Development and Calibration of an s-Tagging Algorithm and Its Application to Constrain the CKM Matrix Elements $|V_{ts}|$ and $|V_{td}|$ in Top-Quark Decays Using ATLAS Run-2 Data

Dissertation
zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften

vorgelegt von

M.Sc. Sonja Verena Zeißner geb. Bartkowski

geboren in Gelsenkirchen

Fakultät Physik

Technische Universität Dortmund

Dortmund, 2021

Abstract

In this thesis, the development and calibration of an algorithm used to identify jets from strange quarks as well as a measurement constraining the CKM matrix elements $|V_{ts}|$ and $|V_{td}|$ in top-quark decays are presented. The thesis considers data from proton-proton collisions at the Large Hadron Collider at a center-of-mass energy of $13\,\mathrm{TeV}$ recorded by the ATLAS Experiment during Run 2. First, the maximally achievable separation between jets from strange quarks and jets from down quarks at hadron colliders given different idealized detector designs is studied using recurrent neural networks containing Long Short-Term Memory layers. Afterwards, an algorithm to select jets from strange quarks for the application at the ATLAS Experiment is developed using deep neural networks. Its efficiency for these jets from strange quarks and mis-tag rates for jets of other flavors is determined in semileptonic decays of top-antitop pairs selected from data. The algorithm to identify jets from strange quarks is then applied in events containing decays of top-antitop pairs with an electron and a muon of opposite-sign electric charge in the final state in order to study its potential to constrain the CKM matrix elements $|V_{ts}|$ and $|V_{td}|$ in the two-dimensional plane spanned by them. In this study, limits of $|V_{ts}|^2 + |V_{td}|^2 < 0.06$, $|V_{ts}| < 0.21$, and $|V_{td}| < 0.24$ are derived at 95% confidence level assuming unitarity of the CKM matrix.

Kurzfassung

Diese Dissertation befasst sich mit der Entwicklung und Kalibrierung eines Algorithmus zur Identifikation von Jets aus strange-Quarks sowie einer Messung zur Bestimmung von oberen Schranken auf die CKM-Matrixelemente $|V_{ts}|$ und $|V_{td}|$ in top-Quark-Zerfällen. In dieser Arbeit werden Daten aus Proton-Proton-Kollisionen am Large Hadron Collider, die bei einer Schwerpunktsenergie von $13\,\mathrm{TeV}$ am ATLAS-Experiment während des Run-2 aufgenommen wurden, verwendet. Zuerst wird die maximal mögliche Trennung von Jets aus strange-Quarks und Jets aus down-Quarks an Hadronenkollidern unter der Annahme einer Nutzung von unterschiedlichen, idealisierten Detektordesigns untersucht. Hierfür werden Recurrent Neural Networks verwendet, deren Hauptkomponente Long Short-Term Memory Layers sind. Anschließend wird ein Algorithmus zur Selektion von Jets aus strange-Quarks zur Anwendung am ATLAS-Experiment entwickelt, der Deep Neural Networks verwendet. Die Identifikationseffizienz dieses Algorithmus in Bezug auf Jets aus strange-Quarks und die Wahrscheinlichkeit, Jets eines anderen Ursprungs fehl zu identifizieren, werden in semileponischen Zerfällen von top-antitop-Paaren, die aus dem aufgenommenen Datensatz selektiert wurden, bestimmt. Schlussendlich wird der Algorithmus zur Identifikation von Jets aus strange-Quarks in Ereignissen angewendet, die dileponischen Zerfällen von top-antitop-Paaren mit einem Myon und einem Elektron im Endzustand beinhalten. Hierbei wird sein Potenzial in der Bestimmung von Schranken auf die CKM-Matrixelemente $|V_{ts}|$ und $|V_{td}|$ im zweidimensionalen Raum untersucht. Daraus folgen Schranken von $|V_{ts}|^2 + |V_{td}|^2 < 0.06$, $|V_{ts}| < 0.21$ und $|V_{td}| < 0.24$ mit einem Konfidenzintervall von 95% unter der Annahme von CKM-Matrix-Unitarität.

To Amalia

Contents

CHAPTER 1

Introduction

In science, the field of physics seeks to understand and explain the fundamental rules that govern the universe. In *particle physics*, the smallest building blocks are studied: Elementary particles which – to our knowledge – are indivisible as well as their interactions with each other. The one description of both used primarily today is called the Standard Model of Particle Physics (SM). It was developed throughout several decades in a joint effort of theorists and experimentalists and allows making good predictions of particle-level interactions.

The further studies venture to smaller scales, the higher are the energies necessary to resolve them. Today, the highest man-made energies per particle are reached by accelerating hadrons in the Large Hadron Collider (LHC) at the CERN research facility in Switzerland. In their collisions, interaction energies of up to several TeV are reached, allowing for the search of *new physics* as well as the precise measurement of some parameters of the SM.

One of the particles studied at the LHC is the top quark. It is the heaviest known elementary particle and was discovered in 1995 at the Tevatron [1, 2]. As described by the SM, at leading order, it can only decay into a down, strange, or bottom quark by emitting a W boson. The probabilities of the decays are proportional to the square of the respective elements of the Cabibbo-Kobayashi-Maskawa (CKM) matrix: V_{td}, V_{ts}, and V_{tb}. Due to the large differences in the magnitude of these matrix elements, the top predominately decays into a W boson and a bottom quark.

The branching ratio of this decay, $R_b = \mathcal{BR}(t \to W^+ b)^1$, has been measured in decays of top-antitop quark pairs [3–6]. The newest and most precise of these measurements [6] has a result of $|V_{tb}| = 1.007 \pm 0.016$ or a limit of $|V_{tb}| > 0.975$ at 95% confidence level assuming $|V_{tb}| < 1.0$.

In an alternative approach, events containing single top quarks produced via t-channel interactions with a W boson have been used to constrain the three CKM matrix elements involving a top quark. Most measurements assume $|V_{ts}|$ and $|V_{td}|$ to be negligible to determine $|V_{tb}|$ [7–23]. The most precise of these measurements and combinations quotes a value of $|V_{tb}| = 0.998 \pm 0.038$ (exp.) ± 0.016 (theo.) [17]. Due to the relation $|V_{tb}|^2 + |V_{ts}|^2 + |V_{td}|^2 = 1$, both of these measurements offer a constraint on $|V_{ts}|^2 + |V_{tb}|^2$. A recent publication by the CMS Experiment at the LHC [23] was able to derive a limit of $|V_{ts}|^2 + |V_{td}|^2 < 0.057$ at 95% confidence level by considering single top quark production in the t-channel from strange and down quarks as well as $t \to W^+ s$ and $t \to W^+ d$ decays while at the same time deriving a lower limit of $|V_{tb}| > 0.970$ at 95% confidence level.

On the other hand, the matrix elements $|V_{ts}|$ and $|V_{tb}|$ can be determined from the mass differences of the mass eigenstates of B_s^0 and $\overline{B_s^0}$ as well as B^0 and $\overline{B^0}$ respectively, which can be determined from their oscillations, cf. Reference [24] and references therein, e.g. References [25, 26]. Reference [24] e.g. quotes $|V_{td}| = (8.1 \pm 0.5) \cdot 10^{-3}$ and $|V_{ts}| = (39.4 \pm 2.3) \cdot 10^{-3}$ based on their own averages. This extrapolation, however, relies on the use of form factors from

---

[1]Whenever in this thesis, a decay $t \to W^+ q$ is mentioned, the decay of its antiparticle, $\bar{t} \to W^- \overline{q}$, is implied as well.

theoretical calculations (e.g. those in Reference [27]). A direct measurement of these matrix elements in either the production or decay of top quarks would constitute a valuable addition to the available measurements.

The direct measurement of $|V_{ts}|$ in $t\bar{t}$ decays at the ATLAS Experiment has been attempted before. In Reference [28], $t\bar{t}$ decays including a $t \to W^+ s$ decay are selected and a limit of $|V_{ts}| < 0.132$ is set at 95% confidence level, however, assuming the background from $t \to W^+ d$ decays to be negligible.

This thesis, on the other hand, aims at disentangling $t \to W^+ s$ and $t \to W^+ d$ decays for an independent extraction of $|V_{ts}|$ and $|V_{td}|$ from top-quark decays at the ATLAS Experiment. For this, an algorithm to distinguish between jets initiated by strange quarks and down quarks is developed. This so-called *s*-tagging algorithm or *s-tagger* is the first *s*-tagger at the ATLAS Experiment. While this *s*-tagger was developed independently, it is similar to the proposed *s*-taggers described in Reference [29], exploiting both information of electrically charged particles measured as tracks and energy deposits of both charged and neutral particles in calorimeters.

It is calibrated in semileptonic decays of top-antitop quark pairs with one lepton in the final state. It is then applied to jets from dileptonic decays of top-antitop quark pairs with an electron and a muon in the final state to extract a limit on $|V_{ts}|$ and $|V_{td}|$.

Furthermore, the development of the *s*-tagger for the ATLAS Experiment is complemented by an independent study on the maximal performance of *s*-taggers which is in principle achievable. This study evaluates the potential of jet characteristics detectable by different detector types such as calorimeters, trackers, or Cherenkov detectors to classify jets as strange or down jets. In this study, neural networks including recurrent Long Short-Term Memory (LSTM) layers are used, similar to what is used for the *s*-tagging algorithm proposed in Reference [30].

This thesis is structured as follows: In Chapter 2, a brief introduction into the SM is given. Chapter 3 gives an overview over the ATLAS detector at the LHC. In Chapter 4, the objects reconstructed from data collected with the ATLAS detector are described, while in Chapter 5.1, the dataset and all simulated samples used in the analysis are detailed. Chapter 6 covers *s*-tagging algorithms. Within this chapter, in Section 6.2, the studies on the maximal performance achievable in the classification of jets from strange and up/down quarks is described, while Section 6.3 details the development of the *s*-tagging algorithm to be used for at the ATLAS Experiment. The calibration of the *s*-tagging algorithm is detailed in Chapter 7. In Chapter 8, it is then applied in a study on the measurement of $|V_{td}|$ and $|V_{ts}|$. Lastly, Chapter 9 offers a summary of the thesis.

A Brief Introduction to the Standard Model of Particle Physics

The SM is a theoretical description of all known elementary particles and their interactions. Of the four fundamental interactions known in physics, it describes all except for gravity. A schematic overview is given in Figure 2.1, showing all elementary particles and their quantum numbers.
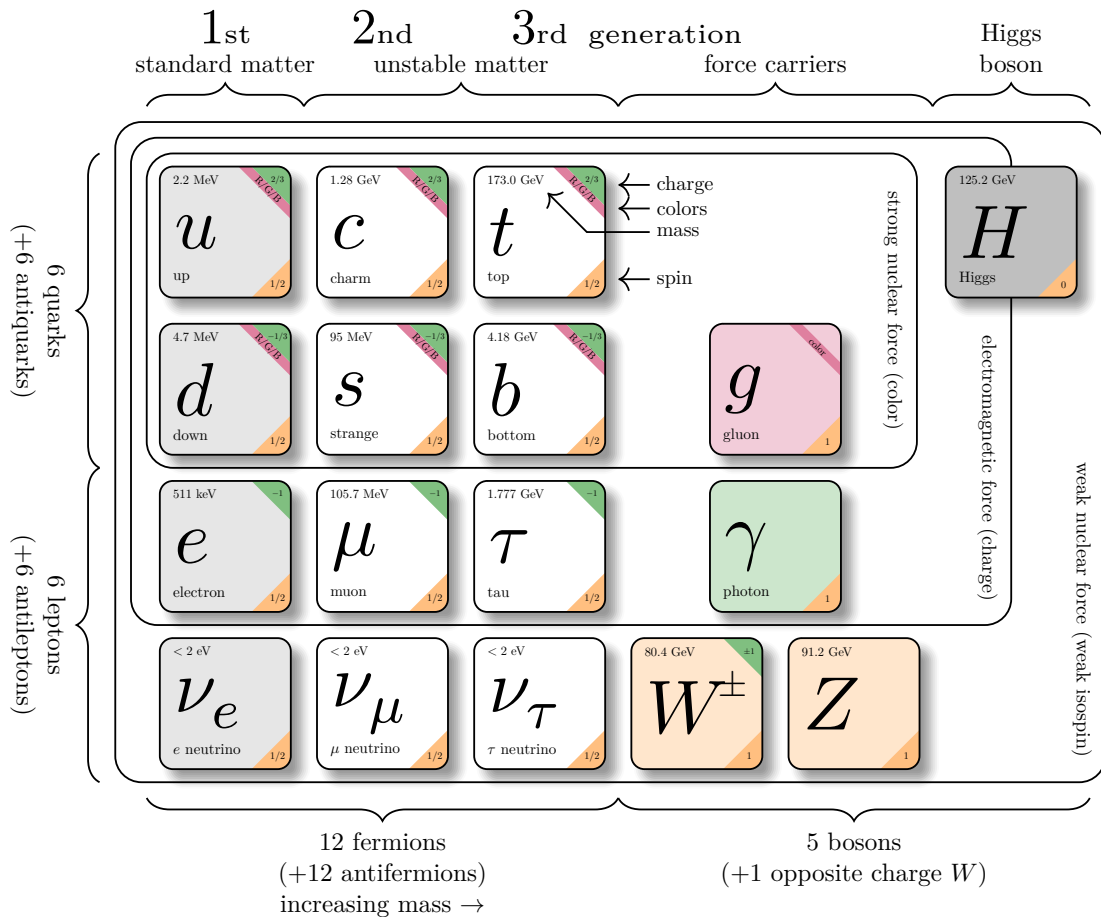
Fig. 2.1: A graphical representation of all elementary particles described by the Standard Model of Particle Physics. The graphic is based on Reference [31], while the cited mass values are taken from Reference [24].

Since the SM is a quantum field theory, all particles are associated with fields and interactions between particles are governed by underlying symmetries. There are three different types of fields: spin-$\frac{1}{2}$ fermion fields, vector boson fields, and a scalar boson field.

Both quarks and leptons are spin-$\frac{1}{2}$ fermions and described by four-component spinor fields containing both particles and antiparticles[1]. There are three electrically charged leptons (*electron (e), muon (μ), tau (τ)*), three electrically neutral neutrinos (ν_e, ν_μ, ν_τ), and six quark flavors (*up (u), charm (c), top (t), down (d), strange (s) and bottom (b)*) plus their antiparticles which carry the opposite charge and are denoted with a dash (e.g. $\bar{t}$). While up-type quarks (u, c, t) carry an electric charge of $+\frac{2}{3}e$, the down-type quarks carry an electric charge of $-\frac{1}{3}e$. There are a total of three fields associated with each quark flavor, each corresponding to a different color charge. Besides their electrical charge, the fermions can be distinguished either by their mass or their weak interaction. The former is done for quarks and charged leptons and the latter for neutrinos since, in the SM, neutrinos are assumed to be massless. However, it was observed that neutrino generations mix [32–34], which means their masses must be non-zero.

Both the strong and the electroweak interaction's force carriers are spin 1 particles, i.e. they are described using vector fields.

The theoretical part of the SM describing the strong interaction is called quantum chromodynamics (QCD). Since its underlying symmetry group is the $SU(3)_C$ where C stands for the color charge, the gluon field consists of an octet, one for each generator of the symmetry. Their structure of interaction results in a coupling between gluons and the color charged quarks as well as a coupling between gluons and gluons. As a result, the coupling strength becomes large for small energies (approximately $\Lambda_{\text{QCD}} \approx 200\,\text{GeV}$). However, for energies significantly larger than Λ_{QCD}, the strong coupling becomes small and perturbative calculations are possible. Due to the nature of the strong interaction, quarks and gluons can only be observed in color-neutral bound states called hadrons – most commonly baryons containing three quarks or mesons containing a quark and an antiquark –, a phenomenon called *confinement*.

The underlying symmetry to the electroweak interaction is the $SU(2)_L \times U(1)_Y$, where L is the left-chirality and Y is the weak hypercharge. It has four vector fields associated with it, one that couples to Y with the coupling strength g' and three that couple to the left-chiral component of other fields with the coupling strength g. Because the coupling of these fields is different for different chiralities, any mass terms of fermionic fields containing both left-chiral and right-chiral components of the field violate the $SU(2)_L$ symmetry. Additionally, the observation of gauge boson masses implies a breaking of gauge symmetry.

Therefore, a scalar Higgs field is introduced to the SM. It has a symmetric *mexican hat* potential and couples to all massive fermions via Yukawa-coupling. Additionally, the electroweak gauge bosons couple to the Higgs field according to their gauge couplings. In a process called sponaneous symmetry breaking (SSB), the Higgs field falls into a non-symmetric ground state, breaking the $SU(2)_L \times U(1)_Y$ to a $U(1)_{\text{em}}$ symmetry. This generates mass terms for both fermions and three of the four bosons.

Additionally, the four gauge fields of the electroweak interaction mix. Two of the resulting bosons are the massive W bosons with opposite-sign electric charge (W^+ and W^-) which couple to the left-chiral components of fields with a coupling strength proportional to g. One is the massive and electrically neutral Z boson, which couples to both the left-chiral components of fields and their hypercharge with a coupling strength proportional to $-g'^2 Y + g^2 T^3$, where T^3 acts on the $SU(2)_L$ spinor doublets. The fourth one is the massless and electrically neutral photon γ which couples to the left-chiral components of fields as well their hypercharge with a coupling strength proportional to $g'g(Y + T^3)$. It is useful to say that the photon couples

---

[1]For convenience, $\hbar = c = 1$.

to the electric charge, which is defined as $Q = Y + T^3$ and is the conserved quantity of the $U(1)_{\text{em}}$.

The existence of the Higgs field has been supported by the discovery of its resonant particle, the Higgs boson [35, 36].

As a result of different Yukawa-couplings for each quark, the mass eigenstates of quarks are different from their eigenstates with regard to their weak interaction. The resulting quark mixing is described by the CKM matrix. Conventionally, the CKM matrix is written as

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \tag{1}$$

Due to its unitarity, it can be expressed using four parameters, all of which are free parameters of the SM. The CKM matrix is almost diagonal, as can be seen in Equation 2, which cites values of a global fit [24]:

$$\begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} = \begin{pmatrix} 0.97446 \pm 0.00010 & 0.22452 \pm 0.00044 & 0.00365 \pm 0.00012 \\ 0.22438 \pm 0.00044 & 0.97359^{+0.00010}_{-0.00011} & 0.04214 \pm 0.00076 \\ 0.00896^{+0.00024}_{-0.00023} & 0.04133 \pm 0.00074 & 0.999105 \pm 0.000032 \end{pmatrix} \tag{2}$$

2.1. Properties of the Top Quark

The t quark was the last quark to be discovered in 1995 [1, 2]. Its most prominent property is its mass of $173.34 \pm 0.27(\text{stat}) \pm 0.71(\text{syst})$ GeV [37] (current world combination of measurements), which makes it the heaviest know elementary particle and roughly 35 times more massive than the second heaviest quark, the b quark.
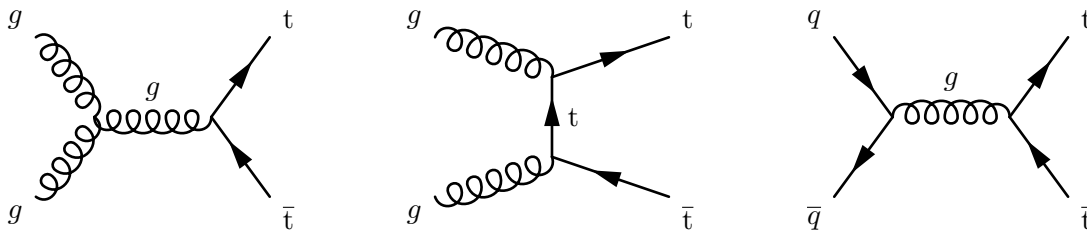
Fig. 2.2: $t\bar{t}$ production on tree level via the strong force.

At pp colliders, the two main production mechanisms for the t quark are $t\bar{t}$ pair production via the strong force and single production via a weak flavor change. Figure 2.2 shows leading order Feynman diagrams of the pair production. Due to its large mass, the t quark decays before hadronizing with a lifetime of $\tau = (3.29^{+0.90}_{-0.63}) \cdot 10^{-25}$ s [24]. At tree level, it always decays into a down-type quark by emitting a W boson. Experimentally, these decays are classed into leptonic decays and hadronic decays, referring to whether the W boson decays into either an electrically charged lepton and a neutrino or into a quark and an antiquark.

The probability for each flavor of the down-type quark the t quark decays into is given by the square of the corresponding CKM matrix element. I.e. approximately 99.8% of all t quarks decay into a b quark, while only approximately 0.16% decay into s quarks and 0.008% into a d

quark. Because of their small probabilities, the latter two decays have not been observed thus far.

CHAPTER 3

The Large Hadron Collider and the ATLAS Detector

3.1. The Large Hadron Collider

The LHC [38] is a circular accelerator, storage ring, and collider for hadrons at the European Council for Nuclear Research (CERN) in Geneva, Switzerland. It is part of the CERN accelerator complex shown in Figure 3.1 and can accelerate charged nuclei to record energies: up to of $6.5\,\mathrm{TeV}$ for protons, up to $2.51\,\mathrm{TeV}$ for lead ions, and up to $2.72\,\mathrm{TeV}$ for xenon ions (cf. e.g. Reference [39]). Several other accelerators are used to pre-accelerate particles before injecting them into the LHC. Protons e.g. are first accelerated in steps by the Linear Accelerator 2 (LINAC2), Proton Synchrotron Booster (BOOSTER), Proton Synchrotron (PS), and Super Proton Synchrotron (SPS) to the injection energy of the LHC, $450\,\mathrm{GeV}$.

Once injected into two separate beam pipes, the charged nuclei circle in opposite directions, held on their trajectories using superconducting magnets. Once per turn, they are accelerated in superconducting cavities, either keeping the *particle beam* at a given energy or increasing it slowly.

The LHC has four interaction points at which the two counter-rotating beams can be put to collision, providing four different detectors with either proton-proton (*pp*), lead-lead (Pb-Pb), proton-lead (*p*-Pb), or xenon-xenon (Xe-Xe) collisions to record. These detectors are the ALICE detector [41] specialized in the recording of collisions involving heavy ions, the LHCb detector [42] with the primary focus on studying the *b* quark, and the two multi-purpose detectors CMS [43] and ATLAS [44].

The LHC saw its first particle beam in 2008. So far, it was operated from 2009 to 2013 (Run 1) at a center-of-mass energy of $\sqrt{s} = 7\,\mathrm{TeV}$ and $8\,\mathrm{TeV}$ and from 2015 to 2018 (Run 2) at a center-of-mass energy of $\sqrt{s} = 13\,\mathrm{TeV}$ for *pp* collisions. During the latter run, a peak instantaneous luminosity (interactions per time and cross section) of $L = 21.0 \cdot 10^{33}\mathrm{cm}^{-2}\mathrm{s}^{-1} = 2.10 \cdot 10^{-5}\mathrm{fb}^{-1}\mathrm{s}^{-1}$ was reached [45]. There was an intermediate upgrade period between 2013 and 2014 called Long Shutdown 1 (LS1) and another still ongoing upgrade period which started 2019 (LS2) before the LHC will recommence with Run 3 in 2022.

3.2. The ATLAS Detector

The ATLAS detector [44] is one of the particle detectors situated around a collision point of the LHC. It is a 4π detector, i.e. it covers as much of the full solid angle as possible (excluding the beam pipe and gaps necessary for the support structure and detector infrastructure) and has a length of approximately $46\,\mathrm{m}$ and a diameter of approximately $25\,\mathrm{m}$, as is illustrated in Figure 3.2. Its geometry is commonly described using cylindrical coordinates, the origin being the center of the detector. The radial coordinate r and the azimuthal angle ϕ span the transverse plane ($\phi \in (-\pi, \pi]$, where $\phi = 0$ points towards the center of the LHC), while
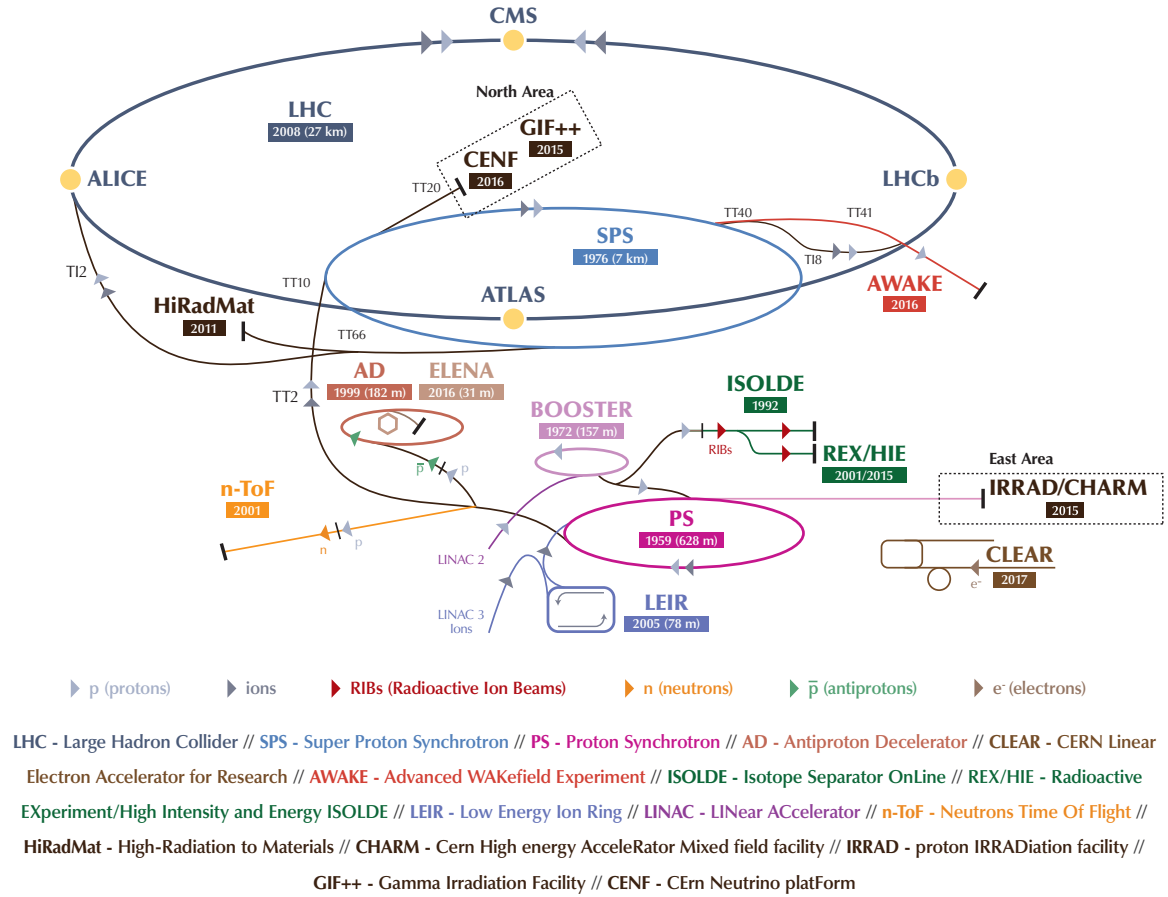
Fig. 3.1: The LHC (dark blue line) as part of the CERN accelerator complex. For proton beams, LINAC2, BOOSTER, PS, and SPS are used as pre-accelerators for the LHC [40].

the polar angle with respect to beam axis is most often expressed using the pseudorapidity $\eta = -\ln\tan\left(\frac{\theta}{2}\right)$, where positive values correspond to a direction clockwise around the LHC if viewed from above. In this thesis, the distance between two objects $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ in the η–ϕ plane is commonly used.

The ATLAS detector consists of different detector components arranged in layers around the collision point. They are described in the following.

3.2.1. The Inner Detector

The innermost part of the detector is the inner detector (ID) [46], which covers the range $|\eta| < 2.5$ and is used for tracking charged particles. It consists of a silicon pixel tracker, a silicon microstrip tracker, and a transition radiion tracker and is illustrated in Figure 3.3.

The pixel detector consists of four layers (the innermost layer having been inserted during the LS1 [47, 48]) arranged concentrically around the vertex area as *barrel* and three disks per *end-cap*. It has the intrinsic resolution of $10\,\mu$m in azimuthal and radial direction and $115\,\mu$m
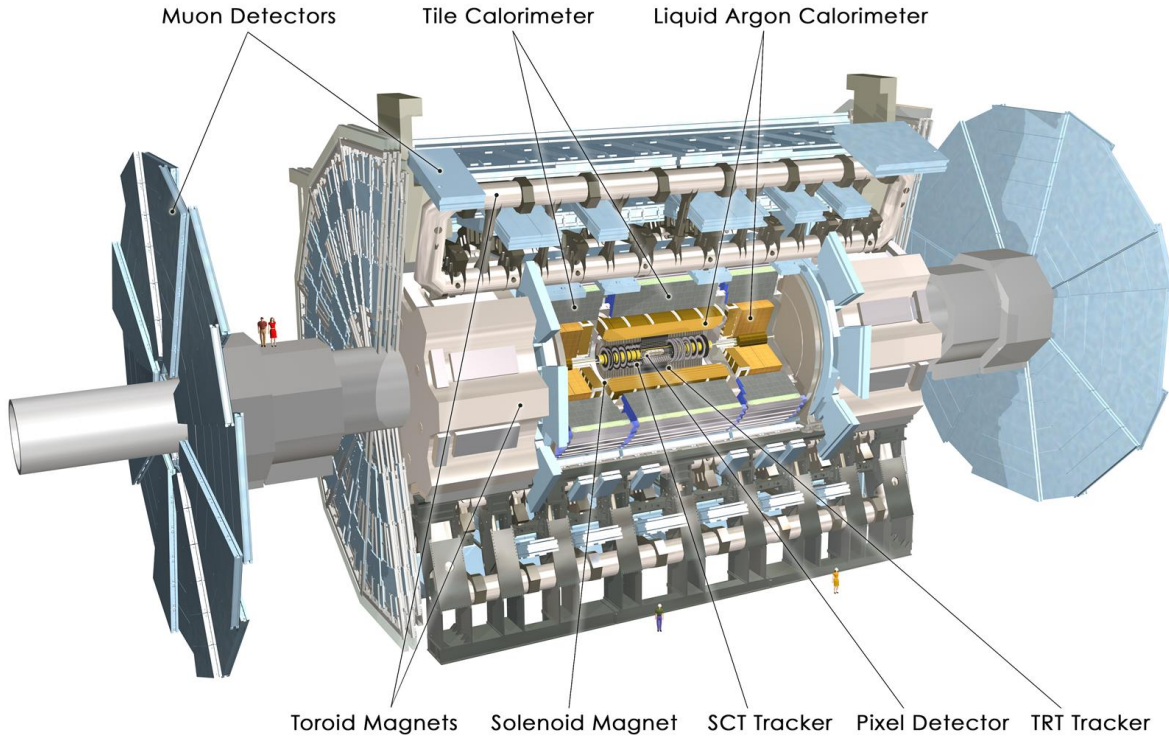
Fig. 3.2: Cut-away overview of the ATLAS detector to illustrate all detector components [44]. The people are shown for scale comparison.

in the z (barrel) and radial direction (end-cap) and provides typically four measurement points called *hits* per track.

The silicon microstrip tracker (SCT) is placed around the pixel detector. It has four double-layers of small-angle stereo strips and nine disks per end-cap and typically provides four additional hits per track. Its intrinsic resolution is $17\,\mu$m in azimuthal and radial direction and $580\,\mu$m in the z (barrel) and radial direction (end-cap).

For $|\eta| < 2.0$, both silicon trackers are surrounded by the transition radition tracker (TRT) which consist of straw tubes filled with a gas mixture containing 70% xenon. It typically provides a set of 36 measuring points per track and additionally offers information aiding electron identification based on transition radiation.

The ID is surrounded by a superconducting solenoid magnet with a $2\,$T axial magnetic field which bends the track of charged particles and therefore enables the reconstruction of the sign of their electric charge and their transverse momentum p_{T}.

3.2.2. The Calorimeter

The ATLAS calorimeter is a sampling calorimeter, i.e. it consists of dense absorber materials in which energetic incoming particles produce particle showers, and active material in which those shower particles produce ionization or photons as detectable signal. A key part of its design is its density, optimized to absorb as much of the particles' energies inside the calorimeter as possible to allow for a good energy resolution even at large energies and to reduce the number
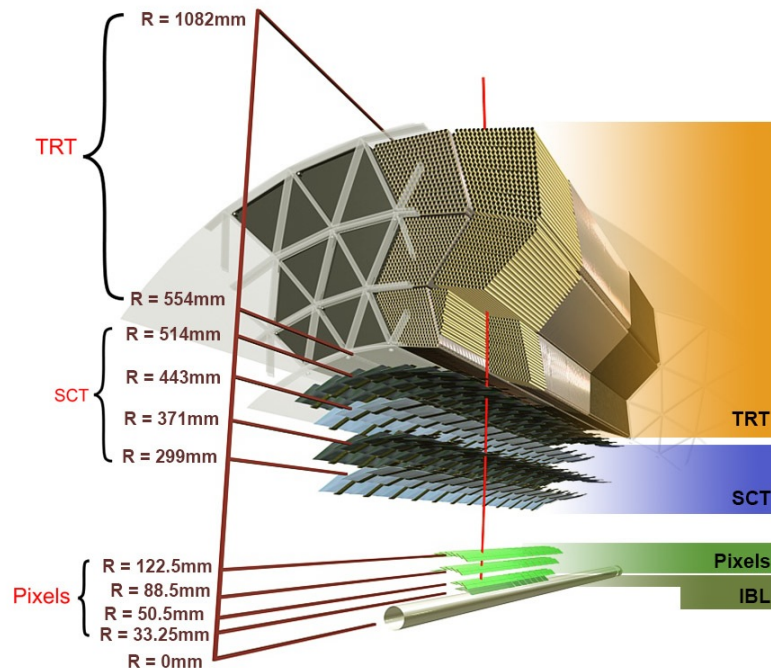
Fig. 3.3: Layout illustrating the scales of the inner detector components in the barrel region of the ATLAS detector [46].

of particles exiting the calorimeter into the muon spectrometer.

Figure 3.4 gives an overview over its four components: the electromagnetic (EM) liquid-argon (LAr) calorimeter, the hadronic LAr end-cap calorimeter, the hadronic tile calorimeter, and the LAr forward calorimeter. The pseudorapidity ranges of each component can be found in Table 3.1.

The EM LAr calorimeter is the innermost calorimeter. It is divided into a barrel with a thickness of at least 22 radiation lengths X_0 and two end-cap calorimeters with $X_0 > 24$. Overall, it consists of a thin LAr presampler and three layers made of lead as absorber material and LAr as active material.

The two hadronic LAr end-cap calorimeter wheels are situated behind the EM end-cap calorimeters. They have four layers built from copper plates as absorber and LAr as active material.

The hadronic tile calorimeter is divided into a barrel and two extended-barrel calorimeters. The barrel part is situated around the barrel part of the EM LAr calorimeter, while the extended-barrel part is situated around the end-cap calorimeters. Both barrel and extend-barrel tile calorimeter have three layers made of steel as absorber and scintillating tiles as active material.

Finally, two forward calorimeters are situated around the beam pipe and between the end-cap calorimeters. They have three layers, the first being built from copper and LAr and optimized for electromagnetic measurements, and the second and third being built from tungsten and LAr and optimized for hadronic measurements.

The entire instrumented calorimeter yields an interaction length λ of 9.7 in the barrel region and 10λ in the end-cap region.
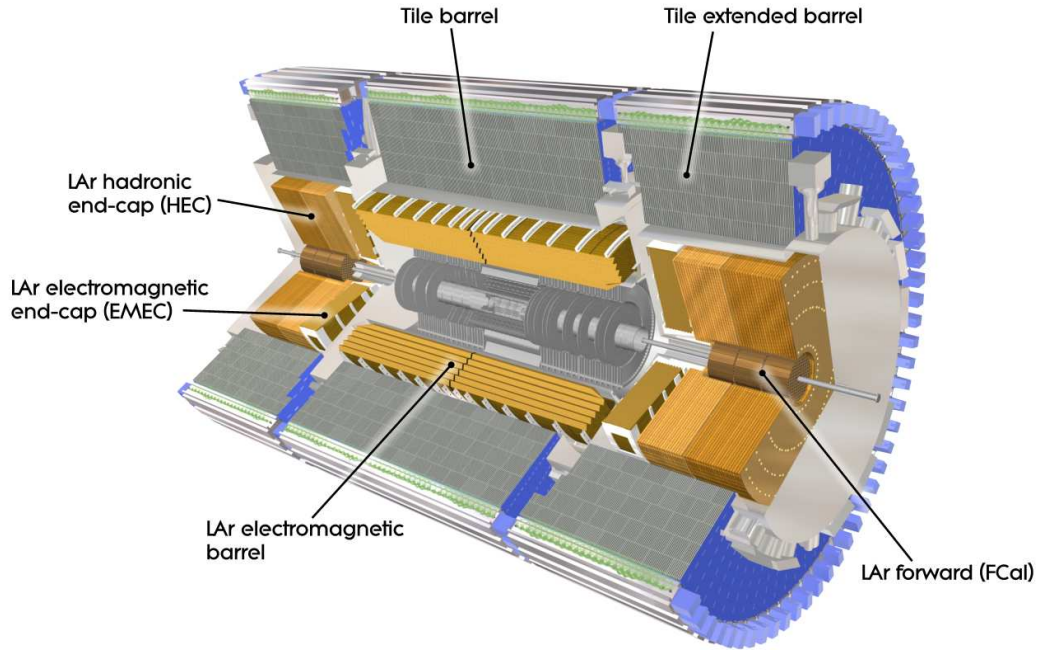
Fig. 3.4: Cut-away overview of the ATLAS calorimeter system [44].

Electromagnetic LAr Calorimeter			
		Barrel	**End-cap**
presampler	(PreSamplerB/E)	$\|\eta\| < 1.52$	$1.5 < \|\eta\| < 1.8$
1st layer	(EMB1/EME1)	$\|\eta\| < 1.475$	$1.375 < \|\eta\| < 3.2$
2nd layer	(EMB2/EME2)	$\|\eta\| < 1.475$	$1.375 < \|\eta\| < 3.2$
3rd layer	(EMB3/EME3)	$\|\eta\| < 1.35$	$1.5 < \|\eta\| < 2.5$
Hadronic Tile Calorimeter			
		Barrel	**Extended barrel**
1st layer	(TileBar0/TileExt0)	$\|\eta\| < 1.0$	$0.8 < \|\eta\| < 1.7$
2nd layer	(TileBar1/TileExt1)	$\|\eta\| < 1.0$	$0.8 < \|\eta\| < 1.7$
3rd layer	(TileBar2/TileExt2)	$\|\eta\| < 1.0$	$0.8 < \|\eta\| < 1.7$
Hadronic LAr End-cap Calorimeter			
1st layer	(HEC0)		$1.5 < \|\eta\| < 3.2$
2nd layer	(HEC1)		$1.5 < \|\eta\| < 3.2$
3rd layer	(HEC2)		$1.5 < \|\eta\| < 3.2$
4th layer	(HEC3)		$1.5 < \|\eta\| < 3.2$
LAr Forward Calorimeter			
1st layer	(FCAL1)		$3.1 < \|\eta\| < 4.83$
2nd layer	(FCAL2)		$3.24 < \|\eta\| < 4.81$
3rd layer	(FCAL3)		$3.29 < \|\eta\| < 4.75$

Tab. 3.1: Pseudorapidities covered by all ATLAS calorimeter layers [44].

3.2.3. The Muon Spectrometer
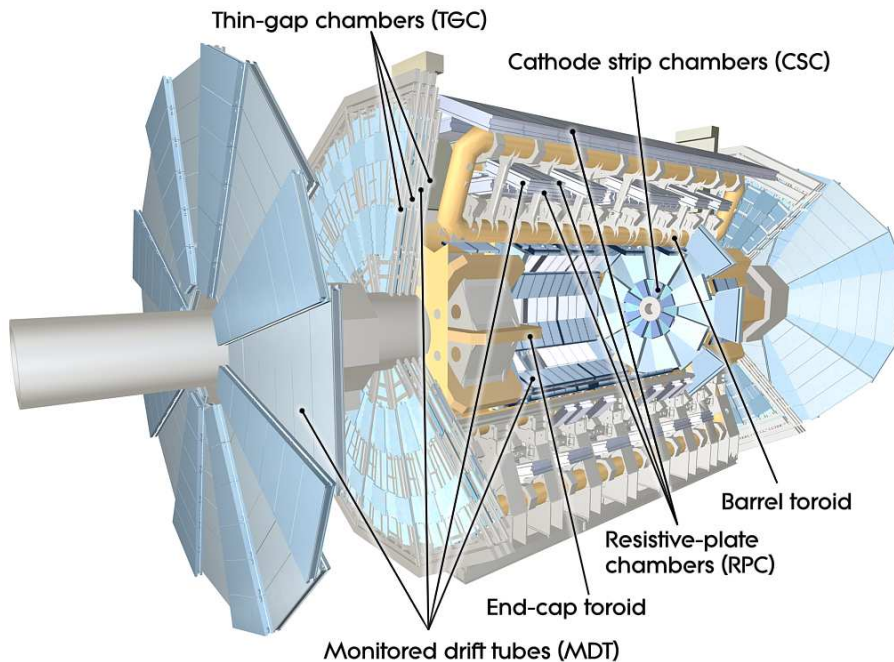
Fig. 3.5: Cut-away overview of the muon system of the ATLAS detector [44].

The outermost layer of the ATLAS detector is the muon spectrometer (MS). It consists of toroidal magnets and separate chambers to trigger event read-out on high-momentum muons and to measure the tracks of muons deflected in the magnetic field.

The barrel magnet system consists of eight long, superconducting air-core coils arranged to form a toroidal magnetic field with a bending strength of 1.5 to 5.5 Tm surrounding the barrel part of the detector. The monitored drift tubes (MDT) used to determine the muons' tracks are immersed in the magnetic field and therefore partially placed into the magnetic coils. For $|\eta| < 1.05$, resistive plate chamber (RPC) are additionally used for dedicated muon triggers.

In the forward direction and per end-cap, the magnetic field is generated by eight smaller coils situated between the coils of the barrel muon system and behind the end-cap calorimeters. They generate a magnetic field with a bending power of 1 to 7.5 Tm. In contrast to the barrel MS, the forward muon detectors are not immersed in the magnetic field of the small coils. The small wheel built from cathode strip chambers (CSCs) is positioned in front of the magnets and covers the high activity region of $2 < |\eta| < 2.7$, while the large wheel consisting of thin gap chambers (TGCs) and MDT, is positioned behind the magnets and covers $1.05 < |\eta| < 2.7$. The large wheel TGCs are used both for measuring muons and triggering on them.

3.2.4. Trigger System

Most collisions during Run 2 were recorded with a spacing of 25 ns between beam packages called *bunches*. This corresponds to an event rate of 40 MHz[1] to record if it were not for dedicated filter systems called triggers, described in detail in Reference [49]. In a first step,

---

[1] The average event rate is smaller than 40 MHz because of gaps in the fill pattern of the LHC.

the hardware trigger (Level-1 trigger, L1) reduces the event rate to 100 kHz. It determines regions of interest (RoIs) based on coarse calorimeter information and input from those muon chambers used for triggering. In a second step, a software trigger (high level trigger, HLT) reduces the event rate to 1 kHz by running a simplified event reconstruction algorithm based on the RoIs and detector information at full granularity.

CHAPTER 4

Object Reconstruction

In analyses, object reconstruction is used to translate the raw information recorded by the detector into objects that can be associated with either single particles (e.g. electrons, photons, muons) or particle jets (e.g. from high-energy quarks).

The main objects in this thesis are electrons, muons, small radius jets, and missing transverse momentum. Additionally, reconstructed tracks and K_S and Λ^0 reconstructed from tracks and matched to jets are part of this thesis. All of these objects are reconstructed and calibrated with the standard ATLAS methods using the Analysis Top framework version 21.2.147 [50].

4.1. Tracks

Charged-particle track reconstruction is described in References [51, 52].

In a first step, clusters of sensors in the ID (pixel detector and SCT detector) in which the measured charge is above a given threshold are defined separately for each ID subsystem. They are then converted to *hits*, i.e. measurements in three dimensions.

The track reconstruction algorithm starts with hits in the pixel and SCT detector. Seeds are formed from combinations of three clusters. They are used in a combinatorial Kalman filter [53] to determine multiple track candidates. An ambiguity resolver is applied to them, selecting tracks based on a quality score and additional quality criteria. The ambiguity resolver is able to handle clusters assigned to more than one track. A neural network helps to determine whether a cluster was caused by one or multiple tracks. Those tracks that pass the ambiguity resolver are then extended to the TRT. Finally, the tracks are fitted using all available information.

4.2. Electrons

Electron reconstruction, calibration, and identification is described in Reference [54].

They are reconstructed from superclusters, i.e. dynamic, variable-size clusters. These superclusters use energy depositions in the EM and hadronic calorimeters that are topologically connected (so-called *topo-clusters* [55]). They are calibrated and matched to reconstructed tracks. Electrons and photons have a similar detector signature. They are reconstructed with similar algorithms. In the final step of their reconstruction, they are classified as either unambiguous electrons or photons or, if both are reconstructed, either one is retained or both electron and photon are retained but marked as ambiguous. Afterwards, electrons are recalibrated, considering additional information of the matched tracks. The energy calibration is done in reconstructed $Z \to e^+e^-$ decays.

To ensure that the quality of the reconstructed electrons matches the requirements of the analyses, additional identification and isolation criteria are utilized. Two different sets of identification criteria are applied in this analysis, one that is more stringent or *tight* (*TightLH*) which is used for the main part of the analysis and one that is less stringent or *loose* (*LooseAndBLayerLH*) which is used for a data driven estimate of fake electrons (most often caused by

jets misidentified as electrons) described in Section 7.3. Both identification working points use a likelihood defined to discriminate between real and fake electrons. In addition to that, the loose identification requires at least one hit in the innermost layer of the pixel detector and all tight electrons have to pass additional isolation requirements w.r.t. other high energy/momentum objects such as energy depositions in the calorimeter and momentum of charged particles not assigned to the electron. The isolation requirement applied in this thesis is called *gradient* isolation. All efficiencies are derived via the tag-and-probe method in $Z \to e^+e^-$ decays.

All electrons used in this thesis are required to have a $p_\mathrm{T} > 27\,\mathrm{GeV}$ and $|\eta| < 1.37$ or $1.52 < |\eta| < 2.47$ which excludes the gap region of the EM calorimeter.

4.3. Muons

The reconstruction, identification, isolation, and calibration of muons is described in References [56, 57].

The muons used in this thesis are reconstructed using a combination of track segments in the MS and tracks in the ID (*combined muons*). Both tracks in the MS and ID are reconstructed independently and combined in a global refit. Most muons are reconstructed by extrapolating tracks in the MS to the ID (*outside-in*), but a complementing inside-out extrapolation is used as well. If it improves the fit quality, hits in the MS can be removed.

Similar to electrons, identification criteria are applied to muons in order to ensure the quality of their reconstruction. In this thesis, all muons have to fulfill criteria with *medium* strictness which minimize the uncertainties associated with the reconstruction and calibration. These criteria use requirements on the hits in the MDT as well as on the significance of the measured ratio between charge and momentum, q/p.

The muons used in the nominal selections in this thesis are additionally required to be isolated following the isolation requirement *FCTight_FixedRad*. It requires the transverse energy of topological cell clusters in a cone of the size $\Delta R = 0.2$ around the position of the muon, $E_\mathrm{T}^{\mathrm{topocone20}}$, divided by the $p_\mathrm{T}(\mu)$ of the muon to be smaller than 15%. Additionally, requirements on the scalar sum of the transverse momenta of ID tracks with $p_\mathrm{T} > 1\,\mathrm{GeV}$ associated with the primary vertex[1] in a cone with radius $\Delta R = 0.2$ (called $p_\mathrm{T}^{\mathrm{cone20}}$) or in a cone with $\Delta R = \min(10\,\mathrm{GeV}/p_\mathrm{T}^\mu, 0.3)$ (called $p_\mathrm{T}^{\mathrm{varcone30}}$) are applied. If the muon has a $p_\mathrm{T}(\mu) < 50\,\mathrm{GeV}$, this requirement is $p_\mathrm{T}^{\mathrm{cone20}}/p_\mathrm{T}(\mu) < 0.04$, otherwise it is $p_\mathrm{T}^{\mathrm{varcone30}}/p_\mathrm{T}(\mu) < 0.04$. Further details on the definition of these variables is given in Reference [57].

The muons used in the estimate of fake muons are not required to fulfill any isolation requirements.

The muons' momenta are calibrated by comparing data and simulation in $Z \to \mu^+\mu^-$ events and scale factors for the scale and resolution are extracted [57].

In this thesis, all muons are required to have a $p_\mathrm{T} > 27\,\mathrm{GeV}$ and an $|\eta| < 2.5$.

4.4. Jets

Particle jets are the result of high-energy quarks or gluons showering and hadronizing. Because of momentum conservation, in the detector, jets appear as groups of energy depositions, which

---

[1]The primary vertex (PV) is the vertex with the highest $\sum p_\mathrm{T}^2$, where the sum includes all tracks fulfilling certain quality criteria and criteria matching them to that vertex. Details are described in Reference [58].

can be clustered into objects. The anti-k_T algorithm [59] used to cluster all jets used in this thesis is *infrared and collinear safe*, i.e. low-energy and collinear radiation does not change the resulting jets. With this clustering algorithm in particular, the resulting jets tend to have a circular shape. The radius parameter used for the clustering is $R = 0.4$ which is equivalent to the jets' maximum radius.

In this thesis, jets are clustered from topological calorimeter clusters [55] calibrated to the electromagnetic scale [60].

After the clustering, the jets are calibrated in four steps as described in Reference [61]. In an initial step, the momentum of jets is corrected w.r.t. the pileup condition of their recording. Afterwards, the energy scale is corrected based on Monte Carlo (MC) simulations. Thereafter, a global sequential calibration is applied which e.g. corrects for differences between the jets' response for different jet flavors (quark or gluon). In this step, a correction for punch throughs[2] is applied as well. The jets are then calibrated in situ in dijet events, multijet events, and Z/γ+jets events.

All jets are required to have a $p_\mathrm{T} > 25\,\mathrm{GeV}$. Additional requirements on the jets' p_T will be discussed in the context of the use of b-tagging algorithms in the following section.

During each bunch crossing, a number of proton pairs interact. The softer interactions that happen simultaneously to a hard interaction that is recorded are referred to as *pileup*. To reduce the number of jets from these pileup interactions, a cut on the jet-vertex tagger (JVT) discriminator [62] is required, which selects jets based on whether tracks matched to them originate in the PV. If the jet has a pseudorapidity of $|\eta| < 2.4$ and a $p_\mathrm{T} < 120\,\mathrm{GeV}$, a value greater than 0.59 is required, and if $|\eta| > 2.4$ and $p_\mathrm{T} < 120\,\mathrm{GeV}$, a value greater than 0.11 is required.

4.5. b-Tagging

b-tagging algorithms are used to identify jets that originate from b quarks.

Most B hadrons, i.e. hadrons containing a b quark, such as $B^\pm$, B^0, B^0_S, etc., have a lifetime of $\tau \approx 1.5\,\mathrm{ps}$ [24] and therefore – at energies typical for collisions at the LHC – tend to traverse a few millimeters inside the detector before decaying. The resulting decay vertices – also called secondary vertices (SVs) – can be reconstructed because of the precise tracking resolution provided by the ID.

In this thesis, the *MV2c10* algorithm discussed in References [63, 64] is used for the identification of jets from b quarks. It uses a boosted decision tree with inputs from other algorithms (*JetFitter*, *IP3D*, and *SV1*), each of which is specialized in the reconstruction of particular characteristics of b-jets such as the decay chain of B hadrons, the impact parameter (IP) of tracks, and SVs. A working point with an 85% efficiency for b-jets and a rejection rate for non-b-jets of $1/\varepsilon_\mathrm{light} = 28$ is chosen. It has the smallest rejection rate for non-b-jets of all calibrated working points and hence the largest purity of non-b-jets if jets are required not to be b-tagged[3].

Two kinds of b-tagger calibrations are used in different parts of this thesis. In the s-tagger calibration described in Chapter 7, the b-tagging calibration detailed in Reference [63] is applied. The calibration of b-jets is done in a likelihood fit in dileptonic $t\bar{t}$ decays and assumes $|V_{tb}| = 1.0$. Therefore, this calibration is biased in such a way that it cannot be used for the

---

[2]Punch through means that a jet is not fully contained in the calorimeter but leaks into the MS.

[3]For the selection of $t \to W^+ s$ decays, it is of great interest to reject b-jets from $t \to W^+ b$ decays.

measurement of $|V_{ts}|$ and $|V_{td}|$ described in Chapter 8. Thus, in this measurement, a different b-tagger calibration is applied. It uses the $p_{\mathrm{T}}^{\mathrm{rel}}$ method which is described in Reference [65] and makes use of b-jets containing μ, which occur in roughly 20% of B hadron decays. The contribution of each jet flavor is fitted in the distribution of the momentum of these μ relative to the jet axis, $p_{\mathrm{T}}^{\mathrm{rel}}$. An ATLAS internal documentation of the calibration for Run-2 data can be found in Reference [66][4]. The resulting scale factors for b-tagged b-jets are compatible with one and have an uncertainty between 2 and 8%. For Run-2 data, currently, the $p_{\mathrm{T}}^{\mathrm{rel}}$ calibration is limited to jet $p_{\mathrm{T}} < 140\,\mathrm{GeV}$, while for Run 1, the limit of this calibration was $p_{\mathrm{T}} < 200\,\mathrm{GeV}$. For this reason, a requirement of $p_{\mathrm{T}} < 200\,\mathrm{GeV}$ was originally imposed on all jets used in this thesis. This requirement is kept for the calibration of the s-tagger. For the measurement of $|V_{ts}|$ and $|V_{td}|$, however, the stricter criterion of $p_{\mathrm{T}} < 140\,\mathrm{GeV}$ is applied to all jets.

4.6. c-Tagging

Similar to b-tagging algorithms, c-tagging algorithms are used to identify jets that originate from c quarks. In comparison to B hadrons, the lifetimes of D hadrons [5] are smaller, e.g. $1.0\,\mathrm{ps}$ for $D^{\pm}$ and $0.4\,\mathrm{ps}$ for D^0 [24]. If they are highly boosted, however, they can still traverse millimeters before decaying, which allows for the use of similar tagging techniques as for b-jets.

The c-tagger employed in this thesis is called *DL1* [63] and uses the same input variables as the MV2c10 algorithm but a deep neural network instead of a boosted decision tree. It is applied at the *tight* working point, which has a c-jet efficiency of 17%, and in combination with an anti-b-tag of the MV2c10 b-tagger at a b-tagging working point that has a 70% efficiency for b-jets. This combination of taggers and its calibration was developed specifically for the search of the process $VH(\to c\bar{c})$ [67]. It is described in detail in an internal documentation, cf. Reference [68]. Scale factors are derived for b-jets, c-jets, light jets, and τ-jets. For b-jets, this is done in a likelihood fit in a sample enriched with b-jets based on a selection of dileptonic $t\bar{t}$ decays. The uncertainty on the scale factors of b-jets is mostly below $\pm 5\%$ for $p_{\mathrm{T}} < 200\,\mathrm{GeV}$. Scale factors for c-jets are derived by considering the two jets from the hadronically decaying W boson in semileptonic $t\bar{t}$ decays, of which one is required to be c-tagged and the other is required not to be c-tagged. The resulting scale factors have an uncertainty of approximately $\pm 10\%$ for $p_{\mathrm{T}} < 200\,\mathrm{GeV}$. These scale factors for c-jets are additionally used in an extrapolation of scale factors for τ-jets, which have a resulting uncertainty of approximately $\pm 22\%$. The extraction of scale factors for light jets makes use of an alternative version of the DL1 tagger (*DL1Flip*), which uses similar input variables as the original tagger. However, some of these input variables are flipped, such as e.g. the sign of the tracks' IPs. This results in a similar efficiency for light jets (their IPs tend to have positive and negative signs with roughly equal probability), while the efficiency is lower for heavy flavor jets (their IPs tend to be positive more often). Details on the procedure can be found in Reference [69]. For the MV2c10 tagger used in this combination of taggers, the standard version and not the flipped version was used for the calibration of light jets. The resulting scale factors have an uncertainty of approximately $\pm 15\%$ for $p_{\mathrm{T}} < 200\,\mathrm{GeV}$.

---

[4]The scale factors derived by this calibration are not yet fully approved by the ATLAS Collaboration. They are applied in this thesis to allow for an estimation of their impact on the measurement of $|V_{ts}|$ and $|V_{td}|$.

[5]D hadrons contain at least one c quark and no b quarks.

4.7. Overlap Removal

Since the reconstruction algorithms are applied independently, it might happen that different objects are reconstructed in the same detector region. The following algorithm is used to resolve the overlap by removing all but one of the objects:

First, electrons that share tracks with muons are removed. Second, if jets fall into the reconstruction radius of an electron ($\Delta R < 0.2$), they are removed. Third, all electrons that are closer than $\Delta R < 0.4$ to a jet are removed. Finally, all muons with a distance $\Delta R < 0.4$ to a jet are removed.

4.8. Missing Transverse Momentum

The cross section of any neutrino interaction is very small, which means they cannot be detected with the ATLAS detector. However, since the momentum of the protons accelerated by the LHC is very small in the transverse plane of the detector, i.e. perpendicular to the beam axis, it is possible to reconstruct the momentum carried away by neutrinos in this direction. This missing transverse momentum, $E_\mathrm{T}^\mathrm{miss}$, is the negative vectorial sum of all reconstructed and calibrated objects plus unclustered soft energy estimated from low-p_T tracks matched to the primary vertex, cf. Reference [70].

4.9. K_S and Λ^0 Reconstruction

Strange hadrons are hadrons that contain at least one s quark and neither c nor b quarks.

One type are kaons, which are mesons containing an s ($\bar{s}$) quark and a $\bar{u}$ or $\bar{d}$ (u or d) quark. While $K^0 \mathrel{\hat=} d\bar{s}$ and $\overline{K}^0 \mathrel{\hat=} \bar{d}s$ are eigenstates of their flavor content, experimentally, it is handy to consider them in the eigenstates of the weak interaction, $K_S \mathrel{\hat=} \frac{d\bar{s}-\bar{d}s}{\sqrt{2}}$ and $K_L \mathrel{\hat=} \frac{d\bar{s}+\bar{d}s}{\sqrt{2}}$. K_S are more short lived and mostly decay into two pions[6] ($\pi^+\pi^-$ or $\pi^0\pi^0$), while K_L are more long lived and mostly decay into three particles (e.g. $\pi^+\pi^-\pi^0$ or $\pi^0\pi^0\pi^0$).

