technische universität
dortmund

**Arbeit zur Erlangung des akademischen Grades
Bachelor of Science**

First studies on the distinction of strange and bottom jets with the ATLAS detector in Run 2

Nils Julius Abicht
geboren in Hamburg

2017

Lehrstuhl für Experimentelle Physik IV
Fakultät Physik
Technische Universität Dortmund

Erstgutachter: Prof. Dr. Kröninger
Zweitgutachter: Prof. Dr. Dr. Rhode
Abgabedatum: 24. Juli 2017

Abstract

In this thesis, the discrimination strength of a set of variables between jets originating from bottom quarks and jets originating from strange quarks is discussed. Variables that describe the jet, variables that describe the jet substructure, variables that describe the jets' calorimeter clusters as well as b-tagging variables are considered. The discrimination between bottom and strange jets is necessary in order to identify the weak decay of a top quark into a strange quark over the large background of weak top quark to bottom quark decay with the ATLAS detector. To evaluate the discrimination strength, Monte Carlo samples are used that simulate top-quark pair production in proton-proton collision at a center of mass energy of $\sqrt{s} = 13\,\mathrm{TeV}$ and their subsequent decay to bottom, strange, as well as light quarks in the final states. It was observed that in addition to b-tagging variables with good discrimination strength, calorimeter cluster variables show some discrimination strength as well. Jet and jet-substructure variables are only useful for low jet p_T.

Kurzfassung

In dieser Arbeit wird die Unterscheidungskraft verschiedener Variablen zwischen Jets, die aus Strangequarks, und Jets, die aus Bottomquarks entstehen, betrachtet. Zusätzlich zu b-tagging-Variablen werden Variablen verwendet, die den Jet selbst, die Jet-Substruktur oder die Cluster des Jets im Kalorimeter beschreiben. Die Unterscheidung zwischen Bottom- und Strangejets ist essentiell für die direkte Messung von Topquark-Zerfällen in Strangequarks und W-Bosonen neben dem dominanten Untergrund aus Topquark-Zerfällen zu Bottomquarks und W-Bosonen mit dem ATLAS Detektor. Um die Unterscheidungskraft zu bewerten, werden Monte-Carlo-Simulationen benutzt, die Topquark-Paarerzeugung in Proton-Proton-Kollisionen mit einer Schwerpunktsenergie von $\sqrt{s} = 13\,\mathrm{TeV}$ sowie deren Zerfall in Bottom-, Strange- und Lightquarks im Endzustand beinhalten. Es wurde beobachtet, dass neben b-tagging-Variablen auch die untersuchten Kalorimeterclustervariablen eine gewisse Unterscheidungskraft besitzen. Die Jet- und Jetsubstrukturvariablen sind hingegen nur bei niedrigem Jet-p_T nützlich.

Contents

1 Introduction

At the Large Hadron Collider (LHC) [1] at CERN, the Standard Model of particle physics (SM) is tested and studied by producing elementary particles in high energy particle collisions. As the top quark has a short lifetime and the largest mass of any particle known today, it is of particular interest to the search for physics beyond the SM (BSM).

The top quark decays almost exclusively via the weak interaction into W bosons and bottom quarks. However, the Cabibbo-Kobayashi-Maskawa matrix (CKM matrix) predicts that $0.16\,\%$ of top quarks instead decay into W bosons and strange quarks [2]. This decay is yet to be measured. In order to measure the decay into strange quarks over the large background of the decay into bottom quarks, it is necessary to develop methods that can discriminate between jets originating from strange quarks (strange jets) and those originating from bottom quarks (bottom jets). There are already various methods that allow for the detection of bottom jets (b-tagging).

This thesis examines the discrimination strength between bottom and strange jets of a set of different variables, such as b-tagging variables, variables that describe the properties of the jet, as well as variables that describe the clusters, the jets are reconstructed from, in the calorimeters. The discrimination strength is obtained from Monte Carlo (MC) simulations of semileptonic top pair decays in which quarks, and hence their flavours, are associated to jets. The semileptonic channel is chosen because it ensures a sufficient number of strange quarks from hadronic W decays in the final state. These strange quarks are assumed to have similar characteristics as strange quarks from the top quark decay.

This thesis is structured as follows: Chapter 2 gives a short overview over the SM, the ATLAS detector and properties of the top quark. Chapter 3 features a summary of the anti-k_{t} clustering algorithm, b-tagging algorithms as well as the Monte Carlo generators and the event preselection used for this thesis. The variables analized in this thesis are presented in chapter 4. In chapter 5, a concluion is given.

2 Top quarks in the Standard Model and the ATLAS detector at the LHC

2.1 The Standard Model of particle physics

The SM is a quantum field theory that describes three of the four fundamental interactions between all elementary particles: The strong, the electromagnetic (EM) and the weak interaction, but not the gravitational force. All elementary particles with spin $s = 1/2\,\hbar$ (fermions) are divided into two groups: fermions with color charge (quarks) and fermions without color charge (leptons).

Quarks with electric charge $q = +2/3\,e$ (up quarks, charm quarks and top quarks) are called up-type quarks while those with $q = -1/3\,e$ (down quarks, strange quarks and bottom quarks) are called down-type quarks. These six quarks are categorized into three generations. Each generation consists of an up- as well as a down-type quark with the first generation containing the quarks with the lowest masses and the third generation those with the highest masses.

Leptons are categorized into three generations as well, sorted by the mass of the charged leptons, similarly to quarks. Each generation contains a charged lepton with $q = -1\,e$ (electrons, muons and tau leptons) and the corresponding uncharged lepton (neutrino).

In addition, for every fermion there is an antifermion with the same mass and the same spin but with opposite charge quantum numbers.

Furthermore, there are particles with integer spin, called bosons. The Higgs boson is the only boson with $s = 0\,\hbar$. Bosons with $s = 1\,\hbar$ (gluons, photons, W and Z bosons) are called gauge bosons and are the force carriers within the SM.

The strong interaction is mediated by gluons, which are massless particles that carry color charge. They exclusively interact with other particles carrying color charge, including themselves. The EM interaction is mediated by the massless and uncharged photon, which only interacts with electrically charged particles. In contrast to those interactions, the weak interaction is carried via gauge bosons with mass, the W and Z bosons. The W bosons have $q = \pm 1\,e$ and allow quarks to

change their flavour as well as charged leptons to become neutrinos and vice versa. Table 2.1 shows an overview of all SM particles.

Table 2.1: Elementary particles of the SM, top to bottom: quarks, leptons, gauge and scalar bosons.

Fermions, $s = 1/2\,\hbar$			
El. charge	First generation	Second generation	Third generation
$+2/3\,e$	Up quark	Charm quark	Top quark
$-1/3\,e$	Down quark	Strange quark	Bottom quark
$-1\,e$	Electron	Muon	Tau lepton
$0\,e$	Electron neutrino	Muon neutrino	Tau neutrino
Bosons			
$s = 0\,\hbar$	Photon, $W^{\pm}/Z^0$ boson, Gluon		
$s = 1\,\hbar$	Higgs boson		

2.2 Top-antitop production

In 1995 the top quark was discovered by the CDF Collaboration [3] and the D0 Collaboration [4]. It has the largest mass of any particle known today, $m_\text{t} = 173.34 \pm 0.27\,\text{(stat.)} \pm 0.71\,\text{(syst.)}\,\text{GeV}$ [5]. Being unstable, with a mean lifetime of $\tau \approx 3.3 \times 10^{-25}\,\text{s}$ [2], in order to study the top quark, it has to be produced, which is done at the Large Hadron Collider (see Chapter 2.4) in proton-proton collisions. For this thesis, top-antitop pairs ($t\bar{t}$) are important. Figure 2.1 shows the leading order Feynman-diagrams for $t\bar{t}$-production from quark-antiquark annihilation as well as from gluon fusion.
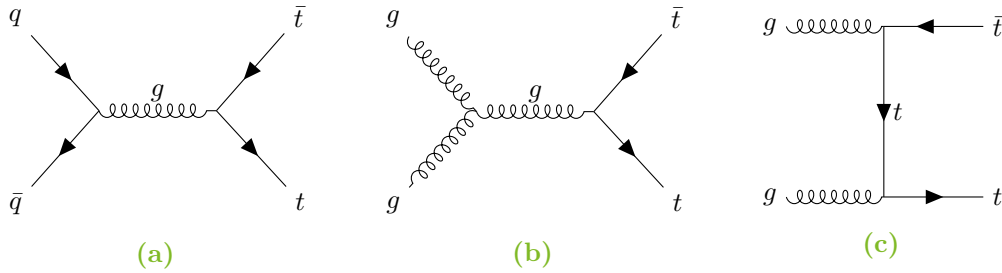

Figure 2.1: Leading order Feynman diagrams for $t\bar{t}$ production in proton-proton collisions through quark-antiquark annihilation (a) and through gluon fusion (b, c).

