

# Measurement of the inclusive and differential $t$ -channel single top quark production cross section at 13 TeV with the CMS experiment

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

von der KIT-Fakultät für Physik des  
Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

**M.Sc. David Seith**

aus Karlsruhe

Tag der mündlichen Prüfung:	18. Dezember 2020
Referent:	Prof. Dr. Thomas Müller
Korreferent:	Prof. Dr. Günter Quast





This document is licensed under a Creative Commons  
Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0):  
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>





---

## Erklärung der selbstständigen Anfertigung der Dissertationsschrift

Hiermit erkläre ich, dass ich die Dissertation mit dem Titel

*Measurement of the inclusive and differential  $t$ -channel single top quark production cross section at 13 TeV with the CMS experiment*

selbstständig angefertigt, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht habe, was aus Arbeiten anderer unverändert oder mit Abänderungen entnommen wurde.

Ich versichere außerdem, dass ich die Dissertation nur in diesem und keinem anderen Promotionsverfahren eingereicht habe und dass diesem Promotionsverfahren keine endgültig gescheiterten Promotionsverfahren vorausgegangen sind.

**Karlsruhe, den 16. November 2020**

.....

(David Seith)



# Introduction

The standard model of particle physics successfully describes all elementary particles and their interaction through the strong and electroweak forces. Despite its success it is not a complete theory of nature as there are phenomena, such as gravity or dark matter, that it cannot explain.

The heaviest known elementary particle is the top quark, which is about as heavy as a gold atom. The production of top quark pairs, which takes place via the strong interaction, was observed for the first time in 1995 at the Tevatron collider. The electroweak production of single top quarks is a much rarer process and it took until 2009 to also observe it at the Tevatron collider. While only about 40 thousand single top quarks were produced at the Tevatron collider during its total operation time, the Large Hadron Collider (LHC) produced roughly 40 millions single top quarks in the years from 2016 to 2018. This large number allows to study the properties of the top quark with unprecedented precision. One such property is the top quark polarization. Due to the V-A structure of the electroweak interaction single top quarks are always produced in a left-handed polarized state. Top quarks decay before they hadronize and therefore the top quark polarization can be measured from its decay products. The polarization of top quarks can be described by a spin density matrix and has three components: longitudinal, transverse and normal. The longitudinal polarization is largest and both the transverse and normal polarizations are expected to be close to zero in the SM.

This thesis measures the inclusive and differential production cross section of single top quarks and top antiquarks in the  $t$  channel. The measurement uses  $137\text{ fb}^{-1}$  of data recorded at the Compact Muon Solenoid (CMS) detector at the LHC from 2016 to 2018 in proton-proton collisions at a center-of-mass energy of 13 TeV. The ratio of the single top quark to top antiquark inclusive production cross section depends on the ratio of up-to-down-type quarks in the proton and is therefore sensitive to the light-quark parton distribution functions. The differential measurement is performed as a function of the top quark transverse momentum, rapidity, and three angular variables, which are sensitive to the three components of the top quark polarization. The angular variables are defined in the top quark rest frame between the charged lepton from the top quark decay and three axes that are defined based on the direction of the spectator quark and the beamline axis. From asymmetries in the distributions of these angular variables, top quark spin asymmetries are extracted, which are related to the top quark polarization. New physics beyond the standard model may change the V-A coupling structure of the electroweak interaction in the standard model. Depending on the specific model of new physics this could result in a decreased longitudinal and an increased transverse

---

or normal top quark polarization. The effect of such new physics can be described in an effective field theory approach by introducing coupling parameters. The magnitude of these coupling parameters is constrained from the measured angular distributions.

This thesis is structured as follows: A description of the standard model of particle physics is provided in Chapter 1. Chapter 2 gives an introduction to the phenomenology of the top quark, its properties, and production at the LHC. Statistical methods that are used in this thesis are described in Chapter 3. The LHC and the CMS experiment are introduced in Chapter 4. The simulation of events is described in Chapter 5 and the reconstruction and definition of physical objects from measured data in Chapter 6. Analyses studying top quarks rely on algorithms to identify jets originating from bottom quarks. This is called b tagging. A study on the performance of b tagging algorithms at the planned High-Luminosity-LHC is presented in Chapter 7. In Chapter 8 the measurement of the inclusive and differential cross section of single top quark and top antiquark production in the  $t$  channel is presented.

# Contents

<b>1</b>	<b>Standard model</b>	<b>1</b>
1.1	Fermions . . . . .	1
1.2	Bosons . . . . .	2
1.3	Quantum field theory . . . . .	4
1.4	Quantum chromodynamics . . . . .	5
1.5	Electroweak theory . . . . .	7
<b>2</b>	<b>Top quark</b>	<b>11</b>
2.1	Production . . . . .	11
2.1.1	Top quark pair production . . . . .	11
2.1.2	Single top quark production . . . . .	12
2.2	Decay . . . . .	16
2.3	Polarization . . . . .	16
2.4	Effective field theory . . . . .	20
<b>3</b>	<b>Statistical methods</b>	<b>25</b>
3.1	Binned maximum likelihood fit . . . . .	25
3.2	Machine learning . . . . .	26
3.2.1	Boosted Decision Trees . . . . .	28
3.2.2	Artificial neural networks . . . . .	29
3.3	Unfolding . . . . .	32
<b>4</b>	<b>The Compact Muon Solenoid detector at the Large Hadron Collider</b>	<b>35</b>
4.1	The Large Hadron Collider . . . . .	35
4.2	The Compact Muon Solenoid . . . . .	40
4.2.1	Tracker . . . . .	42
4.2.2	Electromagnetic calorimeter . . . . .	43
4.2.3	Hadronic calorimeter . . . . .	44
4.2.4	Solenoid . . . . .	46
4.2.5	Muon system . . . . .	46
4.2.6	Data acquisition, trigger and computing . . . . .	46
4.2.7	The CMS detector at the High Luminosity LHC . . . . .	48
<b>5</b>	<b>Event simulation</b>	<b>51</b>
5.1	Event generation . . . . .	51
5.1.1	Hard process . . . . .	54
5.1.2	Parton shower . . . . .	56
5.1.3	Hadronization . . . . .	57
5.1.4	Underlying event . . . . .	57

5.2	Detector simulation . . . . .	57
<b>6</b>	<b>Event reconstruction and object definition</b>	<b>59</b>
6.1	Tracks and vertices . . . . .	59
6.2	Calorimeter clusters . . . . .	60
6.3	Muon reconstruction and identification . . . . .	61
6.4	Electron reconstruction and identification . . . . .	62
6.5	Hadrons and photons reconstruction . . . . .	64
6.6	Jet reconstruction and identification . . . . .	65
6.6.1	Jet reconstruction . . . . .	65
6.6.2	Jet identification . . . . .	66
6.6.3	b jet identification . . . . .	66
6.6.4	Jet energy corrections . . . . .	69
6.6.5	Jet energy resolution corrections . . . . .	69
6.7	Missing transverse momentum . . . . .	70
<b>7</b>	<b>Performance of b tagging algorithms at the HL-LHC</b>	<b>73</b>
7.1	Training . . . . .	73
7.2	Performance . . . . .	76
7.3	Timing information . . . . .	80
7.4	Conclusion . . . . .	84
<b>8</b>	<b>Measurement of the <math>t</math>-channel single top quark production cross section</b>	<b>85</b>
8.1	Introduction . . . . .	85
8.2	Event topology . . . . .	86
8.2.1	Signal process . . . . .	86
8.2.2	Background processes . . . . .	88
8.3	Event selection . . . . .	89
8.4	Event simulation . . . . .	93
8.5	Corrections to simulated events . . . . .	93
8.5.1	Number of pileup interactions . . . . .	93
8.5.2	MET-phi modulation corrections . . . . .	93
8.5.3	b tagging efficiencies . . . . .	95
8.5.4	Lepton efficiencies . . . . .	98
8.5.5	Level-1 trigger ECAL prefiring . . . . .	100
8.6	QCD multijet background modeling . . . . .	101
8.7	Top quark reconstruction . . . . .	104
8.8	Event classification . . . . .	111
8.9	Parton and particle level objects . . . . .	120
8.10	Systematic uncertainties . . . . .	122
8.10.1	Experimental uncertainties . . . . .	123
8.10.2	Theory uncertainties . . . . .	125
8.11	Inclusive cross section measurement . . . . .	126
8.11.1	Signal extraction . . . . .	127
8.11.2	Results . . . . .	128
8.12	Differential cross section measurement . . . . .	132
8.12.1	Observables . . . . .	132

---

8.12.2	Unfolding . . . . .	135
8.12.3	Measured differential cross sections . . . . .	138
8.12.4	Top quark polarization . . . . .	143
8.13	Limits on EFT coefficients . . . . .	144
<b>9</b>	<b>Conclusion</b>	<b>147</b>
<b>A</b>	<b>Appendix</b>	<b>149</b>
A.1	List of samples . . . . .	149
A.2	Muon scale factors 2018 . . . . .	153
A.3	Modeling of the reconstructed top quark in the 2j1m0t region . . . . .	155
A.4	Modeling of BDT input and output variables in the control regions . . .	157
A.5	Particle level unfolding . . . . .	162
A.6	Systematic uncertainty contributions . . . . .	168





# 1 Standard model

The Standard model (SM) of particle physics describes all known elementary particles as well as their interactions via the electromagnetic, weak and strong force. Gravity is the only fundamental force that is not included. In the SM matter is made up of spin  $1/2$  particles called fermions. The forces are mediated via spin 1 particles called bosons. All particles that are predicted by the SM have been discovered. The last one was the Higgs boson, which was discovered in 2012 [1, 2].

## 1.1 Fermions

Fermions are spin  $1/2$  particles and thus follow the Pauli exclusion principle according to the spin-statistics theorem. There are twelve fermions in the standard model. Each fermion also has a corresponding antiparticle, which has the same properties but opposite quantum numbers. Fermions are grouped according to how they interact. There are six quarks, which carry color charge and interact via all interactions in the SM. Fermions that do not interact via the strong interaction are called leptons. There are three charged leptons, which interact via the electromagnetic and weak interaction. To each charged lepton there corresponds a neutral lepton, called neutrino, which interacts only via the weak interaction. Table 1.1 gives an overview on the fermions in the SM. Shown are their masses, electric charge, weak isospin and the forces they interact with. Neutrinos are massless in the SM. However, since the observation of neutrino oscillations [3, 4] it is known that they must have a mass but the exact value is unknown. Therefore, upper limits on neutrino masses are given.

Free fermions are described by the Dirac equation:

$$(i\gamma^\mu\partial_\mu - m)\psi = 0 \quad (1.1)$$

Here,  $\gamma^\mu$  are the gamma matrices,  $m$  is the mass and  $\psi$  the fermion field, which is a four component spinor. The fermion field  $\psi$  can be decomposed into a right-handed  $\psi_R$  and left-handed  $\psi_L$  chirality component:

$$\begin{aligned} \text{right handed :} \quad & \psi_R = \frac{(1 + \gamma_5)}{2}\psi \\ \text{left handed :} \quad & \psi_L = \frac{(1 - \gamma_5)}{2}\psi, \end{aligned}$$

where  $\gamma_5 = i\gamma_0\gamma_1\gamma_2\gamma_3$  is the fifth gamma matrix. The weak interaction couples only to the left-handed fields. The left-handed particles are thus grouped into isospin doublets, whereas right-handed particles form singlets.

### 1.2 Bosons

Bosons are particles that have an integer valued spin. The SM includes five bosons, which are listed in table 1.2. Interactions are mediated by exchanging spin 1 gauge bosons, which act as a carrier of the force. Each gauge boson couples with specific charges and mediates the interaction associated with that charge. The electromagnetic interaction is mediated via the massless photon which couples to the electric charge. The massive  $W^\pm$  and  $Z$  bosons mediate the weak interaction. The  $W^\pm$  bosons couple only to left-handed particles and right-handed antiparticles. They also carry an electric charge and therefore interact with photons. Additionally, the  $W^\pm$  bosons and the  $Z$  boson interact also with themselves. The strong interaction is mediated by the eight massless gluons which couple to color charge. As gluons carry color charge themselves, they interact with each other.

In addition to the gauge bosons, there is also the Higgs boson, which has spin 0. The Higgs boson arises as a consequence of electroweak symmetry breaking, as described in the Higgs mechanism [5, 6]. It is therefore important in order to explain why particles have mass.

**Table 1.1: The fermions of the SM.** There are six leptons and six quarks which are each arranged in three generations. Charged leptons are grouped together with neutrinos into a doublet. Neutrinos do not have an electric charge and thus interact only via the weak interaction (w). Charged leptons interact also via the electromagnetic interaction (e). Quarks carry color charge and thus interact via the strong interaction (s) in addition to electromagnetic and weak interaction. For each fermion the mass, the electric charge and the weak isospin are shown. For neutrino masses upper limits are given as the exact masses are unknown. The mass limits shown for the electron neutrino correspond to the mass limits on the electron neutrino  $m_{\nu_e}$  and in brackets on the electron antineutrino  $m_{\bar{\nu}_e}^2$ . The values are taken from reference [7].

		Mass	Electric charge / e	Weak isospin	Interactions
electron neutrino	$\nu_e$	$< 225 \text{ eV } (< 2 \text{ eV}^2)$	0	$+1/2$	w
electron	e	0.511 MeV	-1	$-1/2$	w, e
muon neutrino	$\nu_\mu$	$< 0.19 \text{ MeV}$	0	$+1/2$	w
muon	$\mu$	105.7 MeV	-1	$-1/2$	w, e
tau neutrino	$\nu_\tau$	$< 18.2 \text{ MeV}$	0	$+1/2$	w
tauon	$\tau$	1.78 GeV	-1	$-1/2$	w, e
down	d	5 MeV	$-1/3$	$+1/2$	w, e, s
up	u	2 MeV	$+2/3$	$-1/2$	w, e, s
strange	s	93 MeV	$-1/3$	$+1/2$	w, e, s
charm	c	1.3 GeV	$+2/3$	$-1/2$	w, e, s
bottom	b	4.2 GeV	$-1/3$	$+1/2$	w, e, s
top	t	173 GeV	$+2/3$	$-1/2$	w, e, s

**Table 1.2: The bosons of the SM.** Listed are the masses, electric charges and spin of the different bosons [7].

		Mass / GeV	Electric charge / e	Spin
photon	$\gamma$	0	0	1
W bosons	$W^\pm$	80.379	$\pm 1$	1
Z boson	Z	91.1876	0	1
gluons	g	0	0	1
Higgs boson	H	125.1	0	0

### 1.3 Quantum field theory

The SM is a quantum field theory, more specifically a non-abelian gauge theory. In quantum field theory particles are considered as excited states of fields. There is an underlying field for each matter and force-carrier particle. The propagation and interactions of the fields are described by different terms in the SM Lagrangian. All terms of the Lagrangian must be invariant under the Poincaré group in order to be independent of the reference frame. In a gauge theory interactions are introduced by requiring the Lagrangian to be additionally locally invariant under a local symmetry or gauge group. The symmetry group of the SM is the  $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ , where the  $SU(3)_C$  part refers to quantum chromodynamics (see section 1.4) and the  $SU(2)_L \otimes U(1)_Y$  part to the electroweak interaction (see section 1.5).

As stated in section 1.1 free fermions may be described by the Dirac equation. The Dirac equation can be derived from the following Lagrangian, via the Euler-Lagrange equations:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi. \quad (1.2)$$

This Lagrangian does not contain any interaction terms. In order to introduce interactions it is required that the Lagrangian is locally invariant under the transformations of the gauge group. The SM is based on Yang-Mills theory [8] and uses it to construct a locally gauge invariant Lagrangian. The Yang-Mills theory is a gauge theory which is based on special unitary groups  $SU(N)$  which have the following Lie algebra:

$$[T^a, T^b] = if^{abc}T^c. \quad (1.3)$$

Here  $T^a$  are the generators of the Lie algebra and the  $f^{abc}$  are their structure constants. Under the Lie algebra the fermion field transforms as:

$$\psi(x) \longrightarrow \psi(x)e^{-i\theta^a(x)T^a}, \quad (1.4)$$

where the  $\theta_a$  are real valued functions. The Lagrangian (1.2) is globally invariant under this transformation but not locally. In order to make the Lagrangian locally invariant under the symmetry group, the partial derivative in equation (1.2) is replaced by the covariant derivative:

$$D_\mu = \partial_\mu - igT^a A_\mu^a \quad (1.5)$$

Here  $A_\mu^a$  are newly introduced gauge fields corresponding to massless gauge bosons and  $g$  is the coupling strength between the fermion and the gauge fields. For each generator

of the Lie algebra one gauge field is introduced. The gauge fields transform under the gauge group as:

$$A_\mu^a(x) \longrightarrow A_\mu^a(x) - \frac{1}{g} \partial^\mu \theta_a(x) + f_{abc} \theta_b(x) A_\mu^c(x), \quad (1.6)$$

The propagator of the gauge fields is given by:

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu}^a F^{a\mu\nu}, \quad (1.7)$$

with the field strength tensors  $F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g f^{abc} A_\mu^b A_\nu^c$ . A general Lagrangian that is locally gauge invariant and describes both the fermion, the bosons, and their interactions, is then given by:

$$\begin{aligned} \mathcal{L} = & \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi && \text{fermion propagator} \\ & - g \bar{\psi} T^a \gamma^\mu A_\mu^a \psi && \text{interaction term} \\ & - \frac{1}{4} F_{\mu\nu}^a F^{a\mu\nu} && \text{boson propagator} \end{aligned} \quad (1.8)$$

The first term in equation (1.8) describes the free propagation of the fermion, the second term the interaction between the fermion and the gauge fields and the third term the propagation of the gauge fields. In case the theory is non-abelian (the structure constants are non-zero) the third term also introduces contributions proportional to  $A^3$  and  $A^4$ , thus describing interactions of the gauge fields with themselves. Adding an additional mass term for the gauge fields to equation 1.8 would break the gauge invariance. The bosons therefore have to be massless.

## 1.4 Quantum chromodynamics

Quantum Chromodynamics (QCD) is the theory that describes the strong interaction. The symmetry group of QCD is the  $SU(3)_C$ . It is a non-abelian group generated by the eight Gell-Mann matrices  $\lambda_a, a \in \{1..8\}$ . QCD therefore contains eight massless gauge bosons  $G_\mu^a$ , called gluons. The gluons couple to the three color charges. With the six different quark flavours, the QCD Lagrangian is given by:

$$\mathcal{L}_{QCD} = \sum_q \bar{\psi}_{q,i} (i\gamma^\mu (\partial_\mu \delta_{ij} - ig_s G_\mu^a T_{ij}^a) - m_q \delta_{ij}) \psi_{q,j} - \frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu} \quad (1.9)$$

Here,  $\psi_{q,i}$  denotes the quark-field for a quark with flavor  $q$ , mass  $m_q$ , and color  $i = r, g, b$ , and  $g_s$  is the strong coupling constant.

The strong coupling constant is not actually constant, but depends on the energy scale of the interaction. Its scaling as a function of the energy scale  $\mu$  is described by the beta function  $\beta(g)$ :

$$\frac{dg_s}{d \ln \mu^2} = \beta(g_s). \quad (1.10)$$

For QCD the one-loop beta function for  $n_f$  active quark flavours and  $N_C$  color charges is [9, 10]:

$$\beta(\alpha_S) = -\left(\frac{11N_C}{3} - \frac{2n_f}{3}\right) \cdot \frac{\alpha_S^2}{2\pi}, \text{ with } : \alpha_S = \frac{g_S^2}{4\pi}, \quad (1.11)$$

with  $n_f \leq 6$  and  $N_C = 3$  as in the SM, the beta function of QCD is negative and therefore, the coupling strength decreases with increasing energy scale. Thus, quarks behave as quasi-free particles at high energies or short distances. This effect is known as asymptotic freedom. At low energies the coupling strength increases. At a scale of about  $\Lambda \approx 200$  MeV, called the QCD scale, the coupling strength becomes approximately one. At this point perturbative calculations are no longer possible and one has to rely on phenomenological approaches.

The low-energy scaling, or long distance behavior of the strong coupling constant gives rise to the confinement property of QCD. The potential between two quarks can be described by a "Coulomb-like" term with an additional term, which increases linearly with the distance between the two quarks:

$$V(r) = -\frac{4\alpha_S}{3} \frac{1}{r} + k \cdot r. \quad (1.12)$$

When two quarks are separated the field strength between them increases, until the energy in the field is large enough to produce a new quark-antiquark pair. The initial quarks did not get isolated but are again bound with another quark. Therefore, quarks do not exist as free isolated particles but are always confined in color neutral composite particles, called hadrons. Hadrons typically consist of a quark and an antiquark (mesons) or of three quarks with different color (baryons). Also hadrons consisting of four or five quarks have been experimentally observed [11, 12]. Protons are stable particles, as baryon number is conserved in the SM and protons are the lightest baryons.

## 1.5 Electroweak theory

The electroweak theory describes the electromagnetic and weak interaction in one unified theory. Its symmetry group is the  $SU(2)_L \otimes U(1)_Y$ . The non-abelian  $SU(2)_L$  is generated by the three Pauli matrices  $\sigma_L^i$ , giving rise to three gauge boson fields  $W_\mu^{1,2,3}$ , which couple to the weak isospin and interact only with left-handed particles. The abelian  $U(1)_Y$  gives rise to one gauge boson  $B_\mu$ , which couples to the hypercharge. The hypercharge is connected with the electric charge  $Q$  and the third component of the weak isospin  $I_3$  via the Gell-Mann-Nishijima relation:  $Y = 2 \cdot (Q - I_3)$  [13–15].

The fields of the electromagnetic ( $\gamma$ ) and the weak ( $W^\pm$ ,  $Z$ ) interaction are obtained after the spontaneous breaking of the electroweak symmetry, which is discussed in more detail below. After symmetry breaking the  $W_\mu^{1,2,3}$  and  $B_\mu$  bosons mix with each other to create the  $\gamma$ ,  $W^\pm$  and  $Z$  [16,17]. The mixing is parametrized by the Weinberg angle [18]  $\sin \theta_W = \frac{g'}{\sqrt{g^2 + g'^2}} \approx 0.472$  [7]. The masses of  $W$  and  $Z$  bosons are then related to each other by the equation  $m_W/m_Z = \cos \theta_W$ .

The Lagrangian of the electroweak interaction before symmetry breaking is given by:

$$\mathcal{L}_{EW} = \sum_{\text{fermions}} \bar{\psi} \gamma^\mu (i\partial_\mu - g' \frac{1}{2} Y B_\mu - g \frac{1}{2} \sigma_L^i W_\mu^i) \psi - \frac{1}{4} W_a^{\mu\nu} W_{\mu\nu}^a - \frac{1}{4} B^{\mu\nu} B_{\mu\nu} \quad (1.13)$$

This Lagrangian does not contain a fermion mass term, as the  $W^i$  couple only to left-handed particles. Therefore, a mass term of the form  $m\bar{\psi}\psi = m(\bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R)$  would break the  $SU(2)_L \otimes U(1)_Y$  gauge invariance. A further problem is that the  $W^\pm$  and  $Z$  bosons of the weak interaction are massive. However, a mass term for the bosons would also break the gauge invariance of the Lagrangian.

Mass terms can be added to the Lagrangian by spontaneously breaking the gauge symmetry. Spontaneously breaking — in contrast to explicitly breaking — means that the Lagrangian of the theory remains invariant under the symmetry group. However, the symmetry of the vacuum - the ground state - on which the theory acts is broken by fixing a particular, non-zero value. In the SM the spontaneous breaking of the electroweak symmetry is described by the Higgs mechanism [5, 6]. It introduces a complex scalar field - the Higgs field - which transforms under the  $SU(2)$  as a doublet.

$$\varphi = \frac{1}{\sqrt{2}} \begin{pmatrix} \varphi_1^+ + i\varphi_2^+ \\ \varphi_1^0 + i\varphi_2^0 \end{pmatrix} \quad (1.14)$$

The Lagrangian of this field is given by:

$$\mathcal{L}_H = (D_\mu \varphi)^\dagger (D^\mu \varphi) - V(\varphi), \quad (1.15)$$

$$V(\varphi) = -\mu^2 \varphi^\dagger \varphi + \lambda (\varphi^\dagger \varphi)^2, \quad (1.16)$$

$$D_\mu = \partial_\mu - ig\sigma^i W_\mu^i - ig' \frac{Y}{2} B_\mu, \quad (1.17)$$

Here,  $D_\mu$  is the covariant derivative, which introduces an interaction between the Higgs field and the electroweak gauge bosons.  $V(\varphi)$  is the Higgs potential. For  $-\mu^2 \geq 0$  the minimum of the potential is at 0 and the potential follows a parabolic shape. The symmetry of the  $SU(2)_L \otimes U(1)_Y$  is spontaneously broken if  $-\mu^2 < 0$ . The Higgs field then takes on a ground state within the new minimum. The choice of the ground state is arbitrary, as in principle any value in the minimum can be chosen. The underlying physics does however not depend on the exact choice of the ground state, meaning the ground state is now degenerated. As the Higgs field has four components, the contour of the minimum visually corresponds to the surface of a 4D sphere, which can be parametrized by three angles. Therefore, there are three free parameters in the choice of the position of the minimum. These degrees of freedom correspond to Goldstone bosons, which arise according to the Goldstone theorem in theories with spontaneously broken symmetries [19, 20]. As the ground state can be freely chosen within the minimum, this corresponds to a gauge freedom. By choosing the unitarity gauge, the minimum is effectively set at  $\langle \varphi_0 \rangle = \frac{1}{\sqrt{2}}(0, \nu)^T$ , with the vacuum expectation value  $\nu = 246$  GeV, and all Goldstone bosons are removed [21, 22]. The corresponding degrees of freedom are absorbed by the  $W^\pm$  and  $Z$  bosons, which acquire mass. The original Higgs field has four components of which three got absorbed, the remaining one gives rise to a massive scalar particle, called the Higgs boson.

In order to introduce a fermion mass term, a Yukawa coupling between the Higgs field  $\varphi$  and the fermions is added. For leptons this is given by:

$$\mathcal{L}_{YW} = -y_f \bar{\Psi}_L \varphi \psi_R + \text{h.c.} \quad (1.18)$$

Where  $\Psi_L$  is a doublet and  $\psi_R$  a singlet. If  $\varphi$  is replaced by its vacuum expectation value, then a fermion mass term is obtained with the mass  $m_f = \frac{y_f \nu}{\sqrt{2}}$ . Masses for quarks are also generated via the Yukawa interaction. However, separate terms for up- and down-type quarks are necessary as when in equation (1.18) the Higgs field is replaced by its vacuum expectation value  $\langle \varphi_0 \rangle = \frac{1}{\sqrt{2}}(0, \nu)^T$  only the isospin-down partner acquires a mass because the up-partner is multiplied with zero. Therefore, this equation is only suitable to explain the masses of charged leptons and down-type quarks. In order to give masses also to up-type quarks, the Higgs field has to be "rotated", which is achieved by using  $\tilde{\varphi} = i\sigma_2 \varphi^*$  and introducing a second term for up-type quarks. The Yukawa interaction between the Higgs field and quarks can be written as:

$$\mathcal{L}_{YQ} = -Y^d \bar{Q}_L \varphi d_R - Y^u \bar{Q}_L \tilde{\varphi} u_R + \text{h.c.}, \quad (1.19)$$



here  $Y^{u,d}$  are complex  $3 \times 3$  matrices,  $Q_L$  are left-handed quark doublets and  $u_R$  and  $d_R$  are right-handed up- and down-type quark singlets. The mass eigenstates of the quarks are obtained by diagonalizing  $Y^{u,d}$  using four unitary matrices  $V_{L,R}^{u,d}$ :

$$M^f = \frac{\nu}{\sqrt{2}} \cdot V_L^f Y^f V_R^{f\dagger}, \quad (1.20)$$

here  $M^f$  are diagonal matrices which have the quark masses as eigenvalues. This procedure introduces a mixing between different quark flavors which alters the coupling strength between quarks and the  $W^\pm$  bosons. This mixing is described by the Cabibbo–Kobayashi–Maskawa (CKM) matrix [17, 23, 24].

$$V_{\text{CKM}} = V_L^u V_L^{d,\dagger} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \quad (1.21)$$

The standard parametrization of the CKM matrix uses three angles and one phase. However, the structure of the CKM matrix is more obvious in the Wolfenstein parametrization [25], which is an approximation of the CKM matrix. It uses four parameters  $\lambda$ ,  $A$ ,  $\rho$  and  $\eta$ . The best measured values for these parameters are [7]:

$$\lambda = 0.2240 \pm 0.0004, \quad A = 0.836 \pm 0.015 \quad (1.22)$$

$$\rho = 0.122 \pm 0.018, \quad \eta = 0.355 \pm 0.012 \quad (1.23)$$

The Wolfenstein parametrization to order  $\lambda^3$  is given by:

$$V_{\text{CKM}} = \begin{pmatrix} 1 - \frac{1}{2}\lambda^2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{1}{2}\lambda^2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + O(\lambda^4) \quad (1.24)$$

The values on the diagonal are approximately one meaning that quarks from the same generation couple more strongly than quarks from different generations. Coupling between different generations are suppressed by powers of  $\lambda$ . The complex term  $\rho - i\eta$  gives rise to CP violation, which means that physical processes are not the same if particles are replaced by their antiparticles (C or charge conjugation) and the coordinate system is inverted or mirrored (P or parity transformation). According to the Sakharov conditions CP violation is necessary to explain the abundance of matter over antimatter in our universe [26]. However, the strength of CP violation measured so far is not sufficient to completely explain the observed abundance.



## 2 Top quark

The top quark is the heaviest particle in the SM with a mass of 173 GeV [7]. It was discovered in 1995 at the Tevatron collider [27, 28]. The top quark is the only quark that decays into a real W boson and a bottom quark. The decay width of the top quark is 1.42 GeV [7], corresponding to a lifetime of only about  $5 \cdot 10^{-25}$  s, which is shorter than the hadronization time of about  $10^{-24}$  s. Therefore, top quarks are the only quarks which decay before they hadronize, which allows to measure the top quark polarization from its decay products.

### 2.1 Production

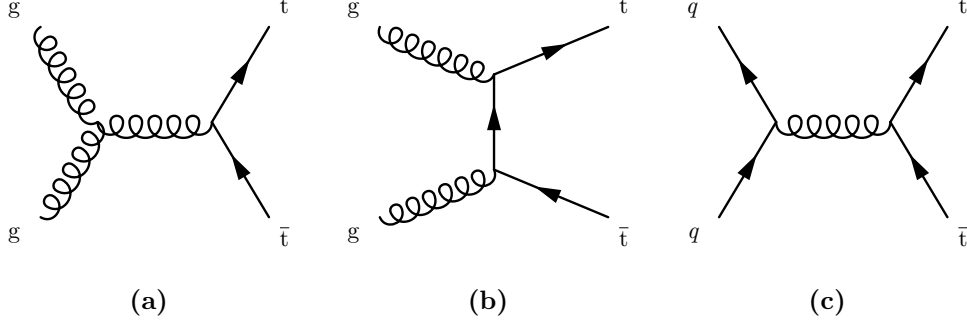
At the LHC the dominant production mode for top quarks is top quark pair production ( $t\bar{t}$ ), which takes place via the strong interaction. The different production modes for top quark pair production are depicted in 2.1. Because gluons make up most of the parton density functions of the proton at TeV scale, the main contribution to the total  $t\bar{t}$  cross section comes from gluon-gluon fusion  $gg \rightarrow t\bar{t}$ , contributing about 90 % to the total cross section. Quark-antiquark annihilation  $q\bar{q} \rightarrow t\bar{t}$  contributes about 10 %.

#### 2.1.1 Top quark pair production

At a center-of-mass energy of  $\sqrt{s} = 13$  TeV the predicted cross section in proton-proton collisions is [29]:

$$\sigma_{t\bar{t}} = 831.76^{+19.77}_{-29.20} (\text{scale}) \pm 35.06 (\text{PDF} + \alpha_S) \text{ pb} \quad (2.1)$$

This cross section was calculated with the Top++2.0 program to next-to-next-to-leading order in perturbative QCD, including soft-gluon resummation to next-to-next-to-leading-log order [30] and assuming a top quark mass of 172.5 GeV. The scale uncertainty comes from the independent variation of the factorisation and renormalisation scales  $\mu_F$  and  $\mu_R$ . The PDF and  $\alpha_S$  uncertainties are obtained following PDF4LHC prescription using the following PDF sets: MSTW2008 68 % CL NNLO, CT10 NNLO and NNPDF2.3 5f FFN [31–34].



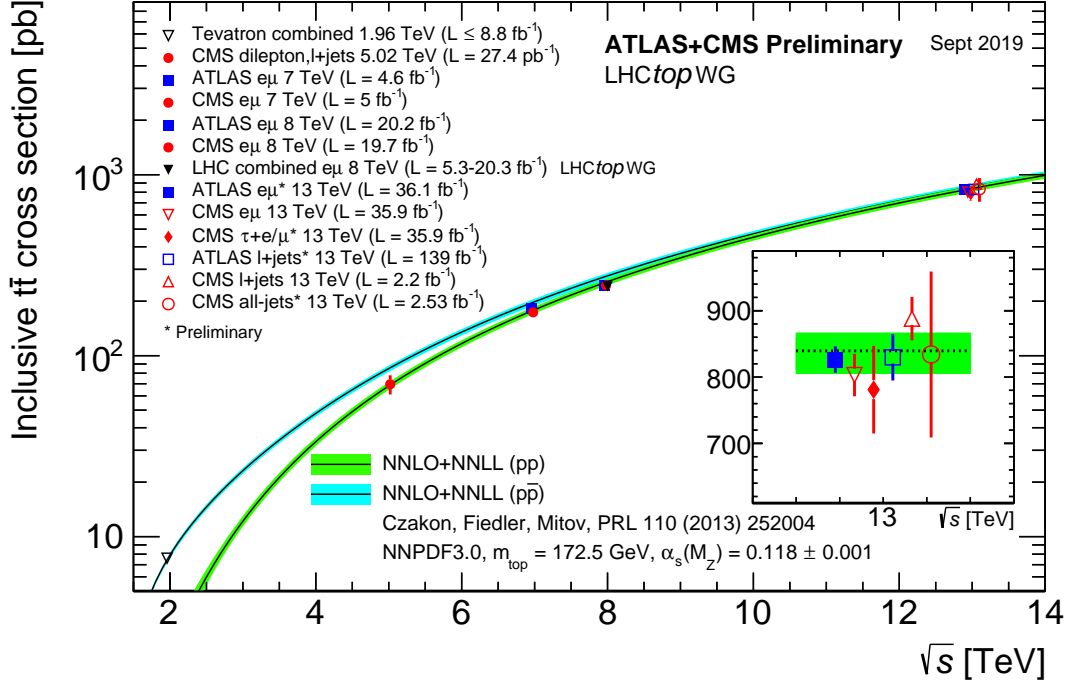
**Figure 2.1: The three dominant production modes for top quark pair production at the LHC.** The largest contribution comes from gluon-gluon fusion (a) and (b). Quark-antiquark annihilation (c) contributes only about 10 % to the total production cross section.

The  $t\bar{t}$  cross section has been measured at the Tevatron at  $\sqrt{s} = 1.96$  TeV in  $p\bar{p}$  collisions and at the LHC at  $\sqrt{s} = 7, 8$  and 13 TeV in  $pp$  collisions. Figure 2.2 shows a comparison between the measured and predicted  $t\bar{t}$  cross sections at different center-of-mass energies.

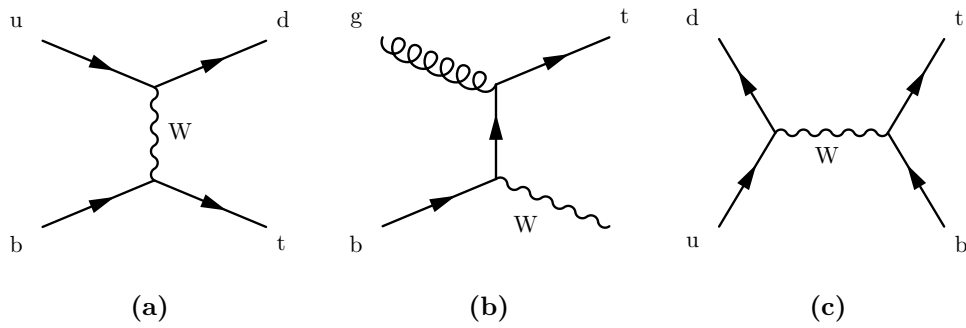
### 2.1.2 Single top quark production

The production of single top quarks takes place via the electroweak interaction and has therefore a smaller cross section compared to top quark pair production at the LHC. There are three production modes for single top quarks at LO, which can be categorized according to the four-momentum  $q_W$  of the involved W boson. These three production modes are depicted in Figure 2.3.

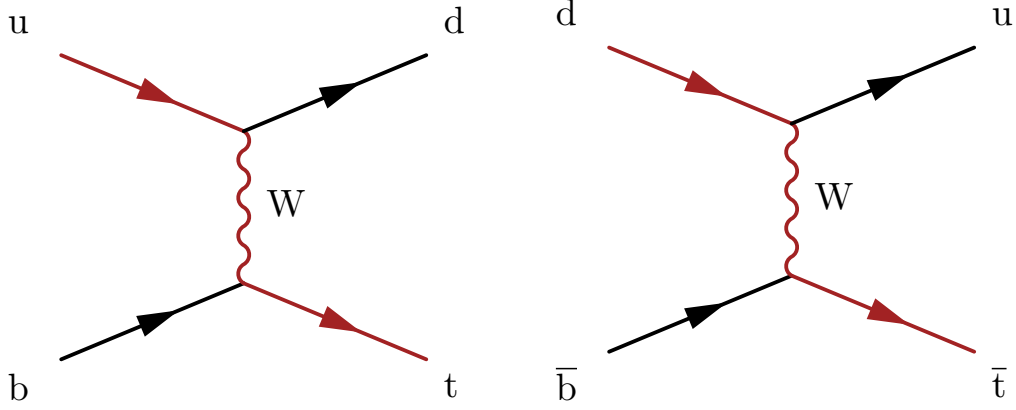
**$t$ -channel production** The  $t$ -channel production mode (figure 2.3a) is characterized by a space-like W boson ( $q_W^2 < 0$ ). It has the largest cross section of the three production modes. Figure 2.4 shows the production of both single top quarks and top antiquarks in the  $t$ -channel. Top quarks are produced if the initial light quark in the proton is an up-type quark and top antiquarks if the initial light quark is a down-type quark. The cross section for producing top quarks in the  $t$ -channel is about 1.7 times larger than the cross section for producing top antiquarks, because there are more up-type than down-type quarks in the proton. A consequence of this is that the cross section ratio  $R = \sigma_{t\text{-ch}}^t / \sigma_{t\text{-ch}}^{\bar{t}}$  is sensitive to the light quark PDF of the proton. A characteristic feature of the  $t$ -channel production mode is the spectator quark, which originates from one of the interacting protons and scatters against the top quark by exchanging a W boson. In the scattering process its direction is only changed slightly and it is therefore predominantly scattered in very forward direction.



**Figure 2.2: Summary of measurements of top quark pair production cross section at the LHC and Tevatron** The measured values are shown as a function of the center-of-mass energy and are compared with theory predictions. The figure is taken from reference [35].



**Figure 2.3: The three dominant production modes for single top quark production at the LHC.** The production of single-top quarks in the t-channel (a) has the largest cross section. The second largest cross section has the associated production with a W boson (b). The smallest contribution stems from the s-channel production mode (c).



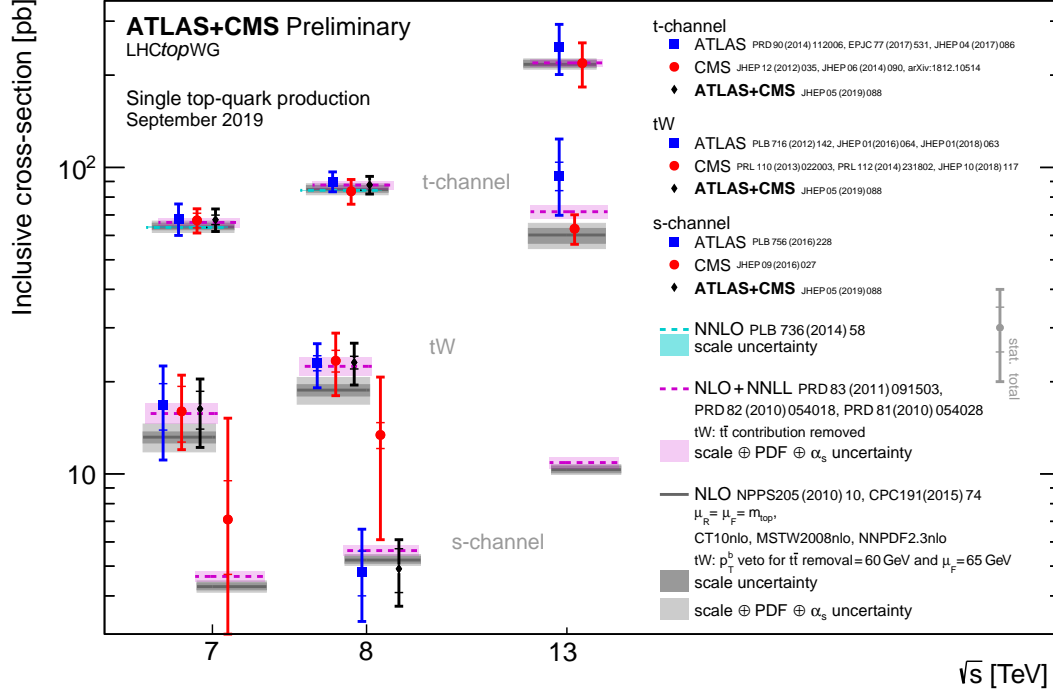
**Figure 2.4: Feynman diagrams for single top quark and top anti-quark production.** The left figure shows the production of single top quarks and the right figure the production of single top antiquarks in the  $t$  channel. In single top quark production the initial light quark is an up type quark and in single top antiquark the initial light quark is a down type quark.

**associated production with a W boson** The second production mode for single-top quarks is the associated production with a W boson, the  $tW$  channel (figure 2.3b). This production mode is characterized by an on-shell W boson ( $q_W^2 = m_W^2$ ).

**s-channel production** The third production mode with the smallest cross section is the  $s$ -channel. It features an time-like W boson ( $q_W^2 > 0$ ) and is illustrated in figure 2.3c.

In single top quark production a difference is made between the four flavor scheme (4FS) and the five flavor scheme (5Fs). In the 4FS bottom quarks are not considered as part of the proton, whereas in the 5FS bottom quarks are part of the proton. Therefore, in the 4FS the bottom quark that interacts with the W boson originates from a gluon splitting. The predicted cross section for the three production channels at the LHC at  $\sqrt{s} = 13$  TeV are shown below. For the  $t$ -channel process, the predicted cross sections for producing a top quark  $\sigma_{t\text{-ch}}^t$  and for producing a top antiquark  $\sigma_{t\text{-ch}}^{\bar{t}}$ , as well as the combined cross section  $\sigma_{t\text{-ch}}^{t+\bar{t}}$  are shown separately. The values are taken from reference [36].

$$\begin{aligned}\sigma_{t\text{-ch}}^{t+\bar{t}} &= 216.99_{-4.64}^{6.62} (\text{scale}) \pm 6.16 (\text{PDF} + \alpha_S) \text{ pb} \\ \sigma_{t\text{-ch}}^t &= 136.02_{-2.92}^{4.09} (\text{scale}) \pm 3.52 (\text{PDF} + \alpha_S) \text{ pb} \\ \sigma_{t\text{-ch}}^{\bar{t}} &= 80.95_{-1.71}^{2.53} (\text{scale}) \pm 3.18 (\text{PDF} + \alpha_S) \text{ pb} \\ \sigma_{tW} &= 71.7 \pm 1.80 (\text{scale}) \pm 3.4 (\text{PDF}) \text{ pb} \\ \sigma_{s\text{-ch}} &= 10.32_{-0.24}^{0.29} (\text{scale}) \pm 0.27 (\text{PDF} + \alpha_S) \text{ pb}\end{aligned}$$



**Figure 2.5: Summary of measurements of the cross section of the different single top quark production channels at the LHC.** The measured cross sections for the three single top quark production modes, the  $t$ -channel, the associated production with a W boson and the  $s$ -channel, are shown as a function of the center-of-mass energy. The measured values are compared with theory predictions based on NLO QCD, NLO QCD complemented with NNLL resummation and NNLO QCD ( $t$ -channel only) [39].

The predicted cross sections have been calculated at next-to-leading order (NLO) in QCD with Hathor v2.1 [37, 38] assuming a top quark mass of 172.5 GeV. The PDF and  $\alpha_s$  uncertainties are calculated using the PDF4LHC prescription [31] using the following PDF sets: MSTW2008 68% CL NNLO, CT10 NNLO and NNPDF2.3 5f FFN [31–34]. In Figure 2.5 the cross section for single-top production, which have been measured at the LHC and Tevatron colliders, are shown for different center-of-mass energies and compared to theory predictions.

Measurements of the single top quark cross section provide an independent way of measuring the CKM matrix element  $V_{tb}$ . Assuming  $|V_{tb}| \gg |V_{td}|$  and  $|V_{tb}| \gg |V_{ts}|$ , the cross section is proportional to  $|f_{LV} V_{tb}|^2$ . The factor  $f_{LV}$  describes possible new-physics contributions. The matrix element is given by:

$$|f_{LV} V_{tb}| = \sqrt{\frac{\sigma_{\text{measured}}}{\sigma_{\text{predicted}}}}. \quad (2.2)$$

## 2.2 Decay

With a branching fraction of more than 99 % [7] top quarks decay almost exclusively into a bottom quark and a W boson. The decays of the top quark are therefore classified according to the decay products of the W boson. The W boson decays in 67.6 % of all cases hadronically into a quark and an antiquark of different types and in 32.4 % leptonically into a charged lepton and a neutrino [7]. In case of top quark pair production the decays are called full-hadronic if both W bosons decay hadronically, semileptonic if exactly one decays hadronically and dileptonic if both decay leptonically.

## 2.3 Polarization

The production time of top quarks is about  $1/m_t \approx 10^{-27}$  s, with  $m_t = 173$  GeV, which is shorter than its lifetime of  $1/\Gamma_t \approx 10^{-25}$  s, which is again shorter than the hadronization time of about  $1/\Lambda_{\text{QCD}} \approx 10^{-24}$  s ( $\Lambda_{\text{QCD}} = 200$  MeV), which is shorter than the spin decorrelation time of about  $m_t/\Lambda_{\text{QCD}}^2 \approx 10^{-21}$  s [40]. Therefore, the helicity of top quarks in their production is kept in their decay and the spin direction of top quarks is correlated with the spin direction of their decay products. In single-top production, top quarks are produced via the electroweak production, coupling to a W boson. Due to the V-A structure of the weak interaction, these top quarks are produced in a left-handed state.

In a reference frame with axis  $(x, y, z)$  where the top quark is at rest, an ensemble of polarized top quarks can be described by a density matrix:

$$\rho = \begin{pmatrix} 1 + P_z & P_x - iP_y \\ P_x + iP_y & 1 - P_z \end{pmatrix} \quad (2.3)$$

Here  $P_i = 2\langle S_i \rangle$  is the top quark polarization along the direction  $i = (x, y, z)$  [41]. Typically, the  $z$  direction is chosen along the direction in which the top quark spin is predominantly oriented, this is also called the spin axis  $\vec{s}$ . Therefore, the  $P_z$  polarization is the largest polarization and is also the one commonly referred to as the top quark polarization. In this thesis it is referred to as longitudinal polarization. The other components are called transverse polarization ( $P_x$ ) and normal polarization ( $P_y$ ).

The polarization  $P_i$  can be measured from an asymmetry  $A_{i,X}$  in the angular distributions of the top quarks decay products. The angle  $\theta_X^i$  is calculated in between the momentum direction  $\vec{p}_X$  of the decay product  $X$  and an appropriately chosen axis  $\vec{a}_i$ :

$$\cos \theta_X^i = \frac{\vec{a}_i \cdot \vec{p}_X}{|\vec{a}_i| \cdot |\vec{p}_X|} \quad (2.4)$$



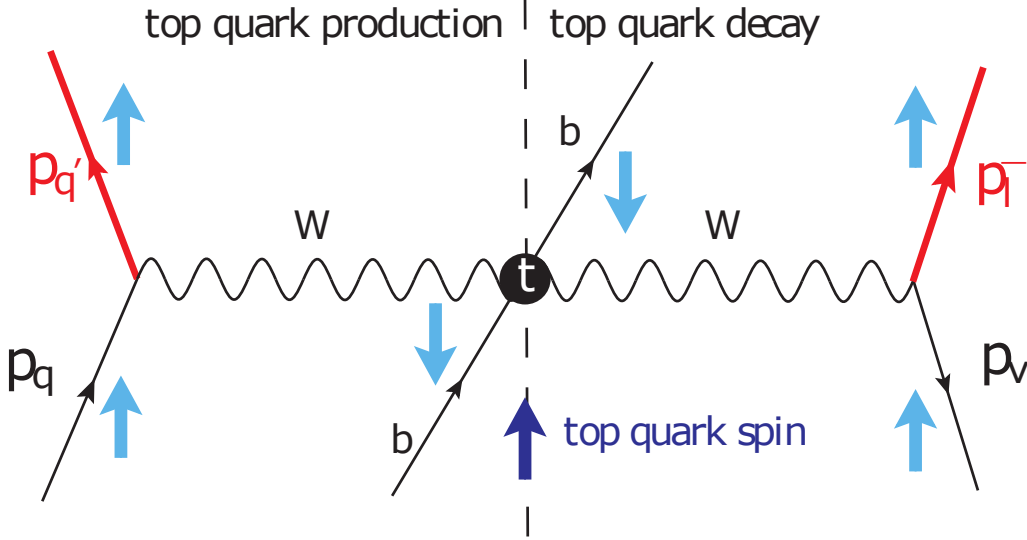
**Table 2.1:** Spin analyzing powers at NLO for different decay products of the top quark [42–46].

decay product	spin analyzing power
charged lepton	0.998
neutrino	-0.33
down-type quark	0.93
up-type quark	-0.31
bottom quark	-0.39
W boson	0.39

$$\frac{1}{\sigma} \frac{d\sigma}{d \cos \theta_X^i} = \frac{1}{2} (1 + 2A_{i,X} \cdot \cos \theta_X^i), \quad \text{with : } A_{i,X} = \frac{1}{2} \alpha_X P_i \quad (2.5)$$

Here,  $\alpha_X$  is the spin-analyzing power of the decay product  $X$  of the top quark. The spin-analyzing power describes how strongly the spin of the decay product is correlated to the spin of the top quark. The spin analyzing powers for the different decay products of the top quark are listed in table 2.1. In the SM it is  $\approx 1$  for the charged lepton for a leptonically decaying top quark.

In order to measure the polarization, an appropriate coordinate system needs to be chosen. The most important part here is to find the direction along which the top quark polarization is maximal, i.e. the spin axis. There are different possibilities for defining the spin axis that are going to be discussed in the following. In the so-called helicity basis, the spin axis is defined in the center-of-mass (COM) frame. It is then taken along the direction of the top quark momentum in this frame. The problem with the helicity basis is that the COM becomes ill-defined beyond LO, due to additional radiations [47]. In Figure 2.6 the single-top quark production and decay are illustrated with the spin directions of the involved particles. A characteristic feature of single top  $t$ -channel production is the spectator quark, which recoils against the top quark. At LO the outgoing spectator quark is a down-type quark in case of top quark production and an up-type quark in case of top antiquark production. In case of top antiquark production the outgoing spectator quark originated from an initial state down-type quark inside the proton. This means that at LO either the initial or final state of the spectator quark is always a down-type quark. The spin directions of both the charged lepton and the down-type quark are maximally correlated with the spin direction of the top quark. Thus, choosing the spin axis along the direction of the down-type quark results in the largest polarization [48]. As the down-type quarks lies either inside the proton or follows its direction one could therefore choose the direction of the beam axis as the spin axis. This is referred to as beamline basis. Which beam line to choose can be decided on an event-by-event basis, based on the direction of the outgoing spectator quark [48, 49]. For top quarks the outgoing spectator quark aligns with the direction of the down-type quark in 80 % of the cases [50]. It is therefore natural to instead choose the spin axis along the direction of the outgoing spectator quark. This choice of the spin axis results in the so-called spectator quark basis. It should be noted that for

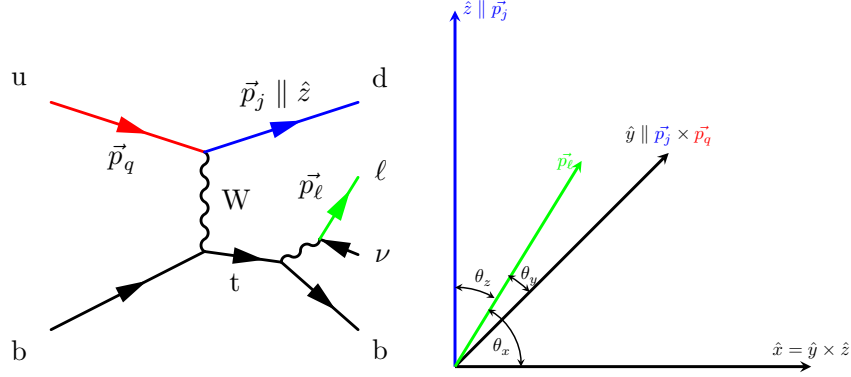


**Figure 2.6: Illustration of the correlation between the spin of the particles involved in the single top quark production.** The black circle represents the top quark and the arrows the spin orientations of the particles. The red colored lines denote the spectator quark  $p_{q'}$  and the charged lepton  $p_\ell$ . As can be seen the spins of the spectator quark and the charged lepton are both aligned with the spin of the top quark. The figure is taken from reference [47].

top antiquarks the outgoing spectator quark is not identical to the down-type quark in 69% of the cases, but the down-type quark is the initial quark inside the proton. In the spectator quark basis therefore, in principle the wrong direction is chosen for top antiquarks. However, as the outgoing spectator quark mostly follows the direction of the initial quark, this choice still results in the correct spin axis most of the time. This has the effect that the polarization for top antiquarks is lower than for top quarks in the spectator basis [49]. This is also true in reverse for the beamline basis. As the down-type quark lies inside the proton for top antiquarks, a higher polarization for top antiquarks is obtained in the beamline basis. The spectator basis results in an overall larger predicted polarization than the beamline basis [47, 49].

A right-handed coordinate system can be constructed by choosing the  $\hat{z}$  direction along the direction  $\vec{p}_j$  of the outgoing spectator quark and the  $\hat{y}$  direction orthogonal to the spectator quark direction and the direction of the initial quark inside the proton  $\vec{p}_q$ , where  $\vec{p}_{j,q}$  are both defined in the top quark rest frame. The  $\hat{x}$  direction is obtained by requiring a right-handed coordinate system [41]. The coordinate system is illustrated in figure 2.7.

$$\hat{z} = \frac{\vec{p}_j}{|\vec{p}_j|}, \quad \hat{y} = \frac{\vec{p}_j \times \vec{p}_q}{|\vec{p}_j \times \vec{p}_q|}, \quad \hat{x} = \hat{y} \times \hat{z}. \quad (2.6)$$



**Figure 2.7: Illustration of the coordinate system for the top quark polarization.** The left figure shows the Feynman diagram for single top quark production in the  $t$  channel with a leptonically decaying top quark. Depicted are also the momentum vectors of the initial state quark, the spectator quark, and the lepton from the top quark decay. The right figure shows the coordinate system that is defined based on the momentum direction of the initial state quark and the spectator quark in the top quark rest frame. The three polarization angles  $\theta_{x,y,z}$  are defined between the coordinate axes and the direction of the charged lepton in the top quark rest frame. In the measurement, the momentum direction of the initial light quark is deduced from the direction of the spectator quark.

The direction of the initial quark cannot be accessed experimentally because it lies within one of the two colliding protons inside the beamline. In order to obtain the direction of the initial light quark an assumption is made. The spectator quark normally follows the direction of the initial light quark. Therefore, the direction of the initial light quark can be chosen as the direction of the spectator quark in the laboratory frame projected on the beamline axis, which means the direction is given by:

$$\hat{p}_q^{\text{Lab}} = \left( 0, 0, \frac{\eta_j^{\text{Lab}}}{|\eta_j^{\text{Lab}}|} \right)^T. \quad (2.7)$$

This choice results in the correct direction for the initial light quark in 97 % of the cases for  $ug \rightarrow dt\bar{b}$  and in 98 % of the cases for  $dg \rightarrow u\bar{t}b$  [41], which are the main production channels for single top quarks and antiquarks respectively. Due to the direction not always being correct, the measured polarizations  $\bar{P}_{x,y}$  are expected to be about 10 % lower, than the actual polarizations  $P_{x,y}$  [41]. At 13 TeV the following values are given for the top quark polarization using the NLO generator POWHEG [51–53]:

$$\begin{aligned} \mathbf{P}(t) &= (-0.02, 0, 0.86) \\ \mathbf{P}(\bar{t}) &= (-0.08, 0, -0.83) \end{aligned}$$

$P_x$  is larger for top antiquarks than for top quarks because the spectator quark is not the down-type quark for top antiquarks in the dominant production channels.

## 2.4 Effective field theory

In an effective field theory (EFT) the effects of physics occurring at higher energy scales are described in a model independent way without explicitly giving the cause for the effects. For this purpose the SM Lagrangian is expanded with additional terms, which are suppressed by powers of an energy scale  $\Lambda$ :

$$\mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{SM}} + \frac{1}{\Lambda} \mathcal{L}_1 + \frac{1}{\Lambda^2} \mathcal{L}_2 + \mathcal{O}\left(\frac{1}{\Lambda^3}\right). \quad (2.8)$$

In principle one averages over degrees of freedom that become relevant at higher energy scales. In this way it is not necessary to specify a certain physical model. The additional term in the Lagrangian can be described by dimensionless coupling parameters  $C_i^{(d)}$ , called Wilson-coefficient and operator  $\mathcal{O}_i^{(d)}$ , which describe the kind of interactions:

$$\mathcal{L}_{d-4} = \sum_i C_i^{(d)} \mathcal{O}_i^{(d)}, \quad (2.9)$$

Here  $d$  denotes the dimension of the operator. The SM consists of dimension two and four operators. There is only one dimension-five operator which is compatible with the gauge group of the SM. This operator violates the lepton number conservation and is therefore not considered here [54]. Therefore, the leading contributions to the effective field theory Lagrangian stem from dimension-six operators. Under the assumption of minimal flavor violation and Baryon number conservation there are a total of 59 dimension-six operators, which can be built from SM fields in the Warsaw Basis [54]. The contributions of the EFT operators to the single top quark production are pictorially illustrated in figure 2.8. The effective operators can either modify the Wtb vertex (figure 2.8a and 2.8c) or create a four fermion contact interaction, which introduces an additional production mode for single top quarks (figure 2.8b). The operators modifying the Wtb vertex are listed in the following. The notation follows that in reference [55]. Flavor indices are labeled by  $ijkl$ , left-handed quark doublets are called  $q$ ; right-handed quark singlets  $u, d$ . The Pauli-matrices are denoted by  $\tau^I$ . The Higgs doublet is denoted by  $\varphi$ ;  $(\varphi^\dagger i \overleftrightarrow{D}_\mu \varphi) = \varphi^\dagger (i D_\mu \varphi) - (i D_\mu \varphi^\dagger) \varphi$ ,  $(\varphi^\dagger i \overleftrightarrow{D}_\mu \varphi)^I = \varphi^\dagger \tau^I (i D_\mu \varphi) - (i D_\mu \varphi^\dagger) \tau^I \varphi$  and  $T^A = \lambda^A/2$  with the Gell-Mann matrices  $\lambda^A$ :

$$O_{\varphi q}^{(3,ij)} = i(\varphi^\dagger i \overleftrightarrow{D}_\mu^I \varphi)(\bar{q}_i \gamma^\mu \tau^I q_j) \quad (2.10)$$

$$O_{tW} = (\bar{q} \sigma^{\mu\nu} \tau^I t) \tilde{\varphi} W_{\mu\nu}^I \quad (2.11)$$

$$O_{\varphi ud}^{(ij)} = i(\varphi^\dagger i \overleftrightarrow{D}_\mu \varphi)(\bar{u}_i \gamma^\mu d_j) \quad (2.12)$$

$$O_{bW} = (\bar{q} \sigma^{\mu\nu} \tau^I b) \tilde{\varphi} W_{\mu\nu}^I \quad (2.13)$$

The Wtb term of the Lagrangian can be written as [56]:

$$\begin{aligned}\mathcal{L}_{Wtb} = & \frac{g}{\sqrt{2}} W^-_\mu \bar{b} \gamma^\mu ((1 + f_1^L) P_L + f_1^R P_R) t \\ & - \frac{g}{\sqrt{2} m_W} \partial_\nu W^-_\mu \bar{b} \sigma^{\mu\nu} (f_2^L P_L + f_2^R P_R) t + \text{h.c.}\end{aligned}$$

with the operators  $P_{L,R} = \frac{1 \pm \gamma_5}{2}$ , which project the fields on their right-handed or left-handed state and the form factors  $f_i^{L,R}$  modifying the coupling structure. They are related to the Wilson-coefficients by the following relations [57]:

$$\begin{aligned}f_1^L &= \frac{C_{\varphi q}^{(3,ij)} \nu^2}{\Lambda^2}, & f_1^R &= \frac{1}{2} C_{\varphi ud}^{(ij)} \frac{\nu^2}{\Lambda^2}, \\ f_2^L &= \sqrt{2} C_{bW} \frac{\nu^2}{\Lambda^2}, & f_2^R &= \sqrt{2} C_{tW} \frac{\nu^2}{\Lambda^2}.\end{aligned}$$

$\nu^2$  is the vacuum expectation value. The coefficients  $C_{tW}$ ,  $C_{bW}$  and  $C_{\varphi ud}^{(ij)}$  are complex valued and can therefore introduce a CP-violating phase. Of the four operators contributing to the  $Wtb$  vertex, the most relevant ones are  $C_{tW}$  and  $C_{\varphi q}^{(3,ij)}$ . The operators  $C_{bW}$  and  $C_{\varphi ud}^{(ij)}$  always appear in the matrix element in terms that are proportional to the mass of the bottom quark  $m_b$ . Therefore, their contributions are suppressed at least by  $\mathcal{O}(m_b)$ . The bottom quark mass is much smaller than any relevant energy scales at the LHC, which means these operators can be neglected [58].

There are a total of 16 four-quark operators contributing to single top quark production. They can be separated into whether they involve four left-handed quarks  $\bar{L}\bar{L}\bar{L}\bar{L}$ , four right-handed quarks  $\bar{R}\bar{R}\bar{R}\bar{R}$ , or a mixture of the form  $\bar{L}\bar{L}\bar{R}\bar{R}$ ,  $\bar{L}\bar{R}\bar{R}\bar{L}$  or  $\bar{L}\bar{R}\bar{L}\bar{R}$  [54, 55]. The dominant contribution of dimension-six operators comes from interferences with the SM Lagrangian as other contributions are at least  $1/\Lambda$  smaller. As the weak interaction involves only left-handed fields there is only a small interference with terms which involve right-handed quarks [59, 60]. Thus, the dominant contributions stem from the  $\bar{L}\bar{L}\bar{L}\bar{L}$  operators, which are:

$$O_{qq}^{(1,ijkl)} = (\bar{q}_i \gamma^\mu q_j) (\bar{q}_k \gamma_\mu q_l) \quad (2.14)$$

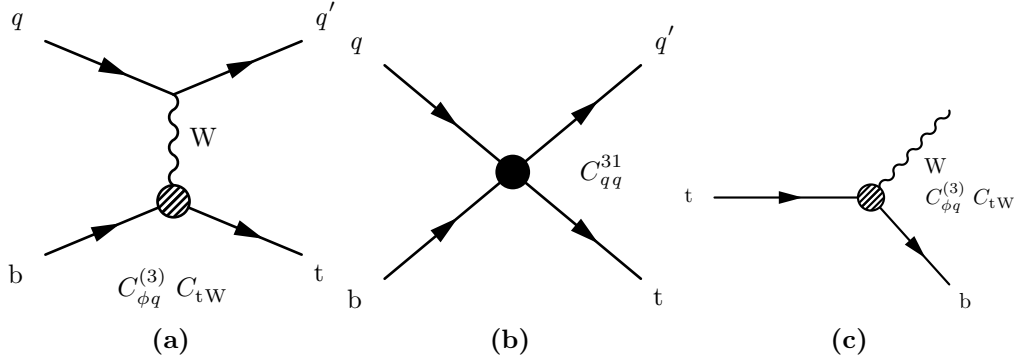
$$O_{qq}^{(3,ijkl)} = (\bar{q}_i \gamma^\mu \tau^I q_j) (\bar{q}_k \gamma_\mu \tau^I q_l) \quad (2.15)$$

$$(2.16)$$

Only the linear combination of these operators is relevant [58]:

$$C_{Qq}^{(3,1,ijkl)} = C_{qq}^{(3,ijkl)} + \frac{1}{6} (C_{qq}^{(1,ijkl)} - C_{qq}^{(3,ijkl)}).$$

Alltogether, there are three dominant operators contributing to single top quark production:  $C_{tW}$ ,  $C_{\varphi q}^{(3,ij)}$  and  $C_{Qq}^{(3,1,ijkl)}$  [61].