Another type of strange hadrons are Λ^0 ($\overline{\Lambda^0}$) baryons containing an s, d, and u ($\bar{s}$, $\bar{d}$, and $\bar{u}$) quark. They mostly decay into either $p+\pi^-$ or $n+\pi^0$ ($\bar{p}+\pi^+$ or $\bar{n}+\pi^0$). In the following, both Λ^0 and their antiparticles $\overline{\Lambda^0}$ will be considered simultaneously and both will be referred to as Λ^0.

K_S decaying to $\pi^+\pi^-$ and Λ^0 decaying to $p+\pi^-$ are reconstructed using the *InDetV0Finder* Tool [71] which considers reconstructed tracks, all of their uncertainties, and their correlations. Parts of the description of the algorithm are based on Reference [72].

In the used algorithm setup, only tracks reconstructed from hits in the silicon tracker (pixel detector and SCT) with a $p_\mathrm{T} > 400\,\mathrm{MeV}$, a longitudinal IP $|z_0| < 250\,\mathrm{mm}$, and a transverse IP divided by its significance $d_0/\sigma(d_0) > 2$ are considered. For each fit, two tracks are selected as input if they carry opposite-sign electric charge. When a track is assumed to be a pion, its mass is set to $m_\pi = 139.57\,\mathrm{MeV}$, and when it is assumed to be a proton, the mass is set to $m_p = 938.272\,\mathrm{MeV}$.

For the reconstruction, at first, a three-dimensional vertex fit without any constraints on the mass or direction of the decaying particle is carried out. To enter the fit, at least one invariant mass of the two tracks – permuting the assumptions of the tracks being pions and protons – is

---

[6]Pions are mesons consisting either of $u\bar{d}$, $\bar{u}d$, $u\bar{u}$, or $d\bar{d}$.

required to be in the following ranges: $400\,\text{MeV} < m(\pi\pi) < 600\,\text{MeV}$ or $1000\,\text{MeV} < m(p\pi) < 1200\,\text{MeV}$ or $1000\,\text{MeV} < m(\pi p) < 1200\,\text{MeV}$. The fit projects the helical trajectories of the two tracks into the (r, φ) plane in cylindrical coordinates and approximates them as circles. If the two circles intersect, the helices are extrapolated to the two intersections and the one with the smaller z difference of the helices is chosen as the vertex. If the circles do not intersect, the point of closest approach is chosen as the approximation for the vertex.

If the invariant mass of the resulting vertex is either $400\,\text{MeV} < m(\pi\pi) < 600\,\text{MeV}$ or $1000\,\text{MeV} < m(p\pi) < 1200\,\text{MeV}$ or $1000\,\text{MeV} < m(\pi p) < 1200\,\text{MeV}$ and the cumulative χ^2 probability is larger than 0.0001, the vertices are further constrained in a fit of the same type but under K_S and Λ^0 mass hypotheses ($m_{K_S} = 497.672\,\text{MeV}$ and $m_\Lambda = 1115.68\,\text{MeV}$). The mass-constrained vertices are retained if the cumulative χ^2 probability of the vertex fit is larger 0.0001, which means that the χ^2 value is smaller than 15.

Chapter 5

Data and Monte Carlo Simulation

5.1. Data

The data studied in this thesis corresponds to the full Run-2 dataset taken by the ATLAS detector during the LHC runs with pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ and a bunch spacing of $25\,\mathrm{ns}$ between 2015 and 2018. It corresponds to an integrated luminosity of $139\,\mathrm{fb}^{-1}$ with an uncertainty of 1.7% [73].

5.2. Monte Carlo Simulation

The following description of simulated samples is based on Reference [74].

The generation of MC simulated events is done in three steps.

First, using perturbation theory, events of a certain process are simulated on parton level according to the process' probability given by its matrix element (ME), folded with the probability of the initial particles of the process described by the parton distribution function (PDF) of two protons colliding at a center-of-mass energy of $13\,\mathrm{TeV}$. Then, the hadronization of all resulting strongly interacting particles is simulated using so-called parton shower algorithms. A detailed description of the software and settings used for this step and for each considered process is given below.

Second, all events are passed through a detailed simulation [75] of the ATLAS detector implemented within the Geant 4 toolkit [76]. This produces so-called *hits*, i.e. energy deposits in sensitive detector volumes, which are converted into *digits* (e.g. digitized voltages or currents in the detector) in a step called *digitization*. During the digitization step, the simulated hard-scattering events, i.e. the pp collisions that produces the processes of interest, are overlaid with a number of *minimum-bias* events, i.e. softer pp collisions that have a large cross section, to account for multiple interactions in the same and neighboring bunch crossings (*pileup*). These minimum-bias events are simulated using the event generator Pythia8.186 [77] with the NNPDF2.3LO PDF set [78] and the tuned parameters from the A3 tune [79].

And third, all events are passed through the same reconstruction algorithms also used for data described in Section 4.

In a procedure called *pileup reweighting*, all MC events are reweighted w.r.t. their distribution of the simulated number of interactions per bunch crossing, μ, to reproduce the distribution of average interactions per bunch crossing, $\langle\mu\rangle$, observed in data. Before, the $\langle\mu\rangle$ in data is divided by 1.03 ± 0.07, a scale factor based on the measurement of the cross section of inelastic pp collisions [80], to increase the agreement between the number of pileup vertices in data and simulation.

5.2.1. Pair-Produced Top Quarks

The production of t quark pairs ($t\bar{t}$) is simulated at next-to-leading order (NLO) accuracy in the strong coupling constant α_s using the ME generator PowhegBox v2 [81–84] with the

NNPDF3.0NLO [85] PDF set. The t quark mass is set to $m_t = 175\,\mathrm{GeV}$ and the h_{damp} parameter[1] is set to $1.5\,m_t$.

The t quark decay simulated at leading order (LO) is handled by MADSPIN [86, 87] in order to preserve all spin correlations and to allow for the simulation of decays to s and d quarks. The parton shower and hadronization is simulated with the PYTHIA8.230 [77] generator, using the tuned parameters from the A14 set [88] and the NNPDF2.3LO PDF set [78]. Bottom and charm hadron decays are simulated with the EVTGEN v1.6.0 program [89]. The nominal $t\bar{t}$ sample contains the decay $t\left(\to W^+ b\right)\bar{t}\left(\to W^-\bar{b}\right)$ and will be referred to as $tt(WbWb)$ in this thesis. Dedicated signal samples contain either the decays $t\left(\to W^+ s\right)\bar{t}\left(\to W^-\bar{b}\right)$ and $t\left(\to W^+ d\right)\bar{t}\left(\to W^-\bar{b}\right)$ with a ratio of 3:1 or the decays $t\left(\to W^+ b\right)\bar{t}\left(\to W^-\bar{s}\right)$ and $t\left(\to W^+ b\right)\bar{t}\left(\to W^-\bar{d}\right)$ with a ratio of 3:1. For simplicity, these decays will be referred to as $tt(WsWb)$ and $tt(WdWb)$ in this thesis.

The cross section of all $t\bar{t}$ samples is corrected to

$$\sigma(t\bar{t})_{\mathrm{NNLO+NNLL}} = 832 \pm 51 \text{ fb}\,,$$

the theory prediction computed with the TOP++2.0 program [90–96] at next-to-next-to-leading order (NNLO) accuracy in QCD, including the resummation of next-to-next-to-leading logarithmic (NNLL) soft gluon terms.

Kinematic distributions of both t and $\bar{t}$ quark are corrected to the NNLO predictions of their p_T and rapidities and the mass and rapidity of the $t\bar{t}$ system given in Reference [97]. This is done using the *TTbarNNLOReweighter* package described in Reference [98].

For samples generated with an alternative parton shower, PYTHIA8.230 is exchanged with HERWIG7.04 [99, 100]. It uses the H7UE tune [100] and the `MMHT2014LO` PDF set [101].

5.2.2. Singly-Produced Top Quarks

The weak production of single t quarks (via t-channel, s-channel, and associated Wt production) is modeled using the POWHEGBOX v2 [82–84, 102] in the production, MADSPIN [86, 87] for the decay of the t quarks, PYTHIA8.230 [77] for the hadronization and the EVTGEN v1.6.0 program [89] for the decay of bottom and charm hadrons.

For the associated Wt and the s-channel production, the five-flavor scheme and therefore the NNPDF3.0NLO [85] PDF set is used. The t-channel production is simulated using the four-flavor scheme with the NNPDF3.0NLOnf4 [85] PDF set. The t quark mass, the renormalization and factorization scale as well as the sets of tuned variables in the shower and hadronization is the same as in the production of $t\bar{t}$. To treat interference between Wt and $t\bar{t}$, the diagram removal scheme described in Reference [103] is applied.

The inclusive cross sections of all three production processes are corrected to their theory predictions, which are

$$\sigma(t,\ t\text{-channel})_{NLO} = 136.02^{+5.40}_{-4.57}\,\mathrm{pb} \text{ and } \sigma(\bar{t},\ t\text{-channel})_{NLO} = 80.95^{+4.06}_{-3.61}\,\mathrm{pb}$$

for t-channel production,

$$\sigma(t,\ s\text{-channel})_{NLO} = 6.35^{+0.23}_{-0.20}\,\mathrm{pb} \text{ and } \sigma(\bar{t},\ s\text{-channel})_{NLO} = 3.97^{+0.19}_{-0.17}\,\mathrm{pb}$$

---

[1]This parameter regulates the momentum of the first additional emission against which the $t\bar{t}$ system recoils.

for *s*-channel production, and

$$\sigma(tW)_{NLO+NNLL} = 71.7 \pm 3.8 \,\text{pb} \tag{3}$$

for associated *Wt* production. They are calculated at NLO in QCD (with NNLL soft gluon corrections in case of *tW* production) [104–107].

5.2.3. *W*+Jets, *Z*+Jets, and Diboson Production

The ME generation, hadronization, and parton shower for the production of *W* and *Z* bosons in association with additional QCD radiation (*W*+jets and *Z*+jets, together called *V*+jets) as well as the production of two bosons called diboson production (*WW*,*WZ*, and *ZZ*) is simulated with the Sherpa v2.2 [108] generator using the NNPDF3.0NNLO PDF set [85] and parton-shower tunes provided by the SHERPA authors.

The *V*+jets samples are of NLO accuracy for up to two parton radiations and of LO accuracy for up to four additional partons in the ME. The *V*+jets cross section is normalized to a differential prediction [109] with a normalization uncertainty of ±5%. Additional information on *V*+jets production is provided in Reference [110].

The diboson samples are of NLO accuracy for up to one additional parton radiation and of LO accuracy for up to three additional partons. Their normalization uncertainty of ±6% is derived by varying the QCD factorization scale [111]. Additional information on the diboson production can be found in Reference [112].

CHAPTER 6

An *s*-Tagging Algorithm

The identification of *s*-jets, i.e. jets from *s* quarks, called *s*-tagging is a relatively new effort at LHC experiments. A promising approach is the use of neural networks (NNs), which is already common for flavor tagging algorithms in use today, cf. e.g. the *c*-tagging algorithm described in Section 4.6. NNs can learn small differences in and correlations between input variables, which is useful as the fragmentation of *u*, *d*, and *s* quarks and hence the resulting jets are very similar.

In the past, at the SLD at the Stanford Linear Accelerator Center (SLAC) and Delphi Experiments at the Large Electron-Positron Collider (LEP) [113], *s*-jets were analyzed by selecting $K^{\pm}$ mesons which were identified with the help of Cherenkov detectors[1] [114, 115].

During this thesis project, two papers on *s*-tagging based on NNs were published: One explored the use of NNs to process track information for *s*-tagging [30], while the other used both track and calorimeter information [29] as input to multivariate methods.

This chapter covers the development of *s*-tagging algorithms in the context of this thesis. In the first section, the used NN structures and training methods are described in detail. The second section features a study exploring an upper bound on the performance of an *s*-tagger assuming a number of simplified detector scenarios. And in the third section, the development of an *s*-tagger for the use at the ATLAS Experiment is presented.

6.1. Deep Neural Networks

In this chapter, two different types of NNs are considered. One is a fully-connected feed-forward deep neural network (DNN) structure and the other is a recurrent structure (RNN) using Long Short-Term Memory (LSTM) layers. All NNs in this thesis are trained using the *TensorFlow* library for machine learning [116] interfaced by the *Keras* library [117].

6.1.1. Fully-Connected Feed-Forward Neural Networks

NNs consist of interconnected units called *nodes*. Commonly, they are organized into distinct layers. In a feed-forward NN, information is always passed in one direction. If a NN is fully connected, it means that each node in one layer is connected to all nodes in the following layer.

Each DNN consists of an input layer, several hidden layers, and an output layer. The first layer consists of *n* nodes, where *n* is the number of input variables. The hidden layers contain an arbitrary number of nodes. However, for the DNNs trained in this thesis, the number of nodes in each hidden layer is chosen to be a power of two, the number of nodes in the first hidden layer defining the *width* (*w*) of the NN, and each following hidden layer contains half as many nodes as the previous hidden layer. The output layer is chosen to contain exactly one node, which is sufficient for a two class (signal vs. background) classification. Figure 6.1 is a graphical representation of what a small-scale example of such a DNN could look like.

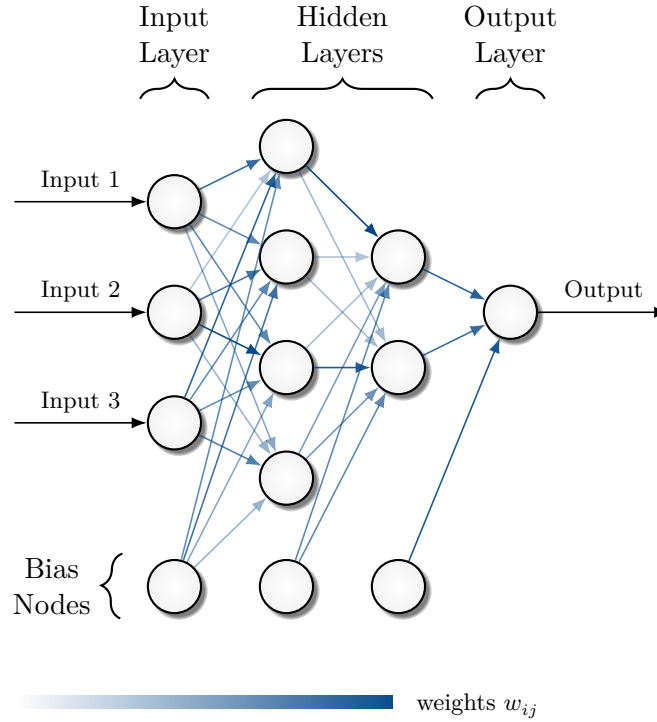[1]The functionality of Cherenkov detectors will briefly be discussed in Section 6.2.3.

Fig. 6.1: Small-scale example of the structure of the deep neural networks trained for the identification of s-jets. Each circle visualizes a node; all circles aligned vertically correspond to a layer. The bottom row shows bias nodes initialized with the value one, each one being connected to the nodes in a hidden layer or the output layer. The weights between nodes are visualized by arrows, their opacity corresponding to the numerical value of the weight and the arrows represent the flow of information. The graphic is based on Reference [118].

The connection between each node i in one layer and another node j in the following layer is characterized by a weight w_{ij}. In addition to the input nodes, there are a number of bias nodes containing the value one connected to the nodes of the hidden layers via connection weights b_j. With this, the numerical input value to each node j in the hidden layers and the output layer is

$$I_j = \sum_i w_{ij}O_i + b_j, \tag{4}$$

where O_i is the output value of the node i in the previous layer. The bias weights b_j shift the input values I_j for more flexibility in the model built by the NN. In the input layer, the value of each node is the value of the corresponding input variable.

Within each node i, an activation function $f(x)$ is used to transform the input value I_i to the output value O_i. In each hidden layer, a *rectifier*

$$f(x) = \max(0, x) = \begin{cases} x \text{ if } x > 0 \\ 0 \text{ if } x \leq 0 \end{cases} \tag{5}$$

is used as activation function because it is non-linear (i.e. it can be used to model non-linear characteristics), it has an easy-to-calculate derivation, and it avoids the problem of vanishing

gradients if the DNN is too large [119]. Each node in the hidden layer is called a rectified linear unit (ReLU). The node in the output layer uses a *sigmoid function*

$$f(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{e^x + 1} \tag{6}$$

as activation. It is monotonic and maps values from $(-\infty, \infty)$ to $(0, 1)$, the selected range of the output values of the NN. Both activation functions are shown in Figure 6.2.
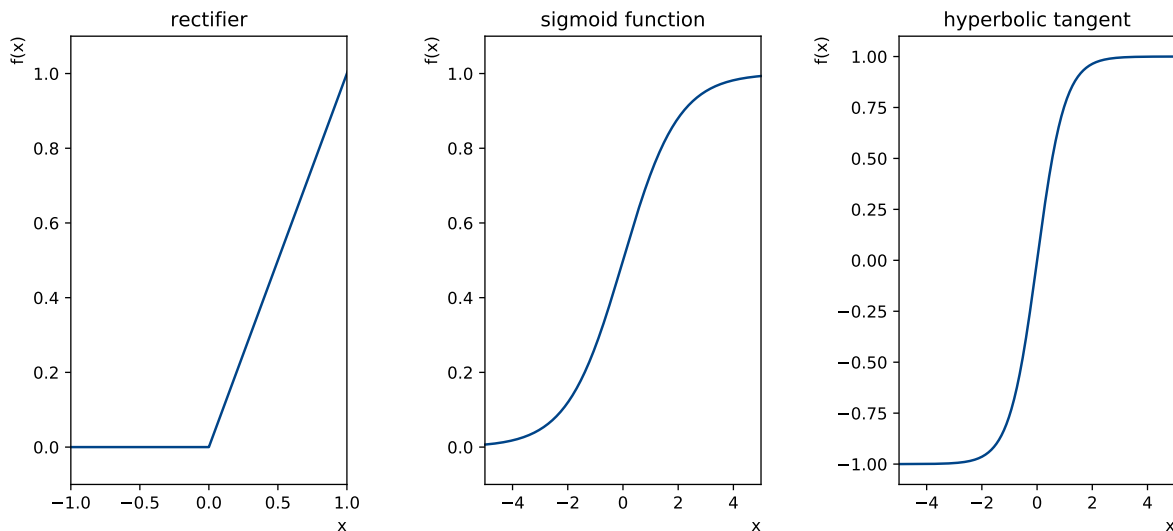

Fig. 6.2: The activation functions used in the neural networks trained for *s*-tagging, a *rectifier* used in the nodes of the hidden layers (left), a *sigmoid function* used in the output node (middle), and a *hyperbolic tangent* used in the recurrent connections of Long Short-Term Memory layers (right) .

6.1.2. Recurrent Neural Networks Using Long Short-Term Memory

In RNNs, an additional dimension is added to the input of the NN, which in most contexts is referred to as *time steps*. By adding this additional dimension, input sequences of arbitrary length (i.e. containing arbitrary number of time steps) can be processed. During the first time step, a recurrent layer calculates its output only based on the input of the first time step. During the following time steps, in addition to the new inputs of the time step, the output of the layer from the previous time step is fed back into the layer. An illustration of this process is shown in Figure 6.3. If the recurrent layer is followed by another recurrent layer, the output of the earlier recurrent layer is passed to the next layer for each time step. If the following layer is not recurrent, only the output of the last time step is passed forward.

There are a variety of ways to take the information from the previous time step into account when processing information in the following time step. LSTM layers [120] do so in a sequence of four steps which enable them to carry information both long-term (i.e. through an arbitrary number of time steps) and short-term (i.e. through a single time step). A node in an LSTM layer can also be called a *cell*, following an analogy of circuit boards. The central component of LSTMs are *cell states* associated to each node, which can be thought of as the layer's memory.
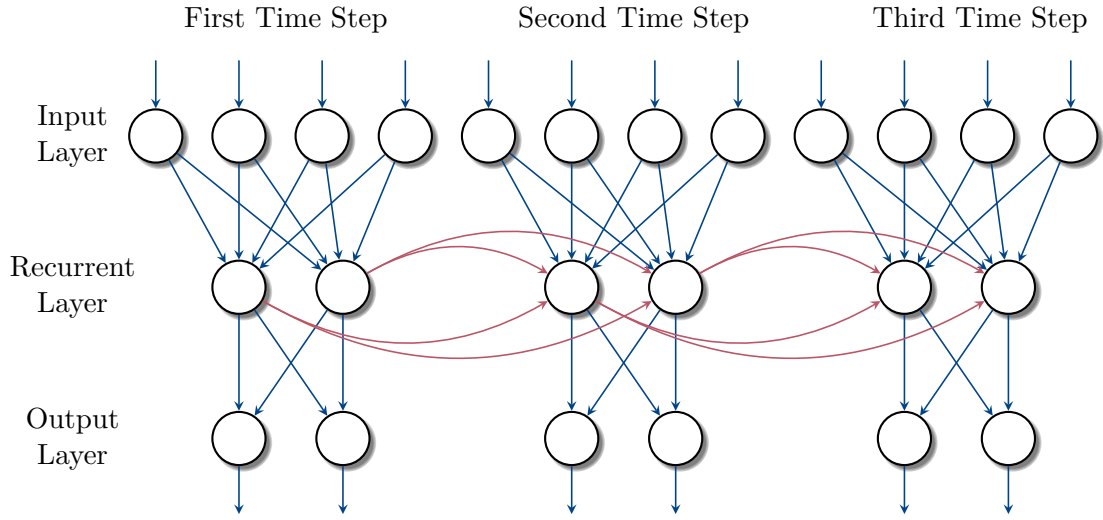
Fig. 6.3: Illustration of a recurrent neural network. Each circle visualizes a node; all circles aligned horizontally correspond to a layer. The weights between nodes are visualized using arrows, showing the flow of information. Red arrows indicate output values being carried to the next time step, while blue arrows indicate the forward feeding of information. For simplicity, no bias nodes are shown.

These cell states are modified during each time step based on the information passed to them. This is done using *gates*, which connect the inputs via a predefined function to a single output.

An LSTM layer consists of n units, each containing a cell state C_{t-1} calculated during the previous time step, which can be written as $\vec{C}_{t-1}$, a vector of length n. Per time step t, it receives an input vector $\vec{x}_t$ of length m. The following operations take place per time step:

In the first step, the per-cell values $f_t \in [0,1]$ for the *forget gate* are calculated based on the output $\vec{h}_{t-1}$ of the previous time step and the input $\vec{x}_t$ of the current time step, deciding how much information residing in the previous cell state is going to be retained:

$$\vec{f_t} = \sigma(W_f \vec{x}_t + U_f \vec{h}_{t-1} + \vec{b}_f) \,, \tag{7}$$

where σ is the sigmoid function applied to each element of its input vector, W_f is a matrix of dimension $n \times m$, U_F is a matrix of dimension $n \times n$ and $\vec{b}_f$ are n values corresponding to the weights of bias nodes.

In the second step, the values $i_t \in [0,1]$ of the *input gate* are calculated, deciding how much new information will be stored in the cell state during this time step:

$$\vec{i_t} = \sigma(W_i \vec{x}_t + U_i \vec{h}_{t-1} + \vec{b}_i) \,, \tag{8}$$

In the third step, the cell state is updated with a temporary cell states

$$\vec{\tilde{C}_t} = \tanh(W_c \vec{x}_t + U_C \vec{h}_{t-1} + \vec{b}_C) \,, \tag{9}$$

and the new cell states of the cell is set to

$$\vec{C_t} = \vec{f_t} \odot \vec{C}_{t-1} + \vec{i_t} \odot \vec{\tilde{C}_t} \,, \tag{10}$$

in which the forget gate and the input gate are applied to the old and the new cell state by using element-wise multiplication $\odot$.

In the fourth step, the output of each cell is calculated as

$$\vec{h}_t = \vec{o}_t \odot \tanh(C_t) , \tag{11}$$

with

$$\vec{o}_t = \sigma(W_o \vec{x}_t + U_o \vec{h}_{t-1} + \vec{b}_0) . \tag{12}$$

One LSTM layer consists of four W matrices, four U matrices, and four bias vectors $\vec{b}$, which makes for a total number or $4(mn + n^2 + n)$ free parameters to be trained.

6.1.3. Preprocessing

For the training of the NNs, the available simulated sample is split into a training sample, a validation sample used to evaluate the performance of the NN during the training, and a test sample used to evaluate the performance once the training is finished, with a ratio of 3:1:1.

The input variables are preprocessed to have similar magnitudes, which increases the speed and stability of the training process. The preprocessing uses scikit-learn's [121] RobustScaler and is done according to the distribution of the training sample. In a first step, all values of an input variable are shifted so that the mean of this variable in the training sample becomes zero. Afterwards, all values are scaled to

$$x_i \rightarrow \frac{x_i - q_{25}(x)}{q_{75}(x) - q_{25}(x)} , \tag{13}$$

where $q_{25}(x)$ and $q_{75}(x)$ are the value at the 25th and 75th percentile of the variable in the training sample.

The same preprocessing is applied to any sample before feeding it into a NN.

6.1.4. Training

During the training of a NN, the weights w_{ij} and b_j (or in case of LSTM layers all weights given by the elements of all matrices W and U and the bias vectors $\vec{b}$) are adjusted to minimize the loss function, which evaluates the classification done by the NN. In this thesis, the *binary cross entropy*

$$L\left(y_{\text{predicted}}|y_{\text{true}}\right) = -\left[y_{\text{true}} \log\left(y_{\text{predicted}}(w_{ij}, b_j)\right) + (1 - y_{\text{true}}) \log\left(1 - y_{\text{predicted}}(w_{ij}, b_j)\right)\right] \tag{14}$$

is chosen as a loss function as it is computationally efficient and provides sufficiently large gradients during the training [122]. Here, y_{true} is the jet label in the simulation (0 for d-jets, i.e. background jets from d quarks, and 1 for s-jets) and $y_{\text{predicted}}$ is the label predicted by the NN.

The adjustment of the weights is done by an optimizer algorithm, namely the *ADAM algorithm* [123], which is based on stochastic gradient descent. One iteration of the algorithm is done using a subset of all training events, a so-called *batch*. First, the gradient of the loss function, i.e. a vector of all partial derivatives w.r.t. all weights, is calculated for each event in the batch and then averaged over these events. Second, this gradient is used to update *moving moment estimators* which average the mean and variance of the gradient over all iterations,

using decay rates β_1 and β_2 to slowly move the moment estimators away from the initial values and towards the terminal values. Third, based on these moving averages, the weights are updated, the step size α moderating the change between the old weights and the new weights in one iteration. Additionally, in one iteration of the optimizer, the step size α is changed in order to slowly decrease. The iteration over all events in the entire training sample is called an *epoch*.

All parameters defining the model that are set before the training are called *hyperparameters*. For the NNs trained in this chapter, the hyperparameters assumed to have the largest impact are the width and the number of hidden layers of the NNs. Additional hyperparameters are the batch size, the initial step size α_{initial} (also called learning rate), and the exponential decay rates β_1 and β_2.

Once a NN configuration is found which has a capacity that is large enough to model the differences between the two classes it is supposed to distinguish, it is necessary to ensure that the effective capacity is not too large and that no features from the training sample are learned that cannot be generalized (*overfitting*). This can be done using a mechanism called *regularization*.

There are several possible regularization techniques. The two highlighted here are the L1 and L2 regularization, which both penalize large connection weights w_{ij} between layers. The loss function of a regularized NN is

$$L_{\text{regularized}} = L(w_{ij}, b_j) + \lambda \sum_{i,j} |w_{ij}|^a, \tag{15}$$

where j are all nodes in the one layer and i are the nodes in the adjacent layer and $L(w_{ij}, b_j)$ is the loss function without regularization. $a = 1$ for L1 regularization and $a = 2$ for L2 regularization, while λ is the regularization strength that can be adjusted. The difference between both types of regularization can be found when considering different magnitudes for the weights w_{ij}. If a weight is small ($w_{ij} < 1.0$), the penalty term in L1 regularization is larger than in L2 regularization. This means L1 regularization can be used to further reduce the impact of already small weights [124]. If a weight is relatively large ($w_{ij} > 1.0$), the penalty term in L2 regularization is larger than in L1 regularization. The quadratic behavior causes large penalty terms for large weights, therefore reducing the importance of single weights and features. For this reason, L2 regularization is applied to reduce overfitting.

6.1.5. Evaluation

In order to evaluate the classification power of the trained NNs, so-called receiver operating characteristic (ROC) curves are used. An example can be found in Figure 6.4. It illustrates the relation between the signal efficiency $\varepsilon_{\text{signal}}$ and background mis-tag rate $\varepsilon_{\text{background}}$ when applying a selection to a dataset by requiring an NN output score larger than a given value. In the chosen representation, the x-axis shows the signal efficiency, while the y-axis shows $1 - \varepsilon_{\text{bkg}}$, both corresponding to the fraction of correctly classified events of the signal and background class. The area under the curve (AUC) is employed as a metric of the NN's performance during the training and optimization process. It is calculated using scikit-learn's [121] ROC AUC score and takes on a value between 0.5 (no separation) and 1.0 (perfect separation).

For better visualization, at times, *modified* ROC curves, which show the rejection rate $1/\varepsilon_{\text{bkg}}$ on the y-axis, are used. However, if an AUC value is quoted, it always refers to the AUC of the standard ROC curve.
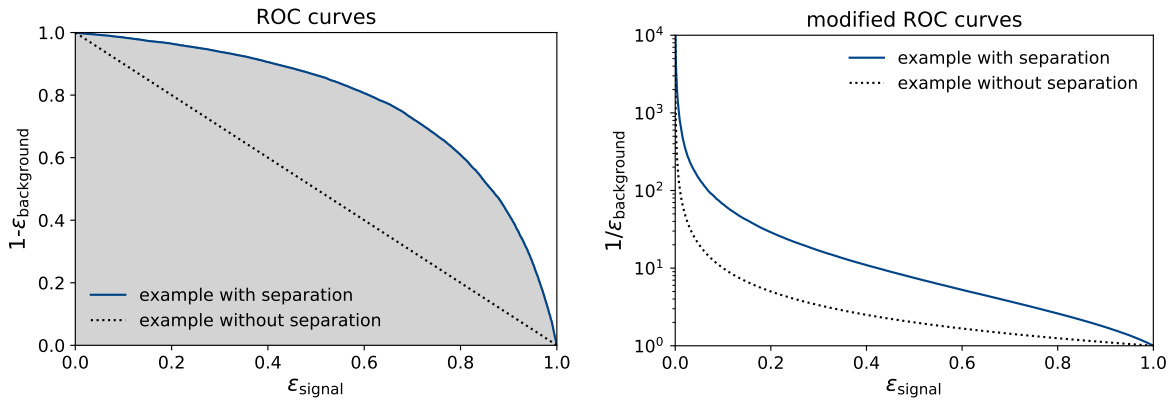
Fig. 6.4: Examples of receiver operating characteristic (ROC) curves and modified ROC curves, illustrating the relation between the efficiency $\varepsilon_{\mathrm{signal}}$ of signal and mis-tag rate $\varepsilon_{\mathrm{background}}$ of background when applying a selection to a dataset by requiring an output score larger than a given value. The grey area represents the area under the ROC curve as it is calculated in the AUC score. The shown blue curves with separation have the same underlying distribution separating signal and background.

To evaluate the performance of the NNs, the epoch with the best AUC value on the validation sample is used. The statistical uncertainty associated with the AUC score calculated on the validation or test sample is estimated as the standard deviation of the AUC score calculated by considering 100 samples of the same size as the sample used for the evaluation. These 100 samples are each sampled from the initial evaluation sample using replacement, a process called *bootstrapping*.

In some instances, the accuracy, which is defined as the fraction of correctly classified events, an *s*-jet efficiency or a *d*-jet rejection rate are quoted in this thesis. If not mentioned otherwise, these values are defined by requiring an NN output score larger than 0.5.

6.2. Considerations on How to Approach *s*-Tagging

In this section, first, basic notions on the identification of *s*-jets are presented. Following this, a study on the maximally achievable separation between *s*-jets and *d*-jets at hadron colliders for a variety of ideal detector scenarios is presented. This study uses simulated particles as input to recurrent neural networks and can be found in Reference [125][2].

Using *b*- and *c*-tagger such as those introduced in Sections 4.5 and 4.6, it is possible to distinguish *s*-jets from *b*- and *c*-jets. With the help of so-called quark-gluon taggers (cf. References [126–138]), it is in principle possible to distinguish between *s*-jets and jets from gluons. The distinction between jets from the lighter quarks (*u*, *d*, and *s*), on the other hand, remains challenging because they are very similar in their hadronization patterns.

In this thesis, only the separation between *s*- and *d*-jets is considered as it is most relevant in the determination of the CKM matrix element $|V_{ts}|$ and $|V_{td}|$ in $t \to W^+ s$ and $t \to W^+ d$

---

[2]This paper was written together with Johannes Erdmann and Olaf Nackenhorst, who supported me as supervisors and editors.

decays. Additionally, s and d quarks carry the same electric charge of $-\frac{1}{3}e$, making them in a way more similar to each other than s and u quarks, the latter carrying an electric charge of $+\frac{2}{3}e$. Hence, it is acceptable to only consider s- and d-jets for studies of the maximally achievable separation.

The section is structured as follows: In Subsection 6.2.1, the MC samples produced for studies on s-tagging performance independent of the ATLAS Experiment are described. Subsection 6.2.2 illustrates properties of s-jets that can be used in the distinction of s- and d-jets and in Subsection 6.2.3, the studies on the estimates of the maximum performance of s-tagging are presented.

6.2.1. Jet Samples Without Detector Simulation

The MC samples used for the studies presented in Section 6.2 contain diquark production in pp collisions at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$. The ME of the process is simulated with MadGraph_aMC@NLO version 2.6.7 [139] at LO in α_S using the NNPDF2.3LO PDF set [78]. The parton shower and hadronization is simulated with PYTHIA8.2.35 [77] and no detector simulation is applied[3].

All particles that are an output of Pythia and have a lifetime $\tau > 0$ are clustered with the anti-k_T algorithm [59] with a jet radius parameter of $R = 0.4$ using FastJet [141]. To determine the flavor of the jets, *ghost matching* [142] is used[4]: The s and d quark from the ME generation are included as input to the clustering; if one of these quarks is clustered into a jet, the jet's flavor is determined by the corresponding quark, and the quark is removed from the list of jet constituents.
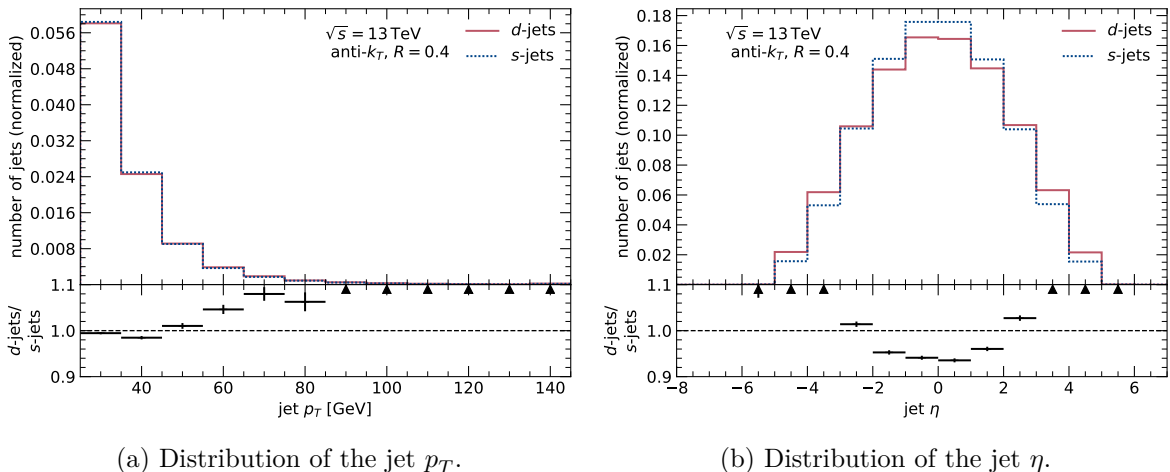

(a) Distribution of the jet p_T. (b) Distribution of the jet η.

Fig. 6.5: Kinematic distributions of s- and d-jets before their reweighting. The lower panel shows the ratio of the two distributions.

The jets' four-momenta are calculated as the sum of the four-vectors of their constituents which have their origin in the primary vertex. After requiring all jets to have a $p_T > 25\,\text{GeV}$,

---

[3]The samples used were generated by Nils Julius Abicht [140].

[4]Ghost matching is a procedure in which "infinitely soft" (and therefore ghost-like) particles are included as input to the clustering input. As they do not carry momentum, they do not change the properties of the resulting jets but can be used to determine e.g. the jets' flavor or dimensions [142].

2.2 million *s*- and *d*-jets are available in the study. Figure 6.5a shows the p_T distribution of *s*- and *d*-jets. It decreases for increasing p_T as expected for jets from diquark production. The difference between the distributions is small. The jet η distribution shown in Figure 6.5b, on the other hand, shows a significant difference between *s*-jets and *d*-jets, with *s*-jets being produced more centrally. To remove these kinematic differences, which should not be learned by the *s*-tagger, all *d*-jets are reweighted in their p_T and η distributions to match the distributions of *s*-jets.

6.2.2. Characteristics of *s*-Jets

Due to the different flavors of the initial quarks, the primary difference between *s*-jets and *d*-jets is their hadron content. One expects the most energetic (*leading*) hadron in an *s*-jet to contain an *s* quark and the leading hadron in a *d*-jet to contain a *d* quark. Figure 6.6 shows the "mass spectrum" of the leading hadrons inside of *s*- and *d*-jets. The leading hadron in *s*-jets tends to be a kaon, either electrically charged ($K^{\pm}$) or electrically neutral (K_S or K_L). In a *d*-jet, the leading hadron tends to be a pion, either electrically charged ($\pi^{\pm}$) or electrically neutral (π^0). In comparison to *s*-jets, the number of *d*-jets with a proton *p* or neutron *n* as leading particle is significantly larger as well; however, their overall number is small.
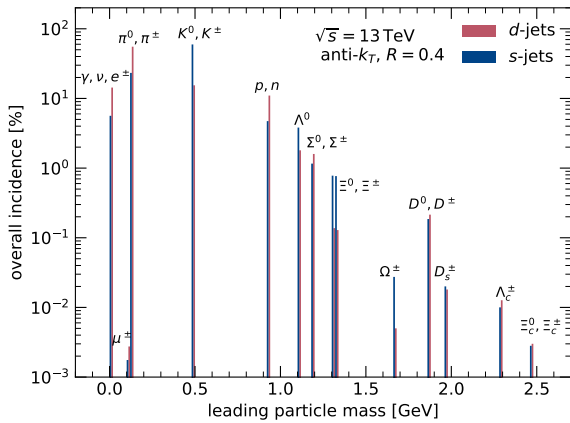

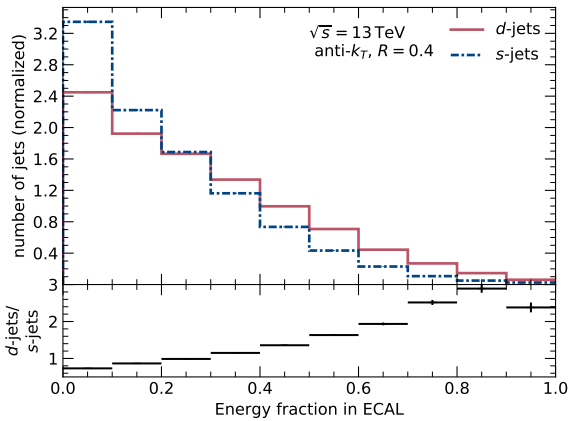Fig. 6.6: Mass spectrum of the jets' constituent particles leading in energy for *s*- and *d*-jets.

Fig. 6.7: Distribution of the fraction of jet energy carried by electrons and photons at a distance of $r = 1\,\text{m}$ from the primary vertex for *s*-jets and *d*-jets as approximation of what is detected in the electromagnetic calorimeter (ECAL). The lower panel shows the ratio of the two distributions.

Considering these differences, four approaches to *s*-tagging can be formulated:

1. As they belong to the most prevalent leading particles, the identification of $\pi^{\pm}$ and $K^{\pm}$ can be used to distinguish *d*- and *s*-jets, respectively. This approach has been taken at the SLD and Delphi Experiment [114, 115]. However, it relies on the use of Cherenkov detectors that are not available or foreseen for multi-purpose collider detectors such as ATLAS and CMS.

2. It is possible to reconstruct $K_S \rightarrow \pi^+\pi^-$ decays from tracks measured in a tracking detector, cf. e.g. the description of such a reconstruction in Section 4.9. Reconstructed K_S can then be used to identify s-jets. However, only approximately 25% of all kaons produced in jets are K_S, and only 69% decay to $\pi^+\pi^-$ [24].

3. K_L, however, have a lifetime that is too long ($\tau \approx 5 \cdot 10^{-8}$ s) to observe their decays with a collider detector at typical LHC energies. As neutral hadrons, they deposit most of their energy in the hadronic calorimeter (HCAL). π^0, on the other hand, are more prevalent in d-jets and decay prominently into two photons, depositing most of their energy in the electromagnetic calorimeter (ECAL). Figure 6.7 shows the fraction of energy carried by either photons or electrons for s-jets and d-jets. As expected, it tends to be larger for d-jets than for s-jets. Thus, it is possible to use the fraction of energy deposited in the ECAL to distinguish s-jets and d-jets.

Besides the flavor of the leading hadron, a number of different jet properties can be considered for the distinction of jet types as well. Figure 6.8a e.g. shows distribution of the p_T of the leading constituent particle. For s-jets, the leading particle tends to carry slightly more momentum than in d-jets, as kaons produced in pp collisions tend to have larger transverse momenta than pions, cf. the fragmentation functions determined at a center-of-mass energy of 13 TeV in Reference [143].

Differences can also be seen in overall jet properties. Figure 6.8b illustrates the jet width defined as

$$\text{width} = \frac{\sum\limits_i p_\mathrm{T}^i \, \Delta R(i, \text{jet})}{\sum\limits_i p_\mathrm{T}^i} \,, \tag{16}$$

where i runs over all constituent particles and

$$\Delta R(i, \text{jet}) = \sqrt{\left(\phi_\text{jet} - \phi_i\right)^2 + \left(\eta_\text{jet} - \eta_i\right)^2} \tag{17}$$

is the distance in the plane spanned by the azimuthal angle ϕ perpendicular to the beam axis and the pseudorapidity η. It illustrates that s-jets tend to be narrower than d-jets, offering another handle to distinguish between s-jets and d-jets.

6.2.3. Estimation of the Maximally Achievable s-Tagging Performance

The performance of any s-tagger will depend on the detector used to record the jets and on the information available to the classification algorithm. When developing an s-tagger for a real collision experiment, it is useful to have an estimate of how much separation between s- and d-jets is in principle achievable. If this estimate is derived as an upper bound, this can be done without considering effects such as detector resolution or noise contamination, which would degrade the separation performance. Applying no specific detector simulation (such as a simulation of the ATLAS detector) to the simulated samples used to derive estimates of the performance allows for a cross-experimental interpretation.

Nonetheless, the principles of particle detection have to be considered for a general picture of the achievable performance of s-taggers. Hence, in this section, six s-taggers are trained for six different types of detectors. The detectors are implemented by imposing detector-specific
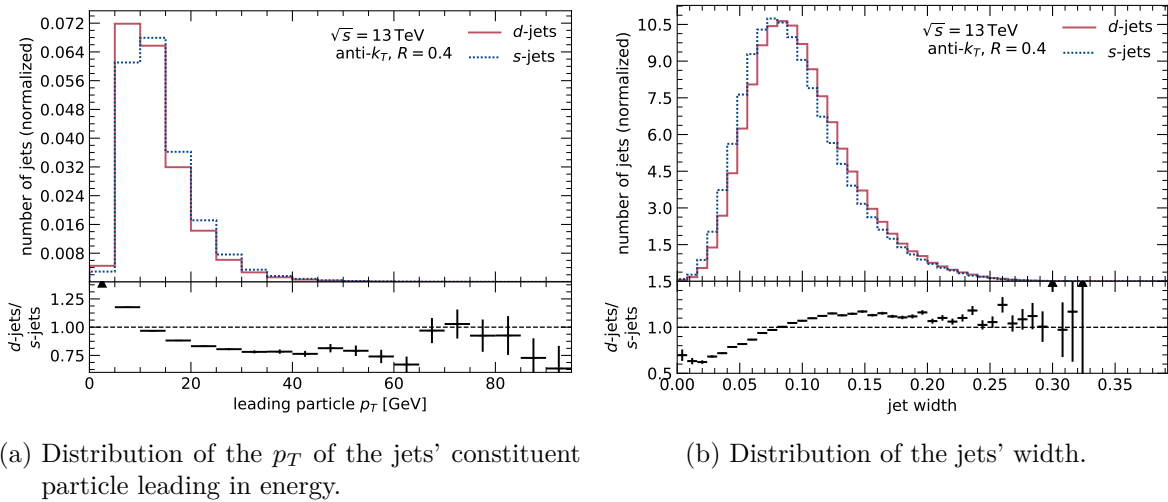
(a) Distribution of the p_T of the jets' constituent particle leading in energy.

(b) Distribution of the jets' width.

Fig. 6.8: Distributions characterizing *s*-jets and *d*-jets. The lower panel shows the ratio of the distributions for *s*-jets and *d*-jets.

selections on the jet constituents. To emulate e.g. a tracking detector, only charged constituent particles are used as input to an *s*-tagger.

In the following, all considered detector scenarios are briefly described. In addition, Table 6.1 shows the selection criteria imposed on the jets' constituent particles and the variables used in the training of the *s*-taggers.

Universal Collider Detector   To estimate the maximum upper bound on the performance of an *s*-tagger, an ideal detector is considered which is able to detect all types of particles created in *pp* collisions with perfect accuracy. This detector (called a perfect *universal collider detector*) can identify particles without ambiguity and measure their kinematic properties with infinite precision. In this detector scenario, no selection is imposed on the jet constituents used as input to the *s*-tagger.

All spatial and kinematic particle properties are expressed in a coordinate system given by (r, ϕ, η), where r is the radial distance from the collision vertex, ϕ is the azimuthal angle in the plane perpendicular to the beam axis, and η is the pseudorapidity. Nine variables are chosen as particle level input features to fully describe the particles. The particles' types[5] and kinematics are given by the four-momentum consisting of the energy E, the pseudorapidity η, the azimuthal angle ϕ, and the mass m. The trajectories of the particles are fully described by the point of creation (r_0, η_0, ϕ_0), the aforementioned four-momentum, and the lifetime τ as measured in the laboratory system. Lastly, the charge q is used as an input feature as well.

Optimistic Collider Detector   When considering a specific type of detector, in the simplified scenarios, it is necessary to make some assumptions. The typical multi-purpose collider detector e.g. found at the LHC is approximately cylindrical and consists of a tracking detector

---

[5]Given that in the simulation used, the mass of particles does not follow a Breit-Wigner distribution but is given as one central value, it is possible to identify almost all particles by their mass and charge. An exceptions to this are neutrinos and neutral hadrons and their antiparticles.

Name	Selection Criteria	Input Variables
universal detector	$\tau > 0$	E, η, ϕ, m (4-momentum), r_0, η_0, ϕ_0 (origin), q (charge), τ (lifetime in lab system)
optimistic collider detector	$\tau > 0$, r_f(charged particle) $> 10\,\mathrm{mm}$, r_f(neutral particle) $> 1\,\mathrm{m}$, no ν	E, η, ϕ, (4-momentum minus mass), r_i, η_i, ϕ_i (initial measurement), q (charge)
tracking detector	$\tau > 0$, $r_f > 10\,\mathrm{mm}$, $q \neq 0$ $r_0 < 1\,\mathrm{m}$	p, η, ϕ (4-momentum minus mass), r_i, η_i, ϕ_i (initial measurement), τ (lifetime in lab system), q (charge)
Cherenkov detector	$\tau > 0$, $r_f > 10\,\mathrm{mm}$, $q \neq 0$, $r_0 < 1\,\mathrm{m}$	p, η, ϕ, m (4-momentum), r_i, η_i, ϕ_i (initial measurement), τ (lifetime in lab system), q (charge)
calorimeter without ECAL/HCAL separation	$\tau > 0$, $r_0 < 1\,\mathrm{m}$, $r_f > 1\,\mathrm{m}$, no ν	E, η, ϕ (3-momentum)
calorimeter with ECAL/HCAL separation	$\tau > 0$, $r_0 < 1\,\mathrm{m}$, $r_f > 1\,\mathrm{m}$, no ν	E, η, ϕ (3-momentum), particle category (γ/e, μ, other)

Tab. 6.1: List of all considered ideal detector scenarios. The second column shows the selection requirements imposed on the constituent particles used as input to the neural networks, where τ is the lifetime of the particles, r_0 is the radial distance between the primary vertex and the point where the particle is created, and r_f is the radial distance between the primary vertex and the decay vertex. The third column describes the variables that are used as input features to the neural network. If the variable carries a subscript 0, it refers to the space point of creation, and if it carries a subscript i, it refers to the space point of initial measurement.

immediately surrounding the beam pipe, a calorimeter positioned around the tracking detector, and a muon system surrounding the calorimeter. This means that charged particles can only be detected once they reach the tracking detector, neutral particles can only be detected once they reach the calorimeter, and neutrinos cannot be detected at all as the probability for them interacting with the detector is negligible. At the LHC, the radius of the innermost instrumentation is $8\,\mathrm{mm}$ at the LHCb detector [42], $29\,\mathrm{mm}$ at the CMS detector [144], $33.25\,\mathrm{mm}$ at the ATLAS detector [46], and $39\,\mathrm{mm}$ at the ALICE detector [41]. The calorimeters are positioned at a radius of $2.56\,\mathrm{m}$ for the ATLAS detector [44], $6\,\mathrm{m}$ for the CMS detector [43], and $4.6\,\mathrm{m}$ for the ALICE detector [145]. Approximating these scales in a conservative way, in the following, if a selection emulating a tracking detector is applied, this tracking detector is assumed to cover a cylindrical volume with a radius of $10\,\mathrm{mm} < r < 1\,\mathrm{m}$, while calorimeters

are assumed to be positioned at a distance of $r = 1\,\mathrm{m}$ from the collision vertex.

In the *optimistic collider detector* scenario used to train a second s-tagger, the combined use of tracking detectors and calorimeters as found at multi-purpose collider detectors is emulated accordingly. This scenario is based on the universal collider detector scenario. However, all electrically charged particles decaying before they reach $r = 10\,\mathrm{mm}$ and all electrically neutral particles decaying before reaching $r = 1\,\mathrm{m}$ are removed from the list of input particles. For all remaining particles, instead of the point of creation, the point of initial detection is used as input feature and the lifetime is removed from the input features as it is typically difficult to reconstruct. This scenario constitutes an optimistic collider detector as it neglects limitations given by the detector structure (such as layers in the tracking detector), reconstruction and identification inefficiencies, and resolution effects.

Tracking Detector Tracking detectors are used to reconstruct the trajectories (*tracks*) of charged particles. They detect a signal (*hit*) when a charged particle passes through one of its components; these components are geometrically arranged in such a way that it is possible to reconstruct a three-dimensional trajectory by connecting these hits. If a tracking detector is embedded in a known magnetic field, it is possible to infer the electric charge and the momentum of the particle from the curvature of the trajectory.

To emulate the characteristic detection abilities of a tracking detector in a third scenario, only charged constituent particles are used as input of the s-tagger. All tracks are assumed to be perfectly measured, i.e. there is neither a limit to spatial resolution, nor noise, nor any inefficiency in the track reconstruction. The particles' three momenta (p, η, ϕ), the points of initial detection (r_i, η_i, ϕ_i), the sign of their charge q – which is either $+1$ or -1 –, and the time τ it takes the particles to traverse the detector are used as input. In case the particle leaves the tracking detector before decaying, a large default value is used for τ. As the tracking detector is assumed to cover $10\,\mathrm{mm} < r < 1\,\mathrm{m}$, all particles that decay before reaching $r = 10\,\mathrm{mm}$ or that are created at $r > 1\,\mathrm{m}$ are removed from the input.

Because tracks of opposite-sign electrically charged particles can be used to reconstruct K_S and Λ^0, in a separate scenario, the use of these reconstructed strange hadron is emulated by replacing two particles of opposite-sign charge with an origin at the same space-time point by a neutral particle. The neutral particle's kinematic properties as well as its mass are determined by adding the four-vectors of the two charged particles.

Cherenkov Detector If the momentum of a charged particle is measured – e.g. by reconstructing its curved trajectory in a known magnetic field – Cherenkov detectors can be used to calculate its mass from the angle of its Cherenkov radiation w.r.t. its axis of momentum. At both the SLD Experiment at SLAC and Delphi Experiment at LEP, this was used to identify $K^{\pm}$ in order to identify s-jets [114, 115]. The use of a Cherenkov detector for particle identification is limited for large particle momenta and small particle masses as the angles of the Cherenkov radiation become more similar. However, since in all considered detector scenarios, resolution effects are neglected, the Cherenkov detector scenario is simply an extension of the tracking detector scenario in which the particle's mass is used as an additional input feature.

Calorimeters With and Without Separation Into ECAL and HCAL Components
Calorimeters are used to determine the energy of both electrically neutral and electrically charged particles except for neutrinos. From a practical point of view, additional detector

components are necessary for the determination of the energy or momentum of muons because they tend to be minimally ionizing at energies prevalent in TeV hadron collisions and therefore do not deposit all of their energy in the calorimeters.

Particles deposit their energy inside calorimeters by producing particle showers. These showers are grouped into two categories: Electromagnetic showers initialized by a photon or electron and hadronic showers initialized by hadrons. Most calorimeters are divided into two successive layers, an ECAL optimized for the detection of electromagnetic showers and an HCAL optimized to absorb and measure the energy of hadrons. This separation can be used for *s*-tagging as π^0 are more prevalent in *d*-jets, decay in approximately 99% of all decays into two γ, and have a relatively short lifetime of approximately $0.085\,\mathrm{fs}$. Thus, on average, the energy fraction carried by γ is larger for *d*-jets than for *s*-jets, as can be seen in Figure 6.7, showing the fraction of energy per jet carried by photons and electrons as measured at a radius of $r = 1\,\mathrm{m}$.

Besides the fraction of energy measured in the ECAL, the geometrical distribution of energy inside the jet cone (*substructure*) can provide information to discriminate between *s*- and *d*-jets.

In order to consider both types of information, two different calorimeter scenarios are considered. Both can determine the energy and position of all particles (except neutrinos) in the η–ϕ plane at a radius of $r = 1\,\mathrm{m}$ perfectly. The first calorimeter scenario has no separation between an ECAL and HCAL and therefore can only use the substructure of the jet to discriminate between *s*- and *d*-jets. This means that the *s*-tagger only uses E, η, and ϕ as input features. The second calorimeter scenario has a separation into an ECAL and an HCAL. This separation is implemented by using an additional input variable for the *s*-tagger which encodes the type of the particle detected: *1* for electrons and photons, *2* for muons, and *0* for all other particles, foremost hadrons but also tau leptons.

The performance of both calorimeter scenarios is compared against the separation provided by the fraction of jet energy that would be deposited in the ECAL, cf. Figure 6.7. This distribution covers a range of $[0.0, 1.0]$ similar to the output range of the NNs trained as *s*-taggers, but on average, *d*-jets have larger values than *s*-jets. Hence, to determine a ROC curve, the distribution of one minus the fraction of energy in the ECAL is used.

6.2.3.1. Implementation

The *s*-tagging algorithms in this section are implemented using NNs containing LSTM layers as they are able to build a model based on variable numbers of input particles and to relate information provided both close to each other and far-apart in the particle input.

The distribution of the number of constituent particles for *s*- and *d*-jets as well as their p_T dependency is shown in Figure 6.9. Most jets contain less than 30 constituent particles. Their number is similar for *s*- and *d*-jets and increases slightly for larger jet p_T.

For the training, all jet constituents are ordered according to their energy (largest to smallest). For computational reasons, the number of particles considered in the training has to be limited. Therefore, if a jet contains more than 50 particles, the additional particles with the smallest energies are removed. If a jet contains less than 50 particles, a mask is applied to the input, effectively removing all empty input particles from the training. To check if the particle ordering has an impact on the performance, an alternative ordering by the radius of the particles' origin is considered for the universal detector scenario.

Prior to the training of the NNs, the available simulated jets are divided into a training sample, a validation sample, and a test sample and all input features are preprocessed as
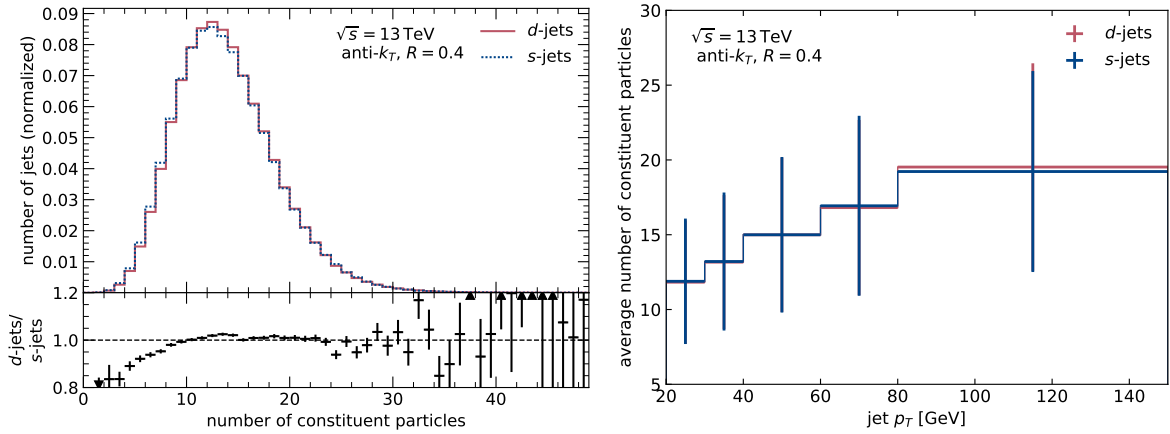
(a) Distribution of the number of constituent parti-
cles for *s*- and *d*-jets. The lower panel shows the
ratio of the distributions.

