ETP-Bachelor-2017-23

Fake rate-based multi-jet QCD background estimation technique for a heavy gauge-boson search at the CMS experiment

Bachelor Thesis

Kevin Ziehl

Department of Physics
Institut für Experimentelle Teilchenphysik (ETP)

Reviewer:	Prof. Dr. Ulrich Husemann
Second reviewer:	Dr. Matthias Schröder

Karlsruhe, 19th of September 2017

Accepted by the first referee of the bachelor thesis.

Karlsruhe, 19.09.2017

. . . .

(**Prof. Dr. Ulrich Husemann**)

I declare that I have developed and written the enclosed thesis completely by myself, and have not used sources or means without declaration in the text.

Karlsruhe, 19.09.2017

. . . .

(**Kevin Ziehl**)

Contents

1 Introduction

The standard model of particle physics is a coherent theory that describes our current knowledge of three of the four fundamental interactions in physics. Nonetheless, several phenomena, for example the nature of dark matter, cannot be explained in the frame of the standard model, leading to the exploration of new theoretical ground. Many theories attempt to merge all different types of matter into one unified structure. The search for a final theory which could explain every possible natural phenomenon was triggered long before one of the prospective final elements of the standard model was found in the year 2012. In this very year, the discovery of the Higgs boson, whose existence is a consequence of the Higgs mechanism, was firstly announced at the LHC by the ATLAS [1] and CMS [2] collaborations with a mass of about $125\,\mathrm{GeV/c^2}$.

But with the discovery of the last missing part of the standard model, the search for the existence of other heavy resonances has not stopped. In fact, many beyond the standard model of particle physics theories predict the existence of massive bosonic resonances [3, 4, 5]. This resonances can be represented by Z' gauge bosons [6, 7], which is object of many investigations.

An extensive effort has been performed by the ATLAS and CMS collaborations to look for massive resonances decaying to top quark-antiquark pairs [8, 9]. These investigations curtailed the possible production cross sections. Other models with a heavy gluon [10], a composite Higgs boson [11], or extra spatial dimensions [12] impose an additional sector which is represented by fermionic vector-like quarks. Finding these predicted particles is a step further towards a Grand Unified Theory (GUT) explaining the behavior of all known particles and forces. This thesis focuses on the kinematic sector, in which the Z' boson decays into a vector-like quark and a top quark. The decay products of the top quark and the heavy T' quark hadronize into a set of particles which will be further reconstructed with appropriate methods. The investigated decay channel is fully hadronic, and the final state consists of at least three jets.

Considering the current exclusion limits, the masses of the hypothesized particles have to be in the TeV range, thus about five times heavier than the top quark, the heaviest particle known in the standard model of particle physics. Owing to the enourmous mass of the Z' boson, the top quark likely receives large transverse momentum, depending on the mass of T' quark. Hence, the decay products of the T' and Z' appear as merged streams, so-called jets, which can be combined as one large jet with a larger radius compared to the jet of a bottom-quark. Knowing precisely which particles are created in each event is a step towards the investigation of such heavy bosonic resonances.

Not only jets originating from the interesting events are detected but also jets which have a different origin. Due to the high mass of the Z′ boson, the production rate at a collider, depending on the model, is small. Therefore, knowing which other processes can mimic the desired decay channel is crucial. These background processes often have a much higher contribution to the signal region than the signal itself. Understanding and knowing the contribution of possible background processes is an important part in every physics analysis. This thesis is devoted to the estimation of the QCD multijet background (often referred to as just QCD background), which is the dominant background in the investigated decay channel, where the Z′ boson decays into a top quark and a vector-like T′ quark. The estimate uses the data-driven top mistag method, which is validated using simulated events. Data-driven methods are preferred to MC simulations, to avoid dependencies on simulated events.

The thesis is structured as follows: Chapter 2 includes an overview of the standard model of particle physics, along with theoretical aspects of the Z′ boson and the vector-like T′ quark. In order to produce such heavy particles, one has to use the current most powerful particle accelerator, the Large Hadron Collider (LHC) at CERN. Chapter 3 is devoted to the experimental setup, giving a brief introduction to the LHC and the CMS experiment. In chapter 4, terms and techniques necessary to understand the following work will be introduced, and the characteristics of the signal and background processes will be discussed. The last chapter 5 describes the development of the technique combined with the experimental results of the analysis. The thesis is concluded by a summary in chapter 6.

2 Theoretical background

In this chapter, a brief introduction to the standard model of particle physics will be given. The first part depicts a rough overview of the current knowlegde of the standard model of particle physics (SM). The second part treats the problems which occur in the SM. These problems give rise to the need for more profound theories, and one of them is briefly discussed. Unless otherwise stated, the information in this chapter is taken from [13].

2.1 The standard model of particle physics

The SM is a consistent, renormalizable Quantum Field Theory and the current most fundamental working theory in the realm of particle physics in our understanding. It describes the interactions of particles on a microscopic scale with the use of force-carrying particles. The interactions that are succesfully included in the SM are the electromagnetic-, strong-, and the weak interactions. The strength of the theory relies on the fact that all phenomena observed so far in the quantum world can be described employing one of these forces. The only force which cannot currently be included in the framework of the SM is the gravitational force. Due to its relative weakness in comparison with the other forces, it can be often neglected when considering particle interactions. Since all fundamental processes are formulated as a field theory, the dynamic evolution can be described by a Lagrangian density.

The well-known part of the matter content in the universe consists of leptons and quarks which interact with each other via bosons. Leptons and quarks are spin-1/2 particles which implies that they are fermions. All twelve fundamental fermions (leptons and quarks) carry the quantum number weak isospin which enables them to interact via the weak interaction. Except the electrically neutral neutrinos, all fermions carry charge and therefore participate in the electromagnetic force. Only quarks have an additional quantum number called colour charge which is why they also participate in the strong interaction. Leptons and quarks can be arranged into generations, reflecting their physical interactions. Each of the leptons and quarks can be assigned a corresponding anti-particle with reversed sign of electric and color charge.

The above particles and those which are composed of them, interact with each other via the aforementioned forces, which are mediated by particles called gauge-bosons. The name implies that these particles carry integer spin. The mediators of the electromagnectic force are photons. The photons couple to electrically charged particles and mediate attractive or repulsive forces, depending on the electric charges of the interacting particles.

The underlying field theory is called Quantum Electrodynamics (QED). Mathematically speaking, QED is an abelian gauge theory with the symmetry group U(1).

Unlike the massless photons of the electromagnetic field, the mediators of the weak interaction, the W and Z bosons, possess significant masses, due to the Higgs mechanism. Another special property is the fact that it is the only interaction that can change the flavour of quarks and leptons. Both field theories are unified in the electroweak interaction. The underlying gauge symmetry is the $\mathrm{U(1)_Y}$ of weak hypercharge and the $\mathrm{SU(2)_L}$ of the weak interaction.

Quantum Chromodynamics is the gauge theory of strong interactions, which describes the interplay between gluons and quarks. It is mediated by gluons of zero mass. It is formulated as a non-abelian gauge theory with symmetry group SU(3).

The final part which completed the SM, namely the discovery of the Higgs boson, was announced in 2012 at the Large Hadron Collider. The Higgs mechanism, namely the spontaneous breaking of the electroweak symmetry, gives mass to massive elementary particles. The Higgs boson is the only fundamental scalar ($\mathrm{S} = 0$) particle known so far. Figure 2.1 shows a complete sketch of all fundamental particle currently included in the SM.

2.2 Physics beyond the standard model

Despite the remarkable verification of the SM, several phenomena have been observed that cannot be explained in the frame of the SM. One unexplained phenomenon is the existence of non-luminous matter, called dark matter, which was first hinted at in the 1930s. Evidence of this mysterious kind of matter can be inferred from the velocity distribution of stars in the galactic plane. This, as well as the unknown nature of dark energy, which was determined by observations of remote supernovae, demand an extension of the SM [15]. According to the ΛCDM-model, which has a similar position in cosmology as the SM in particle physics, about 85% of all matter in the universe is made up of cold dark matter (CDM). To explain the observations using CDM, the expected masses of this matter must lie in GeV-TeV range. Such particle masses arise naturally in extensions to the SM, such as supersymmetry (SUSY) [16].

Another highly-discussed approach for a speculative extension of the SM, besides SUSY, are composite Higgs models (CHM)[17]. This thesis uses in particular the minimal CHM as a reference point [11]. Starting point of this theory is the assumption that the Higgs boson is a composite state favored by some new strong interaction. Similar to resonances in common particle physics, heavy resonances would arise from the composite Higgs as an excitation of the field. The topic of this thesis is the search for heavy spin-1 resonances. From theoretical considerations follows that these $\mathrm{Z'}$ particles can decay into a top quark and a massive vector-like top quark partner $\mathrm{T'}$ ($\mathrm{Z'} \to \mathrm{tT'}$). This work focuses on the kinematic range where the decay to $\mathrm{T'}$ $\mathrm{T'}$ is energetically prohibited. Despite the lack of experimental evidence for vector-like quarks, they are predicted by several new physics theories including the composite Higgs models. They emerge as excited resonances of composed states of particles. They are called vector-like because their current is composed of left- and right-handed charged currents, leading to a vector current. This is opposite to the current of the electroweak interaction with its V-A structure. The left- and right-handed chirality transformation of vector-like fermions are the same under the gauge groups $\mathrm{SU(3)_C} \bigotimes \mathrm{SU(2)_L} \bigotimes \mathrm{U(1)_Y}$ of the SM. Their mass is generated by a direct mass term (known as Dirac mass term) and their interactions between SM quarks are still via Yukawa couplings [18]. This leads to an adaptation of the couplings between the quarks and Z-, W-, and H-boson [19]. Usually they decay under mixing with third-generation SM quarks, thus the decay mode $\mathrm{Z'} \to \mathrm{tT'}$ gets predestined.

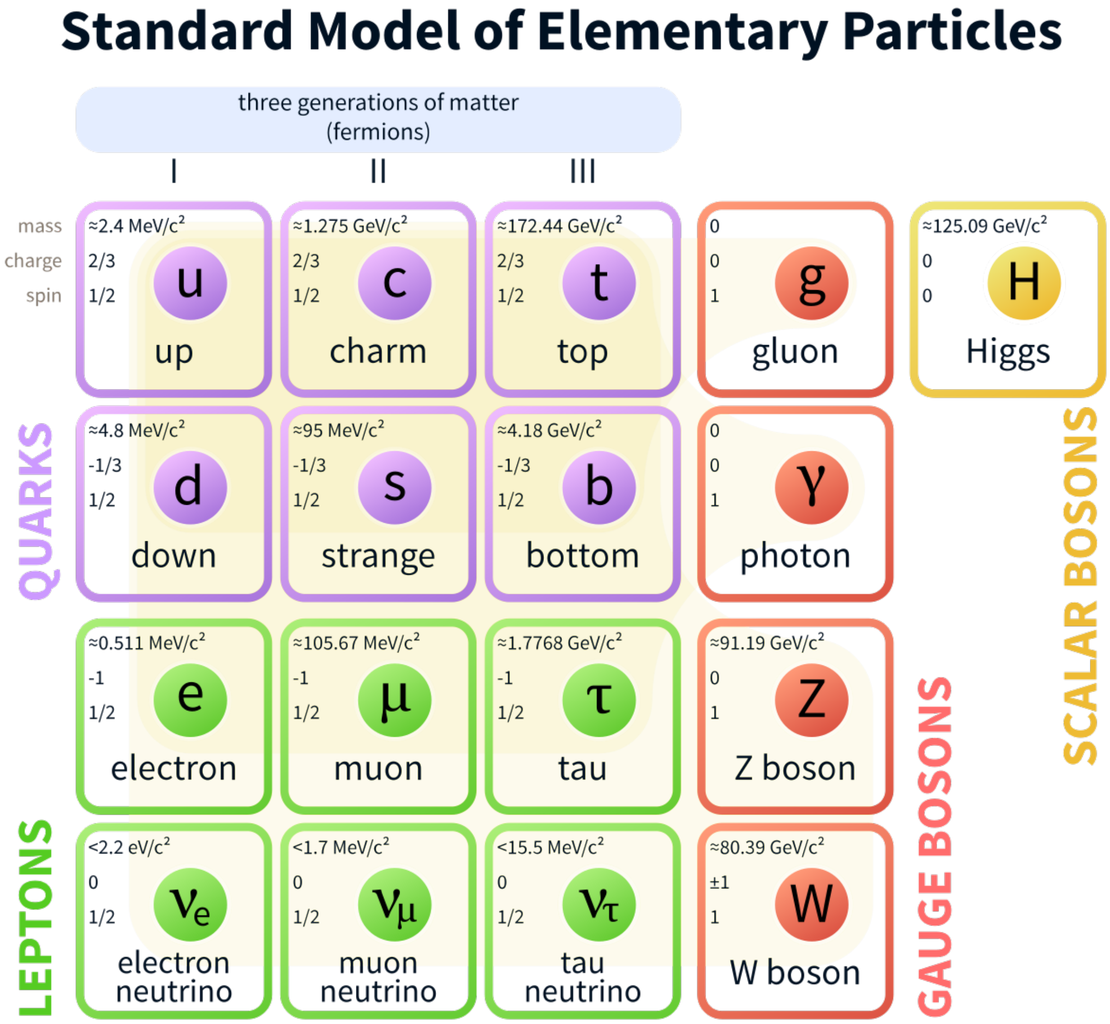

Figure 2.1: **The known fundamental particles along with their mass, charge, and spin.** The first three columns arrange the particles (leptons and quarks) in generations. The fourth column right lists all known mediators and the fifth the Higgs boson. For each particle, the mass, charge, and the spin are also shown. Taken from [14].

