Thesis

Machine learning-based identification of boosted objects and search for pair production of heavy vector-like quarks in fully-hadronic final states with the ATLAS detector

AKILLI, Ece

Abstract

In this thesis, the data recorded by the ATLAS experiment at the LHC in 2015 and 2016, corresponding to an integrated luminosity of 36.1 fb-1, is analyzed to study the performance of the boosted object identification algorithms and to search for pair-produced heavy vector-like quarks in fully hadronic final states. The first part of the thesis focuses on the use of jet moments as inputs to deep neural networks to build binary jet classifiers which discriminate W-boson or top-quark jets from the gluon and light-quark jet background. The second part of the thesis presents a search for pair-produced vector-like quarks in fully-hadronic final states with small missing transverse momentum. The analysis strategy is optimized assuming that the pair produced vector-like quarks decay into a Standard Model boson and a third-generation quark. No significant deviation from the Standard Model expectation is observed and upper limits are set on the production cross-sections.

Reference

UNIVERSITÉ DE GENÈVE

Machine learning-based identification of boosted objects and search for pair production of heavy vector-like quarks in fully-hadronic final states with the ATLAS detector

THÈSE

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

par

Ece Akıllı

de Antalya (Turquie)

Thèse N° 5418

UNIVERSITÉ DE GENÈVE
FACULTÉ DES SCIENCES

DOCTORAT ÈS SCIENCES, MENTION PHYSIQUE

Thèse de Madame Ece AKILLI

intitulée :

«Machine Learning-Based Identification of Boosted Objects and Search for Pair Production of Heavy Vector-like Quarks in Fully-Hadronic Final States with the ATLAS Detector»

La Faculté des sciences, sur le préavis de Monsieur T. GOLLING, professeur associé et directeur de thèse (Département de physique nucléaire et corpusculaire), Monsieur F. SANCHEZ NIETO, professeur ordinaire (Département de physique nucléaire et corpusculaire), Monsieur C. CLEMENT, professeur (Department of physics, Stockholm University, Sweden), autorise l'impression de la présente thèse, sans exprimer d'opinion sur les propositions qui y sont énoncées.

Genève, le 5 décembre 2019

Thèse - 5418 -

Le Décanat

N.B. -   La thèse doit porter la déclaration précédente et remplir les conditions énumérées dans les "Informations relatives aux thèses de doctorat à l'Université de Genève".

Acknowledgments

A PhD in physics was beyond my imagination as a young student and it would not have been possible without the ones who surrounded me. I would like to thank everyone who surrounded and supported me during this journey.

First, I would like to thank my PhD advisor Tobias Golling for accepting me to his group halfway through my master thesis and giving me the opportunity to stay for my PhD. I am grateful for his guidance, support and confidence in me during these challenging years. During my PhD I appreciated the freedom that he gave me and his encouragement on taking on exciting and challenging tasks while keeping me grounded. I also would like to thank my thesis committee members Christophe Clement and Federico Sanchez for reading my thesis and their very helpful comments to improve it.

During my time at UNIGE I had the opportunity to collaborate with the brilliant members of the Golling group, thank you all for insightful discussions. I thank Andrea Coccaro for getting me started with my PhD and jet tagging, Marie Lanfermann for getting me started with deep neural networks, related framework and our discussions. Many thanks to Olaf Nackenhorts for constantly encouraging and motivating me and at the same time giving me freedom and space. I always felt inspired and supported while Olaf and I worked on projects together and had a challenging review process. I thank Stefan Gadatsch for discussing new ideas, beginning a new exciting analysis and his support while I was reaching the end of this path and starting to prepare for my next steps.

I would like to thank all the past and current members of the ATLAS group at UNIGE for interesting discussions and also for fun times. I thank Steven Schramm for his endless support and guidance. Our meetings and at times spontaneous chats made my PhD easier and more enjoyable. I would also like to thank him for finding the time to provide useful comments on my thesis during particularly busy times, as always. I thank Chris Delitzsch, who was an incredible office mate and continued to be an inspiring colleague and friend after she left UNIGE. I thank her for always answering my questions and motivating me. JetEtMiss and workshops would not be the same without her. I am also grateful for her feedback on my thesis. I would also like to thank my other wonderful office mates: Francesco Di Bello, Marco Valente, Riccardo Poggi, Sofia Adorni Braccesi, Moritz Kiehn and Juno Nindhito for their company.

I had the opportunity to work on several exciting projects within different groups in ATLAS. I would like to thank the JSS group, particularly all the W/top tagging team members. I feel grateful that I had the opportunity to work with such a team for several years. Many thanks to Amal Vaidya, Tatsumi Nitta and Tobias Kupfer for our hands-on collaboration. Special thanks

Abstract

In this thesis, the data recorded by the ATLAS experiment at the LHC in 2015 and 2016, corresponding to an integrated luminosity of 36.1 fb^{-1}, is analyzed to study the performance of the boosted object identification algorithms and to search for pair-produced heavy vector-like quarks in fully hadronic final states.

The first part of the thesis focuses on the use of jet moments as inputs to deep neural networks to build binary jet classifiers which discriminate W-boson or top-quark jets from the gluon and light (non-top)-quark jet background, referred to as taggers. The optimization of the deep neural networks (DNN) and the procedure of defining the tagging working points are presented for DNN W and DNN top taggers. The performance of the deep neural network-based taggers is compared with the performance of other W-boson and top-quark tagging techniques in Monte Carlo simulation. Finally, the collected data is used to study the performance of these tagging techniques. The tagging signal efficiency and background rejection in data are measured using boosted lepton+jet $t\bar{t}$, dijet and γ + jet topologies.

The second part of the thesis presents a search for pair-produced heavy vector-like quarks, $T\bar{T}$ and $B\bar{B}$, in fully hadronic final states with small missing transverse momentum. In this analysis, the strategy is optimized assuming that the pair produced vector-like quarks decay into a Standard Model boson and a third-generation quark, leading to $T \rightarrow Wb, Ht, Zt$ or $B \rightarrow Wt, Hb, Zb$ decay modes. A deep neural network is used to identify W/Z, Higgs boson, top quark, and background jets. A novel discriminant based on the matrix element method is used to discriminate vector-like quark signal events from background events. No significant deviation from the Standard Model expectation is observed. Finally, upper limits are set on the production cross-section of $T\bar{T}$ and $B\bar{B}$ events.

Résume

Dans cette thèse, les données récoltées par l'expérience ATLAS au LHC en 2015 et 2016, correspondant à une luminosité integrée de 36.1 fb^{-1}, sont analysées afin d'étudier la performance d'algorithmes d'identification d'objects boostés et pour la recherche de quarks de type vecteur produits en paires dans des états finaux complètement hadroniques.

La première partie de cette thèse se concentre sur l'utilisation des variables des jets comme variables d'entrées données à des réseaux neuronaux profonds de façon à construire des classifieurs de jets binaires, appelés étiqueteurs (*taggers*), qui différencient les bosons W ou les quarks tops du bruit de fond qui se compose de gluons et de jets de quarks légers (autre que le top). L'optimisation des réseaux neuronaux profonds (DNN) et la procédure de définition de la sélection de *tagging* sont présentés pour les étiqueteurs DNN W et DNN top. La performance des étiqueteurs basés sur des réseaux neuronaux profonds est comparée à la performance d'autres techniques de reconnaissance de bosons W et de quarks top dans des simulations de Monte Carlo. Finalement, les données collectées sont utilisées afin d'étudier la performance de ces techniques d'étiquetage. L'efficacité d'étiquetage du signal et celle de rejet du bruit de fond sont mesurées grâce à l'utilisation des topologies de lepton boosté+jet tt, de dijet et de γ + jet.

La seconde partie de cette thèse présente une recherche pour la production en paire de quarks lourds de type vecteur, $T\bar{T}$ and $B\bar{B}$, dans des états finaux complètement hadroniques avec une petite $E_{\mathrm{T}}^{\mathrm{miss}}$. Dans cette analyse, la stratégie est optimisée en supposant que les quarks de type vecteur produits en paires se désintègrent en un boson du Modèle Standard et en un quark de troisième génération, menant aux modes de désintégration $T \rightarrow Wb, Ht, Zt$ ou $B \rightarrow Wt, Hb, Zb$. Un réseau neuronal profond est utilisé afin d'identifier W/Z, le boson de Higgs, le quark top et les jets composant le bruit de fond. Un nouveau disciminant basé sur la méthode de l'élément de matrice est utilisé afin de discriminer le signal de quarks de type vecteur des évènements du bruit de fond. Aucune déviation significative par rapport aux prédictions du Modèle Standard n'est observée. Finalement, des limites supérieures sont établies pour la section efficace de la production d'évènements $T\bar{T}$ et $B\bar{B}$.

Contents

Chapter 1

Introduction

Our current understanding of fundamental particles and their interactions is embodied in the Standard Model. The Standard Model is proven to be a very successful model by the majority of experimental observations, such as the discovery of the Higgs boson and precision measurements. Although it is a successful model, some experimental observations are left unexplained and theoretical questions unanswered. Two notable examples are the existence of dark matter and the mass of the Higgs boson. The luminous matter predicted by the Standard Model is not sufficient to account for some of the observations in large astrophysical systems suggesting the existence of dark matter. Although dark matter is thoroughly searched for, it has not been found yet. Regarding the observed mass of the Higgs boson, it is not well understood why its mass is very low while the theoretical mass calculation predicts it to have large corrections. Physics beyond the Standard Model aims to address such limitations of the Standard Model. Vector-like quarks are hypothetical particles which are predicted by several extensions to the Standard Model.

The Large Hadron Collider at CERN accelerates proton beams to unprecedented high-energies, collides them at four interaction points and generates a large number of proton-proton collisions which are recorded by the experiments located at the interaction points. ATLAS is one of these experiments, recording the outcome of high-energy, high-luminosity collisions to be analyzed. The large datasets of collisions at unprecedented center-of-mass energies at ATLAS presents unique opportunities to study properties of the Standard Model with great precision and to search for physics beyond the Standard Model. While presenting unique opportunities, such high luminosities and collision energies also introduce new challenges which require employing novel object-level and event-level algorithms to analyze the collected data. This thesis addresses such challenges within the context of search for physics beyond the Standard Model in hadronic final states. In particular, in a search for vector-like quarks, which are hypothetical particles predicted by several extensions to the Standard Model.

At high collision energies, large samples of heavy particles are produced with a transverse momentum that exceeds their rest mass. Such boosted heavy particles play an important role in search for signs of new physics. Hence, it is of great importance to efficiently and robustly reconstruct and identify boosted heavy particles. In this thesis, performance of machine learning-based boosted object identification techniques are studied and machine learning-based identification is

applied within the context of a search for pair-produced heavy vector-like quarks in fully hadronic final states in ATLAS. The data recorded by ATLAS in 2015 and 2016, corresponding to an integrated luminosity of 36.1 fb^{-1}, is used.

This thesis is organized as follows. The Standard Model, its limitations and vector-like quarks are presented in chapter 2. An overview of machine learning techniques for classification is presented in chapter 3. The Large Hadron Collider and the ATLAS experiment are described in chapter 4, followed by the description of physics objects in ATLAS in chapter 5. Large-radius jets and its traditional identification techniques are introduced in chapter 6. Machine-learning based large-radius jet identification and performance of taggers in data and simulation are presented in chapter 7. The object definitions, identification techniques, the analysis strategy and results of the search for pair-produced heavy vector-like quarks, $T\bar{T}$ and $B\bar{B}$, in fully hadronic final states are presented in chapter 8. Finally, concluding remarks and outlook are presented in chapter 9.

1.1 Contributions

The ATLAS collaboration consists of more than 3000 scientists who work together in tasks such as construction of the detector, ensuring smooth detector operation, good quality data collection and analysis of the collected data. The studies presented in this thesis were thus carried out in collaboration. My own contributions are listed in this section.

- Chapters 6 – 7: I was involved in all aspects of the large-radius jet identification analysis as a co-leader of the analysis starting from the jet labelling presented in section 6.3. My main contributions were on the simulation-based studies and machine learning-based taggers. Initially I set up a framework which provides common simulated samples for the effort. This framework facilitates extensive jet observables and taggers allowing for a consistent comparison of variables and taggers. Consequently, I also contributed to testing the performance of combined mass, a new mass definition in ATLAS. My major contribution was the development of deep neural network-based W-boson and top-quark taggers which use jet moments as input. This work has been published in Refs. [1, 2].

- Chapter 8: My main contribution was the development of a deep neural network-based multi-class boosted object tagger to distinguish W/Z-boson, Higgs-boson, top-quark and background jets. Moreover, I compared the performance of the chosen and the alternative large-radius jet collections. Finally, I contributed to harmonization of systematic uncertainties for the ATLAS combination of pair-produced vector-like quark searches. This work has been published in Refs. [3, 4].

Chapter 2

The Standard Model

Our current understanding of particle physics is embodied in the Standard Model [5]. The precision tests of the Standard Model and the discovery of the Higgs boson validate that the Standard Model is a very successful model. Although the Standard Model is very successful, it cannot answer all questions in particle physics. Physics beyond the Standard Model attempts to address its shortcomings and the observations which cannot be explained by it. This chapter presents a review of the Standard Model, some of its limitations and theories of physics beyond the Standard Model.

2.1 The Standard Model

The Standard Model (SM) is formulated as a quantum field theory (QFT) with an underlying symmetry. It provides information about fundamental particles and their interactions in terms of particle properties and particle exchange between them [6]. The type of interactions that the particles are subject to are determined by the quantum numbers such as the spin and charge. In the SM, there are two types of particles: fermions and bosons. Fermions have half integer spin whereas bosons have integer spin. The fermions of the SM are quarks and leptons. In addition, there exists a corresponding antiparticle for each fermion which has the same mass and spin but opposite charge and weak isospin. The bosons of the SM are the gauge bosons (spin-1) and the scalar Higgs boson (spin-0). Gauge bosons are the force carriers of the SM and are responsible for three of the four interactions in nature. These three interactions are strong, electromagnetic and weak interactions. Gravity, the weakest of all four interactions in nature, and graviton, the associated hypothetical particle, are not included in the SM. The gravitational force between fundamental particles can be neglected at the energy scale of particle physics. The particles of the SM, their important properties and the interactions are summarized in Fig. 2.1.

2.1.1 Quarks and leptons

The fermions are categorized based on the types of interactions that they are subject to. They are first divided into two main groups as quarks and leptons based on the strong interaction. Unlike leptons, quarks carry color charge and are subject to the strong interaction. The fermions are arranged in three generations where each generation consists of one up-type quark with 2/3
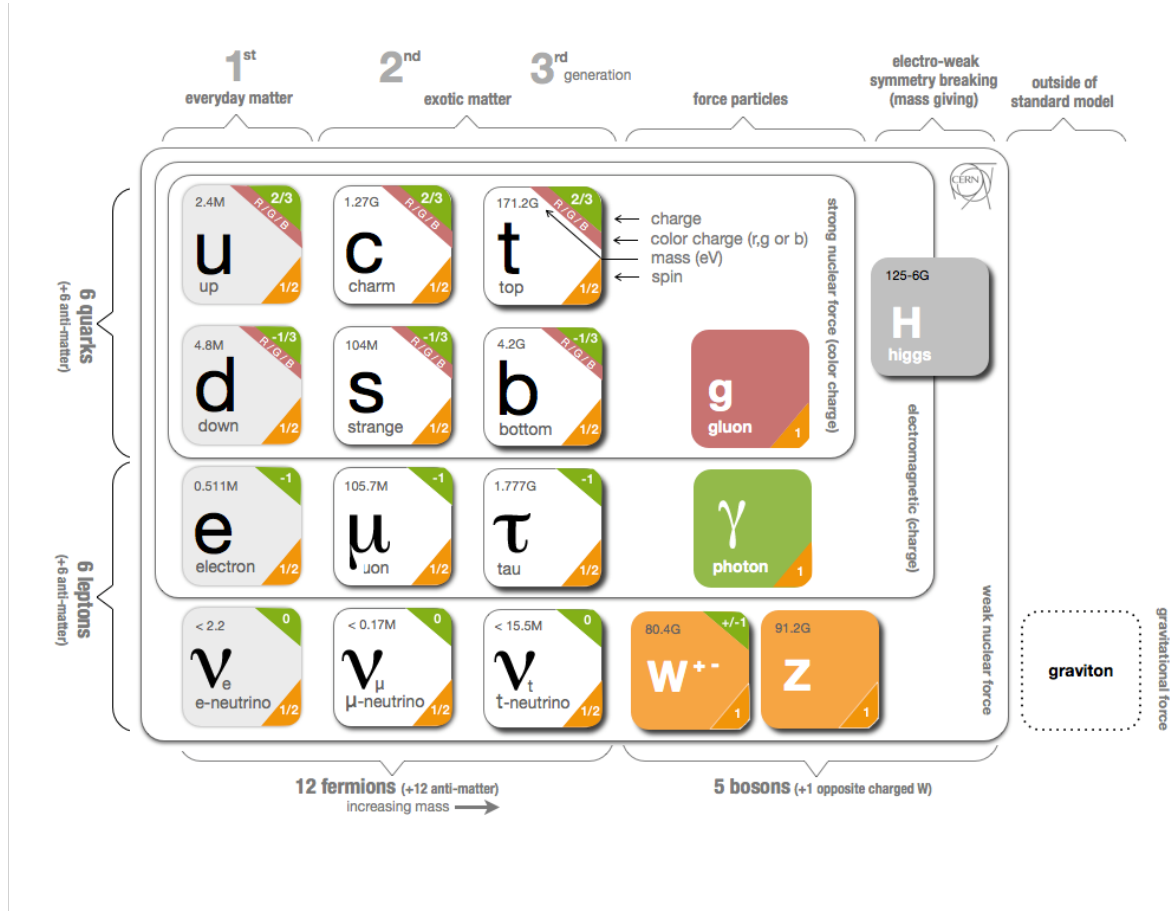
Figure 2.1: Particles and interactions of the Standard Model [7]. Electric charge, color charge, mass and spin are presented for each of the Standard Model particles.

electric charge (up, charm, top) and one down-type quark with $-1/3$ electric charge charge (down, strange, bottom), one electrically charged lepton (electron, muon, tau) and the corresponding electrically neutral lepton (ν_e, ν_μ, ν_τ neutrinos). The quantum numbers and interactions of each generation are identical while the masses of the corresponding fermions increase from the first generation to the third generation. The number of generations is not limited by the theory, however, currently there is no evidence for fourth generation of fermions. Only the first generation of fermions are stable and they make up the everyday matter around us. All fermions are subject to weak interaction. Electrically charged leptons additionally are subject to electromagnetic interaction whereas neutrinos are not. Since quarks carry electric and color charge, they are the only fermions which are subject to all three interactions of the SM. The couplings of the SM interactions define the strength of the interactions.

2.1.2 Interactions

In QFT, the models are built by postulating symmetries where for each symmetry, there is an associated conserved quantity. Noether's theorem [8] associates a symmetry of a Lagrangian

with a corresponding conserved current. The Standard Model is formalized as a renormalizable, gauge and Lorentz invariant QFT. The underlying symmetry of the SM is

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

where $SU(3)_C$ represents the color group for the strong interaction and $SU(2)_L \otimes U(1)_Y$ the electroweak interaction. The subscripts C and Y refer to the conserved quantum numbers associated with the group: C is the color charge and Y is the hypercharge. The interactions between the observed fundamental particles can be extracted from the SM Lagrangian. The subscript L represents that only left-handed fermions are affected by the $SU(2)_L$ group where the weak isospin, I, is conserved. Since each $SU(n)$ group has $n^2 - 1$ and each $U(n)$ group has n generators, the SM group has in total 12 generators, corresponding to 12 gauge bosons.

Electroweak interaction

The SM describes the electromagnetic and weak interactions by a unified theory. Initially these two interactions were described individually where the electromagnetic interaction was described by quantum electrodynamics (QED) and the strong interaction by the Fermi theory. However, it was observed that this approach did not hold at high energies (~ 100 GeV) and the electroweak theory was proposed. The electroweak interactions are described by the group $SU(2)_L \otimes U(1)_Y$ where the symmetry leads to singlets of right-handed charged fermions[1] and doublets of left-handed fermions under $SU(2)_L$. The relation between the electroweak quantum numbers and electric charge (Q) is given by $Y = 2(Q - I_3)$, where I_3 is the weak isospin along the z-axis. The $SU(2)_L \otimes U(1)_Y$ group has four generators. One of the four generators of this group is the isosinglet that belongs to the $U(1)_Y$ group and denoted as B_μ. Three of these generators are the isotriplet of vector bosons which belong to the $SU(2)_L$ group and denoted as $W_\mu^{1,2,3}$. Four mediators of the electroweak interactions are formed by mixing these generators. These mediators are the photon (γ), $W^\pm$ and Z^0 bosons and are obtained by,

$$W^\pm = \frac{1}{\sqrt{2}}(-W_\mu^1 \pm iW_\mu^2) \tag{2.2a}$$

$$\begin{bmatrix} \gamma \\ Z^0 \end{bmatrix} = \begin{bmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{bmatrix} \begin{bmatrix} B_\mu \\ W_\mu^3 \end{bmatrix} \tag{2.2b}$$

where θ_W is the Weinberg (weak mixing) angle. The photon is the mediator of the electromagnetic interaction, it is massless and neutral. Since the photon is massless, it has an infinite range where the strength of the interaction decreases with distance and since it is neutral, it does not have self-interactions. The electrically charged W bosons are responsible for the weak charged current and the neural Z boson is responsible for the weak neutral current. Although in the SM, gauge bosons are imposed to be massless by local gauge invariance, the W and Z bosons

---

[1]Right-handed neutrinos are not observed.

are massive particles. The mediators of the weak interaction obtain mass by the spontaneous breaking of electroweak gauge symmetry, this is referred to as the Brout-Englert-Higgs mechanism [9, 10]. Due to the mass scale of its mediators, the weak interaction has a limited range and is effective only at short distances and high energies. It was observed that only left-handed fermions couple to the $W^{\pm}$, while right-handed fermions do not. Hence, the $SU(2)_L$ acts only on left-handed fermions and violates parity. Finally, unlike the other interactions, the weak interaction allows for flavor changing charged currents. An example of such an interaction is the decay of the top quark to a bottom quark by the emission of a W boson. This is possible due to the fact that the mass eigenstates and the weak eigenstates are not the same, and the Cabibbo-Kobayashi-Maskawa (CKM) matrix [11] defines the transformation between the two sets.

Higgs mechanism

The spontaneous symmetry breaking (SSB) of the electroweak group is introduced to allow for massive W and Z bosons. To break $SU(2)_L \otimes U(1)_Y$, a scalar $SU(2)$-doublet can be introduced. Brout-Englert-Higgs mechanism, or shortly Higgs mechanism, describes how weak bosons and fermions acquire mass by introducing an additional scalar field doublet.

Lagrangian of a scalar field is composed of the kinematic term and the potential term. When the additional scalar field doublet, ϕ, is introduced, the associated Lagrangian,$\mathcal{L}_{\phi}$, is defined in terms of the kinetic term and the potential term, $V(\phi)$, as:

$$V(\phi) = \mu^2 \phi^{\dagger}\phi + \lambda(\phi^{\dagger}\phi)^2 \tag{2.3a}$$

$$\mathcal{L}_{\phi} = \frac{1}{2}(D_{\mu}\phi)^{\dagger}(D^{\mu}\phi) - V(\phi). \tag{2.3b}$$

In these equations, the parameters λ and μ^2 define the behavior of the potential. For the potential to be physical and bounded from below, λ is required to be real and positive whereas μ^2 can be positive or negative. Positive μ^2 results in the trivial case where there is only one choice for ϕ which gives the minimum (vacuum expectation). Negative μ^2 results in the case, referred to as the Mexican Hat, where the potential $V(\phi)$ has a local maximum at the center and has a ring of minima around it. Since the physical vacuum state will correspond to a particular point on this ring, one specific solution is chosen which minimizes the potential and this choice breaks the symmetry. This symmetry breaking and considering expansion around the vacuum introduces the Higgs field and an additional physical scalar particle with spin-0 in the SM, the Higgs boson. Fermions and weak bosons of the SM acquire mass by interacting with the Higgs field. The mass of the Higgs boson is related to this expansion around the vacuum state.

Strong interaction

The strong interaction is described by the quantum chromodynamics (QCD). The gauge group of QCD is $SU(3)_C$ and the related quantum number is the color charge. There are three color

charges: red (r), green (g) and blue (b). The fundamental representation of $SU(3)_C$ is the six color triplets of right-handed and left-handed quarks. The $SU(3)_C$ group has eight generators which correspond to QCD's massless spin-1 force carriers, the gluons. Since $SU(3)_C$ is a non-abelian group, gluons carry color and are self-interacting. Consequently, the theory includes gluon-fermion and gluon-gluon interactions. The strong interactions are vectorial and parity-conserving. One coupling g_s defines the strength of the interaction where the strength is independent of the particular color of the quark or the gluon. Two distinct properties of QCD are color confinement and asymptotic freedom. These properties are described briefly.

Particles which carries color charge are not observed directly individually but observed in hadrons such as proton. Color confinement of QCD is a hypothesis which is proposed to explain this observation. It states that the colored objects are always confined to colorless (singlet) states and that no particle which carry color charge can propagate as a free particle. This feature is also related to the running coupling constant of the QCD.

At the energy scale of an interaction, q, the strong coupling constant $\alpha_s(q^2)$ is used to characterize the strength of the interaction. It is observed that $\alpha_s(q^2)$ decreases as a function of energy and equivalently increases as a function of distance. At low energies $\alpha_s(q^2)$ is too large that the traditional perturbation approach cannot be applied. In this regime, lattice QCD is used to study the strong interaction. At such low energies, high coupling constant results in the confinement of quarks and gluons into bound state. At the high energies, $\alpha_s(q^2)$ gets smaller, and goes down to ~ 0.1 at $|q| > 100$ GeV. Hence, at the high energy hadron collisions, the strong coupling becomes very weak and the quarks within the hadrons can be considered quasi-free particles. This is referred to as the asymptotic freedom.

Another important phenomena is the hadronization. Although quarks and gluons can't exist individually in nature, they are produced freely at colliders and are hadronized. Consequently, the quarks and gluons are not directly observed at the detector. The precise process of hadronization is still not very well defined. It can be understood by the qualitative description that when the distance between two colored particles increases, the coupling strength also increases. Once the energy stored in this interaction is sufficient to produce new particles, these are produced if it is favorable. This process is iterated until the resulting particles have low energy to form hadrons.

2.1.3 Heavy bosons and top quark

The W,Z, Higgs bosons and the top quark are the heavy fundamental particles of the SM which also have short lifetimes. When they are produced, they decay very rapidly and are consequently only indirectly observed from their decay products. In this thesis, hadronic decays of W,Z, Higgs bosons and top quarks are of utmost importance. To conclude this chapter, some properties of these particles and their relevant decay modes are briefly described. The quoted numbers such as the mass and branching ratios of the particles are obtained from the Particle Data Group [12].

W and Z bosons

The W and Z bosons were discovered in 1983 by UA1 and UA2 experiments [13–15]. Since their discovery, studying W and Z bosons has been playing an important role at testing the consistency of the SM. The W boson has a mass of 80.379 ± 0.012 GeV and the Z boson has a mass of 91.1876 ± 0.0021 GeV [12].

The W boson decays to a lepton, neutrino pair (leptonic) or to a quark, antiquark pair (hadronic):

$$W^+ \to \bar{l}\,\nu \,,\, W^+ \to \bar{q}\,q' \,.$$

The dominant decay mode is the hadronic decay which is observed in $\sim 67\%$ of the cases. The Z decays to fermion, anti-fermion pairs:

$$Z \to l\bar{l} \,,\, Z \to \nu\bar{\nu} \,,\, Z \to q\bar{q}.$$

The dominant decay mode is the hadronic decay that is observed in $\sim 70\%$ of the cases. The Feynman diagrams of W and Z decays are presented below.
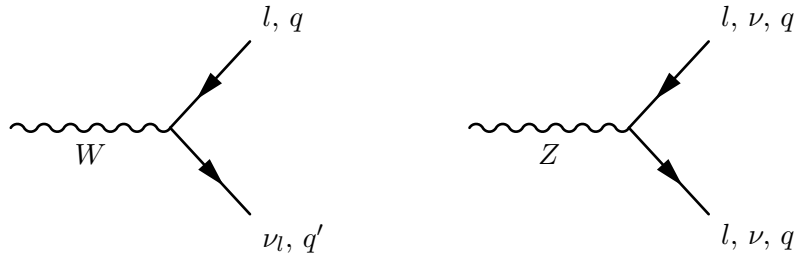
Figure 2.2: Feynman diagrams of W and Z decays.

Higgs boson

The Higgs boson was discovered in 2012 by the ATLAS and CMS experiments [16, 17]. Since its discovery, the Higgs boson has been thoroughly investigated to test the experimental observations' compatibility with the SM and its electroweak symmetry breaking component. All measurements of the Higgs boson properties are so far indicating that the observations are compatible with the theoretical predictions. The mass of the Higgs boson is a free parameter of the SM and is determined experimentally as 125.10 ± 0.14 GeV [12]. The strength of the Higgs boson interactions increases with the mass of the particle that it couples to. This property of the Higgs boson is reflected in its production and decay mechanisms. Some of the important Higgs boson decay modes are

$$H \to l\bar{l} \,,\, H \to \nu\bar{\nu} \,,\, H \to q\bar{q} \,,$$

$$H \to W^+W^- \,,\, H \to ZZ \,,$$

$$H \to \gamma\gamma$$

where W and Z bosons decay further according to their decay modes described above. Its most dominant decay modes are the decays into b quark ($\sim 58\%$) and W boson($\sim 21\%$) pairs. The Feynman diagrams of dominant decay modes are presented below.
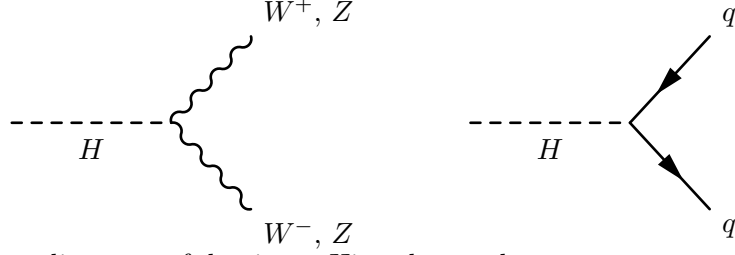
Figure 2.3: Feynman diagrams of dominant Higgs boson decays.

Top quark

The top quark was discovered in 1995 by the CDF and D0 experiments [18, 19]. The top quark is the heaviest fundamental particle of the SM with 172.9 ± 0.4 GeV mass [12] and is very different than the other quarks. It has a very short lifetime and it decays rapidly by first order weak interaction before any hadronization takes place. The top quark decays to a W boson and a b quark in more than 99% of the cases. The W boson then decays leptonically or hadronically resulting in two decay modes

$$t \to W^+ b \ \to \bar{l}\,\nu\,b\,,$$

$$t \to W^+ b \ \to \bar{q}\,q'\,b\,.$$

The Feynman diagram of these decay modes is presented below.
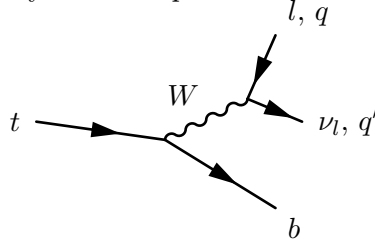
Figure 2.4: Feynman diagram of dominant top-quark decays.

2.2 Beyond the Standard Model

The SM is developed based on series of experimental discoveries and observations. The SM predictions have been tested by a large number of experiments. The predictions and the experimental observations are found to be compatible. However, there are observations at different astrophysical length scales that cannot be explained by the SM. Moreover, the SM has various parameters and fundamental assumptions which lead to theoretical limitations. In this section some of these limitations and a small selection of theories beyond the Standard Model are presented.

2.2.1 Limitations of the Standard Model

Observations in large astrophysical systems with different scales indicate that the luminous matter predicted by the SM is insufficient to account for the observed gravitational effects. Examples of such observations are the rotational velocity curves [20], gravitational lensing, the temperature and polarization anisotropies of the Cosmic Microwave Background [21]. Different hypotheses have been formulated to understand the nature of the unexpected observations. One of these approaches suggest that there exists a large amount of dark matter in the Universe which interacts with the SM particles only through gravity. Although dark matter is thoroughly searched for by direct and indirect detection experiments, it is still not observed.

In the SM, neutrinos are predicted to be massless. However, observation of neutrino oscillations showed that this is not the case in nature [22]. Due to non-zero neutrino masses and neutrino mixing [2], neutrinos can change their flavor in time. This phenomena is referred to as the neutrino oscillations. Observation of neutrino oscillations and non-zero neutrino masses challenge the SM. Non-zero neutrino masses and different ordering between the neutrino mass eigenstates are predicted by some of the theories beyond the SM. There are other fundamental open questions such as the nature of the neutrinos (Dirac or Majorana) and theories beyond the SM are based on different assumptions. Hence, understanding the origin of the neutrino masses and the hierarchy between them are fundamental to test models beyond the SM.

The experimental data so far indicate that in the SM there are three generations of fermions. From the theoretical point of view, the choice of three generations is not favored compared to any other choice. The reason behind it and the existence of other generations are open questions.

One of the fundamental questions of the SM is related to the measured mass of the Higgs boson. From the theoretical point of view, it is not yet well understood why the mass of the Higgs boson is ~ 125 GeV and this choice is not considered to be natural. A given theory is said to be natural if its underlying parameters are all of the same order unless there is a good explanation. There are several naturalness problems in the SM, and one of them is related to the hierarchy between the Planck scale and the electroweak scale. The hierarchy between the Planck scale and the electroweak scale can be quantified by considering the ratio of the Planck mass $(M_{\mathrm{Pl}}) \sim 10^{19}$ GeV and the Higgs mass (m_H). The Higgs mass receives quantum-loop corrections (contributions) from other SM particles it couples to. Although these corrections to the Higgs mass are predicted to quadratically diverge, the measured mass of the Higgs boson is very low. The ratio m_H/M_{Pl} is observed to be much smaller than 1, which introduces a hierarchy and a naturalness issue to the SM. Hence, the measured mass of the Higgs boson indicate that either some of the SM parameters are fine-tuned or other degrees of freedom cancel the large SM quantum-loop corrections in its calculation. There are many theories beyond the SM which address this issue, such as the Little Higgs [23], Composite Higgs [24, 25] and Supersymmetry [26]. In order to cancel the divergences top partners can be introduced. In this thesis, a class of top

---

[2]The neutrino flavor eigenstates and mass eigenstates do not coincide. The flavor of a neutrino is determined as a superposition of the three mass eigenstates at a given time, referred to as the neutrino mixing.

partners, vector-like quarks, which appear in some of these models are searched for. Vector-like quarks are introduced in the next section.

2.2.2 Vector-like quarks

Vector-like quarks (VLQs) are commonly produced in beyond the SM. Little Higgs and Composite Higgs models are widely studied models which address the naturalness of the SM and include VLQs. In these models, the Higgs mass additionally receives loop contributions from the hypothetical VLQs, which cancel the diverging contributions from the SM particles. Coupling between the Higgs boson and the VLQs depends on the model.

VLQs are hypothetical color triplet, spin-half fermions. Unlike the SM quarks, VLQs' left-handed and right-handed components have the same electroweak transformations, making them *vector-like*. VLQs can mix with the SM quarks and consequently modify the couplings to the massive bosons of the SM [27]. The VLQs are expected to mix strongly with the third generation SM quarks and consequently are called top partners. Although mixing with the lighter SM quarks is not forbidden, in many models the VLQs are assumed to mix with the third generation SM quarks for simplicity. The VLQ models used in this thesis also use this assumption. The mixing of the VLQs with the SM quarks allows the VLQs to decay into a pair of a third-generation quark and a Higgs or a gauge boson.

In this thesis, four different types of heavy vector-like quarks are considered. Two of them are vector-like T and B quarks, which have the same charge as their SM partners top $(2/3)$ and b $(-1/3)$ quarks. The other two are vector-like X and Y quarks with charge $+5/3$ and $-4/3$, respectively. Depending on the model, these four VLQs can appear as weak isospin singlets, doublets or multiplets. In this thesis, only the singlets (isospin 0) and doublets (isospin 1/2) are considered:

- Singlets: T, B

- Doublets: (X, T), (T, B), (B, Y).

At the Large Hadron Collider, heavy VLQs can be single or pair produced. The single production of the VLQs is expected to become more favorable for VLQs which have mass higher than ~ 1 TeV, however, the single production cross-section is considered to be more model dependent than the pair production. The cross-section of the single production depends on the coupling of the VLQ to the SM particles (third generation quarks and massive bosons) and the mass of the VLQ. The cross-section of the pair production depends only on the mass of the VLQ. In this thesis, a search is carried out for pair-produced heavy vector-like quarks, $T\bar{T}$ and $B\bar{B}$.

All considered VLQs can be pair produced through the decay of a virtual gluon. Once they are produced, VLQs decay to a third generation SM quark accompanied by a W, Z or Higgs boson. The branching ratios of different decay modes are not known and they depend on the

model that is considered. In this thesis, the following decay modes are considered:

$$T \to Wb,\, T \to Zt,\, T \to Ht\,,$$

$$B \to Wt,\, W \to Zb,\, W \to Hb.$$

Chapter 3

Machine Learning for Classification

A machine learning algorithm is an algorithm that is able to learn from data to solve a practical problem [28, 29]. Machine learning techniques are used in different branches of science to handle complex problems and large datasets. Over the past decades, highly optimized advanced algorithms have been introduced. Supervised deep learning algorithms are one of these techniques. In this chapter an overview of machine learning techniques is provided, with a particular emphasis on the classification algorithms using neural networks since those will be widely used in this thesis.

3.1 Machine Learning Basics

Machine learning (ML) algorithms learn from data to solve a problem by gathering an input dataset and algorithmically building a statistical model based on that dataset [29]. Goals of machine learning algorithms include building a very accurate model or speeding up a computationally-expensive model building or evaluation process even though it may require sacrificing the accuracy of the predictions. In this section ML basics are covered which are essential to understand the ML algorithms used throughout this thesis.

3.1.1 Task type

There are many types of tasks which can be solved with dedicated ML algorithms. Classification and regression are some of the most common tasks. Classification has the aim of assigning a label to an unlabeled dataset where the label is a member of a finite set of classes. Hence, generally a discrete valued output is expected from a classification algorithm. Regression on the other hand is a task where the algorithm predicts a real-valued continuous label.

In this thesis, ML algorithms are applied to solve classification problems. There are two main types of classification problems: binary and multi-class. A widely used example of a binomial classification algorithm is in e-mail spam detection where each e-mail is classified as spam or not-spam. Widely used example of a multi-class classification is handwritten numerical digit recognition where the classes go from zero to nine.

3.1.2 Types of machine learning

Machine learning algorithms can be categorized into three main types based on the information and type of experience they are allowed to have during the learning process: supervised, unsupervised and reinforcement learning. In unsupervised learning, each example in the dataset contains input features, which are also referred to as input variables. The useful properties of a dataset are learned by the algorithm without any labels. In supervised learning, each example in the dataset does not only contain features but is also associated with a label (or target). The goal of the supervised learning is to use this dataset to build a model which can predict the labels well. Supervised and unsupervised learning can be used for semi-supervised learning. The goal of the semi-supervised learning is generally the same as the supervised learning, however, the dataset contains both labeled and unlabeled examples. In some cases, semi-supervised learning is expected to improve the statistical model built by the ML algorithm. Reinforcement learning is different than supervised and unsupervised learning because although there are no labels, an agent explores the environment and updates its policy depending on the responses.

3.1.3 Supervised learning for classification tasks

It is important to give an overall idea of how supervised machine learning classification algorithms generally work before going into details of specific algorithms. In supervised learning, the goal of the ML algorithm is to build a model which takes the input features represented by the vector $\mathbf{x}$ and output the target $\mathbf{y}$. This can be represented by

$$\mathbf{x} \xrightarrow{\mathcal{M}} \mathbf{y} \tag{3.1}$$

where $\mathcal{M}$ is the statistical model built by the ML algorithm. Hence, for supervised learning it is essential to first collect the training data which is a collection of input ($\mathbf{x}$) and output ($\mathbf{t}$) pairs. As an example, in the case of spam detection, the dataset consists of a set of e-mails where the input is the received e-mail and output is the spam or not-spam label of each e-mail. The first step is to collect a sample of e-mails which are already labeled as spam/not-spam. The second step is to process the e-mails and their labels such that they can be interpreted by the machine learning algorithm as a set of features[1]. After preparing this labeled dataset, the learning procedure can be started. Generally, $\mathcal{M}$ has an initial starting point. Since the model $\mathcal{M}$ cannot map the input features to the target perfectly, the mapping results in a quantifiable error. During learning, an optimization algorithm updates the model to minimize the error and improve the accuracy of the model.

---

[1]An example of such processing is defining output binary variables for the spam/not-spam labels and input binary variables which indicate if a word is included in an e-mail or not.

3.1.4 Inputs to the supervised ML algorithms

The task of transforming raw data into a collection of labeled examples is essential for the ML algorithms. This task generally requires domain knowledge. For example in experimental particle physics, signatures in the detector are analyzed to build physics objects such as jets. Jets have engineered features such as the jet mass and other jet moments introduced in section 6.4. These jet moments contain useful information which can be used to classify each jet. Moreover, in simulation, jets can be labeled as signal or background depending on their truth properties as it will be described in detail in section 6.3. Once the labeled examples are collected, in order to prepare the data for ML algorithm training and testing, data is split into several subsets, some criteria are applied on these subsets and transformations are applied to the input features, referred to as pre-processing. These necessary steps are described in this section.

Datasets and pre-processing

It is important to test the performance of a model built by the ML algorithm on a set of examples which are different than the set of example that was used to train the ML algorithm. Hence, in general there are three sets used: training set, validation set and testing set. These sets are obtained by randomly splitting the labeled dataset into three subsets. The training set is used by the ML algorithm to optimize the model and it is generally the largest set. The performance of different algorithms are evaluated using the validation and testing sets. The validation set is generally used to choose the learning algorithm and its configuration as well as to ensure that the model does not only predict the training data well but also works well on unseen data. The testing set is used to evaluate the final performance.

In experimental particle physics, it is common to use only training and testing sets for supervised learning. When the training and testing sets are obtained from the simulation, the final performance test of the ML algorithms are carried out in the collected data, which has not been seen by the ML algorithm during its training or optimization. Hence, the collected data can be interpreted as the ultimate testing set.

Undefined or missing input features

It is common for some datasets to have undefined or missing input features for some its members. There are several ways of handling this problem. If the fraction of such examples are low and if the dataset is big enough, a simple option is to remove the examples with missing features from the training set. If the fraction of such examples is not low, it is important to either use a learning algorithm which can handle missing feature values or varying input feature size or to assign values to the missing features.

Input transformations

Often the numerical range of the input variables are different. Hence, for some algorithms like deep neural networks, the numerical inputs are transformed to have a similar range thus helping the algorithm. There are several commonly used transformations. One of them is the normalization where the input feature values are transformed to have the same range, typically [-1,1] or [0,1]. Another commonly used transformation is the standardization where the input values are shifted and rescaled such that their distributions are centered at 0 (mean $\mu = 0$) and their standard deviation is 1 ($\sigma = 1$). The values needed to apply the transformation are obtained from the training set and are applied to the validation and testing sets consistently.

3.1.5 Learning algorithms

The statistical model built to solve a problem consists of trainable parameters such as the weights and biases of a neural network. A learning algorithm's task is to minimize or maximize an objective function with respect to the trainable parameters. The building blocks of a learning algorithm can be summarized as an objective function, an optimization criterion based on the chosen objective function (minimization or maximization) and an optimization routine extracting information from the training data, where the optimization routine defines how the trainable parameters of the model are updated. The procedure of optimizing the parameters of the model is referred to as the training.

The loss function (or cost function) is a real-valued function which is related to the error of the model with respect to the objective, and is minimized to optimize a model. Gradient-descent based optimization algorithms are frequently used where the loss function is differentiable. They are iterative optimization algorithms which find the minimum of a loss function by using its gradient. Widely used gradient-based algorithms include the gradient descent and stochastic gradient descent algorithms. The gradient descent algorithm starts at a random point defined by a random initialization of the function parameters, and updates the function by taking steps proportional to the negative of the gradient of the function at the current point, thus moving towards the minimum. The gradient descent algorithm proceeds in epochs where an epoch consists of one pass over the entire training set. Hence, number of epochs corresponds to the number of passes over all the training set. The gradient descent algorithm evaluates and updates the weights for each epoch. In contrast to gradient descent, the stochastic gradient descent (SGD) algorithm estimates the gradient of the training set by evaluating the gradients of random subsets of the training dataset, referred to as batches. The gradient is calculated for each batch and the parameters of the model are updated accordingly. The configurable settings of the learning algorithm, such as the batch size, are referred to as hyper-parameters.

3.1.6 Hyper-parameters

Most machine learning algorithms have settings which are determined to control the algorithm's behavior. They are referred to as the hyper-parameters. The hyper-parameters are not learned by the training but are defined by the user before running the algorithm. It is generally important to find the optimal hyper-parameters for a specific problem and dataset, which requires an optimization. There are different ways of doing this optimization; one simple approach is to conduct a grid search where one hyper-parameter is varied at a time and the performance of the resulting models are compared in order to find the best set of hyper-parameters.

3.1.7 Underfitting and overfitting

In supervised learning, if a trained model cannot predict well the targets of the training set, it means that the model underfits. If a trained model predicts very well the targets of the training set but it cannot predict well the targets of the validation or testing set, the model overfits (or overtrains). Overfitting can be a risk for both very simple and advanced machine learning models. Some reasons for overfitting include employing models with too much complexity, too many trainable parameters or having too many features to be learned with a small training set. Finding good hyper-parameters is important to avoid underfitting and overfitting.

3.2 Deep Neural Networks

A neural network (NN) is a type of machine learning model which maps input features of a dataset to an output using a mathematical function. Neural networks are organized in layers which consist of processors that are called neurons or nodes [30]. They can learn and model non-linear and complex relations between the input features to solve a classification or a regression problem. In this thesis, neural networks are used for classification.

3.2.1 Neural networks

A neural network is a mathematical function which maps the input features represented by the vector $\mathbf{x}$ to the output $\mathbf{y}$. In analogy of the Eqn. 3.1, such neural networks can be defined as

$$\mathbf{x} \xrightarrow{f_{\mathcal{NN}}} \mathbf{y} \tag{3.2}$$

or equivalently as

$$\mathbf{y} = f_{\mathcal{NN}}(\mathbf{x}) \tag{3.3}$$

where $f_{\mathcal{NN}}$ is a nested function which approximates a function that perfectly maps $\mathbf{x}$ to the truth target $\mathbf{t}$.

There are many different types of neural networks which are used to tackle different problems. Feed-forward neural networks are important, commonly used models which form the basis for

many applications. In feed-forward neural networks, the function receives the input $\mathbf{x}$, performs intermediate computations defined by $f_{\mathcal{NN}}$, and finally gives the output $\mathbf{y}$. The middle layers which are not input or output layers are referred to as hidden layers. A neural network's depth is characterized by the number of hidden layers it has. If a neural network has more than one hidden layer, it is referred to as a deep neural network (DNN). In this thesis, feed-forward neural networks with fully connected layers, in which the outputs of the previous layer are inputs to the next layer, are used[2]. For the sake of brevity, this section will focus on feed-forward neural networks with fully-connected layers.

As an example, a DNN with an input layer, two hidden layers and an output layer can be described as

$$\mathbf{y} = f_{\mathcal{NN}}(\mathbf{x}) = \mathbf{f}_3(\mathbf{f}_2(\mathbf{f}_1(\mathbf{x}))) \tag{3.4}$$

where each $\mathbf{f}_l$ is defined as

$$\mathbf{f}_l(\mathbf{z}) = \mathbf{g}_l(\mathbf{W}_l\mathbf{z} + \mathbf{b}_l). \tag{3.5}$$

In Eqn. 3.5, l is called the layer index, the function $\mathbf{g}_l$ is called an activation function, and $\mathbf{W}_l$ and $\mathbf{b}_l$ are the neural network's weights and bias in layer l. The components of $\mathbf{W}_l$ and $\mathbf{b}_l$ are the parameters of the DNN which are learned during training. In Fig. 3.1 a feed-forward neural network with a two-dimensional input, two hidden layers and one-dimensional output layer is presented. The neural network is represented graphically as a connected combination of nodes of each sequential layer where the arrows connecting the nodes represent the inputs (inbound arrows) and outputs (outbound arrows) of the nodes. Each input feature is represented by a circle, while each node where an activation function is applied is presented as a box. The output of the each node is the result of the function written inside the box where $\mathbf{g}_l$ is the chosen activation function of each layer[3] and $\mathbf{w}_{l,i}$, $\mathbf{b}_{l,i}$ are components of $\mathbf{W}_l$, $\mathbf{b}_l$.

3.2.2 Deep neural networks and training essentials

In this section, the building blocks of deep neural networks and methods used for training a robust model are described in more detail.

Weight initialization

Since DNN learning is an iterative process, the DNN parameters are required to have initial values to start the process. Stochastic optimization algorithms used for DNNs, like SGD, start at a random point as explained in section 3.1.5. It was observed that finding a good random initialization can impact the performance of the DNNs. Hence, weight initializers such as Glorot uniform (also called Xavier uniform) [31] and He normal [32] were introduced.

---

[2]This neural network architecture is also referred to as multilayer perceptron (MLP).
[3]It is not necessary to use the same activation function in all the nodes of a layer.
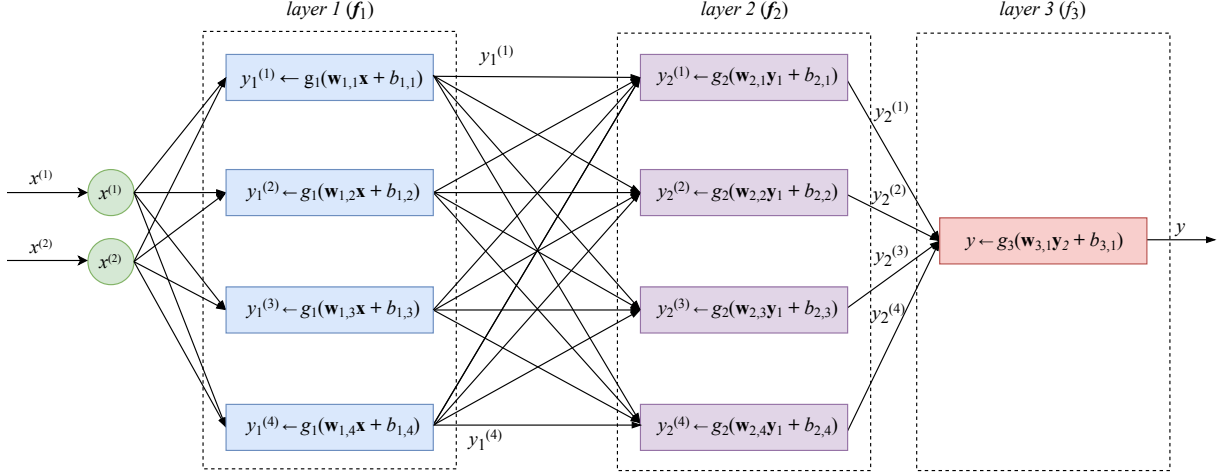
Figure 3.1: A feed-forward neural network with two-dimensional input, two hidden layers and one-dimensional output layer [29].

Activation functions

Activation functions define the basic units of neural networks by determining the output of a node based on its input. They play an important role in extracting non-linear correlations between the input features and obtaining the output appropriate for the problem. Moreover, the choice of the activation function affects how fast a model converges towards the optimum model.

The dimension (number of nodes) and the activation function of the output layer are chosen based on the application. If the DNN is used to build a binary classifier, the output layer generally has one node which tells if an example should be labeled as signal or background. Since the input features received by the DNNs are generally continuous, and DNNs employ functions $(\mathbf{g}_l(\mathbf{W}_l\mathbf{z} + \mathbf{b}_l))$ which also produce continuous outputs, the output layer needs to map this information into a probability to estimate how likely each example is to be signal or background. This operation which converts continuous signals into binary output is called a logistic regression. The sigmoid function is the standard logistic function and is defined as,

$$g(z) = \frac{1}{1 + e^{-z}}. \tag{3.6}$$

The graph of the sigmoid function is presented in Fig. 3.2. For binary classification, the sigmoid is generally used as the activation function of the output node since it is continuous and differentiable. Binary classification using sigmoid can be extended to multi-class and multi-label classification problems by using either softmax or sigmoids (allowing multiple labels per example).

In modern neural networks, non-linear sigmoids, hyperbolic tangent function and rectified linear units are widely used as activation functions for the hidden layers. The usage of non-linear activation functions is required to extract non-linear information since all of the other DNN operations are linear. Hidden layers with sigmoid functions are generally harder to train
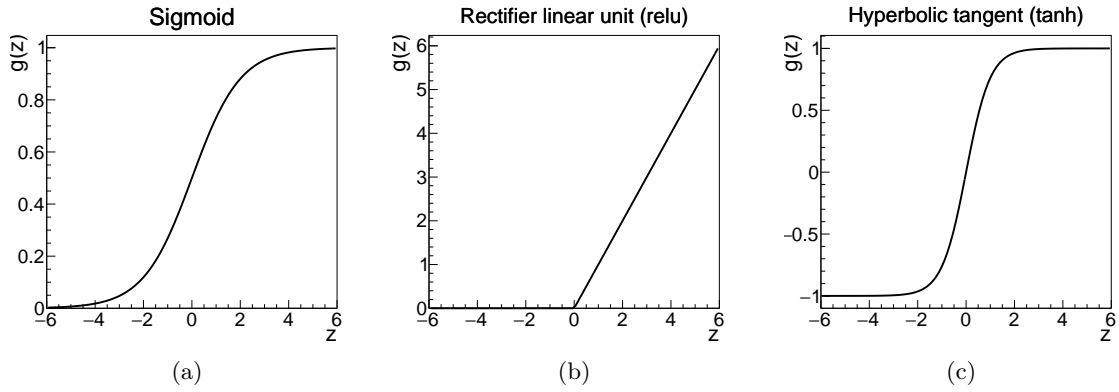
Figure 3.2: Sigmoid (a), rectified linear unit (b) and hyperbolic tangent (c) activation functions.

compared to the other two functions due to vanishing gradient[4] and computationally expensive nature. The hyperbolic tangent function is an activation function which compresses all of its inputs to the range [-1, 1]. It is defined as,

$$g(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} = \frac{2}{1 + e^{-2z}} - 1. \tag{3.7}$$

It can be seen in Fig. 3.2 that tanh has a similar behavior to the sigmoid. Rectified linear units (relu) use the function $g(z) = max(0, z)$. This means that if the input is positive, the output is equal to the input value itself, while if the input is negative the output is zero. Relu is one of the most preferred functions since it avoids the vanishing gradient problem and it is not computationally expensive. Overall it is easier to train and it converges faster than sigmoid and tanh. The graphs of the tanh and relu activation functions are also presented in Fig. 3.2.

Maxout units, which can generalize different functions, are widely used. Maxout units divide the input vector $\mathbf{z}$ into groups of k values and gives the maximum element of one of these groups as output. This provides a way to learn the activation function itself rather than restricting it to a pre-defined relationship between nodes.

Training and backpropagation

While training feed forward deep neural networks, the neural network accepts the input features and produces an output where a loss is calculated, this is called the feed forward propagation. The backpropagation algorithm receives the information of the loss and propagates it backward through the network to compute the gradient. Once the gradient is computed, an algorithm such as SGD is used to optimize the parameters of the neural network. Training using SGD can be presented by the following mapping and optimization step. Starting from Eqn. 3.3, at a given iteration, the neural network's output results in $\mathbf{y} = f_{\mathcal{N}\mathcal{N}}(\mathbf{x})$. As the neural network cannot map the input to the target perfectly, the mapping results in an error, which the neural

---

[4]As it can be seen in the Fig. 3.2, the sigmoid function becomes constant for very large and small values of z. This causes the gradient to be zero at such values, referred to as the vanishing gradient problem.

network tries to minimize during training. In order to optimize the model, the algorithm first calculates the gradient of a loss function ($\nabla_\theta \Phi$) with respect to all the weights of the network, where loss function Φ is a chosen function of **y** and target **t**, and gives information about the resulting cumulative error of up-to-now considered training instances[5]. The gradient is then used to update the weights aiming to minimize the loss function. If the set of weights (w, b) is referred to as θ and the set of weights at iteration t is referred to as θ_t, then the update rule of a network can be described as

$$\theta_{t+1} = \theta_t - \Delta\theta_t \tag{3.8}$$

with

$$\Delta\theta_t = \gamma\nabla_\theta\Phi \tag{3.9}$$

where γ is the learning rate that defines the step size of the update. Since a learning rate too small can lead to very slow convergence and a learning rate too large can cause the loss to fluctuate around a local minimum, the learning rate needs to be optimized for the problem and the training set in order to achieve the best model. The SGD algorithm has shortcoming such as slow convergence and the difficulty choosing the optimal learning rate. To overcome the shortcomings of the gradient-based optimization algorithms, new methods have been proposed. One such method is the introduction of the momentum where information from the past iterations is added to the current update. The new update rule for SGD with momentum can be described by defining another $\Delta\theta_t$ as,

$$\Delta\theta_t = (1 - \mu)\gamma\nabla_\theta\Phi + \mu\Delta\theta_{t-1} \tag{3.10}$$

where μ is the momentum hyper-parameter which is defined as a fraction. Introducing momentum helps neural networks to be more resistant to noise in the training sample and reduces some of the issues, but it introduces another hyper-parameter. It was observed that using a separate learning rate for each parameter and automatically adapting these learning rates during learning could also provide good performance. Hence, adaptive learning methods were proposed and they were observed to perform better in some cases.

A method for stochastic optimization (Adam) [33] is one these methods. It is a method for first-order gradient based optimization of stochastic loss functions, based on adaptive estimates of lower order moments. The method is aimed towards machine learning problems with large datasets or high dimensional parameter spaces. It only requires first order gradients and it computes individual adaptive learning rates for different parameters ($\theta : (w, b)$) from estimates of the first and second moment of the gradients. It naturally performs a form of step size annealing and it automatically adjusts the two important training parameters: learning rate and momentum. Adam was tested on different machine learning problems and for most of the problems it provides good default hyper-parameters settings.