2.3 Top quark decay

The top quark decays almost exclusively into a bottom quark and a W boson. The latter decays further into either a quark-antiquark pair (hadronic decay) or into a lepton and the corresponding neutrino (leptonic decay) [1]. The branching ratio (BR) is measured as $BR_{\mathrm{had}} = (67.4 \pm 0.3)\,\%$ for the hadronic decay and $BR_{\mathrm{lep}} = (21.3 \pm 0.2)\,\%$ for the leptonic decay[2], respectively [2]. From this, the BRs of the two W bosons in $t\bar{t}$ events decaying either both leptonically (dileptonic $t\bar{t}$ decay) or one leptonically and one hadronically (semileptonic $t\bar{t}$ decay) are calculated as $BR_{\mathrm{dilep}} = (4.5 \pm 0.1)\,\%$ and $BR_{\mathrm{semilep}} = (28.7 \pm 0.3)\,\%$, respectively. As the CKM matrix postulates, the product of the weak-interaction top-quark decay is a superposition of down-type quarks. Therefore, with a BR of $BR_{t\bar{t}\to WWbs} = (0.164 \pm 0.009)\,\%$ [2], a $t\bar{t}$ pair is expected to decay into a W boson and a strange quark instead of a bottom quark. Figure 2.2 shows a corresponding Feynman diagram.
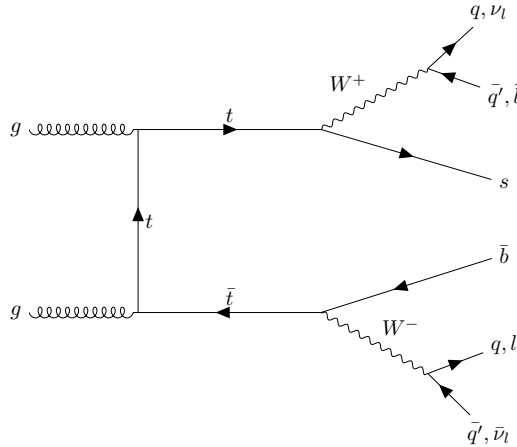

Figure 2.2: Leading order Feynman diagram for $t\bar{t}$ production and decay with a strange quark in the final state.

---

[1] The leptonic decay either consists of a charged lepton and the corresponding antineutrino or a charged antilepton and the corresponding neutrino.

[2] In this case, leptonic refers to electrons or muons.

2.4 The Large Hadron Collider and the ATLAS detector

At the European Organization for Nuclear Research (CERN) laboratory in Geneva, elementary particles are produced and studied in particle collision. With $27\,\text{km}$ circumference, the LHC is the largest collider worldwide. At the LHC, heavy ions as well as protons are brought to collision with a maximum center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$.

The ATLAS experiment [6] is one of the four large experiments located at the LHC, the others being ALICE, the LHCb and the CMS experiments. Figure 2.3 shows the layout of the ATLAS detector. The detector is built symmetrically around the beam pipe with the azimuthal angle around the pipe ϕ ranging from 0 to 2π and the pseudorapidity η ranging from 0 to 4.9, defined as $\eta = -\ln{(\tan{\theta})}$, with θ being the angle to the beam axis. A distance ΔR in the ϕ-η plane can be defined as $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$. The detector consists of an inner detector, an electromagnetic (EM) as well as a hadronic calorimeter, and muon chambers, all of which are divided into central barrel and end-cap regions.

The inner detector measures particle tracks on a micrometer scale and is divided into three subdetectors: A pixel detector with three layers and an additional insertable B-layer [7, 8], a subdetector working with silicon strips and the transition radiation tracker, which uses gas filled tubes to track particles. Those subdetectors are surrounded by a solenoid magnet with a magnetic field of $2\,\text{T}$ [6], which is used to bend the tracks of charged particles and thereby measure their momentum.

The sampling calorimeters measure the energy deposited in clusters by electrons, positrons, photons and hadrons. They consist of alternating layers of absorber material and active material. In the EM calorimeter, lead, or copper in the foward region, is used as absorber material with liquit argon (LAr) as scintillator. The barrel part covers pseudorapidities of $|\eta| < 1.475$, while the end-cap components cover $1.375 < |\eta| < 3.2$. The hadronic calorimeter uses steal absorber with scintillating tiles as active material in the region $|\eta| < 1.7$. In the end-cap region, copper/LAr as well as tungsten/LAr calorimeter moduls cover pseudorapidities of $1.5 < |\eta| < 4.9$.

The muon chambers are located in a magnetic field created by superconducting toroid magnets and measure muon tracks in the range of $|\eta| < 2.7$. For the central part this is done with monitored drift tubes but for the range $2.0 < |\eta| < 2.7$, cathode strip chambers are used instead to withstand the higher flux in this region.
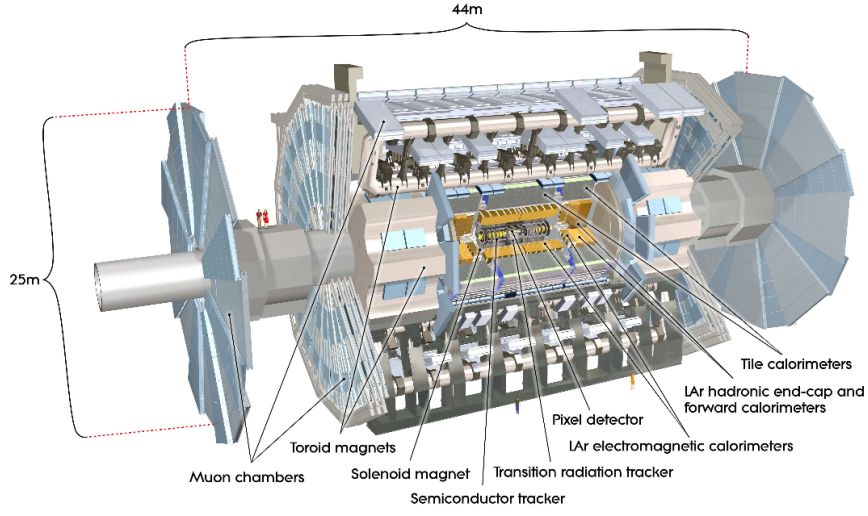
Figure 2.3: Schematic cut-away view of the ATLAS detector [6].

3 Object reconstruction and Monte Carlo simulation

3.1 Jet reconstruction with the anti-k_t clustering algorithm

The ATLAS Collaboration uses the anti-k_t jet clustering algorithm [9] to reconstruct jets from energy deposits in the calorimeters. The algorithm calculates the distances d between all calorimeter clusters i and j as well as their distance to the beam (B). The distances are defined as:

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

where $p_{\mathrm{T},i}$ is the transverse momentum of the cluster i, ΔR_{ij}^2 is the distance between the clusters i and j in the ϕ-η plane and R is a radius parameter, set to 0.4 for jets studied in this thesis. If the smallest distance is the distance between two clusters, they are combined into one new cluster by adding their four momenta, whereas, if the smallest distance is the distance between a cluster and the beam, the cluster is removed and called a jet. This is repeated until no clusters remain.

Advantages of the anti-k_{t} algorithm, compared to for example cone algorithms used at the Tevatron [10], are that it is infrared and collinear safe as well as resilient to soft radiation as the jet cones solely depend on hard particles.

3.2 b-tagging algorithms

The identification of bottom jets (*b*-tagging) plays a vital part in the discrimination between strange and bottom jets. The most important information for *b*-tagging algorithms are the tracks of charged particles reconstructed in the inner detector up to a pseudorapidity of $\eta = |2.5|$. Three basic kinds of *b*-tagging algorithms are used in ATLAS that produce variables that can be used for identifying bottom jets, which are described in the following. The variables produced by these algorithms are combined into one multivariate (MV) variable with optimized discrimination strength (Section 3.2.4).

3.2.1 Inclusive secondary vertex reconstruction algorithm (SV)

Bottom quarks form bound states (*B* mesons) that are rather long lived with lifetimes of about 1.5×10^{-12} s [2]. Therefore, they propagate a certain distance of up to a few millimeters before they decay. Secondary vertex algorithms [11] try to match tracks to a secondary vertex inside the jet. If the secondary vertex is reconstructed, a set of variables that describe the properties of this vertex can be used to discriminate between bottom and non-bottom jets.

3.2.2 Impact parameter based algorithms (IP2D, IP3D)

Secondary-vertex tracks have an impact parameter d_0 with respect to the primary vertex, defined as the track's closest distance to the primary vertex. The algorithm labels the sign of an impact parameter positive, if the point of closest approach lies in front of the primary vertex with respect to the jet direction, and negative, if behind. Tracks from bottom-related decays tend to have a positively signed impact

parameter. A likelihood ratio technique is used on the signed impact parameter significances d_0/σ_{d_0} of all secondary vertex tracks to identify bottom jets. The IP2D algorithm [11] only uses the impact parameter in the ϕ-η plane (transverse impact parameter), whereas the IP3D algorithm [11] also considers where the closest point of approach lies on the beam axis (longitudinal impact parameter). While the IP3D algorithm has a potentially stronger discrimination strength than the IP2D algorithm, this comes at the expense of being more susceptible to pile up events, which typically have a large longitudinal impact parameter.

3.2.3 Decay chain multi-vertex reconstruction algorithm (JetFitter)

The JetFitter algorithm [11] tries to reconstruct the flight path of the B meson and its decay products, which shows a unique topology. With this flight path, it is possible to reconstruct the secondary vertex as well as the vertex from the charm quark decay in situations where it would not have been possible otherwise, for example with only one track matched to the vertex. Reconstructing vertices with only one associated track has the advantage of allowing for a higher efficiency but weakens the discrimination strength against non-bottom jets.