(b) Jet p_T dependency of the number of con-
stituent particles. The error bars show the
standard deviation of the distribution and not
the uncertainty on the mean.

Fig. 6.9: Number of constituent particles per jet for *s*- and *d*-jets.
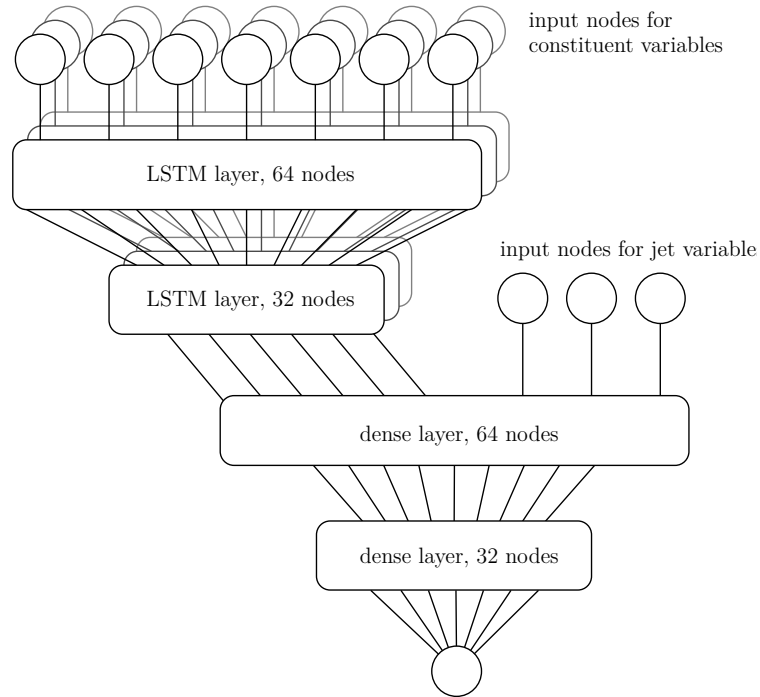

Fig. 6.10: Visualization of the chosen neural network architecture. Circles represent input
and output nodes and boxes represent both LSTM and dense layers. The number of
connections between layers represented as lines does not represent the actual number
of connections.

described in Section 6.1.3. For the training of all NNs, an initial learning rate of 0.001 and a batch size of 1024 jets is chosen.

Figure 6.10 illustrates the NN structure used for all detector scenarios. It combines both LSTM and feed-forward *dense* layers and has two input layers. The first input layer has as many input nodes as there are input features associated with the jets' constituent particles. In the second input layer, the NN is additionally provided with the jets' p_T, η, and ϕ.

The particle level input layer feeds into a set of LSTM layers. Their output – together with the input from the second input layer containing the jet level input features – then feeds into a set of feed-forward dense layers. Their output in turn feeds into a single output node.

The structure was optimized using the input of the universal detector scenario as this scenario is the one with the most complex input. Both the number of layers as well as the number of nodes in the layers are varied. To keep the variation as simple as possible, each hidden NN layer contains half as many nodes as the previous layer.
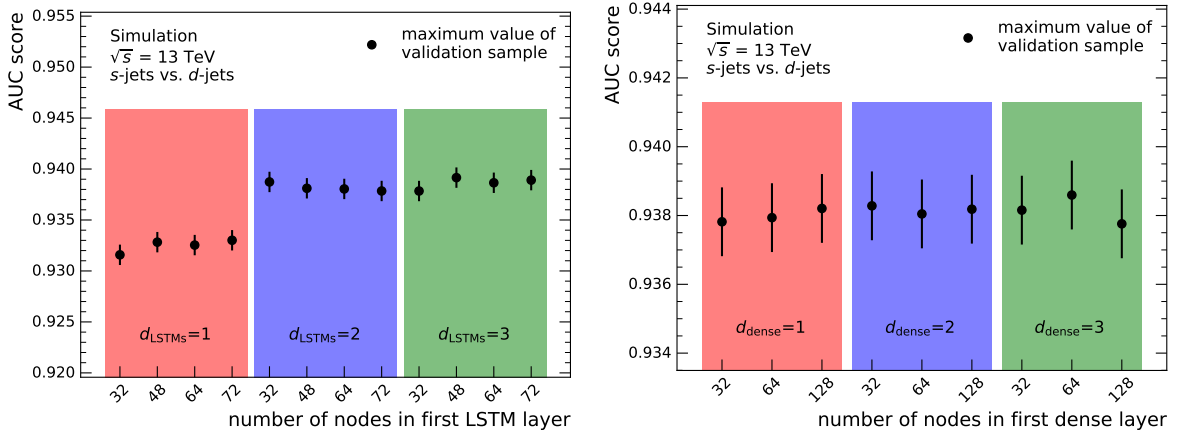
(a) Comparison of the AUC scores when varying the number of LSTM layers (d_{LSTMs}) and the number of nodes in the LSTM layers. The number of nodes shown is the number of nodes in the first LSTM layer; all succeeding LSTM layers contain half as many nodes as the previous layer. The structure of the dense layers is the structure in the final setup and is not varied for the comparison shown.

(b) Comparison of the AUC scores when varying the number of dense layers (d_{dense}) and the number of nodes in the dense layers. The number of nodes shown is the number of nodes in the first dense layer; all succeeding dense layers contain half as many nodes as the previous layer. The structure of the LSTM layers is the structure in the final setup and is not varied for the comparison shown.

Fig. 6.11: Comparison of the AUC scores of neural networks trained using the input of the universal detector scenario and varying the neural network structure. The AUC scores are determined on the validation sample and the uncertainties are the statistical uncertainty.

Figure 6.11 compares the AUC scores achieved for the universal detector scenario with different NN structures. In Figure 6.11a, the number of LSTM layers and the number of nodes therein are varied, while the structure of the dense layers is not varied. It can be seen that the classification performance increases when the number of LSTM layers is increased from one to two, but the performance does not improve by adding a third LSTM layer. The variation of the number of nodes in the LSTM layers does not have a significant impact. Figure 6.11b shows a similar variation of the structure of dense layers while retaining the structure of the

LSTM part of the NN. Both the variation of the number of dense layers as well as their width has no impact on the separation performance.

Following this, a structure containing two LSTM layers with 64 and 32 nodes and two feed-forward layers with 64 and 32 nodes, respectively, was chosen.

The structure of the NNs chosen, however, has a capacity that is too large for the classification problem at hand, which means that the NNs show signs of overfitting during the initial training without regularization applied. In Figure 6.12, e.g. an increasing difference between the AUC score of the training and validation sample during the training can be observed.
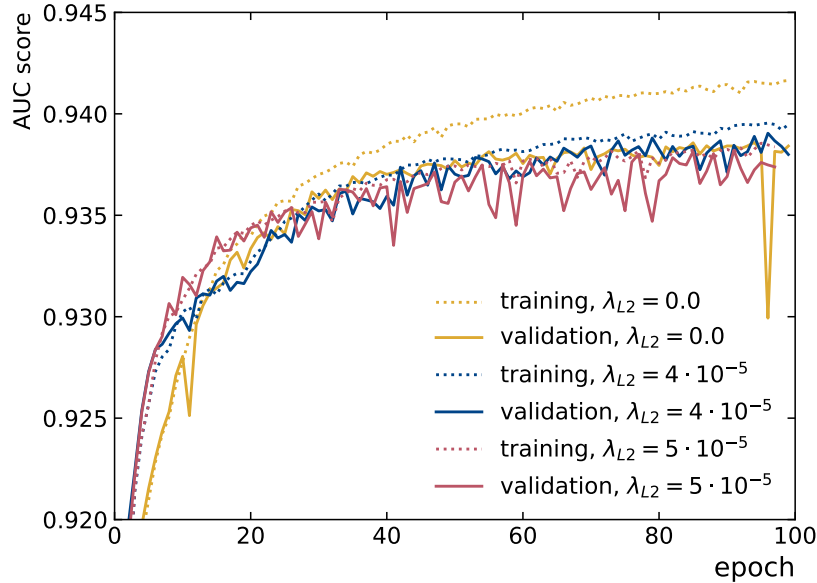
Fig. 6.12: Comparison of the AUC scores of the neural network of the universal detector scenario during the training as evaluated on the training and validation sample. For this particular plot – due to technicalities in the training – the AUC value is calculated without considering the weights of the jets.

In a first step, to reduce the overfitting, the training of the NNs is always stopped once the AUC as evaluated on the validation sample does not increase by more than 0.0001 for 20 consecutive epochs.

Furthermore, to reduce the capacity of the NNs, L_2 regularization is applied in each layer. For all layers, the regularization strength parameter λ_{L2} is set to the same value which is optimized for each detector scenario. The aim of this optimization is to reduce the difference between the output for the training and the test sample without reducing the AUC score as evaluated on the validation sample. Figure 6.12 illustrates this for the universal detector scenario: When applying a regularization strength of $\lambda_{L2} = 4 \cdot 10^{-5}$, the increasing difference between the AUC score for the training and the validation sample during the training is significantly reduced. Increasing the regularization strength further to $\lambda_{L2} = 5 \cdot 10^{-5}$, however, decreases the AUC value for both the training and the validation sample. Therefore, the regularization strength $\lambda_{L2} = 4 \cdot 10^{-5}$ is used.

The regularization strength used for each scenario is given in Table 6.2.

Scenario	Regularization Strength
universal collider detector	$4 \cdot 10^{-5}$
typical collider detector	$3 \cdot 10^{-5}$
tracking detector	$3 \cdot 10^{-5}$
tracking detector $+ K_S$ reconstruction	$5 \cdot 10^{-5}$
Cherenkov detector	$8 \cdot 10^{-5}$
calorimeter with separation	$1 \cdot 10^{-5}$
calorimeter without separation	$8 \cdot 10^{-6}$

Tab. 6.2: Detector scenarios and the strength of the L2 regularization applied to each layer.

6.2.3.2. Results

The following discussion of performances of the *s*-taggers trained for all different detector scenarios is divided into three parts: First, the universal collider detector and the optimistic collider detector are compared. Second, the tracking detector and the Cherenkov detector scenario are discussed. And third, the calorimeter scenarios are presented.

(a) Universal detector scenario.

(b) Optimistic detector scenario.
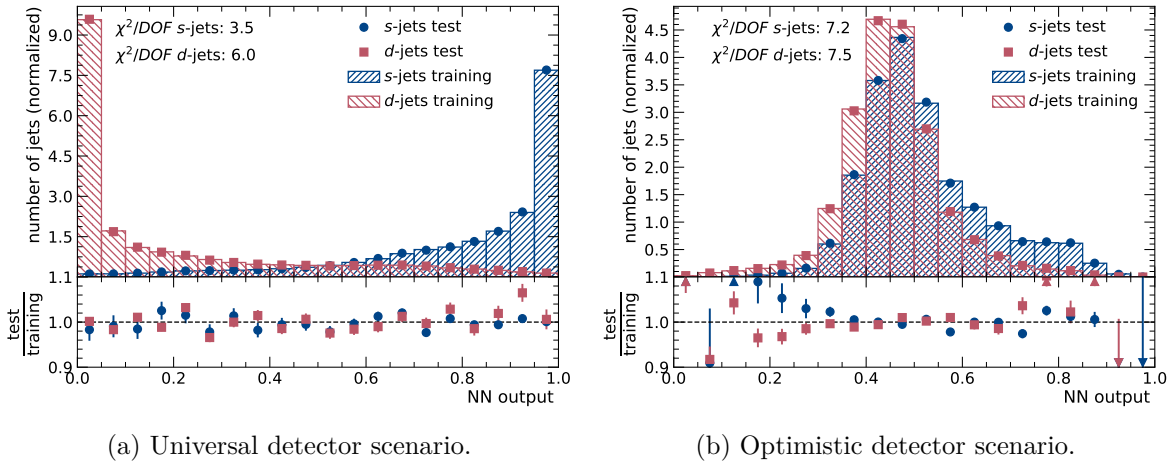
Fig. 6.13: Neural network output distributions for *s*- and *d*-jets in the training sample and the independent test sample (top panel) as well as their ratio (bottom panel).

Universal and Optimistic Collider Detector Figure 6.13a shows the output distribution of the NN in the universal collider detector scenario. The NN is able to separate *s*-jets and *d*-jets well: most *s*-jets receive a score closer to 1 and most *d*-jets receive a score closer to 0. To quantify overfitting, the output distributions of the training and test sample are compared for both jet flavors. For both *s*-jets and *d*-jets, the χ^2 per degrees of freedom (DOF) are calculated according to

$$\chi^2 = \sum_{i} \frac{(t_i - T_i)^2}{T_i} \,, \tag{18}$$

where t_i is the number of jets of the test sample in bin i and T_i is the number of jets of the training sample in bin i. The DOF are the number of bins of the output distributions shown.

While the values of $\chi^2/\text{DOF} = 3.5$ for s-jets and $\chi^2/\text{DOF} = 6.0$ for d-jets do not imply good agreement between both samples, the ratios of the output distribution for the training sample and the test sample do not show a trend, which means that after applying L2 regularization, no strong overfitting can be observed.

Figure 6.13b shows the output distribution of the NN trained with the input of the optimistic collider detector scenario. In this scenario, the separation between s-jets and d-jets is significantly reduced compared to the universal detector scenario as both jet flavors have an average output score that lies much closer to 0.5. The χ^2/DOF values of 7.2 for s- and 7.5 for d-jets are slightly larger than for the universal detector scenario and in the ratio between the training and the test sample, slightly larger deviations are visible for both jet flavors. However, these deviations mainly occur in those bins containing fewer jets.
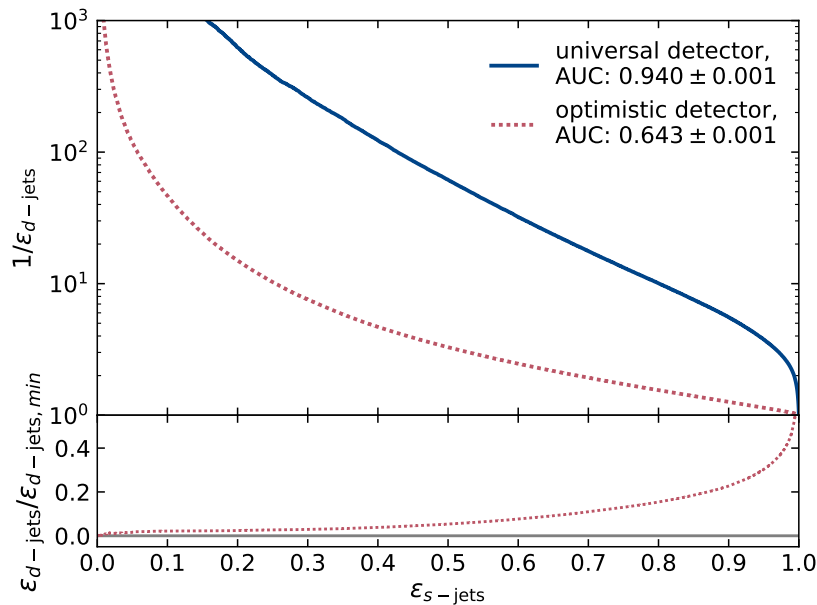
Fig. 6.14: Modified ROC curves illustrating the classification power of neural networks of the universal detector scenario and the optimistic detector scenario. The efficiencies are determined on the test sample and the uncertainties of the AUC scores are the statistical uncertainties. The ratio beneath the ROC curves shows the efficiency for d-jets in the optimistic collider detector scenario divided by the efficiency of d-jets in the universal detector scenario.

Figure 6.14 shows the modified ROC curves for the universal collider detector scenario as well as the optimistic collider detector scenario. The AUC score for the universal detector scenario is 0.940 ± 0.001, which means that if the type and kinematic properties of all constituent particles in a jet were detectable, very good separation between s- and d-jets could be achieved. However, in the optimistic collider detector that is an optimistic version of a typical multi-purpose detector, the separation is significantly reduced. It has an AUC score of 0.643 ± 0.001, which can be interpreted as a high upper limit on the achievable performance of an s-tagger at collider detectors such as ATLAS and CMS.

Figure 6.15a illustrates the p_T dependency of the separation in the universal detector scenario. The separation increases up to a p_T of $80\,\text{GeV}$ as more particles move into the jet cone

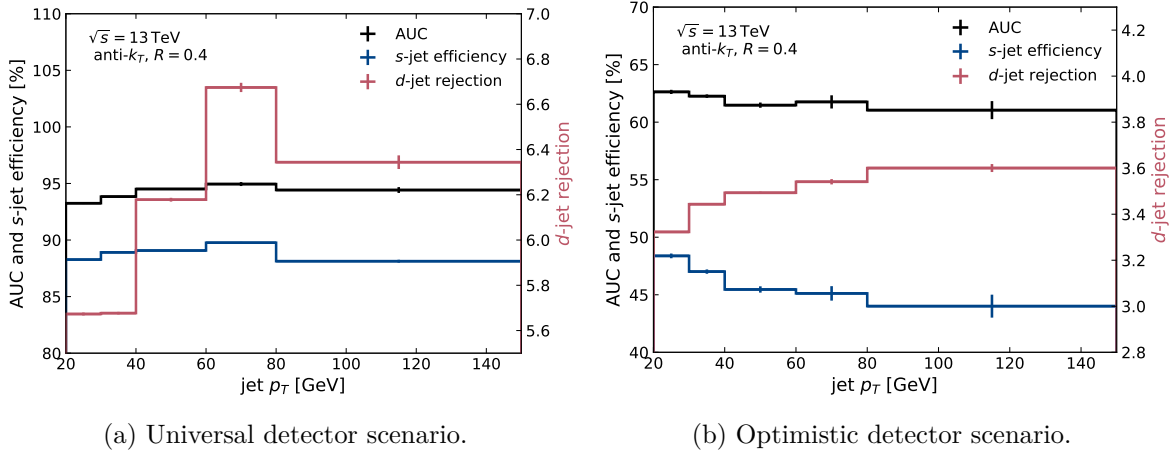(a) Universal detector scenario.

(b) Optimistic detector scenario.

Fig. 6.15: p_T dependencies of the AUC scores as well as the s-jet efficiencies and d-jet mis-tag rates requiring a classification value greater than 0.5.

because the initial quark becomes more boosted. For larger jet p_T, the separation decreases as the number of particles increases. Figure 6.15b shows the p_T dependency of the output of the optimistic detector scenario. In this scenario, the achievable separation decreases for increasing jet p_T as the number of particles in the jets increases.

(a) Tracking detector scenario.

(b) Cherenkov detector scenario.
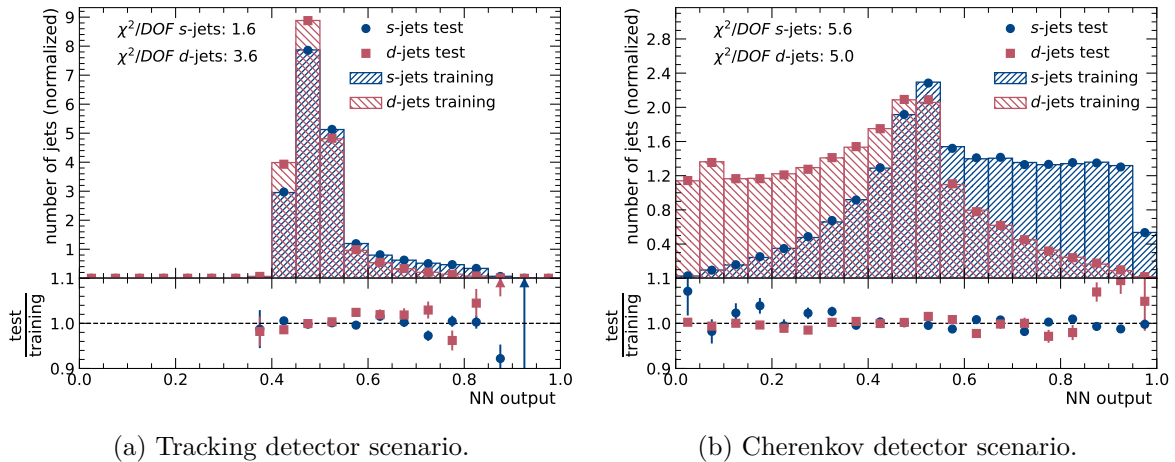
Fig. 6.16: Neural network output distributions for s- and d-jets in the training sample and the independent test sample (top panel) as well as their ratio (bottom panel).

Tracking and Cherenkov Detectors Figure 6.16a shows the output distribution of the NN trained with the input of the tracking scenario. The ratio between the training and test sample of the output distributions of s- and d-jets shows no significant trends, indicating little overfitting after applying the optimized $L2$ regularization. The output distributions for both s- and d-jets are narrow, have large overlap, and both peak around 0.5, which means that the separation is relatively small and significantly reduced w.r.t. what is e.g. seen in the optimistic detector scenario. In contrast, the NN output distributions of the Cherenkov

detector scenario shown in Figure 6.16b cover the entire range of NN output scores and show better separation than what is seen in the optimistic detector scenario. There is a small peak at values of approximately 0.5, which indicates that there are some jets that cannot be classified in this scenario. These jets are found to contain only low momentum ($p_T < 5\,\text{GeV}$) particles, suggesting that in these jets, the leading particle was either not electrically charged or not clustered into the jet.

Figure 6.17 shows the modified ROC curves of both scenarios as well as their AUC values. For the Cherenkov detector scenario, an AUC score of 0.783 ± 0.001 is reached, while the tracking detector has an AUC score of 0.574 ± 0.001. The plot additionally shows a scenario in which neutral particles decaying to two particles with opposite electrical charge are reconstructed before handing them to the neural network. This scenario has an AUC score of 0.572 ± 0.001, which means that the performances of this and the tracking detector scenario agree within the statistical uncertainty. Therefore, it can be concluded that the NN is able to reconstruct decays such as $K_S \to \pi^+ \pi^-$ without the use of an external reconstruction algorithm.
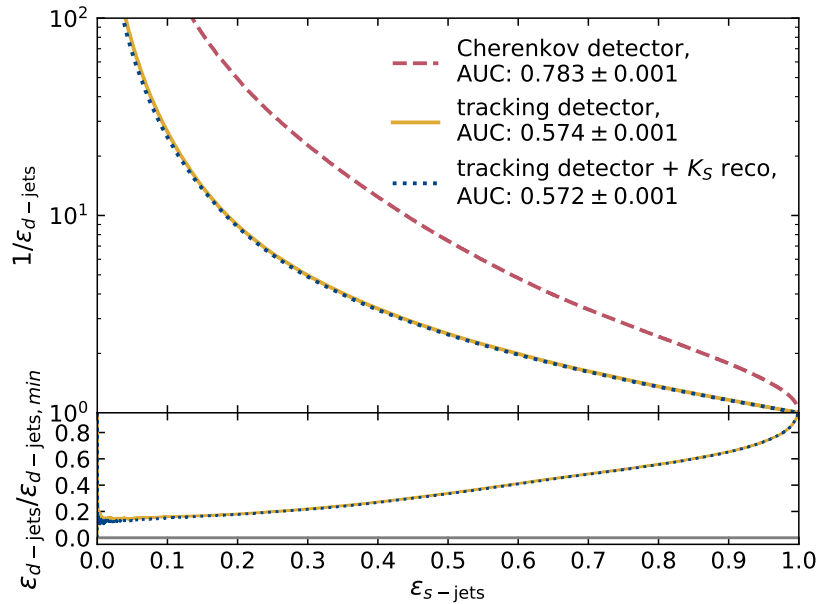
Fig. 6.17: Modified ROC curves illustrating the classification power of the neural networks of the tracking detector, the tracking detector with preceding K_S reconstruction, and the Cherenkov tracking detector. The efficiencies are determined on the test sample and the uncertainties of the AUC scores are the statistical uncertainties. The ratio beneath the ROC curves shows the efficiencies for d-jets for all illustrated scenarios divided by the efficiency for d-jets in the Cherenkov detector scenario.

Figure 6.18 illustrates the jet p_T dependency of the classification in the tracking detector scenario. The AUC score decreases continuously for increasing jet p_T as the number of particles in the jet increases. This behavior is expected to be amplified in a real detector as the p_T resolution of tracks worsens for larger p_T. A similar decrease in the classification power can also be assumed for the Cherenkov detector scenario as the quality of the particle identification decreases for increasing particle momenta.

The performance of the tracking detector scenario can be compared to the performance
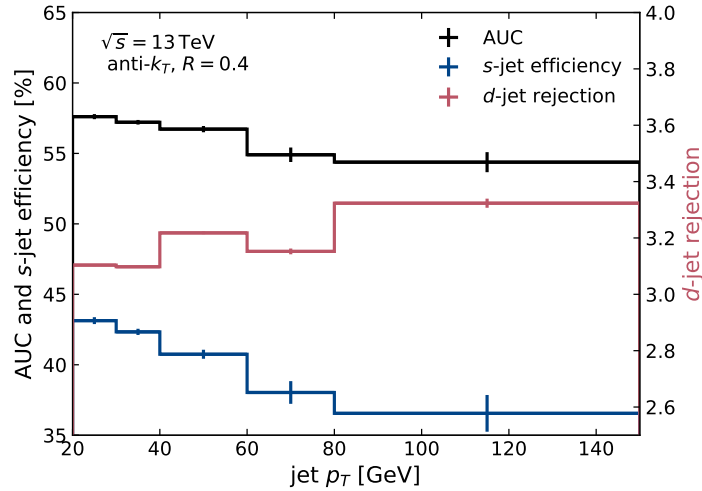
Fig. 6.18: p_T dependency of the AUC score as well as the s-jet efficiency and d-jet mis-tag rate requiring a classification value greater than 0.5 for the tracking detector scenario.

reached in Reference [30] in which a NN based on LSTMs is trained to separate s- and d-jets using tracks as input. For this, a simplified detector simulation using Delphes [146] and realistic pile-up conditions is used. A comparison of the d-jet rejection rates for fixed s-jet efficiencies is shown in Table 6.3. For this more realistic tracking detector scenario, for an s-jet selection efficiency of 30%, the rejection rate decreases by roughly 8% compared to the perfect tracking detector scenario, while for an s-jet selection efficiency of 70%, the rejection rate decreases by 2.5%. This suggests that the largest limitation to s-tagging based on tracks is the lack of information about the particles' type; detector effects such as reconstruction efficiencies, on the other hand, play a smaller role.

	s-Jet Efficiency	**d-Jet Rejection**
tracking detector scenario	30%	4.9
	70%	1.6
Reference [30]	30%	4.5
	70%	1.56

Tab. 6.3: Comparison of s-jet efficiencies and d-jet rejection rates of neural networks trained in the ideal tracking detector scenario and using a more realistic tracking detector simulation, cf. Reference [30].

Calorimeters Figure 6.19 shows the output distribution of the NN trained in the calorimeter scenario that includes a separation into ECAL and HCAL. No strong overfitting is observed since there is no trend visible in the ratio of the output distributions of the training and the test sample. The output distributions for both s- and d-jets have large overlap and peak around 0.5 but have a larger width than the output distributions of the tracking detector scenario.

Figure 6.20 illustrates the jet p_T dependency of the classification in the calorimeter scenario that includes a separation into ECAL and HCAL. The AUC score decreases only slightly for increasing jet p_T, suggesting that the use of a calorimeter with a separation into ECAL and
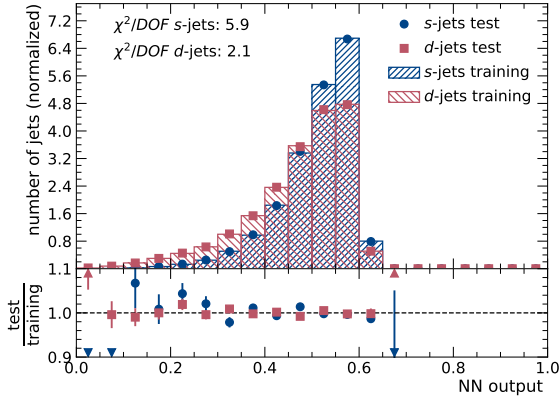
Fig. 6.19: Neural network output distributions of the calorimeter scenario with a separation into ECAL and HCAL for *s*- and *d*-jets in the training sample and the independent test sample (top panel) as well as their ratio (bottom panel).
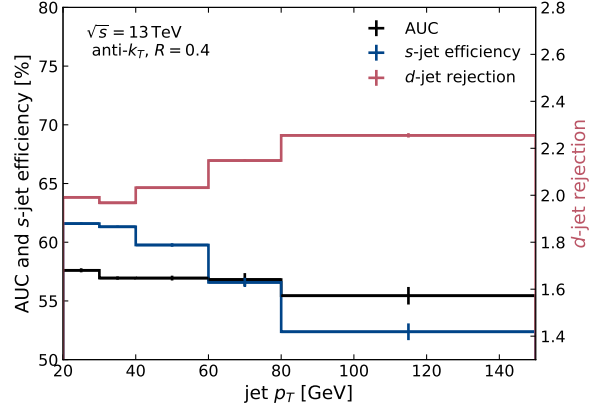
Fig. 6.20: p_T dependency of the AUC score as well as the *s*-jet efficiency and *d*-jet mis-tag rate requiring a classification value greater than 0.5 for the calorimeter scenario with a separation into ECAL and HCAL.
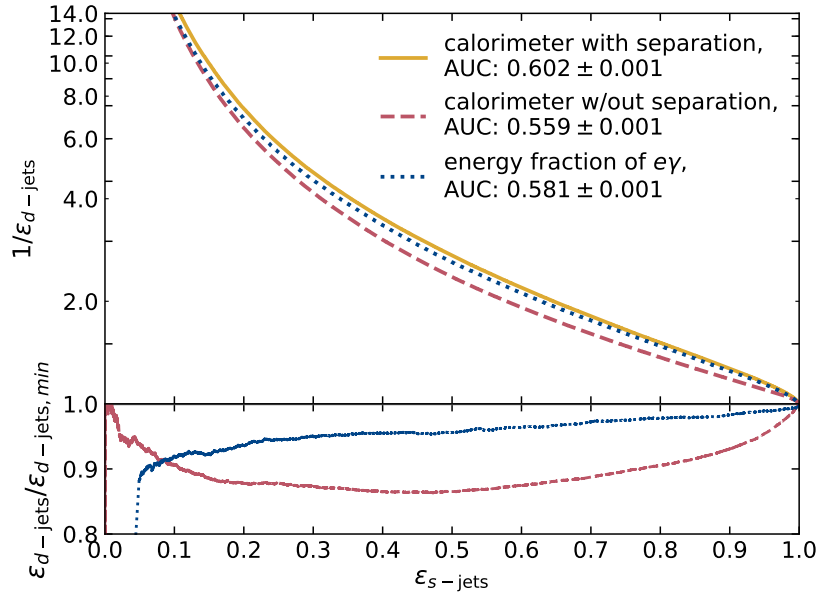

Fig. 6.21: Modified ROC curves illustrating the classification power of neural networks in the calorimeter scenarios without and with a separation into ECAL and HCAL components, and for the distribution of the fraction of energy deposited in the ECAL. The efficiencies are determined on the test sample and the uncertainties of the AUC scores are the statistical uncertainties. The ratio beneath the ROC curves shows the efficiencies for *d*-jets for all illustrated scenarios divided by the efficiency for *d*-jets in the calorimeter scenario with a separation into ECAL and HCAL components.

HCAL for the classification of *s*- and *d*-jets can be especially beneficial for larger jet p_T.

In Figure 6.21, the modified ROC curves of both calorimeter scenarios - a calorimeter without separation into ECAL and HCAL and a calorimeter with such a separation - as well as a ROC curve based on the fraction of the jet energy carried by electrons and photons are shown.

The separation power given by the jets' substructure corresponds to the separation power of the calorimeter scenario without a separation into ECAL and HCAL. Its NN has an AUC score of 0.559 ± 0.001. By adding the separation into ECAL and HCAL, the AUC score increases slightly to 0.602 ± 0.001. The AUC of the ROC curve given by the fraction of energy carried by electrons and photons is 0.582 ± 0.001. This shows that the jets' substructure – as is resolvable by a high-resolution calorimeter – has a separation power smaller than the separation given by the fraction of energy deposited in the ECAL.

6.2.4. Subsummary

In this section, it was found that the main difference between *s*- and *d*-jets is given by the flavor of their constituent particles. If there were detectors that could identify the particle type of both charged and uncharged particles and additionally determine their kinematic properties perfectly, it would be possible to distinguish both jet flavors well. An AUC score of 0.940 can be used as an estimate for the absolute upper bound on the achievable separation between *s*- and *d*-jets.

Considering several detector scenarios, it was found that the most promising approach to *s*-tagging is the use of Cherenkov detectors to identify $K^{\pm}$. With Cherenkov detectors, the maximally achievable separation was estimated to have an AUC score of 0.783.

At multi-purpose detectors as they are currently used at the LHC, an *s*-tagger will have an AUC score smaller than 0.643. For such detectors, the information from the tracking detectors and calorimeters – especially the fraction of energy deposited in the ECAL – should be considered when constructing an *s*-tagger as they both can contribute to the classification of *s*- and *d*-jets.

6.3. An *s*-Tagging Algorithm for the ATLAS Experiment

In this section, an *s*-tagger for the use at the ATLAS Experiment is developed. The NN structure chosen for the *s*-tagger is a DNN because they are suited to learn complex correlations and small differences between the input variables. The use of RNNs is not necessary because the input variables used are not jet constituents but jet attributes.

The NNs are trained on two different sets of input variables.

The first set of variables are related to reconstructed K_S and Λ^0 that are matched to jets. These strange hadrons are the most abundant strange hadrons in *pp* collisions that can be reconstructed with the ATLAS detector. The training of a NN with input features purely based on K_S and Λ^0 matched to jets is described in Section 6.3.2.1. Since not all *s* quarks hadronize into either of these two strange hadrons and they do not exclusively decay into charged hadrons, and since furthermore, the efficiency of their reconstruction is not very large, the efficiency of an algorithm purely based on reconstructed strange hadrons matched to jets is expected to be rather small.

The second set of input variables consists of jet attributes characterizing small differences in the shower shape of jets inside the calorimeter as well as properties related to tracks associated with the jets. Since these differences are empiric, they are able to cover a wide range of

hadronization scenarios. An additional advantage is the availability of these variables for every reconstructed jet and not only those that have a reconstructed strange hadron matched to them. A NN with these input variables is optimized in Section 6.3.2.2.

In a third NN optimization presented in Section 6.3.2.3, both sets of of input variables are combined to train an *s*-tagger that will be used in the measurement of $|V_{ts}|$ and $|V_{td}|$.

6.3.1. Training Samples

For the training of the NNs, jets from MC simulations of semileptonically decaying $t\bar{t}$ as described in Section 5.2.1 are used. If the hadronically decaying W boson decays into an s quark (and an additional c quark) and this s quark is found inside of the jet cone ($\Delta R(s, \text{jet}) < 0.4$), this jet is used as a signal jet. Similarly, if the W boson decays into a d quark (and an additional u quark) and this d quark is found inside of the jet cone ($\Delta R(d, \text{jet}) < 0.4$), this jet is used as a background jet. The training sample consists of 12 million s-jets and 12 million d-jets.

For the small differences in the spectrum of the p_T of these jets, which can be seen in Figure 6.22a, all d-jets are reweighted so that their distribution in p_T matches the distribution of s-jets. The jets are nor reweighted in their η distribution, which is shown in Figure 6.22b and shows only small differences between s-jets and d-jets. Both of these variables plus the azimuthal angle ϕ of the jets are always used as inputs to the NNs.
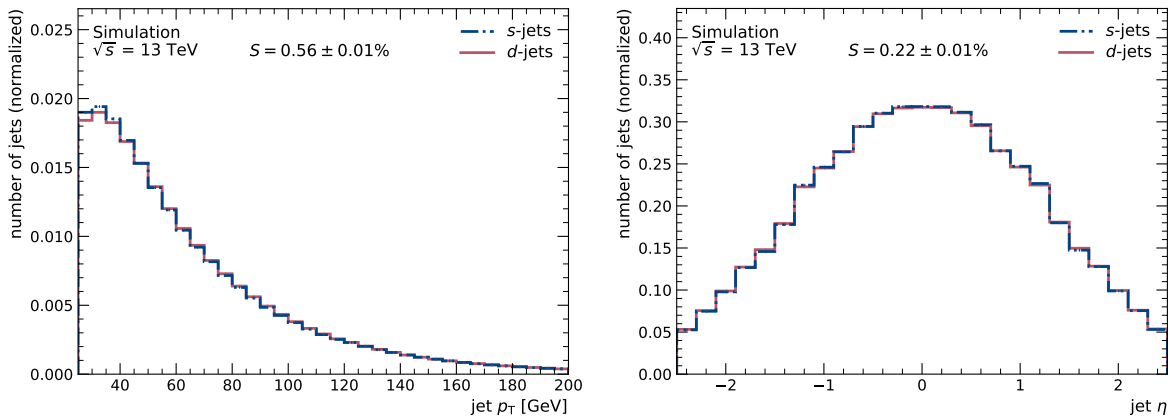
(a) Transverse momentum of jets before applying a reweighting to *d*-jets.

(b) Pseudorapidity of jets after reweighting *d*-jets to have the same distribution of transverse momentum as *s*-jets.

Fig. 6.22: Distributions of kinematic properties of *s*- and *d*-jets.

The discrimination strength of all input variables is evaluated by calculating the separation power in normalized histograms according to

$$S = \sqrt{\frac{1}{4} \sum_{i}^{N_{\text{bins}}} \frac{(s_i - b_i)^2}{s_i + b_i}} \, , \tag{19}$$

where s_i is the fraction of s-jets (signal) in bin i and b_i is the fraction of d-jets (background) in bin i [147].

The first set of input variables is based on K_S and Λ^0 reconstructed as described in Section 4.9. These hadrons are matched to jets via the requirement $\Delta R(K_S \text{ or } \Lambda^0, \text{jet}) \leq 0.4$. In case two strange hadrons are matched to the same jet and the hadron whose reconstructed mass is closest to the theoretical mass of the strange hadron is chosen.

1 million s-jets and 0.9 million d-jets in the training sample have a reconstructed K_S matched to them, while 316 000 s-jets and 364 000 d-jets have a Λ^0 matched to them. Approximately 0.04% of the s-jets and 0.05% of d-jets contain both a K_S and a Λ^0. Requiring at least one K_S or Λ^0 matched to a jet has an efficiency of 11.1% for s-jets and 10.2% for d-jets.

A list of all 12 input variables based on reconstructed K_S and Λ^0 can be found in Table 6.4. Plots of all input variables are presented and discussed in Appendix A.

	Separation Power [%]	
Input Variable	**For Jets Containing K_S**	**For Jets Containing Λ^0**
$p_T(K_S \text{ or } \Lambda^0)/p_T(\text{jet})$	10.79 ± 0.04	4.07 ± 0.05
$p_T(K_S \text{ or } \Lambda^0)$	8.88 ± 0.04	1.69 ± 0.07
$\Delta R(K_S \text{ or } \Lambda^0, \text{jet})$	5.88 ± 0.04	3.25 ± 0.07
jet contains Λ^0	4.57 ± 0.04	
$R_{xy}(K_S \text{ or } \Lambda^0)$	4.41 ± 0.04	2.92 ± 0.07
jet contains K_S	4.43 ± 0.04	
$\tau(K_S \text{ or } \Lambda^0)$	3.58 ± 0.04	2.87 ± 0.07
$p_T^{\text{rel}}(K_S \text{ or } \Lambda^0)$	3.14 ± 0.04	4.83 ± 0.06
$p_T(\text{jet})$	2.77 ± 0.04	1.44 ± 0.07
$\chi^2(K_S \text{ or } \Lambda^0)$	2.58 ± 0.04	0.96 ± 0.07
$\eta(\text{jet})$	2.36 ± 0.04	3.20 ± 0.07
$m(K_S \text{ or } \Lambda^0)$	2.11 ± 0.04	4.22 ± 0.07

Tab. 6.4: A list of all input variables that describe K_S and Λ^0 matched to a jet. The second column shows the separation power and its statistical uncertainty if the hadron matched to the jet is a K_S, while the third column shows the separation power and its statistical uncertainty if the matched hadron is a Λ^0. They are calculated using the training, validation, and test sample.

The most discriminating variable for jets with a K_S matched to them is the p_T of the K_S divided by the p_T of the jet. Its distribution is shown in Figure 6.23a and is on average larger for K_S matched to s-jets than K_S matched to d-jets.

For jets with a Λ^0 matched to them, the most discriminating variable is the momentum of the Λ^0 relative to the jet axis, p_T^{rel}, shown in Figure 6.23b. It is larger for jets from an s quark because a Λ^0 baryon contains a u and d quark in addition to an s quark. These u and d quarks can be both valence and sea quarks of the colliding protons. As valence quarks tend to carry larger momentum, they can contribute more to the momentum of the Λ^0 w.r.t. the jet axis than sea quarks. If the jet originates from a d quark that then hadronizes into a Λ^0, while the additional u quark can still be a valance or sea quark of the protons, the additional s quarks has to be a sea quark which tends to carry less momentum and hence contributes less to the momentum of the Λ^0.

The second set of input variables are more general jet attributes. They can be divided into three categories: the fraction of energy the jet deposits in each calorimeter layer, the
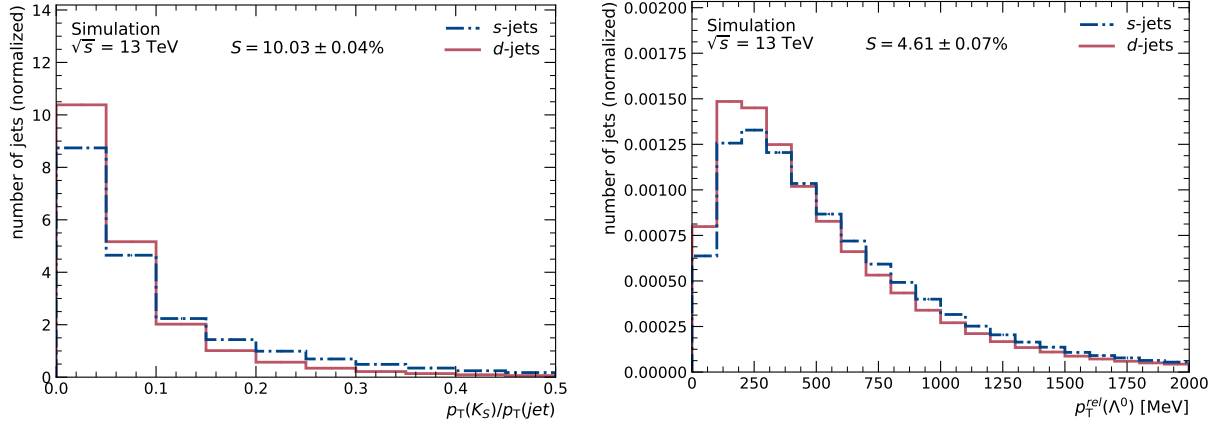
(a) Fraction of the transverse jet momentum carried by K_S.

(b) Momentum of Λ^0 transverse to the jet axis.

Fig. 6.23: Distributions of dependencies between the momenta of reconstructed strange hadrons and jets for s- and d-jets.
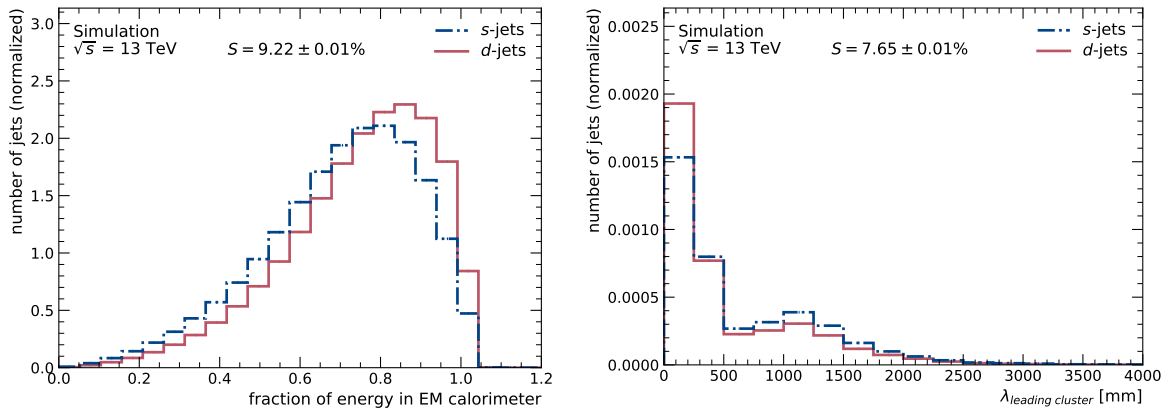
(a) Fraction of energy deposited in the EM calorimeter.

(b) Longitudinal distance between the leading cluster in the jet and its geometric center.

Fig. 6.24: Distributions of jet attributes of s-jets and d-jets.

geometrical moments of the jets' constituent clusters, and characteristics derived from tracks associated with the jets. A list of all 32 variables and their separation powers can be found in Table 6.5. Plots and descriptions of all jet attributes used as input variables can be found in Appendix B.

The most discriminating subset of variables are the sampling fractions, i.e. the energy fractions deposited in one (or several) calorimeter layer(s). The variable with the largest overall separation power is the sampling fraction of the EM calorimeter shown in Figure 6.24a. It is larger for d-jets than for s-jets because d-jets contain more π^0 decaying into $\gamma\gamma$ than s-jets which contain more K_L. Since the sampling fractions overall have the largest separation powers, all sampling fractions are included as input variables in the training.

The second subset of variables are topological cluster moments, which are described in

Type	Variable	Separation Power
sampling fractions (in)	EM calorimeter	9.22 ± 0.01
	PreSamplerB	3.82 ± 0.01
	EMB1	6.65 ± 0.01
	EMB2	5.10 ± 0.01
	EMB3	3.01 ± 0.01
	PreSamplerE	0.88 ± 0.01
	EME1	2.92 ± 0.01
	EME2	2.71 ± 0.01
	EME3	1.42 ± 0.01
	HEC0	1.98 ± 0.01
	HEC1	2.14 ± 0.01
	HEC2	1.16 ± 0.01
	HEC3	0.97 ± 0.01
	TileBar0	4.70 ± 0.01
	TileBar1	4.17 ± 0.01
	TileBar2	2.05 ± 0.01
	TileGap1	1.32 ± 0.01
	TileGap2	0.39 ± 0.01
	TileGap3	0.96 ± 0.01
	TileExt0	2.22 ± 0.01
	TileExt1	2.31 ± 0.01
	TileExt2	1.39 ± 0.01
	maximal sampling fraction	4.46 ± 0.01
shower shape	$\lambda_{\text{leading cluster}}$	7.65 ± 0.01
	$\langle r^2 \rangle$	6.16 ± 0.01
	$\langle \lambda^2 \rangle$	5.65 ± 0.01
	centroid R	5.61 ± 0.01
track related	ghost-matched tracks p_{T}	2.78 ± 0.01
	number of ghost-matched tracks	2.78 ± 0.01
	charge	1.31 ± 0.01
jet kinematics	η	0.22 ± 0.01
	p_{T}	0.00 ± 0.01

Tab. 6.5: Jet attributes used as input variables to the NNs. The first column groups the variables listed in the second column into categories. The third column names their separation power and its statistical uncertainty as calculated on all jets in the training, validation, and test sample.

Reference [148]. Figure 6.25 helps to illustrate their construction. It shows a jet's principle axis $\vec{s}$, the vector $\vec{c}$ to the jet's geometrical center, and the vector $\vec{x}_i$ to any cluster i in the jet, as well as how the longitudinal distance λ_i and the lateral distance r_i between any cluster i and the jet center is defined. The jet moments used as input variables are the longitudinal distance between the cluster with the largest energy (leading cluster) and the jet's geometrical center, $\lambda_{\text{leading cluster}}$, shown in Figure 6.24b, the variance of the energy distribution in the lateral dimension r of the jet, $\langle r^2 \rangle$, the variance of the energy distribution in the longitudinal

dimension, λ, $\langle \lambda^2 \rangle$, and the radial distance R between the primary vertex and the jet's centroid (i.e. the barycenter weighted in E), called *centroid R*. All of these quantities are larger for s- than for d-jets since s-jets tend to deposit their energy further down the calorimeter (more energy in the hadronic calorimeter in comparison to d-jets) and therefore have a larger extension.
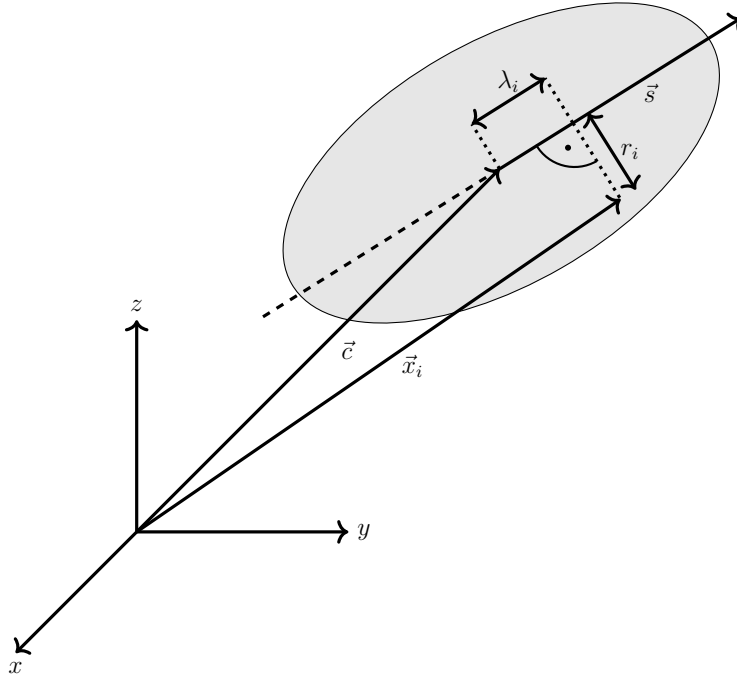

Fig. 6.25: Graphical representation of the jet geometry considered in the calculation of topological cluster moments. $\vec{c}$ is the vector to the geometrical center of the jet, $\vec{s}$ is the jet's principal axis, and $\vec{x}_i$ is the vector to a cluster i within the jet. λ_i is the longitudinal distance between the cluster i and the jet center, and r_i is the radial distance between the cluster and the jet axis. The image is based on the descriptions in Reference [148].

The final subset of input variables are related to reconstructed tracks which are associated with the jet via ghost-matching [142]: These variables are the total number of tracks associated to the jet, the p_{T} carried by these tracks, and their p_{T}-weighted charge defined as

$$Q = \frac{\sum\limits_{i \in \text{tracks}} q_i \cdot p_{\mathrm{T}i}}{\sum\limits_{i \in \text{tracks}} p_{\mathrm{T}i}} \; . \tag{20}$$

The absolute value of all of these variables is slightly larger for d-jets, but their separation power is relatively small.

6.3.2. Optimization of the Neural Networks Used for *s*-Tagging

6.3.2.1. An *s*-Tagging Algorithm Based on Reconstructed K_S and Λ^0

This section covers the training and optimization of the NN based on reconstructed K_S and Λ^0 matched to jets. To reduce the computational resources necessary for the training, all jets which do not have such a strange hadron matched to them are removed from the input samples instead of using default values for the input features if no strange hadrons are matched to the jets.

For the sake of simplicity, during the hyperparameter optimization, only the width of the NN and its depth are varied. The other hyperparameters remain constant: The batch size is set to 1024 events, the initial learning rate is $\alpha_{\text{initial}} = 10^{-4}$, and the exponential decay rates are set to $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The metric evaluated during the training is the AUC score. The training is stopped once the AUC score calculated on the validation sample does not improve over 30 epochs. The final NN weights are set to those of the epoch with the best AUC score as evaluated on the validation sample.

For the optimization, the following NN structures are considered: The input layer always contains 12 input nodes corresponding to the input variables, while the width w of the NN (i.e. the number of nodes in the first hidden layer) is either 128, 256, 512, 1024, or 2048, and the depth d (i.e. the number of hidden layers) is varied between two, three, and four.

The aim of the first step of the optimization is to find a NN configuration which has a capacity that is large enough to make use of all information available in the input variables and their correlations for the best possible discrimination. The majority of the configurations that have such a capacity are actually too large and will show signs of overfitting during the training and evaluation. This overfitting can e.g. be observed as an increase of the loss or a
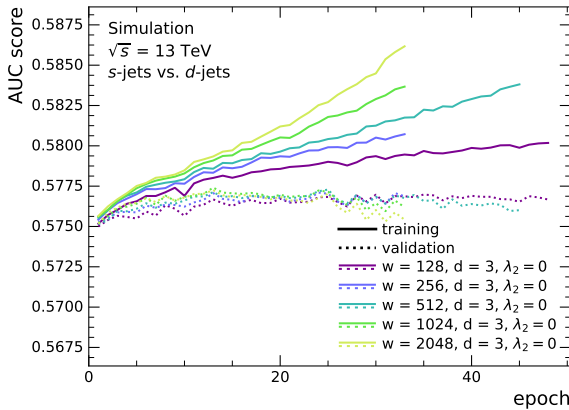
Fig. 6.26: AUC scores of neural networks based on K_S and Λ^0 with a depth d of three hidden layers and different widths w during the training. The solid lines show the AUC as evaluated on the training sample, while the dashed lines show it as evaluated on the validation sample.
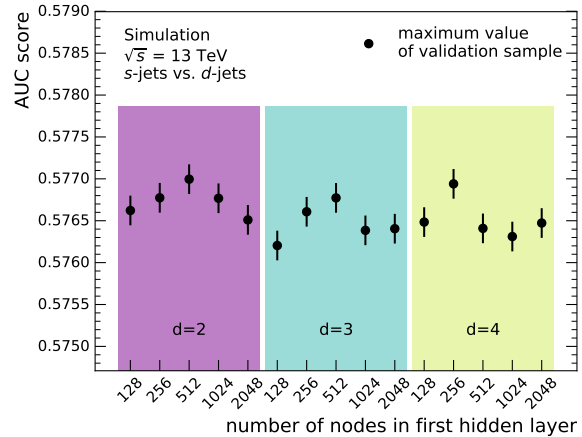
Fig. 6.27: Comparison of the maximal AUC of the output of the validation sample for neural networks based on K_S and Λ^0 with different numbers of nodes in the first hidden layer (widths) and different numbers of hidden layers d. The uncertainties are the statistical uncertainties.

decrease of the AUC score of the validation sample during the training.

Figure 6.26 shows the AUC score during the training of the NN configurations with three hidden layers for the training and the validation sample. The AUC score of the validation sample per epoch looks very similar for all NNs but starts to decrease slightly throughout the training, especially for those NNs with a larger width. The AUC score calculated on the training sample, on the other hand, increases more quickly during the training for an increasing width of the NNs. Both the increasing difference between the AUC scores of the training and validation sample and that the AUC score of the validation sample is not increasing significantly are signs of overfitting during the training, which means that all of the considered NN structures have a sufficient capacity.

A summary plot comparing the maximal AUC score during the training calculated using the NN output distribution of the validation sample for each NN configuration is shown in Figure 6.27. This comparison of AUC scores shows that there are only small differences in the performance of the different NN structures; all fluctuations are covered by the statistical uncertainties. For the following studies, the NN with the largest AUC as calculated on the validation sample is chosen. It has two hidden layers and 512 nodes in the first hidden layer.

λ_{L2}	Accuracy	AUC Score
$5 \cdot 10^{-6}$	0.550 ± 0.002	0.5771 ± 0.0006
$1 \cdot 10^{-5}$	0.551 ± 0.002	0.5773 ± 0.0006
$5 \cdot 10^{-5}$	0.550 ± 0.002	0.5772 ± 0.0006
$1 \cdot 10^{-4}$	0.551 ± 0.002	0.5770 ± 0.0006

Tab. 6.6: Accuracy and AUC score of the validation sample for the neural network based on K_S and Λ^0 when varying λ_{L2} applied in the L2 regularization. The neural network has two hidden layers and 512 nodes in the first hidden layer. The uncertainties are the statistical uncertainties.

Because the best chosen NN model shows signs of overfitting, it is regularized using L2 regularization to find the optimal effective capacity for this classification problem. This is done by adding L2 regularization to every hidden layer of the NN as described in Section 6.1.4. Table 6.6 shows the considered regularization strength values λ_{L2} and their impact on the accuracy and the AUC score of the NN. Following this table, $\lambda_{\mathrm{L2}} = 1 \cdot 10^{-5}$ is chosen for the regularization of the NN as it is the largest of the considered regularization strengths that does not degrade the performance of the NN.

To evaluate the impact of the remaining overfitting, the NN output distributions for s- and d-jets are compared for the training and the independent test sample in Figure 6.28. Differences in the output of the two samples are only visible in the ratio between the two samples and lie within 1% except in those bins containing small numbers of jets. The NN output distribution has a large peak around 0.5, which means that there are a significant number of jets which cannot be distinguished based on strange hadrons, even if they are reconstructed and matched to the jets.