---

[5]In the calculation of the loss Φ, each example in the training and validation set can contribute to the loss with coefficients different from 1, these coefficients are referred to as sample weights or training weights in section 7.2.1.

Regularization techniques

It is essential that a trained model performs well not only on the training data, but also on new inputs. Many methods are developed and used in machine learning to accomplish this, although sometimes it slightly worsens the performance on the training data. These methods, which are used to avoid overfitting and improve the generalization, are collectively referred to as regularization. Some of the regularization techniques which are used in this thesis for neural networks are briefly described here.

One approach to regularization is to add a penalizing term while calculating the loss. L1 and L2 regularizers [28] are examples of such regularization techniques. An L2-regularized loss function includes a penalty term which is proportional to the L2-regularization hyper-parameter and the sum of squared values of the weights. In contrast, an L1-regularized loss function includes a penalty term which is proportional to the L1-regularization hyper-parameter and the sum of absolute values of the weights. L1 regularization can perform feature selection and produce a sparse model by allowing the weights to be reduced to zero.

Batch normalization [34] is a technique which transforms the outputs of each layer in order to standardize the outputs. It reduces the shift in the hidden layer node output values and helps the neural network to converge faster, and to improve its performance and stability. Although batch normalization was not introduced as a regularization technique, it also has a regularizing effect.

Finally, early stopping can be used to avoid overfitting by controlling the number of epochs. While training neural networks, the loss development of the training set and validation set are monitored as a function of epochs. Since the learning algorithm aims to minimize the loss of the training set, it can be observed that after certain number of iterations, the loss of the validation set starts to increase although the loss of the training set is still decreasing. This behavior implies that the model predicts the training set well but it can no longer predict the validation set as well as it could in the previous iterations. Hence, overfitting can be detected by investigating the loss of the validation set. Early stopping makes use of this information by ending the training procedure if the loss of the validation set does not improve for a chosen number of epochs.

3.2.3 Software

Keras [35] is a high-level neural network application program interface. It can be run on top of several machine learning libraries such as Theano [36]. In this thesis, neural networks are trained and optimized with Keras using a Theano backend. Unlike core ATLAS software, both Keras and Theano are developed in Python. Hence, lwtnn [37] is used to provide an interface which can predict the output of the Keras-trained models within the ATLAS framework.

3.3 Boosted Decision Trees

A boosted decision tree (BDT) is a type of supervised learning technique which consists of decision trees and a boosting algorithm. Although they can be used for regression or classification, in this thesis the focus will be on classification. A decision tree classifier is a binary tree-structured classifier which takes a set of input features and in each node of the tree, one input feature is used to make a decision. Based on this decision, the right or the left branch of the tree is followed. This is repeated until a stop criterion is fulfilled at which point a classification decision is made. A graphic representation of a boosted decision tree is shown in Fig. 3.3. Boosting algorithms employ several such decision trees to enhance and stabilize the performance of a classifier. BDTs are well established and widely used in ATLAS. This section provides an overview of BDTs and related concepts.

3.3.1 Decision tree

The training of a decision tree aims to find the optimal splitting criteria for each node. The training starts with the root node, where a decision criterion for the full training sample is determined. This decision splits the training set in two subsets which are delivered to the next node where the same algorithm defines the next splitting criterion for each node. At each node, the criterion is determined by identifying the variable and corresponding requirement which provides the best separation between signal and background. There are several separation criteria which provide good performance; one such criterion used in this thesis is referred to as the Gini Index. The Gini Index is an error metric which estimates the purity of the region by evaluating how much of the training data in a particular region belongs to a single class. The procedure of splitting is repeated until a configurable stop criterion is fulfilled, such as the minimum number of events in a node. The classification decision is made on this final node based on the fraction of signal and background training examples in that node.

3.3.2 Boosting

In boosting algorithms, multiple decision trees are derived from the same training set by reweighting the examples. The weighted average of the outputs of the trained decision trees is taken to provide a single classifier. There are several boosting algorithms which provide good performance, such as GradientBoost. In some cases GradientBoost can benefit from the introduction of a procedure which uses random subsets of the training set for growing the trees, similar to the batches introduced for the stochastic gradient descent. In this case, the algorithm is called stochastic gradient boosting and the number of training examples in the subsets is a hyper-parameter for which an appropriate fraction of events is chosen.
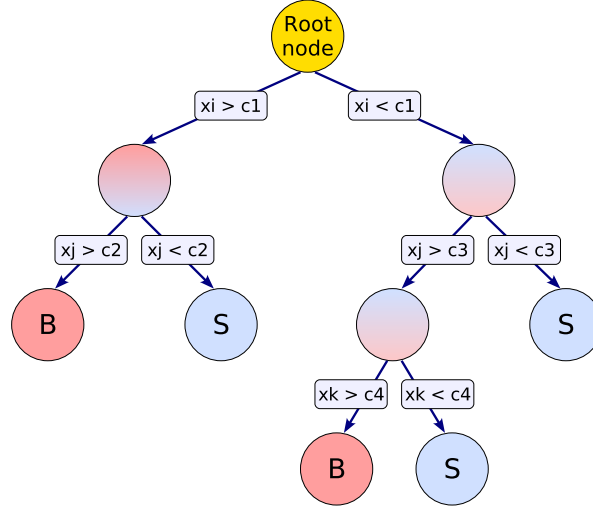
Figure 3.3:   A graphics representation of a binary decision tree [38].

3.3.3 Software

The toolkit for multivariate data analysis (TMVA) [38] is a toolkit which is integrated into the analysis framework ROOT [39]. TMVA hosts a large number of classification algorithms including numerous types of BDTs. The BDTs used in this thesis are implemented using the TMVA 4.2.1 package. More information can be found in Ref. [38] and the TMVA users guide.

Chapter 4

The ATLAS Experiment at the LHC

4.1 The LHC

The Large Hadron Collider (LHC) [40] is a circular, two-ring superconducting hadron accelerator and collider located at the European Organization for Nuclear Research (CERN) complex. The LHC is installed in a 27-km tunnel located underground near Geneva, Switzerland. It is designed to collide proton beams with an unprecedented center of mass energy ($\sqrt{s}$) of up to 14 TeV and produce very large datasets to study the SM and beyond. The LHC physics program ranges from more precise measurements of SM parameters to the search for new physics phenomena. In addition to the proton-proton (pp) collisions, the LHC is designed to produce heavy-ion collisions as well. In this thesis only the pp collisions are considered.

The accelerator complex at CERN is a chain of machines as shown in Figure 4.1. Each machine injects the beam into the next one and sequentially increases the energy. The LHC is the last ring in this chain of particle accelerators. The LHC beams circulating in opposite directions are collided at four interaction points where four major LHC experiments ALICE [41], ATLAS [42], CMS [43], LHCb [44] are located. The LHC was successfully operated to deliver pp collisions with a center of mass energy of 7 TeV during 2010 and 2011, 8 TeV during 2012; this period is referred to as Run 1. In next period of operation from 2015 to 2018, referred to as Run 2, the LHC delivered pp collisions with a center of mass energy of 13 TeV. Currently the LHC is in the long shutdown 2 (LS2) during which the machines and experiments are being upgraded in preparation for the next run referred to as Run 3. It is expected that the LHC will reach a center of mass energy of 14 TeV in Run 3. The Run 2 data is used in this thesis.

In order to produce pp collisions with a center of mass energy of 13 TeV, first the protons are obtained by stripping electrons from hydrogen atoms which are taken from a bottle containing hydrogen gas. The obtained protons are sent to the Linac2 where they are collected into bunches and accelerated to energy of 50 MeV. The proton bunches are then injected to the Proton Synchrotron Booster where they reach 1.4 GeV. Next, the protons are accelerated by the Proton Synchrotron to 25 GeV and by Super Proton Synchrotron to 450 GeV before they are injected to the LHC to be circulated in opposite directions. Finally the LHC accelerates the protons to 6.5 TeV. These accelerated protons circulating in opposite directions are collided at the interaction points.
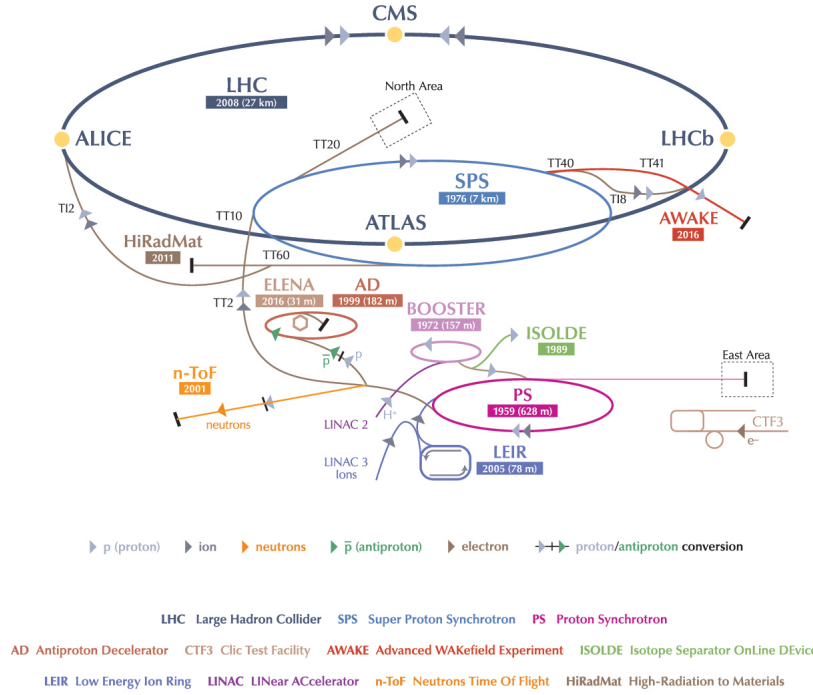
Figure 4.1: Illustration of the CERN accelerator complex and experiments. The LHC is the last ring in a chain of particle accelerators [45].

Large datasets are needed to study the SM with greater precision and to explore new physics beyond the SM. Therefore, in addition to the energy reach, the number of collisions per second is an extremely important property of an accelerator. For a process under study, the expected number of events per second is given by

$$dN/dt = \sigma \cdot L \tag{4.1}$$

where σ is the cross section for the process under study and L is the machine luminosity. The machine luminosity depends only on the beam parameters and for a Gaussian beam distribution it can be defined as

$$L = \frac{N_b^2 n_b f_{\text{rev}} \gamma_r}{4\pi \varepsilon_n \beta^\star} F \tag{4.2}$$

where N_b is the number of particles per bunch, n_b is the number of bunches per beam, f_{rev} the revolution frequency, γ_r the relativistic gamma factor, ε_n the normalized transverse beam emittance, $\beta^\star$ the beta function at the collision point and F the geometric luminosity reduction factor due to the crossing angle at the interaction point. Due to several factors such as the beam loss from collisions, the instantaneous luminosity is not constant over a physics run but decays.
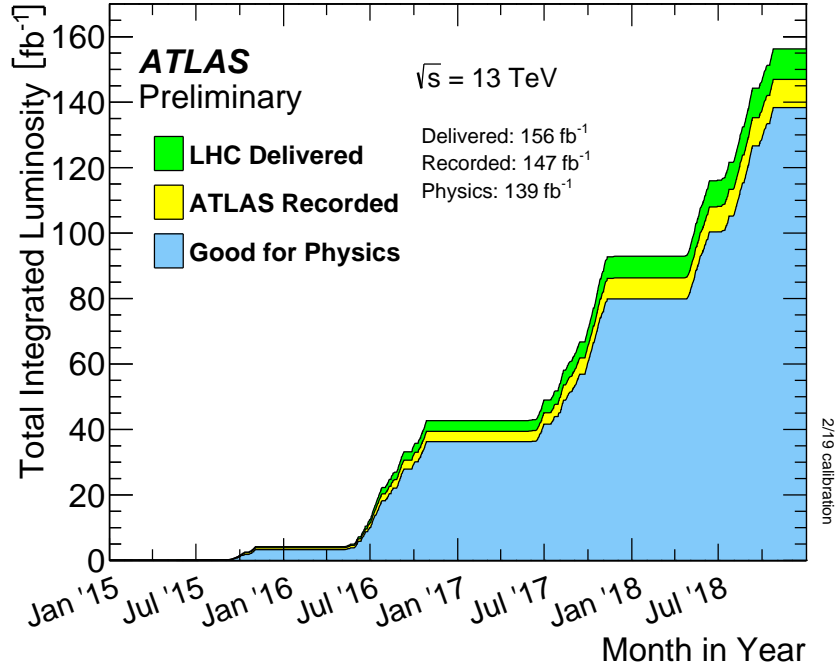
Figure 4.2: Total integrated luminosity versus time delivered to ATLAS (green), recorded by ATLAS (yellow), and recorded good quality data for physics (blue) during stable beams for pp collisions at $\sqrt{s} = 13$ TeV in Run 2.

The collected data size in a period of time is characterized by the integral of the instantaneous luminosity over that period of time and it is called the integrated luminosity, $\mathcal{L} = \int L \, dt$. Peak luminosity and integrated luminosity are two important values that characterize the performance of an accelerator. LHC reached the highest peak luminosity of $19 \cdot 10^{33} \, \mathrm{cm}^{-2} \, \mathrm{s}^{-1}$ and delivered integrated luminosity of $63.4 \, \mathrm{fb}^{-1}$ in 2018 as presented in Ref. [46]. Total integrated luminosity versus time delivered to ATLAS, recorded by ATLAS and recorded good quality data for physics during stable beams for pp collisions at $\sqrt{s} = 13$ TeV in Run 2 are presented in Fig. 4.2.

In order to obtain high luminosities, number of particles per bunch and the number of bunches per beam is increased, which leads to the occurrence of simultaneous pp collisions. The over-lapping of pp interactions from the adjacent bunch crossing is referred to as out-of-time pileup, whereas the overlapping of pp interactions from the same bunch crossing is referred to as in-time pileup. Pileup is either measured as the number of primary vertices, which does not consider out-of-time pileup, or as the average number of interactions per bunch crossing, which is based on the measured luminosity. Unless it is specified, both in-time and out-of-time pileup will be referred to as pileup in this thesis.

4.2 The ATLAS Detector

The ATLAS experiment [42, 47] is one of the two general-purpose, high-luminosity experiments of the LHC. It is designed considering the challenges of high interaction rates, radiation doses, particle multiplicities and energies. The search for the SM Higgs boson was used as a benchmark to establish the performance of important subsystems of ATLAS. The detector is forward-backward symmetric with respect to the interaction point. It has cylindrical geometry and it covers almost the entire solid angle around the interaction point.

4.2.1 Coordinate system

The ATLAS detector uses a right-handed coordinate system where the nominal interaction point is defined as the origin of the coordinate system. The x-axis points towards the center of the LHC, the z-axis is defined by the beam direction and y-axis is orthogonal to both such that the $x - y$ plane is transverse to the beam direction. The azimuthal angle ϕ is measured around the z-axis in the $x - y$ plane and the polar angle θ is the angle from the beam axis in the $x - z$ plane. The pseudo-rapidity is defined as $\eta = -\ln\left(\tan\left(\theta/2\right)\right)$ and it is generally used instead of θ as θ is not Lorentz invariant. Additionally, the rapidity is defined as $Y = \frac{1}{2}\ln\frac{E+p_z}{E-p_z}$ where E is the energy and p_z is the longitudinal momentum of the particle. The angular distance (ΔR) is used to quantify the distance between two objects and is defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$. Another important and widely used variable is the momentum of particles in the $x - y$ plane, the transverse momentum (p_{T}). It is defined as $p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2}$ where p_x and p_y are the momenta in the x and y directions.

4.2.2 Detector components

ATLAS is built in layers and it has three main subsystems: Inner Detector, calorimeter and muon spectrometer. The detector layout is shown in Figure 4.3. In this section brief descriptions of the subsystems are given.

Inner Detector

The Inner Detector (ID) is the innermost component of the ATLAS detector. It is surrounded by a superconducting solenoid magnet that provides a 2 T field in which the particle trajectories, also called tracks, are bent. It has complete azimuthal coverage and spans the pseudo-rapidity region $|\eta| < 2.5$. It records the particle trajectories and measures the direction, momentum, and charge of electrically-charged particles produced in collisions. Fine detector granularity is needed to achieve high momentum and vertex resolution required by the benchmark physics processes. Additionally, in order to minimize its impact on the energy measurement in the calorimeter, as little material as possible needs to be used. ID consists of four subsystems to be able to
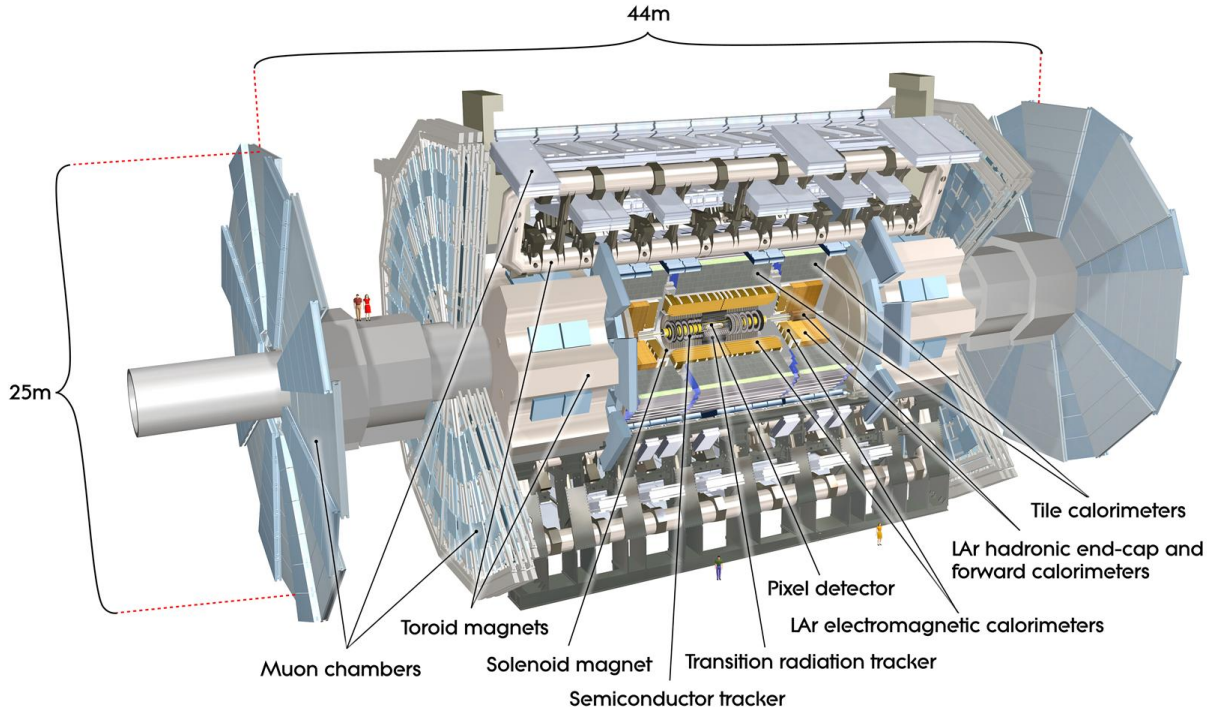
Figure 4.3: Cut-away view of the ATLAS detector [42].

satisfy the requirements. These subsystems are: Insertable B-layer (IBL)[1] [47], pixel detector, semiconductor tracker (SCT), and transition radiation tracker (TRT).

Calorimeter

The calorimeter measures the energy carried by particles, except muons and neutrinos. The AT-LAS calorimeter surrounds the ID and it is placed between the solenoid and toroid magnets. It consists of electromagnetic and hadronic sampling calorimeters with alternating layers of absorbing and active material. The calorimeter was designed such that the calorimeter provides good containment for electromagnetic and hadronic showers, and limits the amount of energy escaping the calorimeter, referred to as punch-through, into other systems. The ATLAS calorimeters have a highly granular lateral and longitudinal segmentation which allows for very precise energy flow reconstruction [48]. The ATLAS calorimeter system and its components are presented in Fig. 4.4.

When particles from the hard-scatter process interact with the calorimeter material, they create cascades of particles, referred to as particle showers. There are two types of showers, electromagnetic and hadronic showers. Electromagnetic showers are mainly initiated by electrons or photons, whereas hadronic showers are initiated by hadrons of different sources. These two types of showers have different properties and require different techniques for detection.

---

[1]IBL was installed between Run 1 and Run 2 to improve vertex reconstruction and B-hadron identification
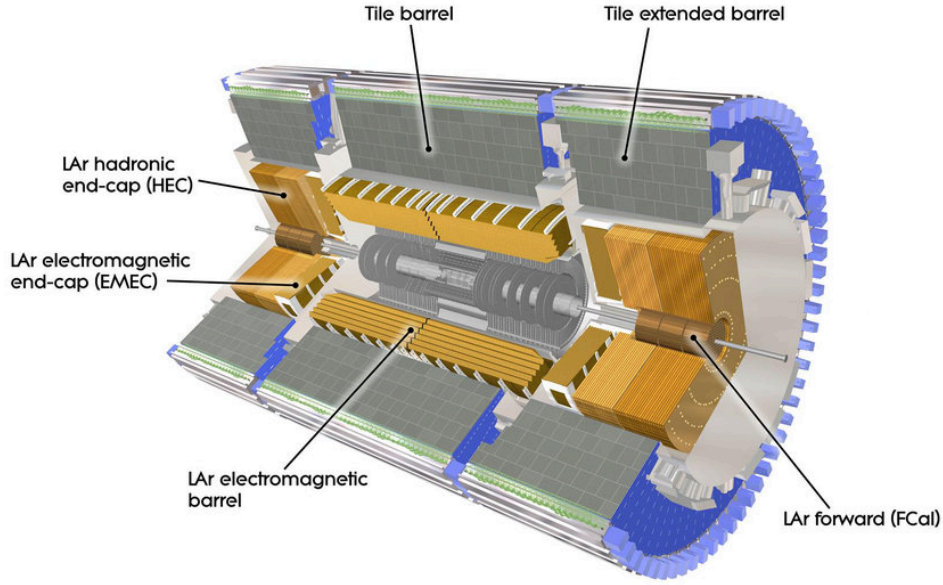
Figure 4.4: Cut-away view of the ATLAS calorimeter system [42].

The electromagnetic calorimeter measures the energy of electrons and photons as they interact with matter. Liquid argon, a material intrinsically resistant to radiations, is used as active material of the electromagnetic calorimeter and it is combined with lead absorbers. The electromagnetic calorimeter is designed in an accordion shape which allows a large acceptance and uniform response over the full azimuthal range without gaps. The electromagnetic calorimeter is divided into a barrel region (EMB) and two end-caps (EMEC) divided into two coaxial wheels.

The electromagnetic calorimeter is surrounded by the tile hadronic calorimeter which is composed of a central barrel and two extended barrels. The tile hadronic calorimeter uses scintillating tiles as the active material and steel as the absorber. It covers up to $|\eta| < 1.7$ and is supplemented by the hadronic end-cap calorimeter (HEC) to cover up to $|\eta| < 3.2$. The HEC is a copper-liquid argon calorimeter which consists of two cylindrical wheels with different granularities. The hadronic calorimeter reaches a thickness of ten interaction lengths such that it minimizes punch-through and is able to stop particles up to energies of several TeV.

Finally, the forward calorimeter combines electromagnetic and hadronic calorimeter, it covers larger $|\eta|$ up to 4.9 and it is composed of liquid argon as active material and copper-tungsten as absorber.

Muon spectrometer

The outermost subsystem of the ATLAS detector is the muon spectrometer (MS) surrounding the calorimeters. The muon spectrometer measures the energy and position of the muons that pass through the inner detector, calorimeters and the muon spectrometer. According to the

Bethe-Bloch formula, muons produced at the LHC are minimum ionizing particles due to their p_T range and as the muons pass through the inner detector and calorimeters, they deposit very little energy at these layers. Muon spectrometer is essential for identifying the muons and their trajectories and it is also used to trigger events with muons. The muons tracks are bent by a superconducting toroid magnet. The toroid magnet configuration produces a field that is mostly orthogonal to the muon tracks while minimizing the degradation of resolution due to multiple scattering. Muon reconstruction in the muon spectrometer starts with a search for hit patterns inside each muon chamber to form segments [49]. Muon tracks candidates in the muon spectrometer are then built by performing a global fit of all the segments in different layers. In order to reconstruct the *combined muons* used in physics analyses, and measure its properties, the tracking information from the inner detector and the muon spectrometer are combined.

4.2.3 Trigger system

Given the high energy, high luminosity and number of *pp* interactions per bunch-crossing, one of the most challenging aspects of a general-purpose high-luminosity detector is the trigger system. The LHC delivers a large number of collisions which need to be read out and stored in a short time. The ATLAS detector cannot be read out sufficiently fast to record all the events. Moreover, the delivered data volume is very large where only a small fraction of these delivered events contain physics processes that are of interest for the analyses. Therefore, a selection needs to be applied to decide which events should be read out and stored. The trigger system is responsible for making this decision. The trigger decision is made based on physics objects such as leptons, photons, jets or event properties such as large energy imbalance (missing energy). In the ATLAS detector, efficient triggering with low p_T thresholds on physics objects and high flexibility on tagging jets are the criteria that need to be fulfilled in order to provide high data-taking efficiencies for most physics processes of interest at the LHC [42].

In Run 1, the trigger system of ATLAS operated efficiently at instantaneous luminosities of up to $\sim 8 \cdot 10^{33}\,\mathrm{cm^{-2}\,s^{-1}}$ and up to $\sim 20 \cdot 10^{33}\,\mathrm{cm^{-2}\,s^{-1}}$ in Run 2. In Run 2 the increased center of mass energy, higher luminosity and pileup resulted in an increased rate which would be impossible to handle with the Run 1 trigger system. Therefore, between Run 1 and Run 2 significant effort was put in to optimizing and upgrading the trigger. As only Run 2 data is used in this thesis, Run 1 trigger system is not covered further in this section.

In Run 2, in order to operate at the designed luminosity, LHC had up to $40\,\mathrm{MHz}$ bunch crossing frequency. The ATLAS trigger system is designed to reduce this input rate to an output recording rate of about $1.5\,\mathrm{kHz}$ for recording and offline processing. Run 2 trigger consists of two components as it is shown in Figure 4.5: a hardware-based Level-1 (L1) trigger and a software-based high-level trigger (HLT). In order to save CPU, each trigger chain is defined as a series of algorithms with the ability to abort execution part way through.

The L1 trigger uses coarse granularity calorimeter and muon detector information to determine region-of-interests (RoI) in the detector. The trigger decision is formed by the Central Trigger
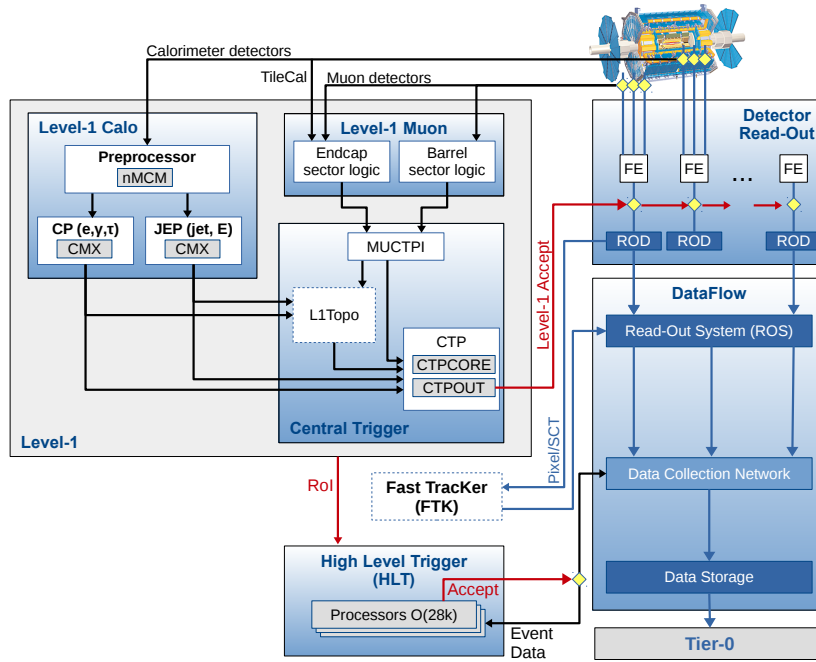
Figure 4.5: Run 2 trigger [50].

Processor (CTP) using these inputs from the L1 calorimeter (L1Calo), muon (L1Muon) triggers as well as several other subsystems. It reduces the rate from approximately 40 MHz to 100 kHz. After the L1 trigger decision, the accepted events and RoI information are transferred to the Read-Out System (ROS) and processed by the HLT.

The HLT trigger uses advanced selection algorithms which use the full-granularity detector information from all subdetectors including inner detector hits, full information from the calorimeter and data from the precision muon detectors. Short execution time is a crucial criteria for the HLT, and RoI-based reconstruction algorithms are employed by time-critical algorithms. The HLT trigger reduces the rate from 100 kHz to a maximum of 1.5 kHz.

The Run 2 ATLAS trigger system has been successfully commissioned in the start-up phase of Run 2 with cosmic ray data and early 13 TeV data. It was shown that it handled the increased rates while preserving the signal efficiencies [50].

Trigger menu

The trigger menu defines the list of available L1 items and HLT trigger chains. One component of the trigger menu is the primary triggers which are used for physics analyses and therefore is

of interest in this thesis. The availability of the triggers depend on the data taking period. To ensure an optimal trigger menu within the rate constraints for a given LHC luminosity, prescale factors can be applied to triggers and they can be changed during data taking such that triggers can be disabled or only a certain fraction of events may be accepted by them. The final trigger decision is the logical OR of many trigger chains meaning if at least one trigger from the trigger menu is fired, the event is saved.

The combined performance studies and analyses presented in this thesis use various different triggers: single electron, muon, photon, jet triggers, multi-muon triggers, a jet-based trigger. For example the jet-based trigger is used in the VLQ analysis. This trigger is called `HT1000_L1J100`, it requires one L1 jet with a p_T threshold of 100 GeV and a minimum total energy of 1000 GeV contained in jets in the event (H_T).

4.3 Simulation

Understanding the final states of pp collisions at the LHC is a challenging problem. Typically hundreds of particles are produced, and in most processes of interest their momenta have a very wide range. When the pp collisions occur, the partons (quarks and gluons) interact and might produce new particles. Depending on the properties of the produced particles, either the produced particles or their decay products and/or shower interact with the detector and leave signatures in the detector.

The Monte Carlo (MC) [51] simulation is an essential part of the high-energy physics experiment software as it provides information on how the actual collected data would look like given different theoretical scenarios and detector properties. The MC samples can be used for the prediction of discriminating observables, estimation of SM backgrounds, signal yield predictions of new physics processes, deriving calibration and estimating systematic uncertainties. Once the necessary calibrations are applied, the MC samples provide an overall good description of the collected data. Thus, MC samples provide labeled data which are for example extremely important for object reconstruction and identification in ATLAS. Additionally, the large labeled and complex datasets provide unique opportunities for the application of some machine learning techniques, one of which is employed in this thesis. The MC simulation in ATLAS is divided into three steps: event generation, physics and detector response, digitization of physics quantities [52].

The event generation is provided by different generators interfaced to the ATLAS common framework called Athena [53]. Some of the widely used event generators specialized on the simulation of the hard scattering process are ALPGEN [54] and MADGRAPH [55]. They are interfaced with the widely used full event simulation generators such as PYTHIA [56], HERWIG [57], SHERPA [58], which are used in this thesis as presented in sections 7.1, 8.3. The event generators cover a wide range of physics processes and give a list of the final-state particles that are present after the collision, to which we refer as the *truth particles*. The main steps of the event generators are the matrix element simulation of the hard process, parton showering and

hadronization[2]. The information from the event generator is then passed to be processed at the next simulation stages. At the second stage of the simulation the generated stable particles are passed through the simulation detector model in GEANT4 [59] which simulates the interaction of the particles with the detector material using the detector description and physics models. The digitization software then transforms the simulated energy deposits into electronic signals. Additionally, digitization is the step where some effects like collisions in addition to the one of interest are introduced to the simulation. Finally, common reconstruction algorithms are used to reconstruct for example calorimeter clusters in MC generated samples and actual collected data. The reconstruction algorithms will be covered in the next chapter.

---

[2]Due to hadronization, quarks and gluons form *jets* and parton shower radiation cascades describe how *jets* are formed. Hence, parton showering and hadronization are particularly important in this thesis and at the LHC.

Chapter 5

Object Definitions in ATLAS

The particles produced in the collisions interact with the ATLAS detector, deposit energy and leave signatures in different parts of the detector. These signatures are transformed into electronic signals. In order to be able to interpret the data, the electronic signals are combined and processed to build physics objects. This process of defining and building physics objects, such as tracks and jets, is called reconstruction. Once the physics objects are reconstructed and calibrated, they form the set of candidates corresponding to different signatures and particles. Certain criteria, referred to as working points, are applied to identify the physics objects of interest.

Reconstruction and identification of physics objects are fundamental components of the data analysis in ATLAS. As their performance impact the signal and background discrimination, they have a large impact on many analyses. Hence, the *combined performance* efforts which provide guidelines for reconstruction and identification working points play an important role. An overview of the physics objects used throughout this thesis are presented in this chapter; the reconstruction and performance of jets will be covered in greater detail as I contributed in this area.

5.1 Tracks and Vertices

Due to the high luminosity and energies, signatures from multiple pp interactions overlap at the ATLAS detector. It is important to reconstruct the point at which a pp interaction occurred. This point of interaction is called a vertex. The input to the vertex reconstruction is a set of reconstructed tracks. Tracks are particle trajectories which were introduced in section 4.2.2.

5.1.1 Tracks

Track reconstruction and selection are fundamental as tracks are used as input to the vertices and they are the inputs to reconstruction of some higher-level physics objects such as muons.

The reconstruction of tracks in the ID is based on fitting a trajectory model to a set of position and direction measurements of a charged particle passing through the detector, referred to as hits. The particle tracks reconstruction requires a minimum number of silicon hits in the ID and then the track is extended by adding hits. Depending on the use case, different criteria

are applied to select or reject tracks. Requiring minimum number of hits and putting an upper threshold on the holes, which are intersections of the track trajectory with a sensitive detector element that do not result in a hit when it is expected, help with the computation, reduces fake tracks from pileup or instrumental effects and improves track quality. There are also p_{T}, η requirements applied. The track efficiency is low at low p_{T} and it improves at higher p_{T}, reaching efficiencies above 90%[1] for tracks which have p_{T} above 2 GeV as presented in Ref. [60].

5.1.2 Vertices

The input to the vertex reconstruction is a collection of selected reconstructed tracks where the possible vertex candidates are identified by extrapolating the found tracks to the beam line. The procedure of vertex reconstruction is divided in two stages: vertex finding and vertex fitting. First, a set of reconstructed tracks passing the vertex selection criteria is defined. Then a seed position for the first vertex is selected. The tracks and the seed position are used to fit the best vertex position in an iterative process. After the vertex position is determined, tracks that are not compatible with the vertex are removed from the vertex. At least two selected tracks are required to define a reconstructed vertex. These steps are repeated and multiple vertices are reconstructed until all the tracks are associated to vertices.

Since multiple *pp* interactions in a short period of time lead to multiple reconstructed vertices, one challenge is to reconstruct and identify the interesting vertex where a hard scatter happened. This point of interaction that we would like to analyze is called a *primary vertex*. Other inter-action points in addition to the one of interest are referred to as pileup vertices. The correct assignment of tracks to their vertex is fundamental to reconstruct the kinematic properties of the hard-scatter and of the soft interactions as a measure of pileup. The primary vertex is then selected comparing the properties of all the reconstructed vertices. As the physics processes of interest generally have a large number of high-p_{T} tracks, the vertex with the largest sum of the squared track p_{T} is generally selected as the primary vertex. It is good to keep in mind that although this is the general case, there are some physics processes where this is not expected, and the choice of the primary vertex can be different. For example, for some rare processes, a large number of soft (low-p_{T}) tracks are expected which can favor either the vertex with the largest sum of the squared track p_{T} or the vertex with largest number of tracks.

5.2 Electrons and Photons

Electrons deposit almost all of their energy in the electromagnetic calorimeter and no significant energy deposit is expected in the hadronic calorimeter. Electrons, as charged particles, leave tracks in the ID. Electron candidates are reconstructed by using the following components: localized clusters of energy deposits found within the electromagnetic calorimeter, charged-particle

---

[1]The exact efficiency depends on the working point and the environment. The quoted value corresponds to the tracks which satisfy the *Loose* criteria.

tracks identified in the inner detector, and close matching in $\eta - \phi$ space of the tracks to the clusters [61].

Inputs from electromagnetic calorimeter and ID tracks are used to form a likelihood function to identify the reconstructed candidate objects as electrons. To ensure that the input tracks originate from the primary vertex, selection criteria on the longitudinal and transverse impact parameter are applied. Different identification criteria are defined for several electron identification working points. Information on the associated tracks, the shower shape variables and the fraction of energy deposited in the hadronic calorimeter are some of the important inputs to the utilized likelihood-based electron identification. The likelihood is optimized to discriminate prompt electrons from the combined background of jets that mimic the signature of prompt electrons, electrons from photon conversions in the detector material, and non-prompt electrons from the decay of hadrons containing heavy flavors.

Photons also deposit all their energy in the electromagnetic calorimeter and no energy deposit is expected in the hadronic calorimeter. Therefore, for photons, a similar strategy as electrons is followed for photon reconstruction and identification. It is important to note that due to Bremsstrahlung of electrons and conversions of photons, their signatures can be hard to distinguish. Since the electrons are charged whereas the photons are neutral, signatures of the electrons and converted photons are different in the tracker. Hence, the tracking information is used to distinguish them.

5.3 Muons

Muons are minimum ionising particles (MIP) in the p_T of interest and passing through the calorimeters, they deposit very little energy. They reach the MS if they have sufficient energy. Therefore, detector information from the ID and MS tracks are combined to reconstruct muon candidates where the input tracks are required to originate from the primary vertex. Additionally, information from the calorimeter is used in regions where measurements from the muon spectrometer are not available due to a MS gap around $|\eta| = 0$.

Different identification working points are defined for different use cases where different efficiencies and fake rates are obtained as a function of muon p_T. The identification criteria consider the quality of ID and MS tracks based on hit requirements and the agreement of the two. Muons have low reconstruction and identification efficiencies at low p_T. One important improvement in Run 2 over Run 1 is the introduction of the identification working point that is optimized to improve the poor efficiency at low p_T (below 5 GeV), referred to as the *LowPt* working point.

5.4 Isolation

For electrons, muons and photons, a common challenge is to differentiate the prompt signal objects from fakes originating from background processes such as heavy quark decays. The

prompt signals often[2] have little activity around them. Whereas, fake electrons, muons and photons originating from background processes have large activity around them. Calorimeter-based and track-based isolation variables are defined as a measure of activity around these objects to suppress backgrounds with high activity. Finally, it is important to keep in mind that the reconstruction and identification of some physics objects can worsen in busy environments.

5.5 Jets

Quarks and gluons are produced in abundance at the LHC. Unlike leptons, produced quarks and gluons are not directly observed in the detector due to color confinement and hadronization [62]. The final state quarks and gluons hadronize, forming baryons and mesons which afterwards can decay in many stages. These hadrons and their decay products lead to *jets* of particles. Therefore, each parton with a sufficiently long lifetime[3] at the ATLAS detector is observed as a jet and the jet's measured properties can be linked to the underlying parton. Jets are one of the most commonly produced physics objects in pp interactions at ATLAS and play a fundamental role in data analysis.

Jets deposit a large amount of energy in the calorimeters and they are defined by jet recombination algorithms such as anti-k_{t} [63] and Cambridge–Aachen (C/A) [64, 65]. It is important to keep in mind that, as mentioned in chapter 4.1, multiple pp collisions happen in a short period of time and only one parton from each incident proton participates in one hard scattering process. As a result, only a fraction of the hadrons in the final state are associated with that hard scattering process of interest, with the remainder ascribed to the underlying event (UE) [66]. This presents a challenge for jet reconstruction in ATLAS.

5.5.1 Inputs

Different type of objects can be used as inputs to the jet recombination algorithms. These inputs are referred to as the constituents of the jets. Constituents of the jets can be based on the calorimeter energy deposits, ID tracks or a combination of the two. The jets used in this thesis are reconstructed using calorimeter energy deposits and sometimes complementary information on these jets' properties is provided by using tracks.

Topoclusters

In ATLAS, inputs to jet reconstruction algorithm are typically topologically connected calorimeter clusters (topoclusters). Topoclusters are formed at the electromagnetic (EM) scale from topologically connected cell signals with the aim of reconstructing three-dimensional energy deposits in the calorimeter [48, 67]. Individual topoclusters are not always expected to contain

---

[2]There are also cases where the prompt electrons, muons and photons are produced in busy environments.
[3]Sufficiently long lifetime such that they don't decay before interacting with the detector. For example the top quark decays before it hadronizes due to its short lifetime.

the full shower of a single particle. Depending on the incoming particle types, energies, spatial separations and cell signal formation, individual topoclusters can represent the full shower or shower fragment, the merged shower of several particles, or a combination of these.

Topoclusters are built using a nearest-neighbor algorithm where the collection of noise-suppressed calorimeter cell signals into topoclusters follows spatial signal-significance patterns generated by particle showers. The noise suppression is obtained by grouping cells with significant energy deposits using an observable controlling the cluster formation called the cell signal significance. The cell signal significance ($\varsigma_{\text{cell}}^{\text{EM}}$) is defined according to Eqn. 5.1 as the ratio of the cell signal (energy deposit $E_{\text{cell}}^{\text{EM}}$) to the expected noise in this cell (σ_{noise}). The noise in calorimeter cells is estimated as the quadratic sum of the measured electronics ($\sigma_{\text{noise}}^{\text{electronic}}$) and pileup ($\sigma_{\text{noise}}^{\text{pileup}}$) noise for each run year.

$$\varsigma_{\text{cell}}^{\text{EM}} = \frac{E_{\text{cell}}^{\text{EM}}}{\sigma_{\text{noise}}} = \frac{E_{\text{cell}}^{\text{EM}}}{\sqrt{(\sigma_{\text{noise}}^{\text{electronic}})^2 + (\sigma_{\text{noise}}^{\text{pileup}})^2}} \tag{5.1}$$

The clustering algorithm starts from the seed cells with the highest significance above $4\sigma_{\text{noise}}$. Then, neighboring cells are iteratively added to the topocluster if their energy passes the $2\sigma_{\text{noise}}$ threshold until there are no longer any adjacent cells with an energy of greater than $2\sigma_{\text{noise}}$. This is followed by the addition of all adjacent cells. This *seed and collect* procedure is repeated for the next topocluster where the procedure starts with the next highest-significance seed cell; the algorithm continues to *seed and collect* until no cells remain. Finally, the cluster splitting algorithm separates the reconstructed topoclusters based on local energy maxima to avoid overlap. The resulting topoclusters are calibrated at the EM scale and are considered to be massless. It is possible to reconstruct topoclusters with negative energies. However, only those with positive energy are used as inputs to the jet reconstruction. Additionally, topoclusters formed at the EM scale are individually calibrated to correct for effects such as the non-compensation of the calorimeter response and signal losses due to clustering and inactive material. This correction is particularly important for hadronic signals because for example the non-compensation of the calorimeter response causes a smaller signal in the calorimeter for hadrons compared to electrons or photons with the same energy. Due to their differences, this calibration is applied on the topoclusters considering whether the cluster is initiated by an electromagnetic or an hadronic signal. The employed calibration attempts to calibrate highly localized and relatively small topoclusters. As the local hadronic calibration includes cell signal weighting, the calibration is referred to as local hadronic cell weighting (LCW or LC) calibration. The LC calibration takes the topoclusters from the electromagnetic scale to LC-scale. Hence, topoclusters reconstructed on the basic EM scale are referred to as EM topoclusters or EMTopo while topoclusters which are further LCW calibrated are referred to as LC topoclusters or LCTopo.

Inputs to truth-jets

In simulation, in addition to jets reconstructed from detector-level observables such as topoclusters, a set of jets based on generator-level information is used to characterize the performance of a given tagging algorithm. These jets are reconstructed with the same jet reconstruction algorithms using stable particles from the hard scatter with lifetimes greater than 10 ps, excluding muons and neutrinos, as inputs. They are referred to as *truth jets* and the related observables are generally denoted by the superscript "true" or "truth".

5.5.2 Reconstruction algorithms

Jet reconstruction algorithms define a procedure employing a set of rules to group together particles in a jet. They are one of the fundamental tools for data analysis in hadronic collisions. Due to the complex nature of the hadronic collisions, the reconstruction algorithms face many theoretical and experimental challenges at the LHC. Examples of such challenges are initial and final state radiations (ISR, FSR), pileup and other contributions from underlying event. An ideal jet algorithm should be easy to implement theoretically and experimentally, and should be infrared and collinear safe [68]. At the LHC, it is also strongly preferred that the reconstruction and properties of the jet are robust against pileup and underlying event.

Infrared safety of the algorithms is achieved if in the presence of soft (low p_{T}) radiation such as a low-p_{T} gluon emission, the jet reconstruction algorithm and the resulting set of jets are not significantly affected. Collinear safety of the algorithms means that the jet reconstruction algorithm and the resulting set of jets are not significantly affected by an additional small-angle splitting. These two criteria together guarantee that the cross section is finite, despite the present divergences as the divergent tree-level matrix elements associated to soft gluon emissions and collinear splittings cancel the divergent loop matrix elements.

There are two main families of algorithms: cone style and successive-combination style, referred as sequential algorithms. The cone algorithms define a jet by grouping a set of particles whose momentum vectors lie within a certain angular cone. The sequential algorithms use some distance measure to recursively groups sets of particles into larger sets of particles. There are algorithms from both families which satisfy the infrared and collinear safety. In this thesis we will focus on the currently widely used sequential algorithms.

The sequential algorithms can be generalized by a set of rules and parameters. First, the distance measures are defined using distances between the entities (particles or pseudojets, where a pseudojet is a collection of one or more constituents defined in a previous step), distances between entities and beam and the momenta of the particles. The distance between entities i and j, d_{ij}, is calculated as shown in Eq. 5.2 where R is the radius parameter, $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ and k_{ti}, y_i, ϕ_i are respectively the transverse momentum, rapidity and azimuth of particle i. The distance between entity i to beam, d_{iB}, is calculated as shown in Eq. 5.3. In these equations p is a parameter used for the sake of generalization of several algorithms and it defines how the input particles are combined.

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R^2} \qquad (5.2)$$

$$d_{iB} = k_{ti}^{2p} \qquad (5.3)$$

The recombination algorithms first find the smallest d_{ij} and compare it with d_{iB}. If d_{ij} is the smallest, they recombine i and j to define a pseudojet. Whereas if d_{iB} is the smallest, the algorithms define i as a jet and removes the entity from the list. The distances are recalculated and the procedure repeated until no entities are left. There are three widely used algorithms corresponding to three different p values: 1, 0, -1. Depending on the value of the p parameter, the algorithms have different features. The case $p = 1$ corresponds to the k_t algorithm, $p = 0$ corresponds to the C/A algorithm and $p = -1$ corresponds to the anti-k_t algorithm. For the C/A algorithm, entities are combined based on their angular separation. For the $p > 0$ cases and the k_t algorithm, soft particles are favored to be merged first, whereas for the anti-k_t algorithm the jet starts growing outwards from a higher-p_T object.

Infrared and collinear safe sequential algorithms generally have irregularities in the boundaries which presents theoretical and experimental challenges. The shape-regular algorithms which result in conical jets are generally simpler for theoretical calculations, easier to calibrate and more UE and pileup robust. However, they also have the disadvantage that they cannot adapt to the branching nature of QCD. Therefore, the sequential anti-k_t algorithm with $p = -1$ was proposed and is currently used by ATLAS. The key feature of this algorithm is that, while the shape of the jet is not modified by the soft particles making it robust against these soft particles, it can be modified by hard particles. As a result, anti-k_t algorithm results in more conical jets compared to the k_t and C/A algorithms and is also infrared and collinear safe. The behaviors of different jet algorithms, reconstructed with a radius parameter of $R = 1.0$, are shown in Fig. 5.1.

The radius parameter of the reconstruction algorithm is defined based on the application. Standard jets in ATLAS use $R = 0.4$ and EM topoclusters to capture the signature of a single particle, these jets will be referred to as the small-radius (small-R) jets in this thesis. In some cases it is desired that a single large jet captures decay products of a heavy object such as the top quark, so it is helpful to use a large radius such as $R = 1.0$. The inputs to large-radius jets can be LC topoclusters or small-radius jets (for reclustered jets [69, 70]). Since the jets with large radius are very susceptible to pileup, dedicated pileup suppression techniques are applied to jets in order to remove the pileup contributions from the jet[4]. Such large-radius jets will be covered and studied in greater detail in this thesis.

---

[4]Large-radius jet grooming is presented in section 6.1.
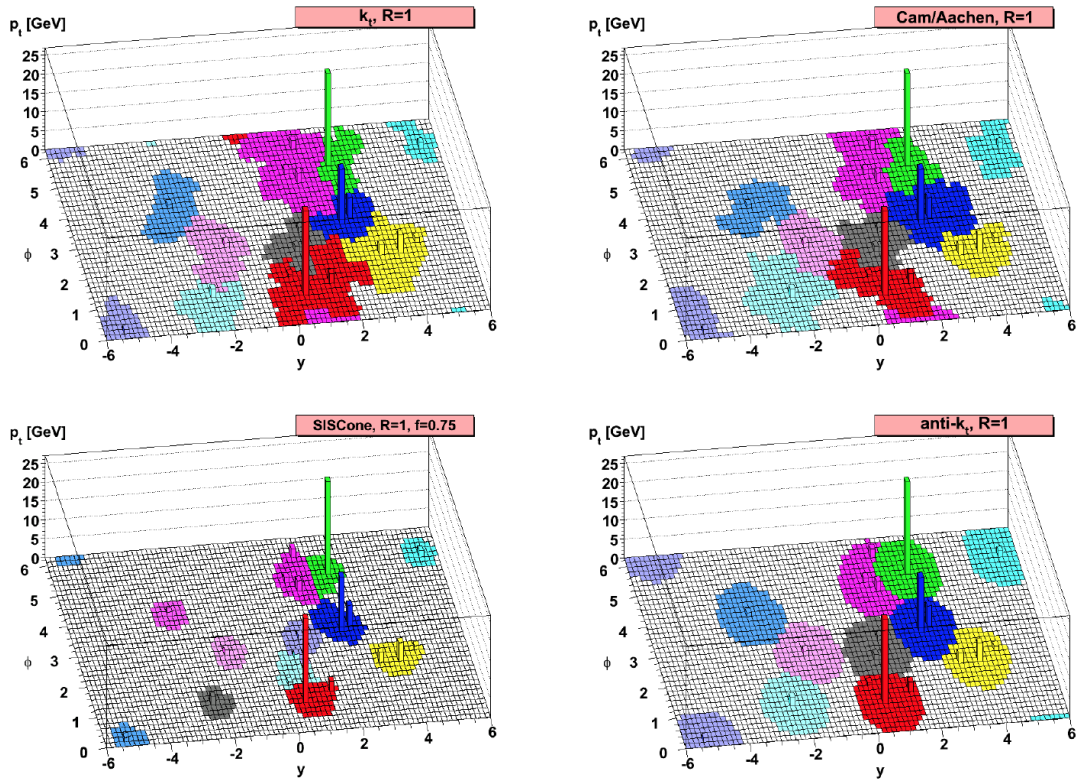
Figure 5.1:   A sample parton-level event, together with many random soft particles, is clustered with four different jets algorithms. The active catchment areas of the resulting jets are presented [63].

5.5.3 Calibration

After the jets are reconstructed, jet calibration is applied to account and correct for various effects such as the pileup, non-compensation of the calorimeter and differences between data and simulation. Series of corrections are applied sequentially, which is referred to as the calibration chain. The applied calibration chain depends on the jet inputs and radius. The calibrations are accompanied by the related systematic uncertainties. Two jet collections which are widely used in ATLAS and have dedicated calibrations are small-radius jets with $R = 0.4$, reconstructed using EM topoclusters and large-radius jets with $R = 1.0$, reconstructed using LC topoclusters.

Small-radius jets reconstructed using EM topoclusters are calibrated to correct the jet energy scale (JES) such that the energy of the reconstructed jet corresponds to the energy of the initial parton at particle level. Hence, this calibration chain is referred to as the JES calibration. Both simulation and data-based methods are used for the small-radius JES calibration as described in Ref. [71]. The data-based calibration methods are referred to as *in-situ*. The calibration chain consists of: origin correction, pileup correction, absolute JES and pseudo-rapidity calibration, global sequential calibration and *in-situ* calibration.

The small-radius JES calibration starts with the simulation-based corrections. First, the origin

correction updates the four-momentum of the jets such that the jets point to the primary vertex rather than the center of the detector, while the jet energy is kept constant. Next, the pileup correction consisting of two components, jet area-based and residual correction, is applied. The pileup correction is followed by a simulation-based JES and pseudo-rapidity calibration where the jet is calibrated to the particle-level jet using truth jets in simulation events. To correct the energy of the reconstructed jets to the truth jet energy scale on average, the jet energy response[5] distribution is fitted to a Gaussian function. The mean of the resulting Gaussian fit and numerical inversion method are used to scale the reconstructed jet four-momentum. Then, in order to correct an observed bias relative to the truth jets in certain regions of the detector, the jet pseudo-rapidity is calibrated using truth jets and the jet transverse momentum is updated accordingly. For the jets which are calibrated up to this point, the jet energy scale is observed to have residual dependencies on several features of the jet. Hence, a global sequential calibration, which uses global jet observables, is applied to reduce such dependencies and improve the resolution of the jet energy scale. Finally, in order to account for differences in jet response between data and simulation, *in-situ* methods are used. *In-situ* energy calibration corrects jets in data using other well-measured reference objects where the transverse momentum of a jet is balanced against other well-measured reference objects in an event. It consists of η-intercalibration, Z + jet balance, γ + jet balance and multijet balance calibrations.

Large-radius jets reconstructed using LC topoclusters are calibrated to correct the JES and the jet mass scale (JMS) to the truth jet scale. In this thesis only simulation-based methods are used to calibrate the large-radius jets. Since pileup suppression techniques are applied to large-radius jets prior to the calibration, large-radius jet calibration chain does not include a dedicated pileup calibration. The calibration chain consists of: absolute JES and pseudo-rapidity calibration and JMS calibration.

The large-radius jet calibration starts with the JES and pseudo-rapidity calibration. The large-radius JES and pseudo-rapidity calibration is performed in a similar manner as the small-radius jet one where the reconstructed jet is calibrated to match the truth jet energy scale on average and the pseudo-rapidity of the jet is updated to correct an observed bias in certain regions of the detector. Next, JMS calibration is applied to correct on average the reconstructed jet mass to the truth jet mass as described in Ref. [72]. The JMS calibration follows a similar approach as the JES calibration where the jet mass response[6] is fitted to a Gaussian distribution and the mean of the Gaussian fit is taken as the average mass response. The average mass response is then used to scale the jet four-momentum. Since the jet mass is an important observable for many physics analyses that use large-radius jets, it is essential to improve the resolution of the jet mass. Hence, the JMS calibration plays an important role for large-radius jets.

As mentioned previously, large-radius jets can also be reconstructed from the fore-mentioned small-radius jets. In this case, since fully-calibrated small-radius jets are the inputs to the large-

---

[5]Jet energy response is defined here as the ratio between the reconstructed jet energy and the associated truth jet energy: $E^{\mathrm{reco}}/E^{\mathrm{truth}}$.

[6]Defined here as the ratio between the reconstructed jet mass and the associated truth jet mass: $m^{\mathrm{reco}}/m^{\mathrm{truth}}$.

radius jet, large-radius jet inherits the small-radius jet calibration and the reclustered jet does not require any additional calibration.

5.5.4 Flavor tagging

The identification of jets containing b-hadrons (*b*-tagging) and c-hadrons (*c*-tagging) to discriminate them from other lighter quarks and gluons is referred to as flavor tagging. Flavor tagging plays an important role in the physics analyses in ATLAS. In this thesis, *b*-tagging is used in several places.

B-hadrons have a large mass, a non-negligible decay length and several lepton decay modes (with non-negligible branching ratios) which result in *b*-jets and their associated tracks having a distinct signature. Therefore, the properties of the related tracks are used to identify *b*-jets. In ATLAS a multivariate approach is taken to build a discriminant using the impact parameter, secondary vertex information and decay chain as inputs. The *b*-tagging discriminants used in this thesis are built by using a Boosted Decision Tree (BDT).

5.6 Missing Transverse Momentum

The missing transverse momentum [73, 74] is an observable used to estimate the transverse momentum carried by undetected particles produced in collisions, such as neutrinos. The momentum conservation in the plane transverse to the beam axis requires that the transverse momenta of all particles in the final state should sum effectively to zero. However, due to undetected particles, an imbalance in the measured momenta might be observed. This momentum imbalance might be caused by SM particles such as neutrinos or by BSM particles which do not leave any signature in the detector due to the type and strength of their interaction with the detector material. This momentum imbalance is referred to as missing transverse momentum ($\vec{E_\mathrm{T}}^{\,\mathrm{miss}}$) and its magnitude is referred to as $E_\mathrm{T}^{\mathrm{miss}}$.

Missing transverse momentum that is commonly used in ATLAS analyses is calculated using the negative vector sum of the p_T of all reconstructed hard objects: jets, electrons, photons, muons and τ-leptons. Additionally, a soft energy term is included to account for non-reconstructed particles originating from the hard scatter. The soft energy term is calculated using only tracks matched to the primary vertex to reduce contamination from pileup.

Chapter 6

Large-Radius Jets, Jet Substructure and Large-Radius Jet Identification Techniques

The center-of-mass energy of the collisions has reached 13 TeV at the LHC. Such high collision energies produce large samples of heavy particles with a p_T that exceeds their rest mass, resulting in very high Lorentz boosts. Since the decay products of such boosted heavy particles are very collimated, reconstruction algorithms which rely on a one-to-one jet-to-parton assignment are often not as successful. As it was introduced in section 5.5, to overcome such challenges jets are reconstructed with a larger radius parameter in order to capture the energy of the complete hadronic decay in a single jet. These jets are referred to as large-R jets. The internal structure and the energy flow of the large-R jets is then analyzed to identify heavy boosted objects among the other abundant jet production at the LHC. In this chapter the focus will be on algorithms which are used in ATLAS to reconstruct and identify jets originating from hadronic decays of W bosons and top quarks.