3 The CMS experiment

The following sections cover a brief description of the experimental setup. In the first part of this chapter the Large Hadron Collider is outlined, followed by a rough overview of one of the general-purpose experiments, the Compact Muon Solenoid. If not otherwise stated, the information is taken from [20] and [21].

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a two-ring hadron collider at the European Organization for Nuclear Research (CERN) located near Geneva in Switzerland. The collider is located in a tunnel 26.7 km in circumference. It contains two separated counter-rotating beams in which protons or lead ions can be collided. The design of the LHC allows accelerating protons up to an energy of 7 TeV per beam which leads to a center-of-mass energy of 14 TeV. Before the proton beam can be injected into the LHC, it must be led through a sequence of linear and circular accelerators in order to reach the needed energy. The entire accelerator chain with the four major experiments is outlined in figure 3.1.

In each of the four crossing points of the beam line, experiments are installed. These four experiments are ALICE[22], ATLAS[23], CMS[21], and LHCb[24]. ALICE is specialized for heavy-ion collisions and attempts to detect the quark-gluon plasma. ATLAS and CMS are general-purpose high-luminosity experiments which examine a wide spectrum of physics questions. Both detectors are aiming for the same scientific goals from the discovery and investigation of the Higgs boson, the search for new heavy particles, up to extra dimensions. Due to their different technical setup, they are ideal for mutual verification. The LHCb experiment is focused on studying CP violation in b-systems in order to study the differences between matter and antimatter and rare particle decays.

An important quantity in particle physics is the number of events per time interval generated in the LHC collisions:

$$\dot{N}_{\text{event}} = \mathcal{L}\sigma_{\text{event}}$$

where $\mathcal{L}$ is the instantaneous luminosity and σ_{event} is the event cross section of the studied physical process. While the cross section depends on the underlying physical process, the luminosity only depends on the beam parameters. Thus, one key factor to study rare processes is to increase the luminosity. The current instantaneous luminosity provided by the LHC is well above $\mathcal{L} = 10^{34}\,\text{cm}^2\text{s}^{-1}$.

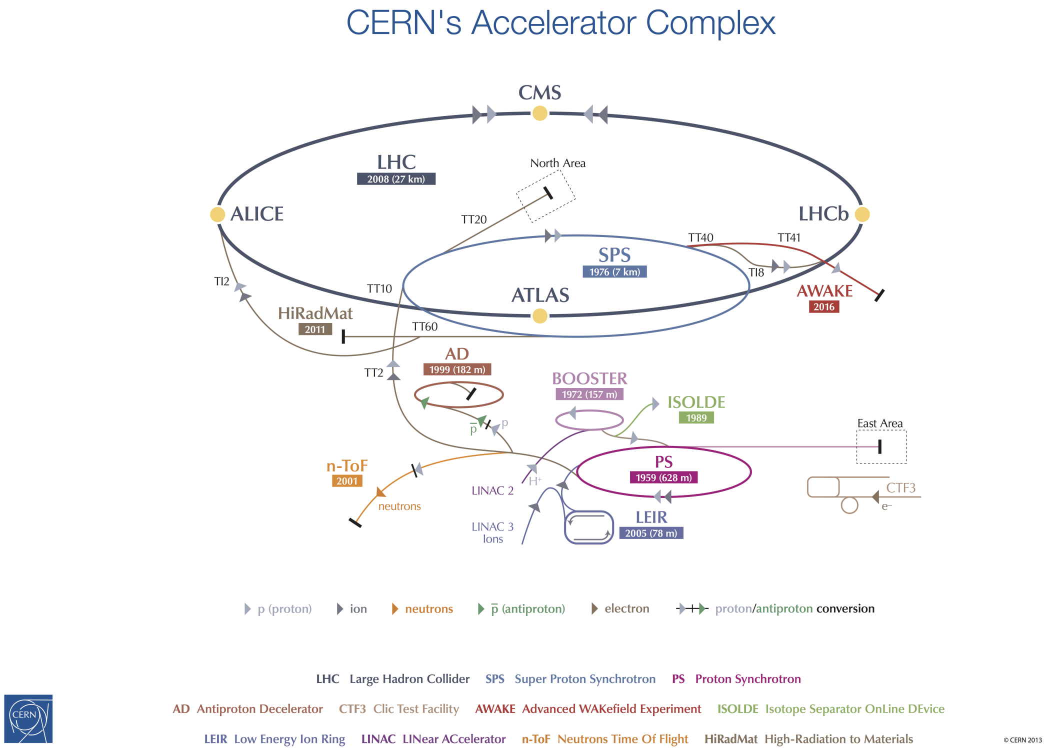

Figure 3.1: **Sketch of the accelerator complex at CERN.** Before the beam gets injected into the LHC, the beam originates in LINAC 2 and is led through the Proton Synchrotron Booster (PBS), the Proton Synchrotron (PS), and the Super Proton Synchrotron (SPS). The four major experiments CMS, ATLAS, LHCb, and ALICE are situated in the intersection points. Taken from [25].

3.2 The CMS detector

The Compact Muon Solenoid (CMS) detector is one of the two multi-purpose machines at the Large Hadron Collider (LHC). The main goal of the CMS experiment is to study the nature of electroweak symmetry breaking which incorporates the Higgs mechanism. In addition, an exhaustive search for physics beyond the standard model at energy scales beyond 1 TeV is performed, which is partly elaborated in this work.

The detector has a diameter of 14.6 m, a length of 21.6 m and a weight of nearly 14 000 t. The coordinate system used by CMS has its origin in the intersection point. The z-axis coincides with the counter-clockwise rotating beam direction, whereas the x-axis points radially inwards towards the center of the LHC, and the y-axis points vertically upward. Considering the cylindrical shape of the detector, one uses also cylindrical coordinates in order to pinpoint a specific direction in it. The radial coordinate, measured in the x-y plane, is denoted by r. The azimuthal angle ϕ is measured from the x-axis also in the x-y plane. The polar angle θ is measured from the z-axis. This enables to determine the pseudorapidity, which is defined as $\eta = -\ln(\tan(\theta/2))$. According to the above defined coordinate system, the transverse momentum and the transverse energy are calculated from the x and y components. Transverse quantities (perpendicular to the beam line) are preferred to avoid considering the non-interacting particles at an event.

Below, a short description of the subdetector system of the CMS detector is given. A slice through the detector barrel is shown in figure 3.2, which illustrates the most important components of the CMS detector. The central part of the CMS detector is the 4 T superconducting solenoid located outside of the calorimeters. A strong magnetic field is required to supply large bending power to measure accurately the momentum of high-energy charged particles. Within the solenoid is the inner tracker and the calorimeter. The tracker consists of 4 layers of silicon pixel detectors and 10 layers of silicon strip detectors. They measure the tracks of charged particles. In combination with the magnetic field, the transverse momentum of charged particles can be computed due to the bending of the trajectory. The precise measurement of the position benefits also the determination of jets originating from b- and t-quarks. The tracker is surrounded by the electromagnetic calorimeter (ECAL). It is made up of lead-tungstate ($PbWO_4$) crystals which cover a polar angle up to $|\eta| < 3.0$. The ECAL measures the energy of particles, especially electrons and positrons, and photons. The ECAL is surrounded by the hadronic calorimeter (HCAL). It has a coverage up to $|\eta| < 3.0$. The HCAL is a sampling calorimeter made of alternating layers of dense absorber and tiles of plastic scintillators. It measures the energy of strongly-interacting particles. The outermost layer is the muon system. Since muons permeate the majority of the detector system almost without interacting, a muon chamber is installed at the very end. The muon system consists of 1400 muon chambers which include drift tubes (DTs), cathode strip chambers (CSCs), and resistive plate chambers (RPCs).

In 2016 alone, the LHC produced about as many collisions as it had in the three years of its first run. The final integrated luminosity totals averaged around 40 fb^{-1} in CMS. The aimed integrated luminosity in 2017 is 45 fb^{-1}.

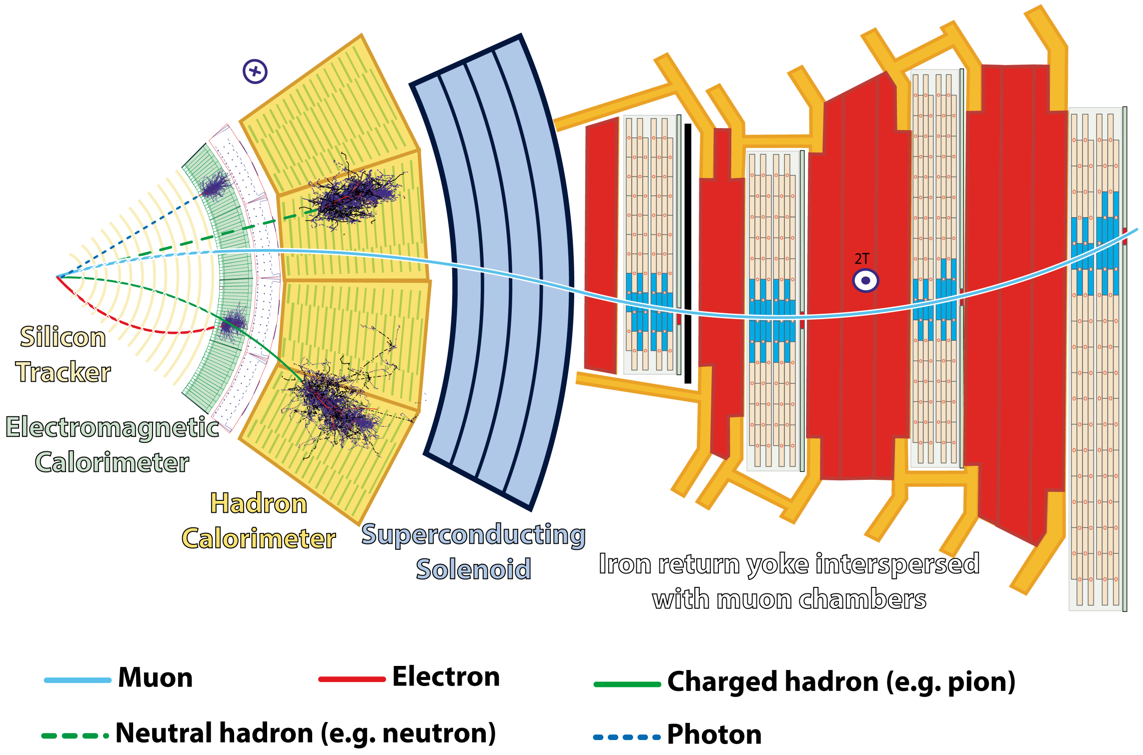
2T
Silicon Tracker
Electromagnetic Calorimeter
Hadron Calorimeter
Superconducting Solenoid
Iron return yoke interspersed with muon chambers
Muon
Electron
Charged hadron (e.g. pion)
Neutral hadron (e.g. neutron)
Photon

Figure 3.2: **Slice through the CMS detector.** The depiction shows all important layers of the detector from the inner tracker to the muon chambers. Each zone is constructed to measure special kinds of particles. Each trajectory is caused by a different particle and they leave specific tracks and signals in the respective detector system. Taken from [26].