The variable with the highest correlation to the NN output is the transverse momentum fraction of the reconstructed strange hadrons, $p_{\mathrm{T}}(K_S \text{ or } \Lambda^0)/p_{\mathrm{T}}(jet)$. It has a correlation coefficient of 0.8 for s-jets and 0.6 for d-jets as it has the largest separation power for jets with a K_S matched to them. Figure 6.29 compares the modified ROC curves of the NN output and $p_{\mathrm{T}}(K_S \text{ or } \Lambda^0)/p_{\mathrm{T}}(jet)$. It shows that the use of the NN yields an improvement in the
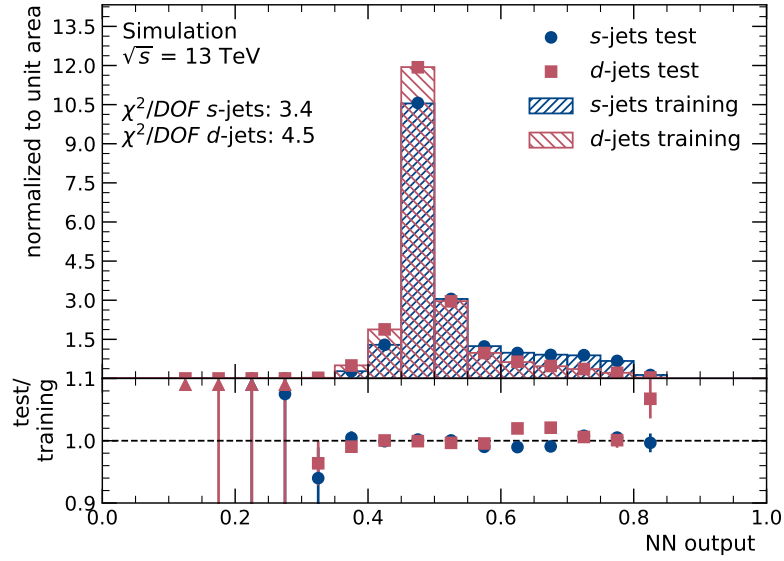
Fig. 6.28: Output distribution of the neural network based on K_S and Λ^0 with two hidden layers, a width of 512 nodes in the first hidden layer, and with regularization. The comparison shows the output for s- and d-jets for the training and test sample (top panel) and their ratio for both jet flavors (bottom panel).
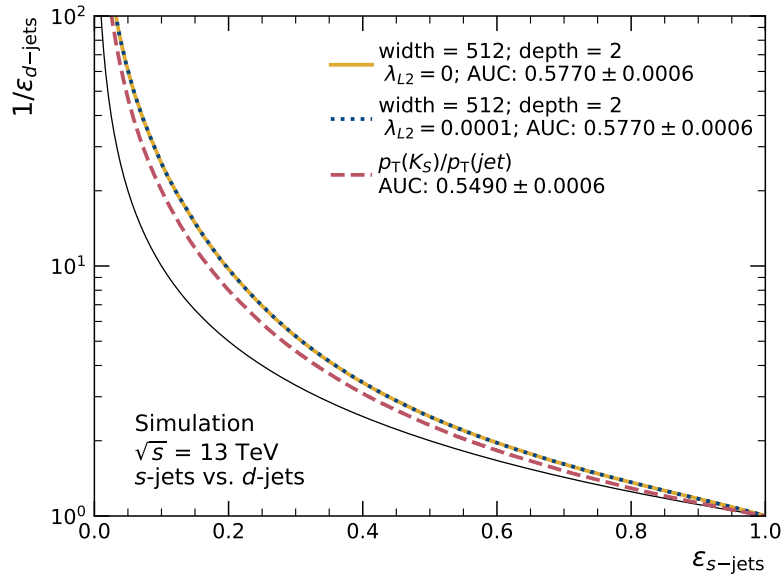

Fig. 6.29: Modified ROC curve of the neural network based on K_S and Λ^0 with two hidden layers and a width of 512 nodes in the first hidden layer, both with and without regularization, and the ROC curve of the fraction of transverse momentum carried by the strange hadron. All modified ROC curves are based on the test sample of jets with a matched reconstructed strange hadron and the uncertainties on the AUCs are the statistical uncertainties.

classification across all s-jet efficiencies and the AUC increases by 59% compared to the use of $p_{\mathrm{T}}(K_S \text{ or } \Lambda^0)/p_{\mathrm{T}}(jet)$. Figure 6.29 also shows that the regularization does not degrade the NN performance but instead slightly improves the AUC score as evaluated on the test sample.

6.3.2.2. An *s*-Tagging Algorithm Based on Jet Attributes

In the following, a second NN based on jet attributes is optimized. For this s-tagger setup, NN depths d of two, three, and four hidden layers are compared, additionally varying the widths w to 128, 256, 512, and 1024 nodes respectively. In the first step of the optimization, no regularization is applied. The batch size is again set to 1024 events, the initial learning rate is $\alpha_{\mathrm{initial}} = 10^{-4}$, and the exponential decay rates are set to $\beta_1 = 0.9$ and $\beta_2 = 0.999$.
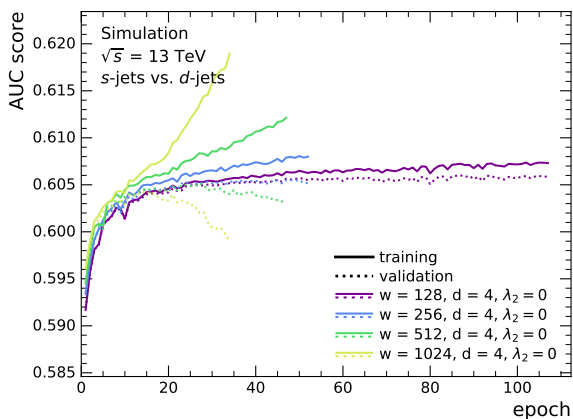
Fig. 6.30: AUC scores of neural networks based on jet attributes with a depth d of four hidden layers and different widths w during the training. The solid lines show the AUC as evaluated on the training sample, while the dashed lines show it as evaluated on the validation sample.
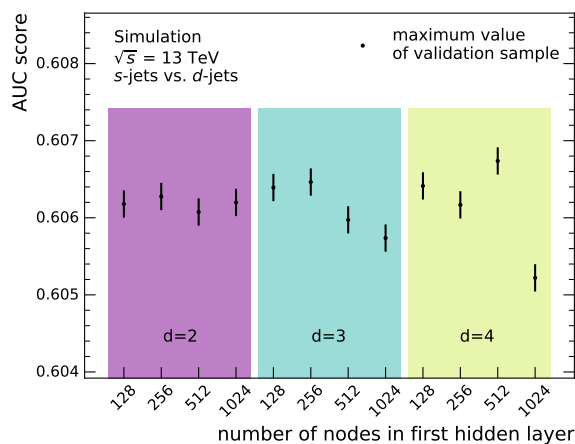
Fig. 6.31: Comparison of the maximal AUC of the output of the validation sample for neural networks based on jet attributes with different numbers of nodes in the first hidden layer (widths) and different numbers of hidden layers d. The uncertainties are the statistical uncertainties.

Figure 6.31 shows the maximal AUC score achieved during the training as evaluated on the validation sample. Similarly to what is seen in Figure 6.26, the differences between different NN configurations appear to be statistical fluctuations. However, the fluctuations increase for larger NN depths. The NN with a depth of four hidden layers and a width of 1024 nodes in the first hidden layer e.g. has a relatively low AUC score. Taking a look at Figure 6.30 which shows the AUC scores of several NNs configurations with four hidden layers, it becomes visible that for this particularly large NN configuration, the AUC scores of the training and validation sample start to diverge very early in training. This means that the NN starts to learn features that cannot be generalized.

In the following, the configuration with the best AUC score evaluated on the validation sample is used. This NN had a depth of four hidden layers and a width of 512 nodes in the first hidden layer. To reduce the effective capacity of this NN configuration, L2 regularization is applied in all hidden layers. A regularization strength of $\lambda_{\mathrm{L2}} = 5 \cdot 10^{-6}$ was found to be
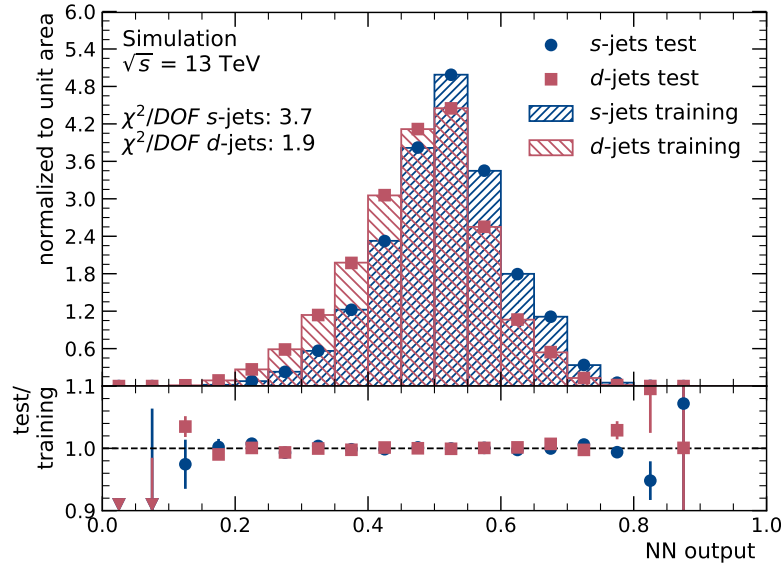
Fig. 6.32: Output distribution of the neural network based on jet attributes with four hidden layers, a width of 512 nodes in the first hidden layer, and with regularization. The comparison shows the output for *s*- and *d*-jets for the training and test sample (top panel) and their ratio for both jet flavors (bottom panel).
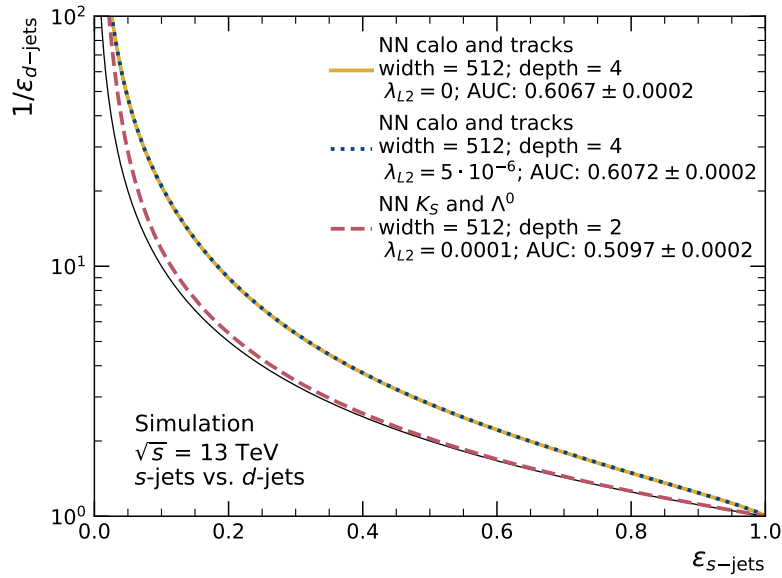

Fig. 6.33: Modified ROC curve of the neural network based jet attributes with four hidden layers and a width of 512 nodes in the first hidden layer, both with and without regularization, and the modified ROC curves of the optimized NN based on reconstructed K_S and Λ^0 is shown. All modified ROC curves are based on the full test sample and the uncertainties on the AUCs are the statistical uncertainties.

optimal, maximizing the AUC score as evaluated on the validation sample.

Figure 6.32 shows the output distributions of the optimized NN for s- and d-jets and for the training and the independent test sample. Except in bins with small numbers of jets, the distributions of the training and the test sample agree well and no trend can be observed in the ratio of the distributions of both samples. Compared to the NN based on K_S and Λ^0, the NN distribution is significantly broader, which means that there are fewer jets that look very similar based on their input variables. However, the output distributions of s- and d-jets have a large overlap, i.e. it is not possible to clearly distinguish these jets. The ROC curves shown in Figure 6.33 illustrate the difference between the use of the two different sets of input variables, the one based on strange hadrons and the one based on jet attributes. This figure accounts for the loss of efficiency due to the requirement of matched K_S and Λ^0, which significantly reduces the discrimination power of the NN based on strange hadrons. Figure 6.33 also shows that the used regularization does not degrade the performance of the NN.

6.3.2.3. *s*-Tagging Algorithm Combining Inputs

For an even better classification performance, the input variables based on reconstructed K_S and Λ^0 and based on jet attributes are combined.

The setup for the optimization of the hyperparameters of the NN is identical to the optimization of the NN based on the jet attributes.

(a) Comparison before applying regularization.

(b) Comparison for several different regularization strengths λ_2 applied in all layers.
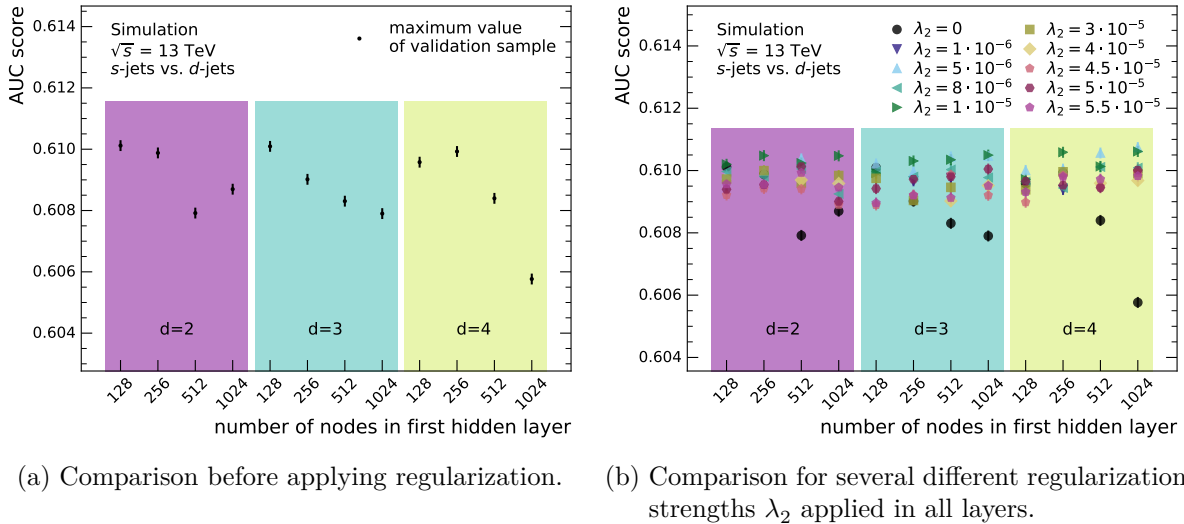
Fig. 6.34: Comparison of the maximal AUCs of the validation sample for neural networks based on jet attributes and reconstructed strange hadrons with different numbers of nodes in the first hidden layer (widths) and different numbers of hidden layers d. The uncertainties are the statistical uncertainties.

Figure 6.34a shows the maximal AUC scores of all NN structures for the validation sample during the training without the use of any regularization. It can be observed that the AUC score decreases for larger capacities of the NN. This is similar to what is observed for the NN with a width of 1024 nodes and four hidden layers trained with the input of jet attributes only and can be interpreted as overfitting. Because the overfitting has such a large impact on the NN output distribution of the validation sample, it is difficult to estimate which NN
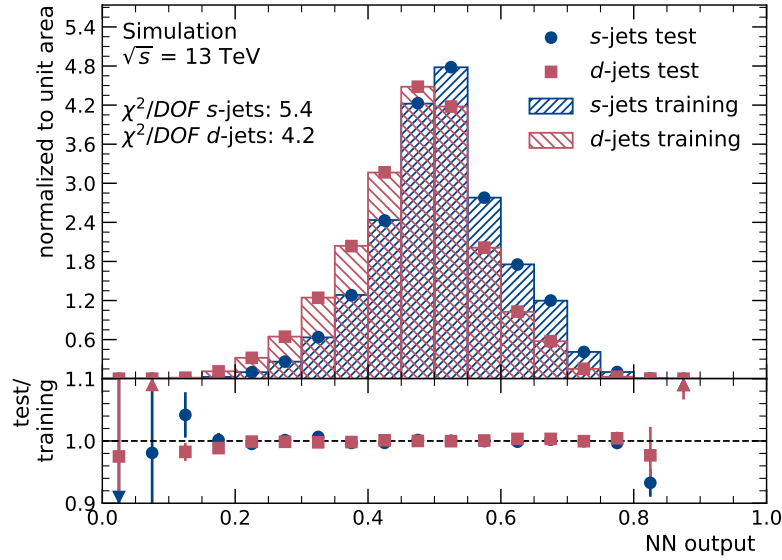
Fig. 6.35: Output distribution of the neural network based on jet attributes and reconstructed strange hadrons with four hidden layers, a width of 1024 nodes in the first hidden layer, and with regularization. The comparison shows the output for *s*- and *d*-jets for the training and test sample (top panel) and their ratio for both jet flavors (bottom panel).
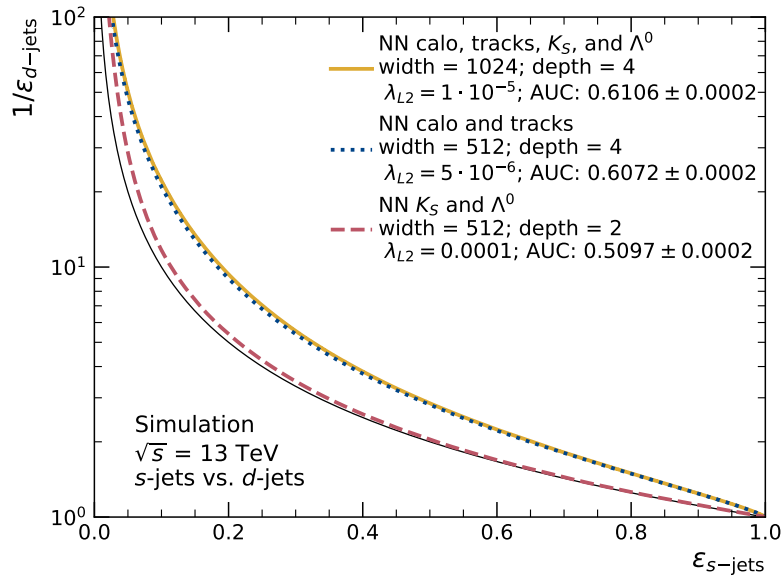

Fig. 6.36: Modified ROC curves of the neural network based on jet attributes and reconstructed strange hadrons with four hidden layers, a width of 1024 nodes in the first hidden layer, and with regularization, the modified ROC curves of the optimized NN based on jet attributes only, and the modified ROC curves of the optimized NN based on reconstructed K_S and Λ^0 is shown. All modified ROC curves are based on the full test sample and the uncertainties on the AUCs are the statistical uncertainties.

has a capacity that is large enough for the classification of *s*- and *d*-jets. Thus, to find the optimal effective capacity, L2 regularization with different strengths λ_{L2} is applied to NNs of all structures. As can be seen in Figure 6.34b, this leads to increased AUC scores for those NNs that have small scores without regularization and similar scores for those of the NNs which achieved better performance without regularization. Overall, the best AUC scores of every NN structure achieved when applying L2 regularization (in most cases, the best regularization strength lies between $\lambda_{L2} = 5 \cdot 10^{-6}$ and $1 \cdot 10^{-5}$) agree within their statistical uncertainties. The network configuration chosen for further studies has four hidden layers, 1024 nodes in the first hidden layer, and L2 regularization with a strength of $\lambda_{L2} = 1 \cdot 10^{-5}$ applied in every layer.

Figure 6.35 shows the output distribution of the final NN used for *s*-tagging both evaluated on the training sample and an independent test sample. No significant trend in the ratio of both distributions can be observed, neither for *s*- nor for *d*-jets, which means that no significant amounts of overfitting remain after the regularization is applied. Overall, the output distribution looks very similar to what is shown in Figure 6.32, which means that the additional information provided by the reconstructed strange hadrons does not yield large additional separation power.

Figure 6.36 shows the modified ROC curves of the optimized NNs based on all three sets of input variables. It illustrates that adding information about reconstructed strange hadrons improves the classification power of the NN based on jet attributes across all efficiencies. However, this improvement is very small, mostly because the overall classification power provided by reconstructed strange hadrons itself is small due to the limited number of reconstructed K_S and Λ^0 matched to jets.

CHAPTER 7

## Efficiency Measurement and Calibration of the *s*-Tagger

When a flavor tagging algorithm – such as the *s*-tagger described in the previous chapter – is trained on simulated events, there is always some uncertainty on how its discrimination power transfers to data. Due to limited precision in the simulation, there are differences between data and simulation both in the distributions of the input variables and their correlations. While the modeling of both can be checked in control regions, it is not obvious whether some differences between simulation and data will add up in the NN used for the tagger, while others might balance each other out. It is therefore necessary to quantify the difference between tagger responses of data and simulation and adjust the simulation to provide a better description of data. This is done by first determining the signal jet efficiency and the mis-tag rates for background jets both in simulation and data. Following this step, scale factors (SFs) that map the efficiency of simulated events to the efficiency measured in data are derived. The uncertainties of these SFs can then be propagated to any analysis, quantifying the uncertainty associated with its use of the tagger.

This chapter will give an overview of how the efficiency and mis-tag rates for the aforementioned *s*-tagger are determined and how the SFs for MC simulations used at the ATLAS Experiment are derived. In Section 7.1, different options to define the flavor of a jet are discussed. Section 7.2 illustrates the event selections used to select semileptonic $t\bar{t}$ decays and to enrich five different jet samples with different flavors. In Section 7.3, the data-driven estimate used to predict the background contribution of events containing fake leptons is shown. Section 7.4 summarizes the method used for the calibration of the *s*-tagger. Finally, in Section 7.5, the determined efficiencies and mis-tag rates of the *s*-tagger together with the resulting SFs are presented.

7.1. Flavor Definition

One conceptual challenge of the calibration of an *s*-tagger is the definition of the jets' true flavor. Ideally, such a flavor label corresponds to the flavor of the one parton the jet results from. However, because of confinement, jets are never the result of only one parton hadronizing; this idea can only in some cases be a good approximation if an initial parton has a relatively large energy contribution to the jet [149]. Therefore, the flavor label of a jet is always an approximation and has to be chosen in the context of its use. While there have been ideas on how to solve this issue, cf. References [149, 150], the flavor labels proposed in these references need to be determined during the clustering of the jets, which is done early in the analysis chains used at the ATLAS Experiment and not easily changed. Therefore, the considerations on flavor labels presented here are limited to the information available after the jet clustering.

One option is the use of *parton labels* to determine the jet's flavor. For this, partons from the ME generation step of the MC simulations (stored in a tree-like truth record) are ghost-
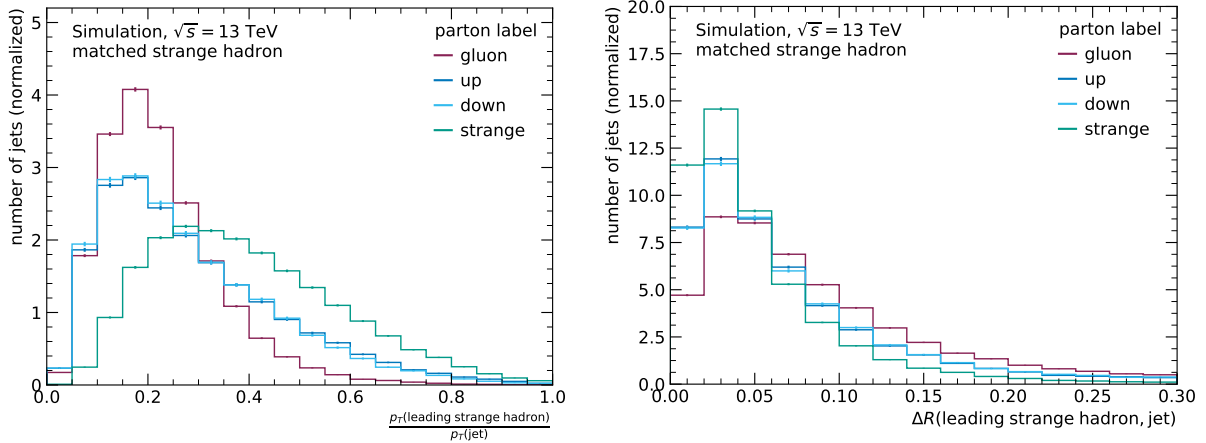
(a) Fraction of the transverse momentum of the jet carried by the leading matched strange hadron.

(b) Distance in the η–ϕ plane between the leading matched strange hadron and the jet axis.

Fig. 7.1: Distributions of kinematic properties of the leading strange hadron matched to a jet requiring $\Delta R(\text{jet}, \text{hadrons}) < 0.3$ and $p_{\text{T}}(\text{hadron}) > 5\,\text{GeV}$. The shower generator used is Pythia. The plots distinguish between jets which have a strange quark, a light quark (up or down), or a gluon ghost-matched to them. The error bars indicate the statistical uncertainty of the sample used.

matched [142] to the jet and the jet flavor is defined as the flavor of the parton with the highest energy matched to the jet. This flavor definition has the advantage of being able to distinguish between all quark flavors and gluons, which is beneficial if the tagger is sensitive to gluon-initiated jets[1]. However, it has plenty of disadvantages: Because of confinement, partons cannot be observed on their own and can therefore not be considered to be accurately modeled physical objects. Since they are purely conceptual objects in the simulation process and as the simulation is limited to the finite number of Feynman diagrams it considers, it is possible that partons of the same "type" stored in the truth record of the simulated events do not have the same properties across different MC generators.

Another disadvantage is that it is not possible to assign a label to a jet if no parton is matched to it, which means a small number of jets will not be labeled at all.

The most common hadron-based flavor definition used for the flavor definition of b-jets and c-jets at the ATLAS Experiment makes use of B and D hadrons. If a B hadron with $p_{\text{T}} > 5\,\text{GeV}$ is matched to the jet ($\Delta R(\text{jet}, \text{hadron}) < 0.3$), this jet is called a b-jet. If no B hadron is matched to the jet, but instead a D hadron with a $p_{\text{T}} > 5\,\text{GeV}$, the jet is called a c-jet. This definition of b- and c-jets is retained in the following discussions about hadron-based flavor labels for all remaining jets.

In principle, this flavor definition could simply be extended to s-jets: A jet would be defined as an s-jet if neither B nor D hadrons are matched to it but instead a strange hadron with $p_{\text{T}} > 5\,\text{GeV}$. All other jets would then be called light jets or l-jets.

Figure 7.1a shows the p_{T} of the leading matched strange hadron matched to a jet that is neither a b- nor c-jet divided by the jet p_{T}. It peaks around 0.2 for jets with a parton label of a light quark or gluon (l- or g-jets) because their strange hadrons tend to be the

---

[1]The s-tagger trained in this thesis contains input variables that are used for quark/gluon tagging, cf. Reference [151].

(a) *s*-Tagger output of jets neither containing a *B*, *D*, or strange hadron.

(b) *s*-Tagger output of jets containing neither a *B*, nor *D* hadron, but a strange hadron with $p_{\mathrm{T}} > 5\,\mathrm{GeV}$.
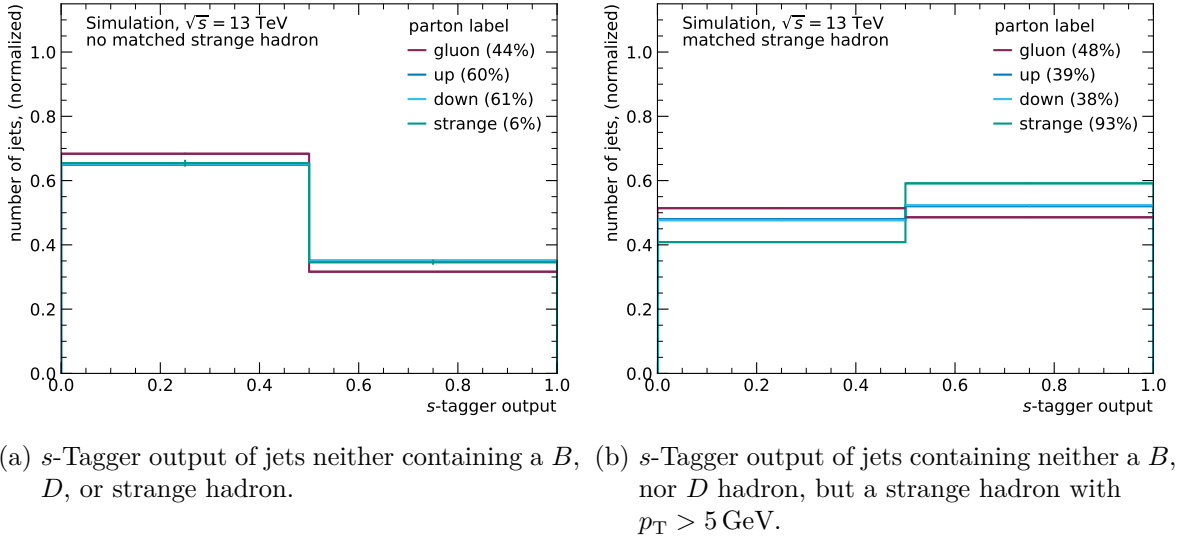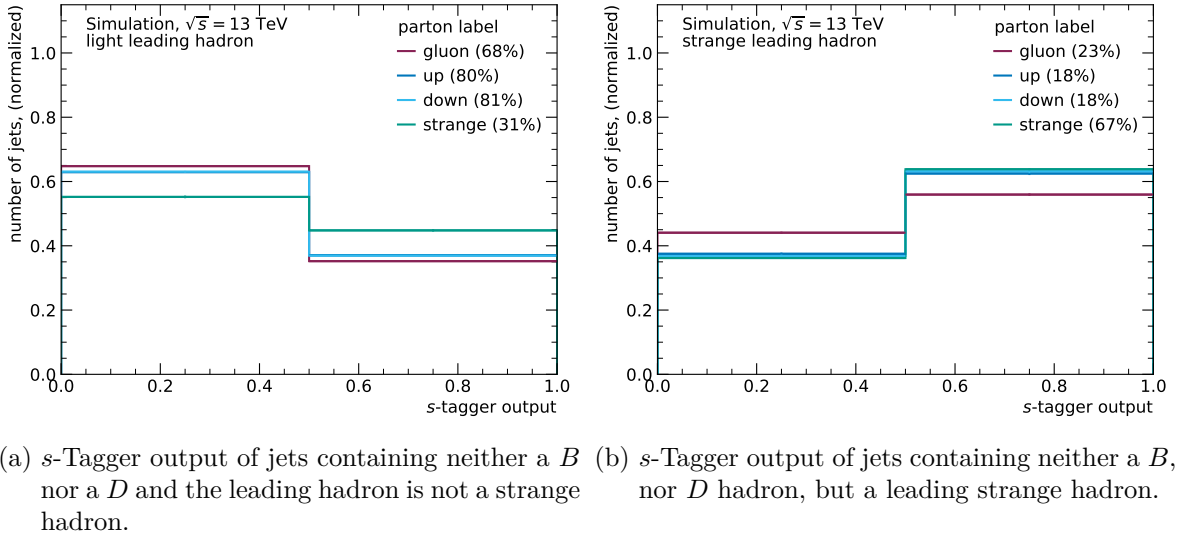
Fig. 7.2: *s*-Tagger output of jets with different parton labels simulated using Pythia as shower generator. The percentages given in the legend illustrate the fraction of jets of the given parton flavor included under the label of the given matched hadron. The error bars indicate the statistical uncertainty of the sample used.

(a) *s*-Tagger output of jets containing neither a *B* nor a *D* and the leading hadron is not a strange hadron.

(b) *s*-Tagger output of jets containing neither a *B*, nor *D* hadron, but a leading strange hadron.

Fig. 7.3: *s*-Tagger output of jets with different parton labels simulated using Pythia as shower generator. The percentages given in the legend illustrate which fraction of jets of the given parton flavor is included under the label based on the leading matched hadron. The error bars indicate the statistical uncertainty of the sample used.

result of radiation processes prior to the hadronization; the relative p_{T} of the strange hadron peaks around 0.35 for jets with a strange parton label (*s*-jets). Figure 7.1b shows the distance between the leading strange hadron and the jet axis, ΔR, in the η–ϕ plane. Because of gluon splitting ($g \to s\bar{s}$), for *g*-jets, this distance is larger than for *s*- and *l*-jets.

These different kinematic properties of the strange hadrons w.r.t. the jet have an impact on the *s*-tagger response. This can be seen in Figure 7.2, showing the *s*-tagger output for jets

that are neither *b*- nor *c*-jets and depending on whether they have a strange hadron matched to them or not. Especially for those jets with a strange hadron matched to them, it can be seen that the *s*-tagger response varies for different parton labels, the largest difference being the difference between jets with a strange parton label and jets with a gluon parton label.

Therefore, it seems necessary to consider the strange hadrons' kinematic properties (especially the fraction of the jet p_T it carries) when defining a hadron-based flavor label.

One option to do this is to only consider the flavor of the leading hadron matched to a jet to define its flavor. There are, however, two issues with this approach. One is that for *g*-jets in which the gluon splits into $s\bar{s}$, the momentum of the gluon is split between the two quarks, resulting in strange hadrons which, while carrying relatively low momentum, can still have the
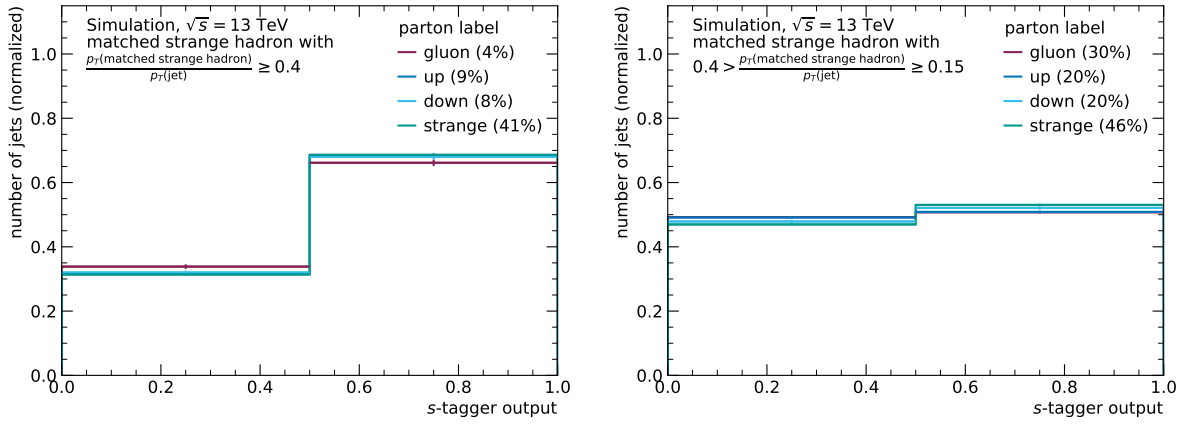
(a) *s*-Tagger output of jets containing neither a *B* or *D* hadron but a strange hadron with $p_T(\text{strange hadron})/p_T(jet) \geq 0.4$.

(b) *s*-Tagger output of jets containing neither a *B*, nor *D* hadron, but strange hadron with $0.4 > p_T(\text{strange hadron})/p_T(jet) \geq 0.15$
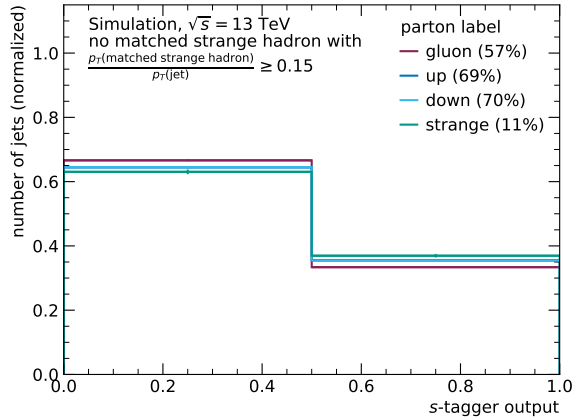
(c) *s*-Tagger output of jets containing neither a *B* nor *D* and no strange hadron with $p_T(\text{strange hadron})/p_T(jet) \geq 0.15$.

Fig. 7.4: *s*-Tagger output for jets of different parton labels simulated using Pythia as shower generator. Additional selections are applied to these jets, requiring them to fulfill the criteria for a hadron-based flavor definition. The percentages given in the legend illustrate which fraction of jets of the given parton flavor passes the hadron-label selection. The error bars indicate the statistical uncertainty of the sample used.

largest momentum of all hadrons in the jet. On the other hand, even if the jet is from an *s* quark, the resulting strange hadron might not be the hadron carrying the largest momentum in the jet. In this case, the jet would be classified as a light jet, even if it contains a strange hadron with relatively large momentum. The impact of both of these caveats on the *s*-tagger response can be observed as different responses for jets of different parton labels in Figure 7.3.

These observations suggest that it would be good to define the flavor according to the relative momentum of the matched strange hadron. Ideally, one would define an arbitrary number of flavors using intervals of its distribution shown in Figure 7.1a, thereby reducing the variance in the *s*-tagger for jets with one hadron label output w.r.t. different parton labels. Figure 7.4 shows such a flavor definition for which the distribution of the relative p_{T} was divided into three bins, $[0.0, 0.15)$, $[0.15, 0.4)$, and $[0.4, 1.0]$. If the leading strange hadron carries a transverse momentum fraction smaller than 0.15, the jet is classed as an *l*-jet along those jets without a strange hadron matched to them. If the momentum fraction is larger than 0.15 but smaller than 0.4, the jets are classed as *ambiguous jets*, which have an *s*-tagger response that is relatively flat. Only if the strange hadron carries a momentum fraction larger than 0.4, it is classed as an *s*-jet. This flavor definition reduces the variance of the *s*-tagger responses for jets of different parton labels significantly. However, in the context of this thesis, this definition cannot be used since for the calibration, the topology of $t\bar{t}$ decays is exploited to enrich calibration regions with specific jet flavors. As the group of ambiguous jets consist of jets from all three light quarks (u, d, s) as well as gluons, it is not possible to define a topology that would enrich a sample in ambiguous jets. Additionally, jets from *s* quarks are split between two labels (*s*-jets and ambiguous jets), meaning that a region enriched with *s*-jets, even if would consist of 50% *s*-jets as defined by the parton label, would only contain 20% *s*-jets as defined by this hadron label.
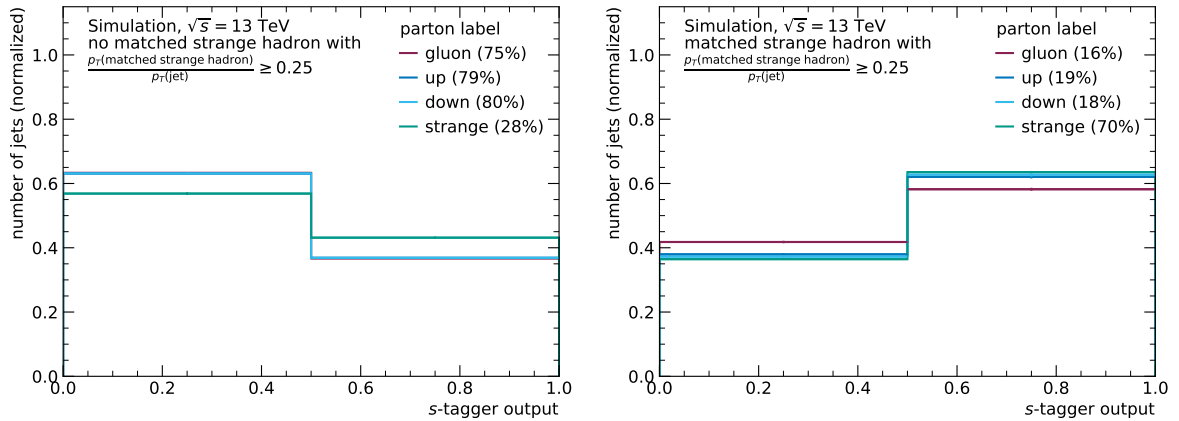
(a) *s*-Tagger output of jets neither containing a *B*, *D*, nor a strange hadron with a relative transverse momentum ≥ 0.25.

(b) *s*-Tagger output of jets containing neither a *B*, nor *D* hadron, but a strange hadron with a relative transverse momentum ≥ 0.25.

Fig. 7.5: *s*-Tagger output of jets with different parton labels simulated using Pythia as shower generator. Additional selections are applied to these jets, requiring them to fulfill the criteria for a hadron-based flavor definition. The percentages given in the legend illustrate which fraction of jets of the given parton flavor are included under the definition of the hadron-based label. The error bars indicate the statistical uncertainty of the sample used.

The final considered option is an approach that divides the distribution of the momentum fraction of the strange hadron into two bins. This is done for the hadron label shown in Figure 7.5. Here, a matched strange hadron with a p_T fraction larger 0.25 is required for a jet to be classified as *s*-jet. Jets with no strange hadron matched to them or if the strange hadron matched to them carries less 25% of the momentum of the jet are classified as *l*-jets. This p_T requirement was chosen to reduce the variance of the *s*-tagger output for jets of different parton labels. It was checked that this variance does not increase when using a different *s*-tagging working point (WP), i.e. a different binning of the *s*-tagger output distribution.

7.2. Jet Flavor Enriching Selections

For the calibration of the *s*-tagger, five regions are defined, each enriched with a certain parton-based jet flavor. Their selections are based on the kinematic reconstruction of semileptonic $t\bar{t}$ decays[2].
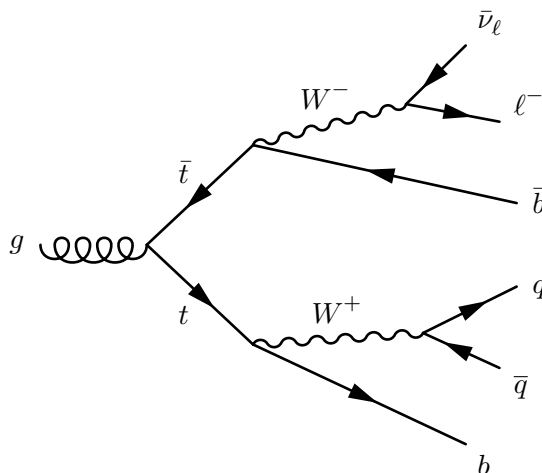
Fig. 7.6: Exemplary Feynman diagram of leading-order $t\bar{t}$ production and a semileptonic decay.

An exemplary Feynman diagram of this decay is shown in Figure 7.6. In the preselection applied to all calibration regions, the following criteria are applied:

- exactly one reconstructed lepton (muon or electron) with $p_\mathrm{T} > 27\,\mathrm{GeV}$ matched to the single-lepton trigger used to select the event for read-out

- exactly four jets with $25\,\mathrm{GeV} < p_\mathrm{T} < 200\,\mathrm{GeV}$

- $E_\mathrm{T}^\mathrm{miss} > 30\,\mathrm{GeV}$ and either $m_\mathrm{T}(W) > 30\,\mathrm{GeV}$ if the lepton is an electron or $m_\mathrm{T}(W) + E_\mathrm{T}^\mathrm{miss} > 60\,\mathrm{GeV}$ if the lepton is a muon, where

$$m_\mathrm{T}(W) = \sqrt{2\left(p_\mathrm{T}(\ell)E_\mathrm{T}^\mathrm{miss} - \vec{p}_\mathrm{T}\vec{E}_\mathrm{T}^\mathrm{miss}\right)}. \tag{21}$$

This preselection results in a relatively clean sample of events containing $t\bar{t}$ decays. Figure 7.7 shows the p_T of the jet leading in p_T. The main background is the associated production of W bosons and jets as well as events containing fake leptons.

---

[2]The definition of the flavor region was done in cooperation with Egor Evsenin during his master thesis project [152].
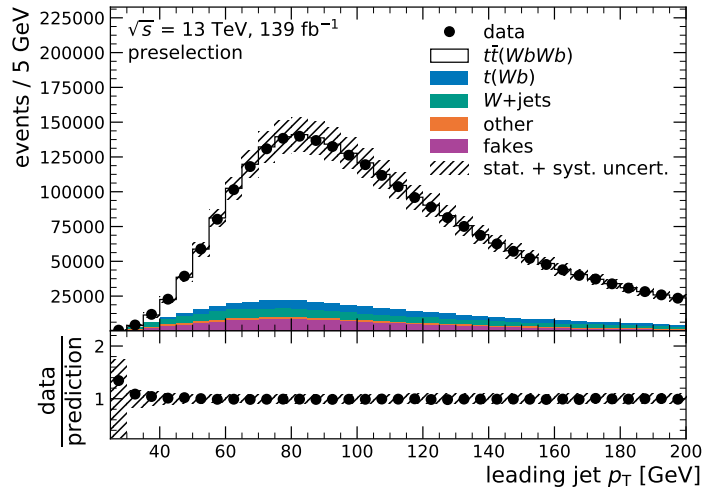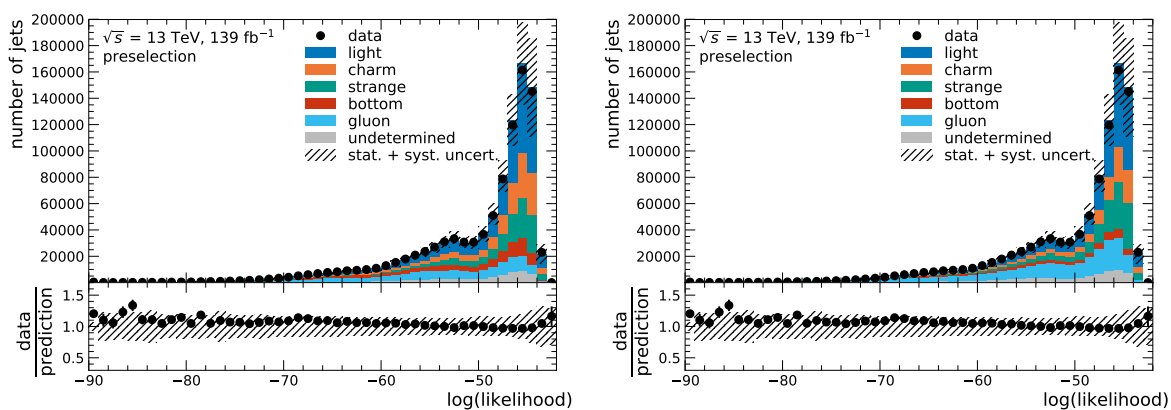
Fig. 7.7: p_T distribution of the jet leading in p_T after applying the preselection in the calibration regions, selecting mainly events containing semileptonic $t\bar{t}$ decays. The hashed band represents both the statistical and systematic uncertainties. The bottom panel shows the ratio between data and prediction.

The event topology given by one t quark decaying hadronically and the other decaying leptonically is reconstruction with the *KLFitter* tool [153] by considering the kinematic information of the lepton, the four jets, and E_T^{miss}. It determines the best matches between reconstructed jets and the quarks in the $t\bar{t}$ decay by maximizing a likelihood consisting of Breit-Wigner functions of the t quark and W boson mass distributions and transfer functions describing the energy response of the ATLAS detector. Further information on the algorithm can be found in Reference [153]. In the setup used, the t mass is fixed to 172.5 GeV and no b-tagging information is considered. In the following, the jet matched to the b quark of the leptonically decaying t quark will be called j_b^{lep}, the jet matched to the b quark of the hadronically decaying t quark will be called j_b^{had}, and the jets matched to the quarks from the W boson decay will be called j_{q1} and j_{q2}, where j_{q1} is the jet with the higher p_T of the two.

The logarithm of the highest likelihood per event given the best permutation is shown in Figure 7.8, the histograms being subdivided according to the parton-label flavors of j_{q1} and j_{q2}. In both figures, the expected ratio of 2:1:1 for l-, s-, and c-jets as given by the CKM matrix elements moderating the W boson decays can be observed. The additional g- and b-jets are the result of imperfect reconstruction of the $t\bar{t}$ topology for some events. As b-jets from t decays tend to have larger momenta, they are more often found contributing to the distribution of the jets with the larger p_T matched to the quarks from the W boson, while the g-jets contribute more to the distribution of jets with the smaller p_T matched to the quarks from the W decay.

In selections striving to enrich the sample of jets with jets originating from quarks, to increase the probability that the jets matched to the quarks from the W boson are matched correctly, the logarithm of the likelihood of the kinematic fit is required to be larger than -48, which increases the overall number of j_{q1} from light, s, and c quarks in these jets. In contrast, j_{q2} is likely to be a g-jet if it was wrongly matched to a quark from the W boson decay. Therefore, a requirement on the logarithm of the likelihood to be smaller than -57

(a) The flavors shown are the parton labels of the jets associated with the decay of the W boson in semileptonic $t\bar{t}$ decays that have the larger p_{T}.

(b) The flavors shown are the parton labels of the jets associated with the decay of the W boson in semileptonic $t\bar{t}$ decays that have the smaller p_{T}.

Fig. 7.8: Logarithm of the likelihood of the best jet permutation calculated using the *KLFitter* tool after applying the preselection for the calibration region selecting mainly events with semileptonic $t\bar{t}$ decays. The hashed band represents both the statistical and systematic uncertainties. The bottom panel shows the ratio between data and prediction.

is applied in the selection used to enrich a sample with jets from g. Both requirements are illustrated in Figure 7.9, showing that the first requirement yields a mass distribution for the reconstructed hadronically decaying W boson which is centered around the actual mass of the W boson, while the second requirement rejects the larger part of events with a reconstructed hadronically decaying W boson in the same mass range. This strategy is inspired by the strategy used in Reference [154] which uses the same requirements on the logarithm of the likelihood ($\log(\text{likelihood}) > -48$) to increase the number of c-jets in semileptonic $t\bar{t}$ decays.

In the following, all additional selection requirements for the five flavor regions are described. A summary is given in Table 7.1.

	Enriched Region				
	Light	**Strange**	**Charm**	**Bottom**	**Gluon**
$\log(\text{likelihood})$	> -48	> -48	> -48	> -48	< -57
j_b^{lep}	b-tagged	b-tagged	b-tagged	***b-tagged***	b-tagged
j_b^{had}	b-tagged	b-tagged	b-tagged	***b-tagged***	b-tagged
j_{q1}	***not b-tagged***	***not b-tagged***	***c-tagged***		
j_{q2}	not c-tagged	c-tagged	not b-tagged		***not b-tagged***

Tab. 7.1: Selection requirements applied in order to enrich the selected jets in the flavors given in the top row of the table. The jets which are selected for the flavor enriched region are highlighted in bold cursive print.
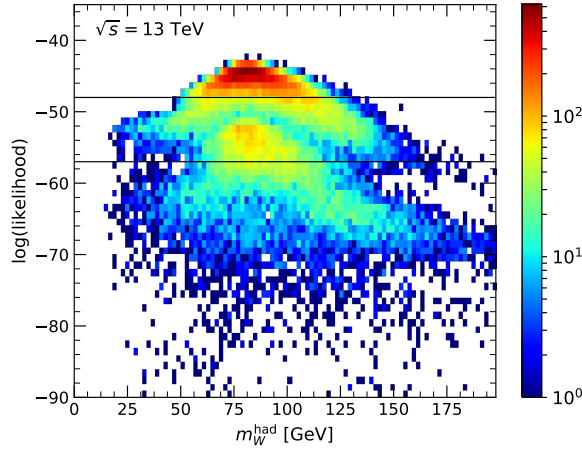
Fig. 7.9: Mass of the reconstructed hadronically decaying W boson and the logarithm of the likelihood value of the best permutation in the kinematic likelihood fit used to match jets to quarks in the event topology of semileptonically decaying $t\bar{t}$ in data.

7.2.1. Bottom-Enriched Region

In the region enriched in b-jets, both of the jets matched to b quarks from a t quark decay are required to be b-tagged. No further requirements are imposed on the other jets. Both b-tagged jets are used for calibration in the bottom-enriched calibration region.

(a) Transverse momentum of jets grouped by their parton labels.

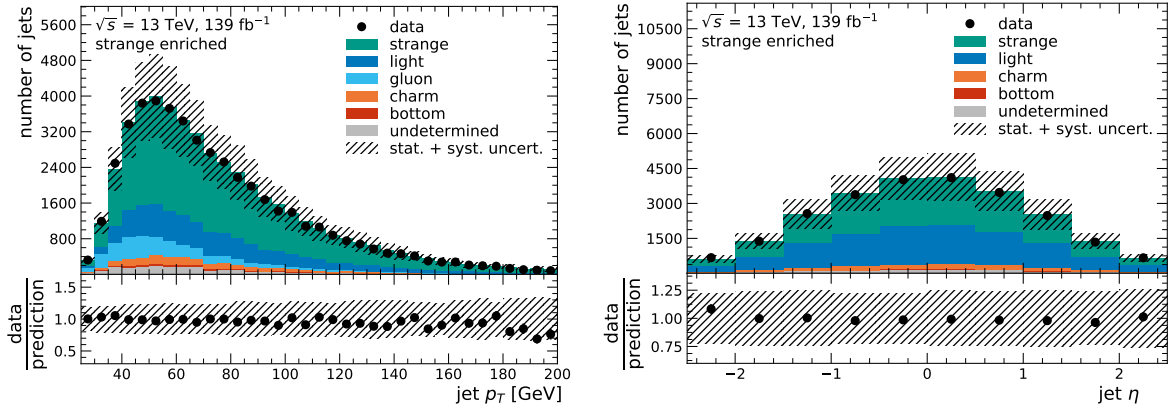(b) Pseudorapidity of jets grouped by their hadron labels.
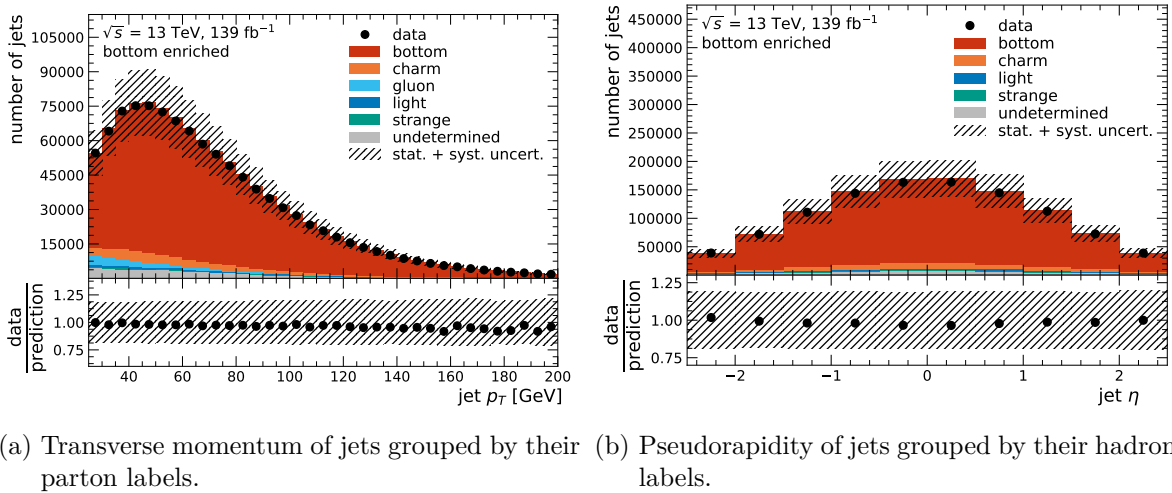
Fig. 7.10: Kinematic distribution of the jets in the region enriched with jets from b quarks. The hashed band represents both the statistical and systematic uncertainties. The bottom panel shows the ratio between data and prediction.

Figures 7.10a and 7.10b show the p_{T} and η distributions, respectively, of the jets in the region enriched with b-jets. For the p_{T} distribution, the parton label is used to group the jets, while for the η distribution, the hadron label is used to give an impression of the similarities and differences between these two labels in the bottom-enriched region. The majority of jets are b-jets, with small contributions from c- and g-jets. In these two distributions, data and

prediction agree well within the statistical and systematic uncertainties, which even appears to be overestimated.

7.2.2. Charm-Enriched Region

In the charm-enriched calibration region, both j_b^{had} and j_b^{lep} are required to have a b-tag. j_{q1} is required to be c-tagged, while j_{q2} is required not to have b-tag. In the charm-enriched region, only the c-tagged j_{q1} is used for the calibration. If an event is part of the charm-enriched region, it will not be part of the other calibration regions to ensure orthogonality between the regions.
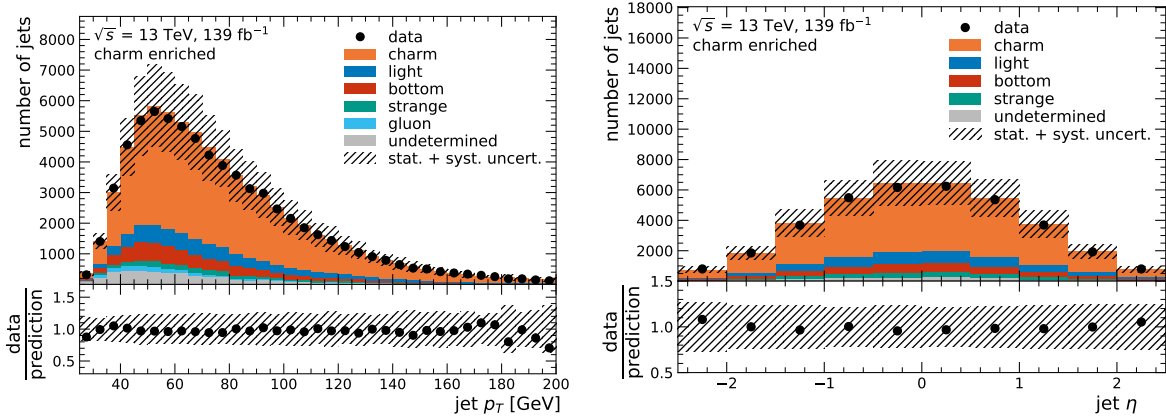
(a) Transverse momentum of jets grouped by their parton labels.

(b) Pseudorapidity of jets grouped by their hadron labels.

Fig. 7.11: Kinematic distribution of the jets in the region enriched with jets from c quarks. The hashed band represents both the statistical and systematic uncertainties. The bottom panel shows the ratio between data and prediction.

Figures 7.11a and 7.11b illustrate the p_{T} and η distributions, respectively, of the jets in the region enriched in c-jets. The majority of jets in this region are c-jets, however, the contribution of jets of other flavors is significantly larger than in the region enriched in b-jets. The largest background contribution is given by l-jets and b-jets. In both distributions shown, data and prediction agree within the statistical and systematic uncertainties.

7.2.3. Strange-Enriched Region

In the strange-enriched calibration region, both j_b^{had} and j_b^{lep} are required to have a b-tag. While the probe jet j_{q1} is required to have no b-tag, j_{q2} is required to have a c-tag, making use of the fact that most hadronically decaying W bosons decay either as $W^+ \to c\bar{s}$ or $W^+ \to u\bar{d}$.

