross-section of $\sigma(t\bar{t}H) = (670 \pm 90(\text{stat.})^{+110}_{-100}\text{syst.})$. Table 5.1 lists the sensitivities of the individual channels

Analysis	Integrated luminosity [fb^{-1}]	$t\bar{t}H$ cross section [fb]	Obs. sign.	Exp. sign.
$H \to \gamma\gamma$	79.8	$710\ ^{+210}_{-190}$ (stat.) $^{+120}_{-90}$ (syst.)	4.1 σ	3.7 σ
$H \to$ multilepton	36.1	790 ± 150 (stat.) $^{+150}_{-140}$ (syst.)	4.1 σ	2.8 σ
$H \to b\bar{b}$	36.1	$400\ ^{+150}_{-140}$ (stat.) $\pm$ 270 (syst.)	1.4 σ	1.6 σ
$H \to ZZ^* \to 4\ell$	79.8	$<$900 (68% CL)	0 σ	1.2 σ
Combined (13 TeV)	36.1$-$79.8	670 ± 90 (stat.) $^{+110}_{-100}$ (syst.)	5.8 σ	4.9 σ
Combined (7, 8, 13 TeV)	4.5, 20.3, 36.1$-$79.8	$-$	6.3 σ	5.1 σ

Table 5.1.: Overview of the results of the single channels used for the ATLAS $t\bar{t}H$ combination indicating their respective cross-section and the observed and expected significance [10].

among which the $H \to \gamma\gamma$ and multilepton decay channel of the Higgs boson have the largest significance. The $H \to \gamma\gamma$ decay channel is mainly dominated by statistical uncertainties while in the multilepton decay channel the statistical and systematic uncertainties have a similar influence. The $H \to b\bar{b}$ channel has a significantly lower sensitivity and is dominated by systematic uncertainties mainly due to large uncertainties in the modelling of the $t\bar{t} + b\bar{b}$ background.



(a) (b)

Figure 5.1.: Signal strength parameter of the individual channels and the combined signal strength of $t\bar{t}H$ shown for the $t\bar{t}H$ observation of ATLAS [10] (a) and CMS [8] (b).

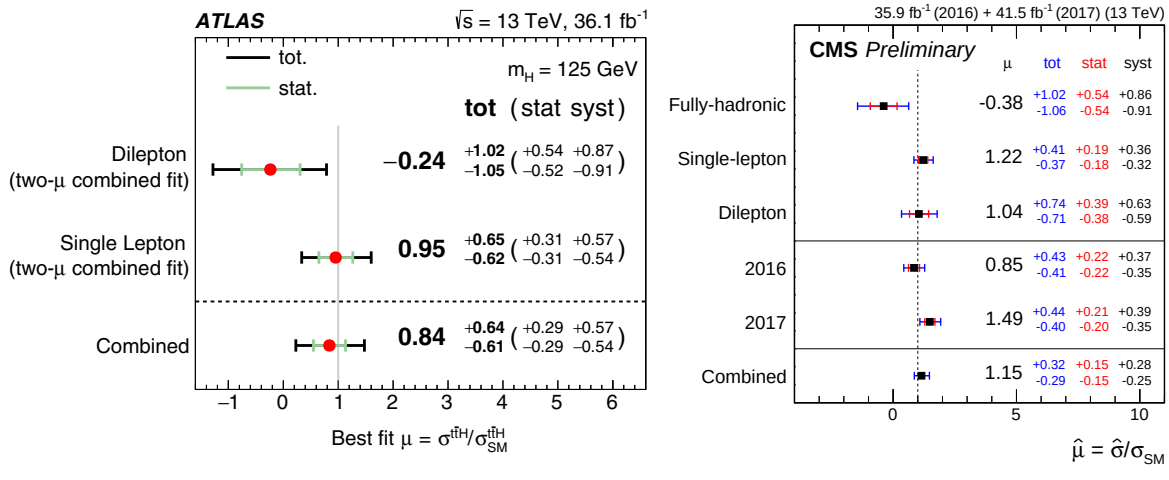
5.1.2 $t\bar{t}H(b\bar{b})$ *Results*

Both ATLAS and CMS have already performed searches for $t\bar{t}H(b\bar{b})$ with a subset of the LHC RUN II dataset shown in Ref. [13] with 36.1 fb^{-1} and [104] with 77.4 fb^{-1}, respectively.
The ATLAS analysis uses events where at least one W-boson from one of the two top quarks decays leptonically and makes use of a complex definition of analysis regions. These analysis regions are mainly defined based on different b-tagging criteria to better extract the information of the signal and background processes in dedicated phase spaces. The analysis is optimised using a set of multivariate analysis techniques such as reconstruction and classification Boosted Decision Trees (BDTs), a

likelihood discriminant and the matrix element method [105]. The combined signal strength[1] was measured with an observed (expected) significance of 1.4 (1.6) standard deviations and a value of $\mu = 0.84 \pm 0.29(\text{stat.})^{+0.57}_{-0.54}(\text{syst.})$ shown in Figure 5.2 (a).

CMS combined two analyses, where one uses the 2016 dataset with 35.9 fb^{-1} and the second analysis is performed with the 2017 dataset with 41.5 fb^{-1}. These two analyses cover the lepton+jets, dilepton and fully hadronic channels. A matrix element method [105] as well as multivariate techniques are employed. In the lepton+jets channel, a Deep Neural Network is employed to categorise the events into signal and multiple background processes via a multi-classification approach. At the same time, the output discriminants of the network are also used in the combined fit. As shown in Figure 5.2, the combined signal strength was measured to $\mu = 1.15 \pm 0.15(\text{stat.})^{+0.28}_{-0.25}(\text{syst.})$ with an observed (expected) significance of 3.9 (3.5) standard deviations.

Both results are mainly dominated by systematic uncertainties and limited by the challenging modelling of the $t\bar{t} + b\bar{b}$ background. The uncertainties are grouped into different sources and compared between the two analyses of ATLAS and CMS in Table 5.2. Overall, the systematic uncertainties of the CMS analysis are considerably smaller than those from the ATLAS analysis. While the uncertainty associated to the $t\bar{t} + \geqslant 1b$ modelling is dominating the ATLAS analysis by far, the combined $t\bar{t} +$ heavy flavour modelling uncertainty is roughly in the same order as for instance the signal modelling uncertainty in the CMS analysis. ATLAS is accounting for the differences between the 4 and 5 flavour scheme in the $t\bar{t} + b\bar{b}$ modelling which is fairly large whereas CMS is not taking them into account.



(a) (b)

Figure 5.2.: Signal strength of the first RUN II $t\bar{t}H(b\bar{b})$ analyses for (a) the ATLAS analysis [13] and (b) CMS analysis [104] showing the signal strengths of the different channels and their combination.

1 The signal strength μ is defined as the cross-section ratio of the measured cross-section σ over the cross-section expected in the SM σ_{SM}: $\mu = \frac{\sigma}{\sigma_{SM}}$.

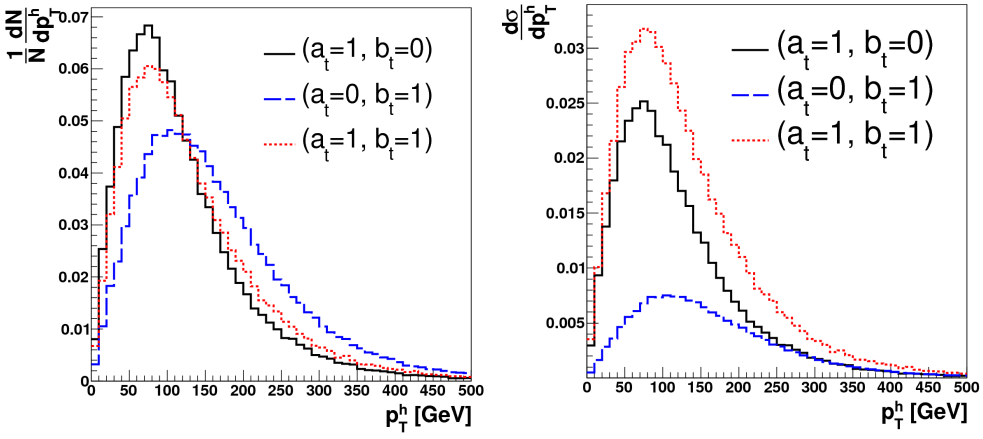
Uncertainty source	$\Delta\mu$	
	ATLAS	CMS
$t\bar{t} + \geqslant 1b$ modelling	$+0.46/-0.46$	
$t\bar{t} + \geqslant 1c$ modelling	$+0.09/-0.11$	
$t\bar{t}$ + hf modelling		$+0.14/-0.15$
Signal modelling	$+0.22/-0.05$	$+0.15/-0.06$
Size of simulated sample	$+0.29/-0.31$	$+0.10/-0.10$
b-tagging	$+0.16/-0.16$	$+0.08/-0.07$
Jet energy scale and resolution	$+0.14/-0.14$	$+0.05/-0.04$
Total systematic uncertainty	$+0.57/-0.54$	$+0.28/-0.25$
Statistical	$+0.29-0.29$	$+0.15/-0.15$
Total	$+0.64/-0.61$	$+0.32/-0.29$

Table 5.2.: Comparison of the breakdown of the contributions to the uncertainties in μ ($\Delta\mu$) for the ATLAS and CMS $t\bar{t}H(b\bar{b})$ analysis. The ATLAS analysis split the group of the $t\bar{t}$ production in association with heavy flavour jets ($t\bar{t}$ + hf) modelling uncertainties into $t\bar{t} + \geqslant 1b$ and $t\bar{t} + \geqslant 1c$ modelling while CMS provided the combined value for it [13, 104].

5.2 SIMPLIFIED TEMPLATE CROSS-SECTION MEASUREMENTS

The Simplified Template Cross-Section (STXS) formalism [106, 107] is a common effort of the LHC experiments to define a consistent basis for comparable differential Higgs kinematic measurements performed in exclusive kinematic phase space regions (STXS bins). This simplifies the combinations of different decay channels as well as between the experiments. Several theory uncertainties are directly folded into the measurements. Thus, the kinematic bins are optimised such that they reduce them as much as possible [107, 108]. For each STXS bin, a separate signal template is defined which is the signal MC prediction in the targeted kinematic region at truth level. After the discovery of the most prominent Higgs decay channels, the statistics of RUN II allows now to perform STXS measurements in the $t\bar{t}H$ channel.

Cross-section measurements in the $t\bar{t}H$ production channel split into bins of the transverse momentum of the Higgs boson p_T^H are sensitive to the CP structure of the Higgs boson [109] and to the Higgs self-coupling [110]. Figure 5.3 shows the normalised p_T^H distribution with the Higgs produced in the $t\bar{t}H$ mode and the normalised differential cross-section as a function of p_T^H for three different CP-scenarios of the Higgs coupling to the top quark: scalar (CP even) case (solid black) which corresponds to the predictions of the SM, pseudo-scalar (CP odd) case (dashed blue) and the CP-violating case (dotted red). In the pseudo-scalar scenario, the values are shifted to higher p_T^H values and the differential cross-section is suppressed with respect to the SM case. In the presented analysis, these properties are, however, not yet investigated due to a too low sensitivity.



(a) (b)

Figure 5.3.: The normalised event distribution (a) and the cross-section (b) as a function of the Higgs transverse momentum p_T^H shown for three different CP-scenarios of the Higgs boson: CP even (solid black), CP odd (dashed blue) and CP-violating (dotted red) [109].

The $H \to b\bar{b}$ decay mode is well suited to probe the differential cross-sections due to their large production rate. Even the high p_T^H regime can be accessed in which for instance the $t\bar{t}H(H \to \gamma\gamma)$ decay mode is lacking statistics [111]. Another advantage is that the $b\bar{b}$ final state of the Higgs can be fully reconstructed.

Similar to the recommendations for the STXS bins for other Higgs production channels given in Ref. [106], taking into account the theory considerations from above, the following STXS bins used for the analysis shown in this thesis are: 0 GeV $\leqslant p_T^H <$ 120 GeV, 120 GeV $\leqslant p_T^H <$ 200 GeV, 200 GeV $\leqslant p_T^H <$ 300 GeV, 300 GeV $\leqslant p_T^H <$ 450 GeV, and $p_T^H \geqslant$ 450 GeV.

Part II

METHODOLOGY

6

OBJECT RECONSTRUCTION AND PARTICLE IDENTIFICATION IN ATLAS

Physics processes produce electrons, muons, taus, quarks, gluons, photons and neutrinos in their final state. Since only position and energy information can be extracted from the detector, these physics objects need to be reconstructed. While quarks are forming jets, neutrinos cannot be directly seen; therefore, it is necessary to define, what the objects measured in the detector are.

Figure 6.1 illustrates the interaction of different particles with the ATLAS detector: charged particles leave a track in the ID, electrons and photons shower in the EM calorimeter while hadrons shower in the hadronic calorimeter.

In particular, the $t\bar{t}H(b\bar{b})$ analysis in Part IV makes use of electrons, muons, taus, jets and missing transverse momentum. The reconstruction of these objects is described in the following chapter. In addition, heavy-flavour identification is also an important tool for the $t\bar{t}H(b\bar{b})$ analysis. This topic is covered in the dedicated Chapter 8.

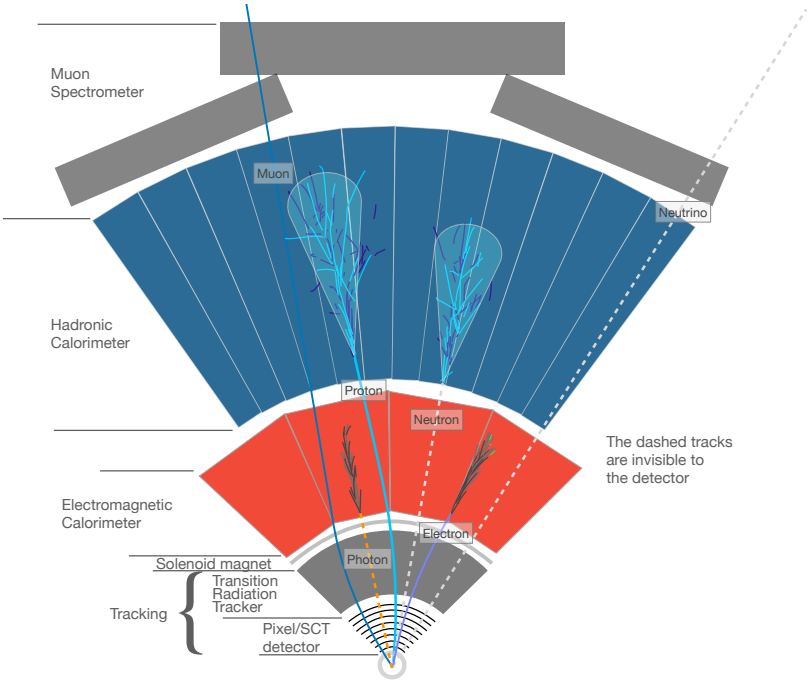


Figure 6.1.: Sketch of a section of the ATLAS detector in the transverse plane showing the interaction of particles with the detector material. Adapted from [112].

6.1 RECONSTRUCTION FROM DETECTOR HITS

All physics objects reconstructed with the ATLAS detector are composed of tracks, vertices or calorimeter energy clusters (topo clusters). They are the fundamental building blocks used in all reconstruction algorithms and are introduced in the following section.

6.1.1 *Tracking*

Charged particles which are passing through ATLAS leave tracks in the ID. They are reconstructed from energy deposits or hits in the ID within the tracking acceptance of $|\eta| < 2.5$. A detailed description of the tracking in RUN II can be found in Ref. [113].

As a first step, hits are assembled by grouping pixels and strips into clusters that reach an energy deposit above a given threshold. In the next step, three-dimensional space points are defined where charged particles traverse the active detector material of the ID.
Next, a combinatorial track finding procedure is applied, starting by forming track seeds from a set of three space points. The combination is done following the preliminary track trajectory adding space points iteratively. Afterwards, a score is associated to each track indicating if the track correctly represents the trajectory of a charged primary particle. According to the track score, the ambiguity solver evaluates the tracks in decreasing order to limit shared clusters which typically indicate a wrong assignment. At this stage, quality criteria are applied where the tracks have to have a minimum transverse momentum $p_T > 500\,\text{MeV}$ and $|\eta| < 2.5$. Moreover a minimum of seven pixel and SCT clusters (twelve are expected), a maximum of either one shared pixel cluster or two shared SCT clusters on the same layer, not more than 2 holes[1] in the combined pixel and SCT detectors and no more than one hole in the pixel detector are required together with impact parameter requirements $|d_0^{BL}| < 2\,\text{mm}$ and $|z_0^{BL} \sin\theta| < 3\,\text{mm}$. Here d_0^{BL} is the transverse impact parameter (IP) calculated w.r.t the measured beamline position and z_0^{BL} is the longitudinal difference along the beamline between the primary vertex (PV) and the point where d_0^{BL} is measured. Finally, the reconstructed tracks are extrapolated into the TRT volume, also adding the TRT hits to the tracks.

6.1.2 *Vertexing*

A vertex is the origin of tracks and therefore the point of particle interactions or particle decay. The primary vertex (PV) is of particular interest denoting the hard interaction of the partons in the colliding protons. Besides the PV, secondary and tertiary vertices are also important, especially in heavy-flavour tagging covered in Chapter 8.
The primary vertices within an event are iteratively reconstructed with an algorithm [114] which is briefly described in the following. Firstly, a set of tracks is defined satisfying certain selection

1 A hole is a missing intersection with a sensitive detector element expected from the track trajectory estimation.

criteria (similar the track requirements in sec. 6.1.1) and a seed position of the first vertex is selected. Secondly, the tracks and the seed are utilised to estimate the best vertex position with an iterative fit. In each step, the vertex position is recomputed after down-weighting less compatible tracks. After the determination of the vertex position, all incompatible tracks are discarded to be used in another vertex. This procedure is then repeated with the remaining tracks in the event. The vertex with the largest quadratic p_T sum is defined as the PV. For the $t\bar{t}H(b\bar{b})$ analysis in this thesis, only events with at least one PV are used, to which two or more tracks are associated with $p_T > 500\,\text{MeV}$.

6.1.3 *Topo Clusters*

Vertices and tracks are reconstructed from the ID information, whereas the topological cell clusters, also called topo clusters, are iteratively reconstructed from calorimeter information [115].
Topologically connected cell signals are reconstructed to form 3-D 'energy blobs' from particle showers in the active calorimeter volume by extracting the significant signals over electronic noise and other fluctuations such as pile-up. This clustering is particularly effective in highly granular calorimeter systems used in ATLAS. The topo clusters are a full or fractional response to a single particle, merged response to several particles or a combination of the two.

6.2 PHYSICS OBJECTS

The physics objects used in this thesis are jets, electrons, muons, taus and missing transverse momentum. The definitions of these objects, their reconstruction algorithms and their performances are described in the following.

6.2.1 *Jets*

Due to the colour charge carried by quarks and gluons, they cannot be observed as free particles and form colourless hadrons. Jets are collimated showers formed by these hadrons. In fact, jets do not have a unique object definition. They rather depend on the chosen clustering algorithm which depend as little as possible on QCD effects [116]. Namely, the jet algorithm has to be *collinear safe*, meaning that the jet configuration does not change by substituting one particle with two collinear particles, and it has to be *infrared safe* for which the configuration should not change by adding soft particles.
In a jet, different detector objects are clustered together such as charged and neutral hadrons, photons (mostly from π^0 decays) as well as electrons and muons can be included. In the detector, the charged particles in a jet first leave tracks in the ID and then deposit energy in the electromagnetic and hadronic calorimeters. Also neutral hadrons and photons deposit energy in the calorimeters. At the LHC, the jet reconstruction is typically performed using the anti-k_t algorithm [117]. It is a clustering

algorithm combining four-vector objects into a cone-like object, a jet. The distance parameter d_{ij} between object i and j defined as:

$$d_{ij} = \min\left(p_{T,i}^{2n}, p_{T,j}^{2n}\right) \frac{\Delta R_{ij}^2}{R^2}, \tag{6.1}$$

with their transverse momentum $p_{T,i/j}$ and the ΔR_{ij} between the two four-vector objects i and j, allows a recursive recombination together with the distance of object i to the beam axis

$$d_{i,B} = p_{T,i}^{2n} R^2, \tag{6.2}$$

with R being the radius parameter and the exponent n set to -1. The advantage of this choice ($n = -1$) is that the clustering prefers high momenta (hard) particles instead of soft ones which leads to an almost circular shape around the hardest particle as shown in Figure 6.2.
Based on the anti-k_t algorithm various jet collections are deployed in ATLAS, three of them will be used in this thesis: *EMTopo* jets, *Particle Flow* jets and *Variable Radius Track* jets.

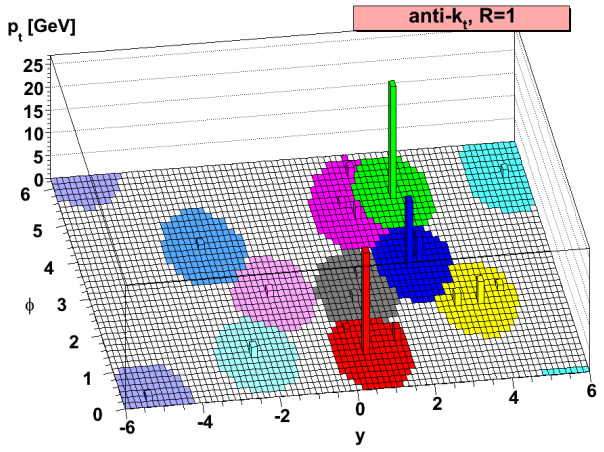


Figure 6.2.: Illustration of the anti-k_t clustering algorithm showing a circular (cone-like) structure around the track with the highest momentum [117].

6.2.1.1 *EMTopo Jets*

The so-called *EMTopo* jets are calorimeter jets reconstructed at the EM energy scale only using topo clusters [118] with the anti-k_t algorithm implemented in the software package FastJet [119]. For the scope of this thesis, the radius parameter $R = 0.4$ is used (for boosted topologies also jets with $R = 1.0$ are used). Additionally, the jets have to satisfy $p_T > 25\,\text{GeV}$ and $|\eta| < 2.5$. Until recently, the *EMTopo* jets were the primary jet collection, used in physics analyses in ATLAS, showing robust energy characteristics.
The calibration of *EMTopo* jets is performed in several steps illustrated in Figure 6.3 correcting the four-momentum of the jet [120]. After the jet reconstruction, the jet direction is modified at the topo cluster level, such that the jet originates from the primary vertex. Then, p_T-density based pile-up

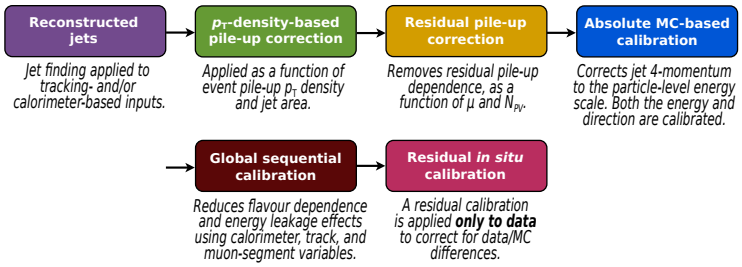


Figure 6.3.: Different steps of the jet four momentum calibration [120].

corrections are applied including jet area information as well as a MC-driven residual correction. The absolute jet energy calibration corrects the jets to agree in energy and direction with dijet MC events. Then a global sequential calibration is set to improve the p_T resolution and the associated uncertainties. The final step is the *in situ* calibration which is only applied to data. If there are still remaining differences between data and MC, they are corrected at this step.

6.2.1.2 *Particle Flow Jets*

During RUN II, ATLAS introduced *Particle Flow jets*, a new jet collection also denoted as *PFlow* jets. They combine tracking and calorimeter information in the jet reconstruction [121] also using the anti-k_t clustering algorithm with a radius parameter of $R = 0.4$.
The first step is to match the tracks from charged particles in the ID to the topo clusters from the calorimeter. In case of a successful match, the energy deposit of the topo cluster is replaced by the corresponding track momentum. The anti-k_t algorithm then takes as input the topo clusters that remain after substitution as well as tracks that match the hard-scattering PV. The calibration follows closely the *EMTopo* scheme performed in the range $20\,\text{GeV} < p_T < 1500\,\text{GeV}$ [120].
The advantage of *PFlow* jets is their improved energy and angular resolution compared to *EMTopo* jets. Also not negligible is the enhanced reconstruction efficiency and pile-up stability.

6.2.1.3 *Variable Radius Track Jets*

In boosted topologies, especially for the boosted $H \to b\bar{b}$ decays, the two b-jets are very collimated and the hadronisation products of the two b-quarks start to overlap at a certain p_T value. In order to avoid these overlaps and to improve b-tagging performance, a jet algorithm with a variable radius (VR) parameter is employed [122] based on track jets formed from charged-particle tracks with $p_T > 0.5$ GeV and $|\eta| < 2.5$. They will be further called *VR Track* jets. The radius parameter from the conventional anti-k_t algorithm from equations (6.1) and (6.2) is now p_T dependent:

$$R \to R_{\text{eff}}(p_T) = \frac{\rho}{p_T}, \tag{6.3}$$

where ρ is a parameter that controls how fast the effective jet size decreases with p_T. In addition, a cut-off parameter is introduced to prevent too large jet radii R_{max} as well as another cut-off to avoid the jet to shrink below the detector resolution R_{min}. These parameters are optimised for $H \to b\bar{b}$

events [123, 124] resulting in the following parameters: $\rho = 30\,\text{GeV}$, $R_{\text{min}} = 0.02$ and $R_{\text{max}} = 0.4$. Furthermore, only jets with at least two constituents, $p_T > 10\,\text{GeV}$ and $|\eta| < 2.5$ are considered [125]. The track jets are not separately energy-calibrated, the energy is calculated via the sum of the track momenta of the tracks associated to the jet.

6.2.2 *Electrons*

Electrons are reconstructed using the information of the ID and the calorimeter system. The typical signature of an electron is that they leave a track in the ID and are then absorbed in the electromagnetic calorimeter where they leave an electromagnetic shower. The $t\bar{t}H(b\bar{b})$ analysis and the flavour tagging studies in this thesis use the algorithms described in detail in Ref. [126, 127].

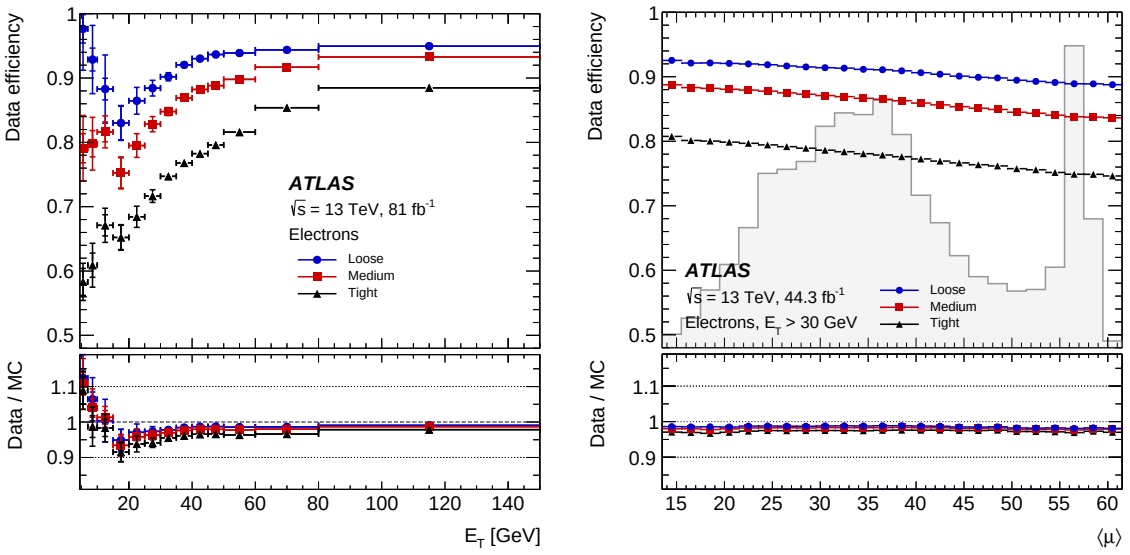
6.2.2.1 *Reconstruction*

The electron object is constructed using a dynamic clustering algorithm with variable-size clusters, so-called superclusters. The reconstruction is performed in the region $|\eta_{\text{cluster}}| < 2.47$ excluding the transition region of the barrel and end-cap ($1.37 < |\eta_{\text{cluster}}| < 1.52$).
At first, topo clusters (described in sec. 6.1.3) are selected and loosely matched to ID tracks. Simultaneously, the conversion vertices matched to the topo clusters are built. Next, the superclusters are built from matched clusters and a first position correction and energy calibration is applied. Tracks are then matched to the electron superclusters. The energy scale and resolution of electrons are calibrated using $Z \to ee$ decays and validated in $Z \to \ell\ell\gamma$ decays [127]. In addition, the energy resolution of the electron is optimised using a multivariate regression algorithm based on the properties of shower developments in the electromagnetic calorimeter.

6.2.2.2 *Identification*

Further quality criteria are required for an electron object, passing several identification selections to improve the purity of the selected objects. The prompt electrons are identified using a likelihood discriminant which uses quantities measured in the ID and the electromagnetic calorimeter. These quantities are chosen such that they discriminate well prompt isolated electrons from other energy deposits like jets, converted photons or genuine electrons stemming from heavy-flavoured hadron decays. Important observables for the likelihood calculation are based on the track quality in the ID, the lateral and longitudinal development of the electromagnetic shower described by shower shape variables as well as the particle identification in the TRT. The algorithm uses probability density functions as input which are derived for the signal from $Z \to ee$ ($E_T > 15\,\text{GeV}$) and $J/\psi \to ee$ ($E_T < 15\,\text{GeV}$) events.
The efficiency of the electron identification is provided in three operating points: *Loose*, *Medium* and *Tight*, yielding different purities. Figure 6.4 shows the data efficiency as a function of E_T and as a function of the average number of bunch crossings for all three operating points. They are all optimised in 9 $|\eta|$ and 12 E_T bins. For this thesis the *Medium* and *Tight* operation points are

used with an average target efficiency of 88% and 80% associated with an increased background process rejection of ~ 2.0 and ~ 3.5 with respect to the inclusive target efficiency, respectively (for $20\,\text{GeV} < E_T < 50\,\text{GeV}$).



(a) (b)

Figure 6.4.: Measured electron identification efficiency in $Z \to ee$ data events as a function of E_T (a) and as a function of the average number of interactions per bunch crossing $\langle\mu\rangle$ (b), for the *Loose*, *Medium* and *Tight* operating point. The right plot also shows in grey the shape of the $\langle\mu\rangle$ distribution. The lower pad shows the Data/MC comparison [127].

6.2.2.3 *Isolation Criteria*

Electrons are typically required to be spatially separated from other particles. There are two kinds of isolation variables: calorimeter-based and track-based.

The calorimeter-based isolation is calculated via the sum of the transverse energy of positive-energy topo clusters with a barycentre falling in a $\Delta R = 0.2$ of the electron barycentre (other than the electron clusters themselves). In addition, leakage and pile-up corrections are applied.

For the track-based isolation the sum of the transverse momentum of tracks within a cone centred around the electron track are considered, where the cone radius decreases with p_T. Moreover, only tracks are taken into account which have $p_T > 1\,\text{GeV}$ and $|\eta| < 2.5$ as well as satisfy certain track quality criteria and have a loose vertex association[2]. In this thesis the *Gradient* isolation working point (WP) is chosen which gives an efficiency of 90% at $p_T = 25\,\text{GeV}$ and 99% at $p_T = 60\,\text{GeV}$ uniform in η.

2 Vertices defined as loose association vertex are either used in the PV fit or satisfy specific IP criteria ($|\Delta z_0| \sin\theta < 3\,\text{mm}$).

6.2.3 *Muons*

Muons leave a track in the detector and traverse the calorimeter system typically without significant energy loss. Therefore, the muon is mainly reconstructed in the ID and the MS sub-detector systems. The RUN II muon reconstruction and performance is described in detail in [128].

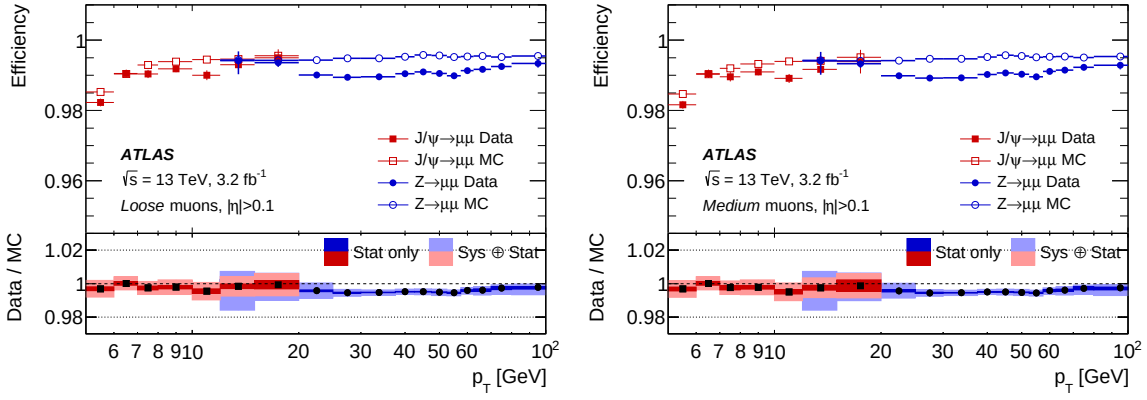
6.2.3.1 *Reconstruction*

The muon reconstruction has two stages: first the independent reconstruction in the ID and MS and secondly the combination of the two to form the muon tracks. The reconstruction in the ID is performed as for any other charged particle.
In the MS at first a search for hit patterns is performed in each muon chamber to form segments. In the MDTs and nearby trigger chambers, hits are aligned on the trajectory in the bending plane of the detector using a Hough transformation and the segments are reconstructed with a straight line fit to hits found in each layer. The hits of the RPCs and TGCs provide measurements for the plane orthogonal to the bending plane and in the CSCs a combinatorial search in the η and ϕ plane is utilised to build the segments. Given this information, muon track candidates are constructed by fitting segments from different layers using a global χ^2 fit.
The combined reconstruction is based on various algorithms defining four different types of muons. The combined (CB) muons are first independently reconstructed in the ID and MS and then their information is combined with an outside-in approach, extrapolating reconstructed tracks from the MS to the ID (complementary an inside-out approach is also used). The segment-tagged (ST) muons are mainly reconstructed from tracks in the ID extrapolated to typically one track segment in the MDTs and CSCs. The third type are the calorimeter-tagged (CT) muons where an ID track is matched to an energy deposit in the calorimeter compatible with a minimal ionising particle. This muon type has the lowest purity and is optimised for the region $|\eta| < 0.1$ with $15\,\text{GeV} < p_T < 100\,\text{GeV}$. There are also extrapolated (ME) muons that are only reconstructed in the MS extending the acceptance to $2.5 < |\eta| < 2.7$ but they are not used in this thesis.

6.2.3.2 *Identification*

Similarly to the electron identification, the muon identification is performed applying quality criteria to suppress background processes.The goal is to identify prompt muons with high efficiency and a good momentum resolution which requires a certain amount of hits in the ID and the MS. To cover different needs of physics analyses, four different muon WPs are available: *Loose*, *Medium*, *Tight* and *high* p_T. For the scope of this thesis, the *Medium* and *Loose* WPs are used.
For the *Medium* WP only combined muons are taken into account. The combined muons are required to have three or more hits in at least two MDT layers except for the $|\eta| < 0.1$ region where only one MDT layer is sufficient combined with no more than one hole layer due to a gap in the MS. This WP tries to minimise systematic reconstruction and calibration systematic uncertainties associated with the muon. The *Medium* WP reconstruction efficiency with $p_T > 20\,\text{GeV}$ is 96.1%.

The *Loose* WP maximises the reconstruction efficiency with good-quality muon tracks. In this case, all muon types are utilised. In fact, the combined muons are used as they are from the *Medium* selection. Additionally, the calorimeter-tagged and segment-tagged muons are taken into account for $|\eta| < 0.1$. The reconstruction efficiency for muons with $p_T > 20\,\text{GeV}$ is 98.1%.
Figure 6.5 shows the reconstruction efficiency measured in data for the two described WPs obtained from $Z \to \mu\mu$ and $J\psi \to \mu\mu$ events.



(a) (b)

Figure 6.5.: The reconstruction efficiency as a function of the muon p_T for the *Loose* (a) and the *Medium* (b) muon selection obtained from $Z \to \mu\mu$ and $J\psi \to \mu\mu$ events with $0.1 < |\eta| < 2.5$. The Data/MC ratio in the lower pad includes systematic and statistical uncertainties while the efficiencies only show statistical uncertainties [128].

6.2.3.3 *Isolation*

Analogously to the electron isolation strategy, the muon isolation is assessed via track- and calorimeter-based variables with very similar definitions. The track-based variable $p_T^{\text{varcone30}}$ is the scalar p_T sum of all tracks, excluding the muon track, with $p_T > 1\,\text{GeV}$ in a radius of $\Delta R = \min(10\,\text{GeV}/p_T^{\mu}, 0.3)$ around the muon transverse momentum p_T^{μ}. The calorimeter isolation variable is constructed from the sum of transverse energies around the muon track, as described for electrons. However, for the scope of this thesis the isolation WP *FixedCutTightTrackOnly* is utilised which is only using the track-based isolation satisfying $p_T^{\text{varcone30}}/p_T^{\mu} < 0.06$. The isolation efficiencies are measured using $Z \to \mu\mu$ events.

6.2.4 *Taus*

τ-leptons can decay either leptonically (into electrons or muons) or hadronically. The leptonic decays are similarly reconstructed as electrons or muons. The τ decays with a hadronic final state are seeded by jets which are required to have $p_T > 10\,\text{GeV}$ and $|\eta| < 2.5$ excluding the barrel-end-cap transition region [129]. Tau leptons are calibrated to correct their energy deposit in the detector to the average value at generator level. The τ identification is based on BDTs discriminating τ-jets from the quark-

and gluon-initiated background jets. Three different efficiency WPs are defined: *Loose*, *Medium* and *Tight*. In this thesis the *Medium* τ-WP and the requirement $p_T > 25\,\text{GeV}$ is used as well as an isolation criterion of $\Delta R_y < 0.2$ between a τ_{had} candidate and any selected electron or muon.

6.2.5 *Missing Transverse Momentum*

The missing transverse momentum, also denoted as E_T^{miss} is the negative vector sum of fully calibrated electrons, muons, photons, hadronically decaying τ-leptons and jets denoted as the hard term as well as soft objects coming from additional tracks associated to the PV [130]. The partons inside the proton are following a momentum distribution (see sec. 4.1.1) and the centre-of-mass system of the hard scattering is not at rest w.r.t. the lab system. Therefore, the known quantity in an ideal detector is the transverse momentum p_T which is 0 at the time of the interaction. However, not all objects are always detected, e.g. neutrinos leave the detector unseen. So E_T^{miss} is a measure of the neutrinos that escape detection. The vector of the missing transverse momentum can be split into a scalar part E_T^{miss} and an azimuthal angle ϕ^{miss} which are defined as

$$E_T^{miss} = \sqrt{(E_x^{miss})^2 + (E_y^{miss})^2}, \tag{6.4}$$

$$\phi^{miss} = \tan^{-1}(E_y^{miss}/E_x^{miss}), \tag{6.5}$$

with $E_{x,y}$ the x and y components of the missing transverse momentum

$$E_{x,y}^{miss} = -\sum_{i \in \text{hard objects}} p_{x,y}^{i} - \sum_{j \in \text{soft signals}} p_{x,y}^{j}. \tag{6.6}$$

In general, overlaps of jets with electrons, muons or photons are taken into account and are corrected. The E_T^{miss} is of interest in this thesis because the leptonic final states in the $t\bar{t}H(b\bar{b})$ analysis contain neutrinos which are not detected and only appear as missing transverse momentum.

7

MACHINE LEARNING

The collection of large datasets requires sophisticated techniques to analyse those. In this context, Machine Learning (ML) is one of the fastest-growing fields in computer science allowing the program to learn patterns in a multi-dimensional phase space.
Especially high energy physics is well suited to apply ML techniques with its broad range of possible applications and its vast amount of labelled data (MC simulation). Already at detector level ML comes into play by deploying neural networks on FPGAs for the trigger [131] or to spot machine failures [132]. In the reconstruction step, new ML techniques are developed for charged particle tracking, in particular to cope for the increasing luminosity during RUN III of the LHC and the HL-LHC [133, 134]. ML is not only able to outperform conventional algorithms but is also able to mimic algorithms which cannot be replaced by ML per se. One example for this case is the detector simulation with GEANT4 [99, 135] which is one of the most CPU consuming tasks within ATLAS. As pointed out in Section 4.2, Generative Adversarial Networks (GANs) and Variational Auto-Encoders (VAEs) are studied to improve the fast calorimeter simulation [103]. Apart from the object reconstruction, also the object identification is improved using neural networks for instance in the τ-identification [136] or in b-tagging algorithms [137], which will be the main application discussed in this thesis. In physics analyses, sophisticated ML methods help to reconstruct and discriminate signal processes [104, 138].

7.1 GENERAL INTRODUCTION

Machine Learning is a very broad umbrella term covering all kinds of algorithms which are not per se optimised for a specific task but are flexible enough to adapt to different problem sets by tuning (training) their parameter set.
ML requires besides the model itself also preparation and follow-up processing steps. In which extent they are necessary always depends on the available data, the model and its later application. Figure 7.1 shows such an example workflow (the single steps are explained in more detail in the dedicated sections e.g. sec. 9.2).

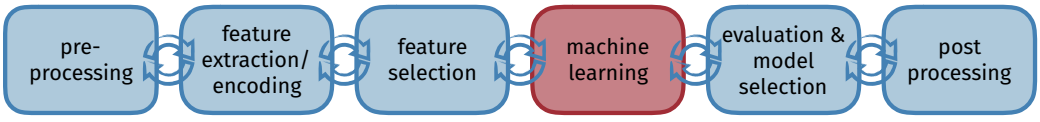


Figure 7.1.: Machine Learning workflow showing additional steps besides machine learning training itself.

Generally, two types of machine learning are distinguished: *Supervised learning* requiring fully labelled training data and *Unsupervised learning* not requiring any labelled data. There are also

intermediate approaches called *Semi-Supervised learning*. In the context of this thesis supervised approaches are used based on Neural Networks (NNs) and Boosted Decision Trees (BDTs).

In the following, a statistical parametric (ML) model is denoted as $P_{\text{model}}(\vec{x}_i; \theta)$ parametrised with the parameters θ while P_{data} is the true but unknown distribution. A data set of length N is given as $\vec{X} = (\vec{x}_1, \vec{x}_2, ..., \vec{x}_N)$, in which each data point i has a feature set $\vec{x}_i = (x_i^1, x_i^2, ..., x_i^M)$ with M features and true labels y_i in case of supervised learning.

7.1.1 *Dataset Handling*

It is important to ensure an unbiased training process. For this purpose, at least three orthogonal datasets are needed as indicated in Figure 7.2. The training sample is utilised for the actual algorithm training. The validation set is typically used to choose between different models and to optimise the model further such as hyperparameter optimisation. While for the training itself mostly a loss function is used (see sec. 7.2.3) to find the best parameter set, on the validation set, the performance measures dedicated for the problem set are evaluated (e.g. signal over background ratio) to fine-tune the model choice. The testing sample is only used to evaluate the final performance and is not involved in the training process. In the case of samples with low statistics, one can use cross-validation or also called k-folding [139] where pairs of training and test-/validation sets are partitioned into k subsets.

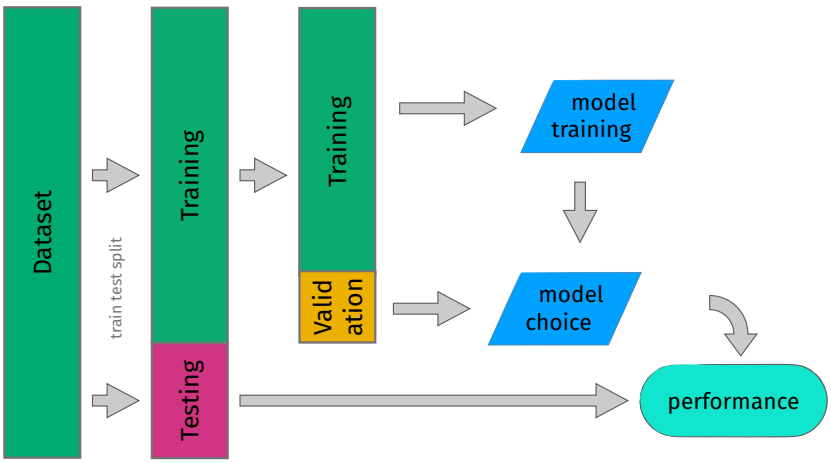


Figure 7.2.: Sketch of the dataset handling for ML training, validating and testing a model.

Typically, in particle physics the event number[1] variable is used to split the dataset into the training and testing set. The advantage is that at every point it is clear which events were used for the training.

1 The event number is a unique integer number associated to each event not correlated with any physical observable.

7.2 PERFORMANCE EVALUATION

Even though every ML application is different, the model performance is the decisive measure in the end. Depending on the task, different metrics are used to judge the performance. In the following, the most common approaches are discussed.

7.2.1 *Maximum Likelihood*

The Maximum Likelihood Estimation (MLE) is an approach to optimise parameters θ of a probability distribution using a likelihood function. A likelihood function $\mathscr{L}(\theta;\vec{X})$, also simply called *likelihood*, is a joint probability distribution of a finite data distribution $\vec{X}$ depending on parameters only. It provides a compatibility measure of the statistical, parametric model $P_{\text{model}}(\vec{x};\theta)$ to data for given values of the unknown parameters

$$\mathscr{L}(\theta;\vec{X}) = \prod_{\vec{x}\in\vec{X}} P_{\text{model}}(\vec{x};\theta). \tag{7.1}$$

The optimum is the maximised likelihood. It is more convenient to minimise the negative log-likelihood

$$\theta^* = \arg\min_{\theta}(-\ln\mathscr{L}(\theta;\vec{X})) = \arg\min_{\theta}\left(-\sum_{\vec{x}\in\vec{X}}\ln P_{\text{model}}(\vec{x};\theta)\right), \tag{7.2}$$

since the product of several terms much smaller than one, is numerically unstable to compute. Even though the solution can be found analytically in some cases, it is mostly computed numerically.

7.2.2 *Multi-Classification Likelihood Discriminant*

A multi-classification model typically has C outputs, one for each class c associating a score for being compatible to that specific class. The multi-classification allows a more detailed interpretation of the results than just a binary classification. Nonetheless, it is often useful to have one single discriminating variable.
Generally, the best discriminant for a neural network with a multiple-class output is given by a monotonically decreasing function combining all output nodes with a signal over background ratio, as an example shown for three classes which will be used later in this thesis

$$\frac{p_{\text{signal}}}{k_1\cdot p_{\text{bkg1}} + (1-k_1)\cdot p_{\text{bkg2}}}, \tag{7.3}$$

with p_{signal}, $p_{\text{bkg1,2}}$ the NN output node corresponding to the signal class and the two background classes, respectively. Hereby, k_1 is considered to be an effective parameter defined between 0 and 1 which allows tuning the relative performance/rejection of background 1. This is only valid if

the values of the output nodes sum up to one. By taking the logarithm of this function it is still monotonically decreasing and results in the following discriminant

$$\mathcal{D} = \log\left(\frac{p_{\text{signal}}}{k_1 \cdot p_{\text{bkg1}} + (1-k_1)\cdot p_{\text{bkg2}}}\right). \tag{7.4}$$

This likelihood discriminant has the advantage that it is tuneable and one can give the emphasis to better discriminate a specific background class.
By adding more output classes it is also necessary to add per additional class another effective parameter k_i satisfying the relation

$$\sum_i k_i = 1. \tag{7.5}$$

7.2.3 *Loss Function*

A loss function $\mathcal{J}$, also called *cost function*, quantifies the deviation of a model with respect to its true values and is minimised during the model training. For supervised learning, loss functions are typically functions of the target labels. The choice of the optimal loss function requires a profound understanding of the problem. Different choices of the loss function could yield other optimal solutions.
Generally, the combined loss of a dataset $\vec{X}$ can be calculated via the average of the individual losses of the single data points $\vec{x}_i$

$$\mathcal{J}(\theta;\vec{X}) = \frac{1}{N}\sum_{i=1}^{N}\mathcal{J}(y_i, P_{\text{model}}(\vec{x}_i;\theta)). \tag{7.6}$$

When dealing with regression problems, the most frequently used loss function is the mean square error (MSE)

$$\mathcal{J}_{\text{MSE}}(\theta;\vec{X}) = \frac{1}{N}\sum_{i=1}^{N}(y_i - P_{\text{model}}(x_i;\theta))^2. \tag{7.7}$$

In the case of binary classification, the negative loglikelihood of a Bernoulli distribution is used, the so-called *binary cross-entropy*

$$\mathcal{J}_{\text{binary-cross-entropy}}(\theta;\vec{X}) = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot \log(P_{\text{model}}(\vec{x}_i;\theta)) + (1-y_i)\cdot\log(1-P_{\text{model}}(\vec{x}_i;\theta)), \tag{7.8}$$

which can be extended to multi-classification with C classes

$$\mathcal{J}_{\text{categorical-cross-entropy}}(\theta;\vec{X}) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_c \log P_{\text{model}}(c|\vec{x}_i;\theta). \tag{7.9}$$

The probability of the data point being of class c is given by $P_{\text{model}}(c|\vec{x}_i;\theta)$ and y_c is a binary indicator specifying whether the class c matches the true class of $\vec{x}_i$.
Often, loss functions are also customised for specific tasks. Another loss is the exponential loss

$$J_{\text{exponential}}(\theta;\vec{X}) = \frac{1}{N}\sum_{i=1}^{N} e^{-y_i P_{\text{model}}(x_i;\theta)}, \tag{7.10}$$

which is of particular interest for the utilised BDT algorithm described in Section 7.4.

7.3 NEURAL NETWORKS

The concept of Neural Networks was already introduced in the 1940s [140] but only became feasible for wide applications with the easy access to large computing power, where especially GPUs were a big step forward. Neural networks are composed of artificial neurons connected via weights to each other, forming a network. The most basic network is a so-called *feed-forward network* as illustrated in Figure 7.3. This simplified example has one input layer, one hidden layer and one output node. Mathematically this network can be described as a matrix function

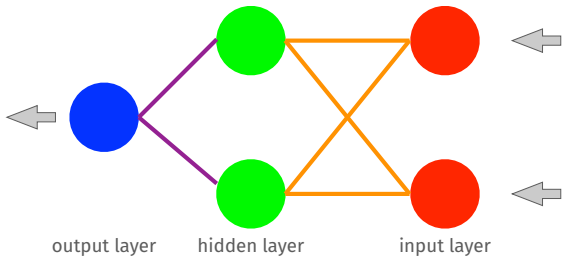


Figure 7.3.: Neural network with two input nodes in red, one hidden layer with two nodes in green and one output node in blue. Inspired from Ref. [141]. The sketch is inverted for illustration purposes to better match the form of Equation (7.11).

$$P^{NN}_{\text{model}} = f_2(b_2 + W_2 f_1(b_1 + W_1\vec{x})), \tag{7.11}$$

with weight matrices W_i, bias terms b_i and activation functions f_i. Expressing P^{NN}_{model} in full matrix notation yields

$$P^{NN}_{\text{model}} = f_2\left(\begin{bmatrix} b^2_{1,1} \end{bmatrix} + \begin{bmatrix} W^2_{1,1} & W^2_{1,2} \end{bmatrix} f_1\left(\begin{bmatrix} b^1_{1,1} \\ b^1_{2,1} \end{bmatrix} + \begin{bmatrix} W^1_{1,1} & W^1_{1,2} \\ W^1_{2,1} & W^1_{2,2} \end{bmatrix} \begin{bmatrix} x^1_{1,1} \\ x^1_{2,1} \end{bmatrix}\right)\right). \tag{7.12}$$

In this notation, it is easier to see that each layer is a linear system in the form $b + W \cdot x$, where the bias is a constant offset. Only the choice of the activation function (more details below) introduces a non-linearity. Consequently, a NN can approximate any arbitrary function by giving the network enough freedom (amount of hidden layers and nodes per hidden layer). The free parameters θ which can be optimised are the weights and the bias values. This simple example already has nine free op-

timisable parameters, more complex networks easily reach several ten-thousands of free parameters. In this thesis mainly deep feed-forward networks are utilised but there are a vast number of different neural network architectures available for every kind of application (see a comprehensive overview of modern machine learning in high energy physics in Ref. [142]).
The terminology of deep learning [143] is used for NNs with multiple hidden layers. Driven by the fast-evolving field and the large industry interest in deep learning, several sophisticated and user-friendly software packages are available, which make NNs accessible to a wide audience. In particular, Tensorflow [144], KERAS [145] and pytorch [146] are the most used and advanced software packages. However, there are many more different frameworks available. In order to easier exchange models between different tools, an open-source format ONNX [147] is used which also gives the future possibility to easier integrate ML models into ATLAS software infrastructure. For the time being, the JSON-based software package lwtnn [148] is used to deploy KERAS models in ATLAS software.

The training of NNs is performed in batches i.e. the training data is divided into equally sized segments. The weights of the NN are updated after every batch. It is important that every batch is an adequate representation of the full dataset, typically realised by shuffling the full dataset before slicing the sample. A full iteration over the entire dataset is called *epoch*.
In general, NNs have so-called *hyperparameters* which are all the non-trainable parameters fixing the architecture and training process of a NN. Besides the number of hidden layers and nodes per hidden layers, the most important hyperparameters are discussed in the following. Also, the batch-size is a hyperparameter which needs to be optimised.

Optimiser

The weights θ of NNs, and therefore the model itself, are optimised using gradient descent. This requires a loss function which has to be sub-differentiable with respect to the inputs $\vec{x_i}$ and parameters θ. After a first, usually random, initialisation of the weights θ_0, they are iteratively updated following the simplified formula

$$\theta_{i+1} = \theta_i - \lambda_i \nabla_\theta \mathcal{J}(\theta; \vec{X}), \tag{7.13}$$

where λ_i represents the learning rate which is tuneable and therefore a hyperparameter. Choosing the learning rate too large can prevent convergence because the optimisation is jumping over the minimum. On the contrary, a too low learning rate slows down the optimisation and the optimiser might get stuck in a local minimum. A learning-rate scheduler can prevent both extremes. In this thesis, the Adam optimiser [149] is used which uses estimates of the first and second moment of the gradient.

Backpropagation

For the gradient-descend method, it is required to be able to calculate the gradient of the loss function with respect to all trainable parameters. However, for NNs it is not feasible to calculate it analytically. The computational graph, which is the mathematical representation of the NN, is very complex and even shows certain discontinuities in some parts. Consequently, lots of nested gradients need to be calculated and the chain rule[2] needs to be employed to fully determine the gradient. To efficiently calculate the gradient, backpropagation [150] is used. So basically, it computes these nested gradients applying the chain rule gradually to the full computational graph. The gradient computation is optimised by reusing sub-expressions and keeping track of parameter dependencies.

Activation Functions

As previously mentioned the activation functions $f(z)$ are essential to allow the NNs to learn non-linear patterns. In the beginnings of NNs, simple step functions were used but then were gradually replaced by monotonically increasing functions such as tanh or logistic functions also denoted as *sigmoid*. These activation functions, however, suffer from vanishing gradient issues significantly slowing down the training. Nowadays, the Rectified Linear Unit (RELU) [151] activation function is widely used, defined as

$$f_{\text{RELU}}(z) = \begin{cases} 0 & \text{for} \quad z < 0 \\ z & \text{for} \quad z \geqslant 0 \end{cases}, \qquad f'_{\text{RELU}}(z) = \begin{cases} 0 & \text{for} \quad z < 0 \\ 1 & \text{for} \quad z \geqslant 0 \end{cases}, \tag{7.14}$$

with the simple derivative $f'(z)$ speeding up computations and not being affected by rapidly vanishing gradients. Besides RELU there are also other improved activation functions such as Leaky RELU [152] or *Softplus* [153].

In general, the output nodes are treated differently and a different activation function is applied compared to the hidden layers. For regression tasks, the linear activation function is used, whereas for classification problems the output should mostly be interpretable as probabilities. For the binary classification, this can be achieved using the *sigmoid* function

$$f_{\text{sigmoid}}(z) = \frac{1}{1 + e^{-z}}, \tag{7.15}$$

returning values between 0 and 1. A generalisation to multi-class classification with $C > 2$ classes c is the *softmax* activation function

$$f_{\text{softmax}}(z_c) = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_c}} \quad \text{fulfilling} \quad \sum_{c=1}^{C} f_{\text{softmax}}(z_c) = 1, \tag{7.16}$$

2 The chain rule states how to compute the derivative of composite functions which is summarised in this formula $\frac{dy}{dx} = \frac{dy}{dz}\frac{dz}{dx}$.

where the output for each class c can be interpreted as the probability of the data point being compatible with class c.

Regularisation

Besides the training performance, an important feature of a ML model is the ability to be generalisable and to not depend on fluctuations in the training data, i.e. avoid overfitting. In order to realise this, the capacity of the model needs to be sometimes limited producing a simpler and more robust model. In practice, mostly stochastic regularisation is used, such as Dropout [154], batch normalisation [155] or early stopping [156]. The Dropout method randomly drops a certain percentage of node connections to neighbouring layers avoiding complex neuron co-adaptions. The batch normalisation re-normalises and re-scales the values of a batch and the early stopping terminates the training process after certain criteria for instance that the loss is not decreasing over a certain amount of epochs.

7.4 BOOSTED DECISION TREES

BDTs were one of the most commonly used multivariate technique in the last years in high energy physics, before NNs became more and more popular, and still have their raison d'être. The availability of BDTs, inside the widely used statistics and data handling package ROOT [157], via the TMVA [158] package made it easily integratable in the analysis software workflows. Therefore, BDTs were deployed for various problem sets, such as in object identification [159] or discriminating signal and background processes in physics analyses.
A decision tree is structured as sketched in Figure 7.4 (left). It has, as the name suggests, a tree-like structure with branches connected via nodes. At each node, a cut decision based on a specific attribute is made. This is repeated until a stop criterion is met such as the maximal tree depth or the minimum events in a leaf node. A decision tree on its own is a weak learner and very sensitive to small changes in the training data.

Boosting allows to create from an ensemble of weak learners, single decision trees, a powerful and robust model illustrated in Figure 7.4 (right). In the scope of this thesis, the Adaptive Boost (AdaBoost) [161] from the TMVA [158] package is used. A boosting algorithm iteratively combines T single decision trees $P^{(t)}_{\text{model}}$ into a single discriminant

$$P^{\text{BDT}}_{\text{model}}(\vec{x}_i) = \sum_{t=1}^{T} \alpha_t P^{(t)}_{\text{model}}(\vec{x}_i), \tag{7.17}$$

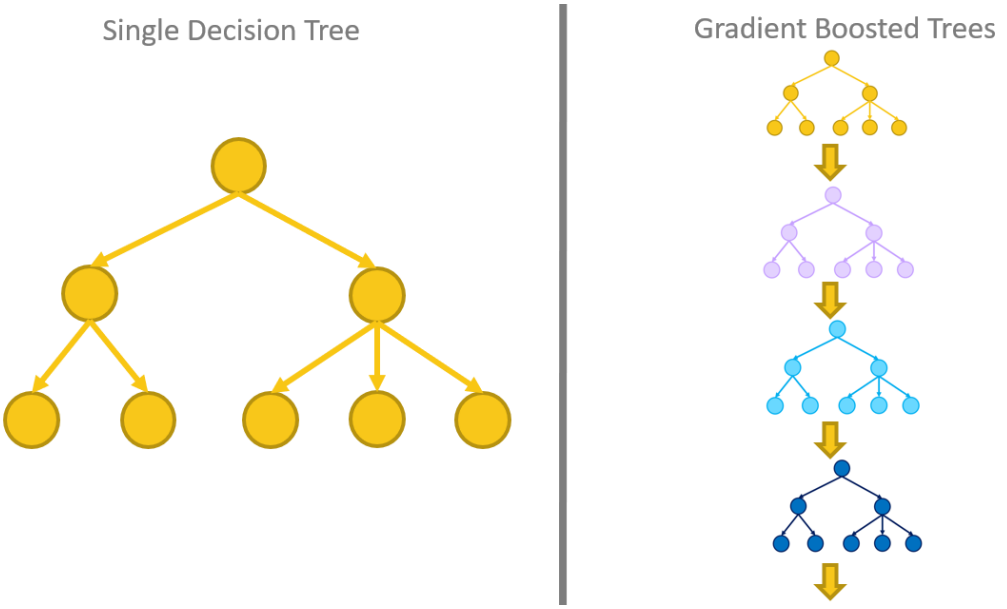


Figure 7.4.: Schematic view of a single decision tree on the left and an illustration of the boosting of trees on the right [160].

with α_t the weights associated to each decision tree. The value of the weight α_t is chosen for AdaBoost such that it minimises the loss function

$$\mathcal{J}_t = \sum_{i=1}^{N} \mathcal{J}_{\text{exponential}}(P_{\text{model}}^{\text{BDT},(t-1)}(\vec{x}_i) + \alpha_t P_{\text{model}}^{(t)}(\vec{x}_i)), \tag{7.18}$$

where the loss function $\mathcal{J}_{\text{exponential}}$ is the exponential loss from equation 7.10. In addition, from one iteration to another, wrongly classified training events get larger weights associated to be more sensitive to those in the following tree.

Nowadays more advanced BDT libraries are available with different boosting algorithms giving a better out-of-the-box performance. The most popular and successful libraries are XGBOOST [162], LIGHTGBM [163] and CATBOOST [164].

Part III

HEAVY-FLAVOUR TAGGING

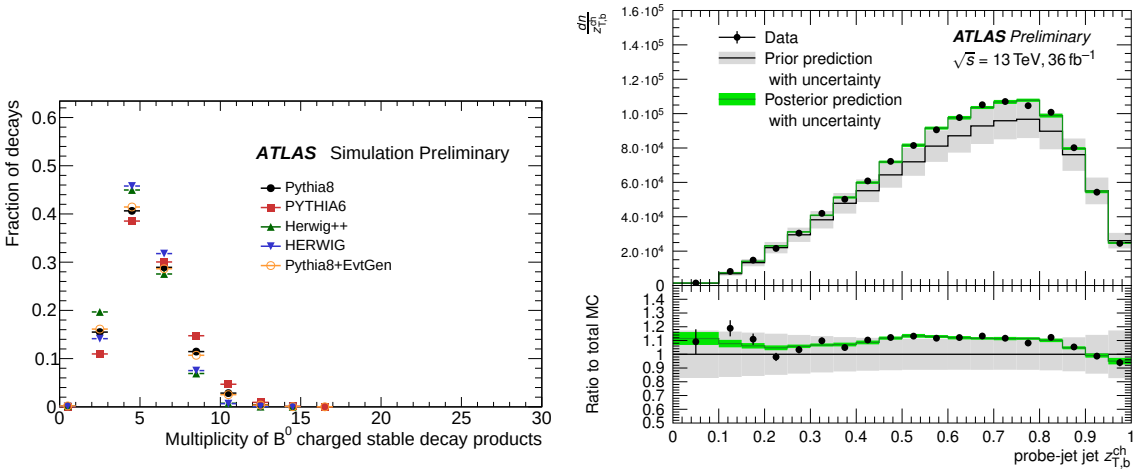
8

INTRODUCTION TO HEAVY-FLAVOUR TAGGING

Heavy-flavour identification, also called *tagging*, plays an important role in particle physics analyses. Several interesting physics processes, such as the $t\bar{t}H(b\bar{b})$ process described in Part IV or the $VH(H \to b\bar{b}, H \to c\bar{c})$ process [165, 166], have b- or c-quarks in their final state. Flavour tagging is a crucial tool to better select the signal and reject the background processes, and is therefore important for both searches and precision measurements.

Since quarks cannot occur as free particles, not the quarks itself are being tagged, but rather the colour neutral states, the b- and c-hadrons. A focus of this thesis is b-tagging in ATLAS and this chapter provides an overview of the techniques used in ATLAS.

The b-hadrons have a lifetime of around 1.5 ps [20], thus they decay only after 2.5 mm when carrying a momentum of 30 GeV. Moreover, b-hadrons have a relatively high mass of $\sim$ 5 GeV [20] as well as a high decay multiplicity of around an average of 5 stable, charged particles illustrated in Figure 8.1 (a) for B^0 hadrons. Another decay property is a relatively high semi-leptonic decay fraction. Figure 8.1 (b) illustrates the p_T fraction a b-hadron carries inside a jet which is large, i.e. peaking around 70%. The typical anatomy of a b-jet looks then as depicted in Figure 8.2 with a decay



(a) (b)

Figure 8.1.: (a) Decay multiplicity of the b-hadron B^0 into charged stable products compared for different MC generators [167]. (b) Jet-fragmentation of the transverse momentum component of b-hadrons with unfolded data [168].

length L of the b-hadron indicated in the red dotted line ending in the secondary vertex where the hadron decays. If the b-hadron decays into a c-hadron, even a tertiary vertex is present in the jet. The displacement of the secondary vertex with respect to the primary vertex is given by the decay length and the displacement of a track with respect to the PV is parametrised with impact parameters (IPs)

denoted as d_0 in the figure. Typically, the tracks coming from b-hadron decays have large d_0 due to the long decay length and the large mass of the b-hadron.
The c-hadrons have similar properties as b-hadrons, but a shorter lifetime, a lower decay multiplicity and are lighter and thus d_0 is smaller, which results in similar but not identical jet topologies of c- and b-jets, making it difficult to distinguish these two kinds of jets. Within the light-flavour jets, most of the tracks are originating directly from the quark fragmentation. This makes b-jets in general well separable from light-flavour jets. Nevertheless, displaced vertices can also occur coming from long-lived particles, photon conversion, poorly measured tracks or detector material interactions.

All these b-hadron/jet properties are targeted by different baseline algorithms extracting different

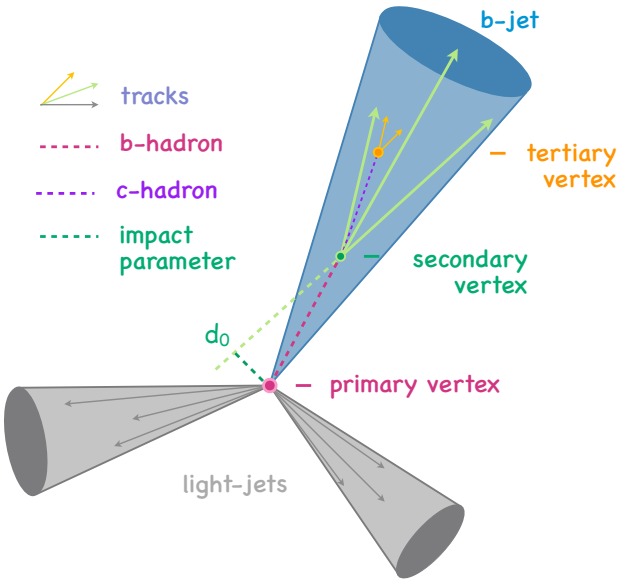


Figure 8.2.: Schematic view of a b-jet (in blue) including a secondary and a tertiary vertex. In gray light jets.

information for b-tagging. The schematic in Figure 8.3 shows the structure of b-tagging in ATLAS during RUN II [169]. The baseline algorithms can be categorised in three categories: the IP based algorithms IP2D/IP3D and RNNIP, the vertex-based algorithms SV1 and JETFITTER, and the soft-muon-based tagger SMT.
The information that the baseline algorithms provide is then combined by the so-called *high-level taggers* MV2 (based on BDTs) and the deep learning-based tagger DL1 which will be studied in detail in Chapter 9 - 12. The baseline algorithms are explained in more detail in the following, together with the description of the used variables for the high-level tagger DL1. All plots will be shown for *PFlow* jets in $t\bar{t}$ events if not differently stated. The *Variable Radius Track jets* distributions will be summarised in the dedicated chapter.

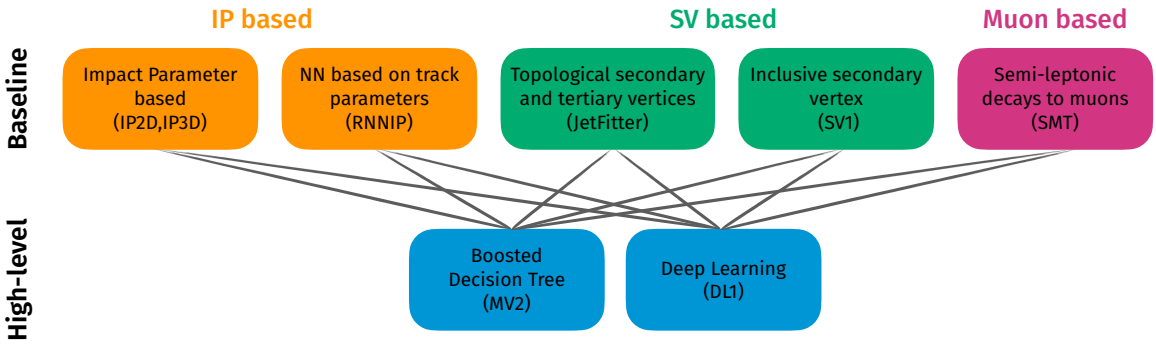


Figure 8.3.: Two level structure of the ATLAS heavy-flavour identification structured into baseline and high-level algorithms. The high-level algorithms combine the output of the baseline algorithms into a final discriminant.

8.1 TRAINING DATASET

All heavy-flavour taggers from RUN II were so far optimised for *EMTopo* jets as described in [169]. The retraining and optimisation of the DL1 tagger is performed in the scope of this thesis for *PFlow* jets and *VR Track* jets.

8.1.1 *Jet Flavour Labelling*

Since jets are clustered objects it is necessary to define a labelling scheme defining criteria when a jet is a b-jet, c-jet or light-flavour jet in the simulation. These so-called *truth-labels* are retrieved iteratively by matching hadrons with a minimum p_T of 5 GeV to the jets. A jet is denoted a b-jet if a b-hadron is within $\Delta R(\text{jet}, \text{b-hadron}) < 0.3$. In the next step, the ΔR matching is repeated for c-hadrons for the jets not labelled as b-jets and afterwards for τ-leptons. All remaining jets not classified as either b-, c- or τ-jets are categorised as light-flavour jets which include besides the light-quarks also gluons.
In addition, there is also an extended labelling available, providing more information about double-b, double-c and bc-categories.

8.1.2 *Modelling*

The algorithm studies are performed using $t\bar{t}$ events generated with POWHEGBOX v2 [88–92] at next-to-leading-order with the NNPDF3.0NLO PDF and h_{damp} =1.5 m_{top} [170]. The hadronisation, MPI and PS settings are described in Section 4.1.3.
The number of jets at high transverse momenta is small in $t\bar{t}$ samples. To populate this phase space with sufficient statistics, $Z' \to b\bar{b}/c\bar{c}/q\bar{q}$ samples are employed. The events are fully simulated with PYTHIA8 (A14 tune) and the leading-order PDF set NNPDF2.3LO. To get a flat jet p_T distribution also throughout the high jet p_T spectrum, the mass of Z' is varied and the natural width of the Z' resonance has to be artificially widened using dedicated weighting factors on an event-by-event basis.

Combined with a branching ratio of one third for each decay mode ($b\bar{b}$, $c\bar{c}$ and light-flavour quark pair) the jet p_T spectrum is equally flat and populated for the different quark classes at high transverse momenta (Fig. 9.3 shows this behaviour).
For both sample setups, the full detector simulation with GEANT4 is used.

8.1.3 *Selection*

The selection of *PFlow* jets and *VR Track* jets is slightly different due to the different clustering procedure. Common for both jet collections is the $|\eta| < 2.5$ requirement. For the $t\bar{t}$ sample lepton+jets and dileptonic events are taken into account (see Fig. 2.7).

8.1.3.1 *PFlow jets*

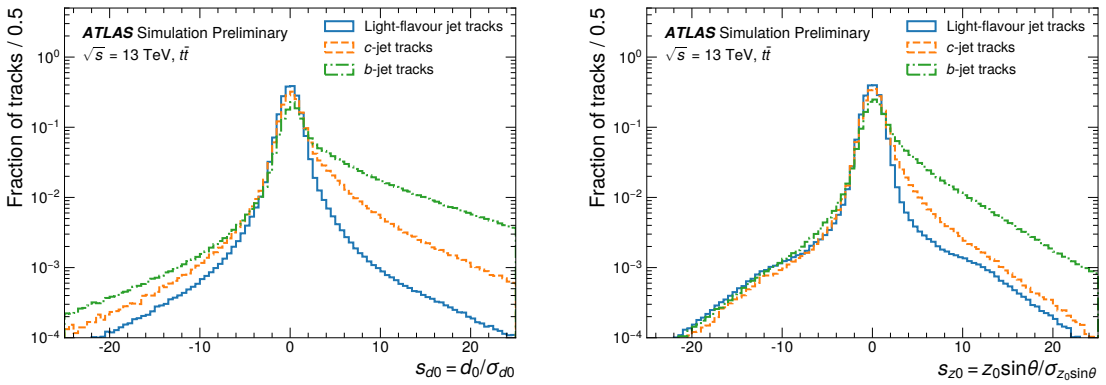
The jets which are re-clustered with the particle flow algorithm using a radius parameter of $R = 0.4$ are required to have a minimum jet p_T of 20 GeV. For jets in the range of $20\,\text{GeV} < p_T < 60\,\text{GeV}$ and $|\eta| < 2.4$, an additional jet vertex tagger (JVT) [171] cut JVT > 0.5 is applied for pile-up suppression. If the jet overlaps with an electron or muon which originates from a W-boson decay (this can only be identified at the generator level), it is removed.