3.2.4 Multivariate algorithms

The combination of the discriminating variables obtained from the basic algorithms is done with a boosted decision tree (BDT) with those variables used as input parameters. For the purpose of training, a set of approximately 5 million simulated $t\bar{t}$-events is used with bottom jets considered signal and charm and light jets[1] considered background. Depending on the ratio of charm to light jets used in the training (20/100, 10/100 or 0/100), three MV algorithms are currently used at the ATLAS experiment: The MV2c20, the MV2c10 and the MV2c00 algorithm [11] respectively. While the use of charm jets in the training slightly decreases the rejection of light jets, it increases the rejection of charm jets.

3.3 Monte Carlo simulation

In order to examine the discrimination strength of variables between bottom jets and strange jets, Monte Carlo simulations are necessary. The matrix element (ME) calculation for the $t\bar{t}$ events is done with the next-to-leading-order (NLO) generator

---

[1]These are jets originating from up and down quarks.

Powheg-Box 2.0 [12–15] and the proton parton distribution function (PDF) set CT10 [16]. This is done assuming a top quark mass of 172.5 GeV and that a top quark always decays into a bottom quark and a W boson. Showering and the underlying event (UE) are simulated with the Pythia 6.428 generator [17] with the CTEQ6L PDF set [18] and the Perugia 2012 UE tune [19]. Minimum bias pile up events are generated by the Pythia 8.186 generator [20] with the MSTW 2017 LO PDF set [21] and the A2 tune [22] and are overlaid on the $t\bar{t}$ events, with a distribution of the number of interactions per bunchcrossing similar to that in the 2015 and 2016 data taking. The obtained samples are processed through a simulation of the detector geometry [23] and the response is simulated with the Geant4 software toolkit [24].

3.4 Event preselection

Generated events have to meet certain selection criteria to be considered. This way, events that would not be relevant in real measurements are cut out and no resources are wasted. Events have to contain one electron with $E_\mathrm{T} > 24$ GeV or muon with $p_\mathrm{T} > 24$ GeV. In addition, electrons are required to have at least a medium likelihood electron identification [25] and only muons with tracks in the inner detector as well as in the muon chambers are accepted. Both electrons and muons have to meet *gradient* isolation criteria, described in Ref. [26] and [27]. Furthermore, at least four jets and missing transverse momentum of $\not{E}_\mathrm{T} > 30$ GeV are required.

4 Discrimination strength of different variables

4.1 Truth-matching

In order to access the discrimination strength between jet flavours of a jet-related variable from the MC samples, truth particles (partons) from the $t\bar{t}$-decay are matched with jets in the same event to determine the jet flavour. Partons and jets are matched, if the distance between their tracks in the ϕ-η plane is smaller than 0.4.

In the Chapters 4.2 to 4.5, receiving operating characteristic (ROC) curves are obtained from the distributions of a set of variables describing the jet itself, the jet calorimeter clusters or the jet substructure, as well as b-tagging variables. To reduce correlation effects between the variables and the transverse jet momentum (jet p_T), the distributions are analyzed separately in five jet-p_T intervals, ranging from 25 GeV to 50 GeV (low jet p_T), 50 GeV to 75 GeV, 75 GeV to 100 GeV (medium jet p_T), 100 GeV to 150 GeV and higher than 150 GeV (high jet p_T). Figure 4.1 shows the jet-p_T distribution for bottom, strange as well as light jets. Bottom jets tend to have larger jet p_T than strange and light jets.
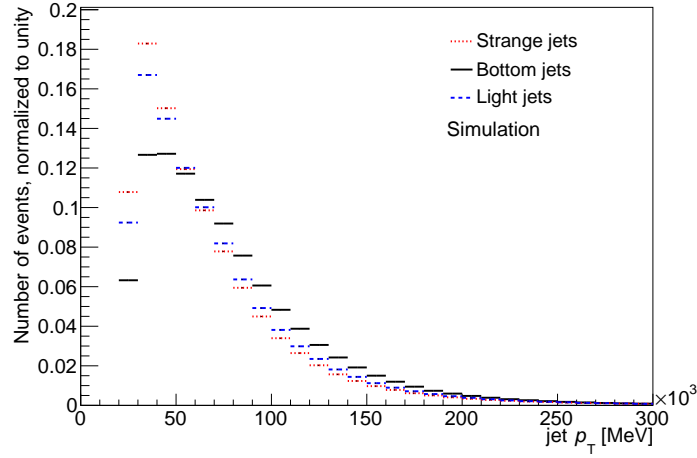

Figure 4.1: Jet-p_T distribution for bottom, strange and light jets. Error bars are included but too small to be visible.

4.2 Jet variables

Figure 4.2 and Figure 4.3 show the jet mass m_J and the jet width W_J, for the jet p_T range of 25 GeV to 50 GeV, 75 GeV to 100 GeV and higher than 150 GeV. m_J is calculated from the sum over the energy of all calorimeter clusters in the jet E_i and their three-momenta $\vec{p}_i$ following the formula

$$m_J = \sqrt{\left(\sum_i E_i\right)^2 - \left(\sum_i \vec{p}_i\right)^2}$$

and W_{J} is defined as

$$W_{\mathrm{J}} = \frac{\sum_{\mathrm{i}} \Delta R_{\mathrm{i}} \, p_{\mathrm{T,i}}}{\sum_{\mathrm{i}} p_{\mathrm{T,i}}},$$

where i runs over all calorimeter clusters in the jet and ΔR_{i} is the distance between the ith calorimeter clusters and the jet axis.

The jet mass as well as the jet width is strongly correlated with the jet p_{T} and on average, jet masses in Figures 4.2a, 4.2c and 4.2e get bigger with higher jet p_{T}. The jet width shows similar correlation effects: the more p_{T} a jet has, the more collimated and smaller in width it is (see Figures 4.3a, 4.3c and 4.3e).

The jet-mass distributions' maxima are located around $6\,\mathrm{GeV}$ for low jet p_{T}, around $11\,\mathrm{GeV}$ for medium jet p_{T} and around $15\,\mathrm{GeV}$ for high jet p_{T}. At low and medium values of jet p_{T}, bottom jets generally have a higher jet mass than strange and light jets. The strange and light jets' distributions are very similar for these jet-p_{T} intervals. This is expected because the quark, the bottom jet originates from is more massive than the others, with the mass difference between strange and light jets not being as significant. The corresponding ROC curves show good discrimination strength between bottom and strange jets.

The maxima of the width distributions are located around 0.13 for low jet p_{T}, around 0.9 for medium jet p_{T} and around 0.7 for high jet p_{T}. In Figures 4.3a and 4.3b, bottom jets tend to have a larger width than strange and light jets, which have similar distributions again. This is expected since the more massive bottom quarks are on average less boosted than strange and light quarks, resulting in their jet cone being less collimated. The corresponding ROC curves (Figures 4.3b and 4.3b) show good discrimination strength.

At high jet p_{T}, the distributions of jet mass and jet width both change in shape. Strange jets have very high jet masses of $30\,\mathrm{GeV}$ more often than the other two jet types with bottom jets having the least highly-massive jets. Strange jets also have a larger number of jets with a width between approximately 0.13 and 0.22 than bottom jets. This unexpected behavior might be due to unwanted correlation effects inside the jet-p_{T} intervals. Additionally, the discrimination based on the bottom jet's higher mass, as described above, becomes less and less relevant at higher jet-p_{T} intervals. As a result, the discrimination strength between bottom and strange jets shown in the high-jet-p_{T} ROC curves for jet mass and jet width is not as meaningful as for lower jet p_{T}.

(a)

