ETP-KA/2020-24

Karlsruhe Institute of Technology

Measurement of the Energy Asymmetry in Top Quark Pair plus Jet Production 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. Johann Vincent Rauser

aus Tübingen

Tag der mündlichen Prüfung:	27. November 2020
Referent:	Prof. Dr. Thomas Müller
Korreferent:	Prof. Dr. Günter Quast

Erklärung der selbstständigen Anfertigung der Dissertationsschrift

Hiermit erkläre ich, dass ich die Dissertation mit dem Titel

Measurement of the Energy Asymmetry in Top Quark Pair plus Jet Production 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 11. November 2020

. . . . . .

(**Johann Vincent Rauser**)

Abstract

The measurement of the energy asymmetry in top quark pair production in association with one additional high-p_{T} jet is presented, using 137.1 fb^{-1} of data recorded with the CMS detector at the CERN LHC in proton-proton collisions at a center-of-mass energy of 13 TeV. The presented measurement focuses on the semileptonic decay process of the top quark pair and requires a boosted event topology of the $\mathrm{t\bar{t}}j$ system. Different types of clustered final-state particles are considered for the reconstruction, having the minimal common requirement of one charged electron or muon, missing transverse momentum due to the corresponding neutrino, and one hard jet in the central region of the detector. The primary focus is to reconstruct events in the boosted regime using top tagged fat jets. If this is not possible, an attempt is taken to reconstruct events in the resolved regime with slim jets under the employment of boosted decision trees. In order to unfold the results, a reconstruction of simulated signal process events on particle level is performed either in the boosted regime, the resolved regime, or using parton information of the event. The unfolding is performed with a maximum likelihood fit, splitting the signal process into different subcategories according to the event kinematic properties on particle level. Both signal and background processes are obtained fully from simulation and the systematic uncertainties are considered as nuisance parameters in the fit.

This analysis is the first measurement of the energy asymmetry and yields an observed value of

$$A_{E,\,\mathrm{unf.}}^{\mathrm{opt}} = -3.0\,\% \,{}^{+4.0\,\%}_{-5.5\,\%}\,(\mathrm{stat} + \mathrm{syst})$$

in a fiducial phase space. This result is in agreement with the corresponding SM expectation of $A_E^{\mathrm{opt}} = -1.59\,\% \pm 1.00\,\%\,(\mathrm{stat}) \pm 0.37\,\%\,(\mathrm{syst})$.

Introduction

Curiosity is one of the most striking forces driving mankind to its modern day development. The desire of gathering knowledge and understanding the principles of nature manifests itself today in different categories of science, with physics being one of the most fundamental. Demystifying the structure and interaction of the smallest particles in the Universe and linking them to the processes at biggest scale, is what particle physicists are aiming for since more than 100 years. The Standard Model of Particle Physics (SM) is the intermediate result of this ongoing process and a story of success, with the latest milestone being the discovery of the Higgs boson at the Large Hadron Collider (LHC) in 2012 [1, 2].

As a variety of observed phenomena, including for example the established existence of dark matter or the baryon-anitbaryon asymmetry in the Universe, are not explained by the SM, further extensions of this model are required. Different approaches are considered for the search of such Beyond Standard Model (BSM) physics in collision experiments, which could show up directly in the production of new particles or indirectly in the precision measurement of known particle properties. The top quark, as the heaviest elementary particle of the SM, is a promising candidate for sensitivity to BSM effects, especially when searching for deviations in its predicted properties. Asymmetries in the production of top quark pairs are hereby of special interest with a significant discovery potential and have been studied already at the Tevatron collider at a center-of-mass energy of 1.96 TeV and at the LHC with center-of-mass energies of 7 TeV and 8 TeV [3, 4]. The energy asymmetry in the production of a top quark pair in association with one high-p_{T} jet ($\mathrm{t\bar{t}}j$) promises to be a sensitive observable to be measured at the LHC with center-of-mass energy of 13 TeV.

The measurement of the energy asymmetry using data collected with the Compact Muon Solenoid (CMS) detector from 2016 to 2018 at a center-of-mass energy of 13 TeV will be presented in this thesis, which is structured as follows:

- An overview of the SM and an introduction to top quark physics at hadron colliders is given in the first chapter.
- The second chapter describes the experimental setup at the LHC and the CMS detector.
- The employed methods for the statistical analysis of the measurement are explained in the third chapter.

- The simulation of events and the reconstruction of physics objects from the raw detector signals is outlined in chapter four.
- The measurement of the energy asymmetry, including the reconstruction of the $t\bar{t}j$ system and the unfolding of the results to particle level, is given in the fifth chapter.
- The thesis concludes in chapter six with a summary and an outlook to future measurements of asymmetries in top quark pair production.

Contents

1. Theoretical Motivation

The level of complexity and detail in modern day science experiments makes it unfeasible to perform any kind of measurement without a profound understanding of the established theoretical knowledge and experimental status in the field. In particle physics these prerequisites are aggregated in the Standard Model of Particle Physics (SM), which will be shortly introduced in this chapter.

Following up on the description of particles and interactions of the SM in Section 1.1, a more detailed introduction to top quark physics will be given in Section 1.2. This will not only cover the most important properties and production modes of the top quark but also give an insight into the asymmetries in top quark pair production, as these are of special interest for the presented analysis.

For reasons of convenience, natural units with $c = \hbar = 1$ are used in this thesis.

1.1 The Standard Model of Particle Physics

The Standard Model of Particle Physics aims at describing matter itself and all its interactions in the Universe. It must not be understood purely as an empiricism, but as a model that developed as an interplay between theoretical prediction and experimental discovery of the smallest particles known by humanity. In order to include the explanation of phenomena that are not yet described by the SM, as for example gravitation, dark matter and the baryon asymmetry in the Universe [5], physicists are developing and testing models beyond the standard model (BSM). Despite the known weaknesses of the SM, its success is tremendous, peaking in the discoveries of the top quark in 1995 [6, 7] and the Higgs boson in 2012 [1, 2].

1.1.1 Particles

The most fundamental division of particles in the SM is the one between fermions and bosons. Whilst fermions account for all matter in the Universe, bosons are mediating the interactions between particles.

Table 1.1: The fermions of the SM. The table gives an overview of all fermions of the SM and the generation they belong to. Quarks and leptons respectively can be further separated according to the third component of the weak isospin. The values are taken from [9].

Fermions	Generation			Electric	3rd Comp.
	1	2	3	Charge (e)	Weak Isospin
Quarks	up (u)	charm (c)	top (t)	$+\frac{2}{3}$	$+\frac{1}{2}$
	down (d)	strange (s)	bottom (b)	$-\frac{1}{3}$	$-\frac{1}{2}$
Leptons	ν_e	ν_μ	ν_τ	0	$+\frac{1}{2}$
	e	μ	τ	-1	$-\frac{1}{2}$

Fermions

Fermions are spin-$\frac{1}{2}$ particles and obey Fermi-Dirac statistics. The SM contains in total twelve fermions (see Table 1.1), which can be categorized according to their specific properties.

Fermions with non-integer electric charge are known as quarks and exist in six different so-called flavors within three generations. Each generation contains one up-type quark with electric charge of $+\frac{2}{3}$ (up, charm and top) and one down-type quark with electric charge of $-\frac{1}{3}$ (down, strange and bottom), with the generations being ordered by ascending quark mass. Quarks do not only have the unique property of interacting with all fundamental forces, but also the possibility to change their flavor through the emission of a W boson. This process alters the electric charge of a quark by one unit of the elementary charge e and hence converts an up-type quark to a down-type quark and vice versa. Flavor-changing neutral currents, which change the flavor of a quark while preserving its electric charge are forbidden in the SM at first order perturbation theory (tree level). The probability that a quark i changes its flavor to a quark j is given by the square of the absolute value of the *Cabibbo-Kobayashi-Maskawa* (CKM) matrix element V_{ij} [8]:

$$V_{\mathrm{CKM}} = \begin{pmatrix} V_{\mathrm{ud}} & V_{\mathrm{us}} & V_{\mathrm{ub}} \\ V_{\mathrm{cd}} & V_{\mathrm{cs}} & V_{\mathrm{cb}} \\ V_{\mathrm{td}} & V_{\mathrm{ts}} & V_{\mathrm{tb}} \end{pmatrix}. \tag{1.1}$$

As the diagonal elements of this matrix are close to one, decays within one generation are preferred.

The remaining six fermions have integer-valued electric charge and are called leptons. As it was the case for quarks, leptons can be grouped into three generations as well, each containing one charged lepton (electron e, muon μ and tau τ) and the corresponding uncharged lepton-neutrino (ν_e, ν_μ and ν_τ). The ordering of the lepton generations is according to the mass of the charged lepton with the lightest being the electron in the first generation.

Table 1.2: Fundamental forces and the corresponding gauge bosons of the SM. The table gives an overview of the three fundamental forces and the corresponding gauge bosons of the SM. J^P refers to the spin J of the particles together with the *eigenvalue* P of the parity operator. The *electromagnetic* (EM) and the *weak* force can be unified in the *electroweak* force. The numerical values are taken from the particle data group (PDG) [9].

Force	Type of Charge	Gauge boson	Mass (GeV)	J^P	Electric Charge (e)
EM	electric charge	photon (γ)	0	1^-	0
weak	weak charge	$W^\pm$	80.385	1	$\pm$ 1
		Z^0	91.188		0
strong	colour charge	gluons (g)	0	1^-	0

As the charged leptons can only decay under the influence of the weak interaction into charged leptons of lower generations, the electron can not decay any further and is therefore stable.

Bosons

Bosons have spin 0 or 1 and obey Bose-Einstein statistics. Except for gravitation, which is not described by the SM yet, all forces in the SM are mediated by gauge bosons. An overview of these fundamental forces and the corresponding spin-1 bosons including their most prominent properties is given in Table 1.2. In addition to these vector bosons, the SM contains one scalar spin-0 boson, which is called Higgs boson. It is an excitation of the Higgs field and does not carry any force but is responsible for giving mass to the SM particles.

1.1.2 Interactions

The SM is a quantum field theory and can be described by a local SU(3) $\times$ SU(2) $\times$ U(1) gauge symmetry group [10]. In the following the respective gauge groups and the resulting fundamental forces of the SM will be discussed.

Strong Interaction

The strong interaction corresponds to the field theory of quantum chromodynamics and is described by the SU(3) gauge group [11]. The strong force acts only on particles that carry a color charge, where this color charge can have the values *red*, *green*, and *blue* with the corresponding anticolors *antired*, *antigreen*, and *antiblue*. It is mediated by gluons, which are the massless gauge bosons of the strong interaction and carry a color and anticolor charge on their own. The three values of color charge can create a neutral state by either the combination of color and corresponding anticolor or by the combination of all three (anti-)color charges. Particles carrying a color charge always form color-neutral objects and there are exactly eight color-carrying gluons in the SM. On short scales color-charged particles can

be considered to be quasi-free (asymptotic freedom), whereas the confinement on larger scales makes it energetically more favorable to create a new quark-antiquark pair when increasing the distance between two strongly bound particles.

Electroweak Interaction and Symmetry Breaking

The remaining SU(2) and U(1) gauge groups represent the weak and the electromagnetic interaction and are combined to the electroweak theory [11–13]. The weak isospin T and the weak hypercharge Y are the generators of the two symmetry groups and contain the massless gauge bosons. These are the three W bosons W_1, W_2 and W_3 for SU(2) and the B boson for U(1). The electric charge Q is given by the *Gell-Mann-Nishijiama* relation and correlated to Y and the third component of the weak isospin T_3 via:

$$Q = \frac{Y}{2} + T_3 \,. \tag{1.2}$$

Due to the Higgs mechanism the electroweak symmetry is spontaneously broken and the massless gauge bosons coalesce to three massive and one massless gauge boson [14, 15]. These are on the one hand the charged, massive W bosons as a mixture of W_1 and W_2:

$$\mathrm{W}^{\pm} = \frac{1}{\sqrt{2}}(W_1 \mp W_2) \,. \tag{1.3}$$

On the other hand W_3 and B form the photon (γ) and the Z^0 (Z) via rotation with the weak mixing angle θ_W:

$$\begin{pmatrix} \gamma \\ \mathrm{Z} \end{pmatrix} = \begin{pmatrix} \cos\theta_{\mathrm{W}} & \sin\theta_{\mathrm{W}} \\ -\sin\theta_{\mathrm{W}} & \cos\theta_{\mathrm{W}} \end{pmatrix} \begin{pmatrix} B \\ W_3 \end{pmatrix} \,. \tag{1.4}$$

The weak mixing angle is also denoted as the *Weinberg* angle and is defined via the coupling constants of the electromagnetic interaction (g') and the weak interaction (g):

$$\cos\theta_W = \frac{g}{\sqrt{g^2 + g'^2}} \,, \tag{1.5}$$

and also connects the masses of the Z^0 and the $W^{\pm}$ particles via:

$$M_{\mathrm{Z}} = \frac{M_{\mathrm{W}}}{\cos\theta_{\mathrm{W}}} \,. \tag{1.6}$$

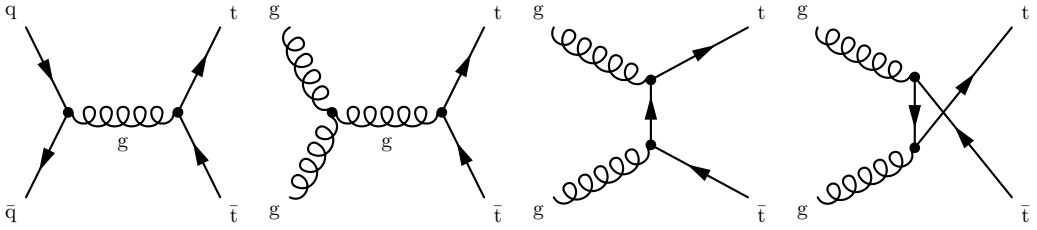

Figure 1.1: Top quark pair production: Shown are the production modes for top quark pairs for two different initial states in leading order. While the first Feynman diagram shows the production via quark-antiquark annihilation, the latter three show the gluon-gluon fusion process via (from left to right) the s, the t and the u channel.

1.2 Top Quark Physics

The top quark was discovered at the Tevatron in 1995 [6, 7] being the last quark to be detected. It is the heaviest particle in the SM with a mass of 173.34 ± 0.76 GeV and the only quark that does not form bound states due to its short lifetime of roughly 5×10^{-25} s [16]. This short lifetime is a result of the high mass of the top quark and the resulting possibility to decay into a real W boson.

The properties of the top quark are of special interest for the presented analysis. These will be discussed in the following together with its production modes and the asymmetries in top quark pair production.

1.2.1 Top Quark Production

There are two important production channels of top quarks with relevance for this thesis, which will be described in the upcoming sections.

Top Quark Pair Production

Depending on the colliding particles, top quark pairs ($t\bar{t}$) are mainly produced in two different processes of the strong interaction at hadron colliders as shown in Figure 1.1. At the Tevatron protons and antiprotons were colliding and the production via quark-antiquark annihilation was the dominant process. The LHC on the other hand is a proton-proton collider, meaning the absence of antiquarks as valence quarks and thus requiring a sea quark as the initial-state antiquark. As a result the gluon-gluon fusion process is the dominant production mode at the LHC. Utilizing the Mandelstam variables [17], this production mode can be split up further into s, t and u channel. The total predicted cross section for $t\bar{t}$ production at next-to-next-to-leading order at the LHC is

$$\sigma_{t\bar{t}} = 831.76^{+19.77}_{-29.20}\,(\text{scale}) \pm 35.06\,(\text{PDF}+\alpha_s)^{+23.18}_{-22.45}\,(\text{mass})\,\text{pb}\,, \tag{1.7}$$

for a center-of-mass energy of 13 TeV and an assumed top quark mass of 172.5 GeV [18, 19].

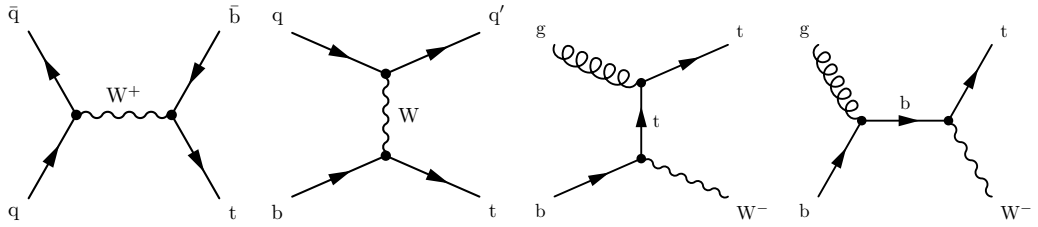

Figure 1.2: Single top quark production: The figures show the four production processes in leading order for a single top quark at the LHC from left two right: *s*-channel production, *t*-channel production and two times tW-channel production.

Single Top Quark Production

About 73% of all top quarks at the LHC are produced as pairs of quark and antiquark and the remaining 27% are produced as single top quarks. The different processes for single top quark production are *s*, *t*, and tW-channel production and shown in the Feynman diagrams in Figure 1.2. The cross sections are calculated at next-to-leading order as:

$$\sigma_{s\text{-ch.}} = 10.32^{+0.29}_{-0.24}\,(\text{scale}) \pm 0.27\,(\text{PDF}+\alpha_\text{s})\,\text{pb}\,, \tag{1.8}$$

$$\sigma_{t\text{-ch.}} = 216.99^{+6.62}_{-4.64}\,(\text{scale}) \pm 6.16\,(\text{PDF}+\alpha_\text{s})\,\text{pb}\,, \tag{1.9}$$

$$\sigma_{\text{tW}} = 71.7 \pm 1.80\,(\text{scale}) \pm 3.40\,(\text{PDF}+\alpha_\text{s})\,\text{pb} \tag{1.10}$$

for a center-of-mass energy of 13 TeV and a top quark mass of 172.5 GeV [20]. The cross section of single top quark production is smaller compared to that of top quark pair production due to the fact that single top quarks at the LHC can only be produced via the electroweak interaction at leading and next-to-leading order.

Top Quark Decays

The lifetime of the top quark is too short to form bound states and it decays via the electroweak interaction. As the matrix element V_{tb} of the CKM matrix is very close to 1, the decay into a real W boson and a bottom quark is favored, while the decay into the remaining two down-type quarks (strange and down) is strongly suppressed. Top quark decays can be further categorized by the subsequent decay chain, mainly by the decay of the W boson. With a probability of about 32% a W boson decays leptonically into a charged lepton and the corresponding neutrino, while the decay into a quark-antiquark pair has a probability of 68% [9] (see Figure 1.3). These branching ratios can be extended to characterize the decay of a top quark pair. The fully-hadronic final state (both W bosons decay hadronically) occurs in 45.7%, the semileptonic final state (exactly one W boson decays leptonically) in 43.8% and the dileptonic final state (both W bosons decay leptonically) in 10.5% of the top quark pair decays.

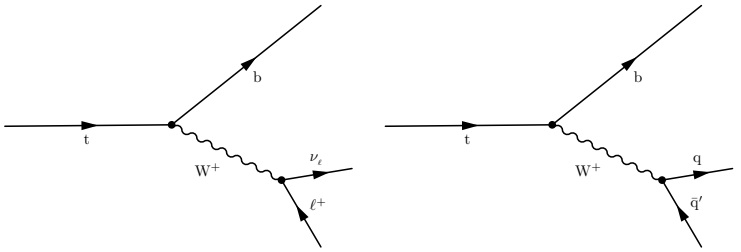

Figure 1.3: Top quark decay: The figures show the decay of a top quark and differ by the subsequent decay of the W boson, which is either leptonic (left) or hadronic (right).

1.2.2 Asymmetries in Top Quark Pair Production

Asymmetries in top quark pair production are generally defined as

$$A_x = \frac{N^+ - N^-}{N^+ + N^-}\,, \tag{1.11}$$

where N^+ and N^- are the numbers of events with positive and negative signs of the sensitive variable x, respectively.

The production of $t\bar{t}$ at leading order is a symmetric process in the SM [21, 22]. Asymmetries in $t\bar{t}$ production are induced by interference terms in next-to-leading order (NLO) processes, which occur on the one side between Born and box diagrams and on the other side between initial-state radiation (ISR) and final-state radiation (FSR) diagrams.

In the following, already established measurements in the production of top quark pairs will be presented. These are mainly the forward-backward asymmetry at the Tevatron and the charge asymmetry at the LHC. The chapter will conclude with the discussion of the energy asymmetry in $t\bar{t}$ + jet production as a new observable for hadron colliders.

Forward-Backward Asymmetry

The sensitive variable for the forward-backward asymmetry at the Tevatron is the rapidity difference between t and $\bar{t}$:

$$\Delta y_{t\bar{t}} = y_t - y_{\bar{t}}\,. \tag{1.12}$$

The rapidity y of a particle is defined under employment of its energy E and its z-momentum component p_z, with z usually being chosen to be parallel with the beam axis:

$$y := \frac{1}{2}\ln\left(\frac{E + p_z}{E - p_z}\right)\,. \tag{1.13}$$

The main production channel of top quark pairs at the Tevatron was quark-antiquark annihilation as both quark and antiquark were present as valence quarks in the colliding protons and antiprotons respectively. As the incoming direction of proton and antiproton are always the same, for a vast majority of the colliding events the momentum direction of quark and antiquark is well-known and allows for the definition of a forward and a backward direction, giving rise to the name forward-backward asymmetry. The definition at the Tevatron is such that a positive asymmetry corresponds to the forward direction pointing in the momentum direction of the incoming proton, resulting in a higher number of top quarks than top antiquarks in the forward direction (see left part of Figure 1.4).

The latest combination from the CDF and DØ Collaborations [3] gives a result for the inclusive forward-backward asymmetry of

$$A_{\mathrm{FB}}^{\mathrm{t\bar{t}}} = 0.128 \pm 0.025\,, \tag{1.14}$$

which is consistent with the SM prediction [23] of

$$A_{\mathrm{FB}}^{\mathrm{t\bar{t}}} = 0.097 \pm 0.007\,. \tag{1.15}$$

Charge Asymmetry

As the LHC is a proton-proton collider, antiquarks can only occur as sea quarks and the collision setup is symmetric in a way that no strict forward direction can be defined. As a result the rapidity distributions of both top quark and top antiquark are symmetric as can be seen in the right part of Figure 1.4. The different momentum fractions of valence and sea quarks however result in a difference of the absolute rapidities of top quark and antiquark. For the charge asymmetry A_{C} a sensitive variable can hence be defined as the difference between these absolute values:

$$\Delta|y_{\mathrm{t\bar{t}}}| := |y_{\mathrm{t}}| - |y_{\bar{\mathrm{t}}}|\,. \tag{1.16}$$

A_{C} is dependent of the center-of-mass energy of the collision [24] as gluon-gluon fusion becomes the more dominant $\mathrm{t\bar{t}}$ production process at higher energies. For center-of-mass energies of 7 and 8 TeV a combined measurement of the CMS and ATLAS Collaborations has been performed [4] with results that are in agreement with the SM predictions:

$$A_{\mathrm{C}}^{\mathrm{LHC7}}(\mathrm{ATLAS + CMS}) = 0.005 \pm 0.007(\mathrm{stat}) \pm 0.006(\mathrm{syst}) \tag{1.17}$$

$$A_{\mathrm{C}}^{\mathrm{LHC7}}(\mathrm{SM\ prediction}) = 0.0123 \pm 0.0005 \tag{1.18}$$

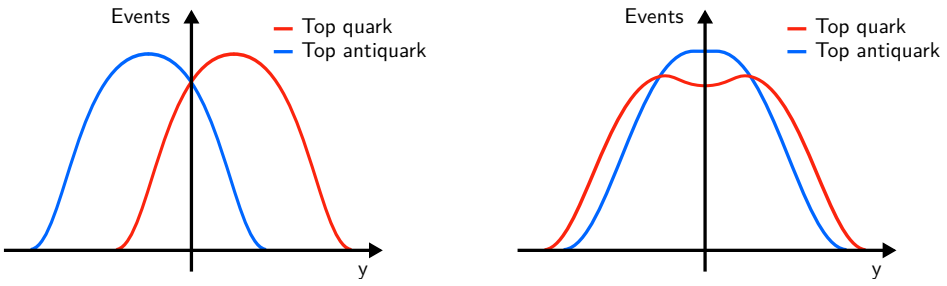

Figure 1.4: Effects on rapidity asymmetries: The figures show an exaggeration of the rapidity distributions of top quark and antiquark at hadron colliders. The left distribution shows a forward-backward asymmetry, which is typical for proton-antiproton colliders, as for example the Tevatron. At proton-proton colliders like the LHC, a central-peripheral asymmetry is expected, as shown in the right figure. Taken from [26].

$$A_C^{\text{LHC8}}(\text{ATLAS + CMS}) = 0.0055 \pm 0.0023(\text{stat}) \pm 0.0025(\text{syst}) \tag{1.19}$$

$$A_C^{\text{LHC8}}(\text{SM prediction}) = 0.0111 \pm 0.0004\,. \tag{1.20}$$

The ATLAS Collaboration recently published a result for A_C at a center-of-mass energy of 13 TeV that differs from zero by four standard deviations [25]:

$$A_C = 0.0060 \pm 0.0015(\text{stat + syst})\,. \tag{1.21}$$

Figure 1.5 gives a comprehensive overview of the latest combined measurements of A_{FB} and A_C including not only the SM predictions but also the variations for different BSM scenarios.

Energy Asymmetry

The previous studies of the charge asymmetry effects in $t\bar{t}$ production reveal the great potential for discovering BSM physics in these precision measurements. As the magnitude of the established asymmetries is decreasing with higher center-of-mass energies it is beneficial to study the influence of other sensitive variables than rapidities on the observable asymmetries. The energy difference $\Delta E_{t\bar{t}} = E_t - E_{\bar{t}}$ between the energies of top quark (E_t) and top antiquark ($E_{\bar{t}}$) in $t\bar{t}$ production in association with one additional jet (in the following denoted as $t\bar{t}j$) is such a variable and gives rise to the energy asymmetry A_E [27–31]:

$$A_E = \frac{N(\Delta E_{t\bar{t}} > 0) - N(\Delta E_{t\bar{t}} < 0)}{N(\Delta E_{t\bar{t}} > 0) + N(\Delta E_{t\bar{t}} < 0)}\,. \tag{1.22}$$

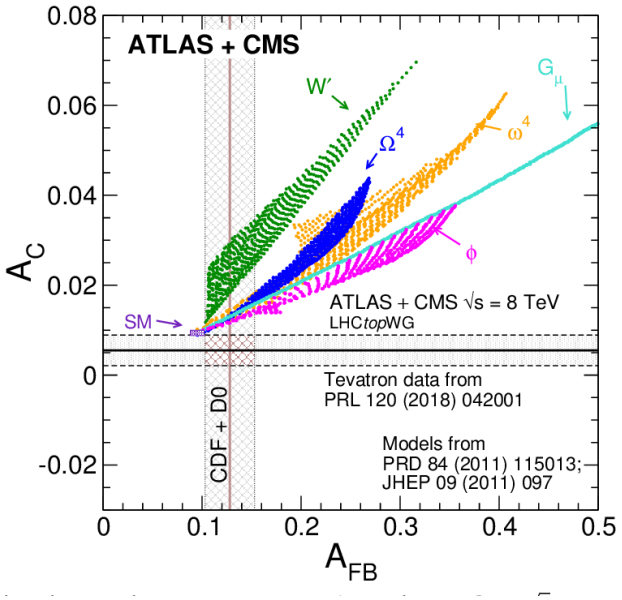

Figure 1.5: Asymmetry measurements at Tevatron and the LHC: A comparison of the combined inclusive measurements of the forward-backward asymmetry A_{FB} at the Tevatron and the charge asymmetry A_{C} at the LHC. Additionally the SM prediction including next-to-next-to-leading order corrections in perturbative quantum chromodynamics (QCD) and NLO electroweak corrections is shown. The plot also shows the predictions for potential BSM contributions: a W' boson, a heavy axigluon (G_μ), a scalar isodoublet (ϕ), a color triplet scalar (ω^4), and a color sextet scalar (Ω^4). Taken from [4].

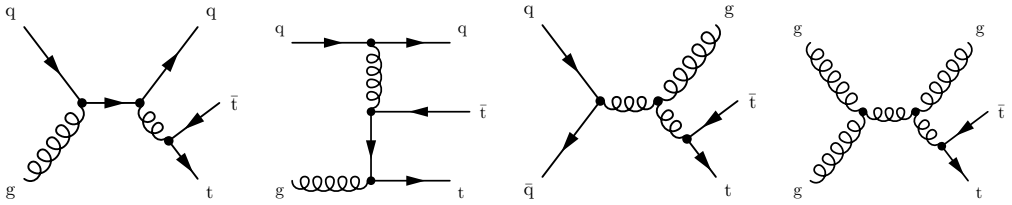

Figure 1.6: Top quark pair + jet production: Shown are Feynman diagrams of leading order $\mathrm{t\bar{t}}j$ production for different initial states. While the qg initial state (left figures, *s* and t channel) has a quark in the final state in addition to the top quark pair, the additional particle producing a jet is a gluon for the qq and gg initial states (right figures), which consequently do not contribute to A_E.

$\Delta E_{t\bar{t}}$ is defined in the $t\bar{t}j$ rest frame. The energy asymmetry arises from the quark-gluon initial-state (qg) production of $t\bar{t}j$ while the gluon-gluon (gg) and quark-quark (qq) initial-states do not contribute to the effect. The Feynman diagrams in Figure 1.6 show the relevant processes of top quark pair production in association with one additional jet at leading order.

A_E has a strong dependence on the scattering angle of the additional jet θ_j (defined in the $t\bar{t}j$ rest frame) and is symmetric in $\theta_j \leftrightarrow \pi - \theta_j$ with a maximum in magnitude around $\theta_j = \frac{\pi}{2}$ (see [28, 29]). To further increase the measurable effect of the energy asymmetry, a differential examination of A_E in θ_j is therefore advantageous:

$$A_E(\theta_j) = \frac{N(\Delta E_{t\bar{t}} > 0, \theta_j) - N(\Delta E_{t\bar{t}} < 0, \theta_j)}{N(\Delta E_{t\bar{t}} > 0, \theta_j) + N(\Delta E_{t\bar{t}} < 0, \theta_j)} \tag{1.23}$$

In proton-proton collisions at the LHC with 13 TeV center-of-mass energy, the incoming quark in the qg process carries on average a higher momentum along the beam axis than the gluon. This transfers to the momentum of the additional jet and the boost of the final state system, which tend to the quark momentum direction. Due to the symmetric proton-proton collision process at the LHC (both qg and gq have the same probability) the direction of the incoming quark is not known but can be guessed by the boost of the $t\bar{t}j$ system in the laboratory frame. This boost is quantified under employment of the rapidity of the $t\bar{t}j$ center-of-mass frame $y_{t\bar{t}j}$ and dedicated selection requirements on $y_{t\bar{t}j}$ allow for an enrichment of events with quark-gluon initial state against gluon-gluon and quark-quark initial states, which do not tend to be boosted in a certain direction.

One can combine the maximum contributions from the qg and the gq channel in the range $0 \leq \theta_j \leq \frac{\pi}{2}$ by defining an optimized scattering angle θ_j^{opt}:

$$\theta_j^{\text{opt}} = \begin{cases} \theta_j & if \;\; y_{t\bar{t}j} > 0 \\ \pi - \theta_j & if \;\; y_{t\bar{t}j} < 0 \end{cases} \tag{1.24}$$

and thus defining an optimized energy asymmetry:

$$A_E^{\text{opt}}(\theta_j^{\text{opt}}) = \frac{N(\Delta E_{t\bar{t}} > 0, \theta_j^{\text{opt}}) - N(\Delta E_{t\bar{t}} < 0, \theta_j^{\text{opt}})}{N(\Delta E_{t\bar{t}} > 0, \theta_j^{\text{opt}}) + N(\Delta E_{t\bar{t}} < 0, \theta_j^{\text{opt}})} \,. \tag{1.25}$$

This optimized energy asymmetry will be considered in the following in three bins of θ_j^{opt}:

- $A_E^{\text{opt},1}$: $0 < \theta_j^{\text{opt}} < 0.3\pi$
- $A_E^{\text{opt},2}$: $0.3\pi < \theta_j^{\text{opt}} < 0.7\pi$

- $A_E^{\text{opt},3}$: $0.7\pi < \theta_j^{\text{opt}} < \pi$

It is shown in Ref. [29] that the energy asymmetry can further be enhanced by applying a selection requirement on the absolute energy difference $\Delta E_{\mathrm{t\bar{t}}}$. The loss in significance due to the reduced cross section that comes in hand with a lower cutoff on $|\Delta E_{\mathrm{t\bar{t}}}|$ is fairly compensated by the increase in magnitude of the asymmetry.

As outlined above, the energy asymmetry has an increased strength in specific phase space regions and the selection, which promises the best combination of measurable magnitude and significance, is given by:

- $|\Delta E_{\mathrm{t\bar{t}}}| > 50\,\text{GeV}$
- $|y_{\mathrm{t\bar{t}}j}| > 0.5$
- $0.3\pi < \theta_j^{\text{opt}} < 0.7\pi$.

Predictions and further discussions of the optimized energy asymmetry will be presented in Section 5.1.

2. The Compact Muon Solenoid Experiment at the Large Hadron Collider

There are in general two main approaches in the search for BSM physics in collision experiments. On the one hand physicists are performing measurements at highest precision, knowing the exact energy of the incoming particles at lepton colliders, as for example currently done at the Belle II experiment in Japan [32] and previously at the LEP collider at the Conseil Européen pour la Recherche Nucléaire (CERN) [33]. On the other hand there are hadron colliders aiming for the direct production of unknown particles at previously unrivaled energies, such as the Tevatron at Fermilab in the USA [34] in the past or the Large Hadron Collider (LHC) [35] in the current epoch of particle physics. Nevertheless also hadron colliders allow for the measurement of precision observables as will be explored in this thesis. This chapter will introduce the LHC and the Compact Muon Solenoid experiment, being the machines that produce and detect the collisions studied in the presented analysis, respectively.

2.1 The Large Hadron Collider

At CERN near Geneva, Switzerland, the Large Hadron Collider has been built in the existing tunnel of the LEP experiment, which has a circumference of 27 km and is located 100 m below the ground level [35–37]. Up to the present day the LHC is the largest and most powerful particle accelerator in the world, colliding protons with center-of-mass energies of up to 13 TeV. In previous stages it has also been operated with lower center-of-mass energies of $\sqrt{s} = 7\,\text{TeV}$ and $\sqrt{s} = 8\,\text{TeV}$ with the plan to perform an update to $\sqrt{s} = 14\,\text{TeV}$ in the near future [38]. In addition to proton-proton collisions the LHC can also be operated with heavy ions, for example in lead-lead or lead-proton collisions.

In order to be injected into the LHC, protons need to pass through various preaccelerator steps that are outlined in Figure 2.1. Subsequent to these steps the protons travel in two

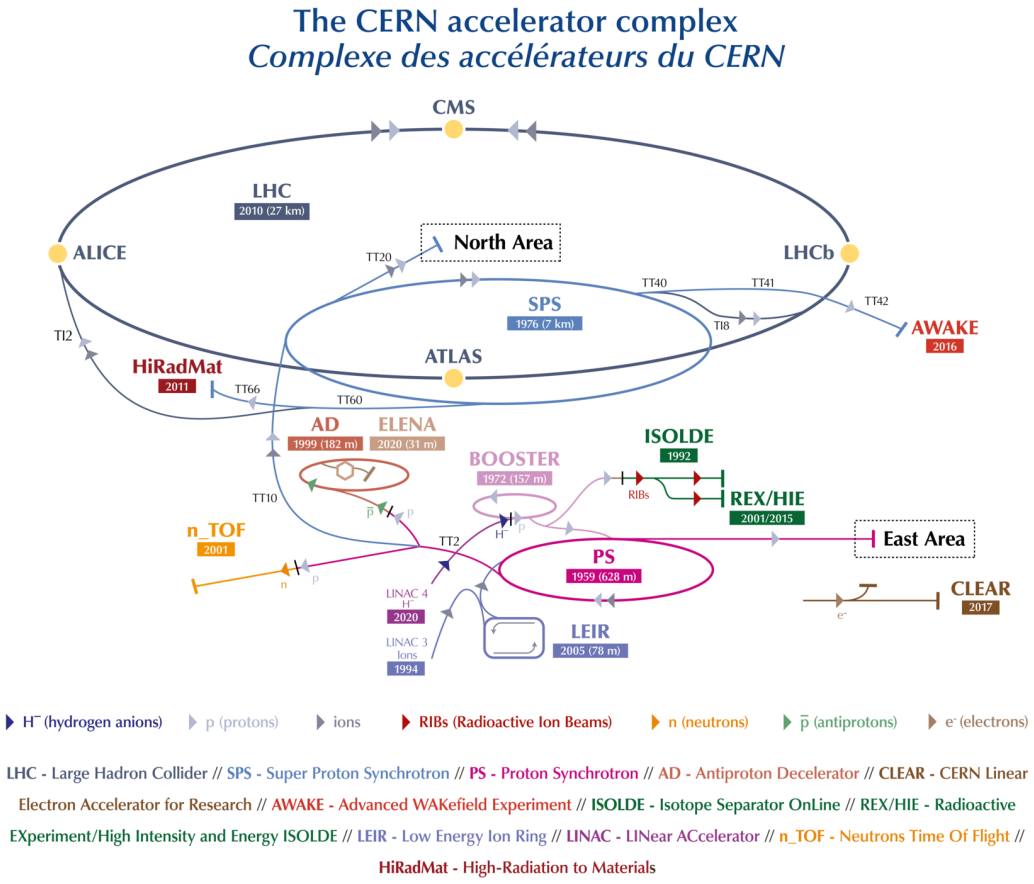

Figure 2.1: The CERN accelerator complex: Shown is a sketch of the main accelerator experiments at CERN. Before being filled into the circular LHC, particles start their accelerating process in the linear accelerator LINAC 4 and traverse the Proton Synchrotron Booster, the Proton Synchrotron (PS), and the Super Proton Synchrotron (SPS), increasing their energy in each of these circular accelerators. The position of the four LHC detectors, namely ATLAS, ALICE, CMS and LHCb is also indicated in the figure. Taken from Ref. [39].

adjacent beamlines in opposite direction to each other through the LHC ring, grouped in up to 2 808 bunches, each containing in the order of 10^{11} protons. The protons are being kept on a circular track through the Lorentz force produced by 1 232 dipole magnets within the LHC and the beam is focused by another 392 quadrupole magnets. In addition a variety of higher multipole order magnets is used to correct smaller imperfections in the field geometry, summing up to about 10 000 superconducting magnets in total. The magnetic field of the dipole magnets has a strength of 8.33 T and they are being cooled to their operation temperature of 1.9 K using approximately 96 t of liquid helium.

For the operation of the LHC it is not sufficient to keep the protons on a bent track, but they need also to be kept in tight bunches and accelerated until reaching their desired collision energy. These tasks are taken care of by eight radiofrequency cavities per beam at Point 4 of the LHC ring. The cavities are operated at 4.5 K and deliver 2 MV each at 400 MHz, and have small energy loss and large stored energy.

A very important figure of merit for a particle accelerator is the instantaneous luminosity L, which is defined as

$$L = \frac{n_\mathrm{B} N_1 N_2 f}{4\pi\sigma_x\sigma_y} \,. \tag{2.1}$$

Here, n_B is the number of bunches and the numbers of particles per bunch are given by N_1 and N_2 for beam 1 and 2 respectively. The frequency of bunch collisions is given by f, while σ_x and σ_y give the size of the bunches normal to the beam axis under the assumption of a Gaussian density distribution of the beams.

The total amount of recordable collision data of a particle accelerator is given by the integrated luminosity

$$L_\mathrm{int} = \int L \,\mathrm{d}t \,, \tag{2.2}$$

which is shown in Figure 2.2 for the LHC at 13 TeV. The amount of events for a specific process with cross section σ is consequently given by

$$N = L_\mathrm{int} \cdot \sigma \,. \tag{2.3}$$

2.2 The Compact Muon Solenoid Experiment

The Compact Muon Solenoid (CMS) experiment is, together with the ATLAS experiment, one of the two multi-purpose detectors at the LHC complex. The detector, being located

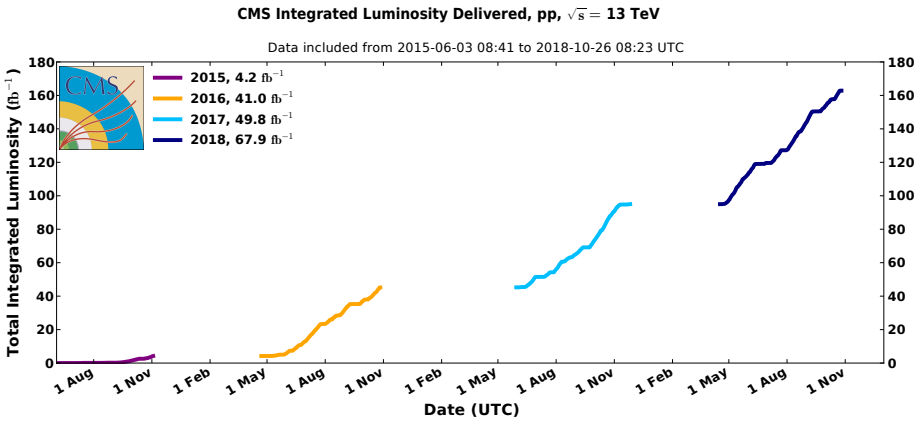

Figure 2.2: CMS Run 2 Integrated Luminosity: The integrated luminosity delivered to the CMS experiment for the four data taking periods from 2015 to 2018 for the LHC at 13 TeV. Taken from Ref. [40].

near Cessy, France, has a weight of 14 000 t and its dimensions are a length of 21 m and a diameter of 15 m [41].

When two protons are colliding with energies of multiple TeV each, there is a variety of physics processes that can take place in such collisions. Due to effects that will be detailed in Chapter 4, it is not possible to determine the particles directly produced in the collision process without misidentification. Rather than this exact detection, a particle detector in high energy physics needs to measure a share of the properties of the particles leaving the collision point at highest precision. This information will allow the physics analyst in later stages to reconstruct the particles taking part in a collision up to a certain level.

The information, which can be directly gained or consecutively derived about a particle focuses on, but is not limited to, the following:

- The momentum direction $\frac{\vec{p}}{|p|}$
- The energy E
- The electric charge
- The particles origin of movement
- The kind of interactions (strong, electromagnetic) it is performing
- The identification of the particle flavor or generation.

The clue to measure as many of these properties as possible, is the layered structure of the detector, which can be seen in Figure 2.3 and will be discussed in detail in the following sections.

For the description of a particle track a dedicated coordinate system is required, which is conventionally defined right-handed for the CMS detector as can be seen in Figure 2.4. As

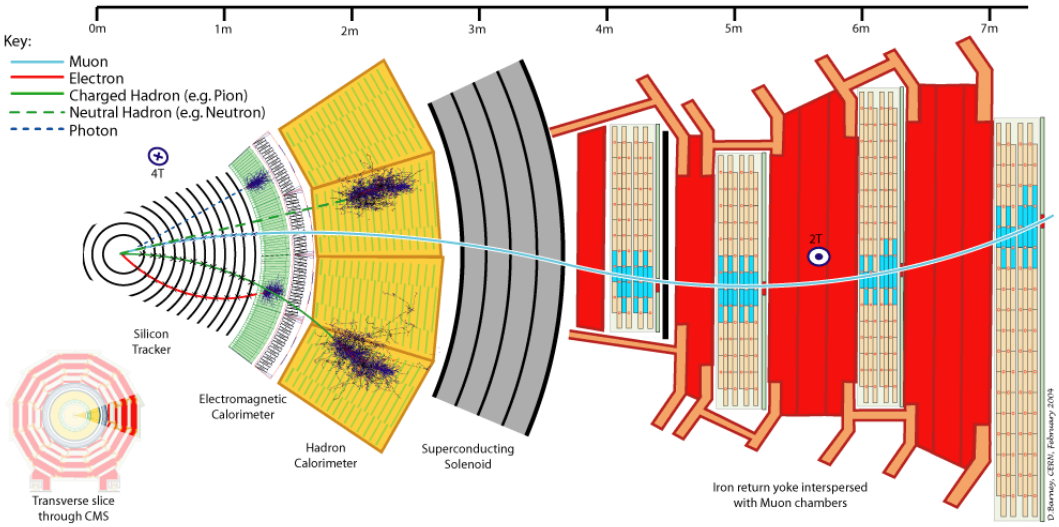

Figure 2.3: CMS Detector Slice: The five main detector components are sketched in a transverse slice through the CMS detector. In addition the signature of different particles in the detector is indicated. Taken from Ref. [42].

the cylindrical detector is symmetric around the beam axis, an angular coordinate system with its origin in the center of the detector is more intuitive than a Cartesian. Therefore the azimuth angle ϕ in the x-y plane and the polar angle θ with regard to the beam axis are defined. In addition to the already introduced Lorentz-invariant rapidity y, a commonly used property is the *pseudorapidity*:

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right), \tag{2.4}$$

which is identical to the rapidity for massless particles. To describe the angular separation of two particles with an observable that is invariant to Lorentz-boosts along the z axis, one defines:

$$\Delta R = \sqrt{(\eta_1 - \eta_2)^2 + (\phi_1 - \phi_2)^2}\,. \tag{2.5}$$

For a hadron collider the exact momentum in z direction of the two colliding partons is unknown, whereas the momentum components in x and y direction are assumed to be roughly zero. It is for this reason that in the later analysis a special role will be assigned to the kinematic observable of the transverse momentum of a particle, which is defined as

$$p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2}\,. \tag{2.6}$$

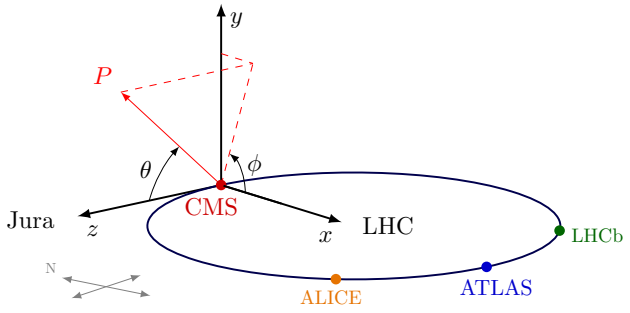

Figure 2.4: Coordinate system of the CMS detector: The z axis of the coordinate system points in the counter-clockwise beam direction, while the x and y axis are defined in a way that they point radially inwards towards the center of the LHC and upwards to the surface, respectively. Taken from Ref. [43].

2.2.1 Tracking System

The innermost part of the CMS detector is the tracking system, which consists of two parts, namely the silicon pixel detector and the silicon strip detector. Both these subdetector systems have the function of precisely measuring the hits of electrically charged particles traversing them. The higher and more accurate the resolution of these hits is, the better the trajectories of particles emerging from the collision point can be reconstructed. The information gain of the tracking system also scales with its coverage of the collision center, which reaches up to $|\eta| < 2.5$ for the CMS detector (see Figure 2.5). The detection system of both tracker components is based on semiconductors, allowing for the measurement of electron-hole pairs, which are produced when a charged particle traverses a detector module. The material of choice for the tracking system at CMS is silicon, which combines the required criteria of robustness against radiation damage, accurate measurement resolution and reasonable cost.

In regard to the aforementioned information gain of particle properties, the tracking system is mainly relevant for the momentum direction, the electric charge and the origin of movement of a particle. The more detailed description of the tracker components in the following is based on Ref. [41] if not stated otherwise.

Silicon Pixel Detector

The analyzed data in this thesis was collected using two different silicon pixel detectors, as by the end of 2016 the existing pixel detector was replaced in the process of the Phase 1 upgrade [44]. While basic measurement principles and pixel properties remained the same, the main difference between the two employed systems is the amount of layers, which increased from three to four in the barrel region and from two to three in the endcap region. In accordance the total amount of pixels increased from about 65 million to roughly 124 million. Each pixel has a surface of approximately 100 µm × 150 µm and a height of 180 µm, allowing in combination for a spatial resolution of 15 to 20 µm. The pixels are

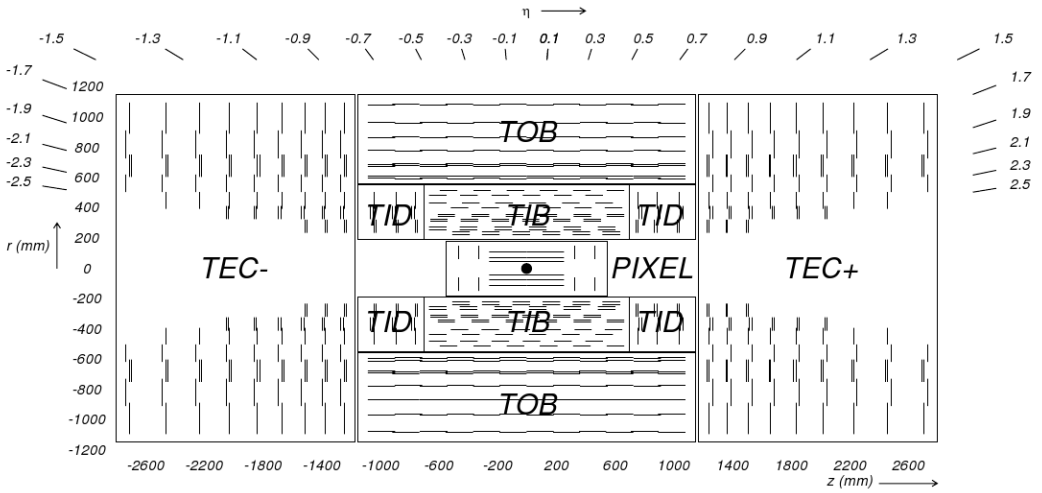

Figure 2.5: The CMS tracker system: The module that is closest to the interaction point is the pixel detector. It is surrounded by the inner strip detector components, which are the tracker inner barrel (TIB) and the tracker inner disks (TID). The outermost layer is composed of the tracker outer barrel (TOB) and the tracker end caps (TEC). Taken from Ref. [41].

organized in groups of 52×80 pixels attached to one readout chip, being grouped further in modules of 8 to 16 readout chips.

Silicon Strip Detector

The silicon strip detector is divided into four parts with a total area coverage of 198 m^2, making it the largest silicon detector in the world. The four submodules are the tracker inner barrel (TIB) and tracker outer barrel (TOB) enclosed by the tracker inner disk (TID) and tracker endcap (TEC) respectively (see Figure 2.5). The TIB consists of four cylindrical layers and is surrounded by the TOB with a total of six layers. The endcap structures contain three (TID) and nine (TEC) wheels on each side of the cylinder respectively. The sensors with sizes between 6×12 cm and 10×9 cm and thicknesses between 320 µm to 500 µm are grouped in over 15 000 modules and sum up to 9.3 million strips in total. The total dimension of the silicon strip detector ranges from a radius of 0.255 m and length of 1.4 m on the inside to a radius of 1.16 m and length of 5.5 m on the outside.

2.2.2 Calorimetry System

The calorimetry system of the CMS detector surrounds the tracking system and consists of two different components, namely the Electromagnetic Calorimeter (ECAL) on the inside and the Hadron Calorimeter (HCAL) on the outside. While electromagnetically interacting particles with small mass, as for example electrons, positrons, and photons, are detected by the ECAL, hadrons, e.g. protons, neutrons, and pions, can traverse the ECAL basically undisturbed and interact with the HCAL via the strong and electromagnetic interactions.

In that sense, the calorimetry system is of main importance for measuring the energy of a particle and for determining whether it can interact via the strong or electromagnetic interactions.

The calorimetry of particles, both in the ECAL and the HCAL, follows an exponential behavior with respect to the distance x that the particles traverse within the detector material. In the case of the ECAL, the remaining energy $E(X_0)$ of an electron, entering the detector with E_0, decreases on average to $\frac{E_0}{e}$ after one *radiation length* X_0. The hadronic *interaction length* λ_0 gives the mean free path of a particle before undergoing an interaction. As X_0 and λ_0 are characteristic for the absorption capability of the detector, these will be used as units for the amount of material in the two components. The purpose of calorimetry is not only to measure the energy of certain particles explicitly, but also to receive an insight in the overall imbalance of the transverse momentum of an event, which points to undetectable particles, for example neutrinos.

Electromagnetic Calorimeter

The electromagnetic calorimeter (see Figure 2.6) consists of lead tungstate crystals ($PbWO_4$, $X_0 = 0.89\,\text{cm}$) as absorber and scintillator material at the same time [45, 46]. It consists of three different components being the ECAL Barrel (EB, coverage up to $|\eta| < 1.48$), the ECAL endcap (EE, coverage from $1.48 < |\eta| < 3.0$) and the preshower detector (ES, coverage from $1.65 < |\eta| < 2.6$). The total number of $PbWO_4$ crystals is 75 848, with each crystal having a length of $25.8\,X_0$ in the EB and $24.7\,X_0$ in the EE, respectively. In front of the EE, the ES is located, which fulfills the task of distinguishing between single high-energy photons and pairs of low-energy photons from neutral pion decays.

The physics effects reducing the energy of traversing electromagnetically interacting particles are bremsstrahlung, Compton scattering, and the production of electron-positron pairs, which create a cascade of particles known as electromagnetic shower. The absorption of these particles by the detector material results in the emission of energy via scintillation light, which can be measured under employment of avalanche photodiodes.

The relative energy resolution of the ECAL indicates the precision that can be reached in the measurement and is given by:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2\,. \tag{2.7}$$

Here S is the stochastic term and refers to uncertainties due to photon statistics and fluctuation of the electromagnetic shower. The noise term N considers the effect of electronics noise in the measurement and the constant term C accounts for miscalibration and non-uniformities, which are dominant for the resolution at high energies. The numerical values have been obtained using an electron test beam [48]:

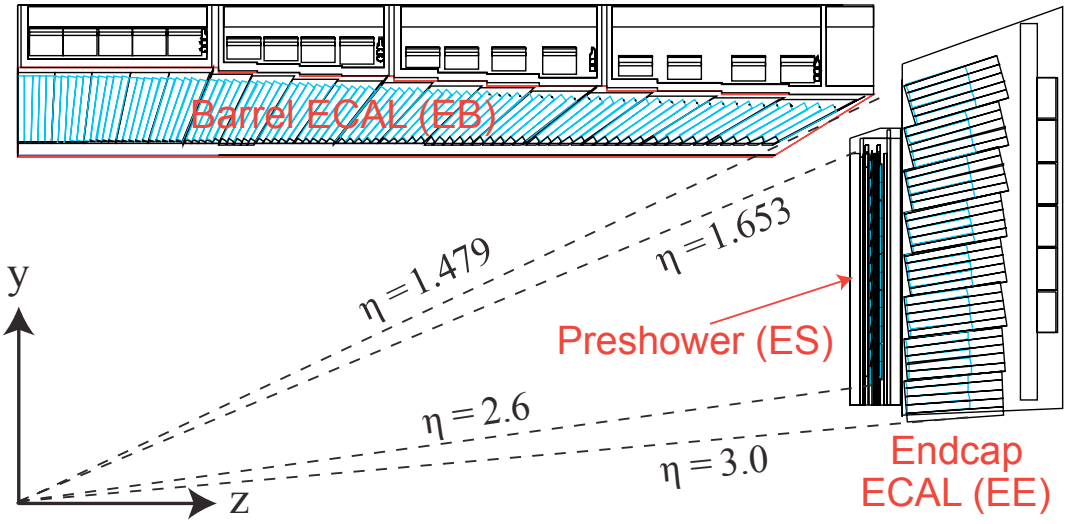

Figure 2.6: The CMS electromagnetic calorimeter: The η-coverage of the three ECAL components is shown. These are the the ECAL Barrel (EB) with 61 200 $PbWO_4$ crystals, the ECAL Endcap (EE) with 14 648 $PbWO_4$ crystals and the Preshower detector (ES), which consists of one layer of lead radiators and one layer of silicon strip detectors. Taken from Ref. [47].

$$\left(\frac{\sigma}{E(\text{GeV})}\right)^2 = \left(\frac{2.8\%}{\sqrt{E(\text{GeV})}}\right)^2 + \left(\frac{12\%}{E(\text{GeV})}\right)^2 + (0.3\%)^2 \,. \tag{2.8}$$

Hadron Calorimeter

In contrast to the previously described ECAL, the HCAL is a sampling calorimeter, this means it consists of layers of different material for absorption and scintillation [49, 50]. A pseudorapidity coverage of up to $|\eta| < 5.2$ is obtained with the four elements of the HCAL. These are the Hadron Barrel (HB) and the Hadron Outer (HO) with a coverage up to $|\eta| < 1.3$ each, the Hadron Endcap (HE) covering $1.3 < |\eta| < 3.0$, and the Hadron Forward (HF) extending the coverage from $|\eta| = 3.0$ to $|\eta| = 5.2$ (see Figure 2.7). The absorption material is brass (70 % copper and 30 % zinc) with an interaction length of $\lambda_\text{I} = 16.42\,\text{cm}$. The secondary particles that are produced through inelastic scattering are of lower energy than the incoming particles and the thereby produced hadronic shower can be detected by the scintillation layers made of plastic material.