8.1.3.2 *VR Track jets*

The jets clustered as *VR Track* jets need to satisfy $p_T > 10\,\text{GeV}$. Moreover, an additional overlap removal is applied, removing jets which coincide i.e. $\Delta R(\text{jet}_i, \text{jet}_j) < \min(R_{\text{jet}_i}, R_{\text{jet}_j})$.

8.2 IMPACT PARAMETER ALGORITHMS

The IP based algorithms are making use of the long decay path of the b-hadron which results in a displaced vertex. Therefore, the IP, which is the point of closest approach of tracks from the b-hadron decay to the primary vertex, is larger than for tracks coming from the PV. The IP can be split into two components, the transverse part d_0 (sketched in Fig. 8.2) and the longitudinal part $z_0 \sin\theta$. From these IP variables, the lifetime signed significances $s_{d_0} = d_0/\sigma_{d_0}$ and $s_{z_0} = z_0 \sin\theta/\sigma_{z_0 \sin\theta}$, shown in Figure 8.4, are calculated, corresponding to the IP divided by its uncertainty. The sign is determined by extrapolating the track to the PV. If the jet axis has to be extended from the PV backwards to cross the track, or its projection, a negative sign is assigned otherwise it is positive. The tracks used in the IP algorithms have to satisfy several track quality criteria. Tracks have to have $p_T^{\text{track}} > 1\,\text{GeV}$, the IPs have to fulfil $|d_0| < 1\,\text{mm}$ and $|z_0 \sin\theta| < 1.5\,\text{mm}$. Additionally, a requirement on the number of hits in the silicon layers is demanded $N_{\text{hits}}^{\text{Si}} \geqslant 7$ as well as an upper limit of silicon and pixel layer holes: $N_{\text{holes}}^{\text{Si}} \leqslant 2$ and $N_{\text{holes}}^{\text{pixel}} \leqslant 1$.



(a) (b)

Figure 8.4.: The signed IP significances shown for light-, c- and b-jet tracks for (a) the transverse and (b) the longitudinal component in $t\bar{t}$ events [137].

8.2.1 *IPxD*

The IPxD taggers [159] comprise two different algorithms: the IP2D which is only using the transverse IP d_0 since it is less prone to pile-up and the IP3D making use of both the transverse and longitudinal IP and their correlations. The track categorisation is based on pixel layer hit patterns defined by reference templates for b, c and light assuming the tracks to be uncorrelated. The final discriminant is a Log-likelihood ratio (LLR) of probabilities of the tracks being b-, c- or light-flavour like defined on jet level

$$\mathrm{IPxD}_{l,c,cl} = \sum_{i \in \mathrm{tracks}} \log\left(\frac{p^i_{b,b,c}}{p^i_{l,c,l}}\right). \tag{8.1}$$

The probability density functions to calculate p_b, p_c and p_l are extracted from reference histograms in MC simulations. Table 8.1 summarises the six different output variables of the IPxD tagger and Figure 8.5 shows the variable distributions.

Variable	Description
	The LLR based on the lifetime signed IP significance to separate:
IPxD_l	b- from light-flavour jets.
IPxD_c	b- from c-jets.
IPxD_{cl}	c- from light-flavour jets.

Table 8.1.: IPxD output variables (x stands for 2 or 3).

8.2.2 *RNNIP*

The second IP-based tagger is the Recurrent Neural Network IP tagger (RNNIP) [172]. It uses the same information as the IPxD algorithms requiring in addition at most one shared hit of multiple

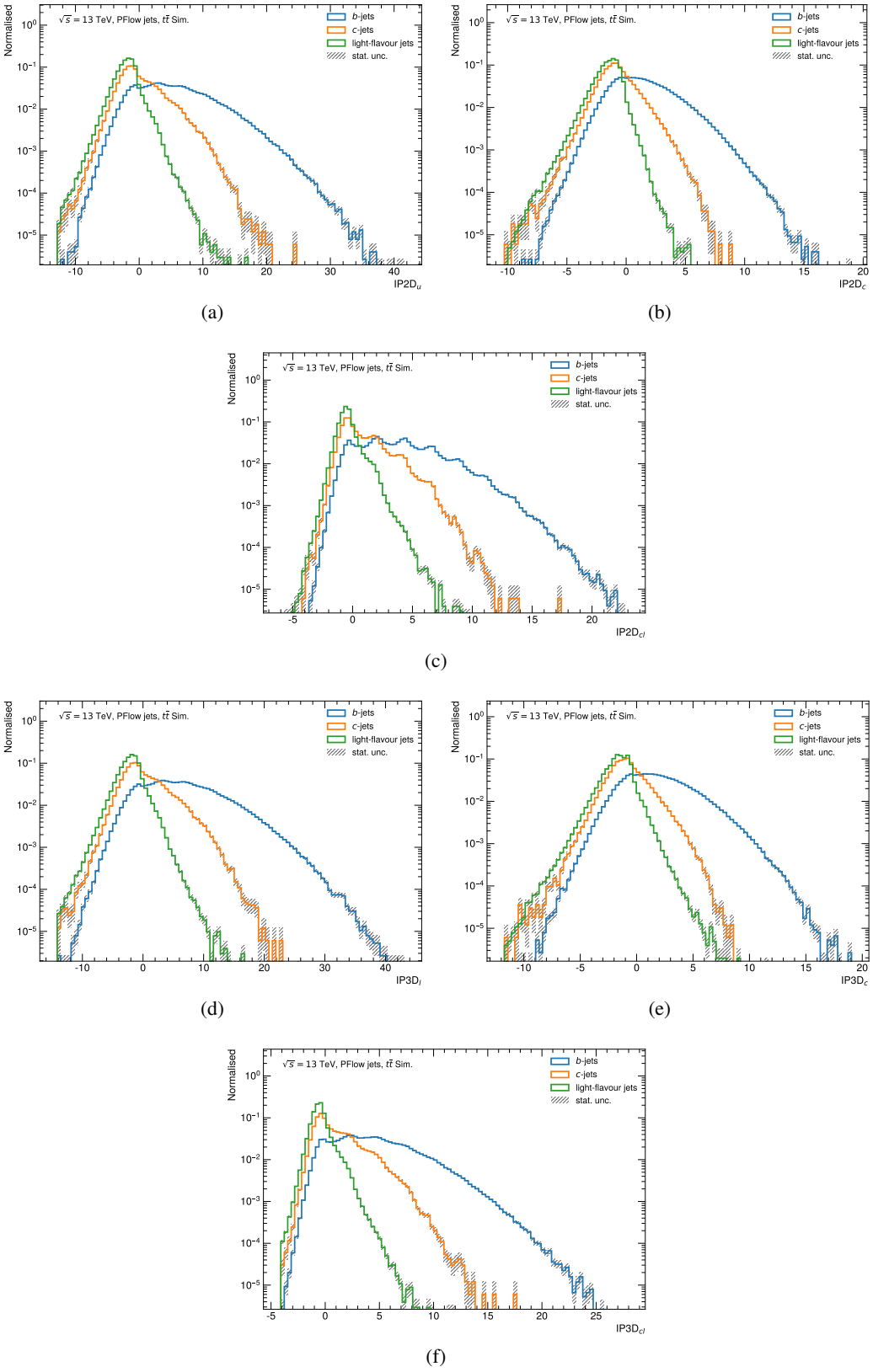


Figure 8.5.: Distributions of the LLR variables from the IP2D (a)-(c) and the IP3D (d)-(f) algorithms. Variables are defined in Table 8.1. Only jets where the algorithm succeeded are plotted.

tracks. A recurrent neural network (RNN) can process input sequences of variable length which is important since jets contain different amounts of tracks. The input sequence to the RNNIP is ordered by the lifetime signed transverse IP significance s_{d_0}. A so-called *long term short term memory* (LSTM) network is employed which allows to preserve correlations between the tracks. This is an important difference compared to IPxD. The other difference is the ability of RNNIP to use hit information instead of the category embedding from IPxD. As shown in [172] it outperforms IPxD especially for decay topologies with higher decay multiplicities and longer decay distances. The RNNIP tagger used for the DL1 training in this thesis (in Chapter 9–11) is an optimised version of the RNNIP in Ref. [172]. The optimisation was done separately for *PFlow* jets and *VR Track* jets. Due to its multi-class output, it provides three different probabilities per jet of being b-, c- or light-flavour like as shown in Table 8.2 and illustrated in the Figure 8.6. The spikes in the RNNIP_{p_l} (Fig. 8.6 (a)) and RNNIP_{p_b} distribution (Fig. 8.6(c)) arise from jets without any tracks. The jets without tracks, which correspond more or less to empty vectors, are also passed through the RNNIP network and thus the spikes occur at these values (~ 0.2 for p_l and ~ 0.7 for p_b) chosen by the network.
Furthermore, there is another tagger improvement made using a Deep Sets architecture resulting in the DIPS tagger [137]. This will be discussed again in more detail in Chapter 12.

Variable	Description
	The output nodes of the RNNIP tagger indicating the probability of the jet being a:
RNNIP_{p_l}	light-flavour jet.
RNNIP_{p_c}	b-jet.
RNNIP_{p_b}	c-jet.

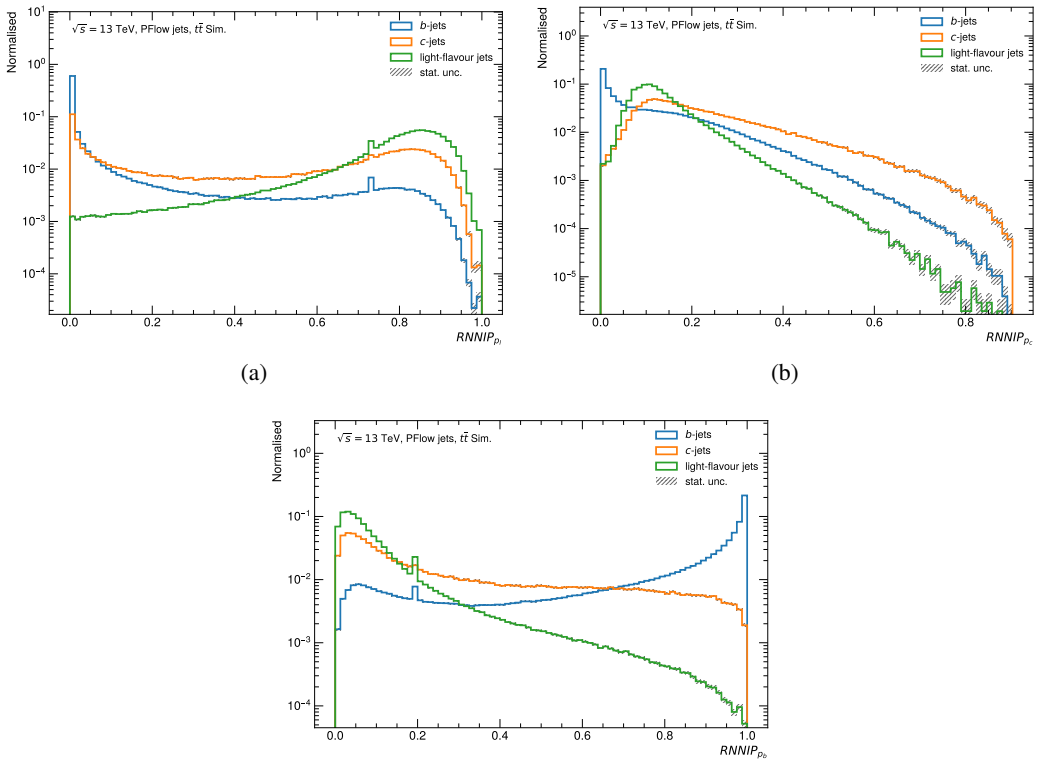
Table 8.2.: RNNIP output variables.

8.3 DISPLACED VERTEX RECONSTRUCTION

It can be extracted from the IP information if a displaced vertex is present in a jet. Nevertheless, it is important to know if it is a real vertex from a b- or c-hadron or originating from mis-reconstructed fragmentation tracks where light-flavour jets are faking b-jet topologies. It is also possible that tracks with low IPs can contribute to a displaced vertex. In ATLAS two different algorithms are used to identify these: the inclusive displaced secondary vertex reconstruction algorithm (SV1) [159, 173] and *JetFitter* [174] a decay chain multi-vertex reconstruction algorithm.

8.3.1 *Secondary Vertex Algorithm*

The SV1 algorithm tries to reconstruct a single displaced vertex in a jet using tracks. Due to the finite tracking resolution in ATLAS, it is not always possible to resolve the full decay cascade of b- or c-hadrons in every jet. This also means that not all decay vertices can be extracted. Therefore, the



(a) (b)

(c)

Figure 8.6.: Output nodes of the RNNIP multi-class network representing the probabilities for the jets being (a) a light-flavour jet, (b) a c-jet and (c) a b-jet.

reconstruction of only a single vertex, if possible, is a good approximation for b-jets.

In the first reconstruction step, all two-track vertices are matched together rejecting those which are compatible with tracks coming from long-lived particles (K_s, Λ), photon conversions or hadronic interactions with the material of the detector. In the next step, the two-track vertices are combined to form the secondary vertex (SV) while all tracks which are not consistent with the SV are removed.

Variable	Description
$N^{SV1}_{TrkAtVtx}$	Number of tracks associated to the SV.
$N^{SV1}_{2TrkVtx}$	Number of reconstructed two-track vertices candidates within the jet.
m^{SV1}_{inv}	Invariant mass of the SV calculated from the associated tracks.
f^{SV1}_{E}	Energy fraction of the SV associated tracks with respect to all tracks of the jet.
$\Delta R(\text{jet, SV})$	ΔR between the jet axis and the direction of the secondary vertex relative to the primary vertex.
L^{SV1}_{xy}	Reconstructed SV transverse decay length.
L^{SV1}_{xyz}	Reconstructed SV decay length.
S^{SV1}_{xyz}	Decay length significance, $L^{SV1}_{xyz}/\sigma_{L^{SV1}_{xyz}}$.

Table 8.3.: Overview of variables extracted from the SV reconstructed with the SV1 algorithm.

From the reconstructed SV, several important properties can be extracted such as the vertex mass, the decay length and its significance or the number of associated tracks and two-track vertices as well as the ΔR between the jet and the SV. All properties are summarised in Table 8.3 and the corresponding distributions are shown in Figure 8.7.

8.3.2 *Decay chain multi-vertex reconstruction*

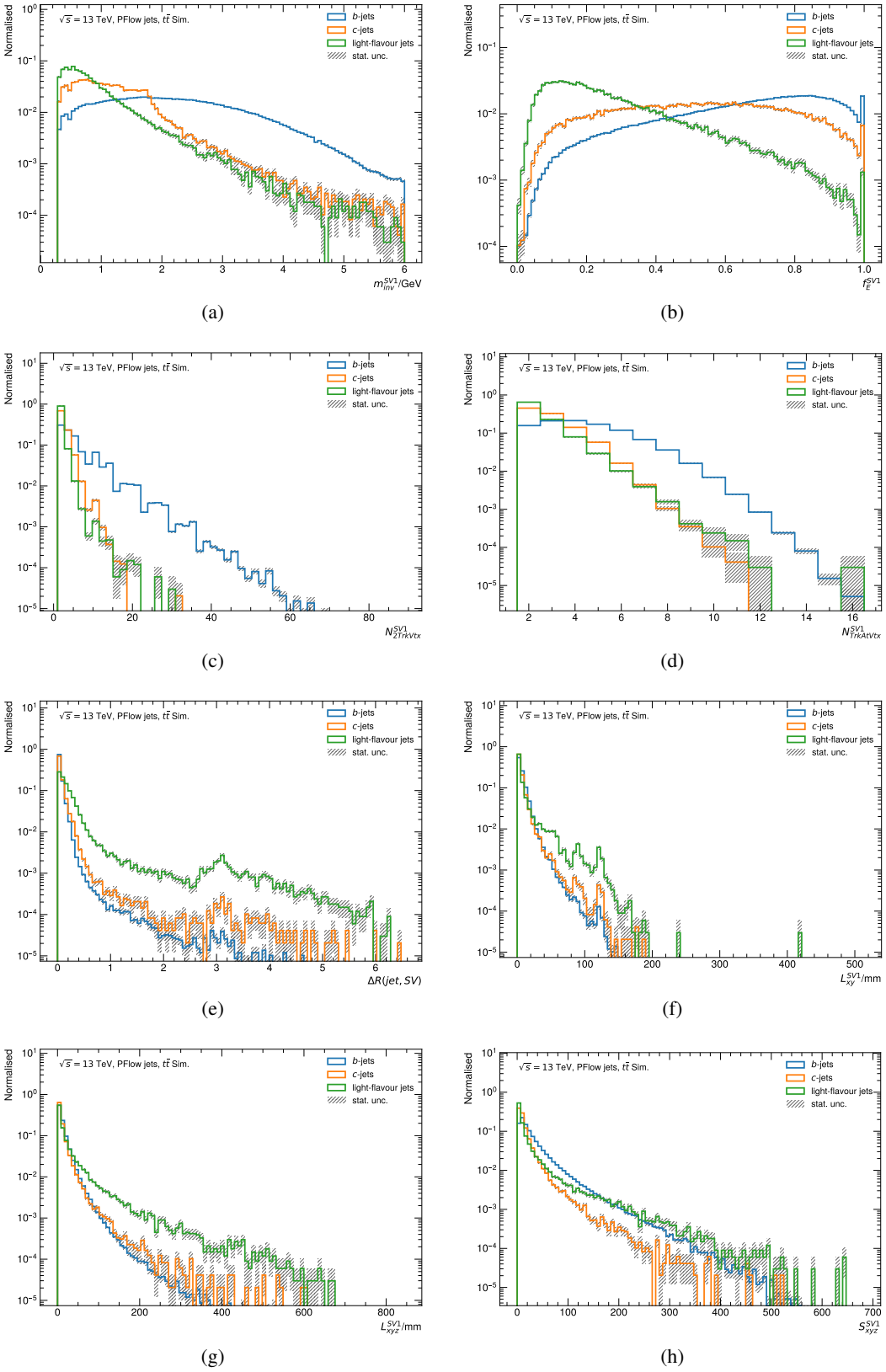
The second displaced vertex finder is JETFITTER (JF) [169, 174] which reconstructs the decay cascade topology of weakly decaying b- and c-hadrons. It assumes that the primary, secondary and tertiary vertices are aligned in one line, in the flight direction of the heavy-flavour hadron. This assumption helps to better cope with the finite detector resolution and also allows e.g. the reconstruction of single track vertices. After an initial track selection removing tracks associated with the PV, the axis through the vertices is retrieved using a modified Kalman Filter [175]. The resulting topology is characterised by several variables listed in Table 8.4 and shown in Figure 8.8.

Variable	Description
m^{JF}_{inv}	Invariant mass of tracks associated to one or more displaced vertices.
f^{JF}_{E}	Charged jet energy fraction in the secondary vertices.
S^{JF}_{xyz}	Decay length significance of the displaced vertex.
$N^{JF}_{\text{1-trk vertices}}$	Number of 1-track displaced vertices.
$N^{JF}_{\geqslant\text{2-trk vertices}}$	Number of vertices with more than one track.
$\Delta R^{JF}(\vec{p}_{jet}, \vec{p}_{vtx})$	ΔR between the jet axis and the vectorial sum of all track momenta associated to displaced vertices.
N^{JF}_{trks}	Number of tracks associated to SV.
$N^{JF}_{vertices}$	Number of reconstructed displaced vertices.

Table 8.4.: JETFITTER variable overview.

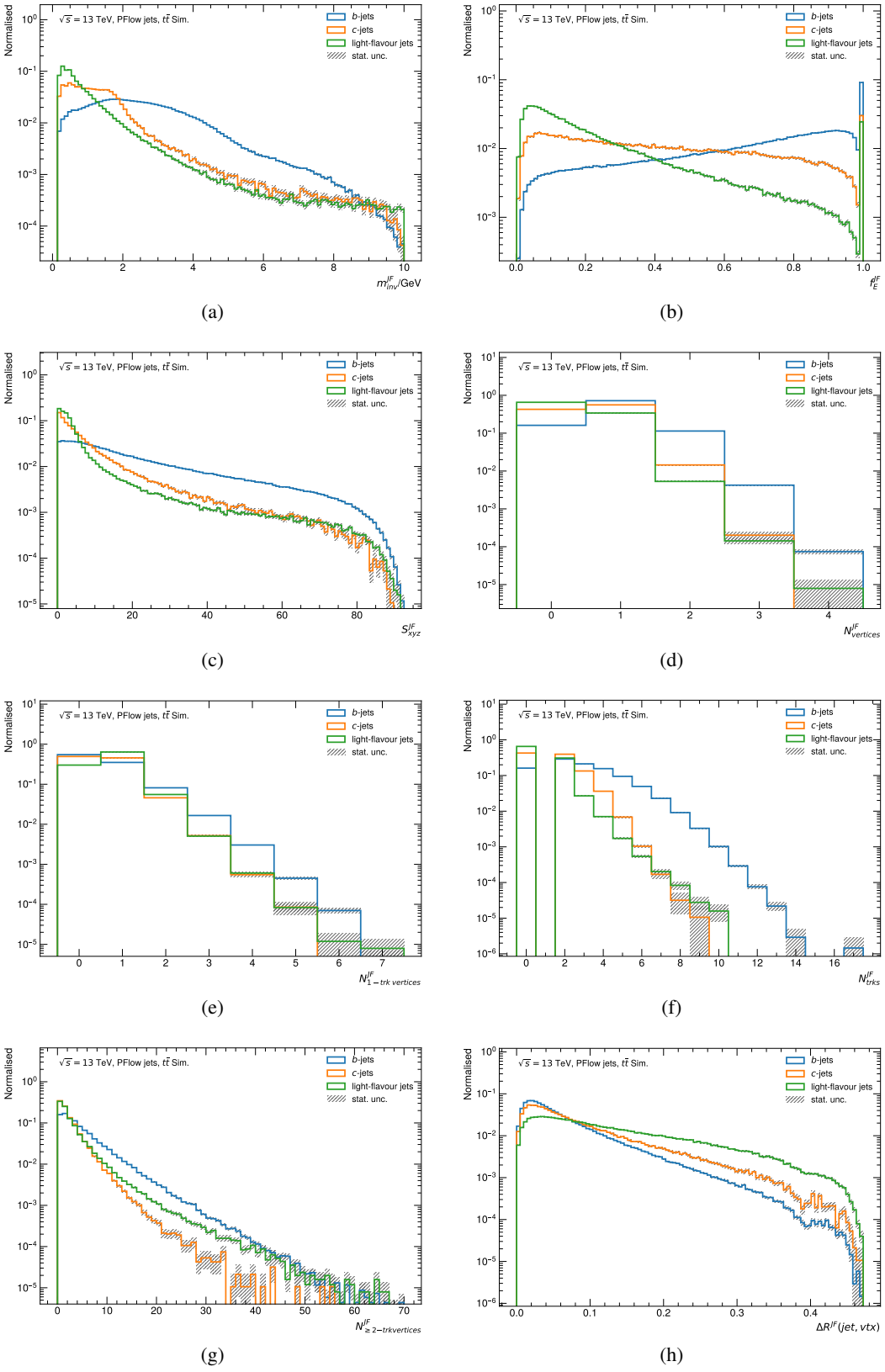
Additionally, special variables for the c-hadron identification are extracted using only the JETFITTER vertex which is closest to the PV. The variables, as shown in Table 8.5 and Figure 8.9, are chosen such that they make use of the different topologies of b- and c-hadron decays. Typically, only a single SV is present in a c-jet justifying the choice of only considering one SV in the calculation of the variables. Moreover, c-hadrons have a lower decay multiplicity due to their lower mass and thus the average energy that a single decay product carries is larger compared to b-hadrons. Consequently, also the rapidity, as defined in Equation (3.3), with respect to the jet axis is larger, visualised in Figure 8.9 (g)-(i).

Recently a new approach is being studied, using Graph Neural Networks [176] to identify the secondary vertex [177], which is showing promising results.



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 8.7.: Variable distributions of the reconstructed SV with SV1 as defined in Table 8.3. Only jets where the algorithm succeeded are plotted.



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 8.8.: JETFITTER variable distributions with the variables from Table 8.4. Only jets where the algorithm succeeded are plotted.

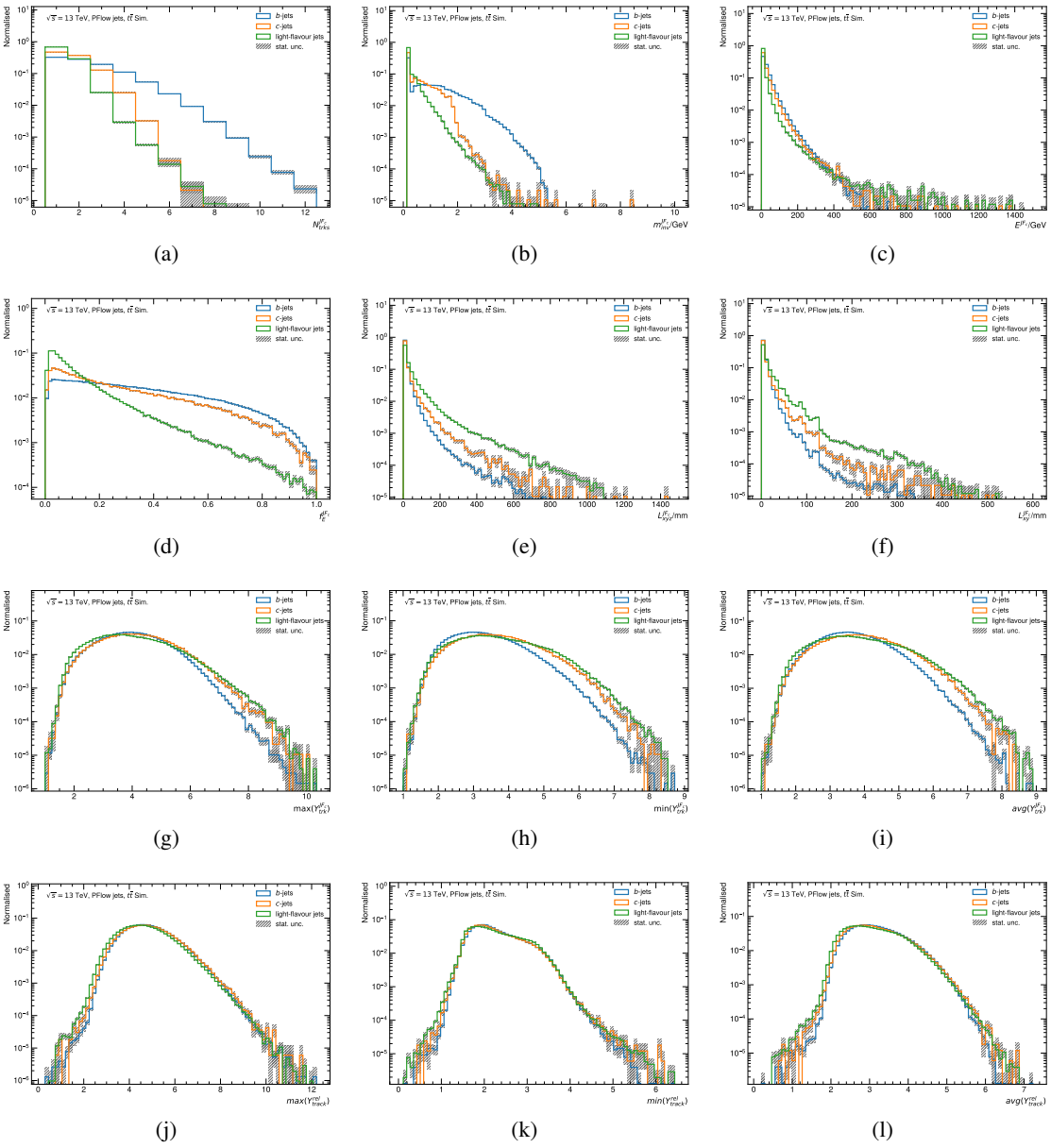


Figure 8.9.: Distributions of JETFITTER variables optimised for c-identification as shown in Table 8.5. Only jets where the algorithm succeeded are plotted.

Variable	Description
$L_{xyz}^{JF_c}$	Displacement of SV from PV.
$L_{xy}^{JF_c}$	Transverse displacement of SV from PV.
$\min(Y_{trk}^{JF_c})$, $\max(Y_{trk}^{JF_c})$, $\text{avg}(Y_{trk}^{JF_c})$	Minimal, maximal and average rapidity of tracks within the jet.
$\min(Y_{trk,\,SV}^{JF_c})$, $\max(Y_{trk,\,SV}^{JF_c})$, $\text{avg}(Y_{trk,\,SV}^{JF_c})$	Minimal, maximal and average rapidity of SV tracks.
$m_{inv}^{JF_c}$	Invariant mass of tracks associated to SV.
E^{JF_c}	Energy of tracks associated to SV.
$f_E^{JF_c}$	Charged jet energy fraction of SV w.r.t all tracks in jet.
$N_{trks}^{JF_c}$	Number of tracks associated to SV.

Table 8.5.: Overview of the JETFITTER variables optimised for c-jet identification [169].

8.4 SOFT MUON TAGGER

As mentioned above, the b-hadrons have a large semi-leptonic decay branching ratio of $\mathcal{BR}(B \to \mu\nu X) \approx 11\%$ and $\mathcal{BR}(B \to C \to \mu\nu X) \approx 10\%$. The Soft Muon Tagger (SMT) [169, 178] is designed to exploit the properties of the muons coming from heavy-flavour decays. The term *soft* is used even though the muon typically carries a non-negligible p_T together with a large relative p_T^{rel} with respect to the jet axis, but is soft compared to muons coming from electroweak boson decays. The usage of the tagger is, however, limited since it is only applicable for semi-leptonic decays and limited by the muon reconstruction and association to jets. Nonetheless, it is a useful supplement next to the IP and SV algorithms.

Originally a BDT was used to combine several variables describing the muon topology [178]. These variables are listed in Table 8.6 and illustrated in Figure 8.10. For the recommendations for *PFlow* jets and *VR Track* jets the d_0 of the soft muon was corrected so that its sign calculation also takes into account the jet axis. In addition, the BDT was replaced by a Neural Network and the longitudinal IP as well as the IP significances of the soft muon were added in the training. The multi-class output of the NN is summarised in Table 8.7 and shown in Figure 8.11.

Variable	Description
$\Delta R(\mu, \text{jet})$	ΔR between the soft muon and the jet
p_T^{rel}	Transverse momentum of muon relative to the jet axis.
$\mathcal{S}$	scattering neighbour significance
$\mathcal{M}$	momentum imbalance significance
$\mathcal{R}$	charge-to-momentum double ratio of ID and MS
$d_0(\mu)$	transvere soft muon impact parameter
$z_0\lvert\sin\theta\rvert(\mu)$	longitudinal soft muon impact parameter
$s_{d_0}(\mu)$	significance of transvere soft muon impact parameter
$s_{z_0}(\mu)$	significance of longitudinal soft muon impact parameter

Table 8.6.: Overview of SMT variables describing the soft muon topology.

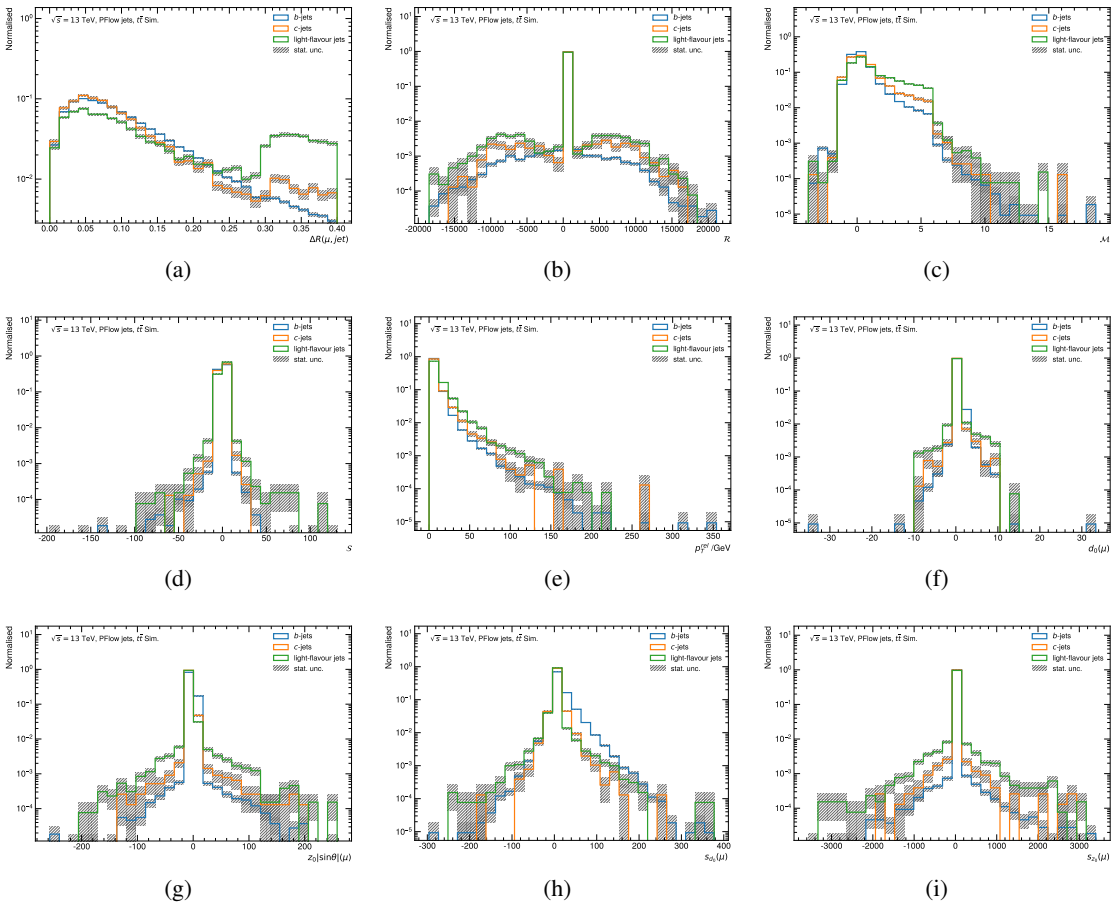
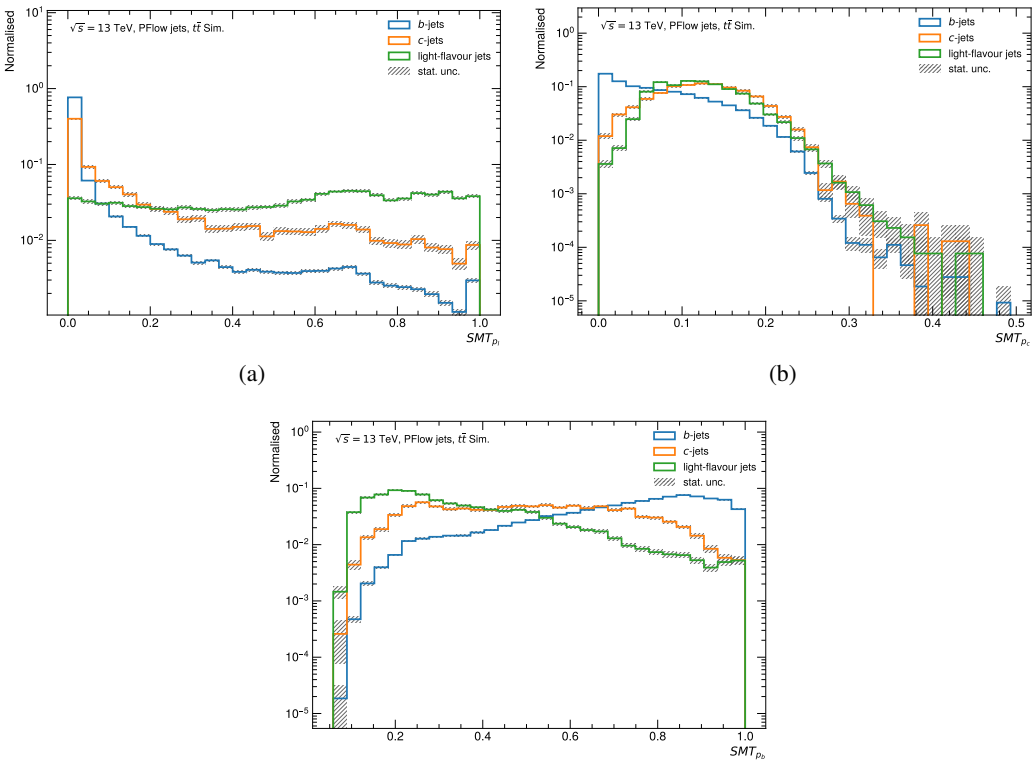


Figure 8.10.: SMT input variable distributions as defined in Table 8.6. Only jets where the algorithm succeeded are shown.

Variable	Description
	Output nodes of the SMT NN tagger indicating the probability of the jet being a:
$\mathrm{SMT}_{\mathrm{p_l}}$	light-flavour jet.
$\mathrm{SMT}_{\mathrm{p_c}}$	b-jet.
$\mathrm{SMT}_{\mathrm{p_b}}$	c-jet.

Table 8.7.: SMT NN output variable distributions. Only jets where the algorithm succeeded are shown.



(a) (b)

(c)

Figure 8.11.: Output nodes of the SMT multi-class network representing the probabilities for the jets being (a) a light-flavour jet, (b) a c-jet and (c) a b-jet.

8.5 HIGH-LEVEL TAGGERS

The baseline taggers described in sections 8.2 – 8.4 are only optimised for specific b-hadron decay properties. In order to combine this information into a single powerful discriminant, so-called *high-level taggers* are designed. As indicated in the schematic 8.3, two different high-level taggers are employed in ATLAS: the BDT based tagger MV2 and the Deep Learning based heavy-flavour tagger (DL1).
The MV2 tagger will be briefly described in Section 8.5.2 and will be employed in the analysis of the $t\bar{t}H(b\bar{b})$ process in Part IV. The DL1 tagger will be studied in full detail in the following chapters.

8.5.1 *Working points*

Ideally, the full spectrum of the final b-tagging discriminant would be calibrated and used in the physics analyses. This continuous calibration would require a separate calibration in very fine efficiency bins leading to an immense complexity and necessary workload which is not feasible to do in the required time scales in ATLAS. Therefore, four different b-tagging working points (WPs) are defined covering various needs of the physics analyses. The efficiency of a specific flavour j (b, c or light) is defined as

$$\varepsilon^{j} = \frac{N^{j}_{\text{pass}}(\mathcal{D} > T_f)}{N^{j}_{\text{total}}}, \tag{8.2}$$

where $N^{j}_{\text{pass}}(\mathcal{D} > T_f)$ are the number of jets of flavour j passing the cut T_f on the tagger discriminant $\mathcal{D}$ and N^{j}_{total} are all jets of flavour j before the cut. The WPs are defined based on the b-jet efficiency $\varepsilon^{\text{b-jet}}$ evaluated on a $t\bar{t}$ sample. The WPs used in ATLAS are listed in Table 8.8. Each WP is characterised besides the b-jet efficiency by its c- and light-flavour jet rejection which is the inverse of the efficiency $\varepsilon^{c,\text{light}}$. The misidentification improves with lower signal efficiency and thus rejects more background with a trade-off against lower signal statistics.

Description	**b-jet efficiency**
loose	85%
medium	77%
tight	70%
very tight	60%

Table 8.8.: Summary of the single cut WPs for b-tagging in ATLAS.

b-jet efficiency	**tag score**
[85, 100]%	1
[77, 85]%	2
[70, 77]%	3
[60, 70]%	4
[0, 60]%	5

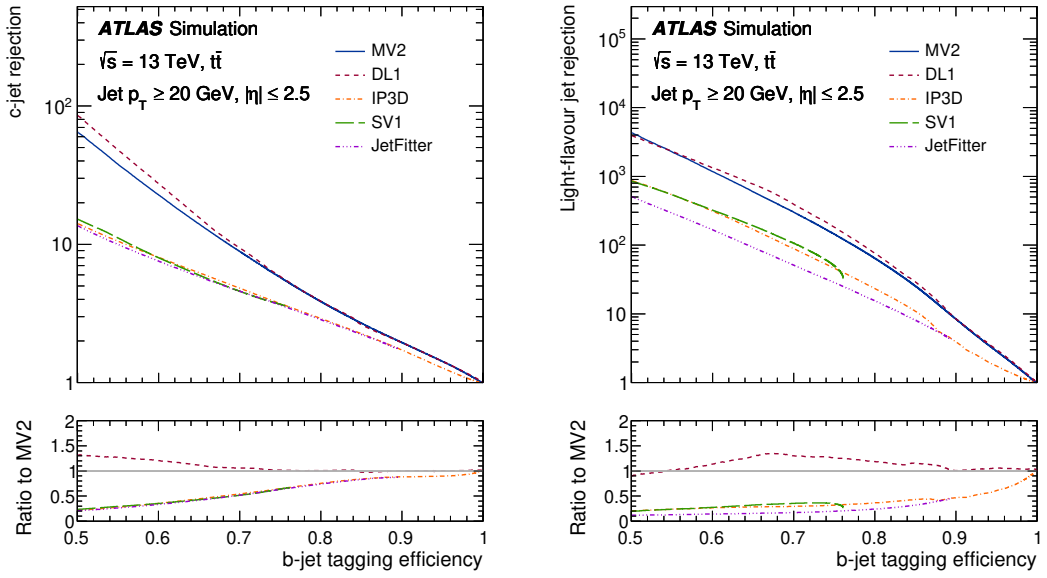
Table 8.9.: Pseudo-continuous WP in ATLAS with the corresponding tag score.

Every jet passing the criteria of a WP is then defined as a b-jet in physics analyses. Additionally, a pseudo-continuous WP is defined with five tag score bins. The tag scores are retrieved based on the single cut WPs listed in Table 8.9. The tag score indicates which b-tagging WP a jet passes. Most of the c- and light-flavour jets end up in the first bin [85, 100]% which is equivalent to not passing

any b-tagging WP, while the last bin $[0, 60]\%$ mainly contains b-jets which corresponds to passing the tightest (60%) WP. The pseudo-continuous b-tagging WP provides important additional information compared to the single cut WPs. It is a kind of an indicator of the quality of the b-jet, which is especially useful for multivariate (Machine Learning) methods. The b-tagging requirement can be loosened and by providing the pseudo-continuous tag score as an additional input, the information is retained while gaining more statistics.

8.5.2 *MV2*

The baseline MV2 algorithm, which used to be the recommended tagger for *EMTopo* jets for RUN II analyses, uses 24 variables as input for the BDT training. These 24 variables comprise kinematic information p_T and $|\eta|$ as well as the output of the baseline taggers IPxD, SV1 and JETFITTER. Figure 8.12 shows the performance overview of MV2, DL1 and their baseline taggers in terms of the background (c- and light-flavour jets) rejection as a function of the b-jet efficiency. The training



(a) (b)

Figure 8.12.: Comparison of b-tagging algorithms optimised and evaluated on *EMTopo* jets. Plot (a) shows the c-jet rejection and (b) the light-flavour jet rejection. The baseline taggers IP3D, SV1 and JETFITTER are shown as well as the two high-level taggers MV2 and DL1 compared in the ratio in the lower pads w.r.t MV2 [179].

of MV2 was performed with a so-called *hybrid sample* which is a mixture of a $t\bar{t}$ and Z' events to cover a large p_T spectrum. The BDT uses b-jets as signal class and c- and light-flavour jets as one background class. To balance the c versus light-flavour performance, a c-jet fraction of 7% and a light-flavour jet fraction of 93% was found to be best at least for a wide range of analyses. This tagger is called MV2c10.

In addition, different sets of input variables were studied adding RNNIP or SMT information. Fur-

thermore, a dedicated c-tagger is constructed with the JETFITTER c-variables and c-jets as signal class in the training.

8.6 CALIBRATION AND MONTE CARLO CORRECTIONS

The heavy-flavour tagging algorithms are optimised on MC simulation. Ideally, data and MC would show consistent flavour efficiencies. However, the simulation typically deviates from data due to detector and modelling effects, and correction factors (scale factors) need to be retrieved including their uncertainties. The calibration is realised for the four single cut WPs in Table 8.8 as well as for the pseudo-continuous tag scores [180–183].
The scale factors

$$\kappa^{j}_{\text{data-MC}} = \frac{\varepsilon^{j}_{\text{data}}}{\varepsilon^{j}_{\text{MC}}}, \tag{8.3}$$

are measured for all three flavours j. These scale factors are provided per jet and they are then combined to get a b-tagging weight defined on event level which can be used in the physics analyses. For b-jets the b-jet efficiency is calibrated while for c- and light-flavour jets the mistag rate is calibrated. All three calibrations are performed independently in data samples which are enriched in the specific jet flavour. The b-jet efficiency is determined on dileptonic $t\bar{t}$ events [179] while for the calibration of the c-jets mis-tag rate lepton+jets $t\bar{t}$ events are employed, investigating the hadronic W-boson decays to c-jets [183].
If a tagger performs very well and the light-flavour rejection is too high, it is not possible to calibrate the light-flavour mistag rate without large uncertainties on the scale factor due to unavoidable heavy-flavour contamination of the light-flavour sample. To overcome this, so-called *flipped taggers* are investigated. The goal is to have a flipped tagger which has a much smaller b-jet efficiency with an unchanged light-flavour mistag rate. So, the flipped tagger can be calibrated and the scale factors then propagated to the non-flipped tagger. The flipping itself is realised by inverting the sign of the s_{d_0} variable (see Fig. 8.4). This *negative-tag method* is described in more detail in Ref. [182] using Z+jets events.

In addition to the data to MC calibrations, differences between different MC generators also have to be corrected for [184]. The MC to MC scale factors $\kappa_{\text{MC-MC}}$ are calculated as a function of the jet p_T. They can be retrieved comparing the nominal MC generator (POWHEGBOX + PYTHIA8) with the alternative generator

$$\kappa_{\text{alternative}}(p_T) = \frac{\varepsilon_{\text{data}}(p_T)}{\varepsilon_{\text{nominal MC}}(p_T)} \frac{\varepsilon_{\text{nominal MC}}(p_T)}{\varepsilon_{\text{alternative MC}}(p_T)} = \frac{\kappa_{\text{data-MC}}(p_T)}{\kappa_{\text{MC-MC}}(p_T)}. \tag{8.4}$$

9

DEEP LEARNING BASED HEAVY-FLAVOUR TAGGER

The Deep Learning based heavy-flavour tagger (DL1) was introduced during RUN II as a second high-level heavy-flavour tagger, besides MV2. The tagger is designed to combine the information of the baseline taggers (introduced in Chapter 8) into a final discriminant. Studies to include more basic detector-level variables rather than by human brain designed observables are also ongoing, shown in the outlook in Chapter 12, including track and hit based information directly in the DL1 training.
The performance of the first version of DL1 [185] optimised for *EMTopo* jets is shown in Figure 8.12. The overall performance of DL1 is slightly better than MV2. The main advantage of DL1 is its multi-class output which means that the network predicts for every jet the probabilities for being compatible with the three main flavour classes: b-jets, c-jets and light-flavour jets. In general, this can be also realised with a BDT, however, NNs are more flexible and have more possibilities to customise their structure.

In this thesis, the DL1 algorithm is being re-optimised and adapted for *PFlow* jets and *VR Track* jets introducing a new machine-learning workflow for the flavour tagging group in ATLAS. The general design will be introduced in Section 9.1 followed by the description of the preprocessing in Section 9.2. The dedicated optimisation for *PFlow* jets is shown in Chapter 10 and for *VR Track* jets in Chapter 11.

9.1 GENERAL DL1 DESIGN

The underlying NN structure of the DL1 tagger is a deep feed-forward neural network with three output nodes corresponding to the b-, c- and light-flavour jet probabilities illustrated in Figure 9.1. The RELU activation function is used for each hidden layer and the last (output) layer makes use of the softmax activation function, such that the resulting network scores can be interpreted as probabilities. The final output score is calculated from the multi-class output as described in Equation (7.4) and results for the b-tagging discriminant into the log-likelihood

$$\mathcal{D}_b(f_c) = \log\left(\frac{p_b}{f_c \cdot p_c + (1-f_c)\cdot p_l}\right), \tag{9.1}$$

with p_b, p_c and p_l being the probabilities for the jet to be a b-jet, c-jet or light-flavour jet, respectively. The c-jet fraction f_c allows to tune how much emphasis is given to the c-jet or to the light-flavour performance (rejection). While the c-jet rejection increases as a function of f_c, the light-flavour jet rejection decreases. This parameter has to be tuned separately for each tagger and depends on the needs of the physics analyses. The advantage compared to MV2 is that this tuning

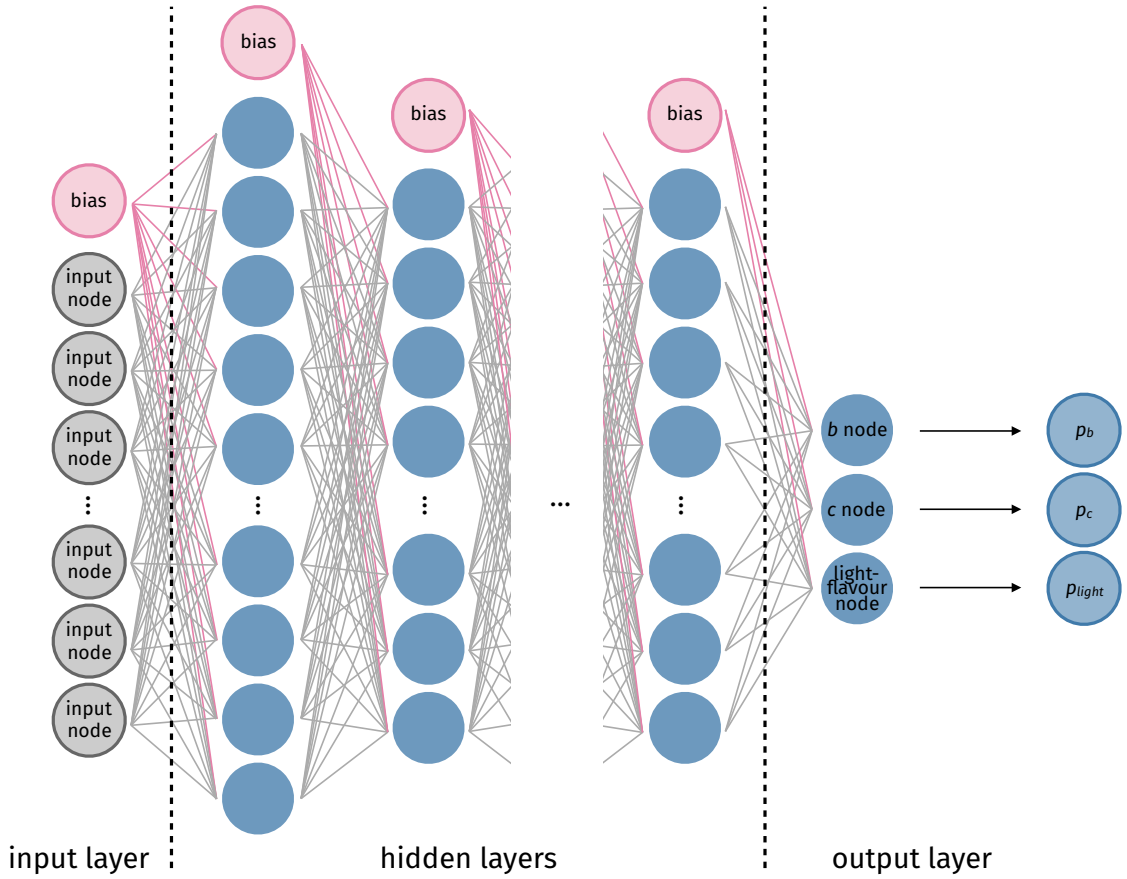


Figure 9.1.: Neural Network structure of the DL1 tagger.

is possible after the training and the c-jet fraction in the training sample does not have to be adapted. Another advantage of the multi-class output is that one can by changing the log-likelihood to

$$\mathcal{D}_c(f_b) = \log\left(\frac{p_c}{f_b \cdot p_b + (1 - f_b) \cdot p_l}\right), \tag{9.2}$$

perform c-tagging without the need of retraining the tagger. Here f_b is now the b-jet fraction. These possibilities have far-reaching positive effects on the workflow within ATLAS. First of all, less person power is necessary since only one tagger has to be trained and maintained. Also, fewer variables have to be calculated and stored in the files used for the physics analyses saving computing and storage resources.

DL1 is a family of three different taggers illustrated in Figure 9.2: *baseline DL1*, *DL1r* and *DL1rmu*. They differ in their input variables used for the NN training. The *baseline DL1* uses the same variables as MV2 with the additional JETFITTER variables optimised for c-jet identification. All the variables are summarised in Table 8.1 for IPxD, in tables 8.4 and 8.5 for the two sets of JETFITTER variables and the SV1 variables are listed in Table 8.3. In addition to the baseline tagger information, also the kinematic variables p_T and $|\eta|$ are passed to the training to explore correlations between the kinematics and the baseline tagger variables. The kinematics are, however, treated differently since it is not intended to classify jets based on differences in the kinematic distributions between the fla-

vours (more details in sec 9.2). The *DL1r* configuration includes in addition the flavour probabilities provided by the RNNIP algorithm as outlined in Table 8.2. The last tagger version *DL1rmu* exploits also the soft-muon information from Table 8.7 besides all variables also used in *DL1r*.
The umbrella term *DL1* is used for the tagger family but typically also the baseline version is called *DL1*. To avoid ambiguities in the text, the baseline version will be denoted as *baseline DL1* while in plots and schematics the expression *DL1* will be kept to be consistent with the official ATLAS naming.
Adam, which is a gradient descent optimiser (see sec. 7.3), is utilised as optimiser together with a learning rate scheduler which reduces the learning rate after a certain amount of epochs if the loss did not change. All other hyperparameters differ slightly between the various trainings and are explained in the dedicated sections.

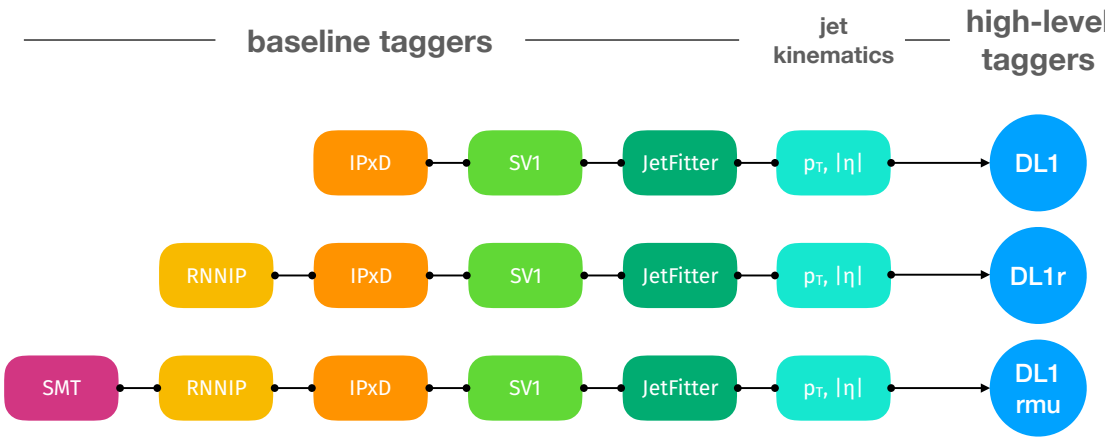


Figure 9.2.: Structure of the different types of DL1 taggers, depending on their variables used in the training.

9.1.1 *Software Chain*

The software chains for training & development and for the application in the ATLAS software ATHENA [186] are completely disentangled from each other. The training software is based on industry-standard open-source software using python3.6 [187]. For data handling the numpy [188] and pandas [189] packages are used together with the file format hdf5 [190]. In general, a heavy use of human readable file formats as JSON [191] and yaml is made to make the code structure well configurable for users. For the training itself TensorFlow [144] is employed with the keras2 [145] frontend. The visualisation package matplotlib [192] as well as the tools from scikitlearn [193] are included in the training process. The full workflow is based on Docker [194] images which allow to run the software on any computing resource without worrying about the package installation and their versions.
The C++ based LightWeight Tagger Neural Network (lwtnn) package [148] integrates the NN based taggers in in the ATLAS software ATHENA.

9.2 PREPROCESSING AND INPUT VARIABLE TREATMENT

To guarantee a robust training, it is necessary to perform several preprocessing steps. The first important step is the choice and preparation of the training sample. For the training of the DL1 taggers a mixture of two samples is taken, as described in Section 8.1.2, denoted as *hybrid sample* [185]. Figure 9.3 shows their jet p_T distributions for each jet flavour. The $t\bar{t}$ sample (solid lines) has a rapid fall in p_T compared to the Z' sample (dashed lines) which has a flat p_T spectrum up to roughly 4.5 TeV and a total range up to 6 TeV[1]. The use of the Z' sample allows a more robust training at high p_T. From these two samples a so-called *hybrid sample* is created. In the following only *PFlow*

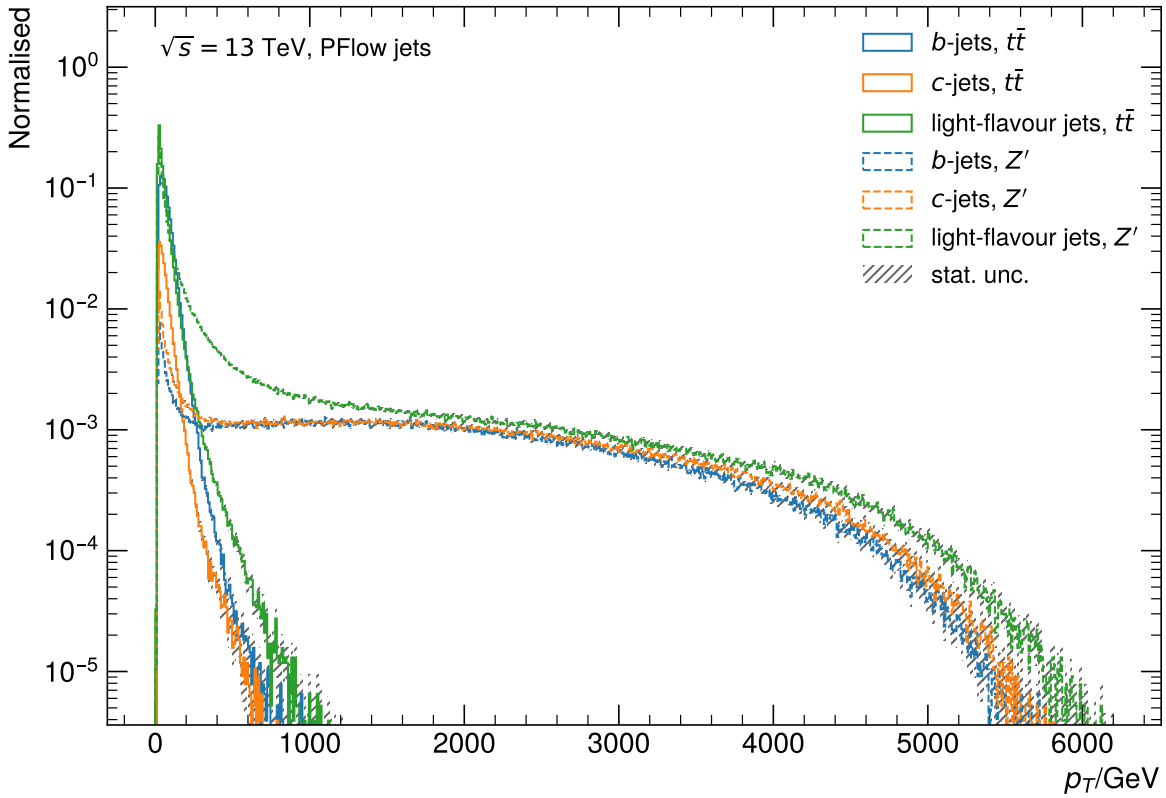


Figure 9.3.: p_T distribution of the $t\bar{t}$ sample (solid lines) and Z' sample (dashed lines). The $t\bar{t}$ b-jet distribution is normalised to unity and all other distributions are normalised to the $t\bar{t}$ b-jet distribution.

jets are considered, the differences for *VR Track* jets are discussed in the dedicated Chapter 11. Since the *extended* Z' sample is only available with the pile-up profile of the 2017 data taking period, only this MC production is used for both $t\bar{t}$ and Z'. The 2017 pile-up profile is a good representation of the overall RUN II pile-up profile as shown in Figure 3.2. The jets from the $t\bar{t}$ sample are chosen to populate the lower p_T-range up to a specific value ξ, ensuring sufficient statistics and the extended Z' sample is populating only the higher p_T-regions larger than ξ. This results in the following selection:

1 This Z' sample is also called *extended* Z', there is also another sample version which only has a p_T range up to roughly 3-4 TeV (which will be important for the *VR Track* jets training), also denoted as *standard* Z'.

$$\text{t}\bar{\text{t}}\text{ selection : } \begin{cases} \text{b-jets} & \text{b-hadron } p_T < 250\,\text{GeV} \\ \text{c-jets} & \text{jet } p_T < 250\,\text{GeV} \\ \text{light-flavour jets} & \text{jet } p_T < 250\,\text{GeV}, \end{cases}$$

and the Z' selections are chosen orthogonal. As indicated, the b-jet selection is using the b-hadron p_T. Together with the choice of the fraction of $t\bar{t}$ to be 70% in the hybrid sample, a smooth transition between the two samples is realised. The p_T distributions including the hybrid selections are shown in Figure 9.4 (a) separately for $t\bar{t}$ and Z'. In Figure 9.4 (b), the p_T distributions from both samples are merged. As intended, the b-jet distribution shows a smooth transition while the c-jet distribution has a kink and the light-flavour distribution a spike. In the previous iteration of the tagger, the

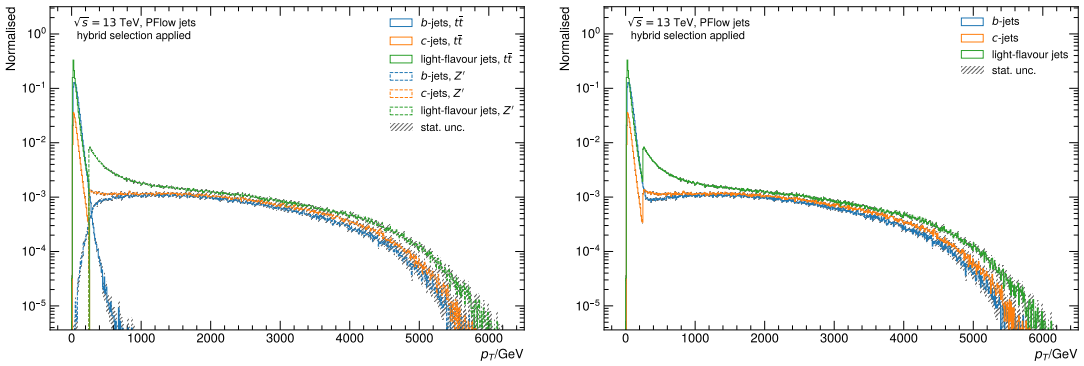


Figure 9.4.: p_T distribution of the $t\bar{t}$ and Z' sample with the hybrid selection applied with a $t\bar{t}$ fraction of 70%. In plot (a) the samples separated normalised to the $t\bar{t}$ b-jet distribution and (b) the samples merged into one distribution per flavour normalised to the b-jet distribution.

p_T and $|\eta|$ distributions[2] were weighted to match the b-jet p_T distribution. The weighting solves three issues: mitigate the discontinuity between samples, solving the problem of having imbalanced classes and reducing the influence to classify w.r.t the jet kinematics. Even though the correlations of the kinematics with the baseline tagger information is important, the goal is to avoid to train on differences in the kinematic distributions to distinguish between the flavours since the tagging is targeting an independent jet-by-jet classification. The classes are imbalanced, i.e. different fractions of b-, c- and light-flavour jets are present in the sample. Figure 9.5 shows the flavour fractions of the hybrid sample. The light-flavour jets are dominating, followed by b-jets and the c-jets constitute the smallest fraction. By reweighting in p_T and $|\eta|$, the weighted sum is $1/3$ for each flavour across the full p_T and $|\eta|$ spectra. However, it turned out that the weighting approach introduced some instabilities in the learning process of the network. Therefore, the resampling method is employed, as illustrated in Figure 9.6. Instead of weighting the distributions, single jets are either removed from the majority classes (*undersampling*) or jets from the minority classes are duplicated (*oversampling*) to match a given distribution. This method is also an indirect weighting but to the NN jets with event

2 Due to the symmetric ATLAS detector and the symmetry of physics processes in $\pm\eta$, the absolute value is used.

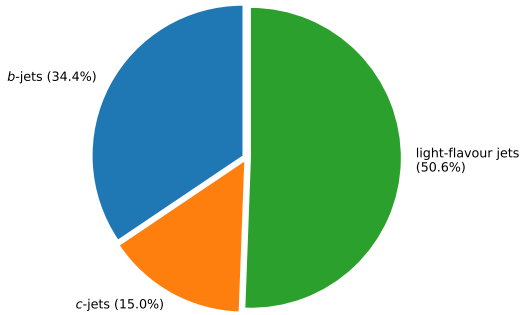


Figure 9.5.: Flavour composition of the hybrid sample.

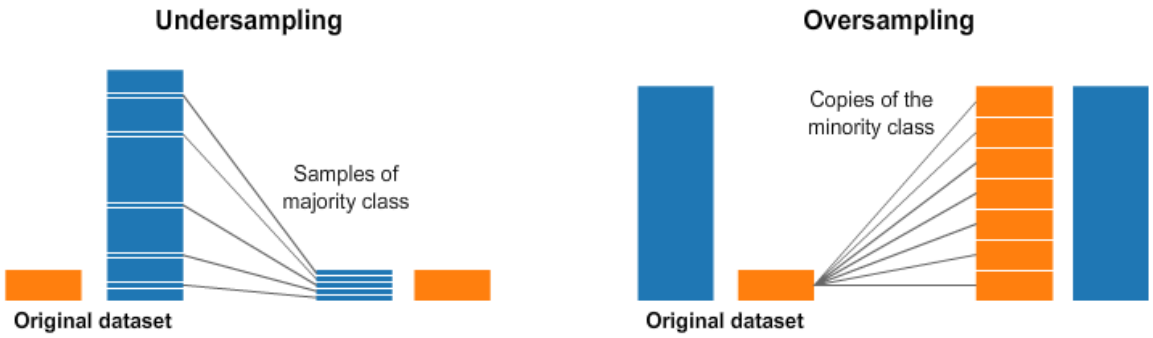


Figure 9.6.: Illustration of the resampling method to cope for imbalanced classes. The undersampling approach is shown on the left and the oversampling approach is shown on the right [195].

weight equals one are passed. Luckily, sufficient MC statistics is available, so it is possible to only make use of the *undersampling* method by removing jets from the majority classes. The b-jet p_T and $|\eta|$ distributions are taken as reference and the c- and light flavour jets are undersampled. From the full sample statistic, a subset from each flavour class is extracted such that in each p_T–$|\eta|$ bin the same amount of jets per flavour is available. The binning used for the undersampling procedure is optimised to be more granular in the lower p_T region, especially in the sample transition region, and to be wider for higher p_T. In total 434 p_T bins and 10 $|\eta|$ bins are utilised for the resampling. The resulting p_T spectrum of the training hybrid sample is depicted in Figure 9.7. The small differences originate from the usage of an equidistant binning in the plot. The hybrid $|\eta|$ distribution is shown before resampling in Figure 9.8 (a) and after resampling in Figure 9.8 (b).

The baseline taggers are not always succeeding, e.g. if there is no secondary vertex in the jet, the SV1 or JETFITTER algorithm cannot extract any information and a default value is returned. The fraction of default values for the IPxD, SV1 and JETFITTER are illustrated for the hybrid sample in Figure 9.9. While the IPxD algorithms have almost no default values, the SV1 and JETFITTER have significantly more cases where they cannot reconstruct a displaced vertex. Due to the design of JETFITTER also single track vertices can be reconstructed and it therefore finds more often displaced vertices than SV1. It is also nicely visible that in both cases the b-jets have the smallest fraction of default values since a secondary vertex is expected to be present in the jet. The next category is

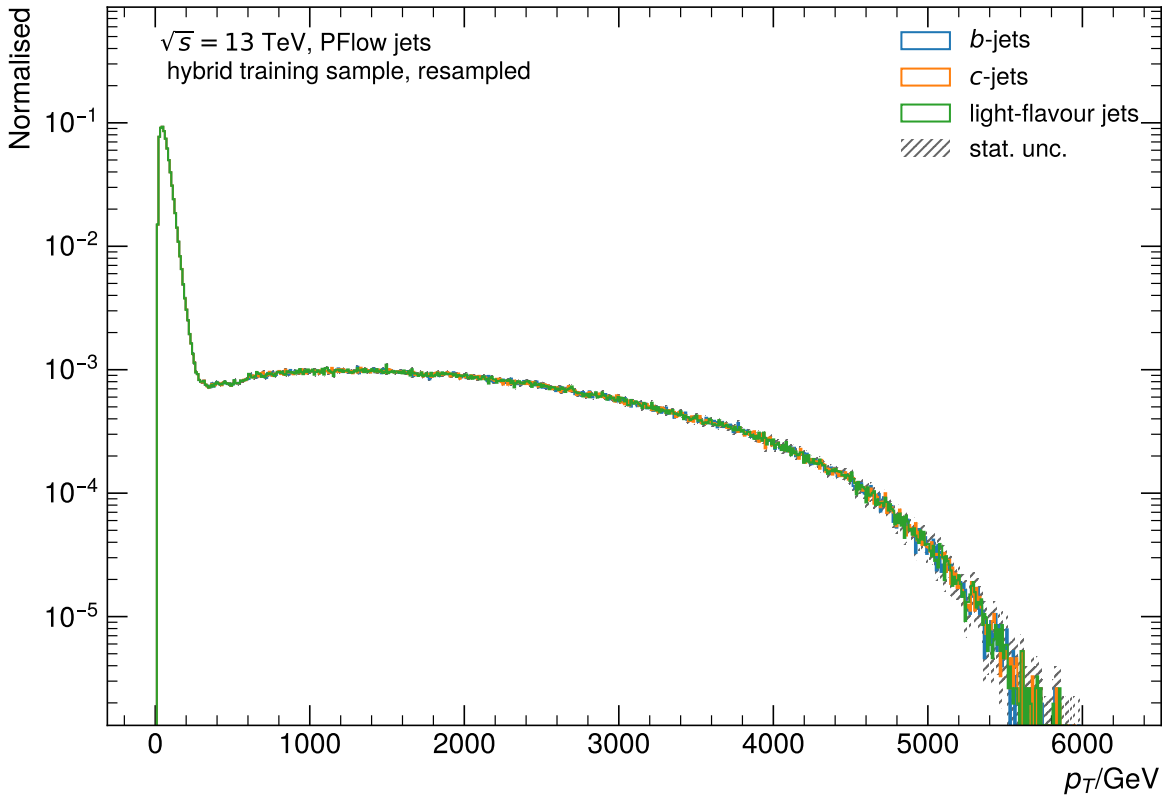
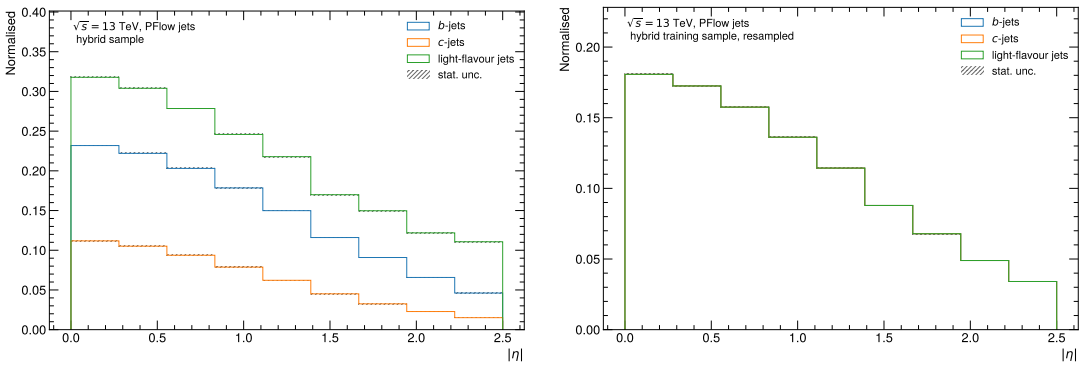


Figure 9.7.: p_T distribution of the undersampled hybrid sample.

c-jets, which usually also have a secondary vertex although it is more difficult to reconstruct them. Light-flavour jets have the most default values since displaced vertices rarely appear.

BDTs can easily deal with default values which are set to a value way off the spectrum which is normally done within ATLAS. However, the learning process of Neural Networks suffers from this kind of default value treatment. Therefore, the default values are replaced by the mean of the inclusive distribution. It is also possible to choose a different approach but the mean value is the simplest solution. In some cases, the default values are also set to the edge value of the distribution for physics reasons. The energy fractions (f_E^{SV1}, f_E^{JF}, $f_E^{JF_c}$) are set to zero if it is not possible to reconstruct a displaced vertex since in this case no energy is carried by a SV. In addition, the default values of the variables S_{xyz}^{JF}, $N_{vertices}^{JF}$, $N_{1\text{-trk vertices}}^{JF}$, N_{trks}^{JF}, $N_{\geqslant 2\text{-trk vertices}}^{JF}$, $N_{trks}^{JF_c}$, $N_{2TrkVtx}^{SV1}$ and $N_{TrkAtVtx}^{SV1}$ are set to the lower edge of the distributions. Two example distributions are shown with the default value set to the mean for m_{inv}^{JF} ($\langle m_{inv}^{JF} \rangle = 2.6\,\text{GeV}$) in Figure 9.10 (a) and to the lower edge value for f_E^{JF} ($f_E^{JF} = 0$) in Figure 9.10 (b). The default values are combined with a binary check variable for each baseline algorithm indicating if the algorithm returned default values. This allows the network to distinguish between the cases where the default value is a true default value or is actually coming from the underlying physics process. The binary indicator variables are listed in Table 9.1.

