RUPRECHT-KARLS-UNIVERSITÄT HEIDELBERG

Lisa Marie Baltes

Quark and Gluon Jet Discrimination using Trigger Jets in Dijet Searches with the ATLAS Experiment

Masterarbeit

HD-KIP-20-55

KIRCHHOFF-INSTITUT FÜR PHYSIK

Department of Physics and Astronomy

University of Heidelberg

Master thesis

in Physics

submitted by

Lisa Marie Baltes

born in Wiesbaden-Dotzheim

2020

Quark and Gluon Jet Discrimination using Trigger Jets in Dijet Searches with the ATLAS Experiment

This Master thesis has been carried out by Lisa Marie Baltes

at the

Kirchhoff-Institute for Physics

under the supervision of

Prof. Dr. Hans-Christian Schultz-Coulon

Unterscheidung von Quark- und Gluon-Jets unter der Verwendung von Trigger Jets in Dijet-Systemen mit dem ATLAS Experiment

Die Analyse von Dijet-Ereignissen offenbart neue Einblicke in neue mögliche Phänomene, die jenseits des Standardmodells (SM) liegen, wie zum Beispiel über dunkle Materie. Massereiche Wechselwirkungsteilchen können als Resonanz im invarianten Massenspektrum des Zerfallsprodukts beobachtet werden. Jedoch sind Dijet-Systeme mit Massen unter 1 TeV aufgrund eines großen Wirkungsquerschnitts der SM Dijet-Erzeugung durch Bandbreiten- und Speicherkapazitäten des Detektors statistisch begrenzt. Die Trigger-Objekt-Level-Analyse (TLA) macht die Suche nach niederenergetischen Resonanzen ab 450 GeV möglich, indem sie nur einen Teil der vollen Detektorinformation der Ereignisse aufzeichnet und rekonstruiert. TLA verwendet den Viererimpuls des online rekonstruierten hadronischen Jets und weitere Jet-Eigenschaften, die auf Informationen des Kalorimetersystems basieren. Informationen des Spurdetektorsystems sind nicht erhalten. Das Signalmodell, welches in dieser Masterarbeit untersucht wird, besteht aus einem massereichen Wechselwirkungsteilchen, welches in zwei Quark-Jets zerfällt, während der QCD Untergrund von Gluon-Gluon- und Quark-Gluon-Endzuständen dominiert wird. Daher führt das Identifizieren des Jet-Flavours der beiden Jets im Endzustand dazu, dass der Untergrund im Vergleich zum Signal unterdrückt wird. Diese Arbeit präsentiert Studien verschiedener multivariater Verfahren, die auf Methoden des maschinellen Lernens basieren, und erzielt eine gute Quark- und Gluon-Jet-Unterscheidung.

Quark and Gluon Jet Discrimination using Trigger Jets in Dijet Searches with the ATLAS Experiment

The analysis of dijet events could reveal new insights on possible phenomena beyond the Standard Model (SM) such as dark matter. Massive particles can be observed as an excess in the invariant mass spectrum of the decay product. However, due to a large SM dijet production cross-section, dijet searches are statistically limited by the detector's bandwidth and storage limitations at masses below 1 TeV. The Trigger-Object Level Analysis (TLA) allows the search for low-mass resonances down to an invariant mass of 450 GeV by recording and reconstructing only a part of the full event. TLA uses the four-momentum of the online reconstructed jet and some jet properties based on calorimeter information. Tracking information is not included. The signal model used in this thesis consists of a massive mediator particle decaying into two quarks, while the QCD background is dominated by gluon-gluon and quark-gluon final states. Therefore, tagging the flavour of the two final state jets suppresses the background compared to the signal. This thesis presents studies of different multivariate analysis methods based on machine learning techniques and achieves a good quark-gluon separation.

Contents

1 Introduction and Motivation

> *It is nevertheless probable that there may be as many as 1000 million stars [within a sphere of radius of 3.09·10^{16} kilometres]; but many of them may be extinct and dark, and nine-tenths of them though not all dark may be not bright enough to be seen by us at their actual distances. [· · ·] Many of our supposed thousand million stars, perhaps a great majority of them, may be dark bodies.*

This is a quote from the "Baltimore lectures on molecular dynamics and the wave theory" written by Lord Kelvin in 1904 [1]. It is referring to one of the first attempts to dynamically estimate the amount of dark matter in the Milky Way. Lord Kelvin describes the correlation between the size of a system and the velocity dispersion of the stars inside it. It took almost another 30 years until the Swiss-American astronomer Zwicky studied the redshifts of various galaxy clusters including the Coma Cluster. Through these studies, he realised that luminous matter is not able to account for the observed average velocity dispersion along the line-of-sight. In his later work, he stated that "if this would be confirmed, we would get the surprising result that dark matter is present in much greater amount than luminous matter" [2].

Down to the present day, scientists wonder about the composition of this unknown component of the universe which is called dark matter. Measurements indicate that dark matter amounts to approximately 23% of the total energy-matter content in the universe, but its nature is still unknown [3]. The search for dark matter links particle physics with cosmology because only elementary particles are potential dark matter candidates. Although it is not possible to embed dark matter into the current model of particle physics, the Standard Model (SM), theories beyond the SM offer many different dark matter candidates.

The challenge of finding dark matter particles (if existent) is the very weak interaction with visible matter. In order to detect a dark matter signal, multiple experiments around the world exploit different detection techniques [4–6]. One such experiment is the Large Hadron Collider (LHC) where dark matter might be produced through interaction with the partons inside the colliding protons via a dark matter mediator. This mediator could then decay back into SM partons, leaving a dijet signature consisting of two quark jets in the detector. The invariant mass of the decay product would form a resonance at the mediator mass above the smoothly falling QCD background. The search for signals in the dijet invariant mass spectrum is not limited to dark matter but also applicable to many other models for physics beyond the SM.

At the ATLAS experiment, dijet searches were able to probe large values for the mediator masses, but there was no excess found so far [7, 8]. Therefore, the aim is to broaden the search towards low mediator masses above approximately 100 GeV. Due to bandwidth and storage limitations, it is only possible to record and later analyse a tiny fraction of all collisions at ATLAS. The trigger system of the ATLAS detector decides which of the collisions are permanently stored. Especially at low masses, the SM dijet production rate exceeds the recording rate of the detector which induces an unwanted loss in statistics of several magnitudes in the invariant mass range below 1 TeV.

A novel approach, called Trigger-Object Level Analysis (TLA), is able to access dijet events in the low-mass region by recording only physics objects that were reconstructed within the trigger system. Therefore, the event information only includes the reconstructed objects with variables that were available at the respective trigger level and other calorimeter-based variables. This reduces the size of each event so that the bandwidth limitations are bypassed and the sensitivity loss can be restored down to invariant masses of approximately 500 GeV. Hence, event recordings at a high rate are possible, while less than 1% of the total trigger bandwidth is used.

Searches for dijet resonances containing two quark jets in the final state are polluted by a large QCD background that is dominated by dijet events in which at least one of the two leading jets is gluon-initiated. Quark and gluon jets possess different properties induced by the fact that a quark carries one colour index, while a gluon carries two colour indices. This leads to gluon jets tending to have more constituents and a broader radiation pattern compared to quark jets. Therefore, tagging the flavour of the jets may enhance the significance of the signal of interest.

This thesis studies multivariate analysis methods based on machine learning techniques in order to achieve a quark-gluon separation. In earlier studies [9–11], most discriminants for the identification of quark- and gluon-initiated jets were based on tracking variables, which are not available in the TLA approach. Thus, these inaccessible variables are replaced by calorimeter-based discrimination variables for this study. Investigating a possible quark-gluon separation, this thesis provides studies for the improvement of the sensitivity for dark matter signals as well as for additional searches for physics beyond the SM.

This thesis is structured as follows. In Chapter 2, an overview of the Standard Model of particle physics is given and it addresses the question, why physics beyond the SM is expected to exist. A special focus lies on dark matter and its detection methods. In Chapter 3, different multivariate analysis methods based on machine learning techniques are introduced. It also presents their implementation in the Toolkit for Multivariate Analysis. The Large Hadron Collider and the ATLAS detector are described in Chapter 4. Chapter 5 explains the reconstruction and calibration procedure of dijet events in the ATLAS detector and provides an introduction into Monte Carlo simulations. Furthermore, the Trigger-Object Level Analysis approach is motivated including a presentation of its calibration scheme. In Chapter 6, quark and gluon jet discrimination is introduced and its application

in TLA is explained. In addition, discrepancies between Monte Carlo simulations and data are illustrated and their correction is presented. The quark-gluon tagging is outlined and its performance is analysed in Chapter 7. In Chapter 8, a statistical analysis of the quark and gluon jet discrimination using the BUMPHUNTER algorithm is performed. The conclusion and outlook of the thesis are presented in Chapter 9.

2 Theoretical Background

The Standard Model [12–14] is the current model of particle physics which successfully describes the known fundamental particles and their interactions: the electromagnetic interaction, the weak interaction and the strong interaction. It does not include the gravitational force. The SM is described in Section 2.1. Despite its huge success, there are some phenomena which cannot be described by this model. This is further explained in Section 2.2.

2.1 The Standard Model of Particle Physics

The SM is a relativistic quantum field theory which unites the fundamental theories of electromagnetism namely quantum electrodynamics (QED), the strong interaction called quantum chromodynamics (QCD) and the weak force. The underlying gauge symmetry is

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

Y is the weak hypercharge including the electromagnetic and weak neural-current couplings and L stands for the left-handed chiral particle states to which the weak charged-current interaction couples. The QCD invariance under transformations in colour space is denoted by C.

Fundamental fermions and their antiparticles are described by interacting fields, which are also responsible for their interactions among each other. These force carriers correspond to so called gauge bosons. The fundamental particles including their relevant properties are shown in Figure 2.1.

The 12 fundamental matter particles of the Standard Model are divided into two categories: quarks and leptons. They are spin-1/2 fermions and can be grouped into three generations, which differ by their increasing masses.

There are six different flavours of *quarks*: up (u), down (d), charm (c), strange (s), top (t) and bottom (b). Up-type quarks (u,c,t) have an electric charge of +2/3 e, while down-type quarks (d,s,b) carry a charge of $q = -1/3$ e. In addition, each quark has one out of three colour indices: red, blue and green (r,b,g). Systems of quarks in which the quarks are bound by the strong QCD force are called hadrons. Combinations of three quarks are labelled as baryons and quark-antiquark systems as mesons. Whereas all mesons are unstable, there is one stable baryon, the proton (uud). Quarks interact via all three interactions of the SM.

Leptons can be divided into charged and neutral particles. The three different types of charged leptons, electron e^-, muon μ^- and tau τ^-, form lepton doublets with the corresponding neutral particles, the massless neutrinos. The charged leptons are

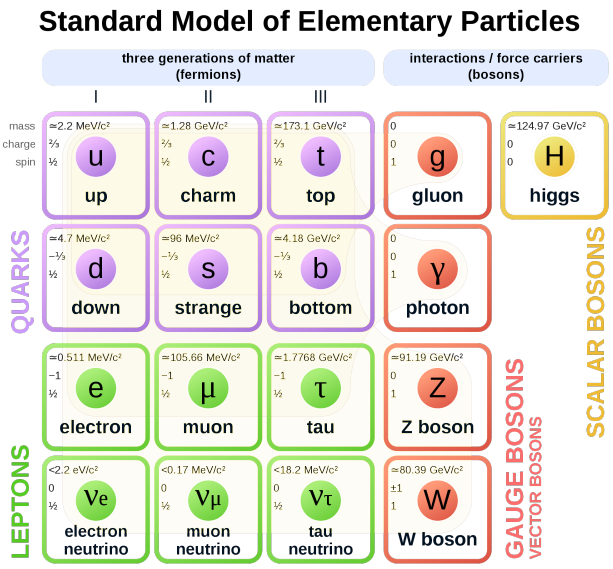


Figure 2.1: The Standard Model of elementary particles including 12 fundamental fermions and 5 fundamental bosons. Taken from [15].

charged negatively ($q = -1e$) and interact through the electromagnetic and the weak force. Of the charged leptons only the electron is stable.

The *gauge bosons*, also called vector bosons, are the interaction particles of the three fundamental forces. All bosons are spin-1 particles. The gluon is the interaction particle of the strong force and it is massless. Since the gluon is colour-charged, it is able to interact with itself. The massless photon is the force-carrying particle of QED. The interaction particles of the weak interaction are the $W^{\pm}$ and the Z boson.

The final element of the Standard Model is the *Higgs boson*, which has a mass of $m_H \approx 125$ GeV. It was discovered by the ATLAS and CMS experiments at the Large Hadron Collider (LHC) at CERN in 2012 [16, 17]. Unlike the other fundamental particles, it is a scalar particle with a spin of 0. The Higgs boson is responsible for the mechanism in which all other particles obtain their mass.

Electroweak Unification

The electroweak theory developed by Glashow, Salam and Weinberg (GSW) [18–20] in the 1960s unites the electromagnetic and weak interaction and is associated with the symmetry group

$$U(1)_Y \times SU(2)_L, \tag{2.2}$$

where the electromagnetic $U(1)$ gauge symmetry of the electromagnetic interaction is replaced by a new $U(1)_Y$ local gauge symmetry. Since the charged-current weak interaction only couples to left-handed chiral particles, the symmetry group $SU(2)_L$

often carries the label L. Right-handed chiral particle states do not change under a local $SU(2)_L$ gauge transformation.

The Lagrange density stays invariant under the above mentioned gauge transformation with the covariant derivative

$$\partial_\mu \rightarrow \mathcal{D}_\mu = \partial_\mu + ig' B_\mu \frac{Y}{2} + ig_W W_\mu \cdot \frac{\tau}{2} \tag{2.3}$$

including the gauge fields B_μ and $W_\mu = \left(W_\mu^1, W_\mu^2, W_\mu^3\right)$ and the coupling constants g' and g_W. The generators of $SU(2)$ are the three Pauli matrices τ.

Of the four interaction bosons, the photon and the Z boson are neutral. Their corresponding fields A_μ and Z_μ can be written as a linear combination of the generator of the hypercharge and the neutral (third) generator of the weak interaction

$$A_\mu = B_\mu \cos\theta_W + W_\mu^3 \sin\theta_W \tag{2.4}$$

$$Z_\mu = -B_\mu \sin\theta_W + W_\mu^3 \cos\theta_W, \tag{2.5}$$

where θ_W is the weak mixing angle. This splitting originates from the Higgs mechanism. The physical $W^\pm$ bosons are identified as a linear combination of the gauge fields W_μ^1 and W_μ^2

$$W_\mu^\pm = \frac{1}{\sqrt{2}} \left(W_\mu^1 \mp i W_\mu^2\right). \tag{2.6}$$

The GSW model shows a relation between the couplings of the electromagnetic and weak interaction

$$e = g_W \sin\theta_W = g' \cos\theta_W = g_Z \sin\theta_W \cos\theta_W. \tag{2.7}$$

Higgs Mechanism

The local $U(1)_Y \times SU(2)_L$ gauge symmetry of the electroweak interaction is broken if mass terms for the massive gauge bosons $W^\pm$ and Z are introduced. The Higgs mechanism is able to preserve the required local gauge invariance and generates masses of the gauge bosons by introducing the Higgs field ϕ. The Lagrangian for the Higgs field is

$$\mathcal{L} = (\mathcal{D}_\mu \phi)^\dagger (\mathcal{D}^\mu \phi) - V(\phi) \tag{2.8}$$

with the covariant derivative defined in Equation 2.3 and the Higgs potential

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2. \tag{2.9}$$

The potential shows for $\mu^2 < 0$ a Mexican hat form. The vacuum expectation value $v = \sqrt{-\mu^2}/\lambda$ of this potential is non-zero ($v = 246$ GeV), which yields a global symmetry breaking of the Lagrangian. This effect is called *spontaneous symmetry breaking*.

Fermion masses

The Higgs mechanism is able to generate masses for the fundamental fermions of the Standard Model. A fermion mass term like $-m_f\bar{\psi}\psi$ breaks the $U(1)_Y \times SU(2)_L$ local gauge symmetry and therefore, can not be included in the Lagrangian. However, a term like

$$-g_f \left(\bar{\Psi}_L \phi \Psi_R + \bar{\Psi}_R \phi^\dagger \Psi_L\right), \tag{2.10}$$

where ϕ is the Higgs field and Ψ_L/Ψ_R indicate the left-/right-handed chiral particle states, is invariant under the required local gauge transformation. This term corresponds to a coupling between the fermion field and the Higgs field with the coupling strength g_f. The coupling is called Yukawa coupling and it is given by

$$g_f = \sqrt{2}\frac{m_f}{v}. \tag{2.11}$$

Quantum Chromodynamics

QCD is the quantum field theory of the strong interaction. It is associated with a $SU(3)$ local gauge invariance. The covariant derivative

$$\partial_\mu \rightarrow \mathcal{D}_\mu = \partial_\mu + i g_S G^a_\mu \cdot \frac{\lambda^a}{2} \tag{2.12}$$

assures the required invariance. The eight generators of the $SU(3)$ symmetry group are the 3×3 Gell-Mann matrices λ^a ($a \in \{1, \cdots, 8\}$). G^a_μ are eight new real fields which correspond to the gluons of QCD and g_S is the strong coupling constant. Since the $SU(3)$ symmetry group is a non-Abelian gauge group, the gluon fields do not only couple to quarks fields but they also self-interact as depicted in Figure 2.2.

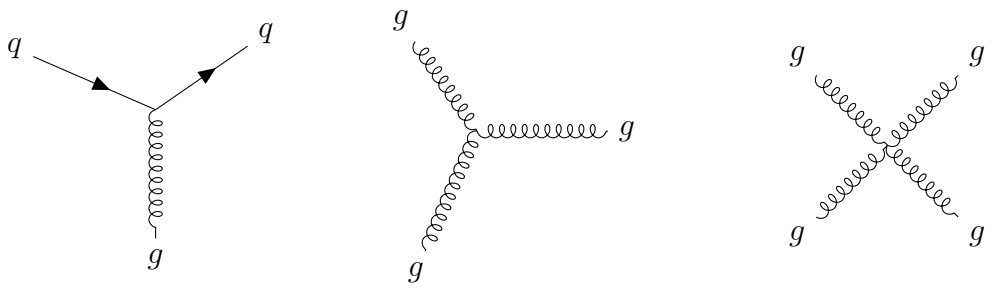


Figure 2.2: QCD interactions predicted from the $SU(3)$ local gauge invariance. The Feynman diagrams show from left to right the coupling of gluons to quarks, the triple gluon vertex and the quartic gluon vertex.

Colour Confinement

Quarks and gluons, which are called partons, have never been observed unbound due to a property called *colour confinement*. Coloured objects always form colour singlet states because particles require a zero colour charge to propagate freely. Therefore, quarks are bound in colourless hadrons. The possible combinations to form hadrons out of quarks and antiquarks are strongly limited by the requirement of forming colour singlet states. Colour confinement is not a proven property but it is believed that its origin lies in the gluon self-interactions.

In processes like $q\bar{q} \to q\bar{q}$ (Figure 2.3), two highly energetic quarks are produced in the final state. They propagate back-to-back in the center-of-mass frame of the interaction. Due to colour confinement, the quarks form collimated streams of colourless particles via a process which is called hadronisation. These particle streams are called *jets*. The two jets of the process in Figure 2.3 follow the direction of the initial quarks.

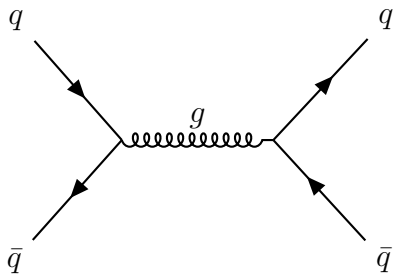


Figure 2.3: Tree-level Feynman diagram of a $q\bar{q} \to q\bar{q}$ interaction.

Running Coupling

The coupling constant of QCD, α_S, is not constant. For low-energy scales, the constant is of $\mathcal{O}(1)$ and it decreases for higher energies. Hence, calculations using perturbation theory are not possible in the low-energy regime. Non-perturbed calculations using lattice QCD are able to probe the predictions of QCD for example the mass of the proton. In the high-energy range or at very small distances, α_S becomes small enough to motivate the use of perturbation theory. This property of QCD is called *asymptotic freedom*. The running of the coupling strength originates from the contribution of higher order Feynman diagrams.

2.2 Physics Beyond the Standard Model

Although the Standard Model is successful in describing the known fundamental particles and their interactions, there are several theoretical and experimental indications that this model is incomplete.

Dark Matter

In the 1930s, Zwicky [21] and Smith [22] observed that luminous matter is not able to account for the masses of the Coma and Virgo clusters. The experimental rotation curves of galaxies orbiting the center of a cluster show significant discrepancies compared to the predictions of the Newtonian/Keplerian gravitation law. The majority of the luminous mass of a cluster is located in its central part. Hence, the velocity distribution is expected to decrease as a function of the distance to the center at radii larger than the luminous part of the cluster. The observations, however, show a uniformly distributed curve. Therefore, Zwicky and Smith postulated the existence of a large amount of non-luminous matter which we now call *dark matter*.

The Standard Model of cosmology predicts that baryonic matter amounts to only 5% of the total energy-matter density in the universe. Dark matter makes up 23%, while the majority is accounted for by dark energy (72%). The existence of dark energy results in a non-vanishing cosmological constant Λ in general relativity, which is responsible for the acceleration of the universe [23–25].

Dark matter is one important contribution for the structure formation of the universe [26]. Depending on its typical velocity in the early universe, three families of dark matter are defined: hot, warm and cold dark matter. Candidates for the hot dark matter (HDM) are the light relativistic neutrinos. The sterile neutrino [27] is a warm dark matter (WDM) candidate. Light particles, like in HDM, remain relativistic until late times of the universe. This induces a typical length scale of matter objects of the size of a large galaxy cluster. Therefore, the model of hot dark matter was ruled out because it could not explain the structures we observe today. Non-relativistic dark matter, such as in the cold dark matter (CDM) model, is today the leading dark matter model because it is able to explain sizes of objects down to the mass of the Earth.

There are several possible candidates of CDM including supersymmetric particles [28], axions [29] and the weakly interacting massive particle (WIMP) [30]. The natural mass of the latter is expected to be in the GeV-TeV range. These non-baryonic particles are expected to interact gravitationally and via the weak force. There is no known particle in the Standard Model which satisfies the required properties of a WIMP.

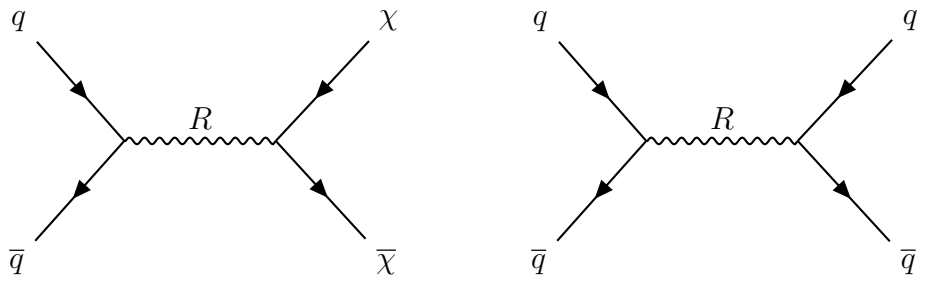
Dark Matter in Collider Experiments

To understand and detect dark matter is one of the important goals of many experiments, e.g. the Large Hadron Collider. If dark matter (χ) interacts weakly with visible matter, e.g. quarks, these particles could be produced at the LHC. However, pair-produced dark matter particles as in Figure 2.4a do not interact with the detectors. ATLAS searches, therefore, focus on two main strategies. The first one implies the requirement of initial state radiation of, e.g. a gluon, which balances the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ from dark matter. This signature is called "Mono-X" search, where X stands for the Standard Model particle, which interacts with the

detector [31].

A second option to search for dark matter is to look for the mediator particle itself. If there is a mediator which connects the Standard Model sector with the dark matter sector then, this mediator R can also decay back into a final state with two Standard Model particles (Figure 2.4b). This motivates the search for dijet final states. Since this diagram is a s-channel production, the invariant mass of the decay product will form a resonance at the mediator mass above the smoothly falling QCD background. The coupling for this new force is the same in both vertices. This dark matter model has two important parameter: the mass of the mediator m_R and its coupling strength to quarks g_q.

One example of such a model introduces the massive Z' gauge boson, whose couplings to leptons are neglected. Therefore, this mediator is called *leptophobic Z'* [32].



(a) $q\overline{q} \to \chi\overline{\chi}$ (b) $q\overline{q} \to q\overline{q}$

Figure 2.4: Feynman diagrams with an unknown mediator particle which couples to the Standard Model sector as well as to the dark matter sector.

3 Multivariate Data Analysis and Machine Learning

The field of pattern recognition is growing in importance, especially in the last 20 years. Pattern recognition addresses the discovery of underlying structure of a data sample with the help of computer algorithms. With these characteristics, it is possible to classify the data into different categories. Nowadays, we have to handle and analyse a huge amount of complex data which makes the characterisation more difficult. The development of *machine learning* enabled the possibility to recognise patterns in data with the help of *multivariate data analysis*. This chapter gives a small overview of different techniques of multivariate analyses [33] and explains how these are embedded in the Toolkit for Multivariate Analysis (TMVA) [34].