(b)
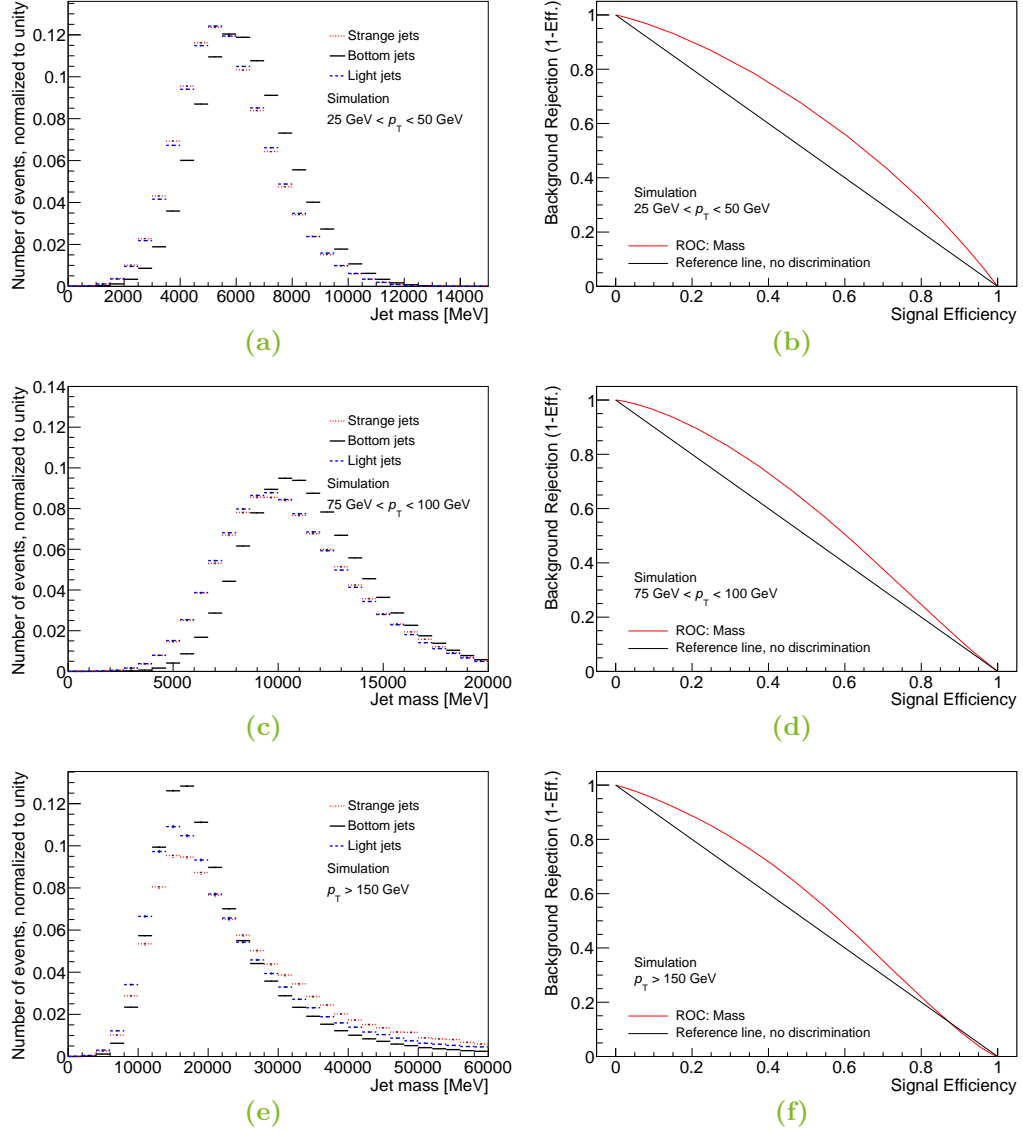

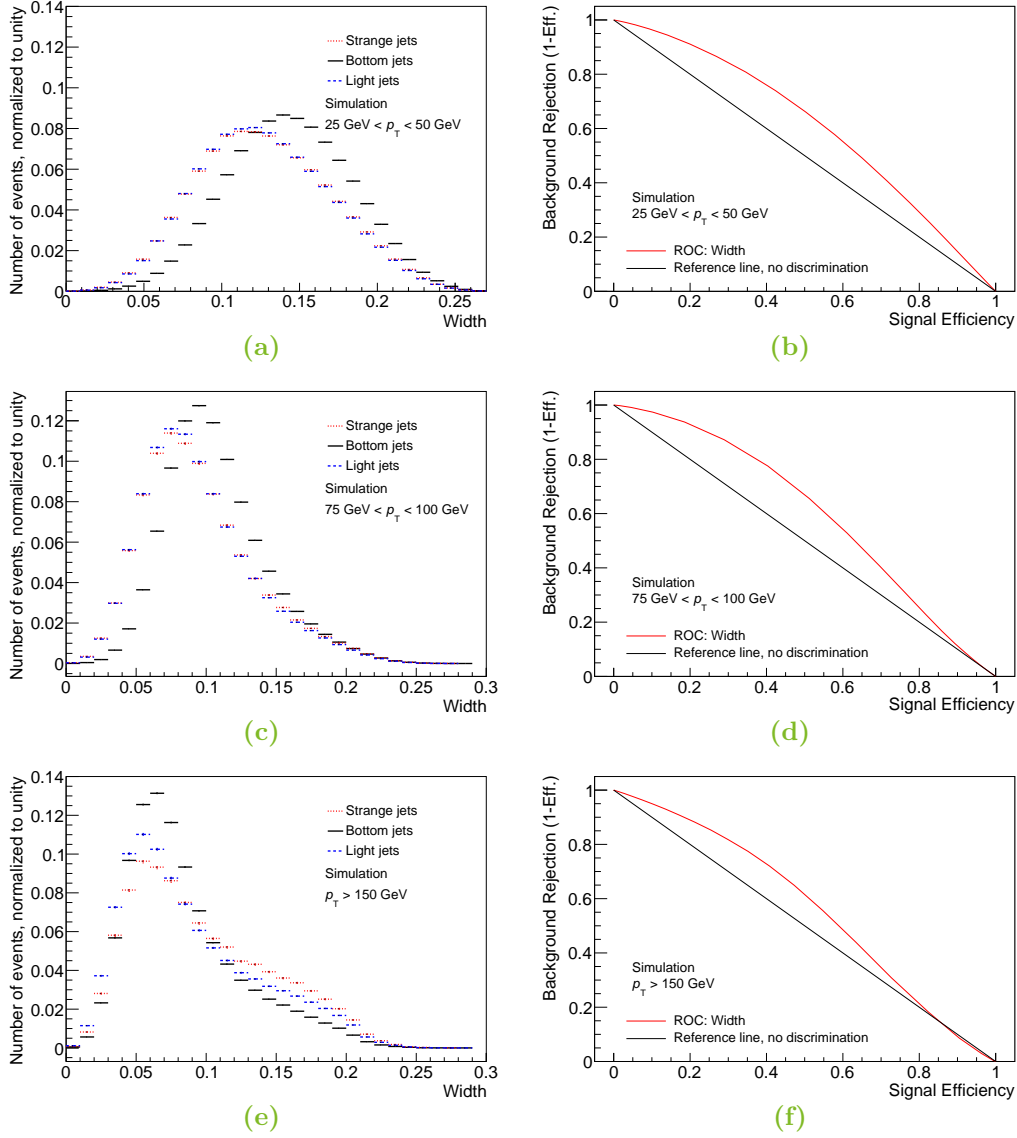
(c)

(d)

(e)

(f)

Figure 4.2: Histograms of the jet mass of light, strange and bottom jets (left) and corresponding ROC curves for strange and bottom quarks (right) for low jet p_T (top), medium jet p_T (middle) and high jet p_T (bottom). Error bars are included but too small to be visible.

(a)

(b)

(c)

(d)

(e)

(f)

Figure 4.3: Histograms of the jet width of light, strange and bottom jets (left) and corresponding ROC curves for strange and bottom quarks (right) for low jet p_{T} (top), medium jet p_{T} (middle) and high jet p_{T} (bottom). Error bars are included but too small to be visible.

4.3 Jet substructure variables

This chapter discusses the N-subjettines τ_N [28, 29]. N-subjettines is defined as

$$\tau_N = \frac{\sum_i p_{T,i} \min(\Delta R_{ik})}{\sum_i p_{T,i} R_0}$$

with N axes $\hat{n}_k$, i runs over all calorimeter clusters in the jet, R_0 is the jet radius parameter and ΔR_{ik} the distance of calorimeter cluster i to the axis $\hat{n}_k$ in the ϕ-η plane. Therefore, the smaller a jet's τ_N, the better it is characterized as having N (or fewer) subjets, whereas larger values of τ_N indicate more than N subjets.

As can be seen in Figure 4.4, 4.5 and 4.6, bottom jets tend to have larger values of τ_N and therefore more subjets than strange jets. This is expected, since the bottom quarks form B mesons, which decay, possibly creating subjets. Bound states formed by strange and light quarks are much less massive and decay more collimated. Strange and light jets show similar τ_N distributions at low and medium jet p_T. The corresponding ROC curves show good discrimination strength between bottom and strange jets, comparable in strength to the jet-variable ROC curves in Section 4.2. However, with higher jet p_T, the whole jet becomes more and more collimated. Therefore, in this case the jet substructure is not as distinguishable as for lower jet p_T and the jet substructure variables lose in significance and discrimination strength between strange and bottom jets.
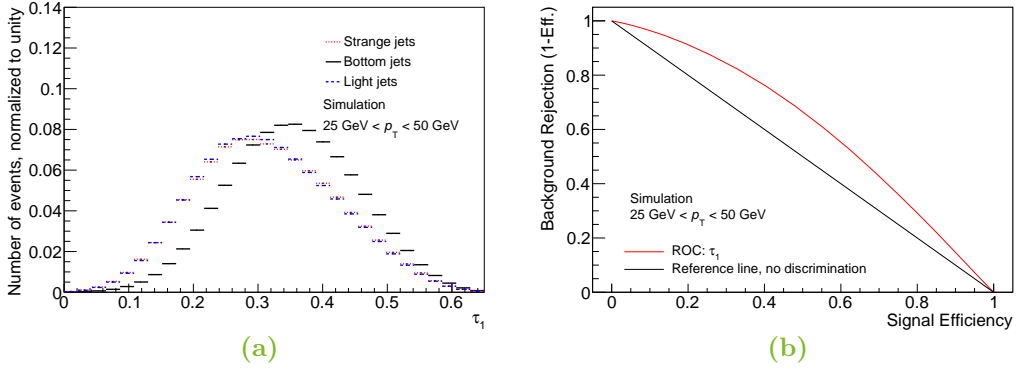