**Figure 2.8: Feynman diagrams with EFT.** Illustration where effective operators can enter into LO diagrams for single-top  $t$ -channel production. The operators  $C_{tW}$  and  $C_{\phi q}^{(3,ij)}$  effect the  $Wtb$  vertex and thus enter in both production (a) and decay (c). The operator  $C_{Qq}^{(3,1,ijkl)}$  enters only in the production, creating a four fermion contact interaction (b).

Table 2.2 lists current constraints on the EFT Wilson coefficients affecting the production and/or decay of single-top quarks. Shown are 95 % confidence level constraints on  $C_{tW}$ ,  $C_{\phi q}^{(3,ij)}$  and  $C_{Qq}^{(3,1,ijkl)}$  which are derived from LHC Run 1 and Run 2 data as well as Tevatron and Fermilab data [58]. Additionally, constraints on the real and imaginary part of  $C_{tW}$  are shown at 90 % confidence level [62]. They have been derived using different observables measured at the LHC, data from flavor physics — especially  $b \rightarrow s \gamma$  — and measurements of electric dipol moments (EDM). Individual constraints on CP violating top couplings are dominated by measurements of EDM and are typically orders of magnitude tighter than the constraints on the real parts of the corresponding operator [62]. These constraints are much stronger compared to what can be achieved in collider physics. However, if marginalized over many operators the constraints become significantly weaker as measurements of the EDM are only sensitive to  $C_\gamma - C_{tW}$  [62]. Uncorrelated measurements of the imaginary part of  $C_{tW}$  so far, stem only from the measurement of the phase between amplitudes for longitudinally polarized and transversely polarised W bosons recoiling against left-handed bottom quarks, which have been carried out by the ATLAS collaboration [63]. The polarization of the top quark is sensitive to both the real and imaginary part of  $C_{tW}$  and can therefore help to further constrain these coupling coefficients [41].

**Table 2.2: Constraints on Wilson coefficients.** Shown are constraints for Wilson coefficients that are relevant to single top quark production. Individual constraints correspond to the case where only one operator is considered to be non-zero. The marginalized constraints correspond to the case where all considered operators are set to float and the constraints are computed by marginalizing over the remaining couplings. See the corresponding references to see over which operators has been marginalized. Here  $\tilde{C}_i = C_i \nu^2 / \Lambda^2$ .

Operator	Individual	Marginalized	Confidence level	Reference
$\tilde{C}_{\varphi q}^{(3)}$	[-0.157, 0.091]	[-0.242, 0.206]	95%	[58]
$\tilde{C}_{Qq}^{3,1}$	[-0.024, 0.036]	[-0.036, 0.073]	95%	[58]
$\tilde{C}_{tW}$	[-0.079, 0.109]	[-0.242, 0.206]	95%	[58]
Re $\tilde{C}_{tW}$	[-0.095, 0.042]	[-0.11, 0.068]	90%	[62]
Im $\tilde{C}_{tW}$	[-0.0012, 0.0012]	[-0.21, 0.42]	90%	[62]





## 3 Statistical methods

### 3.1 Binned maximum likelihood fit

Maximum likelihood fits are commonly used tools for parameter estimations in High Energy Physics. The goal is to estimate some unknown parameters  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)$ , given some observed values  $\boldsymbol{x} = (x_1, x_2, \dots, x_N)$ . The observations are assumed to be statistically independent and to follow a probability density function (PDF)  $f(x|\boldsymbol{\theta})$ . Then the likelihood function is given as the product of all probability density functions:

$$L(\boldsymbol{x}; \boldsymbol{\theta}) = \prod_{i=1}^N f(x_i|\boldsymbol{\theta}). \quad (3.1)$$

The values  $\hat{\boldsymbol{\theta}}$  that maximize the likelihood are the maximum-likelihood estimates for the parameters  $\boldsymbol{\theta}$ . As it is more efficient to work with sums instead of products, normally the negative logarithm of the likelihood is minimized

$$-\ln L(\boldsymbol{x}; \boldsymbol{\theta}) = -\sum_{i=1}^N \ln f(x_i; \boldsymbol{\theta}). \quad (3.2)$$

In this thesis observations are grouped into bins in order to reduce the complexity of the likelihood. The expected number of events  $\lambda_i$  in each bin  $n_i$  follows a Poisson distribution.

$$f(k|\lambda) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (3.3)$$

Thus the binned likelihood function for  $B$  bins is given by:

$$L(\boldsymbol{n}; \boldsymbol{\theta}) = \prod_{i=1}^B \frac{\lambda_i(\boldsymbol{\theta})^{n_i}}{n_i!} e^{-\lambda_i(\boldsymbol{\theta})}. \quad (3.4)$$

The expected number of events in each bin is the sum of the expected number of signal events  $\lambda_{s,i}$  and the expected number of background events  $\lambda_{b,i}$ . These numbers are generally derived from MC simulation or some data-driven background estimation

method. In the maximum-likelihood fit, the expected number of signal events are then scaled by the signal strength modifiers  $\mu$ :

$$\lambda_i(\mu) = \mu \cdot \lambda_{s,i} + \lambda_{b,i}. \quad (3.5)$$

The likelihood then becomes a function of  $\mu$ . There can be multiple signal strength modifiers in case many signal processes are fitted simultaneously.

The expected number of signal and background events in each bin is typically subject to statistical and systematic uncertainties. The different systematic uncertainties that effect the expected number of events can be included in the maximum-likelihood fit in the form of nuisance parameters  $\alpha$ . The expected number of events will then additionally depend on the nuisance parameters  $\lambda(\mu) \rightarrow \lambda(\mu, \alpha)$ . The likelihood can be extended by a log-normal PDF for each nuisance parameter to take the nuisance parameters into account. In the example of a binned likelihood function with two nuisance parameters  $\alpha_S$  and  $\alpha_B$  with default values  $\alpha_{S/B}^{\text{In}}$ , the likelihood can be expressed as:

$$L(\mathbf{n}; \mu, \alpha_S, \alpha_B) \propto \prod_i \frac{\lambda_i^{n_i}(\mu, \alpha_S, \alpha_B)}{n_i!} \cdot e^{-\lambda_i(\mu, \alpha_S, \alpha_B)} \cdot e^{-\frac{1}{2}(\alpha_S - \alpha_S^{\text{In}})^2} \cdot e^{-\frac{1}{2}(\alpha_B - \alpha_B^{\text{In}})^2}. \quad (3.6)$$

In the example of a counting experiment the expected number of events will depend on the nuisance parameters, as

$$\lambda_i(\mu, \alpha_S, \alpha_B) = \mu \cdot \lambda_{s,i} \cdot (1 + p_s)^{\alpha_S} + \lambda_{b,i} \cdot (1 + p_b)^{\alpha_B}. \quad (3.7)$$

Where,  $p_{s,b}$  are the relative uncertainties in the signal/background prediction. Often the likelihood function is only known at discrete values of the parameters it depends on. For minimizing the likelihood it is however easier to work with a continuous function. Therefore, template morphing techniques are used to parameterize the likelihood as a function of the nuisance parameters. The parameterization is done by interpolating between the expected number of events for a one standard deviation up and down variation of the systematic uncertainty.

## 3.2 Machine learning

Machine learning methods are used to classify events into different categories for example signal and background. In this thesis machine learning algorithms like boosted decision trees and neural networks are used as classifiers. These algorithms have a set of parameters, which need to be fitted to the data. This is also called training or learning. They are trained via supervised learning, which means data with known labels is used. In the training an objective function - called loss function - is minimized. A typical loss

function used for classification problems is the cross-entropy. In information theory the entropy - or Shannon entropy - gives the minimum number of bits that are necessary to encode some data [64]. For some data  $\mathbf{x}$  following a probability distribution  $p$  the entropy is defined as:

$$H(\mathbf{x}) = - \sum_i p(x_i) \log p(x_i). \quad (3.8)$$

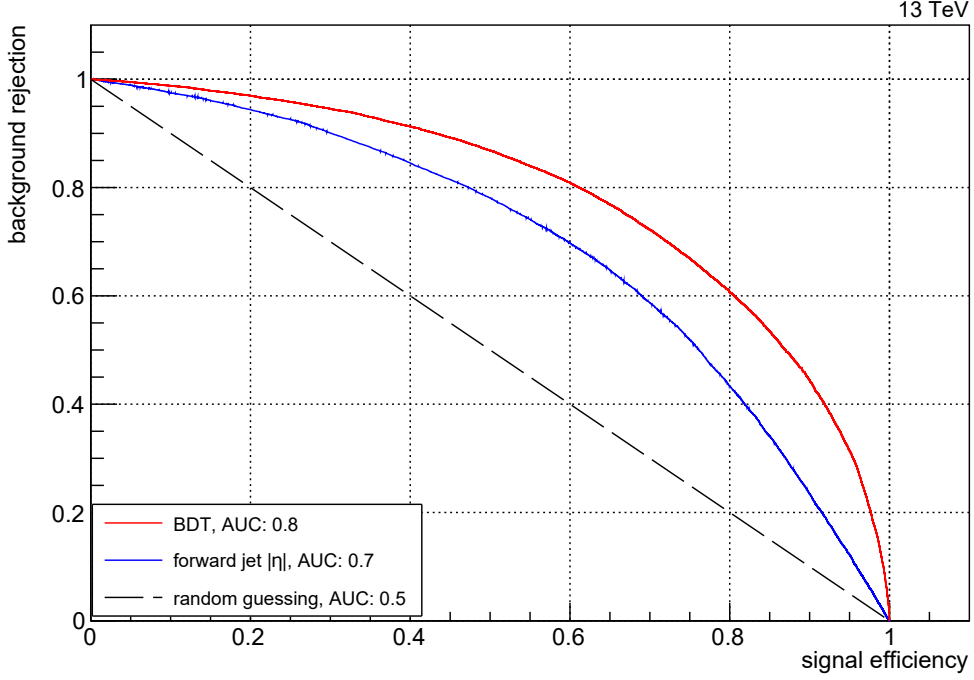
The cross-entropy  $H(p, q)$  between two probability distributions  $p$  and  $q$  is defined as:

$$H(p, q) = - \sum_i p(x_i) \log q(x_i). \quad (3.9)$$

Generally  $H(p, q) \geq H(p)$ , where they are equal only if  $q$  can perfectly describe  $p$ . The difference between the entropy and the cross-entropy tells one essentially how much information is lost if  $p$  is decoded using  $q$ . In supervised learning  $p$  will correspond to the truth labels of some classes, which can be interpreted as a probability distribution. The prediction of the model for each class label corresponds to  $q$ . Thus, the cross-entropy is a measure for how similar the predicted and true distributions are.

Classifiers will have a certain efficiency (or probability) of correctly classifying a signal event as signal (true positive rate) and a certain probability of wrongly classifying background event as a signal (false positive rate). Plotting the true positive rate against the false positive rate for all possible values of the discriminator gives the receiver operating characteristic (ROC) curve. From the shape and the area under the ROC curve the quality of a classifier can be judged. The ROC curve can also be used to define working points, which are used in a cut-based analysis. The working point is an output value of the discriminator that corresponds to a certain false positive rate. Output values larger than the working point are defined as signal and output values lower than the working point as background. In figure 3.1 an example ROC curve for a BDT trained to separate single-top  $t$ -channel events from background events is shown together with the ROC curve obtained from only using the pseudorapidity of the spectator jet as discriminator and a ROC for a discriminator that is randomly guessing. An important figure of merit that can be derived from ROC curves is the area-under-curve (AUC) value, which is the area under the ROC curve. An AUC value of 0.5 corresponds to a discriminator that is randomly guessing and hence has no separation power. The larger  $|AUC - 0.5|$  the better the discrimination power of the classifier.

During the training of machine learning algorithms one has to ensure that they do not overfit the training data set. Overfitting happens if the network essentially learns the training data by heart but fails to generalize to new, unseen data. In order to avoid overfitting first of all a good training setup is necessary. This requires splitting the data in at least three data sets, namely a training, a validation and a testing data set. The training data set is used for adjusting the parameters of the model. The validation data set for tuning hyperparameters and to check for overtraining during the training phase.



**Figure 3.1: Illustration of ROC curves.** Different ROC curves for separating  $t$ -channel single-top-quark signal events from background events are shown. Shown are the ROC curves for an optimized BDT (red) and for the absolute value of the pseudorapidity of the spectator jet (blue). The dashed line corresponds to a discriminator which is randomly guessing.

Finally, the testing data set is used to validate the performance of the final model. During the training overtraining can be spotted by monitoring the performance of the model on both the training and the validation data set. Overtraining will show itself as a divergence between the two were the performance on the training data set increases but the performance on the validation data set decreases. If the model is susceptible to overtraining, regularization techniques should be used or the model complexity has to be decreased.

### 3.2.1 Boosted Decision Trees

Decision trees are simple predictive models that can be trained using machine learning techniques. In this thesis decision trees are used for classification of events into different categories. However, decision trees can also be trained for regression, e.g. to predict a continuous variable. A visual representation of a decision tree is shown in figure 3.2. Decision trees start from a root node and then branch into more nodes until they end up in the final nodes, also called leaf nodes. The splits are made based on the value of a variable. During the training of a decision tree the variable and value that maximize

an objective function are determined at each splitting step. In this thesis the objective function that is maximized is the cross-entropy (see section 3.2).

Individual small decision trees are very weak classifiers, while large decision trees are very susceptible to overtraining. Therefore, typically an ensemble of multiple smaller decision trees are constructed to build a forest of decision trees. This ensemble is then a much more robust and powerful classifier. There are generally many different techniques to build ensembles out of individual classifiers, one of the simplest being to simply average their output. Individual decision trees are weak classifiers, which means they are only weakly correlated with the objective. Therefore, decision trees naturally lend themselves to an ensemble building technique called boosting [65,66]. The general idea behind boosting is to iteratively train weak classifiers and then combine them to a final strong classifier. During the combination each classifier is weighted based on how well it performed during the training. There exist many different boosting algorithms. In this thesis gradient boosting is used. In gradient boosting the gradient descent algorithm is used to find a function  $F(x)$  (the model prediction) that minimizes the specified loss function  $L(y, F(x))$ , with the true values  $y$ . The model prediction is the weighted sum of the outputs  $h_i$  of a set of individual decision trees:

$$F(x) = \sum_i^M \gamma_i h_i(x). \quad (3.10)$$

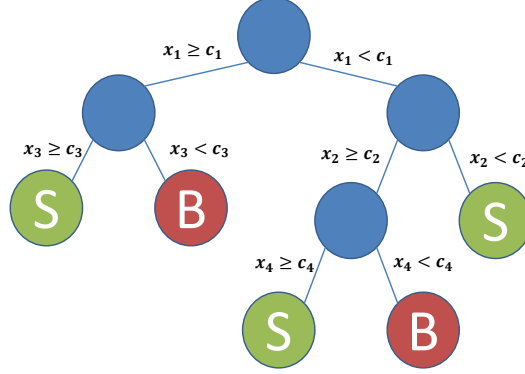
During the training the model is build in an iterative way. This is done by starting with a constant prediction and then adding the output  $h_m$  of one decision tree at each step  $m$ . The output of the decision tree  $h_m$  at step  $m$  is optimized on the residual difference  $h_m(x) = y - F_{m-1}(x)$  between the model predictions at the previous step  $F_{m-1}(x)$  and the true values  $y$ .

### 3.2.2 Artificial neural networks

Artificial neural networks (ANN) are function that are used for various classification and regression tasks in machine learning applications. Their architecture is inspired from biological neural networks in human brains. ANNs consists of multiple artificial neurons. Each neuron receives inputs  $x_i$  and produces a single output that it sends to multiple connected neurons. The importance of each connection between neurons is represented by a weight  $w_i$ . The output is further passed through an activation function  $\sigma$ . Commonly used activation function are for example the rectifier linear unit (ReLU)  $\sigma(x) = \max(0, x)$ , or the softmax function:

$$\sigma_i(x) = \frac{e^{-x_i}}{\sum_j e^{-x_j}}. \quad (3.11)$$

The output of a single neuron is computed as:



**Figure 3.2: Illustration of a decision tree.** Decision trees have a tree structure. It starts from a root node and at each node a binary decision based on the value of one variable is applied and the event is forwarded to the left or right node depending on the result. This is done until it ends up in a final node or leaf node. The event is classified as signal if the final node it ends up in is a signal node (S) and as background if it ends up in a background final node (B). Whether a final node is a signal or a background node depends on whether more signal or more background events ended up in that node during the training.

$$f(x) = \sigma\left(\sum_i w_i x_i\right) \quad (3.12)$$

A network is built of multiple stacked layers of neurons. The output of a neuron in the next layer can be computed via function composition:

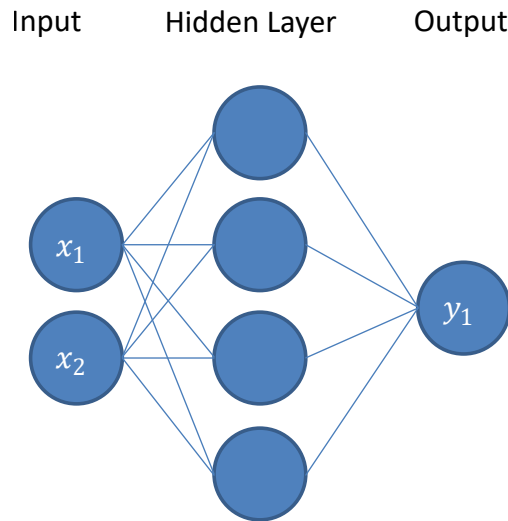
$$g(f) = \sigma\left(\sum_i w_i f_i\right), \quad (3.13)$$

where  $f_i$  are the outputs of the neurons in the previous layer. An illustration of an ANN is given in figure 3.3.

The weights of the ANN need to be adapted to produce the desired output. This happens during the training. The gradient descent algorithm is typically used to minimize the loss function. Parameters are iteratively updated in the direction of the steepest descent by computing the gradient of the neural network with respect to each parameter:

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \eta \nabla f(\mathbf{w}_n). \quad (3.14)$$

Here  $\eta$  is the learning rate that determines the step size in each iteration. In practice often variation of pure gradient descent are used such as stochastic gradient descent, which computes an approximate gradient over a subset of the entire data set. Finding a good learning rate is important in order to converge to a good minima and also



**Figure 3.3: Illustration of an artificial neural network (ANN).** An ANN consists of several neurons which are connected with each other. The neurons are arranged in layers where each neuron is connected to all neurons in the previous layer.

determines the number of iterations needed in order to converge. In the training of neural networks often adaptive optimizers like Adam [67] are used. These optimizers adjust the learning rate automatically during the training based on the gradients in the previous iteration. They also use a set a different learning rate for each weight in the neural network. Computing the gradient for each weight individually is computationally very inefficient. Instead ANNs are trained using backpropagation. Backpropagation makes use of the chain rule to compute the gradients iteratively one layer at a time starting from the output layer.

ANNs may be suspect to overtraining. In order to avoid this different regularization techniques can be used. Commonly used method include  $L_1$  and  $L_2$  regularization that add a linear ( $L_1$ ) or quadratic ( $L_2$ ) penalty terms to the loss function. Another method is dropout, which randomly drops neurons and their connections from the network [68].

### 3.3 Unfolding

Distributions at reconstruction level are distorted due to the detector's limited resolution, finite acceptance and due to reconstruction inefficiencies. Unfolding aims at undoing these distortions in order to obtain the true level distribution. In this chapter the statistical basis of unfolding is established. The following description follows mostly the book [69].

Given some measured values  $s$  and true level variable  $t$ . The measured distribution  $g(s)$  is related to the true level distribution  $f(t)$  by migrations, distortions and transformations that are described by a kernel function  $K(s, t)$ . They are related to each other via the Fredholm integral equations of the first kind:

$$\int_{\Omega} K(s, t) f(t) dt + b(s) = g(s), \quad (3.15)$$

here  $b(s)$  describes the background distribution and  $\Omega$  denotes the phase space.

In practice one deals with binned distributions. Thus the continuous distributions are replaced by vectors:  $f(t) \rightarrow \mathbf{x}$ ,  $g(s) \rightarrow \mathbf{y}$  and the kernel function by a response matrix  $K(s, t) \rightarrow \mathcal{A}$ . In matrix form equation (3.15) reads:

$$\mathbf{y} = \mathcal{A}\mathbf{x} + \mathbf{b}. \quad (3.16)$$

The element  $\mathcal{A}_{ij}$  can be interpreted as the probability of an event belonging into the truth bin  $j$  to be measured in bin  $i$ :  $\mathcal{A}_{ij} = P(\text{measured in bin } i | \text{truth in bin } j)$ . The sum over  $\mathcal{A}_{ij}$  over all bins of the measured distribution  $i$  gives the detection efficiency:

$$\epsilon_j = \sum_i \mathcal{A}_{ij} \quad (3.17)$$

Solving equation (3.16) by simple matrix inversion leads to unstable solutions with large fluctuations. The origin of these fluctuations can be seen by orthogonalisation of the response matrix through singular value decomposition (SVD).

$$\mathcal{A} = \mathcal{U}\mathcal{S}\mathcal{V}^T = \sum_i \sigma_i \mathbf{u}_i \mathbf{v}_i^T. \quad (3.18)$$

Here  $\mathcal{S} = \text{diag}(\sigma_1, \dots, \sigma_n)$  is a diagonal matrix and  $\mathcal{U}$  and  $\mathcal{V}$  are matrices with orthonormal columns. The  $\sigma_i$  are called singular values and are non-negative real numbers, which are ordered in decreasing order. The column-vectors  $\mathbf{u}$  and  $\mathbf{v}$  are called



left and right singular vectors of  $\mathcal{A}$ . The condition number of  $\mathcal{A}$  is defined as the ratio between the largest and the smallest singular value:

$$\text{cond}(\mathcal{A}) = \frac{\sigma_1}{\sigma_n}. \quad (3.19)$$

When inverting  $\mathcal{A}$  one obtains contributions proportional to  $1/\sigma_i$ . Thus, if the smallest singular value  $\sigma_i$  is very small the unfolded distribution is amplified leading to large, unphysical fluctuations. In order to reduce these fluctuations, regularization techniques are used. A simple approach is to introduce a parameter  $\tau$ , which serves as a cutoff. Singular values that are smaller than this cutoff are simply removed from the response matrix  $\mathcal{A}$ , thus reducing fluctuations [70]. In Tikhonov regularization [71–73] some penalty term  $\|\mathcal{L}\mathbf{x}\|^2$ , which ensures the smoothness of the solution, is added to the minimization. Regularization generally requires to find a proper choice of the regularization strength parameter, for which there exists no general method. In case of large regularization strengths it also severely biases the results.

As equation (3.16) often cannot be solved using simple inversion, some expression is defined which is minimized. Some tools, like TUnfold [74] minimize a  $\chi^2$  expression, which in its base form can be written as:

$$\chi^2 = (\mathbf{y} - \mathbf{b} - \mathcal{A}\mathbf{x})^T \Sigma^{-1} (\mathbf{y} - \mathbf{b} - \mathcal{A}\mathbf{x}) + \chi_{\text{regularization}}^2 \quad (3.20)$$

with the covariance matrix  $\Sigma$ . Performing the unfolding via a  $\chi^2$ -minimization requires a simple inversion of the Hesse matrix, which makes the minimization very fast. It is also well suited for unfolding multi-differential distributions.

However, in this thesis the unfolding is performed using a maximum-likelihood fit, which is technically implemented using the HIGGS COMBINE tool [75]. This method is significantly slower as it requires a numerical minimisation with Minuit [76]. However, it also has a couple of advantages over the faster  $\chi^2$ -minimization:

- The number of background events is estimated simultaneously with the number of signal events. Therefore, no separate background subtraction is needed during the unfolding process.
- The signal and background components can be constrained in the fit by adding additional signal or background enriched event categories or ancillary dimensions to the fit. In this thesis, this also allows to use all selected events for the unfolding, instead of defining a small, signal enriched phase space region. Thereby, reducing possible biases on the result from the selection of a signal enriched phase space and uncertainties due to the extrapolation to the full phase space.
- Systematic uncertainties can be added to the fit as nuisance parameters. This allows for a consistent treatment of all systematic uncertainties. Also systematic uncertainties due to the limited size of the simulated samples can be included using the Barlow Beeston lite method [77, 78].

- The fitted number of events actually correspond to the observed number of events, whereas the Neyman  $\chi^2$  function used for example in TUnfold contains a normalization bias term, which requires using an additional area constraints.

The likelihood that is minimized in this case can be written as:

$$\mathcal{L} = \prod_{i \in \text{reco}} \mathcal{P}(y_i | \sum_{j \in \text{gen}} \mu_j \mathcal{A}_{ij}(\boldsymbol{\theta}) + b_i) \quad (3.21)$$

Here,  $\mathcal{P}$  is the Poisson-PDF,  $\boldsymbol{\theta}$  correspond to nuisance parameters that may effect the response matrix  $\mathcal{A}$ , and  $\mu_j = \frac{x_j^{\text{obs}}}{x_j^{\text{true}}}$  are the signal strength parameters. The measured number of events in each bin can be computed from the fitted signal strength parameters.

It can be useful in a maximum likelihood fit to add additional regions or dimensions to the fit that can be used to constraint signal or background components. The maximum-likelihood approach allows to add an ancillary dimension  $k$  to the fit:

$$\mathcal{L} = \prod_{i \in \text{reco}} \prod_k \mathcal{P}(y_{ik} | \sum_{j \in \text{gen}} \mu_j \mathcal{A}_{ijk} + b_{ik}) \quad (3.22)$$

In this thesis, the distributions are additionally binned in bins of the output of a multiclassification BDT, which has been trained to categorize events into different process categories. The output bins of the BDT provide the ancillary dimension  $k$  in this case.

## 4 The Compact Muon Solenoid detector at the Large Hadron Collider

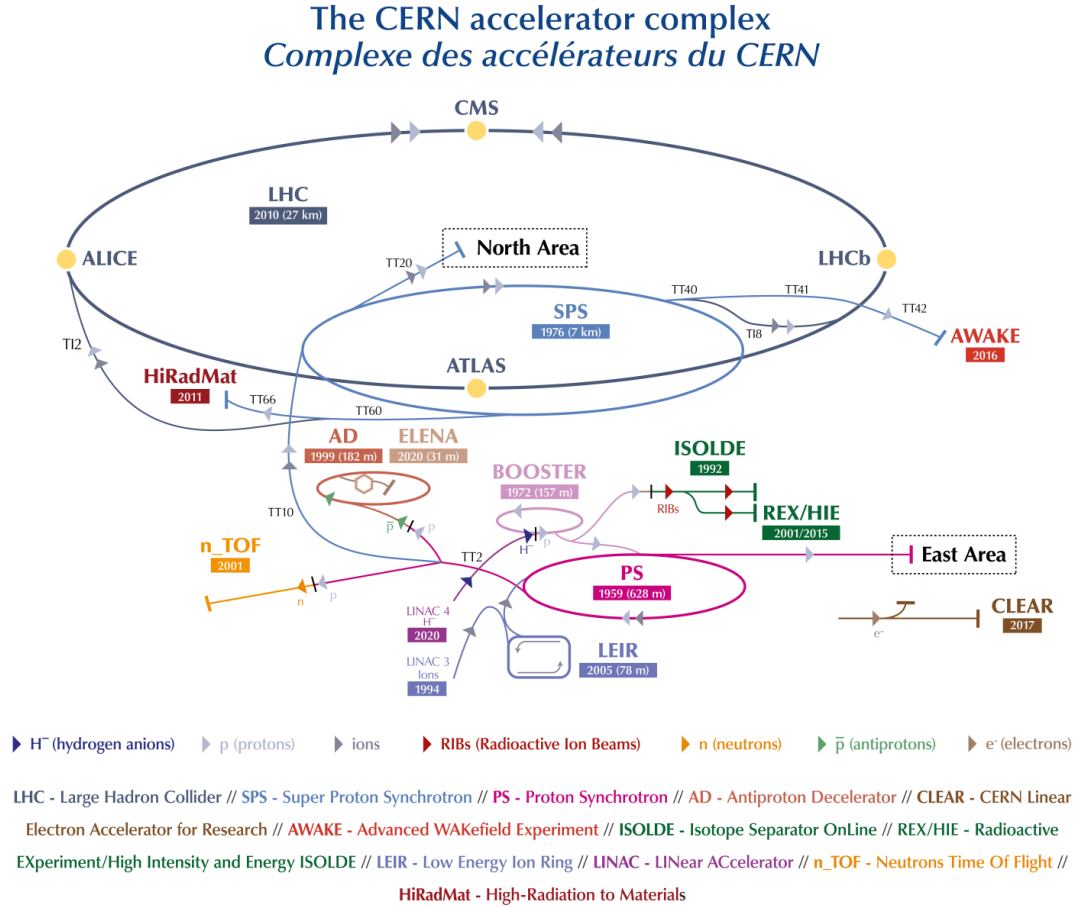
### 4.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is the largest and most powerful particle accelerator ever built. Its construction started in 1998 and first operations began in 2008. The LHC is located at the European Organization for Nuclear Research (CERN) in Geneva.

The LHC is located in a circular 27 km long tunnel 50 m to 175 m below the ground. Before the LHC, the tunnel was used by the Large Electron-Positron Collider [79]. LEP operated from 1989 to 2000 at a center-of-mass energy of 209 GeV. The targeted center-of-mass energy for proton-proton collisions of the LHC is more than fifty times larger with 14 TeV. However, due to technical difficulties Run 1 started in 2010 with a center-of-mass energy of only 7 TeV. In 2012 the energy was increased to 8 TeV. From 2013 to 2015 the LHC was shutdown, upgrading it to enable collisions at 14 TeV. Run 2 started in 2015 and lasted until 2018. In this period the LHC operated at a center-of-mass energy of 13 TeV. The collision energy is planned to be raised in future runs to reach the design center-of-mass energy of 14 TeV. In 2018 the LHC was shutdown for its second upgrade period, which will last until 2021. The goal of this upgrade is to prepare the LHC for runs with about 10 times higher collision rate.

Several experiments are located at the LHC. There are the ATLAS [80] and the Compact Muon Solenoid (CMS) [81] experiment, which are general purpose detectors. They have been built to search for the Higgs boson and potential new particles not predicted by the SM, but also to measure more precisely the properties of already known particles. The advantage of having two independent detectors is the possibility to cross check their results. The goal of the A Large Ion Collider Experiment (ALICE) [82] is to study the strong interactions and quark-gluon plasma from heavy-ion collisions. The Large Hadron Collider Beauty (LHCb) [83] experiment measures properties of B mesons and CP violation. These four experiments are located at different interaction points around the LHC ring. Their position can be seen in figure 4.1.

The protons that are accelerated in the LHC stem from a hydrogen bottle. The hydrogen atoms are ionized in an electric field and then accelerated. Before being injected into the LHC, the protons go through a series of preaccelerators. They are first accelerated to 50 MeV using the Linear Accelerator 2. Three synchrotron accelerators are then used to increase the energy of the protons to 450 GeV. These are the Proton



**Figure 4.1: Overview of the CERN accelerator complex.** Shown is a sketch of the LHC and other accelerators at CERN. The Linear Accelerator 2 (LINAC 2), Proton Synchrotron BOOSTER, Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS), are used in this order as preaccelerators for the LHC. Shown are also the collision points around the LHC ring (yellow dots). At each collision point one of the four major experiments at the LHC is located: ALICE, ATLAS, CMS and LHCb [84].

Synchrotron Booster (up to 1.4 GeV), the Proton Synchrotron (up to 25 GeV), and the Super Proton Synchrotron (SPS). An overview of the LHC and the preaccelerators is shown in figure 4.1. From the SPS the protons are injected into the two beam pipes of the LHC in opposite directions. Ramping up the energy from 450 GeV to the collision energy of 6.5 TeV takes about 20 min. After a successful fill a stable beam can circulate in the LHC ring for more than 10 hours. During one fill the instantaneous luminosity gradually decreases, as protons are lost in the collisions. Once the instantaneous luminosity becomes too low, the fill ends and the beam is dumped. For this purpose the beam is steered into a 750 m long tunnel using kicker magnets. There the beam is diluted and dumped into a thick water cooled graphite block.

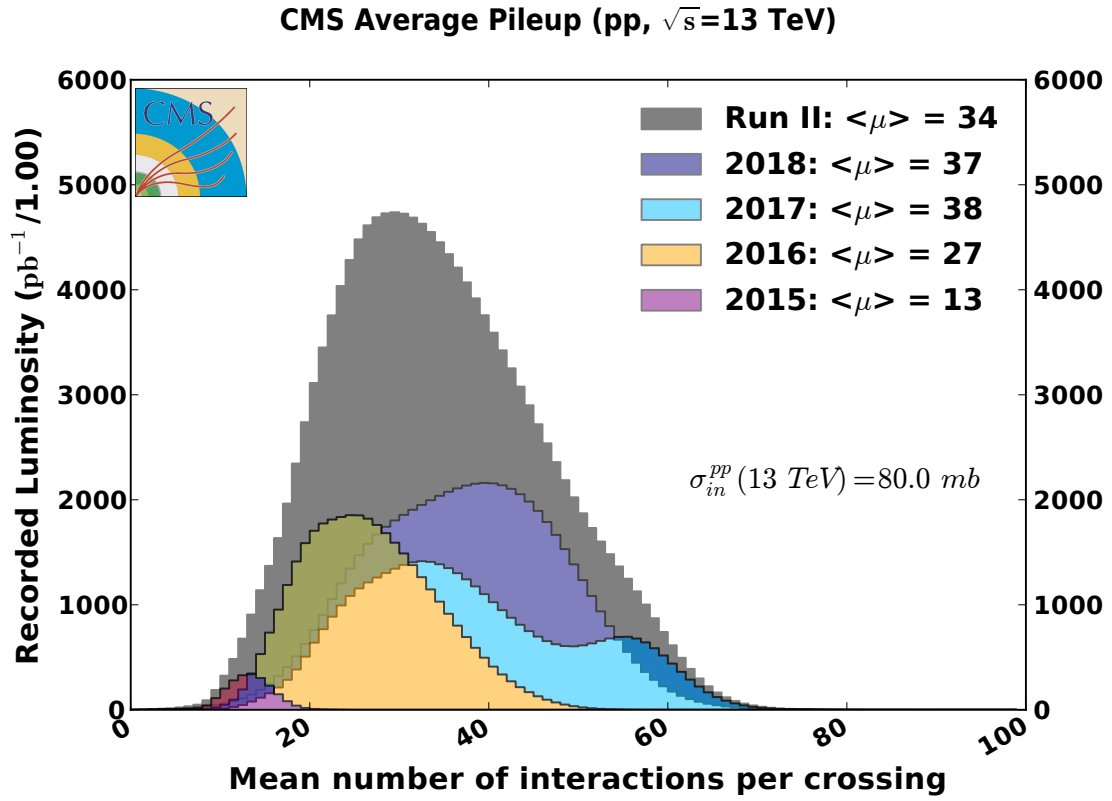
Superconducting dipole magnets are used in order to keep the protons on a circular path. Quadrupole and higher multipole magnets are used as correctors and to control the beam size. The magnets used at the LHC are based on NbTi, cooled to 2 K using superfluid helium and create magnetic fields of more than 8 T. In the LHC the protons are accelerated using 400 MHz superconducting cavities. The cavities provide a field strength of  $5.5 \text{ MV m}^{-1}$ . In order to avoid collisions of the protons with gas molecules, an ultra high vacuum is kept in the beam pipes with the pressure being in the range of  $10^{-10}$  mbar to  $10^{-11}$  mbar.

At the LHC protons are not accelerated individually but in bunches. Each bunch consists of about  $1.15 \cdot 10^{11}$  protons. The number of bunches per fill depends on the filling scheme used during the injection. In Run 2 the time between two consecutive bunches is 25 ns. This corresponds to a distance between two bunches of about 7.5 m. With the circumference of the LHC being a bit less than 27 km, a total of 3564 bunches could maximally fit into the LHC ring. However, there have to be gaps between the bunches to take into account the rise time of the SPS and LHC kicker magnets, which are used for injection. Therefore, for 25 ns schemes the bunches can be arranged in 39 batches of 72 bunches resulting in a maximum of 2808 total bunches [85]. In practice, the number of bunches per fill is lower. So far, the highest number of bunches per fill have been 2556, which was achieved for the first time in the first run period of 2018.

As bunches consist of multiple protons, typically more than one proton-proton interaction takes place per bunch crossing. This is called (in-time) pileup and is quantified by the average number of interactions per bunch crossing. The pileup distribution for the full Run 2 period is shown in figure 4.2. The average number of interactions per bunch crossing over the full Run 2 period was 34. In addition to the in-time pileup there can also be out-of-time pileup, where the signals from two consecutive bunch crossings mix in the detector. This happens if the readout time of the signals in the detector is larger than the time between two bunch crossings.

The production rate of a physical process depends on the cross section  $\sigma$  for the particular process and the instantaneous luminosity  $\mathcal{L}$ :

$$\frac{dN}{dt} = \sigma \mathcal{L} \quad (4.1)$$



**Figure 4.2: Run 2 pileup profile.** Shown is the distribution of the average number of interactions per crossing (pileup) for proton-proton collisions in 2015 (purple), 2016 (orange), 2017 (light blue), and 2018 (dark blue). In grey the combined pileup profile of all data taken in Run 2 is shown. [86].

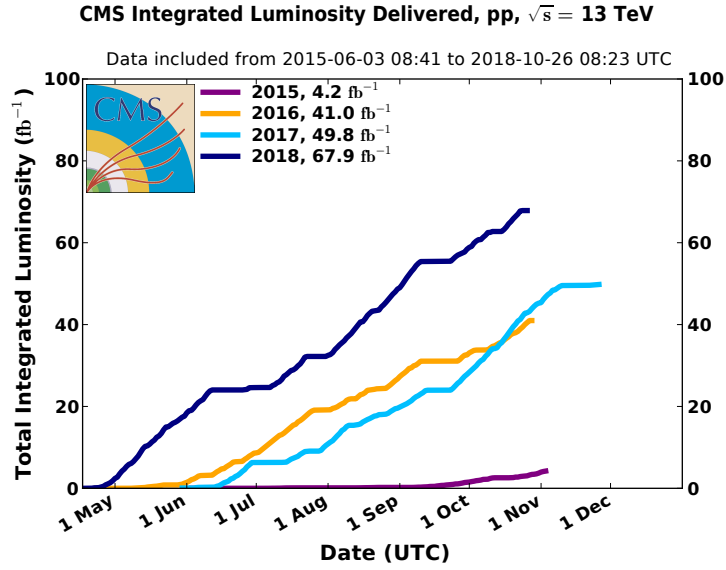
The total number of events produced is given by integrating equation 4.1 over time. As the cross section depends only on the center-of-mass energy, which is constant over the time of a fill, the number of events is given by:

$$N = \sigma \int dt \mathcal{L} = \sigma L. \quad (4.2)$$

Here  $L$  is called the integrated luminosity and is a measure of the number of collisions. The instantaneous luminosity depends on the parameters of the LHC:

$$\mathcal{L} = \frac{N_b^2 n_b f_{\text{rev}} \gamma}{4\pi \epsilon_n \beta^*} F, \quad (4.3)$$

Here  $\gamma$  is the relativistic gamma factor,  $N_b$  the number of particles per bunch,  $n_b$  the number of bunches per beam and  $f_{\text{rev}}$  is the revolution frequency of the bunches. The normalized beam emittance  $\epsilon_n$  gives the spread of the beam in position and momentum space. The beta function  $\beta^*$  describes the focusing of the beams at the interaction point. The two beams are collided at a nonzero angle, reducing the effective area. This is taken into account by the form factor  $F$ . The design instantaneous luminosity of the LHC is  $2.0 \cdot 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ . It was reached for the first time in 2017. Figure 4.3 shows the total integrated luminosity for 2015–2018 as a function of time. The total integrated luminosity delivered by the LHC in Run 2 was  $162.9 \text{ fb}^{-1}$ . More details on the LHC can be found in reference [87].



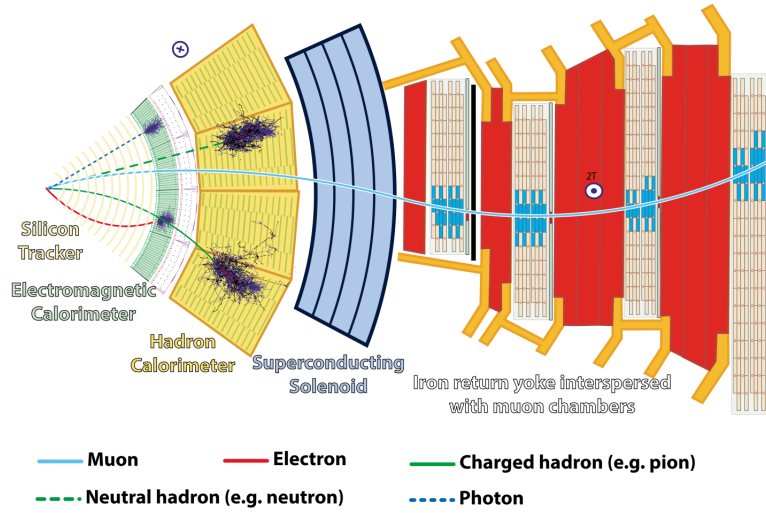
**Figure 4.3:** Delivered luminosity versus time for data taken from 2015 to 2018 at  $\sqrt{s} = 13$  TeV. The delivered luminosity is shown for 2015 (purple), 2016 (orange), 2017 (light blue), and 2018 (dark blue). The different years are all shown over the same range. The luminosities are shown for runs with stable beams and proton-proton collisions only. [88].

## 4.2 The Compact Muon Solenoid

The Compact Muon Solenoid (CMS) detector is a general purpose detector at the LHC. It is designed and used for both proton-proton collisions and heavy ion collisions. The CMS detector has a length of 21.6 m, a diameter of 14.6 m and weighs 14 000 t. The CMS detector is build cylindrical around the beam pipe with the forward calorimeters covering a pseudorapidity range up to  $|\eta| \leq 5$ . Figure 4.4 shows a vertical slice of the CMS detector. From inside to outside the detector consists of a silicon tracker, an electromagnetic calorimeter, an hadronic calorimeter, a superconducting solenoid and muon chambers inside the iron return yoke of the magnet. The superconducting solenoid provides a homogeneous magnetic field of 3.8 T.

The coordinate system used at CMS has the center at the collision point. The  $y$ -axis points vertically upwards, the  $x$ -axis radially inwards towards the center of the LHC ring and the  $z$ -axis along the beam direction, forming a right-handed coordinate system. The  $\phi$  angle is measured from the  $x$ -axis to the  $y$ -axis in the  $x - y$  plane. The radial coordinate in the  $x - y$  plane is called  $r$ . Instead of the polar angle  $\theta$ , which is measured from the  $z$ -axis, typically the pseudorapidity  $\eta = -\ln(\tan(\theta/2))$  is used. If not specified differently, information provided in this chapter is taken from the CMS technical design report [81].





**Figure 4.4: Slice of the CMS detector** Shown is a slice of the CMS detector and the signals that different particles create in the detector. The particle tracks originate from the collision point at the very left and pass through the silicon tracker, where electrically charged particles leave a track. Next comes the electromagnetic calorimeter, where electromagnetically interacting particles deposit energy and may get fully absorbed. Charged and neutral hadrons then deposit the majority of their energy in the hadron calorimeter. Muons pass through the whole detector and are measured in the muon system. The inner detector components are encapsulated by a superconducting solenoid magnet, which creates a magnetic field of about 3.8 T in the inner part of the detector and of about 2 T in the muon system [89].

### 4.2.1 Tracker

The first detector layer around the interaction point is formed by the tracker [81]. Its purpose is to precisely measure the trajectories of charged particles and the reconstruction of vertices. Inside the tracker a homogeneous magnetic field of 3.8 T is provided by the solenoid. This magnetic field bends the trajectory of charged particles, which allows to measure their transverse momenta from the curvature radius.

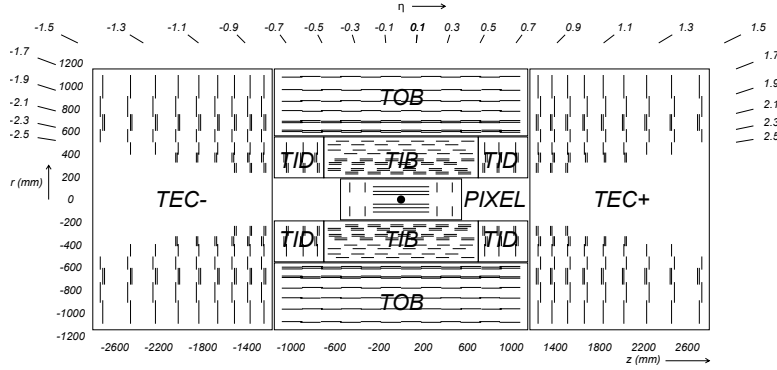
The tracker is 5.8 m long and has a diameter of 2.5 m and is completely made out of silicon with a total active silicon area of 200 m<sup>2</sup>. The tracker covers a pseudorapidity range of  $|\eta| < 2.5$ . There are two subdetectors inside the tracker. The first is a silicon pixel detector that directly surrounds the collision point. The second is a silicon strip detector built around the pixel detector. Both subdetectors have an independent cooling system, power supply, and readout system.

Both silicon detectors use a p-n junction. A negative bias voltage is applied, increasing the depletion region. Incoming charged particles create electron-hole pairs, which move to the electrodes where the charges are collected and readout.

The pixel detector has about 66 million pixels each with a size of 100  $\mu\text{m}$  by 150  $\mu\text{m}$ . Due to the vicinity to the interaction point the flux of particles through the pixel detector is very high. Therefore, an n+ pixel on n-substrate detector design is chosen. This design allows for partial depleted operation and thus increases the radiation resistance. The pixel detector achieves a spatial resolution of 15  $\mu\text{m}$  to 20  $\mu\text{m}$ . In 2017 the pixel detector was updated [90] because radiation damage would have lead to high inefficiencies in the innermost parts of the tracker over time. The new pixel detector is closer to the interaction point with the first layer at a radius of 2.9 cm, due to the reduced size of the beam pipe. The number of barrel layers has been increased from three to four. Also the number of endcap disks have been increased from two to three. Despite the additional layers, the new pixel detector is lighter than the previous one due to an ultra-light mechanical design. The new pixel detector also has a new readout chip and a new cooling system.

The strip detector consists of about 15000 strip detector modules. There are a total of ten detection layers in the barrel region. Four in the tracker inner barrel (TIB) and six in the tracker outer barrel (TOB). The endcaps consist of nine layers, three in the tracker inner disks (TID) and six in the tracker endcap (TEC). Figure 4.5 shows the layout of the CMS tracker before the 2017 pixel update.

For charged hadrons with  $p_T < 10 \text{ GeV}$  the transverse momentum resolution of the tracker is below 1% in the barrel region. At these  $p_T$  values the momentum resolution is limited by multiple scattering. At higher momentum the resolution gets worse as the curvature of the particle trajectories decreases. Thus hits in different layers are closer together and the momentum resolution becomes limited by the resolution of the tracker [91].



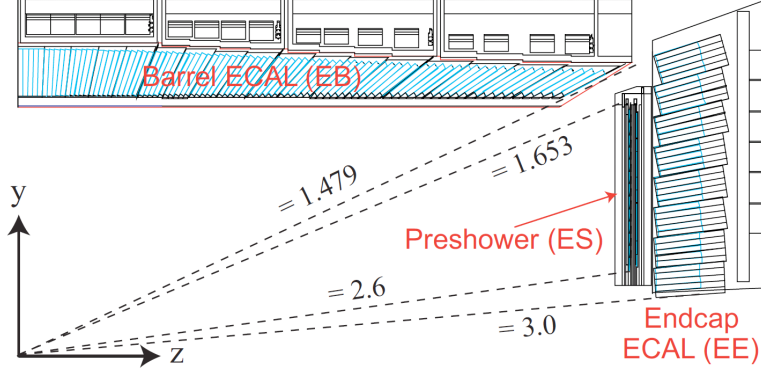
**Figure 4.5: Layout of the CMS tracker before the Phase 1 pixel update in 2017.** The tracker consists of a silicon pixel detector (PIXEL) directly around the collision point and a silicon strip detector surrounding the pixel detector. The pixel detector, as shown here, had three barrel layers and two endcap disks (illustrated by the lines). Since the pixel update in 2017 the pixel detector has four barrel layers and three end-cap disks. The silicon strip detector consists of the tracker inner barrel (TIB), tracker outer barrel (TOB), tracker inner disks (TID) and tracker endcap (TEC). The tracker covers up to  $|\eta| = 2.5$  [81,90].

#### 4.2.2 Electromagnetic calorimeter

The CMS electromagnetic calorimeter (ECAL) [81] is a homogeneous calorimeter that consists of a single layer of lead tungstate ( $\text{PbWO}_4$ ) crystals. Its purpose is to measure the energy of electromagnetically interacting particles, mainly photons, electrons, and positrons. High-energetic particles traversing through the ECAL produce an electromagnetic shower, which is a cascade of secondary particles. Photons start this shower via pair production of electrons and positrons, whereas charged particles emit secondary photons via bremsstrahlung. The crystals produce scintillation light proportional to the number of secondary particles, which is proportional to the energy of the initial particle. This light is measured through avalanche photodiodes.

In order to measure the full energy of the particles, the particles should ideally be fully absorbed in the ECAL.  $\text{PbWO}_4$  has a high density of  $8.28 \text{ g cm}^{-3}$  and a high atomic number resulting in a short radiation length of 0.89 cm. The crystals have a length of about 230 mm, which corresponds to about 25.8 radiation lengths. This is sufficient to absorb approximately 98 % of the energy of electrons and photons with an energy below 1 TeV in the ECAL. Only the remaining energy is deposited in the hadron calorimeter. The size of the ECAL corresponds to one nuclear interaction length for hadrons. Therefore, hadrons deposit some of their energy already in the ECAL [92]. A further advantage of  $\text{PbWO}_4$  is its short scintillation decay time. About 80 % of the light is emitted in the time between two bunch crossings, which is 25 ns.

61 200 lead tungstate crystals are located in the barrel part (EB) of the ECAL, which covers a pseudorapidity range of  $|\eta| < 1.479$ . The two ECAL endcaps (EE) cover the



**Figure 4.6: Layout of the CMS ECAL.** Shown are the ECAL barrel (EB), preshower (ES) and endcap (EE) detector. [94].

range  $1.479 < |\eta| < 3.0$  and consist of 7324 crystals each. In order to improve the neutral pion identification, an ECAL preshower detector (ES) is placed in front of the endcaps, covering the pseudorapidity range of  $1.653 < |\eta| < 2.6$ . The ES consists of a lead absorber and silicon strip sensors and is about 3 radiation lengths thick. The layout of the ECAL is shown in figure 4.6

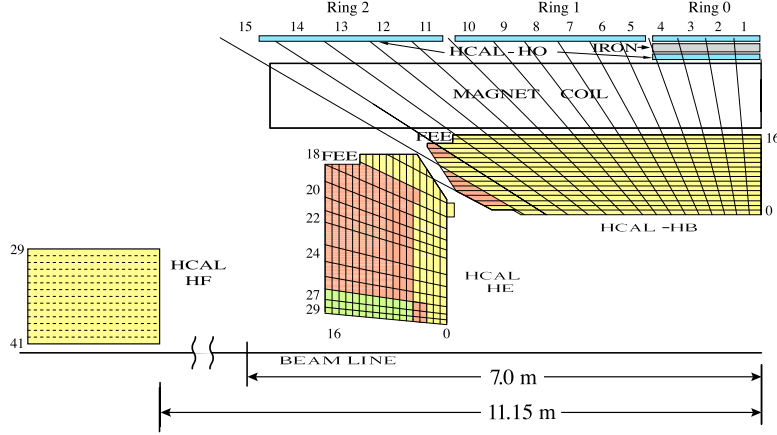
The energy resolution of the ECAL can be parametrized as:

$$\frac{\sigma}{E} = \frac{2.8\%}{\sqrt{E}} \oplus \frac{12\%}{E} \oplus 0.3\%. \quad (4.4)$$

Here, the first term is called the stochastic term because it describes statistical fluctuations in the showering process and in the detected number of photons in the photomultipliers. This term was measured in electron test beams to be about 2.8% in the barrel part for electrons of 20 GeV to 250 GeV. The second term is the noise term and describes electronic noise in the readout chain. At the CMS experiment it was measured to be about 12%. The third term is called the constant term and describes contributions that are not energy dependent. These can arise from the detector geometry, temperature gradients, aging or radiation damage. It is measured to be 0.3% [92, 93].

### 4.2.3 Hadronic calorimeter

The CMS hadron calorimeter (HCAL) [81] is used to measure the energy of neutral and charged hadrons. The HCAL is a sampling calorimeter consisting of alternating layers of brass absorbers and plastic scintillator. Incoming particles interact with the dense brass material creating hadronic showers. When the shower reaches the scintillator, light is created, which is readout by photodetectors.



**Figure 4.7: Layout of the CMS HCAL.** Shown is the HCAL barrel (HB), end-cap detector (HE), outer (HO) and forward detector (HF) [96].

The layout of the HCAL is shown in figure 4.7. The barrel (HB) part of the detector covers the pseudorapidity range of  $|\eta| < 1.3$ . The HCAL has to fit between the ECAL and the solenoid, limiting the material budget for the HB. The effective thickness of the HB depends on the polar angle. It contains between 5.82 (at  $\eta = 0$ ) and 10.6 (at  $|\eta| = 1.3$ ) nuclear interaction lengths ( $\lambda_l$ ). The thickness of the HB in the barrel region is too low to contain the entire hadron shower. Therefore, it is extended by an outer (HO) calorimeter, which sits outside the solenoid. It is placed as the first layer before the muon system inside each ring of the iron yoke. At  $\eta = 0$ , where the effective thickness of the HB is lowest, two HO layers are placed around a 19.5 cm thick tail catcher iron piece. This extends the thickness of the HCAL in the barrel region to a minimum of  $11.8 \lambda_l$ .

The pseudorapidity range between  $1.3 < |\eta| < 3.0$  is covered by the HCAL endcap (HE) detector. To provide maximum possible coverage, a Cherenkov-based forward calorimeter (HF) is placed 11.2 m away from the interaction point, which extends the pseudorapidity coverage to  $|\eta| = 5.2$ . A Cherenkov detector was chosen due to the high flux of particles in the forward region, requiring a radiation tolerant design. The HF consists of steel absorbers and quartz fibres as active medium.

The combined energy resolution of the ECAL and HCAL in the barrel part was measured in pion test beams to be [95]:

$$\frac{\sigma}{E} = \frac{110\%}{\sqrt{E[\text{GeV}]}} \oplus 9\%, \quad (4.5)$$

#### 4.2.4 Solenoid

The superconducting solenoid bends charged particles, resulting in curved tracks, which allows to measure their transverse momentum and to identify the sign of their charge. The solenoid has a diameter of 6 m and is 12.5 m long with a cold mass of 220 t. When the magnet is turned on it creates a magnetic field of 3.8 T and stores an energy of 2.6 GJ. A 12 500 t iron return yoke is used to close the field lines. Outside the solenoid the magnetic field has a strength of 2 T in opposite direction, covering the muon system. The coil of the solenoid is cooled to 4.7 K using 6000 l of liquid helium. Cooling the magnet down from room temperature takes three weeks [81].

#### 4.2.5 Muon system

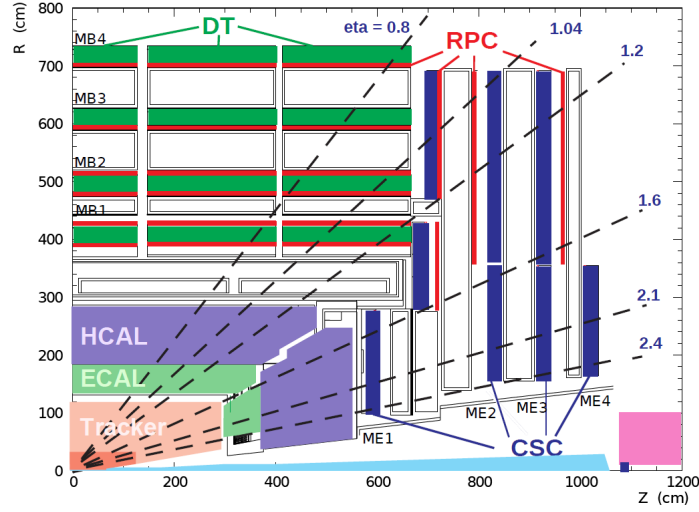
Muons generally traverse through the whole detector without decaying or being absorbed, as high energy muons interact with matter mostly through ionization and are minimal ionizing at the typical energy range at the LHC. Therefore, the muon system is the most outer part of the CMS detector. Muons can be unambiguously identified, because other particles decay or are stopped within the detector. The CMS muon system consists of three gaseous tracking detectors. All three are arranged in four layers between the three layers of the iron yoke. The muon system has a pseudorapidity coverage up to  $|\eta| < 2.4$ .

In general, gaseous detectors consist of two electrodes that are separated by a gas. Ionizing particles with enough energy may ionize the gas between the electrodes. This results in a current drift towards the electrodes where an electric signal is created.

Figure 4.8 shows the layout of the muon system. In the barrel region up to  $|\eta| < 1.2$  drift tubes (DTs) are used. The DTs are filled with a mixture of Ar and CO<sub>2</sub> gas and have an anode wire in the middle. Some of the tubes are aligned parallel to the  $z$  axis, measuring the  $(r, \phi)$  coordinates and some orthogonal to the  $z$  axis, measuring the  $(r, z)$  coordinates. In the pseudorapidity region of  $0.9 < |\eta| < 2.4$  cathode strip chambers (CSCs) are used. The CSCs are multiwire proportional chambers filled with a mixture of argon, CO<sub>2</sub> and CF<sub>4</sub>. They have several anode wires and cathode strips arranged orthogonal to each other. This allows to measure 2D coordinates from each hit. The last component of the muon system are the resistive plate chambers (RPCs). They cover the rapidity range up to  $|\eta| < 1.6$ . The RPCs consist of two parallel plates with high resistivity. Their main features is their excellent time resolution of about 3 ns or better, therefore they are mostly used for triggering [97].

#### 4.2.6 Data acquisition, trigger and computing

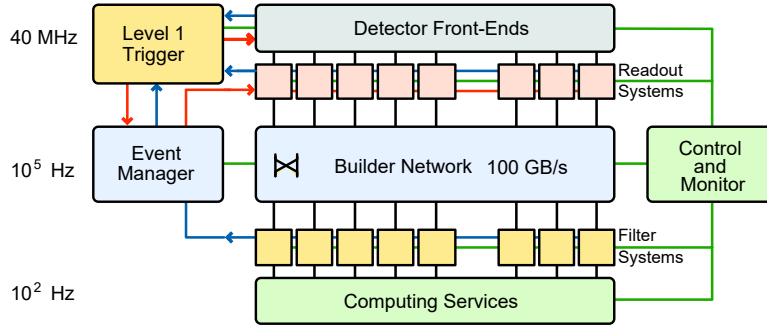
The bunch crossing frequency at the LHC is 40 MHz with about 35 collisions taking place on average per bunch crossing in Run 2. The event size per crossing is about



**Figure 4.8: Layout of the CMS muon system.** Drift tubes (DT) are used in the barrel region and cathode strip chambers (CSC) in the 2 end-cap regions. Resistive plate chambers (RPC) are used in both the barrel and end-cap regions. [98].

1 MB, resulting in a data rate of several TB/s. The amount of data that can be read out is technically limited by the bandwidth of the Data Acquisition system (DAQ), which is about 100 GB/s. Therefore, fast hardware trigger (Level 1 trigger), which are directly incorporated into the detector, are used to decrease the event rate to around 100 kHz. A second technical limitation is the speed at which data can be written to tape, which is about 1 kHz. A system of software based High level Trigger (HLT) that runs on a computing farm is used to reduce the event rate of the Level 1 trigger by about a factor of 100. Figure 4.9 shows an schematic overview over the CMS DAQ and trigger system.

Besides the above mentioned technical limitations most events are dominated by soft multi-jet scattering processes, which are not of much interest to most physics analysis. The triggers are therefore designed to select events that contain physically interesting events and discard others. The Level 1 trigger builds for this purpose simple objects using information from the calorimeters and the muon system. Then cuts on kinematic quantities of these objects are applied. Tracking information is not used in the Level 1 triggers due to the computational complexity of the track reconstruction. The HLT uses information from the entire detector to perform a full event reconstruction using an optimized version of the same software that is also used during offline reconstruction. This allows to apply higher quality selection criteria and to trigger on more complex quantities. The trigger algorithms are organized in so-called paths. Each path consists of several modules that are run in a predefined order and an event needs to pass all filters in a path in order to be written to tape. Each module performs a distinct task and is for efficient computation shared between multiple trigger paths that require the same task.



**Figure 4.9: Overview over the CMS DAQ and Trigger system.** From the Detector Front-Ends and the Readout System data is sent to the hardware-based Level 1 trigger. The Level 1 trigger decides which events to read out. The Builder Network then combines event fragments belonging to the same Level 1 trigger decision. Software based High Level Trigger, running on a computing farm decide which events should be written to tape. The HLT was originally designed for an output rate of about  $10^2$  Hz, as depicted in this figure. However, the output rate has steadily increased over the runtime of the LHC and is now larger than  $10^3$  Hz [81].

Events passing the HLTs are distributed to various computing centers for further processing and storage. The computing centers are organized in the so-called Worldwide LHC Computing Grid (WLCG) [99,100]. The WLCG has a hierarchical structure with several Tiers. The Tier-0 computing center is located at CERN at which all raw data is stored on tape. Also first event reconstructions are done at the Tier-0. From the Tier-0 data gets further distributed to 13 Tier-1 computing centers. One of them is GridKa [101], which is hosted at the Steinbuch Center for Computing (SCC) [102] at the Karlsruher Institute of Technology (KIT) [103]. The Tier-1 centers take care of further data processing, storing of processed data and each of them creates a backup of parts of the raw data. They also distribute the data to the over 150 Tier-2 computing centers worldwide. Tier 2 centers store relevant data samples and take care of production of simulated data and other specialized tasks.

#### 4.2.7 The CMS detector at the High Luminosity LHC

The High Luminosity LHC (HL-LHC) [104] is an upgrade to the LHC which is scheduled for 2027. At the HL-LHC the instantaneous luminosity will increase to about  $10^{35} \text{ cm}^{-2} \text{ s}^{-1}$ , which is a tenfold increase compared to the  $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  achieved in 2018. This increase in instantaneous luminosity will also result in an increased maximum number of 140 or even up to 200 pileup events per bunch crossing in the highest performance scenarios. The increase in instantaneous luminosity is achieved in two ways. The focus of the beam at the interaction point is adjusted during each run, such that the instantaneous luminosity remains constant for a longer duration. In addition crab cavities are used to rotate bunches before the collision such that the colliding bunches have a larger overlapping area. The upgrade will increase the sensitivity to rare physi-



cal processes because of the large amount of data that will be recorded. To guarantee good operation at the HL-LHC, components of the CMS detector need to be upgraded in order to be more radiation tolerant, better at suppressing pileup and at handling high collision rates. In the following a short description of the planned upgrades to the CMS detector is given. The upgraded CMS detector is also called Phase-2 detector, while the current detector is called Phase-1 detector.