3.1 Introduction

The goal of machine learning is to build a model which learns from and predicts data. It takes an arbitrary data set including multiple variables as input and returns the desired characterisation as output. To qualify the output of the model, the whole data set is split into three disjoint parts: a training set, a validation set and a test set.

The *training set* consists of observations and their target characterisation. These are used to train a chosen model and to determine the best parameters by learning the underlying structure of the given input data set.

After the model has been trained, its performance is evaluated by the *validation set*. This data set is used to fine-tune the model parameters but it is never used for the learning process itself. Since the parameters are chosen according to the performance on the validation set, the model is indirectly affected by the validation process.

The *test set* is used to evaluate the final model setup once the training process is complete. The division of the full data set into training, evaluation and test set assures the generalisation of the data analysis. A large training set assures a good model estimate, while a small validation set induces a noisy estimate of the model performance. If the size of the data set is limited, *cross-validation* is used. This technique splits the whole data into k partitions, trains the model with all but one fraction and performs evaluation on the held-out part. This procedure is repeated such that every partition is used once in the evaluation step. Mean and variance of all runs are averaged and at the end, the full data set is retrained.

The training of data that includes the input variables as well as their target

characterisation is called *supervised learning*. If the desired output consists of one or more continuous variable(s), the process is referred to as regression. A classification problem, however, returns one finite output out of a discrete number of output possibilities. In Section 3.2, the classification process is further explained.

Contrary to supervised learning, *unsupervised learning* comprises the input data without any knowledge of the desired output. One possible application of this machine learning paradigm is to find groups of similar properties within the data sample (clustering). Furthermore, it is possible to reconstruct the probability density function for a given data set by a process, which is called density estimation. In addition, the dimensionality reduction from a multi-dimensional input variable space to a two- or three-dimensional space can be used to visualize the data.

The paradigm of *reinforcement learning* is not based on an output known a priori, but instead it discovers the optimal output throughout the learning process itself. The goal is that an agent learns the best possible action by maximising the reward which is provided by the environment for each action selected by the agent. This optimisation process founds on the balance between exploration (of all possible actions) and exploitation (of the already gained knowledge).

In order to optimise a model for a characterisation problem, an *error function* is minimised. This function determines the deviation of the prediction of the model from the training data set for a chosen set of model parameters $\mathbf{w}$. The training set consists of a data set $\{\mathbf{x}_n, t_n\}$ including input vectors $\mathbf{x}_n$ and corresponding target values t_n for $n = 1, ..., N$, whereby N is the number of observations. In two-class classification problems, it is convenient to use a binary representation such that $t_n = 1$ represents class 1 and $t_n = 0$ represents class 2. A common choice for an error function for such a problem is the *binary cross-entropy*

$$E(\mathbf{w}) = -\frac{1}{2}\sum_{n=1}^{N}\{t_n \ln(y_n) + (1 - t_n)\ln(1 - y_n)\}, \tag{3.1}$$

which measures the performance of a two-class classification model. This quantity would be zero in case of no deviation between the model prediction $y_n(\mathbf{x}_n, \mathbf{w})$ and the target value of the data set in any data point. The factor of $1/2$ is a convention.

The number of free parameters, also known as hyperparameters, in a chosen model has a direct influence on the model's flexibility and capability to capture the underlying complex structure of the data. This *model selection* induces a possible range of performances from *under-fitting* over the best selection until *over-fitting*. The former refers to a model which is neither able to model the training data set nor to generalise new data because it does not capture the underlying structure. The latter describes a model with too much flexibility which induces an adjustment of the model to statistical fluctuations of the target values in the input data set. Since the random noise of the validation set varies from the one in the training set, over-fitting can be identified if the model performs well on the training set but shows a poor performance if applied to the validation set. Over-fitting decreases with an increase of the size of the data set because the number of free parameters in the model is

then significantly smaller than the number of data points. Another approach to reduce over-fitting is *regularisation*. The error function in Equation 3.1 is modified by including a penalty term, which yields

$$\widetilde{E}(\mathbf{w}) = -\frac{1}{2}\sum_{n=1}^{N}\{t_n \ln(y_n) + (1-t_n)\ln(1-y_n)\} + \frac{\lambda}{2}||\mathbf{w}||^2, \tag{3.2}$$

where λ determines the regularisation strength. For $\lambda = 0$, the error function simplifies to Equation 3.1. The type of regularisation including the L2 norm of the model parameter $||\mathbf{w}||^2 = \mathbf{w}^{\mathrm{T}}\mathbf{w}$ is called *Ridge regression*. λ controls the relative importance of the penalty term and therefore, has a direct influence on the magnitude of the regularisation.

Besides finding the appropriate number of free parameters, the choice of type of model is indispensable in order to achieve a good performance. In many practical applications, it is beneficial to apply some preprocessing techniques to the data; a topic which is further covered in Section 3.3.

3.2 Event Classification

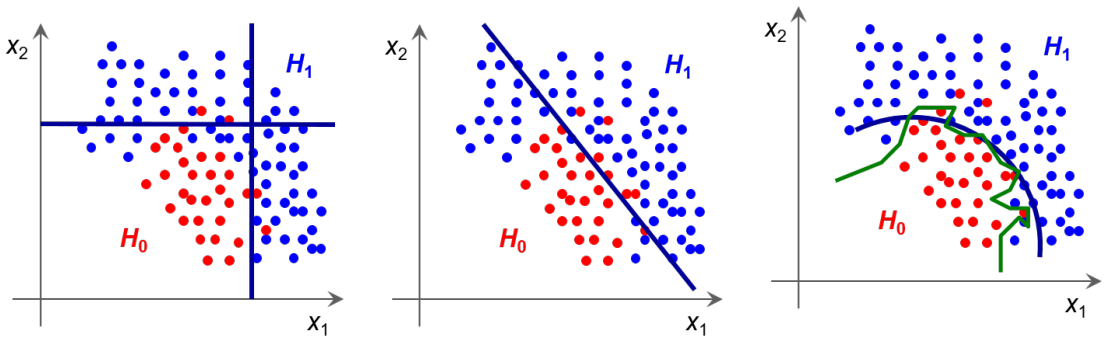
In machine learning, classification aims to divide the training data set into different categories (out of a finite number of options) and thus is able to predict the membership of an unseen data point to a category based on the training. The categories, also referred to as classes, are mostly chosen to be disjoint such that each observation is assigned to only one discrete class. Therefore, the input data set can be split in decision regions, whose borders are called *decision boundaries*. In Figure 3.1, a two-class classification problem with different choices of decision boundaries can be seen. Depending on the underlying nature of the training set, the selection of a rectangular (Figure 3.1a), a linear (Figure 3.1b) or a quadratic decision boundary (Figure 3.1c) yields the best separation. If the two classes within a data set can be completely separated by a linear decision boundary, it is called *linearly separable*. In Figure 3.1c, the green line represents a higher-order decision boundary which clearly shows the effect of over-fitting. This decision boundary will perform significantly worse on a newly created training set with the same underlying distribution but a different random noise.

Linear Models

There are many different classes of models which are used to solve classification problems. Linear models are linear in the input vector $\mathbf{x}$ and can be written as

$$y(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\mathbf{x} + w_0, \tag{3.3}$$

where $\mathbf{w}$ is called *weight vector*, w_0 is the *bias* and y is the real target value. Given that classification problems require discrete class labels, Equation 3.3 is transformed



(a) Rectangular cut.

(b) Linear cut.

(c) Quadractic (blue) and non-linear (green) cut.

Figure 3.1: Comparison of different types of decision boundaries in a two-class classification problem. The classes are denoted by red (H_0) and blue points (H_1). The decision boundaries are marked by a blue line. The green line shows a higher-order decision boundary, whose large number of free parameters induces over-fitting. Taken from [35].

by a non-linear function f called *activation function*. Three examples for activation functions can be seen in Figure 3.2, whereas only the logistic sigmoid can be used for (binary) classification.

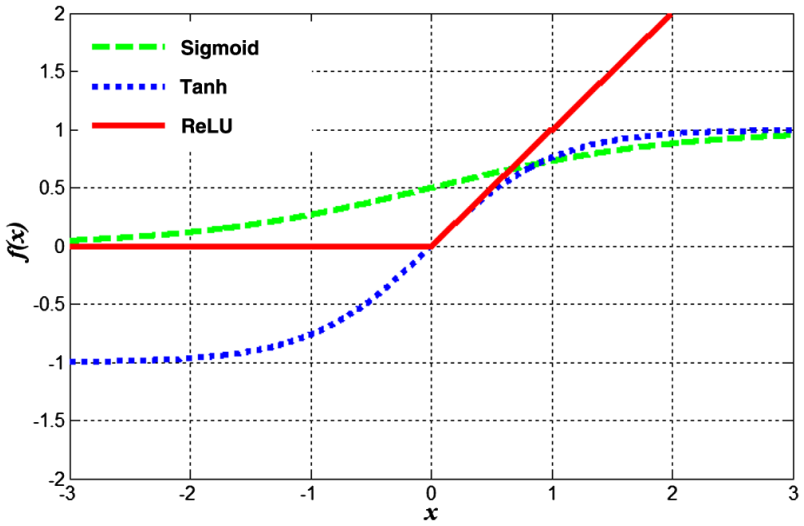


Figure 3.2: Comparison of different activation functions in the range $[-3, 3]$. The logistic sigmoid is shown in green, the hyperbolic tangent in blue and the rectified linear unit (ReLU) in red. Taken from [36].

The perceptron algorithm is another example of a linear model. It further transforms the input vector $\mathbf{x}$ of input variables into non-linear basis functions $\phi(\mathbf{x})$. This strategy leaves the model linear in its coefficients $\mathbf{w}$ but provides a non-linear relationship to the input variables. The generalised linear model for an input vector

can then be expressed by the formula

$$y(\mathbf{x}) = f\left(\widetilde{\mathbf{w}}^{\mathrm{T}}\phi(\mathbf{x})\right), \tag{3.4}$$

whereby the bias w_0 has been absorbed into the definition of the weight vector $\widetilde{\mathbf{w}} = (w_0, \mathbf{w})$. Hence, the first component of $\phi(\mathbf{x})$ is set to 1. In further calculations, the tilde is omitted. The perceptron algorithm uses the sign function[1] as activation function f.

Neural Networks

Feed-forward neural networks that use parametrised basis functions $\phi(x)$, whose parameters are optimised during the training process, are called *multilayer perceptrons*. This model consists of multiple layers of logistic regression models. The basis functions are chosen to be non-linear functions of linear combinations of the input. All neural networks consist of an input and an output layer. In between, there can be several *hidden layers*. A network with one hidden layer, for instance, is called two-layer network. Its network function can be written as

$$y_k(\mathbf{x}, \mathbf{w}) = \sigma\left(\sum_{j=0}^{M} w_{kj}^{(2)} \sigma\left(\sum_{i=0}^{D} w_{ji}^{(1)} x_i\right)\right), \tag{3.5}$$

whereby x_i and y_k represent the input and output variables of the network. The terms $z_j = \sigma\left(\sum_{i=0}^{D} w_{ji}^{(1)} x_i\right)$ are called hidden units. The superscript of the weight parameters $\mathbf{w}$ indicates the corresponding layer. The logistic sigmoid function

$$\sigma(a) = \frac{1}{1 + \exp(a)} \tag{3.6}$$

is used as non-linear activation function. The layout of a two-layer feed-forward neural network can be seen in Figure 3.3. These neural networks are called multilayer networks due to their concatenation of several perceptron models.

The network shown in Figure 3.3 is a fully-connected neural network. To generalise the model, *skip-layer* connections [37], which skip one hidden layer, can be embedded into the network.

3.3 Data Preprocessing

In machine learning, data preprocessing is a step in which the data set gets transformed such that the features of the data can be easily extracted by the training algorithm. The data set, which the machine analyses, often contains correlated variables of different orders of magnitude. To simplify and to speed up the training process, it is often beneficial to preprocess the data. This preprocessing transforms

[1] The sign function has a value of -1 for negative and $+1$ for positive numbers.

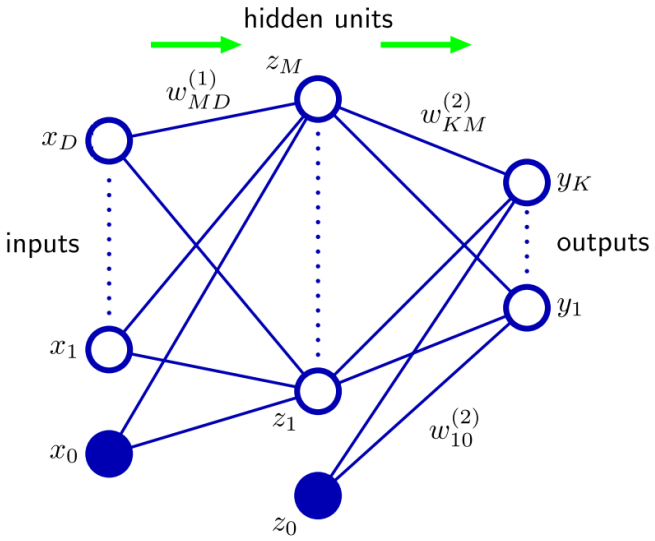


Figure 3.3: Diagram of a two-layer feed-forward neural network. The variables x_i, z_j and y_k are shown as nodes; the weight coefficients are illustrated as links between the nodes; bias parameters are represented as filled nodes. The arrow indicates the direction of information flow inside the network. Taken from [33, p. 228].

the input variables into a new space of variables. The goal is to extract the most important features of the data and feed them into the machine. It is important to keep all the useful information otherwise discriminating power and accuracy of the classification model are reduced. The new feature space is then used as new input to train the machine. The testing data set is preprocessed in the same way as the training set.

Normalisation

The process which determines the minimum and maximum value of all input variables and linearly transforms them into the range of $[-1, 1]$ is called variable normalisation. With this scaling, a comparison between the weights of the model is possible. Compared to smaller weights, larger ones represent a stronger separation power.

Decorrelation

If the features of the data set are strongly dependent on each other, the model will not be able to extract the underlying structure of the data which may lead to a poor performance. Linear correlations are taken into account through the calculation of the square-root of the covariance matrix ($\sqrt{C}$). The new input variables are then

computed via

$$\mathbf{x'} = \left(\sqrt{C}\right)^{-1} \mathbf{x}. \tag{3.7}$$

This transformation has to be applied separately for subsets of different distributions in the data set because their correlations may differ. For most real-world applications, the correlation of the input variables is not linear and the decorrelation process may even worsen the performance of the model.

Principle Component Analysis

Principle Component Analysis (PCA) is used to approximate and/or visualise a data set. It reduces the dimensionality of the input space by finding the linear subspace with minimal information loss. PCA is a linear transformation which maximises the variance of the projected data. The maximal variance is reached with the transformation featuring the largest eigenvalue. The largest variance of this transformation defines the first new coordinate (first principle component). The second-largest variance defines the second coordinate and so on. Lower-order principle components with insignificant variance are usually omitted.

3.4 Toolkit for Multivariate Analysis

The Toolkit for Multivariate Analysis (TMVA) is a framework for multivariate classification and regression techniques available in ROOT [38]. This section focuses on the configuration possibilities for classification. TMVA was designed to perform training, testing and performance evaluation for high energy physics applications especially to find small signals in large data samples. The training methods available are based on supervised learning algorithms; hence, the target sample characterisation has to be known (for classification).

The TMVA analysis is split into a training phase (including training, testing and evaluation) and an application phase. In the latter, the chosen method of the training phase is used to solve the classification or regression problem of an unknown data set.

The *Factory* class (ROOT::TMVA::Factory) is responsible for the organisation of the training phase. It performs preanalysis and preprocessing of input variables as well as calculates their linear correlations. Variable transformations are defined with the booking option *VarTransform*, which offers among others the choices normalisation (*Norm*), decorrelation (*Deco*) and principle component analysis (*PCA*). Furthermore, it reveals a preliminary ranking of the variables based on their discrimination power. In the Factory, training and testing data sets are specified and in addition, the separation into signal and background events is performed for each set, e.g. by applying selection cuts. Afterwards, the multivariate analysis techniques including their chosen configuration are *booked* and the Factory runs the training,

testing and evaluation process. Each executed multivariate analysis stores the training results in so-called weight files. During evaluation, control and performance plots for the training and test data set are created. This includes the ranking of the input variables according to their separation for classification. The separation of a classifier y is defined as

$$\langle S^2 \rangle = \frac{1}{2} \int \frac{(p_s(y) - p_b(y))^2}{p_s(y) + p_b(y)} \mathrm{d}y, \tag{3.8}$$

where p_s and p_b are the distributions for signal and background of y. The separation is zero for identical and one for disjoint distributions. For each classifier, the output distribution of signal and background is used to determine the optimal TMVA cut. The Receiver Operating Characteristic (ROC) curve shows the background rejection as a function of the signal efficiency for different thresholds of the respective cut of the test sample. The Area Under the ROC Curve (AUC) gives a performance measure for each method.

In the application phase, the weight files of the best performing training method are chosen and imported by the *Reader* class (ROOT::TMVA::Reader). The weight files of each method contain the full information of the configuration and the results of the training process. The Reader applies the results of the training phase including the optimal cut to increase the signal purity to an unspecified data set.

TMVA provides several multivariate analysis methods used to solve classification or regression problems. Below a selection is introduced and their booking options are specified.

Rectangular Cuts

One of the simplest classification methods is the application of a rectangular cut on a set of discrimination variables. In a strict sense, this classifier is not part of a multivariate analysis but rather the combination of several univariate ones because the variable cuts are independent of each other. The cut-based classifier has – compared to other discrimination methods – only two possible outputs: signal or background. The optimal cuts for each variable are determined by maximising the background rejection for a fixed signal efficiency. This procedure is repeated for a number of signal efficiencies.

TMVA offers several *FitMethod*s including Monte Carlo Sampling [39], Genetic Algorithm [40] and Simulated Annealing [41]. In case that the distributions of the discrimination variables are known beforehand, minimum (*FMin*) or maximum (*FMax*) cut requirements can be applied using the option *VarProp* in order to remove obsolete requirements. An automatic detection of obsolete requirements is enabled with *FSmart*. The performance and discrimination power of the cut-based optimisation analysis depends strongly on the signal and background distributions because this method is based on the assumption that the signal is clustered in the variable space. If this is not the case, this classifier underperforms.

Linear Discriminant Analysis

A linear discriminant provides the classification of data based on the linear model according to Equation 3.3. Bias and weights are determined by the method of least squares [42]. For gaussian distributed variables with linear correlation, linear classifiers provide a reasonable good performance that is comparable with results obtained with non-linear discriminants. In case of discrimination variables with equal sample mean but different variance, the discrimination power vanishes. The *Fisher* discriminant [43] is one well-known example of a linear classifier available in TMVA.

Boosted Decision Trees

A decision tree is an ordered tree-like model which splits the training data set at each node into two subnodes by the application of a learned decision rule. Each decision rule is only based on one discrimination variable and the splitting is repeated until a stopping criterion is fulfilled. The end nodes (leaves) of the decision tree are labelled as signal or background, based on the majority of training events in the particular node. TMVA fixes the maximal depth of a decision tree with the hyperparameter *MaxDepth*, while requiring a minimum size of *MinNodeSize* for each leaf (in per cent of the training data).

In order to weaken the dependency of the tree on statistical fluctuations in the training sample and to approve the performance of the classifier, a *boosting* algorithm is introduced. It fits many small or large trees to reweighted versions of the training set, whereas the total number of trees is given by *NTrees*. In the end, an event is classified by the majority vote. The boosting type for the decision trees is set by *BoostType* and TMVA offers, e.g., AdaBoost [44] and Bagging [45] as choices. A decision tree that includes a boosting algorithm is called *boosted decision tree*.

Multilayer Perceptron

The feed-forward multilayer perceptron (MLP) is one possible implementation of an artificial neural network [46] and therefore, belongs to the class of non-linear discriminant analyses. The configuration options for the MLP implemented in TMVA are given in [34, p. 97 f.].

For classification, the number of input nodes of the network is given by the number of discriminating variables. The number of *HiddenLayers* determines the hidden layer architecture of the network, whereas the output layer consists of only one neuron: the estimator. The architecture of the network is often given as a function of the number of input variables N. For example, a network specified by the expression $(N+1, N)$ contains two hidden layers with $N+1$ and N nodes respectively. The neuron's response function is given by a concatenation of a synapse function *NeuronInputType* and an activation function *NeuronType*. The training algorithm of the feed-forward neural network is set by *TrainingMethod*, which offers Backpropa-

gation [47], the Broyden-Fletcher-Goldfarb-Shannon (BFGS) Algorithm [48–51] and the Genetic Algorithm as choices. Given that it may not be possible to know the necessary number of training epochs *NCycles* in advance, the convergence of the network is monitored in epoch steps. *ConvergenceTests* sets the number of training cycles in which the improvement has to be smaller than the number *ConvergenceImprove*. The convergence test along with tests for overtraining are performed after each epoch that is set by the *TestRate* argument.

Due to its highly flexible and complex network architecture, a MLP has a good performance but also tends to overtrain the data set. To avoid the latter, a Bayesian extension can be selected by the *UseRegulator* parameter. This option allows a complex network layout, while regularising the model at the same time. The regulariser is equivalent to a Gaussian prior over the network weights [33, p. 24 ff.].

4 The ATLAS Experiment at the Large Hadron Collider

The ATLAS[1] experiment [52] is one of four large experiments at the Large Hadron Collider [53] at CERN[2] close to Geneva. It studies proton-proton collision up to a center of mass energy of 14 TeV to test the predictions of the Standard Model and to search for effects of new physics beyond this model. From 2009 until 2013 during Run-1, the LHC successfully took data at different center of mass energies ranging from 900 GeV to 8 TeV. In Run-2, which started in 2015 and lasted until 2018, the center of mass energy increased to 13 TeV.

4.1 The Large Hadron Collider

The LHC consists of a circular tunnel with circumference of 27 km which was built for the Large Electron-Positron Collider (LEP). The tunnel has eight straight sections and eight arcs. At four interaction points detectors are placed with different research purposes. The ATLAS and CMS[3] detector were build for general physics research. The ALICE[4] detector is mainly used for heavy-ion physics, whereas the LHCb[5] detector investigates B-physics.

Pre-accelerated protons are injected into the LHC tunnel every 25 ns in bunches of up to 10^{11} protons. These bunches are further accelerated until they reach their final interaction energy and then collide 40 million times per second.

The number of events for a given process for a given period of data taking is given by the product of *cross-section* and *integrated luminosity*:

$$N = \sigma \int \mathcal{L}(t)\mathrm{d}t. \tag{4.1}$$

The cross-section σ measures the quantum mechanical probability for the interaction to occur and is dependent on the fundamental physics, whereas the luminosity $\mathcal{L}$ is determined by the rate of the collisions in the collider. The LHC reaches peak luminosities of the order of 10^{34} cm^{-2}s^{-1}.

Proton-proton collisions at the LHC occur up to 10 hours until the proton beam becomes unstable or the luminosity decreased too much. Then, the beam gets dumped into a lead block and the LHC is filled again with new proton bunches.

[1] **A** **T**oroidal **L**HC **A**pparatu**S**
[2] **C**onseil **E**uropéen pour la **R**echerche **N**ucléaire
[3] **C**ompact **M**uon **S**olenoid
[4] **A** **L**arge **I**on **C**ollider **E**xperiment
[5] **L**arge **H**adron **C**ollider **b**eauty

The dense packing of colliding protons increases the number of simultaneously occurring events in each bunch crossing. The additional collisions, which are called *pileup* events, contribute to the background of the main high-energy collision, which is referred to as *hard scatter* event. Two different sources of pileup are distinguished: in-time pileup occurring in the same bunch crossing and out-of-time pileup, which accounts for signals from previous and subsequent bunch crossings. The online and offline event reconstruction is affected by pileup events such that effective algorithms are necessary to distinguish hard-scatter from soft-QCD pileup events. Especially the reconstruction of hadronic jets including their energy and structure is biased by pileup because these interactions can generate additional jets, so-called *pileup jets* [54, 55]. The expected amount of pileup μ is directly correlated to the luminosity $\mathcal{L}$ by

$$\mu = \frac{\mathcal{L}\sigma_{inelastic}}{n_c f_{rev}}, \tag{4.2}$$

where $\sigma_{inelastic}$ is the inelastic cross-section in proton-proton collision, n_c is the number of colliding pairs of bunches and f_{rev} is the revolution frequency of the LHC. The average pileup amounts to approximately 13.4 for an integrated luminosity of 3.2 fb^{-1} at the beginning of Run-2 in 2015 and it increased to around 25.1 for $\int \mathcal{L}\mathrm{d}t = 32.9$ fb^{-1} in 2016. As the luminosity increases during Run-2, the average pileup increases as well [56, 57].