Figure 4.4: Histograms of light, strange and bottom jets' τ_1 (left) and corresponding ROC curves for strange and bottom quarks (right) for low jet p_T.
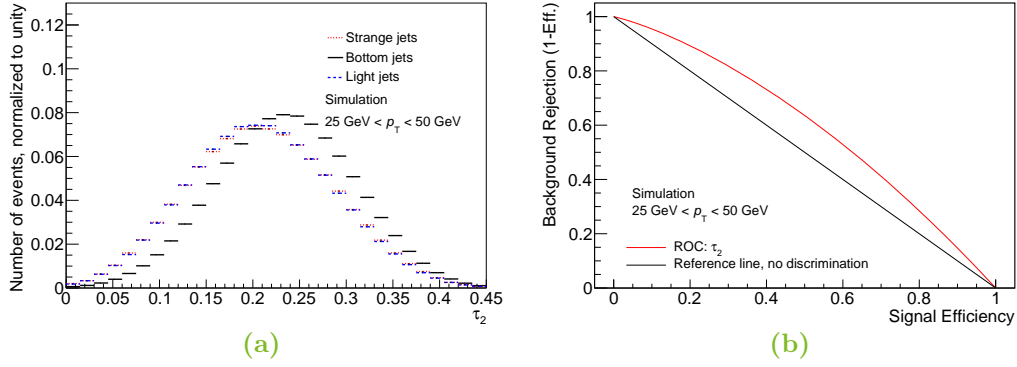
(a)

(b)

Figure 4.5: Histograms of light, strange and bottom jets' τ_2 (left) and corresponding ROC curves for strange and bottom quarks (right) for low jet p_T. Error bars are included but too small to be visible.
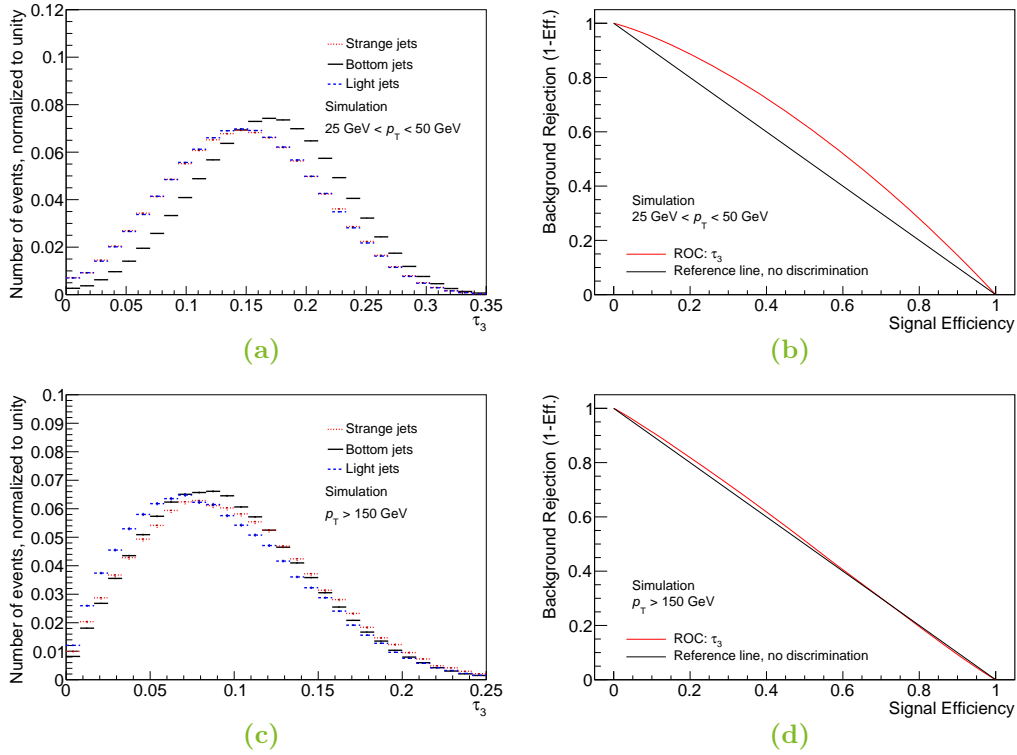

(a)

(b)

(c)

(d)

Figure 4.6: Histograms of light, strange and bottom jets' τ_3 (left) and corresponding ROC curves for strange and bottom quarks (right) for low jet p_T (top) and high jet p_T (bottom). Error bars are included but too small to be visible.

4.4 Calorimeter cluster variables

Variables discussed in this section are the longitudinal distance between the leading cluster's center of gravity and the edge of the calorimeter λ_{clus}, the leading cluster's second moment of longitudinal spread $\langle \lambda^2 \rangle$, the leading cluster's second moment of radial spread $\langle R^2 \rangle$ and the leading cluster's $\langle p_{\text{T}} \rangle$ [30]. The leading cluster is the calorimeter cluster with the highest energy. The leading cluster moments $\langle \nu^{\text{n}} \rangle$ are defined as

$$\langle \nu^{\text{n}} \rangle = \frac{1}{E_{\text{norm}}} \sum_{i, E_{\text{i}} > 0} \omega_{\text{i}}^{\text{geo}} E_{\text{i}} \, \nu_{\text{i}}^{\text{n}},$$

where i runs over all cluster cells with deposited energy $E_{\text{i}} > 0$, ν_{i} is either the distance of the cell to the clusters' center of gravity in shower axis direction λ_{i} with n = 2, in radial direction R with n = 2 or the cell's p_{T} with n = 1. $\omega_{\text{i}}^{\text{geo}}$ is a weight factor that is used, if a cell is shared by two clusters. This is the case, if a cluster contains two or more energy maxima, which happens if two or more particles deposit energy in the calorimeter in close proximity, and the cluster is therefore split. If a cell is neighboring more than one maximum, the cell is shared by the two clusters with the highest energy after splitting. E_{norm} is defined as:

$$E_{\text{norm}} = \sum_{i, E_{\text{i}} > 0} \omega_{\text{i}}^{\text{geo}} E_{\text{i}},$$

with the energie of the i$^{\text{th}}$ cell E_{i}.

Figures 4.7 to 4.10 show histograms as well as ROC curves for low, medium and high jet p_{T} intervals. As expected, the leading cluster $\langle p_{\text{T}} \rangle$ of jets gets higher with higher jet p_{T}. The distributions have maxima at approximately $9\,\text{GeV}$, $24\,\text{GeV}$ and $70\,\text{GeV}$ for the jet p_{T} between $25\,\text{GeV}$ and $50\,\text{GeV}$, $75\,\text{GeV}$ and $100\,\text{GeV}$ or above $150\,\text{GeV}$, respectively. Bottom jets tend to have smaller values of leading cluster $\langle p_{\text{T}} \rangle$, which is expected because bottom jets tend to have more jet substructure (Chapter 4.3). Therefore, they possibly have more clusters, each containing a smaller fraction of the jet's overall p_{T}. The ROC curves show a discrimination strength between bottom and strange jets, comparable to the strength of the jet substructure variables studied in the Chapters 4.2 and 4.3. However, the discrimination strength does not decline with higher jet p_{T}.

The variables leading cluster $\langle \lambda^2 \rangle$ and leading cluster $\langle R^2 \rangle$ show similar distributions across different jet-p_{T} intervals. Their maxima correspond to very roughly about 15 cm longitudinal cluster length and 5 cm cluster width. The bottom jets tend to have leading clusters smaller in size because their substructure (Chapter 4.3) effectuates more clusters, each individually being smaller on average. The variables have discrimination strength between bottom and strange jets across the whole jet-p_{T} spectrum as well, however, less than the leading cluster $\langle p_{\mathrm{T}} \rangle$.

Figure 4.10 shows the leading-cluster-λ_{clus} distributions and ROC curves. For every jet-p_{T} interval, two maxima can be seen, one at approximately 200 mm, in the EM calorimeter, and one at approximately 1250 mm, in the hadronic calorimeter. Bottom jets form B mesons that decay via the weak interaction into a quark and a W boson. The W boson might decay leptonically with the lepton depositing energy in the EM calorimeter. Therefore, it is more likely, that the bottom jet's leading cluster is located in the EM calorimeter rather than in the hadronic calorimeter, which leads on average to smaller leading-cluster-λ_{clus}. The discrimination strength between bottom and strange jets is somewhere between the leading cluster $\langle p_{\mathrm{T}} \rangle$ and leading cluster $\langle \lambda^2 \rangle$ discrimination strength. All calorimeter cluster variables show a certain discrimination strength across all p_{T} intervals.