It is of note that the HB corresponds to a radiation length of only $5.82 \cdot \lambda_\text{I}$ and is therefore not able to fully absorb the hadronic shower, while both the HE and HF have a radiation length of roughly $10 \cdot \lambda_\text{I}$. As the HB could not be enlarged due to the geometric restrictions of the ECAL on the inside and the solenoid magnet on the outside, this problem is tackled by the installation of the HO outside of the solenoid. It is employing the solenoid magnet

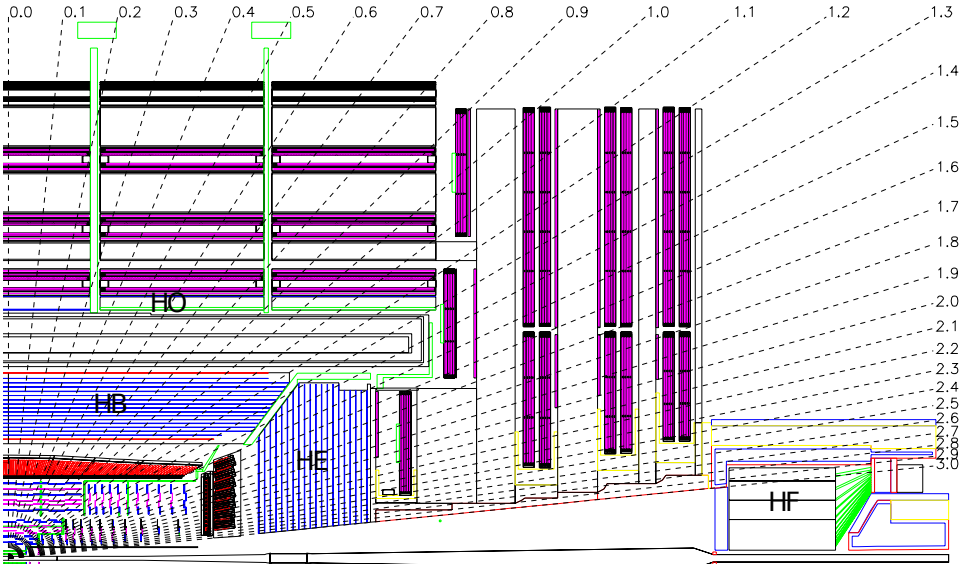

Figure 2.7: The CMS hadron calorimeter: Shown is one quarter of the CMS detector with the subdetectors of the HCAL and their η coverage being highlighted. These are the Hadron Barrel (HB), the Hadron Endcap (HE), the Hadron Outer (HO), and the Hadron Forward (HF) calorimeter. Taken from Ref. [41].

as an additional absorber and thereby extending the radiation length in the central region to more than $10 \cdot \lambda_I$.

2.2.3 Solenoid Magnet

The Lorentz force is bending the track of charged particles that move perpendicular to the field lines of a magnetic field. The direction and strength of a particle's trajectory curvature in conclusion reveal information about its electric charge and momentum. Due to the high energy of particles produced in proton-proton collisions at 13 TeV center-of-mass energy, a strong magnetic field is required to obtain a measurable deviation in their movement with respect to a scenario without a magnetic field applied. At the CMS detector this is achieved by a 220 t solenoid magnet, being the largest magnet of this kind in the world with a diameter of 6.3 m and a length of 12.5 m [51]. Cooled down to a temperature of 4.45 K the solenoid enters a superconducting state and creates a magnet field strength of 3.8 T, while storing 2.6 GJ of energy. In order to control the magnetic flux on the outside of the magnet and thereby strengthening the field on the inside, a 12 500 t steel return yoke is installed surrounding the solenoid. It is by far the heaviest part of the CMS detector and has a diameter of 14 m with a barrel length of 13 m.

2.2.4 Muon System

The muon has not only been of special importance in the search for the Higgs boson (e.g. $H \rightarrow ZZ \rightarrow 4\mu$) or various BSM physics searches, but it is also the only detectable particle

that can easily pass through the already described subdetector components of the CMS detector without being fully absorbed. It is for these reasons that the largest and outermost part of the detector is dedicated to the detection and identification of this minimum ionizing particle. The three different kinds of muon subdetectors, which are embedded in the return yoke of the solenoid magnet, have the purpose of identifying muons and of rapidly deciding whether to store a collision event for later processing (triggering). As the muon chambers are penetrated by a magnetic field of 2 T due to their positioning within the return yoke, the muon momentum measurement can be improved through their trajectory therein.

The muon system is divided into a barrel region and an endcap region as can be seen in Figure 2.8 and consists of Drift Tube (DT) chambers, Cathode Strip Chambers (CSCs) and Resistive Plate Chambers (RPCs) [52–54]. In addition, a first layer of Gas Electron Multiplier (GEM) chambers has been installed in 2017 [55], a process that will be continued in the second and third long shutdown of the LHC [56]. The basic principle of the different muon chambers is inert gas being ionized by charged particles traversing the detector volume. Anodes and cathodes with high voltage accelerate the free electrons and heavy ions, resulting in an avalanche of free electrons that can be measured as a current at the anode wire.

The coverage of the muon system reaches up to $|\eta| < 2.4$, where individual regions are covered by the different types of chambers. 250 DT chambers are part of the barrel region, covering the detector up to $|\eta| < 1.2$, a region where the particle flux is relatively low. The drift time in these subdetectors is of the order of 380 ns, which is not fast enough for making a trigger decision. They are therefore complemented by the RPCs, which are resident in both the barrel and the endcap region and cover the detector up to $|\eta| < 1.9$. In total 1 056 RPCs are installed, enabling fast track building at trigger level and compensating the slower detection response of the DT chambers. The remaining coverage of $|\eta| < 2.4$ is given through the 540 CSCs in the endcaps. With their high radiation resistance and fine segmentation, these detector components can handle the higher background rates and the nonuniform magnetic field in the high-$|\eta|$ regions.

2.2.5 Trigger System

As described in Section 2.1 the LHC delivers proton-proton collisions with a rate of 40 MHz, corresponding to a bunch crossing interval of 25 ns. The sheer amount of data that is being produced in these collisions cannot be recorded by present-day technology. It is not only due to this constraint, but also due to the fact that most collisions contain physical processes of little interest, that the events need to be filtered in real-time. This is achieved by the trigger system with its two main components, the hardware-based Level-1 (L1) trigger and the software-based High-Level Trigger (HLT), which are presented in Figure 2.9.

Only information from the calorimeter and muon systems is accessed by the L1 trigger and the readout of all remaining detector data is delayed until the trigger made a decision on the retention of an event. The data rate is at this stage reduced to the order of 100 kHz

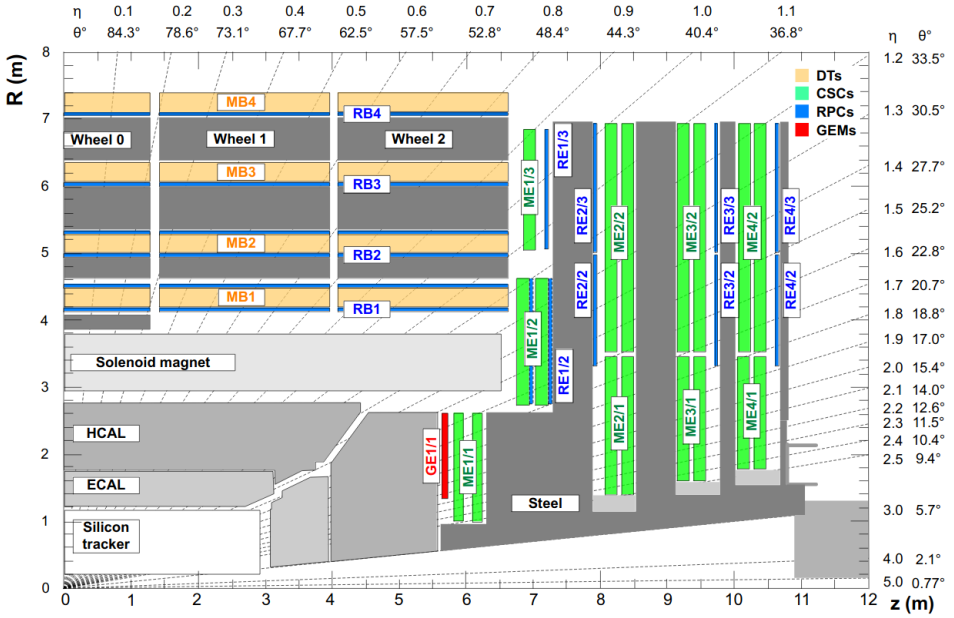

Figure 2.8: The CMS muon system: The figure shows a quarter of the CMS detector including all subsystems with the muon chambers being highlighted. These are the Drift Tube (DT) chambers in the barrel region, the Cathode Strip Chambers (CSCs) in the endcap region and the Resistive Plate Chambers (RTCs) in both regions. In addition, the already installed Gas Electron Multiplier Chambers (GEMs) in the endcap are indicated. Taken from Ref. [55].

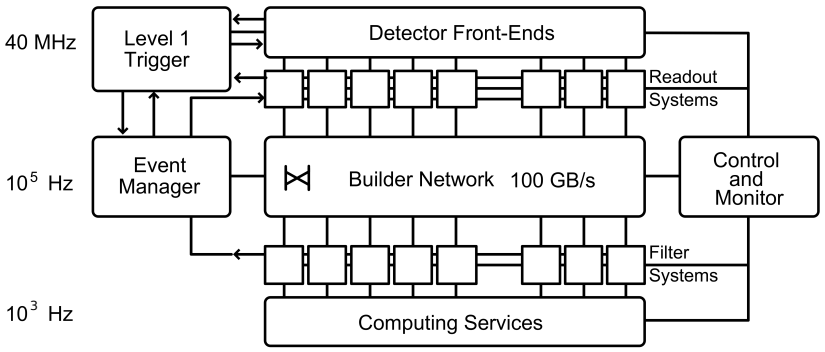

Figure 2.9: The CMS trigger system: In the data acquisition system of CMS the data rate is reduced from 40 MHz to the order of 1 kHz. This is achieved by two subsequent trigger systems, being at first the hardware-based Level-1 (L1) trigger and following up the software-based High-Level Trigger (HLT). Adapted from Ref. [57].

[57, 58]. In a second step the prefiltered events need to pass the requirements of the HLT, which is also the final decision on whether to store an event or not. This is achieved with customized CMS software on dedicated computing farms and reduces the data rate to the order of 1 - 2 kHz.

2.2.6 Computing

The computing power and storage capacities required for analyzing and preserving the collected data of the LHC experiments are enormous. The Worldwide LHC Computing Grid (WLCG) [59, 60] serves these tasks for the ATLAS, ALICE, CMS, and LHCb experiments. It is organized in a hierarchical structure with different tier layers from 0 to 3 (see Figure 2.10), each of them having specific tasks and optimizations. The Tier-0 center is directly located at CERN and stores the raw data collected by the particle detectors, as well as performing a primary processing. From here on the reconstructed data is distributed to the 13 Tier-1 centers around the world, one of which is located at the Karlsruhe Institute of Technology (KIT). These perform a further reprocessing of the data and thus reduce the required storage amount per collision event. In addition the Tier-1 centers serve as dataset storage in various processing stages, which can be accessed by Tier-2 centers. The final physics analyses are mostly performed at the roughly 160 Tier-2 centers, which do not provide storage capacities as large as the Tier-1 centers, but high CPU power. The Tier-3 centers are only indirectly related to the WLCG and provide resources for local analysis tasks with moderate requirements.

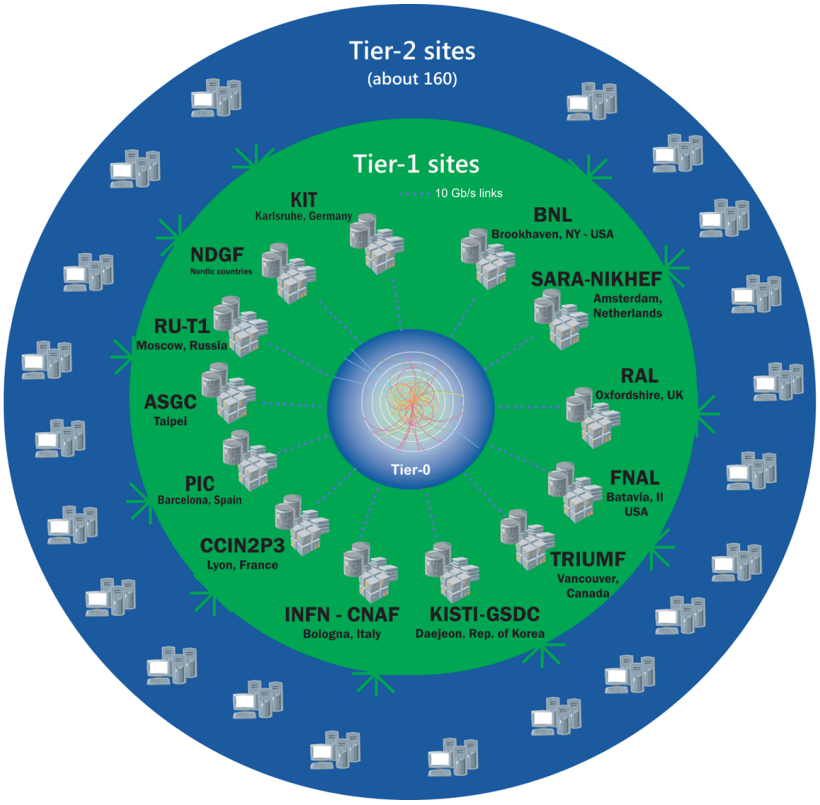

Figure 2.10: The Worldwide LHC Computing Grid (WLCG): Shown are the Tier-0 site at CERN and its connection to the 13 Tier-1 centers around the world. The indicated Tier-2 cites are mostly hosted at institutes and universities as it is the case for the Tier-3 sites, which are not shown at all in the figure. Taken from Ref. [61].

3. Statistical Methods and Multivariate Analysis Techniques

Results in particle physics experiments are always a comparison between the prediction on theory level and the measurement of the experiment, as both of these components are worthless in a solitary contemplation. High precaution is demanded in this process as not only the measurement but also the prediction is subject to statistical fluctuations. Following these premises a thorough understanding of the statistical methods, which are employed for the derivation of a result, is required in order to conduct a physics analysis.

The methods of importance for this thesis are partially integrated in the employed software frameworks and will be outlined in sections 3.1 and 3.2. As multivariate analysis techniques (MVAs) are applied in various steps of data processing with a special focus on boosted decision trees (BDTs) within the analysis, their basic functionality will be discussed in Section 3.3.

3.1 Maximum Likelihood Estimation

The assumed statistical model and the observed data are the two basic ingredients, that need to be brought in relation in order to derive a meaningful result of an analysis. The statistical model is defined as a probability density function (PDF) $f(\vec{x}|\vec{y})$, with the measurements $\vec{x} = \{x_1, x_2, \ldots, x_N\}$ of N independent values and the parameters $\vec{y}$ of the model that need to be estimated.

The maximum likelihood estimation (MLE) is a maximum a posteriori estimation method and can be employed to find the parameters that maximize the likelihood of observing the given measurement [62]. The likelihood function is defined as the product of the PDF for each observed value x_i:

$$L(\vec{y}) = \prod_{i=1}^{N} f(x_i|\vec{y})\,. \tag{3.1}$$

By maximizing this function, the best parameter set $\hat{y}$, based on the measured data, can be determined. As this is a challenging task in practice, one often simplifies the terminology to the minimization of the negative log-likelihood:

$$F(\vec{y}) = -\ln L(\vec{y}) = -\sum_{i=1}^{N} \ln f(x_i|\vec{y})\,, \tag{3.2}$$

and thereby converting the product to a computationally less intensive sum and the maximization to a more stable minimization avoiding large function values. For the best parameter set $\hat{y}$ the following two conditions hold true:

$$\left.\frac{\partial(F)}{\partial y_i}\right|_{y=\hat{y}} \stackrel{!}{=} 0, \tag{3.3}$$

$$\left.\frac{\partial^2(F)}{\partial y_i \partial y_j}\right|_{y=\hat{y}} \text{ is negative definite.} \tag{3.4}$$

The probability of observing n statistically independent events is Poisson distributed and accordingly given by:

$$P(n|\lambda) = \frac{\lambda \cdot e^{-\lambda}}{n!}\,. \tag{3.5}$$

As in principle a counting experiment is performed in this thesis, binned histograms are used for the distributions and the likelihood for a total of m bins is given by the Poisson probabilities of each bin i:

$$L(\vec{\mu}) = \prod_{i=1}^{m} P(n_i|\lambda_i)\,. \tag{3.6}$$

As it is outlined in more detail in sections 5.2 and 5.7, the observed events in this analysis arise as a combination of signal and background processes and the expectation value λ is parameterized as:

$$\lambda(\vec{\mu}) = \lambda_{\mathrm{b}} + \sum_{k=1}^{K} \mu_k \cdot \lambda_{\mathrm{s},k} \,. \tag{3.7}$$

The contributions of background processes are given by λ_{b} while $\lambda_{\mathrm{s},k}$ corresponds to the contribution of the specific signal process k. The signal strength modifiers

$$\vec{\mu} = \{\mu_1, \mu_2, \ldots, \mu_K\} \tag{3.8}$$

give the relation between the observed and predicted number of events for each signal process ($\mu_k = N_k / N_{k,\mathrm{SM}}$) and are the parameter of interest to be adjusted in the fit.

3.1.1 Nuisance parameters

Each measurement underlies a variety of uncertainties of both theoretical and experimental nature, which are not of immediate interest when determining the fit parameters via MLE. Nevertheless these systematic uncertainties need to be taken into account as they alter the prediction and therefore the outcome of the analysis. They are introduced in the fitting procedure via nuisance parameters θ influencing both the signal $s(\theta)$ and background $b(\theta)$ processes.

Two different types of uncertainties need to be considered, as they differ in their way of altering the prediction. While rate uncertainties do not have any impact on the overall shape of a histogram (in the following called template), shape uncertainties can modify the template through the correlated or uncorrelated shift of events per bin. Rate uncertainties change the normalization of a template and are considered in the fit by extending the likelihood with a log-normal prior:

$$\pi(n) = \frac{1}{n\sigma_n\sqrt{2\pi}} \exp\left[-\frac{(\ln(n) - \overline{n})^2}{2\sigma_n^2}\right] . \tag{3.9}$$

Here n is the number of observed events with $\overline{n}$ being the mean number of predicted events and σ_n the corresponding uncertainty.

Shape uncertainties, which can change the shape and the rate of a template, are implemented with a template morphing method [63, 64]. For each uncertainty two additional templates are generated, corresponding to a shift of one standard deviation in upwards and downwards direction, respectively. These templates can for example be obtained by reweighting the simulated events or by repeating the complete analysis chain with systematically varied observables. The interpolation and extrapolation between the templates and the nominal distribution is done using a polynomial function.

In addition to the described nuisance parameters, the uncertainty due to the finite size of the simulated samples also needs to be taken into account, which is done via the Barlow-Beeston light method [65].

The methods as discussed above are evaluated under employment of the software package COMBINE [66, 67], which itself relies on the RooFit library [68].

3.2 Unfolding

The motivation for performing an analysis is not only the immediate comparison of the result with a given theoretical prediction. It is also beneficial to preserve the outcome of the measurement in a way that it can be used for the validation of future theory models or compared and combined with the observations of other experiments. Providing a result that allows for these possibilities is not a trivial task, as it basically means that all dependencies specifically induced by the measurement itself need to be removed. One could argue that at least for theory models this could be avoided by simulating the detector response in addition and comparing the measurement with the very model at hand, but in practice this is rarely done. The procedure applied for receiving such an undiluted result is commonly known as unfolding and aims at correcting for effects of both detector resolution and background contributions, as well as for reducing the impact of statistical fluctuations [69].

The true distribution of an observable may be given by y with each event contributing to a certain bin i. As most distributions in high energy physics cannot be determined analytically, but only through simulation, the only information at hand is the expected truth distribution $\tilde{y}$ with its corresponding bin contents. Due to the inevitable detector effects, the same observable can only be measured with a finite precision, and each event contributes to a certain bin j of the average expected distribution $\tilde{x}$ of the measurement. For simplicity an equal number of bins n for the true and measured distribution is assumed and the probability that an event in $\tilde{y}_i$ is expected in $\tilde{x}_j$ is given by R_{ji}, summing up to:

$$\tilde{x}_j = \sum_{i=1}^{n} R_{ji}\tilde{y}_i \,. \tag{3.10}$$

The matrix R will in the following be denoted as response matrix or migration matrix. The actual measurement x does not only underlie detector effects, but also statistical fluctuations and background contributions b, which need to be taken into account:

$$x_j = \sum_{i=1}^{n} R_{ji}\tilde{y}_i + b_j \,. \tag{3.11}$$

The procedure is schematically shown in Figure 3.1. For the above mentioned reasons there is a general interest in the distribution of $\tilde{y}$ and one is tempted to simply solve the equation via matrix inversion. This imposes various problems [62] and a variety of methods [70–72], which will not be further discussed in this thesis, have been developed to reduce the impact of statistical fluctuations and to derive unbiased results. It should be mentioned at this stage that there is no general consensus or strict recipe in the high energy physics community on how to apply unfolding or whether it should be used at all [73].

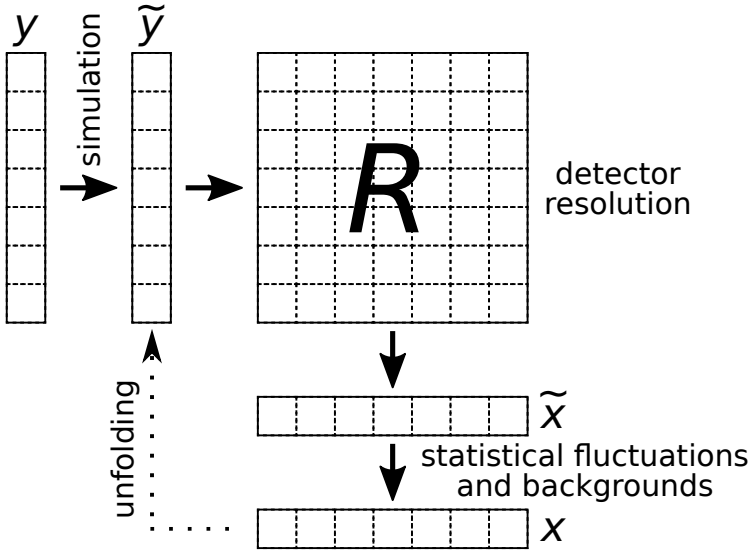

Figure 3.1: Unfolding procedure: Shown are the various effects smearing the true distribution y. Unfolding procedures in general aim at determining $\tilde{y}$ from the measured distribution x.

3.2.1 Unfolding with MLE

The method of choice in the presented analysis is unfolding via MLE, where it should be noted that this does not correspond to the so-called EM iteration method [74], which also uses MLE. Apart from the comparably easy and straightforward implementation within the COMBINE framework, there is a series of further advantages in this method with respect to for example χ^2-minimization based unfolding. The fit performed when determining the parameters of interest in the MLE allows not only for a direct inclusion of the background processes but also for all systematic uncertainties to be covered in terms of nuisance parameters. Additionally, the uncertainties due to the finite size of the simulated samples are directly considered in the fitting procedure and the response matrix is allowed to have dimension $n \times n$, which cannot be handled by various other methods, but is of importance for the analysis in this thesis.

As the true distribution y is not accessible in any way, the term *truth-level* distribution will in the following refer to the simulated distribution $\tilde{y}$. In contrast to the already introduced measurement x, the histogram $\tilde{x}$ of the observable without background contributions but with detector effects applied will be called *detector-level* distribution.

In order to construct the relevant model for a multidimensional MLE, each bin i of the truth-level histogram will be considered as a separate signal process s_i. The contribution of such a signal process to a detector-level bin j is accordingly given by the relevant matrix element of R_{ji}. In that sense the content of a detector-level bin is simply the sum of various signal processes, similar to equation 3.10. By associating a signal strength parameter μ_i to each signal process s_i, the likelihood can be defined as outlined in Equations 3.6 - 3.7. The

unfolded distribution u is then given by the truth-level distribution $\tilde{y}$ where the content of each bin is multiplied with the respective signal strength parameter as determined from the fitting procedure.

A more concrete example and the explicit setup for the presented analysis will be given in Section 5.7.

3.2.2 Regularization

For the sake of completeness the term regularization will be shortly outlined here, as it is an important inclusion to many unfolding algorithms. When solving equation 3.11 for $\tilde{y}$ by matrix inversion, the statistical fluctuations of the measurement x are amplified, which is an undesired behavior. To compensate for this, certain smoothness conditions are often imposed on the truth-level distribution $\tilde{y}$. This is especially important if one has to deal with an ill-conditioned problem, meaning the condition number of the response matrix is large ($O(10^5)$), as in this case the unfolded distribution is biased towards the simulation. Technically this is often achieved by introducing a penalty-term $P(\tilde{y})$ in the likelihood, which depends on the truth-level values (so-called Tikhonov-regularization [75]). It will be shown that regularization does not need to be applied in the presented analysis for both qualitative and quantitative reasons.

3.3 Boosted Decision Trees

Within the analysis, the reconstruction of events is an important task and as for many events there is more than one possibility of reconstruction, a decision has to be taken on which reconstruction is more likely to be correct. Due to the large amount of events both in simulation and in observed data, this decision has to be made in an automated way by an algorithm. The method of choice for this task is to employ boosted decision trees, which are a type of multivariate analysis technique and have been shown in the past to outperform χ^2-minimization based algorithms in similar tasks. The implementation of the BDTs is taken care of by the TMVA [76] toolkit within ROOT [77].

3.3.1 Decision Trees

A decision tree, as sketched in Figure 3.2, is in general used to categorize an event based on a predefined set of input variables. The tree is structured in a root node, multiple internal nodes and ending nodes called leafs. While the leafs classify an event to be of a specific category (e.g. signal or background, correct or wrong), the root and internal nodes make a decision according to the value of a specific input variable. Each node divides a set of events into two subsets for which this process is repeated until reaching a leaf. The structure of the tree is limited by the maximum number of internal nodes and leafs and the decision whether a leaf classifies an event as signal or background is determined in the training process of the tree.

The training procedure of a decision tree corresponds to the optimization of selection requirements on the input variables and is performed with a sample for which the correct

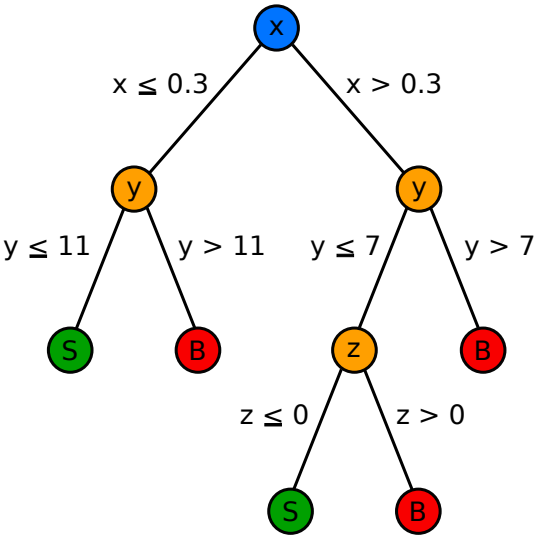

Figure 3.2: Sketch of a decision tree: Events are being classified according to the three values x, y and z. After the decision in the root node (blue), further decisions are taken in the internal nodes (orange) until events are classified in leafs as signal (green) or background (red).

classification into signal or background is known. As a first step the sample is scanned for the variable which is best in order to separate the events and the best separation value is determined. This process is being repeated for the two remaining subsets until the structural limits of the tree or the required separation gain are reached. For each leaf the classification into signal or background is according to the majority of events with respective classification. As events i often have a specific weight w_i, this weight is considered instead of pure event counting. The purity P of a leaf, which gives a measure of the misclassified events in it, is correspondingly defined as the fraction of the weight sum of signal (n_s) and background (n_b) events:

$$P = \frac{\sum_{i=1}^{n_\mathrm{s}} w_i}{\sum_{i=1}^{n_\mathrm{s}} w_i + \sum_{i=1}^{n_\mathrm{b}} w_i} \,. \tag{3.12}$$

The statistical dispersion of a node is given through the Gini impurity:

$$G = P \cdot (1 - P) \,. \tag{3.13}$$

The sample that is used for the training must be statistically independent of the sample that is used for later evaluation in order to avoid a bias towards the statistical fluctuations in the training sample in the final result.

3.3.2 Boosting

The general idea of boosting is to not merely use a single decision tree for the classification of an event, but to utilize the average of many ($O(100)$) classification trees and train them subsequently. By considering the weighted average over all trees, the impact of statistical fluctuations in the training samples can be damped with respect to a single decision tree. This set of trees is called boosted decision tree (BDT) and is obtained in this thesis using the AdaBoost algorithm [78].

After the training of each tree, misclassified events are reweighted and applied for the training process of the next tree, making them more important in this training iteration. The score of a tree m is defined as α_m using the misclassification rate ϵ_m:

$$\alpha_m = \beta \cdot \ln \frac{1 - \epsilon_m}{\epsilon_m} \,. \tag{3.14}$$

The strength of the boost is defined here by β and using $Y_i(x) = 1$ for misclassified events and $Y_i(x) = 0$ for correctly classified events the event weights for the training of the next tree are given by:

$$w_i \rightarrow w_i \cdot e^{\alpha_m \cdot Y_i(x)} \,. \tag{3.15}$$

The final BDT output is given by the weighted sum of all outputs:

$$T(x_i) = \sum_{m=1}^{N} \alpha_m \cdot T_m(x_i) \,, \tag{3.16}$$

where N is the total number of trees and $T_m(x_i)$ is the output of a single tree with value 1 for signal and -1 for background.

3.3.3 Overtraining

In the training process of any multivariate analysis technique special precautions have to be taken so that the method does not learn the statistical fluctuations of the training sample and loses its ability to generalize. This effect is known as overtraining and indicated in Figure 3.3. In order to avoid it, the performance of the discriminator during the training process is evaluated on a statistically independent testing sample. Overtraining is present if the separation power still increases on the training sample while it decreases on the testing sample. An equivalent performance on both samples is desired and accordingly the training process has to be stopped before overtraining occurs.

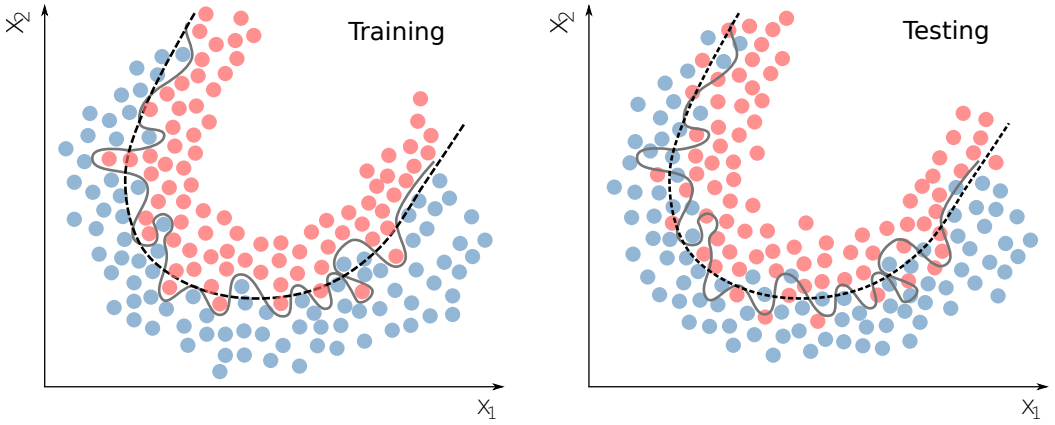

Figure 3.3: Effect of overtraining: The figures indicate the separation power of multivariate discriminators on a training (left) and a testing (right) sample. The dashed line shows an equal separation between signal (red) and background (blue) events on both samples with no overtraining present. The solid line on the other hand is overtrained and has a perfect separation on the training sample but performs significantly worse on the testing sample, as it learned the statistical fluctuations of the training sample. Taken from Ref. [79].

4. Simulation and Reconstruction

To benefit in any way from the signals that can be recorded with the CMS detector, one does not only need to combine and cluster these signals to form physics objects, but also needs to simulate the expected observations. It is only for the comparison of the actual measurement to the expected outcome, which allows for any reliable statement on the involved processes.

Apart from the desired interaction of two partons, there is a variety of other processes happening in parallel at the collision of two protons, which need to be understood for a correct simulation of the complete scenario. These processes are described in Section 4.1. The state-of-the-art tools employed for the simulation of events are outlined in Section 4.2. The reconstruction of pure detector signals to physics objects is given in the last section of this chapter, which also defines the quality requirements applied on these objects in order to be considered in the analysis.

4.1 Proton-Proton Scattering Process

The collision of two protons at high energies gives rise to various phenomena, as can be seen in Figure 4.1. A successful comparison of simulation and observation is only possible if all of these phenomena are accounted for appropriately when simulating an event.

Hard Scattering Process and Parton Distribution Functions

The hard scattering process is the interaction of two partons of the colliding protons and the target of most physics analyses at the CMS experiment. The cross section of such a process depends on the proton momentum fractions x_1 and x_2 of the two partons, which at hadron colliders are not known a priori. The probability of observing a parton with momentum fraction x of the proton momentum is given by the parton distribution function (PDF) $f(x, \mu_F^2)$, which is valid for valence quarks, gluons, and sea quarks, and depends on the given energy scale (called factorization scale) μ_F. As the PDFs at any desired scale

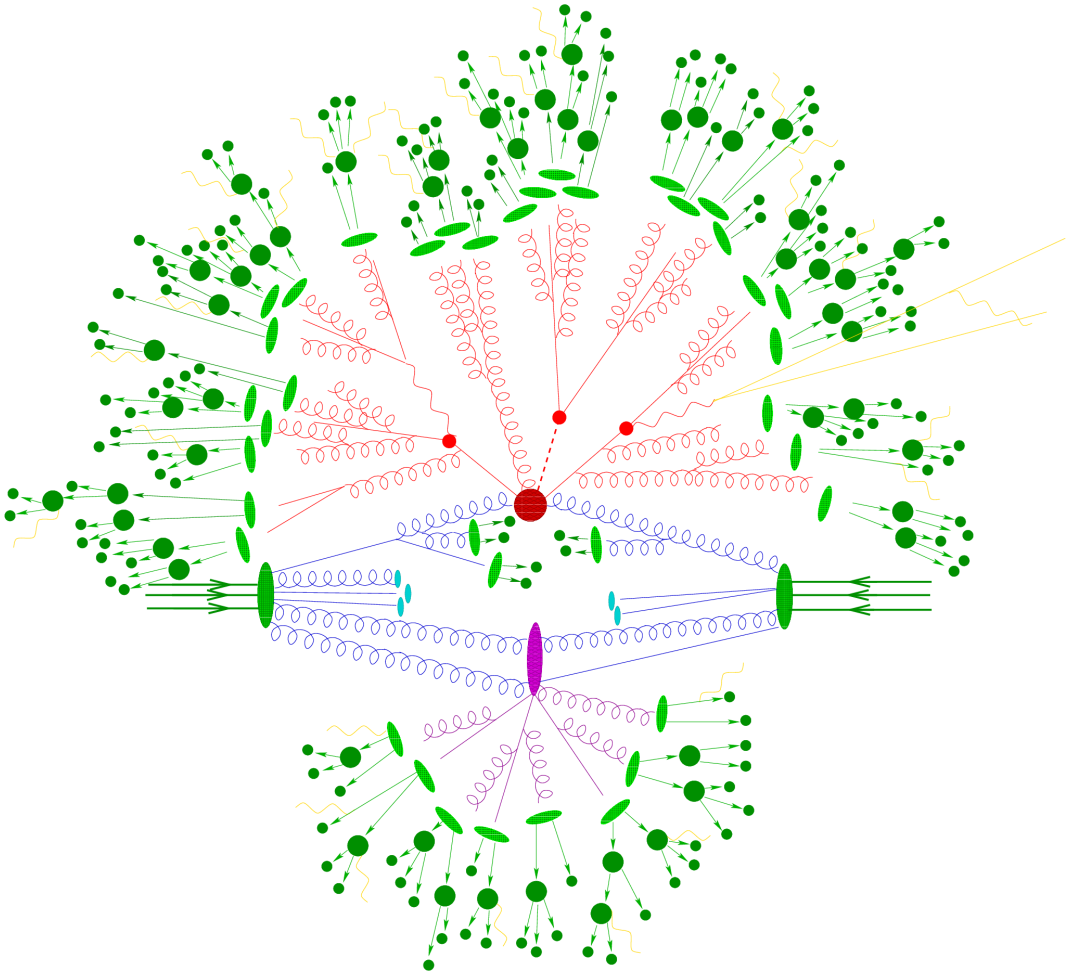

Figure 4.1: Sketch of a proton-proton scattering process: The highest interest is in the collision of the two partons indicated with the big red circle, also called the hard proton-proton scattering process, which is usually the physics process of interest. For the simulation of the full proton collision one also needs to consider the second scattering process, called underlying event and indicated by the purple ellipse. Subsequent processes like the radiation of gluons (red) and photons (yellow) need to be taken into account as well as the green structures, which refer to the hadronization process and the decay of hadrons. Taken from Ref. [80].

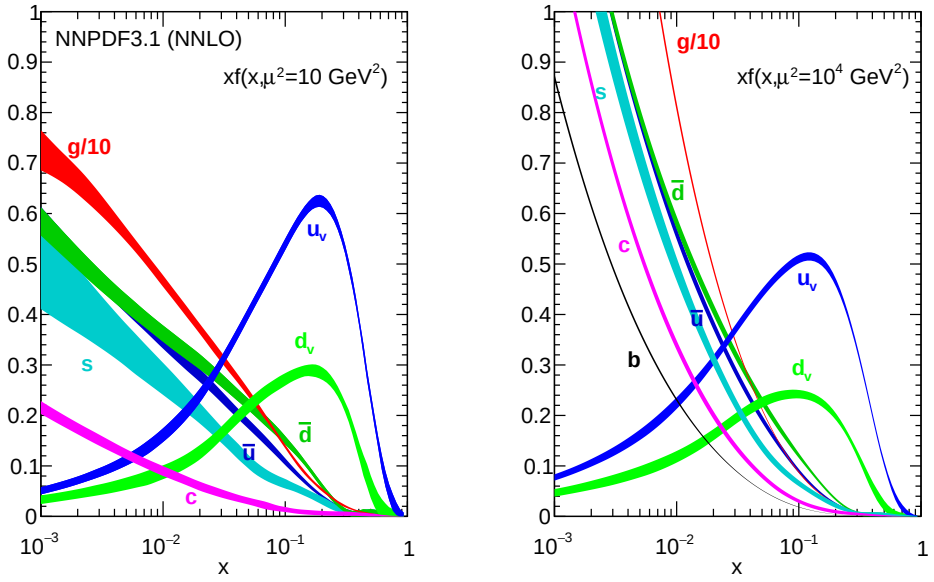

Figure 4.2: The Neural Network Parton Distribution Function (NNPDF) set: For the two values of the factorization scale of $\mu_F^2 = 10\,\text{GeV}^2$ and $\mu_F^2 = 10\,000\,\text{GeV}^2$ the proton PDFs of the NNPDF3.1 set at next-to-next-to-leading order are shown. Taken from Ref. [85].

are unknown a priori and cannot be directly calculated, one needs to evaluate the PDFs measured at a given scale and transfer them in perturbation theory under employment of the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations [81–83]. In Figure 4.2 the PDF set NNPDF [84, 85], which is used in this thesis, is shown for two factorization scale values of $\mu_F^2 = 10\,\text{GeV}^2$ and $\mu_F^2 = 10\,000\,\text{GeV}^2$.

To determine the cross section of producing a particle X in the hard process of a proton-proton collision, all possible initial state combinations with the partonic process $jk \rightarrow \text{X}$ need to be considered, where j and k refer to the possible initial state partons. The matrix element and therefore the cross section of such a partonic process is determined by evaluating all corresponding Feynman diagrams. As one can always add the radiation of gluons, internal lines, or loops, the number of possible Feynman diagrams becomes infinite and an exact calculation is impossible. A good approximation, however, can be made by a perturbation series in α_S, for energies at which $\alpha_S \ll 1$:

$$\sigma = \sigma_{LO}\left(1 + \sum_i \sigma_i \left(\frac{\alpha_S}{2\pi}\right)^i\right). \tag{4.1}$$

The leading-order cross section σ_{LO} includes only Feynman diagrams with the smallest possible number of strong vertices. Each subsequent order i corresponds to one additional

vertex with α_S, either through a real emission or virtual corrections. As the strong coupling constant is larger than the coupling constant of the electromagnetic interaction by approximately two orders of magnitude, the impact of electromagnetic corrections is small compared to QCD corrections at the LHC. It should be noted that α_S is not constant, but depends itself on the renormalization scale μ_R, which is considered in perturbative calculations. The calculations for most processes of relevance in this thesis are either done in next-to-leading order (NLO) or next-to-next-to-leading order (NNLO) in perturbation theory, with the calculations being more precise the higher the order.

Parton Shower

Any particle within the hard scattering process that carries a color charge can radiate off additional gluons. Depending on whether the emitting particle is part of the initial or the final state of the Feynman diagram, this is known as initial-state radiation (ISR) or final-state radiation (FSR). These additional gluons can decay into quark-antiquark pairs or further emit gluons themselves, resulting in a cascade of particles denoted as parton shower. The emission of these partons is of lower energy than the hard scattering process and thus at increasing α_S. This also means that a simulation of the parton shower is not possible via perturbative matrix element calculation and instead it is carried out using Sudakov form factors [86, 87] and the Altarelli-Parisi splitting functions [83]. Special care needs to be taken as for example a leading order process with the simulation of one additional radiation has the same final state particles as the same process at NLO without radiation and thereby double counting is possible. This is considered via merging and matching of the matrix elements and the parton shower, which is in practice done via algorithms as for example MLM [88] or FxFx [89] merging.

Hadronization

Due to the confinement of particles carrying a color charge, these cannot propagate freely but have to form color-neutral bound states, being either mesons or baryons. This process is known as hadronization and has to be included in the simulation process via phenomenological models. Due to the relatively low energy scale at which hadronization takes place and the correspondingly high value of α_S, no perturbative calculation of the processes is possible. One of the most commonly employed models is the Lund string model [90], which describes the gluons as field lines, resulting in tubes of these lines between color-charged objects due to the self-interaction of gluons. The energy of these strings can increase, as for example by spatial separation of particles, and at a certain energy threshold a new quark-antiquark pair is created. This continues until color-neutral states only are reached. As the newly generated particles do not need to be stable themselves, the decay of these particles needs to simulated as well.

Underlying Event and Pileup

When two protons collide it is not sufficient to consider only the processes involved with the hard scattering of two partons, but also the color-charged remnants of the protons need

to be taken into account. These induce further hadronization and cause additional signals within the detector, which is known as underlying event and needs to be simulated.

The high luminosity at the LHC cannot be reached by accelerating and colliding single protons, but bunches of them with 10^{11} protons per bunch. This leads in general to more than one interaction per bunch crossing as multiple proton collisions take place at the same time. This is known as in-time pileup and needs to be taken into account in the simulation as well as out-of-time pileup. The latter corresponds to signals in the detector of previous bunch crossings, as they happen in an interval of only 25 ns, which is short in respect to the detector response.

4.2 Generation of Simulated Events

The simulation of the processes described in the previous section is performed by Monte Carlo (MC) event generators. Multiple of these computing tools are used and interfered for the creation of events and will be shortly introduced in the following. They all rely on the Monte Carlo method [91] but differ in the order of perturbation theory for which they can be employed. It is in general important to simulate a vast number of events for each investigated process, as due to the statistical nature of quantum mechanics the outcome of a single collision cannot be predicted, but only probability density functions can be derived. This section also covers a short description of the detector response simulation tool GEANT 4.

MadGraph5_aMC@NLO

The two existing MC event generators MADGRAPH5 [92] and MC@NLO [93] have been combined in the MADGRAPH5_aMC@NLO [94] package, which can be used for the simulation of events and the calculation of cross sections at LO and NLO precision. It includes the matching of calculated matrix elements to the parton shower simulation and resolves possible double counting effects at NLO with the MC@NLO technique. Therefore, negative event weights are introduced in order to subtract parton shower contributions and thereby obtaining correct NLO predictions. This induces the negative side-effect of reducing the effective number of events and can lead to unphysical predictions with negative event counts in special phase spaces. The advantage of the generator is the rather easy usability, as the user only needs to provide the initial state and final state particles for the process of interest and the calculation of all possible diagrams is carried out automatically.

POWHEG

The positive weight hardest emission generator POWHEG [95–97] is the second important MC generator used for the simulation of events in this thesis. It can be used for the simulation up to NLO accuracy and has the characteristic of calculating the hardest emission already in the matrix element. The subsequent parton shower simulator therefore needs to provide p_T-ordered showers. The method used by POWHEG has the advantage that no negative event weights are introduced and all simulated events can be used for an analysis, but the generator comes with the drawback that only predefined processes can be simulated.

PYTHIA

The MC generator Pythia [98, 99] provides a full event simulation, including matrix element calculation and simulation of underlying event, hadronization, and parton shower. Its main disadvantage is that the calculation of matrix elements can only be done at leading order precision, which does in general not fulfill the precision requirements of most analyses. Pythia is therefore often interfaced with the previously introduced matrix element generators and applies the remaining steps from parton shower simulation onwards. It uses the Lund string model for hadronization and is compatible with Powheg as the simulated parton shower is p_{T}-ordered. Different tunes with dedicated parameters for the modeling of underlying event and hadronization are available for the generator, where the events in this thesis mostly have the CP5 tune [100] employed.

GEANT 4

The last step that needs to be simulated in order to obtain a result that can be compared to measured data is the response of the CMS detector. The software package Geant 4 [101–103] is used for the simulation of both the interaction of particles with the detector material, and the signals of the readout electronics. After this stage the simulation and the measured data are on the same level and are reconstructed in the same way for the further analysis.

4.3 Reconstruction of Objects

The raw electric signals, which are obtained from either the full event simulation or the actual measurement with the detector, cannot directly reveal any insight to the physics processes of interest. It is the combination of these signals and the reconstruction to physics objects that allow for a comparison between prediction and experiment. In the following the algorithms and reconstruction methods for the most important physics objects of the analysis will be outlined. In addition, the applied quality requirements and preselection steps on the different kinds of objects will be presented. As the analysis aims at unfolding the observations to a detector-independent level, the corresponding truth-level counterpart will be introduced for selected objects.

4.3.1 Particle Flow Algorithm

Instead of considering each component of the CMS detector solitary for the reconstruction of objects, the information of all subdetectors is combined for reconstructing the stable particles of an event. The algorithm that is applied for this holistic procedure is the particle-flow (PF) algorithm [104], which requires a high granularity of the detector components and has first been successfully used by the ALEPH experiment at the LEP collider [105]. The first step of the algorithm is the creation of so-called *PF elements*, which can be either charged-particle tracks in the tracking detector or energy clusters in the ECAL or HCAL. These elements are then combined by a dedicated linking algorithm to form *PF blocks*, as indicated in Figure 4.3. The remaining component of the CMS detector, the muon system,

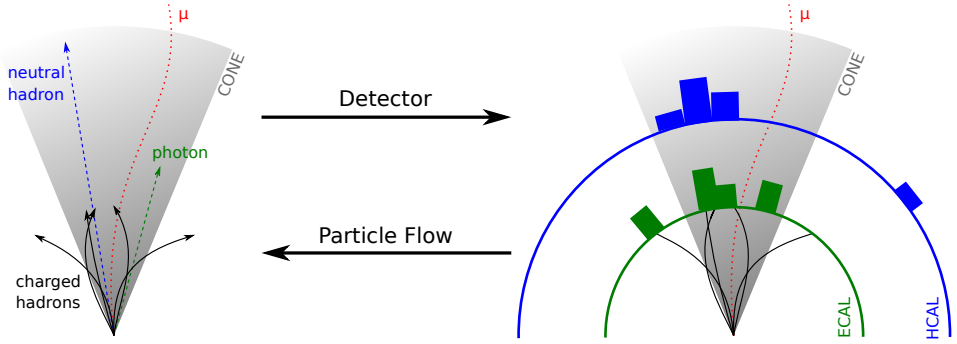

Figure 4.3: The Particle Flow algorithm: In a first step, the detector signals are clustered to *PF elements*, which are then linked together to *PF blocks*. These can be assigned and reconstructed to the different high-level physics objects. Taken from Ref. [106].

comes to play now and its information is utilized in order to identify muons within the PF blocks. The identified blocks are then removed from the full list of blocks and a similar sequential procedure is applied in the following for the identification of electrons and photons, as well as for neutral and charged hadrons. A further postprocessing step is applied and high-level objects, as for example jets or the missing transverse momentum $p_{\mathrm{T}}^{\mathrm{miss}}$, can now be built.

4.3.2 Particle Tracks and Primary Vertices

One of the inputs for the previously described PF elements are the trajectories of charged particles within the tracker components of the CMS detector. They are reconstructed by the combination of hits, induced from charged particles traversing the silicon pixel and the silicon strip detector. The precision of these tracks crucially depends on the alignment of the tracker modules, and possible misalignment needs to be taken into account, in order to reach a resolution in the order of μm [107]. The Combinatorial Track Finder (CTF) algorithm [108], which is based on Kalman filters [109–111], is applied for the reconstruction procedure. With this iterative approach, tracks nearby the interaction region and a transverse momentum of $p_{\mathrm{T}} > 0.8\,\mathrm{GeV}$ are reconstructed first, followed by a removal of the corresponding hits from the remaining event in order to reduce the combinatorics of subsequent iterations. A four step procedure is executed for each iteration, starting with the generation of a track seed from only two or three hits. Such a track is extrapolated to outer layers using Kalman filters in a second step, and additional hits within the vicinity of this extrapolated track are assigned to it. The third step consists of determining an improved trajectory to the assigned hits, again with the use of Kalman filters. Tracks that do not fulfill certain quality criteria are discarded in a final step.

Not only direction and momentum of the reconstructed tracks are of interest for the reconstruction of objects, but it is especially important to find their origin and to investigate the spatial adjacency of multiple such points. If more than one track can be associated

to a common origin, such a point is denoted as a vertex, with a special focus given in the following to the primary vertex (PV), which has the highest sum of p_T^2 of associated tracks. Vertices other than the PV can indicate contributions of in-time pileup from other proton-proton collisions and using the charged hadron subtraction (CHS) technique [112], PF elements are thus removed from the collection if they are assigned to a vertex along the beam-line other than the PV.

The algorithm for the reconstruction of vertices is taking only high-quality tracks as input and is clustering them in a first step by their z position in relation to the beam spot. Special care needs to be taken in order to not split a valid vertex into multiple subclusters of tracks, which is ensured by a deterministic annealing algorithm [113]. After the optimal number of vertices has been found by this procedure, the exact position of each vertex is determined using the adaptive vertex fitter algorithm [114]. Here, vertex candidates need to have at least two associated tracks and a weighting procedure gives the probability of a track being compatible with a certain vertex.

To be considered for the analysis, an event needs to have a primary vertex candidate that fulfills certain quality requirements, for example being located within a cylinder of radius 2 cm and length 24 cm around the center of the detector. Additionally, a minimum of four tracks need to be assigned to the PV, giving it at least five degrees of freedom.

The ability for both the reconstruction of tracks and the derivation of primary vertices as described here is a characteristic of the detector and consequently the definition of a truth-level counterpart is not applicable. This does not impose a problem on the desired unfolding procedure, as the primary vertex does not play any further role in the reconstruction of events, and the impact of the track reconstruction will be covered by the higher-level objects introduced in the following.

4.3.3 Muons

The CMS detector is, as outlined in previous sections, optimized for the detection, identification and reconstruction of muons. Accordingly these are the first particles to be reconstructed by the PF algorithm using the information of the tracker and the muon system. A separate reconstruction of possible muon tracks is performed with the information of each component, leading to the so-called tracker tracks from tracker signals and standalone-muon tracks from the muon system signals [54]. For the combination of these tracks two different approaches are considered, differing by the set of information they start the reconstruction with. The *inside-out* approach takes all tracks with $p_T > 0.5\,\text{GeV}$ into account and reconstructs so-called tracker muons by matching these tracks to a short muon segment built up from hits in the DTs or CSCs. All possible interactions along the path are considered in this algorithm, which relies only on small segments in the muon system and provides high efficiency for low p_T muons ($p_T < 5\,\text{GeV}$) that do not penetrate the muon system very deeply. For the reconstruction of a global muon, the *outside-in* approach is employed, starting with a standalone-muon track and searching for a matching tracker track. In case of a successful match, the associated hits in both detector systems are fitted simultaneously using a Kalman filter. This reconstruction algorithm significantly improves

Table 4.1: Criteria for identifying tight muons. Only global muons with well-fitted tracks can fulfill the tight muon ID requirements.

Criterion	tight muon ID
Global muon	yes
PF muon	yes
χ^2/ndof of global-muon track fit	< 10
Number of muon chamber hits	≥ 1
Number of muon segments in muon stations	≥ 2
Transverse impact parameter d_{xy} wrt. primary vertex	$< 2\,\text{mm}$
Longitudinal distance d_z wrt. primary vertex	$< 5\,\text{mm}$
Number of pixel hits	> 0
Number of tracker layers with hits	> 5

the resolution for muons with high transverse momentum ($p_\text{T} > 200\,\text{GeV}$). In general the information of the standalone-muon track can be used solitary for the reconstruction of a standalone muon, but due to the worse resolution with respect to global muons and higher background of cosmic muons, standalone muons are not considered in the analysis.

In addition to the quality requirements included within the PF algorithm, a further set of identification conditions is applied on muons to be considered for in later analysis stages. These are according to the recommendations of the Muon Physics Objects Group (MUO POG) of the CMS Collaboration [115] and summarized in Table 4.1. Only muons that pass all criteria of the tight muon identification (ID) flag are utilized in order to further suppress both cosmic muons and muons stemming from kaon or pion decays. Depending on the η range, the tight muon ID has an efficiency between 95% to 99%.