One of the main goals of the LHC was the discovery of the Higgs boson which was postulated by Peter Higgs [58] in 1964 and discovered by the ATLAS [16] and CMS [17] experiments in 2012.

4.2 The ATLAS Detector

The ATLAS detector is a 44 m long and 25 m in diameter expanded cylinder with a mass of 7000 tonnes, which is located approximately 100 m under ground. It uses a right-handed coordinate system in which the collision point determines the center of the coordinate system and the z-axis coincides with the beam axis. The x-axis points towards the center of the LHC, whereas the y-axis is perpendicular to the before mentioned axis and points into the sky. Typical measured variables are the angle θ, which determines the angle to the z-axis, and the azimuth angle ϕ. The transverse momentum p_T and the transverse energy $E_\mathrm{T} = E \sin\theta$ of a particle with momentum p and energy E are defined in the x-y plane. p_z is the z-component of the momentum. The *pseudorapidity* η is a spacial coordinate used to describe angles to the beam axis:

$$\eta = -\ln\tan\left(\frac{\theta}{2}\right). \tag{4.3}$$

The angular separation between two particles is measured by the Lorentz invariant quantity for massless particles $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$. The *rapidity* is defined as

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right). \tag{4.4}$$

The ATLAS detector is consisting of several different subdetectors which are concentrically arranged around the interaction point. The three major components of the detector are the inner detector, the calorimeters and the muon spectrometer (see Figure 4.1).

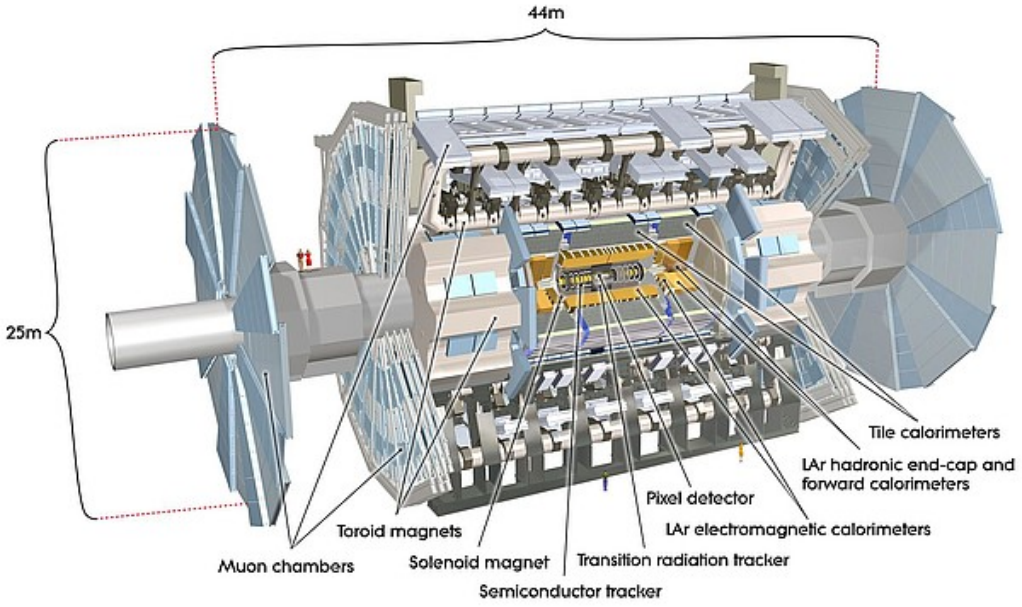


Figure 4.1: The onion-like structure of the ATLAS detector with concentrically arranged subdetectors. Taken from [52].

Inner detector

The inner detector is composed of a silicon pixel detector, a silicon micro-strip detector and a transition-radiation straw-tube tracker. The silicon sensors have a fine granularity with which direction, momentum and electric charge of particles can be measured. With this information, vertex finding and particle identification are performed. Charged particles traversing the tracking material ionise it and loose energy according to the Bethe-Bloch formula [12, p. 13]. This energy loss is strongly dependent on the type and thickness of the material. The inner detector is embedded into a 2 T superconducting solenoid magnet, which induces a measurable bending of the tracks of charged particles due to the Lorentz force. The inner detector is able to track particles up to a pseudorapidity of $|\eta| = 2.5$.

Calorimeters

The calorimeter system uses different sampling techniques with varying granularities to measure the energy and position of incident particles. The combination of passive and active layers induces the total absorption of these particles that results in particle showers, which can be measured by the active material. Furthermore, the passive material assures a short length between interactions of these showers.

A characteristic quantity for electromagnetic showers is the *radiation length*. It is defined as the average distance in which an electron looses a factor of $1/e$ of its energy via bremsstrahlung. The development of hadronic showers is characterised by the *interaction length*, which is defined as the mean distance between hadronic interactions of relativistic hadrons. The interaction length is significantly larger than the radiation length for the same material.

The whole calorimeter system of the ATLAS detector covers a range up to $|\eta| = 4.9$. In the central region, the interaction length of the calorimeters amounts to 9.7. The calorimeter system is divided into several subcomponents which exploit the different properties of particle interactions via the electromagnetic and the strong force. It is split into the Liquid Argon calorimeter and the Tile calorimeter. The layout of the calorimeter system can be seen in Figure 4.2.

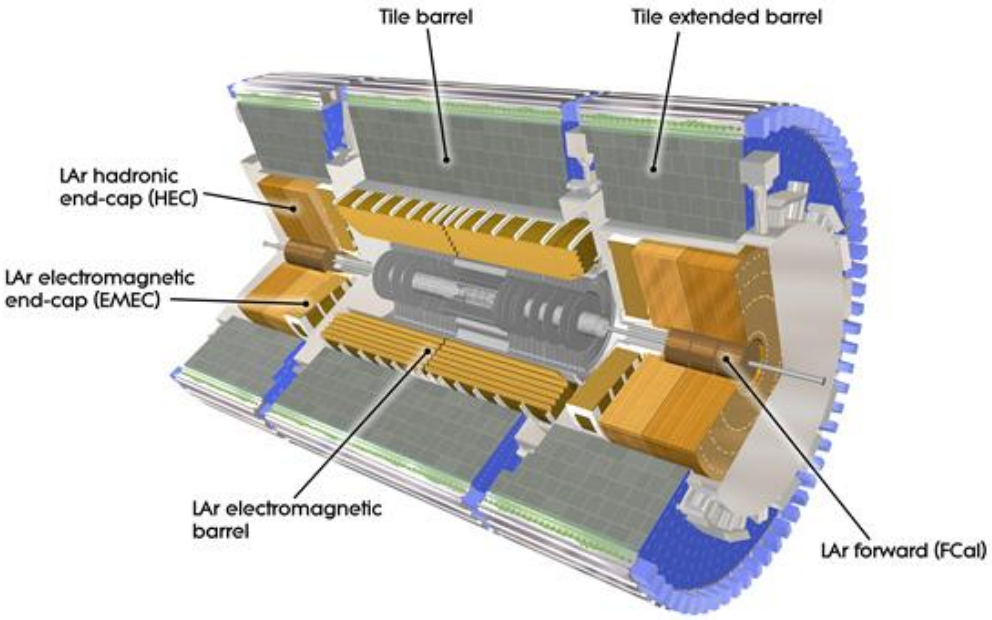


Figure 4.2: The ATLAS calorimeter system including its subcomponents. Liquid argon calorimeters are indicated in yellow, tile calorimeters in grey. Taken from [52].

The Liquid Argon calorimeter [59] is a sampling calorimeter which uses lead as passive and Liquid Argon (LAr) as active material at a temperature of 8 K. LAr has

a linear energy response and it is very fault-tolerant in respect to radiation. Particles passing through the lead absorber induce a particle shower. The resulting particles ionise the LAr and charged particles are created, which drift inside an electric field to the electrodes. There, the electric current is measured and read out. The LAr calorimeter is used to detect both electromagnetic and hadronic showers. It consists of four subdetectors: the electromagnetic barrel, the electromagnetic end-caps, the hadronic end-caps and the forward calorimeter.

The electromagnetic barrel covers a pseudorapidity range of $|\eta| < 1.475$. The end-caps consist of two coaxial wheels covering a range of $1.375 < |\eta| < 3.2$. The combination of electromagnetic barrel and end-caps, which detect electromagnetic showers, is called *electromagnetic calorimeter* (ECal). The thickness of the LAr region in the ECal amounts to 2.1 mm. Both, barrel and end-caps, are sectioned into one presampler and three sampling layers. The presampler detector is a single thin layer of LAr which is responsible for the correction in energy loss by electrons and photons in the inner detector. The sampling layers are made of alternating samples of lead and LAr. The granularity varies throughout the layers. Table 4.1 shows the granularities of the electromagnetic barrel. The slicing of the end-caps is slightly larger. The energy deposit in the first layer with its fine granularity is used for particle reconstruction and identification. Most of the energy of incident particles is deposited in the second layer. The third layer contains the tail of the electromagnetic shower. Due to the fine granularity, the ECal is mostly used for precision measurements for electrons, photons, jets and missing E_{T} [60].

	ECal barrel
Presampler	0.025×0.1
1st layer	$0.025/8 \times 0.1$
2nd layer	0.025×0.025
3rd layer	0.05×0.025

Table 4.1: Granularity for the barrel of the ECal in $\Delta\eta \times \Delta\phi$ slices.

The hadronic end-cap calorimeter (HEC) is a hadronic calorimeter and covers a pseudorapidity range of $1.5 < |\eta| < 3.2$. It is directly located behind the end-caps of the ECal and consists of four lead-LAr layers.

The LAr forward calorimeter (FCal) is approximately 10 interaction lengths deep and it covers a range of $3.1 < |\eta| < 4.9$. It consists of three modules which are designed to measure hadronic interactions.

The Tile calorimeter (TileCal) [61] is a large hadronic sampling calorimeter, which is located outside the ECal. Its coarse granularity is mainly used for jet reconstruction and $E_{\mathrm{T}}^{\mathrm{miss}}$ measurements. It consists of a barrel ($|\eta| < 1.0$) and two extended barrels ($0.8 < |\eta| < 1.7$), which are made out of steel as the absorber and scintillating tiles as the active material. There are three layers of material with interaction lengths of 1.5, 4.1 and 1.8 for the barrel and 1.5, 2.6 and 3.3 for the extended barrel.

The scintillation light produced in the tiles is transmitted by wavelength shifting fibres to photomultiplier tubes [62].
The combination of TileCal, HEC and FCal forms the hadronic calorimeter (HCal).

Muon system

The muon spectrometer identifies and measures the momenta of muons. It consists of large superconducting toroid magnets which bend the muon tracks in high-precision tracking chambers. The provided magnetic field is mostly perpendicular to the muon trajectories. In a barrel region ($|\eta| < 1.4$), the muon tracks are measured by chambers of cylindrical layers around the beam axis. For $1.4 < |\eta| < 2.7$, the muon chambers are orientated perpendicular to the beam axis. The precision tracking chambers consist of Monitored Drift Tubes and Cathode Strip Chambers. Fast detectors provide information for the L1 trigger system.

Trigger system

The collision rate in Run-2 at the LHC is approximately 40 MHz. The rate at which the ATLAS detector can record data is only a few kHz. Disc space and bandwidth limit the number of events which can be processed. Therefore, the trigger and data acquisition system is an essential component of the ATLAS detector. It is responsible to reduce the enormous event rates online by deciding whether an event is interesting and hence, is stored offline for later analysis.

The trigger system in Run-2 [63] is a two-level trigger with a hardware-based *first level* (L1) *trigger* and a software-based *High Level Trigger* (HLT). Both trigger together build a so-called trigger chain.

The L1 trigger reduces the initial collision rate of the LHC down to 100 kHz. The decision time for L1 is 2.5 μs. It uses coarse calorimeter and muon detector information to reconstruct high energetic objects (photons, jets, electrons, muons, taus and $E_{\mathrm{T}}^{\mathrm{miss}}$). The decision thresholds whether to keep an event are mostly based on the energies of the objects in this event. For example, the J100 trigger lets an event pass if it contains a jet with an energy over 100 GeV, otherwise it gets rejected. The energy of the object used for the trigger decision is the calibrated energy at detector level. The performance of the L1 trigger is often studied in turn-on curves, where e.g. the differences in online and offline calibration are compared. In addition, most triggers have a prescale. A prescale of, e.g. 10 means that only every 10th events passing this L1 trigger is kept. The L1 trigger further identifies regions of interest (RoI), which are sent to the HLT. The RoI contain the geographical coordinates (in η and ϕ) of regions with interesting features.

The HLT analyses events that passed the L1 trigger and reduces the rate to 1 kHz with a decision time of 200 ms. It reconstructs events in full granularity in the RoI. The online event reconstruction of the HLT is very similar to the one which is applied to the stored offline data. Similar to L1, different HLT trigger thresholds apply energy cuts to decide if an event is recorded or not.

Events passing the trigger requirements are stored offline in different data *streams* according to how they need to be processed and what they are used for. Each stream is based on a set of trigger selections and prescales and contains all events that satisfy those selections. Most data used for physics analyses is saved in the so-called physics *Main* stream, which is the largest recording stream. It records data that includes the event information from the full detector read-out. There are many other types of streams which are used for calibration purposes or special analyses. One example of the latter is the *Data Scouting* stream, which is further discussed in Section 5.4.

5 Data Sample, Event Selection and Trigger-Object Level Analysis

5.1 Monte Carlo Simulation

Monte Carlo (MC) simulations are a tool used in high-energy physics to emulate the reconstruction of jets from individual particles being generated in proton-proton collisions. The predictions of the MC simulations are then compared to the measured data of a detector. The simulation chain used in the ATLAS experiment consists of different stages from the event generation over the detector simulation to the digitisation of energy deposits in the detector and it is integrated into the ATLAS software framework Athena [64].

The first step of the simulation chain is the physics event generation in which final state particles of proton-proton collisions at the LHC are randomly generated. The properties of the particles are based upon theoretical models such as the SM or supersymmetric models depending on the physics analysis of interest. In this thesis, the event generator employed to generate the QCD $2 \rightarrow 2$ process is Pythia8 [65] using the A14 tune [66] and the NNPDF23LO particle distribution function (PDF) [67]. It simulates incoming and outgoing particles of the hard scatter event, initial- and final-state radiations as well as the hadronisation process. The generated final state particles are the input of the subsequent step, the detector simulation.

The detector simulation is responsible to simulate the interactions between the final state particles and the ATLAS detector. It computes trajectories of particles, their interaction with the matter of the detector and their decays. In ATLAS, the Geant4 [68] simulation toolkit is used to simulate the full detector. Combination methods such as AltfastII, which combines a fast parametrised calorimeter simulation (FastCaloSim [69]) with Geant4, are ten times faster than the full detector simulation. In this thesis, signal events of the leptophobic Z' model simulated by AltfastII are used. The model consists of two partons in the final state. The signal samples were generated for mediator masses of $m_{Z'} = 600$ GeV and $m_{Z'} = 1$ TeV with a coupling of $g_q = 0.02$.

After the detector simulation, the simulated detector hits are transformed into measurable quantities of the ATLAS detector. In addition, pileup events can be added during this digitisation step.

5.2 Jet Reconstruction

The collimated streams of particles, which result from quarks and gluons undergoing the hadronisation process, deposit energy in the calorimeters of the ATLAS experiment. These energy deposits are reconstructed as jets. Different definitions of jets exist and depending on the underlying physical process of interest, the most suitable version is used. The jet reconstruction procedure has a crucial impact on the identification of known resonances decaying into jets as well as in a search for new physics [70].

Jet reconstruction at the ATLAS experiment is primarily based on three-dimensional topological clustering of calorimeter cell signals [71]. These *topo-clusters* are seeded from cells whose energy deposit is larger than 4σ, where σ is the total cell noise, which consists of the electronics and the pileup noise. In Run-2, the cell noise is dominated by noise from pileup. Neighbouring cells are added to a topo-cluster when the cell energy passes 2σ. This process is repeated until all seed clusters are constructed. In addition, all cells neighbouring a seed cell with energies larger than zero are taken into account. Cells with negligible energy deposits far from crucial cell entries are removed to suppress the topological noise. In general, the calibration of topo-clusters is performed at the electromagnetic (EM) scale to correct the jet energy scale to that of particles produced in electromagnetic showers. In addition, topo-clusters can be calibrated using the local cell weighting (LCW) method which adjusts the energy response for hadronic showers with the help of MC simulations of charged and neutral pions at different energies and in all regions of the ATLAS calorimeter system.

ATLAS typically reconstructs jets with the anti-k_t [72] algorithm which uses topo-clusters as input. The reconstruction algorithm employs EM topo-clusters for small-R jets and LCW topo-clusters for large-R jets. R is the parameter for the radius in $\eta \times \phi$ of the reconstructed jet cone. Small-R jets are used for QCD jets induced by a quark or a gluon. These jets are reconstructed with $R = 0.4$. $R = 1$ is used for the reconstruction of jets from hadronical decays of massive particles such as $W^{\pm}$, Z and Higgs bosons and the top quark. These jets are called large-R jets. The parameter of the latter type of jets is chosen such that the decay products are collected into the same cone.

Jets which are reconstructed with the anti-k_t algorithm for $R = 0.4$ using simulated stable particles after the hadronisation as input are called *truth* jets. Muons, neutrinos and non-interacting particles are not included in this definition. Therefore, the energy scale of truth jets is at particle-level.

5.3 Jet Calibration

After the jet reconstruction, jets have to be calibrated to correct for several effects. The calibration steps for jets at EM scale are shown in Figure 5.1. The goal of the calibration is to correct the four-momentum of the jets such that the jet energy

agrees with the one of truth jets.

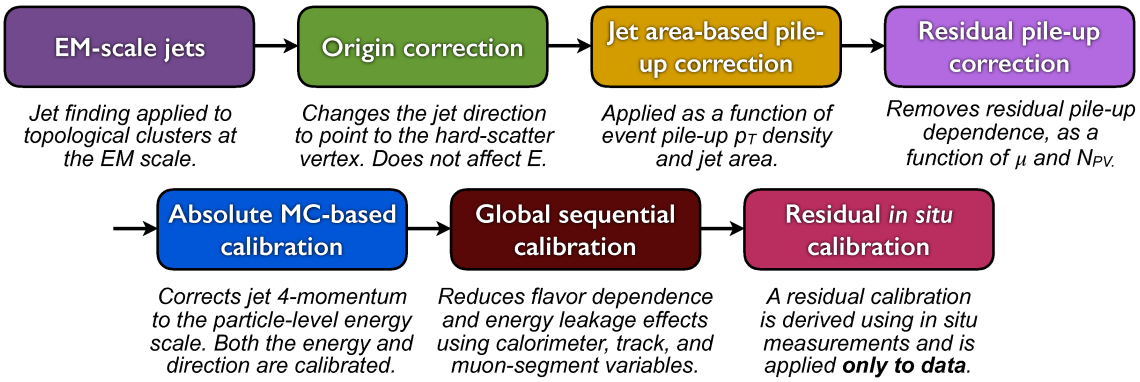


Figure 5.1: Flow diagram representing the calibration steps for EM scale jets which are applied to the four-momentum of each jet. Taken from [73].

The first step in the calibration procedure is the *origin correction*. Initially, jets are reconstructed at the origin of the coordinate system of the ATLAS detector. Now, the four-momentum of each jet is recalculated with respect to the associated hard-scatter vertex. This correction keeps the jet energy constant. The contributions from pileup are corrected in two ways: the *jet-area-based correction* [74] and the *residual pileup correction*. The former removes the contribution from pileup using the average energy density ρ in the event and the jet area $\mathcal{A}_{\text{jet}}$. The latter parametrises the influence of the number of reconstructed primary vertices N_{PV} and the average number of interactions per bunch crossing μ on the p_{T} of a jet. The total pileup correction can be written as

$$p_{\text{T}}^{\text{corr}} = p_{\text{T}} - \rho\mathcal{A}_{\text{jet}} - \alpha\,(N_{\text{PV}} - 1) - \beta\mu \tag{5.1}$$

with the correction coefficients α and β. Linear fits of the jet p_{T} as a function of N_{PV} and μ determine the parameters α and β. Next, the simulation-based calibration corrects the *absolute jet energy scale* (JES) and jet direction by comparing reconstructed jets and truth jets. This substantial discrepancy results from absorbed or undetected particles in the hadronic showers. The average energy response, defined as the ratio $E^{\text{reco}}/E^{\text{truth}}$, sets the inverse jet energy calibration factor which corrects the four-momentum of a reconstructed jet to the truth scale on average. In addition, the correction using the signed difference $\text{sgn}(\eta^{\text{reco}})\left(\eta^{\text{reco}} - \eta^{\text{truth}}\right)$ accounts for biases in the jet η reconstruction. Then, the JES is further corrected using the *Global Sequential Calibration* (GSC) which accounts for the dependence of the jet responses using five track, calorimeter and muon-segment variables [75]. The last step of the calibration procedure includes a *residual in situ calibration* which corrects for the differences in the jet response between data and MC simulations. The calibration factors are obtained by exploiting the p_{T} balance of different physics objects in the transverse plane depending on their p_{T} spectra. Jets at low p_{T} are calibrated using Z bosons decaying into electrons or muons. Photons are used for higher p_{T} jets,

whereas multijets are employed for the calibration of high-p_{T} jets. For each method, the difference in the data-to-MC ratio between reconstructed jets and well-calibrated objects, against which the reconstructed jets recoil, is modelled using second-order polynomial splines. The in situ calibration curve is then defined as a combination of the above mentioned curves favouring the method with the greatest precision in the overlapping regions. This calibration step is applied to data and corrects influences of the detector which are not modelled in simulations.

5.4 Trigger-Object Level Analysis

One possibility to search for physics beyond the Standard Model at the LHC is to study the decay of a new boson into two quark jets. This new mediator particle would interact with the quarks inside the colliding protons and decay back into two SM quarks, leaving a dijet signature in the detector. The jet with the highest p_{T} in the event is called *leading jet*, while the second-highest p_{T} jet is referred to as *subleading jet*. The invariant mass of the decay product (m_{jj}) would form a resonance at the mediator mass above the predictions of the Standard Model. The energy of the LHC beam defines a kinematic limit to producing very high mediator masses [76, 77]. This thesis focuses on searches for mediator masses in the low mass regime, which are limited because the event rate at the LHC exceeds the recording capability of the ATLAS detector.

The ATLAS trigger system selects only a small fraction of all events for later offline analysis. This selection process is based on high-p_T objects such as muons, electrons, photons, jets, τ leptons and missing transverse momentum. Events which do not pass the selection cut are not recorded. Hence, the trigger system plays an important role concerning which events are available for later analysis. The trigger thresholds are the harshest for physics events containing jets because events from QCD processes are the most frequent. These high-rate processes require high trigger thresholds and limit the statistics of the search for signal events in the low-mass regime.

Triggers with a low energy threshold have to be prescaled in order to record the full event information. Prescaling means that out of all events passing the trigger only the inverse fraction of the prescale is kept and stored. The prescales of the single-jet trigger limit the statistics for searches below this mass threshold, which is shown in Figure 5.2. Only a fraction of the events with a leading jet p_T below 380 GeV is being kept, which results in a drastic decrease in statistics of several magnitudes in the low-mass regime.

The *Trigger-Object Level Analysis* approach was developed to circumvent these statistical limitations of searches for dijet resonances in the low-mass regime. TLA stores only physics objects which are reconstructed online within the HLT system instead of using the event information from the total detector read-out. Therefore, the size of an event reduces to around 5% of the total detector event size, which amounts to approximately 1.6 MB. This allows a recording of events at a high HLT

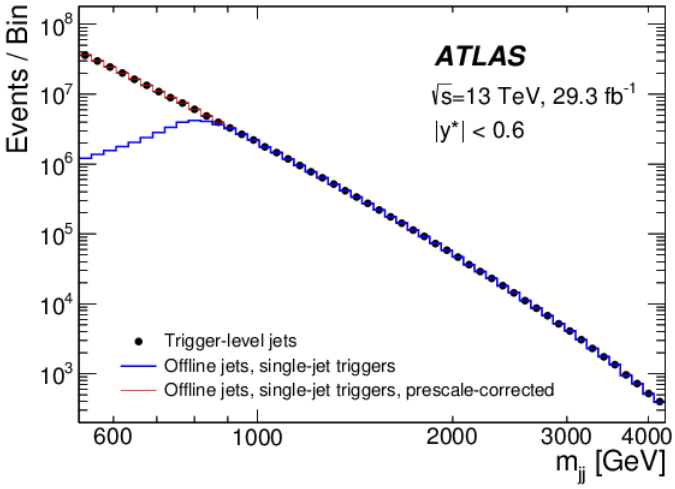


Figure 5.2: Invariant mass distribution for dijet events which pass the single-jet trigger (blue), which pass the single-jet trigger and are corrected for their prescale (red) and for dijet events using trigger-level jets (black). Prescales induce a loss of sensitivity of several magnitudes. Taken from [78].

rate, while only using a small fraction of its total bandwidth. The data for the TLA analysis approach is stored in a dedicated recording stream.