6.1 Grooming

Dedicated jet grooming techniques have been developed to remove the pileup, soft radiation contamination and other underlying event contributions from large-R jets in order to improve the reconstructed jet mass and the analysis of the internal structure of the jet. Grooming aims to improve the jet resolution and object discrimination. Three grooming techniques are commonly used: filtering [75], trimming [76], pruning [77] and soft drop [78]. Although all these techniques are powerful, only trimming is presented in this chapter as trimmed large-R jets are used in this thesis.

   The trimming procedure uses the constituents of the large-R jet formed with radius R and reclusters its constituents into smaller jets (subjets) with radius R_{sub} by using a jet clustering algorithm. Typically the k_{t} reclustering algorithm is used. The ratio of the p_T^i of subjets to the p_T of the large jet (p_T^{jet}) $f = p_T^i/p_T^{\mathrm{jet}}$, is used as a selection criterion. If the ratio f is smaller than a chosen value f_{cut}, the subjet is rejected. The subjets which pass the selection criterion are then recombined into a groomed jet. Schematics of the trimming process is shown in Fig. 6.1. Trimming significantly improves the resolution of the jet energy measurement, mass resolution
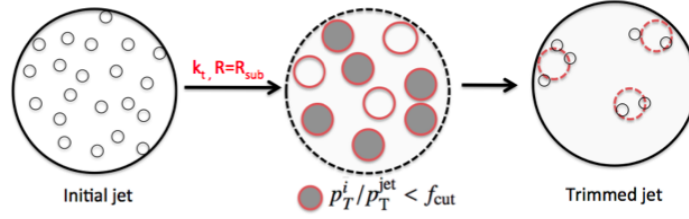
Figure 6.1: Schematics of trimming [80].

and object identification. In ATLAS, pileup mitigation techniques including grooming are studied extensively for Run 2 in Ref. [79]. In Fig. 6.2, jet mass distributions of a sample of W bosons and background multijet jets are presented in two p_T^{true} bins before and after trimming with a variety of trimming parameters. The jets compared are reconstructed using anti-k_t algorithm with $R = 1.0$ and are uncalibrated (prior to any calibration).

6.2 Jet Collections

In this thesis three different large-R jet collections are used. The focus of this and the next chapters rests on two of these jet collections where the inputs to the jet reconstruction algorithms are LC topoclusters.

Among two of these large-R-jet collections, particular focus is on the trimmed anti-k_t jets reconstructed with radius $R = 1.0$ and trimmed with parameters $R_{\text{sub}} = 0.2$, $f_{\text{cut}} = 5\%$. These reconstructed large-R jets are then calibrated in a two-step procedure that first corrects the jet energy scale and then the jet mass scale as described in section 5.5.3. Trimmed anti-k_t large-R jet collections with different R_{sub} parameters have been extensively studied and used by ATLAS for W, Z boson and top quark reconstruction and identification since Run 1 [81–84]. Moreover, the recent studies performed in Run 2 summarized in Ref. [79] demonstrate that under current high energy and pileup conditions, the conventional reconstruction algorithm of large-R trimmed jets is one of the most performant jet reconstruction algorithms in terms of pileup robustness and signal-background discrimination. The second jet collection makes use of the C/A algorithm with a radius parameter of $R = 1.5$ which aims to identify top-quark jets across a broad p_T range, in particular reaching low p_T. These jets, used in conjunction with the HEPTopTagger [85, 86] algorithm described in section 6.5.3, are also trimmed to mitigate the effects of pileup. Trimming with a parameter of $R_{\text{sub}} = 0.2$ and $f_{\text{cut}} = 5\%$, the same as those used in the trimming of the anti-k_t $R = 1.0$ jet collection, is found to provide pileup-robust jet reconstruction and identification performance.

As it was introduced in section 5.5.1, in simulation *truth jets* can be used to characterize the performance of reconstruction and identification algorithms. Truth jets that are reconstructed with the anti-k_t algorithm with a radius parameter $R = 1.0$ but not modified with the trimming algorithm are used to characterize the performance of the above two jet collections and related identification algorithms.
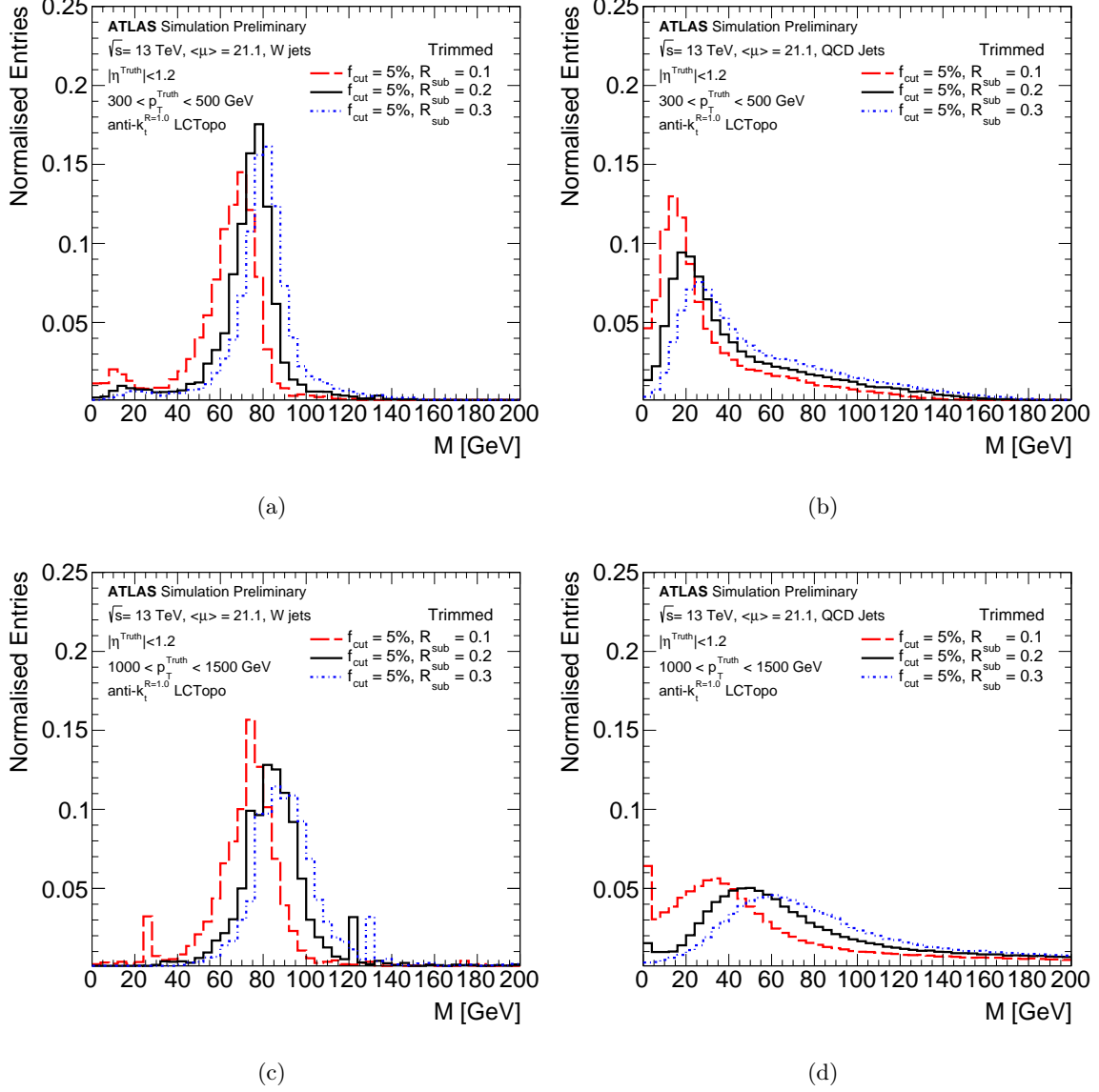
Figure 6.2: The MC-based comparison of W boson (a),(c) and multijet background (b),(d) jet mass distributions for variety of trimming options. The jets are reconstructed with anti-k_{t} algorithm, $R = 1.0$ and using LC-calibrated topoclusters as inputs. The trimming parameter choices are $R_{\mathrm{sub}} = 0.1$, 0.2, and 0.3 for a fixed f_{cut}=5% [79].

6.3 Jet Labeling

Identifying the jets originating from hadronic decays of heavy boosted objects (signal) and discriminating them from other abundant jet production (background) is one of the main purposes of large-R jets. Therefore, the labelling of the particle that initiated the jet as a particular signal or background is highly important to study the performance of the jet reconstruction and identification in MC. For signals, this labelling is achieved by truth matching that is defined under a hypothesis of a signal model for the jets. In the W-boson and top-quark identification studies presented in the next chapters, signal jets are defined as hadronically-decaying W bosons or top quarks when all partonic decay products are fully contained within the region of interest of the reconstructed jet in a three-step truth matching process. It is important to remember that the partonic decay products are two quarks for W bosons following the $W \to qq$ decay and two quarks and a b-quark for top quarks following the $t \to Wb \to qqb$ decay as discussed in section 2.1. First, reconstructed jets are matched to truth jets with a matching criterion of $\Delta R(j_\text{true}, j_\text{reco}) < 0.75$. Next, those truth jets are matched to truth W bosons and top quarks (W, t) with a matching criterion of $\Delta R(j_\text{true}, \text{particle}) < 0.75$. Finally, the partonic decay products of the parent W boson (two quarks) or top quark (two quarks and b quark) are matched to the reconstructed jet. A reconstructed jet is labelled as a W-boson or top-quark jet if the parent particle and all of its decay products are contained within a region in (η, ϕ) with $\Delta R < 0.75 \times R$, where R is the jet radius parameter. This requirement is referred to as the containment and the labeling scheme which requires it is referred to as the contained jet labeling. The value $0.75 \times R$ for the jet labelling criteria is chosen as a compromise between the resulting labelling efficiency and the resolution of the top-quark and W-boson jet mass peak. As the p_T of the W and top quark directly impact how collimated the decay products are, containment and labeling efficiency are expected to be p_T dependent. In Fig. 6.3, the p_T dependence is demonstrated by evaluating the jet labelling efficiency of truth jets with different containment categories as a function of truth W boson and top quark p_T. For jets matched to the parent W boson, at $p_\text{T} \sim 200$ GeV only $\sim 50\%$ of the jets are fully contained while for $p_\text{T} \sim 500$ GeV the containment already rises to nearly 100%. In the case of top-quark jets, the efficiencies are much lower. Only $\sim 20\%$ of the jets are fully contained at $p_\text{T} \sim 350$ GeV and only $\sim 80\%$ of jets are fully contained at $p_\text{T} \sim 1$ TeV.

Background jets are obtained from the relevant background samples and a reconstructed jet is kept and labelled as background if the reconstructed jet is matched to a truth jet with a matching criterion of $\Delta R(j_\text{true}, j_\text{reco}) < 0.75$.

It is important to keep in mind that the chosen signal definition is one of the available choices. This choice generally depends on the analysis and in some other studies more inclusive definitions are also used, where the partonic decay products are not required to be contained within the reconstructed jet.
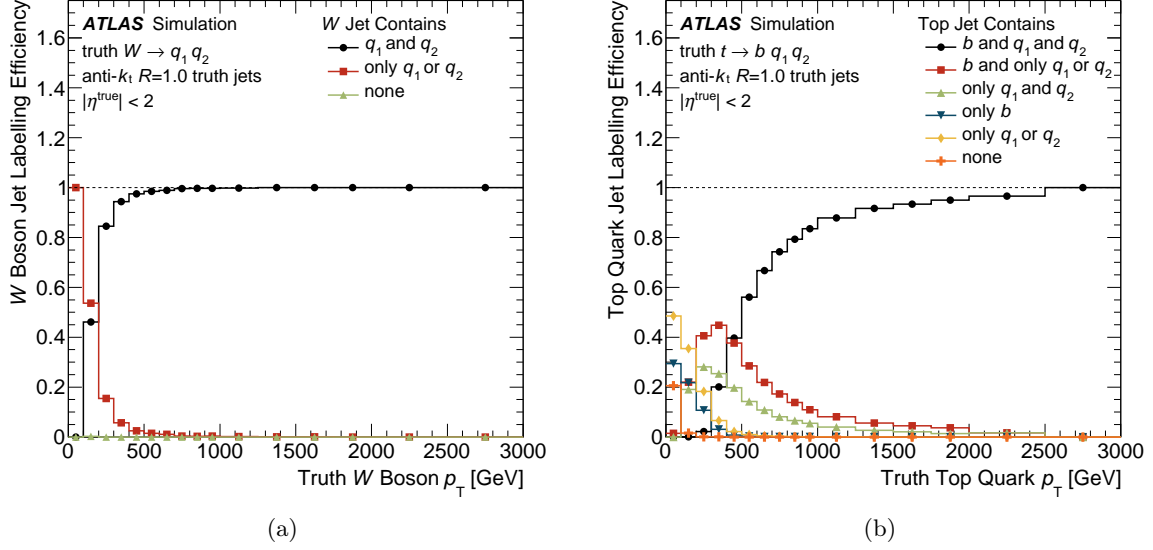
Figure 6.3: Containment of the W-boson (a) and top-quark (b) decay products in a single anti-k_t $R = 1.0$ truth jet as a function of the particle's p_T.

6.4 Substructure Moments

Once the jet is reconstructed as a collection of constituents and calibrated, a variety of methods can be used to classify a jet as originating from a heavy particle decay as opposed to a light jet originating from gluons and quarks of all flavors other than top quarks. These identification algorithms are collectively referred to as large-R jet tagging algorithms. One method of building a powerful tagging algorithm is to use one or more substructure moments.

Substructure moments are a broad class of variables which provide information about the characteristics and the substructure of the large-R jets by analyzing the energy flow of the jet. As multijet background jets and jets from boosted heavy objects have different energy patterns, substructure moments can be used for tagging large-R jets. The set of constituents of the trimmed jets are used to calculate the substructure moments. There is a large number of substructure moments. Boosted object tagging algorithms use different variables depending on the problem. The full set of jet moments studied in this thesis is summarized in Table 6.1 and a comprehensive description and study of jet moments can be found in Ref. [81, 82]. Substructure moments which are widely used for W boson and top tagging in ATLAS are the jet mass, energy correlation functions, N-subjettiness and k_t splitting scales. These moments are briefly described in this section [1].

---

[1]Further variables studied in this thesis are defined in appendix A.

Observable	Variable	Used for	References
Calibrated jet kinematics	p_{T}, m^{comb}	top,W	[72]
N-subjettiness	τ_1,τ_2,τ_{21}	top, W	[87, 88]
	τ_3,τ_{32}	top	
Energy correlation functions	e_3,C_2,D_2	top,W	[89, 90]
Fox–Wolfram moment	R_2^{FW}	W	[91, 92]
Splitting measures	z_{cut}	W	[93, 94]
	$\sqrt{d_{12}}$	top, W	
	$\sqrt{d_{23}}$	top	
Planar flow	$\mathcal{P}$	W	[95]
Angularity	a_3	W	[96]
Aplanarity	A	W	[92]
KtDR	$KtDR$	W	[97]
Qw	Q_w	top	[93]

Table 6.1: Summary of jet moments studied along with an indication of the applicable tagger topology. In the case of the energy correlation observables, the angular exponent β is set to 1.0 and for the *N*-subjettiness observables, the winner-take-all [78] configuration is used.

6.4.1 Jet mass

The jet mass provides the most powerful single large-R jet moment. In Run 1 typically a purely calorimeter-based jet mass was used. The calorimeter-based jet mass (m^{calo}) [72] for a large-R jet with topocluster constituents i with energy E_i, momentum $\vec{p}_i$ (with $|\vec{p}_i| = E_i$) is defined as:

$$m^{\mathrm{calo}} = \sqrt{\left(\sum_i E_i\right)^2 - \left(\sum_i \vec{p}_i\right)^2}. \tag{6.1}$$

At very high p_{T}, the separation of the decay particles of the large-R jet becomes comparable with the calorimeter granularity and the resolution of this observable decreases as energy depositions from individual particles begin to merge in clusters. At such high p_{T}, it becomes beneficial to use the spatial granularity of tracks reconstructed in the ID. Hence, the track-assisted jet mass (m^{TA}) is constructed combining tracking and calorimeter information, it is defined as

$$m^{\mathrm{TA}} = m^{\mathrm{track}} \times \frac{p_{\mathrm{T}}^{\mathrm{calo}}}{p_{\mathrm{T}}^{\mathrm{track}}}, \tag{6.2}$$

where m^{track} and $p_{\mathrm{T}}^{\mathrm{track}}$ are the invariant mass and p_{T} calculated from tracks associated with the large-R trimmed calorimeter jet and $p_{\mathrm{T}}^{\mathrm{calo}}$ is the p_{T} of the same large-R trimmed calorimeter jet. After the JES calibration, JMS calibration is carried out for two jet mass definitions, m^{calo} and m^{TA}, individually. It is expected and observed that the resolution of m^{calo} and m^{TA} have different behavior as a function of p_{T}. Hence, in order to take advantage of their complementary information and behavior, a combined mass (m^{comb}) is introduced as a linear combination of the
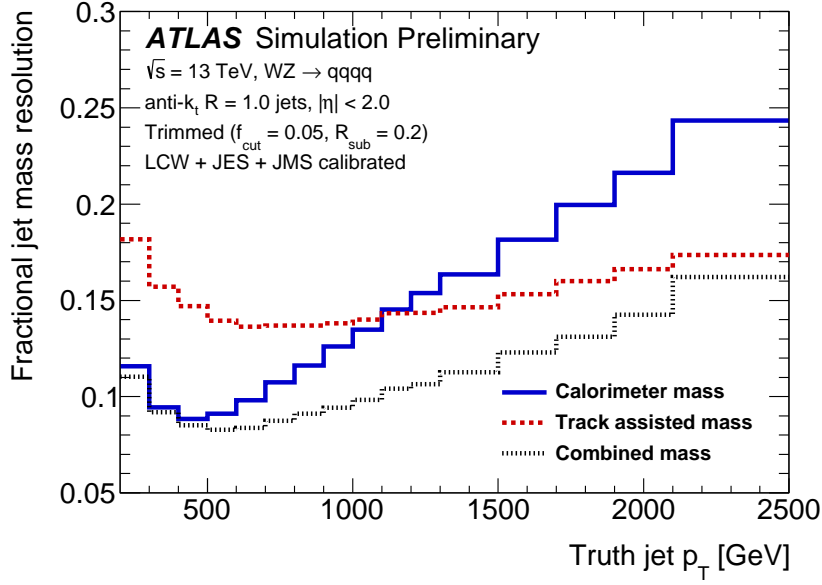
Figure 6.4: Truth jet p_T dependence of fractional jet mass resolution for three mass definitions after calibration.

two where the variables are combined in a weighted average as

$$m^{\text{comb}} = w_{\text{TA}} m^{\text{TA}} + w_{\text{calo}} m^{\text{calo}}. \tag{6.3}$$

The assigned weight w_{TA} for the track-assisted mass is defined as

$$w_{\text{TA}} = \frac{\sigma_{\text{TA}}^{-2}}{\sigma_{\text{calo}}^{-2} + \sigma_{\text{TA}}^{-2}} \tag{6.4}$$

where σ_{TA} and σ_{calo} are the m^{TA} and m^{calo} resolutions evaluated using jets which are JES and JMS calibrated in QCD multijet events, respectively. The calorimeter weight is defined as $w_{\text{calo}} = 1 - w_{\text{TA}}$, such that the sum of weights equals unity. Since m^{comb} is obtained by this linear combination of calibrated m^{calo} and m^{TA}, m^{comb} is also calibrated.

In Fig. 6.4, the fractional jet mass resolution of W and Z boson jets are compared for three different mass definitions after calibration as a function of p_T^{true}: calorimeter-based jet mass, track-assisted mass and combined mass. As it is seen in this figure for fully calibrated (JES+JMS) large-R jets built from topoclusters, m^{comb} provides better resolution across the whole p_T^{true} range. Due to its superior performance, m^{comb} is chosen to be used for the W boson and top quark tagging studies. The relevant full large-R jet calibration (JES+JMS) is applied on jets for the following large-R studies and in Fig. 6.5, m^{comb} distributions for W bosons, top quarks and background events are shown in two p_T^{true} bins.
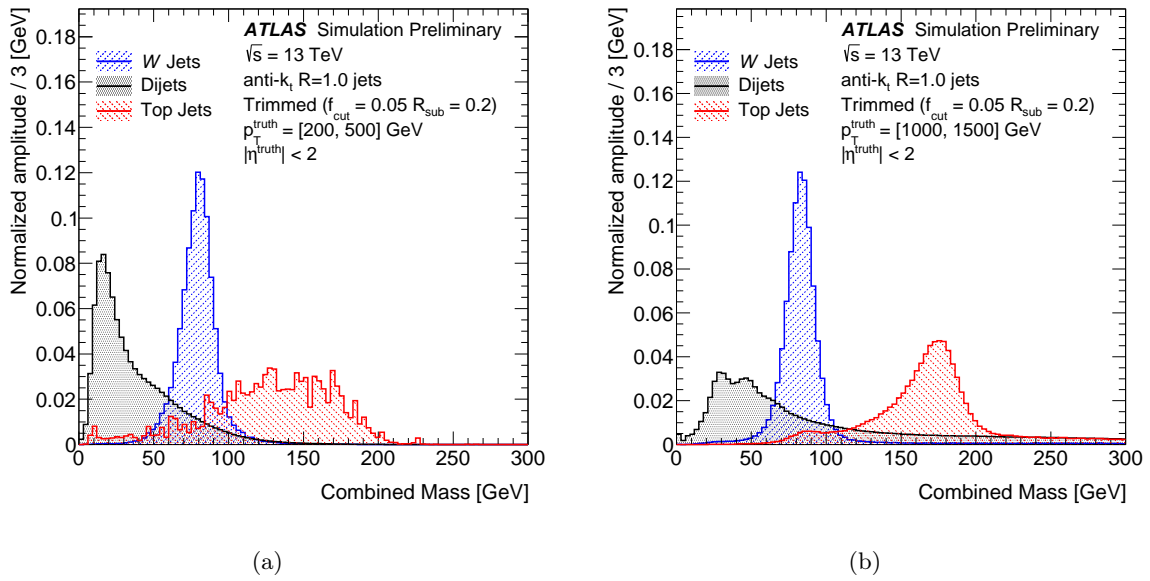
Figure 6.5:   The MC-based comparison of jet mass for top-quark, W-boson jets and light jets from multijet production presented in a low-$p_{\mathrm{T}}^{\mathrm{true}}$ (a) and a high-$p_{\mathrm{T}}^{\mathrm{true}}$ (b) bin.

6.4.2 N-subjettiness

N-subjettiness variables [87, 88], τ_N, are observables related to the subjet multiplicity. Each τ_N is a measure of how well jets can be described as containing N or fewer subjets. First, the subjets are reconstructed within the large-R jet using the k_{t} algorithm requiring exactly N subjets. If a significant fraction of the jets energy is not aligned with the subjet directions, the hypothesis of containing N subjets is unlikely. The τ_N variables are defined by

$$\tau_N = \frac{1}{d_0} \sum_i p_{\mathrm{T}i} \times \Delta R_i^{\min} \ , \ \ d_0 \equiv \sum_i p_{\mathrm{T}i} \times R \tag{6.5}$$

where R is the jet radius parameter in the k_{t} algorithm, $p_{\mathrm{T}i}$ is the p_{T} of jet constituent i and $\Delta R_i^{\min}$ is the angular distance from constituent i to the axis of the closest subjet. In this thesis, the winner-takes-all (WTA) [78] axis is used. The WTA axis uses the hardest constituent inside the subjet k to calculate the angular distance instead of the subjet axis, where $\Delta R_i^{\min} = \min(\Delta R_{k,i})$ and $k = 1, 2, ...N$.

Ratios of τ_N,

$$\tau_{NM} = \frac{\tau_N}{\tau_M}, \tag{6.6}$$

where $M = N - 1$, are generally used to discriminate hadronic W boson and top quark jets from QCD jets as they give information about how likely the large-R jet has N or fewer subjets and improve the discrimination power. For each τ_N, the value of 0 corresponds to a perfect and 1 to a poor description by N subjets. If the large-R jet contains less than N constituents, τ_N is not defined as the k_{t} algorithm fails to define the subjets. Consequently, if there are exactly N

Figure 6.6: The MC-based comparison of τ_{32} after a loose m^{comb} cut for top-quark, W-boson jets and light jets from multijet production presented in a low-$p_{\mathrm{T}}^{\mathrm{true}}$ (a) and a high-$p_{\mathrm{T}}^{\mathrm{true}}$ (b) bin.

constituents in a jet, $\tau_{NM} = 0$ and, if there are less than N constituents in a jet, τ_{NM} is undefined and it is set to a negative default value. W-boson jets are expected to have a two-prong structure whereas top-quark jets are expected to have a three-prong structure. Hence, for W tagging τ_{21} is used, whereas for top quark tagging τ_{32} is used. For top tagging, τ_{32} is found to be a very good discriminant when it is accompanied by a mass cut. In Fig. 6.6, the comparison of τ_{32} after a loose m^{comb} cut for W-boson, top-quark jets and light jets from multijet are presented in two $p_{\mathrm{T}}^{\mathrm{true}}$ bins. In this figure, the jets which have negative default τ_{32} values are not visible since they reside in the underflow bin. Due to cluster merging, it is expected that the number of constituents decrease at the presented high $p_{\mathrm{T}}^{\mathrm{true}}$ bin and the fraction of jets with $\tau_{32} = 0$ and negative default values decrease. As expected, it is observed that for top-quark jets, τ_{32} peaks at lower values compared to the multijet jets and there are only a few jets with $\tau_{32} = 0$ and default values in the lower-$p_{\mathrm{T}}^{\mathrm{true}}$ bin.

6.4.3 Energy correlation functions

Energy correlation functions follow a similar logic as the N-subjettiness and aim to quantify the number of subjets in a system [89, 90]. Unlike N-subjettiness and many jet substructure methods, energy correlation functions do not require the explicit identification of subjets. Additionally, the correlation functions better probe some soft and collinear features.

The first variables are the N-point energy correlation functions (ECF) that run over all constituents of a given jet J and are based on the p_{T} of the constituents and the angular separation R_{ij} of constituent pairs i, j. In this thesis only up to 3-point correlation functions are considered

and they are defined as,

$$\text{ECF}(0, \beta) = 1, \tag{6.7a}$$

$$\text{ECF}(1, \beta) = \sum_{i \epsilon J} p_{\text{T}i}, \tag{6.7b}$$

$$\text{ECF}(2, \beta) = \sum_{i < j \epsilon J} p_{\text{T}i} \, p_{\text{T}j} \, (\Delta R_{ij})^{\beta}, \tag{6.7c}$$

$$\text{ECF}(3, \beta) = \sum_{i < j < k \epsilon J} p_{\text{T}i} \, p_{\text{T}j} \, p_{\text{T}k} \, (\Delta R_{ij} \, \Delta R_{ik} \, \Delta R_{jk})^{\beta}, \tag{6.7d}$$

where the angular exponent β is a free parameter which takes different values depending on the problem and the particle. If a jet contains exactly N regions of high energy-density, the correlation function $\text{ECF}(N+1, \beta)$ is expected to be significantly smaller than $\text{ECF}(N, \beta)$. From these 3-point correlation functions, one can define powerful dimensionless observables which can be used to quantify if a system has N subjets. D_2^{β} is such a dimensionless variable with $N = 2$ which is used to identify two-prong jets using energy correlation double ratios. D_2^{β} is defined as

$$D_2^{\beta} = \frac{\text{ECF}(3, \beta)\text{ECF}(1, \beta)^3}{\text{ECF}(2, \beta)^3}. \tag{6.8}$$

$D_2^{1.0}$ is found to be a very good discriminant for W boson tagging when it is accompanied by a mass window and it is commonly used in ATLAS for W boson tagging. Therefore, in this thesis the simplified notation D_2 is used instead of $D_2^{1.0}$. In Fig. 6.7, the comparison of D_2 after a loose m^{comb} cut for W-boson, top-quark jets and light jets from multijet are presented in two $p_{\text{T}}^{\text{true}}$ bins. In this figure, the jets which have negative default D_2 values are not visible since they reside in the underflow bin. Similar to what is observed for τ_{32} the fraction of jets with $D_2 = 0$ and negative default values decrease due to cluster merging. As expected, it is observed that for the two-prong W boson jets, D_2 peaks at lower values compared to the multijet jets and there are only a few jets with $D_2 = 0$ and default values in the lower-$p_{\text{T}}^{\text{true}}$ bin.

In addition to D_2, energy correlation ratios e_3^{β} and C_2^{β}, defined as

$$e_3^{\beta} = \frac{\text{ECF}(3, \beta)}{\text{ECF}(1, \beta)^3}, \tag{6.9a}$$

$$C_2^{\beta} = \frac{\text{ECF}(3, \beta)\text{ECF}(1, \beta)}{\text{ECF}(2, \beta)^2} \tag{6.9b}$$

are studied. The parameter β is set to 1.0 and the corresponding observables are simply referred to as e_3 and C_2.

6.4.4 k_{t} Splitting scales

The k_{t} sequential recombination algorithm's splitting scales are determined by clustering objects together where the high-p_{T} constituents with large distances tend to be clustered last. The split-
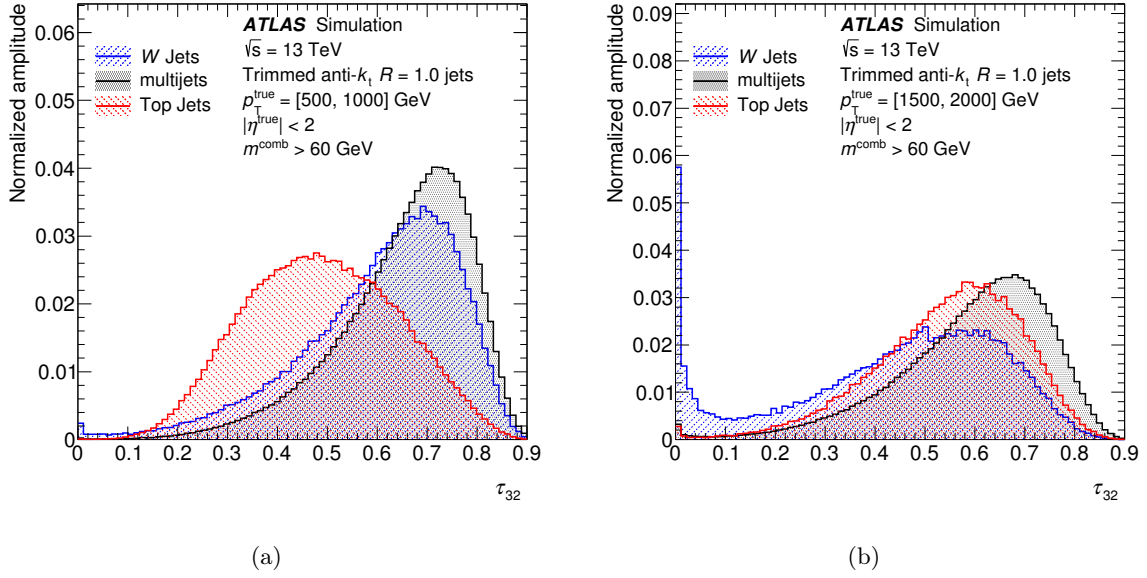
Figure 6.7: The MC-based comparison of D_2 after a loose m^{comb} cut for top-quark, W-boson jets and light jets from multijet production presented in a low-$p_{\mathrm{T}}^{\mathrm{true}}$ (a) and a high-$p_{\mathrm{T}}^{\mathrm{true}}$ (b) bin.

ting scale substructure variables are determined by reclustering the constituents of the trimmed large-R jet with the k_{t} algorithm and are defined by

$$\sqrt{d_{ij}} = \min(p_{\mathrm{T}i}, p_{\mathrm{T}j}) \times \Delta R_{ij}$$

where ΔR_{ij} is the distance between two subjets in (η, ϕ) space and $p_{\mathrm{T}i}$, $p_{\mathrm{T}j}$ are transverse momenta of these two subjets. The k_{t} splitting scales give information about the last recombination steps and whether last recombination steps merge the decay products of massive particles. Hence, indexing of subjets i and j for the splitting scales start with the last step of the k_{t} clustering history. Consequently, $\sqrt{d_{12}}$ is determined by the last clustering step and provides information about the two hardest subjets. Whereas, $\sqrt{d_{23}}$ is determined from the second-to-last clustering step. The expected values of k_{t} splitting scales are typically higher than k_{t} splitting scales from multijet events, soft, wide angle radiation and splitting scales are widely used for top-quark tagging. In case of top-quark jets, the hardest subjets captured by $\sqrt{d_{12}}$ are expected to represent the b-quark and the W boson whereas $\sqrt{d_{23}}$ is expected to cover the subjets originating from the W boson. In Fig. 6.8, the comparison of $\sqrt{d_{23}}$ after a loose m^{comb} cut for W-boson, top-quark jets and light jets from multijet are presented in two $p_{\mathrm{T}}^{\mathrm{true}}$ bins. As expected for top-quark jets, $\sqrt{d_{23}}$ peaks around $m_W/2$ where m_W is the mass of the W boson.
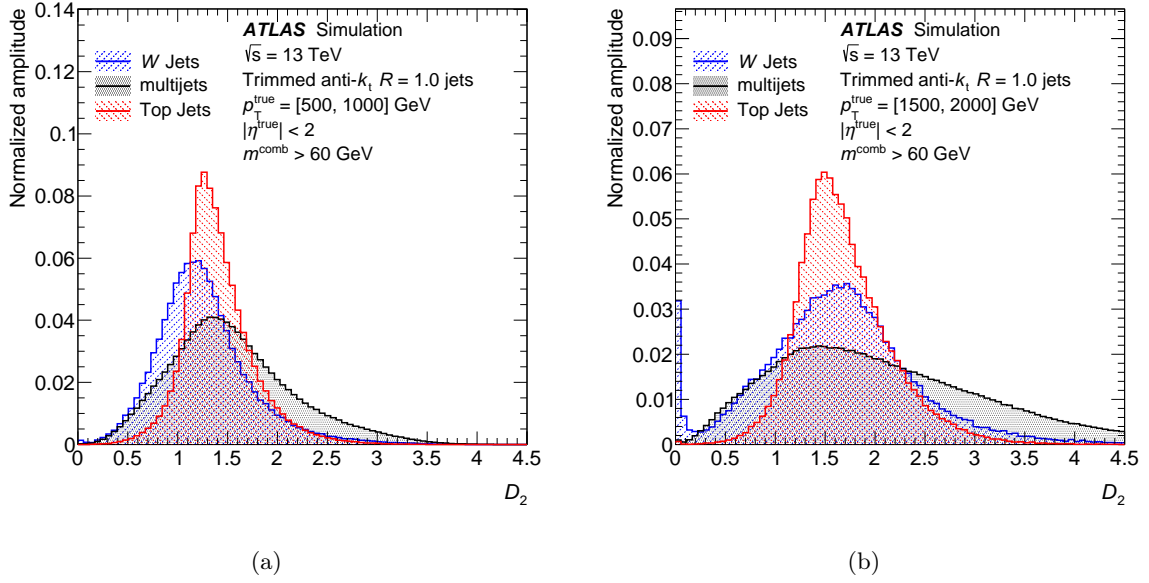
(a)   (b)
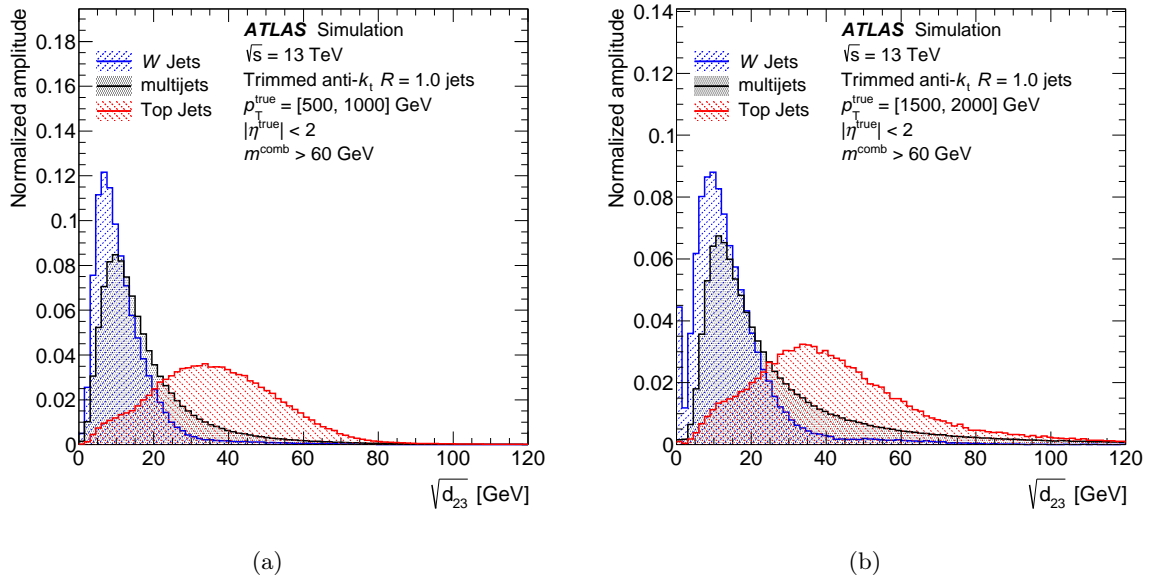
Figure 6.8: The MC-based comparison of $\sqrt{d_{23}}$ after a loose m^{comb} cut for top-quark, W-boson jets and light jets from multijet production presented in a low-$p_{\mathrm{T}}^{\mathrm{true}}$ (a) and a high-$p_{\mathrm{T}}^{\mathrm{true}}$ (b) bin.

6.5 Large-R Jet Identification Techniques and Optimization

In this section, traditional large-R jet tagging techniques (which do not employ a machine-learning method) and their optimization methods are briefly described. Machine-learning based taggers are described in chapter 7. The performance of each tagger is evaluated using, the two primary quantitative figures of merit: the signal efficiency and background rejection, are used and are defined as

$$\text{Signal efficiency}: \epsilon_{\mathrm{sig}} = \frac{N_{\mathrm{sig}}^{\mathrm{tagged}}}{N_{\mathrm{sig}}^{\mathrm{total}}} \tag{6.10a}$$

$$\text{Background rejection}: \frac{1}{\epsilon_{\mathrm{bkg}}} = \frac{N_{\mathrm{bkg}}^{\mathrm{total}}}{N_{\mathrm{bkg}}^{\mathrm{tagged}}}, \tag{6.10b}$$

where $N_{\mathrm{sig}}^{\mathrm{tagged}}$ is the (weighted) number of all signal jets passing the tagging criteria and $N_{\mathrm{sig}}^{\mathrm{total}}$ is the (weighted) number of all signal jets; in a kinematic region. The same notation is followed for background jets. Generally, the performance is quantified in several kinematic regimes based on the p_{T} of the associated anti-k_{t} $R = 1.0$ truth jet ($p_{\mathrm{T}}^{\mathrm{true}}$) to more closely resemble the kinematics of the parent particle and allow the comparison of taggers employing different jet clustering algorithms. To reduce any bias in the tagging performance due to differences between the p_{T} spectra of the signal and background jet samples, the simulated signal samples described in section 7.1 are combined and weighted (separately for W bosons and top quarks) such that the

$p_{\mathrm{T}}^{\mathrm{true}}$ distribution of the ensemble of signal jets matches that of the light-jet background[2]. The performance of all the taggers are compared in section 7.4.

6.5.1 Two-variable cut-based taggers

One of the most common approaches to tagging is employing selection requirements on two jet substructure moments. This approach was extensively studied in Run 1 and also in preparation for Run 2. These two-variable cut-based taggers determine a simple set of criteria on two jet substructure moments which provide a constant signal efficiency as a function jet p_{T} across a broad p_{T} range. This is achieved by defining optimized cut values in each p_{T} bin and defining smooth cut values by fitting the optimized cut values to a smooth function. The strategy and performance of these W-boson and top-quark taggers provide a benchmark for the other tagging strategies.

The taggers are optimized using a sample of signal W-boson or top-quark jets and background light jets extracted from the samples described in section 7.1. In each event the two reconstructed jets matched to the two highest-p_{T} truth jets within $|\eta| < 2.0$ are studied. In the case of signal, W-boson (top-quark) jets are retained if they are truth labelled and have a p_{T} greater than 200 GeV (350 GeV). The optimization is performed as a function of the p_{T} of the associated anti-k_{t} $R = 1.0$ truth jet for both W-boson and top-quark tagging following the labeling described in section 6.3.

For each $p_{\mathrm{T}}^{\mathrm{true}}$ bin, the two-dimensional variable plane of two substructure variables is considered. In order to find the set of optimal cut values, a two-dimensional scan is applied as follows. The cut on the first variable is fixed and the cut on the second variable is scanned until the signal efficiency drops under the desired working point. Next, the cut on the first variable is varied and the second variable is scanned in the same manner. This procedure is repeated until all possibilities providing the chosen signal efficiency have been scanned. The set of cuts providing the largest background rejection is considered to be the optimal set of cuts. For W-boson tagging, a single sided upper or a lower cut on a substructure moment is accompanied by a m^{comb} window. For top-quark tagging, pairs of substructure variables are investigated.

The choice of the substructure moments is based on an overall good background rejection as a function of truth jet p_{T} at a fixed signal efficiency. In this thesis, one working point for each tagger (W-boson or top-quark tagger) is chosen to be investigated and presented for simplicity. For W-boson taggers working points providing 50% fixed signal efficiency and for top-quark taggers working points providing 80% fixed signal efficiency are presented. This choice of signal efficiency working points are mainly based on the commonly used working points in BSM searches.

Once a set of most powerful variable combinations are identified, smooth cut values are defined to provide smooth-cut taggers for these powerful variable combinations as follows. The smooth cut values are provided by fitting the optimized single sided cut values as a function of jet p_{T} with a polynomial function for each variable. For W-boson tagging, a more physically motivated

---

[2]The weighting procedure is explained in section 7.2.1.

(a)   (b)
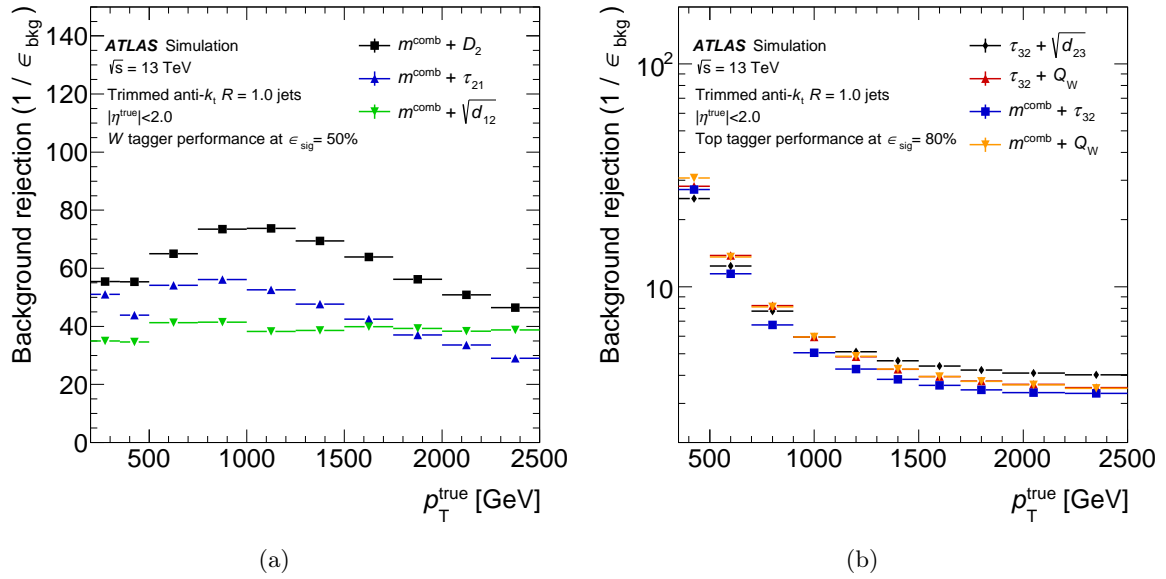
Figure 6.9:   The W-boson (a) and top-quark tagging (b) background rejection as a function of jet p_T^{true} for the best performing two-variable combinations at 50% fixed signal efficiency for W-boson tagging and 80% fixed signal efficiency for top-quark tagging.

cut function is used to fit the two sided mass cuts as

$$m_{\text{cut}}^{\text{comb}} = \sqrt{\left(\frac{p_0}{p_T} + p_1\right)^2 + (p_2 \cdot p_T + p_3)^2} \qquad (6.11)$$

where p_0, p_1, p_2 and p_3 are parameters of the fit. This procedure delivers good performance in terms of a fixed signal efficiency and smooth variable cuts as a function of p_T.

The resulting background rejections as a function of the jet p_T^{true} for the most powerful two-variable combinations of W-boson tagger and for top-quark tagger are presented in Fig. 6.9. In the case of W tagging, the combination of m^{comb} window cut and one-sided D_2 cut is the best performing tagger across the entire p_T range and it is taken as the baseline tagger for W tagging. In the case of top-quark tagging, the most powerful taggers all provide a large background rejection at low p_T^{true}, plateauing at a lower value for high jet p_T^{true} mostly due to the migration of the light-jet mass distribution to higher values and a looser τ_{32} cut to maintain the constant signal efficiency. The two-variable combinations that do not involve mass perform marginally better than those with mass across the entire p_T range. The specific two-variable cut-based top-quark tagger used in an analysis may depend on the context of the analysis and not only on the resulting background rejection. Therefore, the baseline two-variable cut-based top-quark tagger is selected to be the one composed of one-sided selections on m^{comb} and τ_{32} since it has been commonly used in ATLAS.

6.5.2 Shower deconstruction

Shower deconstruction (SD) [98] is a method that attempts to classify jets according to the compatibility of the radiation pattern of the jet with a predefined set of parton shower hypotheses. For a set of subjets, consistency with the decay of a top quark is ensured by requiring that the jet has at least three subjets, that two or more subjets have a mass in a window centered around the W-boson mass, and that at least one more subjet can be added to obtain a total mass in a window centered around the top-quark mass. If the jet passes these consistency requirements, then a set of potential shower histories is constructed for the signal and background models. To quantify the compatibility with the signal or background hypotheses, a probability is assigned to each shower history based on the parton shower model. From these probabilities, the likelihood ratio, χ, of signal and background hypotheses is defined. Finally, $\log \chi$ is used as the discriminant to tag top quarks.

The shower deconstruction tagging method was studied extensively in Run 1. In Run 2 shower deconstruction was recommissioned in the context of the search for a heavy W' boson decaying to a top quark and a bottom quark as published in Ref. [99]. One limitation of the algorithm was observed to be the number of large-R jets which fail the consistency criteria at high p_T due to the smaller number of subjets in that kinematic region. To be able to reduce this effect, the method in which subjets are constructed is updated in Run 2. The SD algorithm presented in this thesis and in Ref. [99] uses the k_t jet clustering algorithm, run on the constituents of the trimmed large-R jet. As k_t splitting scales are less dependent on the large-R jet p_T than the geometric distance between the jet and its constituents, a criterion is imposed to stop clustering if k_t splitting scales larger than 15 GeV are found. Then, the resulting set of subjets are used as subjet inputs to shower deconstruction. The set of subjets is limited to a maximum of the six highest-p_T subjets due to the computationally-expensive nature of SD. The parameters controlling the top-quark topology check using subjet pairings and triplets are fixed to 20 GeV and 40 GeV.

6.5.3 HEPTopTagger

Another approach to top-quark tagging is the HEPTopTagger algorithm [85, 86]. HEPTopTagger relies on reconstructing jets using the C/A algorithm with $R = 1.5$ to allow the tagging of boosted top quarks with p_T as low as 200 GeV. The tagger takes advantage of the C/A clustering sequence which by definition attempts to reverse the decay structure of the top-quark decay.

The HEPTopTagger presented in performance comparisons in this thesis is the original algorithm, introduced in Ref. [85], executed with the settings found to be optimal in Ref. [82], where further details can be found.

Chapter 7

Machine Learning Based Identification of Top Quarks and W Bosons

Some of the jet observables introduced in section 6.4 contain complementary information. It was shown in section 6.5.1 that employing selection requirements on two jet substructure moments provides strong discrimination between signal and background both for W-boson tagging and top-quark tagging. It has also been previously shown that the use of machine learning techniques to combine multiple jet moments provides stronger discrimination [100, 101].

This chapter focuses on the use of jet moments as inputs to deep neural networks to build binary jet classifiers which discriminate W-boson or top-quark jets from the gluon and light(all non-top)-quark jet background. These taggers are referred to as DNN W and DNN top taggers, respectively. The optimization of the deep neural networks and the procedure of defining the tagging working points are presented for DNN W and DNN top taggers. Additionally, other machine learning based identification methods are presented.

The performance of the DNN W and the DNN top taggers is compared with the performance of other W-boson and top-quark tagging techniques in MC simulation. Finally, the data recorded in 2015 and 2016 is used to study the performance of these tagging techniques in data. The tagging signal efficiency and background rejection in data are measured using boosted lepton+jet $t\bar{t}$, dijet and γ + jet topologies.

7.1 Samples

In this section the simulated and collected data samples used in the W-boson and top-quark tagging studies are described. The taggers studied in this thesis are initially designed using MC simulated samples for two signal processes (providing hadronically-decaying W bosons and top quarks) and one background process (providing gluons and non-top quarks).

The dijet process was used to simulate jets from gluons and non-top quarks. It was modelled using the leading-order PYTHIA8 (v8.186) [56] generator with the NNPDF2.3LO [102] PDF set and the A14 tune [103]. Dijet events were generated in slices of leading anti-k_t $R = 0.6$ truth jets, allowing to populate the kinematic region of interest (trimmed anti-k_t $R = 1.0$ jet p_T up to 2.5 TeV). Event-by-event weights were applied to correct for this generation method and to

produce the expected smoothly falling jet p_T distribution of the multijet background.

Signal samples containing either high-p_T top-quark or W-boson jets were obtained from two simulated beyond the Standard Model processes. For obtaining the W-boson signal sample, high-mass sequential standard model [104] $W' \to WZ \to q\bar{q}q\bar{q}$ events were used. For obtaining the top-quark signal sample, high-mass sequential standard model $Z' \to t\bar{t}$ events were used. W bosons were required to decay hadronically at the simulation level, whereas hadronically-decaying top quarks were obtained by selecting the hadronic decays at the truth-matching level introduced in section 6.3. The two signal processes were simulated using the PYTHIA8 generator with the NNPDF2.3LO PDF set and A14 tune. In order to populate the kinematic region of interest, multiple values of the W' and Z' boson mass between 400 and 5000 GeV.

For the study of W-boson and top-quark jets in data, described in section 7.5, simulated samples are needed to model the $t\bar{t}$ signal and backgrounds.

The POWHEG-BOX v2 generator [105–107] was used to simulate $t\bar{t}$ and single-top-quark production in the Wt- and s-channels at next-to-leading order (NLO). The single-top-quark t-channel process used the NLO POWHEG-BOX v1 generator and the CT10 [108] NLO PDF set. For all processes involving top quarks, the parton shower, fragmentation, and the underlying event were simulated using PYTHIA6 (v6.428) [109] with the CTEQ6L1 [110] PDF set and the corresponding Perugia 2012 tune (P2012) [111]. The top-quark mass (m_top) was set to 172.5 GeV and the parameter which controls the matching of the matrix element to the parton shower (h_damp) was set to the mass of the top quark. The $t\bar{t}$ process is normalized to the cross-sections predicted to next-to-next-to-leading order (NNLO) in α_S and next-to-next-to-leading logarithm (NNLL) in soft-gluon terms while the single-top-quark processes are normalized to the NNLO cross-section predictions [112].

Several additional variations of the $t\bar{t}$ generator are used for the estimation of modeling uncertainties. Estimates of the parton showering, hadronization modeling and underlying-event uncertainty are derived by comparing results obtained with the POWHEG-BOX v2 generator interfaced to HERWIG++ (v2.7.1) [113] instead of PYTHIA6. To estimate the hard-scattering modeling uncertainty, the NLO MADGRAPH5_AMC@NLO (v2.2.1) generator [114], referred to as MC@NLO, is used with PYTHIA6. To estimate the uncertainty in the modeling of additional radiation, the POWHEG-BOX v2 generator with PYTHIA6 is used with modified renormalization and factorization scales ($\times2$ or $\times0.5$) and a simultaneously modified h_damp parameter value ($h_\mathrm{damp} = m_\mathrm{top}$ or $h_\mathrm{damp} = 2 \times m_\mathrm{top}$) as described in Ref. [115].

Samples of W/Z+jets and Standard Model diboson ($WW/WZ/ZZ$) production were generated with final states that include either one or two charged leptons. The SHERPA [58] generator with version 2.1.1 and version 2.2.1 were used to simulate these processes at NLO with the CT10 PDF set for the diboson and W/Z+jets production processes, respectively. The W/Z+jets events are normalized to the NNLO cross-sections [116].

For the study of γ+jet events in data, events containing a photon with associated jets were simulated using the SHERPA 2.1.1 generator, requiring a photon p_T above 140 GeV. Matrix elements were calculated with up to four partons at LO and merged with the SHERPA parton shower [117]

using the ME+PS@LO prescription [118]. The CT10 PDF set was used in conjunction with the dedicated parton shower tune.

The MC samples were passed through the full ATLAS detector simulation [119] based on the simulation detector model in GEANT4 [59]. Additional simulated pp collisions which were generated using PYTHIA8 (v8.186) with the A2M [103] tune and MSTW2008LO PDF set [120] were overlaid to simulate the effects of pileup, with a mean number of 24 collisions per bunch crossing. All simulated events were then processed using the same reconstruction algorithms and analysis chain as data as explained in section 4.3.

Data were collected in three broad categories to study the signal efficiency and background rejection. For the signal, a set of top-quark and W-boson jet candidates is obtained from a sample of $t\bar{t}$ candidate events with the lepton + jets decay signature in which one top quark decays semileptonically and the other decays hadronically. The background is studied using data samples enriched in dijet events and γ + jet events. In addition to covering different p_T regions, in the γ + jet topology the jets are mostly initiated by quarks over the full p_T range studied, whereas for the dijet topology the fraction of quarks initiating jets is slightly smaller than the gluon fraction at low p_T and becomes large at high p_T. The data for the lepton + jets $t\bar{t}$ and γ + jet studies were collected during normal operations of the detector while all relevant detector subsystems were fully functional; corresponding to an integrated luminosity of 36.1 fb^{-1}. For the dijet analysis, additional data where the toroid magnet was turned off are used, corresponding to an additional luminosity of 0.6 fb^{-1}. For all datasets, only the events in which at least one primary vertex was reconstructed are used [121].

The lepton + jets events were collected with a set of single-electron and single-muon triggers. These triggers became fully efficient for a lepton p_T greater than 28 GeV. The dijet events were collected with a single large-R (anti-k_t $R = 1.0$) jet trigger. This trigger became fully efficient for an offline jet p_T of approximately 450 GeV. The γ + jet events were collected with a single-photon trigger that became fully efficient for an offline photon p_T of approximately 155 GeV.

7.2 Optimization and Definition of the DNN *W* and DNN top taggers

This section presents the MC simulation-based optimization and definition of the DNN-based taggers which use jet substructure moments as inputs. First the jet selections and weighting schemes employed for different purposes are described. Then the optimization of DNN hyperparameters and input variables are presented. This is followed by the definition of the tagging working points that are used in the later sections.

7.2.1 Jet selection and jet-based weights

To start with, the combined samples of background, W-boson and top-quark jets, which are obtained in section 7.1 and used as inputs for the reference taggers introduced in section 6.5, are

randomly split into training (59.5%), validation (10.5%) and testing (30%) sets for the DNN-based taggers under study. It is important to note that the background sample has many more events compared to the signal samples and this difference affects some of the choices made in this section. As a reminder, the training set is used to minimize the loss, the validation sample is used to ensure that there is no overtraining and the testing set is used to evaluate the performance of the taggers. The same relevant p_{T} thresholds (200 GeV for W-boson tagging, 350 GeV for top-quarks tagging) and $|\eta|$ requirement ($|\eta| < 2.5$) are applied for the DNN W and DNN top taggers as in the reference taggers introduced in section 6.5. Due to the limited sample size and small number of jets at high $p_{\mathrm{T}}^{\mathrm{true}}$, the DNNs are trained in a single $p_{\mathrm{T}}^{\mathrm{true}}$ bin; they are trained and evaluated up to $p_{\mathrm{T}}^{\mathrm{true}}$ of 2 TeV. Additional object selections and weighting schemes for the testing, training, validation sets are presented next and summarized in Tab. 7.1.

Testing set and event weights

When evaluating the performance of the taggers, 30% of each signal and background sample are used for the testing set. The simulated signal samples (separately for W bosons and top quarks) are weighted in $p_{\mathrm{T}}^{\mathrm{true}}$ bins such that the $p_{\mathrm{T}}^{\mathrm{true}}$ distribution of the ensemble of signal jets matches that of the dijet background. This weighting scheme prevents the irregular unphysical jet p_{T} distributions arising from the combination of the signal samples with different generated heavy resonance masses and removes any bias on the tagging performance due to the difference in the p_{T} spectrum of the signal and background jet samples. These weights are referred to as the testing weights.

Training and validation sets, weights and preprocessing

The training and validation sets employ the same cuts and follow the same weighting scheme. To ensure that all used jet substructure features are well defined for the training and validation jets, two additional selection criteria are applied on the jet mass ($m^{\mathrm{comb}} > 40$ GeV) and number of constituents ($N^{\mathrm{const}} > 2$) as summarized in Tab. 7.1. These additional selection criteria are referred to as the training criteria collectively. The jets which fail the training criteria are removed from the training and validation sets since this choice does not reduce the size of the training and validation sets significantly.

The training and validation signal sets in total correspond to 70% of the signal samples. Starting from 70% of the background samples, two separate background samples are created by randomly downsampling to match the number of training and validation jets in each of the signal training and validation sets (taking into account only the jets that pass the training criteria). The requirement of equal numbers of signal and background jets for the training and validation sets prevents learning one flavor more than the other because of the limited signal sample size and avoids using very large weights during training which are not optimal for the DNNs[1].