4 Heavy bosonic resonances decaying via a top quark and a vector-like quark

This section covers an overview of the physics process examined in this thesis as long as some basic terms. For the analysis of jets, a clustering algorithm is needed, which is described in Section 4.1.1. Section 4.1.2 outlines the technique of b-tagging, as it is currently used at the CMS experiment. Section 4.1.3 describes a jet shape quantity, which is of special interest in these algorithms. Section 4.1.4 introduces the soft drop algorithm which is incorporated in all the taggers in this analysis. The fundamental Feynman diagram for the production of the Z' boson as long as the other event characteristics are presented in Section 4.2. The subsequent part 4.2.1 explains the required cuts to analyze the process and the needed reconstruction algorithms to identify the particles in the event. Due to the small cross section of the investigated process, one has to accurately determine the major background sources. The background sources considered are introduced in chapter 4.2.2. An overview of the properties of the data is given in chapter 4.3.

4.1 Jet reconstruction and identification

Owing to color confinement, quarks cannot be observed as free particles. Instead they group together to form colorless hadrons — a process referred to as hadronization. The creation of quarks in high-enery collisions therefore leads to the spontaneous creation of quarks and antiquarks to forms hadrons. The resulting narrow cone of hadrons originating from the hadronization of a single quark is known as a jet. Considering a decay process of a heavy particle, the resulting decay products possess large momenta. Due to the large momenta, even heavy SM objects are so strongly Lorentz boosted that their decay products can be clustered into a single so-called fat-jet. Thus, a technique is required to analyze the substructure of these objects to assign them to the correct particles.

4.1.1 Jet-clustering

Since jets are the only way to gather information about the object they evolve from, the investigation and perfection of reconstruction algorithms is inevitable. One possible algorithm which is used in this analysis is a special realization of a sequential recombination jet algorithm, known as anti-k_{T} algorithm [27]. For the sake of completeness there also exist so-called cone algorithms which will not be further discussed in this thesis. The main idea of a sequential recombination jet algorithm is to combine particles which are nearest

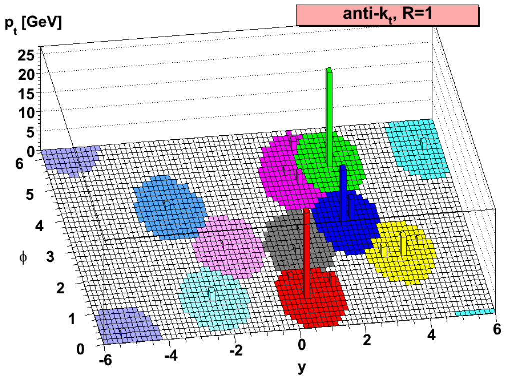

Figure 4.1: **Demonstration of the anti-k_T algorithm.** Depiction of jets in a parton-level event (generated with the MC event generator Herwig) in the rapidity-azimuth plane. As shown, the algorithm creates almost circular jets around hard seeds. Taken from [28].

to each other according to a distance measure which is invariant under longitudinal Lorentz boosts. The great advantage here is that it produces circular jet shapes around high-p_T jets as indicated in figure 4.1. In this work, the top quarks and W-bosons come with such a large boost that all their decay products can be clustered into a single jet with a radius of $R = 0.8$. The b-quarks are clustered into jets with a radius of 0.4. These jets are commonly referred to as AK8- and AK4-jet, respectively.

4.1.2 b-tagging

The identification of jets originating from b-quarks is an important and necessary tool to explore new physics. In particular in this thesis, the correct assignment of reconstructed jets to the final-state b-quarks is mandatory. B-hadrons have a relatively long lifetime in the order of $1.5 \cdot 10^{-12}$ s. This, combined with time dilation due to their large Lorentz boost allows them to travel a small distance ($\sim$ mm) before decaying. Thus, the signature of a b-quark is a jet of particles emerging from the point of the collision (primary vertex) and a secondary vertex from the b-quark decay (compare figure 4.2). The current b-jet identification algorithm from the CMS experiment is called Combinded Seconday Vertex v2 (CSVv2) [29, 30].

4.1.3 N-subjettiness

A jet shape to distinguish between boosted hadronic objects and QCD jets is “N-subjettiness”. It is usually denoted by τ_N and can be thought of as how likely it is that a specified jet contains at least N+1 subjets. It is defined as

$$\tau_\mathrm{N} = \frac{1}{d_0} \sum_k p_{\mathrm{T,k}} \mathrm{min}\left\{\Delta\mathrm{R}_{1,\mathrm{k}}, \Delta\mathrm{R}_{2,\mathrm{k}}, ..., \Delta\mathrm{R}_{\mathrm{N,k}}\right\}$$

with the summation index k, which runs over all constituent particles for a particular jet, their corresponding momenta $p_{\mathrm{T,k}}$, and $\Delta\mathrm{R}_{\mathrm{j,k}} = \sqrt{(\Delta\eta_{\mathrm{j,k}})^2 + (\Delta\phi_{\mathrm{j,k}})^2}$ which equates to the angular distance of a candidate subjet j and a constituent particle k. The normalization factor d_0 is defined as

$$d_0 = \sum_k p_{\mathrm{T,k}} \mathrm{R}_0$$

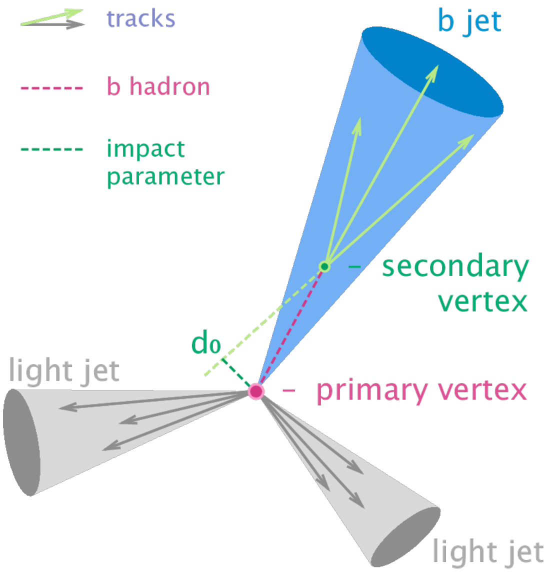

Figure 4.2: **Illustration of the principle of b-tagging.** Due to the relatively long lifetime of the b-quark, a secondary vertex develops (dotted violet line). The jets orginating from the decay point of the b-quark (green dot) can be measured and thus indicate the creation of a b-quark in the collision. Taken from [31].

wheras R_0 is the jet radius used in the jet clustering algorithm. It follows from the definition, that a given jet with N (or less) subjets correspond to a value of $\tau_\mathrm{N} \approx 0$. On the other hand, if the jet is composed of at least N+1 subjets, $\tau_\mathrm{N} \gg 0$. Since the absolute value of N-subjettiness is not a comparable quantity to separate a composed jet from the background, ratios are often used as discriminating variables. For instance, the top tagger CMSv2 used in this thesis, incorporates the ratio τ_3/τ_2.

4.1.4 Soft drop algorithm

Another jet substructure algorithm is called "soft drop declustering". A more detailed introduction can be found in [32]. In order to reduce the effects of contamination from initial state radiation, the tagging method removes wide-angle soft radiation from the jet. The technique is applied to a jet with fixed radius. The soft drop algorithm uses two parameters, a soft threshold z_cut and an angular exponent β. Roughly speaking, the procedure breaks up the jet into two subjets by undoing the last clustering process. If these subjets satisfy a particular condition, called soft-drop condition, which involves the before mentioned parameteres the jet is the final soft-drop jet. The soft-drop mass m_SD is then defined as the invariant mass of all jet constituents. The soft-drop condition is defined as

$$\frac{\min(p_\mathrm{T,i}, p_\mathrm{T,j})}{p_\mathrm{T,i} + p_\mathrm{T,j}} > z_\mathrm{cut} \left(\frac{\Delta R_\mathrm{i,j}}{R_0}\right)^\beta$$

where the indices i and j relate to the two subjets with $p_\mathrm{T,i} > p_\mathrm{T,j}$ and R_0 is the predetermined radius of the jet. As long as this condition is not fulfilled the procedure will be repeated with the highest p_T subjet until the jet cannot be further declustered. An important feature of the soft-drop procedure is for $\beta \neq 0$, the soft drop condition provides a relation between energies and angular distances.

4.2 Process characteristics

This analysis is designed to identify the decay of neutral Z′ resonances decaying to a top quark and a vector-like quark T′. The decay products are further detected as jets. Only

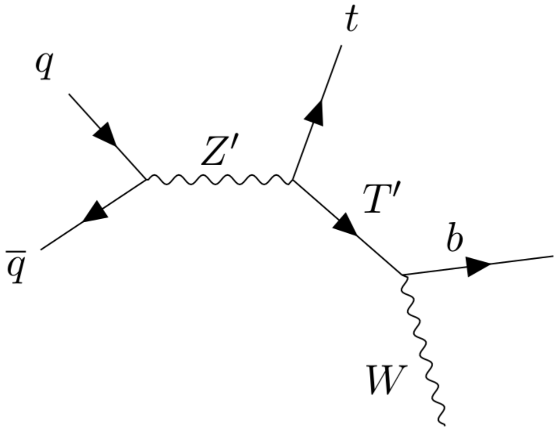

Figure 4.3: **Depiction of the leading order Feynman diagram.** Production of the Z′ boson by a quark-antiquark pair interaction and the decay Z′ → tT′ → tbW.

events with all-hadronic final states are considered suppressing events in the leptonic or semi-leptonic decay channels. Furthermore, the analysis is optimized for the decay mode T′ → bW, but it can also be applied to T′ → tH and T′ → tZ decays. Beside the restriction to an all-hadronic final state, the event is required to have at least three jets. These evolve from the top quark and the decay products of the T′.

The reference point [11] for Z′ resonances is an effective Lagrangian theory build on the minimal composite Higgs model. The neutral resonance introduced in this work is a possible candidate for the Z′ of this thesis. It is produced by quark anti-quark fusion at the LHC. At leading order, the corresponding Feynman diagram for the production as well as the decay mode is shown in figure 4.3. Considering a large margin between the W-boson mass and the T′ quark mass, the W-boson and as well its decay products are highly Lorentz boosted due to momentum conservation. The decay products of the W-boson are merged into a single AK8-jet (fat-jet). Under the assumption that the mass difference between the Z′ and T′ is large enough, the top quark also results in a highly Lorentz boosted jet. The following part describes the methods used for reconstruction and identification of the decay products.

4.2.1 Selection criteria and event reconstruction

The reconstructed particle candidates in each event are clustered into jets, once using the $R = 0.4$ (AK4-jets) and the 0.8 (AK8-jets) algorithm. These different types of jets are independently reconstructed. The clustering process only considers charged hadrons which are not associated to a non-primary vertex. Only jets with $|\eta| < 2.4$ are considered.

Due to the investigated decay channel of the Z′ only events with no leptons and at least three jets are evaluated. Since this analysis focuses on heavy-resonances, only events with a total momentum $H_\mathrm{T} > 850\,\mathrm{GeV}$ are taken into account. The total momentum H_T is defined as the sum over all transverse momenta of the particles in an event. The fully-hadronic state is characterized by two AK8-jets and one AK4-jet. One AK8-jet is associated with the Lorentz boosted top quark originating from the decay of the Z′, another AK8 corresponds to the W-boson from the decay of the vector-like quark T′. The remaining AK4-jet originates from the b-quark, also emitted by the T′ particle.

The identification of t-quarks is achieved by employing the CMS top tagger v2 algorithm [33]. In order to be addressed as a top quark candidate, the jet has to fulfill specific characteristics. The first condition that must be met is that the p_T of the jets must exceed a threshold of 400 GeV to appear as single merged AK8-jets. The soft-drop mass m_SD, calculated using all particle constituents, is required to satisfy the constraint $110 < m_\mathrm{SD} < 210\,\mathrm{GeV}$.