Figures 7.12a and 7.12b show the p_{T} and η distributions, respectively, of the jets in the region enriched in s-jets. The selection ensures that the majority of jets are parton-labeled s-jets. However, the region contains considerable amounts of l-jets and g-jets as well. When considering the hadron label, the number of jets classified as s-jets is significantly smaller. In both distributions, data and prediction agree within statistical and systematic uncertainties.
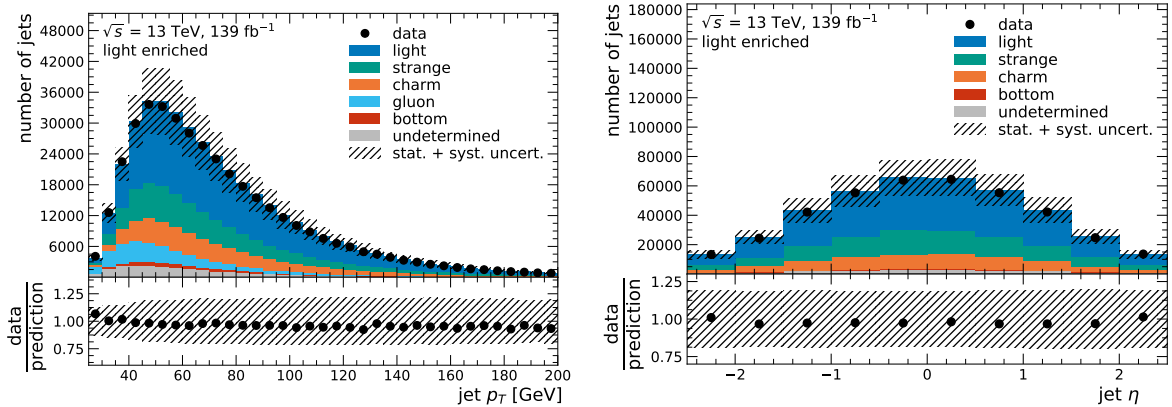
(a) Transverse momentum of the jets grouped by their parton labels.
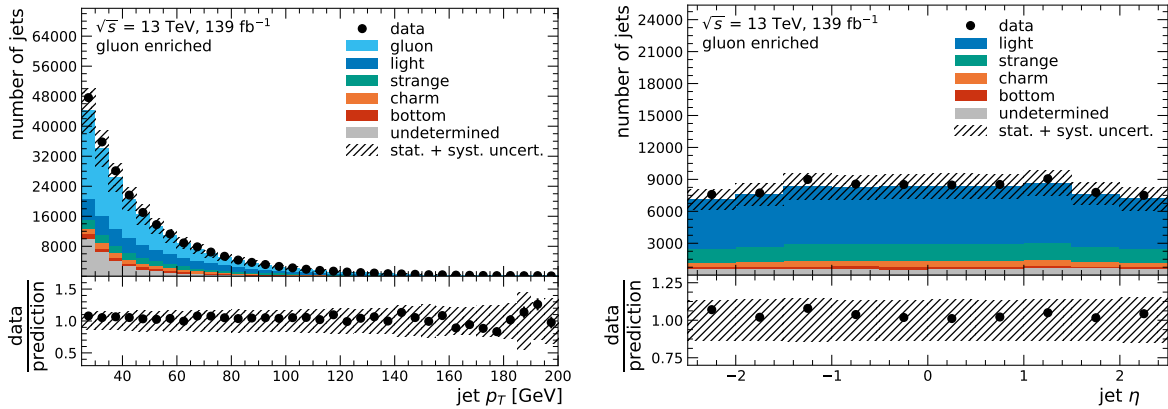
(b) Pseudorapidity of jets grouped by their hadron labels.

Fig. 7.12: Kinematic distribution of the jets in the region enriched with jets from strange quarks. The hashed band represents both the statistical and systematic uncertainties. The bottom panel shows the ratio between data and prediction.

7.2.4. Light-Enriched Region

In the region enriched with l-jets, again, both j_b^{had} and j_b^{lep} are required to have b-tags. Additionally, the probe jet j_{q1} is required not to be b-tagged and j_{q2} is required not to be c-tagged, attempting to reject $W^+ \to c\bar{s}$ decays.

(a) Transverse momentum of the jets grouped by their parton labels.

(b) Pseudorapidity of jets grouped by their hadron labels.

Fig. 7.13: Kinematic distribution of the jets in the region enriched with jets from up or down quarks. The hashed band represents both the statistical and systematic uncertainties. The bottom panel shows the ratio between data and prediction.

Figures 7.13a and 7.13b show the p_{T} and η distributions, respectively, of the jets in the region enriched in l-jets. While in this region, the jet flavor with the largest contribution is the light flavor, it does not account for the absolute majority of all jets, especially when considering the parton label. The second largest contributions are given by s-jets and c-jets,

followed by a considerable amount of g-jets, especially for lower p_{T}. Again, for both p_{T} and η distributions, data and prediction events agree within statistical and systematic uncertainties.

7.2.5. Gluon-Enriched Region

In the gluon-enriched region, j_{q2} is considered as the probe jet since the subleading jet matched to a quark from the W boson is more likely to stem from gluon radiation. Both j_b^{had} and j_b^{lep} are required to have a b-tag, while the probe jet j_{q2} is required not to be b-tagged. No b-tagging requirement is applied to j_{q1}.

(a) Transverse momentum of the jets grouped by their parton labels.

(b) Pseudorapidity of jets grouped by their hadron labels.

Fig. 7.14: Kinematic distribution of the jets in the region enriched with jets from gluons. The hashed band represents both the statistical and systematic uncertainties. The bottom panel shows the ratio between data and prediction.

Figures 7.14a and 7.14b show the p_{T} and η distributions, respectively, of the jets in the region enriched in g-jets. In this region, the majority of jets have a gluon parton label. However, in this region, there is a significant contribution of events without a determined jet flavor. These jets stem either from events containing a fake lepton whose contribution is estimated using a data-driven method, or they are simulated but do not have a parton label. In this region, the majority of jets have a relatively low p_{T} compared to the other regions. The overall number of jets, however, is large compared to the number of jets in the regions enriched with s- and c-jets. In both distributions, the difference between data and prediction is covered well by the combination of statistical and systematic uncertainties.

7.3. Data Driven Background Estimates

In rare cases, events only containing jets can be misreconstructed to contain electrons or muons. This happens either because the properties of a jet resemble those of an electron, or because non-prompt muons from decays inside the jet are reconstructed as prompt muons. While the rejection of such misreconstructed objects using the identification criteria of electrons and muons is very effective, due to the large cross section of hadronic processes in pp collisions, a background contribution from events containing fake leptons is almost inevitable. However, be-

cause the rejection rates are large, such background contributions usually become significantly smaller with every lepton required. Hence, in this analysis, an estimate of this background is only necessary for the calibration requiring one lepton and not for the measurement of $|V_{ts}|$ and $|V_{td}|$ requiring two leptons.

The efficient rejection of fake leptons necessary for most analyses makes the estimation of this background from simulation almost impossible, since most events generated would be rejected, leading to an almost impossible large number of simulation events necessary for a sound background estimation.

This issue is most often circumvented by using data-driven estimates instead. The method applied in this thesis is the *matrix method* [155]. It makes use of different grades of identification: Loose identification has a larger efficiency for both real and fake leptons but a smaller purity of real leptons; tight identification has a smaller efficiency for both real and fake leptons but a larger purity for real leptons. Therefore, if loose identification criteria are applied on a sample, the number of fake leptons is increased. It is then possible to determine the number of leptons passing tight identification criteria based on the number of leptons passing loose criteria (cf. e.g. Reference [156]).

The number of all events with loose leptons N^l and with tight leptons N^t can be split into those events containing real leptons N_r and those containing fake leptons N_f:

$$N^l = N_r^l + N_f^l, \tag{22}$$

$$N^t = N_r^t + N_f^t, \tag{23}$$

$$= \varepsilon_r N_r^l + \varepsilon_f N_f^l, \tag{24}$$

with the efficiencies

$$\varepsilon_r = \frac{N_r^t}{N_r^l} \text{ and } \varepsilon_f = \frac{N_f^t}{N_f^l} \tag{25}$$

for tightening the requirements on the leptons from loose to tight. These equations can be rearranged to express the number of fake leptons that fulfill the tight criteria by the total number of loose and the total number of tight leptons and the efficiencies ε_r and ε_f:

$$N_f^t = \frac{\varepsilon_f}{\varepsilon_r - \varepsilon_f} \left(\varepsilon_r N^l - N^t \right). \tag{26}$$

In practice, to estimate the background contribution from fake leptons, all data events passing the loss selection criteria are used, and a weight

$$w = \frac{\varepsilon_f}{\varepsilon_r - \varepsilon_f} \left(\varepsilon_r - \delta \right) \tag{27}$$

is applied to them, where δ is either 1 for events that additionally pass the tight selection criteria or 0 otherwise.

The critical point of the matrix method is the determination of the efficiencies, especially the fake efficiencies, which depend on the region in which the matrix method is applied. The exact definition of loose and tight leptons was described previously in Section 4.

The real efficiencies are determined in simulated samples of Z+jets and $t\bar{t}$ events. These events are required to contain at least two electrons or muons, respectively, both having opposite electric charge and $p_T > 27\,\text{GeV}$. Their invariant mass is required to lie within a window
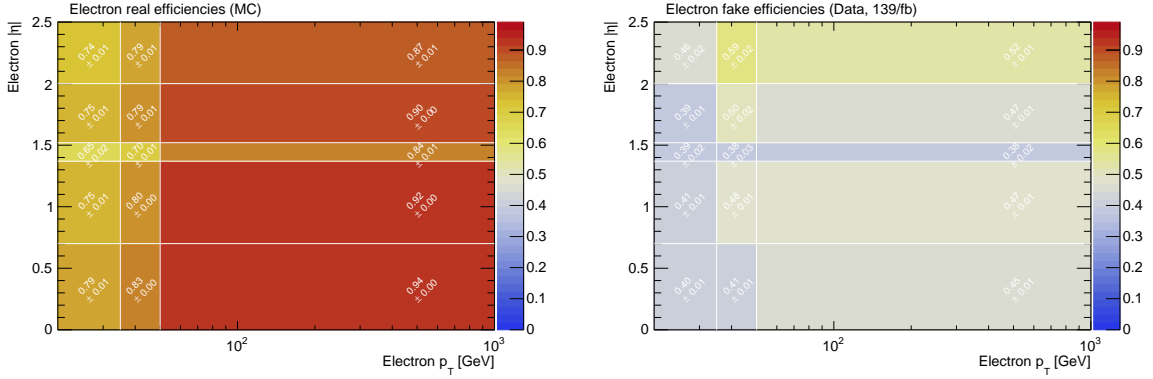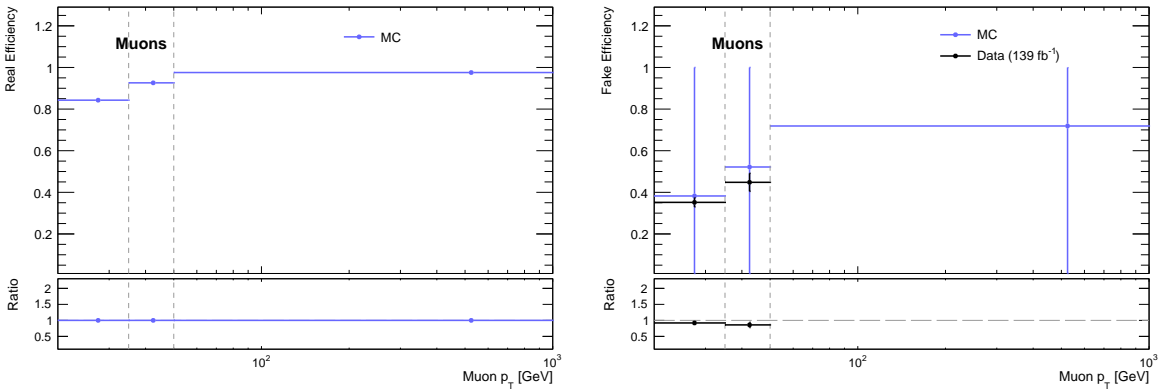
(a) Efficiencies (loose → tight) of real electrons derived in MC simulations using the tag-and-probe method.

(b) Efficiencies of fake electrons derived on data, subtracting contributions of real electrons as given by MC simulations.

Fig. 7.15: Efficiencies (loose → tight) of electrons as used in the fake estimate.

of 10% to the left and right of the Z mass. One lepton (the *tag*) is required to fulfill the tight criteria, while the other lepton (the *probe*) is required to fulfill the loose criteria and to trigger a trigger used in the selection of semileptonic $t\bar{t}$ decays. Additionally, four jets with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ have to be present and at least two jets have to be b-tagged.

The fake efficiencies are determined in data. However, since data contains real contributions as well, these are subtracted as given in simulations ($t\bar{t}$, single t, V+jets, diboson, $t\bar{t} + X$). The selection applied to both data and MC simulations requires one lepton (either electron or muon), $E_{\mathrm{T}}^{\mathrm{miss}} < 10\,\mathrm{GeV}$, at least four jets with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and exactly two of them being b-tagged. This selection is similar to the preselection applied for the calibration, however, the

(a) Efficiencies of real muons derived in MC simulations using the tag-and-probe method.

(b) Efficiencies of fake muons derived on data subtracting contributions of real muons as given by MC simulations and the same efficiency derived from MC simulations. For the last (empty) bin, the efficiency of the middle bin is used in the background estimation.

Fig. 7.16: Efficiencies (loose → tight) of muons as used in the fake estimate.

alternative requirement on the $E_{\mathrm{T}}^{\mathrm{miss}}$ increases the number of fake leptons as events containing only jets tend to be balanced in their p_T because they do not contain neutrinos detected as $E_{\mathrm{T}}^{\mathrm{miss}}$.

Figure 7.15 shows the efficiencies for real and fake electrons as used in the calibration and Figure 7.16 shows them for muons. The real and fake efficiencies are typically determined in bins of $|\eta|$ and p_{T}. However, because the number of fake muons is much smaller than the number of fake electrons, the muon efficiencies are only binned in p_{T}. Because of the small number of fake muons for high p_{T} and because the modeling of the real contributions is not good enough, it is not possible to derive fake efficiencies for a muon $p_{\mathrm{T}} > 50\,\mathrm{GeV}$. Therefore, as an approximation, the efficiency derived in the adjacent bin $35\,\mathrm{GeV} < p_{\mathrm{T}} < 50\,\mathrm{GeV}$ is used.

7.4. The Calibration Method

In this section, the method used for the calibration of the *s*-tagger is discussed. First, profile-likelihood fits, which are used to estimate the flavor composition of jets with and without an *s*-tag, are introduced in Section 7.4.1. After that, in Section 7.4.2, the strategy to measure the efficiencies and mis-tag rates of the *s*-tagger for all jet flavors is discussed. In Section 7.4.3, the systematic uncertainties considered in the calibration are described.

7.4.1. Profile-Likelihood Fits

In this thesis, so-called *profile-likelihood fits* are used to estimate parameters of a given model, such as e.g. the flavor composition of the jets used in the calibration of the *s*-tagger. Profile-likelihood fits are especially suitable for this task as they are able to include systematic uncertainties in the form of nuisance parameters in the fit. In some cases, if enough data points are provided, the fit can even constrain these uncertainties given the available data, effectively reducing the uncertainty on the estimated parameters.

If $\vec{\theta}$ are the parameters of interest in a statistical analysis and $f(x|\vec{\theta})$ is the probability density function to measure x given the values $\vec{\theta}$, then $\mathcal{L}$ is the likelihood of the parameters $\vec{\theta}$ given the measured value x, defined as

$$\mathcal{L}(\vec{\theta}|x) = f_\theta(x)\,. \tag{28}$$

For more than one measured value x_i, $i \in 1, 2, \ldots$, together denoted as $\vec{x}$, the likelihoods of each measurement multiply so that

$$\mathcal{L}(\vec{\theta}|\vec{x}) = \prod_{i=0}^{N} f(x_i|\vec{\theta})\,. \tag{29}$$

The most probable values of the parameters of interest $\vec{\theta}$ considering the measured values $\vec{x}$ can be found by maximizing the likelihood. If the estimators are unbiased, the uncertainty of all fitted parameters is given by a covariance matrix, which is the inverse of the Hessian of the Lagrangian. Alternatively, assuming that the likelihood follows a Gaussian distribution around its maximum, the uncertainties of the parameters are given by its standard deviation σ and fulfill $\ln L(\theta_{\max} \pm \sigma) = \frac{1}{2} \ln L(\theta_{\max})$.

Uncertainties on the model can be taken into account during the optimization by including nuisance parameters (NPs) in the likelihood, as will be discussed in the following.

In this thesis, all profile-likelihood fits are implemented using the *HistFactory* software package [157] interfaced by the *TRExFitter* software package version 4.12 [158]. The likelihood implemented in it has a form of

$$
\mathcal{L}(\phi_f, \alpha_p, \gamma_i | n_i, a_p) = \prod_i \mathrm{Pois}(m_i | \gamma_i \tau_i) \mathrm{Pois}\left(n_i \middle| \gamma_i \prod_f \phi_f \eta_f(\vec{\alpha}) \sigma_{i,f}(\vec{\alpha}) \nu_i \right) \prod_p \mathrm{Gaus}(a_p | \alpha_p) ,
$$
(30)

where n_i is the number of data events in bin i and a_p are the values used to parameterize the probability of the NPs. ϕ_f, α_p, and γ_i are the free parameter of the profile-likelihood fit: ϕ_f are either free-floating parameters of interest or constant normalization parameters of a subgroup of processes f, α_p are the values the NPs take on, and γ_i are normalization factors of bin i.

The latter γ_i are primarily necessary because the number of simulated events available are finite, which means that the actual number of unweighted simulated events m_i per bin i is expected to take on any value according to the probability of a Poisson distribution with the central value $\gamma_i \tau_i$[3]. For lack of better knowledge, τ_i is set to m_i, assuming that the number of MC events is sufficiently large to not be significantly biased [159]. The number of unweighted events is extracted from the relative statistical uncertainty of the nominal MC estimate ν_i in bin i, $m_i = \left(\frac{\delta_i}{\nu_i} \right)^2$, with δ_i being the absolute uncertainty.

The number of events per bin i in data, n_i, are expected to be Poisson distributed around the true value which is supposed to be estimated by the fitted model. The number of events in each bin as described by the model is given by

$$
\gamma_i \prod_f \phi_f \eta_f(\vec{\alpha}) \sigma_{i,f}(\vec{\alpha}) \nu_i ,
$$

where $\eta_f(\alpha_p)$ and $\sigma_{i,f}(\alpha_p)$ parameterize the uncertainties.

All uncertainties are separated into an overall normalization component $\eta_f(\alpha_p)$ and a per-bin shape component $\sigma_{i,f}(\alpha_p)$. The shape component is interpolated linearly such that

$$
\sigma_{i,f}(\alpha_p) = \sigma_{i,f}^0 + I_{i,f,\mathrm{lin}} \quad \text{with} \quad I_{i,f,\mathrm{lin}} = \begin{cases} \alpha_p(I_{i,f}^+ - I_{i,f}^0) & \alpha_p \geq 0 \\ \alpha_p(I_{i,f}^0 - I_{i,f}^-) & \alpha_p < 0 \end{cases} .
$$
(31)

$I_{i,f}^0$, $I_{i,f}^+$, and $I_{i,f}^-$ are the nominal yield, the yield of the up, and the yield of the down variation, respectively. The normalization component is interpolated using a combination of an exponential form and a polynomial form:

$$
\eta_f(\alpha_p) = I_{f,\mathrm{exp}} \quad \text{with} \quad I_{f,\mathrm{exp}} = \begin{cases} (I_f^+/I_f^0)^{\alpha_p} & \alpha_p \geq 1 \\ 1 + \sum_{k=1}^{6} a_k \alpha_p^k & |\alpha_p| < 1 \\ (I_f^-/I_f^0)^{-\alpha_p} & \alpha_p \leq -1 \end{cases} .
$$
(32)

The a_k are determined by matching $d\eta/d\alpha_p \big|_{\alpha_p = \pm a_p}$ and $d^2\eta/d\alpha_p^2 \big|_{\alpha_p = \pm a_p}$ to the exponential interpolations. Details on both interpolations can be found in Reference [157].

---

[3]To be precise, this normalization factor γ_i would have to be derived per simulated sample. However, to reduce computational overhead, this simplification looking only at the overall number of simulated events is used unless stated otherwise.

The probabilities of the NPs α_p are set to follow a Gaussian distribution $\text{Gaus}(a_p|\alpha_p)$ with the center at 0 and a width $a_p = 1$.

To minimize the profile likelihood, the `MIGRAD` algorithm [160] is used. It is based on the minimization approach developed by Davidon [161], Fletcher [162], and Powell [163] in which an iterative search for a function's minimum is performed along the direction of its gradient. The covariance matrix, initially assumed to be unitary, is used in and adjusted during the minimization process. With the final covariance matrix, the fit provides an estimate of the correlations between the parameters of the function.

7.4.2. Measurement of Flavor Compositions and Efficiency Calculation

For the measurement of the *s*-tagging efficiencies and mis-tag rates, the five flavor-enriched regions are divided into 11 bins, two $|\eta|$ bins ($|\eta| \in [0.0, 1.5]$ and $(1.5, 2.5)$) times either seven p_T bins ($p_T \in [25, 50]$, $(50, 60]$, $(60, 70]$, $(70, 90]$, $(90, 120]$, $(120, 150]$, and $(150, 200]$ GeV) in the central region and four p_T bins in the forward region ($p_T \in [25, 50]$, $(50, 70]$, $(70, 120]$, and $(120, 200]$ GeV). The $|\eta|$ bins were chosen to depict the split between barrel and end-cap in the calorimeters situated roughly at $|\eta| = 1.5$ as can be seen in Figure 7.17.
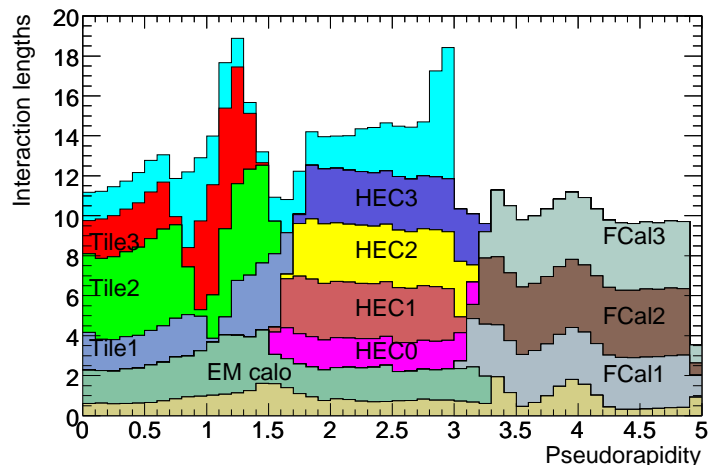
Fig. 7.17: Distribution of the material in the ATLAS detector in $|\eta|$, particularly in the calorimeter layers [44].

The p_T bins were chosen to even out varying numbers of jets across the p_T distributions of all flavor-enriched regions. For central η, this demands a larger p_T bins for low p_T due to the small numbers of jets in the light-, strange-, and charm-enriched regions, and also for larger p_T due to the small number of jets in the gluon-enriched region. For more forward η, to reduce statistical fluctuations in the resulting calibration, it was necessary to merge the higher-p_T bins pairwise, resulting in only four p_T bins.

The measurement is done independently in each p_T and $|\eta|$ bin and follows roughly the same strategy the p_T^{rel} calibration method uses [65]. It considers the momentum component of muons reconstructed inside a jet that is perpendicular to the jet axis in order to determine the contribution of each jet flavor. Similarly, in this calibration, the distribution of flavors across all flavor-enriched regions is used to determine the flavor composition of the samples.

For this, it is assumed that the number of jets in each region i in data, N_i, is the sum over the number of jets of a given flavor f in this region multiplied by a flavor-dependent correction

(a) Pre-fit distribution.
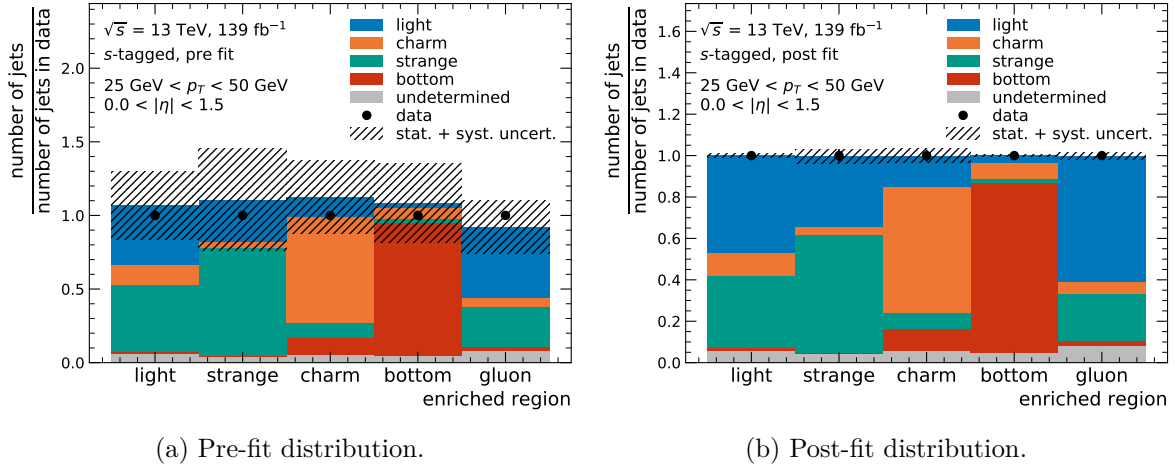
(b) Post-fit distribution.

Fig. 7.18: Distribution of s-tagged jets arranged by their hadron-based flavors in all flavor-enriched calibration regions for $25\,\text{GeV} < p_\text{T} < 50\,\text{GeV}$ and $|\eta| \le 1.5$, pre- and post-fit. The distributions show data and prediction normalized to the data yields in each bin. The hashed bands represent both the statistical and systematic uncertainties of the prediction.

factor c_f that is the same across all regions:

$$N_i = \sum_{j \in \text{flavors}} c_j n_i^j \, . \tag{33}$$

This correction factor can be determined for both tagged and untagged jets using profile-likelihood fits described in the previous section. Figure 7.18 illustrates the use of distribution of jet flavors across the flavor-enriched regions in the determination of normalization factors. It shows the distributions before and after fitting data in an exemplary p_T and η bin for s-tagged jets in that bin. The fit is able to adjust the flavor fractions for a better agreement between data and prediction and to reduce the total uncertainty.

The flavor fraction for the flavor j in each region i after solving the equation is

$$f_j^i = \frac{c^j n_i^j}{N_i} \, . \tag{34}$$

The tagging efficiency for each flavor is derived in its name-sake region, where, in principle, this flavor is most abundant. The efficiency/mis-tag rate of a flavor j is defined as

$$\varepsilon_j^\text{data} = \frac{f_j^\text{tag} N_\text{data}^\text{tag}}{f_j^\text{tag} N_\text{data}^\text{tag} + f_j^\text{tag} N_\text{data}^\text{tag}} \, . \tag{35}$$

The SFs to be applied to all jets in MC simulations when applying the s-tagger are determined by dividing the efficiency in data by the efficiency in MC simulations:

$$\kappa_j = \frac{\varepsilon_j^\text{data}}{\varepsilon_j^\text{MC}} \, . \tag{36}$$

7.4.3. Estimation of Systematic Uncertainties

The systematic uncertainties associated with the calibration are usually estimated in simulations and present themselves as either two-sided (up and down) or one-sided variation with respect to the nominal input. They are taken into account as NPs in the previously discussed profile-likelihood fit. Since for them, a Gaussian probability distribution is assumed, the variations per bin i are symmeterized according to

$$\sigma_i^{\text{symmetric}} = \begin{cases} \frac{1}{2}\left(\left|x_i^{\text{up}} - x_i^{\text{nominal}}\right| + \left|x_i^{\text{down}} - x_i^{\text{nominal}}\right|\right) & \text{for a two-sided variation and} \\ \left|x_i^{\text{variation}} - x_i^{\text{nominal}}\right| & \text{for a one-sided variation,} \end{cases} \tag{37}$$

where x_i^{nominal} is the process's nominal bin content and x_i^{up}, x_i^{down}, and $x_i^{\text{variation}}$ are the bin content of a given variation. The resulting symmetric variations are

$$x_i^{\text{up, symmetric}} = x_i^{\text{nominal}} + \text{sign}(x_i^{\text{up/variation}} - x_i^{\text{nominal}})\sigma_i^{\text{symmetric}}, \tag{38}$$

$$x_i^{\text{down,symmetric}} = x_i^{\text{nominal}} - \text{sign}(x_i^{\text{up/variation}} - x_i^{\text{nominal}})\sigma_i^{\text{symmetric}}, \tag{39}$$

were the sign of the variation (up or down) w.r.t. to other bins is preserved. The uncertainties are symmeterized before they are split into a normalization and shape component.

As no continuous distributions are considered in the calibration, no smoothing is applied to the input. In each region, normalization components that have a variation that is smaller than 1% of the nominal event yield and shape components that have a variation smaller than 1% are removed (*pruned*) prior to the fit. All uncertainties are detailed in the following.

7.4.3.1. Modeling Uncertainties

There are several different uncertainties associated with the modeling of simulated processes. Their description here follows Reference [74].

The cross sections of all simulated processes (except for diboson production) are corrected to a theoretical value of better accuracy than what is used in the ME calculation of their MC simulation. The uncertainties of all cross sections are given in Section 5.2 and used as uncertainties on the processes' normalization.

The uncertainty on the proton's PDF is estimated by evaluating all 100 variations of the NNPDF3.0 [85] PDF set. These 100 variations are gathered into a single two-sided uncertainty by determining the standard deviation of all 100 variations in each bin. This variation is applied in the analysis by varying the nominal sample up and down by this standard deviation.

The uncertainty on the shower generation is only evaluated in the $tt(WbWb)$ samples and by exchanging the shower generator from PYTHIA8.230 [77] to HERWIG7.04 [99, 100] and using the resulting distributions as one-sided variation.

Effects of scale variations in $t\bar{t}$ and single t production are studied by simultaneously varying the factorization and renormalization scale, the h_{damp} parameter, and the hadronization tune. The variations are described in Reference [164]. The up variation increases the h_{damp} parameter by a factor of two to $3m_t$, sets $\mu_r = \mu_f = 0.5$ w.r.t. the nominal renormalization and factorization scale in the ME, and uses the *Var3cUp* tune [88] which increases the α_s^{ISR} coupling in the parton shower of the initial-state radiation (ISR). The down variation uses the same value for h_{damp} as the nominal sample ($1.5m_t$) while setting $\mu_r = \mu_f = 2.0$ w.r.t. the nominal scales and using the *Var3cDown* tune [88] which decreases α_s^{ISR} in the parton shower of the ISR.

The uncertainty of the final-state radiation (FSR) in $t\bar{t}$ and single t production is evaluated by varying the renormalization scale for final-state QCD emissions (α_s^{FSR}) by a factor of 0.5 and 2.0, respectively.

For V+jets and diboson processes, the renormalization and factorization scale are set to same value of $\mu_{r,f} = 0.5$ or $\mu_{r,f} = 2.0$ w.r.t. the nominal value at the same time or one to $\mu_{r,f} = 0.5$ and the other to $\mu_{r,f} = 2.0$. In each bin, the maximal and minimal number of events of all four variations is used as up and down variation, respectively.

Because a reweighting is applied to the $t\bar{t}$ samples to alter the kinematic properties of the t quarks from what is simulated in NLO accuracy to what it was predicted at NNLO accuracy [97], the kinematic distributions without the reweighting are used as one-sided systematic uncertainty w.r.t. to the reweighted distributions used as nominal input.

7.4.3.2. Object Uncertainties

Object uncertainties are the uncertainties associated with the reconstruction and identification of any object used in the analysis.

The systematic uncertainties of electrons have two components. One is the uncertainty on the determination of their energy, i.e. both the correction of the nominal energy value by a scale factor and the energy resolution given by the detector. They are both determined in $Z \to e^+e^-$ decays as described in Reference [54]. The other uncertainty is the uncertainty on the efficiency of the electrons used. This efficiency is split into four different efficiencies, the trigger efficiency, the reconstruction efficiency, the isolation efficiency, and the identification efficiency. They are derived in $Z \to e^+e^-$ as described in References [54, 165]. All six uncertainties are independently included in the fit.

For muons, the uncertainties on the momentum measurement are divided into six sub-components. The p_{T} of simulated muons is corrected by applying a scale correction as well as a smearing [56]. The nominal values of the corrections are determined in $Z \to \mu^+\mu^-$ decays [56] and the uncertainty on these corrections are given by independently varying the scale correction, the smearing of the ID tracks used for the reconstruction, and the smearing of the MS tracks. Additionally, charge-dependent uncertainties on the scale of the momentum are applied, both for the bias of the sagitta measurement due to misalignment of the MS and for the overall resolution of the sagitta measurement [166]. The uncertainty on the muons' total efficiency is divided into five efficiencies: the reconstruction efficiency, the identification efficiency, the isolation efficiency, the trigger efficiency, and the uncertainty on the track-to-vertex association. All of them are determined in $Z \to \mu^+\mu^-$ decays [57, 167]. The latter two are combined into one uncertainty. All are split into a systematic and a statistical uncertainty, which are included in the fit independently.

The $E_{\mathrm{T}}^{\mathrm{miss}}$ is calculated from all reconstructed and calibrated objects and additional soft energy contributions. The uncertainties of the reconstructed objects are propagated to $E_{\mathrm{T}}^{\mathrm{miss}}$. The additional uncertainty of the soft energy contributions has two components. A two-sided variation of the scale of soft contributions is estimated in events without any expected $E_{\mathrm{T}}^{\mathrm{miss}}$ (Z+0 jets events) by looking at the $E_{\mathrm{T}}^{\mathrm{miss}}$ component parallel to the p_{T} of the hardest object (Z). Two one-sided variations of the resolution are derived in the same events by considering the variance of the $E_{\mathrm{T}}^{\mathrm{miss}}$ components parallel or perpendicular to the hardest object. Further details can be found in Reference [70].

To describe the uncertainties of the calibration of jets, a total of 30 up and down variations are used as uncertainties. Details on them can be found in Reference [61] and the references

therein, but the calibration process and the resulting subcategories of uncertainties will be briefly discussed here: Four variations are used as uncertainty for the initial correction of the jets' momenta w.r.t. to pile-up conditions. In the following global sequential calibration procedure [61, 168, 169], another uncertainty w.r.t. to punch-through corrections arises, as well as three uncertainties in the context of the flavor composition (quark or gluon) of the jets. As the flavor composition of jets varies per analysis, the flavor composition was specifically derived in $t\bar{t}$ simulations for both the calibration and the measurement of $|V_{ts}|$ and $|V_{td}|$ after the preselection was applied. This composition was then propagated to the application of the correction factors. The resulting uncertainties have a component for the overall flavor composition and response as well as an additional uncertainty for b-jets. Another six uncertainties arise in the context of the η intercalibration of forward jets in dijet events. The 98 uncertainty components of the remaining in-situ calibration are reduced in the global reduction scheme to a total of 8 effective up and down variations of the jet energy scale (JES) and jet energy resolution (JER).

The uncertainty on the use of the JVT algorithm used for pile-up suppression [170] is the uncertainty of its calibration and provided as one two-sided variation.

The SFs for the application of b-tagging and c-tagging algorithms [63, 68] described in Sections 4.5 and 4.6 are applied per jet. Their variations are given as eigenvariations for the different jet flavors (b, c, and light) and applied separately for each tagger but simultaneously to all jets tagged by it. There are three variations for c-jets, four variations for light jets, nine variations for b-jets, and one variation for τ-jets.

7.4.3.3. Experimental Uncertainties

All uncertainties associated with the conditions of the pp collisions are summarized as experimental uncertainties here.

All simulated events are reweighted to agree in their distribution of the number of pile-up events with the profile observed in data. The uncertainty associated with this procedure is given by the variation of the additional scaling applied to data described in Section 5.2.

The uncertainty on the determination of the integrated luminosity as given in Section 5.1 is applied as a variation of the overall normalization of the event prediction.

7.4.3.4. Uncertainties Specific to the *s*-Tagger Calibration

While requiring b-tags for the jets j_b^{lep} and j_b^{had} reduces the contribution of $tt(WsWb)$ and $tt(WdWb)$ decays in the calibration to $\mathcal{O}(10^{-3})$ and $\mathcal{O}(10^{-6})$, respectively, it is still necessary to assume values of $|V_{ts}|$ and $|V_{td}|$. For this, the nominal value and the 1σ variations are chosen according to Equation 2.

For the estimate of the background induced by fake leptons, an overall normalization uncertainty of 50% is assumed, which results in good agreement between the background estimate and data within this uncertainty plus the statistical uncertainty. This rough approach is justifiable as fake induced background has only a relatively minor contribution. To account for the extrapolation of the fake efficiency of muons from $35\,\text{GeV} < p_\text{T} < 50\,\text{GeV}$ to $p_\text{T} > 50\,\text{GeV}$, an additional 50% uncertainty on said efficiency in the high-p_T bin of the fake efficiency of muons is applied and propagated to the calibration.

An additional uncertainty on the shape of the s-tagger output is applied. This is done because it cannot be ensured that the selections of the flavor-enriched regions – especially the

use of *b*- and *c*-taggers – does not bias the *s*-tagger response, and because there is an inherent variance in the definition of the jets flavors. This uncertainty is derived by comparing the *s*-tagger distributions of the jet flavor in each region.
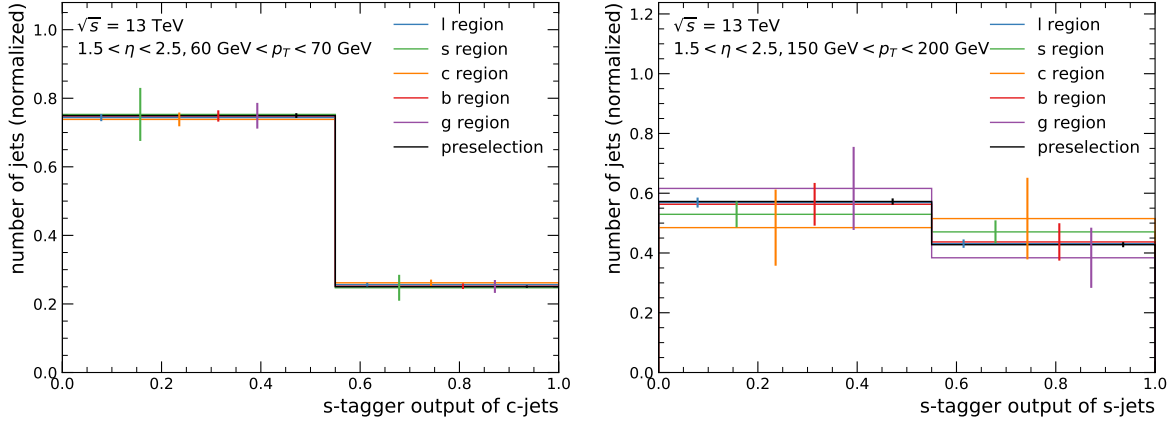
(a) *s*-Tagger response of *c*-jets for $1.5 < |\eta| < 2.5$ and $60\,\text{GeV} < p_\text{T} < 70\,\text{GeV}$.

(b) *s*-Tagger response of *s*-jets for $1.5 < \eta < 2.5$ and $150\,\text{GeV} < p_\text{T} < 200\,\text{GeV}$.

Fig. 7.19: Exemplary *s*-tagger responses after the preselection and in the different flavor-enriched regions. The error bars indicate the statistical uncertainty of the distribution in the given region. All regions enriched in a certain flavor are orthogonal to each other and a subset of the preselection region.

Figure 7.19 shows two example comparisons, one of an *s*-tagger response of *c*-jets with very little variation between regions and one response of *s*-jets on the other end of the spectrum, displaying a large variance in the observed responses. For all flavors, the largest absolute difference between the responses across all regions is taken as systematic uncertainty. To reduce the influence of statistical fluctuation, regions that contain less than ten simulated jets of the given flavor are not included.

7.5. Efficiencies, Mis-Tag Rates, and Scale Factors

In the following, the derived efficiencies of selecting *s*-jets and the mis-tag rates, i.e. the efficiency for all other jet flavors, when requiring an *s*-tagger output larger than 0.55 are discussed. The choice of this working point in light of the measurement of $|V_{ts}|$ and $|V_{td}|$ is described in Chapter 8.2.

Figure 7.20 shows the efficiencies for *s*-jet and mis-tag rates for *l*-, *c*-, and *b*-jet for $|\eta| \leq 1.5$.

The *s*-jet efficiency lies in a range between 25% and 50%, increases for increasing p_T, and has an absolute uncertainty between 6% and 15%. For $p_\text{T} < 60\,\text{GeV}$, it is significantly smaller than the efficiency seen in simulation. The mis-tag rate for *l*-jets is relatively constant across all p_T and lies in the order of 20%. It is very similar to the mis-tag rate determined in simulation and has an absolute uncertainty of up to 10%. Only for $p_\text{T} < 50\,\text{GeV}$, it is slightly larger than simulations suggest. The mis-tag rates for *c*-jets and *b*-jets lie in a range between 18% and 31%. They have an absolute uncertainty of up to 10% for *c*-jets and up to 4% for *b*-jets.

Overall, the distribution of the efficiencies and mis-tag rates across different jet p_T are relatively smooth, which suggests that the choice of bins is able to reduce the influence of
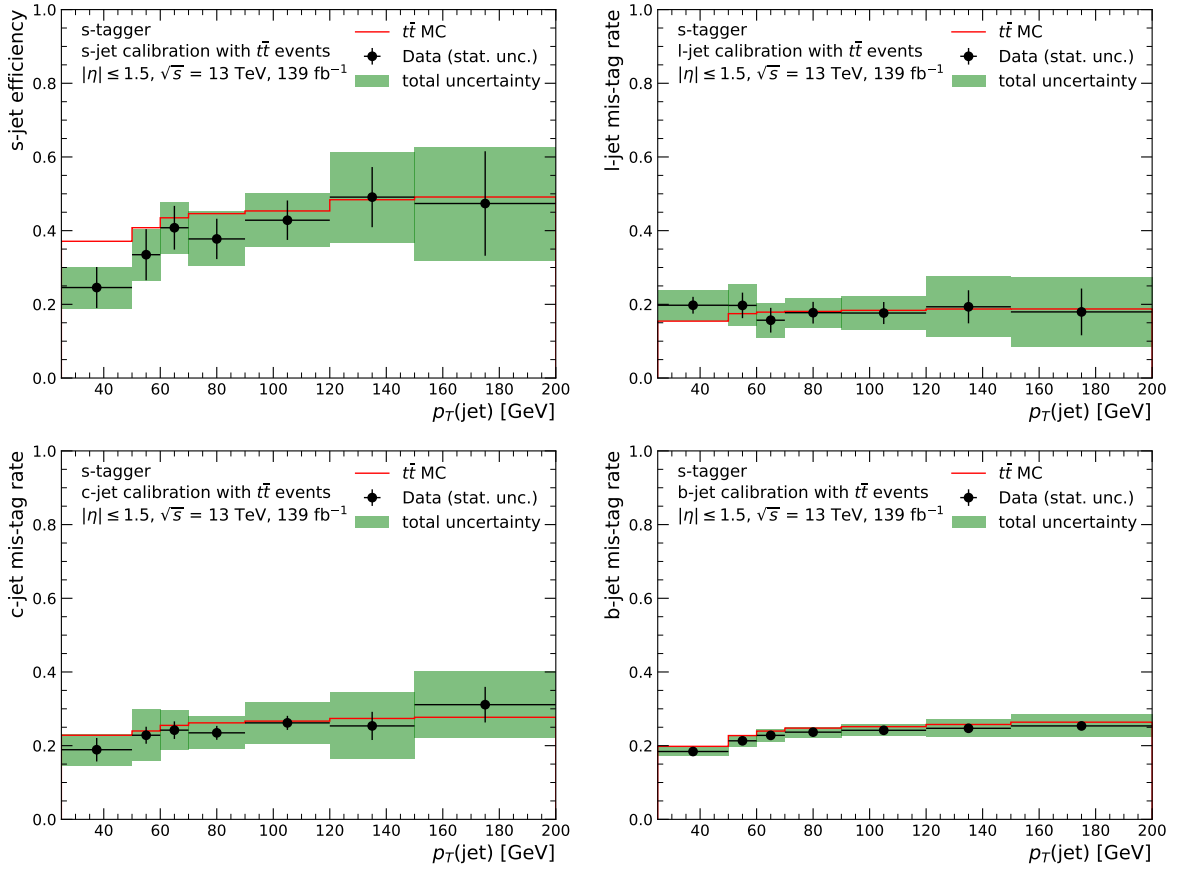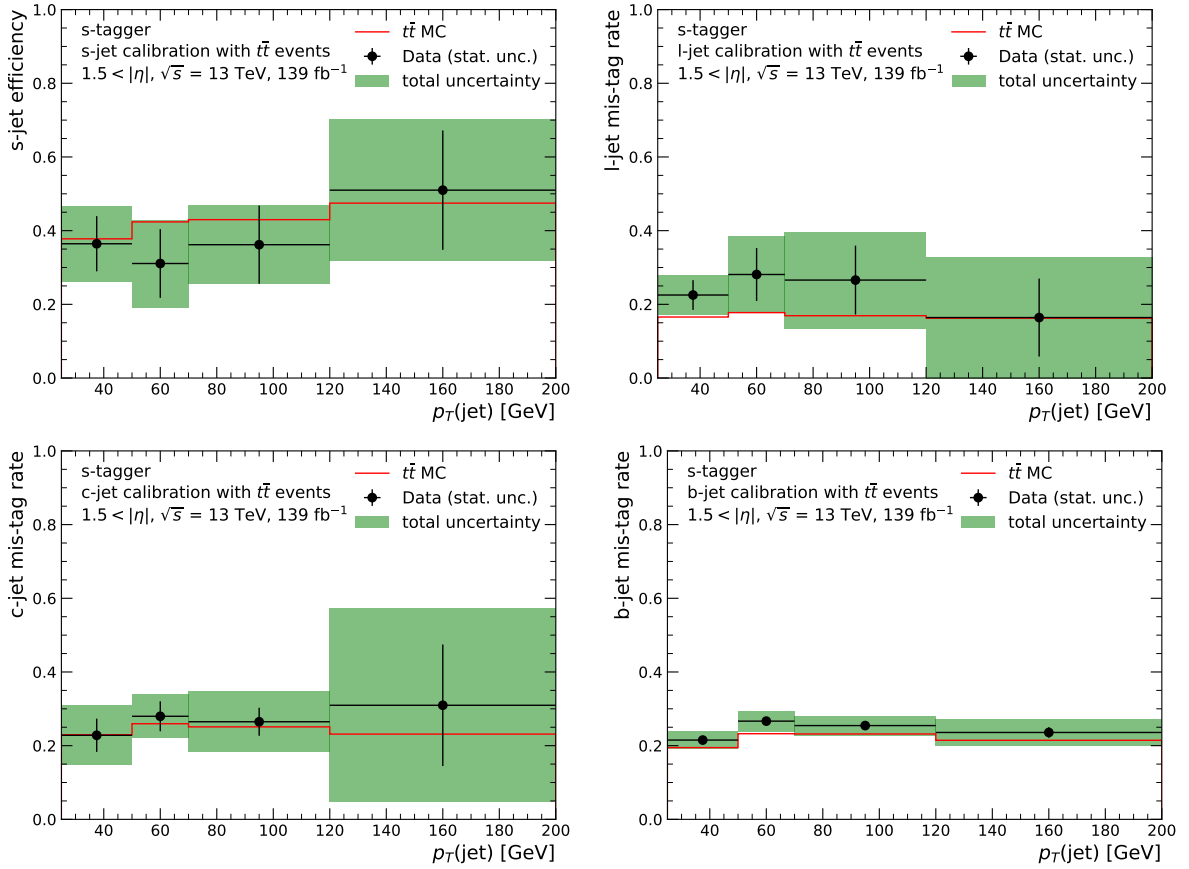
Fig. 7.20: Mis-tag rates for *l*-, *c*-, and *b*-jets and efficiencies of selecting *s*-jets when requiring an *s*-tagger output score larger than 0.55 for jets with $|\eta| \leq 1.5$. The efficiencies and mis-tag rates shown are determined both in MC simulated jets and in data. The error bars represent the statistical uncertainty and the green band represents both the statistical and systematic uncertainties.

statistical fluctuation in the determination of the efficiencies. For low p_{T}, the *s*-tagger is not able to distinguish between *s*-jets and jets of other flavors because of differences between data and the simulations used for the training of the *s*-tagger. For $p_{\mathrm{T}} \gtrsim 60\,\mathrm{GeV}$, however, the efficiencies of *s*-jets and mis-tag rates of jets of the other flavors are distinguishable.

Derived from the efficiencies and mis-tag rates, Figure 7.21 illustrates the SFs that will be applied to all jets with $|\eta| \leq 1.5$ if they are *s*-tagged. All SFs are compatible with one within statistical and systematic uncertainties except for those that are applied to *s*-jets with a low p_{T}.

Especially for *s*-jets and *l*-jets, the statistical uncertainty dominates the uncertainties for efficiencies, mis-tag rates, and SFs. This uncertainty can mainly be attributed to the lack of purity of jet flavors in the regions enriched with *g*-, *s*-, and *l*-jets and the limited number of flavor-enriched regions used in the calibration, resulting in the fit's inability to prediction the flavor composition across all regions.

The systematic uncertainties with the largest overall impact on the efficiencies and mis-tag rates are the uncertainty on the bias of the NN output due to the used flavor definition and selection criteria, as well as the uncertainties on *c*-tagging because of it is application in the
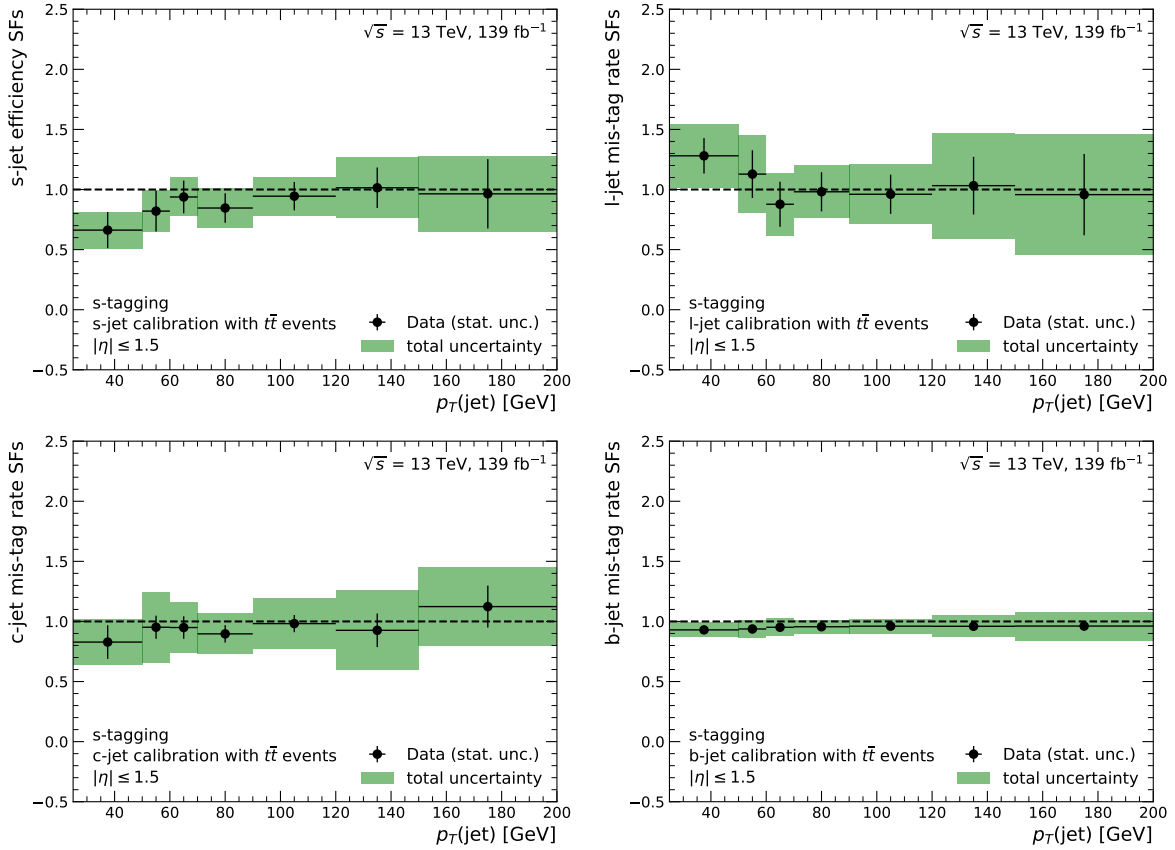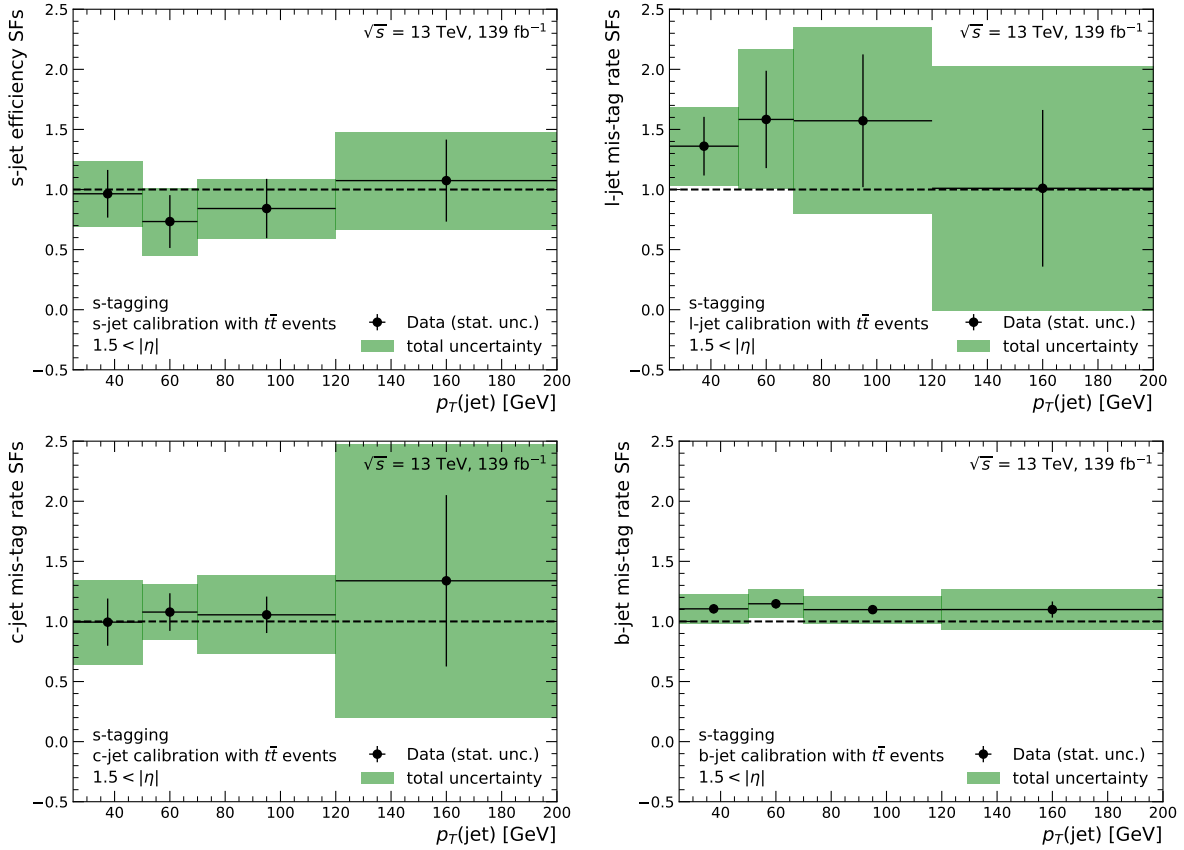
Fig. 7.21: Scale factors for *l*-, *c*-, *b*-, and *s*-jets with $|\eta| \leq 1.5$ and an *s*-tagger output score larger than 0.55. The error bars represent the statistical uncertainty and the green band represents both the statistical and systematic uncertainties.

definition of three of the five flavor-enriched regions.

Figure 7.22 shows the efficiencies for *s*-jets and mis-tag rates for jets of all other flavors for jet $|\eta| > 1.5$. The *s*-jet efficiency lies between 30% and 50% across the entire p_T spectrum, with absolute uncertainties of up to 20%. The mis-tag rates for *l*-jets lie in a range between 17% and 28% and have absolute uncertainties between 5% for low p_T and 16% for high p_T. For *c*-jets, the mis-tag rates lie between 24% and 31% with absolute uncertainties of up to 8% in the lower p_T regions and a large uncertainty of 26% for $p_T > 120\,\mathrm{GeV}$. The mis-tag rates of *b*-jets lie in a range between 22% and 28% and have the smallest absolute uncertainties of all flavors of up to 5%. All efficiencies and mis-tag rates are compatible with the efficiencies seen in simulation within their statistical and systematic uncertainties.

For jets with $|\eta| > 1.5$, the difference between the efficiencies for *s*-jets and jets of all other flavors are not significant, especially because the uncertainties of the *s*-jet efficiencies are large. Nonetheless, the central values of the mis-tag rates for *l*- and *b*-jets are smaller than the central values of the efficiencies for *s*-jets. *c*- and *s*-jets, on the other hand, do not have efficiencies and mis-tag rates that show any separation between them.

Figure 7.23 illustrates the SFs calculated from to the previously shown efficiencies and mis-tag rates for jets with $|\eta| > 1.5$. While they are compatible with one in all bins, due to the large uncertainties in the efficiencies and mis-tag rates, the uncertainties of the SFs reach up to 100%.

Fig. 7.22: Mis-tag rates for *l*-, *c*-, and *b*-jets and efficiencies of selecting *s*-jets when requiring an *s*-tagger output score larger than 0.55 for jets with $|\eta| > 1.5$. The efficiencies and mis-tag rates shown are determined both in MC simulated jets and in data. The error bars represent the statistical uncertainty and the green band represents both the statistical and systematic uncertainties.

Again, especially for *s*-jets and *l*-jets, the statistical uncertainty dominates. The systematic uncertainties with the largest impact overall are the uncertainty of the bias of the NN output as well as the uncertainty on the *c*-tagging used.

Because the uncertainties on the SFs are so large, it becomes very important to properly include correlations between the SFs when applying them in any analysis. These correlations will be discussed in the following section.

Fig. 7.23: Scale factors for *l*-, *c*-, *b*-, and *s*-jets with $|\eta| > 1.5$ and an *s*-tagger output score larger than 0.55. The error bars represent the statistical uncertainty and the green band represents both the statistical and systematic uncertainties.

7.5.1. Correlations Between Scale Factors

The normalization factors determined for each flavor in the profile-likelihood fits are correlated. One type of correlations are the correlations of the normalization factors of each flavor between tagged and untagged jets. This correlation is given due to an overall correction of the normalization that applies to the jets of one flavor independent of their *s*-tag, hence the normalization of tagged and untagged jets of one flavor is correlated. These correlations are already taken into account and propagated to the uncertainties shown in Figures 7.20-7.23. The other type of correlations are the correlations of the normalization factors between jets of different flavors. As an estimate for the total number of jets of all flavors can be fitted from the number of jets in data, an up variation of one flavor contribution (e.g. *s*-jets) implies a down variation of sum of jets that have a different flavor. One can understand this correlation as a lack of knowledge on the real flavor composition that remains after the profile-likelihood fit.