---

[1] While calculating the training weights in $p_{\mathrm{T}}^{\mathrm{true}}$ bins, large weights would be obtained if one reweights the signal to match the background not only in shape but also in scale.

Set	Training and validation		Testing	
Tagger type	W boson	Top Quark	W boson	Top Quark
p_T weighting	uniform	uniform	to multijet	to multijet
Truth p_T [GeV]	[200,2000]	[350,2000]	[200,2000]	[350,2000]
m^comb [GeV]	> 40	> 40	$-$	$-$
N^const	> 2	> 2	$-$	$-$

Table 7.1: Summary of the selections used for training, validation and testing samples.

The resulting signal and background sets are weighted individually in $p_\mathrm{T}^\mathrm{true}$ bins for each topology such that the $p_\mathrm{T}^\mathrm{true}$ distribution is uniform. Employing a weighting scheme which provides a uniform $p_\mathrm{T}^\mathrm{true}$ distribution for both signal and background jets prevents the algorithms from discriminating between signal and background using differences in their p_T distributions. Moreover, since the DNNs are trained in one inclusive $p_\mathrm{T}^\mathrm{true}$ bin, a uniform $p_\mathrm{T}^\mathrm{true}$ aims to give similar importance to jets in different p_T ranges during training.

As described in section 3.2, the transformation of inputs is an important part of build a good model using DNNs. For the DNN W and DNN top taggers, the input variables are standardized prior to training.

7.2.2 Training and optimization

DNN W and DNN top taggers are obtained by training fully-connected feed-forward neural networks with multiple hidden layers. The performance of the DNN taggers is sensitive to the hyper-parameters and the input variables that are used to train the DNN. Therefore, optimizing the hyper-parameters and input variables is of high importance.

The DNN is trained on different sets of input variables in order to find the optimum configuration. In order to identify the most performant DNN for each set of inputs, a grid search of hyper-parameters is carried out. The list of scanned hyper-parameters is summarized in Tab. 7.2. In addition to the listed learning hyper-parameters and architecture types, the number of nodes in the hidden layers are also varied. All of the DNNs are trained for 100 epochs, with early-stopping. For all of the DNN trainings, the optimizer algorithm Adam [33] is used as it has been proven to be successful for numerous similar machine learning problems. Although the activation function of the input and hidden layers is varied during the grid search, the activation function of the output layer is chosen to be a sigmoid due to its applicability to binary classification problems.

Sets of input variables are defined by grouping the inputs related to the corresponding signal. A summary of all of the input variable groups tested for the DNN training is shown in Tab. 7.3 and 7.4 for W-boson and top-quark tagging, respectively. The grouping is chosen by selecting variables based on what features of the substructure they describe, their dependence on the momentum scale of the jet's substructure objects, the amount of information they provide on the jet-energy and jet-mass scale and their dependence on other variables. This approach provides

input variables which contain complementary information to train DNNs and gives insights on the importance of different variables for the training and tagger performance as illustrated with a few examples in the following. For top-quark tagging, Group 1 is constructed by the substructure variables which aim to quantify the number of prongs in a jet. Moreover, these substructure variables do not have any dependence on the momentum scale of the jet's substructure objects, they do not provide any information on the jet-energy and jet-mass scale and are all defined as ratios of other substructure variables [2]. In order to observe the impact of including m^{comb} and p_{T} on the training and tagger performance, Groups 2 and 3 are constructed by extending Group 1 with m^{comb} and p_{T} successively. Next, whether the DNN could reconstruct the higher-level substructure variables in Group 3 and extract the necessary information they contain is tested by defining Group 4 with the variables which are the building blocks of the variables in Group 3 [3]. Group 8 is then constructed by adding information on scale of the substructure objects, k_{t} splitting scale and Q_w, to Group 3 and hence allows to observe the impact of this new information.

During the input and hyper-parameter optimization, the performance of the DNNs is characterized by the relative signal efficiency and relative background rejection in a single inclusive $p_{\text{T}}^{\text{true}}$ bin. Relative signal efficiency and relative background rejection are evaluated using the jets from the testing set and take into account only the jets that satisfy the training criteria. Moreover, the training weights which provide uniform $p_{\text{T}}^{\text{true}}$ spectra are used to evaluate the relative signal efficiency and background rejection such that the overall performance can be characterized by a single $p_{\text{T}}^{\text{true}}$ bin for optimization. Relative signal efficiency is defined as

$$\epsilon_{\text{sig}}^{\text{rel}} = \frac{N_{\text{signal},m^{\text{comb}}>40 \text{ GeV},N^{\text{const}}>2}^{\text{tagged}}}{N_{\text{signal},m^{\text{comb}}>40 \text{ GeV},N^{\text{const}}>2}^{\text{total}}}$$

and in a similar manner, relative background rejection is defined as $1/\epsilon_{\text{bkg}}^{\text{rel}}$. Relative background rejection at a fixed relative signal efficiency of 50% (*W*-boson tagging) or 80% (top-quark tagging) are used to evaluate the performance of different taggers.

The relative background rejection achieved with the training weights described in section 7.2.1 and inclusively in jet $p_{\text{T}}^{\text{true}}$ is presented in Fig. 7.1. It is confirmed that the performance of the DNN tagger depends on both the number of variables and how much of the necessary information is stored in different groups. Within statistical uncertainties, the number of variables necessary for maximum rejection at a fixed relative signal efficiency of 50% (*W*-boson tagging) and 80% (top-quark tagging) is found to be 12 variables for *W*-boson tagging (Group 8 in Tab. 7.3) and 13 variables for top-quark tagging (Group 9 in Tab. 7.4). The chosen DNN hyper-parameters and inputs are summarized in Tab. 7.5.

DNNs are prone to overtraining, therefore it is necessary to test the robustness of the obtained

---

[2]Example: $\tau_{32} = \tau_3/\tau_2$, D_2 is a function of other ECFs as defined in Eqn. 6.8.

[3]ECF_1 and ECF_2 variables are not included exclusively since jet p_{T} and m^{comb} are very similar to these variables and they have the additional advantage of being calibrated.

	W-Boson Tagging		Top-Quark Tagging	
Layer type	Dense	Maxout	Dense	Maxout
Number of hidden layers	3, 4, 5, 6	3, 4, 5, 6	3, 4, 5, 6	3, 4, 5, 6
Activation function	relu, tanh	-	relu, tanh	-
Learning rate	10^{-5}, 10^{-4}, 10^{-3}	10^{-5}, 10^{-4}, 10^{-3}	10^{-5}, 5×10^{-5}, 10^{-4}	10^{-5}, 5×10^{-5}, 10^{-4}
L1 Regularizer	10^{-3}, 10^{-2}	10^{-3}, 10^{-2}	10^{-3}, 10^{-2}	10^{-3}, 10^{-2}
NN weight initialization	Glorot uniform, He normal	Glorot uniform	Glorot uniform, He normal	Glorot uniform
Batch size	200	200	200	200
Batch normalization	Yes	Yes	Yes	Yes
Number of Maxout layers	-	5, 10, 15, 20, 25	-	5, 10, 15, 20, 25
Training groups	9 groups	7 groups	9 groups	7 groups

Table 7.2: Hyper-parameter and architecture scan for the DNN grid search.

Group 1	τ_1, τ_2, e_3, m^{comb}, p_{T}
Group 2	τ_1, τ_2, e_3, m^{comb}, p_{T}, $\sqrt{d_{12}}$, KtDR
Group 3	τ_{21}, C_2, D_2, R_2^{FW}, $\mathcal{P}$, a_3, A, z_{cut}
Group 4	τ_{21}, C_2, D_2, R_2^{FW}, $\mathcal{P}$, a_3, A, z_{cut}, m^{comb}
Group 5	τ_{21}, C_2, D_2, R_2^{FW}, $\mathcal{P}$, a_3, A, z_{cut}, m^{comb}, p_{T}
Group 6	τ_1, τ_2, e_3, m^{comb}, p_{T}, R_2^{FW}, $\sqrt{d_{12}}$, KtDR, a_3, A
Group 7	τ_{21}, C_2, D_2, R_2^{FW}, $\mathcal{P}$, a_3, A, z_{cut}, m^{comb}, $\sqrt{d_{12}}$, $KtDR$
Group 8	τ_{21}, C_2, D_2, R_2^{FW}, $\mathcal{P}$, a_3, A, z_{cut}, m^{comb}, p_{T}, $\sqrt{d_{12}}$, $KtDR$
Group 9	τ_1, τ_2, τ_{21}, $\sqrt{d_{12}}$, C_2, D_2, e_3, m^{comb}, p_{T}, R_2^{FW}, $\mathcal{P}$, a_3, A, z_{cut}, KtDR

Table 7.3: W-boson tagging input groups for DNN as in Fig. 7.1.

Group 1	C_2, D_2, τ_{21}, τ_{32},
Group 2	C_2, D_2, τ_{21}, τ_{32}, m^{comb}
Group 3	C_2, D_2, τ_{21}, τ_{32}, m^{comb}, p_{T}
Group 4	τ_1, τ_2, τ_3, e_3, m^{comb}, p_{T}
Group 5	C_2, D_2, τ_{21}, τ_{32}, $\sqrt{d_{12}}$, $\sqrt{d_{23}}$, Q_w
Group 6	C_2, D_2, τ_{21}, τ_{32}, $\sqrt{d_{12}}$, $\sqrt{d_{23}}$, Q_w, m^{comb}
Group 7	τ_1, τ_2, τ_3, e_3, m^{comb}, p_{T}, $\sqrt{d_{12}}$, $\sqrt{d_{23}}$, Q_w
Group 8	C_2, D_2, τ_{21}, τ_{32}, $\sqrt{d_{12}}$, $\sqrt{d_{23}}$, Q_w, m^{comb}, p_{T}
Group 9	τ_1, τ_2, τ_3, τ_{21}, τ_{32}, $\sqrt{d_{12}}$, $\sqrt{d_{23}}$, Q_w, C_2, D_2, e_3, m^{comb}, p_{T}

Table 7.4: Top-quark tagging input groups for DNN as in Fig. 7.1.

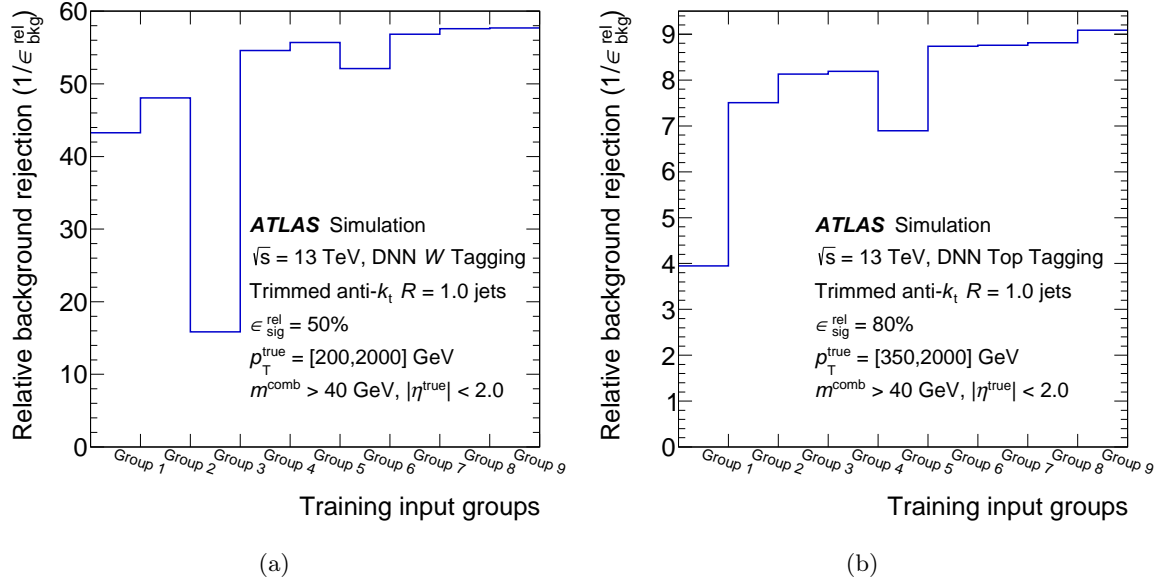(a)   (b)

Figure 7.1: Distributions showing the training with different sets of variables and relative improvement in performance for the DNN W-boson (a) and top-quark (b) taggers at the 50% and 80% relative signal efficiency working point, respectively. Only jets which satisfy the training criteria are considered when calculating the relative signal efficiency and relative background rejection. The performance is evaluated with uniform $p_{\mathrm{T}}^{\mathrm{true}}$ spectra. Uncertainties are not presented.

	W-Boson Tagging	Top-Quark Tagging
Software package	Keras 1.0.8 with Theano backend, lwtnn 2.0
Layer type	Dense	Dense
Number of hidden layers	4	5
Number of nodes per hidden layer	16, 14, 9, 6	18, 16, 14, 10, 5
Activation function	rectified linear unit (relu)	rectified linear unit (relu)
Optimizer	Adam	Adam
Learning rate	0.0001	0.00005
L1 Regularizer	0.001	0.001
NN weight initialization	Glorot uniform	Glorot uniform
Batch size	200	200
Batch normalization	Yes	Yes
Number of epochs	100 with early stopping	100 with early stopping
Training input group	Group 8	Group 9

Table 7.5: Summary of DNN software, chosen hyper-parameters and input variables for W-boson and top-quark tagging.
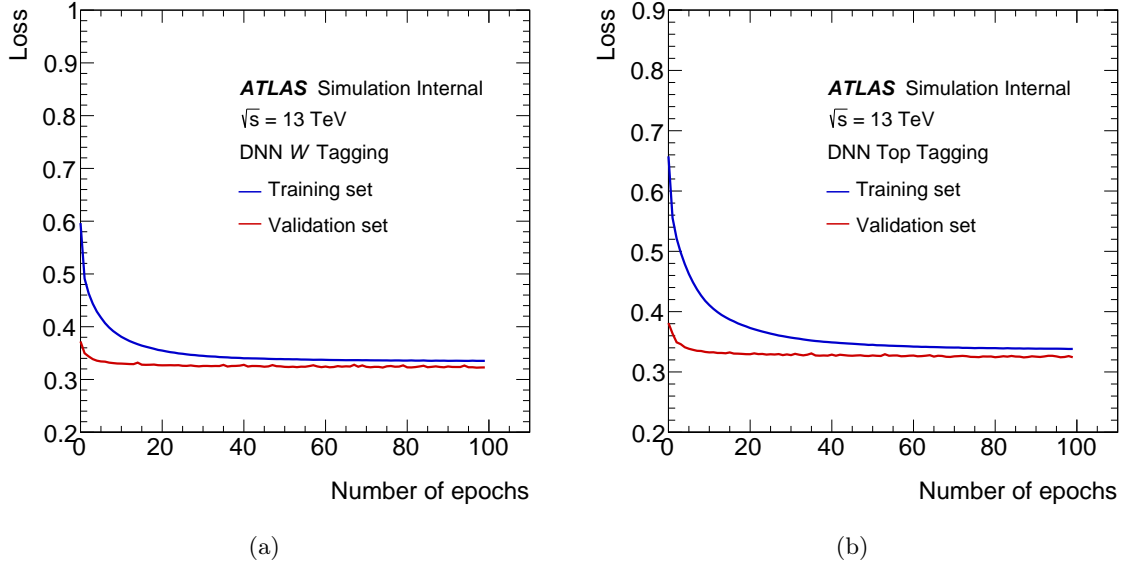
Figure 7.2: Training set and validation set loss distributions for W-boson tagging (a) and top-quark tagging (b) shows the robustness of DNNs with respect to overtraining.

DNNs. To ensure that there is no overtraining, the loss development of the chosen DNNs is studied. The losses on the training and validation sets of these DNNs are compared as a function of the number of epochs in Fig. 7.2. The training set loss and validation set loss are both observed to decrease over time and reach a plateau. This indicates that there is no overtraining. Moreover, as expected, the validation set loss is found to be systematically less than the training set loss. This is expected as the training set loss is calculated continuously during training whereas the validation set loss is calculated after each epoch and after the loss has been regularized.

7.2.3 Chosen taggers, definition of tagging working points

The output of each DNN is a single discriminant that allows for the classification of a jet as either a signal-like jet (top-quark or W boson) or background-like jet (gluon/other quark). The distributions of the DNN W and DNN top discriminants are shown in Fig. 7.3 in a single $p_{\mathrm{T}}^{\mathrm{true}}$ bin for W-boson, top-quark and background jets where relevant.

Similar to the baseline taggers presented in section 6.5, for the chosen DNN taggers the working points are defined as a function of the reconstructed jet p_{T} so that they yield a constant signal efficiency as a function of p_{T}. These taggers are composed of a smooth single-sided cut on the relevant DNN discriminant and a fixed m^{comb} requirement of $m^{\mathrm{comb}} > 40$ GeV. Moreover, if a jet passes the $m^{\mathrm{comb}} > 40$ GeV criteria but fails the $N^{\mathrm{const}} > 2$ criteria, it is tagged as signal independent of the discriminant value. In order to obtain the smooth cut, first discrete requirements for each p_{T} bin are defined by varying the DNN discriminant to obtain the desired signal efficiency. Once the DNN discrete discriminant cut values are defined for each p_{T} bin, smooth cut values on the DNN discriminants are obtained by fitting the obtained discrete cut
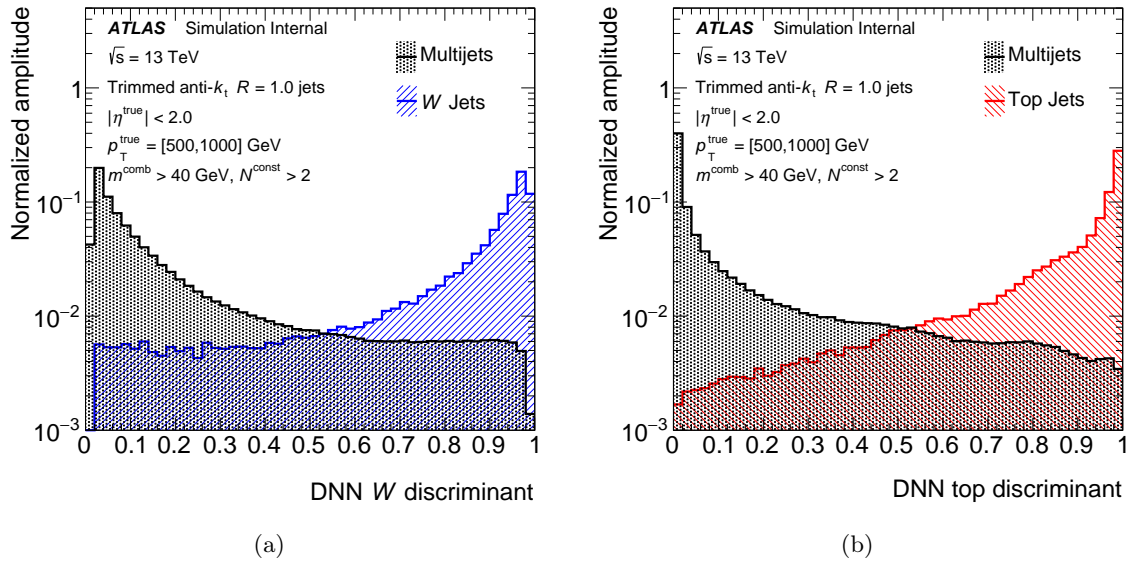
Figure 7.3: Discriminant distributions of DNN W (a) and DNN top (b) taggers.

values as a function of jet p_T with a polynomial function. This procedure provides the tagging working points with a signal efficiency of 50% for W-boson tagging or 80% for top-quark tagging, where the relative variation of the signal efficiency, calculated by comparing the variation from the design signal efficiency with he design signal efficiency for the taggers is less than 5%.

7.3 Other Machine Learning-Based Taggers

7.3.1 Boosted decision tree

An alternative machine learning technique which can be used to combine multiple jet moments is the BDT. The usage of BDTs is thus investigated in parallel following a procedure similar to the DNNs. The training and validation sets used by the DNN are combined to form the training set of the BDTs because BDTs do not employ a validation set in these studies.

The set of observables used to train the BDT is determined using a procedure in which the observables applicable to each topology, listed in Tab. 6.1, are sequentially added as an input. For each successive observable that is to be added to the classifier, the BDT classifier is trained with jets from the training set and the relative performance is evaluated using jets from the testing sample. The variable which gives the greatest increase in relative background rejection at a fixed relative signal efficiency of 50% (W-boson tagging) or 80% (top-quark tagging) is retained. The smallest set of variables which reaches the highest relative background rejection within statistical uncertainties is selected. The training parameters of the BDT were determined and fixed based on an optimization approach similar to that of the DNN. The main difference is that once the optimal BDT settings were found, the settings were not varied for different set of input variables. The BDT software and chosen settings are presented in Tab. 7.6. The

minimum number of selected variables is 11 for W-boson tagging and 10 for top-quark tagging. The relative background rejection achieved at each stage for both classifiers is shown in Fig. 7.4.

Setting Name	Description	Choice
BoostType	Type of boosting technique	GradientBoost
NTrees	Number of trees in the forest	500
MaxDepth	Max depth of the decision tree allowed	20
MinimumNodeSize	Minimum fraction of training events required in a leaf node	1.0%
Shrinkage	Learning rate for GradientBoost algorithm	0.5
UseBaggedBoost	Use only a random (bagged) subsample of all events for growing the trees in each iteration	True
BaggedSampleFraction	Relative size of bagged event sample to original size of the data sample	0.5
SeparationType	Separation criterion for node splitting	GiniIndex
nCuts	Number of grid points in variable range used in finding optimal cut in node splitting	500

Table 7.6: Brief description of the BDT parameters and the chosen values.

7.3.2 Topocluster-based tagger

The substructure moments and taggers previously presented use high-level variables which have specific physical motivation or interpretation. Although combining multiple variables with different physical interpretations lead to improvements, there still might be unutilized information related to the jet energy flow. Moreover, often jet substructure moments lose their discrimination power due to merging of clusters and loss of granularity at very high p_T. Some simulation-based studies have shown that the direct use of jet constituents as inputs to an ML algorithm can lead to improvements in discrimination power compared to jet-moment-based discriminants [122–125]. Therefore, "TopoDNN", a tagger that makes use of lower-level inputs, is investigated. TopoDNN focuses on the identification of high-p_T top quarks with $p_T > 450$ GeV.

The TopoDNN algorithm studied here employs feed-forward NNs and follows the tagger described in Ref. [124]. The input features used in this tagger are the four-vectors of a fixed-number of topoclusters in the individual large-R anti-k_t trimmed jet. Since the topoclusters are defined to be massless, the p_T, η and ϕ of the topoclusters are used as inputs to train the DNN. In contrast with some other taggers which use jet constituents as inputs, this tagger does not employ pixelation nor an architecture which employ sequenced, variable-length inputs. As a preprocessing step, the p_T of each topocluster is normalized to bring the scale of the input variables to the same magnitude as the topocluster η and ϕ. The (η, ϕ) location of the set of topoclusters is then transformed by a translation, a rotation, and a flip based on the assumed three-prong topology of a top-quark decay. The 10 highest-p_T topoclusters are used as input to the DNN. If a jet
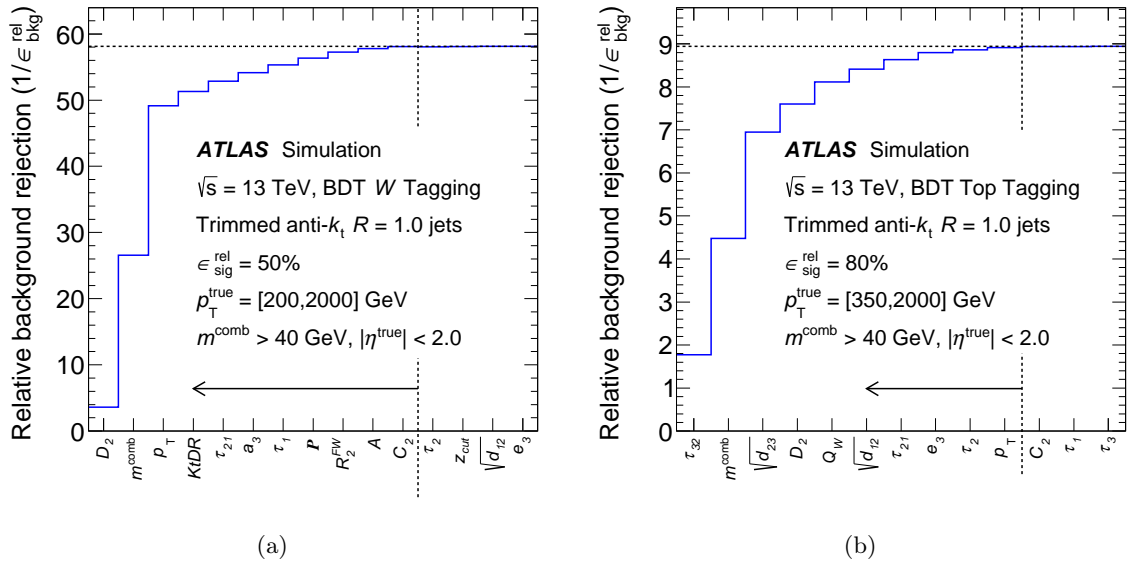
(a)   (b)

Figure 7.4: The relative background rejection of the BDT for different sets of variables, with more variables added successively for W-boson tagging at the 50% (a) and top-quark tagging at the 80% (b) relative signal efficiency working point. Only jets which satisfy the training criteria are considered to calculate the relative signal efficiency and relative background rejection. The performance is evaluated with uniform p_T^{true} spectra. Uncertainties are not presented. The vertical dashed lines and solid arrow represent the set of jet moments used in the final construction of the discriminant.

has fewer than 10 topoclusters, the remaining inputs to the DNN are assigned a default value of zero. Using the 10 highest-p_T topoclusters as inputs was found to provide optimal background rejection for this particular architecture. In order to understand this more qualitatively, the fraction of the jet p_T carried by each of the topoclusters is presented in Fig. 7.5, where the distribution of the p_T-fraction for a subset of the 10 highest-p_T clusters is shown along with the mean value of each of the 20 highest-p_T cluster distributions. It is observed that the 10 highest-p_T topoclusters on average carry more than 99% of the p_T of the jet. This indicates that the 10 highest-p_T topoclusters carry most of the relevant information. The preprocessed p_T, η and ϕ of the 10 highest-p_T topoclusters are used as input to a fully connected NN with four hidden layers composed of 300, 102, 12 and 6 nodes. The hyper-parameters were determined through manual hyper-parameter tuning and are exactly the same as the one used in [124]. The DNN is trained on jets where only the initial top parton is required to be matched to the reconstructed jet obtained from the Z' (signal) and light jets (background) in the high-p_T region from 450 GeV to 2400 GeV in p_T. To remove bias in the training due to the difference in p_T between the signal and the background samples, background sample is subsampled in jet p_T bins such that the p_T distribution is the same in both signal and background, different than the BDT and DNN taggers described previously, which use downsampling and jet-by-jet reweighting.
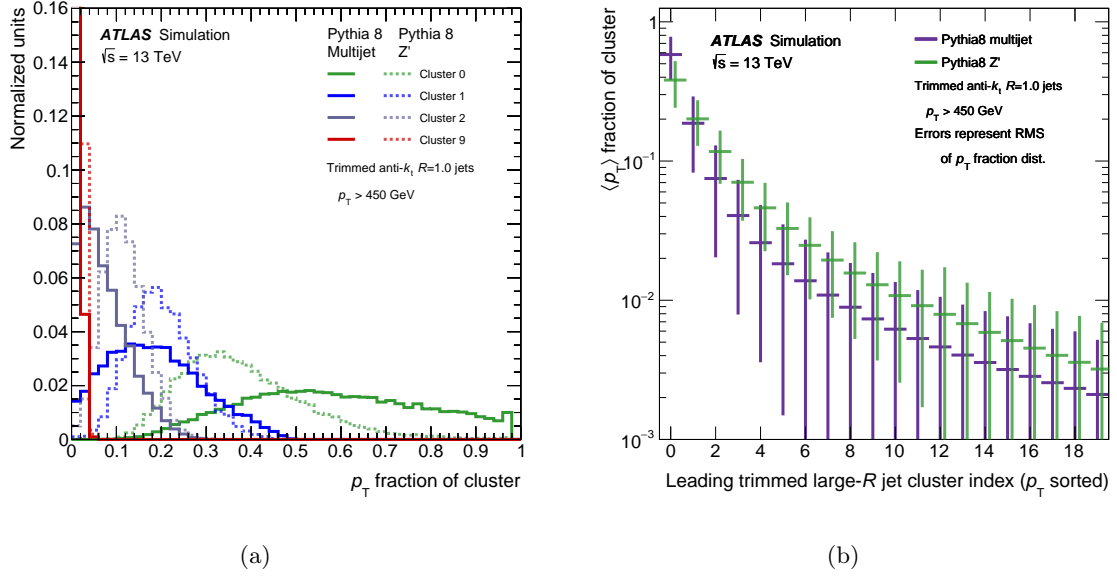
(a)   (b)

Figure 7.5: The distribution of the fraction of p_{T} carried by the highest-p_{T} cluster (Cluster 0) along with the next-highest (Cluster 1), third-highest (Cluster 2), and tenth-highest-p_{T} (Cluster 9) clusters (a) along with the average value of the ratio of the cluster p_{T} to the jet p_{T} for the 20 highest-p_{T} clusters (b). The dashed lines in (a) show distributions for signal jets, and the full lines show distributions for background jets. The vertical lines on each point in (b) represent the RMS of the corresponding distribution of the fraction of p_{T} of a given cluster in (a). In (a), the distribution for the tenth-highest-p_{T} cluster (Cluster 9) extends beyond the maximum value of the vertical axis.

7.4 Performance Comparison of Taggers

It is important to make a direct comparison of the performance of all of the tagging techniques presented in sections 6.5, 7.2 and 7.3. In this section the performance of the taggers is evaluated and compared using two different metrics: the background rejection as a function of $p_{\mathrm{T}}^{\mathrm{true}}$ at the relevant fixed signal efficiency working points and the background rejection as a function of the signal efficiency in the form of receiver operating characteristic (ROC) curves.

The performance of the obtained DNN, BDT and baseline two-variable cut taggers providing a signal efficiency of 50% for W-boson tagging and 80% for top-quark tagging are compared by evaluating the background rejection as a function of $p_{\mathrm{T}}^{\mathrm{true}}$. The performance comparison is presented in Fig. 7.6. It can be seen in this figure that in the case of W-boson tagging, the performance improvements beyond the cut-based taggers are approximately 20%. In the case of top-quark tagging, the improvements in performance are more significant, showing increases in background rejection of a factor of two over the entire kinematic range studied. For both the W and top-quark taggers it is observed that the background rejection at high p_{T} is lower compared to medium p_{T}, mostly due to the migration of the light-jet mass distribution to higher values and more frequent merging of topoclusters at high p_{T}.

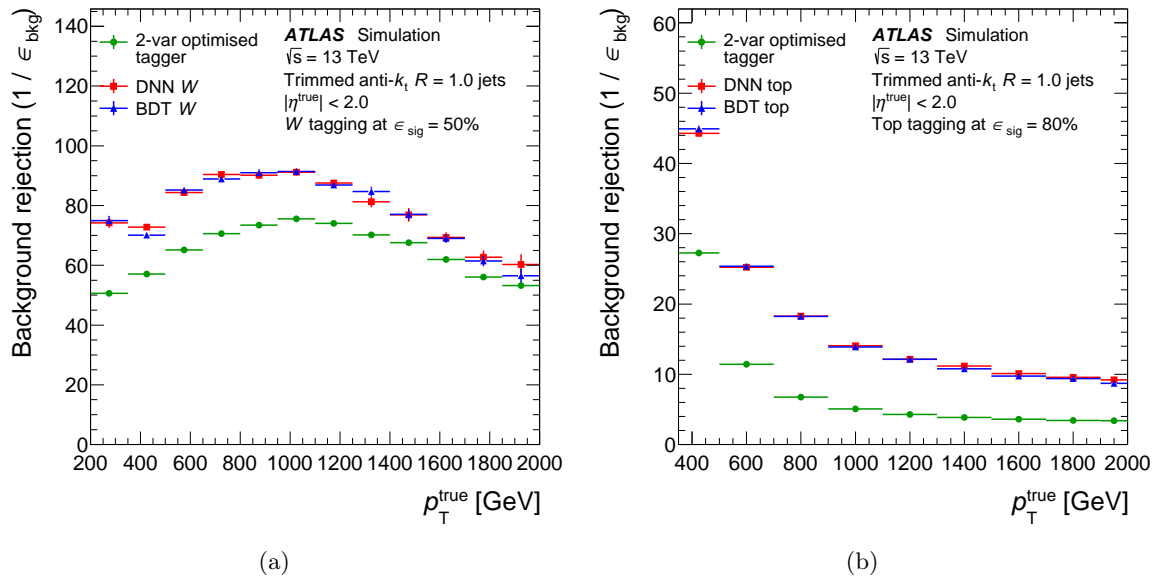(a)   (b)

Figure 7.6: The background rejection comparison of W-boson taggers at a fixed 50% signal efficiency working point (a) and top-quark taggers at a fixed 80% signal efficiency working point (b) for the multivariate jet-shape-based taggers as well as the two-variable optimised taggers, which are composed of a selection on m^{comb} and D_2 in the case of W-boson jet tagging and m^{comb} and τ_{32} for top-quark jet tagging. The performance is evaluated with the $p_{\mathrm{T}}^{\mathrm{true}}$ distribution of the signal jets weighted to match that of the multijet background samples. Statistical uncertainties of the background rejection are presented.

Second, in order to observe the performance of the taggers at different signal efficiencies, the performance of a variety of W-boson and top-quark taggers are compared in two wide jet $p_{\mathrm{T}}^{\mathrm{true}}$ bins in means of ROC curves. For W-boson-tagging, four taggers are compared as listed below.

- Simple $m^{\mathrm{comb}} + D_2$ tagger which is composed of a pre-defined fixed mass requirement of $60 < m^{\mathrm{comb}} < 120$ GeV and a varying D_2 requirement chosen to obtain the desired signal efficiency. This simple tagger is included for comparison purposes.

- The two-variable optimized 50% signal efficiency W-boson tagger which is composed of a varying m^{comb} and D_2 requirement. As this tagger is designed and optimized for the specific fixed signal efficiency working points, it is represented by a point in Figure 7.7 that gives the desired signal efficiency chosen for comparison.

- DNN W and BDT W taggers which are both composed of a requirement on the relevant discriminant and a fixed m^{comb} requirement of $m^{\mathrm{comb}} > 40$ GeV. Moreover, if a jet passes the $m^{\mathrm{comb}} > 40$ GeV criteria but fails the $N^{\mathrm{const}} > 2$ criteria, it is tagged as signal independent of the discriminant value. The requirement on the BDT or DNN discriminant is varied to obtain the desired signal efficiency.

For top-quark tagging, seven taggers are compared as listed below.

- Simple $m^{\mathrm{comb}} + \tau_{32}$ tagger which is composed of a fixed $m^{\mathrm{comb}} > 60$ GeV and a varying maximum τ_{32} cut to obtain the desired signal efficiency. This simple tagger is included for comparison purposes.

- The two-variable optimized 80% signal efficiency top-quark tagger which is composed of a varying m^{comb} and τ_{32} requirement. As this tagger is optimized for the specific fixed signal efficiency working points, it is represented by a point in Figure 7.8 that gives the desired signal efficiency chosen for comparison.

- DNN top and BDT top taggers which are both composed of a requirement on the relevant discriminant and a fixed m^{comb} requirement of $m^{\mathrm{comb}} > 40$ GeV. Moreover, if a jet passes the $m^{\mathrm{comb}} > 40$ GeV criteria but fails the $N^{\mathrm{const}} > 2$ criteria, it is tagged as signal independent of the discriminant value. The requirement on the BDT or DNN discriminant is varied to obtain the desired signal efficiency.

- TopoDNN tagger which is composed of a requirement on the relevant discriminant. The requirement on the discriminant is varied to obtain the desired efficiency. This tagger targets the identification of high-p_{T} inclusive top quarks, and therefore it is only included in the high p_{T} bin for comparison where the details of the different signal sample used for training are less relevant.

- The shower deconstruction tagger which is composed of a requirement on the $\log\chi$ variable described and a fixed m^{comb} requirement of $m^{\mathrm{comb}} > 60$ GeV. Similar to the BDT and DNN taggers, the requirement on $\log\chi$ is varied to obtain the desired signal efficiency.

- The HEPTopTagger which consists of identifying a top-quark candidate followed by a mass requirement of $140 < m^{\mathrm{comb}} < 210$ GeV on this candidate. This tagger is represented by a point in Figure 7.8 as it provides a working point in each p_{T} bin.

The comparisons in this section show that although two-variable combinations and advanced taggers such as shower deconstruction can provide good signal, background discrimination, the discrimination power can be improved by introducing ML-based techniques. The ML-techniques which use multiple jet substructure moments as input explore the richer and non-linear correlations between these observables. When the substructure moments are used as inputs, the DNNs and BDTs are observed to provide similar performance in terms of signal efficiency and background rejection. This is somewhat expected due to the relatively small number of inputs. Although the performance is similar, usage of DNNs have some other benefits such as setting up an infrastructure for the usage of modern machine learning libraries, shorter evaluation time.

Finally, in the case of top quark tagging, when the ROC curve of TopoDNN is compared to the other ML-based taggers at high p_{T}, it is observed that using topoclusters as input provides slightly better discrimination at high p_{T}. This behavior is expected since the frequency of merging of calorimeter energy depositions at high p_{T} increases and leads the substructure variables to lose their discrimination power significantly whereas the direct usage of topoclusters can extract
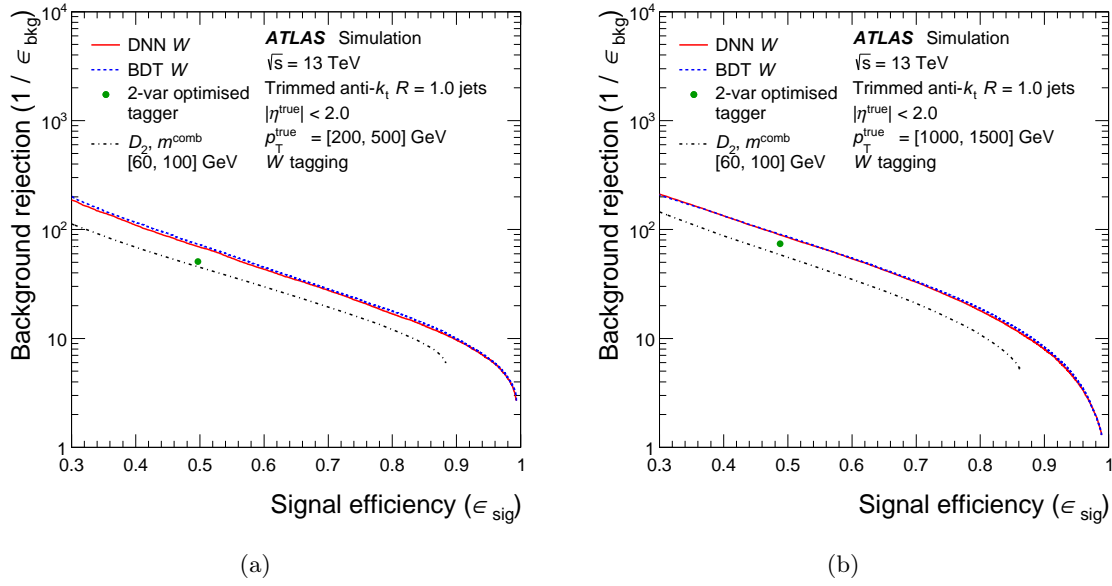
(a)

(b)

Figure 7.7: A performance comparison of the W-boson taggers in low-$p_{\mathrm{T}}^{\mathrm{true}}$ (a) and high-$p_{\mathrm{T}}^{\mathrm{true}}$ (b) bins. The performance is evaluated with the $p_{\mathrm{T}}^{\mathrm{true}}$ distribution of the signal jets weighted to match that of the dijet background samples.

(a)

(b)

Figure 7.8: A performance comparison of the top-quark taggers in low-$p_{\mathrm{T}}^{\mathrm{true}}$ (a) and high-$p_{\mathrm{T}}^{\mathrm{true}}$ (b) bins. The performance is evaluated with the $p_{\mathrm{T}}^{\mathrm{true}}$ distribution of the signal jets weighted to match that of the dijet background samples.
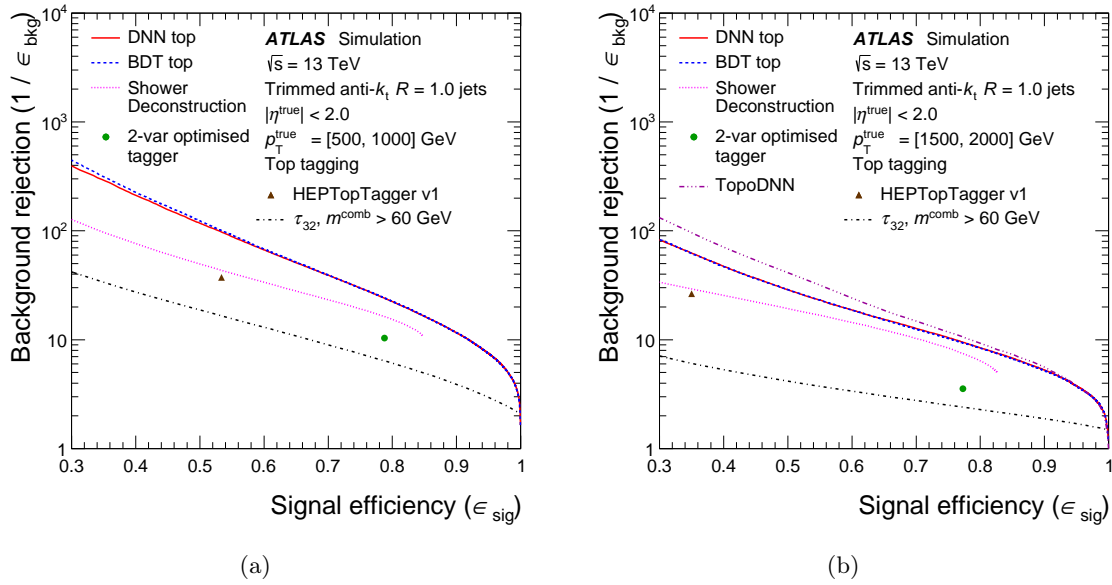
more information. This observation can be confirmed by comparing the performance of the DNN top tagger with TopoDNN tagger only for the jets which have all the substructure variables defined (for jets with more than 2 clusters, $N^{\mathrm{const}} > 2$). By this test it is observed that using the topoclusters as inputs where the substructure variables are not defined leads to significant performance improvements at high p_{T}.

7.5 Performance of Taggers in Data

In the previous sections, the taggers were studied purely in MC. In order to understand which techniques would be beneficial to use in physics analyses, it is essential to evaluate the performance of the tagging techniques also in collected data. In this section, the performance of the taggers are validated using signal and background-enriched data samples collected during 2015 and 2016, corresponding to an integrated luminosity of 36.1 fb^{-1} and 36.7 fb^{-1} as described in section 7.1. W-boson and top-quark jets are obtained from the lepton + jets samples, where the obtained signal jets cover a p_T range of approximately 200 GeV to 1000 GeV. Background jets are obtained from two samples: γ + jet and multijet samples. The γ + jet sample is enriched in light-quark jets and spans a jet p_T range of approximately 200 GeV to 2000 GeV. The multijet sample is enriched in light-quark and gluon jets and spans a jet p_T range approximately 500 GeV to 3500 GeV.

The modeling of the MC in data is validated by studying the jet observables and by measuring the signal efficiencies and background rejections in data. The signal efficiency and background rejection for fixed signal efficiency working points are measured as a function of the jet p_T and the average number of interactions per bunch crossing (μ) to test the robustness of the tagger against pileup. Although these studies are carried out for all the taggers and variables covered in the previous sections, an emphasis will be put on the results of the DNN W and DNN top taggers that I have developed. Results on the baseline two-variable cut-based W-boson, top-quark taggers as well as TopoDNN top-quark tagger are presented for the sake of comparison.

7.5.1 Signal efficiency in boosted $t\bar{t}$ events

A sample of data enriched in $t\bar{t}$ events where one top quark decays hadronically and the other semileptonically in both the electron and the muon decay channel is selected to study the modeling of signal W-boson and top-quark large-R jet tagging. The events are decomposed into two exclusive subsamples which are enriched in W-boson jets and top-quark jets. First, the selection and analysis methods are described and tagging discriminants are presented. This is followed by the signal efficiency measurement for a set of fixed signal efficiency working points.

Analysis and selection

To select the set of lepton-plus-jets $t\bar{t}$ events, the data and MC events are required to pass either an inclusive electron trigger or an inclusive muon trigger, where the thresholds are varied between the 2015 and 2016 datasets due to increases in instantaneous luminosity. In the electron channel, events from the 2015 data-taking period were required to pass at least one of three triggers: one isolated electron with $p_\text{T} > 24$ GeV, one electron with $p_\text{T} > 60$ GeV without any isolation requirement, or one electron with $p_\text{T} > 120$ GeV without any isolation requirement and looser identification criteria. In the 2016 data-taking period, the p_T thresholds of these electron triggers were increased to 26 GeV,60 GeV, and 140 GeV, respectively. In the muon channel, events from

the 2015 data-taking period are required to pass at least one of two muon triggers: one isolated muon with $p_\text{T} > 20$ GeV or one muon with $p_\text{T} > 50$ GeV without any isolation requirement. In the 2016 data-taking period, the p_T thresholds of these triggers were increased to 26 GeV and 50 GeV, respectively.

Events are then required to contain exactly one electron or muon candidate with $p_\text{T} > 30$ GeV that is matched to the trigger-level lepton associated with the appropriate trigger. Electrons are required to have $|\eta| < 2.47$, excluding the calorimeter transition region from $1.37 < |\eta| < 1.52$, and satisfy the "tight" likelihood-based identification criterion [126, 127]. Muons are required to have $|\eta| < 2.5$ and are required to satisfy the "medium" muon identification criteria [128]. For both electrons and muons, the reconstructed lepton candidate is required to be isolated from additional activity as described in section 5.4.

In addition to leptons, small-R jets reconstructed from topoclusters using the anti-k_t algorithm with $R = 0.4$ are used to reconstruct the E_T^miss and identify the signal topology. The small-R jets are required to have $p_\text{T} > 20$ GeV, $|\eta| < 2.5$. To suppress the low-p_T small-R jets produced in pileup interactions, small-R jets with $p_\text{T} < 60$ GeV and $|\eta| < 2.4$, jets are required to originate from the primary vertex by using the Jet Vertex Tagger (JVT) algorithm [129].

For the identification of b-quark jets, jets reconstructed from ID tracks using the anti-k_t algorithm with $R = 0.2$ are used. These jets are b-tagged using the 70% signal efficiency working point of the MV2c10 algorithm [130].

The missing transverse momentum is reconstructed as descried in 5.6. In this case, the negative vector sum is calculated using the single lepton and the full set of small-R calorimeter jets as well as ID tracks not associated with the lepton or jets.

In order to select the lepton + jets $t\bar{t}$ events, events which contain a leptonically decaying W boson are pre-selected by requiring one electron or muon candidate with $p_\text{T} > 30$ GeV and vetoing events which contain additional electrons or muons with $p_\text{T} > 25$ GeV. The E_T^miss is required to be greater than 20 GeV and the sum of E_T^miss and the transverse mass of the leptonically-decaying W-boson candidate(m_T^W) are required to be greater than 60 GeV, where $m_\text{T}^W = \sqrt{2p_\text{T}^\ell E_\text{T}^\text{miss}(1 - \cos\Delta\phi)}$ is calculated from the p_T of the lepton (p_T^ℓ) and $\Delta\phi$ is the azimuthal angle between the lepton momentum and the E_T^miss direction. At least one small-R jet is required to have $p_\text{T} > 25$ GeV and to be close to the lepton by requiring $\Delta R(\text{lepton}, \text{jet}) < 1.5$.

The highest-p_T trimmed anti-k_t $R = 1.0$ jet, referred to as leading jet, is studied. The signal top-quark jet candidate is required to be separated from the semileptonic top-quark decay. This separation is achieved by requiring $\Delta R > 1.5$ between the large-R jet and the small-R jet close to the lepton and by requiring the angular separation in the transverse plane between the lepton and the large-R jet, $\Delta\phi$, to be greater than 2.3.

Additional selections are applied on the pre-selected events to decompose these events into two exclusive subsamples which are enriched in fully contained hadronically-decaying top-quark jets or fully contained hadronically-decaying W-boson jets. The sample enriched in top-quark jets, referred to as "top-quark selection", is defined by requiring a jet p_T greater than 350 GeV and a

b-tagged track jet to be in proximity to the large-R jet. The latter is achieved by a requirement on the angular separation from the b-jet as $\Delta R(b\text{-jet}, \text{large-}R \text{ jet}) < 1.0$. The sample enriched in W-boson jets,referred to as "W-boson selection", is defined by requiring a jet p_T greater than 200 GeV and a b-tagged track jet to be away from the large-R jet. The latter is achieved by a requirement on the angular separation from the b-jet as $\Delta R(b\text{-jet}, \text{ large-}R \text{ jet}) > 1.0$. Since the geometrical separation of the top parton and its b-quark decay product decreases with increasing p_T, this geometrical separation requirement limits the efficiency of the W-boson selection at high p_T and results in limiting the kinematic reach to approximately 600 GeV.

The $t\bar{t}$ and single-top MC samples are divided into three subsamples based on the jet labelling described in section 6.3. The fraction of events in each sample of interest are decomposed into: "$t\bar{t}$ (top)", "$t\bar{t}$ (W)"), and "$t\bar{t}$ (other)", which includes all other events in these samples. The backgrounds are derived from the MC samples, with the exception of the multijet background. The multijet background is estimated by a data-driven method based on looser lepton selection criteria. The statistical uncertainty of the background prediction, referred to as "Stat. uncert." in the figures, results from limited MC sample size as well as the limited size of the data sample used in the data-driven estimation of the multijet background. The total event yield in the MC is normalized to that in the data at this stage of the selection.

The large-R jet mass is presented for the W-boson selection and top-quark selection in Fig. 7.9 including the full set of systematic uncertainties summarized in section 7.5.3. As it can be seen in this figure, these selections result in relatively pure samples of W-boson and top-quark jets. The disagreement between the peak positions in MC and data observed near the W-boson mass and top-quark mass is attributed to a mismodeling of the jet mass scale as studied previously in Ref. [131].

The DNN W and DNN top discriminants are presented in Fig. 7.10. Different than the m^{comb} distributions, for these discriminants, no dedicated experimental systematic uncertainty in the scale or resolution of the discriminant itself is included. Instead, a similar approach that is taken in the context of the identification of $b-$ and $c-$ jets [132] is followed. The mismodelling of the simulation relative to data is taken into account as a derived uncertainty in the *in situ* measurement of the signal efficiency of the tagger itself. It is observed that for nearly all regions of phase space, the overall relative yield of data is well-described by the MC prediction within the theoretical uncertainties which are derived from the comparison of various $t\bar{t}$ MC generators.
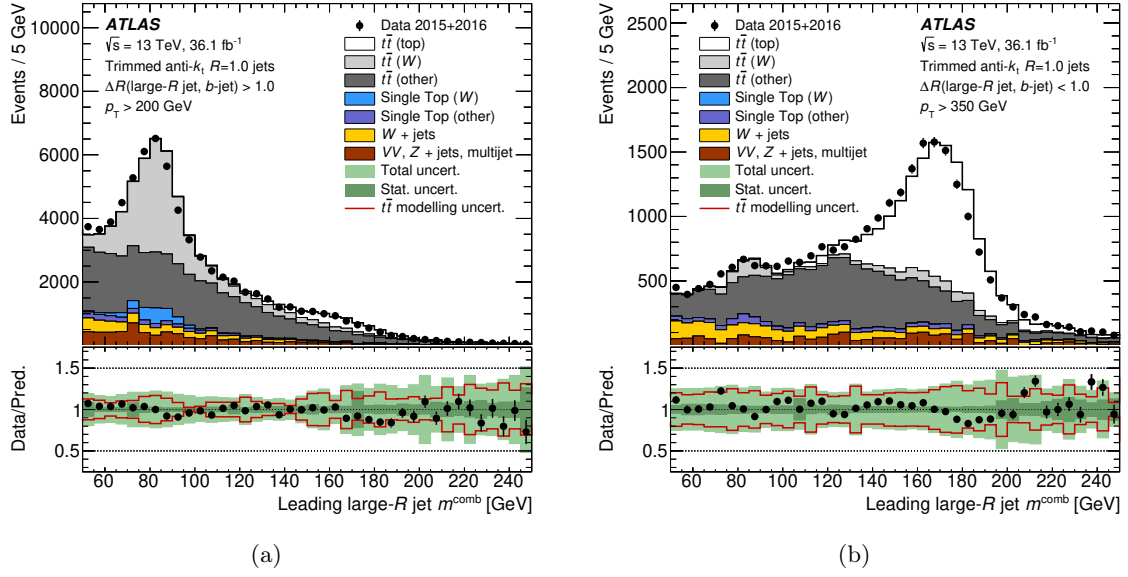
Figure 7.9: A comparison of the observed data and predicted MC distributions of the mass of the leading large-R jet in the event for the W boson (a) and top quark (b) selections in a sample enriched in lepton+jets $t\bar{t}$ events. Simulated distributions are normalized to data.
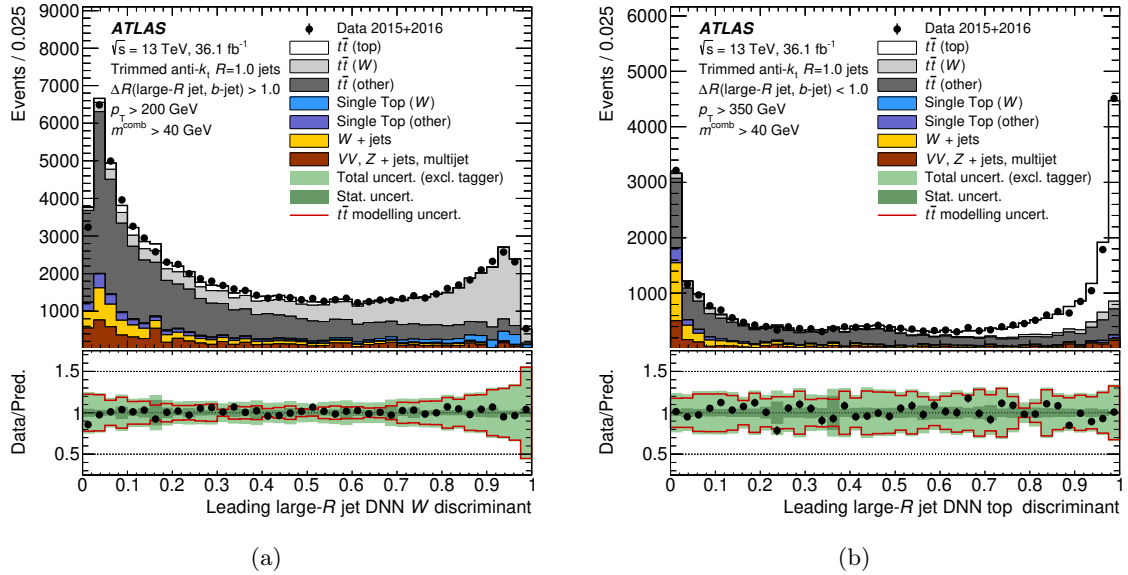

Figure 7.10: A comparison of the observed data and predicted MC distributions of the leading large-R jet DNN W (a) and DNN top (b) discriminants for the respective event selections in a sample enriched in lepton+jets $t\bar{t}$ events. Simulated distributions are normalized to data.

Signal efficiencies

In this section, the signal efficiency in data is studied. The signal efficiency measurement in data is compared with the MC efficiency prediction to test the agreement between data and to estimate the relevant systematic uncertainty. The efficiency measurement for a subset of taggers in signal-enriched data is compared to the MC prediction. The five tagger working points for which the signal efficiency is measured are:

- $D_2 + m^{\mathrm{comb}}$ (*W* boson): The two-variable optimized 50% signal efficiency fully-contained *W*-boson tagger as defined in section 6.5.1;

- DNN (*W* boson): The DNN-based 50% signal efficiency fully-contained *W*-boson tagger as defined in section 7.2.3;

- $m^{\mathrm{comb}} + \tau_{32}$ (top quark): The two-variable optimized 80% signal efficiency fully-contained top-quark tagger as defined in section 6.5.1;

- DNN (top quark): The DNN-based 80% signal efficiency fully-contained top-quark tagger as defined in section 7.2.3;

- TopoDNN (top quark): The TopoDNN tagger which applies a selection on the DNN discriminant to give a fixed 80% signal efficiency as a function of p_{T} for fully contained top-quark jets in MC.

The numbers of signal-like events in data that pass and fail each of these tagging requirements are obtained from a chi-square template fit of "signal" and "background" distributions predicted by MC to the data to correct for mismodeling of the cross-section of the various processes contributing to the phase space of interest. The labelling of "signal" events follows section 6.3, 7.5.1 and is based on simulations of $t\bar{t}$ and single-top-quark events. Background templates whose shapes are similar are merged to increase the stability of the fit.