The rate of the L1 triggers is planned to be increased to 750 kHz with a latency of only 12.5  $\mu$ s. This requires an update of most of the electronic systems installed at the detectors and also of the DAQ system. The L1 trigger will for the first time include information from the tracker and high-granularity calorimeter information [105].

The pixel detector will be upgraded with 6 times smaller pixels and its coverage is increased up to  $|\eta| \leq 4$  [106]. The geometry of the strip tracking detector is redesigned using significantly less material they will also get  $p_T$  modules that allow the usage of the transverse momentum information already at L1 trigger level.  $p_T$  modules consist of two layers that can quickly compute an estimate of a track's curvature from the offset of the hits between the two layers.

Between the tracker and the calorimeter a new timing detector is installed [107]. It measures minimum ionizing particle (MIP) and allows timing resolutions of about 30 ps. In the barrel region up to  $|\eta| < 1.5$  it will consist of crystal scintillators read out by silicon photo multipliers and in the endcap region up to  $|\eta| < 3$  Low Gain Avalanche Diodes will be used. By using the timing information, collisions can be sliced into time windows of about 30 ps. At the HL-LHC pileup collisions will occur roughly every 180 ps. Therefore, the pileup in each window would be about as large as the pileup in Phase-1. The addition of timing also allows for the reconstruction of 4D vertices (3D space + 1D time), which could improve vertex and track reconstruction.

In the HCAL barrel (HB) detector the photodiodes will be replaced by silicon photo-multipliers, which have a higher photon detection efficiency [108]. The endcap calorimeters are completely redesigned as the current design will not be able to withstand the radiation damage at the HL-LHC. Both the ECAL endcap (EE) and HCAL endcap (HE) detectors are replaced by a high granularity calorimeter (HGCAL) [109]. The HGCAL will consist of 28 longitudinal segments in the ECAL part and 24 segments in the HCAL part. The ECAL and a large part of the HCAL will use silicon cells with a size between 0.5 cm<sup>2</sup> and 1 cm<sup>2</sup> and the rest of the HCAL will consist of highly-segmented plastic scintillators with a size of 4 cm<sup>2</sup> to 30 cm<sup>2</sup>. The muon system was found to work well at HL-LHC conditions, therefore mostly the electronics will be upgraded [110].



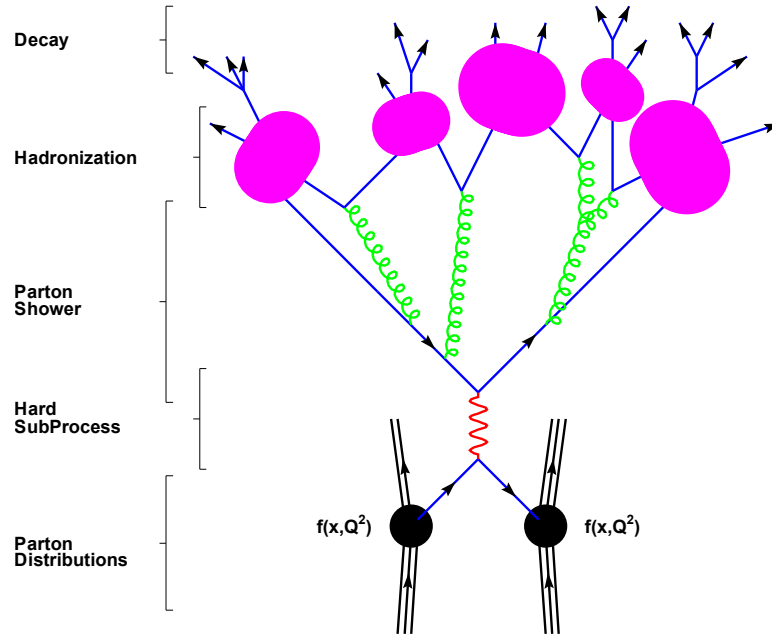
## 5 Event simulation

Accurate simulations of proton-proton collisions are crucial for any physics analysis. The simulations provide predictions for the rates and kinematical variables of physical processes. These predictions are used to model the signal and background processes, which are considered in the analysis. In the search for new physics, simulations are used to predict the kinematical variables of new physical processes.

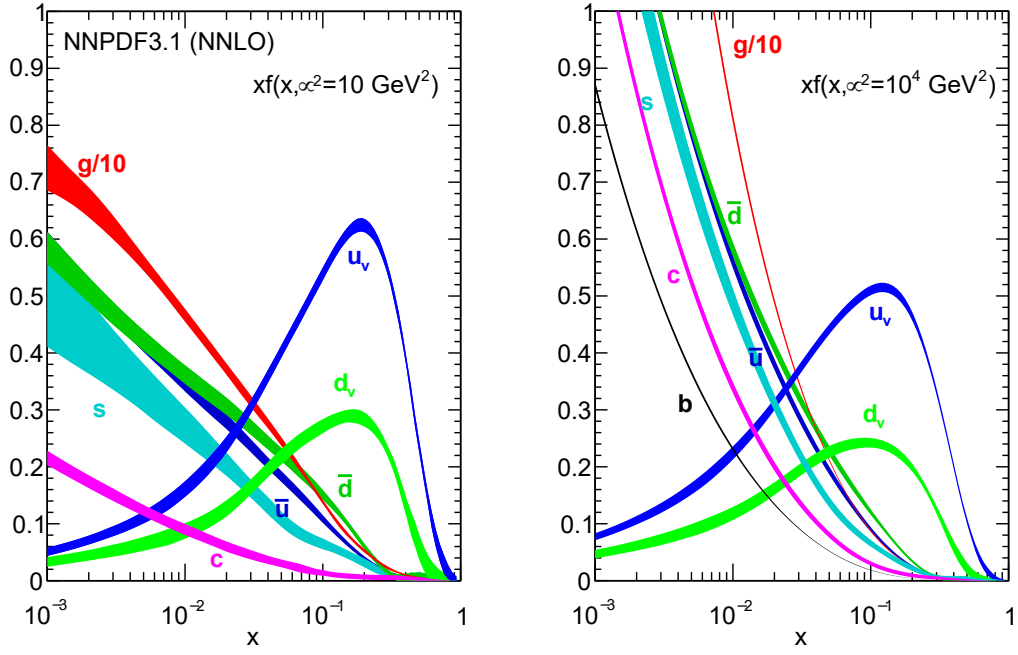
Simply simulating billions of proton-proton collisions would not result in a sufficient amount of events for rare physical processes, as most collisions at the LHC consist of only soft QCD interactions. Hence, simulations assume a specific physical process taking place in the hard interaction of the collision. The first step of the simulation is to generate all particles that are created in the hard process as well as all secondary particles arising from subsequent showers and hadronization. Next the interactions of the particles in the detector and the electric signals that these interactions create are simulated.

### 5.1 Event generation

The generation of simulated events commonly follows a workflow consisting of three distinct steps. Namely, the simulation of the hard subprocess, the parton showering in which additional radiation is simulated, and the hadronization step in which partons are confined into color-neutral bound states. Specific software packages are used to carry out these steps. The packages MADGRAPH5\_AMC@NLO [112] and POWHEG [51–53] are capable of simulating the hard subprocess. Whereas, the general-purpose event generators PYTHIA [113] and HERWIG++ [114] can be used to perform all three steps. Often, MADGRAPH5\_AMC@NLO or POWHEG are used to simulate the hard subprocess and their output is stored in the Les Houches events format (LHE) [115]. This format can be used as input to PYTHIA or HERWIG++ that subsequently perform the parton showering and hadronization steps. In figure 5.1 the different steps are depicted. In the following a description of the different event generation steps is given, following [116].



**Figure 5.1: The basic steps of event generation in a proton-proton collision.** The partons in the protons are described by parton distribution functions and their interaction is called the hard sub-process. In the parton shower step the emission of additional radiation is simulated. In the hadronization step all partons are confined into color-neutral hadrons. Unstable hadrons decay into long-lived particles. The figure is based on [111].



**Figure 5.2:** The Neural Network parton distribution functions (NNPDF3.1) at NNLO. The PDFs for gluons and light-flavored quarks are shown. They are evaluated at  $\mu^2 = 10 \text{ GeV}^2$  (left) and at  $\mu^2 = 10^4 \text{ GeV}^2$  (right) [117].

### 5.1.1 Hard process

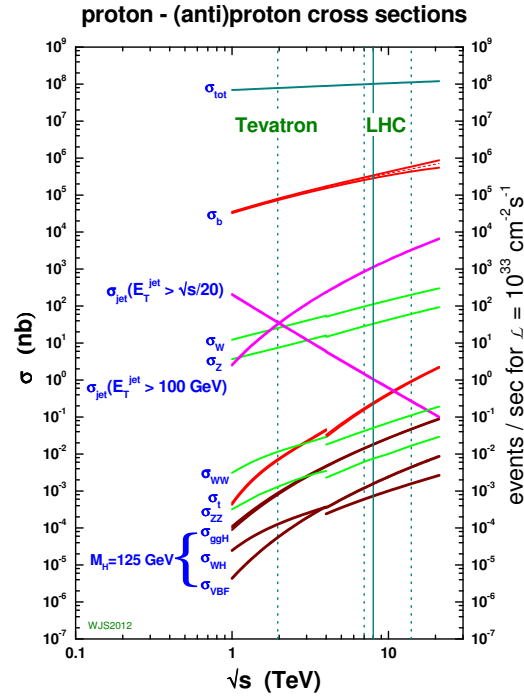
In each bunch crossing and in each proton-proton collision several interactions take place simultaneously. The interaction with the highest momentum transfer is called the hard process. The cross section  $\sigma$  for a  $ab \rightarrow n$  process in a hadron collision can be written as [116, 118]:

$$\sigma = \sum_{a,b} \int_0^1 dx_a dx_b \int d\Phi_n f_a^{h_1}(x_a, \mu_F) f_b^{h_2}(x_b, \mu_F) \frac{1}{2x_a x_b s} |\mathcal{M}_{ab \rightarrow n}|^2(\Phi_n; \mu_F, \alpha(\mu_R)). \quad (5.1)$$

Here, the functions  $f_a^h(x, \mu)$  are the parton-distribution-functions (PDFs) that describe the probability of finding a parton  $a$  with a momentum fraction  $x$  inside its parent hadron  $h$  if the interaction takes place at an energy scale  $\mu$ . The PDFs cannot be calculated from first principle, but must be determined from data for a given energy scale  $\mu$ . Once the PDFs are known for one energy scale, they can be calculated for other energy scales using the DGLAP equations [119–121]. The PDF sets used in this thesis are the NNPDF3.0 [122] and NNPDF3.1 [117] sets. The parton densities for the NNPDF3.1 set are shown for two different scales in figure 5.2.

The term  $\frac{1}{2x_a x_b s}$  in equation 5.1 is the incoming parton flux and  $s$  is the center-of-mass energy squared.  $\mathcal{M}_{ab \rightarrow n}$  corresponds to the matrix element (ME) for the process  $ab \rightarrow n$ , which depends on the renormalization scale  $\mu_R$ , the factorization scale  $\mu_F$  and on the final-state phase space  $\Phi_n$  that determines the momenta of the final-state partons. The renormalization scale is introduced to cope with ultraviolet divergencies that arise due to large momenta in the loops of the Feynman diagrams corresponding to the amplitude. The coupling constant becomes a function of the renormalization scale. The factorization scale is introduced due to infrared divergences that arise from massless particles radiating other massless particles. The PDFs are a function of the factorization scale. The values of  $\mu_F$  and  $\mu_R$  are chosen based on experience depending on the physical process being simulated. Also, the combination with the parton shower restricts the range of allowed values. The ME can be written as a sum of Feynman diagrams. As processes at the LHC take place at energy scales of 10 GeV to TeV and QCD is asymptotically free, the MEs can be calculated using perturbation theory. The first element in the perturbation series is called leading order (LO) and the second Next-to-leading order (NLO) and so forth. The cross section of different SM processes as a function of the collision energy at the LHC and Tevatron can be seen in figure 5.3.

PYTHIA and HERWIG++ can be used to calculate the LO MEs for a wide range of SM processes and some BSM scenarios. MADGRAPH5\_AMC@NLO can be used to calculate LO and NLO MEs for all processes that can be described by the Universal FeynRules Output (UFO) format [124]. The integral over the phase space in equation 5.1 is carried out using Monte Carlo (MC) integration techniques. Their main advantage compared to deterministic integration methods is that they converge faster for high dimensional phase spaces.



**Figure 5.3: Standard Model cross sections as a function of collider energy.** The total proton-(anti)proton cross section  $\sigma_{\text{tot}}$  and the cross sections for different SM processes as a function of the center-of-mass energy are shown for the LHC and the Tevatron. The figure is taken from reference [123].

### 5.1.2 Parton shower

In an interaction additional radiation occurs as gluons are emitted. The gluons then split into further secondary particles creating a parton shower. The parton showers are modeled using perturbative QCD and are implemented as a series of probabilistic  $1 \rightarrow 2$  splittings. The probability for a particle not to split between two energy scales is given by the Sudakov form factors [125]. The evolution of the parton shower is implemented using Monte Carlo techniques. At each step a random number is generated and compared with the Sudakov form factor to determine if the shower evolution should be terminated. The splitting is done until an energy scale of about 1 GeV is reached at which hadronization begins due to confinement.

The radiation can stem either from the initial-state particles of the hard process and is then called initial-state-radiation (ISR) or from the final-state particles of the hard process and is then called final-state radiation (FSR). The FSR is modeled as a series of forward splittings as described above. The ISR however, is implemented as a backward evolution. The particles in the hard process are dressed with additional ISR and the shower is propagated backwards from particles at a higher energy scale and lower momentum fraction down to the partons in the protons at a lower energy scale but with higher momentum fraction.

The implementation of parton showers differs slightly in HERWIG++ and PYTHIA with respect to the order in which the splittings are done. In PYTHIA the emissions with the highest transverse momentum are split first, producing a  $p_T$ -ordered dipole shower. Whereas, HERWIG++ performs an angular-ordered parton showering where splittings with the largest angles are performed first [116].

Radiation can also occur in the hard interaction if the matrix-element is evaluated at NLO or higher. These radiations at matrix-element level may interfere with radiation at the parton shower step. A careful treatment is therefore required in order to avoid double counting. At the same time the precision can be increased by combining the two radiation sources. This is because, with the parton shower soft and collinear emissions can be well described, whereas hard and wide-angle emissions are not. With the fixed-order ME calculations done in the hard process step it is the other way round. There are two general approaches to combine the two: matching and merging. In the matching approach higher-order corrections are provided by the parton shower. In order to avoid double counting they have to be subtracted from the ME. Both MADGRAPH5\_AMC@NLO [126] and POWHEG use methods to reweight the matrix-element calculated at NLO to correct for the first emission in the parton shower. MADGRAPH5\_AMC@NLO introduces negative event weights to account for double counting between the hard process and the parton shower, which requires to generate more events. POWHEG adds a LO-NLO correction factor, but requires a  $p_T$  ordered parton shower. In the case of multi-jet topologies the merging procedure is employed. In this approach, partons produced above a merging scale are generated with a corresponding higher-order ME and partons produced below the merging scale are provided by the parton shower. Merging can be used in POWHEG by combining



it with the CKKW [127] merging algorithm. The FxFx [128] merging scheme can be used to implement merging in MADGRAPH5\_AMC@NLO.

### 5.1.3 Hadronization

Once the energy of the partons is of order  $\Lambda_{\text{QCD}} \approx 200 \text{ MeV}$  they start to form color-neutral bound states due to the confinement of the QCD. As QCD is non-perturbative at low energy scales, one has to rely on phenomenological models to describe the process of hadronization. There are two main models used: the string or Lund model [129,130] and the cluster model [131]. The string model is based on the confinement property of QCD, which leads to a linearly rising potential between two quarks. The color field-lines between two quarks are depicted as strings or tubes. If the quarks are separated, the strings are stretched until they break and a new quark-antiquark pair is created in between them. The cluster model is based on the preconfinement [131] property of parton showers. This property allows to find clusters of partons which form color singlets and have a computable mass distribution. Clusterization starts by splitting of gluons into quark-antiquark pairs. Color-connected pairs then form clusters. Once such a cluster is built, it subsequently decays into hadrons or possibly lighter clusters that further decay into hadrons. The string model usually provides a slightly better description than the cluster model but has more tunable parameters that need to be determined from data. In HERWIG++ the hadronization is implemented via the cluster model and in PYTHIA via the string model [116].

### 5.1.4 Underlying event

The term underlying event is used to describe physical process that happen during a proton-proton collision, which are not part of the hard interaction or of ISR and FSR. The underlying event consists of beam-beam remnant and particles arising from multiple-parton-interactions. The modeling of the underlying event is controlled by tunes. In this thesis the CP5 [132] and CUETP8M1 [133] tune are used.

## 5.2 Detector simulation

After the four-vectors of the final-state particles have been generated by the Monte Carlo generators, a full detector simulation is performed. The interaction of the particles with the detector material is simulated with GEANT4 [134,135]. A detailed model of the CMS detector with accurate geometry and materials is built into GEANT4. The software traces the generated particles stepwise through the detector and simulates their interaction with the detector material using MC techniques. GEANT4 can simulate a wide range of electromagnetic, hadronic, and optical processes over a large energy range from a few hundred eV to TeV.

In order to simulate a pileup profile, special minimum-bias events [136], consisting mostly of soft QCD interactions, are simulated using PYTHIA. They undergo an event generation and GEANT4 simulation and are then mixed with the events from the hard process. The detector responses simulated with GEANT4 are then digitalized and random electric noise is added using dedicated packages of the CMS software. After this step the output of the simulation matches the electric signals measured during real data-taking. A more detailed description of the detector simulation in CMS can also be found in reference [137]. In addition to the full simulation chain described above the CMS collaboration has also developed a technique called fast simulation. Fast simulation does not rely on a full GEANT4 simulation but instead uses a simplified representation of the detector and analytic models to significantly speed up the simulation [138].

## 6 Event reconstruction and object definition

In a proton-proton collision many particles are created. These particles are not directly visible; only the signals they create as they interact with the detector material are measurable. In order to figure out which kind of particles were created in a collision and what kinematic properties they had, one must interpret the measured data. For this purpose sophisticated algorithms exist, which are described in this section.

CMS uses the particle flow (PF) algorithm to reconstruct events. The PF algorithm combines information from all subdetectors in order to achieve the best possible reconstruction of physics objects. The PF algorithm works, by first reconstructing basic PF elements based on information from the different sub-detectors. The PF elements include tracks, vertices, calorimeter clusters, and signals in the muon chambers. Then these PF elements are combined by a linking algorithm into PF blocks. The links are created by extrapolating the reconstructed tracks into the calorimeters and matching them to ECAL and HCAL clusters in the  $(\eta, \phi)$  plane. Similarly, cluster-to-cluster links between HCAL clusters and ECAL clusters and between ECAL clusters and preshower clusters are established. At the last step trajectories originating from the central tracker are combined with information from the muon system. Once the links are established PF candidates of physics objects are identified and reconstructed. The algorithm iterates over the possible PF objects in a fixed order, reconstructs them and then removes their PF elements from the PF block. Objects are reconstructed in the following order: First muons are reconstructed, as described in section 6.3, then electrons (see section 6.4) and isolated photons, and lastly non-isolated photons and hadrons as explained in section 6.5. After all PF blocks are processed, a post-processing step is performed to reduce the risk of particle misidentification. In the following sections the reconstruction of the different PF elements and PF objects are described in more detail. The descriptions follows reference [92].

### 6.1 Tracks and vertices

The reconstruction of tracks is performed iteratively with a combinatorial track finder algorithm, which is an adaption of the combinatorial Kalman Filter [139–143] (KF). The first step in the track reconstruction is the seeding, which provides initial track candidates. Seeds are hits in the tracking detector that can be combined to a charged particle track. Thus, either three hits or two hits and the position of the primary vertex

are necessary. The next task is to find additional hits matching the flight path. For this purpose, the initial seeds are extrapolated outwards and a KF is used for pattern recognition. In order to find the best estimates of track parameters, the tracks are then fitted using a KF. Finally, quality flags are set on the tracks and the one that fail necessary quality criteria are discarded. These steps are repeated in ten consecutive iterations, each time using different seeding and quality criteria. High quality tracks are masked after each iteration in order to reduce the complexity. This iterative approach allows to apply more complex criteria in seeding, filtering, and fitting and therefore significantly improves the tracking efficiency. For tracks with  $p_T > 1 \text{ GeV}$  the tracking efficiency is close to 100 % for muons and above 80 % for charged pions and electrons [139, 144, 145].

Once tracks are reconstructed, the primary vertex (PV) and any additional vertices arising from pileup are reconstructed. First tracks are selected that originate from the primary interaction region. The tracks are then clustered using a deterministic annealing algorithm [146]. Candidates containing at least two tracks are then fitted using the adaptive vertex fitter algorithm [147], in order to determine the best vertex parameters. For tracks with  $p_T > 1 \text{ GeV}$  primary vertices can be reconstructed with a resolution below  $100 \mu\text{m}$  and the average resolution for all tracks in the barrel region is around  $50 \mu\text{m}$  [139, 144, 145].

## 6.2 Calorimeter clusters

Energy deposits in the calorimeter cells are clustered in order to improve the reconstruction of the energy and direction of neutral hadrons, photons and charged hadrons, to separate neutral particles from charged hadrons, and to improve the reconstruction of electrons and bremsstrahlung. The clustering is performed separately in each sub-detector but following the same approach everywhere. The first step in the clustering is to find seed cells. The seeds must have an energy deposit larger than a seed threshold and larger than the surrounding cells. From the seeds topological clusters are grown by adding neighboring cells with an energy deposit larger than 2 times the noise level. Within the topological cluster, clusters are reconstructed using an iterative expectation-maximization algorithm based on a Gaussian-mixture model. It is assumed that the energy deposits observed in the cells of the topological cluster arise from  $N$  Gaussian energy deposits, where  $N$  is equal to the number of seeds. The parameters of this Gaussian energy deposits are determined in a maximum-likelihood fit.

As electrons emit energy in form of bremsstrahlung before reaching the ECAL, multiple ECAL clusters are combined into superclusters in order to collect the entire energy. The superclusters have a small coverage in  $\eta$  and a larger coverage in  $\phi$  in order to account for the bending of the electrons in the magnetic field.

## 6.3 Muon reconstruction and identification

Muons are minimum ionizing particles and are therefore the only particles traversing the whole detector. Thus, they are the only particles leaving a signal in the muon detector, which allows to uniquely identify them. There are three different types of muons being reconstructed, depending on how the information from the inner tracker and the muon chambers is combined.

**Standalone muons**, are reconstructed using only information provided by the muon system.

**Global muons**, are reconstructed by matching a track from a standalone muon with a track from the inner tracker with a Kalman-filter. The matching is performed inwards.

**Tracker muons**, are reconstructed by extrapolating a track from the inner tracker outwards to the muon system. At least one segment in the muon system has to match the track for the muon to be successfully reconstructed.

About 99 % of all muons produced within the detector acceptance are reconstructed as either global muon or tracker muon or both. Muons below 10 GeV are more often reconstructed as tracker muons, because of multiple scattering in the iron return yoke, which results in fewer segments in the muon system being hit by the muon. The reconstruction of global muons improves the momentum resolution for large transverse momenta  $p_T \geq 200$  GeV as compared to only using tracker muons.

Different identification criteria are used for muons. Loose muons must only be reconstructed by the PF algorithm as either global or tracker muons. However, muons can arise from sources like decays in flight ( $\pi \rightarrow \mu\nu$ ), hadron shower remnants, which penetrate into the muon system, inaccurate track-to-segment matches or cosmic muons. To suppress such contributions tight muon identification criteria are defined. These are listed in table 6.1

The efficiency of the loose and tight muon identification criteria can be measured with a tag-and-probe method in  $Z \rightarrow \mu\mu$  events. For loose muons the efficiency is  $\geq 99$  % for tight muons the efficiency is between 95 % and 99 %. The transverse momentum resolution is between 1 % and 6 % for muons below 100 GeV and less than 10 % for muons below 1 TeV [98, 149].

To suppress hadrons and muons in jets additionally the relative isolation  $I_{\text{rel},\mu}$  of the muon is used. The relative isolation of the muon corresponds to the amount of energy deposited from other particles in a cone of  $\Delta R = 0.4$  around the muon relative to the muon momentum. It is defined as:

**Table 6.1: Muon identification criteria.** Shown are the criteria for a muon to be qualified as either a loose or a tight muon. The criteria are based on the recommendations from CMS Muon Physics Object Group within the CMS Collaboration [148].

Variable	Loose muons	Tight muons
Is a global muon	true	true
Is a tracker muon or global muon	true	-
$\chi^2/\text{ndof}$ of the global-muon fit	-	$< 10$
Number muon chamber hits	-	$\geq 1$
Number of muon stations with segments	-	$\geq 2$
Transverse impact parameter $d_{xy}$	-	$< 2 \text{ mm}$
Longitudinal impact parameter $d_z$	-	$< 5 \text{ mm}$
Number of pixel hits	-	$\geq 1$
Number of tracker layers with hits	-	$\geq 6$

$$I_{\text{rel},\mu} = \frac{1}{p_{T,\mu}} \left[ \sum p_T^{\text{ch. had.}} + \max\left(0, \sum p_T^{\text{neutr. had.}} + \sum p_T^\gamma - \Delta\beta \sum p_T^{\text{ch. had., PU}}\right) \right]. \quad (6.1)$$

Here,  $p_{T,\mu}$  is the muon transverse momentum,  $p_T^{\text{ch. had.}}$ ,  $p_T^{\text{neutr. had.}}$ ,  $p_T^\gamma$  and  $p_T^{\text{ch. had., PU}}$  are the transverse momenta of charged hadrons, neutral hadrons, photons and charged hadrons from pileup, respectively.  $\Delta\beta = 0.5$  is the estimated fraction of neutral to charged hadrons from pileup.

## 6.4 Electron reconstruction and identification

Electrons deposit energy in the ECAL and leave a visible track in the inner tracker, allowing them to be reconstructed by combining information from the tracker and the calorimeters. Electrons can emit bremsstrahlung photons when interacting with the material in the tracker. The emitted bremsstrahlung makes track finding for electrons relatively challenging compared to muons. The track finding using an iterative KF generally leads to large  $\chi^2$  values, as the electron momentum changes due to soft photon emissions. Therefore, tracks are selected based on number of hits and the fit  $\chi^2$  value. The selected tracks are fit again with a Gaussian-sum filter (GSF) [150]. The GSF accounts for sudden energy losses along the trajectory and is therefore better suited for electrons. A boosted-decision-tree (BDT) classifier using information from both the GSF and KF track fit is trained to identify good tracks. Its output is used as a quality criteria for each track.

**Table 6.2: Selection criteria for tight electrons and veto electrons.**

Different selection criteria are used in the ECAL barrel ( $|\eta_{\text{SC}}| \leq 1.479$ ) and ECAL endcap ( $|\eta_{\text{SC}}| > 1.479$ ). Here,  $\eta_{\text{SC}}$  is the pseudorapidity of the ECAL supercluster (SC). The selection criteria follow the recommendations from the EGamma Physics Object Group of the CMS collaboration [151].

barrel	tight electron	veto electron
$\sigma_{\eta\eta}$	$< 0.0104$	$< 0.0126$
$ \Delta\eta(\text{SC}, \text{track}) $	$< 0.00255$	$< 0.00463$
$ \Delta\phi(\text{SC}, \text{track}) $	$< 0.022$	$< 0.148$
H/E	$< 0.026 + 1.15/E_{\text{SC}} + 0.0324\rho/E_{\text{SC}}$	$< 0.05 + 1.16/E_{\text{SC}} + 0.0324\rho/E_{\text{SC}}$
Electron isolation	$< 0.0287 + 0.506/p_{\text{T}}$	$< 0.198 + 0.506/p_{\text{T}}$
$ 1/E_{\text{SC}} - 1/p_{\text{track}}  \text{ (GeV}^{-1}\text{)}$	$< 0.159$	$< 0.209$
Expected missing inner hits	$\leq 1$	$\leq 2$
Rejection of converted photons	yes	yes
endcap	tight electron	veto electron
$\sigma_{\eta\eta}$	$< 0.0353$	$< 0.0457$
$ \Delta\eta(\text{SC}, \text{track}) $	$< 0.00501$	$< 0.00814$
$ \Delta\phi(\text{SC}, \text{track}) $	$< 0.0236$	$< 0.19$
H/E	$< 0.0188 + 2.06/E_{\text{SC}} + 0.183\rho/E_{\text{SC}}$	$< 0.05 + 2.54/E_{\text{SC}} + 0.183\rho/E_{\text{SC}}$
Electron isolation	$< 0.0445 + 0.963/p_{\text{T}}$	$< 0.203 + 0.963/p_{\text{T}}$
$ 1/E_{\text{SC}} - 1/p_{\text{track}}  \text{ (GeV}^{-1}\text{)}$	$< 0.0197$	$< 0.132$
Expected missing inner hits	$\leq 1$	$\leq 3$
Rejection of converted photons	yes	yes

If the corresponding ECAL supercluster is not linked to more than two tracks, the electrons are seeded from GSF tracks. Otherwise, they are seeded from ECAL superclusters. To ensure that the energy is contained in the ECAL, the energy measured in close HCAL cells must be less than 10% of the energy in the ECAL supercluster, for ECAL-based electron candidates and photon candidates.

Depending on different quality criteria of the reconstruction, electrons are reconstructed as veto electrons or tight electrons. The selection criteria necessary for an electron to qualify as either a veto or a tight electron are listed in table 6.2. The selection criteria depend on the following variables:

$\sigma_{\eta\eta} = \frac{|\sum(\eta - \bar{\eta})^2 w_i|}{\sum w_i}$ : This is the lateral extension of the shower along the  $\eta$  direction.  $w_i$  is a weight, which depends on the energy [152]. The weighted average is computed by summing over the  $5 \times 5$  matrix of crystals around the highest  $E_{\text{T}}$  crystal of the supercluster.

$\Delta\eta(\text{SC}, \text{track})$  and  $\Delta\phi(\text{SC}, \text{track})$ : are the difference in pseudorapidity and azimuthal angle between the supercluster position and the electron's track.

**H/E**: This is the ratio of the energy deposited in the HCAL to the energy in the supercluster.

**Electron isolation:** This is the amount of energy from other particles deposited in a cone of  $\Delta R = 0.3$  around the electron relative to the electron's transverse momentum.

$1/E_{\text{SC}} - 1/p_{\text{track}}$ : Here  $E_{\text{SC}}$  is the energy deposited in the supercluster and  $p_{\text{track}}$  the momentum of the track.

**Expected missing inner hits:** number of tracker layers with no hits. This is used to reject electrons from photon conversions.

**Rejection of converted photons:** Electrons can be produced in conversions of photons in the tracker material. Electrons from photon conversion can be rejected by requiring no missing hits in the inner parts of the tracker [152].

The momentum resolution for electrons is measured to be better than 5 %. The reconstruction efficiency for electrons was determined using a tag and probe method in  $Z \rightarrow e^- e^+$  to be larger than 85 % [152].

## 6.5 Hadrons and photons reconstruction

After muons and electrons are reconstructed, isolated photons are reconstructed. Like electrons, photons deposit energy in the ECAL, but they do not leave a visible track in the inner tracker. Isolated photons are reconstructed from an ECAL supercluster with  $E_T > 10 \text{ GeV}$ .

After muons, electrons, and isolated photons are reconstructed and removed from the PF blocks, the remaining particles include charged hadrons ( $\pi^\pm$ ,  $K^\pm$ , protons), neutral hadrons ( $K_L^0$  or neutrons), and non-isolated photons stemming from  $\pi^0$  decays, and rarely additional muons from early decays of charged hadrons.

In the first step, photons and neutral hadrons are reconstructed from ECAL and HCAL clusters that are not linked to any tracks. Within the tracker acceptance, all such ECAL clusters are associated to photons and all such HCAL clusters to neutral hadrons. Outside the tracker acceptance all ECAL clusters that are not linked to an HCAL cluster are assigned to photons. All HCAL clusters, regardless of whether they are linked to an ECAL cluster or not, are associated to neutral hadrons.

In the second step, photons, neutral hadrons, and charged hadrons are reconstructed from ECAL and HCAL clusters that are linked to tracks. After the previous step, all remaining HCAL clusters are linked to tracks and the tracks are combined with the remaining ECAL clusters. Then the sum of the track momenta is compared to the calibrated energy. If the energy is larger than the track momenta, the excess is interpreted as the presence of photons and neutral hadrons. Photons are reconstructed from the excess in the ECAL clusters and neutral hadrons from the excess in the



HCAL clusters. Charged hadrons are reconstructed from the tracks assuming them to be charged pions. If there is no energy excess, only charged hadrons and no neutral particles are reconstructed. The momenta of the charged hadrons is recalculated by performing a  $\chi^2$  fit to the tracks and calorimeters in order to improve the momentum resolution.

There are also rare cases where there is a significant deficiency of energy, e.g., the calibrated energy is much lower than the sum of momenta. In this situation, it is tried to reconstruct additional global muons that failed previous reconstruction steps. The tracks of the such reconstructed global muons are then removed. If there is still a significant energy deficiency, it is mostly due to wrongly reconstructed tracks with a large  $p_T$  uncertainty of  $> 1$  GeV. Starting from the track with the largest uncertainty, this tracks are then iteratively removed, until no such tracks are remaining or until energy and momentum are balanced.

## 6.6 Jet reconstruction and identification

Decaying gluons and quarks create a shower of secondary particles that decay and shower further and eventually hadronize due to the confinement of QCD. The particles created in this way are typically located in a cone around the initial particle forming a jet.

### 6.6.1 Jet reconstruction

In CMS, jets are reconstructed by clustering the reconstructed PF particles. Clustering algorithms must be robust against the addition of additional very soft particles (infrared safety) and must cluster two collinear particles in the same way as a single particle with the same momentum (collinear safety). The algorithm used in this thesis for clustering jets is the anti- $k_T$  algorithm [153]. The anti- $k_T$  algorithm is a sequential algorithm and is both infrared and collinear safe. A distance metric  $d_{ij}$  between two PF objects and between a PF object and the beam axis  $d_{iB}$ , is defined:

$$d_{ij} = \min(k_{t,i}^{2p}, k_{t,j}^{2p}) \frac{\Delta_{ij}^2}{R^2},$$

$$d_{iB} = k_{t,i}^{2p}$$

Here,  $k_{t,i}$  is the transverse momentum of object  $i$ .  $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ , where  $y_i$  and  $\phi_i$  are the rapidity and azimuthal angle of particle  $i$ , respectively.  $R$  is a radius parameter that is chosen to be 0.4 in this thesis. And  $p$  is a parameter regulating the relative power between the energy and geometric ( $\Delta_{ij}$ ) scales. In the case of the

anti- $k_T$  algorithm  $p = -1$  is chosen, resulting in an increased  $d_{ij}$  between soft particles. Therefore, soft particles tend not to be clustered among themselves but rather together with hard particles. The clustering is done iteratively. The distance parameters  $d_{ij}$  between all objects and between the objects and the beam axis  $d_{iB}$  are calculated. If  $d_{ij}$  is the minimum, then particle  $i$  and  $j$  are combined into one particle  $k$  and their four-vectors are added. After that particle  $i$  and  $j$  are removed from the list of particles and particle  $k$  is added. If  $d_{iB}$  is the minimum, then  $i$  is called a jet and  $i$  is removed from the list of particles. This is done until all particles are clustered into jets.

### 6.6.2 Jet identification

Quality criteria are applied on reconstructed jets, in order to reject misidentified jets, jets from detector noise and badly reconstructed jets. The criteria are based on the number of PF candidates clustered into a jet and the fraction of the jet energy carried by muons, electrons (charged EM), photons (neutral EM), charged hadrons and neutral hadrons. All requirements for 2016-2018 are listed in table 6.3. The criteria are designed to keep about 98 % to 99 % of the real and well reconstructed jets. The efficiencies are estimated using a tag and probe method in dijet events [154].

### 6.6.3 b jet identification

Jets stemming from b hadrons are called b jets and the process of identifying them is called b tagging. The algorithms that are used for b tagging are called b-taggers. A unique characteristic of b jets is that they often contain secondary vertices. The reason for this is the long life time of b hadrons, which makes them travel some distance before they decay. Thus, they form a secondary vertex that is displaced from the primary vertex. The currently best performing b tagging algorithms used at CMS are deep neural networks named DeepCSV and DeepJet. They are briefly described in the following.

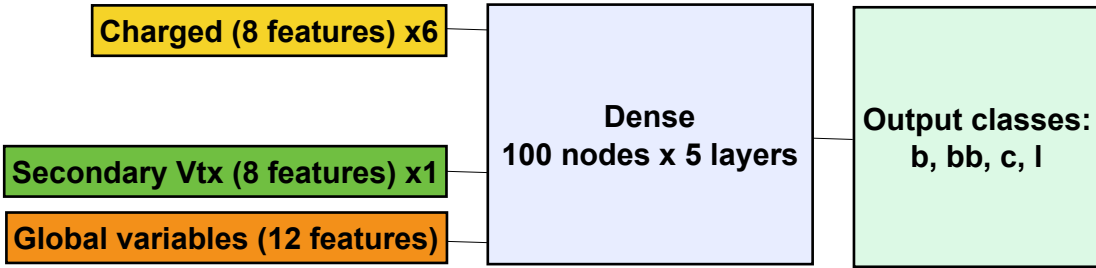
**Table 6.3: Criteria for the identification of AK4 jets.** The selection criteria differ for each year of data-taking. The selection criteria follow the recommendations from the JetMET Physics Object Group of the CMS collaboration [155–157].

2016	$ \eta  \leq 2.4$	$2.4 <  \eta  \leq 2.7$	$2.7 <  \eta  \leq 3.0$	$ \eta  > 3.0$
Number of constituents	$> 1$	$> 1$	–	–
Neutral hadron fraction	$< 0.90$	$< 0.90$	$< 0.98$	–
Neutral EM fraction	$< 0.90$	$< 0.90$	$> 0.01$	$< 0.90$
Number of neutral particles	–	–	$> 2$	$> 10$
Charged hadron fraction	$> 0$	–	–	–
Charged EM fraction	$< 0.99$	–	–	–
Charged multiplicity	$> 0$	–	–	–
2017	$ \eta  \leq 2.4$	$2.4 <  \eta  \leq 2.7$	$2.7 <  \eta  \leq 3.0$	$ \eta  > 3.0$
Number of constituents	$> 1$	$> 1$	–	–
Neutral hadron fraction	$< 0.90$	$< 0.90$	–	$> 0.02$
Neutral EM fraction	$< 0.90$	$< 0.90$	$> 0.02$ and $< 0.99$	$< 0.90$
Number of neutral particles	–	–	$> 2$	$> 10$
Charged hadron fraction	$> 0$	–	–	–
Charged EM fraction	–	–	–	–
Charged multiplicity	$> 0$	–	–	–
2018	$ \eta  \leq 2.6$	$2.6 <  \eta  \leq 2.7$	$2.7 <  \eta  \leq 3.0$	$ \eta  > 3.0$
Number of constituents	$> 1$	–	–	–
Neutral hadron fraction	$< 0.90$	$< 0.90$	–	$> 0.2$
Neutral EM fraction	$< 0.90$	$< 0.99$	$> 0.02$ and $< 0.99$	$< 0.90$
Number of neutral particles	–	–	$> 2$	$> 10$
Charged hadron fraction	$> 0$	–	–	–
Charged EM fraction	–	–	–	–
Charged multiplicity	$> 0$	$> 0$	–	–

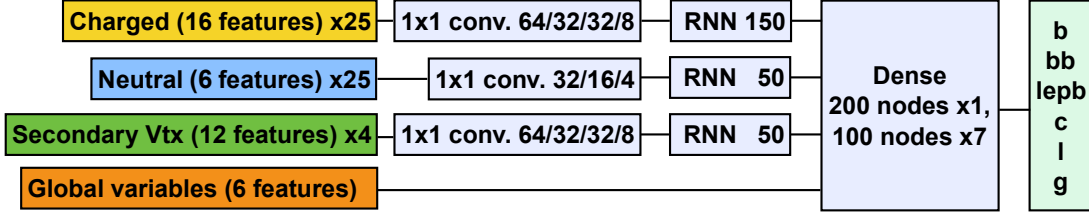
**DeepCSV** is a neural network, which consists of five fully connected layers and four output nodes. The ReLu activation function is used in each layer except the output layer, which uses the softmax activation function. As input it uses in total 68 features including information about charged PF candidates, secondary vertices and some global variables. Jets are classified into originating from bottom quarks, from charm quarks or from light quarks or gluons. An illustration of the DeepCSV algorithm is given in figure 6.1.

**DeepJet** is a neural network, which consists of multiple layers [158]. The first layer consists of 1x1 convolutional networks, which performs automatic feature engineering. This layer is followed by three layers consisting of long-short-term-memory (LSTM) networks, which are a special case of recurrent neural networks (RNN). After the LSTMs there are five fully connected layers followed by six output nodes. ReLu is used as activation function except for the output layer that uses the softmax activation function. About 650 features are used as input including information about charged PF candidates, neutral PF candidates, secondary vertices and some global variables. Like DeepCSV, DeepJet also performs a multiclassification. It does not only identify jets as stemming from b hadrons but also classifies them as originating from charm hadrons, light-flavoured hadrons or gluons. An illustration of the DeepJet algorithm is given in figure 6.2.

DeepJet achieves a significantly better performance compared to DeepCSV with the performance difference being about 10 % at a  $10^{-3}$  misidentification probability. Different working points are defined each corresponding to a certain misidentification probability of light jets. The working points are called loose, medium and tight and the corresponding misidentification probabilities are 10 %, 1 % and 0.1 %, respectively. The average efficiency to correctly tag b jets as b jets in simulated  $t\bar{t}$  events for the DeepJet tagger has been determined to be about 95 %, 85 % and 65 % for the loose, medium and tight working point respectively.



**Figure 6.1: Illustration of the DeepCSV architecture** DeepCSV uses information about charged PF candidates, secondary vertices and some global variables as input. It consists of five fully connected layers with four output nodes. DeepCSV classify jets into different categories, as originating from bottom quarks (b, bb), from charm quarks (c) or from light quarks or gluons (l). The figure is taken from reference [159].



**Figure 6.2: Illustration of the DeepJet architecture** Deepjet uses as input information about charged PF candidates, neutral PF candidates, secondary vertices and some global features. The 1x1 convolutional layers perform a feature engineering and the output is combined using recurrent neural networks (RNN). DeepJet classifies jets into different categories, as originating from bottom quarks (b, bb, lepb), from charm quarks (c), from light quarks (l) or from gluons (g). The figure is taken from reference [158].

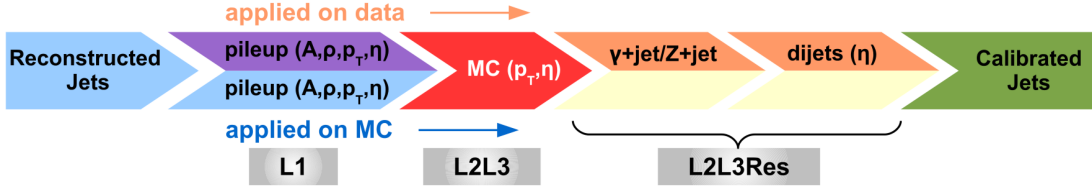
#### 6.6.4 Jet energy corrections

The reconstructed jet energy differs from the particle-level jet energy due to contaminations from pileup and detector response effects. Therefore, the reconstructed jet energy needs to be calibrated to compare it with the energy of particle-level jets from simulation. Particle-level jets are clustered using the anti- $k_T$  algorithm using all stable particles except neutrinos from simulated events.

The correction of the jet energy follows a factorized approach with four steps. At each step the four momenta of the jets are scaled by a scale factor. The corrections are applied sequentially where the output of the previous step is the input for the next step. Figure 6.3 shows the different steps. In the first step additional energy coming from pileup events is removed. The corrections are derived using simulated QCD dijet events with and without pileup overlay. In the next step the detector response between reconstructed and particle-level jets is compared using a simulated QCD dijet sample. The corrections make the detector response uniform as a function of the jet transverse momentum and pseudorapidity. In the last step small differences in the response are corrected. This is done by comparing the momentum of a jet with a more precisely reconstructed jet or particle ( $Z/\gamma$ ) and making use of momentum conservation. Corrections depending on the pseudorapidity can be derived from dijet events by comparing jets in the endcaps, where the measurement is less precise, to more precisely measured jets in the barrel region. Leptonically decaying  $Z$  bosons and photons can be precisely reconstructed. Therefore,  $Z/\gamma$  +jet events are used to correct the absolute scale of the jet  $p_T$ . More details on jet energy corrections can be found in reference [160].

#### 6.6.5 Jet energy resolution corrections

The jet energy resolution in data is worse compared to that in simulation. Therefore, the jet energy in simulation is smeared to match the resolution observed in data. To



**Figure 6.3: Scheme of the jet energy corrections at the CMS experiment.** In the L1 corrections data and MC are corrected for pileup offset. In the L2L3 corrections, differences in the detector response are corrected for using simulated events. The L2L3Res corrections correct for residual differences in the response using dijet and  $Z/\gamma$  +jet events. The figure is taken from reference [161].

smear the jets in simulation it is first checked if a particle-level jet can be matched to the reconstructed jet. The matching criteria is:

$$\Delta R < 0.4/2, \quad |p_T - p_T^{\text{ptcl}}| < 3\sigma_{\text{JER}} \cdot p_T. \quad (6.2)$$

Here,  $\sigma_{\text{JER}}$  is the relative  $p_T$  resolution in simulation. The  $p_T$  of the jet is then rescaled with:

$$c_{\text{JER}} = 1 + (s_{\text{JER}} - 1) \cdot \frac{p_T - p_T^{\text{ptcl}}}{p_T}. \quad (6.3)$$

Here,  $s_{\text{JER}}$  are dedicated data-to-simulation resolution scale factors that are determined by measuring the transverse momentum balance in dijet and  $Z/\gamma$  + jet events [162]. The scale factors are provided by the JetMET Physics Object Group of the CMS collaboration [citejetenergyresolutionwiki](https://cms-jetmet.github.io/jetenergyresolutionwiki/). If no particle-level jet can be matched to the reconstructed jet, a stochastic smearing is applied. The correction factor is then calculated according to:

$$c_{\text{JER}} = 1 + \mathcal{N}(0, \sigma_{\text{JER}}) \sqrt{\max(s_{\text{JER}}^2 - 1, 0)}. \quad (6.4)$$

Here  $\mathcal{N}(0, \sigma)$  are random numbers sampled from a normal distribution.

## 6.7 Missing transverse momentum

Missing transverse momentum  $\mathbf{p}_T^{\text{miss}}$  arises from only weakly interacting particles, like neutrinos or possible yet undiscovered particles. But it can also arise due to finite

acceptance and resolution of the detector or because of pileup.  $\mathbf{p}_T^{\text{miss}}$  is calculated as the negative vector sum of all objects reconstructed from the PF algorithm but it also considers the energy deposits that are not clustered into any PF objects.

$$\mathbf{p}_T^{\text{miss}} = - \sum_i \mathbf{p}_{T,i} \quad (6.5)$$

The missing transverse momentum is corrected by the so-called Type-0 and Type-I corrections. The Type-0 corrections account for effects coming from pileup interactions by removing charged hadrons coming from pileup vertices. The neutral pileup contribution is estimated from the charged pileup using the neutral to charged pion fraction and is removed from the missing transverse momentum. The Type-I corrections correspond to the propagation of the jet energy corrections to the missing transverse momentum. The vector sum of particles that can be clustered into jets are replaced with the vector sum of the corrected jets. Hereby, only jets with  $p_T > 15 \text{ GeV}$  are considered to suppress jets stemming from pileup. The jets also must deploy less than 90 % of their energy in the ECAL and not overlap with PF muon candidates.

Care must be taken that no artificially large  $p_T^{\text{miss}} = |\mathbf{p}_T^{\text{miss}}|$  values are reconstructed, which may be misinterpreted by physics searches as new physics. Artificially large  $p_T^{\text{miss}}$  values often stem from misreconstructed high- $p_T$  muons. These can originate from cosmic rays, severe misreconstruction of the muon momentum or particle misidentification that can for example be caused by punch-through charged hadrons. In the postprocessing step of the PF algorithm, particles that give rise to large  $p_T^{\text{miss}}$  values are identified and the correlation between their momenta and  $\mathbf{p}_T^{\text{miss}}$  is evaluated. Their identification and reconstruction is reconsidered and modified if the modification would reduce  $p_T^{\text{miss}}$  by more than 50 %.





## 7 Performance of b tagging algorithms at the HL-LHC

Identifying jets that originate from bottom quarks is essential for many analyses, especially those studying top quarks, as the top quark decays into bottom quarks with a branching ratio of about 100 %. The current b tagging algorithms used at CMS are DeepJet and DeepCSV (see section 6.6.3), which achieve a b tagging efficiency of 65 % and 55 % respectively at a misidentification probability of 0.1 %. The LHC will be upgraded to the High Luminosity LHC (HL-LHC) [104], which will increase the instantaneous luminosity by about a factor of 10 and is currently scheduled to start operation in 2027. This upgrade will pose a challenge for b tagging algorithms because the increase in instantaneous luminosity will also result in up to 200 pileup events per bunch crossing. To cope with the new environment, the CMS detector will get several upgrades to its individual components. A short description of the planned upgrades for the Phase-2 detector is given in section 4.2.7. For the design of these upgrades it is of interest to study in advance how they might affect the outcome of b-jet identification algorithms. Thus, the performance of the DeepJet and DeepCSV b tagging algorithms are evaluated for the HL-LHC. The results of this study are used as direct inputs to various ongoing Phase-2 sensitivity studies within the CMS collaboration. A description of the retraining of DeepJet and DeepCSV algorithms using simulated events for Phase-2 is given in section 7.1 and the performance of the b tagging algorithms is shown in section 7.2. The Phase-2 CMS detector will include a timing detector, which allows to add timing information for the reconstruction of tracks. A study of the usage of additional timing variables is shown in section 7.3.

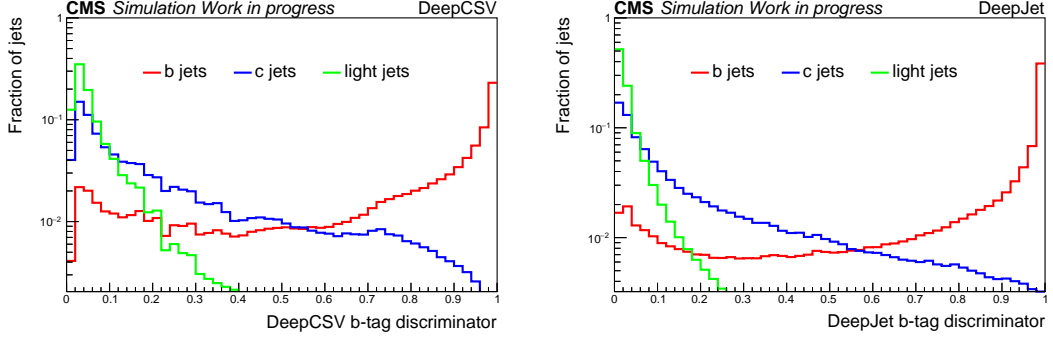
### 7.1 Training

For the training of b tagging algorithms QCD multijet events and events containing top quarks are used. The QCD multijet events provide mostly gluon induced light flavored jets, whereas the jets from the top quark events are enriched in jets originating from bottom quarks. For the Phase-2 training, dedicated simulated QCD multijet, inclusive  $t\bar{t}$ , and single top quark samples have been generated. These samples were simulated using the Phase-2 detector layout and a pileup of 200. The training was performed in version 9.3 of the CMS software, which did not include timing information from the new timing detector. The samples are listed in table 7.1. In Phase-2 it is planned to use the pileup per particle (PUPPI) [163] algorithm for mitigating the effects of large amounts of pileup on jets. Therefore, PUPPI-jets are used for the training of the taggers. Jets

used in the training must be matched to a generator level jet. Further, jets must have  $20 \text{ GeV} < p_T < 1000 \text{ GeV}$  and  $|\eta| < 2.5$ . In total around 40 million jets are selected for training. The jets are randomly split into a training, a validation and a testing data set. 10 % of all jets are used for testing, from the remaining jets 15 % are used for validation and the rest for training. Jets are labeled by ghost association [164]. If at least one  $b$  hadron is clustered in the jet, they are labeled as  $b$  jets. If they contain no  $b$  hadrons but at least one  $c$  hadron, they are labeled as  $c$  jets. And if neither  $b$  hadrons nor  $c$  hadrons are clustered in a jet, the jet is labeled as light jet.

Because of the larger pseudorapidity coverage of the CMS tracker in Phase-2, jets up to  $|\eta| = 3.5$  are considered in the training. The weights for DeepCSV are initialized randomly. As the training of DeepJet takes several days, the weights are not initialized randomly. Instead transfer learning [165] is used by initializing the weights to the ones from the DeepJet model used during Phase-1. The Phase-1 DeepJet model has been trained on about 130 million jets and it has already learned complex features allowing it to distinguish  $b$  jets from jets of different flavors. Even though, the detector geometry in Phase-2 is different and a different algorithm is used for jet identification, this learned features can still be transferred to the training on the Phase-2 geometry. Therefore, transfer learning leads to a faster convergence of the model. The DeepCSV model was trained for 50 epochs and the DeepJet model for 100 epochs. The categorical cross-entropy (see section 3.2) is used as loss function during the training. For the training of both neural networks the Adam [67] optimizer is used. For DeepCSV an initial learning rate of 0.003 and batchsize of 5000 is set, whereas DeepJet uses an initial learning rate of 0.0001 and batchsize of 10 000. The learning rate is halved if the loss on the validation sample stagnates for at least 10 epochs.

In order to test whether the models have been overtrained, the performance of the final models are compared to the performance in the validation and testing data set. In addition the performance is also compared as a function of the number of training epochs. An overtraining would manifest as a divergence, where the performance on the training data set increases and the performance on the validation and testing data set decreases as a function of the number of epochs. From this checks no sign of overtraining is observed. The  $b$ -discriminator outputs of the DeepCSV and DeepJet models are shown in figure 7.1. The figures show the output for  $b$  jets,  $c$  jets and light jets. As can be seen the  $b$  jet distribution is well separated from the others.



**Figure 7.1: Output of Phase-2 b tagging discriminators.** Shown are the output distribution for DeepCSV (left) and DeepJet (right) for b jets, light flavored jets and c jets. The distributions are shown on a mixture of QCD and inclusive  $t\bar{t}$  samples simulated using the CMS detector setup for Phase-2 and both DeepCSV and DeepJet have been retrained for Phase-2.

**Table 7.1: Phase-2 data sets used for the retraining of the b taggers in CMS software version 9.3.** Listed are simulated inclusive  $t\bar{t}$ , QCD multijet and single top quark samples for the CMS Phase-2 upgrade, which have been used to retrain the DeepCSV and DeepJet models. Samples are simulated using a pileup of 200 and version 9.3 of the CMS software the tag is PhaseIITDRFall17MiniAOD-PU200\_93X\_upgrade2023\_realistic.

Name	Number of events
TT_TuneCUETP8M2T4_14TeV-powheg-pythia8	1 764 010
TT_TuneCUETP8M2T4_14TeV-powheg-pythia8	2 874 776
QCD_Flat_Pt-15to7000_TuneCUETP8M1_14TeV_pythia8	7 334 392
QCD_Flat_Pt-15to7000_TuneCUETP8M1_14TeV_pythia8	2 675 000
ST_tW_DR_14TeV_antitop_incl-powheg-pythia8	192 500
ST_tW_DR_14TeV_antitop_incl-powheg-pythia8	312 500
ST_tW_DR_14TeV_top_incl-powheg-pythia8	404 897
ST_tch_14TeV_antitop_incl-powheg-pythia8-madspin	123 318
ST_tch_14TeV_antitop_incl-powheg-pythia8-madspin	336 450
ST_tch_14TeV_top_incl-powheg-pythia8-madspin	372 925

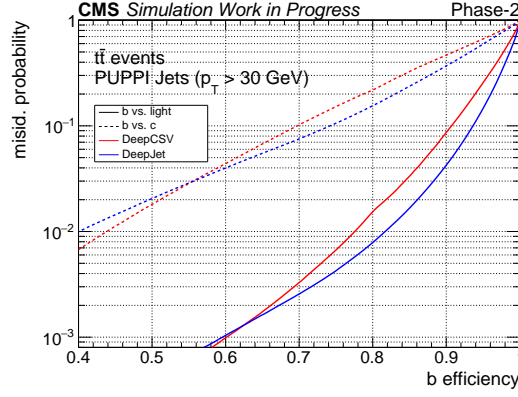
## 7.2 Performance

The performance of the trained taggers is evaluated on inclusive  $t\bar{t}$  and QCD multijet events, which are simulated for Phase-2. Only jets with at least one good track and  $p_T > 30$  GeV are used. The performance of both models is evaluated by computing ROC curves (see section 3.2). All ROC curves show misidentification probability for  $b$  jets vs. light jets or  $c$  jets on the  $y$  axis and the  $b$  jet efficiency on the  $x$  axis. The larger the  $b$  jet efficiency at a given misidentification probability, the better the performance of the tagger.

The performance of the Phase-2 DeepCSV and Phase-2 DeepJet models on events simulated for Phase-2 is shown in figure 7.2. Figure 7.3 shows the performance of the Phase-1 DeepCSV and Phase-1 DeepJet models on events simulated for Phase-1, which corresponds to their performance during Run 2. This allows to directly compare the performance of the taggers at the HL-LHC with the performance they achieved during Run 2. As can be seen the performance of the Phase-2 DeepCSV model is similar to the performance of the Phase-1 DeepCSV model. The  $b$  vs. light jet discrimination of the Phase-2 DeepCSV model is even slightly better compared to the Phase-1 DeepCSV model. The Phase-2 DeepJet model performs better than the Phase-2 DeepCSV model, except for  $b$  vs. light misidentification probabilities below roughly  $10^{-3}$ . However, while DeepCSV sees no deterioration in performance, the performance of the Phase-2 DeepJet model is slightly worse compared to the Phase-1 DeepJet model. Also the performance difference between DeepCSV and DeepJet is not nearly as large in Phase-2 as in Phase-1.

A cross-check is performed to ensure that the deterioration in performance of DeepJet in Phase-2 as compared to Phase-1 is because of the more challenging environment at the HL-LHC and not due to issues in the training. For this purpose also the Phase-1 DeepJet model is applied to events simulated for Phase-2 in order to compare the performance of the Phase-1 and Phase-2 DeepJet model on the same events. The performance of both models are evaluated by computing ROC curves in bins of jet  $|\eta|$ . The Phase-2 DeepJet model shows an improved performance of about 10-15 % over the Phase-1 DeepJet model. The performance difference increases for larger pseudorapidity values which can be explained by the fact that the Phase-1 DeepJet model was only trained with jets up to  $|\eta| < 2.5$  whereas in the Phase-2 training jets with  $|\eta| < 3.5$  were used.

The performance of the Phase-2 DeepCSV and DeepJet models are also studied as a function of  $p_T$  and  $|\eta|$  and corresponding ROC curves are shown in figure 7.4. As stated above, in Phase-2 DeepCSV performs better than DeepJet for low misidentification probabilities. As can be seen in the binned ROC curves this is true in the pseudorapidity region up to  $|\eta| < 2.5$ , which corresponds to the acceptance of the tracker in Phase-1. However, in the region of  $2.5 < |\eta| < 3.5$  DeepJet performs consistently better. The performance of DeepJet does also depend less on the transverse momentum of the Jet than the performance of DeepCSV. The impact of pileup on the performance of the taggers is studied by computing ROC curves in bins of the number of primary

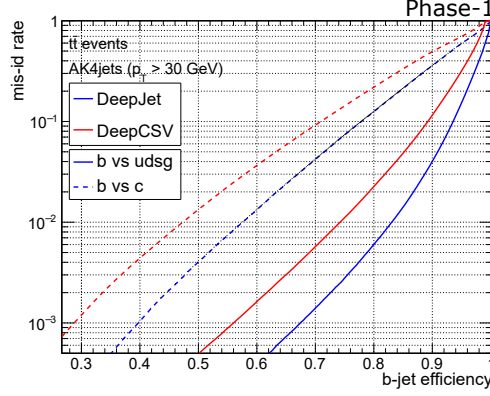


**Figure 7.2: Performance of  $b$  tagging discriminators in Phase-2.**

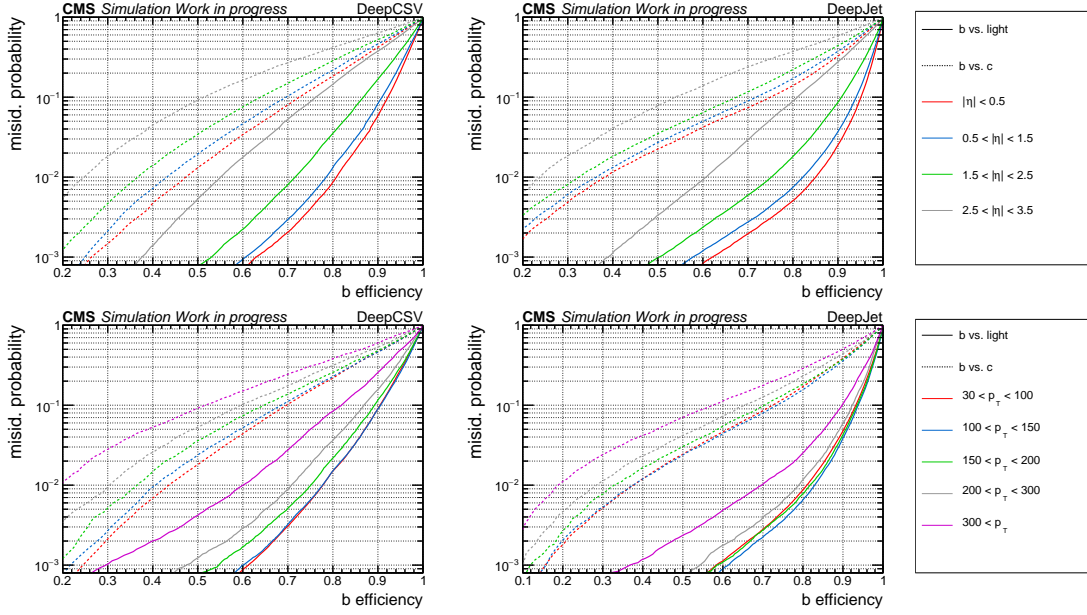
Shown are the ROC curves for DeepCSV and DeepJet. The figure shows the performance of DeepCSV and DeepJet on  $t\bar{t}$  events simulated using the Phase-2 CMS detector geometry and both DeepCSV and DeepJet have been retrained for Phase-2. The shown performance is of illustrative nature, as the exact performance depends on the event topology and on the number of  $b$  jets from gluon splitting.

vertices. However, no significant dependency of the performance on the number of primary vertices has been found. At present only simulated events with a maximum number of 200 pileup events are available. In the future also simulated events with a maximum number of 140 pileup will be produced, which will allow to perform more in depth studies on the pileup dependency.