The N-subjettiness variable is required to satisy $\tau_3/\tau_2 < 0.86$. These conditions lead to a misidentification rate of 10% and a tag-efficiency above 70% [34]. A jet which passes all the above requirements is considered as "top-tagged". An additional requirement, such as including a b-tagged subjet, is not considered in this thesis. The analysis of W-jets employs the same algorithm as in the t-quark identification. In order to be referred to as "W-tagged", the jets must satisfy similar conditions. These contraints are $70 < m_{\mathrm{SD}} < 100\,\mathrm{GeV}$, $\tau_2/\tau_1 < 0.6$ and $p_{\mathrm{T}} > 200\,\mathrm{GeV}$. These selection criteria lead to a misidentification rate of about 5% for QCD, and a tag-efficiency of approximately 60%[34]. Lastly, to identify AK4 jets evolving from b-quarks, the CSVv2 algorithm is used. Using the "medium" working point, 99% of light-flavour jets are rejected, whereas the tag-efficiency lies around 70% for b-jets. The jets are required to have a transverse momentum greater than 100 GeV and $|\eta| < 2.4$.

Since the three different jets should not overlap, the possible AK4 jets are required to have an angular separation ΔR greater than 0.8 with respect to the t-tagged jet and W-tagged jet. AK8-jets originating from top quarks and W-bosons are mutually exclusive due to the different cuts in the corresponding taggers. Despite that the AK8 top and W-jet cannot overlap, one demands, that the angular separation satisfies $\Delta R > 0.4$ to avoid possible double counting. For the subsequent reconstruction of the T$'$ and the Z$'$, from the tagged jets, only the ones with the highest p_{T} are selected. The T$'$ quark four-momentum is calculated as the sum of the four-vector of the highest separated b-jet and W-jet. Furthermore, only events with a T$'$ mass above 500 GeV are considered. These cuts reduce the examination of uninteresting low-energy events and help to reject the background. Similarly to the vector-like T$'$ quark, the four-momentum of the Z$'$ boson is constructed as the sum of the four-vectors of the T$'$ and the highest p_{T} t-tagged jet. Finally, the Z$'$ invariant mass is computed from the reconstructed decay products. It is the sensitive variable in this analysis.

4.2.2 Background processes

Due to the small expected cross section of the investigated process compared to SM processes, the background contributions must be precisely determined. In particular for this process, the two dominant background sources are QCD multijet production as well as top quark production. The latter one contains both top quark-antiquark and single top quark contributions. The major part of the background originates from multijet QCD production. The contribution of QCD events is a consequence of the mistag probability of the algorithms. This will be later used to develop a technique for the estimate of the QCD multijet background. Due to the large contribution of the QCD multijet background to the signal region, the correct estimation of this background is necessary to learn about the desired process. There exist several techniques to estimate the QCD background in the signal region. This thesis explores a new method, referred to as "top mistag method", which will be discussed in detail in the next chapter.

4.3 Monte Carlo simulation data

The method developed in this thesis uses MC-generated samples. The signal samples are generated using MADGRAPH v5.2.2.2 [35]. In these samples the Z$'$ only decays to a top quark and a heavy vector-like quark (T$'$). Signal samples are simulated for three different mass values of the neutral spin-1 resonances, namely $1.5, 2$, and 2.5 TeV with a width of 1% respectively. The width of the T$'$ quark mass is also selected as 1%, with masses of $0.7, 0.9, 1.2$ and 1.5 TeV. The values of the resonance width are selected to lie below the detector resolution. The vector-like quark is simulated with left- and right-handed chirality. The underlying theory which was used for the simulation is based on a minimal composite Higgs model, introduced in section 2.2. The data samples are created for three decay modes of the T′ quark: $\mathrm{T}' \to \mathrm{bW}, \mathrm{T}' \to \mathrm{tH}$, and $\mathrm{T}' \to \mathrm{tZ}$. To ensure that the Z′ boson decay channel in a $\mathrm{T}'\overline{\mathrm{T}}'$ pair is kinematically suppressed, the ratio between the masses of the T′ quark and the Z′ boson in the generated samples are roughly chosen to be 1/2, 2/3, or 5/6. Depending on the relative masses of Z′ and T′, the top quark originating from the Z′ decay cannot be treated as a highly-boosted jet. Two different QCD multijet background samples are used generated with two different MC generators, namely MADGRAPH v5.2.2.2 and PYTHIA 8.2 [36]. The method will be evaluated using the samples generated by MADGRAPH. The dependence of the method on the underlying MC generator can then later be tested by applying the method to the PYTHIA samples. The top quark anti-quark production background is estimated by a simulation with the next-to-leading-order (NLO) generator POWHEG V2 [37]. Single top quark production events, which make up 20% [34] of the top quark background are likewise simulated with POWHEG V2.

5 Estimation of the multi-jet background with the top mistag method

In the following section, the method developed for the QCD multijet background estimation is described. It is also termed "top mistag" method in the following. The top mistag method is developed on simulated data. The main idea of this approach is to estimate the QCD multijet background in the signal region by simulating the fake rate of the top tagger. This serves as a correction factor (event weight), which is a measure of how well the non-top rejection of the top tagger works. It is used to weight all false top jet candidates, which should result in an estimation of the background when considering them as signal-like events in the signal region. Section 5.1 covers the assumptions necessary for the integration of the technique and gives a rough overview of the implementation. The method relies on the proper simulation of two distributions gained from MC which is dicussed in section 5.2. The method is incorporated in a larger framework. The implementation of the method in the framework is treated in section 5.3. A consistency test is performed to analyze how well the prediction estimates the background which is treated in section 5.4. Finally, the results of the analysis are covered in section 5.5.

5.1 Assumptions and idea

To apply the method to the decay channel, a few assumptions must be met in order to perform the procedure. As explained in section 4.2.2 the top quark contribution to the background is evaluated via MC simulations and therefore it does not have to be considered in the subsequent analysis. This is necessary, since the investigated background MC-events should not contain real top quarks. As a consequence, each top-tagged jet in the background sample (only QCD) corresponds to a misidentification. This of course only holds if the method is merely used on MC-samples. Dealing with real data, the QCD background is given by subtracting the simulated top quark background from the data set. That means, the method will be applied for data and the top quark MC sample and the resulting distribution of the MC sample will then subsequently subtracted from the distribution of the data (compare figure 5.1). Furthermore, the signal is assumed to be comparatively small with respect to the background, otherwise this technique would not be applicable, since this method involves also a partial contribution of the signal sample. Since this method is developed on MC data, this insinuates the prerequisit that the distribution collected from MC simulation are still valid for the corresponding distributions gained from data, taking into account MC-Data scale factors. This concerns in particular the QCD

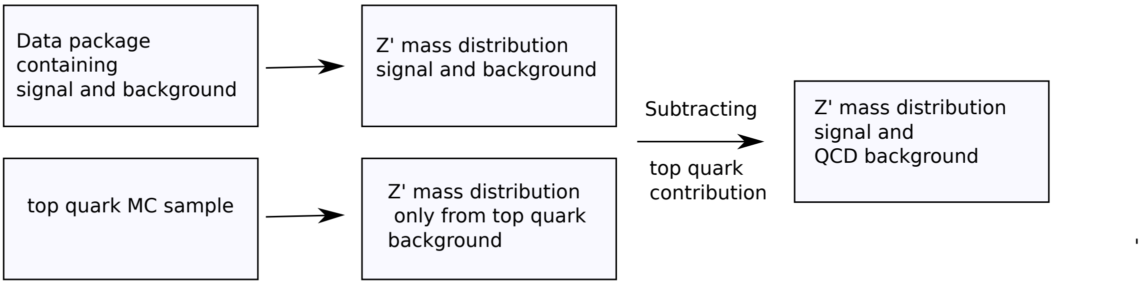
Data package containing signal and background
Z' mass distribution signal and background
top quark MC sample
Z' mass distribution only from top quark background
Subtracting
top quark contribution
Z' mass distribution signal and QCD background

Figure 5.1: **Sketch for the application to real data**. Dealing with real data, the data package contains both the signal and the background. Under the assumption that the top quark background is described by MC simulations, the Z′ mass distribution (following from the top mistag method) of the top quark background is subtracted from the one of the whole data package.

soft-drop mass (SDM) distribution and the mistag rate, which are both entirely taken from MC. The correctness of this quantities rely on the proper simulation of showering processes and jet substructures. The "mistag rate" will be explained in the next section. It is the probability that a jet which does not originate from a top quark is tagged as such. The general idea of the method is as follows: Firstly, the SDM distribution and the mistag rate have to be obtained from the QCD MC-samples. The QCD SDM distribution is needed to manually set the jet mass according to the distribution, since the top tagger is also mass-dependent. Of special interest in this thesis is the mass of the Z′ boson. The weight of each event will be calculated from the mistag rate and subsequently written in a histogram. This distribution should then lead to a proper estimation for the QCD background in the signal region. The details about the implementation are given in section 5.3. The next section explains how to the SDM distribution and the mistag rate are obtained from the QCD MC-samples.

5.2 Determination of the mistag rate and the SDM distribution from MC

Since the top tagger will always misidentify non-top jets, a small part of jets will be considered as top-tagged, although they do not really evolve from top quarks. As a result, the signal region also contains background events which overshadow the actual signal. This whole method is based on a "simulation" of the top tagger, to estimate the number of fake tags as good as possible. These fake tags are also called mistags, since they correspond to a misidentifaction of a jet, meaning a jet is considered as a top jet even if it is not. Recalling from the previous sections, a jet is considered top-tagged only if certain kinematic criteria are fulfilled. For the sake of simplicity, in this thesis the mistag rate is obtained from the MC-QCD samples. Since all the events are simulated in the MC samples, one knows which particles have been created in each event and can futher tell if the particles are correctly tagged. This is often refered to as using MC truth information. It is important to notice that this truth information is only accessible in simulations and not in real proton-proton collisions. The mistag rate is given by the number of AK8-jets which do not orignate from a top quark, but are mistakenly marked as one. The actual implementation is slightly more subtle: In each event, there can be several jets that can fake a top jet. Firstly, the AK8-jet has to have a transverse momentum larger than 400 GeV and $|\eta| < 2.4$ similar to the top-jet requirements described in section 4.2.1. Using truth information at matrix element level, one can determine if the considered AK8 jet intersects with the top quarks from the top quark background or with the top quark evolving from the decay of the Z′. If the angular distance of an AK8-jet with respect to the top quark background jets and the Z′ top quarks is larger than 0.8 in both cases, the jet is a possible mistag candidate. Applying the top tagger to these AK8-jets leads to the number of mistagged jets which will fake a top signal. Ultimately, dividing the number of mistagged jets by the number of fake AK8 candidates gives the probability to mistakenly report a top signal. Since the top tagger also depends on the transverse momentum p_{T} and the pseudo rapidity η, the mistag rate is evaluated in dependence of these two variables. Figure 5.2 shows the result of this calculation. Despite a few statistical outliers the mistag rate is very small as expected.

The weighting of the events is more complicated as it might seem at the beginning. As later explained, one has to carefully examine which AK8-jets will be taken for the reconstruction. For instance, the implementation of the reconstruction of the Z′ in the framework only considers the jet with the highest p_{T}. On the contrary, the procedure explained above to get the mistag rate takes all AK8 jets for the evaluation. This inconsistency will lead (as seen in chapter 5.4) to considerable deviations between the background prediction and the QCD background. Thus, to achieve the best possible description of the background one has to ensure consistency throughout the procedure. Several approaches exist, such as the determination of the mistag rate taking all AK8 jets or only the one with the highest p_{T}. The fake rate is determined from all selected jets, while for the actual signal event selection, only the jet with the highest p_{T} is taken. The interesting quantity is therefore the probability by which a given AK8-jet in an event is the one with the highest p_{T} which will be mistagged.

This can be made clearer by considering first an event containing only one AK8-jet. The probability that a jet will fake a top quark is simply given by the corresponding mistag rate $w_1 = \mathrm{mtr}_1$. In the next step an event containing two AK8-jets is considered. The recorded AK8-jets in an event are always ordered according to their transverse momentum, from highest to lowest. The probability that the first AK8-jet is the one of choice is again simply given by the mistag rate. The only case in which the second one is taken, is when the first one is correctly not tagged as a top jet leading to $w_2 = \mathrm{mtr}_2 \cdot (1 - \mathrm{mtr}_1)$. The second term is just the complementary probability. This result can be easily generalized to

$$w_{\mathrm{n}} = \mathrm{mtr}_{\mathrm{n}} \cdot \prod_{\mathrm{i}=1}^{\mathrm{n}-1} (1 - \mathrm{mtr}_{\mathrm{i}}).$$

The next important part is to derive the SDM distribution from the processed QCD-samples. Since each fake AK8-jet is later used to reconstruct the Z′ boson and the top tagger is mass-dependent, the mass of these AK8 jets has to be manually set according to the QCD SDM distribution in figure 5.3. Since the majority of the QCD background originates from low mass particles, the SDM of the QCD background grows for low masses. On the contrary, AK8-jets originating from top quarks result in an SDM distribution which has a distinct peak around 172 GeV. It follows that the mistagged jets from QCD follow a different SDM distribution than the SDM distribution of jets originating from real top quarks. As an important note, the intent is not to set the mass of the jets according to the SDM of the tops rather than according to the SDM distribution of the QCD jets which pass the top tagger. This has to be considered by the estimation of the number of fake AK8 jets. The next chapter explains the implentation of the top mistag method into the existing framework.