In the case of *W*-boson tagging, this procedure results in a signal $t\bar{t}(W)$ and single top(W) and background $t\bar{t}(\mathrm{top})+t\bar{t}(\mathrm{other})+$single top(other)$+$non-$t\bar{t}$ component template. In the case of top-quark tagging, it results in a signal $t\bar{t}(\mathrm{top})$ and two background $t\bar{t}(W)+t\bar{t}(\mathrm{other})$ and non-$t\bar{t}$ component templates. The normalization of each template is allowed to float freely in the fit. The fit is performed using m^{comb} distributions of the leading anti-k_{t} trimmed jet.Distributions of events that either pass or fail the tagger of interest are fit simultaneously. The total normalization of each grouped background component is allowed to float and is extracted in the fit, while the efficiency of the tagger on background events is fixed to the value in Monte Carlo simulation. Normalizations of signal distributions in the pass and fail categories ($N_{\mathrm{fitted\ signal}}^{\mathrm{tagged}}$ and $N_{\mathrm{fitted\ signal}}^{\mathrm{not\ tagged}}$) are extracted from the fit. The tagger efficiency for signal events in data (ϵ_{data}) is extracted as

$$\epsilon_{\mathrm{data}} = \frac{N_{\mathrm{fitted\ signal}}^{\mathrm{tagged}}}{N_{\mathrm{fitted\ signal}}^{\mathrm{tagged}} + N_{\mathrm{fitted\ signal}}^{\mathrm{not\ tagged}}}.$$

This is compared to the tagger efficiency in MC (ϵ_{MC}), which is based on the numbers of predicted signal events that are tagged as signal ($N_{\mathrm{signal}}^{\mathrm{tagged}}$), and are not tagged as signal ($N_{\mathrm{signal}}^{\mathrm{not\ tagged}}$) by the tagger under study. The tagger efficiency in MC is calculated as

$$\epsilon_{\mathrm{MC}} = \frac{N_{\mathrm{signal}}^{\mathrm{tagged}}}{N_{\mathrm{signal}}^{\mathrm{tagged}} + N_{\mathrm{signal}}^{\mathrm{not\ tagged}}}.$$

The signal efficiency is measured in data and in simulations as a function of the p_{T} of the large-R jet and average number of interactions per bunch crossing (μ). It is important to note that when studying the signal efficiency as a function of μ, the systematic uncertainties are correlated between bins. The results for the W-boson taggers are presented in Figs. 7.11 and 7.12, for the top-quark taggers in Figs. 7.13, 7.14, and 7.15. The signal efficiency for the W-boson and top-quarks taggers in MC agree with the measured efficiency in data within uncertainties. In the case of the W-boson taggers, there is a difference between the design 50% signal efficiency and that measured in $t\bar{t}$ events due to event topology differences between W-boson jets in the samples considered in this section and the MC samples used to design the taggers. The two-variable cut-based taggers and DNN-based taggers show similar behavior in terms of the measured signal efficiency as a function of p_{T} and its related uncertainty. The total uncertainty of the measured signal efficiency is generally about 50% for the W-boson tagger efficiencies and 15% for the top-quark tagger efficiencies. Overall, the total uncertainty is dominated by systematic uncertainties, described in section 7.5.3. The systematic uncertainties are dominated by theoretical uncertainties in $t\bar{t}$ modeling, mainly due to the subtraction of the non-W-boson jets or non-contained top-quark jets.

Signal efficiency of all the taggers are observed to be relatively robust against pileup and the tagger's behavior as a function of μ in data is agrees with the behavior in MC within uncertainties. In the case of the W taggers, the signal efficiency degrades slightly at high μ. Whereas in the case of the top-quark tagging, the signal efficiency is approximately constant as a function of μ.
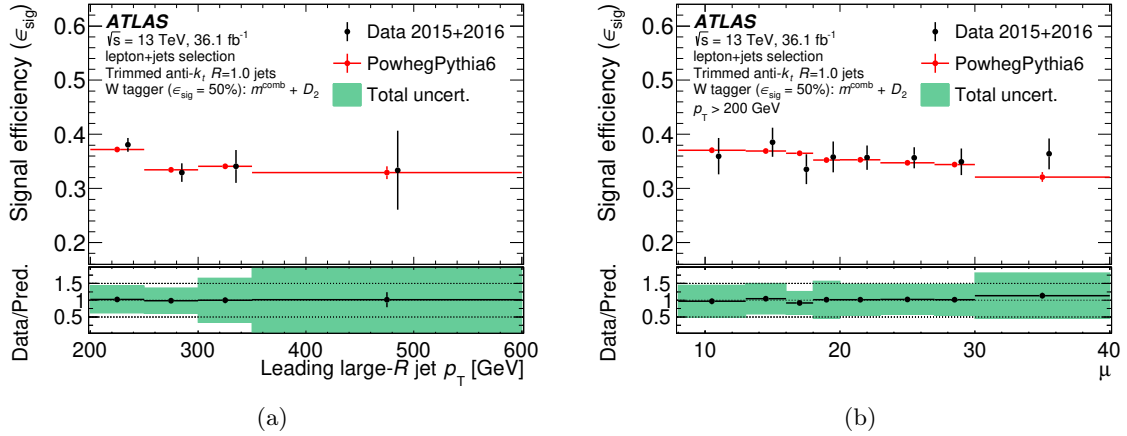
(a)   (b)

Figure 7.11: The signal efficiency of contained W-boson jets for the two-variable $m^{\text{comb}} + D_2$ W-boson tagger as a function of the large-R jet p_{T} (a) and the average number of interactions per bunch crossing μ (b) in data and simulation. The statistical uncertainties of the signal efficiency measurement in data and simulation are shown as error bars in the top panel. In the bottom panel, the ratio of the measured signal efficiency in data to that estimated in Monte Carlo simulation is shown with statistical uncertainties as error bars on the data points and the sum in quadrature of statistical and systematic uncertainties as a shaded band. When considering experimental uncertainties arising from the large-R jet, only those coming from the jet energy scale and resolution are considered.

(a)   (b)

Figure 7.12: The signal efficiency of contained W-boson jets for the jet shape-based DNN W-boson tagger as a function of the large-R jet p_{T} (a) and the average number of interactions per bunch crossing μ (b) in data and simulation. The statistical uncertainties of the signal efficiency measurement in data and simulation are shown as error bars in the top panel. In the bottom panel, the ratio of the measured signal efficiency in data to that estimated in Monte Carlo is shown with statistical uncertainties as error bars on the data points and the sum in quadrature of statistical and systematic uncertainties as a shaded band. When considering experimental uncertainties arising from the large-R jet, only those coming from the jet energy scale and resolution are considered.
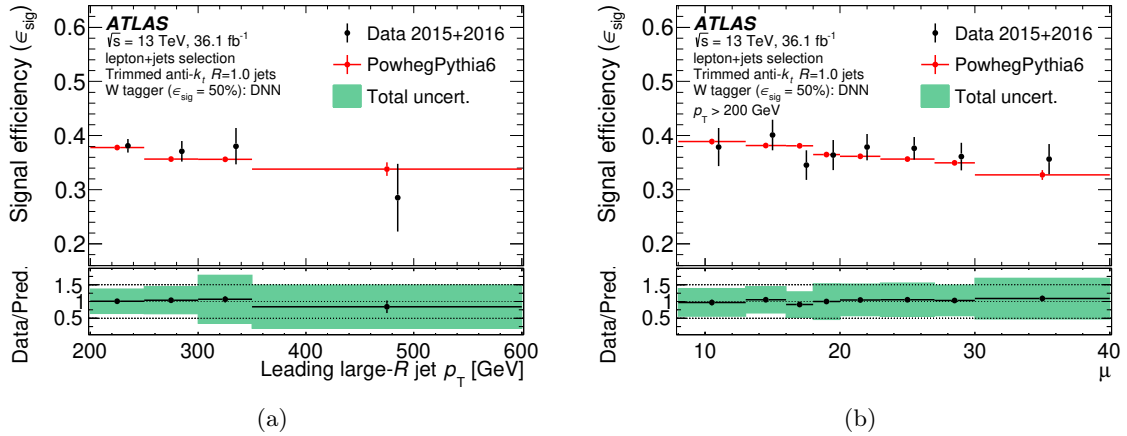
Figure 7.13: The signal efficiency of contained top-quark jets for the two-variable $m^{\mathrm{comb}} + \tau_{32}$ top-quark tagger as a function of the large-R jet p_{T} (a) and the average number of interactions per bunch crossing μ (b) in data and simulation. The statistical uncertainties of the signal efficiency measurement in data and simulation are shown as error bars in the top panel. In the bottom panel, the ratio of the measured signal efficiency in data to that estimated in Monte Carlo is shown with statistical uncertainties as error bars on the data points and the sum in quadrature of statistical and systematic uncertainties as a shaded band. When considering experimental uncertainties arising from the large-R jet, only those coming from the jet energy scale and resolution are considered.
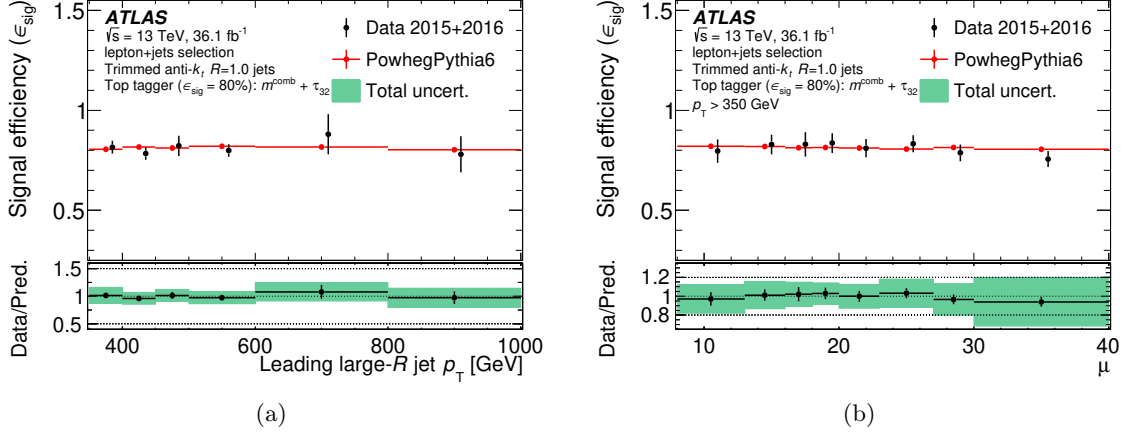

Figure 7.14: The signal efficiency of contained top-quark jets for the jet shape-based DNN top-quark tagger as a function of the large-R jet p_{T} (a) and the average number of interactions per bunch crossing μ (b) in data and simulation. The statistical uncertainties of the signal efficiency measurement in data and simulation are shown as error bars in the top panel. In the bottom panel, the ratio of the measured signal efficiency in data to that estimated in Monte Carlo is shown with statistical uncertainties as error bars on the data points and the sum in quadrature of statistical and systematic uncertainties as a shaded band. When considering experimental uncertainties arising from the large-R jet, only those coming from the jet energy scale and resolution are considered.
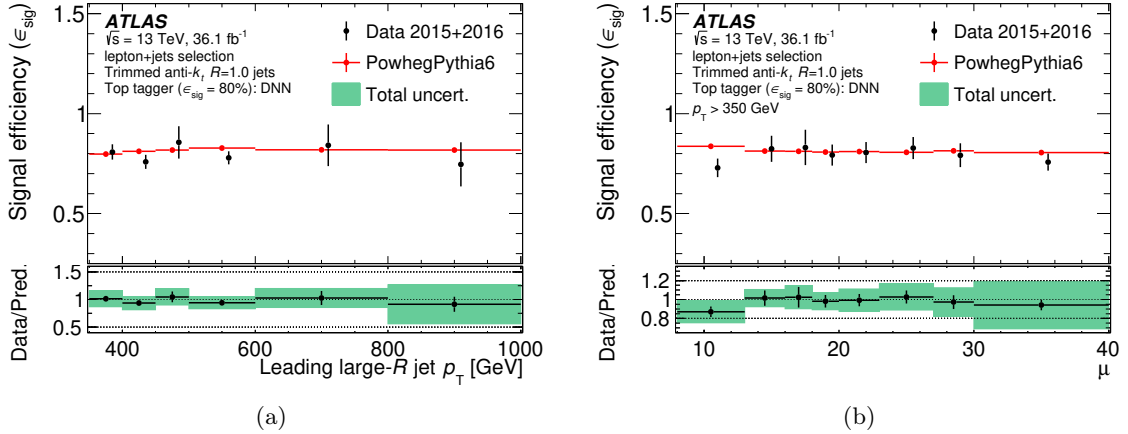
Figure 7.15: The signal efficiency of contained top-quark jets for the TopoDNN top-quark tagger as a function of the large-R jet p_T (a) and the average number of interactions per bunch crossing μ (b) in data and simulation. The statistical uncertainties of the signal efficiency measurement in data and simulation are shown as error bars in the top panel. In the bottom panel, the ratio of the measured signal efficiency in data to that estimated in Monte Carlo is shown with statistical uncertainties as error bars on the data points and the sum in quadrature of statistical and systematic uncertainties as a shaded band. When considering experimental uncertainties arising from the large-R jet, only those coming from the jet energy scale and resolution are considered.
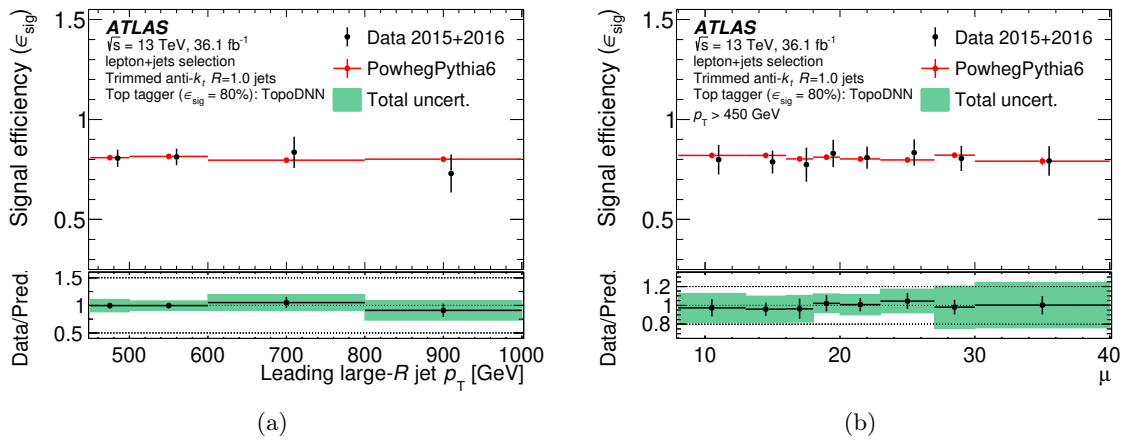
7.5.2 Background rejection in multijet and $\gamma + $ jet events

In this section, the background jets are studied in two sets of events: multijet and $\gamma + $ jet events. Similar content presented for the study of signal W-boson and top-quark jets is presented. First, the selection and analysis methods are described and tagging discriminants are presented in data and MC. This is followed by the background rejection measurement for the same set of fixed signal efficiency working points as in the previous section.

Analysis and selection

In the case of the multijet sample, events are selected in both data and MC using a single anti-k_t trimmed jet with $R = 1.0$ trigger where online transverse energy is required to be greater than 360 GeV during 2015 data taking and to be greater than 420 GeV in 2016 data taking. Events are then required to have at least one large-R anti-k_t trimmed jet with $R = 1.0$ which has p_T greater than 450 GeV. For the events which pass this pre-selection, the modeling of the highest-p_T large-R jet in the event is investigated with respect to two generators, PYTHIA and HERWIG++.

In the case of the $\gamma + $ jet sample, events are selected in both data and MC using a single photon trigger. The single photon trigger requires that photons satisfy "loose" quality criteria and require E_T to be greater than 120 GeV during 2015 data taking and to be greater than 140 GeV during 2016 data taking. Photon candidates are required to be within $|\eta| < 2.5$ and pass the "tight" identification and isolation criteria [133]. Large-R jets are required to have $p_\mathrm{T} > 200$ GeV, $|\eta| < 2.0$ and to be separated from the reconstructed photon with $\Delta\phi(\mathrm{jet}, \gamma) > \frac{\pi}{2}$. Finally, events with at least one photon which has $E_\mathrm{T} > 155$ GeV are selected.

The normalization of the simulated multijet and $\gamma + $ jet predictions is obtained from data after the above selections, taking into account the small signal contamination from the hadronically-decaying W-boson, Z-boson and $t\bar{t}$ events. First, the predicted signal contamination is subtracted from data. Then, the total event yield in the MC is normalized to that in the background-subtracted data.

The distributions of the leading anti-k_t $R = 1.0$ jet mass in the inclusive multijet and $\gamma + $ jet events are presented in Fig. 7.16. The DNN tagging discriminants are presented in Fig. 7.17 for multijet and $\gamma + $ jet events. Systematic uncertainties include all experimental uncertainties related to the selection of events, as well as the reconstruction and calibration of the large-R jet. A good agreement between data and MC is observed within uncertainties where the uncertainties are dominated by MC modeling.
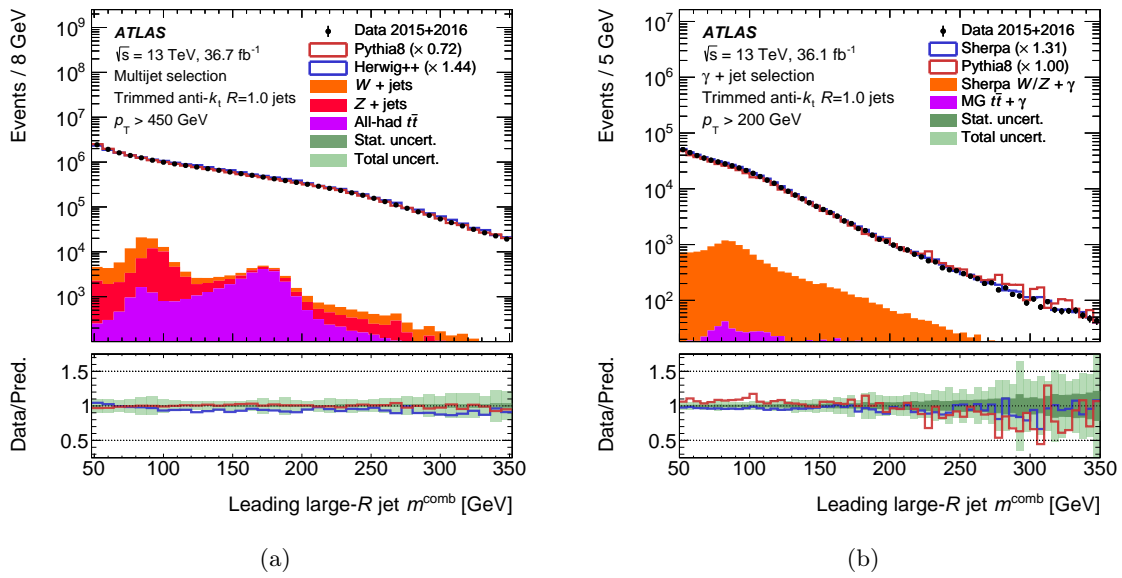
Figure 7.16: A comparison of the observed data and predicted MC distributions of the mass of the leading p_T anti-k_t trimmed jet in events for the multijet (a) and γ + jet (b) selections. The relevant data-driven normalization factor is shown in the legend for each MC generator. Systematic uncertainties are indicated as a band in the lower panel.

Figure 7.17: A comparison of the observed data and MC predictions in the multijet and $\gamma + \mathrm{jet}$ event samples for the anti-k_{t} $R = 1.0$ trimmed jet spectra of the W-boson (a)(c) and top-quark (b)(d) DNN discriminants. The relevant data-driven normalization factor is shown in the legend for each MC generator. Systematic uncertainties are indicated as a band in the lower panel.
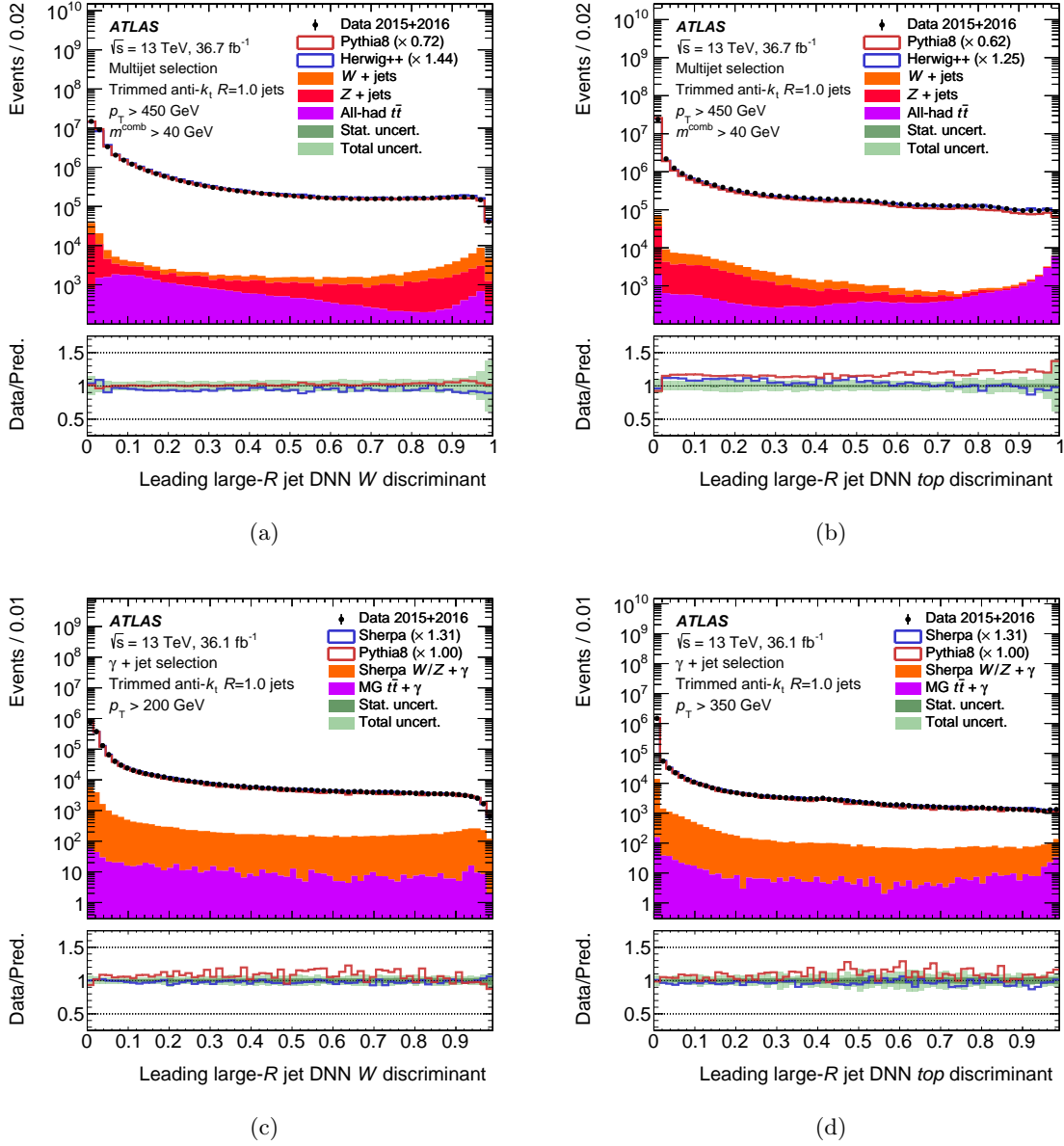
Background rejections

In this section, the background rejection in data is measured. The background rejection measurement in data is compared with the one predicted by the MC samples to test the agreement between the two.

A simple approach is taken where the background rejection is calculated directly from fraction of events. The fraction of events are obtained by subtracting the signal contamination from data and applying the relevant normalization to the multijet and γ + jet samples. This simple approach is possible thanks to the purity of the samples. The background rejections ($1/\epsilon_{\mathrm{bkg}}$) are presented for the four tagger working points listed in 7.5.1. The results for the W-boson taggers are presented in Figs. 7.18 and 7.19, for the top-quark taggers in Figs. 7.20– 7.22.

The background rejection for the W-boson and top-quarks taggers in MC generators shown an overall good agreement with the measured rejection in data. The background rejection as a function of p_{T} is observed to follow the same expected trend for all taggers. The background rejection as a function of μ for W-boson taggers are observed to have an increasing background rejection for higher μ for both background topologies. For top-quark taggers, background rejection as a function of μ follow no clear trend. It is important to note that different than the W-boson taggers, the top-quark taggers have significantly lower background rejection at high p_{T}. Consequently, more background jets are tagged at high p_{T} and enter the background rejection calculation as a function of μ. In some kinematic regions large differences between generators are observed for both topologies.

In the case of W-boson tagging, for the multijet topology, the background rejection in data is predicted well by PYTHIA, while the HERWIG++ prediction is lower than the rejection in data. For the γ + jet topology, large differences are observed in the predictions by the two generators especially at high p_{T}. It is observed that the generator differences are larger for the DNN W tagger, shown in Fig. 7.19, compared to the simpler $D_2 + m^{\mathrm{comb}}$ tagger, shown in Fig. 7.18. Overall, the background rejection of W-boson taggers in γ + jet data is predicted well by SHERPA.

In the case of top-quark tagging, for the multijet topology, the background rejection in data is predicted well by PYTHIA, while the HERWIG++ prediction is significantly lower than the rejection in data. For the γ + jet topology, the background rejection of top-quark taggers show significantly different agreement for the generators. SHERPA background rejection prediction agrees more with the background rejection in data for $m^{\mathrm{comb}} + \tau_{32}$ and DNN top taggers. Whereas PYTHIA background rejection prediction agrees more with the background rejection in data for the TopoDNN tagger. When the differences between generators is compared for different taggers, the DNN top tagger, shown in Fig. 7.21 has larger differences than the simpler cut-based tagger, shown in Fig. 7.20. Similarly, the TopoDNN tagger, shown in Fig. 7.22, has larger differences between generators than the DNN top tagger.

Figure 7.18: The background rejection as a function of the jet p_T and the average number of interactions per bunch crossing μ for the two-variable W-boson tagger in the multijet (a) (c) and γ + jet (b) (d) selection.

Figure 7.19: The background rejection as a function of the jet p_{T} and the average number of interactions per bunch crossing μ for the DNN W-boson tagger in the multijet (a) (c) and γ + jet (b) (d) selection.

Figure 7.20: The background rejection as a function of the jet p_T and the average number of interactions per bunch crossing μ for the two-variable top-quark tagger in the multijet (a) (c) and $\gamma + $ jet (b) (d) selection.
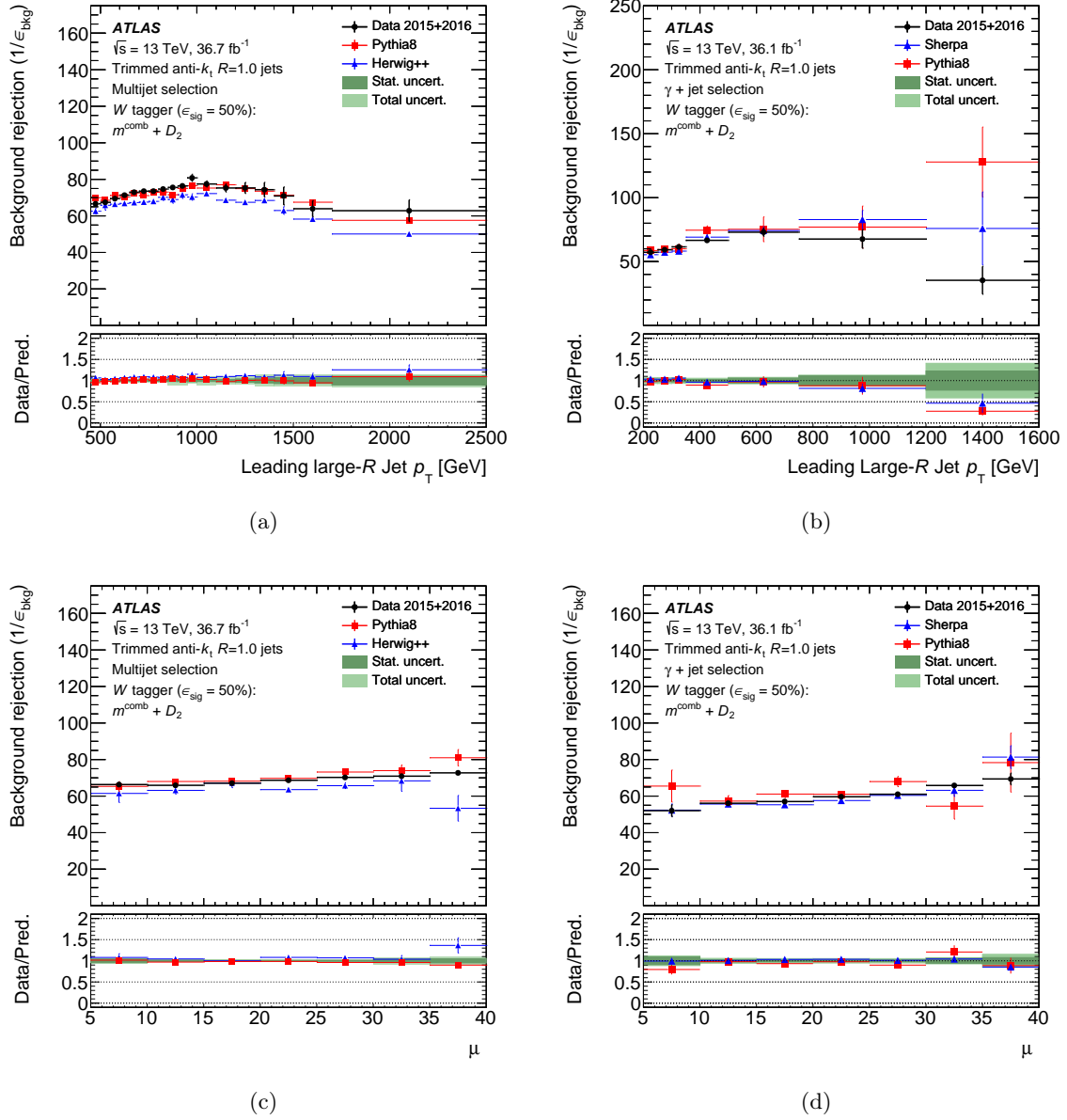
Figure 7.21:   The background rejection as a function of the jet p_T and the average number of interactions per bunch crossing μ for the DNN top-quark tagger in the multijet (a) (c) and $\gamma +$ jet (b) (d) selection.
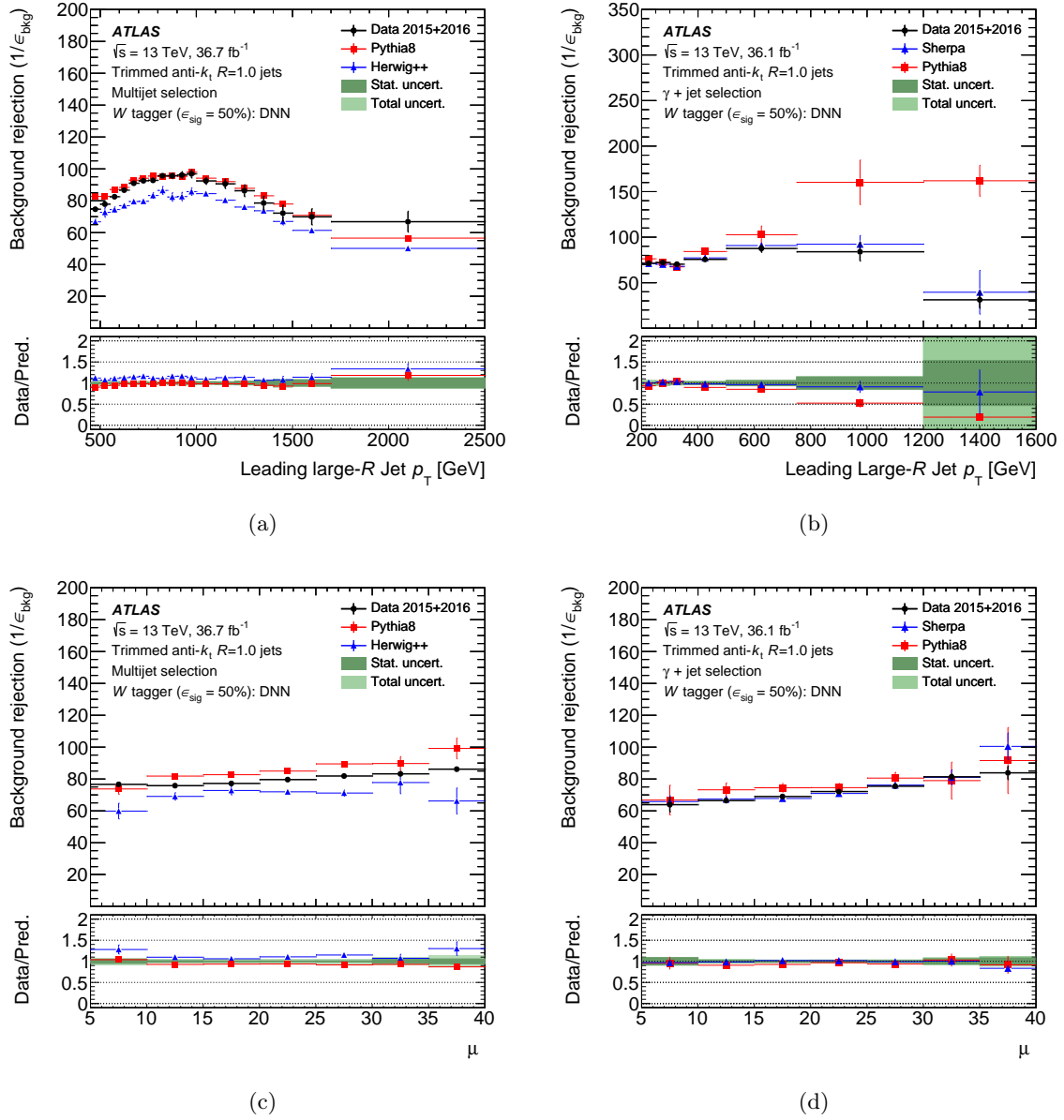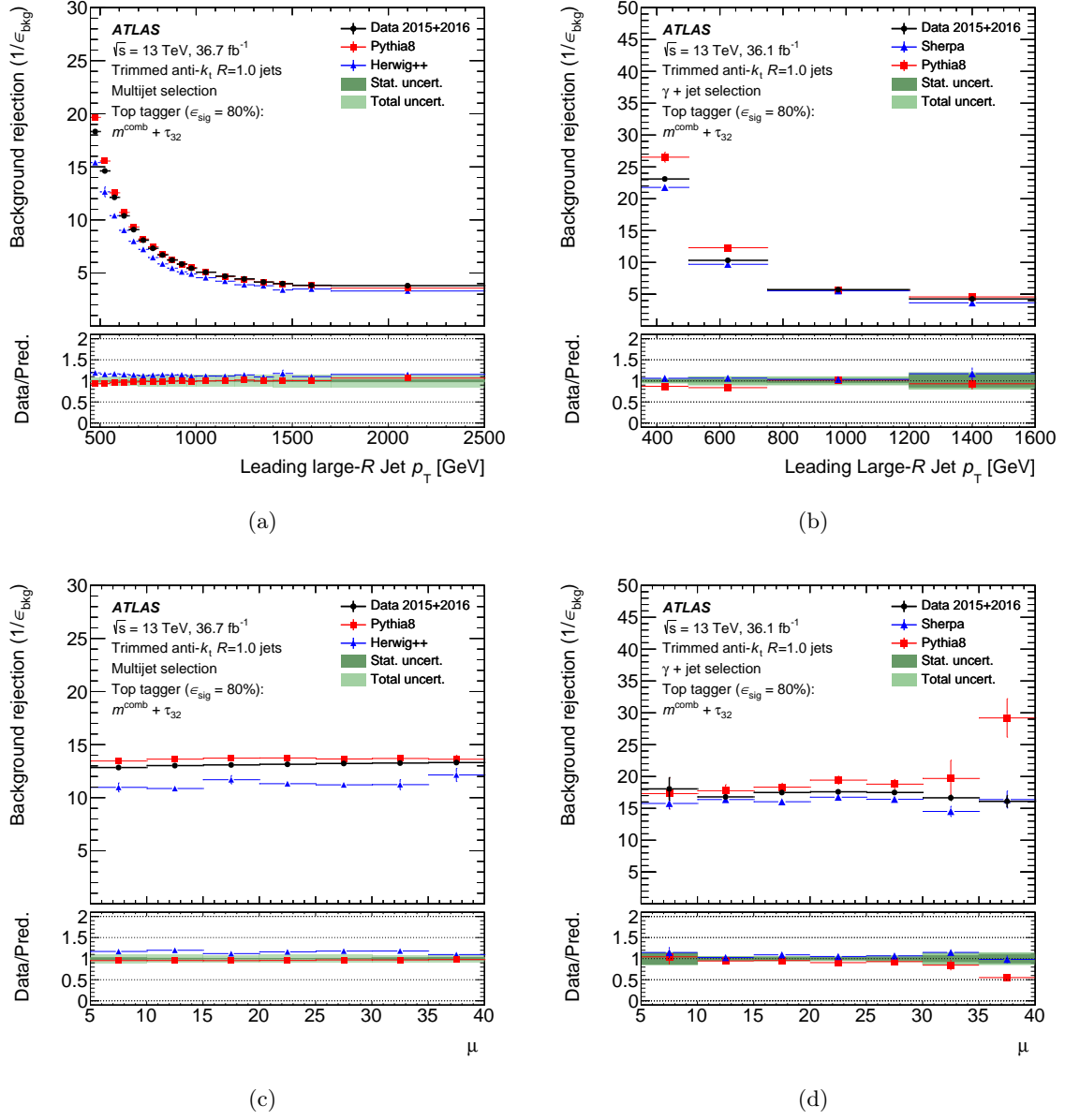
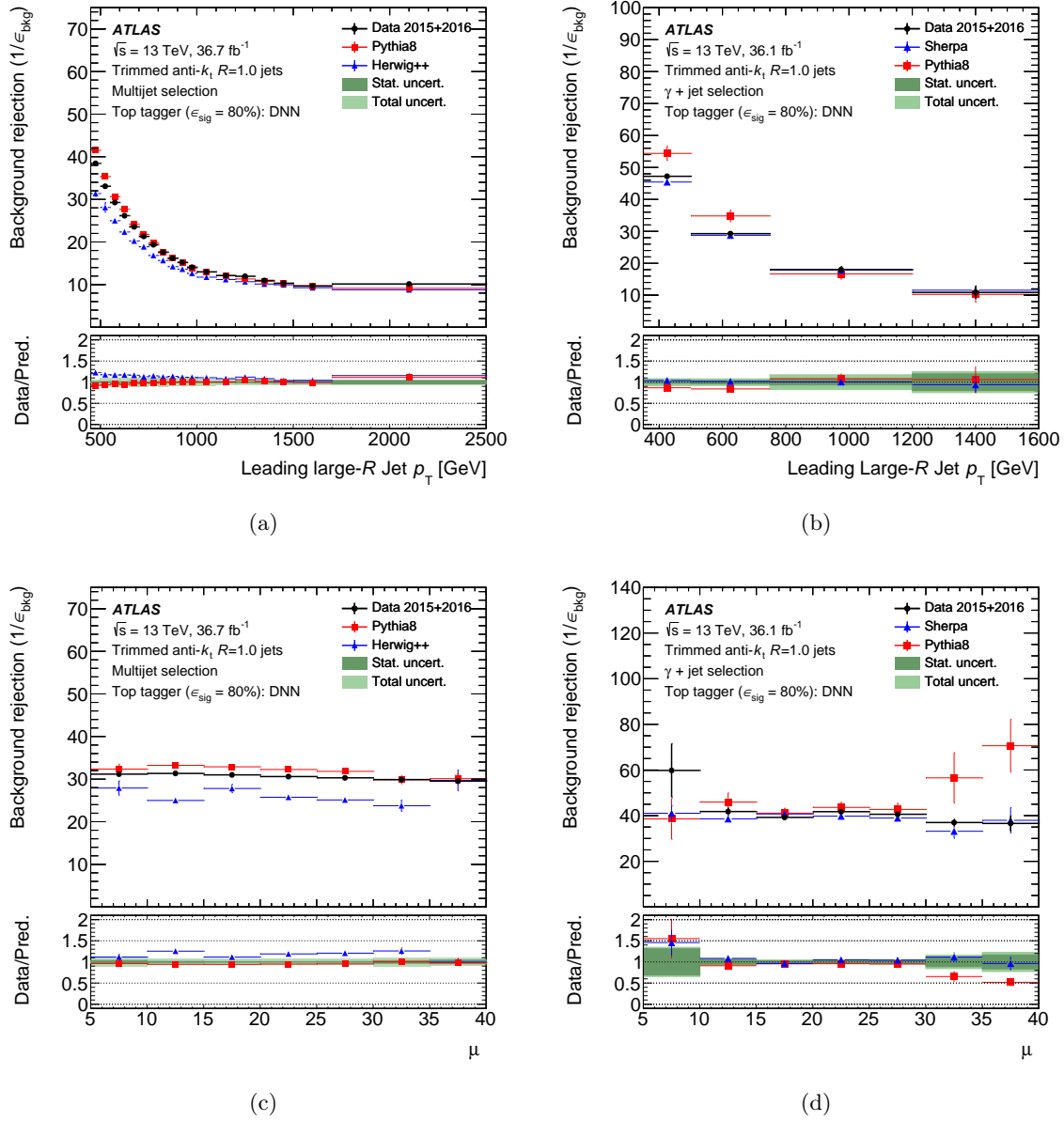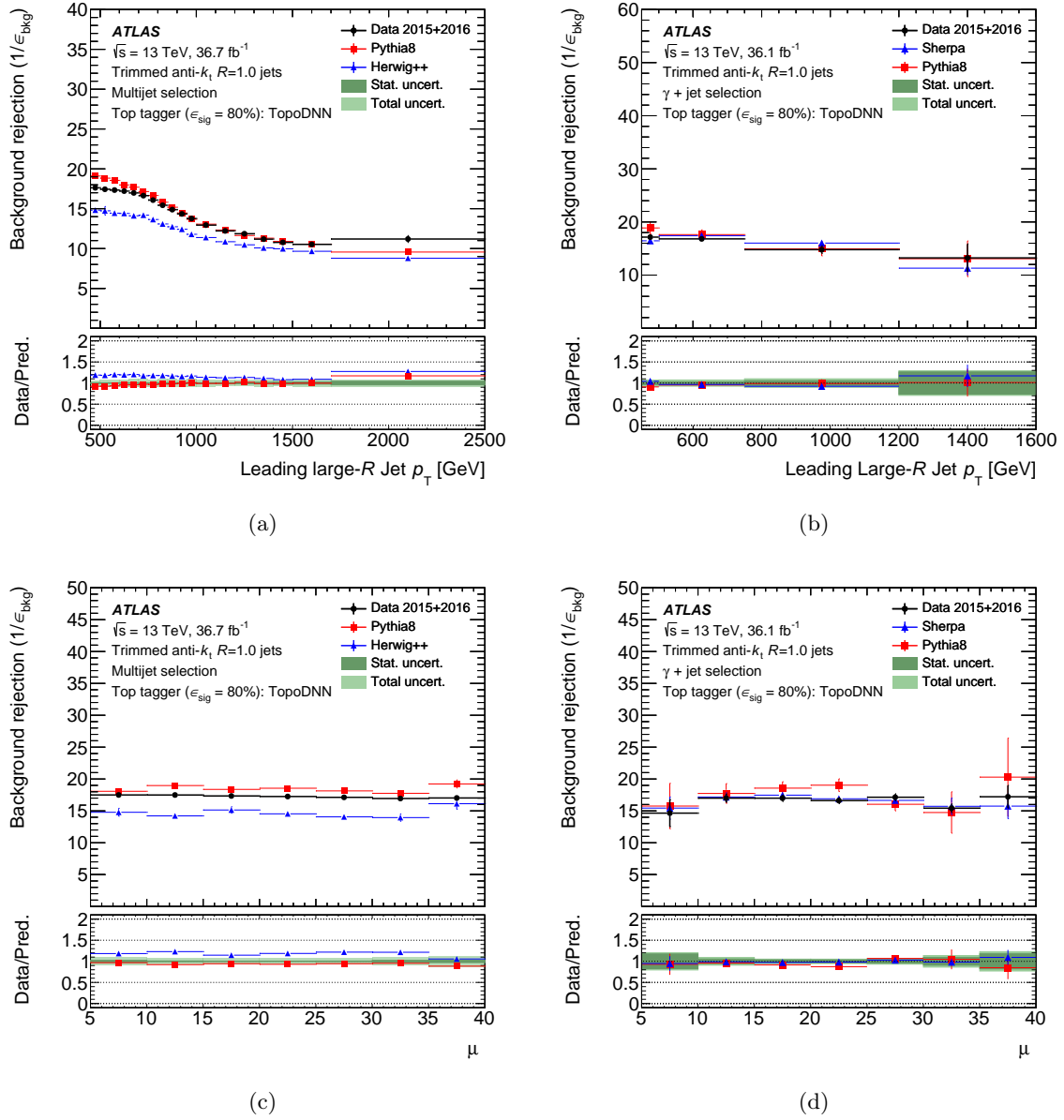Figure 7.22: The background rejection as a function of the jet p_{T} and the average number of interactions per bunch crossing μ for the TopoDNN top-quark tagger in the multijet (a) (c) and $\gamma + \mathrm{jet}$ (b) (d) selection.

7.5.3 Systematic uncertainties

Different sources of systematic uncertainties contribute to the evaluation of the modeling of data by MC. The sources of systematic uncertainties include the theoretical assumptions used for the MC predictions and the reconstruction and calibration of the detector response to the physics objects. The sources of systematic uncertainties, their effect and their estimation procedures are summarized in Tabs. 7.7 and 7.8. Systematic uncertainties are propagated to the signal efficiency measurement by repeating the fit for varied templates that correspond to each systematic uncertainty source and comparing the extracted efficiency for the nominal and varied templates.

In the analysis of signal jets with $t\bar{t}$ events, flavor tagging uncertainties are observed to be a subdominant uncertainty. For the DNN W- and top- tagging classifiers, presented in Fig. 7.10, flavor tagging uncertainties have a large effect at low values of the discriminants where non-top-quark jet contributions are more dominant due to misidentification of non-b-jets as b-jets. Although this effect is seen in distribution of the discriminants, it does not affect the signal efficiency measurements significantly mainly due to the higher discriminant cut values required by the tagging working points. Theoretical modeling of the MC predictions is the dominant source of systematic uncertainties in both the analysis of signal jets with $t\bar{t}$ events and multijet and $\gamma+$ jet background events. The contribution of the uncertainty on the modeling of parton shower and hadronization effects is dominant in all cases, leading to variations in the yield of the MC when investigating the distributions of the DNN discriminants. In the case of $t\bar{t}$ events, this effect is observed as large $t\bar{t}$ systematic uncertainties. In the case of background events, this effect is observed by the differences in predictions from PYTHIA8 and HERWIG++. The same behavior is also observed for substructure variables although they are not presented here for the sake of brevity. These dominant systematic uncertainties propagate to the measured signal and background efficiencies and lead to large variations as it can be seen in Figs. 7.11–7.22. The measured signal efficiency of the top-quark tagger is observed to be susceptible to the truth-level labelling of the top quark and the particular working point chosen for the tagger.

It was observed in sections 7.5.1 and 7.5.2 that while the behavior of the two-variable cut-based taggers and DNN-based taggers were similar in terms of signal efficiency and its related uncertainty, the behavior of the taggers were different in terms of background rejection and its uncertainty. Although the background rejection is significantly larger for the DNN-based taggers, the evaluated systematic uncertainties and differences between generators are also larger compared to the simple two-variable cut-based taggers, implying overall larger systematic uncertainties. When the taggers are used in a physics analysis, the tagger uncertainties are propagated to the final results of the analysis. Hence, larger tagger systematic uncertainties can lead to larger systematic uncertainties on analysis results and degrade the precision of a given analysis. Therefore, it is important for physics analyses to consider the trade-off between the improved discrimination power and larger associated systematic uncertainties of the DNN-based taggers. Since the implications of the tagger discrimination power and tagger systematic uncertainties might be different for each physics analysis, ideally each analysis should test if the tagger dis-

crimination power improvement is significant enough to compensate the implications of larger tagger uncertainties on the final analysis results. As the DNN-based taggers presented in this chapter overall provide significant performance improvement ($\sim$ factor of two for top-quark taggers and $\sim 20\%$ for W-boson taggers) over two-variable cut-based taggers while the increase in systematic uncertainties is not as sizable, it is in general expected that these DNN-based taggers would improve physics analysis results.

Source	Affected topologies	Description
Event generator choice	$t\bar{t}$	Hard-scattering modeling uncertainty estimated as the difference between POWHEG+HERWIG and MC@NLO+HERWIG [115].
Showering choice	$t\bar{t}$	Parton shower and hadronization modeling uncertainty estimated from the difference between POWHEG+PYTHIA6 and POWHEG+HERWIG [115].
Modelling of extra QCD radiation	$t\bar{t}$	Uncertainty in amount of initial/final-state radiation estimated as the difference between the nominal POWHEG+PYTHIA6 generator and *radLo* and *radHi* tunes of POWHEG+PYTHIA6. The radiation variations include variation of renormalization and factorization scales and the h_{damp} parameters [115].
$t\bar{t}$ total cross-section uncertainty	$t\bar{t}$	Uncertainty in normalization of $t\bar{t}$ MC contribution of magnitude $\pm5.5\%$ [112]
Single-top total cross-section uncertainty	$t\bar{t}$	Uncertainty in normalization of single-top MC contribution of magnitude $\pm5.3\%$ (a conservative estimate enveloping the uncertainties on t-channel, s-channel and Wt-channel)
W+jets total cross-section uncertainty	$t\bar{t}$	Uncertainty in normalization of W +jets MC contribution of magnitude $\pm5.0\%$ [134]
W+jets theory scale uncertainties	$t\bar{t}$	Uncertainty arising from the choice of renormalization and factorization scale, CKKW matching scale and QSF scale [135]. For the renormalization and factorization scale, $\times0.5$ and $\times2$ variations are considered and the renormalization and factorization scales are varied independently as well as in correlated and anti-correlated ways. The envelope of the variations is considered as the final renormalization+factorization scale uncertainty.
Signal normalization	multijet, $\gamma +$ jet	The uncertainty in the subtraction of processes containing a hadronically-decaying top quark or vector boson, conservatively taken to be 25%.

Table 7.7: Summary of theoretical systematic uncertainties considered in the performance measurements in data.

Source	Affected topologies	Description
anti-k_t R =1.0 trimmed jet moment scale	$t\bar{t}$, multijet, γ + jet	The uncertainty in the scale of the detector response for all jet moments derived by comparing the calorimeter quantity with the reference track jet [81].
anti-k_t R =1.0 trimmed jet moment resolution	$t\bar{t}$, multijet, γ + jet	The uncertainty in the resolution of the detector response conservatively estimated as a 2% absolute uncertainty in p_T, a 20% relative uncertainty in jet mass, and a 15% relative uncertainty in all other jet moments [131].
anti-k_t $R = 0.4$ jet energy scale and resolution	$t\bar{t}$	The uncertainty in the scale and resolution of the detector response for the jet p_T derived from simulation and *in situ* calibration measurements [71].
$E_\mathrm{T}^\mathrm{miss}$ track soft term	$t\bar{t}$	The uncertainty on the component of the $E_\mathrm{T}^\mathrm{miss}$ calculation due to energy flow that is unassigned to a calibrated physics object, estimated in-situ in Z+jet events [73].
Flavor tagging	$t\bar{t}$	The uncertainty in the scale factor correcting the efficiency response of the detector to identify heavy-flavor b- and c-jets as well as light-flavor jets derived *in situ* using $t\bar{t}$ events [130, 132].
Lepton reconstruction and calibration	$t\bar{t}$	The uncertainty in the scale factor correcting the efficiency to trigger on, reconstruct, and identify leptons as well as uncertainties in their energy and p_T scale and resolution [128, 136, 137].
Photon reconstruction and calibration	γ + jet	The uncertainty in the scale factor correcting the efficiency to trigger on, reconstruct, and identify photons [138] as well as uncertainties in their energy scale and resolution [139].
Multijet background normalization	$t\bar{t}$	The uncertainty in the data-driven prediction of the yield of multijet events, conservatively taken to be 50% based on the estimate in Ref. [140].
Multijet lepton misreconstruction efficiencies	$t\bar{t}$	The statistical uncertainty of the real and fake/non-prompt lepton reconstruction efficiencies estimated in Ref. [140] is propagated through the matrix method.
Luminosity uncertainty	$t\bar{t}$, multijet, γ + jet	A 2.1 % relative uncertainty in the MC yield, based on the luminosity uncertainty of the combined 2015+2016 dataset based on [141].
Pileup uncertainty	$t\bar{t}$, multijet, γ + jet	Uncertainty in the reweighting of MC pileup profile to the measured pileup profile in data based on disagreement between instantaneous luminosity in data and in simulation [141].

Table 7.8: Summary of experimental systematic uncertainties considered in the performance measurements in data.

7.6 Outlook

As it is generally the case, there is room for improvement in several aspects of the W-boson and top-quark tagging studies. Some possible improvements are listed.

- Jet labeling:

  The W-boson and top-quark tagging studies presented in this thesis employ jet labeling in different parts of the analysis such as the optimization, working point definition and categorization of events in signal-enriched data. The chosen jet labeling procedure is not only based on the particle of interest but on its partonic decay products. As a reminder the partonic decay products of the generator-level top quark or W boson are required to contained within the reconstructed large-R jets. It was observed that this choice introduced a strong generator dependence and resulted in larger systematic uncertainties. In the future, choosing a labeling scheme which is less generator dependent could reduce the systematic uncertainties. Moreover, contained jet labeling scheme has overall low acceptance for top-quark initiated jets. Hence, this choice is not ideal for some of the physics analyses since it results in loss of signal. The choice of jet labeling affects the performance of all introduced taggers and the favored choice (contained or more inclusive labeling) depends on the analysis.

- Deep neural network training signal sample:

  The W-boson signal sample and top-quark signal sample were obtained from simulated $W' \to WZ \to q\bar{q}q\bar{q}$ and $Z' \to t\bar{t}$ events, respectively. The obtained signal samples had significantly fewer jets in high p_T region. Although this problem was overcome by training the DNN up to 2 TeV and applying jet-based weights to obtain uniform p_T spectra, training on signal samples which provide a balanced jet p_T distribution would have advantages. One advantage of using such a signal sample would be allowing the training to be extended above 2 TeV, other would be performance improvements at high p_T.

- Jets with undefined substructure variables:

  As mentioned previously, if a jet has fewer than three constituents, some of the substructure variables are not defined. Since the fraction of such jets were sufficiently low, these jets were removed from the training sample to avoid undefined input variables. Since these jets were not used in training, these jets were assigned a pre-defined tagging decision during evaluation time. It was observed that this pre-defined decision was not optimal at high-p_T. Hence, better handling of missing variables (such as assigning an optimized default value) or an alternative pre-defined tagging decision can improve background rejection at high p_T.

Chapter 8

Search for Pair Production of Heavy Vector-Like Quarks in Hadronic Final States

Heavy vector-like quarks are predicted by several extensions to the Standard Model. This chapter presents a search for pair-produced heavy vector-like quarks, $T\bar{T}$ and $B\bar{B}$, in fully-hadronic final states with small missing transverse momentum. This chapter also presents the results of the ATLAS combination [4] of the pair-produced vector-like quark searches, to which this analysis contributes to. The data recorded by the ATLAS experiment in 2015 and 2016 corresponding to an integrated luminosity of 36.1 fb^{-1} is used.

8.1 Introduction

VLQs have been searched for at the LHC by the ATLAS and CMS experiments in Run 1 [142–153] and Run 2 [154–169] prior to the search presented in this chapter. The majority of the searches focus on vector-like T and B quark production and assume that once produced they decay into a Standard Model boson and a third-generation quark, leading to $T \rightarrow Wb, Ht, Zt$ or $B \rightarrow Wt, Hb, Zb$ decay modes. As W, Z, Higgs bosons and top quark further decay and are indirectly observed from their decay products, the VLQ decays provide a rich set of possible detector signatures.

The vector-like T and B quark production cross-sections and branching ratios of the relevant decay modes depend on the parameters of the considered model and are not known a priori. Hence, it is important for the experiments to cover the parameter space as much as possible. The ATLAS experiment has a comprehensive VLQ search program which covers a wide range of event topologies both for pair production and single production. The searches are mainly organized by the production mechanism and event topology. Each search primarily targets a certain set of VLQ decay channels and has different sensitivity to the VLQ decay modes. In Run 1, the ATLAS VLQ searches primarily focused on pair-production of vector-like T and B quarks. Since the Run 1 searches observed no significant deviation from the Standard Model predictions, upper and lower limits were set on the VLQ production cross-sections and VLQ masses, respectively [142–147]. The data recorded in 2015 and 2016 during Run 2 of the LHC has presented unique opportunities for the VLQ searches as the center-of-mass energy of the

collisions and the integrated luminosity increased significantly compared to Run 1 data. The pair-produced VLQ searches have benefited from the higher energy reach, larger datasets and dedicated novel techniques in Run 2 while the single-produced VLQ searches have become more accessible and important since the single production may become dominant for heavier (with mass above ~1 TeV) VLQs.

The analysis presented in this chapter searches for pair-produced vector-like T and B quarks in fully-hadronic final states using the data recorded by the ATLAS experiment in 2015 and 2016. While previous pair-produced vector-like T and B quark searches by the ATLAS and CMS experiments have focused on decay channels with one or more leptons in the final state, few searches by CMS in Run 1 [150–152] and one search by ATLAS in Run 2 [156] have included decay channels with fully-hadronic final states. The previous fully-hadronic ATLAS analysis in Run 2 [156] searches for pair-produced vector-like T quark events in final states with large missing transverse momentum ($E_\mathrm{T}^\mathrm{miss} > 200$ GeV), whereas the analysis presented in this chapter searches for both pair-produced vector-like T and B quark events and focuses on fully-hadronic final states with small missing transverse momentum ($E_\mathrm{T}^\mathrm{miss} < 200$ GeV). The search presented in this chapter is complementary to the other ATLAS pair-produced VLQ searches in Run 2 [154–157, 170, 171] and while this search is sensitive to all final states involving hadronic decays of W, Z, Higgs bosons and top quarks, it is especially powerful for the $B \to Hb$ decay mode.

8.2 Overview of the Search

The fully-hadronic search presented in this chapter assumes that the pair-produced vector-like quarks decay only via $T \to Wb, Ht, Zt$ or $B \to Wt, Hb, Zb$ decay modes. The analysis strategy is optimized by considering all hadronic decay modes of pair-produced vector-like quarks. Hence, final states involving b-jets and hadronic decays of W, Z, and Higgs bosons and top quarks are expected.