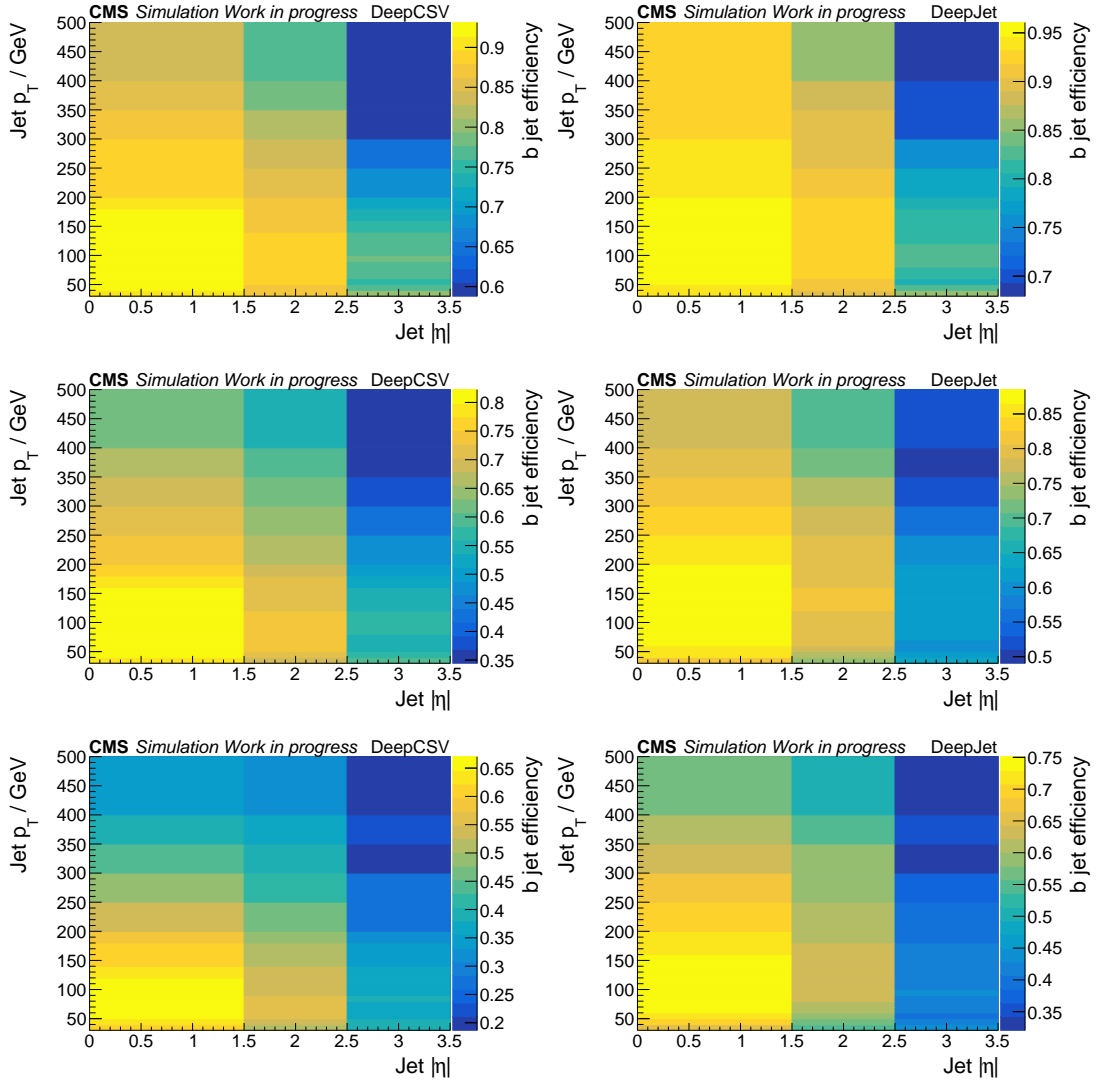
From ROC curves binned in  $p_T$  and  $|\eta|$ , efficiency maps are computed. The efficiency maps correspond to the efficiency of the model at a given working point in bins of  $p_T$  and  $|\eta|$ . The working points are defined in terms of light jet misidentification probability, which is the probability to wrongly classify a light jet as  $b$  jet. There are three different working points: loose, medium and tight that correspond to 10 %, 1 %, and 0.1 % misidentification probability respectively. It should be noted that in these efficiency maps the misidentification probability is fixed in each bin. This is different to how the working points are defined in Phase-1, where they correspond to a global misidentification probability, which is essentially an average over all bins. The efficiency maps are shown in figure 7.5.



**Figure 7.3: Performance of  $b$  tagging discriminators in Phase-1.** Shown are the ROC curves for DeepCSV and DeepJet. The figure is taken from reference [158] and shows the performance of DeepCSV and DeepJet on  $t\bar{t}$  events simulated using the Phase-1 CMS detector geometry and both DeepCSV and DeepJet have been trained on the Phase-1 geometry. This corresponds to the performance of the taggers during Run 2. The shown performance is of illustrative nature, as the exact performance depends on the event topology and on the number of  $b$  jets from gluon splitting.



**Figure 7.4: Performance of Phase-2  $b$  tagging discriminators in bins of  $p_T$  and  $|\eta|$ .** Shown are the ROC curves for DeepCSV (left) and DeepJet (right) for different ranges in jet  $\eta$  (top) and jet  $p_T$  (bottom). The performance is evaluated on a mixture of simulated QCD multijet and inclusive  $t\bar{t}$  samples simulated using the CMS detector setup for Phase-2 and both DeepCSV and DeepJet have been retrained for Phase-2.



**Figure 7.5: Efficiency maps for Phase-2 b tagging discriminators.** Shown are the efficiency maps for DeepCSV (left) and DeepJet (right). The efficiency maps are shown for the loose (top), medium (middle) and tight (bottom) working point. The maps are binned in jet  $|\eta|$  and jet  $p_T$ . The performance is evaluated on QCD multijet and inclusive  $t\bar{t}$  samples simulated using the CMS detector setup for Phase-2 and both DeepCSV and DeepJet have been retrained for Phase-2.

### 7.3 Timing information

As explained before the CMS detector in Phase-2 will include a timing detector that will enable the measurement of the flight time of particles. The timing information adds an additional dimension to reconstructed vertices and tracks, because not only their spacial  $x$ ,  $y$  and  $z$  coordinates are reconstructed but also a time coordinate. Therefore, vertices reconstructed using timing information are referred to as 4D vertices and vertices without timing information as 3D vertices. It is interesting to study how the addition of timing information may help  $b$  tagging algorithms. Unfortunately, in CMS software version 10.6 there were some issues that lead to a significant deterioration in the 4D vertex reconstruction. These issues are fixed and the reconstruction improved in newer software versions. However, simulating new samples with the fixed version takes several months and was in fact not done in time for this thesis. In the meantime preliminary studies were conducted to evaluate the viability of including additional timing information as input to the DeepCSV model. 3D primary vertices were used and timing information was included by defining dedicated timing variables. Also an additional variable is introduced to stabilize the performance against pileup. The additional variables are:

**Event time:** The event time is computed from all tracks attached to jets in an event with a transversal distance  $d_{xy} < 0.05$  cm and longitudinal distance  $d_z < 0.1$  cm to a 3D primary vertex. The event time is then the  $p_T^2$  weighted average time:

$$t_{\text{Event}} = \frac{\sum_i t_i p_{T,i}^2}{\sum_i p_{T,i}^2}. \quad (7.1)$$

**Jet time:** The total jet time is computed for each jet. It is calculated like the event time but only from tracks associated to a given jet. As the sign and the offset relative to the event time do not contain useful information, the absolute difference between the total jet time and event time is taken as the jet time.

**Jet secondary vertex time:** The jet secondary vertex time is computed like the jet time, but only from tracks attached to a secondary vertex.

**Pileup density:** The pileup density is obtained by dividing the number of pileup vertices in a window of 1 cm around the primary vertex's  $z$ -position by 20. This gives the number of pileup vertices per cm.

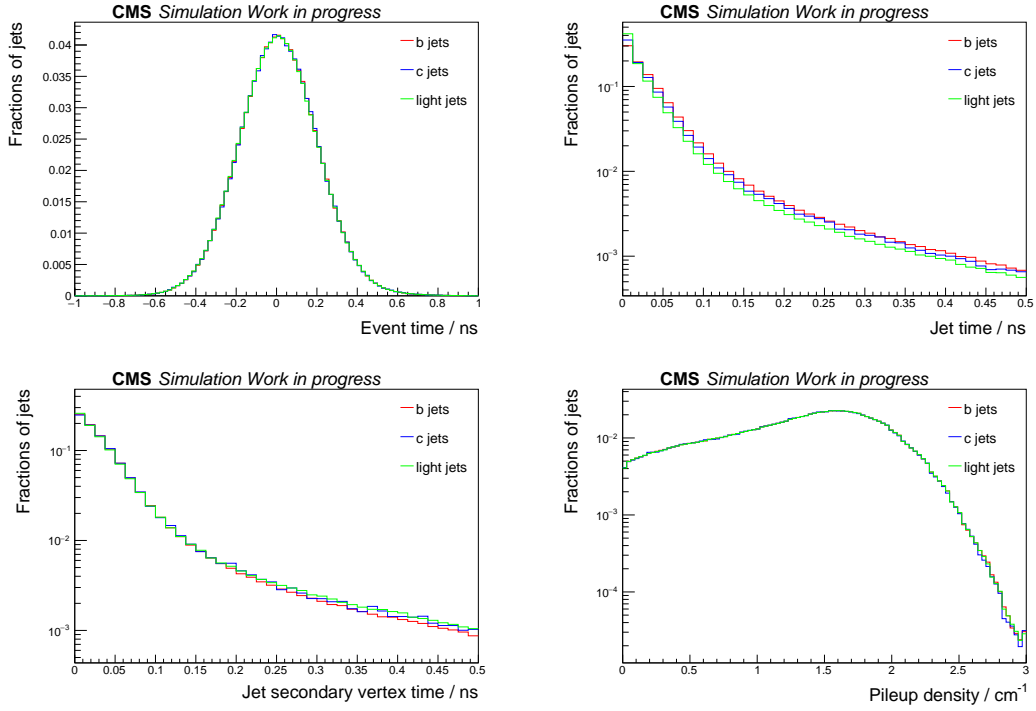
The distributions of these variables are shown in figure 7.6. The RMS of the event time spread is about 200 ps, for the jet and jet secondary vertex time it is about 40 ps relative to the event time. The event time is very similar to the 4D primary vertex time with the RMS of the spread of the difference between the 4D primary vertex time and the event time being about 30 ps. The pileup density is not a timing related variable but it was included for the purpose of stabilizing the performance as a function of pileup. As can be seen, the jet time has larger values for  $b$  jets especially compared to light jets, which



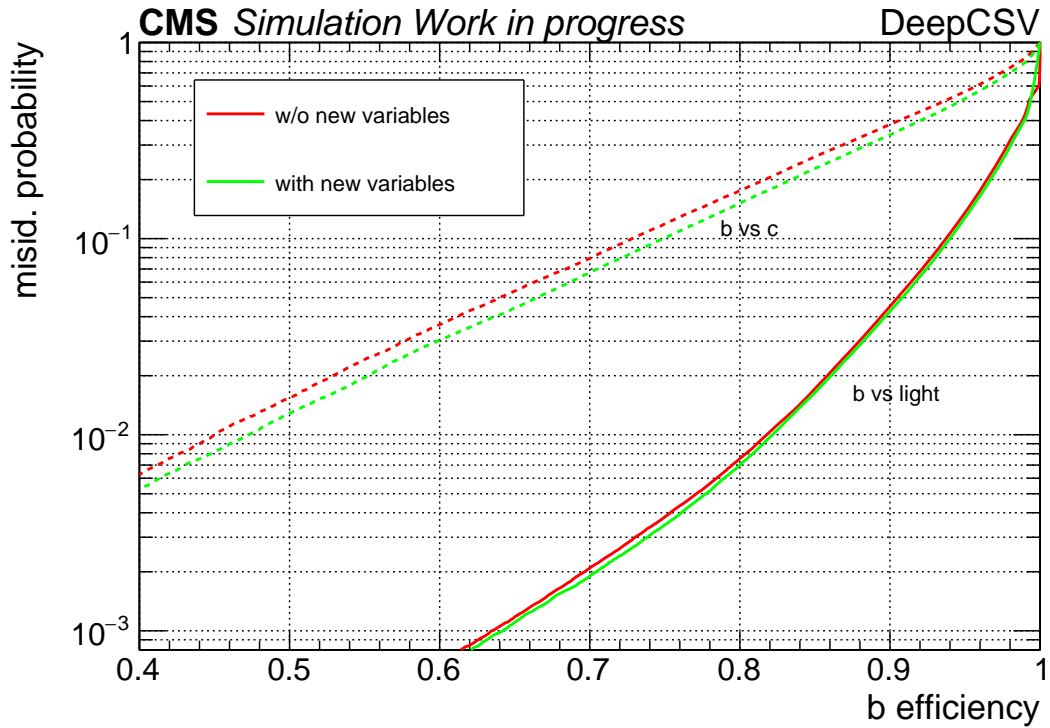
**Table 7.2: Phase-2 data sets used for the retraining of the b taggers in CMS software version 10.6.** Listed are simulated inclusive  $t\bar{t}$  and QCD multijet samples for the CMS Phase-2 upgrade which have been used to retrain the DeepCSV model. Samples are simulated using a pileup of 200 and version 10.6 of the CMS software the tag is PhaseIITDRSpring19MiniAOD-PU200\_106X\_upgrade2023\_realistic.

Name	Number of events
TTbar_14TeV_TuneCP5_Pythia8	98 704
TT_TuneCP5_14TeV-powheg-pythia8	300 000
TTTo2L2Nu_TuneCP5_14TeV-powheg-pythia8	4 935 651
QCD_Pt_0_1000_14TeV_TuneCUETP8M1	499 073
QCD_Pt-15to3000_TuneCP5_14TeV-pythia8	4 972 820
QCD_Pt-15to3000_TuneCP5_Flat_14TeV-pythia8	96 958
QCD_Pt_30to50_TuneCP5_14TeV_pythia8	496 632
QCD_Pt_50to80_TuneCP5_14TeV_pythia8	291 159
QCD_Pt_80to120_TuneCP5_14TeV_pythia8	95 790
QCD_Pt_120to170_TuneCP5_14TeV_pythia8	50 000
QCD_Pt_170to300_TuneCP5_14TeV_pythia8	50 000
QCD_Pt_300to470_TuneCP5_14TeV_pythia8	50 000
QCD_Pt_470to600_TuneCP5_14TeV_pythia8	50 000
QCD_Pt_600oInf_TuneCP5_14TeV_pythia8	50 000

can be explained by the longer lifetime and therefore longer flight time of b hadrons. The jet secondary vertex time shows a similar distribution for b jets and other jets with a tendency to slightly lower values for b jets. The position of a secondary vertex in a b jet corresponds to the location where the b hadron decayed. Therefore, measured from a secondary vertex the flight times are not expected to be significantly larger for b jets, compared to jets of different flavors. Thus, it is understandable that the jet secondary vertex time is not larger for b jets compared to other jet flavors. However, even variables with low individual discrimination power still provide information to the tagger through their correlation with other variables. The DeepCSV model was retrained using the samples listed in table 7.2 with and without including the additional variables listed above. Figure 7.7 shows a comparison of the performance of the model with and without the additional variables. The model with additional timing variables achieves a slightly better performance especially for b jet vs. c jets discrimination.



**Figure 7.6:** Distributions of additional variables for  $b$  tagging. Shown are the event time (top left), the jet time (top right), the jet secondary vertex time (bottom left), and the pileup density (bottom right) for  $b$  jets,  $c$  jets and light jets. The timing variables make use of the new timing detector, which will be deployed in the CMS Phase 2 detector.



**Figure 7.7: Performance of DeepCSV in Phase-2 with timing information.** Shown are ROC curves for DeepCSV with and without additional timing variables. The performance is evaluated on inclusive  $t\bar{t}$  samples simulated using the CMS detector setup for Phase 2 and DeepCSV has been retrained for Phase2.

## 7.4 Conclusion

Both DeepCSV and DeepJet achieve good performance at the HL-LHC. DeepCSV proves to be quite robust and achieves very similar performance as in Run 2. The performance of DeepJet is on average better than the performance of DeepCSV, but the performance difference is not as large as during Run 2. Preliminary studies show that timing information provides additional discrimination between  $b$  jets and jets of different flavors, which can help to improve the performance of  $b$  tagging algorithms. However, further studies on the feasibility of timing information with full 4D primary vertex reconstruction are needed before drawing final conclusions.

# 8 Measurement of the $t$ -channel single top quark production cross section

## 8.1 Introduction

Measurements of single top quark production can be used to probe the electroweak sector of the SM as single top quarks are produced by the weak interaction. Therefore, single top quark production is suitable for constraining new physics contributions to the  $Wtb$  vertex. Single top quark production also gives insights into the structure of the proton and provides an independent measurement of the Cabibbo-Kobayashi-Maskawa (CKM) matrix element  $|V_{tb}|$ . Further, due to the V-A structure of the weak interaction, single top quarks are produced in a left-handed state and are therefore polarized. As top quarks decay before they hadronize, their polarization can be measured through their decay products. There are three production modes for single top quarks: the  $t$ -channel, the  $tW$  channel and the  $s$ -channel. The dominant production mode of single top quarks in proton-proton collisions at the LHC is the  $t$  channel, which is studied in this thesis.

The inclusive single top quark  $t$ -channel production cross section at the LHC has been measured at  $\sqrt{s} = 7$  TeV by the ATLAS Collaboration [166, 167] and the CMS Collaboration [168, 169]. At 8 TeV a fiducial, inclusive and differential cross section measurement was performed by the ATLAS Collaboration [170] and an inclusive cross section measurement and measurement of the top quark polarization by the CMS Collaboration [171, 172]. The inclusive cross section at  $\sqrt{s} = 13$  TeV has been measured by the ATLAS Collaboration [173]. The CMS Collaboration has provided an inclusive and a differential cross section measurement at 13 TeV [174, 175] using  $35.9 \text{ fb}^{-1}$  of data.

This analysis provides the first measurement of the inclusive and differential  $t$ -channel single top quark production cross section using the full CMS Run 2 data set of  $137.1 \text{ fb}^{-1}$  at 13 TeV. It is the first measurement of the transverse ( $x$ ) and normal ( $y$ ) top quark polarization components in single top quark production. Further, an improved technique that employs multivariate techniques for reconstructing the top quark, has been developed and used in this analysis for the first time.

Specifically, this analysis aims at measuring the total production cross section of  $t$ -channel single top quarks and antiquarks, as well as their ratio. Further, differential cross sections of single top quark and antiquarks are measured as a function of the top quark transverse momentum and rapidity and of three polarization angular variables

$\cos(\theta_x)$ ,  $\cos(\theta_y)$ ,  $\cos(\theta_z)$ , which have been defined in section 2.3. The transverse, normal and longitudinal top quark spin asymmetries  $A_x$ ,  $A_y$  and  $A_z$  — which are also defined in section 2.3 — are extracted by performing a linear fit to the measured angular distributions. Finally, the measurement of the differential angular distributions are used to constraint EFT coupling parameters that are highly sensitive to the top quark polarization.

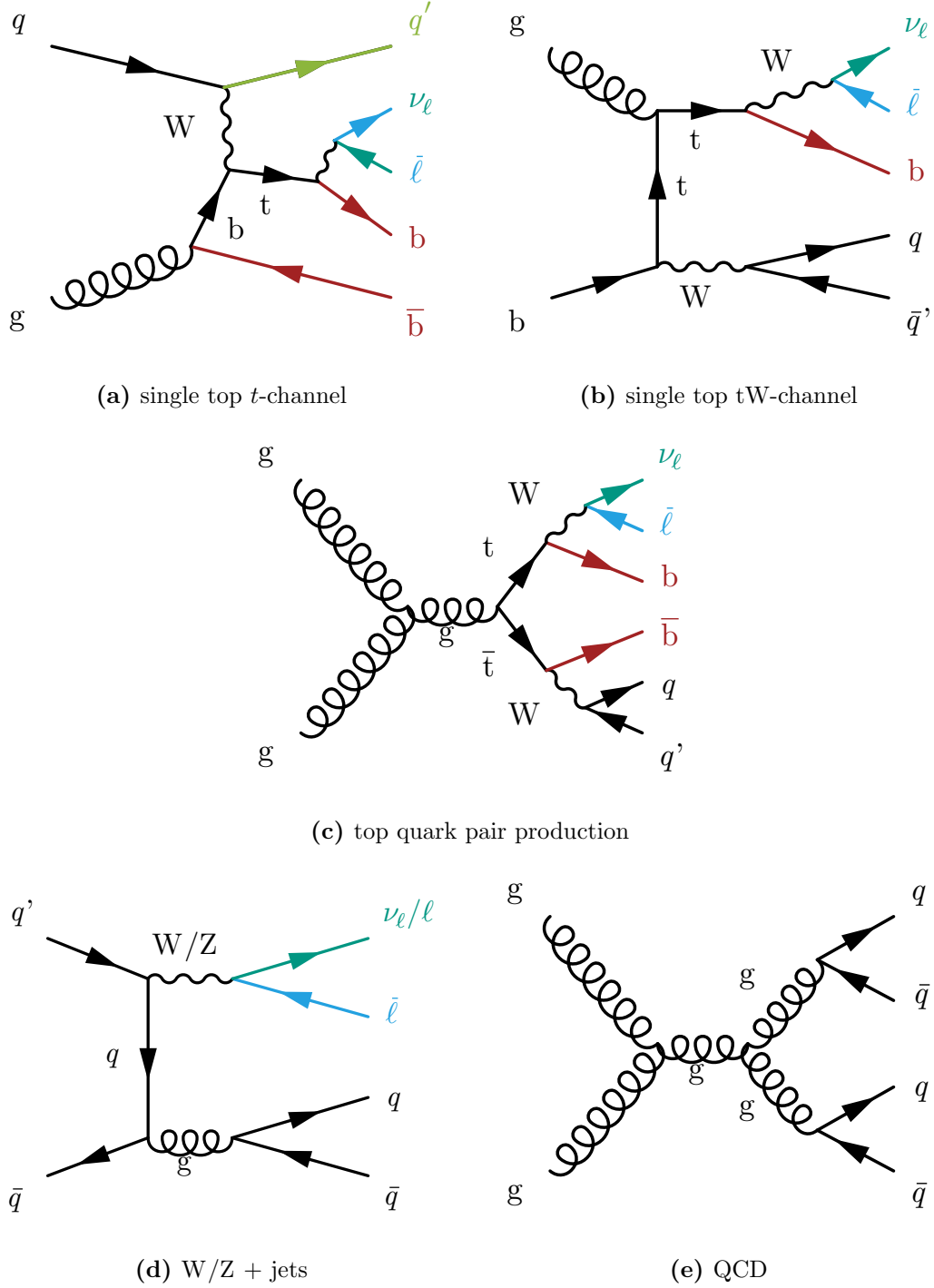
In this analysis, events with one isolated muon or electron and two or three jets are selected. The modeling of the QCD multijet background is obtained from a sideband region in data. A multiclassification BDT is trained to categorize events into  $t$ -channel single top quark, top quark background, W/Z-jets or QCD multijet events. The inclusive  $t$ -channel single top quark production cross section is obtained by performing a ML fit to data on a distribution derived from the multiclassification BDT. The differential cross sections are obtained via unfolding, which is technically implemented via a two dimensional ML fit. For each variable, templates are derived for signal and background processes and fitted to the data. A possible right-handed coupling between the W boson and the top quark may significantly alter the top quark polarization. In order to constraint the magnitude of such couplings, the angular distributions are simulated for different coupling points using MADGRAPH5\_AMC@NLO. A  $\chi^2$ -minimization is performed in order to find the set of couplings that best matches the measured differential angular distributions and to derive constraints.

## 8.2 Event topology

This analysis aims at measuring the properties of top quarks. However, top quarks themselves are not observable in the detector, as they decay after around  $5 \cdot 10^{-25}$  s. Measurable are only the final-state particles that live long enough to be detected in the CMS detector. These final-state particles make up the distinct event topology of single top  $t$ -channel events. A good understanding of the signal's event topology is therefore necessary for the reconstruction of the signal process and to make precise measurements. There are also various background processes that have a similar final state as the signal process. A good knowledge of the event topology allows to define a kinematic phase-space region that is enriched in signal events and suppresses contributions from background processes. The Feynman diagrams for the signal process and the dominant background processes are shown in figure 8.1.

### 8.2.1 Signal process

The signal process in this analysis is the  $t$ -channel production of single top quarks. The LO Feynman diagram is shown in 8.1a. In this analysis leptonic decays of the top quark are considered. The final state therefore contains a charged lepton and missing transverse momentum which is due to the corresponding neutrino. In addition there is a jet originating from the bottom quark from the top quark decay. An additional b jet



**Figure 8.1: Feynman diagrams for signal and background processes**  
 Shown are Feynman diagrams for single top  $t$ -channel production (a), for single top  $tW$  production (b), for top quark pair production (c), for  $W/Z$ -jets production (d) and for QCD multijet production (e).

arises from the splitting of the initial state gluon. This  $b$  jet is typically softer than the  $b$  jet from the top quark decay and often times lies outside the detector acceptance. A distinct feature of the single top  $t$ -channel production mode is the light-flavored forward jet which recoils against the top quark and is therefore also called spectator jet. This jet normally has a large pseudorapidity. It can be distinguished from additional radiation from the parton shower by its larger transverse momentum. The final state of the signal process thus consists of a single charged lepton, missing transverse momentum, one light-flavored forward jet and one to two  $b$  jets.

### 8.2.2 Background processes

**Additional single top quark production modes** Besides the  $t$ -channel production mode, there are two additional production modes of single top quarks. The first is the  $s$ -channel production that has an about 20 times lower cross section than the signal process and is therefore neglected. The second is the associated production of a top quark with a  $W$  boson, which has an about 3 times lower cross section than the signal process and is considered as a relevant background process in this analysis. Figure 8.1b shows a possible Feynman diagram for single top quark  $tW$  production. It can mimic the final state of the signal process if one of the two  $W$  bosons decays hadronically and the other leptonically. In this case, its final state consists of a charged lepton, missing transverse momentum and three jets of which one originates from a bottom quark.

**Top quark pair production** Top quark pairs ( $t\bar{t}$ ) are produced via the strong interaction and have therefore an about 4 times larger production cross section than single top quarks. Figure 8.1c shows a Feynman diagram for a semileptonic  $t\bar{t}$  event. To resemble the signal event topology at least one of the top quarks needs to decay leptonically. Fully-hadronic  $t\bar{t}$  events are negligible in this analysis. The main contribution stems from semileptonic  $t\bar{t}$  events. In this case there is a charged lepton, missing transverse momentum, two  $b$  jets and two additional jets in the event — making four jets in total. If one of the none  $b$  jets fails the detector acceptance the final state looks exactly like the signal process. Despite having two leptons, dileptonic  $t\bar{t}$  events are still a relevant background process for this analysis. If one of the leptons is not properly reconstructed or fails the lepton identification criteria, the final state has one charged lepton, missing transverse momentum and two jets, resembling the final-state of the signal process.

**$W$  and  $Z$  production in association with jets** The associated production of a  $W$  or  $Z$  boson with jets ( $W/Z$ -jets) has an at least one order of magnitude larger cross section than single top quark production. Contributions from this background process can be significantly reduced by requiring at least one  $b$  jet, as the processes do not contain a top quark and  $b$  jets can only arise from gluon splitting. An exemplary Feynman diagram for  $W/Z$ -jets production is shown in figure 8.1d. This process can involve an arbitrary number of jets that arise from gluon splitting. Though, the production cross section decreases with the number of addi-



tional jets. In W+jets events the W boson needs to decay into a charged lepton and a neutrino. If there are two or three additional jets from which at least one is identified as a b jet then the final state resembles the final state of the signal process, making it very difficult to reject. In Z+jets events there is no missing transverse momentum as the Z boson decays into a lepton and an antilepton. One of the leptons may fail the identification criteria and missing transverse momentum can arise from misreconstruction. But Z+jets production is of minor importance compared to W+jets production.

**QCD multijet production** The total production cross section for QCD interactions is several orders of magnitude larger than the production cross section of single top quarks. Therefore, QCD multijet events are a relevant background to this analysis, although the event topology differs greatly from the event topology of the signal process. A Feynman diagram of a typical QCD multijet event is depicted in figure 8.1e. The final state includes multiple jets. Leptons can originate from misidentified jets.

### 8.3 Event selection

All events considered in this analysis must have a well reconstructed primary vertex with a maximum vertical distance of 2 cm with respect to the beam axis and a maximum longitudinal distance of 24 cm with respect to the detector center. Also at least four tracks must be associated with the primary vertex. Events are rejected if the missing transverse momentum is badly reconstructed due to detector noise, bad calibration, beam halo effects or badly-reconstructed muons or other PF objects.

As explained in section 8.2, the final-state of single top  $t$ -channel production consists of a charged lepton, missing transverse momentum, one light-flavored forward jet and one or two b jets. In this analysis, events with one muon or electron are selected. Events must pass the corresponding single-muon or single-electron trigger paths. The trigger paths used are different in each year of data-taking because different trigger menus have been used in each year due to the increased instantaneous luminosity. In general, the single-muon and single-electron trigger paths with the lowest  $p_T$  thresholds that are still unscaled are used. Table 8.1 lists the high-level trigger (HLT) paths used in each year of data-taking. In 2017 and 2018 an electron+jet cross trigger path has been used in addition to the single-electron trigger paths. These cross trigger paths allow for a lower electron  $p_T$  threshold without significantly increasing the total rate of the HLT menu, as the requirement of the additional jet reduced the number of events passing the trigger path. In 2016 the single-electron trigger path is restricted to  $|\eta| < 2.1$ , whereas in 2017 and 2018 no restriction on the pseudorapidity of the electron is applied.

In this analysis, muons with loose and tight ID, as defined in section 6.3, are used. Isolated tight muons must have  $p_T > 26/30/26$  GeV in 2016/2017/2018,  $|\eta| < 2.4$  and a relative isolation of  $I_{\text{rel.}} < 0.06$ . Loose muons must have  $p_T > 10$  GeV,  $|\eta| < 2.4$

**Table 8.1: High-level trigger paths used in the analysis.** This analysis uses single-muon and single-electron triggers that differ in each year of data taking. In 2017 and 2018 electron+jet cross-triggers are used in addition to the single-electron triggers as they allow for a lower electron  $p_T$  threshold.

lepton	year	trigger paths
$\mu$	2016	HLT_IsoTkMu24, HLT_IsoMu24
	2017	HLT_IsoMu27
	2018	HLT_IsoMu24
e	2016	HLT_Ele32_eta2p1_WPTight_Gsf
	2017	HLT_Ele35_WPTight_Gsf, HLT_Ele30_eta2p1_WPTight_Gsf_CentralPFJet35_EleCleaned
	2018	HLT_Ele32_WPTight_Gsf, HLT_Ele30_eta2p1_WPTight_Gsf_CentralPFJet35_EleCleaned

**Table 8.2: Electron impact parameter selections.** Selections applied on the transverse ( $d_{xy}$ ) and longitudinal ( $d_z$ ) impact parameters of electrons. The selection criteria depend on the difference between the electron's pseudorapidity  $\eta$  and the pseudorapidity  $\eta_{SC}$  of the corresponding supercluster.

$ \eta - \eta_{SC} $	$d_{xy} / \text{cm}$	$d_z / \text{cm}$
$\leq 1.479$	$< 0.05$	$< 0.1$
$> 1.479$	$< 0.1$	$< 0.2$

and  $I_{\text{rel.}} < 0.2$ . The choices of the  $p_T$  cuts are driven by the trigger thresholds. The cuts are chosen such that the triggers operate at their maximum efficiency and not in the turn-on regime. Events with muons must have exactly one isolated tight muon, no additional loose muons and no veto electrons.

The definition of tight and veto electrons can be found in section 6.4. In this analysis, tight electrons must have  $p_T > 35/32/32 \text{ GeV}$  in 2016/2017/2018 and  $|\eta| < 2.1$  in all years. Additionally, both tight and veto electrons must pass selection criteria based on their transverse and longitudinal impact parameters as detailed in table 8.2. The selection depend on the electron's pseudorapidity and of the pseudorapidity  $\eta_{SC}$  of the corresponding supercluster. Electrons in the transition region between the ECAL barrel and endcap  $1.442 < |\eta_{SC}| < 1.566$  are rejected. Events with electrons are required to have exactly one tight electron, no additional veto electrons and no loose muons.

All jets considered in this analysis must have  $p_T > 40 \text{ GeV}$  and  $|\eta| < 4.7$ , further they must pass the quality criteria described in section 6.6 and have a minimum distance  $\Delta R > 0.4$  with respect to the selected lepton in the event. B jets must pass the tight b tagging working point of the DeepJet algorithm and must lie within the acceptance of the tracker.

Depending on the number of selected jets and b-tagged jets three different event categories are defined:

**2j1t** - signal region. Events need to contain exactly two jets, of which one is b tagged

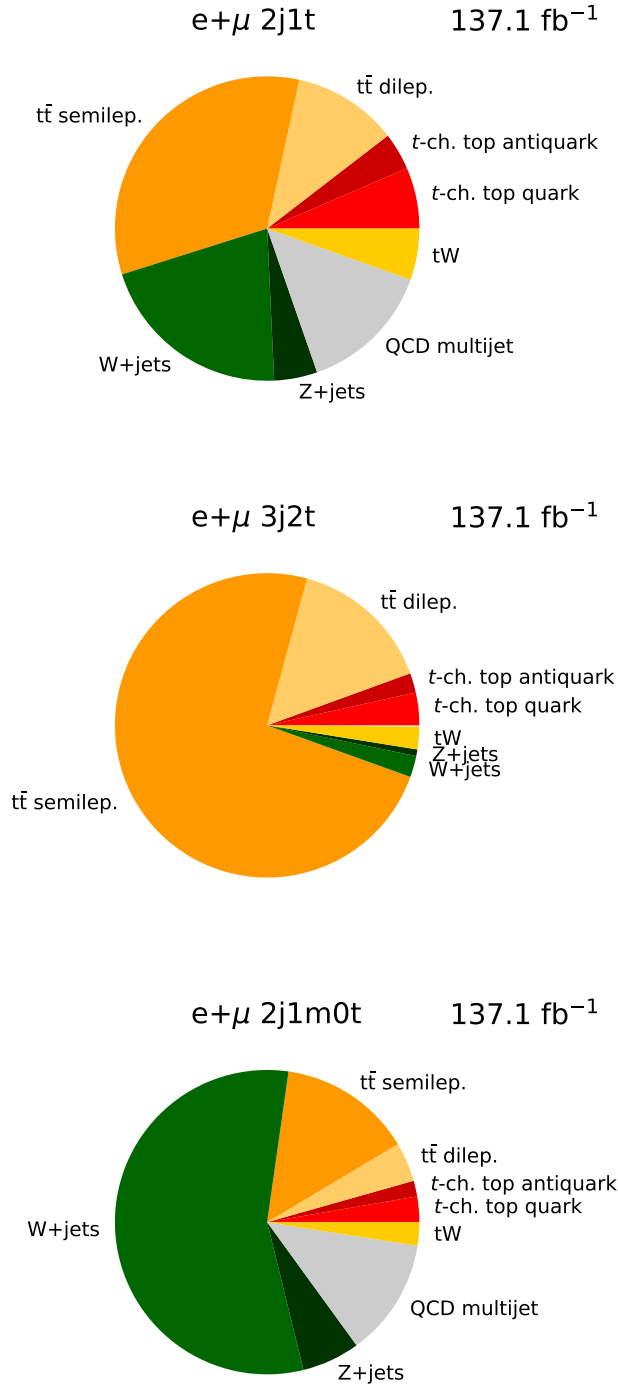
**3j2t** -  $t\bar{t}$  control region. Events must have exactly three jets, of which two are b tagged

**2j1m0t** - W/Z-jets and QCD control region. Events must have exactly two jets, of which none pass the tight b tagging working point. One of the two jets must fail the tight b tagging working point but must pass the medium b tagging working point.

The flavor composition of W/Z-jets events in the signal region is about 90 % W/Z + heavy-flavored jets and only about 10 % W/Z + light-flavored jets. The 2j1m0t region has a W/Z + heavy-flavored jet fraction of about 75 %, which is close enough to the composition in the signal region to control the modeling.

In all categories events must pass either the muon or electron selections described above. The strict isolation and identification criteria employed on the leptons help to significantly reduce background contributions from QCD multijet events by about a factor of  $10^{-4}$  compared to having no selection criteria, as estimated from simulated QCD multijet events. The tight b tagging criteria on the selected b jet reduces the amount of QCD multijet events further by about a factor of  $10^{-4}$  to  $10^{-5}$  and also significantly reduces the contributions from W/Z-jets background events by a factor of about  $10^{-5}$ .

Figure 8.2 shows the relative contributions of semileptonic and dileptonic  $t\bar{t}$ ,  $tW$  production, single top quark and antiquark production in the  $t$  channel, W+jets, Z+jets and QCD multijet events to the 2j1t, 3j2t and 2j1m0t event categories. The fraction of signal events in the 2j1t event category is about 10.5 % - about 6.5 % from top quark and 4 % from top antiquark production. The largest background stems from semileptonic  $t\bar{t}$  production making up about one third of the total events. In the  $t\bar{t}$ -enriched 3j2t event category semileptonic  $t\bar{t}$  contributes about 74 % and dileptonic about 15 %. The fraction of W+jets events in the 2j1m0t event category is about 56 %, the contribution of Z+jets is 6 %, while QCD multijet contributes about 13 %.



**Figure 8.2: Relative contributions of signal and background processes to each event category.** Shown are the contributions of semileptonic and dileptonic  $t\bar{t}$ ,  $tW$  production, single top quark and antiquark production in the  $t$  channel,  $W$  +jets,  $Z$  +jets and QCD multijet events to the  $2j1t$  (top),  $3j2t$  (middle) and  $2j1m0t$  (bottom) event categories.

## 8.4 Event simulation

The single top quark production in the  $t$  channel, the single top quark production in association with a W boson and the top quark pair production are generated at NLO accuracy using POWHEG and the NNPDF3.1 NNLO PDF set. For all three top quark production modes the top quark mass is set to  $m_t = 172.5 \text{ GeV}$ . The single top quark production in the  $t$  channel is simulated in the four flavor scheme (4FS) and the single top quark production in association with a W boson is simulated in the five flavor scheme. For the associated production of W and Z bosons with jets, the event generator MADGRAPH5\_AMC@NLO is used with the FxFx merging scheme. In 2017 and 2018 the NNPDF 3.1 NNLO PDF set is used and in 2016 the NNPDF3.0 NNLO PDF set. In all simulated events the parton shower is modeled using PYTHIA. For the underlying event, the tune CP5 [132] is used, except for W/Z-jets events in 2016 for which the tune CUETP8M1 [133] is used. A list of all used data and simulated samples can be found in appendix A.1.

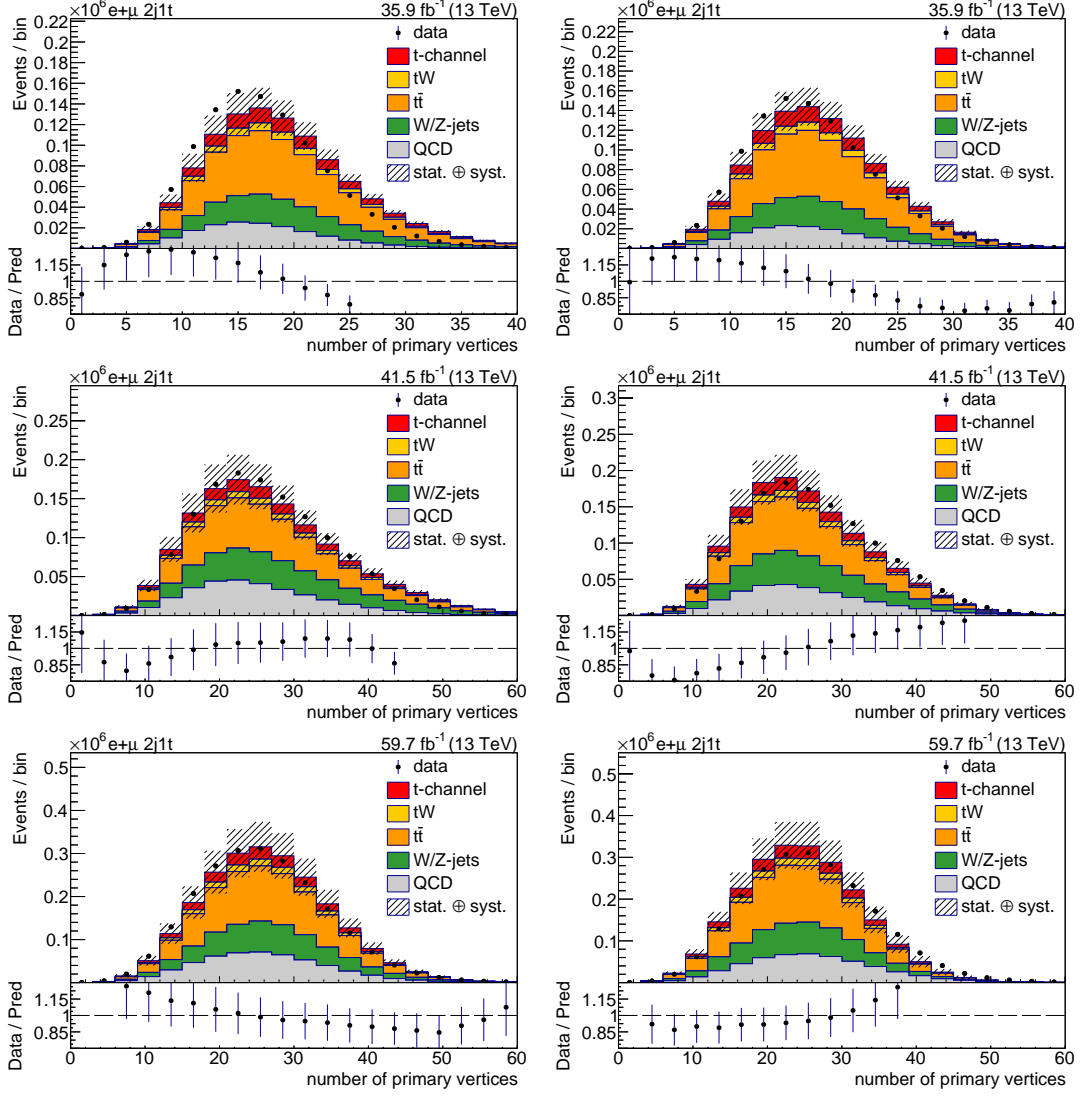
## 8.5 Corrections to simulated events

### 8.5.1 Number of pileup interactions

The number of simultaneous proton-proton collisions (pileup) can be estimated by multiplying the measured instantaneous luminosity with the total inelastic proton-proton cross section of  $69.2 \pm 3.2 \text{ mb}^{-1}$  [176]. The instantaneous luminosity varies between different runs and is therefore not exactly known beforehand. When simulating events, a pileup distribution has to be assumed. It is sampled from a Poisson distribution with the mean set to the expected number of pileup events. As simulation start before the end of data-taking, the pileup distribution in data and simulation normally do not match. Therefore, the simulated events are reweighted to match the pileup distribution measured in data. Figure 8.3 shows the number of primary vertices in simulation and data for each year of data-taking before and after applying the reweighting. As can be seen, the distributions do not agree well between data and simulation. After reweighting the agreement is slightly improved.

### 8.5.2 MET-phi modulation corrections

Due to the rotational symmetry of collisions around the beam axis, the distribution of the missing transverse momentum should be independent of the azimuthal angle  $\phi$ . However, the measured transverse momentum does depend on  $\phi$  and shows a sinusoidal modulation. This can be seen in figure 8.4. This modulation stems from effects such as anisotropic detector responses, inactive detector regions, detector misalignment or displacement of the beam spot. The amplitude of the modulation depends on the



**Figure 8.3: Effect of pileup reweighting.** Shown are the distributions of the primary vertices before applying the pileup reweighting (left) and after applying the pileup reweighting (right). The distributions are shown in the  $2j1t$  event category for data taken in 2016 (top), 2017 (middle) and 2018 (bottom). All other event weights are applied in both cases. The prediction is scaled to match the number of observed events.

number of pileup interactions. The modulation can be reduced by shifting the center of the coordinate system dependent on the number of primary vertices in the event. The corrections are derived for each run period and separately for data and simulation, as different modulations are observed in data and simulation. Figure 8.4 shows the  $\phi$  distribution of the missing transverse momentum after applying the MET-phi modulation corrections. As can be seen, the modulations are significantly reduced and the agreement between data and simulation is improved.

### 8.5.3 b tagging efficiencies

The efficiency of the b tagging algorithms differs between data and simulation. Therefore, event weights are applied to correct the simulated events. The event weights for an event with  $n$  b-tagged jets and  $m$  not-b-tagged jets are calculated as:

$$w = \frac{P(\text{Data})}{P(\text{MC})}, \quad (8.1)$$

where  $P(\text{Data})$  ( $P(\text{MC})$ ) is the probability to observe  $n$  b-tagged jets and  $m$  not-b-tagged jets in data (simulation). These probabilities are calculated for each event as:

$$P(\text{MC}) = \prod_{i=\text{b-tagged}} \epsilon_i \prod_{j=\text{not b-tagged}} (1 - \epsilon_j), \quad (8.2)$$

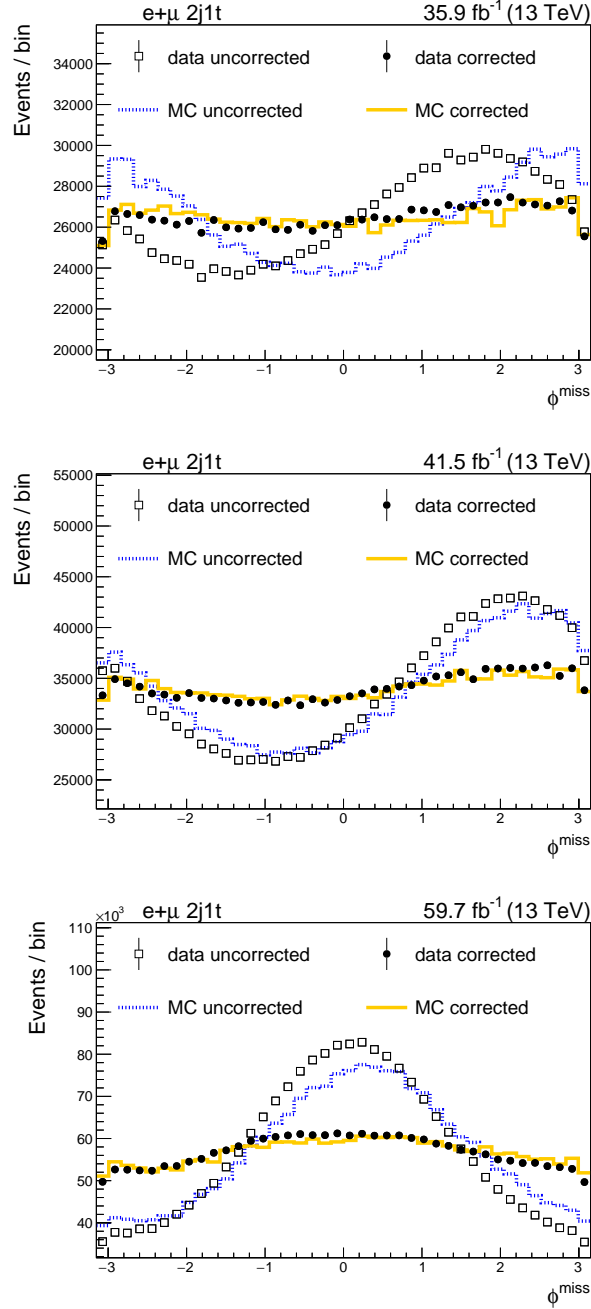
$$P(\text{Data}) = \prod_{i=\text{b-tagged}} \text{SF}_i \epsilon_i \prod_{j=\text{not b-tagged}} (1 - \text{SF}_j \epsilon_j). \quad (8.3)$$

Here  $\text{SF}_i$  are the scale factors for the jet flavor  $i$ . These scale factors are provided by the b Tag & Vertexing Physics Object Group within the CMS Collaboration [177]. The  $\epsilon_i$  are the b tagging efficiencies, which are derived from simulation for each process independently. They are calculated according to:

$$\epsilon_i = \frac{N_i^{\text{b-tagged}}}{N_i^{\text{total}}}, \quad (8.4)$$

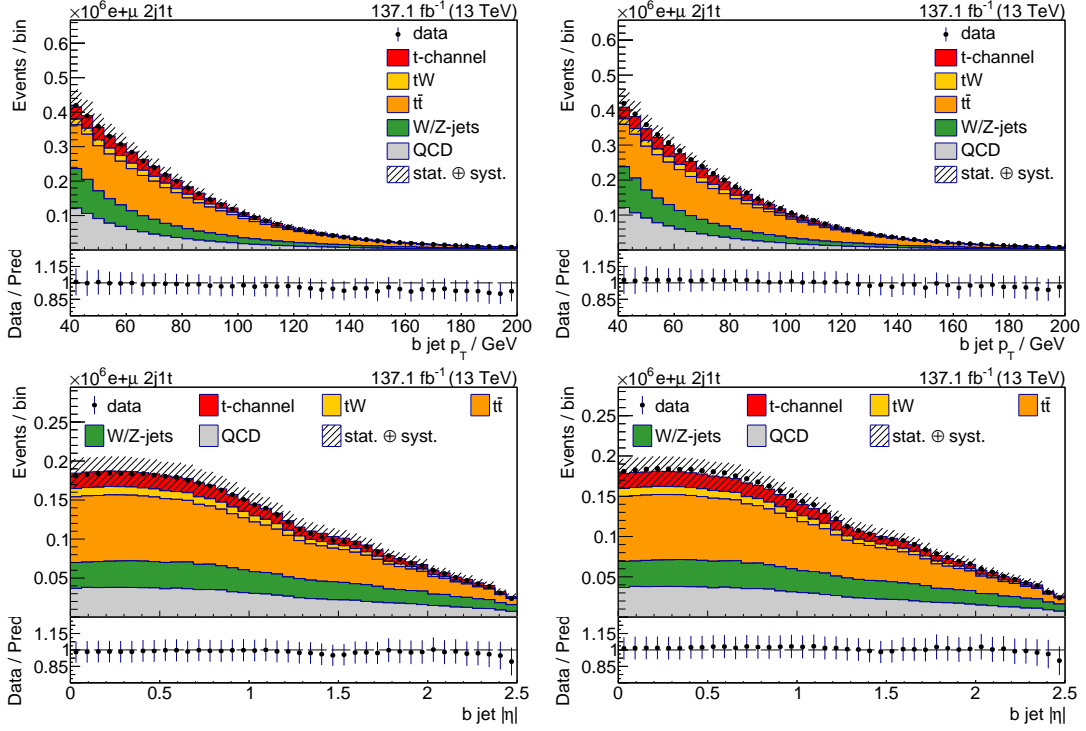
here  $N_i^{\text{b-tagged}}$  and  $N_i^{\text{total}}$  are the number of b-tagged jets and total number of jets with flavor  $i$ , respectively. The jet flavor is here the true flavor, which is known in the simulation.

The distributions of the transverse momentum and pseudorapidity of the reconstructed b jet associated to the top quark decay before and after applying the b tagging efficiency corrections are shown in figure 8.5.



**Figure 8.4: Effect of MET- $\phi$  modulation corrections** Shown are the azimuthal angle ( $\phi$ ) distributions of the missing transverse momentum before and after applying the MET- $\phi$  modulation corrections. The distributions are shown for data and simulation for 2016 (top), 2017 (middle) and 2018 (bottom) data sets. The uncorrected distributions show a sinusoidal modulation that is different between data and simulation. After applying the MET- $\phi$  modulation corrections the  $\phi$  distributions are mostly flat.





**Figure 8.5:  $b$  jet kinematic variables.** Shown are the distributions of the transverse momentum (top) and absolute value of the pseudorapidity (bottom) of the reconstructed  $b$  jet associated to the top quark decay. On the left the distributions are shown without applying the  $b$  tagging scale factors and on the right after applying the  $b$  tagging scale factors. All other event weights are applied in both cases.

### 8.5.4 Lepton efficiencies

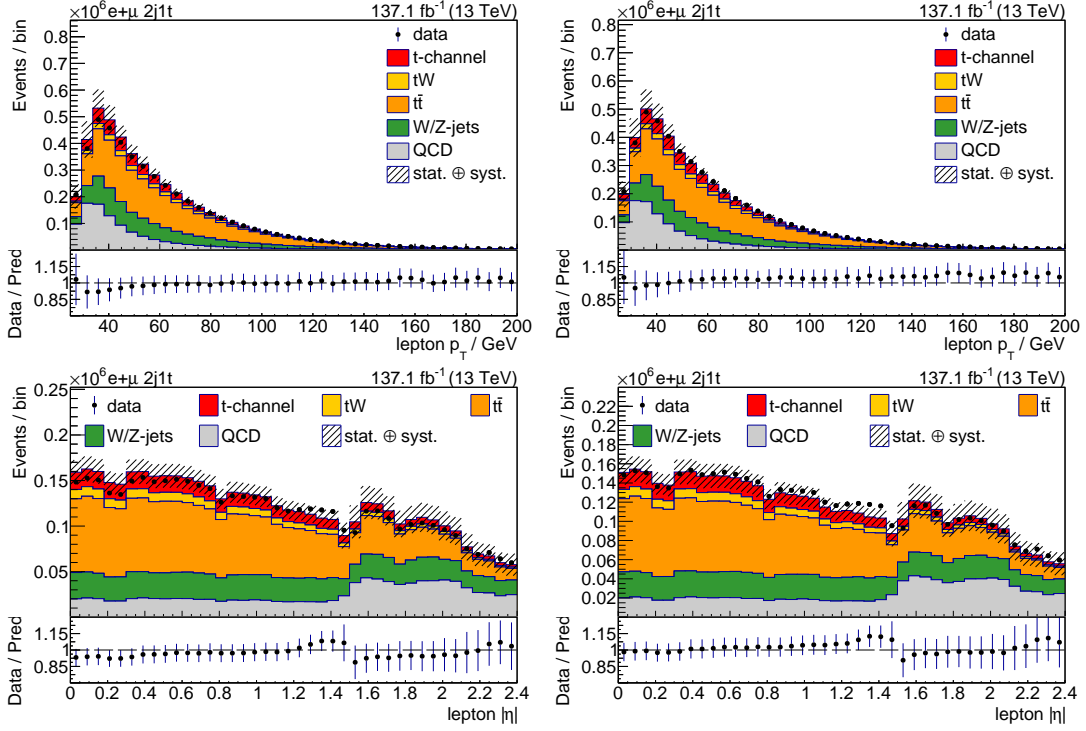
In the analysis events with a single lepton are selected, which need to pass the selection criteria of a single-lepton trigger. Additionally, only leptons with a certain ID, isolation and kinematic properties are considered (see also sections 8.3 and 6.3 and 6.4 for more details). Applying those selection criteria on the lepton influences the overall efficiency of the event selection. The kinematic variables and reconstruction efficiencies of leptons differ in simulation and data leading to different selection efficiencies and also shape differences in the distributions. Therefore, scale factors are applied to correct the simulation for these differences. For muons the total efficiency is the product of the efficiency of the muon ID, isolation and the trigger efficiency. It can be calculated according to:

$$\epsilon = \epsilon_{\text{ID}}^{\mu} \cdot \epsilon_{\text{Iso|ID}}^{\mu} \cdot \epsilon_{\text{Trigger|Iso}}^{\mu}. \quad (8.5)$$

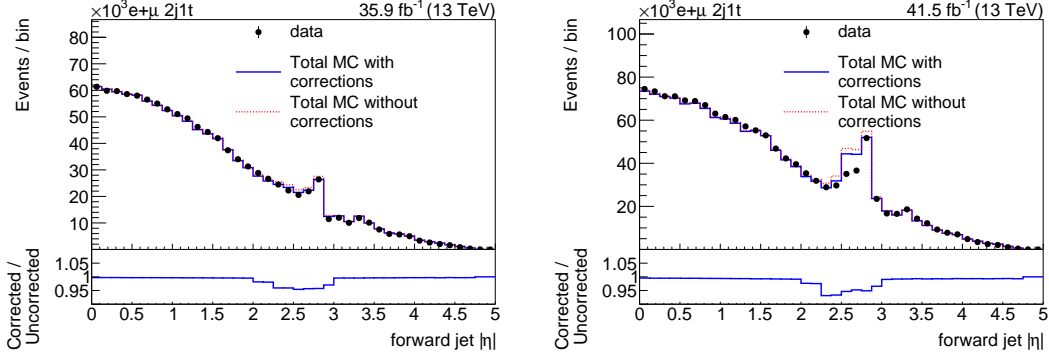
Each subsequent efficiency depends on the previous. The ID, isolation and trigger scale factors are derived using a tag-and-probe method in a dilepton data set close to the Z boson or  $J/\psi$  resonance. Scale factors are provided by the Muon Physics Object Group (Muon POG) of the CMS Collaboration [178]. This analysis uses a very tight muon selection criteria on the relative isolation of 0.06 in order to strongly suppress contributions from QCD multijet background. Isolation and trigger scale factors for this working point are not provided by the Muon POG. Instead the isolation and trigger scale factors for 2016 used by the inclusive and differential single-top quark  $t$ -channel production cross section measurements [174,175] are used, as these analysis use the same isolation cut. For 2017 scale factors derived in the search for the single-top  $s$ -channel production [179] are used, as the lepton selection criteria are the same as in this analysis. Muon isolation scale factors for 2018 have been derived specifically for this analysis. For details see appendix A.2.

Electron scale factors to correct for the identification and reconstruction efficiencies of electrons are provided by Electron/ $\gamma$  Physics Object Group of the CMS Collaboration [180]. They are derived using a tag-and-probe method in a dielectron data set near the Z boson resonance. As for muons, the electron trigger scale factors derived for the inclusive and differential single-top quark  $t$ -channel production cross section measurements [174,175] are used for 2016 data. For 2017 and 2018 the electron trigger scale factors used in the search for the single-top  $s$ -channel production [179] are used, as this analysis uses the same triggers and the same lepton selection criteria.

The effect of applying the lepton efficiency scale factors can be seen in figure 8.6, which shows the transverse momentum and pseudorapidity distributions of the reconstructed tight lepton in the 2j1t region.



**Figure 8.6: Lepton kinematical variables.** Shown are the distributions of the transverse momentum (top) and absolute pseudorapidity (bottom) of the reconstructed tight lepton. On the left the distributions are shown without applying the lepton scale factors and on the right after applying the lepton scale factors. All other event weights are applied in both cases.



**Figure 8.7: Effect of prefire reweighting.** Shown are the distribution of the pseudorapidity of the reconstructed spectator jet in the 2j1t region for the 2016 (left) and 2017 (right) data set. The dotted red line shows the simulated distribution before applying the L1 ECAL prefire-correction-weights and the solid blue line after applying the corrections. In both cases all other event weights are applied. The bottom panel shows the ratio between the corrected and uncorrected distribution, which corresponds to the average event weight. The prefiring issue affects mostly the region  $2 \leq |\eta| \leq 3$ .

### 8.5.5 Level-1 trigger ECAL prefiring

The Level-1 trigger ECAL prefiring issue was present in 2016 and 2017 data. The issue was that the gradual timing shift of the ECAL was not correctly propagated to the Level-1 trigger primitives (TP). A significant amount of high- $|\eta|$  TPs were therefore associated to the previous bunch crossing. This resulted in events self vetoing themselves, as Level-1 trigger rules forbid to fire in two consecutive bunch crossings. The vetoing occurred if a large amount of energy was deposited in the ECAL region of  $2 \leq |\eta| \leq 3$ . As this effect is not present in simulation, dedicated event weights are applied to correct for this effect. The event weights are calculated as the product of all objects in the event with the probability not to prefire:

$$w = 1 - P(\text{prefiring}) = \prod_{i=\text{photons, jets}} (1 - \epsilon^{\text{prefire}_i}). \quad (8.6)$$

The probabilities as a function of jet  $p_T$  and  $\eta$  are provided by the JetMET Physics Object Group of the CMS Collaboration [181]. The distribution of the pseudorapidity of the reconstructed spectator jet in 2016 and 2017 with and without the L1 ECAL prefiring corrections and for data are shown in figure 8.7. The difference between the corrected and uncorrected distribution are generally smaller than 5%. After applying the corrections the agreement between data and simulation is improved. However, especially for 2017 some discrepancies still remain.

## 8.6 QCD multijet background modeling

Although, the contributions from QCD multijet events can be reduced strongly by requiring at least one b jet and isolated leptons, this background process still remains important to the analysis due to its high cross section. The final event selection's efficiency of selecting QCD multijet events is only about  $10^{-8}$  to  $10^{-9}$ . Therefore, modeling the QCD multijet background with MC simulation would require to generate an unfeasible high number of events in order to achieve a sufficient modeling. Also, as simulation is not perfect, the modeling from MC simulation is not always reliable. Instead, the QCD multijet background is modeled using a data driven approach.

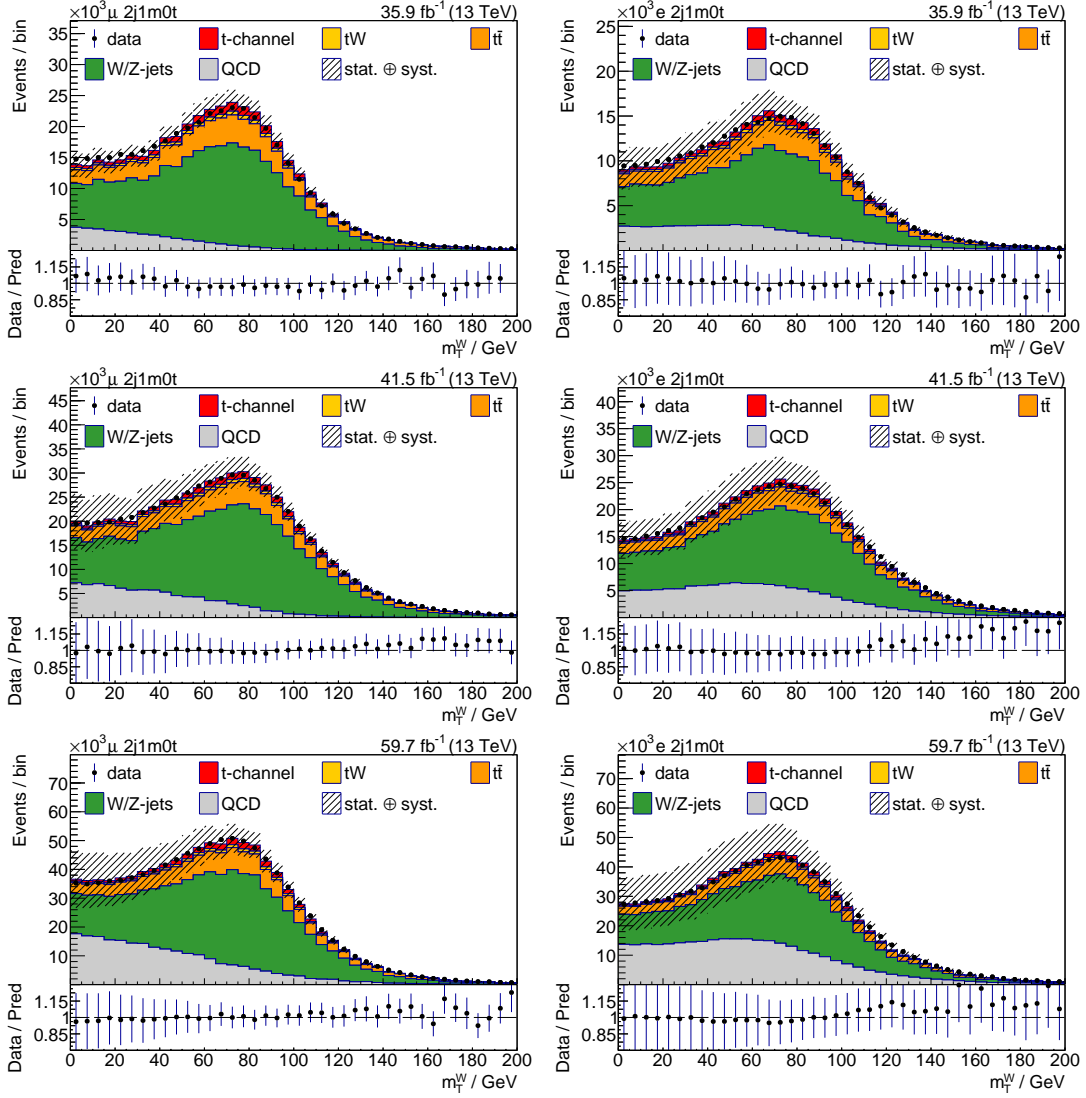
The QCD multijet background is estimated from a QCD-enriched sideband region in data. The definitions of the sidebands differ between events with muons and events with electrons. Anti-isolated tight muons are defined by inverting the isolation requirement on isolated tight muons (see section 8.3). They must be tight muons with  $I_{\text{rel.}}^{\mu} > 0.2$  and fulfill the same  $p_T$  and  $\eta$  selection criteria as isolated tight muons. In 2017 and 2018 the usage of isolated trigger paths has been found to cause a mismodeling in the anti-isolated muon's  $\phi$  distribution. Therefore, in 2017 and 2018 events in the muon sideband region must, instead of the isolated trigger paths, pass the trigger paths HLT\_Mu27 or HLT\_Mu20. This mismodeling has not been observed in 2016 and therefore the trigger requirements are not changed in 2016. The sideband region for muons is defined by requiring exactly one anti-isolated tight muon, and no additional loose muons with  $p_T > 10 \text{ GeV}$  and  $|\eta| < 2.4$ . Further, there must be no veto electrons and no electrons that pass the selection criteria for the electron sideband as defined in the following. For electrons simply inverting the isolation is not sufficient, as there can be multiple reasons for QCD multijet events passing the event selection criteria. For example an undetected photon conversion or wrong linkages between ECAL cells and tracks during the reconstruction. Instead, a sideband is defined by requiring the electrons to fail the cut-based electron veto identification criteria. Additionally, they must have a relative isolation  $I_{\text{rel.}}^e < 1$ . Events in the electron sideband must further have no veto electrons, no loose muons and no anti-isolated tight muons. The purity of QCD multijet events in the barrel region is estimated from simulation to be 91 % for muons and 88 % for electrons and in the endcap to be 94 % for muons and 92 % for electrons. The normalization of the QCD multijet background is extracted simultaneously with the number of signal events in a maximum-likelihood fit (see section 8.11).