The Data Scouting Stream

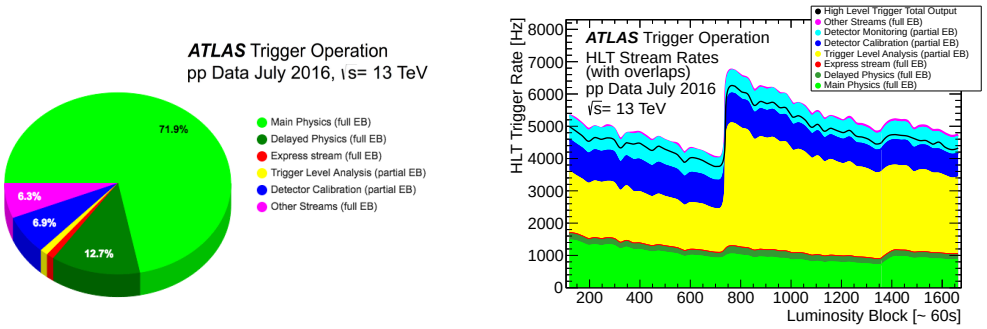
Data recording for TLA at the ATLAS Data Scouting (DS) stream started in 2015. Compared to the Main stream, it saves a reduced amount of all the event information including the four-momenta of online reconstructed objects and additional calorimeter-based variables used for the jet calibration. However, the access to the direct output of the detector including tracking information is no longer possible.

The main limitations at the ATLAS detector originate from the transfer bandwidth, processing time and storage space of offline data. The bandwidth

$$\text{bandwidth} = \text{event size} \times \text{trigger rate} \tag{5.2}$$

depends strongly on the trigger threshold, which determines the trigger rates, and the size of each event.

The reduction of the event size and partial event building (EB) result in a small impact of only 1% of TLA on the total bandwidth (see Figure 5.3a) despite the performance at high trigger rates. The trigger rates of the HLT for different physics streams are shown in Figure 5.3b. While the Main physics HLT rate amounts to approximately 1 kHz, TLA is able to record events at a much higher rate. In conclusion, TLA is able to restore the loss of sensitivity down to masses of approximately 500 GeV, while only occupying a small fraction of the total bandwidth.



(a) (b)

Figure 5.3: (a) Contribution to the total available bandwidth of different physics streams at the HLT for a fill taken in July 2016 with a peak luminosity of $\mathcal{L} = 1.02 \cdot 10^{34}$ cm^{-2} s^{-1} and a peak pileup of $\mu = 35$. (b) Stream rates at the HLT as a function of the number of luminosity blocks for different physics streams for a fill taken in July 2016 with a peak luminosity of $\mathcal{L} = 1.02 \cdot 10^{34}$ cm^{-2} s^{-1} and an average pileup of $\mu = 24.2$. Taken from [79].

Jet Calibration in TLA

The jets saved in the DS stream are called *trigger jets*, while jets that were reconstructed with the full event information are referred to as *offline jets*. The reconstruction of jets in the trigger system is performed as close as possible to the offline jet reconstruction explained in Section 5.2.

The calibration procedure for trigger jets [78] is similar to the one applied in offline analysis described in Section 5.3. In the HLT jet reconstruction, no tracking information is available such that calibrations requiring this input like origin correction and residual pileup correction are not applied. The track-based variables used in the offline calibration during the GSC are replaced by the trigger jet energy fractions in the calorimeters and the minimum number of topo-clusters, which contain 90% of the trigger jet energy. Before in situ calibrations are carried out, the energy of trigger jets is scaled to the one of offline jets by calculating the p_{T} response, which is defined as the p_{T} ratio of a trigger jet to its corresponding[1] offline jet. This calibration step is only applied to data. The in situ calibration curve of the final step is, compared to the offline calibration, not derived by the combination of spline functions but rather by a polynomial fit of the jet p_{T} response on a logarithmic p_{T} axis.

[1] Trigger jets and offline jets are matched within a distance of $\Delta R = 0.4$ if the corresponding offline jet exists.

Event Selection

The dijet event selection for TLA is based on the one used in [8]. Pileup jets are usually suppressed by track-based jet features which are not available in TLA. Therefore, the two leading jets are required to have $p_{\mathrm{T}} > 85$ GeV because the influence of pileup cannot be neglected for low-p_{T} jets. They are further required to lie within the pseudorapidity range of $|\eta| < 2.8$ in order to avoid the forward detector region, where the jet energy resolution increases. Additional requirements for the leading and subleading jets depend on the employed L1 trigger. In this thesis, the jets are selected by the L1 J100 trigger. In Figure 5.4, the efficiency curves of four different L1 trigger are shown. The leading jet p_{T} is required to have an efficiency of more than 99.5% of the L1 J100 trigger[2], which sets its transverse momentum requirement to $p_{\mathrm{T}} > 220$ GeV. In addition, the L1 J100 trigger applies a cut of $E_{\mathrm{T}} > 100$ GeV for each jet.

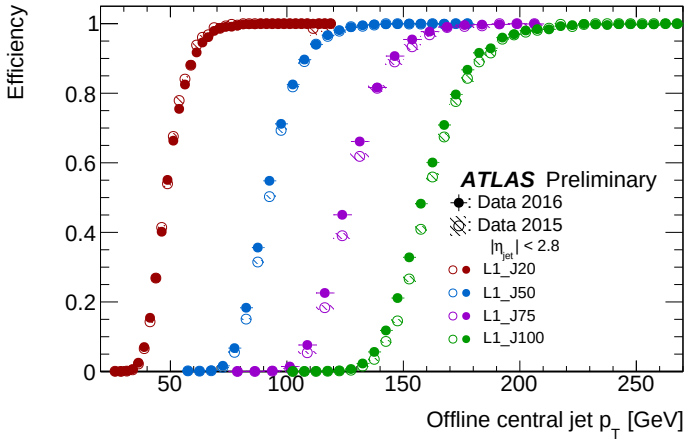


Figure 5.4: Efficiency curves of different L1 trigger as a function of the leading jet p_{T}.

Furthermore, the two leading jets are required to have an angular separation of $|y^*| < 0.6$, where $y^* = (y_1 - y_2)/2$ is the mean of the rapidities of the two trigger jets with the highest p_{T}. In addition, MC simulations are usually generated in slices of the leading truth jet p_{T} and each slice is assigned an individual scaling factor. After the addition of pileup events corresponding to measurements in data, one of the two leading jets might be a pileup jet with a higher transverse momentum and thus, a wrong scaling factor might be assigned to this event. Therefore, a MC quality cut of $0.6 < \bar{p}_{\mathrm{T}}/p_{\mathrm{T}}^{\mathrm{truth}} < 1.4$, where $\bar{p}_{\mathrm{T}}$ is the mean of the two leading trigger jet transverse momenta, is applied.

This thesis uses 819.2 pb^{-1} of proton-proton collisions at $\sqrt{s} = 13$ TeV which were recorded in 2016 by the DS stream of the ATLAS detector.

[2] The efficiency of a trigger is determined by a comparison of the number of events that were triggered to the total amount of events which is given by a reference.

6 Quark-Gluon Tagging

Quark and gluon jet discrimination is a powerful tool in searches for physics beyond the SM, which motivates the desire to label the flavour of a jet. For this thesis, quark-gluon tagging with the help of machine learning techniques on an event-by-event basis by using different discrimination variables of trigger jets in dijet events is performed.

6.1 Introduction

For many different physics models, it is important to understand the properties of jets which are either initiated by a quark or a gluon. Whether it is to test QCD predictions or to extract a signal of interest from a background dominated sample, the separation of quark and gluon jets has many applications. With the help of suitable jet properties, it is possible to apply cuts on data which enhance, e.g. the quark content of a given sample. This approach is especially useful in hadron colliders where a large fraction of the jets produced in hard-scatter processes is initiated by gluons. The signals of interest in, e.g. dijet searches are often dominated by quarks, while the background is gluon-dominated.

Given that there is no hadron-level definition of quark and gluon jets [80], any jet flavour tagging has to involve a connection between the initial parton and the jet itself [81]. In this thesis, the definition of quark and gluon jets is based on the event record of the applied parton shower event generator.

Most quark-gluon tagging techniques are predicated on QCD features based upon the fact that quarks and gluons carry different colour charges. The quark has a single colour index, while the gluon carries two colour indices. This property is expressed in different colour factors for quarks and gluons of $C_F = 4/3$ and $C_A = 3$. The coupling strengths of quarks and gluons to an additional emitted gluon, therefore, differ because the bremsstrahlung process is directly proportional to the coupling strengths. Consequentially, the hadronisation process of quarks and gluons differs such that the resulting jets possess different properties, which can be identified and exploited in order to achieve a good quark-gluon separation. The quark and gluon jet discrimination of previous studies often made use of jet properties such as particle multiplicities, energies and angular distributions [81–88] and in addition, there are several discriminating variables investigated at the LHC [10, 89–93]. A selection of these variables is discussed below.

QCD predicts that the mean multiplicities of quark and gluon jets differ by a factor of $C_A/C_F = 2.25$. Thus, gluon jets are expected to have a more than twice as high particle multiplicity of any type of particle (charged or neutral) compared

to quark jets. However, the theoretical prediction and the experimental results are expected to differ due to varying definitions of the investigated objects. While the measurements in detectors are based on hadrons, QCD relates its calculations on partons and does not account for energy-momentum conservation and higher order perturbative terms. Indeed, this difference was determined by measuring the mean charged particle multiplicity ratio of gluon over quarks jets, which increases with increasing jet energy [94, 95]. Furthermore, it can be shown that there is almost a factor of two difference between the mean charged multiplicities of gluons and light (u,d,s) quark jets at small rapidities and with energies around 40 GeV [96]. Thus, the measurement of particle multiplicities establishes a good foundation for a quark-gluon discrimination process.

Additionally, quark and gluon jets tend to have differing energy distributions. For example, the scaled energy distribution of a charged particle in a jet, which is also called fragmentation function, is sensitive to quark-gluon separation. This property founds on the fact that gluon jets are suppressed in the high momentum region with respect to quark jets [94, 96]. The reason for this behaviour is that jets which have a higher momentum fraction are more likely initiated by a valence quarks [12, p. 196 ff.]. Quark jets retain most of their initial momenta after a gluon radiation and therefore, possess more fast particles on average than gluon jets. A quark in a gluon-initiated jet has to be created during the showering process and tends thus to have a lower momentum fraction.

Studies for flavour tagging also often rely on angular distributions of jets such as jet width, the family of radial moments [97], angularities [98] and the pseudorapidity. In general, gluon jets are wider than quark jets and therefore, are observed in larger angular cones in the detector. With increasing energy, both jet types become narrower [94]. The linear radial moment also called girth, which is based on the sum of transverse momenta of the jet constituents weighted by its distance from the jet axis, combined with the charged particle multiplicity induces a gluon rejection of 90%, while 60% of the quark jets are kept [81]. Jet properties based on the location of the jet inside the detector such as the pseudorapidity also differ for quark- and gluon-initated jets. The pseudorapidity of quark jets extends to higher values compared to gluon jets of equivalent energy and thus, quark jets are suppressed in the central region [94]. This feature is caused since more forwards jets tend to have a higher total jet energy and hence a higher momentum fraction where the PDF of valence quarks dominates.

Compared to electron-positron[1] colliders, quark-gluon tagging in proton-proton colliders is hindered due to initial-state radiation, multi-parton interactions and pileup. However, the radiation pattern of quark and gluon jets depends on the colour connection between the initial partons and the jets and thus, the measured samples of many event topologies of collider experiments possess the property of being either quark or gluon enriched [99]. The label "enriched" refers to the dominating jet flavour in the Born-level process [100]. For example, the production of a Z boson

[1] The antiparticle of the electron is called positron.

in combination with a jet is quark enriched, while dijet final states contain a larger fraction of gluon jets compared to quark jets [9]. These enriched data samples can be used to extract differences in the jet properties of quarks and gluons.

In conclusion, the colour-dependent behaviour of quark and gluon jets can be exploited in order to achieve a good quark-gluon separation using different track- and calorimeter-based variables. The separation power between quark and gluon jets according to Equation 3.8 is shown in Figure 6.1 as a function of the jet p_{T} for selection of these variables. The number reconstructed tracks (n_{trk}) as well as the track-based energy-energy-correlation (EEC) angularity [101] with $\beta = 0.2$ show the best separation. The EEC angularity is a 2-point energy correlation function and measures the kinematic properties of an additional radiated jet. Thus, quark jets tend to have smaller values compared to gluon jets according to the colour fraction C_A/C_F.

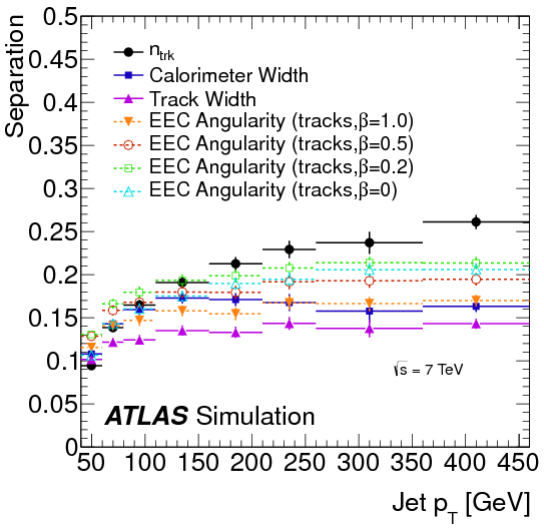


Figure 6.1: Separation power of different track- and calorimeter-based discrimination variables as a function of the jet p_{T}. Taken from [90].

6.2 Quark Gluon Tagging in TLA

The quark and gluon jet discrimination in TLA is limited by the available discrimination variables because the events of trigger jets only contain information about the reconstructed objects of the HLT and additional calorimeter-based variables. Thus, no tracking information is available. This makes a number discrimination variables such as the charged track multiplicity n_{trk} and the track width unusable. Therefore, the tracking variables have to be replaced by calorimeter-based discrimination variables in order to perform quark-gluon tagging in TLA. Below, a number of possible variables is presented.

The multiplicity of jets can be described by the N_{90} variable, for which the E_{T} of all jet constituents are added in descending order and all constituents are counted until 90% of the total trigger jet energy is reached. The leading jet N_{90} distributions of signal and background events using a dijet MC simulation can be seen in Figure 6.2a. The mean number of jet constituents in signal events is smaller compared to the ones in background events because in general, gluon jets tend to have more jet elements. Thus, N_{90} seems to have a good discrimination power. With increasing jet energy, the average N_{90} value for both signal and background events decreases.

In Figure 6.2b, the η distribution of the leading jet is depicted. A different behaviour for signal compared to background events can be observed. There tend to be more gluon jets in the central region compared to in the forward region. This means that the more forward a jet in a dijet event, the more quark-like it is. This feature is induced by the fact that more forwards jets tend to have a higher total jet energy and hence a higher momentum fraction which makes it more likely for them to be initiated by a quark.

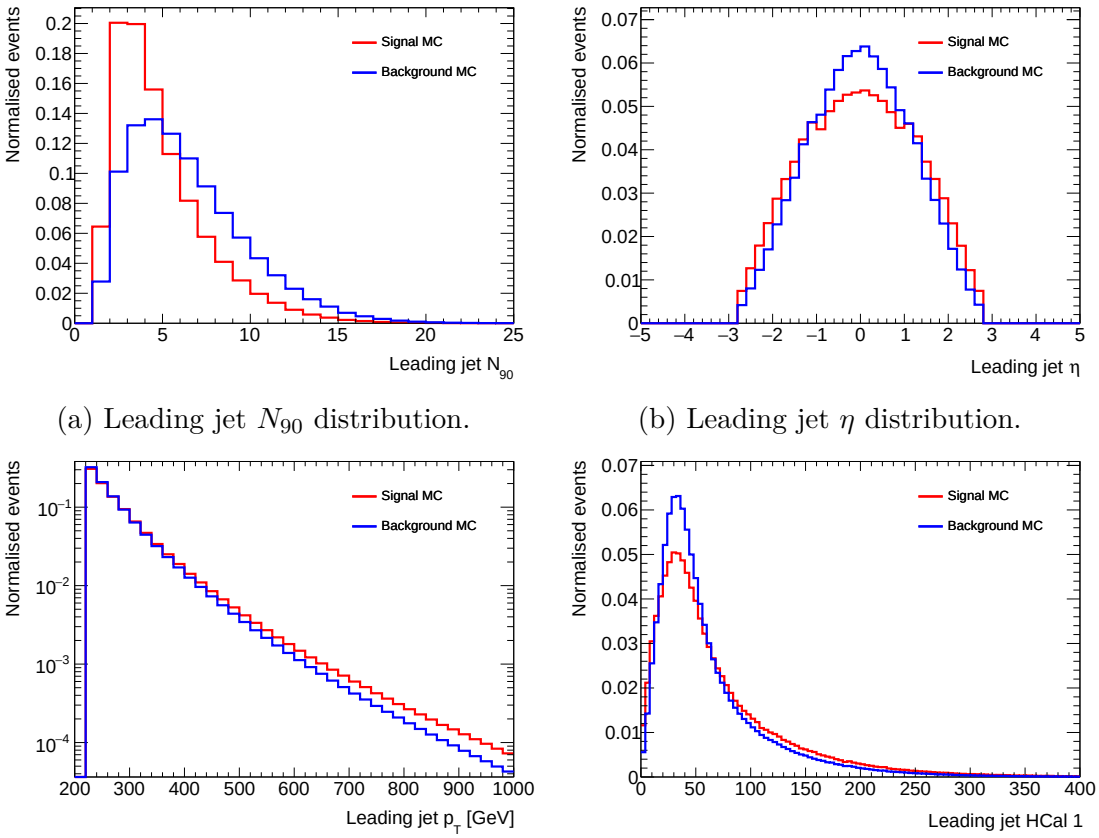
The energy spectra of quarks and gluon jets are dominated by soft gluons and differ because these gluons tend to have a smaller momentum-fraction. Thus, dijet signal events tend to have more high-p_{T} jets in comparison to background samples, which can be seen in Figure 6.2c. This increases also the possibility for quark jets to penetrate further into the calorimeter system until the HCal. Therefore, the parton from which the jets originate can be identified using the shower depth of the jets, which deposit energy in the calorimeter layers of the ATLAS detector.

The depth of a jet can be measured by comparing the energy deposits in the layers of the calorimeter system of the detector. Since quark jets have on average a higher p_{T}, these jets tend to hit layers which are located further away from the interaction vertex. This behaviour can be seen in Figure 6.2d, where the normalised distributions for signal and background events of the leading jet energy deposit in the first hadronic calorimeter layers are depict. Here, the energy deposits of the first layers of TileCal and HEC are added in order to generate an artificially induced energy layer which covers the whole TLA pseudorapidity range of $|\eta| < 2.8$. The long tail of the signal distribution towards higher energies indicates that a quark dijet event deposits in general more energy in this layer compared to the dijet background composed of gluons.

In summary, despite the constraint of the available event information, there are a several possible discrimination variables on hand, of which N_{90} appears to have the highest separation power.

6.3 Reweighting of Monte Carlo

Different MC event generators for high energy proton-proton collisions follow varying models to describe the resulting particles and how they interact with the detector. This induces a potential bias of the applied generator on the quark-gluon tagging result. To assure that the generator in this work (Pythia8) induces no unwanted



(a) Leading jet N_{90} distribution.

(b) Leading jet η distribution.

(c) Leading jet p_{T} distribution.

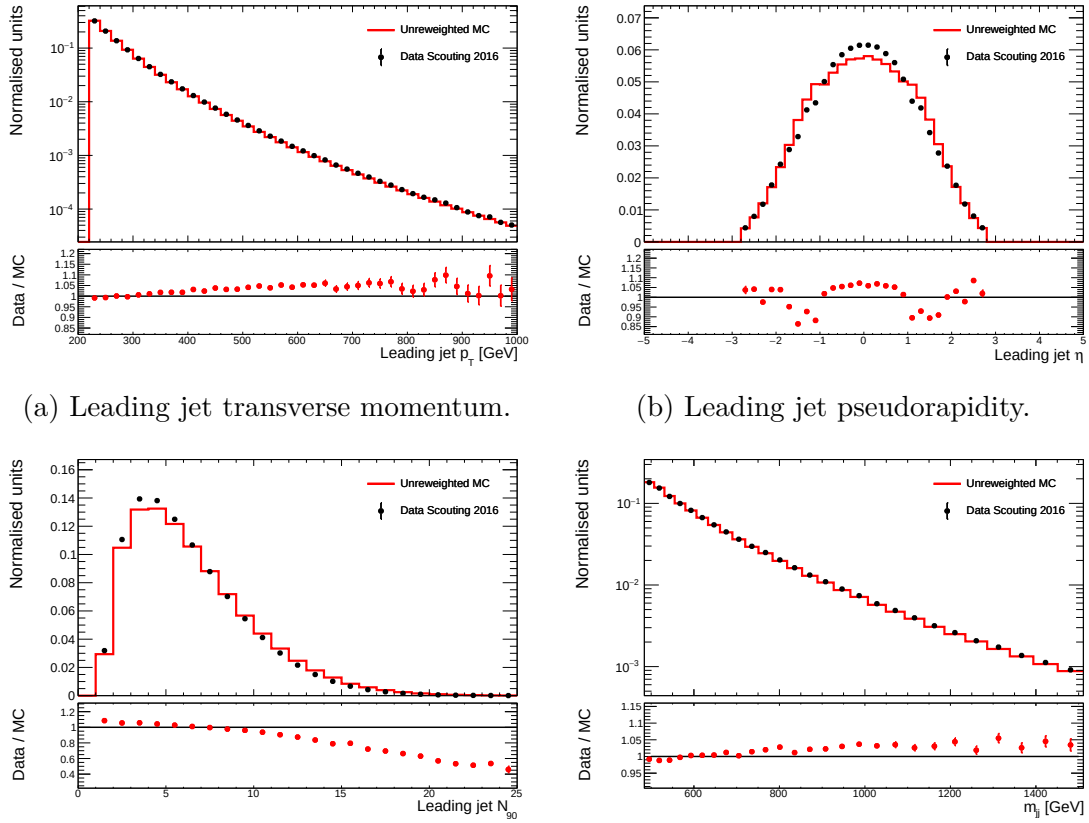
(d) Leading jet HCal layer 1 distribution.

Figure 6.2: Comparison of the different normalised distributions of quark dijet events (signal) and dijet events containing at least one gluon jet (background) using PYTHIA8 dijet MC simulations for jets with $|\eta| < 2.8$.

MC dependent behaviour, the distributions of certain variables of trigger jets in data and MC are compared. Variables such as N_{90}, pseudorapidity η, transverse momentum p_{T} for the leading jet and the invariant mass m_{jj} are depicted in Figure 6.3 and show a mis-modelling between data and MC simulations. Subleading jet variables show a similar behaviour.

The concrete reason for this mis-modelling is unknown and the distributions might look different using other MC generators. A possibility to correct this mis-modelling is to reweight the MC simulation in such a way that the distribution of interest matches the data sample. This ensures that the choice of MC generator does not affect our training process.

The reweighting procedure is based on a correction of the ratio of data over MC of different variables, from which the reweighting factors are determined. Afterwards these factors are assigned to each event separately. Then, this factor is applied to the respective event by a multiplication with the event weight. There are two possible reweighting procedures: bin-by-bin and fit reweighting. The former determines the



(a) Leading jet transverse momentum.

(b) Leading jet pseudorapidity.

(c) Leading jet N_{90}.

(d) Invariant mass m_{jj}.

Figure 6.3: Normalised distributions of trigger jets for data (black) and MC simulation (red) after applying TLA cuts. The lower plot presents the data-to-MC ratio.

reweighting factor through a bin-wise evaluation of the data-over-MC ratio, while the latter fits a polynomial function of varying order to the ratio and its behaviour is then used to identify the corresponding reweighting factor. The choice of reweighting procedure along with the number of free fit parameters has a non-negligible impact on the correction factors of the MC simulation. This can be illustrated as follows.

For the fit reweighting, a function f is sought, whose fit values agree well with the values of the N data points in the histogram. The form of the fit function is selected to be a polynomial, whose order is chosen on the basis of the χ^2 fit-value[2]. This value is based on difference between the input distribution $y_i(x_i)$ and the fit prediction $f(x_i; \mathbf{p})$ for a set of fit parameters $\mathbf{p}$ weighted by the error for each bin σ_i and is given by

$$\chi^2(\mathbf{p}) = \sum_{i=1}^{N} \left(\frac{y_i - f(x_i; \mathbf{p})}{\sigma_i} \right)^2 . \tag{6.1}$$

[2] The χ^2 fit is evaluated using the function ROOT:TF1::GetChisquare().

The χ^2 value is then minimised in order to find the optimal set of fit parameters $\mathbf{p}$, whose error sizes strongly depend on the second derivative of $\chi^2(\mathbf{p})$.