(a)

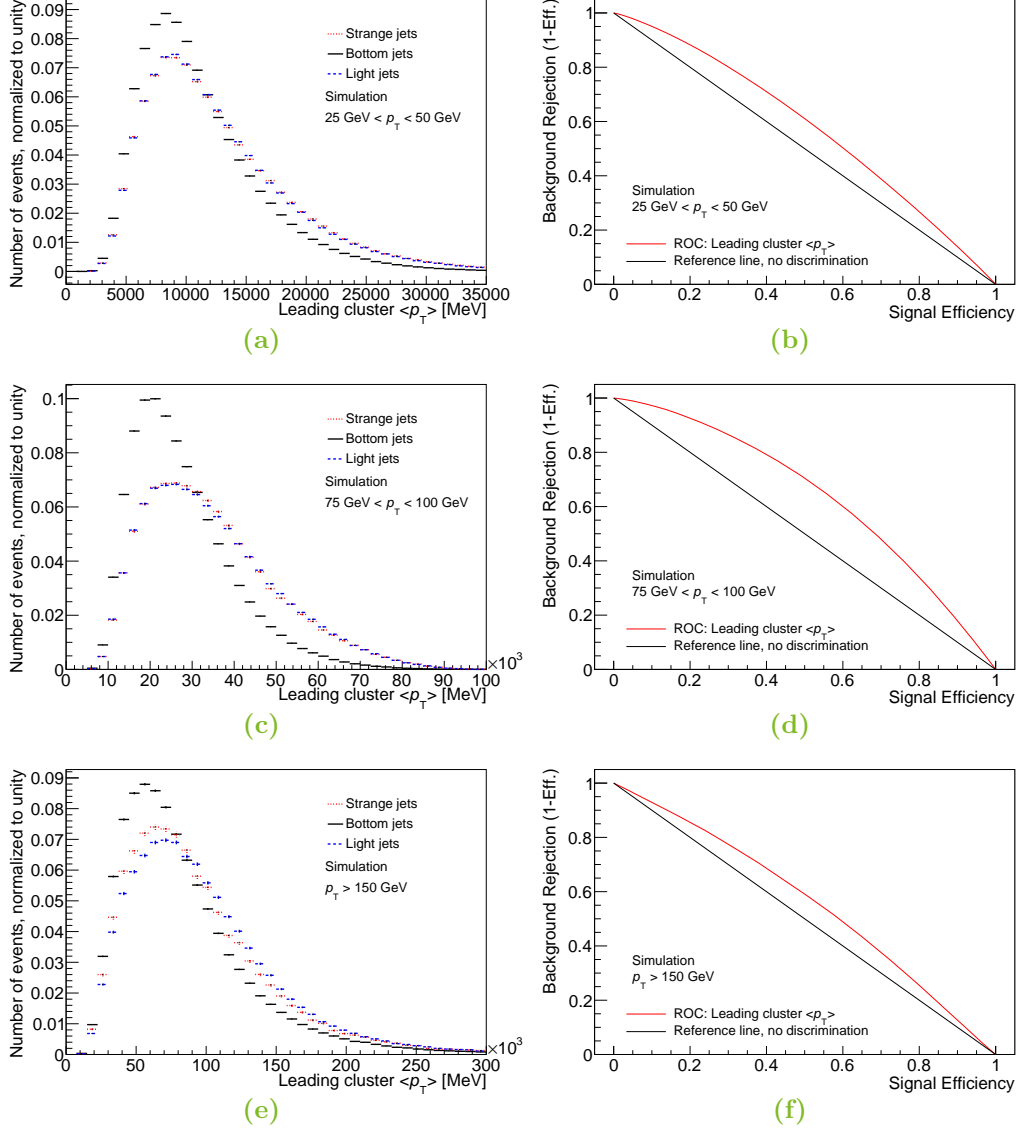**(b)**

(c)

(d)

(e)

(f)

Figure 4.7: Histograms of light, strange and bottom jets' leading cluster $\langle p_T \rangle$ (left) and corresponding ROC curves for strange and bottom quarks (right) for low jet p_T (top), medium jet p_T (middle) and high jet p_T (bottom). Error bars are included but too small to be visible.
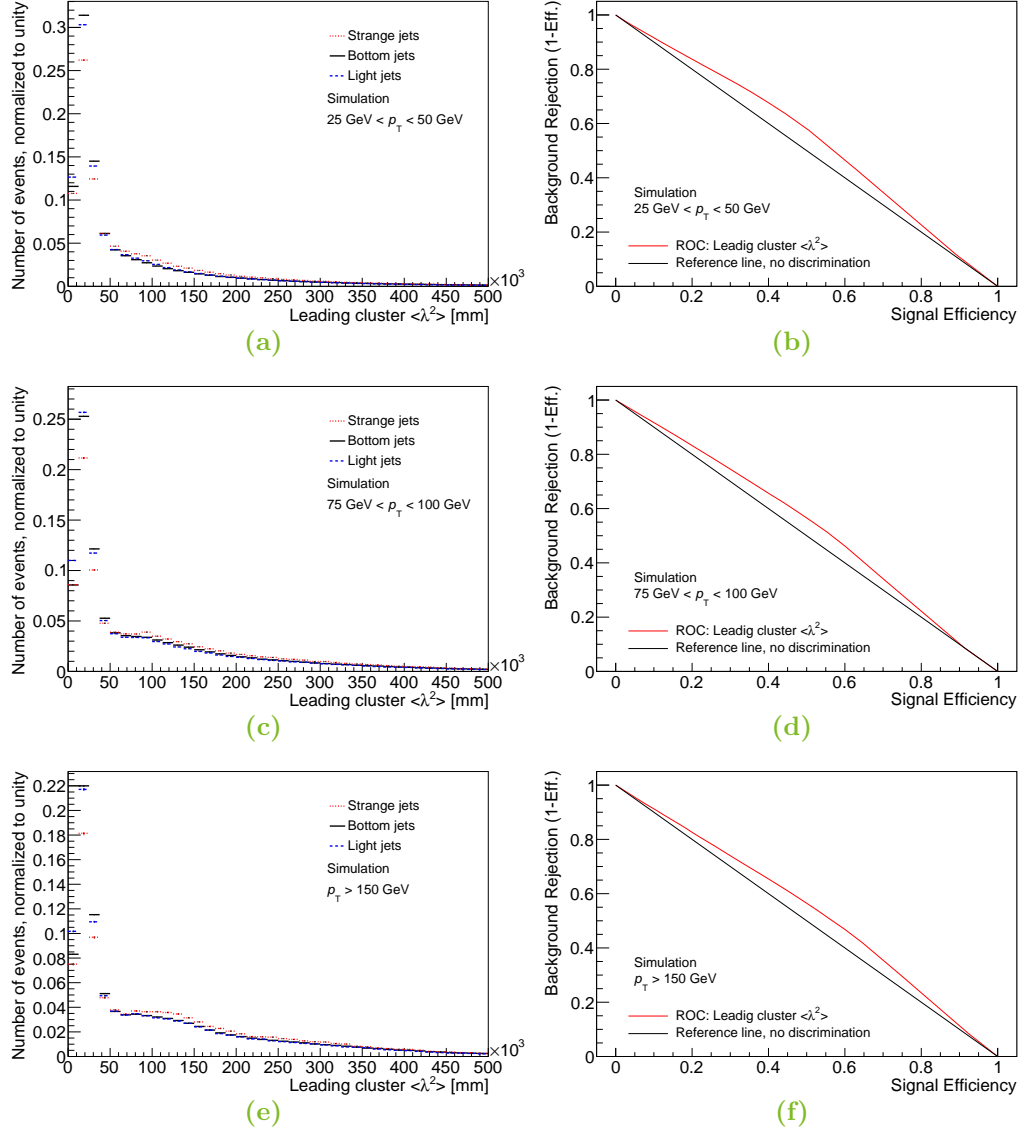
Figure 4.8: Histograms of light, strange and bottom jets' leading cluster $\langle \lambda^2 \rangle$ (left) and corresponding ROC curves for strange and bottom quarks (right) for low jet p_T (top), medium jet p_T (middle) and high jet p_T (bottom). Error bars are included but too small to be visible.

(a)