The procedure of modeling the QCD multijet background from data is validated in the 2j1m0t event category. With about 15 % the relative amount of QCD multijet events in this category is similar to the amount in the signal region. Separately for muons and electrons and for each year of data-taking a fit is performed on the transverse mass  $m_T^W$  of the reconstructed W boson. For the fits two templates are constructed, one for the QCD multijet background  $Q$  that is taken from the corresponding sideband region and one for all other non-QCD processes  $B$ , which is taken from simulation in the 2j1m0t

event category. They are fitted to data using a simple fit with two parameters, one for the amount of QCD ( $N_{\text{QCD}}$ ) and one for the amount of non-QCD ( $N_{\text{non-QCD}}$ ):

$$F(m_{\text{T}}^{\text{W}}) = N_{\text{QCD}} \cdot Q(m_{\text{T}}^{\text{W}}) + N_{\text{non-QCD}} \cdot B(m_{\text{T}}^{\text{W}}). \quad (8.7)$$

The fit is performed in a QCD multijet enriched region, requiring  $m_{\text{T}}^{\text{W}} < 50 \text{ GeV}$ , and the result is extrapolated to the full analysis phase space. The distributions with the predictions scaled to the fit results are shown in figure 8.8 separately for events with muons and electrons and for each year of data-taking. Overall a good agreement between the prediction and data is observed. This fit is only performed for validation purposes, as the estimation of the QCD multijet background normalization is performed simultaneously with the signal extraction.



**Figure 8.8: Validation of the data-driven QCD multijet estimation.** Shown is the transverse W boson mass distribution in the 2j1m0t event category for events with muons (left) and events with electrons (right) for data taken in 2016 (top), 2017 (middle) and 2018 (bottom). The prediction is scaled to the result of a dedicated fit which has been performed separately for muons, electrons, and each year of data taking.

## 8.7 Top quark reconstruction

For the top quark reconstruction it is assumed that the top quark decays leptonically and that there is only one top quark in the event, as is the case for single top  $t$ -channel production. The top quark is reconstructed by adding up the four-momenta of the reconstructed decay product candidates. In the 3j2t event category there are two b-tagged jets that may be associated to the top quark decay. In all events the non b-tagged jet is associated to the spectator jet. Here, the one with the larger transverse momentum of the two is chosen as it is more likely to originate from the top quark decay and not from gluon-splitting. The charged lepton is either a muon or an electron coming from the decay of the W boson or from the decay of a  $\tau$ , which originated from the W boson. The neutrino is not observable in the detector as it is only interacting via the weak interaction. The presence of the neutrino results in missing transverse momentum  $p_T^{\text{miss}}$  in the event. It is assumed that the entire missing transverse momentum in an event is due to the neutrino. However, as the longitudinal component of the missing momentum is not known at a hadron collider only the transverse components of the neutrino momentum can be determined in this way. To determine the longitudinal component  $p_{\nu,z}$  of the neutrino momentum the mass of the W boson is fixed to its currently best known value of 80.385 GeV [7]. Further it is assumed that the neutrino and the charged lepton are massless. The W boson mass is given by:

$$m_W^2 = p_W^2 \quad (8.8)$$

$$= (p_\ell + p_\nu)^2 \quad (8.9)$$

$$= 2E_\ell E_\nu - 2\mathbf{p}_\ell \cdot \mathbf{p}_\nu \quad (8.10)$$

$$= 2E_\ell E_\nu - 2p_{\ell,T}p_{\nu,T} \cos \Delta\phi - 2p_{\ell,z}p_{\nu,z}, \quad (8.11)$$

here  $\Delta\phi$  is the azimuthal angle difference between the charged lepton and the neutrino. The energy of the neutrino is given by  $E_\nu = \sqrt{p_{\nu,T}^2 + p_{\nu,z}^2}$ . Squaring equation (8.11), a quadratic equation in  $p_{\nu,z}$  is obtained, which has the two solutions:

$$p_{\nu,z}^\pm = \frac{1}{p_{\ell,T}} \left( \mu p_{\ell,z} \pm E_\ell \sqrt{\mu^2 - p_{\ell,T}^2 p_{\nu,T}^2} \right), \quad (8.12)$$

Where  $\mu$  is defined as:

$$\mu = \frac{m_W^2}{2} + p_{\ell,T}p_{\nu,T} \cos \Delta\phi. \quad (8.13)$$

In about one third of the cases the radicand  $R = \mu^2 - p_{\ell,T}^2 p_{\nu,T}^2$  in equation 8.12 becomes negative, resulting in complex-valued solutions. In this case it is assumed that the



missing transverse momentum is misreconstructed and therefore  $p_{\nu,x}$  and  $p_{\nu,y}$  are varied independently until the radicand becomes zero, while still keeping the W boson mass fixed [182]. It is tried to leave the variations with respect to the measured transverse momentum small, meaning the difference:

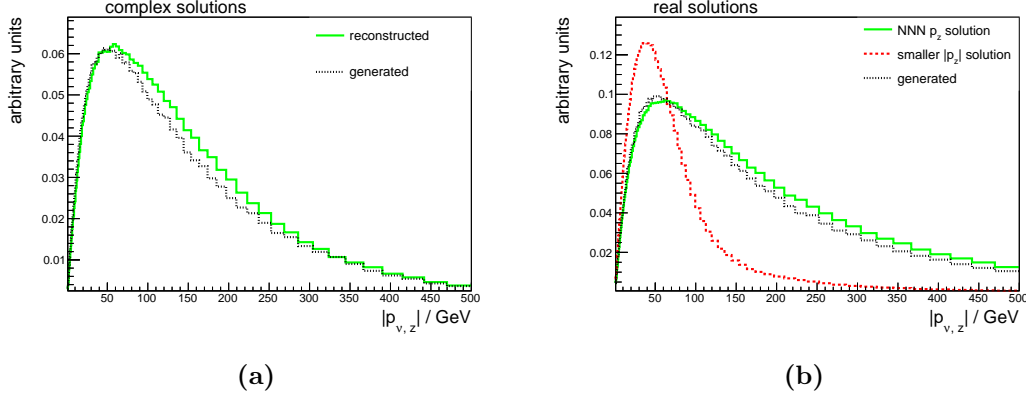
$$\delta = \sqrt{(\cancel{E}_x - p_{\nu,x})^2 + (\cancel{E}_y - p_{\nu,y})^2} \quad (8.14)$$

should be minimal. A comparison between the reconstructed and generated  $p_{\nu,z}$  for simulated single top quark  $t$ -channel events with complex valued solutions is given in figure 8.9a. In two third of the events the radicand in equation 8.12 is positive resulting in two positive solutions of which one has to be chosen. The common choice used in previous analyses is to always pick the solution that yields the smaller absolute value for  $p_{\nu,z}$ . As  $p_{\nu,z}$  follows a steeply falling distribution this gives a better description on average than always picking the larger of the two solutions. However, as can be seen in figure 8.9b choosing always the smaller solutions results in a  $p_{\nu,z}$  distribution that is significantly softer than the generated distribution, as the criterion biases the solution towards smaller values. Instead of always picking the smaller of the two solutions in this thesis the choice is based on the output of a neural network. The specifics of the setup and training of the neural network discriminator will be discussed in more detail below. The  $p_{\nu,z}$  distribution when picking the solution based on the output of the neural network discriminator is also shown in figure 8.9b. Using this approach a good description of the  $p_{\nu,z}$  distribution is achieved. Picking the  $p_{\nu,z}$  solution based on the output of the discriminator also results in a better reconstruction of the top quark, especially the pseudorapidity of the top quark, as is shown in figure 8.10. In addition the reconstruction of the top quark mass in simulated single top  $t$ -channel events is found to be improved with this approach. The reconstructed top quark mass, transverse momentum and pseudorapidity in the 2j1t region and 3j2t  $t\bar{t}$ -enriched event categories are shown in figure 8.11. The modeling in the W/Z-jets-enriched 2j1m0t event category is shown in appendix A.3.

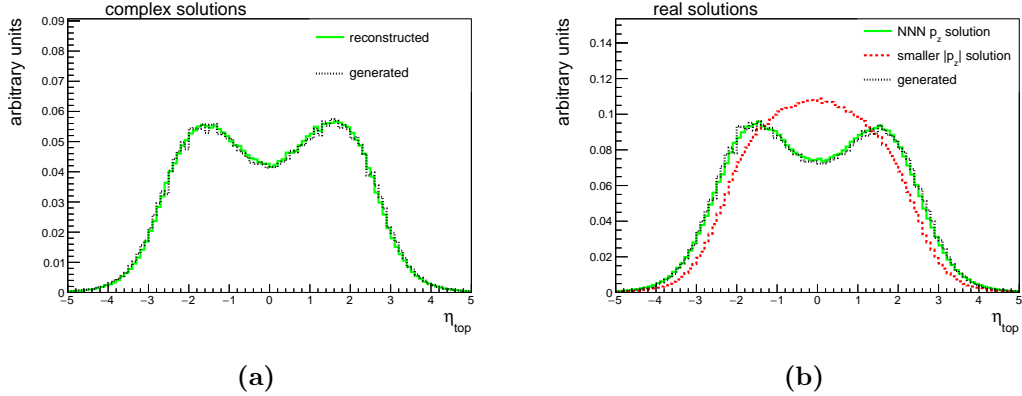
The neural network used in this thesis to choose one of the two  $p_{\nu,z}^{\pm}$  solutions is a simple feed-forward neural network with two hidden layers and 100 neurons per layer. The neural network is used as a discriminator, it classifies each of the two solutions  $p_{\nu,z}^{\pm}$  as being "correct" or "wrong". The neural network is in the following referred to as the neutrino neural network (NNN).

For the training of the NNN, only simulated single top  $t$ -channel events where equation (8.12) results in two real solutions are used. The training events are randomly chosen from all matching simulated single top  $t$ -channel samples, including samples that correspond to systematic variations. Only a small subset of about 1.5% of all simulated single top  $t$ -channel events are used for the training. From those events 20% are used for validation purposes. A single discriminator is trained for the full Run 2 data set. For the training the *TensorFlow* [183] framework is used.

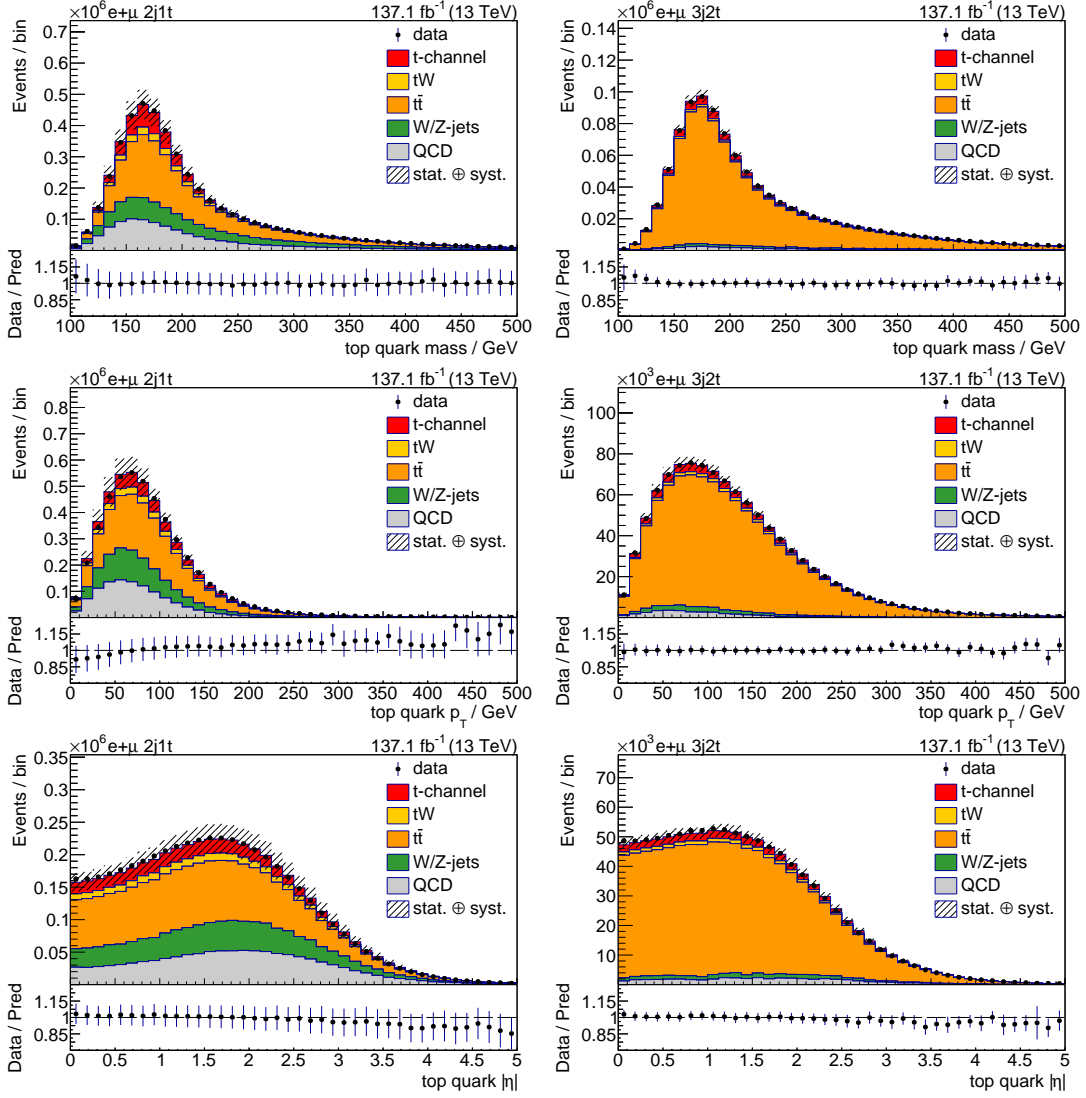
The correct solution is defined as the solution that minimizes the absolute difference in pseudorapidity between the generated and reconstructed neutrino. The pseudorapidity  $\eta$  and the transverse and longitudinal momentum  $p_T$  and  $p_z$  of any particle are related to each other via the relation:



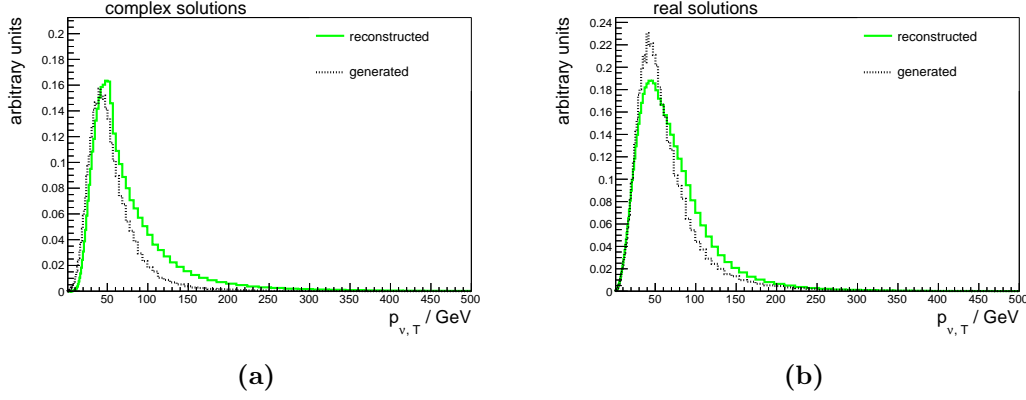
**Figure 8.9: Longitudinal neutrino momentum.** Generated and reconstructed distribution of the longitudinal component of the neutrino momentum  $p_{\nu,z}$  for simulated single top  $t$ -channel events with a leptonic W boson decay. The left figure (a) shows the distribution for events which have complex solutions of equation (8.12). The right figure (b) shows the distribution for events in which two real solutions are obtained. Different methods of choosing one of the two solutions are compared: choosing always the solution with the smaller absolute value and choosing the solution based on the output of a neural network (NNN).



**Figure 8.10: Comparison of top quark pseudorapidity.** Shown are the generated and reconstructed distribution of the pseudorapidity of the top quark for simulated leptonic single top  $t$ -channel events. Figure (a) shows the distribution for events which have complex solutions of equation (8.12). Figure (b) shows the distribution for events for which two real solutions are obtained. Two different methods of reconstructing the longitudinal momentum of the neutrino are compared. Either picking always the smaller solution in equation (8.12) or by choosing the solution based on the output of a neural network (NNN).



**Figure 8.11: Reconstructed top quark kinematical variables.** The reconstructed top quark mass (top), transverse momentum (middle) and absolute pseudorapidity (bottom) are shown in the 2j1t (left) and 3j2t (right) event categories. The hashed area corresponds to the combined statistical and systematic uncertainties of the prediction. The prediction is scaled to match the data.



**Figure 8.12: Transverse neutrino momentum.** Generated and reconstructed distribution of the transverse component of the neutrino momentum  $p_{\nu,T}$  for simulated single top  $t$ -channel events with a leptonic W boson decay. The left figure (a) shows the distribution for events which have complex solutions of equation (8.12). The right figure (b) shows the distribution for events for which two real solutions are obtained.

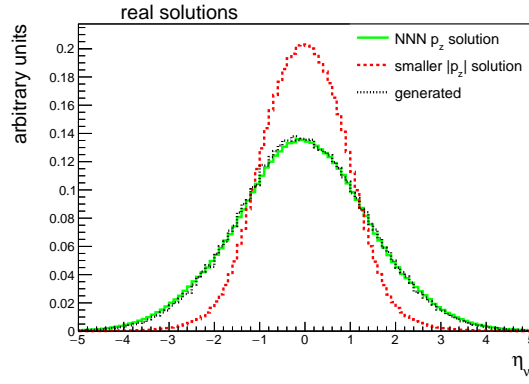
$$p_z = p_T \cdot \sinh(\eta) \quad (8.15)$$

From this equation the  $\eta_{\nu}^{\pm}$  values corresponding to the two real  $p_{\nu,z}^{\pm}$  can be computed. If  $\eta^{\text{gen}}$  is the pseudorapidity of the generated neutrino, then the differences  $\delta^{\pm}$  are defined as:

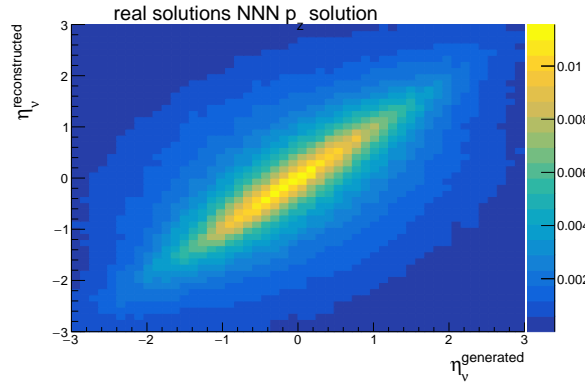
$$\delta^{\pm} = |\eta^{\text{gen}} - \eta_{\nu}^{\pm}| \quad (8.16)$$

The solution with the smaller  $\delta$  is then chosen as the correct solution. Using this figure of merit a good agreement between both the reconstructed and generated  $p_z$  and  $\eta$  distributions is observed. The measure could also be defined in terms of minimizing the difference in  $p_z$  instead of  $\eta$ . However, doing so results in a classifier that favors smaller values of  $\eta$  and therefore does not describe the pseudorapidity distribution of the neutrino as well as a classifier that has been trained minimizing the differences in pseudorapidity. This behaviour can be understood by recalling that differences in  $p_z$  can arise from either differences in  $p_T$  or in  $\eta$ . In figure 8.12 the generated and reconstructed  $p_T$  distributions of the neutrino are shown. As can be seen the reconstructed  $p_T$  distribution tends towards larger values than the generated one. A classifier that is trained with the objective to find the solution that minimizes the differences in  $p_z$  will therefore try to compensate too large values of reconstructed  $p_T$  by picking smaller  $\eta$  values.

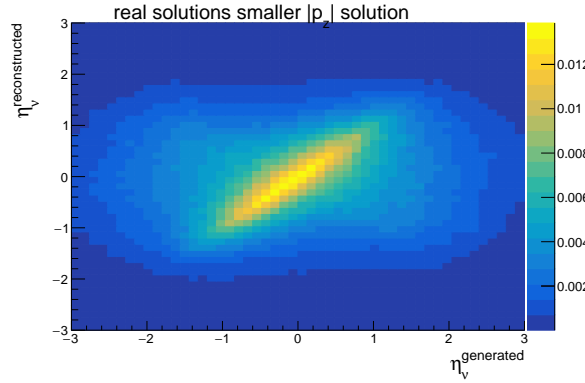
For the training of the NNN a total of 14 input variables are used, which are listed in table 8.3. The variables are sorted according to their impact on the classification result



(a)



(b)



(c)

**Figure 8.13: Neutrino Pseudorapidity.** Generated and reconstructed distribution of the pseudorapidity of the neutrino for simulated single top  $t$ -channel events with a leptonic W boson decay. The figures show only events with two real solutions of equation (8.12). Figure (a) shows the 1D distribution of the generated neutrino pseudorapidity, as well as the neutrino pseudorapidity reconstructed using the NNN and by picking always the smaller of the  $|p_{\nu,z}^{\pm}|$  solutions. Figure (b) shows the 2D distribution of the generated and the reconstructed pseudorapidity, where the reconstruction is done using the NNN. Figure (c) is done reconstructing the neutrino by always picking the solution with the smaller absolute value.

of the NNN. To evaluate the impact of each variable first an inference is run on a data set. From the predictions of the NNN, the AUC value of the ROC  $a$  is computed. Then the inference is repeated on the same data set 14 times — once for each variable — and the values of the corresponding variable is replaced with random numbers drawn from a uniform distribution between the smallest and largest value of the variable. The smaller the AUC value  $a_i$  when replacing variable  $i$  with random noise, the more important is the variable for the performance of the NNN. The ranking  $r_i$  for variable  $i$  is computed as the relative difference in AUC value:  $r_i = \frac{a - a_i}{a}$ . The four input variables with the highest rank are the energy and pseudorapidity of the b jet from the top quark decay and the pseudorapidities corresponding to the two solutions of equation (8.12).

The model achieves an accuracy of about 70 % for picking the correct solution and an AUC value of 0.78 on the validation data set. Figure 8.13 compares the generated and reconstructed pseudorapidity distributions of the neutrino. If the  $p_{\nu,z}$  is reconstructed by picking the solution with the larger NNN output the 1D distributions show a good agreement and there is a clear correlation between the reconstructed and generated distribution as can be seen from the scatter plot 8.13b. Choosing always the smallest solution results in  $\eta$  spectrum that is more biased towards zero. Also there is a weaker correlation between the reconstructed and generated distribution as visible in figure 8.13c. The reconstructed pseudorapidity of the top quark for real and complex solutions is shown in figure 8.10. For events with complex solutions the agreement between the reconstructed and generated solution is good. For real solutions, picking always the solution with smaller  $|p_{\nu,z}|$  results in a reconstructed shape that significantly differs from the generated one. Choosing the solution based on the NNN output leads to a significantly better agreement between the reconstructed and generated solutions.

Using the NNN may result in a bias as it could learn the underlying distribution during training and then only pick solutions that are biased towards the SM. To study a possible bias, the model was retrained while reweighting the pseudorapidity distribution such that it follows a uniform distribution. This model showed no significant deterioration in performance compared to the nominal model. The nominal model was further applied to simulated  $t$ -channel single top quark samples that had been generated with non-SM settings and had a top quark pseudorapidity distribution that is different from the SM prediction. Also in this case the reconstructed distribution using the NNN agreed well with the generated one and no bias towards the SM distribution was observed in the non-SM samples.

**Table 8.3: Input variables used for the neutrino neural network.** The input variables are sorted according to the relative difference in the AUC value, which is obtained if the variable is replaced with random noise for computing the predictions.

Rank	Variable	Relative difference in AUC
1	energy of the b jet associated to the top quark decay	0.28
2	$\eta$ of the b jet associated to the top quark decay	0.23
3	$\eta_\nu^+$ solution of equation (8.12)	0.22
4	$\eta_\nu^-$ solution of equation (8.12)	0.22
5	$\eta$ of the charged lepton	0.22
6	$p_T$ of the charged lepton	0.17
7	$p_T$ of the b jet associated to the top quark decay	0.16
8	missing transverse momentum	0.13
9	$\phi$ of the b jet associated to the top quark decay	0.13
10	$\phi$ of the charged lepton	0.09
11	$\phi$ of the missing transverse momentum	0.06
12	flavor and charge of the charged lepton	0.03
13	mass of the charged lepton	0.02
14	b tagging score of the b jet associated to the top quark decay	< 0.01

## 8.8 Event classification

The event selection is optimized to select a large number of single top  $t$ -channel events while rejecting most background contributions. However in the signal region, single top  $t$ -channel production still contributes only about 10 % to the total number of selected events. The signal purity can be increased using machine learning techniques. For the signal extraction it is not only useful to have regions dominated by the signal process, but also to have background-dominated regions that can be used by a ML fit to constraint each background process individually. Therefore, one multiclassification BDT is trained to categorize events into four process categories: single top  $t$ -channel events, top background events — which includes  $t\bar{t}$  and  $tW$ —,  $W/Z$ -jets events and QCD multijet events. The output of the BDT are four numbers  $BDT_c \in [0, 1]$ , one for each of the four process categories. Those numbers are correlated with the probability of an event to belong to a given process. The larger the BDT output for a process category the larger the probability that the event belongs to this process. A BDT is chosen for this task instead of a neural network, as a low number of input variables is used and BDTs are very robust discriminators.

In the following the design, training and evaluation procedure of the BDT is elaborated in more detail. The hyperparameters and input variables of the BDT are chosen such as to achieve descent discrimination power while being robust, avoiding overtraining and having a low correlation with the variables as a function of which a differential cross section measurement is performed; namely the top quark  $p_T$  and rapidity and the three angular polarization variables  $\cos(\theta_x)$ ,  $\cos(\theta_y)$  and  $\cos(\theta_z)$ , which are defined in section 8.12.1. For the training of the BDT the TMVA [184] framework is used. The BDT consist of 1000 shallow trees with a maximum depth of three and a minimum

node size of 5 %, no pruning is used during the training. The cross-entropy is used as loss function and gradient boosting is employed as boosting algorithm. For the training events from the 2j1t event category are used. As events with negative event weights can negatively effect the boosting during the training, only events with positive event weights are considered for the training data set. This is mostly relevant for W/Z-jets as the W/Z-jets simulated samples are generated with AMC@NLO that use negative event weights to account for the differences between LO and NLO. Around 1 % of the total number of simulated events are used for training. For QCD multijet, data from the sideband region is used for the training. All events are reweighted during the training such that there is an equal amount of events for each process category. Of the chosen events 70 % are used for the training and 30 % for the testing data set.

A good selection of input variables is of critical importance for the performance of the discriminator. Only variables that are well modeled and that show a good agreement between data and prediction are considered as input variables. Choosing badly modeled input variables would result in worse performance on data than on simulation and an inability of the model to generalize to unknown data. A second important consideration is that the variables should be only weakly correlated to the variables in which a differential cross section measurement is performed, i.e., the three polarization angles and the top quark transverse momentum and rapidity. This reduces the possible bias of the BDT on the measurement. In total six input variables are used for the training of the BDT all of them fulfill above requirements. The variables are listed in table 8.4.

One of the six variables is the event shape. It is a measure of the three-jet structure of an event and is computed from the eigenvectors  $\lambda_i$  of the momentum tensor  $S^{ab}$ :

$$S^{ab} = \frac{\sum_i^{\text{jets}, \ell, \mathbf{p}_T^{\text{miss}}} p_i^a p_i^b}{\sum_i^{\text{jets}, \ell, \mathbf{p}_T^{\text{miss}}} |\mathbf{p}_i|^2} \quad (8.17)$$

$$\text{event shape} = 3(\lambda_1 \lambda_2 + \lambda_1 \lambda_3 + \lambda_2 \lambda_3), \quad (8.18)$$

Here  $p_i^a$  is the  $a$ -th three-momentum component of particle  $i$ . The eigenvectors  $\lambda_i$  are ordered in descending order and are normalized such that  $\lambda_1 + \lambda_2 + \lambda_3 = 1$ . Figures 8.14 and 8.15 show the distributions of the input variables for the different process and also compares the total distributions from simulation to data in the 2j1t event category. In appendix A.4 the modeling of the input variables in the  $t\bar{t}$ -enriched 3j2t and W/Z-jets-enriched 2j1m0t event categories are shown. In general good agreement between data and simulation is found. Figure 8.16 shows the correlation between the BDT input and output variables and among the polarization angles, as well as the top quark  $p_T$  and  $y$ . All input variables as well as the outputs of the BDTs, exhibit low correlation with the polarization angles and the top quark transverse momentum and rapidity. The largest one being between  $\cos(\theta_z)$  and  $\Delta R(\text{light jet}, \text{b jet})$  with a correlation of 0.28.

The output distributions of the BDT for each process category and for the training and testing data set are shown in figure 8.17. There is generally good agreement in the output distributions of the training and testing data set, which allows to conclude that no overtraining occurred. As can be seen, good discrimination power for  $t$ -channel,  $t\bar{t}$  +tW and QCD against the corresponding remaining processes can be achieved, with



**Table 8.4: BDT input variables.** Listed are the input variables used for the training of the multiclassification BDT. The variables are sorted by their importance during the training.

Description	Variable	Score in training
Absolute value of the pseudorapidity of the spectator jet	$ \eta_{j'} $	0.19
$\Delta R$ between the momentum vectors of the spectator jet and the b-tagged jet associated to the top quark decay.	$\Delta R$ (light jet, b jet)	0.18
Absolute difference in the pseudorapidity of the charged lepton and the b-tagged jet associated to the top quark decay	$ \Delta\eta(\ell, \text{b jet}) $	0.17
See equation (8.17)	event shape	0.17
Invariant mass of the top quark reconstructed from the charged lepton, the neutrino, and the b-tagged jet associated to the top quark decay	$m_t$	0.15
Transverse mass of the reconstructed W boson	$m_T^W$	0.14

**Table 8.5: Confusion matrix of the multiclassification BDT.** The rows show the true categories and the columns the predicted event categories. The number in row  $i$  and column  $j$  gives the relative amount of events that are predicted by the BDT as event category  $j$  but really belong to event category  $i$ . The numbers are given relative to the number of true events, such that numbers in each row sum up to one.

truth \ predicted	$t$ -channel	$t\bar{t} + tW$	W/Z-jets	QCD multijet
$t$ -channel	64 %	15 %	7 %	14 %
$t\bar{t} + tW$	17 %	54 %	12 %	17 %
W/Z-jets	18 %	26 %	32 %	24 %
QCD multijet	15 %	14 %	14 %	57 %

AUC values of 0.81, 0.73 and 0.72 respectively. The classification into W/Z-jets is the most difficult one with an AUC value of only 0.66. Table 8.5 shows the confusion matrix of the multiclassification BDT. The confusion matrix shows how large the migrations are among the different classes. As can be seen from the table 64 % of all signal events are correctly classified as signal events. Most misclassified signal events are classified as  $t\bar{t}$  or  $tW$  events (15 %) or as QCD multijet events (14 %). Only 7 % of the signal events are wrongly classified as W/Z-jets events. Figure 8.18 compares the output distributions of the BDTs between simulation and data in the 2j1t event category. The BDT output distributions in the  $t\bar{t}$ -enriched 3j2t and W/Z-jets-enriched 2j1m0t event categories can be found in appendix A.4. Overall good agreement between data and prediction is found.

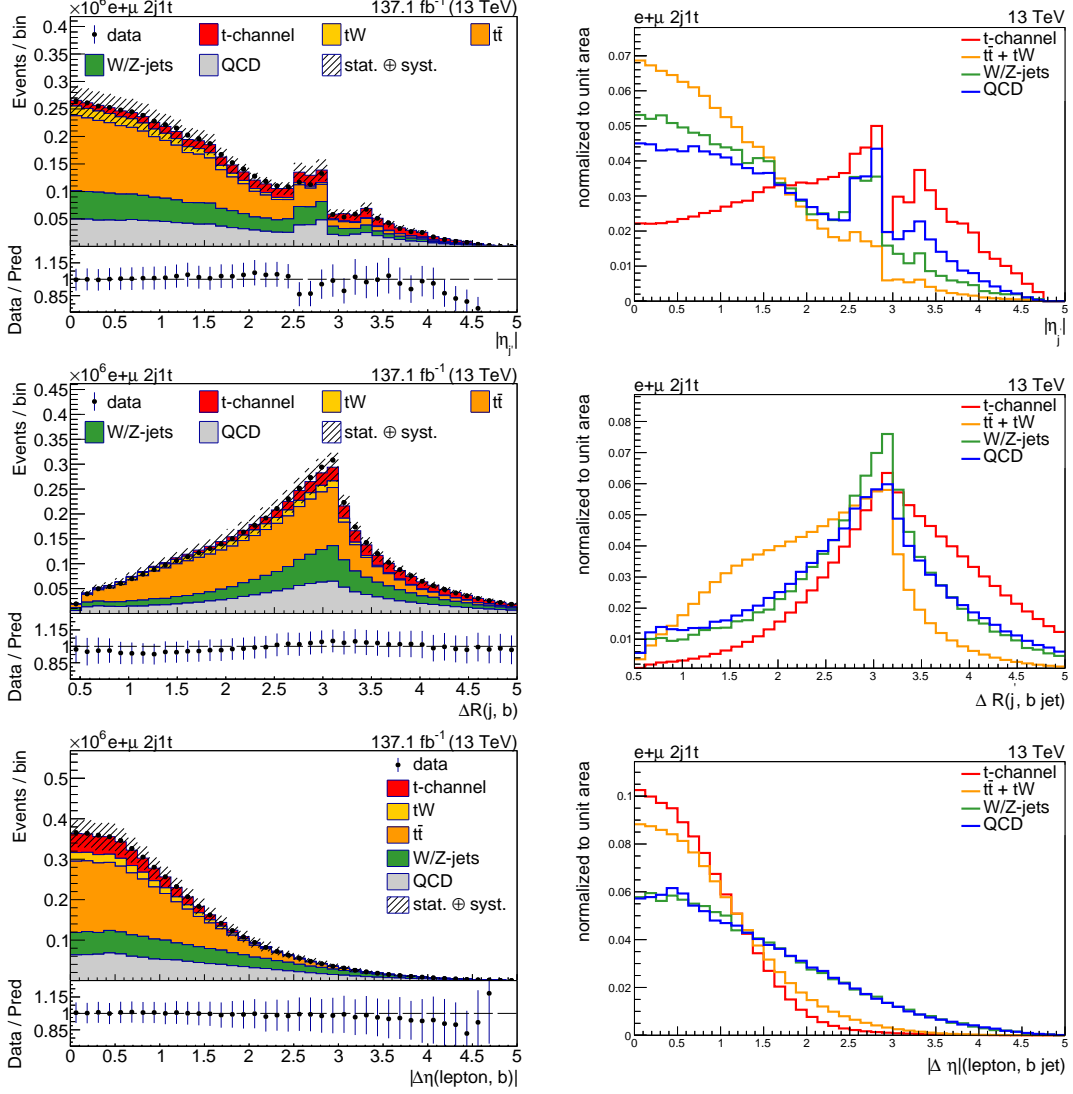
From the four different output nodes of the BDT a single output  $BDT_S$  is defined as follows. Of the four output nodes of the BDT the output node  $BDT_{\max}$  with the largest value is chosen and the corresponding process category is called  $c_{\max}$ . Based on  $c_{\max}$  an offset for each category is defined:

$$\text{offset} = \begin{cases} 0, c_{\max} = \text{QCD multijet} \\ 1, c_{\max} = \text{W/Z-jets} \\ 2, c_{\max} = t\bar{t} \text{ and } tW \text{ background} \\ 3, c_{\max} = \text{single top } t\text{-channel} \end{cases} \quad (8.19)$$

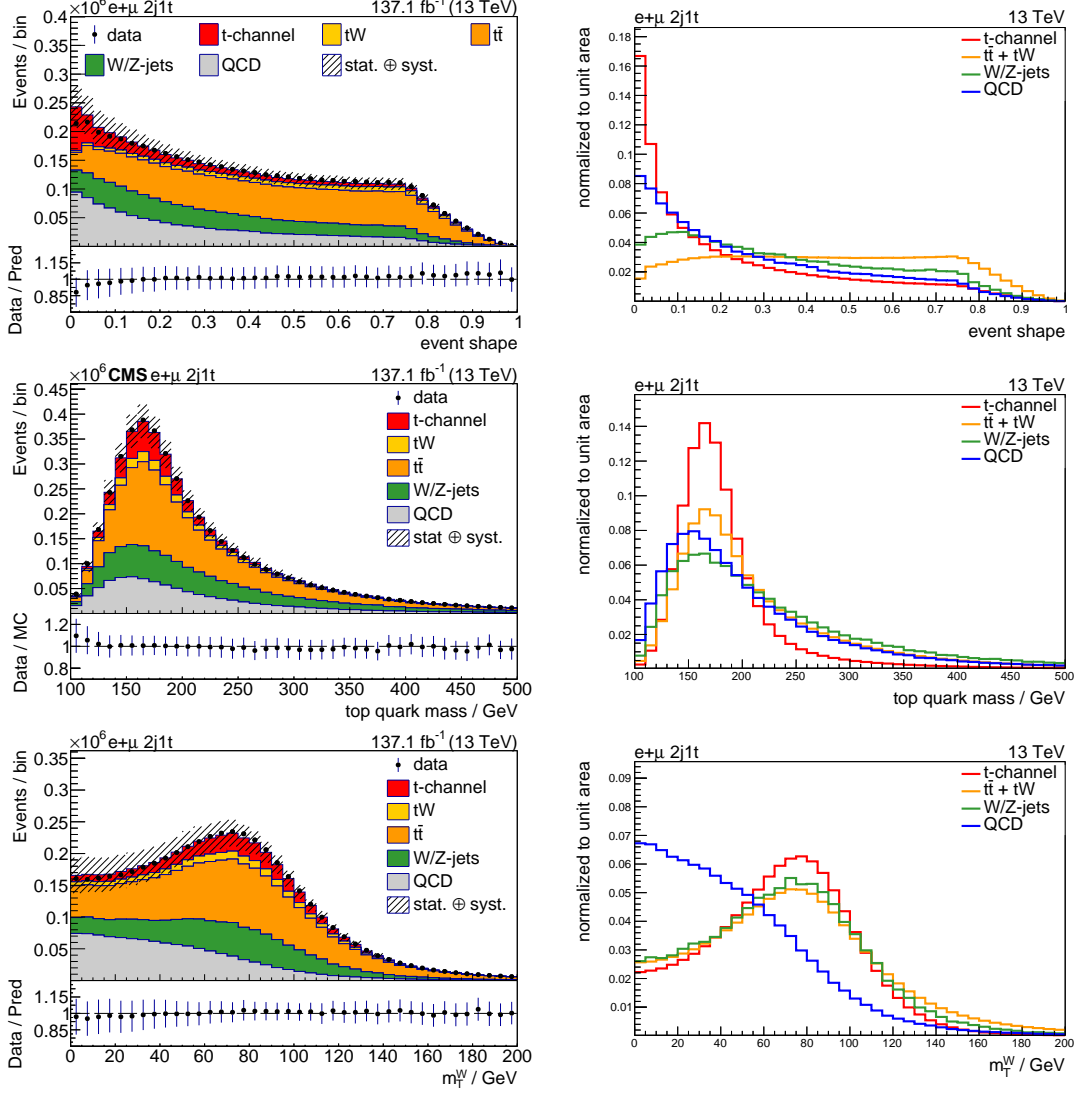
For example, if the output node that corresponds to single top  $t$ -channel events is the largest, which means the event has a high probability to be a single top  $t$ -channel event, then the offset is 3. The final BDT discriminator value is calculated as:

$$BDT_S = BDT_{\max} + \text{offset}(c_{\max}) \quad (8.20)$$

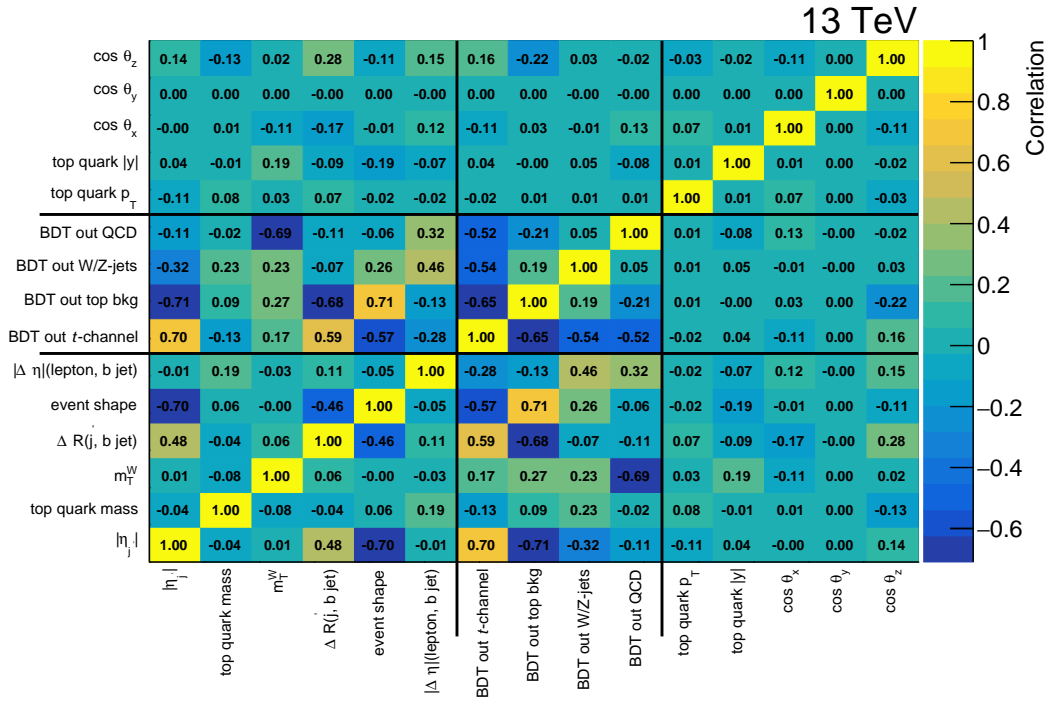
The result is a discriminator in which the values from 0 to 1 are enriched in QCD multijet events, the values from 1 to 2 in W/Z-jets events, the values from 2 to 3



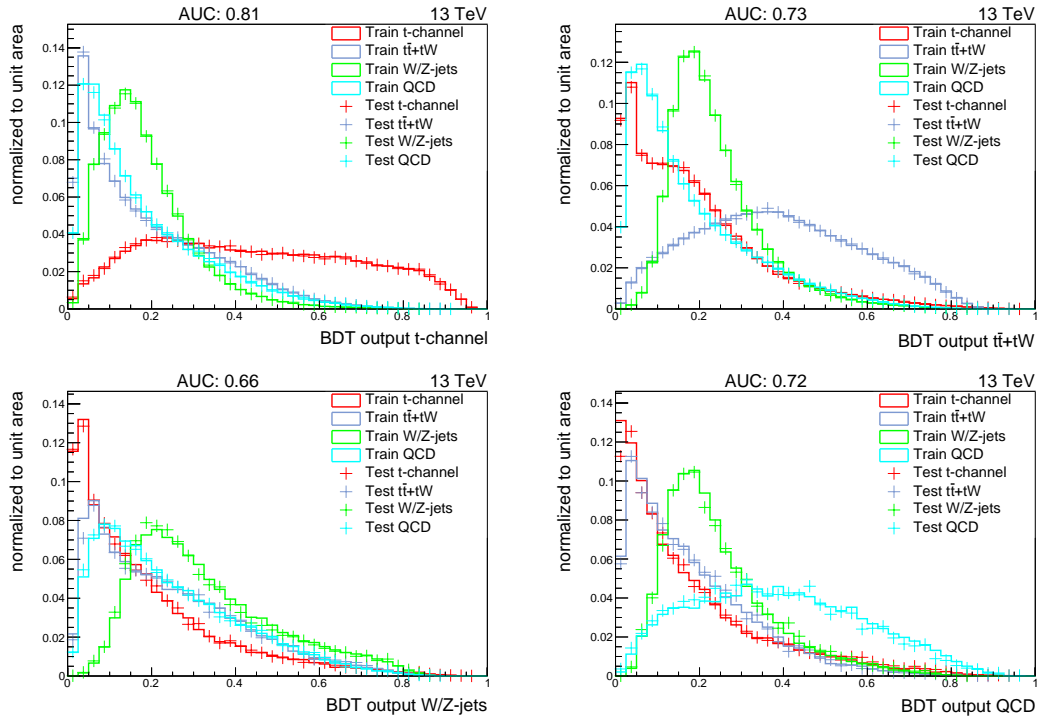
**Figure 8.14: Modeling of BDT input variables.** Shown are the shapes for the three highest ranked input variables used as input for the BDT. The left figures show the comparison between data and prediction for each input variable. The prediction is scaled to match the amount of data. The right figures show the distributions for the different processes for each input variable where each process is normalized to 1. The figures correspond to the full Run 2 data set in the 2j1t event category.



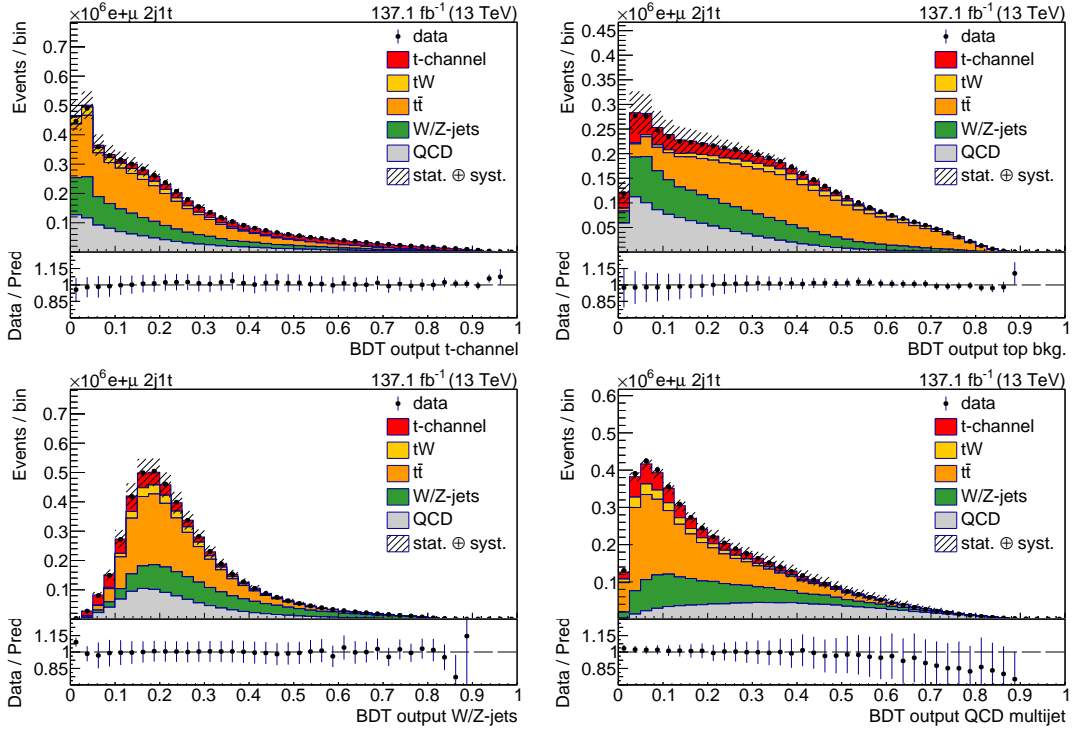
**Figure 8.15: Modeling of BDT input variables.** Shown are the shapes for the three lowest ranked input variables used as input for the BDT. The left figures show the comparison between data and prediction for each input variable. The prediction is scaled to match the amount of data. The right figures show the distribution for the different processes for each input variable where each process is normalized to 1. The figures correspond to the full Run 2 data set in the 2j1t event category.



**Figure 8.16: BDT Correlations.** Correlation between the BDT input variables, the BDT outputs, the top quark transverse momentum and rapidity, and the three polarization angular variables.

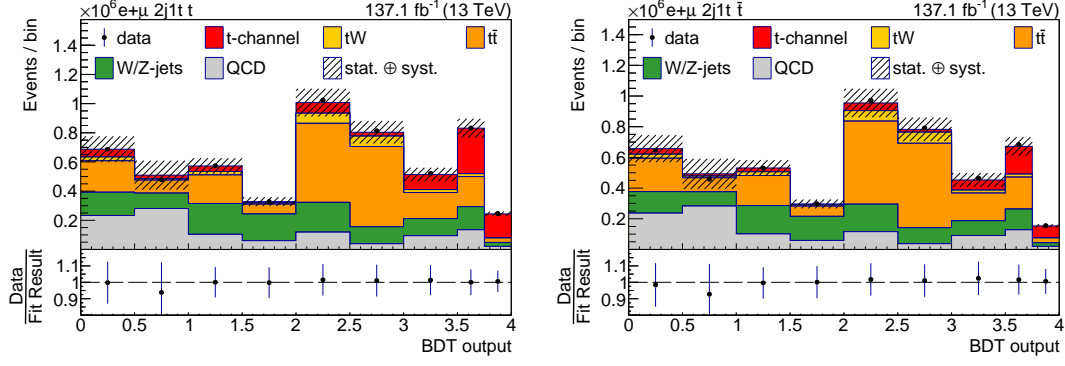


**Figure 8.17: BDT output distributions.** Shown are the distribution for each of the BDT output nodes for each process for both the training and the testing data set. All distributions are normalized to unit area.



**Figure 8.18: Comparison of the BDT output distributions to data.**

The top left figure shows the output distribution for the  $t$ -channel node, the top right figure the  $t\bar{t}$  and  $tW$  node, bottom left the  $W/Z$ -jets node and bottom right the QCD multijet node. The prediction is scaled to match the amount of data.



**Figure 8.19: Prefit  $BDT_S$  distributions.** Shown are the distributions of  $BDT_S$  for events with positively charged leptons (left) and negatively charged leptons (right). The distribution corresponds to events in the 2j1t event category. The prediction is scaled to match the amount of data.

in  $t\bar{t}$  and  $tW$  background events and the values from 3 to 4 in single top  $t$ -channel signal events. The distribution of the discriminator separately for events with positively charged leptons and for events with negatively charged leptons is shown in figure 8.19.

## 8.9 Parton and particle level objects

As explained in chapter 5, events are generated in steps. First, the hard interaction is simulated, then the parton shower followed by the hadronization and last the detector response. In order to compare predictions with measured data all steps of the simulation have to be performed. However, this makes it difficult to compare measurements between different experiments - with different detector responses - or between different event generators or theory models. In order to make these comparisons the effects of the detector response need to be undone. This is achieved using unfolding techniques (see section 3.3). For the unfolding it is necessary to properly define the objects and observables at the different levels of simulation. Physics objects after detector simulation (or actual measured object in data) are referred to as detector level objects and those detector level objects that are selected and reconstructed in the analysis are called reconstruction level objects. Those objects have already been defined in chapter 6 and the selection criteria have been laid out in section 8.3. Besides those there are also two further levels at which objects are defined; the parton level and the particle level. Physical objects contributing to the hard interaction before hadronization are referred to as parton level objects in this thesis. Objects after hadronization, but before simulating the detector response and particles from additional proton-proton collisions (pileup) are called particle level objects. Naturally, the particle level objects are more similar to the reconstructed objects. Additional selection criteria on the particle level objects are applied, which are similar to the selections at reconstruction level. The selection criteria are defined such that they are close to the fiducial phase space of the detector.



In the following a detailed description of objects at parton level, which are relevant to single-top  $t$ -channel production, is given.

**Top quark:** The top quark must originate from the hard interaction and has to be on-shell. Some generators simulate the top quark in multiple stages and store a copy of the top quark after each stage. In this case the top quark after accounting for QCD and QED radiations is taken.

**Prompt charged lepton:** Charged leptons can be muons or electrons. They must be part of the hard interaction and have to originate from a  $W$  boson. In about 11 % of single-top  $t$ -channel events, the  $W$  boson decays into a  $\tau$ . In these cases the muon or electron from the  $\tau$  decay is taken.

**Prompt neutrinos:** Neutrinos must also be associated to the hard interaction and have to originate from a  $W$  boson. If the  $W$  boson decays into a  $\tau$ , a pseudo-neutrino is constructed, whose four-momentum is computed by summing the four-momenta of the neutrino from the  $W$  boson decay and the subsequent decay.

**Spectator quark:** The spectator quark must be part of the hard interaction and must originate from the initial-state partons. As it should be of light flavor (u, d, s, c) it must not be a bottom quark. If the generator stores multiple copies of the quark, the copy after accounting for QCD and QED radiations is taken.

**Initial light quark:** At LO the single-top  $t$ -channel production in the 4FS has a gluon and a quark in the initial state (quark-gluon process). The initial light quark is then the quark in this process that lies inside the proton. The initial light quark must be part of the hard interaction and should not originate from any other particles. Further it must be a light-flavored quark. At NLO or at higher orders, the initial state can also consist of two quarks or two gluons. In case of two quarks in the initial state one is interested in the quark that scatters against the top quark, emitting the spectator quark. The other typically radiates a gluon. However, it is not always possible to find the correct quark of the two from the ancestor history as not always enough details from the event generator are stored in the event record. If the quarks have different flavor, the correct one can be chosen by matching the flavor to the charge of the top quark. However, in about 2 % of all events, the ambiguity cannot be resolved. In about 5 % of all events the initial state contains two gluons. In those events no initial light quark is reconstructed.

The following gives a detailed description of the definition of objects at particle level, which are relevant to single-top  $t$ -channel production. In general, objects at particle level use only final state particles with a mean lifetime of at least 30 ps. Further, all objects must lie within the detector acceptance:

**Dressed leptons:** Charged leptons may radiate photons. These photons are clustered with the lepton. Dressed leptons are clustered from photons and leptons within

a radius of  $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \leq 0.1$  around the charged lepton. No isolation conditions are imposed. The lepton must not originate from a hadron or quark.

**Missing transverse momentum / Neutrino:** The missing transverse momentum is computed by summing the four-momentum of all neutrinos from W boson, Z boson and  $\tau$  decays. The neutrino is reconstructed from the transverse momentum by utilizing the W boson mass constraint.

**Jets:** After hadronization quarks and gluons form jets, which contain different hadrons, leptons and photons. Jets are clustered using the anti- $k_T$  algorithm with  $\Delta R = 0.4$ . They are clustered from all stable particles excluding all neutrinos, photons, electrons, and muons used to define the dressed leptons.

**b jets:** All particle-level jets in which a b hadron is clustered are treated as b jets. Only b hadrons with an initial transverse momentum  $p_T \geq 5 \text{ GeV}$  are considered.

**Spectator jet:** At particle level the spectator jet is taken as the not-b-tagged jet with the largest pseudorapidity in an event.

**Pseudo Top quark:** The pseudo top quark is reconstructed by summing the four-momenta of the dressed lepton, neutrino and b jet. If there are multiple b jets in the event the one that results in a top quark mass closer to  $172.5 \text{ GeV}$  is used.

After defining the particle-level objects, a fiducial phase space is defined using selection criteria close to the selection criteria at reconstruction level. The selection criteria are based on dressed leptons and particle-level jets. A good muon must be a dressed muon with  $p_T \geq 25 \text{ GeV}$  and  $|\eta| < 2.4$  and a good electron a dressed electron with  $p_T \geq 30 \text{ GeV}$  and  $|\eta| \leq 2.1$ . Veto muons must be dressed muons with  $p_T \geq 15 \text{ GeV}$  and  $|\eta| < 2.5$ , and veto electrons are dressed electrons with  $p_T \geq 15 \text{ GeV}$  and  $|\eta| \leq 2.5$ . Jets must have  $p_T \geq 40 \text{ GeV}$  and  $|\eta| < 4.7$ . Further, there must be no dressed lepton in a cone of  $\Delta R = 0.4$  around the jet. b jets must fulfill a tighter pseudorapidity criteria of  $|\eta| < 2.4$ , in order to restrict it to the acceptance of the tracker. Events in the fiducial phase space must have exactly one good muon or electron and no additional veto muons or electrons. Additionally, there must be at least two jets in the event of which at least one must be b-tagged.

## 8.10 Systematic uncertainties

The analysis is subject to several sources of systematic uncertainties that are discussed in this section.

**Table 8.6: Luminosity systematic uncertainties correlations.** Uncorrelated and correlated contributions in the different years to the uncertainty in the luminosity measurement. The numbers are given in percent and are provided by the Luminosity Physics Object Group of the CMS Collaboration [188].

Source	2016 / %	2017 / %	2018 %
Uncorrelated 2016	1.8	0	0
Uncorrelated 2017	0	2	0
Uncorrelated 2018	0	0	1.5
X-Y factorization	0.9	0.8	2
Length scale 17-18	0	0.3	0.2
Beam current calibration 17-18	0	0.3	0.2
Dynamic inefficiency	0.3	0	0

### 8.10.1 Experimental uncertainties

**Luminosity:** The total uncertainty in the measured luminosity per year is 2.5 % in 2016, 2.3 % in 2017 and 2.5 % in 2018 [185–187]. These uncertainties are compromised of several sources of which some are correlated between different years and some are uncorrelated. A detailed breakdown can be found in table 8.6.

**Pileup:** As described in section 8.5.1 dedicated event weights are used to correct the difference in the pileup profiles in data and simulation. The inelastic proton-proton cross section that is used for the pileup reweighting is  $69.2 \pm 3.2$  mb. The relative uncertainty of 4.6 % is determined by the Luminosity Physics Object Group of the CMS collaboration [176]. This uncertainty is taken into account by computing dedicated sets of weights with the one standard deviation up and down varied cross section.

**Lepton and b tagging efficiencies:** The differences in the efficiencies between data and simulation in identifying leptons and b jets are taken into account using dedicated scale factors (see sections 8.5.4 and 8.5.3). Applying these scale factors introduces systematic uncertainties. To take these uncertainties into account the events are reweighted using shifted scale factors. The scale factors are shifted up and down by one standard deviation according to their measured systematic and statistical uncertainties. The uncertainties are considered fully correlated between all signal and background processes. The uncertainties in the lepton efficiencies are considered independently for electrons and muons but each fully correlated for all years. The uncertainties in the b tagging efficiencies are considered as partially correlated between years.

**Level 1 Trigger ECAL prefiring:** Dedicated event weights are used to correct for the prefiring issue discussed in section 8.5.5. The uncertainties in these weights

are computed for each  $(p_T, \eta)$  bin as either a flat 20 % in the prefiring probability or the statistical uncertainty in the corresponding bin, depending on which is larger. Events are reweighted using the up and down varied event weights.

**Corrections to the jet energy and missing transverse momentum:** As discussed in section 6.6 the jet energies are corrected in order to obtain a better agreement between data and simulation. Each of these corrections introduces systematic uncertainties that have to be taken into account. The systematic uncertainties in the jet energy corrections can be decomposed into different sources, such as high- $p_T$  extrapolation uncertainties, absolute scale uncertainties, relative jet energy resolution, jet flavor dependent uncertainties or uncertainties due to pileup. The quadratic sum of these individual sources is taken as uncertainty. A detailed explanation can for example be found in reference [160].

The jet energy in simulation is smeared using resolution scale factors. Uncertainties due to the jet energy resolutions contribute already to the total uncertainties of the jet energy corrections. However, the jet energy resolution uncertainties used as part of the uncertainties of the jet energy corrections are stemming from residual differences in dijet events that are mainly caused by the differences in resolution between the barrel and the forward region. This effect is zero in the barrel region in which both jets have the same resolution. Thus, additional uncorrelated uncertainties are used for the jet energy resolution scale factors. They are implemented by smearing the jet energy using the up and down variations of the resolution scale factors.

As described in section 6.7 the jet energy corrections also affect the missing transverse momentum, as a change in the jets energies also translates into changes in the reconstructed missing transverse momentum of the event. Therefore, the uncertainties imposed due to the jet energy corrections are also propagated to the missing transverse momentum.

Changes in the jet energy affects both the selection efficiency as well as the reconstruction of physics objects. Therefore, all uncertainties described above are taken into account by rerunning all analysis steps on dedicated samples with altered jet energies and momentum. Systematically varied templates are then derived from these samples.

**Limited size of simulated samples:** This analysis relies on MC simulation for the prediction of the signal and background templates. Due to computational limitations only a limited amount of simulated events can be generated. The limited size of the MC samples causes statistical fluctuations in each bin. These fluctuations are taken into account using the Barlow-Beeston-lite method [77,78]. This method introduces one additional nuisance parameter per bin.

**Background normalization:** Normalization uncertainties in the cross sections of the background processes are applied. The uncertainty in the  $t\bar{t}$  cross section is taken to be 6 % [30]. For the  $tW$  cross section the experimental uncertainty of

11 % [189] instead of the theory uncertainty is used to account for possible overlap with the  $t\bar{t}$  process at NLO accuracy. An uncertainty of 10 % in the W/Z-jets cross section is applied. This value is chosen larger than the theory uncertainty in the NNLO cross section of about 4 %, to account for uncertainties stemming from the selection of heavy-flavored jets. Finally, an uncertainty of 50 % is applied in the normalization of the data-driven QCD multijet background.

### 8.10.2 Theory uncertainties

**Top quark  $p_T$ :** In differential measurements it is observed that the distribution of measured top quark  $p_T$  in  $t\bar{t}$  events tends towards larger values, than predicted [190, 191]. This effect is taken into account by reweighting the nominal distributions with weights that correspond to the difference in the observed and predicted  $p_T$  spectrum. This uncertainty is only considered for  $t\bar{t}$  events and not for other top quark processes, i.e. single-top  $t$  channel and  $tW$  processes. The uncertainty is taken as fully correlated between all years.

**PDF and  $\alpha_S$ :** Events are simulated using the NNPDF3.1 [117] PDF set. This PDF set is provided with 100 Hessian eigenvector variations of the nominal PDF value. For each eigenvector variation dedicated event weights are stored in each event. Those weights can be used to reweight the events in order to obtain systematically shifted templates. The nominal value for the strong coupling constant used in this PDF set is  $\alpha_S = 0.118$ . In addition to the eigenvalue variations there are two weights in each event that correspond to up and down variations of the strong coupling constants, namely  $\alpha_S = 0.120$  and  $\alpha_S = 0.116$ , respectively. Using these weights systematic varied templates are constructed following the PDF4LHC recommendations for Run II [192]. This uncertainty is evaluated for each process individually and is considered correlated between all years of data-taking for same processes.