5.3 Development and implementation of the technique

From section 4.2.1 it is clear that the top tagger output depends on the SDM, the transverse momentum and the pseudo rapidity. Therefore to properly simulate the top tagger, one has to create a “mistag estimation device”, which takes AK8-jets as input and outputs the

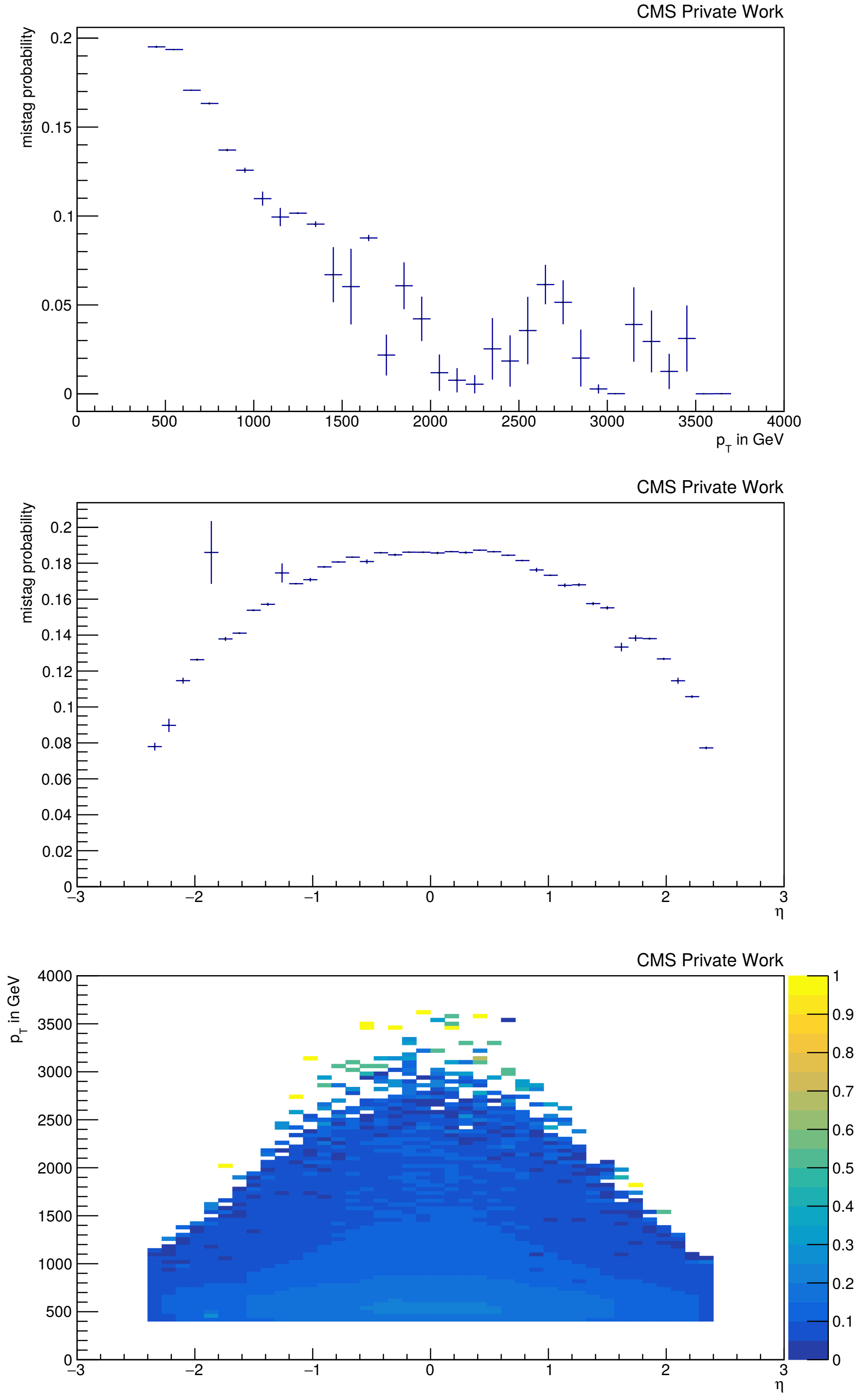

Figure 5.2: **Top-tag mistag rate as a function of the jet's η and p_T.** The jet's p_T and η dependence are separately shown in the upper and middle plot. The average mistag probability is 15%.

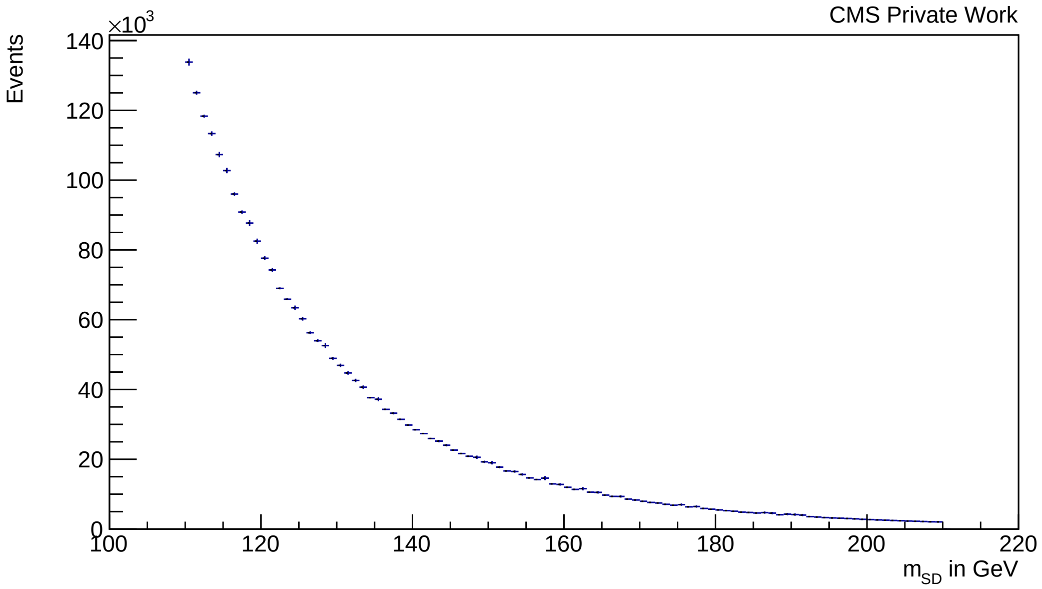

Figure 5.3: **Soft drop mass distribution of QCD**. Since the majority of the contribution comes from low mass particles, the contribution of the QCD distribution decreases rapidly. The QCD SDM distribution has only to be considered in the range $110 < m_{\mathrm{SD}} < 210\,\mathrm{GeV}$ in correspondence with the top tagger.

probability on which these jets will be mistagged in dependence of these three jet quantities. As outlined in the previous section, the mistag rate is determined in dependence of p_{T} and η. The mass dependence is taken from the SDM distribution and is manually set according to the distribution. This means that before the reconstruction of the Z′ each AK8-jet's mass is set to a random number distributed according to the contents of the QCD SDM distribution. This idea is sketched in figure 5.4. In each event in the QCD sample a set of AK8-jets may be detected. This AK8-jets can be assigned a p_{T}, η, and m_{SD}. If a AK8-jet has the right shape, it will be mistakenly tagged as a top. This top jets will than subsequently used to recontruct a Z′ boson, which is actually not existent. For the simulated top tagger (mistag estimation), the mistag rate includes the dependence of p_{T} and η. This combined with the SDM distribution should deliver a accurate estimation of the number of mistagged jets.
For the actual implementation, one has to recall that this thesis focuses on the decay channel $\mathrm{Z}' \rightarrow \mathrm{tT}' \rightarrow \mathrm{tbW}$. The AK8-jets mentioned earlier have already passed several kinematic cuts before they are ultimately go through the top tagger. For a proper simulation of the number of mistagged jets this has to be done for the simulation as well. Kinematic cuts are part of the selection criteria of the investigated decay channel. Not all gathered jets are useful for the reconstruction, since they can overlap or have too small momentum etc. Therefore, imposing cuts on the collection of determined jets reduces the number of jets that have to be evaluated. The greatest number of jets are those with low p_{T}. Since the masses of the Z′ and T′ are considered to be very large, the most basic cut to apply is to impose the condition that $H_{\mathrm{T}} > 850\,\mathrm{GeV}$. H_{T} is the sum of all tranverse momenta in an event, and using this condition, all low center-of-mass-energy events are automatically not considered. As known from the decay structure of the Z′, the interesting events are those which include a top, a W-,and a b-jet candidate which fulfill the selection criteria defined earlier. These three jets correspond to a t-quark, W-boson, and a b-quark, respectively.

First the AK8-jets from the W-boson and the t-quark are considered. Since both particles have different masses, double-counting is automatically ruled out by the W- and t-tagger in the original procedure. For the top mistag method this does not hold and the W-bosons and AK8-jets have to be properly separated. This can be solved by considering the case that from the collection of AK8-jets first all W-boson are determined. After passing the W-tagger, mostly only a subset of the AK8-jets is tagged as a W-boson. Only jets that have not been tagged as a W-boson can be considered as a t-quark. For the separation of the t-quarks from the W-jets, one can use the angular distance introduced in 4.1.1. Since the jets are treated as they have circular shape, two jets are logically considered as separated, when they do not intersect. Since the t-quark and the W-boson result both in AK8-jets, the jets are considered as separated if $\Delta R > 0.8$. This selection leaves a set of W-bosons and a set of separated AK8-jets, which could fake top jets.For the reconstruction only the jets with the highest p_{T} are considered, so that only the W-boson with the highest p_{T} is further evaluated.

Now the b-quarks have to be considered. They must also be properly separated from the W-jets and the separated AK8 jets. This is performed in the same way as for the W-jets and the AK8-jets. Only those b-quarks are kept which do not overlap with the W-jets or any AK8-jet ($\Delta R > 0.8$). This leaves three set of jets: separated b-quarks, W-bosons, and separated AK8-jets. Similar to the set of W-bosons, only the b-quark with the highest p_{T} is considered.

The next step is the reconstruction of the T′ and the Z′. The T′ can be reconstructed from the W-jet and b-jet by adding the four-momenta. Only if the mass of the reconstructed T′ exceeds 500 GeV, the Z′ boson mass is calculated. The Z′s are reconstructed by considering all separated AK8-jets rather than just the one with the highest p_{T}. The reason for this is as follows: One has to recall that this whole procedure is done to mimic the actual reconstruction process. Thus, all these applied kinematic cuts should resemble the original procedure. And in the original procedure, all AK8-jets are taken for the reconstruction of the Z′, that is why in the simulation all separated AK8-jets have to be considered as well. Before the execution of the four-vector addition the mass of the AK8-jet is manually set to a random number distributed according to the contents of the QCD SDM distribution. This whole process is illustrated in figure 5.5.
In order to make a statement about the quality of the procedure, the mass of the Z′ boson is taken as the representative physical quantity. The mistag rate is firstly used when the data is written into a histogram, wherein each event is weighted with a weight calculated with the mistag rate. The comparison of the prediction and the actual background is discussed in the next chapter.