An additional requirement is applied on muons in order to filter out so-called nonprompt muons stemming from a semileptonic decay inside a jet. Therefore the relative isolation is defined as follows:

$$I^\mu_\text{rel} = \frac{1}{p^\mu_\text{T}} \left[\sum p^\text{ch. had.}_\text{T} + \max\left(0, \sum p^\text{neutr. had.}_\text{T} + \sum p^\gamma_\text{T} - \Delta\beta \sum p^\text{ch. had., PU}_\text{T}\right)\right] . \tag{4.2}$$

The transverse momentum of the muon is given by $p_{\text{T},\mu}$, while the other terms correspond to the transverse momentum of charged hadrons from the primary vertex ($p^\text{ch. had.}_\text{T}$), neutral hadrons ($p^\text{neutr. had.}_\text{T}$), photons ($p^\gamma_\text{T}$), and charged hadrons from pileup ($p^\text{ch. had., PU}_\text{T}$). Only contributions within a cone with radius $R = 0.4$ are considered for the relative isolation of a muon and the correction factor $\Delta\beta$ is applied to estimate the contributions of neutral hadrons originating from pileup. As it is not possible to determine whether a neutral hadron is stemming from the primary vertex or from pileup, $\Delta\beta$ is required as correction and estimated to be 0.5 [116]. The tight requirement of $I^\mu_\text{rel} < 0.15$ is applied in order to

reduce the amount of nonprompt muons from QCD multijet production. The efficiency of this criterion has been determined to be approximately 95% [115].

Within the further analysis workflow, only muons will be considered that are classified to be tight muons. In addition to the already outlined selection criteria of passing the tight muon ID flag and having a relative isolation smaller than 0.15, a tight muon needs to be within the central region of the detector ($|\eta| < 2.4$) and have a transverse momentum of at least $p_T > 27$ GeV.

The truth-level counterpart for a reconstructed muon is a dressed lepton. The definition of dressed leptons is given at the end of the following section, as these are also the truth-level counterpart of reconstructed electrons.

4.3.4 Electrons

The reconstruction of electrons is performed subsequent to the reconstruction of muons and the subtraction of corresponding PF elements from the event and is far more challenging due to the low mass of electrons [117, 118]. As they can lose already a significant amount of their energy in terms of bremsstrahlung within the tracker material due to their relatively low mass, they do not traverse as many detector components as muons and leave a broad shower in azimuthal direction in the ECAL. Narrow showers are in general easier to reconstruct and dedicated algorithms thus have to be applied for the more complicated electron reconstruction. As it is the case for muons, two different approaches are considered for the reconstruction of an electron.

In a first attempt neighboring energy deposits in the ECAL crystals (clusters) are combined to so-called *superclusters*, where the cluster with highest energy deposit is the seed for the combination. Employing Gaussian Sum Filters (GSF) [119, 120], tracks of the tracker detector are matched to the superclusters of an event, using the respective seeds as starting point for the matching. This technique yields a good reconstruction of electrons with transverse momentum above 5 GeV. The use of CTF algorithms with Kalman filters for track reconstruction is not applicable for electrons, as their curvature changes too much due to bremsstrahlung and it is accordingly difficult to assign hits to the track.

Electrons with p_T of less than 5 GeV can be better reconstructed using Kalman filters together wit a GSF for track reconstruction and matching these tracks to the energy deposits of the ECAL.

In order to increase the quality of reconstructed electrons a set of requirements is applied to them. These requirements are provided by the E/gamma Physics Object Group of the CMS Collaboration and are bundled in an identification (ID) flag [121]. Table 4.2 summarizes the selection requirements for the tight electron ID, which needs to be passed by electrons to be considered in the analysis. Due to the different background conditions in the barrel and the endcap region of the ECAL, a dedicated set of selection requirements is applied for an electron, depending on the absolute value of its supercluster pseudorapidity.

Table 4.2: Identification of tight electrons. The selection requirements that are applied on an electron in order to pass the tight ID are different in the barrel and the endcap region.

Criterion	$\lvert\eta_{SC}\rvert \leq 1.479$	$\lvert\eta_{SC}\rvert > 1.479$
SC shower shape	< 0.0104	< 0.0353
$\lvert\Delta\eta$(SC, track)$\rvert$	< 0.00255	< 0.00501
$\lvert\Delta\phi$(SC, track)$\rvert$	< 0.022	< 0.0236
Hadronic energy/EM energy	$< 0.026 + 1.15/E_{SC} + 0.0324\rho/E_{SC}$	$< 0.0188 + 2.06/E_{SC} + 0.183\rho/E_{SC}$
Electron isolation	$< 0.0287 + 0.506/p_T$	$< 0.0445 + 0.963/p_T$
$\lvert 1/E_{SC} - 1/p_{track}\rvert$	< 0.159	< 0.0197
Expected missing inner hits	≤ 1	≤ 1
Pass conversion veto	yes	yes

The isolation I^{e}_{rel} of an electron is given by:

$$I^{e}_{rel} = \frac{1}{p^{e}_{T}} \left[\sum p_T^{\text{ch. had.}} + \max\left(0, \sum p_T^{\text{neutr. had.}} + \sum p_T^{\gamma} - \rho A_{\text{eff}}\right)\right]. \tag{4.3}$$

It gives a measure for the energy deposited by other particles within a cone radius of 0.3 around the electron. Contributions from pileup interactions are included via the average transverse momentum density ρ and the effective area A_{eff} of the cone. A selection requirement on the electron isolation is of high importance for the electron identification and is therefore already included in the tight ID requirement.

In the final selection stages of the analysis only electrons will be considered that are classified as tight electrons. In order to be classified as tight, an electron needs not only to pass the tight ID flag but also to be within the central region of the detector ($|\eta| < 2.1$) and have a transverse momentum of at least $p_T > 35\,\text{GeV}$.

The truth-level counterpart of both muons and electrons is a dressed lepton, which is defined in a detector-independent way with the same characteristics as applied in the Rivet framework [122, 123]. The collection of dressed leptons of an event is accessible in all simulation samples of the CMS Collaboration at level of the NanoAOD event format [124, 125]. Dressed leptons include the effect of QED FSR in a model-independent way and are constructed by final state leptons from the simulation with the energy of all photons within a cone of $\Delta R = 0.1$ around the lepton. Double counting of photons is avoided as a photon is only added to the closest final state lepton if it happens to be within the capture radius of more than one.

4.3.5 Photons and Hadrons

The PF elements that have been assigned to reconstructed muons or electrons are removed from the collection of an event and the remaining PF blocks are assigned to either hadrons

or photons. Similar to the reconstruction of electrons, but without the matching to associated tracks, superclusters in the ECAL can be assigned to isolated photons [126]. The remaining PF blocks are mainly due to charged hadrons (charged pions, kaons and protons), neutral hadrons (neutral kaons and neutrons), and nonprompt photons (neutral pion decays).

For the remaining clusters in the ECAL and the HCAL the coverage of the tracker is of special importance. Clusters within the tracker acceptance of $|\eta| < 2.5$ that cannot be linked to any track, will be reconstructed as nonisolated photons and neutral hadrons. In case a matching track can be assigned to one of these clusters, the corresponding PF elements are assigned to a charged hadron. As the information of the tracker is not accessible for clusters in the ECAL or HCAL in the region with $|\eta| > 2.5$, a further division between neutral and charged hadrons is not possible for the remaining clusters. In case of a cluster in the ECAL that cannot be linked to any HCAL cluster, this is assigned to a photon. Connected clusters between the ECAL and the HCAL are in a final step assigned to neutral or charged hadrons.

4.3.6 Jets

For most physics analyses at the LHC the solitary charged and neutral hadrons are not of particular interest, as they do not directly correspond to the final state particles of the investigated collision processes. Rather than that, a collection of multiple such objects within a narrow cone, called jet, can contain all constituents of a particle shower caused by the hadronization of a final state parton. Even though jets are rather abstract objects without a direct physics meaning, they are crucial for most analyses, allowing for an easier comparison of data and simulation. In addition to hadrons, jets can also contain nonisolated leptons and photons, which are produced via the weak or electromagnetic decay of hadrons.

The main algorithms that can be applied for the reconstruction of jets are either cone-based, meaning they cluster all particles within a given cone size, or sequential, meaning they combine particles to a jet in a sequential procedure. Any jet reconstruction algorithm needs to fulfill the requirement of being both collinear and infrared safe. Collinear safety is assured if the result of an algorithm is independent of the collinear splitting of a hadron. If the algorithm does not show sensitivity to the soft emission of gluons, infrared safety is given. Sequential recombination algorithms are most widely used in high-energy physics and provide robust results that can be compared to theory predictions, as they are collinear and infrared safe.

The two main ingredients for the sequential jet algorithm applied in this thesis are the distance parameter d_{ij} between two particles i and j, and the beam distance $d_{i\mathrm{B}}$ between a particle i and the beam:

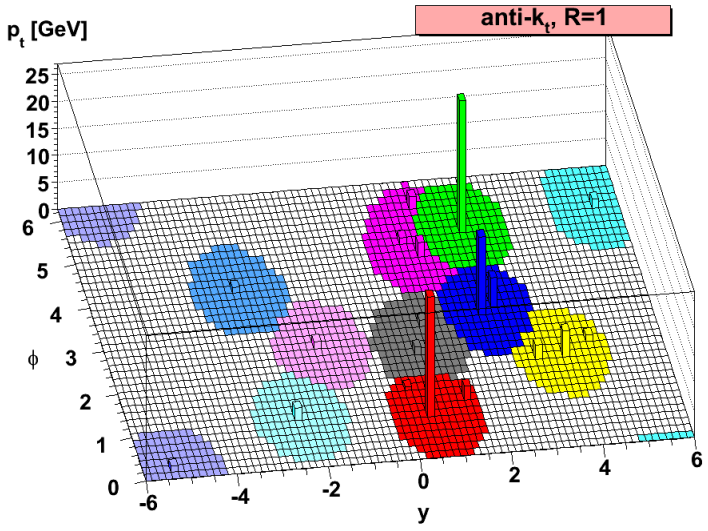

Figure 4.4: The anti-k_T jet clustering algorithm: The clustering of partons into jets with the anti-k_T algorithm is shown in the y-ϕ plane. The transverse momentum of the partons is indicated by their z component and each color corresponds to one reconstructed jet. Taken from Ref. [131].

$$d_{ij} = \min\left(k_{\mathrm{T},i}^{2n}, k_{\mathrm{T},j}^{2n}\right)\frac{\Delta_{ij}^2}{R^2}, \tag{4.4}$$

$$d_{i\mathrm{B}} = k_{\mathrm{T},i}^{2n}\,. \tag{4.5}$$

The transverse momentum of the particles is given as k_T and the distance between two particles in the y-ϕ plane is defined as $\Delta_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}$. The value of R defines the desired radius parameter of the cone and the parameter n further defines the reconstruction procedure and is commonly chosen to have a value of either $n = 1$ (k_T algorithm [127, 128]), $n = 0$ (Cambridge/Aachen algorithm [129, 130]), or $n = -1$ (anti-k_T algorithm [131]). The jets used in the presented analysis are reconstructed with the anti-k_T algorithm and have a distance parameter of either $R = 0.4$ or $R = 0.8$. The anti-k_T algorithm calculates the distance parameter between all particles and the two particles with the smallest distance $d_{ij} < d_{i\mathrm{B}}$ and $d_{ij} < d_{j\mathrm{B}}$ are combined to a new particle. All distance parameters are now recalculated and the procedure is repeated in an iterative procedure until the distance parameter of a particle d_{ij} to all remaining particles is bigger than its beam distances $d_{i\mathrm{B}}$. At this stage the particle is removed from the set and accounts for a jet. The procedure, which is illustrated in Figure 4.4, is repeated until all particles are clustered into a jet.

As outlined, two different cone radius parameters of $R = 0.4$ and $R = 0.8$ are considered for jet clustering in this thesis. In order to distinguish between them, jets clustered with

$R = 0.4$ will be denoted as *slim jets* in the following, while jets clustered with $R = 0.8$ will be named *fat jets*.

Certain quality requirements are applied on slim jets in order to be considered within the later analysis steps. As it was the case for muons and electrons, these requirements are bundled in identification (ID) flags as defined by the CMS Collaboration [132]. All slim jets in the presented analysis need to pass the tight ID flag, which ensures an efficiency in high-quality jets of roughly 99 % by rejecting misidentified jets, jets with low reconstruction quality, and fake jets stemming from detector noise. The most prominent selection requirements of the tight jet ID are dependent on the absolute pseudorapidity value of the jet and take the number of different constituents and their relative energy fractions into account.

In addition to passing the tight jet ID flag, a slim jet needs to be classified as tight jet in order to be used for the selection and reconstruction of events. A tight jet needs to have $p_T > 30\,\text{GeV}$ if it is in the central region of the detector ($|\eta| \leq 2.4$), or $p_T > 40\,\text{GeV}$ if it is in the outer region of the detector ($2.4 < |\eta| \leq 4.7$). Only slim jets that are not within a cone of radius $\Delta R = 0.4$ around a tight lepton can be classified as tight jets.

Fat jets are only considered in the analysis if they are in the central region of the detector ($|\eta| \leq 2.0$) and have a transverse momentum of $p_T > 300\,\text{GeV}$. There is no selection requirement similar to the tight ID flag applied on fat jets.

For simulated samples a jet clustering can also be performed without the simulation of the detector effects, taking into account the particles of an event after simulating the hadronization process. It is notable that only visible particles are considered for the clustering process, e.g. neutrinos are excluded. The clustering is based on the same definitions as applied in the Rivet framework and dressed leptons are excluded from this jet clustering on truth level as well as photons. In order to obtain the most suitable truth-level counterpart of jets for the unfolding procedure, these jets are clustered using the anti-k_T algorithm with the two cone radius parameters of $R = 0.4$ and $R = 0.8$. The truth-level jets will in the following be denoted *slim generator jets* ($R = 0.4$) and *fat generator jets* ($R = 0.8$).

Jet Energy Corrections

Reconstructed jets are subject to various effects that affect their energy measurement in an undesired way both in measured data and simulated events. The CMS detector is not completely homogeneous and therefore a different jet energy can be reconstructed for the same particles depending on their flight direction. In addition, the resolution of the detector is only on a finite level and contributions to the jet reconstruction due to pileup need to be taken into account.

To correct for these and further effects, a set of jet energy corrections (JEC) is applied in a factorized approach, where a dedicated correction is taking care of each disturbance [133–135]. Both the slim and the fat jets, which are considered in this thesis, undergo the following corrections:

- **L1 pileup:** The level 1 (L1) pileup correction is the first correction to be applied on both data and simulation and aims at removing contributions from pileup to the jet energy. The strength of the correction is determined by comparing simulated QCD dijet events with and without pileup overlay, and is a function of the energy density, area, transverse momentum, and pseudorapidity of the jet. To correct in addition for residual differences between data and simulation, an η-dependent correction is applied. It is derived with a random-cone method in so-called zero-bias events, that do not have a potential trigger bias.
- **L2L3 MC truth:** The purpose of the L2L3 MC truth correction is to resolve differences between the reconstructed jet p_T and the corresponding generator jet p_T. Simulated QCD dijet events are used to derive the correction factors based on the jet p_T and η.
- **L2L3 residual corrections:** To resolve the remaining differences between data and simulation, the L2L3 residual corrections are applied on data only. Using dijet events with two jets that have similar transverse momentum and one of them being in the barrel region, the L2 residual corrections are determined as a function of the jet pseudorapidity. The absolute L3 residuals are correcting the jet energy depending on the jet p_T. These are obtained from Z+jets, γ+jets and multijets processes.

b Tagging

Due to its short life time, the top quark decays before it can form bound states and this decay results according to the matrix element V_{tb} in almost every case in a W boson and a bottom quark. As it is the case for any final state quark or gluon in an event, such a bottom quark can only be reconstructed from a jet. For the reconstruction of top quarks in an event it is therefore of special interest to distinguish between jets that originate from a bottom quark and jets that originate from gluons or one of the four remaining light quarks (up, down, charm, strange).

Using dedicated algorithms, the identification of such a jet flavor is possible within a certain level of accuracy. The identification of slim jets originating from bottom quarks is crucial for many analyses at the CMS Collaboration and various *b tagging* algorithms have been developed for the different periods of data taking [136, 137]. Many of these algorithms are based on a multivariate approach and exploit the properties of a secondary vertex as one of their input features. If a secondary vertex can be reconstructed within a jet (see Figure 4.5), this is in combination with other jet properties a good indicator that the jet may originate from a bottom quark. While the decay of bottom quarks into top quarks is not allowed, the decay into a charm or an up quark is suppressed due to the matrix elements V_{ub} and V_{cb} respectively. As a result bottom quarks can form bound states with a comparably long life time of around 1.6 ps [9], called B mesons. The distance they travel from the primary vertex before decaying is in the order of a few mm up to 1 cm and leads to the reconstruction of a secondary vertex.

The algorithm that is deployed within this thesis is the DeepJet algorithm [139] and relies on a deep neural network. The performance of this tagger is shown in Figure 4.6 in com-

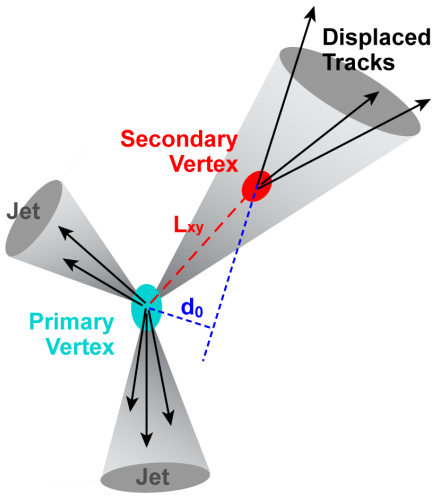

Figure 4.5: Displaced tracks and secondary vertex: Due to their long lifetime, B mesons can travel a distance of up to 1 cm before decaying into different particles. The tracks of the decay products can be assigned to a secondary vertex, which is displaced in respect to the primary vertex of the collision. Taken from Ref. [138].

parison to other b tagging algorithms employed for analyses at the CMS Collaboration. In order to classify a slim jet as stemming from a bottom quark in this thesis, in the following denoted as b jet, its DeepJet classifier value needs to be higher than the year-dependent value of the DeepJet medium working point. A working point (WP) of a b tagging algorithm is in general defined by the probability of misidentifying a light-flavored jet from either an up, down or strange quark or a gluon as b jet. There are commonly three working points of interest, having a misidentification probability of 10 % (loose WP), 1 % (medium WP) and 0.1 % (tight WP). For the presented analysis only the medium WP is applied, which has a b tagging efficiency of around 83 % for the DeepJet algorithm. The values of the medium WP are 0.3093 (2016), 0.3033 (2017) and 0.2770 (2018) [140].

For fat jets it is not investigated in this thesis whether they stem from a bottom quark. A slim generator jet is considered to be a generator b jet if a B hadron with an initial transverse momentum of $p_T > 5$ GeV is included in the jet clustering.

top Tagging

As it will be discussed in detail in Section 5.2, the presented analysis investigates top quarks that can be highly Lorentz-boosted. Due to the high boost of such a top quark, which is transferred to the momentum of its decay particles, these can lie within a cone with relatively small radius size. In case the decay of the top quark is fully hadronic, meaning the W boson decays into two light quarks, the clustering of slim jets does not necessarily yield the expected three slim jets for the bottom quark and the two light quarks but only

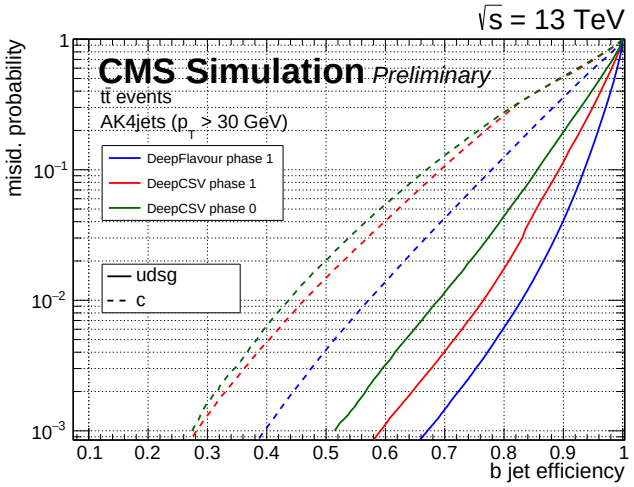

Figure 4.6: b tagging efficiency: The figure shows the performance of the DeepJet tagging algorithm (called DeepFlavour here) in comparison to the DeepCSV tagging algorithm before and after the upgrade of the pixel detector. The DeepJet algorithm yields the lowest misidentification probability at any fixed b jet efficiency value for jets both from light-flavored quarks or gluons and from charm quarks. Taken from Ref. [141].

one or two slim jets. This significantly complicates the reconstruction of top quarks using slim jets, but can be compensated by employing fat jets for the top quark reconstruction.

A fat jet can contain all particles produced in the decay and the subsequent hadronization process of a highly Lorentz-boosted hadronically decaying top quark. In order to distinguish such a top jet from a fat jet prodcued in a QCD multijet event through the collinear splitting of a highly Lorentz-boosted gluon, the DeepAK8 tagging algorithm has been developed by the CMS Collaboration [142]. It is a tagging algorithm based on a deep neural network that takes track and particle information as inputs and classifies from which kind of decay process a fat jet most likely stems. Different specifications of the tagging algorithm are available for analyzers, depending on the desired decay processes to be investigated.

Within this thesis, fat jets are being selected and classified as top jets according to their classifier value of the DeepAK8 mass-decorrelated top vs. QCD tagger and need to pass the year-dependent working point that guarantees a mistagging rate of 2.5 % for fat jets in QCD multijet events. These working points are determined to be 0.177 (2016), 0.117 (2017) and 0.174 (2018) [143] and the performance of the tagging algorithm is illustrated in Figure 4.7. As many analyses apply a selection requirement on the mass of a fat jet, the tagging algorithm is mass-decorrelated in order to avoid a possible bias. On fat generator jets no such selection criterion is applied.

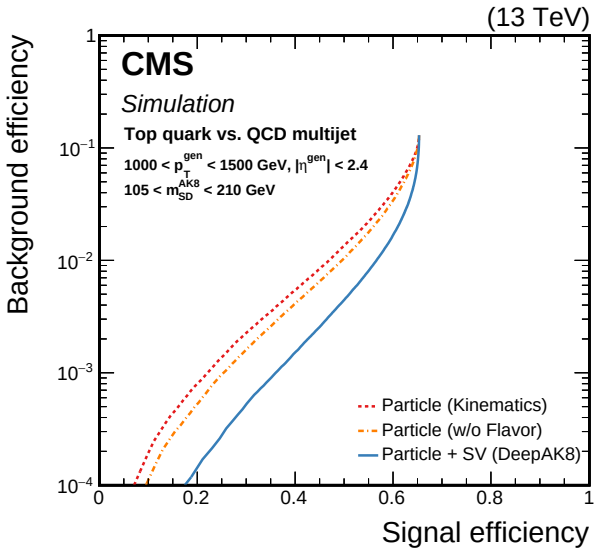

Figure 4.7: Top vs. QCD tagging algorithm performance: The performance of the DeepAK8 mass-decorrelated top vs. QCD tagging algorithm is shown on simulated samples for the 2016 data taking period. In addition the performance of the tagging algorithm with only a subset of the input information is indicated in the plot. Taken from Ref. [142].

4.3.7 Missing Transverse Momentum

The conservation of momentum is one of the most fundamental principles in physics and needs to hold true at particle collisions in high-energy physics. As a result the transverse momenta of all particles produced in a proton-proton collision at the LHC should sum up to roughly zero, as the two initial protons have basically only momentum in longitudinal direction while their momentum in transverse direction is negligible. Nevertheless this is not for every event the case when summing up the transverse momenta of all reconstructed PF particles and one can define the missing transverse momentum vector as [144]:

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} = -\sum_{i=1}^{N_{\mathrm{particles}}} \vec{p}_{\mathrm{T},\,i}\,. \tag{4.6}$$

The magnitude of this vector is denoted as missing transverse momentum $p_{\mathrm{T}}^{\mathrm{miss}}$. There are two main reasons for $p_{\mathrm{T}}^{\mathrm{miss}} > 0$, one being the imperfection of the detector coverage and the smearing of reconstructed particle momenta due to the detector resolution. This in fact has only minor impact on the overall missing transverse momentum, while the main impact is due to particles that leave the detector without being detected. Apart from possible BSM particles, the neutrinos of the SM show this behavior as they are only very weakly

interacting with the detector material. The missing transverse momentum of an event can hence point to neutrinos being produced in the collision and hadronization processes.

As the missing transverse momentum is defined as the vector sum of the PF particles, the corrections that are applied on jets need to be propagated accordingly:

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss,\,corr}} = \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} - \sum_{j=1}^{N_{\mathrm{PF\,jets}}} \left(\vec{p}_{\mathrm{T},j}^{\,\mathrm{corr}} - \vec{p}_{\mathrm{T},j} \right) . \tag{4.7}$$

A set of quality filters is applied on the missing transverse momentum of events in this thesis [145], but these are only of very minor impact for the event selection.

The truth-level counterpart for the missing transverse momentum is defined by the magnitude of the vector sum of the transverse momenta of all neutrinos simulated in the collision and hadronization process.

5. Measurement of the Energy Asymmetry in Top Quark Pair plus Jet Production

The energy asymmetry in $t\bar{t}$ production in association with one additional jet is a successor of previously studied asymmetries in $t\bar{t}$ production. It is an observable that is tailor-made for the LHC with proton-proton collisions at a center-of-mass energy of 13 TeV and predictions for the observable have been given on both parton [29] and particle level [31]. The predictions are given up to NLO accuracy in QCD and deviations from the prediction in the measurement could point to possible BSM contributions in the couplings to top quarks. It has been shown in addition, that within the framework of Standard Model Effective Field Theory (SMEFT), the energy asymmetry is highly sensitive to effective four-quark interactions and resolves insensitive directions in current LHC fits. On particle level the prediction for the energy asymmetry promises to be measurable with a significance of 3σ during Run 3 of the LHC and with a significance of 5σ at the HL-LHC. The possibility for observing deviations from the SM is given already with the data collected in Run 2 of the LHC and thereby enhances the need for the presented analysis. It will in addition enforce the inclusion of the energy asymmetry as one of the important observables in global searches for new physics in the top quark sector [31].

The analysis presented in this thesis focuses on the semileptonic decay of the top quark pair system, which is produced in association with one hard jet. The results are unfolded to truth level and compared to the corresponding predictions in the fiducial phase space. Within this chapter the optimal phase space for the measurement of the energy asymmetry is discussed in Section 5.1, followed by an overview of the signal process topology and the most important background processes in Section 5.2. The selection and reconstruction of events, both on truth and detector level, is explained in Sections 5.3 and 5.4. The corrections that are applied on simulated events are discussed in Section 5.5, while Section 5.6 lists the systematic and statistic uncertainties of the measurement together with a

description of their treatment in the extraction of results. The unfolding procedure from reconstruction to truth level via maximum likelihood estimation is covered in detail in Section 5.7, followed by Section 5.8, which concludes the chapter with the presentation of the final results.

5.1 Energy Asymmetry

The presented measurement will focus purely on the optimized energy asymmetry as introduced in Section 1.2.2:

$$A_E^{\text{opt}}(\theta_j^{\text{opt}}) = \frac{N(\Delta E_{\text{t}\bar{\text{t}}} > 0, \theta_j^{\text{opt}}) - N(\Delta E_{\text{t}\bar{\text{t}}} < 0, \theta_j^{\text{opt}})}{N(\Delta E_{\text{t}\bar{\text{t}}} > 0, \theta_j^{\text{opt}}) + N(\Delta E_{\text{t}\bar{\text{t}}} < 0, \theta_j^{\text{opt}})} \,. \tag{5.1}$$

The strength of the effect strongly depends on the investigated phase space and the magnitude of the observable can be enhanced by dedicated selection requirements on the final-state particles of $\text{t}\bar{\text{t}}j$ production. When determining the corresponding cutoffs it is not only important to maximize the strength of the effect, but also to guarantee the experimental accessibility and a significant cross section in the fiducial phase space. The energy asymmetry is mostly induced by the quark-gluon initial state and increases with the energy difference $\Delta E_{\text{t}\bar{\text{t}}}$ and is therefore largest in phase space regions with boosted top quarks. Requiring boosted top quarks allows in addition for a better reconstruction of the $\text{t}\bar{\text{t}}$ system and the additional hard jet in the event, especially if a high selection criterion on the transverse momentum of the hadronically decaying top quark (t_{had}) is applied in the semileptonic decay channel. In association with criteria on the the leptonically decaying top quark (t_{lep}) and the additional jet (j_{add}), such selection enriches the amount of events with boosted topology. The final-state particles are required to be within the central region of the detector and a good spatial separation is obtained by applying a selection criterion on the ΔR values between them. In total the following requirements for a boosted topology are applied in order to enhance the energy asymmetry:

$$\begin{aligned}
&p_{\text{T}}(\text{t}_{\text{had}}) > 300\,\text{GeV}\,, && |\eta(\text{t}_{\text{had}})| < 2.0\,, && \\
&p_{\text{T}}(\text{t}_{\text{lep}}) > 50\,\text{GeV}\,, && |\eta(\text{t}_{\text{lep}})| < 2.5\,, && \Delta R(\text{t}_{\text{lep}}, \text{t}_{\text{had}}) > 1.5\,, \\
&p_{\text{T}}(j_{\text{add}}) > 100\,\text{GeV}\,, && |\eta(j_{\text{add}})| < 2.5\,, && \Delta R(j_{\text{add}}, \text{t}_{\text{had}}) > 1.5\,.
\end{aligned}$$

This set of conditions is the minimal selection for the fiducial phase space of the analysis and in further stages denoted as the *boosted* phase space. It is applied on all events on detector level, independently of the reconstruction algorithm in use, in order to be considered for the energy asymmetry determination.

For the final extraction of results different phase space regions can be considered by applying criteria on the three observables of importance for the energy asymmetry, being

θ_j^{opt}, $\Delta E_{\text{t}\bar{\text{t}}}$, and $y_{\text{t}\bar{\text{t}}j}$. The most prominent deviation of A_E^{opt} from zero can be obtained in the regions with $0.3\pi < \theta_j^{\text{opt}} < 0.7\pi$. It has already been discussed in Section 1.2.2 that A_E^{opt} is in addition strongly dependent on the absolute value of $\Delta E_{\text{t}\bar{\text{t}}}$ and therefore a requirement of $|\Delta E_{\text{t}\bar{\text{t}}}| > 50\,\text{GeV}$ is applied. A selection criterion on $y_{\text{t}\bar{\text{t}}j}$ can enhance the strength of the asymmetry but significantly reduces the cross section of the fiducial phase space and therefore the results are derived for both a region with $|y_{\text{t}\bar{\text{t}}j}| > 0.5$ and a region without a requirement on $y_{\text{t}\bar{\text{t}}j}$.

5.2 Event Topology

A thorough understanding of the signal process final state and the corresponding signatures in the detector is essential in order to perform the selection and reconstruction of events. In addition, various background processes with the same or a similar final state need to be taken into account as they contribute significantly to the amount of events observed in data and need to be handled accordingly. With the information derived from these considerations, the optimal phase space for the selection and reconstruction of events can be studied in more detail. The simulated samples for the various processes are not discussed within the following sections, but the full list of simulated samples for the three data taking periods can be found in the appendix in Tables A.1, A.2 and A.3.

5.2.1 Signal Process

While the energy asymmetry strongly depends on the initial state of $\text{t}\bar{\text{t}}j$ production, the final state after the decay of the $\text{t}\bar{\text{t}}$ system does not have any impact on the expected outcome on parton level. Nevertheless, the decay channel of the top quark pair, being either dileptonic, semileptonic or fullhadronic, plays an important role for the presented analysis, as an unfolding of the results to truth level is desired. The predictions for the analysis on truth level are given on particle level as well, meaning the $\text{t}\bar{\text{t}}j$ system is reconstructed using truth-level objects like generator jets and dressed leptons. This reconstruction is, as will be discussed in more detail in Section 5.4, a challenging, yet feasible task in the semileptonic decay channel, while the dileptonic and fullhadronic decay channels impose severe problems, due to the presence of two neutrinos in the final state and the high amount of possible jet combinatorics, respectively. In that sense, only $\text{t}\bar{\text{t}}j$ production with semileptonic decay of the $\text{t}\bar{\text{t}}$ system accounts as signal process of the analysis, while events with dileptonic or fullhadronic $\text{t}\bar{\text{t}}$ decay account as background.

A leading order Feynman diagram for $\text{t}\bar{\text{t}}j$ production with qg initial state and semileptonic decay of the $\text{t}\bar{\text{t}}$ system is shown in Figure 5.1. The final state of this process consists of five quarks in total with two of them being bottom quarks and hence providing a high chance of being reconstructed as b jets. The three remaining light quarks lead in general to three additional slim jets, which are not b tagged. In addition, there is one charged lepton in the final state and one neutrino that manifests in the event record as missing transverse momentum. Due to the boosted topology in which the energy asymmetry is studied and the corresponding criteria on the kinematic properties of the hadronically decaying top

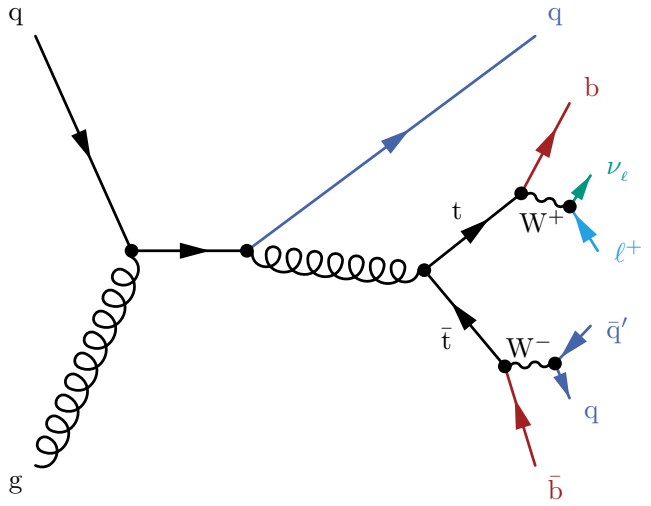

Figure 5.1: Production of $t\bar{t}j$ with a semileptonic decay of the top quark pair system: An exemplary Feynman diagram for born-level $t\bar{t}j$ production with qg initial state is shown in the semileptonic decay channel of the top quark pair. The final state comprises of three light-flavored quarks, two bottom quarks, one charged lepton and the corresponding neutrino.

quark, it is not guaranteed that three resolved jets can be reconstructed for the decay products of this top quark. In fact there is a high chance that these Lorentz-boosted decay products and their hadronization products are bundled within one fat jet, which reduces the expected jet signature from three light-flavored jets and two b jets to one fat jet, one light-flavored jet and one b jet. The probability for this scenario will be studied in more detail in Section 5.4.2, while in Section 5.4 a focus on the reconstruction within these two reconstruction hypotheses is given.

In order to increase the amount of simulated events in the boosted phase space, a dedicated sample with a requirement of 250 GeV on the scalar sum of the transverse momenta of all generator fat jets has been produced for each year of data taking. These samples will in the following be denoted as filtered samples and will be in use with the samples produced centrally by the CMS Collaboration (unfiltered samples), which cover the entire phase space. An orthogonal selection will be applied on these samples in order to avoid double-counting of events while still maximizing the amount of simulated events to be used in the analysis in each selection.

5.2.2 Background Processes

There is a huge variety of possible background processes, especially in higher order of QCD, which can mimic the same final state in the detector as the production of $t\bar{t}j$ with a semileptonic $t\bar{t}$ decay. The most prominent groups of processes will be described here and representative Feynman diagrams are given in Figure 5.2. In later analysis stages the background processes will be treated bundled for the presentation of results according to

the group definitions below. Even though some of the described processes do not have the same final state as the semileptonic decay channel of $t\bar{t}j$ production, they can pass the applied selection criteria due to the following effects:

- A charged lepton is produced in the hadronization process of a final-state quark and hence the quark is misreconstructed as charged lepton.
- A charged lepton is misreconstructed as jet.
- A jet stemming from a light-flavored quark is classified to be a b jet and vice versa.
- Due to imperfections in the detector resolution a high value for $p_{\mathrm{T}}^{\mathrm{miss}}$ is reconstructed despite the absence of neutrinos in the final state of the event.
- A particle leaves the detector without being fully absorbed and causes a high value for $p_{\mathrm{T}}^{\mathrm{miss}}$.

Despite the rather low probability of the described effects, they can be of importance due to the high cross section of some of the background processes.

Top Quark Pair Production with Dileptonic and Fullhadronic Decay

The production of a top quark pair with the subsequent decays of the W bosons being either both leptonic or both hadronic, is the dominant background process in the analysis. The dileptonic decay channel is especially important if an additional gluon is emitted, which can split up into two quarks and hence cause additional jets in the event. The event then consists of two b jets, two light-flavored jets, two charged leptons and missing transverse momentum. The fullhadronic decay channel has already at tree level two b jets and four light-flavored jets in the final state. The branching ratio of these two decay channels is in combination higher than the signal process branching ratio but their contributions can mainly be damped by the imposed quality criteria on the reconstructed jets and leptons. The combination of the two decay channels is in the following denoted as "$t\bar{t}$ DL + FH".

Single Top Quark Production

The production of a single top quark is another important background process in the analysis as it can have the same final state as the signal process in case of an additional gluon emission. Within this thesis only the two dominant single top quark production processes at the LHC, being the t-channel production and the tW associated production, are considered as background process. The contribution of s-channel single top quark production is negligible as the cross section of this production channel is a factor of 20 times smaller than that of the t-channel. The modeling of t-channel single top quark production is within the four-flavor-scheme (4FS), while the tW associated production is modeled in the five-flavor-scheme (5FS). In the 4FS scheme bottom quarks cannot directly stem from the valence or sea quarks of the proton but need to be produced via gluon splitting, while in the 5FS they can be part of the initial-state partons. Single top quark production processes are in the following bundled as "ST".

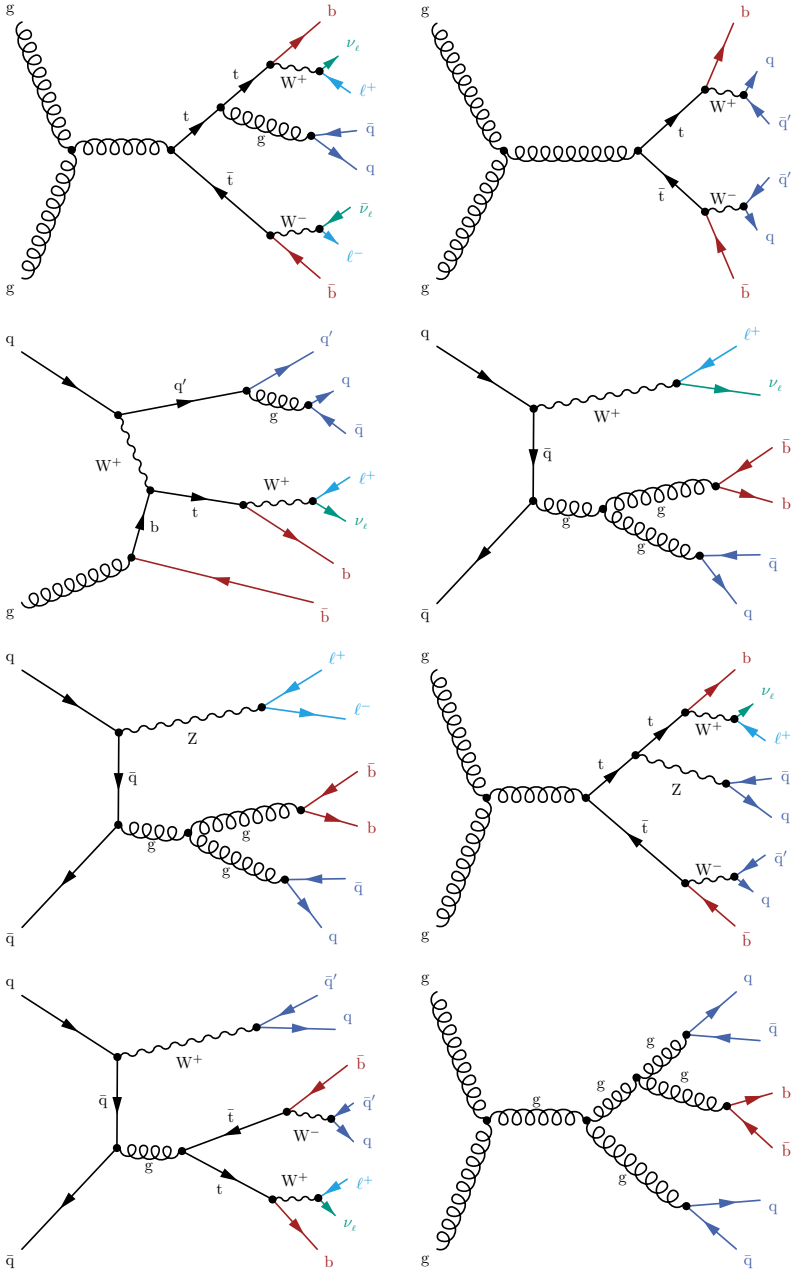

Figure 5.2: Representative Feynman diagrams for different background processes: Examples of all background process groups with the same or a similar final state as the signal process are shown. In the upper row $t\bar{t}$ production is shown at NLO with dileptonic decay (left) and at LO with fullhadronic decay (right), followed in the second row by NLO t-channel single top quark production (left) and NLO W+jets production (right). In the third row the production of a Z boson in association with jets at NLO (left) and with a top quark pair at LO (right) is shown. The processes in the lowest row are the production of $t\bar{t}$ in association with a W boson (left) and QCD multijet production (right).

Production of W and Z Bosons in Association with Jets

There are in total five different processes of electroweak W and Z boson production in association with jets produced via strong interaction that are considered as background in the analysis. The production of a single W boson (W+jets) or Z boson (Z+jets) with jets at NLO can mimic the signal process in case of a leptonic W/Z decay. The three combinations of two heavy vector bosons produced together (WW, WZ, and ZZ) in association with jets can have a similar final state as the signal process in case of a semileptonic decay of the vector boson pair. The group of these processes is called "V(V) + jets"

Top Quark Pair Production in Association with a W or Z Boson

The production of a top quark pair in association with either a W boson or a Z boson comprises in total three heavy vector bosons, of which each can decay either leptonically or hadronically. A final state very similar to the signal process is obtained in case exactly one of these bosons decays leptonically. As the additional vector boson in the event can only be produced via electroweak interaction, the cross section of such processes is comparably low in respect to pure $t\bar{t}$ production. Top quark pair production in association with a W or Z boson will in the following be named "$t\bar{t}V$ ".

QCD Multijet Production

The background process with the smallest impact for the analysis is the production of QCD multijet events. Even though the cross section of these processes is extremely high and they can mimic various final states due to the high jet multiplicity, they can be filtered out quite effectively by the quality criteria on tight leptons and a low cutoff on $p_{\mathrm{T}}^{\mathrm{miss}}$. Nevertheless, they have a small contribution that needs to be taken into account and is in the following denoted as "QCD".

5.3 Event Selection

The main purpose of an event selection is to obtain a signal-enriched phase space and to improve the signal-to-background ratio within the selected phase space. Applying the event selection at an early stage of the analysis in addition significantly reduces the required computing power for further processing steps like the reconstruction of top quark kinematic properties. Due to the boosted phase space, which is investigated for the measurement of the energy asymmetry, there are in principle two different collections of final-state objects possible for the signal process as discussed in Section 5.2.1. In case the decay products of the hadronically decaying top quark can be resolved in a slim jet each, the event contains in total three light-flavored jets, two b jets, one charged lepton and missing transverse momentum due to the neutrino. If the resolved reconstruction of the slim jets is not possible and all products of the hadronically decaying top quark are bundled within one fat jet, the event comprises only one light-flavored jet, one b jet, one top jet, one charged lepton and missing transverse momentum. Due to this different conditions,

Table 5.1: HLT Trigger paths. The paths of the HLT Trigger that are used for events with one muon or one electron.

Lepton flavor	Trigger path
	2016
μ	HLT_IsoTkMu24
	HLT_IsoMu24
e	HLT_Ele32_eta2p1_WPTight_Gsf
	2017
μ	HLT_IsoMu27
e	HLT_Ele35_WPTight_Gsf
	HLT_Ele30_eta2p1_WPTight_Gsf_CentralPFJet35_EleCleaned
	2018
μ	HLT_IsoMu24
e	HLT_Ele32_WPTight_Gsf
	HLT_Ele30_eta2p1_WPTight_Gsf_CentralPFJet35_EleCleaned

no final event selection is performed at the early stage of event processing, but only a minimum set of requirements is applied, which is valid for both scenarios described here.

The data sets that are analyzed in this thesis are listed in Tables A.4, A.5, and A.6 in the appendix and contain only events that pass specific HLT paths for electrons or muons. In a first step of the event selection, events both in simulation and in measured data are required to pass at least one of the applied HLT paths. The HLT paths differ for each year of data taking and are listed in Table 5.1. For the 2016 data taking periods these paths are selecting events that contain at least one electron or muon with a low cutoff on its transverse momentum and an additional cutoff on the absolute pseudorapidity for electrons. For the 2017 and 2018 data taking periods the electron HLT path has been extended in a way that the p_T cutoff on the electron is lowered if a jet with $p_T > 35\,\text{GeV}$ is present in the event.

In addition to applying the HLT paths, an offline requirement on the leptons in an event is avoiding possible double-counting of events that pass both the electron and the muon HLT paths. To be considered in the analysis, an event must contain exactly one tight electron or exactly one tight muon and no additional veto lepton. A veto lepton can be either an electron or a muon and has the same selection criteria applied as tight electrons and muons respectively, with the only difference being the cutoff on the transverse momentum. This is lowered to $p_T > 20\,\text{GeV}$ for both electrons and muons in order to account as a veto lepton. This cutoff mainly reduces the contributions from $t\bar{t}$ production in the dileptonic decay channel and further processes with two charged leptons in the final state. To account for the neutrino in the final state of the signal process, a requirement on the missing transverse momentum of $p_T^{\text{miss}} > 40\,\text{GeV}$ is applied, which has a strong impact on the reduction of fullhadronic $t\bar{t}$ production and QCD multijet production.

The boosted phase space contains at least one slim jet with $p_{\mathrm{T}} > 100\,\mathrm{GeV}$ and $|\eta| < 2.5$, which is not a b jet, to account for the additional jet in $\mathrm{t\bar{t}}j$ production. Each event is required to contain such a jet in order to pass the event selection, as it can already prior to the reconstruction of the $\mathrm{t\bar{t}}j$ system be guaranteed that it would not pass the asymmetry-enhancing selections of the boosted phase space without the presence of such a jet.

At this stage of event processing there are no further selection requirements applied and each event that passes the selection is in the subsequent step tested for a possible reconstruction under the hypothesis of being a signal event.

5.4 Event Reconstruction

The observables that enter the calculation of the energy asymmetry, being θ_j^{opt}, $y_{\mathrm{t\bar{t}}j}$, and $\Delta E_{\mathrm{t\bar{t}}}$, are all strongly dependent on the kinematic properties of the top quark pair and the additional jet in $\mathrm{t\bar{t}}j$ production. It is therefore obvious that the correct reconstruction of the final-state particles to the $\mathrm{t\bar{t}}j$ system is of main importance both on truth level and on detector level. While misreconstruction of events on truth level leads to the prediction of worthless results without a physics meaning, misreconstruction on detector level reduces the achievable precision of the measurement and complicates the unfolding procedure. Hence, thorough studies have been performed in this thesis in order to optimize the reconstruction on both levels discussed.

Within the following sections, the common reconstruction of leptonically decaying W bosons will be presented, followed by a discussion on the phase space to be employed for reconstruction, which slightly differs from the definitions made in theory predictions [31]. Further, the different reconstruction algorithms applied on truth level and on detector level are presented. Subsequent to the reconstruction on detector level, the fiducial phase space of the analysis can be defined and the event yields, as well as the predicted results in this phase space, can be determined.

5.4.1 W Boson Reconstruction

Independent of whether the reconstruction of the $\mathrm{t\bar{t}}$ system is performed on truth level or on detector level, there is in principle only three different kinds of objects that can be employed for the reconstruction, being jets, leptons, and missing transverse momentum. The semileptonic decay channel of the top quark pair system is characterized by one leptonically decaying W boson, which manifests itself in exactly one charged lepton and one neutrino in the final state of the event. The charged lepton can be, according to the event selection, an electron or a muon, stemming either directly from the decay of the W boson or indirectly from the subsequent leptonic decay of a tau lepton from the W boson decay. The four-momentum of the W boson can be constructed by the four-momenta of the charged lepton and the neutrino, with the drawback that only the transverse component of the neutrino momentum can be estimated by $p_{\mathrm{T}}^{\mathrm{miss}}$, while its longitudinal component remains unknown. This can be compensated by introducing an additional constraint and fixing the mass of the W boson to its literature value of $m_{\mathrm{W}} = 80.385\,\mathrm{GeV}$ [9]. With the

assumption, that the x and y components of the neutrino momentum are given by the missing transverse momentum, the W boson mass can thus be written as:

$$m_{\mathrm{W}}^2 = 2 \cdot \left(E_\ell \sqrt{{p_{\mathrm{T}}^{\mathrm{miss}}}^2 + p_{z,\nu}^2} - p_{\mathrm{T},\ell} p_{\mathrm{T}}^{\mathrm{miss}} \cos\Delta\phi - p_{z,\ell} p_{z,\nu} \right) = (80.385\,\mathrm{GeV})^2 \,. \tag{5.2}$$

Here the energy of the lepton is given as $E_\ell = \sqrt{p_{\mathrm{T},\ell}^2 + p_{z,\ell}^2}$ and the azimuthal angle between the charged lepton ℓ and $p_{\mathrm{T}}^{\mathrm{miss}}$ is denoted as $\Delta\phi$. This equation can be refactored and solved for the unknown longitudinal component of the neutrino momentum:

$$p_{z,\nu} = \frac{\Lambda p_{z,\ell}}{p_{\mathrm{T},\ell}^2} \pm \sqrt{\frac{\Lambda^2 p_{z,\ell}^2}{p_{\mathrm{T},\ell}^4} - \frac{E_\ell^2 \left(p_{\mathrm{T}}^{\mathrm{miss}}\right)^2 - \Lambda^2}{p_{\mathrm{T},\ell}^2}}, \tag{5.3}$$

with

$$\Lambda = \frac{m_{\mathrm{W}}^2}{2} + p_{\mathrm{T},\ell} p_{\mathrm{T}}^{\mathrm{miss}} \cos\Delta\phi. \tag{5.4}$$

For the quadratic equation two different cases are possible, depending on the value of the discriminant. In case of a positive value, two different solutions are obtained for $p_{z,\nu}$ and the one with the smaller absolute value is chosen. The reconstruction of $p_{\mathrm{T}}^{\mathrm{miss}}$ underlies imperfections and the discriminant can thus become negative, leading to complex solutions for $p_{z,\nu}$. In this case exactly one real solution can be obtained by varying the x and y component of $p_{\mathrm{T}}^{\mathrm{miss}}$ in a way that the radicand becomes zero [146].

This reconstruction procedure for the leptonically decaying W boson is applied on both truth and detector level. On detector level it is guaranteed that the event contains exactly one tight lepton that accounts for the charged lepton of the reconstruction algorithm. This is not ensured on truth level and the algorithm can thus only be applied in case of exactly one dressed lepton in the event. In case the number of dressed leptons is unequal to one, parton information of the event record will be employed for the reconstruction as will be further discussed in the following sections.

5.4.2 Reconstruction Phase Space

The predictions for the energy asymmetry on particle level [31] are obtained for a fiducial phase space with at least one fat generator jet in the event. The clustering of this fat generator jet has been performed with a cone radius parameter of $R = 1.0$, which differs from the value of $R = 0.8$, which is commonly used in analyses of the CMS Collaboration and accordingly also in this thesis. As all decay products of the hadronically decaying top

quark are estimated to be collimated within this one fat generator jet and the kinematic properties of this jet are directly assigned to the respective top quark, this difference in jet clustering has an high impact on the analysis. This has been studied using the RIVET framework and indicated the need for a separate prediction using fat generator jets clustered with $R = 0.8$.

The energy asymmetry is mostly present in a boosted topology with highly Lorentz-boosted top quarks and the corresponding phase space can be reached by a cutoff on the transverse momenta of t_{had} and the additional jet (j_{add}). While enhancing the magnitude of the observable effect of the energy asymmetry, these cutoffs on the other hand also significantly reduce the cross section in the fiducial phase space and lead to a huge statistical uncertainty of the measurement with the available LHC Run 2 data. Within this thesis it has therefore been studied whether an extension of the fiducial phase space can still yield to a magnitude of the energy asymmetry that differs significantly from zero while increasing the sensitivity at the same time. In order to justify such a phase space extension, the unfiltered semileptonic $t\bar{t}$ samples have been tested for the coexistence and the overlay of a hadronically decaying top quark with $p_T > 300\,\text{GeV}$ and a fat generator jet with $p_T > 300\,\text{GeV}$. Prior to this test the samples have been filtered with the HLT paths of the analysis as these impose the minimal set of applied requirements, which cannot be removed or modified offline. The results of this study are given in Tables 5.2 and 5.3. The same sets of requirements have been applied in both tables, while the order differs between them. As the same set of events is obtained after the first three requirements in each permutation, the subsequent requirements on $\Delta R(t_{had}, j_{AK8})$ are shown only once.

The first permutation of requirements can be summarized in a way that about 72.5 % of the events in the unfiltered semileptonic $t\bar{t}$ simulation sample with $p_T(t_{had}) > 300\,\text{GeV}$ comprise a fat generator jet with $p_T > 300\,\text{GeV}$. In association with the studies on $\Delta R(t_{had}, j_{AK8})$ this means that in only about 56.4 % of events with $p_T(t_{had}) > 300\,\text{GeV}$, the decay products of the hadronically decaying top quark are collimated within the selected fat generator jet of the event. The second studied permutation of the same requirements shows, that even though a fat generator jet with $p_T > 300\,\text{GeV}$ is present in the event, only in about 59.1 % the hadronically decaying top quark is boosted with $p_T > 300\,\text{GeV}$.

In sum these studies show that the pure selection of a fat generator jet with $p_T > 300\,\text{GeV}$ does not guarantee to obtain a phase space with the desired boosted topology and allow for an extension of the phase space to also consider events without the presence of a fat generator jet. Accordingly in the following reconstruction algorithms an alternative approach for the event reconstruction with slim jets will be exploited both on truth level and on detector level. It should be mentioned though, that the primary focus is still on the reconstruction employing a fat jet as proposed in Ref. [31] and the alternative reconstruction is only tested for events that fail this reconstruction approach. The alternative reconstruction approach in that sense corresponds to an extension and not a redefinition of the fiducial phase space.

Within all presented reconstruction methods, the primary focus is always set on the reconstruction of the $t\bar{t}$ system. After the successful reconstruction of $t\bar{t}$, the remaining jets

Table 5.2: Studies on the hadronically decaying top quark on parton level. The table shows the result of the studies on the hadronically decaying top quark on parton level and the presence of a corresponding fat generator jet (j_{AK8}) in the unfiltered semileptonic $\mathrm{t\bar{t}}$ simulation sample. Shown are the numbers of events that pass specific requirements. The listed requirements are applied sequentially from top to bottom and the quoted efficiencies are in relation to the amount of events that pass the previous requirement (in relation to the total number of simulated events).