Overview of the analysis steps is presented along with references to the related sections in Fig. 8.1. As it can be seen in the overview, jet tagging decisions are used in different stages of the analysis and play a fundamental role. The small-R jets in the final state are $b-$ tagged, whereas the large-R jets in the final state are classified as arising from hadronically-decaying W/Z, Higgs bosons, top quarks, or background (light quark/gluon-initiated jets) using a deep neural network. Since the search focuses on final states with small $E_\mathrm{T}^\mathrm{miss}$ and b quarks, hadronically-decaying W, Z, Higgs bosons and top quarks are expected in the final state, a pre-selection is applied to select events with low $E_\mathrm{T}^\mathrm{miss}$, b-jets, W-, Z-, Higgs- boson and top-quark jets. Twelve signal regions are defined according to the expected numbers of W-,Z-, Higgs-boson, top-quark and b-jets from different VLQ decay channels. A signal log-likelihood is calculated via the matrix element method and is used as the final discriminant. In each signal region, the analysis discriminant distribution is evaluated for the VLQ signal and SM background events where the dominant multijet background contribution is estimated by a data-driven method. Finally, the statistical analysis is carried out to test the compatibility of the data with the VLQ signal in twelve signal

regions.

Reconstruct physics objects
The physics objects used in the analysis are summarized in section 8.4.

↓

Tag physics objects
The taggers used in the analysis are summarized in section 8.4.
Small-R jets are $b-$tagged.
Large-R jets are tagged as W/Z-, Higgs-boson, top-quark or background jets.

↓

Apply pre-selection
The pre-selection is presented in section 8.5.1.
Pre-selection is based on: H_{T}, $E_{\mathrm{T}}^{\mathrm{miss}}$, high-$p_{\mathrm{T}}$ small-R jets and jet tagging decisions.

↓

Define 12 signal regions
The signal regions are defined based on jet tagging decisions as presented in section 8.5.1.

↓

Evaluate the discriminant in 12 signal regions
The analysis discriminant is presented in section 8.5.1. The methods used to estimate the background contributions are presented in section 8.5.2.

↓

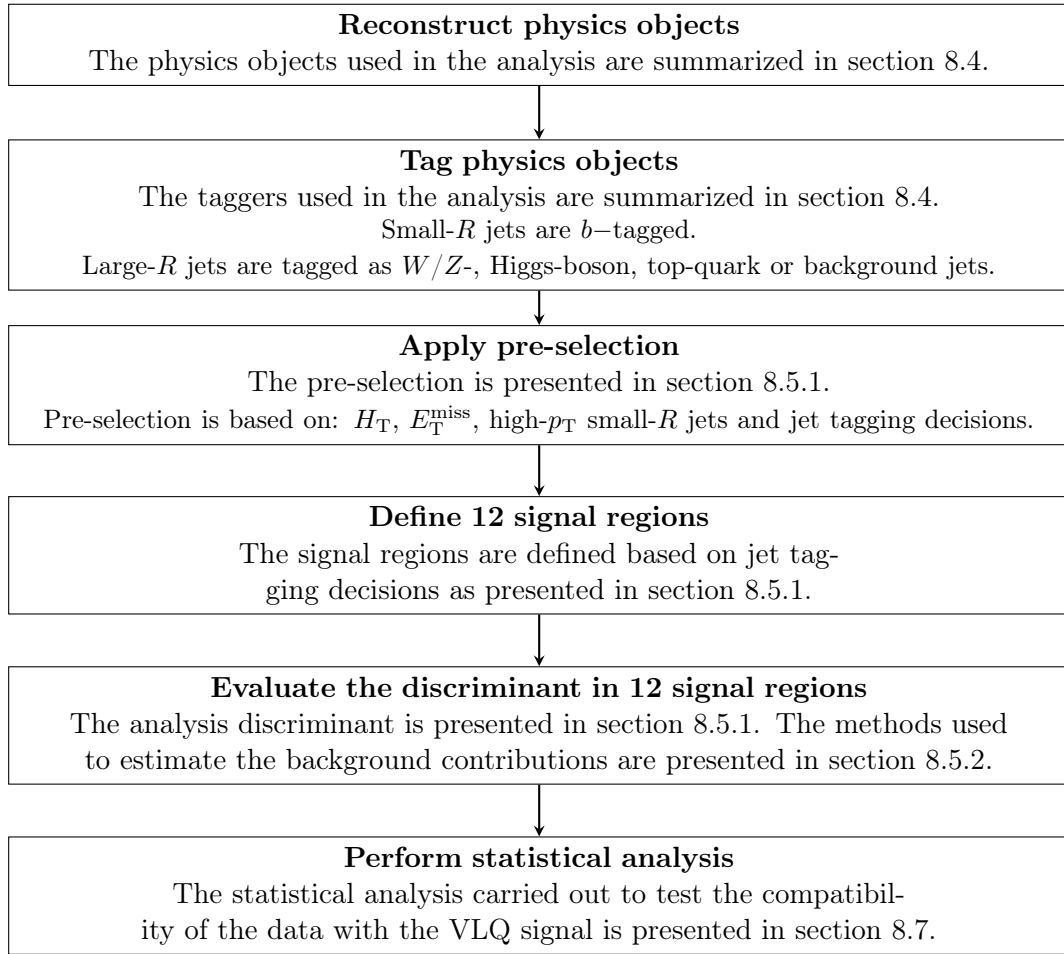| Perform statistical analysis |
|---|
| The statistical analysis carried out to test the compatibility of the data with the VLQ signal is presented in section 8.7. |

Figure 8.1: Overview of the analysis steps along with references to the related sections.

This chapter is organized as follows. After the Monte Carlo and data samples are described in section 8.3, the physics objects are described in section 8.4 with an emphasis on the development and performance of the dedicated large-R jet multi-class tagger in section 8.4.2. The analysis strategy is presented in section 8.5 which is followed by a summary of systematic uncertainties in section 8.6 and the statistical analysis in section 8.7. Finally, the results of the analysis and the results of the ATLAS combination [4] of the pair-produced vector-like quark searches, to which this analysis also contributes, are presented in sections 8.8 and 8.9.

8.3 Samples

The dataset used in this analysis was recorded in 2015 and 2016 and corresponds to an integrated luminosity of 36.1 fb^{-1}.

The VLQ pair production events, $Q\bar{Q}$, were generated using the LO generator PROTOS v2.2 [172, 173] with the NNPDF2.3 LO PDF set [174]. PYTHIA 8.186 [175] with A14 tune [103]

is used for parton showering and fragmentation. VLQs were produced for the isospin singlet scenario with a narrow width and for masses between 700 – 1200 GeV in steps of 50 GeV and for 500, 600, 1300, and 1400 GeV. Finally, additional samples were produced assuming a doublet scenario for VLQ masses of 700, 950 and 1200 GeV, in order to study differences from the different chirality of VLQs arising in singlet and doublet models.

The $Q\bar{Q}$ pair-produced cross-section varies between 3.38 ± 0.25 pb ($m_Q = 500$ GeV) and 3.50 ± 0.43 fb ($m_Q = 1400$ GeV). It is computed using `top++` v2.0 [176] at NNLO in QCD, including resummation of NNLL soft-gluon terms, and using the MSTW 2008 NNLO set of PDFs [177]. Modeling uncertainties are estimated by variations of the factorization and renormalization scales and by taking uncertainties of the PDF and strong coupling constant, α_{S}, into account. The latter two represent the largest contribution to the overall theoretical uncertainty in the predicted cross-section and are calculated using the PDF4LHC [178] prescription with the MSTW 2008 68% CL NNLO, CT10 NNLO [179, 180] and NNPDF2.3 5f FFN PDF sets.

The main backgrounds of this analysis are: multijet, $t\bar{t}$, single-top-quark and $t\bar{t} + \mathrm{X}$ ($X = W$, Z, Higgs) events. Although a data-driven method is used to estimate the QCD multijet background, simulated multijet samples are used to optimize the object selection and to validate the data-driven method. The simulated multijet events were generated using the same settings as in section 7.1, the PYTHIA8 (v8.186) [56] generator with the NNPDF2.3LO [102] PDF set and the A14 tune [103]. The $t\bar{t}$ events were generated using POWHEG-BOX v2 + PYTHIA 8.210 [181, 182] at NLO with the CT10 PDF set and and the corresponding Perugia 2012 tune (P2012) tune [183] for parton showering. The NLO radiation factor, h_{damp}, was set to $1.5m_{\mathrm{top}}$. The $t\bar{t}$ background is decomposed into $t\bar{t}$ + light-flavor jets ($t\bar{t} + \mathrm{light}$) and $t\bar{t} + c$- or b-flavor jets ($t\bar{t} + \mathrm{HF}$). Single-top-quark production (Wt and t-channel) was generated using POWHEG-BOX v1 + PYTHIA 6.428 [184–186] at NLO with the CT10 PDF set and the Perugia 2012 tune for parton showering. The $t\bar{t} + V$ ($V{=}W, Z$) and $t\bar{t} + H$ background was modeled using MADGRAPH5_aMC@NLO v2.3.2 [187] as the generator at LO with up to two additional partons and at NLO precision, respectively. The parton showering and fragmentation is performed using PYTHIA 8.210 [175] (PYTHIA 8.186) for $t\bar{t} + Z$ and $t\bar{t} + H$ ($t\bar{t} + W$).

8.4 Objects

In this analysis, hadronic final states are investigated. Consequently, the main physics objects used in this analysis are jets. Due to the expected final-state particles, small-R jets (b-quarks) and large-R jets (heavy bosons and top quark) are used to investigate the final state. Detailed information on the small-R jets and large-R jets is given later in this section. $E_{\mathrm{T}}^{\mathrm{miss}}$ is also used for event selection, it is reconstructed as described in 5.6. In this case, the negative vector sum is calculated using reconstructed jets, electrons and muons as well as ID tracks which are not associated with the leptons or jets. Finally, a lepton veto is applied to ensure orthogonality with ATLAS VLQ searches which include leptons. Same requirements are applied on the electron ("tight", isolated) and muon ("medium", isolated) candidates as it was in section 7.5.1. Events

which contain electron or muon candidates that pass the identification requirements and have $p_{\mathrm{T}} > 20$ GeV are rejected.

Small-R jets are used to reconstruct and identify the b-jets. Moreover, they are also used as inputs to large-R jet reconstruction which capture and identify the hadronic decays of boosted heavy objects (W, Z, Higgs bosons and top quarks) as it will be described in detail in 8.4.1. Small-R jets are reconstructed from EM-scale topoclusters using the anti-k_{t} algorithm with R = 0.4. The small-R jets are required to have $p_{\mathrm{T}} > 20$ GeV, $|\eta| < 2.5$. Similar to the approach in section 7.5.1, the low-p_{T} small-R jets produced in pileup interactions are suppressed by using the JVT algorithm for small-R jets with $p_{\mathrm{T}} < 60$ GeV and $|\eta| < 2.4$. Finally, a jet or a lepton is vetoed if an electron or muon selected with loosened criteria is in the proximity of a jet to avoid overlap and misidentification of objects [188, 189].

Small-R jets are tagged as b-jets if they satisfy the 77% signal efficiency working point of the MV2c10 algorithm [130]. In addition to the 77% b-tagging working point, three other b-tagging working points of MV2c10 (60%, 70%, 85%) are used in the context of the large-R jet tagging.

8.4.1 Variable-radius reclustered jets

In this analysis, variable-radius reclustered (vRC) jets are used to reconstruct the hadronic decays of W, Z, Higgs bosons and top quarks. Variable-reclustered jets are reconstructed using two methods: jet reclustering [69, 70] and variable-radius [190] algorithm.

Different than the large-R jets introduced in section 6.2 and used in the previous chapters, reclustered (RC) jets use calibrated small-R jets rather than the topoclusters as inputs to the anti-k_{t} algorithm. Since the small-R jets are calibrated, the calibration of the RC jets is automatic without the need of a dedicated RC jet calibration. Moreover, related uncertainties are easily propagated from small-R jets to the RC jets. RC jets are observed to have similar performance as the trimmed anti-k_{t} $R = 1.0$ jets [191] and have the additional benefit that the calibration and uncertainties are simplified. Due to these features and flexibility to choose the preferred radius parameter, RC jets are commonly used in ATLAS analyses.

For vRC jets, the reclustering technique is expanded to include the variable-radius algorithm (VR). The VR jet algorithm takes into account that the separation of the decay products of a heavy object scales inversely with the p_{T} of the heavy object. Therefore, the VR jet algorithm uses a varying radius, R_{eff}, instead of a fixed radius parameter. For vRC jets, the radius of each jet is defined by,

$$R_{\mathrm{eff}} = \min(R_{\mathrm{max}}, \rho/p_{\mathrm{T}}) \qquad (8.1)$$

where R_{max} is the maximum allowed size of the jet, ρ is the clustering scale constant and p_{T} is the transverse momentum of the jet with respect to the beam axis. Since the choice of the clustering scale constant can impact the performance of the vRC jet algorithm, several values are tried for this analysis and the one which gave an overall good performance, characterized by mass distribution and signal/background discrimination, for W, Z, Higgs bosons and top quarks is chosen. vRC jets are reconstructed using the small-R jets with $R = 0.4$ as inputs,

$\rho = 315$ GeV, $R_{\mathrm{max}} = 1.2$ and are trimmed with $f_{\mathrm{cut}} = 0.05$. The vRC jets are retained if they satisfy $p_{\mathrm{T}} > 150$ GeV, $|\eta| < 2.5$ and have mass greater than 40 GeV.

For the MC-based studies, the reconstructed vRC jets are labeled to indicate the particle that initiated the jet. All the vRC jets retained from the multijet events are labeled as *background.* Labeling of signal vRC jets is done by a simple truth matching approach. Reconstructed vRC jets (V-boson, Higgs-boson or top-quark jet) retained from the VLQ signal samples are matched to a hadronically decaying boson or top quark at generator level within a geometrical distance of $\Delta R = 0.75 \cdot \rho/p_{\mathrm{T}}$[1]. For the Higgs boson, only direct decays into quark pairs are considered. It is important to note that, this inclusive labeling scheme does not require the decay products of the bosons or top quark to be contained within the vRC jet. All vRC jets matched to multiple generator-level V bosons, Higgs bosons or top quarks are discarded, the fraction of such jets was found to be negligible.

The jet mass distribution is studied to demonstrate the advantages of the vRC jets over RC jets that are reconstructed with a fixed radius parameter of $R = 0.8$. In Fig. 8.2, the median of the jet mass and 68% interquantile range mass windows are presented for truth-labelled signal (W, Z, Higgs-boson and top quark) RC and vRC jets. It is observed in this figure that unlike RC jets, vRC jets do not have large tails and that the jet mass is more stable as a function of p_{T} for vRC jets.

A multi-class DNN is used to tag vRC jets as W/Z (V)-boson, Higgs-boson, top-quark, or background-jets. Since the W boson and Z boson have very similar properties and distinguishing them is challenging, they are considered as one class in the vRC-jet tagging and throughout the analysis.

---

[1]Similar procedure is followed for the studied RC jets with $\Delta R(\mathrm{parton, jet}) < 0.75R$, $R = 0.8$.
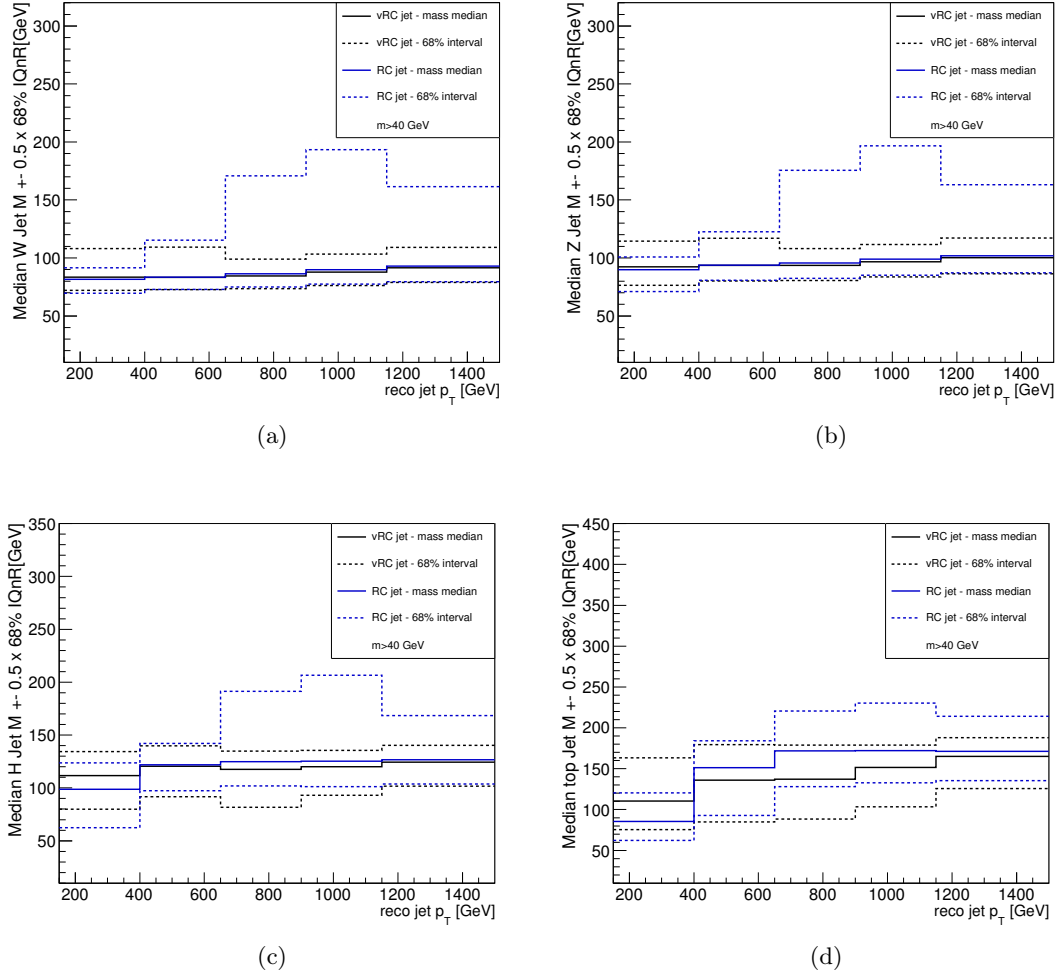
Figure 8.2: The median of the jet mass (solid) and 68% interquantile range (IQnR) mass windows (dashed) as a function of jet p_{T} for vRC jets with $\rho = 315$ GeV and RC jets with $R = 0.8$ presented for W-boson (a), Z-boson (b), Higgs-boson (c) and top-quark (d) jets.

8.4.2 Identification of vRC jets using a multi-class DNN

Two of the fundamental challenges of this analysis are to unambiguously tag the vRC jets and suppress the multijet background. In order to achieve this, it is crucial to employ a powerful boosted object tagger. It was shown in chapter 7 that usage of ML can significantly improve the signal vs background discrimination of boosted object taggers. Since the powerful ML-based taggers presented in chapter 7 are not yet calibrated and do not allow unambiguous identification of V-boson, Higgs-boson and top-quark jets, a novel multi-class boosted object tagger is developed for this analysis. A multi-class fully-connected DNN is trained to tag vRC jets using properties of the vRC jets and its calibrated constituents as inputs to the tagger, has the benefit of simplified calibration and systematic uncertainties as described in section 8.4.1.

Inputs

The signal vRC jets (boosted V-boson, Higgs-boson and top-quark) are obtained from simulated VLQ samples. The background vRC jets are obtained from simulated multijet samples. First, in order to obtain a balanced dataset for training and validation, the simulated samples are downsampled as follows. V-boson, Higgs-boson and top-quark jets are downsampled based on their truth label to have equal number of signal jets from each of the three signal flavors. Additionally, the background sample is downsampled to match the total number of signal jets. The sample of background and signal jets are then randomly split into training (42.5%), validation (7.5%) and testing (50%) sets. Finally, the p_T distribution of the background is reweighted to match the p_T distribution of the signal[2] to prevent DNN from learning the differences between the p_T distributions of signal and background jets. This weighting scheme is used both to train the DNN and to evaluate the performance of the resulting tagger throughout this section.

The DNN is trained using the mass, p_T, and number of constituent jets of the vRC jet, as well as the four-momentum vectors and b-tagging information of the three highest-p_T (leading) constituent small-R jets (subjets) as input. If a vRC jet has fewer than 3 subjets, default values are assigned to the remaining subjets. The following physically-motivated default values are assigned for the missing subjet properties: $p_T = 0$, $\eta = $ vRC jet η, $\phi = $ vRC jet ϕ, energy $= 0$, b-tag discriminant $= -1$. Although the p_T distributions of the signal and background are the same, the p_T of the vRC jet is used as input to allow learning relations between the p_T of the vRC jets and other input features. The b-tagging information of the small-R jets is not continuous. This is due to the restriction that only pre-defined b-tagging working points are calibrated for analysis use. Consequently, the MV2c10 output is binned according to the calibrated pseudo-continuous b-tagging working point and the resulting binned pseudo-continuous b-tagging information is used as an input to the tagger. By effectively using only properties of the calibrated small-R jets as input to the tagger, all jet-related systematic uncertainties can be propagated through the DNN by varying the corresponding properties of the small-R jets. The input variables are standardized prior to training as explained in section 3.1.4.

---

[2]p_T distributions of signal objects are observed to be very similar, hence did not require reweighting.

Name	Description
vRC jet mass	Mass of the reconstructed vRC jet
vRC jet p_{T}	Transverse momentum of the reconstructed vRC jet
Number of subjets	Number of small-R jets re-clustered in the vRC jet
Subjet p_{T}	Transverse momentum of the three leading subjets
Subjet η	η of the three leading subjets
Subjet ϕ	ϕ of the three leading subjets
Subjet energy	Energy of the three leading subjets
Subjet b-tag discriminant	Pseudo-continuous b-tagging information of the three leading subjets

Table 8.1: Summary of the 18 input variables used in the training of the VLQ DNN tagger.

The training input variables are summarized in Tab. 8.1 and four of the input variables are presented in Fig. 8.3. As it can be seen in Fig. 8.3, the vRC jet mass distribution of top-quark jets has two significant peaks around the top-quark mass and W-boson mass as expected due to the p_{T} range of the vRC jets and the inclusive truth-labeling. In Fig. 8.3 it is also seen that there are bumps in the p_{T} spectrum of the leading small-R jet for V- and Higgs-boson jets, these bumps are observed due to the merging of the constituent small-R jet.

Training

The DNN consists of four fully-connected hidden layers and a four-dimensional output layer where four nodes of the output layer correspond to four jet flavors: V, Higgs, top and background. The DNN is trained using the *Adam* [33] optimizer algorithm. Hidden layers of the DNN use batch normalization [34] and relu activation functions, whereas the output layer uses a sigmoid function. The hyper-parameters of the DNN are optimized by a grid search where the architecture, learning rate, L1 regularizer and batch size are varied. To ensure that there is no overtraining, in addition to monitoring the losses on the training and validation sets, the DNN output distributions are studied in different sets. The chosen hyper-parameters for the DNN architecture and training are presented in Tab. 8.2.

Definition and performance of the tagger working point

To reduce the four-dimensional DNN output information (D_{DNN}), the outputs of the different classes are combined by building a discriminant function, $P(X)$, for each signal: $P(V)$ for V-boson tagging, $P(H)$ for Higgs-boson tagging and $P(t)$ for top-quark tagging. A selection on the $P(X)$ is applied to make an independent binary decisions for each signal. Once this decision is made independently by each $P(X)$, there are vRC jets which are double or triple tagged as V, Higgs or top. Hence, while defining $P(X)$ for each signal, although rejecting background jets is prioritized, discrimination amongst the three signals is also considered. In order to reject background jets strongly while obtaining good discrimination between signal classes, a similar
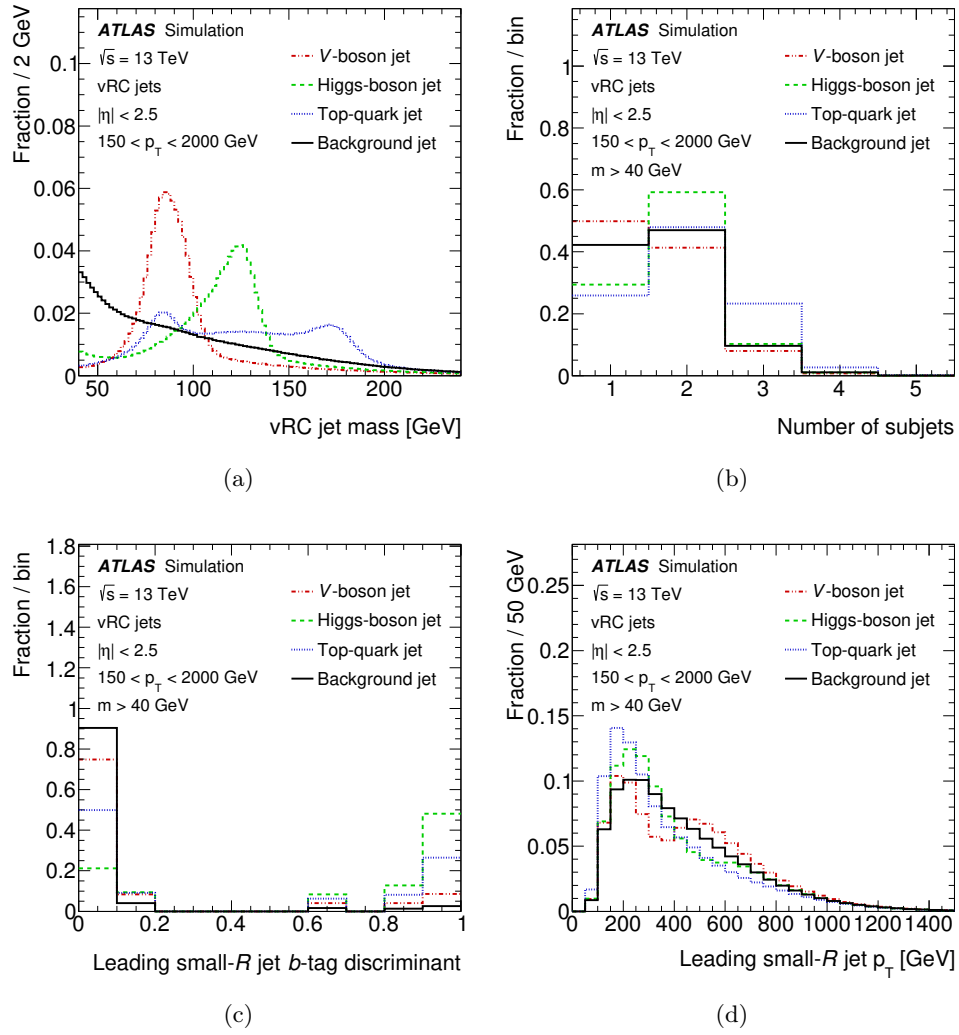
Figure 8.3: The distributions of four of the input variables used to train the DNN to tag vRC jets: The vRC jet mass (a), number of subjets to the vRC jet (b), *b*-tagging discriminant of the leading small-*R* jet (c) and the p_T of the leading small-*R* jet (d).

Name	Value
Software package	KERAS 1.0.8 WITH THEANO BACKEND, LWTNN 2.0
Number of hidden layers	4
Number of nodes per hidden layer	32, 27, 14, 12
Optimizer	Adam
Learning rate	0.00001
L1 regularizer	0.001
Batch size	400
Epochs	100 with early stopping

Table 8.2: Settings used for the DNN architecture and training.

approach as the ATLAS DL1 flavor tagging algorithms [192] is followed and a discriminant function in the form of

$$P(X) = \log_{10}\left(\frac{D_{\mathrm{DNN}}^{X}}{c_1 \cdot D_{\mathrm{DNN}}^{\mathrm{background}} + c_2 \cdot D_{\mathrm{DNN}}^{sig_2} + c_2 \cdot D_{\mathrm{DNN}}^{sig_3}}\right) \tag{8.2}$$

is built for each signal X, where sig_2 and sig_3 represent the other two signal classes [3]. This functional form is chosen since it gives different importance to background jet discrimination and discrimination between signal classes by employing coefficients (c_1, c_2) for each $P(X)$ which are then tuned. In order to tune the coefficients, c_1 and c_2 are varied and the performance of the resulting taggers are compared. Three sets of coefficients are studied where the background jet discrimination is always given the priority: $\{c_1 = 1.0, c_2 = 0\}$, $\{c_1 = 0.9, c_2 = 0.05\}$, $\{c_1 = 0.8, c_2 = 0.1\}$. The discriminant functions $P(V)$, $P(H)$ and $P(t)$ are then defined as

$$
\begin{aligned}
P(V) &= \log_{10}\left(\frac{D_{\mathrm{DNN}}^{V}}{0.9 \cdot D_{\mathrm{DNN}}^{\mathrm{background}} + 0.05 \cdot D_{\mathrm{DNN}}^{t} + 0.05 \cdot D_{\mathrm{DNN}}^{H}}\right), \\
P(H) &= \log_{10}\left(\frac{D_{\mathrm{DNN}}^{H}}{0.9 \cdot D_{\mathrm{DNN}}^{\mathrm{background}} + 0.05 \cdot D_{\mathrm{DNN}}^{V} + 0.05 \cdot D_{\mathrm{DNN}}^{t}}\right), \\
P(t) &= \log_{10}\left(\frac{D_{\mathrm{DNN}}^{t}}{0.9 \cdot D_{\mathrm{DNN}}^{\mathrm{background}} + 0.05 \cdot D_{\mathrm{DNN}}^{H} + 0.05 \cdot D_{\mathrm{DNN}}^{V}}\right)
\end{aligned}
\tag{8.3}
$$

where the coefficient factors of $c_1 = 0.9$ for background jets and $c_2 = 0.05$ for V-boson, Higgs-boson or top-quark jets are chosen as a compromise between background rejection and the ability to discriminate amongst the three signals. For each signal discriminant $P(X)$, an optimized working point is defined to obtain a boosted-object tagger with a chosen signal efficiency.

Working points for each signal are defined in a single inclusive p_{T} bin and optimal set of tagging working points is chosen by studying the impact on the expected sensitivity of the VLQ analysis. The 70% working point is chosen for V- and Higgs-boson taggers, which correspond to the thresholds $P(V) > -0.2$ and $P(H) > 0.35$. The 60% working point is chosen for top-quark tagger which corresponds to a threshold of $P(t) > 0.1$.

The distributions of discriminants and the corresponding thresholds for these working points are shown in Fig. 8.4. The multi-peak behavior in some of the discriminants is a result of differences in important vRC jet properties used as input to the DNN, such as the mass, number of constituent small-R jets, and b-tagging information of the subjets. These properties relate to, for example, whether or not all of the decay products of the V boson, Higgs boson or top quark are contained within the vRC jet. The double-peak structure in the distribution of Higgs-boson jets observed in Fig. 8.4 arises predominantly from whether the vRC jet contains the two expected subjets from the Higgs-boson decay or if it contains additional hadronic energy. The dependence of the discriminants on input variables is studied by investigating relevant

---

[3]For example if X represents V boson, sig_2 represents Higgs boson and sig_3 represents top quark.
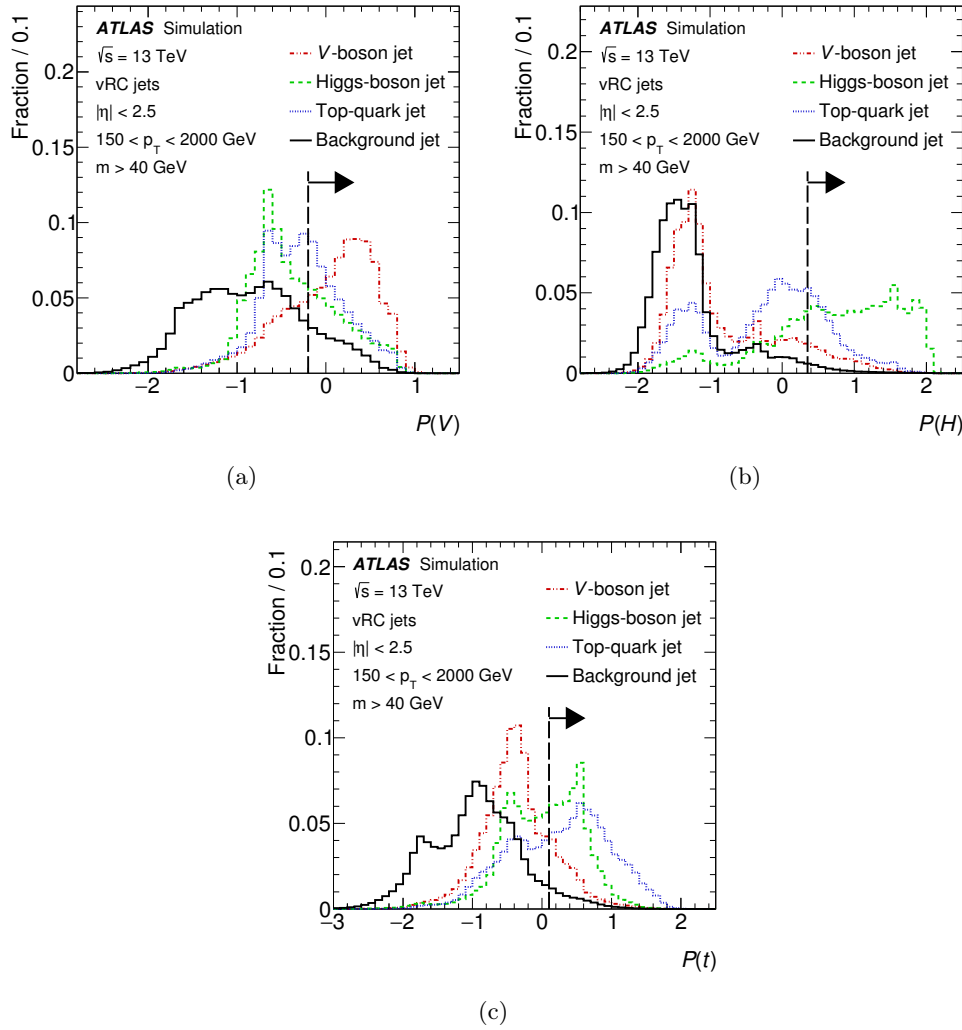
Figure 8.4: The distributions of discriminant P for the V-tagger (a), Higgs-tagger (b) and top-tagger (c). The dashed vertical line indicates the applied tagging selection.

two-dimensional distributions. The distributions are presented in App. C. The resulting signal efficiency and background rejection (estimated from weighted simulated multijet events) for each boosted-object tagger is shown as a function of p_T in Fig. 8.5.

To handle the ambiguities due to double- or triple-tagged vRC jets, additional discriminants are defined to choose a single tag for each vRC jet. These discriminant are presented in Fig. 8.6 along with the optimized thresholds, which aims to maximize the efficiency of both objects by maximizing their sum, chosen to resolve double-tagged vRC jets. It is observed that the jets with triple-tag are hard to distinguish and Higgs-boson jets are more frequently triple-tagged than V-boson jets or top-quark jets. Hence, triple-tagged vRC jets are tagged as a Higgs boson. The vRC jets that are tagged as a V boson, top quark, or Higgs boson are referred to as V-tagged, top-tagged, and Higgs-tagged, respectively.

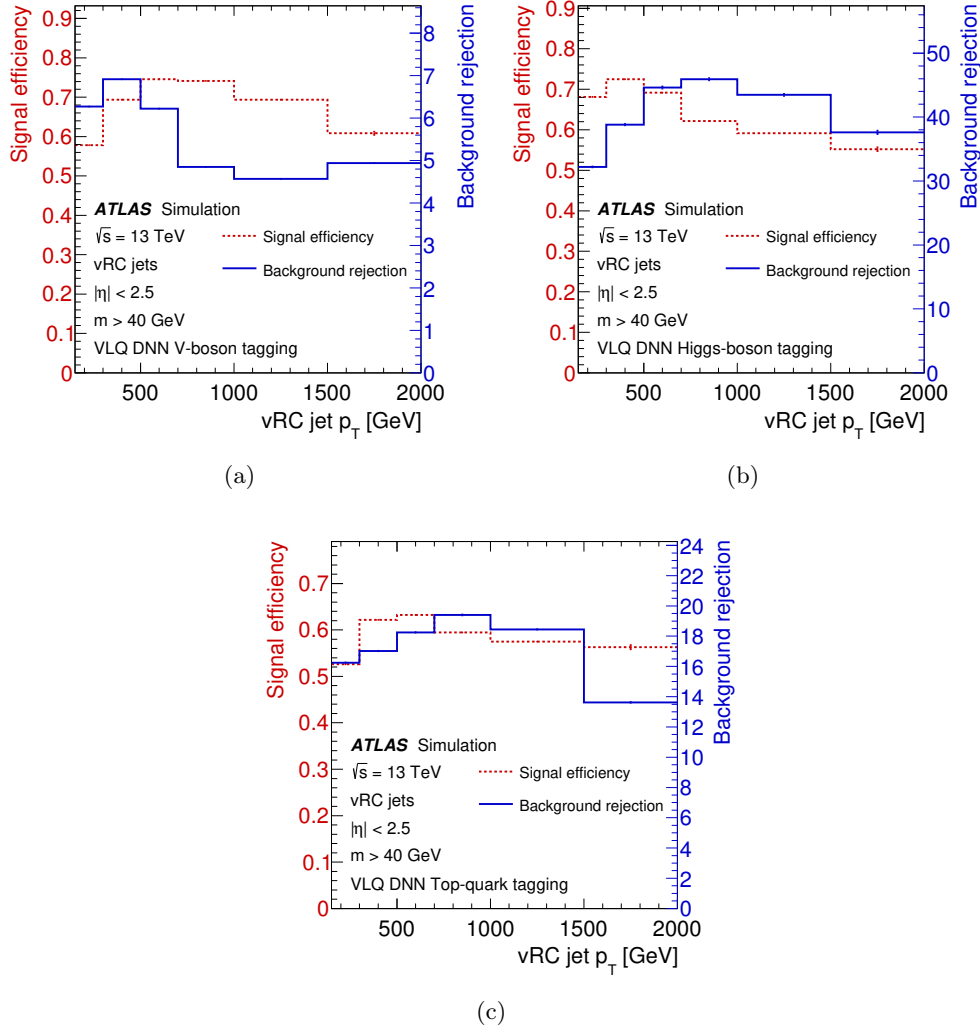It is observed that the VLQ DNN tagger provides a strong signal, background discrimination.

Figure 8.5: The signal efficiency (dashed) and multijet background rejection (solid) as a function of vRC jet p_T for the DNN V-boson tagger (a), Higgs-boson tagger (b) and top-quark tagger (c). The dashed lines refer to the left y-axis scale, and the solid lines refer to the right. Background jets are taken from simulated weighted multijet events. Statistical uncertainties are shown for signal and background.
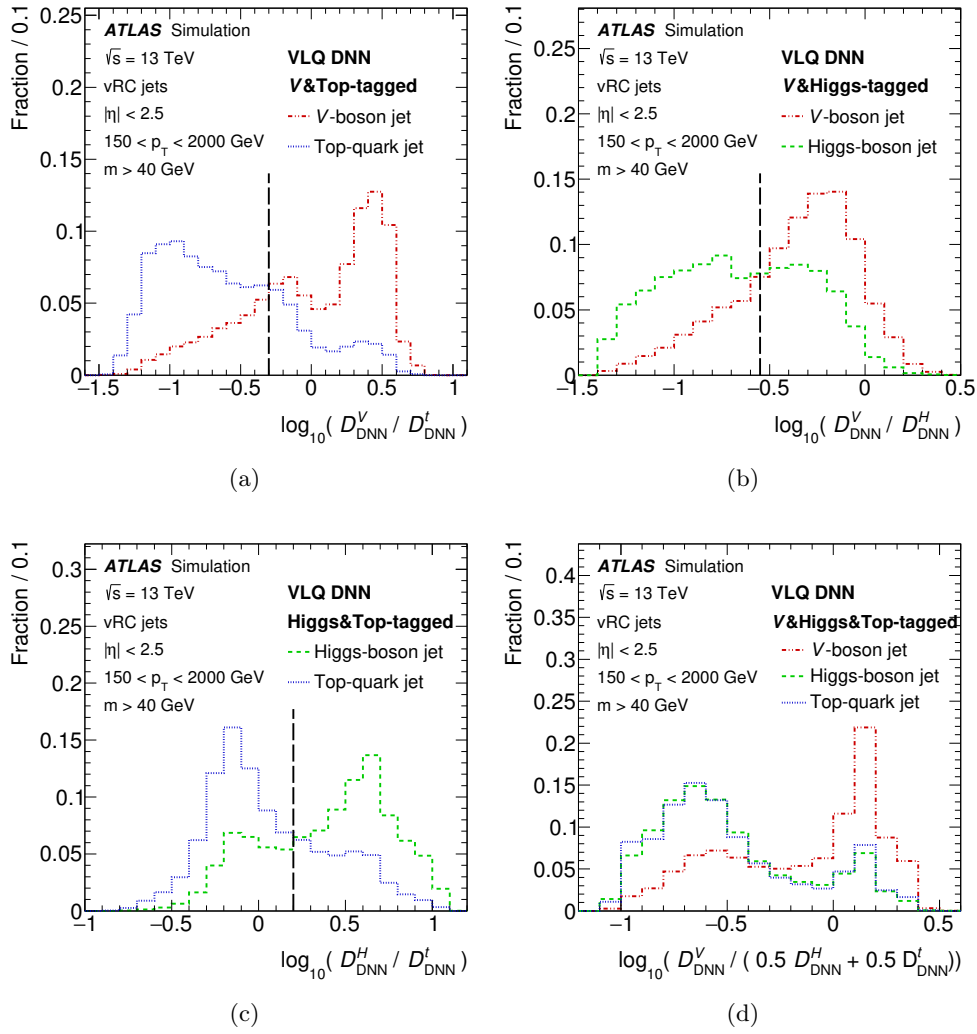
Figure 8.6: The additional discriminant functions defined to resolve multiple-tagged vRC jets such as V- and top-tagged (a), V- and Higgs-tagged (b), Higgs- and top-tagged (c) and V-, Higgs- and top-tagged (d). The dashed vertical line indicates the applied selection.

Finally, to study the impact of the tagger, the shape of the vRC jet mass distribution before and after the final boosted-object tagging after resolving ambiguities is shown in Fig. 8.7 for each jet type. As expected, each tagger preferentially selects vRC jets with a mass near the mass of the target particle. In the case of the top-quark jets, vRC jets with a mass near the W-boson mass are often V-tagged (dominant at low p_{T}) and Higgs-tagged.
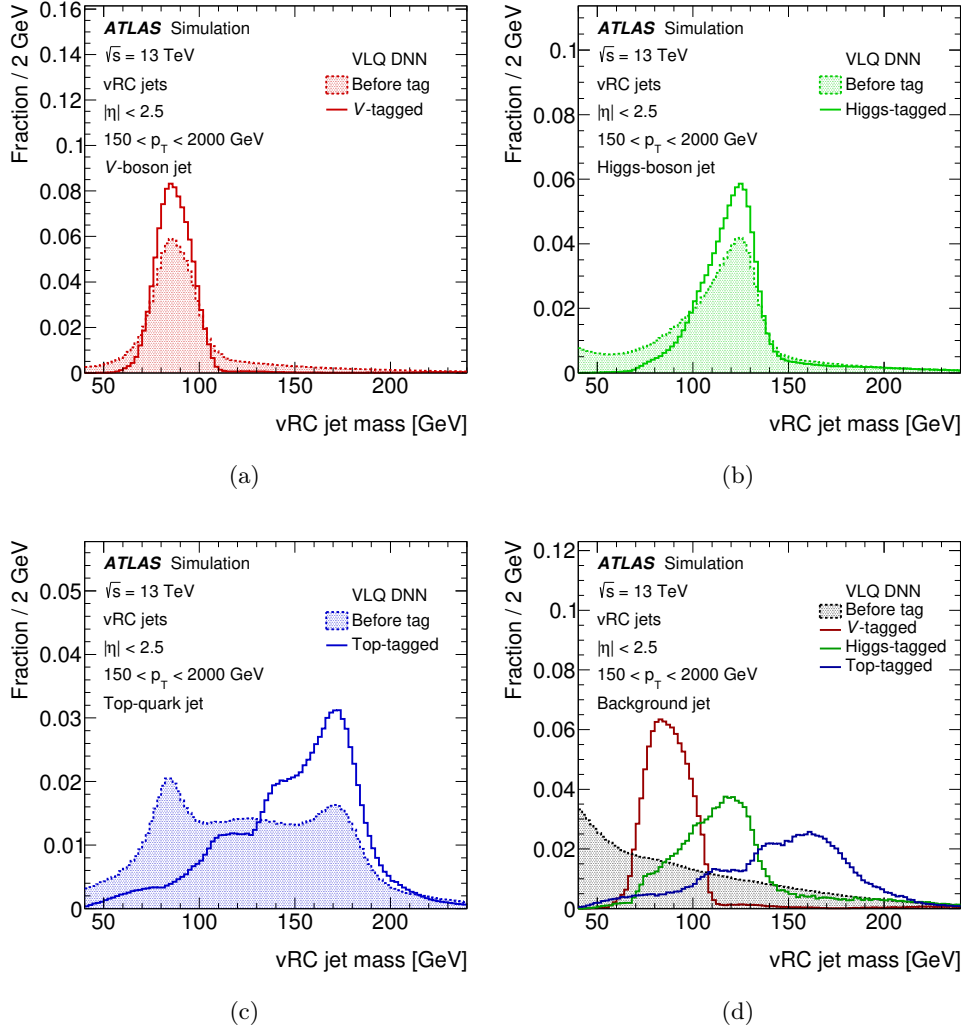
Figure 8.7: The mass distribution of *V*-boson jets (a), Higgs-boson jets (b), top-quark jets (c) and background jets (d) are shown before and after the final VLQ DNN boosted-object tagging. For signal jets, only the impact of the correct tag is shown, while for background jets the impact of each boosted-object tag is shown. All distributions are normalized to unit integral.

8.5 Analysis Strategy

In this analysis fully-hadronic final states with small $E_\mathrm{T}^\mathrm{miss}$ are analyzed, where all possible hadronic decays of W, Z, Higgs bosons and top quarks are considered. The analysis strategy is optimized by considering all considered hadronic decay modes of pair-produced VLQs. The fundamental aspects of this search are to suppress the multijet background and to model the remaining multijet background contribution in the signal regions well. In order to suppress the multijet background, first multiple high-p_T and b-tagged small-R jets are required. Then, as a pre-selection events are required to contain at least one V-tagged, top-tagged, or Higgs-tagged vRC jet, identified by the VLQ DNN tagger. Selected events are categorized into twelve orthogonal signal regions according to the numbers of V-, Higgs- and top-tagged vRC jets and b-tagged small-R jets. The signal regions are designed to cover all possible VLQ decays and to improve the signal to background ratio, referred to as the signal significance. Finally, the matrix element method [193] is used to calculate the signal probability in each signal region in order to discriminate VLQ signal and multijet events. A binned likelihood fit is performed using the signal probability to extract the signal strength and improve the modeling of the background. The multijet background contribution is estimated in twelve signal regions for each bin of the likelihood distribution by using a bin-by-bin $ABCD$ method.

8.5.1 Event selection and categorization

To select the events of interest for the analysis, the data and MC events are first required to pass a jet-based trigger. The chosen jet-based trigger requires a single jet with $p_\mathrm{T} > 100$ GeV at the first trigger level and a total scalar sum of the transverse momenta of all track particles and energy deposits, H_T, to be greater than 1000 GeV at the high-level trigger. An offline requirement of $H_\mathrm{T} > 1250$ GeV is applied to ensure the full-efficiency of the trigger. Events are then required to have no leptons, as defined in section 8.4, and to have low $E_\mathrm{T}^\mathrm{miss}$, $E_\mathrm{T}^\mathrm{miss} < 200$ GeV, to suppress background, maximize the signal significance and to be orthogonal with other ATLAS VLQ searches. The resulting events enter the twelve signal regions if they have at least four high-p_T small-R jets with p_T thresholds of 300, 200, 125, and 75 GeV, and at least two b-tagged small-R jets. In addition, the events are required to have at least two V- or Higgs- tagged vRC jets and $E_\mathrm{T}^\mathrm{miss} > 40$ GeV.

The twelve signal regions, listed in Tab. 8.3, are defined by the number of V- and Higgs- tagged vRC jets, top-tagged vRC jets and b-tagged small-R jets. In Fig. 8.8, fraction of events from each background source that contribute to each signal region after the event selection and the background-only fit to data described in section 8.7 is presented. In addition to the twelve signal regions, nine validation regions are defined to evaluate the closure uncertainty of the multijet background estimation (seven) and for its validation (two) as described later in section 8.5.2.

Region Name	V-tags	H-tags	top-tags	b-tags	MEM final states
(VV,0t,2b)	2	0	0	2	WbWb,ZbZb
(VV,0t,3b)	2	0	0	≥ 3	WbWb,ZbZb
(VV,1t,2b)	2	0	1	2	ZtWb,WtZb
(VV,1t,3b)	2	0	1	≥ 3	ZtWb,WtZb
(VH,0t,2b)	1	1	0	2	WbWb,ZbZb
(VH,0t,3b)	1	1	0	≥ 3	WbWb,ZbZb
(VH,1t,2b)	1	1	1	2	HtWb,WtHb
(VH,1t,3b)	1	1	1	≥ 3	HtWb,WtHb
(HH,0t,3b)	0	2	0	≥ 3	HbHb
(HH,1t,3b)	0	2	1	≥ 3	HtWb,WtHb
(XX,2t,2b)	≥ 0	≥ 0	≥ 2	2	HtHt,ZtZt,WtWt,HtZt
(XX,2t,3b)	≥ 0	≥ 0	≥ 2	≥ 3	HtHt,ZtZt,WtWt,HtZt

Table 8.3: The definition of the twelve signal regions in the analysis. The number of b-tags is based on all small-R jets, including those used to construct vRC jets with V-, Higgs-, or top-tags. The last two signal regions require two bosons of any type X (V or Higgs boson). The rightmost column lists the matrix element method (MEM) final states used to define the signal hypothesis in section 8.5.1.
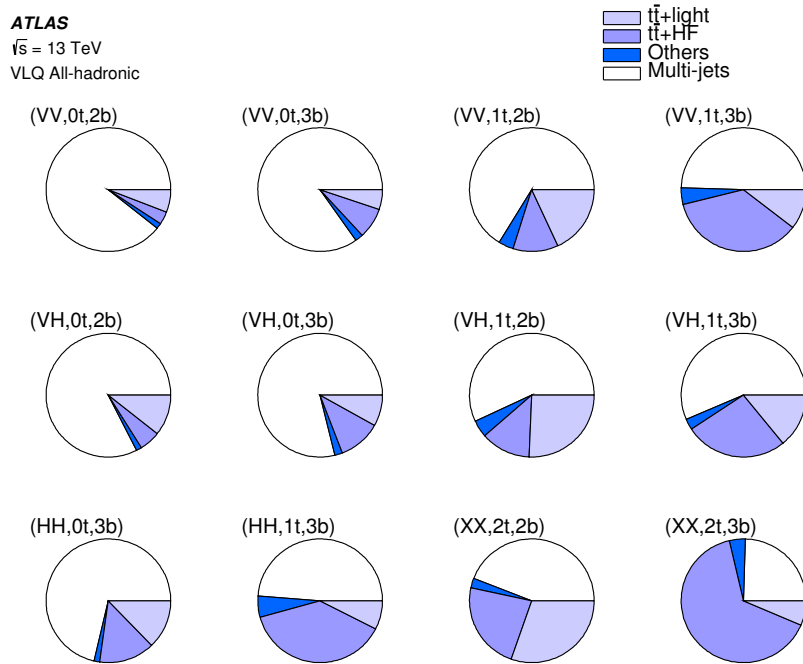

Figure 8.8: Fraction of events from each background source that contributes to each signal region after the event selection and the background-only fit to data as described in section 8.7. Others refers to backgrounds from single-top-quark and $t\bar{t} + \mathrm{X}$ production.

Matrix element method-based discriminant

In this search, matrix element method (MEM) is used to define a discriminant which provides good discrimination between VLQ signal and SM background events. MEM exploits the theoretical and experimental information to assign each observed event a probability. This probability quantifies the compatibility of an observed event with a theoretical hypothesis. First, an event-based probability density function for a physics process i, $P_i\left(\boldsymbol{x}|\boldsymbol{\alpha}\right)$, is calculated by computing the matrix element of the process and a set of theoretical and experimental parameters $\boldsymbol{\alpha}$ as,

$$P_i\left(\boldsymbol{x}|\boldsymbol{\alpha}\right) = \frac{(2\pi)^4}{\sigma_i^{\mathrm{eff}}\left(\boldsymbol{\alpha}\right)} \int d\Phi_N\left(\boldsymbol{y}\right) \boldsymbol{f}\left(p_{\mathrm{A}}\right) \boldsymbol{f}\left(p_{\mathrm{B}}\right) \frac{\left|\mathcal{M}_i\left(\boldsymbol{y}|\boldsymbol{\alpha}\right)\right|^2}{\mathcal{F}} W\left(\boldsymbol{y}|\boldsymbol{x}\right). \tag{8.4}$$

In Eqn. 8.4, $\boldsymbol{x}$ and $\boldsymbol{y}$ represent the four-momentum vectors of initial- and final- state particles at reconstruction and parton level, respectively. The transfer functions $W\left(\boldsymbol{y}|\boldsymbol{x}\right)$ map the reconstruction level quantities to the parton level quantities. Since the MEM is a computationally expensive method, the transfer functions in this thesis follow the simple definition as in the ATLAS detector simulation of the Delphes framework [194] and are parameterized as single-Gaussian functions. Several more complex and computationally more expensive parameterizations such as the one in the KLFitter package [195] were investigated as alternative transfer functions and the loss on the expected analysis sensitivity was found to be small. Hence, the single-Gaussian transfer function used in this search is observed to be a good compromise between discrimination power and computation time. $\mathcal{M}_i(\boldsymbol{y}|\boldsymbol{\alpha})$ is the transition matrix element, which is defined by the Feynman diagrams of the hypothesis hard-scattering process. It is calculated using MAD-GRAPH 5 [187] in LO. The VLQ pair production matrix element is calculated using the Feynrules model [196]. The functions $\boldsymbol{f}(p_{\mathrm{A}})$ and $\boldsymbol{f}(p_{\mathrm{B}})$ are the PDFs for the initial-state partons which have momenta p_{A} and p_{B}, respectively. The effective cross-section $\sigma_i^{\mathrm{eff}}(\boldsymbol{\alpha})$ takes into account the experimental acceptance and efficiency and normalizes P_i to unity. Finally, the flux factor, $\mathcal{F}$, and phase-space element $d\Phi_N$ describe the kinematics of the process. The numerical integration is performed over the phase space of the initial- and final-state particles where the integration variables are the energies.

The V-, Higgs-, top-tagged vRC jets and b-tagged small-R jets are assigned to final states. Since the reconstructed jets can form multiple candidate VLQ final states, the process probability density is calculated for each allowed assignment permutation of the jets to the final-state partons. In the signal regions, up to two boson tags are expected at the final state, if there are more than two boson-tagged vRC jets, only the two highest-p_{T} ones are used. If some of the vRC jets have the same tag, they are permuted. If b-quarks are expected in the final state under hypothesis test, maximum five b-tagged small-R jets are assigned to the final state and are permuted. In order to consider all the allowed assignment permutations, a likelihood function is defined as,

$$\mathcal{L}_i\left(\boldsymbol{x}|\boldsymbol{\alpha}\right) = \sum_{p=1}^{N_{\mathrm{P}}} P_i^{\mathrm{p}}\left(\boldsymbol{x}|\boldsymbol{\alpha}\right), \tag{8.5}$$
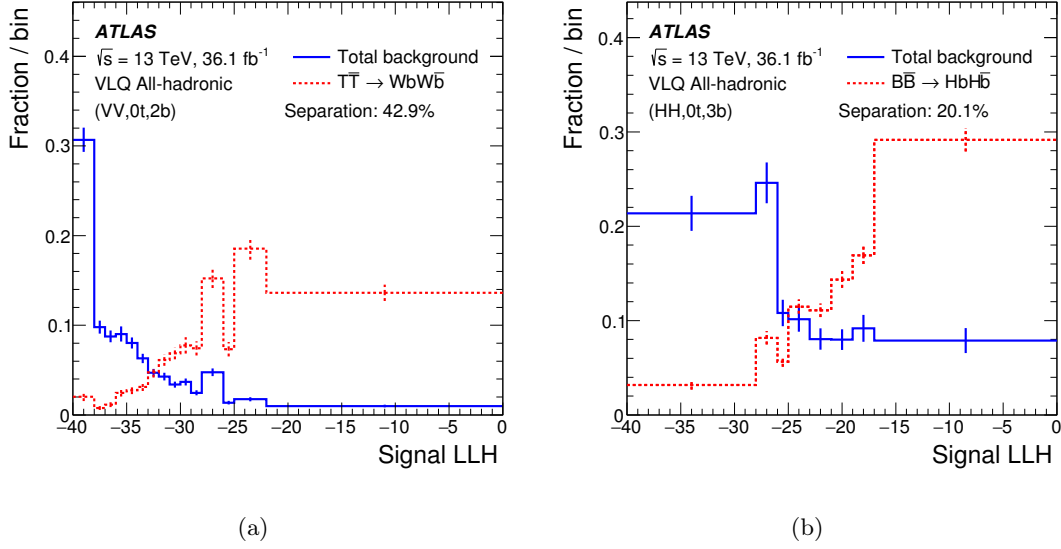
Figure 8.9: The signal LLH distributions of the simulated background and vector-like quark pair production ($m_{T/B} = 1\,\text{TeV}$) in (VV,0t,2b) and (HH,0t,3b) signal regions. Exclusive decays of $T \to Wb$ (a) and $B \to Hb$ (b) are assumed. Distributions are normalized to unit integral.

where N_P is the number of allowed assignment permutations.

In this analysis only probabilities of signal hypotheses are calculated. The signal hypothesis is computed from all Feynman diagrams of $T\bar{T}$ or $B\bar{B}$ production which have the same number of top quarks, V bosons, and Higgs bosons at the final stage as defined in Tab. 8.3. Based on early sensitivity studies of the analysis the masses of the vector-like B and T quarks are set to 900 GeV in the calculation of the matrix element. Moreover, V bosons, Higgs bosons and top quarks are assumed to be reconstructed as vRC jets in the calculation.