5.4 Consistency test

To test the reliability of the method to actually estimate the background contribution in the signal region, one has to perform a consistency test. The consistency test tests whether the prediction agrees with true distribution. In the following figures the “prediction” is the mass distribution following from the top mistag method, whereas the actual background is always referred to as “QCD mistag”. Figure 5.6 shows the result of this consistency test. One can infer from the plot that the prediction agrees with the truth within the range of 1 TeV to 3 TeV with a weighted root mean squared error of 1.2%. To lower masses the deviation increases gradually. The total rate is overestimated by a factor of 1.05 or 4.7% relative to the background contribution. To conclude if this approximation is sufficient for a serious application using real data is not easy. The consistency test can only check if the method works consistently within one MC simulation. This is a necessary but not a sufficient requirement for the method to work on data. What can be deduced is that the shape of the Z′ mass distribution within the same “MC-structure” is well approximated. MC-structure refers to the MC-generator which was used to produce the file samples (compare section

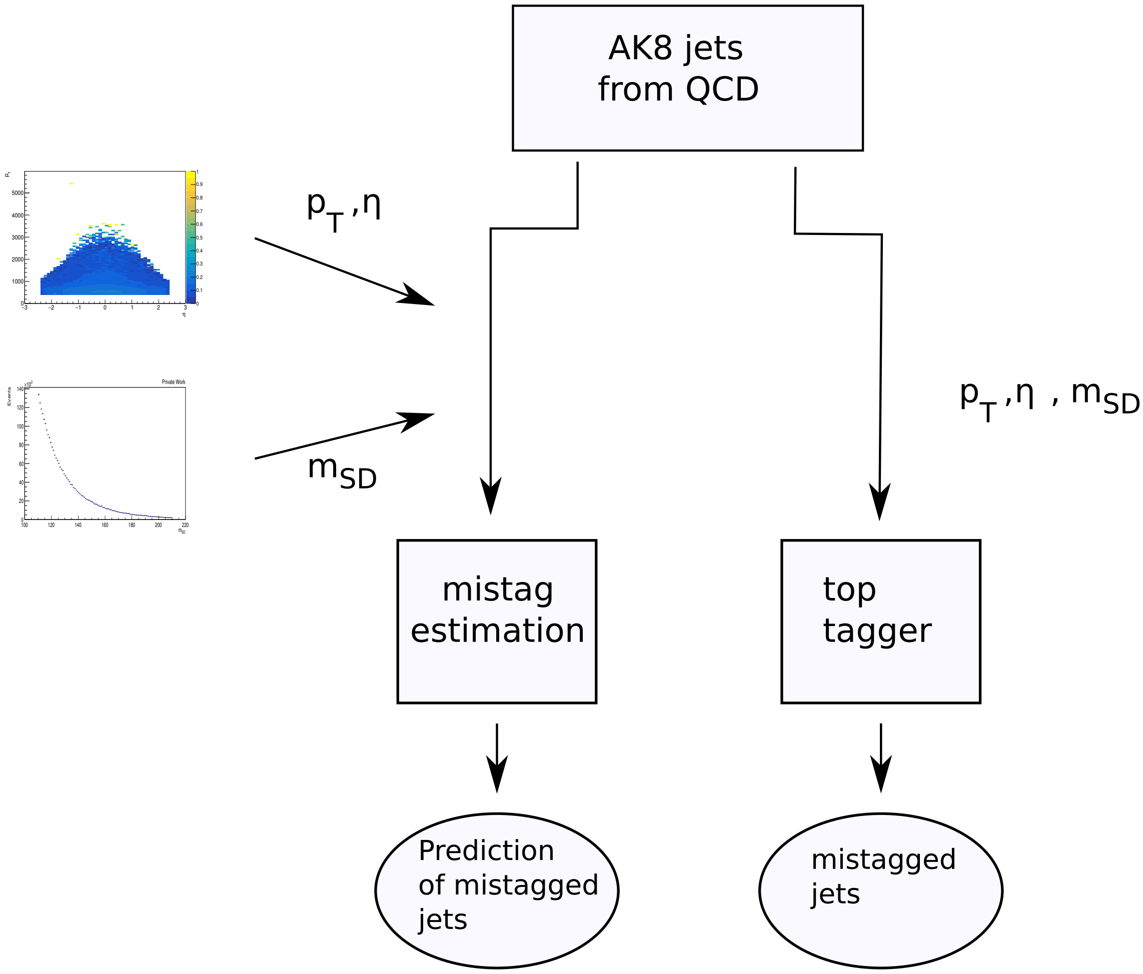

Figure 5.4: **Sketch of the top tagger simulation**. The set of AK8 jets in each event in the QCD sample is “fed” in the top tagger. If the jet (with a certain p_{T}, η, m_{SD}) satisfies the criteria of the top tagger it will be mistagged. This process is mimicked by the mistag rate (p_{T},η) and the SDM distribution.

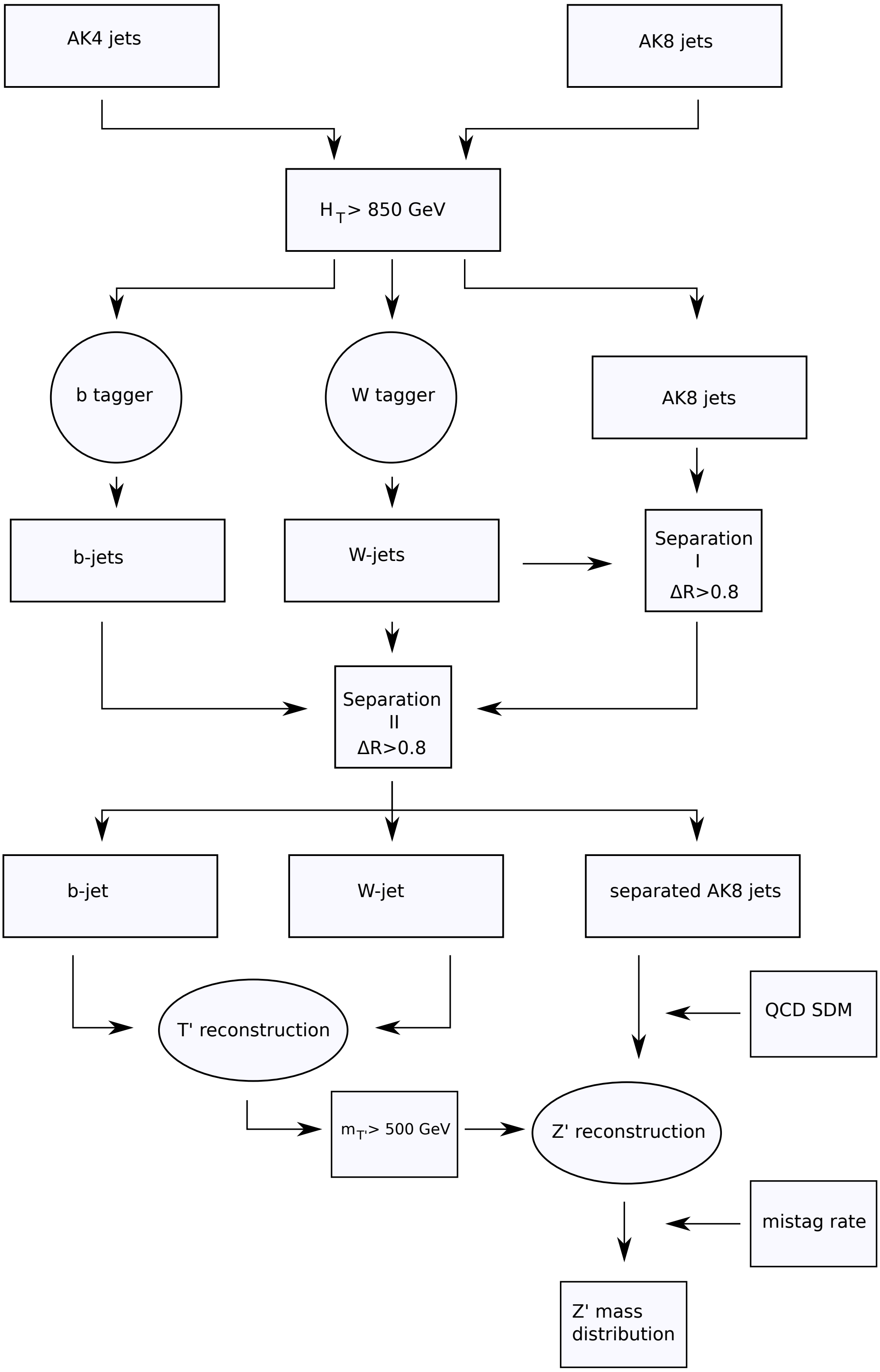

Figure 5.5: **Sketch of the workflow implementation**. This sketch depicts the different steps which have to be executed as explained in the text. Separation I and II denote the separation between W-jets and AK8 jets and the W-jet, b-jet and the AK8 jets, respectively. Finally, from the Z′ reconstruction the mass of the Z′ boson can be calculated.

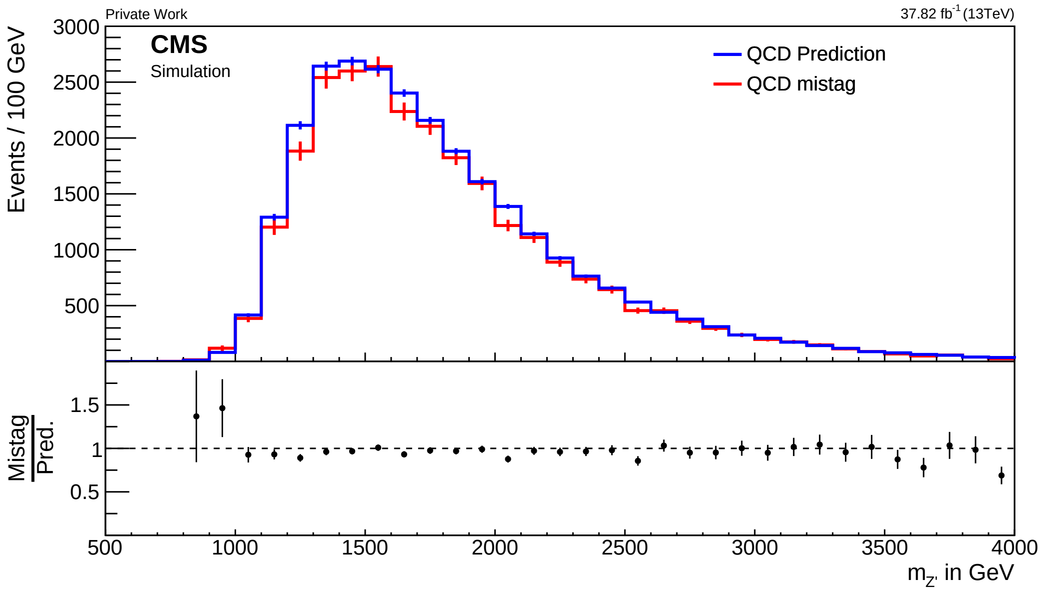

Figure 5.6: **Consistency test.** The ratio is obtained from the mass distribution of the normal procedure (entitled QCD mistag) and the mass distribution of the top mistag method (entitled QCD Prediction). The prediction approximates the best if the ratio is equal to one.

4.3). Figure 5.7 also shows the normalized distribution. Normalizing the histograms enables to compare the shape of both distributions. For the normalized distribution, the deviation of the estimation lies between 0.9% – 14.4% in the interesting interval from 1 to 3 TeV. The good match between both distributions also indicate that the top mistag method is consistent in describing the background within the same MC-structure.
Figure 5.8 exemplifies the comparison between the used and an alternative approach. In the alternative approach, each event is weighted with the corresponding mistag rate as seen in figure 5.2. It can be shown that the alternative approach has a weighted root mean squared error of 1.6% within the range of 1 TeV to 3 TeV. That means that the alternative approach has a 36% greater deviation to the unity-ratio than the main approach.

5.4.1 Signal contamination

The method has a subtle peculiarity. The whole signal sample will also contribute to the prediction of the background. Hence, all top quarks in the signal samples will thus contribute to the background prediction. This leads to a large signal contamination of the prediction and thus to an overestimation of the background (only if a signal is observed). If the contribution is large enough to overshadow the actual signal, the method loses its sensitivity. The final Z′ mass distribution, including the top quark background, signal contamination, and the signal sample itself, is shown in figure 5.9. For the evaluation, a signal sample involving a Z′ mass of 2.5 TeV and a T′ mass of 1.5 TeV is used. As indicated in the plot, despite the signal contamination, the method is still sensitive to the signal assuming a Z′ cross section of 1 pb. The ratio and as well the difference plot support the conclusion that the top mistag method is still sensitive to signal events.