Requirement	**2016**	**2017**	**2018**	**all**
None	107 305 100	110 014 744	101 550 000	318 869 844
HLT paths	38 273 676 35.66 % (35.66 %)	40 579 357 36.89 % (36.89 %)	38 780 020 38.18 % (38.18 %)	117 633 053 36.89 % (36.89 %)
$p_{\mathrm{T}}(\mathrm{t_{had}}) > 300\,\mathrm{GeV}$	1 576 142 4.12 % (1.47 %)	1 667 581 4.11 % (1.52 %)	1 572 195 4.05 % (1.55 %)	4 815 918 4.09 % (1.51 %)
$N(j_{\mathrm{AK8}}) > 0$	1 576 131 100.00 % (1.47 %)	1 667 570 100.00 % (1.52 %)	1 572 188 100.00 % (1.55 %)	4 815 889 100.00 % (1.51 %)
$p_{\mathrm{T}}(j_{\mathrm{AK8}}) > 300\,\mathrm{GeV}$	1 143 079 72.52 % (1.07 %)	1 208 829 72.49 % (1.10 %)	1 140 149 72.52 % (1.12 %)	3 492 057 72.51 % (1.12 %)
$\Delta R(\mathrm{t_{had}}, j_{\mathrm{AK8}}) < 1.2$	890 588 77.91 % (0.83 %)	940 143 77.77 % (0.85 %)	886 373 77.74 % (0.87 %)	2 717 104 77.81 % (0.85 %)
$\Delta R(\mathrm{t_{had}}, j_{\mathrm{AK8}}) < 0.8$	890 046 99.94 % (0.83 %)	939 609 99.94 % (0.85 %)	885 850 99.94 % (0.87 %)	2 715 505 99.94 % (0.85 %)
$\Delta R(\mathrm{t_{had}}, j_{\mathrm{AK8}}) < 0.4$	887 702 99.74 % (0.83 %)	937 141 99.74 % (0.85 %)	883 472 99.73 % (0.87 %)	2 708 315 99.74 % (0.85 %)

of the event are tested to account for the additional jet in $\mathrm{t\bar{t}}j$ production. The objects used for the reconstruction are defined in Section 4.3.

5.4.3 Event Reconstruction on Truth Level

The reconstruction of an event is being separated into the reconstruction in the boosted event regime and the resolved event regime. If an event can be reconstructed in the boosted regime, this reconstruction is taken as to be the correct one. If it cannot be reconstructed in the boosted regime it is attempted to reconstruct the event in the resolved regime and the resolved reconstruction will be considered the correct one. In case an event can be neither reconstructed in the boosted regime nor in the resolved regime, the observables required for the unfolding procedure will be determined under the employment of the parton level information. This ensures that a truth-level definition exists for each event of the signal process.

In order to be tested for a possible reconstruction in the boosted or resolved regime, an event needs to fulfill the following common requirements between both regimes:

Table 5.3: Studies on the presence of fat generator jet. The table shows the result of the studies on the presence of fat generator jet (j_{AK8}) and the presence of a corresponding hadronically decaying top quark on parton level in the unfiltered semileptonic $\mathrm{t\bar{t}}$ sample. Shown are the numbers of events that pass specific requirements. The listed requirements are applied sequentially from top to bottom and the quoted efficiencies are in relation to the amount of events that pass the previous requirement (in relation to the total number of simulated events).

Requirement	2016	2017	2018	all
None	107 305 100	110 014 744	101 550 000	318 869 844
HLT paths	38 273 676 35.66 % (35.66 %)	40 579 357 36.89 % (36.89 %)	38 780 020 38.18 % (38.18 %)	117 633 053 36.89 % (36.89 %)
$N(j_{\mathrm{AK8}}) > 0$	27 065 208 70.71 % (25.22 %)	28 738 986 70.82 % (26.12 %)	27 352 424 70.53 % (26.93 %)	83 156 618 70.69 % (26.10 %)
$p_{\mathrm{T}}(j_{\mathrm{AK8}}) > 300\,\mathrm{GeV}$	1 929 049 7.13 % (1.80 %)	2 046 834 7.12 % (1.90 %)	1 929 275 7.05 % (1.90 %)	5 905 158 7.10 % (1.85 %)
$p_{\mathrm{T}}(\mathrm{t_{had}}) > 300\,\mathrm{GeV}$	1 143 079 59.26 % (1.07 %)	1 208 829 59.06 % (1.10 %)	1 140 149 59.10 % (1.12 %)	3 492 057 59.14 % (1.12 %)

- The event contains exactly one dressed lepton (ℓ) with $p_{\mathrm{T}} > 27\,\mathrm{GeV}$ and $|\eta| < 2.5$.
- Missing transverse momentum $p_{\mathrm{T}}^{\mathrm{miss}} > 20\,\mathrm{GeV}$.
- It is possible to reconstruct the leptonically decaying W boson using the procedure as described in Section 5.4.1 utilizing the dressed lepton and the missing transverse momentum.
- The transverse mass of the reconstructed W boson $m_{\mathrm{T}}^{\mathrm{W}}$ and $p_{\mathrm{T}}^{\mathrm{miss}}$ sum up to more than 60 GeV.

Reconstruction in the Boosted Regime

The reconstruction of events in the boosted regime follows the description in Ref. [31] as closely as possible. In order to be reconstructed in the boosted regime, an event needs to have at least one fat generator jet (j_{AK8}) with $p_{\mathrm{T}}(j_{\mathrm{AK8}}) > 300\,\mathrm{GeV}$ and $120\,\mathrm{GeV} < m(j_{\mathrm{AK8}}) < 220\,\mathrm{GeV}$, which is separated from the dressed lepton by requiring $\Delta\phi(j_{\mathrm{AK8}}, \ell) > 1.0$. This jet is assumed to contain all decay products of the hadronically decaying top quark and will therefore be assigned to $\mathrm{t_{had}}$.

The event is now tested for a slim generator jet with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 2.5$ to account for the bottom quark of the leptonically decaying top quark $\mathrm{t_{lep}}$. This jet ($j_{\mathrm{b,lep}}$) needs to be within a cone of $\Delta R(j_{\mathrm{b,lep}}, \ell) < 2.0$ with the dressed lepton and separated by $\Delta R(j_{\mathrm{AK8}}, j_{\mathrm{b,lep}}) > 1.5$ from the selected fat generator jet. In case there are multiple jets fulfilling these requirements the b jet with the highest p_{T} is chosen. In case none of the jets is a b jet, the jet with the highest p_{T} is chosen.

The remaining slim jet with the highest p_T, fulfilling $p_T > 100\,\text{GeV}$ and $|\eta| < 2.5$, is selected to be the additional jet j_{add} of the event. This jet in addition needs to be separated from j_{AK8} and ℓ by requiring $\Delta R(j_{\text{add}}, j_{\text{AK8}}) > 1.5$ and $\Delta R(j_{\text{add}}, \ell) > 0.4$.

If any of the described particles is not matchable it will be tried to reconstruct the event in the resolved regime.

Reconstruction in the Resolved Regime

In the resolved reconstruction of an event no fat generator jet is taken into account for the reconstruction, but only slim generator jets and the leptonically decaying W boson. In order to be reconstructable in the resolved regime an event needs to have at least five slim generator jets, of which at least two need to be b jets. The jets in the event are now being assigned to the final-state partons of the top quark pair decay. These final-state partons are the bottom quark b_{lep} from the leptonic decay of t_{lep}, the bottom quark b_{had} from the decay of t_{had} and the two light quarks q_1 and q_2 from the decay of the hadronically decaying W boson from t_{had}.

Each possible permutation of jets to final-state partons is now investigated with the only restriction that only b jets may be assigned to b_{lep} and b_{had}. The permutation with the minimal sum of ΔR values is chosen to be the correct assignment. The sum of ΔR values is defined as:

$$\Delta R_{\text{sum}}(j_a, j_b, j_c, j_d) = \Delta R(\text{b}_{\text{lep}}, j_a) + \Delta R(\text{b}_{\text{had}}, j_b) + \Delta R(\text{q}_1, j_c) + \Delta R(\text{q}_2, j_d)\,, \tag{5.5}$$

where the indices a, b, c, d range from one to the number of jets and are not allowed to be identical. The leptonically decaying W boson is used together with the jets of the correct assignment to reconstruct t_{had} and t_{lep}.

Following up this assignment, the hardest remaining jet with $p_T > 100\,\text{GeV}$ and $|\eta| < 2.5$ is selected to be the additional jet j_{add} of the event. This jet in addition needs to be separated from t_{had} and ℓ by requiring $\Delta R(j_{\text{add}}, \text{t}_{\text{had}}) > 1.5$ and $\Delta R(j_{\text{add}}, \ell) > 0.4$. If no jet is fulfilling these requirements the event is not reconstructable in the resolved regime.

It is notable here that this reconstruction definition differs from the pseudo top recommendations of the LHC Top Working Group [147]. The definition as described in Ref. [147] is optimized for the production of $\text{t}\bar{\text{t}}$ without an additional high-p_T jet in the event and uses the four jets with highest p_T in the event for the reconstruction of the $\text{t}\bar{\text{t}}$ system. To pass the reconstruction requirements as described in this chapter using the pseudo top recommendations, an event would need to have at least five jets with $p_T > 100\,\text{GeV}$, which is highly unlikely and not eligible for the investigated phase space.

Reconstruction using Parton Information

In case the reconstruction on truth level is possible neither in the boosted regime nor in the resolved regime, the parton information of the event record will be employed for the

reconstruction of the $t\bar{t}$ system. As the analysis aims at unfolding the results to particle level, the described procedure is not desirable but still unavoidable in order to reach a reasonable signal-to-background ratio in the fiducial phase space. If the parton information was not employed, events in the fiducial phase space without possible reconstruction on truth level in the boosted or resolved regime would need to be accounted as background and would further reduce the reachable precision of the measurement.

Using the parton information of the event, the top quark and antiquark are defined and according to their decay information they are assigned to t_{had} and t_{lep}. Hence, the slim and fat generator jets are not used for the reconstruction of the $t\bar{t}$ system. As the leptonically decaying W boson does not enter the reconstruction, dressed leptons and missing transverse momentum of the event are also not considered for the $t\bar{t}$ reconstruction. The additional jet of the event is being searched for in a sequential procedure using the search patterns below and considering only slim generator jets. In case a matching slim generator jet for j_{add} is found within one of the processing steps, the subsequent steps are not processed. In case multiple slim generator jets in an event fulfill the same requirement, the jet with the highest transverse momentum is chosen for j_{add}. The following order of selection steps is applied on each jet j to find j_{add}:

- $|\eta(j)| < 2.5, \quad p_T(j) > 100\,\text{GeV}, \quad \Delta R(t_{had}, j) > 1.2, \quad \Delta R(t_{lep}, j) > 1.2$
- $|\eta(j)| < 2.5, \quad p_T(j) > 100\,\text{GeV}, \quad \Delta R(t_{had}, j) > 0.8, \quad \Delta R(t_{lep}, j) > 0.8$
- $\Delta R(t_{had}, j) > 1.2, \quad \Delta R(t_{lep}, j) > 1.2$
- $\Delta R(t_{had}, j) > 0.8, \quad \Delta R(t_{lep}, j) > 0.8$
- $\Delta R(t_{had}, j) > 0.4, \quad \Delta R(t_{lep}, j) > 0.4$

If an event has no slim generator jet fulfilling any of the requirements above, the slim generator jet with the highest transverse momentum is assigned to j_{add}. In case an event has no slim generator jet, the event will be discarded. This has not been observed for any event passing the preselection as defined in Section 5.3.

Figure 5.3 shows the distribution of the reconstructed masses of the top quark and antiquark and the transverse momentum of the additional jet for the unfiltered and the filtered signal samples. Approximately one third of the events is reconstructed in each the boosted regime, the resolved regime or using parton information. The distributions of the shown kinematic properties are similar for the boosted and the resolved reconstruction, while they differ for the reconstruction with parton information. Using parton information, a narrow peak for the reconstructed top quark masses at the literature value is observed and the transverse momentum of the additional jet is allowed to reach values lower than 100 GeV. The same distributions for each truth-level reconstruction algorithm separately can be found in Figures C.1, C.2 and C.3 in the appendix.

The fraction of roughly one third of the events being not reconstructable on truth level in the boosted or resolved regime emphasizes the need for using the parton information as additional reconstruction algorithm. Using parton information for events that have no

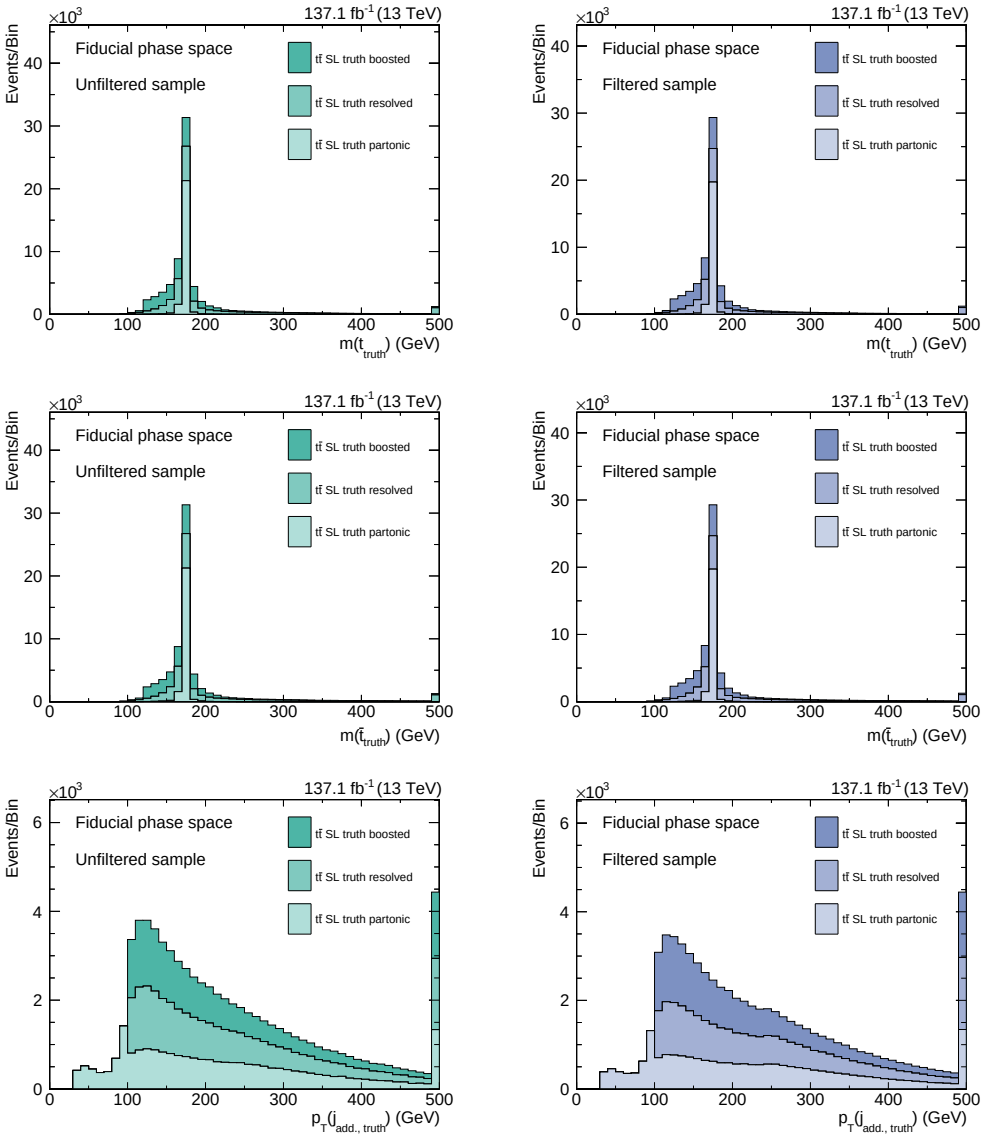

Figure 5.3: Event reconstruction on truth level: The mass of the reconstructed top quark, top antiquark, and the transverse momentum of the reconstructed additional jet are shown on truth level for the fiducial phase space. The distributions in the left column are obtained by utilizing exclusively the unfiltered signal sample, while the distributions in the right column use the filtered signal sample only. The distributions are splitted by the respective reconstruction algorithm as described in the main text. The total amount of events is lower for the distributions of the filtered sample, as this sample does not provide a full coverage of the fiducial phase space (see Section 5.4.5).

truth-level reconstruction in the boosted or resolved regime may impose the question, why the same procedure is not applied for events of $t\bar{t}j$ production with dileptonic or fullhadronic decay of $t\bar{t}$. This is not considered in the analysis as these decay channels do not contribute to the signal process and are guaranteed to obtain a wrong reconstruction on detector level.

5.4.4 Event Reconstruction on Detector Level

Even though the final state of the analysis contains a high-p_T slim jet in addition to the $t\bar{t}$ system, the reconstruction focuses primarily on the $t\bar{t}$ system in order to facilitate and optimize the reconstruction algorithms. The additional slim jet will be selected after the reconstruction of the $t\bar{t}$ system from the remaining slim jets in the event. As the energy asymmetry observable is optimized for a boosted phase space, an attempt is taken for each event to be reconstructed via a boosted reconstruction algorithm. In case a boosted reconstruction is not possible, an event will be reconstructed in the resolved regime without the use of fat jets. Events that can be reconstructed successfully neither in the boosted nor in the resolved regime will be discarded for the further analysis.

Event Reconstruction in the Boosted Regime

An event can only be reconstructed in the boosted regime if a fat jet j_{AK8} is present with $p_T > 300\,\text{GeV}$, $|\eta| < 2.0$ and $120\,\text{GeV} > m_{SD} > 220\,\text{GeV}$ where m_{SD} is the soft-drop corrected jet mass [148]. This fat jet needs to be a top jet and to be well separated from the tight lepton of the event by requiring $\Delta\phi(j_{AK8}, \ell) > 1.0$. In case multiple fat jets fulfill these requirements, the fat jet with the highest transverse momentum is selected. This fat jet is considered to contain all decay products of the hadronically decaying top quark and will therefore be assigned to t_{had}.

In order to reconstruct the leptonically decaying top quark the slim jet from the bottom quark of the decay ($j_{b,lep}$) needs to be determined. All tight jets are considered that have $|\eta| < 2.5$ and lie within $\Delta R(\ell, j_{b,lep}) < 2.0$ and $\Delta R(t_{had}, j_{b,lep}) > 1.5$. The b jet with the highest transverse momentum is chosen for the reconstruction. In the case that none of the above described jets is a b jet, the jet with the highest transverse momentum is chosen. The leptonically decaying top quark is reconstructed by the vectorial sum of the leptonically decaying W boson and $j_{b,lep}$. The charge of the tight lepton defines whether top quark or antiquark decayed leptonically and hence the final assignment. In case an event cannot be reconstructed in the boosted regime via the described algorithm, it is attempted to reconstruct the event in the resolved regime without using fat jets.

Event Reconstruction in the Resolved Regime

The reconstruction in the resolved event regime is using only the leptonically decaying W boson and the slim jets of the event for the reconstruction of the $t\bar{t}$ system. The basic idea of the resolved reconstruction is to assign four slim jets to the final-state quarks of the event, which are the two bottom quarks from the top quark and antiquark decays (b_t and $b_{\bar{t}}$) and the two light-flavored quarks from the hadronically decaying W boson (q_1 and

Table 5.4: Setup of the BDT used for the jet assignment. Various configuration settings and different parameters have been tested for the setup, with the listed values showing the be best performance in average. A separate training is performed for each year.

Parameter	Value
NTrees	450
MaxDepth	8
MinNodeSize	1 %
BoostType	AdaBoost
AdaBoostBeta	0.3
SeparationType	GiniIndex
nCuts	20
PruneMethod	NoPruning

q_2). This assignment is performed by training, testing and evaluating a dedicated BDT, implemented using the TMVA interface of ROOT with a commom setup for all processed years as listed in Table 5.4.

A separate training is performed for each year and in order to avoid the loss of simulated events for evaluation due to the training and testing process of the BDT a further split into two orthogonal sets is performed respectively. This makes in total six separate BDTs that are trained independent of each other. Both in data and simulation events are split up by their event number into sets of events with even and odd event number to employ a two-fold approach in the BDT training, which allows to use all simulated events in the analysis without the necessity to discard a subset. The event number is uncorrelated to any other property of interest of an event. By training and testing a BDT on one of these sets in simulation and evaluating it exclusively on the orthogonal set both in simulation and data it can be ensured that no simulated event has to be discarded in the analysis.

For the resolved reconstruction each possible assignment of the jets of the event to the four final-state particles is investigated with the restriction that only b jets can be assigned to the two bottom quarks. In order to be considered for the BDT training an event needs to be reconstructable in the resolved regime on truth level (see Section 5.4.3) and needs to have at least two b jets and three additional slim jets. The assignment of the jets j_a (b jets only), j_b (b jets only), j_c, and j_d is considered to be correct if it minimizes the sum of ΔR values:

$$\Delta R_{\text{sum}} = \Delta R(j_a, \mathrm{b_t}) + \Delta R(j_b, \mathrm{b_{\bar{t}}}) + \Delta R(j_c, \mathrm{q_1}) + \Delta R(j_d, \mathrm{q_2})\,. \tag{5.6}$$

For this matching procedure not the partons of the event record are assigned to b_t, $b_{\bar{t}}$, q_1, and q_2, but the slim generator jets that have been assigned in the resolved reconstruction on truth level to the respective partons. For each event the correct assignment and one random other assignment are stored and given as signal and background sample to the BDT training procedure, respectively. Using the reconstructed leptonically decaying W boson, the top quark and antiquark are reconstructed for both assignments and their kinematic properties are given as input to the BDT among other variables. As already employed for the boosted reconstruction, the decision whether the top quark or antiquark decays leptonically is taken by the charge of the tight lepton. The full list of variables, which are used for the BDT training, testing, and evaluation, can be found in Table 5.5.

The AUC-values for the different trainings are listed in Table 5.6 and are of the same order for all years and equal within each year for the two event categories with even and odd event numbers. Figures B.1, B.2, and B.3 in the appendix show the signal and background distributions of the BDT training and testing processes together with the corresponding ROC-curves. A good separation between correct and wrong reconstruction can be achieved and a similar performance for the training and testing data sets is observed.

The reconstruction of events is done by evaluating the classifier value for each possible slim jet to final-state particle assignment with the restriction that only b jets can be assigned to the two bottom quarks. To ensure the statistical independence of training, testing, and evaluation data set, events with even event number will be evaluated using the BDT trained on oddly numbered events and vice versa. The permutation with the highest score in the evaluation of the BDT is considered to be the correct one and top quark and antiquark will be reconstructed using the selected slim jets in association with the leptonically decaying W boson.

Events that can be neither reconstructed in the boosted nor in the resolved regime will be discarded at this stage.

Reconstruction of the $t\bar{t}j$ system

Events that have been reconstructed successfully in the boosted or the resolved regime are in a subsequent step investigated for the presence of an additional high-p_T slim jet in the central region to account for j_{add} and thereby complete the reconstruction of the $t\bar{t}j$ system. Any slim jet that has been employed for the reconstruction of the $t\bar{t}$ system is not being taken into account therefore. In addition, this slim jet needs to have $p_T > 100\,\text{GeV}$ and $|\eta| < 2.5$ and needs to be well separated from the top quark and antiquark by imposing $\Delta R(j_{add}, t_{had}) > 1.5$ and $\Delta R(j_{add}, \ell) > 0.4$. In case multiple such slim jets are present in an event, the slim jet with the highest transverse momentum is assigned to j_{add}. Events without a candidate for j_{add} will be discarded at this stage.

Table 5.5: Description of variables used in the BDT training and evaluation. The variables are ranked by their importance in the BDT training. The ranking represents the average importance across all years and configurations.

Variable	Description
$m(\bar{t})$	Invariant mass of the reconstructed top antiquark
$m(t)$	Invariant mass of the reconstructed top quark
$\Delta R(W_t, b_t)$	ΔR between W boson and jet assigned to the b quark from the top quark decay
$\Delta R(W_{\bar{t}}, b_{\bar{t}})$	ΔR between W boson and jet assigned to the b quark from the top antiquark decay
$m(W_{had})$	Invariant mass of the reconstructed hadronically decaying W boson
$\Delta R(q_1, q_2)$	ΔR between the jets assigned to the two quarks from the hadronic W boson decay
t_{lep} decay	Boolean information whether top quark decays leptonically
$p_T(t)$	Transverse momentum of the reconstructed top quark
$p_T(\bar{t})$	Transverse momentum of the reconstructed top antiquark
$\Delta R(t, \bar{t})$	ΔR between the reconstructed top quark and antiquark
$p_T(q_1)$	Transverse momentum of the jet assigned to the higher p_T quark from the hadronic W boson decay
$p_T(q_2)$	Transverse momentum of the jet assigned to the lower p_T quark from the hadronic W boson decay
$p_T(b_t)$	Transverse momentum of the jet assigned to the b quark from the top quark decay
$p_T(b_{\bar{t}})$	Transverse momentum of the jet assigned to the b quark from the top antiquark decay
$p_T(W_{had})$	Transverse momentum of the reconstructed hadronically decaying W boson
$Idx(b_t)$	Index of the jet assigned to the b quark from the top quark decay
$Idx(b_{\bar{t}})$	Index of the jet assigned to the b quark from the top antiquark decay
$Idx(q_1)$	Index of the jet assigned to the higher p_T quark from the hadronic W boson decay
$Idx(q_2)$	Index of the jet assigned to the lower p_T quark from the hadronic W boson decay
$m(q_1)$	Invariant mass of the jet assigned to the higher p_T quark from the hadronic W boson decay
$m(q_2)$	Invariant mass of the jet assigned to the lower p_T quark from the hadronic W boson decay

Table 5.6: AUC-Values for the BDT performance. For the different years and event categories the AUC-Values for the BDT performance are shown.

Year	Event number	AUC
2016	even	0.866
	odd	0.866
2017	even	0.859
	odd	0.859
2018	even	0.855
	odd	0.855

5.4.5 Fiducial Phase Space

Subsequent to the reconstruction of events on detector level, the final event selection can be performed and hence the fiducial phase space can be defined. For the determination of the energy asymmetry only events are taken into account for which the reconstruction of the $t\bar{t}j$ system is possible on detector level, and which pass the spatial requirements as already outlined in Section 5.1, summarized as follows:

$$
\begin{aligned}
&p_{\mathrm{T}}(\mathrm{t_{had}}) > 300\,\mathrm{GeV}\,, && |\eta(\mathrm{t_{had}})| < 2.0\,, && \\
&p_{\mathrm{T}}(\mathrm{t_{lep}}) > 50\,\mathrm{GeV}\,, && |\eta(\mathrm{t_{lep}})| < 2.5\,, && \Delta R(\mathrm{t_{lep}}, \mathrm{t_{had}}) > 1.5\,, \\
&p_{\mathrm{T}}(j_{\mathrm{add}}) > 100\,\mathrm{GeV}\,, && |\eta(j_{\mathrm{add}})| < 2.5\,, && \Delta R(j_{\mathrm{add}}, \mathrm{t_{had}}) > 1.5\,.
\end{aligned}
$$

It has been outlined in Section 5.2.1 that an additional simulation sample has been produced for the signal process in order to increase the number of generated events in the boosted topology. As the filtered and the unfiltered signal samples for the analysis have an overlap in the phase space they cover, it needs to be ensured that this overlap is removed prior to comparing simulation to data. This is achieved by a splitting of the events in each of these samples depending on the reconstruction algorithm applied for an event on detector level. From the unfiltered signal sample only events are considered that are reconstructed in the resolved regime on detector level while only events with reconstruction in the boosted regime on detector level are taken into account from the filtered signal sample. The filtered sample is a subset of the unfiltered sample (which covers the full phase) and the validity of the procedure has been verified by comparing the expected event yields from both samples in the boosted reconstruction regime. This comparison shows an equal number of events scaled to the expectation in measured data and an equal distribution of the sensitive variables in the boosted reconstruction regime between both signal samples. The amount of unscaled simulated events in the boosted reconstruction regime is significantly higher for the filtered sample. The resolved regime on the other hand is not fully covered by the filtered sample. In the following parts of the analysis, events from the unfiltered sample, which are reconstructed in the resolved regime on detector level will be denoted as "$t\bar{t}$ SL resolved". The term "$t\bar{t}$ SL boosted" is used for the events from the filtered sample, which are reconstructed in the boosted regime on detector level.

In Table 5.7 the event yields for both simulated samples and data are given including the different groups of background processes. A separate listing for the three periods of data taking is given as well as for the combination of all data taking periods. Deviations between simulation and data are observed for 2016 and 2017, pointing in different directions and yielding less events than expected for 2016 and more events than expected for 2017. Both deviations are moderate and the difference between simulation and data is less than 10 % and well covered by the systematic uncertainties (see Section 5.6). For the 2018 data taking period a good agreement between the predicted and observed numbers of events is given, which is also the case for the combination of all data taking periods.

Table 5.7: Event yields. The table lists the event yields for the simulated data sets and for observed data in the fiducial phase space. The event yields for the simulated samples are obtained by scaling the total number of simulated events in the fiducial phase space to the respective cross sections of the processes, the selection efficiencies, and the integrated luminosity of the data taking periods. The background processes are grouped as defined in Section 5.2.2.

Sample	**2016**	**2017**	**2018**	**all**
$t\bar{t}$ SL boosted	6758.29	8759.19	12116.05	27633.55
$t\bar{t}$ SL resolved	11212.05	12914.35	20142.08	44268.50
$t\bar{t}$ DL + FH	2087.57	2459.31	3684.99	8231.87
ST	1322.87	1394.37	2454.17	5171.42
V(V) + jets	883.39	839.95	1125.47	2848.82
$t\bar{t}V$	289.38	344.77	514.41	1148.57
QCD	158.01	281.34	572.66	1012.03
Expected	22711.61	26993.31	40609.87	90314.80
Observed	**20747**	**28374**	**40366**	**89487**

One of the justifications for the extension of the fiducial phase space in Section 5.4.2 has been the increased amount of events when removing the constraint on the reconstruction in the boosted regime only. The event yields in the fiducial phase space for the signal samples clearly emphasize this extension as the signal events with reconstruction in the resolved regime make up nearly 50 % of the total events, while only about 30 % of the total events are signal events with reconstruction in the boosted regime. This is a significant increase in available events, which reduces the corresponding statistical uncertainty on the energy asymmetry measurement.

In Figure 5.4 the distributions of the kinematic variables representing the reconstructions on truth level and detector level are shown for the signal process in the fiducial phase space. These are the masses of the reconstructed top quark and antiquark and the transverse momentum of the additional jet. The top quark masses show a broader distribution on detector level than on truth level, which is on the one hand explained through the resolution of the detector and on the other hand due to possible misreconstruction of the $t\bar{t}$ system. The peak of the reconstructed top quark masses at the literature value on detector level is given for both the reconstruction method in the boosted and the resolved regime, where the peak is more narrow in the resolved regime, while the cutoff applied on the soft-drop corrected fat jet mass at 120 GeV and 220 GeV is clearly visible for the boosted regime. The transverse momentum of the additional jet shows a similar distribution be-

tween truth level and detector level, with the biggest difference being the strict cutoff at 100 GeV, which is only present for the reconstruction on detector level.

The distributions in Figure 5.5 show the same kinematic distributions as in Figure 5.4 on detector level only, including the contributions from background processes and the observations from measured data. In addition, the three main variables of interest for the calculation of the energy asymmetry are shown, being the energy difference between top quark and top antiquark $\Delta E_{\mathrm{t\bar{t}}}$, the absolute of the rapidity of the $\mathrm{t\bar{t}}j$ center-of-mass system $|y_{\mathrm{t\bar{t}}j}|$, and the optimized scattering angle of the additional hard jet in the event θ_j^{opt}. The shapes of the background processes are equal to those of the signal processes and the step at the top masses at 120 GeV is present in these distributions as well. For all distributions of interest there is a good agreement between simulation and measured data as can be seen in the figure. In the appendix the distributions are shown separately for each year of data taking in Figures C.4, C.5, and C.6. Due to the relatively small fiducial phase space of the analysis, which contains roughly 90 000 events in combination for all periods of data taking, the statistical uncertainty can play a dominant role in the measurement of the energy asymmetry and it does not make sense to enhance this effect by a further split of the phase space. Therefore, a separate contemplation for each data taking period will not be given in the remaining parts of the presented analysis.

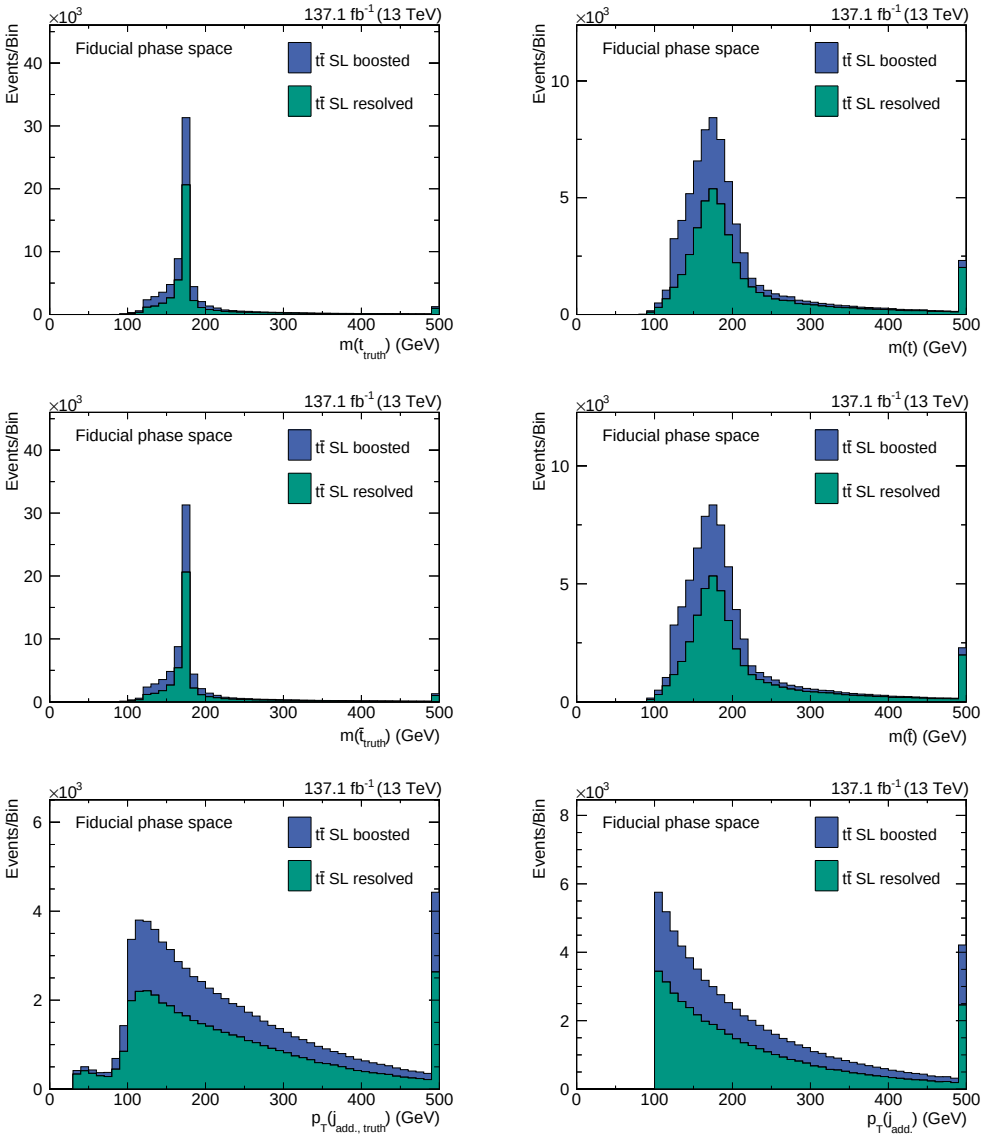

Figure 5.4: Reconstruction in the fiducial phase space: The distribution of the masses of the reconstructed top quark and antiquark are shown together with the transverse momentum of the additional jet of $t\bar{t}j$ production. Only events from the two signal samples as selected in the fiducial phase space are considered here. The left column shows the distributions for the reconstruction on truth level, while the right column shows the same distribution for the reconstruction on detector level.

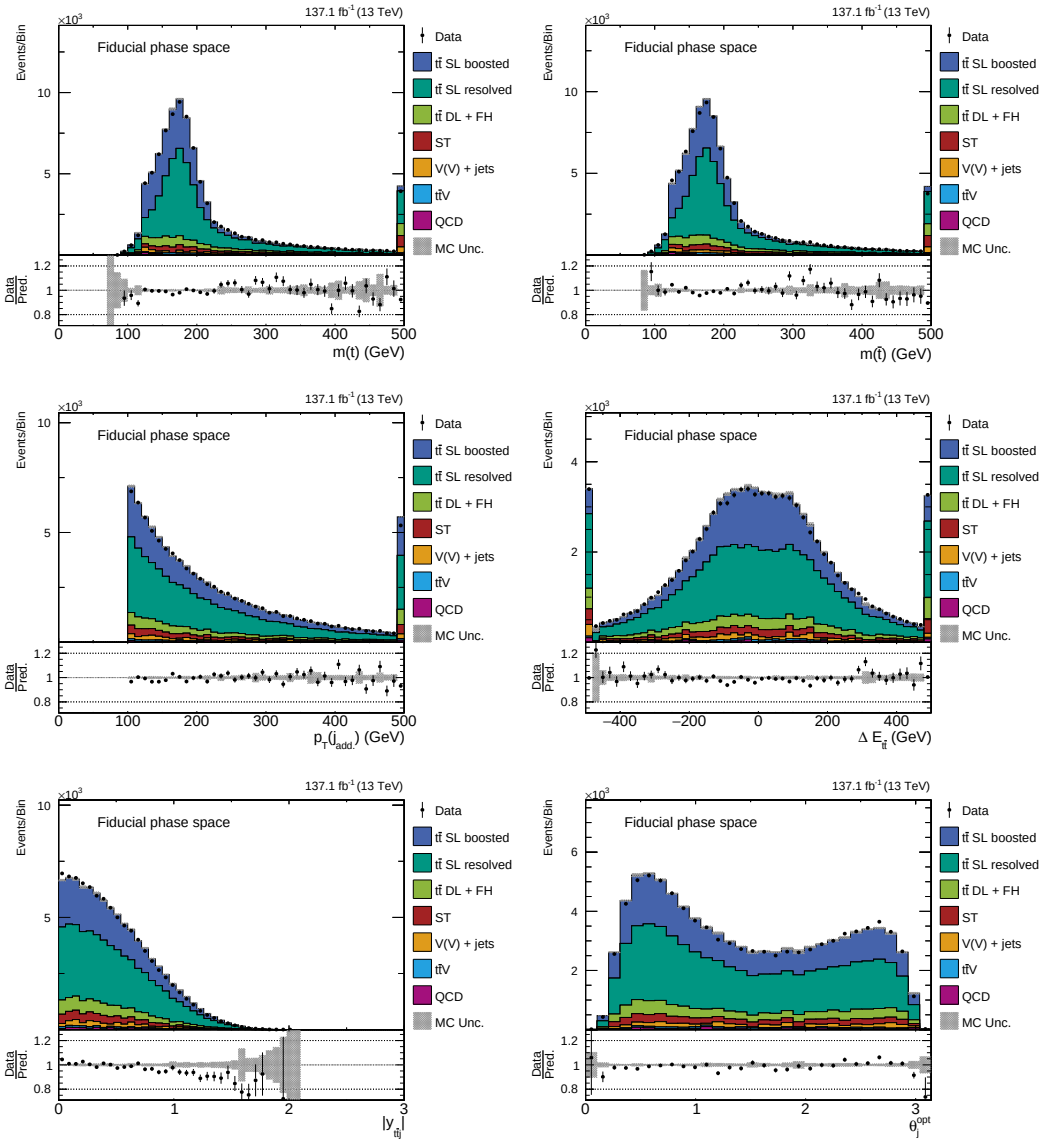

Figure 5.5: Variables of interest in the fiducial phase space: The distributions of the kinematic properties of the reconstruction are shown for simulation and measured data. In addition the distributions for the three variables of interest for the energy asymmetry are shown. Good agreement between simulation and measured data can be observed. Small deviations are within the uncertainty due to the limited size of the simulated samples and the relatively small fiducial phase space of the analysis.

5.4.6 Energy Asymmetry

Having defined the fiducial phase space for the analysis, the energy asymmetry on different levels can be investigated within this phase space. Values for different levels are given in Tables 5.8, 5.9, 5.10, and 5.11, and are shown in addition in Figure 5.6. Further combinations of requirements are applied for enhancing the strength of the observable, where the combinations with $|y_{t\bar{t}j}| > 0.5$ and $|\Delta E_{t\bar{t}}| > 0\,\text{GeV}$ are not presented, as they have shown to decrease both the amount of events and the strength of the effect on all levels. The given values are quoted with a statistical and a systematic uncertainty if applicable. The statistical uncertainty is obtained by assuming a Poisson distribution for $N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$ and $N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$ and by error propagation of the corresponding uncertainty, which can be done for all investigated levels. The systematic uncertainty is the error propagation of the uncertainty due to the limited size of simulation samples and therefore not quoted for observed data in Table 5.11. Further systematic uncertainties as listed in Section 5.6 are not included at this stage.

Table 5.8 shows the energy asymmetry on truth level for the combination of the two signal samples and gives the predictions for the fiducial phase space of the analysis. The deviation from zero is the strongest for the energy asymmetry when applying the requirements $|y_{t\bar{t}j}| > 0.5$ and $|\Delta E_{t\bar{t}}| > 50\,\text{GeV}$ and considering only events with $0.3\pi < \theta_j^{\text{opt}} < 0.7\pi$, which is in agreement with the observations made in Ref. [29] and Ref. [31]. The obtained value on truth level of $\mathbf{A_E^{opt} = -1.59\,\% \pm 1.00\,\%\,(stat) \pm 0.37\,\%\,(syst)}$ will be the target of the analysis on detector level and denoted as the prediction in the fiducial phase space.

The values for the combination of the two signal samples on detector level are given in Table 5.9 and show a very similar behavior compared to the values on truth level within the given uncertainties. The inclusion of the background processes leads to the values in Table 5.10. Due to the higher number of expected events when including the background processes, the statistic uncertainty becomes smaller in comparison to the detector-level values without backgrounds. The systematic uncertainty due to the limited simulation samples size increases on the other hand, which is a result of the comparably low number of unscaled events in the background simulation samples in the fiducial phase space. For some sets of requirements the magnitude of the energy asymmetry becomes higher when including the background processes. Due to the high systematic uncertainty when performing an evaluation of the energy asymmetry on detector level with the background processes only, it cannot be further determined whether this is a systematic effect due to the reconstruction algorithms, or a purely statistical fluctuation within the events in the fiducial phase space.

The values for the energy asymmetry for measured data are given in Table 5.11 and allow in principle for a direct comparison to the detector-level values of the signal and background processes combined. In this comparison an acceptable agreement within the given uncertainties is observed for most sets of requirements. It needs to be stressed that for the simulation samples further systematic uncertainties need to taken into account and that a valuable result can only be obtained by a removal of the background influences and an unfolding to the truth-level predictions.

Table 5.8: Asymmetry values on truth level. The energy asymmetry values on truth level for the combination of the signal samples $t\bar{t}$ SL boosted and $t\bar{t}$ SL resolved. The quoted uncertainties refer to the propagated Poisson uncertainty on $N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$ and $N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$ (stat) and the propagated error due to the limited size of the simulation samples (syst).

θ_j^{opt}	$\lvert y_{t\bar{t}j}\rvert$	$\lvert\Delta E_{t\bar{t}}\rvert$	$N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$	$N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$	A_E^{opt}
$0 - 0.3\pi$	> 0	$> 0\,\text{GeV}$	12642.1	12710.9	$-0.272\,\% \pm 0.628\,\%\,(\text{stat}) \pm 0.231\,\%\,(\text{syst})$
	> 0	$> 50\,\text{GeV}$	10115.0	10158.1	$-0.213\,\% \pm 0.702\,\%\,(\text{stat}) \pm 0.258\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	4362.2	4342.6	$0.226\,\% \pm 1.072\,\%\,(\text{stat}) \pm 0.394\,\%\,(\text{syst})$
$0.3\pi - 0.7\pi$	> 0	$> 0\,\text{GeV}$	13864.7	14130.0	$-0.948\,\% \pm 0.598\,\%\,(\text{stat}) \pm 0.221\,\%\,(\text{syst})$
	> 0	$> 50\,\text{GeV}$	10037.2	10288.6	$-1.237\,\% \pm 0.701\,\%\,(\text{stat}) \pm 0.257\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	4882.0	5039.6	$\mathbf{-1.588}\,\% \pm \mathbf{1.004}\,\%\,(\text{stat}) \pm \mathbf{0.371}\,\%\,(\text{syst})$
$0.7\pi - \pi$	> 0	$> 0\,\text{GeV}$	9306.6	9302.9	$0.020\,\% \pm 0.733\,\%\,(\text{stat}) \pm 0.269\,\%\,(\text{syst})$
	> 0	$> 50\,\text{GeV}$	7570.5	7537.9	$0.216\,\% \pm 0.814\,\%\,(\text{stat}) \pm 0.299\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	2668.0	2668.4	$-0.008\,\% \pm 1.369\,\%\,(\text{stat}) \pm 0.505\,\%\,(\text{syst})$

Table 5.9: Asymmetry values on detector level. The energy asymmetry values on detector level for the combination of the signal samples $t\bar{t}$ SL boosted and $t\bar{t}$ SL resolved. The quoted uncertainties refer to the propagated Poisson uncertainty on $N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$ and $N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$ (stat) and the propagated error due to the limited size of the simulation samples (syst).

θ_j^{opt}	$\lvert y_{t\bar{t}j}\rvert$	$\lvert\Delta E_{t\bar{t}}\rvert$	$N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$	$N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$	A_E^{opt}
$0 - 0.3\pi$	> 0	$> 0\,\text{GeV}$	12837.7	12890.1	$-0.204\,\% \pm 0.623\,\%\,(\text{stat}) \pm 0.228\,\%\,(\text{syst})$
	> 0	$> 50\,\text{GeV}$	10503.8	10518.6	$-0.070\,\% \pm 0.690\,\%\,(\text{stat}) \pm 0.253\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	4447.0	4480.9	$-0.380\,\% \pm 1.058\,\%\,(\text{stat}) \pm 0.388\,\%\,(\text{syst})$
$0.3\pi - 0.7\pi$	> 0	$> 0\,\text{GeV}$	13780.1	14124.7	$-1.235\,\% \pm 0.599\,\%\,(\text{stat}) \pm 0.222\,\%\,(\text{syst})$
	> 0	$> 50\,\text{GeV}$	10620.8	10909.8	$-1.342\,\% \pm 0.681\,\%\,(\text{stat}) \pm 0.256\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	5050.2	5234.2	$\mathbf{-1.788}\,\% \pm \mathbf{0.986}\,\%\,(\text{stat}) \pm \mathbf{0.373}\,\%\,(\text{syst})$
$0.7\pi - \pi$	> 0	$> 0\,\text{GeV}$	9170.7	9153.8	$0.092\,\% \pm 0.739\,\%\,(\text{stat}) \pm 0.271\,\%\,(\text{syst})$
	> 0	$> 50\,\text{GeV}$	7708.6	7670.1	$0.250\,\% \pm 0.806\,\%\,(\text{stat}) \pm 0.298\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	2562.4	2575.1	$-0.248\,\% \pm 1.395\,\%\,(\text{stat}) \pm 0.518\,\%\,(\text{syst})$

Table 5.10: Asymmetry values on detector level including backgrounds. The energy asymmetry values on detector level for the combination of the signal samples $t\bar{t}$ SL boosted and $t\bar{t}$ SL resolved, and all simulation samples for the background processes. The quoted uncertainties refer to the propagated Poisson uncertainty on $N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$ and $N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$ (stat) and the propagated error due to the limited size of the simulation samples (syst).

θ_j^{opt}	$\lvert y_{t\bar{t}j}\rvert$	$\lvert\Delta E_{t\bar{t}}\rvert$	$N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$	$N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$	A_E^{opt}
$0 - 0.3\pi$	> 0	$> 0\,\text{GeV}$	15741.7	16049.2	$-0.967\,\% \pm 0.561\,\%\,(\text{stat}) \pm 0.344\,\%\,(\text{syst})$
	> 0	$> 50\,\text{GeV}$	12986.2	13251.8	$-1.012\,\% \pm 0.617\,\%\,(\text{stat}) \pm 0.390\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	5458.7	5593.7	$-1.222\,\% \pm 0.951\,\%\,(\text{stat}) \pm 0.561\,\%\,(\text{syst})$
$0.3\pi - 0.7\pi$	> 0	$> 0\,\text{GeV}$	17178.2	18028.8	$-2.416\,\% \pm 0.533\,\%\,(\text{stat}) \pm 0.466\,\%\,(\text{syst})$
	> 0	$> 50\,\text{GeV}$	13473.7	14053.4	$-2.106\,\% \pm 0.603\,\%\,(\text{stat}) \pm 0.523\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	6419.0	6707.2	$\mathbf{-2.195}\,\% \pm \mathbf{0.873}\,\%\,(\text{stat}) \pm \mathbf{0.791}\,\%\,(\text{syst})$
$0.7\pi - \pi$	> 0	$> 0\,\text{GeV}$	11671.8	11699.6	$-0.119\,\% \pm 0.654\,\%\,(\text{stat}) \pm 0.498\,\%\,(\text{syst})$
	> 0	$> 50\,\text{GeV}$	9874.6	9890.8	$-0.082\,\% \pm 0.711\,\%\,(\text{stat}) \pm 0.552\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	3284.4	3300.0	$-0.236\,\% \pm 1.232\,\%\,(\text{stat}) \pm 0.975\,\%\,(\text{syst})$

Table 5.11: Asymmetry values for measured data. The energy asymmetry values for measured data in the fiducial phase space. The quoted uncertainty is obtained by assuming a Poisson distribution for $N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$ and $N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$ and by error propagation of the corresponding uncertainty.

θ_j^{opt}	$\lvert y_{t\bar{t}j}\rvert$	$\lvert\Delta E_{t\bar{t}}\rvert$	$N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$	$N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$	A_E^{opt}
$0 - 0.3\pi$	> 0	$> 0\,\text{GeV}$	15359.0	15916.0	$-1.781\,\% \pm 0.565\,\%\,(\text{stat})$
	> 0	$> 50\,\text{GeV}$	12731.0	13137.0	$-1.570\,\% \pm 0.622\,\%\,(\text{stat})$
	> 0.5	$> 50\,\text{GeV}$	5132.0	5404.0	$-2.582\,\% \pm 0.974\,\%\,(\text{stat})$
$0.3\pi - 0.7\pi$	> 0	$> 0\,\text{GeV}$	16991.0	17481.0	$-1.421\,\% \pm 0.539\,\%\,(\text{stat})$
	> 0	$> 50\,\text{GeV}$	13193.0	13682.0	$-1.820\,\% \pm 0.610\,\%\,(\text{stat})$
	> 0.5	$> 50\,\text{GeV}$	6146.0	6501.0	$\mathbf{-2.807}\,\% \pm \mathbf{0.889}\,\%\,(\text{stat})$
$0.7\pi - \pi$	> 0	$> 0\,\text{GeV}$	11876.0	11863.0	$0.055\,\% \pm 0.649\,\%\,(\text{stat})$
	> 0	$> 50\,\text{GeV}$	10095.0	10051.0	$0.218\,\% \pm 0.705\,\%\,(\text{stat})$
	> 0.5	$> 50\,\text{GeV}$	3201.0	3264.0	$-0.974\,\% \pm 1.244\,\%\,(\text{stat})$

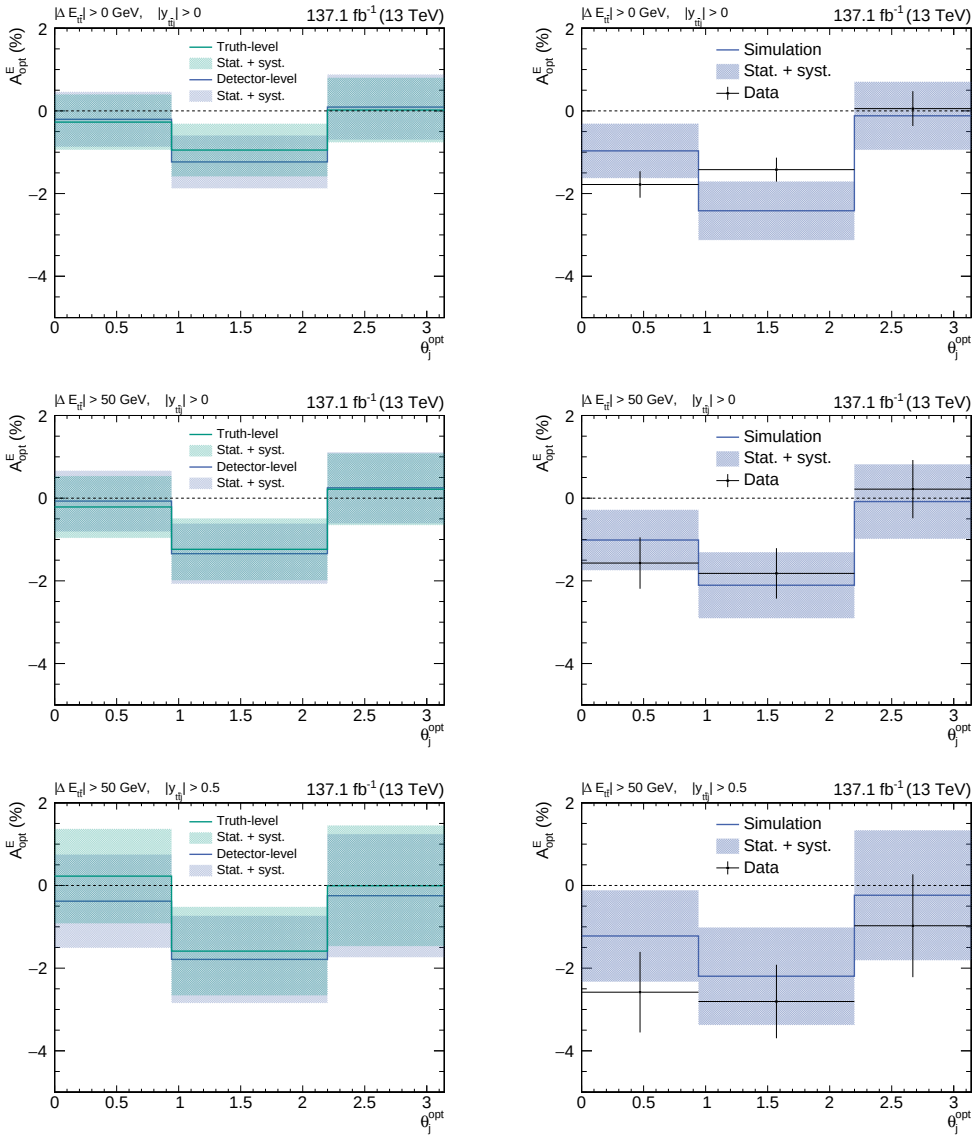

Figure 5.6: Asymmetry values in the fiducial phase space: The energy asymmetry values in the fiducial phase space are shown for different requirements on $|\Delta E_{t\bar{t}}|$ and $|y_{t\bar{t}j}|$. On the left hand side the values on truth level and detector level are shown under consideration of only the signal process. The figures on the right hand side include all background processes for simulation and in addition the observations on measured data are shown.