The last preprocessing step is to balance the ranges of the input variables such that they are all in the same order of magnitude to improve the learning process of the NN. Therefore, all variables are shifted to a mean of zero and a standard deviation of one, with the exception of the binary check variables. The two kinematic variables, the JETFITTER mass m_{inv}^{JF} and the SV1 energy fraction f_E^{JF} are



(a) (b)

Figure 9.8.: The $|\eta|$ distributions of the hybrid sample (a) $t\bar{t}$ and Z' merged without any further processing and (b) resampled to match the b-jet spectrum. The b-jet distributions are normalised to unity and the c-jet and light-flavour jet distributions to the normalised b-jet distribution.

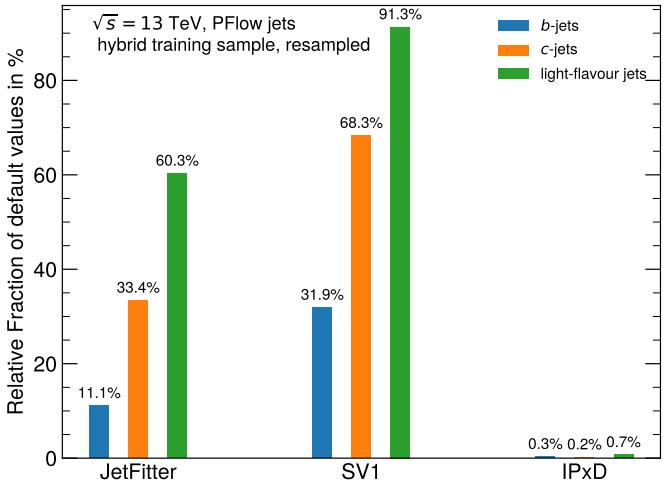
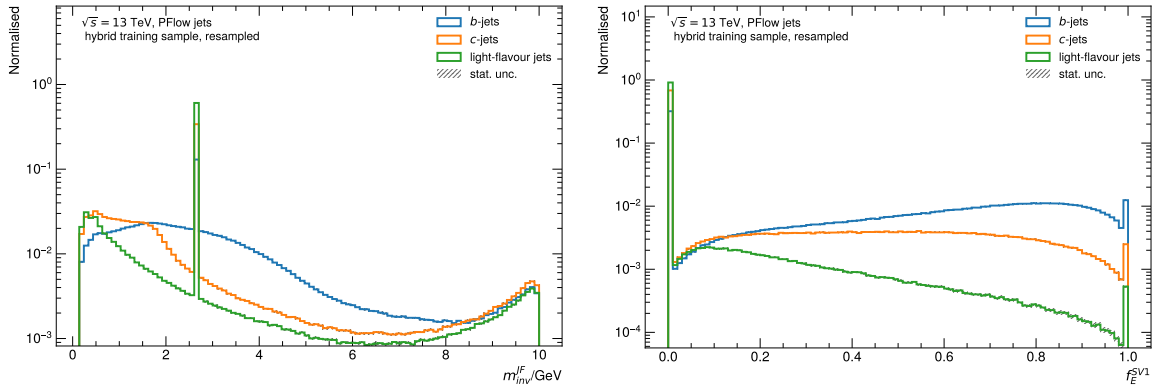


Figure 9.9.: Bar chart indicating the amount of default values per baseline tagger and flavour.

Variable	Description
	Binary check variables for the baseline tagger:
$SV1_{isDefault}$	SV1
$IP2D_{isDefault}$	IP2D
$IP3D_{isDefault}$	IP3D
$JF_{isDefault}$	JetFitter
$JF^{c}_{isDefault}$	c-identification optimised JetFitter
$SMT_{isDefault}$	Soft Muon Tagger

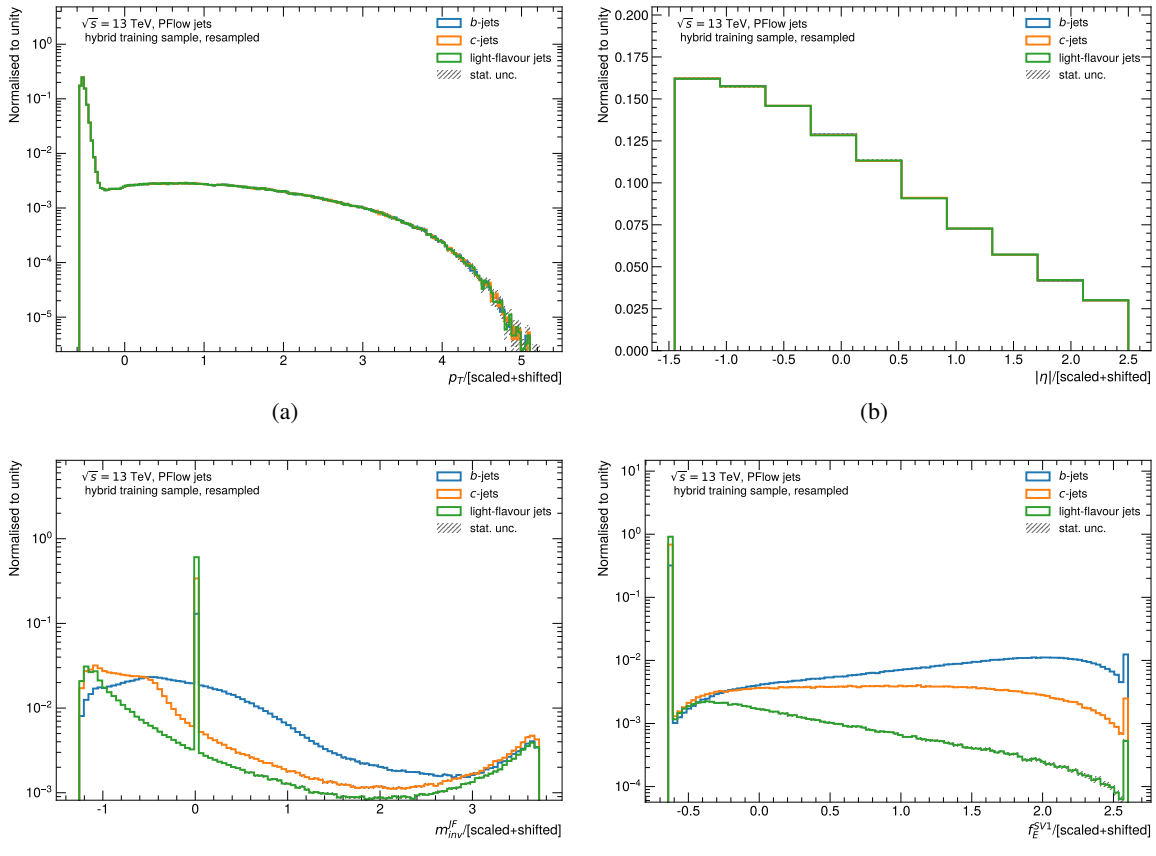
Table 9.1.: Binary check variables indicating if a baseline algorithm returned default values.

shown in Figure 9.11 with the full preprocessing applied. All remaining variables which are summarised in Figure 9.12 are also fully preprocessed. Furthermore, outliers coming from jets in extreme phase spaces are removed from certain distributions, which are far off the spectrum and would only disturb the training process.



(a) (b)

Figure 9.10.: The variable distribution of (a) the JETFITTER mass m_{inv}^{JF} with its default value set to its mean $\langle m_{inv}^{JF} \rangle = 2.6\,\text{GeV}$ and (b) the SV1 energy fraction f_E^{JF} with its default value 0. All distributions normalised to unity.



(a) (b)

(c) (d)

Figure 9.11.: Fully preprocessed distributions (resampled, scaled and shifted) of (a) the jet p_T (b) $|\eta|$ (c) m_{inv}^{JF} and (d) f_E^{JF}. All distributions normalised to unity.

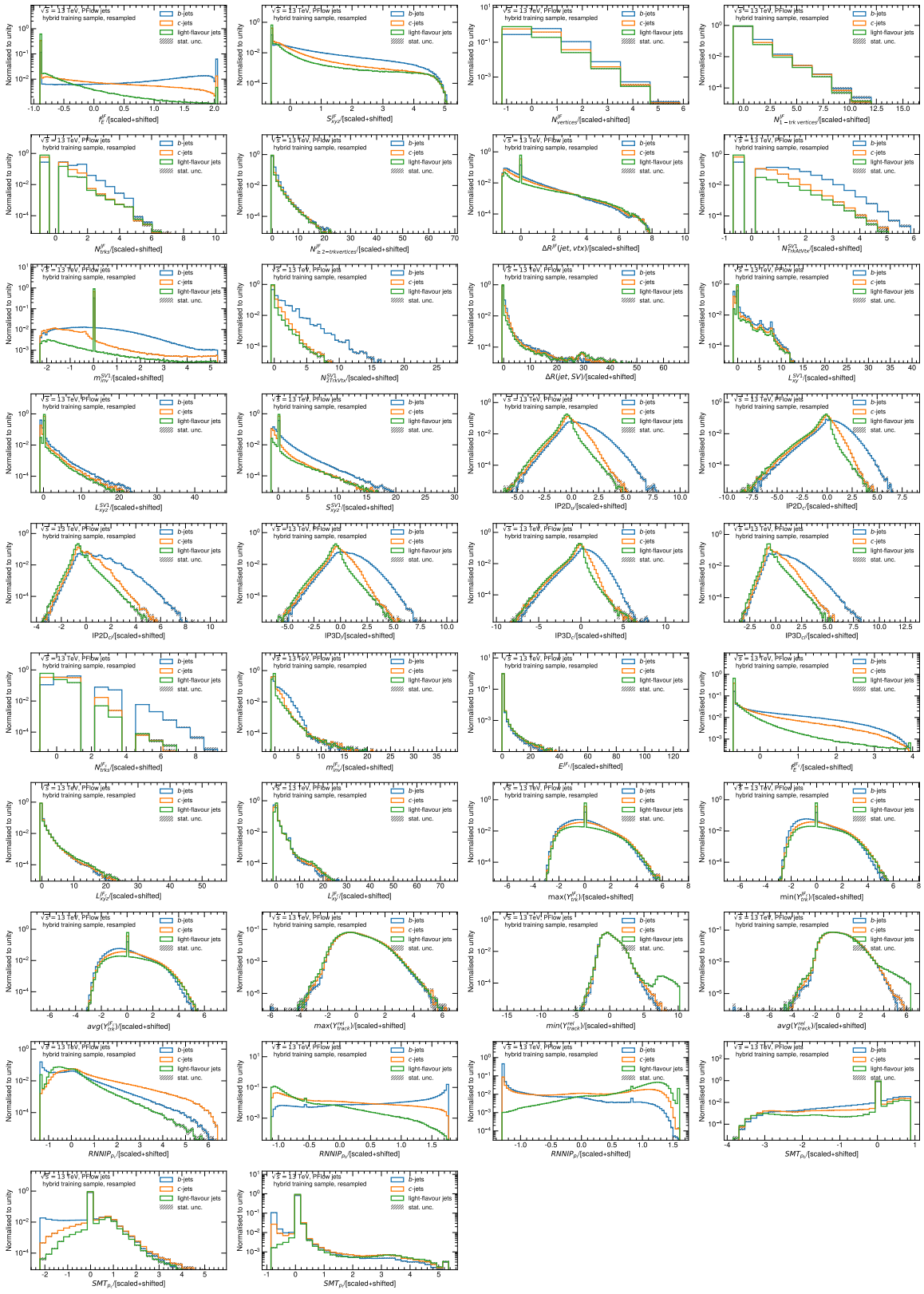


Figure 9.12.: All remaining variables used for the DL1 tagger family after fully preprocessing (resampled, scaled and shifted). All distributions normalised to unity.

10

PARTICLE FLOW JETS TRAINING OF THE DEEP LEARNING BASED HEAVY-FLAVOUR TAGGER

b-tagging is crucial for most physics analyses in ATLAS. Therefore it is important to provide a collaboration-wide tagger recommendation as well as its calibration. In this thesis, the b-tagging algorithm is studied and optimised to be provided for the whole collaboration. All studies in this chapter use particle flow jets, which is the default jet collection in ATLAS since 2020.
The previous b-tagging algorithm was optimised for *EMTopo* jets. To improve the b-tagging performance even further the algorithms are reoptimised (retrained) for *PFlow* jets. The main focus lies on the *DL1r* tagger, which includes the RNNIP information and should become the new default b-tagger in ATLAS.
First, some general training optimisation studies are shown in Section 10.1, followed by a dedicated hyperparameter optimisation on ATLAS GRID GPUs in Section 10.2, continued with a comparison of the different tagger variants and the final performance overview in Section 10.3.

10.1 TRAINING OPTIMISATION

This section describes several smaller studies dealing with different improvements and physics implications for the training of the tagger.

10.1.1 *Smoothing Final Discriminant*

The discriminant $\mathcal{D}_\mathrm{b}$ as defined in Equation (9.1) is used to determine the working points (WPs) which define if a jet is considered to be a b-jet. In Figure 10.1 the discriminant of the 2018 *baseline DL1* version [185] (previous tagger iteration) is shown. The WPs are indicated as vertical dashed lines. Especially the 60% WP is critical since it is exactly at an edge of a peak in the b-jet distribution. This can lead to a jump in the efficiency if a slight change occurs. Ideally, the distribution should be smoother to avoid such edge cases. Several studies were performed adapting the NN architecture to trace down the cause of this effect. It turned out that the usage of *Dropout* caused this peaky distribution. This most probably happens since the information about the default value is crucial for a good training and if this information does not get properly propagated through the network (which can happen by dropping certain node connections) it can influence the training process. Typically, *Dropout* makes the training more robust but in this specific case, it causes an effect which should be avoided. Therefore, in the following model optimisation *Dropout* is not used.
However, *Dropout* was not the only reason for this effect though the dominant one. Switching to the

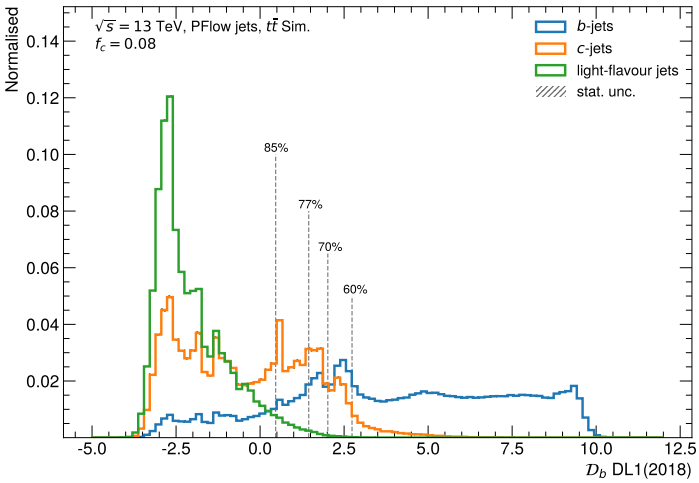


Figure 10.1.: *Baseline DL1* discriminant from the previous tagger iteration (2018), the vertical dashed lines are indicating the official b-tagging working points.

newer keras version (from keras1 to keras2) and using the resampling approach was also resulting in a smoother output distribution. Figure 10.2 shows the final discriminant for a training with the new setup and a comparable sample size to what was used in the previous iteration. The hyperparameters for this setup are summarised in Table 10.1. The distribution is much smoother, especially for b-jets and thus a more robust working point definition is possible.

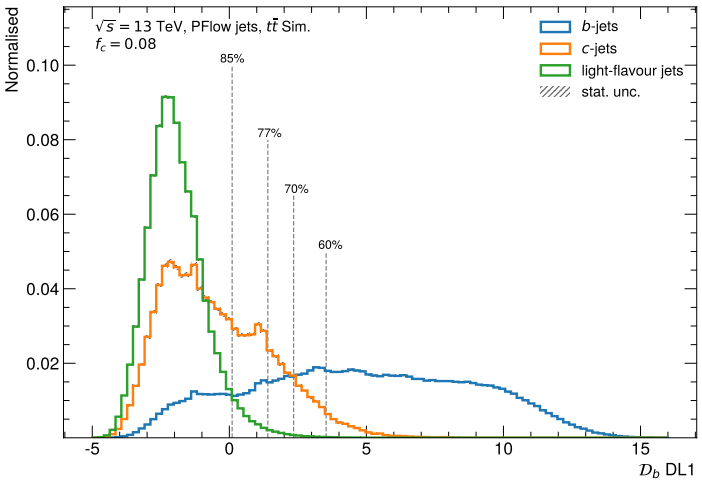


Figure 10.2.: *Baseline DL1* discriminant for a revised training after removing Dropout, updating to Keras2 and making use of resampling. The dashed lines show different b-jet efficiencies corresponding to the b-tagging WPs.

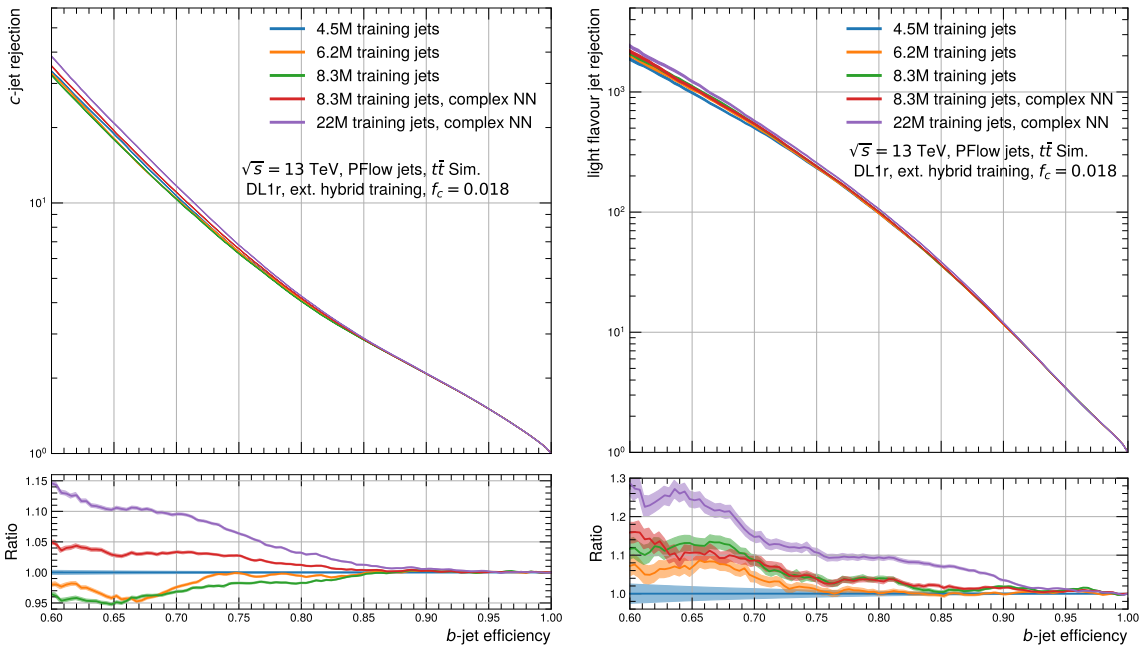
10.1.2 *Training Sample Statistic Implications*

In the previous iteration of the DL1 tagger [185], a training set statistic of 5.1M jets was used and the reweighting approach was applied. In this section, the implications of the training sample statistic are being tested. The preprocessing is the same for all samples as described in Section 9.2 and is based on the resampling approach.

Hyperparameter	Value
$N_{\text{hidden layers}}$	8
$N_{\text{nodes/layer}}$	[72, 57, 60, 48, 36, 24, 12, 6]
learning rate	0.01
Training batch size	4000
Training sample size	5.1M jets

Table 10.1.: Hyperparameters used for the revised *baseline DL1* training using Keras2 and resampling.

The training sample size is altered between 4.5M and 22M jets to study its influence on the learning process. In a first step, three different training statistics are tested (4.5M, 6.2M and 8.3M) while keeping the network architecture identical. The resulting performances are shown in Figure 10.3 as the blue (4.5M), orange (6.2M) and green (8.3M) lines. For both the c-jet and light-flavour jet rejection no big differences are visible between the trainings with 4.5M or 6.2M input jets. There is a small improvement in the light-flavour jet rejection by using 8.3M jets in the training. In general, all plots of this type have binomial uncertainties indicated as their width. They are larger for the light-flavour jet rejection since more jets are rejected yielding a lower statistic, especially for lower b-jet efficiencies.



(a) (b)

Figure 10.3.: Performance comparison of different *DL1r* tagger setups for different training sample statistics, (a) shows the c-jet rejection and (b) the light-flavour jet rejection as a function of the b-jet efficiency. The coloured band indicates the binomial uncertainty.

In the next step, the NN architecture is adapted to be more complex, leaving the NN more freedom to adjust the weights with larger training statistics, in this case 8.3M jets. Clearly, the performance

improves for both c-jet and light-flavour jet rejection. Finally, the maximally available sample size of 22M jets is employed which is limited by the amount of available c-jets in the $t\bar{t}$ sample due to the undersampling approach. The performance improves by almost 20% in the light-flavour jet rejection and in the c-jet rejection at a b-jet efficiency of 60% compared to the training with 4.5M training jets shown in Figure 10.3.

Hyperparameter	**Baseline NN**	**Complex NN (8.3M)**	**NN (22M)**
$N_{\text{hidden layers}}$	8	8	8
$N_{\text{nodes/layer}}$: 1^{st}, 2^{nd}	[72, 57]	[128, 72]	[256, 128]
$N_{\text{nodes/layer}}$: $3^{\text{rd}} - 8^{\text{th}}$		[60, 48, 36, 24, 12, 6]	
Learning rate	0.01	0.01	0.01
Activation function	ReLu	ReLu	ReLu
Training batch size	3000	5000	15000

Table 10.2.: Overview of hyperparameters used for the training statistic studies.

The hyperparameters of the different networks are listed in Table 10.2. The complexity of the networks are enhanced by increasing the number of nodes in the first and second hidden layers. In addition, the training batch size is raised for larger statistics. Besides speeding up the training, since fewer iterations over the training sample are necessary, also the performance increased by adapting the batch size.

To perform the NN training with 22M jets it is necessary to change the training procedure. Usually, the training sample is loaded into memory and the computation unit can directly access the data. These 22M jets, however, do not fit into the memory anymore and a different procedure needs to be established. The training sample needs to be read from the disk in batches during the training. In order to speed this up, the process is parallelised with so-called *pipelining*. The CPU is reading the batches from the disk and preparing it. The training itself is then performed on the GPU in parallel as illustrated in Figure 10.4. This can be realised with tools from keras and TensorFlow.

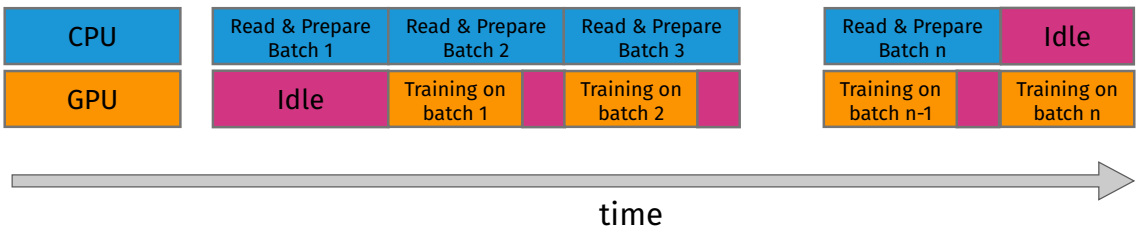


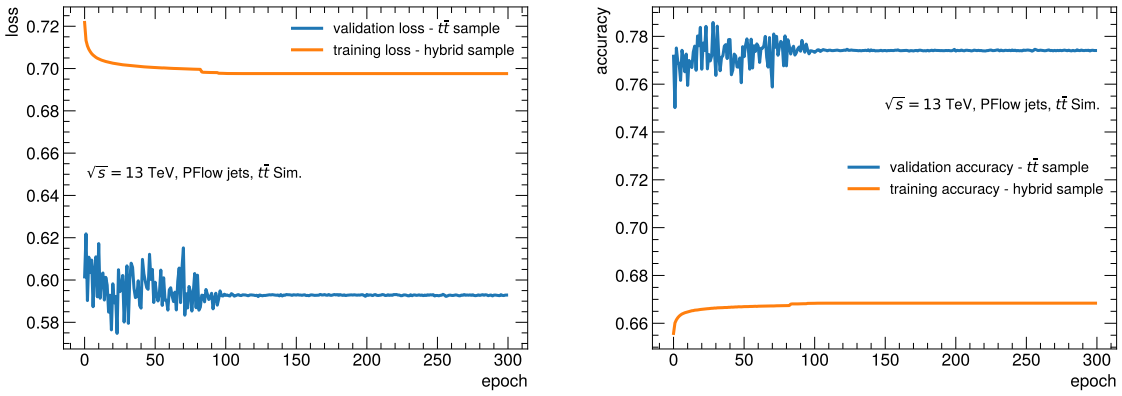
Figure 10.4.: Schematic of the parallel training with CPUs and GPUs. The CPU reads the batches from the disk and prepares them to feed them to the GPU, which is performing the training.

However, not only the training procedure but also the preprocessing needs to be adapted. The shuffling of the training dataset which is normally done by the keras framework, needs to be done

now by hand before the training. The shuffling is done to ensure an adequate representation of the training sample in each training batch.

10.1.3 *Performance Monitoring*

To monitor the training process and directly check for possible overfitting as well as getting a first performance measure, several quantities are evaluated and stored during the training. The monitoring is designed such that it is possible to check the evolution already during the training.
Besides the loss and accuracy shown in Figure 10.5, also the c-jet and light-flavour jet rejection for the 77% WP are evaluated, as illustrated in Figure 10.6. The presented training is performed with the



(a) (b)

Figure 10.5.: Training and validation (a) loss and (b) accuracy as a function of the training epoch. *DL1r* training with 22M training jets as indicated in Table 10.2.

input variables for *DL1r* and 22M training jets using the hyperparameters listed in Table 10.2. All distributions evaluated on the $t\bar{t}$ validation set show fluctuations up to epoch 100 and then converge. The Z' sample is only used for the high p_T extension in the training while the physics sample of interest is $t\bar{t}$. Therefore $t\bar{t}$ only is used as validation sample to monitor the performance. The worse performance of the hybrid training sample in terms of the loss in Figure 10.5 (a) and the accuracy in Figure 10.5 (b) is due to the different sample composition. Generally, the $t\bar{t}$ sample shows a better performance.

To further check the stability of the training, several epochs are evaluated in more detail in Figure 10.7. Also there, the same trend is visible: while the first epochs still show fluctuations with respect to each other, the results for the epochs 120 and 150 are almost identical. For instance, the improved c-jet rejection in epoch 3 is coming with a trade-off in its light-flavour jet rejection. In addition, the two epochs 120 and 150 are performing best overall for the light-flavour jet rejection and also very well in the c-jet rejection.

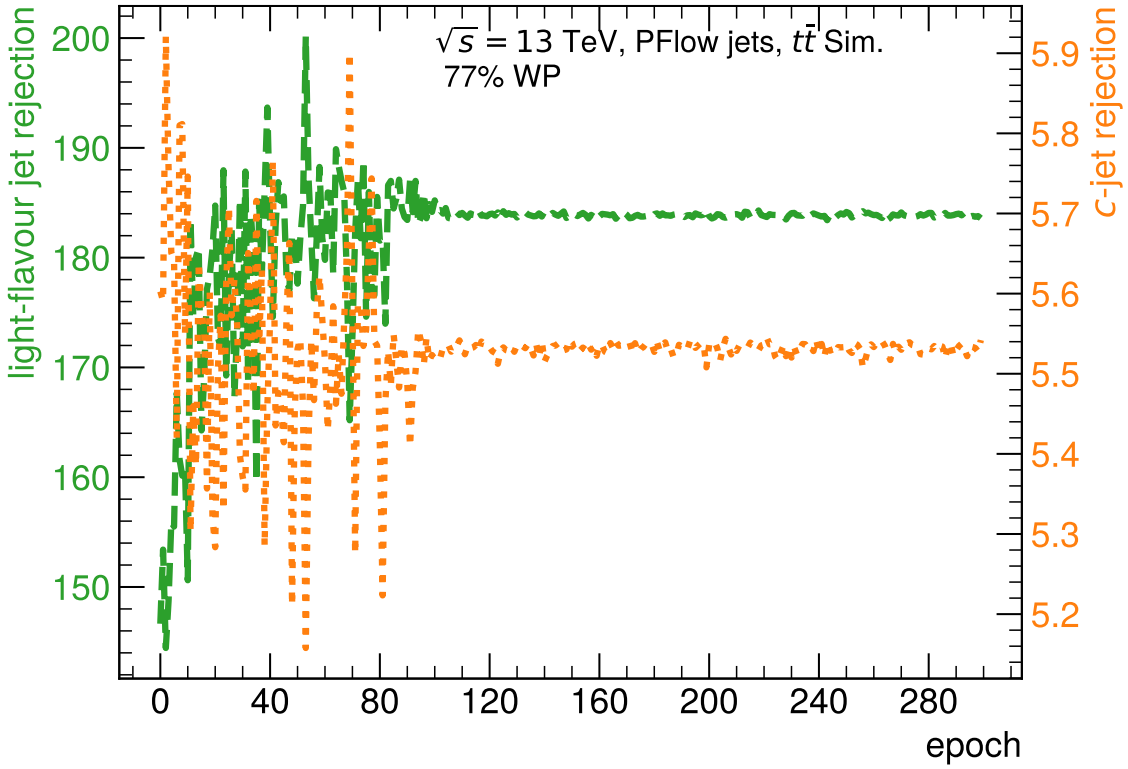
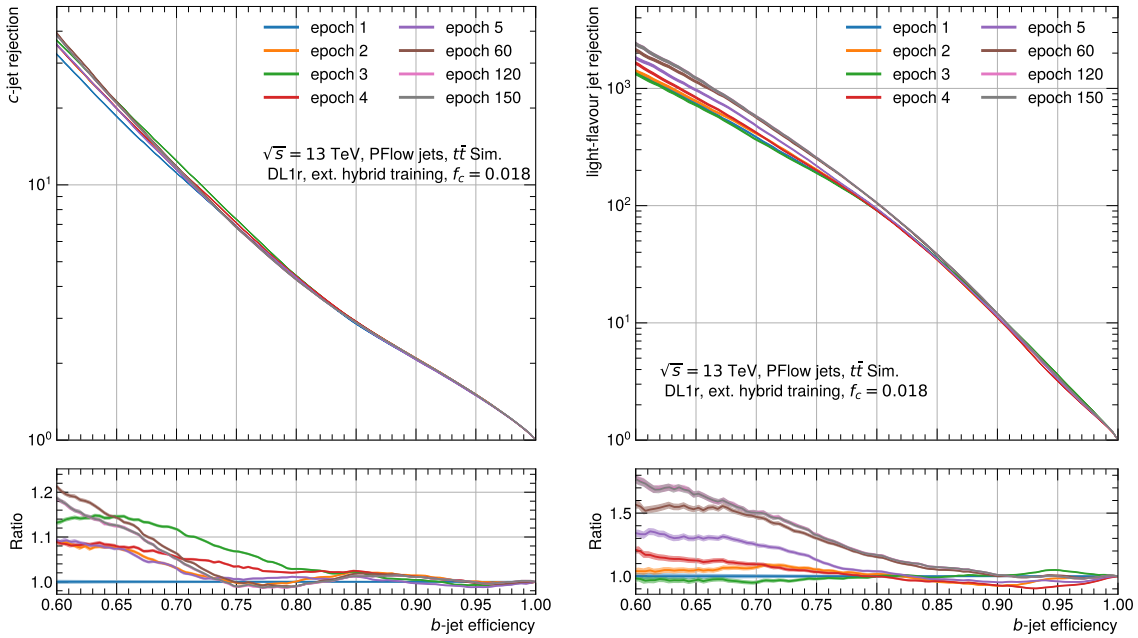


Figure 10.6.: Light-flavour jet rejection (green) and c-jet rejection (orange) as a function of the training epoch evaluated on the validation $t\bar{t}$ sample during the training. *DL1r* training with 22M training jets as indicated in Table 10.2.



(a) (b)

Figure 10.7.: Epoch comparison of *DL1r* training for (a) the c-jet rejection and (b) light-flavour jet rejection.

Due to the convergence of the background rejections, overfitting can be excluded. This can be also shown by evaluating the c-jet rejection for a subset of $t\bar{t}$ of the training sample and the validation sample shown in Figure 10.8. There are differences visible but there is no clear trend of the validation set being less performant than the training sample.

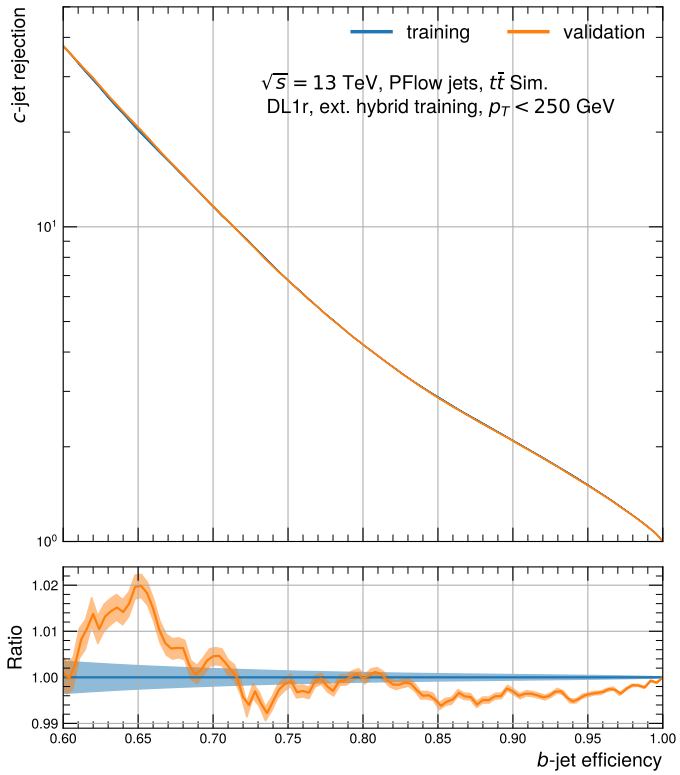
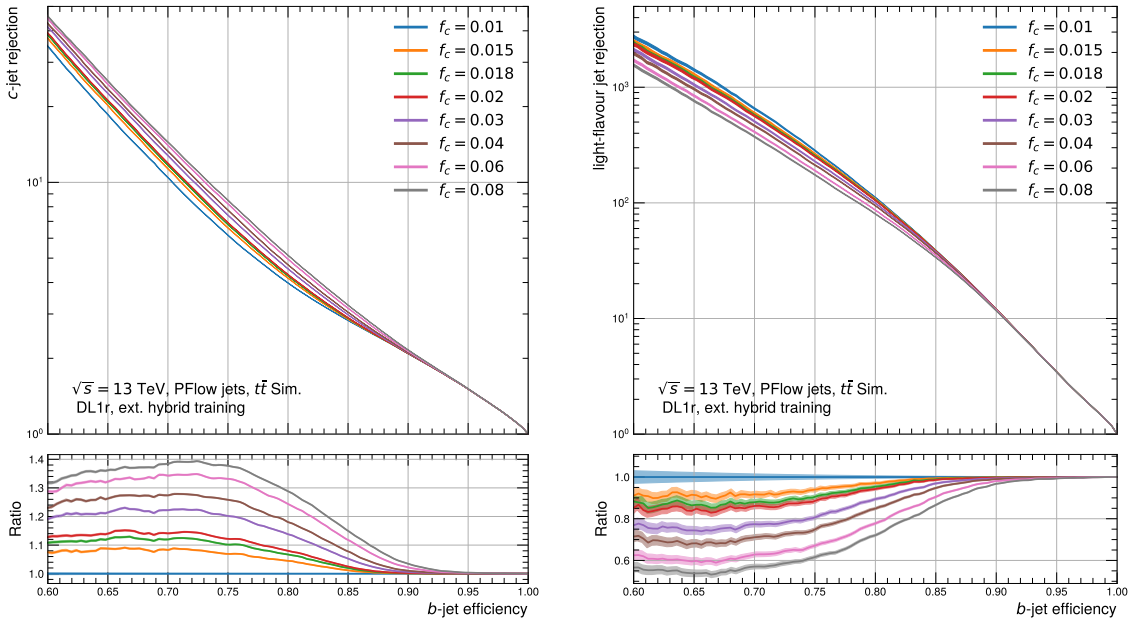


Figure 10.8.: c-jet rejection as a function of the b-jet efficiency evaluated on the training and validation $t\bar{t}$ sample for $p_T < 250$ GeV. *DL1r* training with 22M training jets as indicated in Table 10.2.

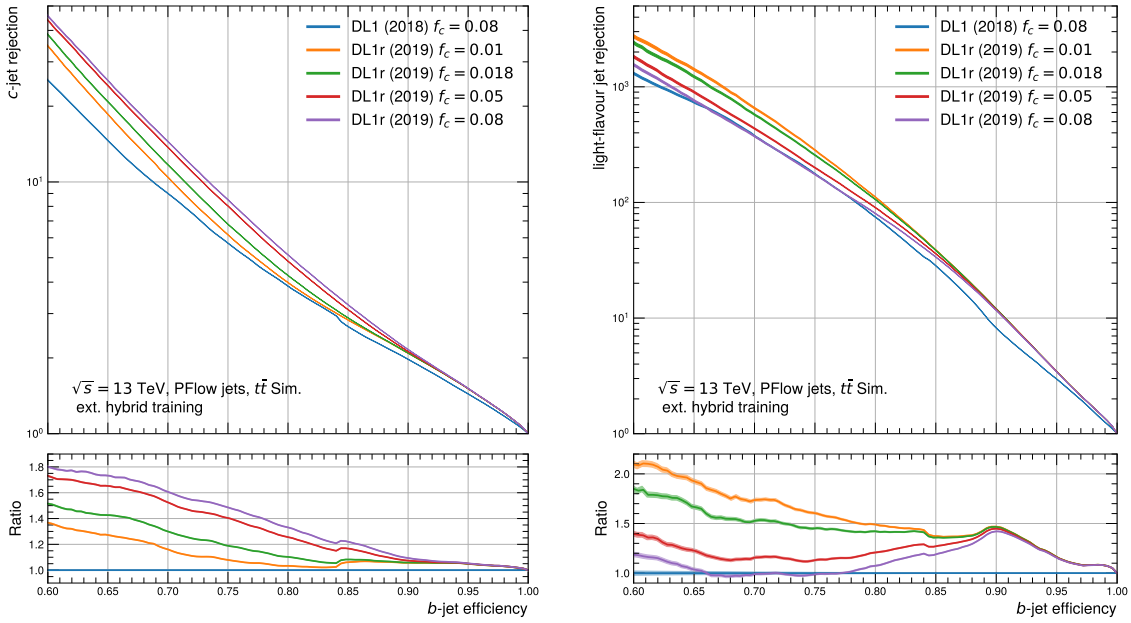
10.1.4 *Optimisation of* c*-jet Fraction*

The free parameter f_c in the final discriminant indicated in Equation (9.1) needs to be optimised for the individual trainings. A scan through different values of f_c is shown in Figure 10.9 for the *DL1r* training with 22M training jets as shown in Table 10.2. It nicely demonstrates the performance shift from the c-jet rejection towards the light-flavour jet rejection by lowering the value of f_c and vice versa. Finding the optimal value for f_c requires the input from physics analyses stating which quantity (c-jet or light-flavour jet rejection) is more important. In the end, a compromise has to be made covering the requirements of most analyses. A comparison of different f_c performances with the previous recommendation of *baseline DL1* is illustrated in Figure 10.10. The recommendations for *EMTopo* jets (DL1 (2018)) used a f_c value of 0.08 which roughly corresponds to the actual fraction of c-jets in $t\bar{t}$. From these plots it turns out that $f_c = 0.018$ is better suited for the new *DL1r* tagger, making a compromise between a better c-jet and light-flavour jet rejection. The fact that the f_c value needs to be smaller is due to the different composition of the training sample. Before, the c-jets



(a) (b)

Figure 10.9.: Scanning of several f_c values for the *DL1r* tagger with 22M training jets. (a) shows the c-jet rejection and (b) the light-flavour jet rejection.



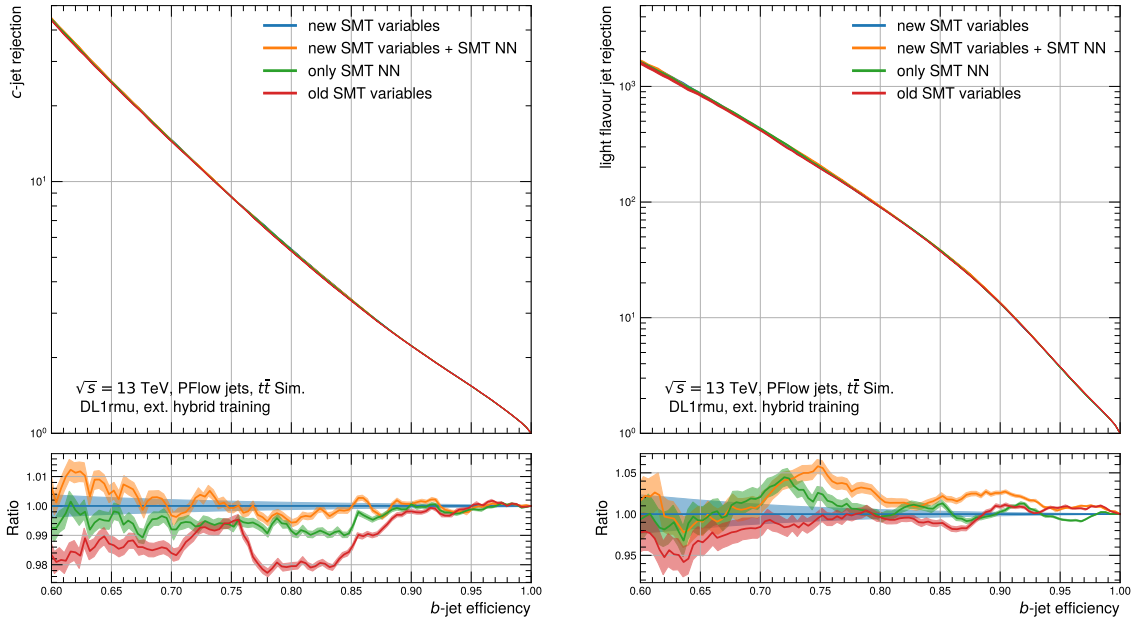
(a) (b)

Figure 10.10.: Comparison of different f_c values of the *DL1r* training with the *baseline DL1* (DL1 (2018)) training on *EMTopo* jets. (a) shows the c-jet rejection and (b) the light-flavour jet rejection.

were only weighted to match the kinematic distributions of the b-jets, now with the undersampling approach, the same amount of c-jets and b-jets (and light-flavour jets) are included in the training. Thus the training sees more different c-jet topologies and can learn them better. To retrieve a similar relative performance increase for c-jets and light-flavour jets, the f_c value needs to be smaller.

10.1.5 *Choice of SMT variables*

The soft muon information as described in Section 8.4 is included in the *DL1rmu* tagger. There are three different variable combinations available to describe the soft muon information as described in Section 8.4: the variable set with the old d_0 sign calculation, the updated d_0 sign calculation (see Table 8.6) and the soft muon NN combining this information (see Table 8.7). To choose the best variable set as input for the *DL1rmu* training, all these variable sets are tested and their results are compared in Figure 10.11. As expected the soft muon variables with the old d_0 sign definitions perform the worst indicated as the red line in the plot. The combination of the new variable set combined with the soft muon NN delivers the best results. Using the new soft muon variables alone gives similar and only marginally worse results than the combination with the NN. The usage of the soft muon NN variables gives a similar performance in terms of the light-flavour jet rejection and slightly worse performance in the c-jet rejection compared to using the new soft muon variables directly. The three latter variable combination sets with the updated d_0 sign calculation differ maximally by roughly 2%.
In the end, the soft muon NN output was chosen as input set to the *DL1rmu* tagger, to not duplicate the transferred information to the network by including the high-level information of the soft muon NN and its input variables. In addition, it is also important to have a standalone tool as the soft muon NN and therefore this set of input variables for the *DL1rmu* is chosen.
In general, it would be also a possibility to add an intermediate (additional) output with a separate loss for the soft muon information to the DL1 tagger which would also provide the soft muon information separately and would also pass more information to the final network in a similar manner to the UMAMI network described later in Section 12.1.



(a) (b)

Figure 10.11.: Performance comparison for different sets of soft-muon variables as input for the *DL1rmu* tagger shown for (a) the c-jet rejection and (b) light-flavour jet rejection.

10.2 HYPERPARAMETER OPTIMISATION ON GRID GPUS

The training optimisation from Section 10.1 gave a significant gain in the tagger performance even though the hyperparameters were so far only optimised by hand and no systematic scan was performed. There are several possibilities to do a hyperparameter scan. The classic approach is the grid-search in which a set of trainings with pre-defined hyperparameter combinations is compared to each other. This allows to cover a large phase space, however, it is also costly in computations. Often a certain phase space of the hyperparameters can be excluded by only evaluating a subset of them.
Besides the hyperparameter optimisation itself, a new technical infrastructure is introduced and studied in cooperation with ATLAS IT [196]. To test this new infrastructure, a heavy computational workload is required, and the grid-search is a good use case. The detailed technical setup of the hyperparameter optimisation is described in Chapter A.
Making use of this workflow, a hyperparameter optimisation is performed for a training of the *DL1r* tagger. A subset of around 300 hyperparameter combinations is tested across five dimensions (3 layers, batch size and learning rate) as listed in Table 10.3 providing in total 450 combinations. The hyperparameters are chosen such that they vary the most important hyperparameters starting from the network architecture of the complex NN for 22M training jets in Table 10.2.

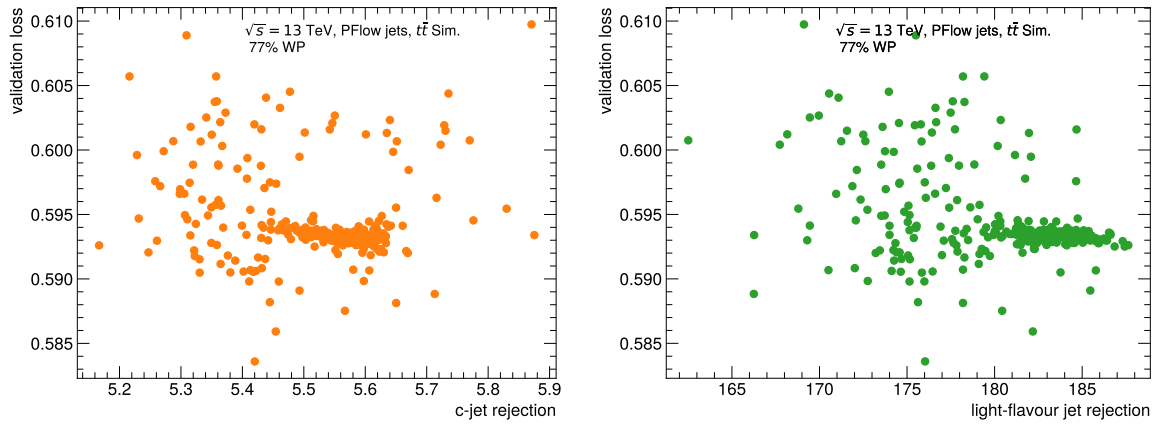
Each training is performed for 130 epochs to reach the convergence as e.g. shown in Figure 10.6 to avoid fluctuations. Figure 10.12 shows the scatter plots of the validation loss and the c-jet and light-flavour jet rejection. Clearly, the validations loss seems not to be a good measure for choosing

Hyperparameter	Values
$N_{\text{hidden layers}}$	8
$N_{\text{nodes/layer}}$: 1^{st}	128, 256, 512
$N_{\text{nodes/layer}}$: 2^{nd}	96, 128, 256
$N_{\text{nodes/layer}}$: 3^{rd}	60, 128
$N_{\text{nodes/layer}}$: $4^{\text{th}} - 8^{\text{th}}$	fixed to [48, 36, 24, 12, 6]
learning rate	5 values between 10^{-4} and 0.01
Training batch size	5 values between 20000 and 40000
Activation function	ReLu
Training sample size	22M jets

Table 10.3.: Hyperparameter combinations yielding in total 450 combinations.

the best set of hyperparameters. The scatter plots show for both the c-jet rejection and the light-flavour jet rejection a big spread and not a common trend. Even worse, several points are clustering horizontally for similar values of the validation loss but spanning a wider range in the background rejections. This means that even if the loss stays the same, the physics results change.

Therefore, another measure to quantify the performance of a hyperparameter set needs to be chosen.



(a) (b)

Figure 10.12.: Results of the hyperparameter optimisation showing a scatter plot between the validation loss and (a) the c-jet rejection and (b) the light-flavour jet rejection at the 77% WP.

The hyperparameter combinations are individually ranked by their c-jet and light-flavour jet rejection at the 77% WP. Typically, in the cases where the c-jet rejection is better the light-flavour jet rejection is worse and vice versa. Afterwards the two rankings are iteratively scanned from the top to the bottom and the first hyperparameter combination which is found in both rankings is chosen as the best combination. This is not necessarily the best ranked hyperparameter set from the individual ranks.

All hyperparameter combinations are visualised in the parallel coordinates plot in Figure 10.13 where

the red line indicates the best hyperparameter combination.

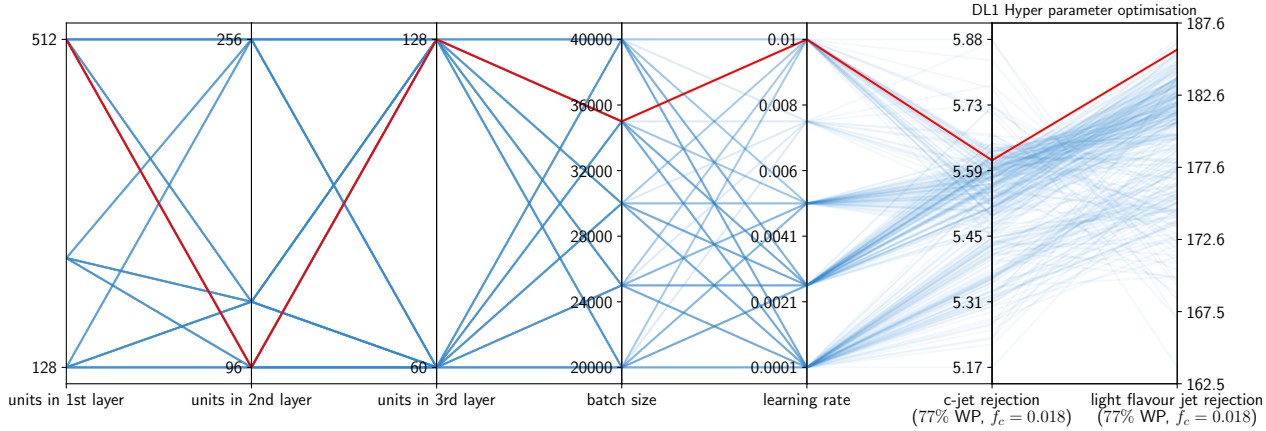
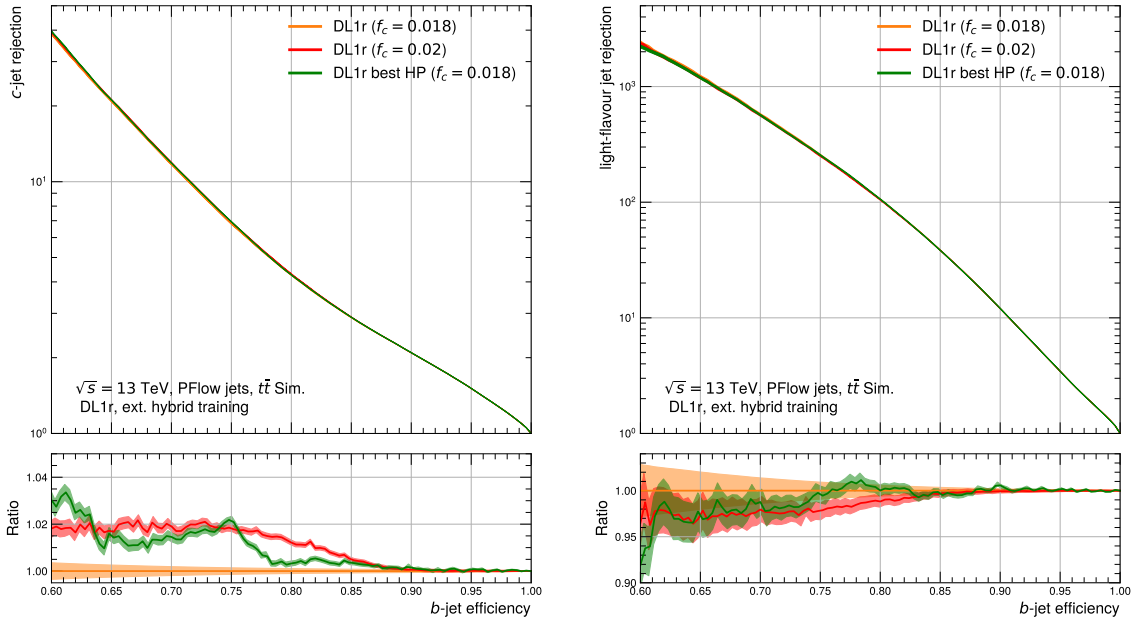


Figure 10.13.: Parallel coordinate plot showing the different hyperparameter combinations for the *DL1r* hyperparameter scan. The red line indicates the best combination.

The best hyperparameter combination compared to the by-hand optimised network is shown in Figure 10.14. The differences are only very small comparing the orange (by-hand optimised NN) and green curve (best combination). While the c-jet rejection improves slightly, the light-flavour jet rejection is a bit worse for the optimised hyperparameters. This difference can be corrected by adapting the f_c value shown in the red curve (by-hand optimised NN with $f_c = 0.02$) delivering the same performance as the optimised hyperparameters. It was also checked that the batch size has no big influence on the performance, as example shown in Figure 10.15, since only values between 20000 and 40000 are scanned and for the by hand-optimised network a batch size of 15000 is used. Since the batch size has no influence and the performance differences are marginal, the by-hand optimised hyperparameters are kept with the certitude that the network is well optimised and the optimised f_c value can be kept.

Hyperparameter	**Values**
$N_{\text{hidden layers}}$	8
$N_{\text{nodes/layer}}$	[256, 128, 60, 48, 36, 24, 12, 6]
learning rate	0.01
Training batch size	15000
Activation function	ReLu
Free (trainable) parameters	59,275
Fixed parameters	1,140
Training sample size	22M jets

Table 10.4.: Final network architecture for the *DL1r* tagger.



(a) (b)

Figure 10.14.: The (a) c-jet rejection and (b) light-flavour jet rejection as a function of the b-jet efficiency shown for the best hyperparameter combination in green and the by-hand optimised network for 22M training jets for two values of f_c in orange and red.

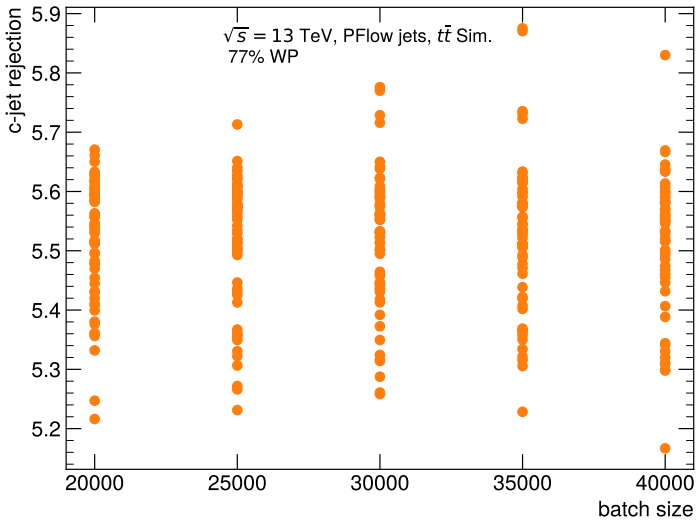
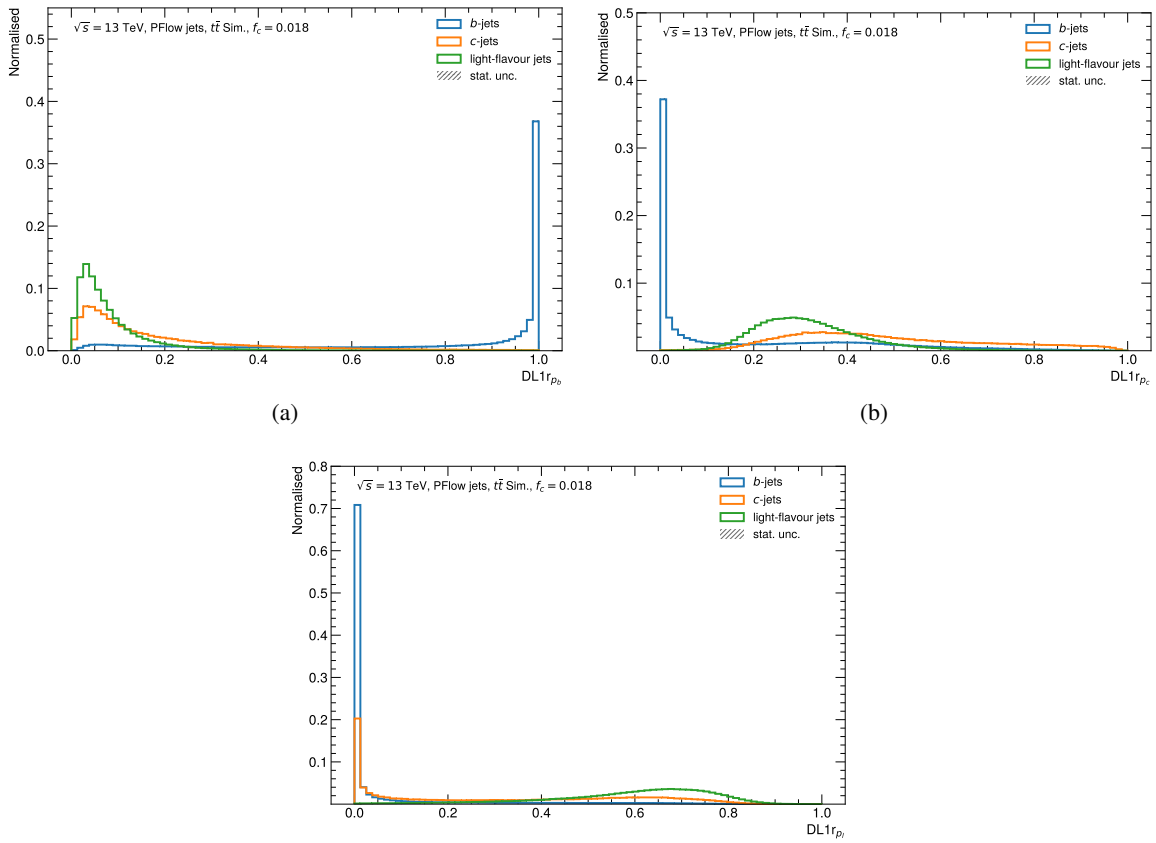


Figure 10.15.: Scatter plot of the results of the different hyperparameter combinations as a function of the c-jet rejection and the batch size.

10.3 PERFORMANCE OVERVIEW AND TAGGER VARIANTS COMPARISON

As shown in the sketch 9.2, three different versions of the DL1 tagger family are investigated. The *DL1r* version was discussed in detail above. The other two variants, the *baseline DL1* and *DL1rmu* are also retrained. The retraining of the *baseline DL1* is done to have a backup solution in case the *DL1r* tagger would have not been calibratable. This luckily turned out to not be the case [197]. The *DL1rmu* variant is also retrained but not calibrated since the light-flavour jet rejection is too high and thus the uncertainties would be too large to ensure a proper calibration. Different studies were performed to adapt the NN architecture for the *baseline DL1* and *DL1rmu*. The *DL1r* architecture described in Table 10.4 was found to perform also well for these other two variants. The only difference is the batch size which is set to 20000 for the *baseline DL1* and *DL1rmu* to speed up the training a bit with no impact on the performance.

The *DL1r* tagger outputs are shown in Figure 10.16 for all three output nodes providing the probabilities for every jet to be a b-, c- or light-flavour jet. In general, b-jets have the largest separation from the other two flavour categories (Fig. 10.16 (a)). While the c-jets still show a small separation (Fig. 10.16 (b)), the light-flavour jets are not well separable (Fig. 10.16 (c)). These outputs are then



(a) (b)

(c)

Figure 10.16.: The network output distributions of the *DL1r* tagger for (a) the b-jet node (b) the c-jet node and (c) the light-flavour jet node.

combined into the final b-tagging discriminant of the *DL1r* tagger which is shown in Figure 10.17. It is a smooth curve with the b-jets taking larger values with respect to the c-jets and light-flavour jets. The standard ATLAS WPs are indicated as dashed lines indicating the cut on the $\mathcal{D}_b$ discriminant. The distributions of the *baseline DL1* and *DL1rmu* tagger discriminants are shown in Figure 10.18.

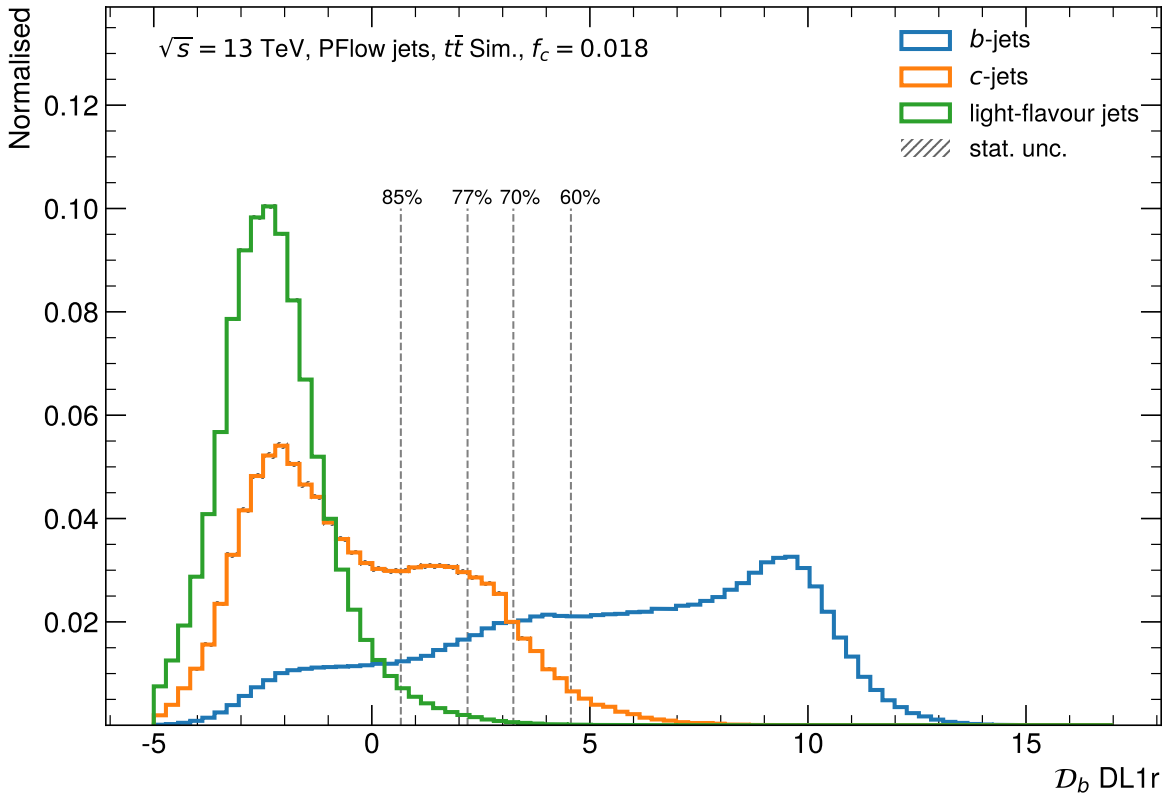
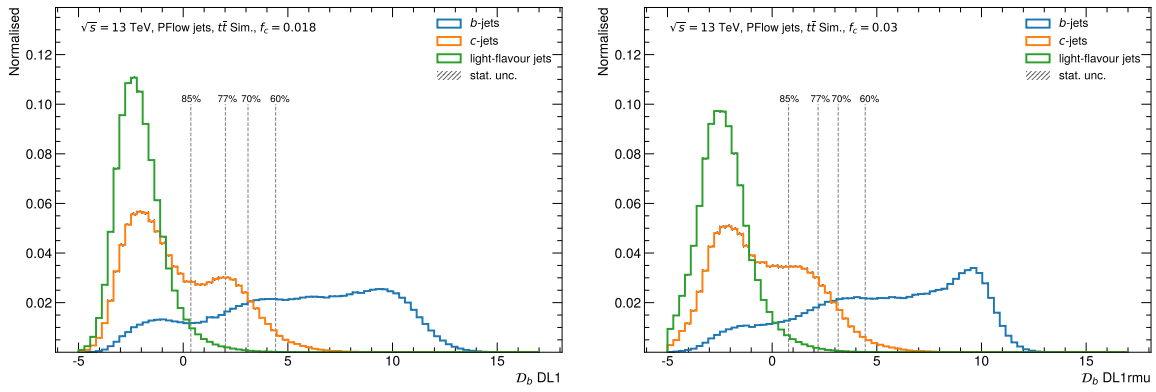


Figure 10.17.: *DL1r* b-tagging discriminant separately shown for b-jets (blue), c-jets (orange) and light-flavour jets (green). The vertical dashed lines are indicating the official b-tagging working points.

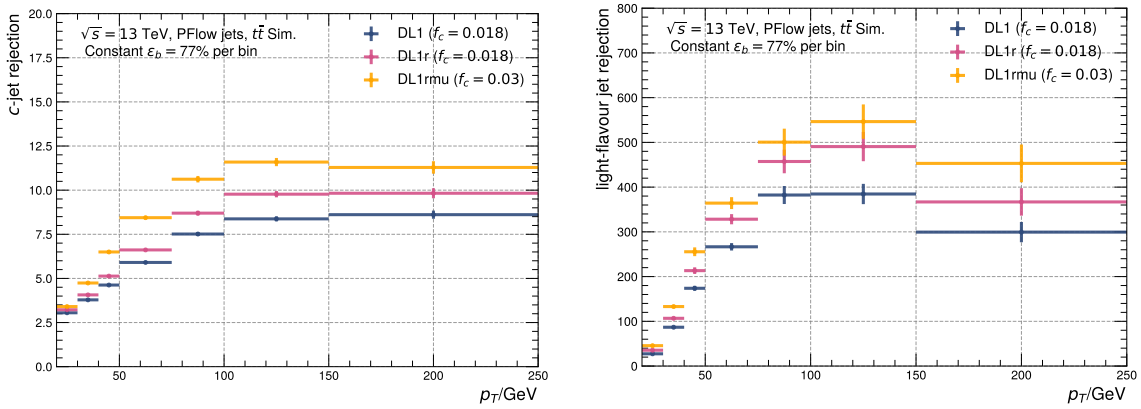
Also, their curves are very smooth and have no sharp edges and thus the WP definition does not suffer from the effect described in Section 10.1.1. The b-jet distributions are peaking more strongly for *DL1r* than for *baseline DL1* for high $\mathcal{D}_b$ and even more for *DL1rmu* which is expected to happen since the networks have more information available to classify b-jets. The f_c values were separately optimised and for the *baseline DL1* the same value as for *DL1r* $f_c = 0.018$ is used, while for *DL1rmu* a value of $f_c = 0.03$ was found to give better performance.

The p_T dependent performance is representatively evaluated for the 77% WP. Figure 10.19 and 10.20 show the p_T dependent background rejection for a $t\bar{t}$ and an *extended* Z' sample, respectively. In this case, the b-jet efficiency of $\epsilon_b = 77\%$ is kept constant per p_T bin. Overall, the expected trend that *DL1r* outperforms the *baseline DL1* and *DL1rmu* outperforms *DL1r* is visible in all four plots in every bin. The p_T spectrum up to 250 GeV is shown for the $t\bar{t}$ sample and the *extended* Z' covers the higher p_T spectrum from 250 GeV upwards. The improvements in the $t\bar{t}$ sample are almost constant over the full p_T range (Fig. 10.19) between the three tagger variants. Therefore, the additional RNNIP and soft muon information contribute to the performance gain in this lower p_T



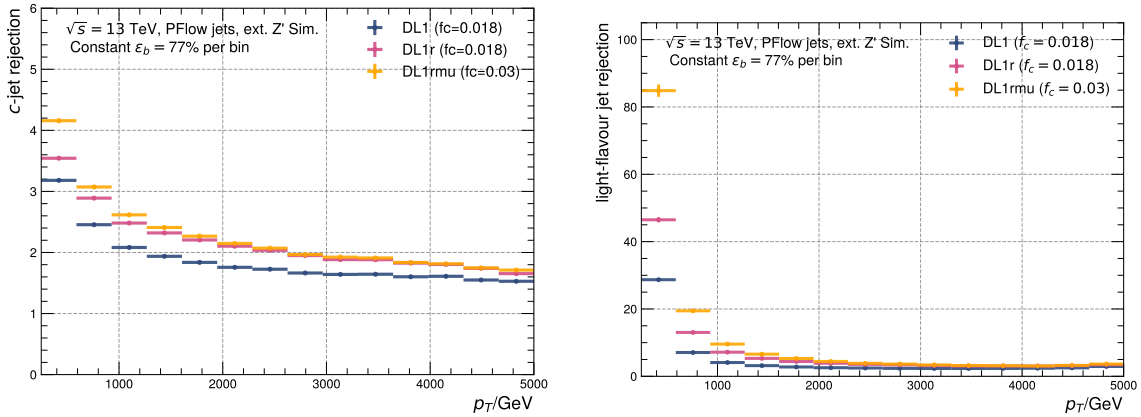
(a) (b)

Figure 10.18.: The (a) *baseline DL1* and (b) *DL1rmu* discriminants, the vertical dashed lines are indicating the different working points.



(a) (b)

Figure 10.19.: p_T dependent performance for a constant b-jet efficiency of 77% per bin for the (a) c-jet rejection and (b) light-flavour jet rejection for the $t\bar{t}$ sample.



(a) (b)

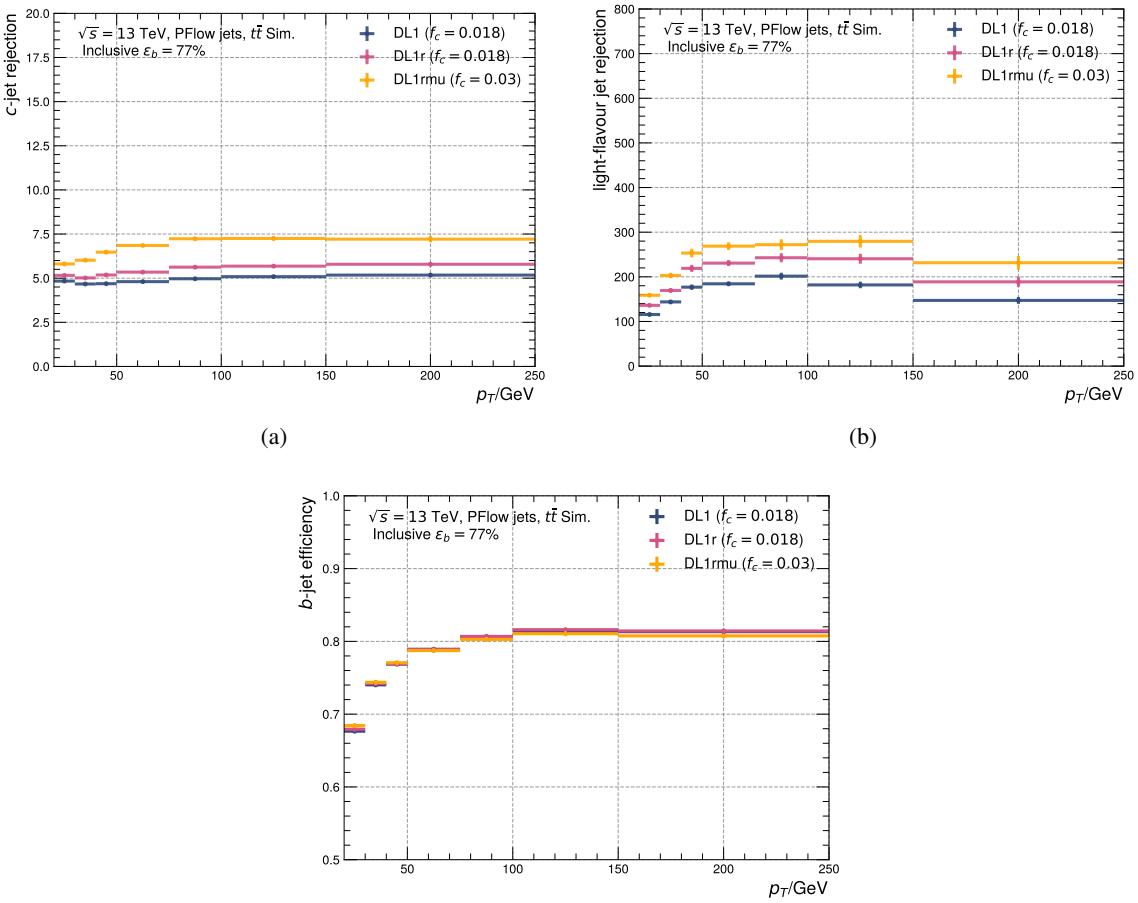
Figure 10.20.: p_T dependent performance for a constant b-jet efficiency of 77% per bin for the (a) c-jet rejection and (b) light-flavour jet rejection for the *extended* Z' sample.

regime. Looking at the higher p_T regime (Fig. 10.20), the first two bins show a similar behaviour while going to higher p_T values the soft muon information is not helping anymore. Hence, the RNNIP information improves the high p_T b-tagging while the soft muon information is not. This comes from the fact, that the RNNIP tagger was optimised on the high p_T *extended* Z' sample while it is getting more difficult to extract and associate the muon information to the b-hadron decay for high p_T.