(b)
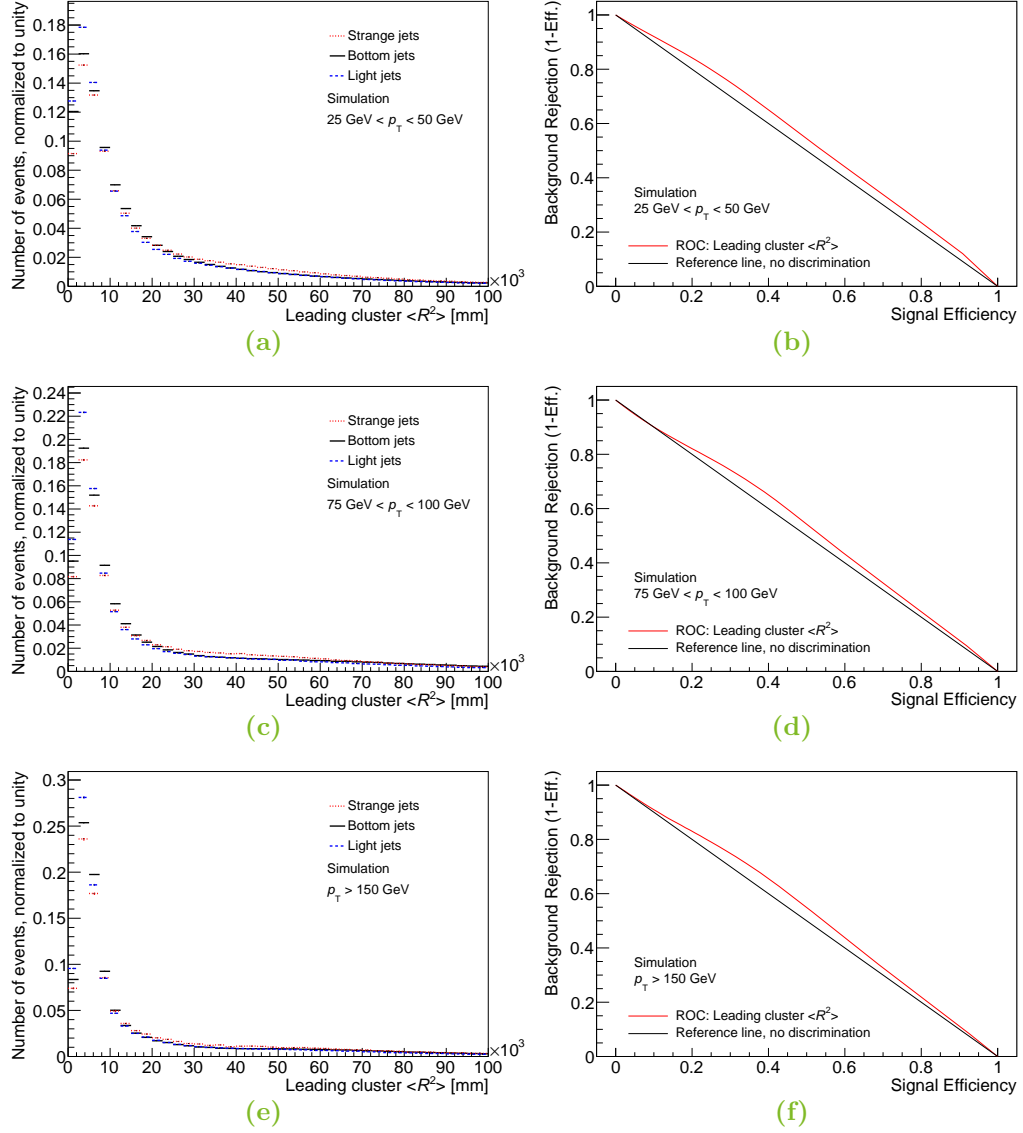

(c)

(d)

(e)

(f)

Figure 4.9: Histograms of light, strange and bottom jets' leading cluster $\langle R^2 \rangle$ (left) and corresponding ROC curves for strange and bottom quarks (right) for low jet p_T (top), medium jet p_T (middle) and high jet p_T (bottom). Error bars are included but too small to be visible.

(a)

(b)

(c)

(d)

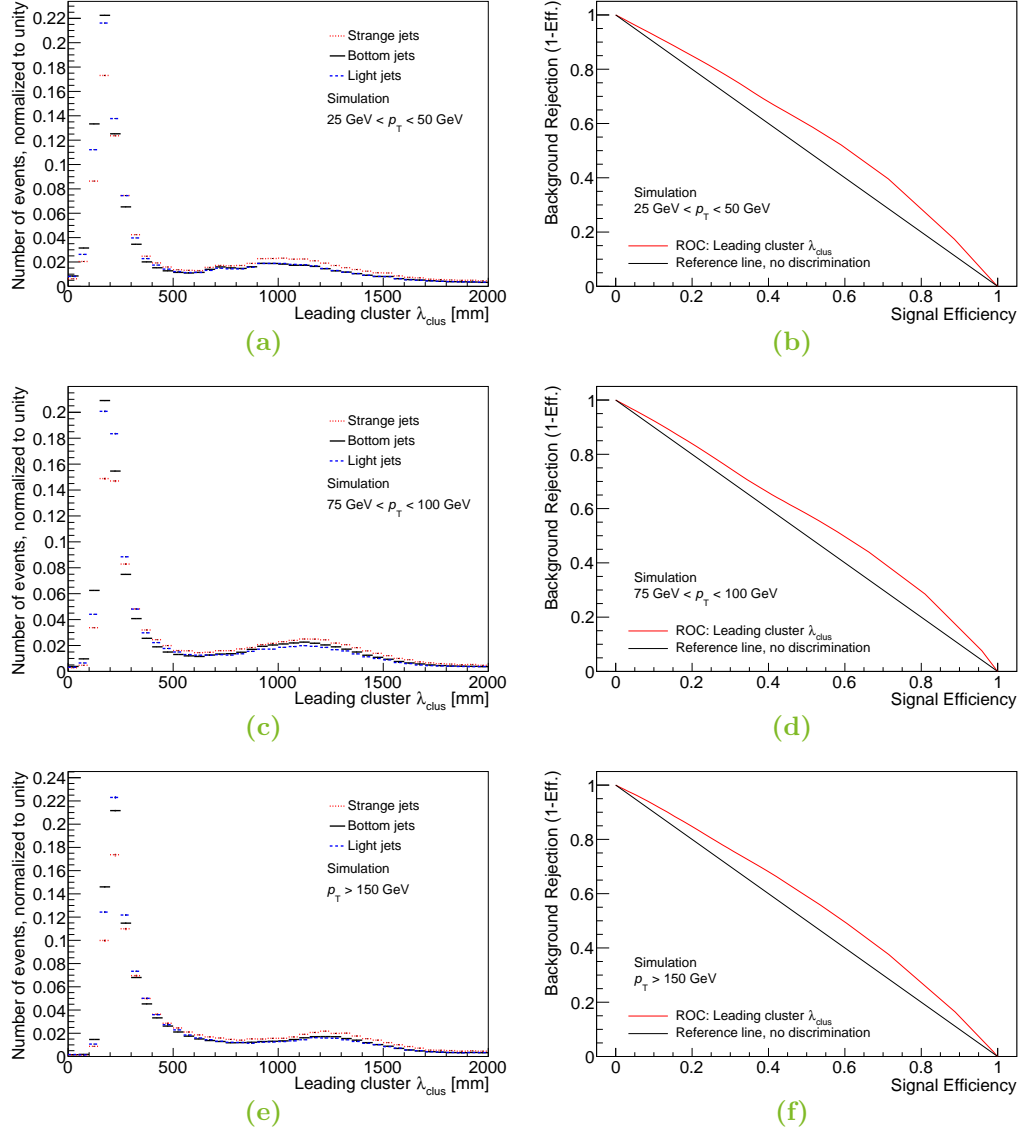**(e)**

(f)

Figure 4.10: Histograms of light, strange and bottom jets' leading cluster λ_{clus} (left) and corresponding ROC curves for strange and bootom quarks (right) for low jet p_{T} (top), medium jet p_{T} (middle) and high jet p_{T} (bottom). Error bars are included but too small to be visible.

4.5 b-tagging variables

As shown exemplarily for the MV2c20 discriminant in Figure 4.11, all multivariate *b*-tagging algorithms described in chapter 3.2.4 show good discrimination strength between bottom and strange jets with almost no difference in discrimination strength between MV2c20, MV2c10 or MV2c20. This is the case over the whole jet-p_T spectrum.

The biggest contribution to the discrimination strength of the multivariate algorithms comes from the JetFitter variables and the SV variables. Figure 4.12 and 4.13 show exemplarily the SV1 3D significance and the Jetfitter P_b distributions.

The SV1 3D significance is defined as the distance between the primary and the secondary vertex divided by the measurement's uncertainty. As one would expect, it is much larger for bottom jets, because if a secondary vertex for a strange jet is falsely reconstructed, its position is random and its measurement uncertainty tends to be large. In contrast, a secondary vertex from a *B* meson decay is mostly in front of the primary vertex in jet direction, if measured correctly. In Figure 4.12a, approximately 95 % of strange jets and only about 30 % of bottom jets are in the first bin, which leads to strong discrimination between those jets (Figure 4.12b). The JetFitter P_b value is the probability calculated by the JetFitter algorithm that a jet is originating from a bottom quark.

Both the JetFitter P_b value and the SV1 3D significance are similarly discriminating between bottom and strange jets for all jet-p_T intervals. Light and strange jets show similar distributions. This is expected because neither light nor strange jets qualify more or less for the criteria, the *b*-tagging variables are based on.
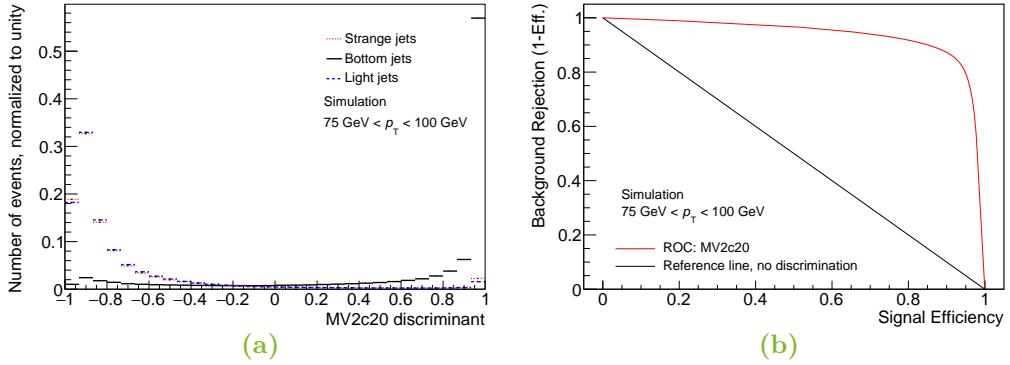