In addition to the already-mentioned simplifications, more simplifications are applied to reduce the complexity and the computation time of the likelihood calculation as described in Ref. [197]. The logarithm of the resulting signal likelihoods (signal LLH) is used in each signal region as the analysis discriminant. The signal LLH for the total background and signal simulations in two of the most sensitive signal regions are presented in Fig. 8.9 where exclusive decays $T \to Wb$ and $B \to Hb$ are assumed. The discrimination power of the discriminant is quantified by the separation. Separation, S, is defined as

$$\frac{1}{2} \int \frac{(S(x) - B(x))^2}{S(x) + B(x)} dx \tag{8.6}$$

where $S(x)$ and $B(x)$ are the signal and background yields per bin and S and B are normalized to unity.

8.5.2 Background estimation

Simulated and data-driven techniques are used to estimate the SM background contributions to signal regions. Data-driven background estimation techniques are often used in analysis when the modeling or the sample size of a simulated samples are limited. Similar to the approach in section 7.5.1, $t\bar{t}$, single top, and $t\bar{t} + \mathrm{X}$ background contributions are obtained from the MC samples, whereas the multijet background contribution is estimated by a data-driven method.

The multijet background contribution is estimated individually for each signal regions using the ABCD method. The ABCD method relies on two uncorrelated event-based features which are used to maximize the signal significance in an analysis. Each of the two event-based features can be a discrete or continuous variable as well as a selection. Two of these features are used to create a two-dimensional plane which is divided into four regions: three control regions and one signal region. Three control regions are defined as background-dominated regions which have sufficient number of background events and negligible signal leakage. The principle of the ABCD method is to use the background yields in three control regions to estimate the small background contribution in the signal region.

In this analysis $E_{\mathrm{T}}^{\mathrm{miss}}$ and boson tagging are used to define the four regions of ABCD. The x-axis of the ABCD plane is defined by an $E_{\mathrm{T}}^{\mathrm{miss}}$ threshold, whereas the y-axis of the ABCD plane is defined by the number of *loose-* and *tight-* tagged bosons as it can be seen in Fig. 8.10. The number of tight-tagged bosons is obtained by the number of DNN VLQ $V-$ or Higgs-tagged vRC jets. The number of loose-tagged bosons is obtained by a selection of a simple mass window for the vRC jets. The vRC jets which have a mass between 69-104 GeV or 104-155 GeV are loose-tagged as V bosons or Higgs bosons, respectively. To validate that the two axes are uncorrelated, the correlation between these two axes is evaluated in simulated multijet events and is found to be negligible. In Fig. 8.10 the region D is the signal region whereas regions A, B, C are the control regions. The four orthogonal regions are defined as:

- Region A: ≥ 2 vRC jets that are {tight-tagged or loose-tagged bosons} and < 2 vRC jets that are {tight-tagged bosons} and $E_{\mathrm{T}}^{\mathrm{miss}} < 40$ GeV,

- Region B: ≥ 2 vRC jets that are {tight-tagged bosons} and $E_{\mathrm{T}}^{\mathrm{miss}} < 40$ GeV,

- Region C: ≥ 2 vRC jets that are {tight-tagged or loose-tagged bosons} and < 2 vRC jets that are {tight-tagged bosons} and $E_{\mathrm{T}}^{\mathrm{miss}} \geq 40$ GeV,

- Region D: 'Signal Region,' ≥ 2 vRC jets that are {tight-tagged bosons} and $E_{\mathrm{T}}^{\mathrm{miss}} \geq 40$ GeV.

The event yields in the three control regions are obtained after subtracting the other SM background contributions from the data. The yields in the region D, N_{D}, is calculated individually for each of the twelve signal regions. Moreover, it is calculated on a bin-by-bin basis in the
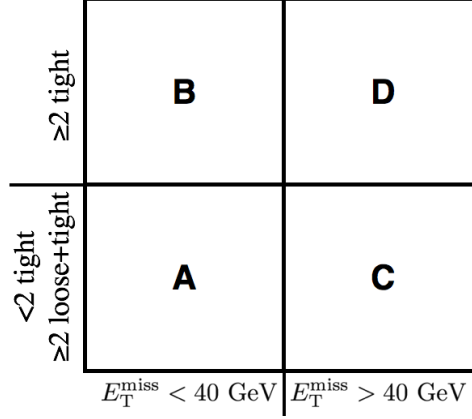
Figure 8.10: ABCD method region definitions for multijet background estimation.

signal LLH discriminant. N_{D} for each signal LLH bin is calculated as,

$$N_{\mathrm{D}} = N_{\mathrm{C}} \times (N_{\mathrm{B}}/N_{\mathrm{A}}) \qquad (8.7)$$

where N_{A}, N_{B} and N_{C} are the observed yields in the A, B, C regions for each of the signal LLH bins. The bin size of the signal LLH distribution is varied in order to have sufficient number of events in each signal LLH bin of the control regions. The bin size is determined such that at least 50 events per bin are in each of the control regions. The total yield in region D is calculated from the integral over the full distribution after the bin-by-bin calculation. This approach estimates both the shape, which may differ for different phase spaces, and the total yield of the multijet background. An uncertainty is assigned to each bin of the multijet background prediction to account for statistical uncertainties in the control regions propagated through the method. In addition to the twelve signal regions, in total nine validation regions are defined. Seven validation regions are defined to evaluate an uncertainty on the closure of the method by changing the number of b-tagged small-R jets to be exactly one while keeping the requirement on the number of boson and top-quark tagged vRC jets the same as in each of the corresponding signal regions. One of these validation regions is presented in Fig. 8.11, where only statistical uncertainties and detector simulation uncertainties are taken into account. Two validation regions are defined for the validation of the multijet background estimation by regions which have two Higgs-tagged vRC jets, exactly two b-tagged small-R jets, and either zero or one top-tagged vRC jet.

8.6 Systematic Uncertainties

The systematic uncertainties in this analysis include the theoretical assumptions used for the simulations, the reconstruction and calibration of the detector response to the physics objects and data-driven background modeling. Systematic uncertainties are propagated to the final results of the analysis by introducing each source of uncertainty as a nuisance parameter in the final likelihood fit, described in next section.
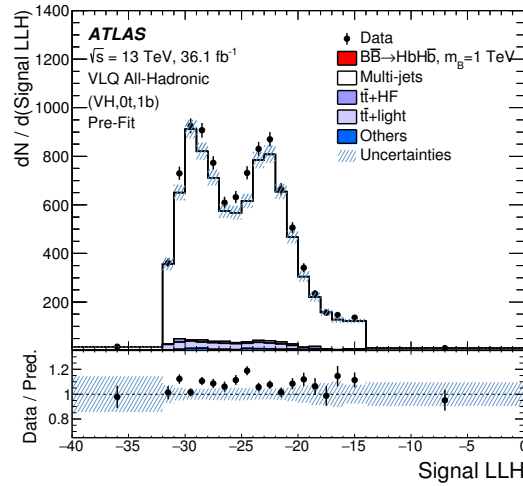
Figure 8.11: Comparison between data and MC prediction for the signal LLH in a validation region. The hatched area represents the uncertainty of the background calculated by adding the statistical and detector-related uncertainties in quadrature. The underflow and overflow events are included in the first and last bins, respectively.

The source of systematic uncertainties can be categorized in three broad categories: luminosity and pileup, object reconstruction and background modeling uncertainties. Each source of uncertainty is assumed to be fully correlated across all regions and samples, different sources of uncertainties are taken to be uncorrelated between themselves. The multi-jet closure, $t\bar{t}$ modeling, small-R jet energy resolution (JER) and jet energy scale (JES), and $E_{\mathrm{T}}^{\mathrm{miss}}$ are found to be the the dominant source of systematic uncertainties.

8.6.1 Luminosity and pileup uncertainties

Similar to the uncertainties in section 7.5.3, a 2.1 % uncertainty is included to account for the integrated luminosity measurement uncertainty of the 2015 and 2016 dataset [141]. Since the simulated events and the collected data have different pileup conditions, a correction is applied to the simulated events to have the same pileup distribution as the collected data. An uncertainty related to this correction is included.

8.6.2 Object reconstruction uncertainties

Small-R and vRC jet uncertainties

Uncertainties on the small-R jet energy and mass resolution, energy and mass scale, and JVT requirement to suppress pileup jets are considered. The dominant small-R JER uncertainties are obtained by comparing dijet events in data and simulation. The dominant small-R JES uncertainties are obtained from simulation and *in situ* calibration measurements [71]. The uncertainties on the vRC jets and the VLQ DNN tagger are directly inherited from the small-R jet

uncertainties.

Flavor tagging uncertainties

Flavor tagging uncertainties are introduced to account for the uncertainties in the scale factors which are applied to correct the b-tagging identification efficiency in simulation. The uncertainties are obtained by comparing simulated $t\bar{t}$ events and enriched data [130, 132]. Where the number of events in data is not sufficient at high-p_T, additional extrapolation uncertainties are included. While most of the vRC jet and VLQ DNN tagger uncertainties are inherited directly from small-R jets, an additional uncertainty is introduced since the VLQ DNN tagger uses additional flavor-tagging working points. This uncertainty is evaluated as a p_T-dependent signal efficiency uncertainty and is considered separately for V-boson, Higgs-boson and top-quark tagging. The uncertainty in the yields varies between 4.0% to 11.0% for simulated samples and between 0.3% to 9.4% for multijet background. Flavor tagging uncertainties have an impact on the final result but it is not one of the dominant uncertainties.

Lepton uncertainties

The lepton uncertainties are considered since a lepton veto is applied in the analysis. The lepton uncertainties are found to have negligible impact.

Missing transverse momentum uncertainties

The $E_\mathrm{T}^\mathrm{miss}$ reconstruction consists of the reconstructed jets, electrons and muons and the soft term. The uncertainties on the reconstructed jets, electrons and muons are accounted for by the above uncertainties. An uncertainty to account for the uncertainty on the soft-term component of the $E_\mathrm{T}^\mathrm{miss}$ is included. This uncertainty accounts for the modeling of the underlying event. The uncertainty in the yields reach up to 18.7% for simulated samples and 8.2% for the multijet background.

8.6.3 Background modeling uncertainties

Multijet background estimation

The closure uncertainty of the multijet background estimation is calculated by comparing the simulated multijet events and data in the seven validation regions. The seven validation regions are defined by changing the number of b-tagged small-R jets to be exactly one while keeping the number of boson and top-quark tagged vRC jets the same as in each of the corresponding signal regions. The difference between the simulated events and data in the validation regions is propagated as a normalization uncertainty to the corresponding two or three b-tag signal regions. In addition to the closure uncertainty, the uncertainties on the simulation-based subtraction of the other SM background contributions is taken into account. Finally, an uncertainty arising from the potential signal contamination in the validation regions is taken into account.

$t\bar{t}$ modeling

Similar to the uncertainties in section 7.5.3, several variations of the $t\bar{t}$ generator are used for the estimation of modeling uncertainties. Since the number of events in the variations of the $t\bar{t}$ samples are limited, uncertainties are taken into account after merging signal regions with two and minimum three b-tagged small-R jets. The uncertainties vary between 1.4% and 33% for the normalization of $t\bar{t}$ + light and between 13% and 51% for the normalization of $t\bar{t}$ + heavy flavor (HF). Systematic uncertainties accounting for variations in matrix element generator, parton shower and initial/final-state radiations are included as separate nuisance parameters in the likelihood fit. Separate normalization factors are assigned to $t\bar{t}$ +light and $t\bar{t}$ +HF contributions and are allowed to float freely in the profile likelihood fit.

Other backgrounds

An additional uncertainty is included to account for the cross-section uncertainty of the other backgrounds. It is taken to be a total uncertainty of 5.3%.

8.7 Statistical Analysis

The statistical analysis is used to quantify the compatibility of the data with the VLQ signal hypothesis. The observed signal LLH distributions in data is compared to the background-only and signal+background hypotheses in the signal regions. Hypothesis testing is performed using a modified frequentist method based on a profile likelihood that includes a treatment of the systematic uncertainties as nuisance parameters. If no significant deviation from the background-only hypothesis is observed, the CL_s method [198, 199] is used to set upper limits on the signal production cross-section of the different signal scenarios.

A binned likelihood function, $\mathcal{L}(\mu, \theta)$, is reconstructed as a product of Poisson probability terms over all bins, where μ is the multiplicative factor to the signal production cross-section and θ is the set of nuisance parameters. For a given signal, the parameter μ represents the signal strength, $\mu = 0$ corresponds to the background-only hypothesis and $\mu = 1$ corresponds to the signal+background hypothesis. There are two additional parameters included in the profile likelihood fit which correspond to the normalization of the $t\bar{t}$ +light and $t\bar{t}$ +HF contributions as mentioned in section 8.6.

The binned likelihood function, $\mathcal{L}(\mu, \theta)$, is defined as [200],

$$\mathcal{L}(\mu, \theta) = \prod_{j=1}^{N} \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \prod_{k=1}^{M} f_k(\theta) \qquad (8.8)$$

where N is the number of bins, n_j is the number of observed events, s_j and b_j are the expected number of signal and background events in bin j. The first product term corresponds to the product of Poisson probability terms over all bins and the second product term $\prod_{k=1}^{M} f_k(\theta)$ is the

constraint term introduced to take into account the systematic uncertainties through nuisance parameters. To test a hypothesized value of μ, the profile likelihood ratio $\lambda(\mu)$ is defined as,

$$\lambda(\mu) = \mathcal{L}(\mu, \hat{\hat{\theta}}_\mu)/\mathcal{L}(\hat{\mu}, \hat{\theta}) \tag{8.9}$$

where $\hat{\mu}$ and $\hat{\theta}$ are the values of the parameters that maximize the likelihood function (with the constraint $0 \leq \hat{\mu} \leq \mu$), and $\hat{\hat{\theta}}_\mu$ are the values of the nuisance parameters that maximize the likelihood function for a given value of μ. The nuisance parameters broadens the profile likelihood as a function of μ relative to the case if their values were fixed. The test statistic, q_μ, is defined as $q_\mu = -2 \ln \lambda(\mu)$ where lower values of q_μ correspond to less incompatibility between the data and hypothesized μ.

The compatibility of the data with the background-only hypothesis ($\mu = 0$) is quantified by the p_0 value as,

$$p_0 = \int_{q_{0,\mathrm{obs}}}^{\mathrm{inf}} f(q_0|0) dq_0 \tag{8.10}$$

where $f(q_0|0)$ is the probability density function of the statistic q_0 under assumption of the background-only hypothesis. To set upper limits on the signal strength parameter, the agreement between the observed data and the signal+background hypothesis with a given μ is quantified by the p-value,

$$p_\mu = \int_{q_{\mu,\mathrm{obs}}}^{\mathrm{inf}} f(q_\mu|\mu) dq_\mu \tag{8.11}$$

where $f(q_\mu|\mu)$ is the probability density function of q_μ under the hypothesis μ.

The estimation of p_0 is performed for each signal scenario. If no significant deviation from the background-only hypothesis is observed, upper limits on the signal production cross-section is set by using the $\mathrm{CL_s}$, which is defined as $\mathrm{CL_s} = \mu/(1 - p_0)$. Values of the production that yield $\mathrm{CL_s} < 0.05$ are excluded at $\geq 95\%$ confidence level (CL) for a given signal.

8.8 Results

The profile likelihood for the background-only hypothesis is performed simultaneously in all signal regions. The post-fit distributions of signal LLH are shown in Fig. 8.12–8.14 for data, background and a hypothetical signal. The distributions show the number of events per width of 1.0 in the x-axis and the underflow and overflow are included in the first and last bins, respectively. In the following figures and tables, the contribution 'Others' refers to the combined single-top-quark and $t\bar{t} + \mathrm{X}$ backgrounds.

The post-fit event yields are summarized in Tab. 8.4 for twelve signal regions along with the predicted signal event yields before the fit for vector-like quark B with a mass of 1 TeV. It is observed that for some signal regions, the uncertainties of the individual background components can be larger than the uncertainty on the sum of the backgrounds due to correlations. The predicted and observed numbers of events in all signal regions before and after the fit are compared

Region	Multijet	$t\bar{t}$ + light	$t\bar{t}$ + HF	Others	Total background	$m_B = 1$ TeV $\mathcal{B}(B \to Hb) = 1$	$m_T = 1$ TeV $\mathcal{B}(T \to Ht) = 1$	Data
(VV,0t,2b)	5890 ± 190	380 ± 170	230 ± 90	92 ± 12	6590 ± 110	8.0 ± 1.0	3.5 ± 0.5	6614
(VV,0t,3b)	1300 ± 60	80 ± 40	130 ± 60	31 ± 8	1540 ± 40	11.5 ± 1.0	3.8 ± 0.6	1534
(VV,1t,2b)	680 ± 80	190 ± 90	130 ± 60	41 ± 11	1040 ± 90	2.2 ± 0.4	11.6 ± 1.4	1044
(VV,1t,3b)	190 ± 40	40 ± 26	130 ± 70	16 ± 5	380 ± 60	3.1 ± 0.4	7.4 ± 1.1	409
(VH,0t,2b)	7500 ± 400	1000 ± 500	500 ± 210	129 ± 15	9150 ± 340	23.4 ± 3.1	1.33 ± 0.33	9202
(VH,0t,3b)	3010 ± 180	310 ± 140	430 ± 200	76 ± 17	3820 ± 170	70 ± 6	6.2 ± 0.7	3778
(VH,1t,2b)	360 ± 60	160 ± 70	80 ± 40	28 ± 6	640 ± 50	3.9 ± 0.7	6.1 ± 0.8	623
(VH,1t,3b)	370 ± 50	100 ± 60	180 ± 80	19 ± 5	660 ± 90	18.2 ± 2.2	37.3 ± 3.3	662
(HH,0t,3b)	990 ± 110	180 ± 90	200 ± 100	19 ± 5	1390 ± 110	77 ± 6	38 ± 4	1407
(HH,1t,3b)	56 ± 13	8 ± 5	44 ± 24	6.4 ± 1.6	115 ± 16	17.1 ± 2.0	39 ± 4	113
(XX,2t,2b)	13 ± 4	8 ± 5	7 ± 5	0.7 ± 0.4	29 ± 7	0.17 ± 0.10	35 ± 4	30
(XX,2t,3b)	11 ± 7	3 ± 4	30 ± 19	2.0 ± 0.8	47 ± 21	2.4 ± 0.5	16.1 ± 2.3	51

Table 8.4: Event yields in twelve signal regions after the fit to data under the background-only hypothesis, as well as the predicted signal event yields before the fit for a B VLQ with a mass of 1 TeV. The uncertainties include statistical and systematic uncertainties.

in Fig. 8.15.

No significant deviation from the background-only hypothesis is observed and upper limits are set on the production cross-section of $T\bar{T}$ and $B\bar{B}$ events in different scenarios. Signal hypotheses are considered in a vector-like quark mass range between 500 and 1400 GeV.

Two benchmark scenarios are defined for the $T\bar{T}$ and $B\bar{B}$ productions. First benchmark is the case where each vector-like quark decays purely via the Higgs decay mode, $T\bar{T} \to HtH\bar{t}$ and $B\bar{B} \to HbH\bar{b}$. Second benchmark is the case where the $B\bar{B}$ production branching ratio is consistent with a weak-isospin doublet, (B, Y). Upper limits at the 95% confidence level (CL) on the $T\bar{T}$ and $B\bar{B}$ production cross-section are set for these two benchmark scenarios as a function of T (B) VLQ mass m_T (m_B) and compared to the theoretical predictions by Ref. [176] in Fig. 8.16.

It is observed that the analysis sensitivity becomes slightly degraded when considering signal samples with a significantly higher VLQ mass. This is due to the fact that in the calculation of the matrix elements, the masses of the vector-like B and T quarks are set to 900 GeV. The limits are found to be very strong for the $B \to Hb$ decay mode. The largest difference between the observed and expected limits is observed to be for $\mathcal{B}(B \to Hb) = 1$ with a vector-like quark mass around 950 GeV. This is due to a deficit in data in the last two bins of the (HH,0t,3b) signal region and the sensitivity of the matrix element calculation in this final state and vector-like quark mass. In the case of the exclusive $B \to Hb$ decay, the observed B mass exclusion limit is 1010 GeV where it is expected to be 970 GeV. In the case of the exclusive $T \to Ht$ decay, the observed and expected mass T exclusion limit are 1010 GeV.

Finally, all combinations of branching ratios ($\mathcal{B}$) are tested by reweighting the relative fractions of the three T and B decay modes: $T \to Wb, Ht, Zt$ and $B \to Wt, Hb, Zb$. The lower limit on the T mass as a function of $\mathcal{B}(T \to Ht)$ versus $\mathcal{B}(T \to Wb)$ and the lower limit on the B mass mass as a function of $\mathcal{B}(B \to Hb)$ versus $\mathcal{B}(B \to Wt)$ are shown in Fig. 8.17.

The statistical uncertainty of the data in the signal regions and the uncertainty which accounts
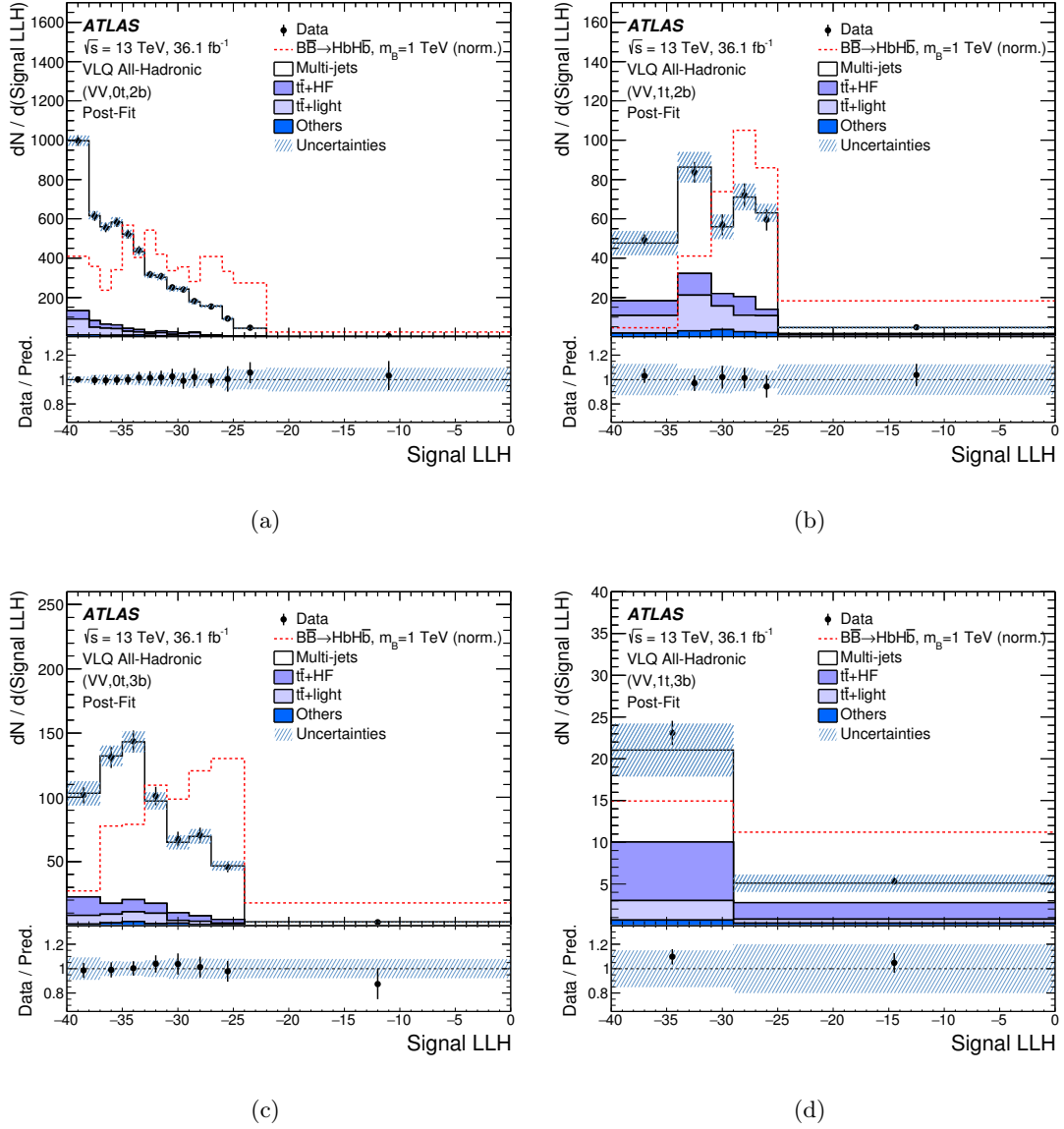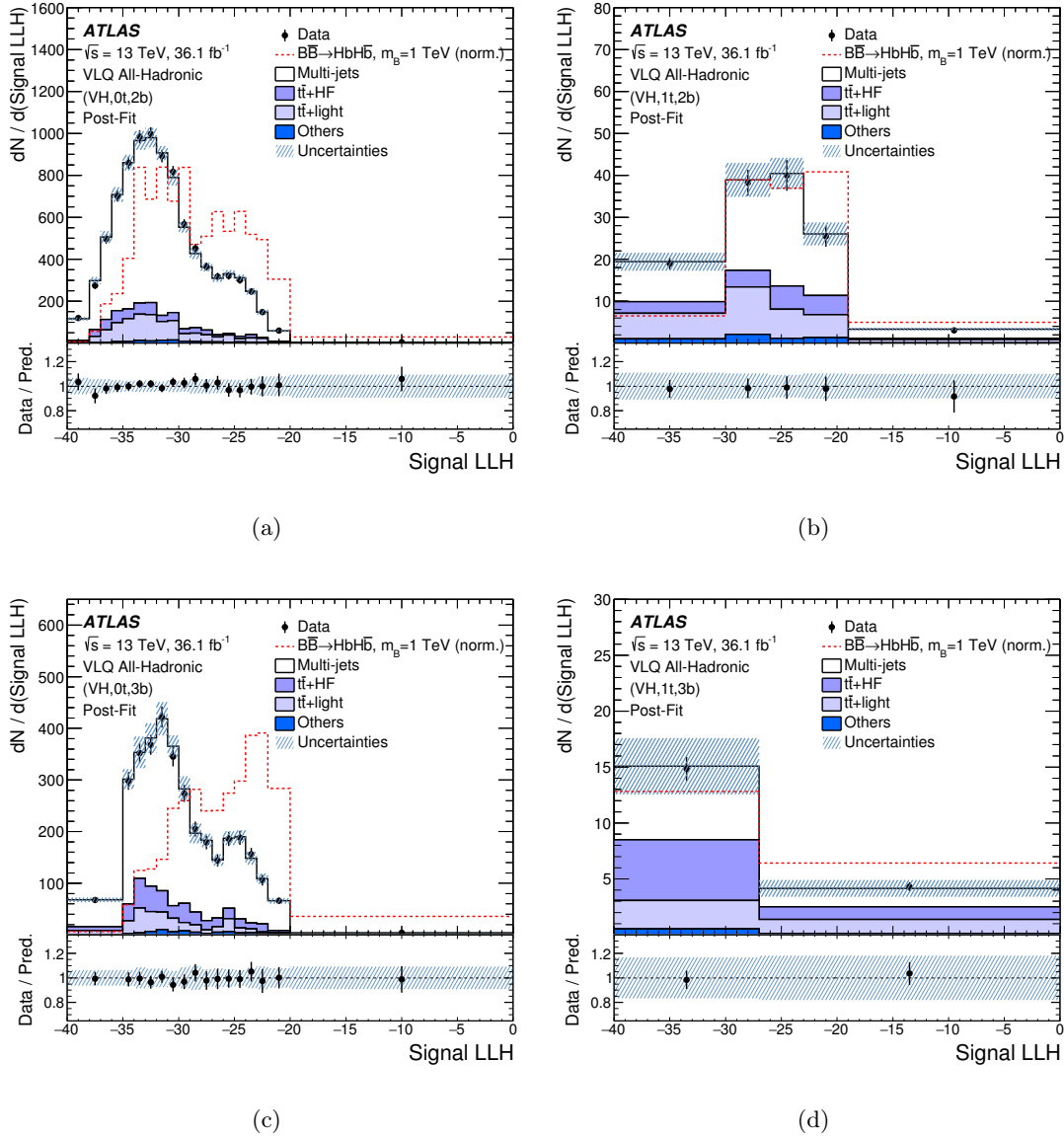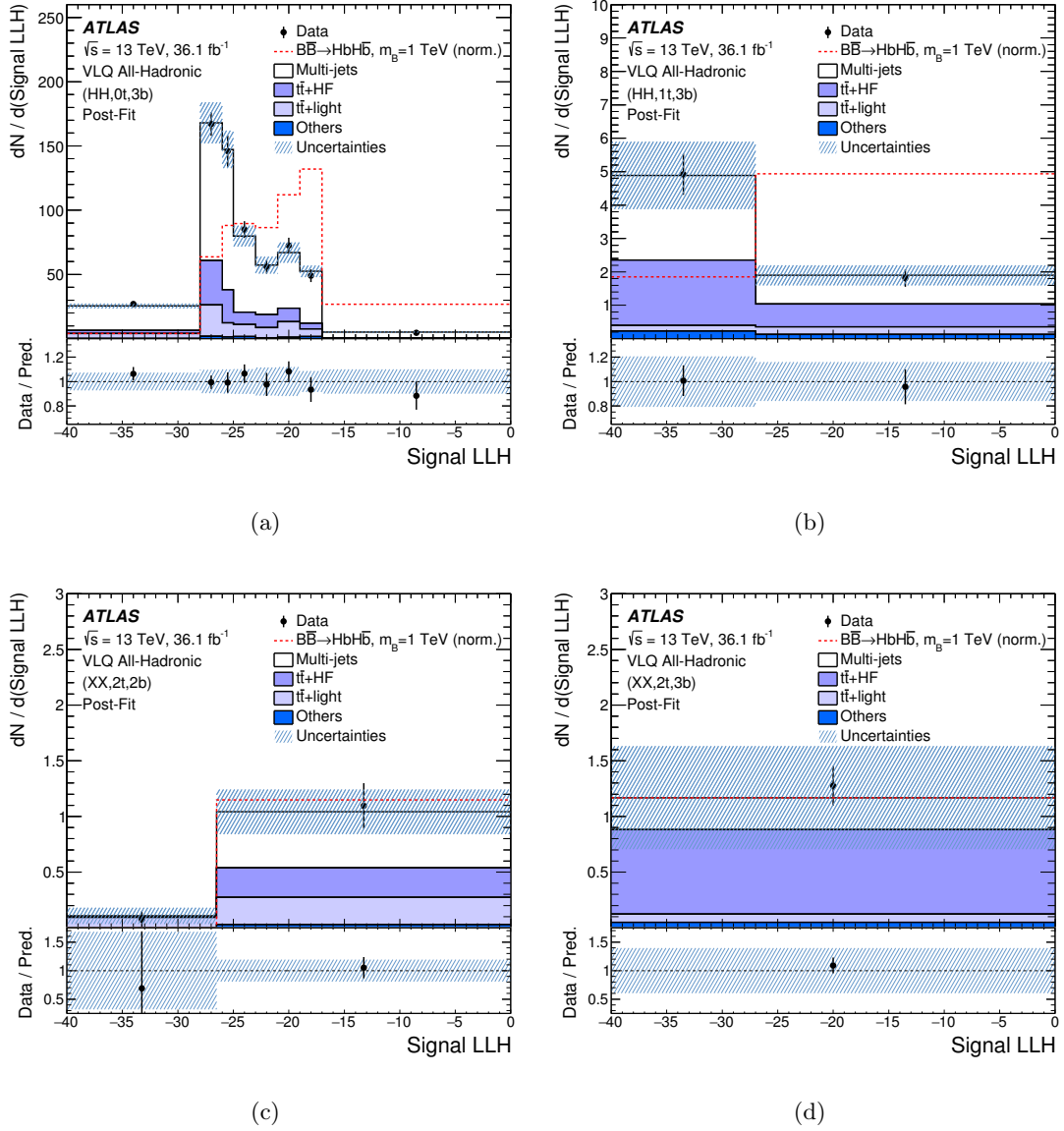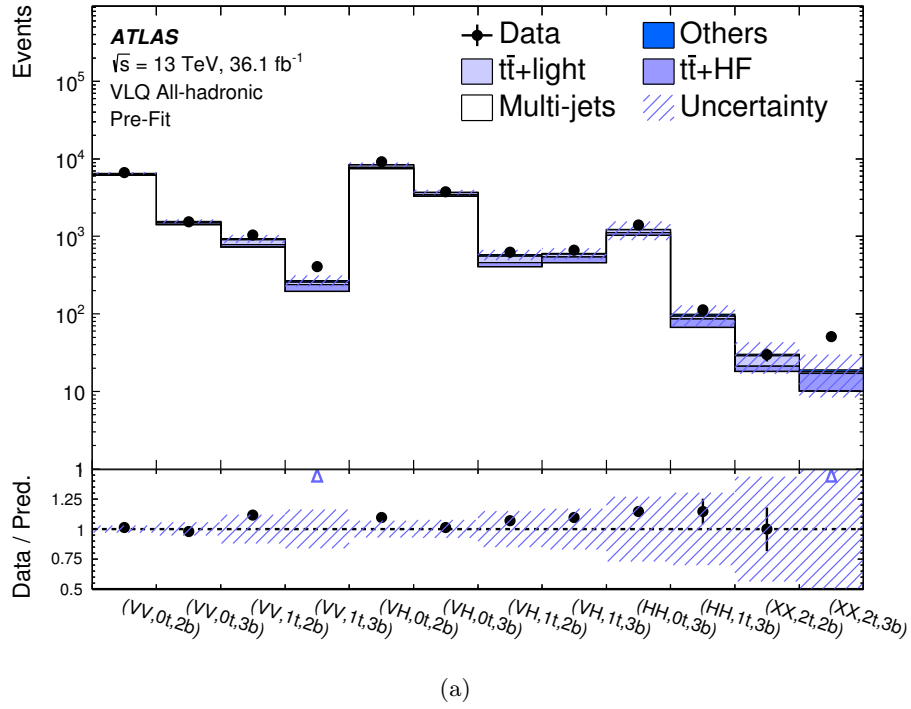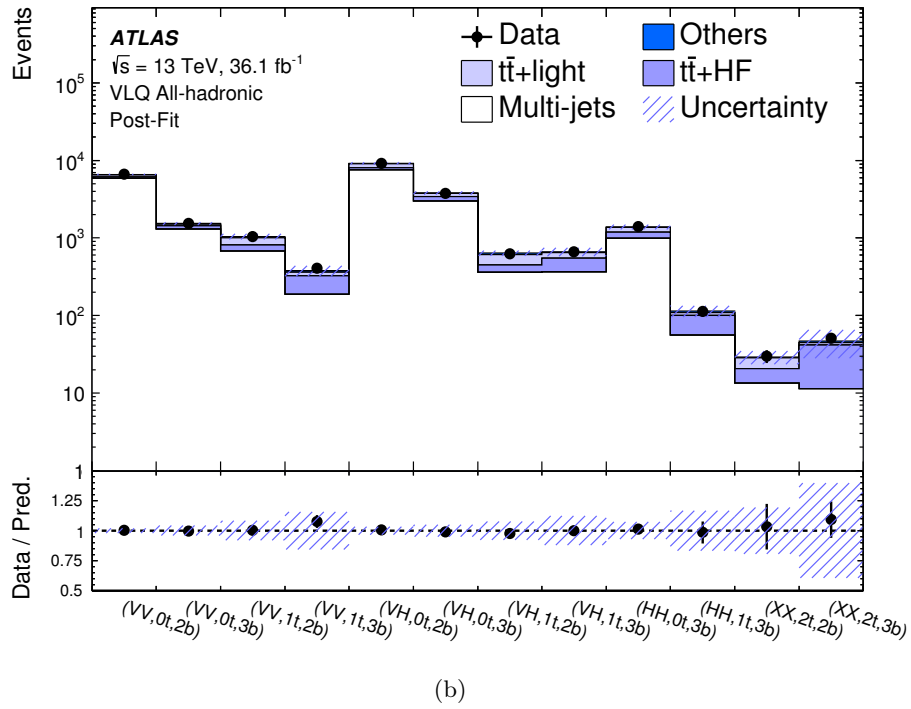
(a)

(b)

(c)

(d)

Figure 8.12: Comparison between data and prediction for the signal LLH distribution after the fit to the data under the background-only hypothesis for the (VV,0t,2b) (a), (VV,1t,2b) (b), (VV,0t,3b) (c) and (VV,1t,3b) (d) signal regions. The hatched area represents the total uncertainty of the background. A hypothetical signal for $\mathcal{B}(B \to Hb) = 100\%$ and $m_B = 1$ TeV is shown overlaid, normalized to the integral of the total background.

(a)

(b)

(c)

(d)

Figure 8.13: Comparison between data and prediction for the signal LLH distribution after the fit to the data under the background-only hypothesis for the (VH,0t,2b)(a), (VH,1t,2b)(b), (VH,0t,3b) (c) and (VH,1t,3b) (d) signal regions. The hatched area represents the total uncertainty of the background. A hypothetical signal for $\mathcal{B}(B \to Hb) = 100\%$ and $m_B = 1$ TeV is shown overlaid, normalized to the integral of the total background.

Figure 8.14: Comparison between data and prediction for the signal LLH distribution after the fit to the data under the background-only hypothesis for the (HH,0t,3b)(a), (HH,1t,3b)(b), (XX,2t,2b) (c) and (XX,2t,3b) (d) signal regions. The hatched area represents the total uncertainty of the background. A hypothetical signal for $\mathcal{B}(B \to Hb) = 100\%$ and $m_B = 1$ TeV is shown overlaid, normalized to the integral of the total background.

(a)

(b)

Figure 8.15: Comparison between data and prediction for the event yields before (a) and after (b) the fit to the data under the background-only hypothesis. The figures show the total normalization across all signal regions. The hatched area represents the total uncertainty of the background.
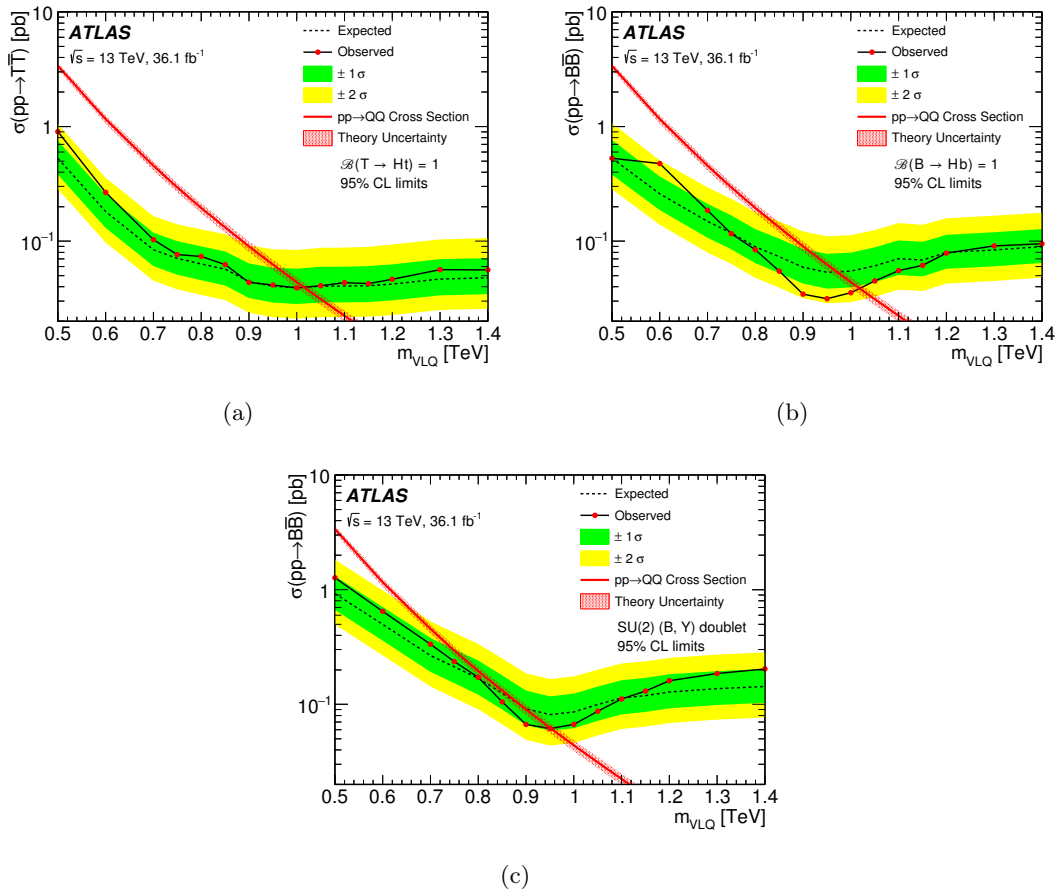
Figure 8.16: Observed (solid) and expected (dashed) upper limits at the 95% CL on the $T\bar{T}$ (a) and $B\bar{B}$ (b) cross-section as a function of the VLQ mass assuming $\mathcal{B}(T \to Ht) = 1$ and $\mathcal{B}(B \to Hb) = 1$, respectively, and on the $B\bar{B}$ (c) cross-section with the assumption of branching ratios consistent with a weak-isospin doublet. In the doublet case, only contributions from the B VLQ are considered. The green and yellow bands correspond to ± 1 and ± 2 standard deviations around the expected limit. The thin red line and band show the theoretical prediction and uncertainties.

for statistical uncertainties in the control regions propagated through the ABCD method are observed to have the largest impact on the sensitivity of the analysis. The dominant systematic uncertainties are observed to be the multijet closure, $t\bar{t}$ modeling, small-R JER and JES, and $E_{\mathrm{T}}^{\mathrm{miss}}$ uncertainties.
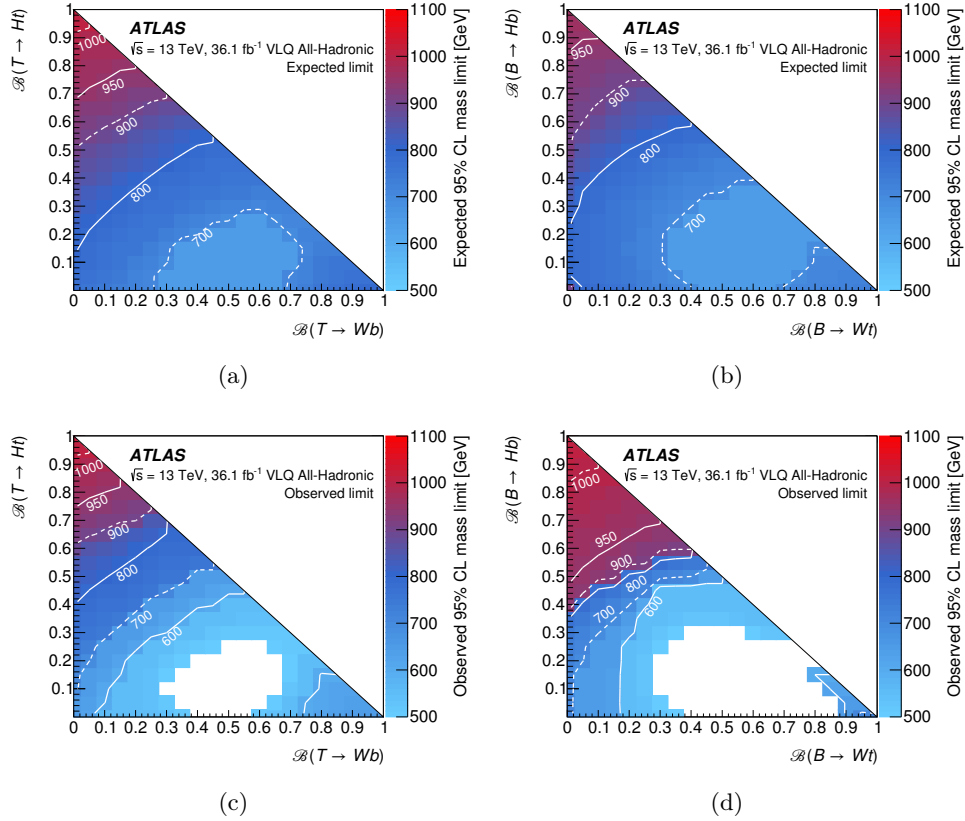
Figure 8.17: Expected and observed 95% CL lower limits on m_T (a,c) and m_B (b,d) in the branching-ratio planes. Contour lines indicate sensitivity to different VLQ masses across the planes. The white space on the observed limit figures corresponds to branching ratios where there is no observed exclusion above a mass of 500 GeV.
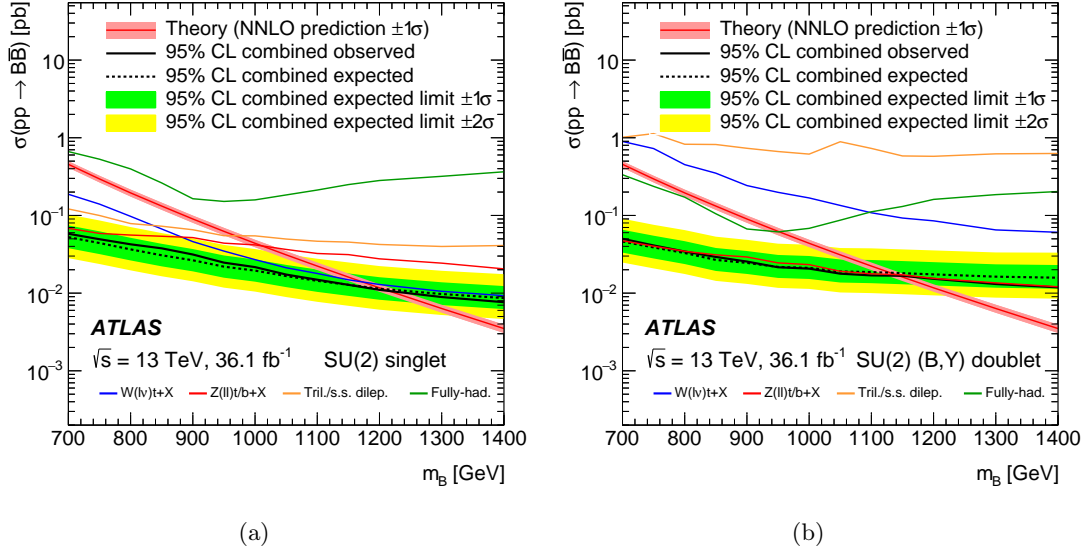
Figure 8.18: Observed (solid) and expected (dashed) upper limits at the 95% CL on the $B\bar{B}$ cross-section versus mass for the combination and the individual analyses in black and colored lines, respectively. The singlet (a), and (B, Y) doublet (b) scenarios are shown. The solid green lines represents the search in fully-hadronic final states. The green and yellow bands correspond to ± 1 and ± 2 standard deviations around the expected limit. The rapidly falling thin red line and band show the theory prediction and corresponding uncertainty [176], respectively.

8.9 Combination of pair-produced vector-like quark searches

Complementary searches for pair-produced vector-like quarks have been performed in ATLAS at $\sqrt{s} = 13$ TeV using data collected in 2015 and 2016 corresponding to an integrated luminosity of 36.1 fb^{-1}. These searches observed no significant deviation from the SM background predictions. Hence, the limits of these searches are combined to set limits on the masses of the vector-like T and B as presented in Ref. [4].

The search in fully-hadronic final states (fully-hadronic or fully-had.) presented in this chapter also contributed to the combination. It is the only contributing analysis with significant sensitivity to $B\bar{B} \to HbH\bar{b}$. In Fig. 8.18, observed and expected upper limits at the 95% CL on the $B\bar{B}$ cross-section as a function of mass, where the B singlet and (B, Y) doublet scenarios are considered, are presented for the combination and the individual analyses.

Similar to what was done in the fully-hadronic analysis, all combinations of branching ratios are tested assuming the three T and B decay modes: $T \to Wb, Ht, Zt$ and $B \to Wt, Hb, Zb$. The resulting lower limits on the VLQ mass as a function of branching ratio are presented in Figure 8.19. For any combination of branching ratios, the observed (expected) lower mass limits are 1.31 (1.22) TeV for T and 1.03 (0.98) TeV for B. These are the strongest limits on the vector-like T and B masses where for the first time vector-like T with masses below 1.31 TeV and vector-like B with masses below 1.03 are excluded for decays into SM particles.
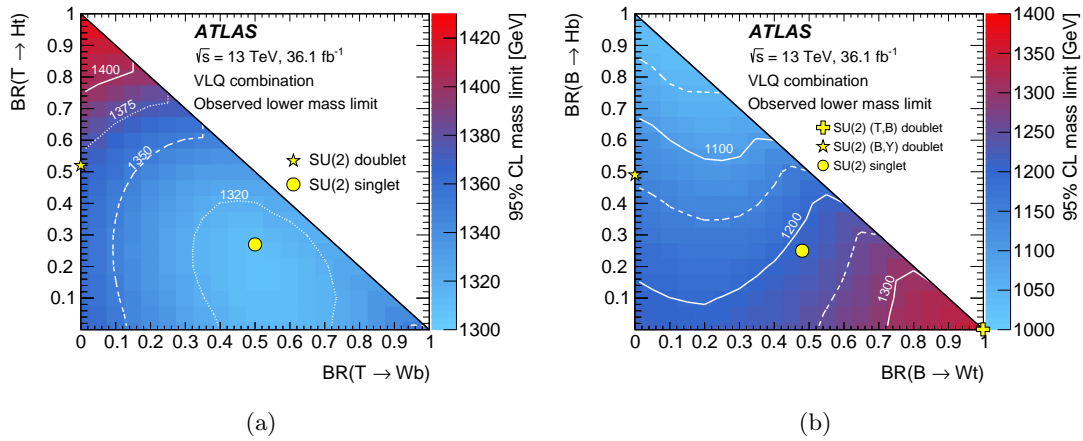
Figure 8.19: Observed 95% CL lower limits on m_T (a) and m_B (b) in the branching-ratio planes. The yellow markers indicate the branching ratios for the SU(2) singlet and doublet scenarios where the branching ratios become approximately independent of the VLQ mass.

8.10 Outlook

In order to obtain a stronger sensitivity, some aspects of the analysis, such as the vRC jet tagger and calculation of the analysis discriminant (signal likelihood), could be improved. Some possible improvements are listed in this section.

- vRC jet tagging:

  In this analysis, subjets of vRC jets are use as inputs to the DNN. Although the number of subjets vary from jet to jet, the employed DNN does not allow for variable length and uses default values for the missing subjet properties. Instead of using feedforward fully-connected DNNs with fixed number of inputs, employing NNs like recurrent neural networks (RNN) which allow variable length inputs would be expected to improve the performance of the tagger. Another improvement could be changing the jet collection used to build the vRC jets. Using smaller radius jets or track-jets could improve the performance. Finally, although the pseudo-continuous b-tagging working point was observed to provide better performance than the fixed-efficiency working point employed in other parts of the analysis, it introduced additional systematic uncertainties in the analysis. In the future it would be beneficial to simply the choice.

- Final discriminant:

  The masses of the vector-like B and T quarks were set to 900 GeV in the calculation of the matrix element. Although this choice simplified the procedure, it resulted in sensitivity loss for VLQs with significantly high mass. This degrading in the analysis sensitivity could be overcome by employing a varying vector-like B and T mass.

Chapter 9

Conclusions

In this thesis, two main topics were discussed: the performance of large-R jet tagging algorithms, with an emphasis on the machine learning-based taggers, and the application of machine learning-based taggers within the context of a vector-like quark pair production search in hadronic final states.

Different methods to tag boosted, hadronically decaying W bosons and top quarks were studied in data and simulation. The simulation-based studies showed that traditional tagging methods such as two substructure variable combinations and shower deconstruction provide good signal and background discrimination. In the attempt of combining complementary information available in various jet moments or extracting information from the jet constituents, different machine learning-based taggers were introduced. It was observed that all machine learning-based taggers, regardless of the type of inputs they use, provide stronger discrimination power compared to the traditional methods. Moreover, it was observed that using jet constituents as inputs to the deep neural networks rather than the jet moments lead to performance improvements for high-p_{T} jets, where the substructure variables begin to lose their discrimination power. The performance of the taggers were then validated using the data recorded by the ATLAS experiment in 2015 and 2016 corresponding to an integrated luminosity of 36.1 fb^{-1}. The modeling of tagging discriminants were studied by comparing the discriminant distributions in signal-enriched (lepton+jet $t\bar{t}$) and background-enriched (dijet and γ + jet) data samples with predictions from the Monte Carlo simulation. Finally, the tagging signal efficiency and the background rejection in data were measured as a function of pileup and jet p_{T} using boosted lepton+jet $t\bar{t}$, dijet and γ + jet topologies. It was observed that the tagging discriminants, tagger signal efficiency and background rejection were overall well-modeled by Monte Carlo simulation. These studies demonstrated the power and robustness of the machine-learning based techniques. Moreover, they demonstrate the first steps taken towards extracting tagging efficiencies and relevant uncertainties for both signal and background from data.

A search for pair production of vector-like quarks, $T\bar{T}$ and $B\bar{B}$, in the all-hadronic final state was presented using the data recorded by the ATLAS experiment in 2015 and 2016, corresponding to an integrated luminosity of 36.1 fb^{-1}. In this analysis, pair produced vector-like quarks were assumed to decay into a Standard Model boson and a third-generation quark, leading to $T \rightarrow Wb, Ht, Zt$ or $B \rightarrow Wt, Hb, Zb$ decay modes. In order to obtain a dataset enriched in signal,

high-p_T small-R jets and multiple b-tagged small-R jets were required. Variable-R reclustered jets were built by using the small-R jets and a multi-class deep neural network was used to identify them as W/Z, Higgs boson, top quark, or background jets. The events were divided into twelve signal regions based on the number of W/Z, Higgs boson, top quark variable-R reclustered jets and b-tagged small-R jets. A signal log-likelihood was calculated via the matrix element method and used as the final discriminant in each signal region. In each signal region, multijet contribution is estimated using the bin-by-bin $ABCD$ method, a data-driven method. No significant deviation from the Standard Model expectation was observed. Upper limits are set on the production cross-section of $T\bar{T}$ and $B\bar{B}$ events in different scenarios. Although the analysis was optimized considering all the decay modes, the limits were found to be strongest for the $B \to Hb$ decay mode. In the case of the exclusive $B \to Hb$ decay, the observed B mass exclusion limit was found to be 1010 GeV and in the case of the exclusive $T \to Ht$ decay, the observed T mass exclusion limit was found to be 1010 GeV at the 95% CL.

Appendix A

Jet substructure moments

Some of the substructure moments which are widely used for W-boson and top-quark tagging in ATLAS were presented in section 6.4, further moments are defined below.

Fox–Wolfram moment ($R_2^{\mathbf{FW}}$)   The Fox–Wolfram moments have been first proposed to characterize the shapes of QCD events [91]. The moments are also used to study the shapes of jets and it was shown that the ratio of the second-order to zeroth-order Fox–Wolfram moment (R_2^{FW}) can be powerful to tag the hadronically-decaying W bosons [92]. The Fox–Wolfram moments are defined as

$$H_l = \sum_{i,j \epsilon J} \frac{|\vec{p_i}||\vec{p_j}|}{E^2} P_l(\cos \theta_{ij}) \tag{A.1}$$

for a given jet J where sum runs over all constituents, $\vec{p_i}$ and $\vec{p_j}$ are the momenta of constituents i and j, θ_{ij} is the distance between those constituents, E is the total energy of all the constituents in the jet rest frame and $P_l(x)$ are the Legendre polynomials.

Aplanarity (A)   Similar to the R_2^{FW}, the aplanarity is also used to characterize the shape of the jet and it is calculated using the constituents of a jet in its rest frame [92, 201]. It is defined as

$$A = \frac{3\lambda_3}{2} \tag{A.2}$$

where λ_3 is an eigenvalue of the sphericity tensor.

Planar flow ($\mathcal{P}$)   The planar flow moment measures the distribution of jet constituents' energies with respect to the jet axis and quantifies to which extent the energy flow inside the jet is linear or planar [95, 96]. First, the shape tensor I_E^{kl} is reconstructed as

$$I_w^{kl} = \frac{1}{m} \sum_{i \epsilon J} \frac{p_{i,k} p_{i,l}}{E_i} \tag{A.3}$$

for a given jet J where sum runs over all constituents, m is the jet mass, E_i is the energy of jet constituent i and pi, k (pi, l) is the k-th (l-th) component of its transverse momentum relative

to the jet axis. Planar flow ($\mathcal{P}$) is then defined as

$$\mathcal{P} = \frac{4\det(I_E)}{\text{tr}(I_E)^2}. \tag{A.4}$$

Angularity (a_3) Angularities are a class of jet shape variables which measure the extent of symmetry in the energy flow inside a jet [96, 202]. The angularity a_3 is used for W-boson tagging and it is defined as

$$a_3 = \frac{1}{m} \sum_{i \epsilon J} E_i \sin^\alpha \left(\theta_i\right) \left[1 - \cos\left(\theta_i\right)\right]^{1-\alpha} \tag{A.5}$$

for a given jet J where sum runs over all constituents, m is the jet mass, θ_i is the angle between the constituent i and the jet axis and α is a parameter that can be chosen to emphasize radiation near the edges or core of the jet. $\alpha = -2$ is used.

Splitting scale z_{cut} Splitting scales were presented in section 6.4.4. The mass(m)-normalized form of the k_t splitting scale substructure variable $\sqrt{d_{12}}$ [93] is referred to as z_{cut} in this thesis. It is defined as

$$z_{\text{cut}} = \frac{d_{12}}{d_{12} + m^2}. \tag{A.6}$$

KtDR *KtDR* is the angular distance between entities i and j in the last step of the k_t algorithm. Similar to the previously presented k_t splitting scale variables, it is determined by reclustering the constituents of the trimmed large-R jet with the k_t algorithm. It is defined in the pseudorapidity–azimuthal space as:

$$KtDR = \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2}. \tag{A.7}$$

Q_w A W boson is expected inside a top-quark jet and the Q_w variable investigates the presence of a W boson inside a large-radius jet for top-quark tagging. The variable is defined by using the k_t algorithm to identify the $M \to ABC$ splitting with three subjets and finding the minimum pair-wise invariant mass between three subjets A, B, and C.