Further predictions and observations on different levels and different subsets of the fiducial phase space are given in the appendix in Tables D.1 to D.13. In addition to the discussed main set of requirements, the analysis has also been performed without applying the requirement of $|y_{t\bar{t}j}| > 0.5$, which predicts an energy asymmetry value of $A_E^{\text{opt}} = -1.24\,\% \pm 0.70\,\%\,(\text{stat}) \pm 0.26\,\%\,(\text{syst})$ on truth level. This value is closer to zero than the prediction for the fiducial phase space, but suffers less from statistic and systematic uncertainties. The values for this alternative set of requirements are presented in the appendix in Chapter F derived with the same procedure as described within this chapter.

5.5 Corrections on Simulated Events

The production of simulation samples is a time-consuming and computing-intensive task, which is usually performed prior and during the respective data taking period. Some of the exact data-taking conditions for the simulation of samples are unknown at the time of production and deviations in observables and selection efficiencies between data and simulation can be a possible result. In order to balance such undesired effects without a complete reproduction of the simulation, dedicated event weights and scaling factors can be applied as correction on the simulation samples. The corrections, which are applied on simulated events in this thesis, are outlined in the following.

Pileup Reweighting

The amount of pileup interactions is estimated prior to the respective data taking period and the corresponding pileup distributions for the simulation samples is modeled from a Poisson distribution with the preliminary estimate as mean. The difference between this estimated value and the actual amount of pileup interactions does not have a significant effect on objects with high transverse momentum, as pileup interactions mainly involve soft interactions only. Certain observables, as for example the number of primary vertices per event, suffer from the mismodeling of pileup interactions and all simulated events are reweighted in order to improve the agreement between data and simulation for such observables. The corresponding event weights are derived by the Luminosity Physics Objects Group (LUM POG) [40] of the CMS Collaboration based on a minimum-bias data set with a cross section of 69.2 mb [149, 150]. A further improvement of the agreement between data and simulation can be obtained by changing the minimum-bias cross section to higher values of up to 80.0 mb, which is nevertheless not employed, as the minimum-bias cross section is measured with high accuracy.

Lepton Efficiencies

The selection of exactly one tight lepton in an event results in different yields in simulation and data, which needs to be corrected accordingly. Various lepton properties are subject to the selection requirements and dedicated scale factors are applied to account for the most important of them.

For muons, efficiencies on the identification (ID), isolation and the trigger are considered and the overall efficiency is given by the product:

$$\epsilon = \epsilon_{\text{ID}} \cdot \epsilon_{\text{Isolation}|\text{ID}} \cdot \epsilon_{\text{Trigger}|\text{Isolation}} \,. \tag{5.7}$$

As indicated in the above equation, the respective efficiencies are not completely independent of each other and the isolation efficiency depends on the ID efficiency, while the trigger efficiency depends on the value of the isolation efficiency. The corrections are provided by the Muon Physics Objects Group (MUO POG) [151–153] and are derived using a tag-and-probe method on J/ψ meson or Z boson resonances [154, 155].

The overall efficiency for the selection of electrons is given by:

$$\epsilon = \epsilon_{\text{Reconstruction}} \cdot \epsilon_{\text{ID}} \cdot \epsilon_{\text{Trigger}} \,, \tag{5.8}$$

which includes corrections for the reconstruction, identification, and trigger efficiency of selected electrons. The corrections for reconstruction and identification are centrally provided by the Electron & Photon Physics Objects Group (EGamma POG) [156] of the CMS Collaboration and have been derived with a tag-and-probe method using $\text{Z} \rightarrow \text{e}^+\text{e}^-$ events [157, 158]. The scale factors for the employed electron triggers are not provided by the EGamma POG, but had to be produced privately following the procedures described in Refs. [159] and [160].

b Tagging Efficiencies

Only events with at least two b jets can be reconstructed in the resolved regime in the analysis and the events passing the preselection are therefore partially filtered according to the amount of b jets in the event record. This selection is known to yield different results between data and simulation and different reweighting procedures are available for correcting this effect [161]. The method, which is employed in the presented analysis, is applying event weights only on simulated events that pass the selection criteria and hence migrations between different b jet multiplicities do not need to be considered. For the determination of the event-specific weight, efficiencies for each simulated process are determined in a first step on the full simulation sample without any preselection. These efficiencies are derived in dependence of specific jet p_{T} and η bins and correspond to the probability of identifying a jet of true flavor f as a b jet:

$$\epsilon_f(p_{\text{T}}, \eta) = \frac{N_f^{\text{b-tagged}}(p_{\text{T}}, \eta)}{N_f^{\text{total}}(p_{\text{T}}, \eta)}. \tag{5.9}$$

The total number of jets with flavor f is given as N_f^{total}, while the number of b-tagged jets with flavor f is represented by $N_f^{\text{b-tagged}}$. For jets with true flavor b, these efficiencies correspond to the probability of correctly identifying such jets with the b tagging algorithm, while for c-flavored or light-flavored jets, the respective mistagging efficiency is represented by ϵ_{c} and ϵ_{udsg}, respectively.

In order to calculate the weight for a specific event, the probabilities in simulation and data for correctly identifying all jets in an event with N_i b tagged jets and N_j untagged jets are used. These probabilities are given via:

$$P(\mathrm{MC}) = \prod_{i=\mathrm{tagged}}^{N_i} \epsilon_{\mathrm{b}}^{i} \prod_{j=\mathrm{untagged}}^{N_j} \left(1 - \epsilon_{\mathrm{c,udsg}}^{j}\right) , \tag{5.10}$$

$$P(\mathrm{Data}) = \prod_{i=\mathrm{tagged}}^{N_i} \mathrm{SF}_{\mathrm{b}}^{i} \epsilon_{\mathrm{b}}^{i} \prod_{j=\mathrm{untagged}}^{N_j} \left(1 - \mathrm{SF}_{\mathrm{c,udsg}}^{j} \epsilon_{\mathrm{c,udsg}}^{j}\right) . \tag{5.11}$$

For the probability in data, the scale factors of the DeepJet algorithm for b jets $\mathrm{SF_b}$ and c/light-flavor jets $\mathrm{SF_{c,udsg}}$ are taken into account. These scale factors are defined as $\mathrm{SF} = \epsilon_{\mathrm{Data}}/\epsilon_{\mathrm{MC}}$ and correct the differences between the b tagging and mistagging efficiencies in simulation and data. They are dependent on the employed working point of the DeepJet algorithm and provided by the b Tag & Vertexing Physics Objects Group (BTV POG) of the CMS Collaboration for each year of data taking [162–164]. The weight that is applied on an event is then given as:

$$w = \frac{P(\mathrm{Data})}{P(\mathrm{MC})} . \tag{5.12}$$

This weight is applied only on events that are reconstructed in the resolved regime. Events that are reconstructed in the boosted regime are not filtered according to the amount of b jets in an event and hence no b tagging correction is required on them.

top Tagging Efficiencies

The reconstruction of an event in the boosted regime requires a top tagged fat jet in the event. This selection suffers from similar differences in the efficiencies between data and simulation as it is the case for b tagged jets. In case of top jets, this undesired effect is balanced by applying an additional event weight, which is given as:

$$w = \prod_{i=1}^{N_i} \mathrm{SF}^{i}(p_{\mathrm{T}}) . \tag{5.13}$$

The number of fat jets in an event is given by N_i, and the p_{T}-dependent scale factor of a fat jet is given by $\mathrm{SF}^{i}(p_{\mathrm{T}})$. These scale factors are provided centrally by the Jet & MET Physics Objects Group (JetMET POG) [165] of the CMS Collaboration for the different years of data taking and the employed working points of the DeepAK8 top vs QCD tagger [166]. The event weight is applied on all events, independent of the reconstruction algorithm.

5.6 Uncertainty Treatment

The measurement of the energy asymmetry in $\mathrm{t\bar{t}}j$ production is affected by various different sources of uncertainty, which can be grouped into experimental and theoretical uncertainties. For the extraction of the final result via unfolding, each uncertainty is introduced as a nuisance parameter in the unfolding procedure.

5.6.1 Experimental Uncertainties

Pileup

The corrections of the pileup reweighting procedure have an uncertainty both due to the reweighting method and due to the measurement of the total inelastic proton-proton cross section. A coverage of both effects is obtained by shifting the nominal proton-proton cross section by 4.6 % as recommended by the LUM POG [167] and propagating this shift to the unfolding distributions.

Lepton Efficiencies

Both the statistical and the systematic uncertainty in the derivation of the lepton efficiencies need to be accounted by adding the corresponding uncertainty on the distributions considered in the unfolding. Each selection efficiency on electrons or muons is associated with a variation in up and down direction and their respective combination is applied instead of the nominal event weight for the lepton efficiency.

b Tagging Efficiency

The inefficiencies of the b tagging algorithm, which are corrected via event weights on the simulation, are affected by various systematic uncertainties. The scale factors, which are centrally provided by the BTV POG with a corresponding uncertainty, are shifted up and down and applied as alternative event weights to account for this uncertainty.

top Tagging Efficiency

The JetMET POG provides uncertainties on the scale factors, which have to be applied to correct the inefficiencies in the selection of top tagged fat jets. These uncertainties are considered by the application of up and down shifted event weights instead of the nominal event weights.

Jet Energy Scale

The corrections on the jet energy, as described in Section 4.3.6, introduce a p_{T}- and η-dependent uncertainty on the four-momentum of each jet [168]. The variation of each jet energy by ± 1 standard deviation of the jet energy correction cannot be covered correctly by the application of an event weight. Instead, the effect is considered by repeating the complete analysis chain with simulation samples for which the energy of each jet in the event is varied accordingly. The variation of jet energies is also propagated to the calculation of the missing transverse momentum.

Jet Energy Resolution

The jet energy resolution is smeared in simulation samples in order to improve the agreement between data and simulation [169]. The uncertainty of this correction is treated similarly to the corrections of the jet energy scale and the complete analysis is reiterated with the respective up and down shifts on the jet energy applied.

Unclustered Energy

The missing transverse momentum of an event is obtained with all high-level objects in an event, as for example photons, electrons, muons, and jets [170]. The uncertainty on the energy measurement of these objects needs to be propagated to $p_{\mathrm{T}}^{\mathrm{miss}}$ and the complete analysis chain is repeated with the corresponding up and down shift on $p_{\mathrm{T}}^{\mathrm{miss}}$ applied.

Luminosity

The integrated luminosity for each data taking period can only be measured within a finite level of precision and has an uncertainty of 2.5 % [171], 2.3 % [172], and 2.5 % [173] for the 2016, 2017, and 2018 data sets, respectively. As a combined fit is performed for all data sets, a conservative normalization uncertainty of 2.5 % is applied globally to consider this effect.

Limited Size of Simulated Samples

The limited size of simulation samples and therefore events in each bin of the unfolding distribution is considered in the unfolding procedure following the Barlow-Beeston method [65]. This method introduces one nuisance parameter for each bin of each simulated sample in every considered region, which requires a very high computation power in the evaluation of the fit. In the presented analysis this is avoided by applying the so-called "Barlow-Beeston-lite" approach, in which only one nuisance parameter is introduced per bin.

5.6.2 Theoretical Uncertainties

PDF and α_{s}

To account for the uncertainties in the choice of the employed PDF set and the value of the strong coupling constant α_{S}, the variations for different eigenvalues of the NNPDF set [84, 85] and α_{S} values are considered. Following the recommended procedure in Ref. [174], the envelope of 30 different shape variations of the nominal PDF and two α_{S} values of $\alpha_{\mathrm{S}} = 0.1195$ and $\alpha_{\mathrm{S}} = 0.1165$ is constructed. The nominal value is $\alpha_{\mathrm{S}} = 0.118$. The up and down variations in the unfolding distribution of this envelope are used in order to determine the uncertainty due to the choice of the PDF set and the value of α_{S}.

Initial-state and Final-state Radiation

The chosen value of the strong coupling constant α_S has an additional impact on the probability for additional gluon radiation in the initial or final state of a simulated event. This is taken into account via event weights, which correspond to doubled and halved probabilities for additional gluon radiation. These event weights are provided by PYTHIA and are available for all simulation samples of $t\bar{t}$ production and single top quark production for all periods of data taking.

Renormalization and Factorization Scales

The uncertainty on the renormalization and factorization scales at matrix-element level is considered via the variation of the values for μ_R and μ_F. The LHE reweighting procedure [175] is employed to create distributions with the values of μ_R and μ_F being multiplied each individually by either 0.5, 1 or 2. Except for the combinations in which one value is multiplied by 0.5 and the other is multiplied by 2, all combinations are considered for determining an envelope, which is used to assign the respective up and down shift of the renormalization and factorization scales.

Matching of Parton Shower and Matrix Element

The simulation samples for the production of $t\bar{t}$ are produced with POWHEG for the calculation of the matrix element and interfered with PYTHIA for the parton shower simulation. The matching of the parton shower to the matrix element is controlled by a damping function with the damping parameter h_{damp}, which regulates in addition the high-p_T radiation of partons [100]. As the analysis is investigating the production of $t\bar{t}$ with one additional jet, this parameter is of special interest and an uncertainty on h_{damp} needs to be considered. This is done via dedicated simulation samples for $t\bar{t}$ production, in which the h_{damp} parameter has been varied with respect to the nominal value of $1.379 \cdot m_t$ [176, 177]. The value of m_t is set to 172.5 GeV and the up and down variations are set to $1.379^{+0.926}_{-0.5052} \cdot m_t$. The production of dedicated samples for the filtered signal sample has not yet been finalized and therefore the unfiltered signal samples with h_{damp} variation are employed for both the event reconstruction in the boosted and the resolved regime.

Underlying Event

The event generator tune CP5 is used for the simulation of $t\bar{t}$ production and the settings of the tune are responsible for the modeling of the underlying event and multi-parton interactions. Uncertainties in the tune settings are covered by dedicated $t\bar{t}$ simulation samples, in which especially the settings for multi-parton interactions and color reconnection are varied. As it is the case for the h_{damp} varied samples, no variation for the filtered signal sample, but only the unfiltered varied signal samples are available for this uncertainty source.

Table 5.12: Normalization uncertainties. The normalization uncertainties that are assigned to the different groups of processes in the analysis are shown, based on the respective cross sections.

Sample	Normalization uncertainty
$t\bar{t}$ SL boosted	5 %
$t\bar{t}$ SL resolved	5 %
$t\bar{t}$ DL + FH	5 %
ST	15 %
V(V) + jets	10 %
$t\bar{t}V$	5 %
QCD	50 %

Normalization of Signal and Background Processes

The normalization of each simulated process depends on the theoretical cross section that is applied to scale the simulated events to the expected amount of events in measured data. These cross sections are only known at a certain precision and a rate uncertainty is assigned to each group of processes. These uncertainties are based on the most recent measurements of the given processes and increased to a rather conservative estimation due to the mixture of different processes within one group. The assigned values for the different groups are given in Table 5.12, the references are given together with the list of simulation samples in Appendix A.1.

Figure 5.7 shows the truth-level variation on the event yields and the corresponding energy asymmetry values for the up and down shifts of all systematic uncertainties that cause a shape variation in the unfolding bin distribution. In Figures E.1 and E.2 in Appendix E the shape variation on the unfolding bin distribution on detector level is shown for each of these uncertainty sources separately. A further discussion of these distributions will be given in Section 5.8.

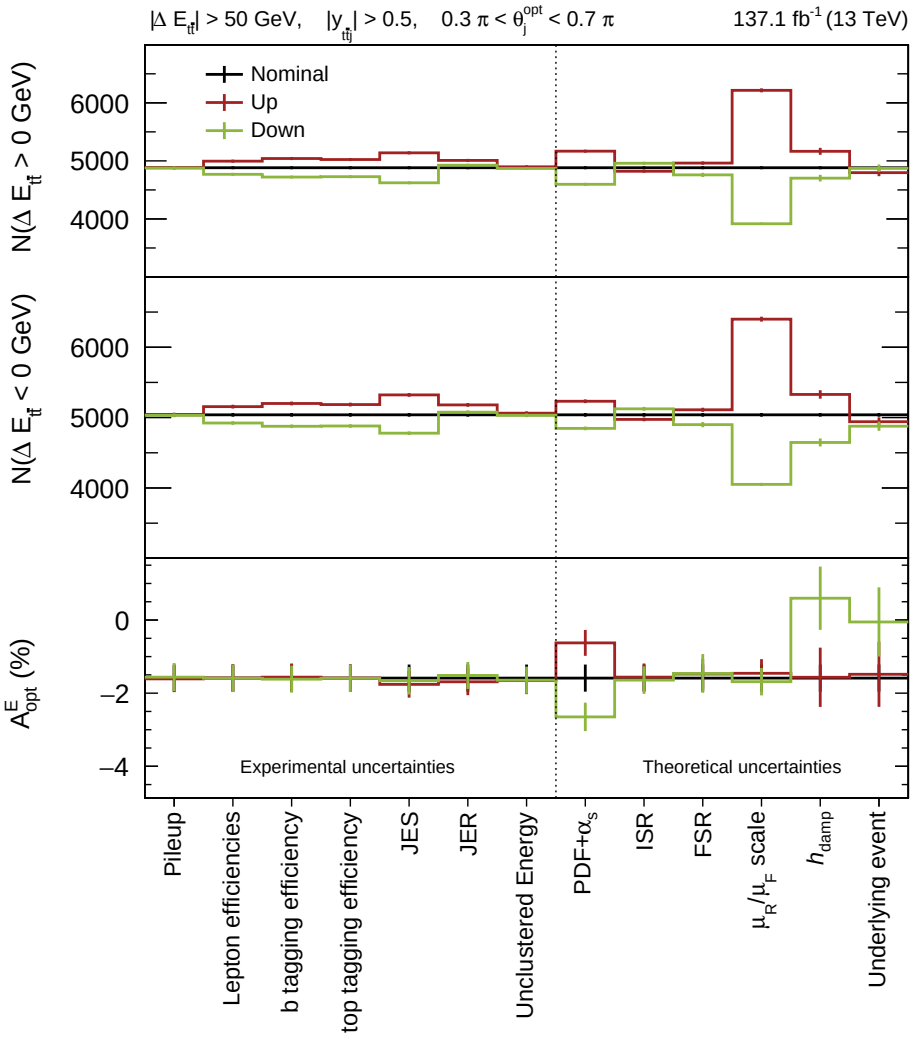

Figure 5.7: Impact of systematic uncertainties: The impact of the systematic uncertainties that yield a shape variation for the unfolding distribution is shown on truth level in respect to the nominal distributions. The two upper rows show the event yields for the up and down shifts of the systematic uncertainties for both events with $\Delta E_{t\bar{t}} > 0\,\text{GeV}$ (upper row) and $\Delta E_{t\bar{t}} < 0\,\text{GeV}$ (middle row). The lower row shows the respective variations for the energy asymmetry value. The error bars indicate the uncertainty due to the limited size of the simulation samples.

Table 5.13: Coarse binning for the variables of interest. The minimal binning on $\Delta E_{t\bar{t}}$, θ_j^{opt}, and $|y_{t\bar{t}j}|$ is applied, which still allows for the subsequent application of selections in order to enhance the strength of the energy asymmetry.

Variable	Bin 1	Bin 2	Bin 3	Bin 4
$\Delta E_{t\bar{t}}$	$-50\,\text{GeV} > \Delta E_{t\bar{t}}$	$-50\,\text{GeV} > \Delta E_{t\bar{t}} > 0\,\text{GeV}$	$0\,\text{GeV} < \Delta E_{t\bar{t}} < 50\,\text{GeV}$	$50\,\text{GeV} < \Delta E_{t\bar{t}}$
θ_j^{opt}	$0 < \theta_j^{opt} < 0.3\pi$	$0.3\pi < \theta_j^{opt} < 0.7\pi$	$0.7\pi < \theta_j^{opt} < \pi$	–
$\lvert y_{t\bar{t}j}\rvert$	$\lvert y_{t\bar{t}j}\rvert < 0.5$	$\lvert y_{t\bar{t}j}\rvert > 0.5$	–	–

5.7 Unfolding to Truth Level

The comparisons of the event yields and the corresponding energy asymmetry values in Tables 5.8 - 5.11 and Figure 5.6 show mostly a good agreement between truth level and detector level for the two signal samples, as well as between the full simulation and measured data. However, this does not mean that a direct comparison of the results on measured data to theory predictions is possible. For this comparison both a subtraction of the contributions from background processes and an unfolding to truth level are required. The need for the background subtraction is obvious from the event yields, while the need for unfolding becomes evident from the distributions and migration matrices in Figures 5.8, 5.9, and 5.10, in addition to the arguments raised in Section 3.2. These figures show the distributions for the three variables of interest for the energy asymmetry, being $\Delta E_{t\bar{t}}$, $|y_{t\bar{t}j}|$, and θ_j^{opt}. It is hereby not only of importance to compare the distributions of these variables between truth level and detector level, but also to investigate this matching on a more event-based level by migration matrices. A different and more coarse binning is chosen for this migration matrices for each of the variables. It has been discussed in Sections 5.1 and 5.4.6 that the strength of the effect can be enhanced by selection criteria on $\Delta E_{t\bar{t}}$, $|y_{t\bar{t}j}|$, and θ_j^{opt}, and these selections give the basic for the binning in the migration matrices. The chosen binning is given in Table 5.13.

For the three variables, similar distributions are observed between truth level and detector level for both the fine and the coarse binning schemes. The energy difference between top quark and antiquark shows the biggest deviations in the fine binning between truth level and detector level, which is expected in a way that this observable is very sensitive to misreconstruction effects, especially in the resolved regime with the comparably high multiplicity of involved slim jets. In the coarse binning scheme this effect becomes moderate, but is still visible. This sensitivity to misreconstruction becomes especially evident in the migration matrices for $\Delta E_{t\bar{t}}$ in both the resolved and the boosted regime and leads to strong migrations for events with $|\Delta E_{t\bar{t}}| < 50\,\text{GeV}$ on truth level. In the resolved regime these events have a probability of about only 32 % to be in the same $\Delta E_{t\bar{t}}$ bin on truth level and detector level, while this probability is about 50 % in the boosted regime. For events

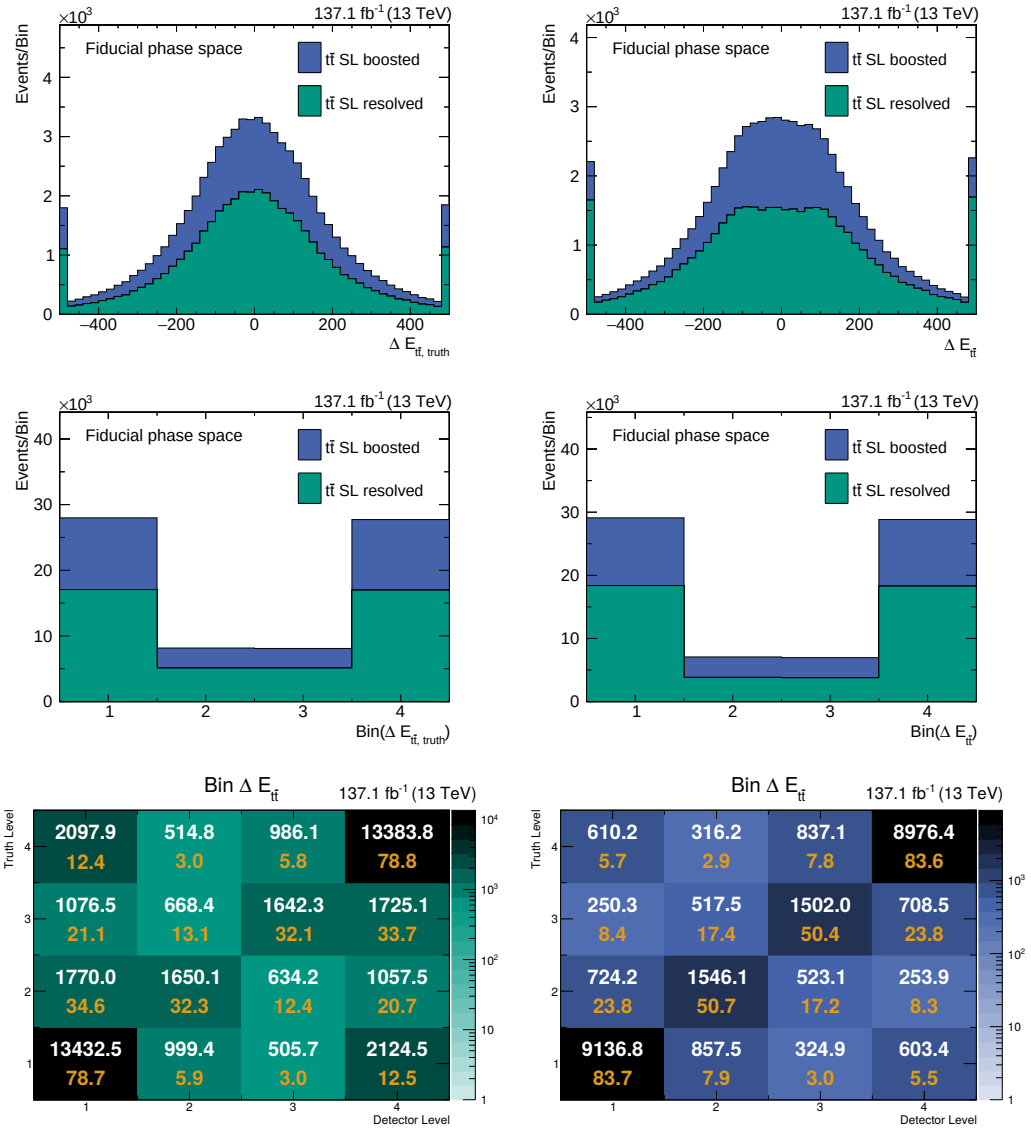

Figure 5.8: Distributions and migration matrices for $\Delta E_{t\bar{t}}$: The upper row shows the distributions for the variable of interest $\Delta E_{t\bar{t}}$ with a granular binning on truth level (left) and detector level (right). The same distributions are shown in the middle row for a coarse binning scheme as detailed in the text. This binning scheme is also deployed for the migration matrices in the lower row, which show the migration of events between truth level and detector level for the signal events in the resolved (left) and boosted regime (right). The white numbers correspond to the event yields, while the orange numbers show the percentage of events of one truth-level bin in the different detector-level bins and are normalized to 100 % for each row.

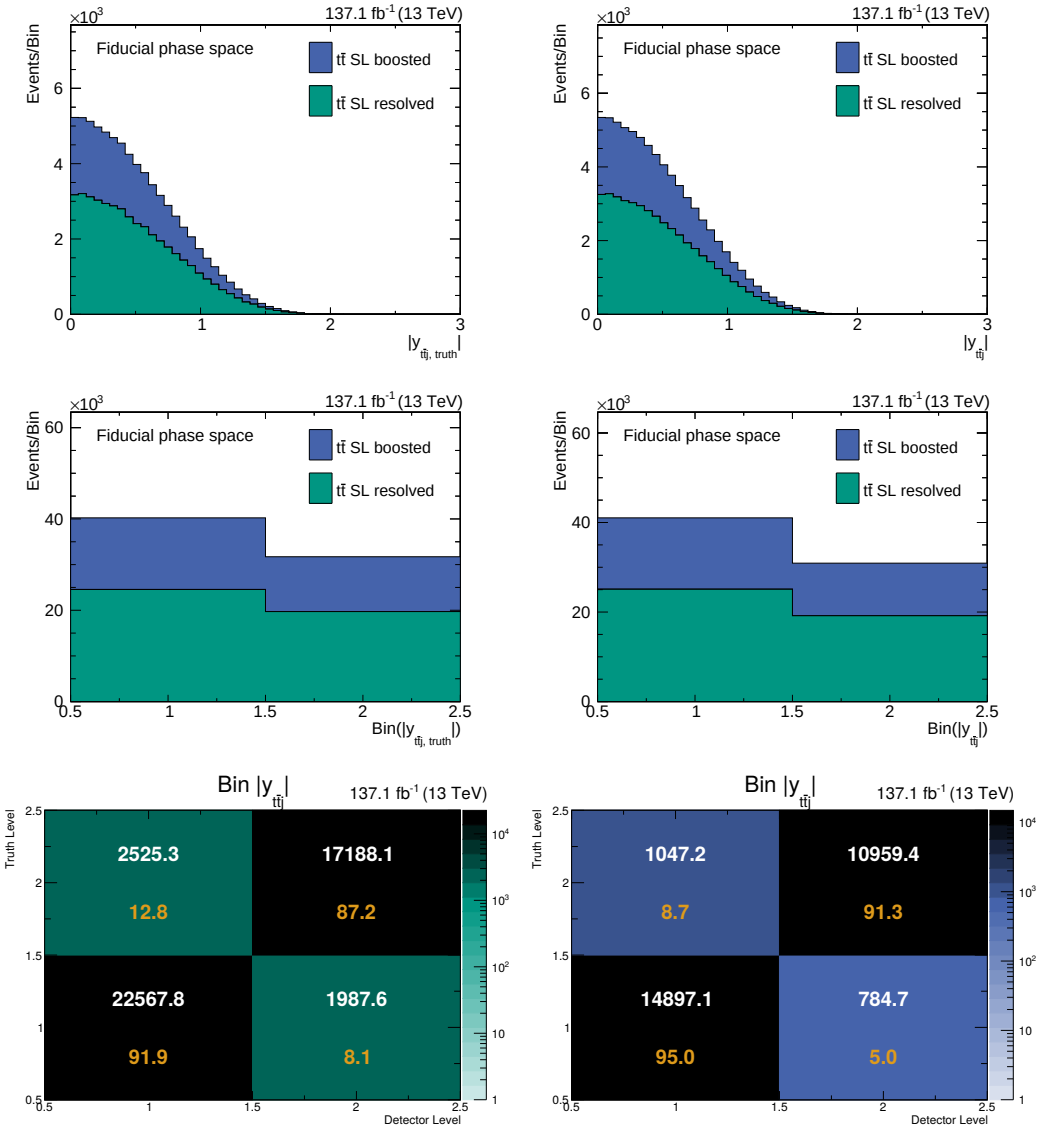

Figure 5.9: Distributions and migration matrices for $y_{t\bar{t}j}$: The upper row shows the distributions for the variable of interest $y_{t\bar{t}j}$ with a granular binning on truth level (left) and detector level (right). The same distributions are shown in the middle row for a coarse binning scheme as detailed in the text. This binning scheme is also deployed for the migration matrices in the lower row, which show the migration of events between truth level and detector level for the signal events in the resolved (left) and boosted regime (right). The white numbers correspond to the event yields, while the orange numbers show the percentage of events of one truth-level bin in the different detector-level bins and are normalized to 100 % for each row.

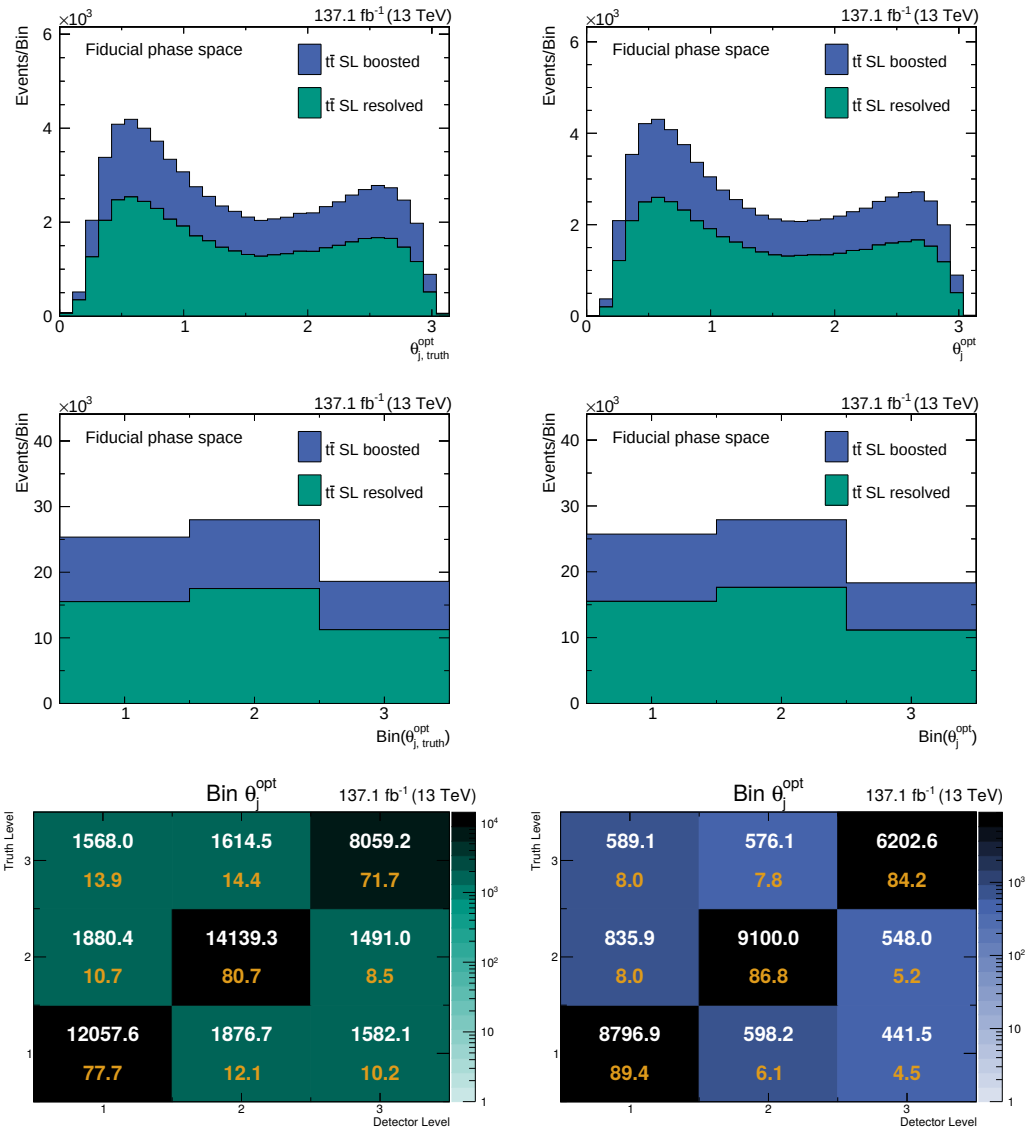

Figure 5.10: Distributions and migration matrices for θ_j^{opt}: The upper row shows the distributions for the variable of interest $y_{t\bar{t}j}$ with a granular binning on truth level (left) and detector level (right). The same distributions are shown in the middle row for a coarse binning scheme as detailed in the text. This binning scheme is also deployed for the migration matrices in the lower row, which show the migration of events between truth level and detector level for the signal events in the resolved (left) and boosted regime (right). The white numbers correspond to the event yields, while the orange numbers show the percentage of events of one truth-level bin in the different detector-level bins and are normalized to 100 % for each row.

with $|\Delta E_{t\bar{t}}| > 50\,\text{GeV}$, this probability is in the order of 80 %, which further promotes the selection requirement on $|\Delta E_{t\bar{t}}|$ in the fiducial phase space. The observable $|y_{t\bar{t}j}|$ is less sensitive to misreconstruction, which is indicated through the high probabilities in the migration matrices, reaching from 87 % to 95 %, to be in the same bin on truth level and detector level. As $y_{t\bar{t}j}$ is the rapidity of the center-of-mass system of the $t\bar{t}j$ system with respect to the laboratory frame, its value only depends on the selection of objects for the reconstruction algorithms, but not on the assignment of these objects to the final-state particles. While swapping the two slim jets used for the reconstruction of j_{add} and the b quark from the t_{lep} decay can have high impact on both $\Delta E_{t\bar{t}}$ and θ_j^{opt}, this does not affect the value of $y_{t\bar{t}j}$ at all, making it a robust observable against misreconstructions in that sense. The migration effects for θ_j^{opt} are less dominant than for $\Delta E_{t\bar{t}}$, but show a high sensitivity to the employed reconstruction algorithm on detector level. The migration into different bins is higher for the reconstruction in the resolved regime, which is expected due to the higher jet multiplicity in this regime and the correspondingly higher probability for misreconstructions due to an incorrect assignment of the slim jets.

The presented migration matrices support the need for unfolding and would be the basic input for a separate unfolding in each of the three variables of interest. However, such a separate unfolding is not desired as it would not allow to determine the energy asymmetry value in the fiducial phase space. Any event-specific information vanishes in the unfolding procedure as the result consists of pure shape information and it is thus not possible to apply a subsequent selection requirement on any observable other than the one unfolded. To avoid this undesired restriction, a dedicated binning, which gives access to all relevant information for the fiducial phase space, needs to be set up prior to the unfolding. The most intuitive choice would be to configure this binning in a way, that no information about the respective bin affiliation for the three variables $\Delta E_{t\bar{t}}$, $|y_{t\bar{t}j}|$, and θ_j^{opt} is lost in the unfolding procedure. This would result in a binning scheme with 24 bins, each of them representing a dedicated configuration out of all possibilities that are given with four bins for $\Delta E_{t\bar{t}}$, three bins for θ_j^{opt}, and two bins for $|y_{t\bar{t}j}|$. Such a binning scheme has been implemented and evaluated on the Asimov data set, showing a tight constraint on many systematic uncertainties and yielding a high uncertainty on the energy asymmetry due to the limited size of simulation samples under employment of the Barlow-Beeston-lite approach. This binning scheme has therefore been disfavored and instead a binning is chosen that gives access only to the most relevant information. For both truth level and detector level, it is in the fiducial phase space only of interest, whether an event has $0.3\pi < \theta_j^{\text{opt}} < 0.7\pi$, $|y_{t\bar{t}j}| > 0.5$, and $|\Delta E_{t\bar{t}}| > 50\,\text{GeV}$, and if so, whether $\Delta E_{t\bar{t}}$ is positive or negative. The corresponding binning is detailed in Table 5.14 and the distributions on truth level and detector level, as well as the migration matrices for the two signal samples, are shown in Figure 5.11. Due to the condition numbers of the two migration matrices being both smaller than 100, regularization is not required in the unfolding procedure.

Each bin of this binning scheme will in the following be denoted as the "unfolding bin" and will be used for the unfolding of the energy asymmetry to truth level. The basic principle of the unfolding procedure via maximum likelihood estimation has already been discussed in Section 3.2 and will be detailed in the following for the determination of the energy

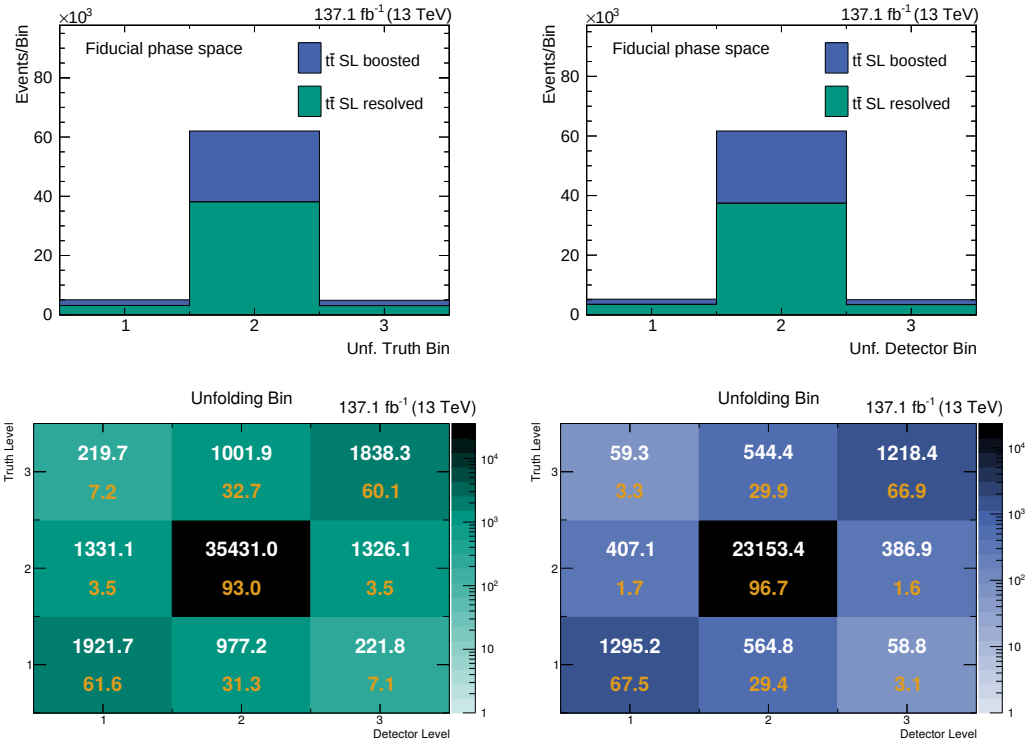

Figure 5.11: Distributions and migration matrices for the unfolding binning: The upper row shows the distributions for the unfolding binning on truth level (left) and detector level (right). This binning scheme is also deployed for the migration matrices in the lower row, which show the migration of events between detector level and truth level for the signal events in the resolved (left) and boosted regime (right). The white numbers correspond to the event yields, while the orange numbers show the percentage of events of one truth-level bin in the different detector-level bins and are normalized to 100 % for each row.

Table 5.14: Binning scheme for the unfolding procedure. Three different bins are defined in order to have a minimum set of bins that provides full access to the relevant information on the three variables of interest for the studied fiducial phase space of the energy asymmetry.

Bin 1	Bin 2	Bin 3
$0.3\pi < \theta_j^{\text{opt}} < 0.7\pi$	$0 < \theta_j^{\text{opt}} < 0.3\pi$ **OR** $0.7\pi < \theta_j^{\text{opt}} < \pi$	$0.3\pi < \theta_j^{\text{opt}} < 0.7\pi$
AND	**OR**	**AND**
$\lvert y_{\mathrm{t\bar{t}}j}\rvert > 0.5$	$\lvert y_{\mathrm{t\bar{t}}j}\rvert < 0.5$	$\lvert y_{\mathrm{t\bar{t}}j}\rvert > 0.5$
AND	**OR**	**AND**
$\Delta E_{\mathrm{t\bar{t}}} < -50\,\text{GeV}$	$-50\,\text{GeV} < \Delta E_{\mathrm{t\bar{t}}} < 50\,\text{GeV}$	$\Delta E_{\mathrm{t\bar{t}}} > 50\,\text{GeV}$

asymmetry. The basic idea in the unfolding procedure is to treat every truth-level bin of the observable to be unfolded as one separate process that contributes to all detector-level bins of the observable. The respective contribution of a truth-level bin to a detector-level bin is given by the elements of the migration matrix. This splitting of the detector-level distribution into different processes based on the truth-level bin contents is shown for the unfiltered signal sample in Figure 5.12.

The distribution of the unfolding bin on detector level for the full simulation and data is shown in Figure 5.13, with the two signal samples being split up according to the contributions of the truth-level bins. The unfolding is performed as a multidimensional maximum likelihood fit of the simulation to observed data with three signal strength parameters to be determined in the fit and the systematic uncertainties considered as nuisance parameters. The purpose of the signal strength parameters is to vary the respective contributions of the three different truth-level bins and to determine the unfolded content of each unfolding bin. Under the assumption of a signal strength parameter r_{neg}, which scales the contributions of truth-level bin 1 ($\Delta E_{\mathrm{t\bar{t}}}$ negative), and a signal strength parameter r_{pos}, which scales the contributions of truth-level bin 3 ($\Delta E_{\mathrm{t\bar{t}}}$ positive), the unfolded bin contents (N_{unf}) of bin 1 and bin 3 of the unfolded distribution are given as:

$$N_{\text{unf}}(\text{Bin 1}) = r_{\text{neg}} \cdot N_{\text{truth}}(\text{Bin 1}) \qquad \text{and}$$
$$N_{\text{unf}}(\text{Bin 3}) = r_{\text{pos}} \cdot N_{\text{truth}}(\text{Bin 3})\,.$$

The unfolded value of the energy asymmetry is given as:

$$A_{E,\text{unf.}}^{\text{opt}} = \frac{N_{\text{unf}}(\text{Bin 1}) - N_{\text{unf}}(\text{Bin 3})}{N_{\text{unf}}(\text{Bin 1}) + N_{\text{unf}}(\text{Bin 3})}\,, \tag{5.14}$$

which can be expressed as:

$$A_{E,\text{unf.}}^{\text{opt}} = \frac{r_{\text{neg}} \cdot N_{\text{truth}}(\text{Bin 1}) - r_{\text{pos}} \cdot N_{\text{truth}}(\text{Bin 3})}{r_{\text{neg}} \cdot N_{\text{truth}}(\text{Bin 1}) + r_{\text{pos}} \cdot N_{\text{truth}}(\text{Bin 3})}\,. \tag{5.15}$$

As the truth-level bin contents are fixed numbers, $A_{E,\text{unf.}}^{\text{opt}}$ varies only with the values of r_{neg} and r_{pos}, which are determined in the fit. The values and uncertainties of r_{neg} and r_{pos} are not of further interest for the result of the analysis and it is thus desirable to have the value of $A_{E,\text{unf.}}^{\text{opt}}$ directly determined in the fit instead. This can be achieved by rearranging the equation above to:

$$r_{\text{pos}} = r_{\text{neg}} \cdot \frac{N_{\text{truth}}(\text{Bin 1})}{N_{\text{truth}}(\text{Bin 3})} \cdot \frac{1 + A_{E,\text{unf.}}^{\text{opt}}}{1 - A_{E,\text{unf.}}^{\text{opt}}}\,. \tag{5.16}$$

The three signal strength parameters for the maximum likelihood fit are then given by:

$\mathbf{r}_{\text{neg}}$: This signal strength parameter scales the contributions from truth-level bin 1 of both signal samples directly and in addition also the contributions from truth-level bin 3 of both signal samples via Equation 5.16. Its prefit value is set to $r_{\text{neg}} = 1.0$.

$\mathbf{r}_{\text{t}\bar{\text{t}}}$: This signal strength parameter scales the contributions from truth-level bin 2 of both signal samples and the contributions from the two $\text{t}\bar{\text{t}}$ background processes with fullhadronic and dileptonic decay of the $\text{t}\bar{\text{t}}$ system. Its prefit value is set to $r_{\text{t}\bar{\text{t}}} = 1.0$.

$\mathbf{r}_{\text{Asym}}$: This signal strength parameter is replacing $A_{E,\text{unf.}}^{\text{opt}}$ in Equation 5.16 and scaling the contributions from truth-level bin 3 of both signal samples via Equation 5.16. Its prefit value is set to $r_{\text{Asym}} = -0.01588$, which corresponds to the truth-level value of $A_E^{\text{opt}} = -1.588\,\%$ in the fiducial phase space as obtained in Section 5.4.6.

This parametrization of the signal strength parameters for the maximum likelihood fit has the additional advantage that the respective uncertainties on $A_{E,\text{unf.}}^{\text{opt}}$ due to the nuisance parameters are directly determined in the fitting procedure. The contributions from background processes are considered in the fitting procedure as well, which corresponds to a direct background subtraction to measured data. It should be noted at this stage that no additional signal strength parameters are induced for the background processes other than $\text{t}\bar{\text{t}}$ production. The contributions from the remaining background processes to bin 1 and bin 3 of the unfolding bin distribution are negligible and hence there is no relevant shape information for these background processes, which would be required in order to constrain

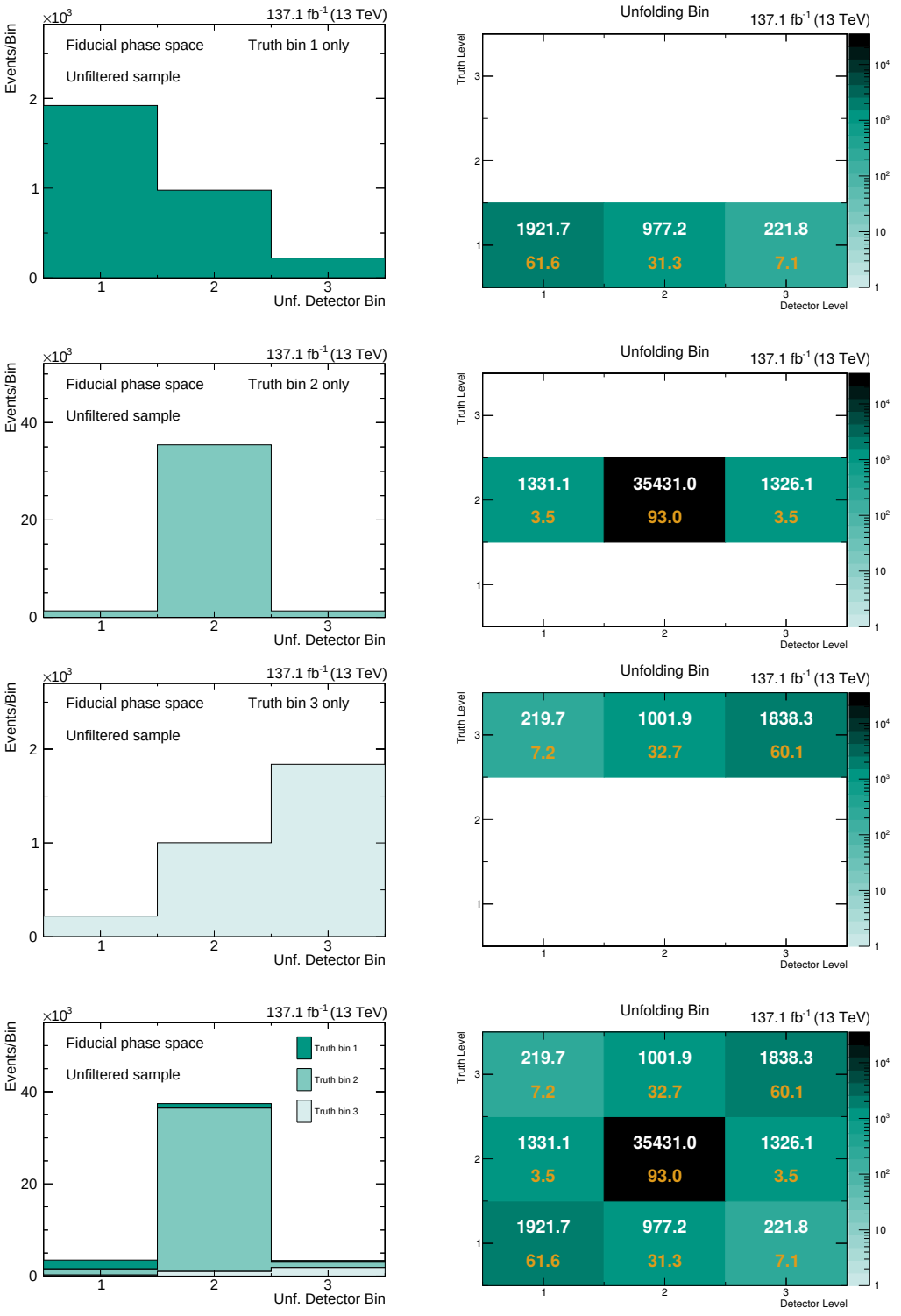

Figure 5.12: Setup of the unfolding procedure: The figure indicates how the unfolding procedure is being set up. The upper three rows indicate the treatment of each truth-level bin (row in the migration matrix) as a separate signal process. The lower row shows the full distribution on detector level split up by the different truth-level bins.

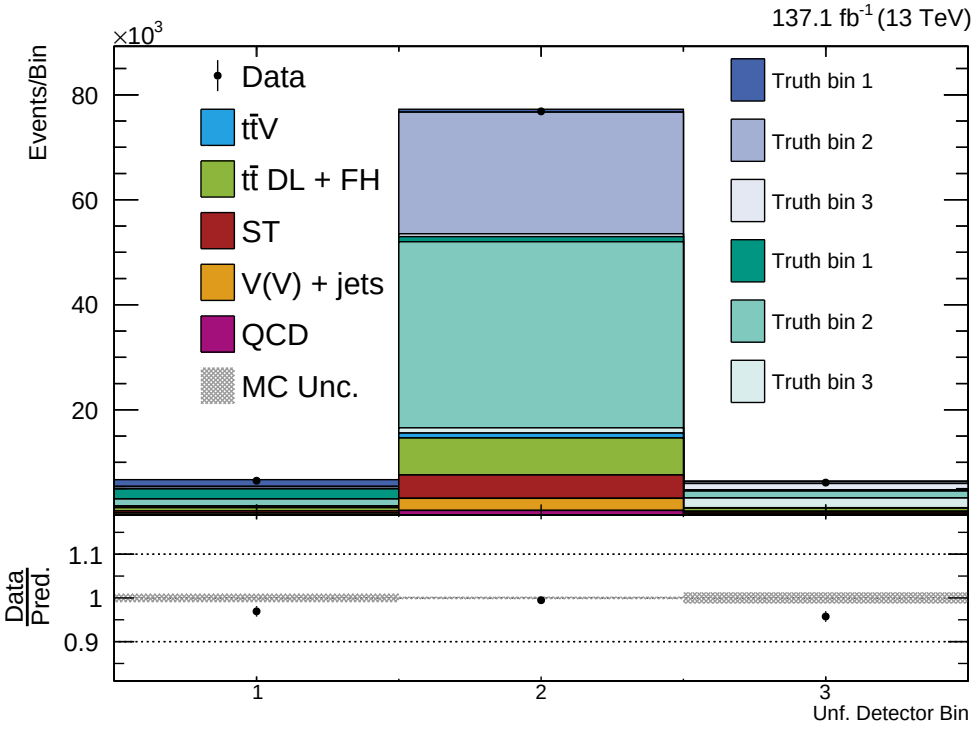

Figure 5.13: Distribution of the unfolding bin: The distribution of the unfolding bin is shown for simulation and measured data. For the simulation, the two signal samples are split by the truth-level contributions and the contents of the respective truth-level bins 1, 2, and 3 are indicated separately for the filtered (dark blue) and unfiltered (dark green) sample.

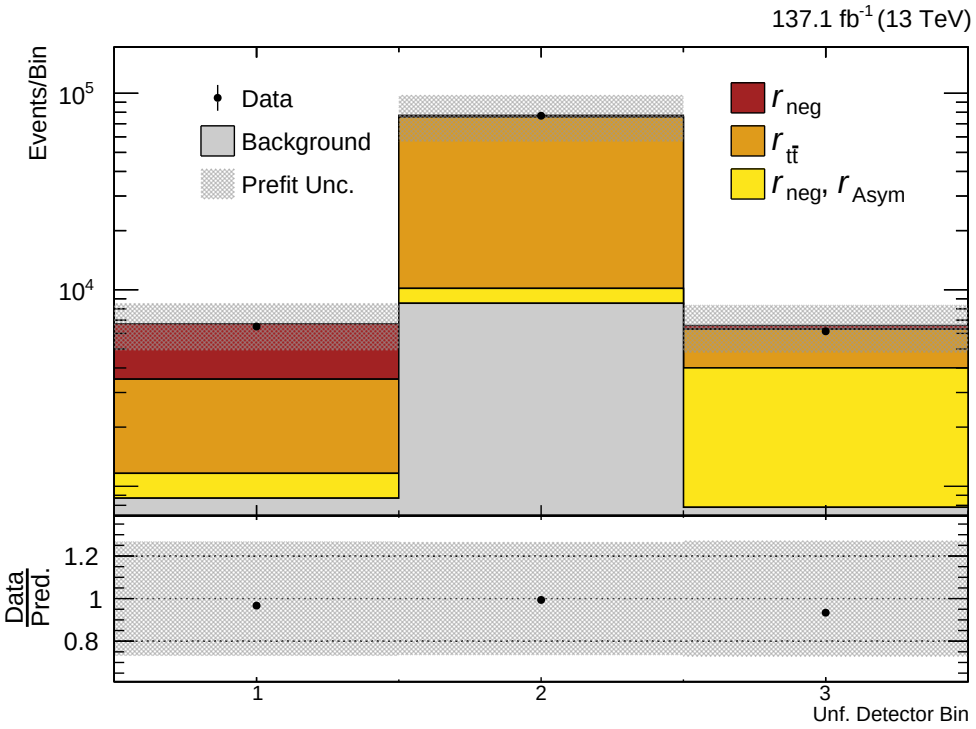

Figure 5.14: Prefit distribution of the unfolding bin: The distribution of the unfolding bin for simulation and measured data is shown together with the prefit uncertainty on simulation due the systematic uncertainty sources. The relevant signal strength parameters for the different contributions as detailed in the text are indicated. The signal strength parameters are set to their prefit values, being $r_{\text{neg}} = 1.0$, $r_{\text{t}\bar{\text{t}}} = 1.0$, and $r_{\text{Asym}} = -0.01588$.

their contributions in the fit. As outlined in Section 5.6, uncertainties on the normalization of the background processes are introduced in the fit as scale uncertainties. Figure 5.14 shows the prefit distribution of the unfolding bin and indicates which contributions are scaled by the three signal strength parameters, respectively.