**Factorization and renormalization scales at matrix-element level:** Simulated events contain event weights that correspond to variations of a factor two up or down of the renormalization  $\mu_R$  and factorization scales  $\mu_F$ . The envelope of all variations in which the renormalization and factorization scale are varied separately up or down is taken as uncertainty. The variations in which both scales are simultaneously varied up or down are not considered in the envelope. This uncertainty is considered correlated for the same process across all years. Further it is considered uncorrelated between QCD-induced ( $t\bar{t}$ bar) and electroweak-induced (single-top) processes.

**Initial-state and final-state radiation:** The emissions in the parton shower are computed at a certain energy scale  $Q^2$  corresponding to a certain value of the strong coupling constant  $\alpha_S$ . To take into account uncertainties due to  $\alpha_S$  in the parton shower, the scale for QCD emissions in initial-state (ISR) and final-state radiations (FSR) is varied. Technically, these variations are implemented

via event weights that correspond to a factor two up and down shifts of the scale  $Q$ , separately for ISR and FSR. For the ISR variation electroweak-induced single top quark processes are not correlated with QCD-induced  $t\bar{t}$  processes. This is done because the parton shower approximates higher order accuracies that are process dependent. The FSR uncertainties are considered correlated between  $t\bar{t}$  and single top quark processes.

**Matching of parton shower and matrix element:** In POWHEG the probability for real emissions is scaled by a damping function  $f = h_{\text{damp}}^2 / (p_T^2 + h_{\text{damp}}^2)$ . This damping function controls the matching between the parton shower and the matrix element. The value of the damping variable is [132].

$$h_{\text{damp}} = 1.379_{-0.5052}^{+0.926} \cdot m_{\text{top}}. \quad (8.21)$$

The uncertainty is taken into account for the  $t\bar{t}$  process by using dedicated simulated samples that have been generated with a one standard deviation up or down shifted damping variable. In all samples a top quark mass of 172.5 GeV is assumed.

**Underlying event tune:** In simulated events, the CP5 tune is used to model the underlying event. Systematic uncertainties due to the choice of the tune settings for single top  $t$ -channel and  $t\bar{t}$  process are taken into account by using dedicated simulated samples that have been generated with different tune settings. A description of the underlying event tune settings can be found in reference [132,193]. Uncertainties are considered correlated between QCD-induced and electroweak-induced processes. The simulated W/Z-jets events in 2016 are generated with the underlying event tune CUETP8M1. For all other samples the CP5 tune is used. This may result in inconsistencies due to different modelling of the underlying event. The change in the selection efficiency in the signal sample when using the CUETP8M1 instead of the CP5 tune is found to be smaller than 5%. Thus, a conservative additional 5% rate uncertainty is applied for W/Z-jets events in 2016, which should cover effects due to the differences in the underlying event tune.

## 8.11 Inclusive cross section measurement

The cross section of any process can be calculated as:

$$\sigma_{\text{total}} = \frac{N_{\text{signal}}}{\epsilon L_{\text{int}}}. \quad (8.22)$$

Here  $N_{\text{signal}}$  are the observed number of signal events,  $L_{\text{int}}$  is the measured integrated luminosity and  $\epsilon$  the event selection efficiency. Given the selected number of events  $N_{\text{selected}}$  and the total number of events  $N_{\text{total}}$  in a sample of simulated events, the

selection efficiency is defined as  $\epsilon = N_{\text{selected}}/N_{\text{total}}$ . Both the integrated luminosity and the selection efficiency is given by the experiment and analysis setup. The missing ingredient in a cross section measurement is therefore the number of signal events that needs to be determined from data.

### 8.11.1 Signal extraction

In order to determine the number of signal events a ML fit using the HIGGS COMBINE tool [75] is performed on the combined output distribution of the multiclassification BDT<sub>S</sub> (see section 8.8). The discriminator is defined in such a way that it defines four intervals of which each is enriched in one of the processes:  $t$ -channel single top quark production,  $t\bar{t}$  or  $tW$  production, W/Z-jets production or QCD multijet production. Thus, not only a signal-enriched region, but also three background enriched regions are used in the ML fit. These background-enriched regions help to constraint the background components in the ML fit and thus reduce the overall uncertainty in the background prediction. Further, the normalization of the QCD multijet background is still undetermined, as only the shape has been taken from the data in the sideband. Defining a QCD multijet enriched region allows to fit the QCD normalization simultaneously with the number of signal events.

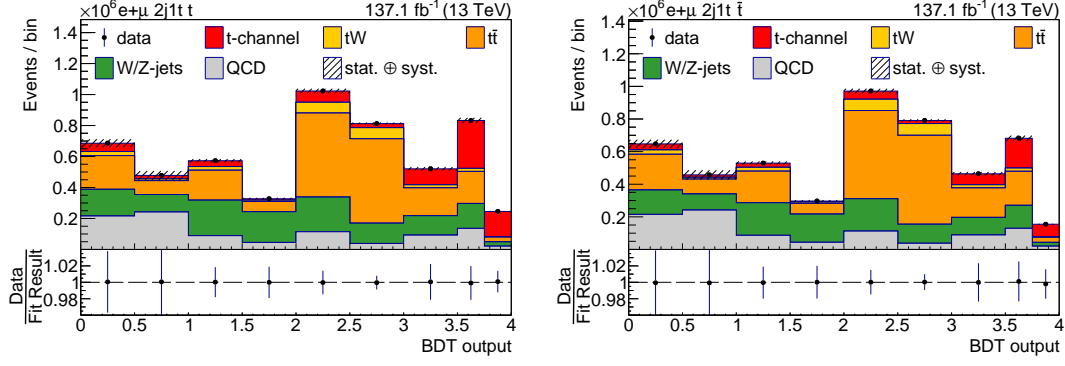
For each year of data-taking and each process templates of the discriminating variable are constructed separately for positively and negatively charged leptons. Top quarks decay always into positively charged leptons and top antiquarks always into negatively charged leptons. Therefore, this setup allows to fit the number of top quark and top antiquark events separately and also to determine the ratio of the top quark to top antiquark production cross section. In the fit one signal strength parameter for the rate of top quark  $\mu_t$  and one for the rate of top antiquark  $\mu_{\bar{t}}$  events is used. The signal strength parameters scale the amount of observed single top quark and single top antiquark events, such that the observed number of events is the product of the predicted number of events and the signal strength parameter. Therefore, using the signal strength parameters the top quark and top antiquark cross sections are computed according to:

$$\sigma_{t,\text{obs.}} = \mu_t \cdot \sigma_{t,\text{theo.}}$$

$$\sigma_{\bar{t},\text{obs.}} = \mu_{\bar{t}} \cdot \sigma_{\bar{t},\text{theo.}}$$

Here  $\sigma_{\text{theo.}}$  are the theoretically predicted cross sections (see chapter 2.1). In addition two signal strength parameter for the amount of QCD multijet events in barrel and endcap are introduced to fit the QCD multijet normalization simultaneously.

The systematic uncertainties considered are described in section 8.10. For each uncertainty and process an additional varied template is constructed. All systematic uncertainties, except the theory uncertainties in the modeling of the signal process, are implemented in the ML fit as nuisance parameters. The only exception is the uncertainty in the luminosity measurement that is added as a flat uncertainty in the measured cross section. Theory uncertainties affecting the modeling of the signal processes are



**Figure 8.20: Postfit  $\text{BDT}_S$  distributions.** Shown are the distributions of  $\text{BDT}_S$  for events with positively charged leptons (left) and negatively charged leptons (right). The distribution corresponds to events in the 2j1t event category. The prediction is scaled to the inclusive fit result.

not implemented as nuisance parameters in the ML fit. These signal modeling uncertainties are the uncertainties due to the choice of PDF and  $\alpha_S$ , due to the factorization and renormalization scales, initial-state and final-state radiation and the modeling of the underlying event tune. These are not profiled in the fit because the measurement is carried out in a limited phase space region and the results are extrapolated to the full phase space based on the predictions of the signal simulation. The signal modeling uncertainties are therefore relevant in the selected phase space region and in the full phase space. Thus, these uncertainties should not get constrained from the fit in the reduced phase space. Therefore, instead of implementing the signal modeling uncertainties as nuisance parameters in the ML fit, a separate fit is performed for each signal modeling uncertainty. In these fits the nominal signal template is replaced by the systematically varied one. The uncertainty due to a source on a signal strength parameter is taken as the larger of the absolute shifts in the fitted signal strength parameter, caused by either the "up" or "down" variation of the source. Figure 8.20 shows the  $\text{BDT}_S$  distribution for events with positively and negatively charged leptons where the predictions are scaled to the fit result.

### 8.11.2 Results

The measured cross sections for  $t$ -channel single top quark production  $\sigma_t$  and top antiquark production  $\sigma_{\bar{t}}$  are:

$$\begin{aligned}
 \sigma_t &= 130 \pm 1 \text{ (stat)} \pm 3 \text{ (prof)} \pm 20 \text{ (sig-mod.)} \pm 3 \text{ (lumi)} \text{ pb} \\
 &= 130 \pm 1 \text{ (stat)} \pm 20 \text{ (syst)} \text{ pb} \\
 &= 130 \pm 20 \text{ pb}
 \end{aligned}$$



$$\begin{aligned}
 \sigma_{\bar{t}} &= 80 \pm 1 \text{ (stat)} \pm 2 \text{ (prof)} \pm 15 \text{ (sig-mod.)} \pm 2 \text{ (lumi)} \text{ pb} \\
 &= 80 \pm 1 \text{ (stat)} \pm 15 \text{ (syst)} \text{ pb} \\
 &= 80 \pm 15 \text{ pb}
 \end{aligned}$$

The systematic uncertainty ("syst") is calculated as the quadratic sum of the uncertainty from the systematic variations that are profiled in the maximum likelihood fit ("prof"), the theory uncertainty in the signal modeling ("sig-mod."), and the uncertainty due to the luminosity ("lumi"). The measured values agree with the predictions from the SM. From  $\sigma_t$  and  $\sigma_{\bar{t}}$  the cross section ratio  $R_{t\text{-ch.}} = \frac{\sigma_t}{\sigma_{\bar{t}}}$  is calculated taking into account the correlation between the uncertainties.

$$\begin{aligned}
 R_{t\text{-ch.}} &= 1.63 \pm 0.01 \text{ (stat)} \pm 0.03 \text{ (prof)} \pm 0.05 \text{ (sig-mod.)} \\
 &= 1.63 \pm 0.01 \text{ (stat)} \pm 0.06 \text{ (syst)} \\
 &= 1.63 \pm 0.06
 \end{aligned}$$

In figure 8.21 a comparison between the measured cross section ratio and the predictions from different PDF sets is shown. The measured cross section ratio is lower than the predictions but still agrees with most of them within the uncertainties of the measurement and the prediction. The largest difference is found between the measured value and the ABM [194, 195] PDF set.

The total single-top  $t$ -channel cross section  $\sigma_{t+\bar{t}}$  is given by the sum of  $\sigma_t$  and  $\sigma_{\bar{t}}$ :

$$\begin{aligned}
 \sigma_{t+\bar{t}} &= 211 \pm 1 \text{ (stat)} \pm 6 \text{ (prof)} \pm 35 \text{ (sig-mod.)} \pm 5 \text{ (lumi)} \text{ pb} \\
 &= 211 \pm 1 \text{ (stat)} \pm 35 \text{ (syst)} \text{ pb} \\
 &= 211 \pm 35 \text{ pb}
 \end{aligned}$$

Here, the profiled uncertainty is calculated using the full covariance matrix of the fit result, taking into account the correlations between the profiled uncertainties of  $\sigma_t$  and  $\sigma_{\bar{t}}$ . The signal modeling uncertainty is calculated by linearly adding up the signal modelling uncertainties of  $\sigma_t$  and  $\sigma_{\bar{t}}$ , assuming them to be fully correlated.

From the total single-top  $t$ -channel cross section the CKM-Matrix element  $V_{tb}$  is calculated. As explained in chapter 2.1, it can be calculated as:

$$|f_{LV} V_{tb}| = \sqrt{\frac{\sigma_{\text{measured}}}{\sigma_{\text{predicted}}}}, \quad (8.23)$$

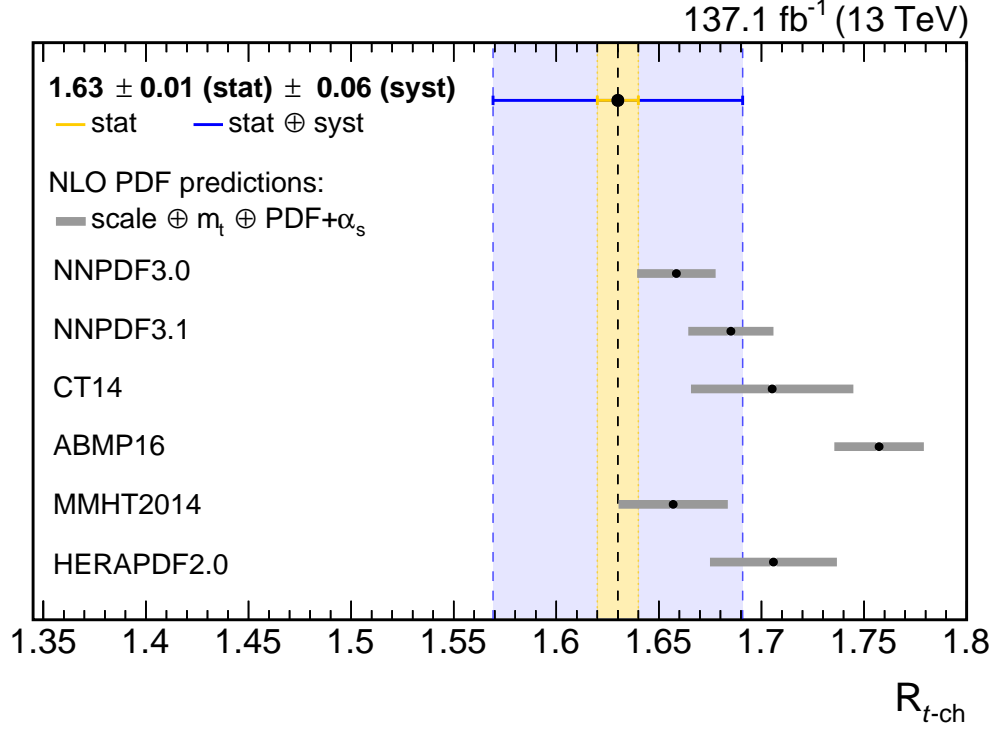
assuming that the top quark decays only in bottom quarks. Here,  $f_{LV}$  is a anomalous form factor that modifies the  $Wtb$  interaction vertex [59].  $f_{LV}$  is 1 in the SM and unequal one assuming new physics that modifies the strength of the left-handed coupling

between the top quark and the W boson. Using the theoretical single-top  $t$ -channel cross section (see chapter 2.1) and the measured cross section it can be computed to be:

$$|f_L V_{tb}| = 0.99 \pm 0.08 (\text{exp}) \pm 0.02 (\text{theo}). \quad (8.24)$$

Here, the experimental uncertainties correspond to the total uncertainties from the measurement, while the theoretical uncertainties correspond to the uncertainty in the theory predictions. Under the assumption that the CKM matrix is unitary, a 95 % confidence-level lower limit can be computed for the CKM matrix element  $V_{tb}$ . The lower limit is determined to be:  $V_{tb} > 0.82$  (95 % C.L.).

The relative contributions of all systematic variations considered in the analysis to the total uncertainty of  $\sigma_t$ ,  $\sigma_{\bar{t}}$  and their ratio is shown in table 8.7. The relative contributions of systematic variations that are profiled in the fit are determined as follows. For each systematic variation, the nuisance parameters corresponding to the variation are kept to their best fit value and all other are allowed to vary freely. The relative difference in the systematic uncertainty as compared to the nominal fit is then quoted as the relative contribution of the systematic variation. For simplicity multiple similar systematic variations are grouped together. The dominant systematic uncertainties are the theory uncertainties in the signal modeling. The three signal modeling uncertainties with the largest contribution to the total uncertainty are the uncertainty due to the underlying event tune, the systematic uncertainties due to the renormalization and factorization scale and the systematic uncertainties due to the signal PDF. From the systematic uncertainties that are profiled in the ML fit, the largest are the systematic uncertainties due to the limited size of simulated samples, the jet energy correction uncertainties and the uncertainties due to the modeling of the background processes.



**Figure 8.21: Comparison of the measured cross section ratio with predictions from different PDF sets.** The measured cross section ratio  $R_{t\text{-ch}}$  is compared with the NLO predictions from different PDF sets. The PDF sets are provided by LHAPDF 6.2.1 [196]: NNPDF3.0 [122], NNPDF3.1 [117], CT14 [197], ABMP16 [194, 195], MMHT2014 [198], HERAPDF2.0 [199]. The uncertainties in the predictions of the different PDF sets are calculated using the combined PDF+ $\alpha_s$  uncertainty, the uncertainty due to the top quark pole mass and the uncertainty due to the renormalization and factorization scale.

**Table 8.7: Uncertainty contributions** Relative contribution of the different systematic uncertainty sources to the total uncertainty. The values are sorted based on their contribution to the uncertainty to the cross section ratio  $R_{t\text{-ch}}$ .

Source	$R_{t\text{-ch}} / \%$	$\sigma(t) / \%$	$\sigma(\bar{t}) / \%$
Signal tune	20.2	20.0	20.4
$\mu_R/\mu_F$ scale signal	16.8	28.2	26.4
PDF signal	15.7	15.1	15.4
FSR signal	15.6	16.8	16.8
ISR signal	13.0	12.8	13.0
Limited size of simulated samples	4.9	1.3	1.6
Jet energy scale corrections	2.3	1.0	1.2
Background modeling	2.2	0.7	0.8
b tagging	2.0	0.6	0.8
Background normalization	2.0	0.8	0.9
Jet energy resolution corrections	1.5	0.6	0.7
Unclustered energy corrections	1.3	0.3	0.5
Pileup corrections	1.2	0.8	0.9
Lepton scale factors	1.1	0.6	0.2
Prefire corrections	0.3	0.4	0.4

## 8.12 Differential cross section measurement

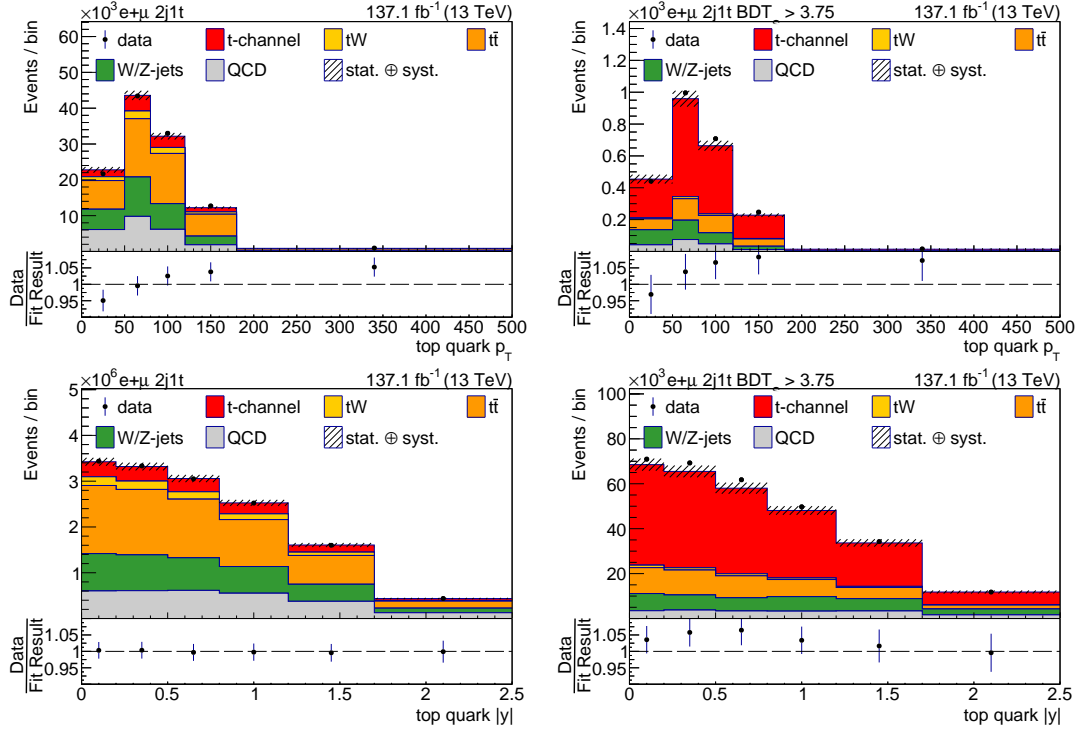
The differential cross section  $d\sigma_k^{\text{gen}}$  in bin  $k$  as a function of a variable  $x$  is defined as:

$$\frac{d\sigma_k^{\text{gen}}}{dx_k} = \frac{1}{\Delta_k^x L_{\text{int}}} \sum_j R_{jk}^{-1} N_j^{\text{signal}}, \quad (8.25)$$

where  $L_{\text{int}}$  is the integrated luminosity,  $\Delta_k^x$  is the width of bin  $k$ ,  $N_j^{\text{signal}}$  is the measured number of signal events and  $R_{jk}$  is the response matrix element that describes migrations between bins  $j$  and  $k$  and also includes the efficiency. For the measurement of a differential cross section first the variables as a function of which it is measured needs to be defined at both reconstruction and generator level, next a binning needs to be chosen, then the response matrix has to be determined from simulation and lastly the cross section is determined using an unfolding procedure (see section 3.3).

### 8.12.1 Observables

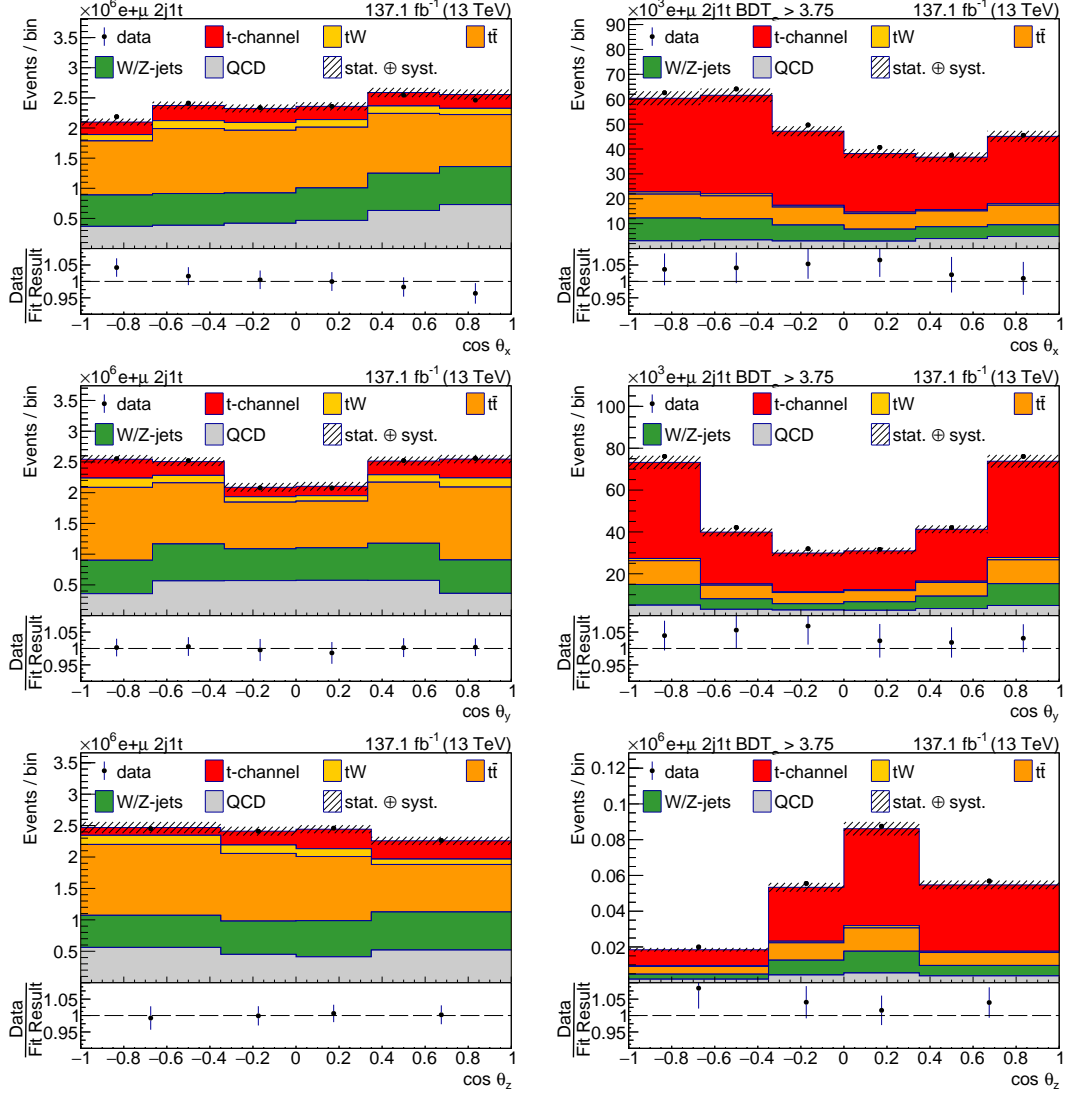
The differential cross sections at parton and particle level on the following five variables are measured, namely, the transverse momentum and rapidity of the top quark and the three polarization angles  $\cos(\theta_x)$ ,  $\cos(\theta_y)$ ,  $\cos(\theta_z)$ . At reconstruction level the top quark is defined by summing the four-momenta of the b-tagged jet, the charged lepton



**Figure 8.22: Prefit distributions for the top quark transverse momentum and rapidity.** The distribution corresponds to events in the 2j1t event category. On the left all events are shown, the right shows only events in a signal-enriched region for which the BDT output is larger than 3.75.

and the reconstructed neutrino (see section 8.7). At parton level the top quark is required to be on-shell and stem from the hard interaction, while at particle level a pseudo-top quark is reconstructed from dressed leptons and jets clustered from stable particles (see section 8.9). The rapidity is computed as  $y = \frac{1}{2} \ln \frac{E+p_z}{E-p_z}$ . The polarization angles are defined in section 2.3. Essentially, a coordinate system in the top quark rest-frame is defined. At reconstruction and particle level the  $z$  axis is chosen along the direction of the reconstructed spectator jet and at parton level along the direction of the generated spectator quark. In both cases the  $y$  axis is defined orthogonal to the  $z$  axis and the direction of the beam axis that matches the direction of flight of the spectator jet/quark. The  $x$  axis is chosen orthogonal to both the  $y$  and  $z$  axis such that a right-handed coordinate system is formed. The polarization angles  $\cos(\theta_x)$ ,  $\cos(\theta_y)$  and  $\cos(\theta_z)$  are then computed between the axis  $x$ ,  $y$  and  $z$ , respectively and the direction of the charged lepton in the top quark rest-frame. Figure 8.22 and 8.23 show the prefit distribution of these variables in the 2j1t event category. The distributions are shown inclusively to illustrate the overall modeling and in a signal enriched region with  $\text{BDT}_S > 3.75$ . Overall good agreement between data and prediction is observed for all variables.

For the unfolding an appropriate binning must be chosen. For  $\cos(\theta_x)$  and  $\cos(\theta_y)$  simply an equidistant binning was chosen as the response matrices for both observables



**Figure 8.23: Prefit distributions for the polarization angles.** Shown are the distributions for  $\cos(\theta_x)$  (top),  $\cos(\theta_y)$  (middle) and  $\cos(\theta_z)$  (bottom). The distribution corresponds to events in the  $2j1t$  event category. On the left all events are shown, the right shows only events in a signal-enriched region for which the BDT output is larger than 3.75.

**Table 8.8: Binning** The number of bins and bin edges are listed in which the differential measurements are performed.

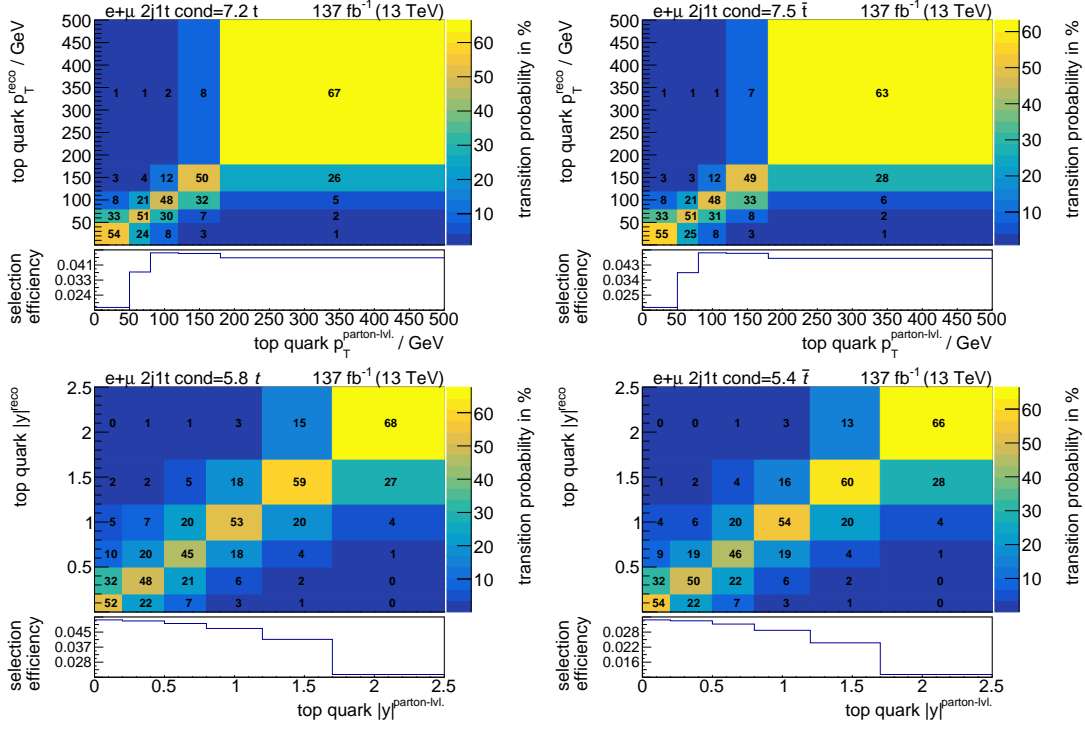
variable	number of bins	bin edges
top quark $p_T$ / GeV	5	0, 50, 80, 120, 180, 500
top quark $ y $	6	0, 0.2, 0.5, 0.8, 1.2, 1.7, 2.5
$\cos(\theta_x)$	6	-1, -0.67, -0.33, 0, 0.33, 0.67, 1
$\cos(\theta_y)$	6	-1, -0.67, -0.33, 0, 0.33, 0.67, 1
$\cos(\theta_z)$	4	-1 -0.35, 0, 0.35, 1

are very diagonal. For the other variables the binning was chosen such that the condition number (see section 3.3) of the response matrix is minimal. Only sensible binnings have been considered, e.g.,  $\cos(\theta_z)$  must be symmetric with respect to zero, the distance between bins must be at least 0.05 for  $\cos(\theta_z)$  and the top quark rapidity, and at least 10 GeV for the top quark  $p_T$ . The chosen bin edges are listed in table 8.8. The same binning and the same number of bins is used at reconstruction and generator level. For the top quark  $p_T$  and  $|y|$  the chosen bins results in an overflow bin. As there are migrations from and to the overflow bin, the overflow bin is considered during the unfolding. The response matrices for all variables at parton level are shown in figure 8.24 and 8.25 and the response matrices at particle level are shown in appendix A.5. All response matrices are well conditioned with condition numbers for all response matrices being smaller than ten.

### 8.12.2 Unfolding

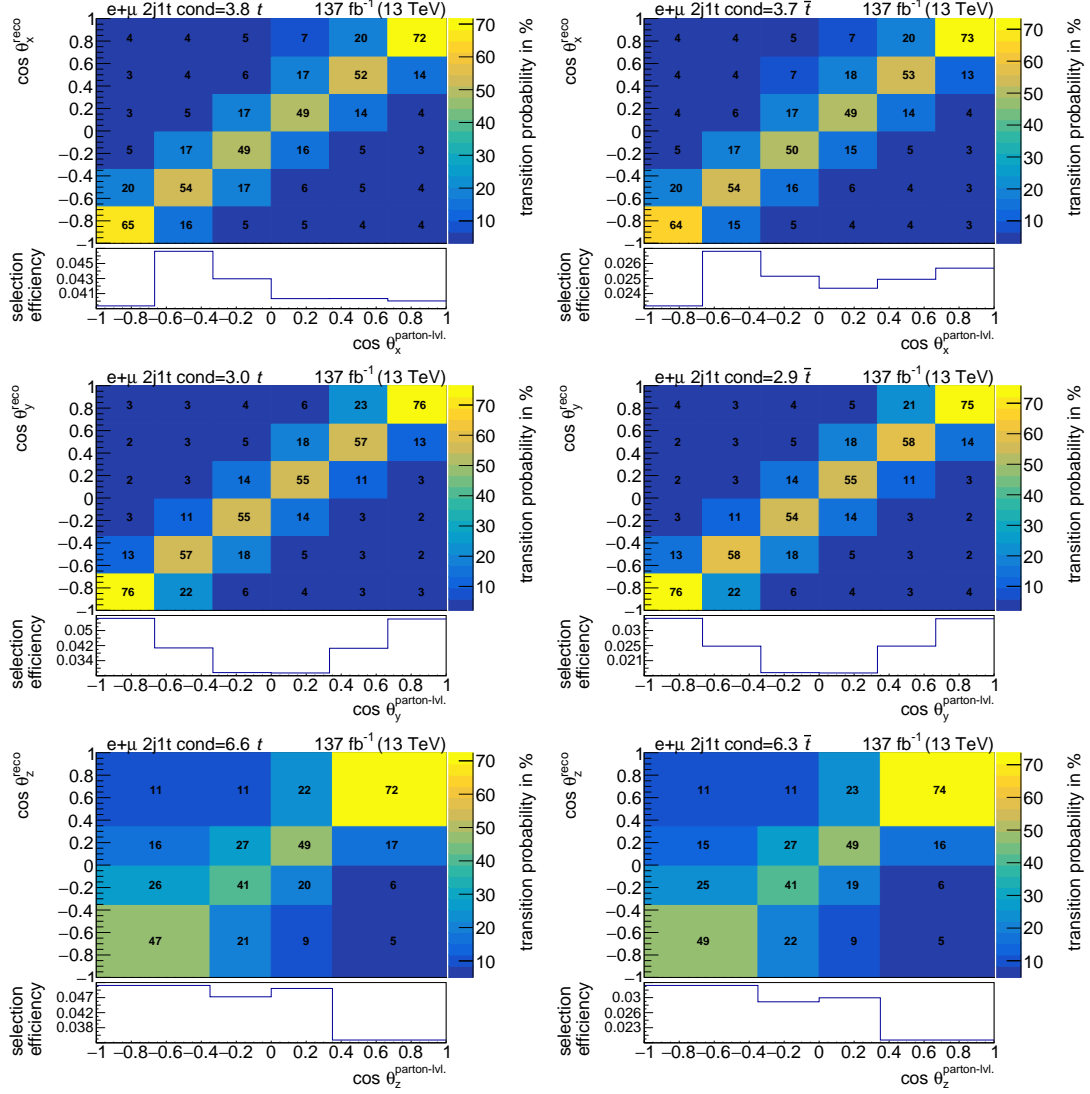
Technically, the unfolding is implemented via a two-dimensional ML fit using the HIGSCOMBINE Tool, as explained in section 3.3. The two-dimensional ML fit is performed to the output distribution of  $\text{BDT}_S$  and the variable as a function of which the differential cross section is measured. The variable  $\text{BDT}_S$  is used as an ancillary variable in the unfolding in order to obtain regions that are dominated by the signal process and regions that are dominated by each background process. Those regions can be used in the ML fit to constraint the background processes and to determine the number of background events. Therefore, no separate estimation of the background contributions and subsequent background subtraction is necessary prior to unfolding. The same binning for  $\text{BDT}_S$  is used as in the inclusive fit.

For the signal process, one two-dimensional fit template for each generator level bin is constructed. This allows to fit the amount of events in each generator level bin by introducing a corresponding signal strength parameter. All fit templates are constructed separately for events with positively and negatively charged leptons in order to be able to fit the yields for top quarks and top antiquarks separately. Therefore, for each generator level bin there are two signal strength modifiers included in the fit, one for the amount of  $t$ -channel top quark events and one for  $t$ -channel top antiquark events. The observed number of events in each generator level bin is computed from the fitted



**Figure 8.24: Response matrices for the polarization angles at parton level.** Shown are the response matrices and selection efficiencies for the top quark  $p_T$  (top) and  $|y|$  (bottom). The response matrices are shown at parton level for top quarks (left) and top antiquarks (right). They are obtained from simulated single-top  $t$ -channel events in the 2j1t event category. The histograms are normalized such that the sum over the parton level bins sums up to 100%.





**Figure 8.25: Response matrices for the polarization angles at parton level.** Shown are the response matrices and selection efficiencies for the polarization angles  $\cos(\theta_x)$  (top),  $\cos(\theta_y)$  (middle) and  $\cos(\theta_z)$  (bottom). The response matrices are shown at parton level for top quarks (left) and top antiquarks (right). They are obtained from simulated single-top  $t$ -channel events in the 2j1t event category. The histograms are normalized such that the sum over the parton level bins sums up to 100 %.

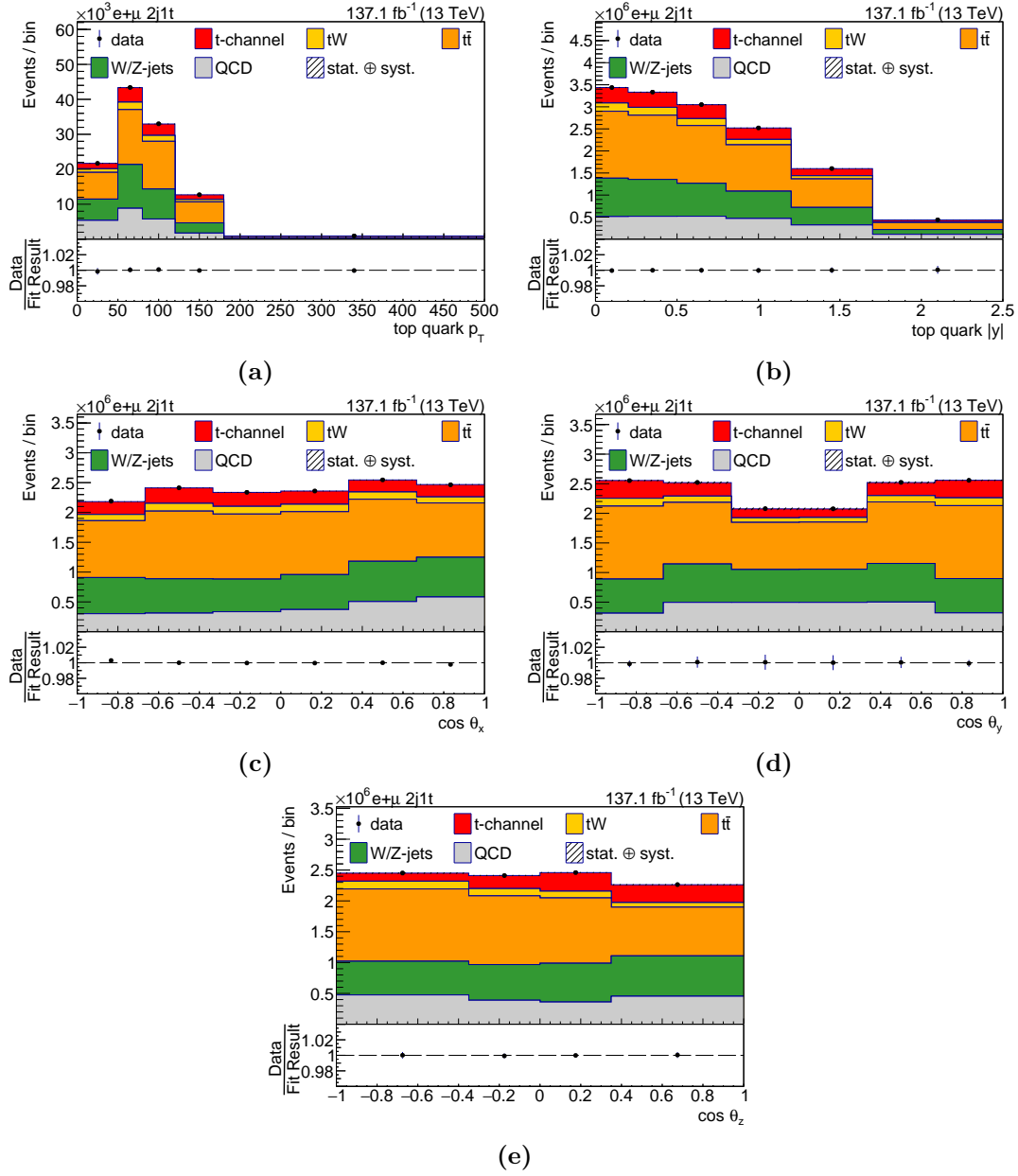
signal strength modifier for that bin. In addition, two parameters are introduced in the fit that scale the amount of QCD multijet events in the barrel and endcap regions. No regularization is used in the unfolding because the response matrices are found to be well conditioned.

All systematic uncertainties that are considered in the fit are described in section 8.10. The systematic uncertainties are treated in the same way as for the inclusive fit. This means all uncertainties except the theory uncertainties in the modeling of the signal process are treated as nuisance parameters in the fit. For each signal modeling uncertainty a separate fit is performed in which the nominal signal template is replaced with the systematically varied one. The postfit distributions for each variable are shown in figure 8.26. For better visibility the distributions are projected on one dimension by summing over bins of  $\text{BDT}_S$ .

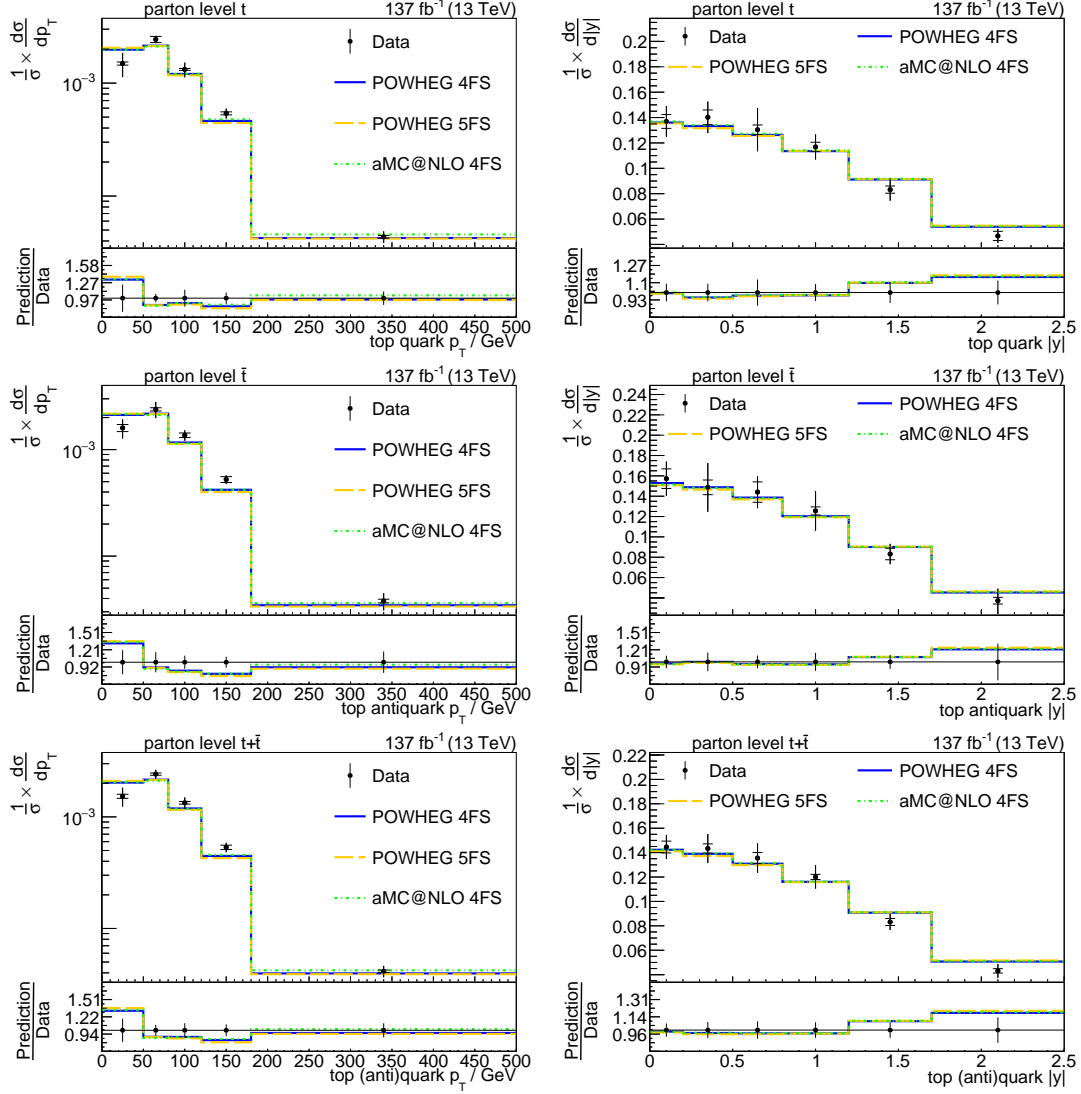
### 8.12.3 Measured differential cross sections

The measured normalized differential cross sections as a function of the top quark transverse momentum and rapidity at parton level are shown in figure 8.27 and for the polarization angles at parton level in figure 8.28 and figure 8.29. The measured normalized differential cross sections at particle level are shown in appendix A.5. The measured results are compared with simulations of the SM predictions from POWHEG in the 4FS and the 5FS, and MADGRAPH5\_AMC@NLO, both interfaced with PYTHIA. The results agree mostly with the predictions from the SM. The distribution of the top quark pseudorapidity follows a trend in the last two bins, where the data tends slightly towards more central values. However, the deviation in the last bin of the normalized distribution is only about 1.3 standard deviations for top antiquarks. For top antiquarks the first bin of the measured  $\cos(\theta_z)$  distribution shows an overfluctuation of about 1.3 standard deviations and the second bin an underfluctuation of about 1.7 standard deviations. The deviations go in opposite directions, whereas a change in the longitudinal top quark polarization would be expected to have a coherent effect in the same direction. Therefore, the deviations are most certainly due to fluctuations in the measurement.

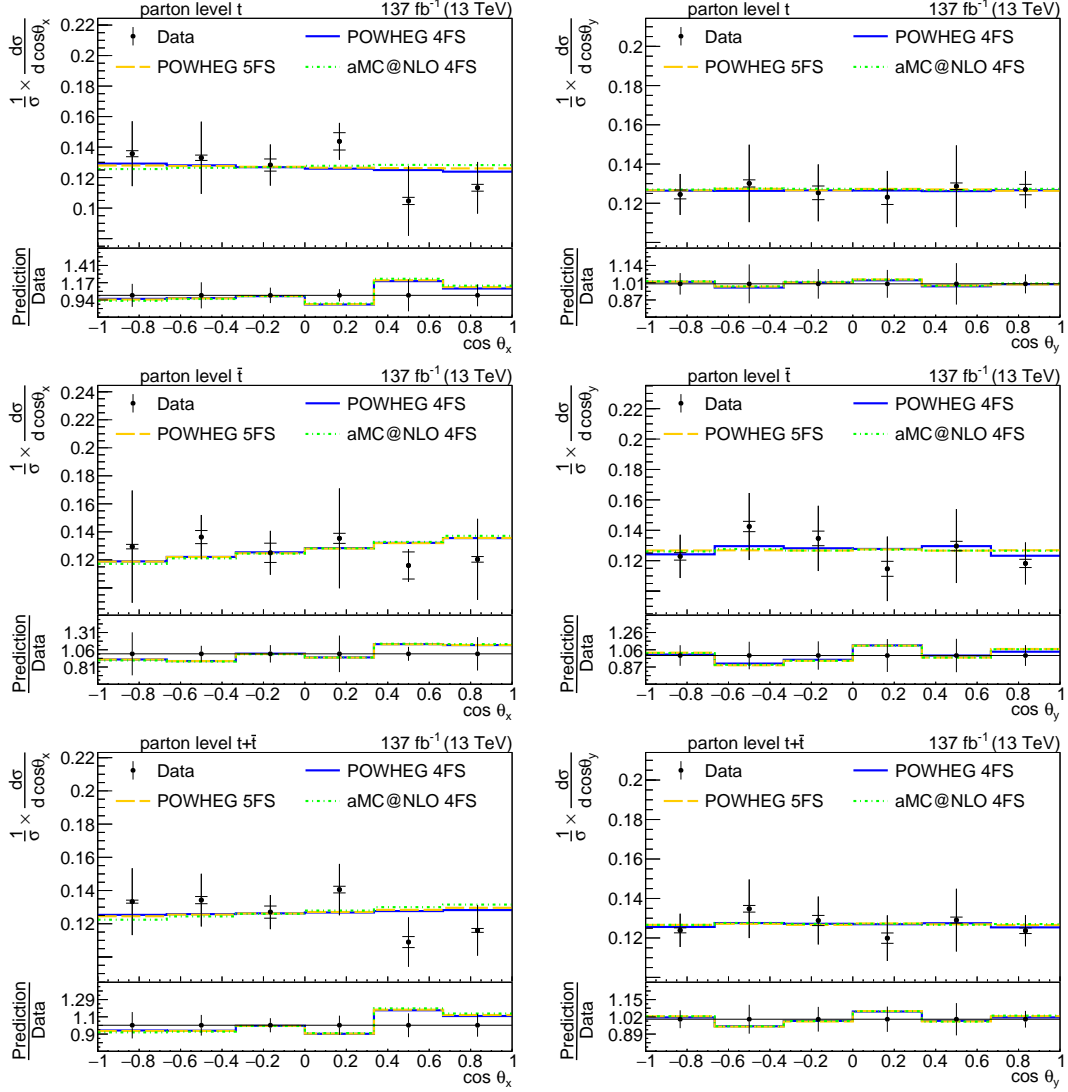
The largest contribution to the systematic uncertainty is due to the theory uncertainty in the signal modeling and the limited number of simulated events. The statistical uncertainties are about 1% in most bins. A detailed breakdown of the contributions from each systematic source to the uncertainty in each bin can be found in appendix A.6. For the theory uncertainties in the signal modeling the contributions correspond to the relative change in the fit result when repeating the fit with the nominal signal template replaced by the systematically varied one. For the other uncertainties, the contributions are derived by repeating the nominal fit while keeping the corresponding nuisance parameters fixed to their best fit value. The contribution is then computed as the relative difference in the systematic uncertainty.



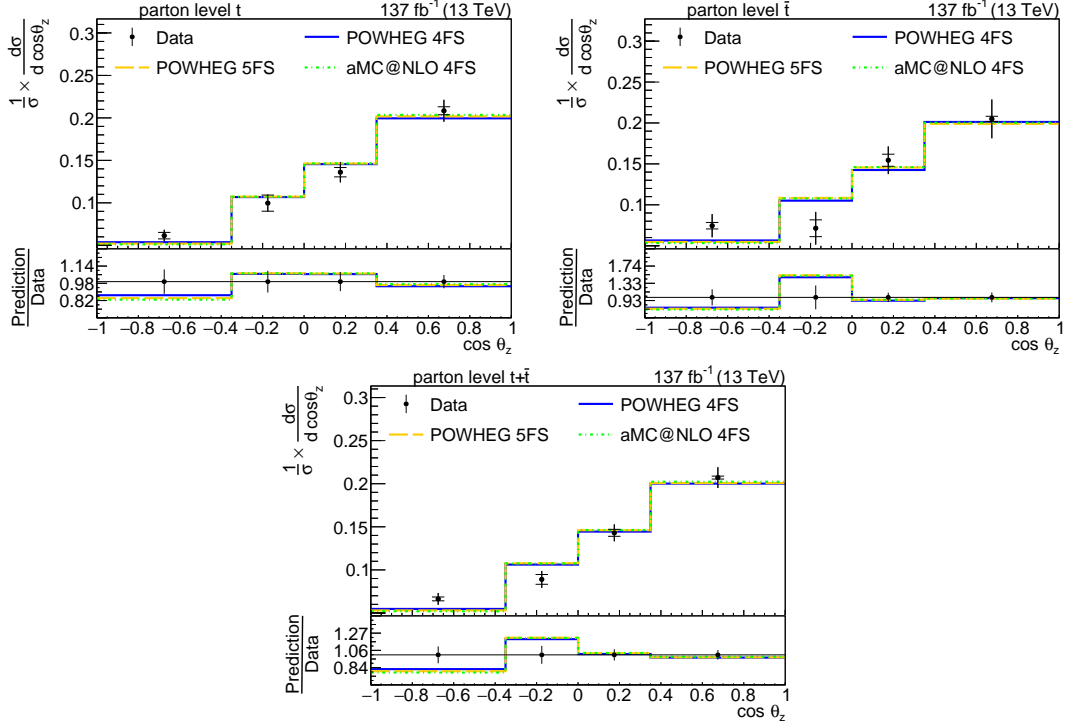
**Figure 8.26: Postfit distributions** Shown are the postfit distributions for the top quark transverse momentum (a), rapidity (b) and for the polarization angles  $\cos(\theta_x)$  (c),  $\cos(\theta_y)$  (d) and  $\cos(\theta_z)$  (e). The distribution corresponds to events in the 2j1t region. The prediction is scaled to the fit result.



**Figure 8.27: Unfolding results for top quark transverse momentum and rapidity at parton level.** Shown are the measured normalized differential cross sections as a function of the top quark  $p_T$  (left) and as a function of the top quark rapidity (right). On the top the differential cross sections for top quarks ( $t$ ) in the middle for top antiquarks ( $\bar{t}$ ) and on the bottom the sum of the two are shown. The measured distributions are compared to POWHEG in the four-flavour scheme (4FS, blue) and the five-flavour scheme (5FS) (orange), and to AMC@NLO in the 4FS scheme (green). In the ratio panel, the black line marks the 1. The horizontal lines on the error bars indicate the uncertainties that are profiled in the fit.



**Figure 8.28: Unfolding results for the  $x$  and  $y$  polarization angles at parton level.** Shown are the measured normalized differential cross sections as a function of the polarization angles  $\cos(\theta_x)$  (left) and  $\cos(\theta_y)$  (right). On the top the differential cross sections for top quarks ( $t$ ) in the middle for top antiquarks ( $\bar{t}$ ) and on the bottom the sum of the two are shown. The measured distributions are compared to POWHEG in the four-flavour scheme (4FS, blue) and the five-flavour scheme (5FS, orange), and to AMC@NLO in the 4FS scheme (green). In the ratio panel, the black line marks the 1. The horizontal lines on the error bars indicate the uncertainties that are profiled in the fit.



**Figure 8.29: Unfolding results for the  $z$  polarization angles at parton level.** Shown are the measured normalized differential cross sections as a function of the polarization angles  $\cos(\theta_z)$ . On the top-left the differential cross sections for top quarks ( $t$ ) in the top-right for top antiquarks ( $\bar{t}$ ) and on the bottom the sum of the two are shown. The measured distributions are compared to POWHEG in the four-flavour scheme (4FS, blue) and the five-flavour scheme (5FS, orange), and to aMC@NLO in the 4FS scheme (green). In the ratio panel, the black line marks the 1. The horizontal lines on the error bars indicate the uncertainties that are profiled in the fit.

### 8.12.4 Top quark polarization

The normalized differential angular distribution at parton level  $\cos(\theta_i)$  is related to the top quark polarization component  $P_i$  via:

$$\frac{d\sigma}{\sigma d\cos\theta_i} = \frac{1}{2}(1 + 2A_i \cos\theta_i), \quad \text{with : } A_i = \frac{1}{2}P_i\alpha_\ell. \quad (8.26)$$

Here,  $A_i$  is the forward-backward asymmetry (also called spin-asymmetry) in the measured angular distributions. There are three asymmetries, corresponding to the transverse ( $A_x$ ), normal ( $A_y$ ) and longitudinal ( $A_z$ ) polarization components of the top quark. Each asymmetry is extracted from the corresponding unfolded  $\cos(\theta_i)$  distributions at parton level by fitting the following linear function to the distribution:

$$f(x) = b \cdot (1 + 2 \cdot A \cdot x), \quad (8.27)$$

where  $A$  is the spin-asymmetry from equation 8.26. The fit is performed using the *optimize* module from *SCIPY* [200]. The measured asymmetries are:

$A_x(t) = -0.06 \pm 0.01$ (prof) $\pm 0.05$ (sig-mod.)	$= -0.06 \pm 0.05$ (syst)
$A_x(\bar{t}) = -0.06 \pm 0.01$ (prof) $\pm 0.06$ (sig-mod.)	$= -0.06 \pm 0.06$ (syst)
$A_x(t + \bar{t}) = -0.07 \pm 0.02$ (prof) $\pm 0.09$ (sig-mod.)	$= -0.07 \pm 0.09$ (syst)
$A_y(t) = +0.00 \pm 0.01$ (prof) $\pm 0.03$ (sig-mod.)	$= +0.00 \pm 0.03$ (syst)
$A_y(\bar{t}) = -0.02 \pm 0.01$ (prof) $\pm 0.04$ (sig-mod.)	$= -0.02 \pm 0.04$ (syst)
$A_y(t + \bar{t}) = -0.00 \pm 0.01$ (prof) $\pm 0.05$ (sig-mod.)	$= -0.00 \pm 0.05$ (syst)
$A_z(t) = +0.42 \pm 0.01$ (prof) $\pm 0.03$ (sig-mod.)	$= +0.42 \pm 0.03$ (syst)
$A_z(\bar{t}) = +0.47 \pm 0.02$ (prof) $\pm 0.08$ (sig-mod.)	$= +0.47 \pm 0.08$ (syst)
$A_z(t + \bar{t}) = +0.42 \pm 0.03$ (prof) $\pm 0.07$ (sig-mod.)	$= +0.42 \pm 0.08$ (syst)

Here "prof" corresponds to the systematic uncertainties that are profiled during the unfolding and "sig-mod." to the theory uncertainties in the signal modeling. The systematic uncertainty "syst" is calculated as the quadratic sum of the two. Assuming  $\alpha_\ell = 1$  the top quark polarization  $P_i$  can be computed from the measured spin asymmetries according to  $P_i = 2 A_i$ . The measured asymmetries for  $A_y$  and  $A_z$  agree well with the SM prediction for both top quarks and top antiquarks. Only the asymmetry  $A_x(\bar{t})$  deviates by about 1.8 standard deviation from the SM prediction of  $A_x^{\text{pred.}}(\bar{t}) = 0.04$ .

### 8.13 Limits on EFT coefficients

In  $t$ -channel single top quark production the top quark is polarized due to the V-A structure of the weak interaction. Because of this structure only left-handed, polarized top quarks are produced. Therefore, new physic effects that introduce a right-handed interaction change the polarization of the top quark. In the EFT approach such effects can be described by the coefficient  $C_{tW}$ . The coefficient  $C_{tW}$  has an imaginary part that describes possible CP-violating interactions. The distributions of  $\cos(\theta_x)$  and  $\cos(\theta_y)$  for different values of the real and imaginary part of  $C_{tW}$  are shown in figure 8.30. The  $\cos(\theta_x)$  distribution has a strong dependence on the real part of  $C_{tW}$  while showing no significant dependence on the imaginary part. For the  $\cos(\theta_y)$  distribution it is the other way around, it shows a strong dependence on the imaginary part and no dependence on the real part. The measured differential distributions of  $\cos(\theta_x)$  and  $\cos(\theta_y)$  are therefore used to individually constraint the real and imaginary part of  $C_{tW}$ .

In the presence of  $N$  dimension six operators the rate of any process is given by:

$$\sigma_{\text{SM+EFT}} = \sigma_{\text{SM}} + \sum_i \frac{C_i}{\Lambda^2} \sigma_i + \sum_{i,j} \frac{C_i C_j^*}{\Lambda^4} \sigma_{ij}. \quad (8.28)$$

Here the  $C_i$  are the Wilson coefficients of the considered operators and  $\Lambda$  is the EFT scale. For a single operator this corresponds to a quadratic dependence. Predictions for different values of the Wilson coefficients are produced with MADGRAPH using the dim6TopEFT [55] model at LO. The SM predictions from MADGRAPH are validated by comparing normalized differential distributions to SM distributions obtained from POWHEG. A  $d$ -dimensional quadratic function has  $N = 1 + 2d + \frac{d}{2} \cdot (d - 1)$  degrees of freedom. Therefore, at least that many EFT predictions with different coefficient values need to be generated. To compute predictions at arbitrary values of the Wilson coefficients a quadratic interpolation function  $f_i(\mathbf{c})$  that is a function of the Wilson coefficients  $\mathbf{c} = (\text{Re } C_{tW}, \text{Im } C_{tW})$  is fitted in each bin  $i$  of the  $\cos(\theta_x)$  and  $\cos(\theta_y)$  distributions:

$$f_i(\mathbf{c}) = \frac{\Delta\sigma_i(\mathbf{c})}{\sigma_{\text{SM},i}}, \quad (8.29)$$

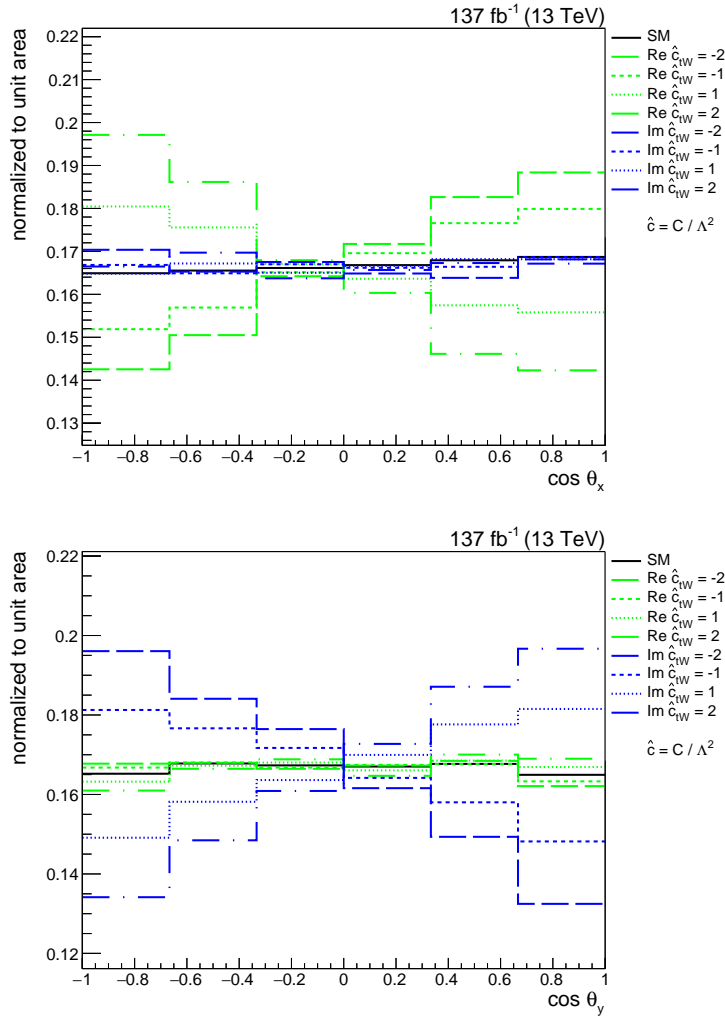
where  $\Delta\sigma_i(\mathbf{c})$  is the difference between the SM prediction in bin  $i$  and the prediction of the EFT model in bin  $i$ . The interpolation procedure is validated by comparing generated distributions for given coefficients with the interpolated ones.