Figure 4.11: Histogram of MV2c20 output for light, strange and bottom jets (left) and corresponding ROC curve for strange and bottom quarks (right), for medium jet p_T. Error bars are included but too small to be visible.
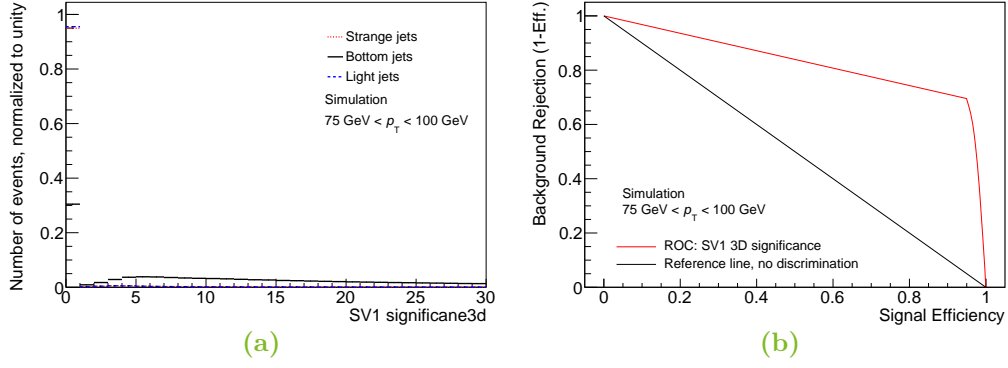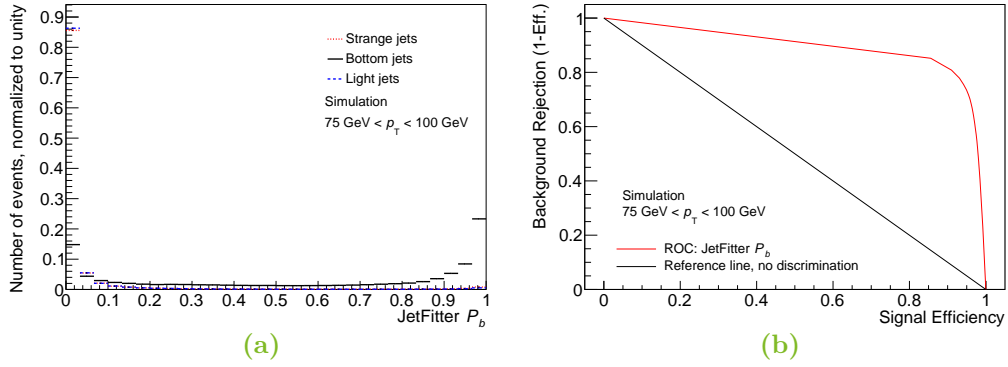
Figure 4.12: Histogram of SV1 3D significance for light, strange and bottom jets (left) and corresponding ROC curve for strange and bottom quarks (right), for medium jet p_{T}. Error bars are included but too small to be visible.

Figure 4.13: Histogram of JetFitter P_b for light, strange and bottom jets (left) and corresponding ROC curve for strange and bottom quarks (right), for medium jet p_{T}. Error bars are included but too small to be visible.

5 Conclusions

Discrimination between bottom and strange jets is an important factor in order to measure strange quarks over bottom-quark background, for example for measuring the direct top-to-strange-quark decay. To evaluate possibilities to discriminate between bottom and strange jets, four types of variables were tested with respect to their discrimination strength: b-tagging variables, jet variables, jet substructure variables and calorimeter cluster variables. This was done with MC simulations. To reduce correlation effects between these variables and the jet p_T, the comparison was done in intervals of jet p_T.

As expected, the studied b-tagging variables show strong discrimination between bottom and strange jets. While the tested jet variables perform quite well at low jet p_T, their discrimination strength declines with higher jet p_T. The jet-substructure variables studied show good discrimination strength at low jet p_T while they are less discriminating at higher values of jet p_T as well. The calorimeter cluster variables have a slightly smaller discrimination strength, which is similarly good for all jet-p_T intervals.

It was shown that differences between bottom and strange jets exist, which could be used to develop a discrimination method between them with a neural network or a boosted decision tree based on the variables tested in this thesis. It is especially noteworthy, that not only b-tagging variables show discrimination strength, which would not be sufficient in order to measure the top-to-strange-quark decay described in Chapter 2.3. Instead, if correlation effects are accounted for, the other variables studied can be used to further cut out bottom-quark background.

Bibliography

[1] L. Evans and P. Bryant, "LHC Machine," *JINST* 3 (2008) S08001.

[2] K. A. Olive et al., "Review of Particle Physics," *Chin. Phys.* C 38 (2014) 090001.

[3] CDF Collaboration, "Observation of top quark production in $\bar{p}p$ collisions," *Phys. Rev. Lett.* 74 (1995) 2626.

[4] D0 Collaboration, "Search for high mass top quark production in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV," *Phys. Rev. Lett.* 74 (1995) 2422.

[5] ATLAS, CDF, CMS and D0 Collaborations, "First combination of Tevatron and LHC measurements of the top-quark mass," ATLAS-CONF-2014-008 (2014), https://cds.cern.ch/record/1669819.

[6] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider," *JINST* 3 (2008) S08003.

[7] ATLAS Collaboration, "ATLAS Insertable B-Layer Technical Design Report", ATLAS-TDR-19 (2010), https://cds.cern.ch/record/1291633.

[8] ATLAS Collaboration, "ATLAS Insertable B-Layer Technical Design Report Addendum," ATLAS-TDR-19-ADD-1 (2012), https://cds.cern.ch/record/1451888.

[9] M. Cacciari, G. P. Salam, G. Soyez, "The anti-k_t jet clustering algorithm," *JHEP* 04 (2008) 063.

[10] G. P. Salam and G. Soyez, "A Practical Seedless Infrared-Safe Cone jet algorithm," *JHEP* 05 (2007) 086.

[11] ATLAS Collaboration, "Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data," ATLAS-CONF-2011-102 (2011), http://cds.cern.ch/record/1369219.

[12] S. Frixione, P. Nason, and G. Ridolfi, "A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction," *JHEP* 09 (2007) 126.

[13] P. Nason, "A New method for combining NLO QCD with shower Monte Carlo algorithms," *JHEP* 11 (2004) 040.

[14] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method," *JHEP* 11 (2007) 070.

[15] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX," *JHEP* 06 (2010) 043.

[16] H.-L. Lai et al., "New parton distributions for collider physics," *Phys. Rev.* D 82 (2010) 074024.

[17] T. Sjöstrand, S. Mrenna, and P. Z. Skands, "Pythia 6.4 physics and manual," *JHEP* 05 (2006) 026.

[18] J. Pumplin et al., "New generation of parton distributions with uncertainties from global QCD analysis," *JHEP* 07 (2002) 012.

[19] P. Z. Skands, "Tuning Monte Carlo Generators: The Perugia Tunes," *Phys. Rev.* D 82 (2010) 074018.

[20] T. Sjöstrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1," *Comput. Phys. Commun.* 178 (2008) 852.

[21] A. Sherstnev and R. S. Thorne, "Parton Distributions for LO Generators," *Eur. Phys. J.* C 55 (2008) 553.

[22] ATLAS Collaboration, "Further ATLAS tunes of PYTHIA6 and Pythia 8," ATL-PHYS-PUB-2011-014 (2011), https://cds.cern.ch/record/1400677.

[23] ATLAS Collaboration, "The ATLAS Simulation Infrastructure", *Eur. Phys. J.* C 70 (2010) 823.

[24] S. Agostinelli et al., "Geant4—a simulation toolkit," *Nucl. Instrum. Meth. A* 506 (2003) 250.

[25] ATLAS collaboration, "Electron identification measurements in ATLAS using $\sqrt{s} = 13$ TeV data with 50 ns bunch spacing," ATL-PHYS-PUB-2015-041 (2015), http://cds.cern.ch/record/2048202.

[26] ATLAS Collaboration, "Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data," ATLAS-CONF-2016-024 (2016), https://cds.cern.ch/record/2157687.

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

[28] J. Thaler and K. Van Tilburg, "Identifying boosted objects with n-subjettiness," *JHEP* 03 (2011) 015.

[29] J. Thaler, K. Van Tilburg, "Maximizing Boosted Top Identification by Minimizing N-subjettiness", *JHEP* 02 (2012) 093.

[30] ATLAS Collaboration, "Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1," CERN-PH-EP-2015-304 (2016), https://arxiv.org/abs/1603.02934.

Eidesstattliche Versicherung

Ich versichere hiermit an Eides statt, dass ich die vorliegende Abschlussarbeit mit dem Titel "First studies on the distinction of strange and bottom jets with the ATLAS detector in Run 2" selbstständig und ohne unzulässige fremde Hilfe erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt, sowie wörtliche und sinngemäße Zitate kenntlich gemacht. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

_____   _____
Ort, Datum   Unterschrift

Belehrung

Wer vorsätzlich gegen eine die Täuschung über Prüfungsleistungen betreffende Regelung einer Hochschulprüfungsordnung verstößt, handelt ordnungswidrig. Die Ordnungswidrigkeit kann mit einer Geldbuße von bis zu 50 000 € geahndet werden. Zuständige Verwaltungsbehörde für die Verfolgung und Ahndung von Ordnungswidrigkeiten ist der Kanzler/die Kanzlerin der Technischen Universität Dortmund. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann der Prüfling zudem exmatrikuliert werden (§ 63 Abs. 5 Hochschulgesetz –HG–).

Die Abgabe einer falschen Versicherung an Eides statt wird mit Freiheitsstrafe bis zu 3 Jahren oder mit Geldstrafe bestraft.

Die Technische Universität Dortmund wird ggf. elektronische Vergleichswerkzeuge (wie z. B. die Software "turnitin") zur Überprüfung von Ordnungswidrigkeiten in Prüfungsverfahren nutzen.

Die oben stehende Belehrung habe ich zur Kenntnis genommen.

_____   _____
Ort, Datum   Unterschrift