The Figures 10.21 and 10.22 show the inclusive 77% WP, which is defined on an inclusive $t\bar{t}$ sample including all data taking periods[1], and is evaluated on the $t\bar{t}$ and *extended* Z' sample, respectively. The b-jet efficiency plot for $t\bar{t}$ in Figure 10.21 (c) reveals that the b-jet efficiency is not flat as a function of p_T but increasing with p_T. All three taggers show the same trend in terms of the increasing b-jet efficiency. Consequently, the c-jet rejection (Fig. 10.21 (a)) and the light-flavour jet rejection (Fig. 10.21 (b)) are much flatter than for the above case where the b-jet efficiency was maintained for every single bin. Similar to the constant b-jet efficiency the three taggers have the same performance behaviour with respect to each other (*baseline DL1* < *DL1r* < *DL1rmu*).
In contrast, the performance behaviour on the *extended* Z' sample changes more drastically with the inclusive b-jet efficiency. As illustrated in Figure 10.22 (c), the b-jet efficiency decreases drastically for higher p_T values. The RNNIP information stabilises this decrease a bit as well as the soft muon information while the *baseline DL1* shows a larger decrease down to about 5% at 5 TeV compared to slightly more than 10% for *DL1r* and *DL1rmu*. This behaviour is also reflected in the background rejection. For both, the c-jet rejection (Fig. 10.22 (a)) and the light-flavour jet rejection (Fig. 10.22 (b)) show an inverted performance compared to what is usually expected. Already in the second p_T bin for the c-jet rejection the *performance order* is inverted which means that the *baseline DL1* has a higher rejection than *DL1r* and *DL1r* higher than *DL1rmu*. For the light-flavour rejection, this effect is starting in the 4^{th} p_T bin. However, appearances are deceiving since the lower b-jet efficiencies in the higher p_T bins cause this inversion. Therefore, the background rejections are higher but also less b-jets are passing this selection.

The comparison of the full b-jet efficiency spectrum is illustrated in Figure 10.23 for the $t\bar{t}$ sample. Including the RNNIP information in the tagger training improves the c-jet rejection constantly by about 15% over a wide b-jet efficiency range (60% to ~ 80%). Similarly, the additional soft muon information gives another ~ 20% between 60% and ~ 75% b-jet efficiency. For higher b-jet efficiencies the performance differences are decreasing for the c-jet rejection but are still around 5% to ~ 15% at a b-jet efficiency of 85%, which is the loosest WP used in ATLAS.
The light-flavour jet rejection (Fig. 10.23 (b)) improves with the RNNIP information by almost 25% over the full b-jet efficiency spectrum relevant for the WP definition, it even increases more for a b-jet efficiency of ~ 84% compared to the *baseline DL1*. Another 10% performance gain is achieved adding the soft muon information, again applicable over the full b-jet efficiency range relevant for the WPs.

1 As a reminder, these performance plots are only evaluated on the 2017 data taking period pile-up profile.



(a) (b)

(c)

Figure 10.21.: p_T dependent performance for an inclusive b-jet efficiency of 77% defined on an inclusive $t\bar{t}$ sample (official ATLAS cut value) for the (a) c-jet rejection, (b) light-flavour jet rejection and (c) the b-jet efficiency as a function of p_T for the $t\bar{t}$ sample.

Even though the *DL1rmu* tagger performs best, it cannot be used in physics analyses since it is not possible to calibrate the light-flavour mis-tag rate. As pointed out in Section 8.6, if the light-flavour jet rejection is too high, the uncertainties on the scale factors are getting really large due to drastically reduced statistics. While this can be overcome for *DL1r* by using a *flipped tagger* (see sec. 8.6), this is not possible for the SMT information.

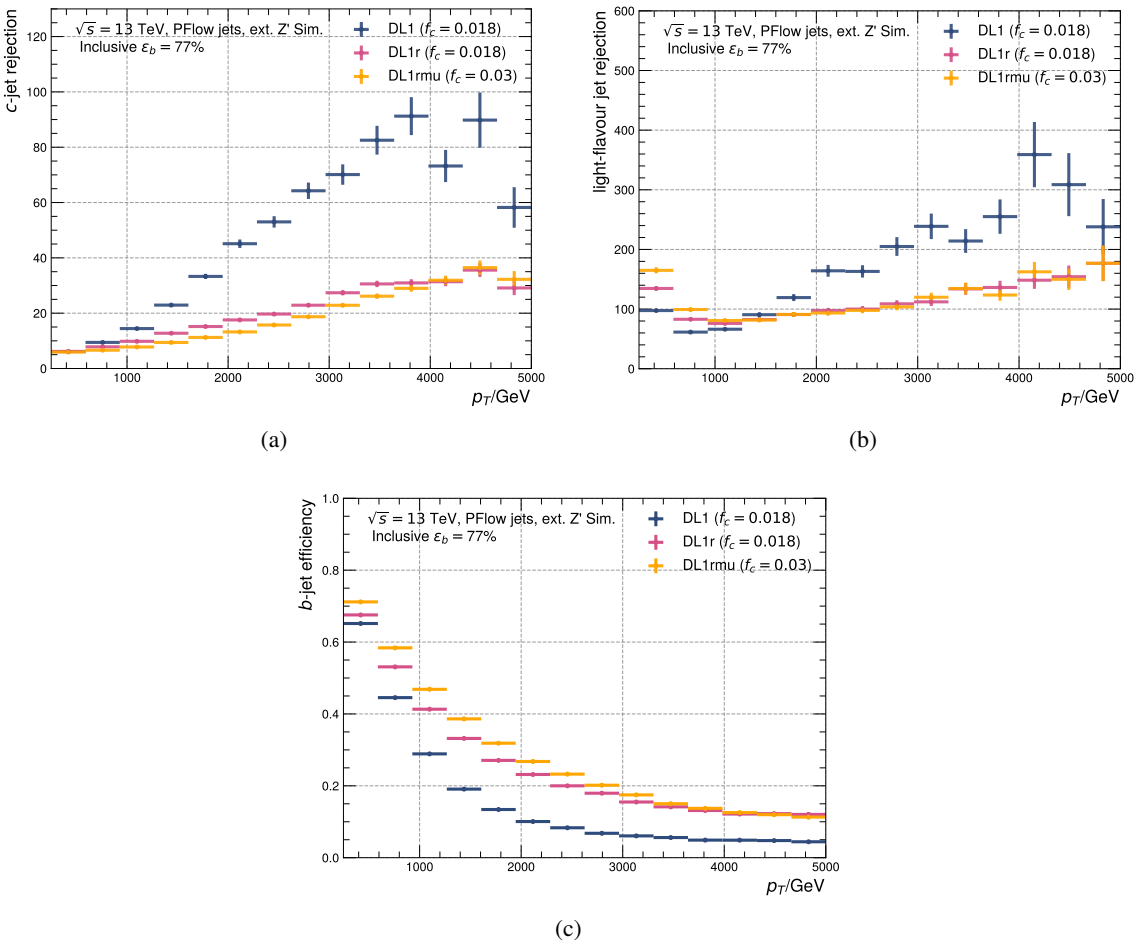
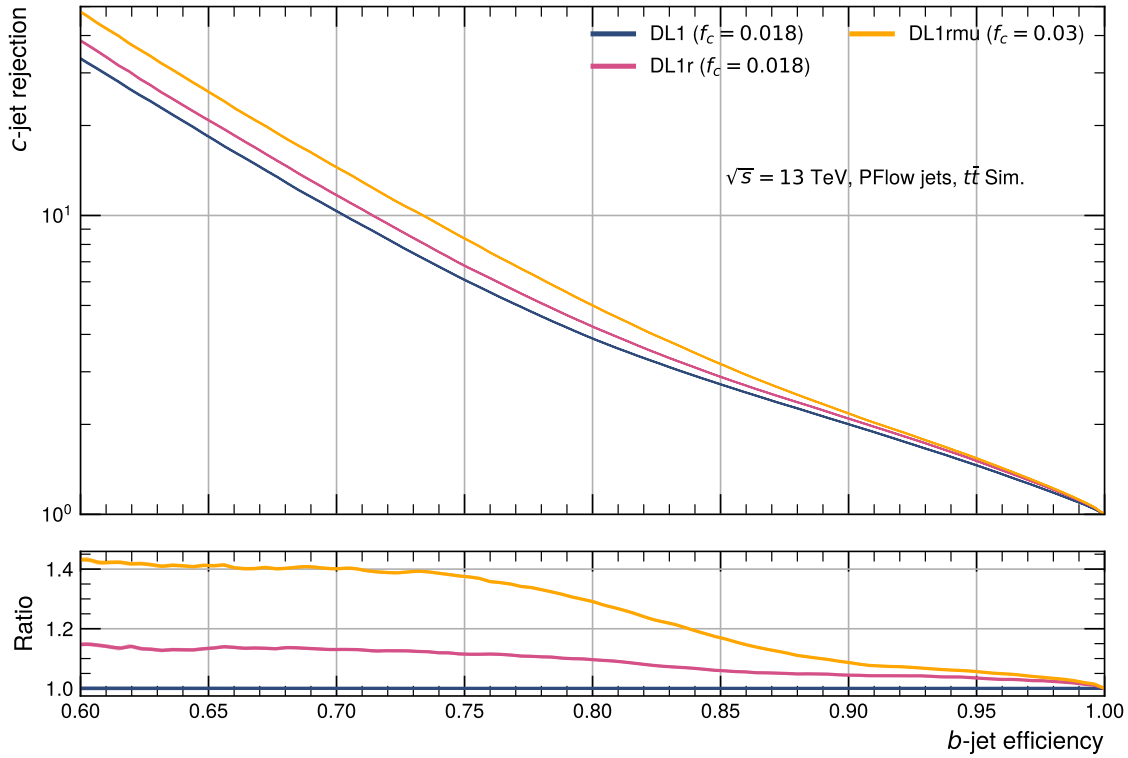
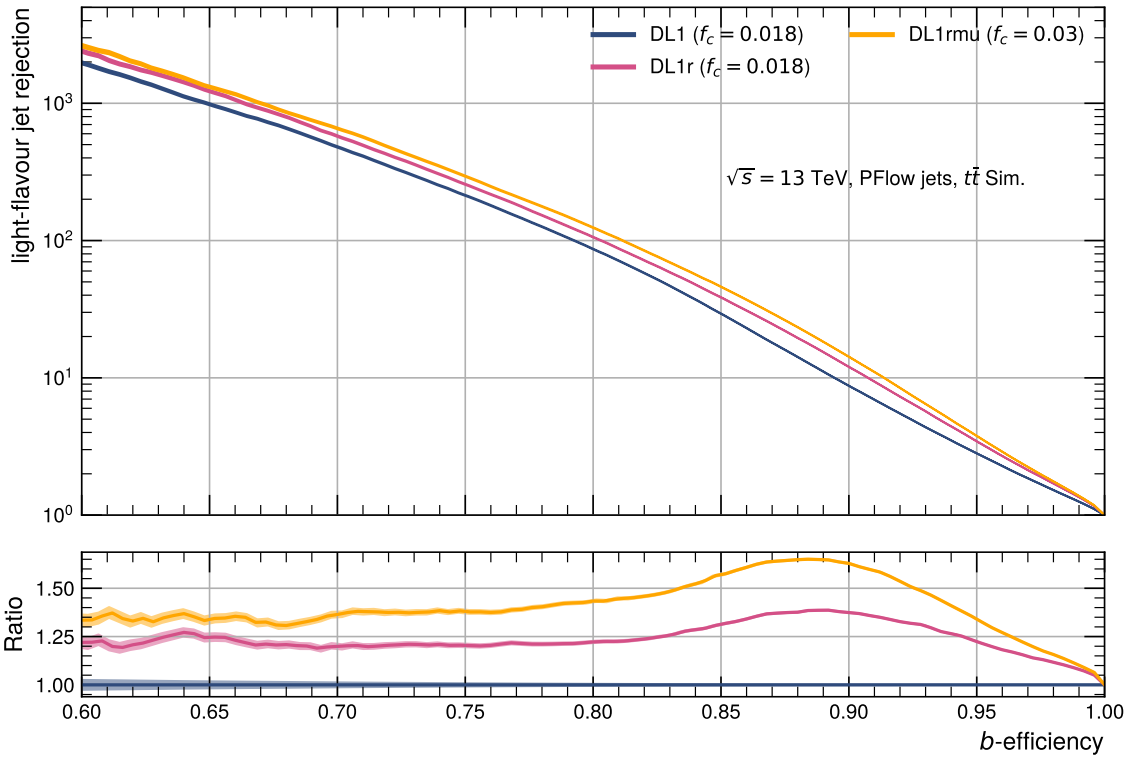


Figure 10.22.: p_T dependent performance for an inclusive b-jet efficiency of 77% defined on an inclusive $t\bar{t}$ sample (official ATLAS cut value) for the (a) c-jet rejection, (b) light-flavour jet rejection and (c) the b-jet efficiency as a function of p_T for the *extended* Z' sample.



(a)

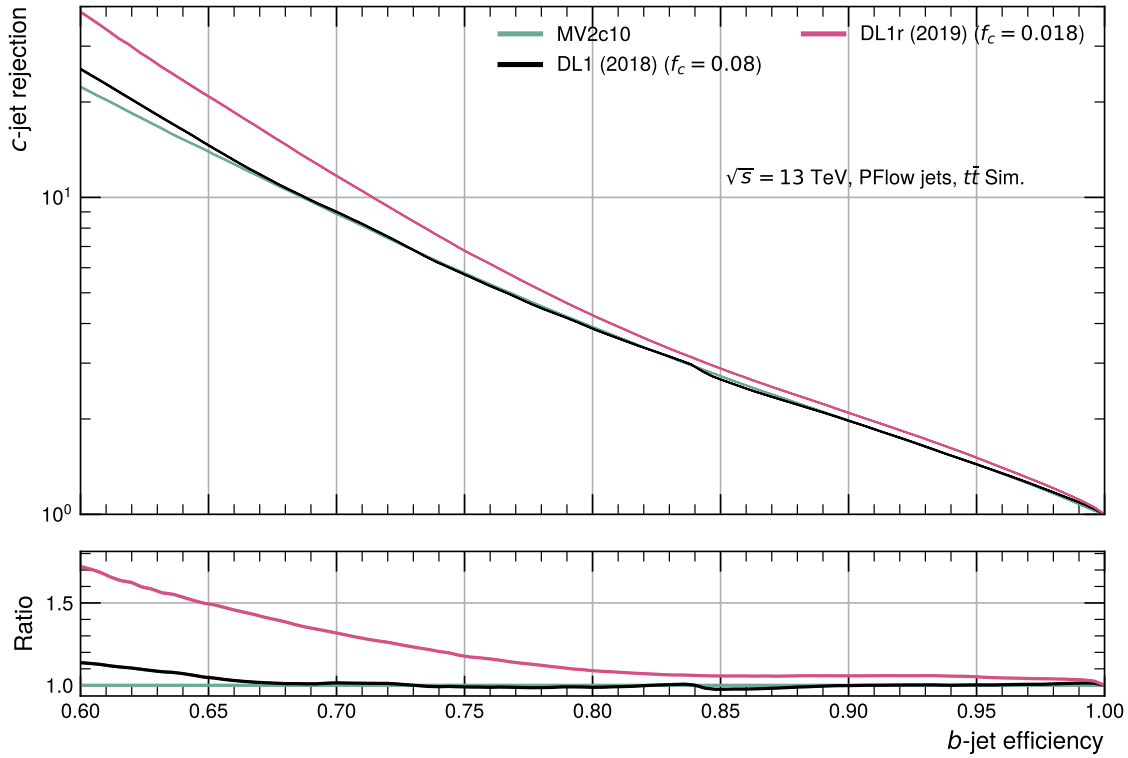


(b)

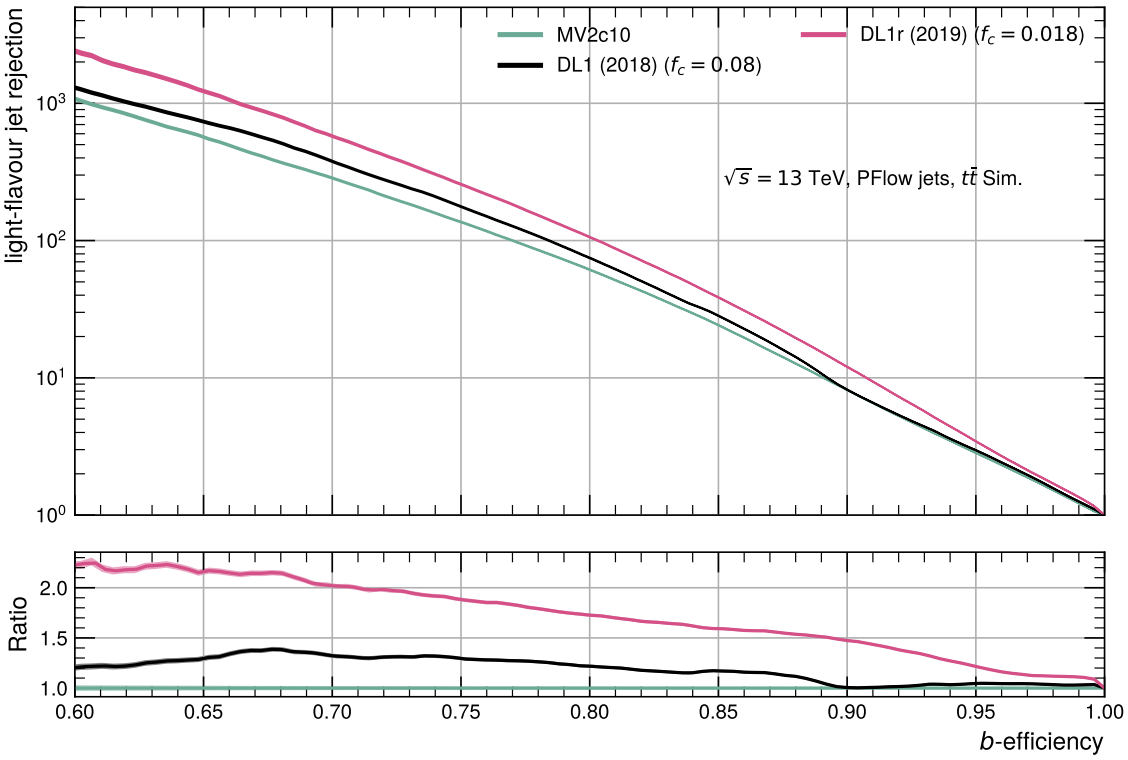
Figure 10.23.: Performance comparison of the three different DL1 tagger versions: *baseline DL1* (blue), *DL1r* (red) and *DL1rmu* (orange) for the (a) c-jet rejection and (b) light-flavour jet rejection as a function of the b-jet efficiency.

10.3.1 *Comparison to previous Recommendations*

Before the *DL1r* taggers were retrained and reoptimised, the recommended taggers were the *MV2c10* tagger and the previous iteration of the *baseline DL1* tagger (DL1 (2018)), both optimised on *EMTopo* jets. The new recommendation is now to use the *DL1r* tagger for particle flow jets, provided to the whole ATLAS collaboration. A comparison of the two previous recommendations and the new recommendation is shown in Figure 10.24. Compared to *MV2c10*, the new *DL1r* improves by up to 70% and 120% for the c-jet and light-flavour jet rejection, respectively, evaluated at the 60% WP. This major improvement leads to gains, especially in analyses which heavily depend on b-tagging such as the $t\bar{t}H(b\bar{b})$ analysis.



(a)



(b)

Figure 10.24.: Performance comparison of the previous recommendations (MV2 and DL1 (2018)) and the newly optimised *DL1r* tagger for the (a) c-jet rejection and (b) light-flavour jet rejection as a function of the b-jet efficiency.

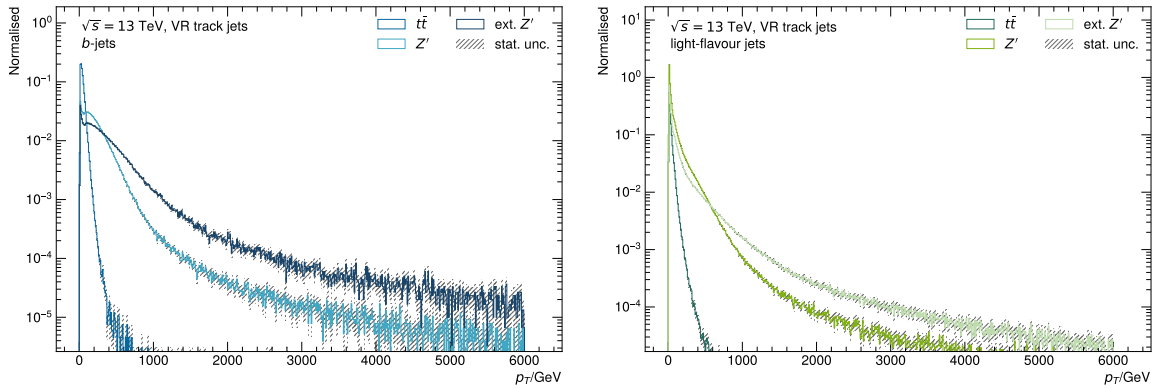
11

VARIABLE RADIUS TRACK JETS TRAINING

Track jets are of special interest for boosted topologies and have a variable radius depending on their jet p_T, described in more detail in Section 6.2.1. The first dedicated training for these jets is presented in the following. The basic training principle with the preprocessing chain and the architecture is taken from the *PFlow* jets training described in chapters 8 and 10. The adaptations for the application to *VR Track* jets and their performance are described in this chapter.

11.1 TRAINING SAMPLE CREATION

The input features for the different tagger versions (see Fig. 9.2) are the same as for *PFlow* jets. The p_T distribution of the *VR Track* jets, as shown in Figure 11.1, is quite different compared to the *PFlow* jets. The p_T spectrum of the *extended* Z' sample is not as flat as for *PFlow* jets for high transverse momenta and consequently, the statistics is not as high in the high p_T regime. Additionally, the *standard* Z' is also shown which has a steeper decrease in p_T than the extended version.



(a) (b)

Figure 11.1.: p_T distributions of the $t\bar{t}$ sample , Z' sample and *extended* Z' for (a) b-jets and (b) light-flavour jets.

The hybrid sample definition for the *VR Track* jets turned out to not be as straight forward as for *PFlow* jets. Already in the retraining of RNNIP for *VR Track* jets, it was found that the *extended* Z' sample hurt the training and the training on the hybrid sample could not reproduce the training on the dedicated samples ($t\bar{t}$ and *extended* Z' separately). The solution in this case was to use only the leading and subleading jets from the *standard* Z' sample as well as lowering the sample transition cut to $p_T = 125\,\text{GeV}$.

For the DL1 training, these studies were repeated and also here it was observed that the hybrid

definition from *PFlow* jets applied to *VR Track* jets degrades the performance (background rejection). The p_T distributions for b- and light-flavour jets in Figure 11.2 reveal several differences among them. The p_T distributions are shown for all jets in an event (inclusive), for the four leading jets (the four jets with the highest p_T in an event) and the two leading jets. The number of light-flavour jets at low p_T is reduced when considering the four leading jets and even more by only considering the first two leading jets. This is expected since soft initial and final state radiation is dominated by light constituents. Since the $t\bar{t}$ sample is used to populate the lower p_T range, the four leading jets are chosen to be taken into account for the training.

The *standard* Z' sample retains more statistics compared to the *extended* Z' by requiring only the first two or four leading jets. Both scenarios: using either the *standard* Z' or the *extended* Z' in the training were investigated. Different training studies were performed to optimise the hybrid composition. The *extended hybrid* sample showed strong overfitting and the training only converged by regulating it with *Dropout*. However, *Dropout* introduced the uneven distributions of the output discriminants as already described for *PFlow* jets in Section 10.1.1. In addition, the training with the *standard* Z' sample showed adequate results when applied to the *extended* Z' sample which was not the case the other way around. The additional advantage is that the *standard* Z' sample is available for all three data taking periods and thus more statistics can be used in the training. Since Z' is used for the higher p_T regime, only the two leading jets are utilised. To ensure a smooth transition between the two samples, the p_T cut for $t\bar{t}$ is chosen to be higher (400 GeV) compared to the *PFlow* jets, the lower p_T cut for Z' is reduced to 125 GeV and an upper limit of 3 TeV is applied. All hybrid selections are summarised in Table 11.1.

The choice of the hybrid training sample turned out to be the most critical part of the *VR Track* jets training. Once this was solved the further training was analogous to the *PFlow* jets procedure.

Jet Flavour	$t\bar{t}$	*standard* Z'
all flavours	4 leading jets	2 leading jets
b-jets	b-hadron $p_T < 400\,\text{GeV}$	$125\,\text{GeV} <$ b-hadron $p_T < 3\,\text{TeV}$
c-jets light-flavour jets	jet $p_T < 400\,\text{GeV}$	$125\,\text{GeV} <$ jet $p_T < 3\,\text{TeV}$

Table 11.1.: Selections used to build the hybrid training sample for *VR Track* jets.

The resampled p_T and $|\eta|$ distributions are illustrated in Figure 11.3. The 2-D binning has been optimised for the *VR Track* jets to cope for the different p_T range with respect to *PFlow* jets. In total, 20M jets are left after the resampling and can be used for the training. The resulting variable distributions with the full preprocessing applied are shown in Figures 11.4 and 11.5.

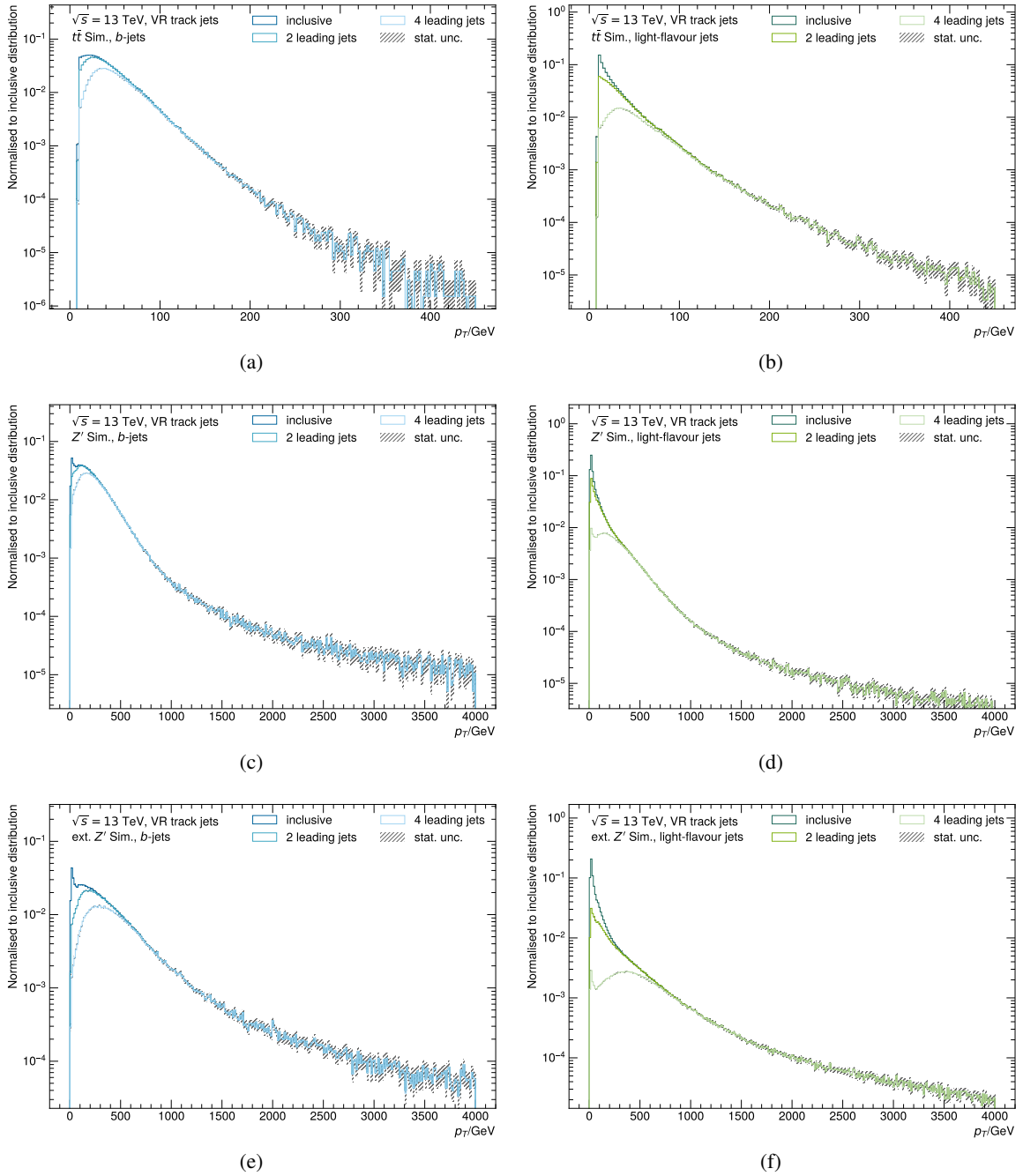
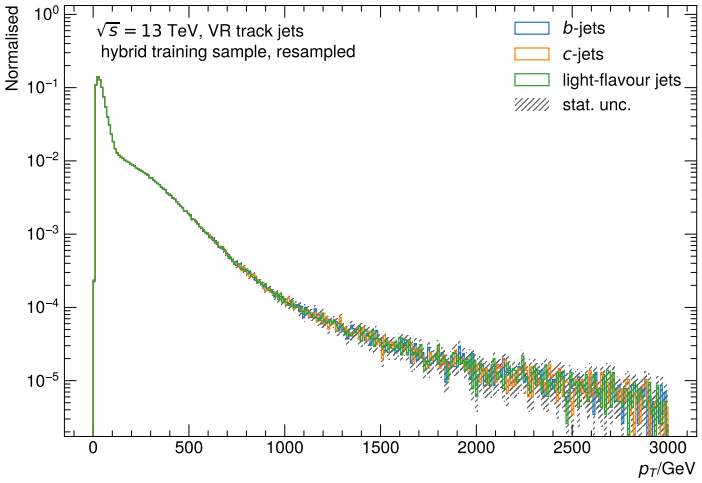
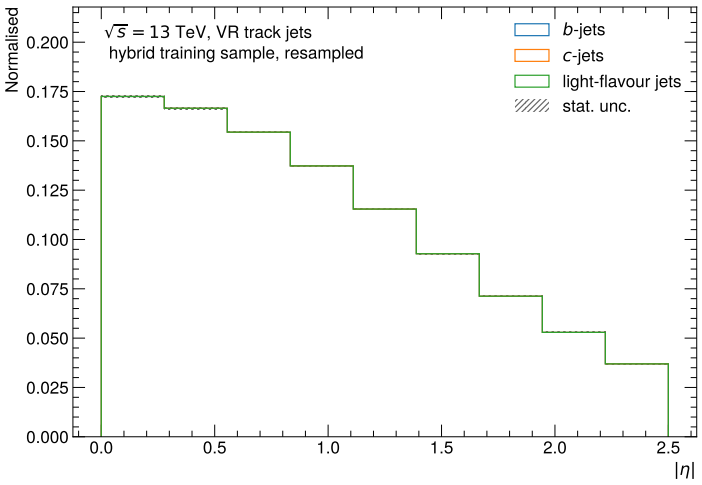


Figure 11.2.: p_T distributions for all jets in an event (inclusive), the four leading jets and the two leading jets for the $t\bar{t}$ sample (a) & (b), Z' (c) & (d) and *extended* Z' (e) & (f). The left column shows the b-jet p_T distributions and the right column the light-flavour jet p_T spectra. The distribution with the label *inclusive* is normalised to unity and all other distributions are then normalised to the *inclusive* integral.



(a)



(b)

Figure 11.3.: (a) the p_T and (b) the $|\eta|$ distribution of the hybrid sample resampled to match the b-jet spectrum.

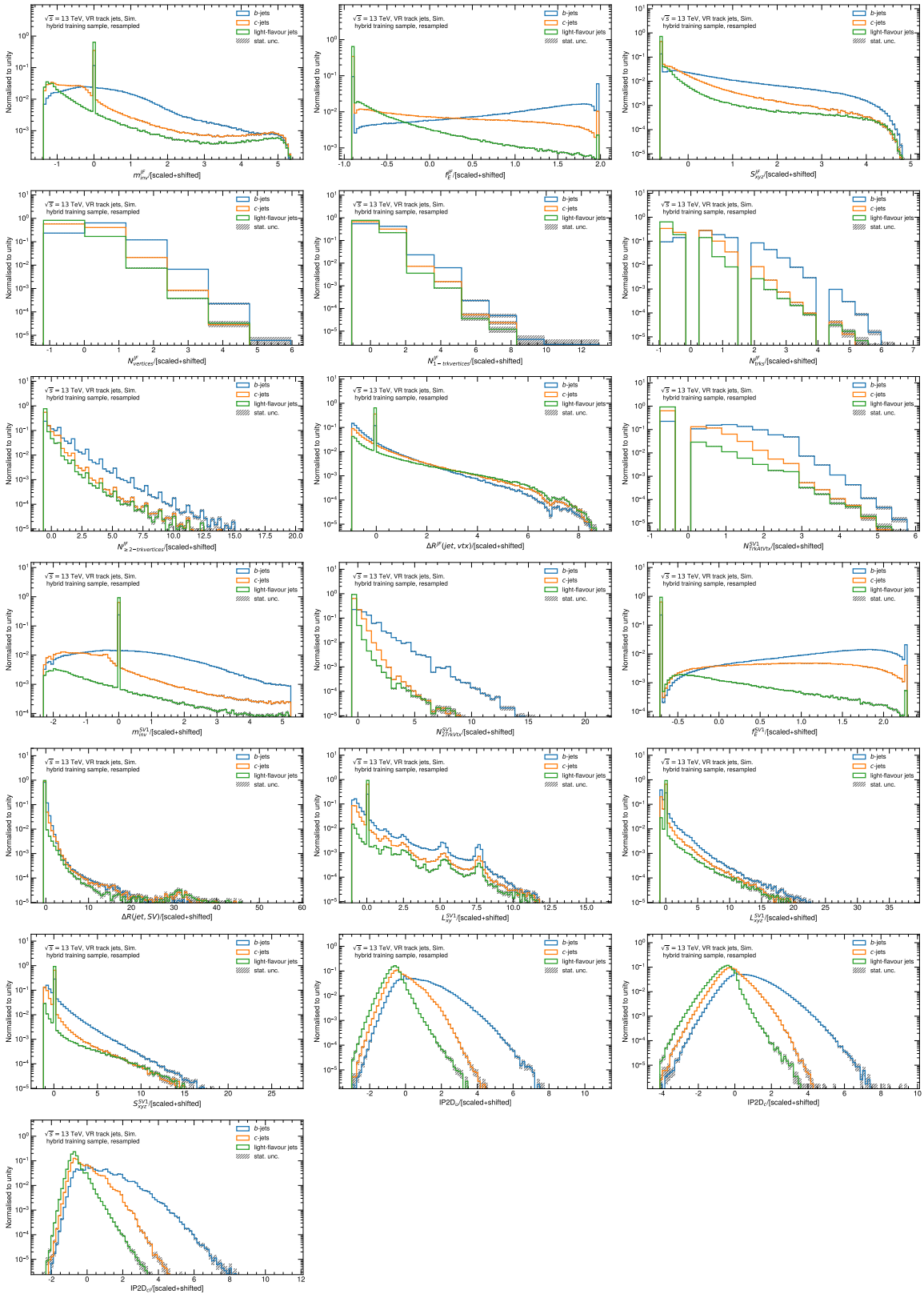


Figure 11.4.: First part (of 2) of input variables used for the DL1 tagger family training, fully preprocessed (resampled, scaled, shifted and default values replaced).

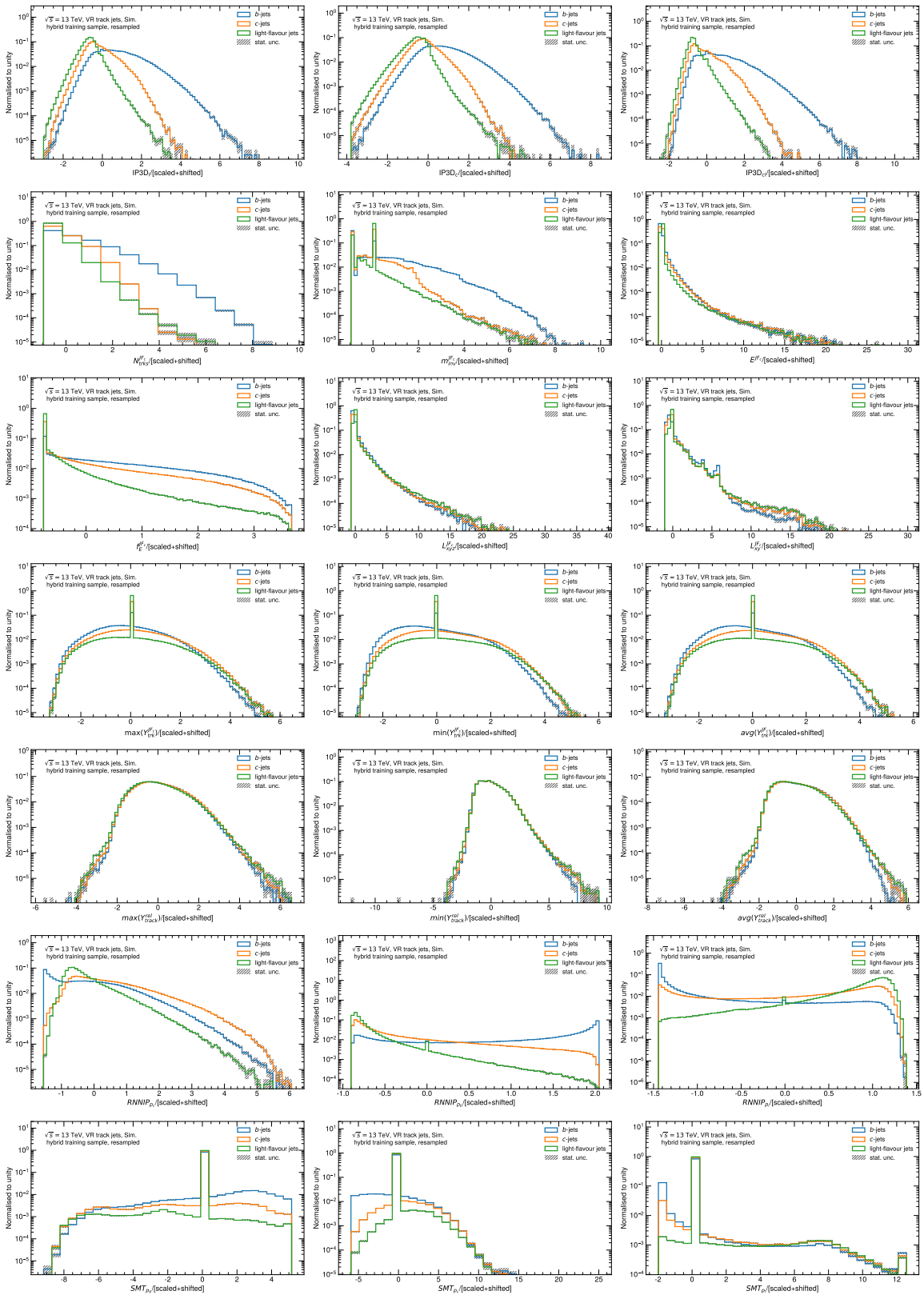


Figure 11.5.: Second part (of 2) of input variables used for the DL1 tagger family training, fully preprocessed (resampled, scaled, shifted and default values replaced).

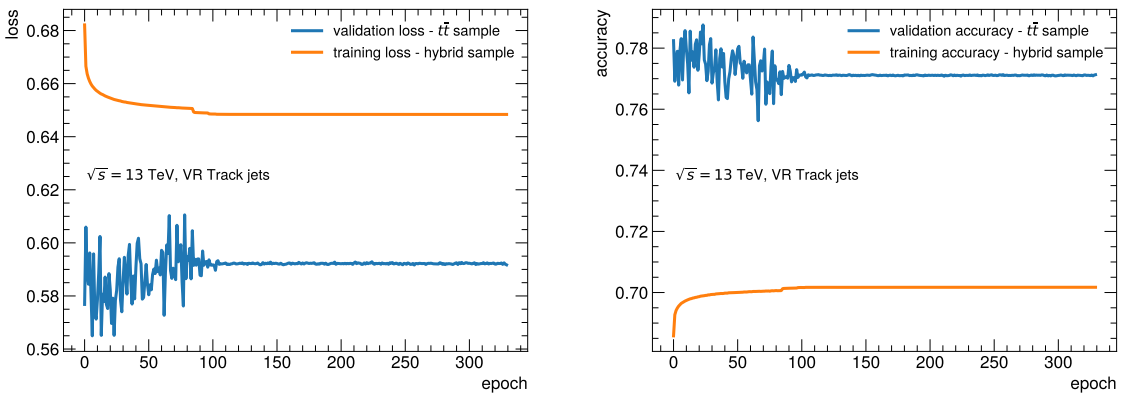
11.2 TAGGER TRAINING

After the hybrid training sample was optimised ensuring a robust training, the training itself is similar to the *PFlow* jets training. Similar to the hyperparameter optimisation described in Section 10.2 a small scan is performed with again no better configuration found than the by-hand designed architecture which is summarised in Table 11.2. In fact, the architecture is the same as for *PFlow* jets (see Table 10.4) except the increased training batch size.

Hyperparameter	**Values**
$N_{\text{hidden layers}}$	8
$N_{\text{nodes/layer}}$	[256, 128, 60, 48, 36, 24, 12, 6]
learning rate	0.01
Training batch size	45000
Activation function	ReLu
Free (trainable) parameters	59,275
Fixed parameters	1,140
Training sample statistic	20M

Table 11.2.: Final network architecture for the *DL1r* tagger for *VR Track* jets.

The training is monitored to ensure that no overfitting is occurring and to see if the training is robust. Figure 11.6 shows the loss and accuracy for both the training and validation sample. Also here, the loss and accuracy converge after around 110 epochs. After a first decrease in the training



(a) (b)

Figure 11.6.: Training and validation (a) loss and (b) accuracy as a function of the training epoch. *DL1r* training with 20M training jets as indicated in Table 11.2.

loss, it stays constant around epoch 60 until a small drop occurs around epoch 85 and again a smaller drop around epoch 95. Interestingly, after these drops, the validation loss stabilises more and more. These drops are coming from the learning rate scheduler which reduces the learning rate, when the training loss did not change after a certain amount of epochs. Often, if the learning rate is too large,

the network is jumping over the minimum and by decreasing the learning rate it can help the network to find the minimum. This seems to exactly happen in this case.
In addition, the background rejection at the 77% WP is monitored for the validation sample illustrated in Figure 11.7. After fluctuations in the first epochs, similar to the loss, the background rejection also converges. The network does not show signs for overfitting, on the one hand verified by the convergence of the validation loss and on the other hand, the same performance comparison as shown for *PFlow* jets in Figure 10.8 also yielded the same conclusion.
The f_c values were chosen to be identical to the values from the *PFlow* jets which provided again a good compromise between the c- and light-flavour jet rejection.

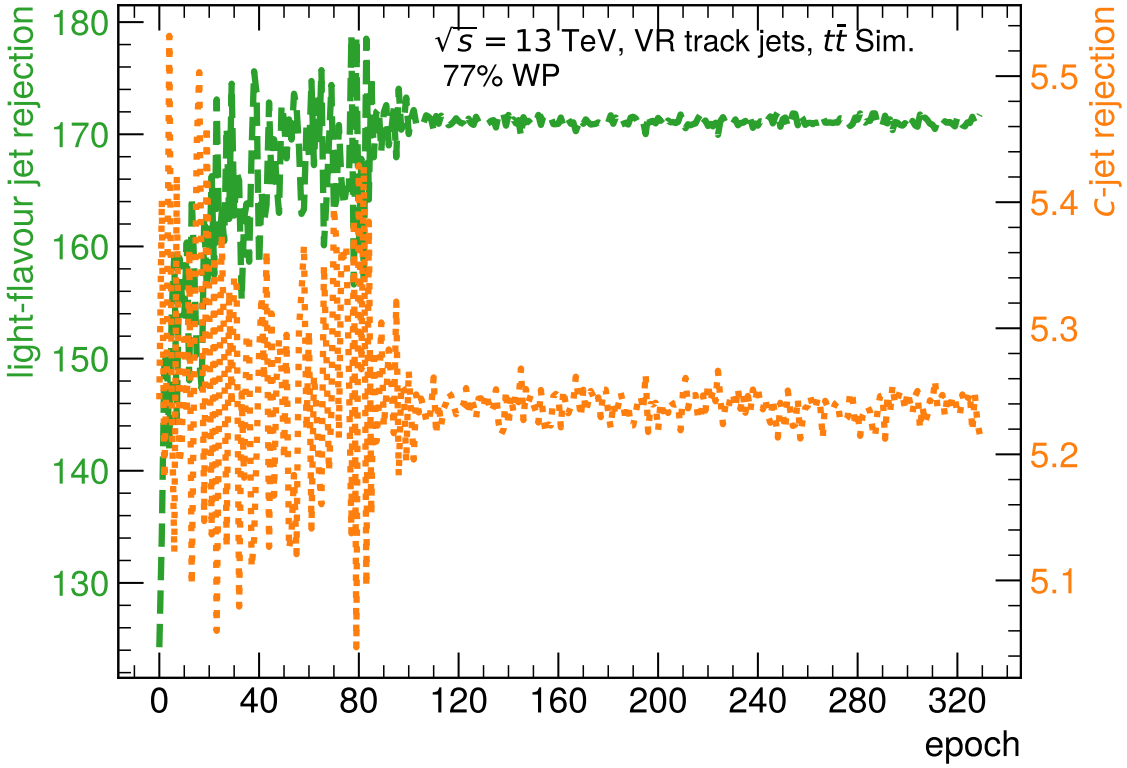


Figure 11.7.: Light-flavour jet rejection (green) and c-jet rejection (orange) as a function of the training epoch evaluated on the validation $t\bar{t}$ sample during the training. *DL1r* training with 20M training jets as indicated in Table 11.2.

11.3 PERFORMANCE OVERVIEW AND TAGGER VARIANTS COMPARISON

The architecture of the *DL1r* tagger defined in Table 11.2 is used also for the *baseline DL1* and *DL1rmu* training. The resulting final b-tagging discriminant is shown in Figure 11.8 for all three variants. The same behaviour as for the *PFlow* jets training is also observed here, the distributions are smooth enough to allow a good WP definition and the b-jets are peaking more towards higher values the more information is used in the training.

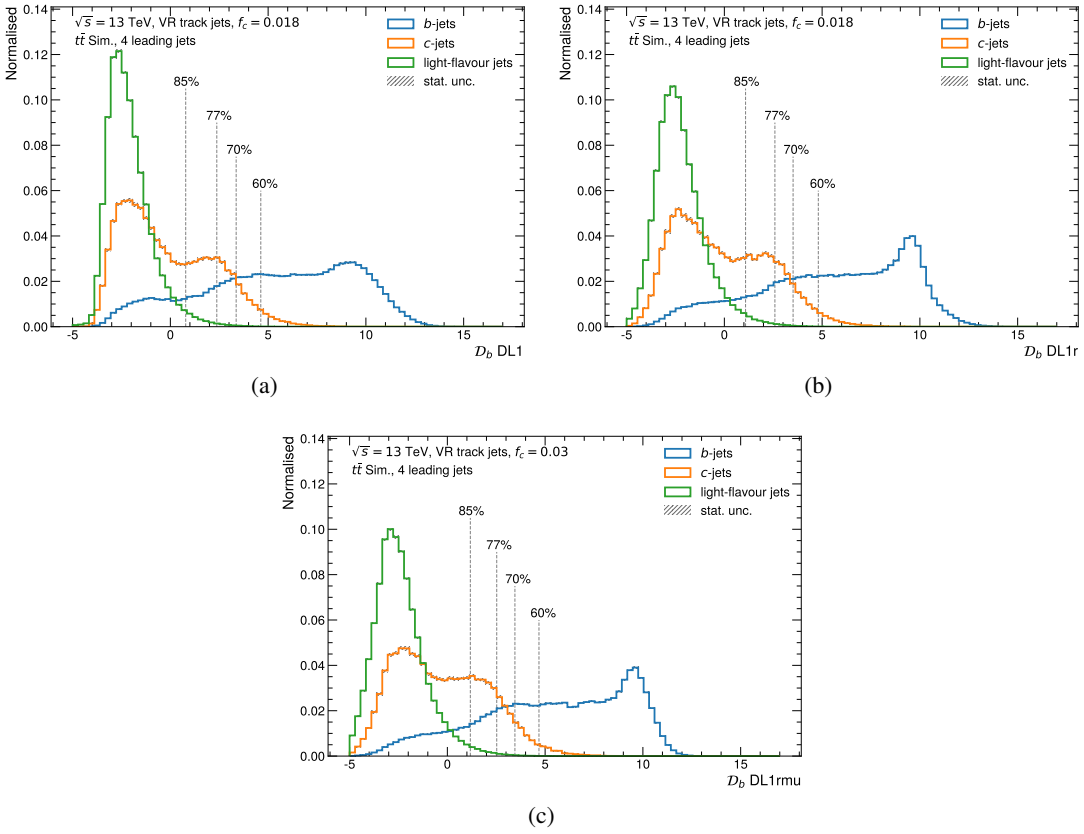
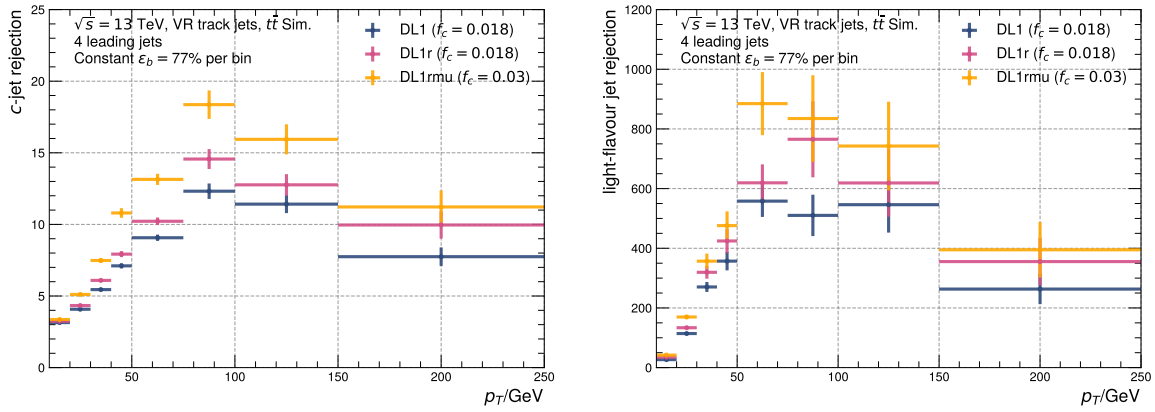


Figure 11.8.: The (a) *baseline DL1*, (b) *DL1r* and (c) *DL1rmu* discriminants, the vertical dashed lines are indicating the WPs.

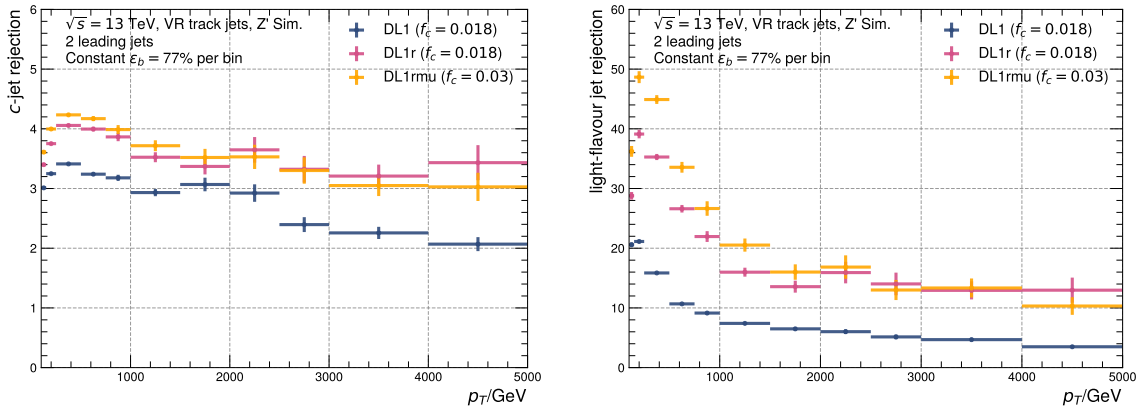
The p_T dependent performance comparisons are shown in Figures 11.9-11.11. A constant efficiency per p_T bin is maintained for the plots in Figure 11.9 and 11.10. For lower transverse momenta, the performance is evaluated on a $t\bar{t}$ sample (Fig. 11.9) where the *DL1r* tagger outperforms the *baseline DL1* tagger and the *DL1rmu* tagger again the *DL1r* tagger in each p_T bin whereby especially for the last two bins the performance differences are well compatible between the three variants within their uncertainties. The higher transverse momenta regime is investigated on a *standard* Z' sample (Fig. 11.10). Also similar to the observations for *PFlow* jets, the biggest performance gain at high p_T is achieved by employing the RNNIP information.

The performance as a function of p_T for the 77% fixed cut WP is shown in Figure 11.11. The b-jet efficiency of the $t\bar{t}$ sample (Fig. 11.11 (e)) is consistent between the three tagger variants, also here it is not constant over the full p_T range but raises in the beginning and flattens out. This is again



(a) (b)

Figure 11.9.: p_T dependent performance for a constant b-jet efficiency of 77% per bin for the (a) c-jet rejection and (b) light-flavour jet rejection for the $t\bar{t}$ sample.



(a) (b)

Figure 11.10.: p_T dependent performance for a constant b-jet efficiency of 77% per bin for the (a) c-jet rejection and (b) light-flavour jet rejection for the *extended* Z' sample.

in contrast to the behaviour at high transverse momenta demonstrated on the *standard* Z' sample (Fig. 11.11 (f)). In this case, the b-jet efficiency decreases with p_T, however, not as drastically as for *PFlow* jets. Similarly, the *baseline DL1* b-jet efficiency falls faster than for *DL1r* and *DL1rmu*. While the performance order is retained on the $t\bar{t}$ sample for the b-jet rejection (Fig. 11.11 (a)) and light-flavour-jet rejection (Fig. 11.11 (c)), i.e. *baseline DL1* $<$ *DL1r* $<$ *DL1rmu*, this order is inverted for the c-jet rejection at high p_T on the *standard* Z' sample (Fig. 11.11 (b)) due to the lower b-jet efficiency. However, the light-flavour jet rejection (Fig. 11.11 (d)) does not suffer so much from this effect. Still, it is also visible at $p_T > 2$ TeV but in this regime, the statistical uncertainties are so large that they cover this effect.

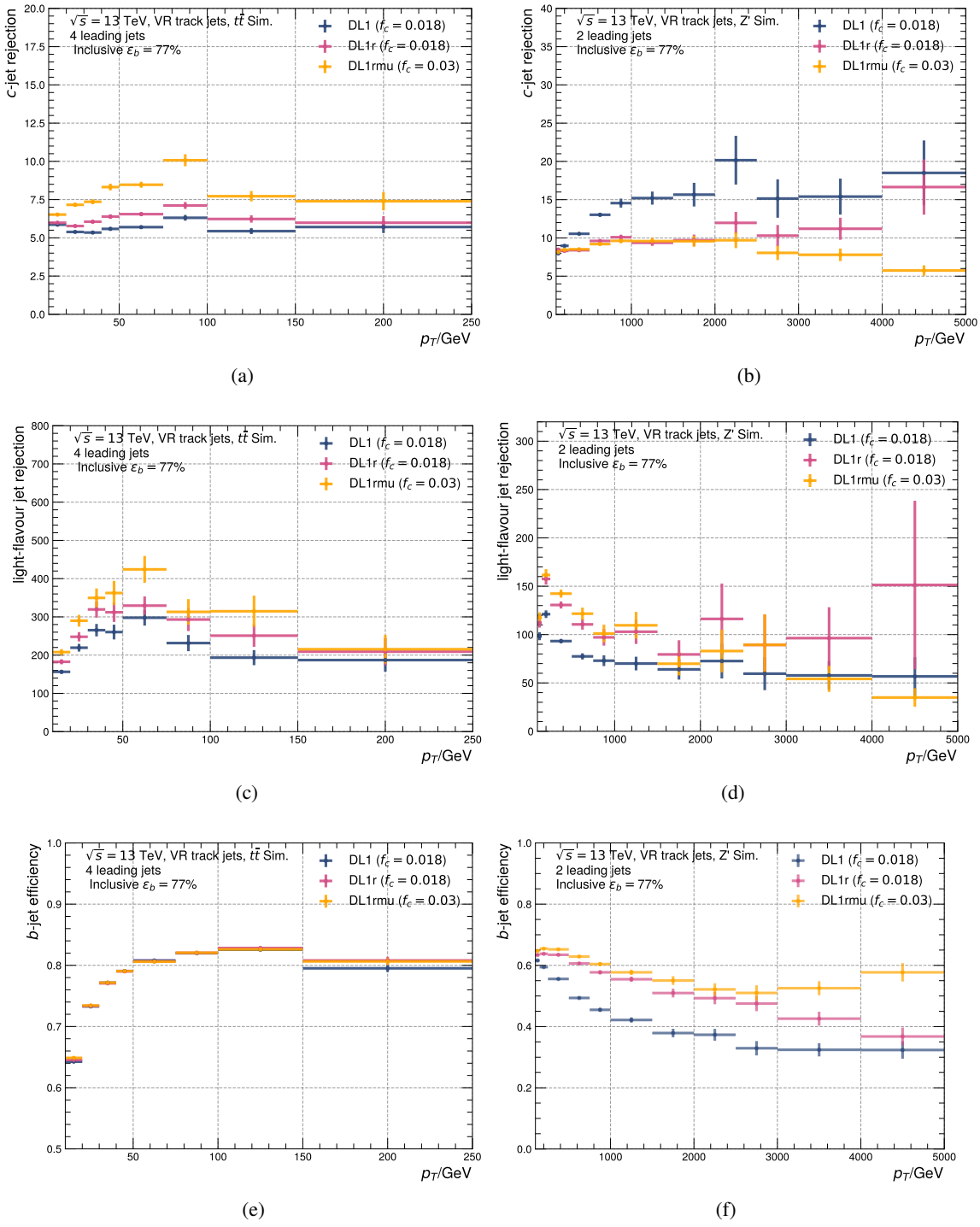
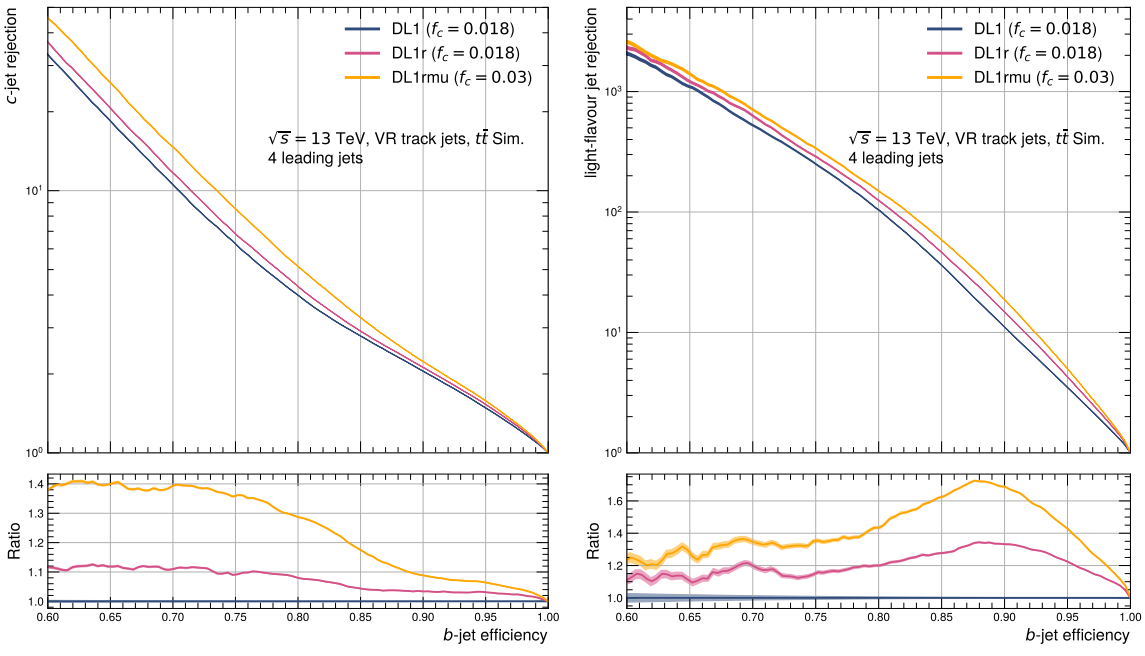


Figure 11.11.: p_T dependent performance for an inclusive b-jet efficiency of 77% defined on an inclusive $t\bar{t}$ sample (official ATLAS cut value) for the (a) & (b) c-jet rejection, (c) & (d) light-flavour jet rejection and (e) & (f) the b-jet efficiency. The evaluation is done on a $t\bar{t}$ sample (left column) and on a *standard* Z' sample (right column).

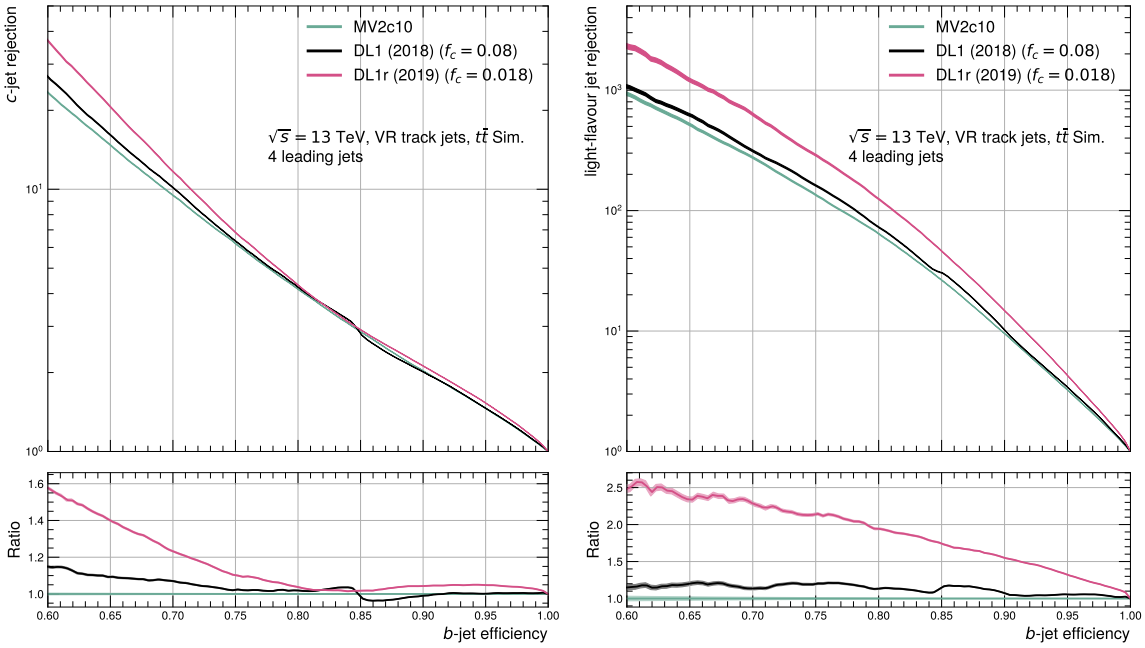
The inclusive p_T performance as a function of the b-jet efficiency is shown in Figure 11.12 comparing the three tagger variants. A similar picture emerges compared to *PFlow* jets. The additional RNNIP information improves the c-jet and the light-flavour jet rejection by about 10% (with some fluctuations) to a b-jet efficiency of 80%. At this point, the light-flavour jet rejection increases while the c-jet rejection decreases. The general trend when including the additional soft muon information follows the one of *DL1r* with an increased performance of about 10% in the light-flavour jet rejection (increases slightly more at a b-jet efficiency of 80%). Additionally, an improvement of around 25% in terms of the c-jet rejection is achieved up to a b-jet efficiency of 75% where it starts decreasing again. The larger improvement in the c-jet rejection compared to the light-flavour rejection is due to the f_c value of 0.03. It was decided to keep this consistent with the *PFlow* jets results since in general the *DL1rmu* tagger cannot be calibrated.

Also, for *VR Track* jets the tagger *DL1r* will become the new recommendation for physics analyses in ATLAS. Thus the comparison to the previous recommendation in ATLAS (MV2c10 and *baseline DL1* optimised on *EMTopo* jets) is important to quantify the improvements achieved with the newly optimised *DL1r* tagger. Figure 11.13 shows this comparison as a function of the b-jet efficiency with large improvements in both c- and light-flavour jet rejection of up to 60% and 120% at 60% WP, respectively. The light-flavour jet rejection performance gain is more constant over the full b-jet efficiency spectrum while for the c-jet rejection there is almost no gain for b-jet efficiencies larger than 80% and the rejection is only slightly better for b-jet efficiencies around 75%.



(a) (b)

Figure 11.12.: Performance comparison of the three different DL1 tagger versions: *baseline DL1* (blue), *DL1r* (red) and *DL1rmu* (orange) for the (a) c-jet rejection and (b) light-flavour jet rejection as a function of the b-jet efficiency.



(a) (b)

Figure 11.13.: Performance comparison of the previous recommendations (MV2 and DL1 (2018)) and the newly optimised *DL1r* tagger for the (a) c-jet rejection and (b) light-flavour jet rejection as a function of the b-jet efficiency.

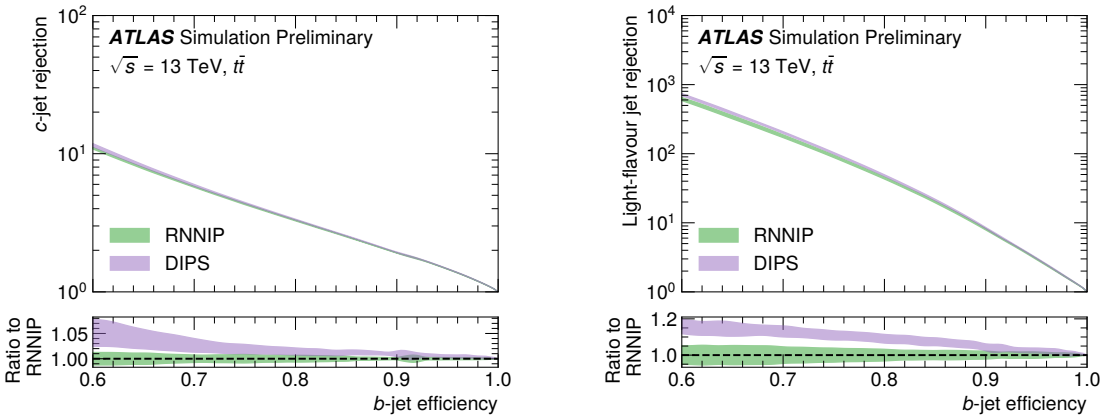
12

ONGOING DEVELOPMENTS IN FLAVOUR TAGGING

The retraining campaign described in the two previous chapters delivered a large performance gain for b-tagging in ATLAS. The NN-based tagger architectures provide a flexible basis for further developments in this domain. The next logical step is to perform a joint training of the DL1 tagger and the successor of RNNIP, described in more detail below.

12.1 UMAMI

Inspired from the Japanese word UMAMI for the fifth basic taste, or the fifth flavour, a new end-to-end tagger is being developed. It joins the developments of the high-level tagger optimised in this thesis and the recent improvements in the track based tagger Deep Impact Parameter Sets (DIPS) [137], the successor of RNNIP. The advantage of the Deep Sets architecture [198] of the DIPS tagger compared to RNNIP is its sum pooling layer. A pooling operation is a lossy summary of a set of features. Mostly an average, maximum or sum pooling is used which takes the average of several layers, the maximum value or their sum. For RNNIP the order in which the tracks are passed to the network matters due to its underlying Recurrent Neural Network structure. The tracks are ordered by the signed impact parameter s_{d_0}. However, the tracks originating from b-hadron decays do not have a physically motivated ordering per se. With the new architecture, the ordering is obsolete since the sum pooling layer is permutation invariant. As demonstrated in Ref. [137], and shown in Figure 12.1, the DIPS architecture outperforms RNNIP. Another important advantage is its parallelisability which reduces the training time by more than a factor three [137] as well as its evaluation time.



(a) (b)

Figure 12.1.: Performance comparison of the RNNIP (green) and the DIPS (purple) algorithms in terms of the (a) c-jet rejection and (b) the light-flavour jet rejection as a function of the b-jet efficiency. The uncertainty bands originate from the standard deviation of five trainings [137].

12.1.1 *DIPS Network*

The DIPS architecture is shown in Figure 12.2 as a part of the UMAMI architecture. The sub-networks ϕ and F represent the DIPS part. Each track is characterised by its track variables listed in Table 12.1. The corresponding variable distributions are shown in Figure 12.3 for the hybrid sample composed of $t\bar{t}$ and *extended* Z' for *PFlow* jets as it was used for the DL1 *PFlow* jets training (see Section 9.2 for more details). The scaling and shifting of the variables is described in Ref. [137]. Due to the long tails in the p_T^{frac} and the ΔR distributions, their logarithms are used. The shifting and scaling is only applied to the following variables: $\log(p_T^{frac})$ (Fig. 12.3 (c)), $\log(\Delta R)$ (Fig. 12.3 (d)), nPixHits (Fig. 12.3 (i)) and nSCTHits (Fig. 12.3 (l)). All other variables have already a mean close to zero and are therefore not shifted and the differences in their values do not necessitate an additional scaling. Each track in a jet is first processed through a network ϕ as indicated in Figure 12.2. In the next step, all n track networks, corresponding to the number of tracks in a jet, are summed up and further processed via the network F. This can be summarised into

$$\vec{p}_i = F\left(\sum_{i=1}^{n} \phi(\vec{x}_i^t)\right), \tag{12.1}$$

where $\vec{x}_i^t$ are the track input features and $\vec{p}_i$ is the vector of the b-, c- and light-flavour jet class probabilities corresponding to the DIPS output nodes.

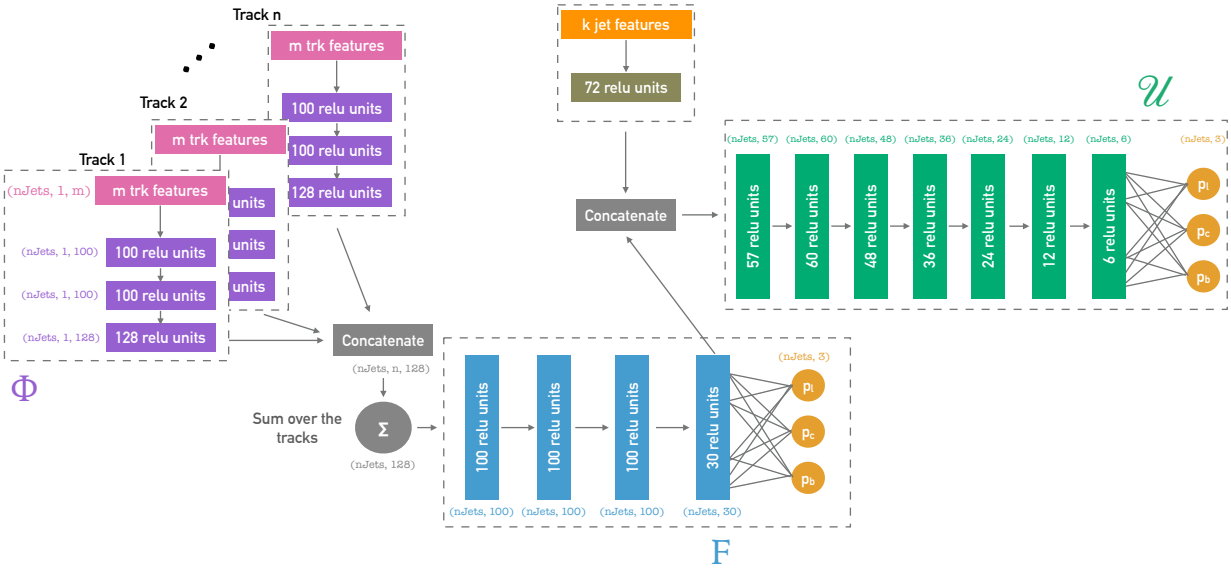


Figure 12.2.: Network structure of the combined DL1 and DIPS tagger UMAMI.