Appendix B

Jet moment correlations

A visual representation of correlations between jet moments, the DNN, BDT W, top discriminants based on the jet moments described in chapter 7 and average number of interactions per bunch crossing (μ). These are shown in logarithmic scale for the event density.
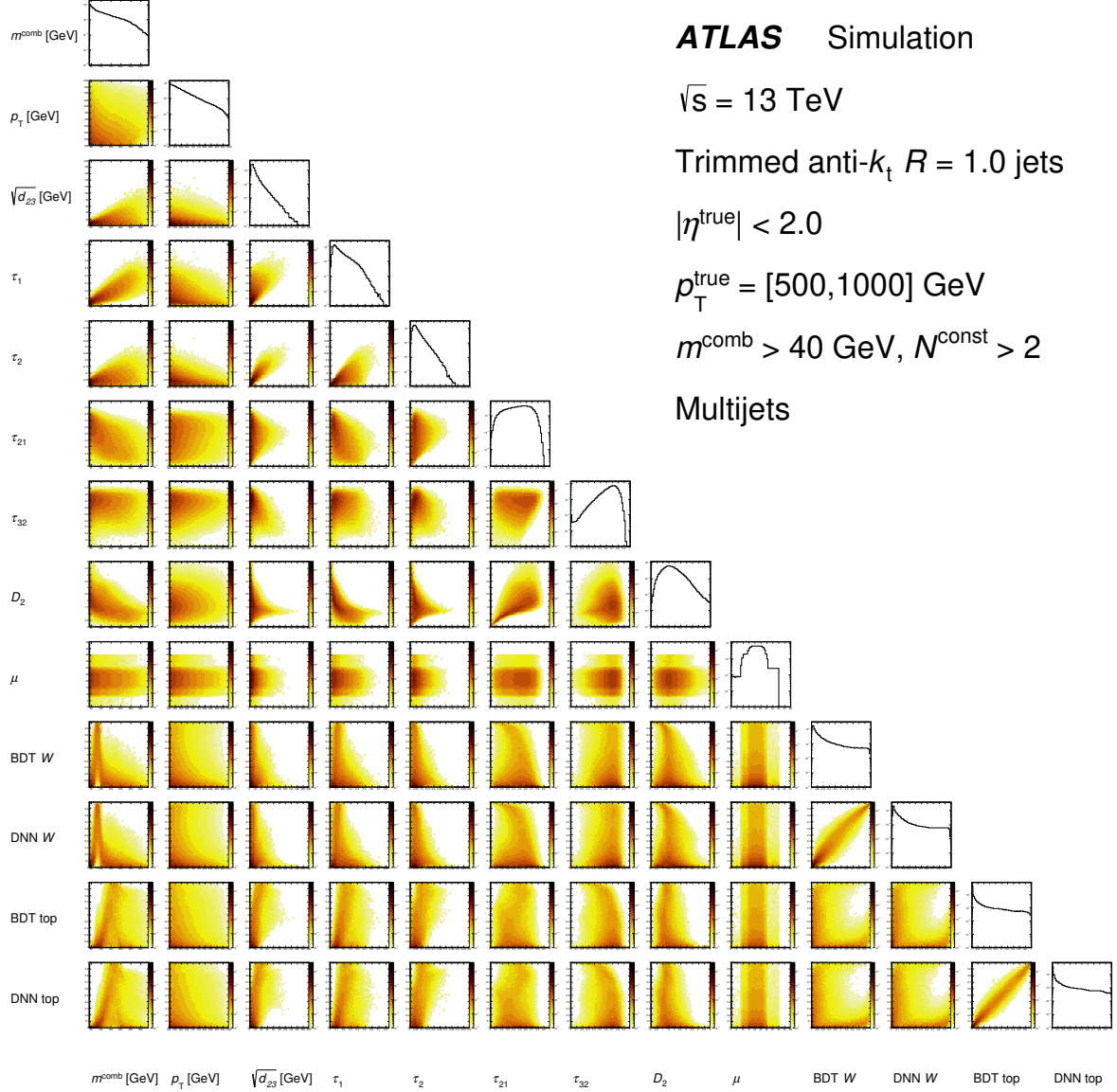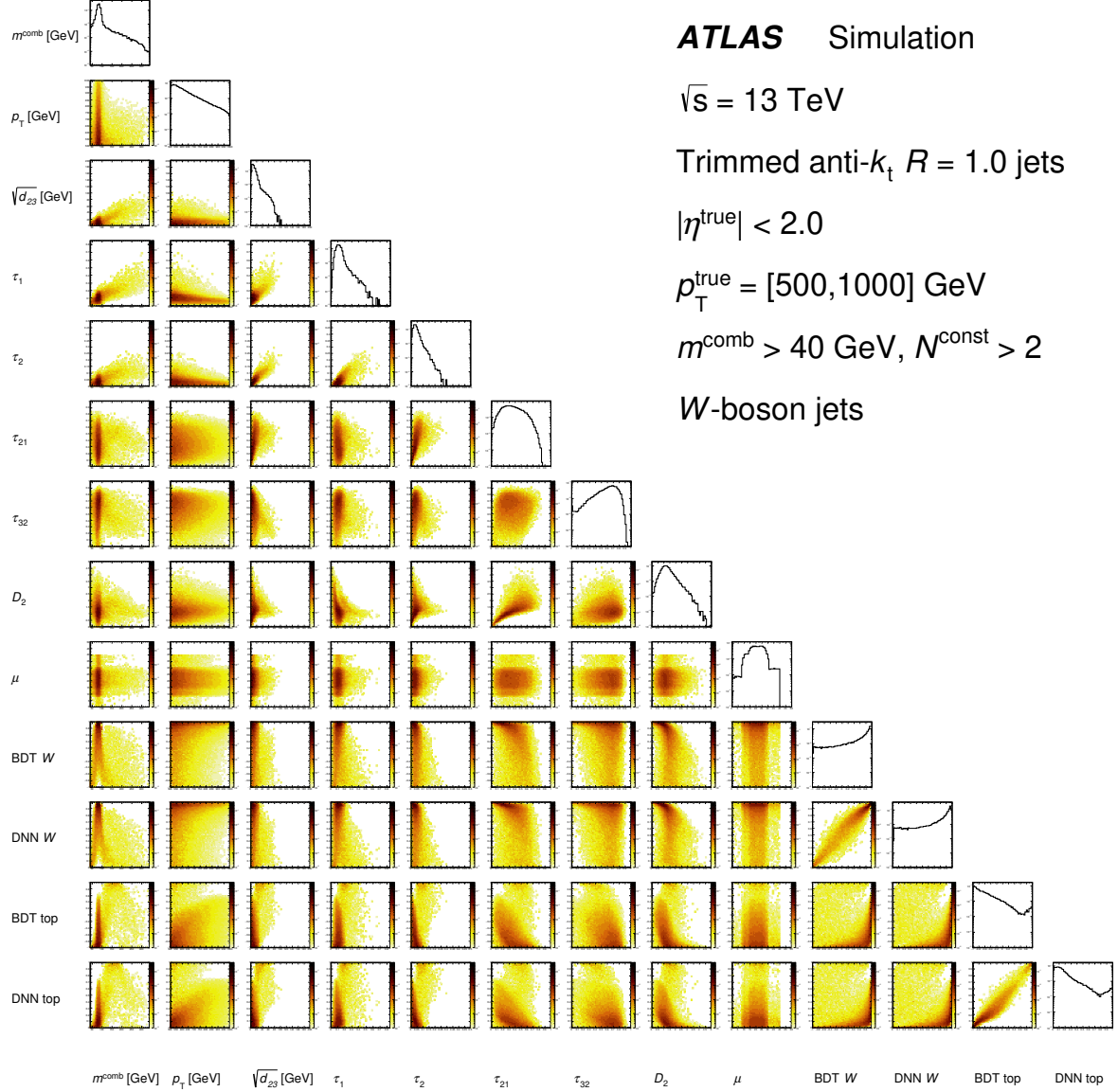
Figure B.1: A visualization of the two-dimensional correlation between a subset of the input observables used in the W boson and top quark multivariate jet moment based taggers in an inclusive sample of light jets in the $p_{\mathrm{T}}^{\mathrm{true}}$ range from 500 GeV to 1000 GeV. The event density is shown in logarithmic scale for the two-dimensional event density.
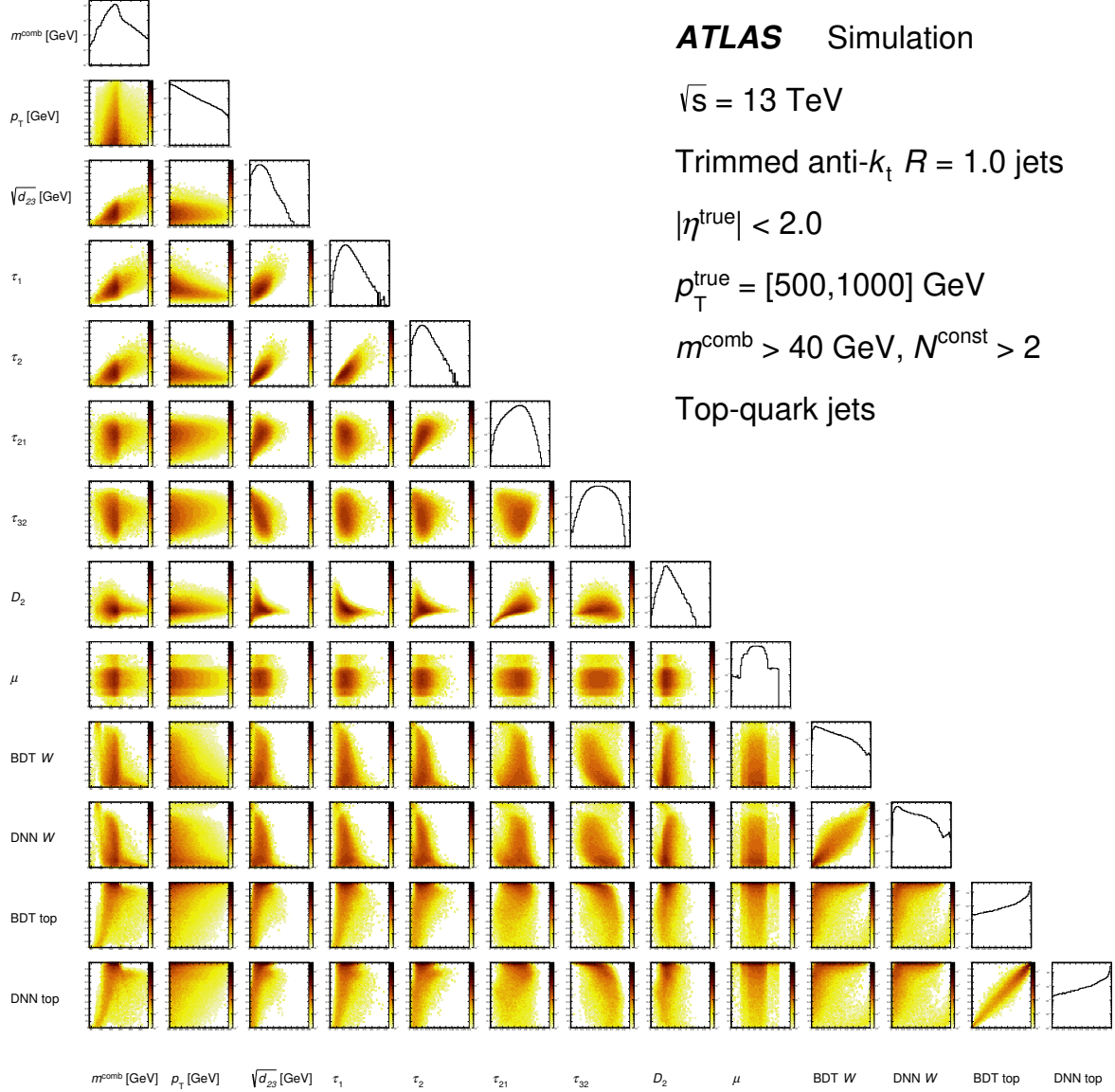
Figure B.2: A visualization of the two-dimensional correlation between a subset of the input observables used in the W boson and top quark multivariate jet moment based taggers in an inclusive sample of W boson jets in the $p_{\mathrm{T}}^{\mathrm{true}}$ range from 500 GeV to 1000 GeV. The event density is shown in logarithmic scale for the two-dimensional event density.

Figure B.3: A visualization of the two-dimensional correlation between a subset of the input observables used in the W boson and top quark multivariate jet moment based taggers in an inclusive sample of fully contained top quark jets in the p_T^{true} range from 500 GeV to 1000 GeV. The event density is shown in logarithmic scale for the two-dimensional event density.

Appendix C

vRC jet tagging peak structures

In order to understand the multi-peak behavior of the VLQ DNN tagging discriminants (P(V), P(H), P(t)) in section 8.4.2, the correlations between these three vrcjet tagging discriminants and some of the powerful input variables are visualized in terms of two-dimensional distributions in Figs. C.1,C.2,C.3,C.4. The correlations between the peaks in P discriminants and inputs for three signals and background. It is observed that the multi-peak behavior in some of the discriminants is caused by the variations in important features, such as the mass and p_T of the vRC jet or the number of b-tagged and non-b-tagged constituent small-R jets, that can cause one type to mimic another type. For example for top-quark jets, one can see that the smaller peak is associated to the low-p_T and low-mass top-quark jets which mimic the W bosons. Another example is the Higgs-boson jets which have either only one constituent or have low b-tagging discriminant score, these jets can mimic the V bosons and also the background.
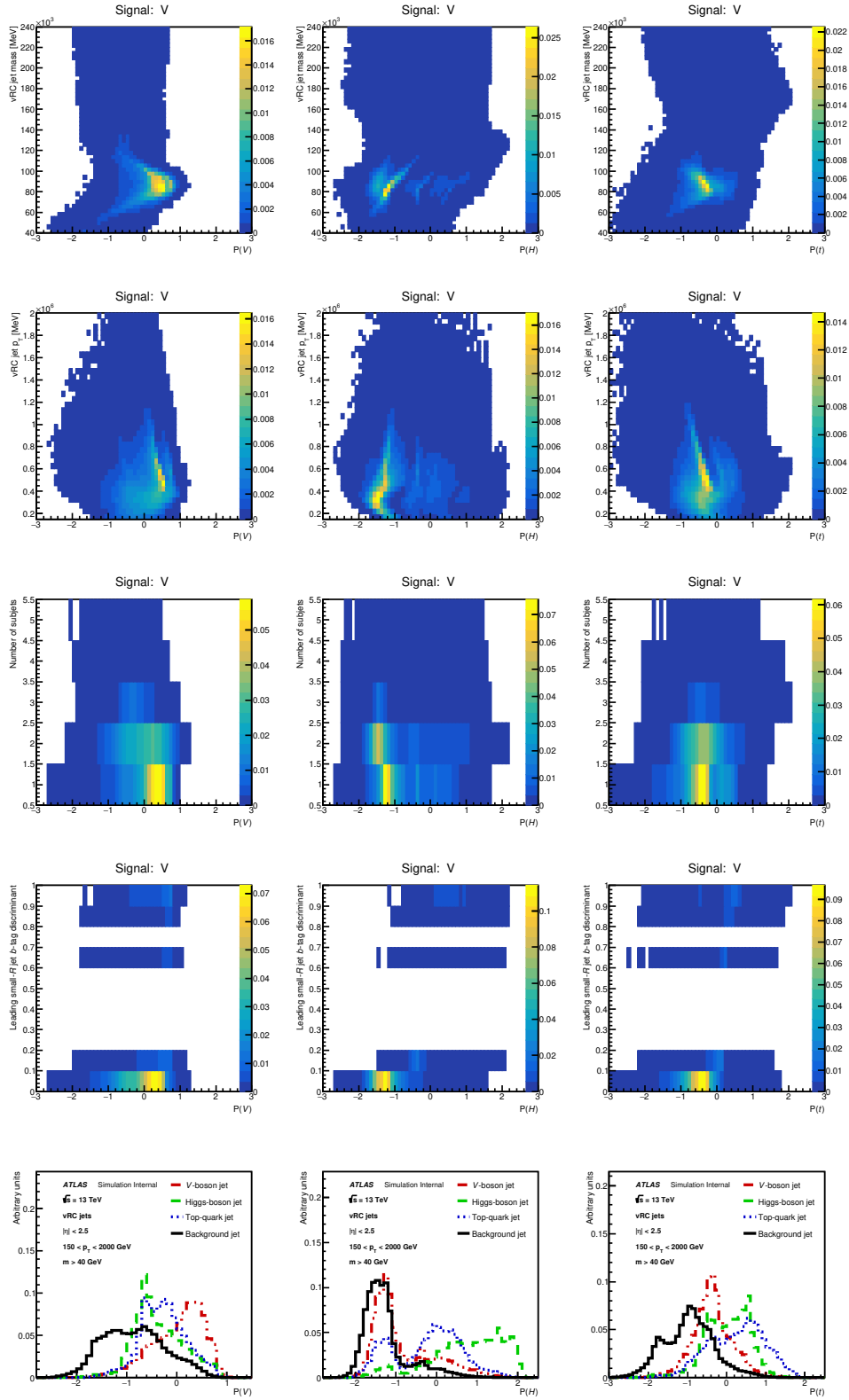
Figure C.1: Two-dimensional distributions of vrcjet tagging discriminants and some input variables for *V*-boson jets.
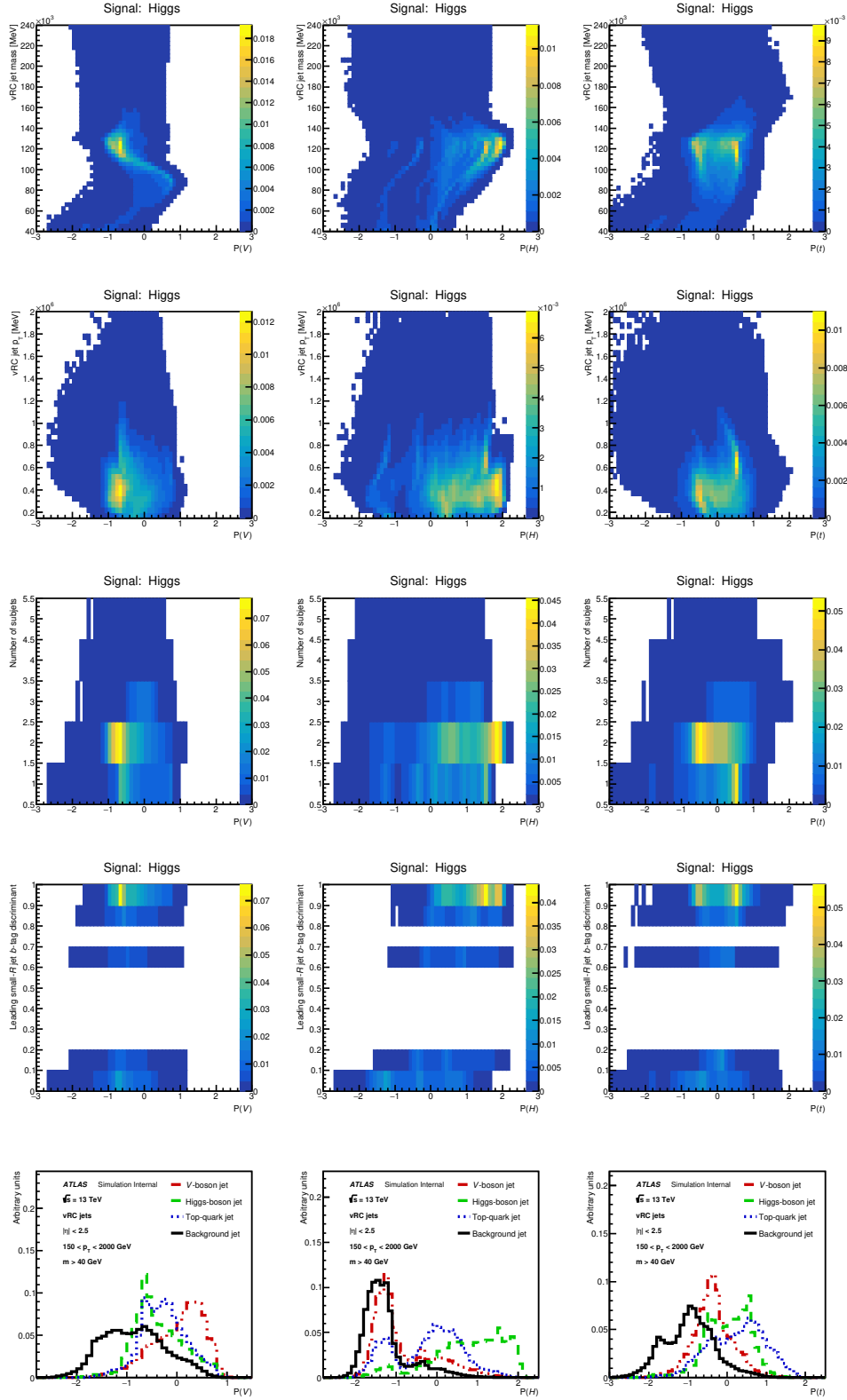
Figure C.2: Two-dimensional distributions of vrcjet tagging discriminants and some input variables for Higgs-boson jets.
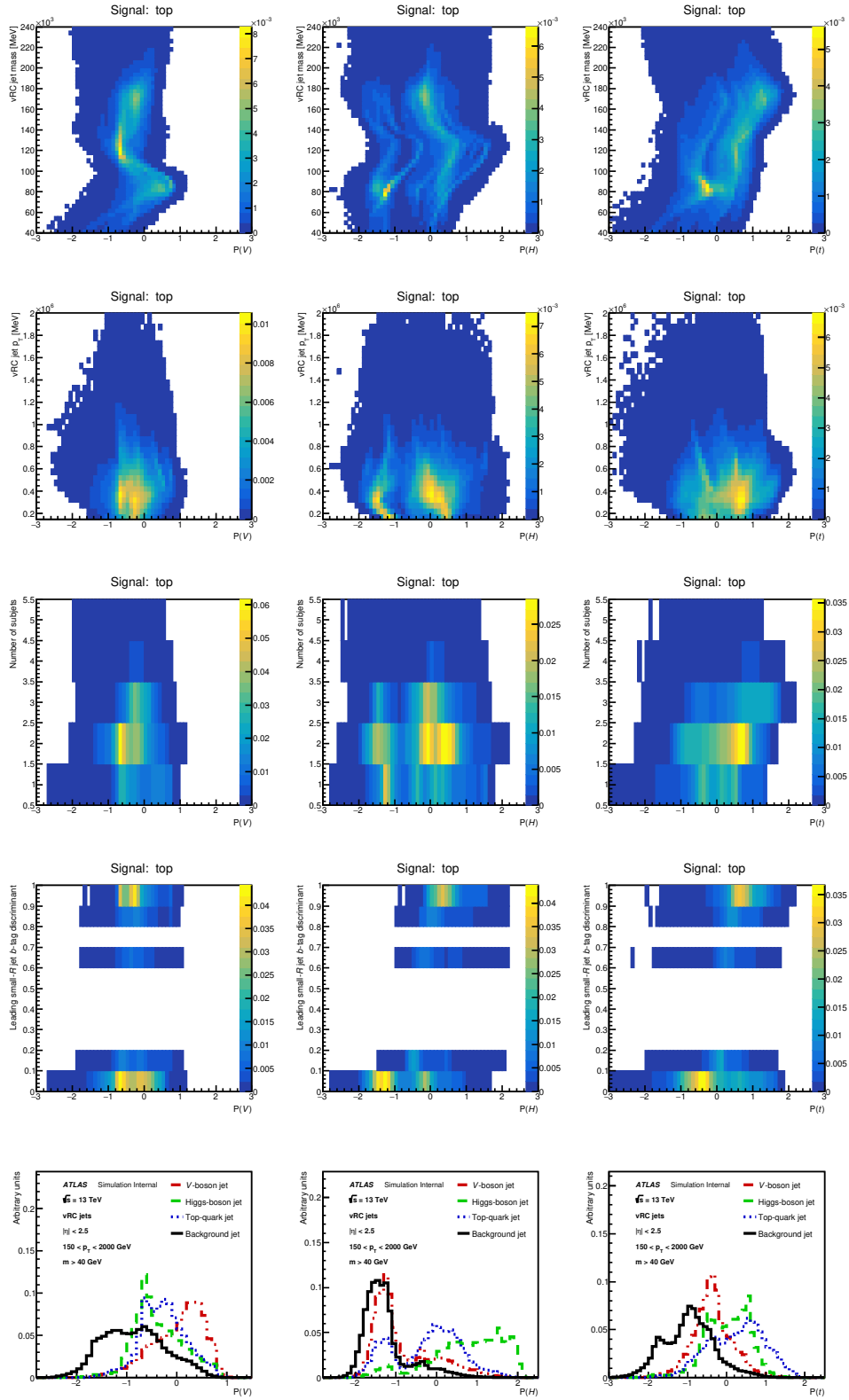
Figure C.3:   Two-dimensional distributions of vrcjet tagging discriminants and some input variables for top-quark jets.
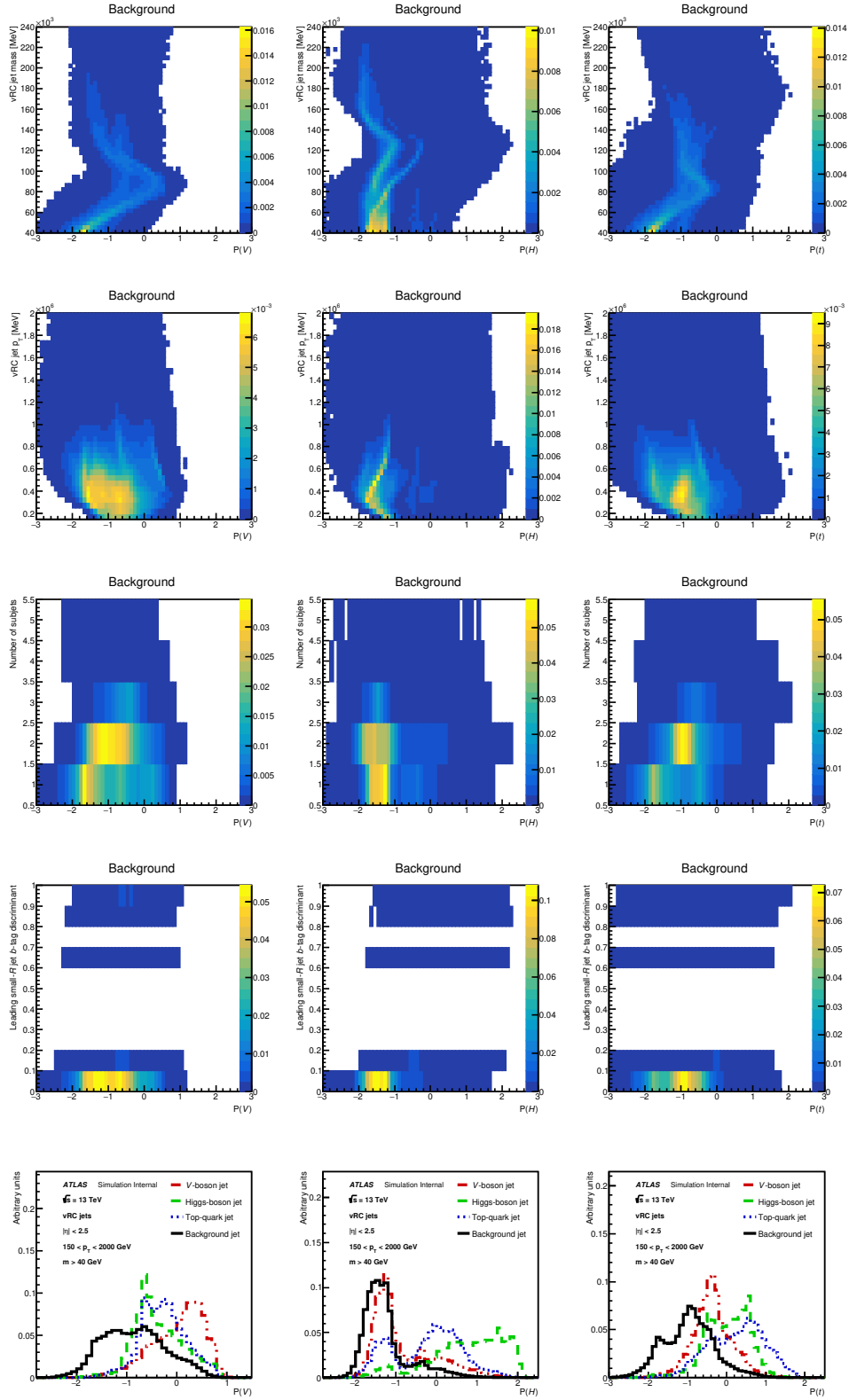
Figure C.4: Two-dimensional distributions of vrcjet tagging discriminants and some input variables for background jets.

Bibliography

[1] ATLAS Collaboration. Identification of Hadronically-Decaying W Bosons and Top Quarks Using High-Level Features as Input to Boosted Decision Trees and Deep Neural Networks in ATLAS at $\sqrt{s} = 13$ TeV. ATL-PHYS-PUB-2017-004, 2017.

[2] ATLAS Collaboration. Performance of top-quark and W-boson tagging with ATLAS in Run 2 of the LHC. *Eur. Phys. J.*, C79(5):375, 2019.

[3] ATLAS Collaboration. Search for pair production of heavy vector-like quarks decaying into hadronic final states in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev. D*, 98(9):092005, 2018.

[4] ATLAS Collaboration. Combination of the searches for pair-produced vector-like partners of the third-generation quarks at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev. Lett.*, 121(21):211801, 2018.

[5] Mark Thomson. *Modern particle physics*. Cambridge University Press, Cambridge, Sep 2013.

[6] B R Martin and G Shaw. *Particle Physics; 4th ed.* The Manchester Physics Series. Wiley, New York, NY, 2017.

[7] Andrew Purcell. Go on a particle quest at the first CERN webfest. Le premier webfest du CERN se lance à la conquête des particules. (BUL-NA-2012-269. 35/2012):10, Aug 2012.

[8] Emmy Noether and M. A. Tavel. Invariant variation problems, 2005.

[9] Peter W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13:508–509, 1964. [,160(1964)].

[10] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.

[11] Nicola Cabibbo. Unitary symmetry and leptonic decays. *Phys. Rev. Lett.*, 10:531–533, Jun 1963.

[12] M. et al Tanabashi. Review of particle physics. *Phys. Rev. D*, 98:030001, Aug 2018.

[13] P. Bagnaia et al. Evidence for Z0 —> e+ e- at the CERN anti-p p Collider. *Phys. Lett.*, 129B:130–140, 1983. [,7.69(1983)].

[14] M. Banner et al. Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider. *Phys. Lett.*, 122B:476–485, 1983. [,7.45(1983)].

[15] G. Arnison et al. Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c**2 at the CERN SPS Collider. *Phys. Lett.*, 126B:398–410, 1983. [,7.55(1983)].

[16] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716:1, 2012.

[17] Serguei Chatrchyan et al. Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC. *Phys. Lett.*, B716:30–61, 2012.

[18] F. Abe et al. Observation of top quark production in $\bar{p}p$ collisions. *Phys. Rev. Lett.*, 74:2626–2631, 1995.

[19] S. Abachi et al. Observation of the top quark. *Phys. Rev. Lett.*, 74:2632–2637, 1995.

[20] V. C. Rubin and W. K. Ford, Jr. Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions. *Astrophys.J.*, 159:379, February 1970.

[21] Y. Akrami et al. Planck 2018 results. I. Overview and the cosmological legacy of Planck. 2018.

[22] Y. et al Fukuda. Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.*, 81:1562–1567, Aug 1998.

[23] Martin Schmaltz and David Tucker-Smith. Little Higgs review. *Ann. Rev. Nucl. Part. Sci.*, 55:229–270, 2005.

[24] Michael J. Dugan, Howard Georgi, and David B. Kaplan. Anatomy of a Composite Higgs Model. *Nucl. Phys.*, B254:299–326, 1985.

[25] Natascia Vignaroli. Early discovery of top partners and test of the Higgs nature. *Phys. Rev.*, D86:075017, 2012.

[26] Melissa van Beekveld, Sascha Caron, and Roberto Ruiz de Austri. The current status of fine-tuning in supersymmetry. 2019.

[27] J. A. Aguilar-Saavedra, R. Benbrik, S. Heinemeyer, and M. PÃ©rez-Victoria. Handbook of vectorlike quarks: Mixing and single production. *Phys. Rev.*, D88(9):094010, 2013.

[28] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning.* MIT Press, 2016. http://www.deeplearningbook.org.

[29] Andriy Burkov. *The hundred-page machine learning book.* Andriy Burkov, 2019.

[30] J. Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117, 2015. Published online 2014; based on TR arXiv:1404.7828 [cs.NE].

[31] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2010, Chia Laguna Resort, Sardinia, Italy, May 13-15*, pages 249–256, 2010,.

[32] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *CoRR*, abs/1502.01852, 2015.

[33] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.

[34] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. 2015.

[35] François Chollet. Keras. https://github.com/fchollet/keras, 2015.

[36] Theano Development Team. Theano: A Python framework for fast computation of mathematical expressions. May 2016.

[37] Daniel Hay Guest, Michela Paganini, Michael Kagan, Joshua Wyatt Smith, and Marie Lanfermann. lwtnn/lwtnn: Release for athena v21, February 2017.

[38] P. Speckmayer, A. Hocker, J. Stelzer, and H. Voss. The toolkit for multivariate data analysis, TMVA 4. *J. Phys. Conf. Ser.*, 219:032057, 2010.

[39] Rene Brun and Fons Rademakers. ROOT - An Object Oriented Data Analysis Framework, 1997.

[40] Lyndon Evans and Philip Bryant. LHC Machine. *JINST*, 3:S08001, 2008.

[41] K. Aamodt et al. The ALICE experiment at the CERN LHC. *JINST*, 3:S08002, 2008.

[42] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.

[43] CMS Collaboration. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[44] A. Augusto Alves, Jr. et al. The LHCb Detector at the LHC. *JINST*, 3:S08005, 2008.

[45] Cinzia De Melis. The CERN accelerator complex. Complexe des accélérateurs du CERN. Jan 2016. General Photo.

[46] Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC. Technical Report ATLAS-CONF-2019-021, CERN, Geneva, Jun 2019.

[47] B. Abbott et al. Production and integration of the ATLAS Insertable B-Layer. *JINST*, 13(05):T05008, 2018.

[48] ATLAS Collaboration. Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1. *Eur. Phys. J. C*, 77:490, 2017.

[49] ATLAS Collaboration. Muon reconstruction performance of the ATLAS detector in proton-proton collision data at $\sqrt{s} =$ 13 TeV. *Eur. Phys. J.*, C76(5):292, 2016.

[50] ATLAS Collaboration. Performance of the ATLAS Trigger System in 2015. *Eur. Phys. J.*, C77(5):317, 2017.

[51] Andy Buckley et al. General-purpose event generators for LHC physics. *Phys. Rept.*, 504:145–233, 2011.

[52] A Rimoldi and. Simulation strategies for the ATLAS experiment at LHC. *Journal of Physics: Conference Series*, 331(3):032026, dec 2011.

[53] ATLAS Collaboration. Athena, April 2019.

[54] Michelangelo L. Mangano, Mauro Moretti, Fulvio Piccinini, Roberto Pittau, and Antonio D. Polosa. ALPGEN, a generator for hard multiparton processes in hadronic collisions. *JHEP*, 07:001, 2003.

[55] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. Madgraph 5: going beyond. *Journal of High Energy Physics*, 2011(6):128, Jun 2011.

[56] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A brief introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008.

[57] Manuel Bahr, Stefan Gieseke, Martyn A. Gigg, David Grellscheid, Keith Hamilton, Oluseyi Latunde-Dada, Simon Platzer, Peter Richardson, Michael H. Seymour, Alexander Sherstnev, and Bryan R. Webber. Herwig++ physics and manual. *The European Physical Journal C*, 58(4):639–707, Dec 2008.

[58] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, F. Siegert, and J. Winter. Event generation with SHERPA 1.1. *JHEP*, 02:007, 2009.

[59] S. Agostinelli et al. GEANT4: a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250–303, 2003.

[60] ATLAS Collaboration. Early Inner Detector Tracking Performance in the 2015 data at $\sqrt{s}$ = 13 TeV. Technical Report ATL-PHYS-PUB-2015-051, CERN, Geneva, Dec 2015.

[61] ATLAS Collaboration. Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s}$ = 13 TeV. *Submitted to: Eur. Phys. J.*, 2019.

[62] Tilman Plehn. *Lectures on LHC physics; 2nd ed.* Lecture Notes in Physics. Springer, Cham, 2015.

[63] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008.

[64] Yuri L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber. Better jet clustering algorithms. *JHEP*, 08:001, 1997.

[65] M. Wobisch and T. Wengler. Hadronization corrections to jet cross-sections in deep inelastic scattering. In *Monte Carlo generators for HERA physics. Proceedings, Workshop, Hamburg, Germany, 1998-1999*, pages 270–279, 1998.

[66] Stephen D. Ellis and Davison E. Soper. Successive combination jet algorithm for hadron collisions. *Phys. Rev. D*, 48:3160–3166, 1993.

[67] A. Hrynevich. ATLAS jet and missing energy reconstruction, calibration and performance in LHC Run-2. *JINST*, 12(06):C06038, 2017.

[68] John E. Huth et al. Toward a standardization of jet definitions. In *1990 DPF Summer Study on High-energy Physics: Research Directions for the Decade (Snowmass 90) Snowmass, Colorado, June 25-July 13, 1990*, pages 0134–136, 1990.

[69] 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. *JHEP*, 02:075, 2015.

[70] Jet reclustering and close-by effects in ATLAS run II. Technical Report ATLAS-CONF-2017-062, CERN, Geneva, Jul 2017.

[71] ATLAS Collaboration. Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev. D*, 96:072002, 2017.

[72] ATLAS Collaboration. Jet mass reconstruction with the ATLAS Detector in early Run 2 data. ATLAS-CONF-2016-035, 2016.

[73] ATLAS Collaboration. Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 13$ TeV. 2018.

[74] $E_\mathrm{T}^\mathrm{miss}$ performance in the ATLAS detector using 2015-2016 LHC p-p collisions. Technical Report ATLAS-CONF-2018-023, CERN, Geneva, Jun 2018.

[75] Jonathan M. Butterworth, Adam R. Davison, Mathieu Rubin, and Gavin P. Salam. Jet substructure as a new Higgs search channel at the LHC. *AIP Conf. Proc.*, 1078:189–191, 2009.

[76] David Krohn, Jesse Thaler, and Lian-Tao Wang. Jet trimming. *JHEP*, 02:084, 2010.

[77] Stephen D. Ellis, Christopher K. Vermilion, and Jonathan R. Walsh. Recombination algorithms and jet substructure: Pruning as a tool for heavy particle searches. *Phys. Rev. D*, 81:094023, 2010.

[78] Andrew J. Larkoski, Simone Marzani, Gregory Soyez, and Jesse Thaler. Soft drop. *Journal of High Energy Physics*, 2014(5):146, May 2014.

[79] ATLAS Collaboration. Impact of Alternative Inputs and Grooming Methods on Large-R Jet Reconstruction in ATLAS. ATL-PHYS-PUB-2017-020, 2017.

[80] ATLAS Collaboration. Performance of jet substructure techniques for large-R jets in proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector. *JHEP*, 09:076, 2013.

[81] ATLAS Collaboration. Identification of Boosted, Hadronically Decaying W Bosons and Comparisons with ATLAS Data Taken at $\sqrt{s} = 8$ TeV. *Eur. Phys. J. C*, 76:154, 2016.

[82] ATLAS Collaboration. Identification of high transverse momentum top quarks in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *JHEP*, 06:093, 2016.

[83] ATLAS Collaboration. Boosted hadronic top identification at ATLAS for early 13 TeV data. ATL-PHYS-PUB-2015-053, 2015.

[84] Identification of boosted, hadronically-decaying W and Z bosons in $\sqrt{s} = 13$ TeV Monte Carlo Simulations for ATLAS. Technical Report ATL-PHYS-PUB-2015-033, CERN, Geneva, Aug 2015.

[85] Tilman Plehn, Gavin P. Salam, and Michael Spannowsky. Fat Jets for a Light Higgs. *Phys. Rev. Lett.*, 104:111801, 2010.

[86] Tilman Plehn, Michael Spannowsky, Michihisa Takeuchi, and Dirk Zerwas. Stop reconstruction with tagged tops. *JHEP*, 10:078, 2010.

[87] Jesse Thaler and Ken Van Tilburg. Identifying boosted objects with N-subjettiness. *JHEP*, 03:015, 2011.

[88] Jesse Thaler and Ken Van Tilburg. Maximizing boosted top identification by minimizing N-subjettiness. *JHEP*, 02:093, 2012.

[89] Andrew J. Larkoski, Gavin P. Salam, and Jesse Thaler. Energy correlation functions for jet substructure. *JHEP*, 06:108, 2013.

[90] Andrew J. Larkoski, Ian Moult, and Duff Neill. Power counting to better jet observables. *JHEP*, 12:009, 2014.

[91] Geoffrey C. Fox and Stephen Wolfram. Observables for the Analysis of Event Shapes in e+ e- Annihilation and Other Processes. *Phys. Rev. Lett.*, 41:1581, 1978.

[92] Chunhui Chen. New approach to identifying boosted hadronically-decaying particle using jet substructure in its center-of-mass frame. *Phys. Rev. D*, 85:034007, 2012.

[93] Jesse Thaler and Lian-Tao Wang. Strategies to identify boosted tops. *JHEP*, 07:092, 2008.

[94] ATLAS Collaboration. Measurement of k_{T} splitting scales in $W \to \ell\nu$ events at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Eur. Phys. J. C*, 73:2432, 2013.

[95] Leandro G. Almeida, Seung J. Lee, Gilad Perez, Ilmo Sung, and Joseph Virzi. Top quark jets at the LHC. *Phys. Rev.*, D79:074012, 2009.

[96] ATLAS Collaboration. ATLAS measurements of the properties of jets for boosted particle searches. *Phys. Rev. D*, 86:072006, 2012.

[97] S. Catani, Yuri L. Dokshitzer, M. H. Seymour, and B. R. Webber. Longitudinally invariant K_t clustering algorithms for hadron hadron collisions. *Nucl. Phys.*, B406:187–224, 1993.

[98] Davison E. Soper and Michael Spannowsky. Finding top quarks with shower deconstruction. *Phys. Rev. D*, 87:054012, 2013.

[99] ATLAS Collaboration. Search for $W' \to tb$ decays in the hadronic final state using pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Lett.*, B781:327–348, 2018.

[100] CMS Collaboration. Identification techniques for highly boosted W bosons that decay into hadrons. *JHEP*, 12:017, 2014.

[101] CMS Collaboration. Top Tagging with New Approaches. CMS-PAS-JME-15-002, Geneva, 2016.

[102] Richard D. Ball, Valerio Bertone, Stefano Carrazza, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Nathan P. Hartland, and Juan Rojo. Parton distributions with QED corrections. *Nucl. Phys. B*, 877:290–320, 2013.

[103] ATLAS Collaboration. ATLAS Pythia 8 tunes to 7 TeV data. ATL-PHYS-PUB-2014-021, 2014.

[104] Guido Altarelli, B. Mele, and M. Ruiz-Altaba. Searching for new heavy vector bosons in $p\bar{p}$ colliders. *Z. Phys. C*, 45:109, 1989. [Erratum: Z. Phys. C 47,676(1990)].

[105] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010.

[106] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with parton shower simulations: the POWHEG method. *JHEP*, 11:070, 2007.

[107] Paolo Nason. A new method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004.

[108] Hung-Liang Lai, Marco Guzzi, Joey Huston, Zhao Li, Pavel M. Nadolsky, Jon Pumplin, and C. P. Yuan. New parton distributions for collider physics. *Phys. Rev. D*, 82:074024, 2010.

[109] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 physics and manual. *JHEP*, 05:026, 2006.

[110] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, Pavel M. Nadolsky, and W. K. Tung. New Generation of Parton Distributions with Uncertainties from Global QCD Analysis. *JHEP*, 07:012, 2002.

[111] Peter Skands. Tuning Monte Carlo generators: The Perugia tunes. *Phys. Rev. D*, 82:074018, 2010.

[112] Michał Czakon, Paul Fiedler, and Alexander Mitov. Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_s^4)$. *Phys. Rev. Lett.*, 110:252004, 2013.

[113] M. Bahr et al. Herwig++ physics and manual. *Eur. Phys. J. C*, 58:639–707, 2008.

[114] 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. *JHEP*, 07:079, 2014.

[115] ATLAS Collaboration. Simulation of top-quark production for the ATLAS experiment at $\sqrt{s} = 13$ TeV. ATL-PHYS-PUB-2016-004, 2016.

[116] Stefano Catani, Leandro Cieri, Giancarlo Ferrera, Daniel de Florian, and Massimiliano Grazzini. Vector Boson Production at Hadron Colliders: A Fully Exclusive QCD Calculation at Next-to-Next-to-Leading Order. *Phys. Rev. Lett.*, 103:082001, 2009.

[117] Steffen Schumann and Frank Krauss. A Parton shower algorithm based on Catani-Seymour dipole factorisation. *JHEP*, 03:038, 2008.

[118] Stefan Höche, Frank Krauss, Steffen Schumann, and Frank Siegert. QCD matrix elements and truncated showers. *JHEP*, 05:053, 2009.

[119] ATLAS Collaboration. The ATLAS Simulation Infrastructure. *Eur. Phys. J. C*, 70:823, 2010.

[120] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt. Parton distributions for the LHC. *Eur. Phys. J. C*, 63:189–285, 2009.

[121] ATLAS Collaboration. Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC. *Eur. Phys. J. C*, 77:332, 2017.

[122] Luke de Oliveira et al. Jet-images – deep learning edition. *JHEP*, 07:069, 2016.

[123] Pierre Baldi, Kevin Bauer, Clara Eng, Peter Sadowski, and Daniel Whiteson. Jet substructure classification in high-energy physics with deep neural networks. *Phys. Rev. D*, 93(9):094034, 2016.

[124] Jannicke Pearkes, Wojciech Fedorko, Alison Lister, and Colin Gay. Jet Constituents for Deep Neural Network Based Top Quark Tagging. 2017.

[125] Gregor Kasieczka, Tilman Plehn, Michael Russell, and Torben Schell. Deep-learning top taggers or the end of QCD? *JHEP*, 05:006, 2017.

[126] ATLAS Collaboration. Electron efficiency measurements with the ATLAS detector using 2012 LHC proton–proton collision data. *Eur. Phys. J. C*, 77:195, 2017.

[127] ATLAS Collaboration. Electron identification measurements in ATLAS using $\sqrt{s} = 13$ TeV data with 50 ns bunch spacing. ATL-PHYS-PUB-2015-041, 2015.

[128] ATLAS Collaboration. Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 76:292, 2016.

[129] ATLAS Collaboration. Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector. *Eur. Phys. J. C*, 76:581, 2016.

[130] ATLAS Collaboration. Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run. ATL-PHYS-PUB-2016-012, 2016.

[131] ATLAS Collaboration. In-situ measurements of the ATLAS large-radius jet response in 13 TeV pp collisions. ATLAS-CONF-2017-063, 2017.

[132] ATLAS Collaboration. Performance of b-jet identification in the ATLAS experiment. *JINST*, 11:P04008, 2016.

[133] ATLAS Collaboration. Photon identification in 2015 ATLAS data. ATL-PHYS-PUB-2016-014, 2016.

[134] ATLAS Collaboration. Measurement of $W^{\pm}$ and Z Boson Production Cross Sections in pp Collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector. ATLAS-CONF-2015-039, 2015.

[135] ATLAS Collaboration. ATLAS simulation of boson plus jets processes in Run 2. ATL-PHYS-PUB-2017-006, 2017.

[136] ATLAS Collaboration. Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton–proton collision data. ATLAS-CONF-2016-024, 2016.

[137] ATLAS Collaboration. Electron and photon energy calibration with the ATLAS detector using data collected in 2015 at $\sqrt{s} = 13$ TeV. ATL-PHYS-PUB-2016-015, 2016.

[138] ATLAS Collaboration. Measurement of the photon identification efficiencies with the AT-LAS detector using LHC Run-1 data. *Eur. Phys. J. C*, 76:666, 2016.

[139] ATLAS Collaboration. Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data. *Eur. Phys. J. C*, 74:3071, 2014.

[140] ATLAS Collaboration. Search for heavy particles decaying into top-quark pairs using lepton-plus-jets events in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Eur. Phys. J. C*, 78(7):565, 2018.

[141] ATLAS Collaboration. Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC. *Eur. Phys. J. C*, 76:653, 2016.

[142] ATLAS Collaboration. Analysis of events with b-jets and a pair of leptons of the same charge in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *JHEP*, 10:150, 2015.

[143] ATLAS Collaboration. Search for pair and single production of new heavy quarks that decay to a Z boson and a third-generation quark in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *JHEP*, 11:104, 2014.

[144] ATLAS Collaboration. Search for production of vector-like quark pairs and of four top quarks in the lepton-plus-jets final state in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *JHEP*, 08:105, 2015.

[145] ATLAS Collaboration. Search for pair production of a new heavy quark that decays into a W boson and a light quark in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Phys. Rev. D*, 92:112007, 2015.

[146] ATLAS Collaboration. Search for single production of vector-like quarks decaying into Wb in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Eur. Phys. J. C*, 76:442, 2016.

[147] ATLAS Collaboration. Search for the production of single vector-like and excited quarks in the Wt final state in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *JHEP*, 02:110, 2016.

[148] CMS Collaboration. Inclusive Search for a Vector-Like T Quark with Charge $\frac{2}{3}$ in pp Collisions at $\sqrt{s} = 8$ TeV. *Phys. Lett. B*, 729:149–171, 2014.

[149] CMS Collaboration. Search for Top-Quark Partners with Charge 5/3 in the Same-Sign Dilepton Final State. *Phys. Rev. Lett.*, 112(17):171801, 2014.

[150] CMS Collaboration. Search for vector-like T quarks decaying to top quarks and Higgs bosons in the all-hadronic channel using jet substructure. *JHEP*, 06:080, 2015.

[151] CMS Collaboration. Search for pair-produced vectorlike B quarks in proton-proton collisions at $\sqrt{s}$=8 TeV. *Phys. Rev. D*, 93(11):112009, 2016.

[152] CMS Collaboration. Search for vector-like charge 2/3 T quarks in proton-proton collisions at sqrt(s) = 8 TeV. *Phys. Rev. D*, 93(1):012003, 2016.

[153] CMS Collaboration. Search for vectorlike light-flavor quark partners in proton-proton collisions at $\sqrt{s}$ =8 TeV. *Phys. Rev. D*, 97:072008, 2018.

[154] ATLAS Collaboration. Search for pair production of vector-like top quarks in events with one lepton, jets, and missing transverse momentum in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector. *JHEP*, 08:052, 2017.

[155] ATLAS Collaboration. Search for pair production of heavy vector-like quarks decaying to high-p_T W bosons and b quarks in the lepton-plus-jets final state in pp collisions at $\sqrt{s}$=13 TeV with the ATLAS detector. *JHEP*, 10:141, 2017.

[156] ATLAS Collaboration. Search for pair production of up-type vector-like quarks and for four-top-quark events in final states with multiple b-jets with the ATLAS detector. *JHEP*, 07:089, 2018.

[157] ATLAS Collaboration. Search for pair production of heavy vector-like quarks decaying into high-p_T W bosons and top quarks in the lepton-plus-jets final state in pp collisions at $\sqrt{s}$=13 TeV with the ATLAS detector. *JHEP*, 08:048, 2018.

[158] ATLAS Collaboration. Search for pair- and single-production of vector-like quarks in final states with at least one Z boson decaying into a pair of electrons or muons in pp collision data collected with the ATLAS detector at $\sqrt{s} = 13$ TeV. *Phys. Rev. D*, 98(11):112010, 2018.

[159] ATLAS Collaboration. Search for single production of a vector-like B quark decaying into a bottom quark and a Higgs boson which decays into a pair of photons. Technical Report ATLAS-CONF-2018-024, CERN, Geneva, Jul 2018.

[160] ATLAS Collaboration. Search for single production of vector-like quarks decaying into Wb in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *JHEP*, 05:164, 2019.

[161] CMS Collaboration. Search for single production of a heavy vector-like T quark decaying to a Higgs boson and a top quark with a lepton and jets in the final state. *Phys. Lett. B*, 771:80–105, 2017.

[162] CMS Collaboration. Search for single production of vector-like quarks decaying to a Z boson and a top or a bottom quark in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 05:029, 2017.

[163] CMS Collaboration. Search for single production of vector-like quarks decaying into a b quark and a W boson in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett. B*, 772:634–656, 2017.

[164] CMS Collaboration. Search for top quark partners with charge 5/3 in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 08:073, 2017.

[165] CMS Collaboration. Search for pair production of vector-like T and B quarks in single-lepton final states using boosted jet substructure in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 11:085, 2017.

[166] CMS Collaboration. Search for single production of a vector-like T quark decaying to a Z boson and a top quark in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett. B*, 781:574–600, 2018.

[167] CMS Collaboration. Search for pair production of vector-like quarks in the bW$\overline{\text{b}}$W channel from proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett. B*, 779:82–106, 2018.

[168] CMS Collaboration. Search for single production of vector-like quarks decaying to a b quark and a Higgs boson. *JHEP*, 06:031, 2018.

[169] CMS Collaboration. Search for vector-like T and B quark pairs in final states with leptons at $\sqrt{s} = 13$ TeV. *JHEP*, 08:177, 2018.

[170] ATLAS Collaboration. Search for pair- and single-production of vector-like quarks in final states with at least one Z boson decaying into a pair of electrons or muons in pp collision data collected with the ATLAS detector at $\sqrt{s} = 13$ TeV. 2018.

[171] ATLAS Collaboration. Search for new phenomena in events with same-charge leptons and b-jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. 2018.

[172] J.A. Aguilar-Saavedra. Protos: Program for top simulations. Technical report.

[173] J. A. Aguilar-Saavedra. Identifying top partners at LHC. *JHEP*, 11:030, 2009.

[174] Richard D. Ball et al. Parton distributions with lhc data. *Nucl. Phys. B*, 867:244–289, 2013.

[175] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A brief introduction to pythia 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008.

[176] Michal Czakon and Alexander Mitov. Top++: A program for the calculation of the top-pair cross-section at hadron colliders. *Comput. Phys. Commun.*, 185:2930, 2014.

[177] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt. Parton distributions for the LHC. *Eur. Phys. J. C*, 63:189–285, 2009.

[178] Jon Butterworth et al. PDF4LHC recommendations for LHC Run II. *J. Phys. G*, 43:023001, 2016.

[179] Hung-Liang Lai, Marco Guzzi, Joey Huston, Zhao Li, Pavel M. Nadolsky, Jon Pumplin, and C. P. Yuan. New parton distributions for collider physics. *Phys. Rev. D*, 82:074024, 2010.

[180] Jun Gao, Marco Guzzi, Joey Huston, Hung-Liang Lai, Zhao Li, Pavel Nadolsky, Jon Pumplin, Daniel Stump, and C. P. Yuan. Ct10 next-to-next-to-leading order global analysis of qcd. *Phys. Rev. D*, 89(3):033009, 2014.

[181] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing nlo calculations in shower monte carlo programs: the powheg box. *JHEP*, 06:043, 2010.

[182] John M. Campbell, R. Keith Ellis, Paolo Nason, and Emanuele Re. Top-pair production and decay at NLO matched with parton showers. *JHEP*, 04:114, 2015.

[183] Peter Zeiler Skands. Tuning monte carlo generators: The perugia tunes. *Phys. Rev. D*, 82:074018, 2010.

[184] Emanuele Re. Single-top Wt-channel production matched with parton showers using the POWHEG method. *Eur. Phys. J.*, C71:1547, 2011.

[185] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. NLO single-top production matched with shower in POWHEG: s- and t-channel contributions. *JHEP*, 09:111, 2009. [Erratum: JHEP02,011(2010)].

[186] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 physics and manual. *JHEP*, 05:026, 2006.

[187] 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. *JHEP*, 07:079, 2014.

[188] ATLAS Collaboration. Search for pair production of vector-like top quarks in events with one lepton, jets, and missing transverse momentum in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector. *JHEP*, 08:052, 2017.

[189] ATLAS Collaboration. Search for pair production of heavy vector-like quarks decaying to high-p_T W bosons and b quarks in the lepton-plus-jets final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *JHEP*, 10:141, 2017.

[190] David Krohn, Jesse Thaler, and Lian-Tao Wang. Jets with variable r. *JHEP*, 06:059, 2009.

[191] ATLAS Collaboration. Jet reclustering and close-by effects in ATLAS run II. 2017.

[192] M. Lanfermann. *Identification of b-jets and c-jets Using Deep Neural Networks with the AT-LAS Detector. The Development and Performance of a Family of DL1 High-level Flavour Tagging Algorithms*. PhD thesis, Université de Genève, 2019.

[193] K. Kondo. Dynamical Likelihood Method for Reconstruction of Events With Missing Momentum. 1: Method and Toy Models. *J. Phys. Soc. Jap.*, 57:4126–4140, 1988.

[194] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. LemaÃ®tre, A. Mertens, and M. Selvaggi. DELPHES 3, A modular framework for fast simulation of a generic collider experiment. *JHEP*, 02:057, 2014.

[195] J. Erdmann et al. A likelihood-based reconstruction algorithm for top-quark pairs and the KLFitter framework. 748:18, 2013.

[196] Mathieu Buchkremer, Giacomo Cacciapaglia, Aldo Deandrea, and Luca Panizzi. Model Independent Framework for Searches of Top Partners. *Nucl. Phys. B*, 876:376–417, 2013.

[197] ATLAS Collaboration. Search for pair production of heavy vector-like quarks decaying into hadronic final states in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev.*, D98(9):092005, 2018.

[198] A L Read. Presentation of search results: the cl_s technique. *J. Phys. G*, 28:2693, 2002.

[199] Thomas Junk. Confidence level computation for combining searches with small statistics. *Nucl. Instrum. Meth. A*, 434:435–443, 1999.

[200] G. Cowan, K. Cranmer, E. Gross, and O. Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71:1554, 2011.

[201] James D. Bjorken and Stanley J. Brodsky. Statistical model for electron-positron annihilation into hadrons. *Phys. Rev. D*, 1:1416–1420, Mar 1970.

[202] Leandro G. Almeida, Seung J. Lee, Gilad Perez, George F. Sterman, Ilmo Sung, and Joseph Virzi. Substructure of high-p_T Jets at the LHC. *Phys. Rev. D*, 79:074017, 2009.