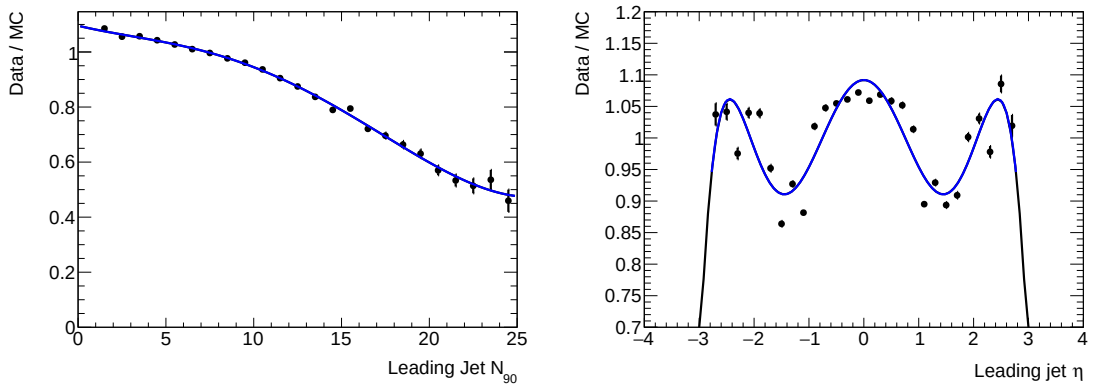
In the first step, a constant function is fitted to the data-over-MC ratio of a selected distribution. The order of the polynomial is then increased step by step in such a way that a new free fit parameter accompanied by a higher order term is added in each step, while previous orders of the polynomial including their parameters are preserved. The reduced χ^2 values, which are defined by the ratio between χ^2 and the number of degrees of freedom n_f, of two subsequent steps are compared to the one of the current fit. This procedure is able to determine the optimal fit function for the data-over-MC ratio even for ratios with completely even or odd functional behaviour. Orders in which the χ^2/n_f value decreases less than 5% are omitted if the subsequent order results in a χ^2 drop of more than this threshold. If χ^2/n_f decreases less than 5% in two subsequent fitting steps, the fit reweighting procedure is stopped and the fit function, whose χ^2/n_f value improvement satisfied the chosen threshold, is chosen.

The reweighting procedure for the N_{90} variable is carried out first due to the fact that it seems to be a good discriminant for quark-gluon tagging. The fit result of the leading jet N_{90} distribution is shown in Figure 6.4a using the polynomial function $a_0 + a_1 \cdot x + a_2 \cdot x^2 + a_3 \cdot x^3 + a_4 \cdot x^4$ in the range $[0, 25]$. The χ^2/n_f value of this fit amounts to approximately 1.84. The reweighting procedure for the leading jet N_{90} variable is chosen to be a fit reweighting instead of a bin-by-bin reweighting because the former reflects the functional behaviour of the data-over-MC ratio well. Bin-by-bin reweightings should in general be avoided because statistical fluctuations of the distributions might then be taken into account.

The polynomial fit of the leading jet η distribution using the fit function $a_0 + a_1 \cdot x^2 + a_2 \cdot x^4 + a_3 \cdot x^6$ in the range $[-2.8, 2.8]$ is shown in Figure 6.4b. Here, the data-over-MC seems to follow a even functional behaviour and fit parameters of odd orders did not satisfy the required χ^2/n_f improvement. The final fit of the leading jet η distribution results in a χ^2/n_f value of ≈ 58.97. This value is more than 32-times larger than the one of the leading jet N_{90} distribution and therefore, a bin-by-bin reweighting procedure is applied for this variable.

In addition, the impact of the reweighting procedure of a certain variable on other kinematic variables has to be checked to account for possible correlations. This check controls as well whether the mis-modelling of the different variables arises from the same origin. The effect of the fit reweighting process of the leading jet N_{90} distribution on different variables can be seen in Figure 6.5. The upper left plot shows the leading jet N_{90} distribution for which the reweighting factors were determined. The reweighted MC simulation corrects the mis-modelling in this variable. In the other three plot of this figure, the induced change of the distributions for a selection of leading and subleading jet variables is shown. It can be seen that the fit reweighting process of the leading jet N_{90} variable has no significant effect on the other kinematic variables.

To assure that the discrepancies of leading and subleading jet N_{90} are not correlated, a two dimensional (2D) reweighting procedure is compared to two one di-



(a) Leading jet N_{90}. (b) Leading jet η.

Figure 6.4: Fit reweighting of the data-to-MC ratio of two leading jet distributions with different polynomial functions. The fit functions are shown in blue.

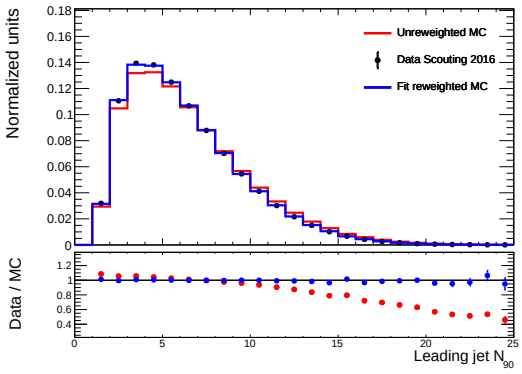
mensional (1D) reweightings applied after each other. For this purpose, a 2D fit of the leading and subleading jet N_{90} with a 2D polynomial is performed. This fit is compared to a 1D fit in which the leading and subleading N_{90} data-over-MC ratios are fitted separately and both reweighting factors are applied after each other for each event. The resulting reweighting factor is calculated as

$$\text{Factor}_{\text{res}} = \text{Factor}_1 \cdot \text{Factor}_2 \tag{6.2}$$

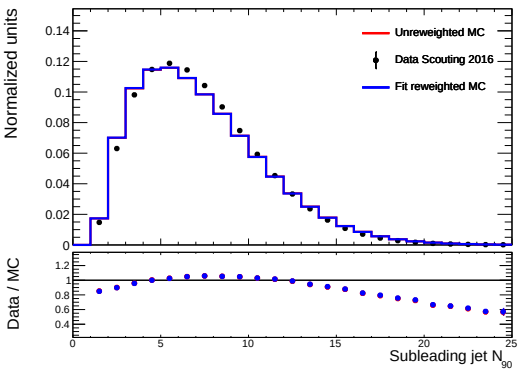
In Figure 6.6 the impact of both reweighting techniques on the leading and subleading jet N_{90} variable can be observed. Not sufficient statistics of a two dimensional technique compared to one dimensional one result in inaccuracies in the 2D fit in the high N_{90} range because there are only a few events produced. Comparing both subfigures, it becomes clear that the 1D reweighting performs in a comparable way compared to the 2D reweighting without a low statistics impact. In addition, this comparison shows that the mis-modelling in N_{90} for leading and subleading jet are not or only slightly correlated. Finally, the one dimensional reweighting for N_{90} was chosen.

The discrepancies in the data-over-MC ratio for the N_{90}, η and p_{T} distributions of leading and subleading jets are corrected in five steps in the following order:

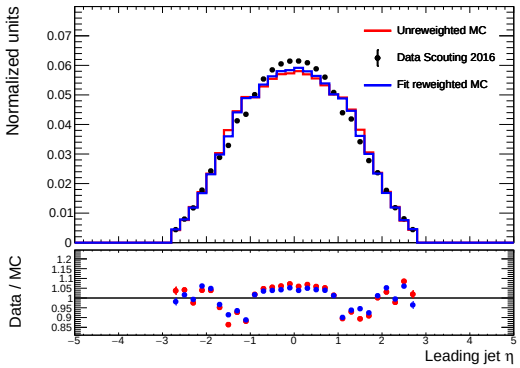
1. Fit reweighting of leading jet N_{90}
2. Fit reweighting of subleading jet N_{90}
3. 2D bin-by-bin reweighting of leading jet and subleading jet η
4. Fit reweighting of leading jet p_{T}
5. Fit reweighting of subleading jet p_{T}



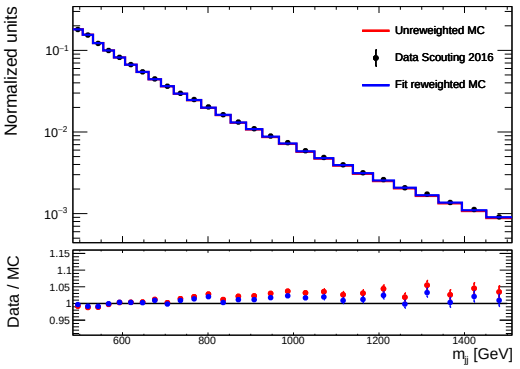
(a) Leading jet N_{90}.



(b) Subleading jet N_{90}.



(c) Leading jet η.



(d) Invariant mass m_{jj}.

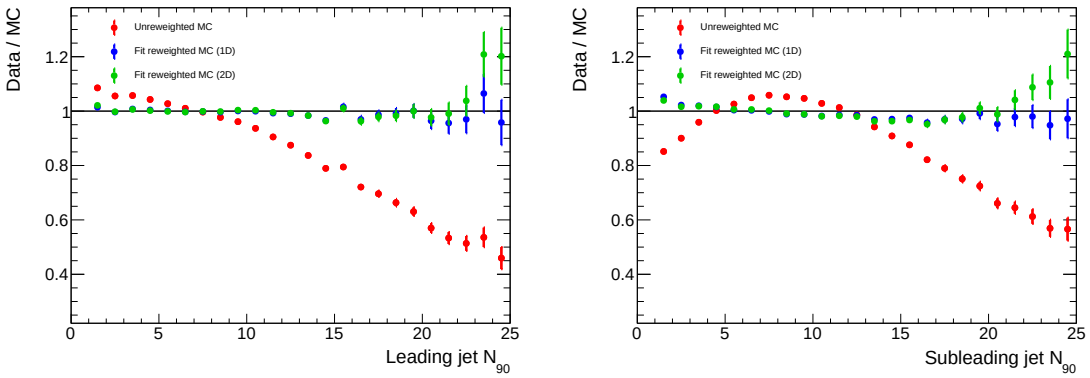
Figure 6.5: Unreweighted (red) and fit reweighted (blue) normalised distributions of trigger jets for MC simulation compared to data (black) after applying TLA cuts. The lower plot presents the data-to-MC ratio for the unreweighted and reweighted MC simulations.

The correction for η is performing using a 2D bin-by-bin reweighting because these two leading and subleading jet variables are highly correlated and no clear functional behaviour was visible. The final result of the reweighting procedure can be seen in Figure 6.7. The mis-modellings of the MC simulation compared to data are diminished such that the reweighted variables now follow the distributions of the data sample.

Further, it was tested whether the total reweighting procedure resulted in any change in other trigger jet variables. The invariant mass m_{jj} of a dijet event depends on the p_{T}, η and ϕ variables of the two leading jets and is given by

$$m_{jj}^2 = 2p_{\mathrm{T},1}p_{\mathrm{T},2}\left[\cosh(\eta_1 - \eta_2) - \cos(\phi_1 - \phi_2)\right], \tag{6.3}$$

where the labels 1 and 2 represent leading and subleading jet variables respectively. This feature induces a correction of the m_{jj} distribution when p_{T} and η are reweighted. In other variables such as the leading jet ϕ and the pseudorapidity

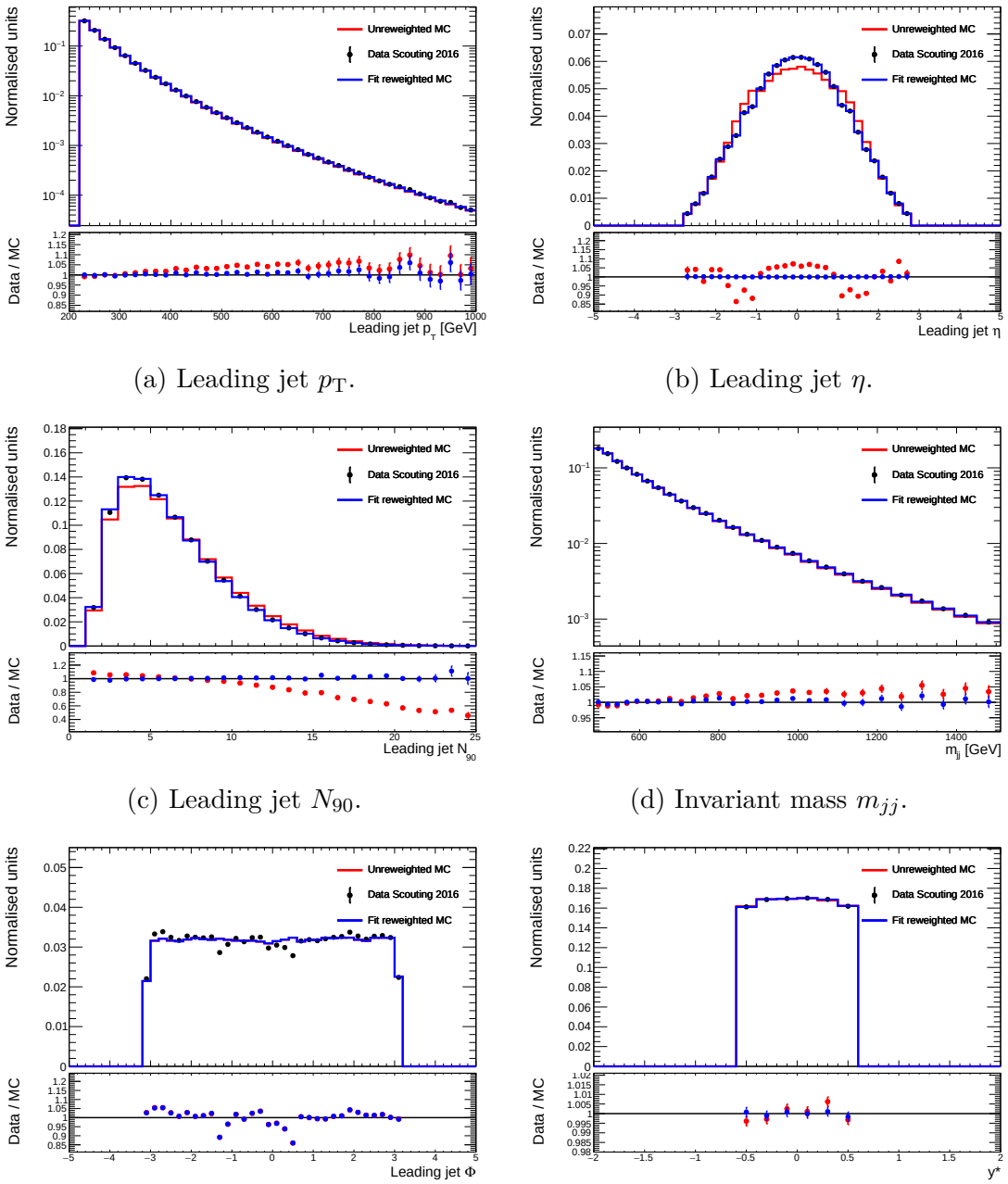


(a) Leading jet N_{90}.

(b) Subleading jet N_{90}.

Figure 6.6: Data-to-MC ratio for the unreweighted (red) and fit reweighted leading and subleading jet N_{90} distribution. The fit reweighting in 1D is shown in blue; the one in 2D is shown in green.

difference of leading and subleading jet $\Delta\eta$, no significant effect was observed (see Figure 6.7). Thus, the resulting distributions can now be used as discrimination variables for a quark and gluon jet discrimination.



(a) Leading jet p_{T}.

(b) Leading jet η.

(c) Leading jet N_{90}.

(d) Invariant mass m_{jj}.

(e) Leading jet ϕ.

(f) y^*.

Figure 6.7: Final result of the reweighting procedure of different kinematic variables. The unreweighted and final reweighted distributions are shown in red and blue respectively. The data is shown in black. The lower plot presents the data-over-MC ratio for the unreweighted and reweighted MC simulations.

7 Performance of Quark-Gluon Tagging in TLA

This thesis engages in multivariate classification methods based on machine learning techniques in order to achieve a good quark-gluon separation on an event-by-event basis. The two leading jets of the investigated dijet signal model are quark jets, while the reducible QCD background is chosen such that at least one of the two leading jets is gluon-initiated. The choice of the training model combined with well-selected discriminating variables is crucial for a good model performance. In order to study the quark-gluon separation, the whole data set is split in training, testing and application data sets following a 60% - 30% - 10% division. The learning process is executed on MC simulations because these samples include the parton identifications of each jet.

7.1 Discrimination Variables

As previously discussed in Section 6.2, only the event information of trigger jet is available as possible discrimination variables in TLA. In this section, the separation power of a number of available variables is discussed.

Transverse momentum p_{T} and pseudorapidity η of a jet allow the differentiation of the dijet events in energy ranges as well as regions of the detector. This enables the possibility for the model to learn the nature of quark and gluon jets with different kinetic properties. Here, the azimuth angle ϕ is excluded as a discrimination variable because the number of jets in dijet events is uniformly distributed over this distribution.

Since the number of calorimeter cells forming a gluon jet is on average larger than the ones forming a quark jet, N_{90} is used as a discrimination variable. The N_{90} distribution is shifted towards higher values as p_{T} increases and thus, choosing a combination of p_{T} and N_{90} as input variables for a model may account for this correlation.

The separation, as defined in Equation 3.8, for the above mentioned input variables of the leading and subleading jets is shown in Table 7.1. The ranking is determined by their discrimination power. The N_{90} variables have the highest discrimination power of this selection of discriminating variables. A comparison with the separations in Figure 6.1 yields that N_{90} does not perform as good as the discrimination variables depicted in the previous chapter. However, a combination of the variables in Table 7.1 might enhance the discrimination power because this work focuses on the separation of signal and background events, not on differentiating

between quark and gluon jets for each jet separately.

	Separation
Subleading jet N_{90}	0.09136
Leading jet N_{90}	0.07188
Subleading jet p_{T}	0.01927
Leading jet p_{T}	0.00265
Subleading jet η	0.00130
Leading jet η	0.00078

Table 7.1: Input variable separation. The top variable is ranked best.

In addition, the energy deposits of different calorimeter layers of ECal, TileCal and HEC can be used to investigate the depth of jets in the ATLAS calorimeters. This is possible, since quark jets tend to have less constituents and a more narrow radiation pattern. As a consequence, more quark jets tend to be detected in parts of the calorimeter system with a higher cumulative interaction length compared to gluon jets with equal jet p_{T} and η. This feature can be exploited by using energy layer information as discrimination variables in the quark-gluon tagging.

The interaction lengths of the material in the different calorimeter layers as a function of the absolute pseudorapidity is shown in Figure 7.1. It does not contain the different layers of the ECal. Additionally, the layers of the FCal are not considered in this analysis because their location exceeds the TLA pseudorapidity range of $|\eta| < 2.8$.

Since the total number of available layers is too large to be fed into a model, seven artificial layers are defined to compensate the original information by combining layers with similar cumulative interaction lengths:

1. ECal0 ≔ Presampler ECal
2. ECal1 ≔ ECal layer 1
3. ECal2 ≔ ECal layer 2
4. ECal3 ≔ ECal layer 3
5. HCal1 ≔ TileCal Layer 1 + HEC Layer 0
6. HCal2 ≔ TileCal Layer 2 + HEC Layer 1-2
7. HCal3 ≔ TileCal Layer 3 + HEC Layer 3

This choice of new layers allows to investigate the depth of a jet almost independently from its η value. To further improve the quark-gluon tagging, the information of the seven energy layers could be combined with the kinematic variables p_{T} and η for both leading jets. The calorimeter width of the jets is not available in the HLT data of 2016.

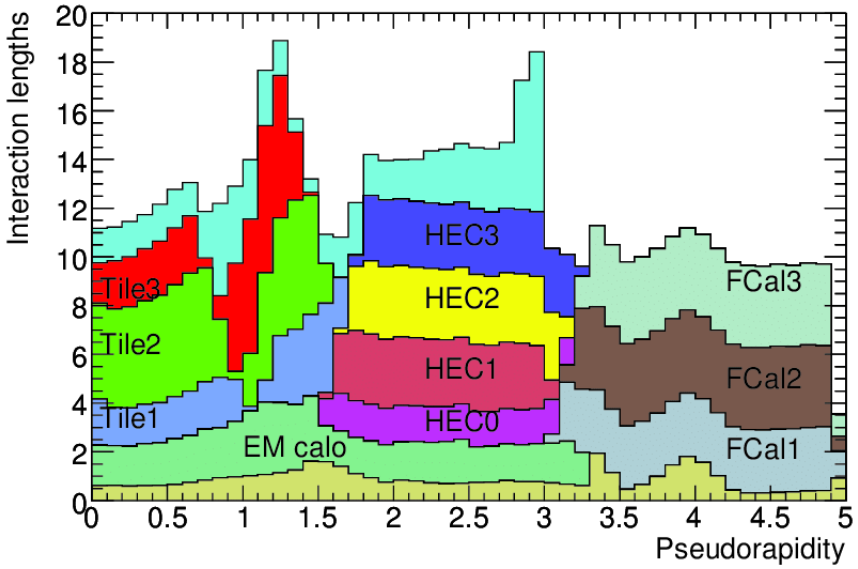


Figure 7.1: Cumulative interaction length of the material as a function of the absolute pseudorapidity in front of the ECal, across the ECal, in each hadronic layer and the total amount at the end of the calorimeter system. Presampler layers (up to $|\eta| < 1.8$) and additional material in front of the calorimeter system are shown in beige; the light blue layer indicates the additional amount of material in front of the first layer of the muon spectrometer (up to $|\eta| < 3.0$). Taken from [52, p. 112].

7.2 Comparison of Multivariate Classification Methods

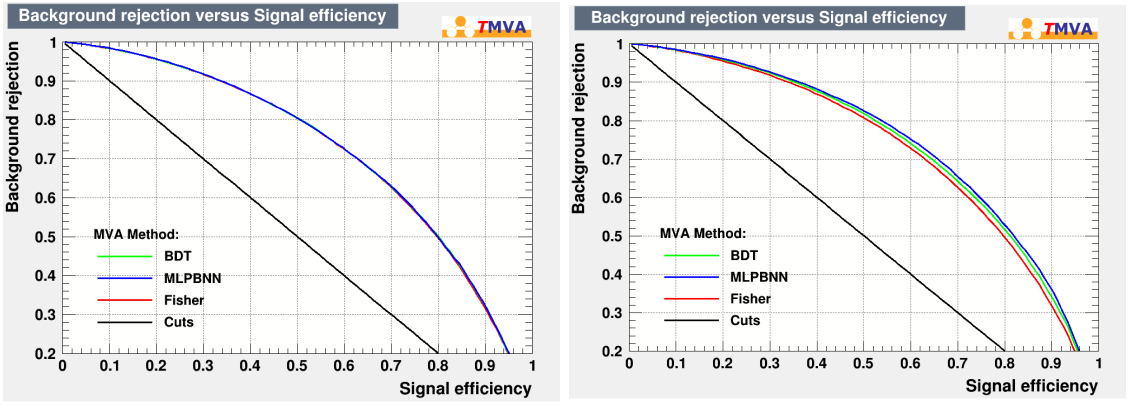
Depending on the structure of the training data set and the set of discrimination variables, different multivariate analysis techniques might result in a better quark and gluon jet discrimination performance. To investigate the best set-up for the quark-gluon tagging, the performance of different models using the same set of training data is compared. For the comparison, a rectangular cut method, a linear discriminant analysis (Fisher), a BDT and a MLP are chosen. For the models where the optimised settings differ from the TMVA default values [34] the chosen settings and variable transformations are displayed in Table 7.2. The Fisher discriminant was booked with default values and no variable transformation was employed.

A good performance is always a tradeoff between a flexible model, which is able to describe complex underlying properties of the data, and a model which shows indications of overtraining. In order to avoid overtraining and to reduce the model complexity, the maximal depth for the BDT is chosen to be 3 and the Bayesian extension for the MLP is turned-on. First, a selection of four variables including the p_{T} and N_{90} distributions for the leading and subleading jet is chosen as input

Cuts	BDT	MLP
VarTransform = None	*VarTransform* = None	*VarTransform* = Norm
FitMethod = MC	*NTrees* = 1000	*NCycles* = 600
VarProp = FSmart	*MinNodeSize* = 3%	*HiddenLayers* = N+5,N
	MaxDepth = 3	*NeuronType* = tanh
	nCuts = 10	*TrainingMethod* = BFGS
		TestRate = 5
		ConvergenceTests = 10
		UseRegulator = True

Table 7.2: Configuration setting for the MLP method. Default settings are chosen for all not mentioned options.

of the different models. Afterwards, the η distributions for both leading jets are added as input variables in order to compare the model performance as a function of the number of input variables. The result of the training processes in form of ROC curves is shown in Figure 7.2.



(a) N_{90} and p_{T}.