5.8 Results

The multidimensional maximum likelihood fit with the three signal strength parameters r_{neg}, $r_{\text{t}\bar{\text{t}}}$, and r_{Asym} on the unfolding bin distribution results in the following postfit values:

$$r_{\text{neg}} = \quad 0.951\,^{+0.026}_{-0.025}\,(\text{stat})\,^{+0.293}_{-0.249}\,(\text{syst})\,,$$

$$r_{\text{t}\bar{\text{t}}} = \quad 1.003\,^{+0.004}_{-0.004}\,(\text{stat})\,^{+0.282}_{-0.242}\,(\text{syst})\,,$$

$$r_{\text{Asym}} = -0.02990\,^{+0.02099}_{-0.02102}\,(\text{stat})\,^{+0.03366}_{-0.05080}\,(\text{syst})\,.$$

The corresponding postfit distribution of the unfolding bin is shown in Figure 5.15, which shows very good agreement between simulation and data and a smaller impact from uncertainties in comparison to the prefit distribution in Figure 5.14. Combining both statistical and systematic uncertainties, the unfolded result for the energy asymmetry in the fiducial phase space, which is obtained directly from the fit, is given as:

$$A^{\text{opt}}_{E,\text{unf.}} = -2.99\,\%\,^{+3.97\,\%}_{-5.50\,\%}\,(\text{stat + syst})\,. \tag{5.17}$$

This result is in agreement with the SM predictions as determined in Table 5.8, but does not provide the required precision in order to make a statement about contributions from either BSM physics or SMEFT effects as discussed in Ref. [31].

One of the main advantages in the measurement of asymmetries in comparison to other precision measurements is the damping of systematic uncertainties, which can occur if an uncertainty source has a similar impact on both event categories that enter the asymmetry calculation. This is estimated to reduce especially the impact of the experimental uncertainties to the final result. In Figure 5.7 the truth-level variation on the event yields and the corresponding energy asymmetry values for the up and down shifts of all systematic uncertainties that cause a shape variation in the unfolding bin distribution have been shown. These distributions and the corresponding shifted energy asymmetry values are not directly transferable to the signal strength parameters of the multidimensional maximum likelihood fit, as the correlations between the signal strength parameters need to be taken into account additionally. Nevertheless they indicate the sensitivity of the energy asymmetry to different sources of systematic uncertainty and give an estimation for the impact of each individual source.

It is notable that within this consideration the expected impact on the energy asymmetry is higher for the theoretical uncertainty sources in comparison to the experimental ones. This is also confirmed by the individual impacts of the uncertainty sources on the signal strength parameters as obtained in the unfolding fitting procedure, which are shown in Table 5.15.

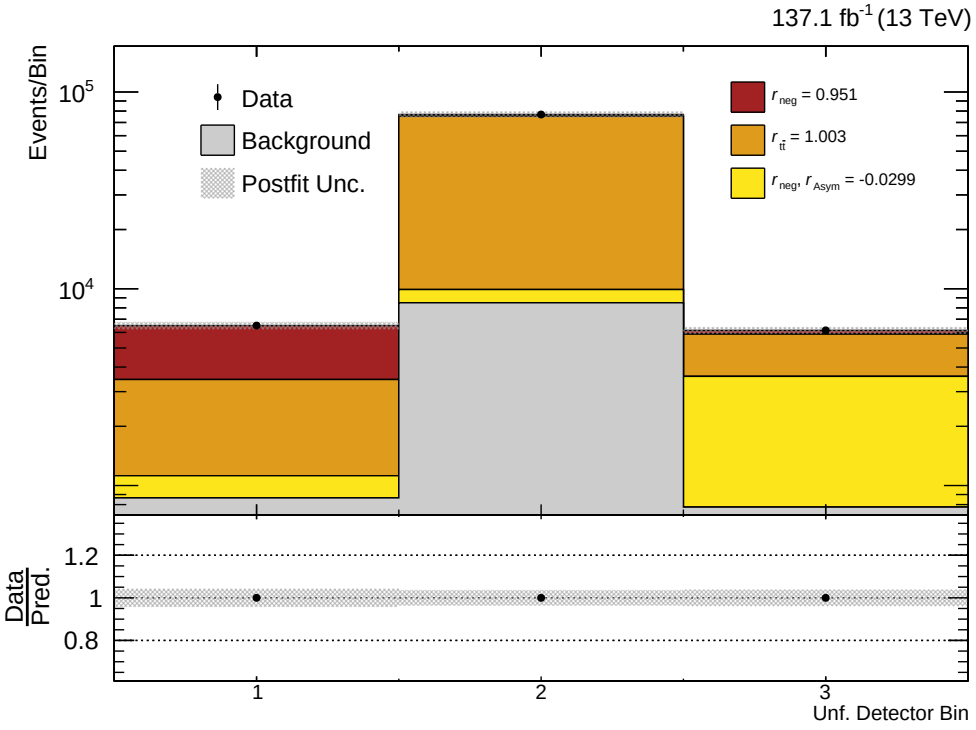

Figure 5.15: Postfit distribution of the unfolding bin: The distribution of the unfolding bin for simulation and measured data is shown together with the postfit uncertainty on simulation due to the systematic uncertainty sources. The relevant signal strength parameters for the different contributions as detailed in the text are indicated and the contributions are scaled with the corresponding postfit values, respectively.

The experimental uncertainty sources mostly have a rather small impact on the signal strength parameters r_{neg} and $r_{\text{t}\bar{\text{t}}}$, while the impact on r_{Asym} is strong only for dedicated sources. Notable is here the limited size of the simulation samples in unfolding bin 1 and unfolding bin 3, as well as the uncertainties on the jet energy resolution and the unclustered energy. With the exception of the $\mu_{\text{R}}/\mu_{\text{F}}$ scale, the impact of the theoretical uncertainty sources on the signal strength parameters r_{neg} and $r_{\text{t}\bar{\text{t}}}$ is small.

The parameter r_{Asym} shows higher sensitivity to the theoretical uncertainty sources, worth mentioning are especially the underlying event tune, PDF+α_{S}, and h_{damp}. This introduces the energy asymmetry as an additional important observable for the tuning of the corresponding parameters in the simulation of events. The most dominating uncertainty source is h_{damp}, which is expected due to the importance of the additional jet kinematic properties in $\text{t}\bar{\text{t}}j$ production and the role of h_{damp} in the matching of the parton shower to the matrix element and the regulation of high-p_{T} parton radiation.

Table 5.15: Impact of systematic uncertainties on signal strength parameters. For each signal strength parameter the up and down shift, as well as the relative change in %, is given for each systematic uncertainty source. The values of r_{Asym} are multiplied by 100 in order to be directly comparable to A_E^{opt}.

Uncertainty source	r_{neg}		$r_{\mathrm{t\bar{t}}}$		$100 \cdot r_{\mathrm{Asym}}$	
Experimental uncertainties						
Pileup	$0.951^{-0.009}_{+0.009}$	$-0.9\,\%$ $+0.9\,\%$	$1.003^{+0.000}_{-0.001}$	$+0.0\,\%$ $-0.1\,\%$	$-2.990^{+0.393}_{-0.483}$	$+13.2\,\%$ $-16.2\,\%$
Lepton efficiencies	$0.951^{-0.014}_{+0.008}$	$-1.4\,\%$ $+0.8\,\%$	$1.003^{-0.009}_{+0.004}$	$-0.9\,\%$ $+0.4\,\%$	$-2.990^{-0.188}_{-0.017}$	$-6.3\,\%$ $-0.6\,\%$
b tagging efficiency	$0.951^{-0.029}_{+0.022}$	$-3.0\,\%$ $+2.3\,\%$	$1.003^{-0.018}_{+0.011}$	$-1.8\,\%$ $+1.1\,\%$	$-2.990^{-0.085}_{-0.141}$	$-2.8\,\%$ $-4.7\,\%$
top tagging efficiency	$0.951^{-0.015}_{+0.010}$	$-1.5\,\%$ $+1.0\,\%$	$1.003^{-0.013}_{+0.008}$	$-1.2\,\%$ $+0.8\,\%$	$-2.990^{-0.236}_{-0.032}$	$-7.9\,\%$ $-1.1\,\%$
JES	$0.951^{-0.052}_{+0.036}$	$-5.5\,\%$ $+3.8\,\%$	$1.003^{-0.035}_{+0.027}$	$-3.5\,\%$ $+2.7\,\%$	$-2.990^{-0.048}_{-0.168}$	$-1.6\,\%$ $-5.6\,\%$
JER	$0.951^{-0.052}_{+0.002}$	$-5.5\,\%$ $+0.2\,\%$	$1.003^{-0.011}_{-0.001}$	$-1.1\,\%$ $-0.1\,\%$	$-2.990^{+1.800}_{-0.559}$	$+60.2\,\%$ $-18.7\,\%$
Unclustered Energy	$0.951^{-0.009}_{-0.006}$	$-0.9\,\%$ $-0.6\,\%$	$1.003^{-0.000}_{+0.000}$	$-0.0\,\%$ $+0.0\,\%$	$-2.990^{+0.453}_{+0.511}$	$+15.2\,\%$ $+17.1\,\%$
Luminosity	$0.951^{-0.012}_{+0.005}$	$-1.2\,\%$ $+0.5\,\%$	$1.003^{-0.011}_{+0.005}$	$-1.1\,\%$ $+0.5\,\%$	$-2.990^{-0.227}_{-0.032}$	$-7.6\,\%$ $-1.1\,\%$
BB-lite (Bin 1)	$0.951^{-0.019}_{+0.019}$	$-2.0\,\%$ $+2.0\,\%$	$1.003^{+0.001}_{-0.001}$	$+0.1\,\%$ $-0.1\,\%$	$-2.990^{+1.071}_{-1.172}$	$+35.8\,\%$ $-39.2\,\%$
BB-lite (Bin 2)	$0.951^{+0.005}_{-0.005}$	$+0.5\,\%$ $-0.5\,\%$	$1.003^{-0.001}_{+0.000}$	$-0.1\,\%$ $+0.0\,\%$	$-2.990^{-0.062}_{-0.052}$	$-2.1\,\%$ $-1.7\,\%$
BB-lite (Bin 3)	$0.951^{+0.001}_{-0.001}$	$+0.2\,\%$ $-0.1\,\%$	$1.003^{-0.000}_{-0.000}$	$-0.0\,\%$ $-0.0\,\%$	$-2.990^{-1.607}_{+1.466}$	$-53.8\,\%$ $+49.0\,\%$
Theoretical uncertainties						
PDF+α_{S}	$0.951^{-0.025}_{+0.014}$	$-2.6\,\%$ $+1.5\,\%$	$1.003^{-0.023}_{+0.013}$	$-2.3\,\%$ $+1.3\,\%$	$-2.990^{-1.489}_{+1.508}$	$-49.8\,\%$ $+50.4\,\%$
ISR	$0.951^{-0.003}_{-0.004}$	$-0.4\,\%$ $-0.4\,\%$	$1.003^{+0.004}_{-0.011}$	$+0.3\,\%$ $-1.1\,\%$	$-2.990^{+0.072}_{-0.298}$	$+2.4\,\%$ $-10.0\,\%$
FSR	$0.951^{+0.000}_{-0.009}$	$+0.0\,\%$ $-0.9\,\%$	$1.003^{-0.006}_{+0.008}$	$-0.6\,\%$ $+0.8\,\%$	$-2.990^{+0.153}_{-0.065}$	$+5.1\,\%$ $-2.2\,\%$
$\mu_{\mathrm{R}}/\mu_{\mathrm{F}}$ scale	$0.951^{-0.218}_{+0.239}$	$-22.9\,\%$ $+25.1\,\%$	$1.003^{-0.214}_{+0.234}$	$-21.3\,\%$ $+23.3\,\%$	$-2.990^{-0.256}_{+0.063}$	$-8.6\,\%$ $+2.1\,\%$
h_{damp}	$0.951^{-0.002}_{+0.046}$	$-0.2\,\%$ $+4.8\,\%$	$1.003^{-0.046}_{+0.039}$	$-4.6\,\%$ $+3.9\,\%$	$-2.990^{-0.173}_{-4.049}$	$-5.8\,\%$ $-135.4\,\%$
Underlying event	$0.951^{+0.024}_{+0.014}$	$+2.5\,\%$ $+1.5\,\%$	$1.003^{-0.001}_{+0.000}$	$-0.1\,\%$ $+0.0\,\%$	$-2.990^{+0.943}_{-1.165}$	$+31.5\,\%$ $-38.9\,\%$
σ(TT Semilep HT250)	$0.951^{-0.000}_{-0.003}$	$-0.0\,\%$ $-0.3\,\%$	$1.003^{-0.006}_{+0.003}$	$-0.6\,\%$ $+0.3\,\%$	$-2.990^{-0.066}_{-0.080}$	$-2.2\,\%$ $-2.7\,\%$
σ(TT Semilep)	$0.951^{-0.014}_{+0.008}$	$-1.5\,\%$ $+0.8\,\%$	$1.003^{-0.011}_{+0.005}$	$-1.1\,\%$ $+0.5\,\%$	$-2.990^{-0.268}_{+0.026}$	$-8.9\,\%$ $+0.9\,\%$
σ(TT FH + DL)	$0.951^{-0.001}_{-0.000}$	$-0.1\,\%$ $-0.0\,\%$	$1.003^{-0.001}_{+0.000}$	$-0.1\,\%$ $+0.0\,\%$	$-2.990^{-0.084}_{-0.031}$	$-2.8\,\%$ $-1.0\,\%$
σ(Single Top)	$0.951^{-0.003}_{-0.001}$	$-0.3\,\%$ $-0.1\,\%$	$1.003^{-0.003}_{+0.001}$	$-0.3\,\%$ $+0.1\,\%$	$-2.990^{-0.158}_{+0.049}$	$-5.3\,\%$ $+1.6\,\%$
σ(V + V/Jets)	$0.951^{-0.004}_{+0.003}$	$-0.4\,\%$ $+0.3\,\%$	$1.003^{-0.001}_{-0.000}$	$-0.1\,\%$ $-0.0\,\%$	$-2.990^{+0.028}_{-0.138}$	$+0.9\,\%$ $-4.6\,\%$
σ(QCD, TT + X)	$0.951^{-0.011}_{+0.005}$	$-1.2\,\%$ $+0.5\,\%$	$1.003^{-0.003}_{+0.000}$	$-0.3\,\%$ $+0.0\,\%$	$-2.990^{+0.065}_{-0.132}$	$+2.2\,\%$ $-4.4\,\%$

6. Conclusion

The measurement of the energy asymmetry in the production of a top quark pair in association with one additional high-p_{T} jet has been presented in this thesis. Data from proton-proton collisions at a center-of-mass energy of 13 TeV, collected with the CMS detector in the years 2016, 2017, and 2018, corresponding to a data set of 137.1 fb^{-1}, have been analyzed. This analysis provides the first result of the energy asymmetry and yields an observed value of

$$A_{E,\mathrm{unf.}}^{\mathrm{opt}} = -2.99\,\%\ ^{+3.97\,\%}_{-5.50\,\%}\,(\mathrm{stat} + \mathrm{syst})$$

in the fiducial phase space, which is in good agreement with the SM prediction.

The presented work introduces the energy asymmetry as an important observable in the measurement of top quark properties and demonstrates its capability to be a successor of the previously studied asymmetries in top quark pair production. It especially extends the phase space of asymmetry measurements to $\mathrm{t\bar{t}}j$ production as additional final state to pure $\mathrm{t\bar{t}}$ production. It has been shown that the proposed phase space from theory predictions [31] can be extended in a way that the amount of available data in the fiducial phase space is more than doubled with only a small decrease of the effect's magnitude. This extension was crucial for the reachable precision in the presented measurement and will significantly improve the sensitivity of the measurement after inclusion of the LHC Run 3 data in comparison to the predictions in Ref. [31].

A further improvement of the energy asymmetry measurement with the amount of the LHC Run 2 data could be obtained by the inclusion of the fullhadronic and dileptonic decay channel of the top quark pair in $\mathrm{t\bar{t}}j$ production. As the variables of interest, as for example $\Delta E_{\mathrm{t\bar{t}}}$ and θ_j^{opt}, are very sensitive to misreconstruction effects, the development of dedicated and efficient reconstruction algorithms is required. A better understanding of

the measured value of the energy asymmetry can be obtained by elaborate studies on the contributions of background processes, which requires a high amount of simulated events in the fiducial phase space.

While not yet being able to make a significant statement about BSM contributions or SMEFT effects, the presented measurement demonstrates the sensitivity of the observable to important theory parameters in the simulation of top quark pair production. The energy asymmetry can hence become an important ingredient for constraining the parameter set of the underlying event tune and the damping parameter h_{damp} in the matching of matrix element and parton shower, which are important systematic uncertainties in many top quark physics analyses.

Apart from theory uncertainties the precision of the measurement is mainly limited due to the available amount of events in the fiducial phase space and the total amount of simulated events. The precision of the analysis will hence significantly improve with the additional amount of 350 fb^{-1} of data, which is expected to be collected during Run 3 of the LHC [178]. After the upgrade of the LHC to the high-luminosity LHC, the designed instantaneous luminosity will increase by a factor of 5 and the recorded data set is desired to be in the order of 3000 fb^{-1} at a center-of-mass energy of 14 TeV. This amount of data will allow for an even more precise measurement of the energy asymmetry and provide the required sensitivity for testing the impact of BSM contributions and SMEFT effects.

A. Simulated Samples and Datasets

A.1 Simulated Samples

Table A.1: Nominal simulation data sets for the analysis for 2016 data. If not stated otherwise, the cross section obtained from the generator is used. The fragment "RunIISummer16NanoAODv5-PUMoriond17_Nano1June2019_102X_mcRun2_asymptotic_v7-v1…" and the postfix "/NANOAODSIM" are omitted everywhere. If no reference is given, the corresponding cross sections are directly obtained from the simulation generator.

Data set name	Events	Cross section × BR (pb)
TTToSemiLeptonic_TuneCP5_PSweights_13TeV-powheg-pythia8	107 305 100	365.3 (NNLO[18])
TTToSemiLeptonic_HT250_TuneCP5_PSweights_13TeV-powheg-pythia8	28 403 407	34.6 (NNLO)
TTToHadronic_TuneCP5_PSweights_13TeV-powheg-pythia8	68 518 800	378.2 (NNLO[18])
TTTo2L2Nu_TuneCP5_PSweights_13TeV-powheg-pythia8	65 852 400	88.2 (NNLO[18])
TTWJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	3 120 397	0.18 (NLO)
TTWJetsToQQ_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	833 298	0.37 (NLO)
TTZToLLNuNu_M-10_TuneCUETP8M1_13TeV-amcatnlo-pythia8	5 934 228	0.26 (NLO)
TTZToQQ_TuneCUETP8M1_13TeV-amcatnlo-pythia8	749 400	0.6 (NLO)
DYJetsToLL_M-50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	120 777 245	5765.4 (NNLO [179])
ST_tW_top_5f_inclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8	4 983 500	35.9 (NNLO[20])
ST_tW_antitop_5f_inclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8	4 980 600	35.9 (NNLO[20])
ST_t-channel_top_4f_InclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8	31 848 000	136.02 (NLO[20])
ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8	17 780 700	80.95 (NLO[20])
WJetsToLNu_Pt-100To250_TuneCP5_13TeV-amcatnloFXFX-pythia8	98 585 849	689.7 (NLO)
WJetsToLNu_Pt-250To400_TuneCP5_13TeV-amcatnloFXFX-pythia8	10 021 205	24.5 (NLO)
WJetsToLNu_Pt-400To600_TuneCP5_13TeV-amcatnloFXFX-pythia8	1 939 947	3.11 (NLO)
WJetsToLNu_Pt-600ToInf_TuneCP5_13TeV-amcatnloFXFX-pythia8	1 974 609	0.468 (NLO)
WW_TuneCUETP8M1_13TeV-pythia8	6 988 168	118.7 (NLO)
WZ_TuneCUETP8M1_13TeV-pythia8	2 997 571	65.54 (NLO)
ZZ_TuneCUETP8M1_13TeV-pythia8	998 034	15.83 (NLO)
QCD_HT300to500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	37 516 961	323600.0 (NLO)
QCD_HT500to700_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	44 061 488	29990.0 (NLO)
QCD_HT700to1000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	21 604 533	6351.0 (NLO)
QCD_HT1000to1500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	10 360 193	1039.0 (NLO)
QCD_HT1500to2000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	7 046 372	99.01 (NLO)
QCD_HT2000toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	4 027 896	20.23 (NLO)

Table A.2: Nominal simulation data sets for the analysis for 2017 data. If not stated otherwise, the cross section obtained from the generator is used. The fragment "RunIIFall17NanoAODv5-PU2017_12Apr2018_Nano1June2019_102X_mc2017_realistic_v7-v1…" and the postfix "/NANOAODSIM" are omitted everywhere. If no reference is given, the corresponding cross sections are directly obtained from the simulation generator.

Data set name	Events	Cross section × BR (pb)
TTToSemiLeptonic_TuneCP5_PSweights_13TeV-powheg-pythia8	110 014 744	365.3 (NNLO[18])
TTToSemiLeptonic_HT250_TuneCP5_PSweights_13TeV-powheg-pythia8	28 357 449	34.6 (NNLO)
TTToHadronic_TuneCP5_PSweights_13TeV-powheg-pythia8	130 091 218	378.2 (NNLO[18])
TTTo2L2Nu_TuneCP5_PSweights_13TeV-powheg-pythia8	69 098 644	88.2 (NNLO[18])
TTWJetsToLNu_TuneCP5_PSweights_13TeV-amcatnloFXFX-madspin-pythia8	4 908 905	0.18 (NLO)
TTWJetsToQQ_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8	811 306	0.37 (NLO)
TTZToLLNuNu_M-10_TuneCP5_PSweights_13TeV-amcatnlo-pythia8	11 092 000	0.26 (NLO)
TTZToQQ_TuneCP5_13TeV-amcatnlo-pythia8	8 940 000	0.6 (NLO)
DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8	182 104 014	5765.4 (NNLO [179])
ST_tW_top_5f_NoFullyHadronicDecays_TuneCP5_PSweights_13TeV-powheg-pythia8	5 103 599	35.9 (NNLO[20])
ST_tW_antitop_5f_NoFullyHadronicDecays_TuneCP5_PSweights_13TeV-powheg-pythia8	5 577 319	35.9 (NNLO[20])
ST_t-channel_top_4f_InclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8	122 630 600	136.02 (NLO[20])
ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8	63 620 800	80.95 (NLO[20])
WJetsToLNu_Pt-50To100_TuneCP5_13TeV-amcatnloFXFX-pythia8	18 242 254	3298 (NLO)
WJetsToLNu_Pt-100To250_TuneCP5_13TeV-amcatnloFXFX-pythia8	97 657 266	689.7 (NLO)
WJetsToLNu_Pt-250To400_TuneCP5_13TeV-amcatnloFXFX-pythia8	9 488 289	24.5 (NLO)
WJetsToLNu_Pt-400To600_TuneCP5_13TeV-amcatnloFXFX-pythia8	1 948 286	3.11 (NLO)
WJetsToLNu_Pt-600ToInf_TuneCP5_13TeV-amcatnloFXFX-pythia8	1 958 230	0.468 (NLO)
WW_TuneCP5_13TeV-pythia8	7 765 828	118.7 (NLO)
WZ_TuneCP5_13TeV-pythia8	3 928 630	65.54 (NLO)
ZZ_TuneCP5_13TeV-pythia8	1 949 768	15.83 (NLO)
QCD_HT300to500_TuneCP5_13TeV-madgraph-pythia8	59 569 132	323600.0 (NLO)
QCD_HT500to700_TuneCP5_13TeV-madgraph-pythia8	56 207 744	29990.0 (NLO)
QCD_HT700to1000_TuneCP5_13TeV-madgraph-pythia8	46 840 955	6351.0 (NLO)
QCD_HT1000to1500_TuneCP5_13TeV-madgraph-pythia8	16 882 838	1039.0 (NLO)
QCD_HT1500to2000_TuneCP5_13TeV-madgraph-pythia8	11 634 434	99.01 (NLO)
QCD_HT2000toInf_TuneCP5_13TeV-madgraph-pythia8	5 941 306	20.23 (NLO)

Table A.3: Nominal simulation data sets for the analysis for 2018 data. If not stated otherwise, the cross section obtained from the generator is used. The fragment "RunIIAutumn18NanoAODv5-Nano1June2019_102X_upgrade2018_realistic_v19-v1…" and the postfix "/NANOAODSIM" are omitted everywhere. If no reference is given, the corresponding cross sections are directly obtained from the simulation generator.

Data set name	Events	Cross section × BR (pb)
TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8	101 550 000	365.3 (NNLO[18])
TTToSemiLeptonic_HT250_TuneCP5_13TeV-powheg-pythia8	22 476 261	34.6 (NNLO)
TTToHadronic_TuneCP5_13TeV-powheg-pythia8	199 098 000	378.2 (NNLO[18])
TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8	64 310 000	88.2 (NNLO[18])
TTWJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8	4 911 941	0.18 (NLO)
TTWJetsToQQ_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8	835 296	0.37 (NLO)
TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8	13 280 000	0.26 (NLO)
TTZToQQ_TuneCP5_13TeV-amcatnlo-pythia8	8 891 000	0.6 (NLO)
DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8	193 094 040	5765.4 (NNLO [179])
ST_tW_top_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8	8 722 734	35.9 (NNLO[20])
ST_tW_antitop_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8	6 909 815	35.9 (NNLO[20])
ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-pythia8	154 307 600	136.02 (NLO[20])
ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-pythia8	79 090 800	80.95 (NLO[20])
WJetsToLNu_Pt-50To100_TuneCP5_13TeV-amcatnloFXFX-pythia8	18 999 100	3298 (NLO)
WJetsToLNu_Pt-100To250_TuneCP5_13TeV-amcatnloFXFX-pythia8	98 034 198	689.7 (NLO)
WJetsToLNu_Pt-250To400_TuneCP5_13TeV-amcatnloFXFX-pythia8	9 655 249	24.5 (NLO)
WJetsToLNu_Pt-400To600_TuneCP5_13TeV-amcatnloFXFX-pythia8	1 967 802	3.11 (NLO)
WJetsToLNu_Pt-600ToInf_TuneCP5_13TeV-amcatnloFXFX-pythia8	1 943 949	0.468 (NLO)
WW_TuneCP5_PSweights_13TeV-pythia8	7 958 000	118.7 (NLO)
WZ_TuneCP5_PSweights_13TeV-pythia8	3 822 000	65.54 (NLO)
ZZ_TuneCP5_13TeV-pythia8	1 979 000	15.83 (NLO)
QCD_HT300to500_TuneCP5_13TeV-madgraphMLM-pythia8	54 661 579	323600.0 (NLO)
QCD_HT500to700_TuneCP5_13TeV-madgraphMLM-pythia8	55 152 960	29990.0 (NLO)
QCD_HT700to1000_TuneCP5_13TeV-madgraphMLM-pythia8	48 158 738	6351.0 (NLO)
QCD_HT1000to1500_TuneCP5_13TeV-madgraphMLM-pythia8	15 466 225	1039.0 (NLO)
QCD_HT1500to2000_TuneCP5_13TeV-madgraphMLM-pythia8	10 955 087	99.01 (NLO)
QCD_HT2000toInf_TuneCP5_13TeV-madgraphMLM-pythia8	5 475 677	20.23 (NLO)

A.2 Datasets

Table A.4: The different data taking periods of the 2016 data taking period used for this analysis.

Period	Run Range	Integrated Luminosity
/SingleMuon(SingleElectron)/Run2016B_ver2-Nano1June2019_ver2-v1/NANOAOD	272007–275376	5.8fb^{-1}
/SingleMuon(SingleElectron)/Run2016C-Nano1June2019-v1/NANOAOD	275657–276283	2.6fb^{-1}
/SingleMuon(SingleElectron)/Run2016D-Nano1June2019-v1/NANOAOD	276315–276811	4.2fb^{-1}
/SingleMuon(SingleElectron)/Run2016E-Nano1June2019-v1/NANOAOD	276831–277420	4.0fb^{-1}
/SingleMuon(SingleElectron)/Run2016F-Nano1June2019-v1/NANOAOD	277772–278808	3.1fb^{-1}
/SingleMuon(SingleElectron)/Run2016G-Nano1June2019-v1/NANOAOD	278820–280385	7.5fb^{-1}
/SingleMuon(SingleElectron)/Run2016H-Nano1June2019-v1/NANOAOD	280919–284044	8.6fb^{-1}
Total	272007–284044	35.9fb^{-1}

Table A.5: The different data taking periods of the 2017 data taking period used for this analysis.

Period	Run Range	Integrated Luminosity
/SingleMuon(SingleElectron)/Run2017B-Nano1June2019-v1/NANOAOD	297046–299329	4.8fb^{-1}
/SingleMuon(SingleElectron)/Run2017C-Nano1June2019-v1/NANOAOD	299368–302029	9.7fb^{-1}
/SingleMuon(SingleElectron)/Run2017D-Nano1June2019-v1/NANOAOD	302030–303434	4.3fb^{-1}
/SingleMuon(SingleElectron)/Run2017E-Nano1June2019-v1/NANOAOD	303824–304797	9.3fb^{-1}
/SingleMuon(SingleElectron)/Run2017F-Nano1June2019-v1/NANOAOD	305040–306462	13.5fb^{-1}
Total	297046–306462	41.5fb^{-1}

Table A.6: The different data taking periods of the 2018 data taking period used for this analysis.

Period	Run Range	Integrated Luminosity
/SingleMuon(EGamma)/Run2018A-Nano1June2019-v1/NANOAOD	315252–316995	14.0fb^{-1}
/SingleMuon(EGamma)/Run2018B-Nano1June2019-v1/NANOAOD	317080–319310	7.1fb^{-1}
/SingleMuon(EGamma)/Run2018C-Nano1June2019-v1/NANOAOD	319337–320065	6.9fb^{-1}
/SingleMuon(EGamma)/Run2018D-Nano1June2019-v1/NANOAOD	320673–325175	31.9fb^{-1}
Total	315252–325175	59.7fb^{-1}

A.3 Systematically Varied Simulated Samples

Table A.7: Systematically varied $t\bar{t}$ samples for the 2016 analysis. The fragment "RunIISummer16NanoAODv5-PUMoriond17_Nano1June2019_102X_mcRun2_asymptotic_v7-v1" and the postfix "/NANOAODSIM" are omitted everywhere.

Data set name	Events
h_{damp}	
TTTo2L2Nu_hdampDOWN_TuneCP5_PSweights_13TeV-powheg-pythia8	14 895 500
TTTo2L2Nu_hdampUP_TuneCP5_PSweights_13TeV-powheg-pythia8	14 514 500
TTToSemiLeptonic_hdampDOWN_TuneCP5_PSweights_13TeV-powheg-pythia8	29 770 400
TTToSemiLeptonic_hdampUP_TuneCP5_PSweights_13TeV-powheg-pythia8	29 671 200
Underlying Event	
TTTo2L2Nu_TuneCP5down_PSweights_13TeV-powheg-pythia8	11 151 500
TTTo2L2Nu_TuneCP5up_PSweights_13TeV-powheg-pythia8	13 881 200
TTToSemiLeptonic_TuneCP5down_PSweights_13TeV-powheg-pythia8	23 359 000
TTToSemiLeptonic_TuneCP5up_PSweights_13TeV-powheg-pythia8	21 697 600

Table A.8: Systematically varied $t\bar{t}$ samples for the 2017 analysis. The fragment "RunIIFall17NanoAODv5-PU2017_12Apr2018_Nano1June2019_102X_mc2017_realistic_v7-v1…" and the postfix "/NANOAODSIM" are omitted everywhere.

Data set name	Events
h_{damp}	
TTTo2L2Nu_hdampDOWN_TuneCP5_PSweights_13TeV-powheg-pythia8	5 476 459
TTTo2L2Nu_hdampUP_TuneCP5_PSweights_13TeV-powheg-pythia8	3 288 128
TTToSemiLeptonic_hdampDOWN_TuneCP5_PSweights_13TeV-powheg-pythia8	26 367 765
TTToSemiLeptonic_hdampUP_TuneCP5_PSweights_13TeV-powheg-pythia8	23 977 012
TTToHadronic_hdampDOWN_TuneCP5_PSweights_13TeV-powheg-pythia8	27 117 982
TTToHadronic_hdampUP_TuneCP5_PSweights_13TeV-powheg-pythia8	27 260 880
Underlying Event	
TTTo2L2Nu_TuneCP5down_PSweights_13TeV-powheg-pythia8	5 500 000
TTTo2L2Nu_TuneCP5up_PSweights_13TeV-powheg-pythia8	5 500 000
TTToSemiLeptonic_TuneCP5down_PSweights_13TeV-powheg-pythia8	27 104 055
TTToSemiLeptonic_TuneCP5up_PSweights_13TeV-powheg-pythia8	20 122 010
TTToHadronic_TuneCP5down_PSweights_13TeV-powheg-pythia8	27 252 808
TTToHadronic_TuneCP5up_PSweights_13TeV-powheg-pythia8	27 108 792

Table A.9: Systematically varied $t\bar{t}$ samples for the 2018 analysis. The fragment "RunIIAutumn18NanoAODv5-Nano1June2019_102X_upgrade2018_realistic_v19-v1…" and the postfix "/NANOAODSIM" are omitted everywhere.

Data set name	Events
h_{damp}	
TTTo2L2Nu_hdampDOWN_TuneCP5_13TeV-powheg-pythia8	5 458 000
TTTo2L2Nu_hdampUP_TuneCP5_13TeV-powheg-pythia8	3 700 000
TTToSemiLeptonic_hdampDOWN_TuneCP5_13TeV-powheg-pythia8	25 904 000
TTToSemiLeptonic_hdampUP_TuneCP5_13TeV-powheg-pythia8	26 892 000
TTToHadronic_hdampDOWN_TuneCP5_13TeV-powheg-pythia8	26 425 000
TTToHadronic_hdampUP_TuneCP5_13TeV-powheg-pythia8	24 965 000
Underlying Event	
TTTo2L2Nu_TuneCP5down_13TeV-powheg-pythia8	4 954 000
TTTo2L2Nu_TuneCP5up_13TeV-powheg-pythia8	3 862 000
TTToSemiLeptonic_TuneCP5down_13TeV-powheg-pythia8	18 929 000
TTToSemiLeptonic_TuneCP5up_13TeV-powheg-pythia8	26 876 000
TTToHadronic_TuneCP5down_13TeV-powheg-pythia8	26 675 000
TTToHadronic_TuneCP5up_13TeV-powheg-pythia8	23 488 000

B. BDT Figures of Merit

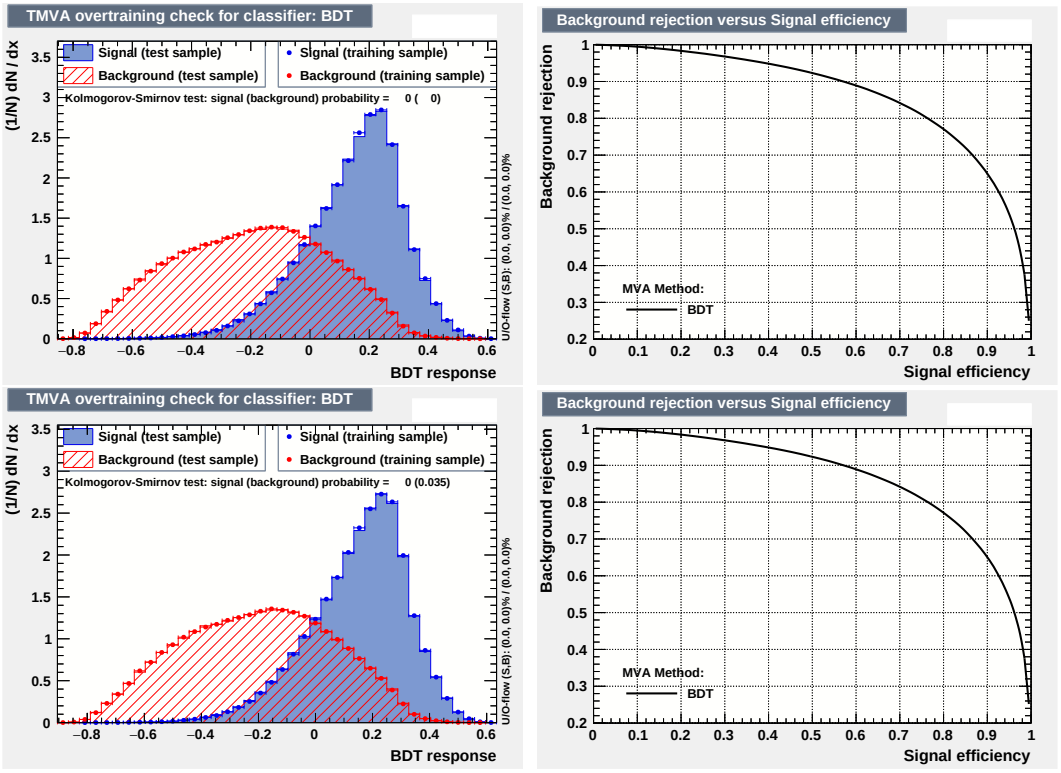

Figure B.1: BDT overtraining check and ROC-curves for 2016: The overtraining checks and the ROC-curves are shown for the two BDTs that are used for the jet assignment in the reconstruction of events in the resolved regime on detector level for simulation samples and data of the 2016 data taking period. The upper row corresponds to the BDT that is trained with even-numbered events, while the lower row corresponds to the BDT that is trained with odd-numbered events.

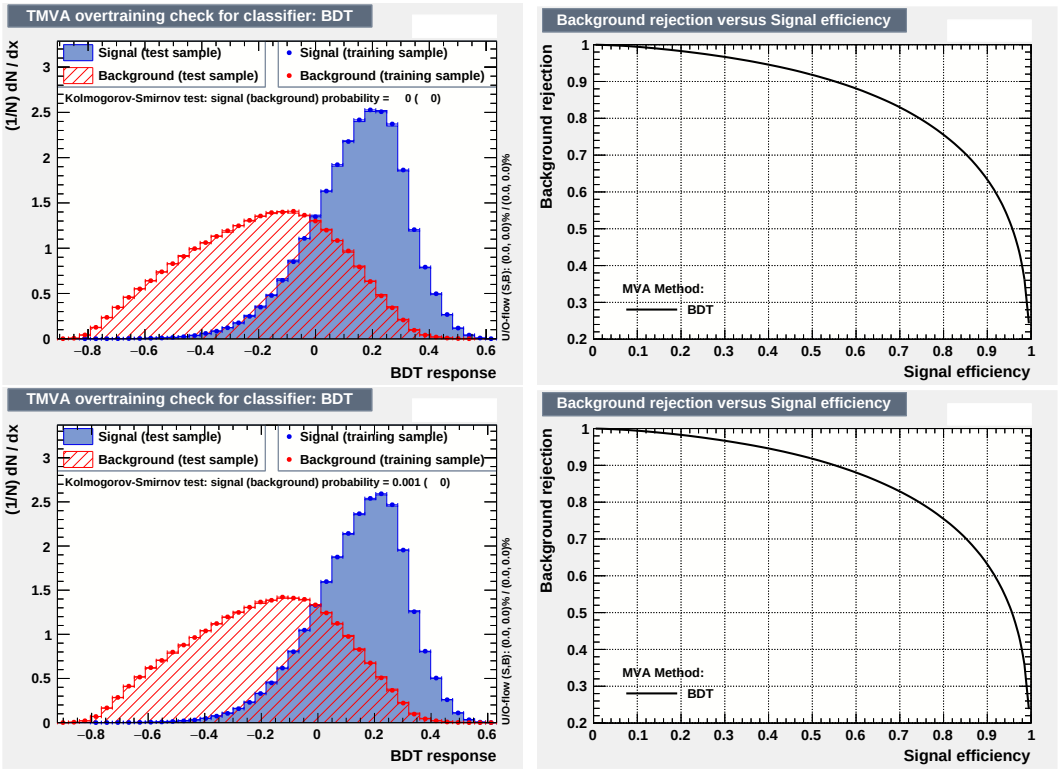

Figure B.2: BDT overtraining check and ROC-curves for 2017: The overtraining checks and the ROC-curves are shown for the two BDTs that are used for the jet assignment in the reconstruction of events in the resolved regime on detector level for simulation samples and data of the 2017 data taking period. The upper row corresponds to the BDT that is trained with even-numbered events, while the lower row corresponds to the BDT that is trained with odd-numbered events.

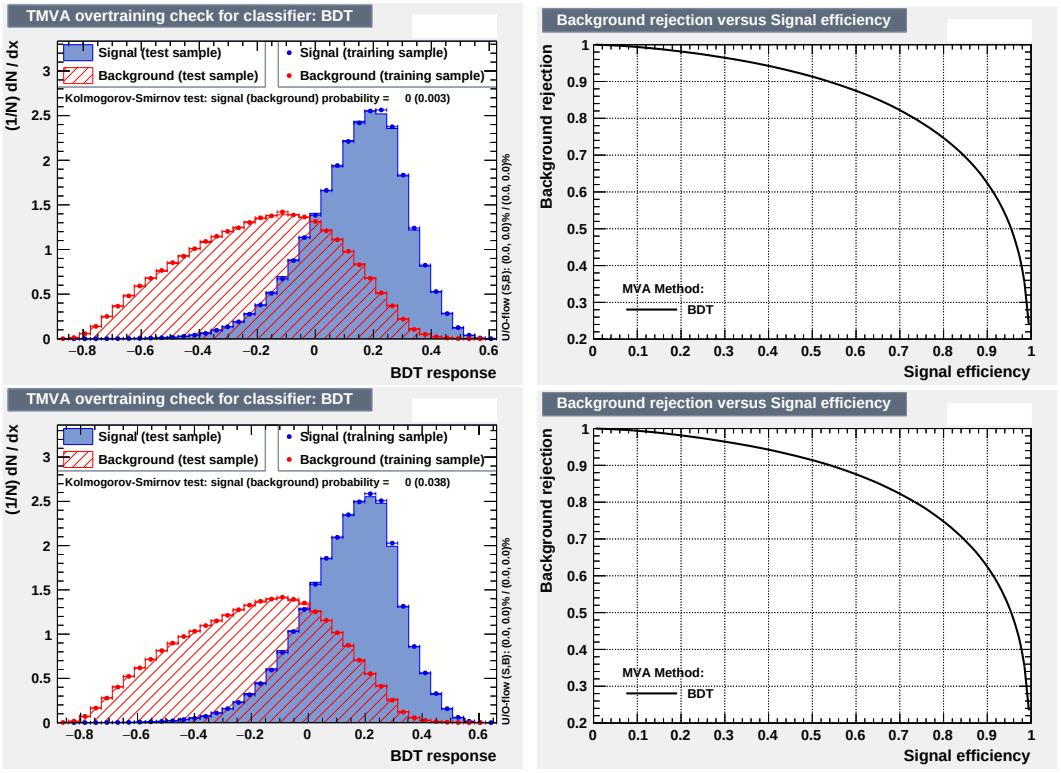

Figure B.3: BDT overtraining check and ROC-curves for 2018: The overtraining checks and the ROC-curves are shown for the two BDTs that are used for the jet assignment in the reconstruction of events in the resolved regime on detector level for simulation samples and data of the 2018 data taking period. The upper row corresponds to the BDT that is trained with even-numbered events, while the lower row corresponds to the BDT that is trained with odd-numbered events.

C. Reconstruction Distributions

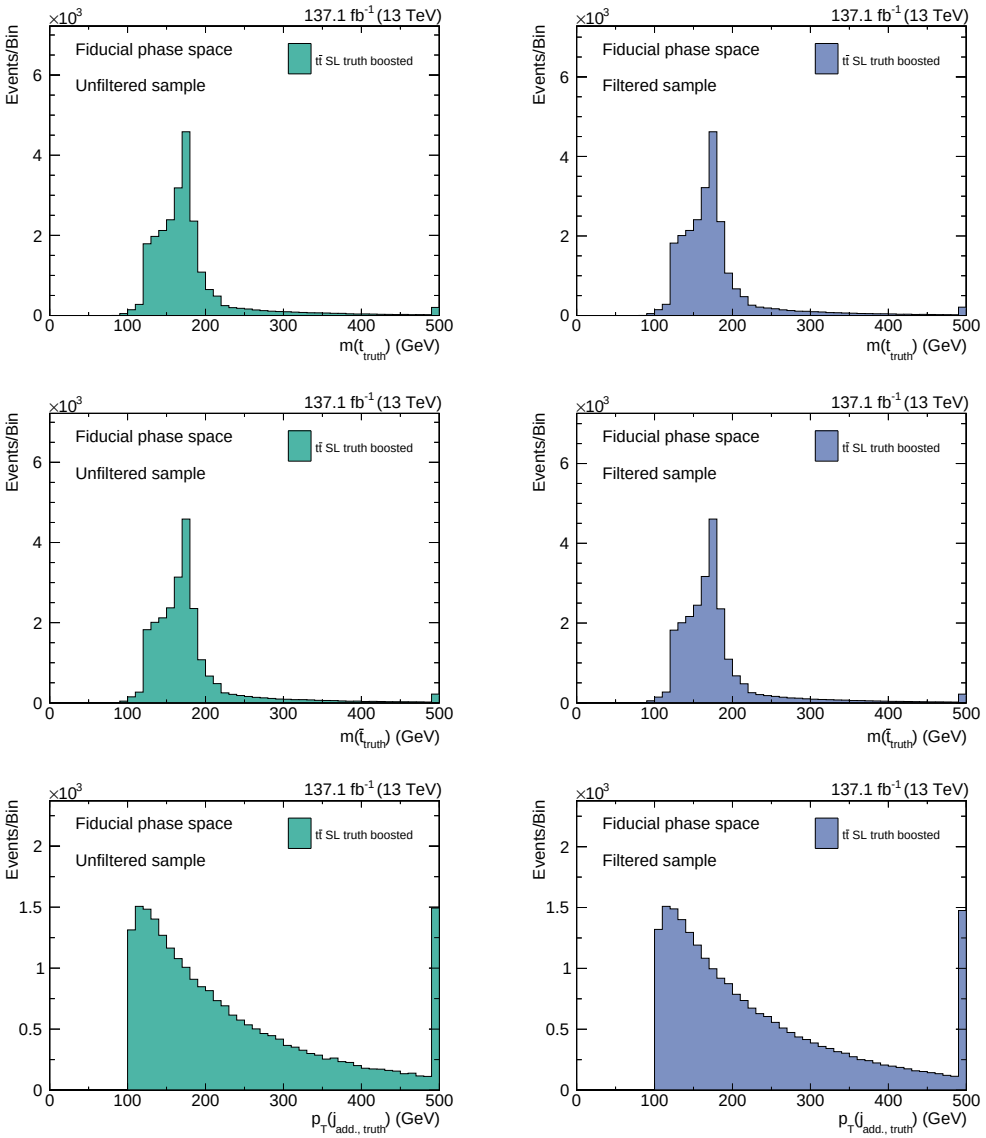

Figure C.1: Event reconstruction on truth level in the boosted regime: The mass of the reconstructed top quark and top antiquark and the transverse momentum of the reconstructed additional jet are shown on truth level for the fiducial phase space. The distributions in the left column are obtained by utilizing exclusively the unfiltered signal sample, while the distributions in the right column use the filtered signal sample only. The distributions show only events that are reconstructed in the boosted regime on truth level.

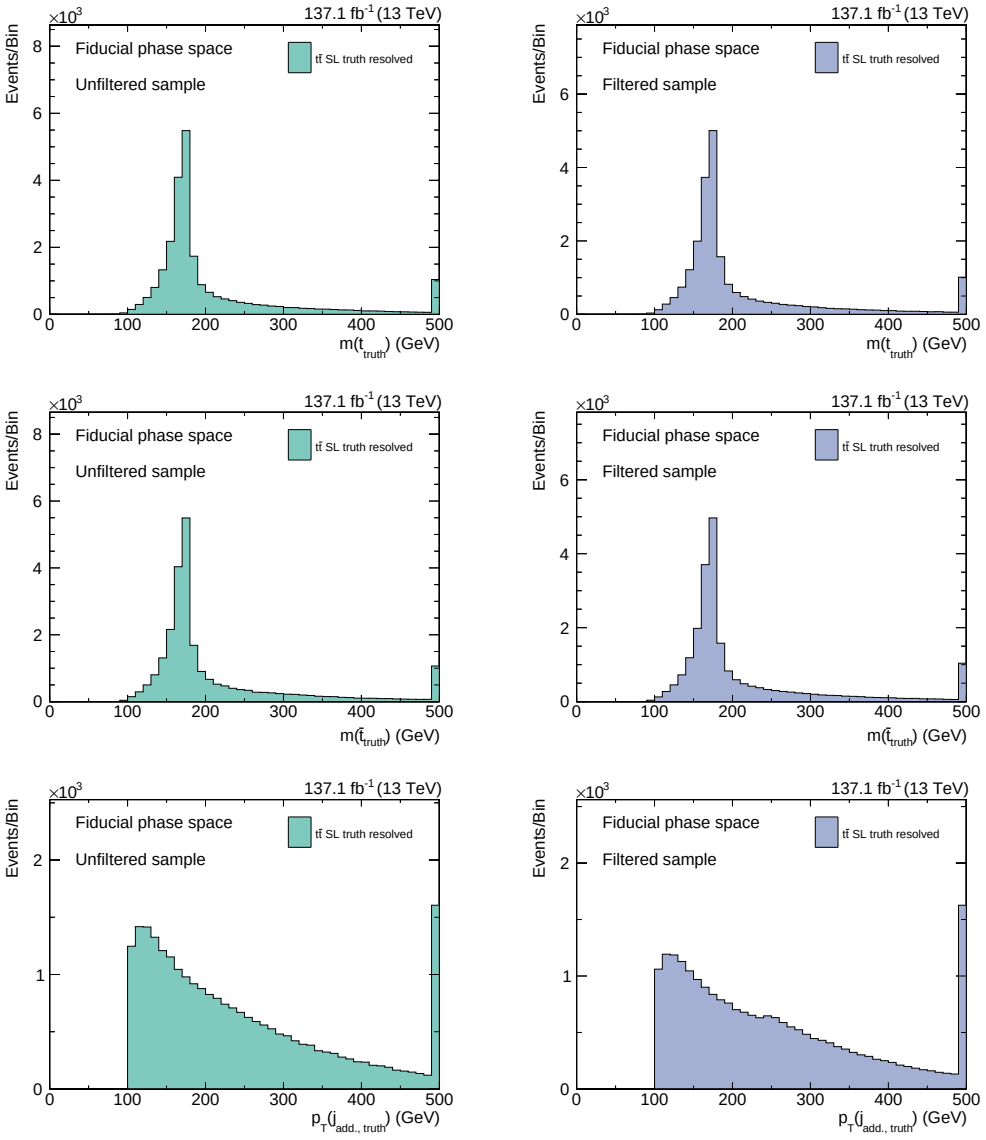

Figure C.2: Event reconstruction on truth level in the resolved regime: The mass of the reconstructed top quark and top antiquark and the transverse momentum of the reconstructed additional jet are shown on truth level for the fiducial phase space. The distributions in the left column are obtained by utilizing exclusively the unfiltered signal sample, while the distributions in the right column use the filtered signal sample only. The distributions show only events that are reconstructed in the resolved regime on truth level. The total amount of events is lower for the distributions of the filtered sample, as this sample does not provide a full coverage of the fiducial phase space (see Section 5.4.5).

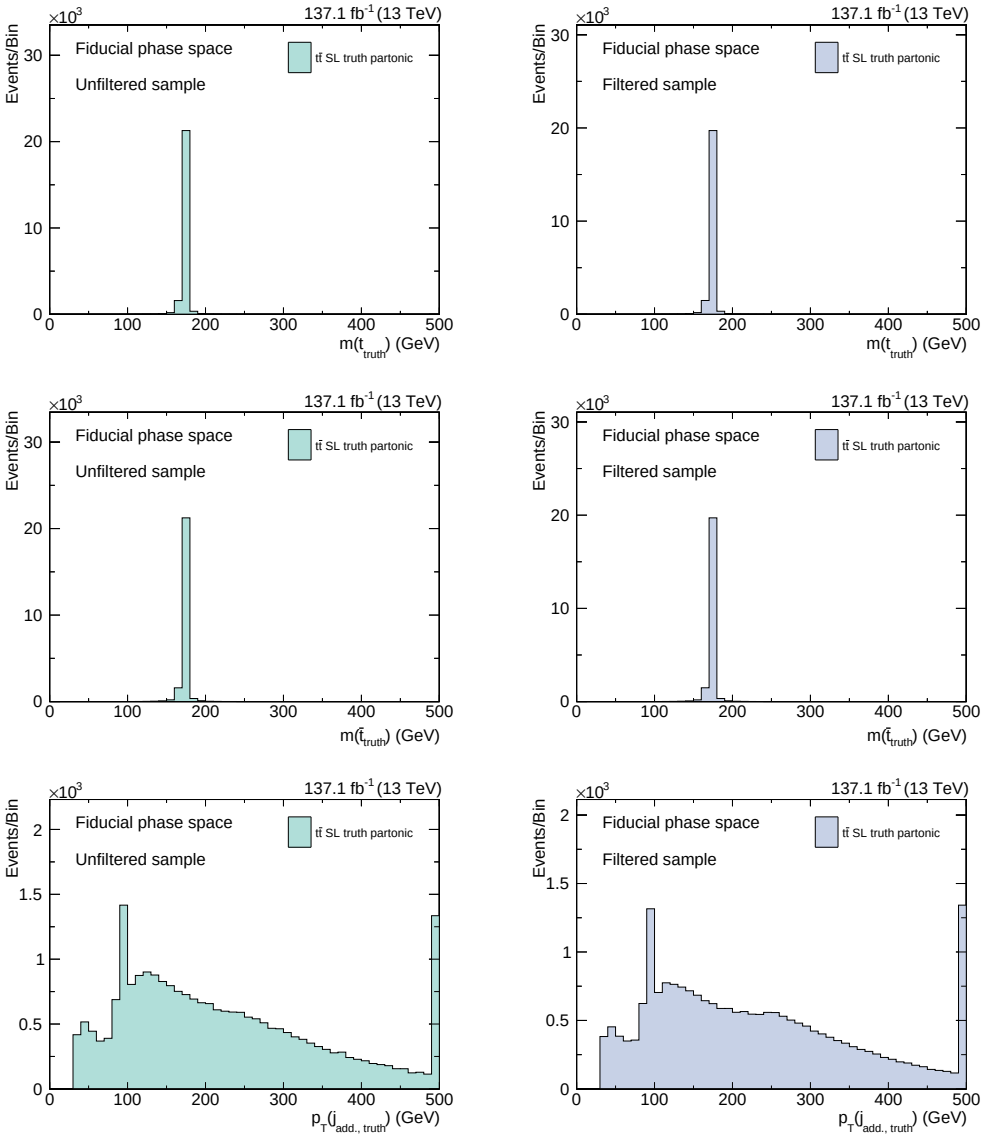

Figure C.3: Event reconstruction on truth level using partonic information: The mass of the reconstructed top quark and top antiquark and the transverse momentum of the reconstructed additional jet are shown on truth level for the fiducial phase space. The distributions in the left column are obtained by utilizing exclusively the unfiltered signal sample, while the distributions in the right column use the filtered signal sample only. The distributions show only events that are reconstructed using partonic information on truth level. The total amount of events is lower for the distributions of the filtered sample, as this sample does not provide a full coverage of the fiducial phase space (see Section 5.4.5).