5.4.2 Cross-check using a different MC-generator

All the results stated so far are obtained by using one set of MC-generated samples. This means, that the mistag rate of the top tagger is determined by using only one set of samples from a specific MC-generator. Since the top mistag method is developed by using MC data,

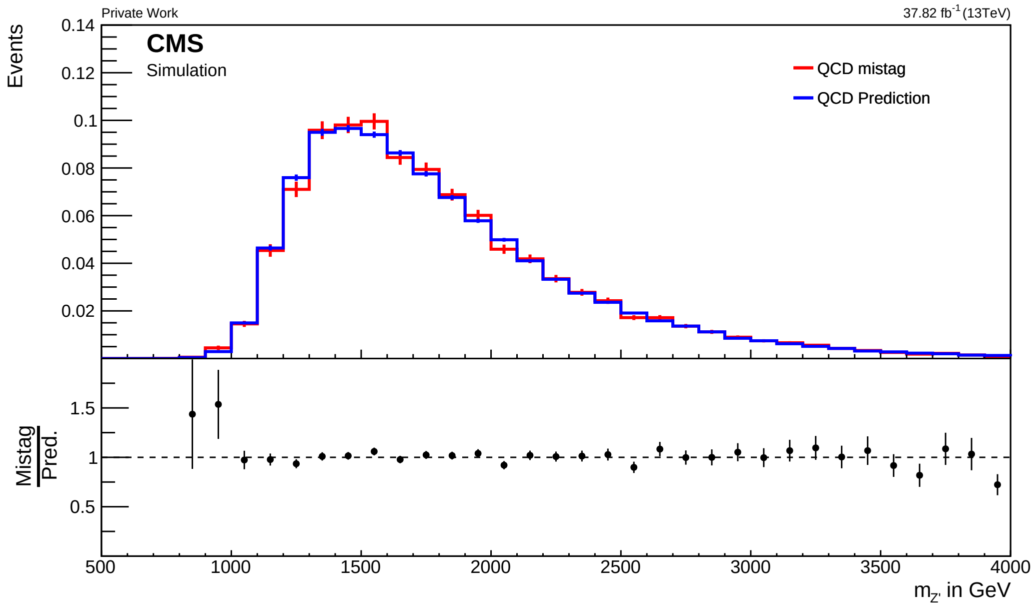

Figure 5.7: **Normalized consistency test.** The ratio is obtained by dividing the normalized normal procedure (entitled QCD mistag) with the normalized mass distribution obtained from the prediction (entitled QCD Prediction).

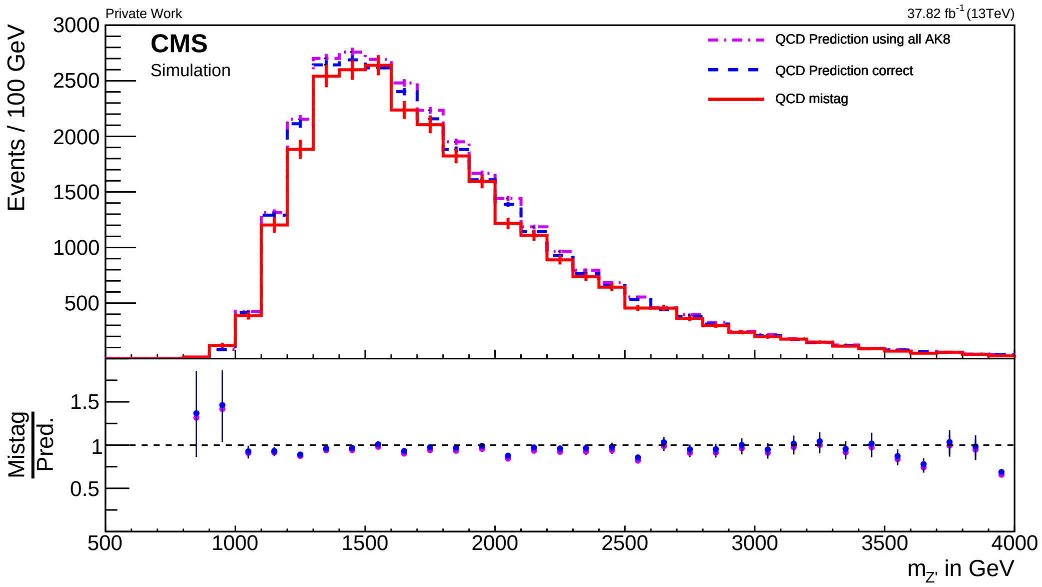

Figure 5.8: **Consistency test using different approaches for the event weight.** The three different mass distributions are the mistagged mass distribution, the prediction gained as explained in section 5.2, and an alternative prediction. This alternative prediction simply weights each event with the corresponding mistag rate. The main prediction delivers a more accurate estimation. The ratio is always executed with respect to the mistagged distribution. In the ratio plot, the triangles correspond to the main prediction whereas the dots corresponds to the alternative prediction.

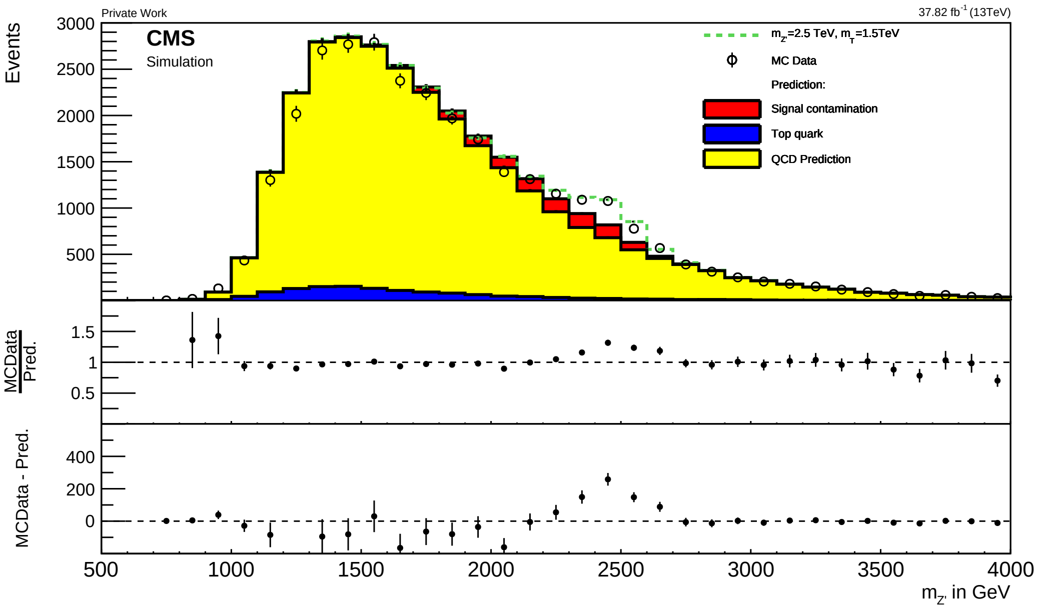

Figure 5.9: **Final consistency test with all contributions.** The total distribution contains all background contributions for Z′ → tT′ → tbW → all hadronic, as well as the influence of the signal contamination. The black circles represent the reconstructed MC events, which will be considered as data. They are composed of the QCD background, top quark background and the signal sample as indicated in the legend. The cross section for the signal sample is set to 1 pb which represents a typical cross section for such searches. The yellow distribution is the prediction of the QCD background. The blue distribution visualizes the top quark antiquark background simulated by MC. The small red part is the number of signal events, which contribute falsely to the QCD prediction. Lastly, the dotted green line reflects the signal. The image further shows the ratio and the difference between the data and the prediciton.

a reasonable question would be whether the method actually depends on the MC samples. As shown above, the results indicate a coherent solution. That means, when restricted to the same MC samples, the method supplies an adequate estimation of the QCD background within the uncertainties as shown in figure 5.6. In order to test if the developed method is independent of the MC-generator, one has to compare the QCD prediction and background from a different MC-generator using the mistag rate from MADGRAPH. The reference samples are produced with PYTHIA (compare section 4.3). Figure 5.10 shows the result of the comparison. As shown, disregarding the statistical deviations of the PHYTHIA sample, the QCD background prediction overestimates the actual QCD background in the signal region of the PYTHIA sample. It turns out that the predicted rate is 1.2 times larger than the rate of the mistagged mass distribution which equates to a deviation of 24.6%. Therefore, the estimation of the total rate in the PYTHIA sample is 19% worse than in the MADGRAPH sample, since the latter one overestimates the background only by a factor of 1.05. The weighted root mean squared error is determined to 56.8% and is therefore 45.6% larger relative to the main approach. Evaluating the normalized distributions lead to a deviation between 2.2% – 116.9% in the mass interval from 1 to 3 TeV. This leads to the conclusion, that the top mistag method still depends on MC-quantities which are not evident and are difficult to determine. Nonetheless, the lower normalized plot in figure 5.10 suggests that the shape of the background distribution can still be estimated by the top mistag method within the determined accuracy.

5.5 Conclusions and outlook

5.5.1 Results

Since the method is based on jet quantities, it has a great potential to be expanded to different decay channels under the condition that possible correlations between the mistag rate and the contributing jets are properly considered. The implementation of the method is straightforward in a way that the applicant can easily control its behavior. Despite the contribution of the signal, the method still provides enough sensitivity for a signal when neglecting systematical uncertainties. The method provides a coherent estimation of the background within the same MC-structure. The total rate of events is overestimated by a factor of 1.05. Thus, when estimating the total number of events, the method struggles and have therefore be taken from a different data-driven method. Applying the mistag rate to different MC-samples leads to the conclusion that the mistag rate is not independent of the MC-generator. The total rate of events is then overestimated by a factor of 1.2. Nevertheless, the uncertainty on the shape of the background suggests that it could still approximately be estimated. Therefore the conclusion can be drawn that the top mistag method is a potential way to estimate the shape of the QCD background within in the uncertainty range of 0.9%-14.4%. Lastly it can be said, that the judgement of the reliability of the top mistag method cannot be completely deduced from these studies.

5.5.2 Limitations

Despite the positive results of the top mistag method, it has some restrictions. The conclusion if the approximation is sufficient for the estimation of the QCD background of real data is difficult. For one reason, there is no evidence, that the simulation of this process matches with the one observed in data. Furthermore, the cross section of the Z′ production rate is a free parameter, thus it can impose a meticulous estimation for the background in order to extract any signal if the cross section is small enough. And since the top mistag method has also contribution coming from the signal itself, it is possible that the method is no longer sensitive to identifying a signal event. As the further analysis revealed, the developed method is not completely independent of the MC generator. As

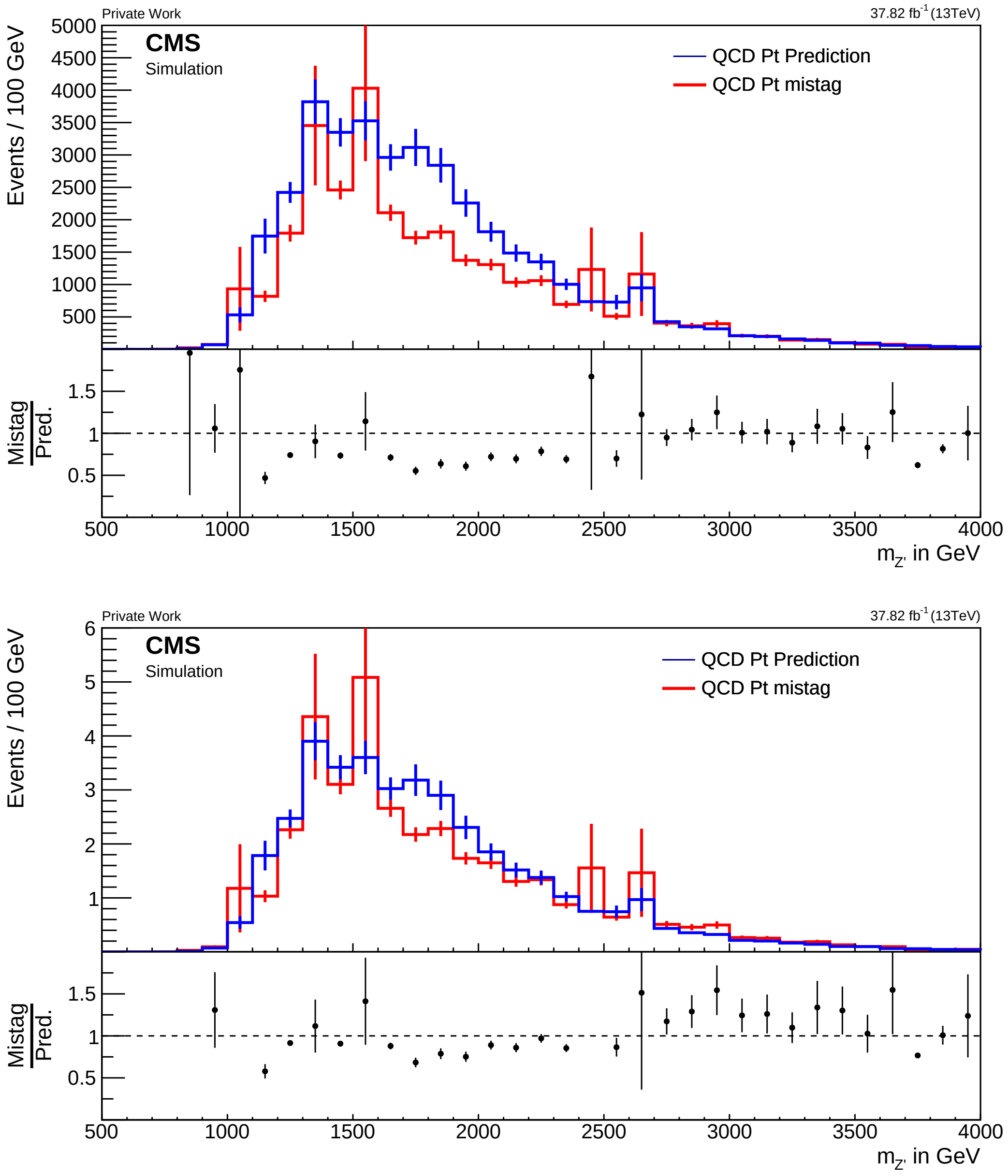

Figure 5.10: **Crosstest of the top mistag method.** The upper image shows the comparison between the background estimation applied to the Pythia QCD sample using the mistag rates derived from the MADGRAPH sample. The lower ratio plot shows the same distribution as normalized distribution.