Input	Description	Preprocessed
s_{d_0}	d_0/σ_{d0}: Transverse IP significance	
s_{z0}	$z_0 \sin\theta/\sigma_{z0\sin\theta}$: Longitudinal IP significance	
$\log p_T^{frac}$	$\log p_T^{track}/p_T^{jet}$: Logarithm of fraction of the jet p_T carried by the track	✓
$\log \Delta R$	Logarithm of opening angle between the track and the jet axis	✓
IBL hits	Number of hits in the IBL: could be 0, 1, or 2	
PIX1 hits	Number of hits in the next-to-innermost pixel layer: could be 0, 1, or 2	
shared IBL hits	Number of shared hits in the IBL	
split IBL hits	Number of split hits in the IBL	
nPixHits	Combined number of hits in the pixel layers	✓
shared pixel hits	Number of shared hits in the pixel layers	
split pixel hits	Number of split hits in the pixel layers	
nSCTHits	Combined number of hits in the SCT layers	✓
shared SCT hits	Number of shared hits in the SCT layers	

Table 12.1.: Input features for the DIPS algorithm [137]. The right column *Preprocessed* indicates if the variables are shifted to mean zero and scaled to significance of one.

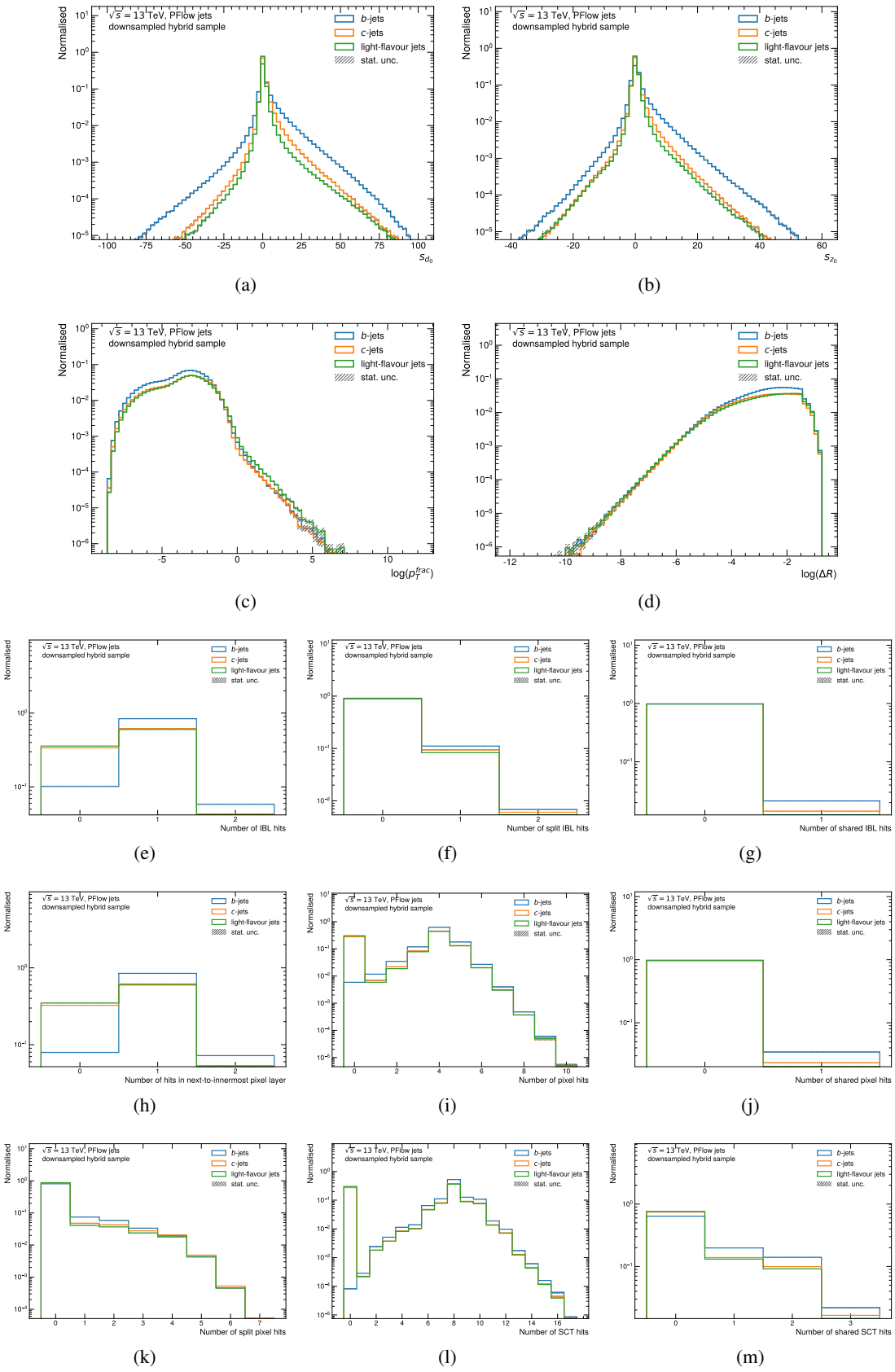


Figure 12.3.: Track variables used for the DIPS training extracted from the downsampled hybrid sample: (a)&(b) the impact parameters, (c) the $\log p_T^{frac}$, (d) $\log \Delta R$ and (e)-(m) the number of hits in the different ID layers.

12.1.2 *Joint Architecture*

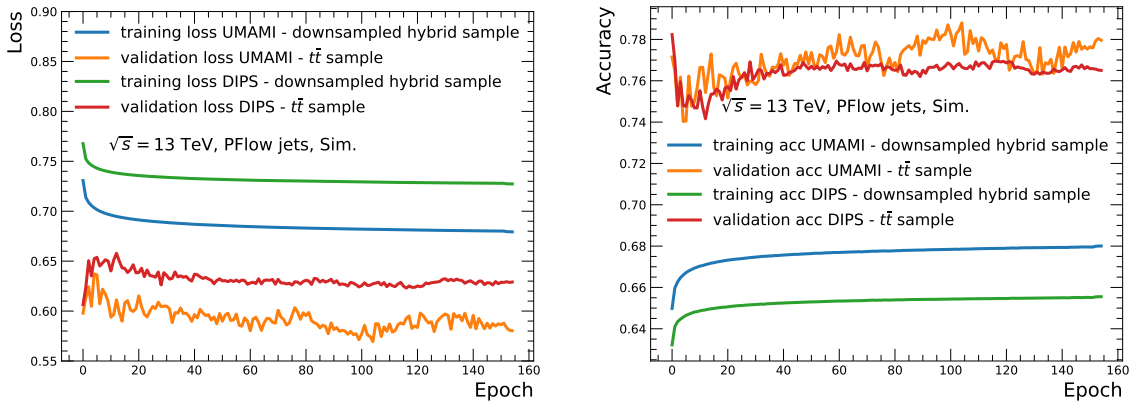
The *DL1r* tagger already used the information of the RNNIP in the training. However, the used RNNIP information was limited to the outputs of the standalone RNNIP algorithm which are the three flavour class probabilities. A joint architecture allows to pass more information to the *jet-network* which is the $\mathcal{U}$ NN in the sketch 12.2. In this case, a layer with 30 nodes (units) is concatenated with the other jet features which are processed through one NN layer with 72 nodes. Apart from this, a full back-propagation up to all the track NNs ϕ is done in the training allowing a joint optimisation. The DIPS part has also an intermediate loss which maintains the possibility to evaluate the DIPS performance separately and also is an implicit help for the network to optimise this network branch. After the combination of the DIPS part and the jet features a final feed-forward network $\mathcal{U}$ is employed inspired by the DL1 architecture with also three output nodes corresponding to the flavour probabilities.
Both networks, F and $\mathcal{U}$ have their dedicated losses. While the loss of the $\mathcal{U}$ network $\mathcal{J}(\mathcal{U})$ is sensitive to the track and the jet features, the loss of the F network $\mathcal{J}(\mathsf{F})$ is only sensitive to the tracks. The overall optimisation is performed on the combined loss

$$\mathcal{J}(\text{comb.}) = \mathcal{J}(\mathcal{U}) + \lambda \cdot \mathcal{J}(\mathsf{F}) = \mathcal{J}(\text{UMAMI}) + \lambda \cdot \mathcal{J}(\text{DIPS}), \tag{12.2}$$

where the parameter λ defines the importance of the two losses.

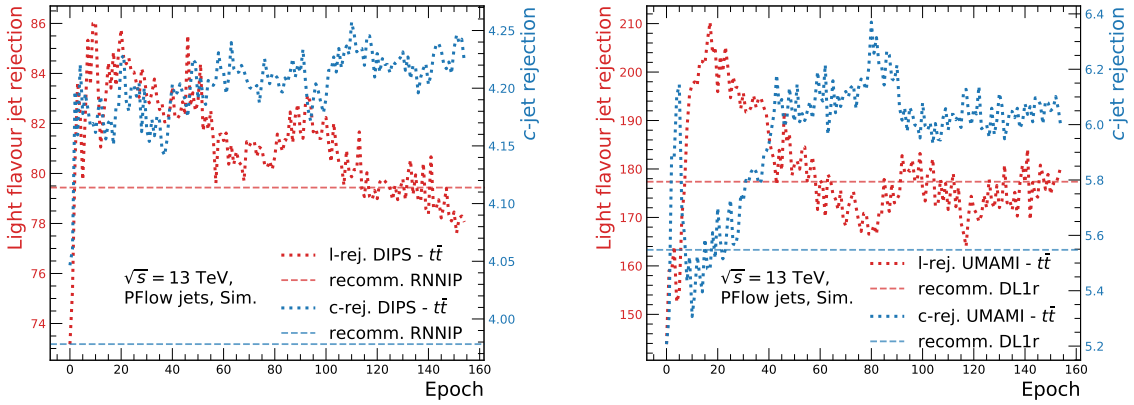
12.1.3 UMAMI *Training*

The training of the UMAMI tagger is only in a preliminary phase and no dedicated optimisation studies were performed so far. Due to technical reasons, only the standard version of DIPS from [137] is studied and not the optimised version with a looser track selection. The λ parameter from Equation (12.2) is set to 1 and a training statistic of 6M jets is used. The corresponding loss and accuracy curves for the training and validation set are shown in Figure 12.4. As expected, the loss $\mathcal{J}(\mathcal{U})$ is better (lower) than the DIPS loss $\mathcal{J}(\mathsf{F})$ since $\mathcal{J}(\mathcal{U})$ encodes the track and jet information. The training losses for DIPS and UMAMI show a constant downwards trend also still after 160 epochs. In general, the DIPS validation loss and accuracy are fluctuating more than for UMAMI. Around epoch 100 it seems that the DIPS part is starting to overfit, the loss is increasing. Figure 12.5 shows the background rejection as a function of the training epochs for both sets of the network outputs (DIPS and UMAMI). Here, the same effect is visible, while the c-jet rejection stays almost constant after epoch 100 (Fig. 12.5 (a)), the light-flavour jet rejection decreases. This behaviour also propagates to the UMAMI performance where the light-flavour jet rejection starts decreasing as well around epoch 100 but seems then to stabilise again, correcting in some sense for this behaviour of the DIPS network. A similar effect is already visible around epoch 45 for the light-flavour jet rejection but recovering again slightly around epoch 80.



(a) (b)

Figure 12.4.: Epoch-dependent training and validation loss (a) and accuracy (b) for the DIPS and UMAMI outputs.



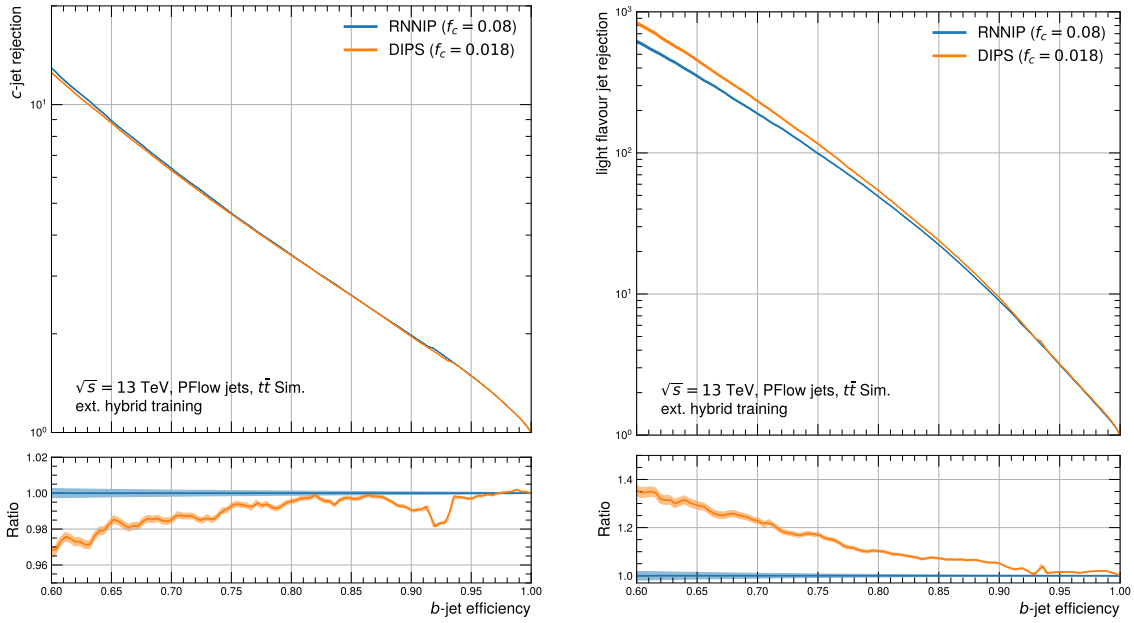
(a) (b)

Figure 12.5.: Light-flavour jet rejection (red) and c-jet rejection (blue) as a function of the training epochs for the (a) DIPS and (b) UMAMI output nodes of the tagger. The vertical dashed lines indicate the performance of the current recommendations RNNIP and *DL1r*.

The UMAMI background rejection reveals a slightly different picture. The light-flavour jet rejection reaches a maximum around epoch 20 and then continuously decreases until it stabilises more around epoch 90 with the small decrease and recovery mentioned above. The c-jet rejection follows an opposite trend, after some larger fluctuations in the beginning, it starts to increase and reaches after certain fluctuations a still pretty noisy plateau after epoch 100.

To avoid the regions where overfitting occurs, an epoch around the maximum for the light-flavour jet rejection in UMAMI is used for further comparisons, corresponding to epoch 17. The background rejection versus b-jet efficiency plots are shown in Figure 12.6 for the intermediate DIPS output and for UMAMI in Figure 12.7. DIPS outperforms RNNIP in terms of the light-flavour jet rejection and is slightly worse for the c-jet rejection. This effect can be compensated by adapting the f_c parameter. Thus the intermediate DIPS output of the joint UMAMI training provides already promising results. The final UMAMI performance (Fig. 12.7) improves compared to *DL1r* in the light-flavour jet rejec-

tion over the full b-jet efficiency range and degrades for the c-jet rejection. Again, these differences can be compensated with the f_c parameter.



(a) (b)

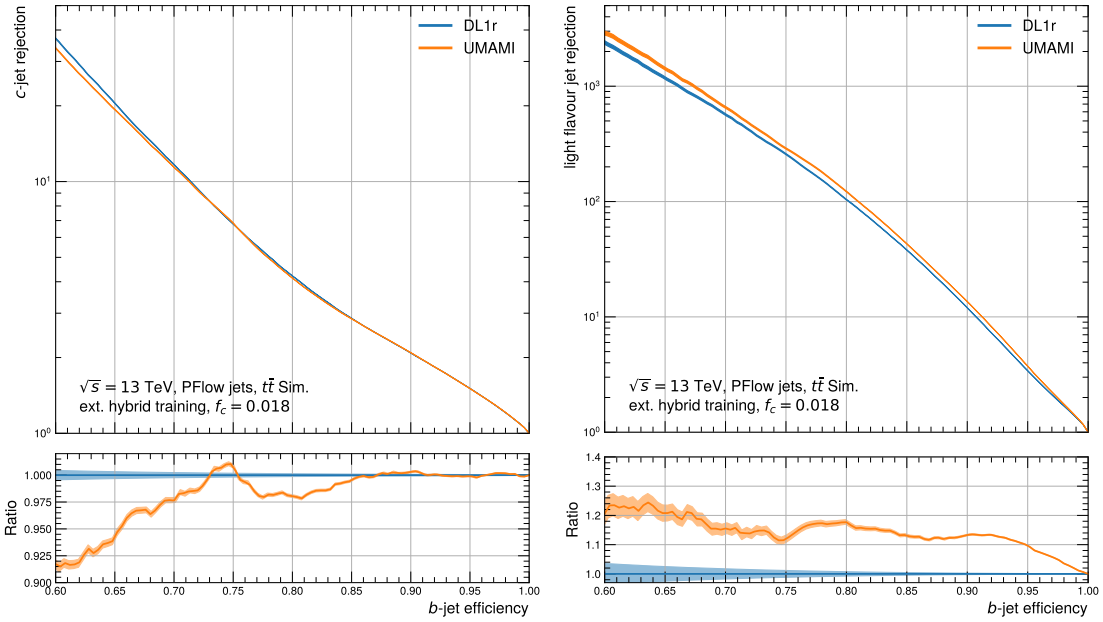
Figure 12.6.: DIPS (orange) performance compared to RNNIP (blue) in terms of (a) the c-jet rejection and (b) the light-flavour jet rejection as function of the b-jet efficiency.

12.1.4 *Possible Improvements and Outlook*

This first study creating a combined architecture of DIPS and DL1 shows already promising results. Nevertheless, there are lots of possibilities to improve and understand the tagger better avoiding overfitting.

The next steps would be to make use of the optimisations described in Ref. [137] by loosening the track selection and including the impact parameters d_0 and $z_0 \sin\theta$ in the training. Furthermore, extensive optimisation studies need to be done on the network itself: the loss function from Equation (12.2) needs to be optimised, in particular the λ parameter. The number of hidden layers and their number of nodes requires some tuning. In addition, attention techniques are worth to be tested. They allow to adapt the importance of the different tracks in the jet by assigning them a weight depending on a certain feature e.g. the impact parameter or p_T. Another possibility is to test different pooling operations besides the sum operation, e.g. the average or the maximum.

In conclusion, this new architecture is most probably the next step in flavour tagging in ATLAS and requires still lots of R&D work which go beyond the scope of this thesis.



(a) (b)

Figure 12.7.: UMAMI (orange) performance compared to *DL1r* (blue) in terms of (a) the c-jet rejection and (b) the light-flavour jet rejection as function of the b-jet efficiency.

Especially at large transverse momenta, further improvements are necessary to achieve a better performance. In that regime methods based on the hit information in the Inner Detector are studied within ATLAS. A charged particle can be identified as hits in the ID and in case a b-hadron decays between two ID layers, an increase in hits is observed. This is an ideal use case for so-called *Graph Neural Networks* [199]. Not only in the tagger algorithm development itself, but also in the calibration, especially in the light-flavour jet calibration, NN techniques would bring a gain. It would be for instance possible to employ *Invertible Neural Networks* [200] to allow a better calibration with the flipped taggers.

Part IV

ASSOCIATED PRODUCTION OF A HIGGS BOSON AND A TOP-QUARK PAIR WITH $H \to b\bar{b}$ DECAY

13

ANALYSIS OVERVIEW

Since the Yukawa coupling of a fermion is directly proportional to its mass, the top quark as heaviest elementary particle has therefore also the largest Yukawa coupling. The measurement of the $t\bar{t}H$ process allows a direct measurement of this coupling. After the first RUN II $t\bar{t}H(H \to b\bar{b})$ analysis from ATLAS [13] was published on a subset of the dataset of $36\,\text{fb}^{-1}$ (see sec. 5.1.2), the analysis presented in this thesis is performed with the full RUN II proton-proton collision dataset of $139\,\text{fb}^{-1}$ at the centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ [5]. A measurement of the Simplified Template Cross-Section (STXS) as a function of the Higgs p_T is performed for the first time using $t\bar{t}H(b\bar{b})$ events.

This chapter describes the $t\bar{t}H(b\bar{b})$ analysis overview giving insights in its motivation, challenges and strategy. After a short analysis summary in Section 13.1, the event selection is explained in Section 13.2 followed by a description of the modelling of the signal and background processes in Section 13.3. The analysis strategy is presented in Section 13.4 and the profile likelihood fit is introduced in Section 13.5. In the last Section 13.6 a summary of the systematic uncertainties is given.

13.1 THE ANALYSIS IN A NUTSHELL

As was shown in the pie-chart in Figure 2.7, the $t\bar{t}$ pair can either decay fully hadronically, semi-leptonically (lepton+jets) or dileptonically. In the scope of this thesis the lepton+jets channel[1] was studied where one W-boson decays leptonically and the other one hadronically as shown in the Feynman diagram in Figure 13.1. This channel offers a large statistics and has a relatively clean topology with the lepton in the final state allowing to suppress the multijet background. Since only one neutrino is present in the final state, the kinematics of the event can be fully determined taking into account the missing transverse momentum E_T^{miss}. Due to the fairly high branching ratio, also higher p_T regimes have sufficient statistics (compared to the dilepton channel) which is important for the STXS measurements. The lepton+jets channel is further split into a resolved (lower p_T) regime and a boosted regime. While this thesis focuses on the analysis of the resolved lepton+jets channel, the dilepton and the boosted channel were optimised separately. Nevertheless, all three channels are combined in a joint likelihood fit for the final results presented in this thesis.

The detector signature of the lepton+jets channel includes exactly one isolated lepton. For this analysis only decays into electrons and muons are considered, thus the term lepton ℓ is exclusively used for electrons and muons in the following. Nonetheless, the leptonic decay of the taus into elec-

1 The lepton+jets channel will be also denoted as *single lepton channel*.

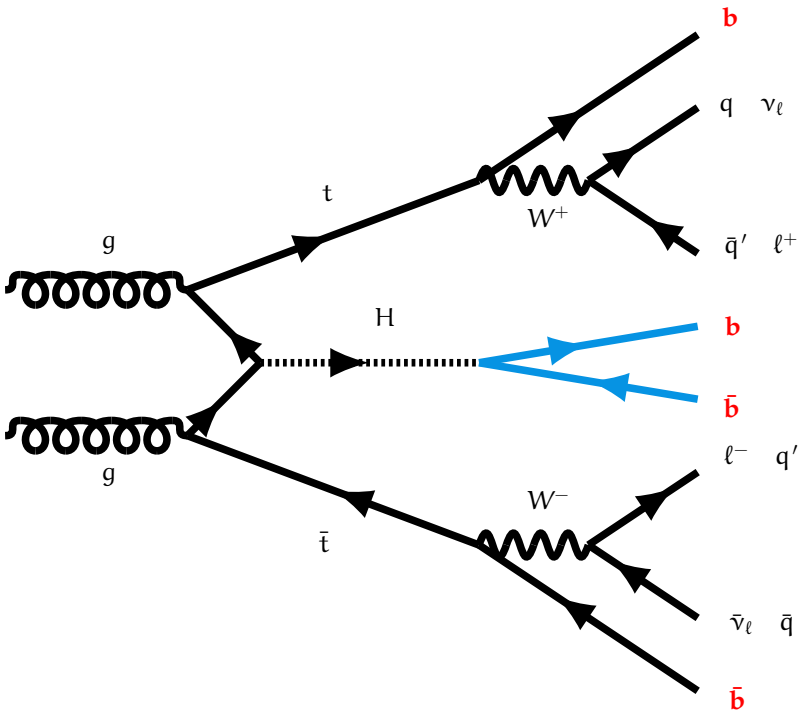


Figure 13.1.: Feynman diagram of the Higgs production with associated top quarks. The Higgs is decaying into a pair of b-quarks. The final state of this process contains at least four b-jets and exactly one lepton from the W decays.

trons and muons is also considered. In addition, six quarks and therefore six jets are present in the final state where at least four of them are b-jets. Consequently, b-tagging is crucial for this analysis and the analysis would heavily benefit from the improvements shown in Part III. However, this analysis was still performed with *EMTopo* jets and therefore does not include these improvements. Nevertheless, a short perspective will be given in Chapter 15 with *PFlow* jets and the new b-tagging improvements. The complex final state topology of $t\bar{t}H(b\bar{b})$ poses great challenges, especially the overwhelming $t\bar{t}+$jets background. In particular, the main irreducible background is coming from $t\bar{t}+b\bar{b}$ production for which an example Feynman diagram is shown in Figure 13.2. This process is poorly constrained by data measurements and has large theory uncertainties which limit the analysis.

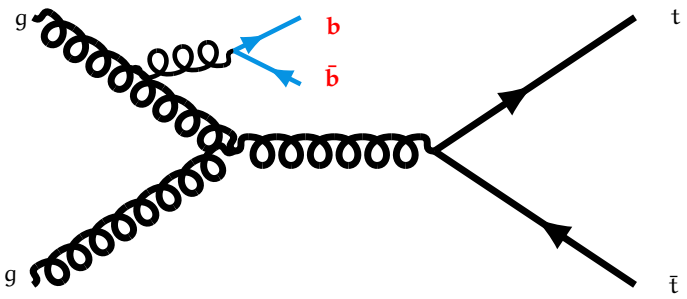


Figure 13.2.: Example Feynman diagram showing the $t\bar{t}$ production associated with a gluon initiated $b\bar{b}$ pair.

A schematic overview of the analysis strategy is illustrated in Figure 13.3. The first step is the event selection (see sec. 13.2), where a first phase space is chosen enhancing the $t\bar{t}H(b\bar{b})$ signal con-

tribution. In the next step, different multivariate techniques are employed which are based on Boosted Decision Trees (BDTs) for the resolved lepton+jets channel. The reconstruction BDT is matching the jets to the final state partons from the top-quark and Higgs decays. The preselected events are split into signal-depleted categories (control region, CR) and signal-enriched categories (signal region, SR) optimised for the STXS measurement. In the signal regions, an additional BDT is used to separate signal and background processes. All regions are then used in a combined profile-likelihood fit taking into account the systematic uncertainties and also including the boosted lepton+jets and dilepton channel.

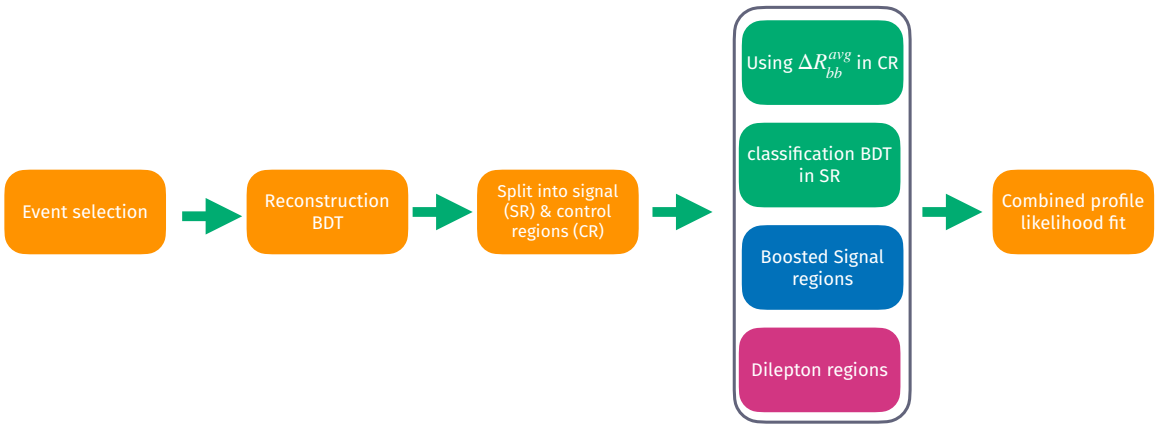


Figure 13.3.: Sketch of the $t\bar{t}H(b\bar{b})$ analysis strategy for the resolved lepton+jets channel. The profile likelihood fit is performed in a combination with the lepton+jets boosted and the dilepton channel.

13.2 EVENT SELECTION

In this section, the selection of the events used in the resolved lepton+jets channel is introduced. The dedicated selections for the boosted lepton+jets and the dilepton channels can be found in Ref. [5] and are orthogonal to the lepton+jets criteria.

The events for this analysis are extracted from the dataset recorded with the ATLAS experiment during the LHC RUN II which corresponds to the proton-proton collision data-taking periods from 2015–2018 with a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ and an integrated luminosity of $(139.0 \pm 2.4)\,\text{fb}^{-1}$ [201].
The physics objects used in this analysis are described in Chapter 6.
Single-lepton triggers are used to record all events for this analysis which show a high efficiency above their trigger threshold. The events have to fulfil a lepton isolation criterion and pass the low p_T trigger threshold for electrons corresponding to 24 (26) GeV in the data taking period 2015 (2016-2018) and for muons 20 (26) GeV. Alternatively, events with a looser identification criterion or even without any isolation criteria are accepted if they satisfy a higher trigger threshold [202]. A summary of the different trigger settings is shown in Table 13.1. Furthermore, at least one primary vertex has to be present in each event.

lepton	p_T threshold		identification requirement		isolation criterion	
	2015	2016–2018	2015	2016–2018	2015	2016–2018
electrons	24 GeV	26 GeV	medium	tight	–	loose
	60 GeV	60 GeV	medium	medium	–	–
	120 GeV	140 GeV	loose	loose	–	–
muons	20 GeV	26 GeV	medium	medium	loose	medium
	50 GeV	50 GeV	medium	medium	–	–

Table 13.1.: Single-lepton trigger settings used for this analysis shown for the data taking periods 2015 and 2016–2018 separately.

EMTopo jets are used as described in Section 6.2.1 with a radius parameter $R = 0.4$ (small-R jets). To reduce pile-up effects, the *Medium* WP of the jet vertex tagger (JVT) is applied [203]. In addition, jets with radius $R = 1.0$ reclustered from small-R jets [204] are employed to veto boosted events which allows to ensure orthogonality between the resolved and boosted lepton+jets channel. Events which pass the boosted selection are removed from the resolved lepton+jets channel.
b-jets are identified using the high-level b-tagging algorithm MV2c10 described in Section 8.5.2, making use of the single-cut WPs and pseudo-continuous b-tagging introduced in Section 8.5.1. The correction factors to data are retrieved as described in Section 8.6 for the b-jet efficiency [179], c-jet [183] and light-flavour jet [182] mis-tag rate separately.

Electrons (see sec. 6.2.2) and muons (see sec. 6.2.3) have to fulfil $p_T > 10\,\text{GeV}$ and the *Medium* and *Loose* identification operation point, respectively.

In order to avoid double-counting of jets or leptons, a so-called *overlap removal* is performed. If a jet is within $\Delta R_y = 0.2$[2] of an electron candidate, the closest jet is removed. An electron is rejected, in case there is still a jet within $\Delta R_y = 0.4$ of the electron. Muons are directly discarded if a jet is within $\Delta R_y < 0.4$ of the muon. This suppresses muons coming from heavy-flavour decays inside jets. If this jet has less than three tracks however, the jet is discarded instead and the muon is kept.

At least five jets are required to be present in the event, where not less than four jets have to be tagged with the 70% b-tagging WP. In addition, exactly one lepton with $p_T > 27\,\text{GeV}$ needs to be in the event satisfying the *Tight* identification criterion for electrons and the *Medium* criterion for muons. Furthermore, no additional lepton with $p_T > 10\,\text{GeV}$ fulfilling the *Medium* and *Loose* identification operation point for electrons and muons, respectively, is allowed. Moreover, events including more than two hadronic τ candidates are removed to ensure orthogonality to other $t\bar{t}H$ channels.

13.3 MODELLING OF SIGNAL AND BACKGROUND PROCESSES

The physics analyses which are searching for new processes are typically performed in a blinded way. This means that the analysis is optimised without looking at the data in regions sensitive to the $t\bar{t}H(b\bar{b})$ signal to avoid introducing a bias. Therefore the optimisation is performed on simulated signal and background samples. Even though the $t\bar{t}H$ production and the $H \to b\bar{b}$ decay channel are already independently discovered [11, 12], the $t\bar{t}H(b\bar{b})$ channel has not yet been observed and it is thus a physics search which is done in a blinded way[3].

To simulate the detector response both the full detector simulation GEANT4 and the fast simulation *AtlFast-II* are utilised (see sec. 4.2). The pile-up interactions are simulated using PYTHIA8.186 [84] (A3 tune [205]) and all events are reweighted to the respective pile-up profiles observed in data during RUN II corresponding to 34 proton-proton interactions per bunch-crossing on average. Generally, the generator settings described in Section 4.1.3 also apply for the modelling of the signal and background processes used for the $t\bar{t}H(b\bar{b})$ analysis.

All simulated samples which are used in this analysis are summarised in Table 13.2. The table shows both the nominal samples which are used for the baseline modelling and the alternative samples which are used for sanity checks and to estimate systematic uncertainties.

If not differently stated the ME generator is used at NLO precision in QCD together with the PDF set `NNPDF3.0NLO` [77] for five-flavour scheme (5FS) samples and the PDF set `NNPDF3.0NLOnf4` [77] for the four-flavour scheme (4FS) samples.

2 For ΔR_y the rapidity instead of the pseudorapidity is used for its calculation.

3 For the optimisation of the analysis, the signal contribution in every bin considered in the analysis has to be $< 7.7\%$, which is the blinding threshold.

Process	ME generator	ME PDF	PS	Normalisation
Higgs boson				
$t\bar{t}H$	PowhegBox v2	NNPDF3.0NLO	Pythia8.230	NLO+NLO (EW) [107]
	PowhegBox v2	NNPDF3.0NLO	Herwig7.04	NLO+NLO (EW) [107]
	MadGraph5_aMC@NLO v2.6.0	NNPDF3.0NLO	Pythia8.230	NLO+NLO (EW) [107]
tHjb	MadGraph5_aMC@NLO v2.6.2	NNPDF3.0NLOnf4	Pythia8.230	–
tWH	MadGraph5_aMC@NLO v2.6.2 [DR]	NNPDF3.0NLO	Pythia8.235	–
$t\bar{t}$ and single-top				
$t\bar{t}$	PowhegBox v2	NNPDF3.0NLO	Pythia8.230	NNLO+NNLL [206–212]
	PowhegBox v2	NNPDF3.0NLO	Herwig7.04	NNLO+NNLL [206–212]
	MadGraph5_aMC@NLO v2.6.0	NNPDF3.0NLO	Pythia8.230	NNLO+NNLL [206–212]
$t\bar{t}+b\bar{b}$	PowhegBoxRes	NNPDF3.0NLOnf4	Pythia8.230	–
	Sherpa v2.2.1	NNPDF3.0NNLOnf4	Sherpa	–
tW	PowhegBox v2 [DR] [213–215]	NNPDF3.0NLO	Pythia8.230	NLO+NNLL [216, 217]
	PowhegBox v2 [DS] [213–215]	NNPDF3.0NLO	Pythia8.230	NLO+NNLL [216, 217]
	PowhegBox v2 [DR] [213–215]	NNPDF3.0NLO	Herwig7.04	NLO+NNLL [216, 217]
	MadGraph5_aMC@NLO v2.6.2 [DR]	CT10NLO	Pythia8.230	NLO+NNLL [216, 217]
t-channel	PowhegBox v2 [213–215]	NNPDF3.0NLOnf4	Pythia8.230	NLO [218, 219]
	PowhegBox v2 [213–215]	NNPDF3.0NLOnf4	Herwig7.04	NLO [218, 219]
	MadGraph5_aMC@NLO v2.6.2	NNPDF3.0NLOnf4	Pythia8.230	NLO [218, 219]
s-channel	PowhegBox v2 [213–215]	NNPDF3.0NLO	Pythia8.230	NLO [218, 219]
	PowhegBox v2 [213–215]	NNPDF3.0NLO	Herwig7.04	NLO [218, 219]
	MadGraph5_aMC@NLO v2.6.2	NNPDF3.0NLO	Pythia8.230	NLO [218, 219]
Other				
W+jets	Sherpa v2.2.1 (NLO [2j], LO [4j]) [87, 220]	NNPDF3.0NNLO	Sherpa [221–225]	NNLO [226]
Z+jets	Sherpa v2.2.1 (NLO [2j], LO [4j]) [87, 220]	NNPDF3.0NNLO	Sherpa [221–225]	NNLO [226]
VV (had.)	Sherpa v2.2.1	NNPDF3.0NNLO	Sherpa [220, 221]	–
VV (lep.)	Sherpa v2.2.2	NNPDF3.0NNLO	Sherpa [220, 221]	–
VV (lep.) + jj	Sherpa v2.2.2 (LO [EW])	NNPDF3.0NNLO	Sherpa [220, 221]	–
$t\bar{t}W$	MadGraph5_aMC@NLO v2.3.3	NNPDF3.0NLO	Pythia8.210	NLO+NLO (EW) [107]
	Sherpa v2.0.0 (LO [2j])	NNPDF3.0NNLO	Sherpa	NLO+NLO (EW) [107]
$t\bar{t}\ell\ell$	MadGraph5_aMC@NLO v2.3.3	NNPDF3.0NLO	Pythia8.210	NLO+NLO (EW) [107]
	Sherpa v2.0.0 (LO [1j])	NNPDF3.0NNLO	Sherpa	NLO+NLO (EW) [107]
$t\bar{t}Z$ ($qq, \nu\nu$)	MadGraph5_aMC@NLO v2.3.3	NNPDF3.0NLO	Pythia8.210	NLO+NLO (EW) [107]
	Sherpa v2.0.0 (LO [2j])	NNPDF3.0NNLO	Sherpa	NLO+NLO (EW) [107]
$t\bar{t}t\bar{t}$	MadGraph5_aMC@NLO v2.3.3	NNPDF3.1NLO	Pythia8.230	NLO+NLO (EW) [227]
tZq	MadGraph5_aMC@NLO v2.3.3 (LO)	CTEQ6L1 [228]	Pythia8.212	–
tWZ	MadGraph5_aMC@NLO v2.3.3 [DR]	NNPDF3.0NLO	Pythia8.230	–

Table 13.2.: Overview of all simulated MC samples used in this analysis. The nominal sample is always the first row for each process. If not differently stated the ME generator is at NLO precision in QCD. The abbreviations [DR] and [DS] stand for the diagram removal scheme [229] and the diagram subtraction scheme [170, 229], respectively. The higher-order cross-section used to normalise these samples is listed in the last column and refers to the order of QCD processes if no additional information is provided. If no information is present in this column, there is no higher-order k-factor applied to this process. The table is taken from [5].

13.3.1 $t\bar{t}H$ *Signal Modelling*

The signal of this analysis is the associated production of the Higgs boson with a $t\bar{t}$ pair ($t\bar{t}H$) which is modelled in the 5FS with the generator PowhegBox [88–92] and the PS and hadronisation is simulated with Pythia8. For the event generation, both the renormalisation and hadronisation scale are set to $\mu_R = \mu_F = \sqrt[3]{m_T(t) \cdot m_T(\bar{t}) \cdot m_T(H)}$, where m_T is the transverse mass of a particle defined as $m_T = \sqrt{m^2 + p_T^2}$ and the h_{damp} parameter (see sec. 4.1) is fixed to $h_{damp} = 3/4 \cdot (m_t + m_{\bar{t}} + m_H) = 352.5$ GeV. In general, all Higgs boson decay modes are taken into account, however the analysis is optimised for the $H \to b\bar{b}$ decay and only small fractions of other decay modes are present in the final selection (at maximum 6% in some regions

of the resolved lepton+jets channel). All $t\bar{t}H$ samples are normalised to the $t\bar{t}H$ cross-section of $\sigma_{t\bar{t}H} = (507 \pm 50)$ fb [107] which is determined at NLO accuracy in QCD and incorporating NLO electro-weak corrections.

13.3.2 $t\bar{t}$ + jets *Background*

The dominant background in this analysis is coming from the $t\bar{t}$ pair production in association with additional jets. Depending on the flavour of the additional jets, these events are categorised accordingly [230]. In simulated events, the labelling of the additional jets is done in a slightly different way than for the flavour-tagging algorithms described in Section 8.1.1. So-called *particle jets* are formed with the anti-k_t algorithm with the radius parameter $R = 0.4$, only taking into account particles with a mean lifetime $\tau > 3 \cdot 10^{-11}$ s which do not originate from the top-quark or W-boson decays. Then, a ΔR matching is performed as for flavour tagging with $\Delta R(\text{jet}, \text{hadron}) < 0.4$ associating the flavour label to the jet. The different components are listed in Table 13.3 with the main categories: $t\bar{t} + \geqslant 1b$, $t\bar{t} + \geqslant 1c$ and $t\bar{t}$ + light.

$t\bar{t}$ + jets category	description
	a $t\bar{t}$ pair and
$t\bar{t} + \geqslant 1b$	at least one additional jet containing at least one b-hadron
$t\bar{t} + 1b$	exactly one additional jet containing exactly one b-hadron
$t\bar{t} + 1B$	exactly one additional jet containing at least two b-hadrons
$t\bar{t} + \geqslant 2b$	at least two additional jets containing at least one b-hadron each
$t\bar{t} + \geqslant 1c$	at least one additional jet matched to at least one c-hadron
$t\bar{t}$ + light	all the other cases excluding the above ones

Table 13.3.: Overview of the different $t\bar{t}$ + jets components. The inclusive $t\bar{t} + \geqslant 1b$ category is again split into three exclusive categories which will be used to estimate systematic uncertainties.

The main background $t\bar{t} + b\bar{b}$ is modelled separately in the 4FS (described below), in contrast to the $t\bar{t} + \geqslant 1c$ and $t\bar{t}$ + light events. The modelling of the two latter categories is performed in the 5FS. Additionally, $t\bar{t} + \geqslant 1b$ events are also simulated in the 5FS but they are only used for studies and to define a subset of the modelling systematics. To generate the 5FS events POWHEGBOX v2 is used with the renormalisation and factorisation scale set to $\mu_R = \mu_F = m_T(\text{top})$ and $h_{\text{damp}} = 1.5 \cdot m_{\text{top}}$. The PS and hadronisation processes are simulated using PYTHIA8.

$t\bar{t} + b\bar{b}$ *Background*

As indicated above, the irreducible $t\bar{t} + b\bar{b}$ background is the dominant background and the main challenge of this analysis. It is modelled in the 4FS using as generator POWHEGBOXRES [231, 232] and OPENLOOPS [233, 234] with the factorisation scale $\mu_F = 1/2 \left(\sum_{i=t,\bar{t},b,\bar{b}} m_T(i) + \sum_j p_T(j) \right)$ where j represents any additional partons. The renormalisation scale is set to

$\mu_R = \sqrt[4]{m_T(t) \cdot m_T(\bar{t}) \cdot m_T(b) \cdot m_T(\bar{b})}$ together with $h_{damp} = 1/2 \Sigma_{i=t,\bar{t},b,\bar{b}} m_T(i)$. For all involved processes the mass of the b-quarks is set to $m_b = 4.95\,\text{GeV}$.

The $t\bar{t} + \geqslant 1b$ process can be further split into three subcategories: $t\bar{t} + 1b$, $t\bar{t} + 1B$ and $t\bar{t} + \geqslant 2b$ (see Table 13.3) which will be used for a systematic uncertainty estimation in Section 13.6.2. Their relative fractions are shown in Figure 13.4 for the nominal modelling in the 4FS and the alternative 5FS setup both described above as well as for POWHEGBOX+HERWIG7.

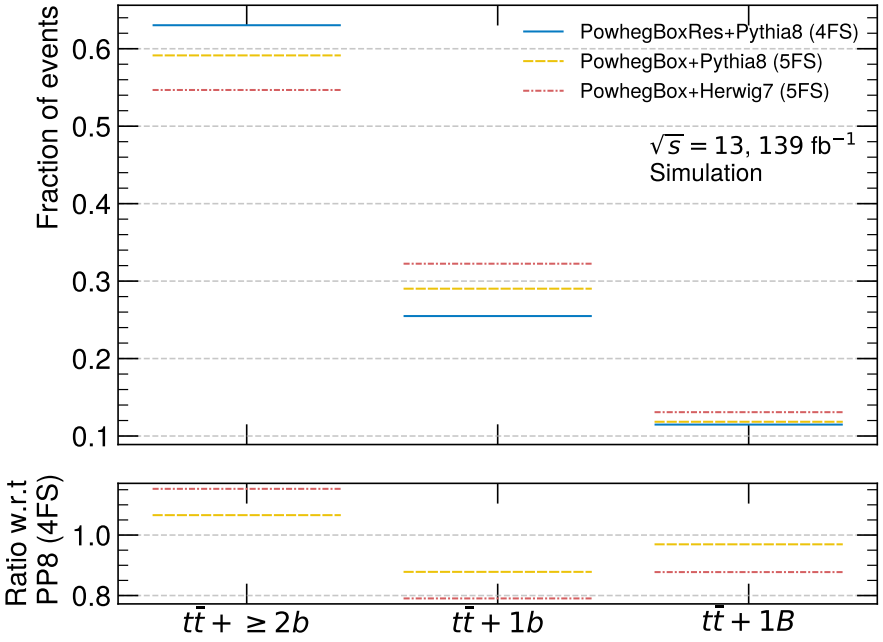


Figure 13.4.: Relative fraction of the $t\bar{t} + \geqslant 1b$ subcomponents: $t\bar{t} + 1b$, $t\bar{t} + 1B$ and $t\bar{t} + \geqslant 2b$ shown for the nominal 4FS (blue) and the two alternative predictions: POWHEGBOX+PYTHIA8 5FS (orange) and POWHEGBOX+HERWIG7 5FS (red).

13.3.3 *Other Processes*

Besides the main backgrounds coming from $t\bar{t}$+jets events described above, there are also other contributing background processes. The $t\bar{t}V$ background which is the production of a vector boson (W, Z) in association with a $t\bar{t}$ pair is simulated at NLO in QCD with the MADGRAPH5_aMC@NLO v2.3.3 generator with $\mu_R = \mu_F = 0.5 \times \sum_i m_T(i)$. All the other backgrounds are listed in Table 13.2 with their generator settings and described in more detail in [5]. In the following, these background processes will be summarised as *Other* since their single contributions are very small.

13.3.4 *Fake Leptons*

Fake leptons are photons or jets which are wrongly reconstructed as leptons. They are mostly originating from multijet processes where one jet is misidentified as electron or muon. The isolation requirements demanded on trigger and event selection level in combination with the required lepton

quality criteria eliminate most fake leptons in the lepton+jets channel. The contribution of the fake background is negligible in the lepton+jets channel and small in the dilepton channel where it is estimated from MC simulation.

13.3.5 *Inclusive Modelling*

The modelling in data is shown in Figure 13.5 for the inclusive resolved lepton+jets selection described in Section 13.2. The illustrated distributions are the $\Delta R_{b\bar{b}}^{avg}$ variable, which is the average ΔR between all possible b-tagged jet (70% WP) pairs in an event, and the number of jets per event as well as the number of jets passing the 70% and 60% b-tagging working points. Both the number of jets (Fig. b) and number of b-tagged jets at the 60% WP (Fig. d) show a clear slope in the data over MC ratio, not fully covered by the uncertainties. The number of b-tagged jets at the 70% WP (Fig. c) has a normalisation offset, a slope is not visible due to only having two bins. Similarly, the $\Delta R_{b\bar{b}}^{avg}$ distribution shows a constant normalisation offset. This offset was already seen before in the previous analysis [13] corresponding to a normalisation factor of the $t\bar{t} + \geqslant 1b$ background of 1.25. Since this background is dominating in this phase space, it matches the offset seen also here.

13.4 ANALYSIS STRATEGY

This analysis is mainly targeting the STXS measurement in bins of p_T^H (see sec. 5.2) but also the inclusive cross-section measurement of the $t\bar{t}H$ signal. The events passing the lepton+jets selection described in Section 13.2 are further divided into two types of analysis regions: signal regions (SRs) and control regions (CRs). All regions are defined to be orthogonal (disjoint) to each other. Furthermore, several multivariate techniques are employed to reconstruct the Higgs candidate and to classify the $t\bar{t}H$ signal outlined in Section 13.4.2.

13.4.1 *Region Definition*

The analysis regions in the lepton+jets channel are categorised as a function of the number of jets per event into SRs and CRs. The analysis region definitions are listed in Table 13.4. The two CRs contain events with exactly five jets and they are split according to the high and low b-tagging criteria, respectively. The $CR^{5j}_{\geqslant 4b\ hi}$ requires not less than four b-tagged jets to pass the 60% WP while in the $CR^{5j}_{\geqslant 4b\ lo}$ at least one of the b-tagged jets (with the 70% WP) is not allowed to meet the 60% WP criterion ensuring orthogonality. For the signal regions, events are selected containing six or more jets and they are further subdivided into five reconstructed p_T^H categories: 0–120 GeV, 120–200 GeV, 200–300 GeV, 300–450 GeV and $\geqslant 450$ GeV. The p_T^H is reconstructed using the information of the reconstruction BDT as described in Section 13.4.2. These p_T^H categories correspond also to the truth Higgs transverse momentum bins used for the STXS measurement, where the truth p_T^H is only defined on simulation level as the p_T of the truth Higgs object. Figure 13.6 shows the background

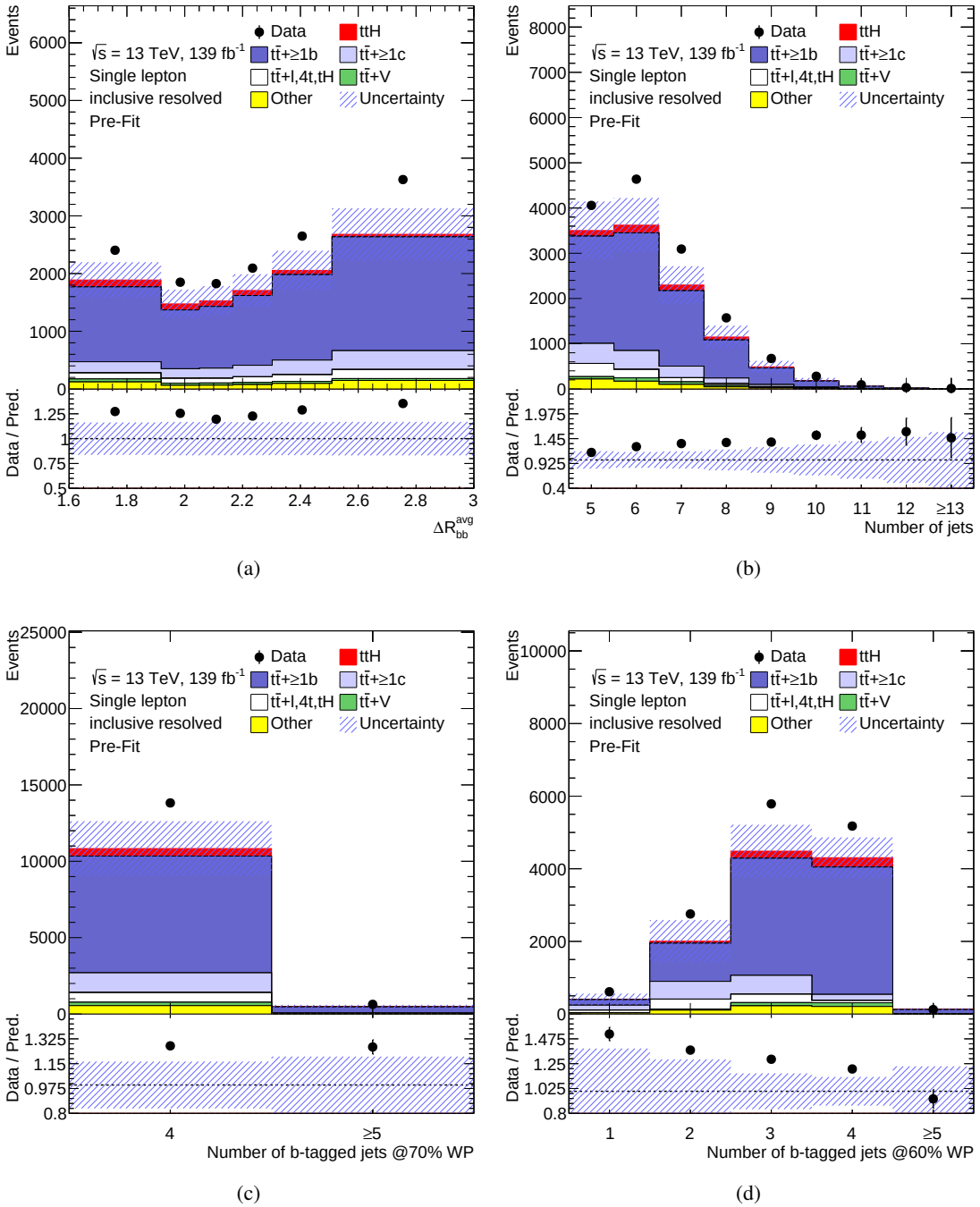


Figure 13.5.: Inclusive lepton+jets distributions showing the (a) ΔR_{bb}^{avg} variable, (b) the number of jets per event, the number of jets passing the (c) 70% and (d) 60% b-tagging WP per event. The uncertainty band contains the systematic and statistical uncertainties described in Section 13.6 except the normalisation of the $t\bar{t} + \geqslant 1b$ background which is only defined after the likelihood fit.

Region	$SR^{\geqslant 6j}_{\geqslant 4b}$					$CR^{5j}_{\geqslant 4b\ hi}$	$CR^{5j}_{\geqslant 4b\ lo}$
p_T^H/GeV	$[0,120)$	$[120,200)$	$[200,300)$	$[300,450)$	$\geqslant 450$	inclusive	
#leptons	$= 1$						
#jets	$\geqslant 6$					$= 5$	
#b-tag @70%	$\geqslant 4$						
#b-tag @60%	–					$\geqslant 4$	< 4
Fit input	classification BDT				Yield	ΔR^{avg}_{bb}	

Table 13.4.: Analysis region definition of the lepton+jets channel split into signal regions with $\geqslant 6$ jets subdivided into five p_T^H bins and control regions with five jets categorised into two regions as a function of the b-tagging WP. In the last row, the variable which is used as fit input is listed.

contributions in each analysis region. The $t\bar{t}+$jets production dominates the regions by far and only smaller fractions are coming from $t\bar{t}V$ or from *Other* processes. The largest fraction of $t\bar{t}+$jets consists of the $t\bar{t}+\geqslant 1b$ background which corresponds to more than 70% of the total background in all SRs followed by the $t\bar{t}+\geqslant 1c$ process which is about 10%. The region $CR^{5j}_{\geqslant 4b\ hi}$ shows an increased $t\bar{t}+\geqslant 1b$ fraction ($\sim$ 84%) while $CR^{5j}_{\geqslant 4b\ lo}$ is enhanced in $t\bar{t}+\geqslant 1c$ and $t\bar{t}+$light compared to the other regions which allows the fit to better extract information and constraints for these processes. The signal contribution in the different lepton+jets regions are illustrated in Figure 13.7. The black solid line shows the signal (S) over background (B) ratio and the red dashed line the $S/\sqrt{B}$ ratio. It is clearly visible that the values for $S/\sqrt{B}$ decrease with p_T^H in the SRs. Especially the two SRs with the highest p_T^H values show lower values which is also due to the boosted veto[4] (the boosted regions are only defined for $p_T^H > 300\,\text{GeV}$).

The dilepton and the boosted lepton+jets channel follow a similar strategy defining the analysis regions. The flow chart including the different selections is shown for all three channels in Figure B.1. In total, 16 regions (11 SRs and five CRs) are defined which will be used in the statistical analysis described in Chapter 14.

13.4.2 *Multivariate Techniques*

Multivariate techniques allow to better reconstruct and classify physics topologies. As demonstrated in Part III, in flavour tagging a heavy use of Neural Networks is made. In this analysis, BDTs are used to reconstruct the Higgs candidate and to classify between the $t\bar{t}H$ signal and background processes. BDTs are introduced in Section 7.4. In addition, a likelihood discriminant method is used to separate

4 The boosted veto makes sure that the events used in the boosted signal regions are not utilised in the resolved regions to maintain orthogonality.

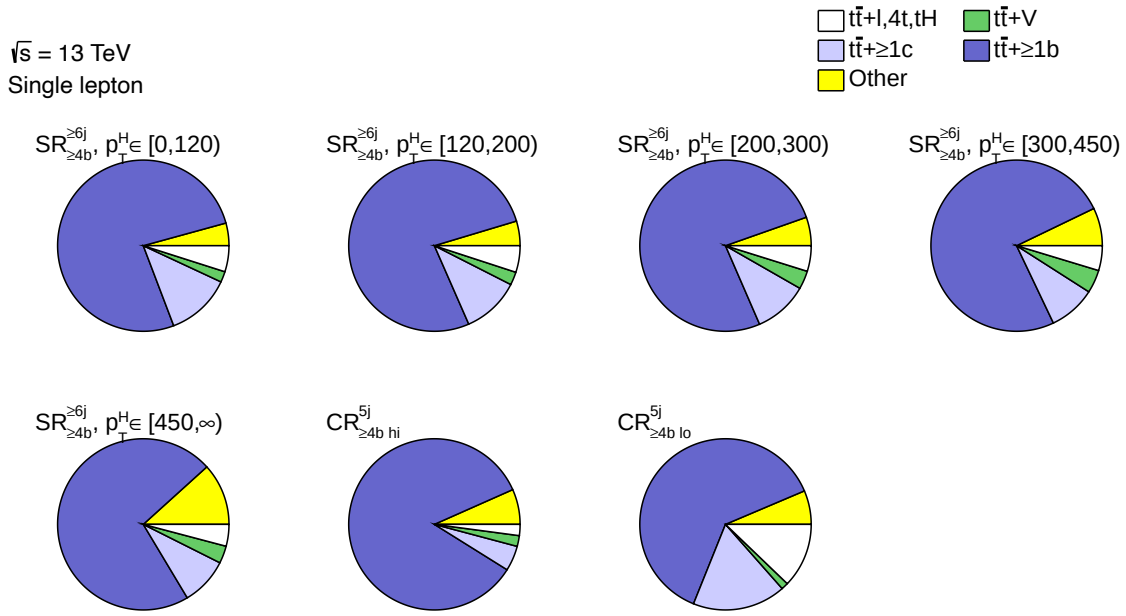


Figure 13.6.: Pie charts showing the background composition in the resolved lepton+jets analysis regions. The $t\bar{t}H$ signal is excluded here.

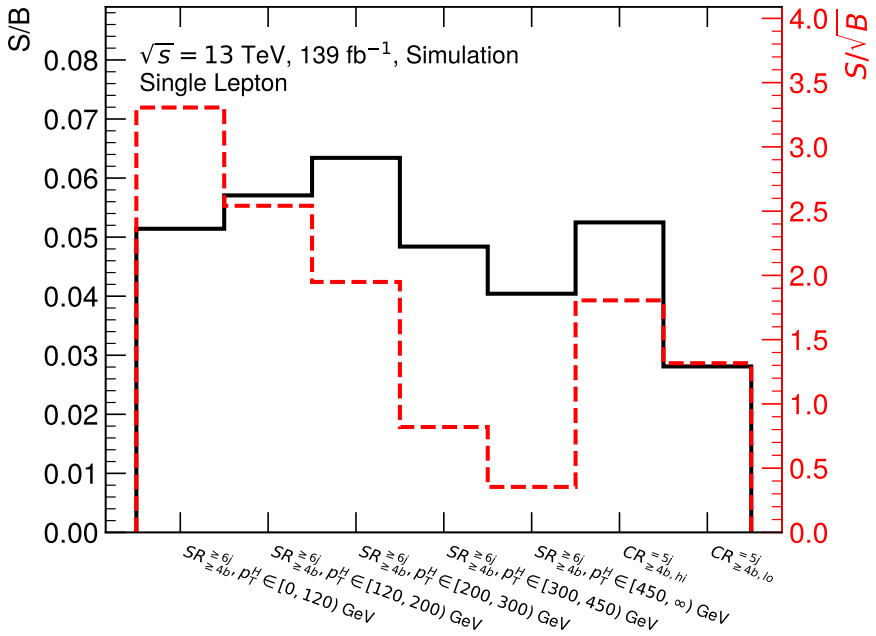


Figure 13.7.: Contribution of the $t\bar{t}H$ signal (S) in the different resolved lepton+jets analysis regions. The black solid line associated to the left vertical axis shows the signal over background (B) ratio and the red dashed line the $S/\sqrt{B}$ distribution corresponding to the right vertical axis.

the signal and background. All the described multivariate techniques are used in the signal regions in the resolved lepton+jets channel, the techniques used in the boosted lepton+jets and dilepton channel are described in [5].

Reconstruction BDT

To extract further information from the complex final state of the $t\bar{t}H$ process (see Fig. 13.1), the different partons in the final state coming from the Higgs and $t\bar{t}$ system need to be matched to the jets in an event. Each possible combination of leptons and jets is a permutation. For an event containing six jets, the possible permutations that would need to be tested would amount to 720, without making any assumptions. With the additional b-tagging information this is reduced to 48 permutations and with a p_T ordering[5] only 12 permutations need to be tested. For each permutation, several quantities are calculated such as the invariant mass of the object candidates and their angular distances. The permutation with the correct parton to jet assignment is used as *signal* class in the BDT training and all other permutations as *background*. In total 15 variables related to the topological information of the $t\bar{t}$ system and four related to the Higgs system are used in the BDT training (the entire variable list is shown in Table B.1). An additional training is performed excluding the topological Higgs information to avoid a bias on background processes. Depending on the use-case either of these two reconstruction BDT versions are used. The training is performed inclusively on all $t\bar{t}H$ signal events modelled with the MADGRAPH5_aMC@NLO v2.3.2 generator in the resolved lepton+jets channel with at least 6 jets where at least 4 are b-tagged at the 85% WP.
The permutation with the largest BDT score is then selected for the event reconstruction. This allows to reconstruct the kinematics of the Higgs boson candidate. The efficiency to correctly reconstruct the Higgs boson candidate is 43% for all signal events passing the lepton+jets selection with at least 6 jets. In the different STXS bins, the Higgs boson reconstruction efficiency is ranging from 35% in the lowest STXS bin up to 59% in the highest STXS bin summarised in Table 13.5. In addition, the performance is also shown in Figure 13.8 as the migration matrix indicating the purity of the truth Higgs p_T^H and the reconstructed p_T^H.

p_T^H [GeV]	$SR^{\geqslant 6j}_{\geqslant 4b}$
Inclusive	43%
$[0, 120)$	35%
$[120, 200)$	45%
$[200, 300)$	57%
$[300, 450)$	59%
$[450, \infty)$	

Table 13.5.: Efficiency of the Higgs boson candidate to be correctly reconstructed in a given STXS bin with the reconstruction BDT. The efficiency is calculated for all signal events which are selected in the resolved lepton+jets signal regions [5].

5 Here, without the loss of generality, it is assumed that one of the two final partons of the Higgs and the hadronic W decay, respectively, carries the higher p_T. This reduces the combinatorics while the physics behind is not affected since the important information is the association of the jets to the Higgs boson and the W.

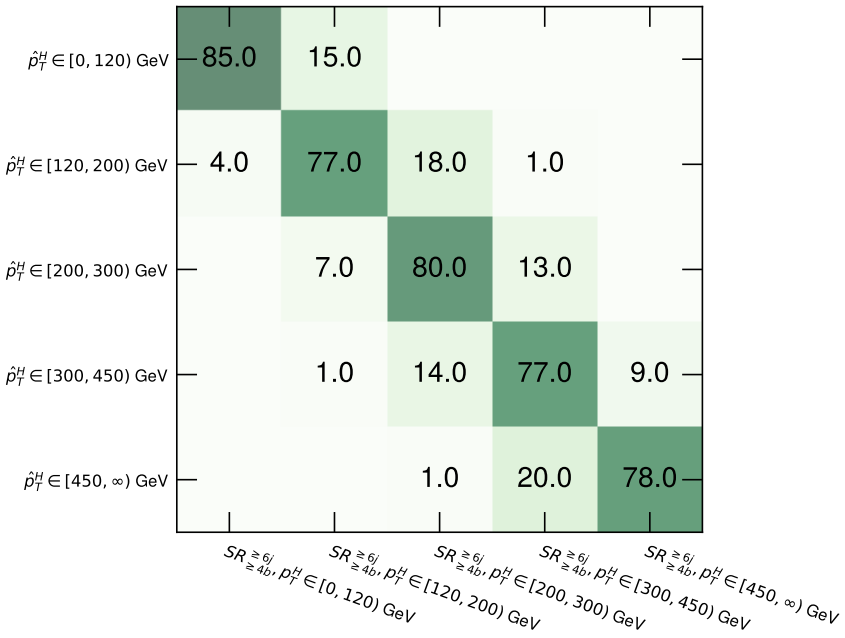


Figure 13.8.: Migration matrix between the $\hat{p}_T^H$ (truth p_T^H) on the y-axis and the reconstructed p_T^H on the x-axis in bins of the lepton+jets signal regions. In each row, the fraction of truth-matched Higgs boson candidates given in percentages is shown for the different reconstruction p_T^H bins and thus the values in the matrix indicate the purity in each bin.

Likelihood Discriminant

Various one-dimensional probability density functions of different variables, like the invariant masses and angular distributions from reconstructed objects are used to calculate the likelihood discriminant [13]. Two background hypotheses are considered for the $t\bar{t}+1b$ and $t\bar{t}+\geqslant 2b$ processes separately and averaged and weighted with their respective fractions in $t\bar{t}$ samples. The signal p^{sig} and background hypotheses p^{bkg} are retrieved from the product of the single variable probability density functions and averaged over all parton permutations, weighted by the b-tagging information. In the likelihood discriminant calculation, however no correlations are considered. As input for the classification BDT the ratio $p^{sig}/(p^{sig}+p^{bkg})$ is utilised per event. The likelihood discriminant is only used in the lepton+jets SR and shown in Figure 13.9.

Classification BDT

As mentioned above, in the lepton+jets SRs a classification BDT is trained to better separate the signal and background processes. The training is performed on the $t\bar{t}H$ signal and the dominant $t\bar{t}+$jets background on events with $\geqslant 6$ jets from which at least 4 have to be b-tagged with the 85% WP. This more inclusive phase space compared to the baseline SR definition provides more statistics that can be used in the training. The classification BDT combines in total 18 different discriminating input variables such as the likelihood discriminant and the ΔR_{bb}^{avg} variable, as shown in Table 13.6, which are the two highest-ranked variables in the training. In addition, the information provided

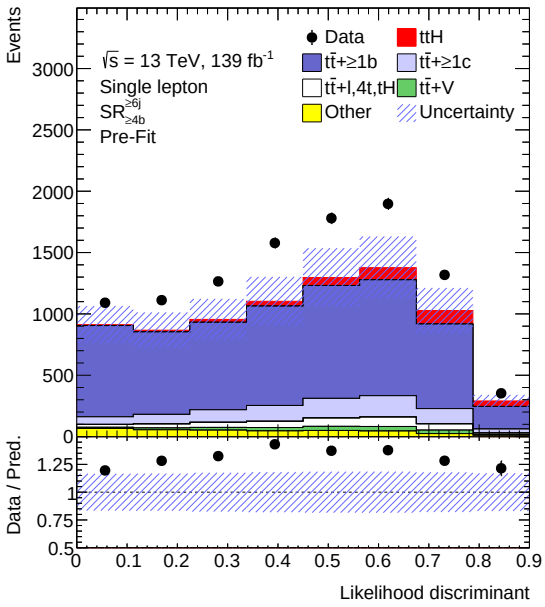


Figure 13.9.: Distribution of the likelihood discriminant in the lepton+jets SRs. All uncertainties including their correlations are considered in the uncertainty band excluding $k(t\bar{t} + \geqslant 1b)$ which is only defined after the fit.

by the reconstruction BDT is used, i.e. the reconstruction BDT score and the resulting kinematic variables of the Higgs and $t\bar{t}$ systems as well as angles between the reconstructed objects. Moreover, the information from the pseudo-continuous b-tagging is part of the classification BDT input. It was verified that all input variables have an adequate modelling. An overview of all input variables ordered by their importance for the training is shown in table 13.6. The importance of a variable is retrieved by evaluating how often it is used for a decision in the BDT. Figure 13.10 shows the BDT distribution for the $t\bar{t}H$ signal and the $t\bar{t} + \text{jets}$ background.

Ranking	Variable	Definition
1	LHD	Likelihood discriminant
2	$\Delta R_{b\,b}^{avg}$	Average ΔR for all b-tagged jet pairs
3	$m_{b\,b}^{min\ \Delta R}$	Mass of the combination of two b-tagged jets with the smallest ΔR
4	$\Delta R_{b\,b}^{max\ p_T}$	ΔR between the two b-tagged jets with the largest vector sum p_T
$5^{\dagger *}$	BDT output	Output of the reconstruction BDT
$6^{\dagger}$	$m_{b\,b}^{Higgs}$	Higgs candidate mass
$7^{\dagger}$	$\Delta R_{b\,b}^{Higgs}$	ΔR between b-jets from the Higgs candidate
$8^{\dagger *}$	$\Delta R_{H,t\bar{t}}$	ΔR between Higgs candidate and $t\bar{t}$ candidate system
$9^{\dagger}$	$m_{H,b_{lep\ top}}$	Mass of Higgs candidate and b-jet from leptonic top candidate
$10^{\dagger}$	$\Delta R_{H,lep\ top}$	ΔR between Higgs candidate and leptonic top candidate
11	$\Delta\eta_{jj}^{max}$	Maximum $\Delta\eta$ between any two jets
12	B_{jet}^{5}	5^{th} largest jet b-tagging discriminant
14	$N_{b\,b}^{Higgs\ 30}$	Number of b-tagged jet pairs with invariant mass within 30 GeV of the Higgs-boson mass
13	B_{jet}^{3}	3^{rd} largest jet b-tagging discriminant
15	B_{jet}^{4}	4^{th} largest jet b-tagging discriminant
17	Aplanarity	$1.5\lambda_2$, where λ_2 is the second eigenvalue of the momentum tensor [235] built with all jets
$16^{\dagger}$	$w_{b\text{-tag}}^{Higgs}$	Sum of b-tagging discriminants of jets from best Higgs candidate from the reconstruction BDT
18	H_1	Second Fox–Wolfram moment computed using all jets and the lepton

Table 13.6.: Variable importance ranking of the input variables to the classification BDT in the lepton+jets signal regions. The variables marked with a † are extracted from the information of the reconstruction BDTs and an additional star indicates that the topological information of the Higgs is used in the reconstruction BDT.

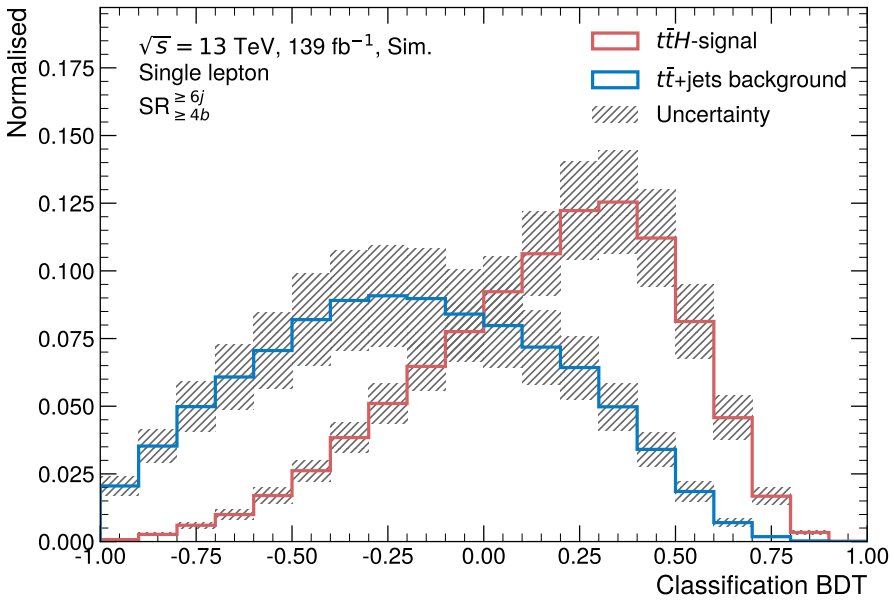


Figure 13.10.: Classification BDT in the lepton+jets SRs shown for the $t\bar{t}H$ signal (red) and the $t\bar{t}+$jets background (blue). The uncertainty band includes all uncertainties except for $k(t\bar{t}+\geqslant 1b)$ which is only defined after the fit. Both distributions are normalised to unity.

13.5 PROFILE LIKELIHOOD FIT

To extract the $t\bar{t}H$ signal from data, a complex fit model is required. A detailed description of the statistical methods for high energy physics can be found in Ref. [236] based on the Neyman-Pearson lemma [237] which states that to reject a hypothesis H_0 in favour of hypothesis H_1 the most powerful test is the ratio of their likelihoods. For the analysis presented here, a profile-likelihood fit is employed to extract the signal strength, incorporating the predicted yields and uncertainties in every bin of the analysis regions to fit them to data.