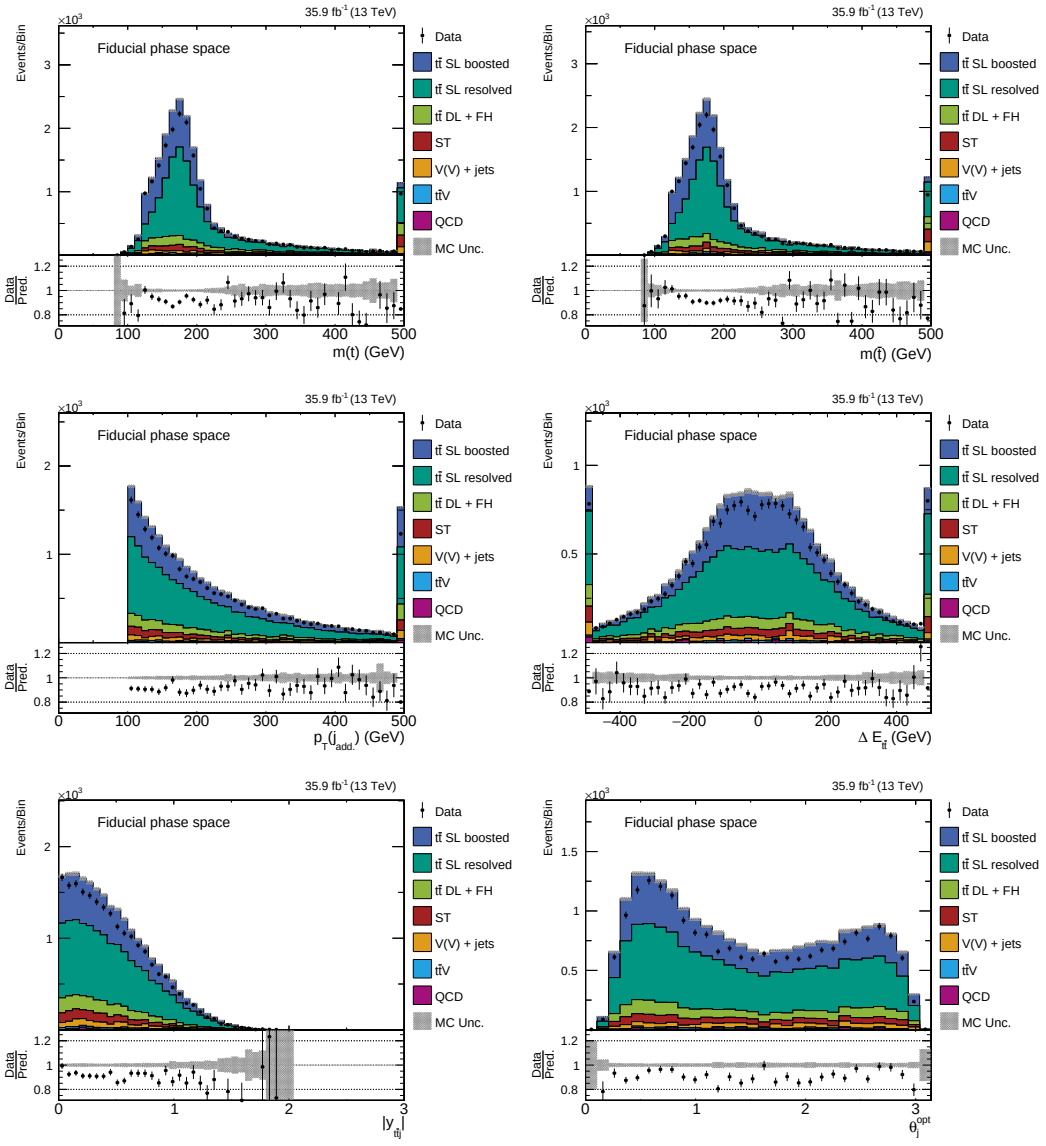

Figure C.4: Variables of interest in the fiducial phase space for the data taking period of 2016: The distributions of the kinematic properties of the reconstruction are shown for simulation and measured data. In addition the distributions for the three variables of interest for the energy asymmetry are shown.

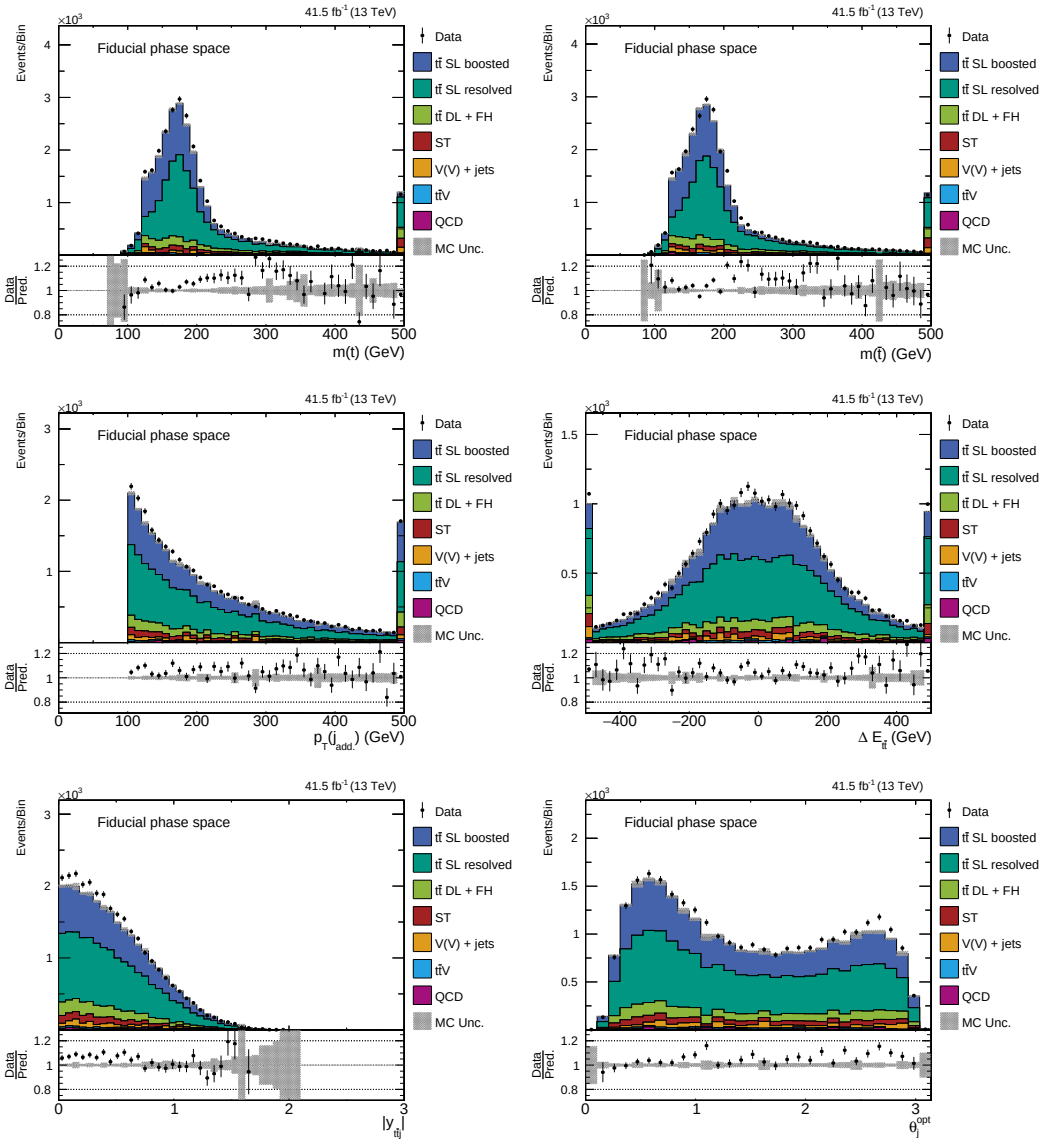

Figure C.5: Variables of interest in the fiducial phase space for the data taking period of 2017: The distributions of the kinematic properties of the reconstruction are shown for simulation and measured data. In addition the distributions for the three variables of interest for the energy asymmetry are shown.

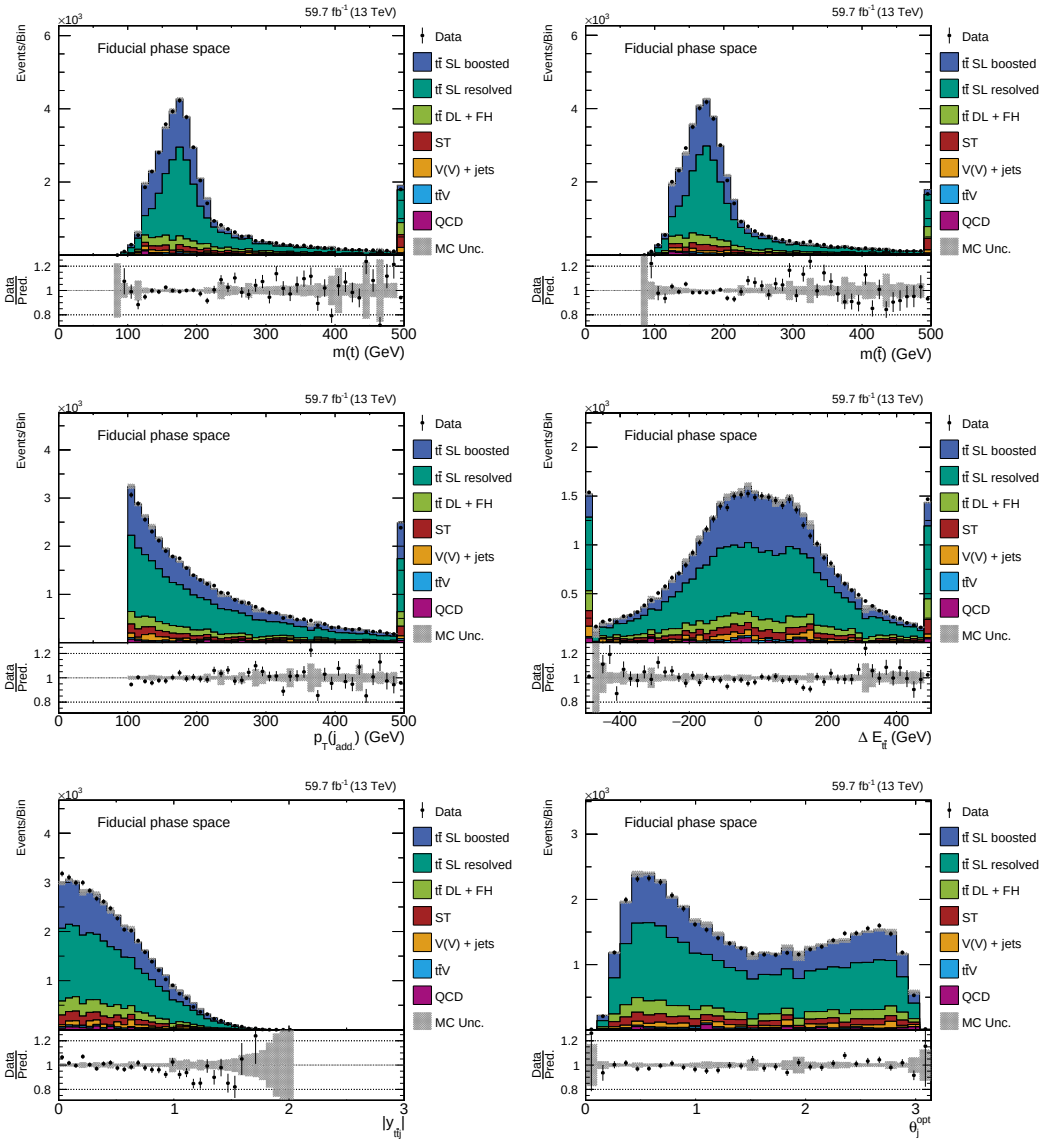

Figure C.6: Variables of interest in the fiducial phase space for the data taking period of 2018: The distributions of the kinematic properties of the reconstruction are shown for simulation and measured data. In addition the distributions for the three variables of interest for the energy asymmetry are shown.

D. Energy Asymmetry Values

Table D.1: Asymmetry values on truth level for unfiltered sample: The energy asymmetry values on truth level for the signal sample TT Semilep with detector-level reconstruction in the boosted and resolved regime. The quoted uncertainties refer to the propagated Poisson uncertainty on $N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$ and $N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$ (stat) and the propagated error due to the limited size of the simulation samples (syst).

θ_j^{opt}	$\lvert y_{t\bar{t}j}\rvert$	$\lvert\Delta E_{t\bar{t}}\rvert$	$N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$	$N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$	$A_E^{\textbf{opt}}$
	> 0	$> 0\,\text{GeV}$	12631.2	12667.3	$-0.143\,\% \pm 0.629\,\%\,(\text{stat}) \pm 0.264\,\%\,(\text{syst})$
$0 - 0.3\pi$	> 0	$> 50\,\text{GeV}$	10100.2	10146.2	$-0.227\,\% \pm 0.703\,\%\,(\text{stat}) \pm 0.295\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	4346.5	4345.9	$0.007\,\% \pm 1.073\,\%\,(\text{stat}) \pm 0.451\,\%\,(\text{syst})$
	> 0	$> 0\,\text{GeV}$	13829.7	14123.0	$-1.049\,\% \pm 0.598\,\%\,(\text{stat}) \pm 0.251\,\%\,(\text{syst})$
$0.3\pi - 0.7\pi$	> 0	$> 50\,\text{GeV}$	10008.1	10280.1	$-1.341\,\% \pm 0.702\,\%\,(\text{stat}) \pm 0.294\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	4860.9	5024.4	$-1.653\,\% \pm 1.006\,\%\,(\text{stat}) \pm 0.422\,\%\,(\text{syst})$
	> 0	$> 0\,\text{GeV}$	9313.8	9283.8	$0.161\,\% \pm 0.733\,\%\,(\text{stat}) \pm 0.309\,\%\,(\text{syst})$
$0.7\pi - \pi$	> 0	$> 50\,\text{GeV}$	7588.9	7527.4	$0.407\,\% \pm 0.813\,\%\,(\text{stat}) \pm 0.343\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	2651.6	2654.8	$-0.062\,\% \pm 1.373\,\%\,(\text{stat}) \pm 0.581\,\%\,(\text{syst})$

Table D.2: Asymmetry values on detector level for unfiltered sample: The energy asymmetry values on detector level for the signal sample TT Semilep with detector-level reconstruction in the boosted and resolved regime. The quoted uncertainties refer to the propagated Poisson uncertainty on $N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$ and $N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$ (stat) and the propagated error due to the limited size of the simulation samples (syst).

θ_j^{opt}	$\lvert y_{t\bar{t}j}\rvert$	$\lvert\Delta E_{t\bar{t}}\rvert$	$N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$	$N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$	A_E^{opt}
	> 0	$> 0\,\text{GeV}$	12803.5	12864.5	$-0.238\,\% \pm 0.624\,\%\,(\text{stat}) \pm 0.262\,\%\,(\text{syst})$
$0 - 0.3\pi$	> 0	$> 50\,\text{GeV}$	10480.9	10501.9	$-0.1\,\% \pm 0.69\,\%\,(\text{stat}) \pm 0.29\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	4429.8	4474.3	$-0.5\,\% \pm 1.06\,\%\,(\text{stat}) \pm 0.445\,\%\,(\text{syst})$
	> 0	$> 0\,\text{GeV}$	13757.5	14104.9	$-1.247\,\% \pm 0.599\,\%\,(\text{stat}) \pm 0.252\,\%\,(\text{syst})$
$0.3\pi - 0.7\pi$	> 0	$> 50\,\text{GeV}$	10600.7	10898.2	$-1.384\,\% \pm 0.682\,\%\,(\text{stat}) \pm 0.287\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	5042.9	5220.7	$-1.732\,\% \pm 0.987\,\%\,(\text{stat}) \pm 0.417\,\%\,(\text{syst})$
	> 0	$> 0\,\text{GeV}$	9189.2	9129.1	$0.328\,\% \pm 0.739\,\%\,(\text{stat}) \pm 0.311\,\%\,(\text{syst})$
$0.7\pi - \pi$	> 0	$> 50\,\text{GeV}$	7721.6	7650.7	$0.461\,\% \pm 0.807\,\%\,(\text{stat}) \pm 0.339\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	2539.7	2560.9	$-0.414\,\% \pm 1.4\,\%\,(\text{stat}) \pm 0.591\,\%\,(\text{syst})$

Table D.3: Asymmetry values on truth level for boosted reconstruction: The energy asymmetry values on truth level for the signal sample TT Semilep with detector-level reconstruction in the boosted regime only. The quoted uncertainties refer to the propagated Poisson uncertainty on $N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$ and $N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$ (stat) and the propagated error due to the limited size of the simulation samples (syst).

θ_j^{opt}	$\lvert y_{t\bar{t}j}\rvert$	$\lvert\Delta E_{t\bar{t}}\rvert$	$N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$	$N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$	A_E^{opt}
	> 0	$> 0\,\text{GeV}$	4887.6	4949.0	$-0.625\,\% \pm 1.008\,\%\,(\text{stat}) \pm 0.272\,\%\,(\text{syst})$
$0 - 0.3\pi$	> 0	$> 50\,\text{GeV}$	3925.8	3951.3	$-0.324\,\% \pm 1.127\,\%\,(\text{stat}) \pm 0.304\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	1657.7	1668.6	$-0.327\,\% \pm 1.734\,\%\,(\text{stat}) \pm 0.467\,\%\,(\text{syst})$
	> 0	$> 0\,\text{GeV}$	5161.9	5322.0	$-1.527\,\% \pm 0.977\,\%\,(\text{stat}) \pm 0.263\,\%\,(\text{syst})$
$0.3\pi - 0.7\pi$	> 0	$> 50\,\text{GeV}$	3815.9	3962.3	$-1.882\,\% \pm 1.134\,\%\,(\text{stat}) \pm 0.305\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	1822.1	1918.9	$-2.587\,\% \pm 1.634\,\%\,(\text{stat}) \pm 0.44\,\%\,(\text{syst})$
	> 0	$> 0\,\text{GeV}$	3669.0	3698.9	$-0.406\,\% \pm 1.165\,\%\,(\text{stat}) \pm 0.316\,\%\,(\text{syst})$
$0.7\pi - \pi$	> 0	$> 50\,\text{GeV}$	2998.3	3009.0	$-0.178\,\% \pm 1.29\,\%\,(\text{stat}) \pm 0.35\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	1066.4	1069.7	$-0.155\,\% \pm 2.164\,\%\,(\text{stat}) \pm 0.586\,\%\,(\text{syst})$

Table D.4: Asymmetry values on detector level for boosted reconstruction: The energy asymmetry values on detector level for the signal sample TT Semilep with detector-level reconstruction in the boosted regime only. The quoted uncertainties refer to the propagated Poisson uncertainty on $N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$ and $N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$ (stat) and the propagated error due to the limited size of the simulation samples (syst).

θ_j^{opt}	$\lvert y_{t\bar{t}j}\rvert$	$\lvert\Delta E_{t\bar{t}}\rvert$	$N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$	$N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$	A_E^{opt}
	> 0	$> 0\,\text{GeV}$	5091.8	5130.0	$-0.374\,\% \pm 0.989\,\%\,(\text{stat}) \pm 0.267\,\%\,(\text{syst})$
$0 - 0.3\pi$	> 0	$> 50\,\text{GeV}$	4050.2	4070.5	$-0.251\,\% \pm 1.11\,\%\,(\text{stat}) \pm 0.299\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	1710.2	1728.9	$-0.542\,\% \pm 1.705\,\%\,(\text{stat}) \pm 0.459\,\%\,(\text{syst})$
	> 0	$> 0\,\text{GeV}$	5064.6	5209.7	$-1.412\,\% \pm 0.986\,\%\,(\text{stat}) \pm 0.266\,\%\,(\text{syst})$
$0.3\pi - 0.7\pi$	> 0	$> 50\,\text{GeV}$	3602.6	3736.0	$-1.818\,\% \pm 1.167\,\%\,(\text{stat}) \pm 0.314\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	1664.1	1761.6	$-2.845\,\% \pm 1.708\,\%\,(\text{stat}) \pm 0.46\,\%\,(\text{syst})$
	> 0	$> 0\,\text{GeV}$	3572.9	3619.3	$-0.644\,\% \pm 1.179\,\%\,(\text{stat}) \pm 0.319\,\%\,(\text{syst})$
$0.7\pi - \pi$	> 0	$> 50\,\text{GeV}$	2889.5	2915.1	$-0.441\,\% \pm 1.313\,\%\,(\text{stat}) \pm 0.355\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	965.4	966.8	$-0.07\,\% \pm 2.275\,\%\,(\text{stat}) \pm 0.615\,\%\,(\text{syst})$

Table D.5: Asymmetry values on truth level for resolved reconstruction: The energy asymmetry values on truth level for the signal sample TT Semilep with detector-level reconstruction in the resolved regime only. The quoted uncertainties refer to the propagated Poisson uncertainty on $N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$ and $N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$ (stat) and the propagated error due to the limited size of the simulation samples (syst).

θ_j^{opt}	$\lvert y_{t\bar{t}j}\rvert$	$\lvert\Delta E_{t\bar{t}}\rvert$	$N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$	$N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$	A_E^{opt}
	> 0	$> 0\,\text{GeV}$	7754.5	7761.9	$-0.048\,\% \pm 0.803\,\%\,(\text{stat}) \pm 0.335\,\%\,(\text{syst})$
$0 - 0.3\pi$	> 0	$> 50\,\text{GeV}$	6189.1	6206.8	$-0.142\,\% \pm 0.898\,\%\,(\text{stat}) \pm 0.375\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	2704.5	2673.9	$0.568\,\% \pm 1.364\,\%\,(\text{stat}) \pm 0.569\,\%\,(\text{syst})$
	> 0	$> 0\,\text{GeV}$	8702.8	8808.0	$-0.601\,\% \pm 0.756\,\%\,(\text{stat}) \pm 0.316\,\%\,(\text{syst})$
$0.3\pi - 0.7\pi$	> 0	$> 50\,\text{GeV}$	6221.2	6326.4	$-0.838\,\% \pm 0.893\,\%\,(\text{stat}) \pm 0.372\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	3059.9	3120.7	$-0.984\,\% \pm 1.272\,\%\,(\text{stat}) \pm 0.532\,\%\,(\text{syst})$
	> 0	$> 0\,\text{GeV}$	5637.7	5604.0	$0.299\,\% \pm 0.943\,\%\,(\text{stat}) \pm 0.395\,\%\,(\text{syst})$
$0.7\pi - \pi$	> 0	$> 50\,\text{GeV}$	4572.3	4528.9	$0.476\,\% \pm 1.048\,\%\,(\text{stat}) \pm 0.439\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	1601.6	1598.7	$0.091\,\% \pm 1.768\,\%\,(\text{stat}) \pm 0.745\,\%\,(\text{syst})$

Table D.6: Asymmetry values on detector level for resolved reconstruction: The energy asymmetry values on detector level for the signal sample TT Semilep with detector-level reconstruction in the resolved regime only. The quoted uncertainties refer to the propagated Poisson uncertainty on $N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$ and $N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$ (stat) and the propagated error due to the limited size of the simulation samples (syst).

θ_j^{opt}	$\lvert y_{t\bar{t}j}\rvert$	$\lvert\Delta E_{t\bar{t}}\rvert$	$N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$	$N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$	A_E^{opt}
	> 0	$> 0\,\text{GeV}$	7746.0	7760.1	$-0.091\,\% \pm 0.803\,\%\,(\text{stat}) \pm 0.335\,\%\,(\text{syst})$
$0 - 0.3\pi$	> 0	$> 50\,\text{GeV}$	6453.6	6448.1	$0.043\,\% \pm 0.88\,\%\,(\text{stat}) \pm 0.367\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	2736.7	2752.0	$-0.279\,\% \pm 1.35\,\%\,(\text{stat}) \pm 0.563\,\%\,(\text{syst})$
	> 0	$> 0\,\text{GeV}$	8715.5	8914.9	$-1.131\,\% \pm 0.753\,\%\,(\text{stat}) \pm 0.316\,\%\,(\text{syst})$
$0.3\pi - 0.7\pi$	> 0	$> 50\,\text{GeV}$	7018.2	7173.8	$-1.096\,\% \pm 0.839\,\%\,(\text{stat}) \pm 0.352\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	3386.1	3472.6	$-1.261\,\% \pm 1.207\,\%\,(\text{stat}) \pm 0.509\,\%\,(\text{syst})$
	> 0	$> 0\,\text{GeV}$	5597.8	5534.6	$0.568\,\% \pm 0.948\,\%\,(\text{stat}) \pm 0.396\,\%\,(\text{syst})$
$0.7\pi - \pi$	> 0	$> 50\,\text{GeV}$	4819.1	4755.0	$0.669\,\% \pm 1.022\,\%\,(\text{stat}) \pm 0.427\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	1597.0	1608.3	$-0.354\,\% \pm 1.766\,\%\,(\text{stat}) \pm 0.742\,\%\,(\text{syst})$

Table D.7: Asymmetry values on truth level for boosted reconstruction on truth level: The energy asymmetry values on truth level for the signal sample TT Semilep with truth-level reconstruction in the boosted regime only. The quoted uncertainties refer to the propagated Poisson uncertainty on $N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$ and $N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$ (stat) and the propagated error due to the limited size of the simulation samples (syst).

θ_j^{opt}	$\lvert y_{t\bar{t}j}\rvert$	$\lvert\Delta E_{t\bar{t}}\rvert$	$N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$	$N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$	A_E^{opt}
	> 0	$> 0\,\text{GeV}$	4113.0	4179.5	$-0.802\,\% \pm 1.098\,\%\,(\text{stat}) \pm 0.465\,\%\,(\text{syst})$
$0 - 0.3\pi$	> 0	$> 50\,\text{GeV}$	3295.0	3366.5	$-1.074\,\% \pm 1.225\,\%\,(\text{stat}) \pm 0.519\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	1347.8	1381.1	$-1.22\,\% \pm 1.914\,\%\,(\text{stat}) \pm 0.81\,\%\,(\text{syst})$
	> 0	$> 0\,\text{GeV}$	4424.3	4529.3	$-1.173\,\% \pm 1.057\,\%\,(\text{stat}) \pm 0.446\,\%\,(\text{syst})$
$0.3\pi - 0.7\pi$	> 0	$> 50\,\text{GeV}$	3231.8	3339.7	$-1.643\,\% \pm 1.233\,\%\,(\text{stat}) \pm 0.521\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	1495.2	1553.7	$-1.919\,\% \pm 1.811\,\%\,(\text{stat}) \pm 0.763\,\%\,(\text{syst})$
	> 0	$> 0\,\text{GeV}$	3072.6	3100.7	$-0.457\,\% \pm 1.273\,\%\,(\text{stat}) \pm 0.539\,\%\,(\text{syst})$
$0.7\pi - \pi$	> 0	$> 50\,\text{GeV}$	2541.4	2547.0	$-0.11\,\% \pm 1.402\,\%\,(\text{stat}) \pm 0.594\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	844.9	868.5	$-1.376\,\% \pm 2.416\,\%\,(\text{stat}) \pm 1.025\,\%\,(\text{syst})$

Table D.8: Asymmetry values on truth level for resolved reconstruction on truth level: The energy asymmetry values on truth level for the signal sample TT Semilep with truth-level reconstruction in the resolved regime only. The quoted uncertainties refer to the propagated Poisson uncertainty on $N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$ and $N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$ (stat) and the propagated error due to the limited size of the simulation samples (syst).

θ_j^{opt}	$\lvert y_{t\bar{t}j}\rvert$	$\lvert\Delta E_{t\bar{t}}\rvert$	$N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$	$N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$	A_E^{opt}
	> 0	$> 0\,\text{GeV}$	4448.5	4423.6	$0.281\,\% \pm 1.062\,\%\,(\text{stat}) \pm 0.441\,\%\,(\text{syst})$
$0 - 0.3\pi$	> 0	$> 50\,\text{GeV}$	3670.0	3648.6	$0.292\,\% \pm 1.169\,\%\,(\text{stat}) \pm 0.485\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	1582.1	1543.3	$1.242\,\% \pm 1.789\,\%\,(\text{stat}) \pm 0.741\,\%\,(\text{syst})$
	> 0	$> 0\,\text{GeV}$	4757.8	4824.6	$-0.698\,\% \pm 1.022\,\%\,(\text{stat}) \pm 0.425\,\%\,(\text{syst})$
$0.3\pi - 0.7\pi$	> 0	$> 50\,\text{GeV}$	3615.3	3703.3	$-1.203\,\% \pm 1.169\,\%\,(\text{stat}) \pm 0.485\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	1728.2	1789.3	$-1.736\,\% \pm 1.686\,\%\,(\text{stat}) \pm 0.703\,\%\,(\text{syst})$
	> 0	$> 0\,\text{GeV}$	3112.6	3099.1	$0.218\,\% \pm 1.269\,\%\,(\text{stat}) \pm 0.53\,\%\,(\text{syst})$
$0.7\pi - \pi$	> 0	$> 50\,\text{GeV}$	2646.1	2616.7	$0.559\,\% \pm 1.378\,\%\,(\text{stat}) \pm 0.575\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	885.5	872.7	$0.725\,\% \pm 2.385\,\%\,(\text{stat}) \pm 1.0\,\%\,(\text{syst})$

Table D.9: Asymmetry values on truth level for partonic reconstruction on truth level: The energy asymmetry values on truth level for the signal sample TT Semilep with truth-level reconstruction using partonic information only. The quoted uncertainties refer to the propagated Poisson uncertainty on $N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$ and $N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$ (stat) and the propagated error due to the limited size of the simulation samples (syst).

θ_j^{opt}	$\lvert y_{t\bar{t}j}\rvert$	$\lvert\Delta E_{t\bar{t}}\rvert$	$N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$	$N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$	A_E^{opt}
	> 0	$> 0\,\text{GeV}$	4069.7	4064.2	$0.067\,\% \pm 1.109\,\%\,(\text{stat}) \pm 0.469\,\%\,(\text{syst})$
$0 - 0.3\pi$	> 0	$> 50\,\text{GeV}$	3135.3	3131.1	$0.067\,\% \pm 1.263\,\%\,(\text{stat}) \pm 0.533\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	1416.6	1421.5	$-0.174\,\% \pm 1.877\,\%\,(\text{stat}) \pm 0.795\,\%\,(\text{syst})$
	> 0	$> 0\,\text{GeV}$	4647.6	4769.0	$-1.29\,\% \pm 1.03\,\%\,(\text{stat}) \pm 0.435\,\%\,(\text{syst})$
$0.3\pi - 0.7\pi$	> 0	$> 50\,\text{GeV}$	3161.0	3237.1	$-1.188\,\% \pm 1.25\,\%\,(\text{stat}) \pm 0.524\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	1637.5	1681.4	$-1.322\,\% \pm 1.736\,\%\,(\text{stat}) \pm 0.731\,\%\,(\text{syst})$
	> 0	$> 0\,\text{GeV}$	3128.6	3084.0	$0.717\,\% \pm 1.269\,\%\,(\text{stat}) \pm 0.537\,\%\,(\text{syst})$
$0.7\pi - \pi$	> 0	$> 50\,\text{GeV}$	2401.3	2363.6	$0.791\,\% \pm 1.449\,\%\,(\text{stat}) \pm 0.613\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	921.2	913.6	$0.412\,\% \pm 2.335\,\%\,(\text{stat}) \pm 0.994\,\%\,(\text{syst})$

Table D.10: Asymmetry values on detector level for boosted reconstruction only including backgrounds: The energy asymmetry values on detector level for the combination of the signal samples TT Semilep HT 250 Boosted and TT Semilep Resolved, and all simulation samples for the background processes with reconstruction in the boosted regime only. The quoted uncertainties refer to the propagated Poisson uncertainty on $N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$ and $N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$ (stat) and the propagated error due to the limited size of the simulation samples (syst).

θ_j^{opt}	$\lvert y_{t\bar{t}j}\rvert$	$\lvert\Delta E_{t\bar{t}}\rvert$	$N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$	$N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$	A_E^{opt}
	> 0	$> 0\,\text{GeV}$	6044.7	6090.6	$-0.378\,\% \pm 0.908\,\%\,(\text{stat}) \pm 0.611\,\%\,(\text{syst})$
$0 - 0.3\pi$	> 0	$> 50\,\text{GeV}$	4832.5	4855.5	$-0.238\,\% \pm 1.016\,\%\,(\text{stat}) \pm 0.708\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	2012.7	2058.3	$-1.119\,\% \pm 1.567\,\%\,(\text{stat}) \pm 0.955\,\%\,(\text{syst})$
	> 0	$> 0\,\text{GeV}$	6157.0	6488.4	$-2.621\,\% \pm 0.889\,\%\,(\text{stat}) \pm 0.764\,\%\,(\text{syst})$
$0.3\pi - 0.7\pi$	> 0	$> 50\,\text{GeV}$	4459.1	4622.7	$-1.801\,\% \pm 1.049\,\%\,(\text{stat}) \pm 0.796\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	2114.5	2170.1	$-1.298\,\% \pm 1.528\,\%\,(\text{stat}) \pm 1.328\,\%\,(\text{syst})$
	> 0	$> 0\,\text{GeV}$	4399.3	4473.3	$-0.834\,\% \pm 1.062\,\%\,(\text{stat}) \pm 0.781\,\%\,(\text{syst})$
$0.7\pi - \pi$	> 0	$> 50\,\text{GeV}$	3558.2	3600.1	$-0.586\,\% \pm 1.182\,\%\,(\text{stat}) \pm 0.857\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	1218.4	1182.9	$1.481\,\% \pm 2.04\,\%\,(\text{stat}) \pm 1.65\,\%\,(\text{syst})$

Table D.11: Asymmetry values for measured data for boosted reconstruction only: The energy asymmetry values for measured data in the fiducial phase space for boosted reconstruction only. The quoted uncertainty is obtained by assuming a Poisson distribution for $N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$ and $N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$ and by error propagation of the corresponding uncertainty.

θ_j^{opt}	$\lvert y_{t\bar{t}j}\rvert$	$\lvert\Delta E_{t\bar{t}}\rvert$	$N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$	$N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$	A_E^{opt}
	> 0	$> 0\,\text{GeV}$	6082.0	6267.0	$-1.498\,\% \pm 0.9\,\%\,(\text{stat})$
$0 - 0.3\pi$	> 0	$> 50\,\text{GeV}$	4839.0	5027.0	$-1.906\,\% \pm 1.007\,\%\,(\text{stat})$
	> 0.5	$> 50\,\text{GeV}$	1934.0	2072.0	$-3.445\,\% \pm 1.579\,\%\,(\text{stat})$
	> 0	$> 0\,\text{GeV}$	6313.0	6570.0	$-1.995\,\% \pm 0.881\,\%\,(\text{stat})$
$0.3\pi - 0.7\pi$	> 0	$> 50\,\text{GeV}$	4441.0	4692.0	$-2.748\,\% \pm 1.046\,\%\,(\text{stat})$
	> 0.5	$> 50\,\text{GeV}$	2060.0	2206.0	$-3.422\,\% \pm 1.53\,\%\,(\text{stat})$
	> 0	$> 0\,\text{GeV}$	4774.0	4695.0	$0.834\,\% \pm 1.028\,\%\,(\text{stat})$
$0.7\pi - \pi$	> 0	$> 50\,\text{GeV}$	3882.0	3809.0	$0.949\,\% \pm 1.14\,\%\,(\text{stat})$
	> 0.5	$> 50\,\text{GeV}$	1255.0	1266.0	$-0.436\,\% \pm 1.992\,\%\,(\text{stat})$

Table D.12: Asymmetry values on detector level for resolved reconstruction only including backgrounds: The energy asymmetry values on detector level for the combination of the signal samples TT Semilep HT 250 Boosted and TT Semilep Resolved, and all simulation samples for the background processes with reconstruction in the resolved regime only. The quoted uncertainties refer to the propagated Poisson uncertainty on $N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$ and $N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$ (stat) and the propagated error due to the limited size of the simulation samples (syst).

θ_j^{opt}	$\lvert y_{t\bar{t}j}\rvert$	$\lvert\Delta E_{t\bar{t}}\rvert$	$N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$	$N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$	$A_E^{\textbf{opt}}$
	> 0	$> 0\,\text{GeV}$	9697.0	9958.6	$-1.331\,\% \pm 0.713\,\%\,(\text{stat}) \pm 0.409\,\%\,(\text{syst})$
$0 - 0.3\pi$	> 0	$> 50\,\text{GeV}$	8153.8	8396.2	$-1.465\,\% \pm 0.777\,\%\,(\text{stat}) \pm 0.458\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	3446.0	3535.5	$-1.282\,\% \pm 1.197\,\%\,(\text{stat}) \pm 0.692\,\%\,(\text{syst})$
	> 0	$> 0\,\text{GeV}$	11021.2	11540.4	$-2.301\,\% \pm 0.666\,\%\,(\text{stat}) \pm 0.589\,\%\,(\text{syst})$
$0.3\pi - 0.7\pi$	> 0	$> 50\,\text{GeV}$	9014.6	9430.7	$-2.256\,\% \pm 0.736\,\%\,(\text{stat}) \pm 0.674\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	4304.5	4537.0	$-2.63\,\% \pm 1.063\,\%\,(\text{stat}) \pm 0.984\,\%\,(\text{syst})$
	> 0	$> 0\,\text{GeV}$	7272.5	7226.3	$0.319\,\% \pm 0.83\,\%\,(\text{stat}) \pm 0.646\,\%\,(\text{syst})$
$0.7\pi - \pi$	> 0	$> 50\,\text{GeV}$	6316.4	6290.7	$0.204\,\% \pm 0.891\,\%\,(\text{stat}) \pm 0.716\,\%\,(\text{syst})$
	> 0.5	$> 50\,\text{GeV}$	2066.0	2117.1	$-1.222\,\% \pm 1.546\,\%\,(\text{stat}) \pm 1.204\,\%\,(\text{syst})$

Table D.13: Asymmetry values for measured data for resolved reconstruction only: The energy asymmetry values for measured data in the fiducial phase space for resolved reconstruction only. The quoted uncertainty is obtained by assuming a Poisson distribution for $N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$ and $N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$ and by error propagation of the corresponding uncertainty.

θ_j^{opt}	$\lvert y_{t\bar{t}j}\rvert$	$\lvert\Delta E_{t\bar{t}}\rvert$	$N(\Delta E_{t\bar{t}} > 0\,\text{GeV})$	$N(\Delta E_{t\bar{t}} < 0\,\text{GeV})$	$A_E^{\textbf{opt}}$
	> 0	$> 0\,\text{GeV}$	9277.0	9649.0	$-1.966\,\% \pm 0.727\,\%\,(\text{stat})$
$0 - 0.3\pi$	> 0	$> 50\,\text{GeV}$	7892.0	8110.0	$-1.362\,\% \pm 0.79\,\%\,(\text{stat})$
	> 0.5	$> 50\,\text{GeV}$	3198.0	3332.0	$-2.052\,\% \pm 1.237\,\%\,(\text{stat})$
	> 0	$> 0\,\text{GeV}$	10678.0	10911.0	$-1.079\,\% \pm 0.681\,\%\,(\text{stat})$
$0.3\pi - 0.7\pi$	> 0	$> 50\,\text{GeV}$	8752.0	8990.0	$-1.341\,\% \pm 0.751\,\%\,(\text{stat})$
	> 0.5	$> 50\,\text{GeV}$	4086.0	4295.0	$-2.494\,\% \pm 1.092\,\%\,(\text{stat})$
	> 0	$> 0\,\text{GeV}$	7102.0	7168.0	$-0.463\,\% \pm 0.837\,\%\,(\text{stat})$
$0.7\pi - \pi$	> 0	$> 50\,\text{GeV}$	6213.0	6242.0	$-0.233\,\% \pm 0.896\,\%\,(\text{stat})$
	> 0.5	$> 50\,\text{GeV}$	1946.0	1998.0	$-1.318\,\% \pm 1.592\,\%\,(\text{stat})$

E. Systematic Uncertainties

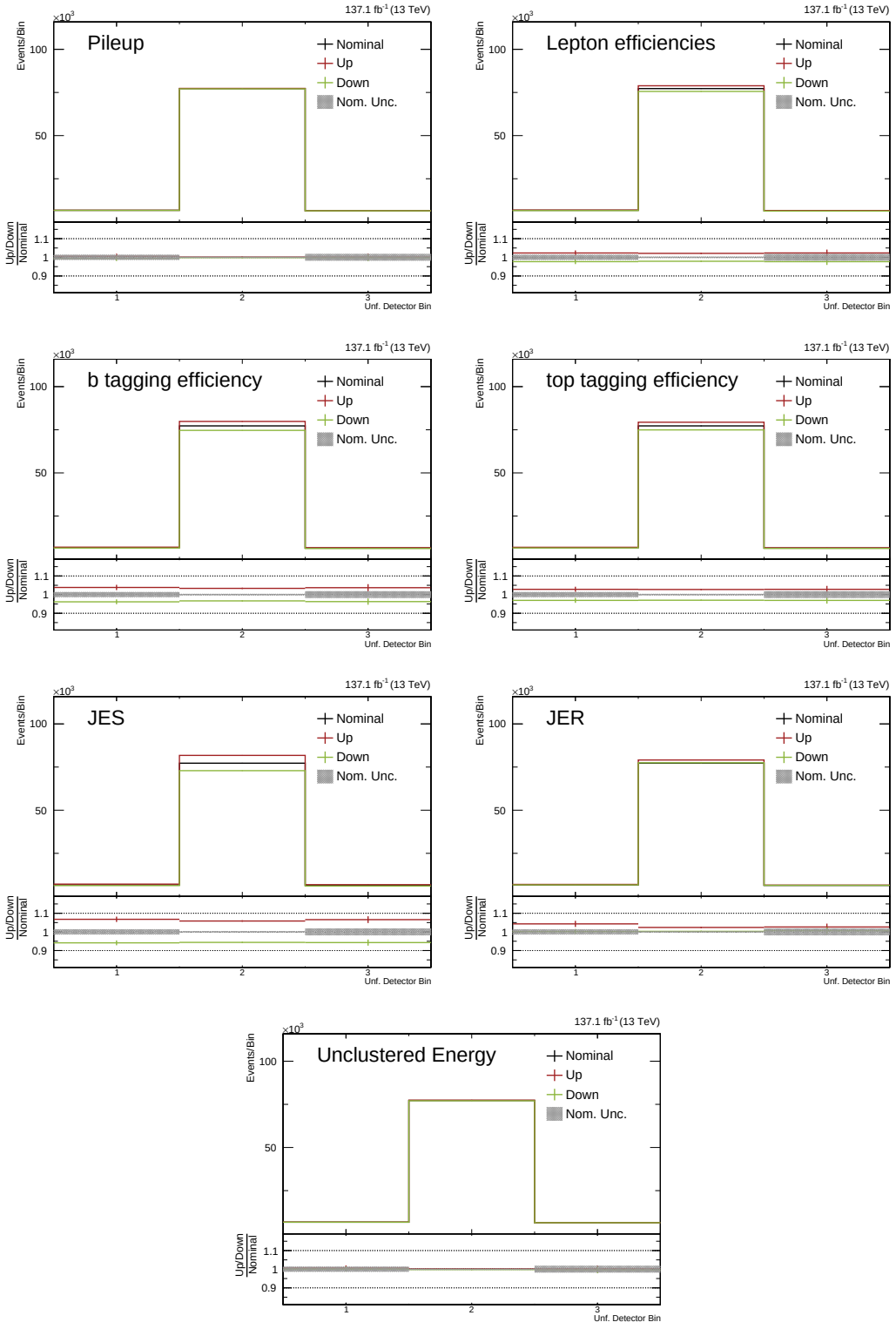

Figure E.1: Shape variations for experimental systematic uncertainties: The up and down shifts for the unfolding bin are shown for the experimental systematic uncertainties that vary the shape and the normalization of the distribution.

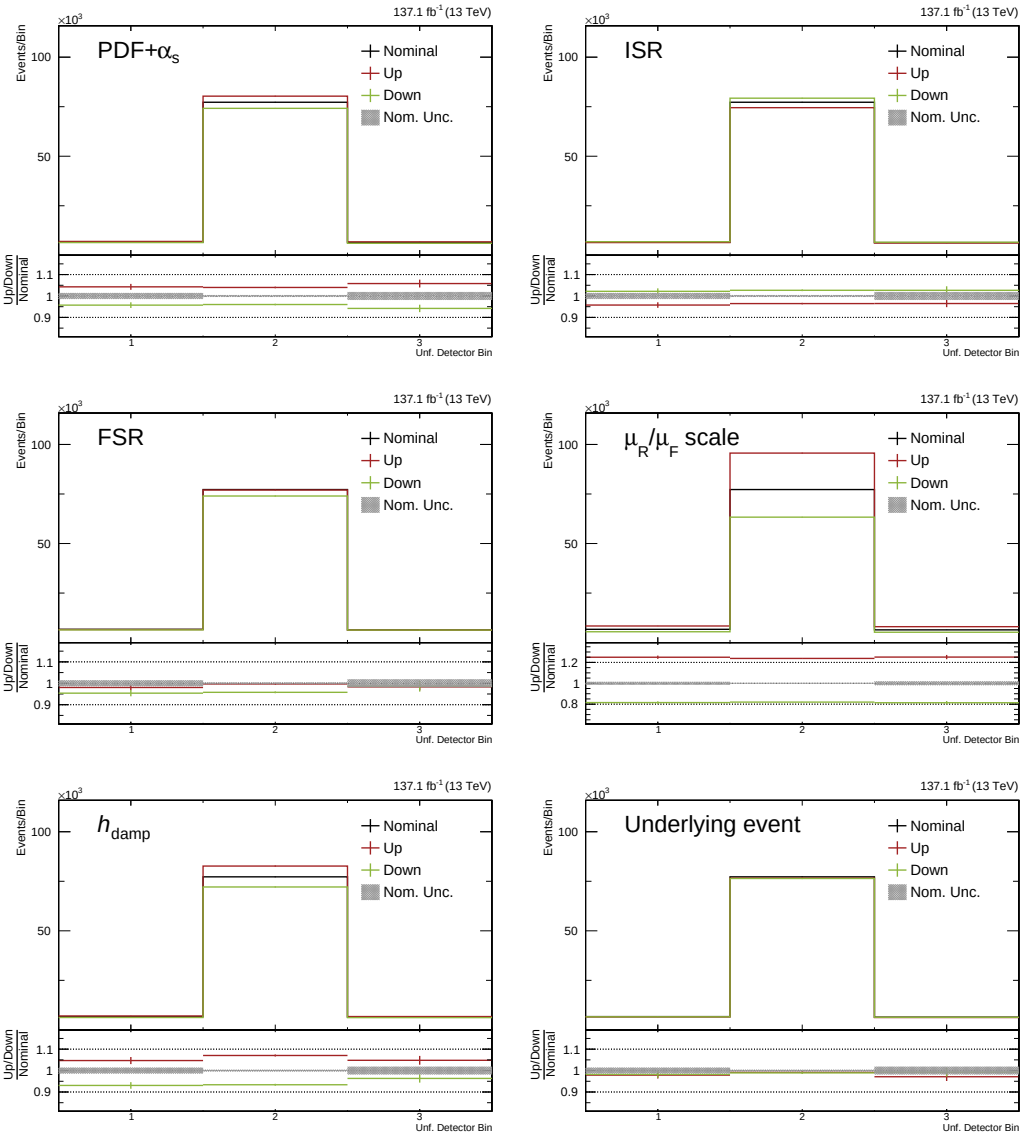

Figure E.2: Shape variations for theoretical systematic uncertainties: The up and down shifts for the unfolding bin are shown for the theoretical systematic uncertainties that vary the shape and the normalization of the distribution.

F. Energy Asymmetry for Alternative Selection

For the alternative selection the selection requirement on $|y_{t\bar{t}j}| > 0.5$ is removed. The corresponding binning for the unfolding bin is detailed in Table F.1. The distribution of the unfolding bin on truth level and detector level as well as the corresponding migration matrices for the two signal samples are given in Figure F.1. Figure F.2 shows the distribution on detector level including all background processes and measured data. The impact due to the systematic uncertainty sources that cause a shape variation for the unfolding bin is shown in Figures F.3, F.4, and F.5. The prefit and postfit distribution for the unfolding bin is shown in Figure F.6. The following postfit values are obtained for the signal strength parameters:

$$r_{\text{neg}} = \quad 0.958\,^{+0.017}_{-0.017}\,(\text{stat})\,^{+0.295}_{-0.254}\,(\text{syst})\,,$$

$$r_{t\bar{t}} = \quad 1.006\,^{+0.006}_{-0.005}\,(\text{stat})\,^{+0.277}_{-0.239}\,(\text{syst})\,,$$

$$r_{\text{Asym}} = -0.00606\,^{+0.01345}_{-0.01341}\,(\text{stat})\,^{+0.01786}_{-0.02116}\,(\text{syst})\,.$$

This yields an value for the unfolded energy asymmetry in the alternative selection for:

$$A^{\text{opt}}_{E,\text{unf.}} = -0.60\,\%\,^{+2.24\,\%}_{-2.51\,\%}\,(\text{stat + syst})\,. \tag{F.1}$$

In Table F.2 the impact of the systematic uncertainty sources is given separately for each of the sources. Further details about the procedure and a discussion of the results is given in 5.7 and 5.8.

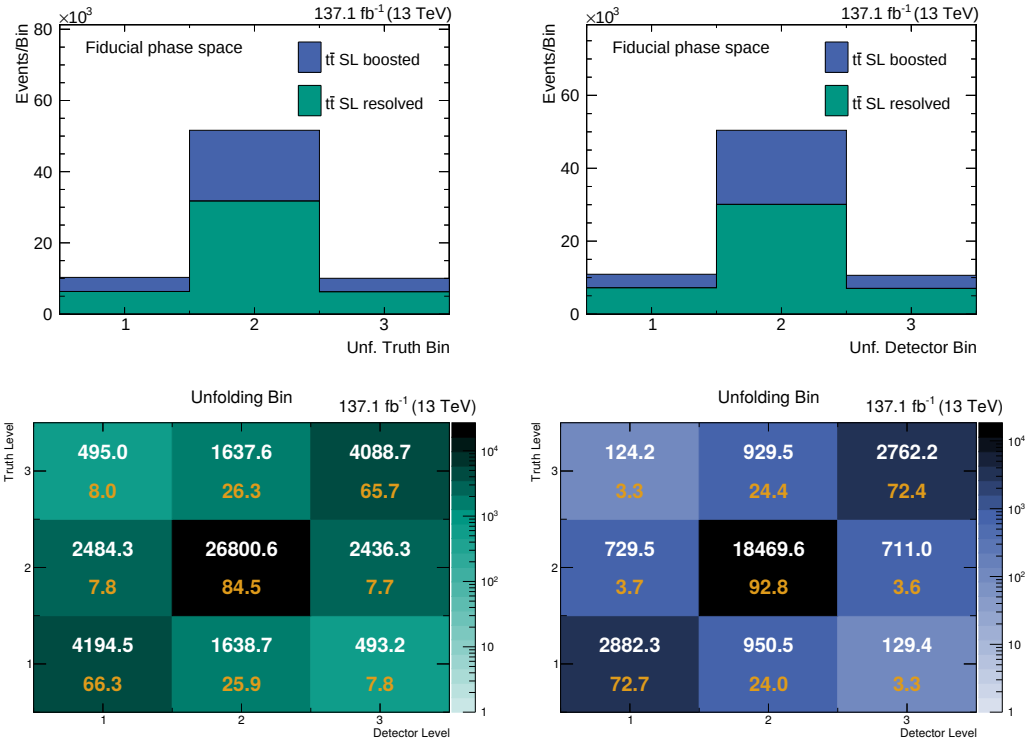

Figure F.1: Distributions and migration matrices for the unfolding binning: The upper row shows the distributions for the unfolding binning on truth level (left) and detector level (right). This binning scheme is also deployed for the migration matrices in the lower row, which show the migration of events between detector level and truth level for the signal events in the resolved (left) and boosted regime (right). The white numbers correspond to the scaled event numbers, while the orange numbers show the percentage of events of one truth-level bin in the different detector-level bins and are normalized to 100 % for each row.

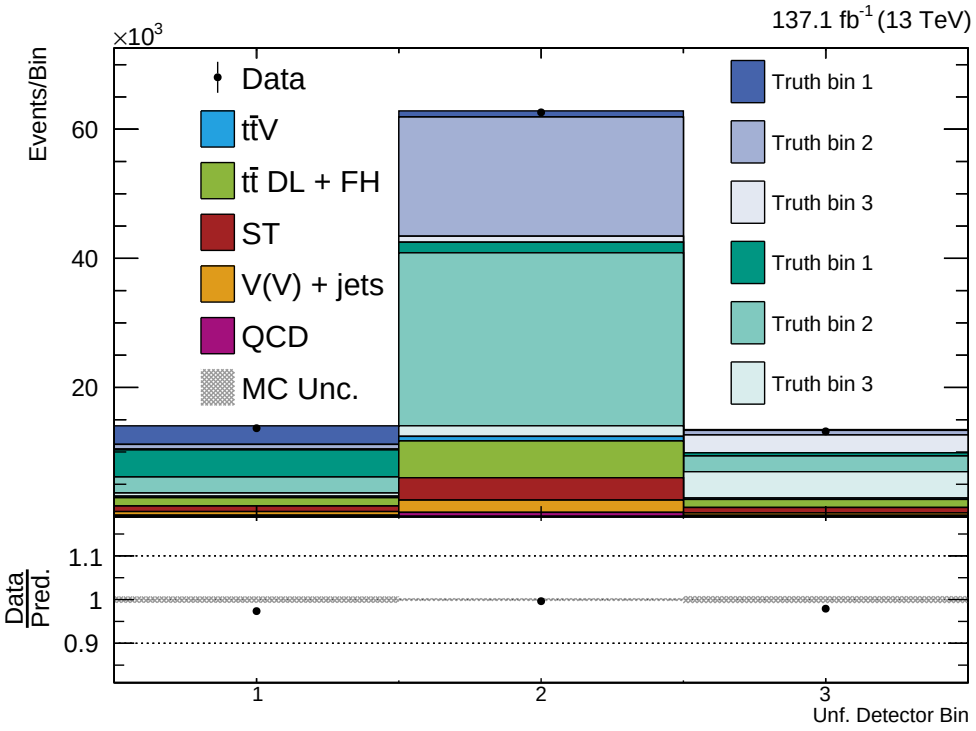

Figure F.2: Distribution of the unfolding bin: The distribution of the unfolding bin is shown for simulation and measured data. For the simulation the two signal samples are split by the truth-level contributions and the contents of the respective truth-level bins 1, 2, and 3 are indicated separately for the filtered (blue) and unfiltered (green) sample.