Figure 7.24a shows an example of the correlations between the normalization factors for jets with $25\,\text{GeV} < p_\text{T} < 50\,\text{GeV}$ and $|\eta| \leq 1.5$ as estimated by the profile-likelihood fit including all uncertainties. All of the normalization factors between tagged and untagged jets of the same flavor are positively correlated as expected, the largest correlation being 31% for the correlation between the normalization factors of charm jets and 15% for the correlation of

(a) Full profile-likelihood fit including all systematic uncertainties.

(b) Profile-likelihood fit including only statistical uncertainties using the nominal samples with Pythia as shower generator.

(c) Profile-likelihood fit including only statistical uncertainties using the shower variation sample with Herwig as shower generator.
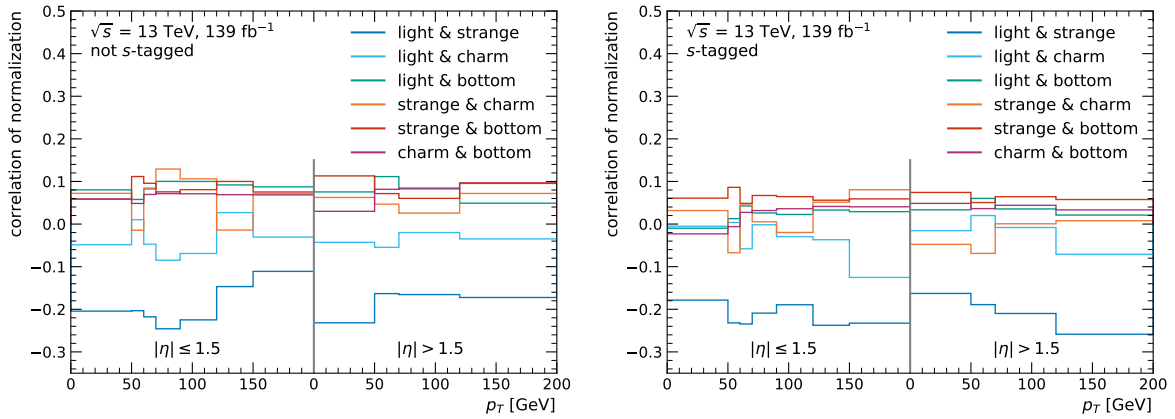
Fig. 7.24: Correlations [%] between the normalization factors derived for jets with $25\,\text{GeV} < p_\text{T} < 50\,\text{GeV}$ and $|\eta| \leq 1.5$ for different configurations used in the profile-likelihood fits.

bottom jets. Also, as suggested previously, the normalization factors for light and strange jets are significantly anti-correlated (-18% for tagged and -20% for untagged jets), as it is not possible to fully determine the flavor composition in the light-, strange-, and gluon-enriched flavor regions.

To verify that the correlations determined during the profile-likelihood fit are stable for

different flavor compositions, the fit is repeated including only statistical uncertainties in order to compare the impact of the use of two different shower generators, Pythia (used in nominal sample) and Herwig (used in the shower variation sample). The resulting correlation matrices are shown in Figures 7.24b and 7.24b. While the derived correlations are different w.r.t. the fit including the full set of systematic uncertainties, it holds that normalization factors of the same flavor are correlated for tagged and untagged jets. As the correlations are very similar for both shower generators used, it can be assumed that the flavor composition of the samples has no significant impact on the correlations.

Figure 7.25 shows the correlations between the normalization factors of different flavors for tagged and untagged jets. Their distribution is relatively flat across different p_T and η bins that are fitted independently, which suggests that the correlations derived in the profile-likelihood fit are trustworthy.

(a) Jets with an *s*-tagger output score ≤ 0.55, i.e. without an *s*-tag.

(b) Jets with an *s*-tagger output score > 0.55, i.e. with an *s*-tag.

Fig. 7.25: Pair correlations between the normalizations of jet flavors derived in the profile-likelihood fit. The plots show the relative stability of the correlations for each η and p_T bin and for all combinations of hadron flavors.

In order to take all correlations properly into account, eight eigenvariations of the normalization factors are determined by diagonalizing the covariance matrix for each p_T and η bin. Using these eigenvariations, the uncertainties are then propagated to the SFs and can be applied in any analysis using the *s*-tagger.

CHAPTER 8

Studies On Constraining $|V_{td}|$ and $|V_{ts}|$

The calibrated s-tagger is applied in a study on constraining the t-quark-related CKM matrix elements $|V_{td}|$ and $|V_{ts}|$. The relative probabilities of the decays $tt(WbWb)$, $tt(WsWb)$, and $tt(WdWb)$ are proportional to $|V_{tb}|^2$, $|V_{ts}|^2$, and $|V_{td}|^2$, respectively. It is therefore possible to simultaneously constrain each of these CKM matrix elements in a profile-likelihood fit. Because the contribution of $tt(WsWb)$ and $tt(WdWb)$ decays is expected to be small, it is reasonable to additionally consider a procedure to determine exclusion limits on the branching fractions $\mathcal{BR}(t \to W^+s)$ and $\mathcal{BR}(t \to W^+d)$ from which limits on $|V_{ts}|$ and $|V_{td}|$ can be deduced.

In the first part of this chapter, a mathematical prescription used to derive exclusion limits is introduced. Section 8.2 illustrates the data selections used to distinguish between the different $t\bar{t}$ decays. Section 8.3 discusses the composition of the background from $tt(WbWb)$ decays as well as a means to control this composition. In Section 8.4, the systematic uncertainties and the modeling of the simulated event prediction are discussed. In Section 8.5, the sensitivity to the relative contributions of $t \to W^+s$ and $t \to W^+d$ decays is discussed and in Section 8.6, the results of the measurement are presented.

8.1. Hypothesis Testing

One mathematical formalism used to derive limits on $|V_{ts}|$ and $|V_{td}|$ is called hypothesis testing. In this thesis, the frequentist CL_s method introduced in Reference [171] is used.

In general, the compatibility of a hypothesis with an observation can be quantified by a p-value, which is the cumulative probability of equal or less compatibility between the observation and the hypothesis. Assuming that the hypothesis follows a Gaussian probability distribution, one can convert the p-value into a significance

$$Z = \Phi^{-1}(1 - p), \tag{40}$$

where $\Phi^{-1}(x)$ is the inverse of the probability density function of the Gaussian distribution.

In particle physics, hypotheses are most often either *background-only* (B-only) hypotheses, i.e. the prediction of background is hypothesized to describe the observation, or *signal-plus-background* (S+B) hypothesis, i.e. a combination of the predicted background and a predicted signal is hypothesized to describe the observation.

It is important to note that a hypothesis cannot be confirmed because it is always possible that an observation could also be described by another hypothesis. However, a hypothesis can be falsified if it is rejected by an observation.

In particle physics, an observation is required to be incompatible with a hypothesis (e.g. a B-only hypothesis) with a significance of $Z = 3\sigma$ in order to claim evidence, where σ is the standard deviation of the Gaussian distribution function, while a significance of $Z = 5\sigma$ is required to claim a discovery.

The rejection of an S+B hypothesis can be used to exclude a signal prediction. If an observation has a p-value smaller 0.05 w.r.t. a S+B hypothesis, i.e. it is incompatible with

the hypothesis at a confidence level (CL) of at least 95%, this S+B hypothesis is said to be excluded by the observation at 95% CL.

If a predicted signal is variable, e.g. if its characteristics depend on one or more parameters, the rejection of S+B hypotheses can be used to exclude ranges of these parameters. Most often, the signal strength μ is such a parameter. It quantifies the overall magnitude of the signal contribution w.r.t. a nominal value.

To test hypotheses, so-called *test statistics* are used, which condense the distributions characterizing the hypotheses into single values. In this thesis, the test statistic suggested by Pearson and Neyman [172] is used because it has the largest possible statistical power. It is based on the ratios of likelihoods

$$\lambda(\mu) = \frac{\mathcal{L}\left(\mu, \vec{\hat{\theta}}(\mu)\right)}{\mathcal{L}\left(\hat{\mu}, \vec{\hat{\theta}}(\hat{\mu})\right)} \, , \tag{41}$$

where μ is the parameter of interest and $\vec{\theta}$ are all other parameters in the likelihood. $\mathcal{L}(\mu, \vec{\hat{\theta}}(\mu))$ is the conditional maximum-likelihood, i.e. $\vec{\hat{\theta}}(\mu)$ are the values of $\vec{\theta}$ that maximize $\mathcal{L}$ for a fixed value μ. $\mathcal{L}(\hat{\mu}, \vec{\hat{\theta}}(\hat{\mu}))$ is the unconditional likelihood, were both $\hat{\mu}$ and $\vec{\hat{\theta}}(\hat{\mu})$ take on those values maximizing $\mathcal{L}$. For the likelihoods, the definition in Section 7.4.1 is used. The test statistic is defined as [171]

$$q_0 = \begin{cases} -2\ln\lambda(0) & \hat{\mu} \geq 0, \\ 0 & \hat{\mu} < 0, \end{cases} = \begin{cases} -2\ln\frac{\mathcal{L}\left(0, \vec{\hat{\theta}}(0)\right)}{\mathcal{L}\left(\hat{\mu}, \vec{\hat{\theta}}(\hat{\mu})\right)} & \hat{\mu} \geq 0, \\ 0 & \hat{\mu} < 0, \end{cases} \tag{42}$$

for the B-only hypothesis and

$$\tilde{q}_\mu = \begin{cases} -2\ln\frac{\mathcal{L}\left(\mu, \vec{\hat{\theta}}(\mu)\right)}{\mathcal{L}\left(0, \vec{\hat{\theta}}(0)\right)} & \hat{\mu} < 0 \\ -2\ln\frac{\mathcal{L}\left(\mu, \vec{\hat{\theta}}(\mu)\right)}{\mathcal{L}\left(\hat{\mu}, \vec{\hat{\theta}}(\hat{\mu})\right)} & 0 \leq \hat{\mu} \leq \mu \\ 0 & \mu \leq \hat{\mu} \end{cases} \, , \tag{43}$$

for the S+B hypotheses. Larger values of q_0 or $\tilde{q}_\mu$ correspond to a greater incompatibility between data and the hypothesis. The cases for different $\hat{\mu}$ are used to handle data with a value of $\hat{\mu} < 0$, even though $\mu \geq 0$ is required by the physics model. The p-value for a given μ is defined as

$$p_\mu = \int_{q_{\text{observed}}}^{\infty} f(q_\mu | \mu) dq_\mu \, , \tag{44}$$

where $f(q_\mu | \mu)$ is the probability density function of the test statistic, μ is the parameter of interest of the hypothesis, q_μ are the possible values the data can express for the test statistic, and q_{observed} are the observed values. Figure 8.1a illustrates the distribution of two test statistics (B-only and S+B hypothesis) and an observed value that is in good agreement with the B-only hypothesis and in less agreement with the S+B hypothesis.
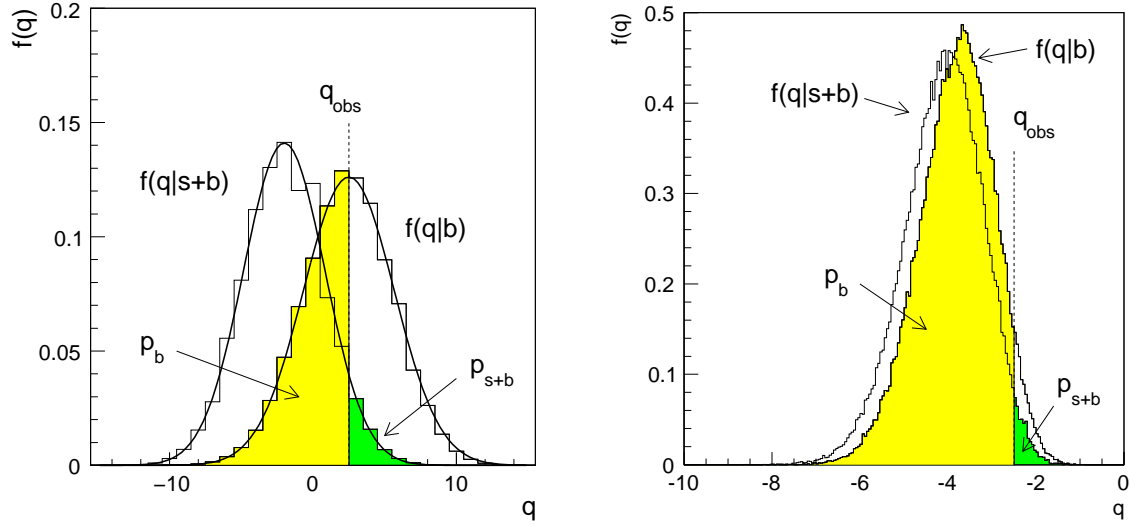
(a) Distributions of two distinctly different test statistics for a background-only and a signal-plus-background hypotheses [171].

(b) Distributions of two very similar test statistics for a background-only and a signal-plus-background hypotheses [173].

Fig. 8.1: Distributions of the test statistics for background-only hypotheses (right Gaussians) and signal-plus-background hypotheses (left Gaussians). Given the observed value q_{obs}, the *p*-value for the background-only hypothesis corresponds to the yellow area, while the *p*-value of the signal-plus-background hypothesis corresponds to the green area.

As Reference [173] illustrates, if the signal contribution is small compared to the background contribution, the B-only hypothesis and the S+B hypothesis can be too similar, resulting in the exclusion of values of μ for which the analysis actually has no sensitivity. This is illustrated in Figure 8.1b, where there is large overlap between the distributions of the test statistics of the B-only hypothesis and a S+B hypothesis. To mitigate this effect, the CL_s method [174] is used for setting exclusion limits. It combines the *p*-value of both B-only (p_b) and S+B (p_{s+b}) hypothesis to

$$\text{CL}_s := \frac{p_{s+b}}{1 - p_b} \, . \tag{45}$$

With this definition, signal contributions can be excluded at 95% CL if $\text{CL}_s < 0.05$.

For a large number of events N [175],

$$-2 \ln \lambda(\mu) = \frac{(\mu - \hat{\mu})^2}{\sigma^2} + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right) \, , \tag{46}$$

the test statistic can be approximated as [171]

$$q_0 \approx \begin{cases} \hat{\mu}^2/\sigma^2 & \hat{\mu} \geq 0, \\ 0 & \hat{\mu} < 0 \end{cases} \tag{47}$$

and

$$
\tilde{q}_\mu \approx \begin{cases} \frac{\mu^2}{\sigma^2} - \frac{2\mu\hat{\mu}}{\sigma^2} & \hat{\mu} < 0, \\ \frac{(\mu-\hat{\mu})^2}{\sigma^2} & 0 \le \hat{\mu} \le \mu, \\ 0 & \hat{\mu} > \mu, \end{cases} \tag{48}
$$

respectively. Assuming $\mu = 0$ realized in data, this *asymptotic approximation* is used to estimate the expected exclusion limits[1].

Whenever the performance and stability of any statistical methods are tested, an *Asimov dataset* is created. It is a stand-in for real data and is built from the available predictions such that the estimators for all parameters yield the true parameter as given by the prediction [171].

All statistical methods described above are used as implemented in the *TRExFitter* software package version 4.12 [158], based on the *RooFit* package [176].

8.2. Signal Region Selection

In order to constrain the t-quark-related CKM matrix elements, a dileptonic event selection is chosen, which requires an electron and a muon of opposite-sign electric charge in the final state. Figure 8.2 shows an exemplary Feynman diagram of $t\bar{t}$ production and a subsequent dileptonic decay, also depicting all flavor combinations for the quarks in the final state as allowed on tree-level by the SM. The $e\mu$ selection yields very few background events from processes other than $t\bar{t}$ since the requirement of two leptons reduces the background contribution of W boson production in association with jets and the requirement of two leptons of different flavors reduces the background contribution of Z boson production in association with jets. To reduce the contribution of background from hadronic resonances decaying to leptons, additionally, an invariant mass $m(e\mu) > 15\,\text{GeV}$ is required.
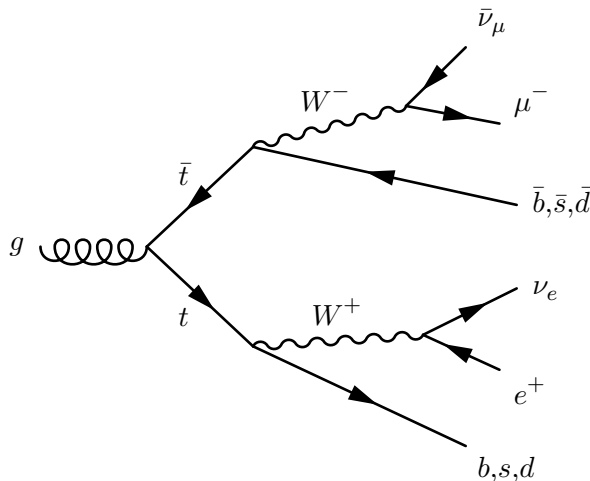

Fig. 8.2: Exemplary Feynman diagram of leading-order $t\bar{t}$ production and decays in the dileptonic channel. All possible flavor combinations of the quarks in the final state are shown.

---

[1]It can also be used to derive the expected significance by assuming $\mu = \mu'$, where μ' is the expected signal strength.

Moreover, the selection requires at least two jets with $25\,\text{GeV} < p_\text{T} < 140\,\text{GeV}$; the lower bound on the p_T is a standardized cut used in analyses targeting processes involving t quarks at the ATLAS Experiment, while the upper bound is placed because of the use of the b-tagging SFs derived by the p_T^rel calibration, which is only valid up to jet $p_\text{T} < 140\,\text{GeV}$.

To constrain $|V_{ts}|$ and $|V_{td}|$, the unitarity of the CKM matrix, i.e. $|V_{tb}|^2 + |V_{ts}|^2 + |V_{td}|^2 = 1$ is assumed. The input distributions considered in the analysis should provide at least three DOF, one for the overall normalization of $t\bar{t}$ decays, and two for the branching fractions $\mathcal{BR}(t \to W^+ s)$ and $\mathcal{BR}(t \to W^+ d)$.
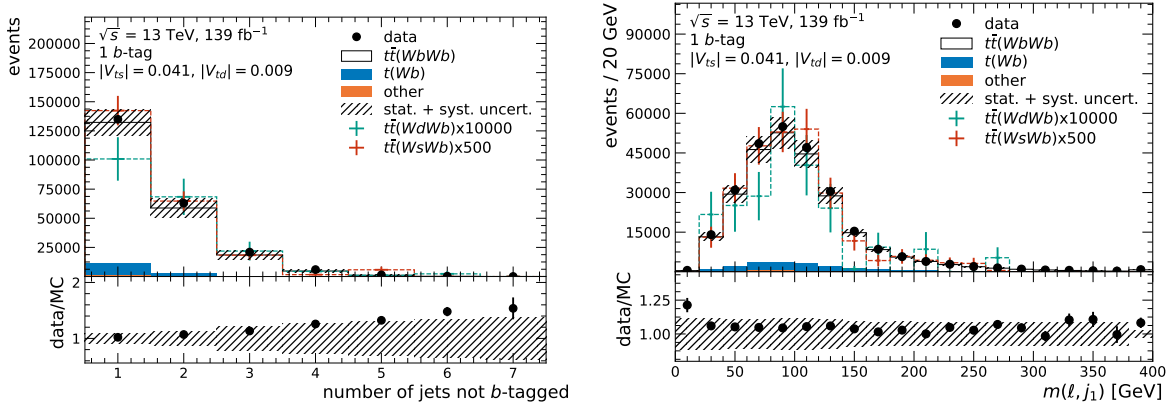
(a) The number of jets without a b-tag with $25\,\text{GeV} < p_\text{T} < 140\,\text{GeV}$.

(b) Invariant mass of the leading jet without a b-tag and the lepton that minimizes $m(\ell_1, b)^2 + m(\ell_2, j_1)^2$.

Fig. 8.3: Distributions illustrating the event yields of background processes and the agreement between prediction and data in the signal region selection requiring one b-tagged jet. The hashed bands show the combined statistical and systematic uncertainties. The signal processes $tt(WsWb)$ and $tt(WdWb)$ are overlaid, assuming $|V_{ts}| = 0.041$ and $|V_{td}| = 0.009$ and scaling them by a factor 500 and 10000, respectively. The lower panel shows the ratio between data and prediction.

As the CKM matrix element $|V_{tb}|$ is known to be close to unity, only $t\bar{t}$ decays with at least one b quark in the final state are considered; decays such as $tt(WsWs)$ and $tt(WdWd)$ and decays involving flavor-changing neutral currents that occur on loop level and are suppressed are neglected.

Two signal region (SR) selections are defined:

In the first SR selection, in order to select $t\bar{t}$ decays with one b quark and one s or d quark in the final state ($tt(WsWb)$ and $tt(WdWb)$), exactly one b-jet is required. Figure 8.3a shows the resulting number of jets without a b-tag. Because this distribution of the jet multiplicity is similar for both the signal $t\bar{t}$ decays, $tt(WsWb)$ and $tt(WdWb)$, and the background $tt(WbWb)$ decay, no further requirements are imposed on the number of jets without a b-tag. Approximately half of all events contain only one jet without a b-tag; however, for the other half of events, it is necessary to find the jet without a b-tag that has the highest probability to originate from a t decay as the s-tagger classification of this jet will yield the main discrimination between $tt(WsWb)$ and $tt(WdWb)$ events in this analysis. A trend in the ratio between data and prediction in Figure 8.3a indicates that the modeling of these jets is not ideal and that the number of jets not b-tagged is underestimated for large jet multiplicities.

The modeling will be further discussed in Section 8.4.

To match one of the jets without a b-tag to a t quark, three approaches are considered. In all of these scenarios, the jet with a b-tag is assumed to be a decay product of one of the t quarks.

- The leading jet (i.e. the jet with the largest p_T) without a b-tag, j_1, is assumed to be the jet from the second t decay as jets from $t\bar{t}$ decays tend to have larger momentum than jets from additional radiation or pile-up. To pair the jets with and without a b-tag each to one lepton in the final state, i.e. to find the pairs of jets and leptons that have their origin in the same t quark, the permutation minimizing $m(\ell_1, b)^2 + m(\ell_2, j_1)^2$ is chosen.

- Alternatively, $m(\ell_1, b)^2 + m(\ell_2, j_i)^2$ is evaluated for each jet j_i without a b-tag and all possible pairings of leptons and jets. The jet j_i that provides the smallest $m(\ell_1, b)^2 + m(\ell_2, j_i)^2$ and the corresponding lepton overall are assigned to the t quark decay without a b quark. This association tends to pair leptons and jets that are close in proximity, as leptons and jets originating from t quark decays tend to be closer together when the t quark carries large momentum.

- In a third considered approach, the jet without a b-tag that minimizes $\Delta R(\ell_1, b) + \Delta R(\ell_2, j_i)$, where the j_i are all jets without a b-jet and all permutations of leptons and jets are allowed, is chosen. Similarly to the previous type of association, this one selects lepton and jets pairs that are close in proximity, while, however, neglecting the momentum of the jets and the leptons in the consideration.

Assuming that the t quark has no intrinsic width and neglecting the masses of both the lepton and the quark from its decay, one finds that the kinematic limit of the invariant mass of the jet and lepton from a t decay, $m(\ell, j)$, is

$$\max(m(\ell, j)) = \sqrt{m_t^2 - m_W^2} \approx 153\,\mathrm{GeV}\,, \tag{49}$$

where m_t is the mass of the t quark and m_W is the mass of the W boson. However, as the t quark has an intrinsic width and mainly because the detector resolution needs to be considered as well, a requirement of $m(\ell, j) < 175\,\mathrm{GeV}$ is imposed on the invariant masses of both lepton and jet pairs. This requirement has been used in other analyses involving dileptonic $t\bar{t}$ decays, cf. e.g. the previously mentioned calibration of the b-tagging algorithm [63]. Figure 8.3b shows the distribution of the invariant mass of the leading jet without a b-tag and the lepton assigned to it, $m(\ell, j_1)$. The distributions are relatively similar for all $t\bar{t}$ decays independent of the flavor of the quarks, which means that $m(l, j_1)$ cannot be used to distinguish between $tt(WbWb)$ and $tt(WsWb)/tt(WdWb)$ decays. Within its statistical and systematic uncertainty, the simulated prediction for $m(l, j_1)$ agrees well with the observed distribution.

As only jets with $p_\mathrm{T} < 140\,\mathrm{GeV}$ can be tested for their b-tag due to the limitations of the $p_\mathrm{T}^\mathrm{rel}$ calibration, b-tagged jets with $p_\mathrm{T} > 140\,\mathrm{GeV}$ could slip through the selection. Therefore, it is additionally tested if the SR selection is improved when all events containing jets with $p_\mathrm{T} > 140\,\mathrm{GeV}$ are rejected.

In the second SR selection, to select events enriched with $tt(WbWb)$ decays, the selection for $tt(WsWb)$ and $tt(WdWb)$ decays is adapted accordingly. First of all, a minimum of two jets, exactly two of them being b-tagged, are required. The leptons and b-tagged jets are paired by

s-Tagger	Jet Classified Using the *s*-Tagger		
Classification Threshold	**Leading**	**Minimizing $\sum m(\ell, j)^2$**	**Closest to Lepton**
0.50	0.00386	0.003861	0.003881
	(0.00411)	(0.004125)	(0.00414)
0.55	0.003820	0.003832	0.00385
	(0.0041074)	(0.004096)	(0.00411)
0.60	0.003829	0.003853	0.00387
	(0.0041072)	(0.004112)	(0.00408)

Tab. 8.1: Expected exclusion limits on $\mathcal{BR}(t \to W^+ s)$ calculated using the CL_s method for different choices of values for the *s*-tagger classification and different ways of associating jets to t decays. The fit assumes $|V_{td}| = 0.009$ and $|V_{ts}| = 0.0$ (no signal injection) and only statistical uncertainties are used. For the values given in parentheses, all events containing jets with $p_T > 140\,\mathrm{GeV}$ are rejected.

minimizing $m(\ell_1, b_i)^2 + m(\ell_2, b_j)^2$, $i, j = 1, 2$, and both individual values $m(\ell, b)$ are required to be smaller than $175\,\mathrm{GeV}$.

To evaluate the performance of the three approaches for the selection and association of the non-*b*-tagged jet used with the *s*-tagger, a fit with a simple input and including only the statistical and no systematic uncertainties is considered. It uses three bins: One contains the event yield of the SR selection requiring two *b*-tagged jets and two contain the event yields of the SR selection requiring one *b*-tagged jet depending on whether the jet without the *b*-tag that was assigned to a t decay has an *s*-tag or not. This way, each bin is enriched in $tt(WbWb)$, $tt(WdWb)$, and $tt(WsWb)$ decays, respectively, which makes it a simplified version of the fit used to constrain the CKM matrix elements $|V_{td}|$ and $|V_{ts}|$ later on.

For this fit, no *s*-tagging SFs are applied, yet, because it is additionally used to choose a WP for the *s*-tagging calibration[2]. The fit does not consider data but estimates the performance based on the Asimov data. Using the CL_s method, it derives an expected upper limit at 95% CL on $\mathcal{BR}(t \to W^+ s)$, assuming $|V_{ts}| = 0.0$, $|V_{td}| = 0.009$, and $|V_{tb}| \approx 1.0$. The resulting limits are shown in Table 8.1. The best expected limit is found if the leading jet without a *b*-tag is selected as input to the *s*-tagger and if 0.55 is used as classification threshold of the *s*-tagger. Table 8.1 also shows the resulting limits on $\mathcal{BR}(t \to W^+ s)$ if all events containing jets with $p_T > 140\,\mathrm{GeV}$ are rejected. As these limits are larger for all considered *s*-tagger WPs and jet assignments, events containing jets with $p_T > 140\,\mathrm{GeV}$ will be allowed to pass the selection, but these jets will not be evaluated with the *b*-tagger.

The *s*-tagger output distribution of the leading jet without a *b*-tag is shown in Figure 8.4a. It shows that the discrimination between all types of $t\bar{t}$ decays is small. It additionally becomes clear that the number of available $tt(WsWb)$ and $tt(WdWb)$ events is not ideal as the discrimination between the two types of signal is not statistically significant and that for a proper measurement, an extension of the signal sample is necessary.

To increase the stability of the fits used in the statistical analysis and to enhance the discrimination power, the *s*-tagger output distribution of the subleading jet without a *b*-tag is included in the analysis as well. This is useful as it is possible for jets from t decays to carry less momentum than jets from additional radiation or pile-up. The distribution of the *s*-tagger

---

[2]This step in the optimization was done prior to the *s*-tagging calibration, hence this optimization was kept deliberately simple.
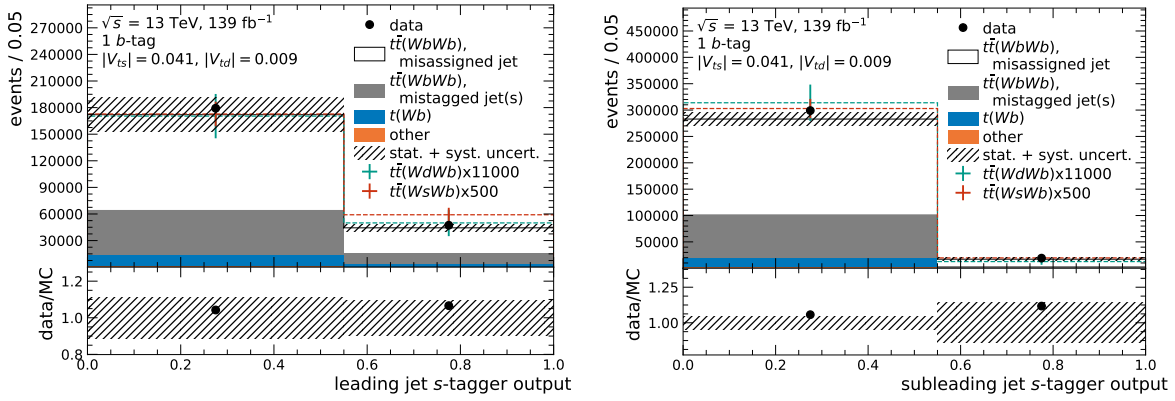
(a) *s*-Tagger output of the leading jet without a *b*-tag.

(b) *s*-Tagger output of the subleading jet without a *b*-tag.

Fig. 8.4: *s*-Tagger output of the two leading jets without a *b*-tag in the signal region requiring one *b*-tagged jet. The background of $tt(WbWb)$ decays if divided into subcategories according to whether the jet without a *b*-tag assigned to a *t* decay was mistagged as non-*b*-jet or misassigned to a *t* decay. The hashed bands show the combined statistical and systematic uncertainties. The signal processes $tt(WsWb)$ and $tt(WdWb)$ are overlaid, assuming $|V_{ts}| = 0.041$ and $|V_{td}| = 0.009$ and scaling them by a factor 500 and 11000, respectively. The lower panel shows the ratio between data and prediction.
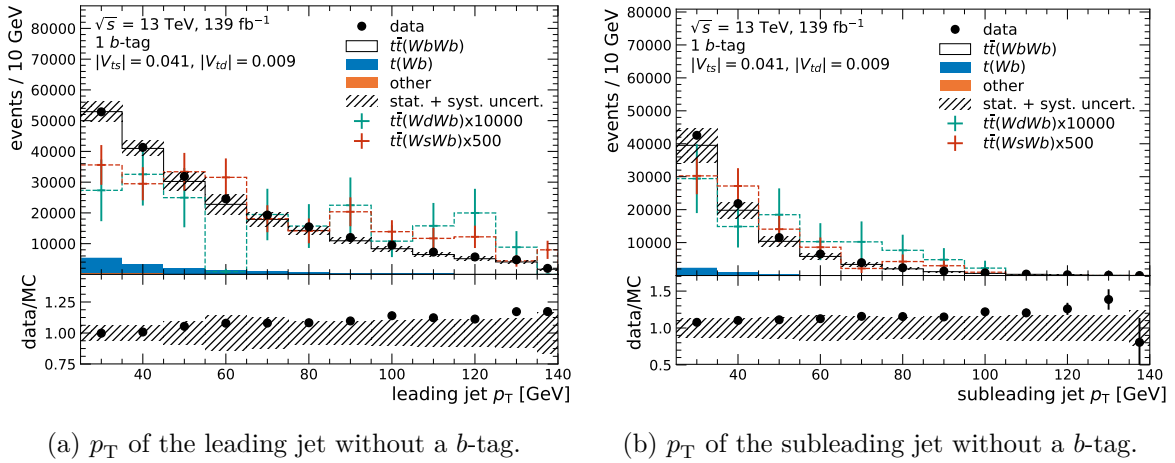
(a) p_T of the leading jet without a *b*-tag.

(b) p_T of the subleading jet without a *b*-tag.

Fig. 8.5: p_T distribution of the two leading jets without a *b*-tag in the signal region requiring one *b*-tagged jet. The hashed bands show the combined statistical and systematic uncertainties. The signal processes $tt(WsWb)$ and $tt(WdWb)$ are overlaid, assuming $|V_{ts}| = 0.041$ and $|V_{td}| = 0.009$ and scaling them by a factor 500 and 10000, respectively. The lower panel shows the ratio between data and prediction.

output of these subleading jets is shown in Figure 8.4b. Similar to the distribution for leading jets without a *b*-tag, the discrimination between all $t\bar{t}$ decays is small and lacks statistical significance.

Figure 8.5 shows the p_T distribution of both the leading and the subleading jet without a

b-tag. For these distributions, the ratio between data and prediction shows a significant trend, which will be discussed in Section 8.4. For the leading jet, there is a significant difference in the p_T distribution between $tt(WbWb)$ decays and $tt(WsWb)$ and $tt(WdWb)$ decays, the latter two having a larger average p_T. For the subleading jet, no significant differences are observed. Considering these distributions, six bins are defined based on the SR selection requiring one *b*-tagged jet:

- Two bins of the *s*-tagger output distribution for leading jets without a *b*-tag and $p_\mathrm{T} \leq 50\,\mathrm{GeV}$

- Two bins of the *s*-tagger output distribution for leading jets without a *b*-tag and $p_\mathrm{T} > 50\,\mathrm{GeV}$

- Two bins of the *s*-tagger output distribution for the subleading jet without a *b*-tag

These six bins are complemented by two additional bins. The seventh bin contains the event yield of the selection requiring two *b*-tagged jets targeting $tt(WbWb)$ decays as it was previously used for the optimization of the SR selection. Figure 8.6a shows the p_T distribution of the leading *b*-jet that passes this selection. It shows overall excellent agreement between data and prediction. The eighth bin is a control region (CR) bin used to analyze and constrain effects that allow for background events containing $tt(WbWb)$ decays to pass the SR selection requiring only one *b*-tagged jet and is discussed in detail in the following section.
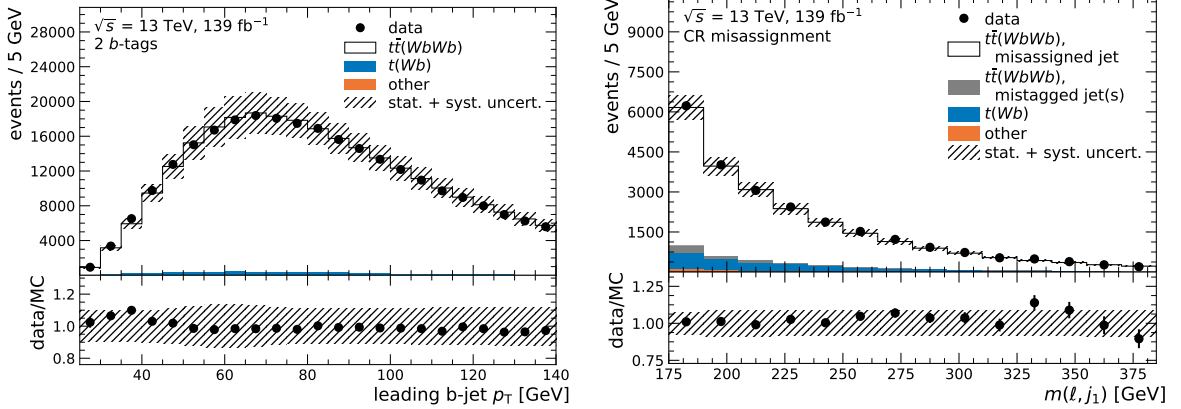
(a) p_T distribution of the leading *b*-tagged jet in the signal region requiring two *b*-jets.

(b) Distribution of the invariant mass of the leading jet without a *b*-tag and the lepton paired with it in the control region dedicated to constraining the contribution of $tt(WbWb)$ background induced by the misassignment of a jet to a *t* quark decay.

Fig. 8.6: Distributions illustrating the agreement between data and prediction used in two of the eight bins of the analysis. The hashed bands show the combined statistical and systematic uncertainties and the lower panel shows the ratio between data and prediction.

8.3. Background Composition in the Signal Region

The main background to the $tt(WsWb)$ and $tt(WdWb)$ decays selected in the SR are events containing $tt(WbWb)$ decays. By including a region requiring two b-tagged jets in the analysis, the overall fraction of $t \to bW^+$ decays can be constrained. The background from $tt(WbWb)$ decays can be further divided into classes depending on whether a jet is mistagged by the b-tagger or whether an additional light jet is misassigned to a t quark decay. Four different combinations of misassignment and mistagging are possible; they are listed – together with their absolute event yields – in Table 8.2. In the SR requiring one b-tag, 31% of the selected $tt(WbWb)$ decays pass the selection because one of the two b-jet is not b-tagged. 67% of the selected $tt(WbWb)$ decays pass the selection because only one of the two b-jets is reconstructed and/or b-tagged and an additional jet without a b-tag is misassigned to a t decay. The remaining 2% of the selected $tt(WbWb)$ events pass the selection with two mistags (a non-b-jet that is misassigned to a t quark decay is additionally wrongly b-tagged and a b-jet correctly assigned to a t decay is wrongly not b-tagged) or because two additional jets were misassigned to the two t decays and one of them is wrongly b-tagged. In the following, all events containing $tt(WbWb)$ are classed as "mistagged" if they contain at least one mistagged jet (this includes the latter three types of misselections), or else they are classed as $tt(WbWb)$ including a misassigned jet. Figures 8.4a and 8.4b illustrate the relative contributions of these two types of $tt(WbWb)$ background in the SR.

Type of $tt(WbWb)$ Background	Signal Region, 1 b-Tagged Jet	Control Region Misassignment
Light jet misassigned to top-quark decay	130000 ± 10000	19000 ± 2000
b-jet from top-quark decay without a b-tag	61000 ± 15000	340 ± 90
Light jet mistagged as b-jet and both jets misassigned to top-quark decay	2400 ± 400	340 ± 70
Both jets assigned to top quarks are mistagged	1400 ± 400	55 ± 20

Tab. 8.2: Number of events containing $tt(WbWb)$ decays passing the signal region selection requiring one b-tagged jet, divided by the reason for their selection as listed in the first column. The number of events are either the sum of the events in all bins containing events passing the signal region selection requiring one b-tagged jet (middle column) or the number of events passing the selection of a control region dedicated to the constraint of background contributions due to jet misassignment (right column). The uncertainties correspond to the statistical and systematic uncertainties and event yields are rounded to significant digits.

The uncertainty on the SR contribution of $tt(WbWb)$ decays passing the selection due to a mistagged jet should be equivalent to the b-tagging uncertainty. It is, on the other hand, not evident that the contribution of background induced by misassigning a jet to a t decay is correctly modeled in simulations. Therefore, in order to verify and correct the number of such events passing the selection, a dedicated control region is defined.

This region uses the same selection requirements as the SR selection detailed in the previous section. However, to increase the fraction of events containing misassigned jets, the requirement on the invariant mass $m(\ell, j_1)$ is inverted to $m(\ell, j_1) > 175\,\text{GeV}$, where j_1 is the leading

jet without a b-tag and ℓ is the lepton assigned to it by minimizing $m(\ell_1, j_1)^2 + m(\ell_2, b)^2$ for the two possible permutations of lepton and jet pairs. Figure 8.6b shows the resulting distribution of $m(\ell, j_1)$ with an overall good agreement between data and simulation, which means that corrections to the normalization of the background contribution due to misassignment will be small. While the overall event yield in the CR is significantly smaller than in the bins containing events passing the SR selection requiring one b-tagged jet, the fraction of events included because of a misassigned jet is significantly increased to 96%. A single bin containing all events passing the selection of this CR is included in the fit to constrain the background contribution due to jet misassignment.

In principle, the amplitude of both the production and decay of single t quarks involve CKM matrix elements. Single t quark decaying to a W boson and an s or d quark, however, do not have a significant event contribution in this analysis since at least one b-tagged jet is required in the selection. For smaller $|V_{tb}|$ and larger $|V_{ts}|$ or $|V_{td}|$, in principle, both cross sections and kinematics of singly produced t quarks would be altered. As single t quark production involving the latter two CKM matrix elements is not available in simulations at this point, these effects were neglected in this analysis, assuming a fixed value of $|V_{tb}| = 1.0$ for single t quark production.

8.4. Systematic Uncertainties

Most systematic uncertainties described in 7.4.3 that are applied in the s-tagger calibration described in Chapter 7 are included in the fit used to constrain $|V_{ts}|$ and $|V_{td}|$ as well. Those uncertainties not applied in the measurement are the uncertainties on the SFs of the c-tagger (which is not used in this measurement), the uncertainties related to the estimation of background from lepton fakes (which has a negligible impact in this measurement), and the uncertainties on the biases imposed on the output of the s-tagger due to the selection of the flavor enriched regions (because this uncertainty is specific to the calibration of the s-tagger).

In the measurement of $|V_{ts}|$ and $|V_{td}|$, the b-tagging SFs derived by the calibration using the p_T^{rel} method are applied to b-jets. These SFs are varied according to the up and down variation available and are for now assumed to not be correlated to the b-tagging SFs of the other jet flavors.

The only additional uncertainty included in the measurement is the uncertainty on the application of the s-tagger as derived in Chapter 7. The eight eigenvariations are applied to all jets evaluated with the s-tagger and simultaneously for jets of all p_T and η. The eigenvariation with the largest impact on the analysis imposes an up to 9% variation on the normalization of jets with and without an s-tag.

As previously mentioned, the number of simulated events in the samples of $tt(WsWb)$ and $tt(WdWb)$ decays is relatively small, which leads to a considerable impact of the statistical uncertainty of these simulated samples on the measurement. This statistical uncertainty is included as a separate uncertainty for both $tt(WsWb)$ and $tt(WdWb)$ decays.

Both the multiplicity (Figure 8.3a) and the p_T distributions of jets without a b-tag (Figures 8.5a and 8.5b) show significant trends that are, however, mostly covered by systematic uncertainties. Several observations indicate an issue in the b-tagging SFs for b-jets without a b-tag (i.e. mistagged jets) as derived by the p_T^{rel} calibration, which has not been fully approved by the ATLAS Collaboration yet. First of all, the disagreement between data and prediction
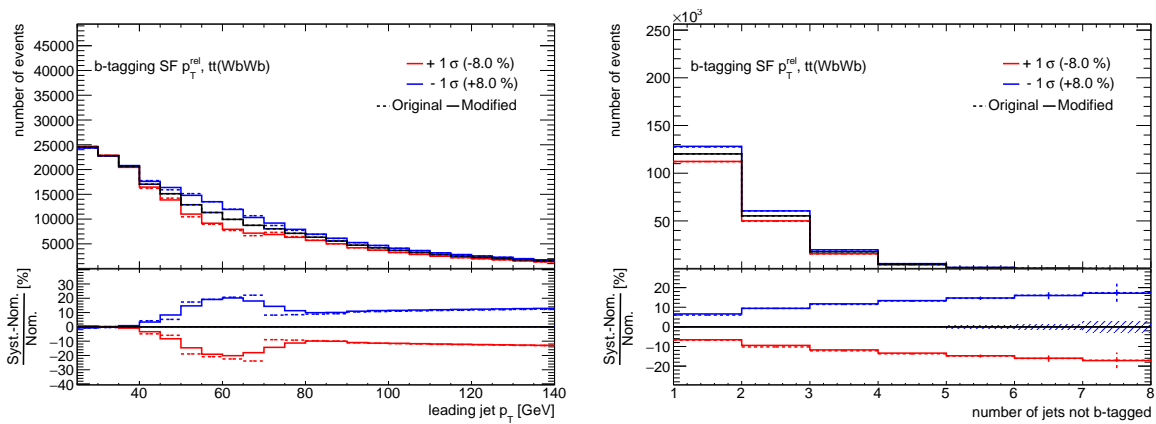
(a) Impact of the uncertainty of the *b*-tagging scale factors derived by the p_T^{rel} calibration on the p_T distribution of leading jets without a *b*-tag.

(b) Impact of the uncertainty of the *b*-tagging scale factors derived by the p_T^{rel} calibration on the number of jets without a *b*-tag.

Fig. 8.7: Absolute (top panel) and relative (bottom panel) impact of the up and down variations of the *b*-tagging scale factors derived by the p_T^{rel} calibration. The plots show the up and down variation of this uncertainty (labeled as *modified*) for $tt(WbWb)$ decays passing the signal region selection requiring one *b*-tagged jet w.r.t. the nominal prediction (labeled as *original*).

is prominent in regions that contain significant contributions of mistagged *b*-jets like the SR requiring one *b*-tagged jet, while there is almost perfect agreement in the SR requiring two *b*-tagged jets and the CR for misassignment. It is notable that the SFs of untagged jets are calculated from the SFs of tagged jets using

$$\kappa_b(\text{untagged jets}) = \frac{1 - \kappa_b(\text{tagged jets})\varepsilon_b}{1 - \varepsilon_b}, \tag{50}$$

where κ_b are the *b*-tagging SFs for tagged and untagged *b*-jets, respectively, and ε_b is the efficiency for *b*-jets. Due to this relation, small variations of the SFs of tagged jets yield large variations for the SFs of untagged jets. And lastly, it is found that the down variation of the *b*-tagging SFs is able to account for the larger part of the deviations observed in the Figures 8.3a, 8.5a, and 8.5b, as can be seen in Figure 8.7.

Figure 8.8 supports all of the above arguments. It shows the distribution of the leading jet without a *b*-tag in the SR requiring one *b*-tagged jet. However, for this figure, not the *b*-tagging SFs for *b*-jets derived by the p_T^{rel} calibration, but the standard *b*-tagging scale factors derived in $t\bar{t}$ decays are used. With these alternative SFs, the trend in the p_T distribution is significantly reduced. However, as previously mentioned, this calibration cannot be used in the measurement of $|V_{ts}|$ and $|V_{td}|$, as it makes assumptions on these CKM matrix elements.

In principle, the profile-likelihood fits used in this analysis are able to compensate for the discrepancies between data and prediction by shifting the nominal value of a systematic uncertainty according to the up or down variation of this uncertainty, which is referred to as pull. Since several systematic variations can have the same impact on the input distribution of the fit, a pull does not necessarily reflect the reason for the discrepancy between data and prediction. The up variation of the scale variation of the $tt(WbWb)$ sample e.g. shows a similar behavior as the down variation of the uncertainty on the *b*-tagging SFs, as can be seen
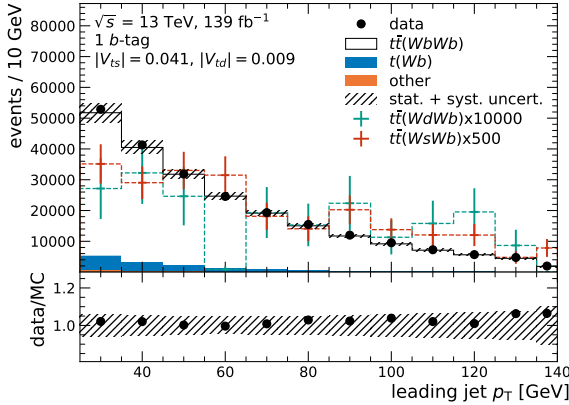
Fig. 8.8: p_T of the leading jet without a b-tag in the signal region requiring one b-tagged jet. The standard and not the p_T^{rel} b-tagging scale factors for b-jets are used. The hashed bands show the combined statistical and systematic uncertainties. The signal processes $tt(WsWb)$ and $tt(WdWb)$ are overlaid, assuming $|V_{ts}| = 0.041$ and $|V_{td}| = 0.009$. The lower panel shows the ratio between data and prediction.
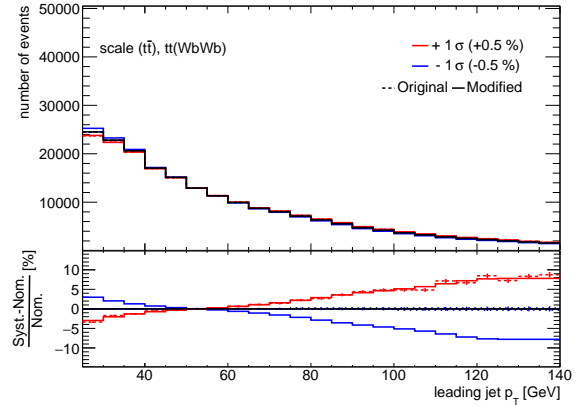
Fig. 8.9: Absolute (top panel) and relative (bottom panel) impact of the uncertainty of the scale variation on the $tt(WbWb)$ sample. The plots show the up and down variation of this uncertainty (labeled as *modified*) for events passing the signal region selection requiring one b-tagged jet w.r.t. the nominal prediction (labeled as *original*).

in Figure 8.9. Furthermore, even if a potential fault in the p_T^{rel} calibration could be compensated by pulls on the systematic uncertainty, one needs to keep in mind that the systematic uncertainties have a Gaussian constraint and therefore will most likely not be pulled enough to compensate the effect fully. Hence, the measurement of $|V_{ts}|$ and $|V_{td}|$ will likely be biased towards larger values.

8.5. Expected Sensitivity

The main objective of this thesis is the separation of $t \to W^+ s$ and $t \to W^+ d$ decays using an s-tagger. Figure 8.10 summarizes all bins used in the analysis. It illustrates that $tt(WbWb)$ decays and the decays to $tt(WsWb)$ and $tt(WdWb)$ can be separated well and that the difference between $tt(WsWb)$ and $tt(WdWb)$ decays is small. In the following, the sensitivity of the analysis setup is explored.

In addition to employing the CL_s method to derive exclusion limits on $|V_{ts}|$ and $|V_{td}|$, the branching fractions $\mathcal{BR}(t \to W^+ s)$ and $\mathcal{BR}(t \to W^+ d)$ are determined in a profile-likelihood fits as introduced in Section 7.4.1. As $\mathcal{BR}(t \to W^+ q) \propto |V_{tq}|^2 \propto tt(WqWb)$ and $|V_{tb}|^2 + |V_{ts}|^2 + |V_{td}|^2 = 1$ is assumed, the contribution of $tt(WbWb)$ decays is scaled down if the contribution of $tt(WsWb)$ or $tt(WdWb)$ decays is scaled up and vice versa. Additionally, the overall normalization of $tt(WbWb)$ background due to the misassignment of a light jet to a t decay is included as a free parameter in the fit. All other backgrounds (including the
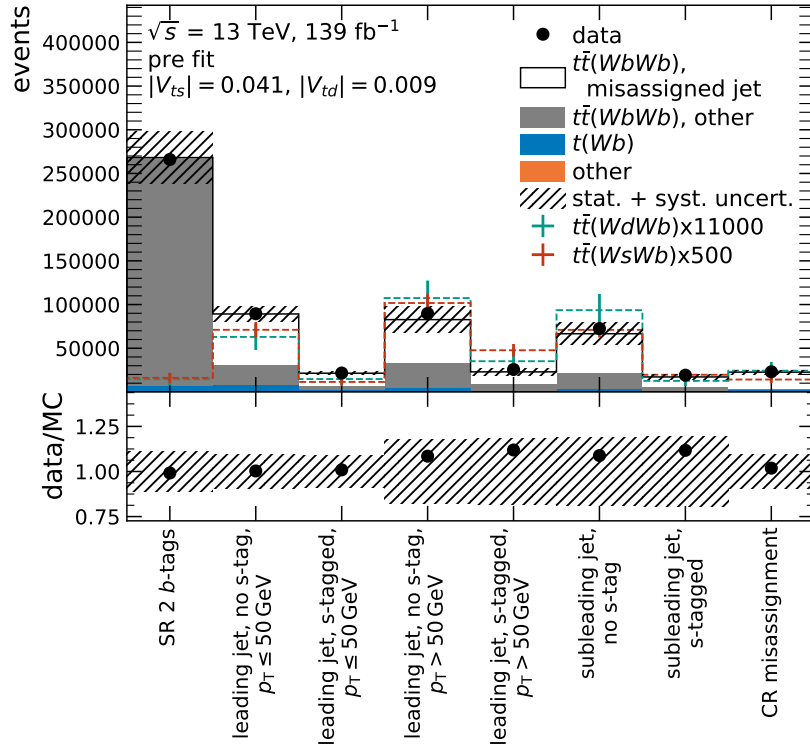
Fig. 8.10: Illustration of all bins used in the measurement of $|V_{ts}|$ and $|V_{td}|$. The $tt(WbWb)$ background induced by one light jet misassigned to t decays is shown separate from other $tt(WbWb)$ background as its contribution is constrained in the dedicated control region bin. The hashed bands show the combined statistical and systematic uncertainties and the lower panel shows the ratio between data and prediction. The signal processes $tt(WsWb)$ and $tt(WdWb)$ are overlaid, assuming $|V_{ts}| = 0.041$ and $|V_{td}| = 0.009$. The lower panel shows the ratio between data and prediction.

background contribution of single t quark decays) are only varied according to the impact of the previously described systematic uncertainties and remain constant otherwise. The fits are parameterized in terms of $\mathcal{BR}(t \rightarrow W^+s) + \mathcal{BR}(t \rightarrow W^+d)$ and $\mathcal{BR}(t \rightarrow W^+d)/(\mathcal{BR}(t \rightarrow W^+s) + \mathcal{BR}(t \rightarrow W^+d))$, both to increase the stability of the fit and to increase legibility of the results.

Figure 8.11 shows the likelihood profiles of profile-likelihood fits to Asimov data that were done to determine the sensitivity of the analysis to $tt(WsWb)$ and $tt(WdWb)$ decays. They are calculated by fixing $\mathcal{BR}(t \rightarrow W^+s) + \mathcal{BR}(t \rightarrow W^+d)$ and $\mathcal{BR}(t \rightarrow W^+d)/(\mathcal{BR}(t \rightarrow W^+s) + \mathcal{BR}(t \rightarrow W^+d))$ to each value represented in the plot and minimizing the negative logarithm of the likelihood. The smallest resulting negative logarithm of the likelihood overall is then subtracted from all resulting negative logarithm of the likelihood values; this $-\Delta \ln(\mathcal{L})$ is shown as heatmap in the figure and is used to define both one-sided and two-sided confidence intervals assuming that the likelihood follows a Gaussian distribution around the minimum.

In Figure 8.11a, the expected sensitivity of the presented measurement strategy when assuming $|V_{ts}|^2 + |V_{td}|^2 = 0.00213$ and a ratio of $9 : 1$ for $|V_{ts}|^2 : |V_{td}|^2$ is shown. Especially for large $\mathcal{BR}(t \rightarrow W^+d)/(\mathcal{BR}(t \rightarrow W^+s) + \mathcal{BR}(t \rightarrow W^+d))$, the sensitivity is limited by the
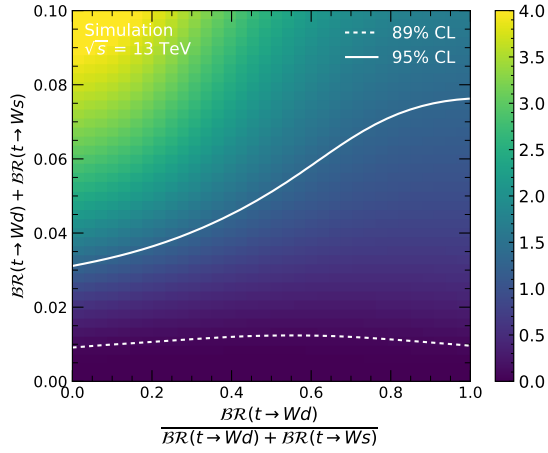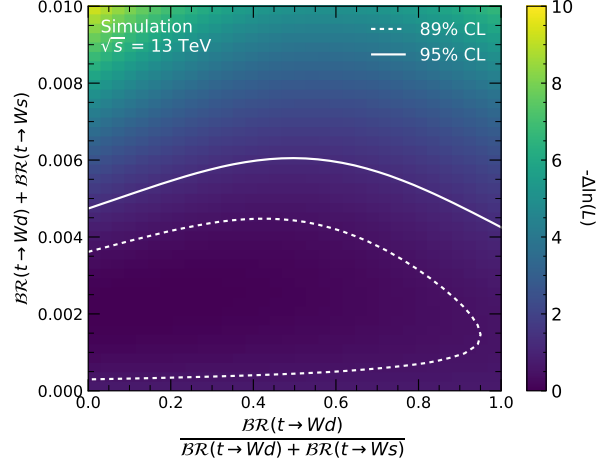
(a) Expected likelihood profile including the statistical uncertainty associated with the simulated samples of $tt(WsWb)$ and $tt(WdWb)$ decays.
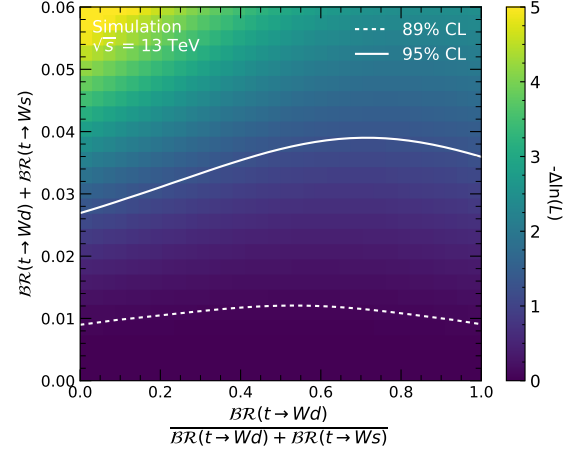
(b) Expected likelihood profile including a statistical uncertainty associated with the simulated samples of $tt(WsWb)$ and $tt(WdWb)$ decays reduced by 50%.
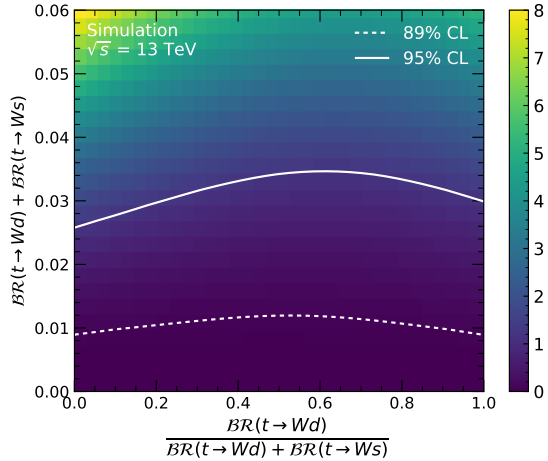
(c) Expected likelihood profile not including the statistical uncertainty associated with the simulated samples of $tt(WsWb)$ and $tt(WdWb)$ decays.