Given a binned data distribution with n_i events per bin, the expectation value of the number of events in a given bin i can be expressed as

$$E[n_i(\boldsymbol{\mu}, \mathbf{k}, \boldsymbol{\theta})] = \sum_{\mu_\alpha \in \boldsymbol{\mu}} \mu_\alpha \cdot s_{\alpha,i}(\boldsymbol{\theta}) + \sum_{k_\beta \in \mathbf{k}} k_\beta \cdot b_{\beta,i}(\boldsymbol{\theta}), \tag{13.1}$$

where $s_{\alpha,i}$ are the predicted signal events of category α and $b_{\beta,i}$ the predicted background events of category β in bin i. The set of signal-strength parameters $\boldsymbol{\mu}$ are the so-called *parameters of interest* where one element μ_α is defined as

$$\mu_\alpha = \frac{\sigma^\alpha}{\sigma^\alpha_{\mathrm{SM}}}, \tag{13.2}$$

with the signal cross-section σ^α and the expected SM cross-section $\sigma^\alpha_{\mathrm{SM}}$. For an inclusive cross-section measurement only one signal-strength parameter is used, while for the STXS measurement in this analysis five signal-strength parameters are employed. The set of background normalisation factors $\mathbf{k}$ with elements k_β allow a freely floating normalisation of a certain background process β. In this analysis, only the $t\bar{t}+\geqslant 1b$ background is chosen to have a freely floating normalisation

which is determined in the fit to data, while all other processes are normalised to their predicted cross-sections and k_β is set to one. Besides these two types of normalisation factors only acting on the normalisation of the signal and background templates without assuming any prior knowledge, the set of nuisance parameters θ provides additional degrees of freedom. The nuisance parameters correspond to the systematic uncertainties acting both on the shape and normalisation of the signal and background templates implemented in the likelihood as Poissonian or Gaussian priors also called *penalty terms*. Their central value is defined to be zero and any deviation from this nominal value is commonly denoted as a *pull* where a ± 1 deviation corresponds to a one standard deviation variation. The binned likelihood function is given as

$$\mathscr{L}(\mu, k, \theta) = \prod_i^m \frac{(E[n_i(\mu, k, \theta)])^{n_i}}{n_i!} e^{E[n_i(\mu,k,\theta)]} \tag{13.3}$$

corresponding to a product of Poisson probabilities for all bins. The likelihood ratio then results in

$$\lambda_\mu = \frac{\mathscr{L}(\mu, \hat{\hat{k}}, \hat{\hat{\theta}})}{\mathscr{L}(\hat{\mu}, \hat{k}, \hat{\theta})}, \tag{13.4}$$

with the single-hat parameters corresponding to the parameter values maximising the likelihood and the double hat indicates that the values of those parameters maximise the likelihood for a given set of μ. As already pointed out in Section 7.2.1, it is statistically more stable to minimise the negative log-likelihood resulting in the test statistic

$$q_\mu = -2 \ln \lambda_\mu. \tag{13.5}$$

In this analysis the compatibility of data to the background-only hypothesis is measured corresponding to $\mu_\alpha = 0 \ \forall \ \mu_\alpha \in \mu$ in Equation (13.4) with the corresponding test statistics denoted as q_0. The discovery significance Z, which is the significance of a deviation from the background-only hypothesis, is given as

$$Z = \sqrt{q_0}. \tag{13.6}$$

The RooStat framework [238, 239] provides the technical implementation of these statistical tools.

The bins i of Equation (13.1) are bins of the classification BDT distribution in the signal regions, and of a simple ΔR variable in the control regions. In each signal region, with the exception of the highest p_T^H region, the shape of the BDT as discriminant variable is used. Due to the low statistics in the highest p_T^H region $SR^{\geqslant 6j}_{\geqslant 4b}$ ($p_T^H \in [450, \infty)$ GeV), only one single bin and therefore only the normalisation is used. The boosted region, however, takes the shape of the BDT distribution into account, since it has sufficient statistics to do so. In the two CRs, the average ΔR between all b-tagged jet pairs in an event (ΔR^{avg}_{bb}) is utilised taking into account both the normalisation and shape in the fit.

13.6 SYSTEMATIC UNCERTAINTIES

The $t\bar{t}H$ analysis is heavily affected by systematic uncertainties from different sources. There are two main categories of systematic uncertainties: the experimental uncertainties originating mainly from the reconstruction of the various physics objects and their calibrations and secondly the modelling uncertainties related to the signal and background process modelling in MC. In total, 216 nuisance parameters, corresponding to the systematic components, and the free-floating $t\bar{t} + \geqslant 1b$ normalisation factor are included in this analysis. They are sorted into subcategories in Table 13.7. The systematic uncertainties can either affect both the shape and the normalisation (SN) or only the normalisation (N) of a process also indicated in Table 13.7.
To each uncertainty component, one nuisance parameter is associated. Especially the experimental uncertainties often have several independent components coming from one type of uncertainty, e.g. the b-jet efficiency calibration provides 45 uncertainty components and thus 45 nuisance parameters are considered in the analysis.
In addition, for every bin considered in the analysis one nuisance parameter is assigned to take into account the uncertainties coming from the finite statistics of the MC samples.

13.6.1 *Experimental Uncertainties*

The experimental uncertainties have in general a rather low impact on the final fit. Only the uncertainties associated to jets and b-tagging have a more important influence. All experimental nuisance parameters are correlated across all analysis channels, regions and processes and typically affect both the shape and normalisation except the luminosity uncertainty.
The total uncertainty on the integrated luminosity of the full RUN II dataset was measured to be 1.7% [201]. To account for differences between data and simulation in the pile-up modelling one additional uncertainty is considered [240].

Jets and Heavy-Flavour Tagging

The uncertainties associated to jets dominate the experimental uncertainties. Even though the single components are in the range of 1%–5% of relative uncertainties, the large number of jets in the targeted final state enhances their effect. The uncertainties on the jet energy scale and resolution amount to 31 and 9 nuisance parameters, respectively [120]. The uncertainties for the jet energy scale are extracted from test-beam and LHC data as well as from simulation. Further uncertainty sources are also considered such as those related to the jet flavour assuming a conservative uncertainty of $\pm 50\%$ on the quark-gluon fraction. Moreover, pileup corrections are taken into account as well as uncertainties from jet kinematics (η-dependence, high p_T jets) as well as detector simulation differences (GEANT4 vs. *AtlFast-II*). The jet energy resolution is measured in dijet events as a function of p_T and rapidity using RUN II data and MC simulation from which also its uncertainties are extracted.
Furthermore, one uncertainty is related to the jet vertex tagger calibration accounting for differences

Systematic uncertainty	Type	Components
Experimental uncertainties		
Luminosity	N	1
Pile-up modelling	SN	1
Physics Objects		
Electrons	SN	7
Muons	SN	15
Jet energy scale	SN	31
Jet energy resolution	SN	9
Jet vertex tagger	SN	1
E_T^{miss}	SN	3
b-tagging		
Efficiency	SN	45
Mis-tag rate (c)	SN	20
Mis-tag rate (light)	SN	20
Modelling uncertainties		
Signal		
$t\bar{t}H$ cross-section	N	2
H branching fractions	N	3
$t\bar{t}H$ modelling	SN	4
$t\bar{t}$ + jets **Background**		
$t\bar{t}$ cross-section	N	1
$t\bar{t} + \geqslant 1c$ normalisation	N	1
$t\bar{t} + \geqslant 1b$ normalisation	N (free floating)	1
$t\bar{t}$ + light modelling	SN	4
$t\bar{t} + \geqslant 1c$ modelling	SN	4
$t\bar{t} + \geqslant 1b$ modelling	SN	17
Other Backgrounds		
$t\bar{t}W$ cross-section	N	2
$t\bar{t}Z$ cross-section	N	2
$t\bar{t}W$ modelling	SN	1
$t\bar{t}Z$ modelling	SN	1
Single top cross-section	N	3
Single top modelling	SN	7
W+jets normalisation	N	3
Z+jets normalisation	N	3
Diboson normalisation	N	1
4t cross-section	N	1
Small backgrounds cross-sections	N	3

Table 13.7.: Overview of all sources of systematic uncertainty considered in the analysis. The expression "SN" means that both the shape and normalisation are taken into account while "N" stands for the normalisation effects only. The right column states the number of components in which the systematic is split for these processes.

between data and simulation measured in $Z \to \mu^-\mu^+$ events analogous to [203].

Since this analysis relies heavily on b-tagging, it is also a source of systematic uncertainties. The b-tagging calibrations are described in Section 8.6 and provide uncertainties as a function of the different b-tagging working points and the jet p_T (the inefficiency calibration depends also on jet $|\eta|$). A principal component analysis (eigenvalue decomposition) yields uncorrelated uncertainties which are in the range of 2%–10% for the b-jet efficiency calibration and between 10% to 25% and 15% to 50% for the c-jets and light-flavour jets mis-tag rate calibration, respectively. In total, the flavour-tagging uncertainties are decomposed into 85 components.

Leptons

Even though the systematic uncertainties related to leptons have a small effect, 22 different uncertainty sources are taken into account [127, 128]. They are coming from the trigger, reconstruction, identification and isolation efficiencies for electrons (four components) and muons (ten components). Moreover, three (five) independent uncertainty components for electrons (muons) are arising from the lepton momentum scale and resolution.

Missing Transverse Momentum

The systematic uncertainties associated to the missing transverse momentum have only a small impact on the final result because E_T^{miss} is only used in the event reconstruction. Since the E_T^{miss} is calculated from the reconstructed physics objects and a soft term (see sec. 6.2.5), the energy scale and resolution uncertainties from the physics objects are propagated to the E_T^{miss} together with an additional component for the soft term.

13.6.2 *Modelling Uncertainties*

In contrast to the experimental uncertainties, the modelling uncertainties are not correlated across all background and signal processes, but typically they are still correlated across channels and analysis regions with some exceptions. The uncertainties are split into several components depending on the signal and background processes as well as into different physics effects in MC generators.
While the cross-section, branching fraction and normalisation uncertainties only affect the normalisation of the physics processes, all other modelling uncertainties are also sensitive to shape effects (see Table 13.7).

SIGNAL MODELLING

To determine the signal cross-section uncertainty, which is only relevant for the measurement of the signal strength (not for the cross-section measurement), the PDF and α_S in the fixed-order

calculation are varied, resulting in an uncertainty of $\pm 3.6\%$ [107, 241–245]. The effect of the PDF variations on the shapes of the distributions used in the analysis is negligible.

To quantify the impact of ISR, a simultaneous variation of the renormalisation μ_R and factorisation scales μ_F by a factor 0.5 (higher parton radiation) and a factor 2 (lower parton radiation) are performed in the ME together with a variation of α_S^{ISR} in the PS to retrieve its uncertainty. Similarly, the uncertainty on the FSR is evaluated, varying α_S^{FSR} in the PS. The effect of these systematic uncertainties in the analysis is evaluated using event weights.

To evaluate modelling uncertainties, often so-called *two-point systematics* are used. They are retrieved by comparing two different MC generators setups, extracting an uncertainty from their differences. The nominal generator setup and the alternative setups are summarised in Table 13.2.
The two systematic uncertainties related to PS & hadronisation and NLO matching are retrieved in this way by comparing the nominal setup POWHEGBOX+PYTHIA8 to POWHEGBOX+HERWIG7 and to MADGRAPH5_aMC@NLO+PYTHIA8, respectively.

The ISR down variation was found to have the largest impact on the total cross-section as well as on the p_T^H shape estimating the uncertainty caused by missing higher-order terms in the perturbative QCD calculations. This amounts to an uncertainty of 9.2% for the total cross-section and to 10%–17% for STXS bin migration uncertainties retrieved using the Stewart-Tackmann procedure [246] which uses scale variations in the fixed-order calculations to estimate the uncertainties.
An uncertainty dedicated to the Higgs boson branching fraction for the $H \to b\bar{b}$ decay mode amounts to 2.2% [107]. Apart from the bin migration uncertainties, all signal process related nuisance parameters are correlated across all STXS bins, i.e. between all signal templates.

$t\bar{t}$ + jets MODELLING

The $t\bar{t}$ + jets modelling uncertainties are categorised in the subcategories $t\bar{t} + \geqslant 1b$, $t\bar{t} + \geqslant 1c$ and $t\bar{t}$ + light since they are typically affected differently by the systematic uncertainties. Thus all systematic uncertainties associated to $t\bar{t}$ + jets are uncorrelated across these three subcategories and therefore have separate nuisance parameters. Nevertheless, the uncertainty of one category is correlated across all bins (with some exceptions as explained below). The $t\bar{t} + \geqslant 1b$ and $t\bar{t} + \geqslant 1c$ processes are fairly sensitive to differences in the precision of the ME calculation or the utilised flavour scheme. The $t\bar{t}$ + light processes profit from already well known precise measurements. Table 13.8 lists all systematic sources related to the $t\bar{t}$ + jets process.

On the inclusive $t\bar{t}$ cross-section (NNLO+NNLL) an uncertainty of $\pm 6\%$ is taken only applied to $t\bar{t}$ + light samples due to their dominance in the inclusive phase space [206–212]. This uncertainty comprises several effects from varying different quantities like the factorisation and normalisation scales, the PDFs, α_S as well as the top-quark mass.
The normalisation of the $t\bar{t} + \geqslant 1c$ component was a free-floating parameter in the previous iteration

Uncertainty source	Description		Components
$t\bar{t}$ cross-section	$\pm 6\%$		$t\bar{t}+$ light
$t\bar{t}+\geqslant 1b$ normalisation	Free-floating		$t\bar{t}+\geqslant 1b$
$t\bar{t}+\geqslant 1c$ normalisation	$\pm 100\%$		$t\bar{t}+\geqslant 1c$
NLO matching	MADGRAPH5_aMC@NLO+PYTHIA8 vs. POWHEGBOX+PYTHIA8		All
PS & hadronisation	POWHEGBOX+HERWIG7 vs. POWHEGBOX+PYTHIA8		All
ISR	Varying α_S^{ISR} (PS), μ_R&μ_F (ME)	in POWHEGBOXRES+PYTHIA8	$t\bar{t}+\geqslant 1b$
		in POWHEGBOX+PYTHIA8	$t\bar{t}+\geqslant 1c$, $t\bar{t}+$ light
FSR	Varying α_S^{FSR} (PS)	in POWHEGBOXRES+PYTHIA8	$t\bar{t}+\geqslant 1b$
		in POWHEGBOX+PYTHIA8	$t\bar{t}+\geqslant 1c$, $t\bar{t}+$ light
$t\bar{t}+\geqslant 1b$ fractions	POWHEGBOX+HERWIG7 vs. POWHEGBOX+PYTHIA8		$t\bar{t}+1b/1B$, $t\bar{t}+\geqslant 2b$
p_T^{bb} shape	Shape mismodelling measured from data		$t\bar{t}+\geqslant 1b$

Table 13.8.: Overview of the systematic uncertainties associated to the $t\bar{t}+$jets modelling grouped in three different section. The first section comprises normalisation and cross-section uncertainties while the uncertainties in the second section are designed such that they do not influence the normalisation of $t\bar{t}+\geqslant 1b$, $t\bar{t}+\geqslant 1c$, and $t\bar{t}+$ light. The third section comprises uncertainties specifically assigned to the $t\bar{t}+\geqslant 1b$ mis-modelling effects. All systematic uncertainty sources are uncorrelated across the three sub-components [5].

of the analysis [13]. In the following, $t\bar{t}+\geqslant 1c$ is normalised to the SM prediction with a 100% uncertainty. The $t\bar{t}+\geqslant 1b$ normalisation is kept free-floating in the fit.

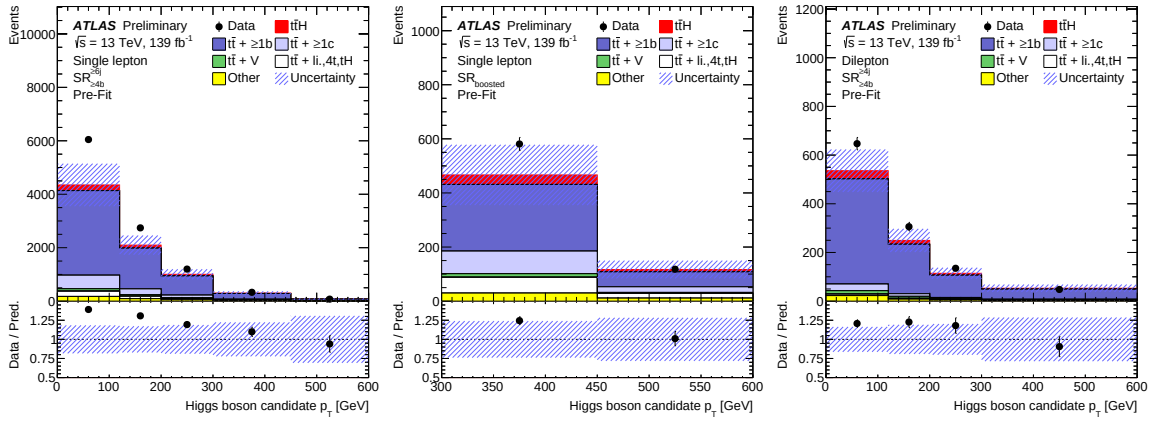
The systematic uncertainties listed in the middle row of Table 13.8 are retrieved as described for the $t\bar{t}H$ modelling, with the exception that in all alternative samples the $t\bar{t}+\geqslant 1b$ fraction is reweighted to be the same as for the nominal generators, leaving it to the fit to extract this information from data via the normalisation factor $k(t\bar{t}+\geqslant 1b)$. For the ISR uncertainty the μ_R and μ_F in the ME are varied by a factor 0.5 (2.0) and α_S^{ISR} in the PS is set to 0.140 (0.115) rather than the nominal value 0.127. To retrieve the FSR uncertainty α_S^{FSR} is changed to 0.1423 and 0.1147 in place of the nominal value $\alpha_S^{FSR} = 0.127$. The variations for both systematic uncertainties (ISR, FSR) are performed on the respective nominal samples, i.e. POWHEGBOXRES+PYTHIA8 (4FS) for $t\bar{t}+\geqslant 1b$ and POWHEGBOX+PYTHIA8 (5FS) for $t\bar{t}+\geqslant 1c$ and $t\bar{t}+$ light. For the determination of the NLO matching and the PS & hadronisation uncertainties, both being two-point systematics, no alternative 4FS generator samples with sufficient statistics are available . This uncertainty is not meant to cover differences between the 4FS and the 5FS modelling since it was found that the 4FS represents data better than 5FS and therefore no dedicated uncertainty coping for this difference is being used in this analysis. Consequently, the relative difference between POWHEGBOX+PYTHIA8 (5FS) and MADGRAPH5_aMC@NLO+PYTHIA8 (5FS) as well as POWHEGBOX+HERWIG7 (5FS) for the NLO matching and PS & hadronisation uncertainty is used instead, respectively.

The predicted fraction of the $t\bar{t}+\geqslant 1b$ subcomponents ($t\bar{t}+\geqslant 2b$ and $t\bar{t}+1b/1B$) as shown in Figure 13.4 are varying for different MC generators. Therefore, an additional uncertainty is associated to account for these differences. The discrepancies between the POWHEGBOX+PYTHIA8 $t\bar{t}$ (5FS) and POWHEGBOX+HERWIG7 $t\bar{t}$ (5FS) models are used to estimate this effect, resulting in

a normalisation difference of $\mp 19.5\%$ and $\pm 41.2\%$ for the $t\bar{t} + \geqslant 2b$ and $t\bar{t} + 1b/1B$ components, respectively[6]. The nuisance parameter associated to this uncertainty is correlated across all analysis regions. The NLO matching and PS & hadronisation uncertainties are adapted to be independent of the respective $t\bar{t} + \geqslant 1b$ sub-component normalisations.

p_T^{bb} *Shape Uncertainty*

The transverse momentum of the reconstructed Higgs candidate p_T^H is not well modelled as shown in Figure 13.11. Especially the p_T^H distribution in the resolved lepton+jets channel (see Fig. 13.11 (a)) shows a clear slope in the data over MC prediction ratio. A similar behaviour also occurs in the dilepton channel (see Fig. 13.11 (c)). Only the signal regions are examined since they are split into p_T^H bins while the control regions are inclusive in p_T^H. To cope with this effect, an additional uncer-



(a) (b) (c)

Figure 13.11.: Distribution of the reconstructed p_T^H before the fit for the (a) resolved and (b) boosted lepton+jets and (c) the dilepton SRs. All uncertainties including their correlations are considered in the uncertainty band excluding $\kappa(t\bar{t} + \geqslant 1b)$ which is only defined post-fit [5].

tainty is retrieved only taking into account shape effects of the $t\bar{t} + \geqslant 1b$ process. In a first step, the inclusive pre-fit normalisation $\kappa(t\bar{t} + \geqslant 1b)$ in the resolved lepton+jets signal regions (Fig. 13.11 (a)) is determined via

$$\kappa(t\bar{t} + \geqslant 1b) = \frac{N_{data} - N_{MC}(\text{non-}t\bar{t} + \geqslant 1b)}{N_{MC}(t\bar{t} + \geqslant 1b)} = 1.52, \tag{13.7}$$

where N_{data} is the number of data events, $N_{MC}(t\bar{t} + \geqslant 1b)$ the number of predicted $t\bar{t} + \geqslant 1b$ events and $N_{MC}(\text{non-}t\bar{t} + \geqslant 1b)$ the amount of background events not originating from the $t\bar{t} + \geqslant 1b$ category (the signal process is excluded in the calculation). Using the data information in this step is legitimate, even though the analysis is performed in a blinded way, since the signal contribution in

6 While writing this thesis, it was found that those values are not correct and the actual differences are smaller ($\mp 7.4\%$ and $\pm 10.8\%$ for $t\bar{t} + \geqslant 2b$ and $t\bar{t} + 1b/1B$) as indicated in Figure 13.4. Given that the systematic uncertainty used for the preliminary result (conference note [5]) is conservative, the fit was not redone for this thesis, but will be redone with the correct values for the paper.

each p_T^H-bin region is small prior to applying the classification BDT. Next, the weight per p_T^H bin i is calculated following the formula

$$\omega_i = \frac{N^i_{data} - N^i_{MC}(\text{non-}t\bar{t}+\geqslant 1b)}{N^i_{MC}(t\bar{t}+\geqslant 1b)} \frac{1}{\kappa(t\bar{t}+\geqslant 1b)}, \tag{13.8}$$

which takes into account the pre-fit normalisation from Equation (13.7). The same procedure is repeated for the dilepton channel and the results from the lepton+jets channel are also applied to the boosted channel. The resulting weights are shown in Figure 13.12. These weights are applied as an additional nuisance parameter (p_T^{bb} shape) in the fit in the corresponding reconstructed p_T^H bin and are correlated between all channels. This uncertainty is constructed such that a pull of one standard deviation in the fit would be equivalent to fully reweighting the $t\bar{t}+\geqslant 1b$ samples with ω_i.

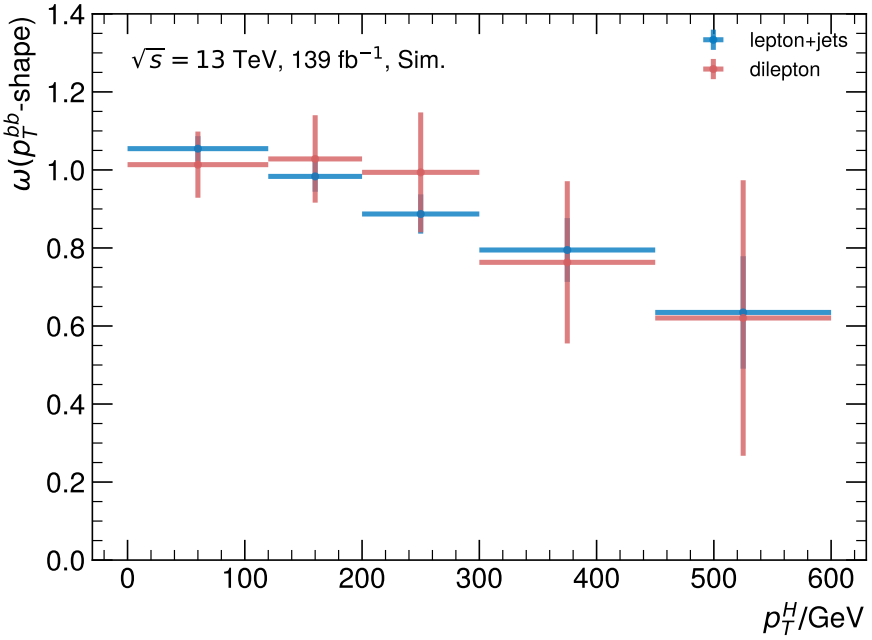


Figure 13.12.: The weights corresponding to the p_T^{bb} shape uncertainty retrieved via Equation (13.8) for lepton+jets (blue) and dilepton (red). The vertical error bars represent the statistical uncertainty and the horizontal error bars indicate the bin width while for the last bin all events with $p_T > 450$ GeV are included.

In order to give the fit enough flexibility for the signal extraction when dealing with the background mismodelling, the NLO matching and PS & hadronisation uncertainties corresponding to the $t\bar{t}+\geqslant 1b$ sub-category are decorrelated between the lepton+jets and dilepton channels. To get additional freedom, the NLO matching uncertainty is decorrelated for $t\bar{t}+\geqslant 1b$ between the p_T^H bins of the SRs. This was studied in blinded fits where only signal depleted bins are used where the criterion was the goodness of fit.

OTHER BACKGROUND PROCESSES

The systematic uncertainties associated to background processes other than $t\bar{t}$ + jets are summarised in Table 13.9 with their respective sources and the corresponding descriptions. These uncertainties play a subordinate role compared to the $t\bar{t}$ + jets uncertainties.

Sample	Uncertainty source	Description
single-top production [76, 218, 219, 247]	cross-section	±5%
	NLO matching	POWHEGBOX+PYTHIA8 vs. MADGRAPH5_aMC@NLO+PYTHIA8
	PS & hadronisation	POWHEGBOX+PYTHIA8 vs. POWHEGBOX+HERWIG7
	tW –$t\bar{t}$ inference@NLO [229]	[DR] vs [DS] in POWHEGBOX+PYTHIA8
$t\bar{t}V$	NLO cross-Section [248]	±15% (split into PDF and scale uncertainties)
	NLO matching	MADGRAPH5_aMC@NLO+PYTHIA8 vs. SHERPA
	PS & hadronisation	
W+ jets	cross-section	±40%
W+HF-jets	normalisation	±30% (uncorrelated between events with = 2 and > 2 HF-jets)
Z+ jets	normalisation	±35% (uncorrelated across jet bins)
Diboson	over-all	±50% (incl. cross-section & additional jet production [249])
four-top	normalisation	±50% (variation of μ_F and μ_R, PDFs and α_S [250])
tZq	cross-section	±0.9% (PDFs) ±7.9% (factorisation & renormalisation)
tWZ	cross-Section [250]	±50%

Table 13.9.: Overview of systematic uncertainties associated to the modelling of all background processes other than $t\bar{t}$ + jets. The abbreviation DR denotes the *diagram removal scheme* (nominal), DS is short for *diagram subtraction scheme* and HF stands for *heavy flavour*.

14

ANALYSIS RESULTS

To extract the signal, a profile likelihood fit is performed as described in Section 13.5. In total, 16 orthogonal analysis regions are used in the combined fit: five $SR^{\geqslant 6j}_{\geqslant 4b}$ and two $CR^{5j}_{\geqslant 4b}$ from the resolved lepton+jets channel as defined in Section 13.4.1, two signal regions from the boosted lepton+jets channel as well as four $SR^{\geqslant 4j}_{\geqslant 4b}$ and three CRs from the dilepton channel (an overview of all analysis regions is shown in Figure B.1). The binning in each region was optimised in the seek of the best sensitivity and together they have 53 bins.
Both, the inclusive cross-section measurement and the STXS measurement use the same strategy. The only difference is that in the STXS case, the signal template is split up into five parts according to the truth Higgs transverse momentum and for every signal template a separate signal normalisation (signal strength) μ is considered in the fit.

14.1 INCLUSIVE CROSS-SECTION MEASUREMENT

14.1.1 *Expected Performance*

The analysis is optimised on MC simulation and the performance is evaluated via the *Asimov* dataset instead of data. This dataset is built from the nominal background and signal simulation. Therefore, by construction, the signal strength and the background normalisation are 1 and the nuisance parameters are not pulled. Nevertheless, uncertainties on the signal strength and the background normalisation as well as the expected significance can be extracted in the profile likelihood fit.
To speed up the fitting and to facilitate its convergence, the shape and normalisation of systematic uncertainties are pruned if they are below a threshold of 1%. The pruning is performed on a bin-by-bin basis for each sample and analysis region. In the previous publication [13] it was shown that this pruning threshold reduces the computing time for the fit significantly and no change in the uncertainty on the signal strength nor in pulls or constraints of the nuisance parameters was found. In addition, to reduce the impact in the calculation of systematic uncertainties coming from statistical fluctuations in MC samples, smoothing algorithms are employed. In Figure 14.1 a selection of systematic uncertainties is shown with their 1σ impact on the nominal modelling.

The expected inclusive signal strength from the combined fit of all three channels results in $\mu_{\text{incl.}} = 1.00^{+0.36}_{-0.31}$ and $k(t\bar{t}+\geqslant 1b) = 1.00 \pm 0.07$ with an expected significance of 3.4 standard deviations. To retrieve the statistical uncertainty, a separate fit is performed with the free parameters $\mu_{\text{incl.}}$ and $k(t\bar{t}+\geqslant 1b)$[1]. The systematic uncertainty is then calculated by quadratically subtracting the stat-

1 For the data fit later, all nuisance parameters are set to their respective post-fit values from the nominal fit to retrieve the statistical uncertainty.

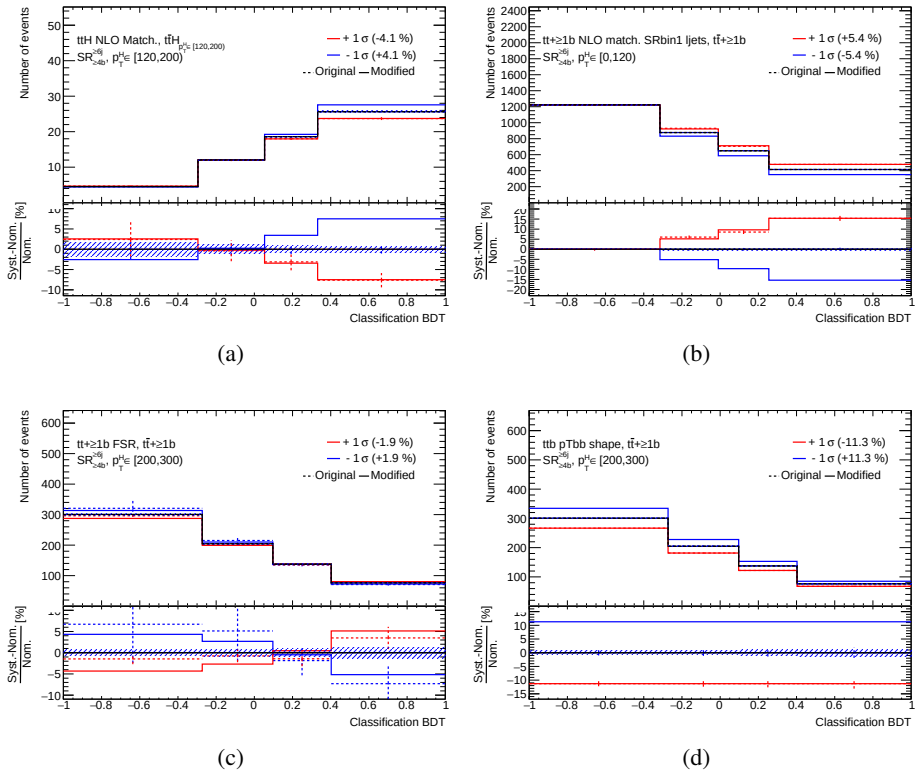


Figure 14.1.: Comparison of the nominal prediction (black) with the one standard deviation up (red) and down (blue) variation of the (a) $t\bar{t}H$ NLO matching uncertainty in the $p_T \in [120, 200)$ GeV bin, (b) the $t\bar{t} + \geqslant 1b$ NLO matching uncertainty in the $p_T \in [0, 120)$ GeV bin, (c) the $t\bar{t} + \geqslant 1b$ FSR uncertainty in the $p_T \in [200, 300)$ GeV bin and (d) the $t\bar{t} + \geqslant 1b$ p_T^{bb} shape uncertainty in the $p_T \in [200, 300)$ GeV bin. The dashed lines indicate the original 1σ variation and the solid lines the 1σ variation after smoothing and symmetrisation.

istical uncertainty from the total uncertainty which results in $\mu_{\text{incl.}} = 1.00 \pm 0.18\,(\text{stat.})\,^{+0.31}_{-0.26}\,(\text{syst.})$ dominated by the systematic uncertainty.

Comparing this result to the result previously published on a subset of the RUN II data ($36.1\,\text{fb}^{-1}$) shown in Figure 5.2 (a), a significant improvement in terms of the systematic uncertainty was made, being almost a factor two smaller. The statistical uncertainty only decreased by about 40% even though the statistics of the dataset increased by about a factor 3.8, caused by a tighter event selection in this analysis. In the previous publication with $36.1\,\text{fb}^{-1}$, the $t\bar{t} + \geqslant 1b$ background was modelled in the 5FS and uncertainties incorporating the differences between the 4FS and 5FS were assigned which had the second-largest impact in the analysis. This time, this uncertainty is not used because the $t\bar{t} + \geqslant 1b$ background is modelled with the 4FS. Together with improvements in the b-tagging calibration allowing a better region definition, this resulted in an increased sensitivity.

In data, the k-factors and nuisance parameter pulls can deviate from 1 and thus modify the sensitivity of the analysis. To calculate a more *realistic* significance, the nuisance parameter pulls and the $k(t\bar{t} + \geqslant 1b)$ value from data are taken into account building a pseudo-dataset and fitting it. This pseudo-dataset is built by using the nuisance parameter pulls from a data fit where the signal strength is set to the SM expectation $\mu = 1$. The realistic expected significance is determined to be 3.0

standard deviations which is an improvement compared to the expected sensitivity of 1.6 standard deviations from the previous publication.

14.1.2 *Fit Results on Data*

The combined profile likelihood fit of the inclusive cross-section to data gives a best-fit signal strength of

$$\mu_{\text{incl.}} = 0.43^{+0.20}_{-0.19}\,(\text{stat.})\,^{+0.30}_{-0.27}\,(\text{syst.}) = 0.43^{+0.36}_{-0.33},$$

measured with an observed significance of 1.3 standard deviations above the background-only hypothesis. The $t\bar{t} + \geqslant 1b$ background normalisation factor results in

$$k(t\bar{t} + \geqslant 1b) = 1.26 \pm 0.09.$$

The inclusive signal strength is compatible with the 36.1 fb^{-1} result (see Fig. 5.2 (a)) within their uncertainties. Even though the expected sensitivity improved from an expected significance of 1.6 to 3.0 standard deviations, the observed significance decreased from 1.4 to 1.3 standard deviations due to the low signal strength.

Moreover, Figure 14.2 shows the comparison of the fit with the combined inclusive signal strength and the fit where each of the three channels (resolved and boosted lepton+jets and dilepton channel) have an individual signal strength, the so-called *3-μ fit*. The fit procedure is otherwise identical to the nominal fit. The single μ values are compatible with the inclusive μ obtained in the nominal fit within their uncertainties. The signal strength from the resolved lepton+jets has the smallest uncertainty and is thus the most sensitive. The $t\bar{t} + \geqslant 1b$ normalisation factor from the 3-μ fit resulted in $k(t\bar{t} + \geqslant 1b) = 1.26^{+0.09}_{-0.08}$ which is in good agreement with the result from the nominal fit. The compatibility is also checked with a χ^2-test comparing the negative log-likelihood values of the 3-μ and nominal fit. A probability of 83% is found to get a difference equal or larger than the one observed between the 3-μ and the nominal fit. A log-likelihood scan of the signal strength in the nominal (combined 1μ) fit is shown in Figure 14.3. It shows the difference in the log-likelihood for different values of the signal strength with respect to the best-fit value. By definition, The best-fit value of the measured signal-strength lies exactly in the minimum of the distribution.
Furthermore, another kind of fit is performed in which a separate fit in each channel is done not considering any correlations between the channels and the following signal strengths were found:

$$\mu_{\text{incl.}}^{\text{l+jets resolved}} = 0.23^{+0.47}_{-0.46},$$
$$\mu_{\text{incl.}}^{\text{l+jets boosted}} = -0.72^{+1.27}_{-1.61},$$
$$\mu_{\text{incl.}}^{\text{dilepton}} = 0.95^{+0.91}_{-0.84}.$$

Generally, the values of μ are smaller in all three cases than in the nominal combined fit but are compatible with each other within their uncertainties. While the result of the dilepton channel is very similar, the two lepton+jets results change more drastically in the separate μ fit. This indicates that the boosted channel needs constraints on $t\bar{t}+\text{jets}$ and on $k(t\bar{t}+\geqslant 1b)$ from the other channels. Especially when only fitting the boosted signal regions, the signal strength turns negative and a more than 100% uncertainty is associated to it with regard to a $\mu = 1$. The observed differences are caused by correlations of nuisance parameters affecting the different channels which are not taken into account when fitting the channels separately.

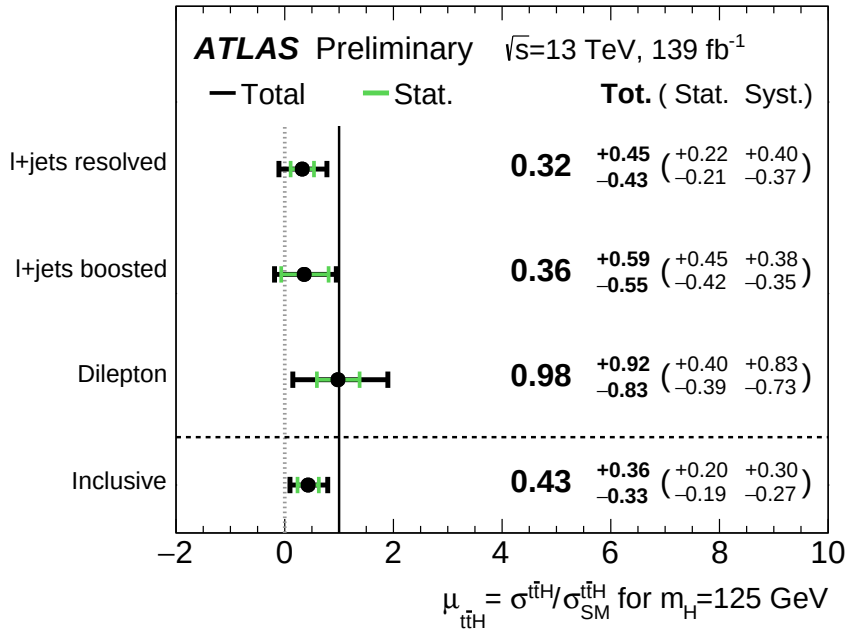


Figure 14.2.: Fitted values of the $t\bar{t}H$ signal strength parameter in the individual channels and in the inclusive signal strength measurement. The results of the three individual channels are retrieved with the 3-μ fit and the inclusive scenario is the nominal combined fit with one signal strength [5].

The correlation matrix of the nuisance parameters and the signal strength for the nominal inclusive data fit is shown in Figure 14.4. The largest correlations are coming from the $t\bar{t}+\geqslant 1b$ fraction nuisance parameter with: the $t\bar{t}+\geqslant 1b$ NLO matching in the dilepton CRs, the $t\bar{t}+\geqslant 1b$ ISR nuisance parameter and the dilepton $t\bar{t}+\geqslant 1b$ PS & hadronisation nuisance parameter, ranging from 48% to 62%. The $t\bar{t}+\geqslant 1b$ NLO matching nuisance parameter for the truth p_T bin $0 \leqslant p_T^H < 120$ GeV is strongly anti-correlated with $\mu_{\text{incl.}}$ (-51%) as well as with the p_T^{bb} shape nuisance parameter (-43%) and the lepton+jets $t\bar{t}+\geqslant 1b$ PS & hadronisation nuisance parameter (-27%). The largest anti-correlation occurs between the $t\bar{t}+\geqslant 1b$ NLO matching nuisance parameter in the lepton+jets CRs and the lepton+jets $t\bar{t}+\geqslant 1b$ PS & hadronisation nuisance parameter (-68%).

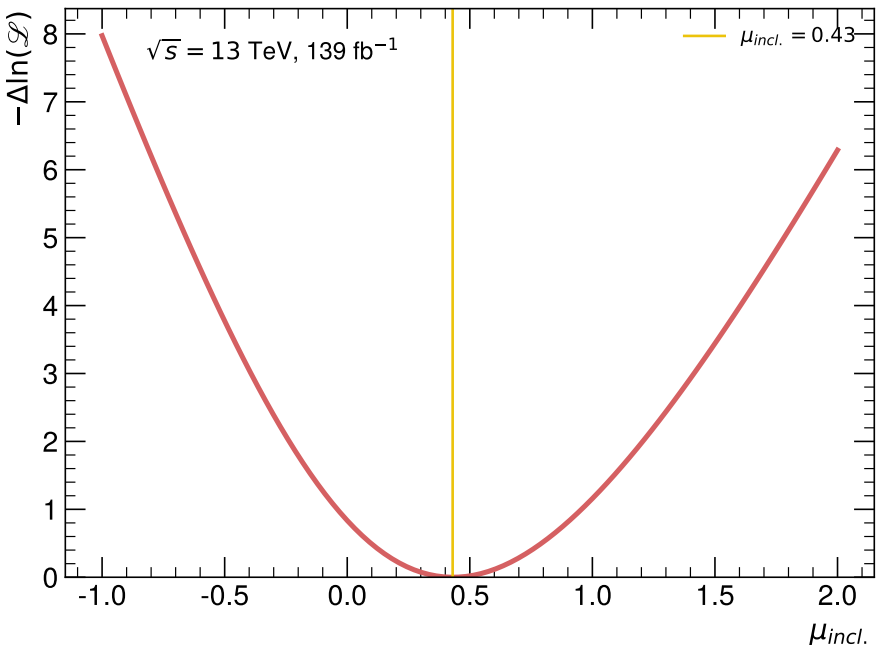


Figure 14.3.: Result of a likelihood scan of the signal strength $\mu_{\text{incl.}}$. The difference in the negative log-likelihood $-\Delta\ln(\mathcal{L})$ is calculated with respect to the best-fit value of the nominal fit indicated as vertical orange line.

14.1.3 *Fit Performance*

In Figure 14.5 a summary of the predicted event yields (pre-fit) and the predicted yields fitted to data (post-fit) compared to the observed event yields in data are shown for all lepton+jets regions. For the pre-fit case, the signal strength and the $t\bar{t}+\geqslant 1b$ normalisation $k(t\bar{t}+\geqslant 1b)$ are fixed to unity and no related uncertainty is considered whereas in the post-fit case the values from the nominal fit results are applied. The post-fit uncertainties then take into account the correlations of all nuisance parameters and their constraints. While the predicted pre-fit event yields show some dis-agreements with data, the post-fit yields agree well in all regions. In several regions, the predicted pre-fit yields are lower than the observed data yields which is well corrected after the fit. The corresponding event yields are also listed in table B.2.

In addition, for all important variable distributions, which are all BDT input variables, all fit inputs and additional kinematic variables, the data/MC agreement was checked and was found to be overall very good post-fit as illustrated in Figure 14.6. To calculate the p-values in this plot, all correlations of the uncertainties are considered. The shown p-values are peaking at one and only a few distributions show up in the tail. Typically, one would expect a flat distribution of the p-values, however, due to the dominance of systematic uncertanties in the analysis, this peak at one occurs. Importantly, the signal (brown) and control (pink) regions have values close to one indicating a good post-fit modelling. The classification BDT distributions of the resolved lepton+jets regions entering the fit are shown in Figure 14.7 and the $\Delta R_{b\bar{b}}^{\text{avg}}$ distributions in Figure 14.8 before the fit and after the fit.

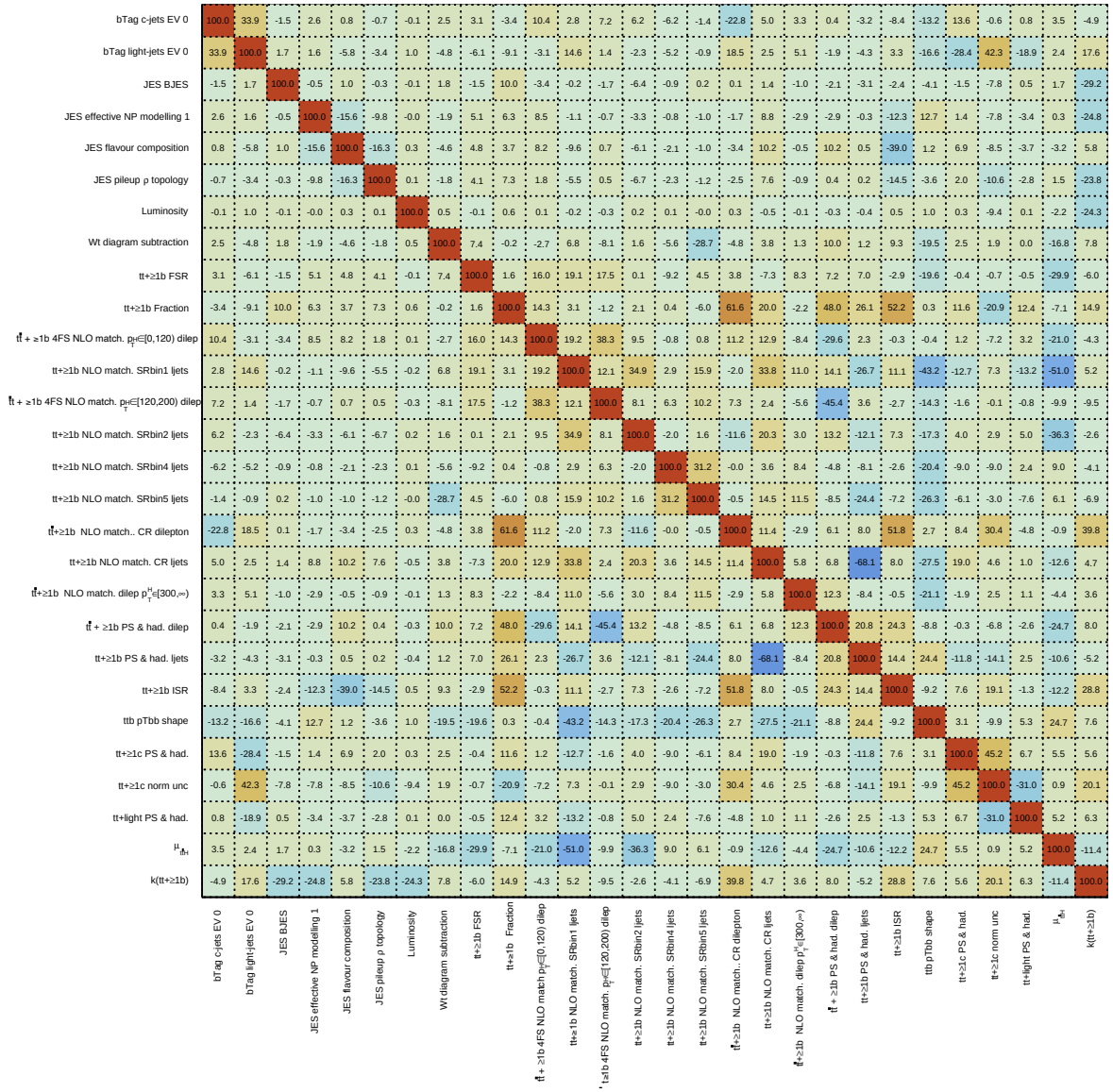
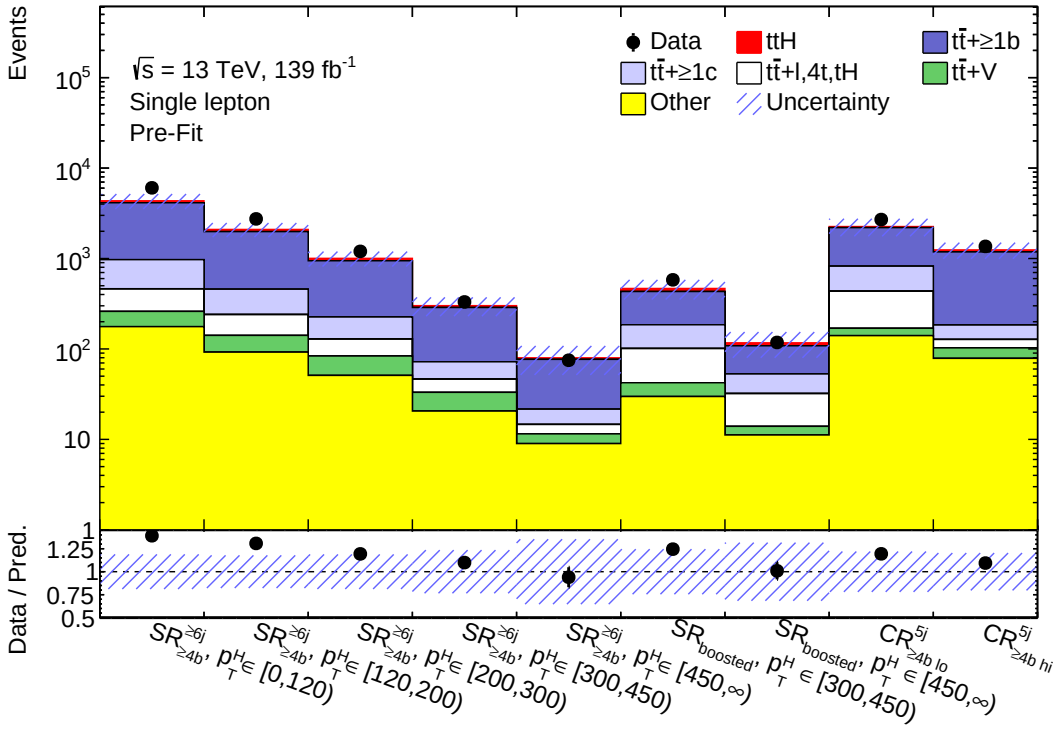
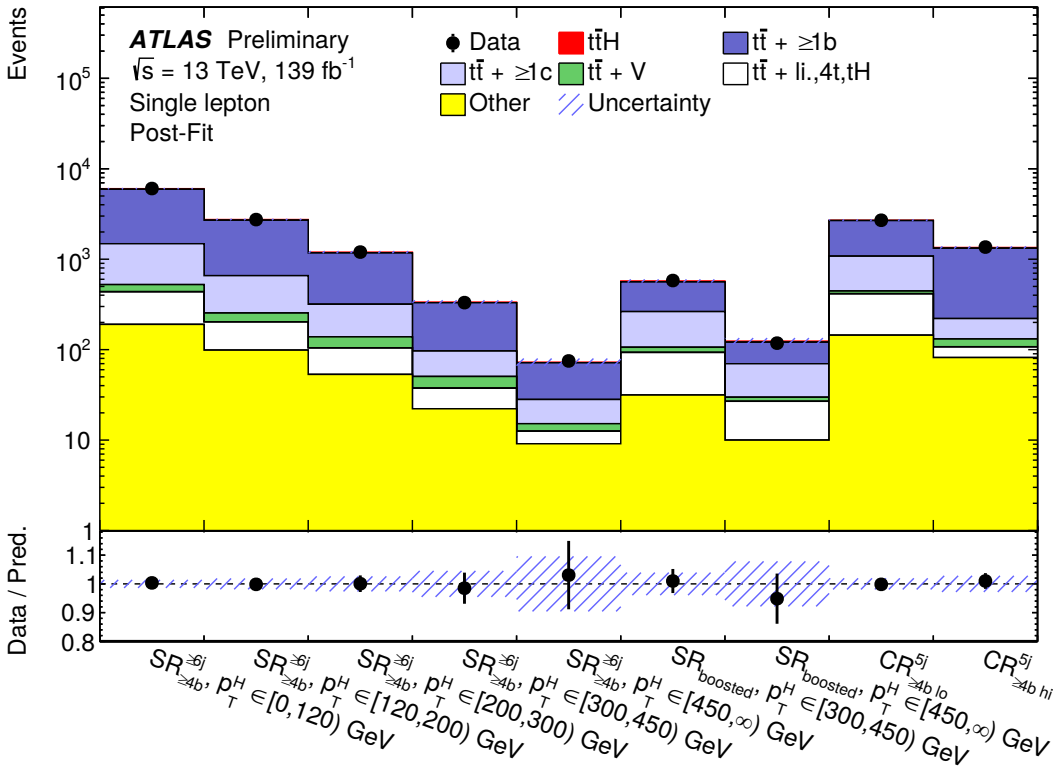


Figure 14.4.: Correlation matrix of the nuisance parameters and signal strength after the nominal inclusive fit to data. All values are given in percent. Each parameter has to have at least one correlation above 20% to be included here. For the $t\bar{t} + \geqslant 1b$ NLO matching uncertainties 'SRbinN', with $N = 1..5$, corresponds to the truth p_T bins $0 \leqslant p_T^H < 120$ GeV, $120 \leqslant p_T^H < 200$ GeV, $200 \leqslant p_T^H < 300$ GeV, $300 \leqslant p_T^H < 450$ GeV and $p_T^H \geqslant 450$ GeV, respectively.



(a)



(b)

Figure 14.5.: Predicted and observed event yields in all regions of the lepton+jets channel (a) before the fit and (b) after the fit to data [5]. The uncertainty band of the pre-fit distribution (a) contains all uncertainties except the uncertainty of $k(t\bar{t}+\geqslant 1b)$ which is only defined post-fit. In the post-fit version (b) all uncertainties with their correlations are considered.

Furthermore, the post-fit distributions of the number of jets and the reconstructed p_T^H are illustrated in Figure 14.9. The agreement with data in the number of jets improved but is still not optimal[2]. In particular, for eight and nine jets the data prefers fewer events and this effect is not covered by the uncertainties. In general, the agreement of the prediction and data improved after the fit as well as the associated model uncertainty caused by correlations and constraints of the nuisance parameters mainly associated to the $t\bar{t} + \geqslant 1b$ background.

The quality of the fit is further examined obtaining the *global goodness of fit* which amounts to 86% indicating a good fit quality. It is evaluated using the so-called *saturated model* [251] which is a model with as many parameters as data points to perfectly describe the data such that it can be used to compare to the actual fit model to evaluate its quality: the *global goodness of fit*.

In Figure 14.10 a summary plot of all events passing the analysis selection is shown as a function of $\log_{10}(S/B)$ which is determined from the signal (S) and background (B) predictions in the different bins entering the fit. Two scenarios are shown: the fit results with the best-fit signal strength (red) and the SM prediction (orange). The data is in good agreement with the nominal fit results nicely visible in the ratio panel. While in most bins the data also agrees well with the SM prediction, in the last three bins, which are most sensitive to the signal, the SM scenario ($\mu_{t\bar{t}H} = 1$) overestimates the event yields.

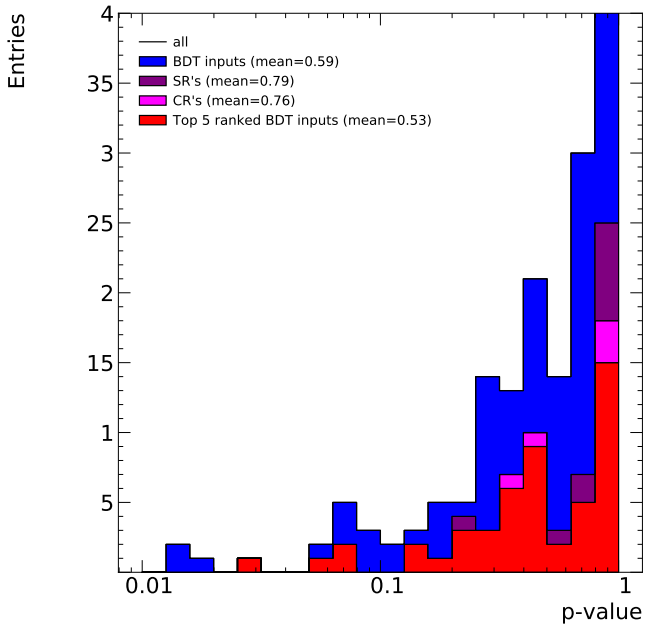


Figure 14.6.: Goodness of fit test as a function of the p-value retrieved from the χ^2 value and the number of degrees of freedom. The p-value is calculated for the BDT input distributions (blue), the signal (brown) and control (pink) regions as well as separately again for the top 5 ranked variables entering the BDT training. Plot retrieved from internal communication with Ana Luisa Carvalho.

2 After the writing of this thesis, an issue was found in the analysis affecting all inclusive plots: certain systematic uncertainties were not drawn in the inclusive distributions. Therefore, the drawn uncertainties, in particular for the number of jets, were too low and, in fact, with the corrected uncertainties the predictions agree within the systematic uncertainties with data. This issue does not affect the fit results, only the inclusive distributions. This will be corrected in the paper publication.

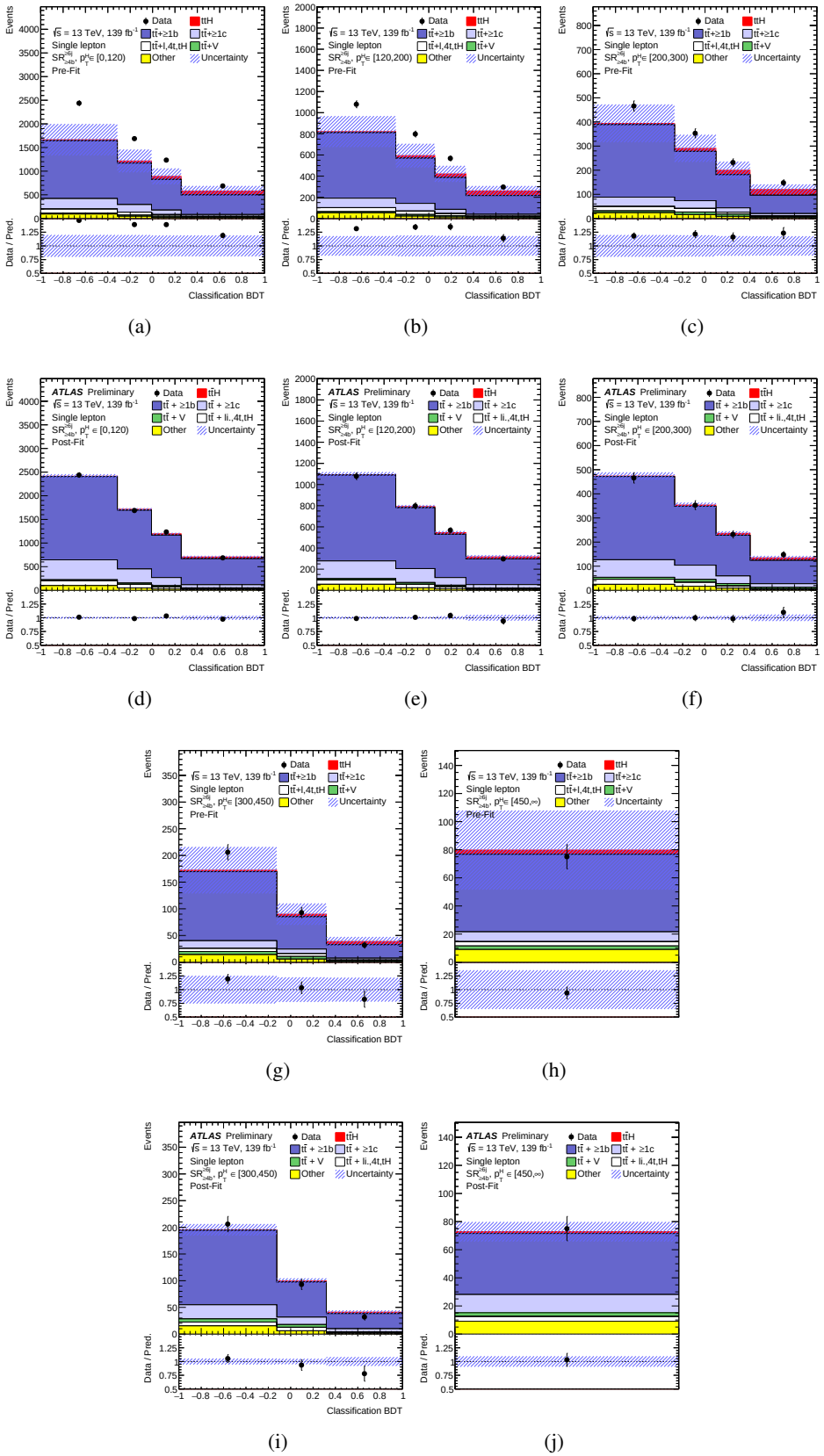


Figure 14.7.: Comparison between data and prediction for the BDT discriminant in the resolved single-lepton SRs before (a-c& g-h) and after (d-f& i-j) the inclusive fit to data shown for (a,d) $0 \leqslant p_T^H < 120$ GeV, (b,e) $120 \leqslant p_T^H < 200$ GeV, (c,f) $200 \leqslant p_T^H < 300$ GeV, (g,i) $300 \leqslant p_T^H < 450$ GeV and (h,j) $p_T^H \geqslant 450$ GeV (yield only). The $t\bar{t}H$ signal yield (solid red) is normalised to the fitted μ value from the inclusive fit. The post-fit uncertainty band includes all uncertainties and their correlations while for the uncertainty on the pre-fit distributions the uncertainty on $k(t\bar{t}+\geqslant 1b)$ is not defined [5].

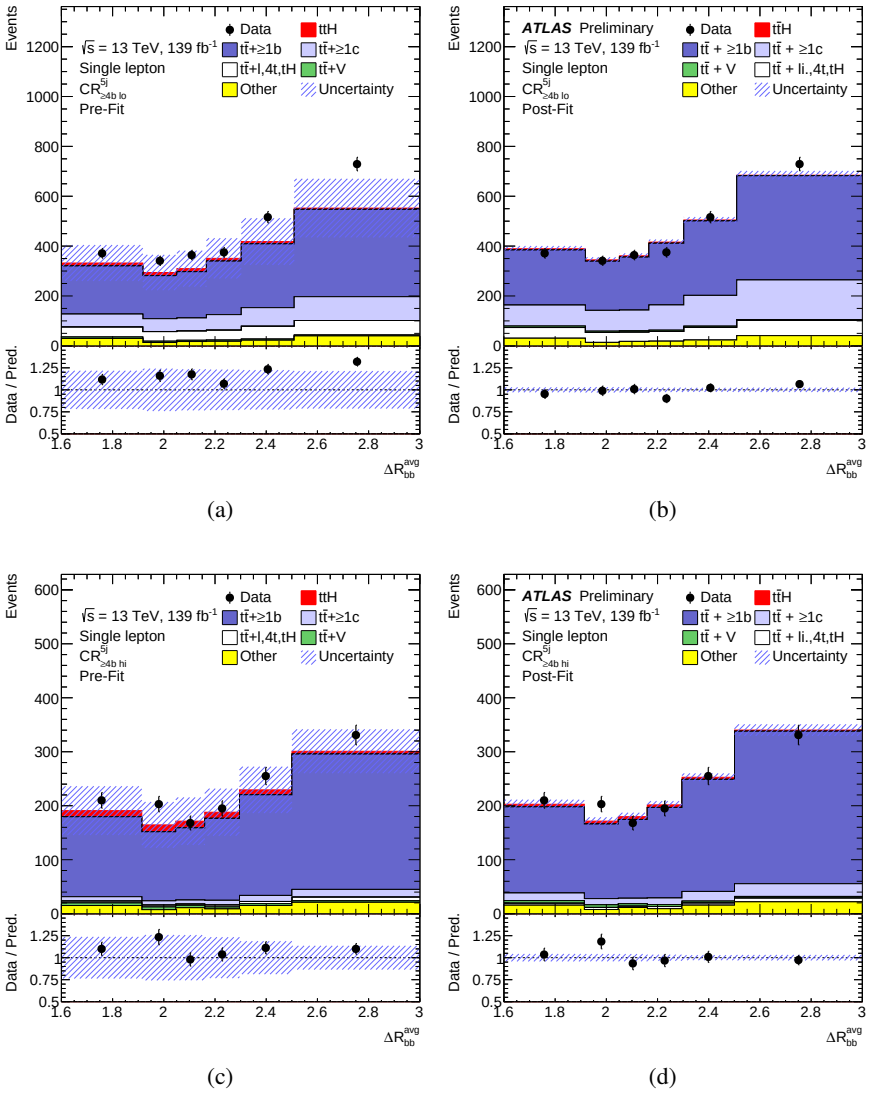
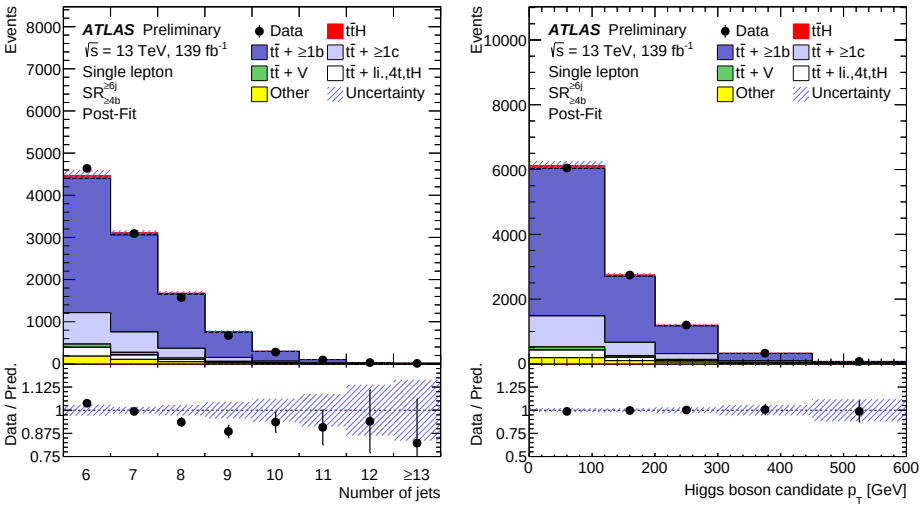


Figure 14.8.: Comparison between data and prediction for the ΔR_{bb}^{avg} discriminant in the resolved single-lepton CRs before (left) and after (right) the inclusive fit to data shown for (a,b) $CR^{5j}_{\geqslant 4b\ lo}$ and (c,d) $CR^{5j}_{\geqslant 4b\ hi}$. The $t\bar{t}H$ signal yield (solid red) is normalised to the fitted μ value from the inclusive fit. The post-fit uncertainty band includes all uncertainties and their correlations while for the uncertainty on the pre-fit distributions the uncertainty on $k(t\bar{t}+\geqslant 1b)$ is not defined [5].



(a) (b)

Figure 14.9.: Post-fit distributions of the (a) number of jets and (b) the reconstructed Higgs boson candidate p_T in the lepton+jets resolved $SR^{\geq 6j}_{\geqslant 4b}$ signal regions. The $t\bar{t}H$ signal yield (solid red) is normalised to the fitted μ value from the inclusive fit. The uncertainty band includes all uncertainties and their correlations [5].

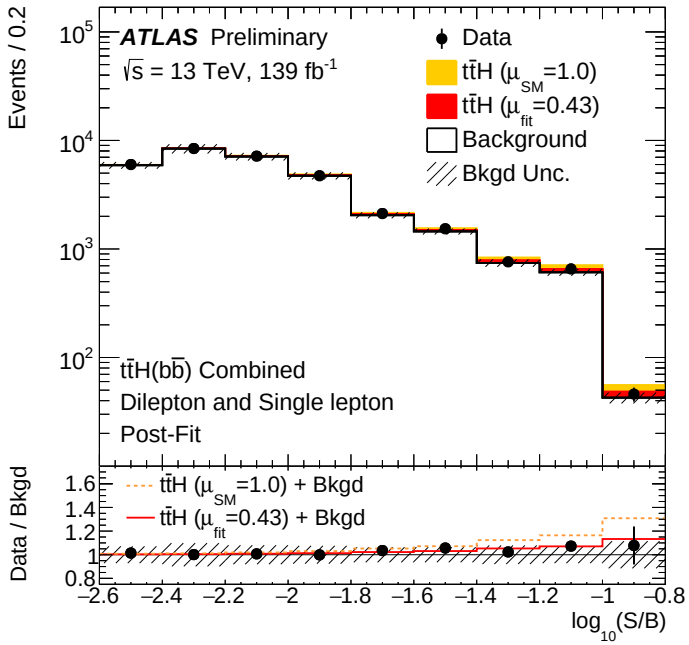


Figure 14.10.: Post-fit yields of signal (S) and total background (B) as a function of $\log_{10}(S/B)$, compared to data. Final-discriminant bins in all dilepton and single-lepton analysis regions are combined into bins of $\log_{10}(S/B)$, with the signal normalised to the SM prediction used for the computation of $\log_{10}(S/B)$. The signal is then shown normalised to the best-fit value and the SM prediction. The lower frame reports the ratio of data over background which is compared to the data over $t\bar{t}H$ signal-plus-background yields for the best-fit signal strength (solid red line) and the SM prediction (dashed orange line) [5].

14.1.4 *Dominant Uncertainties*

The uncertainty associated to the fit result is mainly driven by systematic uncertainties which are described in Section 13.6. The impact of the different nuisance parameters and hence the systematic uncertainty sources are evaluated and ranked by their impact on the signal strength $\Delta\mu$ which is the shift in $\mu_{\text{incl.}}$ evaluated in a separate fit, with fixed nuisance parameter $\hat{\theta} \pm \Delta\hat{\theta}$ [3], with respect to the nominal fit. $\Delta\hat{\theta}$ is the shift on $\mu_{\text{incl.}}$ when shifting a nuisance parameter from its fitted value $\hat{\theta}$ by one standard deviation $\Delta\hat{\theta}$ upwards and downwards. The 20 highest ranked nuisance parameters according to their post-fit impact are shown in Figure 14.11. The upper axis represents the scale for the pre-fit and post-fit impact on μ. The pre-fit (post-fit) impact is given as $\hat{\theta} \pm \Delta\theta$ ($\hat{\theta} \pm \Delta\hat{\theta}$), with $\Delta\theta$ ($\Delta\hat{\theta}$) the pre-fit (post-fit) uncertainties. While $\Delta\theta$ is set to one which is the pre-fit prior corresponding to one standard deviation, the post-fit value of $\Delta\hat{\theta}$ is typically smaller due to constraints from the fit. Both the pre-fit and post-fit impacts are shown as empty and filled rectangles, respectively. The lower axis indicates the scale of the pull of a nuisance parameter defined as $\frac{\hat{\theta}-\theta_0}{\Delta\theta}$ with θ_0 the nominal pre-fit value. The pulls are indicated as black points with their respective error bar. The background normalisation $k(t\bar{t}+\geqslant 1b)$ is drawn with its actual value and since its pre-fit impact is not properly defined, it is not shown. The six highest-ranked nuisance parameters are all associated to the $t\bar{t}+\geqslant 1b$ modelling where the two dominant systematic uncertainties are coming from the NLO matching which are retrieved from the comparison of the two generators MADGRAPH5_aMC@NLO+PYTHIA8 and POWHEGBOX+PYTHIA8. Besides the uncertainties from the $t\bar{t}+\geqslant 1b$ modelling, also tW and signal modelling related nuisance parameters are showing up in the ranking. However, their impact is small compared to the $t\bar{t}+\geqslant 1b$ nuisance parameters. In addition, the impact on the signal strength is evaluated in groups of systematic uncertainty sources listed in Table 14.1. A consistent picture is drawn, the $t\bar{t}+\geqslant 1b$ modelling dominates the systematic uncertainties followed by the signal modelling and the tW modelling. The largest instrumental uncertainty is originating from the flavour-tagging calibration. Moreover, the available MC statistics for the background (Background-model statistical uncertainty) is of similar size as the flavour-tagging uncertainties, which can be reduced by generating more events.

The largest pulls are coming from the $t\bar{t}+\geqslant 1b$ ISR and the p_T^{bb} shape uncertainty. The $t\bar{t}+\geqslant 1b$ ISR nuisance parameter is pulled by about 1.4σ, mainly correcting for the mismodelling of extra radiation in $t\bar{t}+\geqslant 1b$ events. Thus a softer renormalisation and factorisation scale is favoured by data in the ME calculation and should be taken into account in the MC production for a future analysis. Extensive studies were performed understanding this pull in detail. In particular, the distribution of the number of jets, which are used to categorise events, is corrected as shown pre-fit in Figure 13.5 (b) and post-fit in Figure 14.9 (a) due to this pull. The shape of the BDT distributions used as input for the fit are not found to be affected by this pull. In addition, it was checked if decorrelating the $t\bar{t}+\geqslant 1b$ ISR nuisance parameter across all analysis regions would have an impact, but no real differences were spotted. The largest pull was seen in the dilepton $\text{CR}^{\geqslant 4j}_{3b\ \text{hi}}$ while all other pulls not

3 The parameters with a hat (e.g. $\hat{\theta}$) correspond to the best-fit values and those without a hat are the corresponding pre-fit values.