A  $\chi^2$ -minimization is performed in order to find the set of coefficients that best describe the measured distributions:

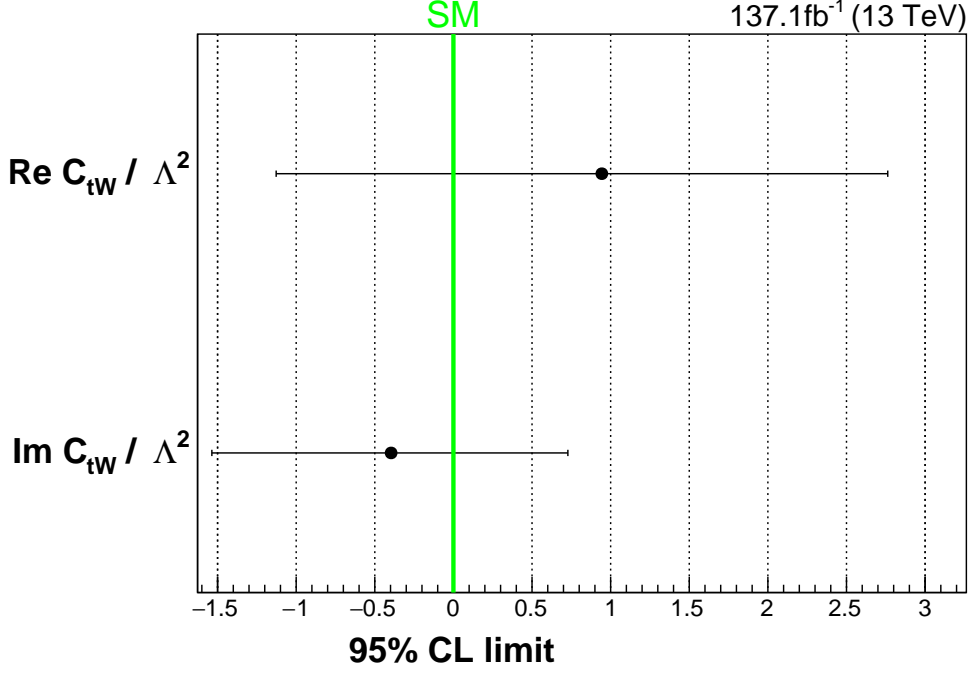
$$\chi^2 = \sum_{i,j} \Delta_i \cdot \Delta_j \cdot V_{ij}^{-1}. \quad (8.30)$$

The sums run over all bins in the corresponding measured angular distribution,  $V_{ij}$  is the covariance matrix of the measurement, and  $\Delta_i = (\sigma_{i,\text{obs}}^k - \sigma_{i,\text{EFT}}^k(\mathbf{c}))$  is the difference between the measured normalized cross section in bin  $i$  and the predicted normalized cross section as a function of the Wilson coefficients  $\mathbf{c}$  in bin  $i$ . The prediction is





**Figure 8.30: Effect of EFT coefficients on angular distributions.** Shown are the normalized angular distributions  $\cos(\theta_x)$  (top) and  $\cos(\theta_y)$  (bottom). The distributions are shown for the SM case and for different values of the real and imaginary part of EFT coefficients  $C_{tW}$ .



**Figure 8.31: Limits on Wilson coefficients.** Shown are the observed limits on the real and imaginary part of the Wilson coefficient  $C_{tW}$ . The nominal point corresponds to the best-fit result from the  $\chi^2$  minimization to the unfolded distribution. Shown are individual limits on the real and imaginary part of  $C_{tW}$ , which have been computed from the  $\cos(\theta_x)$  and  $\cos(\theta_y)$  differential distributions, respectively.

computed from the interpolation function. Distributions are normalized and only the shape information is used. The  $\cos(\theta_x)$  distribution is used to find the best fit value for  $\text{Re } C_{tW}$  and the  $\cos(\theta_y)$  distribution is used to find the best fit value of  $\text{Im } C_{tW}$ . The best fit values are  $(\text{Re } C_{tW}, \text{Im } C_{tW}) = (0.94, -0.40)$ . From the distribution of the  $\chi^2$  functions 95 % confidence levels on the real and imaginary parts of  $C_{tW}$  are computed. The constraints are shown in figure 8.31 and the values are:  $-1.13 < \text{Re } C_{tW}/\Lambda^2 < 2.8$  and  $-1.54 < \text{Im } C_{tW}/\Lambda^2 < 0.73$ .

## 9 Conclusion

A measurement of the inclusive and differential  $t$ -channel single top quark production cross section is performed in this thesis. The measurement uses  $137\text{ fb}^{-1}$  of data recorded at the CMS experiment at the LHC with a center-of-mass energy of 13 TeV. Events are selected with exactly one muon or electron and two or three jets, of which at least one is identified as originating from a bottom quark. In the analysis an improved technique for reconstructing the top quark has been developed that makes use of a neural network in order to achieve a better description of the top quark's kinematic variables. A multiclassification BDT is used to classify events into different process categories.

The cross sections are extracted from a fit to the output distribution of the multiclassification BDT. The inclusive cross section of  $t$ -channel single top quark production was measured to be  $\sigma_t = 130 \pm 20\text{ pb}$  and the cross section of top antiquark production to be  $\sigma_{\bar{t}} = 80 \pm 15\text{ pb}$ . Both numbers are in agreement with the predictions of the SM. The ratio of the single top quark to top antiquark production is computed from the measured cross sections to be  $R_{t\text{-ch.}} = \sigma_t/\sigma_{\bar{t}} = 1.63 \pm 0.06$ . The computed value is compared with the predictions of different PDF sets and agrees with all of them within 2 standard deviations. From the sum of the measured cross sections the CKM matrix element  $V_{tb}$  is determined to be:  $|f_L V_{tb}| = 0.99 \pm 0.08\text{ (exp)} \pm 0.02\text{ (theo)}$ .

The differential cross section measurement is performed via unfolding, which is implemented as a two-dimensional fit to the output distribution of a multiclassification BDT and to the variable for which the differential cross section measurement is performed. The usage of the BDT as an ancillary dimension allows all events selected in the analysis to be used in the fit instead of performing the unfolding in a signal-enriched phase-space region. Performing the unfolding through a fit allows to determine the amount of background events simultaneously with the amount of signal events and it is therefore not necessary to perform a separate background estimation and subtraction prior to unfolding. The measured differential cross sections as a function of the top quark transverse momentum and rapidity agree with the predictions of the SM. Three angular variables,  $\cos(\theta_x)$ ,  $\cos(\theta_y)$ , and  $\cos(\theta_z)$ , are defined in the top quark rest frame between the charged lepton from the top quark decay and three axes. These axes are defined based on the direction of the spectator quark and the beamline axis. The transverse ( $x$ ), normal ( $y$ ), and longitudinal ( $z$ ) components of the top quark polarization are proportional to an asymmetry in these angular variables. The asymmetries are measured to be:  $A_x(t + \bar{t}) = -0.07 \pm 0.09$ ,  $A_y(t + \bar{t}) = 0.00 \pm 0.05$ , and  $A_z(t + \bar{t}) = 0.42 \pm 0.08$ . All three are in agreement with the predictions of the SM.

The transverse and normal components of the top quark polarization are very sensitive to new physics contributions that would introduce a right-handed coupling in the  $Wtb$  interaction. Such contributions can be described in an EFT approach by the Wilson coefficient  $C_{tW}$ . Using MC simulation, predictions for the distributions of  $\cos(\theta_x)$  and  $\cos(\theta_y)$  for different values of the real and imaginary part of  $C_{tW}$  are generated. A  $\chi^2$ -minimization is performed to determine the values that agree best with the measured distributions. From the  $\chi^2$  distribution the following 95% confidence level intervals are extracted:  $-1.13 < \text{Re } C_{tW}/\Lambda^2 < 2.8$  and  $-1.54 < \text{Im } C_{tW}/\Lambda^2 < 0.73$ .

The uncertainties in the measurement are dominated by systematic uncertainties with the largest uncertainty being the theory uncertainties in the signal modeling. The statistical uncertainty in the inclusive cross section measurement is less than 1%. It is therefore not expected that the precision can be significantly increased with more data alone. Instead a better understanding of the signal modeling uncertainties is required. The larger amount of data that will be available in Run 3 of the LHC and at the planned HL-LHC will still be useful for multidifferential cross section measurements, measurements in boosted regimes, and for rarer single top quark production modes, such as the  $s$ -channel production, and associated productions with a Z or Higgs boson. At the HL-LHC not only the luminosity will be increased, but the CMS detector will be upgraded. One of the major upgrades will be an extension of the tracker up to pseudorapidity values of 4. This will allow for both lepton reconstruction and b tagging to be performed in a larger fiducial volume. Therefore, also increasing the fiducial volume for future top quark measurements. This could reduce the systematic uncertainties due to the extrapolation from the fiducial phase space to the full phase space. As has been demonstrated as part of this thesis, the performance of b tagging algorithms at the HL-LHC is expected to be comparable to their performance during Run 2. Current b tagging algorithms are in fact flavor tagging algorithms, as they are capable of identifying jets originating from bottom quarks, charm quarks, and light quarks (u,d,s) or gluons. Especially light quark tagging is currently not used in analysis. If its efficiency could be further improved, the use of light quark tagging could prove to be beneficial to tag the spectator jet in  $t$ -channel single top quark production. This could help to reduce the  $t\bar{t}$  background, which is the dominant background process for single top  $t$ -channel production. This will be particularly interesting at the HL-LHC because tagging information will be available up to large pseudorapidity values and the spectator jet is produced predominantly in forward direction. Therefore, the study of single top quark production will remain an interesting field of research at the LHC.

# **A Appendix**

## **A.1 List of samples**

**Table A.1: List of simulated samples used for 2016.** Shown are the simulated samples as well as the number of events and the cross section in pb.

Name	Events	Cross section / pb
ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8	17 780 700	80.95
ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5down_PSweights_13TeV-powheg-pythia8	998 800	80.95
ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5up_PSweights_13TeV-powheg-pythia8	1 000 000	80.95
ST_t-channel_top_4f_InclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8	31 848 000	136.02
ST_t-channel_top_4f_InclusiveDecays_TuneCP5down_PSweights_13TeV-powheg-pythia8	1 000 000	136.02
ST_t-channel_top_4f_InclusiveDecays_TuneCP5up_PSweights_13TeV-powheg-pythia8	1 000 000	136.02
ST_tW_antitop_5f_inclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8	4 980 600	35.85
ST_tW_top_5f_inclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8	4 983 500	35.85
TTTo2L2Nu_TuneCP5_PSweights_13TeV-powheg-pythia8	67 860 400	88.23
TTTo2L2Nu_TuneCP5down_PSweights_13TeV-powheg-pythia8	13 903 300	88.23
TTTo2L2Nu_TuneCP5up_PSweights_13TeV-powheg-pythia8	14 838 600	88.23
TTTo2L2Nu_hdampDOWN_TuneCP5_PSweights_13TeV-powheg-pythia8	14 836 800	88.23
TTTo2L2Nu_hdampUP_TuneCP5_PSweights_13TeV-powheg-pythia8	14 889 100	88.23
TTToSemiLeptonic_TuneCP5_PSweights_13TeV-powheg-pythia8	105 539 600	365.33
TTToSemiLeptonic_TuneCP5down_PSweights_13TeV-powheg-pythia8	26 388 000	365.33
TTToSemiLeptonic_TuneCP5up_PSweights_13TeV-powheg-pythia8	26 050 600	365.33
TTToSemiLeptonic_hdampDOWN_TuneCP5_PSweights_13TeV-powheg-pythia8	29 702 200	365.33
TTToSemiLeptonic_hdampUP_TuneCP5_PSweights_13TeV-powheg-pythia8	29 376 200	365.33
WJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	257 931 694	61 526.7
WToLNu_0J_13TeV-amcatnloFXFX-pythia8	49 350 414	50 131.98
WToLNu_1J_13TeV-amcatnloFXFX-pythia8	41 432 585	8426.09
WToLNu_2J_13TeV-amcatnloFXFX-pythia8	251 576 435	3172.96
DYJetsToLL_M-50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	120 777 245	6077.22

**Table A.2: List of simulated samples used for 2017.** Shown are the simulated samples as well as the number of events and the cross section in pb.

Name	Events	Cross section / pb
ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8	129 541 600	80.95
ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5down_PSweights_13TeV-powheg-pythia8	23 440 500	80.95
ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5up_PSweights_13TeV-powheg-pythia8	22 394 800	80.95
ST_t-channel_top_4f_InclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8	243 648 400	136.02
ST_t-channel_top_4f_InclusiveDecays_TuneCP5down_PSweights_13TeV-powheg-pythia8	46 085 400	136.02
ST_t-channel_top_4f_InclusiveDecays_TuneCP5up_PSweights_13TeV-powheg-pythia8	40 331 600	136.02
ST_tW_antitop_5f_NoFullyHadronicDecays_TuneCP5_PSweights_13TeV-powheg-pythia8	5 577 319	19.55
ST_tW_top_5f_NoFullyHadronicDecays_TuneCP5_PSweights_13TeV-powheg-pythia8	5 103 599	19.55
TTTo2L2Nu_TuneCP5_PSweights_13TeV-powheg-pythia8	69 155 808	88.23
TTTo2L2Nu_TuneCP5down_PSweights_13TeV-powheg-pythia8	15 109 000	88.23
TTTo2L2Nu_TuneCP5up_PSweights_13TeV-powheg-pythia8	15 403 000	88.23
TTTo2L2Nu_hdampDOWN_TuneCP5_PSweights_13TeV-powheg-pythia8	15 439 459	88.23
TTTo2L2Nu_hdampUP_TuneCP5_PSweights_13TeV-powheg-pythia8	13 161 127	88.23
TTToHadronic_TuneCP5_PSweights_13TeV-powheg-pythia8	130 262 440	378.19
TTToHadronic_TuneCP5down_PSweights_13TeV-powheg-pythia8	27 252 808	378.19
TTToHadronic_TuneCP5up_PSweights_13TeV-powheg-pythia8	27 108 792	378.19
TTToHadronic_hdampDOWN_TuneCP5_PSweights_13TeV-powheg-pythia8	27 022 682	378.19
TTToHadronic_hdampUP_TuneCP5_PSweights_13TeV-powheg-pythia8	27 070 600	378.19
TTToSemiLeptonic_TuneCP5_PSweights_13TeV-powheg-pythia8	110 014 744	365.33
TTToSemiLeptonic_TuneCP5down_PSweights_13TeV-powheg-pythia8	50 015 727	365.33
TTToSemiLeptonic_TuneCP5up_PSweights_13TeV-powheg-pythia8	20 122 010	365.33
TTToSemiLeptonic_hdampDOWN_TuneCP5_PSweights_13TeV-powheg-pythia8	53 217 628	365.33
TTToSemiLeptonic_hdampUP_TuneCP5_PSweights_13TeV-powheg-pythia8	23 977 012	365.33
WJetsToLNu_0J_TuneCP5_13TeV-amcatnloFXFX-pythia8	180 935 349	50 131.98
WJetsToLNu_1J_TuneCP5_13TeV-amcatnloFXFX-pythia8	259 274 926	8426.09
WJetsToLNu_2J_TuneCP5_13TeV-amcatnloFXFX-pythia8	195 908 401	3172.96
DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8	209 747 524	6077.22

**Table A.3: List of simulated samples used for 2018.** Shown are the simulated samples as well as the number of events and the cross section in pb.

Name	Events	Cross section / pb
ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8	79 090 800	80.95
ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5down_13TeV-powheg-madspin-pythia8	23 386 400	80.95
ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5up_13TeV-powheg-madspin-pythia8	23 979 400	80.95
ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8	154 307 600	136.02
ST_t-channel_top_4f_InclusiveDecays_TuneCP5down_13TeV-powheg-madspin-pythia8	48 101 600	136.02
ST_t-channel_top_4f_InclusiveDecays_TuneCP5up_13TeV-powheg-madspin-pythia8	49 377 100	136.02
ST_tW_antitop_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8	1 086 487	19.55
ST_tW_top_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8	8 722 734	19.55
TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8	64 310 000	88.23
TTTo2L2Nu_hdampDOWN_TuneCP5_13TeV-powheg-pythia8	15 398 000	88.23
TTTo2L2Nu_hdampUP_TuneCP5_13TeV-powheg-pythia8	14 974 000	88.23
TTToHadronic_TuneCP5_13TeV-powheg-pythia8	330 548 000	378.19
TTToHadronic_TuneCP5down_13TeV-powheg-pythia8	26 675 000	378.19
TTToHadronic_TuneCP5up_13TeV-powheg-pythia8	23 488 000	378.19
TTToHadronic_hdampDOWN_TuneCP5_13TeV-powheg-pythia8	26 425 000	378.19
TTToHadronic_hdampUP_TuneCP5_13TeV-powheg-pythia8	24 965 000	378.19
TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8	301 219 998	365.33
TTToSemiLeptonic_hdampDOWN_TuneCP5_13TeV-powheg-pythia8	25 904 000	365.33
TTToSemiLeptonic_hdampUP_TuneCP5_13TeV-powheg-pythia8	26 964 000	365.33
WJetsToLNu_0J_TuneCP5_13TeV-amcatnloFXFX-pythia8	192 288 265	50 131.98
WJetsToLNu_1J_TuneCP5_13TeV-amcatnloFXFX-pythia8	171 669 288	8426.09
WJetsToLNu_2J_TuneCP5_13TeV-amcatnloFXFX-pythia8	98 362 049	3172.96
DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8	194 213 235	6077.22

**Table A.4: List of data damples.** Shown are the data sets with their run ranges, as well as the integrated luminosity corresponding to the run range.

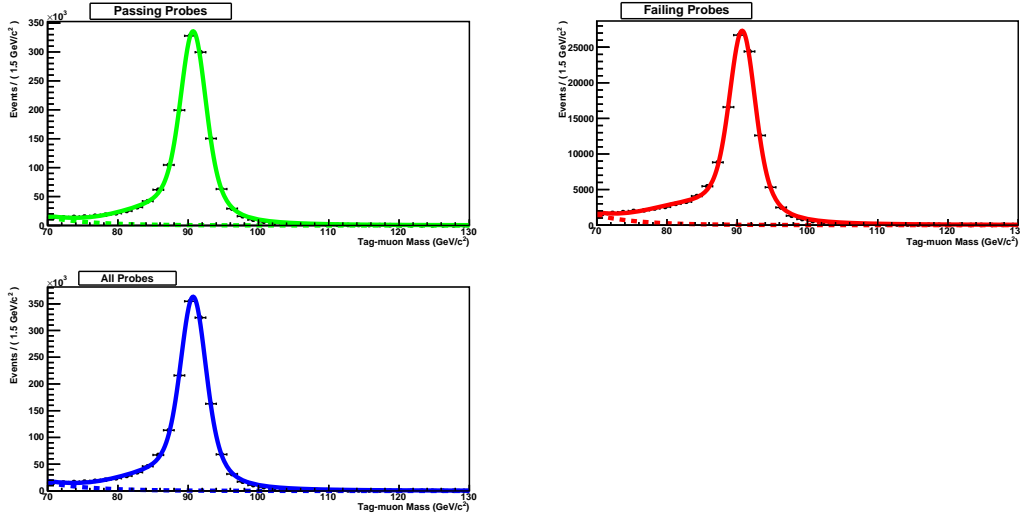
Dataset	Run name	Run range	Integrated luminosity
SingleMuon(Electron)	Run2016B	272007–275376	$5.8 \text{ fb}^{-1}$
SingleMuon(Electron)	Run2016C	275657–276283	$2.6 \text{ fb}^{-1}$
SingleMuon(Electron)	Run2016D	276315–276811	$4.2 \text{ fb}^{-1}$
SingleMuon(Electron)	Run2016E	276831–277420	$4.0 \text{ fb}^{-1}$
SingleMuon(Electron)	Run2016F	277772–278808	$3.1 \text{ fb}^{-1}$
SingleMuon(Electron)	Run2016G	278820–280385	$7.5 \text{ fb}^{-1}$
SingleMuon(Electron)	Run2016H	280919–284044	$8.6 \text{ fb}^{-1}$
Total 2016:		272007–284044	$35.9 \text{ fb}^{-1}$
2017			
SingleMuon(Electron)	Run2017B	297046–299329	$4.8 \text{ fb}^{-1}$
SingleMuon(Electron)	Run2017C	299368–302029	$9.7 \text{ fb}^{-1}$
SingleMuon(Electron)	Run2017D	302030–303434	$4.3 \text{ fb}^{-1}$
SingleMuon(Electron)	Run2017E	303824–304797	$9.3 \text{ fb}^{-1}$
SingleMuon(Electron)	Run2017F	305040–306462	$13.5 \text{ fb}^{-1}$
Total 2017:		272007–284044	$41.5 \text{ fb}^{-1}$
2018			
SingleMuon(EGamma)	Run2018A	315252–316995	$14 \text{ fb}^{-1}$
SingleMuon(EGamma)	Run2018B	317080–319310	$7.1 \text{ fb}^{-1}$
SingleMuon(EGamma)	Run2018C	319337–320065	$6.94 \text{ fb}^{-1}$
SingleMuon(EGamma)	Run2018D	320673–325175	$31.93 \text{ fb}^{-1}$
Total 2018:		315252–325175	$59.7 \text{ fb}^{-1}$
Total:			$137.1 \text{ fb}^{-1}$



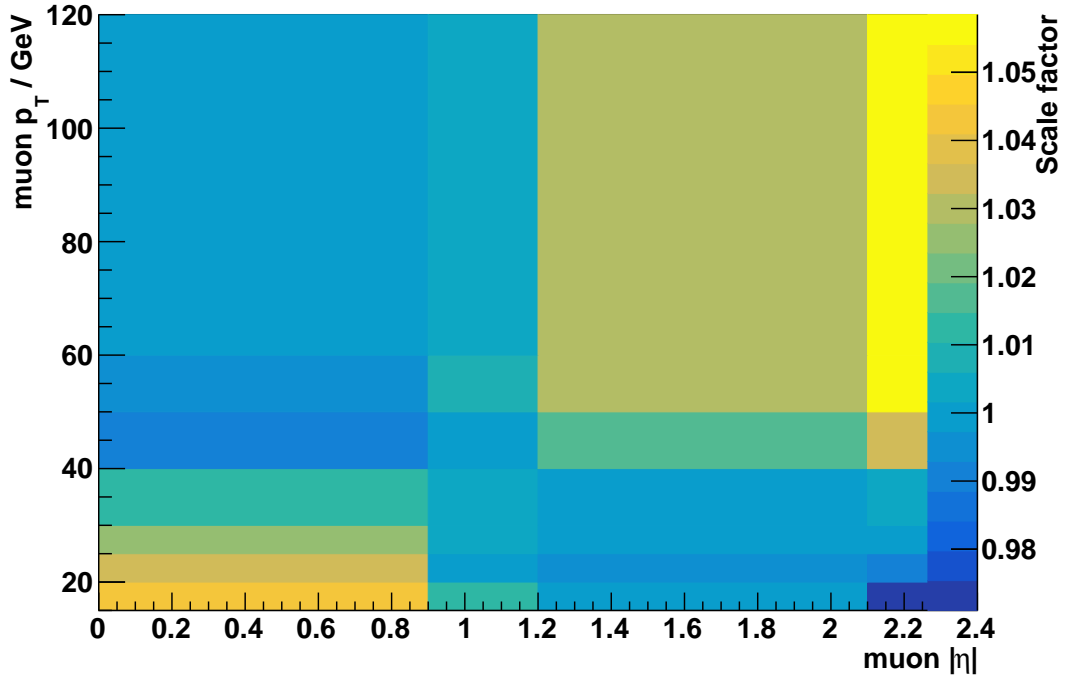
## A.2 Muon scale factors 2018

The muon isolation and trigger efficiencies in data and simulation are determined using a tag-and-probe method [98]. In this method "tag" muons are selected using strict selection criteria and "probe" muons using relaxed criteria. The relative amount of probe muons passing the selection gives an estimate of the efficiency. The efficiency in data is estimated on events taken in 2018 passing at least one muon trigger, while the efficiency in simulation is estimated using  $Z \rightarrow \ell^+ \ell^-$  events.

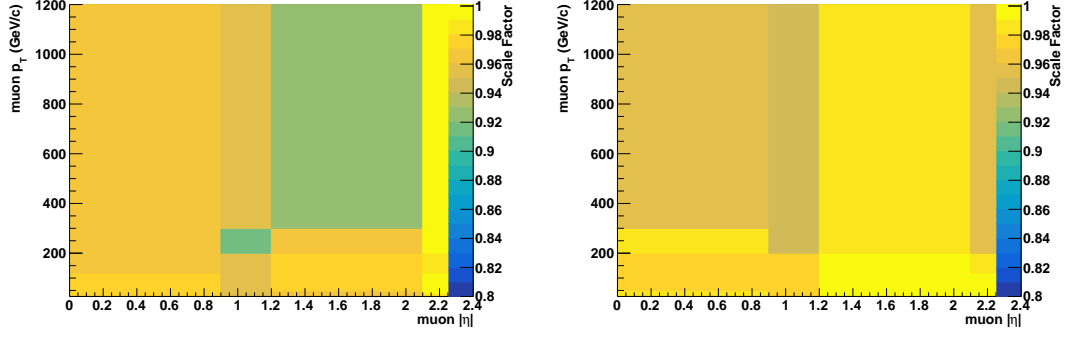
Muons must pass the trigger path HLT\_IsoMu27 and the tight muon identification criteria and their relative isolation must be  $I_{\text{rel.}} < 0.06$ . The transverse momentum of the tag must be larger than 26 GeV. For the trigger efficiencies the tag must also pass the trigger path HLT\_IsoMu24. The angular distance between the tag and the probe muon must be  $\Delta R(\text{tag, probe}) > 0.3$ . The invariant mass  $m_{\mu\mu}$  of the tag and the probe muon must lie within the Z boson mass window:  $70 \text{ GeV} < m_{\mu\mu} < 130 \text{ GeV}$ . The invariant mass is fitted in this range using a Voigtian function for the signal process and an exponential falling function for background processes (such as  $\gamma \rightarrow \mu\mu$ ). The efficiencies are extracted from the fit according to  $\epsilon = N_{\text{pass}}/N_{\text{pass+fail}}$ . An example fit is shown in figure A.1. The scale factors are computed according to  $\text{SF} = \epsilon_{\text{data}}/\epsilon_{\text{MC}}$ . The computed muon isolation scale factors are shown in figure A.2 and computed trigger scale factors in figure A.3. Systematic uncertainties are considered due to the choice of the mass window and the binning in the fit, and the relative isolation working point. The SFs are derived with systematically varied settings and the largest difference in each bin to the nominal SF is taken as systematic uncertainty.



**Figure A.1: Example fit to the Z boson mass.** Shown are the results of the fit to the Z boson mass in data events for events with  $p_T < 20$  GeV and  $|\eta| < 0.9$ . The dotted lines correspond to the fitted background. One such fit for both data and simulation is performed in each  $p_T$ - $|\eta|$  bin.

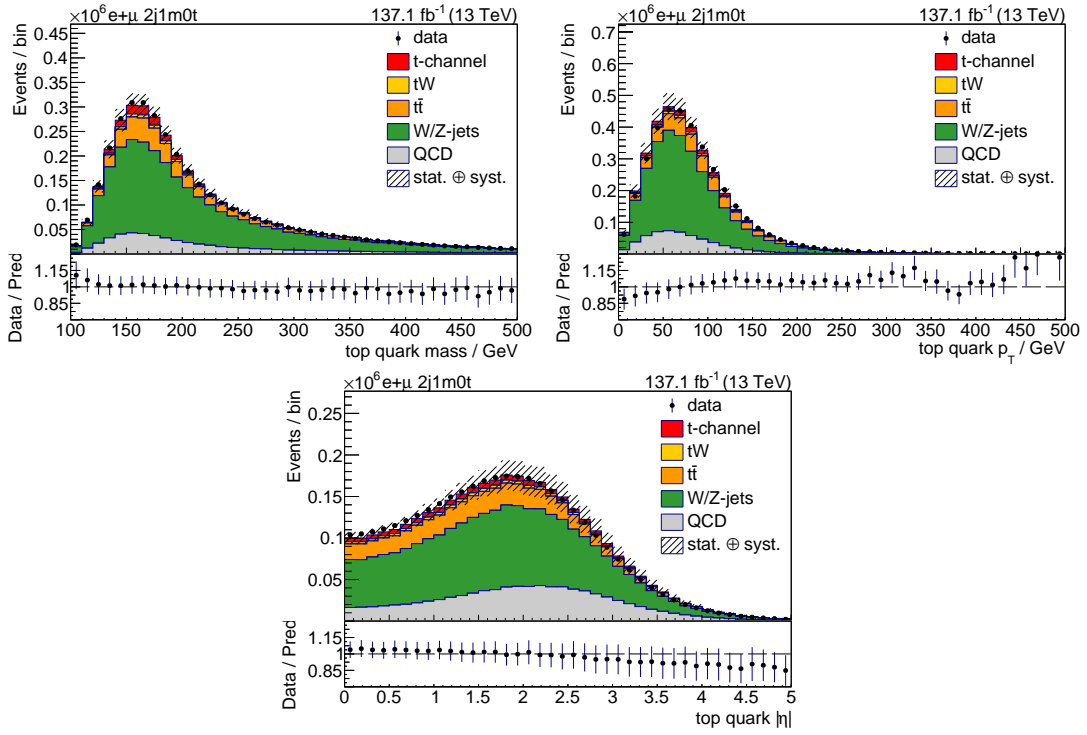


**Figure A.2: Muon isolation scale factors for 2018.** Shown are the muon isolation scale factors for muons with a relative isolation of  $I_{\text{rel.}} < 0.06$  for data taken in 2018. The scale factors have been derived using a tag-and-probe method.



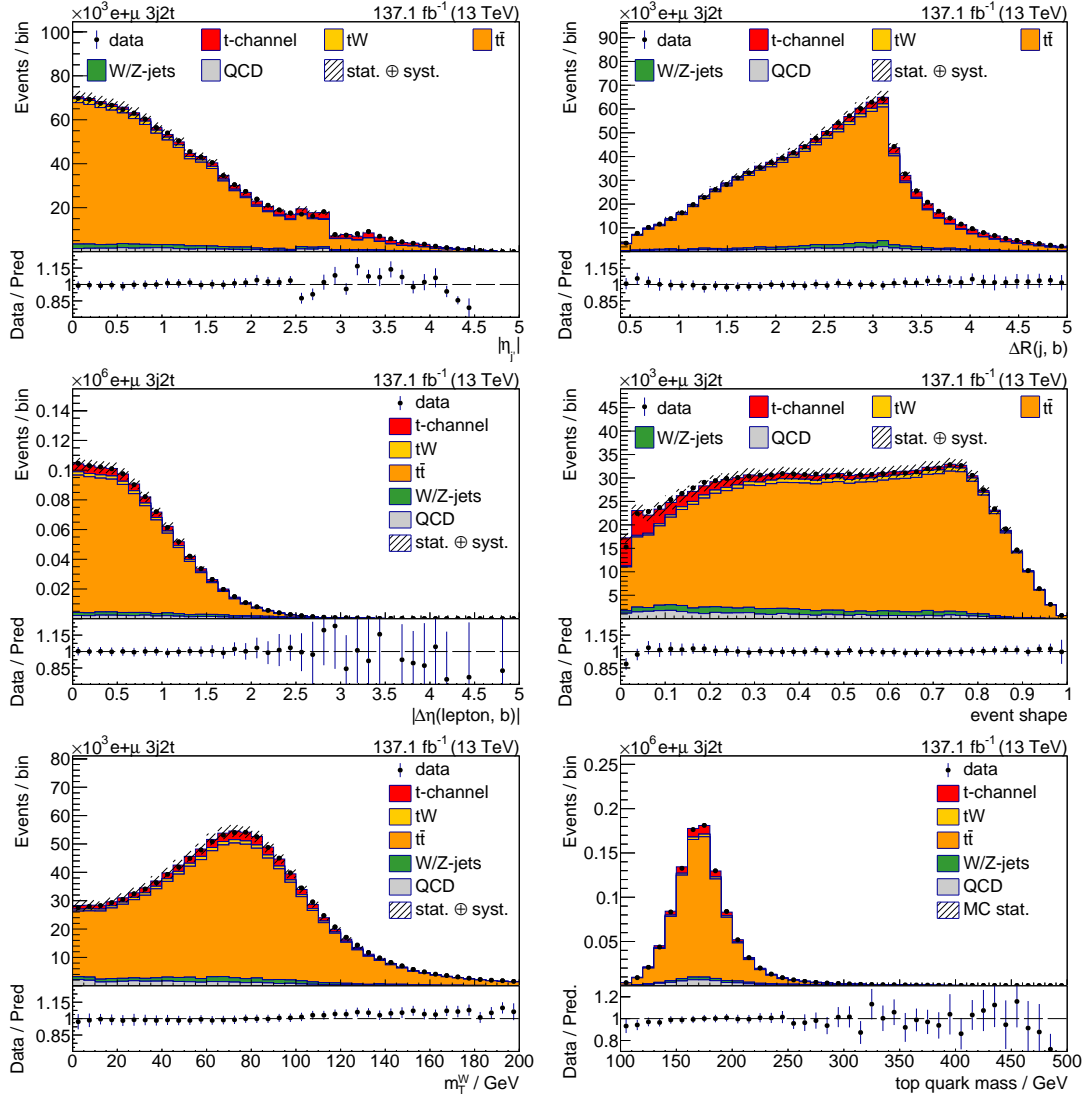
**Figure A.3: Muon trigger scale factors for 2018.** Shown are the muon trigger scale factors for muons with a relative isolation of  $I_{\text{rel.}} < 0.06$  for data taken in 2018. The scale factors are shown separately for data recorded before the trigger update in run number 316360 (left) and for data recorded after the update (right). The scale factors have been derived using a tag-and-probe method.

### A.3 Modeling of the reconstructed top quark in the 2j1m0t region

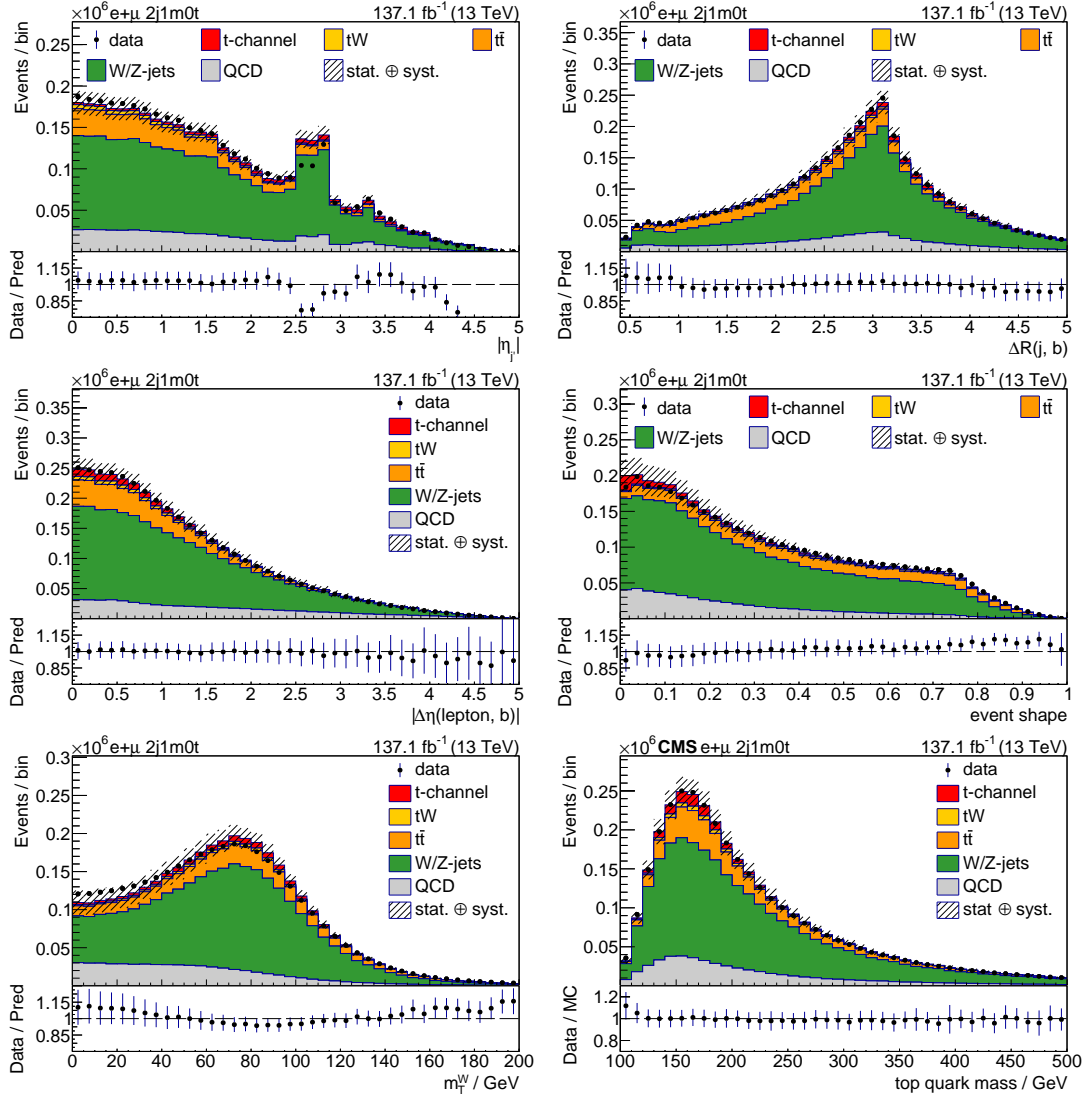


**Figure A.4: Reconstructed top quark kinematic variables in the 2j1m0t region.** The reconstructed top quark mass (top), transverse momentum (middle) and absolute pseudorapidity (bottom) are shown in the 2j1m0t region. The dashed area corresponds to the combined statistical and systematic uncertainties from the prediction. The prediction is scaled to match the data.

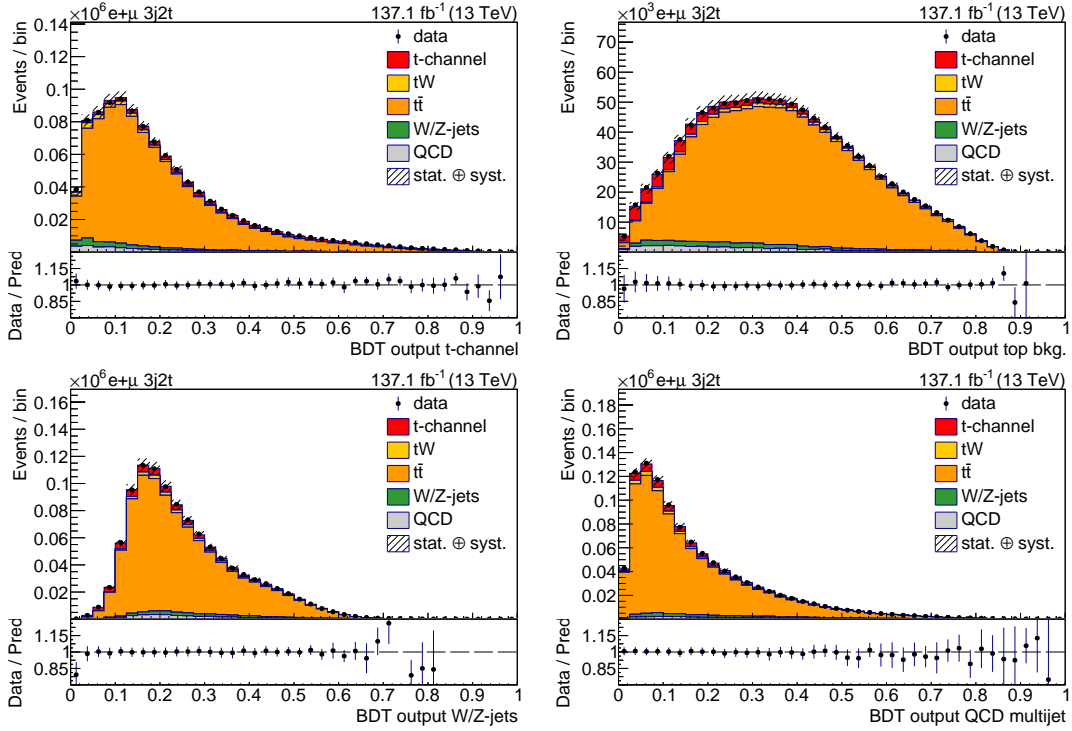
## **A.4 Modeling of BDT input and output variables in the control regions**



**Figure A.5: Modeling of BDT input variables in the  $t\bar{t}$  control region.** Shown are the shapes for the input variables used for the BDTs in the  $3j2t$   $t\bar{t}$ -enriched event category. The prediction is scaled to match the amount of data.

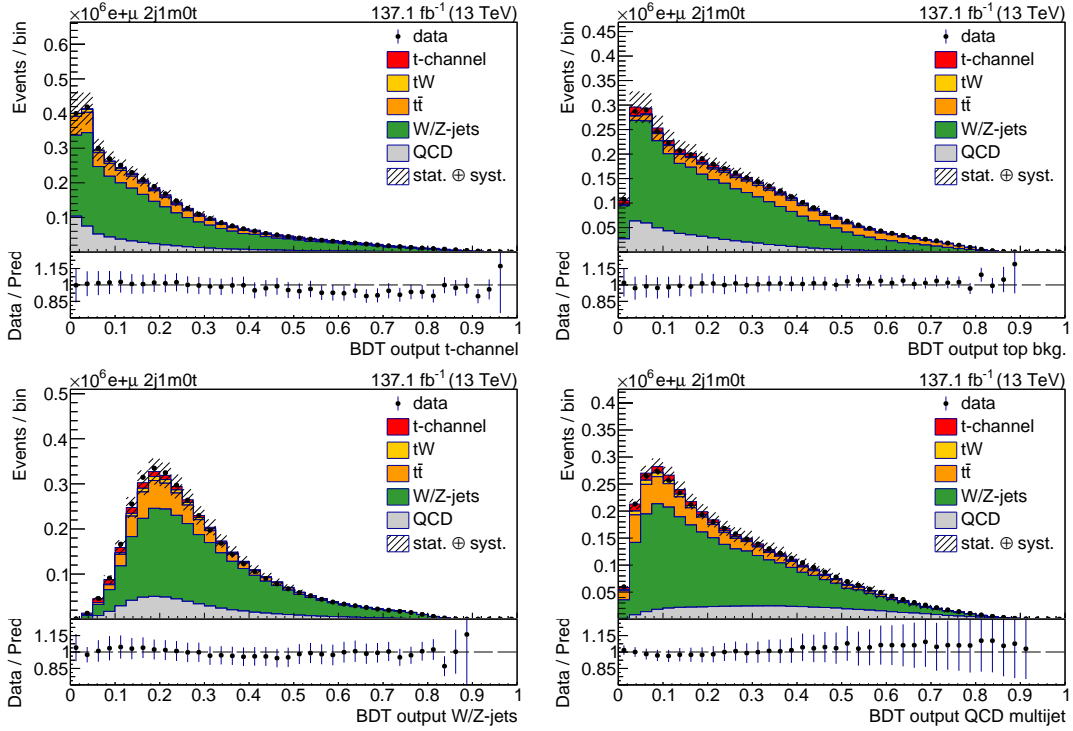


**Figure A.6: Modeling of BDT input variables in the W/Z-jets control region.** Shown are the shapes for the input variables used for the BDTs in the 2j1m0t W/Z-jets-enriched event category. The prediction is scaled to match the amount of data.



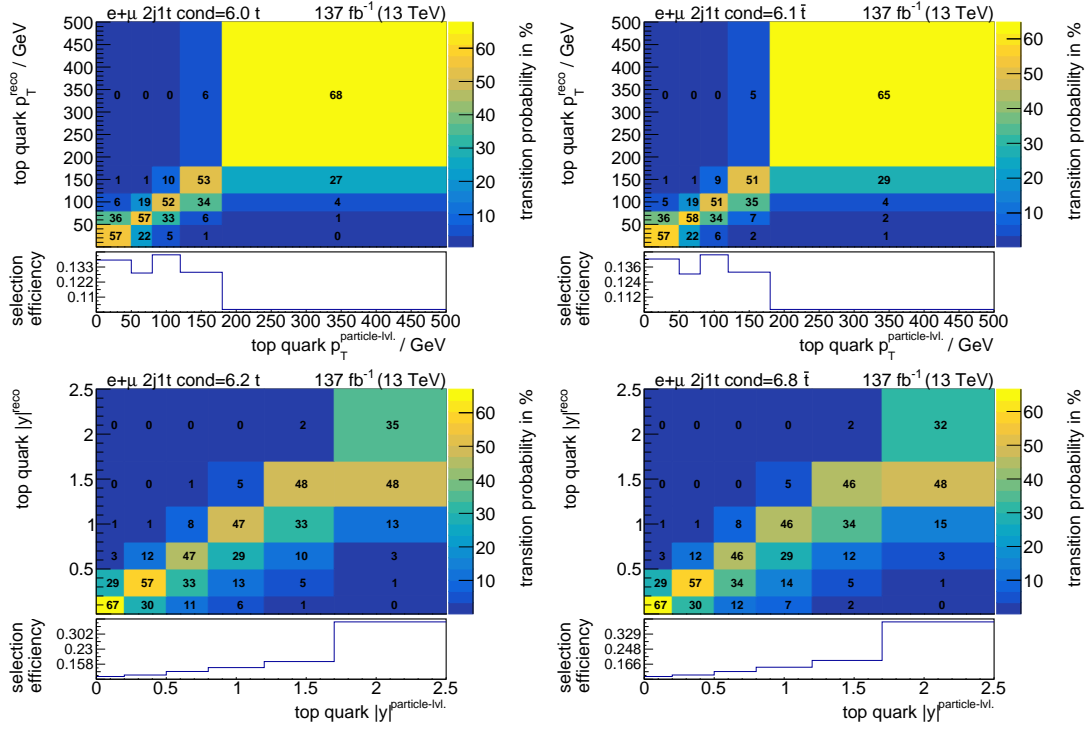
**Figure A.7: Comparison of the BDT output distributions to data in the  $t\bar{t}$  control region.** The top left figure shows the output distribution for the  $t$ -channel node, the top right the  $t\bar{t}$  and  $tW$  node. The bottom left figure shows the W/Z-jets node and the bottom right figure the the QCD multijet node.



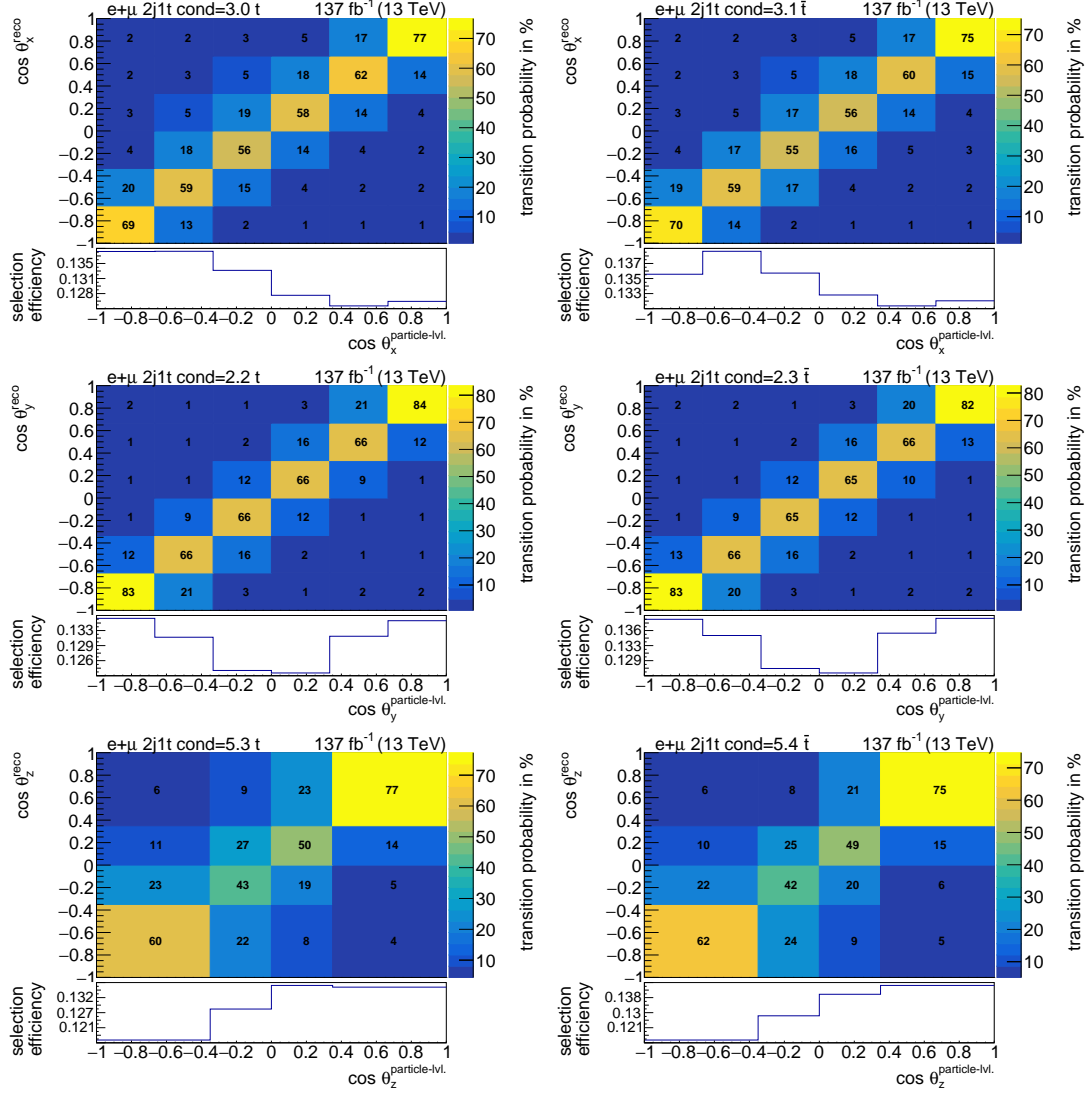


**Figure A.8: Comparison of the BDT output distributions to data in the W/Z-jets control region.** The top left figure shows the output distribution for the  $t$ -channel node, the top right figure the  $t\bar{t}$  and  $tW$  node. The bottom left figure shows the W/Z-jets node and the bottom right figure the QCD multijet node.

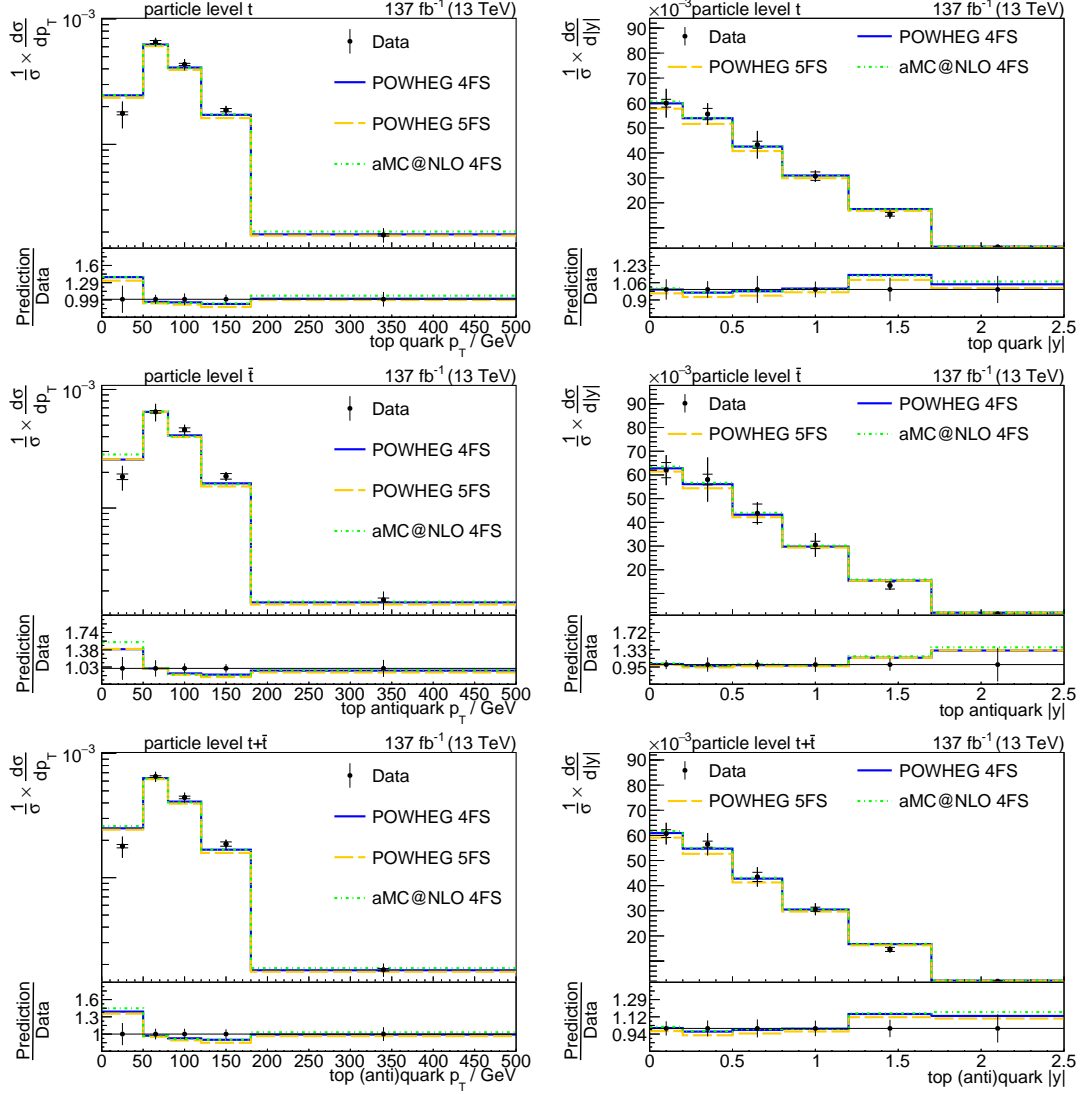
## **A.5 Particle level unfolding**



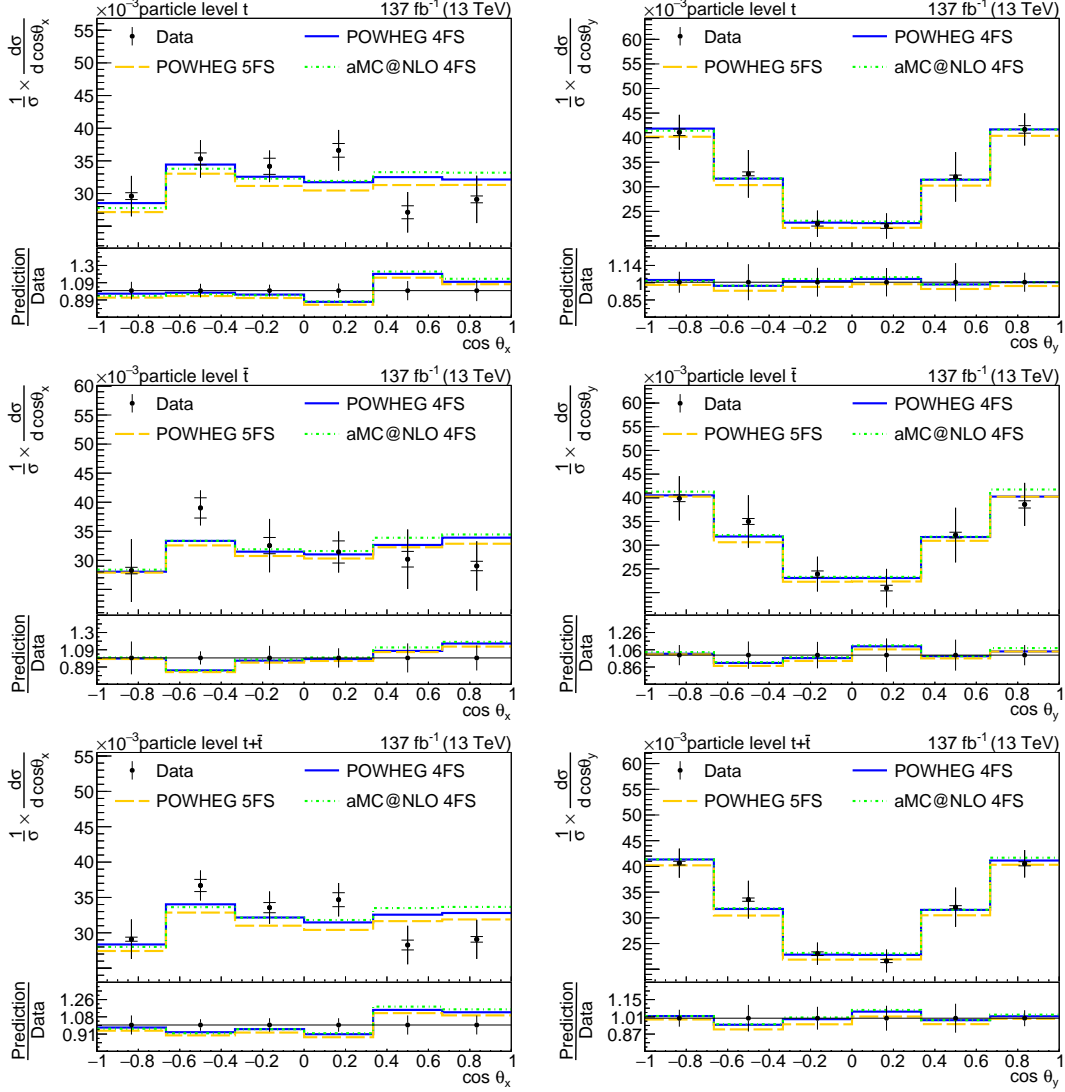
**Figure A.9: Response matrices for top quark transverse momentum and rapidity at particle level.** Shown are the response matrices and selection efficiencies for the top quark  $p_T$  (top) and  $|y|$  (bottom). The response matrices are shown at particle level for top quarks (left) and top antiquarks (right). They are obtained from simulated single-top  $t$ -channel events in the 2j1t region. The histograms are normalized such that the sum over the particle level bins sums up to 100%.



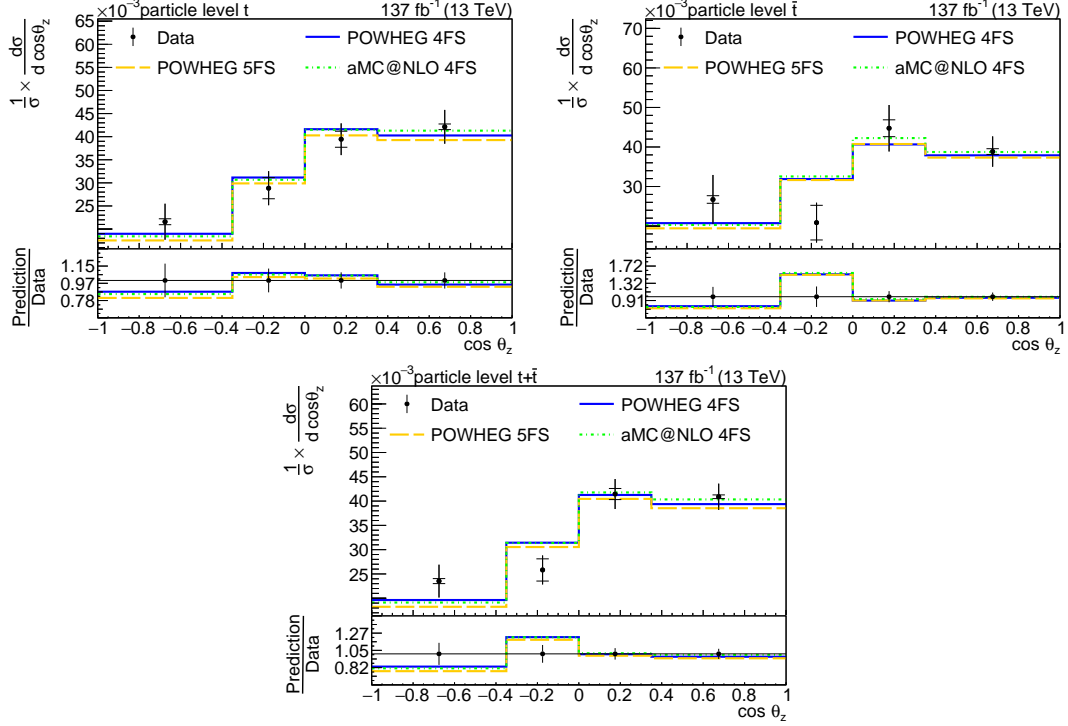
**Figure A.10: Response matrices for the polarization angles at particle level** Shown are the response matrices and selection efficiencies for the polarization angles  $\cos(\theta_x)$  (top),  $\cos(\theta_y)$  (middle) and  $\cos(\theta_z)$  (bottom). The response matrices are shown at particle level for top quarks (left) and top antiquarks (right). They are obtained from simulated single-top  $t$ -channel events in the 2j1t region. The histograms are normalized such that the sum over the particle level bins sums up to 100 %.



**Figure A.11: Unfolding results for top quark transverse momentum and rapidity at particle level.** Shown are the measured normalized differential cross sections as a function of the top quark  $p_T$  (left) and  $|y|$  (right). On the top the differential cross sections for top quarks ( $t$ ) in the middle for top antiquarks ( $\bar{t}$ ) and on the bottom the sum of the two are shown. The measured distributions are compared to POWHEG in the four-flavour scheme (4FS, blue) and the five-flavour scheme (5FS) (orange), and to AMC@NLO in the 4FS scheme (green). In the ratio panel, the black line marks the 1. The horizontal lines on the error bars indicate the uncertainties that are profiled in the fit.



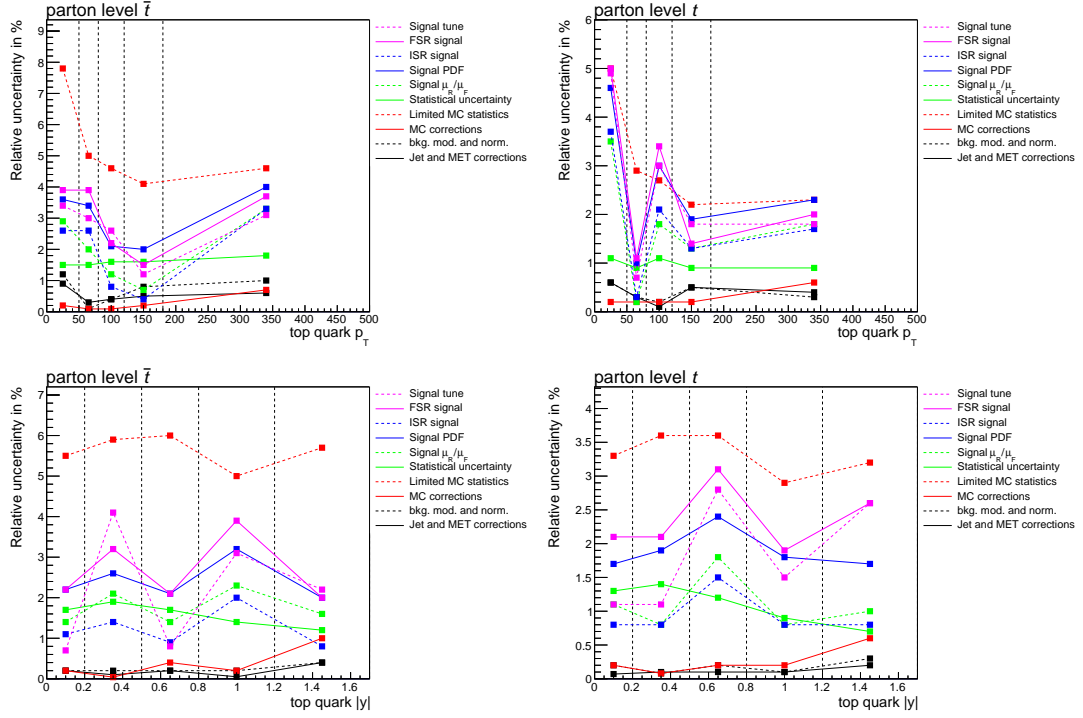
**Figure A.12: Unfolding results for the polarization angles  $x, y$  at particle level.** Shown are the measured normalized differential cross sections as a function of the polarization angles  $\cos(\theta_x)$  (left),  $\cos(\theta_y)$  (right). On the top the differential cross sections for top quarks ( $t$ ) in the middle for top antiquarks ( $\bar{t}$ ) and on the bottom the sum of the two are shown. The measured distributions are compared to POWHEG in the four-flavour scheme (4FS, blue) and the five-flavour scheme (5FS) (orange), and to aMC@NLO in the 4FS scheme (green). In the ratio panel, the black line marks the 1. The horizontal lines on the error bars indicate the uncertainties that are profiled in the fit.