(d) Expected likelihood profile not including the statistical uncertainty associated with the simulated samples of $tt(WsWb)$ and $tt(WdWb)$ decays and scaling the expected background events to 4.5% of the nominal contribution.

Fig. 8.11: Expected two-dimensional likelihood profiles in the plane spanned by $\mathcal{BR}(t \to W^+s) + \mathcal{BR}(t \to W^+d)$ and $\mathcal{BR}(t \to W^+d)/(\mathcal{BR}(t \to W^+s) + \mathcal{BR}(t \to W^+d))$. The difference between the negative logarithm of the likelihood for each point and its global minimum is illustrated as a heatmap. The fit assumed $|V_{ts}|^2 + |V_{td}|^2 = 0.00213$ and a ratio of $9:1$ for $|V_{ts}|^2 : |V_{td}|^2$. The confidence intervals correspond to 1 and 2σ variations from the best fit value.

available number of simulated $tt(WdWb)$ decays, resulting in a large 95% confidence interval covering values of $\mathcal{BR}(t \to W^+s) + \mathcal{BR}(t \to W^+d) < 0.075$.

To determine the impact of the statistical uncertainty of the prediction of $tt(WsWb)$ and $tt(WdWb)$ decays, this uncertainty is artificially reduced by 50% for both decays (i.e. assuming the availability of signal samples that are four times larger), the resulting likelihood profile being shown in Figure 8.11b, and removed altogether, resulting in the expected likelihood profile shown in Figure 8.11c. Both of these estimates assume that the few available simulated $tt(WsWb)$ and $tt(WdWb)$ decays are a good representation of a larger number of $tt(WsWb)$ and $tt(WdWb)$ decays. An increase of 300% of the simulated $tt(WsWb)$ and $tt(WdWb)$ decays available to construct the prediction reduces the impact of the statistical uncertainty of the prediction significantly; the resulting 95% confidence interval covers values of $\mathcal{BR}(t \to W^+s) + \mathcal{BR}(t \to W^+d) < 0.036$. However, even when the statistical uncertainty of the signal samples is removed completely, it is not possible to distinguish $tt(WsWb)$ and $tt(WdWb)$ decays.

Figure 8.11d shows that the total background contribution of $tt(WbWb)$ and other processes would have to be reduced by 95.5% or more so that the 89% confidence interval would not cover the entire range of $\mathcal{BR}(t \to W^+d)/(\mathcal{BR}(t \to W^+s) + \mathcal{BR}(t \to W^+d))$ when assuming $|V_{ts}|^2 + |V_{td}|^2 = 0.00213$ and a ratio of $9 : 1$ for $|V_{ts}|^2 : |V_{td}|^2$. As the main background of $tt(WbWb)$ is induced by misassignment of an additional jet to t quark decays, the reduction of this background could be the starting point of further improvements of the analysis strategy. However, as dileptonic $t\bar{t}$ decays have two neutrinos in their final states, a kinematic reconstruction of the decay is difficult and will, most likely, not enable a 95.5% background reduction.

Based on the studies presented in Chapter 6.2, it can be assumed that the construction of an s-tagger with increased classification power at the ATLAS Experiment is not likely because the capacity for particle identification is limited. However, the calibration of the s-tagger might be improved by reconsidering or extending the calibration strategy (e.g. by introducing additional regions to it), by including an improved c-tagging calibration if it becomes available, or by including a larger set of data when it becomes available.

8.6. Constraints On $|V_{ts}|$ and $|V_{td}|$

While it is not possible to determine $|V_{ts}|^2 + |V_{td}|^2$ or the ratio of $|V_{ts}|^2$ and $|V_{td}|^2$ in this analysis, it is possible to constrain them.

In a first step, a profile-likelihood fit as described in the previous section is done using data. Figure 8.12 summarizes the constraints and pulls on all uncertainties used in this fit.

The uncertainty of the SFs of the p_T^{rel} calibration is pulled towards the down variation and the scale uncertainty of the $tt(WbWb)$ sample is pulled towards larger values. These pulls are expected based on the previous discussion on the modeling of the simulated prediction. The uncertainty on the p_T^{rel} calibration is constrained, which means that the fit is able to reduces the uncertainty of the b-tagging SFs used because the input distributions – comprised of bins containing b-jets with and without a b-tag – provide it with sensitive to these b-tagging SFs.

The largest pull and the third largest constraint is observed in the uncertainty of the cross section of $t\bar{t}$ production. It occurs because the fit is able to distinguish between $t\bar{t}$ events and background from other processes based on their different multiplicities of b-tagged jets and their relative fractions of the leading and subleading jets included in the SR. The resulting
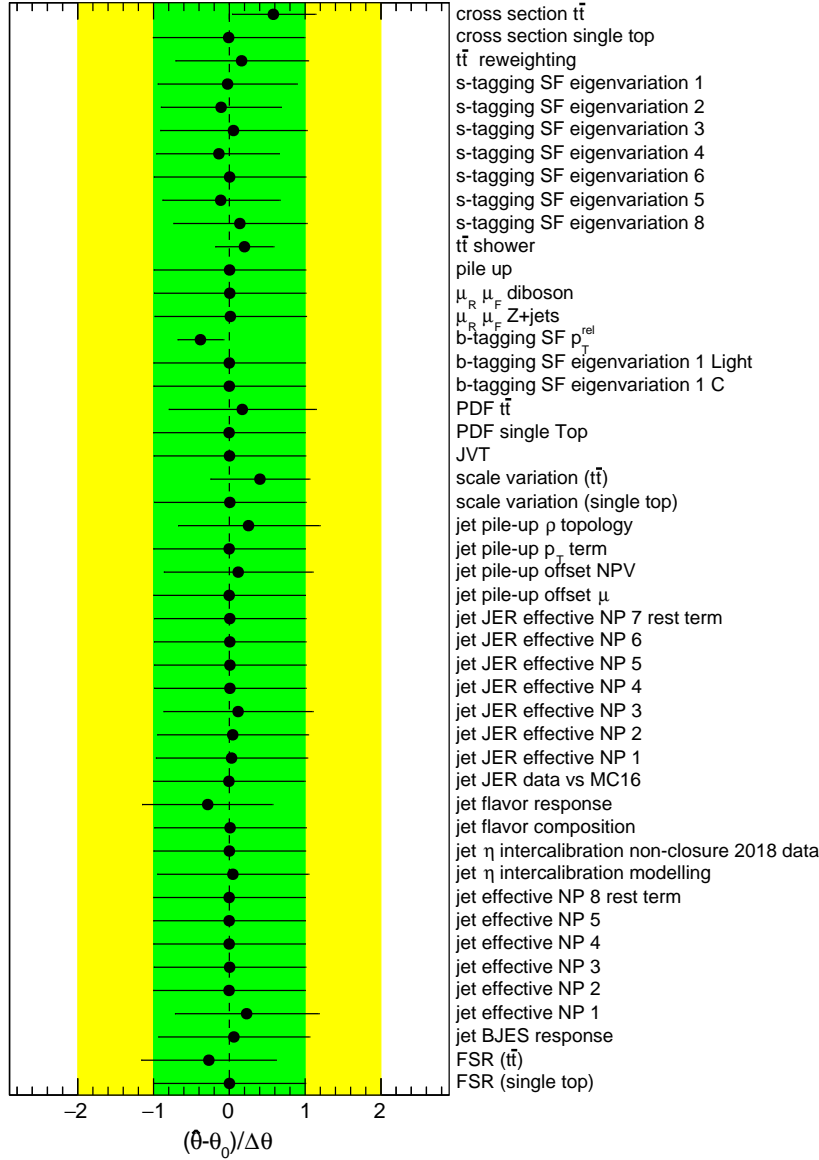
Fig. 8.12: Pull plot of all systematic uncertainties that have either a normalization or shape component with an impact larger than 1%. This plot is the result of a profile-likelihood fit to find the best fit values of $|V_{ts}|$ and $|V_{td}|$ using data.

central value for the cross section of $t\bar{t}$ production still lies within the uncertainty of the theory prediction.

The second largest constraint is observed for the uncertainty on the shower generator used for $tt(WbWb)$ samples. The shower uncertainty is constrained because the fit is sensitive to the different flavor compositions of jets that are the result of using different shower generators. However, the shower variation used as input to the fit is not founded on physical principles; it is based on the symmeterized one-sided variation given by exchanging the shower generator from the nominally used generator Pythia to Herwig. Therefore, a constrained shower uncertainty
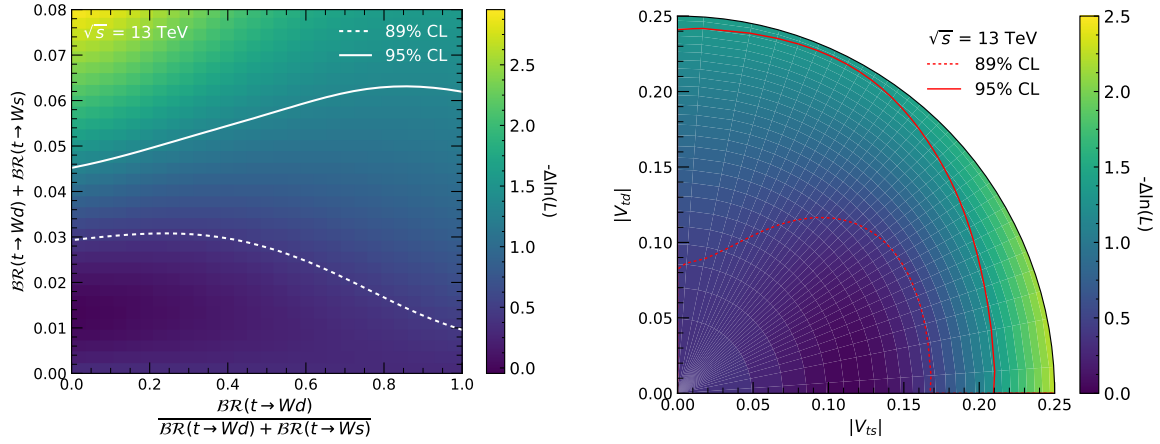
(a) Likelihood profiles in the plane spanned by $\mathcal{BR}(t \to W^+s) + \mathcal{BR}(t \to W^+d)$ and $\mathcal{BR}(t \to W^+d)/(\mathcal{BR}(t \to W^+s) + \mathcal{BR}(t \to W^+d))$.

(b) Likelihood profile in the plane spanned by the CKM matrix elements $|V_{ts}|$ and $|V_{td}|$.

Fig. 8.13: Two-dimensional likelihood profiles derived in data. The difference between the negative logarithm of the likelihood for each point and its global minimum is illustrated as a heatmap. The confidence intervals correspond to 1 and 2σ variations from the best fit value.

means that the analysis is sensitive to quantities that are potentially not well modeled, which needs further investigation in the future.

All other pulls and constraints shown in Figure 8.12 are not significant compared to the previously mentioned nuisance parameters.

When interpreting the results presented in the following, all previously described caveats of the analysis strategy have to be kept in mind.

Figure 8.13a shows the likelihood profile in the plane spanned by $\mathcal{BR}(t \to W^+s) + \mathcal{BR}(t \to W^+d)$ and $\mathcal{BR}(t \to W^+d)/(\mathcal{BR}(t \to W^+d) + \mathcal{BR}(t \to W^+s))$ derived in data. The 89% confidence interval is shifted towards larger values of $\mathcal{BR}(t \to W^+s)$ compared to the likelihood profile based on Asimov data because there is relatively more data in bins including jets with an s-tag than in those bins including jets without an s-tag. The 89% confidence interval is additionally shifted to larger values of $\mathcal{BR}(t \to W^+d) + \mathcal{BR}(t \to W^+s)$, because of the bias of the b-tagging SFs. The best-fit value for $\mathcal{BR}(t \to W^+s) + \mathcal{BR}(t \to W^+d)$ is found to be approximately 0.01. This value is roughly one order of magnitude larger compared to the value expected based on the values of $|V_{ts}|$ and $|V_{td}|$ derived in a global fit of the CKM matrix [24]. The 95% confidence interval is shifted slightly towards lower values of $\mathcal{BR}(t \to W^+d) + \mathcal{BR}(t \to W^+s)$ compared to the likelihood profile based on Asimov data because, in the fit to data, larger $\mathcal{BR}(t \to W^+s)$ are preferred. Based on this result, a constraint of $|V_{ts}|^2 + |V_{td}|^2 < 0.06$ can be given at a 95% CL.

Figure 8.13b shows the likelihood profile transformed into the plane spanned by $|V_{ts}|$ and $|V_{td}|$. While the fit is still not sensitive to the ratio of $|V_{ts}|$ and $|V_{td}|$, it shows that $|V_{ts}| < 0.21$ and $|V_{td}| < 0.24$ at 95% CL.

105

In an alternative approach, because both $tt(WsWb)$ and $tt(WdWb)$ decays are not in the reach of observation, 95% CL_s on $\mathcal{BR}(t \to W^+s) + \mathcal{BR}(t \to W^+d)$are derived. This is done for 21 equidistant values of $\mathcal{BR}(t \to W^+d)/(\mathcal{BR}(t \to W^+s) + \mathcal{BR}(t \to W^+d)) \in [0.0, 1.0]$. Similar to the setup of the profile-likelihood fits, for the 95% CL_s method, the background contribution of $tt(WbWb)$ decays is adjusted according to the different contributions of $tt(WsWb)$ and $tt(WdWb)$ included in the hypothesis test in order to preserve the unitarity of the CKM matrix. The contribution of $tt(WbWb)$ background due to jet misassignment is adjusted to the value determined in the profile-likelihood fit. All other backgrounds are only varied according to the impact of the previously described systematic uncertainties and remain constant otherwise.

The resulting limits are shown in Figure 8.14a. The main difference between considering a likelihood profile and the CL_s method to derive limits on $\mathcal{BR}(t \to W^+s) + \mathcal{BR}(t \to W^+d)$ is that the CL_s method only uses the nominal values provided by the simulated prediction to derive limits. The uncertainties, especially the large statistical uncertainties of the samples containing $tt(WsWb)$ and $tt(WdWb)$ decays, only become apparent as large uncertainty bands for the expected CL_s limit. The observed 95% CL_s limit lies within the 1σ confidence interval of the expected upper limit for almost all ratios of $\mathcal{BR}(t \to W^+s)$ and $\mathcal{BR}(t \to W^+d)$. It is slightly larger for large $\mathcal{BR}(t \to W^+s)$ but still well contained in the 2σ confidence interval. The maximum observed 95% CL_s limit excludes $\mathcal{BR}(t \to W^+s) + \mathcal{BR}(t \to W^+d) > 0.06$.
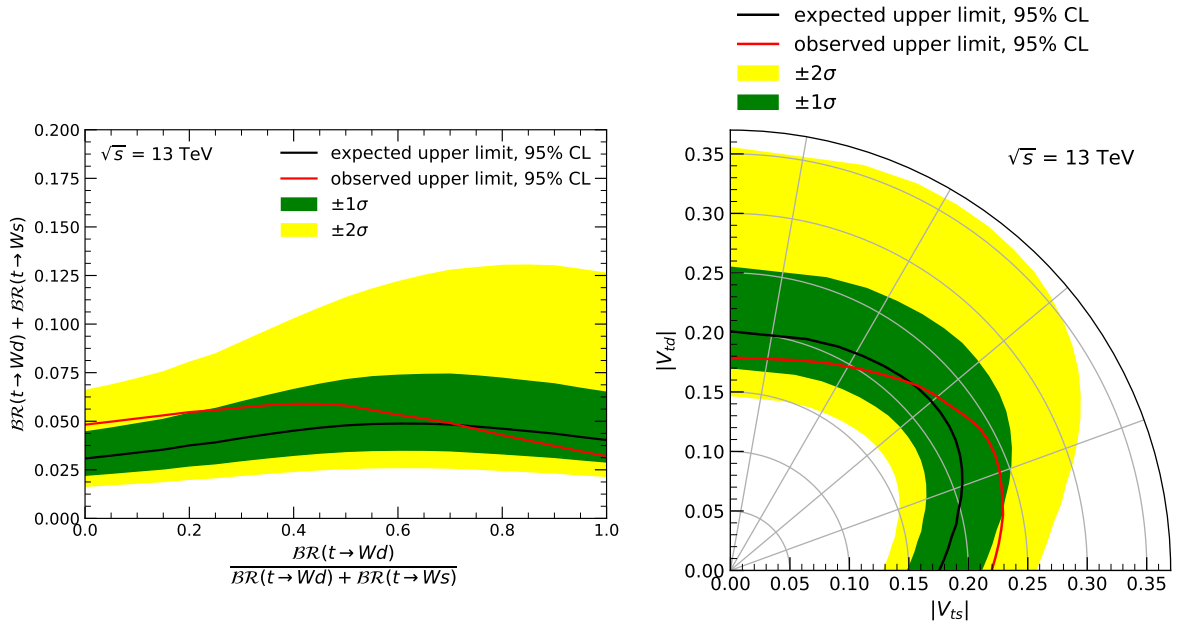
(a) 95% CL_s limits on the sum of the branching ratios $\mathcal{BR}(t \to W^+s)$ and $\mathcal{BR}(t \to W^+d)$ for different relative contributions of them.

(b) 95% CL_s limits on $|V_{ts}|$ and $|V_{td}|$.

Fig. 8.14: Expected and observed 95% CL_s limits. 1 and 2σ variations from the expected value are shown as green and yellow bands.

In Figure 8.14b, the 95% CL_s limits are transformed to be represented in the two-dimensional plane spanned by $|V_{ts}|$ and $|V_{td}|$. At 95% CL, $|V_{ts}| > 0.23$ and $|V_{td}| > 0.18$ are excluded.

The constraints on $|V_{ts}|$ and $|V_{td}|$ derived in the presented study are quantitatively slightly worse than the most stringent value obtained by an R_b measurement, which is $|V_{tb}|^2 > 0.95$ at 95% CL [6]. They are similar to the limit of $|V_{ts}|^2 + |V_{td}|^2 < 0.057$ at 95% CL derived in single t quark productions [17].

However, with a limit of $|V_{ts}| < 0.132$ at 95% CL assuming $|V_{td}| = 0.0$, Reference [28] is able to set a much stricter limit on $|V_{ts}|$ than the study presented in this thesis because its main focus is the reduction of background from $t \to W^+ b$ decays. In this thesis, on the other hand, the separation of $t \to W^+ s$ and $t \to W^+ d$ decays was the main objective.

It is also apparent that the limits on $|V_{ts}|$ and $|V_{td}|$ derived in t decays, while being independent of the use of hadronic form factors from theory calculations, cannot yet compete with the measurements of $|V_{ts}|$ and $|V_{td}|$ in B_s^0–$\overline{B_s^0}$ and B^0–$\overline{B^0}$ oscillations [24], which have a per-mill precision.

CHAPTER 9

Summary and Conclusion

In this thesis, an algorithm for the identification of jets from strange quarks – a so-called s-tagger – was developed, calibrated, and applied in a measurement constraining the CKM matrix elements $|V_{ts}|$ and $|V_{td}|$ in top-quark decays at the ATLAS experiment.

The attainable performance of an s-tagger at hadron colliders was explored in a study considering different idealized detector designs. For this, simulated particles were used as input to recurrent neural networks including Long Short-Term Memory layers to derive estimates on the achievable discrimination power between jets from strange and down quarks. This study revealed that the main difference between these jets are the types of the most energetic constituent particles. Hence, detectors that enable particle identification – such as Cherenkov detectors – allow for the best separation of jets from strange and down quarks. If particle identification is not available, the combined use of information from tracking detectors as well as electromagnetic and hadronic calorimeters is most promising to separate the two jet flavors.

The presented studies support the results of the other available publications on s-tagging and can be used as a reference for the performance of s-taggers at actual collider detectors. Should future detectors desire to the identification of jets from strange quarks, the studies can also be used in the consideration of different detector setups.

The first s-tagger at the ATLAS Experiment was built with the aim of being able to distinguish between jets from $t \to W^+ s$ and $t \to W^+ d$ decays. It uses a deep neural network to combine information from reconstructed K_S and Λ^0, from energy deposited in the calorimeters, and from reconstructed tracks. For an identification efficiency for jets from strange quarks of 40%, it provides a rejection rate of 4.0 for jets from down quarks.

The s-tagger was calibrated in dependence of the jets' transverse momentum and pseudorapidity using semileptonic $t\bar{t}$ decays including a hadronically decaying W boson. Using dedicated selections, five regions were defined, each enriched with jets from either bottom, charm, strange, or light quarks or from gluons. The relative contributions of jets of different flavors in each region were determined from data for jets with and without an s-tag by using a profile-likelihood fit. The tagging efficiency for jets from strange quarks and the mis-tag rates for all other jet flavors were determined for a working point with an efficiency of 40% for jets from strange quarks. The derived uncertainties of these efficiencies and mis-tag rates reach up to 50% for jets with a pseudorapidity $|\eta| \leq 1.5$ and up to 100% for jets with a pseudorapidity $|\eta| > 1.5$. For their reduction, it is necessary to extend the calibration strategy to be more sensitive to the flavor composition of jets from light quarks.

It was found that for jet with a pseudorapidity $|\eta| \leq 1.5$, the s-tagger has the ability to separate between s-jets and jets of other flavors, while for $|\eta| > 1.5$, the large uncertainties degrade the separation. Based on the measured efficiencies and mis-tag rates, scale factors to adjust the s-tagger response of simulated jets to match the response of jets in data were calculated, which allow for its application in data analyses. Thus, it provides a valuable addition to the flavor tagging algorithms already in use at the ATLAS Experiment.

The calibrated s-tagger was applied in an analysis setup designed to constrain the CKM matrix elements $|V_{ts}|$ and $|V_{td}|$ in top-quark decays. This setup uses events containing pair-produced top quarks decaying leptonically with an electron and a muon of opposite-sign electric charge in the final state. It was demonstrated that, in principle, it is possible to disentangle $t \to W^+ s$ and $t \to W^+ d$ decays using an s-tagger, which can be used to individually determine or constrain $|V_{ts}|$ and $|V_{td}|$ in top-quark decays. However, because $|V_{tb}|$ is close to unity, in the presented analysis setup, the background from $t \to W^+ b$ decays is so large that the sensitivity of the s-tagger is not sufficient to determine the ratio of the branching fractions of $t \to W^+ s$ and $t \to W^+ d$ decays.

Nonetheless, assuming unitarity of the CKM matrix, it was possible to derive a limit of $|V_{ts}|^2 + |V_{td}|^2 < 0.06$ at 95% confidence level. This limit has a similar size as other constraints on the top-quark related CKM matrix elements derived in top-quark decays. Using the s-tagger, it was also newly possible to derive limits on $|V_{ts}|$ and $|V_{td}|$ in the two-dimensional plane spanned by them. These limits can be interpreted as individual limits on $|V_{ts}|$ and $|V_{td}|$ without making any additional assumptions on the CKM matrix except for its unitarity. $|V_{ts}|$ and $|V_{td}|$ are found to be smaller than 0.21 and 0.24 at 95% confidence level, respectively.

For a better sensitivity to $t \to W^+ s$ and $t \to W^+ d$ decays, future attempts to determine or constrain $|V_{ts}|$ and $|V_{td}|$ in top-quark decays using an s-tagger will have to focus on the reduction of background from $t \to W^+ b$ decays.

Acknowledgements

Bibliography

[1] D0 Collaboration, *Observation of the top quark*, Phys. Rev. Lett. **74** (1995) 2632. arXiv: `hep-ex/9503003`.

[2] CDF Collaboration, *Observation of top quark production in $\bar{p}p$ collisions*, Phys. Rev. Lett. **74** (1995) 2626. arXiv: `hep-ex/9503002`.

[3] D0 Collaboration, *Precision measurement of the ratio $\mathcal{B}(t \to Wb)/\mathcal{B}(t \to Wq)$ and Extraction of V_{tb}*, Phys. Rev. Lett. **107** (2011) 121802. arXiv: `1106.5436 [hep-ex]`.

[4] CDF Collaboration, *Measurement of $R = \mathcal{B}(t \to Wb)/\mathcal{B}(t \to Wq)$ in Top–quark–pair Decays using Lepton+jets Events and the Full CDF Run II Data set*, Phys. Rev. D **87** (2013) 111101. arXiv: `1303.6142 [hep-ex]`.

[5] CDF Collaboration, *Measurement of $B(t \to Wb)/B(t \to Wq)$ in Top-Quark-Pair Decays Using Dilepton Events and the Full CDF Run II Data Set*, Phys. Rev. Lett. **112** (2014) 221801. arXiv: `1404.3392 [hep-ex]`.

[6] CMS Collaboration, *Measurement of the ratio $\mathcal{B}(t \to Wb)/\mathcal{B}(t \to Wq)$ in pp collisions at $\sqrt{s} = 8$ TeV*, Phys. Lett. B **736** (2014) 33. arXiv: `1404.2292 [hep-ex]`.

[7] D0 Collaboration, *Evidence for production of single top quarks and first direct measurement of $|Vtb|$*, Phys. Rev. Lett. **98** (2007) 181802. arXiv: `hep-ex/0612052`.

[8] D0 Collaboration, *Evidence for production of single top quarks*, Phys. Rev. D **78** (2008) 012005. arXiv: `0803.0739 [hep-ex]`.

[9] CDF Collaboration, *Measurement of the Single Top Quark Production Cross Section at CDF*, Phys. Rev. Lett. **101** (2008) 252001. arXiv: `0809.2581 [hep-ex]`.

[10] CDF Collaboration, *First Observation of Electroweak Single Top Quark Production*, Phys. Rev. Lett. **103** (2009) 092002. arXiv: `0903.0885 [hep-ex]`.

[11] CDF and D0 Collaboration, *Combination of CDF and D0 Measurements of the Single Top Production Cross Section*, 2009. arXiv: `0908.2171 [hep-ex]`.

[12] CDF Collaboration, *Observation of Single Top Quark Production and Measurement of $|V_{tb}|$ with CDF*, Phys. Rev. D **82** (2010) 112005. arXiv: `1004.1181 [hep-ex]`.

[13] CDF Collaboration, *Measurement of the Single Top Quark Production Cross Section and $|V_{tb}|$ in Events with One Charged Lepton, Large Missing Transverse Energy, and Jets at CDF*, Phys. Rev. Lett. **113** (2014) 261804. arXiv: `1407.4031 [hep-ex]`.

[14] CMS Collaboration, *Measurement of the t-channel single top quark production cross section in pp collisions at $\sqrt{s} = 7$ TeV*, Phys. Rev. Lett. **107** (2011) 091802. arXiv: `1106.3052 [hep-ex]`.

[15] ATLAS Collaboration, *Measurement of the t-channel single top-quark production cross section in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*,
Phys. Lett. B **717** (2012) 330. arXiv: `1205.3130` [`hep-ex`].

[16] ATLAS Collaboration, *Comprehensive measurements of t-channel single top-quark production cross sections at $\sqrt{s} = 7$ TeV with the ATLAS detector*,
Phys. Rev. D **90** (2014) 112006. arXiv: `1406.7844` [`hep-ex`].

[17] CMS Collaboration, *Measurement of the t-channel single-top-quark production cross section and of the $|V_{tb}|$ CKM matrix element in pp collisions at $\sqrt{s} = 8$ TeV*,
JHEP **06** (2014) 090. arXiv: `1403.7366` [`hep-ex`].

[18] ATLAS Collaboration,
Measurement of the inclusive cross-sections of single top-quark and top-antiquark t-channel production in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector,
JHEP **04** (2017) 086. arXiv: `1609.03920` [`hep-ex`].

[19] CMS Collaboration, *Cross section measurement of t-channel single top quark production in pp collisions at $\sqrt{s} = 13$ TeV*, Phys. Lett. B **772** (2017) 752.
arXiv: `1610.00678` [`hep-ex`].

[20] ATLAS Collaboration, *Fiducial, total and differential cross-section measurements of t-channel single top-quark production in pp collisions at 8 TeV using data collected by the ATLAS detector*, Eur. Phys. J. C **77** (2017) 531. arXiv: `1702.02859` [`hep-ex`].

[21] ATLAS and CMS Collaboration,
Combinations of single-top-quark production cross-section measurements and $|f_{LV} V_{tb}|$ determinations at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS experiments,
JHEP **05** (2019) 088. arXiv: `1902.07158` [`hep-ex`].

[22] CMS Collaboration,
Measurement of the single top quark and antiquark production cross sections in the t channel and their ratio in proton-proton collisions at $\sqrt{s} = 13$ TeV,
Phys. Lett. B **800** (2020) 135042. arXiv: `1812.10514` [`hep-ex`].

[23] CMS Collaboration, *Measurement of CKM matrix elements in single top quark t-channel production in proton-proton collisions at $\sqrt{s} = 13$ TeV*,
Phys. Lett. B **808** (2020) 135609. arXiv: `2004.12181` [`hep-ex`].

[24] Particle Data Group, *Review of Particle Physics*, Phys. Rev. D **98** (2018) 030001.

[25] LHCb Collaboration,
Precision measurement of the B_s^0-$\bar{B}_s^0$ oscillation frequency with the decay $B_s^0 \to D_s^- \pi^+$,
New J. Phys. **15** (2013) 053021. arXiv: `1304.4741` [`hep-ex`].

[26] HFLAV Collaboration,
Averages of b-hadron, c-hadron, and τ-lepton properties as of summer 2016,
Eur. Phys. J. C **77** (2017) 895. arXiv: `1612.07233` [`hep-ex`].

[27] S. Aoki *et al.*, *Review of lattice results concerning low-energy particle physics*,
Eur. Phys. J. C **77** (2017) 112. arXiv: `1607.00299` [`hep-lat`].

[28] C. G. Unverdorben, *Measurement of the CKM matrix element $|V_{ts}|^2$*,
doctoral thesis, Ludwig-Maximilians-Universität München, 2015. [Online]. Available:
`https://edoc.ub.uni-muenchen.de/18227/1/Unverdorben_Christopher_G.pdf`.

[29] Y. Nakai, D. Shih, and S. Thomas, *Strange Jet Tagging*, 2020.
arXiv: 2003.09517 [hep-ph].

[30] J. Erdmann,
A tagger for strange jets based on tracking information using long short-term memory,
JINST **15** (2020) 01021. arXiv: 1907.07505 [physics.ins-det].

[31] D. Galbraith and C. Burgard, *Standard Model of the Standard Model*, 2012.
[Online]. Available: http://www.texample.net/tikz/examples/model-physics/,
accessed 19.02.2021.

[32] Super-Kamiokande Collaboration, *Evidence for oscillation of atmospheric neutrinos*,
Phys. Rev. Lett. **81** (1998) 1562. arXiv: hep-ex/9807003.

[33] SNO Collaboration, *Measurement of the rate of $\nu_e + d \rightarrow p + p + e^-$ interactions
produced by ^8B solar neutrinos at the Sudbury Neutrino Observatory*,
Phys. Rev. Lett. **87** (2001) 071301. arXiv: nucl-ex/0106015.

[34] SNO Collaboration, *Direct evidence for neutrino flavor transformation from neutral
current interactions in the Sudbury Neutrino Observatory*,
Phys. Rev. Lett. **89** (2002) 011301. arXiv: nucl-ex/0204008.

[35] ATLAS Collaboration, *Observation of a new particle in the search for the Standard
Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1.
arXiv: 1207.7214 [hep-ex].

[36] CMS Collaboration, *Observation of a New Boson at a Mass of* 125 GeV *with the CMS
Experiment at the LHC*, Phys. Lett. B **716** (2012) 30. arXiv: 1207.7235 [hep-ex].

[37] ATLAS, CDF, CMS, and D0 Collaboration,
First combination of Tevatron and LHC measurements of the top-quark mass, 2014.
arXiv: 1403.4427 [hep-ex].

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

[39] ALICE Collaboration, *Investigations of Anisotropic Flow Using Multiparticle
Azimuthal Correlations in pp, p-Pb, Xe-Xe, and Pb-Pb Collisions at the LHC*,
Phys. Rev. Lett. **123** (2019) 142301. arXiv: 1903.01790 [nucl-ex].

[40] E. Mobs, *The CERN accelerator complex*, August 2018.
[Online]. Available: http://cds.cern.ch/record/2636343.

[41] ALICE Collaboration, *The ALICE experiment at the LHC*,
Nucl. Phys. A **805** (2008) 511.

[42] LHCb Collaboration, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005.

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

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

[45] ATLAS Collaboration, *Peak luminosity per fill*, 2018.
[Online]. Available: https://atlas.web.cern.ch/Atlas/GROUPS/DATAPREPARATION/
PublicPlots/2018/DataSummary/figs/peakLumiByFill.pdf, accessed 19.2.2021.

[46] ATLAS Collaboration,
      Track Reconstruction Performance of the ATLAS Inner Detector at $\sqrt{s} = 13$ TeV,
      ATL-PHYS-PUB-2015-018, 2015.
      [Online]. Available: `https://cds.cern.ch/record/2037683`.

[47] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*,
      ATLAS-TDR-19, 2010. [Online]. Available: `https://cds.cern.ch/record/1291633`,
      addendum: `https://cds.cern.ch/record/1451888`.

[48] B. Abbott *et al.*, *Production and integration of the ATLAS Insertable B-Layer*,
      JINST **13** (2018) T05008. arXiv: 1803.00844 [`physics.ins-det`].

[49] ATLAS Collaboration, *Operation of the ATLAS trigger system in Run 2*,
      JINST **15** (2020) P10004. arXiv: 2007.12539 [`physics.ins-det`].

[50] ATLAS Collaboration, *AnalysisTop 21.2.147*, July 2020. [Online]. Available:
      `https://gitlab.cern.ch/atlas/athena/-/tree/release/21.2.147`.

[51] T. Cornelissen, M. Elsing, I. Gavrilenko, *et al.*,
      The new ATLAS track reconstruction (NEWT),
      J. Phys. Conf. Ser. **119** (2008) 032014.

[52] ATLAS Collaboration, *Performance of the ATLAS Track Reconstruction Algorithms
      in Dense Environments in LHC Run 2*, Eur. Phys. J. C **77** (2017) 673.
      arXiv: 1704.07983 [`hep-ex`].

[53] R. Fruhwirth, *Application of Kalman filtering to track and vertex fitting*,
      Nucl. Instrum. Meth. A **262** (1987) 444.

[54] ATLAS Collaboration, *Electron and photon performance measurements with the
      ATLAS detector using the 2015–2017 LHC proton-proton collision data*,
      JINST **14** (2019) P12006. arXiv: 1908.00005 [`hep-ex`].

[55] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its
      performance in LHC Run 1*, Eur. Phys. J. C **77** (2017) 490.
      arXiv: 1603.02934 [`hep-ex`].

[56] 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.
      arXiv: 1603.05598 [`hep-ex`].

[57] ATLAS Collaboration, *Muon reconstruction and identification efficiency in ATLAS
      using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV*,
      Eur. Phys. J. C **81** (2021) 578. arXiv: 2012.00578 [`hep-ex`].

[58] F. Meloni, *Primary vertex reconstruction with the ATLAS detector*,
      JINST **11** (2016) C12060.

[59] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-kt jet clustering algorithm*,
      JHEP (2008) 063.

[60] ATLAS Collaboration,
      *Properties of Jets and Inputs to Jet Reconstruction and Calibration with the ATLAS
      Detector Using Proton-Proton Collisions at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-036,
      2015. [Online]. Available: `https://cds.cern.ch/record/2044564`.

[61] ATLAS Collaboration, *Jet energy scale and resolution measured in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, 2020.
arXiv: `2007.02645 [hep-ex]`.

[62] ATLAS Collaboration, *Tagging and suppression of pileup jets*,
ATLAS-CONF-2014-018, 2014.
[Online]. Available: `https://cds.cern.ch/record/1700870`.

[63] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV*,
Eur. Phys. J. C **79** (2019) 970. arXiv: `1907.05120 [hep-ex]`.

[64] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*, ATL-PHYS-PUB-2017-013, 2017.
[Online]. Available: `https://cds.cern.ch/record/2273281`.

[65] ATLAS Collaboration, *Measurement of the b-tag Efficiency in a Sample of Jets Containing Muons with 5 fb^{-1} of Data from the ATLAS Detector*,
ATLAS-CONF-2012-043, 2012.
[Online]. Available: `https://cds.cern.ch/record/1435197`.

[66] V. Vecchio, R. A. Creager, B. Liu, *et al.*,
Measuring the b-jet Identification Efficiency in a Sample of Jets Containing Muons in pp Collisions at $\sqrt{s} = 13$ TeV recorded with the ATLAS Detector,
ATL-COM-PHYS-2019-096, 2021.
[Online]. Available: `https://cds.cern.ch/record/2658016`, internal documentation.

[67] ATLAS Collaboration, *Direct constraint on the Higgs-charm coupling from a search for Higgs boson decays to charm quarks with the ATLAS detector*,
ATLAS-CONF-2021-021, 2021.
[Online]. Available: `https://cds.cern.ch/record/2771724`.

[68] A. Chisholm, T. J. Neep, P. Thompson, *et al.*,
Search for $H \to c\bar{c}$ decays in associated VH production with the ATLAS experiment,
ATL-COM-PHYS-2020-013, 2020.
[Online]. Available: `https://cds.cern.ch/record/2706457`, internal documentation.

[69] ATLAS Collaboration, *Calibration of light-flavour jet b-tagging rates on ATLAS proton-proton collision data at $\sqrt{s} = 13$ TeV*, ATLAS-CONF-2018-006, 2018.
[Online]. Available: `https://cds.cern.ch/record/2314418`.

[70] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 13$ TeV*,
Eur. Phys. J. C **78** (2018) 903. arXiv: `1802.08168 [hep-ex]`.

[71] ATLAS Collaboration, *Indetv0finder*, 2014. [Online]. Available:
`https://gitlab.cern.ch/atlas/athena/blob/21.2/InnerDetector/`
`InDetRecAlgs/InDetV0Finder/InDetV0Finder/InDetV0FinderTool.h`.

[72] H. C. Cheng, *Correlation Measurement of $\Lambda^0\overline{\Lambda}^0$ $\Lambda^0\Lambda^0$ and $\overline{\Lambda}^0\overline{\Lambda}^0$ with the ATLAS Detector at $\sqrt{s} = 7$ TeV*, CERN-THESIS-2017-042, doctoral thesis, 2017.
[Online]. Available: `https://cds.cern.ch/record/2265715`.

[73] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*, ATLAS-CONF-2019-021, 2019.
[Online]. Available: `https://cds.cern.ch/record/2677054`.

[74] S. Amoroso, F. Siegert, J. Kretzschmar, and C. Gutschow, *PMG references document*, ATL-COM-PHYS-2019-701, 2019.
[Online]. Available: `https://cds.cern.ch/record/2678867`, internal documentation.

[75] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*,
Eur. Phys. J. C **70** (2010) 823. arXiv: `1005.4568` [`physics.ins-det`].

[76] S. Agostinelli *et al.*, *GEANT4: A Simulation toolkit*,
Nucl. Instrum. Meth. A **506** (2003) 250.

[77] T. Sjöstrand, S. Ask, J. R. Christiansen, *et al.*, *An Introduction to PYTHIA 8.2*,
Comput. Phys. Commun. **191** (2015) 159. arXiv: `1410.3012` [`hep-ph`].

[78] R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. Del Debbio, *et al.*,
Parton distributions with LHC data, Nucl. Phys. B **867** (2013) 244.
arXiv: `1207.1303` [`hep-ph`].

[79] ATLAS Collaboration, *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie-Landshoff diffractive model*, ATL-PHYS-PUB-2016-017, 2016.
[Online]. Available: `https://cds.cern.ch/record/2206965`.

[80] ATLAS Collaboration, *Measurement of the Inelastic Proton–Proton Cross Section at $\sqrt{s} = 13$ TeV with the ATLAS Detector at the LHC*,
Phys. Rev. Lett. **117** (2016) 182002. arXiv: `1606.02625` [`hep-ex`].

[81] S. Frixione, P. Nason, and G. Ridolfi, *A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, JHEP **09** (2007) 126.
arXiv: `0707.3088` [`hep-ph`].

[82] P. Nason,
A New method for combining NLO QCD with shower Monte Carlo algorithms,
JHEP **11** (2004) 040. arXiv: `hep-ph/0409146`.

[83] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **11** (2007) 070.
arXiv: `0709.2092` [`hep-ph`].

[84] 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. arXiv: `1002.2581` [`hep-ph`].

[85] NNPDF Collaboration, *Parton distributions for the LHC Run II*,
JHEP **04** (2015) 040. arXiv: `1410.8849` [`hep-ph`].

[86] S. Frixione, E. Laenen, P. Motylinski, and B. R. Webber, *Angular correlations of lepton pairs from vector boson and top quark decays in Monte Carlo simulations*,
JHEP **0704** (2007) 081. arXiv: `hep-ph/0702198`.

[87] P. Artoisenet, R. Frederix, O. Mattelaer, and R. Rietkerk,
Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations,
JHEP **1303** (2013) 015. arXiv: `1212.3460` [`hep-ph`].

[88] ATLAS Collaboration, *ATLAS Pythia 8 tunes to* 7 TeV *data*,
ATL-PHYS-PUB-2014-021, 2014.
[Online]. Available: `https://cds.cern.ch/record/1966419`.

[89] D. J. Lange, *The EvtGen particle decay simulation package*,
Nucl. Instrum. Meth. A **462** (2001) 152.

[90] M. Beneke, P. Falgari, S. Klein, and C. Schwinn,
Hadronic top-quark pair production with NNLL threshold resummation,
Nucl. Phys. B **855** (2012) 695. arXiv: `1109.1536 [hep-ph]`.

[91] M. Cacciari, M. Czakon, M. Mangano, A. Mitov, and P. Nason, *Top-pair production
at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*,
Phys. Lett. B **710** (2012) 612. arXiv: `1111.5869 [hep-ph]`.

[92] P. Bärnreuther, M. Czakon, and A. Mitov, *Percent Level Precision Physics at the
Tevatron: First Genuine NNLO QCD Corrections to* $q\bar{q} \to t\bar{t} + X$,
Phys. Rev. Lett. **109** (2012) 132001. arXiv: `1204.5201 [hep-ph]`.

[93] M. Czakon and A. Mitov, *NNLO corrections to top-pair production at hadron
colliders: the all-fermionic scattering channels*, JHEP **12** (2012) 054.
arXiv: `1207.0236 [hep-ph]`.

[94] M. Czakon and A. Mitov,
NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction,
JHEP **1301** (2013) 080. arXiv: `1210.6832 [hep-ph]`.

[95] M. Czakon, P. Fiedler, and A. Mitov,
The total top quark pair production cross-section at hadron colliders through $O(\alpha_S^4)$,
Phys. Rev. Lett. **110** (2013) 252004. arXiv: `1303.6254 [hep-ph]`.

[96] M. Czakon and A. Mitov, *Top++: A Program for the Calculation of the Top-Pair
Cross-Section at Hadron Colliders*, Comput. Phys. Commun. **185** (2014) 2930.
arXiv: `1112.5675 [hep-ph]`.

[97] M. Czakon, D. Heymes, A. Mitov, *et al.*,
Top-pair production at the LHC through NNLO QCD and NLO EW,
JHEP **10** (2017) 186. arXiv: `1705.04105 [hep-ph]`.

[98] L. Serkin, J. Margo, M. Faraj, G. Panizzo, and M. Pinamonti,
Ttbarnnloreweighter package, 2019.
[Online]. Available: `https://gitlab.cern.ch/lserkin/TTbarNNLOReweighter`,
accessed September 2020.

[99] M. Bahr et al., *Herwig++ Physics and Manual*, Eur. Phys. J. C **58** (2008) 639.
arXiv: `0803.0883 [hep-ph]`.

[100] J. Bellm *et al.*, *Herwig 7.0/Herwig++ 3.0 release note*, Eur. Phys. J. C **76** (2016) 196.
arXiv: `1512.01178 [hep-ph]`.

[101] L. Harland-Lang, A. Martin, P. Motylinski, and R. Thorne,
Parton distributions in the LHC era: MMHT 2014 PDFs,
Eur. Phys. J. C **75** (2015) 204. arXiv: `1412.3989 [hep-ph]`.

[102] E. Re, *Single-top Wt-channel production matched with parton showers using the
POWHEG method*, Eur. Phys. J. C **71** (2011) 1547. arXiv: `1009.2450 [hep-ph]`.

[103] S. Frixione, E. Laenen, P. Motylinski, B. R. Webber, and C. D. White,
Single-top hadroproduction in association with a W boson, JHEP **0807** (2008) 029.
arXiv: 0805.3067 [hep-ph].

[104] M. Aliev, H. Lacker, U. Langenfeld, *et al.*,
HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR,
Comput. Phys. Commun. **182** (2011) 1034. arXiv: 1007.1327 [hep-ph].

[105] P. Kant, O. M. Kind, T. Kintscher, *et al.*,
*HatHor for single top-quark production: Updated predictions and uncertainty estimates
for single top-quark production in hadronic collisions*,
Comput. Phys. Commun. **191** (2015) 74. arXiv: 1406.4403 [hep-ph].

[106] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated
production with a W^- or H^-*, Phys. Rev. D **82** (2010) 054018.
arXiv: 1005.4451 [hep-ph].

[107] N. Kidonakis, *Top Quark Production*,
Helmholtz International Summer School on Physics of Heavy Quarks and Hadrons,
2013. arXiv: 1311.0283 [hep-ph].

[108] T. Gleisberg, S. Hoeche, F. Krauss, *et al.*, *Event generation with SHERPA 1.1*,
JHEP **02** (2009) 007. arXiv: 0811.4622 [hep-ph].

[109] C. Anastasiou, L. J. Dixon, K. Melnikov, and F. Petriello, *High precision QCD at
hadron colliders: Electroweak gauge boson rapidity distributions at NNLO*,
Phys. Rev. D **69** (2004) 094008. arXiv: hep-ph/0312266.

[110] ATLAS Collaboration, *ATLAS simulation of boson plus jets processes in Run 2*,
ATL-PHYS-PUB-2017-006, 2017.
[Online]. Available: http://cds.cern.ch/record/2261937.

[111] E. Bothmann, M. Schönherr, and S. Schumann,
Reweighting QCD matrix-element and parton-shower calculations,
Eur. Phys. J. C **76** (2016) 590. arXiv: 1606.08753 [hep-ph].

[112] ATLAS Collaboration, *Multi-Boson Simulation for* 13 TeV *ATLAS Analyses*,
ATL-PHYS-PUB-2017-005, 2017.
[Online]. Available: http://cds.cern.ch/record/2261933.

[113] S. Myers and E. Picasso, *The Design, construction and commissioning of the CERN
Large Electron Positron collider*, Contemp. Phys. **31** (1990) 387.

[114] SLD Collaboration, *Light quark fragmentation in polarized Z0 decays at SLD*,
Nucl. Phys. B Proc. Suppl. **96** (2001) 31. arXiv: hep-ex/0008032.

[115] DELPHI Collaboration,
Measurement of the strange quark forward backward asymmetry around the Z0 peak,
Eur. Phys. J. C **14** (2000) 613.

[116] M. Abadi *et al.*,
TensorFlow: Large-scale machine learning on heterogeneous distributed systems, 2015.
[Online]. Available: https://www.tensorflow.org.

[117] F. Chollet *et al.*, *Keras*, 2015. [Online]. Available: https://keras.io.

[118]   *Visualization of a deep neural network*, 2017. [Online]. Available:
`https://tex.stackexchange.com/questions/353993/neural-network-graph`,
accessed 1.9.2019.

[119]   X. Glorot, A. Bordes, and Y. Bengio, *Deep sparse rectifier neural networks*,
Proceedings of the Fourteenth International Conference on Artificial Intelligence and
Statistics, Proceedings of Machine Learning Research **15** (2011) 315.
[Online]. Available: `http://proceedings.mlr.press/v15/glorot11a.html`.

[120]   S. Hochreiter and J. Schmidhuber, *Long Short-Term Memory*,
Neural Comput. **9** (1997) 1735.

[121]   F. Pedregosa *et al.*, *Scikit-learn: Machine Learning in Python*,
J. Machine Learning Res. **12** (2011) 2825. arXiv: `1201.0490`.

[122]   I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016,
`http://www.deeplearningbook.org`.

[123]   D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, 2014.
arXiv: `1412.6980 [cs.LG]`.

[124]   R. Tibshirani, *Regression Shrinkage and Selection via the Lasso*,
Journal of the Royal Statistical Society. Series B (Methodological) **58** (1996) 267.

[125]   J. Erdmann, O. Nackenhorst, and S. V. Zeißner,
Maximum performance of strange-jet tagging at hadron colliders, JINST (in press).
arXiv: `2011.10736 [hep-ex]`.

[126]   J. Pumplin, *How to tell quark jets from gluon jets*, Phys. Rev. D **44** (1991) 2025.

[127]   CMS collaboration, *Performance of quark/gluon discrimination in* 8 TeV *pp data*,
CMS-PAS-JME-13-002, 2013.
[Online]. Available: `https://cds.cern.ch/record/1599732`.

[128]   A. Collaboration, *Light-quark and gluon jet discrimination in pp collisions at*
$\sqrt{s} = 7$ TeV *with the ATLAS detector*, Eur. Phys. J. C **74** (2014) 3023.
arXiv: `1405.6583 [hep-ex]`.

[129]   P. T. Komiske, E. M. Metodiev, and M. D. Schwartz,
Deep learning in color: towards automated quark/gluon jet discrimination,
JHEP **01** (2017) 110. arXiv: `1612.01551 [hep-ph]`.

[130]   T. Cheng, *Recursive Neural Networks in Quark/Gluon Tagging*,
Comput. Softw. Big Sci. **2** (2018) 3. arXiv: `1711.02633 [hep-ph]`.

[131]   E. M. Metodiev, B. Nachman, and J. Thaler,
Classification without labels: Learning from mixed samples in high energy physics,
JHEP **10** (2017) 174. arXiv: `1708.02949 [hep-ph]`.

[132]   L. M. Dery, B. Nachman, F. Rubbo, and A. Schwartzman,
Weakly Supervised Classification in High Energy Physics, JHEP **05** (2017) 145.
arXiv: `1702.00414 [hep-ph]`.

[133]   E. M. Metodiev and J. Thaler,
Jet Topics: Disentangling Quarks and Gluons at Colliders,
Phys. Rev. Lett. **120** (2018) 241602. arXiv: `1802.00008 [hep-ph]`.

[134] P. T. Komiske, E. M. Metodiev, and J. Thaler,
An operational definition of quark and gluon jets, JHEP **11** (2018) 059.
arXiv: 1809.01140 [hep-ph].

[135] P. T. Komiske, E. M. Metodiev, and J. Thaler,
Energy Flow Networks: Deep Sets for Particle Jets, JHEP **01** (2019) 121.
arXiv: 1810.05165 [hep-ph].

[136] G. Kasieczka, N. Kiefer, T. Plehn, and J. M. Thompson,
Quark-Gluon Tagging: Machine Learning vs Detector, SciPost Phys. **6** (2019) 069.
arXiv: 1812.09223 [hep-ph].

[137] V. Mikuni and F. Canelli, *ABCNet: An attention-based method for particle tagging*,
Eur. Phys. J. Plus **135** (2020) 463. arXiv: 2001.05311 [physics.data-an].

[138] G. Kasieczka, S. Marzani, G. Soyez, and G. Stagnitto,
Towards Machine Learning Analytics for Jet Substructure, JHEP **09** (2020) 195.
arXiv: 2007.04319 [hep-ph].

[139] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer,
MadGraph 5 : Going Beyond, JHEP **06** (2011) 128. arXiv: 1106.0522 [hep-ph].

[140] N. Abicht, *Study of convolutional neural networks for strange-tagging based on jet
images from calorimeters*, master thesis, TU Dortmund, Fakultät Physik, 2020.

[141] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet User Manual*,
Eur. Phys. J. C **72** (2012) 1896. arXiv: 1111.6097 [hep-ph].

[142] M. Cacciari, G. P. Salam, and G. Soyez, *The Catchment Area of Jets*,
JHEP **04** (2008) 005. arXiv: 0802.1188 [hep-ph].

[143] CMS Collaboration, *Measurement of charged pion, kaon, and proton production in
proton-proton collisions at $\sqrt{s} = 13$ TeV*, Phys. Rev. D **96** (2017) 112003.
arXiv: 1706.10194 [hep-ex].

[144] CMS Collaboration, *The Phase1 CMS Pixel detector upgrade*,
JINST **11** (2016) C12010.

[145] ALICE Collaboration, *The ALICE experiment at the CERN LHC*,
JINST **3** (2008) S08002.

[146] J. de Favereau, C. Delaere, P. Demin, *et al.*, *DELPHES 3: a modular framework for
fast simulation of a generic collider experiment*, JHEP **02** (2014) 057.
arXiv: 1307.6346 [hep-ex].

[147] G. Punzi, *Notes on statistical separation of classes of events*, 2006.
arXiv: physics/0611219 [physics.data-an].

[148] T. Barillari, E. Bergeaas Kuutmann, T. Carli, *et al.*, *Local Hadronic Calibration*,
ATL-LARG-PUB-2009-001-2, 2008.
[Online]. Available: https://cds.cern.ch/record/1112035.

[149] A. Banfi, G. P. Salam, and G. Zanderighi, *Infrared safe definition of jet flavor*,
Eur. Phys. J. C **47** (2006) 113. arXiv: hep-ph/0601139.

[150] A. Buckley and C. Pollard,
QCD-aware partonic jet clustering for truth-jet flavour labelling,
Eur. Phys. J. C **76** (2016) 71. arXiv: 1507.00508 [hep-ph].

[151] ATLAS Collaboration, *Quark versus Gluon Jet Tagging Using Charged Particle Multiplicity with the ATLAS Detector*, ATL-PHYS-PUB-2017-009, 2017.
[Online]. Available: `https://cds.cern.ch/record/2263679`.

[152] E. Evsenin-Gutschank, *Development of a calibration strategy for the first strange tagger with the ATLAS detector at $\sqrt{s} = 13$ TeV using Run-2 data*,
master thesis, TU Dortmund, Fakultät Physik, 2019.

[153] J. Erdmann, S. Guindon, K. Kröninger, *et al.*, *A likelihood-based reconstruction algorithm for top-quark pairs and the KLFitter framework*,
Nucl. Instrum. Meth. A **748** (2014) 18. arXiv: `1312.5595 [hep-ex]`.

[154] ATLAS Collaboration, *Measurement of b-tagging Efficiency of c-jets in $t\bar{t}$ Events Using a Likelihood Approach with the ATLAS Detector*, ATLAS-CONF-2018-001,
2018. [Online]. Available: `https://cds.cern.ch/record/2306649`.

[155] D0 Collaboration, *Measurement of the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV using kinematic characteristics of lepton + jets events*,
Phys. Rev. D **76** (2007) 092007. arXiv: `0705.2788 [hep-ex]`.

[156] ATLAS Collaboration,
Estimation of non-prompt and fake lepton backgrounds in final states with top quarks produced in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector,
ATLAS-CONF-2014-058, 2014.
[Online]. Available: `https://cds.cern.ch/record/1951336`.

[157] K. Cranmer, G. Lewis, L. Moneta, A. Shibata, and W. Verkerke,
HistFactory: A tool for creating statistical models for use with RooFit and RooStats,
CERN-OPEN-2012-016, 2012.
[Online]. Available: `https://cds.cern.ch/record/1456844`.

[158] M. Pinamonti, L. Valery, A. Held, and T. Dado, *Trexfitter version 4.12*,
December 2020. [Online]. Available:
`https://gitlab.cern.ch/TRExStats/TRExFitter/-/blob/TRExFitter-00-04-12`.

[159] R. J. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*,
Comput. Phys. Commun. **77** (1993) 219.

[160] F. James and M. Goossens, *Minuit – function minimization and error analysis*.
[Online]. Available: `https://root.cern.ch/download/minuit.pdf`,
accessed 29.05.2021.

[161] W. C. Davidon, *Variable metric method for minimization*,
SIAM Journal on Optimization **1** (1991) 1.
[Online]. Available: `https://doi.org/10.1137/0801001`.

[162] R. Fletcher, *A new approach to variable metric algorithms*,
The Computer Journal **13** (1970) 317.
[Online]. Available: `https://doi.org/10.1093/comjnl/13.3.317`.

[163] M. J. D. Powell, *Variable metric methods for constrained optimization*,
Mathematical Programming – The State of the Art [Bonn 1982], p. 288,
Springer Berlin Heidelberg, 1983.
[Online]. Available: `https://doi.org/10.1007/978-3-642-68874-4_12`.

[164] ATLAS Collaboration,
Studies on top-quark Monte Carlo modelling with Sherpa and MG5_aMC@NLO,
ATL-PHYS-PUB-2017-007, 2017.
[Online]. Available: `https://cds.cern.ch/record/2261938`.

[165] ATLAS Collaboration,
Performance of electron and photon triggers in ATLAS during LHC Run 2,
Eur. Phys. J. C **80** (2020) 47. arXiv: `1909.00761 [hep-ex]`.

[166] D. Vannicola, *Reconstruction and identification of high-pT muons in $\sqrt{s} = 13$ TeV proton-proton collisions with the ATLAS detector*, ATL-PHYS-PROC-2019-110, 2019.
[Online]. Available: `https://cds.cern.ch/record/2691453`.

[167] ATLAS Collaboration, *Performance of the ATLAS muon triggers in Run 2*,
JINST **15** (2020) P09015. arXiv: `2004.13447 [physics.ins-det]`.

[168] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*,
Phys. Rev. D **96** (2017) 072002. arXiv: `1703.09665 [hep-ex]`.

[169] ATLAS Collaboration, *Determination of jet calibration and energy resolution in proton-proton collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*,
Eur. Phys. J. C **80** (2020) 1104. arXiv: `1910.04482 [hep-ex]`.

[170] ATLAS Collaboration, *Tagging and suppression of pileup jets*,
ATLAS-CONF-2014-018, 2014.
[Online]. Available: `https://cds.cern.ch/record/1700870`.