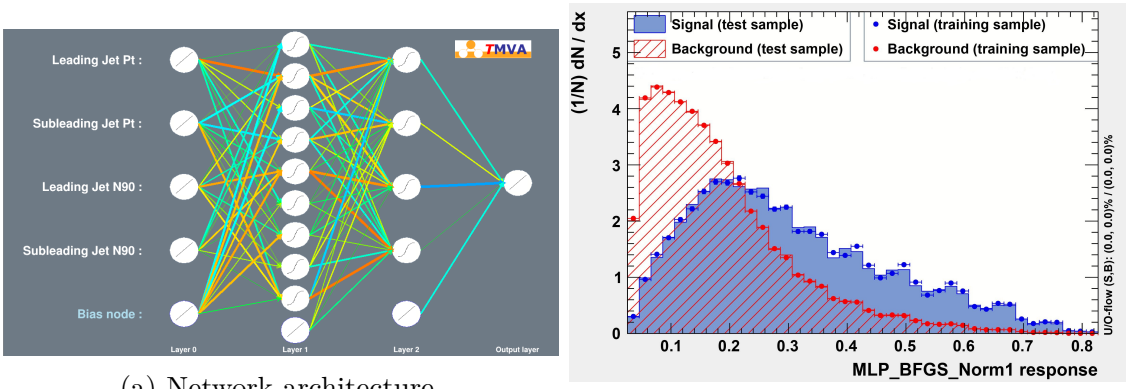
(b) N_{90}, p_{T} and η.

Figure 7.2: ROC curves of a selection of trained models using different input variables.

Figure 7.2a depicts the ROC curves of the trained models using the four input variables mentioned above. The rectangular cut classifier performs worst compared to the other three models because it simply applies cuts on every variable individually. The signal and background distributions are not clustered so that the cut-based classifier is not able to separate them. The Fisher discriminant performs comparable (AUC = 0.723) to the non-linear classifiers BDT and MLP (both AUC = 0.725) because the information input is focused in only four variables from which most are either not or only linearly correlated. A selection of 2D distributions out of which correlations can be read is depicted in the Figure B.1.

In Figure 7.2b, the performance of the different classifiers using the mentioned six variables can be observed. While the Fisher discriminant performs similar (AUC = 0.725), the BDT and the MLP are able to combine the additional information of the pseudorapidity with the transverse momentum and N_{90} to improve the quark-gluon tagging performance. As a result, the AUC value increases for the BDT to 0.735 and for the MLP to 0.743. In this thesis, the MLP, therefore, is chosen as the optimal model used for the quark and jet gluon discrimination. The MLP classifier and its performance is further elaborated in the following sections.

7.3 Tagging using p_{T} and N_{90}



(a) Network architecture.

(b) Overtraining check.

Figure 7.3: Control and performance figures of the MLP provided by the graphical user interface (GUI) of TMVA. (a) Three-layer network architecture with four input variables. (b) MLP output distributions for signal and background events from the training and test samples.

The architecture of the multilayer perceptron introduced in Section 7.2 is shown in Figure 7.3a. It consists of an input layer, two hidden layer and an output layer. The number of nodes in the input layer is composed of p_{T} and N_{90} of the leading and subleading jet as input variables and one bias. The hidden layers consist of nine and four nodes respectively and additionally, one bias node for each inner layer. The number of output layers is directly connected to the number of classes in a classification problem. Thus, there is one output node which determines the class of a dijet event either to be signal or background. The links between the nodes illustrate the weight coefficients of the trained network. The weights were normalised to the range $[-1, 1]$ prior to this visualisation. The thickness of a connecting arrow depends on the absolute value of the weights assigned to it: the larger the absolute value of the weight, the thicker the arrow connecting the two nodes. The colour of each connecting line is determined by the weight itself such that normalised weights with a value of -1 are depict in blue, while red is assigned to a weight of 1. Zero weights are shown in green. The colours for intermediate values follow accordingly.

In Figure 7.3a, there are some more important connections with larger absolute weights visible but no input variable or specific weight stands out. This allows the conclusion that the MLP combines all input variables as well as the bias more or less equally in order to achieve the optimal quark and gluon jet discrimination.

Figure 7.3b shows the response of the multilayer perceptron to signal and background events compared for the training and test sample. It allows a comparison of the network performance of the training and test sample and therefore, provides a possibility to identify overtraining. The response distributions of the training and test sample agree very well for signal and background, respectively, such that it is unlikely that this network is heavily overtrained. In addition, the numerical data of this figure can be used to retrieve the optimal neural network cut (NNCut), which separates signal and background events with a maximal significance gain.

In order to determine the significance gain, the impact of the quark-gluon tagging on signal and background events has to be compared. Therefore, a selection of different NNCuts is applied to two Z' dijet samples as well as to the QCD background separately. Then, the significance

$$\text{significance} = \frac{\text{S}}{\sqrt{\text{S+B}}} \tag{7.1}$$

is calculated using the number of signal (S) and backgrounds (B) events in the full width at half maximum (FWHM) range of the Z'-signal respectively. The gain in significance after the quark-gluon tagging for different NNCuts can be seen in Figure 7.4.

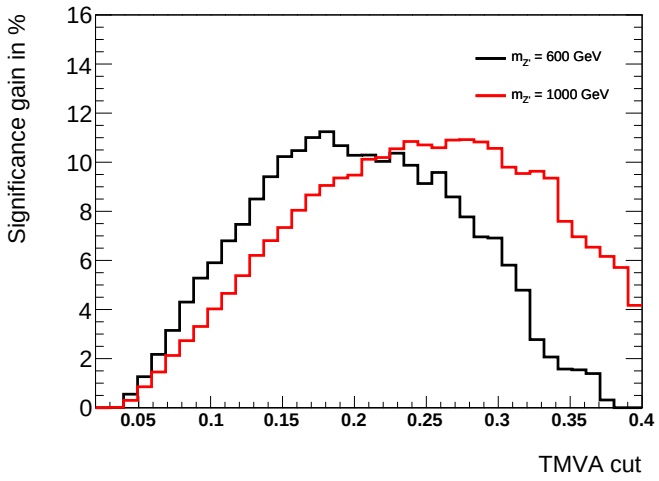


Figure 7.4: Significance gain in % of the signal FWHM for two Z' signal samples with $m_{Z'} = 600$ GeV and $m_{Z'} = 1$ TeV over a range of NNCuts of the MLP network with N_{90} and p_{T} as discriminating variables.

The optimal NNCut for the 600 GeV dijet sample is identified to be 0.18 and it yields a significance gain of approximately 11.2%. The significance gain resulting

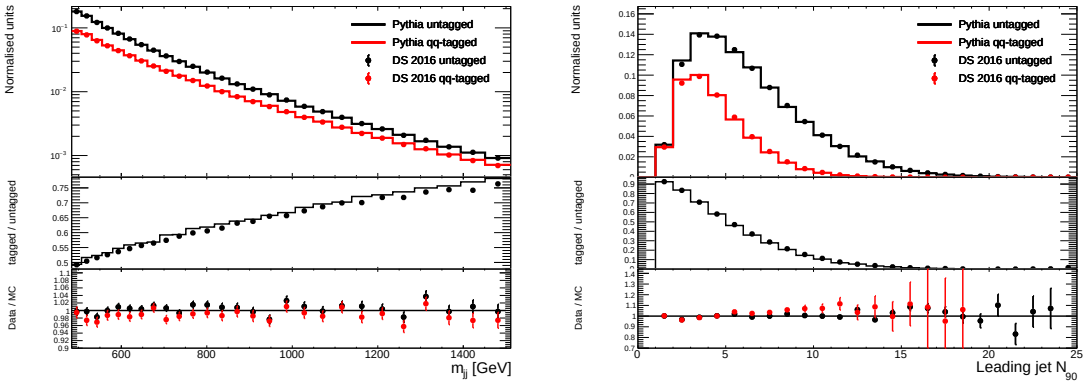
from this cut amounts to approximately 9.1% for the signal sample with $m_{Z'} = 1$ TeV. As soon as new signal samples with the invariant masses between 600 GeV and 1 TeV are produced, it is possible to determine the optimal NNCut as a function of m_{jj}. Due to the fact that the goal is to enhance dijet signal samples in the low-mass regime, the NNCut at this moment is chosen to be 0.18.

A comparison of the m_{jj} distribution of the DS data sample of 2016 and the MC simulation before and after the quark-gluon tagging is shown in Figure 7.5a. The MC simulation is illustrated as a solid line, while circular markers represent the data sample. The distributions before the tagging are shown in black; the tagged events can be seen in red. In the upper panel, the normalised distributions for data and MC before and after the tagging process are compared and their ratios for the distributions before and after the tagging can be seen in the middle part of the figure. The uncertainties are computed via binomial error calculation [102, p. 26 ff.] because the numerator is a subset of the denominator in each case. Depending on the invariant mass of the dijet event, the quark-gluon tagging rejects 22% to 50% of the QCD background with a trend of decreasing rejection for higher invariant masses. At an invariant mass of 600 GeV, approximately 45% of the events get rejected, while the background rejection amounts to 66% around $m_{jj} = 1$ TeV. The increase of the tagged-over-untagged ratio in the middle panel agrees with the observations in Figure 6.2c because quark jets tend to have higher transverse momenta and thus, a higher invariant mass. The lower part of the figure depicts the data-over-MC ratios with and without tagging, which show a discrepancy in the depicted m_{jj} range. The quark-gluon tagging seems to have a different impact on data and MC such that in the course of the tagging, more events from data are rejected compared to the ones from MC simulations. Thus, the data-over-MC ratio after the flavour tagging seems to be smaller than the original one. The reason for this behaviour might lie in the small but still existent mis-modelling of the data from MC simulation even before the tagging. Although the concrete origin of the discrepancy is unknown, the deviation is in most cases smaller than 2% and agrees essentially with the size of the uncertainties.

In Figure 7.5b, the comparison of the tagged and untagged leading jet N_{90} distributions demonstrates the fact that the N_{90} distribution for quark jets tends to peak at lower values. The quark-gluon tagging removes only approximately 10% to 20% of the events in the three lowest N_{90} bins and towards higher N_{90} values, the background rejection increases further. This demonstrates that signal events with two quark jets are less likely to be rejected compared to gluon jet.

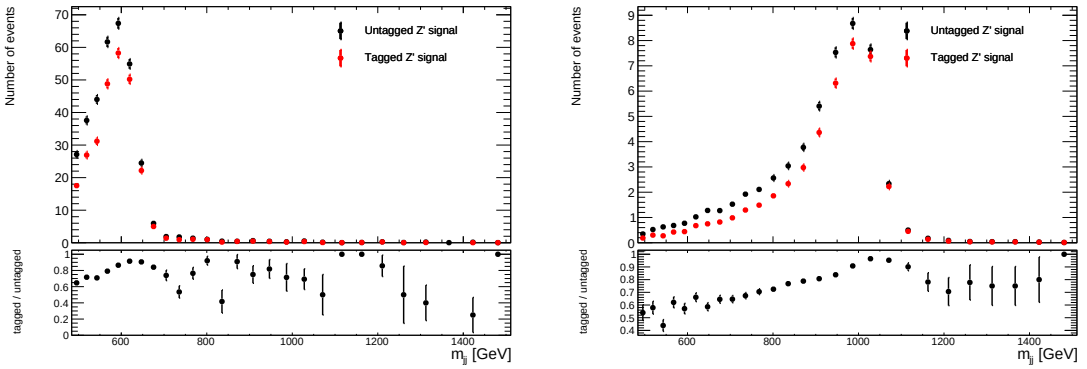
The quark-gluon tagging is also applied on the Z' dijet sample with $m_{Z'} = 600$ GeV and the resulting invariant mass distributions can be seen in Figure 7.6a. The upper part shows the number of events before and after the tagging process, while the tagged-over-untagged ratio is depicted in the lower part of the figure. The error for this ratio is as well computed using the binomial error calculation. The invariant mass distribution peaks, as expected, at the mass of the Z' particle of 600 GeV. After the quark-gluon tagging, approximately 70% to 90% of the signal in the FWHM region are sustained. In Figure 7.6b, the tagging application of the

invariant mass distribution for the Z' signal with $m_{Z'} = 1$ TeV is shown. With about 80% to 95% in the FWHM region, more signal events remain compared to the lower mass signal sample. However, only around 34% of the QCD background are rejected in this mass range. A summary of the quark-gluon tagging performance of the three-layer multilayer perceptron using N_{90} and p_{T} of the two leading jets as input variables is shown in Table 7.3.



(a) Invariant mass m_{jj}. (b) Leading jet N_{90}.

Figure 7.5: Comparison of different distributions of data (circles) and MC simulation (solid line) before (black) and after (red) the quark-gluon tagging using N_{90} and p_{T} as input variables. Upper part: normalised distributions. Middle Part: ratio of distributions before and after the tagging for data and MC. Lower part: data-over-MC ratio before and after the tagging.



(a) $m_{Z'} = 600$ GeV, $g_q = 0.02$ (b) $m_{Z'} = 1$ TeV, $g_q = 0.02$

Figure 7.6: Invariant mass distribution of different Z' dijet signals before (black) and after (red) tagging including the tagged-over-untagged ratio.

	600 GeV	1 Tev
QCD rejection	$\approx 45\%$	$\approx 34\%$
Z' efficiency	$\approx 80\%$	$\approx 90\%$

Table 7.3: QCD rejection and Z' signal efficiency after the quark-gluon tagging in two different FWHM mass regions using N_{90} and p_{T} as discriminating variables.

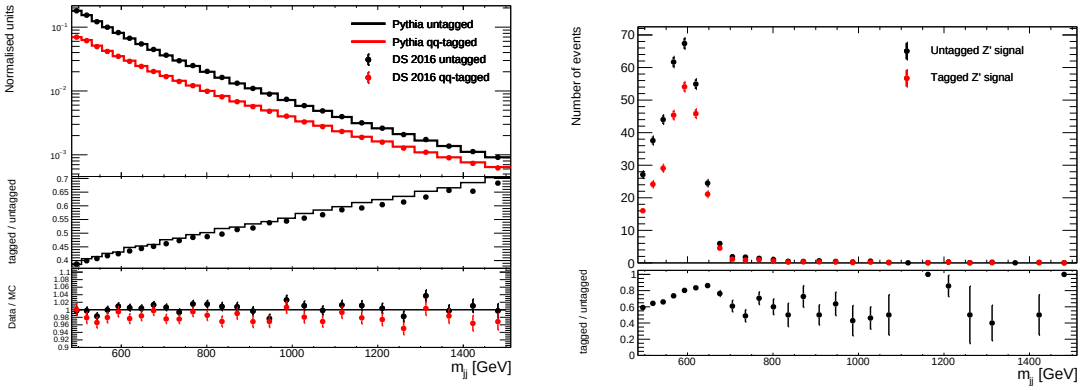
7.4 Tagging using p_{T}, η and N_{90}

In this section, the training of the MLP is based on the same discrimination variables as in the previous section, but as additional information, the pseudorapidity η of the two leading jets is added. This results in six input variables of the neural network and thus, the number of input nodes adds up to seven including the bias. The configuration settings for the network were chosen just as in Section 7.3 and can be seen in Table 7.2. The network architecture is again specified as $(N + 5, N)$, thus, the number of hidden layers stays constant, while the number of nodes in each layer is adjusted according to the number of input variables. A selection of evaluation figures of the trained network can be seen in the Appendix B. The AUC value becomes 0.743 such that a higher significance gain of the signal is expected.

After the training, the optimal NNCut for this network was identified as 0.21 by comparing the significance gains of the two tagged Z' dijet signal samples compared to the MC simulation for a variety of selection cuts. Here, the significance gain amounts to 15.2% for the 600 GeV and to 12.5% for the 1 TeV sample (see Figure B.3). Compared to the neural network in the previous section, the significance gain increases by taking the η distributions into account, which by themselves do not possess a high separation power (see Table 7.1). One can thus conclude that the neural network links the discrimination variables so that the total separation excels the individual ones by taking correlation among these variables into account. The corresponding figure of the leading jet N_{90} distribution is shown in Figure B.4a. The characteristic tagged-over-untagged ratio with a peak toward small N_{90} values can be seen there as well. However, a higher optimal value of NNCut increases the total amount of rejected signal events yielding a lower efficiency especially in the first bins of the N_{90} distribution because the quark jets in a signal event tend to have a lower average N_{90} value.

Figure 7.7a depicts the result of the quark-gluon tagging for the m_{jj} distributions of data and MC. The tagged-over-untagged ratio of the distributions of data and MC shows the expected rising shape towards higher invariant masses. At an invariant mass of 600 GeV, approximately 57% of the events are rejected after the tagging process, while the QCD background rejection amounts to around 55% at $m_{jj} \approx 1$ TeV. Compared to the tagging performed in the previous section, the background efficiency decreases by a factor of approximately 1.24 and 1.20 in the FWHM regions of the Z' dijet samples with $m_{Z'} = 600$ GeV and $m_{Z'} = 1$ TeV.

The signal efficiency of the Z' dijet sample with $m_{Z'} = 600$ GeV after the quark-gluon tagging with the six discriminating variables is depicted in Figure 7.7b. In the range of the FWHM, it decreases to values between 65% and 85 %. Thus, the signal peak in this mass range is further reduced than in the study before by 6%. Table 7.4 displays the background rejections and signal efficiency for this set-up also including the results of the Z' sample with $m_{Z'} = 1$ TeV, which can be seen in Figure B.4b.



(a) (b)

Figure 7.7: (a) Comparison of the invariant mass distribution of data (points) and MC simulation (histogram) before (black) and after (red) the quark-gluon-tagging using N_{90}, p_{T} and η as input variables. Upper part: normalised distributions. Middle Part: ratio of distributions before and after the tagging. Lower part: data over MC ratio before and after the tagging. (b) Invariant mass distribution of the Z' dijet signal with $m_{Z'} = 600$ GeV and $g_q = 0.02$ before and after tagging including the tagged-over-untagged ratio.

	600 GeV	**1 Tev**
QCD rejection	$\approx 57\%$	$\approx 45\%$
Z' efficiency	$\approx 75\%$	$\approx 80\%$

Table 7.4: QCD rejection and Z' signal efficiency after the quark-gluon tagging in two different FWHM mass regions using N_{90}, p_{T} and η as discriminating variables.

7.5 Tagging using Energy Layer Information

In order to further improve the quark and gluon jet separation, the energy deposits in the energy layers defined in Section 7.1 for the two leading jets are included as additional discrimination variables into the training process. This results in $N = 20$ input variables for the MLP. The configuration settings for the neural network are chosen as depicted in Table 7.2 but instead of a $(N+5, N)$ network architecture, the size of the network was reduced to $(N-9, N-14)$. Since the network architecture includes two hidden layers with 11 and 6 nodes just as in the previous section, a comparison of the separation power of two networks with the same architecture but a different number of input nodes is possible. The goal is to investigate how a neural network handles additional information using a fixed network complexity.

The network architecture of the MLP is shown in Figure 7.8. There are some more important connections with larger absolute network weights visible but no input variable or specific weight stands out. This implies that the MLP considers all input variables for the training process although e.g. the subleading jet N_{90} variable has a larger separation than the leading jet p_{T}. The separation for all 20 discrimination variables is given in Table B.1. Although the separation for most energy layer variables exceeds the one for the η variables, the AUC value of the quark-gluon tagging amounts to 0.738, which is a smaller value compared to the one obtained in the previous section. This demonstrates that the performance of a MLP depends not only on the selection of input variables but also on the architecture of the network and other network parameters. Furthermore, a larger number of discrimination variables drastically increases the training time. Hence, a preselection of variables is required in order to balance the gain in separation power and the increase in training time. Here, a decrease in significance gain of the signal is expected compared to the previous section.

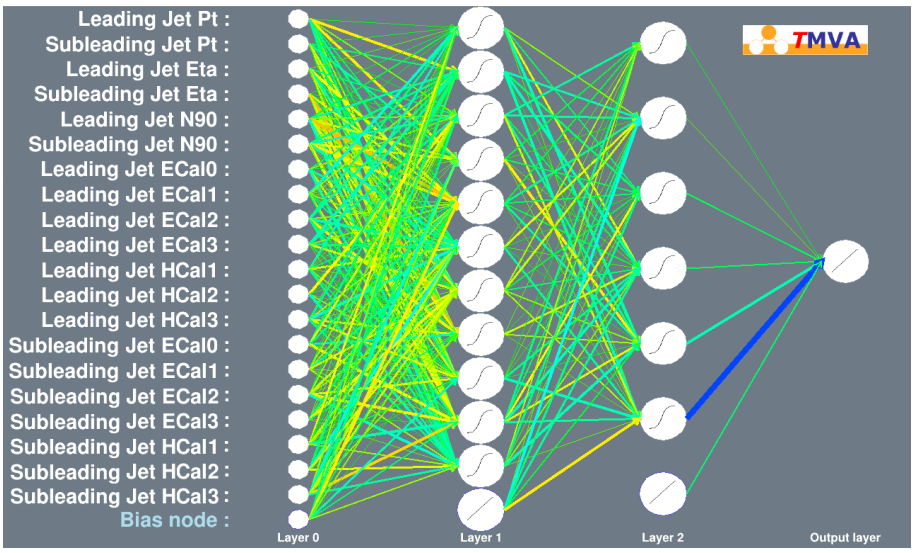


Figure 7.8: Three-layer network architecture with 20 input variables.

The significance gain of the two investigated signal samples as a function of the applied NNCut is shown in Figure 7.9. The optimal cut value for the $m_{Z'} = 600$ GeV signal sample is identified as 0.24 and it yields a significance gain of 15.1%. Compared to the tagging without the energy layer information, the significance gain degrades. This shows that the network is not able to increase the quark and gluon jet discrimination with the additional information of the energy deposits in the different layers of the calorimeter. However, Figure 7.8 suggests that all input variables are considered with similar importance. This implies that the network using energy layer information learns a different underlying property for quark and gluon jets compared to the network without the additional information.

The significance gain of the dijet signal sample with $m_{Z'} = 1$ TeV amounts to 13.0% using the optimal NNCut value for the lower mass signal sample. In comparison to the network using six discriminating variables, the significance gain increases for a signal in this mass regime. However, since the focus of this thesis is the mass regime of the invariant mass distribution below 1 TeV, the subsequent analyses omit the information of the energy deposits in different layers in the calorimeter system.

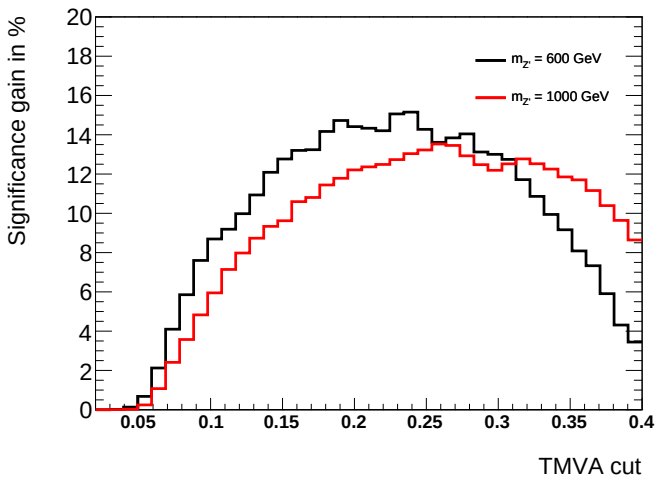


Figure 7.9: Significance gain in % of the signal FWHM for two Z' signal samples with $m_{Z'} = 600$ GeV and $m_{Z'} = 1$ TeV over a range of NNCuts of the MLP network with N_{90}, p_{T}, η and energy layer information as discriminating variables.

8 Statistical Evaluation of Quark-Gluon Tagging Performance

There are several statistical data analysis methods which can be used in order to determine the statistical significance of a result. In this thesis, the BUMPHUNTER algorithm is used to statistically analyse the performance of the quark-gluon tagging conducted in the previous chapter. The goal is to determine the impact of the tagging process on the required cross-section of a signal to be identified as one. The following analysis is based on a qualitative evaluation of the quark and gluon jet discrimination using the MC simulation as a background estimate.

8.1 Hypothesis Tests and P-Values

The goal of a statistical test can be to determine whether a deviation between a data set and a theoretical model is explainable by statistical fluctuations. If the deviation is not attributed to random chance, the result of such a test is called *statistically significant*. In order to verify or exclude a statistical significance, the data sample is compared to a so-called *null hypothesis* (H_0), which is the assumption of the background model fully describing the system. Therefore, these tests are called *hypothesis tests* [102]. Each hypothesis test includes a *test statistic* t, which assigns the deviation between data and H_0 a number. A famous hypothesis test is Pearson's χ^2 test, whose test statistic was already introduced in Equation 6.1, where $f(\mathbf{p})$ was the background hypothesis for a fixed set of parameters. The BUMPHUNTER algorithm [103, 104], which is described in detail in Section 8.2, is a hypothesis test using a different test statistic. Although different hypothesis tests use varying test statistics, all follow the principle that the larger the difference between data and the background distribution, the larger the test statistic.

After the choice of a test statistic, *pseudo-data* is generated which features H_0 as underlying background hypothesis. Then, the test statistic is calculated for each set of pseudo-data by comparing the new generated data sets with H_0. The results are combined to form the distribution of test statistics.

The next step involves the calculation of the *p-value*. The p-value, defined in the interval $[0, 1]$, is given by the probability of having a test statistic that is at least as big as the observed one assuming that H_0 is true:

$$\text{p-value} \equiv P(t \geq t_0 | H_0), \tag{8.1}$$

where t is a random variable following the distribution of test statistics and t_0 is the test statistic comparing the observed data set and the background distribution based on H_0.