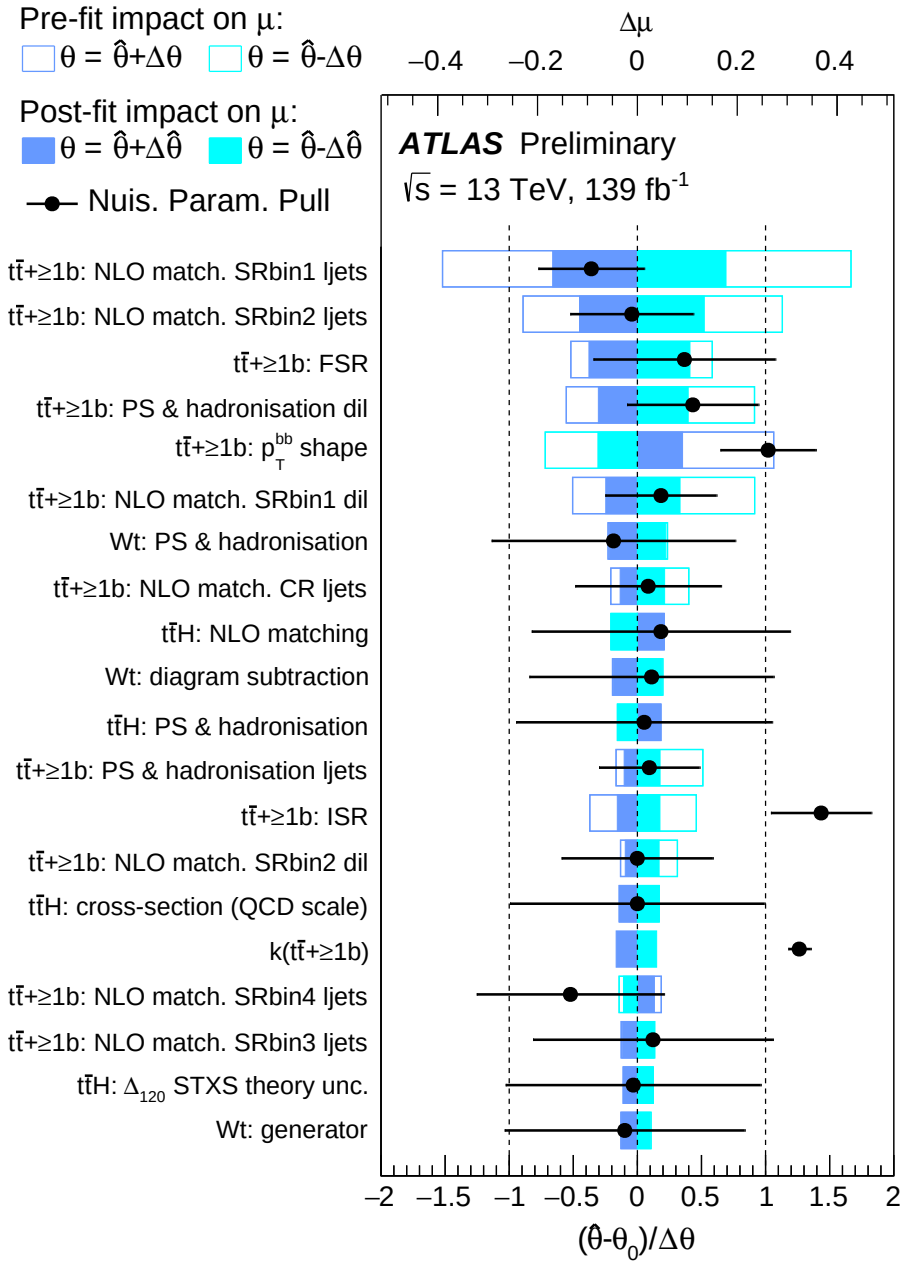


Figure 14.11.: Ranking of the 20 nuisance parameters with the largest post-fit impact on μ in the fit. Nuisance parameters corresponding to MC statistical uncertainties are not included. The empty blue rectangles correspond to the pre-fit impact on μ and the filled blue ones to the post-fit impact on μ, both referring to the upper scale. The impact of each nuisance parameter, $\Delta\mu$, is computed by comparing the nominal best-fit value of μ with the result of the fit when fixing the considered nuisance parameter to its best-fit value, $\hat{\theta}$, shifted by its pre-fit (post-fit) uncertainties $\pm\Delta\theta$ ($\pm\Delta\hat{\theta}$). The black points show the pulls of the nuisance parameters relative to their nominal values, θ_0. For $k(t\bar{t}+\geqslant 1b)$ the pre-fit prior is 1 and not 0 as for the uncertainties and thus the pull is also w.r.t 1. These pulls and their relative post-fit errors, $\Delta\hat{\theta}/\Delta\theta$, refer to the lower scale. For the $t\bar{t}+\geqslant 1b$ NLO matching uncertainties 'SRbinN', with N = 1..5, corresponds to the truth p_T bins $0 \leqslant p_T^H < 120$ GeV, $120 \leqslant p_T^H < 200$ GeV, $200 \leqslant p_T^H < 300$ GeV, $300 \leqslant p_T^H < 450$ GeV and $p_T^H \geqslant 450$ GeV, respectively. The 'ljets' ('dil') label refers to the single-lepton (dilepton) channel [5].

Uncertainty source	$\Delta\mu$	
$t\bar{t} + \geqslant 1b$ modelling	+0.249	-0.236
$t\bar{t}H$ modelling	+0.142	-0.057
tW modelling	+0.083	-0.076
b-tagging efficiency and mis-tag rates	+0.053	-0.046
Background-model statistical uncertainty	+0.045	-0.046
Jet energy scale and resolution	+0.031	-0.030
$t\bar{t} + \geqslant 1c$ modelling	+0.029	-0.031
$t\bar{t}$ + light modelling	+0.020	-0.020
Luminosity	+0.013	-0.002
Other sources	+0.027	-0.026
Total systematic uncertainty	+0.30	-0.27
$t\bar{t} + \geqslant 1b$ normalisation	+0.029	-0.054
Total statistical uncertainty	+0.20	-0.19
Total uncertainty	+0.36	-0.33

Table 14.1.: Breakdown of the contributions to the uncertainties in μ. The contribution of the different sources of uncertainty is evaluated after the fit. The $\Delta\mu$ values are obtained by repeating the fit after having fixed a certain set of nuisance parameters corresponding to a group of systematic uncertainties, and subtracting in quadrature the resulting total uncertainty of μ from the uncertainty from the full fit. The same procedure is followed when quoting the effect of the $t\bar{t} + \geqslant 1b$ normalisation. The total uncertainty is different from the sum in quadrature of the different components due to correlations between nuisance parameters existing in the fit [5].

related to the $t\bar{t} + \geqslant 1b$ ISR nuisance parameters stayed the same. The nuisance parameter related to the p_T^{bb}shape uncertainty in the $t\bar{t} + \geqslant 1b$ background is pulled by about one standard deviation, which is covering the mismodelling of the p_T^H distribution as shown in Figure 13.11 by construction. Therefore, this pull is equivalent to applying a data-driven weight as a function of p_T^H. The post-fit modelling illustrated in Figure 14.9 (b) is hence very good. In a bias-study, the influence on the sensitivity of the p_T^{bb}shape uncertainty was evaluated by decorrelating the free-floating parameter $k(t\bar{t} + \geqslant 1b)$ in every STXS bin, and the bias was found to be negligible. Furthermore, the $t\bar{t} + \geqslant 1c$ normalisation uncertainty is pulled by about a factor 0.6 and strongly constrained. In the previous publication, the $t\bar{t} + \geqslant 1c$ normalisation factor was a free-floating parameter in the fit with a best-fit value of $k(t\bar{t} + \geqslant 1c) = 1.63 \pm 0.23$ [13]. In the analysis presented here, $t\bar{t} + \geqslant 1c$ is not free-floating which is now reflected in this pull. Also, the largest constraints are seen for the $t\bar{t} + \geqslant 1c$ normalisation uncertainty as well as for the $t\bar{t} + \geqslant 1b$ modelling uncertainties.

14.2 STXS MEASUREMENT

Since the analysis regions were designed for the STXS measurement, no big changes are necessary to perform the fit compared to the inclusive cross-section measurement. As described in Section 5.2, the signal template is split into five truth p_T^H bins, also corresponding to the reconstructed p_T^H bins of the SRs and the STXS bin migration uncertainties are removed since each signal template has now a

dedicated signal strength parameter. The resulting best-fit values for the signal strength are shown in Figure 14.12. And the $t\bar{t} + \geqslant 1b$ background normalisation was measured to be

$$k(t\bar{t} + \geqslant 1b) = 1.25^{+0.09}_{-0.08},$$

which is very similar to the observed value in the inclusive fit. In Figure B.2 the pre-fit and post-fit yields are summarised for the lepton+jets regions. Moreover, the pre-fit and post-fit distributions in the lepton+jets signal regions are shown in Figure 14.13. Likewise, the pre-fit distributions show some normalisation off-sets while the post-fit distributions are well modelled. The good overall modelling is also confirmed by the good global goodness of fit of 83%. With a χ^2-test the compatibility of the STXS fit with the SM prediction is evaluated, where all signal strength parameters are fixed to one. This test gives a compatibility of 42.5%.

The uncertainties associated to the signal strength parameter are overall fairly large. All μ values are consistent with each other within their uncertainties as well as with the inclusive measurement. Even though some signal strength parameters are negative, the total signal yield in every single bin used in the fit is never zero or negative after the fit since this is compensated by the other signal strength parameters. The first two signal strength parameters ($\mu_{t\bar{t}H\in[0,120)\,\mathrm{GeV}}$ and $\mu_{t\bar{t}H\in[120,200)\,\mathrm{GeV}}$) are dominated by systematic uncertainties. In contrast to this, the three remaining signal strength parameters are mainly dominated by the statistical uncertainty.

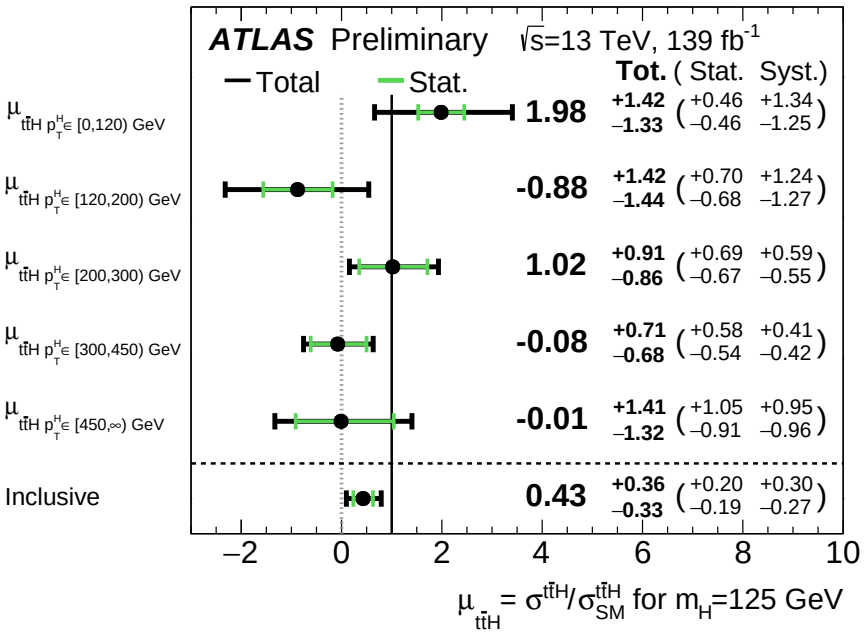


Figure 14.12.: Signal-strength measurements in the individual STXS p_T^H bins, as well as the inclusive signal strength [5]. Both results are retrieved with a combined fit of all three channels.

The impact on the signal strength is separately retrieved for each individual signal strength parameter. The ranking of the 20 most impacting nuisance parameters is shown in Figures 14.14 and 14.15 for each μ separately. In general, the pulls and constraints are very similar to the inclusive

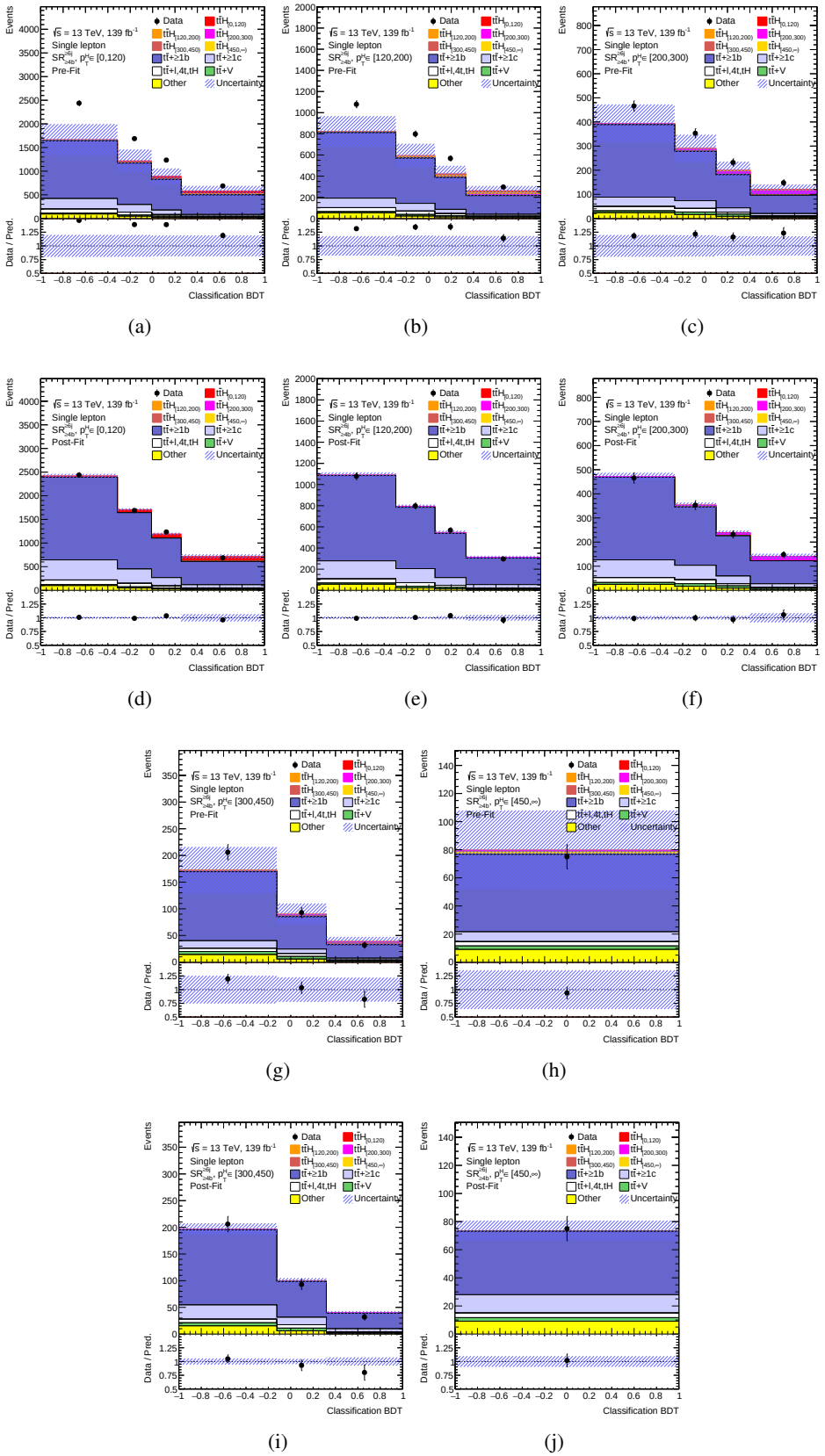


Figure 14.13.: Comparison between data and prediction for the BDT discriminant in the resolved single-lepton SRs before (a-c & g-h) and after (d-f & i-j) the STXS fit to data shown for (a,d) $0 \leqslant p_T^H < 120$ GeV, (b,e) $120 \leqslant p_T^H < 200$ GeV, (c,f) $200 \leqslant p_T^H < 300$ GeV, (g,i) $300 \leqslant p_T^H < 450$ GeV and (h,j) $p_T^H \geqslant 450$ GeV (yield only). The $t\bar{t}H$ signal yields (solid red-ish lines) are normalised to the fitted μ values from the STXS fit. The post-fit uncertainty band includes all uncertainties and their correlations while for the uncertainty on the pre-fit distributions the uncertainty on $k(t\bar{t}+\geqslant 1b)$ is not defined.

cross-section measurement. Again the $t\bar{t} + \geqslant 1b$ ISR uncertainty has the largest pull followed by the p_T^{bb} shape uncertainty. Even though certain instrumental nuisance parameters show up in the ranking, the dominant contributions are originating from the $t\bar{t} + \geqslant 1b$ modelling. The p_T^{bb} shape uncertainty is generally getting more dominant in the higher STXS bins, since the shape effect gets more prominent for larger p_T^H values as shown in Figure 13.12.

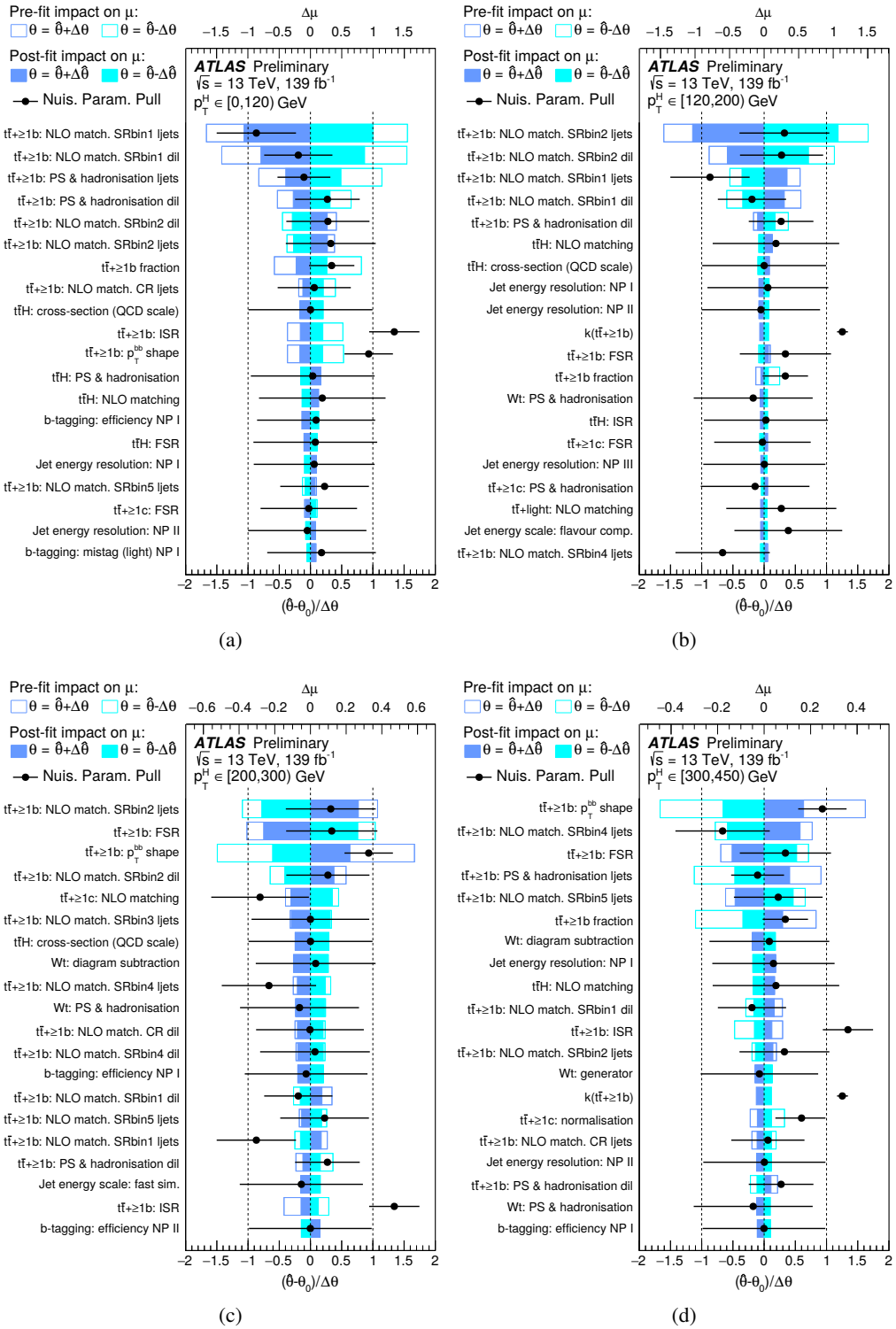


Figure 14.14.: Ranking of the 20 nuisance parameters with the largest post-fit impact on μ in the STXS fit, for (a) $0 \leqslant p_T^H < 120$ GeV, (b) $120 \leqslant p_T^H < 200$ GeVand (c) $200 \leqslant p_T^H < 300$ GeVand (d) $300 \leqslant p_T^H < 450$ GeV. Nuisance parameters corresponding to MC statistical uncertainties are not included. For experimental uncertainties that are decomposed into several independent sources, NP X corresponds to the X^{th} nuisance parameter, ordered by their impact on μ. For the $t\bar{t} + \geqslant 1b$ NLO matching uncertainties 'SRbinN', with $N = 1..5$, corresponds to the truth p_T bins $0 \leqslant p_T^H < 120$ GeV, $120 \leqslant p_T^H < 200$ GeV, $200 \leqslant p_T^H < 300$ GeV, $300 \leqslant p_T^H < 450$ GeV and $p_T^H \geqslant 450$ GeV, respectively. The 'ljets' ('dil') label refers to the single-lepton (dilepton) channel [5].

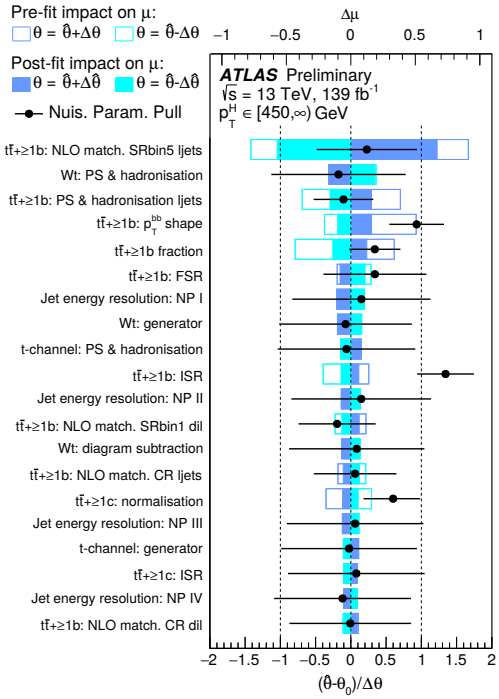


Figure 14.15.: Ranking of the 20 nuisance parameters with the largest post-fit impact on μ in the STXS fit, for $p_T^H \geqslant 450$ GeV. Nuisance parameters corresponding to MC statistical uncertainties are not included. For experimental uncertainties that are decomposed into several independent sources, NP X corresponds to the X$^{\text{th}}$ nuisance parameter, ordered by their impact on μ. For the $t\bar{t}+\geqslant 1b$ NLO matching uncertainties 'SRbinN', with $N = 1..5$, corresponds to the truth p_T bins $0 \leqslant p_T^H < 120$ GeV, $120 \leqslant p_T^H < 200$ GeV, $200 \leqslant p_T^H < 300$ GeV, $300 \leqslant p_T^H < 450$ GeV and $p_T^H \geqslant 450$ GeV, respectively. The 'ljets' ('dil') label refers to the single-lepton (dilepton) channel [5].

15

OUTLOOK

The $t\bar{t}H(b\bar{b})$ analysis presented in Chapters 13 and 14 is dominated by systematic uncertainties among which the modelling of the $t\bar{t} + b\bar{b}$ background is by far the largest uncertainty. In the following, ideas to further reduce the systematic uncertainties and ideas to improve the general analysis strategy will be discussed. The multivariate techniques, namely the reconstruction and classification BDTs, are well suited to be replaced by Neural Networks opening new possibilities in the reconstruction and classification of events. And in particular, the improvements achieved within this thesis in the b-tagging algorithm performance described in detail in Part III are a great basis for further improvements. In the following two different optimisation possibilities are demonstrated: a new tagger incorporating the $t\bar{t} + 1B$ category and a first look into possible improvements using the newly optimised b-tagging algorithm *DL1r*.

15.1 FIRST STUDIES WITH THE NEW DEEP LEARNING BASED $b\bar{b}$-TAGGER

Depending on the chosen MC generator, the fraction of $t\bar{t} + 1B$ events contained in the $t\bar{t} +$ jets category as shown in Figure 13.4 varies between 11% and 13%. The $t\bar{t} + 1B$ events contain two b-hadrons within one jet in addition to the $t\bar{t}$ system as shown in the Feynman graph in Figure 13.3. Due to the flexible structure of the deep-learning-based heavy flavour tagger presented in Part III, it is possible to introduce an additional category to the multi-classification. The extended flavour labelling as described in Section 8.1.1 can be used to identify jets containing two b-hadrons, further denoted as bb-*jets*.

15.1.1 *Training Dataset*

The $t\bar{t}$ samples described in Section 8.1.2 are complemented by $Z(\rightarrow \mu\mu)+$ bb-jets events from which the bb-jets are extracted. The Z+jets events are simulated with the SHERPA v2.2.1 generator using the PDF set `NNPDF3.0NNLO` as indicated in Table 13.2. While the light-flavour, c- and b-jets are extracted from the $t\bar{t}$ sample, the bb-jets are taken from the Z+jets sample only.
The jets are clustered with the *Particle Flow* algorithm.

15.1.2 *Training Setup*

The neural network architecture is adapted from the DL1 tagger described in Section 9.1 adding a fourth classification category, the bb-jets category. The variables used for the training are the same

as for the *DL1rmu* tagger (see. fig 9.2) and the p_T and $|\eta|$ distributions are reweighted to match the b-jet distributions. The same preprocessing steps are performed as for the DL1 tagger trainings. Four example variable distributions are shown in Figure 15.1 which are the most promising input variables for the classification of bb-jets. The charged energy fraction of the secondary vertex, calculated from the JETFITTER algorithm optimised for c-jet identification (a), is shifted to lower values for bb-jets compared to b-jets. This can happen since only the SV closest to the primary vertex is taken into account and hence often only one b-hadron is considered in the calculation which is typically the softer b-hadron since it is closer to the PV and thus more likely to be less boosted than the other. The mass of the vertex from the SV1 algorithm (c) is flatter for bb-jets compared to b-jets and also reaches larger values. In the SV1 algorithm only one secondary vertex is reconstructed and hence in some cases only one b-hadron is part of the SV and for the higher values both b-hadrons are reconstructed in one SV. Unfortunately, the distribution is cut at 6 GeV since the algorithm is optimised for single b-jets and thus not optimal for bb-jets.

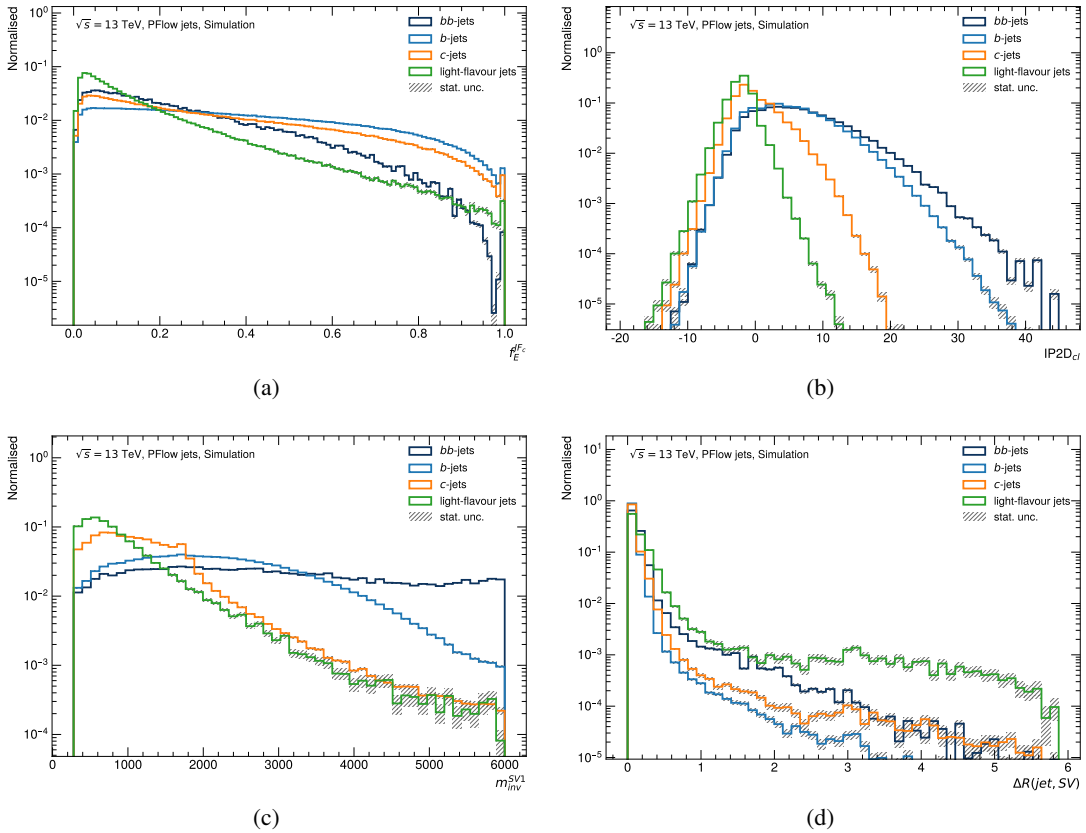
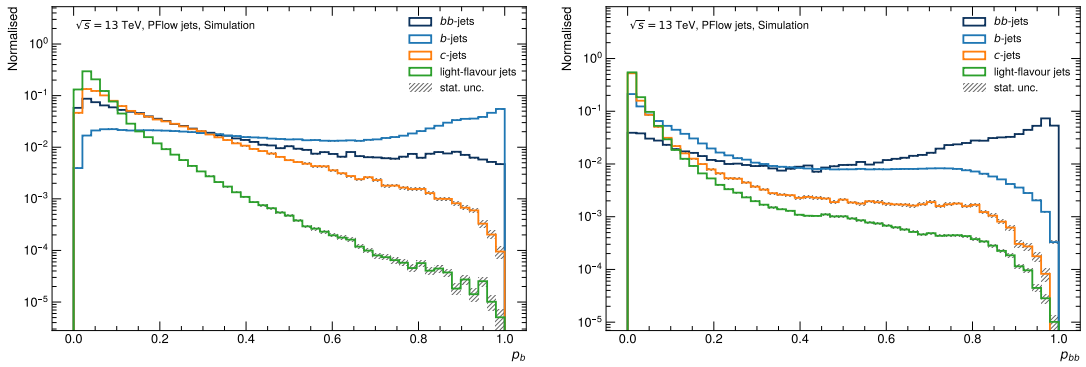


Figure 15.1.: Example input variable distributions for the extended *DL1rmu* tagger training with a bb-category: (a) charged jet energy fraction of JF w.r.t. all tracks in the jet $f_E^{JF_c}$ (b) $\mathrm{IP2D_{cl}}$ (c) the invariant mass of the SV calculated from the associated tracks and (d) $\Delta R(\mathrm{jet}, \mathrm{SV})$.

15.1.3 *First Training Results*

A first training was performed for 200 epochs. Figure 15.2 shows the distribution of the b-jet output node (a) and the bb-jet output node (b) which indicate the probability of a jet being a b- or bb-jet, respectively. Both distributions show a nice separation of the b- and bb-jets. To further evaluate the performance, a simplified version of the log-likelihood discriminant from Equation (7.4) is used $\mathcal{D}_{bb} = \log\left(\frac{p_{bb}}{p_b}\right)$, only taking into account the output-nodes of the b- (p_b) and bb-jets (p_{bb}). This simplification is made since the main goal is to compare the performance between these two classes. For further studies, all output-nodes should be taken into account. At a bb-jet efficiency of 30%, a b-jet rejection of 64 is obtained. Moreover, a rejection of around 5 for bb-jets is achieved at a b-jet efficiency of 70%.



(a) (b)

Figure 15.2.: Output distributions of the extended *DL1r* tagger with an additional bb-category for the (a) b-jet class output node and (b) the bb-jet output node.

15.1.4 *Outlook*

This first study of an extended heavy-flavour tagger incorporating an additional bb-category showed already promising results. Nevertheless, there are still many improvements possible. Especially, the resampling approach presented in Section 9.2 offers a more stable training. Moreover, the newly introduced UMAMI tagger (see sec. 12.1) exploiting track information in an end-to-end training can extract more information to better classify bb-jets.

15.2 FIRST LOOK AT PFLOW JETS AND THE NEW DL1R TAGGER IN $t\bar{t}H(b\bar{b})$

The $t\bar{t}H(b\bar{b})$ analysis benefits from improvements in b-tagging in many ways: smaller calibration uncertainties reflect in smaller systematic uncertainties in the analysis and an improved tagging performance allows to better classify events. The improvements made in this thesis in the b-tagging performance are therefore crucial for further analysis improvements. In this section, a first look is

taken into possible sensitivity improvements due to the new *DL1r* tagger using *PFlow* jets.

A simplified setup is used: no systematic uncertainties are taken into account, only the resolved lepton+jets regions are considered and the sensitivity is only evaluated using MC simulation.
The region definitions are the same as described in Section 13.4.1 with the exception that events from the boosted lepton+jets channel are not vetoed. For *EMTopo* jets, the same simplifications apply, but apart from that the same setup as described in Chapter 13 is used for them (i.e. the MV2c10 tagger), to have a setup for a direct comparison here. Instead of the MV2c10 b-tagging algorithm, the *DL1r* tagger developed in Chapter 10 is employed for *PFlow* jets. To incorporate the new b-tagging algorithm in the full analysis chain, the reconstruction and classification BDTs are also evaluated using the information of *DL1r* but no retraining is performed with respect to the version for *EMTopo* jets. The signal (S) over background (B) ratio (black) as well as $S/\sqrt{B}$ (red) are shown in Figure 15.3 for both the *EMTopo* jets (dotted-dashed lines) and *PFlow* jets (solid lines). Overall, the new setup with *PFlow* jets and *DL1r* shows an increased S/B and $S/\sqrt{B}$ ratio in each resolved lepton+jets region with respect to the *EMTopo* jets.

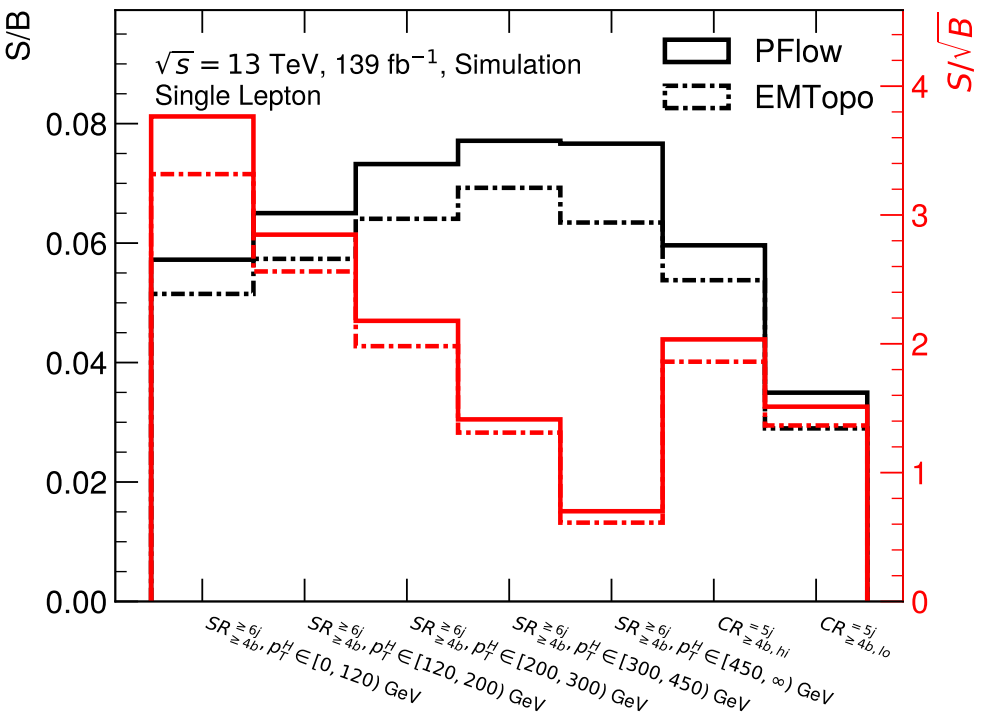


Figure 15.3.: Contribution of the $t\bar{t}H$ signal (S) in the different resolved lepton+jets analysis regions shown for *PFlow* jets (solid lines) which use the new *DL1r* tagger and *EMTopo* jets (dotted-dashed lines) which still use the MV2c10 tagger. The black lines are associated to the left vertical axis showing the signal over background (B) ratio and the red lines show the $S/\sqrt{B}$ distribution corresponding to the right vertical axis.

A statistical only Asimov fit is performed yielding a signal strength of

$$\mu_{\text{incl.}}^{\text{PFlow}} = 1.00 \pm 0.19. \tag{15.1}$$

Moreover, the same fit is carried out for *EMTopo* jets resulting in

$$\mu_{\text{incl.}}^{\text{EMTopo}} = 1.00 \pm 0.22. \tag{15.2}$$

The new jet collection and the improvements from *DL1r* are already reflected in the results for this simplified setup. These results are a good basis for further improvements which are under development towards a RUN II legacy paper.

Part V

CONCLUSION

16

SUMMARY AND CONCLUSION

In this thesis, two closely linked topics were presented: the reoptimisation of the deep-learning based heavy-flavour tagging algorithm in ATLAS and the search for the Higgs boson production in association with a pair of top quarks in the $H \to b\bar{b}$ decay channel ($t\bar{t}H(b\bar{b})$) with the full LHC RUN II dataset. The final state of the $t\bar{t}H(b\bar{b})$ process contains at least four b-jets and thus heavily benefits from improvements in b-tagging.

The reoptimisation of the deep-learning based heavy flavour tagger was performed for two different jet collections: *ParticleFlow* jets and *Variable Radius Track* jets. All future physics analyses in ATLAS will be using *ParticleFlow* jets and thus a well optimised b-tagging algorithm is crucial for these physics results. The input feature handling for the neural network training as well as the network architecture itself were optimised. Significant improvements were achieved, with up to a factor of two in the background rejection for certain phase space regions. Moreover, a hyperparameter optimisation workflow was developed using GRID GPUs to be provided for the use in the collaboration. In addition, a new idea for a heavy-flavour tagger was presented, combining track and jet information in an end-to-end training. This will most probably be the new direction of flavour tagging in ATLAS and is currently being optimised.

The $t\bar{t}H(b\bar{b})$ analysis is carried out with a dataset of $139\,\text{fb}^{-1}$ of proton-proton collisions at a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ recorded with the ATLAS detector. The measurement is directly sensitive to the top Yukawa coupling which is the strongest Yukawa coupling in the Standard Model. The targeted channel in this thesis is the resolved lepton+jets channel which has at least six jets in the final state, where at least four are b-jets. One W-boson from the decay of the associated top-quark pair is decaying leptonically. The final state partons are matched to the jets in the event using a reconstruction BDT. In this analysis, the dominant background process is coming from $t\bar{t}+\text{jets}$ events, especially from $t\bar{t}+\geqslant 1b$ events. For the final result, this channel is combined with the boosted lepton+jets channel optimised for $p_T^H > 300\,\text{GeV}$ and the dilepton channel. The signal strength of the inclusive cross-section measurement resulted in

$$\mu_{\text{incl.}} = 0.43^{+0.20}_{-0.19}\,(\text{stat.})\,^{+0.30}_{-0.27}\,(\text{syst.}),$$

with an observed (expected) significance of 1.3 (3.0) standard deviations. The normalisation factor of the $t\bar{t}+\geqslant 1b$ background is free-floating in the fit and was measured to be $k(t\bar{t}+\geqslant 1b) = 1.26 \pm 0.09$. The measurement is mainly dominated by systematic uncertainties. Particularly, the uncertainties associated to the modelling of the $t\bar{t}+\geqslant 1b$ background process have the

largest impact on the overall uncertainty.

Furthermore, a differential measurement of the Higgs boson transverse momentum was performed, for the first time in the $t\bar{t}H(b\bar{b})$ channel. This was done in the STXS framework which allows an easier comparison between different channels and between experiments. The signal strength was measured in the five bins: $0\text{ GeV} \leqslant p_T^H < 120\text{ GeV}$, $120\text{ GeV} \leqslant p_T^H < 200\text{ GeV}$, $200\text{ GeV} \leqslant p_T^H < 300\text{ GeV}$, $300\text{ GeV} \leqslant p_T^H < 450\text{ GeV}$, and $p_T^H \geqslant 450\text{ GeV}$. The signal strengths associated to the first two p_T^H bins are dominated by their systematic uncertainty while the remaining signal strength parameters are limited by the statistical uncertainty. In general, the uncertainties are fairly large and the different signal strength parameters from the STXS measurement are in agreement with the measured signal strength in the inclusive cross-section.
It was possible to reduce the systematic uncertainties compared to the previous analysis by almost a factor of two, mainly related to the background modelling of $t\bar{t} + \geqslant 1b$. This improvement was mainly achieved by using a 4 flavour scheme based modelling of $t\bar{t} + b\bar{b}$ events, more statistics in the MC simulation and reducing double-counting of certain uncertainty sources.
To further improve the analysis, the multivariate techniques will need to be revised. In particular, the reconstruction BDT could be replaced by a neural network to improve the event reconstruction using customised architectures. In addition, the improvements achieved in this thesis concerning the b-tagging will be important for the future analysis.

Part VI

APPENDICES

A

HYPERPARAMETER OPTIMISATION ON GRID GPUS - TECHNICAL SETUP

To process all the data from the LHC, including simulation, reconstruction etc., powerful computing resources are necessary. The Worldwide LHC Computing Grid (WLCG) combines about 170 computing centres around the world to one powerful computing infrastructure with about 1 million computer cores and 1 exabyte of storage [252]. The vast majority of these computing resources are CPUs. However, in the last couple of years, more and more shared GPU resources became available. Generally, not every institute which is doing ML has GPUs available. Therefore a setup usable by the whole ATLAS collaboration is tested together with ATLAS IT, giving everyone access to GPU resources.

In general, the WLCG software stack is not suitable for ML since it is not flexible enough to cope for the quickly evolving and diverse ML software. At this point, the Linux container images, often called *Docker images*, come into play. They contain a full software stack packed in a container image and are fully isolated from the host environment. This allows the user to choose its custom software. Starting with freely accessible base-images, GitLab can automatically build such a custom image and the site administrators only have to provide the virtualisation software *singularity* [253] to allow the user to execute the container images. Typically, the base images are already optimised for GPU usage.

An example workflow is shown in Figure A.1. The hyperparameter optimisation software [3] generates JSON configuration files containing the information about the hyperparameters of the network architecture. Each configuration file contains a different set of hyperparameters. These configuration files, as well as the training and validation dataset, are uploaded to the WLCG sites (via the rucio data management [254]). A container image built from Gitlab containing the training software is then deployed to various WLCG sites running a large number of jobs over the different hyperparameters (config files). This procedure is illustrated in Figure A.2. Each job uses the same training and validation sample and a certain amount of hyperparameters are tested. The output is again JSON files containing the information about the performance of the NN such as the training and validation loss and the background rejection at the 77% WP. This heavily parallelisable workload is running in a fraction of the time which would be necessary running them one after each other.

The final workflow for ML developments would then for instance look as displayed in Figure A.3. The interactive development is done on the laptop accessing shared and centrally provided computing resources such as a *JupyterHub*[1]. The code is pushed into a GitLab repository where a container image is automatically built and this can then be deployed to the WLCG. In general, the hyperpara-

1 At CERN this is available via `https://hub.cern.ch`.

meter scan can be also performed on any other computing cluster with a job scheduling system and singularity available.

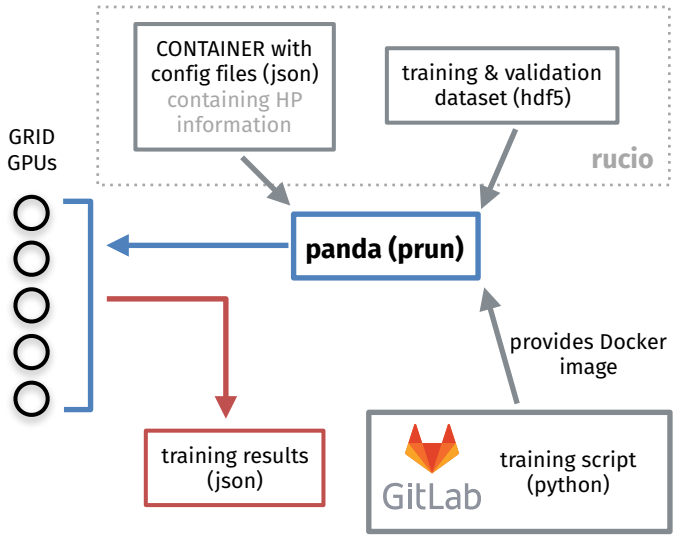


Figure A.1.: Schematic workflow deploying hyperparameter optimisations to the WLCG.

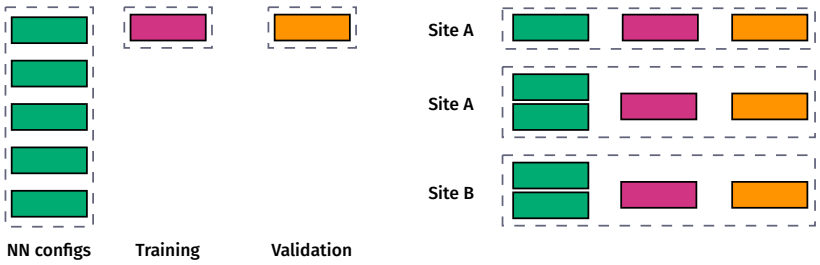


Figure A.2.: Job brokering illustration of the hyperparameter optimisation in GRID GPUs.

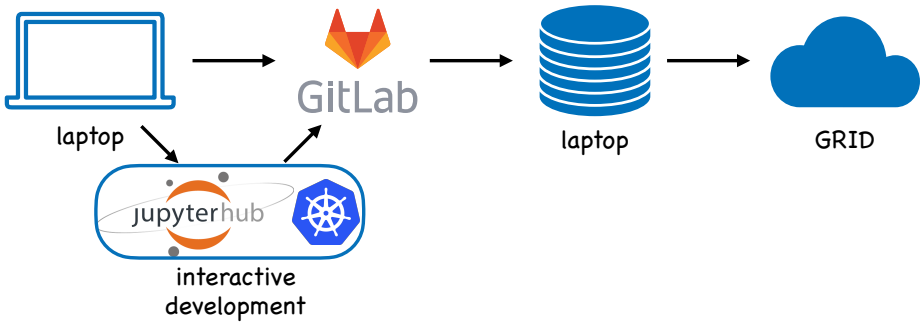


Figure A.3.: Workflow for machine learning model development.

B

ADDITIONAL MATERIAL $t\bar{t}H(b\bar{b})$ ANALYSIS

Variables associated to the topological information of the $t\bar{t}$ system
Mass of the leptonically decaying top quark $t_{lep.}$
Mass of the hadronically decaying top quark $t_{had.}$
Mass of the hadronically decaying W-boson $W_{had.}$
Mass of $W_{had.}$ and the b-quark from the leptonically decaying top quark $b^t_{lep.}$
Mass of $W_{lep.}$ and the b-quark from the hadronically decaying top quark $b^t_{had.}$
$\Delta R(W_{had.}, b^t_{lep.})$
$\Delta R(W_{had.}, b^t_{had.})$
$\Delta R(\ell, b^t_{lep.})$
$\Delta R(\ell, b^t_{had.})$
$\Delta R(b^t_{lep.}, b^t_{had.})$
$\Delta R(q_1$ from $W_{had.}, q_2$ from $W_{had.})$
$\Delta R(b^t_{had.}, q_1$ from $W_{had.})$
$\Delta R(b^t_{had.}, q_2$ from $W_{had.})$
$\mathrm{Min}(\Delta R(b^t_{had.}, q_1$ from $W_{had.}), \Delta R(b^t_{had.}, q_2$ from $W_{had.}))$
$\Delta R(\ell, b^t_{lep.}) - \mathrm{Min}(\Delta R(b^t_{had.}, q_1$ from $W_{had.}), \Delta R(b^t_{had.}, q_2$ from $W_{had.}))$
Variables associated to the Higgs boson candidate
Mass of the Higgs boson candidate M_H
M_H and q_1 from $W_{had.}$
$\Delta R(b_1$ from Higgs candidate, b_2 from Higgs candidate)
$\Delta R(b_1$ from Higgs candidate, $\ell)$

Table B.1.: Input variables of the reconstruction BDT in the lepton+jets channel.

	$SR^{\geq 6j}_{\geq 4b}$, $p_T^H \in$ [0,120) GeV		$SR^{\geq 6j}_{\geq 4b}$, $p_T^H \in$ [120,200) GeV		$SR^{\geq 6j}_{\geq 4b}$, $p_T^H \in$ [200,300) GeV		$SR^{\geq 6j}_{\geq 4b}$, $p_T^H \in$ [300,450) GeV		$SR^{\geq 6j}_{\geq 4b}$, $p_T^H \in$ [450,∞) GeV	
	Pre-fit	Post-fit	Pre-fit	Post-fit	Pre-fit	Post-fit	Pre-fit	Post-fit	Pre-fit	Post-fit
$t\bar{t}H$	213 ± 29	93 ± 74	113 ± 15	49 ± 39	59.9 ± 7.8	26 ± 21	13.9 ± 2.0	5.9 ± 4.6	3.09 ± 0.49	1.26 ± 1.00
$t\bar{t}$ + ≥1b	3160 ± 500	4450 ± 160	1530 ± 240	2040 ± 85	720 ± 140	855 ± 43	215 ± 60	234 ± 20	55 ± 26	43.4 ± 8.2
$t\bar{t}$ + ≥1c	510 ± 540	960 ± 210	220 ± 230	404 ± 87	96 ± 100	179 ± 38	26 ± 27	46 ± 11	6.9 ± 7.5	12.9 ± 3.3
$t\bar{t}$ + light, 4t, tH	200 ± 120	250 ± 140	100 ± 59	105 ± 57	46 ± 24	52 ± 26	13.5 ± 7.9	15.4 ± 8.8	3.2 ± 2.2	3.5 ± 2.2
$t\bar{t}$ + W	7.0 ± 1.2	7.3 ± 1.1	4.31 ± 0.90	4.46 ± 0.87	2.47 ± 0.52	2.54 ± 0.48	1.05 ± 0.32	1.09 ± 0.31	0.47 ± 0.15	0.48 ± 0.14
$t\bar{t}$ + Z	77 ± 11	79 ± 10	44.6 ± 6.6	46.0 ± 6.4	30.1 ± 4.9	31.1 ± 4.9	11.5 ± 2.4	11.8 ± 2.3	2.05 ± 0.64	2.12 ± 0.64
Single top Wt	71 ± 40	80 ± 43	40 ± 26	44 ± 27	17.9 ± 7.6	18.7 ± 7.8	8.5 ± 7.9	9.5 ± 9.0	6.0 ± 5.3	6.1 ± 5.4
Other top sources	46 ± 24	48 ± 25	23 ± 16	24 ± 16	13 ± 10	14 ± 10	4.3 ± 2.8	4.5 ± 2.7	1.08 ± 0.54	1.09 ± 0.54
V & VV + jets	60 ± 24	63 ± 24	29 ± 11	30 ± 11	19.7 ± 8.3	20.6 ± 8.2	7.8 ± 3.4	8.1 ± 3.4	1.90 ± 0.88	1.92 ± 0.84
Total	4350 ± 820	6026 ± 84	2100 ± 370	2747 ± 52	1000 ± 190	1198 ± 31	301 ± 71	336 ± 15	80 ± 28	72.8 ± 7.0
Data	6047		2742		1199		331		75	

	$SR_{boosted}$, $p_T^H \in$ [300,450) GeV		$SR_{boosted}$, $p_T^H \in$ [450,∞) GeV		$CR^{5j}_{\geq 4b\ lo}$		$CR^{5j}_{\geq 4b\ hi}$	
	Pre-fit	Post-fit	Pre-fit	Post-fit	Pre-fit	Post-fit	Pre-fit	Post-fit
$t\bar{t}H$	35.1 ± 4.1	15 ± 12	8.5 ± 1.1	3.6 ± 2.8	61.7 ± 8.1	26 ± 20	62.1 ± 8.6	26 ± 21
$t\bar{t}$ + ≥1b	246 ± 46	297 ± 27	55 ± 23	51.0 ± 9.8	1370 ± 180	1595 ± 80	1000 ± 240	1102 ± 51
$t\bar{t}$ + ≥1c	84 ± 90	157 ± 37	21 ± 23	40 ± 11	390 ± 410	630 ± 140	56 ± 59	90 ± 23
$t\bar{t}$ + light, 4t, tH	59 ± 26	62 ± 25	18 ± 10	16.9 ± 7.6	270 ± 120	270 ± 100	25 ± 16	26 ± 16
$t\bar{t}$ + W	1.86 ± 0.39	1.89 ± 0.36	0.55 ± 0.18	0.57 ± 0.17	2.53 ± 0.53	2.62 ± 0.46	0.54 ± 0.13	0.53 ± 0.12
$t\bar{t}$ + Z	10.7 ± 2.1	11.0 ± 2.1	2.21 ± 0.60	2.34 ± 0.60	26.4 ± 3.7	25.9 ± 3.5	23.5 ± 3.4	22.8 ± 3.1
Single top Wt	13.1 ± 8.0	14.0 ± 8.3	6.1 ± 5.8	4.9 ± 4.3	58 ± 32	60 ± 32	27 ± 20	28 ± 20
Other top sources	4.3 ± 3.2	4.4 ± 3.0	0.80 ± 0.78	0.88 ± 0.78	41 ± 16	41 ± 16	27 ± 11	28 ± 11
V & VV + jets	12.4 ± 5.7	13.1 ± 5.6	4.3 ± 2.3	4.2 ± 2.0	42 ± 16	43 ± 15	24.2 ± 8.8	24.9 ± 8.8
Total	470 ± 110	575 ± 23	117 ± 37	124.4 ± 9.7	2257 ± 500	2700 ± 52	1250 ± 250	1348 ± 38
Data	581		118		2696		1362	

Table B.2.: Summary of the pre-fit and post-fit event yields in the lepton+jets SRs (top) as well as the CRs and boosted SRs (bottom). Post-fit yields are after the inclusive fit in all channels. All uncertainties are included, taking into account correlations in the post-fit case. The uncertainty in the $\mathrm{t\bar{t}} + \geqslant 1\mathrm{b}$ is not defined pre-fit and therefore only included in the post-fit uncertainties. For the $\mathrm{t\bar{t}H}$ signal, the pre-fit yield values correspond to the theoretical prediction and corresponding uncertainties, while the post-fit yield and uncertainties correspond to those in the inclusive signal-strength measurement [5].

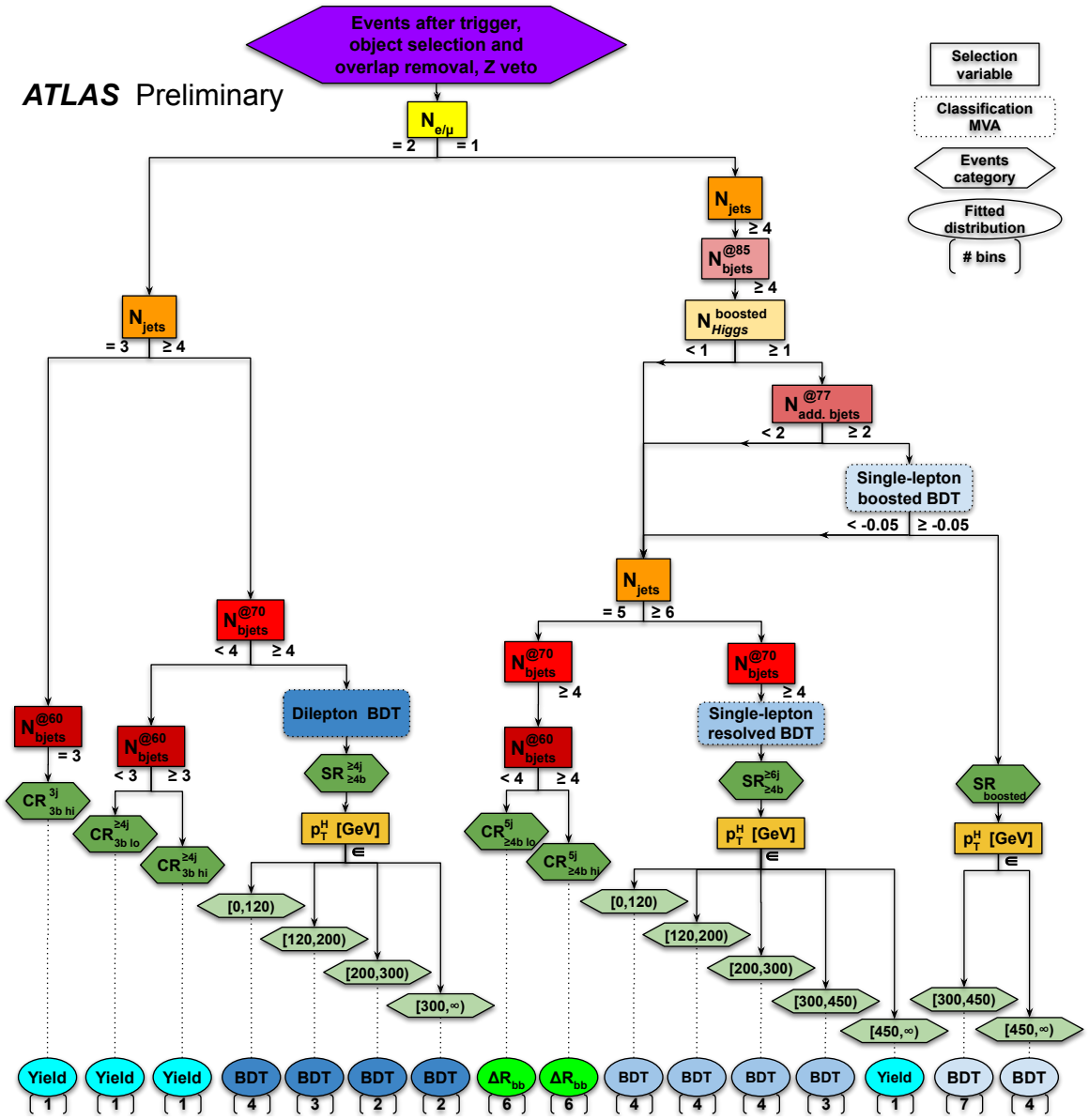
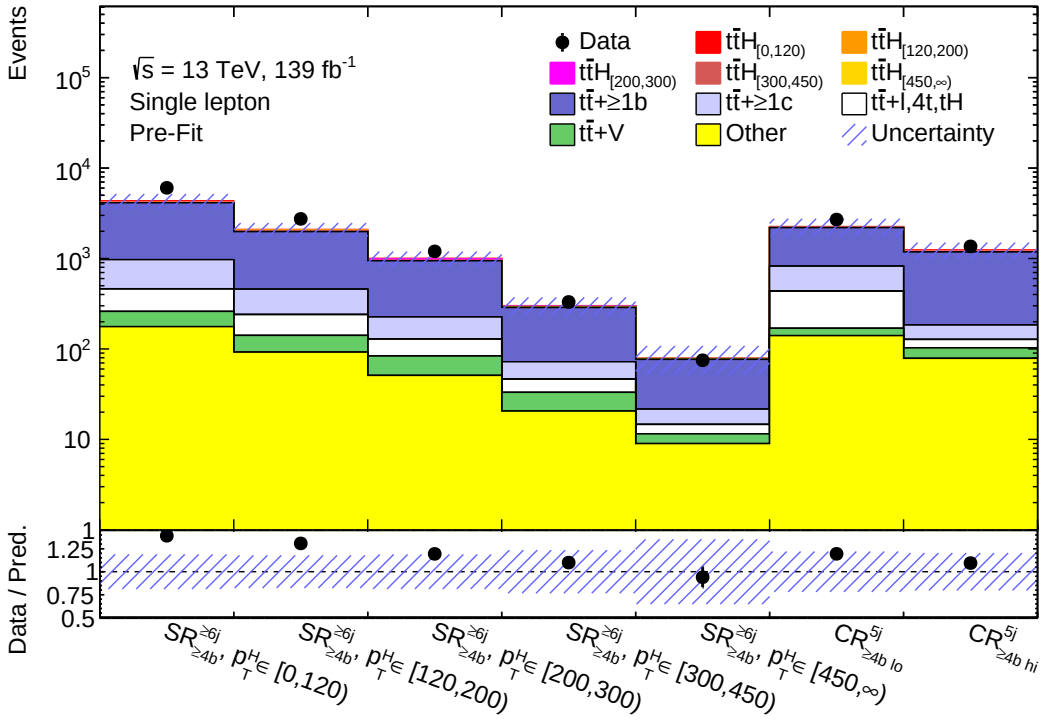
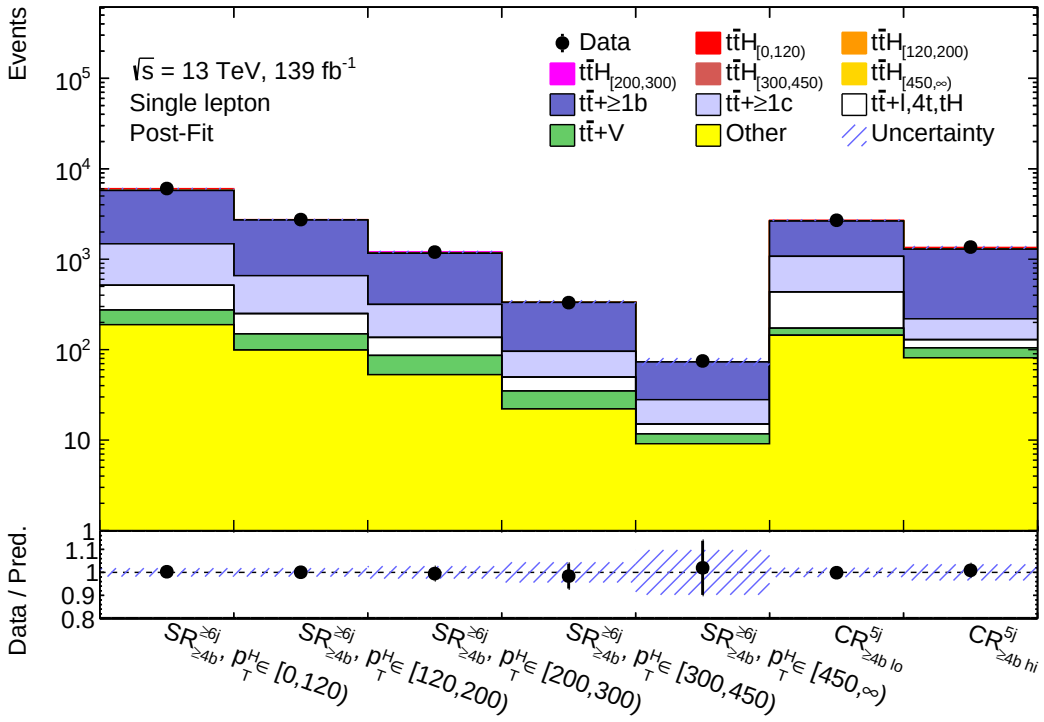


Figure B.1.: Schematic illustration of the event selection presented in a tree structure describing the different selection steps concluding in the different analysis regions for the dilepton, lepton+jets resolved and boosted channels. The numbers in the last row are indicating how many bins are used in the final fit and above the variable is listed used for the fit [5].



(a)



(b)

Figure B.2.: Predicted and observed event yields in all regions of the lepton+jets channel (a) before and (b) after the STXS fit to data. The uncertainty band of the pre-fit distribution (a) contains all uncertainties except the uncertainty of $\mathrm{k}(\mathrm{t\bar{t}} + \geqslant 1\mathrm{b})$ which is only defined post-fit. In the post-fit version (b) all uncertainties with their correlations are considered. The $\mathrm{t\bar{t}H}$ signal is split into the five STXS truth $\mathrm{p_T^H}$ bins.

C

FRENCH SUMMARY

Depuis plusieurs décennies, les prédictions du Modèle Standard (MS) de la physique des particules sont testées et validées. L'un des grands succès du Grand collisionneur de hadrons (LHC) au CERN a été la découverte du boson de Higgs en 2012. Avec le grand nombre de collisions proton-proton enregistrées avec les expériences au LHC, des mesures précises du boson de Higgs sont désormais possibles.

ATLAS et CMS ont récemment découvert le processus de production $t\bar{t}H$ en utilisant les données prises durant le RUN II du LHC. Le processus $t\bar{t}H(H \to b\bar{b})$ permet de mesurer directement le couplage de Yukawa du quark top, qui est le couplage fermion-Higgs le plus grand du modèle standard et joue donc un rôle important dans la physique du boson du Higgs. L'état final de ce processus contient au moins 4 jets provenant de quarks b ce qui nécessite d'établir une stratégie d'analyse avancée ainsi que de développer des méthodes sophistiquées pour l'identification des jets provenant de quarks b. L'étiquetage des quarks b n'est pas seulement crucial pour l'analyse $t\bar{t}H(b\bar{b})$, mais aussi pour la plupart des analyses de physique au sein de l'expérience ATLAS. La ré-optimisation de l'étiquetage des quarks de saveurs lourdes basé sur un apprentissage profond dans ATLAS est présentée dans cette thèse pour deux collections de jets différentes. Le processus $t\bar{t}H(b\bar{b})$ est mesuré en utilisant $139\,\mathrm{fb}^{-1}$ de données enregistrées par ATLAS durant le RUN II à une énergie dans le centre de masse de $\sqrt{s} = 13\,\mathrm{TeV}$.

C.1 LE TAGUEUR DE SAVEURS DE QUARK LOURD BASÉ SUR UN APPRENTISSAGE APPROFONDI

L'identification des saveurs lourdes, également appelée *tagging*, joue un rôle important dans les analyses de physique des particules. Plusieurs processus physiques intéressants, tels que le processus $t\bar{t}H(b\bar{b})$ [5] ou le processus $VH(H \to b\bar{b}, H \to c\bar{c})$ [165, 166], contiennent des quarks b ou c dans leur état final. L'étiquetage des saveurs est un outil crucial pour mieux sélectionner le signal et rejeter les processus de bruit de fond, ce qui est important pour la recherche de nouveaux processus ainsi que les mesures de précisions.

Puisque les quarks ne peuvent pas se présenter sous forme de particules libres, ce ne sont pas les quarks eux-mêmes qui sont étiquetés, mais plutôt les états de couleur neutre correspondant aux hadrons b et c. Dans cette thèse l'étiquetage des saveur dans l'éxpérience ATLAS est présenté.

Les hadrons b ont une durée de vie d'environ 1.5 ps [20]. Ainsi ces hadrons produits à des impulsions de 30 GeV ne se désintègrent donc qu'après avoir traversé 2.5 mm dans le détecteur. De plus, les hadrons b ont une masse relativement élevée de $\sim 5\,\mathrm{GeV}$ [20] ainsi qu'une grande multiplicité dans leur désintégration environ 5 particules stables et chargées en moyenne. Une grande fraction

de ces désintégration correspond à la création d'un boson W et d'un quark connu sous le nom de désintégration semi-leptonique.

La réoptimisation des utils d'étiquetage des saveurs lourdes basée sur un apprentissage profond a été réalisée pour deux collections de jets différentes : les jets *ParticleFlow* et les jets *variable radius track*. Dans le future, toutes les analyses de physique dans ATLAS utiliseront des jets *ParticleFlow*. Ainsi avoir un algorithme d'étiquetage bien optimisé est crucial pour mener ces analyses. La gestion des fonctions d'entrée pour la formation du réseau de neurones utilisé ainsi que l'architecture du réseau elle-même ont été optimisées. Un aperçu de l'architecture du réseau de neurones utilisé dans l'étude est présenté Figure C.1.

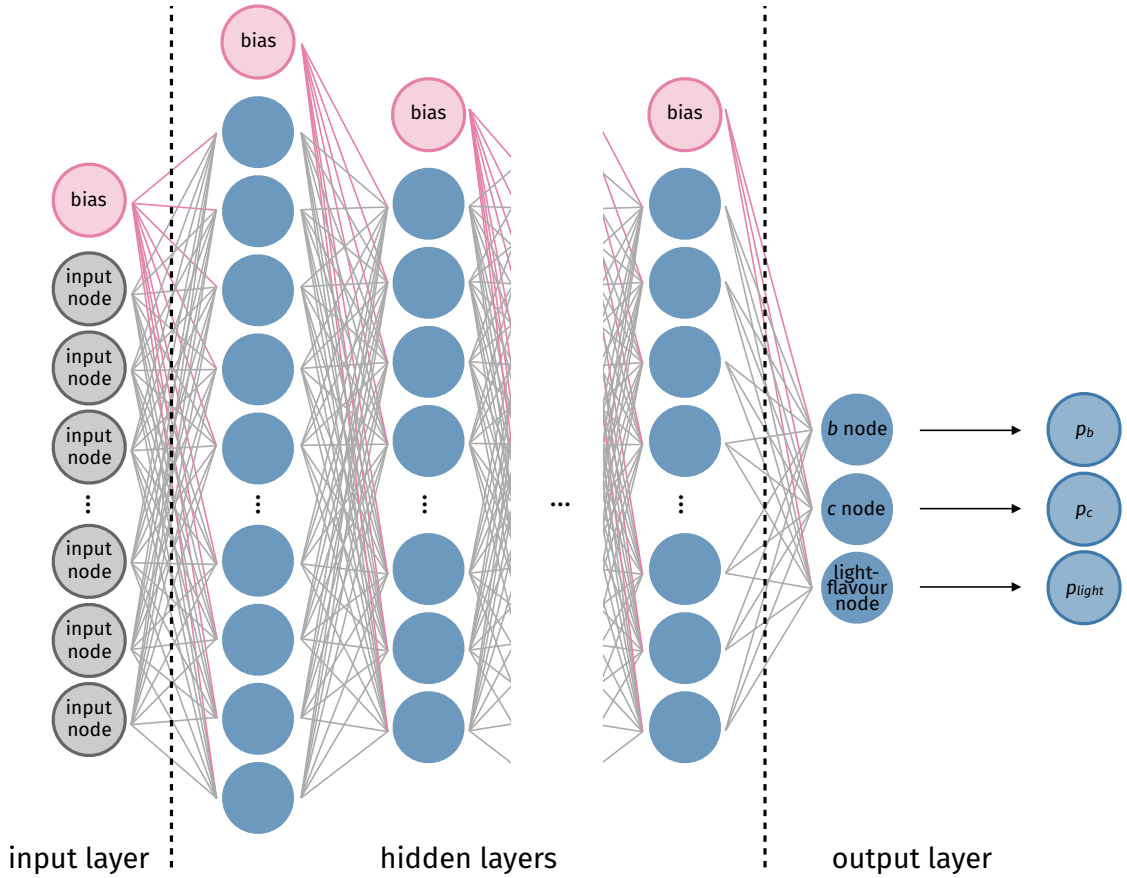
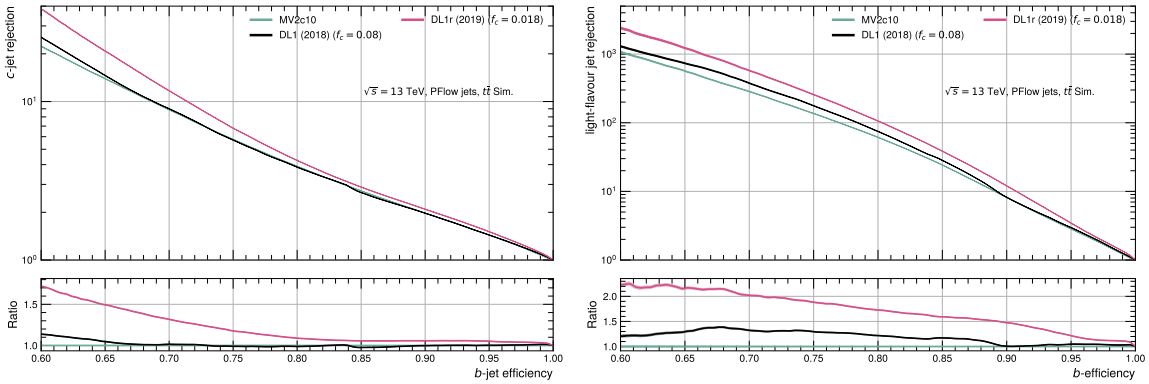


Figure C.1.: Structure du réseau de neurones de l'outil DL1.

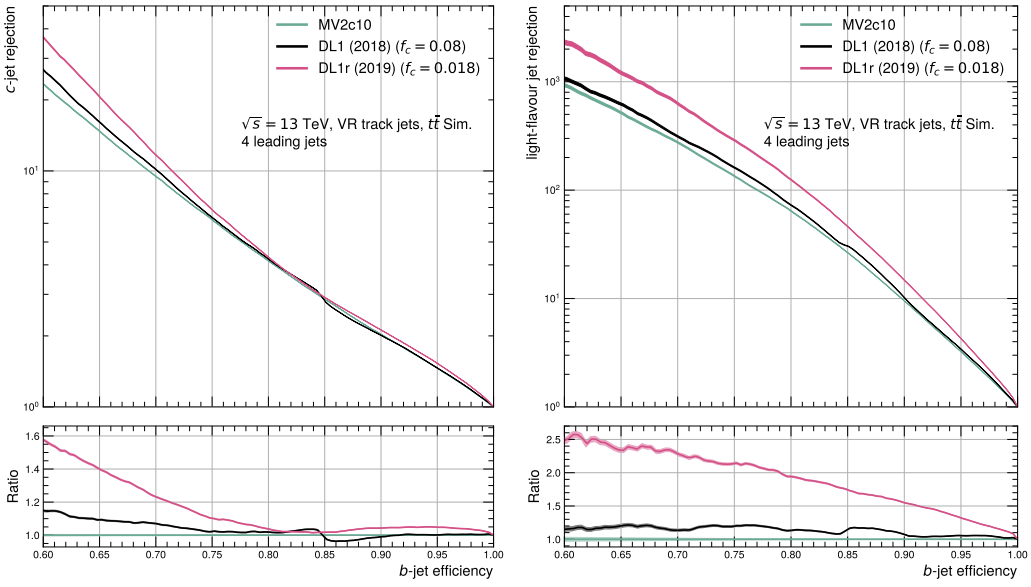
Des améliorations significatives ont été obtenues, pouvant atteindre un facteur deux sur le rejet du bruit de fond pour certaines régions de l'espace de phase, comme le montre la Figure C.2 pour *PFlow* jets et la Figure C.3 pour *EMTopo* jets.
De plus, un travail d'optimisation des hyperparamètres a été développé à l'aide de GPU sur le grille de calcul pour l'utilisation dans la collaboration. En outre, une nouvelle idée d'étiquetage de saveurs lourdes a été présentée, combinant les informations sur les traces déposées dans le détectuer et les

informations sur les jets dans un entraînement complet. C'est une stratégie prometteuse d'étiquetage des saveurs qui pourra être utilisée éventuellement dans l'expérience ATLAS et doit encore faire l'objet d'études plus approfondies.