[171] G. Cowan, K. Cranmer, E. Gross, and O. Vitells,
Asymptotic formulae for likelihood-based tests of new physics,
Eur. Phys. J. C **71** (2011) 1554, [Erratum: Eur. Phys. J. C **73** (2013) 2501].
arXiv: `1007.1727 [physics.data-an]`.

[172] J. Neyman and E. S. Pearson,
On the Problem of the Most Efficient Tests of Statistical Hypotheses,
Phil. Trans. Roy. Soc. Lond. A **231** (1933) 289.

[173] G. Cowan, *The CL_s method: information for conference speakers*, 2011.
[Online]. Available: `http://www.pp.rhul.ac.uk/~cowan/atlas/CLsInfo.pdf`,
accessed on 15.02.2021.

[174] T. Junk, *Confidence level computation for combining searches with small statistics*,
Nucl. Instrum. Meth. A **434** (1999) 435. arXiv: `hep-ex/9902006`.

[175] A. Wald, *Tests of statistical hypotheses concerning several parameters when the number of observations is large*,
Transactions of the American Mathematical Society **54** (1943) 426.
[Online]. Available: `http://www.jstor.org/stable/1990256`.

[176] W. Verkerke and D. Kirkby, *The RooFit toolkit for data modeling*, 2003.
arXiv: `physics/0306116`.

Acronyms

AUC   area under the curve
BOOSTER Proton Synchrotron Booster
CERN European Council for Nuclear
        Research
CKM   Cabibbo-Kobayashi-Maskawa
CL   confidence level
CR   control region
CSC   cathode strip chamber
DNN   deep neural network
DOF   degrees of freedom
ECAL electromagnetic calorimeter
EM   electromagnetic
FSR   final-state radiation
HCAL hadronic calorimeter
ID   inner detector
IP   impact parameter
ISR   initial-state radiation
JER   jet energy resolution
JES   jet energy scale
JVT   jet-vertex tagger
LAr   liquid-argon
LEP   Large Electron-Positron Collider
LHC   Large Hadron Collider
LINAC2 Linear Accelerator 2
LO   leading order
LS1   Long Shutdown 1
LS2   Long Shutdown 2
LSTM Long Short-Term Memory

MC   Monte Carlo
MDT   monitored drift tubes
ME   matrix element
MS   muon spectrometer
NLO   next-to-leading order
NNLL next-to-next-to-leading logarithmic
NNLO next-to-next-to-leading order
NN   neural network
NP   nuisance parameter
PDF   parton distribution function
PS   Proton Synchrotron
PV   primary vertex
QCD   quantum chromodynamics
ReLU rectified linear unit
RNN   recurrent neural network
ROC   receiver operating characteristic
RoI   region of interest
RPC   resistive plate chamber
SCT   silicon microstrip tracker
SF   scale factor
SLAC Stanford Linear Accelerator Center
SM   Standard Model of Particle Physics
SPS   Super Proton Synchrotron
SR   signal region
SSB   sponaneous symmetry breaking
SV   secondary vertex
TGC   thin gap chamber
TRT   transition radiation tracker
WP   working point

APPENDIX A

s-Tagger Input Variables Based on Reconstructed K_S and Λ^0

This chapter of the appendix gives an overview over the variables based on reconstructed K_S and Λ^0 which are used as input features of the s-tagger trained for the ATLAS Experiment. All distributions are shown twice, once for jets containing at least one reconstructed K_S and once for jets containing at least one reconstructed Λ^0.

The most prominent physical differences between K_S and Λ^0 are their lifetimes ($\tau(K_S) = 0.08954 \pm 0.00004$ ns, $\tau(\Lambda^0) = 0.2632 \pm 0.00020$ ns [24]) and their masses ($m(K_S) = 497.611 \pm 0.013$ MeV, $m(\Lambda^0) = 1115.683 \pm 0.006$ MeV [24]).
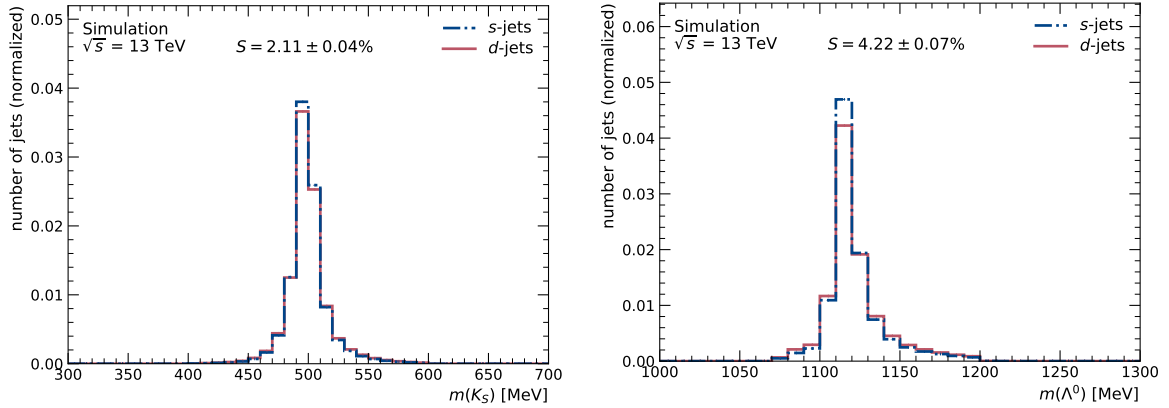
(a) Mass of the reconstructed K_S.

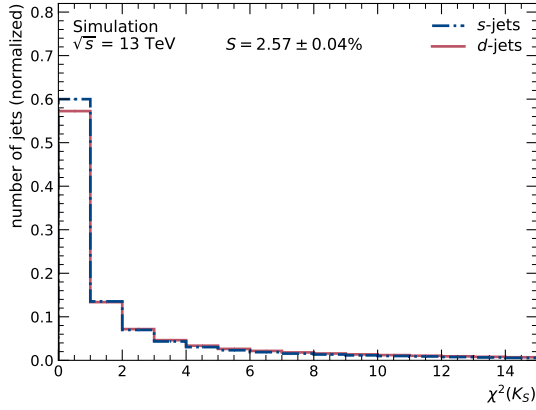(b) Mass of the reconstructed Λ^0.

Fig. A.1: Distributions of the masses of the reconstructed strange hadrons for d-jets and s-jets.

Figure A.1a and A.1b show the reconstructed mass of K_S and Λ^0 matched to s- and d-jets. Both mass distributions peak around the central mass value of each hadron. For s-jets, the peaks are slightly more pronounced since these jets should contain fewer fake K_S and Λ^0. The width of the mass peaks is dominated by the resolution of the detector.

Figures A.2a and A.2b illustrate the quality of the reconstruction of K_S and Λ by showing the χ^2 value of the reconstruction. For s-jets, χ^2 tends to be smaller which, again, hints at fewer fake K_S and Λ^0 matched to s-jets than to d-jets.

Figures A.3a and A.3b show the reconstructed decay times τ of the K_S and Λ^0 in their inertial frame. All curves show an exponential behavior in line with the expected lifetimes. However, there are so-called *turn-on* effects for small decay times, i.e. an inefficiency of reconstruction for those K_S or Λ^0 that decay too early to be reconstructed. Since in s-jets, the strange hadrons tend to be more energetic, on average, smaller decay times can be reconstructed.

Figures A.4a and A.4b illustrate that strange hadrons inside of s-jets have higher p_T than those reconstructed in d-jets. Similar behavior can be seen for the fraction of p_T of an s- or
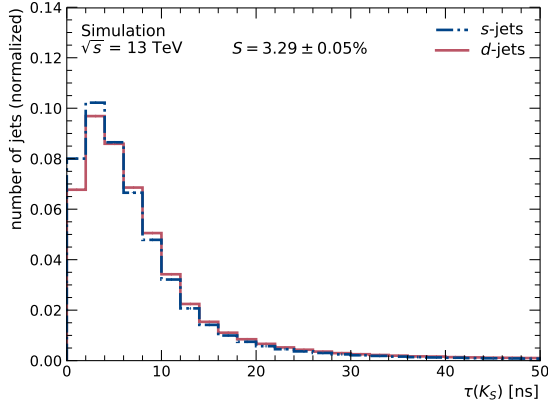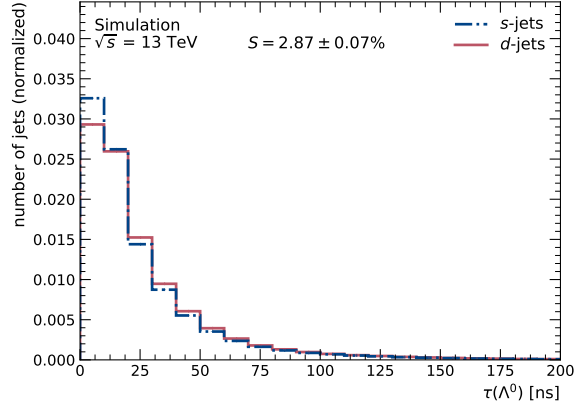
(a) χ^2 of the reconstruction of K_S.
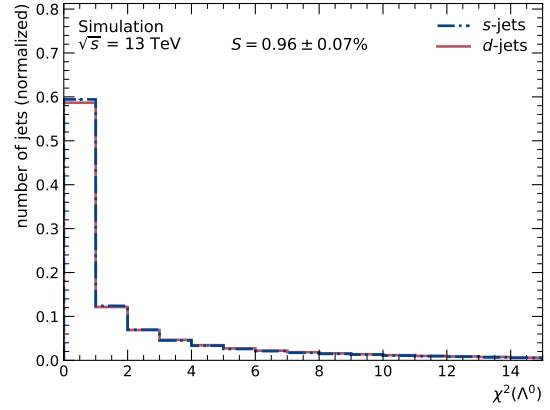
(b) χ^2 of the reconstruction of Λ^0.

Fig. A.2: Distributions of the χ^2 of the reconstructed strange hadrons matched to d-jets and s-jets.

(a) Distribution of the decay times of reconstructed K_S.

(b) Distribution of the decay times of reconstructed Λ^0.

Fig. A.3: Distributions of the decay times of the reconstructed strange hadrons in their inertial frame for strange hadrons matched to either s-jets or d-jets. The decay time is calculated from their reconstructed transverse momentum and the radial distance of the decay vertex from the primary vertex.

d-jet that is carried by a K_S or Λ^0, cf. Figures A.4c and A.4d.

Now taking a look at the kinematic distributions of the s- and d-jets, it is found that the selection of jets with either K_S or Λ^0 matched to them changes the distributions of the p_T and η of these jets. Figures A.5a and A.5b show the η distributions for jets containing K_S and Λ^0, respectively. The η distribution of jets containing K_S are almost as isotropically distributed as all s- and d-jets from W decays available for the training of the NN (cf. Figure 6.22b), which means that the lifetime of the K_S is short enough and their momentum is large enough to not see large effects of decay lengths in the detector. The reconstruction of Λ^0, however, is more efficient in the forward direction (i.e. for large $|\eta|$), because in this direction, the decay length

(a) Transverse momentum of reconstructed K_S.

(b) Transverse momentum of reconstructed Λ^0.

(c) Transverse momentum of jets that is carried by a reconstructed K_S.

(d) Transverse momentum of jets that is carried by a reconstructed Λ^0.
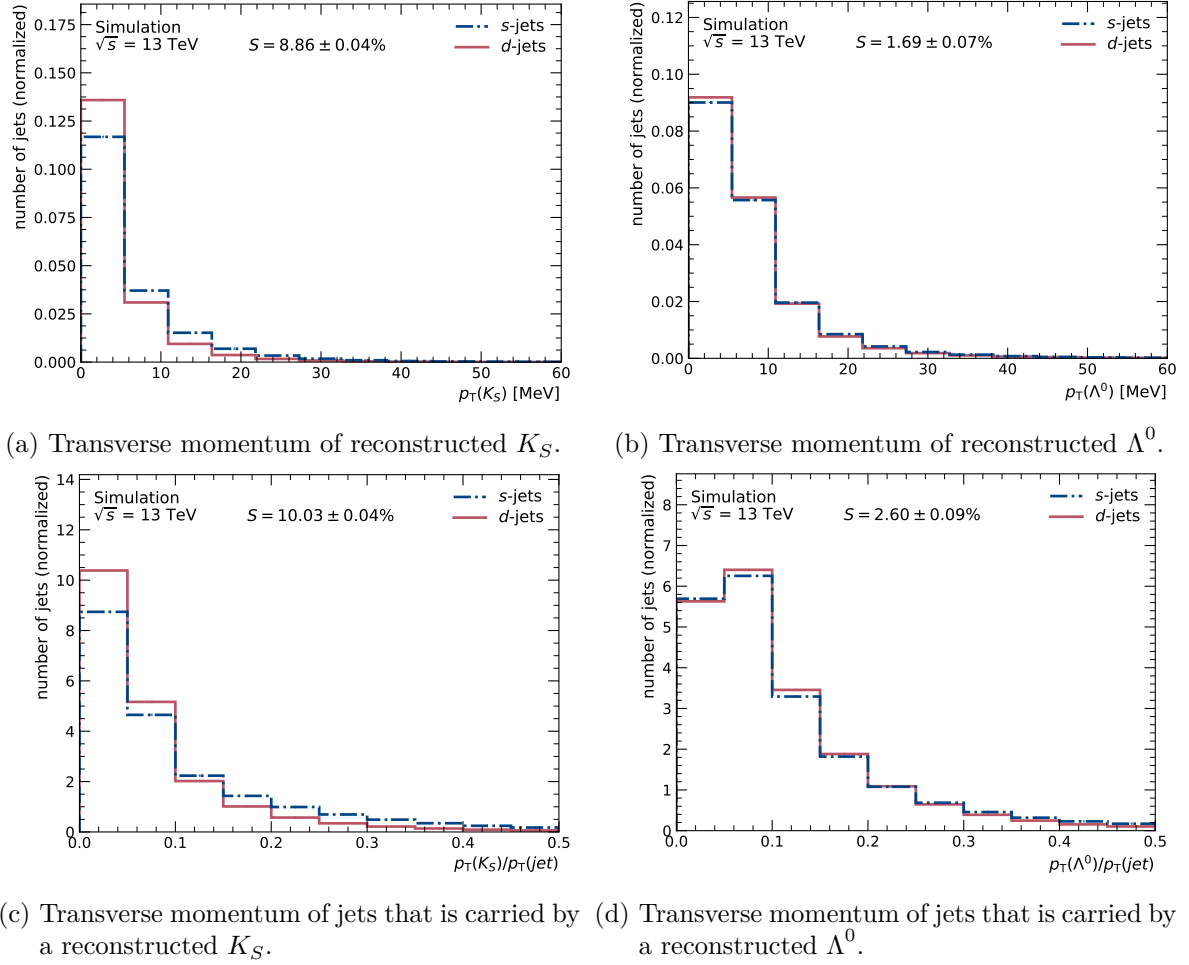
Fig. A.4: Distributions of the transverse momenta of the reconstructed strange hadrons matched to s- and d-jets and their relative contributions to the jet momenta for both s- and d-jets.

covered by the inner detector is larger.

Generally, s-jets are more central than d-jets if they contain a reconstructed strange hadron, most likely because they contain more strange hadrons from the isotropic hard scattering process and less from underlying processes can be observed in the more forward direction. The same trend can be seen in the p_T distributions of the s-and d-jets containing strange hadrons, where d-jets tend to have larger transverse momenta, cf. Figures A.6a and A.6b.

R_{xy}, the radial distance between the primary vertex and the reconstructed decay vertex of the K_S or Λ^0, is shown in Figures A.7a and A.7b. The maximal measured value R_{xy} is $500\,\mathrm{mm}$ and corresponds to the radius of the outermost layer of the ID, and the peak structure visible in the distribution corresponds to the layers of the ID. Λ^0 tend to decay farther away from the primary vertex than K_S due to their larger lifetime. For K_S, R_{xy} tends to be shifted to higher values for those K_S matched to s-jets since they typically have higher momenta. However, reconstructed Λ^0 in s-jets seem to be reconstructed slightly closer to the primary
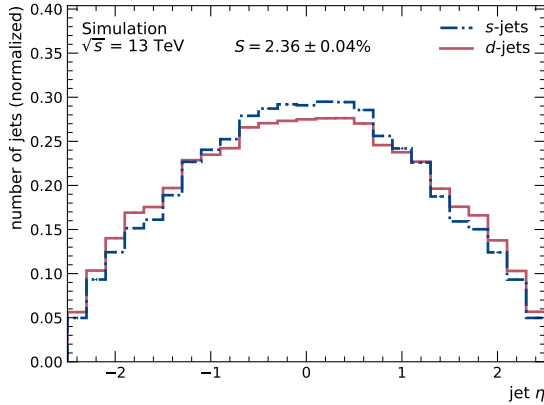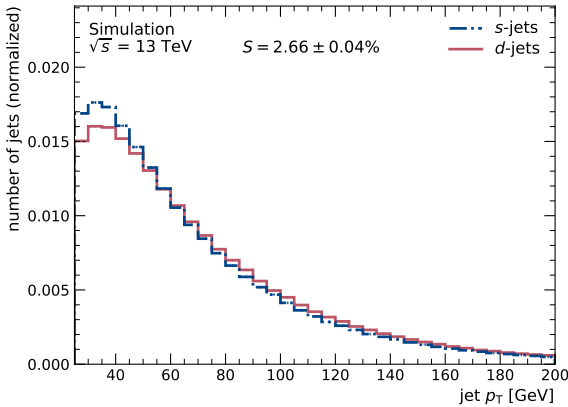
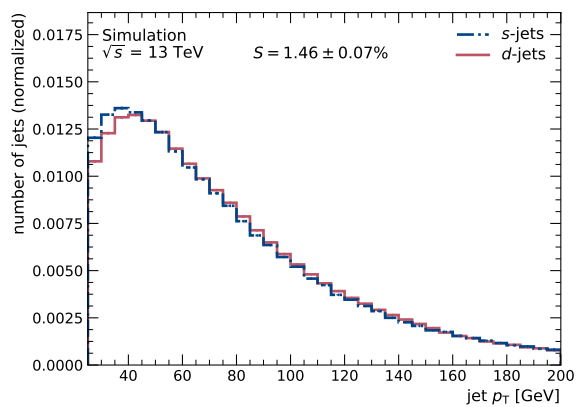(a) Pseudorapidity of s- and d-jets with a matched reconstructed K_S.
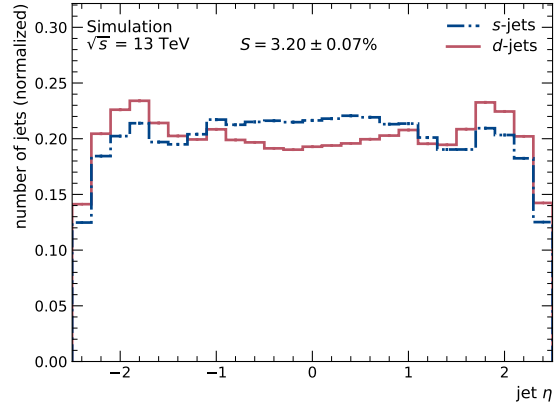
(b) Pseudorapidity of s- and d-jets with a matched reconstructed Λ^0.

Fig. A.5: Distributions of the pseudorapidity of s- and d-jets with a given reconstructed strange hadron matched to them.

(a) Transverse momentum of s-jets and d-jets which a matched K_S.

(b) Transverse momentum of s-jets and d-jets which a matched Λ^0.

Fig. A.6: Distributions of the transverse momentum of s-jets and d-jets with a given strange hadron matched to them.

vertex than those in d-jets, which is the result of the complex interplay between kinematics and reconstruction efficiencies in different regions of the detector.

The momentum of the reconstructed K_S or Λ^0 perpendicular to the jet axis, p_T^{rel}, as well as the angular distance between K_S or Λ^0 and the jet axis (ΔR) are shown in Figures A.8a–A.9b.

For both K_S and Λ^0, p_T^{rel} is larger for s-jets than for d-jets. One can reason that, if the original high-momentum quark is an s-quark, for hadronization to result in a K_S, one u quark is needed in addition to the s quark, while for Λ^0, a u- and a d-quark are needed. These additional quarks add momentum in a direction different to the jet axis and cause a larger ΔR. Since u and d quarks from the colliding protons can be both valence and sea quarks, they can add a large momentum relative to the jet direction. However, if the original high-momentum quark is a d quark, the s quarks necessary in production of a strange hadron cannot

(a) Radial distance of the decay vertex of K_S.

(b) Radial distance of the decay vertex of Λ^0.
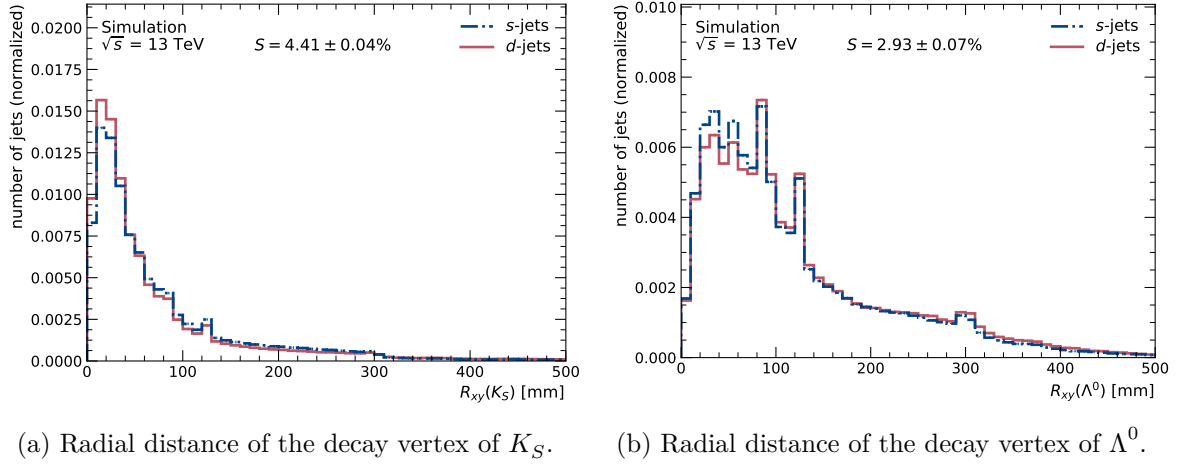
Fig. A.7: Radial distance between the primary vertex and the reconstructed decay vertex of the matched strange hadron for both s-jets and d-jets.

(a) p_T^{rel} of K_S.

(b) p_T^{rel} of Λ^0.
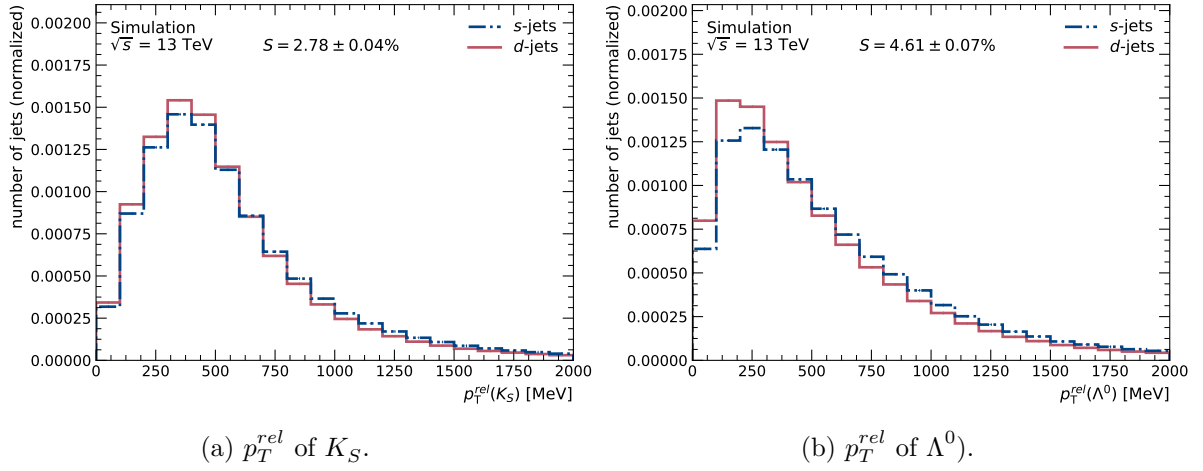
Fig. A.8: Momentum component of the reconstructed strange hadrons to the axis of the s-jet or d-jet they are matched to.

be a valence quark and therefore tends to add less momentum transverse to the jet axis. This explains the difference of the shapes of p_T^{rel} for s-and d-jets.

For K_S, the ΔR of s-jets is smaller than for d-jets, as it is expected for a particle which carries a large part of the jets' momentum. For Λ^0 matched to s-jets, however, ΔR is larger than for d-jets, most likely because here, the relative momentum transverse to the jet axis has a larger impact on ΔR.
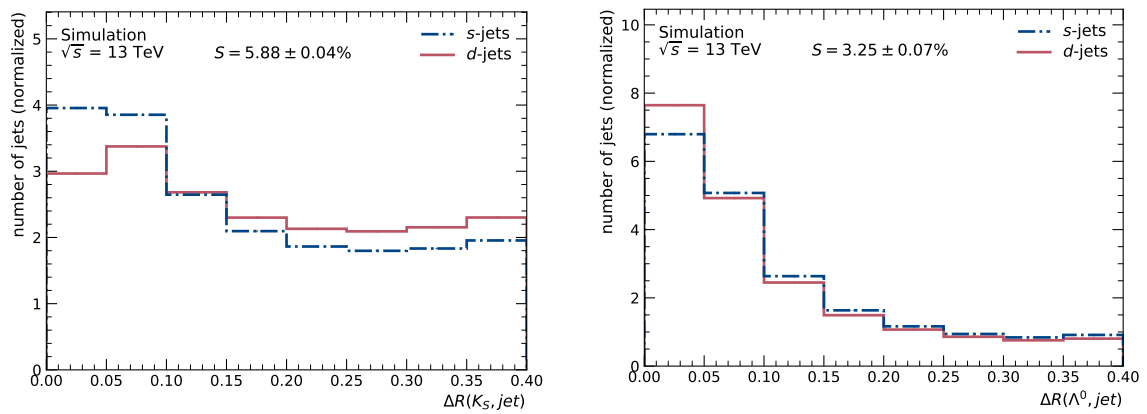
(a) ΔR between the reconstructed K_S and the jet
    axis.

(b) ΔR between the reconstructed Λ^0 and the jet
    axis.

Fig. A.9: ΔR between the reconstructed strange hadron and the axis of the s- or d-jets they
         are matched to.

APPENDIX B

Jet Attributes Used as *s*-Tagger Input Variables

This chapter of the appendix illustrates the jet attributes used to train the *s*-tagger for the ATLAS Experiment. They are based on either energies measured in the calorimeters or tracks associated to the jets and can be split into three categories: The energies a jet deposits in each calorimeter layer, the dimensions of the jets' shower in the detector (i.e. the geometrical moments), and their composition of electrically charged and neutral particles.
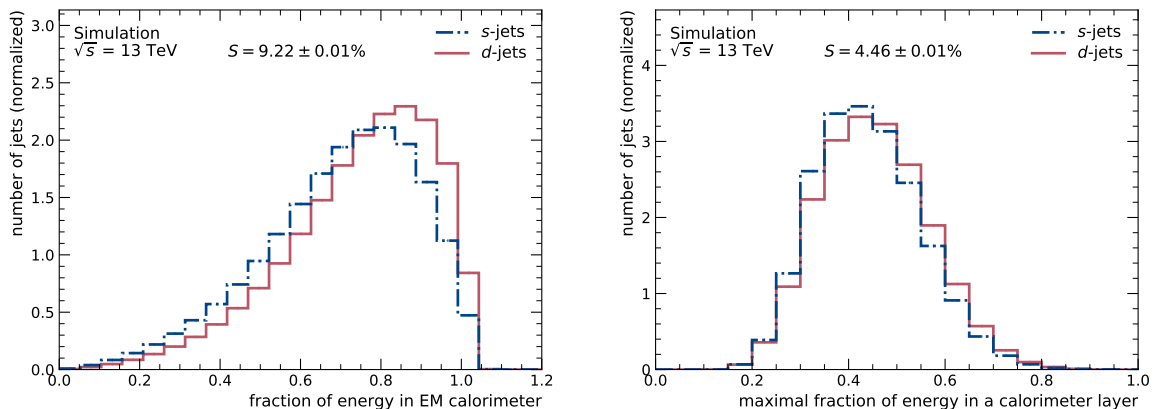
(a) Fraction of the total jet energy deposited in the EM calorimeter.

(b) Maximum fraction of the total energy deposited in a single calorimeter layer.

Fig. B.1: Distributions of jet attributes for *s*-jets and *d*-jets.

The *sampling fractions* are the energy per jet measured in one calorimeter layer divided by the total energy of the jet. All of them can be seen in Figures B.2 and B.3. Jets are only included in the plots if they could have deposited any energy in the calorimeter layer based on the direction of their jet axis in. Most sampling fraction distributions peak around 0.0 because there are only a few layers in which every jet deposits energy, such as the very first calorimeter layer called presampler. It is remarkable that small negative energy entries can occur. This is due to statistical fluctuations in the energy measurement the calibration cannot fully take into account. While some sampling fractions show larger separation power than others, all separations are larger than statistical fluctuations. The largest separation power given by sampling fractions can be found for jets in the barrel region, and mainly in the sampling fractions of the presampler, the layers of the EM calorimeter as well as the first two layers of the tile calorimeter. *d*-jets tend to deposit slightly more energy in the presampler and the first two layers of the EM calorimeter, while *s*-jets tend to deposit slightly more energy in the third layer of the EM calorimeter and the tile calorimeter.

The most discriminating variable overall is the energy fraction deposited in all layers of the EM calorimeter shown in Figure B.1a.

To provide the *s*-tagger with a full picture of the jets' showers in the calorimeter, all sampling fractions are included as inputs to the NN. Additionally, the maximal fraction of energy deposited in a single layer is used as an input variable. It is shown in Figure B.1b and tends to be smaller for *d*-jets.

Another group of input variables are the jet moments, described in detail in Reference [148]. They include the longitudinal distance between the cluster with the largest energy (leading cluster) in the jet and the jet's geometrical center $\lambda_{\text{leading cluster}}$ shown in Figure B.4a, the average of the square of the lateral distance between the jet axis and the jet's clusters $\langle r^2 \rangle$ shown in Figure B.5a, the average of the longitudinal distance between the jet's geometric center and the jet's clusters $\langle \lambda^2 \rangle$ shown in Figure B.4b, as well as the radial distance R between the primary vertex and the jet's centroid (i.e. barycenter weighted in E), shown in Figure B.5b. All of these quantities tend to be larger for *s*-jets than for *d*-jets since *s*-jets tend to deposit their energy further down the calorimeter (more energy in the hadronic calorimeter in comparison to *d*-jets) and therefore have a larger extension.

The final subset of input variables are related to the tracks associated with the jet via ghost matching [142]. It includes the number of tracks associated to the jet shown in Figure B.5c, the total p_T carried by these tracks shown in Figure B.5d, and the p_T-weighted charge of the jet calculated via

$$Q = \frac{\sum\limits_{i \in \text{tracks}} q_i \cdot p_{Ti}}{\sum\limits_{i \in \text{tracks}} p_{Ti}} \,. \tag{51}$$

All of these variables are larger (in their absolute value) for *d*-jets, which means that for *d*-jets, more momentum is carried by charged particles than for *s*-jets.
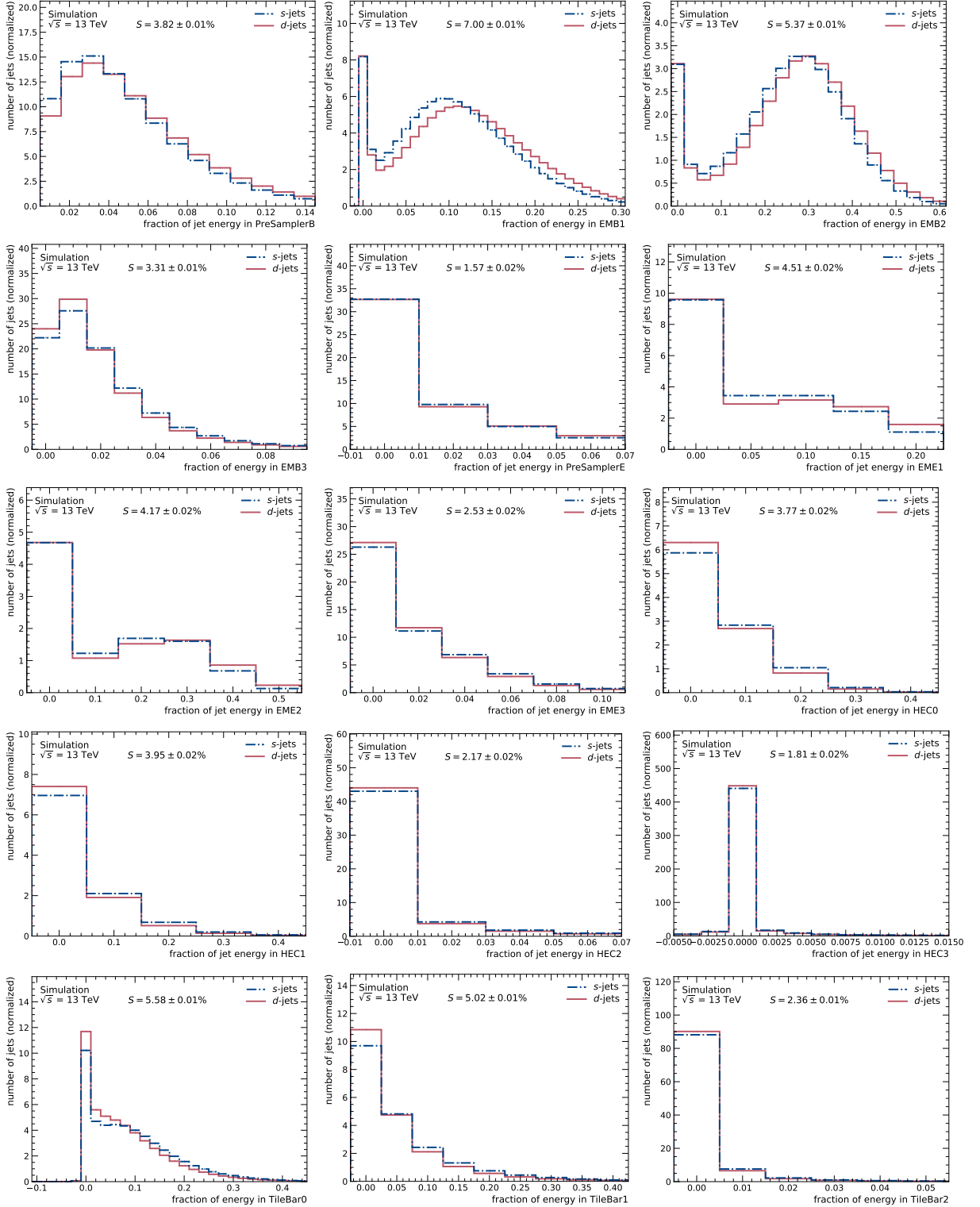
Fig. B.2: Distribution of the fraction of the total energy of the jet deposited in each calorimeter layer for both *s*- and *d*-jets. Only jets that geometrically touch the given calorimeter layers are included.
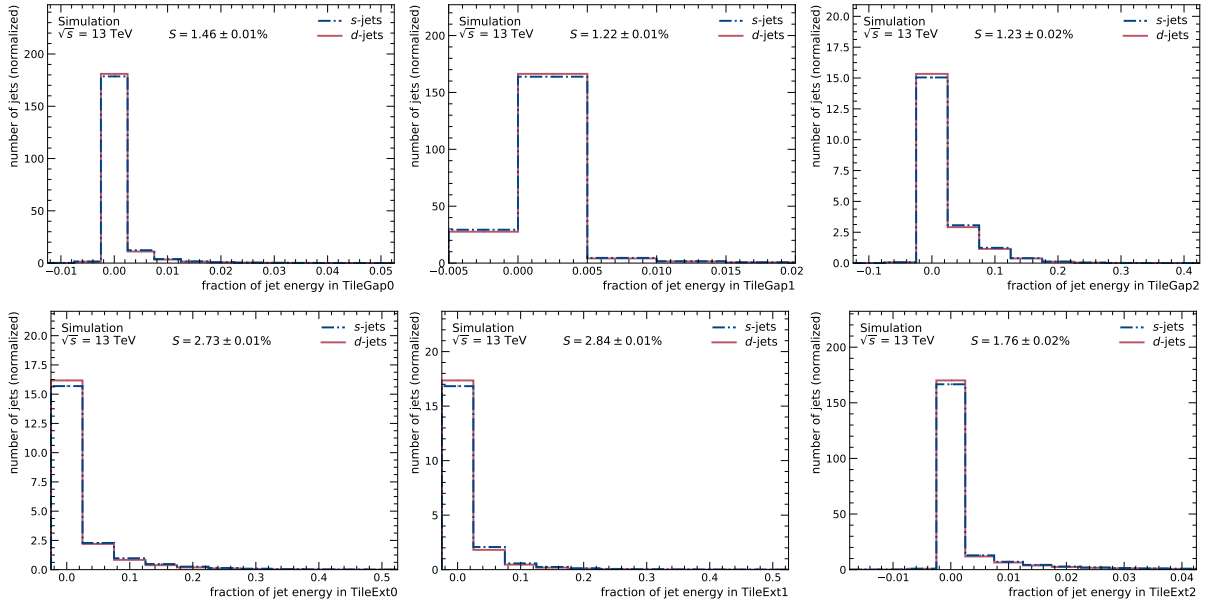
Fig. B.3: Distribution of the fraction of the total energy of the jet deposited in each calorimeter layer for both s- and d-jets. Only jets that geometrically touch the given calorimeter layers are included.
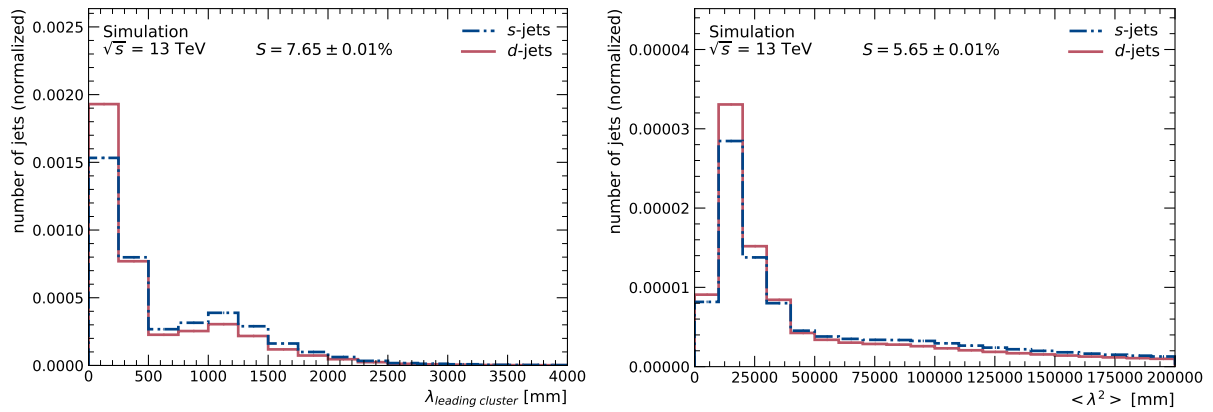
(a) Longitudinal distance between the geometrical center of the jet and the most energetic cluster of the jet.

(b) Average of the square of the longitudinal distance between clusters and the geometrical center of the jet.

Fig. B.4: Distributions of jet attributes for s-jets and d-jets.

133
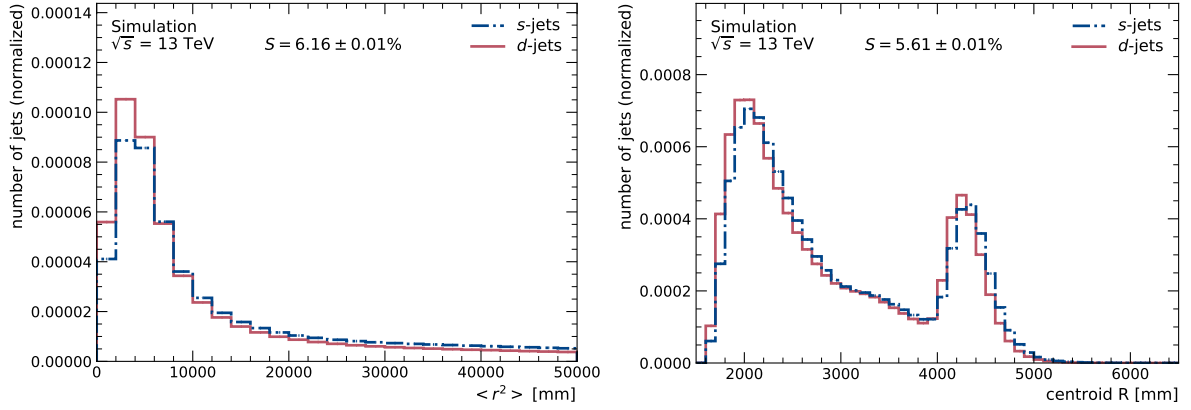
(a) Average of the square of the lateral distance between clusters and the jet axis.

(b) Distances between the primary vertex and the jets' centroids, i.e. their barycenters calculated from the clusters weighted by their energies.
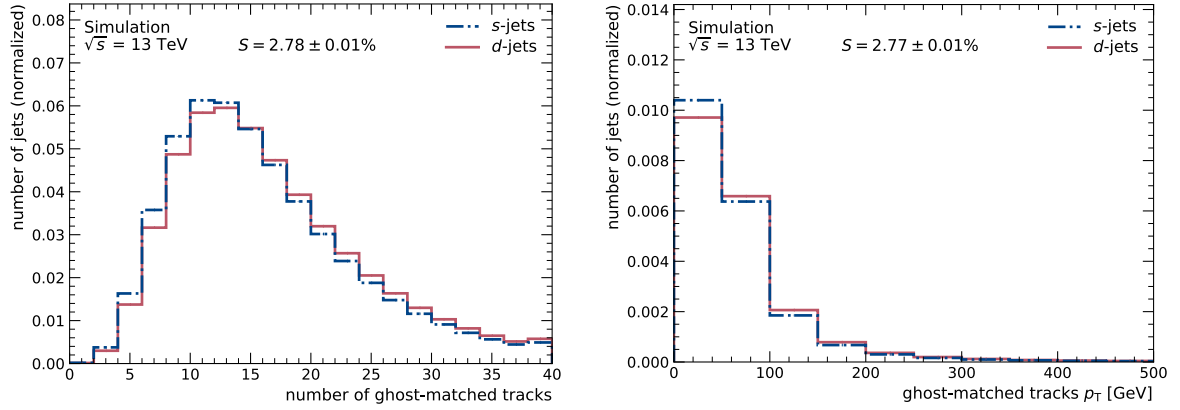
(c) Number of ghost-matched tracks.

(d) Sum of the transverse momenta carried by the ghost-matched tracks.
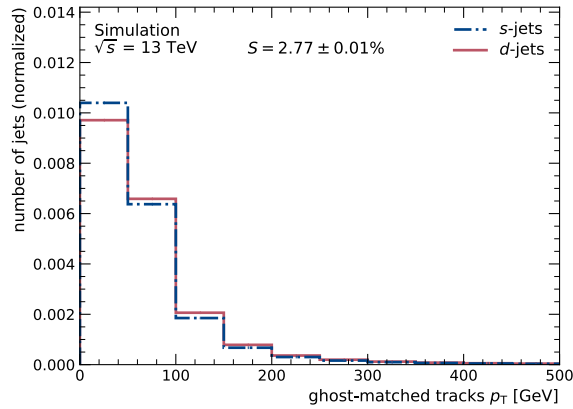
(e) Sum of the charges of all tracks weighted by the fraction of their transverse momentum.

Fig. B.5: Distributions of jet attributes for *s*-jets and *d*-jets.

Appendix C

s-Tagger Feature Selection

In this chapter, the option to remove features from the NN trained with both strange hadron features and jet attributes for an *s*-tagger at the ATLAS Detector is discussed.

The *s*-tagger setup that was found to be optimal in Section 6.3.2.3 is used as a baseline. To determine the contribution each variable has to the NN output, $L1$ regularization is applied in the input layer only. The regularization strength λ_{L1} is gradually turned up until the AUC score of the NN evaluated on the validation sample decreases. A regularization strength of $\lambda_1 = 7 \cdot 10^{-6}$ is the strongest regularization that does not result in a decrease of the AUC score. For a NN with $L1$ regularization of this strength, the connection weights between the input layer and the first hidden layer were evaluated. For this, the mean of the absolute value of all connecting weights between an input node and all nodes in the first hidden layer is calculated. Table C.1 shows the derived order of importance of the input features using these average weights.

To remove the influence of the less important variables in the NN on the output, one by one, the weights between an input node and the first hidden layer are set to zero, starting with those weights connected to the input variable with the smallest contribution to the neural network and then continuing towards those weights of the input nodes with larger contributions. Table C.2 shows the resulting accuracy, separation power, and AUC scores as evaluated on the validation sample. Both statistical uncertainty on the accuracy and AUC scores are derived by bootstrapping the validation sample. The uncertainty on the separation power is derived in a binned distribution of the NN output and assuming the number of jets in each bin to be Poisson distributed. It can be seen that only the sampling fraction in the EME3 calorimeter can be removed without impacting the performance of the NN. Since the full jet kinematics are supplied to the NN, it is able to reconstruct the sampling fraction of a single calorimeter layer from the sampling fractions in all other calorimeter layers.

In an additional step, the NN is retrained after removing variables from the input cumulatively, once again starting from the variable with the smallest contribution. Table C.3 shows the resulting values for the accuracy, separation power, and AUC score. Once again, only one variable can be removed without seeing an impact on the performance. All other variables contribute to the classification power of the *s*-tagger.

Variable	Weights	Relative Weights	Correlation to NN Output	
			Signal	Background
Sampling fraction in `EME3`	0.000019	$1.01 \cdot 10^{-5}$	0.068	0.105
$\Delta R(K_S \text{ or } \Lambda^0, \text{jet})$	0.000020	$1.06 \cdot 10^{-5}$	-0.127	-0.071
Sampling fraction in `HEC0`	0.000020	$1.10 \cdot 10^{-5}$	0.111	0.15
$p_T(\text{jet})$	0.000023	$1.23 \cdot 10^{-5}$	0.049	-0.012
Sampling fraction in `EMB3`	0.000025	$1.34 \cdot 10^{-5}$	0.163	0.217
jet contains K_S	0.000026	$1.42 \cdot 10^{-5}$	0.140	0.117
Sampling fraction in `EMB2`	0.000056	$3.02 \cdot 10^{-5}$	-0.233	-0.265
Sampling fraction in `TileGap1`	0.000068	$3.69 \cdot 10^{-5}$	0.073	0.081
Sampling fraction in `TileBar1`	0.000127	$6.87 \cdot 10^{-5}$	0.295	0.312
Sampling fraction in `TileExt1`	0.000209	$1.13 \cdot 10^{-4}$	0.158	0.17
$R_{xy}(K_S \text{ or } \Lambda^0)$	0.000861	$4.67 \cdot 10^{-4}$	0.084	0.028
Sampling fraction in `TileExt2`	0.001700	$9.23 \cdot 10^{-4}$	0.115	0.119
$\chi^2(K_S \text{ or } \Lambda^0)$	0.002706	$1.47 \cdot 10^{-3}$	-0.116	-0.081
$p_T(K_S \text{ or } \Lambda^0)/p_T(\text{jet})$	0.003614	$1.96 \cdot 10^{-3}$	0.131	0.084
jet contains Λ^0	0.003752	$2.03 \cdot 10^{-3}$	-0.039	-0.049
$m(K_S \text{ or } \Lambda^0)$	0.004184	$2.27 \cdot 10^{-3}$	0.057	0.026
Sampling fraction in `TileExt0`	0.004919	$2.67 \cdot 10^{-3}$	0.141	0.165
$p_T(K_S \text{ or } \Lambda^0)$	0.006863	$3.72 \cdot 10^{-3}$	0.192	0.099
Sampling fraction in `EME2`	0.007352	$3.99 \cdot 10^{-3}$	-0.078	-0.082
$p_T^{\text{rel}}(K_S \text{ or } \Lambda^0)$	0.007709	$4.18 \cdot 10^{-3}$	0.129	0.105
Sampling fraction in `HEC3`	0.011791	$6.40 \cdot 10^{-3}$	0.057	0.055
Sampling fraction in `HEC2`	0.012443	$6.76 \cdot 10^{-3}$	0.099	0.102
Sampling fraction in `HEC1`	0.013005	$7.06 \cdot 10^{-3}$	0.134	0.156
$\eta(\text{jet})$	0.013524	$7.34 \cdot 10^{-3}$	-0.006	0.001
Sampling fraction in `TileGap3`	0.021911	$1.19 \cdot 10^{-2}$	0.059	0.077
Sampling fraction in `EME1`	0.033682	$1.83 \cdot 10^{-2}$	-0.136	-0.135
Sampling fraction in `TileGap2`	0.037186	$2.02 \cdot 10^{-2}$	0.054	0.058
$\langle r^2 \rangle$	0.038149	$2.07 \cdot 10^{-2}$	0.305	0.363
centroid R	0.045042	$2.44 \cdot 10^{-2}$	0.166	0.174
Sampling fraction in `PreSamplerB`	0.048505	$2.63 \cdot 10^{-2}$	-0.261	-0.224
Sampling fraction in EM calorimeter	0.050015	$2.71 \cdot 10^{-2}$	-0.606	-0.647
Sampling fraction in `PreSamplerE`	0.058469	$3.17 \cdot 10^{-2}$	-0.09	-0.077
Maximal sampling fraction	0.069258	$3.76 \cdot 10^{-2}$	-0.252	-0.367
number of ghost-matched tracks	0.082644	$4.49 \cdot 10^{-2}$	-0.198	-0.163
Sampling fraction in `TileBar0`	0.102472	$5.56 \cdot 10^{-2}$	0.281	0.323
Sampling fraction in `EMB1`	0.109051	$5.92 \cdot 10^{-2}$	-0.368	-0.379
Sampling fraction in `TileBar2`	0.110464	$6.00 \cdot 10^{-2}$	0.141	0.141
$\lambda_{\text{leading cluster}}$	0.119074	$6.46 \cdot 10^{-2}$	0.432	0.464
ghost-matched tracks p_T	0.156317	$8.49 \cdot 10^{-2}$	0.000	0.000
$\langle \lambda^2 \rangle$	0.161973	$8.80 \cdot 10^{-2}$	0.280	0.326
charge	0.191738	$1.04 \cdot 10^{-1}$	0.000	0.000
$\tau(K_S \text{ or } \Lambda^0)$	0.309544	$1.68 \cdot 10^{-1}$	-0.019	-0.018

Tab. C.1: Input features of the neural network combining information about reconstructed K_S and Λ^0 as well as jet attributes, sorted by their importance (lowest to highest). The second column shows the mean of the absolute values of the weights connecting the input node of the variable named in the first column and all nodes in the first hidden layer. The third column shows the relative contribution these average absolute weights have w.r.t. all other variables. The fourth and fifth column quotes the correlation between the distribution of the input variables and the output distribution of the neural network both in signal and background.

Variable Removed (Cumulative)	Accuracy	Separation [%]	Validation AUC
none	0.5761 ± 0.0002	13.86 ± 0.02	0.6106 ± 0.0002
Sampling fraction in `EME3`	0.5763 ± 0.0002	13.90 ± 0.02	0.6106 ± 0.0002
$\Delta R(K_S \text{ or } \Lambda^0, \text{jet})$	0.5757 ± 0.0002	13.77 ± 0.02	0.6099 ± 0.0002
Sampling fraction in `HEC0`	0.5756 ± 0.0002	13.77 ± 0.02	0.6099 ± 0.0002
$p_T(\text{jet})$	0.5757 ± 0.0002	13.78 ± 0.02	0.6099 ± 0.0002
Sampling fraction in `EMB3`	0.5756 ± 0.0002	13.78 ± 0.02	0.6099 ± 0.0002
jet contains K_S	0.5756 ± 0.0002	13.77 ± 0.02	0.6099 ± 0.0002
Sampling fraction in `EMB2`	0.5755 ± 0.0002	13.76 ± 0.02	0.6097 ± 0.0002
Sampling fraction in `TileGap1`	0.5755 ± 0.0002	13.75 ± 0.02	0.6097 ± 0.0002
Sampling fraction in `TileBar1`	0.5753 ± 0.0002	13.73 ± 0.02	0.6095 ± 0.0002
Sampling fraction in `TileExt1`	0.5750 ± 0.0002	13.71 ± 0.02	0.6094 ± 0.0002
$R_{xy}(K_S \text{ or } \Lambda^0)$	0.5750 ± 0.0002	13.70 ± 0.02	0.6093 ± 0.0002
Sampling fraction in `TileExt2`	0.5737 ± 0.0002	13.60 ± 0.02	0.6084 ± 0.0002
$\chi^2(K_S \text{ or } \Lambda^0)$	0.5737 ± 0.0002	13.60 ± 0.02	0.6084 ± 0.0002
$p_T(K_S \text{ or } \Lambda^0)/p_T(\text{jet})$	0.5738 ± 0.0002	13.62 ± 0.02	0.6085 ± 0.0002
jet contains Λ^0	0.5736 ± 0.0002	13.58 ± 0.02	0.6082 ± 0.0002
$m(K_S \text{ or } \Lambda^0)$	0.5686 ± 0.0002	12.94 ± 0.02	0.6032 ± 0.0002
Sampling fraction in `TileExt0`	0.5637 ± 0.0002	12.28 ± 0.02	0.5978 ± 0.0002
$p_T(K_S \text{ or } \Lambda^0)$	0.5650 ± 0.0002	12.37 ± 0.02	0.5986 ± 0.0002
Sampling fraction in `EME2`	0.5642 ± 0.0002	12.34 ± 0.02	0.5984 ± 0.0002
$p_T^{\text{rel}}(K_S \text{ or } \Lambda^0)$	0.5629 ± 0.0002	11.55 ± 0.03	0.5932 ± 0.0002
Sampling fraction in `HEC3`	0.5611 ± 0.0002	11.47 ± 0.03	0.5925 ± 0.0002
Sampling fraction in `HEC2`	0.5577 ± 0.0002	11.22 ± 0.03	0.5902 ± 0.0002
Sampling fraction in `HEC1`	0.5585 ± 0.0002	11.32 ± 0.03	0.5911 ± 0.0002
$\eta(\text{jet})$	0.5595 ± 0.0002	12.19 ± 0.02	0.5964 ± 0.0002
Sampling fraction in `TileGap3`	0.5557 ± 0.0002	10.81 ± 0.02	0.5851 ± 0.0002
Sampling fraction in `EME1`	0.5470 ± 0.0002	10.63 ± 0.02	0.5839 ± 0.0002
Sampling fraction in `TileGap2`	0.5509 ± 0.0002	11.83 ± 0.02	0.5933 ± 0.0002
$\langle r^2 \rangle$	0.5509 ± 0.0002	11.69 ± 0.02	0.5921 ± 0.0002
centroid R	0.5487 ± 0.0002	11.64 ± 0.02	0.5916 ± 0.0002
Sampling fraction in `PreSamplerB`	0.5471 ± 0.0002	11.30 ± 0.02	0.5886 ± 0.0002
Sampling fraction in EM calorimeter	0.5467 ± 0.0002	10.86 ± 0.03	0.5858 ± 0.0002
Sampling fraction in `PreSamplerE`	0.5439 ± 0.0002	10.70 ± 0.03	0.5852 ± 0.0002
Maximal sampling fraction	0.5437 ± 0.0002	10.23 ± 0.03	0.5809 ± 0.0002
number of ghost-matched tracks	0.5395 ± 0.0002	10.19 ± 0.02	0.5792 ± 0.0002
Sampling fraction in `TileBar0`	0.5473 ± 0.0002	10.58 ± 0.03	0.5838 ± 0.0002
Sampling fraction in `EMB1`	0.5405 ± 0.0002	9.62 ± 0.03	0.5774 ± 0.0002
Sampling fraction in `TileBar2`	0.5348 ± 0.0002	9.37 ± 0.03	0.5751 ± 0.0002
$\lambda_{\text{leading cluster}}$	0.5371 ± 0.0002	8.17 ± 0.03	0.5669 ± 0.0002
ghost-matched tracks p_T	0.5269 ± 0.0002	6.95 ± 0.03	0.5570 ± 0.0002
$\langle \lambda^2 \rangle$	0.5298 ± 0.0002	5.55 ± 0.03	0.5508 ± 0.0002
charge	0.4994 ± 0.0002	0.14 ± 0.03	0.5014 ± 0.0002
$\tau(K_S \text{ or } \Lambda^0)$	0.5000 ± 0.0002	0.00 ± 0.03	0.5000 ± 0.0002

Tab. C.2: Accuracy, separation power and AUC scores of the s-tagger after setting the connection weights between the input nodes and the first hidden layer to zero. Setting the weights to zero is done cumulatively starting from the top of the list of variables. Both statistical uncertainty on the accuracy and AUC scores are derived by bootstrapping the validation sample. The uncertainty on the separation power is derived in a binned distribution of the NN output and assuming the number of jets in each bin to be Poisson distributed.

Variable Removed (Cumulative)	Accuracy	Separation [%]	Validation AUC
none	0.5761 ± 0.0002	13.82 ± 0.02	0.6106 ± 0.0002
Sampling fraction in `EME3`	0.5762 ± 0.0002	13.83 ± 0.02	0.6106 ± 0.0002
$\Delta R(K_S$ or $\Lambda^0, \text{jet})$	0.5761 ± 0.0002	13.81 ± 0.02	0.6100 ± 0.0002
Sampling fraction in `HEC0`	0.5761 ± 0.0002	13.80 ± 0.02	0.6100 ± 0.0002
$p_T(\text{jet})$	0.5733 ± 0.0002	13.12 ± 0.02	0.6048 ± 0.0002
Sampling fraction in `EMB3`	0.5732 ± 0.0002	13.09 ± 0.02	0.6045 ± 0.0002

Tab. C.3: Accuracy, separation power and AUC scores of the *s*-tagger after retraining without the variables in the given line and all variables in the lines above. Both statistical uncertainty on the accuracy and AUC scores are derived by bootstrapping the validation sample. The uncertainty on the separation power is derived in a binned distribution of the NN output and assuming the number of jets in each bin to be Poisson distributed.