The p-value can be interpreted as a false-positive probability, which is the probability of rejecting H_0 although it is true. This type of probability is also called *Type-I error* and it is given by

$$\alpha = \int_{t_{cut}}^{\infty} \rho_t(t)dt, \tag{8.2}$$

where ρ_t is the probability distribution function of t given H_0 and t_{cut} is a cut value that determines the threshold above which H_0 is rejected. The goal is to minimise α so that the false-positive probability becomes small. If the test statistic of a hypothesis test returns a p-value of x, then the discovery algorithm, which has the smallest α still declaring a discovery, would reject H_0 with a probability of x that H_0 is actually true. Thus, the p-value is not the probability of H_0 being wrong, but it is smallest α value of an algorithm that would reject H_0 [103].

Due to the different test statistics of different hypothesis tests, the resulting p-values of these tests might vary strongly. However, the resulting p-values do not contradict each other, but rather focus on different discrepancies. For a significant fluctuation of the data from H_0 in a certain bin of a histogram, there is a hypothesis test which would declare a discovery with a low p-value even if H_0 is true. This behaviour is called "look elsewhere effect" because each bin has its own possibility to be declared as discovery. Hence, in order to evaluate a deviation between data and H_0 in a proper way, many hypothesis test have to be considered simultaneously. A hypothesis test that considers all bins of the histogram in an equal way is called *hypertest*. The test statistic t of a hypertest is based on the smallest p-value of all considered hypothesis tests in such a way that t increases monotonically when the smallest p-value decreases.

8.2 Bumphunter Algorithm

The Bumphunter (BH) algorithm is a hypertest combining different hypothesis tests which focus on the discovery of deviations of different widths in different regions of a spectrum. In TLA, it is used to find localised excesses in the invariant mass spectrum that are caused by the resonant production of massive particles. The algorithm determines the most significant bump, which results from a deviation between the measured data and H_0, by scanning over the whole spectrum using windows of varying widths. The window size first includes two bins of the spectrum and it increases until the window extends to half the range of spectrum.

For every choice of window size and every location of the window in the spectrum, all bins in the current window $[n, m]$ are combined to define a new bin, whose numbers of expected (B) and observed (D) events are calculated via

$$B = \sum_{i=n}^{m} b_i, \qquad D = \sum_{i=n}^{m} d_i, \tag{8.3}$$

where b_i and d_i are the expected and observed events in the original bins. Afterwards, the p-value for the new bin is calculated using the localised hypothesis test, that focuses on this width and location. Since the BH algorithm assumes that the number of entries in each bin are following a Poisson distribution, the p-value is given by

$$\text{p-value} = \begin{cases} \sum_{n=D}^{\infty} \frac{B^n}{n!} \mathrm{e}^{-B} & D \geq B \\ \sum_{n=0}^{\infty} \frac{B^n}{n!} \mathrm{e}^{-B} & D < B \end{cases}, \tag{8.4}$$

where B and D are the number of expected and observed events in the new bin as defined above. The calculation of the p-value is repeated for every window size at each position of the spectrum.

Next, the BH test statistic is calculated according to

$$t = -\log\left(\min_i\{\text{p-value}_i\}\right), \tag{8.5}$$

where $\min_i\{\text{p-value}_i\}$ is the smallest p-value found in the step above. The functional behaviour of the negative logarithm function assures that t increases with increasing discrepancy.

The test statistic of Equation 8.5 is then used to determine t_0 by comparing the observed data to H_0. With the help of pseudo-data generated according to H_0, the BH p-value is calculated according to Equation 8.1. The uncertainty of the BH p-value

$$\sigma_{\text{p-value}} = \sqrt{\frac{p(1-p)}{N}} \tag{8.6}$$

is given by the standard error of a Bernoulli experiment, in which the p-value is identified as the probability p and, where N is given by the number of pseudo-experiments. Thus, increasing the number of pseudo-experiments decreases the uncertainty of the BH p-value. A BH p-value smaller or equal to 0.01 is setting the limit for a potential discovery.

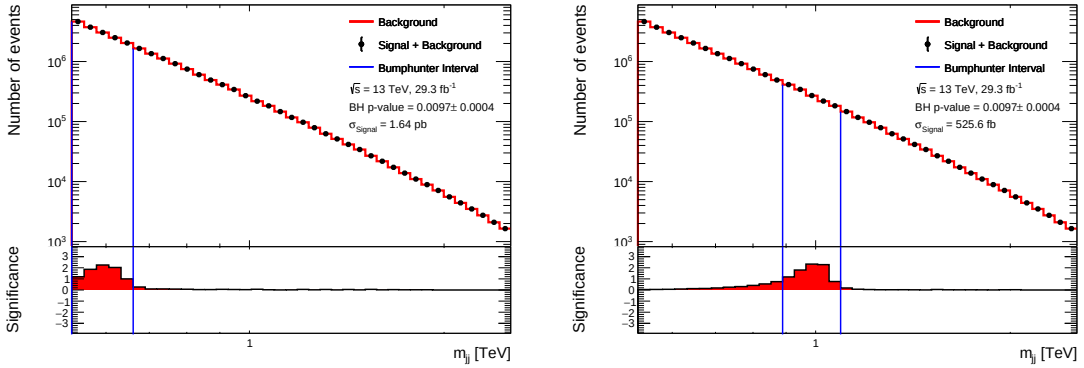
8.3 Study of Significance Improvement by Quark-Gluon Tagging

In order to detect potential signals, recorded data is compared to the predictions of a theoretical model, like the SM. In TLA, the SM background distribution is usually determined using a sliding-window fit [8]. Since a qualitative evaluation of the quark-gluon tagging should be studied, the background estimation is not required in this work. Instead, the MC simulation itself is used as the background estimate. Z' signal samples are artificially integrated into the MC simulation in

order to imitate signal excesses at different invariant mass values $m_{Z'}$. These newly created MC samples are used as input data samples for the BH algorithm.

The goal of this analysis is to determine the cross-section of a signal, that is required in order to be identified as statistically significant by the BH algorithm using a perfect background estimate H_0 at a luminosity of 29.3 fb^{-1}. The focus is especially on the impact of the background rejection induced by the quark-gluon tagging on the required signal cross-section. For that, the cross-sections of the Z' signal samples are increased until the BH algorithms identifies the signal as a potential discovery. 50,000 pseudo-data sets are generated for each investigated signal cross-section and are then used to calculate the respective BH p-value and its uncertainty. The number of pseudo-experiments was chosen such that the relative uncertainty of the BH p-value amounts to less than 5%.

The results of the BH algorithm applied to the untagged invariant mass distributions for two different Z' signal injections are given in Figure 8.1. The figures show respectively the minimal signal cross-section required in order to be considered a potential discovery. In the upper panel, the background estimate is depicted in red, while the simulated data sample is shown in black. The latter includes the Z' signal samples, whose signal cross-section is listed in the figure. The sample with a mediator mass of 600 GeV requires a higher cross-section ($\sigma_{Signal} \approx 1.64$ pb) compared to the one with $m_{Z'} = 1$ TeV ($\sigma_{Signal} \approx 525.6$ fb) because the number of background events in the invariant mass distribution increases with decreasing m_{jj}. The lower part of the figure shows the significance distribution as defined in Equation 7.1 for each bin. It clearly illustrates the injected bump of the signal samples, which is not visible by eye in the upper distribution. The low and high edge of the most significant deviation window is displayed in blue.



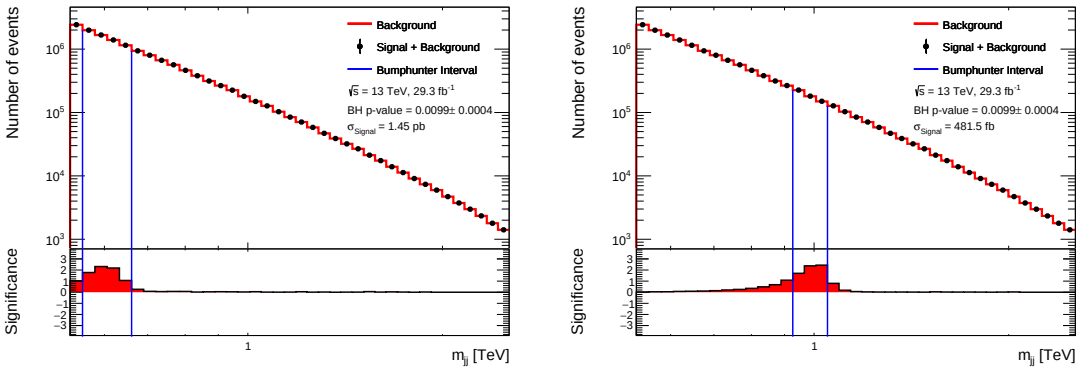
(a) $m_{Z'} = 600$ GeV and $g_q = 0.02$.

(b) $m_{Z'} = 1$ TeV and $g_q = 0.02$.

Figure 8.1: Bumphunter results for the untagged m_{jj} distribution for signal samples with $m_{Z'} = 600$ GeV and $m_{Z'} = 1$ TeV. The background (red) and simulated data (black) distribution as well as the significance distribution (lower part) can be seen. The BH p-values including their uncertainties and the required signal cross-sections to trigger BH are given as well.

The BH results after the quark-gluon tagging using the leading and subleading jet N_{90} and p_{T} variables of the two leading jets are shown in Figure 8.2. On the left, the Z' signal sample with a mass of 600 GeV was injected. A comparison of the required cross-sections for this signal sample before and after the tagging demonstrates that with tagging a reduction of σ_{Signal} of 12.1% can be achieved. Thus, the quark and gluon jet discrimination reduces the QCD background in such a way that the significance of the signal is enhanced. This shows that, although the tagging process reduces the signal efficiency (Figure 7.6a), the reduction of the QCD background prevails and the classification of signal and background events successfully improves the significance.

Figure 8.2b shows that the signal significance found by the BH algorithm for the $m_{Z'} = 1$ TeV signal sample is improved by 8.4% compared to the untagged scenario. Thus, the improvement is smaller compared to the lower mass signal sample, which might be induced by the fact that the NNcut of the MLP training output was chosen such that the significance gain shown in Figure 7.4 was maximised for the 600 GeV sample. However, since the quark jet fraction in dijet events is enhanced at larger invariant masses, it is in general more difficult to improve the resonance significance of a more massive Z' particle.



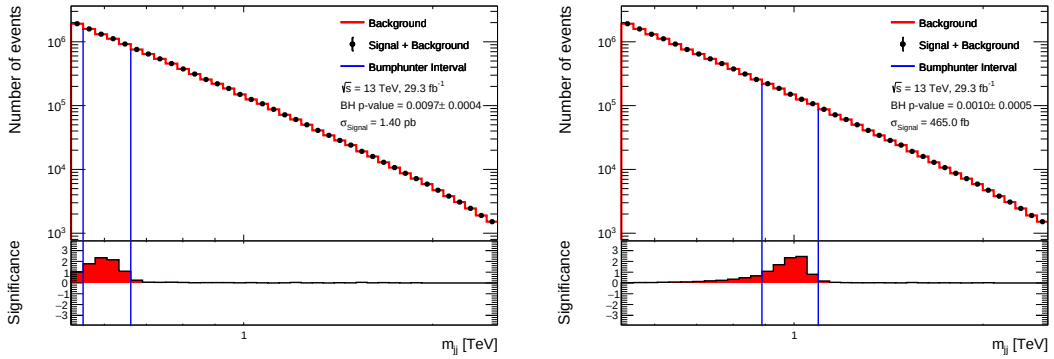
(a) $m_{Z'} = 600$ GeV and $g_q = 0.02$.

(b) $m_{Z'} = 1$ TeV and $g_q = 0.02$.

Figure 8.2: Bumphunter results for the tagged m_{jj} distribution using N_{90} and p_{T} in the quark-gluon tagging for signal samples with $m_{Z'} = 600$ GeV and $m_{Z'} = 1$ TeV. The background (red) and simulated data (black) distribution as well as the significance distribution (lower part) can be seen. The BH p-values including their uncertainties and the required signal cross-sections to trigger BH are given as well.

The statistical analysis results of BH algorithm evaluating the quark und gluon jet discrimination using in addition to N_{90} and p_{T} also the η variables of the two leading jets are given in Figure 8.3. The BH p-value assigns a statistical significance to the 600 GeV Z' sample with a cross-section of 1.39 pb, which corresponds to a significance enhancement of 15.1% compared to the untagged distribution. In relation to the quark-gluon tagging evaluation using four variables in the discrimination

process, the signal significance gain improves by additional 3% with respect to the untagged reference with the additional usage of the η variables. For the $m_{Z'} = 1$ TeV signal sample, the required cross-section can be reduced to 465 fb, while still triggering the BH algorithm. This is related to an increase in significance of the signal sample of 11.5% compared to the untagged case.



(a) $m_{Z'} = 600$ GeV and $g_q = 0.02$.

(b) $m_{Z'} = 1$ TeV and $g_q = 0.02$.

Figure 8.3: Bumphunter results for the tagged m_{jj} distribution using N_{90}, p_{T} and η in the quark-gluon tagging for signal samples with $m_{Z'} = 600$ GeV and $m_{Z'} = 1$ TeV. The background (red) and simulated data (black) distribution as well as the significance distribution (lower part) can be seen. The BH p-values including their uncertainties and the required signal cross-sections to trigger BH are given as well.

In summary, this statistical evaluation shows that quark-gluon tagging is successful in distinguishing quark induced dijet signals from QCD background events using a variety of discrimination variables. In general, a combination of these variables outmatches the independent separation power because the neural network is able to account for their correlations. The best found signal significance gain can be achieved by using the N_{90}, p_{T} and η variables of the two leading jets as discrimination variables, which exploit the differences in kinematics and multiplicities of quark and gluon jets. The significance gain of the two investigated signal samples induced by the quark and gluon jet discrimination using different discrimination variables is summarised in Table 8.1.

	600 GeV	**1 Tev**
N_{90}, p_{T}	$\approx 12.1\%$	$\approx 8.4\%$
N_{90}, p_{T}, η	$\approx 15.1\%$	$\approx 11.5\%$

Table 8.1: Significance gain of the respective signal samples using different discrimination variables for the quark-gluon tagging.

9 Conclusion

Searches for dijet resonances at the Large Hadron Collider provide the opportunity to study multiple models describing physics beyond the Standard Model. At the ATLAS experiment, bandwidth and storage capacities result in statistical limitations for the low-mass regime of the dijet invariant mass spectrum. The Trigger-Object Level Analysis approach circumvents these limitation by recording only physics objects that are reconstructed online within the ATLAS trigger system instead of using the event information from the total detector read-out. This reduces the size of an event such that event recordings at a high trigger rate are possible, while only a small fraction of the total bandwidth is used. However, down to the present day no significant excess has been found within the TLA regions.

In order to enhance the significance of a potential signal, this work engages in studies of quark and gluon jet discrimination using multivariate analysis methods based on machine learning techniques. The goal of the quark-gluon tagging is to reduce the gluon-dominated QCD background such that dijet data samples, which contain two quark jets in the final state, might show a significant local excess especially in the low-mass regime below 1 TeV.

In order to separate quark from gluon jets, features based upon the different colour factors of the initial partons are exploited. Since TLA does not have access to the full event information, this thesis investigates a possible quark-gluon separation based on trigger jet variables. This motivates the use of the N_{90} variable, which is the number of jet constituents whose energy adds up to 90% of the total jet energy, as well as the transverse momentum and the pseudorapidity of the two leading jets as discriminating variables. However, these variables show a significant mismodelling by comparing data and MC simulation, which may induce a potential bias in the training towards the chosen MC generator. Therefore, these variables are corrected by a reweighting procedure applied on the MC simulation on a event-by-event basis. The corresponding reweighting factors are determined via either a bin-by-bin reweighting or a polynomial fit reweighting, in which the order of the polynomial fit function is chosen on the basis of the χ^2 fit-value.

The choice of discriminating variables as well as the selection of the training model including its hyperparameters are crucial for a good model performance. In this thesis, it is shown that the multilayer perceptron is a good choice for a quark-gluon tagging model because it is able to take non-linear correlations between the discrimination variables into account. The highly flexible and complex network architecture of this model has a good performance but also tends to overtrain the data set. Therefore, the model is regularised using a Gaussian prior for the network weights. The optimal quark and gluon jet discrimination is accomplished by combining the N_{90},

p_{T} and η variables of the two leading jets. Additional energy layer information does not improve the network performance. The tagging process is statistically evaluated using the BUMPHUNTER algorithm, which determines the statistical significance of a Gaussian-like signal at 600 GeV and 1 TeV using a certain background estimate. This evaluation is based on a qualitative approach in which the MC simulation is used as the background estimate. For the former signal sample, the significance gain amounts to approximately 15.1%, while increase in significance for the latter sample adds up to 11.5%. Thus, the required cross-section for signals in the lower mass region is reduced to about 84.9%.

In conclusion, the result of this thesis shows that it is possible to enhance the significance of a dijet signal with two quark jets in the final state by rejecting the QCD background, which is dominated by dijet events in which at least one of the two leading jets is gluon-initiated. A qualitative analysis using a sliding-window fit as background estimate in combination with the quark-gluon tagging performed in this thesis might reveal features in the invariant mass spectrum of dijet events recorded by the Data Scouting stream in Run-2 that without the tagging are not identified as statistically significant.

A Lists

A.1 List of Figures

A.2 List of Tables

B Additional Plots and Tables

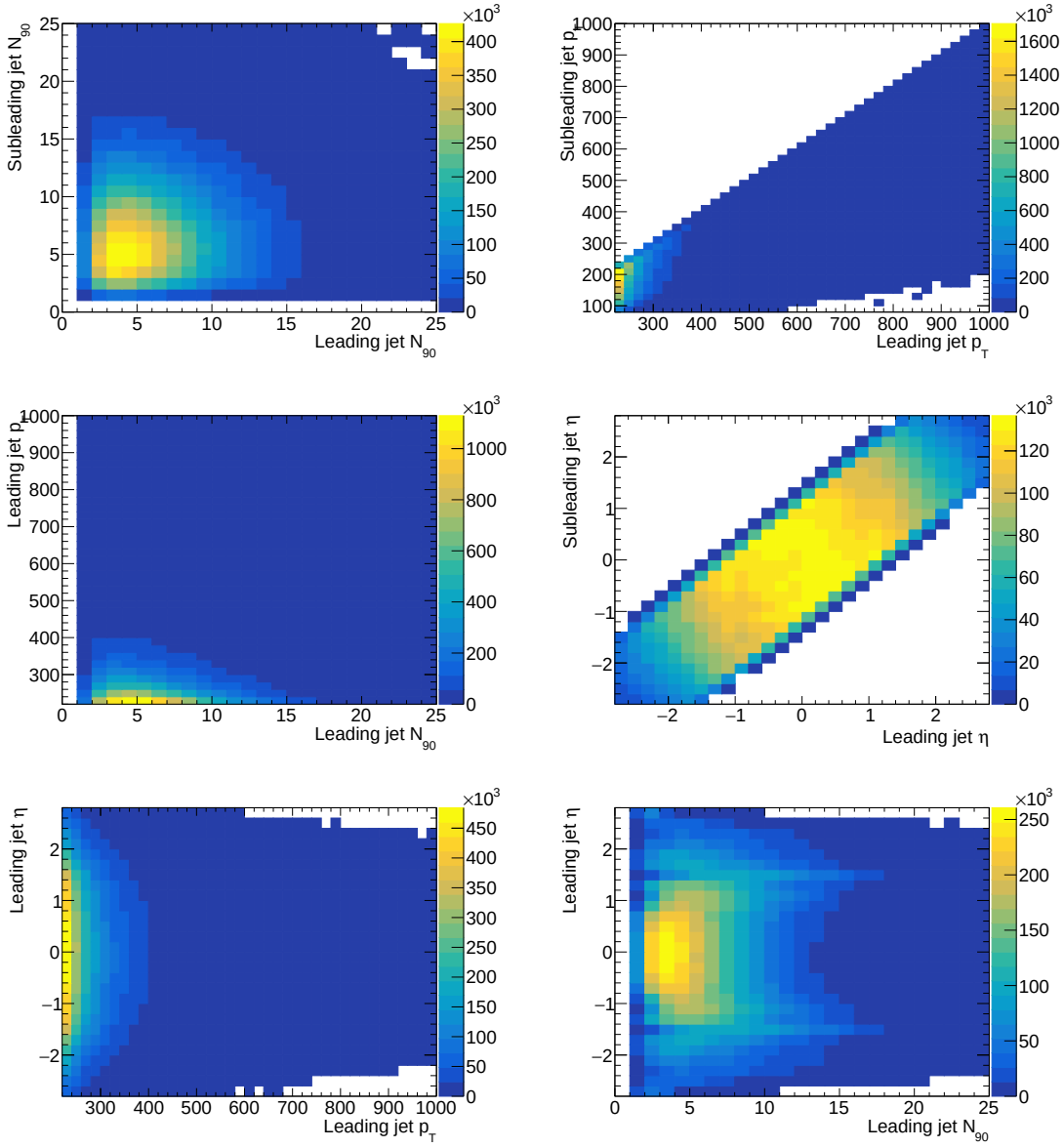
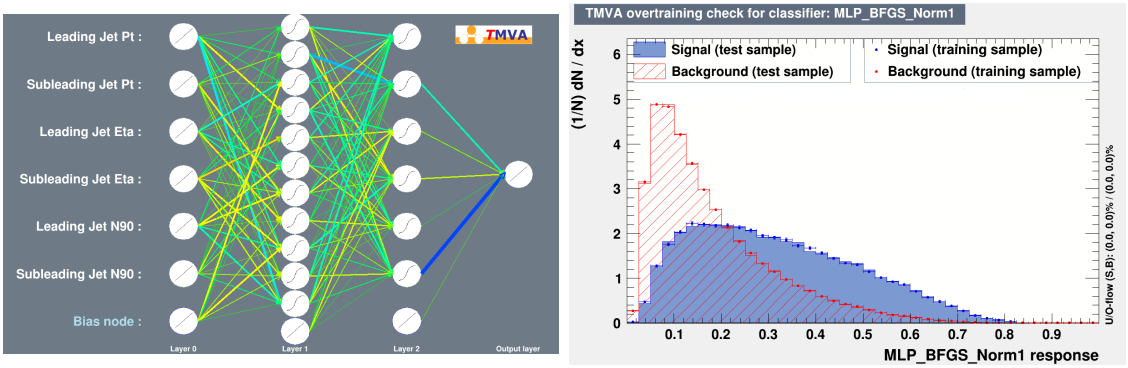


Figure B.1: Correlation plots of a selection of leading and subleading jet variables after the application of TLA cuts using MC simulations.



(a) Network architecture.

(b) Overtraining check.

Figure B.2: Control and performance plots of the MLP provided by the graphical user interface (GUI) of TMVA. (a) Three-layer network architecture with six input variables. (b) MLP output distributions for signal and background events from the training and test samples.

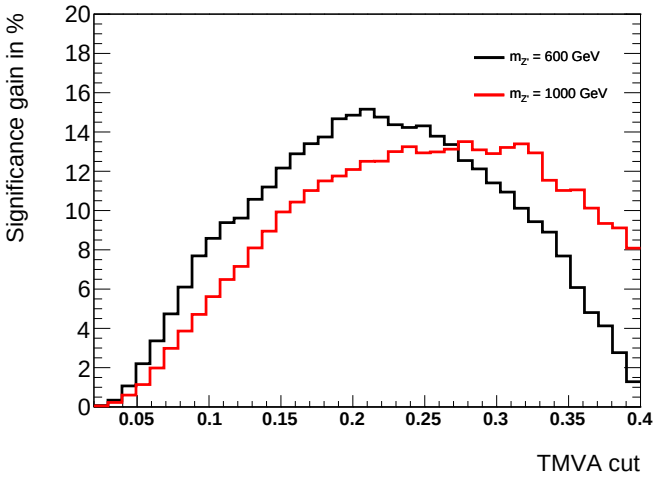
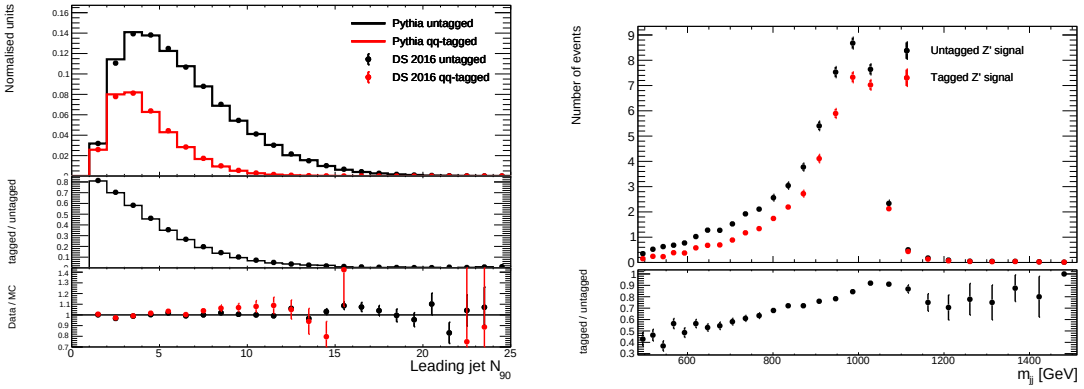


Figure B.3: Significance gain in % of the signal FWHM for two Z' signal samples with $m_{Z'} = 600$ GeV and $m_{Z'} = 1$ TeV over a range of NNCuts of the MLP network with N_{90}, p_{T} and η as discriminating variables.