a result, both MC-distributions, namely the SDM and the mistag rate depend on the underlying MC-generator.

5.5.3 Room for improvement

The crucial parts of the analysis, where an enhancement can be performed, is the determination of the mistag rate from MC data and the treatment of signal and background. To avoid the MC-dependence, a prospectively better approach would be to gain the mistag rate through di-jet top events, based on real measurements. Furthermore, the mistag rate is obtained in dependence of η and p_{T}. It could be also possible that the mistag rate is correlated to a subsequent W-tag or b-tag. Even if such a correlation seems unlikely, a complete investigation of the impact of the various parameters on the mistag rate could lead to more insight. Another way to improve the method is to properly distinguish between signal and background such that the signal no longer contributes to the prediction.

6 Summary

Despite the success of the SM of particle physics, many unanswered questions still exist. These unresolved questions entail for instance the nature of dark matter and dark energy. Many BSM theories attempt to improve the SM of particle physics to incorporate these unresolved phenomena. These theories have often in common that they predict the existence of massive bosonic resonances as well as vector-like quarks. The search of for any heavy bosonic resonances covers many models. Nevertheless, the minimal composite Higgs model is used as a benchmark. In this thesis, the focus lies on a hypothetical decay of heavy resonances produced in proton-proton collisions into vector-like quarks and SM particles with a fully hadronic final state. In order to identify the decay products of such heavy resonances, algorithms to identify the boosted decay products are used. The decay products can be considered as boosted objects since the mass ranges of the Z′ and the T′ lie in the TeV range. Of particular importance in this thesis are the jets orginating from the W-boson, b-quark, and the t-quark. The W-boson and b-quark originate from the decay of the vector-like T′. Only events with a high center-of-mass energy are considered due to the interesting kinematic range. If a three-jet event is found with one t-tag, b-tag and W-tag, the Z′ and T′ can be reconstructed. Since every physics process is overlayed with background contributions, understanding the latter is an important part in every particle physics analysis. Specifically in this process, the contribution of the top quark anti-quark and the QCD background can not be neglected. The background evolving from top quark anti-quark pair production is sufficiently described in MC simulations and is not further examined.

The focus of this thesis is a data driven estimation of the contribution of the QCD background in the signal region by applying the top mistag method. The method is developed by using MC generated samples. The intent of this method is to estimate the number of events which are falsely considered as signal events due to mis-tagged jets reconstructed by the top tagger. The actual method is implemented in an already existing framework. Firstly, the top mistag rate and the QCD SDM distribution have to be obtained from MC data. The mistag rate is needed to weight the distribution to simulate the misidentifactions of the top tagger, and the SM distribution is taken to manually set the mass of the jets accordingly to it. After the application of all kinematic cuts the Z′ can be reconstructed. The mass of the reconstructed Z′ is of special interest in this thesis. An estimation of the QCD background in the signal region is obtained by weighting each event with a weight calculated by the mistag rate. A consistency test is performed to compare the reliability of the developed estimation with MC data. The test shows a deviation of a factor of 1.05 between the prediction and the QCD background in terms of the total number of events. The shape of the mass distribution still shows good agreement with deviations ranging from 0.9% up to 14.4% in the Z′ mass interval from 1 to 3 TeV. Despite the fact that the signal events are a subset of the events used to predict the background (signal contamination), the method is still able to detect a signal (neglecting systematic uncertainties). The cross-check of the method with a different MC data set revealed that the accomplished accuracy does not hold for the other sample. The shape has the lowest deviation with respect to the main approach which suggests that it can still approximately be estimated.

This analysis is only the first attempt of applying this method to one specific decay channel of the Z′. Since the method is merely built using jet quantities, it can be expanded to different physics processes as well. The application of this method to real data would involve more complications. It is not certain that the method will be still sensitive for a signal after including all systematic uncertainties. The method can be improved by taking the mistag rate from real di-jet top events rather than from MC samples, to assure a higher consistency. Even if the discovery of heavy bosonic resonances cannot be verified, one can at least set new limits, which will lead to new theories trying to fit the data.

Bibliography

[1] G. Aad et al. „Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC“. In: *Phys. Lett.* (2012). doi: 10.1016/j.physletb.2012.08.020.

[2] S. Chatrchyan et al. „Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC“. In: *Phys. Lett.* (2012). doi: 10.1016/j.physletb.2012.08.021.

[3] L. Susskind. „Dynamics of spontaneous symmetry breaking in the Weinberg-Salam theory“. In: *Phys. Lett. D* (1979). doi: 10.1103/PhysRevD.20.2619.

[4] S. Weinberg. „Implications of dynamical symmetry breaking“. In: *Phys. Rev. D* (1976). doi: 10.1103/PhysRevD.13.974.

[5] S. Dimopoulos and H. Georgi. „Softly broken supersymmetry and SU(5)“. In: *Nucl. Phys. B* (1981). doi: 10.1016/0550-3213(81)90522-8.

[6] J.L. Rosner. „Prominent decay modes of a leptophobic Z'“. In: *Phys. Rev. B* (1996). doi: 10.1016/0370-2693(96)01022-2.

[7] K.R. Lynch, M. Narain, E.H Simmons, and S. Mrenna. „Finding Z' bosons coupled preferentially to the third family at CERN LEP and the Fermilab Tevatron“. In: *Phys. Rev. D* (2001). doi: 10.1103/PhysRevD.63.035006.

[8] CMS Collaboration. „Search for anomalous $t\bar{t}$ production in the highly-boosed all-hadronic final state“. In: *JHEP* (2012). doi: 10.1007/JHEP09(2012)029.

[9] CMS Collaboration. „Searches for resonant $t\bar{t}$ production in proton-proton collisions at $\sqrt{s} = 8$ TeV“. In: *Phys. Rev. D* (2016). doi: 10.1103/PhysRevD.93.012001.

[10] C. Bini, R. Contino, and N. Vignaroli. „Heavy-light decay topologies as a new strategy to discover a heavy gluon“. In: *JHEP* (2012). doi: 10.1007/JHEP01(2012)157.

[11] D. Greco and D. Liu. „Hunting composite vector resonances at the LHC: naturalness facing data“. In: *JHEP 12(2014) 126* (arXiv:1410.2883). DOI: `10.1007/JHEP12(2014)126`.

[12] N. Vignaroli. „New W' signals at the LHC“. In: *Phys. Rev. D* (2014). doi: 10.1103/PhysRevD.89.095027.

[13] M. Thomson, *Modern Particle Physics,* New York: Cambridge University Press, 2013.

[14] Wikimedia Commons. *Standard Model of Elementary Particles.* general Photo. [Online; accessed May 18, 2017]. June 2006. URL: `https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg`.

[15] B. P. Schmidt and S. Perlmutter. „Measuring Cosmology with Supernovae“. In: *Lect.Notes Phys.598:195-217,2003* (arXiv:astro-ph/0303428).

[16] S.P. Martin. „A Supersymmetry Primer“. In: *Adv. Ser. Dir. High Energy Phys.* 18 (2016). doi: 10.1142/9789812839657001.

[17] E. Pontón et al. „Minimal Composite Higgs Models at the LHC“. In: *JHEP06(2014)159* (arXiv:1402.2987). DOI: `10.1007/JHEP06(2014)159`.

[18] Y. Okada and L. Panizzi. „LHC signatures of vector-like quarks“. In: *arxiv* (2012). arXiv:1207.5607 [hep-ph].

[19] J. A. Aguilar-Saavedra et al. „A handbook of vector-like quarks: mixing and single production“. In: *Phys. Rev. D 88, 094010* (2013). arXiv:1306.0572.

[20] L.Evans and P. Bryant. „LHC Machine“. In: *JINST,* vol. 3 (2008), p. S08001.

[21] CMS Collaboration. „The CMS experiment at the CERN LHC“. In: *JINST,* vol. 3 (2008), p. S08004.

[22] ALICE Collaboration. „The ALICE experiment at the CERN LHC“. In: *JINST,* vol. 3 (2008), p. S08002.

[23] ATLAS Collaboration. „The ATLAS Experiment at the CERN Large Hadron Collider“. In: *JINST,* vol. 3 (2008), p. S08003.

[24] LHCb Collaboration. „The LHCb Detector at the LHC“. In: *JINST,* vol. 3 (2008), p. S08005.

[25] C. De Melis. *The CERN accelerator complex. Complexe des accélérateurs du CERN.* general Photo. [Online; accessed May 17, 2017]. Jul 2016. URL: `https://cds.cern.ch/record/2197559`.

[26] D. Barney. *CMS Detector Slice.* general Photo. [Online; accessed May 17, 2017]. Jan 2016. URL: `https://cds.cern.ch/record/2120661`.

[27] M. Cacciari et al. „The anti-kt jet clustering algorithm“. In: *JHEP* (2008). doi: 10.1088/1126-6708/2008/04/063.

[28] G.P. Salam. „Towards jetography“. In: *Eur.Phys.J.C* (2010). doi: 10.1140/epjc/s10052-010-1314-6.

[29] CMS Collaboration. „Identification of b-quark jets with the CMS experiment“. In: *JINST* 8 (2013). doi: 10.1088/1748-0221/8/04/P04013, P04014.

[30] CMS Collaboration. „Identification of b quark jets at the CMS experiment in the LHC Run 2“. In: *CMS Physics Analysis Summary CMS-PAS-BTV-15-001* (2016).

[31] Wikimedia Commons. *B-tagging diagram.* general Photo. [Online; accessed July 30, 2017]. May 2016. URL: `https://commons.wikimedia.org/wiki/File:B-tagging_diagram.png`.

[32] A.J. Larkoski et al. „Soft drop“. In: *JHEP* (2014). doi: 10.1007/JHEP05(2014)146.

[33] CMS Collaboration. „Top Tagging with New Approaches“. In: *CMS Physics Analysis Summary CMS-PAS-JME-15-002* (2016).

[34] CMS Collaboration. „Search for a heavy resonance decaying to a top quark and a vector-like top quark at $\sqrt{s} = 13$ TeV“. In: *CMS Physics Analysis Summary CMS-B2G-16-013* (2017). arxiv: 1703.06352.

[35] J. Alwall et al. „The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations“. In: *JHEP* (2014). doi: 10.1007/JHEP07(2014)079.

[36] T. Sjöstrand et al. „An introduction to PYTHIA 8.2“. In: *Comput. Phys. Commun.* 191 (2015). doi: 10.1016/j.cpc.2015.01.024, pp. 159–177.

[37] P. Nason. „A new method for combining NLO QCD with shower Monte Carlo algorithms“. In: *JHEP 11* (2004). doi: 10.1088/1126-6708/2004/11/040.