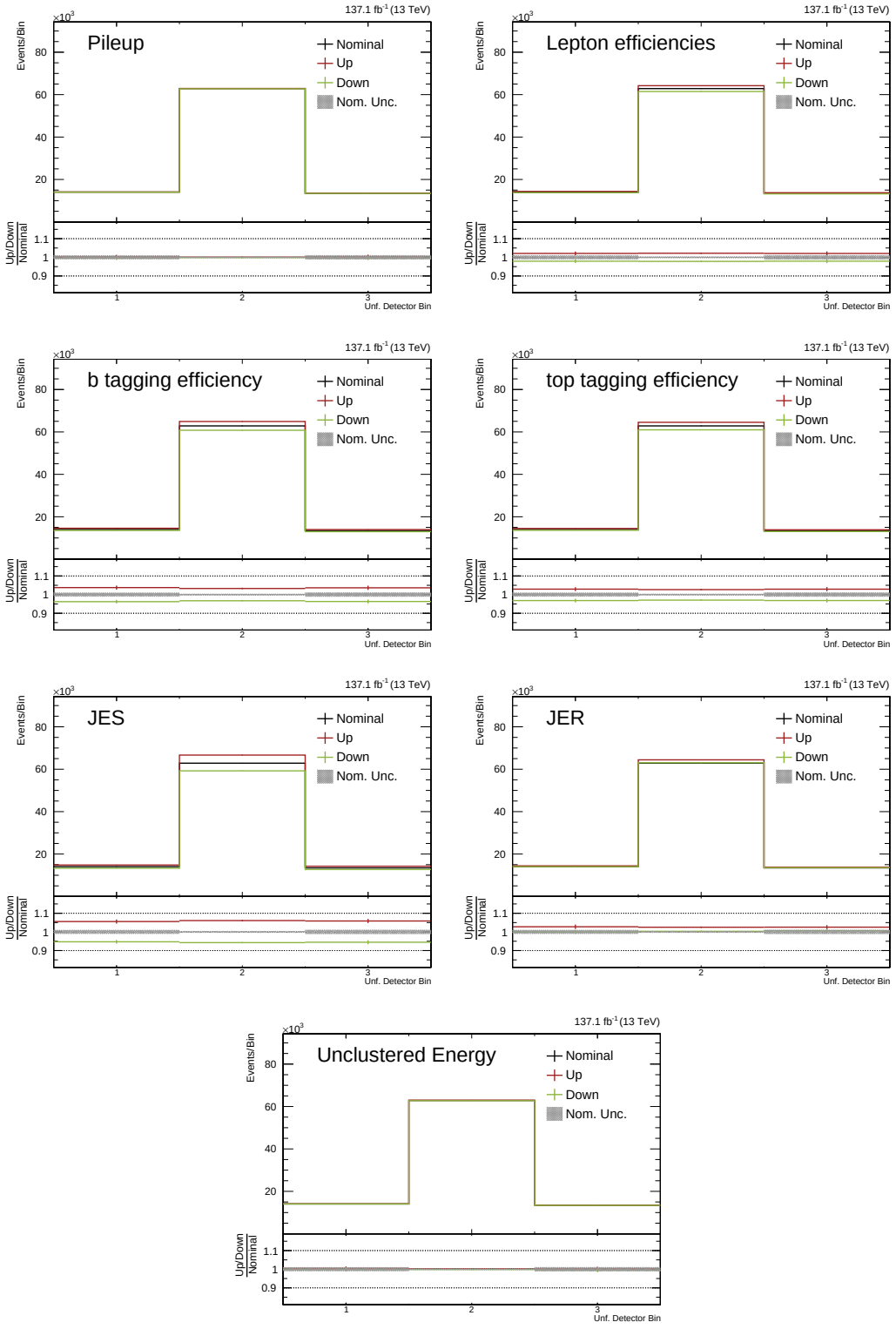

Figure F.3: Shape variations for experimental systematic uncertainties: The up and down shifts for the unfolding bin are shown for the experimental systematic uncertainties that vary the shape and the normalization of the distribution.

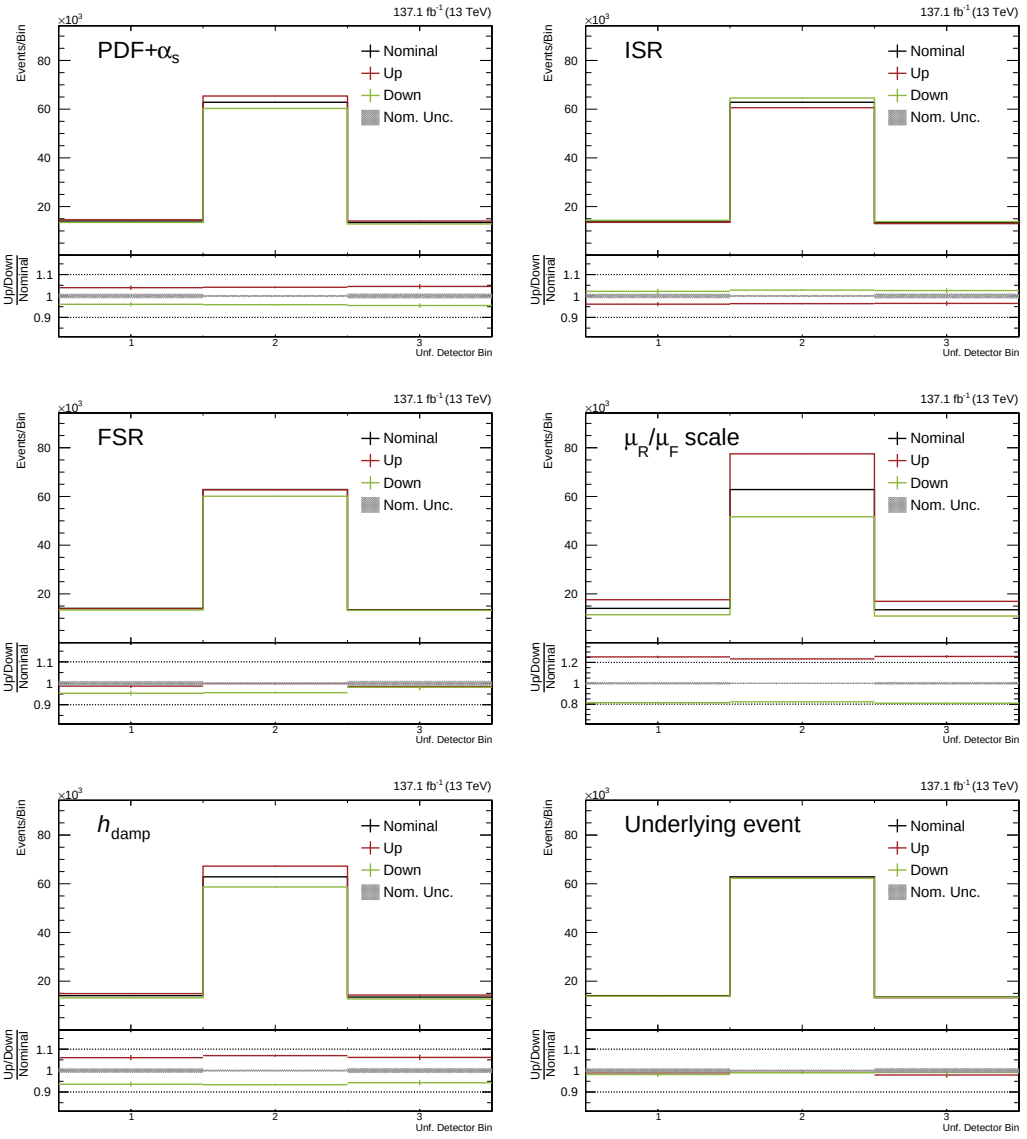

Figure F.4: Shape variations for theoretical systematic uncertainties: The up and down shifts for the unfolding bin are shown for the theoretical systematic uncertainties that vary the shape and the normalization of the distribution.

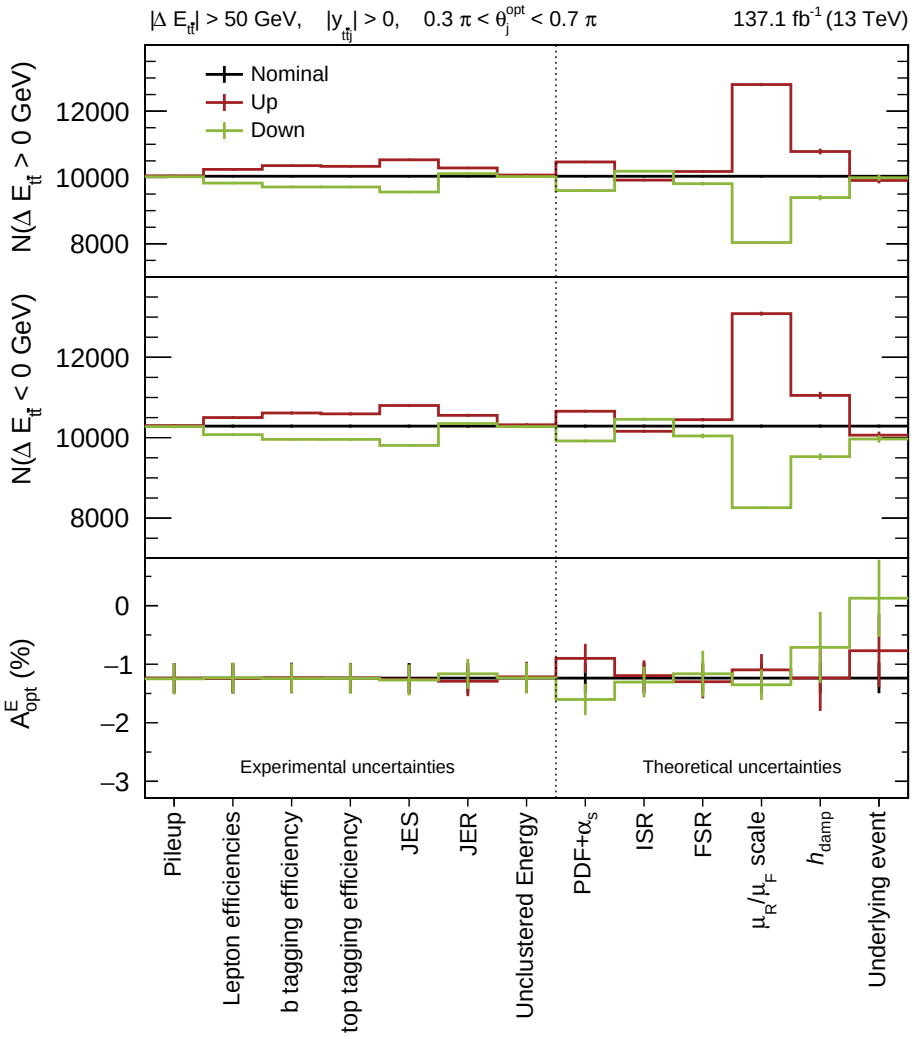

Figure F.5: Impact of systematic uncertainties: The impact of the systematic uncertainties that yield a shape variation for the unfolding distribution is shown on truth level in respect to the nominal distributions. The two upper rows show the event yields for the up and down shifts of the systematic uncertainties for both events with $\Delta E_{t\bar{t}} > 0\,\text{GeV}$ (upper row) and $\Delta E_{t\bar{t}} < 0\,\text{GeV}$ (middle row). The lower row shows the respective variations for the energy asymmetry value. The error bars indicate the uncertainty due to the limited size of the simulation samples.

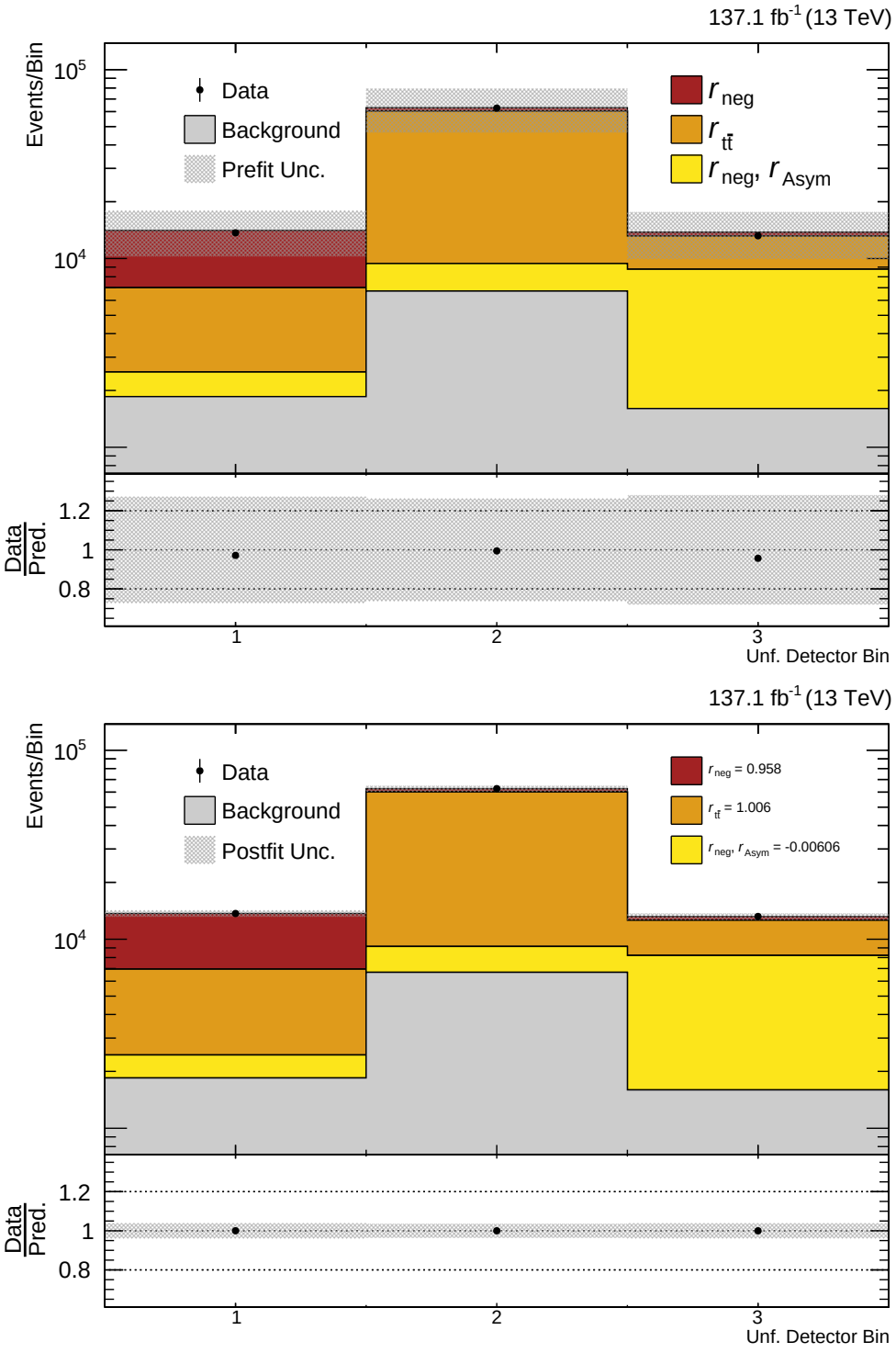

Figure F.6: Prefit and postfit distribution of the unfolding bin: The distribution of the unfolding bin for simulation and measured data is shown together with the prefit (upper row) and postfit (lower row) uncertainty on simulation due the systematic uncertainty sources.

Table F.1: Binning scheme for the unfolding procedure: Three different bins are defined in order to have a minimum set of bins that provides full access to the relevant information on the three variables of interest for the studied phase space of the energy asymmetry.

Bin 1	**Bin 2**	**Bin 3**
	$0 < \theta_j^{\text{opt}} < 0.3\pi$	
$0.3\pi < \theta_j^{\text{opt}} < 0.7\pi$	**OR**	$0.3\pi < \theta_j^{\text{opt}} < 0.7\pi$
	$0.7\pi < \theta_j^{\text{opt}} < \pi$	
AND	**OR**	**AND**
$\Delta E_{\text{t}\bar{\text{t}}} < -50\,\text{GeV}$	$-50\,\text{GeV} < \Delta E_{\text{t}\bar{\text{t}}} < 50\,\text{GeV}$	$\Delta E_{\text{t}\bar{\text{t}}} > 50\,\text{GeV}$

Table F.2: Impact of systematic uncertainties on signal strength parameters: For each signal strength parameter the up and down shift as well as the relative change in % is given for each systematic uncertainty source. The values of r_{Asym} are multiplied by 100 in order to be directly comparable to A_E^{opt}.

Uncertainty source	r_{neg}		$r_{t\bar{t}}$		$100 \cdot r_{\mathrm{Asym}}$	
Experimental uncertainties						
Pileup	$0.958^{-0.005}_{+0.005}$	$-0.5\,\%$ $+0.5\,\%$	$1.006^{+0.000}_{-0.001}$	$+0.0\,\%$ $-0.1\,\%$	$-0.606^{+0.033}_{+0.007}$	$+5.4\,\%$ $+1.1\,\%$
Lepton efficiencies	$0.958^{-0.011}_{+0.005}$	$-1.1\,\%$ $+0.6\,\%$	$1.006^{-0.011}_{+0.007}$	$-1.1\,\%$ $+0.7\,\%$	$-0.606^{+0.073}_{-0.021}$	$+12.1\,\%$ $-3.5\,\%$
b tagging efficiency	$0.958^{-0.027}_{+0.022}$	$-2.8\,\%$ $+2.3\,\%$	$1.006^{-0.016}_{+0.011}$	$-1.6\,\%$ $+1.1\,\%$	$-0.606^{+0.208}_{-0.124}$	$+34.3\,\%$ $-20.5\,\%$
top tagging efficiency	$0.958^{-0.019}_{+0.014}$	$-2.0\,\%$ $+1.4\,\%$	$1.006^{-0.012}_{+0.008}$	$-1.2\,\%$ $+0.8\,\%$	$-0.606^{+0.085}_{-0.038}$	$+13.9\,\%$ $-6.3\,\%$
JES	$0.958^{-0.032}_{+0.023}$	$-3.4\,\%$ $+2.4\,\%$	$1.006^{-0.043}_{+0.032}$	$-4.2\,\%$ $+3.2\,\%$	$-0.606^{-0.248}_{+0.214}$	$-40.9\,\%$ $+35.3\,\%$
JER	$0.958^{-0.018}_{+0.005}$	$-1.9\,\%$ $+0.6\,\%$	$1.006^{-0.011}_{-0.002}$	$-1.1\,\%$ $-0.2\,\%$	$-0.606^{+0.307}_{-0.727}$	$+50.7\,\%$ $-119.8\,\%$
Unclustered Energy	$0.958^{-0.005}_{-0.001}$	$-0.5\,\%$ $-0.1\,\%$	$1.006^{-0.000}_{-0.000}$	$-0.0\,\%$ $-0.0\,\%$	$-0.606^{+0.264}_{+0.395}$	$+43.5\,\%$ $+65.2\,\%$
Luminosity	$0.958^{-0.015}_{+0.006}$	$-1.6\,\%$ $+0.6\,\%$	$1.006^{-0.014}_{+0.005}$	$-1.4\,\%$ $+0.5\,\%$	$-0.606^{+0.054}_{-0.011}$	$+8.9\,\%$ $-1.7\,\%$
BB-lite (Bin 1)	$0.958^{-0.013}_{+0.012}$	$-1.4\,\%$ $+1.3\,\%$	$1.006^{+0.002}_{-0.003}$	$+0.2\,\%$ $-0.3\,\%$	$-0.605^{+0.818}_{-0.768}$	$+135.3\,\%$ $-127.0\,\%$
BB-lite (Bin 2)	$0.958^{+0.004}_{-0.005}$	$+0.4\,\%$ $-0.5\,\%$	$1.006^{-0.001}_{+0.001}$	$-0.1\,\%$ $+0.1\,\%$	$-0.606^{+0.004}_{+0.036}$	$+0.6\,\%$ $+6.0\,\%$
BB-lite (Bin 3)	$0.958^{-0.000}_{-0.001}$	$-0.0\,\%$ $-0.1\,\%$	$1.006^{-0.000}_{-0.000}$	$-0.0\,\%$ $-0.0\,\%$	$-0.605^{-0.822}_{+0.855}$	$-135.8\,\%$ $+141.2\,\%$
Theoretical uncertainties						
PDF+α_S	$0.958^{-0.018}_{+0.009}$	$-1.8\,\%$ $+0.9\,\%$	$1.006^{-0.022}_{+0.014}$	$-2.2\,\%$ $+1.4\,\%$	$-0.606^{-0.406}_{+0.514}$	$-67.0\,\%$ $+84.8\,\%$
ISR	$0.958^{-0.002}_{-0.005}$	$-0.3\,\%$ $-0.5\,\%$	$1.006^{+0.004}_{-0.013}$	$+0.4\,\%$ $-1.3\,\%$	$-0.606^{+0.009}_{+0.044}$	$+1.5\,\%$ $+7.3\,\%$
FSR	$0.958^{+0.001}_{-0.008}$	$+0.1\,\%$ $-0.9\,\%$	$1.006^{-0.010}_{+0.012}$	$-1.0\,\%$ $+1.2\,\%$	$-0.606^{+0.398}_{-0.481}$	$+65.7\,\%$ $-79.4\,\%$
μ_R/μ_F scale	$0.958^{-0.229}_{+0.249}$	$-23.9\,\%$ $+26.0\,\%$	$1.006^{-0.207}_{+0.225}$	$-20.5\,\%$ $+22.4\,\%$	$-0.606^{-0.015}_{+0.040}$	$-2.5\,\%$ $+6.7\,\%$
h_{damp}	$0.958^{-0.030}_{+0.035}$	$-3.2\,\%$ $+3.6\,\%$	$1.006^{-0.050}_{+0.042}$	$-5.0\,\%$ $+4.1\,\%$	$-0.606^{-0.009}_{-0.852}$	$-1.5\,\%$ $-140.5\,\%$
Underlying event	$0.958^{+0.004}_{+0.016}$	$+0.4\,\%$ $+1.7\,\%$	$1.006^{-0.000}_{-0.001}$	$-0.0\,\%$ $-0.1\,\%$	$-0.606^{+1.143}_{-0.986}$	$+189.7\,\%$ $-163.6\,\%$
σ(TT Semilep HT250)	$0.958^{-0.003}_{-0.003}$	$-0.3\,\%$ $-0.3\,\%$	$1.006^{-0.008}_{+0.003}$	$-0.8\,\%$ $+0.3\,\%$	$-0.606^{+0.030}_{+0.053}$	$+5.0\,\%$ $+8.8\,\%$
σ(TT Semilep)	$0.958^{-0.017}_{+0.009}$	$-1.7\,\%$ $+0.9\,\%$	$1.006^{-0.013}_{+0.005}$	$-1.3\,\%$ $+0.5\,\%$	$-0.606^{-0.000}_{+0.050}$	$-0.0\,\%$ $+8.2\,\%$
σ(TT FH + DL)	$0.958^{-0.002}_{+0.000}$	$-0.2\,\%$ $+0.0\,\%$	$1.006^{-0.001}_{+0.000}$	$-0.1\,\%$ $+0.0\,\%$	$-0.606^{+0.030}_{+0.038}$	$+5.0\,\%$ $+6.3\,\%$
σ(Single Top)	$0.958^{-0.006}_{+0.002}$	$-0.6\,\%$ $+0.2\,\%$	$1.006^{-0.002}_{+0.000}$	$-0.2\,\%$ $+0.0\,\%$	$-0.606^{+0.048}_{+0.037}$	$+7.9\,\%$ $+6.2\,\%$
σ(V + V/Jets)	$0.958^{-0.003}_{+0.002}$	$-0.3\,\%$ $+0.2\,\%$	$1.006^{-0.000}_{-0.000}$	$-0.0\,\%$ $-0.0\,\%$	$-0.606^{+0.127}_{-0.107}$	$+20.9\,\%$ $-17.7\,\%$
σ(QCD, TT + X)	$0.958^{-0.014}_{+0.006}$	$-1.5\,\%$ $+0.7\,\%$	$1.006^{-0.003}_{-0.001}$	$-0.3\,\%$ $-0.1\,\%$	$-0.606^{+0.171}_{-0.048}$	$+28.3\,\%$ $-7.9\,\%$

List of Figures

List of Tables

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. B*, vol. 716, pp. 1–29, 2012.

[2] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett. B*, vol. 716, pp. 30–61, 2012.

[3] CDF and DØ Collaborations, "Combined Forward-Backward Asymmetry Measurements in Top-Antitop Quark Production at the Tevatron", *Phys. Rev. Lett.*, vol. 120, p. 042001, Jan 2018.

[4] ATLAS and CMS Collaborations, "Combination of inclusive and differential $t\bar{t}$ charge asymmetry measurements using ATLAS and CMS data at $\sqrt{s}$ = 7 and 8 TeV", *JHEP*, vol. 04, p. 033, 2018.

[5] L. Canetti, M. Drewes, and M. Shaposhnikov, "Matter and antimatter in the universe", *New Journal of Physics*, vol. 14, p. 095012, Sep 2012.

[6] CDF Collaboration, "Observation of top quark production in $\bar{p}p$ collisions", *Phys. Rev. Lett.*, vol. 74, pp. 2626–2631, 1995.

[7] DØ Collaboration, "Observation of the top quark", *Phys. Rev. Lett.*, vol. 74, pp. 2632–2637, 1995.

[8] M. Kobayashi and T. Maskawa, "CP-Violation in the Renormalizable Theory of Weak Interaction", *Progress of Theoretical Physics*, vol. 49, no. 2, pp. 652–657, 1973.

[9] Particle Data Group, "Review of Particle Physics", *Phys. Rev. D*, vol. 98, p. 030001, 2018.

[10] R. Oerter, "The theory of almost everything : the Standard Model, the unsung triumph of modern physics", Pi Press, New York, 2006.

[11] D. Griffiths, "Introduction to elementary particles", Wiley, New York, 1987.

[12] R. Feynman, "QED : the strange theory of light and matter", Princeton University Press, Princeton, N.J, 2006.

[13] S. L. Glashow, "Partial-symmetries of weak interactions", *Nuclear Physics*, vol. 22, no. 4, pp. 579 – 588, 1961.

[14] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.*, vol. 13, pp. 508–509, 1964.

[15] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.*, vol. 13, pp. 321–323, 1964.

[16] ATLAS, CDF, CMS and DØ Collaborations, "First combination of Tevatron and LHC measurements of the top-quark mass", 2014. arXiv:1403.4427 [hep-ex].

[17] S. Mandelstam, "Determination of the Pion-Nucleon Scattering Amplitude from Dispersion Relations and Unitarity. General Theory", *Phys. Rev.*, vol. 112, pp. 1344–1360, 1958.

[18] LHC Top Physics Working Group, "NNLO+NNLL top-quark-pair cross sections", https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO. Accessed: 2020-10-30.

[19] M. Czakon and A. Mitov, "Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders", *Comput. Phys. Commun.*, vol. 185, p. 2930, 2014.

[20] LHC Top Physics Working Group, "NLO single-top channel cross sections", https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopRefXsec. Accessed: 2020-10-30.

[21] J. H. Kühn and G. Rodrigo, "Charge asymmetry of heavy quarks at hadron colliders", *Phys. Rev. D*, vol. 59, p. 054017, Feb 1999.

[22] J. H. Kühn and G. Rodrigo, "Charge asymmetry in hadroproduction of heavy quarks", *Phys. Rev. Lett.*, vol. 81, pp. 49–52, Jul 1998.

[23] M. Czakon, P. Fiedler, and A. Mitov, "Resolving the tevatron top quark forward-backward asymmetry puzzle: Fully differential next-to-next-to-leading-order calculation", *Phys. Rev. Lett.*, vol. 115, p. 052001, Jul 2015.

[24] J. H. Kühn and G. Rodrigo, "Charge asymmetries of top quarks at hadron colliders revisited", *JHEP*, vol. 01, p. 063, 2012.

[25] ATLAS Collaboration, "Inclusive and differential measurement of the charge asymmetry in $t\bar{t}$ events at 13 TeV with the ATLAS detector", ATLAS-CONF-2019-026, CERN, Geneva, Jul 2019.

[26] F. Roscher, "Charge Asymmetry Measurements in Top Quark Pair Production at $\sqrt{s} = 8$ TeV with the CMS Experiment", 2015. CERN-THESIS-2015-242.

[27] S. Berge and S. Westhoff, "Charge Asymmetry in Top Pair plus Jet Production – A Snowmass White Paper", in *Community Summer Study 2013: Snowmass on the Mississippi*, 7 2013.

[28] S. Berge and S. Westhoff, "Top-Quark Charge Asymmetry Goes Forward: Two New Observables for Hadron Colliders", *JHEP*, vol. 07, p. 179, 2013.

[29] S. Berge and S. Westhoff, "Observing the Top Energy Asymmetry at the LHC", *Phys. Rev. D*, vol. 95, no. 1, p. 014035, 2017.

[30] S. Westhoff, "Review of top charge asymmetries for LHC run II", *PoS*, vol. CKM2016, p. 120, 2017.

[31] A. Basan, P. Berta, L. Masetti, E. Vryonidou, and S. Westhoff, "Measuring the top energy asymmetry at the LHC: QCD and SMEFT interpretations", *JHEP*, vol. 03, p. 184, 2020.

[32] Belle-II Collaboration, "The Belle II Physics Book", *PTEP*, vol. 2019, no. 12, p. 123C01, 2019.

[33] G. Bachy, A. Hofmann, S. Myers, E. Picasso, and G. Plass, "The LEP collider: construction, project status and outlook", *Part. Accel.*, vol. 26, pp. 19–32, 1990.

[34] S. Holmes, "Tevatron Run II Handbook", 1998.

[35] L. Evans and P. Bryant, "LHC Machine", *JINST*, vol. 3, p. S08001, 2008.

[36] CERN, "LHC Guide", CERN-Brochure-2017-002-Eng, 2017.

[37] CERN, "The Large Hadron Collider", https://home.cern/science/accelerators/large-hadron-collider. Accessed: 2020-10-30.

[38] CERN, "A new schedule for the LHC and its successor", https://home.cern/news/news/accelerators/new-schedule-lhc-and-its-successor. Accessed: 2020-10-30.

[39] E. Mobs, "The CERN accelerator complex - 2019. Complexe des accélérateurs du CERN - 2019", 2019. General Photo, CERN-GRAPHICS-2019-002.

[40] CMS Luminosity Physics Object Group, "Public CMS Luminosity Information", https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults. Accessed: 2020-10-30.

[41] CMS Collaboration, "The CMS Experiment at the CERN LHC", *JINST*, vol. 3, p. S08004, 2008.

[42] D. Barney, "CMS Detector Slice", https://cds.cern.ch/record/2120661, Oct. 2016.

[43] UZH CMS Group, "CMS Wiki Pages", https://wiki.physik.uzh.ch/cms/latex:example_spherical_coordinates. Accessed: 2020-10-30.

[44] A. Dominguez, D. Abbaneo, K. Arndt, N. Bacchetta, A. Ball, E. Bartz, W. Bertl, G. M. Bilei, G. Bolla, H. W. K. Cheung, *et al.*, "CMS Technical Design Report for the Pixel Detector Upgrade", CERN-LHCC-2012-016. CMS-TDR-11, 2012.

[45] CMS Collaboration, "The CMS electromagnetic calorimeter project: Technical Design Report", CERN-LHCC-97-033, 1997.

[46] P. Bloch, R. Brown, P. Lecoq, and H. Rykaczewski, "Changes to CMS ECAL electronics: addendum to the Technical Design Report", CERN-LHCC-2002-027, 2002.

[47] A. Benaglia, "The CMS ECAL performance with examples", *JINST*, vol. 9, no. 02, p. C02008, 2014.

[48] Q. Ingram, "Energy resolution of the barrel of the CMS electromagnetic calorimeter", *JINST*, vol. 2, no. 04, pp. P04004–P04004, 2007.

[49] CMS Collaboration, "The CMS hadron calorimeter project: Technical Design Report", CERN-LHCC-97-031, 1997.

[50] V. Kryshkin and A. Ronzhin, "An optical fiber readout for scintillator calorimeters", *Nucl. Instrum. Methods Phys. Res. A*, vol. 247, no. 3, pp. 583 – 585, 1986.

[51] CMS Collaboration, "Precise mapping of the magnetic field in the CMS barrel yoke using cosmic rays", *JINST*, vol. 5, no. 03, pp. T03021–T03021, 2010.

[52] CMS Collaboration, "The CMS muon project: Technical Design Report", CERN-LHCC-97-032, 1997.

[53] G. Abbiendi, "The CMS muon system in Run2: preparation, status and first results", in *Proceedings, 2015 European Physical Society Conference on High Energy Physics (EPS-HEP 2015): Vienna, Austria, July 22-29, 2015*, p. 237, 2015.

[54] CMS Collaboration, "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV", *JINST*, vol. 13, no. 06, p. P06015, 2018.

[55] CMS Collaboration, "CMS Technical Design Report for the Muon Endcap GEM Upgrade", CERN-LHCC-2015-012. CMS-TDR-013, 2015.

[56] CMS Collaboration, "The Phase-2 Upgrade of the CMS Muon Detectors", CERN-LHCC-2017-012. CMS-TDR-016, CERN, 2017.

[57] CMS Collaboration, "CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project", CERN-LHCC-2002-026, 2002.

[58] CMS Collaboration, "CMS Technical Design Report for the Level-1 Trigger Upgrade", CERN-LHCC-2013-011. CMS-TDR-12, 2013.

[59] I. Bird, K. Bos, N. Brook, D. Duellmann, C. Eck, I. Fisk, D. Foster, B. Gibbard, C. Grandi, F. Grey, *et al.*, "LHC computing Grid. Technical design report", CERN-LHCC-2005-024, 2005.

[60] I. Bird *et al.*, "Update of the Computing Models of the WLCG and the LHC Experiments", CERN-LHCC-2014-014, LCG-TDR-002, 2014.

[61] Worldwide LHC Computing Grid, "WLCG Document Repository", http://wlcg-docs.web.cern.ch/wlcg-docs/?dir=outreach/images. Accessed: 2020-10-30.

[62] V. Blobel and E. Lohrmann, "Statistische und numerische Methoden der Datenanalyse", Teubner, Stuttgart Leipzig, 1998.

[63] A. L. Read, "Linear interpolation of histograms", *Nucl. Instrum. Methods Phys. Res. A*, vol. 425, pp. 357–360, 1999.

[64] J. S. Conway, "Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra", in *Proceedings, PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN, Geneva, Switzerland 17-20 January 2011*, pp. 115–120, 2011.

[65] R. Barlow and C. Beeston, "Fitting using finite Monte Carlo samples", *Comput. Phys. Commun.*, vol. 77, no. 2, pp. 219 – 228, 1993.

[66] ATLAS and CMS Collaborations, LHC Higgs Combination Group, "Procedure for the LHC Higgs boson search combination in Summer 2011", CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, 2011.

[67] CMS Higgs Working Group, "Combine", https://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/. Accessed: 2020-10-30.

[68] W. Verkerke and D. P. Kirkby, "The RooFit toolkit for data modeling", *eConf*, vol. C0303241, p. MOLT007, 2003.

[69] G. Cowan, "Statistical data analysis", Oxford science publications, 1998.

[70] A. Hocker and V. Kartvelishvili, "SVD approach to data unfolding", *Nucl. Instrum. Meth. A*, vol. 372, pp. 469–481, 1996.

[71] S. Schmitt, "TUnfold: an algorithm for correcting migration effects in high energy physics", *JINST*, vol. 7, p. T10003, 2012.

[72] G. D'Agostini, "A multidimensional unfolding method based on Bayes' Theorem", DESY-94-099, DESY, Hamburg, Jun 1994.

[73] R. D. Cousins, S. J. May, and Y. Sun, "Should unfolded histograms be used to test hypotheses?", Jul 2016.

[74] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm", *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.

[75] A. N. Tikhonov and V. Y. Arsenin, "Solutions of ill-posed problems", 1977.

[76] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss, "TMVA: Toolkit for Multivariate Data Analysis", *Proceeding of Science*, vol. ACAT, p. 040, 2007.

[77] R. Brun and F. Rademakers, "ROOT: An object oriented data analysis framework", *Nucl.Instrum.Meth.A*, vol. 389, pp. 81–86, 1997.

[78] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting", *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119 – 139, 1997.

[79] S. Fink, "Probing the Top-Yukawa Coupling by Searching for Associated Higgs Boson Production with a Single Top Quark at the CMS Experiment", 2016. CERN-THESIS-2016-361.

[80] S. Höche, "Introduction to parton-shower event generators", in *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders (TASI 2014): Boulder, Colorado, June 2-27, 2014*, pp. 235–295, 2015.

[81] 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*, vol. 46, pp. 641–653, 1977. [Zh. Eksp. Teor. Fiz.73,1216(1977)].

[82] V. N. Gribov and L. N. Lipatov, "Deep inelastic e p scattering in perturbation theory", *Sov. J. Nucl. Phys.*, vol. 15, pp. 438–450, 1972. [Yad. Fiz.15,781(1972)].

[83] G. Altarelli and G. Parisi, "Asymptotic freedom in parton language", *Nucl.Phys. B*, vol. 126, no. 2, pp. 298 – 318, 1977.

[84] R. D. Ball *et al.*, "Parton distributions for the LHC Run II", *JHEP*, vol. 04, p. 040, 2015.

[85] R. D. Ball *et al.*, "Parton distributions from high-precision collider data", *Eur. Phys. J. C*, vol. 77, p. 663, 2017.

[86] V. V. Sudakov, "Vertex parts at very high-energies in quantum electrodynamics", *Sov. Phys. JETP*, vol. 3, pp. 65–71, 1956. [Zh. Eksp. Teor. Fiz.30,87(1956)].

[87] A. Sen, "Asymptotic Behavior of the Sudakov Form-Factor in QCD", *Phys. Rev. D*, vol. 24, p. 3281, 1981.

[88] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, "Matching matrix elements and shower evolution for top-quark production in hadronic collisions", *JHEP*, vol. 01, p. 013, 2007.

[89] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO", *JHEP*, vol. 12, p. 061, 2012.

[90] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, "Parton Fragmentation and String Dynamics", *Phys. Rept.*, vol. 97, pp. 31–145, 1983.

[91] N. Metropolis and S. Ulam, "The Monte Carlo Method", *Journal of the American Statistical Association*, vol. 44, no. 247, pp. 335–341, 1949.

[92] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, "MadGraph 5 : Going Beyond", *JHEP*, vol. 06, p. 128, 2011.

[93] S. Frixione and B. R. Webber, "Matching NLO QCD computations and parton shower simulations", *JHEP*, vol. 06, p. 029, 2002.

[94] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP*, vol. 07, p. 079, 2014.

[95] P. Nason, "A new method for combining NLO QCD with shower Monte Carlo algorithms", *JHEP*, vol. 11, p. 040, 2004.

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

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

[98] T. Sjöstrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1", *Comput. Phys. Commun.*, vol. 178, pp. 852–867, 2008.

[99] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, "An Introduction to PYTHIA 8.2", *Comput. Phys. Commun.*, vol. 191, pp. 159–177, 2015.

[100] CMS Collaboration, "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements", *Eur. Phys. J. C*, vol. 80, no. 1, p. 4, 2020.

[101] S. Agostinelli *et al.*, "Geant4—a simulation toolkit", *Nucl. Instrum. Methods Phys. Res. A*, vol. 506, no. 3, pp. 250 – 303, 2003.

[102] J. Allison *et al.*, "Geant4 developments and applications", *IEEE Trans. Nucl. Sci.*, vol. 53, p. 270, 2006.

[103] J. Allison *et al.*, "Recent developments in Geant4", *Nucl. Instrum. Methods Phys. Res. A*, vol. 835, pp. 186 – 225, 2016.

[104] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *JINST*, vol. 12, no. 10, p. P10003, 2017.

[105] ALEPH Collaboration, "Performance of the ALEPH detector at LEP", *Nucl. Instrum. Meth. A*, vol. 360, pp. 481–506, 1995.

[106] D. Müller, "Search for single top quark production in the s channel at 13 TeV with the CMS experiment", 2020. CERN-THESIS-2020-064.

[107] CMS Collaboration, "Alignment of the CMS tracker with LHC and cosmic ray data", *JINST*, vol. 9, p. P06009, 2014.

[108] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST*, vol. 9, no. 10, p. P10009, 2014.

[109] P. Billoir, "Progressive track recognition with a Kalman-like fitting procedure", *Comput. Phys. Commun.*, vol. 57, no. 1, pp. 390 – 394, 1989.

[110] P. Billoir and S. Qian, "Simultaneous pattern recognition and track fitting by the Kalman filtering method", *Nucl. Instrum. Methods Phys. Res. A*, vol. 294, no. 1, pp. 219 – 228, 1990.

[111] R. Mankel, "A concurrent track evolution algorithm for pattern recognition in the HERA-B main tracking system", *Nucl. Instrum. Methods Phys. Res. A*, vol. 395, no. 2, pp. 169 – 184, 1997.

[112] CMS Collaboration, "Pileup Removal Algorithms", no. CMS-PAS-JME-14-001, 2014.

[113] K. Rose, "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems", *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2210–2239, 1998.

[114] R. Frühwirth, W. Waltenberger, and P. Vanlaer, "Adaptive Vertex Fitting", CMS-NOTE-2007-008, 2007.

[115] CMS Collaboration, "Baseline muon selections for Run-II", https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideMuonIdRun2. Restricted webpage. Accessed: 2020-10-30.

[116] CMS Collaboration, "Search for neutral Higgs bosons decaying to tau pairs in pp collisions at $\sqrt{s} = 7$ TeV", *Phys. Lett. B*, vol. 713, pp. 68–90, 2012.

[117] CMS Collaboration, "Electron reconstruction in CMS", *Eur. Phys. J. C*, vol. 49, pp. 1099–1116, 2007.

[118] CMS Collaboration, "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV", *JINST*, vol. 10, p. P06005, 2015.

[119] R. Frühwirth, "Track fitting with non-gaussian noise", *Comput. Phys. Commun.*, vol. 100, no. 1, pp. 1 – 16, 1997.

[120] CMS Collaboration, "Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC", CMS-NOTE-2005-001, 2005.

[121] CMS Collaboration, "Cut Based Electron ID for Run 2", https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedElectronIdentificationRun2. Restricted webpage. Accessed: 2020-10-30.

[122] C. Bierlich *et al.*, "Robust Independent Validation of Experiment and Theory: Rivet version 3", *SciPost Phys.*, vol. 8, p. 026, 2020.

[123] A. Buckley, J. Butterworth, L. Lonnblad, D. Grellscheid, H. Hoeth, J. Monk, H. Schulz, and F. Siegert, "Rivet user manual", *Comput. Phys. Commun.*, vol. 184, pp. 2803–2819, 2013.

[124] CMS Collaboration, "NanoAOD Collection", https://cms-nanoaod-integration.web.cern.ch/integration/master-102X/mc102X_doc.html. Accessed: 2020-10-30.

[125] CMS Collaboration, "NanoAOD Documentation", https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookNanoAOD. Accessed: 2020-10-30.

[126] CMS Collaboration, "Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV", *JINST*, vol. 10, no. 08, p. P08010, 2015.

[127] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, "Longitudinally invariant K_t clustering algorithms for hadron hadron collisions", *Nucl. Phys. B*, vol. 406, pp. 187–224, 1993.

[128] S. D. Ellis and D. E. Soper, "Successive combination jet algorithm for hadron collisions", *Phys. Rev. D*, vol. 48, pp. 3160–3166, 1993.

[129] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, "Better jet clustering algorithms", *JHEP*, vol. 08, p. 001, 1997.

[130] M. Wobisch and T. Wengler, "Hadronization corrections to jet cross-sections in deep inelastic scattering", in *Monte Carlo generators for HERA physics. Proceedings, Workshop, Hamburg, Germany, 1998-1999*, pp. 270–279, 1998.

[131] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", *JHEP*, vol. 04, p. 063, 2008.

[132] CMS Collaboration, "Jet Identification", https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetID. Restricted webpage. Accessed: 2020-10-30.

[133] CMS Collaboration, "Introduction to Jet Energy Corrections at CMS", https://twiki.cern.ch/twiki/bin/view/CMS/IntroToJEC. Restricted webpage. Accessed: 2020-10-30.

[134] CMS Collaboration, "Determination of jet energy calibration and transverse momentum resolution in CMS", *JINST*, vol. 6, no. 11, pp. P11002–P11002, 2011.

[135] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *JINST*, vol. 12, no. 02, pp. P02014–P02014, 2017.

[136] CMS Collaboration, "Identification of b-Quark Jets with the CMS Experiment", *JINST*, vol. 8, p. P04013, 2013.

[137] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", *JINST*, vol. 13, p. P05011, 2018.

[138] DØ Collaboration, "B-Jet Identification", http://www-d0.fnal.gov/Run2Physics/top/singletop_observation/, July 2020.

[139] CMS Collaboration, "Deep learning in jet reconstruction at CMS", *Journal of Physics: Conference Series*, vol. 1085, p. 042029, 2018.

[140] CMS Collaboration, "Recommendation for Using b-tag Objects in Physics Analyses", https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation. Restricted webpage. Accessed: 2020-10-30.

[141] CMS Collaboration, "Performance of b tagging algorithms in proton-proton collisions at 13 TeV with Phase 1 CMS detector", CMS-DP-2018-033, 2018.

[142] CMS Collaboration, "Machine learning-based identification of highly Lorentz-boosted hadronically decaying particles at the CMS experiment", CMS-PAS-JME-18-002, Geneva, 2019.

[143] CMS Collaboration, "Boosted Jet Tagging using Deep Neural Networks and low level variables: DeepAK8 Top/W, WPs and SFs for V2 Training.", https://indico.cern.ch/event/877167/contributions/3744193/attachments/1989744/3379280/DeepAK8_Top_W_SFs_V2.pdf. Restricted presentation. Accessed: 2020-10-30.

[144] CMS Collaboration, "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s}$ = 13 TeV using the CMS detector", *JINST*, vol. 14, no. 07, p. P07004, 2019.

[145] CMS Collaboration, "MET Filter Recommendations for Run II", https://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETOptionalFiltersRun2. Restricted webpage. Accessed: 2020-10-30.

[146] 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", 2010. CERN-THESIS-2010-255.

[147] LHC Top Physics Working Group, "Particle level objects and pseudo-top-quark definitions", https://twiki.cern.ch/twiki/bin/view/LHCPhysics/ParticleLevelTopDefinitions. Accessed: 2020-10-30.

[148] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, "Soft Drop", *JHEP*, vol. 05, p. 146, 2014.

[149] CMS Collaboration, "CMS Luminosity – Public Results", https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults. Accessed: 2020-10-30.

[150] CMS Collaboration, "Pileup Reweighting Utilities", https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupMCReweightingUtilities. Restricted webpage. Accessed: 2020-10-30.

[151] CMS Collaboration, "Reference muon id, isolation and trigger efficiencies for 2016 legacy re-reco data.", https://twiki.cern.ch/twiki/bin/view/CMS/MuonReferenceEffs2016LegacyRereco. Restricted webpage. Accessed: 2020-10-30.

[152] CMS Collaboration, "Reference muon id, isolation and trigger efficiencies for 2017 data.", https://twiki.cern.ch/twiki/bin/view/CMS/MuonReferenceEffs2017. Restricted webpage. Accessed: 2020-10-30.

[153] CMS Collaboration, "Reference muon id, isolation and trigger efficiencies for 2018 data.", https://twiki.cern.ch/twiki/bin/view/CMS/MuonReferenceEffs2018. Restricted webpage. Accessed: 2020-10-30.

[154] CMS Collaboration, "Tag and Probe", https://twiki.cern.ch/twiki/bin/view/CMS/MuonTagAndProbe. Restricted webpage. Accessed: 2020-10-30.

[155] CMS Collaboration, "Muon T&P Instructions for Run-II", https://twiki.cern.ch/twiki/bin/view/CMS/MuonTagAndProbeTreesRun2. Restricted webpage. Accessed: 2020-10-30.

[156] CMS Collaboration, "E/gamma RunII Recommendations", https://twiki.cern.ch/twiki/bin/view/CMS/EgammaRunIIRecommendations. Restricted webpage. Accessed: 2020-10-30.

[157] CMS Collaboration, "Tag and Probe", https://twiki.cern.ch/twiki/bin/view/CMSPublic/TagAndProbe. Accessed: 2020-10-30.

[158] CMS Collaboration, "Electron Tag-and-Probe", https://twiki.cern.ch/twiki/bin/view/CMSPublic/ElectronTagAndProbe. Accessed: 2020-10-30.

[159] P. Ott, "Inclusive cross section measurement for the production of single top quarks in the t channel at 13 TeV with the CMS experiment", 2017. Master thesis, Karlsruhe Institute of Technology (KIT), ETP-KA/2017-32.

[160] CMS Collaboration, "Measurement of the 2016 Trigger Efficiencies for a dilepton selection for a ttbar analysis", CMS Analysis Note AN-2016/392, CERN, 2016.

[161] CMS Collaboration, "Methods to apply b-tagging efficiency scale factors", https://twiki.cern.ch/twiki/bin/view/CMS/BTagSFMethods. Restricted webpage. Accessed: 2020-10-30.

[162] CMS Collaboration, "Heavy flavour tagging for 13 TeV 2016 Legacy data", https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation2016Legacy. Restricted webpage. Accessed: 2020-10-30.

[163] CMS Collaboration, "Heavy flavour tagging for 13 TeV data in 2017 and 94X MC", https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation94X. Restricted webpage. Accessed: 2020-10-30.

[164] CMS Collaboration, "Heavy flavour tagging for 13 TeV data in 2018 and 10_2_X MC", https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation102X. Restricted webpage. Accessed: 2020-10-30.

[165] CMS Collaboration, "JetMET Physics Object Group (JME POG)", https://twiki.cern.ch/twiki/bin/view/CMS/JetMET. Restricted webpage. Accessed: 2020-10-30.

[166] CMS Collaboration, "DeepAK8 (V2) Working Points and Scale Factors for 13 TeV data", https://twiki.cern.ch/twiki/bin/view/CMS/DeepAK8Tagging2018WPsSFs. Restricted webpage. Accessed: 2020-10-30.

[167] CMS Collaboration, "Estimating Systematic Errors Due to Pileup Modeling", https://twiki.cern.ch/twiki/bin/view/CMS/PileupSystematicErrors. Restricted webpage. Accessed: 2020-10-30.

[168] CMS Collaboration, "Jet energy scale uncertainty sources", https://twiki.cern.ch/twiki/bin/view/CMS/JECUncertaintySources. Restricted webpage. Accessed: 2020-10-30.

[169] CMS Collaboration, "Jet Energy Resolution", https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution. Restricted webpage. Accessed: 2020-10-30.

[170] CMS Collaboration, "MET Corrections and Uncertainties for Run-II", https://twiki.cern.ch/twiki/bin/view/CMS/MissingETRun2Corrections. Restricted webpage. Accessed: 2020-10-30.

[171] CMS Collaboration, "CMS Luminosity Measurements for the 2016 Data Taking Period", CMS-PAS-LUM-17-001, 2017.

[172] CMS Collaboration, "CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s}$ = 13 TeV", CMS-PAS-LUM-17-004, 2018.

[173] CMS Collaboration, "CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s}$ = 13 TeV", CMS-PAS-LUM-18-002, 2019.

[174] J. Butterworth *et al.*, "PDF4LHC recommendations for LHC Run II", *J. Phys. G*, vol. 43, p. 023001, 2016.

[175] A. Kalogeropoulos and J. Alwall, "The SysCalc code: A tool to derive theoretical systematic uncertainties", 2018. arXiv:1801.08401 [hep-ph].

[176] CMS Collaboration, "hdamp for TuneCP5", https://gitlab.cern.ch/cms-gen/Tuning/merge_requests/11. Restricted webpage. Accessed: 2020-10-30.

[177] CMS Collaboration, "Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of $\mathrm{t\bar{t}}$ at $\sqrt{s} = 8$ and 13 TeV", CMS-PAS-TOP-16-021, 2016.

[178] CERN, "The HL-LHC project", https://hilumilhc.web.cern.ch/content/hl-lhc-project. Accessed: 2020-10-30.

[179] CMS Collaboration, "Summary table of samples produced for the 1 Billion campaign, with 25ns bunch-crossing", https://twiki.cern.ch/twiki/bin/viewauth/CMS/SummaryTable1G25ns. Restricted webpage. Accessed: 2020-10-30.

Danksagung

An erster Stelle möchte ich mich bei Prof. Dr. Thomas Müller dafür bedanken, dass er mich bereits für meine Masterarbeit in seine Arbeitsgruppe aufgenommen hat und mir im Anschluss daran die Möglichkeit zur Promotion geboten hat. Ich bin ihm insbesondere dafür dankbar, dass er mir die Teilnahme an Schulen, Workshops und Konferenzen im Laufe der Promotion ermöglicht hat, sowie für den sechsmonatigen Aufenthalt am CERN. Besonders hervorheben möchte ich das großzügige Zeitangebot und die Freiheiten, die mir in der selbstständigen Forschung geboten wurden.

Bei Prof. Dr. Günter Quast möchte ich mich nicht nur für die Übernahme des Korreferats bedanken, sondern auch für die vielen interessanten Diskussionen und die Integration in seine Arbeitsgruppe und den damit verbundenen Blick über den Tellerrand in das moderne Computing bei einem Experiment in der Hochenergiephysik.

Diese Arbeit profitierte signifikant von den Computing-Resourcen, welche vom bwForCluster NEMO an der Universität Freiburg zur Verfügung gestellt wurden.

Mein besoderer Dank gilt meinen beiden hervorragenden Betreuern Dr. Thorsten Chwalek und Dr. Nils Faltermann, die mir zu jeder Situation mit Rat und Tat, sowie ausreichend Humor zur Seite standen. Ohne ihre fachkundigen Empfehlungen und richtungsweisenden Beiträge wäre diese Arbeit nicht möglich gewesen.

Des weiteren möchte ich mich bei allen aktuellen und ehemaligen Mitgliedern meiner Arbeitsgruppe für die konstruktive und angenehme Arbeitsatmosphäre, sowie die gegenseitige Unterstützung bedanken. Dr. Nils Faltermann, Dr. Daniela Schäfer und Marco Link gilt ein besonderer Dank für das Korrekturlesen der Arbeit. Auch den weiteren Mitgliedern des Instituts für experimentelle Teilchenphysik möchte ich für das angenehme Miteinander und die arbeitsgruppenübergreifende Unterstützung danken. Hervorzuheben ist hierbei Frau Bräunling für die Übernahme von unzähligen organisatorischen Angelegenheiten, sowie das Admin-Team für die exzellente Bereitstellung und Wartung der IT-Infrastruktur.

Abschließend möchte ich mich bei meiner Familie und meinen Freunden für die Unterstützung vom ersten Studientag bis zum Abschluss der Promotion bedanken. Im Besonderen geht dieser Dank an meine Eltern und meine Freundin Rebecca, die mit ihren stets aufmunternden Worten jeglichen Abschwung meiner Motivation zunichte machten.