(a) (b)

Figure C.2.: Comparaison des performances des recommandations précédentes (MV2 et DL1 (2018)) et de l'outil *DL1r* nouvellement optimisé pour *PFlow* jets pour la rejection des jets de c (a) et des jet de saveur légères (*light-jets*) (b) en fonction de l'efficacité des jets de b.



(a) (b)

Figure C.3.: Comparaison des performances des recommandations précédentes (MV2 et DL1 (2018)) et de l'outil *DL1r* nouvellement optimisé pour *VR Track* jets pour la rejection des jets de c (a) et des jet de saveur légères (*light-jets*) (b) en fonction de l'efficacité des jets de b.

C.2 ANALYSE $t\bar{t}H(b\bar{b})$

Le mode de production du boson de Higgs en association avec une paire de quarks top ($t\bar{t}H$) a été récemment observé [8–10], marquant une découverte importante. La mesure de ce canal de production est directement sensible au couplage de Yukawa, décrivant l'interaction du quark top avec le boson de Higgs. Comme le couplage de Yukawa augmente proportionnellement à la masse du fermion, le couplage de Yukawa du quark top est le plus fort dans le MS.

Dans cette thèse, l'analyse $t\bar{t}H(b\bar{b})$ [5] est présentée, réalisée avec l'ensemble des données prises durant le RUN II correspondant à $139\,fb^{-1}$ enregistrées avec l'expérience ATLAS à une énergie dans le centre de masse de $\sqrt{s} = 13\,TeV$.
Même si le Modèle Standard est testée avec une grande précision, des observations astrophysiques prouvent qu'il y a des phénomènes physiques au-delà du Modèle Standard. La mesure des sections efficaces différentielle du processus $t\bar{t}H$ dans le cadre des mesures différentielles "Simplified Template Cross-Section" (STXS) [106, 107] est sensible à de tels effets. Ces mesures permettent d'accéder à la structure CP du boson de Higgs et de sonder les auto-couplages de boson de Higgs anormaux conduisant à une sensibilité élevée de la mesure.

Le canal analysé dans cette thèse est le canal lepton+jets dans le régime résolu (*resolved*) comprenant au moins six jets dans l'état final, dont quatre au moins sont des jets de b. Dans ce canal, un boson W provenant de la désintégration des paires de quark top doit se désintégrer en leptons. Les partons de l'état final sont appariés aux jets dans l'événement en utilisant des arbres de décision boostés (BDT). Dans cette analyse, le bruit de fond dominant provient des événements $t\bar{t}$ + jets, et en particulier des événements $t\bar{t} + \geqslant 1b$. Pour le résultat final, ce canal est combiné avec le canal lepton+jets optimisé pour $p_T^H > 300\,GeV$ et le canal dilepton. L'intensité du signal, qui est le rapport entre la section mesurée et la section prévue dans le Modèle Standard, est présenté pour les différents canaux dans la Figure C.4. La mesure inclusive de la section transversale a pour valeur:

$$\mu_{\text{incl.}} = 0.43^{+0.20}_{-0.19}\,(\text{stat.})\,^{+0.30}_{-0.27}\,(\text{syst.}),$$

avec une signification observée (attendue) de $1,3$ $(3,0)$ déviations standard. Le facteur de normalisation du bruit de fond $t\bar{t} + \geqslant 1b$ est flottant dans l'ajustement et sa valeur mesurée est de $k(t\bar{t} + \geqslant 1b) = 1.26 \pm 0.09$. La mesure est principalement dominée par les incertitudes systématiques. En particulier, les incertitudes associées à la modélisation du bruit de fond $t\bar{t} + \geqslant 1b$ ont un impact le plus important sur l'incertitude globale.

De plus, une mesure différentielle de la section efficace en fonction de l'impulsion transverse du boson de Higgs pour le processus $t\bar{t}H$ a été effectuée, pour la première fois dans le canal $t\bar{t}H(b\bar{b})$. Cette mesure a été effectuée dans le cadre des mesures STXS qui permet une comparaison plus facile entre les différents canaux et entre les expériences. L'intensité du signal a été mesurée dans les cinq

régions (*bins*): $0\ \text{GeV} \leqslant p_T^H < 120\ \text{GeV}$, $120\ \text{GeV} \leqslant p_T^H < 200\ \text{GeV}$, $200\ \text{GeV} \leqslant p_T^H < 300\ \text{GeV}$, $300\ \text{GeV} \leqslant p_T^H < 450\ \text{GeV}$, et $p_T^H \geqslant 450\ \text{GeV}$. Les résultats de cette mesure sont présentés sur la Figure C.5. Les intensités de signal associées aux premiers bins de p_T^H sont dominées par les incertitudes systématiques tandis que les autres paramètres d'intensité de signal sont limités par les incertitude statistiques. En général, les incertitudes sont assez grandes et les intensités du signal mesuré à partir des sections efficaces différentielles sont en accord avec l'intensité du signal mesuré inclusivement.

Les incertitudes systématiques dans ces mesures ont été reduites par rapport à l'analyse précédente d'à peu près un facteur deux, grâce aux modifications apportés à la modélisation du bruit de fond $t\bar{t} + \geqslant 1b$ et à l'estimation des incertitudes associées.

L'analyse peut encore être améliorée en utilisant d'autres techniques multivariées. En particulier, la reconstruction grâce à des BDT pourrait être remplacée par un réseau de neurones pour améliorer la reconstruction des événements en utilisant des architectures personnalisées.

Finalement, les améliorations réalisées dans cette thèse concernant l'étiquetage des jets de quark b amélioreront davantage les mesures futures.

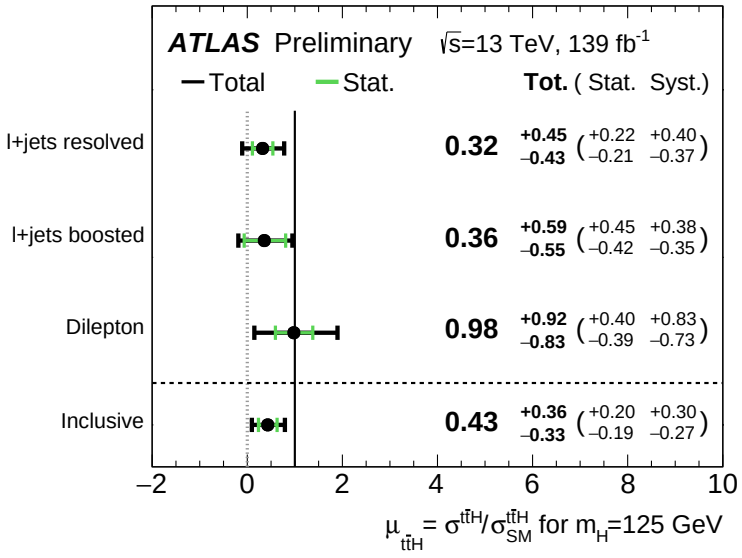


Figure C.4.: Valeurs ajustées du paramètre d'intensité du signal $t\bar{t}H$ dans les canaux individuels et pour la mesure inclusive. Les résultats des trois canaux individuels sont mesur par un ajustement 3-μ et le scénario inclusif est l'ajustement combiné nominal [5].

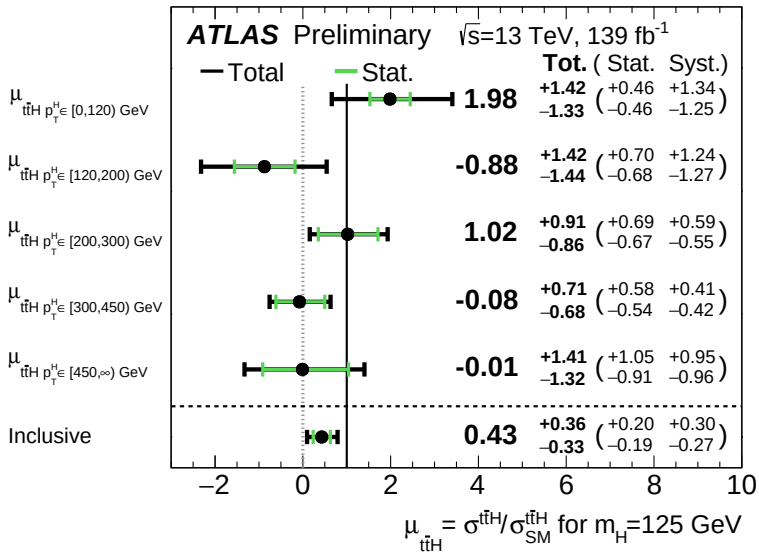


Figure C.5.: Mesures de la force du signal dans les différents bins STXS p_T^H, ainsi que la force du signal inclusive [5]. Les deux résultats sont obtenus par un ajustement combiné des trois canaux.

D

ACRONYMS

CERN Organisation européenne pour la recherche nucléaire

SM Standard Model

QFT Quantum Field Theory

QCD Quantum Chromodynamics

QED Quantum Electrodynamics

CKM Cabibbo-Kobayashi-Maskawa

EWSB electroweak symmetry breaking

VEV vacuum expectation value

LHC Large Hadron Collider

HL-LHC High Luminosity LHC

ID Inner Detector

SCT semiconductor tracker

TRT transition radiation tracker

ITk Inner Tracker

IBL insertable b-layer

EM electromagnetic

LAr liquid argon

MS muon spectrometer

RPCs resistive plate chambers

TGCs thin gap chambers

MDTs monitored drift tubes

CSCs cathod strip chambers

HLT high level trigger

RoI region of interest

L1 Level-1

PDF Parton Distribution Function

DGLAP Dokshitzer–Gribov–Lipatov–Altarelli–Parisi

MC Monte Carlo

MPI multi-parton interaction

PS parton shower

ME matrix element

ISR initial state radiation

FSR final state radiation

4FS four-flavour scheme

5FS five-flavour scheme

NLO next-to-leading order

PV primary vertex

JVT jet vertex tagger

ML Machine Learning

MLE Maximum Likelihood Estimation

LLR Log-likelihood ratio

BDT Boosted Decision Tree

NN Neural Network

ReLU Rectified Linear Unit

AdaBoost Adaptive Boost

DL1 Deep Learning based heavy-flavour tagger

WP working point

VR variable radius

IP impact parameter

SV secondary vertex

SV1 inclusive displaced secondary vertex reconstruction algorithm

JF JETFITTER

SMT Soft Muon Tagger

DIPS Deep Impact Parameter Sets

SR signal region

CR control region

STXS Simplified Template Cross-Section

WLCG Worldwide LHC Computing Grid

ACKNOWLEDGEMENTS

Preparing and writing a PhD thesis is a big project which would not be possible without the support of several people. Especially, in an international joint PhD (cotutelle) the coordination and team work is essential.
First I want to thank Prof. Dr. Gregor Herten for giving me the chance to do my PhD following my Bachelor's and Master's thesis in his group. I appreciate a lot the big support, also during the lengthy administrative challenges for the cotutelle agreement. I enjoyed a lot working in the Freiburg group with all the members and in particular with the other PhD students Fabio, Manfredi, Patrick, Peter, Shalu, Simone, Thor and Veronika.
I want to express a big thank you to you, Andrea, for the day to day supervision. You were always there if I had any kind of question. It was a pleasure working with you and to have some non-physics beers especially in workshops and conferences.
And also a big thanks to my two french supervisors Henri and Frédéric for always supporting me and answering my questions. Without you Fred, I would have not beeen able to manage all this french bureaucratic jungle.
It was with great joy working in the FTAG group with lots of great people. A special thanks to Chris for always supporting the young students and to Nicole and Dan for the nice discussions.
Also a thank you to the $\mathrm{t\bar{t}H(b\bar{b})}$ group, for the nice team work finally publishing the analysis.

During my three years in my PhD I spent one year in Freiburg, CERN and Paris. It was great to meet lots of nice colleagues not only to discuss physics but also for having some beers together, thanks Aishik, Brian, Mykola, Lukas, Aaron, Julia, Frank, Jan and Franky.
I want to thank all my non-physics friends for also distracting me from the work from time to time and helping to free up my mind, especially Fabian, Marina, Jessi, Johannes and Alicia.
A special thanks goes to my whole family who always gave me an incredible support and a safe heaven also during the time of the pandemic.
Last but not least I want to thank a special person, Konie, for supporting me and backing me during these years.

REFERENCES

[1] ATLAS Collaboration. *Expected performance of the 2019 ATLAS* b*-taggers*. URL: http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2019-005/ (visited on 27/01/2021).

[2] ATLAS Collaboration. *Performance of 2019 recommendations of ATLAS Flavor Tagging algorithms with Variable Radius track jets*. URL: http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2019-006/ (visited on 27/01/2021).

[3] Manuel Guth. https://gitlab.cern.ch/atlas-flavor-tagging-tools/dl1-hyperparameter-optimisation.

[4] ATLAS Collaboration. *Hyper Parameter Scan with the Deep Learning Heavy Flavour Tagger (DL1)*. URL: http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2019-001/ (visited on 27/01/2021).

[5] ATLAS Collaboration. *Measurement of the Higgs boson decaying to* b*-quarks produced in association with a top-quark pair in* pp *collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*. Tech. rep. ATLAS-CONF-2020-058. Geneva: CERN, 2020. URL: https://cds.cern.ch/record/2743685.

[6] 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.* B716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[7] CMS Collaboration. 'Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC'. In: *Phys. Lett.* B716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[8] CMS Collaboration. 'Observation of $t\bar{t}H$ Production'. In: *Phys. Rev. Lett.* 120 (2018), p. 231801. DOI: 10.1103/PhysRevLett.120.231801. arXiv: 1804.02610 [hep-ex].

[9] ATLAS and CMS Collaborations. 'Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV'. In: *JHEP* 08 (2016), p. 045. DOI: 10.1007/JHEP08(2016)045. arXiv: 1606.02266 [hep-ex].

[10] 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. DOI: https://doi.org/10.1016/j.physletb.2018.07.035.

[11] CMS Collaboration. 'Observation of Higgs Boson Decay to Bottom Quarks'. In: *Phys. Rev. Lett.* 121 (2018), p. 121801. DOI: 10.1103/PhysRevLett.121.121801. arXiv: 1808.08242 [hep-ex].

[12] ATLAS Collaboration. 'Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector'. In: *Phys. Lett. B* 786 (2018), p. 59. DOI: 10.1016/j.physletb.2018.09.013. arXiv: 1808.08238 [hep-ex].

[13] ATLAS Collaboration. 'Search for the standard model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector'. In: *Phys. Rev. D* 97 (2018), p. 072016. DOI: 10.1103/PhysRevD.97.072016. arXiv: 1712.08895 [hep-ex].

[14] CMS Collaboration. 'Search for $t\bar{t}H$ production in the $H \to b\bar{b}$ decay channel with leptonic $t\bar{t}$ decays in proton-proton collisions at $\sqrt{s} = 13$ TeV'. In: *JHEP* 03 (2019), p. 026. DOI: 10.1007/JHEP03(2019)026. arXiv: 1804.03682 [hep-ex].

[15] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. Reading, USA: Addison-Wesley, 1995. ISBN: 978-0-201-50397-5.

[16] C.P. Burgess and G.D. Moore. *The standard model: A primer*. Cambridge University Press, Dec. 2006. ISBN: 978-0-511-25485-7, 978-1-107-40426-7, 978-0-521-86036-9.

[17] W. Hollik. 'Quantum field theory and the Standard Model'. In: *2009 European School of High-Energy Physics*. Dec. 2010. arXiv: 1012.3883 [hep-ph].

[18] CERN Bulletin (2012). *GO ON A PARTICLE QUEST AT THE FIRST CERN WEBFEST*. URL: http://cds.cern.ch/journal/CERNBulletin/2012/35/News%20Articles/1473657 (visited on 16/09/2020).

[19] ATLAS Collaboration. 'Measurement of the top quark mass in the $t\bar{t} \to$ lepton+jets channel from $\sqrt{s} = 8$ TeV ATLAS data and combination with previous results'. In: *Eur. Phys. J. C* 79 (2019), p. 290. DOI: 10.1140/epjc/s10052-019-6757-9. arXiv: 1810.01772 [hep-ex].

[20] M. Tanabashi et al. 'Review of Particle Physics'. In: *Phys. Rev. D* 98 (3 2018), p. 030001. DOI: 10.1103/PhysRevD.98.030001.

[21] H. David Politzer. 'Reliable Perturbative Results for Strong Interactions?' In: *Phys. Rev. Lett.* 30 (26 1973), pp. 1346–1349. DOI: 10.1103/PhysRevLett.30.1346.

[22] Roel Aaij et al. '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. DOI: 10.1103/PhysRevLett.115.072001. arXiv: 1507.03414 [hep-ex].

[23] Murray Gell-Mann. 'Symmetries of Baryons and Mesons'. In: *Phys. Rev.* 125 (3 1962), pp. 1067–1084. DOI: 10.1103/PhysRev.125.1067.

[24] Gavin P. Salam. 'Elements of QCD for hadron colliders'. In: *2009 European School of High-Energy Physics*. Nov. 2010. arXiv: 1011.5131 [hep-ph].

[25] David J. Gross and Frank Wilczek. 'Asymptotically Free Gauge Theories. I'. In: *Phys. Rev. D* 8 (10 1973), pp. 3633–3652. DOI: 10.1103/PhysRevD.8.3633. URL: https://link.aps.org/doi/10.1103/PhysRevD.8.3633.

[26] Sheldon L. Glashow. 'Partial-symmetries of weak interactions'. In: *Nuclear Physics* 22.4 (1961), pp. 579 –588. ISSN: 0029-5582. DOI: http://dx.doi.org/10.1016/0029-5582(61)90469-2.

[27] Abdus Salam. 'Gauge unification of fundamental forces'. In: *Rev. Mod. Phys.* 52 (3 1980), pp. 525–538. DOI: 10.1103/RevModPhys.52.525.

[28] Steven Weinberg. 'A Model of Leptons'. In: *Phys. Rev. Lett.* 19 (21 1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.

[29] Tadao Nakano and Kazuhiko Nishijima. 'Charge Independence for V-particles*'. In: *Progress of Theoretical Physics* 10.5 (1953), pp. 581–582. DOI: 10.1143/PTP.10.581.

[30] F. Englert and R. Brout. 'Broken Symmetry and the Mass of Gauge Vector Mesons'. In: *Phys. Rev. Lett.* 13 (9 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321.

[31] P.W. Higgs. 'Broken symmetries, massless particles and gauge fields'. In: *Physics Letters* 12.2 (1964), pp. 132 –133. ISSN: 0031-9163. DOI: https://doi.org/10.1016/0031-9163(64)91136-9.

[32] Peter W. Higgs. 'Broken Symmetries and the Masses of Gauge Bosons'. In: *Phys. Rev. Lett.* 13 (16 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508.

[33] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble. 'Global Conservation Laws and Massless Particles'. In: *Phys. Rev. Lett.* 13 (20 1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585.

[34] Peter W. Higgs. 'Spontaneous Symmetry Breakdown without Massless Bosons'. In: *Phys. Rev.* 145 (4 1966), pp. 1156–1163. DOI: 10.1103/PhysRev.145.1156.

[35] T. W. B. Kibble. 'Symmetry Breaking in Non-Abelian Gauge Theories'. In: *Phys. Rev.* 155 (5 1967), pp. 1554–1561. DOI: 10.1103/PhysRev.155.1554.

[36] F. Sauerburger. *LHC cross-section plot*. URL: https://lhc-xsecs.org/ (visited on 16/09/2020).

[37] Virginia Trimble. 'Existence and Nature of Dark Matter in the Universe'. In: *Annual Review of Astronomy and Astrophysics* 25.1 (1987), pp. 425–472. DOI: 10.1146/annurev.aa.25.090187.002233. eprint: https://doi.org/10.1146/annurev.aa.25.090187.002233.

[38] Gerson Goldhaber. 'The Acceleration of the Expansion of the Universe: A Brief Early History of the Supernova Cosmology Project (SCP)'. In: *AIP Conference Proceedings* 1166.1 (2009), pp. 53–72. DOI: 10.1063/1.3232196.

[39] Super-Kamiokande Collaboration. 'Evidence for Oscillation of Atmospheric Neutrinos'. In: *Phys. Rev. Lett.* 81 (8 1998), pp. 1562–1567. DOI: 10.1103/PhysRevLett.81.1562.

[40] SNO Collaboration. 'Measurement of the Rate of $\nu_e + d \rightarrow p + p + e^-$ Interactions Produced by ^{8}B Solar Neutrinos at the Sudbury Neutrino Observatory'. In: *Phys. Rev. Lett.* 87 (7 2001), p. 071301. DOI: 10.1103/PhysRevLett.87.071301. URL: https://link.aps.org/doi/10.1103/PhysRevLett.87.071301.

[41] ATLAS Collaboration. *Standard Model Summary Plots Spring 2020*. Tech. rep. ATL-PHYS-PUB-2020-010. Geneva: CERN, 2020. URL: http://cds.cern.ch/record/2718937.

[42] Lyndon Evans and Philip Bryant. 'LHC Machine'. In: *Journal of Instrumentation* 3.08 (2008), S08001–S08001. DOI: 10.1088/1748-0221/3/08/s08001.

[43] O. Brüning, H. Burkhardt and S. Myers. 'The large hadron collider'. In: *Progress in Particle and Nuclear Physics* 67.3 (2012), pp. 705 –734. ISSN: 0146-6410. DOI: https://doi.org/10.1016/j.ppnp.2012.03.001.

[44] Brüning et al. *LHC Design Report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2004. DOI: 10.5170/CERN-2004-003-V-1.

[45] J. W. G. Thomason, R. Garoby, S. Gilardoni, L. J. Jenner and J. Pasternak. 'Proton driver scenarios at CERN and Rutherford Appleton Laboratory'. In: *Phys. Rev. ST Accel. Beams* 16 (5 2013), p. 054801. DOI: 10.1103/PhysRevSTAB.16.054801.

[46] The ATLAS Collaboration. 'The ATLAS Experiment at the CERN Large Hadron Collider'. In: *Journal of Instrumentation* 3.08 (2008), S08003. DOI: 10.1088/1748-0221/3/08/s08003.

[47] The CMS Collaboration. 'The CMS experiment at the CERN LHC'. In: *Journal of Instrumentation* 3.08 (2008), S08004. DOI: 10.1088/1748-0221/3/08/s08004.

[48] *PEOPLE STATISTICS*. URL: https://cms.cern/collaboration/people-statistics (visited on 06/10/2020).

[49] *The Collaboration*. URL: https://atlas.cern/discover/collaboration (visited on 06/10/2020).

[50] The LHCb Collaboration. 'The LHCb Detector at the LHC'. In: *Journal of Instrumentation* 3.08 (2008), S08005. DOI: 10.1088/1748-0221/3/08/s08005.

[51] The ALICE Collaboration. 'The ALICE experiment at the CERN LHC'. In: *Journal of Instrumentation* 3.08 (2008), S08002. DOI: 10.1088/1748-0221/3/08/s08002.

[52] ATLAS Collaboration. *Luminosity Public Results for Run-2*. URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2 (visited on 16/09/2020).

[53] Joao Pequenao. 'Computer generated image of the whole ATLAS detector'. https://cds.cern.ch/record/1095924. 2008.

[54] *ATLAS inner detector: Technical Design Report, 1*. Technical Design Report ATLAS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/331063.

[55] 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.

[56] 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].

[57] M Capeans, G Darbo, K Einsweiller, M Elsing, T Flick, M Garcia-Sciveres, C Gemme, H Pernegger, O Rohne and R Vuillermet. *ATLAS Insertable B-Layer Technical Design Report*. Tech. rep. CERN-LHCC-2010-013. ATLAS-TDR-19. 2010. URL: https://cds.cern.ch/record/1291633.

[58] ATLAS Collaboration. *Impact Parameter Resolution*. URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/IDTR-2015-007/ (visited on 17/01/2021).

[59] ATLAS Collaboration. 'Performance of the ATLAS Transition Radiation Tracker in Run 1 of the LHC: tracker properties'. In: *JINST* 12 (2017), P05002. DOI: 10.1088/1748-0221/12/05/P05002. arXiv: 1702.06473 [hep-ex].

[60] *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.

[61] *ATLAS liquid-argon calorimeter: Technical Design Report*. Technical Design Report ATLAS. Geneva: CERN, 1996. URL: https://cds.cern.ch/record/331061.

[62] *Technical Design Report for the Phase-II Upgrade of the ATLAS LAr Calorimeter*. Tech. rep. CERN-LHCC-2017-018. ATLAS-TDR-027. Geneva: CERN, 2017. URL: http://cds.cern.ch/record/2285582.

[63] *ATLAS tile calorimeter: Technical Design Report*. Technical Design Report ATLAS. Geneva: CERN, 1996. URL: https://cds.cern.ch/record/331062.

[64] A Artamonov et al. 'The ATLAS Forward Calorimeter'. In: *Journal of Instrumentation* 3.02 (2008), P02010. DOI: 10.1088/1748-0221/3/02/p02010. URL: https://doi.org/10.1088%2F1748-0221%2F3%2F02%2Fp02010.

[65] *Plots of b-tagging performance before and after the installation of the Insertable B-Layer*. URL: https://hep.uchicago.edu/images/Calorimeters_labels.png (visited on 16/10/2020).

[66] ATLAS Collaboration. 'Operation and performance of the ATLAS Tile Calorimeter in Run 1'. In: *Eur. Phys. J. C* 78.12 (2018), p. 987. DOI: 10.1140/epjc/s10052-018-6374-z. arXiv: 1806.02129 [hep-ex].

[67] N Ilic for the benefit of the ATLAS Collaboration. 'Performance of the ATLAS Liquid Argon Calorimeter after three years of LHC operation and plans for a future upgrade'. In: *Journal of Instrumentation* 9.03 (2014), p. C03049. DOI: 10.1088/1748-0221/9/03/c03049. URL: https://doi.org/10.1088%2F1748-0221%2F9%2F03%2Fc03049.

[68] *ATLAS muon spectrometer: Technical Design Report*. Technical Design Report ATLAS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/331068.

[69] E. Diehl. 'ATLAS Muon Detector Commissioning'. In: *Meeting of the Division of Particles and Fields of the American Physical Society (DPF 2009)*. 2009. arXiv: 0910.2767.

[70] Peter Jenni, Marzio Nessi, Markus Nordberg and Kenway Smith. *ATLAS high-level trigger, data-acquisition and controls: Technical Design Report*. Technical Design Report ATLAS. Geneva: CERN, 2003. URL: https://cds.cern.ch/record/616089.

[71] ATLAS Collaboration. *The Run-2 ATLAS Trigger System*. Tech. rep. ATL-DAQ-PROC-2016-003. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2133909.

[72] J C Collins and D E Soper. 'The Theorems of Perturbative QCD'. In: *Annual Review of Nuclear and Particle Science* 37.1 (1987), pp. 383–409. DOI: 10.1146/annurev.ns.37.120187.002123.

[73] Manjunath Bhat, Krzysztof Cichy, Martha Constantinou and Aurora Scapellato. 'Parton distribution functions from lattice QCD at physical quark masses via the pseudo-distribution approach'. In: (May 2020). arXiv: 2005.02102 [hep-lat].

[74] Sayipjamal Dulat, Tie-Jiun Hou, Jun Gao, Marco Guzzi, Joey Huston, Pavel Nadolsky, Jon Pumplin, Carl Schmidt, Daniel Stump and C.-P. Yuan. 'New parton distribution functions from a global analysis of quantum chromodynamics'. In: *Phys. Rev. D* 93 (3 2016), p. 033006. DOI: 10.1103/PhysRevD.93.033006.

[75] 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.

[76] A. D. Martin, W. J. Stirling, R. S. Thorne and G. Watt. 'Parton distributions for the LHC'. In: *Eur. Phys. J. C* 63 (2009), pp. 189–285. DOI: 10.1140/epjc/s10052-009-1072-5. arXiv: 0901.0002 [hep-ph].

[77] Richard D. Ball et al. 'Parton distributions for the LHC run II'. In: *JHEP* 04 (2015), p. 040. DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ph].

[78] 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.

[79] G. Altarelli and G. Parisi. 'Asymptotic freedom in parton language'. In: *Nuclear Physics B* 126.2 (1977), pp. 298 –318. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(77)90384-4.

[80] 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.

[81] Bo Andersson, G. Gustafson, G. Ingelman and T. Sjöstrand. 'Parton fragmentation and string dynamics'. In: *Phys. Rept.* 97 (1983), pp. 31–145. DOI: 10.1016/0370-1573(83)90080-7.

[82] Jan-Christopher Winter, Frank Krauss and Gerhard Soff. 'A modified cluster-hadronization model'. In: *Eur. Phys. J. C* 36 (2004), pp. 381–395. DOI: 10.1140/epjc/s2004-01960-8. arXiv: hep-ph/0311085.

[83] Stefan Höche. 'Introduction to parton-shower event generators'. In: *Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders*. 2015, pp. 235–295. DOI: 10.1142/9789814678766_0005. arXiv: 1411.4085 [hep-ph].

[84] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen and Peter Z. Skands. 'An introduction to PYTHIA 8.2'. In: *Comput. Phys. Commun.* 191 (2015), p. 159. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].

[85] M. Bähr et al. 'Herwig++ physics and manual'. In: *Eur. Phys. J. C* 58 (2008), p. 639. DOI: 10.1140/epjc/s10052-008-0798-9. arXiv: 0803.0883 [hep-ph].

[86] Johannes Bellm et al. 'Herwig 7.0/Herwig++ 3.0 release note'. In: *Eur. Phys. J. C* 76.4 (2016), p. 196. DOI: 10.1140/epjc/s10052-016-4018-8. arXiv: 1512.01178 [hep-ph].

[87] Enrico Bothmann et al. 'Event generation with Sherpa 2.2'. In: *SciPost Phys.* 7.3 (2019), p. 034. DOI: 10.21468/SciPostPhys.7.3.034. arXiv: 1905.09127 [hep-ph].

[88] Paolo Nason. 'A new method for combining NLO QCD with shower Monte Carlo algorithms'. In: *JHEP* 11 (2004), p. 040. DOI: 10.1088/1126-6708/2004/11/040. arXiv: hep-ph/0409146.

[89] Stefano Frixione, Paolo Nason and Carlo Oleari. 'Matching NLO QCD computations with parton shower simulations: the POWHEG method'. In: *JHEP* 11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].

[90] 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. DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581 [hep-ph].

[91] 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. DOI: 10.1103/PhysRevD.91.094003. arXiv: 1501.04498 [hep-ph].

[92] Stefano Frixione, Paolo Nason and Giovanni Ridolfi. 'A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction'. In: *JHEP* 09 (2007), p. 126. DOI: 10.1088/1126-6708/2007/09/126. arXiv: 0707.3088 [hep-ph].

[93] Johan Alwall, Pavel Demin, Simon de Visscher, Rikkert Frederix, Michel Herquet, Fabio Maltoni, Tilman Plehn, David L. Rainwater and Tim Stelzer. 'MadGraph/MadEvent v4: The New Web Generation'. In: *JHEP* 09 (2007), p. 028. DOI: 10.1088/1126-6708/2007/09/028. arXiv: 0706.2334 [hep-ph].

[94] ATLAS Collaboration. *ATLAS Pythia 8 tunes to 7 TeV data*. ATL-PHYS-PUB-2014-021. 2014. URL: https://cds.cern.ch/record/1966419.

[95] D. J. Lange. 'The EvtGen particle decay simulation package'. In: *Nucl. Instrum. Meth. A* 462 (2001), p. 152. DOI: 10.1016/S0168-9002(01)00089-4.

[96] Richard D. Ball et al. 'Parton distributions with LHC data'. In: *Nucl. Phys. B* 867 (2013), p. 244. DOI: 10.1016/j.nuclphysb.2012.10.003. arXiv: 1207.1303 [hep-ph].

[97] L.A. Harland-Lang, A.D. Martin, P. Motylinski and R.S. Thorne. 'Parton distributions in the LHC era: MMHT 2014 PDFs'. In: *Eur. Phys. J. C* 75.5 (2015), p. 204. DOI: 10.1140/epjc/s10052-015-3397-6. arXiv: 1412.3989 [hep-ph].

[98] ATLAS Collaboration. 'The ATLAS Simulation Infrastructure'. In: *Eur. Phys. J. C* 70 (2010), p. 823. DOI: 10.1140/epjc/s10052-010-1429-9. arXiv: 1005.4568.

[99] S. Agostinelli et al. 'GEANT4 – a simulation toolkit'. In: *Nucl. Instrum. Meth. A* 506 (2003), p. 250. DOI: 10.1016/S0168-9002(03)01368-8.

[100] ATLAS Collaboration. *The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim*. Tech. rep. ATL-PHYS-PUB-2010-013. Geneva: CERN, 2010. URL: https://cds.cern.ch/record/1300517.

[101] Flavia Dias. 'The new ATLAS Fast Calorimeter Simulation'. In: *PoS* ICHEP2016 (2016), p. 184. DOI: 10.22323/1.282.0184.

[102] Jana Schaarschmidt. 'The new ATLAS Fast Calorimeter Simulation'. In: *J. Phys. Conf. Ser.* 898.4 (2017). Ed. by Richard Mount and Craig Tull, p. 042006. DOI: 10.1088/1742-6596/898/4/042006.

[103] Aishik Ghosh on behalf of the ATLAS Collaboration. 'Deep generative models for fast shower simulation in ATLAS'. In: *Journal of Physics: Conference Series* 1525 (2020), p. 012077. DOI: 10.1088/1742-6596/1525/1/012077. URL: https://doi.org/10.1088%2F1742-6596%2F1525%2F1%2F012077.

[104] 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.

[105] Frank Fiedler et al. 'The Matrix Element Method and its Application in Measurements of the Top Quark Mass'. In: *Nucl. Instrum. Meth. A* 624 (2010), pp. 203–218. DOI: 10.1016/j.nima.2010.09.024. arXiv: 1003.1316 [hep-ex].

[106] Nicolas Berger et al. 'Simplified Template Cross Sections - Stage 1.1'. In: (June 2019). arXiv: 1906.02754 [hep-ph].

[107] D. de Florian et al. 'Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector'. In: (2016). DOI: 10.23731/CYRM-2017-002. arXiv: 1610.07922 [hep-ph].

[108] J. R. Andersen et al. 'Les Houches 2015: Physics at TeV Colliders Standard Model Working Group Report'. In: (2016). arXiv: 1605.04692 [hep-ph]. URL: http://lss.fnal.gov/archive/2016/conf/fermilab-conf-16-175-ppd-t.pdf.

[109] Fawzi Boudjema, Rohini M. Godbole, Diego Guadagnoli and Kirtimaan A. Mohan. 'Lab-frame observables for probing the top-Higgs interaction'. In: *Phys. Rev. D* 92.1 (2015), p. 015019. DOI: 10.1103/PhysRevD.92.015019. arXiv: 1501.03157 [hep-ph].

[110] Fabio Maltoni, Davide Pagani, Ambresh Shivaji and Xiaoran Zhao. 'Trilinear Higgs coupling determination via single-Higgs differential measurements at the LHC'. In: *Eur. Phys. J. C* 77.12 (2017), p. 887. DOI: 10.1140/epjc/s10052-017-5410-8. arXiv: 1709.08649 [hep-ph].

[111] ATLAS Collaboration. *Measurement of Higgs boson production in association with a* $\mathrm{t\bar{t}}$ *pair in the diphoton decay channel using* $139\,fb^{-1}$ *of LHC data collected at* $\sqrt{s} = 13\,TeV$ *by the ATLAS experiment*. ATLAS-CONF-2019-004. 2019. URL: https://cds.cern.ch/record/2668103.

[112] *How ATLAS detects particles: diagram of particle paths in the detector*. 2013. URL: https://cds.cern.ch/record/1505342 (visited on 26/09/2020).

[113] 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. DOI: 10.1140/epjc/s10052-017-5225-7. arXiv: 1704.07983 [hep-ex].

[114] 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. DOI: 10.1140/epjc/s10052-017-4887-5. arXiv: 1611.10235 [physics.ins-det].

[115] ATLAS Collaboration. 'Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1'. In: *Eur. Phys. J. C* 77 (2017), p. 490. DOI: 10.1140/epjc/s10052-017-5004-5. arXiv: 1603.02934 [hep-ex].

[116] Gavin P. Salam. 'Towards Jetography'. In: *Eur. Phys. J. C* 67 (2010), pp. 637–686. DOI: 10.1140/epjc/s10052-010-1314-6. arXiv: 0906.1833 [hep-ph].

[117] 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. DOI: 10.1088/1126-6708/2008/04/063. URL: https://doi.org/10.1088%2F1126-6708%2F2008%2F04%2F063.

[118] ATLAS Collaboration. 'Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector'. In: *Phys. Rev. D* 96 (2017), p. 072002. DOI: 10.1103/PhysRevD.96.072002. arXiv: 1703.09665 [hep-ex].

[119] Matteo Cacciari, Gavin P. Salam and Gregory Soyez. 'FastJet user manual'. In: *Eur. Phys. J. C* 72 (2012), p. 1896. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph].

[120] ATLAS Collaboration. 'Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector'. In: (2020). arXiv: 2007.02645 [hep-ex].

[121] ATLAS Collaboration. 'Jet reconstruction and performance using particle flow with the ATLAS Detector'. In: *Eur. Phys. J. C* 77.7 (2017), p. 466. DOI: 10.1140/epjc/s10052-017-5031-2. arXiv: 1703.10485 [hep-ex].

[122] David Krohn, Jesse Thaler and Lian-Tao Wang. 'Jets with variable R'. In: *JHEP* 06 (2009), p. 059. DOI: 10.1088/1126-6708/2009/06/059. arXiv: 0903.0392 [hep-ph].

[123] ATLAS Collaboration. *Variable Radius, Exclusive-k_T, and Center-of-Mass Subjet Reconstruction for Higgs($\to b\bar{b}$) Tagging in ATLAS*. ATL-PHYS-PUB-2017-010. 2017. URL: https://cds.cern.ch/record/2268678.

[124] ATLAS Collaboration. 'Identification of boosted Higgs bosons decaying into b-quark pairs with the ATLAS detector at 13 TeV'. In: *Eur. Phys. J. C* 79.10 (2019), p. 836. DOI: 10.1140/epjc/s10052-019-7335-x. arXiv: 1906.11005 [hep-ex].

[125] ATLAS Collaboration. 'Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV'. In: *JHEP* 08 (2018), p. 089. DOI: 10.1007/JHEP08(2018)089. arXiv: 1805.01845 [hep-ex].

[126] ATLAS Collaboration. *Electron and photon reconstruction and performance in ATLAS using a dynamical, topological cell clustering-based approach*. ATL-PHYS-PUB-2017-022. 2017. URL: https://cds.cern.ch/record/2298955.

[127] ATLAS Collaboration. 'Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data'. In: *JINST* 14 (2019), P12006. DOI: 10.1088/1748-0221/14/12/P12006. arXiv: 1908.00005 [hep-ex].

[128] 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 (2016), p. 292. DOI: 10.1140/epjc/s10052-016-4120-y. arXiv: 1603.05598 [hep-ex].

[129] ATLAS Collaboration. *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using* pp *collisions at* $\sqrt{s} = 13$ *TeV*. ATLAS-CONF-2017-029. 2017. URL: https://cds.cern.ch/record/2261772.

[130] ATLAS Collaboration. 'Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13$ TeV'. In: *Eur. Phys. J. C* 78 (2018), p. 903. DOI: 10.1140/epjc/s10052-018-6288-9. arXiv: 1802.08168 [hep-ex].

[131] Javier Duarte et al. 'Fast inference of deep neural networks in FPGAs for particle physics'. In: *JINST* 13.07 (2018), P07027. DOI: 10.1088/1748-0221/13/07/P07027. arXiv: 1804.06913 [physics.ins-det].

[132] Maciej Wielgosz, Matej Mertik, Andrzej Skoczeń and Ernesto De Matteis. 'The model of an anomaly detector for HiLumi LHC magnets based on Recurrent Neural Networks and adaptive quantization'. In: *Engineering Applications of Artificial Intelligence* 74 (2018), pp. 166–185. ISSN: 0952-1976. DOI: https://doi.org/10.1016/j.engappai.2018.06.012. URL: http://www.sciencedirect.com/science/article/pii/S095219761830143X.

[133] Sabrina Amrouche, Tobias Golling, Moritz Kiehn, Claudia Plant and Andreas Salzburger. 'Similarity hashing for charged particle tracking'. In: *IEEE International Conference on Big Data 2019*. 2019, pp. 1595–1600. DOI: 10.1109/BigData47090.2019.9006316.

[134] Aristeidis Tsaris et al. 'The HEP.TrkX Project: Deep Learning for Particle Tracking'. In: *Journal of Physics: Conference Series* 1085 (2018), p. 042023. DOI: 10.1088/1742-6596/1085/4/042023.

[135] 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. DOI: https://doi.org/10.1016/j.nuclphysbps.2011.03.142. URL: http://www.sciencedirect.com/science/article/pii/S092056321100212X.

[136] ATLAS Collaboration. *Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment*. ATL-PHYS-PUB-2019-033. 2019. URL: https://cds.cern.ch/record/2688062.

[137] 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.

[138] Jack Collins, Kiel Howe and Benjamin Nachman. 'Anomaly Detection for Resonant New Physics with Machine Learning'. In: *Phys. Rev. Lett.* 121 (24 2018), p. 241803. DOI: 10.1103/PhysRevLett.121.241803.

[139] 'Cross-Validation'. In: *Encyclopedia of Machine Learning*. Ed. by Claude Sammut and Geoffrey I. Webb. Boston, MA: Springer US, 2010, pp. 249–249. ISBN: 978-0-387-30164-8. DOI: 10.1007/978-0-387-30164-8_190. URL: https://doi.org/10.1007/978-0-387-30164-8_190.

[140] Walter Pitts Warren S. McCulloch. 'A logical calculus of the ideas immanent in nervous activity'. In: *The bulletin of mathematical biophysics 5 115–133* (1943). DOI: 10.1007/BF02478259.

[141] Stefan Wunsch. *Fermilab Keras Workshop*. https://github.com/stwunsch/fermilab_keras_workshop. 2017.

[142] IML WG. *A Living Review of Machine Learning for Particle Physics*. URL: https://iml-wg.github.io/HEPML-LivingReview/ (visited on 13/11/2020).

[143] Ian Goodfellow, Yoshua Bengio and Aaron Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.

[144] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: http://tensorflow.org/.

[145] François Chollet et al. *Keras*. 2015. URL: https://keras.io.

[146] Adam Paszke et al. 'PyTorch: An Imperative Style, High-Performance Deep Learning Library'. In: Curran Associates, Inc., 2019, pp. 8024–8035. URL: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[147] Junjie Bai, Fang Lu, Ke Zhang et al. *ONNX: Open Neural Network Exchange*. https://github.com/onnx/onnx. 2019.

[148] Daniel Hay Guest, Joshua Wyatt Smith, Michela Paganini, Michael Kagan, Marie Lanfermann, Attila Krasznahorkay and Daniel Edison Marley. *lwtnn/lwtnn*. Version v2.9. June 2019. DOI: 10.5281/zenodo.3249317. URL: https://doi.org/10.5281/zenodo.3249317.

[149] Diederik P Kingma and J Adam Ba. 'A method for stochastic optimization. arXiv 2014'. In: *arXiv:1412.6980* 434 (2019).

[150] David E. Rumelhart, Geoffrey E. Hinton and Ronald J. Williams. 'Learning Representations by Back-propagating Errors'. In: *Nature* 323.6088 (1986), pp. 533–536. DOI: 10.1038/323533a0. URL: http://www.nature.com/articles/323533a0.

[151] Vinod Nair and Geoffrey E. Hinton. 'Rectified Linear Units Improve Restricted Boltzmann Machines'. In: ICML'10. Haifa, Israel: Omnipress, 2010, 807–814. ISBN: 9781605589077.

[152] Andrew L. Maas, Awni Y. Hannun and Andrew Y. Ng. 'Rectifier nonlinearities improve neural network acoustic models'. In: *in ICML Workshop on Deep Learning for Audio, Speech and Language Processing*. 2013.

[153] Charles Dugas, Yoshua Bengio, François Bélisle, Claude Nadeau and René Garcia. 'Incorporating Second-Order Functional Knowledge for Better Option Pricing'. In: *Proceedings of the 13th International Conference on Neural Information Processing Systems*. NIPS'00. Denver, CO: MIT Press, 2000, 451–457.

[154] 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.

[155] 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.

[156] Rich Caruana, Steve Lawrence and Lee Giles. 'Overfitting in Neural Nets: Backpropagation, Conjugate Gradient, and Early Stopping'. In: *IN PROC. NEURAL INFORMATION PROCESSING SYSTEMS CONFERENCE*. 2000, pp. 402–408.

[157] 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. DOI: 10.1016/S0168-9002(97)00048-X.

[158] Andreas Hoecker et al. *TMVA - Toolkit for Multivariate Data Analysis*. 2007. arXiv: physics/0703039 [physics.data-an].

[159] ATLAS Collaboration. *Optimisation of the ATLAS* b*-tagging performance for the 2016 LHC Run*. ATL-PHYS-PUB-2016-012. 2016. URL: https://cds.cern.ch/record/2160731.

[160] Rosaria Silipo. *Ensemble Models: Bagging & Boosting*. URL: https://medium.com/analytics-vidhya/ensemble-models-bagging-boosting-c33706db0b0b (visited on 12/11/2020).

[161] Yoav Freund and Robert E. Schapire. 'A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting'. In: *J. Comput. Syst. Sci.* 55.1 (1997), pp. 119–139. DOI: 10.1006/jcss.1997.1504.

[162] Tianqi Chen and Carlos Guestrin. 'XGBoost: A Scalable Tree Boosting System'. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. San Francisco, California, USA: ACM, 2016, pp. 785–794. ISBN: 978-1-4503-4232-2. DOI: 10.1145/2939672.2939785. URL: http://doi.acm.org/10.1145/2939672.2939785.

[163] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye and Tie-Yan Liu. 'LightGBM: A Highly Efficient Gradient Boosting Decision Tree'. In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett. Vol. 30. Curran Associates, Inc., 2017, pp. 3146–3154. URL: https://proceedings.neurips.cc/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf.

[164] Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush and Andrey Gulin. 'CatBoost: unbiased boosting with categorical features'. In: *Advances in Neural Information Processing Systems*. Ed. by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi and R. Garnett. Vol. 31. Curran Associates, Inc., 2018, pp. 6638–6648. URL: \url{https://proceedings.neurips.cc/paper/2018/file/14491b756b3a51daac41c24863285549-Paper.pdf}.

[165] ATLAS collaboration. 'Measurement of the associated production of a Higgs boson decaying into b-quarks with a vector boson at high transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector'. In: (Aug. 2020). arXiv: 2008.02508 [hep-ex].

[166] ATLAS Collaboration. 'Search for the Decay of the Higgs Boson to Charm Quarks with the ATLAS Experiment'. In: *Phys. Rev. Lett.* 120 (2018), p. 211802. DOI: 10.1103/PhysRevLett.120.211802. arXiv: 1802.04329 [hep-ex].

[167] ATLAS Collaboration. *Comparison of Monte Carlo generator predictions for bottom and charm hadrons in the decays of top quarks and the fragmentation of high* p_T *jets*. ATL-PHYS-PUB-2014-008. 2014. URL: https://cds.cern.ch/record/1709132.

[168] ATLAS Collaboration. *Measurements of* b*-jet moments sensitive to* b*-quark fragmentation in* $t\bar{t}$ *events at the LHC with the ATLAS detector*. Tech. rep. ATLAS-CONF-2020-050. Geneva: CERN, 2020. URL: http://cds.cern.ch/record/2730444.

[169] ATLAS Collaboration. *Optimisation and performance studies of the ATLAS* b*-tagging algorithms for the 2017-18 LHC run*. ATL-PHYS-PUB-2017-013. 2017. URL: https://cds.cern.ch/record/2273281.

[170] ATLAS Collaboration. *Studies on top-quark Monte Carlo modelling for Top2016*. ATL-PHYS-PUB-2016-020. 2016. URL: https://cds.cern.ch/record/2216168.

[171] ATLAS Collaboration. *Tagging and suppression of pileup jets with the ATLAS detector*. ATLAS-CONF-2014-018. 2014. URL: https://cds.cern.ch/record/1700870.

[172] ATLAS Collaboration. *Identification of Jets Containing* b*-Hadrons with Recurrent Neural Networks at the ATLAS Experiment*. ATL-PHYS-PUB-2017-003. 2017. URL: https://cds.cern.ch/record/2255226.

[173] ATLAS Collaboration. *Secondary vertex finding for jet flavour identification with the ATLAS detector*. ATL-PHYS-PUB-2017-011. 2017. URL: https://cds.cern.ch/record/2270366.

[174] G Piacquadio and C Weiser. 'A new inclusive secondary vertex algorithm for b-jet tagging in ATLAS'. In: *Journal of Physics: Conference Series* 119.3 (2008), p. 032032. DOI: 10.1088/1742-6596/119/3/032032.

[175] Rudolph Emil Kalman et al. 'A new approach to linear filtering and prediction problems'. In: *Journal of basic Engineering* 82.1 (1960), pp. 35–45.

[176] Jie Zhou et al. 'Graph Neural Networks: A Review of Methods and Applications'. In: *CoRR* abs/1812.08434 (2018). arXiv: 1812.08434. URL: http://arxiv.org/abs/1812.08434.

[177] Jonathan Shlomi et al. 'Secondary Vertex Finding in Jets with Neural Networks'. In: (Aug. 2020). arXiv: 2008.02831 [hep-ex].

[178] Andrea Sciandra. *Development of a new Soft Muon Tagger for the identification of* b*-jets in ATLAS*. Tech. rep. ATL-PHYS-PROC-2017-190. Geneva: CERN, 2017. DOI: 10.22323/1.314.0768. URL: https://cds.cern.ch/record/2287545.

[179] 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 (2019), p. 970. DOI: 10.1140/epjc/s10052-019-7450-8. arXiv: 1907.05120 [hep-ex].

[180] ATLAS Collaboration. *Calibration of the* b*-tagging efficiency on charm jets using a sample of* $W + c$ *events with* $\sqrt{s} = 13$ *TeV ATLAS data*. ATLAS-CONF-2018-055. 2018. URL: https://cds.cern.ch/record/2652195.

[181] ATLAS Collaboration. *Calibration of the ATLAS* b*-tagging algorithm in* $t\bar{t}$ *semileptonic events*. ATLAS-CONF-2018-045. 2018. URL: https://cds.cern.ch/record/2638455.

[182] ATLAS Collaboration. *Calibration of light-flavour* b*-jet mistagging rates using ATLAS proton–proton collision data at* $\sqrt{s} = 13$ *TeV*. ATLAS-CONF-2018-006. 2018. URL: https://cds.cern.ch/record/2314418.

[183] ATLAS Collaboration. *Measurement of* b*-tagging efficiency of* c*-jets in* $t\bar{t}$ *events using a likelihood approach with the ATLAS detector*. ATLAS-CONF-2018-001. 2018. URL: https://cds.cern.ch/record/2306649.

[184] ATLAS Collaboration. *Monte Carlo to Monte Carlo scale factors for flavour tagging efficiency calibration*. ATL-PHYS-PUB-2020-009. 2020. URL: https://cds.cern.ch/record/2718610.

[185] ATLAS Collaboration. *Search for new phenomena in a lepton plus high jet multiplicity final state with the ATLAS experiment using* $\sqrt{s} = 13$ *TeV proton–proton collision data*. ATLAS-CONF-2017-013. 2017. URL: https://cds.cern.ch/record/2257717.

[186] ATLAS Collaboration. *Athena*. Version 22.0.1. Apr. 2019. DOI: 10.5281/zenodo.2641997. URL: https://doi.org/10.5281/zenodo.2641997.

[187] Guido Van Rossum and Fred L. Drake. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009. ISBN: 1441412697.

[188] Charles R. Harris et al. 'Array programming with NumPy'. In: *Nature* 585.7825 (Sept. 2020), pp. 357–362. DOI: 10.1038/s41586-020-2649-2.

[189] The pandas development team. *pandas-dev/pandas: Pandas*. Version latest. Feb. 2020. DOI: 10.5281/zenodo.3509134. URL: https://doi.org/10.5281/zenodo.3509134.

[190] The HDF Group. *Hierarchical data format version 5*. 2000-2010. URL: http://www.hdfgroup.org/HDF5.

[191] Felipe Pezoa, Juan L Reutter, Fernando Suarez, Martín Ugarte and Domagoj Vrgoč. 'Foundations of JSON schema'. In: *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee. 2016, pp. 263–273.

[192] J. D. Hunter. 'Matplotlib: A 2D graphics environment'. In: *Computing in Science & Engineering* 9.3 (2007), pp. 90–95. DOI: 10.1109/MCSE.2007.55.

[193] F. Pedregosa et al. 'Scikit-learn: Machine Learning in Python'. In: 12 (2011), pp. 2825–2830.

[194] Dirk Merkel. 'Docker: lightweight linux containers for consistent development and deployment'. In: *Linux journal* 2014.239 (2014), p. 2.

[195] Rafael Alencar. *Resampling strategies for imbalanced datasets*. URL: https://www.kaggle.com/rafjaa/resampling-strategies-for-imbalanced-datasets (visited on 27/11/2020).

[196] A.C. Forti, L. Heinrich and M. Guth. 'Hardware Accelerated ATLAS Workloads on the WLCG Grid'. In: *J. Phys. Conf. Ser.* 1525.1 (2020), p. 012059. DOI: 10.1088/1742-6596/1525/1/012059.

[197] ATLAS Collaboration. *Flavour-tagging efficiency corrections for the 2019 ATLAS PFlow jet* b*-taggers with the full LHC Run II dataset*. URL: http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2020-001/ (visited on 04/12/2020).

[198] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Ruslan Salakhutdinov and Alexander Smola. *Deep Sets*. 2018. arXiv: 1703.06114 [cs.LG].

[199] Jie Zhou et al. *Graph Neural Networks: A Review of Methods and Applications*. 2018. arXiv: 1812.08434. URL: http://arxiv.org/abs/1812.08434.

[200] Lynton Ardizzone, Jakob Kruse, Sebastian Wirkert, Daniel Rahner, Eric W. Pellegrini, Ralf S. Klessen, Lena Maier-Hein, Carsten Rother and Ullrich Köthe. *Analyzing Inverse Problems with Invertible Neural Networks*. 2019. arXiv: 1808.04730 [cs.LG].

[201] ATLAS Collaboration. *Luminosity determination in* pp *collisions at* $\sqrt{s} = 13$ *TeV using the ATLAS detector at the LHC*. ATLAS-CONF-2019-021. 2019. URL: https://cds.cern.ch/record/2677054.

[202] ATLAS Collaboration. 'Performance of the ATLAS trigger system in 2015'. In: *Eur. Phys. J. C* 77 (2017), p. 317. DOI: 10.1140/epjc/s10052-017-4852-3. arXiv: 1611.09661 [hep-ex].

[203] ATLAS Collaboration. 'Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8\,\text{TeV}$ using the ATLAS detector'. In: *Eur. Phys. J. C* 76 (2016), p. 581. DOI: 10.1140/epjc/s10052-016-4395-z. arXiv: 1510.03823 [hep-ex].

[204] 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. DOI: 10.1007/JHEP02(2015)075. arXiv: 1407.2922 [hep-ph].

[205] 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.

[206] M. Beneke, P. Falgari, S. Klein and C. Schwinn. 'Hadronic top-quark pair production with NNLL threshold resummation'. In: *Nucl. Phys. B* 855 (2012), pp. 695–741. DOI: 10.1016/j.nuclphysb.2011.10.021. arXiv: 1109.1536 [hep-ph].

[207] Matteo Cacciari, Michal Czakon, Michelangelo Mangano, Alexander Mitov and Paolo Nason. 'Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation'. In: *Phys. Lett. B* 710 (2012), pp. 612–622. DOI: 10.1016/j.physletb.2012.03.013. arXiv: 1111.5869 [hep-ph].

[208] Peter Bärnreuther, Michal Czakon and Alexander Mitov. 'Percent-Level-Precision Physics at the Tevatron: Next-to-Next-to-Leading Order QCD Corrections to $q\bar{q} \to t\bar{t} + X$'. In: *Phys. Rev. Lett.* 109 (2012), p. 132001. DOI: 10.1103/PhysRevLett.109.132001. arXiv: 1204.5201 [hep-ph].

[209] Michal Czakon and Alexander Mitov. 'NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels'. In: *JHEP* 12 (2012), p. 054. DOI: 10.1007/JHEP12(2012)054. arXiv: 1207.0236 [hep-ph].

[210] Michal Czakon and Alexander Mitov. 'NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction'. In: *JHEP* 01 (2013), p. 080. DOI: 10.1007/JHEP01(2013)080. arXiv: 1210.6832 [hep-ph].

[211] Michal Czakon, Paul Fiedler and Alexander Mitov. 'Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$'. In: *Phys. Rev. Lett.* 110 (2013), p. 252004. DOI: 10.1103/PhysRevLett.110.252004. arXiv: 1303.6254 [hep-ph].

[212] Michal Czakon and Alexander Mitov. 'Top++: A program for the calculation of the top-pair cross-section at hadron colliders'. In: *Comput. Phys. Commun.* 185 (2014), p. 2930. DOI: 10.1016/j.cpc.2014.06.021. arXiv: 1112.5675 [hep-ph].

[213] Rikkert Frederix, Emanuele Re and Paolo Torrielli. 'Single-top t-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO'. In: *JHEP* 09 (2012), p. 130. DOI: 10.1007/JHEP09(2012)130. arXiv: 1207.5391 [hep-ph].

[214] Simone Alioli, Paolo Nason, Carlo Oleari and Emanuele Re. 'NLO single-top production matched with shower in POWHEG: s- and t-channel contributions'. In: *JHEP* 09 (2009), p. 111. DOI: 10.1088/1126-6708/2009/09/111. arXiv: 0907.4076 [hep-ph].

[215] Emanuele Re. 'Single-top *W*t-channel production matched with parton showers using the POWHEG method'. In: *Eur. Phys. J. C* 71 (2011), p. 1547. DOI: 10.1140/epjc/s10052-011-1547-z. arXiv: 1009.2450 [hep-ph].

[216] Nikolaos Kidonakis. 'Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^-'. In: *Phys. Rev. D* 82 (2010), p. 054018. DOI: 10.1103/PhysRevD.82.054018. arXiv: 1005.4451 [hep-ph].

[217] Nikolaos Kidonakis. 'Top Quark Production'. In: *Proceedings, Helmholtz International Summer School on Physics of Heavy Quarks and Hadrons (HQ 2013): JINR, Dubna, Russia, July 15-28, 2013*. 2014, pp. 139–168. DOI: 10.3204/DESY-PROC-2013-03/Kidonakis. arXiv: 1311.0283 [hep-ph].

[218] M. Aliev, H. Lacker, U. Langenfeld, S. Moch, P. Uwer and M. Wiedermann. 'HATHOR – HAdronic Top and Heavy quarks crOss section calculatoR'. In: *Comput. Phys. Commun.* 182 (2011), pp. 1034–1046. DOI: 10.1016/j.cpc.2010.12.040. arXiv: 1007.1327 [hep-ph].

[219] P. Kant, O. M. Kind, T. Kintscher, T. Lohse, T. Martini, S. Mölbitz, P. Rieck and P. Uwer. '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. DOI: 10.1016/j.cpc.2015.02.001. arXiv: 1406.4403 [hep-ph].

[220] Tanju Gleisberg and Stefan Höche. 'Comix, a new matrix element generator'. In: *JHEP* 12 (2008), p. 039. DOI: 10.1088/1126-6708/2008/12/039. arXiv: 0808.3674 [hep-ph].

[221] Steffen Schumann and Frank Krauss. 'A parton shower algorithm based on Catani–Seymour dipole factorisation'. In: *JHEP* 03 (2008), p. 038. DOI: 10.1088/1126-6708/2008/03/038. arXiv: 0709.1027 [hep-ph].

[222] Stefan Höche, Frank Krauss, Marek Schönherr and Frank Siegert. 'A critical appraisal of NLO+PS matching methods'. In: *JHEP* 09 (2012), p. 049. DOI: 10.1007/JHEP09(2012)049. arXiv: 1111.1220 [hep-ph].

[223] Stefan Höche, Frank Krauss, Marek Schönherr and Frank Siegert. 'QCD matrix elements + parton showers. The NLO case'. In: *JHEP* 04 (2013), p. 027. DOI: 10.1007/JHEP04(2013)027. arXiv: 1207.5030 [hep-ph].

[224] S. Catani, F. Krauss, R. Kuhn and B. R. Webber. 'QCD Matrix Elements + Parton Showers'. In: *JHEP* 11 (2001), p. 063. DOI: 10.1088/1126-6708/2001/11/063. arXiv: hep-ph/0109231.

[225] Stefan Höche, Frank Krauss, Steffen Schumann and Frank Siegert. 'QCD matrix elements and truncated showers'. In: *JHEP* 05 (2009), p. 053. DOI: 10.1088/1126-6708/2009/05/053. arXiv: 0903.1219 [hep-ph].

[226] Charalampos Anastasiou, Lance J. Dixon, Kirill Melnikov and Frank Petriello. 'High precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at next-to-next-to leading order'. In: *Phys. Rev. D* 69 (2004), p. 094008. DOI: 10.1103/PhysRevD.69.094008. arXiv: hep-ph/0312266.

[227] Rikkert Frederix, Davide Pagani and Marco Zaro. 'Large NLO corrections in $t\bar{t}W^{\pm}$ and $t\bar{t}t\bar{t}$ hadroproduction from supposedly subleading EW contributions'. In: *JHEP* 02 (2018), p. 031. DOI: 10.1007/JHEP02(2018)031. arXiv: 1711.02116 [hep-ph].

[228] J. Pumplin et al. 'New Generation of Parton Distributions with Uncertainties from Global QCD Analysis'. In: *JHEP* 07 (2002), p. 012. DOI: 10.1088/1126-6708/2002/07/012. arXiv: hep-ph/0201195.

[229] 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. DOI: 10.1088/1126-6708/2008/07/029. arXiv: 0805.3067 [hep-ph].

[230] ATLAS Collaboration. 'Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s} = 8\,\text{TeV}$ with the ATLAS detector'. In: *Eur. Phys. J. C* 75 (2015), p. 349. DOI: 10.1140/epjc/s10052-015-3543-1. arXiv: 1503.05066 [hep-ex].

[231] 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. DOI: 10.1140/epjc/s10052-018-5956-0. arXiv: 1802.00426 [hep-ph].

[232] Tomas Ježo. *Powheg-Box-Res ttbb source code*. 2019. URL: https://gitlab.cern.ch/tjezo/powheg-box-res_ttbb/.

[233] Fabio Cascioli, Philipp Maierhöfer and Stefano Pozzorini. 'Scattering Amplitudes with Open Loops'. In: *Phys. Rev. Lett.* 108 (2012), p. 111601. DOI: 10.1103/PhysRevLett.108.111601. arXiv: 1111.5206 [hep-ph].

[234] Ansgar Denner, Stefan Dittmaier and Lars Hofer. 'COLLIER: A fortran-based complex one-loop library in extended regularizations'. In: *Comput. Phys. Commun.* 212 (2017), pp. 220–238. DOI: 10.1016/j.cpc.2016.10.013. arXiv: 1604.06792 [hep-ph].

[235] Vernon D. Barger, J. Ohnemus and R. J. N. Phillips. 'Event shape criteria for single lepton top signals'. In: *Phys. Rev. D* 48 (1993), R3953–R3956. DOI: 10.1103/PhysRevD.48.R3953. arXiv: hep-ph/9308216 [hep-ph].

[236] 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. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an].

[237] 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.

[238] Wouter Verkerke and David Kirkby. *The RooFit toolkit for data modeling*. 2003. arXiv: physics/0306116 [physics.data-an].

[239] Lorenzo Moneta et. al. *The RooStats Project*. 2011. arXiv: 1009.1003 [physics.data-an].

[240] ATLAS Collaboration. 'Measurement of the Inelastic Proton–Proton Cross Section at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS Detector at the LHC'. In: *Phys. Rev. Lett.* 117 (2016), p. 182002. DOI: 10.1103/PhysRevLett.117.182002. arXiv: 1606.02625 [hep-ex].

[241] Risto Raitio and Walter W. Wada. 'Higgs-boson production at large transverse momentum in quantum chromodynamics'. In: *Phys. Rev. D* 19 (1979), p. 941. DOI: 10.1103/PhysRevD.19.941.

[242] W. Beenakker, S. Dittmaier, M. Krämer, B. Plümper, M. Spira and P.M. Zerwas. 'NLO QCD corrections to $t\bar{t}H$ production in hadron collisions'. In: *Nucl. Phys. B* 653 (2003), pp. 151–203. DOI: 10.1016/S0550-3213(03)00044-0. arXiv: hep-ph/0211352.

[243] 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. DOI: 10.1103/PhysRevD.68.034022. arXiv: hep-ph/0305087.

[244] 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. DOI: 10.1016/j.physletb.2014.09.022. arXiv: 1407.1110 [hep-ph].

[245] 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. DOI: 10.1007/JHEP06(2015)184. arXiv: 1504.03446 [hep-ph].

[246] Iain W. Stewart and Frank J. Tackmann. 'Theory uncertainties for Higgs and other searches using jet bins'. In: *Phys. Rev.* D85 (2012), p. 034011. DOI: 10.1103/PhysRevD.85.034011. arXiv: 1107.2117 [hep-ph].

[247] 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. DOI: 10.1140/epjc/s10052-009-1164-2. arXiv: 0905.3531 [hep-ph].

[248] John M. Campbell and R. Keith Ellis. '$t\bar{t}W^{+-}$ production and decay at NLO'. In: *JHEP* 07 (2012), p. 052. DOI: 10.1007/JHEP07(2012)052. arXiv: 1204.5678 [hep-ph].

[249] ATLAS Collaboration. *Multi-boson simulation for* 13 *TeV ATLAS analyses*. ATL-PHYS-PUB-2016-002. 2016. URL: https://cds.cern.ch/record/2119986.

[250] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli and M. Zaro. 'The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations'. In: *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph].

[251] Robert D. Cousins. *Generalization of chisquare goodness-of-fit test for binned data using saturated models, with application to histograms*. URL: http://www.physics.ucla.edu/~cousins/stats/cousins_saturated.pdf (visited on 11/04/2021).

[252] *Worldwide LHC Computing Grid*. URL: https://wlcg-public.web.cern.ch/ (visited on 01/12/2020).

[253] Gregory M. Kurtzer, Vanessa Sochat and Michael W. Bauer. 'Singularity: Scientific containers for mobility of compute'. In: *PLOS ONE* 12.5 (May 2017), pp. 1–20. DOI: 10.1371/journal.pone.0177459. URL: https://doi.org/10.1371/journal.pone.0177459.

[254] Martin Barisits et al. 'Rucio - Scientific Data Management'. In: *CoRR* abs/1902.09857 (2019). arXiv: 1902.09857. URL: http://arxiv.org/abs/1902.09857.

université PARIS-SACLAY

ÉCOLE DOCTORALE
Particules, hadrons, énergie et noyau : instrumentation, imagerie, cosmos et simulation (PHENIICS)

TITRE: 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

MOTS CLÉS: boson de Higgs, LHC, ttH, quark top, ATLAS

RÉSUMÉ: ATLAS et CMS ont récemment découvert le processus de production $t\bar{t}H$ en utilisant les données prises durant le RUN II du LHC. Le processus $t\bar{t}H(H \to b\bar{b})$ permet de mesurer directement le couplage de Yukawa du quark top, qui est le couplage fermion-Higgs le plus grand du modèle standard et joue donc un rôle important dans la physique du boson du Higgs. L'état final de ce processus contient au moins 4 jets provenant de quarks b ce qui nécessite d'établir une stratégie d'analyse avancée ainsi que de développer des méthodes sophistiquées pour l'identification des jets provenant de quarks b. L'étiquetage des quarks b n'est pas seulement crucial pour l'analyse $t\bar{t}H(b\bar{b})$, mais aussi pour la plupart des analyses de physique au sein de l'expérience d'ATLAS. La ré-optimisation de l'étiquetage des quarks de saveurs lourdes basé sur un apprentissage profond dans ATLAS est présentée dans cette thèse pour deux collections de jets différentes. Diverses améliorations ont été apportées, entraînant une augmentation importantes des performances allant jusqu'à un facteur deux dans certaines régions de l'espace des phases. L'analyse $t\bar{t}H(b\bar{b})$ est effectuée en utilisant $139\,fb^{-1}$ de données enregistrées par ATLAS durant le RUN II à une énergie dans le centre de masse de $\sqrt{s} = 13\,TeV$. L'intensité du signal, qui est le rapport entre la section efficace mesurée et la section efficace prédite par le modèle standard, a été mesurée à $0,43^{+0,20}_{-0,19}(stat.)^{+0,30}_{-0,27}(syst.)$ avec une signification observée (prévue) de $1,3$ $(3,0)$ déviations standard pour la mesure de la section efficace inclusive. En outre, une mesure simplifiée de la section efficace utilisant des gabarits Monte Carlo en fonction de l'impulsion transverse du boson de Higgs est effectuée. Cette mesure est limitée par la difficulté de simuler correctement le bruit de fond dominant $t\bar{t} + b\bar{b}$ ainsi que par de grandes incertitudes systématiques.

TITLE: 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

KEYWORDS: Higgs boson, LHC, ttH, top quark, ATLAS

ABSTRACT: Since several decades, the predictions of the Standard Model (SM) of particle physics are being probed and validated. One major success of the Large Hadron Collider (LHC) at CERN was the discovery of the Higgs boson in 2012. With the increasing amount of proton-proton collisions recorded with the experiments located at the LHC, precise Higgs measurements are now possible and rare processes are accessible.
ATLAS and CMS recently discovered the production process of a Higgs boson in association with a pair of top quarks using LHC RUN II data. The $t\bar{t}H(H \to b\bar{b})$ process allows for a direct measurement of the Top-Yukawa coupling which is the strongest fermion-Higgs coupling in the Standard Model and plays therefore an important role in Higgs physics. The challenging final state with at least 4 b-jets requires an advanced analysis strategy as well as sophisticated b-jet identification methods. b-tagging is not only crucial in the $t\bar{t}H(b\bar{b})$ analysis, but most physics analyses within ATLAS are making use of it. The reoptimisation of the deep-learning-based heavy flavour tagger in ATLAS is shown in this thesis for two different jet collections. Various improvements were made resulting in a drastic performance increase up to a factor two in certain regions of the phase space. The $t\bar{t}H(b\bar{b})$ analysis is performed using $139\,fb^{-1}$ of RUN II ATLAS data at a centre-of-mass energy of $\sqrt{s} = 13\,TeV$. The signal strength, being the ratio of the measured cross-section over the predicted cross-section in the SM, was measured to be $0.43^{+0.20}_{-0.19}(stat.)^{+0.30}_{-0.27}(syst.)$ with an observed (expected) significance of 1.3 (3.0) standard deviations in the inclusive cross-section measurement. In addition, a simplified template cross-section (STXS) measurement in different Higgs p_T bins is performed which is possible because of the ability to reconstruct the Higgs boson. The measurement is limited by the capability to describe the challenging irreducible $t\bar{t} + b\bar{b}$ background and by systematic uncertainties.

ZUSAMMENFASSUNG: Seit mehreren Jahrzehnten werden die Vorhersagen des Standardmodells (SM) der Teilchenphysik erprobt und validiert. Mit der zunehmenden Anzahl von Proton-Proton-Kollisionen, die mit den Experimenten am LHC aufgezeichnet werden, sind nun präzise Higgs-Messungen möglich.
ATLAS und CMS haben kürzlich den $t\bar{t}H$-Produktionsprozess mit Hilfe von LHC RUN II-Daten entdeckt. Der $t\bar{t}H(H \to b\bar{b})$-Prozess ermöglicht eine direkte Messung der Top-Yukawa-Kopplung, welche die stärkste Fermion-Higgs-Kopplung ist und daher eine wichtige Rolle im SM einnimmt. Der anspruchsvolle Endzustand mit mindestens 4 b-Jets erfordert eine fortschrittliche Analysestrategie sowie elaborierte b-Jet-Identifikationsmethoden. b-Tagging ist nicht nur in der $t\bar{t}H(b\bar{b})$-Analyse von entscheidender Bedeutung, sondern die meisten Physik-Analysen innerhalb von ATLAS machen davon Gebrauch. Die Re-Optimierung des Deep-Learning-basierten Heavy-Flavour Taggers in ATLAS wird in dieser Arbeit für zwei verschiedene Jet-Definitionen gezeigt. Es wurden verschiedene Änderungen vorgenommen, die zu einer signifikanten Verbesserung von bis zu einem Faktor zwei in der Untergrundunterdrückung in bestimmten Phasenraumregionen führten. Die $t\bar{t}H(b\bar{b})$-Analyse wurde mit $139\,fb^{-1}$ RUN II ATLAS-Daten bei einer Schwerpunktsenergie von $\sqrt{s} = 13\,TeV$ durchgeführt. Die Signalstärke, d.h. das Verhältnis des gemessenen Wirkungsquerschnitts zum vorhergesagten Wirkungsquerschnitt im SM, wurde mit $0,43^{+0,20}_{-0,19}(stat.)^{+0,30}_{-0,27}(syst.)$ mit einer beobachteten (erwarteten) Signifikanz von $1,3$ $(3,0)$ Standardabweichungen für den inklusiven Wirkungsquerschnitt gemessen. Zusätzlich wurde zum ersten Mal eine vereinfachte differenzielle Wirkungsquerschnittsmessung in verschiedenen Higgs p_T-Bereichen durchgeführt. Die Messung wird durch systematische Unsicherheiten begrenzt, hauptsächlich im Zusammenhang mit dem anspruchsvollen irreduziblen $t\bar{t} + b\bar{b}$ Untergrund.

Maison du doctorat de l'Université Paris-Saclay
2ème étage aile ouest, Ecole normale supérieure Paris-Saclay
4 avenue des Sciences,
91190 Gif sur Yvette, France