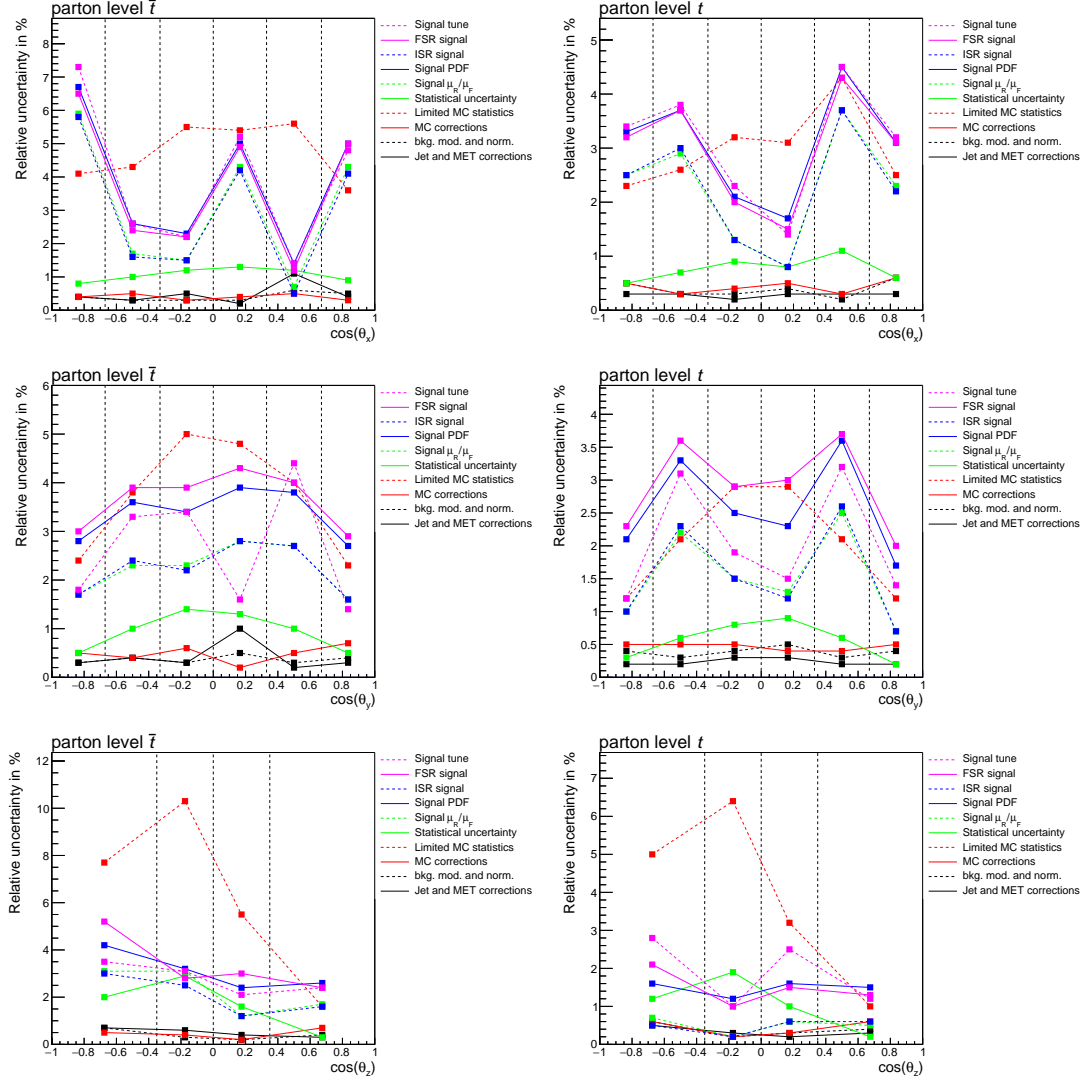
**Figure A.13: Unfolding results for the polarization angles  $z$  at particle level.** Shown are the measured normalized differential cross sections as a function of the polarization angle  $\cos(\theta_z)$ . On the top-right the differential cross sections for top quarks ( $t$ ) in the top-left for top antiquarks ( $\bar{t}$ ) and on the bottom the sum of the two are shown. The measured distributions are compared to POWHEG in the four-flavour scheme (4FS, blue) and the five-flavour scheme (5FS) (orange), and to aMC@NLO in the 4FS scheme (green). In the ratio panel, the black line marks the 1. The horizontal lines on the error bars indicate the uncertainties that are profiled in the fit.

## **A.6 Systematic uncertainty contributions**

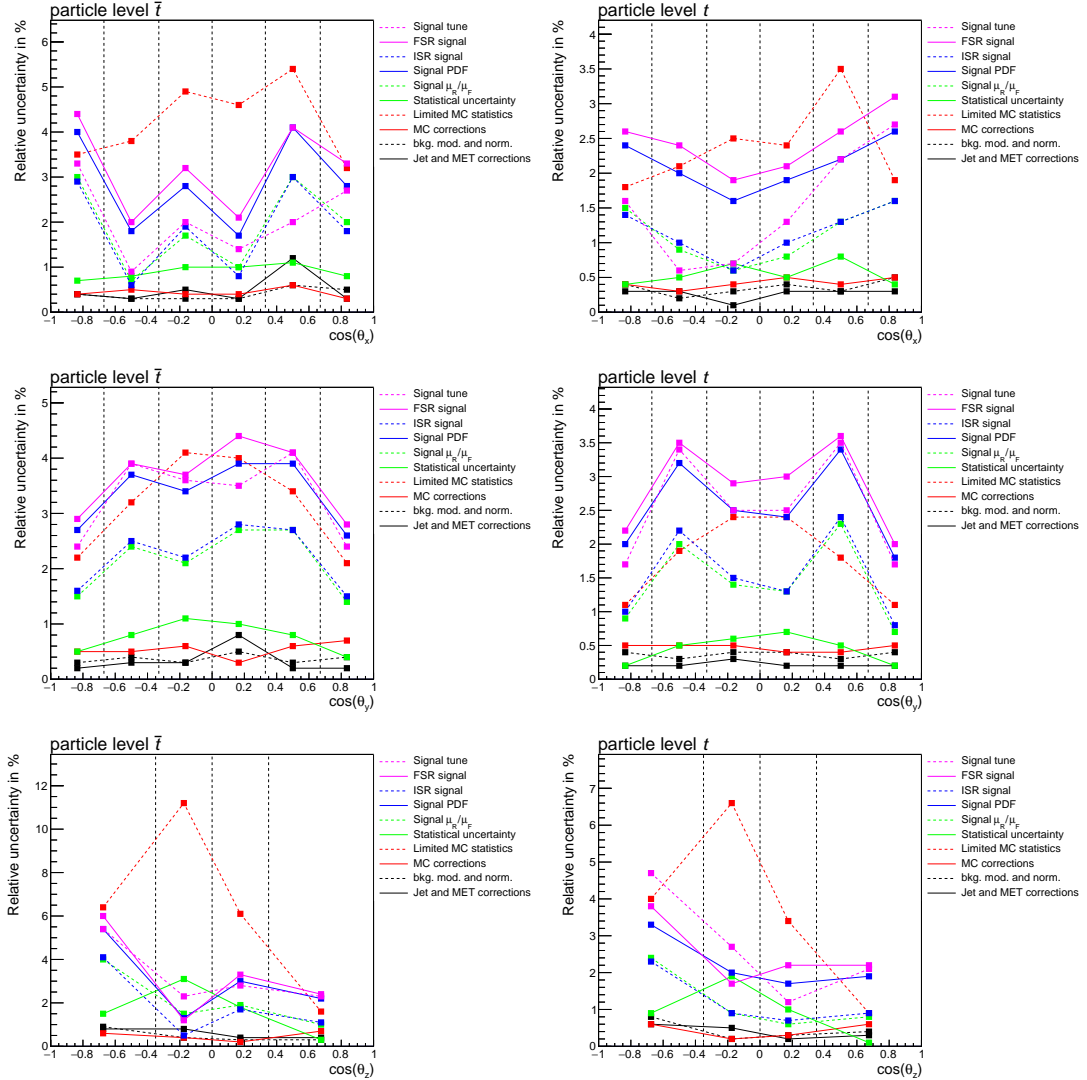




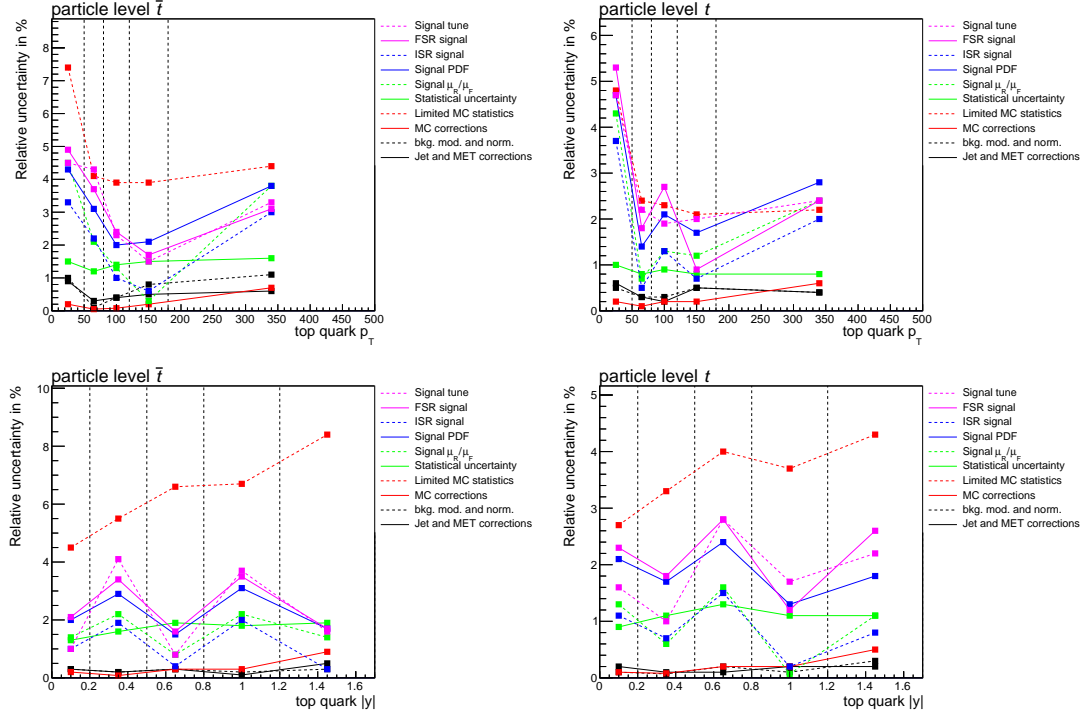
**Figure A.14: Systematic contributions for parton level top quark transverse momentum and rapidity.** Shown are the relative contributions to the systematic uncertainties for each bin of the parton level top quark  $p_T$  (top) and  $|y|$  (bottom), for top antiquarks (left) and top quarks (right). Under MC corrections, the contributions due to the lepton SFs, b tagging SFs, pileup reweighting and prefire corrections are summarized. The Jet and MET corrections are the combined contributions due to the jet energy scale corrections, the jet energy resolution corrections and the unclustered energy corrections. The marker show the bin centers and the dashed lines the bin edges.



**Figure A.15: Systematic contributions for parton level angular distributions.** Shown are the relative contributions to the systematic uncertainties for each bin of the parton level  $\cos(\theta_x)$  (top),  $\cos(\theta_y)$  (middle) and  $\cos(\theta_z)$  (bottom), for top antiquarks (left) and top quarks (right). Under MC corrections, the contributions due to the lepton SFs, b tagging SFs, pileup reweighting and prefire corrections are summarized. The Jet and MET corrections are the combined contributions due to the jet energy scale corrections, the jet energy resolution corrections and the unclustered energy corrections. The marker show the bin centers and the dashed lines the bin edges.



**Figure A.16: Systematic contributions for particle level top quark transverse momentum and rapidity.** Shown are the relative contributions to the systematic uncertainties for each bin of the particle level top quark  $p_T$  (top) and  $|y|$  (bottom), for top antiquarks (left) and top quarks (right). Under MC corrections, the contributions due to the lepton SFs, b tagging SFs, pileup reweighting and prefire corrections are summarized. The Jet and MET corrections are the combined contributions due to the jet energy scale corrections, the jet energy resolution corrections and the unclustered energy corrections. The marker show the bin centers and the dashed lines the bin edges.



**Figure A.17: Systematic contributions for parton level angular distributions.** Shown are the relative contributions to the systematic uncertainties for each bin of the parton level  $\cos(\theta_x)$  (top),  $\cos(\theta_y)$  (middle) and  $\cos(\theta_z)$  (bottom), for top antiquarks (left) and top quarks (right). Under MC corrections, the contributions due to the lepton SFs, b tagging SFs, pileup reweighting and prefire corrections are summarized. The Jet and MET corrections are the combined contributions due to the jet energy scale corrections, the jet energy resolution corrections and the unclustered energy corrections. The marker show the bin centers and the dashed lines the bin edges.

# List of Figures

2.1	The three dominant production modes for top quark pair production at the LHC. . . . .	12
2.2	Summary of measurements of top quark pair production cross section at the LHC and Tevatron . . . . .	13
2.3	The three dominant production modes for single top quark production at the LHC. . . . .	13
2.4	Feynman diagrams for single top quark and top antiquark production. .	14
2.5	Summary of measurements of the cross section of the different single top quark production channels at the LHC. . . . .	15
2.6	Illustration of the correlation between the spin of the particles involved in the single top quark production. . . . .	18
2.7	Illustration of the coordinate system for the top quark polarization. . . .	19
2.8	Feynman diagrams with EFT. . . . .	22
3.1	Illustration of ROC curves. . . . .	28
3.2	Illustration of a decision tree. . . . .	30
3.3	Illustration of an artificial neural network (ANN). . . . .	31
4.1	Overview of the CERN accelerator complex. . . . .	36
4.2	Run 2 pileup profile. . . . .	38
4.3	Delivered luminosity versus time for data taken from 2015 to 2018 at $\sqrt{s} = 13$ TeV. . . . .	40
4.4	Slice of the CMS detector . . . . .	41
4.5	Layout of the CMS tracker before the Phase 1 pixel update in 2017. . .	43
4.6	Layout of the CMS ECAL. . . . .	44
4.7	Layout of the CMS HCAL. . . . .	45
4.8	Layout of the CMS muon system. . . . .	47
4.9	Overview over the CMS DAQ and Trigger system. . . . .	48
5.1	The basic steps of event generation in a proton-proton collision. . . . .	52
5.2	The Neural Network parton distribution functions (NNPDF3.1) at NNLO. .	53
5.3	Standard Model cross sections as a function of collider energy. . . . .	55
6.1	Illustration of the DeepCSV architecture . . . . .	68
6.2	Illustration of the DeepJet architecture . . . . .	69
6.3	Scheme of the jet energy corrections at the CMS experiment. . . . .	70
7.1	Output of Phase-2 b tagging discriminators. . . . .	75
7.2	Performance of b tagging discriminators in Phase-2. . . . .	77

7.3	Performance of b tagging discriminators in Phase-1. . . . .	78
7.4	Performance of Phase-2 b tagging discriminators in bins of $p_T$ and $ \eta $ . .	78
7.5	Efficiency maps for Phase-2 b tagging discriminators. . . . .	79
7.6	Distributions of additional variables for b tagging. . . . .	82
7.7	Performance of DeepCSV in Phase-2 with timing information. . . . .	83
8.1	Feynman diagrams for signal and background processes . . . . .	87
8.2	Relative contributions of signal and background processes to each event category. . . . .	92
8.3	Effect of pileup reweighting. . . . .	94
8.4	Effect of MET-phi modulation corrections . . . . .	96
8.5	b jet kinematic variables. . . . .	97
8.6	Lepton kinematical variables. . . . .	99
8.7	Effect of prefire reweighting. . . . .	100
8.8	Validation of the data-driven QCD multijet estimation. . . . .	103
8.9	Longitudinal neutrino momentum. . . . .	106
8.10	Comparison of top quark pseudorapidity. . . . .	106
8.11	Reconstructed top quark kinematical variables. . . . .	107
8.12	Transverse neutrino momentum. . . . .	108
8.13	Neutrino Pseudorapidity. . . . .	109
8.14	Modeling of BDT input variables. . . . .	115
8.15	Modeling of BDT input variables. . . . .	116
8.16	BDT Correlations. . . . .	117
8.17	BDT output distributions. . . . .	118
8.18	Comparison of the BDT output distributions to data. . . . .	119
8.19	Prefit BDT <sub>S</sub> distributions. . . . .	120
8.20	Postfit BDT <sub>S</sub> distributions. . . . .	128
8.21	Comparison of the measured cross section ratio with predictions from different PDF sets. . . . .	131
8.22	Prefit distributions for the top quark transverse momentum and rapidity. .	133
8.23	Prefit distributions for the polarization angles. . . . .	134
8.24	Response matrices for the polarization angles at parton level. . . . .	136
8.25	Response matrices for the polarization angles at parton level. . . . .	137
8.26	Postfit distributions . . . . .	139
8.27	Unfolding results for top quark transverse momentum and rapidity at parton level. . . . .	140
8.28	Unfolding results for the $x$ and $y$ polarization angles at parton level. . .	141
8.29	Unfolding results for the $z$ polarization angles at parton level. . . . .	142
8.30	Effect of EFT coefficients on angular distributions. . . . .	145
8.31	Limits on Wilson coefficients. . . . .	146
A.1	Example fit to the Z boson mass. . . . .	154
A.2	Muon isolation scale factors for 2018. . . . .	154
A.3	Muon trigger scale factors for 2018. . . . .	155
A.4	Reconstructed top quark kinematic variables in the 2j1m0t region. . . .	156
A.5	Modeling of BDT input variables in the $t\bar{t}$ control region. . . . .	158
A.6	Modeling of BDT input variables in the W/Z-jets control region. . . . .	159

A.7	Comparison of the BDT output distributions to data in the $t\bar{t}$ control region. . . . .	160
A.8	Comparison of the BDT output distributions to data in the W/Z-jets control region. . . . .	161
A.9	Response matrices for top quark transverse momentum and rapidity at particle level. . . . .	163
A.10	Response matrices for the polarization angles at particle level . . . . .	164
A.11	Unfolding results for top quark transverse momentum and rapidity at particle level. . . . .	165
A.12	Unfolding results for the polarization angles $x, y$ at particle level. . . . .	166
A.13	Unfolding results for the polarization angles $z$ at particle level. . . . .	167
A.14	Systematic contributions for parton level top quark transverse momentum and rapidity. . . . .	169
A.15	Systematic contributions for parton level angular distributions. . . . .	170
A.16	Systematic contributions for particle level top quark transverse momentum and rapidity. . . . .	171
A.17	Systematic contributions for parton level angular distributions. . . . .	172





# List of Tables

1.1	The fermions of the SM. . . . .	3
1.2	The bosons of the SM. . . . .	3
2.1	Spin analyzing powers . . . . .	17
2.2	Constraints on Wilson coefficients. . . . .	23
6.1	Muon identification criteria. . . . .	62
6.2	Selection criteria for tight electrons and veto electrons. . . . .	63
6.3	Criteria for the identification of AK4 jets. . . . .	67
7.1	Phase-2 data sets used for the retraining of the b taggers in CMS software version 9.3. . . . .	75
7.2	Phase-2 data sets used for the retraining of the b taggers in CMS software version 10.6. . . . .	81
8.1	High-level trigger paths used in the analysis. . . . .	90
8.2	Electron impact parameter selections. . . . .	90
8.3	Input variables used for the neutrino neural network. . . . .	111
8.4	BDT input variables. . . . .	113
8.5	Confusion matrix of the multiclassification BDT. . . . .	114
8.6	Luminosity systematic uncertainties correlations. . . . .	123
8.7	Uncertainty contributions . . . . .	132
8.8	Binning . . . . .	135
A.1	List of simulated samples used for 2016. . . . .	150
A.2	List of simulated samples used for 2017. . . . .	150
A.3	List of simulated samples used for 2018. . . . .	151
A.4	List of data damples. . . . .	152



# Bibliography

- [1] 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.* **B716** (2012) 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.
- [2] CMS Collaboration, “Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC”, *Phys. Lett.* **B716** (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.
- [3] Super-Kamiokande Collaboration, “Evidence for oscillation of atmospheric neutrinos”, *Phys. Rev. Lett.* **81** (1998) 1562–1567, doi:10.1103/PhysRevLett.81.1562, arXiv:hep-ex/9807003.
- [4] 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, doi:10.1103/PhysRevLett.87.071301, arXiv:nucl-ex/0106015.
- [5] F. Englert and R. Brout, “Broken symmetry and the mass of gauge vector mesons”, *Phys. Rev. Lett.* **13** (1964) 321–323, doi:10.1103/PhysRevLett.13.321.
- [6] P. W. Higgs, “Broken symmetries and the masses of gauge bosons”, *Phys. Rev. Lett.* **13** (1964) 508–509, doi:10.1103/PhysRevLett.13.508.
- [7] Particle Data Group Collaboration, “Review of Particle Physics”, *Phys. Rev.* **D98** (2018), no. 3, 030001, doi:10.1103/PhysRevD.98.030001. and 2019 update.
- [8] C. N. Yang and R. L. Mills, “Conservation of isotopic spin and isotopic gauge invariance”, *Phys. Rev.* **96** (Oct, 1954) 191–195, doi:10.1103/PhysRev.96.191.
- [9] D. J. Gross and F. Wilczek, “Asymptotically Free Gauge Theories - I”, *Phys. Rev.* **D8** (1973) 3633–3652, doi:10.1103/PhysRevD.8.3633.
- [10] H. D. Politzer, “Reliable Perturbative Results for Strong Interactions?”, *Phys. Rev. Lett.* **30** (1973) 1346–1349, doi:10.1103/PhysRevLett.30.1346.

- [,274(1973)].
- [11] Belle Collaboration, “Observation of a narrow charmonium - like state in exclusive  $B^\pm \rightarrow K^\pm \pi^+ \pi^- J/\psi$  decays”, *Phys. Rev. Lett.* **91** (2003) 262001, doi:10.1103/PhysRevLett.91.262001, arXiv:hep-ex/0309032.
- [12] LHCb Collaboration, “Observation of  $J/\psi p$  Resonances Consistent with Pentaquark States in  $\Lambda_b^0 \rightarrow J/\psi K^- p$  Decays”, *Phys. Rev. Lett.* **115** (2015) 072001, doi:10.1103/PhysRevLett.115.072001, arXiv:1507.03414.
- [13] M. Gell-Mann, “The interpretation of the new particles as displaced charge multiplets”, *Il Nuovo Cim.* **4** (1956), no. S2, 848–866, doi:10.1007/BF02748000.
- [14] K. Nishijima, “Charge Independence Theory of V Particles”, *Prog. Theor. Phys.* **13** (1955), no. 3, 285–304, doi:10.1143/PTP.13.285.
- [15] T. Nakano and K. Nishijima, “Charge Independence for V-particles”, *Prog. Theor. Phys.* **10** (1953) 581–582, doi:10.1143/PTP.10.581.
- [16] S. Weinberg, “A model of leptons”, *Phys. Rev. Lett.* **19** (1967) 1264–1266, doi:10.1103/PhysRevLett.19.1264.
- [17] S. L. Glashow, J. Iliopoulos, and L. Maiani, “Weak Interactions with Lepton-Hadron Symmetry”, *Phys. Rev.* **D2** (1970) 1285–1292, doi:10.1103/PhysRevD.2.1285.
- [18] S. L. Glashow, “Partial Symmetries of Weak Interactions”, *Nucl. Phys.* **22** (1961) 579–588, doi:10.1016/0029-5582(61)90469-2.
- [19] J. Goldstone, “Field theories with superconductor solutions”, *Nuovo Cim.* **19** (1961), no. 1, 154–164, doi:10.1007/BF02812722.
- [20] J. Goldstone, A. Salam, and S. Weinberg, “Broken symmetries”, *Phys. Rev.* **127** (1962) 965–970, doi:10.1103/PhysRev.127.965.
- [21] S. Weinberg, “Physical processes in a convergent theory of the weak and electromagnetic interactions”, *Phys. Rev. Lett.* **27** (1971) 1688–1691, doi:10.1103/PhysRevLett.27.1688.
- [22] S. Weinberg, “General theory of broken local symmetries”, *Phys. Rev. D* **7** (1973) 1068–1082, doi:10.1103/PhysRevD.7.1068.
- [23] N. Cabibbo, “Unitary Symmetry and Leptonic Decays”, *Phys. Rev. Lett.* **10** (1963) 531–533, doi:10.1103/PhysRevLett.10.531. [,648(1963)].

- 
- [24] M. Kobayashi and T. Maskawa, “CP Violation in the Renormalizable Theory of Weak Interaction”, *Prog. Theor. Phys.* **49** (1973) 652–657, doi:10.1143/PTP.49.652.
- [25] L. Wolfenstein, “Parametrization of the Kobayashi-Maskawa Matrix”, *Phys. Rev. Lett.* **51** (1983) 1945, doi:10.1103/PhysRevLett.51.1945.
- [26] A. Sakharov, “Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe”, *Sov. Phys. Usp.* **34** (1991), no. 5, 392–393, doi:10.1070/PU1991v034n05ABEH002497.
- [27] CDF Collaboration, “Observation of top quark production in  $\bar{p}p$  collisions”, *Phys. Rev. Lett.* **74** (1995) 2626–2631, doi:10.1103/PhysRevLett.74.2626, arXiv:hep-ex/9503002.
- [28] DØ Collaboration, “Observation of the top quark”, *Phys. Rev. Lett.* **74** (1995) 2632–2637, doi:10.1103/PhysRevLett.74.2632, arXiv:hep-ex/9503003.
- [29] CMS Collaboration, “NNLO+NNLL top-quark-pair cross sections”. url:<https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO>.
- [30] 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, doi:10.1016/j.cpc.2014.06.021, arXiv:1112.5675.
- [31] M. Botje et al., “The PDF4LHC Working Group Interim Recommendations”, arXiv:1101.0538.
- [32] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, “Uncertainties on  $\alpha(S)$  in global PDF analyses and implications for predicted hadronic cross sections”, *Eur. Phys. J.* **C64** (2009) 653–680, doi:10.1140/epjc/s10052-009-1164-2, arXiv:0905.3531.
- [33] J. Gao et al., “CT10 next-to-next-to-leading order global analysis of QCD”, *Phys. Rev.* **D89** (2014), no. 3, 033009, doi:10.1103/PhysRevD.89.033009, arXiv:1302.6246.
- [34] R. D. Ball et al., “Parton distributions with LHC data”, *Nucl. Phys.* **B867** (2013) 244–289, doi:10.1016/j.nuclphysb.2012.10.003, arXiv:1207.1303.
- [35] LHC Top Working group, “Summary of Top quark pair production cross section”. CC-BY-4.0 license, February, 2019. url:[https://twiki.cern.ch/twiki/pub/LHCPhysics/TopPairCrossectionSqrtsHistory/tt\\_curve\\_toplhcg\\_sep19.pdf](https://twiki.cern.ch/twiki/pub/LHCPhysics/TopPairCrossectionSqrtsHistory/tt_curve_toplhcg_sep19.pdf).
- [36] CMS Collaboration, “NLO single-top channel cross sections”. url:<https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopRefXsec>.

- [37] M. Aliev et al., “HATHOR: HAdronic Top and Heavy quarks crOSS section calculatoR”, *Comput. Phys. Commun.* **182** (2011) 1034–1046, doi:10.1016/j.cpc.2010.12.040, arXiv:1007.1327.
- [38] P. Kant 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) 74–89, doi:10.1016/j.cpc.2015.02.001, arXiv:1406.4403.
- [39] LHC Top Working group, “Summary of ATLAS and CMS measurements of the single top production cross-sections”. CC-BY-4.0 license. url:[https://twiki.cern.ch/twiki/pub/LHCPhysics/SingleTopAllChannelsHistory/singletop\\_allchannels\\_sep19.pdf](https://twiki.cern.ch/twiki/pub/LHCPhysics/SingleTopAllChannelsHistory/singletop_allchannels_sep19.pdf).
- [40] A. F. Falk and M. E. Peskin, “Production, decay, and polarization of excited heavy hadrons”, *Phys. Rev. D* **49** (1994) 3320–3332, doi:10.1103/PhysRevD.49.3320, arXiv:hep-ph/9308241.
- [41] J. A. Aguilar-Saavedra and S. Amor Dos Santos, “New directions for top quark polarization in the  $t$ -channel process”, *Phys. Rev.* **D89** (2014), no. 11, 114009, doi:10.1103/PhysRevD.89.114009, arXiv:1404.1585.
- [42] J. A. Aguilar-Saavedra and J. Bernabeu, “W polarisation beyond helicity fractions in top quark decays”, *Nucl. Phys.* **B840** (2010) 349–378, doi:10.1016/j.nuclphysb.2010.07.012, arXiv:1005.5382.
- [43] M. Jezabek, “Top quark physics”, *Nucl. Phys. B Proc. Suppl.* **37** (1994), no. 2, 197, doi:10.1016/0920-5632(94)90677-7, arXiv:hep-ph/9406411.
- [44] W. Bernreuther, A. Brandenburg, Z. Si, and P. Uwer, “Top quark pair production and decay at hadron colliders”, *Nucl. Phys. B* **690** (2004) 81–137, doi:10.1016/j.nuclphysb.2004.04.019, arXiv:hep-ph/0403035.
- [45] A. Brandenburg, Z. Si, and P. Uwer, “QCD corrected spin analyzing power of jets in decays of polarized top quarks”, *Phys. Lett. B* **539** (2002) 235–241, doi:10.1016/S0370-2693(02)02098-1, arXiv:hep-ph/0205023.
- [46] A. Czarnecki, M. Jezabek, and J. H. Kuhn, “Radiative corrections to  $b \rightarrow c \tau \text{ anti-}\tau\text{-neutrino}$ ”, *Phys. Lett. B* **346** (1995) 335–341, doi:10.1016/0370-2693(94)01681-2, arXiv:hep-ph/9411282.
- [47] R. Schwienhorst, C. P. Yuan, C. Mueller, and Q.-H. Cao, “Single top quark production and decay in the  $t$ -channel at next-to-leading order at the LHC”, *Phys. Rev.* **D83** (2011) 034019, doi:10.1103/PhysRevD.83.034019, arXiv:1012.5132.

- 
- [48] G. Mahlon and S. J. Parke, “Improved spin basis for angular correlation studies in single top quark production at the Tevatron”, *Phys. Rev.* **D55** (1997) 7249–7254, doi:10.1103/PhysRevD.55.7249, arXiv:hep-ph/9611367.
- [49] G. Mahlon and S. J. Parke, “Single top quark production at the LHC: Understanding spin”, *Phys. Lett.* **B476** (2000) 323–330, doi:10.1016/S0370-2693(00)00149-0, arXiv:hep-ph/9912458.
- [50] W. Bernreuther, “Top quark physics at the LHC”, *J. Phys.* **G35** (2008) 083001, doi:10.1088/0954-3899/35/8/083001, arXiv:0805.1333.
- [51] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method”, *J. High Energy Phys.* **11** (2007) 070, doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.
- [52] C. Oleari, “The POWHEG-BOX”, *Nucl. Phys. Proc. Suppl.* **205-206** (2010) 36–41, doi:10.1016/j.nuclphysbps.2010.08.016, arXiv:1007.3893.
- [53] P. Nason, “A new method for combining NLO QCD with shower monte carlo algorithms”, *J. High Energy Phys.* **2004** (nov, 2004) 040–040, doi:10.1088/1126-6708/2004/11/040.
- [54] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek, “Dimension-Six Terms in the Standard Model Lagrangian”, *J. High Energy Phys.* **10** (2010) 085, doi:10.1007/JHEP10(2010)085, arXiv:1008.4884.
- [55] D. Barducci et al., “Interpreting top-quark LHC measurements in the standard-model effective field theory”, arXiv:1802.07237.
- [56] C.-R. Chen, F. Larios, and C. P. Yuan, “General analysis of single top production and  $W$  helicity in top decay”, *Phys. Lett.* **B631** (2005) 126–132, doi:10.1016/j.physletb.2005.10.002, arXiv:hep-ph/0503040.
- [57] K. Whisnant, J.-M. Yang, B.-L. Young, and X. Zhang, “Dimension-six CP conserving operators of the third family quarks and their effects on collider observables”, *Phys. Rev.* **D56** (1997) 467–478, doi:10.1103/PhysRevD.56.467, arXiv:hep-ph/9702305.
- [58] A. Buckley et al., “Constraining top quark effective theory in the LHC Run II era”, *J. High Energy Phys.* **04** (2016) 015, doi:10.1007/JHEP04(2016)015, arXiv:1512.03360.
- [59] J. Aguilar-Saavedra, “A Minimal set of top anomalous couplings”, *Nucl. Phys. B* **812** (2009) 181–204, doi:10.1016/j.nuclphysb.2008.12.012, arXiv:0811.3842.

- [60] C. Zhang and S. Willenbrock, “Effective-Field-Theory Approach to Top-Quark Production and Decay”, *Phys. Rev. D* **83** (2011) 034006, doi:10.1103/PhysRevD.83.034006, arXiv:1008.3869.
- [61] C. Zhang, “Single Top Production at Next-to-Leading Order in the Standard Model Effective Field Theory”, *Phys. Rev. Lett.* **116** (2016), no. 16, 162002, doi:10.1103/PhysRevLett.116.162002, arXiv:1601.06163.
- [62] V. Cirigliano, W. Dekens, J. de Vries, and E. Mereghetti, “Constraining the top-Higgs sector of the Standard Model Effective Field Theory”, *Phys. Rev. D* **94** (2016), no. 3, 034031, doi:10.1103/PhysRevD.94.034031, arXiv:1605.04311.
- [63] ATLAS Collaboration, “Search for anomalous couplings in the  $Wtb$  vertex from the measurement of double differential angular decay rates of single top quarks produced in the  $t$ -channel with the ATLAS detector”, *J. High Energy Phys.* **04** (2016) 023, doi:10.1007/JHEP04(2016)023, arXiv:1510.03764.
- [64] C. E. Shannon, “A mathematical theory of communication”, *Bell Labs Tech. J.* **27** (1948), no. 3, 379–423, doi:10.1002/j.1538-7305.1948.tb01338.x.
- [65] R. Schapire, “The strength of weak learnability”, *Mach Learn* **5** (1990) 197–227, doi:10.1007/BF00116037.
- [66] Y. Freund, “Boosting a Weak Learning Algorithm by Majority”, *Inf. Comput.* **121** (1995), no. 2, 256 – 285, doi:10.1006/inco.1995.1136.
- [67] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization”, 2017. arXiv:1412.6980.
- [68] N. Srivastava et al., “Dropout: A simple way to prevent neural networks from overfitting”, *J. Mach. Learn. Res.* **15** (2014), no. 56, 1929–1958, url:<http://jmlr.org/papers/v15/srivastava14a.html>.
- [69] O. Behnke, K. Kröninger, G. Schott, and T. Schörner-Sadenius, “Data analysis in high energy physics: a practical guide to statistical methods”. Wiley-VCH, Weinheim, 2013. doi:10.1002/9783527653416.
- [70] A. Hocker and V. Kartvelishvili, “SVD approach to data unfolding”, *Nucl. Instrum. Meth. A* **372** (1996) 469–481, doi:10.1016/0168-9002(95)01478-0, arXiv:hep-ph/9509307.
- [71] A. Tikhonov, “Solution of incorrectly formulated problems and the regularization method”, *Soviet Meth. Dokl.* **4** (1963) 1035–1038.
- [72] A. Tikhonov, “On the stability of inverse problems”, *Dokl. Akad. Nauk SSSR* **39** (1943) 195–198.



- 
- [73] A. Tikhonov, “On the stability of the functional optimization problem”, *USSR Comput. Math. Math. Phys* **6** (1966) 28–33.
- [74] S. Schmitt, “TUnfold: an algorithm for correcting migration effects in high energy physics”, *J. Instrum.* **7** (2012) T10003, doi:10.1088/1748-0221/7/10/T10003, arXiv:1205.6201.
- [75] CMS Collaboration, “HiggsAnalysis-CombinedLimit”.  
url:<https://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/>.
- [76] F. James and M. Roos, “Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations”, *Comput. Phys. Commun.* **10** (1975) 343–367, doi:10.1016/0010-4655(75)90039-9.
- [77] R. J. Barlow and C. Beeston, “Fitting using finite Monte Carlo samples”, *Comput. Phys. Commun.* **77** (1993) 219–228, doi:10.1016/0010-4655(93)90005-W.
- [78] J. Conway, “Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra”, in *PHYSTAT 2011*, pp. 115–120. 2011. arXiv:1103.0354. doi:10.5170/CERN-2011-006.115.
- [79] S. Myers, “The LEP Collider, from design to approval and commissioning”. John Adams’ Lecture. CERN, Geneva, 1991. Delivered at CERN, 26 Nov 1990. url:<http://cds.cern.ch/record/226776>.
- [80] ATLAS Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider”, *J. Instrum.* **3** (2008) S08003, doi:10.1088/1748-0221/3/08/S08003.
- [81] CMS Collaboration, “The CMS experiment at the CERN LHC”, *J. Instrum.* **3** (aug, 2008) S08004–S08004, doi:10.1088/1748-0221/3/08/s08004.
- [82] ALICE Collaboration, “The ALICE experiment at the CERN LHC”, *J. Instrum.* **3** (2008) S08002, doi:10.1088/1748-0221/3/08/S08002.
- [83] LHCb Collaboration, “The LHCb Detector at the LHC”, *J. Instrum.* **3** (2008) S08005, doi:10.1088/1748-0221/3/08/S08005.
- [84] E. Mobs, “The CERN accelerator complex - 2019. Complexe des accélérateurs du CERN - 2019”, url:<https://cds.cern.ch/record/2684277>. General Photo.
- [85] R. Bailey and P. Collier, “Standard Filling Schemes for Various LHC Operation Modes”, Technical Report LHC-PROJECT-NOTE-323, CERN, Geneva, Sep, 2003. url:<https://cds.cern.ch/record/691782>.

- [86] CMS Collaboration, “Interactions per crossing (pileup) for 2015-2018”, 2019. Accessed: 27. October 2019. [url:https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults](https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults).
- [87] Lyndon Evans and Philip Bryant, “LHC Machine”, *J. Instrum.* **3** (2008), no. 08, S08001, doi:10.1088/1748-0221/3/08/s08001.
- [88] CMS Collaboration, “Cumulative delivered and recorded luminosity versus time for 2015-2018 (pp data only)”, 2019. Accessed: 27. October 2019. [url:https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults](https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults).
- [89] D. Barney, “CMS Detector Slice”, Jan, 2016. CMS Collection. Accessed: 27. October 2019. [url:http://cds.cern.ch/record/2120661](http://cds.cern.ch/record/2120661).
- [90] CMS Collaboration, “Phase 1 upgrade of the CMS pixel detector”, *J. Instrum.* **12** (2017), no. 02, C02033, doi:10.1088/1748-0221/12/02/C02033.
- [91] CMS Collaboration, “The CMS tracker system project: Technical Design Report”. Technical Design Report CMS. CERN, Geneva, 1997. [url:http://cds.cern.ch/record/368412](http://cds.cern.ch/record/368412).
- [92] CMS Collaboration, “Particle-flow reconstruction and global event description with the CMS detector”, *J. Instrum.* **12** (2017), no. 10, P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.
- [93] CMS Collaboration, “Energy resolution of the barrel of the CMS electromagnetic calorimeter”, *J. Instrum.* **2** (apr, 2007) P04004–P04004, doi:10.1088/1748-0221/2/04/p04004.
- [94] CMS Collaboration, “The CMS ECAL performance with examples”, *J. Instrum.* **9** (feb, 2014) C02008–C02008, doi:10.1088/1748-0221/9/02/c02008.
- [95] CMS Collaboration, “The CMS barrel calorimeter response to particle beams from 2-GeV/c to 350-GeV/c”, *Eur. Phys. J.* **C60** (2009) 359–373, doi:10.1140/epjc/s10052-009-0959-5. [Erratum: *Eur. Phys. J.* C61,353(2009)].
- [96] CMS Collaboration, “Performance of the CMS hadron calorimeter with cosmic ray muons and LHC beam data”, *J. Instrum.* **5** (Mar, 2010) T03012–T03012, doi:10.1088/1748-0221/5/03/t03012, arXiv:0911.4991.
- [97] CMS Collaboration, “The performance of the CMS muon detector in proton-proton collisions at  $\sqrt{s}=7$  TeV at the LHC”, *J. Instrum.* **8** (nov, 2013) P11002–P11002, doi:10.1088/1748-0221/8/11/p11002.

- 
- [98] CMS Collaboration, “Performance of CMS Muon Reconstruction in  $pp$  Collision Events at  $\sqrt{s} = 7$  TeV”, *J. Instrum.* **7** (2012) P10002, doi:10.1088/1748-0221/7/10/P10002, arXiv:1206.4071.
  - [99] K. Bos et al., “LHC computing Grid: Technical Design Report. Version 1.06 (20 Jun 2005)”. Technical Design Report LCG. CERN, Geneva, 2005. url:<https://cds.cern.ch/record/840543>.
  - [100] I. Bird et al., “Update of the Computing Models of the WLCG and the LHC Experiments”, Technical Report CERN-LHCC-2014-014. LCG-TDR-002, Apr, 2014. url:<https://cds.cern.ch/record/1695401>.
  - [101] GridKa. url:<http://www.gridka.de/>.
  - [102] Steinbuch Centre for Computing. url:<http://www.scc.kit.edu>.
  - [103] Karlsruhe Institute of Technology. url:<http://www.kit.edu>.
  - [104] Apollinari G. and Béjar Alonso I. and Brüning O. and Fessia P. and Lamont M. and Rossi L. and Taviani L., “High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1”. CERN Yellow Reports: Monographs. CERN, Geneva, 2017. doi:10.23731/CYRM-2017-004.
  - [105] CMS Collaboration, “The Phase-2 Upgrade of the CMS Level-1 Trigger”, Technical Report CERN-LHCC-2020-004. CMS-TDR-021, CERN, Geneva, Apr, 2020. url:<https://cds.cern.ch/record/2714892>, Final version.
  - [106] CMS Collaboration, “The Phase-2 Upgrade of the CMS Tracker”, Technical Report CERN-LHCC-2017-009. CMS-TDR-014, CERN, Geneva, Jun, 2017. url:<https://cds.cern.ch/record/2272264>.
  - [107] CMS Collaboration, “A MIP Timing Detector for the CMS Phase-2 Upgrade”, Technical Report CERN-LHCC-2019-003. CMS-TDR-020, CERN, Geneva, Mar, 2019. url:<https://cds.cern.ch/record/2667167>.
  - [108] CMS Collaboration, “The Phase-2 Upgrade of the CMS Barrel Calorimeters”, Technical Report CERN-LHCC-2017-011. CMS-TDR-015, CERN, Geneva, Sep, 2017. url:<https://cds.cern.ch/record/2283187>, This is the final version, approved by the LHCC.
  - [109] CMS Collaboration, “The Phase-2 Upgrade of the CMS Endcap Calorimeter”, Technical Report CERN-LHCC-2017-023. CMS-TDR-019, CERN, Geneva, Nov, 2017. url:<https://cds.cern.ch/record/2293646>.
  - [110] CMS Collaboration, “The Phase-2 Upgrade of the CMS Muon Detectors”, Technical Report CERN-LHCC-2017-012. CMS-TDR-016, CERN, Geneva, Sep, 2017. url:<https://cds.cern.ch/record/2283189>, This is the final version,

approved by the LHCC.

- [111] M. Dobbs et al., “Les Houches guidebook to Monte Carlo generators for hadron collider physics”, in *3rd Les Houches Workshop on Physics at TeV Colliders*, pp. 411–459. 3, 2004. [arXiv:hep-ph/0403045](#).
- [112] J. Alwall et al., “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”, *J. High Energy Phys.* **07** (2014) 079, [doi:10.1007/JHEP07\(2014\)079](#), [arXiv:1405.0301](#).
- [113] T. Sjöstrand et al., “An Introduction to PYTHIA 8.2”, *Comput. Phys. Commun.* **191** (2015) 159–177, [doi:10.1016/j.cpc.2015.01.024](#), [arXiv:1410.3012](#).
- [114] M. Bahr et al., “Herwig++ Physics and Manual”, *Eur. Phys. J.* **C58** (2008) 639–707, [doi:10.1140/epjc/s10052-008-0798-9](#), [arXiv:0803.0883](#).
- [115] J. Alwall et al., “A Standard format for Les Houches event files”, *Comput. Phys. Commun.* **176** (2007) 300–304, [doi:10.1016/j.cpc.2006.11.010](#), [arXiv:hep-ph/0609017](#).
- [116] A. Buckley et al., “General-purpose event generators for LHC physics”, *Phys. Rept.* **504** (2011) 145–233, [doi:10.1016/j.physrep.2011.03.005](#), [arXiv:1101.2599](#).
- [117] NNPDF Collaboration, “Parton distributions from high-precision collider data”, *Eur. Phys. J.* **C77** (2017), no. 10, 663, [doi:10.1140/epjc/s10052-017-5199-5](#), [arXiv:1706.00428](#).
- [118] R. K. Ellis, W. J. Stirling, and B. R. Webber, “QCD and collider physics”, *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.* **8** (1996) 1–435, [doi:https://doi.org/10.1017/CB09780511628788](#).
- [119] Y. L. Dokshitzer, “Calculation of the Structure Functions for Deep Inelastic Scattering and  $e^+e^-$  Annihilation by Perturbation Theory in Quantum Chromodynamics.”, *Sov. Phys. JETP* **46** (1977) 641–653, [url:http://www.jetp.ac.ru/cgi-bin/dn/e\\_046\\_04\\_0641.pdf](#). [*Zh. Eksp. Teor. Fiz.*73,1216(1977)].
- [120] V. N. Gribov and L. N. Lipatov, “Deep inelastic  $e p$  scattering in perturbation theory”, *Sov. J. Nucl. Phys.* **15** (1972) 438–450, [url:https://cds.cern.ch/record/427157](#). [*Yad. Fiz.*15,781(1972)].
- [121] G. Altarelli and G. Parisi, “Asymptotic Freedom in Parton Language”, *Nucl. Phys.* **B126** (1977) 298–318, [doi:10.1016/0550-3213\(77\)90384-4](#).

- 
- [122] NNPDF Collaboration, “Parton distributions for the LHC Run II”, *J. High Energy Phys.* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.
- [123] J. Stirling, “Standard model cross sections as a function of collider energy”, 2012. Accessed: 01-October-2020.  
url:http://www.hep.ph.ic.ac.uk/~wstirling/plots/plots.html.
- [124] C. Degrande et al., “UFO - The Universal FeynRules Output”, *Comput. Phys. Commun.* **183** (2012) 1201–1214, doi:10.1016/j.cpc.2012.01.022, arXiv:1108.2040.
- [125] V. V. Sudakov, “Vertex parts at very high-energies in quantum electrodynamics”, *Sov. Phys. JETP* **3** (1956) 65–71,  
url:http://cds.cern.ch/record/478063. [Zh. Eksp. Teor. Fiz.30,87(1956)].
- [126] S. Frixione and B. R. Webber, “Matching NLO QCD computations and parton shower simulations”, *J. High Energy Phys.* **06** (2002) 029, doi:10.1088/1126-6708/2002/06/029, arXiv:hep-ph/0204244.
- [127] S. Catani, F. Krauss, R. Kuhn, and B. R. Webber, “QCD matrix elements + parton showers”, *J. High Energy Phys.* **11** (2001) 063, doi:10.1088/1126-6708/2001/11/063, arXiv:hep-ph/0109231.
- [128] R. Frederix and S. Frixione, “Merging meets matching in MC@NLO”, *J. High Energy Phys.* **12** (2012) 061, doi:10.1007/JHEP12(2012)061, arXiv:1209.6215.
- [129] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, “Parton Fragmentation and String Dynamics”, *Phys. Rept.* **97** (1983) 31–145, doi:10.1016/0370-1573(83)90080-7.
- [130] B. Andersson, “The Lund model”, volume 7. Cambridge University Press, 7, 2005. doi:https://doi.org/10.1017/CB09780511524363, ISBN 978-0-521-01734-3, 978-0-521-42094-5, 978-0-511-88149-7.
- [131] D. Amati and G. Veneziano, “Preconfinement as a Property of Perturbative QCD”, *Phys. Lett.* **83B** (1979) 87–92, doi:10.1016/0370-2693(79)90896-7.
- [132] CMS Collaboration, “Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements”, *Eur. Phys. J. C* **80** (2020), no. 1, 4, doi:10.1140/epjc/s10052-019-7499-4, arXiv:1903.12179.
- [133] CMS Collaboration, “Event generator tunes obtained from underlying event and multiparton scattering measurements”, *Eur. Phys. J.* **C76** (2016), no. 3, 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.

- [134] GEANT4 Collaboration, “GEANT4: A Simulation toolkit”, *Nucl. Instrum. Meth.* **A506** (2003) 250–303, doi:10.1016/S0168-9002(03)01368-8.
- [135] J. Allison et al., “Geant4 developments and applications”, *IEEE Trans. Nucl. Sci.* **53** (2006) 270, doi:10.1109/TNS.2006.869826.
- [136] CMS Collaboration, “Measurement of the inelastic proton-proton cross section at  $\sqrt{s} = 13$  TeV”, *J. High Energy Phys.* **07** (2018) 161, doi:10.1007/JHEP07(2018)161, arXiv:1802.02613.
- [137] S. Banerjee, “CMS simulation software”, *J. Phys. Conf. Ser.* **396** (2012) 022003, doi:10.1088/1742-6596/396/2/022003.
- [138] CMS Collaboration, “Fast simulation of the CMS detector”, *J. Phys. Conf. Ser.* **219** (2010) 032053, doi:10.1088/1742-6596/219/3/032053.
- [139] CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *J. Instrum.* **9** (2014), no. 10, P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.
- [140] R. Fruhwirth, “Application of Kalman filtering to track and vertex fitting”, *Nucl. Instrum. Meth.* **A262** (1987) 444–450, doi:10.1016/0168-9002(87)90887-4.
- [141] P. Billoir and S. Qian, “Simultaneous pattern recognition and track fitting by the Kalman filtering method”, *Nucl. Instrum. Meth.* **A294** (1990) 219–228, doi:10.1016/0168-9002(90)91835-Y.
- [142] P. Billoir, “Progressive track recognition with a Kalman like fitting procedure”, *Comput. Phys. Commun.* **57** (1989) 390–394, doi:10.1016/0010-4655(89)90249-X.
- [143] R. Mankel, “A Concurrent track evolution algorithm for pattern recognition in the HERA-B main tracking system”, *Nucl. Instrum. Meth.* **A395** (1997) 169–184, doi:10.1016/S0168-9002(97)00705-5.
- [144] CMS Collaboration, “CMS tracking results”, July, 2020. Accessed: 01-October-2020. url:<https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsTRK>.
- [145] CMS Collaboration, “Tracking and Primary Vertex Results in First 7 TeV Collisions”, Technical Report CMS-PAS-TRK-10-005, CERN, Geneva, 2010. url:<http://cds.cern.ch/record/1279383>.
- [146] K. Rose, “Deterministic annealing for clustering, compression, classification, regression, and related optimization problems”, *Proceedings of the IEEE* **86** (Nov, 1998) 2210–2239, doi:10.1109/5.726788.

- 
- [147] CMS Collaboration, “Adaptive vertex fitting”, *J. Phys.* **G34** (2007) N343, doi:10.1088/0954-3899/34/12/N01.
- [148] CMS Collaboration, “Baseline muon selections for Run-II”.  
url:<https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonIdRun2>.
- [149] CMS Collaboration, “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at  $\sqrt{s} = 13$  TeV”, *J. Instrum.* **13** (2018), no. 06, P06015, doi:10.1088/1748-0221/13/06/P06015, arXiv:1804.04528.
- [150] CMS Collaboration, “Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC”, *J. Phys. G* **31** (Jul, 2005) N9–N20, doi:10.1088/0954-3899/31/9/n01.
- [151] CMS Collaboration, “Cut Based Electron ID for Run 2”. url:<https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedElectronIdentificationRun2>.
- [152] CMS Collaboration, “Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at  $\sqrt{s} = 8$  TeV”, *J. Instrum.* **10** (2015), no. 06, P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.
- [153] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- $k_t$  jet clustering algorithm”, *J. High Energy Phys.* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.
- [154] CMS Collaboration, “Jet algorithms performance in 13 TeV data”, Technical Report CMS-PAS-JME-16-003, CERN, Geneva, 2017.  
url:<https://cds.cern.ch/record/2256875>.
- [155] CMS Collaboration, “Jet Identification for the 13 TeV data Run2016”.  
url:<https://twiki.cern.ch/twiki/bin/view/CMS/JetID13TeVRun2016>.
- [156] CMS Collaboration, “Jet Identification for the 13 TeV data Run2017”.  
url:<https://twiki.cern.ch/twiki/bin/view/CMS/JetID13TeVRun2017>.
- [157] CMS Collaboration, “Jet Identification for the 13 TeV data Run2018”.  
url:<https://twiki.cern.ch/twiki/bin/view/CMS/JetID13TeVRun2018>.
- [158] E. Bols et al., “Jet Flavour Classification Using DeepJet”, arXiv:2008.10519.
- [159] M. Verzetti, “Machine learning applications to jet tagging in CMS”, in *Machine learning for phenomenology*. 2018.  
<https://conference.ippp.dur.ac.uk/event/660>.

- [160] CMS Collaboration, “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV”, *J. Instrum.* **12** (2017), no. 02, P02014, doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.
- [161] CMS Collaboration, “Jet Energy Calibrations at CMS experiment with 13 TeV collisions”, *PoS EPS-HEP2017* (2017) 805, doi:10.22323/1.314.0805.
- [162] CMS Collaboration, “Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS”, *J. Instrum.* **6** (2011) P11002, doi:10.1088/1748-0221/6/11/P11002, arXiv:1107.4277.
- [163] D. Bertolini, P. Harris, M. Low, and N. Tran, “Pileup Per Particle Identification”, *J. High Energy Phys.* **10** (2014) 059, doi:10.1007/JHEP10(2014)059, arXiv:1407.6013.
- [164] CMS Collaboration, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”, *JINST* **13** (2018), no. 05, P05011, doi:10.1088/1748-0221/13/05/P05011, arXiv:1712.07158.
- [165] L. Torrey and J. Shavlik, “Transfer Learning”. Handbook of Research on Machine Learning Applications. IGI Global. (2009). url:<https://ftp.cs.wisc.edu/machine-learning/shavlik-group/torrey.handbook09.pdf>.
- [166] 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) 330–350, doi:10.1016/j.physletb.2012.09.031, arXiv:1205.3130.
- [167] 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), no. 11, 112006, doi:10.1103/PhysRevD.90.112006, arXiv:1406.7844.
- [168] 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, doi:10.1103/PhysRevLett.107.091802, arXiv:1106.3052.
- [169] CMS Collaboration, “Measurement of the Single-Top-Quark  $t$ -Channel Cross Section in  $pp$  Collisions at  $\sqrt{s} = 7$  TeV”, *J. High Energy Phys.* **12** (2012) 035, doi:10.1007/JHEP12(2012)035, arXiv:1209.4533.
- [170] 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), no. 8, 531, doi:10.1140/epjc/s10052-017-5061-9, arXiv:1702.02859.



- 
- [171] 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”, *J. High Energy Phys.* **06** (2014) 090, doi:10.1007/JHEP06(2014)090, arXiv:1403.7366.
- [172] CMS Collaboration, “Measurement of Top Quark Polarisation in T-Channel Single Top Quark Production”, *J. High Energy Phys.* **04** (2016) 073, doi:10.1007/JHEP04(2016)073, arXiv:1511.02138.
- [173] 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”, *J. High Energy Phys.* **04** (2017) 086, doi:10.1007/JHEP04(2017)086, arXiv:1609.03920.
- [174] 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, doi:10.1016/j.physletb.2019.135042, arXiv:1812.10514.
- [175] CMS Collaboration, “Measurement of differential cross sections and charge ratios for  $t$ -channel single top quark production in proton-proton collisions at  $\sqrt{s}=13$  TeV”, *Eur. Phys. J. C* **80** (2020), no. 5, 370, doi:10.1140/epjc/s10052-020-7858-1, arXiv:1907.08330.
- [176] CMS Collaboration, “Utilities for Accessing Pileup Information for Data”. url:https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupJSONFileforData.
- [177] CMS Collaboration, “Recommendation for Using b-tag Objects in Physics Analyses”. url:https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation.
- [178] CMS Collaboration, “Reference muon id, isolation and trigger efficiencies for Run 2 legacy data.”. url:https://twiki.cern.ch/twiki/bin/view/CMS/MuonReferenceEffsRun2Legacy.
- [179] D. Müller, “Search for single top quark production in the s channel at 13 TeV with the CMS experiment”. PhD thesis, Karlsruhe Institute of Technology (KIT), 2020. url:https://publish.etp.kit.edu/record/21968.
- [180] CMS Collaboration, “Egamma Recommendations for Run 2”. url:https://twiki.cern.ch/twiki/bin/view/CMS/EgammaRunIIRecommendations.
- [181] CMS Collaboration, “Reweighting recipe to emulate Level 1 ECAL prefiring.”. url:https://twiki.cern.ch/twiki/bin/viewauth/CMS/L1ECALPrefiringWeightRecipe.

- [182] T. Chwalek, “Messung der  $W$  -Boson-Helizitätsanteile in Top-Quark-Zerfällen mit dem CDF II Experiment und Studien zu einer frühen Messung des  $t\bar{t}$  -Wirkungsquerschnitts mit dem CMS Experiment”. PhD thesis, Karlsruhe U., 2010. doi:10.2172/1369264.
- [183] M. Abadi et al., “TensorFlow: Large-scale machine learning on heterogeneous systems”, 2015. Software available from tensorflow.org. url:<http://tensorflow.org/>.
- [184] J. Therhaag, “TMVA Toolkit for multivariate data analysis in ROOT”, *PoS ICHEP2010* (2010) 510, doi:10.22323/1.120.0510.
- [185] CMS Collaboration, “CMS Luminosity Measurements for the 2016 Data Taking Period”, Technical Report CMS-PAS-LUM-17-001, CERN, Geneva, 2017. url:<https://cds.cern.ch/record/2257069>.
- [186] CMS Collaboration, “CMS luminosity measurement for the 2017 data-taking period at  $\sqrt{s} = 13$  TeV”, Technical Report CMS-PAS-LUM-17-004, CERN, Geneva, 2018. url:<https://cds.cern.ch/record/2621960>.
- [187] CMS Collaboration, “CMS luminosity measurement for the 2018 data-taking period at  $\sqrt{s} = 13$  TeV”, Technical Report CMS-PAS-LUM-18-002, CERN, Geneva, 2019. url:<https://cds.cern.ch/record/2676164>.
- [188] CMS Collaboration, “Luminosity Physics Object Group (Lumi POG)”. url:<https://twiki.cern.ch/twiki/bin/view/CMS/TWikiLUM>.
- [189] CMS Collaboration, “Measurement of the production cross section for single top quarks in association with W bosons in proton-proton collisions at  $\sqrt{s} = 13$  TeV”, *J. High Energy Phys.* **10** (2018) 117, doi:10.1007/JHEP10(2018)117, arXiv:1805.07399.
- [190] CMS Collaboration, “Measurement of differential cross sections for top quark pair production using the lepton+jets final state in proton-proton collisions at 13 TeV”, *Phys. Rev. D* **95** (2017) 092001, doi:10.1103/PhysRevD.95.092001, arXiv:1610.04191.
- [191] CMS Collaboration, “Measurement of normalized differential  $t\bar{t}$  cross sections in the dilepton channel from pp collisions at  $\sqrt{s} = 13$  TeV”, *J. High Energy Phys.* **04** (2018) 060, doi:10.1007/JHEP04(2018)060, arXiv:1708.07638.
- [192] J. Butterworth et al., “PDF4LHC recommendations for LHC Run II”, *J. Phys. G* **43** (2016) 023001, doi:10.1088/0954-3899/43/2/023001, arXiv:1510.03865.
- [193] CMS Collaboration, “Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of  $t\bar{t}$  at  $\sqrt{s} = 8$  and 13 TeV”, CMS Physics

- Analysis Summary CMS-PAS-TOP-16-021, 2016.  
`url:http://cds.cern.ch/record/2235192`.
- [194] S. Alekhin, J. Blümlein, S. Moch, and R. Placakyte, “Parton distribution functions,  $\alpha_s$ , and heavy-quark masses for LHC Run II”, *Phys. Rev. D* **96** (2017), no. 1, 014011, doi:10.1103/PhysRevD.96.014011, arXiv:1701.05838.
- [195] S. Alekhin, J. Blümlein, and S. Moch, “NLO PDFs from the ABMP16 fit”, *Eur. Phys. J. C* **78** (2018), no. 6, 477, doi:10.1140/epjc/s10052-018-5947-1, arXiv:1803.07537.
- [196] A. Buckley et al., “LHAPDF6: parton density access in the LHC precision era”, *Eur. Phys. J. C* **75** (2015) 132, doi:10.1140/epjc/s10052-015-3318-8, arXiv:1412.7420.
- [197] S. Dulat et al., “New parton distribution functions from a global analysis of quantum chromodynamics”, *Phys. Rev. D* **93** (2016), no. 3, 033006, doi:10.1103/PhysRevD.93.033006, arXiv:1506.07443.
- [198] 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), no. 5, 204, doi:10.1140/epjc/s10052-015-3397-6, arXiv:1412.3989.
- [199] H1, ZEUS Collaboration, “Combined Measurement and QCD Analysis of the Inclusive  $e^+p$  Scattering Cross Sections at HERA”, *J. High Energy Phys.* **01** (2010) 109, doi:10.1007/JHEP01(2010)109, arXiv:0911.0884.
- [200] P. Virtanen et al., “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”, *Nat. Methods* **17** (2020) 261–272, doi:10.1038/s41592-019-0686-2.



# Danksagung

Ich bedanke mich bei Prof. Dr. Thomas Müller für die Aufnahme in seine Arbeitsgruppe nach meinem Studium sowie für die Ermöglichung meiner Promotion. Weiterhin danke ich ihm für seine Unterstützung bei meiner Teilnahme an verschiedenen Workshops, Konferenzen und Schulen. Besonders dankbar bin ich jedoch für die Möglichkeit ein halbes Jahr am CERN in Genf verbracht haben zu können, um mich dort mit anderen Wissenschaftlern auszutauschen und wertvolle Erfahrungen zu sammeln.

Ich danke auch Prof. Dr. Quast für die Übernahme des Korreferats sowohl für meine Promotion als auch für meine Masterarbeit.

Fürderhin bedanke ich mich bei Dr. Thorsten Chwalek und Dr. Nils Faltermann, die mich seit meinem Eintritt in die Arbeitsgruppe betreut haben und stets mit Rat und Tat zur Seite standen.

I'm thanking PhD. Soureek Mitra for his support and supervision in BTV tasks and for proof reading my thesis.

Mein Dank geht zudem an Marco Link, Max Neukum, Komal Tauqueer, Ansar Iqbal und Dr. Nils Faltermann für das sorgfältige Korrekturlesen meiner Doktorarbeit. Dankbar bin ich auch Johann Rauser, Sebastian Wozniewski, Michael Wassmer, Stefan Wunsch und Matthias Schnepf für die gemeinsame Prüfungsvorbereitung. Dr. Denise Müller und Johann Rauser danke ich für eine schöne Zeit am CERN.

Ich bedanke mich außerdem bei allen aktuellen und ehemaligen Mitgliedern der Arbeitsgruppe sowie allen Mitgliedern des Instituts für Experimentelle Teilchenphysik für eine angenehme Atmosphäre. Den Admins danke ich für die Instandhaltung unserer Computinginfrastruktur und für die gute Zusammenarbeit während meiner eigenen Admintätigkeit.

Zum Schluss bedanke ich mich bei meiner Familie und meinen Freunden für ihre Unterstützung in allen Lebenslagen, vor allem während des Studiums.