(a) (b)

Figure B.4: (a) Comparison of leading jet N_{90} distribution of data (points) and MC simulation(histogram) before (black) and after (red) the quark-gluon-tagging using N_{90}, p_{T} and η as input variables. Upper part: normalised distributions. Middle Part: ratio of distributions before and after the tagging. Lower part: data over MC ratio before and after the tagging. (b) Invariant mass distribution of the Z' dijet signal with $m_{Z'} = 1$ TeV and $g_q = 0.02$ before and after tagging including the tagged-over-untagged ratio.

	Separation
Subleading jet N_{90}	0.09136
Leading jet N_{90}	0.07188
Subleading jet p_{T}	0.01927
Subleading jet ECal2	0.01324
Subleading jet HCal2	0.01147
Subleading jet HCal1	0.01034
Leading jet ECal1	0.00779
Leading jet ECal0	0.00763
Leading jet ECal2	0.00724
Subleading jet ECal1	0.00615
Subleading jet ECal0	0.00600
Subleading jet HCal3	0.00554
Leading jet HCal1	0.00531
Leading jet HCal2	0.00395
Subleading jet ECal3	0.00331
Leading jet p_{T}	0.00265
Leading jet HCal3	0.00134
Subleading jet η	0.00130
Leading jet ECal3	0.00099
Leading jet η	0.00078

Table B.1: Input variable separation. The top variable is ranked best.

C Bibliography

[1] Baron Kelvin. Baltimore lectures on molecular dynamics and the wave theory of light, 1907. URL https://archive.org/details/baltimorelecture00kelviala/page/n8/mode/2up.

[2] F. Zwicky. Republication of: The redshift of extragalactic nebulae. *Gen. Relativ. Gravit.*, 41(1):207–224, 2009. ISSN 00017701. doi: 10.1007/s10714-008-0707-4.

[3] Planck Collaboration. Planck 2015 results - XIII. Cosmological parameters. *Astron. Astrophys.*, 594:A13, oct 2016. ISSN 1432-0746. doi: 10.1051/0004-6361/201525830.

[4] Jan Conrad. Indirect Detection of WIMP Dark Matter: a compact review. Technical report, nov 2014.

[5] Teresa Marrodán Undagoitia and Ludwig Rauch. Dark matter direct-detection experiments. *J. Phys. G*, 43(1):013001, dec 2016. ISSN 0954-3899. doi: 10.1088/0954-3899/43/1/013001.

[6] Antonio Boveia and Caterina Doglioni. Dark Matter Searches at Colliders. *Annu. Rev. Nucl. Part. Sci.*, 68(1):429–459, oct 2018. ISSN 0163-8998. doi: 10.1146/annurev-nucl-101917-021008.

[7] ATLAS Collaboration. Search for new phenomena in dijet mass and angular distributions from pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Lett. B*, 754:302–322, mar 2016. ISSN 03702693. doi: 10.1016/j.physletb.2016.01.032.

[8] ATLAS Collaboration. Search for new phenomena in dijet events using 37 fb^{-1} of pp collision data collected at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev. D*, 96(5):052004, sep 2017. ISSN 24700029. doi: 10.1103/PhysRevD.96.052004.

[9] Philippe Gras, Stefan Höche, Deepak Kar, Andrew Larkoski, Leif Lönnblad, Simon Plätzer, Andrzej Siódmok, Peter Skands, Gregory Soyez, and Jesse Thaler. Systematics of quark/gluon tagging. *J. High Energy Phys.*, 2017(91), 2017. doi: 10.1007/JHEP07(2017)091.

[10] ATLAS Collaboration. Quark versus Gluon Jet Tagging Using Charged-Particle Constituent Multiplicity with the ATLAS Detector. Technical report, may 2017.

[11] Gregor Kasieczka, Nicholas Kiefer, Tilman Plehn, and Jennifer M. Thompson. Quark-Gluon Tagging: Machine Learning vs Detector. *SciPost Phys.*, 6(6): 069, jun 2019. doi: 10.21468/SciPostPhys.6.6.069.

[12] Mark Thomson. *Modern particle physics*. Cambridge University Press, 2013. ISBN 9781107034266.

[13] Antonio Pich. The Standard Model of Electroweak Interactions. Technical report, may 2007.

[14] William N. Cottingham and Derek A. Greenwood. *An introduction to the standard model of particle physics*. Cambridge University Press, 2. ed., 1. edition, 2007. ISBN 978-0-521-85249-4.

[15] Standard Model - Wikipedia, 2019. URL https://en.wikipedia.org/wiki/Standard_Model.

[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):1–29, jul 2012. doi: 10.1016/j.physletb.2012.08.020.

[17] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B*, 716(1):30–61, sep 2012. ISSN 03702693. doi: 10.1016/j.physletb.2012.08.021.

[18] Sheldon L. Glashow. Partial-symmetries of weak interactions. *Nucl. Phys.*, 22 (4):579–588, 1961. ISSN 00295582. doi: 10.1016/0029-5582(61)90469-2.

[19] A. Salam and J. C. Ward. Electromagnetic and weak interactions. *Phys. Lett.*, 13(2):168–171, nov 1964. ISSN 00319163. doi: 10.1016/0031-9163(64)90711-5.

[20] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19(21):1264–1266, 1967. ISSN 00319007. doi: 10.1103/PhysRevLett.19.1264.

[21] F. Zwicky. Die Rotverschiebung von extragalaktischen Nebeln. *Helv. Phys. Acta*, 6:110–127, 1933. ISSN 0018-0238. doi: 10.5169/seals-110267.

[22] Sinclair Smith. The Mass of the Virgo Cluster. *Astrophys. J.*, 83:23, jan 1936. ISSN 0004-637X. doi: 10.1086/143697.

[23] Steven Weinberg. The cosmological constant problem. *Rev. Mod. Phys.*, 61 (1):1–23, jan 1989. ISSN 00346861. doi: 10.1103/RevModPhys.61.1.

[24] Adam G. Riess, Alexei V. Filippenko, Peter Challis, Alejandro Clocchiatti, Alan Diercks, Peter M. Garnavich, Ron L. Gilliland, Craig J. Hogan, Saurabh Jha, Robert P. Kirshner, B. Leibundgut, M. M. Phillips, David Reiss, Brian P. Schmidt, Robert A. Schommer, R. Chris Smith, J. Spyromilio, Christopher Stubbs, Nicholas B. Suntzeff, and John Tonry. Observational Evidence from

Supernovae for an Accelerating Universe and a Cosmological Constant. *Astron. J.*, 116(3):1009–1038, sep 1998. ISSN 00046256. doi: 10.1086/300499.

[25] S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, J. C. Lee, N. J. Nunes, R. Pain, C. R. Pennypacker, R. Quimby, C. Lidman, R. S. Ellis, M. Irwin, R. G. McMahon, P. RuizLapuente, N. Walton, B. Schaefer, B. J. Boyle, A. V. Filippenko, T. Matheson, A. S. Fruchter, N. Panagia, H. J. M. Newberg, W. J. Couch, and The Supernova Cosmology Project. Measurements of Ω and Λ from 42 HighRedshift Supernovae. *Astrophys. J.*, 517(2):565–586, jun 1999. ISSN 0004-637X. doi: 10.1086/307221.

[26] Carlos S. Frenk and Simon D. M. White. Dark matter and cosmic structure. *Ann. Phys.*, 524:507–534, oct 2012. doi: 10.1002/andp.201200212.

[27] Mikhail Shaposhikov. Sterile neutrinos in cosmology and how to find them in the lab Related content. *J. Phys. Conf. Ser.*, 136(2):022045, nov 2008. doi: 10.1088/1742-6596/136/2/022045.

[28] John Ellis, J. S. Hagelin, D. V. Nanopoulos, K. Olive, and M. Srednicki. Supersymmetric relics from the big bang. *Nucl. Phys. B.*, 238(2):453–476, jun 1984. ISSN 05503213. doi: 10.1016/0550-3213(84)90461-9.

[29] John Preskill, Mark B. Wise, and Frank Wilczek. Cosmology of the invisible axion. *Phys. Lett. B*, 120(1-3):127–132, jan 1983. ISSN 03702693. doi: 10.1016/0370-2693(83)90637-8.

[30] Gary Steigman and Michael S. Turner. Cosmological constraints on the properties of weakly interacting massive particles. *Nucl. Phys. B.*, 253:375–386, 1985. ISSN 05503213. doi: 10.1016/0550-3213(85)90537-1.

[31] ATLAS Collaboration. Constraints on mediator-based dark matter and scalar dark energy models using $\sqrt{s} = 13$ TeV pp collision data collected by the ATLAS detector. *J. High Energy Phys.*, 2019(5):142, may 2019. ISSN 10298479. doi: 10.1007/JHEP05(2019)142.

[32] Bogdan A. Dobrescu and Felix Yu. Coupling-mass mapping of dijet peak searches. *Phys. Rev. D*, 88(3):035021, aug 2013. ISSN 15507998. doi: 10.1103/PhysRevD.88.035021.

[33] Christopher M. Bishop. *Pattern recognition and machine learning*. Springer, New York [u.a.], 8. (corr.) edition, 2009. ISBN 0-387-31073-8 and 978-0-387-31073-2.

[34] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss, M. Backes, T. Carli, O. Cohen, A. Christov, D. Dannheim, K. Danielowski, S. Henrot-Versille, M. Jachowski, K. Kraszewski, A. Krasznahorkay, M. Kruk,

Y. Mahalalel, R. Ospanov, X. Prudent, A. Robert, D. Schouten, F. Tegenfeldt, A. Voigt, K. Voss, M. Wolter, and A. Zemla. TMVA - Toolkit for Multivariate Data Analysis. *J. Phys. Conf. Ser.*, mar 2007.

[35] Jörg Stelzer, A. Höcker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Törne, and H. Voss. Data Analysis Techniques with TMVA. Technical report, 13th International Workshop on Advanced Computing and Analysis Techniques in Physics Research, Jaipur, India, feb 2010.

[36] Xin Wang, Zhiqiang Hou, Wangsheng Yu, Yang Xue, Zefenfen Jin, and Bo Dai. Robust visual tracking via multiscale deep sparse networks. *Opt. Eng.*, 56(4): 043107, apr 2017. ISSN 0091-3286. doi: 10.1117/1.oe.56.4.043107.

[37] A. Emin Orhan and Xaq Pitkow. Skip Connections Eliminate Singularities. International Conference on Learning Representations, ICLR, jan 2017.

[38] Rene Brun and Fons Rademakers. ROOT - An object oriented data analysis framework. *Nucl. Instrum. Methods Phys. Res. A*, 389(1-2):81–86, apr 1997. ISSN 01689002. doi: 10.1016/S0168-9002(97)00048-X.

[39] Jean-Charles Walter and Gerard Barkema. An introduction to Monte Carlo methods. *Physica A*, 418:78–87, apr 2014. doi: 10.1016/j.physa.2014.06.014.

[40] Colin Reeves. Genetic Algorithms. In Fred Glover and Gary A. Kochenberger, editors, *Handbook of Metaheuristics*, chapter 3, pages 55–82. Springer US, Boston, MA, 2003. ISBN 978-0-306-48056-0. doi: 10.1007/0-306-48056-5_3.

[41] Darrall Henderson, Sheldon H. Jacobson, and Alan W. Johnson. The Theory and Practice of Simulated Annealing. In Fred Glover and Gary A. Kochenberger, editors, *Handbook of Metaheuristics*, chapter 10, pages 287–319. Springer US, Boston, MA, apr 2006. ISBN 978-0-306-48056-0. doi: 10.1007/0-306-48056-5_10.

[42] Stephen M. Stigler. Gauss and the Invention of Least Squares. *Ann. Stat.*, 9 (3):465–474, 1981. ISSN 0090-5364. doi: 10.1214/AOS/1176345451.

[43] R. A. Fisher. The use of multiple measurements in taxonomic problems. *Ann. Eugen.*, 7(2):179–188, sep 1936. ISSN 20501420. doi: 10.1111/j.1469-1809.1936.tb02137.x.

[44] Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In Paul Vitanyi, editor, *Comput. Learn. Theory*, volume 904, pages 23–37, Berlin, Heidelberg, 1995. Springer Verlag. ISBN 978-3-540-49195-8. doi: 10.1007/3-540-59119-2_166.

[45] Leo Breiman. Bagging predictors. *Mach. Learn.*, 24(2):123–140, 1996. ISSN 08856125. doi: 10.1007/bf00058655.

[46] Ajith Abraham. Artificial Neural Networks. In *Handbook of Measuring System Design*, chapter 129. John Wiley & Sons, jul 2005. doi: 10.1002/0471497398.mm421.

[47] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986. ISSN 00280836. doi: 10.1038/323533a0.

[48] C. G. Broyden. The Convergence of a Class of Double-rank Minimization Algorithms. *IMA J. Appl. Math.*, 6(1):76–90, mar 1970. ISSN 0272-4960. doi: 10.1093/imamat/6.1.76.

[49] R. Fletcher. A new approach to variable metric algorithms. *Comput. J.*, 13 (3):317–322, jan 1970. ISSN 0010-4620. doi: 10.1093/comjnl/13.3.317.

[50] Donald Goldfarb. A family of variable-metric methods derived by variational means. *Math. Comput.*, 24(109):23–23, jan 1970. ISSN 0025-5718. doi: 10.1090/s0025-5718-1970-0258249-6.

[51] D. F. Shanno. Conditioning of quasi-Newton methods for function minimization. *Math. Comput.*, 24(111):656, sep 1970. ISSN 0025-5718. doi: 10.1090/s0025-5718-1970-0274029-x.

[52] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:8003, 2008. doi: 10.1088/1748-0221/3/08/S08003.

[53] Lyndon Evans and Philip Bryant. LHC Machine. *J. Instrum.*, 3(08):S08001, aug 2008. doi: 10.1088/1748-0221/3/08/s08001.

[54] ATLAS Collaboration. Constituent-level pile-up mitigation techniques in ATLAS. Technical report, aug 2017.

[55] ATLAS Collaboration. Performance of pileup mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector. *Eur. Phys. J. C*, 76(581), dec 2016. ISSN 1434-6052. doi: 10.1140/epjc/s10052-016-4395-z.

[56] ATLAS Collaboration. ATLAS data quality operations and performance for 2015-2018 data-taking. *JINST*, 15(04):P04003, apr 2020. doi: 10.1088/1748-0221/15/04/P04003.

[57] ATLAS Collaboration. Performance of electron and photon triggers in ATLAS during LHC Run 2. *Eur. Phys. J. C*, 80:47, feb 2020. doi: 10.1140/epjc/s10052-019-7500-2.

[58] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13(16):508–509, oct 1964. ISSN 00319007. doi: 10.1103/PhysRevLett.13.508.

[59] ATLAS Collaboration. ATLAS liquid-argon calorimeter : Technical design report. Technical report, 1996.

[60] Stefanie Morgenstern. LAr Calorimeter Performance in LHC Run-2. Technical report, jun 2018.

[61] ATLAS Collaboration. ATLAS tile calorimeter: Technical design report. Technical report, 1996.

[62] Pawel Klimek. Calibration and performance of the ATLAS Tile Calorimeter during the LHC Run 2 Pawel Klimek on behalf of the ATLAS Collaboration. *J. Phys. Conf. Ser.*, 1162:012003, jan 2019. doi: 10.1088/1742-6596/1162/1/012003.

[63] ATLAS Collaboration. Performance of the ATLAS Trigger System in 2015. *Eur. Phys. J. C*, 77:317, may 2017. doi: 10.1140/epjc/s10052-017-4852-3.

[64] ATLAS Collaboration. The ATLAS Simulation Infrastructure. *Eur. Phys. J. C*, 70:823–874, 2010. doi: 10.1140/epjc/s10052-010-1429-9.

[65] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. A brief introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178(11):852–867, jun 2008. ISSN 00104655. doi: 10.1016/j.cpc.2008.01.036.

[66] ATLAS Collaboration. ATLAS Pythia 8 tunes to 7 TeV datas. Technical report, nov 2014.

[67] Stefano Carrazza, Stefano Forte, and Juan Rojo. Parton Distributions and Event Generators. Technical report, nov 2013.

[68] GEANT4 Collaboration. GEANT4 - A simulation toolkit. *Nucl. Instrum. Methods Phys. Res. A*, 506(3):250–303, 2003. ISSN 01689002. doi: 10.1016/S0168-9002(03)01368-8.

[69] Takashi Yamanaka. The ATLAS calorimeter simulation FastCaloSim . *J. Phys. Conf. Ser.*, 331(3):032053, 2011. doi: 10.1088/1742-6596/331/3/032053.

[70] Steven Schramm. ATLAS Jet Reconstruction, Calibration, and Tagging of Lorentzboosted Objects. *EPJ Web Conf.*, 182:02113, 2018. ISSN 2100-014X. doi: 10.1051/EPJCONF/201818202113.

[71] ATLAS Collaboration. Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1. *Eur. Phys. J. C*, 77(490), 2017. doi: 10.1140/epjc/s10052-017-5004-5.

[72] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-kt jet clustering algorithm. *J. High Energy Phys.*, 2008(04):063–063, apr 2008. doi: 10.1088/1126-6708/2008/04/063.

[73] 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(7), oct 2017. ISSN 24700029. doi: 10.1103/PhysRevD.96.072002.

[74] Matteo Cacciari and Gavin P. Salam. Pileup subtraction using jet areas. *Phys. Lett. B*, 659(1-2):119–126, jul 2007. doi: 10.1016/j.physletb.2007.09.077.

[75] ATLAS Collaboration. Determination of jet calibration and energy resolution in proton-proton collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector. Technical report, oct 2019.

[76] ATLAS Collaboration. Search for New Physics in the Dijet Mass Distribution using 1 fb^{-1} of pp Collision Data at $\sqrt{s}$ = 7 TeV collected by the ATLAS Detector. *Phys. Lett. B*, 708(1-2):37–54, aug 2011. doi: 10.1016/j.physletb.2012.01.035.

[77] CMS Collaboration. Search for Resonances in the Dijet Mass Spectrum from 7 TeV pp Collisions at CMS. *Phys. Lett. B*, 704(3):123–142, jul 2011. doi: 10.1016/j.physletb.2011.09.015.

[78] ATLAS Collaboration. Search for Low-Mass Dijet Resonances Using Trigger-Level Jets with the ATLAS Detector in pp Collisions at $\sqrt{s}$ = 13 TeV. *Phys. Rev. Lett.*, 121(8):081801, aug 2018. ISSN 10797114. doi: 10.1103/PhysRevLett.121.081801.

[79] ATLAS Collaboration. Trigger Operation Public Results. URL https://twiki.cern.ch/twiki/bin/view/AtlasPublic/TriggerOperationPublicResults.

[80] Patrick T. Komiske, Eric M. Metodiev, and Jesse Thaler. An operational definition of quark and gluon jets. *J. High Energy Phys.*, 2018(59), 2018. doi: 10.1007/JHEP11(2018)059.

[81] Jason Gallicchio and Matthew D. Schwartz. Quark and Gluon Tagging at the LHC. *Phys. Rev. Lett.*, 107(17), 2011. doi: 10.1103/PhysRevLett.107.172001.

[82] H. P. Nilles and K. H. Streng. Quark-gluon separation in three-jet events. *Phys. Rev. D*, 23(9):1944–1950, may 1981. ISSN 05562821. doi: 10.1103/PhysRevD.23.1944.

[83] Lorella M. Jones. Tests for determining the parton ancestor of a hadron jet. *Phys. Rev. D*, 39(9):2550–2560, may 1989. ISSN 05562821. doi: 10.1103/PhysRevD.39.2550.

[84] Z. Fodor. How to see the differences between quark and gluon jets. *Phys. Rev. D*, 41(5):1726–1730, mar 1990. ISSN 05562821. doi: 10.1103/PhysRevD.41.1726.

[85] Lorella Jones. Towards a systematic jet classification. *Phys. Rev. D*, 42(3): 811–814, aug 1990. ISSN 05562821. doi: 10.1103/PhysRevD.42.811.

[86] Leif Lönnblad, Carsten Peterson, and Thorsteinn Rögnvaldsson. Using neural networks to identify jets. *Nucl. Phys. B.*, 349(3):675–702, feb 1991. ISSN 05503213. doi: 10.1016/0550-3213(91)90392-B.

[87] Jon Pumplin. How to tell quark jets from gluon jets. *Phys. Rev. D*, 44(7): 2025–2032, oct 1991. ISSN 05562821. doi: 10.1103/PhysRevD.44.2025.

[88] Andrew J. Larkoski, Jesse Thaler, and Wouter J. Waalewijn. Gaining (mutual) information about quark/gluon discrimination. *J. High Energy Phys.*, 2014 (11):1–56, nov 2014. ISSN 10298479. doi: 10.1007/JHEP11(2014)129.

[89] CMS Collaboration. Performance of quark/gluon discrimination using pp collision data at $\sqrt{s} = 8$ TeV. Technical report, 2013. URL https://cds.cern.ch/record/1599732.

[90] ATLAS Collaboration. Light-quark and gluon jet discrimination in *pp* collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Eur. Phys. J. C*, 74(8):3023, 2014. doi: 10.1140/epjc/s10052-014-3023-z.

[91] CMS Collaboration. Search for the standard model Higgs boson produced through vector boson fusion and decaying to $b\bar{b}$. *Phys. Rev. D*, 92(3):032008, aug 2015. ISSN 15502368. doi: 10.1103/PhysRevD.92.032008.

[92] CMS Collaboration. Measurement of electroweak production of two jets in association with a Z boson in proton-proton collisions at $\sqrt{s} = 8$ TeV. *Eur. Phys. J. C*, 75(2), feb 2015. doi: 10.1140/epjc/s10052-014-3232-5.

[93] ATLAS Collaboration. Measurement of the charged-particle multiplicity inside jets from $\sqrt{s} = 8$ TeV *pp* collisions with the ATLAS detector. *Eur. Phys. J. C*, 76(6):1–23, 2016. doi: 10.1140/epjc/s10052-016-4126-5.

[94] DELPHI Collaboration. Energy dependence of the differences between the quark and gluon jet fragmentation. *Zeitschrift für Phys. C - Part. Fields*, 70: 179–195, 1996. ISSN 01709739. doi: 10.1007/s002880050095.

[95] DELPHI Collaboration. The Scale Dependence of the Hadron Multiplicity in Quark and Gluon Jets and a Precise Determination of C_A/C_F. *Phys. Lett. B*, 449:383–400, 1999. doi: 10.1016/S0370-2693(99)00112-4.

[96] OPAL Collaboration. Experimental properties of gluon and quark jets from a point source. *Eur. Phys. J. C*, 11:217–238, 1999. doi: 10.1007/s100520050628.

[97] Kevin Black, Jason Gallicchio, John Huth, Michael Kagan, Matthew D. Schwartz, and Brock Tweedie. Multivariate discrimination and the Higgs + W/Z search. *J. High Energy Phys.*, 2011(4), apr 2011. doi: 10.1007/JHEP04(2011)069.

[98] Carola F. Berger, Tibor Kúcs, and George Sterman. Event shapeenergy flow correlations. *Phys. Rev. D*, 68(1):014012, jul 2003. ISSN 15502368. doi: 10.1103/PhysRevD.68.014012.

[99] CLEO Collaboration. Comparison of Particle Production in Quark and Gluon Fragmentation at $\sqrt{s} \sim 10$ GeV. *Phys. Rev. D*, 76:012005, 2007. doi: 10.1103/PhysRevD.76.012005.

[100] Peter Skands. Introduction to QCD. pages 341–420. World Scientific Pub Co Pte Lt, 2013. doi: 10.1142/9789814525220_0008.

[101] Andrew J. Larkoski, Gavin P. Salam, and Jesse Thaler. Energy correlation functions for jet substructure. *J. High Energy Phys.*, 2013(6):1–43, jun 2013. ISSN 11266708. doi: 10.1007/JHEP06(2013)108.

[102] Glen Cowan. *Statistical Data Analysis [with applications from particle physics]*. Clarendon Press, Oxford, 1998. ISBN 0-19-850156-0, 0-19-850155-2, 978-0-19-850156-5, 978-0-19-850155-8.

[103] Georgios Choudalakis. On hypothesis testing, trials factor, hypertests and the BumpHunter. Technical report, jan 2011.

[104] G. Choudalakis and D. Casadei. Plotting the differences between data and expectation. *Eur. Phys. J. Plus*, 127(2):1–11, feb 2012. ISSN 2190-5444. doi: 10.1140/EPJP/I2012-12025-Y.

Erklärung:

Ich versichere, dass ich diese Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, den 04.05.2020 .....